

6
25



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE INGENIERIA

**PRINCIPIOS DE ANALISIS DE RIESGO
EN INGENIERIA PETROLERA**

T E S I S

QUE PARA OBTENER EL TITULO DE:

INGENIERO PETROLERO

P R E S E N T A:

Norma Espinoza de los Monteros Rebolledo



MEXICO, D. F.

1993.

**TESIS CON
FALLA DE ORIGEN**



Universidad Nacional
Autónoma de México



UNAM – Dirección General de Bibliotecas Tesis Digitales Restricciones de uso

DERECHOS RESERVADOS © PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

Todo el material contenido en esta tesis está protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.

LISTA DE CONTENIDO

	Página
RESUMEN	i
LISTA DE FIGURAS	ii
LISTA DE TABLAS	iii
I. INTRODUCCION	
II. METODOS DE ANALISIS DE RIESGO EN INGENIERIA PETROLERA	2
II.1 Métodos para Estimar Probabilidades en Exploración Petrolera	7
II.2 Estimación de Probabilidades de un Descubrimiento de Aceite	11
II.3 Modelos de análisis de Riesgo basados en las Últimas Distribuciones de Reserva	18
II.4 Probabilidad de Resultados en Programas de Perforación Multipozo	24
II.4.1 Análisis basado en la Condición de Eventos Dependientes	
II.4.2 Análisis basado en la Condición de Eventos Independientes	
II.4.3 Análisis de un Ejercicio que Incluya Varios Niveles de Reserva en un Programa Multipozo	
II.4.4 Tres Niveles Estimados de Riesgo	
III. TECNICA DE SIMULACION EN ANALISIS DE RIESGO	57
III.1 La Lógica de Simulación	57
III.2 Los Mecanismos de Análisis de Simulación	62
III.2.1 Definiendo las Variables (paso 1)	
III.2.2 Definiendo el Modelo de Análisis (paso 2)	
III.2.3 Clasificando las Variables en Grupos (paso 3)	
III.2.4 Definiendo Distribuciones para las Incógnitas, Variables Aleatorias (paso 4)	
III.2.5 Ejecutar los Pasos de Simulación (paso 5)	
III.2.6 Cálculo de EMV y preparación del Despliegue de Gráficas (paso 6)	
IV. EI USO DE COMPUTADORAS PARA ANALISIS DE SIMULACION	122
IV.1 Procedimientos para Distribuciones de Lectura en la Computadora (parte 1)	123
IV.2 Procedimientos de lectura en los Pasos de Computación, Valores Numéricos para parámetros conocidos y relaciones de dependencia parcial (parte 2)	131
IV.3 Generador de Números Aleatorios (parte 3)	131
IV.4 Programando los Pasos de Simulación (parte 4)	131
IV.5 Procedimientos para Datos de Salida (parte 5)	132
IV.6 Defensa Filosófica de la Simulación	133
V. EJEMPLOS Y APLICACIONES	136
CONCLUSIONES	166
NOMENCLATURA	167
REFERENCIAS	168

RESUMEN

En este trabajo se presentan, algunos aspectos en el Análisis de Decisiones cuyos métodos o técnicas ayudan, al que toma la decisión, a escoger la más adecuada bajo condiciones de incertidumbre.

El trabajo se basa en lo establecido por el Dr. Paul D. Mewendorp¹ relacionado con el "Análisis de Decisión en Exploración Petrolera" y pretende mostrar la importancia de considerar el riesgo y la incertidumbre en ciertas áreas de Ingeniería Petrolera.

El capítulo II cubre lo relacionado a los métodos para la estimación de las probabilidades de un descubrimiento de aceite, así como el análisis de los resultados bajo las condiciones de eventos dependientes e independientes.

En el capítulo III se presenta la lógica de la simulación y los mecanismos utilizados para su análisis.

El capítulo IV proporciona los procedimientos para realizar el análisis de simulación utilizando computadores.

Finalmente, en el capítulo V se presenta también, de manera ejemplificada, el procedimiento de un estudio de simulación que da una clara idea de los pasos necesarios para llevar a cabo este tipo de análisis económico de un proyecto con riesgo o incertidumbre.

LISTA DE FIGURAS

	Página
Fig. 1 Arbol de decisión para el caso de eventos dependientes.	15
Fig. 2 Arbol de decisión para el caso de eventos independientes.	16
Fig. 3 Gráficas de distribución de frecuencia de reservas a varios puntos en el tiempo.	20
Fig. 4 Posibles formas de la distribución de reservas por campo en una cuenca.	25
Fig. 5 Gráfica de probabilidad binomial para tener al menos tres descubrimientos.	38
Fig. 6 Método combinando probabilidades del programa multipozo y reservas descubiertas	43
Fig. 7 Gráfica de frecuencia acumulativa de distribución de reserva para varios descubrimientos.	51
Fig. 8 Esquema del análisis de simulación de la ecuación 6.	59
Fig. 9 Método gráfico para "forzar" una distribución lognormal a través de valores límites.	68
Fig. 10 Gráfica de distribución triangular, caso I.	72
Fig. 11 Gráfica de distribución triangular, caso II.	73
Fig. 12 Aproximación de una distribución triangular para una variable aleatoria.	75
Fig. 13 Distribución de la probabilidad discreta de una variable aleatoria.	79
Fig. 14 Distribución de la frecuencia acumulativa de la probabilidad discreta.	80
Fig. 15 Representaciones no válidas de distribuciones de frecuencia acumulativas.	81
Fig. 16 Histograma de la distribución de valores de la variable aleatoria X.	84
Fig. 17 Gráfica de frecuencia acumulativa del histograma de la Fig. 16.	85
Fig. 18 Distribución de una variable aleatoria y su gráfica de frecuencia acumulativa.	87
Fig. 19 Esquema de uso incorrecto del mismo número aleatorio para probar más de una distribución.	89
Fig. 20 Ejemplo de dos variables aleatorias sin relación.	92
Fig. 21 Ejemplo de dos variables aleatorias completamente dependientes.	93
Fig. 22 Ejemplo de dos variables aleatorias con una dependencia parcial.	94
Fig. 23 Ejemplos de varios tipos de variabilidad con regiones limitadas.	99
Fig. 24 Representación tridimensional de la dependencia parcial de la Fig. 22.	100
Fig. 25 Gráfica de espesor neto productivo y potencial inicial (IP).	106
Fig. 26 Gráfica de frecuencia acumulativa de la variable normalizada YNORM.	107

LISTA DE FIGURAS (cont.)

	Página
Fig. 27 Gráfica de frecuencia acumulativa de la variable aleatoria X.	109
Fig. 28 Ejemplo de estimación de la línea de tendencia y asignación de un rango a la misma.	111
Fig. 29 Gráfica de datos X-Y cuyos valores más probables están entre X_{min} y X_{max} .	112
Fig. 30 Gráfica de dos variables aleatorias parcialmente dependientes (ajuste mínimos cuadrados).	114
Fig. 31 Gráfica de dos variables aleatorias parcialmente dependientes X y Y.	116
Fig. 32 Histograma de frecuencia relativa del análisis de simulación de un prospecto.	119
Fig. 33 Gráfica de frecuencia acumulativa de la distribución de la utilidad de la Fig. 32.	120
Fig. 34 Ejemplo de una distribución uniforme.	126
Fig. 35 Gráfica de frecuencia acumulativa de la variable aleatoria X.	129
Fig. 36 Esquema de una decisión parcial con tres componentes básicos de un análisis de simulación.	137
Fig. 37 Dibujo esquemático de las partes principales del modelo de simulación de la Fig. 36.	140
Fig. 38 Distribuciones y gráficas de dependencia parcial entre espesor de arenas e IP por pozo.	142
Fig. 39 Correlación para determinar el número de pozos de desarrollo secos.	143
Fig. 40 Árbol de decisión parcial dado un descubrimiento.	145
Fig. 41 Ejemplo de una distribución triangular simétrica.	146
Fig. 42 Gráfica de frecuencia acumulativa para dos distribuciones, pesimista y optimista.	151
Fig. 43 Gráfica de VPN esperado contra la probabilidad de un descubrimiento de aceite.	153
Fig. 44 Curvas de declinación exponencial.	156
Fig. 45 Correlación de IP/pozo contra reservas a partir de esquemas de declinación.	157
Fig. 46 Mapa esquemático de una cuenca hipotética.	160
Fig. 47 Diagrama de flujo del modelo de simulación de la reserva de una cuenca.	162

LISTA DE TABLAS

	Página
Tabla 1 Reporte secuencial de la exploración del Canal de Santa Bárbara.	3
Tabla 2 Métodos para análisis de programas de perforación multipozo basado en eventos dependientes e independientes.	27
Tabla 3 Ejemplo de diez formas de alcanzar dos descubrimientos de aceite en cinco intentos.	30
Tabla 4 Cálculo del número de pozos correspondientes a cada categoría.	32
Tabla 5 Cálculo de probabilidad binomial para varios eventos.	36
Tabla 6 Probabilidad binomial de tener al menos tres descubrimientos como una función de "n".	37
Tabla 7 Tabulación de varios niveles de reservas posibles.	40
Tabla 8 Cálculo del tamaño mínimo del campo (R_{min}) para varios descubrimientos.	49
Tabla 9 Cálculo de la probabilidad de alcanzar el VPN.	50
Tabla 10 Asignación de valores para tres variables independientes.	53
Tabla 11 Ejemplo de una distribución dentro de un rango de valores de utilidad.	54
Tabla 12 Ejemplo de datos de probabilidad de ocurrencia.	55
Tabla 13 Características de la frecuencia acumulativa para diferentes distribuciones de probabilidad	77
Tabla 14 Ejemplo de una variable aleatoria teniendo valores discretos.	78
Tabla 15 Preparación de datos para dependencias parciales entre variables aleatorias.	95
Tabla 16 Pasos para probar valores de dos variables parcialmente dependientes.	96
Tabla 17 Pasos para probar valores de dos variables parcialmente dependientes en una computadora	97
Tabla 18 Tabla de espesor neto productivo y potencial inicial.	105
Tabla 19 Tabla de frecuencia acumulativa de la variable aleatoria X.	108
Tabla 20 Resultados de un análisis de simulación.	115
Tabla 21 Tabulación de resultados de los valores esperados de la Tabla 20.	118
Tabla 22 Resultados de las distribuciones del VPN.	150
Tabla 23 Código de símbolos usados en el modelo de la reserva de la cuenca de la Fig. 46.	159

I. INTRODUCCION

Tradicionalmente, los métodos de análisis económico establecen medidas de rentabilidad sin riesgo, los cuales son solo indicadores para conocer las ganancias de una inversión dada.

Se sabe que no existe, probablemente, un solo parámetro de rentabilidad que considere todos los aspectos de un proyecto de inversión. Por lo tanto, cada compañía debe seleccionar el parámetro de rentabilidad que más represente a sus intereses.

La mayoría de las decisiones que se toman en las empresas, se hacen bajo condiciones de incertidumbre, lo que implica que hay por lo menos dos posibles resultados que pudiesen ocurrir de una situación particular. La teoría de Análisis de Decisión establece metodologías que permiten evaluar, de una manera más fundamentada, el grado de riesgo e incertidumbre involucrado en proyectos de la Industria Petrolera.

En realidad el proceso de Análisis de Decisión conduce a visualizar, de forma más explícita, los posibles resultados que puedan ocurrir de un proyecto dado.

Finalmente, es importante señalar que este tipo de teoría no es sino solamente un suplemento en la toma de decisiones y no reemplaza o intenta hacerlo el buen juicio profesional del que toma las decisiones.

II. METODOS DE ANALISIS DE RIESGO EN EXPLORACION PETROLERA

Teniendo como patrones los principios de probabilidad y estadística se pueden aplicar estos principios al análisis de riesgo en Exploración Petrolera. En este contexto, el riesgo es un término un tanto engañoso, de tal manera que se tiene la necesidad de cuantificar o estimar muchos tipos de riesgos:

- Riesgo de perforar un pozo seco exploratorio o de desarrollo.
- Riesgo político, riesgo económico.
- Riesgo relacionado a los precios futuros del aceite y gas.
- Riesgo de daño a instalaciones costa fuera por tormenta.
- Riesgo de decidir si un descubrimiento es lo suficientemente grande para recuperar el costo inicial de la exploración.
- Riesgo de un número mínimo de descubrimientos en un programa de exploración múltiple de pozos.
- Riesgo ambiental.
- Riesgo de ruina.
- Etc.

En este capítulo se discutirán algunos de los métodos usados para evaluar estos tipos de riesgo.

Los problemas relativos a tomar decisiones de exploración bajo condiciones de riesgo e incertidumbre, existen desde que el petróleo comenzó como negocio. Anteriormente se han hecho pruebas para definir el riesgo de una manera informal y usualmente envuelve adjetivos más que probabilidades.

Hace algunos años, la nueva disciplina de teoría de decisión estadística (análisis de decisión) comenzó a emerger y los que toman decisión de exploración comenzaron a tomar nota de los beneficios potenciales del análisis de decisión. El problema involucrado en análisis de decisión fue, y es todavía, obtener las probabilidades requeridas para resolver un árbol de decisión o un cálculo del valor medio esperado (EMV)? El análisis de riesgo es ciertamente el más débil eslabón (un mínimo contexto en la exploración del aceite).

Pero ¿cuál es la alternativa? ¿Ignorar el riesgo?. Como ejemplo, en 1968 la industria comenzó a perforar en el Canal de Santa Barbara en las costas de California. Al tiempo de la venta inicial en el consenso general parece que la industria tuvo la certeza virtual del potencial del petróleo en el Canal.

Muchas de las estructuras productivas de aceite costa dentro, fuera de las aguas federales del Canal, proporcionaron razonablemente un buen control ecológico. Pero cuando perforaron la historia cambió, como se resume en la serie de reportes dados en la tabla 1. Estos comentarios no se dan para hacer un juicio sobre su comercio reportado sino para ilustrar que la "cosa segura" después puede no ser tan segura. Falta reconocer que el riesgo puede ser desastre cuando se juega "alto riesgo" de Exploración Petrolera.

Tabla 1.

Reporte secuencial de la exploración del canal de Santa Barbara

- Diciembre 11, 1967 - Reservas bajo las aguas federales (Canal de Santa Barbara) han sido estimadas mayores que las de la cuenca de los Angeles (OGJ).
- Enero 22, 1968 - Por primera vez las compañías tienen la oportunidad de licitar en una larga oferta en zonas de alto potencial de aceite La venta involucra el área más explorada en aguas profundas jamás ofertadas ... El canal es mucho más conocido que el Golfo de México por ejemplo. Se cree que existe una vasta cantidad de aceite en el Canal.
- Marzo, 1968 - La licitación de Santa Barbara rompe todos los records, esto debido en parte a los amplios conocimientos del área la cual produce de pozos cercanos a la costa (PE).
- Mayo, 1968 - El Canal de Santa Barbara, California, con sus aguas profundas, altos costos y super potencial no va a decepcionar a nadie (PE).
- Mayo 27, 1968 - Se encuentra aceite profundo (OGJ).
- Junio 17, 1968 - Tercer agujero seco adicional a la lista de las fallas del Canal de Santa Barbara (OGJ).
- Agosto 12, 1968 - Después de \$602.7 millones en gastos y \$10.5 millones para perforación de 12 pozos en el canal - se tiene:
- 1 descubrimiento de multizonas,
 - 1 descubrimiento indicado, 700 pies de arena con aceite
 - 4 pozos anunciados oficialmente secos
 - 4 abandonados - no son comerciales
- Otro pozo suspendido después de dificultades con el agujero... si los próximos 6 meses de perforación no se producen mejores resultados que los 6 meses anteriores algunas compañías desearían que las ventas federales tuvieran una garantía de regreso de capital (OGJ).
- Agosto 19, 1968 - Abandono de más pozos, taponados (OGJ).
- Agosto 26, 1968 - Otro pozo abandonado ... quinceavo pozo abandonado o suspendido (OGJ).
- Septiembre 9, 1968 - Se han perforado 5 pozos más en el Canal de Santa Barbara todos negativos. Aunque los resultados son decepcionantes, la rápida perforación continúa (OGJ).
- Septiembre 23, 1968 - La sorpresa mayor ha sido la escases de arenas productivas. Sólo dos de quince objetivos perforados fueron encontrados comerciales (OGJ).
- Septiembre 30, 1968 - En resumen, pozos taponados, abandonados sin reserva comercial de aceite.

No hay que creer que se pueden ignorar los elementos de riesgo. Sin embargo, a través del análisis de riesgo la alternativa de ignorarlo es insostenible. Se debe aceptar las realidades de los riesgos de exploración y adoptar un enfoque positivo para aprender todo acerca del análisis de riesgo. Esto requiere que se pruebe y evalúe de una manera experta el riesgo y comunicación para hacer de una manera clara y concisa, la decisión.

No existen fórmulas que reditúan en probabilidades estimadas. Pero los métodos discutidos en este capítulo, junto con el importante método nuevo llamado simulación (capítulo siguiente) provee las bases para usar análisis de decisión más efectivamente en decisiones de exploración.

Primero se dará un breve resumen del problema fundamental y se establecerá el análisis de riesgo involucrado en exploración. El resto del capítulo se hará un compendio de los métodos de análisis de riesgo que puede ser usado en exploración petrolera. Más de estos métodos se documentan en literatura de ingeniería petrolera. El propósito es discutir la fuerza y debilidad de cada método usado.

Hay que recordar que estos métodos son muy absolutos en análisis de riesgo de exploración petrolera. El valor de estos métodos es el que puede entrar en el aumento del análisis de sensibilidad. Soluciones a las preguntas "que es" son muy importantes y pueden proveer las bases para una decisión racional, siempre a través de algunas de las probabilidades específicas.

EL PROBLEMA FUNDAMENTAL

Desde que una inversión aceptable tiene una oportunidad razonablemente buena de alcanzar un beneficio, los resultados internos son expresados en niveles de probabilidad, más que reservas tal como $2 \text{ bls}/\text{pie}^3$, $3 \text{ bls}/\text{pie}^3$, etc. Pero por supuesto, la incertidumbre de obtener buenos niveles posibles de reserva, en muchos casos más, es la incógnita mayor en la decisión.

Como se consideran soluciones al problema se tiene que tener en mente ciertas características que son únicas en el proceso de decisión de Exploración Petrolera.

1.

Con respecto a las probabilidades del nivel de reserva, no se pueden describir los procesos que originalmente generaron la distribución de acumulaciones de aceite. Esta es una desventaja porque probablemente nunca se pueda desarrollar un modelo exacto probabilístico (ecuación) que sirva análogamente al proceso de exploración.

En teoría, el primer paso para solucionar algún problema de probabilidad, tratado como fenómeno de oportunidad, es determinar el proceso fundamental; ésto es, generando los resultados de interés. Una vez que se define el proceso, los matemáticos usualmente pueden desarrollar una ecuación describiendo el proceso que puede usarse en el cálculo de probabilidades en pruebas futuras. Quizás la mejor ilustración de este enfoque está en la existencia del control de calidad de un producto fabricado por una máquina automática.

Suponga que el operario de la máquina ha observado que ésta produce periódicamente una pieza defectuosa, y que la proporción de ocurrencia de estos defectos es de una en cada mil. Suponga que él tiene que fomentar lo observado, los defectos al ocurrir parecen no seguir un

patrón. Su problema es poder calcular la probabilidad de decir, hay menos de cinco piezas defectuosas en una cantidad de 10 000 piezas.

El matemático probablemente concluya con esta evidencia, que la ocurrencia de defectos de la máquina es equivalente a un proceso de Bernoulli: solamente dos resultados; la probabilidad de un defecto parece ser constante en el tiempo, y toda la ocurrencia aleatoria de defectos significa un proceso de Bernoulli. De estas conclusiones se escribe la ecuación binomial de probabilidad como una exacta analogía de la oportunidad del fenómeno y soluciona la ecuación usando valores de $p=0.001$, $n= 10000$ y $x < 5$ para contestar la pregunta del operario de la máquina.

En la exploración del aceite no se sabe el proceso fundamental que controla el origen, acumulación, migración, y entrapamiento de reservas de aceite. Y probablemente nunca se sabra esto. Así que no se puede decir que las distribuciones de reserva en una cuenca son el resultado de la generación de un proceso "lognormal", o un proceso de Poisson, o algún otro proceso para esta materia. El efecto neto es determinar probabilidades de "no perforar" varios niveles de reservas en un prospecto sin tener algún modelo a usar como análogo.

Cualquier modelo "teórico" se desarrollará solamente como una aproximación a lo que se piensa o percibe como el proceso fundamental -pero probablemente nunca tendrá una exacta probabilidad análoga. Esta es la razón principal por la que no hay "reglas de" fórmula para probabilidades de exploración.

2.

Como una relación producto del punto 1. El perforar una serie de pozos es una secuencia de acontecimientos dependientes basados en un proceso "probando con reemplazo". Los pocos modelos que se han propuesto en análisis de riesgo de exploración (como por ejemplo la ecuación binomial) usualmente están basados en la noción de acontecimientos independientes. Aunque sucesivamente los pozos son una serie de eventos dependientes, las probabilidades para varios niveles de reservas en pozos sucesivos continuamente cambian. Esto hace el problema mucho más complejo, y los modelos que aproximadamente representan este proceso serán mucho más complicados que modelos basados en pruebas independientes.

3.

La probabilidad estimada tiene a menudo que basarse en muy pequeños o ningún dato estadístico o experiencias. Datos adicionales en exploración usualmente provienen de pozos o sísmica, y normalmente no se pueden proporcionar decisiones hasta que haya una cantidad suficiente de datos sobre que basar la probabilidad a estimar. En otras industrias frecuentemente es posible obtener datos estadísticos relativamente baratos de muestras anteriores, decisiones mayores de inversión (encuestas de opinión, estudios, etc). Pero en este negocio se tiene más tiempo con probabilidades basadas en información deficiente.

4.

Con respecto a los factores económicos y la predicción de probabilidades se tiene una situación igualmente compleja de tratar. Como una evidencia nunca antes experimentada, la inestabilidad del precio del crudo desde 1973 los factores económicos relacionados con el desarrollo y exploración de un descubrimiento pueden representar serios problemas. No es más válido que suponer que los futuros precios del crudo pueden predecirse con certeza, o

que sólo fluctúan debido a la exigencia de la demanda del mercado local y patrones existentes.

Ahora es claro, estos precios del crudo son un tema aún mucho más largo, por condiciones del juego mundial de factores políticos y económicos. Y no se han tenido precedentes para pronosticar futuros precios de crudo en tal compleja red de impulsos e influencias en el sistema. Se presenta otra situación, tener que pronosticar o calcular las probabilidades sin la utilización del beneficio de un modelo realista o análogo.

Los efectos de la inflación, escases de acero para plataformas, tuberías de revestimiento para pozos, etc. se tienen que considerar cuando se determine el valor de un descubrimiento. Estos factores son también parte de un patrón mucho más complejo, de comercio en el mundo y condiciones económicas. El porvenir es desconocido y la predicción de costos futuros puede también ser extremadamente difícil.

Todos estos comentarios probablemente señalan el fin de la simple evaluación de un prospecto de perforación en que el "sistema de evaluación" terminaba en la salida de una batería a la venta de crudo, en la que es fácil pronosticar el precio. El "sistema de evaluación" ahora tiene la necesidad de considerar muchos más parámetros económicos y políticos.

5.

La geología es una ciencia muy abstracta. Algunos exploradores tienen que recrear pruebas en su mente como la existencia de antiguos ríos, deltas, el retroceso de mares, cuencas submarinas, la erosión ocurrida, etc. en la consideración de un proyecto exploratorio. Pero al final de su análisis existe la pregunta al cuantificar sus (intuitivas?) sensaciones del prospecto en términos de probabilidades. No es una tarea fácil, decirlo.

En un extremo la visión de geología no puede ser cuantificada en probabilidades. Desde otro punto de vista no puede hacerse por intuición -la exploración es un negocio de decisión que se hace en base a manejar datos de negocios, criterios probabilísticos y probabilidades de cuantificar riesgo e incertidumbre. Probablemente cada extremo es parcialmente válido, y se tiene que tratar de hallar una balanza entre los extremos.

Un estudiante imaginó esto como un cambio el cuál puede ser movido de ser necesario para lograr una balanza realista entre la "geología pura" de un prospecto y la necesidad de expresar un valor en términos de beneficio y probabilidades. No es fácil cuantificar la incertidumbre relativa a los factores abstractos que un explorador puede considerar en un análisis del prospecto. Sin embargo se tiene que probar si se puede hacer efectivo el uso del análisis de decisión en exploración.

Estas consideraciones sin duda hacen la tarea de estimar probabilidades en exploración más complejas y difíciles. Sin embargo, los métodos discutidos en el resto del capítulo pueden a menudo usarse eficazmente o por lo menos empiezan a evaluar sensaciones desconocidas. La técnica de simulación es discutida en el capítulo siguiente que provee otro paso significativo, analizando los efectos de riesgo e incertidumbre.

Aunque la lista de características arriba sugerida todavía tiene muchas fallas, se pueden usar con experiencia, para hacer mejores decisiones en exploración.

II.1 Métodos para Estimar Probabilidades en Exploración Petrolera

Estimación Subjetiva de Probabilidades

La estimación subjetiva de probabilidad es un modo común de expresar el grado de riesgo e incertidumbre relacionando parámetros inciertos en exploración. Los cálculos representan una opinión personal de la probabilidad de que un resultado ocurra. La opinión puede basarse en datos estadísticos disponibles, o puede simplemente basarse en la opinión y sensaciones del analizador.

Ejemplo: "Se han estudiado todos los datos geológicos disponibles del pozo y se construyó este mapa de isopacas. Un pozo perforado en el centro de la sección 14 deberá encontrar 150 pies de zona productiva y se estima una ganancia neta de \$1,200,000.

Se estima la oportunidad de que esto ocurra en 60 por ciento ".

Tal declaración probablemente representa su "grado de credibilidad", o la "veracidad de interpretación" que se asocia con el prospecto.

Pero un juicio de probabilidad de este tipo en realidad, ¿qué significa? "Por ejemplo, se tiene un 60 por ciento de posibilidad de una zona productiva de 150 pies y un 40 por ciento de posibilidad de que la zona no sea productiva? O un 40 por ciento de probabilidad de una zona productiva pequeña menor a dichos 145 pies? "Se tiene un 60 por ciento de oportunidad de una ganancia de \$1,200,000? Y así, cuál nivel de ganancia ocurrirá en el otro 40 por ciento del tiempo?"

Estas preguntas señalan un punto del problema llamado "estimación puntual" de riesgo. Con una apropiada calificación puede decir la oportunidad de un resultado (tal como lograr una ganancia de \$1,200,000 y/o 150 pies de zona productiva) pero no le dice nada al que decide acerca de los resultados que pueden ocurrir. Un sólo valor estimado de riesgo acerca de cuántos factores pueden ocurrir no da mucha información acerca del rango de otros posibles resultados. Otras implicaciones subjetivas de probabilidad estimada incluyen:

- a). Subjetivas estimaciones de probabilidad se basan en juicios individuales. Dos personas, dan los mismos datos fundamentales, pero no alcanzan las mismas conclusiones en cuanto al grado de riesgo. La parcialidad personal, emociones y la experiencia influyen en juicios subjetivos.
- b). Una probabilidad subjetiva estimada del tipo ilustrado en el ejemplo es usualmente hecha al final del análisis y significa representar la probabilidad de cómo muchos inciertos parámetros pueden en realidad ocurrir para el nivel declarado de probabilidad. Para algunas personas una probabilidad estimada pueda ser más difícil de acceder. Por ejemplo, suponga que el resultado favorable de una ganancia de \$1,200,000 depende de la ocurrencia simultánea de cuatro factores independientes, cada uno de ellos sujeto a un grado de incertidumbre.

Los factores pueden ser espesor neto productivo 150 pies, los costos de perforación que tienen que predecirse en el análisis, etc. Si cada uno de estos cuatro factores tiene un 0.60 de probabilidad de tener sólo los valores usados para calcular la ganancia de \$1,200,000 toda la probabilidad de que ocurran los cuatro simultáneamente para producir la ganancia no

es 0.60, sino $(0.6)^4 = 0.1296$ ¿Es lo que realmente significa la recomendación del ejemplo? Esto es incierto.

El punto aquí es si hay muchas variables aleatorias separadas que puedan ser subjetivamente bastante difícil de estimar la probabilidad de que esos valores favorables de todas las variables puedan ocurrir simultáneamente.

c). Grupos subjetivos de probabilidad a veces son predispuestos por las sensaciones del analizador de las consecuencias del resultado. Por ejemplo, si el potencial de ganancia o pérdida monetaria es muy grande las preferencias y actitudes acerca del dinero pueden influir al estimar la probabilidad de que el evento ocurra. Las complejidades del sistema de valor y los juicios acerca de probabilidades pueden atrapar al analizador aún antes de suceder.

La literatura relaciona a la psicología al hacer los juicios que contienen ejemplos numéricos de como los sistemas de valor pueden influir en estimar la probabilidad. La extensión posible, de probabilidades subjetivas estimadas tiene que hacerse sin consideración de los resultados monetarios consecuencia del resultado.

A pesar de las limitaciones obvias de la estimación subjetiva de probabilidad probablemente se tendrán que usar probabilidades subjetivas en el extenso análisis de riesgo de exploración - simplemente porque no se puede tener alternativa para determinar las probabilidades. Este puede ser el mejor camino; sin embargo, aplicar subjetividad al análisis de la recomendación del ejemplo (por ejemplo, simulación), y se tiene que esforzar continuamente para mejorar las destrezas al hacer estimaciones subjetivas de probabilidad. Esto requerirá que se sea consciente de las inferencias y limitaciones subjetivas de probabilidad siempre que se tenga que usarlas.

Decisión Arbitraria "Mínima" para cubrir Incertidumbre

En algunas ocasiones el riesgo es tratado simplemente para elevar los "mínimos" standares requeridos de no-riesgo para aceptar un proyecto.

Ejemplo: Una compañía puede ganar 10% en su inversión de capital, pero arbitrariamente especifica una tasa de retorno mínima para perforar prospectos de 25%. La lógica es que los prospectos que son aceptados y que resultan en terminaciones exitosas puede ser bastante provechoso llevar la balanza con un 10% de tasa de retorno para el programa total a perforar.

Por esta razón el enfoque es que se intenta proteger contra la incertidumbre de aumentar el ritmo de "cortes" a la inversión. Un procedimiento así, presumiblemente refleja la necesidad de tener el retorno proporcionado al grado de riesgo. El razonamiento está reflejado a un punto, pero la debilidad es que el ritmo económico mínimo arbitrario (no-riesgo), no considera los varios niveles de riesgo entre inversiones.

Quién dice, por ejemplo, que un prospecto de perforación que tiene un alto riesgo con una tasa de retorno del 27%, sea exitosamente mejor que un prospecto con 23% de tasa de retorno con alta oportunidad de éxito? Si la mínima tasa de retorno aceptable de no-riesgo es 25% el prospecto debió haber sido ejecutado cuando, en realidad, pudo haber tenido más inversión deseable en plazos

con el propósito de tratar de balancear los objetivos de maximizar ganancias y minimizar riesgo e incertidumbre.

Otra debilidad al usar mínimos arbitrarios es el hecho desafortunado, que muchos los han usado para regular el flujo de capital de perforación. Cuando el capital de perforación es escaso algunos mínimos son elevados así como más selectivos en la elección del número de prospectos que puede ser aceptado. O si el período de algunos mínimos gastos es alargado así más prospectos son seleccionados.

Esto, por supuesto, acaba con el propósito de emplear los mínimos arbitrarios en el primer plano. Variando la mínima tasa de retorno regular de no-riesgo, el flujo del capital evidentemente conduce a estrategias inconsistentes de decisión -proyectos que fueron rechazados anteriormente ahora se aceptan, o proyectos rechazados ahora pueden ser aceptados después.

En la balanza final este enfoque incorpora riesgo e incertidumbre en el proceso de decisión, haciendo un sustituto pobre para un cuidadoso y cuantitativo análisis de riesgo.

Modificaciones a Tasas de Retorno para Riesgos

Las probabilidades no pueden usarse en la definición estricta de la tasa de retorno. Sin embargo una aproximación de los efectos de riesgo pueden ser conseguidos calculando una "psuedo" inversión para la tasa de retorno calculada. El enfoque es usado por varias compañías como un medio parcial de contabilidad para la probabilidad de un agujero seco en una tasa normal de retorno.

Relaciones Beneficio/Riesgo permitidas en Pozos Secos

Algunos exploradores sienten que una medida válida del grado relativo de riesgo en perforación de prospectos exploratorios es el parámetro definido como *hoyas secos permitidos*. Para este enfoque el analizador calcula una ventaja estimada VPN como si fuera exitoso. Este VPN es dividido entre el costo de un pozo exploratorio seco. El resultado ahora es el múltiplo de veces que la ganancia de un descubrimiento excede el costo de un agujero seco.

Ejemplo: Estime el VPN de existencia de una perspectiva productiva (calcular después de la consideración de impuestos, arrendamiento geofísico inicial y gastos de operación):

\$5,000,000

El costo de un agujero exploratorio es: \$500,000

Agujeros secos permitidos = $\$5,000,000/\$500,000=10$

En este ejemplo la ganancia potencial es 10 veces más grande que lo que cuesta un agujero seco. O se puede proporcionar tantas perforaciones como 10 agujeros secos se encuentren en el descubrimiento y aún se tiene una tasa de retorno igual a la tasa descontada usada en los cálculos del VPN.

El múltiplo que aquí ha sido definido como agujeros secos permitidos también se ha llamado Ganancia/Riesgo en algunas referencias.

El empleo de este enfoque no da alguna información acerca de la probabilidad del descubrimiento, pero ofrece una idea de cuántos pozos exploratorios pueden perforarse para localizar las reservas. Como tal, provee un relativo indicador de conveniencia en las compañías competitivas en áreas de exploración. Presumiblemente, la más cara es la mejor área múltiple.

Por ejemplo, suponga que se considera la exploración en dos cuencas diferentes, A y B, con aproximadamente las mismas oportunidades para un descubrimiento. Si los agujeros secos permitidos en la cuenca A son 10 y en la cuenca B son 35 lo posterior será una elección preferencial. En la cuenca B se puede perforar más pozos y encontrar un descubrimiento, todos estos factores son iguales para emprender la perforación, la oportunidad más alta de hallar por lo menos un descubrimiento.

El enfoque puede ser útil en ciertos casos, pero otras veces se enfrenta con varias limitaciones: no explica explícitamente la probabilidad de descubrimiento y no provee una decisión específica "hacer/no hacer". La segunda limitación viene del hecho que este método no dice que los agujeros secos permitidos tienen que ser mayores en orden para lograr un nivel objetivo de probabilidad.

Técnicas Numéricas de Grado

Otro método para evaluar el riesgo es una técnica que emplea una escala graduada arbitraria. El resultado es otro indicador relativo de riesgo, que explica el grado de riesgo numérico usando probabilidad.

La técnica consiste primero en registrar todos los factores que indicarían una posible presencia de aceite en un prospecto. Ejemplos de algunos de estos factores son: proximidad a rocas generadoras, cierre estructural, permeabilidad adecuada, etc. El analizador considera cada factor en su prospecto particular y asigna una escala relativa, o grado al factor. La escala puede extenderse de uno a cuatro, con cuatro excelente y uno pobre.

Por ejemplo, si hay una indicación de una roca generadora en la proximidad cercana a la zona productiva puede asignarse un grado de 3 ó 4, si tiene indicaciones de permeabilidad muy baja puede asignarse un grado de solamente 1 ó 2 al factor de permeabilidad. Después de asignar un grado a varios o a todos los factores posibles (una de las referencias incluye una lista de factores que cubre cerca de seis páginas!), se asigna un factor de peso para esos factores que son particularmente importantes al tener una acumulación de aceite. Los valores son entonces agregados y comparados a prospectos alternativos. Al tener un resultado mayor (presumiblemente) es el prospecto preferido.

Este enfoque sin duda tiene una ventaja distinta, que forza a los exploradores a pensar acerca de cada factor individualmente, asignando más factores de riesgo al final del análisis en que todos los factores ocurrirán simultáneamente. Sin embargo, las limitaciones son también bastante evidentes. La escala numérica graduada es bastante arbitraria, como lo son también los factores de peso usados para incrementar el grado de algunos factores.

Después que se tienen hechos 20 ó 30 juicios arbitrarios acerca del grado para asignar algunos de los factores esotéricos relacionados con geología de petróleo, la pregunta es si no es más sencillo una conjetura justa a la respuesta y evitar toda molestia.

Las experiencias ganadas por técnicas numéricas graduadas se hacen bastante más oscuras por el nuevo método de análisis de riesgo llamado simulación (capítulo siguiente). Si un analizador tiene que asignar un grado de porosidad probablemente considerará las porosidades de pozos cercanos en la formación equivalente.

Esto involucrará mirar datos estadísticos, encontrar valores mínimos, máximos, de promedio, etc. Teniendo observados todos estos datos se tienen que asignar valores arbitrarios. "Por qué no, como una alternativa, se expresan los datos disponibles de estadística como una distribución de valores posibles de porosidad e incluye la distribución real en el análisis, bastante más arbitrario?". La simulación permite hacer esto, como se verá en el capítulo próximo.

II.2 Estimación de Probabilidades de un Descubrimiento de Aceite

La probabilidad de que en un pozo (o pozos) resulte un descubrimiento o terminación es, por supuesto, un eslabón importante al calcular valores esperados de prospectos de perforación. Es una de las probabilidades más difíciles de estimar y se tiene que recordar la importancia de la capacidad de expresar la sensibilidad de este parámetro, por último se espera graficar la ganancia de un valor.

Probablemente uno de los métodos más comúnmente empleados al estimar una probabilidad de descubrimiento es analizar la razón de éxito de un pozo exploratorio (o uno en desarrollo, buena razón de éxito se ha visto al perforar uno adicional dentro de un campo). La razón es analizada dividiendo el número de descubrimientos anteriores entre el número total de pozos exploratorios perforados en la cuenca. Se publican frecuentemente datos de razón de éxito en diarios comerciales y otras publicaciones profesionales de sociedad.

Los datos importantes de suposición se hacen cuando se usan los resultados de estadística de pozos pasados. Estimar la probabilidad de perforar el descubrimiento en el pozo siguiente es que la probabilidad de descubrimiento no cambie con el tiempo.

Esto significa más acontecimientos independientes y probar con un proceso de reemplazo. Es un efecto decir qué lo que ha sucedido en el pasado es un reflejo de lo que pasará en los pozos siguientes a ser perforados.

El problema no reconoce lo finito del espacio de muestra de los prospectos en una cuenca y el hecho es que una vez que una estructura (o prospecto) ha sido perforado es quitado del espacio de muestra de prospectos no perforados, para el pozo siguiente a perforar. El espacio de muestra cambia sucesivamente después de que cada pozo exploratorio ha sido perforado y esto resulta en que la probabilidad de descubrimiento continuamente cambie.

Las relaciones de éxito son sin duda las "primeras aproximaciones" útiles, pero se tiene que saber que el uso de relaciones de éxito significa un fenómeno probabilístico, esto estrictamente no se aplica al mundo-real de exploración petrolera. Ejemplo donde el uso de relaciones de éxito pasadas pueden guiar a desviaciones, incluye:

1.

Cincuenta pozos han sido terminados en el Campo A, con un promedio de reservas por pozo de 80,000 barriles. Quince pozos han sido secos y abandonados, resultando con una relación de éxito en la terminación por pozo de 77%. Es la oportunidad de terminar el pozo siguiente

77%? Si, "cuántos pozos más se pueden perforar con una razón de éxito de 77%?. Qué pasa cuando se desarrollan métodos desconociendo los límites?

2.

Doscientos treinta pozos exploratorios han sido perforados en la provincia geológica B y treinta y cinco (15.2%) han sido descubrimientos prósperos de campos nuevos. Es correcto suponer que los sobrantes para encontrar un campo nuevo con doscientos treinta pozos es 15.2%? Qué pasaría si hubiera solamente treinta y cinco campos del total en la provincia - cuántos pozos más serían requeridos para comprobar esto? Haga de las relaciones de éxito de exploración una distinción entre el total de campos encontrados y el total que serán éxito económico?

3.

Se descubrió el Campo ABC hace seis meses. Desde entonces dos pozos han sido exitosos, es la probabilidad de éxito para el pozo siguiente de desarrollo 1.0? (ésto significa un cierto acontecimiento, no hay riesgo en todo!)

Se tiene que recordar que no se puede perforar por siempre con una razón de éxito. Justamente porque en 10% de los pozos petroleros exploratorios anteriormente perforados se encontró aceite o gas necesariamente no se puede continuar perforando con la esperanza de que ese 10% de los pozos petroleros exploratorios futuros también resulten en descubrimientos. Después todos los prospectos productivos habrán sido encontrados y la probabilidad de descubrimiento en pozos subsiguientes será cero, a pesar de que se ha tenido relaciones de éxito históricas. Las relaciones de éxito pueden ser una aproximación útil de probabilidades de descubrimiento pero deben usarse con cautela!

Una observación calificada. Allí pueden existir ciertas áreas donde factores estratigráficos son tan complejos o inter-relacionados, en esencia, cada pozo es una entidad separada. Un posible ejemplo es el que juega el yacimiento Anadrako de N.W. Oklahoma. Cada pozo perforado es casi un acontecimiento independiente.

En estos casos las probabilidades pueden ser imparcialmente estables con el tiempo, en cada caso los datos pasados de frecuencia pueden proveer una base realista para el análisis de riesgo. Concluir que un juego es verdaderamente aleatorio requiere un análisis cuidadoso de estadística, sin embargo, y de la naturaleza del juego de exploración de aceite esta situación casual no ocurre muy frecuentemente.

Algunos exploradores prefieren usar un método de alternativa para probabilidades estimadas de descubrimiento. Racionalizaron que un descubrimiento próspero requiere que una trampa se presente y la posición indica la evidencia geológica y geofísica, de que la formación objetiva tiene que tener suficiente espesor y tiene que contener hidrocarburos que pueden ser fácilmente producidos en el pozo. Todos estos factores tienen que producir el descubrimiento próspero presente, y presuntoso cada uno de los factores que contribuye es independiente, la probabilidad de descubrimiento es el producto de las probabilidades de cada uno de los factores individuales a ocurrir. Esto es:

Probabilidad de un descubrimiento en un pozo.

$$= \left(\text{Probabilidad de trampa estructural} \right) \times \left(\text{Probabilidad de trampa en la posición proyectada por la sísmica} \right) \times \left(\text{Probabilidad de espesor productivo} \right) \times \left(\text{Probabilidad de hidrocarburos en la formación} \right) \quad (1)$$

Para mostrar cómo la ecuación 1 puede ser usada para suponer que se ha considerado un prospecto estructural se tienen los siguientes parámetros:

La probabilidad de una trampa de depósito = 0.9	De experiencia se tiene buena interpretación de éxito en el área sísmica.
Probabilidad de existencia de trampa en la posición indicada = 0.7	(Un factor sísmico veraz?)
Probabilidad de un espesor productor suficiente = 0.8	Posiblemente de extrapolación geológica con estructuras cercanas.
Probabilidad de hidrocarburo en la formación = 0.25	Quizás de analogía con campos cercanos o quizás supuesto?

Para estos factores se puede solucionar la ecuación 1 y producir una probabilidad de descubrimiento:

$$\text{Probabilidad de un Descubrimiento en un pozo exploratorio} = (0.9) \times (0.7) \times (0.8) \times (0.25) = 0.126$$

Conceptualmente, este enfoque tiene mérito. Uno de los problemas en aplicarlo, sin embargo, es determinar o estimar las relaciones en términos individuales en la parte correcta de la ecuación 1. Si están disponibles datos válidos no hay problema, pero el prospecto de una cuenca nueva, puede ser muy difícil de estimar la probabilidad para los cuatro factores. En estas situaciones se suponen cuatro valores juntos multiplicados lo cual es mejor a que se adivine la probabilidad del descubrimiento.

Otra situación que a veces surge es analizar una probabilidad de descubrimiento en un prospecto múltiple. Se supone un pozo exploratorio existente considerando que hay tres secciones separadas geológicamente las cuales pueden ser productivas, A, B y C. ¿Cuál es la probabilidad de que todas esas tres zonas serán productivas?

En ésta situación el análisis tiene que determinar primero si la ocurrencia de encontrar aceite en una zona dada es independiente o dependiente de que si o no las otras zonas son productivas. "Se requiere que se hagan preguntas tales como: Si la zona poco profunda A es productiva eso significa que las oportunidades de producción en las zonas más profundas B y C son mejores? "Si zonas A y B son secas es todavía probable que la zona C sea productiva? etc.

Si la probabilidad de ocurrencia en una zona es dependiente o no, de que las otras zonas sean productivas, el árbol de decisión muestra los resultados posibles de perforar tres formaciones en un

pozo, como se muestra en la figura 1. La barra sobre el símbolo es para un acontecimiento o probabilidad del acontecimiento (ejemplo, $P(\bar{A})$) significa que el acontecimiento no ocurre, y la probabilidad de que el acontecimiento no ocurra respectivamente.

Al solucionar el árbol, el analizador tiene que especificar o estimar siete términos de probabilidad -el conjunto de probabilidades en rectángulos.

Los términos calculados en cada una de las probabilidades del punto terminal son dados en la columna de la figura 1. La probabilidad que ninguna de las zonas sea productiva es simplemente la probabilidad del punto terminal 8. La probabilidad de que todas esas tres zonas sean productivas corresponde a la probabilidad de ocurrencia del punto terminal 1.

La probabilidad de existencia de dos o más zonas productivas es la suma de las probabilidades para puntos terminales 1, 2, 3 y 5. (Los cuatro términos de probabilidad se suman porque cada uno era un acontecimiento mutuamente exclusivo.)

Si la probabilidad de producción en una zona es independiente de sí o no las otras zonas son productivas, el análisis es mucho más sencillo y requiere que el analizador complete específicamente solamente tres probabilidades, en lugar de siete. El árbol para el caso de acontecimientos independientes se muestra en la figura 2. La interpretación y empleo del árbol son las mismas en cuanto al caso de acontecimientos dependientes. Las tres probabilidades que tiene que especificar son las probabilidades de hidrocarburos en cada zona individual.

Este enfoque para probabilidades analizadas en un prospecto multiple es relativamente rápido una vez que las probabilidades requeridas son especificadas. Pero esto es probablemente un ejemplo de donde el modelo de probabilidad es más sofisticado y muestra la habilidad de describir el prospecto que se ha visto.

Así, solucionar el caso dependiente mostrado en la figura 1 se tendría que especificar o calcular términos tales como:

$P(C|\bar{A}\bar{B})$ - La probabilidad condicional que la Zona C sea productiva dado que la Zona A es productiva y la Zona B no es-productiva.

$P(C|\bar{A}B)$ - La probabilidad condicional que la Zona C sea productiva dado que las Zonas A y B no son productivas. etc.

Y estos términos pueden ser bastante difíciles de especificar, particularmente en áreas nuevas de exploración donde hay poca información de perforación sobre que basar el juicio. En áreas que han sido extensamente perforadas se puede tener información suficiente para determinar estas probabilidades, para lo cuál los métodos de la figura 1 ó 2 pueden ser útiles para el caso.

Sin embargo otra técnica para determinar probabilidades de descubrimiento fue propuesta, por algunas investigaciones recientes del Examen Geológico a Kansas (Referencia 7.14). Esta técnica se llamo una "re-experiencia" técnica y contempla lo relativo a los resultados de la post-perforación a la percepción de la pre-perforación de la estructura geológica densamente perforada en el área central de Kansas.

Cuando se explora por aceite o gas en un área, normalmente se prepara o calcula la interpretación geológica de la estructura de los prospectos basados en el control disponible de datos sísmicos del pozo. Suponga que se percibe la interpretación de un anticlinal cercano. Cuando

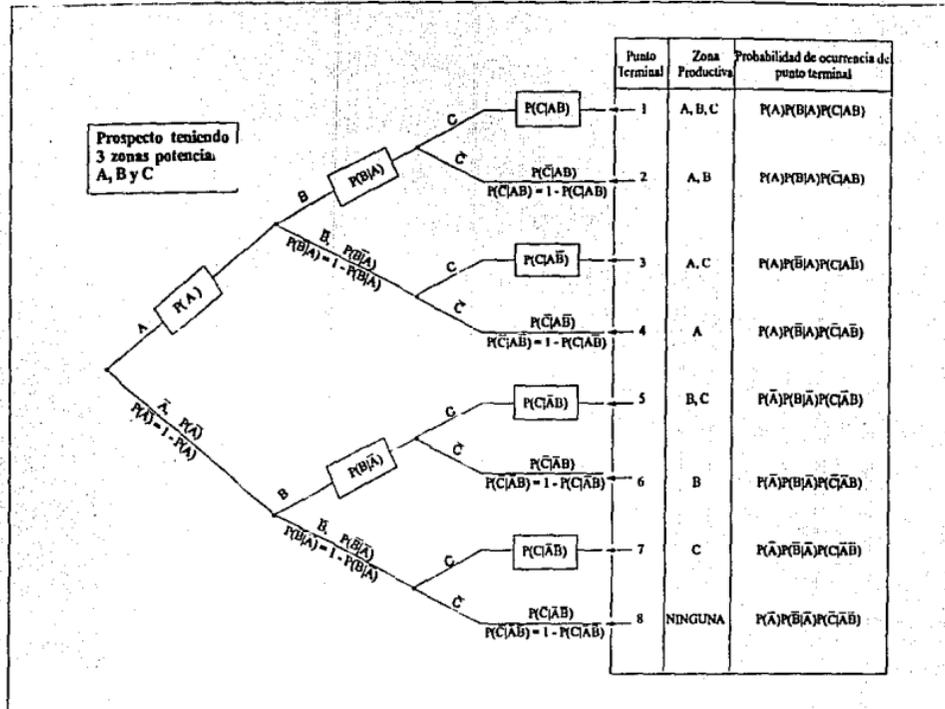


Figura 1. *Árbol de decisión mostrando posibles resultados en un prospecto de perforación teniendo tres zonas potenciales productivas A, B y C. Para usar este método los exploradores definen o estiman siete probabilidades encerradas en rectángulos. La barra sobre cada evento (A, B, etc.) significa la zona que no es productiva. P(A) significa la probabilidad ya sea que la zona A no sea productiva. Con este árbol las probabilidades de producción en las zonas son dependientes ya sea que las zonas en el prospecto son productivas. El mismo árbol muestra para el caso de eventos independientes en la figura 2.*

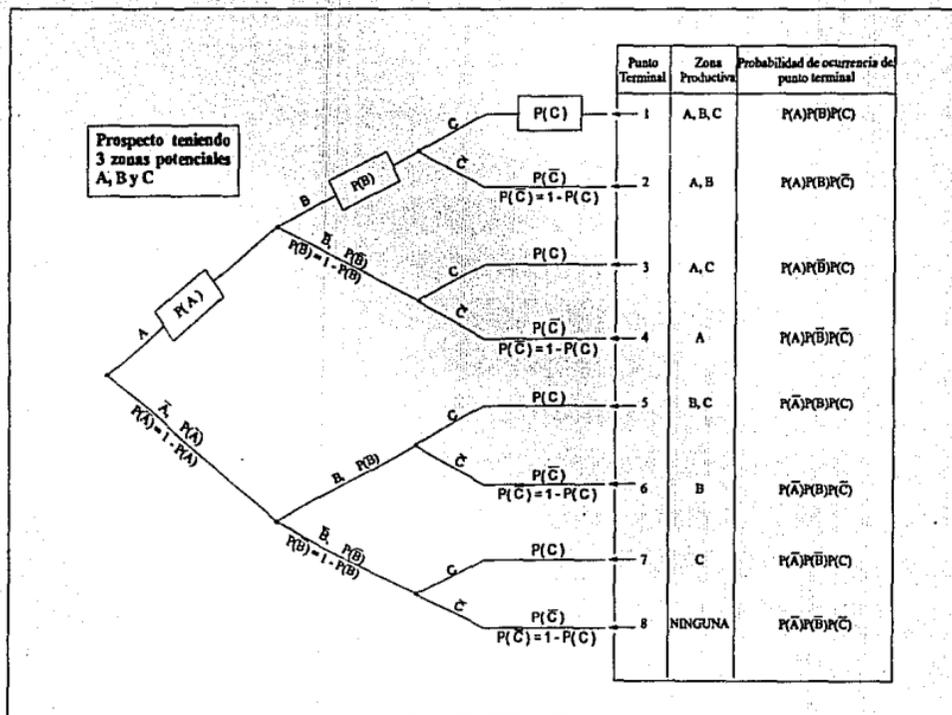


Figura 2. Árbol de decisión mostrando posibles resultados en un prospecto de perforación teniendo tres zonas potenciales A, B y C en el cual la probabilidad de producción en cualquier zona dada es independiente de que las otras zonas sean o no productivas. Se asume que A, B y C son resultados independientes en este caso especial del más general de la figura 1. Solamente tres probabilidades pueden ser especificadas para resolver el árbol - las probabilidades encerradas en los rectángulos. Si A, B y C no son eventos independientes este árbol no se aplica, y se usa el de la figura 1.

tratando de estimar la probabilidad de la existencia de un anticlinal productivo una pregunta lógica que puede hacerse es "Que ha resultado de perforar anticlinales cerrados (que se percibió antes de perforar) en esta formación en áreas semejantes cercanas?"

La investigación desarrollada responde a preguntas de este tipo por estudiar un área de 24 por 24 millas en el Condado septentrional de Stafford de Kansas central. En el periodo de tiempo de 1930-1970, ésta área de control había sido extensamente explorada (765 pozos exploratorios comprenden un área de 576 millas cuadradas) e información extensa en formación para sobrepasar, reservas etc. era disponible.

El primer paso en la técnica en preparar mapas estructurales (usando computadoras con el contorno por algoritmos) para cada año se basó en datos disponibles para el año en particular. (El mapa para 1935, por ejemplo, incluye todos los pozos perforados anteriormente a 1938 pero ninguno de los pozos perforados subsecuentemente). Estos mapas estructurales evidentemente cambiaron con el tiempo conforme nuevos pozos se perforaron y se disponía de más control.

Los pozos exploratorios eran indentificados por el año en que fueron perforados, y un ensayo se hizo para definir la interpretación estructural (anticlinal, nariz, monoclinal, sinclinal, etc) a ese sitio antes de que el pozo fuera perforado por computadora y medios visuales. Esta "pre-perforación" se comparó entonces con los resultados de la post-perforación del pozo exploratorio (seco, aceite, etc)

Para relacionar esta información para cada pozo exploratorio perforado durante el período de 40 años se pueden establecer frecuentemente los resultados de "post-perforación" como los relacionados con "pre-perforar" percepciones de la estructura. La idea era entonces usar estos resultados (probabilidades) del área de control densamente perforada con otras áreas nuevas de comparables condiciones geológicas en la misma formación como aproximaciones de las probabilidades de descubrimiento.

Esta "re-experiencia" técnica, por supuesto, es dependiente de tener un área de control densamente perforada disponible sobre que derivar las probabilidades de descubrimiento para un área nueva. Esta limitación restringe la aplicación de áreas maduras de exploración. Probablemente es muy difícil usar esto con confianza en áreas de exploración de pozo petrolero tál como en Greenland.

La investigación con esta técnica fué también cuestionada por problemas de encontrar métodos consistentes clasificados de "pre-perforación" con interpretaciones estructurales alrededor de cada pozo exploratorio. Se tiene también que recordar que tál enfoque implícito requiere que el área nueva de interés de exploración sea en la misma formación y geológicamente semejante a las condiciones del área de control.

Las probabilidades del área de control no deberán ser aplicadas, por ejemplo, a un mapa de contorno de una arena de agua poco profundas!

Para resumir, probablemente el método más común de probabilidades estimadas de un descubrimiento es usar relaciones pasadas de éxito. Cuando se usa una relación de éxito para determinar la probabilidad de un descubrimiento, en la prueba próxima se infiere que la serie de pozos perforados en un yacimiento es una serie de acontecimientos independientes. Mientras que las relaciones pasadas de éxito son útiles para estimar las primeras aproximaciones de probabilidades de un descubrimiento éstas se tienen que usar con cautela porque la inferencia de

pruebas independientes no es estrictamente realista del contexto real del mundo de exploración petrolera.

Probabilidades de descubrimientos pueden analizarse directamente usando la ecuación 1 si se tiene suficiente información acerca del prospecto para poder especificar o calcular los términos de probabilidad de cada factor que contribuyen a la ecuación. Si un prospecto tiene secciones múltiples productivas la probabilidad de descubrimiento puede ser calculada usando el método mostrado en las figuras 1 y 2.

La figura 1 es usada si la probabilidad de ocurrencia de aceite o gas en las zonas múltiples productivas es dependiente o ninguna de las otras zonas son productivas, y la figura 2 es usada si se considera que las probabilidades de producción en cada zona son independientes una de otra zona. El caso dependiente requiere especificación, siete términos de probabilidad y su uso puede limitarse debido a los problemas de estimar estos valores. Finalmente se tiene que mencionar un matiz sutil de interpretar el sentido de una probabilidad de descubrimiento.

Si se dice que la probabilidad de un descubrimiento es 0.3 en un programa de diez pozos perforados, normalmente se interpretaría éste significado como que es más probable que el treinta por ciento de los pozos perforados serán productivos - por ejemplo, tres descubrimientos buenos en un programa de diez. Pero ¿qué pasa si se considera un pozo - qué tiene $p=0.30$ que significa esto? La mayoría probablemente interpreta esto como la probabilidad de que en el pozo se descubra aceite o gas, cual es la interpretación correcta.

Esta interpretación es problemática para algunos, sin embargo argumentan que el pozo exploratorio es cualquiera de los dos: seco ($p=0$) o productivo ($p=1.0$) y ese $p=0.3$ es realmente un número sin sentido. - No se puede tener 0.3 descubrimiento - es o no es aceite! Para responder a preguntas de este tipo uno puede rápidamente perder de vista las definiciones básicas de probabilidad por seguir esta lógica.

Regresando al viejo ejemplo de monedas lanzadas. Cuando se lanza una moneda la probabilidad de cara es 0.5 y es convincente. Pero cuando en realidad se lanza la moneda es cualquiera de las dos probabilidades cara o no cara. "Así, cuál es la diferencia entre esto y el sentido de $p=0.3$ en el pozo simple? Ninguna. ¿Cuando se interpreta una probabilidad de descubrimiento para un solo pozo se tiene que interpretar cómo la probabilidad de descubrir aceite.

Si el análisis involucra pozos múltiples se puede interpretar la probabilidad de descubrimiento como la fracción del total de muchas pruebas que serán descubrimientos. Estas dos interpretaciones pueden darse en ese orden - la lógica no trabaja en la dirección inversa.

IL3 Modelos de Análisis de Riesgo basados en las Últimas Distribuciones de Reserva.

Se tienen varios modelos para definir niveles de probabilidades de reservas en varias etapas de la exploración en una cuenca usando una última distribución de reserva. Estos modelos están basados en la hipótesis de que la función de densidad de probabilidad de reservas remanentes por descubrir en una cuenca es proporcional al área entre la última distribución de reserva y la distribución de reservas descubiertas a la fecha. Lo último aquí es tomar algo medio después de que todos los prospectos han sido perforados. La variable aleatoria de estas distribuciones de reserva es reserva por campo.

Suponer que se puede estimar una distribución de probabilidad de reservas por campo en una cuenca geológica. Esta distribución será llamada la última distribución de reserva de la cuenca. Puede ser de cualquier forma y tener límites que son aplicables para las magnitudes de las reservas por campo.

Ahora se puede suponer más allá de esto, con el uso de una apropiada constante de proporcionalidad, se convierte la distribución última de probabilidad en una distribución de frecuencia en la que el eje vertical es ahora el número de campos en cada rango de reserva, más que la función de densidad de probabilidad de la distribución. Esta distribución de frecuencia puede aparecer como una parte de la figura 3.

La forma de la distribución última en la figura 3(a) es aproximadamente lognormal, pero la distribución de frecuencia puede tener cualquier forma que sea representativa. De esta distribución de frecuencia se puede leer que hay N_1 campos en la reserva en el rango de R_0 a R_1 barriles por campo, N_2 campos en la reserva en el rango R_1 a R_2 barriles por campo, etc.

Ahora se supone que alguna exploración ha ocurrido en la cuenca y en el tiempo T_1 , se tienen descubrimientos de campos de varios tamaños. Estos datos pueden ser graficados como histogramas en la misma gráfica, como se muestra en la parte b de la figura 3. Por el método descrito las probabilidades para encontrar varios tamaños de campos a ser descubiertos son proporcionales al área entre los histogramas de frecuencia que se encuentran a la fecha (Tiempo T_1) y la última distribución de frecuencia.

Se muestra el área entre estas dos curvas achurada en la figura 3 (b). Conceptualmente esto es equivalente a decir que si hay (últimamente) veinte campos en el rango de reserva de R_i a R_{i+1} y ocho campos se han sido encontrado como de T_1 , son doce los encontrados en subsiguientes pozos exploratorios.

Como avanza el tiempo, T_2 , más campos se encuentran y el histograma de frecuencia de reservas a la fecha se mueve más cerca al límite superior (última distribución). Si todos los campos en un rango particular de reserva son descubiertos el histograma de frecuencia y la última distribución de frecuencia coinciden. El área entre ellas para el rango de reserva es cero, así la probabilidad de encontrar más campos en ese rango es cero. Eventualmente, cómo (y si) todos los prospectos son perforados las dos curvas coincidirán sobre el rango completo de reservas por campo.

Para T_0 la última distribución de frecuencia superior es constante y permanece fija en la gráfica de distribución. La única curva que cambia con el tiempo es la frecuencia inferior del histograma de frecuencia de los descubrimientos a la fecha. Y a cualquier punto en el tiempo la función de densidad de probabilidad (y por lo tanto las probabilidades) de campos remanentes a ser descubiertos permanecen proporcionales al área entre las dos curvas de frecuencia.

La explicación racional para este enfoque es un poco simplista. Las propuestas originales para este método de determinación de las probabilidades de nivel de reserva no hicieron el paso adicional que era usar una constante proporcional para convertir de distribuciones de probabilidad a distribuciones de frecuencia de la figura 3. Fueron formuladas usando notas sobre la función de densidad de probabilidad de la última distribución de reserva y la distribución de descubrimientos a la fecha. La descripción aquí en términos de frecuencias era describir la lógica del método mientras se evita involucrar matemáticas.

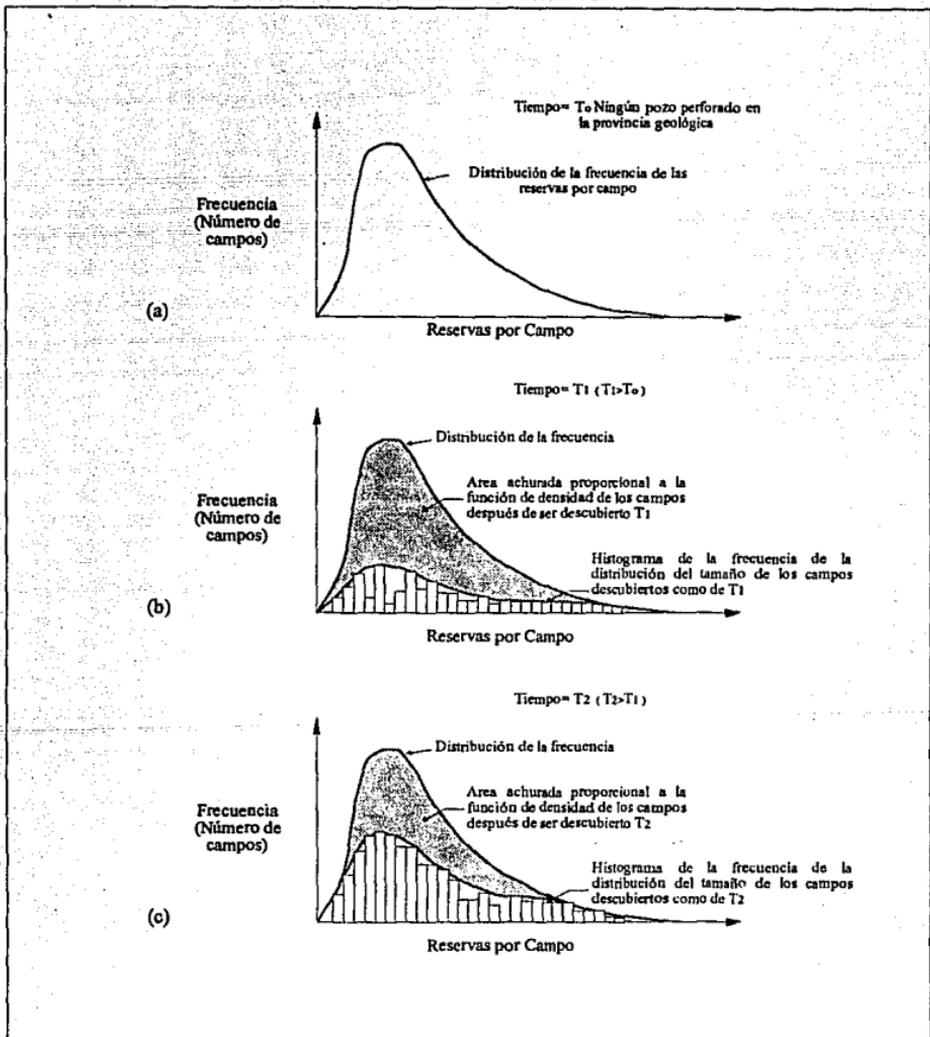


Figura 3. Gráficas de la distribución de frecuencia de reservas descubiertas en una provincia geológica a varios puntos en el tiempo.

¿Es realista el método? ¿Funciona? Las respuestas son sí y no respectivamente. Conceptualmente, la hipótesis parece válida. Reconoce la "prueba con reemplazo" de explorar naturalmente para aceite. Explica el hecho evidente de que una vez que se hace un descubrimiento hay un campo menos en esa reserva para ser hallado en subsiguientes pozos exploratorios.

Usando este método en cualquier punto en el tiempo se podría calcular cuántos campos quedan para encontrar en cada rango de reserva y expresar los valores como frecuencias relativas condicionales, y multiplicando las oportunidades de encontrar aceite en la primera localización por las frecuencias condicionales relativas se obtienen las probabilidades de descubrimientos de varios tamaños. Estas probabilidades pueden ser usadas como una norma del valor medio (EMV) calculado, y la exploración continúa hasta el tiempo tal en que el EMV, para continuar la exploración, se hace negativo y/o es menor que el EMV para representar la exploración de una alternativa.

Ahora lo malo. No funciona probablemente para el mundo real a causa de la pregunta evidente - Cómo se estima la última distribución de frecuencia en primer lugar? Si la curva superior no puede definirse a cero de tiempo (o antes de la exploración de una cuenca nueva) el esquema entero es pequeño y en vano. El hecho es que es virtualmente imposible estimar esta distribución última hasta que la cuenca ha sido completamente perforada -en éste punto las probabilidades de reserva se hacen dudosas de todas formas. - Si se sabe la última distribución de reservas por campo al tiempo cero muchas de las decisiones de inversión de exploración son considerablemente simplificadas!

Así es que, con este método de análisis de riesgo se requiere especificar una última distribución de reserva, y a menos que ésta sea el método a utilizar, es de valor pequeño y práctico. Un investigador propone que la última distribución de reserva es lognormal y puede ser definida por el analista para estimar sus dos parámetros μ y σ . No ayuda mucho.

Otros tienen ensayos de re-estudios de exploración con perforación secuencial de muchas cuencas perforadas para tratar de determinar si hay cualquier señal o tendencia de la cual se pueda estimar la última distribución de reserva de un nuevo juego que represente un punto temprano en el tiempo. Los prospectos para un rompimiento aquí son muy débiles realmente.

Probablemente la situación del mundo-real donde éste método puede ser usado, efectivamente en un juego en el que la geología, estratigrafía, etc. son correlacionadas a una cuenca cercana (o similar) que ha sido extensamente explorada. Un caso al respecto en este punto es la presente (1974/1975) extensión en el suroeste de la cuenca Denver-Julesburg. Esta representa un nuevo activo que parece estar correlacionado con una más antigua (1950 y 1960) actividad al norte y este.

En la porción más antigua más de 700 campos hablan sido encontrados antes en 1970 y parece que la distribución de reservas por campo para esta área puede aplicarse para estimar los probables resultados de la nueva actividad en la porción sureste de la cuenca (cercana a Denver). Por supuesto que por tal "analogía" la extrapolación significa que ambas áreas tienen similares depositaciones históricas y geológicas.

Ese método de análisis de riesgo es conceptualmente válido pero demasiado difícil, si no imposible de aplicar a causa de la inhabilidad para definir la última distribución de reserva en una cuenca nueva. Sin embargo, está predispuesta por un público más crítico. Aunque se puede, por alguna pequeña brillante percepción, determinar la última distribución de probabilidad de reservas por campo cerca del tiempo cero, ¿puede esto resolver el problema? ¿Sería suficiente saber por

ejemplo, que la probabilidad de un campo más grande que un billón de barriles es 0.10 (el área en la cola derecha de la distribución de reserva para reservas más grandes que un billón de barriles por campo?)

La respuesta es no. Para saber las probabilidades de varios niveles de reservas como para leer la última distribución de probabilidad es solo una parte del problema. El eslabón perdido es saber de ese 0.10 ¿cuántos campos son más grandes que un billón de barriles?. Esto es, se requiere saber el parámetro por el que se pueden obtener probabilidades para el número real de campos en cada rango de reserva. Este parámetro es la constante de proporcionalidad que antes se mencionó para convertir una distribución de probabilidad en una distribución de frecuencia.

Los parámetros son el número total de prospectos en la cuenca que tienen aceite o gas. Se llamará a N constante, el número total de prospectos productivos en la cuenca. -Esto es absolutamente vital en análisis de riesgo de exploración!. Sin especificar este parámetro no se pueden determinar probabilidades de qué es lo que se encuentre en un punto en el tiempo. El solo decir que el 0.10 (10 %) de los campos productivos tendrá reservas más grandes que un billón de barriles no es suficiente.

Se puede decir que 10% de N campos tendrán por lo menos un billón de barriles. Si N es 10 esto significa que hay $0.10 \times 10 = 1.0$ campo de este tamaño en la cuenca, y si uno ya ha sido descubierto sería mejor dejar de perforar - no hay más que encontrar. Si $N = 100$ esto significa que $0.10 \times 100 = 10$ campos de este tamaño en la cuenca. Si se ha encontrado uno se tiene que mantener la perforación no obstante porque, hay nueve por ser descubiertos.

El punto aquí es que el conocimiento de la última distribución de reservas por campo en una cuenca puede proveer un paso importante en determinar probabilidades de qué se pueda encontrar un punto en el tiempo. Pero en suma para conocer la última distribución se tiene también que conocer (o poder estimar) el número total de prospectos productivos que existen en la cuenca. Esta importante proporcionalidad es, en la mayoría de los casos, difícil de estimar en el juego como la última distribución de reserva. Uno u otro de estos términos probablemente requieren un enfoque para determinadas probabilidades. Se hablará acerca de estimar N un poco más, en la sección siguiente de este capítulo.

Antes de ir a esta breve discusión acerca de distribuciones de reserva hay que poner atención a varios puntos sutiles acerca de analizar datos de reserva. Una declaración muy común es que las reservas por campo en una cuenca sedimentaria son distribuidas lognormalmente. Esta declaración es, en el sentido estricto de la palabra, probablemente incorrecta. Quizás se tiene que decir que en muchas cuencas parece que la distribución de reservas por campo puede ser aproximada a una distribución lognormal.

La declaración que las reservas son distribuidas lognormalmente es simplemente muy fuerte. Un factor es el punto de que no se sabe si el origen, acumulación, migración, y entrapamiento de reservas de aceite y gas fueron controladas por (o el resultado de) una distribución lognormal. Mientras que la distribución lognormal parece ser una aproximación razonablemente buena de reservas por campo en muchas ocasiones hay algunas cuencas en donde la distribución de reservas es simplemente no una lognormal. Así que se tiene que ser muy cuidadoso y no suponer más que las reservas son distribuidas lognormalmente.

Otra inquietud, que parte de tal declaración, es que el parámetro usual como variable aleatoria para éstas distribuciones de reserva son reservas recuperables - el único valor del tipo de reservas.

Reservas recuperables son una función de cómo hoy, operan las ganancias y el precio obtenido para el aceite y gas. Reservas recuperables son una función de la producción económica del aceite y gas.

Es razonable suponer que naturalmente los depósitos distribuidos de aceite y gas consisten en una exacta distribución lognormal por costumbre, durante millones de años se ha sabido cómo hoy, reservas recuperables se definen económicamente?. Esto es extremadamente dudoso. Quizás el volumen de hidrocarburos en el poro puede ser muy bien definido de una manera lognormal, pero reservas recuperables son una función de cómo se operan los campos de aceite y parece dudoso que las influencias de hoy, impuestos, precios de aceite, costos, etc. tengan un buen comportamiento.

Y hasta lleva varios pasos del razonamiento, cuándo se evalúan reservas recuperables por campo ¿cómo se explica para los volúmenes de gas asociado producido con el aceite?. "Se convierten volúmenes producidos de gas a volúmenes equivalentes de yacimiento a una presión base? "O en una base equivalente al valor del dolar?

"O se ignoran los volúmenes de gas producidos, y se grafican las reservas de aceite determinadas así como producción acumulativa más reservas remanentes estimadas? "¿Qué pasa si un campo de aceite tiene gas inicial - como se trata el volumen de gas y el gas producido en determinadas reservas para el campo?

Otro problema cuando se trata con datos de reserva se relacionan con lo que es llamado el truncamiento de datos. Por ejemplo si se tiene una lista de todos los campos en la cuenca que tienen reservas mayores que 100 millones de barriles cada uno, puede con estas reservas graficar la probabilidad en papel logaritmico pero éstas no se ajustarían a una distribución lognormal. La razón de esto es que en virtud de los 100 millones de barriles de límite, los datos representarían antes solamente el extremo derecho de la distribución.

Los valores bajos de reservas no fueron incluidos en los datos. Estos valores bajos de reservas cuando se incluyen con las de 100 millones de barriles pueden en realidad ser distribuidos lognormalmente, pero si inicialmente solo son usados los datos de los campos grandes, no sorprende que los datos no quepan en una distribución lognormal. Los valores bajos eran truncados, o suprimidos.

Así es que cuando se analizan datos de reserva se tiene que estar seguro de que se tienen los valores más pequeños de reserva al lado izquierdo es decir el extremo izquierdo de la distribución lognormal. Pero ésta declaración entonces conduce a otro dilema. "¿Que tan pequeños tienen que ser los valores de reservas por campo para asegurar que se ha evitado el "truncamiento de datos"? Campos con reservas bajas ¿como 100,000 barriles? ¿10,000 barriles? ¿1,000 barriles?

Algunos exploradores han visto que si se incluyen todos los depósitos de aceite en la cuenca bajo el límite de cero barriles por campo la curva de distribución nunca desciende al lado izquierdo (como en una distribución lognormal). La distribución continuaría más arriba conforme las reservas por campo se hacen más y más pequeñas. Tal distribución, casi es descrita como una distribución exponencial (o una distribución gamma teniendo los parámetros apropiados), y que nunca se vería una distribución lognormal.

La distribución puede aparecer como en la figura 4 (b). Las propuestas observadas, en esencia, son que si alguna vez se es capaz (y le interesa) tabular las reservas por campo para los muy pequeños menores a R_{min} no habría campos en este rango (como en la figura 4 (a)) habría un

número muy grande de ellos, como se muestra en el área de la figura 4 (b) entre cero reservas y reservas de R_{\min} .

Y por supuesto es claro que implicaría que las reservas nunca pueden distribuirse lognormalmente si se truncan los datos de reserva a algún valor mínimo, R_{\min} . ¿Conviene o no conviene?

La obvia consecuencia práctica de tal hipótesis es que nunca se sabrá, porque una vez que las reservas por campo obtienen menos que cualquier mínimo nivel económico éstas no interesan. No puede ser rentable una muestra de aceite o un depósito de aceite que tiene sólo un pie de zona productiva. En la práctica todo R_{\min} más pequeño es clasificado como un agujero seco exploratorio y se busca aceite en otra parte. Pero la hipótesis con algún grado de validez tiene que decir que reserva tiene que ser lognormalmente distribuida.

Los números teóricamente son buenos tópicos, pero desde un punto de vista práctico se puede dar paso a los números en los análisis de riesgo. Para reservas menores que R_{\min} probablemente no se termine el pozo descubierto, se pueden clasificar estos resultados como agujeros secos exploratorios. Para descubrimientos más grandes que R_{\min} se calcula un valor económico asociado, y la distribución de estas reservas (lognormal o de otra forma) sería una entrada al análisis. Esto elimina la necesidad de definir las distribuciones de reserva para valores de reservas menores que R_{\min} .

Resumiendo, en muchas cuencas del mundo las reservas por campo pueden ser razonablemente aproximadas por una distribución lognormal - pero hay algunas áreas donde las reservas descubiertas simplemente no entran en una distribución lognormal. Así es que se tiene que evitar suponer que todas las reservas están distribuidas lognormalmente.

II.4 Probabilidad de Resultados de Programas de Perforación Multipozo

Las distribuciones de probabilidad pueden ser usadas en programas de perforación multipozo. Estas incluyen las distribuciones binomial, multinomial, y la hypergeométrica. Ahora se discutirá en detalle substancial cómo estos métodos pueden ser aplicados a la exploración petrolera. En particular, se mostrará como éstas distribuciones (modelos análogos) pueden ser usados para calcular probabilidades de resultados de perforar una serie de pozos exploratorios.

El primer punto que interesa analizar en el programa de exploración pozo-múltiple es la perforación de los pozos como una serie de eventos independientes o una serie de eventos dependientes. La razón que se tiene para hablar sobre esto es que los cálculos son diferentes para cada caso. El número es vital ya que se tiene que especificar uno u otro para aplicar al juego de pozos considerado. Esto es, tomar una posición neutral.

Si el juicio de los exploradores es que la serie de pozos es una serie de eventos independientes, esto implica que las probabilidades de varios resultados pueden ser constantes con el tiempo, y que la probabilidad de que cualquier resultado dado no sea afectado por no afectar la probabilidad de un resultado en alguna otra prueba. La noción de independencia implica el fenómeno de probar con replazo, y por lo tanto es un espacio de muestra que permanece sin cambiar en todo el tiempo.

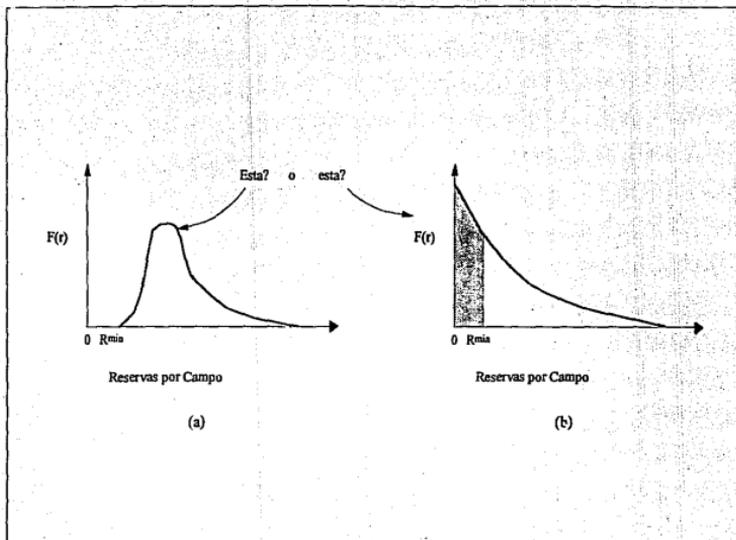


Figura 4. Posibles formas de la distribución de reservas por campo en una cuenca. R_{min} representa las mínimas reservas requeridas para justificar la terminación del pozo descubridor.

A la inversa, si este es un juicio mejor que el propuesto, es decir una serie de eventos dependientes, significa que las probabilidades de varios resultados cambian sucesivamente después de que cada pozo exploratorio ha sido perforado, dependiendo qué ha sido hallado. Eventos dependientes significa probar sin reemplazo, y el espacio de muestra tiene solamente un número finito de elementos.

Cuando se define el espacio de muestra como consistente de todos los prospectos geológicos, la representación base de dependencia será la representación aplicable de la serie de pozos múltiples de perforación. En ciertas ocasiones, sin embargo, cuando la noción de eventos independientes es correcta, se puede elegir analizar las opciones de decisión en esta base para otra vez aprovechar los cálculos más sencillos.

En la tabla 2 se presentan los dos tipos de fenómenos probabilísticos y sus aplicaciones. La tabla muestra dos columnas para distintos métodos de análisis del programa de perforación pozo-multiple - uno basado en la condición de eventos dependientes y otro basado en la condición de eventos independientes.

II.4.1. Análisis basado en la Condición de Eventos Dependientes

Se ha dicho en varios puntos de ésta discusión que la perforación de una serie de pozos es una situación de eventos dependientes. ¿Qué es lo que realmente significa esto? ¿Cómo se puede decir si una serie de pozos en el mundo-real es un evento independiente ó dependiente? ¿Se hace tal distinción en esta materia?

Para contestar la primera de éstas preguntas considere un experimento en el cuál un grupo de 52 cartas son completamente barajadas y puestas cara hacia abajo en una mesa. Para esta analogía se supondrá que cada una de las cartas representa un prospecto geológico en un área de la cuenca en exploración. Por lo tanto hay 52 "prospectos" o elementos en el espacio de muestra.

Ahora se supone que todos los diamantes representan el prospecto productivo de aceite y todos los corazones, espadas, y tréboles corresponden a prospectos que no tienen aceite o gas. Finalmente, para completar la analogía se supone que se elige una carta (que está cara abajo inicialmente) y se le da vuelta hacia arriba lo que es equivalente a perforar un prospecto con un pozo exploratorio. Si alguna de las otras tres cartas se da vuelta hacia arriba el prospecto se considera pozo seco.

¿Algunos pozos son perforados (esto es, algunas cartas son volteadas hacia arriba) ¿cuál es la probabilidad de encontrar aceite? ¿Es 13/52, realmente? Hay 52 elementos en el espacio de muestra cada uno tiene una oportunidad de ocurrencia igual, y de estos 52 elementos 13 son definidos como diamantes. Así, se selecciona una de las 52 cartas y se le da vuelta hacia arriba, la probabilidad de un diamante es exactamente 13/52.

Suponga que se da vuelta a una carta y es un diamante. ¿Ahora ¿cuál es la probabilidad de que la segunda carta a la que se da vuelta también sea un diamante? Como se han dado las 51 cartas hacia abajo sin tocar y la carta original que está todavía hacia arriba (por lo tanto no se puede seleccionar en alguna prueba subsiguiente) la respuesta es 12/51. De las 51 cartas que quedan hacia abajo hay $13 - 1 = 12$ diamantes, la oportunidad de que un diamante sea la segunda carta vuelta hacia arriba "descubierta", es 12/51.

Tabla 2.
Métodos para análisis de programas de perforación multipozo
basado en eventos independientes y dependientes.

Eventos Dependientes		Eventos Independientes
<p>Los pozos se consideran como una serie de eventos dependientes. Las probabilidades de resultados en cualquier pozo son dependientes de que los descubrimientos se hayan dado. Estas probabilidades cambian sucesivamente después de que cada prospecto ha sido perforado. Una prueba sin reemplazo. Un número finito de prospectos en la cuenca (o área explorada), N y después de que cada prospecto es perforado, hay un elemento menos en el espacio de la muestra de prospectos remanentes no perforados. El espacio de muestra cambia después de que cada pozo ha sido perforado.</p>	<p align="center">Características generales de la probabilidad análoga.</p>	<p>Los pozos se consideran como series de eventos independientes. Las probabilidades de resultados en cualquier pozo son independientes de nada que haya pasado antes o que vaya a pasar después. Estas probabilidades permanecen constantes después de que cada prospecto ha sido perforado. Una prueba con reemplazo. Implica un número infinito de prospectos en la cuenca - y tan largo como se continúe perforado hasta tener una proporción constante de éxitos y fracasos. No se conoce el número finito de prospectos en la cuenca (o área explorada). El espacio de muestra remanente es el mismo después de que cada prospecto ha sido perforado, aunque siempre hay un prospecto menos a perforar, después que un evento exitoso es terminado.</p>
<p>Cuando el espacio de muestra de resultados está definido como todos los prospectos de exploración en una cuenca (o área explorada) esta condición de dependencia se aplica en todos los casos. Ésta es la real representación de la secuencia de perforación de pozos. Es particularmente importante para "prueba sin reemplazo" los aspectos de exploración si el número total de prospectos en la cuenca o área, N, es pequeño (del mismo orden de magnitud como el número de pozos perforados en el programa de exploración, n). Si el número de prospectos es muy grande comparado al número de pozos perforados (para valores de N menores a 10n) las probabilidades calculadas usando la población de dependencia pueden aproximarse a su equivalente "evento independiente" método que aventaja a un simple cálculo.</p>	<p align="center">Aplicación a programas de perforación multipozo de exploración (o de desarrollo)</p>	<p>Debido a que las características y realidades de exploración es un proceso de muestra sin reemplazo (más que el muestreo con reemplazo) este método de análisis tiene aplicación únicamente en las situaciones especiales:</p> <p>Cada uno de los pozos en el programa multipozo está en una cuenca diferente. En una cuenca geológica de interés donde se existe una variación y complejidad estratigráfica, estos pozos perforados siempre, para los intentos y propósitos, caen en un yacimiento separado. Éste es un caso muy especial, el cual solamente ocurre en algunas etapas maduras de exploración.</p> <p>El método de análisis basado en la condición de eventos independientes provee una aproximación de las probabilidades derivadas usando métodos de eventos dependientes, si el número de pozos en el programa de perforación, n, es pequeño comparado al número total de prospectos no perforados, N, en la cuenca o área de exploración. Para ésta aproximación N deberá ser al menos 10 veces el valor de n (es decir, $N \geq 10n$). Como N es muy grande la probabilidad obtenida por métodos para eventos independientes aproximadamente será igual a la probabilidad correspondiente basado en una serie de eventos dependientes.</p>
<p>Distribución Hipergeométrica (Ecuación 15).</p>	<p align="center">Modelo de distribución de probabilidad aplicable</p>	<p>Solamente se consideran 2 resultados (ejemplo éxito o fracaso, productor o pozo seco)</p> <p>Distribución Binomial (Ecuación 3).</p> <p>Distribución Multinomial (Ecuación 4).</p>
<p>N = Número total de prospectos no perforados en la cuenca o área de exploración antes de que comience el programa de perforación. n = Número de eventos (pozos) en el programa multipozo. r = número de posibles resultados en cualquier evento. d_1, d_2, \dots, d_r = número de elementos en el espacio muestra designados como cada resultado antes de r eventos con la condición de $(d_1 + d_2 + \dots + d_r) = N$ $x_1, x_2, x_3, \dots, x_r$ = número de resultados que ocurren en cada una de las r categorías, con la condición de que $(x_1 + x_2 + x_3 + \dots + x_r) = n$.</p>	<p align="center">Factores que pueden ser especificados por los analistas al usar el método</p>	<p>n = Número de eventos (pozos) en el programa multipozo. x = número de éxitos de interés ($0 \leq x \leq n$) p = probabilidad de éxito en algún evento dado ($0 \leq p \leq 1.0$)</p> <p>a = Número de eventos (pozos) en el programa multipozo. r = número posible de resultados en cualquier evento. $x_1, x_2, x_3, \dots, x_r$ = número de veces que los resultados ocurren en n, eventos, con la condición $(x_1 + x_2 + x_3 + \dots + x_r) = n$. $p_1, p_2, p_3, \dots, p_r$ = probabilidad de ocurrencia en cualquier evento dada la condición $(p_1 + p_2 + p_3 + \dots + p_r) = 1.0$</p>

Esta probabilidad es una probabilidad condicional - es condicional por el hecho que se tiene que dar vuelta arriba a un diamante. La serie de dar vuelta hacia arriba a dos cartas en la manera descrita, es una serie de eventos dependientes. Las otras para encontrar un diamante en la prueba próxima cambian después de cada prueba, depende de lo que se ha dado volteado hacia arriba.

Suponga que se habla seleccionado y volteado hacia arriba a cinco cartas y ninguna de las cinco cartas es un diamante. ¿Qué probabilidad hay de que la sexta carta volteada hacia arriba sea un diamante? La respuesta es 13/47. Hay 47 cartas aún para darles vuelta hacia arriba tan sólo 13 de los diamantes originales permanecen en el grupo de 47 cartas cara abajo en la mesa.

La analogía de las 52 cartas cara abajo en una mesa es exactamente igual al mundo-real de la exploración petrolera. Naturalmente se han distribuido N prospectos en la cuenca o área de exploración, algunos de los cuáles tienen aceite y otros que están secos. Como se perfora cada prospecto se esta, en esencia, dando vuelta a cada una de las cartas para ver si tiene aceite o está seco. Pero de estas cartas volteadas hacia arriba hay una menos en el espacio de muestra de resultados en la próxima prueba. En el mundo-real una vez que un prospecto ha sido perforado se quita del total y permanece perforado. Cada sistema es probado sin reemplazo.

¿Qué analogía equivalente sería si la carta experimental que representa es una serie de eventos independientes? El experimento sería exactamente el mismo, y en la primera carta escogida para el diamante sería lo mismo 13/52. Pero después para la serie de eventos independientes se tendrían que recoger todas las 52 cartas, completas, y se ponen cara abajo antes de elegir otra carta.

El resultado de la segunda carta de un diamante es 13/52. Cualquiera que sea la primera carta que es colocada de espaldas en la cubierta y se dispone para selección otra vez en la segunda prueba del experimento. El segundo 13/52 es una probabilidad que es independiente de lo que ocurrió en la primera prueba. Continuar esta serie, después de que la segunda prueba ha sido realizada, ejecutar todas las 52 cartas otra vez, hacia abajo en la mesa y tomar otra carta elegida y darle vuelta hacia arriba.

El resultado para esto ¿es un diamante?. Exactamente el mismo - 13/52. El resultado para un diamante por este proceso siempre será 13/52. ¿Aunque se repitan las pruebas 20 veces, o 100 o 1 millón de veces? Se llama a ésta prueba sin reemplazo y las pruebas reiterativas son una serie de eventos independientes.

Ahora claramente esta analogía pasada no se aplica al mundo-real a causa de que una vez perforado un prospecto y observado si tiene aceite, no se puede "desperforarlo" y ponerlo atrás en el espacio de muestra disponible para ser elegido y perforado en una prueba subsiguiente. Un prospecto que se ha perforado se queda así. Se lleva fuera del espacio de muestra de prospectos que pueden ser perforados en pruebas subsiguientes.

Como un resultado de todo esto se tiene claramente que convenir que la serie de perforación de varios pozos de exploración es un proceso de prueba sin reemplazo en una serie de eventos dependientes en la que el resultado cambia después de que cada prospecto ha sido perforado. Las probabilidades de que se pueda encontrar algún punto son dependientes del hallazgo. Y para reconocimiento de estas condiciones se puede, por razonamiento inverso, concluir que esa serie de pozos de exploración perforados dentro de una cuenca o área de exploración no es una serie de acontecimientos independientes.

Con esta base se puede mirar un problema sencillo. En este ejemplo el área de exploración de interés contiene 10 anomalías sísmicas de las cuales 30% se estima tienen aceite. La pregunta de interés es: "¿Cuál es la probabilidad de encontrar dos descubrimientos en un programa de exploración de 5 pozos?" la respuesta es 0.417 y fue calculada usando la ecuación 2 de probabilidad hipergeométrica de dos resultados. Pero se puede determinar esto por lógica en vez de usar la ecuación.

Suponga que la primera de las dos anomalías que se prueba es aceite y la siguiente es pozo seco. Usando los símbolos P = productivo y D = seco, la serie sería P, P, D, D, D. ¿Cuál es la probabilidad de ocurrencia de la serie? La probabilidad de que la primera anomalía seleccionada tenga producción de aceite, con todos los otros factores iguales, son $3/10$. La probabilidad condicional de que la segunda anomalía tenga aceite, dado que la primera tuvo aceite, es $2/9$. (Recuerde la analogía original de las 52 cartas!)

La probabilidad condicional de que la tercera sea seca es $7/8$. (Las ocho anomalías no perforadas incluyen las siete anomalías originalmente secas). La probabilidad condicional de que la cuarta anomalía sea seca, dado que dos anomalías son aceite más una anomalía seca, es $6/7$. Finalmente, la probabilidad condicional de la existencia de la quinta anomalía seca, dadas esas dos de cuatro ya probadas había sido encontrar seco es $5/6$. La probabilidad de que esta serie entera de P, P, D, D, D, pueda ocurrir es el producto de estos cinco términos de probabilidad. La probabilidad de dos descubrimientos seguidos por tres fracasos es $3/10 \times 2/9 \times 7/8 \times 6/7 \times 5/6 = 3/72$.

¿Esta es la contestación a la pregunta original? - Definitivamente no! Esta es la probabilidad de una manera particular de obtener 2 descubrimientos en 5 pozos, pero hay otras series que también pueden ocurrir. Por ejemplo PDPDD, o PDDPD, etc. Así es que ahora se tienen que enumerar las maneras mutuamente exclusivas en que en la serie de cinco pruebas puede incluirse dos descubrimientos, y analizar la probabilidad de que cada serie en particular ocurra. El paso final es la suma de las probabilidades de las maneras mutuamente exclusivas en las que se puede obtener dos descubrimientos en cinco pruebas, y esta suma será la respuesta que se busca.

Para este ejemplo hay 10 caminos mutuamente exclusivos de lograr dos descubrimientos en cinco pruebas, como se registra en la columna 1 de la tabla 3. Las probabilidades de cada serie, se basan en la condición de prueba sin reemplazo - eventos dependientes- las cuales se muestran en la columna 2 de la misma tabla.

Se observa que en esta serie de cálculos la probabilidad para dos descubrimientos en cinco pozos perforados es $30/72 = 0.417$. En general no se necesita registrar tal detalle porque contestar preguntas de este tipo se puede solucionar con la ecuación de probabilidad hipergeométrica (ecuación. 2) directamente. La tabla 14 muestra lo que la ecuación hace.

Se notan otras dos cosas acerca de la pregunta del ejemplo. La exacta analogía con las cartas (y continuando la suposición de que un diamante es igual a aceite y algún otro igual a un pozo seco) es tomar 10 cartas (había 10 anomalías sísmicas) de tal forma que 3 sean diamantes y las otras 7 cartas sean algún otro. Las 10 cartas son completamente cambiadas y puestas cara arriba en la mesa. La probabilidad de que dos de estas cinco cartas sean diamantes cambia.

Tabla 3.

Ejemplo de diez formas de alcanzar dos descubrimientos de aceite en cinco intentos.

Secuencias mutuamente exclusivas (P = anomalía productiva D = anomalía seca)	Probabilidad de ocurrencia de la secuencia dado que de las anomalías iniciales N = 10, hipotéticamente 3 tienen aceite y 7 son secas*
1. PDDDD	$3/10 \times 2/9 \times 7/8 \times 6/7 \times 5/6 = 3/72$
2. PDPDD	$3/10 \times 7/9 \times 2/8 \times 6/7 \times 5/6 = 3/72$
3. PDDPD	$3/10 \times 7/9 \times 6/8 \times 2/7 \times 5/6 = 3/72$
4. PDDDP	$3/10 \times 7/9 \times 6/8 \times 5/7 \times 2/6 = 3/72$
5. DPDD	$7/10 \times 3/9 \times 2/8 \times 6/7 \times 5/6 = 3/72$
6. DPDP	$7/10 \times 3/9 \times 6/8 \times 2/7 \times 5/6 = 3/72$
7. DDPDD	$7/10 \times 3/9 \times 6/8 \times 5/7 \times 2/6 = 3/72$
8. DDPDP	$7/10 \times 6/9 \times 3/8 \times 2/7 \times 5/6 = 3/72$
9. DDDPD	$7/10 \times 6/9 \times 3/8 \times 5/7 \times 2/6 = 3/72$
10. DDDPP	$7/10 \times 6/9 \times 3/8 \times 3/7 \times 2/6 = 3/72$
	$\Sigma = 30/72 = 0.417$

* Mientras las probabilidades cambian después de cada evento para reflejar la relación de dependencia en el siguiente evento, cada uno de los prospectos restantes no perforados se asumen igualmente probables de seleccionarse. Este esquema es llamado "prueba igualmente probable sin reemplazo".

Y hay que notar también que las probabilidades de que la siguiente anomalía sea productiva o seca en las 10 series siempre cambian. Las oportunidades de que pueda subir o bajar el petróleo - siempre cambian como un resultado de un mecanismo de prueba sin reemplazo. Si éstos fueron una serie de eventos independientes las probabilidades para la serie 4, por ejemplo, (PDDDP) pueden ser $3/10 \times 7/10 \times 7/10 \times 7/10 \times 3/10$. Esto es claramente un valor diferente al valor (correcto) de $3/10 \times 7/9 \times 6/8 \times 5/7 \times 2/6$.

Todo esto es una serie de eventos dependientes - prueba sin reemplazo - se tiene que usar el modelo hipergeométrico de probabilidad, ecuación 2 para calcular las probabilidades de resultados de pruebas múltiples.

$$\text{Probabilidad de } x_1, x_2, x_3, \dots, x_r \text{ descubrimientos en una prueba de } n \text{ eventos} = \frac{\binom{d_1}{x_1} \binom{d_2}{x_2} \binom{d_3}{x_3} \dots \binom{d_r}{x_r}}{\binom{d_n}{x_n}} \quad (2)$$

donde

n = número de pruebas (pozos) en el programa múltiple de pozo.

N = número total de prospectos no perforados en la cuenca o área de exploración antes de perforar n pozos.

r = número de resultados posibles en alguna prueba.

d₁, d₂, ... d_r = número de elementos en el espacio de muestra señalados como cada resultado antes de perforar los n pozos (d₁ + d₂ + ... + d_r = N)

x₁, x₂, ... x_r = número de resultados que ocurre en cada una de las r categorías, tal que (x₁ + x₂ + ... + x_r = n)

Los números equivalentes para el problema de ejemplo son los siguientes:

$n = 5$, el número de pozos perforados (uno de cada cinco anomalías sísmicas).

$N = 10$, el número de prospectos no perforados inicialmente en la área de exploración.

$d_1 = 3$ } El número de elementos en el espacio original de muestra, donde el subíndice 1
 $d_2 = 7$ } corresponde al resultado con aceite y el subíndice 2 a la categoría de seco.

$x_1 = 2$ } El número actual de resultados en la categoría de aceite y seco anomalías
 $x_2 = 3$ } (El subíndice 1= aceite, 2= seco).

Substituyendo directamente se tiene:

$$\begin{array}{l} \text{Probabilidad de } x_1 = 2 \\ \text{descubrimientos y } x_2 = 3 \\ \text{fracasos en un programa} \\ \text{de } n = 5 \text{ pozos} \end{array} = \frac{\binom{3}{2} \binom{7}{3}}{\binom{10}{5}} = \frac{\binom{3!}{2!1!} \binom{7!}{3!4!}}{\binom{10!}{5!5!}} = \underline{\underline{0.417}}$$

Se tiene otro ejemplo. Se ha examinado algunos reconocimientos sísmicos a lo largo de una nueva área y se tienen identificadas 20 anomalías del tamaño suficiente para ser consideradas viables prospectos de exploración. Por comparación, en un área geológica similar se estima que el 70% de las anomalías son secos, 15 % tiene reservas menores que 100 millones de barriles, 10 % tienen reservas de 100 millones de barriles a 500 millones de barriles, y solamente 5 % de las anomalías tienen reservas de 500 millones a 1000 millones de barriles.

Si se perforan seis de las anomalías que tienen probabilidad $\{a\}$ seis serán todos secos, $b\}$ solamente se tendrá un campo grande y dos campos pequeños, $c\}$ 3 campos de reservas menores de 100 millones de barriles, y un campo con reservas en el rango de 100-500 millones de barriles?

Primero se puede definir N como 20 y $n = 6$, el número de pozos existentes perforados en el programa de exploración multipozo. Se puede también asignar los subíndices, categorías y se calculan los términos "d" como se muestra en la tabla 4.

Note que en la columna 3 se obtiene el término "d" multiplicando el número total de prospectos en el área ($N = 20$) por la proporción del total estimado de cada categoría. Se obtienen estas proporciones por analogía al área cercana. Del área cercana estimada había un 5% de probabilidad de que las reservas en el rango de 500-1000 millones de barriles (por ejemplo la porción derecha de una distribución de reserva), pero esta fracción relativa o porcentaje no es suficiente. Se tiene también que decir 5% del total de los prospectos tendrán reservas de esta cantidad. Por lo tanto $N = 20$ es la proporcionalidad vital constante mencionada antes, con respecto a las últimas distribuciones de reserva.

Tabla 4.

Cálculo del número de pozos correspondientes a cada categoría.

Subíndice	Categoría correspondiente	Término "d", número de elementos originalmente en cada categoría
1	Anomalia seca	$20 \times 0.70 = 14 \leftarrow d_1$
2	Anomalia con reservas <100 MM bis	$20 \times 0.15 = 3 \leftarrow d_2$
3	Anomalia con reservas en el rango de 100-500 MM bis	$20 \times 0.10 = 2 \leftarrow d_3$
4	Anomalia con reservas en el rango de 500-1000 MM bis	$20 \times 0.05 = 1 \leftarrow d_4$

Finalmente, contestar a las preguntas específicas:

- a. La probabilidad de que los seis sean secos.
 Para este caso $x_1 = 6, x_2 = 0, x_3 = 0, x_4 = 0$
 Resolviendo la ecuación 2 se tiene:

Probabilidad de

$$\left. \begin{array}{l} x_1 = 6 \text{ secos} \\ x_2 = 2 \text{ descubrimientos en} \\ \quad <100 \text{ MM.} \\ x_3 = 0 \text{ descubrimientos en} \\ \quad 100-500 \text{ MM.} \\ x_4 = 0 \text{ descubrimientos en} \\ \quad 500-1000 \text{ MM.} \end{array} \right\} = \frac{(C_6^{14})(C_0^3)(C_0^2)(C_0^1)}{(C_6^{20})} = 0.0775$$

en $n = 6$ pozos perforados.

(redondeado a 4 dígitos)

- [Nota: Recuerde (C_x^n) está definido como $\left[\frac{n!}{x!(n-x)!} \right]$, donde $n!$ es el producto de todos los enteros incluso n , y $0!$ está definido como 1.]

- b. La probabilidad de un campo grande ($x_4 = 1$), 2 campos pequeños ($x_2 = 2$) y el resto seco (por ejemplo $x_1 = 3, x_3 = 0$). Así se soluciona la misma ecuación pero con valores diferentes del término "x".

Probabilidad de

$$\left. \begin{array}{l} x_1 = 3 \text{ secos} \\ x_2 = 2 \text{ descubrimientos en} \\ \quad <100 \text{ MM.} \\ x_3 = 0 \text{ descubrimientos en} \\ \quad 100-500 \text{ MM.} \\ x_4 = 1 \text{ descubrimiento en} \\ \quad 500-1000 \text{ MM.} \end{array} \right\} = \frac{(C_3^{14})(C_2^3)(C_0^2)(C_1^1)}{(C_6^{20})} = 0.0282$$

en $n = 6$ pozos perforados.

Finalmente, la pregunta c) concerniente a la probabilidad de que en los seis pozos perforados se encuentren 3 campos con reservas < 100 millones de barriles (por ejemplo $x_2 = 3$), un campo con un rango de reservas de 100-500 millones de barriles ($x_3 = 1$), y un campo con reservas en un rango de 500-1000 millones de barriles ($x_4 = 1$). Esto significa por supuesto, que una de las seis anomalías tendría que ser seca ($x_1 = 1$). Otra vez resolviendo la ecuación 2 se tiene:

Probabilidad de

$$\left. \begin{array}{l} x_1 = 1 \text{ seco} \\ x_2 = 3 \text{ descubrimientos en} \\ < 100 \text{ MM.} \\ x_3 = 1 \text{ descubrimiento en} \\ 100-500 \text{ MM.} \\ x_4 = 1 \text{ descubrimiento en} \\ 500-1000 \text{ MM.} \end{array} \right\} = \frac{(C_1^{14})(C_3^3)(C_3^2)(C_1^1)}{(C_6^{20})} = 0.0007$$

en $n = 6$ pozos perforados.

Estos dos ejemplos dan una idea de cómo el método de análisis se basa en la condición de eventos dependientes y puede ser aplicado a los programas de perforación multipozo. La ecuación 2 es tediosa de resolver, pero es la única alternativa si se quiere que el análisis de probabilidad sea consistente "probar sin reemplazo" la naturaleza de la perforación de una serie de pozos de exploración.

Como un comentario para este trabajo se tiene que saber cuántos de los elementos del espacio de muestra existen en cada categoría antes de perforar n pozos. Es decir, se tiene que poder estimar o calcular el d_1, d_2, \dots etc, términos. Se puede hacer esto en el ejemplo porque se multiplican las proporciones en cada rango de reserva (como se observa en el área cercana) por N , el número total de prospectos en el área.

En general se encuentra que cuando se aplica esto al mundo-real la estimación de N puede ser un problema muy tedioso. A lo largo del ambiente estructural sísmico probablemente se determinaría un número de anomalías más grandes que el número de prospectos no perforados de interés y no el número total de prospectos en la cuenca.

Los mismos comentarios también se aplicarían en áreas estructurales costa-dentro con la excepción de que la exploración y costos de desarrollo son inferiores y se pueden probablemente ver proporciones de estructuras más pequeñas. Esto probablemente resultaría en N más grande. En una cuenca o área donde el mecanismo de entrapamiento es puramente estratigráfico probablemente no se podría enumerar N para control sísmico y/o geológico. Se tiene que calcular o extrapolar un valor de un área análoga perforada. Tal que la extrapolación pueda ser en la forma de un número de los prospectos por 100 millas cuadradas, o un número de prospectos por milla cúbica de estrato sedimentario.

Estas relaciones deberán ser multiplicadas por el área correspondiente o volumen de la cuenca nueva por obtener por lo menos un orden de cálculo de magnitud de N . En estos tipos de ensayos, sin embargo, N es un número difícil de cuantificar, y el analista requiere hacer un análisis de sensibilidad que use varios valores de N para ver su efecto en las estrategias completas de inversión.

Al principio de esta sección se hacen tres preguntas acerca de la hipótesis de que el perforar una serie de pozos es una serie de acontecimientos dependientes. El experimento de cartas discutido,

junto con los dos ejemplos numéricos ilustra la noción de eventos dependientes. ¿Se puede decir que una serie de pozos en el mundo-real es dependiente o independiente? La respuesta es que siempre será dependiente excepto para las posibles ocasiones registradas en la porción derecha de la tabla 2.

La regla general es que una serie de pozos de exploración es una serie de eventos dependientes. Se puede seleccionar el decir si son dependientes o independientes, pero si se elige usar como posibilidad el modelo más complicado se usa la ecuación 2 hipergeométrica. La tercera pregunta es qué sistema se usa.

Las probabilidades numéricas calculadas para cada sistema serán diferentes, particularmente si N es del mismo orden de magnitud como n, el número existente de pozos perforados. Como N se hace muy grande (relativamente a n) las diferencias se hacen mínimas, y los resultados de cada método serán equivalentes. Si se desea basar las decisiones en resultado de probabilidades que sean representativas con la secuencia de eventos, se debe elegir el método dependiente de prueba sin reemplazo.

II.4.2. Análisis basados en la Condición de Eventos Independientes.

Mientras que se acaba de demostrar que la realidad de la perforación de una serie de pozos de exploración es un acontecimiento dependiente, hay varias ocasiones especiales donde el sistema paralelo de eventos independientes se aplica y es registrado en la tabla 2 e incluye el caso donde cada pozo exploratorio en el programa de n pozos está en una cuenca diferente, y el caso donde estratigráficamente varía tan rápidamente (en una base areal) que cada pozo perforado es casi un yacimiento separado. Estos son casos muy especiales.

Quizás una situación más útil para considerar la serie de pozos como eventos independientes es cuando N es mucho más grande que n - específicamente, cuando N es por lo menos diez veces más grande que n. En estas ocasiones los resultados obtenidos de un evento análogo independiente pueden ser aproximadamente igual al que se habría obtenido usando la distribución hipergeométrica. El motivo para elegir la aproximación es que en general, las ecuaciones de eventos independientes son un trabajo mucho más fácil desde el punto de vista computacional.

Los modelos matemáticos análogos de interés aquí se basaron en la condición de que una serie de eventos son estadísticamente independientes de cualquier otro. Los modelos de probabilidad son la distribución multinomial y un caso especial de distribución multinomial para dos resultados es la llamada distribución binomial. El modelo de probabilidad binomial es usado mucho en estadística y control de calidad, y de los métodos de probabilidad para analizar programas para perforar programas múltiples de pozos (hipergeométrico, multinomial, binomial) son simples y fáciles de usar. Así se discutirá el modelo binomial de probabilidad, como se relaciona con programas para perforar pozos múltiples.

La ecuación binomial de probabilidad es:

$$\text{Probabilidad de } x \text{ sucesos en } n \text{ eventos independientes} = \binom{n}{x} (p)^x (1-p)^{n-x} \quad (3)$$

donde

n = número de pruebas independientes.

x = número de éxitos en n pruebas ($0 \leq x \leq n$)

p = probabilidad de éxito en alguna prueba dada ($0 < p < 1.0$)

$$\binom{n}{x} = \text{combinación de } n \text{ cosas tomando } x \text{ a un tiempo} = \frac{n!}{x!(n-x)!}$$

Puesto que como la ecuación binomial es un modelo de dos-resultados puede solamente emplearse un modelo para distinguir entre éxito o fracaso, descubrimiento o agujero seco. No se puede incluir en el modelo varios niveles de reservas (o niveles de probabilidad) dado que se hace un descubrimiento. Los otros requisitos específicos del modelo son que cada prueba sea un evento independiente (por ejemplo en un proceso basado en probar con reemplazo), y la probabilidad de un éxito o fracaso permanece constante en el tiempo, no importa cuántas veces se repitan los procesos. Por supuesto es una severa restricción a los negocios con aceite la noción de eventos independientes. Pero ninguno al menos puede ser útil para ganar experiencias en ciertos tipos de escenarios en exploración. Una aplicación tal puede ser como la siguiente.

Ejemplo: El personal de exploración evalúa un programa de perforación multipozo en un relativamente nuevo ejercicio de una cuenca en que las trampas estratigráficas son el objetivo de exploración primaria. La cuenca es grande y se cree que el número de trampas posibles (prospectos) en la cuenca N , es grande - quizás del orden de 200 o más. Pero cálculos muy preliminares de algunas correlaciones con un área análoga proporciona la probabilidad de que encontrar reservas es cómo de 0.12.

Los ingenieros han advertido que del programa de 10 pozos de exploración se tendrían que hacer por lo menos dos descubrimientos para alcanzar un retorno aceptable del programa. ¿Cuál es la probabilidad de no tener descubrimientos en 10 pozos? ¿De solamente un descubrimiento en 10 pozos?

Solución: Dado que N excede 200 prospectos en la cuenca y se considera un $n = 10$ pozos perforados en el programa, es claro que $N \gg 10$ n ; por lo tanto se puede emplear el modelo de evento independiente como una aproximación al probar un proceso real sin reposición. Los otros parámetros que se requieren para solucionar este problema son la probabilidad de éxito (descubrimiento), $p = 0.12$, y x , el número de éxitos de interés. Para la primera pregunta lo que interesa conocer es la probabilidad de por lo menos $x = 2$ éxitos en las 10 pruebas, dado que $p = 0.12$.

Se puede leer la respuesta a esta pregunta de tablas de probabilidad binomial acumulativa.

Probabilidad Binomial de $x \geq 2$ éxitos en $n = 10$ pruebas, $p = 0.12 = 0.3417$.

La ecuación se puede resolver para esta pregunta, en lugar de usar tablas de probabilidad. Probabilidad Binomial de $x \geq 2$ éxitos en $n = 10$ pruebas, $p = 0.12$

$$\sum_{x=2}^{x=10} \binom{n}{x} (p)^x (1-p)^{n-x}$$

Hay un grupo de aritmética involucrado para resolver esta ecuación; sin embargo, es mucho más conveniente para trabajar con probabilidades binomiales.

Para la segunda pregunta interesa determinar la probabilidad de no descubrimientos, $x=0$, en los diez pozos. Se puede leer esto también de tablas como:

Probabilidad Binomial de $x=0$ éxitos en $n=10$ pruebas, $p=0.12=0.2785$

Y otra vez, si se prefiere puede solucionarse la ecuación 3 o directamente emplear las tablas. Los parámetros que se usarían en la ecuación para esta segunda pregunta son $n=10$, $x=0$, y $p=0.12$.

Finalmente, la probabilidad de solamente un descubrimiento en 10 pozos sería la solución de la ecuación 3 para los parámetros $n=10$, $x=1$, y $p=0.12$. O se puede leer este valor de tablas como:

Probabilidad Binomial de $x=1$ éxito en $n=10$ pruebas, $p=0.12=0.3798$

Resumiendo este ejemplo se tiene cerca de un 28% de probabilidad de no tener descubrimientos en los 10 pozos del programa, un 38% de oportunidad de solo un descubrimiento, y un 34% de oportunidad de dos o más descubrimientos, requeridos para encontrar la mínima utilidad objetivo.

La compañía después de observar que se tiene solamente un 34% de oportunidad de por lo menos dos descubrimientos en 10 pozos, se hace la siguiente pregunta. ¿Cuántos pozos tendrían que perforarse para tener por lo menos un 70% de oportunidad de por lo menos dos descubrimientos?

En esta pregunta el valor desconocido es n , el número de pozos requerido para una probabilidad de 70% o más de por lo menos tener dos descubrimientos, donde $p=0.12$. Se puede usar el modelo binomial para preguntas de este tipo, también. Todo éso es requerido para determinar probabilidades binomiales para valores sucesivamente más altos de n hasta que la oportunidad de por lo menos dos descubrimientos exceda 70%. Solamente se usan las tablas de probabilidad binomial un número de veces como se muestra en la tabla 5:

Tabla 5.

Cálculo de probabilidad binomial para varios eventos.

n, Número total de Pozos perforados	Probabilidad Binomial de que al menos dos éxitos ($x \geq 2$) en los n eventos, dado $p = 0.12$
$n = 10$	0.3417 ← Calculado previamente
$n = 14$	0.5141
$n = 16$	0.5885
$n = 18$	0.6540
$n = 19$	0.6835
$n = 20$	0.7109

De este análisis se puede concluir que se tendría que perforar $n=20$ pozos o más para tener una oportunidad por lo menos del 70% de dos o más descubrimientos. Analizando este tipo es útil obtener un orden de magnitud de número de pozos exploratorios que serían requeridos para especificar una probabilidad de lograr un objetivo, dado un número de descubrimientos.

Esta lógica básica puede ser extendida a una aplicación posible real. Suponga que se perfora en una cuenca en la cual N , el número total de prospectos en la cuenca, es muy grande, quizás mayor que 200. El estrato objetivo produce gas, y se ha negociado un programa de perforación en el cual se pueden encontrar reservas de gas para producir un gasto específico diario durante los primeros dos años del contrato.

Los ingenieros han estimado que se necesitarán por lo menos tres descubrimientos de gas para alcanzar los gastos mínimos establecidos. No se conoce que oportunidad hay de encontrar gas pero se estima un rango de $p = 0.20$ a $p = 0.40$. El interés es tratar de determinar un orden de magnitud estimada del número de pozos que se tendrían que perforar (y por lo tanto la cantidad de capital de exploración requerido) para tener una oportunidad razonablemente buena (o sea 80 o 90%) de por lo menos tener tres descubrimientos.

Se puede ganar experiencias muy útiles de este problema con el modelo de probabilidad binomial en la misma manera que el ejemplo anterior. Los pasos involucrados incluyen estimar un valor de p (para este ejemplo, $p = 0.20$) y entrando a las tablas acumulativas binomiales para leer la probabilidad de tener por lo menos tres éxitos ($x \geq 3$) para varios valores de n , el número total de pozos. Estos datos son registrados y usados en las gráficas resumiendo datos. Entonces el valor de p se cambia y se hacen otra vez las entradas reiterativas a las tablas mencionadas. Se repite esto para varios valores de p . Se registran los resultados de los pasos en la tabla 6. Estos datos son graficados en la figura 5 que da un resumen útil de los datos obtenidos para el modelo de probabilidad binomial.

Tabla 6.

Probabilidad binomial de tener al menos tres descubrimientos
como una función de "n".

CASO 1		CASO 2		CASO 3	
n, el número de eventos	Probabilidad binomial de $x \geq 3$, dado $p = 0.20$	n, el número de eventos	Probabilidad binomial de $x \geq 3$, dado $p = 0.30$	n, el número de eventos	Probabilidad binomial de $x \geq 3$, dado $p = 0.40$
5	0.0579	5	0.1631	4	0.1792
8	0.2031	7	0.3529	6	0.4557
10	0.3222	9	0.5372	8	0.6846
14	0.5519	11	0.6873	10	0.8327
16	0.6482	13	0.7975	12	0.9166
20	0.7939	15	0.8732	14	0.9602
25	0.9018	17	0.9226	16	0.9817
		19	0.9538		

La interpretación de la figura 5 es como sigue. Suponga que el director quiere alcanzar una probabilidad de 85% de tener por lo menos tres descubrimientos. Si la probabilidad de encontrar gas es alta como 0.40 tiene que, por lo menos, perforar 10 pozos exploratorios. Si la probabilidad de encontrar gas es baja como 0.20 el número mínimo de pozos requeridos tendría que ser 22. Estos valores son leídos de la figura 5 entrando en la escala vertical a un valor de 0.85 y en forma horizontal hacia la derecha a las curvas apropiadas de $p = 0.40$ y 0.20, entonces verticalmente hacia abajo en la escala para encontrar n . Así si se quiere por lo menos un 85% de oportunidad de 3 o más descubrimientos de gas, un mínimo de 10 pozos serían requeridos bajo la peor condición (valor

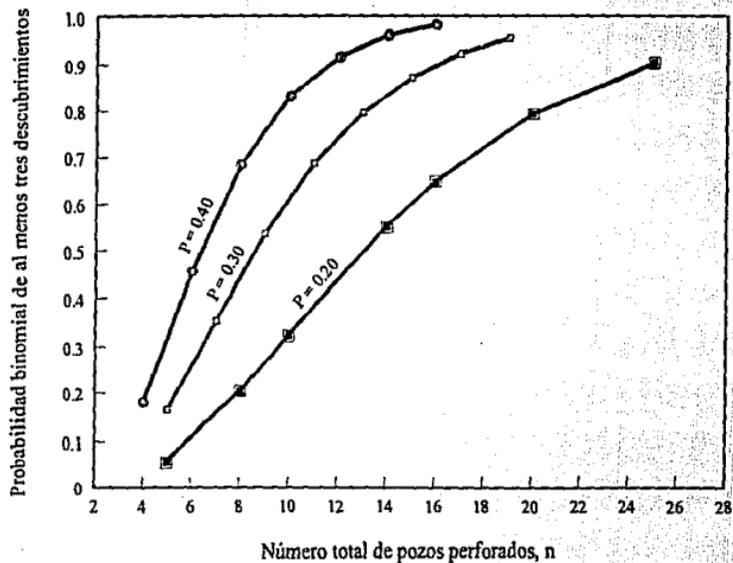


Figura 5. Gráfica de la probabilidad binomial de tener al menos tres descubrimientos ($x \geq 3$) como una función de n , el número total de pozos perforados para varios valores de p , la probabilidad de éxito en cualquier evento.

alto de p) y un mínimo de 22 pozos bajo la mejor condición (valor bajo de p). O alternativamente, se puede usar la gráfica en modo inverso. Suponga que se arregla la financiación para la perforación de solamente 15 pozos. ¿Cuáles son las oportunidades de lograr el objetivo de obtener por lo menos tres descubrimientos?

Para determinar esto se consulta la gráfica en el eje horizontal con un valor de $n=15$, entonces hacia arriba a los varios valores de p, y lee las probabilidades correspondientes de por lo menos 3 descubrimientos en el eje vertical. Para este caso, si perforan 15 pozos se tiene un 60% de oportunidad de por lo menos 3 descubrimientos si $p=0.20$, un 87% de oportunidad de por lo menos tres descubrimientos si $p=0.30$, y un 98% de oportunidad de por lo menos 3 descubrimientos si p es alto como 0.40.

Una gráfica tal como la figura 5 expresa ¿cuál es el valor verdadero de p? Sin duda no. Otra vez los conceptos de análisis de decisión estadístico no pueden decir la información geológica tal como la oportunidad de encontrar gas en esta cuenca particular. Se tiene que especificar cuál p es, o el rango de valores posibles de p. Teniendo especificado esto se puede entonces usar el análisis de decisión, estadísticas y gráficas tales como la figura 5 para ayudar a determinar las estrategias factibles de inversión.

Estos ejemplos dan ideas acerca del uso de probabilidad binomial en el análisis de perforar programas multipozo. Hay que recordar que el modelo binomial se basa en la noción de eventos independientes, y no es una analogía exacta a la perforación de una serie de pozos de exploración. La razón de poder usarla en estos ejemplos es que en ambos casos el número total de prospectos no perforados, N, era relativamente mayor a n, el número de pozos considerado. Y para valores de N por lo menos diez veces más grandes que el de n el modelo de evento independiente (binomial) puede producir probabilidades aproximadamente equivalentes a las que se pueden obtener usando el modelo probado sin reemplazo (hipergeométrico)

El otro evidente defecto del modelo binomial es que puede únicamente relacionar dos posibles resultados. Este defecto puede evitarse si se considera el caso más general del modelo de evento independiente - la distribución de probabilidad multinomial.

La ecuación multinomial de probabilidad es:

$$\begin{array}{l} \text{Probabilidad de } x_1, \\ x_2, x_3, \dots, x_r \text{ resultados} \\ \text{en } n \text{ eventos} \\ \text{independientes} \end{array} = \frac{n!}{x_1! x_2! x_3! \dots x_r!} (P_1)^{x_1} (P_2)^{x_2} (P_3)^{x_3} \dots (P_r)^{x_r} \quad (4)$$

donde:

n = número total de eventos independientes.

r = número posible de resultados.

$x_1, x_2, x_3, \dots, x_r$ = número de veces que ocurren los resultados en n eventos, donde $(x_1 + x_2 + x_3 + \dots + x_r = n)$.

$P_1, P_2, P_3, \dots, P_r$ = probabilidad de ocurrencia de los resultados de cualquier evento independiente, con la condición de que $(P_1 + P_2 + P_3 + \dots + P_r = 1.0)$.

El modelo de probabilidad multinomial esta basado en la condición de que a cada n pruebas hay un acontecimiento independiente y las probabilidades de ocurrencia de cada r resultados posibles,

quedan constantes con el tiempo. Esto está basado en el esquema de prueba con reemplazo, así este modelo no es una exacta analogía del proceso de prueba sin reemplazo, característico en el esquema de perforación de pozos exploratorios.

Pero si N es muy grande con relación al número de pozos considerados, los resultados del modelo multinomial serán aproximadamente equivalentes a las probabilidades obtenidas con el modelo hipergeométrico de probabilidad de la ecuación 2. La ventaja del modelo multinomial (sobre el hipergeométrico) es que puede ser más fácil de manejar desde un punto de vista computacional.

Para citar un ejemplo de cómo esta distribución puede ser usada, se supone que se ha considerado será en el área de una cuenca en el actual el mecanismo de entrapamiento definido por la perforación extensa. La nueva extensión del área es muy grande (en términos de área) pero ha tenido poca exploración porque es substancialmente más profunda y el aumento de los precios del crudo aunado a la escases, han estimulado un interés nuevo en el área profunda.

No se conoce el número de prospectos en el área nueva, N, pero si se extrapola una densidad equivalente de prospectos por unidad de área del área vieja a la nueva, N sería del orden de 150-170. Se considera un programa de exploración de 8 pozos en algunas áreas recientemente compradas. Puesto que N= 150-170 es substancialmente más grande que n= 8 pozos se puede usar la distribución multinomial como una aproximación al proceso de prueba sin reemplazo. Se estima que la probabilidad de descubrimiento en el área nueva sea de 12%

Para obtener una idea de los varios niveles de reservas posibles por campo el personal tiene tabuladas reservas de la vieja y densamente explorada porción de la cuenca, en la tabla 7.

Tabla 7.

Tabulación de varios niveles de reservas posibles.

Rango* de reservas bls/campo	Número de campos con reservas en el rango	Proporción de campos con reservas en el rango	Probabilidad de encontrar reservas en el rango en una area nueva, dado que la oportunidad de encontrar aceite es 0.12
Pequeño	35	35/50 = 0.70	0.70 x 0.12 = 0.084
Medio	10	10/50 = 0.20	0.20 x 0.12 = 0.024
Grande	5	5/50 = 0.10	0.10 x 0.12 = 0.012
	50	1.00	

* Para simplificar aritméticamente, el rango de reserva se divide entre los tres rangos. Se puede tener subdividido el rango de reservas por campo entre el número mayor de intervalos si se designa. Los rangos pueden definirse también numéricamente.

Esta tabulación es, en esencia, el resultado de dividir una distribución de reservas por campo en tres intervalos o rangos y determinar las probabilidades condicionales de cada rango por áreas de lectura debajo de la curva de distribución. Estas probabilidades condicionales son los datos de la columna 3, y están condicionadas a que el aceite sea encontrado en la primera localización. En la

columna 4 se multiplica las probabilidades condicionales por la probabilidad de encontrar aceite, $p=0.12$, para obtener una probabilidad incondicional de encontrar reservas de varias cantidades en el nuevo rango del área. Con estos datos se quiere determinar las respuestas a las siguientes preguntas:

Presuntamente se cree que la información del área más vieja sea representativa de lo que se puede esperar encontrar en el nuevo rango de área, cuál es la probabilidad de que:

- "Todos los ocho pozos sean secos?"
- "Solamente se encuentren las reservas en dos campos pequeños?"
- "El perforar los ocho pozos resulta en el descubrimiento de un campo grande, un campo mediano, y dos campos pequeños?"

Se puede contestar a estas preguntas usando la ecuación 4 de probabilidad multinomial. Si se asigna a los subíndices en x y p términos como 1= agujero seco, 2= reservas pequeñas, 3= campo de reservas mediano, y 4= campo de reservas grandes, se pueden tabular los parámetros necesarios para solucionar la ecuación:

$$\begin{aligned}
 n &= 8 \\
 p_1 &= \text{Probabilidad de pozo seco} = 0.880 \\
 p_2 &= \text{Probabilidad de reservas pequeñas} = 0.084 \\
 p_3 &= \text{Probabilidad de reservas medias} = 0.024 \\
 p_4 &= \text{Probabilidad de reservas grandes} = 0.012 \\
 x_1 &= 8 \text{ prospectos secos} \\
 x_2 &= x_3 = x_4 = 0 \text{ prospectos que tienen aceite}
 \end{aligned}$$

Ahora se sustituye directamente en la ecuación 4 y se calcula la probabilidad de que en ninguno de los ocho pozos se encuentre aceite.

a. Probabilidad de

$$\left. \begin{array}{l} x_1 = 8 \\ x_2 = 0 \\ x_3 = 0 \\ x_4 = 0 \end{array} \right\} \text{resultados} = \frac{8!}{8! 0! 0! 0!} (0.88)^8 (0.084)^0 (0.024)^0 (0.012)^0$$

en $n = 8$ eventos = 0.3596

Para la pregunta b, solamente los términos que cambian son el término x . En este caso $x_2 = 2$, $x_3 = 0$, $x_4 = 0$, y x_1 , el número de agujeros secos = 6. Otra vez se sustituye directamente en la ecuación 4:

b. Probabilidad de

$$\left. \begin{array}{l} x_1 = 6 \\ x_2 = 2 \\ x_3 = 0 \\ x_4 = 0 \end{array} \right\} \text{resultados} = \frac{8!}{6! 2! 0! 0!} (0.88)^6 (0.084)^2 (0.024)^0 (0.012)^0$$

en $n = 8$ eventos = 0.0918

Para la pregunta c, los términos de x que cambian son $x_4=1$, $x_3=1$, $x_2=2$, y $x_1=4$.

c. Probabilidad de

$$\left. \begin{array}{l} x_1 = 4 \\ x_2 = 2 \\ x_3 = 1 \\ x_4 = 1 \end{array} \right\} \text{resultados} = \frac{8!}{4!2!1!1!} (0.88)^4 (0.084)^2 (0.024)^1 (0.012)^1$$

$$\text{en } n = 8 \text{ eventos} = 0.0010$$

Recuerde que la razón de usar solamente la distribución multinomial en este ejemplo (bastante más realista que el hipergeométrico) es que el personal estima que N era mayor comparado con el valor de $n=8$. Para esta condición el producto multinomial da los mismos resultados como si se usara la ecuación hipergeométrica, y es más fácil de calcular. Se tiene también que recordar que la distribución multinomial es un modelo de acontecimientos independientes y las probabilidades de cada resultado (el término p) permanecen constantes con el tiempo.

II.4.3. Análisis de un Ejercicio que Incluya Varios Niveles de Reserva en un Programa Multipozo

En 1961 un trabajo de análisis de riesgo de exploración con multipozos fue presentado (Referencia 7.5) sugiriere un modo de combinar los métodos de análisis de multipozo que se acaban de discutir con las probabilidades de algunos de los descubrimientos que se hacen para producir por lo menos un nivel mínimo especificado de probabilidad.

Al usar este método el analizador especifica el número de pozos exploratorios que serán considerados, sus costos, algunos costos sísmicos y venta de bonos que se tienen que recuperar, el nivel mínimo de probabilidad deseado, y la distribución de describir las reservas dentro de un rango posible por campo que pudiera ocurrir, dado un descubrimiento.

Teniendo especificados estos parámetros el método puede ser usado para determinar la probabilidad que resulta de perforar el programa multipozo y producir al menos el nivel mínimo de rentabilidad. Como tal puede ser una técnica útil de análisis y comparar alternativas de programas múltiples de exploración de pozos.

El mismo método de análisis tiene una serie de cálculos preliminares que requieren producir los datos numéricos para el cálculo final. Se presenta este método para aquellos exploradores que obtienen estos cálculos preliminares y pierden de vista todos los pasos restantes, hay que ser capaz de todo. Para tratar de evitar que esto pase se ha preparado un diagrama ilustrado, o mapa, del método completo de análisis en la figura 6. Así antes que se discuta una aplicación numérica se muestra cómo se usa este método, se tiene probablemente que estudiar el diagrama de la figura 6 con un poco más de detalle.

En éste método se especifica el número de existencia de pozos de exploración considerados en el programa múltiple de pozos, el costo por perforar un pozo C, la cantidad de dólares del valor presente neto (VPN_1) que se quiere recobrar además de una tasa de retorno igual a i_0 , descontando el valor de reservas expresado en una base por barril, y la distribución de valores posibles de reservas por campo que se espera en el área de exploración considerada.

ESPECIFICACION INICIAL DE PARAMETROS PROPUESTOS PARA EL PROGRAMA DE PERFORACION

n = Número de pozos exploratorios considerados en el programa.
 C = Costos de perforación exploratoria por pozo.
 VPNI = Gastos iniciales (tiempo cero) por venta de bonos y/o costos por sismica los cuales pueden ser recuperados.
 D = VPN descontado compuesto de reservas descubiertas expresado en una base por barril y usando una tasa de retorno de i.
 Distribución de reserva = La distribución de reservas por campo que puede esperarse en el área en exploración (Puede ser cualquier tipo de distribución).
 Objetivo = El nivel mínimo de utilidad deseado del programa de exploración de n-pozos.

(PASO I)

Calcular la probabilidad de x descubrimientos en n eventos.

$$(0 < x < n)^*$$

* Si el No. hipotético de prospectos productivos es < n entonces el rango para x será de 0 al No. máximo de estructuras productivas.

(PASO II)

Calcular R_{min} el tamaño promedio del campo para alcanzar el objetivo de utilidad, dado x descubrimientos.

$$R_{min} = \frac{(n)(C) + VPNI}{(x)(D)}$$

(Ecuación 5)

(PASO III)

1. Graficar la distribución de reserva como una frecuencia acumulativa.
2. Desarrollar las distribuciones de R el tamaño promedio del campo dado x descubrimientos para valores de x > 2. Graficar esas distribuciones como frecuencias acumulativas.

X, Número posible de descubrimientos en un programa de n-pozos.	Probabilidad de X descubrimientos en n pozos P(X) (Paso I)	R _{min} el tamaño promedio del campo para alcanzar el objetivo de utilidad, dado x descubrimientos. (Paso II)	Probabilidad de que el tamaño promedio del campo descubierta, R sea al menos igual a R _{min} . P(R > R _{min}). (leído de las gráficas del Paso III)	Probabilidad conjunta. [P(X)][P(R > R _{min})] (col. 2 x col. 4)
0	1.00			
1				
2				
3				
⋮				
⋮				
				$\Sigma [P(X)][P(R > R_{min})]$

PROBABILIDAD DE ALCANZAR EL OBJETIVO DE UTILIDAD

Figura 6. Método de análisis que combina las probabilidades de resultados del programa multipozo de exploración con las probabilidades de que las reservas descubiertas, sean de suficiente magnitud como para alcanzar el objetivo de utilidad.

El objetivo de utilidad está expresado como un valor presente neto positivo suficiente para recuperar el costo de perforar un pozo exploratorio, iniciar desembolsos para arrendar y sísmica (VPN_1) con una tasa de retorno igual a i_0 , la tasa de descuento se usa para calcular D. El método entonces producirá la probabilidad de lograr la ganancia declarada como objetivo con el programa n-pozos. Esta probabilidad es la suma de las probabilidades en conjunto de la columna 5 del cálculo final mostrado en la figura 6. La suma de la columna 5 en el lado inferior derecho pasa al rincón de la figura 6 el cual es el término de probabilidad que se trata de determinar por este método.

En cuanto a los parámetros iniciales, n y C son términos evidentes y no tienen que requerir más explicación. VPN_1 es la cantidad de dólares del valor presente neto en exceso de una tasa de retorno igual a i_0 que se desea recobrar del descubrimiento anticipado, del programa de exploración de n-pozos. Puede estar fijo a algún nivel y puede incluir gastos iniciales de exploración de venta de bonos, y programas sísmicos. Si estos gastos no se hacen en el programa considerado, el VPN_1 puede ser fijo igual a cero.

El término D, es decir el valor presente neto descontado de reservas, es calculado y se expresa el parámetro en un valor presente neto de reservas después de la consideración del costo de pozo de desarrollo e impuestos.

El aspecto de este método se hace con respecto al único modelo anterior de probabilidad que se ha discutido en esta sección en programas de pozo múltiple en el cuál se especificó la distribución esperada de reserva. Los exploradores pueden usar algún tipo de distribución que sientan aplicable (lognormal o de otra clase) y la distribución puede tener algún rango de valores.

La variable aleatoria de esta distribución de reservas es por campo, o si se prefiere, reservas por proyecto. El método puede ser usado en un yacimiento de aceite o un yacimiento de gas. Y si existe incertidumbre considerable en cuanto a la forma y rango (límites) de la distribución de reserva el analizador sin duda querrá hacer el análisis usando varias distribuciones posibles para averiguar la sensibilidad de varias distribuciones factibles de reserva en la estrategia completa de exploración.

Las partes preliminares o intermedias del método consisten de tres pasos separados "preparativo de datos", como se muestra en la figura 6. Paso I simplemente un cálculo de las probabilidades de varios números de descubrimientos en un programa de perforación de n-pozos usando las técnicas anteriormente discutidas en esta sección en programas de pozo múltiple.

Se resumen estos métodos en la tabla 2, y consisten en los modelos hipergeométrico, binomial, y multinomial dependiendo de si se considera que la serie de pozos es una serie de acontecimientos dependientes o acontecimientos independientes. Para las razones esbozadas anteriormente la palabra real usada será el modelo de "evento-dependiente" (hipergeométrico). Habrá unas ocasiones, sin embargo, en que el modelo evento independiente puede dar una aproximación adecuada y en el cuál los cálculos pueden ser algo más sencillos.

El paso II introduce en el análisis el concepto de cómo mayores reservas tienen que ser alcanzadas en el objetivo si se hace $x=1,2,3,..$ descubrimientos en el programa de n-pozos. La relación entre estas dos dimensiones del programa está en la ecuación 5:

$$R_{\min} = \frac{(n)(C) + VP_{N_1}}{(x)(D)} \quad (5)$$

donde R_{\min} está definido como el tamaño mínimo de campo promedio requerido para alcanzar el objetivo, dado x descubrimientos. Los otros términos se definen en la figura 6.

El término numerador de la ecuación 5 representa el (cero de tiempo) costo del programa del pozo exploratorio perforado (el número de pozos multiplicado por los costos por pozo) más el valor presente neto con una tasa de retorno de i_0 . El numerador sumado tiene unidades de dólares. Dividiendo esta suma entre D , el valor presente neto de reservas por barril, produce el número de barriles que tiene que alcanzar el objetivo. Por ejemplo, se supone planear perforar diez pozos a los costos de \$400,000 cada uno, se había gastado \$1 millón para un bono de firma, y el valor estimado del crudo descubierto es \$1.75 por barril descontado a $i_0 = 15\%$.

Los términos del numerador serán $[(10)(\$400,000) + \$1,000,000] = \$5,000,000$. Dividiendo entre \$1.75 por barril da 2.86 millones de barriles. Entonces, si se hace $x = 1$ descubrimiento este debe de tener por lo menos 2.86 millones de barriles para recuperar los \$4000,000 de costos de perforar, \$1000,000 por bonos, y todavía una tasa mínima de retorno de $i_0 = 15\%$, la tasa de descuento usada para calcular D .

Observe que la ecuación 5 también incluye un término x en el denominador. Este es el término que hace R_{\min} condicional al número de descubrimientos que en realidad se realizó del programa n -pozos. Si se hacen dos descubrimientos todavía se necesita tener hallado un mínimo de 2.86 millones de barriles (para la ilustración del párrafo anterior), pero los campos individuales tienen que ser más grandes que 2.86 millones de barriles.

Solamente se requiere que el promedio de los dos campos tenga que ser más grande que $2.86/2 = 1.43$ millones de barriles. Por lo tanto el parámetro R_{\min} es un promedio por campo, y se hace progresivamente más pequeño cuando el número de descubrimientos aumenta. Esto por supuesto, es lógico. Como el número de descubrimientos aumentan el tamaño promedio de campo puede ser más pequeño, no obstante producirá un total de todos los campos de por lo menos $[(n)(c) + VP_{N_1}]/D$ barriles. La solución de la ecuación 5 para los diversos valores posibles de x genera los datos numéricos de la columna 3 del cálculo final.

En el paso III el método requiere algunas distribuciones en una base acumulativa de frecuencia y un grupo separado de cálculos para generar una serie de distribuciones de los tamaños promedio de campo, dados x descubrimientos. Se usan entonces estas gráficas acumulativas de frecuencia para determinar las probabilidades de que el tamaño del campo, dando x descubrimientos, exceda R_{\min} , el tamaño promedio de campo requerido para lograr el objetivo.

Se explicará el Paso III usando los números del ejemplo de los párrafos anteriores. Si se hace un descubrimiento ($x = 1$) se razona que tendría que ser por lo menos 2.86 millones de barriles para lograr el objetivo. "Cuál es la probabilidad que ésto pase?" Bueno, anteriormente se había especificado una distribución esperada de reserva y la probabilidad de que un descubrimiento mayor a 2.86 millones de barriles, para simplemente sería que la cola derecha de la distribución de reserva correspondiente a las reservas de por lo menos 2.86 millones de barriles.

Se puede, por supuesto, leer este valor directamente si se ha graficado la distribución esperada de reserva en una base acumulativa de frecuencia. Por lo tanto, la primera parte del Paso III es graficar la distribución de la reserva original hipotética como una frecuencia acumulativa.

El caso de 2 o más descubrimientos provee un problema. Si se hacen 2 descubrimientos se razona el promedio de los dos tamaños de campo que tiene que exceder 1.43 millones de barriles. "¿Cuál es la probabilidad si se hacen dos descubrimientos de que el promedio (aritmético) de los dos excederá 1.43 millones de barriles? "¿Tiene alguna idea?

Sin duda se puede decir una cosa evidente que no es cierta. Esto es, puede entrar a la distribución original de reserva y leer el área al lado derecho de 1.43 millones de barriles. Pero ésto daría la probabilidad de que los campos sean por lo menos 1.43 millones de barriles. - Y eso no es lo que se pregunto!. Se necesita la probabilidad de que el tamaño del campo promedio de dos descubrimientos sea más grande que 1.43 millones de barriles.

El objetivo en este punto es que primero se tienen que desarrollar distribuciones del tamaño del campo promedio de $x = 2$, $x = 3$, $x = 4$, etc. estos descubrimientos, se definen como los valores posibles de reservas por campo en la distribución original de reserva. Al hacer ésto se necesita proseguir con una serie de pasos en la cuál se prueban valores posibles de reservas para grupos de dos campos, de tres campos, etc. Considere por ejemplo que la distribución inicial de reserva tenía un rango de reservas por campo bajo un límite inferior como de 50 M barriles y un límite superior de 15 000 M de barriles. Y se supone más, que se hablan hecho dos descubrimientos cuyas reservas reales tenían 175 M barriles y 4,200 M barriles.

Las reservas totales de estos dos descubrimientos son $175 M + 4,200 M = 4,375 M$ barriles. El tamaño de campo promedio de estos dos descubrimientos es $4375 M \text{ barriles} / 2 = 2187.5 M$ barriles. Este valor representa una muestra, y una combinación de los valores posibles de reservas si se hablan hecho dos descubrimientos. Con un cuidadoso procedimiento de pruebas se puede hacer "muestras" de muchos tamaños posibles de campos promedio para dos descubrimientos.

Se repite entonces éste mismo proceso probando para el caso de $x = 3$ descubrimientos. En esta ocasión se prueban tres valores de la distribución original de reservas. Se agregan estos tres valores de reserva juntos y la suma dividida entre tres para producir un valor posible del tamaño promedio del campo, dados tres descubrimientos. Después de muchas muestras la distribución de estos valores promedio esta determinada y graficada como una frecuencia acumulativa. Se repite entonces el proceso para $x = 4$, $x = 5$, $x = 6$, etc. hasta que x alcance un valor numérico de n o el número máximo de prospectos considerados a ser productivo, lo que sea primero.

Esta técnica especial de probar puede sonar bastante misteriosa hasta el momento pero el concepto será explicado con mucho más detalle en el capítulo siguiente cuando se discute la noción de probar una distribución. Aunque esto puede sonar complicado, se puede hacer extremadamente eficaz y rápido en una computadora y a un costo muy bajo. Se verá en un ejemplo numérico.

Teniendo completo el paso III de la fase "preparativo de datos" de esta técnica, se tiene solamente que combinar todos los datos numéricos en los cálculos de la columna cinco mostrados en la parte baja de la figura 6. La columna 5 junto con los términos de probabilidad, representa la probabilidad de hacer x descubrimientos y al mismo tiempo teniendo que el tamaño del campo promedio de x descubrimientos excedió R_{min} el mínimo tamaño medio de campo requerido por lo menos para alcanzar el objetivo, dado x descubrimientos. Estas probabilidades mutuamente exclusivas son entonces sumadas a todos los valores posibles de x para alcanzar la probabilidad

deseada del número de n-pozos que el programa de exploración produce para alcanzar el propósito u objetivo.

Todo esto suena muy largo y complicado, pero realmente no es tan difícil. El siguiente ejemplo numérico va a ayudar a clarificar los mecanismos del método. Se tiene que recordar que las dos dimensiones importantes de la estrategia de decisión consideradas por este método de análisis son (1) proveer un modo de unir una distribución de reserva con varios números de descubrimientos que son posibles en un programa de n-pozos, y (2) proveer un medio para determinar la probabilidad de por lo menos alcanzar un objetivo declarado del programa total de exploración.

Estas consideraciones hacen un método útil de análisis a la probabilidad hipergeométrica o de modelos binomiales. El tamaño del campo encontrado evidentemente afecta el valor de un descubrimiento. El modelo hipergeométrico puede decir la probabilidad de x descubrimientos en n pruebas, pero la dimensión agregada que se tiene también que considerar si la probabilidad de éstos x descubrimientos serán de tamaño suficiente para alcanzar un objetivo rentable.

Un ejemplo: Suponga que se está considerando 5-pozos exploratorios programados en una cuenca en la cuál el costo de perforación estimado es \$500 000 por pozo. Recién se adquirió los derechos de perforación en el área para la firma de \$200,000. Entonces se hacen corridas sísmicas, con un costo de \$400,000. Así además de los \$2,500,000 a ser gastados para los cinco pozos exploratorios se tiene un adicional de \$600,000 los cuales se desean recuperar de las ventas de cualquiera de los descubrimientos hechos.

Se han identificado 20 prospectos, y por algunas comparaciones con una cuenca cercana, se estima que 30% de los prospectos tendrán aceite. (Esto implica que $N \times 0.30 = 20 \times 0.30 = 6$ de los prospectos hipotéticos no contienen hidrocarburos). El VPN descontado de cualquier reserva encontrada, expresado en una base por barril es \$1.29/ bbl, por barril.

Se calculó esta figura de un programa estimado de producción de los descubrimientos y descuentos al 18%. Las utilidades netas usadas en el descuento incluyen la consideración de los precios del crudo, costo de pozos de desarrollo, operaciones e impuestos. Si se hace un descubrimiento, en magnitud de reservas, esta es estimada aproximadamente a partir de una aproximación lognormal en la cuál habrá un 60% de posibilidad de que la reserva exceda 150 M barriles por campo y una posibilidad del 10% de que exceda 7500 M barriles por campo.

El objetivo es recobrar los costos de la perforación exploratoria (\$2,500,000), los bonos iniciales y desembolsos sísmicos (\$600,000) y hacer por lo menos una tasa de retorno del 18% por año, la tasa de descuento usada en los cálculos del VPN. "Cuál es la probabilidad de que los 5 pozos del programa de exploración resultarán por lo menos en este nivel de utilidad?"

La solución: El primer paso en la solución de este ejemplo es que se tienen que calcular las probabilidades de encontrar aceite en los cinco prospectos en primer lugar. Se tienen que calcular las probabilidades de que todos los cinco prospectos sean secos, y que uno de los cinco tenga aceite. Para obtener estas probabilidades se requiere solucionar la ecuación 2 hipergeométrica de probabilidad usando los parámetros siguientes:

$N=20$, el número total de prospectos no perforados.

$n=5$, el número de prospectos que se probarán en el programa de exploración multipozo considerado.

$r=2$, el número de resultados posibles de cualquier evento ya sea pozo seco o productor de aceite (Los varios niveles de reservas, dado que se hace un descubrimiento, y sus probabilidades de ocurrencia serán contadas para más tarde). Se asignará el subíndice 1 si la estructura es seca y el subíndice 2 para el caso donde el prospecto contiene aceite.

$d_1=14$, el número de prospectos no perforados hipotéticamente sin contenido de aceite (calculados al multiplicar N por la proporción de los prospectos considerados a ser secos, $N \times (1-0.30) = 20 \times (1-0.30) = 14$)

$d_2=6$, el número de prospectos que hipotéticamente contienen aceite, ($N \times 0.3 = 20 \times 0.3 = 6$)

Ahora se tienen que calcular las probabilidades del número posible de varios descubrimientos, x_2 en los cinco pozos perforados.

a. La probabilidad de no descubrimientos:

$$x_2 = 0$$

$$x_1 = n - x_2 = 5 - 0 = 5$$

Resolviendo la ecuación 2:

$$\left. \begin{array}{l} \text{Probabilidad de} \\ x_1 = 5, x_2 = 0 \text{ en} \\ n=5 \text{ eventos} \end{array} \right\} = \frac{(C_5^{14})(C_0^6)}{(C_5^{20})} = 0.1291$$

b. La probabilidad de un descubrimiento:

$$x_2 = 1, \quad x_1 = n - x_2 = 5 - 1 = 4. \text{ Resolviendo la ecuación 2:}$$

$$\left. \begin{array}{l} \text{Probabilidad de} \\ x_1 = 4, x_2 = 1 \text{ en} \\ n=5 \text{ eventos} \end{array} \right\} = \frac{(C_4^{14})(C_1^6)}{(C_5^{20})} = 0.3874$$

c. La probabilidad de dos descubrimientos:

$$x_2 = 2, \quad x_1 = n - x_2 = 5 - 2 = 3. \text{ Resolviendo la ecuación 2:}$$

$$\left. \begin{array}{l} \text{Probabilidad de} \\ x_1 = 3, x_2 = 2 \text{ en} \\ n=5 \text{ eventos} \end{array} \right\} = \frac{(C_3^{14})(C_2^6)}{(C_5^{20})} = 0.3522$$

d. La probabilidad de tres descubrimientos:

$$x_2 = 3, \quad x_1 = n - x_2 = 5 - 3 = 2. \text{ Resolviendo la ecuación 2:}$$

$$\left. \begin{array}{l} \text{Probabilidad de} \\ x_1 = 2, x_2 = 3 \text{ en} \\ n=5 \text{ eventos} \end{array} \right\} = \frac{(C_2^{14})(C_3^6)}{(C_5^{20})} = 0.1174$$

e. La probabilidad de cuatro descubrimientos:

$x_2 = 4$, $x_1 = n - x_2 = 5 - 4 = 1$. Resolviendo la ecuación 2:

$$\left. \begin{array}{l} \text{Probabilidad de} \\ x_1 = 1, x_2 = 4 \text{ en} \\ n=5 \text{ eventos} \end{array} \right\} = \frac{(C_1^{14})(C_4^6)}{(C_5^{20})} = 0.0135$$

f. Finalmente, la probabilidad de cinco descubrimientos:

$x_2 = 5$, $x_1 = n - x_2 = 5 - 5 = 0$. Resolviendo la ecuación 2:

$$\left. \begin{array}{l} \text{Probabilidad de} \\ x_1 = 0, x_2 = 5 \text{ en} \\ n=5 \text{ eventos} \end{array} \right\} = \frac{(C_0^{14})(C_5^6)}{(C_5^{20})} = 0.0004$$

El siguiente paso de la fase "preparación de datos" (Paso II), es calcular el tamaño mínimo promedio de campo R_{min} , que es requerido para lograr la utilidad objetivo. Al hacer ésto se soluciona la ecuación 5 usando los parámetros siguientes:

$n = 5$, número de pozos exploratorios que han sido perforados.

$C = \$500,000$; costos de los pozos exploratorios.

$VPN_1 = \$600,000$; valor presente neto deseado además de hacer una tasa de retorno de "i₀" y recuperar los costos de perforación, venta de bonos y costo sísmico.

$D = \$1.29$ por barril; valor presente neto de las reservas descontadas al 18% y expresadas en una base por barril.

Resolviendo la ecuación 5 para valores de $x = 1, 2, 3, 4$ y 5 se obtienen los valores de R_{min} en la tabla 8:

Tabla 8.

Cálculo del tamaño mínimo del campo (R_{min}) para varios descubrimientos.

x , Número posible de descubrimientos en los $n=5$ pozos del programa	R_{min} , el mínimo tamaño promedio del campo requerido para alcanzar el objetivo (ecuación 5) $R_{min} = \frac{(5)(\$500,000) + \$600,000}{(x)(\$1.29/b)}$
1	2,400,000 bls = 2,400 M bls
2	1,200,000 bls = 1,200 M bls
3	800,000 bls = 800 M bls
4	600,000 bls = 600 M bls
5	480,000 bls = 480 M bls

Estos resultados significan en palabras que, si se hace un descubrimiento, tiene que ser de por lo menos 2,400 Mbls para alcanzar la utilidad objetivo; si se hacen 3 descubrimientos el promedio (aritmético) de los tres tamaños de campo tiene que exceder a 800 Mbls, etc. Estos cálculos de

valores de R_{min} ocurren en la columna 3 del último cálculo. Lo último en el "preparativo de datos" es graficar las frecuencias acumulativas de la distribución inicial de reserva y las distribuciones del tamaño promedio del campo para $x = 2, 3, 4$ y 5 descubrimientos. De estas gráficas entonces se pueden leer las probabilidades de que los tamaños promedio de campo, R , excederá R_{min} para los varios valores de x . Estos son los números de la tabulación final columna 4.

En la descripción de este ejemplo numérico se declaró que la distribución de reserva era aproximadamente lognormal con el 60% de probabilidad de reservas que exceden 150 Mbls. por campo y un 10% de probabilidad de exceder 7500 Mbls. por campo.

Graficando estos dos porcentajes acumulativos y los valores de reserva en una gráfica con escala logarítmica en la probabilidad, se proporcionan los puntos necesarios para construir la distribución lognormal de reserva en una frecuencia acumulativa base, como se muestra en la figura 7. (La gráfica de distribución de una frecuencia acumulativa lognormal es una línea recta en papel logarítmico, y solamente se requieren esos dos valores de la variable aleatoria y su correspondiente porcentaje acumulativo para construir esta línea recta.)

Para representar una prueba de la distribución de reserva para grupos de $x = 2, 3, 4$ y 5 descubrimientos se pueden determinar las correspondientes distribuciones del promedio de reservas por campo dado x descubrimientos. Estas distribuciones tienen solamente graficada una frecuencia acumulativa base en la figura 7 y se identifica a lo largo de la porción baja de las curvas como 5, 4, 3 y 2 -representando las distribuciones del tamaño promedio del campo para el caso de $x = 5, 4, 3$ o 2 descubrimientos respectivamente.

Hay que notar que mientras la distribución original de reserva en este ejemplo era lognormal las distribuciones de las reservas promedio por campo no son lognormal, aunque los valores de reserva usados en el cálculo de los promedios fueron probados de una distribución lognormal.

Ahora se ha completado el paso tres "preparación de datos" de este método de análisis, y se puede proceder al cálculo final de la columna cinco como se muestra en la tabla 9.

Tabla 9.

Cálculo de la probabilidad de alcanzar el objetivo de retorno VPIN de $\{(n)(C) + VPIN\}$ más una tasa de retorno de i_0 % para el problema del ejemplo (5 pozos del programa de exploración, $C = \$500,000$ por pozo, $VPIN = \$600,000$ y $i_0 = 18$ %)

x, posible número de descubrimientos en un programa de n-pozos	Probabilidad de x descubrimientos en n pozos $P(x)$ (Paso I)	R_{min} tamaño promedio del campo requerido para alcanzar el objetivo dado x descubrimientos (Paso II)	Probabilidad de que el tamaño promedio del campo, R, sea al menos igual a R_{min} . $P(R > R_{min})$ (de la Figura 7)	Probabilidad conjunta $[P(x)][P(R > R_{min})]$ col. 2 x col. 4
0	0.1291	-	-	-
1	0.3874	2400 M bls	0.20	0.0775
2	0.3522	1200 M bls	0.41	0.1444
3	0.1174	800 M bls	0.57	0.0669
4	0.0135	600 M bls	0.71	0.0096
5	0.0004	480 M bls	0.81	0.0003
	1.0000			0.2987

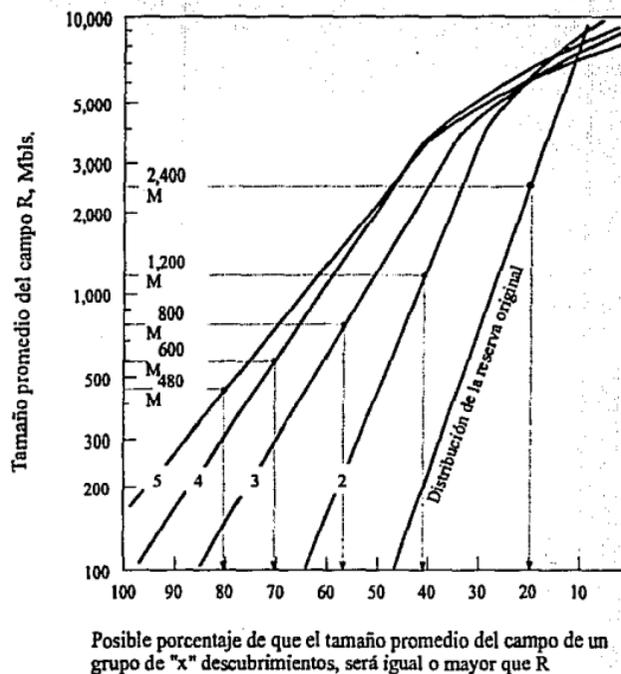


Figura 7. Gráfica de frecuencia acumulativa de la distribución de reserva usada para el ejemplo y las distribuciones del tamaño promedio del campo, dado $x=2,3,4$ y 5 descubrimientos.

Los datos para las columnas 2 y 3 se calcularon resolviendo el modelo hipergeométrico y la ecuación 5 respectivamente. Las probabilidades de la columna 4 son leídas de la gráfica acumulativa de frecuencia de la figura 7. Para el caso de $x=1$ descubrimientos el valor leído es 0.20 como la probabilidad de las reservas por campo la cuales excederán (R_{min}) de 2400 Mbls. (Se entra al eje vertical a $R=2400$ Mbls. como se muestra en la figura 7, se lee a través de la línea de frecuencia de distribución original de reserva, y hacia abajo hasta el eje horizontal).

Si se hacen dos descubrimientos del tamaño del campo promedio se tiene que exceder 1200 Mbls. Se entra al eje vertical a 1200 Mbls. se lee a través a la curva acumulativa de frecuencia correspondiente a $x=2$ descubrimientos, y hacia abajo al eje vertical se lee 41%. El resto de valores de la columna 4 son leídos de la figura 7 de manera semejante.

El paso final en la tabla 9 es multiplicar las probabilidades de las columnas 2 y 4, y sumar los términos del producto. El término 0.1444 en la columna 5 representa la probabilidad conjunta de hacer dos descubrimientos en el programa de exploración de cinco pozos y al mismo tiempo el tamaño promedio de campo de aquellos dos que serán más grandes que $R_{min}=1200$ Mbls., lo mínimo que se requiere para alcanzar utilidad del objetivo. Cada una de las entradas en la columna 5 representa la probabilidad de una forma mutuamente exclusiva de reunión de la utilidad objetivo, y la suma de estas probabilidades mutuamente exclusivas es la probabilidad total del objetivo. Por lo tanto, se puede concluir de este ejemplo que se tiene un 30% de oportunidad (29.87% más precisamente) de recuperar los costos de perforación de \$2,500,000 el costo sísmico y bonos de venta de \$600,000 y todavía tener una tasa de retomo del programa de exploración de cinco pozos de $i_o=18\%$, QED. Esto significa, por supuesto, que se tiene un 70% de oportunidad de no lograr la utilidad declarada. Este ejemplo tiene que ser útil para ilustrar cómo el concepto puede ser aplicado al análisis del programa de exploración de pozos múltiples. Como comentario final se tiene que observar que hay varias modificaciones que pueden hacerse al método fundamental de análisis para ciertos tipos de opciones de exploración.

Los términos de probabilidad de la columna 2 (probabilidades de x descubrimientos en n pruebas) pueden ser calculadas usando en vez del modelo binomial de probabilidad el hipergeométrico si satisface el análisis, el perforar la serie de n pozos, aproximadamente una serie de acontecimientos independientes. Y el término D , el valor presente neto descontado de las reservas expresadas en una base por barril, puede considerarse que varía con respecto a x más que si es tratado como una constante como en el ejemplo.

La razón única por la que se quería hacer D una función de x deberá ser para explicar la posibilidad de operar muchos campos pequeños que no pueden ser tan rentables como operar uno o dos campos más grandes. Con este esquema, valores diferentes de D serían usados para resolver la ecuación 5 para R_{min} , dependiendo del valor de x . Normalmente se espera que D disminuya cuando x se incremente. Esto resultará en R_{min} sea más alto (como x se incremente) que los R_{min} valores que serían calculados usando un valor constante de D .

II.4.4 Tres Niveles Estimados de Riesgo

A través de los años el enfoque usual en el análisis de prospectos ha sido calcular el "promedio" o valor "más probable" de probabilidad para un descubrimiento. El valor simple de probabilidad es calculado usando un simple valor "promedio" de espesor, y valor "promedio" del factor de recuperación, "promedio" de los costos de perforación, precio de la venta de crudo más probable.

Todos estos valores se combinan para cada parámetro y se produce un valor llamado "promedio", o "más probable".

Entonces se compara este valor contra los costos de un agujero seco de una manera arbitraria y/o subjetiva para establecer la conveniencia relativa del prospecto perforado.

Las limitaciones de tal análisis son imparcialmente evidentes. Primero solamente se presenta a la compañía dos niveles discretos de valores -un agujero seco y un promedio, o un valor más probable, dado un descubrimiento. Y fácilmente se puede convenir en que hay muchos, muchos niveles posibles entre estos dos valores discretos.

Otra limitación crítica es que esos valores simples de espesor neto productivo, porosidad, factor de recuperación, costos de perforación, precios del crudo, etc, cuando se combinan producen un solo valor que puede o no representar un valor que es "más probable".

Por lo tanto, en un ensayo se puede definir más niveles discretos posibles que han propuesto exploradores describiendo tres valores de cada variable (más que un promedio, o valor más probable) y entonces calcular toda la combinación de formas posibles de los tres valores de cada variable que pueden ocurrir con las otras variables.

Típicamente los tres valores especificados fueron clasificados como un valor mínimo, más probable, y máximo. Tal enfoque es definitivamente un paso en la dirección correcta en la que se provee a la compañía más información acerca del rango de niveles posibles de rentabilidad que se esperarían si el prospecto es perforado. Esta aproximación generalmente se llama tres-niveles estimados de riesgo.

Para ilustrar cómo trabaja el método se supone que se tiene una oportunidad de inversión en la que las utilidades es una función de tres variables independientes A, B y C. La relación algebraica que involucra las variables es Utilidad = (A) (B) (C). Ahora se supone que se pueden especificar tres valores discretos de cada variable aleatoria - un valor mínimo o bajo, un valor más probable y un valor alto o máximo. Estos se muestran en la tabla 10.

Tabla 10.

Asignación de valores para tres variables independientes.

Variable	Valor mínimo	Valor más probable	Valor máximo
A	A ₁ = 2	A ₂ = 5	A ₃ = 10
B	B ₁ = 6	B ₂ = 7	B ₃ = 8
C	C ₁ = 2	C ₂ = 8	C ₃ = 20

En este planteamiento la incertidumbre estriba en que todavía no se sabe cuál valor de cada variable ocurrirá de los tres valores posibles. Esto conduce a la posibilidad de que haya muchas combinaciones diferentes de A, B, y C. Veintisiete, en realidad, para este ejemplo. Estos veintisiete valores de beneficio incluyen:

$$\begin{aligned} \text{Beneficio}_1 &= (A_1)(B_1)(C_1) = (2)(6)(2) = 24 \\ \text{Beneficio}_2 &= (A_1)(B_1)(C_2) = (2)(6)(8) = 96 \\ \text{Beneficio}_3 &= (A_1)(B_1)(C_3) = (2)(6)(20) = 240 \\ \text{Beneficio}_4 &= (A_1)(B_2)(C_1) = (2)(7)(2) = 28 \\ \text{Beneficio}_5 &= (A_1)(B_2)(C_2) = (2)(7)(8) = 112 \end{aligned}$$

Si se calculan todos los 27 valores de utilidad representando las combinaciones posibles de A, B, y C y se tabulan los resultados en la tabla 11.

Tabla 11.

Ejemplo de una distribución dentro de un rango de valores de utilidad.

Rango de valores de utilidad	Número de veces caídos en el rango	Porcentaje de veces caídos en el rango
0-200	11	40.7 %
201-400	7	26.0 %
401-600	3	11.1 %
601-800	3	11.1 %
801-1000	0	0 %
1001-1200	1	3.7 %
1201-1400	1	3.7 %
1401-1600	1	3.7 %
	27	100.0 %

Es claro que se tiene más información de los niveles posibles que si se acabara de calcular un promedio, o valor más probable de beneficio de $(A_2)(B_2)(C_2) = 280$. Ahora se puede observar el rango, o límites, de valores posibles de utilidad en adición a la información acerca de las demás probabilidades de varios rangos de utilidad. Y todo lo que se tiene que hacer es especificar tres valores discretos posibles de cada variable (más que un promedio) y se calculan valores de utilidad que pueden surgir de todas las maneras posibles de combinaciones de variables que puedan ocurrir.

Para un prospecto de inversión que tiene r variables hay 3^r combinaciones posibles de valores. En este sencillo ejemplo había solamente tres variables, el número total de valores posibles de utilidad es $3^3 = 27$. Si se había considerado tres valores de cinco diferentes variables aleatorias habrían $3^5 = 243$ combinaciones diferentes.

En términos de árboles de decisión, las 27 combinaciones de valores de utilidad representan los 27 puntos terminales mutuamente exclusivos de un árbol de decisión teniendo una serie de posibles nodos de la secuencia. El primer nodo posible tendría tres ramas correspondiente a los tres valores posibles de la variable B.

Habría un total de nueve puntos terminales aquí. Por último al final de cada una de estas nueve ramas se tiene otro nodo posible que tiene tres ramas correspondientes a los tres valores posibles de la variable C. Esto entonces, resulta en $9 \times 3 = 27$ puntos terminales representando el número de combinaciones posibles de caminos que tres valores discretos de cada una de las tres variables puede ocurrir.

La sencilla ilustración discutida demuestra la lógica general del enfoque. Hay, sin embargo, una consideración más usualmente incluida en el análisis. Las probabilidades relativas de ocurrencia de cada uno de los tres valores de cada variable. En el ejemplo de arriba se supuso que cada valor será igualmente probable de ocurrir. Esto es, $P(A_1) = P(A_2) = P(A_3) = 1/3$, $P(B_1) = P(B_2) = P(B_3) = 1/3$, y $P(C_1) = P(C_2) = P(C_3) = 1/3$. Como resultado los 27 valores calculados tienen exactamente la misma probabilidad de ocurrencia, $1/27$.

Pero suponiendo que el analizador desea "comparar" los tres valores de cada variable de la manera mostrada en la tabla 12.

Tabla 12.

Ejemplo de datos de probabilidad de ocurrencia.

Variable A		Variable B		Variable C	
Valor posible	Probabilidad de ocurrencia	Valor posible	Probabilidad de ocurrencia	Valor posible	Probabilidad de ocurrencia
A1 = 2	0.20	B1 = 6	0.10	C1 = 2	0.30
A2 = 5	0.50	B2 = 7	0.80	C2 = 8	0.60
A3 = 10	0.30	B3 = 8	0.10	C3 = 20	0.10
	1.00		1.00		1.00

Cómo funcionaría el método ahora?

La respuesta es exactamente la misma, excepto que para cada uno de los 3^r valores de beneficio también se podría calcular una probabilidad específica de ocurrencia. Por ejemplo:

$$\begin{aligned} \text{Beneficio}_1 &= (A_1)(B_1)(C_1) = (2)(6)(2) = 24 \text{ (como)} \\ \text{Probabilidad de Beneficio}_1 &= (0.20)(0.10)(0.30) = \underline{0.006} \\ \text{Beneficio}_2 &= (A_1)(B_1)(C_2) = (2)(6)(8) = 96 \text{ (como)} \\ \text{Probabilidad de Beneficio}_2 &= (0.20)(0.10)(0.60) = \underline{0.012} \end{aligned}$$

·
·
·
etc.

La suma de todos los términos de probabilidad 3^r así calculados son 1.0. Con esta modificación, se encuentra la probabilidad de utilidad siendo menor o igual a por decir 100, solamente se suman todos los términos de probabilidad para utilidad de 100 o menos. Y se puede calcular un valor medio de utilidad (EMV) simplemente multiplicando los 3^r valores de utilidad por sus probabilidades correspondientes de ocurrencia y sumando todos los 3^r términos del producto.

Tres niveles estimados de riesgo, el método descrito, claramente da más información acerca de los varios niveles posibles de rentabilidad, más que un solo valor promedio. El analista solo necesita describir dos valores numéricos adicionales de cada variable (además el valor "promedio" que se había venido usando), más los factores de peso o probabilidades de ocurrencia de los tres valores de cada variable. No hay problema en describir los tres valores numéricos.

Realmente, el analista probablemente va a través de un proceso de definir un valor mínimo o bajo, y un valor máximo o elevado, cuando define el promedio o valor más probable. Los factores de peso causan pequeñas molestias al analizador. Tendrá que comparar si el valor intermedio es más probable de ocurrir que cualquiera de los otros dos, etc. Pero para el que probablemente es el caso excepcional, cuando los tres valores son todos igualmente probables de ocurrir esta es una parte necesaria del análisis.

Mientras que tres niveles estimados de riesgo claramente dicen más acerca del prospecto que un sólo valor más probable de utilidad más el costo de un agujero seco, el método tiene que ser reemplazado ampliamente por otro método nuevo de análisis de riesgo llamado simulación. Se hablará más acerca de éste en el capítulo siguiente.

Brevemente, la razón por la cual la simulación es usualmente elegida sobre tres niveles estimados de riesgo es que la simulación permite la oportunidad de considerar la distribución completa de valores posibles de cada variable. No solamente un sólo valor promedio, o tres valores discretos - sino la distribución completa de valores posibles! Así, mientras que el tercer nivel calculado de riesgo da mucha más información que un solo valor promedio, la simulación puede proveer hasta más información que tres niveles de cálculo de riesgo.

Déside un punto de vista histórico el método de tres niveles estimados de riesgo es un precursor de la simulación. Y ahora esa simulación gana aceptación y el uso del programa útil de tres niveles estimados de riesgo probablemente decae algo.

III. TECNICA DE SIMULACION EN ANALISIS DE RIESGO

En este capítulo se discute un enfoque nuevo al análisis de riesgo de exploración llamado simulación. El concepto de simulación permite al analizador la opción de riesgo, describiendo la incertidumbre en forma de distribución de valores posibles tal que los parámetros desconocidos como espesor productivo, recuperación, costos de perforación, etc. pueden tener.

Estas distribuciones se combinan para producir una distribución de los niveles posibles de probabilidad la cuál pueda suponer el proyecto. De tal distribución solamente un pequeño, y último paso es calcular un parámetro esperado de valor para usar en el proceso de hacer la decisión. El método es un modelo continuo de riesgo e incertidumbre, opuesto a los modelos de resultado discreto que se han discutido.

La simulación, como un medio de análisis de riesgo en decisiones de baja incertidumbre, es un proceso imparcialmente nuevo. Fue propuesto para decisiones generales de negocios en 1964, y su uso en inversiones de exploración de aceite empezó a aparecer en la literatura a principios de 1967. Hoy su existencia es usada por casi cada compañía de crudo o agencia mayor, involucrada en la exploración activa de aceite y gas.

El método de simulación tiene varios sinónimos, incluso simulación aleatoria, Simulación Monte Carlo, y el método Monte Carlo. Se tiene que señalar, sin embargo, que la simulación discutida aquí no tiene absolutamente nada que ver con un método de análisis de ingeniería llamado, simulación numérica. Ni lo que se hace tiene nada que ver con cierta computadora programando lenguajes que tienen la palabra simulación en su nombre. El método de simulación combina distribuciones de probabilidad y variables aleatorias.

El sujeto que emplea simulación para análisis de riesgo de exploración de petróleo es imparcialmente liberal. Así probablemente no se podrá discutir en completo detalle cada posible faceta del método.

III.1 La Lógica de Simulación

Antes de sumergirse en detalles es importante, otra vez, dar una visión de conjunto del método y como se relaciona con la exploración del petróleo. La primera razón es hacer un análisis de simulación en el contexto de decisión bajo incertidumbre, definiendo la distribución de utilidad la cual se puede anticipar. Una vez que se sabe ésta distribución se puede fácilmente hallar su valor medio. Y desde que el valor medio de una distribución es, por definición, el mismo valor esperado se tiene entonces que el valor medio de una distribución del valor presente neto descontado a i_0 es también el valor presente neto esperado, o EMV.

Éste es el parámetro de dirección que se puede usar para determinar la viabilidad de la perforación proyectada. Además de ser capaz de calcular el EMV directamente, la distribución de utilidad ofrece numerosas opciones para graficar información presentada al que toma decisiones, acerca del rango y probabilidades de ocurrencia de niveles posibles de utilidad y pérdida. Un cuadro tal como una distribución de probabilidad a veces vale más que miles de palabras.

Así el objetivo de la simulación es determinar la distribución de utilidad de un proyecto o propuesta para perforar. Ahora se toma un paso hacia atrás para ver como la distribución de utilidad viene en primer lugar en el cuadro. Se supone una inversión hipotética en la que la última ganancia neta es una función de tres variables independientes separadas, X, Y y Z. Y se supone la relación que combina estos parámetros juntos es la siguiente:

$$\text{Utilidad} = 3.4X + [(1 / \sqrt{14Y}) + 17.9] [0.4 Z^2] \quad (6)$$

Esto por supuesto es una situación en la que X, Y y Z se llaman variables independientes aleatorias y la Utilidad es la variable dependiente. En los negocios de crudo la relación probablemente tendría el VPN en el lado izquierdo del signo igual y en el lado derecho debería ser la ecuación por la que se calculan las reservas recuperables, convertir las reservas a ventas netas, se sustraen los gastos de operación, costos de perforación, impuestos, derechos, etc. y se descuentan las utilidades netas a t_0 . Las variables independientes incluirían todos los factores empleados para calcular la ganancia VPN: espesor neto productivo, factores de recuperación, área productiva, costos de perforación, precios de crudo, gastos de operación, etc. Pero para esta visión de conjunto se piensa en términos más sencillos, un resumen del modelo tal como en la ecuación 6.

Ahora se supone que se conocen los valores exactos que las variables aleatorias tendrán en una inversión. ¿Se sabe la ganancia resultante? Sin duda; solamente se substituyen el (conocido) 9 valores de X, Y y Z en la ecuación 6 y se soluciona directamente para la utilidad. Esto sería llamado por matemáticos una computación determinista. No hay incertidumbre con respecto a las variables o al valor de la ganancia resultante.

Pero tal caso ideal, nunca existe en análisis de prospectos. Al tiempo de hacer una decisión usualmente no se saben los valores exactos para la mayor parte de las variables aleatorias afectando a la utilidad VPN. Quizás se pueda saber acerca de que rangos de valores posibles están. En este caso ¿cuál valor del rango se usaría en el cálculo de la ganancia?

La respuesta usando el concepto de simulación es que se pueden considerar todos los valores dentro del rango para cada variable aleatoria. Y todos éstos son combinados para producir el rango y distribución de la ganancia. La situación usual en análisis de prospectos de exploración es que no se sabe que valores exactos de cada variable aleatoria se pueden usar, la simulación deriva la distribución resultando valores de ganancia. El escenario, entonces, es como en la figura 8. El modelo que se trata de analizar es la ecuación de utilidad para la inversión hipotética de la ecuación 6.

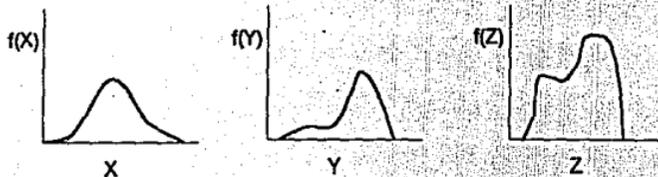
Se dan las expresiones cuantitativas de incertidumbre como distribuciones de valores posibles de las variables aleatorias X, Y y Z. Se tiene que notar aquí que se habla del rango completo y distribución de valores posibles de la variable aleatoria, y no de un promedio justo o valor más probable, o tres valores discretos tal como se discutió en el capítulo pasado bajo el encabezado de tres niveles de riesgo. Se ha considerado el rango completo y continuo de valores posibles de X, Y, y Z.

Se tiene que observar un punto adicional. El estado real será que la variable X tendrá un valor específico, la variable Y tendrá solamente un valor específico, y la variable Z solamente tendrá un valor. Pero el apuro es que antes que la inversión sea aceptada no se sabe que valores tendrán. Se puede hablar acerca de las variables, el rango y probabilidades relativas de valores

Modelo o sistema de Interés } →

$$\text{Utilidad} = 3.4 X - \left[\frac{1}{\sqrt{14 Y}} + 17.9 \right] [0.4 Z^2]$$

Expresiones de riesgo e
incertidumbre:
Distribuciones de los
posibles valores de las
variables aleatorias X, Y y
Z.



Resultado del análisis de
simulación: Definición de
los resultados de la
distribución de la utilidad.

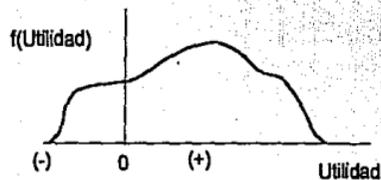


Figura 8. Esquema de Análisis de Simulación de la inversión hipotética de la ecuación 6.

posibles de cada variable.

Definir el escenario en este punto no es nada nuevo. Se sabe del modelo o sistema que se trata de analizar en el análisis de prospecto. Y se puede convenir que usualmente no se sabe, de antemano, el espesor neto productivo real (a pesar de que un geólogo muestre orgulloso un mapa de isopacas); pero se puede estimar un posible rango de valores de, 0 a 100 pies.

El problema es solucionar la ecuación, una vez que se han definido las distribuciones para cada variable aleatoria. No es inmediatamente evidente qué cada una de las tres distribuciones está en la figura 8, y probablemente no es muy evidente como la ecuación puede aún ser resuelta si la variable aleatoria es expresada como distribución, más, que sólo determinados valores.

Aquí es precisamente en donde el análisis de simulación entra al cuadro. Es sencillo y fácil el camino para analizar un prospecto de inversión en donde algunas (o todas) las variables aleatorias son expresadas como distribuciones de probabilidad. Se usa simulación en esta situación, probablemente porque es más evidente que los dos caminos alternativos para resolver el problema a trabajar. Primero se mencionarán estos dos caminos.

Las distribuciones de probabilidad, también se les llama función de densidad de probabilidad. Son las $f(X)$, $f(Y)$ y $f(Z)$ funciones ordinales que, cuando se integran sobre el rango de valores posibles de cada variable se producen dimensiones de 1.0, del área bajo todas las distribuciones de probabilidad. Así una cosa se puede hacer si se tiene un problema tal como el mostrado en la figura 8 sería consultar a un matemático y preguntarle como determinar las funciones de densidad de probabilidad, insertar éstas funciones analíticas en la ecuación y solucionar para la función correspondiente de densidad de la variable dependiente, ganancia.

Una idea fina; pero simplemente no funcionaría porque las complejidades de una solución analítica tal como ésta son totalmente intratables. Mientras que hay algunas situaciones especiales en donde el enfoque analítico puede ser posible, el enfoque analítico será, en general, no factible en análisis de prospecto de exploración de petróleo.

El segundo método trata de simplificar el problema determinando los valores medios de las distribuciones X, Y y Z e insertar estos valores discretos en la ecuación y resolver para la utilidad. La idea deberá ser que este valor de utilidad deberá corresponder al valor medio de la distribución de la utilidad. Y como la media de la distribución y EMV, es similar, entonces se tendría el parámetro deseado para el análisis de decisión. Pero esta lógica no funciona tampoco;

En general, no se puede calcular el valor medio de la variable dependiente para substituir los valores medios de cada distribución de la variable aleatoria en la ecuación. Simplemente no trabajará para análisis de proyectos de exploración. Otra cosa que no funciona es tomar los valores más probables de cada distribución (los modos), substituyendo estos valores en la ecuación y resolver la ecuación en espera de calcular el valor más probable de ganancia. Tanto estos métodos "algo cortos" serán de uso.

De todo esto se dice que se usa la mecánica y lógica de simulación por default, solamente de este modo se puede solucionar el problema. Si se ha definido una oportunidad de inversión como se muestra arriba de la línea horizontal que divide a la figura 8, la única manera de convertir todo esto en una descripción de la distribución del resultado de la ganancia es por

simulación.

En pocas palabras es como se define la distribución de la variable dependiente (ganancia). Se hacen una serie de cálculos iterativos de valores posibles de ganancia. Se calcula cada valor usando un valor de X, Y y Z escogido dentro de sus rangos respectivos de distribución. Cada valor de ganancia calculado en esta manera representa un estado posible de naturaleza, o combinación posible de X, Y y Z. Los valores específicos de X, Y y Z para cada cálculo de ganancia son escogidos para la misma distribución de frecuencia como la distribución original especificada por el analizador. Cada uno de estos reiterados valores calculados de la variable dependiente es llamado un paso de simulación, o más simplemente, un paso. El valor de la variable independiente usado para cada paso es obtenido por muestreo de su distribución original en una manera que honra la forma y el rango de la distribución.

Se continúan estos cálculos iterativos o pasos, hasta que un número suficiente de valores de la variable dependiente están disponibles para definir su distribución. Esto usualmente requiere por lo menos 100 pasos y quizás tantos como 1000 o 1500 pasos. Evidentemente estas muchas soluciones de aritmética de una ecuación significan una cantidad grande de trabajo - y verdaderamente la manera práctica de llevar a cabo en forma rápida, eficaz y con menores costos de tiempo es usando la computadora.

Y éso es todo. Una serie de soluciones reiterativas de la ecuación o modelo de interés - cada vez con valores de la variable independiente que son probados (elegidos) de sus distribuciones respectivas. Después de qué un número suficiente de pasos se ha hecho, el cálculo de valores de la variable dependiente es tabulado como una distribución - el objetivo del análisis.

Por supuesto, hay muchos detalles mecánicos en los que se tendrá que ahondar más tarde, tal como probar valores de la distribución de la variable independiente para cada paso, cuántos pasos se requieren, si son muchos los pasos requeridos, como son las distribuciones de la variable independiente definidas en primer lugar, qué si se relacionan las variables X y Y de alguna manera, etc. Pero la lógica general de simulación es simplemente definir la distribución objetiva por una serie de ejemplos reiterativos.

Este esquema para análisis de riesgo de exploración tiene varias ventajas muy importantes:

- Permite a los exploradores describir riesgo e incertidumbre como un rango y distribución de valores posibles para cada factor desconocido, más que un solo valor medio o más probable. El resultado de la distribución de utilidad reflejará todos los valores posibles de la variable. El método explica el grado de riesgo e incertidumbre visto para cada variabilidad de las variables.
- Puede ser aplicado a algún tipo de cálculo involucrando variables aleatorias. Se habla primeramente de VPN del prospecto a perforar, pero la lógica puede ser aplicada para describir una distribución de reservas recuperables, una distribución de datos de saturación de agua, la posición del agujero de un pozo perforado direccionalmente, etc. La lógica es la misma - solamente cambia la ecuación del modelo de interés y su correspondiente cambio de variables.
- No hay límite al número de variables que puede ser considerado. Si la ganancia es una función de 45 diferentes variables, y se pueden definir distribuciones para cada variable la simulación puede todavía usarse más eficazmente. La contabilidad en la computadora será un poco más grande, y el programa puede requerir algunos

- segundos más de tiempo de maquina - pero ésa es la única desventaja.
- Las distribuciones usadas para definir los valores posibles para cada variable aleatoria, no tienen que ser de una forma específica como lognormal, normal, etc. Más aún, las distribuciones pueden ser del tipo que sea. Si el analizador puede dibujar el cuadro de distribución en papel, eso es todo lo que se requiere. Esto es importante porque el analizador no tiene que ser capaz de describir la distribución, más que con un cuadro.
- La experiencia puede ser usada más eficazmente porque los juicios acerca de la distribución de valores posibles de cada variable pueden hacerse por la mayoría de las personas que saben acerca del parámetro. El geólogo puede definir el intervalo productor neto y distribuciones de áreas productivas, el ingeniero puede definir la distribución de factores de recuperación y programas de producción, y el ingeniero que perfora puede especificar la distribución del costo probable de perforar, etc. El método éste no requiere que una persona provea toda la información al análisis.
- El programa de computadora del análisis de simulación es relativamente corto y para correr un análisis de simulación se requiere muy poco tiempo de computadora. En realidad, la lectura inicial de distribuciones y la impresión de los datos de frecuencia de la variable dependiente usualmente requerirán más tiempo de computadora que los cálculos reales.
- El método presta el mismo análisis de sensibilidad. Verdaderamente uno de los aspectos importantes de hacer un análisis de simulación, es ser capaz de definir uno, dos o tres factores que tengan un efecto más significativo en los valores resultantes de utilidad. Se hacen tales análisis al usar distribuciones diferentes (ya sea en términos de la forma y/o el rango) para cada variable y observar, el rango de la distribución variable dependiente en el cual está cambiando.

Finalmente, se tiene que observar que la parte crítica para los exploradores es definir las distribuciones de cada variable aleatoria. Se hablará de esto en detalle más tarde en el capítulo, pero por el momento se tiene que ver que las distribuciones pueden estar basadas en datos de estadística si están disponibles, por analogía con otras áreas semejantes productivas, o quizás aún por un juicio subjetivo. Si se forza a confiar en opiniones subjetivas las opiniones entran al análisis en forma de distribuciones. Una variable a la vez - un paso a la vez.

Este es el enfoque nuevo mencionado en el capítulo anterior de la discusión de la estimación de la probabilidad subjetiva. Si se tiene que confiar en juicios subjetivos parece razonable esperar que se hará un empleo mejor de lo considerado si es un parámetro a la vez, más que tratar de usar todos los parámetros inciertos que ocurrirán colectivamente en una perforación proyectada.

Teniendo ya una visión del método, ahora se tocarán algunos de los detalles mecánicos involucrados en un análisis de simulación.

III.2 Los Mecanismos de Análisis de Simulación

Esta sección del capítulo es concierne a las "tuercas y pernos" de la técnica de simulación de análisis de riesgo. Hay mucha miscelánea y detalles que se tienen que considerar al organizar el análisis y preparar los datos. Y es importante que el analizador tenga una buena experiencia para comprender y en la cuál poder basarse. Sin esta comprensión, la

simulación puede ser un instrumento peligroso.

Para emplear simulación antes se debe tener un conocimiento general y/o no tener un problema para continuar en estos detalles. Si el concepto es completamente nuevo y la sección anterior parece un poco misteriosa se sugiere hojear esa sección y proceder a la sección próxima de las aplicaciones y ejemplos. Entonces se tiene que seguir por medio de algunos ejemplos para tener un mejor entendimiento de como todos los pedazos se juntan. A este tiempo se puede regresar a ésta sección para perseguir los detalles de cada una de las piezas individuales del análisis.

Si se analiza una perforación proyectada y se quiere cuantificar el grado de riesgo e incertidumbre usando un análisis de simulación se requiere seguir los seis pasos generales:

Paso 1: Defina todas las variables. Se tiene que especificar la medida del valor de interés (ejemplo valor presente neto, después de impuestos), y todas las variables que afectan el valor.

Paso 2: Defina la relación que empaata todas las variables juntas. La relación de interés aquí es una ecuación o serie de cálculos numéricos para calcular el valor de la probabilidad de una perforación proyectada. Esta relación, o serie de cálculos, son el modelo o sistema que se analizará.

Paso 3: Sortear las variables afectando el valor en dos grupos - las variables cuyos valores se saben con certeza y las desconocidas, o variables aleatorias para las cuales no se pueden especificar valores exactos al tiempo de hacer la decisión.

Paso 4: Defina distribuciones para todas las variables desconocidas, o aleatorias.

Paso 5: Haga los pasos reiterativos así como describa la distribución de resultante del valor.

Paso 6: Calcule el Valor Esperado de la distribución de utilidad y prepare el despliegue de gráficos para el procedimiento de análisis y resultados.

Teniendo enumerados los pasos o piezas de un análisis de simulación se observan los detalles involucrados en cada uno de estos pasos.

III.2.1 Definiendo las Variables (Paso 1)

Este paso inicial es el punto evidente para empezar algún análisis cuantitativo de un proyecto de perforación: define la medida del valor de interés y los factores que lo afectan. Típicamente, la medida del valor es elegida después de impuestos VPN descontado a i_0 , sin embargo, se es libre de elegir alguna medida de valor que agrade más (tasa de retorno, ganancia descontada/relación de inversión, etc.) Las variables que afectan el valor de los parámetros son tratados a diario: espesor productivo neto, porosidad, saturación de agua, factores de recuperación, número de pozos a ser perforados, productividad de los pozos o del campo como una totalidad, costos de perforación, gastos de operación, costos de plataforma y oleoductos facilidad si son costa fuera, asociar precios de ventas de crudo y gas, impuestos, etc.

Mientras que el interés principal aquí es el análisis desde el punto de vista del prospecto hay muchas ocasiones en donde existe interés en determinar la distribución de algunas de las dimensiones intermedias tales como reservas. En éstas ocasiones la variable dependiente de interés puede ser las reservas recuperables de la estructura (más que la utilidad), y las variables que afectan a las reservas incluirían solamente los parámetros relacionados con las reservas. Al

final de la sección siguiente se da una breve referencia a algunos de estos esquemas intermedios auxiliares de simulación.

III.2.2 Definiendo el Modelo de Análisis (Paso 2)

Este paso es el seguimiento lógico de la lista de las variables - definiendo la relación o ecuación que involucra todas las variables juntas. Puede llevar la forma de una sola ecuación, varias ecuaciones, o aún una serie de cálculos. Para la inversión Hipotética de la sección inicial de este capítulo la relación era la ecuación 6. En análisis de prospecto es la serie de cálculos por los que se pueden calcular las reservas recuperables, convertir las reservas a ganancias descontadas, sustraer todos los costos por desarrollo y operación y gastos.

Ya que la relación cambiará de proyecto a proyecto (dependiendo en que país del mundo se esté, si el prospecto es costa dentro o costa fuera, etc.) probablemente no serviría con un mismo propósito útil el registrar ejemplos específicos en este punto. La mayoría de los exploradores están íntimamente familiarizados con estos pasos, y la relación específica a usar para la simulación sigue siendo mucho más lógica para cada prospecto.

Como una observación, los primeros dos pasos son fundamentales para cualquier análisis, de simulación o de cualquier otro. En complemento, estos dos pasos son en esencia, los que definen, qué factores y como afectan. Es solamente cuando se procede al Paso 3 que comienza a llevarse una dirección nueva para analizar riesgo e incertidumbre.

III.2.3 Clasificando las Variables en Grupos (Paso 3)

En este paso se tienen que dividir todas las variables que se registraron en el Paso 1 como valor afectado (ganancia) en dos grupos, consistentes de todas las variables o parámetros para saber sus valores exactos. Un grupo consiste de todas las variables o parámetros para las cuales su valor exacto es conocido. El segundo grupo incluye todos los parámetros o variables para los cuales los valores exactos no se saben exactamente al tiempo del análisis. En el análisis del prospecto de exploración las variables afectadas, desafortunadamente, caerán en el segundo grupo!

Ejemplos de algunos de los parámetros que se pueden incluir en el grupo "conocido" son las tasas de impuestos y los precios. Si ya se ha firmado un arrendamiento se pagan los derechos al dueño del mineral - está especificado en el contrato. Y presumiendo que se opera en un campo políticamente estable probablemente se sabrían las tasas exactas para calcular los impuestos debidos.

El grupo "desconocido" de parámetros incluye usualmente todas las variables involucradas relativas al tamaño de la estructura o prospecto, las cantidades de gas y aceite a recuperar, programas de productividad, factores involucrados con costos y futuros gastos de operación, y los futuros precios de crudo y ventas de gas. Estas son las variables aleatorias que serán consideradas en el análisis en la forma de sus respectivas distribuciones.

Cuando se hace esto con algunas de las variables se requiere recordar que cuando se dice de una variable que es "desconocida" significa que no se sabe su valor exacto al tiempo del

análisis. Por ejemplo - espesor neto productivo. Es claramente una variable que afecta las magnitudes de reservas recuperadas, y por lo tanto la ganancia, pero es probablemente una variable que se tiene que clasificar en el grupo "desconocido". Esto es porque no se sabe el espesor productivo neto de la formación objetivo antes de perforar el pozo.

No se puede decir si serán 45 metros de espesor, o 71 metros. Se puede, sin embargo, especificar un rango y distribución de valores posibles de espesor neto productivo supuesto en el proyecto, en la cuál se clasificaría la variable espesor neto productivo, como una variable desconocida o aleatoria. El punto aquí será "desconocido"/"conocido" en este sentido se relaciona con o sin que se pueda especificar el valor numérico exacto de la variable con certeza al tiempo de hacer el análisis existente.

Otro sutil, pero importante punto al clasificar las variables es la pregunta de, si o no se sabe que bastante cerca de una variable alguna vez, se puede describir una distribución de valores posibles. Algunos analizadores dirán, por ejemplo, que los parámetros como factor de recuperación son un factor desconocido, pero puesto que se está en un área de pozos petroleros exploratorios en los cuales no se tienen bases para definir un rango o distribución de valores posibles, se usa un factor de recuperación promedio de aceite, de 15% en el lugar.

Como consecuencia se deberá "conocer" el factor de recuperación en el grupo de variables "conocidas" para el cuál se usan valores específicos en los cálculos sencillos de utilidad. En el sentido estricto de qué se llevará a cabo un análisis de simulación para complementar, tal punto de vista (aunque bastante usado) no es válido. Los enunciados implican la lógica siguiente:

- a. Si se tiene información confiable acerca del factor de recuperación se sabría su valor numérico preciso, es decir RF_1 .
- b. Si se tiene información poco confiable, pero buen control y experiencia acerca de un área correlacionada se puede estimar un rango de valores posibles del factor de recuperación de RF_{MIN} a RF_{MAX} .
- c. Como el control y datos correlacionados se hacen cada vez más pobre y/o menores el número del rango puede agrandarse e incluir un rango mayor de valores posibles de factores de recuperación. El rango más ancho puede ser representativo de una incertidumbre más grande.
- d. Si se está en un área densa de pozos exploratorios y no se haya manera de estimar RF no se tiene una base para determinar un rango y distribución; por lo tanto, se usa un solo promedio, el valor más probable en los cálculos.

Y esta lógica es claramente circular!. Significa que lo menos que se conoce sobre algo es lo más que se conoce de ello!.

El propósito de un análisis de simulación total es poder explicar la variabilidad en el análisis. Incertidumbre significa variabilidad, y esta variabilidad es expresada como un rango y distribución de valores posibles de la variable. Hacerlo de otra manera es ignorar la dimensión del riesgo e incertidumbre de lo que se trata de analizar.

La razón por la que se recalca este punto, es que se observó repetidamente que los exploradores que empiezan a trabajar con simulación, tienen esta equivocación. Cuando definen una variable tal, como factor de recuperación como un solo valor se preguntan si significa, entonces, que saben (de antemano) que valor preciso de recuperación será. Su respuesta es: "No, no sé exactamente cual será, pero no tengo datos disponibles o información

sobre como establecer un rango por lo que solo uso el valor de 15% "se han escuchado tales declaraciones tan frecuentemente que se está convencido que este es uno de los puntos en que más personas cometen fallas cuando usan simulación.

Si no se sabe un valor exacto de una variable, tiene que clasificarse en el grupo de "desconocida" y la variable es aleatoria. Si prefiere decirse que se sabe tan poco sobre una variable que no se le puede asignar o definir un rango - y se elige, en lugar de eso, un solo valor "promedio" entonces no se necesita simulación.

III.2.4 Definiendo Distribuciones para las Incógnitas, Variables Aleatorias (Paso 4)

Este es el paso en donde el profesional con experiencia y juicio entra. Una cosa es decir reservas recuperables como una variable aleatoria (Paso 3), pero otra cosa es decir "OK, pero cuál es el rango y distribución de reservas que se puede esperar?" Todos convienen inequívocamente en que no se sabe, al tiempo de hacer la decisión, la cantidad exacta de reservas (algunas) que el prospecto producirá. Pero rápidamente cuando se trata de especificar, cuál es la distribución de las reservas esto es una empresa difícil de resolver.

La variabilidad resulta de no conocer las reservas al hacer la simulación. Así este paso es la clave, el punto crítico del análisis completo.

Para definir o estimar parámetros para cada distribución hay que tener las siguientes normas generales:

- Las distribuciones pueden ser de cualquier forma, perfil o rango. No se tienen que usar distribuciones de uso "normal" de estadística tal como la normal, lognormal, etc. Las distribuciones pueden ser discretas o continuas. Algunas de las variables aleatorias pueden relacionarse una con otra y en tal caso, la relación de dependencia tiene que definirse.
- Los juicios acerca de la distribución para cada variable no tienen que hacerse por una sola persona. La experiencia del personal y los más familiarizados para cada parámetro pueden establecer el juicio.
- Las distribuciones pueden estar basadas en histogramas o en datos de frecuencia de distribuciones de campos cercanos; las distribuciones pueden basarse en el conocimiento de ciertas variables características que siguen una distribución común; o pueden basarse solamente en un juicio subjetivo.
- No se tiene que ser "cerrado" a una distribución específica. Si las opiniones varían sobre el rango y distribución de una variable se ensaya varios juicios de distribuciones posibles. Como resultado de tal análisis de sensibilidad se puede encontrar, en primer lugar, que la variable ni siquiera sea crítica.

Es, por supuesto, útil preparar una lista de las variables aleatorias usuales en el análisis de un prospecto y definir un rango específico de distribución que tenga que usarse para cada variable.

Un parámetro que puede ser imparcialmente exacto para describir una distribución normal es la porosidad de un núcleo. La distribución lognormal es frecuentemente una representación imparcialmente buena para distribuciones de permeabilidad de núcleo, espesores de estrato

sedimentario en un yacimiento, recuperación de aceite (barriles por pie-acre) en una formación dada, produciendo por un mecanismo común, y en algunas ocasiones, reservas por campo. Cuando se analiza un prospecto se puede elegir usar estas formas funcionales de distribución si no se tiene otros datos o bases en que definir otro tipo de distribución.

Conociendo la forma de la distribución (existente como normal o lognormal para las correspondientes variables) solamente se tendría que determinar la posición de la distribución en el eje horizontal - eso es, el rango. Si se tienen datos estadísticos disponibles de un área cercana se correlaciona con cuidado la misma extensión. O, se puede especificar una (por analogía) desviación media y normal, en la cuál se puede definir la distribución entera.

Una tercera alternativa que es a veces útil, es calcular o estimar un valor bajo y uno alto de la variable aleatoria que se piensa estará cerca de los puntos limitados por un rango de valores posibles de la variable aleatoria. Conocer dos valores discretos y estimar su correspondiente porcentaje acumulativo puede "forzar" a los puntos a una distribución lognormal o normal.

Por ejemplo, suponiendo que se desea describir la recuperación de aceite en barriles por acre-pie (BAF), como una variable desconocida aleatoria que tiene la forma de una distribución lognormal. Pero el problema es que no se tienen datos de estadística para establecer su rango.

Se puede posiblemente resolver este problema calculando un valor bajo cerca del límite y un valor alto cerca del límite superior del rango, usando la ecuación 7 para recuperación de aceite:

$$\text{Rec. de Aceite} = \frac{(7758) (\phi) (1 - S_w) (R.F.)}{B_{0i} \text{ barriles por acre-pie}} \quad (7)$$

ϕ será la porosidad expresada como una fracción decimal; S_w será la saturación de agua congénita como una fracción decimal; R.F. es la fracción (decimal) del aceite que es recuperado y B_{0i} es el factor de volumen de aceite inicial de formación (adimensional).

Se estima el límite inferior de porosidad probable y R.F. para el prospecto y los probables límites superiores de S_w y B_{0i} . Substituyendo estos límites numéricos en la ecuación se obtendrá un valor bajo de BAF cerca del final de la recuperación de aceite. Se puede calcular similarmente un valor alto de ϕ y R.F. y valores bajos de S_w y B_{0i} resultan en un valor alto de BAF. Para ser más específico, se suponen los valores de recuperación de aceite calculados en esta manera 50 BAF y 450 BAF.

Enseguida, se tiene que calcular la probabilidad de que la recuperación de aceite sea menor de 50 BAF o más grande que 450 BAF. Se supone que se estiman que éstas sean 2% y 3% respectivamente. Se pueden ahora combinar todos estos segmentos de información y se habla de una distribución única específica, para recuperación de aceite: una distribución lognormal en la que hay un 2% de probabilidad de que sea menor o igual a 50 BAF y 97% de probabilidad de que sea menor que 450 BAF. La distribución puede ser obtenida por gráfica, como en la figura 9, graficando opuestamente el 50 BAF y 450 BAF en papel logarítmico el 2% y 97% porcentajes respectivamente, y relacionar los puntos con una línea recta.

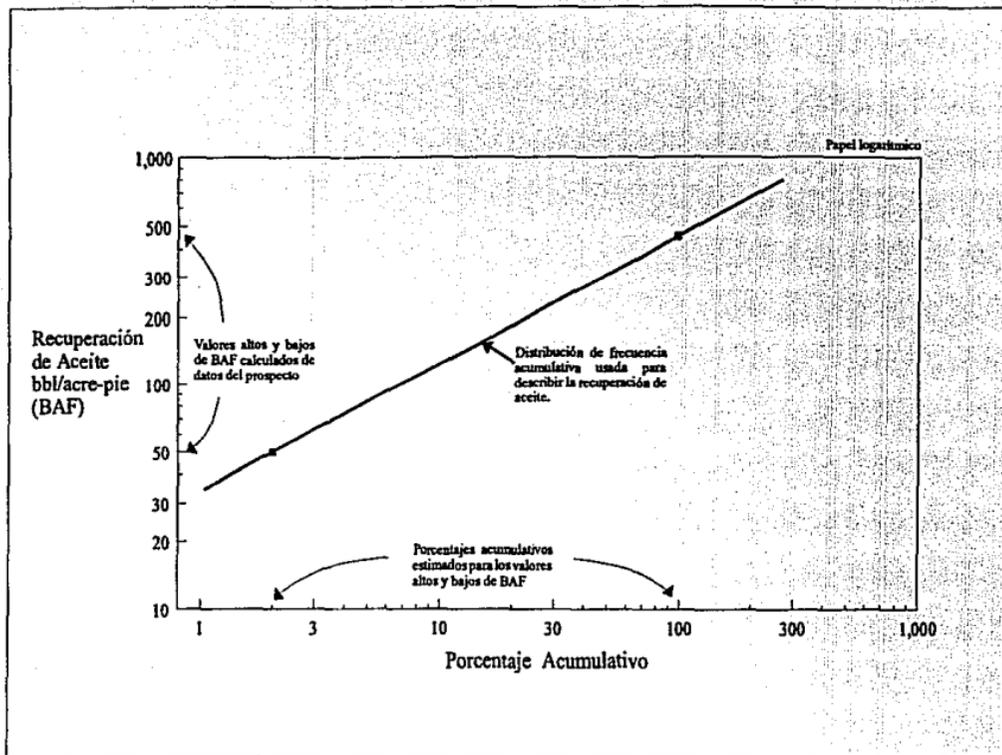


Figura 9. Método gráfico para "forzar" una distribución lognormal a través de los valores límites de 50 BAF y 450 BAF. Estos valores son calculados usando datos del prospecto. Estos son graficados contra los porcentajes estimados menores o iguales a la recuperación BAF. Finalmente, una línea recta es dibujada a través de los puntos de la frecuencia acumulativa en la distribución lognormal. Esta gráfica de frecuencia acumulativa se usa en análisis de simulación para representar la posible distribución de los valores de recuperación de aceite del prospecto.

Así, en esencia, se tiene una distribución lognormal "forzada" por dos valores cerca de la variable aleatoria eso fue calculado usando valores mínimos y máximos de los factores que influyen en la recuperación. En resumen, hay algunas variables aleatorias en análisis de prospecto que frecuentemente pueden ser representadas por una distribución normal o lognormal. Si no se tienen evidencias para sugerir que tipo de distribución diferente puede aplicarse por lo menos se tiene una base sobre que estimar la forma de la distribución.

El rango, o posición en la escala horizontal de la variable aleatoria pueden ser establecidos por datos de estadística, si están disponibles; por conocimiento de la distribución media y desviación estándar, o por "ajuste" de la distribución lognormal o normal usando dos valores discretos cerca de su correspondiente porcentaje acumulativo.

Algunas variables pueden definirse como distribuciones discretas. Generalmente, éstas consistirán de variables aleatorias que pueden tener solamente ciertos valores discretos o contables. Cuando se define distribuciones discretas se tiene que recordar que la suma de las probabilidades de ocurrencia, de todos los posibles valores discretos que pueden ocurrir tiene que ser 1.0. Con simulación se puede usar siempre una distribución discreta en la que algunos acontecimientos ya sea que ocurra o no.

Por ejemplo, se supone que se analiza a lo largo de una exploración una aventura en una área de tempestades severas. Una de las incertidumbres que puede afectar ésta última probabilidad es que una tempestad destruya nuestra plataforma de producción en cualquier año dado. Si la probabilidad de que esto ocurra es 3%, es decir, se puede describir una distribución discreta de probabilidad al tener dos valores: "Sí, había un tempestad," y "No, no había una tempestad durante el año". Las alturas de las líneas de probabilidad arriba de estos dos puntos entonces serían 3% y 97% respectivamente.

En el análisis de simulación la distribución sería probada en cada paso a determinar si una tempestad ha ocurrido. Y el 3% del tiempo el "valor" de la variable sería sí, en cual caso el programa sería directo a la serie de cálculos que consideran las pérdidas de una plataforma. El otro 97% del tiempo el "valor" sería el daño no ocurrido por la tempestad, y estaría calculado por consiguiente. Esta ilustración tiene que sugerir que se tiene flexibilidad considerable en describir distribuciones discretas en un análisis de simulación.

Para algunas variables se pueden tener algunos datos de estadística de áreas cercanas correlacionadas, en que basar las distribuciones para la existencia del prospecto considerado.

La manera usual para representar datos de estadística es mediante un histograma o distribución relativa de frecuencia. Cualquiera de las dos formas es aceptable. Si hay una cantidad grande de datos y las barras del histograma son estrechas se puede suavizar una curva de distribución por la parte de arriba de las barras. La decisión de si la distribución se representa como un histograma o una curva suave es relativamente poco importante si hay un número grande de barras del histograma.

Es fundamental la pregunta de si representar la distribución como una curva suave de la forma que aproximadamente describe el que observó los datos. Desde que el trabajo es más fácil con un histograma, se prefiere usar mucho más que una curva suave que entre por la parte de arriba de las barras.

Cuando se empleen datos de estadística se tiene que recordar que al describir una

distribución no se tiene que discernir en su forma. El histograma puede ser de cualquier forma y rango imaginable. Este tiene que ser evidente, pero sin embargo frecuentemente se tiene la siguiente pregunta:

"Se tienen algunos datos de estadística de un área cercana de espesor productivo. Se arreglan los datos en término de frecuencias acumulativas y se graficaron en papel probabilístico normal. Pero no resultaban una línea-recta. Los datos se graficaron en papel logprobabilístico. Pero tampoco eran una línea-recta. Así que no se distribuyen normalmente o lognormalmente los datos distribuidos, no cabe ninguna opción, que se puede hacer?"

La respuesta - no se preocupe!. Si tiene datos de estadística que se piense son representativos de lo que se espera del prospecto considerado, se usan estos como exactos. Las distribuciones que se usan en simulación no tienen que ser forzosamente normales o lognormales - pueden tener cualquier forma si se sabe que es representativa. Parece ser que tal punto evidente no requiere mucha explicación. Pero es una flaqueza humana adivinar y tratar de obtener de cualquier manera una línea recta, y algunos exploradores parecen tener problemas si los datos son exactos y no, funcionan bien como una línea recta.

Una situación un poco más complicada es cuando se desea definir una distribución para una variable aleatoria pero no se tienen datos disponibles y no se tiene idea de que forma o como tiene que ser la distribución. ¿Qué pasa entonces?. En estos casos se requiere como primer intento por lo menos establecer el rango de valores - un valor mínimo y un valor máximo. Lo siguiente es determinar un valor, o rango de valores dentro de los límites que pueden ocurrir más probablemente que otros valores. ¿Esto es, que la distribución tenga un modo, o un valor más probable?

Si la respuesta es sí, se puede entonces representar la variable como una distribución triangular. Si no hay razón para suponer que un valor sea algo más probable que otro valor, quizás lo que mejor se puede hacer es describir la variable como una distribución uniforme, o rectangular.

Y en la realidad se usan frecuentemente éstas dos distribuciones para representar la variabilidad (incertidumbre) de una variable aleatoria cuando solamente se puede decir acerca del parámetro que sea un valor mínimo, un máximo, y sí o no la distribución tienen un modo. En esencia, se usan las distribuciones como expresiones de ignorancia sobre la variable aleatoria. Pero si se desea explicar la variabilidad puede no tenerse otra elección que usar las distribuciones uniformes o triangulares.

Algunos analistas de simulación han tomado la posición de que raramente se sabe mucho más que un valor mínimo, más probable, y máximo. Desde su punto de vista se debe considerar usar algunas distribuciones más explícitas o detalladas que una distribución triangular. Sobre esas distribuciones triangulares se tienen que usar si no se tiene nada más a seguir, pero si se tienen datos o información se sugiere un tipo diferente de distribución que sin duda se puede usar.

Más específicamente, si solo se puede calcular un valor mínimo y un valor máximo se tiene probablemente que usar la distribución uniforme o rectangular. Estos dos valores límites definen completamente la distribución. Se hace entonces el análisis con la premisa de que algún valor de la variable aleatoria entre los dos límites, probablemente puede ocurrir como cualquier otro valor. Como nota hay algunas variables que pueden actualmente ser distribuidas

de esta manera; en ese caso la distribución uniforme es un poco más precisa que una aproximación.

La distribución triangular requiere que se especifique un valor mínimo, un valor más probable, y un valor máximo. El valor del modo puede ser hallado en algún punto entre los límites, o puede ser coincidente con el valor mínimo o máximo. Estos tres parámetros definen completamente la distribución triangular. ¿Bastante sencillo, correcto?.

Pero se tiene que advertir de algunos peligros de usar la distribución triangular de una manera "indiscriminada". Primero de todo, frecuentemente los exploradores fallan en su definición de los valores mínimos y máximos. Para ilustrar el problema, se supone que se tiene la siguiente colección de datos de una variable aleatoria x :

10,11,12,12,12,12,16,17,19, y 24

¿Si tiene que definir la colección de datos como una distribución triangular qué valores se tienen que usar para los tres parámetros?. Todos conciden en que 10 es el mínimo, 12 es el más probable, y 24 es el máximo?. El resultado de la distribución triangular sería como el que se muestra en la figura 10.

Pero ahora suponga, que los datos disponibles de la variable aleatoria x son:

10,10,10,11,11,12,12,12,12,14,17,18,20,23,24

¿Como se tendría una distribución triangular a través de estos datos?

Muchos observarían a 10 como el valor mínimo, 24 el valor máximo, y $x=12$ el valor más frecuente. Así se termina describiendo otra distribución triangular teniendo exactamente los mismos parámetros (valores mínimos, más probables, y máximos) como en la figura 10 que representa la primera colección de datos. ¿Pero, son los mismos?

Tal vez la respuesta tenga que ser no. La razón es que ese mínimo valor numérico para la segunda colección de datos será, realmente 10, pero ahí parece ser una probabilidad significativa que el valor 10 ocurra. Ocurre justamente tres veces en las dieciséis observaciones. Pero si se asigna a 10 como el valor mínimo de una distribución triangular, la probabilidad de tener un valor de X dentro del rango de $(10 + \alpha)$ se minimiza a cero conforme se acerca a cero. Así quizás una representación más significativa de la segunda colección de datos numéricos sería una distribución tal como la figura 11.

El punto a recordar será que cuando se hable de valores mínimos y máximos de una distribución triangular, significa que tienden a cero para valores la probabilidad de ocurrencia, al considerar rangos más cercanos a los límites. Esto es diferente que hablar acerca de un valor numérico mínimo o máximo de la variable aleatoria cuya probabilidad de ocurrencia puede no ser cero. Esta es una equivocación muy común, y se tiene que estar alerta del sentido de los valores mínimos y máximos de una distribución triangular.

Otro punto que recordar cuando se describen distribuciones de una variable, es como definir el valor mínimo, más probable, y máximo. Se puede suponer indudablemente que se especifica el valor más probable en el punto medio del rango. La lógica para esto, según algunos

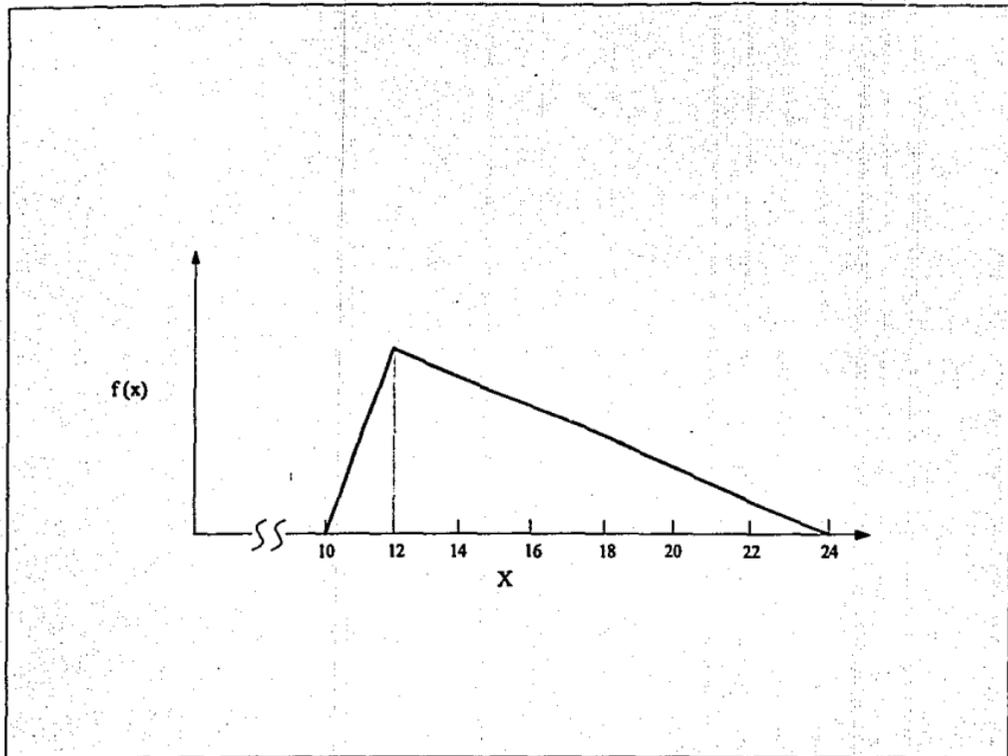


Figura 10. Gráfica de distribución triangular, Caso I.

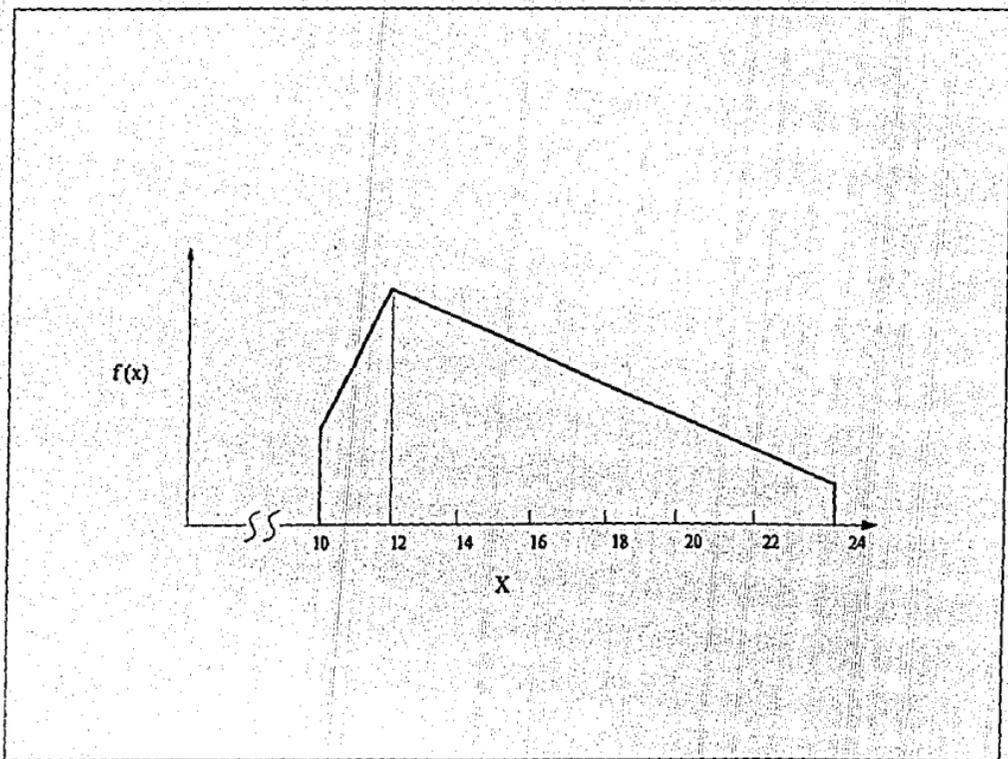


Figura 11. Gráfica de distribución triangular, Caso II.

psicólogos, es que se interpreta como más probable al "mejor cálculo" y se esfuerzan para hacer un pronóstico imparcial.

Si se elige el mejor cálculo en el rango medio se razona como buena la probabilidad de que la variable sea más grande que su mejor cálculo, como menor su estimación. La idea es si siempre se selecciona el valor más probable a medio rango menospreciando las oportunidades de una estimación consistente alta o baja. El punto que se tiene que recordar aquí, es decir el valor más probable, más que un mejor cálculo, y reconocer que una distribución triangular no tiene que ser simétrica. El valor más probable puede ser cualquiera de los dos puntos finales del rango o algún punto entre ellos.

Una última nota que recordar sobre usar distribuciones triangulares es que son, en general, representaciones muy pobres de una desviación alta. Por ejemplo si se distribuye una variable aleatoria como en la parte (a) de la figura 12, los valores mínimos, más probables, y máximos de la distribución son y_1 , y_2 y y_7 respectivamente.

Si se elige representar esta distribución como una distribución triangular, los parámetros serían, y_1 , y_2 y y_7 , como en la figura 12(b). Pero esta es una aproximación pobre porque da mucho más (probabilidad) peso a valores más altos que la distribución original.

Si se representan datos de desviación alta de (a) como una aproximación, sería mucho más aconsejable usar uno de los cuatro lados del polígono mostrado en la parte (c). Así se advierte mucha cautela en usar distribuciones triangulares para representar una variable aleatoria teniendo valores de desviación altos.

Un último comentario acerca de definir distribuciones de la variable aleatoria desconocida, es estimar el número de dependencias. En la simple inversión hipotética de la sección anterior se asume que las variables aleatorias X, Y y Z eran independientes más o menos una de otra. Esto significa que en los reiterativos cálculos de los pasos de simulación los valores usados para X no tenían que afectar el valor de Y o Z.

En un paso en particular el valor de X puede estar cerca del fin alto de la distribución X y el valor de Y puede ser bajo. Y cualquier valor X y Y, no tiene que afectar al valor numérico de Z seleccionado para el paso. Pero en realidad, si ahí existe una relación, o dependencia entre variables X y Y tal que ¿si una tiene un valor alto, la otra también tenderá a un valor alto?

Cuando se clasifican las variables en el paso 3 el grupo (desconocido) de variables aleatorias en el análisis del prospecto muy bien puede tener algunas variables que son dependientes una de otra. Por ejemplo, el espesor neto productivo es una incógnita, y el pozo (o campo) produce a un ritmo. Ambos parámetros serían necesarios en la lista de variables para describir distribuciones.

Pero espesor y buena productividad se relacionan en virtud de la ecuación de Darcy (velocidad de flujo es directamente proporcional al espesor del depósito). Así, esto significaría de algún modo que cuando se describen distribuciones para estos dos parámetros se tendrá que explicar la dependencia. Otros ejemplos de posibles dependencias: saturación de agua congénita (S_w) generalmente aumenta según disminuye la porosidad (ϕ), espesor productivo se relaciona con el campo, la capacidad física productiva con el número de pozos en el campo, y la capacidad de almacenaje y programas de llegada de buque cisterna, se relacionan con los costos

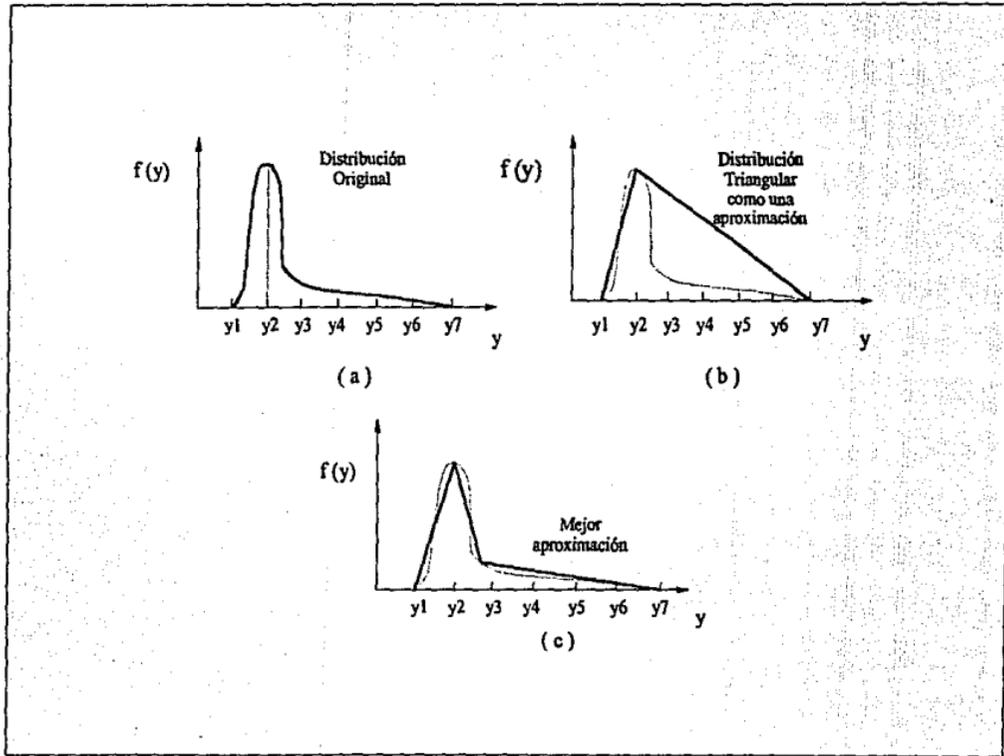


Figura 12. En la ilustración se muestra que una distribución triangular es una aproximación pobre para una variable aleatoria grande. Una mejor aproximación es un polígono de cuatro lados como se muestra en (c).

de tubería y con el campo productivo, etc.

Todo esto sugiere que uno de los intereses que se tendrá que tomar en cuenta cuando se definen distribuciones, es la posible relación de dependencias entre algunas de las variables aleatorias. Este es un número extremadamente importante, y los resultados obtenidos de un análisis de simulación pueden ser muy erróneos si el análisis no hace honor a la relación de dependencia.

Se requiere primero poder reconocer si existe una dependencia entre varias variables y tener algunos caminos para explicar las dependencias en los cálculos de simulación. Es fácil de hacer, pero antes de obtener los detalles se requiere hablar de algunos otros aspectos de simulación. Así se volverá al importante número de dependencia en una parte adelante de la discusión de los detalles mecánicos de como se ejecutan los cálculos reales de simulación.

III.2.5 Ejecutar los Pasos de la Simulación (Paso 5)

Antes de este punto del análisis de simulación, se tiene definido el valor del parámetro de interés y las variables que lo afectan (Paso 1), definida la relación entre todas las variables juntas (Paso 2), se han clasificado las variables como conocidas o variables para las que no se saben valores exactos (Paso 3), y se han especificado distribuciones para todas las variables aleatorias desconocidas. Hasta este punto, se ha definido el modelo o sistema de interés junto con juicios de variabilidad (incertidumbre).

Se completa el trabajo del explorador por el momento y lo que sigue mecánicamente es resolver el problema. Es mecánico en el sentido de que los cálculos y procedimientos de este paso siguen reglas exactas fijas que requieren no iterar más por parte del analista. Estos cálculos mecánicos y procedimientos pueden realizarse por un técnico, pero es mucho tiempo y es más eficiente completar este paso del análisis con la ayuda de la computadora. Por lo tanto los comentarios en este paso serán en el sentido general de qué la computadora para generar los datos necesita definir la distribución final de utilidad.

Como comentario, un análisis de simulación consiste de una serie reiterativa de cálculos de valor, o utilidad. Cada valor de utilidad que es calculado representa una combinación de valores de cada una de las variables aleatorias que afectan la utilidad. Aunque hay, en muchos casos, un número infinitamente grande de combinaciones posibles la variable de la distribución dependiente será bien definida por un número reiterativo de cálculos en un rango de 100-150. Para cada uno de estos reiterativos pasos se obtiene un valor, para cada variable aleatoria para probar su distribución especificada originalmente.

Probar este proceso es la parte clave del análisis, y tiene que hacerse de tal forma que la serie de valores probados sea distribuida en exactamente la misma forma que su distribución original. Eso es, si se especifica la recuperación de aceite será una distribución lognormal pasando por los puntos de 50 BAF y 450 BAF con los porcentajes 2% y 97% respectivamente, se requiere tener en los pasos reiterativos de simulación, la distribución en exactamente la misma manera.

Un esquema probado que lleva a cabo este objetivo es usar un número aleatorio como punto de entrada en la escala acumulativa de porcentaje (o probabilidad acumulativa) de una gráfica acumulativa de frecuencia de la distribución de la variable aleatoria. El número aleatorio es numéricamente igual al porcentaje equivalente y el valor correspondiente de la variable

aleatoria se lee de la gráfica.

Es el valor de la variable aleatoria para un paso particular. En el paso siguiente se obtiene otro número aleatorio que es numéricamente igual a su porcentaje equivalente, y otro valor de la variable aleatoria se lee de la gráfica. Sobre una serie de muestreos los valores probados de la variable aleatoria, se distribuyen en exactamente la misma forma que la distribución especificada originalmente. Este esquema probado, aunque ajusta una de varias maneras posibles, se aleja de la mayoría de los métodos eficaces en una computadora y es usado casi exclusivamente para realizar los pasos de simulación.

Los detalles involucrados en realidad para realizar lo descrito son los siguientes. La primera cosa que se tiene que hacer es convertir todas las distribuciones a sus correspondientes distribuciones acumulativas de frecuencia.

Para el caso de distribuciones continuas de probabilidad las características de y/o los procedimientos requeridos se resumen en la tabla 13, para los tipos principales de distribuciones. Se grafican usualmente las distribuciones acumulativas de frecuencia en papel normal. Sin embargo, se usa a veces papel normal y lognormal de probabilidad cuando se trabaja con distribuciones normal y lognormal esto es para hacer uso de las características de la línea recta de estas distribuciones cuando se grafican en papel probabilístico.

Tabla 13.

Características de la frecuencia acumulativa para diferentes distribuciones de probabilidad.

Tipo de Distribución de Probabilidad	Características de la correspondiente Distribución de Frecuencia Acumulativa, y/o Procedimientos para calcular la Frecuencia Acumulativa.
Normal	Gráficas de Frecuencia Acumulativa como línea recta en papel probabilístico.
Lognormal	Gráficas de Frecuencia Acumulativa como línea recta en papel logprobabilístico (Ejemplo Figura 9).
Uniforme (Rectangular)	Gráficas de Frecuencia Acumulativa como línea recta en papel gráfico coordenado. El porcentaje 0% corresponde al valor mínimo de la variable aleatoria, y el porcentaje 100 % al valor máximo.
Triangular	Convierta la distribución a su equivalente frecuencia acumulativa. Calcule áreas acumulativas para valores de la variable aleatoria a la izquierda como a la derecha del valor más probable.
Histograma, o Distribución relativa de frecuencia	Los puntos graficados para la frecuencia acumulativa son determinados por acumulación de frecuencias relativas (o porcentajes) para los valores mínimo y máximo de la variable aleatoria.
Distribución continua (de cualquier forma)	Convierta la distribución a su equivalente frecuencia acumulativa. Esta técnica puede usarse para cualquier distribución continua, no importa su forma.

Si la distribución de probabilidad es discreta, la correspondiente gráfica acumulativa de frecuencia tiene una forma ligeramente diferente. Para ilustrar, como datos fijos representan una distribución discreta de probabilidad se convierten a su forma acumulativa de frecuencia equivalente suponiendo incluir en el análisis una variable aleatoria, D, que puede tener los valores numéricos discretos dados en la primera columna de la tabla 14. Las probabilidades de ocurrencia de la columna 2 fueron estimados por los exploradores como representativas para el prospecto considerado.

Tabla 14.

Ejemplo de una Variable Aleatoria teniendo valores discretos.

Valores posibles de la Variable Aleatoria D	Probabilidad de Frecuencia Estimada del valor D (Frecuencia Relativa)
D = 1	0.20
D = 2	0.40
D = 3	0.25
D = 4	0.10
D = 5	0.05
	1.00

Si se construyera un cuadro de esta distribución discreta de probabilidad aparecería como en la figura 13. La gráfica acumulativa de frecuencia de esta distribución tendría la forma de la figura 14. Las alturas de los segmentos verticales de la frecuencia acumulativa corresponden a las probabilidades de ocurrencia de cada uno de los valores posibles de la variable aleatoria. Cada segmento vertical sucesivo empieza en la frecuencia acumulativa relativa en la que el anterior segmento vertical termina. El segmento vertical correspondiente al máximo valor posible de la variable aleatoria termina en una frecuencia acumulativa relativa de 1.0. El eje vertical puede, alternativamente, ser expresado como porcentaje acumulativo, multiplicando los valores relativos de frecuencia por 100%.

Hay varias formas para estos tipos de gráficas, que no son gráficas válidas de frecuencia acumulativa, y se necesita evitar la equivocación de usar estos tipos de gráficas en análisis de simulación. Se muestran estas gráficas en la figura 15. La gráfica en la figura 15(a) no es una distribución de frecuencia acumulativa válida pues significa probabilidades negativas. Una distribución acumulativa de frecuencia es una gráfica de área bajo la distribución de probabilidad acumulativa a la izquierda del valor de la variable aleatoria para el rango de valores mínimo al máximo.

Como se mueve del valor mínimo al valor máximo, el área acumulativa continúa crece. Se observa en la figura 15(a) el área acumulativa menor que la variable aleatoria mayor de RV_{MN} a RV_1 (como tiene que ser), pero entonces empieza a disminuir de RV_1 a RV_2 . Esto puede solamente ocurrir si se comienza a sustraer área bajo la curva. Pero esto viola el hecho de que las probabilidades son siempre positivas - no hay probabilidades negativas valuadas. Por lo tanto, una curva acumulativa de distribución de frecuencia tiene siempre que aumentar monotonamente, en función de los rangos de la variable aleatoria de la mínima a la máxima. La figura 15(a) viola este requisito.

La figura 15(b) no es representación válida, de una variable aleatoria discreta a causa de las porciones horizontales de la gráfica de valores acumulativos de frecuencia CF_1 y CF_2 . Los segmentos verticales tienen que tener discontinuidades sobre las porciones de la escala de la variable aleatoria que no puede ocurrir. La técnica usada en simulación se hace inoperable cuando una gráfica acumulativa de frecuencia tiene un segmento horizontal. La razón es que, si el número aleatorio seleccionado corresponde a CF_1 o CF_2 se puede determinar un valor único de la variable aleatoria. Eso es, el número aleatorio correspondiente a CF_1 no se puede

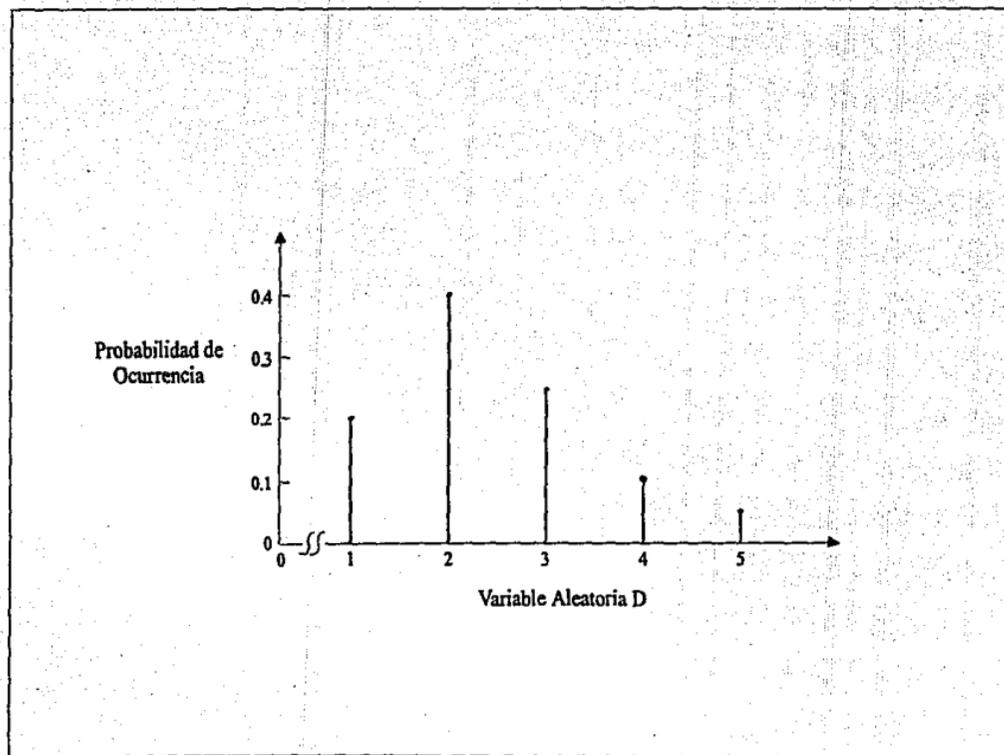


Figura 13. Distribución de la probabilidad discreta de la variable aleatoria D. (Datos de la Tabla 14).

ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA

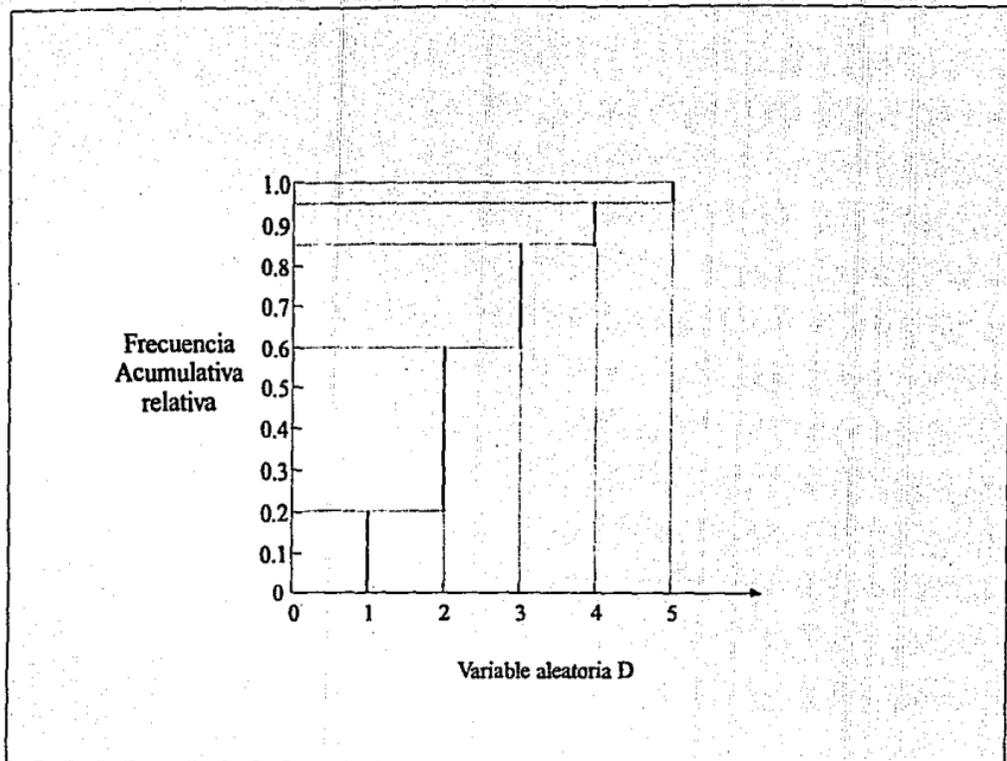
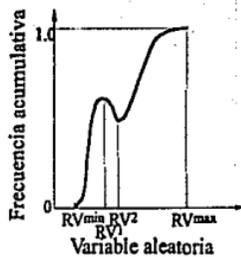
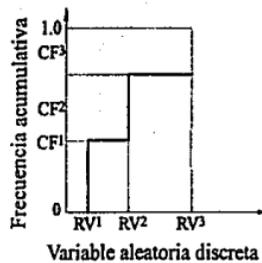


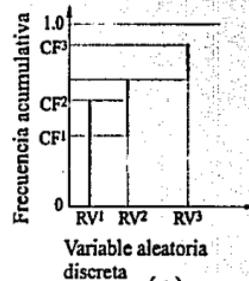
Figura 14. Distribución de la frecuencia acumulativa relativa de la distribución de la probabilidad discreta de la Tabla 14 y figura 13. Las líneas trazadas son unicamente para ayuda en la interpretación de la gráfica.



(a)



(b)



(c)

Figura 15. Las gráficas no son representaciones válidas de distribuciones de frecuencia acumulativa.

determinar, el valor correspondiente a la variable aleatoria es RV_1 , RV_2 o algún valor entre ellos. La regla es, no líneas horizontales en las gráficas acumulativas de frecuencia!

La gráfica en la figura 15(c) no es válida por dos razones. Primero, no cumple que la suma de las frecuencias relativas para los tres valores sea 1.0, al agregar CF_3 se obtiene más que 1.0. La suma de las probabilidades de ocurrencia de todos los posibles valores numéricos de una variable aleatoria discreta tienen que sumar 1.0 por definición. Así cualquiera de las tres probabilidades mostradas en (c) tiene que ser normalizadas para que sumen 1.0, al agregar CF_3 o se tiene que especificar qué ocurrirá la proporción del tiempo correspondiente a $(1.0 - CF_3)$. La segunda razón por la que la gráfica de la figura 15(c) es inválida, es que los segmentos verticales que representan las frecuencias relativas de RV_1 y RV_2 se traslapan en el rango CF_1 a CF_2 . Esto otra vez viola la monotonía de la regla.

La gráfica acumulativa de frecuencia puede ser expresada en términos de fracciones acumulativas (probabilidades) o porcentajes acumulativos. En éste caso, la escala vertical es solamente la fracción acumulativa o probabilidad multiplicada por 100%. Análizando la mayoría de la simulación hecha en computadora operará con fracciones acumulativas, pero es un poco más sencillo explicar simulación usando porcentajes acumulativos. Así en discusiones subsiguientes pueden usarse estas dos escalas igualmente válidas intercambiables. También, es perfectamente satisfactorio graficar las frecuencias acumulativas en un "porcentaje acumulativo" la base más que un "acumulativo por ciento menos que o igual a" la base. Cualquiera de los dos tipos de gráficas serán satisfactorio para los proyectos probados de simulación.

El siguiente detalle requiere concernir con la definición y sentido de los números aleatorios. Se han mostrado números aleatorios en algunas de las descripciones anteriores para probar que la variable aleatoria o distribución sea terminada, sin definir realmente qué son y porque son usadas. El uso de números aleatorios en simulación es estrictamente con el propósito de tener un fácil, imparcial método de probar valores de las distribuciones de las variables aleatorias para cada paso de simulación tal que los valores probados son distribuidos en exactamente la misma forma como la distribución de la variable aleatoria originalmente especificada.

Usando números aleatorios para hacer este proceso de prueba, es solo un camino de varios para probar varias distribuciones. Pero el método más eficaz y usado universalmente en análisis de simulación, así se usa y no molesta explorar algunos de los proyectos probados alternativos.

Los números aleatorios son números positivos, adimensionales. Una serie de números de 2-dígitos (los dígitos 00,01,02,... 98,99) son una serie de números aleatorios si: a) no hay un patrón en el orden en que los números aparecen, y b) cada uno de los 100 números de los dígitos es igualmente probable que ocurra.

En cualquier punto de una serie de números aleatorios, un número es probable que ocurra como cualquier otro número. Se puede hablar de números aleatorios de 2-dígitos, 3-dígitos aleatorios (000 en adelante e incluso 999), o números aleatorios expresados como fracciones decimales sobre el rango 0.0 a 1.0. Será un poco más cómodo hablar en términos de números aleatorios de 2-dígitos en las discusiones. Sin embargo, en las computadoras se expresa normalmente a los números aleatorios como 8 o 16 dígitos o fracciones decimales.

Los números aleatorios pueden ser obtenidos de varias fuentes. Si se tiene interés en obtener números aleatorios de 2-dígitos se puede simplemente tirar un dado que tenga 100 caras, uno de

los 100 números de 2-dígitos aparece en la cara resultante. Reiterativos tiros del dado de 100 caras - genera una serie de números aleatorios de 2-dígitos. Otra fuente de números aleatorios es la serie de dígitos al lado derecho del punto decimal de números irracionales (tal como π , e, raíz de 2, etc).

No obstante otra fuente es el uso de lo que se llama generadores aleatorios de números. Estos son una serie de sencillas, pero muy cuidadosamente diseñadas, ecuaciones que cuando se resolvieron mucho tiempo resultarán una serie de números que son aproximadamente aleatorios o llamados "números pseudo-aleatorios". Realmente ésta es la manera como se obtienen las tablas aleatorias de números, y es la manera como los números aleatorios son generados por una computadora.

El generador algorítmico de números aleatorios es normalmente programado hoy en día en más computadoras científicas, en una biblioteca, como una función de subrutina normal. El programador ni siquiera tiene que estar enterado de como los números son generados, sólo se accesa la subrutina y se obtiene un número aleatorio. La decisión de sí o no una serie de números aleatorios satisfacen varias pruebas que involucran unos conceptos altamente matemáticos y estadísticos no se tratará aquí. Basta decir al respecto que aunque las tablas de números aleatorios y los generadores de números aleatorios producen series de números pseudo-aleatorios, para el propósito se usaran números aleatorios de la edición "pseudo", contra los "verdaderamente" aleatorios no son de interés aquí.

Para probar las distribuciones de probabilidad se puede con seguridad considerar los números aleatorios obtenidos de una tabla o de una computadora (algoritmo) como esencialmente aleatorios.

Ya que se han definido los números aleatorios, la próxima pregunta evidente es ¿como es que ellos prueban la distribución en un análisis de simulación?. La respuesta es que los números aleatorios son iguales numéricamente equivalentes al porcentaje (o si se usan números aleatorios expresados como fracciones decimales) sirven como el punto de entrada en la gráfica acumulativa de frecuencia de la variable aleatoria existente a probar. El valor aleatorio probado de la variable es el valor correspondiente al porcentaje acumulativo, punto de entrada.

Se ilustra esto con un ejemplo. Se supone una variable aleatoria que afecta la probabilidad x , se distribuye en forma de la frecuencia relativa, o histograma de la figura 16. Este histograma puede representar, para el ejemplo, datos de estadística que pueden ser obtenidos al observar x -datos de un área correlacionada cercana. El histograma entonces es convertido a su equivalente frecuencia acumulativa como se muestra en la figura 17. Ahora se supone que se hacen los pasos reales de simulación en los que se necesita probar un valor de x para cada uno de los pasos.

En un paso dado se supone que el número aleatorio obtenido de las tablas es 55. Entonces se entra al eje vertical ordenado al porcentaje de 55 (el porcentaje numéricamente equivalente al número aleatorio), se lee a través a la curva acumulativa de frecuencia, entonces abajo en la abscisa se lee el valor correspondiente de la variable aleatoria x como de 4.1.

Este es el valor de x que entonces es usado en la ecuación para calcular el provecho en el paso de simulación. En el paso próximo se probará otro valor de x para obtener otro número aleatorio, entrar al eje ordenado al porcentaje numéricamente equivalente, y leer el valor correspondiente de x en la abscisa.

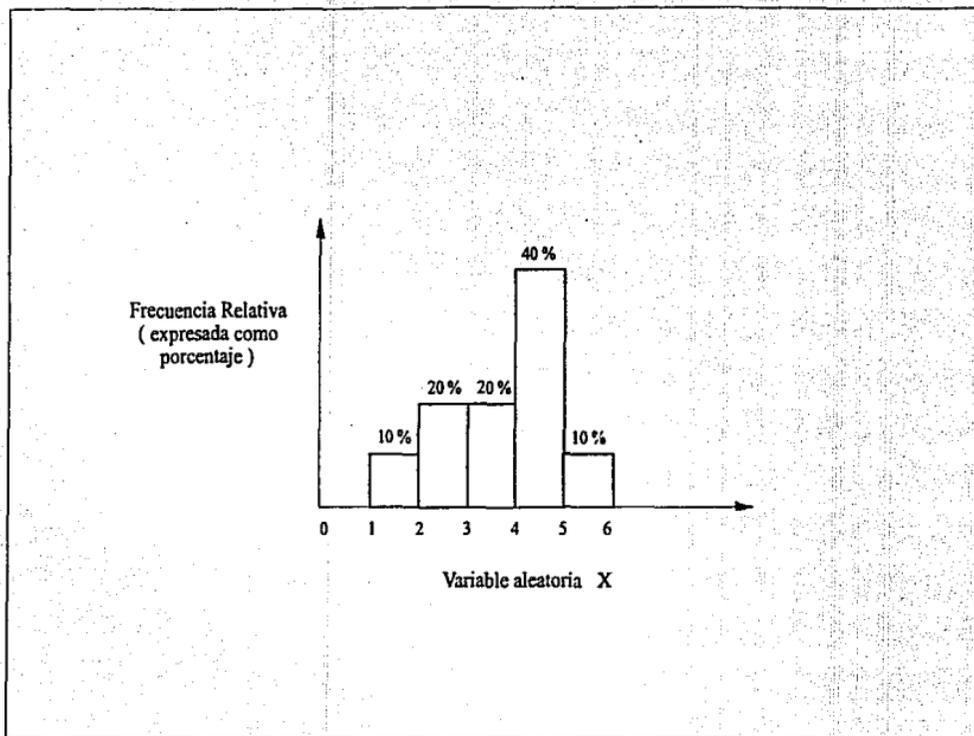


Figura 16. Histograma de la probable distribución de valores los cuales deberá asumir la variable aleatoria X en la perspectiva considerada. La distribución fué definida en el paso 4 del análisis usando datos correlacionados de campos cercanos.

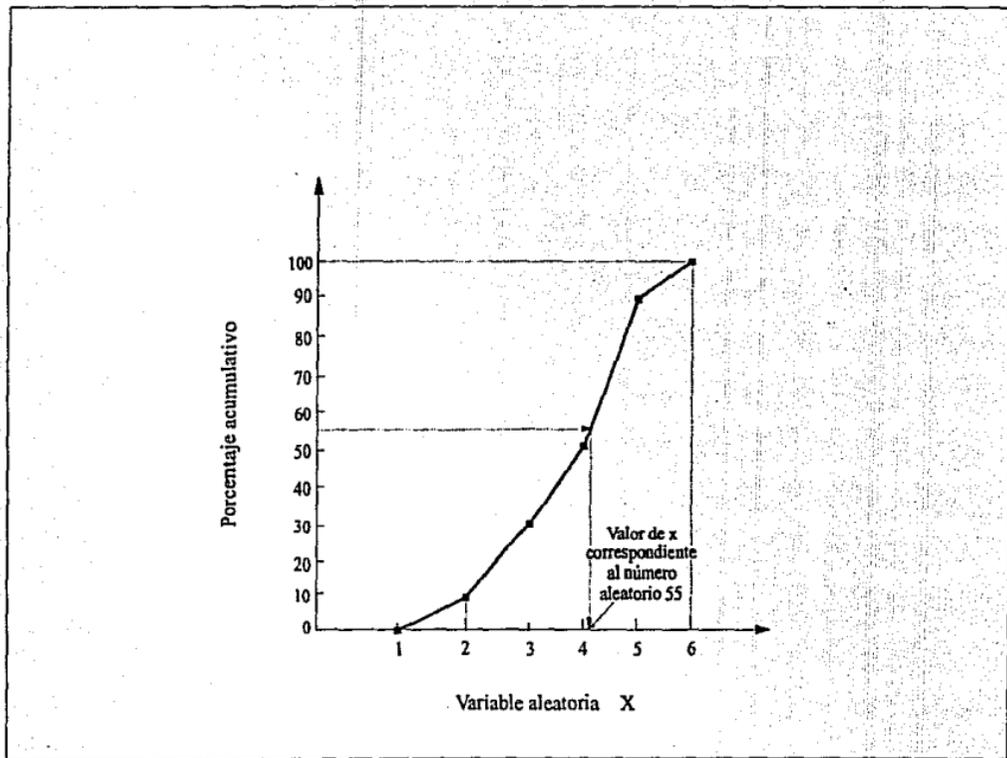


Figura 17. Gráfica de Frecuencia Acumulativa del histograma de la figura 16.

Si el procedimiento para probar valores de x fuera repetido para 200 pasos, es decir, se habría obtenido 200 valores de x . Y si se tiene el tiempo de tabular y graficar los 200 valores probados el histograma de valores probados resultante tendrá la forma exacta como el histograma de la figura 16. El proyecto mecánico que se acaba de describir usando números aleatorios asegura que los valores probados de una distribución de la variable aleatoria sobre una serie de pasos repetitivos será distribuida exactamente como se especifica en la distribución de variables del Paso 4 del análisis.

Este es un punto muy significativo, y tiene que asegurarse completamente de que se entiendan los mecanismos probados de este proyecto. Algunas personas, por ejemplo, obtendrán este punto y entonces se despiden de la simulación porque sienten que cuando se introducen números aleatorios en el análisis repentinamente se reduce el sistema de incertidumbre, a uno casual completamente.

Pero es necesario recordar que los números aleatorios no tienen efecto en el sistema o probabilidad de resultado -pero, no se usan sino para ajustar el paso mecánico de probar valores de cada distribución de variable aleatoria de cada uno de los pasos. Es un método el cual asegura el hecho de que, en realidad, la incertidumbre específica del sistema se ha descrito por la vía de las distribuciones de las variables aleatorias. Se puede determinar el mismo resultado aún sin usar números aleatorios - pero los procedimientos alternativos son mucho más tediosos y no se prestan fácilmente para su solución en la computadora.

Para demostrar quizás de otra manera que el número aleatorio probado del proyecto trabaja, se considera la distribución de la figura 18(a). Pero al observar la distribución se puede probablemente convenir en que la mayor parte de los valores de W se probaría que tienen que estar dentro del rango de W_1 y W_2 . Solamente rara vez se tiene que los valores probados de W son menores que W_1 o más grandes que W_2 . Se da la forma aproximada de la distribución de frecuencia acumulativa en la parte (b) de la figura 18.

Si se entra en la abscisa al rango W_1 y W_2 y se procede hacia arriba a la curva, a través del eje ordenado se observa que se ha circundado casi la escala vertical entera entre las correspondientes frecuencias acumulativas CF_1 y CF_2 . Desde el punto de entrada en la escala vertical es aleatorio, esto significa entonces que la mayor parte del tiempo el punto de entrada estará dentro del rango CF_1 - CF_2 . Y esto vuelve a significar que la mayor parte del tiempo los valores correspondientes de W estarán dentro del rango W_1 - W_2 como originalmente se hizo en la hipótesis. Solamente la oportunidad de probar un valor de W menor que W_1 es cuando el número aleatorio obtenido es numéricamente menor que CF_1 .

Se pueden generalizar estos comentarios, es decir que, aunque el punto de entrada sea aleatorio en el eje ordenado, el valor correspondiente de la variable aleatoria no será aleatorio. Como el declive de los incrementos acumulativos de la curva de frecuencia expone una proporción más grande del eje vertical para un rango dado en el eje horizontal. Como el declive se hace menor hay una proporción más pequeña del eje vertical asociado con una extensión dada en la abscisa.

La pendiente de la curva acumulativa de frecuencia es el ritmo al cual se agrega área acumulativa bajo la curva de distribución de probabilidad, dentro de un rango, desde un valor mínimo a un valor máximo de la variable aleatoria. Como se obtiene cerca del centro de la distribución W de la figura 18(a) se agrega área bajo la curva a un ritmo rápido - por lo tanto la pendiente de la acumulativa es mayor, significando que tiene que ser mayor la oportunidad de

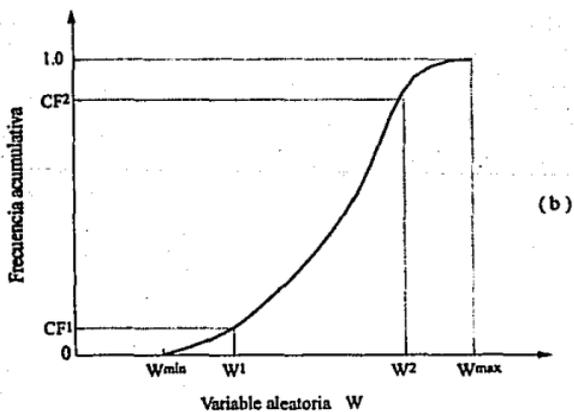
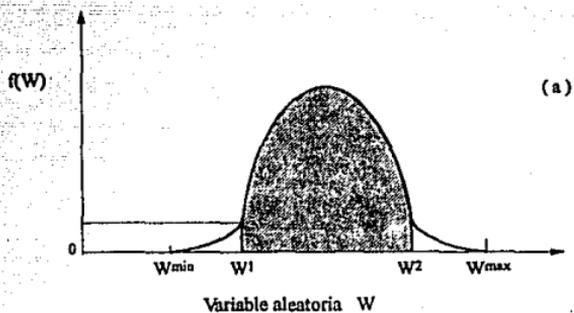


Figura 18. Se muestra una distribución de la variable aleatoria W y su correspondiente gráfica de frecuencia acumulativa.

probar un valor de W en ese rango.

A medida que se aproxime a W_{\max} el ritmo al cual se agrega área acumulativa bajo la distribución disminuye - y la pendiente correspondiente de la frecuencia acumulativa disminuye, exponiendo una distancia menor en la escala vertical para un rango de W dado. Aunque el punto de entrada en el eje ordenado es aleatorio, el control del declive acumulativo de frecuencia relativa prueba la posición del valor de la abscisa. Y este proyecto asegura, por lo tanto, que los valores probados son exactamente de la misma forma que la distribución original.

La única ocasión en la que los valores de la variable aleatoria también serán aleatoriamente distribuidos es cuando la curva acumulativa de frecuencia tiene una pendiente constante del valor mínimo al máximo. Pero esto corresponde a la curva acumulativa de frecuencia uniforme, o distribución aleatoria en que se habría esperado que los valores probados tienen que ser en primer lugar aleatoriamente distribuidos.

La serie de pasos repetidos de simulación se hace probando un valor de cada una de las variables aleatorias, substituyendo estos valores en la ecuación de utilidad y resolviendo para el beneficio. Cada paso requiere que se pruebe cada una de las distribuciones de la variable aleatoria. Como se tiene que usar un número aleatorio diferente en cada una de las series se prueba la distribución. Así, si hay cinco variables aleatorias afectando el provecho se requiere de cinco números aleatorios para cada paso.

Notar que no es correcto usar el mismo número aleatorio para probar todas las distribuciones en un paso. La razón para esto es que usando el mismo número aleatorio deberá automáticamente implicar mezclar valores para todas las variables. Por ejemplo, si el número aleatorio para probar las distribuciones X , Y y Z , fue 95, como en la figura 19, se obtendrían valores de X , Y y Z que fueran cercanos al límite superior, X_{95} , Y_{95} y Z_{95} .

Si el número aleatorio es bajo, los tres valores son bajos. Usar el mismo número aleatorio para probar que una distribución no permite la posible combinación, por decir un valor alto de X , un valor bajo de Y , y un valor de rango-medio de Z . Así la regla es que hay que emplear un número aleatorio y probar separadamente cada distribución!

Después de cada paso se completa el cálculo del valor de la variable dependiente a anotar, o almacenar en la computadora. Cada valor o beneficio calculado de ésta manera representa un estado, o combinación posible de maneras en que las variables aleatorias interactúan. Se continúan los pasos hasta que un número suficiente de valores de provecho están disponibles para poder definir la distribución.

Desafortunadamente, hay reglas no prescritas para decir exactamente cuántos pasos se requieren. La razón está en que el número de pasos requeridos para definir la variable dependiente de la distribución es una función de: a) el número de existencia de distribuciones probadas en cada paso, b) las formas y magnitudes de las distribuciones individuales, c) la relación que reúne todas las variables juntas, etc. Como una regla general, sin embargo, el número de existencia de distribuciones probadas crece con el número de incrementos de pasos requeridos. Sea como sea, el número mínimo de pasos no es menor de cerca de 100, y muchos rangos arriba, tantos como 1000 ó 1500.

En la ausencia de una regla del número mínimo de pasos, es decir, es práctica usual hacer un paso y calcular las frecuencias relativas de varios rangos de beneficio y significa el valor

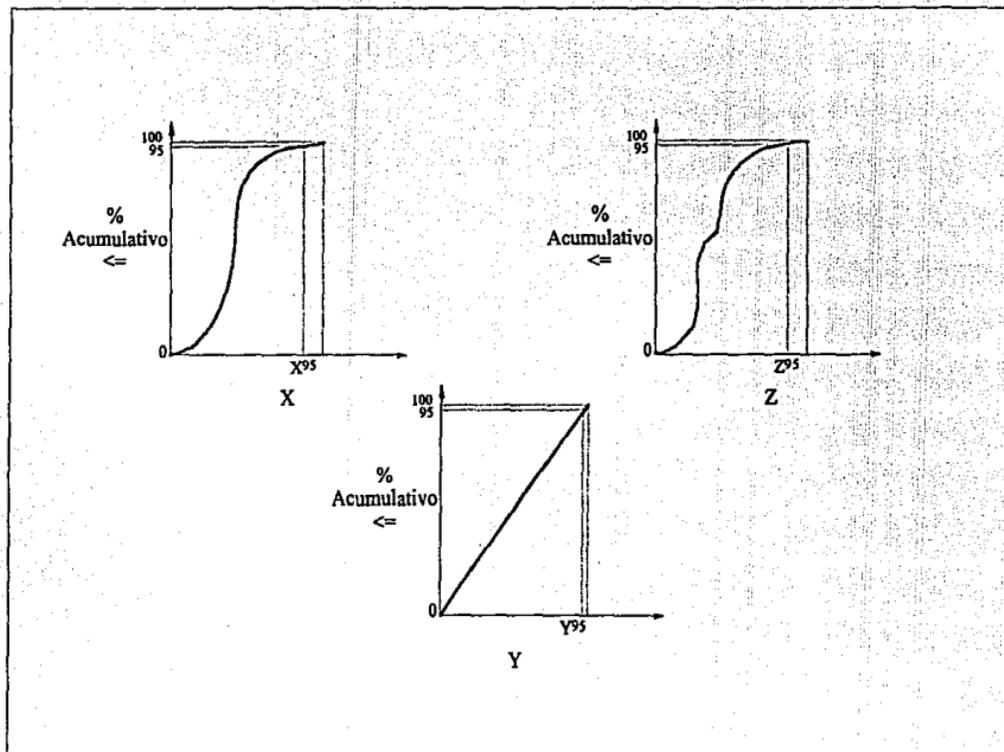


Figura 19. La ilustración muestra que no es correcto el uso del mismo número aleatorio para probar más de una distribución. Cómo el esquema implica que si el número aleatorio es alto (95) todos los valores X, Y y Z son altos. Si el número es bajo los valores probados pueden todos ser bajos. Se usan por separado números aleatorios para cada distribución.

correspondiente. El otro incremento, 100 o se hacen 200 pasos y se comparan las frecuencias relativas y bajo los $(n+100)$ pasos a las frecuencias relativas anteriormente calculadas después de n justos pasos. Si son esencialmente la misma distribución de provecho se han estabilizado (en términos de su forma y extensión) y ya se han hecho bastantes pasos. Si las frecuencias relativas y valores bajos después de n y $(n+100)$ pasos no se han ajustado más de cerca se requieren más pasos.

En una computadora tal "prueba de computadora" es una serie que puede llevar más tiempo en términos de entrada y capacidad, que hacer simplemente un número muy grande de pasos, en el primer paso, es decir 1000 ó 1500. El tiempo de cómputo para hacer los pasos reales es muy pequeño, y puede requerir menos tiempo de computadora en una corrida larga, hacer un exceso de pasos y eliminar además los pasos de entrada y salida de datos.

Generalmente estos trabajos con simulación son una base regular para obtener experiencia en el número de pasos requerido. Típicamente comenzarán altos y gradualmente trabajarán con números descendentes sobre una serie de análisis semejantes de simulación.

Se tiene ahora esencialmente que tapar todos los mecanismos de análisis de simulación excepto para el número importante de dependencia. Se mencionan al final de la discusión del Paso 4 algunas de las variables aleatorias que se pueden afectar, en realidad, se relacionan una con otra. Si éste es el caso se requiere poder probar distribuciones de las variables aleatorias en una manera que duplicará la dependencia.

Así las preguntas anteriores son (a) ¿como se reconoce que puede existir una dependencia entre dos o más variables aleatorias?; y (b) dado que se puede reconocer una relación de dependencia ¿como se modifica para probar que la técnica puede explicar la dependencia en el proyecto probado?.

En este punto se puede relatar la manera de lo que se hizo primero sabiendo de la necesidad de dependencia. Uno de los análisis tempranos se hizo en 1966 se relacionó con juego y exploración representando, entre otras cosas, las distribuciones de espesor neto productivo y el gasto inicial de producción por pozo.

Por ignorancia se puso cada una de estas distribuciones en el análisis separadamente, y en cada paso se probaron las dos distribuciones separadas como si fueran independientes una de la otra. Los resultados del análisis eran exagerados, se tuvo que seguir la pista hacia atrás para ver en donde se había desviado. Solamente llevo un corto tiempo mientras se encontró la equivocación. El programa probó el espesor neto productivo separado de IP (potencial inicial, bls/dla/pozo) y fué a través de un análisis de curva de declinación que se calculó la vida del campo, para cada paso puede determinarse un VPN de renta.

Si el valor del espesor fuera alto (lo que implica grandes reservas) y al mismo tiempo el valor probado de IP fuera bajo, la vida resultante del campo era ...como 243 años! y al otro extremo si los números aleatorios fueron tales que el espesor fué bajo (significando reservas bajas) e IP era alto, el campo fué agotado en como 17 días!. Innecesario es decir, que se aprendió una importante lección en la importancia de tratar dependencias en análisis de simulación de riesgo e incertidumbre.

Hay varias medidas de estadística de correlación (o a falta de esto) puede averiguarse si dos o más variables son dependientes una de otra. Pero una manera más fácil que practican los

exploradores puede ser usar o tener que aprender teoría de estadística, hacer un gráfica de valores de dos variables de interés.

Por ejemplo, fue de interés si se relacionaban las variables A y B una y otra, y tenían algunos datos numéricos de pozos o campos cercanos que se consideran representativos: se puede hacer un gráfica de variables. Una variable B graficada contra los valores numéricos de A y B para cada punto. Y si se hace ésto la gráfica se parece a la gráfica de la figura 20.

¿Qué se concluiría acerca de cualquier dependencia entre las variables?. Bien, se puede probablemente convenir que parecen no estar relacionadas. Se pueden tener valores altos de A y valores bajos de B, valores altos de B y valores bajos de A, y cerca entre todos. Se puede probablemente asegurar y concluir en esta ocasión que no hay relación o dependencia, así se puede mostrar cada distribución separadamente en el análisis. Este es un extremo donde la variable aleatoria es independiente.

Ahora si se hace algo semejante al graficar entre otras dos variables aleatorias, C y D y los resultados son como se muestra en la figura 21. Esto representa el otro extremo de la dependencia completa. Si se conoce C, automáticamente se conoce D (o viceversa). Así en términos de análisis de simulación, solamente se necesita especificar la distribución de una de las variables (cualquiera que sea fácil de definir). Para cada paso se probaría la distribución y se determina su valor numérico.

Entonces se entraría en la gráfica con ese valor, para leer el valor correspondiente de la otra variable a usar en el paso. O, alternamente si ésta se determina en una computadora, se determinaría la función (ecuación) de la curva y se usa la ecuación, para establecer el valor de D dando el de C (o viceversa).

Estos dos maneras representan cualquiera de los dos casos o la dependencia, y son sencillos de manejar. Pero ahora se tratará el caso donde hay algún "medio modo" de dependencia. Otra vez se supone que se tienen datos numéricos de dos variables aleatorias (desconocidas) X y Y de un área cercana, y que al graficar los datos X-Y aparecen como en la figura 22. Los datos X-Y no son claramente independientes (como en la figura 20), y ahí no parece ser explícita la relación de dependencia tal como en la figura 21.

La variable Y tiene un rango considerable de valores, dado que X sugiere la necesidad de probar una distribución de Y para cada paso. Pero también tiene que ser bastante evidente que el rango de variación de Y cambie conforme X se incrementa. Por estos medios la distribución Y necesita probar que tiene de alguna manera de compatibilidad con el valor específico de X que se muestra en cada paso.

La situación mostrada en la figura 22, representa "en gris" el área entre el extremo negro y blanco, casos de completa dependencia o completa independencia. Así la pregunta evidente entonces es, ¿como se modifica esto, para que probando proyectos se muestren ambas distribuciones X y Y en cada paso pero hasta cierto punto cuál explicará la dependencia parcial observada de la figura 22?

La manera de hacer esto es muy sencilla, pero es un poco engañosa e involucra explicaciones. Así, antes de zambullirse o "abrumarse", con esto por lo menos ya se sabe el camino hacia el que se va.

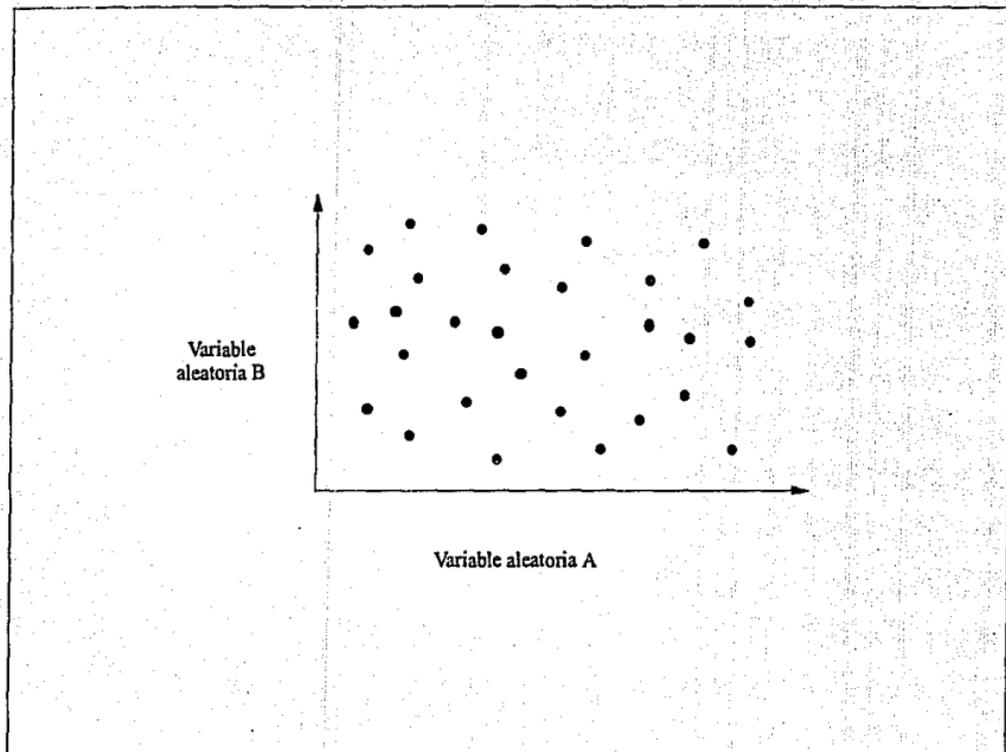


Figura 20. La gráfica muestra que no hay relación entre las variables aleatorias A y B. En este caso se deberán describir las distribuciones para cada variable y probar cada distribución separadamente en los pasos de simulación.

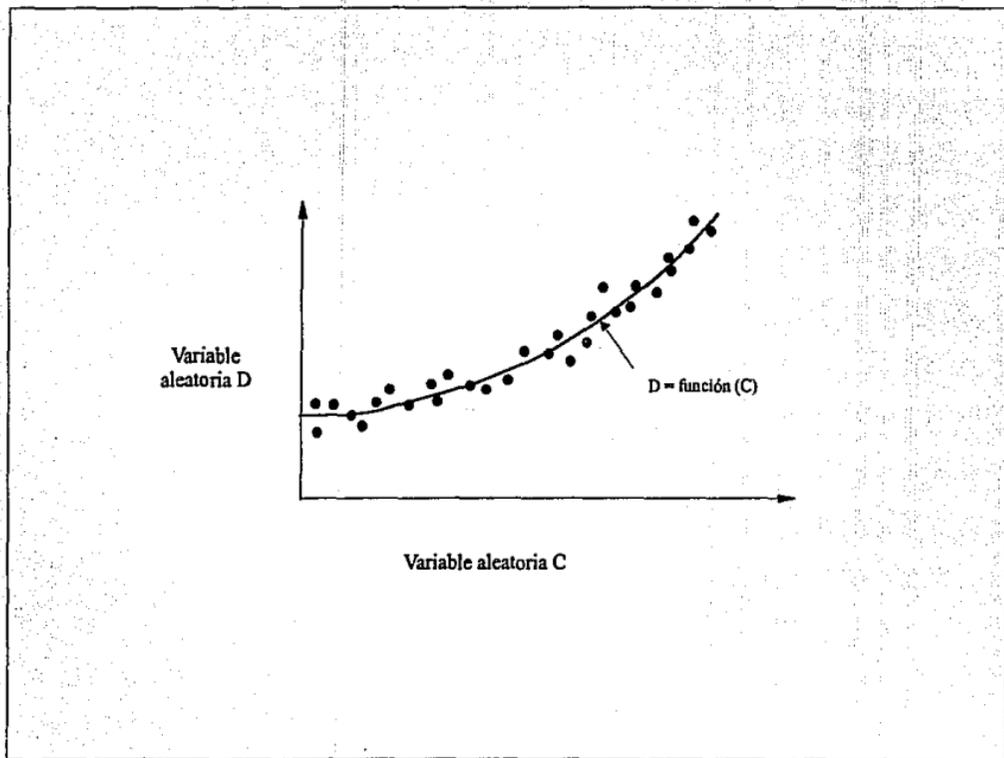


Figura 21. La gráfica muestra las variables aleatorias C y D las cuales son completamente dependientes. En este caso se necesita especificar solamente la distribución de C . Conociendo un valor de C se puede entrar a la gráfica (o resolver la ecuación $D = f(C)$) con el valor de C el correspondiente valor de D .

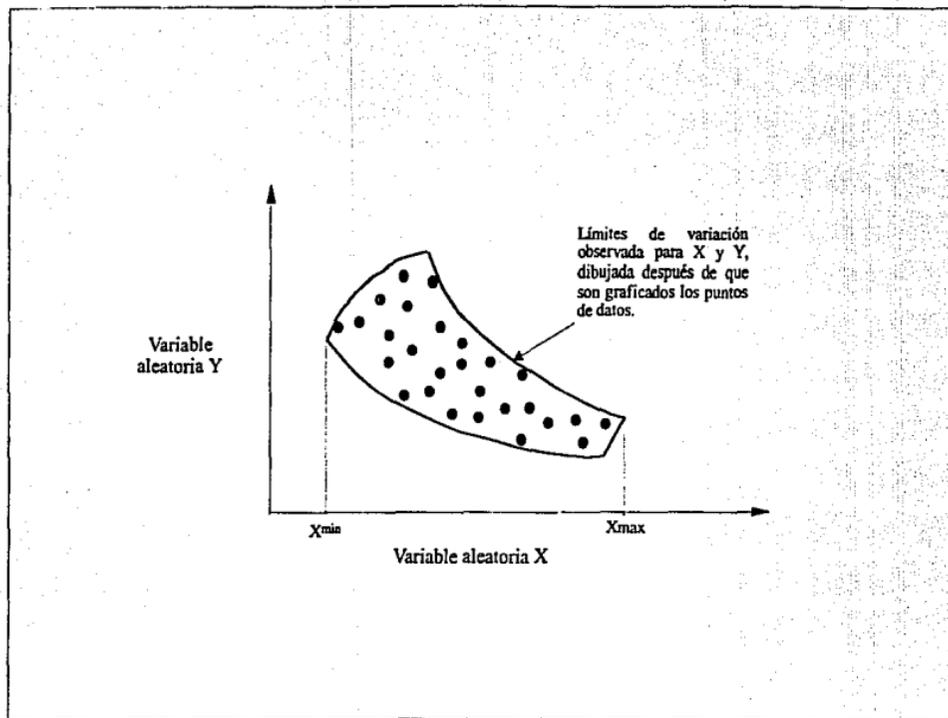


Figura 22. Gráfica de las variables aleatorias X y Y en la cuál hay una dependencia parcial entre X y Y, pero la dependencia no es una función específica como en la figura 21. Después de que los puntos se dibujan, un límite o envolvente se dibuja alrededor de los datos observados. Con base en esta información cualquier combinación de valores X-Y fuera de los límites no deberá incluirse en el análisis de simulación.

Primero, se explicarán los pasos iniciales listados en la tabla 15, la preparación de datos que se tiene que seguir para manejar la situación de dependencia parcial. Entonces se seguirán los pasos conceptuales, probando en realidad para dos variables que exhibirán una dependencia parcial. Esta necesidad está dada en la tabla 16. Lo siguiente a través de los mismos pasos como si se hiciera en una computadora (tabla 17).

Esto entonces es seguido por un ejemplo numérico. Finalmente, se mencionan algunos puntos generales para reconocer y tratar dependencias parciales en simulación. Hasta éste momento cualquiera es un experto en dependencia parcial o está completamente confundido (o ambos casos!). Es de veras muy sencillo manejar dependencias una vez que se obtiene el significado.

Para explicar dependencias parciales en análisis de simulación primero se tienen que seguir algunos pasos como "preparación de datos". Se dan estos pasos en la tabla 15. Los primeros dos pasos se explican por sí mismo, y al terminar estos dos se tendría un límite, como en la figura 22. Tiene que aclararse que el límite alrededor de los puntos graficados X-Y, no tiene que dibujarse como líneas rectas. Se dibuja, sin embargo, para incluir todos los posibles valores X-Y (observados). La razón es que el siguiente proyecto probado excluirá todas las combinaciones X-Y fuera del límite.

Tabla 15.
Preparación de datos a contar para dependencias parciales
entre Variables Aleatorias en Análisis de Simulación.

-
1. Prepare una gráfica de datos X-Y disponibles como en la figura 22.
 2. Dibuje el límite o envoltorio alrededor de la prueba observada de los puntos X-Y. Esto define los límites junto con los cuales X y Y pueden variar. Cualquier combinación de valores de X y Y fuera de los límites puede excluirse y se considera que tiene una probabilidad de ocurrencia de cero.
 3. Determine la variación de Y con el límite como una función de X, y defina una distribución normalizada de Y la cual representa la variación observada. Los límites de ésta distribución normalizada son 0 y 1.0 dónde 0 corresponde al valor mínimo posible de Y y 1.0 corresponde al valor máximo posible de Y para un valor de X dado. La distribución normalizada adimensional se da por el símbolo YNORM en el eje de la variable aleatoria. La distribución puede ser de cualquier forma.
 4. Los datos del modelo de simulación son:
 - a. La distribución de frecuencia acumulativa X.
 - b. La frecuencia acumulativa Y de la distribución normalizada YNORM.
 - c. La siguiente ecuación es usada para calcular el valor de Y para cada paso:

$$Y = YMINx + (YMAXx - YMINx)(YNORM) \quad (8)$$

dónde:

YMINx = Valor mínimo posible de Y, dado un valor probado de X.

YMAXx = Valor máximo posible de Y, dado un valor probado de X.

YNORM = Valor de la distribución normalizada Y adimensional, probada para cada paso.

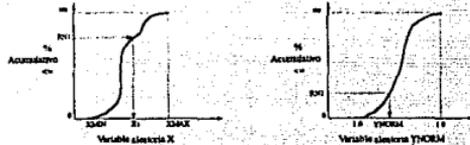
Y = Valor calculado de la variable aleatoria Y para usar en cada paso.

El tercer paso de la tabla 15 tiene que especificar el tipo de variación de Y dentro del límite como una función de X. Conceptualmente esto es equivalente a cortar el papel de la gráfica fuera del límite de la figura 22 y observar dentro del límite de Y solamente como es distribuido, como una función de X. Eso es, dentro del límite ¿los valores de Y dispersos verticalmente, son aleatorios? ¿Están la mayor parte de los valores cerca del límite superior o inferior? ¿Se agrupan los datos al centro entre los límites?.

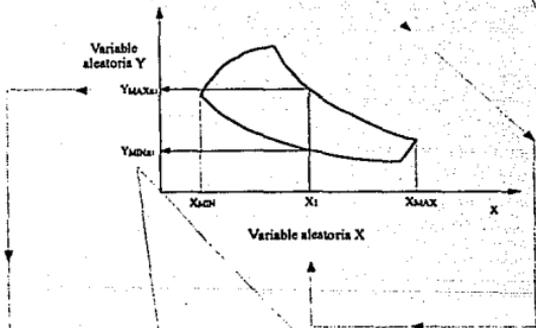
Tabla 16.

PASOS CONCEPTUALES PARA PROBAR VALORES EN CADA PASO PARA DOS VARIABLES X y Y PARCIALMENTE DEPENDIENTES.

1) Probar la frecuencia acumulativa de la distribución de la variable aleatoria X y la variable normalizada YNORM en la manera usual.



2) Usando el valor probado (X¹) localice el mismo valor en el eje de la abscisa de la gráfica X-Y, en forma recta hasta el eje ordenado sobre X¹, y lea los valores numéricos de YMIN y YMAX en la intersecciones del eje ordenado con los límites superior e inferior.



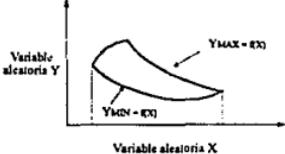
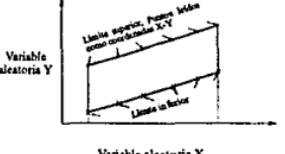
3) Resuelva la ecuación (8) para el valor de Y. Para este paso los valores usados en el cálculo son X¹ probado en el paso 1 y el valor de Y calculado en este paso.

$$Y = YMIN x + (YMAX x - YMIN x) (YNORM) \dots \quad (8)$$

donde $\left\{ \begin{array}{l} YMIN x = \text{Valor mínimo posible de Y, dando X} \\ YMAX x = \text{Valor máximo posible de Y, dando X} \\ YNORM = \text{Valor probado de la distribución normalizada, adimensional.} \end{array} \right.$

Tabla 17.

PASOS PARA PROBAR VALORES DE DOS VARIABLES ALEATORIAS X y Y PARCIALMENTE DEPENDIENTES EN UNA COMPUTADORA.

<p>METODO # 1 (Ecuaciones de los límites en la gráfica X-Y que son conocidos)</p>	<p>METODO # 2 (Ecuaciones de los límites en la gráfica X-Y que no son conocidos)</p>
<p>Paso 1). Introducir a la computadora:</p> <ol style="list-style-type: none"> Distribución de la frecuencia acumulativa de la variable aleatoria X. Distribución de la frecuencia acumulativa de la distribución normalizada Y, YNORM. Ecuación 8. Ecuaciones de los límites superior e inferior como funciones de X.  <p>Paso 2). Probar los valores de X y YNORM en la manera usual.</p> <p>Paso 3). Sustituir el valor de X en las ecuaciones para YMIN y YMAX para calcular YMIN* y YMAX*.</p> <p>Paso 4). Sustituir los valores calculados de YMIN* y YMAX* y el valor probado de YNORM en la ecuación 8, y resolver para Y.</p>	<p>Paso 1). Introducir a la computadora:</p> <ol style="list-style-type: none"> Distribución de la frecuencia acumulativa de la variable aleatoria X. Distribución de la frecuencia acumulativa de la distribución normalizada Y, YNORM. Ecuación 8. De cuatro a ocho puntos de la gráfica X-Y para los límites superior e inferior.  <p>Paso 2). Usando valores X-Y para el límite superior se tiene que calcular una ecuación polinomial a través de los puntos de la forma YMIN=f(X). Repetir para los puntos del límite inferior la ecuación polinomial, YMAX= f(X).</p> <p>Paso 3). Probar los valores de X y YNORM en la manera usual.</p> <p>Paso 4). Sustituir el valor de X en las ecuaciones para YMIN y YMAX para calcular YMIN* y YMAX*.</p> <p>Paso 5). Sustituir los valores calculados de YMIN* y YMAX* y el valor probado de YNORM en la ecuación 8, y resolver para Y.</p>
<p>Nota: El paso 1) se realiza únicamente una vez. Los pasos 2), 3) y 4) se repiten para cada paso.</p>	<p>Nota: Los pasos 1) y 2) se realizan únicamente una vez. Los pasos 3), 4) y 5) se repiten para cada paso.</p>

Para los datos de la figura 22 hay que notar aquí la variación de Y como una función esencialmente aleatoria de X. Estos mecanismos que describen la variabilidad de Y como una función de X son distribuciones uniformes, rectangulares.

Se tiene la elección de definir la variabilidad de una variable como una función con algún tipo de distribución que aproxime los datos. En la figura 23 se dan esbozos de algunas posibilidades de esta consideración. En la parte (a) los datos X-Y graficados sugieren que la variabilidad de Y dentro del límite es uniforme. La distribución particular uniforme Y correspondiente al valor específico X_1 es mostrada en la parte insertada arriba. La intersección del eje ordenado X_1 con el límite superior e inferior determina y define los límites específicos, unifica la única distribución Y, y está dada en la gráfica como YMIN y YMAX.

Para la figura 23(b), la variabilidad de Z como una función de W no es aleatoria. Más, parece ser una concentración de puntos de datos con un grupo de puntos aproximadamente intermedio entre el límite superior e inferior. En éste caso se puede elegir representar la variabilidad de Z dentro de la región como una serie de distribuciones simétricas triangulares. La distribución particular correspondiente a un valor W_1 aparece en la parte insertada arriba de la figura 23(b). Se puede posiblemente tener también representada la variabilidad de Z como una serie de distribuciones normales.

La figura 23(c) muestra la variabilidad de T, como las funciones de S que pueden representarse por distribuciones triangulares o desviación de la línea recta hacia los valores máximos de T y reflejar la concentración, o agrupación de datos en la porción superior del área resaltada. La desviación de la línea recta correspondiente a una distribución triangular de S_1 es mostrada en la inserción. En lugar de la desviación de la línea recta de la distribución triangular se puede quizás tener seleccionada y representar la variabilidad con una distribución lognormal que tiene su valor de manera cercana a los máximos valores de T.

En este punto, presumiblemente se reconoce qué es significativo describir la variación de Y dentro de la región mostrada como una función de X. Es simplemente una manera de tratar de caracterizar la gráfica de puntos datos X-Y por un tipo de distribución uniforme, triangular, etc. Pero tener descrita la forma, o perfil de la distribución de variabilidad de Y solamente es medio problema.

La otra dimensión que se tiene que concernir es el hecho de que allí existe una distribución específica única de Y, (en la figura 23(a) por ejemplo) para cada valor de X!. Y desde que X es dibujada como la variable aleatoria continua, es un número infinito de posibles valores de X entre los valores mínimo y máximo, X_{MIN} y X_{MAX} . Para cada valor de X hay una única distribución uniforme de Y, causada por el hecho de que los límites numéricos de Y, YMIN y YMAX, varían como varía X. Para que semejante cosa ocurra, se piensa en 3 dimensiones, la dimensión 3 de superficie se obtiene cuando se agrega una escala de frecuencia a la gráfica X-Y, figura 22, como se muestra en la figura 24.

La forma en que una figura se desarrolla poder ser como sigue. Se supone que se erige una pared vertical alrededor del área resaltada de datos X-Y en la figura 22. Entonces desde la variabilidad de Y con el área limitada se considera aleatorio poner un "techo" en el límite vertical. El resultado es una superficie 3 - dimensional de la figura 24.

El problema de probar valores X y Y para cada paso, como se relata en la figura 24, una vez que se prueba (en la manera usual) un valor de X, es decir X_i , de la distribución X se ha

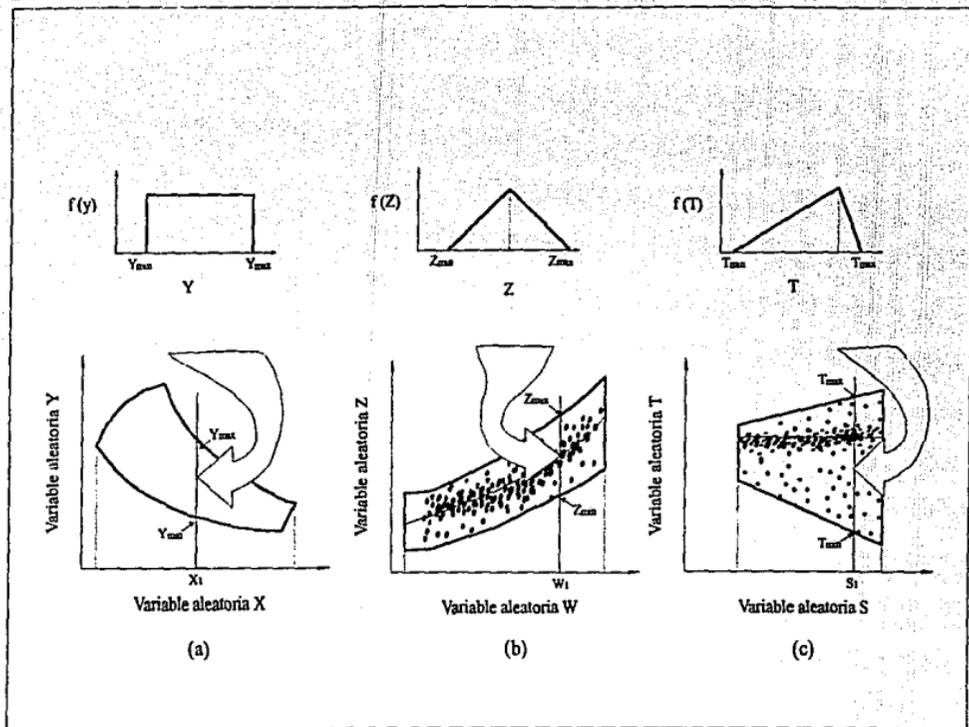


Figura 23. Ejemplos de varios tipos de variabilidad con las regiones limitadas de las variables aleatorias parcialmente dependientes. En la gráfica (a) la variación de Y como una función de X es aleatoria. En la gráfica (b) la variación de Z como una función de W puede aproximarse a una distribución triangular simétrica (o quizás alternativamente, una distribución normal). La gráfica (c) muestra la variación de T como una función de S la cuál puede representarse por una distribución triangular con su valor más probable, el valor máximo refleja la densidad de puntos correspondiente al área de la gráfica. Las distribuciones superiores representan conceptualmente las distribuciones Y , Z y T correspondientes al eje vertical dibujado como X_1 , W_1 y S_1 respectivamente.

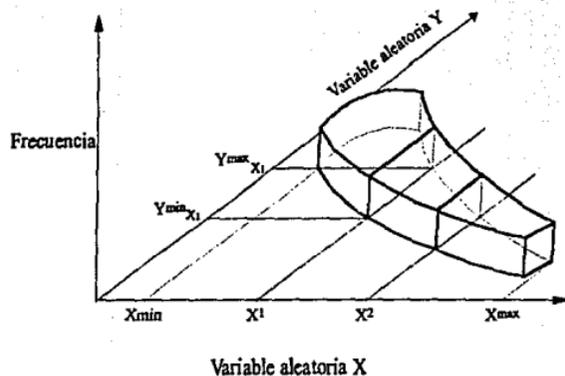


Figura 24. Representación Tridimensional de la dependencia parcial de la figura 22. El tercer eje es la escala de frecuencia. La proyección de la parte de arriba de la figura sobre el plano X-Y es el límite de la figura 22. Esta figura representa el caso donde la variación de Y como una función de X es uniforme, o aleatoria.

arreglado la posición en el eje X de la gráfica 3 - adimensional. Para probar un valor de Y se puede, en esencia, cortar la superficie con un cuchillo para observar el perfil de distribución por la cara del corte. Sería una distribución rectangular que tiene por "paredes" a Y_{MIN} y Y_{MAX} y un "techo" plano.

Ésta sería la distribución exacta que probaría un valor de Y. En el próximo paso el valor de X probado puede ser X_2 . Si se corta la superficie de la figura 24 en este punto (X_2) se ve sin embargo otra distribución uniforme. Su forma es la misma pero los valores numéricos de sus límites son diferentes. Verdaderamente para cada posible "corte" de la superficie se expondría una distribución Y diferente. Las formas serían las mismas (dos "paredes" verticales y un "techo" plano!) pero los valores numéricos de todos los límites son diferentes.

El alcance que se acaba de describir significa que se requiere guardar, o apuntar la distribución específica Y para cada valor posible de X. De esta manera se establece un valor de X de manera que cada paso puede mostrar el valor de Y para dicho paso.

Si hay muchos valores (infinitos?) posibles de X por este medio se tiene que graficar muchas (infinitas?) distribuciones acumulativas de Y. Y eso puede ser un trabajo extremadamente tedioso!). Así qué se requiere de alguna forma para no tener que trabajar el resto de la vida en graficar distribuciones Y.

Afortunadamente, la técnica para hacer eso es relativamente rígida, y este método es realmente el que tiene un procedimiento sencillo para tratar dependencias parciales conjuntas. Involucra normalizar la distribución mencionada en el paso 3 de la tabla 15. Al normalizar una distribución media se convierte la escala de la variable aleatoria a una escala equivalente adimensional en el rango de cero a uno.

La ecuación que lleva a cabo esta normalización, en forma general, es:

$$X' = \frac{X - X_{MIN}}{X_{MAX} - X_{MIN}} \quad (9)$$

dónde X es la variable aleatoria, X_{MIN} es el valor mínimo de la distribución, X_{MAX} es su valor máximo, y X' es la variable aleatoria transformada al normalizar la distribución. Cuando $X = X_{MIN}$ la solución de la ecuación 9 da un valor de X' de cero. Cuando $X = X_{MAX}$ se obtiene de la ecuación un valor de X' de 1.0. Entonces se ha probado mantener ésta discusión en términos de variables generales parcialmente dependientes X y Y, y desde que se tiene interés en normalizar la variabilidad de Y, se tiene que describir la ecuación con los símbolos de la tabla 15 como:

$$Y_{NORM} = \frac{Y - Y_{MIN}}{Y_{MAX} - Y_{MIN}}$$

dónde:

Y_{MIN} = valor mínimo de Y

Y_{MAX} = valor máximo de Y

Y_{NORM} = Variable aleatoria normalizada, adimensional

La razón será extremadamente útil al tratar con una distribución normalizada que es una distribución representativa que describe todas las distribuciones infinitas posibles Y como

funciones de X. Para cada paso se puede mostrar un solo valor de YNORM de la distribución representativa, y éste nos diría la distancia fraccional a lo largo de la línea de YMIN a YMAX para localizar el valor probado deseado.

El paso final solamente sería asociar esta distancia fraccional a los valores numéricos específicos de YMIN y YMAX asociados con el valor de X, probado anteriormente. Este último paso se usa al terminar la ecuación (8) de la tabla 15.

Con referencia en la figura 23, todo número de distribuciones Y infinito puede tener como distribución normalizada equivalente, una distribución uniforme de parámetro YNORM, teniendo como valor mínimo cero y valor máximo 1.0. Para el caso de las distribuciones triangulares simétricas de la figura 23(b) la distribución equivalente normalizada sería triangular, teniendo como parámetros un valor mínimo de cero, un valor más probable de 0.5 (por simetría el modo es la mitad entre el valor mínimo y el máximo), y un valor máximo de 1.0.

Para el caso de la figura 23(c) en la que los valores más probables parecen estar a tres cuartos de la distancia superior, obliga a la distribución triangular equivalente a tener un valor mínimo de cero, un valor más probable de 0.75, y valor máximo de 1.0.

En todos los casos el valor numérico mínimo de la variable Y corresponderá a cero sobre la distribución normalizada y el valor numérico máximo de Y corresponderá a 1.0. Y teniendo determinadas las dimensiones de la distribución, normalizada la variable YNORM expresa la distribución en su forma equivalente acumulativa de frecuencia probando un valor de YNORM para cada paso de simulación usando un número aleatorio.

Todo lo descrito al tratar dependencias parciales asegura que sobre una serie de ejemplos reiterativos de X y Y todos los valores X-Y caen originalmente dentro del área definida resaltada, y la distribución de Y como una función de X será en la forma exacta originalmente especificada en el Paso 3 de la tabla 15. O más simplemente, el proyecto que se acaba de describir asegura que los valores probados de X y Y sobre una serie de pasos de simulación tendrán precisamente la misma dependencia parcial que se observaría cuando se hace la primera marca en una gráfica como la de la figura 22 y 23.

Ahora se verá como es este trabajo en el análisis real de simulación. En la tabla 16 se ha preparado un dibujo esquemático de los pasos conceptuales probando dos variables, exhibiendo la dependencia parcial. El objetivo conceptual es la manera de probar valores de antemano con las gráficas reales acumulativas de frecuencia anteriores. Como se verá más adelante, la serie real de pasos en una computadora tiene que modificar a algunos.

La tabla 16 muestra como se procedería en un paso. Desde que se prueban dos variables aleatorias se necesitan dos números aleatorios, aunque en este caso las variables se relacionen una con otra. Se puede emplear el primer número aleatorio, RN_1 , para probar la distribución X y el segundo, RN_2 , para probar la distribución normalizada YNORM (paso 1 de la tabla 16). Teniendo determinado un valor de X para el paso se usa X como punto de entrada en la gráfica X-Y.

Con el valor de X en el eje ordenado se puede determinar YMIN y YMAX de la intersección de los puntos de los límites superior e inferior. Estos valores límite de Y prueban el valor de YNORM que se utilizan en la ecuación (2) y que resuelve el valor de Y para el paso. Para los

pasos sucesivos de la tabla 16 solamente son repetidos como se muestra para los valores nuevos de X y Y.

Si el análisis de simulación se hace por computadora (como es normalmente) los pasos de la tabla 16 se modifican algo. La razón principal que se necesita para modificar el procedimiento es que la gráfica real empleada en el paso 2 o tabla 16 no estará disponible en la computadora. Así en lugar de eso, se tendrá que proporcionar a la computadora las ecuaciones del límite inferior y el límite superior como funciones de X.

De esta manera, una vez que se prueba el valor de X, este es substituido en la ecuación del límite inferior y se calcula el valor de $YMIN_x$, (como opuesto a la lectura del valor a graficar directamente). Se determina similarmente, el valor de $YMAX_x$, para el paso por resolver la ecuación del límite superior para el valor probado de X. Estos valores se usan en el valor probado de YNORM entonces la ecuación (8) se soluciona para Y en la misma forma anterior.

Si las ecuaciones del límite superior bajan líneas no conocidas por el analizador se puede leer en algunas computadoras puntos X-Y en cada límite y dejar que la computadora derive ecuaciones usando un polinomio para aproximar la subrutina (una biblioteca normal con subrutina de algoritmo en computadoras más científicas). Se resumen todos estos pasos para aplicaciones de computadora en la tabla 17.

Este punto se observa práctico, en una situación del mundo-real involucrando dependencia parcial entre dos variables aleatorias.

Ejemplo Numérico

Un geólogo con "éxito" de una Compañía de Aceite, ha preparado una nueva recomendación exploratoria de un pozo en un Campo ABC. Aunque tenía algunos datos de un campo cercano, consideró el proyecto de "clasificar un pozo de petróleo en exploración" en categorías. Para evaluar algunas incertidumbres decidió usar el método de simulación.

El primer paso en su análisis era formular un modelo económico para el proyecto. Después de estudiar cuidadosamente la técnica de simulación inventó un modelo para tener tres sub-secciones fundamentales:

1. Una rutina generaliza una distribución del total de reservas recuperables del campo. Variables aleatorias: acres productivos, espesor neto productivo, y factor primario de recuperación en unidades bl/ acre-pie.
2. Una contabilidad de desembolsos totales de desarrollo. Variables aleatorias: número de pozos produciendo, número de agujeros secos desarrollados, costos de perforación.
3. Una proyección de rentas futuras y para operar gastos se puede calcular un valor neto presente discontinuo de rentas futuras. Variables aleatorias: potencial inicial del pozo, costo de operación.

Cada paso de simulación incluiría estas tres sub-secciones, junto con parámetros relacionados a encontrar aceite, precios de crudo, velocidades de descuento, etc. La última variable dependiente sería un valor neto presente del proyecto. Después de correr suficientes

pasos se puede estimar la distribución total, así como el valor de la utilidad esperada.

El paso próximo en el análisis es averiguar las distribuciones de valores posibles de cada variable aleatoria que pueden suponerse en el área del proyecto. Para un repaso del campo cercano concluyó que el espesor productivo neto inicial en datos del potencial de los 31 pozos en el campo, probablemente sería representativo del espesor productivo y potenciales en la área del proyecto. Su primer paso era tabular los datos de los 31 pozos en el campo, como se muestra en la tabla 18.

El geólogo estudia el potencial inicial (IP) datos y notables variaciones en el espesor productivo neto y potencial. Por ejemplo, el pozo C-5 con 20 pies de espesor tiene solamente 75 BPD, mientras el pozo B-1 con 10 pies de espesor fue terminado para 197 BPD. Reconoció eso cuando se probaron valores potenciales iniciales en la parte 3 del modelo, al probar tiene que incluir una dependencia en pago neto si tal dependencia existe. Pero al observar los datos no era cierto, si se relacionaron con el pago neto unos, y algunos de otra manera.

Pregunta: ¿Hay una dependencia entre el espesor neto productivo y el potencial inicial? Si es así, ¿como puede la dependencia parcial ser descrita y usada en este análisis de simulación?

Solución: Ahora que ya se tiene experiencia en manejar dependencias parciales se le puede aconsejar a este geólogo como proceder. Primero, se requiere hacer una gráfica de la productividad neta y datos de IP para observar si aún existen datos de dependencia parcial. Para que sea consistente la nomenclatura se deja al espesor neto productivo como la variable aleatoria X, y el IP la variable aleatoria Y. La gráfica de datos X-Y de la tabla 18 está dada en la gráfica de la figura 25.

Se observa al graficar que hay, realmente, una dependencia parcial entre las dos variables. Como incrementos de espesor por pozo la productividad se incrementa (como tiene que ser en virtud de la ecuación de Darcy), pero hay, sin embargo una considerable "banda" de variación de productividad para valores de espesores productivos dados.

Así probablemente sería aconsejable probar valores de espesor y de potencial usando las técnicas para dependencia parcial que han sido discutidas. Para llevar a cabo esto se necesita seguir los pasos de preparación de datos de la tabla 15.

El primer paso, el de graficar, se cumplió. El límite alrededor del área de posibles combinaciones X-Y ha sido dibujado en la figura 25 como un polígono de cuatro caras. La selección de líneas rectas para la parte superior e inferior resultan ser, por supuesto, discretionales. Pero parece ser imparcialmente bueno "entrar", dado el límite de número de puntos observados de los datos. En el paso 3 se necesita describir la variabilidad dentro de la región limitada como una función de productividad neta.

Se juzgaría que la variación es esencialmente aleatoria. Se asume que la distribución normalizada de una función de espesor, YNORM, será distribución uniforme o rectangular, al tener su valor mínimo en cero y su valor máximo 1.0. Y desde que la frecuencia acumulativa de una distribución uniforme es una línea recta se puede graficar la frecuencia acumulativa de la variable normalizada YNORM directamente, figura 26.

Tabla 18.**Tabla de espesor neto productivo y potencial inicial, datos para 31 pozos de un campo**

Número de Pozo	Espesor neto productivo, pies	Potencial Inicial del Pozo, BPD
B-3	5	25
C-1	5	102
D-4	6	160
C-7	9	72
B-1	10	197
D-6	10	175
C-8	10	112
E-1	11	134
A-5	12	68
D-7	15	98
C-6	17	195
D-1	19	172
E-7	20	150
C-5	20	75
E-4	23	110
D-8	26	172
A-3	28	200
B-2	30	140
D-2	31	93
C-4	37	190
E-3	38	139
E-6	40	225
A-2	45	245
A-4	47	182
C-3	54	230
E-2	59	170
A-1	65	225
E-5	66	185
D-3	70	275
D-5	78	237
C-2	80	300

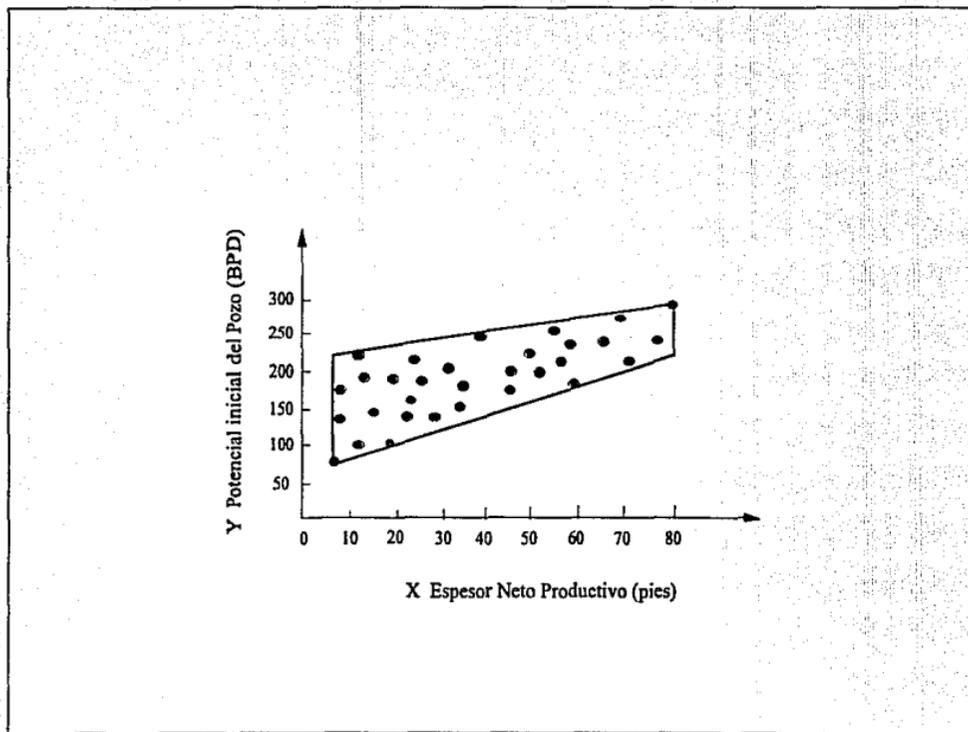


Figura 25. Gráfica de Espesor neto productivo y potencial inicial (IP), datos de la tabla 18. Cada punto graficado representa uno de los 31 pozos listados. Una dependencia parcial parece evidente, y el área de los datos X-Y ha sido limitada por el polígono de cuatro, como se muestra.

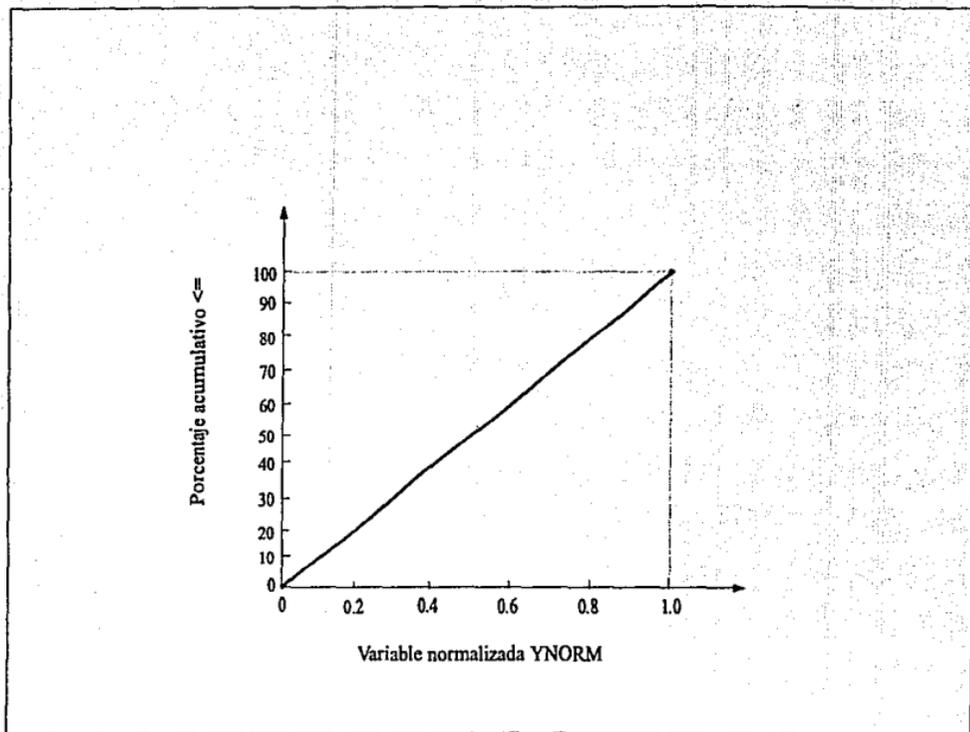


Figura 26. Gráfica de Frecuencia Acumulativa de la variable normalizada YNORM representando la variabilidad del potencial inicial como una función del espesor neto productivo de la figura 25.

Es poca la siguiente información que necesita dar el geólogo, la distribución de X, espesor neto productivo, expresada en una base acumulativa de frecuencia. El grueso de datos de la tabla 18 define la distribución de la variable X, pero se requiere solamente una gráfica de distribución acumulativa de frecuencia equivalente. En simulación se puede ir directamente a una acumulativa y ni siquiera incomoda con la distribución misma.

El grueso de los datos es arreglado en orden numéricamente ascendente y se puede obtener la frecuencia acumulativa graficando los puntos con solamente un mínimo de cálculos por lado. Más específicamente como se observa en la tabla 19.

Tabla 19.

Tabla de frecuencia acumulativa de la variable aleatoria X
(espesor neto productor).

Valores de la Variable Aleatoria, X, Espesor Neto Productivo (pies)	Número de Pozos que tienen espesor igual o menor que X	Porcentaje de Pozos que tienen espesor igual o menor que X
5	2	6.5 %
10	7	22.6 %
15	10	32.3 %
20	14	45.0 %
30	18	58.0 %
40	22	71.0 %
54	25	80.7 %
70	29	93.5 %
80	31	100.0 %

Se puede graficar estos porcentajes calculados contra el valor de X en coordenadas cartesianas, como se muestra en la figura 27.

Ahora todas esas gráficas se analizan a través de los pasos de la tabla 16. Se requieren dos números aleatorios de la tabla de números aleatoria y para el ejemplo se supone que los números aleatorios obtenidos son 65 y 31. Los pasos para determinar un valor de X (espesor) y Y (IP) para este paso son:

1. Probar valores de X y Y NORM:

Entrar a la figura 27 con el número aleatorio 65 resulta un valor de X de 35 pies.

Entrar a la figura 26 con el número aleatorio 31 da un valor de Y_{NORM} de 0.31 (tienen que checar estos valores, claro)

2. Usando el valor de X= 35 pies se entra a la gráfica X-Y de la figura 25 en la abscisa a 35 pies y con una línea recta (imaginaria) ordenado arriba un valor de X= 35 pies. La intersección del eje ordenado con los valores mínimos y máximos de Y da valores de:

$$Y_{MIN_{35}} = 102 \text{ BPD}$$

$$Y_{MAX_{35}} = 234 \text{ BPD}$$

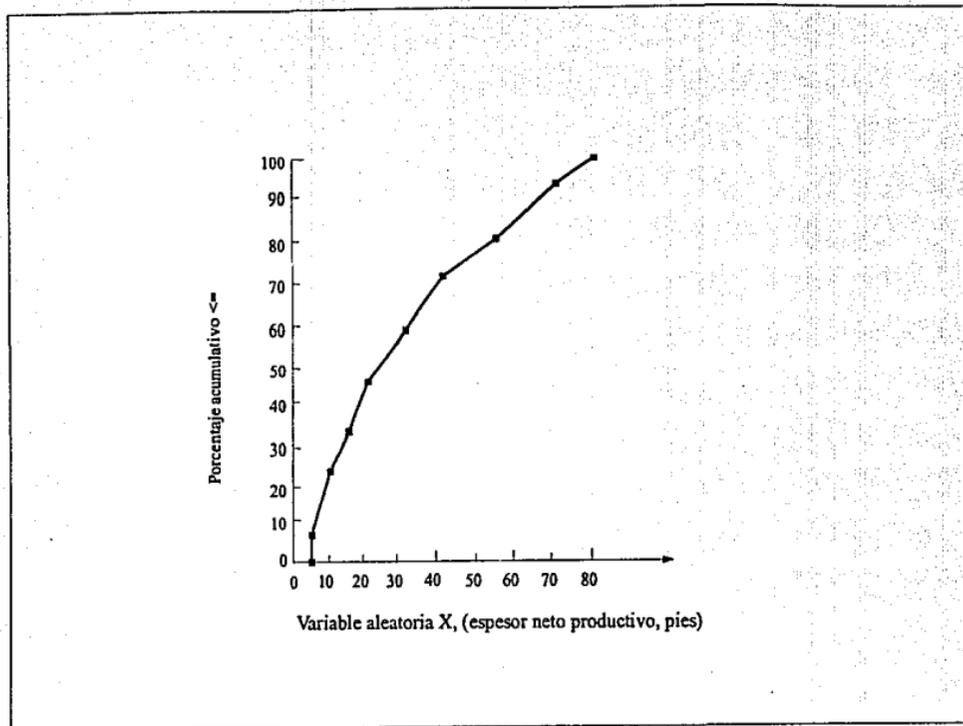


Figura 27. Gráfica de Frecuencia Acumulativa de la variable aleatoria X , espesor neto productivo. Datos de la Tabla 18.

3. Resolviendo la ecuación 8 para Y:

$$Y = Y_{MINx} + (Y_{MAXx} - Y_{MINx}) (Y_{NORM})$$

$$Y = 102 \text{ BPD} + (234 \text{ BPD} - 102 \text{ BPD}) (0.31)$$

$$Y = 143 \text{ BPD}$$

Los valores usados en el paso para X y Y serían $X = 35$ pies y $Y = 143$ BOPD. Se pueden practicar los pasos con otros números aleatorios usando 80 y 55. Tiene que finalizarse con valores de $X = 52$ pies y $Y = 208$ BPD.

Si el geólogo planeó correr este análisis en una computadora se puede analizar este punto, y discutir un ejemplo real en la sección siguiente en aplicaciones y ejemplos.

Como una alternativa, se puede juzgar una tendencia probable entre dos variables y entonces asignar un rango, o banda en cualquiera de los dos lados de la línea de la tendencia que permite la variabilidad de Y como una función de X en el análisis. Se ilustra esto en la figura 28. Con tal proposición se puede definir la región limitada de valores dentro de cuáles X y Y pueden ocurrir. Lejos de describir la distribución de Y dentro de la solución como una función de X probablemente tendría que usarse una distribución uniforme si algún valor de Y se considera dentro del rango igualmente probable de ocurrir. Si la línea de tendencia del centro es más probable se puede representar la variación de Y dentro de los límites con una distribución triangular.

3. En la figura 23(b) y (c) se sugirió que una distribución triangular simétrica y una desviación de la línea recta pueden ser usadas en las respectivas ocasiones para representar la variación observada de Z y T. Estas distribuciones específicas pueden ser usadas porque la mayoría del grupo de los valores probables quedan siempre a la misma distancia del límite inferior al límite superior para todos los valores de W y S. Pero ¿qué sucede si la posición del grupo de valores probables cambia más (con respecto a su distancia entre la solución) con X como los rangos de X_{MIN} a X_{MAX} , como en la figura 29?

No se puede representar todo el infinito de distribuciones Y (una para cada valor de X) por una sola, distribución representativa. En realidad, cada distribución de Y tiene una forma separada, específica. Cerca del rango inferior de X el modo se desvía de la línea recta hacia Y_{MIN} . Como X se incrementa el valor más probable se mueve hacia Y_{MAX} , como se mostró en la inserción superior en la figura 29.

La solución es que en una entrada extra dada a la computadora: la ecuación de la línea representa el grupo de valores más probables como una función de X. Así, se requiere encontrar la ecuación de la curva como una función de X, o dar a la computadora un grupo de puntos coordenados para la línea y dejar que la computadora ajuste una ecuación polinomial a través de los puntos.

Conociendo la ecuación del grupo de puntos más probables no significa que se ha reducido el número infinito de distribuciones triangulares Y_{NORM} a una, pero soluciona el problema porque la computadora puede probar para la distribución específica Y_{NORM} dado un valor de X. Se explican los mecanismos de como la computadora hace esto en la Sección IV de este capítulo.

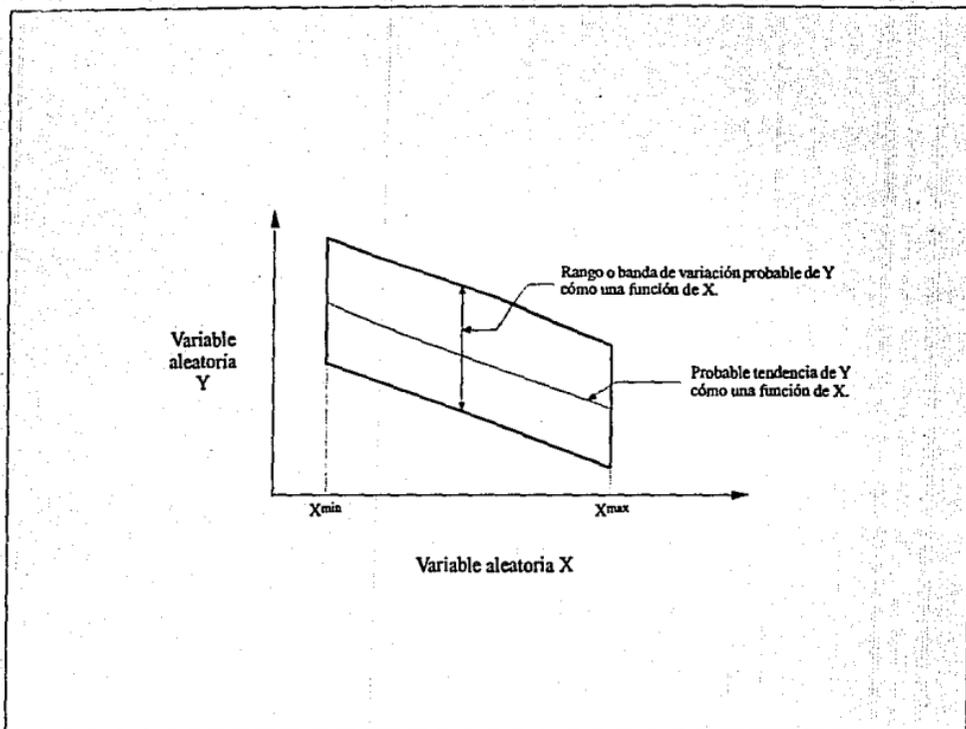


Figura 28. Ejemplo de estimación de la línea de tendencia y asignación de un rango en cada lado de la tendencia aproximando la probable variación de Y como una función de X.

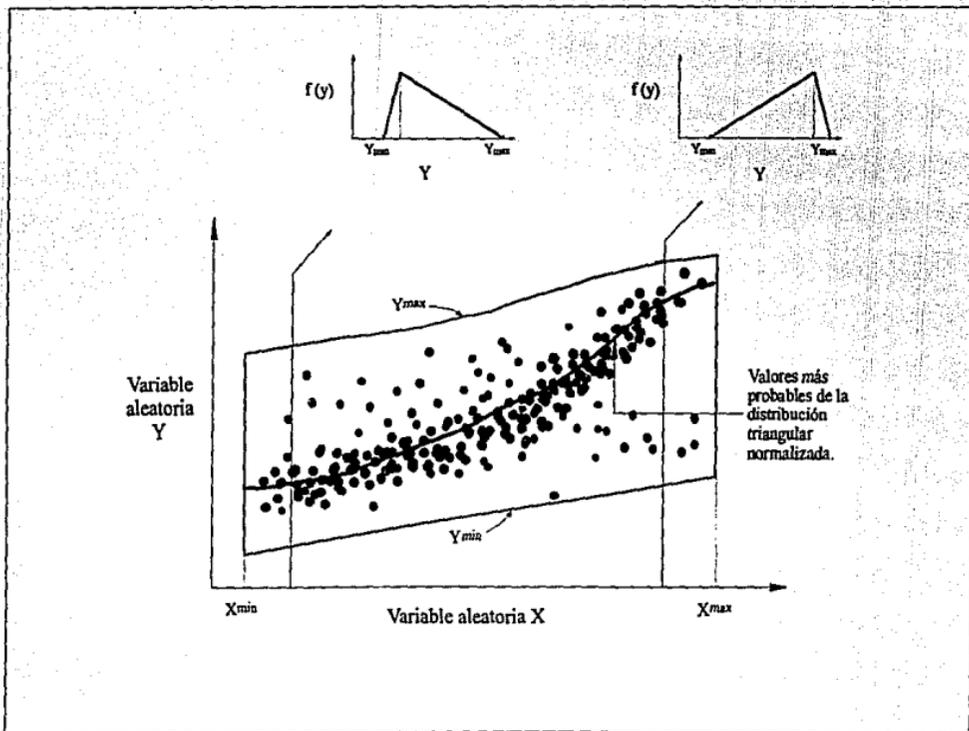


Figura 29. Gráfica de datos X-Y en la cuál la posición de los valores más probables cambia entre X_{\min} y X_{\max} . Para este caso la posición de los valores más probables puede ser tratada como una función de X.

El punto aquí es si la variabilidad de Y se da como una función de X. En la figura 29 se puede manejar la dependencia parcial observada para especificar la ecuación del grupo de valores más probables como una función de X.

4. En las discusiones de dependencia parcial siempre se ha acentuado el hecho concerniente a la gráfica X-Y que la variabilidad de Y es como una función de X. Pero ¿qué hay acerca de la variabilidad de X? ¿Como se explica el hecho aparente de que en la figura 25 la mayor parte de los datos (se ven como una variación a lo largo del eje X) ocurre en el rango inferior de valores de X? Si se estudia a través de la gráfica se puede observar ésto agrupando los datos en el rango de $X=5$ pies a $X=40$ pies.

Esto llevado cuidadosamente por el hecho que la primera muestra de X para la distribución determina el punto de entrada en la figura 25. Y la observación hecha solamente significa que la distribución de productividad neta es la desviación de la línea recta con la mayor parte de su área cerca del fin inferior del rango. Pero por los mecanismos descritos más temprano de porque el procedimiento probado usa una gráfica acumulativa de frecuencia se trabaja exactamente duplicando esta variación observada.

La distribución acumulativa de frecuencia controlará el valor de X para cada paso. Y controlará la variabilidad de X. El proyecto pasa la dependencia parcial llevando cuenta completa de esto, y sobre las series de paso los valores X-Y probados exhibirán las relaciones exactas de dependencia parcial de la figura 25.

5. Los procedimientos para tratar dependencias parciales en análisis de simulación involucran un poco de trabajo extra por parte del analizador. ¿Lo vale?. Por ejemplo, si allí existe una dependencia parcial tal como se mostró en la figura 30 ¿por qué no se puede ajustar una función de mínimos cuadrados por los puntos de datos observados y tratar a Y como una función directa de X usando la ecuación de la función?

La respuesta es sí, se puede hacer esto. Y realmente, simplificaría el análisis. Para cada paso simplemente se probaría un valor de X, se substituye en la función de mínimos cuadrados y se soluciona para Y (en lugar de probar un valor de Y NORM y resolver la ecuación 8).

Seguir tal curso, sin embargo, es desconocer lo explicado por simulación usando - variabilidad. El resultado de valores no reflejarían la variación substancial de valores Y que se pudieron tener en cualquiera de los dos lados de la función.

Es casi como si se hubiera decidido construir un garaje nuevo para proteger un brillante automóvil nuevo y entonces cuando se termina el garaje se decide dejar el auto estacionado en el exterior de la entrada para coches porque es demasiado trabajo extra el abrir la puerta del garaje cada vez que se requiere meter el cochel. Puede ilustrar sus propias conclusiones como la viabilidad de una proposición poco-corta.

6. Finalmente, se mencionará adelante en esta sección, en el mecanismo de simulación que no es válido usar el mismo número aleatorio, al probar dos o más distribuciones (vea la figura 19 y discusión anexa). Algunos autores sobre simulación proponen que usar el mismo número aleatorio para probar diversas distribuciones es, en realidad, un modo de tratar las dependencias parciales, en materia del procedimiento tal como se ha discutido. Pero la propuesta de usar el mismo número aleatorio dos veces, al explicar la dependencia no es válida!

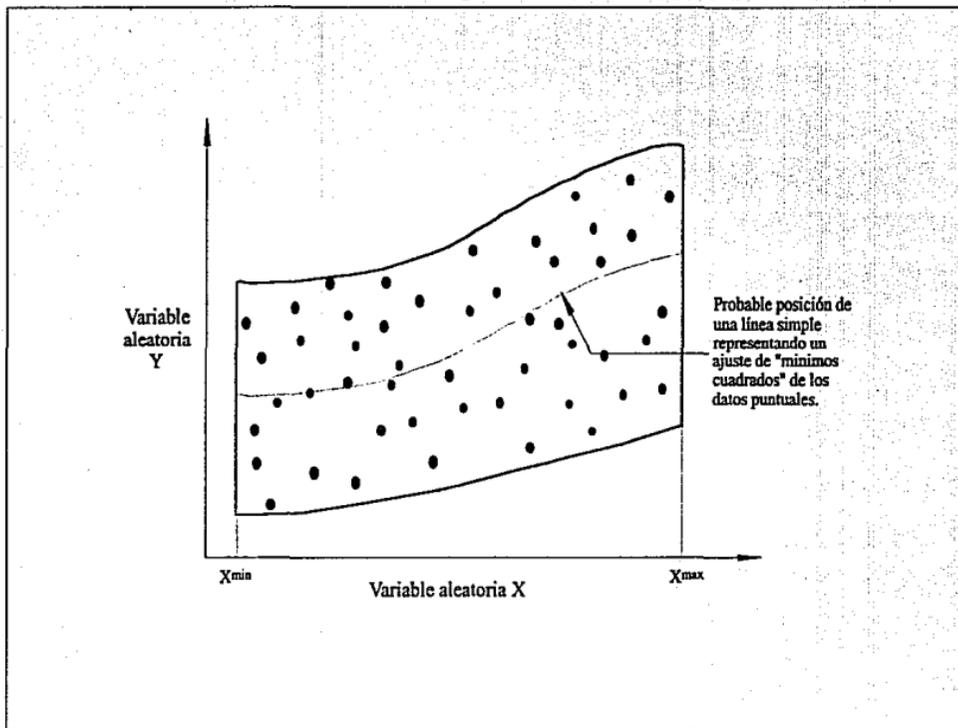


Figura 30. Gráfica de dos variables aleatorias parcialmente dependientes en la cuál la línea punteada representa el ajuste por mínimos cuadrados de los datos.

La figura 19 supone que las variables X y Y son parcialmente dependientes y el límite de la región dentro de la cual las dependencias existen, se muestra en la figura 31. Si se usa el mismo número aleatorio al probar las distribuciones X y Y por separado (como en la figura 19) esto hará que sobre una serie de pasos los valores probados de X y Y mientan, a lo largo de una línea representada por la curva discontinua. Si el número aleatorio es bajo, ambos X y Y tendrían valores bajos. Si el número aleatorio de X fuera grande, Y debería ser grande.

La curva por sí misma puede tener una forma ligeramente diferente (dependiendo de la forma de las distribuciones X y Y), pero todos los puntos probados mentirían exactamente a lo largo de la curva. Esto, por supuesto, otra vez derrota el objetivo de tratar de explicar la variabilidad de Y como una función de X. Si se usa el mismo número aleatorio dos veces (con la esperanza de tratar la dependencia parcial), no se puede para relacionar la dependencia parcial observada en la figura 31. Y la regla aún existe - no use el mismo número aleatorio para probar dos o más distribuciones.

III.2.6 Cálculo de EMV y Preparación del Despliegue de Gráficas (Paso 6)

Después de un número suficiente de pasos repetidos de simulación, esencialmente se ha completado el análisis. La única tarea pendiente es usar la probabilidad de valores para determinar el valor esperado de aprovechamiento (EMV) y hacer un despliegue gráfico más útil y manejable. Para ilustrar como se termina esto, suponga que se hace un análisis de simulación de un prospecto de perforación exploratoria en el cual la utilidad fue definida en términos de un beneficio VPN usando una tasa de descuento del 15 por ciento. Y suponga aún, que después de hacer 150 pasos de simulación los valores actuales calculados de la ganancia VPN se dan en la tabla 20, columna 2.

Tabla 20.

Resultados de 150 pasos de un Análisis de Simulación de un prospecto de perforación exploratoria. Utilidad expresada como VPN (miles de dólares), tasa de descuento $i_0=15\%$.

Rango* de NPV (M\$)	Número de Valores calculados de VPN en cada rango.	Frecuencia relativa en cada rango
-\$75 a -\$50	3	3/150 = 0.020
-\$50 a -\$25	6	6/150 = 0.040
-\$25 a 0	12	12/150 = 0.080
0 a +\$25	15	15/150 = 0.100
+\$25 a +\$50	60	60/150 = 0.400
+\$50 a +\$75	30	30/150 = 0.200
+\$75 a +\$100	12	12/150 = 0.080
+\$100 a +\$125	10	10/150 = 0.067
+\$125 a +\$150	2	2/150 = 0.013
	150	1.000

* Valor mínimo: -\$72 M
Valor máximo: +\$148 M

Tal tabulación resulta de dividir el rango completo de las ganancias calculadas entre una serie de sub-rangos (nueve en este caso) y contando el número de valores calculados los cuales caen en cada uno de los nueve rangos. Por ejemplo, tres de los pasos resultaron en VPN

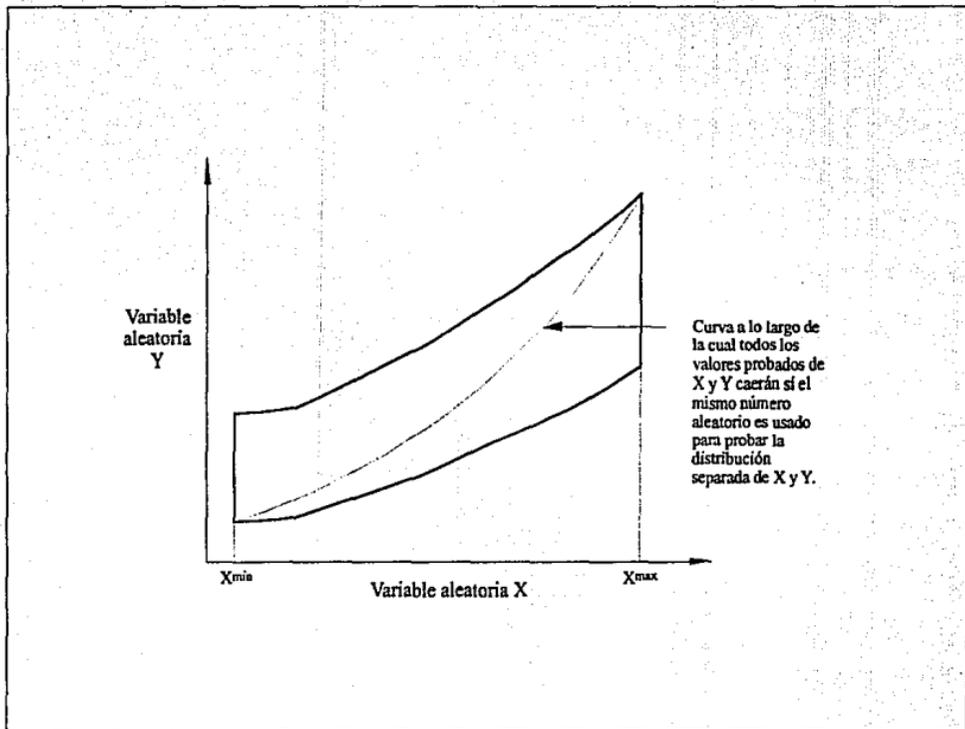


Figura 31. Gráfica de dos variables aleatorias parcialmente dependientes X y Y. Si el mismo número aleatorio es usado al probar las distribuciones separadas (como en la figura 19), la localización de los puntos probados caerán a lo largo de la curva discontinua. En todos los valores que caigan a lo largo de la curva no podrá ser posible una variabilidad de Y como una función de X.

pérdidas de entre -\$50 M y -\$75 M, seis pasos resultaron con pérdidas de entre -\$25 M y -\$50 M, etc. Y por supuesto, la suma de todas las frecuencias de la columna dos es el número total de pasos que se hicieron.

La columna 3 simplemente expresa las frecuencias de la columna 2 como una fracción del número total de pasos. A este punto todo lo que se ha hecho es arreglar los 150 valores numéricos de la ganancia en una tabla de frecuencias relativas.

El paso siguiente -y realmente importante en el esquema entero de simulación- involucra una suposición. La suposición es que las frecuencias relativas actuales de la columna 3 son representativas de la probabilidad de tener una utilidad en cada rango. Se dice aquí que las frecuencias relativas observadas por un diseño de 150 combinaciones cuidadosamente "probado", pueden ocurrir al ser representadas por sus proporciones, de modo que, varias variables que se habrían obtenido probarán todas las combinaciones posibles (las cuales para variables aleatorias continuas, es un número infinitamente grande).

Esto es similar a probar las costumbres de lectura de 200,000,000 adultos. Se puede hablar a cada persona, si se tiene el tiempo y recursos, pero un procedimiento más sencillo sería entrevistar a algunas personas seleccionadas como una manera de eliminar cualquier perjuicio en el procedimiento de muestreo. Si 12% de ésta muestra cuidadosamente seleccionada regularmente lee la edición de París del Herald-Tribune probablemente se estaría seguro al proyectar que 12% de los 200 millones de personas lee el Herald-Tribune. Conceptualmente se hace la misma cosa con simulación.

Los valores calculados de utilidad para cada uno de los reiterativos representantes de una muestra de utilidad es para el mismo espacio teórico (o población) de valores posibles de ganancia. Pero los valores calculados no son solo cualquier valor, son valores cuidadosamente muestreados tal que las distribuciones de valores usados al calcular la ganancia son de la misma forma exacta, como originalmente fueron especificados (Paso 4).

Así, cuando se termina un análisis de simulación y se calculan las frecuencias relativas de los valores actuales de ganancia, se supone que éstos son representativos de las probabilidades de ocurrencia de varios niveles de utilidad. Esta suposición es perfectamente válida si se hace un número suficiente de pasos. Tan grande como cuando se hacen las verificaciones para asegurar que un número suficiente de pasos han sido hechos, nada se tiene que temer en cuanto a esta sutil, pero importante suposición.

Con esta pequeña explicación, y la seguridad de que para este análisis 150 pasos son una cantidad suficiente, se pueden traducir, o interpretar las frecuencias relativas de la columna 3 de la tabla 20 como probabilidades. Al final el cálculo de EMV es entonces en forma directa (tabla 21), y la utilidad esperada VPN de + \$41.33 M que es el parámetro de decisión con el cual se determinaría si o no perforar el prospecto. Desde que la expectación es positiva el prospecto deberá ser una inversión aceptable. Por supuesto, esta inversión tiene que compararse con otras alternativas disponibles de inversión si alguna repercute sobre el capital existente.

El valor esperado de una decisión alternativa, es la base fundamental sobre la cual se aceptan o se rechazan las decisiones que involucran incertidumbre. Consecuentemente, teniendo calculado un EMV para la salida de un análisis de simulación (tal como en la tablas 20 y 21) se cumple el objetivo. Es útil, sin embargo, también presentar para tomar la decisión los resultados del análisis de simulación se despliegan en varias gráficas. Un despliegue evidente

es una "fotografía", mostrando la distribución completa de los resultados obtenidos del análisis de la simulación.

Tabla 21.

Tabulación de resultados de los valores esperados para el ejemplo de la tabla 20.

Rango de VPN (M\$)	Probabilidad de ocurrencia (col. 3 de la tabla 6)	Valor del punto medio del VPN (M\$)	VPN esperado (col. 2 x col. 3)
-\$ 75 a -\$ 50	0.020	-\$ 62.5 %	-\$ 1.25 %
-\$ 50 a -\$ 25	0.040	-\$ 37.5 %	-\$ 1.50 %
-\$ 25 a 0	0.080	-\$ 12.5 %	-\$ 1.00 %
0 a +\$ 25	0.100	+\$ 12.5 %	+\$ 1.25 %
+\$ 25 a +\$ 50	0.400	+\$ 37.5 %	+\$ 15.00 %
+\$ 50 a +\$ 75	0.200	+\$ 62.5 %	+\$ 12.50 %
+\$ 75 a +\$100	0.080	+\$ 87.5 %	+\$ 7.00 %
+\$100 a +\$125	0.067	+\$112.5 %	+\$ 7.54 %
+\$125 a +\$150	0.013	+\$137.5 %	+\$ 1.79 %
	1.000		+\$ 41.33 %
VPN esperado con $i = 15\%$ (EMV)			

Para la salida de datos de la tabla 20, la distribución aparecería como en la figura 32. Un cuadro tal como este, puede dar al que decide una ayuda para los rangos de las posibles utilidades, las probabilidades relativas de pérdidas contra ganancias, una idea del rango de utilidades las cuales son más probables, etc.

Otro despliegue útil es convertir la distribución de utilidad a su equivalente en frecuencia acumulativa. La ventaja de tener la gráfica acumulativa de frecuencia disponible es que se pueden leer las probabilidades y dar niveles de ganancia directamente, sin tener que detenerse y agregar todas las áreas de cada barra del histograma.

Estas gráficas pueden ser impresas en cualquiera de las dos maneras: la probabilidad acumulativa de una utilidad o pérdida dada, o la probabilidad acumulativa de un nivel o más de utilidad dado. La previa es construida por acumulación de áreas bajo la distribución, como incrementos de utilidad. La posterior es construida por acumulamiento de las áreas bajo la distribución, como decrementos de utilidad. Usualmente el que decide está conciente de la probabilidad de por lo menos un nivel dado de utilidad así la posterior de las dos opciones es usualmente la más útil. Para la distribución de tal figura 32 como una frecuencia acumulativa puede aparecer como en la figura 33.

La decisión puede usar la figura 33 para contestar todas las preguntas de algo que le pueda interesar, por ejemplo, acerca de la probabilidad de hacer al menos un 15% de retorno. Desde la tasa de descuento usada en los cálculos de VPN se tenía un 15% esto es, en esencia, que se interesa acerca de la probabilidad de que el beneficio VPN sea por lo menos 0. Para determinar esto, puede entrar en la abscisa con 0 de utilidad VPN, se lee arriba a través de la ordenada hasta un valor de por lo menos 0 de VPN (por ejemplo, por lo menos una tasa de retorno de 15%) de 86%.

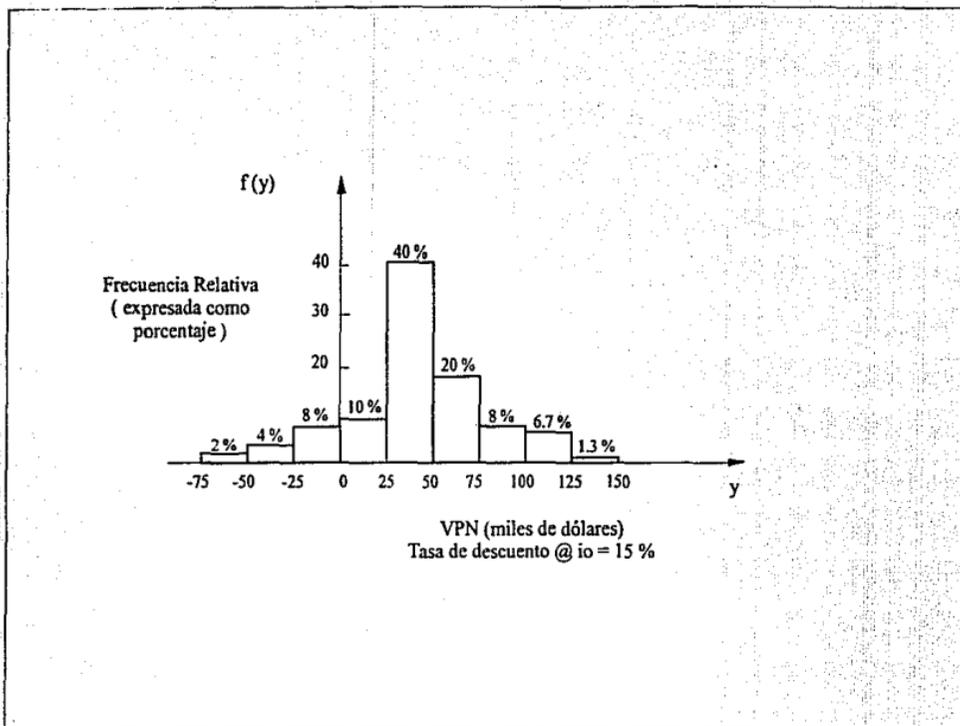


Figura 32. Histograma de Frecuencia Relativa de los resultados de la simulación del análisis del prospecto, tabla 20.

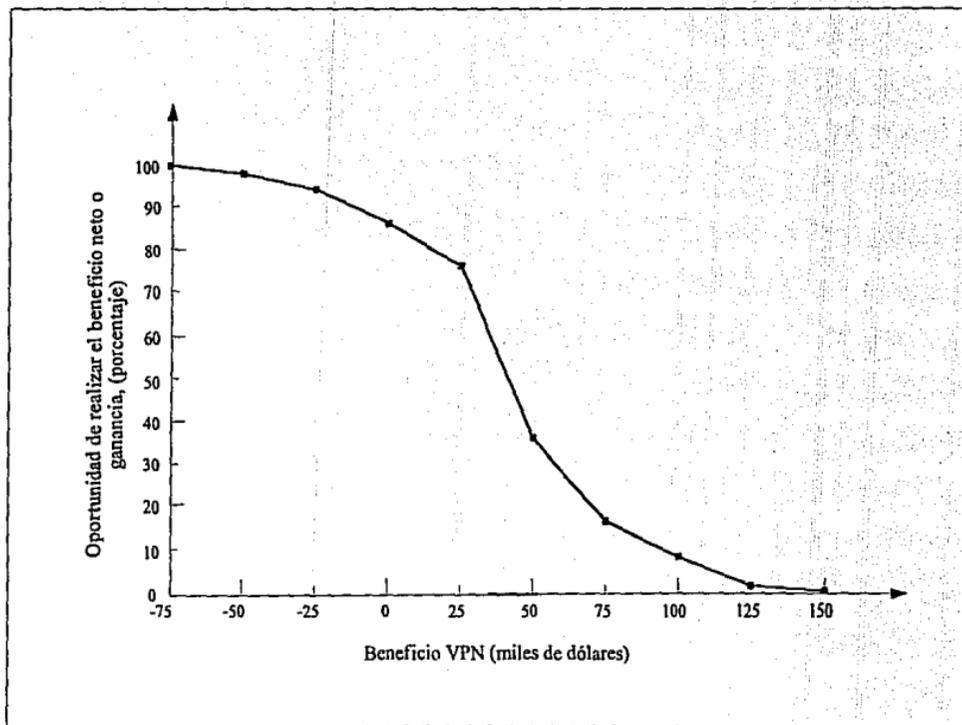


Figura 33. Gráfica de Frecuencia Acumulativa de la distribución de la utilidad de la figura 32.

Suponga que al asegurar los derechos de perforación se tiene que pagar \$50,000 por bono de firma. Si se carga al proyecto con este desembolso al tiempo cero ¿cuál es la probabilidad de por lo menos tener una tasa de retorno del 15%?. La respuesta es 36%. Si se paga un extra de \$50 mil para un bono, el proyecto deberá producir por último una utilidad VPN de por lo menos \$50 mil. Se encuentra la probabilidad de que esto ocurra desde la lectura de la gráfica de frecuencia acumulativa directamente.

Finalmente, se tiene que notar que una vez obtenidos los resultados de los 150 pasos de simulación (tabla 20), el análisis no requirió hacer cualquier riesgo final subjetivo. Todos los juicios y experiencias profesionales en el análisis de los pasos 1-4 en forma de distribuciones de probabilidad, relaciones de dependencia, etc. Todos los puntos anteriores (pasos 5 y 6) eran simplemente aritmética y contabilidad.

Tal vez éste sea el factor más favorable para la simulación. Si se tiene que confiar en los juicios subjetivos, parece razonable esperar que los resultados finales sean menos oscuros, parciales y emocionales, si se hacen estos juicios sobre una variable a la vez antes en el análisis.

Es difícil pronosticar en el tiempo, como todas las distribuciones de probabilidad, interactuarán y afectarán la última utilidad, por lo tanto es menos probable que un acto individual pueda influenciar o forzar una respuesta. Inversamente, usando un juicio subjetivo al final de un análisis discreto de resultados, puede influenciar mucho al analizador. Puede romper un trato simplemente por la (arbitraria?) estimación de un factor de riesgo.

Con estos comentarios se tiene también que notar que en general no es posible pronosticar el rango y forma de la distribución dependiente de utilidad. A veces se ha dicho que la distribución del beneficio será normal, o de algún otro tipo. Hay que ser cauteloso acerca de tales declaraciones porque es muy riesgoso estimarlo realmente.

La forma de la distribución de utilidad final está influenciada por las formas de las distribuciones de las variables aleatorias, la magnitud de los valores numéricos de la distribución, y de la relación (ecuación) la cual describe todas las variables a aprovechar.

IV. EL USO DE COMPUTADORAS PARA ANALISIS DE SIMULACION

En esta sección se ofrecerán los comentarios a cerca de algunos programas de computadora para análisis de simulación. Claro, es solamente una manera de hacer todos los reiterativos cálculos requeridos para definir las distribuciones de beneficio. Aunque algunos exploradores pueden perder el control del problema, una vez que éste va a la "caja negra", no se tiene alternativa factible. Los comentarios aquí serán generales.

No se planea obtener un programa específico porque cada uno parece tener un empleo un poco diferente controlando procedimientos y/o programando lenguajes. Estos comentarios generales deben ser útiles, sin embargo, como guía a los programadores de la compañía que deben instalar los procedimientos para manejo de simulación.

Primero, se expresa como propósito general los programas sobre riesgo. Con esto aumenta el uso y aceptación de simulación y muchas compañías se proponen generalmente desarrollar, el programa de análisis de riesgo (utilizado en simulación) que se puede usar durante toda la organización. Una casa de comercio consulta un número que obtuvo una acción y la empresa realizarla el análisis general de riesgo ofreciendo un costo nominal de \$1.50 por corrida.

La meta es admirable, pero el resultado práctico en muchas ocasiones ha sido un sacrificio en la flexibilidad o para opciones generales. A menudo el geologo en Casper, Wyoming, para realizar sus hallazgos puede tener que "cazar" su prospecto en el formato general del programa de computadora en la oficina, en su casa o su compañía en Houston

El punto de vista opuesto es por quién ha tenido experiencia programando y tiene acceso a facilidades de computadora. Como se verá en un momento el programa de computadora queda esencialmente intacto de prospecto a prospecto excepto por toda la descripción importante de las ecuaciones que se relatan, todas las variables aprovechan la función analítica. A causa de esto usualmente se prefiere cambiar las pocas tarjetas marcadas necesarias para el análisis de simulación específica para cada prospecto. La ventaja evidente es que el programa encaja exactamente dentro de cada prospecto particular de exploración.

Entonces no todo el mundo en exploración es un programador, el programa general agranda la accesibilidad de la técnica a más exploradores. Pero también el programa general tiene que diseñarse para permitir cambios fáciles a las ecuaciones calculadas de beneficio. Así, es más útil tener por lo menos un explorador en cada compañía que pueda hacer el programa con los cambios necesarios en el lugar.

Pero cualquier posición a cerca del propósito general de programas contra programación individual se puede convenir en que es necesario diseñarlos para que la entrada de datos (para el punto de vista del usuario) sea lo más sencillo y eficiente posible. Esto es particularmente real en cuanto a las opciones mecánicas y para lectura en la computadora de las varias distribuciones de las variables aleatorias que el programador desea usar. Se hablará más de esto en un momento.

Un programa de computadora para hacer un análisis de simulación consiste de cinco partes, o facetas:

1. Procedimientos para leer en la computadora la variable aleatoria para distribuciones de

probabilidad.

2. Procedimientos para leer en la computadora las ecuaciones (o serie de pasos) usados para calcular los valores numéricos, para conocer los parámetros, y algunas ecuaciones que se relacionan con dependencias parciales.

3. Un generador de números aleatorios.

4. El programa requiere hacer los pasos reales de simulación.

5. Procedimientos para calcular la relación de datos de salida con distribuciones de beneficio, calcular el valor medio de la distribución de beneficio, y desplegar algunas gráficas discretas generadas por la computadora.

De estas cinco partes solamente la segunda parte varía de análisis a análisis. Las restantes partes 1,3,4 y 5 esencialmente son las mismas para cada análisis.

IV.1 Procedimientos para Distribuciones de Lectura en la Computadora (Parte I)

Esta es probablemente la parte más crítica del programa por lo que se tratará de simplificar los procedimientos de entrada para el usuario. De hecho la consideración y diplomacia usada aquí puede a veces significar la diferencia entre el éxito o el fracaso. Se desea, por supuesto, que el preprograma proporcione las capacidades para poder manejar muchos tipos diferentes de distribuciones de probabilidad tal que el usuario pueda alguna vez emplearlos.

Este puede incluir las distribuciones uniforme (rectangular), triangular, normal, lognormal e histogramas de frecuencia relativa. Pero ¿que pasa si el usuario tiene una distribución la cuál no puede ajustar a una distribución normal? Para leer distribuciones normal y lognormal se puede dar a la computadora el valor de μ y el de σ y poder manejar internamente las distribuciones. Pero ¿si algunos de los miembros del personal no conocen como calcular σ ?

Años atrás, se propuso repasar un programa general de simulación de análisis de riesgo desarrollado por los matemáticos del centro de investigación, del departamento de exploración de una compañía. En este programa la opción para usar cualquier otro tipo de distribuciones continuas en lugar de la uniforme y triangular, era la distribución Weibull.

La respuesta inmediata era probablemente la misma - ¿Qué es exactamente una distribución Weibull? ¡Nunca se habla oído hablar de esto, y hubo que revisar seis libros antes de hallarla mencionada! Bueno, como es una distribución útil en representar la distribución de la vida de una pieza de maquinaria. Es usada ampliamente en sistemas de análisis de veracidad y con una selección cuidadosa de los parámetros numéricos de la función de densidad, la distribución Weibull puede tener cualquier forma imaginable.

Esto fue una brillante elección como una distribución de propósito general -desde un punto de vista matemático. Pero ¿qué hace el pobre del geólogo localizado en Monahans, Texas, si se supone que nunca ha oído hablar de una distribución Weibull?

El punto de todo esto es que se tiene que hacer un esfuerzo para hacer los procedimientos de entrada no-técnicos y tan simples como sea posible, mientras se tenga aún la capacidad y flexibilidad de acomodar muchos otros tipos diferentes de distribuciones como sea posible. Un balance final, cuando el valor de simplicidad es comparado contra el valor de flexibilidad se convence que el juego más versátil de opciones de entrada es:

- A.- La distribución uniforme- El usuario solamente especifica los valores mínimo y máximo de la variable aleatoria.
- B.- La Distribución triangular - El usuario sólo necesita especificar los valores mínimo, más probable, y máximo de la variable aleatoria.
- C.- Las coordenadas X-Y de la gráfica de frecuencia acumulativa de cualquier distribución (otra que no sea uniforme o triangular)

Para permitir que el usuario lea en un juego de coordenadas X-Y de la distribución acumulativa de frecuencia, el programa puede usar virtualmente cualquier tipo de distribución que haya elegido. El único requisito desde el punto de vista del usuario, es que primero se tiene que convertir la distribución que se desea usar a su equivalente frecuencia acumulativa tal que se pueda leer en las coordenadas X-Y.

La hoja de datos de entrada para distribuciones uniformes sólo necesita incluir espacios para la información siguiente:

Nombre de la Variable	NPAY	(espesor neto productor, pies)
Valor mínimo	20	(pies)
Valor máximo	50	(pies)

Para leer una distribución triangular tal como la distribución del costo de perforación todo lo que se necesita es:

Nombre de la Variable	DRLCOST	Costos de perforación
Valor de la variable	100	(miles de dólares)
Valor más probable	130	(miles de dólares)
Valor máximo	200	(miles de dólares)

Para entrar a otra distribución, se provee como valores de entrada de la gráfica acumulativa de frecuencia:

Nombre de la Variable	DRLCOST (costos de perforación miles de dólares.)					
Valor de la variable	100	108.5	115	123	131	140
Corresp. frec. acumul.	0	.0178	.0696	.1719	.3127	.5126
Valor de la variable	146	154	163	174	181	190
Corresp. frec. acumul.	.6459	.7959	.9006	.9674	.9916	1.0000

¡Sencillo! Y esto ofrece completa flexibilidad para manejar cualquier tipo de distribución. El usuario no tiene que conocer como se calcula μ o σ , o conocer qué es una distribución Weibull. Cualquiera que sea la distribución, sólo se necesita convertirla a una frecuencia acumulativa y leer en una serie de coordenadas X-Y. Si se ha dibujado una variable en una

gráfica de probabilidad en papel lognormal, todo lo que se tiene que hacer es leer una serie de puntos X-Y de la línea recta. Estas tres opciones son tan sencillas y versátiles que es difícil recomendar e ir mucho más allá de estos tipos de distribuciones.

En este punto, puede ser de interés entender como la computadora actualmente opera en estos datos así como los valores probados con números aleatorios. Esto es, ¿qué es lo que el programador tiene que hacer en términos de pre-programar estas opciones?. Esto solo consiste en tener la computadora programada con algunas ecuaciones y mantener la codificación. Para esta explicación breve, se usa el símbolo X como la variable aleatoria, CF como la frecuencia acumulativa y RN como un número aleatorio expresado como una fracción decimal.

Distribución Uniforme: Para esta distribución la frecuencia acumulativa es solamente una línea recta como se muestra en la figura 34. La ecuación de una línea recta, dadas las coordenadas de X_1 - CF_1 , X_2 - CF_2 es:

$$\frac{(CF - CF_1)}{(X - X_1)} = \frac{(CF_2 - CF_1)}{(X_2 - X_1)}$$

Rearreglando la solución para X se tiene:

$$X = X_1 + \left[\frac{(CF - CF_1)(X_2 - X_1)}{(CF_2 - CF_1)} \right]$$

Pero para la distribución uniforme $X_1 = X_{MIN}$, el valor mínimo especificado en la hoja de entrada $X_2 = X_{MAX}$, el valor máximo especificado por el usuario, $CF_1 = 0$, y $CF_2 = 1$. Así se pueden substituir estos valores en la ecuación como:

$$X = X_{MIN} + \left[\frac{(CF - 0)(X_{MAX} - X_{MIN})}{(1.0 - 0)} \right]$$

o

$$X = X_{MIN} + [(CF) (X_{MAX} - X_{MIN})] \quad (11)$$

Esta es la ecuación que es programada de antemano. Se especifican X_{MIN} y X_{MAX} como datos de entrada. Para probar la distribución X la computadora obtiene un número aleatorio RN, lo iguala a CF y resuelve para X en cada paso.

Distribución Triangular: es esencialmente lo mismo en este caso. Las ecuaciones que la computadora usa son formas rearrregladas. Estas ecuaciones se resuelven para la probabilidad

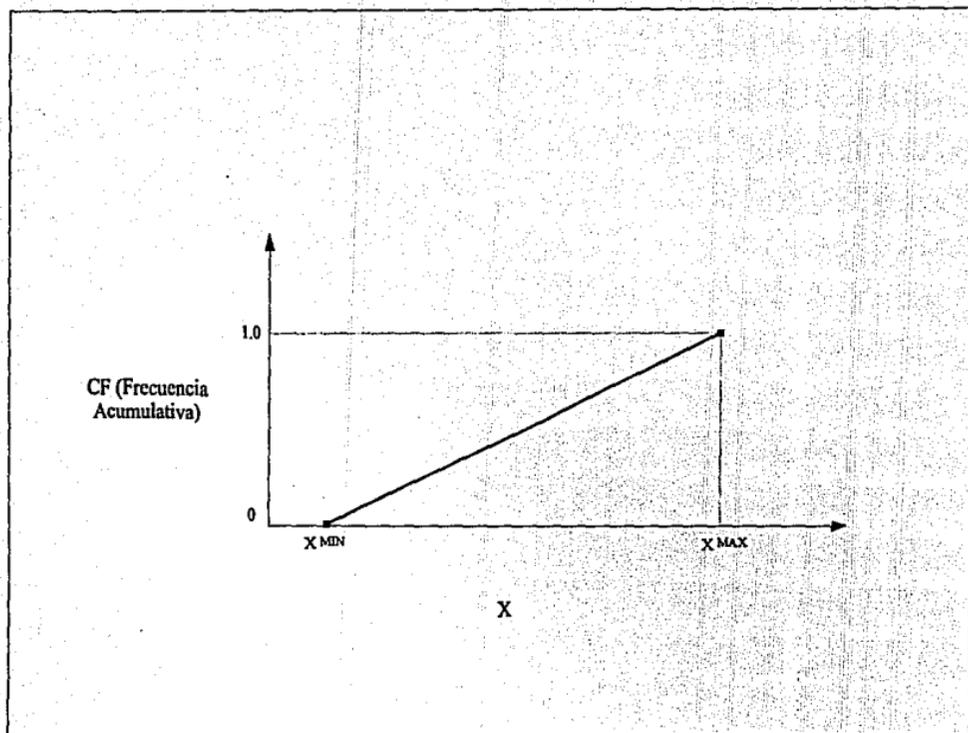


Figura 34. Ejemplo de una distribución uniforme.

acumulativa menor o igual que un valor de X. Al probar valores se hace lo inverso. Se iguala el número aleatorio a la frecuencia acumulativa y se resuelve para X. Por lo tanto las ecuaciones rearrregladas para resolver X son:

Para valores de $X \leq X_2$:

$$X = X_1 + (X_3 - X_1) \left[\sqrt{CF} \frac{(X_2 - X_1)}{(X_3 - X_1)} \right]$$

Para valores de $X > X_2$:

$$X = X_1 + (X_3 - X_1) \left[1 - \sqrt{(1 - CF) \left(1 - \frac{(X_2 - X_1)}{(X_3 - X_1)} \right)} \right]$$

Donde X_1 = valor mínimo de la distribución triangular, X_2 = valor más probable, y X_3 = valor máximo.

Pero se codifican los valores mínimo, más probable y máximos especificados por el usuario como XMIN, XMODE y XMAX, se pueden substituir estos parámetros para X_1 , X_2 , X_3 respectivamente:

Para valores de $X \leq XMODE$

$$X = X_{MIN} + (X_{MAX} - X_{MIN}) \left[\sqrt{CF} \left(\frac{X_{MODE} - X_{MIN}}{X_{MAX} - X_{MIN}} \right) \right] \quad (12)$$

Para valores de $X > XMODE$

$$X = X_{MIN} + (X_{MAX} - X_{MIN}) \left[1 - \sqrt{(1 - CF) \left(1 - \frac{X_{MODE} - X_{MIN}}{X_{MAX} - X_{MIN}} \right)} \right] \quad (13)$$

El único problema ahora es que cuando se obtiene un número aleatorio y se iguala a CF se tiene que conocer cuál ecuación hay que resolver. No se conoce X y así no se conoce si X es menor que XMODE o más grande que XMODE. Pero el truco es que la frecuencia acumulativa menor o igual a un valor de $X = X_2$ siempre será igual a $(X_2 - X_1)/(X_3 - X_1)$. Así se puede reescribir la prueba como $CF \leq (XMODE - XMIN) / (XMAX - XMIN)$ o mayor. Como un resultado de esta lógica, el procedimiento usado por la computadora para probar un valor de una distribución triangular teniendo como parámetros XMIN, XMODE, y XMAX es:

Obtener un número aleatorio, RN
 Igualar CF = RN

$$\text{Si CF es } \leq \frac{(X_{MODE} - X_{MIN})}{(X_{MAX} - X_{MIN})} \quad \text{Resuelva la ec.12 para X}$$

$$\text{Si CF es } > \frac{(X_{MODE} - X_{MIN})}{(X_{MAX} - X_{MIN})} \quad \text{Resuelva la ec.13 para X}$$

Para la Distribución General: El dato de entrada especificado por el usuario es puesto en coordenadas X - CF. Estas coordenadas deberán ser leídas como una serie de puntos de la gráfica acumulativa de frecuencia, de la variable aleatoria como en la figura 35. Se puede observar ahora que en vez de sólo una o dos ecuaciones para resolver X, hay en este caso seis segmentos rectos específicos en la curva. Mientras cada segmento es una línea recta, las pendientes son diferentes. Por lo tanto, cuando se obtiene un número aleatorio y se iguala a CF se tiene que determinar cuál segmento recto de la línea se usará.

Para llevar a cabo esto se necesita programar la ecuación general de una línea recta, aplicable para cada juego de coordenadas X - CF:

$$X = X_N + \left\{ \frac{(CF - CF_N)(X_{N+1} - X_N)}{(CF_{N+1} - CF_N)} \right\} \quad (14)$$

para N = 1,2,3,4,..., (r - 1) y donde CF = RN
 el número aleatorio, (0 <= RN <= 1.0)

Los valores numéricos de X₁, CF₁, X₂, CF₂,..., X_r, CF_r son leídos como datos de entrada y almacenados por la computadora. Cuando se prueba una vez un valor de X, la computadora seguirá esta secuencia:

Obtener un número aleatorio, RN

Igualar CF = RN

Probar: Es CF <= CF₂? En caso de Si, hacer N= 1 y resolver la ecuación 14 para X

En caso de No vaya a la prueba 2.

Prueba 2: Es CF < CF₃? En caso de Si, hacer N= 2 y resolver la ecuación 14 para X.

En caso de no, vaya a la prueba 3.

Prueba 3: Es CF < CF₄? En caso de Si, hacer N= 3 y resolver la ecuación 14 para X.

En caso de no, vaya a la prueba 4.

etc.

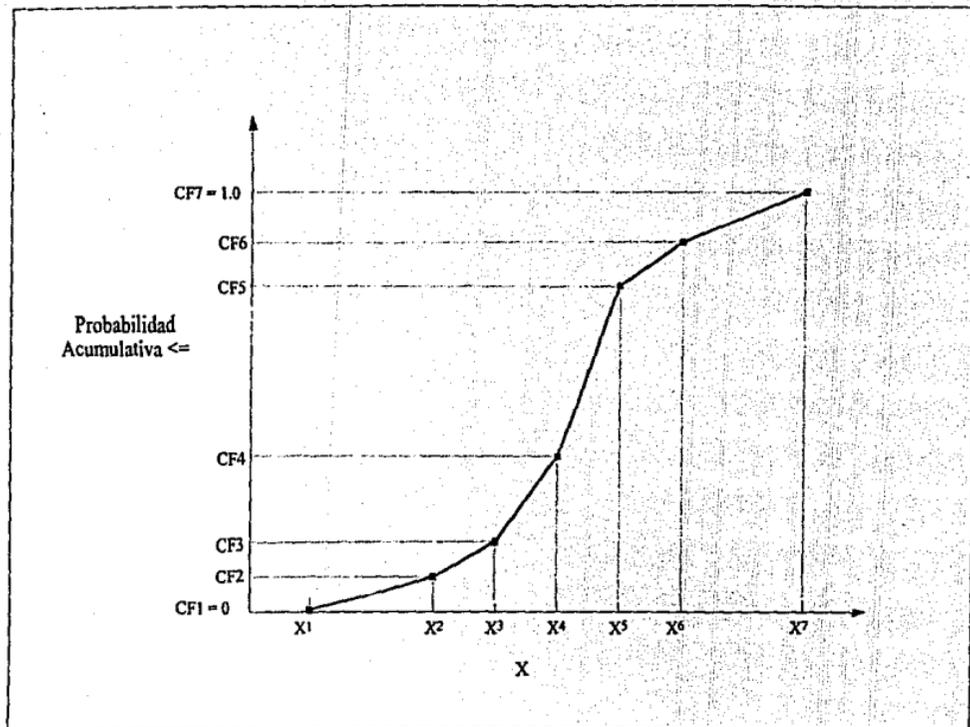


Figura 35. Gráfica de Frecuencia Acumulativa de la variable aleatoria X . Las coordenadas $X_1, CF_1, X_2, CF_2, X_3, CF_3, \dots, X_7, CF_7$ son leídas por la computadora como datos de entrada.

Prueba r-2: Es $CF < CF_{r-1}$? En caso de Si, hacer $N = r-2$ y resolver la ecuación 14 para X
 En caso de no, hacer $N = r-1$ y resolver la ecuación 14 para X.

fin del muestreo de X.

Esta serie de pasos, junto con la ecuación 14 proveerá un medio para probar cualquier distribución expresada en su frecuencia acumulativa en coordenadas X - CF.

Si se está familiarizado con la programación, se deberá tener suficiente detalle en el programa para recibir las coordenadas de la frecuencia uniforme, triangular y acumulativa como datos de entrada. Para aquellos que no son programadores los comentarios probablemente son un poco confusos. Pero el punto importante es observar que cuando una computadora prueba una distribución normalmente soluciona las ecuaciones 10, 10, 12, y 13.

En algunas ocasiones puede ser computacionalmente más eficaz leer (o generar internamente) 100 valores de la frecuencia acumulativa y almacenarlos como valores de la variable aleatoria. Entonces al probar un valor de la variable, el número aleatorio es igualado al subscrito y el correspondiente valor numérico de la variable aleatoria es llamado de la memoria. Si su método es más o menos eficiente dependerá de la capacidad de memoria y de la configuración de la unidad de proceso de la computadora y su equipo periférico.

Se tienen que mencionar algunos puntos finales acerca de lectura en la computadora de distribuciones continuas de probabilidad.

1. Si el programa tiene la capacidad de recibir coordenadas X - CF datos de la frecuencia acumulativa las opciones extra de una distribución uniforme y triangular en forma redundante. La distribución uniforme puede ser lida en datos X - CF como:

Nombre de la Variable	X	
Valor de la variable	XMIN	XMAX
CF, frecuencia acumulativa correspondiente.	0	1.0

La frecuencia acumulativa de la distribución triangular puede ser usada para leer cerca de 6-8 juegos de coordenadas X - CF y entrar con el mismo tipo de dato. Pero la razón para dejar las distribuciones uniforme y triangular como opciones separadas es que se usan ambas frecuencias en el análisis de simulación, y es probablemente un poco más fácil para el usuario registrar sólo los parámetros como se describió previamente.

2.- Anteriormente en la discusión sobre el tratamiento de las dependencias parciales se mencionó que el procedimiento para tratar una dependencia tal como la de la figura 23(b) era especificar una distribución normalizada ZNORM que describa todas las posibles distribuciones de Z como funciones de W. Los parámetros de la distribución triangular normalizada tendrían que ser 0, 0.5 y 1.0 - el 0.5 refleja el hecho de que para cada distribución de Z siempre se posiciono en forma media entre ZMIN a ZMAX

Entonces más adelante, se discutirá la situación en que la posición del modo cambia como la variable aleatoria de la figura 29 se incremente. En este caso no puede entrar una específica distribución normalizada triangular. El procedimiento, en lugar de esto, es proporcionar a la

computadora la ecuación de la línea a través de los modos de las distribuciones Y, figura 29, como una función de X (en adición a las ecuaciones de los límites superior e inferior de Y como una función de X).

Cuando la computadora prueba un valor de X y Y para cada paso los pasos siguientes en la tabla 17 tienen una parte adicional al paso 4. Es resolver para el valor modo como una función del valor probado de X. Este valor calculado es entonces insertado en las ecuaciones 8 ó 9 para probar el valor correspondiente de la distribución triangular apropiada de la variable parcial dependiente.

IV.2 Procedimientos de lectura en los Pasos de Computación, Valores Numéricos para parámetros conocidos, y relaciones de dependencia parcial (Parte 2 del programa)

Esta es la parte del programa de simulación en la cuál se puede cambiar de prospecto a prospecto. Incluye las ecuaciones o serie de pasos involucrados en las etapas hechas, la entrada de valores numéricos para todos los parámetros conocidos (constantes), y alguna ecuación especial relacionada con las dependencias parciales tratadas. Debido a estos aspectos del cambio de programa, es inútil tratar de sugerir formatos específicos de entrada o procedimientos.

Para aquellas compañías que eligen un modelo de simulación de análisis de riesgo con un propósito general esta parte del programa será específica, para los tipos de prospectos de perforación considerados (por ejemplo un modelo para prospectos costa-dentro, costa-fuera, etc). Para ello se puede preferir tener un acceso más flexible, en el cuál el explorador pueda modificar el análisis por así convenir a sus necesidades, lo que significa cambiar el programa por completo.

IV.3 Generador de Números Aleatorios (Parte 3)

Como se ha mencionado en varias ocasiones anteriores cualquier computadora tiene una subrutina generadora de números aleatorios ya programada como una función de biblioteca. Así lo único aquí es asegurarse de que la codificación se mantenga recta y el procedimiento apropiado para proceder a llamar a la subrutina.

IV.4 Programando los Pasos de Simulación (Parte 4)

Esta parte del programa consiste de todas las instrucciones programadas requeridas que actualmente hacen los pasos repetitivos, haciendo una comparación para ver si suficientes pasos se han hecho, y el acceso de los datos de entrada y memoria de valores calculados de probabilidad de cada paso. Hay trucos no usados que se pueden ofrecer aquí, pero esto es usualmente una sección del programa que no puede causar problemas particulares para un programador competente.

IV.5 Procedimientos para Datos de Salida (Parte 5)

Hay muchas maneras diferentes de hacer el despliegue de salida de un análisis. El analista normalmente no desea tener todos los 2000 valores de utilidad imprimidos (se ahogaría en papel!), así uno de los procedimientos usuales de salida es arreglar los valores calculados de probabilidad en forma tabular, como en la tabla 20. El analista normalmente no conoce el rango de valores que se tendrán de utilidad, por lo que es necesario programar un esquema para establecer automáticamente los rangos de cada intervalo de la columna 1, tabla 20. Esto se puede tener haciendo que la computadora busque el valor negativo más grande de utilidad y el valor positivo mayor. Entonces todo lo que se requiere es dividir este rango en 15-20 intervalos iguales. De tal forma que un esquema automáticamente fijará los rangos de los intervalos para cada corrida.

Una vez que se establecen los intervalos, la computadora está entonces programada para contar cuántos valores caen en cada rango, calcula las frecuencias relativas, (como en la tabla 20), multiplica las frecuencias relativas por el punto medio de cada intervalo (como en la tabla 21), y suma los términos del producto para calcular el valor esperado. Todos estos datos tabulados son impresos automáticamente.

Dependiendo de la capacidad de la computadora y preferencias personales, hay quién hace su despliegue de la distribución del utilidad como una impresión gráfica, así como una impresión de la frecuencia acumulativa de la utilidad.

Todos estos procedimientos y formatos de salida son programados y no cambiarán de análisis a análisis.

Cuando se escribe un nuevo programa y se hace todo esto hay varias verificaciones que se pueden hacer a lo largo del camino para checar la integridad del programa. Se pueden programar la parte 1 y la 3 y entonces verificar si las distribuciones de entrada están siendo manejadas correctamente o probar valores de la variable e imprimirlos. Pueden entonces graficarse los valores probados para ver si se duplican las distribuciones que se leyeron en el inicio. Siguiendo ésta verificación se puede programar el procedimiento de salida (Parte 5).

Esto es entonces verificado por lecturas de distribuciones específicas, y los valores de cada variable, entonces teniendo los valores probados dados en forma tabular como frecuencias relativas y un valor principal. Si la distribución de capacidad es igual a la distribución de entrada se ha confirmado que la parte 5 del programa es correcta. Finalmente, las partes 2 y 4 son agregadas.

Para verificar esto lo mejor que se puede hacer es leer en un análisis de prospecto (tal como el del ejemplo numérico de la sección anterior) y tener la impresión de salida, cada 20 pasos, todos los cálculos internos. Esta impresión deberá incluir los números aleatorios usados para los pasos y los correspondientes valores probados de las variables aleatorias.

En esta forma se puede ir a través de los cálculos para ver si la computadora hizo lo que se esperaba que hiciera. De acuerdo a la experiencia, se tiene que, al hacer chequeos como estos usualmente se encuentran más errores en el esquema lógico que en el programa mismo. De cualquier manera, la revisión periódica de todos los cálculos internos de un paso sirve como una manera útil de verificar la caja negra!

IV.6 Defensa Filosófica de la Simulación

Ahora que se ha discutido toda la gama de detalles relativos a la simulación es importante analizar el concepto desde un punto de vista diferente para mantener una perspectiva de las cosas. Sin duda la razón primaria para el uso de la simulación es obtener la experiencia la cual resulta de expresar la incertidumbre como una distribución y rango de valores posibles de las incógnitas.

Los exploradores siempre han tenido incertidumbre en declaraciones tales como, "no se conoce exactamente el espesor que tiene el yacimiento, puede ser tan pequeño como 20 pies o tanto como 300 pies", e históricamente se resuelve el dilema usando un solo valor el cual puede ser un valor probable típico, o promedio, o de rango medio. Pero la simulación ofrece la oportunidad de incluir una entera declaración inicial (que el espesor deba tener un rango de 20 a 300 pies) en el análisis.

¿No se está forzado a hacer un juicio subjetivo? (¿arbitrario?) sobre un valor "promedio" representativo, y la información resultante presenta al que decide sólo un valor de probabilidad.

Otras ventajas importantes son las siguientes:

- La Simulación permite aplicar un juicio profesional a una variable una vez, tal que todos los parámetros desconocidos ocurrirán juntos.
- Los juicios profesionales sobre cada parámetro pueden hacerse por cualquiera que sea conocedor sobre el tema. Es un método de análisis de riesgo el cual combina eficazmente la experiencia de la compañía en análisis de exploración de prospectos.
- El método está conducido a un análisis de sensibilidad. Por variación de las distribuciones el analista puede aislar una o más variables críticas. Se puede entonces mostrar al que decide sus mejores juicios (en la forma de distribuciones) de cada variable. Esto ayuda a guiar al que decide a través de los datos y la información así se puede enfocar la atención en la consideración de las variables.
- El despliegue gráfico de las distribuciones puede proveer mucha más información al que decide sobre los niveles de ganancia los cuales son probables y su probabilidad correspondiente de ocurrencia.
- El concepto es completamente general y puede ser usado para analizar cualquier situación que involucre incertidumbre. No hay restricciones en el número de distribuciones que puede ser considerado o las formas de las distribuciones. Solamente una cantidad mínima de experiencia estadística es requerida al usar simulación.

Mientras sus ventajas son aparentes, la aplicación del concepto de exploración petrolera pronto o más tarde derivará dos preguntas importantes:

- 1.- ¿Puede la simulación ser usada donde no hay datos de estadística o de información disponible sobre qué basar una distribución?
- 2.- ¿Es más profesional expresar el juicio como un rango y distribución de valores posibles a mejorar o empeorar que un solo valor estimado más pobre?

Es muy importante que se planten estas preguntas antes de iniciar la simulación. Si se prepara un análisis de simulación pero este realmente no esta sometido a la consideración del director, no dudará en seguida al detectar su aprensión. Y si no se puede defender el mérito del

método, probablemente no se deberá de usar.

En cursos enseñados por todo el mundo en materia de análisis de riesgo en muchas ocasiones, se da a los estudiantes un problema de simulación en el cual se dan los datos siguientes de un campo cercano:

Area: 6,000 acres
Espesor medio = 50 pies. Rango es de 15 a 90 pies.
Porosidad: Varía de 10% a 30% Valor promedio: 18%
Saturación de agua: varía de 20% a 30%
Factor de recuperación primaria: 16% del aceite original
Factor de Volumen del aceite: 1.4
etc.

Invariablemente formularon el modelo de simulación relativo a las reservas como sigue:

"Primero se genera una distribución de reserva usando la ecuación del volumen de poro con parámetros de área de drenaje, espesor, porosidad, saturación de agua, factor de recuperación, y factor de volumen del aceite. Al hacer esto se prueba una distribución triangular de espesor (mínimo 15 pies, modo 50 pies, máximo 90 pies), una distribución triangular de porosidad, y una distribución uniforme de Sw. Se considerará el área, factor de recuperación y factores de volumen como constantes.

En este punto se interrumpe, y los siguientes cambios usualmente toman lugar:

Porque se consideró el área, factor de recuperación, etc. como constantes? Se sabe por cierto que el factor de recuperación para este prospecto no perforado será exactamente 16%?

No se sabe, pero no se tenía otra información acerca del factor de recuperación, así no se pudo describir una distribución a usar. Esto es el porque, el análisis se basa sobre un sólo valor de 16% del volumen original de aceite.

Esta lógica significa que la simulación está bien cuando se tienen datos de estadística, pero en la ausencia de datos esto no puede ser usado. Y esto, conduce a menudo a la conclusión que expresa que la simulación es un instrumento bueno de análisis en áreas fuertemente perforadas tal como el Oeste de Texas pero no se puede usar en áreas vírgenes de exploración.

Pero esta conclusión claramente pierde los puntos de simulación. ¡Esto es análogo al razonamiento circular mencionado antes en el cual al menos se supo acerca de una variable, la más precisa de los valores estimados! Un sólo punto estimado de las variables aleatorias puede significar certeza cuando no puede existir esta. En realidad el factor de recuperación de este prospecto no es conocido con certeza que sea del 16%. ¡De hecho, no se puede ni siquiera decir con seguridad que la estructura tenga aceite!

Es una natural (pero incorrecta) tendencia, dejar la presencia o ausencia de datos ya sea que se trate una variable como un parámetro conocido o como una variable aleatoria. Pero la decisión deberá estar basada en lo que se pueda pronosticar, o definir un valor exacto del parámetro antes de perforar el pozo. Si la respuesta es que no se puede, entonces se necesita tratar y describir la variable como un rango y distribución -ya sea que haya datos disponibles o no.

La respuesta a la primera pregunta tiene que ser un rotundo sí. Esto en áreas donde hay pocos datos y que la incertidumbre es grande. Y no se puede obtener mucho de una impresión de incertidumbre (variabilidad) si se basan los cálculos en solo valores promedio de cada parámetro.

Se ve una expresión de incertidumbre en la forma que una distribución es más débil que un solo valor.

Ejemplo: "El señor geólogo, dice que el espesor puede ser en cualquier parte de 20 a 200 pies, los costos de perforación de \$500,000 a \$800,000, etc...
Esto parece agrandar lo que no pudo hacer su mente - todo se debería decir con rangos".

El director hace un comentario acerca de lo que probablemente el señor geólogo había dicho que el espesor será de 85 pies, costos por pozo de \$627,495, etc. esto habría sido mucho más preciso, valoración confiada de experiencia profesional.

¿Qué piensa? Este punto es claramente una selección personal, y si se aprueba sólo lo expresado por el director no hay mucho provecho de lo ganado en la simulación. Esto es, se tomará lo visto que expresando un rango de valores acerca de un parámetro que no se podrá medir o evaluar antes de que el pozo sea perforado, esto es más honrado y realista en valoración de la incertidumbre.

Uno de los estudiantes del curso (un geólogo, a propósito) lo dijo bastante elocuente. Declaró que para un geólogo ignorar o rehusar considerar la variabilidad es ser completamente ingenuo acerca del mundo geológico. Desafió a los geólogos escépticos a ir a una estructura geológica y tomar medidas de ángulo de echado, espesor, etc, entonces moverse 100 pies y medir los mismos parámetros. El punto era que el sistema completo de geología es uno de variabilidad y expresar esta variabilidad es hacer un contorno más realista del sistema geológico. Esta pregunta es una que se tiene que resolver.

Se cree firmemente que expresar la incertidumbre como un rango es mucho más realista que decir que el espesor será exactamente de 85 pies. Pero esta opinión no es la única que tiene que prevalecer. Se tiene que formular una posición propia en el punto en cuestión. Y sería un buen consejo el considerar este punto antes de empezar a usar la simulación. Si psicológicamente no se tiene noción de que distribuciones son las mejores expresiones de un mejor juicio profesional, el trabajo tedioso involucrado tendrá un pequeño beneficio.

En el balance final, la simulación no es un cúralo todo. Quienes toman la posición de "Cuando se duda de la simulación" son probablemente aquellos con puntos de vista extremos como de aquellos que desesperan ante la venida de la simulación en el juego de la exploración el cual es ahora "Estrangulación por simulación". Esto obliga a hacer cuidadosas y explícitas consideraciones de todas las variables y facetas del prospecto. Esto puede proveer conocimiento con mucho más experiencia acerca de varios niveles de probabilidad de ocurrencia, que muchos otros conceptos de análisis. Pero no se puede esperar que la simulación diga cuál es la distribución de la reserva en una cuenca, o cuál es la probabilidad de encontrar aceite, o si se tiene que explorar en el norte o extremo sur de la cuenca. Esto no es un replazamiento para un experto profesional. Esto es un poderoso método nuevo de análisis el cual puede grandemente ensanchar las capacidades de un explorador para comunicar el valor del prospecto al que toma decisiones.

V. EJEMPLOS Y APLICACIONES

En un proyecto de análisis de perforación exploratoria hay muchas, muchas variaciones diferentes, o maneras de organizar y estructurar el análisis. Esto es porque las variables descritas como distribuciones pueden variar de un prospecto a otro, una dependencia parcial entre varias variables se puede aplicar a un prospecto pero no a otros, etc. En cada prospecto usualmente se requerirán cambios menores en la estructura del esquema de la lógica de la simulación. Pero a pesar del hecho de que hay muchas variaciones, los análisis incluyen mayores semejanzas.

Estos atributos comunes incluyen la previsión de explicar un éxito en un pozo exploratorio y las probabilidades de un fracaso, la necesidad de convertir reservas descubiertas a un efectivo contante equivalente que fluye VPN, procedimientos para averiguar el número de pozos de desarrollo a ser perforados, y una contabilidad para todo el costo de la inversión, gastos e impuestos.

Estas consideraciones se aplicarán si el prospecto que está considerado es costa-dentro o costa-fuera, un prospecto de gas o un prospecto de aceite, o si el prospecto está en América del Norte o en el Sudeste de Asia. Las leyes específicas de impuesto y regulaciones del gobierno pueden cambiar sin duda si las magnitudes de escala cambian, y sin duda las magnitudes de cambio de escala del prospecto. Pero analizando un prospecto involucran básicamente la misma serie de pasos por todas partes.

En este capítulo se considerará el armazón general de un análisis de simulación de un prospecto serán y discutidos varios dibujos esquemáticos para proveer una visión de conjunto, de como organizar el análisis de la simulación. Entonces se observará un ejemplo real numérico de un análisis de prospecto. Este ejemplo tiene que ser útil para enseñar como se formula el esquema de lógica del análisis. En seguida se harán algunas observaciones generales acerca de incorporar el parámetro de productividad inicial en el análisis. Este es un parámetro, el cuál se tiene que tratar con mucho cuidado a causa de su usual dependencia en otras dimensiones del análisis.

Finalmente, brevemente se cubrirán algunas aplicaciones de simulación y otras de análisis de factibilidad. Una de éstas es un modelo de simulación, el cuál puede estimar la distribución de las reservas totales en una área nueva de exploración. Las otras aplicaciones serán mencionadas en forma de una breve revisión de literatura.

La Organización General de un Análisis de Simulación

Los análisis de la mayoría de los prospectos de perforación pueden ser organizados en forma de un parcial árbol de decisiones, como se muestra en la figura 36. El árbol es parcial en el sentido de que no se incluyen las otras alternativas de inversión. El objetivo de analizar el prospecto es determinar el beneficio esperado VPN (o valor medio equivalente) a la oportunidad nodo A. Este valor esperado sería el parámetro de decisión que aceptará o rechazará el proyecto y/o compara su conveniencia relativa a otros disponibles proyectos de perforación.

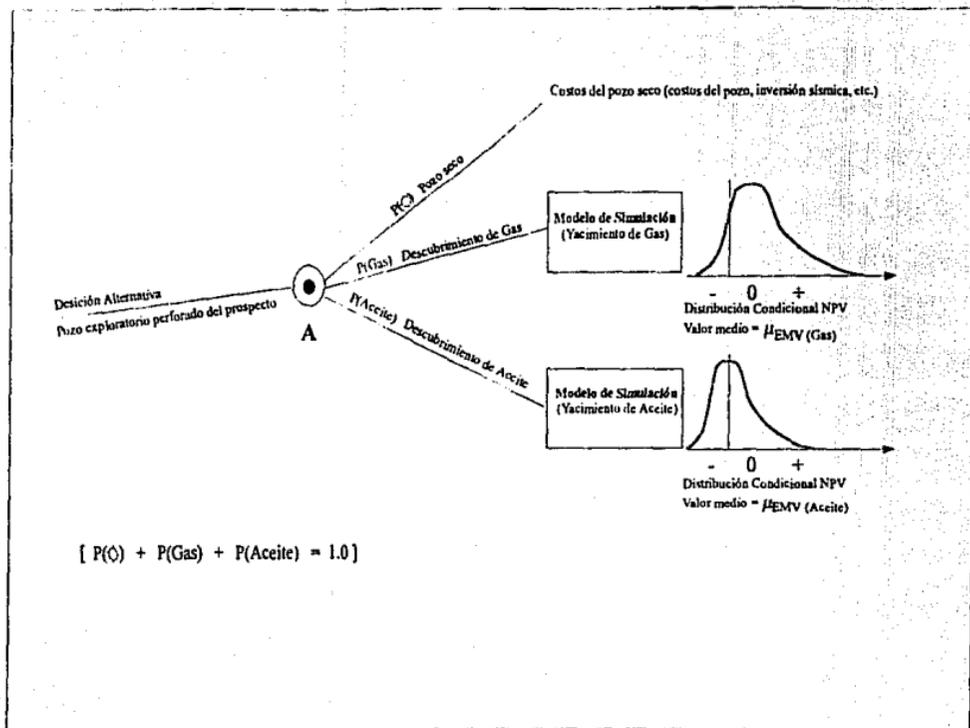


Figura 36. Esquema de una decisión parcial mostrando tres componentes básicos de un análisis de simulación de un prospecto de perforación exploratoria.

Para el resultado de un "pozo seco" el punto final incluye el costo del pozo seco más cualquier documentación a ser aplicada al proyecto para bonos de arrendamiento, estudio sísmico, etc. Si se hace un descubrimiento de gas o de aceite los valores del punto final no son solo valores, sino distribuciones de factibilidad. La razón de ésto es apropiada al considerar distribuciones de beneficio que al tiempo decidirán si se acepta el proyecto, no se sabe el tamaño exacto y valor que tendrá un descubrimiento de aceite o gas.

Esto es, si se encuentra aceite ¿serán 100 millones de barriles o 10 millones de barriles? ¿Requerirá el descubrimiento dos plataformas (costa-fuera) o sólo una? etc.. Al determinar estas distribuciones de utilidades, se podrá usar un modelo de simulación para el caso de aceite y gas. Los resultados de la simulación serán una distribución condicional de la utilidad, obteniendo un descubrimiento de gas o un descubrimiento de aceite.

El valor medio de estas respectivas distribuciones, representa el valor esperado, dado que era una certeza que el gas o el aceite serían encontrados. Pero como esto es obvio, no es el caso que se tenga que hacer un cálculo final en el cuál los valores medios condicionales de EMV_{gas} y EMV_{aceite} estén pesando por sus probabilidades respectivas de ocurrencia, $P(Gas)$ y $P(Aceite)$, en forma de un cálculo de la ecuación 14:

$$MV_A = P(\text{O}) \times \left(\begin{array}{l} \text{Costos de Pozos} \\ \text{secos y pago de} \\ \text{deudas} \end{array} \right) + P(Gas) \times \left(\begin{array}{l} \text{Cocondicional} \\ \text{valor medio} \\ \mu_{EMV(gas)} \end{array} \right) + P(Aceite) \times \left(\begin{array}{l} \text{Cocondicional} \\ \text{valor medio} \\ \mu_{EMV(aceite)} \end{array} \right) \quad (14)$$

En esta ecuación EMV_A es el valor esperado a oportunidad del nodo A (el objetivo), y se define como los otros términos en la figura 36.

El siguiente efecto de la formulación del análisis, como en la figura 36, es dividir el análisis del prospecto en dos partes: una parte que consiste de un análisis de simulación para determinar una distribución condicional del beneficio dado un descubrimiento, y la segunda parte combinar la probabilidad condicional esperada de valores con el descubrimiento y costo del pozo seco, por vía de la ecuación 14. Estas dos partes del método tienen ciertas ventajas computacionales.

Para eso se reduce a uno el número de parámetros, el cual tiene que ser probado con un número aleatorio en cada paso. Esto es porque el análisis de simulación sólo empieza una vez hecho un descubrimiento y no requiere que el analizador incluya un esquema de muestreo aleatorio para determinar cada paso, ya sea que en el pozo exploratorio se haya encontrado aceite o gas.

Otra ventaja es, que las dos partes tratan de aproximarse más a un análisis de sensibilidad de las probabilidades de un descubrimiento. La condición de valores medios de un descubrimiento dado es independiente de las probabilidades de un descubrimiento. Así, si se desea evaluar el efecto en EMV_A de las diferentes oportunidades de un descubrimiento, solamente se cambian los términos de $P(O)$, $P(Gas)$, y $P(Aceite)$ y se resuelve la ecuación 14. Si se incluye la probabilidad de un descubrimiento junto con la simulación probablemente se tendría que regresar al análisis de simulación para hacer una sensibilidad. Con este estudio se requiere

correr la simulación sólo una vez, con las probabilidades del descubrimiento, tratado por separado en la ecuación 14.

Se tiene que mencionar que el árbol parcial de la figura 36 no es una representación completamente clara de como se desean analizar algunas situaciones. Por ejemplo, se puede quizás desear incluir una opción al considerar la perforación de un segundo pozo exploratorio, si el primer pozo en el prospecto resulta seco. Esto significaría que otro nodo de decisión y oportunidad sería agregado a éste que ahora es el punto terminal del pozo seco de la figura 36.

También, se tiene que notar que si se está en un medio predominantemente de aceite, el término $P(\text{Gas})$ puede ser igualado a cero, haciendo los resultados en cualquiera de los dos, un descubrimiento de aceite o de un pozo seco. El yacimiento de gas de la figura 36 significa un descubrimiento de gas no-asociado. El gas asociado producido en conjunto con el aceite debería ser tomado en cuenta junto con el modelo de simulación del yacimiento.

Con esta descripción breve de como las piezas de todo el conjunto se unen se habla un poco más acerca de las características generales del "modelo de simulación" mostrado en la figura 36. El modelo consiste de todas las variables y de los pasos de computación que son usados en la determinación de la distribución condicional de valor de un descubrimiento de aceite o de gas. Estos pasos computacionales esencialmente consisten de las tres secciones principales mostradas en la figura 37. Los términos y pasos dados en la figura 37 tienen que ser razonablemente claros para alguien que jamás ha calculado el valor de un descubrimiento.

En la parte baja de cada sección están listados algunos de los factores que probablemente se tendrían que considerar en los cálculos. Las listas no son necesariamente completas, y no necesariamente implican que todos los factores tengan que ser usados como distribuciones en el análisis. Allí probablemente también estaría un número de factores que deberán ser considerados en más de una sección. Por ejemplo, se puede muestrear un valor de espesor productor en la primera sección como uno de los parámetros para calcular reservas. Este mismo valor de espesor productor neto puede ser retenido como un punto de entrada en una gráfica de dependencia parcial al probar un valor de IP por pozo en la segunda sección.

Como se menciona al principio de ésta sección, hay muchas variaciones posibles en el análisis de los prospectos de exploración, pero usualmente tendrán las características generales esbozadas en la figura 37. Un diagrama tal como el que aparece en la figura 37, obviamente está simplificado, pero generalmente este deberá proveer por lo menos el armazón sobre el cual se construya un análisis específico.

Para interponer un pequeño "mundo-real" de acuerdo al significado de las figuras 36 y 37, se observará ahora un ejemplo numérico del análisis de un prospecto de perforación exploratoria usando simulación.

Ejemplo:

En esta sección se observará un análisis de simulación que fue hecho para determinar si se ofrecía \$1 millón de bonos por un bloque de 20,000 acres. De algunas correlaciones sísmicas y geológicas primarias el personal de exploración había hallado una anomalía estructural grande bajo el bloque. El intervalo productivo esperado fue la arena A. Cerca de la arena A se tuvo producción a 50 millas al norte en la misma cuenca. El personal estimó las probabilidades para éste prospecto llamado, el prospecto de estudio, el cuál tendría aceite en un rango de 0.10 a

0.20. Con el análisis se ensaya para determinar si se podría justificar \$1 millón de bonos, y así, la probabilidad mínima del descubrimiento que se deberá requerir para obtener por lo menos un 16% de retorno esperado (por ejemplo un VPN esperado @ $i_0 = 0.16 >= + \$1 \text{ MM}$).

Primero, algunos comentarios generales acerca del análisis y como este se organizó:

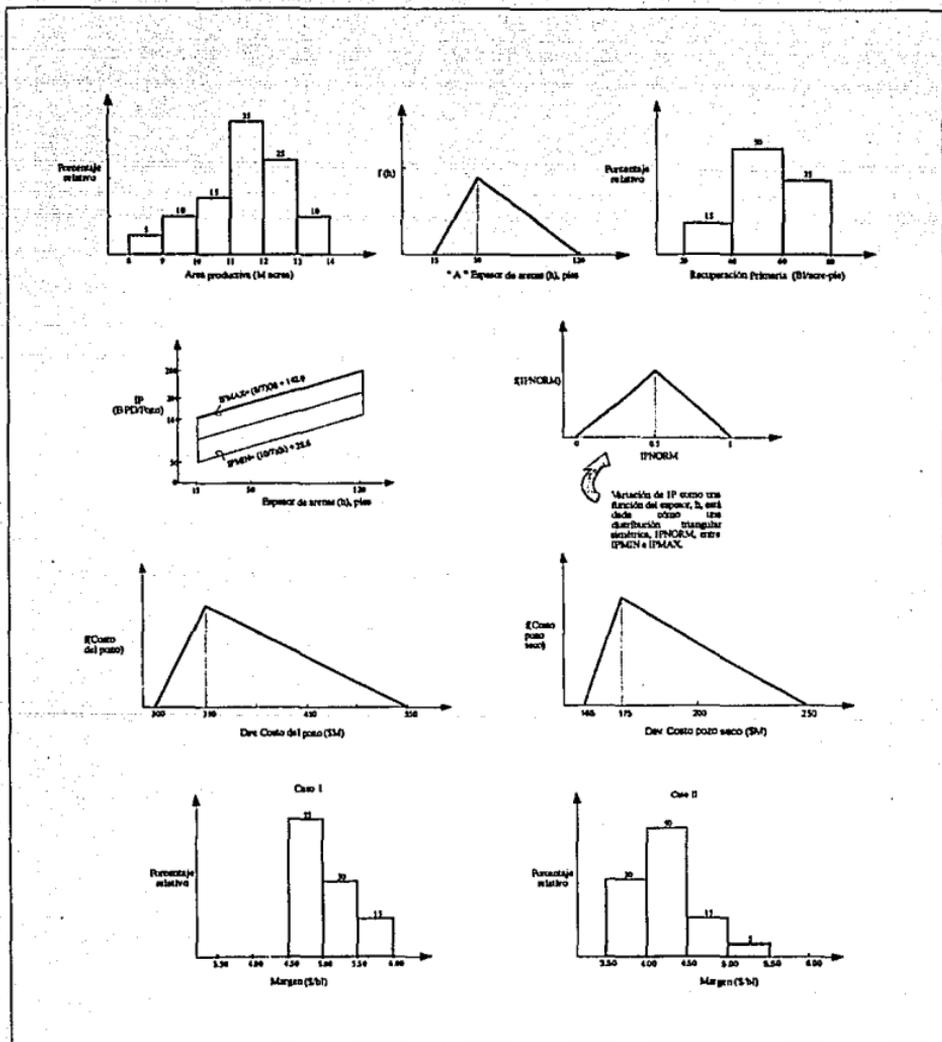
1. Se consideraron dos pozos exploratorios secos como máximo que se debían perforar sobre la anomalía. Estos pozos tenían un costo esperado de \$200 M cada uno.
2. Se definieron los siguientes parámetros como variables aleatorias descritas como distribuciones de probabilidad:
 - a. Área productiva - acres.
 - b. Espesor neto productor de la Arena "A", h-pies.
 - c. Recuperación primaria - barriles/acre-pie).
 - d. Potencial Inicial, IP, por pozo - BPD/pozo.
 - e. Costos de perforación de pozos de desarrollo - \$/pozo
 - f. Costos de perforación de pozos de desarrollo secos - \$/pozo
 - g. Margen - \$/ barril
(se definió el Margen como el precio bruto de venta por barril para aceite y gas asociado menos impuestos y gastos operativos.)

De análisis de datos del norte del campo, se determinó que una dependencia parcial definitiva existió entre el potencial (IP) y el espesor (h). De una gráfica de h contra IP por pozo se pudo establecer un límite alrededor de la región de valores posibles. Por lo tanto, se observó que la variabilidad de IP como una función de h, aproximadamente se representaba por una distribución simétrica triangular. Estas relaciones parciales dependientes, junto con las otras distribuciones usadas en el análisis se muestran en la figura 38.

3. El método para establecer el número de pozos de desarrollo requerido fue probar un valor de IP por pozo, calcular la recuperación por pozo sobre 20 años la declinación exponencial, y dividir la recuperación por pozo entre el total de las reservas del campo. Esto tenía el efecto de conteo para la variabilidad de productividad con un IP/ pozo, entonces perforar suficientes pozos hasta agotar el campo en 20 años. Se supuso que los pozos estarían en declinación desde el comienzo de la producción, y serían abandonados en 20 años a un gasto de 10 BPD por pozo.

En un momento se discutirán los cálculos reales de los pasos de simulación y las ecuaciones del paso 5, 6 y 13 relativas a la declinación exponencial.

4. Teniendo determinado el número de pozos de desarrollo a perforar, el número de pozos secos de desarrollo probable a ser perforados alrededor del límite productivo del campo se determinó la correlación de la figura 39.
5. Se estimaron que 16 pozos por año pueden ser perforados si se descubre aceite en el proyecto. El tiempo requerido para terminar la perforación fue determinado dividiendo el número total de pozos del campo (terminaciones y pozos secos de desarrollo) entre 16 pozos por año. Para propósitos de simplificación, los descuentos de los costos de perforación, asumieron que el total de los desembolsos por perforación ocurriera al tiempo en que la mitad de los pozos hayan sido perforados.



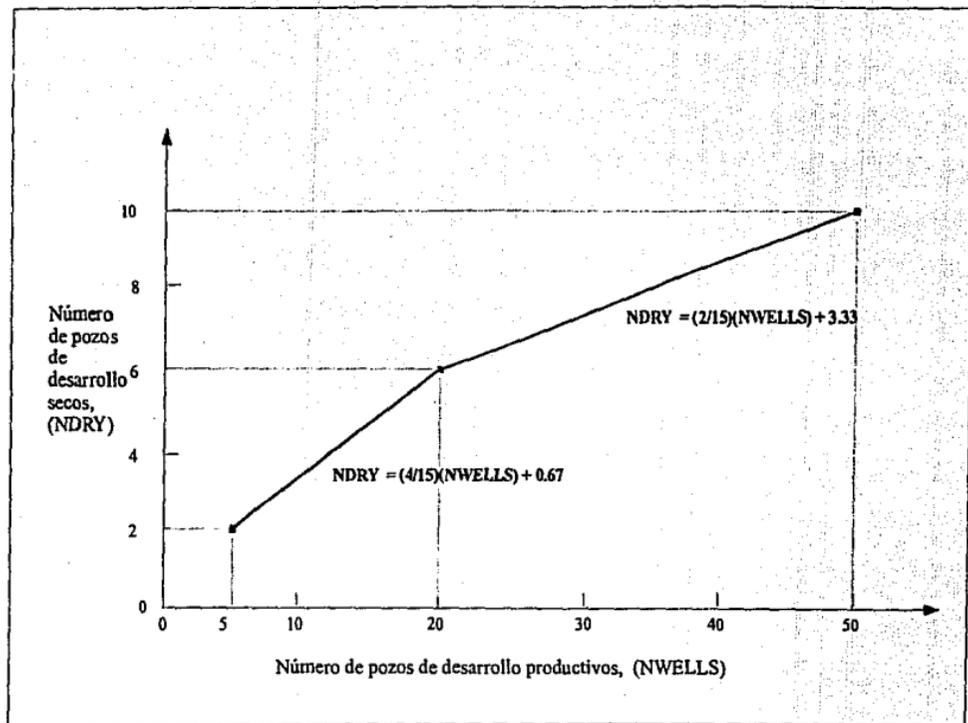


Figura. 39 Correlación para determinar el número de pozos de desarrollo secos.

Se supuso también que este punto fuera cuando comenzara la producción, para propósitos de encontrar VPN ganancias. (Como comentario: esta suposición de simplificación fue hecha para hacer este pequeño ejemplo más fácil de seguir. En la práctica actual la perforación extendida sobre un período substancial de tiempo, deberá basarse en los tiempos reales cuando se terminen los pozos y comiencen a producir.)

6. Los cálculos de VPN se basaron en descuentos continuos usando una tasa del 16% por año; que es, $i_0 = 0.16$.
7. El método de análisis de simulación fue usado para desarrollar y describir una distribución condicional de beneficio de VPN, dado un descubrimiento. El valor medio de ésta distribución condicional de beneficio, $\mu_{VPN|aceite}$, fue comparado con la probabilidad de encontrar aceite en el primer lugar, $P(\text{ACEITE})$ esperar el valor de el resultado de un descubrimiento. El suponer el valor producido esperado de la salida de la estructura seca fue solamente la pérdida de dos pozos secos, - \$0.4 MM, multiplicado por $P(\text{POZO SECO})$. El árbol parcial de decisiones (del tipo de la figura 36) describe todo esto la figura 40.
8. Se consideró la estimación de un margen bastante especulativo, debido a las incertidumbres de las tendencias futuras del precio del crudo de aceite mundial y las condiciones económicas. Como un resultado, se corrieron dos casos - uno usando el Caso I de margen de distribución, dado en la esquina inferior izquierda de la figura 38 y otro usando el Caso II de distribución de esquina inferior derecha de la figura 38. El Caso I consideró la situación optimista y el Caso II fue considerado el caso más pesimista. Cuando el caso II fue corrido la distribución de todas las otras variables fue mantenida como en el caso I. Consecuentemente una comparación de resultados de las dos simulaciones debió mostrar los efectos de las diferentes suposiciones del margen por barril.

Esquema Lógico del Análisis de Simulación

La serie de pasos que representa una simulación es la siguiente:

1. Probar el área productiva, arena de espesor "A", y la distribución de la recuperación primaria (barriles por acre-pie) y calcular las reservas totales del campo:

$$\text{TOTAL RESERVES} = (\text{Area}) (\text{Espesor}) [\text{Bbls/acre-pie}]$$

2. Usando el valor probado del espesor neto de arena, (h), calcular el valor mínimo y máximo de la distribución del potencial inicial:

$$\text{IPMIN} = (10/7) (h) + 28.6$$

$$\text{IPMAX} = (8/7) (h) + 142.9$$

3. Probar un valor de IPNORM de la distribución normalizada de IP, teniendo como perfil una distribución simétrica triangular. Esto es, mostrado en la figura 41.

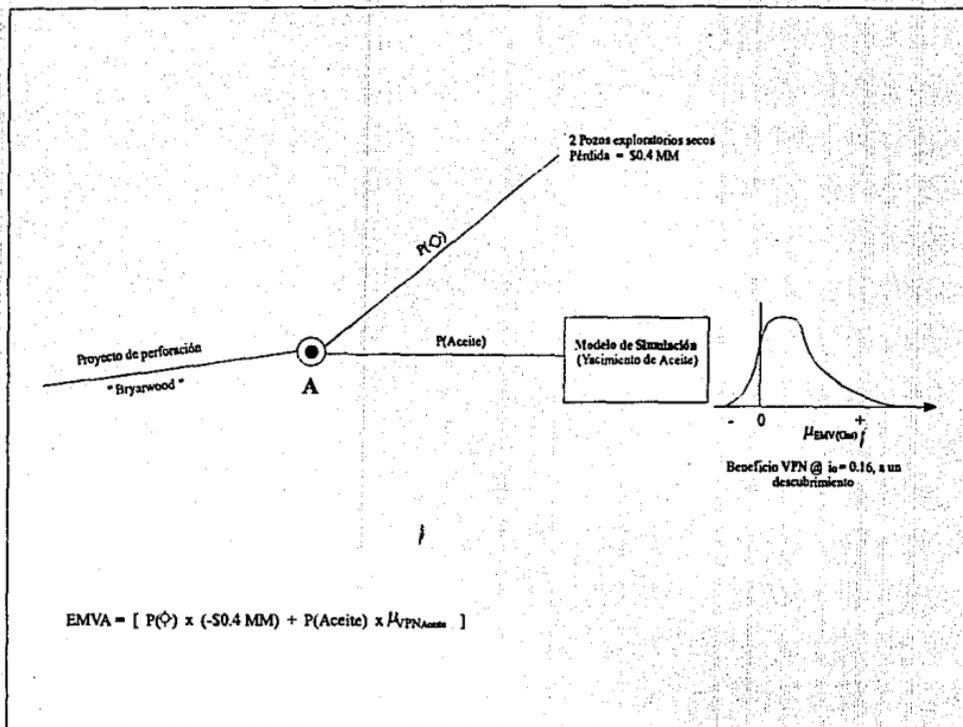


Figura. 40 Arbol de decisión parcial dado un descubrimiento.

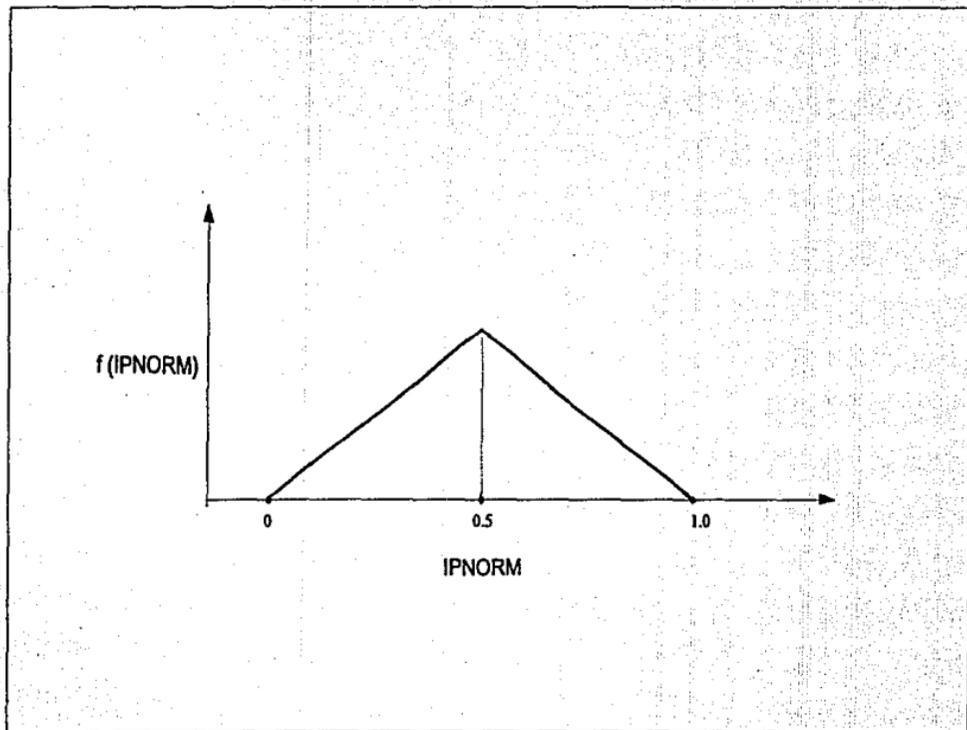


Figura. 41 Ejemplo de una distribución triangular simétrica.

4. Calcular el IP por pozo:

$$IP \text{ [bpd/pozo]} = IPMIN + (IPMAX - IPMIN) (IPNORMAL)$$

(Nota: Los pasos 2,3 y 4 involucran los procedimientos para probar un valor de IP/pozo dada una parcial dependencia como la variable aleatoria espesor. Estos son los pasos descritos en general en la tabla 16 y específicamente en la tabla 17 para el caso donde las variables X y Y son espesor e IP/pozo respectivamente.)

5. Calcular el ritmo de declinación exponencial:

$$a = \ln[q_i / q_f] / t$$

donde:

q_i = ritmo de declinación inicial, bbls/año

= IP x 365.

q_f = gasto de abandono, bbls/año,

=(10 bpd x 365).

t = tiempo de declinación

= 20 años

6. Calcular la reserva recuperable por pozo:

$$DELN = \Delta N = (1 / a) [q_i - q_f]$$

$$DELN = \Delta N = (1 / a) [(IP \times 365) - (10 \times 365)]$$

7. Calcular el número de pozos de desarrollo del campo, NPOZOS

$$NPOZOS = \text{TOTAL RESERVES} / DELN \text{ (Redondeado al valor entero más cercano)}$$

8. Determinar el correspondiente número de pozos de desarrollo secos, NSECO:

$$\text{SI } NPOZOS > 20 \text{ entonces } NSECO = (2/15) (NPOZOS) + 3.33$$

$$\text{SI } NPOZOS \leq 20 \text{ entonces } NSECO = (4/15) (NPOZOS) + 0.67$$

(Rodondeado al valor entero más cercano.)

9. Probar distribuciones de costos de perforación.

10. Calcular los costos totales de perforación requeridos, DCOSTOS:

$$DCOSTOS = (NPOZOS) \text{ (Valor probado de Desv. del costo por perforación de pozos.)} \\ + (NSECO) \text{ (Valor probado de Desv. de costos de pozos secos.)}$$

11. Calcular el tiempo requerido para terminar la perforación de todos los pozos del campo, basado en 16 terminaciones por año:

$$TIEMPO = (NPOZOS + NSECO) / 16 \text{ (El Tiempo tendrá unidades de años.)}$$

12. Descuentos de costos de perforación al tiempo cero a $i_0 = 0.16$ asumiendo que el total de los desembolsos ocurren cuando la mitad de los pozos del total del campo han sido perforados:

$$VPNDCOSTOS = (DCOSTO) \times (e^{-0.16 \times \text{tiempo} / 2})$$

13. Probar el valor de margen y calcular el valor de descuento de ganancia así como del principio de la producción, REV:

$$REV = (\text{RESERVAS TOTALES}) (\text{MARGEN}) [a / (a+j) \times (1e^{-t} \times (a+j)) / (1-e^{-(a+j)t})]$$

donde:

a = ritmo de declinación exponencial (paso 5)

j = tasa de descuento, 0.16

t = años de declinación exponencial, 20 años

14. Descuento REV Al tiempo cero, asumiendo que la producción comienza a TIEMPO / 2:

$$VPNREV = (REV) \times (e^{-0.16 \times \text{tiempo} / 2})$$

15. Calcular la utilidad VPN:

$$\text{UTILIDAD VPN} = \text{VPNREV} - \text{VPNDCOST}$$

16. Se termina el paso, se guarda el valor de UTILIDAD VPN y se regresa al paso 1.

Adelante se ilustra este esquema lógico, se hacen los cálculos numéricos de un paso hipotético. Se debe recordar de nuevo que el análisis de simulación es sólo usado para describir una distribución de beneficio dado un descubrimiento. Como resultado de los costos de un pozo exploratorio seco y las probabilidades de que no entre en el análisis de simulación.

1. Valores empleados de área, espesor, y recuperación primaria son:

área = 11,840 acres

h, espesor = 48 pies

recuperación = 50 barriles/acre-pie

$$\text{RESERVAS TOTALES} = (11,840 \text{ acres}) (48 \text{ pies}) (50 \text{ barriles/acre-pie}) = 28,416,000 \text{ bls.}$$

2. Para un valor de espesor, h, de 48 pies calcular IPMIN e IPMAX:

$$IPMIN = (10/7) (48) + 28.6 = 97 \text{ Bpd / Pozo}$$

$$IPMAX = (8/9) (48) + 142.9 = 198 \text{ Bpd / Pozo}$$

3. Valor empleado de IPNORM: 0.51

4. Calcular IP por pozo:

$$IP = IPMIN + (IPMAX - IPMIN) (IPNORM)$$

$$= 97 + (198 - 97) (0.51) = 148.5 \text{ BPD/pozo}$$

(Para este paso se tiene que descubrir un campo con 28.4 millones de barriles teniendo pozos con un promedio de productividad inicial de 148 BPD. Los pasos 2, 3 y 4 obtienen un valor de IP para la distribución condicional de IP dado que el espesor fue 48 pies. La distribución condicional específica de IP que probó una distribución triangular simétrica tiene como valores mínimo y máximo a 97 y 198 BPD respectivamente.)

5. Se calcula el flujo de declinación exponencial para IP = 148.5 BPD:

$$a = \ln[(148.5 \times 365) / (10 \times 365)] / (20) = 0.135$$

6. Calcule las reservas recuperables por pozo, DELN

$$\text{DELN} = (1/0.135) [(1 - 8.5 \times 365) - (10 \times 365)] = 375,000 \text{ Bbls/Pozo}$$

7. Calcule el número de pozos, NPOZOS

$$\text{NPOZOS} = \text{TOTAL RESERVES} / \text{DELN} = 28,416,000 \text{ bbls} / 375,000 \text{ bbls/Pozo} = 76 \text{ Pozos}$$

(Una vez determinado el promedio de productividad por pozo en el paso 4 el programa calcula el siguiente promedio de recuperación por pozo, dado el IP a 20 años de declinación exponencial, y a 10 BPD de gasto de abandono. Para este paso se tiene 375 000 bbls/pozo. Conociendo el total de reservas del campo y el promedio de recuperación por pozo en el paso 7 se calcula el número de terminaciones necesarias para agotar al campo en 20 años.)

8. Determine el número de pozos secos de desarrollo, dado que NPOZOS = 76:

$$\text{NSECO} = (2/15) (76) + 3.33 = 13 \text{ dev. pozos secos.}$$

9. Valor Probado de costos de pozos (para este paso):

Costos de Terminación del pozo: \$355,000 cada una
Desv. Costos de pozos secos: \$181,000

10. Calcular los costos totales de perforación:

$$\text{DCOSTOS} = (76) (\$355,000) + (13) (\$181,000) = \$29,333,000$$

11.- Calcular el tiempo requerido para terminar de perforar el campo:

$$\text{TIEMPO} = (76+13) / (16) = 5.56 \text{ años}$$

12. Calcular los costos VPN de perforación:

$$\text{VPND COST} = (\$29,333,000) (e^{-0.16 \times 5.56/2}) = \$18,797,000$$

13. Valor Probado de margen (caso 1): \$4.80/bbl

Calcular el valor discontinuo de ganancia como comienzo de la producción:

$$\begin{aligned} \text{REV} &= (28,416,000 \text{ bbls}) (\$4.80/\text{bbls}) [0.135 / (0.135 + 0.16)] * [(1 - e^{-20(0.135+0.16)}) / -e^{-(20 \times 0.135)}] \\ &= \$66,713,000 \end{aligned}$$

14.- Descuento REV, al tiempo cero desde TIEMPO/2:

$$VPNREV = (\$66,713,000) (e^{-0.16 \times 5.56 / 2}) = \$42,751,000$$

15.- Calcular UTILIDAD VPN:

$$UTILIDAD VPN = \$42,751,000 - \$18,797,000 = \$23,954,000$$

16.- Fin del paso.

En éste caso hipotético, se descubrió un campo que tiene 28.4 millones de barriles de reservas primarias, tomando 76 pozos para producir el campo en 20 años, se requiere perforación cerca de 5 1/2 años, y el descubrimiento costó \$23.9 millones precio neto al tiempo cero usando una tasa de descuento de 16% por año.

Resultados del Análisis de Simulación

El modelo de simulación puede correrse usando 500 pasos. Los resultados de las distribuciones de beneficio condicional VPN de un caso ejemplo presentado en la literatura, se muestran en la tabla 22:

Tabla 22.

Resultados de las distribuciones del VPN a partir de un modelo de simulación.

	CASO I (Rangos de margen de \$4.50 a \$6.00 por barril)	CASO II (Rangos de margen de \$3.50 a \$5.50 por barril)
VPN MINIMO	+\$2.414 MM	-\$2.084 MM (perdida)
VPN MAXIMO	+\$ 74.880 MM	+ \$ 63.562 MM
VALORES Modo	+\$ 24.154 MM a + \$ 27.777 MM	+ \$ 24.174 MM a + \$ 27.456 MM
VALOR Medio ($\mu_{VPN_{aceite}}$)	+\$ 30.063 MM	+ \$ 21.541 MM

Las gráficas acumulativas de frecuencia de cada VPN (distribución de beneficio) están dadas en la figura 42. De esta gráfica quien toma decisiones puede asociar las probabilidades de lograr varios niveles de utilidad, dado un descubrimiento. Los cálculos finales de EMV, asumiendo una probabilidad de descubrimiento de 0.10, son los siguientes:

$$CASO I P(\diamond) = 0.9 P(ACEITE) = 0.1 \mu_{VPN_{aceite}} = +\$30.06 \text{ millones}$$

$$\text{Pérdidas por pozo exploratorio seco} = -\$0.4 \text{ millones.}$$

$$EMV_A = [(0.9)(-\$0.4 \text{ MM}) + (0.1)(+\$30.06 \text{ MM})] = +\$2.646 \text{ MM}$$

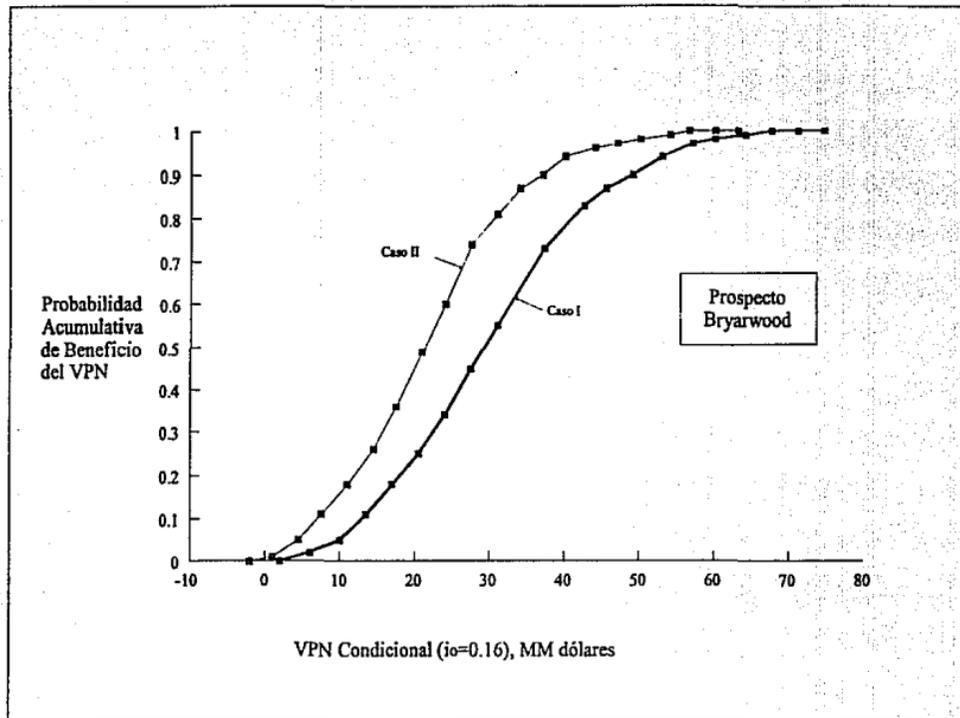


Figura 42. Gráfica de Frecuencia Acumulativa para el ejemplo de análisis de simulación. El Caso I es para la distribución optimista y el Caso II está basado en la menor y más pesimista distribución.

CASO II $P(\diamond) = 0.9 P(\text{ACEITE}) = 0.1 \mu \text{VPN}_{\text{Aceite}} = +\21.54 millones

Pérdidas por pozo exploratorio seco = - \$0.4 millones.

$EMV_A = [(0.9)(- \$0.4 \text{ MM}) + (0.1)(+ \$21.54 \text{ MM})] = + \$1.794 \text{ MM}$

Así en respuesta a la primer pregunta, es factible gastar \$1 millón para un bono de venta y tener todavía un retorno esperado de por lo menos 16%. Esto es porque el EMV para ambos casos (a una probabilidad de descubrimiento de 0.10) es mayor que un VPN esperado de cero (correspondiente a una tasa de retorno de 16%) más \$1 millón por los bonos. Esto es, EMV es $>$ $VPN = + \$1.00 \text{ MM}$.

Para la respuesta a la segunda pregunta acerca de la probabilidad mínima de que un descubrimiento requiera tener una tasa de retorno de por lo menos 16%, se construyó una gráfica de EMV contra la probabilidad de encontrar hidrocarburos, figura 43. Los puntos graficados usados para construir la gráfica, fueron los valores de EMV calculados antes para una probabilidad de descubrimiento de 0.10 y de - \$0.4 millones de pérdida cuando la probabilidad de descubrimiento es cero.

Los valores sobre la ordenada de EMV son exclusivos de los bonos de renta. De esta gráfica se puede determinar que la probabilidad mínima del descubrimiento requerido para obtener una tasa de retorno esperado del 16%, si se paga \$1 millón de bonos se está entre 0.06 y 0.07 para este caso II, y entre 0.04 y 0.05 para el caso I (Nota: estas probabilidades son los valores de la abscisa correspondientes a un EMV de + \$1.0 millón). Desde que el cálculo personal de probabilidad de descubrimiento esta en el rango de 0.10 a 0.20 la conclusión que, en realidad, justifica el desembolso de \$1 millón es segura la perforación y procede a una prueba de perforación exploratoria en el Prospecto de estudio.

Este ejemplo numérico es útil para mostrar qué se parece a un análisis real de simulación, y como la distribución de los resultados de simulación deriva en beneficio, se calcula el valor esperado para varias estrategias de decisión. Y se menciona otra vez la importancia de una presentación grafica tal como la de la figura 43 que despliega la sensibilidad de la probabilidad de descubrimiento en EMV.

Estimar la probabilidad de descubrimiento es un proceso nebuloso, y de gráficas como la figura 43 y se debe ser capaz de formular estrategias positivas de decisión, aunque no se pueda saber el valor exacto de la probabilidad de descubrimiento al tiempo en el que se evalúan las opciones de decisión.

Tratando Productividad Inicial y determinando desarrollo de Programas en el Modelo de Simulación.

La productividad inicial de los pozos o campos es una variable aleatoria que tiene un efecto evidente en el VPN de rentas, y puede ser un factor importante en establecer el número de pozos de desarrollo a ser perforados en el campo. También es una variable que rara vez se puede pronosticar con certeza antes de perforar los pozos. Como un resultado deseado normalmente al describir la variabilidad (incertidumbre) asociada con la productividad como una distribución de posibles valores.

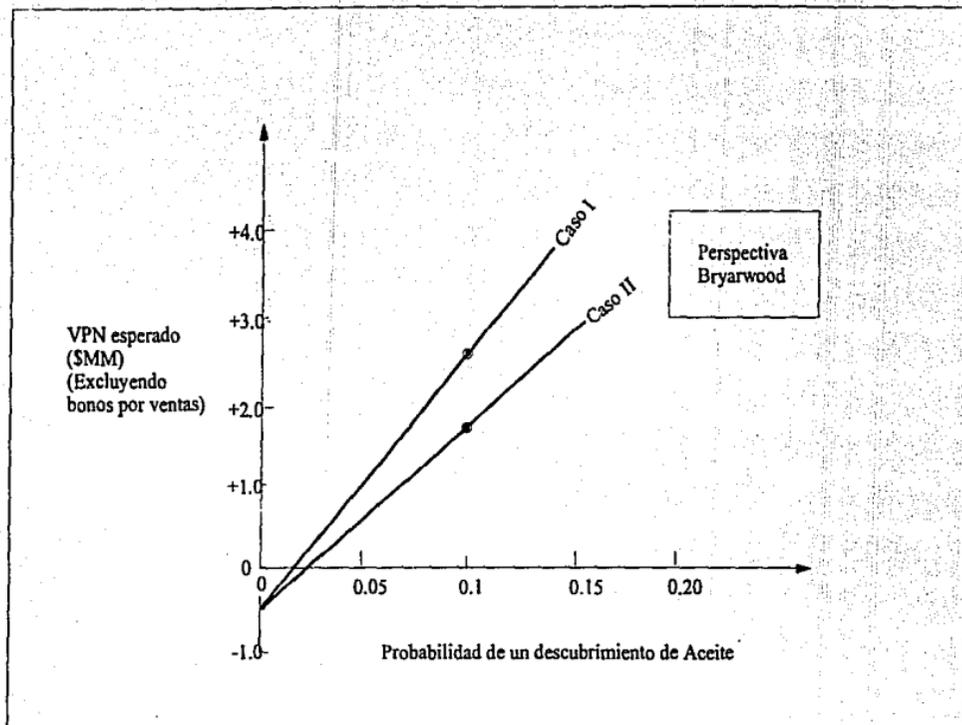


Figura 43. Gráfica de VPN esperado (excluyendo bonos por ventas) contra la probabilidad de un descubrimiento de aceite para cada caso del análisis del Prospecto Bryarwood.

Pero aquí es en donde el problema se hace un poco complicado. Como se ha mencionado en varias ocasiones anteriores, en este capítulo la productividad es usualmente dependiente de otros parámetros o dimensiones en el análisis, tal como el espesor neto productivo (por la ley de Darcy), límites físicos para producir a ritmos debido al pozo y/o superficie para levantar el equipo, el número de pozos en el campo, y posiblemente limitaciones aún de capacidad de almacenamiento debido al retraso en la llegada a los tanques.

Todo esto es para decir que los primeros dos pasos de la sección 2 de la figura 37 no son a veces sencillos de manejar, como puede parecer al iniciar la revisión. El problema es poder desarrollar un programa de producción y determinar el número de pozos, de manera que se requiera reconocer todas las correlaciones dependientes y dimensiones del prospecto en cada paso.

Hay, por supuesto, un número de maneras de llevar a cabo esto. Se registrarán varios enfoques posibles para dar una idea de algunas de las formas para manejar IP así como los programas de pozos de desarrollo. Pero la elección variará de prospecto a prospecto, por lo que se requiere ser muy cuidadoso y explícito al estructurar esta parte del análisis.

1. Si dispone de datos para preparar una gráfica de IP/pozo contra el espesor neto productivo, se pueden seguir los pasos usados en el ejemplo descrito anteriormente. El valor de la productividad neta probado para el cálculo de la reserva, es usado como el punto de entrada en la gráfica de dependencia parcial mostrando el IP/pozo. Conociendo el IP/pozo y los parámetros del gasto de abandono, vida productiva, y el patrón de declinación, se puede calcular las reservas recuperables por pozo. Este parámetro entonces es dividido entre las reservas para obtener el número de pozos productores que serán perforados para este paso.

2. Conociendo la reserva recuperable del campo (de la sección 1) se puede calcular el gasto de producción inicial de este que deberá requerirse para agotarlo en 15 ó 20 años a un límite económico específico. Entonces se puede probar por separado un valor de IP/pozo de una gráfica de IP/pozo contra el espesor. Finalmente, dividiendo el IP total del campo entre el valor mostrado de IP/ pozo se obtiene el número de pozos a ser perforados.

3. Si un modelo de espaciamento se aplica (como 80 acres por pozo, 160 acres por pozo, etc.) se puede dividir el espaciamento por pozo entre el valor del área productiva total del campo usada en el cálculo de la reserva para dar el número de pozos. Se puede entonces dividir la reserva total del campo (de la sección 1) entre el número de pozos para obtener la reserva recuperable promedio por pozo.

El paso siguiente sería usar una correlación de IP/pozo contra reservas por pozo para probar un IP dentro de un rango factible, dadas las reservas por pozo, para poder estimar el programa de producción. Esta correlación puede requerir que se hagan algunos cálculos preliminares para establecer las dependencias parciales, y en la discusión anterior sobre dependencias parciales se mencionó que se tendría que decir como establecer los límites de una gráfica X-Y sin tener acceso a un grupo de datos estadísticos.

Así es que se desglosará y se mostrará una forma de como hacer esto. Se supone que se puede especificar que el rango posible de potencial inicial por pozo es de, 50 BPD/pozo a 400 BPD/ pozo. Y suponiendo que el rango de vida productiva de los pozos es de 8 a 24 años y que los pozos producirán a una capacidad sobre un patrón de declinación exponencial a un

límite económico de 5 BPD. Lo que se ha especificado con estos enunciados es la variabilidad del gasto inicial, vida productiva, patrón de declinación y un límite económico.

Estos factores son suficientes para calcular las reservas recuperables totales por pozo, y por una serie de cálculos de ingeniería se puede establecer la correlación entre IP/pozo y reservas por pozo. Los cálculos son directos y consisten de repetitivas soluciones de las ecuaciones que relacionan reservas a IP, vida productiva, y gasto de abandono. Por ejemplo, se puede primero calcular las reservas producidas por un pozo que comienza a un gasto máximo de 400 BPD y se agota el mismo en un tiempo mínimo de 8 años, el programa está dado en la curva (1) en la figura 44. Entonces la vida puede aumentar a 10 años y calcular las reservas resultantes.

Cada cálculo sucesivo sería para una vida productiva más larga, hasta el punto máximo de alcanzar 24 años. Para cada programa posible de producción se habría calculado un valor de reservas recuperables. Entonces se puede disminuir el gasto inicial es decir, 350 BPD, dar un tiempo fijo de 8 años y repetir los cálculos para un rango de vida productiva de 8-24 años. Entonces el gasto sería disminuido a 300 BPD y el proceso se repite, a 250 BPD, 200 BPD,.... y finalmente 50 BPD.

Los resultados obtenidos son finalmente mostrados en una gráfica de valores de IP/pozo usados en los cálculos repetitivos, contra las recuperaciones por pozo que fueron calculadas para varios valores de vida productiva. La gráfica aparecería como en la figura 45.

Esta correlación establece los límites junto con los cuales el IP/pozo puede variar entre valores dados de reservas por pozo y vida productiva, las reservas totales para el paso calculado en la sección 1, y las recuperaciones promedio por pozo calculadas previamente a dividir el valor de área probada entre el patrón de espaciamiento. El método de cálculos preliminares descrito no ayuda a definir la variabilidad de IP/pozo dentro de los límites, pero provee un medio para establecer los límites de la dependencia parcial. Así, esto sugiere una manera alternativa de definir la gráfica de dependencia parcial si no se tiene un grupo de puntos X-Y disponibles.

Para recapitular, esta alternativa de manejar el número de pozos y potencial inicial es como sigue:

- a. Probar área, espesor, distribuciones bls/acre-pie y cálculos de reservas totales de campo.
- b. Dividir el valor probado de área entre el patrón de espaciamiento (acres entre acres por pozo) determinar el número requerido de pozos de desarrollo.
- c. Dividir las reservas totales de campo entre el número de pozos de desarrollo para obtener las reservas por pozo.
- d. Usando los valores calculados de reservas por pozo como punto de entrada en la gráfica de IP/pozo contra reservas por pozo se obtiene un valor de IP/pozo dentro de la región limitada. Este IP entonces es usado para definir el programa completo de producción por años.

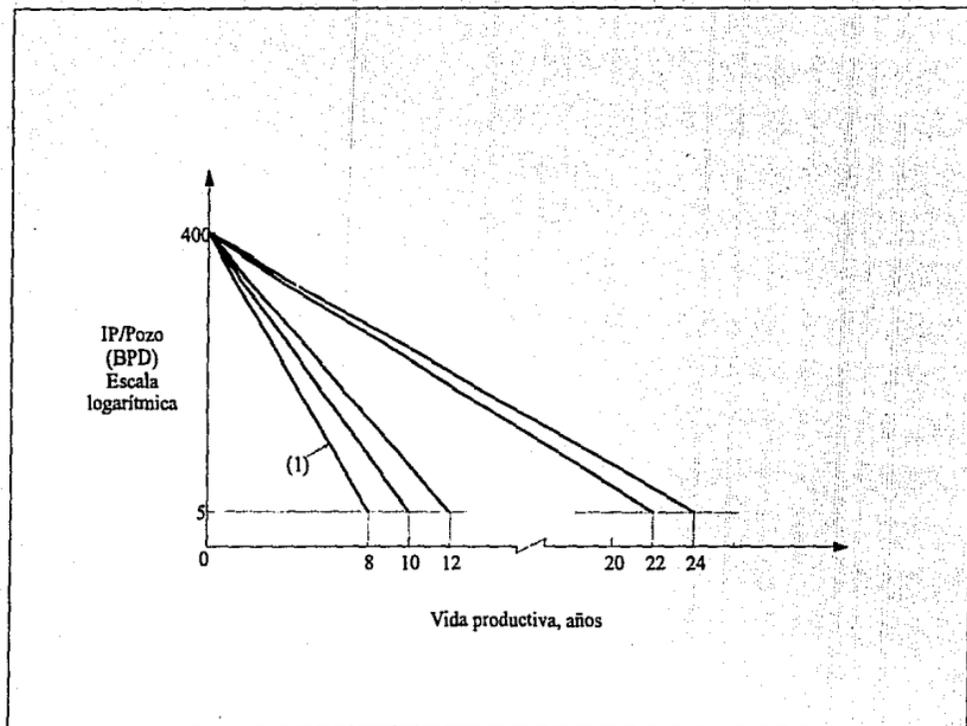


Figura 44. Curvas de declinación exponencial calculadas preliminarmente para determinar la correlación de IP/Pozo contra reservas por pozo.

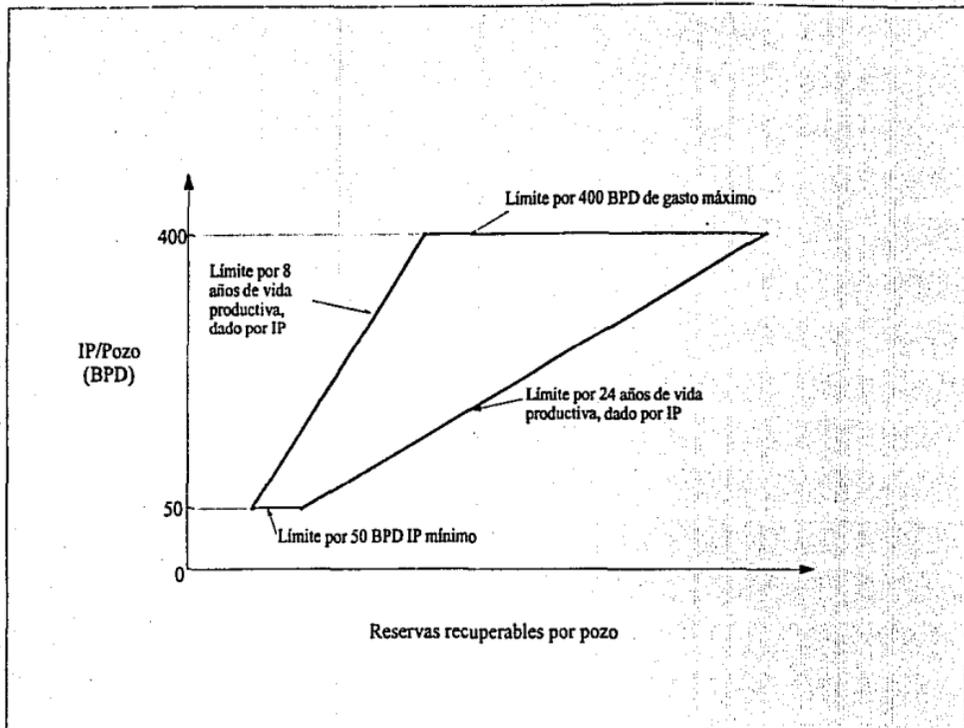


Figura 45. Correlación de IP/Pozo contra reservas por pozo obtenidas de los cálculos preliminares de los esquemas de declinación.

Esta serie de pasos sería repetida en cada paso de simulación, y debe proporcionar una manera de tratar el IP/pozo como una variable aleatoria- pero en forma tal que esta considere todas las dimensiones inter-relacionadas en cada paso.

4. Conociendo la reserva recuperable total del campo se puede determinar el número de pozos a perforar por una prueba iterativa en la cuál los pozos de desarrollo adicionales (para recuperar la misma cantidad de reservas) deberán perforarse hasta el tiempo en que el incremento de la utilidad VPN de cualquier pozo sea menor que el costo de perforación del pozo adicional.

Este criterio para establecer el número de pozos es, en esencia, continuar agregando pozos para acelerar la recuperación de las reservas hasta que el incremento de la ganancia de un pozo adicional sea excedido por el costo del mismo. Este esquema está hecho para optimizar el aprovechamiento maximizo de utilidad VPN sin considerar el patrón de espaciamiento.

Estos cuatro enfoques proveen algunas ideas de como manejar el potencial inicial y el programa del pozo de desarrollo en cada paso del análisis de simulación. Se recalca otra vez que ésto es una parte muy importante del modelo, y una cuidadosa consideración debe darse para éste con alternativas para manejar el IP y el número de pozos a ser perforados en cada paso.

Otras Aplicaciones de Simulación

En este capítulo se ha presumido que la variable dependiente descrita por simulación es algún valor medido como VPN beneficio discontinuo a t_0 por año. Pero hay otros tipos útiles de análisis para que el concepto de simulación pueda ser empleado eficazmente. En estas aplicaciones alternativas, la metodología de simulación restante es la misma (por ejemplo describir variables aleatorias como distribuciones, probar las distribuciones empleando números aleatorios, etc.) y solamente los tipos de variables cambian.

Un ejemplo es emplear análisis de simulación, para tratar de trazar el rango y distribución probable de las reservas totales a recuperar, en un yacimiento sedimentario después de solamente una cantidad de perforación exploratoria limitada. Considere el siguiente ejemplo:

Varias estructuras han sido perforadas en un área nueva de exploración, resultando varios descubrimientos aislados de gas y dos agujeros secos. Los descubrimientos confirman que los hidrocarburos están presentes en la cuenca que contiene muchas anomalías estructurales, semejante a algunas lejanas que han sido probadas. Parecen las condiciones para que exista un meritorio programa de exploración, pero el problema es que el área es remota y no hay facilidades para transportar (oleoductos) y para vender el gas.

La empresa está renuente a invertir cantidades grandes de capital hasta que se confirme que próximamente se tendrá disponible un mercado. Las compañías de oleoductos, en cambio, están renuentes a desembolsar gastos enormes para poner una línea hasta ellos, hasta tener una firme indicación de si la cuenca contendrá bastantes reservas para justificar la línea. ¿Como puede resolverse este callejon sin salida? ¿Como se puede obtener en el presente alguna idea del rango de reservas posibles en la cuenca?

Ejemplos de áreas en donde tal escenario está presente (antes en 1975) incluye el área del Noreste del banco de Arena de Australia y las cuencas en las Regiones Árticas de Canadá y Alaska. Las preguntas hechas en el escenario son realmente importantes y oportunas de que ocurran en más áreas donde las reservas son un número relativamente remoto para mercados de aceite y gas.

Lo siguiente es hacer un modelo de simulación que pueda ser usado para ganar experiencia en las preguntas. El modelo, llamado un modelo de reserva de yacimiento, se verá más tarde que tiene la flexibilidad de ser usado para evaluar muchas otras dimensiones de una exploración. Probablemente la única característica del modelo es el uso de un número aleatorio para simular, el proyecto, los resultados de perforar (hipotéticamente), todo se observó y/o proyectó en los prospectos del yacimiento.

El modelo mismo es más tarde desplegado como un diagrama de flujo del programa de una computadora. Para seguir la lógica involucrada se requiere primero definir los códigos, símbolos a emplear, como se muestra en la tabla 23. Ahora se supone que el yacimiento sedimentario consiste de un número grande (100 por ejemplo) de prospectos estructurales como en la figura 46. Estas anomalías pueden ser definidas por reconocimiento sísmico o puede ser que la existencia de algunas de las anomalías haya sido por hipótesis extrapolada de los sitios y densidad de anomalías identificadas por las porciones del yacimiento para las cuales, sin embargo, no se tienen corridas sísmicas.

Tabla 23.

Código de símbolos usados en el modelo de la reserva de la cuenca de la figura 46.

RESERVE	= Reserva calculada para una estructura productiva dada (aceite o gas).
TOTRESERVE	= Reserva (total) acumulativa para las estructuras probadas teniendo gas o aceite.
N	= Número de estructuras no perforadas.
NZERO	= Número total de estructuras en la cuenca.
P	= Fracción de las estructuras no perforadas que tienen hidrocarburos.
PZERO	= Relación de éxito (es decir la relación de las estructuras productivas al total de las estructuras después que toda la cuenca han sido perforada).
PROD	= Número de estructuras no perforadas que tienen hidrocarburos.
I	= Índice del número de los pasos de simulación
K	= Índice del número de las estructuras separadas (K=1,2,3,...,NZERO)

Para cada anomalía se describe una distribución de área y espesor neto productivo que refleja la variabilidad (incertidumbre) del tamaño físico de los prospectos y una distribución de barriles por acre-pie (o MPC por acre-pie si es una provincia de gas) que refleja los rangos de posibles de recuperaciones de hidrocarburos. Cada anomalía es identificada por un número K, y las tres distribuciones para cada anomalía son indentificadas por el apropiado subíndice K.

Los datos de entrada que el analizador tiene que especificar para el uso de éste modelo son:

- NZERO (Número total de estructuras en el yacimiento, 100, por ejemplo. Éste número sería calculado o determinado de los datos regionales sísmicos.)
- PZERO (Último gasto de producción en las estructuras totales, 0.15 por ejemplo. Este 15% medio de las 100 estructuras que tendrán gas o aceite por ejemplo, $100 \times 0.15 = 15$ estructuras. Así PROD al tiempo cero = NZERO x PZERO)

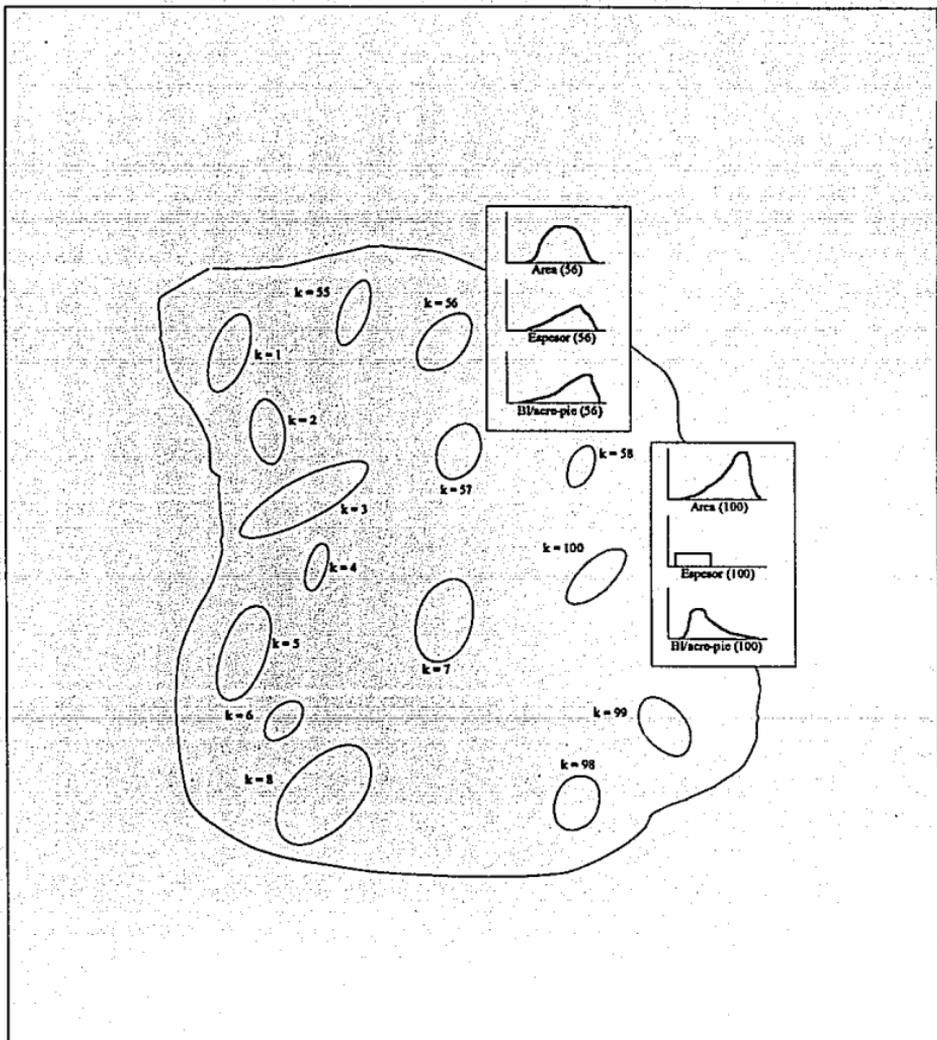


Figura 46. Mapa esquemático de una cuenca hipotética mostrando algunas de las 100 perspectivas no perforadas. Para cada anomalía se puede describir las distribuciones productivas del área, espesor y recuperación de aceite y gas por acre-pie. El valor k es solamente un índice numérico para mantener identificado cada prospecto en el modelo.

- c. Distribución del área
Distribución del espesor
neto productivo
Distribución
MPC/acre-pie (o
bls/acre-pie para aceite)
- } Para cada una de las NZERO estructuras.

El modelo mismo se describe por el diagrama de flujo del programa para una computadora en la figura 47. Para quién no esté acostumbrado a pensar en términos de diagramas de flujo, probablemente la figura 47 sea muy compleja e imposible. Pero no es muy complicada si se detiene a andar paso a paso. Al hacer estas suposiciones se estima que hay 100 estructuras en la cuenca y que 15% tienen hipotéticamente aceite. Esto es NZERO igual a 100 y PZERO es 0.15.

Para cada paso, lo primero es fijar algunas de las condiciones iniciales de los parámetros. Para iniciar aquí es en el sentido de cualquiera de los NZERO prospectos perforado anteriormente.

- TOTRESERVE se fija a 0 - refleja esas reservas acumulativas encontradas así, es cero.
- N se fija a NZERO - N es el número de estructuras no perforadas, que inicialmente es igual al número total en la cuenca, NZERO
- P se fija a PZERO - P es la probabilidad de descubrimiento en la estructura próxima probada, y corresponde al tiempo cero de 0.15, el valor como PZERO especificado en este ejemplo.
- PROD se fija e iguala al producto de NZERO multiplicado por PZERO - PROD es el número remanente de estructuras que tienen aceite, 15 en este caso al tiempo cero.

Lo siguiente en el programa es fijar K igual a 1, en el sentido ahora a "perforar" K = 1 estructura. Un número aleatorio expresado como un decimal es obtenido enseguida y es probado para ver si es menor que (1-P). P es la probabilidad de éxito, así (1-P) es la probabilidad de agujero seco. Para la ilustración numérica P es 0.15 para la estructura K=1 así (1-P) es 0.85. Por lo tanto si el número aleatorio es menor que 0.85 análogamente se ha perforado un agujero seco.

Siguiendo las flechas fuera del diamante hacia la caja inferior el programa se calcula un nuevo valor de N como N anterior menor 1 (para el ejemplo este valor nuevo de N ahora sería $100 - 1 = 99$; hay 99 estructuras no perforadas.). Después de examinar y observar que N no es cero se calcula una probabilidad nueva de éxito para la estructura próxima como $P = PROD / N$ (Para el ejemplo ésta sería $15/99$). Entonces $K = K + 1 = 1 + 1 = 2$ y se repite la serie para "perforar" la estructura $k = 2$.

Si el número aleatorio para la estructura inicial fuera igual a o más grande que 0.85 análogamente sería que se había hecho un descubrimiento, y la salida el diamante sería a la derecha al rectángulo identificado como B. Para determinar la magnitud de reservas descubiertas, la computadora prueba (una vez) valores de área, espesor, y recuperación por acre-pie de las $k = 1$ distribuciones.

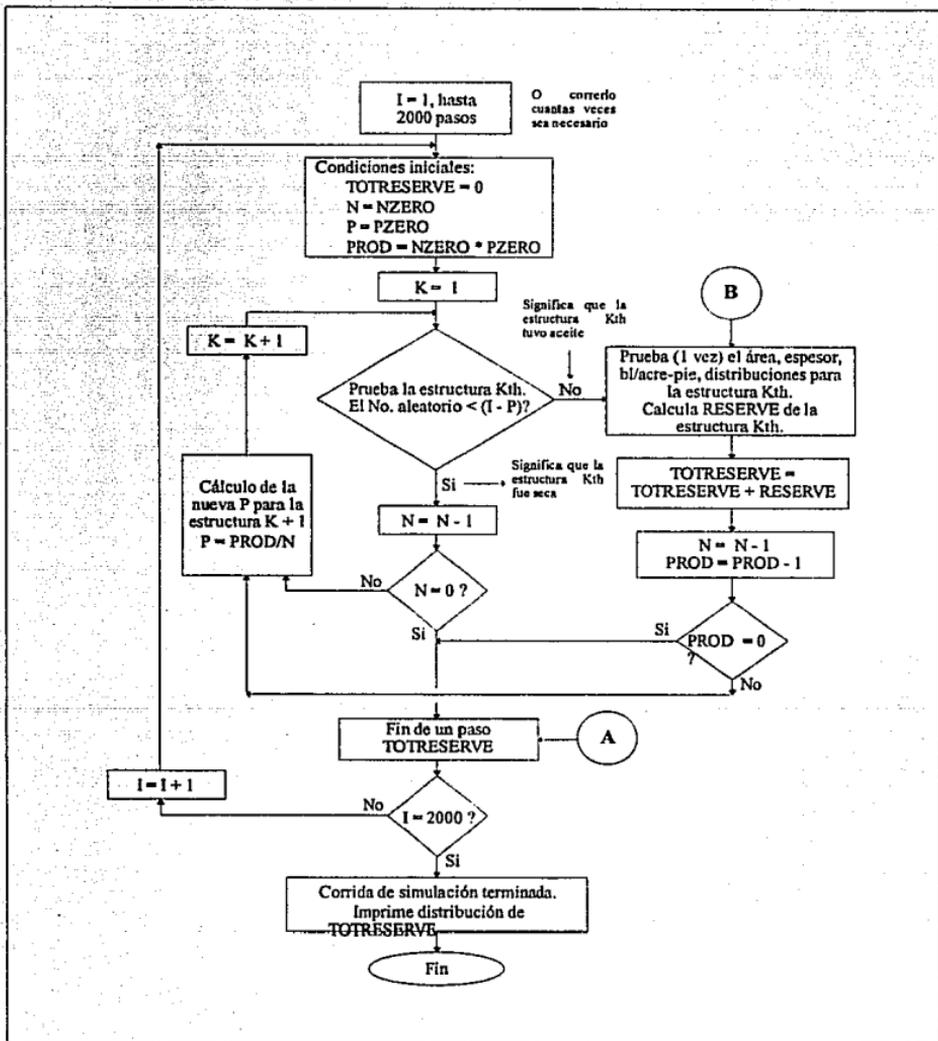


Figura 47. Diagrama de Flujo del modelo de Simulación de la Reserva de la cuenca.

Estos tres parámetros son multiplicados para obtener la RESERVA por campo, recuperable de la estructura. Siguiendo las flechas descendentes el programa, agrega ésta cantidad de reservas al total anterior (TOTRESERVE) para obtener las nuevas reservas acumulativas totales encontradas en el paso. Los siguientes parámetros N y PROD, son reducidos una unidad, para reflejar hecho de que hay una estructura menos no perforada y el número de estructuras productivas es reducido a uno a causa del descubrimiento que se acaba de realizar.

El paso próximo es probar si PROD es igual a cero. Si es así todas las estructuras productivas han sido encontradas y se finalizan los pasos. Si PROD no es cero, la secuencia de perforación continúa, para la perforación de la estructura $k + 1$ se usa una nueva probabilidad de descubrimiento correspondiente a $PROD / N$.

Finalizan los pasos cuando PROD es cero y el programa deberá imprimir el valor de TOTRESERVE, el hallazgo total de reservas. Entonces l es igual a $l + 1$ para el próximo paso y la secuencia entera vuelve a la parte de arriba del diagrama de flujo de condiciones iniciales para el paso próximo. Después de hacer, 2000 pasos los valores de TOTRESERVE son tabulados como salida para describir la distribución de reservas totales en la cuenca. Los 2000 pasos totales pueden ser corridos, hoy en día en computadoras en solo segundos, así el tiempo de computadora y gasto es insignificante.

Características importantes de este modelo:

- Las probabilidades de descubrimiento (P) se revisan después de que cada estructura ha sido examinada por fuera para reflejar el proceso de reemplazo.

- En cada paso las estructuras que tengan aceite varía. En el primer paso el aceite se puede encontrar en las estructuras, $K = 3, 10, 11, 14, 21, \dots$ en el paso siguiente los descubrimientos pueden ocurrir en las estructuras, $K = 1, 9, 12, 14, 36, 41, 56, \dots$ etc. Puesto que solamente en este análisis se trata de definir la distribución de reservas totales no es de interés cuáles estructuras específicamente eran productivas.

En virtud de que el número aleatorio usado para el proyecto si la K_{th} estructura es productiva, en esencia, se redistribuye aleatoriamente las reservas en varias estructuras después de cada paso.

- La secuencia de perforar continuará en cada paso hasta que todas las NZERO estructuras han sido probadas o hasta que todas las (NZERO x PZERO) estructuras hipotéticas hayan sido productivas, lo que ocurra primero. En el ejemplo (NZERO x ZERO) era 15, así la secuencia de perforación continuará hasta que se hayan hecho 15 descubrimientos. Cuando esto ocurre los pasos terminan.

Descrita la lógica fundamental del modelo se tienen algunas opciones.

Opción 1. El programa como se describe en la figura 47 imprimirá (o tabulará) las reservas totales para cada paso, recuadro A. Pero se puede también imprimir el valor de K una vez que la secuencia de computación se alcanza al final del paso. Este valor representa el número de estructuras que habían sido perforadas en cada paso antes de descubrir todas las (NZERO x PZERO) estructuras productivas. Después de 2000 pasos se tendrían 2000 valores de K que se describen como una distribución del número de estructuras probadas, antes del descubrimiento de los hidrocarburos en todas las estructuras. El valor mínimo de ésta distribución sería

$K = NZERO \times PZERO$, el caso donde no se perforaron agujeros secos. El valor máximo de la distribución sería $K = NZERO$, el caso en donde no se encontró la última estructura productiva hasta la estructura perforada anteriormente. Con una distribución de K valores se puede, con la experiencia ganada, obtener la cantidad de perforación exploratoria requerida para encontrar todas las reservas en el yacimiento. Se puede responder a la pregunta como:

a. - "Cuántos pozos exploratorios se perforarán para tener una certeza del 95% de que se habrían encontrado todas las reservas de la Cuenca?"

b. - "Si se perforan 30 pozos exploratorios ¿cuál es la probabilidad de que se hayan encontrado todas las reservas de la cuenca?"

Opción 2. Suponga que intereza evaluar la distribución de reservas encontradas, 20 en el programa de exploración de pozos (donde $20 < NZERO$). No hay problema. Solamente se necesita agregar otro punto de salida a los pasos y volver después de $K = 20$ pozos que habrían sido perforados, a pesar de sí originalmente, todas (o alguna) de las estructuras hipotéticas productivas había sido hallada. La distribución de reserva de capacidad proveería experiencia de los resultados que se pueden esperar de perforar 20 pozos. Y un más sí se pudiera perforar, 4 pozos por año ésto nos daría alguna idea del tiempo requerido para involucrar las reservas halladas antes de justificar el comienzo de un oleoducto.

Opción 3. ¿Qué sucede si algunas de las estructuras son mucho más grandes, que otras y pueden probablemente probarse primero? - ¿como puede el modelo ser usado para reflejar ésto? ¡sencillo! El programa prueba estructuras en el orden secuencial, en la serie de $K = 1, 2, 3, 4, \dots$, etc.. Pero es libre de indentificar las estructuras en la forma elegida. ¡Así si una estructura dada es probable de perforar primero (debido a consideraciones tales como - es grande la estructura o pequeña u oleoductos cerrados,...) simplemente se designa como la estructura $K = 1$! Si hay un orden específico de perforación se anticipa un subíndice a las estructuras en el mismo orden.

Opción 4. Se tienen muchas opciones en cuanto a la adición económica e inversión de capital, consideraciones para el análisis. Por ejemplo, al final de cada paso se sabrían las reservas totales descubiertas (recuadro A) y K , el número de estructuras probadas. De estos K pozos perforados el número de pozos cerrados descubiertos sería igual a $(NZERO \times PZERO)$, y el número de pozos exploratorios perforados secos en el paso sería $[K - (NZERO \times PZERO)]$. Estos números de pozos pueden ser multiplicados por el pozo cerrado descubierto, obtener lo que cuesta la inversión total de exploración y encontrar las reservas totales del paso. Al final de los 2000 pasos se puede tener una distribución de inversión de exploración, requerida para encontrar las reservas del yacimiento.

También, las reservas totales del yacimiento calculadas al final de cada paso pueden ser multiplicadas por un factor VPN / b (calculado después de los costos de operación, impuestos, costo de transporte, costos de desarrollo de perforación, etc.) para dar un VPN de renta de la reserva total del yacimiento. Después de repetir los pasos resulta una distribución de VPN Renta. Con ésta distribución y la distribución de capital de perforación exploratoria la empresa pudo obtener una idea de los valores esperados de un capital total de x dólares desembolsados al perforar.

Opción 5. Para ser aún más preciso, un valor o distribución de costos de perforación y VPN / b puede ser asignado a cada estructura. Estos considerarían la profundidad específica perforada, distancias de un probable oleoducto, de cada estructura etc.. Con este esquema, los

costos del aceite de estructura K_{th} probada y las rentas pueden calcularse al mismo tiempo como las reservas (figura 47, recuadro B). Estos costos, rentas, y reservas serían acumulados en el paso siguiente, y al final de cada paso el valor total de las reservas, costos totales de perforación exploratoria, y ganancia VPN total serían almacenados en la computadora. Después de los 2000 pasos las distribuciones de capacidad de reservas totales de la cuenca, costos totales de perforación exploratoria, y el total VPN de reservas encontradas se imprime.

Con esta opción los costos y rentas serían específicos para cada estructura, dada su localización geográfica, profundidad de sedimentos, etc. El factor VPN/bl usado en esta opción, indudablemente es dependiente de la cantidad de reservas por estructura. Construir esta correlación puede requerir algunos casos de estudios hipotéticos de varios proyectos posibles de desarrollo para dar los niveles de reservas en la estructura.

No hay duda en este punto de que el modelo de simulación de la figura 47 tiene muchas aplicaciones posibles para evaluar resultados de programas de perforación múltiple de pozos, así como ensayar, juzgar, trazar el rango probable y distribución de las reservas en una cuenca nueva.

Desde que el concepto de simulación empezó a aparecer primero en la literatura de petróleo en 1967, han habido numerosas aplicaciones surgidas para otro análisis. Algunas de estas otras aplicaciones incluyen:

- El análisis de la variabilidad de la saturación de agua congénita S_w , calculado de medidas de registros de pozo. (Referencia 8.2).
- La evaluación de si gastar más dinero en continuar probando en el laboratorio un nuevo refinamiento al proceso o llevar el riesgo de ir en escala completamente fuera, saber con seguridad si el proceso trabajará. (referencia 8.9).
- El largo de tiempo requerido en poner oleoductos. La única parte de esta aplicación es modelar las condiciones de tiempo de la simulación proyecto de lógica. (referencia 8.12).
- La determinación del rango de variación que puede ocurrir al calcular la posición del agujero perforado en un pozo direccional. (referencia 8.14)

En el último análisis el concepto de simulación describe la combinación analítica e interrelaciones de las variables aleatorias expresadas como distribuciones de probabilidad así, es que en general, el número de aplicaciones posibles de simulación solamente está limitado por la habilidad de definir y describir los sistemas de incertidumbre.

CONCLUSIONES

Los detalles de los métodos de análisis descritos en este capítulo muestran la complejidad y frustraciones en el análisis de riesgo de explotación petrolera. Parece como si todos los métodos tuvieran ciertas restricciones, suposiciones y/o limitaciones, y ninguno de los métodos representarían una alternativa ideal al sistema del mundo-real que se estudia y analiza.

Los datos disponibles no son suficientes, los cálculos son tediosos y complejos, el que decide no entiende los conceptos, y cualquier número de obstáculos semejantes empiezan a aparecer. A veces existe la sensación de que se hacen estos análisis con el pensamiento de que si esto confirma las ideas y se tiene el prejuicio a usarlas - de otra manera se quiere confiar en formas más antiguas de analizar los prospectos.

Todos estos comentarios son recalcados al principio del capítulo - cuando se abruma por la esterilidad de probabilidades y análisis de riesgo se pregunta "¿Cuál es la alternativa?" Seguro esto es difícil. Pero si se desconocen las probabilidades de riesgo e incertidumbre no se acaba en nada, y las estrategias de inversión, sin cualquiera de las consideraciones de éstos factores, serán bastante austeras, a menos que se sea excepcionalmente afortunado! Así, a pesar de las complejidades y frustraciones, los conceptos proveen solamente la verdadera base racional para avanzar en condiciones de riesgo e incertidumbre.

En una nota positiva, mientras que hay algunos huecos al determinar ciertos términos de probabilidad usando los métodos discutidos en este capítulo, hay un concepto nuevo que en muchos casos puede llenar el vacío donde los métodos presentes no son adecuados. Es el método de análisis de riesgo llamado simulación. El método puede extender las capacidades para evaluar probabilidades. Así hay más en cuanto al análisis de riesgo de explotación, y quizás el concepto de simulación puede mitigar algunas de las frustraciones y la sensación de duda en lo que se refiere a prospectos con el objeto de analizar riesgos de Exploración Petrolera.

En métodos numéricos de cuantificación de riesgo e incertidumbre se necesita recordar que los números no son necesariamente el fin de ellos mismos. Estos no son absolutos en análisis de riesgo.

NOMENCLATURA

N	Número total de prospectos productivos en una cuenca.
EMV	Valor medio esperado.
R _{min}	Reserva mínima requerida para justificar la terminación del pozo descubridor.
n	Número de eventos (pozos) en el programa multi pozo.
r	Número de resultados posibles en cualquier evento.
d _i	Número de elementos en el espacio de muestra designados para cada resultado con la condición de $(d_1 + d_2 + \dots + d_r) = N$.
x _i	Número de resultados que ocurren en cada una de las "r" categorías, con la condición de $(x_1 + x_2 + \dots + x_r) = n$.
X	Variable aleatoria
P	Probabilidad de éxito en algún evento dado $(0 \leq P \leq 1)$ (fracción).
p _i	Probabilidad de ocurrencia de los resultados en cualquier evento, dada la condición $(p_1 + p_2 + \dots + p_r) = 1$ (fracción).
VPN	Valor presente neto (dólares).
Y,Z	Variable independientes.
RF	Factor de recuperación (fracción).
BAF	Recuperación de aceite, (barriles por acre-pie).
Sw	Saturación de agua (%).
B _{oi}	Factor de volumen de aceite inicial (bls/bls).
φ	Porosidad (fracción).
RV	Variable aleatoria.
CF	Frecuencia acumulativa.
X'	Variable aleatoria transformada de la distribución normalizada.
H ^{EMV(aceite)}	Valor medio condicional.
RN	Número aleatorio (fracción).

REFERENCIAS

1. Newendorp, Paul D. and Root, Paul J., "Risk Analysis in Drilling Investment Decisions" *Journal of Petroleum Technology*, (Jun. 1968), pp. 579-585.
2. Hayward, J. T., "Probabilities and Wildcats tested Through Mathematical Manipulation" *Oil and Gas Journal*, Vol. 33, No. 26 (Nov. 15, 1934), pp. 129-131.
3. Mabra, D. Allen, Jr., "Here's a New Way to Evaluate Drilling Prospects" *Oil and Gas Journal*, Vol. 55, No. 5 (Feb. 4, 1957), pp. 186-188.
4. Schwade, Irving T., "Geologic Quantification. Description Numbers Success Ratio" *AAPG Boletín*. Vol. 51, No. 7 (Jul. 1967), pp. 1225-1239.
5. Arps, J. J. and Roberts, T. G., "Economics of Drilling for Cretaceous Oil on East Flank of Denver-Julesburg Basin" *AAPG Boletín*. Vol. 42, No. 11 (Nov. 1958), pp. 2549-2566.
6. Campbell, W. M. and Schuh, F. J., "Risk Analysis: Over all Chance of Success Related to Number of Ventures" *SPE Paper No. 218 presentado en el 36avo. Congreso Anual de la Society of Petroleum Engineers*, (Oct. 8-11, 1961), Dallas.
7. Walstrom, John E., "A Statistical Method for Evaluating Functions Containing Indeterminate Variables and Its Application to Recoverable Reserves Calculations and Water Saturation Determinations" *Geological Sciences, Publicaciones Stanford University*, Vol. IX, No. 2, p. 823.
8. Northern, P. G. "Risk, Probability and Decision-Making in Oil and Gas Operations" *Journal of Canadian Petroleum Technology*, Vol. 6, No. 4 (Dic. 1967), pp. 150-154.
9. Smith, Marvin B., "Probability Estimates for Petroleum Drilling Decisions" *Journal of Petroleum Technology*, (Jun 1974), pp. 687-695.
10. Profat, Alfredo, "Statistical Estimation of Wildcat Well Outcome Probabilities by Visual Analysis of Structure Contour Maps of Stafford County, Kansas" *Publicacion Kansas Geological Survey, Lawrence, Kansas*, 1974.
11. Megill, Robert E., "An Introduction To Risk Analysis" *The Petroleum Publishing Co. Tulsa* 1977.
12. Harbaugh, J.W., J. H. Doveton, and J. C. Davis, "Probability Methods in Oil Exploration" *Wiley Interscience, New York*, 1977.
13. Hertz, David B., "Risk Analysis in Capital Investment" *Harvard Business Review*, Vol. 42, No. 1 (Ene.-Feb., 1964), pp. 95-106.
14. Walstrom, J. E., Mueller, T. D. and McFarlane, R. C., "Evaluating Uncertainty in Engineering Calculations" *Journal of Petroleum Technology*, (Dic. 1967), pp. 1595-1603.
15. Smith, Marvin B., "Estimate Reserves by Using Computer Simulation Method" *Oil and Gas Journal*, (Mar. 11, 1968), pp. 81-84.

16. **Thorngren, J. T., "Probability Technique Improves Investment Analysis"**
Chemical Engineering, Vol. 74, No. 17, (Aug. 14, 1967), pp. 143-151.
17. **Sprow, Frank B., "Evaluation of Research Expenditure Using Triangular Distribution Functions"**
Industrial and Engineering Chemistry, Vol. 59, No. 7 (Jul., 1967), pp. 35-38.
18. **McCarron, J. K. "Computer simulation may lower costs for offshore pipelines"**
Oil and Gas Journal, (Feb. 17, 1969), pp. 84-93.