103

UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

ALGUNAS RELACIONES ENTRE DISTRIBUCIONES DE VARIABLES ALEATORIAS UNIVARIADAS DISCRETAS Y CONTINUAS

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I A

P R E S E N T A :

LUZ MARIA URIBE VARGAS

DIRECTOR DE TESIS: MAT. MARGARITA ELVIRA CHAVEZ CANO

MEXICO, D. F. 2001

ERSIDAD NACIONAL
AVPNºMA DE
MEXICO

MAT. MARGARITA ELVIRA CHÁVEZ CANO
Jefa de la División de Estudios Profesionales
P r e s e n t e

Comunicamos a usted que hemos revisado el trabajo de Tesis:

"ALGUNAS RELACIONES ENTRE DISTRIBUCIONES DE VARIABLES ALEATORIAS UNIVARIADAS DISCRETAS Y CONTINUAS"

realizado por: Luz María Uribe Vargas

Con número de cuenta 8925897-4 , pasante de la carrera de Actuaría

Dicho trabajo cuenta con nuestro voto aprobatorio.

A t e n t a m e n t e

Director de tesis Propietario	Mat. Margarita Elvira Chávez Cano	M E Ch Cano
Propietario	M en C. Beatriz Eugenia Rodríguez Fernández	Beatriz Rodríguez F.
Propietario	Dr. José Rubén Hernández Cid	Rubén Hernández C.
Suplente	M. en C. Inocencio Rafael Madrid Ríos	
Suplente	Act. Martha Guadalupe Rodríguez Rodelo	

Consejo Departamental de Matemáticas.

M. en C. José Antonio Flores Díaz.

A mi madre

A quienes son duales, como la *Bernoulli*. A quienes gustan contar eventos raros, como la *Poisson*. A quienes olvidan, cuando es pertinente, como la *geométrica* y la *exponencial*. A quienes pueden adoptar toda una *gamma* de formas inimaginadas, como la *beta*, y son de cuando en cuando *uniformes*. A quienes no se apellidan *Rayleigh* o *Weibull*, mas ejercen también funciones de riesgo. A quienes sorprenden pareciendo *normales*, siendo en realidad *Cauchy*: con momentos infinitos y lenta convergencia. A quienes en su construcción subjetiva siempre toman en cuenta la varianza, como la *t de Student*, la F o la χ^2...

A todas las *variables aleatorias* que deambulan en el espacio paramétrico, generando momentos finitos e infinitos y distribuyéndose a todo lo que da la masa o densidad de su discreta *y* continua existencia

AGRADECIMIENTOS

A mi madre, ***Lucía***: lucidez y amor infinitos, por T O D O

A ***mi familia***: mis ti@s, mis prim@s y mis 6 "sobris". Muy especialmente a mi ***padrino Pedrito*** por ser tan dicharachero y a mi ***madrina Lupita*** por la veladora que encendió para mí; a ambos por procurarnos tanto. A mi ***Mamalala*** y a mi ***Papajuan*** por mirar siempre por nosotras, incluso ahora desde una estrella. A mi primo ***Juan Manuel*** por la buena onda de prestarme sus inseparables estilógrafos.

A las maestras y los maestros que, ya sea oficial o extraoficialmente, me dieron clases tan disfrutables, y a sus ayudantes de ese entonces; especialmente a ***Adalberto García-Maynez***, ***Agustín Cano***, ***Agustín Román***, ***Ángel Carrillo***, ***Beatriz Rodríguez***, ***Begoña Fernández***, ***Enrique Andrade***, ***Fernando Alonso***, ***Javier Fernández***, ***Margarita Chávez***, ***Miguel Lara*** y ***Roberto Cánovas***.

A la ***Universidad*** y a quienes con sus impuestos me permitieron estudiar

A quienes este trabajo de algo les sirva

A esa energía revitalizadora que ronda en el aire y que tal vez (no me consta) se llame "dios"

A mis amig@s de la Facultad

A ***Aida*** Morales por mantener su solidaridad *siempre* despierta,
por recordarme tantas veces de mirar al cielo
para redescubrir la belleza del día, del atardecer o de la noche
y por disfrutar tanto con los postres

A la ***Chabelita*** Alonso por hacerme reír tantas veces
y por las múltiples tortas de pollo y chismes que compartimos

A ***Maru*** Rizo por su creatividad y por incorporarse
finalmente a este tránsito del miedo a la locura y el disfrute

A ***Mede*** Soriano por ese cumplido tan hermoso que guardo
y por siempre llamarme "Lucecita". A su amiga ***Cris*** Alvarado
por su papiroflexia feliz y por ser tan risueña.

A ***Lalo*** M. Naranjo por ser tan diferente a mí y, sin embargo,
saber "darme el avión" sin que me dé cuenta

A ***Noé*** Celestino por su sentido del humor y por ser tan auténtico.

A ***Pablito*** Bernal por enseñarme que los *ángeles de la guarda* ¡en verdad existen!

A ***Pau*** Rivera por su paciencia para con mi dispersión en las tareas de muestreo
y por ser compañera (¡quién lo dijera bonita!) del mismo "dolor" *hasta el último momento*

A ***Angie*** Mendoza, a ***Moni*** Ángeles, a ***Norma*** Reyes, a ***Lety*** Ramírez,
y a ***Raúl*** Pérez por ser tan dedicad@s al estudio y por el trabajo conjunto
que alguna vez desarrollamos Lo mismo para ***Fanny*** Jasso,
a quien también le agradezco que me haya prestado el Mood.

A mis "otr@s" amiguit@s hallados

A la "banda" CCH-era: *Luz, Raulito, Fer, Ricardo, Elizabeth Omar y..* ¡uff!, a tod@s l@s
compañer@s, y todos l@s *profes*, por todo lo compartido y porque con ell@s aprendí tantísimo
del trabajo en equipo Lo mismo para *el José*, a quien también le agradezco que,
en su momento, haya estado "en la esquina de ningún lugar"

A ***Ángel*** Bernal, por toda su amabilidad, por dibujarme el *Diagrama Final*
y por sorprenderme con que *casi* todo es posible, *incluso* con una computadora

A ***Adrianita*** López, ***Alfonso*** Mejía, ***Bety*** Zubieta, ***Ivonne*** Orellana,
Lulú Herrera, ***Norma*** Contreras, ***Pili*** Rivera, ***Ricardo*** Aparicio y
Vicky Muñoz, compañer@s del CONAPO que de una u otra manera
estuvieron al pendiente de este trabajo

A ***Xio*** Peinador por todo el consuelo y la disposición para ayudar que me brindó
con este "asunto tesil" y por sus bellos amuletos

A tod@s aquell@s que alguna vez preguntaron con interés "¿cómo vas?"
y que tal vez por ahora desafortunadamente olvido.

A mis sinodales

A ***Bety***, por su jovialidad y su entusiasmo hacia la vida,
que tanto tiempo de "diván" me ahorraron

A ***Rubén***, por su paciencia inmensa,
su ¡¡yaaa!! en algún momento tan justo y necesario y por su risa.

A ***Martha***, por su buena disposición y su destanteadora seriedad.

A ***Rafael***, por su comprensión para con la revisión de este trabajo

(A ***Maru***, por comprender la situación de (*casi*) última hora)

Y, por supuesto, a ***Margarita***, por adoptar ese corte de cabello actual,
por sus vivires y desvivires compartidos durante este proceso
y por *todo* su apoyo, no sólo *académico* sino también *técnico, administrativo y moral.*

* * * * * * * * * *

A *todas* y *todos*, por su colaboración para con esta tesis

t l a h z o c a m a t i
("de mi boca con amor" ≅ *gracias*)

Luz

ÍNDICE

pág.

ÍNDICE DE FIGURAS

pág

Algunas relaciones entre distribuciones de variables aleatorias univariadas discretas y continuas

Presentación

Este trabajo está basado fundamentalmente en un artículo de Lawrence M Leemis publicado en 1986 en la revista *The American Statistician*, cuyo título puede traducirse como "Relaciones entre distribuciones univariadas comunes", mismo que se corresponde con el título de este trabajo de tesis[1]. La síntesis de este artículo es un diagrama en el que se esquematizan 56 relaciones encontradas entre 28 distribuciones –9 discretas y 19 continuas– Estas relaciones son de tres tipos. distribuciones límite, transformaciones y casos especiales

Leemis propone dos aplicaciones del diagrama, mismas que aquí se suscriben. Primero, dice, después de presentar las distribuciones univariadas comunes en un curso introductorio de probabilidad, puede ser utilizado para indicar cómo las distribuciones se relacionan una con otra Segundo, en un curso avanzado, provee un repaso rápido de las distribuciones univariadas importantes Es en este sentido que se plantearon los contenidos del presente trabajo, organizándolo en tres capítulos.

El objetivo principal de esta tesis es presentar el desarrollo matemático de cada una de las relaciones, lo cual se consolida en el último capítulo. Así en el primer capítulo se plantean los conceptos y resultados que sustentan teóricamente tres de las cuatro herramientas que se utilizarán para verificar las relaciones del esquema Estas técnicas son el *método de la función de distribución*, el *método de las transformaciones*, el *método de la función generadora de momentos* y el *teorema del límite central* El segundo es un capítulo de presentación de las 28 distribuciones que entrarán en juego; además de la definición –cuando fue posible– se incluyeron otros elementos tales como algunos hechos históricos sobre ella, algunos ejemplos de aplicación y algunas propiedades interesantes asociadas

Durante el desarrollo del trabajo se encontraron algunas relaciones adicionales Aquí se da cuenta de aquellas (12) que responden a los mismos criterios de las relaciones que se presentan en el esquema de Leemis. De aquí que el *Diagrama Final* presentado al final de la sección de *Consideraciones finales y conclusiones* sea una conclusión global de este trabajo de tesis

Finalmente, sólo queda expresar el deseo de que este trabajo encuentre alguna utilidad entre quienes estén vinculados con el estudio o el uso de las distribuciones de variables aleatorias; en particular, entre los estudiantes de los cursos de probabilidad y estadística que se imparten en la Facultad de Ciencias

---

[1] Leemis, Lawrence M , "Relationships among common univariate distributions", en *The American Statistician*, Mayo 1986, Vol 40, No 2, pp 143-146 En 1985, el Profesor Leemis se encontraba en la Escuela de Ingeniería Industrial de la Universidad de Oklahoma Una copia del artículo completo puede consultarse en el ANEXO 8

1

CONCEPTOS Y RESULTADOS PRELIMINARES

En este capítulo se presentan los principales conceptos y definiciones, así como los teoremas y propiedades que se manejan dentro del lenguaje de la teoría de las distribuciones de las variables aleatorias, y que se requieren para la comprensión de los contenidos de los dos capítulos siguientes. Una vez definidos los términos "variable aleatoria", "función de distribución acumulativa", y "función de densidad de probabilidad", los contenidos del capítulo están encaminados a la presentación de las tres técnicas que se utilizarán en el Capítulo 3 para verificar las distintas relaciones entre las distribuciones. Estas técnicas son: la de la *técnica función de distribución acumulativa*, la *técnica de la transformación* y la *técnica de la función generadora de momentos*; las tres permiten determinar la distribución de una función de una o más de una variable aleatoria El Teorema del Límite Central que es otra de las herramientas a utilizar se presenta en el Capítulo 2, dentro del apartado de presentación de la distribución normal.

Se buscó incluir todos y cada uno de los resultados pertinentes para sustentar teóricamente cada uno de los supuestos o procedimientos que se involucran en las tres técnicas, con el fin de contrarrestar, en lo posible, el que se perciba una técnica como una mera "receta". Si bien en la práctica se suelen aplicar herramientas acabadas sin mayor reflexión sobre su sustento teórico –lo cual finalmente es comprensible–, se consideró que este trabajo era una buena oportunidad para detenerse un poco en su asimilación Este interés trajo consigo un riesgo: el de la dispersión excesiva. Explorar el por qué de las cosas abre rutas infinitas que, paradójicamente, por cuestiones de tiempo y espacio hay que tratar de acotar. En este caso, eso no resultó fácil y es por ello que se incluye un buen número de notas al pie; su finalidad es no omitir elementos que se consideraron útiles e interesantes, pero que dan lugar a la dispersión o dificultan el enlace de las ideas centrales

El orden elegido para la presentación de los contenidos de este capítulo procuró ser inductivo, es decir, que los conceptos y resultados expuestos fueran cada vez más generales Es por ello que la "técnica de la función generadora de momentos" se pospuso hasta el final del capítulo, pues requiere de elementos adicionales a las otras dos técnicas. La "técnica de la función de distribución acumulativa", así como la "técnica de la transformación" se presentan primero para el caso de funciones univariadas, para después generalizarse. Se incluyen ejemplos de algunos de los conceptos y resultados. Por supuesto, puede decirse que en el Capítulo 3 se encuentra un número más amplio de "ilustraciones" de las tres grandes herramientas (de las cuatro, de hecho, incluyendo al Teorema del Límite Central).

VARIABLE ALEATORIA

La variables aleatorias están asociadas a experimentos aleatorios. Se entiende como un "experimento aleatorio" aquel cuyo resultado es incierto, por ejemplo.

1. lanzar un dado y observar la cara que queda hacia arriba;
2. el lanzamiento de 3 monedas;
3. el registro del número de muertes por accidentes automovilísticos en una zona,
4. el registro del tiempo en horas que funciona un foco; y,
5. contar el número de veces que llueve y registrar en pulgadas la precipitación pluvial durante algún mes en alguna región.

Como motivación previa para definir "variable aleatoria" considere el ejemplo siguiente.

Ejemplo 1.1 (*ejemplo introductorio de variable aleatoria*). Suponga un estudio de opinión en el que interesa preguntar a 50 personas si están de acuerdo o en desacuerdo con cierta cuestión Si se registra con un "1" el acuerdo y con un "0" el desacuerdo, el espacio muestral para este experimento tiene 2^{50} elementos, cada uno de los cuales es un vector de unos y ceros con 50 entradas. Ante tales magnitudes, se hace necesaria una forma de resumir ese escenario, pues puede ser que la única cantidad de interés sea el número de personas que están de acuerdo (o, equivalentemente, en desacuerdo) del total de 50. En ese caso, si se define una variable X como el *número de unos registrados dentro de un total de* 50 *posibilidades*, se habrá captado la esencia del problema. Note que el espacio muestral de X es el conjunto de enteros $\{0, 1, 2, \ldots, 50\}$, el cual es mucho más fácil de manipular que el espacio muestral original.

Observe que al definir la cantidad X se está definiendo una *función* del *espacio muestral original* a un *nuevo espacio muestral*: el conjunto de enteros $\{0, 1, 2, \ldots, 50\}$ que es un subconjunto de los números reales.

◆

Una variable aleatoria se define sobre un espacio de probabilidad (Ω, $\mathcal{A}$, $P[\cdot]$), donde Ω es un espacio muestral, $\mathcal{A}$ es una sigma-álgebra de eventos y $P[\cdot]$ es una función de probabilidad con dominio $\mathcal{A}$ [1]

Definición 1.1
Variable aleatoria[2]

Dado un espacio de probabilidad (Ω, $\mathcal{A}$, $P[\cdot]$), una *variable aleatoria*, denotada por $X(\cdot)$ o, simplemente, X, es una función definida del espacio muestral Ω a los números reales.[3] La función $X(\cdot)$ debe ser tal que el conjunto A_r, definido como $A_r = \{\omega : X(\omega) \leq r\}$, pertenezca a $\mathcal{A}$ para todo número real r

◈

La segunda parte de la *Definición 1.1* quiere decir que se requiere que cualquier colección de ω 's para las cuales $X(\omega) \leq r$ sea un evento (es decir, un elemento de $\mathcal{A}$) para cada número real. Una definición más sencilla de variable aleatoria (pero suficiente para efectos de lo que en este trabajo se desarrollará) es la siguiente.

Definición 1.1 bis
Variable aleatoria[4]

Una *variable aleatoria*, X, es una función que va del espacio muestral Ω a los números reales.

◈

Usualmente, una variable aleatoria se describe en términos de un experimento aleatorio, en lugar de hacerlo especificando su forma funcional, como podrá observarse más adelante en los *Ejemplos 1.2* y *1.3*. Así, en esos términos, Ω es la totalidad de los resultados de ese experimento y la variable aleatoria $X(\cdot)$ con dominio Ω hace corresponder algún número real a cada resultado del experimento.

---

[1] Para mayor referencia sobre un espacio muestral puede consultarse ANEXO 1

[2] Definición presentada por Mood *et al* [1974], p 53

[3] Más formalmente, es una funcion medible (es decir, que puede encontrarse su función inversa) que asigna valores reales a los eventos que conforman Ω

[4] Definición presentada por Casella y Berger [1990], p 26

Al codominio de la función $X(\cdot)$ se le denomina ***rango*** de la variable aleatoria y aquí se le denotará como $\Re_X$. El rango de una variable aleatoria puede ser numerable (ya sea finito o infinito) o no numerable La numerabilidad o no numerabilidad del rango dará lugar a una clasificación de las variables aleatorias en "discretas" y "continuas", respectivamente, como se verá más adelante

Aunque la notación $X(\cdot)$ resulta más adecuada porque enfatiza que una variable aleatoria es una función, por simplicidad en la literatura se utiliza la notación abreviada X, y lo mismo se hará en este trabajo de aquí en adelante. En general, se utilizan las últimas letras mayúsculas del alfabeto, con o sin subíndices, para denotar a las variables aleatorias; y las correspondientes letras minúsculas para denotar sus valores. Así, por ejemplo, la variable aleatoria X puede tomar el valor x[5] A continuación se presentan algunos ejemplos de variables aleatorias.

Ejemplo 1.2 (*ejemplo de variable aleatoria*)
Considere el experimento de lanzar una moneda tres veces. Sea X la variable aleatoria que denota el *número de águilas obtenidas*. Representando con una A el resultado "águila" y con una S el resultado "sol", el espacio muestral de este experimento es $\Omega = \{(A, A, A), (A, A, S), (A, S, A), (S, A, A), (A, S, S), (S, A, S), (S, S, A), (S, S, S)\}$. Una enumeración completa del valor de X para cada punto del espacio muestral se presenta en la *Figura 1*.

ω	$X(\omega)$
(A, A, A)	3
(A, A, S)	2
(A, S, A)	2
(S, A, A)	2
(A, S, S)	1
(S, A, S)	1
(S, S, A)	1
(S, S, S)	0

Figura 1

[5] Sobre las palabras "variable" y "aleatoria", Mood *et al* [1974] comentan que no existe *una justificación convincente* para su uso. Consideran que la expresión "variable aleatoria" es poco apta, pero que ha ganado tal difusión que sería desatinado tratar de cambiarla. Vale la pena comentar que las *Definiciones 1.1* y *1.1.bis* son definiciones "modernas" de variable aleatoria, pues antiguamente, se le definía, atendiendo al sentido común, como algo *variable* que tomaba sus valores al *azar*.

El rango de la variable aleatoria X es $\Re_X = \{0, 1, 2, 3\}$.[6]

◊

Ejemplo 1.3 (*otro ejemplo de variable aleatoria*) Considere el experimento de lanzar dos dados Ω consta de 36 puntos y puede escribirse como $\Omega = \{(i,j) \quad i = 1, \ldots, 6 \ \text{y} \ j = 1, \ldots, 6\}$ Distintas variables aleatorias pueden ser definidas. Por ejemplo, sea X la suma de las caras que caen hacia arriba. Entonces $X(\omega) = i + j$ si $\omega = (i, j)$, y así $\Re_X = \{2, 3, \ldots, 12\}$.

Asimismo, sea Y la diferencia en valor absoluto de las caras que caen hacia arriba; entonces $Y(\omega) = |i - j|$ si $\omega = (i, j)$. En este caso, $\Re_Y = \{0, 1, \ldots, 5\}$

◊

Ejemplo 1.4 (un *ejemplo más de variable aleatoria*). Tómese un foco y considere la variable aleatoria X como el tiempo (en horas) de duración hasta que se funde. El rango de esta variable aleatoria es $\Re_X = \{x: x \geq 0\}$.

◊

Las variables aleatorias pueden clasificarse en "discretas" y "continuas", según sea su *rango*

Definición 1.2
Variables aleatorias discretas y continuas

Una variable aleatoria X se dice que es *discreta* si su rango es numerable y se dice que es *continua* si su rango es no numerable

◊

Con esta definición puede verse que en los *Ejemplos 1.2* y *1.3* se definieron variables aleatorias discretas, mientras que la variable aleatoria definida en el *Ejemplos 1.4* es continua. Las variables

[6] Mood *et al* presentan el ejemplo de lanzar una moneda una sola vez y contar el numero de águilas Este no sólo es más sencillo, sino que resulta muy útil para ilustrar la segunda parte de su definición de variable aleatoria (*Definición* 1 1), es decir, para mostrar que $\{\omega : X(\omega) \leq r\}$ pertenece a $\mathcal{A}$ para cualquier numero real r.

En ese caso el espacio muestral es $\Omega = \{\text{águila, sol}\}$; $X(\omega) = 1$ si $\omega =$ águila y $X(\omega) = 0$ si $\omega =$ sol. es decir, el rango de X es $\Re_X = \{0, 1\}$. $\mathcal{A}$ consiste de los cuatro subconjuntos ϕ, $\{\text{águila}\}$, $\{\text{sol}\}$ y Ω Ahora, si $r < 0$, $\{\omega \quad X(\omega) \leq r\} = \phi$, si $0 \leq r < 1$, $\{\omega : X(\omega) \leq r\} = \{\text{sol}\}$; si $r \geq 1$, $\{\omega : X(\omega) \leq r\} = \Omega = \{\text{águila, sol}\}$. Por lo tanto, para cada r el conjunto $\{\omega : X(\omega) \leq r\}$ pertenece a $\mathcal{A}$, entonces X es una variable aleatoria.

aleatorias también pueden caracterizarse como discretas o continuas dependiendo de la forma de su "función de distribución de probabilidad" la cual se definirá más adelante

❖ *Sobre la notación*. A partir de este momento, cuando resulte conveniente la abreviatura se escribirá "v.a." en lugar de variable(s) aleatoria(s).

FUNCIÓN DE DISTRIBUCIÓN ACUMULATIVA Y FUNCIÓN DE DENSIDAD [MASA] DE PROBABILIDAD

Puesto que el valor de una variable aleatoria X puede estar determinado por el resultado de un experimento, pueden asignarse probabilidades a los posibles valores de la variable aleatoria, $P[X(\omega) = x] = P[X = x]$ (donde $\omega \in \Omega$ y $x \in \Re_X$). No sólo eso, pueden calcularse también probabilidades de eventos como $\{X(\omega) \leq x\} = \{X \leq x\}$ o $\{X = x\}$, para *cualquier* número real x, o de manera general $\{X \in A\}$ para algún conjunto A [7]

A cada variable aleatoria X puede asociársele una función llamada la "función de distribución acumulativa" de X o más comúnmente llamada "función de distribución", la cual puede utilizarse para responder cualquier pregunta de probabilidad referida a la variable aleatoria.

Definición 1.3
Función de distribución acumulativa

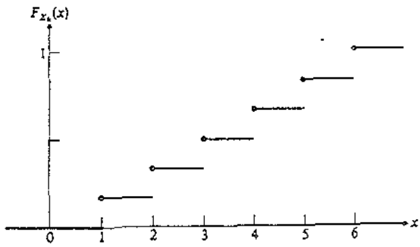
La *función de distribución acumulativa* (*f.d.a.*) de una variable aleatoria X, denotada por $F_X(\cdot)$, se define como aquella función cuyo dominio es el conjunto de los números reales, su codominio es el intervalo $[0,1]$ y que satisface que $F_X(x) = P[X \leq x] = P[\{\omega : X(\omega) \leq x\}]$ para todo número real x.

◆

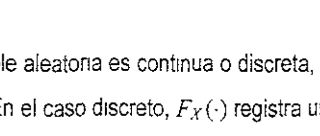
A continuación se presentan dos ejemplos de funciones de distribución acumulativas, uno discreto y uno continuo.

[7] Un ejemplo de esos cálculos se presenta en el ANEXO 2.

Ejemplo 1.5. *(función de distribución acumulativa discreta)*



Ejemplo 1.6. *(función de distribución acumulativa continua)*



Dependiendo de si la variable aleatoria es continua o discreta, la función de distribución acumulativa es continua o escalonada. En el caso discreto, $F_X(\cdot)$ registra un brinco en cada valor del rango de X $(\Re_X)$.[8]

Toda función de distribución acumulativa satisface ciertas propiedades, algunas de las cuales resultan inmediatas si se piensa en la definicion de $F_X(\cdot)$ en términos de probabilidades (véase ANEXO 3).

[8] Precisamente, algunos autores definen si una variable aleatoria es discreta o continua con base en la forma de la grafica de la función de distribucion acumulativa

Una característica fundamental de la función de distribución acumulativa es que está definida de manera *única* para cada variable aleatoria. En otras palabras, la función de distribución acumulativa permite *caracterizar por completo* una variable aleatoria. El teorema que establece que $F_X(\cdot)$ determina por completo la distribución de probabilidad de una variable aleatoria X, requiere de la noción variables *aleatorias idénticamente distribuidas* que se presenta a continuación

Definición 1.4
V.a.'s idénticamente distribuidas

Se dice que las variables aleatorias X y Y están idénticamente distribuidas si para todo conjunto A, $P[X \in A] = P[Y \in A]$.

◆

Nota: Observe que dos variables aleatorias que están idénticamente distribuidas no necesariamente son iguales. Es decir, la *Definición 1.4* no dice que $X = Y$.[9]

Teorema 1.1
Unicidad de la f.d.a.

Sean X y Y variables aleatorias con funciones de distribución acumulativas $F_X(\cdot)$ y $F_Y(\cdot)$, respectivamente. Entonces X y Y están idénticamente distribuidas si y sólo si $F_X(z) = F_Y(z)$ para todo z.

◆

❖ A partir de este momento se abreviará el término "función de distribución acumulativa" como *función de distribución*.

La función de distribución describe la distribución de los valores de una variable aleatoria. A su vez, asociada con una variable aleatoria y con su función de distribución está otra función llamada la "función de masa de probabilidad" o la "función de densidad de probabilidad", según se trate de un caso discreto o continuo, respectivamente. Ambas permiten calcular "probabilidades puntuales".

[9] Por ejemplo, considere el experimento de lanzar una moneda tres veces. Si se define a X como el número de soles obtenidos y a Y como el número de águilas obtenidas, se puede comprobar fácilmente que la distribución de X y de Y es la misma; es decir, que para $k = 0, 1, 2, 3$ se tiene que $P[X = k] = P[Y = k]$. Entonces X y Y están idénticamente distribuidas. Sin embargo, para ningún punto muestral ω se tiene que $X(\omega) = Y(\omega)$.

Definición 1.5
Función de masa de probabilidad

La *función de masa de probabilidad (f.m.p.)*, $f_X(x)$, de una variable aleatoria discreta X está dada por

$$f_X(x) = P[X = x] \qquad \text{para todo número real } x$$

◈

Definición 1.6
Función de densidad de probabilidad

La *función de densidad de probabilidad (f.d.p.)*, $f_X(x)$, de una variable aleatoria continua X es la función que satisface

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt \quad \text{para todo número real } x$$

◈

En el ANEXO 3 se presentan las propiedades que cumple toda función de densidad [masa] de probabilidad.

La función de densidad de probabilidad o función de masa de probabilidad de una variable aleatoria X es una función no negativa: es *estrictamente positiva* sólo sobre su *rango* ($\Re_X$) y es *cero* en *cualquier otro caso*. Esto significa que el rango de X puede verse como $\Re_X = \{x : f_X(x) > 0\}$. Definido así se le conoce como el *conjunto soporte* o simplemente *soporte* de $f_X(\cdot)$.

$F_X(x)$ y $f_X(x)$ modelan el comportamiento de un fenómeno que puede identificarse con una variable aleatoria. El siguiente teorema expone la relación que guardan la función de distribución y la función de densidad [o de masa] de una variable aleatoria X De manera particular, la relación [T.1.2.4] será de suma utilidad en este trabajo.

Teorema 1.2
Relación entre la f.d.a. y la f.m.p. o la f.d.p.

Sea X una variable aleatoria. Entonces $F_X(\cdot)$ puede obtenerse de $f_X(\cdot)$ y viceversa.

Demostración

Caso discreto. Sean $x_1, x_2, \ldots, x_n, \ldots$ los puntos de masa de una variable aleatoria discreta X. Suponga que $f_X(\cdot)$ está dada, entonces

$$F_X(x) = \sum_{\{j: x_j \le x\}} f_X(x_j). \qquad \text{[T.1.2.1]}$$

A la inversa, suponga que $F_X(\cdot)$ está dada, entonces

$$f_X(x_j) = F_X(x_j) - \lim_{h \to 0} F_X(x_j - h), \quad h > 0, \qquad \text{[T.1.2.2]}$$

para cada punto de masa x_j. Además, como $f_X(x_j) = 0$ para cada $x \neq x_j$, $j = 1, 2, \ldots$ entonces $f_X(x)$ queda determinada para todo número real.

Caso continuo. Sea X una variable aleatoria continua. Suponga que $f_X(\cdot)$ está dada, entonces $F_X(\cdot)$ se obtiene integrando $f_X(\cdot)$:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\,dt. \qquad \text{[T.1.2.3]}$$

A la inversa, si $F_X(\cdot)$ está dada entonces –por el Teorema Fundamental del Cálculo– $f_X(\cdot)$ puede obtenerse por diferenciación.

$$f_X(x) = \frac{d}{dx} F_X(x) \qquad \text{[T.1.2.4]}$$

para todos aquellos puntos para los cuales $F_X(x)$ es diferenciable.

◆

Por lo tanto, una variable aleatoria X puede identificarse o caracterizarse plenamente por su función de distribución de probabilidad $F_X(x)$ o por su función de densidad [masa] de probabilidad $f_X(x)$, ya que éstas permiten definir cuál es su distribución. (Más adelante se verá que también es posible caracterizar la distribución de una variable aleatoria mediante la "función generadora de momentos".) A lo largo del desarrollo de la probabilidad y la estadística se han encontrado *múltiples* distribuciones de probabilidad, a las cuales, bajo distintos criterios, se les ha reconocido con algún nombre particular. En el Capítulo 2 se presentan 24 distribuciones de probabilidad.

❖ *Sobre la notación*. En adelante, al decir que "una variable aleatoria X tiene distribución D" o "una variable aleatoria X se distribuye D" o "X es una variable aleatoria D" se entenderá que la función de distribución de probabilidad de X es $F_X(x)$ y que su función de densidad de probabilidad es $f_X(x)$. Una forma sintética de decir esto es mediante la notación $X \sim$ D.

Una buena parte de las relaciones entre distribuciones que se verifican en el Capítulo 3 se dan a través de funciones o transformaciones de las variables aleatorias En la siguiente sección se estudian tres técnicas para determinar la distribución de una función o transformación de una o más variables aleatorias.

DISTRIBUCIÓN DE UNA FUNCIÓN O TRANSFORMACIÓN DE UNA VARIABLE ALEATORIA

Al modelar fenómenos en términos de una variable aleatoria X, es frecuente que se esté interesado en el comportamiento de alguna función de X; por ejemplo, $Y = g(X)$. Es decir, dada la función de densidad de probabilidad $f_X(x)$ de una variable aleatoria X, se busca encontrar la distribución de probabilidad de $Y = g(X)$. Precisamente, en esta sección se presentan tres técnicas para encontrar dicha distribución

Ejemplo 1.7 (*ejemplo práctico de una función de una variable aleatoria*). Existen muchos problemas físicos en los que la deducción de la densidad de probabilidad de una forma funcional de una variable dada es sumamente importante. Por ejemplo, la velocidad de una molécula de gas (Ley de Maxwell–Boltzmann) se comporta como una variable aleatoria V con distribución conocida como *gamma*. Resulta de interés determinar la distribución de $E = mV^2$, la energía cinética de la molécula de gas, que es una nueva variable aleatoria

◈

☞ *La técnica de la función de distribución acumulativa: caso univariado*

El siguiente teorema engloba lo que se conoce como la *técnica de la función de distribución acumulativa*. No sólo eso, más adelante representará un *paso intermedio* fundamental para establecer la *técnica de la transformación* en el caso de variables aleatorias continuas.

(...continúa *La técnica de la función de distribución acumulativa: caso univariado)*

Precisamente, la técnica de la función de distribución resulta particularmente útil para el caso *continuo*. Se empieza por presentar dicha técnica para el caso univariado para posteriormente extenderla al caso bivariado.

Teorema 1.3
La técnica de la función de la distribución para el caso continuo univariado

Sea X una variable aleatoria continua con función de densidad de probabilidad $f_X(x)$. Sea $g(\cdot)$ una transformación definida de los reales a los reales. Entonces $Y = g(X)$ es una nueva variable aleatoria, cuya función de distribución puede ser obtenida integrando $f_X(x)$ *sobre la región definida* por $\{x \in \Re_X : g(x) \leq y\}$ como sigue

$$F_Y(y) = P[Y \leq y] = P[g(X) \leq y] = \int_{\{x:g(x)\leq y\}} f_X(x)dx \qquad [T.1.3].$$

◆

Una vez resuelta la integral en el término derecho de la expresión de [T.1.3], puede aplicarse el *Teorema 1.2* para obtener $f_Y(y)$ e identificar así (más fácilmente) la distribución de Y.

(termina *La técnica de la función de distribución acumulativa: caso univariado*)

Ejemplo 1.8 (*aplicación de la técnica de la función de distribución*). Sea X una variable aleatoria con función de densidad de probabilidad $f_X(x)$ dada por [10]

$$f_X(x) = \begin{cases} 1 & \text{si} \quad x \in (0,1) \\ 0, & \text{e.o.c.} \end{cases}$$

Sea $Y = X^2$. Aplicando el *Teorema 1.3*, la densidad de Y se obtiene calculando

[10] A esta distribución se le conoce como *distribución uniforme estándar* y también se abordará más ampliamente en el Capítulo 2.

$$F_Y(y) = P[Y \le y] = P[X^2 \le y] = \int_{\{x: x^2 \le y\}} f_X(x)dx = \int_0^{\sqrt{y}} dx = \sqrt{y}\,, \qquad \text{para} \quad 0 < y < 1$$

Por lo tanto

$$F_Y(y) = \sqrt{y}\, I_{(0,1)}(y) + I_{[1,\infty)}(y)$$
$$= \begin{cases} 0 & \text{si} \quad x \le 0 \\ \sqrt{y} & \text{si} \quad 0 < x < 1 \\ 1 & \text{si} \quad x \ge 1 \end{cases}$$

Ahora bien, en este caso, derivando $F_Y(y)$ (*Teorema 1.2*) se obtiene la función densidad de probabilidad de Y:

$$f_Y(y) = \frac{1}{2\sqrt{y}} I_{(0,1)}(y)\,.$$

◈

Ejemplo 1.9 (*aplicación de la técnica de la función de distribución*). Sea X una v.a. con función de densidad de probabilidad dada por

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\} \qquad \text{para} \quad -\infty < x < \infty, -\infty < \mu < \infty \quad \text{y} \quad \sigma > 0$$ [11]

Suponga que se quiere determinar la distribución de $Y = g(X) = X^2$

$$\begin{aligned} F_Y(y) &= P[Y \le y] \\ &= P[X^2 \le y] = P[-\sqrt{y} \le X \le \sqrt{y}] \\ &= \Phi(\sqrt{y}) - \Phi(-\sqrt{y}) = 2\int_0^{\sqrt{y}} \phi(u)du \\ &= 2\int_0^{\sqrt{y}} \tfrac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2} du = \tfrac{2}{\sqrt{2\pi}} \int_0^{y} \tfrac{1}{2\sqrt{z}} e^{-\frac{1}{2}z} dz \\ &= \int_0^{y} \tfrac{1}{\Gamma(\frac{1}{2})} \tfrac{1}{2\sqrt{z}} e^{-\frac{1}{2}z} dz \end{aligned}$$

para $y > 0$. (Como se verificará en los Capítulos 2 y 3, esta expresión puede ser identificada como la función de distribución de una distribución gamma con parámetros $r = ½$ y $\lambda = ½$.)

◈

[11] Como se verá en el Capítulo 2, esta función de densidad de probabilidad corresponde a la llamada *distribución normal*.

La técnica de la transformación: caso univariado

Como ya se había anticipado, una aplicación de la *técnica de la función de distribución* (y del *Teorema 1.2*) para encontrar la densidad de $Y = g(X)$ produce la *técnica de la transformación* para el caso continuo. En principio, se presenta esta técnica para el caso univariado y más adelante se extiende al caso bivariado. Al igual que la *técnica de la función de distribución*, la *técnica de la transformación* será particularmente útil en los contextos continuos. No obstante, se dedica primero un espacio al contexto discreto.

Teorema 1.4
La técnica de la transformación para el caso discreto univariado

Sea X una variable aleatoria discreta con rango $\Re_X = \{x_1, x_2,\ldots, x_i,\ldots\}$ y función de masa de probabilidad $f_X(x)$, de tal forma que X toma los valores $x_1, x_2,\ldots, x_i,\ldots$ con probabilidades $f_X(x_1), f_X(x_2),\ldots, f_X(x_i),\ldots$. Sea $g(\cdot)$ una transformación definida de los reales a los reales. Entonces $Y = g(X)$ es una nueva variable aleatoria cuyos posibles valores $y_1, y_2,\ldots,$ se determinan sustituyendo los valores sucesivos de X en $g(\cdot)$. La función de masa de probabilidad $f_Y(y)$ puede ser determinada por la transformación $g(\cdot)$, por la función de masa de probabilidad $f_X(x)$ de X y por las leyes de la probabilidad como sigue:

$$f_Y(y) = \begin{cases} f_X(g_{x_i}^{-1}(y)), & \text{si} \quad y = g(x_i) \quad \text{para} \quad i = 1,2,\ldots \\ 0, & \text{e.o.c.} \end{cases}$$

donde $g_{x_i}^{-1}(y)$ es la función inversa de $g(x)$ que mapea a y de regreso a x_i.

Si alguna $g(x_i)$ es igual a otra se suman las partes de la función de masa de probabilidad que correspondan a esas funciones que sean iguales. Por ejemplo, si $g(x_{j1}) = g(x_{j2}) = \ldots = g(x_{jn}) = g(x_i)$, entonces la parte de la densidad de $y = g(x_i)$ es

$$\sum_{i=1}^{n} (g_{xj_i}^{-1}(y)).$$

◆

(...continuará *La técnica de la transformación: caso univariado*)

Ejemplo 1.10 *(aplicación de la técnica de la transformación, para determinar la distribución de una función de una variable aleatoria discreta)* Suponga que X toma los valores $0, 1, 2, 3, 4$ y 5 con probabilidad $f_X(0) = p_0$, $f_X(1) = p_1$, $f_X(2) = p_2$, $f_X(3) = p_3$, $f_X(4) = p_4$ y $f_X(5) = p_5$, respectivamente. Sea $Y = g(X) = (X-2)^2$ Y puede tomar los valores $0, 1, 4$ y 9. Entonces

$f_Y(0) = f_X(2)$,

$f_Y(1) = p_1 + p_3$ (porque los valores 1 y 3 de x dan lugar al valor 1 de y)

$f_Y(4) = p_0 + p_4$ (porque los valores 0 y 4 de x dan lugar al valor 4 de y)

y $f_Y(9) = p_5$

◈

Ejemplo 1.11 *(aplicación de la técnica de la transformación, para determinar la distribución de una función de una variable aleatoria discreta)*. Sea X una variable aleatoria con f.m.p. dada por

$$f_X(x) = P[X = x] = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n$$

donde n es un entero positivo y $0 \leq p \leq 1$. A los términos n y p que pueden ser fijados en valores distintos, produciendo distintas distribuciones de probabilidad se les llama *parámetros* Se dice que X tiene *distribución binomial* con parámetros n y p.[12] Considere la variable aleatoria $Y = g(X)$, donde $g(x) = n - x$. Es decir $Y = n - X$. En este caso $\Re_X = \{0, 1, \ldots, n\}$ y $\Re_Y = \{ y: y = g(x),\ x \in \Re_X\} = \{0, 1, \ldots, n\}$. Para cualquier $y \in \Re_Y$, $n - x = g(x) = y$ si y sólo si $x = n - y$. Por lo tanto, $g^{-1}(y)$ es el punto único $x = n - y$, y

$$\begin{aligned} f_Y(y) &= \sum_{x=g^{-1}(y)} f_X(x) \\ &= f_X(n-y) \\ &= \binom{n}{n-y} p^{n-y} (1-p)^{n-(n-y)} \\ &= \binom{n}{y} (1-p)^y p^{n-y} \end{aligned}$$

Por tanto, Y tiene también distribución binomial con parámetros n y $1 - p$.

◈

[12] En el Capítulo 2 se ahondará sobre ésta y otras distribuciones de probabilidad

(...continúa *La técnica de la transformación: caso univariado*)

Teorema 1.5[13]
La técnica de la transformación para el caso continuo univariado

Sea X una variable aleatoria con función de densidad de probabilidad $f_X(x)$. Suponga que $g(x)$ es una función continua de x *estrictamente monótona* (creciente o decreciente). Entonces la variable aleatoria Y definida como $Y = g(X)$ tiene una función de densidad de probabilidad dada por

$$f_Y(y) = \begin{cases} f_X[g^{-1}(y)] \dfrac{d}{dy} g^{-1}(y) & \text{si } \; y = g(x) \; \text{ para alguna } \; x \\ 0 & \text{si } \; y \neq g(x) \; \text{ para toda } \; x \end{cases}$$

donde $g^{-1}(y)$ se define como el valor de x tal que $g(x) = y$.

(...continuará)

Demostración (del *Teorema 1.5*).

Caso (a): $g(x)$ es creciente.

Se calculará primero la función de distribución de $Y = g(x)$

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= P[g(X) \leq y] \\ &= P[X \leq g^{-1}(y)], \qquad \text{ya que } g^{-1} \textit{ es (también) creciente} \\ &= F_X(g^{-1}(y)) \end{aligned}$$

[13] Por sencillez y porque es suficiente para los fines de este trabajo, se expone la versión que presenta Ross [1988]. Una versión general de este resultado puede consultarse en Mood *et al.* [1974], en la cual se pide que la transformación $y = g(x)$ sea uno a uno (en particular, una transformación estrictamente monótona es uno a uno) y que $\frac{d}{dy} g^{-1}(y)$ sea continua y no negativa (en efecto, una transformación estrictamente monótona cumple ambas características) [véase Mood *et al* [1974], p. 200].

Ahora, derivando

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \frac{d}{dy} F_X(g^{-1}(y)) \\
&= F'_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) \\
&= f_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y)
\end{aligned}
$$

Caso (b) $g(x)$ es decreciente.

$$
\begin{aligned}
F_Y(y) &= P[Y \le y] \\
&= P[g(X) \le y] \\
&= P[X \ge g^{-1}(y)], \qquad \text{ya que } g^{-1} \text{ es (también) decreciente} \\
&= 1 - P[X < g^{-1}(y)] \\
&= 1 - F_X(g^{-1}(y))
\end{aligned}
$$

Derivando

$$
\begin{aligned}
f_Y(y) &= \frac{d}{dy} F_Y(y) \\
&= \frac{d}{dy}\left[1 - F_X(g^{-1}(y))\right] \\
&= -F'_X(g^{-1}(y)) \frac{d}{dy} g^{-1}(y) \\
&= f_X(g^{-1}(y))\left[-\frac{d}{dy} g^{-1}(y)\right]
\end{aligned}
$$

◈

Entonces, si se conoce la distribución de probabilidad de una variable aleatoria X y se está interesado en determinar la distribución de alguna función monótona de esa variable aleatoria, $g(X)$, basta expresar el evento $\{g(X) \le x\}$ en términos de X dentro de algún conjunto y hacer uso después del *Teorema 1.2* para determinar la función de densidad de probabilidad de $g(X)$, lo que equivale a aplicar el *Teorema 1.5*.

◈ ◈

(...continúa *La técnica de la transformación: caso univariado*)

Para el caso de funciones *univariadas monótonas resulta más* sencillo recurrir a la técnica de la transformación para evitar el cálculo de integrales. Más adelante, al abordar el caso de funciones multivariadas, la técnica de la función de distribución adquirirá mayor relevancia. Pero independientemente de la técnica que se utilice, es muy importante definir claramente el rango de la nueva variable aleatoria $Y = g(X)$

Es importante señalar que la condición de que $g(x)$ sea estrictamente monótona para determinar la f.d.p. [f.m.p.] de $g(X)$ no necesariamente es restrictiva, como se establece en el *Teorema 1.5 bis*.

Teorema 1.5bis
La técnica de la transformación para el caso continuo univariado

Sea X una variable aleatoria con función de densidad de probabilidad $f_X(x)$. Suponga que $y = g(x)$ es una transformación tal que es creciente para algunos subconjuntos de $\Re_X = \{x: f_X(x) > 0\}$ y decreciente para otros, pero es posible descomponer a $\Re_X$ en un número finito (o incluso numerable) de conjuntos ajenos $\Re_X^1, \ldots, \Re_X^m$, tales que $g(x)$ es estrictamente monótona sobre cada $\Re_X^i$, entonces la densidad de la variable aleatoria Y definida como $Y = g(X)$ puede encontrarse. Sea $x = g_i^{-1}(y)$ para cada $x \in \Re_X^i$. Entonces la densidad de Y está dada por

$$f_Y(y) = \sum_i \left|\frac{d}{dy} g_i^{-1}(y)\right| f_X\left(g_i^{-1}(y)\right)$$

donde la suma corre sobre aquellos valores de i para los cuales $g(x) = y$ para algún valor de x en $\Re_X^i$.

◆

(...termina *La técnica de la transformación: caso univariado*)

Para finalizar esta sección dedicada a las transformaciones univariadas de variables aleatorias apartado se presenta a continuación una serie de ejemplos generales de aplicación de los *Teoremas 1.5* y *1.5bis* con su correspondiente verificación "a pie". Observe que en todas y cada una de las verificaciones se hace *uso de la técnica de la función de distribución.* Puesto que todos estos ejemplos de aplicación son resultados a retomar como herramientas en el Capítulo 3, en lugar de enumerarlos como ejemplos, se enumerarán como resultados finales.

Ejemplos de aplicación de la técnica de la transformación (Teoremas 1.5 y 1.5bis).

Resultado 1.1
Distribución de X^2
(aplicación del *Teorema 1.5 bis*)

Sea X una variable aleatoria continua, $-\infty < x < \infty$, con función de densidad de probabilidad $f_X(x)$. entonces la distribución de $Y = X^2$ está dada por:

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})], \qquad y \geq 0. \qquad [1.1]$$

Demostración

Observe que dada cualquier x, $y \geq 0$. Entonces,

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= P[X^2 \leq y] = P[|X|^2 \leq y] \\ &= P[|X| \leq \sqrt{y}] = P[-\sqrt{y} \leq X \leq \sqrt{y}] \\ &= F_X(\sqrt{y}) - F_X(-\sqrt{y}) \end{aligned}$$

Por lo tanto, derivando se obtiene

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})], \qquad y \geq 0.$$

◇

Observe que el *Resultado 1.1* se obtendría directamente aplicando el *Teorema 1.5bis*, considerando la región donde $y = x^2$ es decreciente, $x < 0$, y la región donde $y = x^2$ es creciente, $x \geq 0$.

◇◇

El siguiente resultado corresponde a la distribución de la transformación inversa definida en el *Resultado 1.1*.

Resultado 1.2
Distribución de $X^{½}$
(aplicación del *Teorema 1.5*)

Sea X una variable aleatoria continua con rango no negativo, $x \geq 0$, y función de densidad de probabilidad $f_X(x)$, entonces la distribución de $Y = X^{½}$ está dada por:

$$f_Y(y) = 2y \cdot f_X(y^2), \quad y \geq 0. \qquad [1.2]$$ [14]

Demostración:

Como $x \geq 0$ entonces $y \geq 0$. Además, $y = g(x) = x^{½}$ es creciente. Calculando $F_Y(y)$:

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= P[X^{1/2} \leq y] \\ &= P[X \leq y^2] \\ &= F_X(y^2) \end{aligned}$$

Derivando:

$$f_Y(y) = [2y] \cdot f_X(y^2), \qquad y \geq 0.$$

◆

Bajo el *Teorema 1.5*:

$$g^{-1}(y) = y^2 = x, \text{ pues } g(x) = x^{1/2} = y$$

y

$$\frac{d}{dy} g^{-1}(y) = 2y \geq 0, \text{ ya que } y = g(x) \text{ es creciente (pues } x \geq 0),$$

con lo cual efectivamente $f_Y(y) \geq 0$.

◆◆

[14] Observe que mientras que para el *Resultado 1.1* no había restricción para el rango de la variable aleatoria X, el *Resultado 1.2* requiere que el rango sea no negativo para que la transformación $g(X)$ esté bien definida.

Resultado 1.3
Distribución de $1/X$
(aplicación del *Teorema 1.5*)

Sea X una variable aleatoria continua con rango positivo, $x > 0$, y función de densidad de probabilidad $f_X(x)$, entonces la distribución de $Y = 1/X$ está dada por

$$f_Y(y) = \frac{1}{y^2} \cdot f_X\left(\frac{1}{y}\right) \quad , \quad y > 0 \text{ [15]} \qquad [1.3]$$

Demostración.

Como $x > 0$ entonces $y > 0$. Además, $y = g(x) = \dfrac{1}{x}$ es decreciente. Calculando $F_Y(y)$:

$$\begin{aligned} F_Y(y) &= P[Y \le y] \\ &= P\left[\frac{1}{X} \le y\right] \\ &= P\left[\frac{1}{y} \le X\right] \qquad \text{(ya que } y > 0\text{)} \\ &= 1 - P\left[X \le \frac{1}{y}\right] \\ &= 1 - F_X\left(\frac{1}{y}\right) \end{aligned}$$

Derivando.

$$\begin{aligned} f_Y(y) &= \left[-\left(-\frac{1}{y^2}\right)\right] \cdot f_X\left(\frac{1}{y}\right) \\ &= \frac{1}{y^2} \cdot f_X\left(\frac{1}{y}\right) \quad , \quad y > 0 \end{aligned}$$

◇

Bajo el *Teorema 1.5*:

$$g^{-1}(y) = \frac{1}{y} = x, \text{ pues } g(x) = \frac{1}{x} = y$$

y $\quad \dfrac{d}{dy} g^{-1}(y) = -\dfrac{1}{y^2} < 0$, ya que $y = g(x)$ es decreciente.

[15] Observe que el *Resultado 1.3* restringe el rango de la variable aleatoria X a valores *estrictamente positivos* para que la transformación $g(X)$ esté *siempre* bien definida. También pueden considerarse sólo valores *estrictamente negativos*; en tal caso la transformación también es decreciente

El *Resultado 1.4* que se presenta a continuación es una forma de generalizar los *Resultados 1.1* a *1.3*, al considerar que el exponente de la transformación es cualquier número real (distinto de cero)

Resultado 1.4
Distribución de X^r
(aplicación del *Teorema 1.5*)

Sea X una variable aleatoria continua con rango positivo, $x > 0$, y función de densidad de probabilidad $f_X(x)$, entonces la distribución de $Y = X^r$, con r un número real distinto de cero, está dada por:

$$f_Y(y) = \frac{1}{r}\, y^{(1/r)-1} f_X(y^{1/r}), \qquad y > 0, \qquad \text{para todo } r \neq 0. \qquad [1.4]$$

O, equivalentemente:

$$f_Y(y) = \left[\frac{1}{r}\, y^{(1/r)-1}\right] f_X(y^{1/r}), \qquad y > 0, \qquad \text{para} \quad r > 0 \qquad [1.4.a]$$

$$f_Y(y) = -\left[\frac{1}{r}\, y^{(1/r)-1}\right] f_X(y^{1/r}), \qquad y > 0, \qquad \text{para} \quad r < 0 \qquad [1.4.b][16]$$

Demostración: Puesto que dependiendo del signo de r la transformación será creciente o decreciente es necesario considerar los dos casos. Para ello, sea $R > 0$.

(a) Caso $r = R > 0$.

Como $x > 0$ entonces $y > 0$. Además, $y = g(x) = x^R$ es *creciente*. Calculando $F_Y(y)$:

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= P[X^R \leq y] \\ &= P[X \leq y^{1/R}] \\ &= F_X(y^{1/R}) \end{aligned}$$

Derivando:

[16] El *Resultado 1.4* restringe el rango de la variable aleatoria X a valores estrictamente positivos para que la transformación $g(X)$ esté *siempre* bien definida

$$
\begin{aligned}
f_Y(y) &= \left[\frac{1}{R}y^{(1/R)-1}\right]f_X\left(y^{1/R}\right) \\
&= \left[\frac{1}{r}y^{(1/r)-1}\right]f_X\left(y^{1/r}\right), \qquad y>0
\end{aligned}
$$

(b) Caso $r=-R<0$

Como $x>0$ entonces $y>0$. Además, $y=g(x)=x^{-R}$ es *decreciente*. Calculando $F_Y(y)$.

$$
\begin{aligned}
F_Y(y) &= P[Y \le y] \\
&= P[X^{-R} \le y] = P\left[\frac{1}{X^R} \le y\right] \\
&= P\left[\frac{1}{X} \le y^{1/R}\right] = P\left[\frac{1}{y^{1/R}} \le X\right] \\
&= 1 - P\left[X \le \frac{1}{y^{1/R}}\right] = 1 - F_X\left(\frac{1}{y^{1/R}}\right) \\
&= 1 - F_X\left(y^{-1/R}\right)
\end{aligned}
$$

Derivando:

$$
\begin{aligned}
f_Y(y) &= -\left[-\frac{1}{R}y^{-(1/R)-1}\right]\cdot f_X\left(y^{-1/R}\right) \\
&= -\left[\frac{1}{(-R)}y^{(1/-R)-1}\right]\cdot f_X\left(y^{1/-R}\right) \\
&= -\left[\frac{1}{r}y^{(1/r)-1}\right]\cdot f_X\left(y^{1/r}\right) \qquad , \qquad y>0
\end{aligned}
$$

◈

Bajo el *Teorema 1.5*

$$g^{-1}(y) = y^{1/r} = x\text{. pues } g(x) = x^r = y$$

Así:

$$\frac{d}{dy}g^{-1}(y) = \frac{1}{r}y^{(1/r)-1} > 0, \quad \text{si } r>0 \text{ ya que } y=g(x) \text{ es creciente (pues } x>0)$$

y $$-\frac{d}{dy}g^{-1}(y) = -\frac{1}{r}y^{(1/r)-1} > 0, \quad \text{si } r<0 \text{ ya que } y=g(x) \text{ es decreciente (pues } x>0).$$

Por lo tanto:

$$\left|\frac{d}{dy}g^{-1}(y)\right| = \left|\frac{1}{r}y^{(1/r)-1}\right| = \frac{1}{|r|}y^{(1/r)-1} > 0, \text{ engloba ambos casos.}$$

◈ ◈

Resultado 1.5
Distribución de $|X|$
(aplicación del *Teorema 1 5 bis*)

Sea X una variable aleatoria continua, $-\infty < x < \infty$, con función de densidad de probabilidad $f_X(x)$, entonces la distribución de $Y = |X|$ está dada por

$$f_Y(y) = f_X(y) + f_X(-y), \qquad y \geq 0. \qquad [1.5]$$

Demostración:

Observe que dada cualquier x, $y \geq 0$. Entonces,

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= P[|X| \leq y] \\ &= P[-y \leq X \leq y] \\ &= F_X(y) - F_X(-y) \end{aligned}$$

Derivando:

$$f_Y(y) = f_X(y) + f_X(-y), \qquad y \geq 0.$$

◆

El *Resultado 1.5* se puede obtener también directamente aplicando el *Teorema 1.5bis*, considerando la región donde $y = |x| = g(x)$ es decreciente, $x < 0$, y la región donde $y = |x| = g(x)$ es creciente, $x \geq 0$ y sumando así las funciones de densidad de Y asociadas a cada región. El desarrollo se muestra a continuación:

Si $\quad x < 0 \Rightarrow \quad y = -x = g_1(x) \Rightarrow \quad x = -y = g_1^{-1}(y)$

$$\Rightarrow \quad -\left[\frac{d}{dy} g_1^{-1}(y)\right] f_X\left(g_1^{-1}(y)\right) = -[-1] f_X(-y) = f_X(-y), \qquad y \geq 0.$$

Si $\quad x \geq 0 \Rightarrow \quad y = x = g_2(x) \Rightarrow \quad x = y = g_2^{-1}(y)$

$$\Rightarrow \quad \left[\frac{d}{dy} g_2^{-1}(y)\right] f_X\left(g_2^{-1}(y)\right) = [1] f_X(y) = f_X(y), \qquad y \geq 0.$$

Por lo tanto:

$$f_Y(y) = \sum \frac{d}{dy} g_i^{-1}(y) f_X\left(g_i^{-1}(y)\right) = f_X(-y) + f_X(y), \qquad y \geq 0.$$

◆◆

A continuación se presenta la distribución de una combinación lineal de una variable aleatoria

Resultado 1.6
Distribución de $cX + d$
(aplicación del *Teorema 1 5*)

Sea X una variable aleatoria continua, $-\infty < x < \infty$, con función de densidad de probabilidad $f_X(x)$, entonces la distribución de una combinación lineal de X, $Y = cX + d$, donde c y d son números reales, $c \neq 0$, está dada por:

$$f_Y(y) = \frac{1}{c} f_X\left(\frac{y-d}{c}\right), \qquad -\infty < y < \infty \qquad [1.6]$$

Demostración Puesto que dependiendo del signo de c la transformación será creciente o decreciente es necesario considerar los dos casos.

(a) Caso $c > 0$.

Como $-\infty < x < \infty$ entonces $-\infty < y < \infty$. Además, $y = g(x) = cx + d$ es *creciente* (geométricamente, la transformación define una recta con pendiente *positiva*) Calculando $F_Y(y)$:

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= P[cX + d \leq y] \\ &= P[cX \leq y - d] \\ &= P\left[X \leq \frac{y-d}{c}\right] \qquad \text{(pues } c > 0) \\ &= F_X\left[\frac{y-d}{c}\right] \end{aligned}$$

Derivando:

$$f_Y(y) = \frac{1}{c} \cdot f_X\left(\frac{y-d}{c}\right), \qquad -\infty < y < \infty .$$

(b) Caso $c < 0$.

Como $-\infty < x < \infty$ entonces $-\infty < y < \infty$. Además, $y = g(x) = cx + d$ es *decreciente* (geométricamente, la transformación define una recta con pendiente *negativa*). Calculando $F_Y(y)$:

$$\begin{aligned} F_Y(y) &= P[Y \le y] \\ &= P[cX + d \le y] \\ &= P[cX \le y - d] \\ &= P\left[X \ge \frac{y-d}{c}\right] \qquad \text{(pues } c < 0\text{)} \\ &= 1 - P\left[X \le \frac{y-d}{c}\right] \\ &= 1 - F_X\left[\frac{y-d}{c}\right] \end{aligned}$$

Derivando:

$$f_Y(y) = -\frac{1}{c} \cdot f_X\left(\frac{y-d}{c}\right), \qquad -\infty < y < \infty .$$

◆

Bajo el *Teorema 1.5*:

$$g^{-1}(y) = \frac{y-d}{c} = x, \text{ pues } g(x) = cx + d = y.$$

Así:

$$\frac{d}{dy} g^{-1}(y) = \frac{1}{c} > 0, \qquad \text{si } c > 0 \text{ ya que entonces } y = g(x) \text{ es creciente}$$

y

$$-\frac{d}{dy} g^{-1}(y) = -\frac{1}{c} > 0, \quad \text{si } c < 0 \text{ ya que entonces } y = g(x) \text{ es decreciente.}$$

Por lo tanto,

$$\left|\frac{d}{dy} g^{-1}(y)\right| = \left|\frac{1}{c}\right| > 0, \text{ engloba ambos casos, y con ello efectivamente } f_Y(y) > 0.$$

◆◆

Resultado 1.7
Distribución de e^X
(aplicación del *Teorema 1.5*)

Sea X una variable aleatoria continua, $-\infty < x < \infty$, con función de densidad de probabilidad $f_Y(x)$, entonces la distribución de la exponencial de X, $Y = e^X$, está dada por

$$f_Y(y) = \frac{1}{y} f_X(\ln y), \quad y > 0. \qquad [1.7]$$

Demostración. Observe que la transformación $y = e^x$ es siempre creciente (para $-\infty < x < \infty$) y además es positiva $y > 0$. Calculando $F_Y(y)$:

$$\begin{aligned} F_Y(y) &= P[Y \le y] \\ &= P\left[e^X \le y\right] \\ &= P\left[X \le \ln y\right] \\ &= F_Y(\ln y) \end{aligned}$$

Derivando.

$$f_Y(y) = \frac{1}{y} f_X(\ln y), \quad y > 0$$

◊

Bajo el *Teorema 1.5*:

$$g^{-1}(y) = \ln y = x\text{, pues } g(x) = e^x$$

Así:

$$\frac{d}{dy} g^{-1}(y) = \frac{1}{y} > 0,\ \text{ ya que } y = g(x) \text{ es creciente.}$$

(Observe que como $y = e^x > 0$ entonces $\frac{d}{dy} g^{-1}(y) = \frac{1}{y}$ está bien definida.)

◊ ◊

Finalmente, dentro de las aplicaciones de la *Técnica de la transformación* para el caso univariado (*Teoremas 1.5* y *1.5bis*) se presenta el *Resultado 1.8* que contempla como un caso particular la transformación inversa asociada al *Resultado 1.7*.

✂ ***Resultado 1.8***

Distribución de $c \cdot \ln X$

(aplicación del *Teorema 1.5*)

Sea X una variable aleatoria continua con rango positivo, $x > 0$, y función de densidad de probabilidad $f_X(x)$, entonces la distribución de c veces el logaritmo de X, $Y = c \cdot \ln X$, donde $c \neq 0$ está dada por.

$$f_Y(y) = \frac{1}{c} e^{y/c} f_X\left(e^{y/c}\right), \qquad -\infty < y < \infty. \qquad [1.8]$$

Demostración. Observe que como $-\infty < \ln X < \infty$ y $c \neq 0$, entonces $-\infty < y = c \cdot \ln X < \infty$. Además, la transformación está bien definida porque $x > 0$; y es creciente o decreciente dependiendo del signo de c. Por ello es necesario considerar los dos casos. Calculando $F_Y(y)$.

(a) Caso $c > 0$. La transformación $y = c \cdot \ln x$ es creciente (porque $\ln x$ es creciente)

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= P[c \ln X \leq y] \\ &= P\left[\ln X \leq \frac{y}{c}\right] \qquad \text{(porque } c > 0) \\ &= P\left[X \leq e^{y/c}\right] \\ &= F_X(e^{y/c}) \end{aligned}$$

Derivando

$$f_Y(y) = \frac{1}{c} f_X\left(e^{y/c}\right), \quad -\infty < y < \infty.$$

(b) Caso $c < 0$. La transformación $y = c \cdot \ln x$ es decreciente (pues el coeficiente c invierte el sentido de $\ln x$ que es creciente).

$$\begin{aligned} F_Y(y) &= P[Y \leq y] \\ &= P[c \ln X \leq y] \\ &= P\left[\ln X \geq \frac{y}{c}\right] \qquad \text{(porque } c < 0) \\ &= P\left[X \geq e^{y/c}\right] = 1 - P\left[X \leq e^{y/c}\right] \\ &= 1 - F_X(e^{y/c}) \end{aligned}$$

Derivando:

$$f_Y(y) = -\left[\frac{1}{c}\right] f_X\left(e^{y/c}\right), \qquad -\infty < y < \infty .$$

◈

Bajo el *Teorema 1.5*

$$g^{-1}(y) = e^{y/c} = x, \text{ pues } g(x) = c \cdot \ln x$$

Así.

$$\frac{d}{dy} g^{-1}(y) = \frac{1}{c} > 0, \qquad \text{si } c > 0, \text{ ya que entonces } y = g(x) \text{ es creciente}$$

y $$-\frac{d}{dy} g^{-1}(y) = -\frac{1}{c} > 0, \qquad \text{si } c < 0, \text{ ya que entonces } y = g(x) \text{ es decreciente}$$

Por lo tanto:

$$\left|\frac{d}{dy} g^{-1}(y)\right| = \frac{1}{|c|} > 0,$$ engloba ambos casos, y con ello efectivamente $f_Y(y) > 0$.

◈ ◈

FUNCIONES O TRANSFORMACIONES DE MÚLTIPLES VARIABLES ALEATORIAS (VECTOR ALEATORIO)

Hasta este momento se ha abordado el caso de distribuciones de probabilidad de funciones de variables aleatorias univariadas Sin embargo, es común que se esté interesado en planteamientos probabilísticos que involucren dos (X y Y, por ejemplo) o más variables ($X_1, X_2, \ldots, X_k$, por ejemplo). Para tratar con tales probabilidades es necesario generalizar los conceptos y resultados definidos para el caso de una sola variable aleatoria. A ello se dedicará brevemente esta sección, comenzando con el concepto de *función de distribución de probabilidad acumulativa conjunta*. El lector que esté familiarizado con el manejo de vectores aleatorios puede sin problema pasar directamente a revisar las tres técnicas para determinar la distribución de una función de un vector aleatorio.

Variables aleatorias conjuntas y sus funciones de distribución y de densidad [masa]

Definición 1.7
F.d.a. conjunta

Sean $X_1, X_2, \ldots, X_k$ variables aleatorias definidas sobre el mismo espacio de probabilidad $(\Omega, \mathcal{A}, P[\cdot])$. La *función de distribución acumulativa conjunta* de $X_1, \ldots, X_k$ denotada por $F_{X_1,\ldots,X_k}(\cdot,\ldots,\cdot)$ se define como $P[X_1 \leq x_1, \ldots, X_k \leq x_k]$ para todo $(x_1, \ldots, x_k)$.

◆

Es decir, la función de distribución acumulativa conjunta es una función con dominio el espacio euclidiano de dimensión k y contradominio el intervalo $[0,1]$.

Al igual que la función de distribución acumulativa de una variable aleatoria unidimensional, la acumulativa conjunta cumple ciertas propiedades, las cuales pueden generalizarse al caso k-dimensional a partir de lo que sucede para dos variables (véase ANEXO 3).

Definición 1.8
V.a.'s conjuntas discretas

Se dice que el vector $(X_1, X_2, \ldots, X_k)$ es un vector aleatorio discreto (o *variable aleatoria k-dimensional discreta*) si puede tomar valores sólo en un número numerable de puntos en el espacio real de dimensión k. Se dice también que las variables aleatorias conjuntas discretas $X_1, X_2, \ldots, X_k$ son *variables aleatorias conjuntas discretas.*

◆

Definición 1.9
Función de densidad discreta conjunta

Si $(X_1, \ldots, X_k)$ es una variable aleatoria k-dimensional discreta, entonces la función de densidad discreta conjunta de $(X_1, \ldots, X_k)$, denotada por $f_{X_1,\ldots,X_k}(\cdot,\ldots,\cdot)$, se define como

$$f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = P[X_1 = x_1, \ldots, X_k = x_k]$$

para $(x_1, \ldots, x_k)$ un valor de $(X_1, \ldots, X_k)$, y se define como 0 en cualquier otro caso.

- *Observación* $f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) \geq 0$ y $\sum f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = 1$, donde la suma es sobre todos los posibles valores de $(X_1, \ldots, X_k)$

◊

Definición 1.10
Función de densidad marginal discreta

Si $X_{i1},\ldots, X_{im}$ es cualquier subconjunto de las variables aleatorias discretas conjuntas $X_1,\ldots, X_k$, entonces a $f_{X_{i1},\ldots,X_{im}}(x_{i1},\ldots,x_{im})$ se le llama *función de densidad marginal discreta*, que puede ser obtenida a partir de la densidad conjunta.

◊

Caso bivariado. si X y Y (en lugar de utilizar la simbología X_1 y X_2) son variables aleatorias discretas conjuntas con valores (x_1,y_1), $(x_2,y_2),\ldots$, entonces $f_X(\cdot)$ y $f_Y(\cdot)$ son las densidades marginales, calculadas como

$$f_X(x) = \sum_{y_i:\ f_{X,Y}(x,y_i)>0} f_{X,Y}(x,y_i) \qquad \text{y} \qquad f_Y(y) = \sum_{x_i:\ f_{X,Y}(x_i,y)>0} f_{X,Y}(x_i,y)$$

Definición 1.11
V.a.'s continuas conjuntas
y su función de densidad

Se dice que el vector $(X_1, X_2,\ldots, X_k)$ es una *vector aleatorio continuo* (*variable aleatoria k-dimensional continua*) si y sólo si existe una función $f_{X_1,\ldots,X_k}(\cdot,\ldots,\cdot) \geq 0$ tal que

$$F_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = \int_{-\infty}^{x_k} \cdots \int_{-\infty}^{x_1} f_{X_1,\ldots,X_k}(u_1,\ldots,u_k)du_1\ldots du_k$$

para todo $(x_1, \ldots, x_k)$. A $f_{X_1,\ldots,X_k}(\cdot,\ldots,\cdot)$ se le define como una *función de densidad de probabilidad conjunta*.

Observación. $f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) \geq 0$ y $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_k}(x_1,\ldots,x_k)dx_1\ldots dx_k = 1$.

◊

La utilidad principal de la función de densidad unidimensional es la de permitir el cálculo de probabilidades. Por ejemplo, para una variable aleatoria continua X con función de densidad $f_X(\cdot)$, $P[a < X < b] = \int_a^b f_X(x)dx$ representa el *área* bajo $f_X(\cdot)$ sobre el intervalo (a,b). En el caso *bidimensional* ($k = 2$) el *volumen* dará las probabilidades.

Definición 1.12
Función de densidad marginal continua

Si $X_{i1},\ldots, X_{im}$ es cualquier subconjunto de las variables aleatorias continuas conjuntas $X_1,\ldots, X_k$, entonces a $f_{X_{i1},\ldots,X_{im}}(x_{i1},\ldots,x_{im})$ se le llama *función de densidad marginal* de la variable aleatoria m-dimensional $(X_{i1},\ldots, X_{im})$.

♦

Caso bivariado. Por ejemplo, si X y Y son variables aleatorias continuas conjuntas, entonces $f_X(\cdot)$ y $f_Y(\cdot)$ son las densidades marginales, que pueden obtenerse como

$$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \qquad \text{y} \qquad f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$

puesto que

$$f_X(x) = \frac{dF_X(x)}{dx} = \frac{d}{dx}\left[\int_{-\infty}^{x}\left(\int_{-\infty}^{\infty} f_{X,Y}(u,y)\,dy\right)du\right] = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\,dy$$

y de manera análoga para $f_Y(y)$.

Con las definiciones anteriores se puede plantear el concepto de independencia entre variables aleatorias, el cual será de suma utilidad al tratar posteriormente con funciones de múltiples variables aleatorias, como se verá en el Capítulo 3.

Variables aleatorias independientes

Definición 1.13
Independencia estocástica

Sea $(X_1, X_2, \ldots, X_k)$ una variable aleatoria k-dimensional con función de densidad conjunta $f_{X_1,\ldots,X_k}(\cdot,\ldots,\cdot)$. Se dice que $X_1,\ldots,X_k$ son variables aleatorias *estocásticamente independientes* (o, simplemente, *variables aleatorias independientes, v.a.i.'s*) si y sólo si

$$f_{X_1,\ldots,X_k}(x_1,\ldots,x_k)=\prod_{i=1}^{k} f_{X_i}(x_i)$$

para todos los valores $(x_1,\ldots,x_k)$ de $(X_1,\ldots,X_k)$.

◈

Esta definición expone una forma sencilla de calcular la función de densidad conjunta de múltiples variables aleatorias, cuando éstas son independientes, como el *producto de sus densidades marginales*. El siguiente teorema hace extensiva la propiedad de independencia de variables aleatorias a funciones de ellas y será clave en el Capítulo 3

Teorema 1.6
Independencia de funciones de v.a.i.'s[17]

Si $X_1,\ldots,X_k$ son variables aleatorias independientes y $g_1(\cdot),\ldots,g_k(\cdot)$ son k funciones tales que $Y_j=g_j(X_j)$, $j=1,\ldots,k$ son variables aleatorias, entonces $Y_1,\ldots,Y_k$ también son independientes.

◈

Funciones de múltiples variables aleatorias

En esta sección se plantea el caso de funciones que dependen de más de una variable aleatoria. Al igual que en la sección previa dedicada a funciones de una sola variable aleatoria, se presentarán técnicas para determinar la distribución de esas funciones multivariadas. Se expone la generalización de la *técnica de la función de distribución* y la generalización de la *técnica de la*

[17] Para la demostración de este teorema véase ANEXO 3

transformación. Además, se presenta el concepto de "esperanza" para posteriormente incorporar una tercera técnica: la *técnica de la función generadora de momentos*.

Ejemplo 1.12 (*ejemplo introductorio de una función de múltiples variables aleatorias y de la necesidad de determinar su distribución*). Uno de los objetivos de la estadística es hacer inferencia acerca de una población. De manera muy breve, para ello se toma como base la información contenida en una muestra tomada de esa población y con ella se construye una medida de aproximación a alguna característica de interés, obteniendo después otra medida de la bondad de la inferencia. Las estadísticas utilizadas para estimar los parámetros de una población o para tomar decisiones con respecto a ella son funciones de las observaciones aleatorias que se presentan en una muestra.

Considere que se desea estimar la media de una población, μ. Para ello se extrae una muestra aleatoria de n observaciones $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$.[18] Intuitivamente, se puede proponer la media muestral

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

como una estimación de μ Interesa saber ahora qué tan buena es esta estimación. Esto depende del comportamiento de las variables aleatorias $X_1, X_2, \ldots X_n$, ya que éste tendrá un efecto sobre la nueva variable aleatoria $Y = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.

La medida de la bondad de una estimación es el *error de estimación*, la diferencia entre la estimación y el parámetro estimado; en este caso $\bar{x} - \mu = y - \mu$. Como Y es una variable aleatoria no puede asegurarse que el error en la estimación sea menor que un valor específico, digamos δ. Sin embargo, si es factible determinar la distribución de probabilidad del estimador Y se podrá utilizar esta distribución para determinar la probabilidad de que el error en la estimación sea menor o igual que δ.

[18] Para una breve referencia sobre cómo puede obtenerse una muestra aleatoria puede consultarse ANEXO 4.

Para determinar la distribución de probabilidad de una función de n variables aleatorias, $X_1, X_2, \ldots X_n$ debe encontrarse primero la distribución de probabilidad conjunta para ese vector aleatorio. Bajo el supuesto de que las variables aleatorias $X_1, X_2, \ldots, X_n$ obtenidas a través de una muestra aleatoria son *independientes* entre sí, la función de densidad [masa] conjunta está dada por el producto de las funciones de densidad [masa] respectivas $f_{X_i}(x_i)$

Así, la función de densidad [masa] de probabilidad conjunta para $X_1, X_2, \ldots, X_n$ es

$$f_{X_1,X_2,\ldots X_n}(x_1,x_2,\ldots,x_n) = \prod_{i=1}^{n} f_{X_i}(x_i).$$

◈

DISTRIBUCIÓN DE UNA FUNCIÓN MULTIVARIADA DE VARIABLES ALEATORIAS

En general, considere que se tienen $X_1, X_2, \ldots X_n$, variables aleatorias, y $g_1(\cdot,\ldots,\cdot), g_2(\cdot,\ldots,\cdot),\ldots,g_k(\cdot,\ldots,\cdot)$, funciones de esas n variables aleatorias. Se quiere encontrar la distribución conjunta de $Y_1, Y_2,\ldots, Y_n$, donde $Y_j = g_j(X_1,\ldots,X_n)$, $j = 1, 2,\ldots, k$. Si la densidad conjunta de las variables aleatorias $X_1, X_2,\ldots, X_n$ está dada, entonces –en teoría– es posible encontrar la distribución conjunta de $Y_1, Y_2,\ldots, Y_n$. Esto se sigue de que la función de distribución acumulativa conjunta de $Y_1, Y_2,\ldots, Y_n$ satisface que

$$\begin{aligned} F_{Y_1,\ldots,Y_k}(y_1,\ldots,y_k) &= P[Y_1 \le y_1,\ldots,Y_k \le y_k] \\ &= P[g_1(X_1,\ldots,X_n) \le y_1,\ldots,g_k(X_1,\ldots,X_n) \le y_k] \end{aligned}$$

para $y_1, y_2,\ldots,y_k$ fijos, que es la probabilidad de un evento descrita en términos de $X_1, X_2,\ldots, X_n$, y teóricamente tal probabilidad puede ser determinada integrando o sumando la densidad conjunta sobre la región que corresponda al evento. El problema es que en general no puede evaluarse fácilmente la probabilidad deseada para cada $y_1, y_2,\ldots,y_k$. Las tres técnicas que se presentan a continuación pueden contrarrestar esta dificultad.

☞ ***La técnica de la función de distribución: generalización***

Si la función de distribución conjunta de las variables aleatorias $X_1, X_2, \ldots, X_n$ está dada, entonces, teóricamente, la distribución conjunta de las variables aleatorias $Y_1, Y_2, \ldots, Y_k$ puede ser determinada, donde $Y_j = g_j(X_1, \ldots X_n)$, $j = 1, 2, \ldots, k$, para funciones dadas $g_1(\cdot, \ldots, \cdot), \ldots, g_k(\cdot, \ldots, \cdot)$. Por definición, la función de distribución *acumulativa conjunta* de $Y_1, Y_2, \ldots, Y_k$ es $F_{Y_1, \ldots, Y_k}(y_1, \ldots, y_k) = P[Y_1 \le y_1; \ldots; Y_k \le y_k]$. Pero para cada $y_1, \ldots, y_k$ el evento $\{Y_1 \le y_1; \ldots; Y_k \le y_k\} \equiv \{g_1(X_1, \ldots, X_n) \le y_1; \ldots; g_k(X_1, \ldots, X_n) \le y_k\}$ Este último evento está descrito en términos de las funciones dadas $g_1(\cdot, \ldots, \cdot), \ldots, g_k(\cdot, \ldots, \cdot)$ y de las variables aleatorias dadas $X_1, X_2, \ldots, X_n$. Como se está suponiendo que se conoce la distribución conjunta de $X_1, X_2, \ldots, X_n$, entonces la probabilidad del evento $\{g_1(X_1, \ldots, X_n) \le y_1; \ldots; g_k(X_1, \ldots, X_n) \le y_k\}$ puede ser calculada y en consecuencia determinar $F_{Y_1, \ldots Y_k}(\cdot, \ldots, \cdot)$. El método descrito anteriormente para obtener la distribución conjunta de $Y_1, Y_2, \ldots, Y_k$ se denomina la *técnica de la función de distribución*.

☜

La aplicación de la *técnica de la función de distribución* permite obtener la distribución de la *suma* y la *diferencia* de *dos* variables *aleatorias, de su producto y su cociente*, así como de la *mínima* y la *máxima estadística de orden*, como se mostrará a continuación. Si bien en el Capítulo 3 no se retomarán todos estos resultados, se considera conveniente presentarlos aquí no sólo como ejemplos de aplicación de la técnica de la función de distribución, sino también como elementos que pudieran *ser útiles para verificar resultados futuros* dentro de la naturaleza de este trabajo.

Resultados 1.9 y 1.10
Distribución de $X+Y$ y de $X-Y$[19]
(aplicación de la *Técnica de la función de distribución*)

Sean X y Y v.a. con función de densidad conjunta continua $f_{X,Y}(x,y)$. Defínase $Z=X+Y$ y $V=X-Y$. Entonces,

$$f_Z(z)=\int_{-\infty}^{\infty} f_{X,Y}(x,z-x)\,dx=\int_{-\infty}^{\infty} f_{X,Y}(z-y,y)\,dy, \qquad [1.9]$$

y

$$f_V(v)=\int_{-\infty}^{\infty} f_{X,Y}(x,x-v)\,dx=\int_{-\infty}^{\infty} f_{X,Y}(v+y,y)\,dy. \qquad [1.10]$$

Demostración Se probará [1.9]

$$F_Z(z)=P[Z\le z]=P[X+Y\le z]=\iint\limits_{x+y\le z} f_{X,Y}(x,y)\,dx\,dy$$

$$=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{z-x} f_{X,Y}(x,y)\,dy\right]dx$$

sustituyendo $y=u-x$

$$=\int_{-\infty}^{\infty}\left[\int_{-\infty}^{z} f_{X,Y}(x,u-x)\,du\right]dx$$

Aplicando el Teorema 1.2 :

$$f_Z(z)=\frac{dF_Z(z)}{dz}=\frac{d}{dz}\left\{\int_{-\infty}^{z}\left[\int_{-\infty}^{\infty} f_{X,Y}(x,u-x)\,dx\right]du\right\}$$

$$=\int_{-\infty}^{\infty} f_{X,Y}(x,z-x)\,dx.$$

Análogamente para probar [1.10].

◊

Resultado 1.11 (corolario de 1.9)
Distribución de $X+Y$[20]
(aplicación de la *Técnica de la función de distribución*)

Sean X y Y v.a.i. continuas y $Z=X+Y$. Entonces,

$$f_Z(z)=f_{X+Y}(z)=\int_{-\infty}^{\infty} f_Y(z-x)f_X(x)\,dx \qquad [1.11a]$$

$$=\int_{-\infty}^{\infty} f_X(z-y)f_Y(y)\,dy. \qquad [1.11b]$$

Demostración. La demostración se sigue inmediatamente del *Resultado 1.9* y de la independencia de X y de Y (véase *Definición 1.13* y *Teorema 1.6*) ◊

---

[19] Tomado de Mood et al. [1974], p 185.

[20] Tomado de Mood *et al.* [1974, p 186, donde además se menciona que en análisis matemático a la función $f_Z(\cdot)$ se le llama la *convolución* de las funciones $f_X(\,)$ y $f_Y(\cdot)$ Este corolario será particularmente importante en el Capítulo 3

Resultados 1.12 y 1.13
Distribución de XY y de X/Y[21]

Sean X y Y variables aleatorias con función de densidad conjunta continua $f_{X,Y}(x,y)$, y defínase $Z = XY$ y $U = X/Y$. Entonces,

- $$f_Z(z) = \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{z}{x}\right) dx = \int_{-\infty}^{\infty} \frac{1}{|y|} f_{X,Y}\left(\frac{z}{y}, y\right) dy, \qquad [1.12]$$

y

- $$f_U(u) = \int_{-\infty}^{\infty} |y| f_{X,Y}(uy, y)\, dy. \qquad [1.13]$$

Demostración. Se probará la *primera parte* del *Resultado 1.12*.

$$
\begin{aligned}
F_Z(z) = P[Z \le z] &= \iint_{xy \le z} f_{X,Y}(x,y)\, dx\, dy \\
&= \int_{-\infty}^{0} \left[\int_{z/x}^{\infty} f_{X,Y}(x,y)\, dy\right] dx + \int_{0}^{\infty} \left[\int_{-\infty}^{z/x} f_{X,Y}(x,y)\, dy\right] dx,
\end{aligned}
$$

sustituyendo $u=xy$

$$
\begin{aligned}
&= \int_{-\infty}^{0} \left[\int_{z}^{\infty} f_{X,Y}\left(x, \frac{u}{x}\right) \frac{du}{x}\right] dx + \int_{0}^{\infty} \left[\int_{-\infty}^{z} f_{X,Y}\left(x, \frac{u}{x}\right) \frac{du}{x}\right] dx \\
&= \int_{-\infty}^{z} \left[\int_{-\infty}^{0} \frac{1}{-x} f_{X,Y}\left(x, \frac{u}{x}\right) dx\right] du + \int_{-\infty}^{z} \left[\int_{0}^{\infty} \frac{1}{x} f_{X,Y}\left(x, \frac{u}{x}\right) dx\right] du \\
&= \int_{-\infty}^{z} \left[\int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{u}{x}\right) dx\right] du;
\end{aligned}
$$

Aplicando el Teorema 1.2:

$$
\begin{aligned}
f_Z(z) &= \frac{dF_Z(z)}{dz} \\
&= \int_{-\infty}^{\infty} \frac{1}{|x|} f_{X,Y}\left(x, \frac{z}{x}\right) dx.
\end{aligned}
$$

◆

[21] Tomado de Mood *et al.* [1974], pp. 187-188

Distribución de las estadísticas de orden [22]

Resultados 1.14 y 1.15
Distribución (f.d.a.) de la máxima
y de la mínima estadística de orden
para el caso discreto o continuo
(aplicación de la *Técnica de la función de distribución*)

Sean $X_1,\ldots,X_n$ v.a.i.i.d. Defínase $Y_n = \text{máx}\,[X_1,\ldots,X_n]$ y $Y_1 = \text{mín}\,[X_1,\ldots,X_n]$ Entonces, las funciones de distribución de la máxima y de la mínima estadística de orden, respectivamente, están dadas por

$$F_{Y_n}(y) = [F_X(y)]^n \qquad [1.14]$$

y

$$F_{Y_1}(y) = 1 - [1 - F_X(y)]^n \qquad [1.15]$$

Demostración. Se seguirá un razonamiento que apele a cuál es el significado de la mínima y de la máxima estadísticas de orden dentro de la muestra aleatoria.

(a) Función de distribución de Y_n.

$$F_{Y_n}(y) = P[Y_n \le y] = P[X_1 \le y, \ldots, X_n \le y\,],$$

donde la última igualdad se sigue de que la *más grande* de las X_i, $i = 1,\ldots, n$, es *menor o igual que* y si y sólo si *todas* las X_i son *menores o iguales* que y. Puesto que integran una muestra aleatoria, las X_i son *independientes*, y entonces

$$P[X_1 \le y, \ldots, X_n \le y\,] = \prod_{i=1}^{n} P[X_i \le y] = \prod_{i=1}^{n} F_{X_i}(y) = \prod_{i=1}^{n} F_X(y)$$

donde la última igualdad se desprende de que $X_1, \ldots, X_n$ tienen la *misma* distribución. $F_X(x)$ (nuevamente, por tratarse de una muestra aleatoria). Por lo tanto

$$F_{Y_r}(y) = [F_X(y)]^n \,.^{23}$$

[22] Para una breve exposición sobre las estadísticas de orden consúltese ANEXO 4 Se presenta también la f d a y la f d.p (f.m.p) de la j-ésima estadística de orden, tanto para el caso discreto como continuo

De manera similar se procede para determinar la función de distribución de Y_1.

(b) Función de distribución de Y_1.

$$F_{Y_1}(y) = P[Y_1 \le y] = 1 - P[Y_1 > y] = 1 - P[X_1 > y, \ldots, X_n > y]$$

ya que la *más pequeña* de las X_i, $i = 1, \ldots, n$, es *mayor* que y si y sólo si *todas* las X_i son *mayores* que y Aplicando el supuesto de independencia de las X_i y de la homogeneidad de la distribución, se obtienen las siguientes equivalencias

$$1 - P[X_1 > y, \ldots, X_n > y] = 1 - \prod_{i=1}^{n} P[X_i > y] = 1 - \prod_{i=1}^{n} [1 - F_{X_i}(y)] = 1 - \prod_{i=1}^{n} [1 - F_X(y)]$$

Por lo tanto,

$$F_{Y_1}(y) = 1 - [1 - F_X(y)]^n .$$ [24]

◆

Los dos resultados siguientes son corolarios de [*1.14*] y [*1.15*] para variables aleatorias continuas. Su demostración se obtiene fácilmente aplicando el hecho de que la derivada de la función de distribución da como resultado la función de densidad (*Teorema 1.2*).

✂ ***Resultados 1.16 y 1.17*** (*corolarios de* ***1.14*** *y* ***1.15***)
Distribución (f.d.p) de la máxima y de la mínima estadísticas de orden para el caso continuo
(aplicación de la *Técnica de la función de distribución*)

Sean $X_1, \ldots, X_n$ v.a.i.i.d. *continuas* Entonces las funciones de densidad de la máxima y de la mínima estadística de orden, respectivamente, están dadas por

$$f_{Y_n}(y) = n\,[F_X(y)]^{n-1} f_X(y) \qquad [1.16]$$

y

$$f_{Y_1}(y) = n\,[1 - F_X(y)]^{n-1} f_X(y) \qquad [1.17].$$

◆

[23] Observe que *si considera que* $X_1, \ldots, X_n$ *es un conjunto* de variables aleatorias con distribución distinta, digamos $F_{X_i}(x)$, hay que limitarse a decir que la f.d.a puede expresarse como producto de las f.d.a. marginales de las X_i. Si además se descarta el supuesto de independencia, la f.d.a. no puede expresarse siquiera en términos de un producto.

[24] Salvedades similares a las señaladas para determinar la f.d.a. de Y_n pueden hacerse también para el caso de Y_1.

La técnica de la transformación: caso bivariado

A continuación se presenta la técnica de la transformación para determinar la distribución conjunta de *dos* funciones de *dos* variables aleatorias y, particularmente, de dos variables aleatorias independientes. El caso de dos funciones bivariadas es el mayor nivel de generalización que se estudiará aquí, no sólo porque con él se satisfacen los requerimientos del Capítulo 3, sino también porque la complejidad que involucra el extender la técnica a n funciones de n variables aleatorias escapa a las pretensiones de este trabajo. Una auténtica generalización de la técnica de la transformación puede consultarse en Casella y Berger [1990] y en Mood *et al* [1974] [25]

Teoremas 1.7 y 1.8
La técnica de la transformación
para el caso bivariado (continuo)

Sean X_1 y X_2 dos variables aleatorias continuas conjuntas con función de densidad de probabilidad conjunta $f_{X_1,X_2}(x_1,x_2)$. Considérense Y_1 y Y_2 dos variables aleatorias tales que $Y_1 = g_1(X_1,X_2)$ y $Y_2 = g_2(X_1,X_2)$. El dominio y la imagen de las funciones g_1 y g_2 son ϑ y Ψ, respectivamente, donde

$$\vartheta = \{(x_1, x_2)\colon f_{X_1,X_2}(x_1,x_2) > 0\}$$

y $$\Psi = \{(y_1,y_2) \text{ para los cuales existe } (x_1,x_2) \in \vartheta, \text{ tal que } (y_1,y_2) = (g_1(x_1,x_2),g_2(x_1,x_2))\}.$$

Además, las funciones g_1 y g_2 satisfacen las condiciones siguientes

(a) Las ecuaciones $y_1 = g_1(x_1,x_2)$ y $y_2 = g_2(x_1,x_2)$ pueden resolverse de manera única para x_1 y x_2 en términos de y_1 y y_2, estando sus soluciones dadas por $x_1 = g_1^{-1}(y_1,y_2)$ y $x_2 = g_2^{-1}(y_1,y_2)$ (es decir, definen una transformación uno a uno que va de ϑ a Ψ).

---

[25] Véase Casella y Berger [1990], *op cit*, pp. 176–178; Mood *et al.* [1974], pp. 211–212. Además, estos últimos autores discuten la situación en que alguna de las transformaciones g_i no sea uno a uno (pp 209–210)

(b) Las primeras derivadas parciales de $x_1 = g_1^{-1}(y_1, y_2)$ y $x_2 = g_2^{-1}(y_1, y_2)$ son continuas en todos los puntos $(x_1, x_2) \in \vartheta$.

(c) El *Jacobiano* de la transformación, denotado por J, y definido como el determinante siguiente

$$J = \begin{vmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{vmatrix}$$

es *distinto* de cero para todo $(y_1, y_2) \in \Psi$.

Entonces la densidad conjunta de Y_1 y Y_2 está dada por

$$f_{Y_1,Y_2}(y_1, y_2) = \begin{cases} J\, f_{X_1,X_2}(g_1^{-1}(y_1,y_2), g_2^{-1}(y_1,y_2)) & \text{si} \quad (y_1,y_2) \in \Psi \\ 0 & \text{e.o.c.} \end{cases} \qquad [T.1.7]$$

lo que en términos de funciones indicadoras equivale a decir que.

$$f_{Y_1,Y_2}(y_1, y_2) = J\; f_{X_1,X_2}(g_1^{-1}(y_1,y_2), g_2^{-1}(y_1,y_2))\, I_\Psi(y_1, y_2) \qquad [T\,1.7\,]^{26,27}$$

Si además se incorpora el supuesto de que X_1 y X_2 son independientes, entonces (aplicando la *Definición 1.13* y el *Teorema 1.6*) la expresión [*T.1.7*'] queda como

$$f_{Y_1,Y_2}(y_1, y_2) = J \cdot f_{X_1}(g_1^{-1}(y_1,y_2)) f_{X_2}(g_2^{-1}(y_1,y_2))\, I_\Psi(y_1, y_2) \qquad [T\,1.8]$$

---

[26] Observe que $I_\Psi(y_1,y_2) = I_\vartheta(g_1^{-1}(y_1,y_2), g_2^{-1}(y_1,y_2))$

[27] La versión de otros autores como Ross [1988] para este teorema considera al Jacobiano como el determinante siguiente:

$$J = \begin{vmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{vmatrix},$$

de tal suerte que la expresión [1.7'] queda como $f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{J} f_{X_1,X_2}(g_1^{-1}(y_1,y_2), g_2^{-1}(y_1,y_2))\, I_\Psi(y_1,y_2)$, o simplificando $f_{Y_1,Y_2}(y_1,y_2) = \frac{1}{J} f_{X_1,X_2}(x_1,x_2)\, I_\Psi(y_1,y_2)$; finalmente los resultados son equivalentes [véase Ross [1988], p. 229]. (Tanto Mood *et al.* (p. 205) como Ross (p. 230) presentan un esbozo de la demostración de este resultado, mencionando que se requiere del manejo de integrales dobles.)

A continuación se presentará un ejemplo de aplicación de la técnica de la transformación que se acaba de presentar Múltiples aplicaciones adicionales de éste se presentarán en el Capítulo 3, donde la expresión *[T.1.8]* será de suma utilidad

Ejemplo 1.13 *(aplicación de la técnica de la transformación, caso continuo bivariado)*. Sean X_1 y X_2 dos variables aleatorias independientes, cada una con distribución de probabilidad dada por

$$f_{X_i}(x) = I_{(0,1)}(x) = \begin{cases} 1 & \text{si} \quad x \in (0,1) \\ 0 & \text{e.o.c} \end{cases}$$

Entonces,

$$f_{X_1,X_2}(x_1,x_2) = I_{(0,1)}(x_1) \cdot I_{(0,1)}(x_2),$$

$$\vartheta = \{(x_1, x_2) \colon 0 < x_1 < 1,\ 0 < x_2 < 1\}.$$

Sean $y_1 = g_1(x_1,x_2) = x_1 + x_2$ y $y_2 = g_2(x_1,x_2) = x_1 - x_2$; entonces

$$x_1 = \tfrac{1}{2}(y_1 + y_2) = g_1^{-1}(y_1,y_2)$$
$$x_2 = \tfrac{1}{2}(y_1 - y_2) = g_2^{-1}(y_1,y_2)$$

Calculando el Jacobiano

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} \dfrac{1}{2} & \dfrac{1}{2} \\ \dfrac{1}{2} & -\dfrac{1}{2} \end{vmatrix} = -\frac{1}{2}$$

ϑ y Ψ se bosquejan en la *Figura 2*.

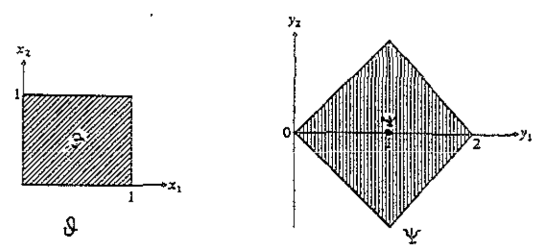


Figura 2

Como la transformación es uno a uno, las primeras derivadas parciales de g_1^{-1} y g_2^{-1} son continuas, y el Jacobiano es distinto de cero:

$$
\begin{aligned}
f_{Y_1,Y_2}(y_1,y_2) &= J\, f_{X_1,X_2}(g_1^{-1}(y_1,y_2), g_2^{-1}(y_1,y_2)) \\
&= \tfrac{1}{2} I_{(0,1)}\left(\frac{y_1+y_2}{2}\right) I_{(0,1)}\left(\frac{y_1-y_2}{2}\right) \\
&= \begin{cases} \tfrac{1}{2} & \text{para} \quad (y_1,y_2)\in\Psi \\ 0 & \text{e.o.c.} \end{cases}
\end{aligned}
$$

♦

Finalmente se presenta a continuación la tercera técnica para determinar la distribución de una función de variables aleatorias: la *técnica de la función generadora de momentos*.

La función generadora de momentos

Una función que está asociada con la distribución de probabilidad es la *función generadora de momentos*, que como su nombre lo indica puede ser utilizada para generar momentos. Sin embargo, su uso principal es el de ser una herramienta que permite caracterizar una distribución.

Definición 1.14[28]
Función generadora de momentos

Si X es una variable aleatoria con función de distribución F_X, entonces la *función generadora de momentos* (*f.g m.*) de X (o de F_X), denotada por $M_X(t)$, se define como

$$M_X(t) = E[e^{tX}],$$

si la esperanza existe para todo valor de t en algún intervalo $-h < t < h,\ h > 0$.

[28] Para abordar el concepto de "función generadora de momentos" se requiere comprender primero los conceptos de "*esperanza de una variable aleatoria*" X ($E[X]$) y de "momento", los cuales se tratan brevemente en el ANEXO 5 Por otra parte, en el ANEXO 7 se presenta una tabla (basada en Mood *et al.* [1974]) que contiene las expresiones de las funciones generadoras de momentos correspondientes a la gran mayoria de las distribuciones que se presentan en el esquema de Leemis, siempre y cuando éstas existan (siendo en tal caso únicas, como se verá más adelante) o sean de utilidad para los fines de este trabajo

De manera más explícita, la f.g.m de X puede escribirse como

$$M_X(t) = \sum_x e^{tx} P[X = x] \qquad \text{si } X \text{ es discreta,}$$

y

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx \qquad \text{si } X \text{ es continua}$$ [29]

◇

El teorema siguiente expone la forma como $M_X(t)$ genera momentos.

Teorema 1.9
Generación de momentos

Si X es una variable aleatoria con función generadora de momentos $M_X(t)$, entonces

$$E[X^r] = M_X^{(r)}(0),$$

donde

$$M_X^{(r)}(0) = \frac{d^r}{dt} M_X(t)\Big|_{t=0}.$$

Es decir, el r-ésimo momento es igual a la r-ésima derivada de $M_X(t)$ evaluada en $t = 0$

◇

- *Observación* Si la función generadora de momentos existe entonces caracteriza un conjunto infinito de momentos. Sin embargo, el caracterizar el conjunto de momentos de una variable aleatoria no es suficiente para determinar su distribución de manera única, ya que puede haber dos variables aleatorias distintas con los mismos momentos.[30]

Sólo cuando el soporte de la función de distribución es acotado, la secuencia infinita de momentos determina de manera única la distribución. Aún más, si la función generadora de momentos existe en una vecindad del cero, la distribución está determinada de manera *única*, sin importar cuál sea su soporte. Por lo tanto, la existencia de todos los momentos no equivale a la existencia de la función generadora de momentos.

---

[29] En el ANEXO 7 se presenta una tabla (basada en Mood *et al* [1974]) que contiene las expresiones de las funciones generadoras de momentos correspondientes a la gran mayoría de las distribuciones que se presentan en el esquema de Leemis, siempre y cuando éstas existan (siendo únicas, como se verá más adelante) o sean de utilidad para los fines de este trabajo.

[30] Casella y Berger [1990] ilustran esto considerando $f_1(x) = \frac{1}{\sqrt{2\pi}x} e^{-(\ln x)^2/2}, 0 \leq x < \infty$ y $f_2(x) = f_1(x)[1 + \text{sen}(2\pi \ln x)], 0 \leq x < \infty$.

En la práctica, en muchos casos es más sencillo calcular los momentos de manera directa que mediante la función generadora de momentos. De hecho, el uso principal de esta función no es la generación de momentos, sino la caracterización de una distribución. Esta propiedad puede conducir a resultados *sumamente poderosos*. El *siguiente* teorema muestra cómo puede ser caracterizada una distribución a través de la función generadora de momentos.

Teorema 1.10
Caracterización de una distribución
mediante la f.g.m.

Sean $F_X(x)$ y $F_Y(y)$ dos funciones de distribución para las cuales existen todos sus momentos.

(*a*) Si F_X y F_Y tienen soporte acotado, entonces $F_X(u) = F_Y(u)$ para toda u si y sólo si $E[X^r] = E[Y^r]$ para todo entero no negativo r.

(*b*) Si las funciones generadoras de momentos existen y $M_X(t) = M_Y(t)$ para toda t en el intervalo $-h < t < h$; $h > 0$, entonces $F_X(u) = F_Y(u)$ para toda u.

♦

Aunque la demostración de este teorema no se presentará, es importante insistir en la utilidad de la parte (*b*): si es posible encontrar la función generadora de momentos de una variable aleatoria entonces, teóricamente, se puede encontrar la distribución de esa variable aleatoria, ya que existe una única función de distribución para cada función generadora de momentos dada.[31]

Así como lo fue el *Teorema 1.2* para las técnicas de la función de distribución y de la transformación, el *Teorema 1.10* es el corazón de la técnica de la función generadora de momentos. Es la unicidad de las funciones de distribución, de densidad [*masa*] de probabilidad y de

[31] Casella y Berger [1990] presentan una discusión acerca de la demostración del *Teorema 1.10*, aclarando que no la presentan porque no proporciona mayor enseñanza, pues se trata de una cuestión técnica. Mencionan que se apoya en la teoría de las transformadas de Laplace. La transformada de Laplace se define como $\int_{-\infty}^{\infty} e^{tx} f_X(x)dx$, que es precisamente la definición de $M_X(t)$ (y se dice que $M_X(t)$ es la transformada de Laplace de $f_X(x)$). *Un hecho clave* sobre las transformadas de Laplace es su unicidad. Si $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)dx$ para toda t tal que $|t| < h$, donde h es algún número positivo, entonces dada $M_X(t)$ existe sólo una función $f_X(x)$ que satisface esa ecuación, lo cual hace que el *Teorema 1.10* sea razonable. (Lo anterior es aplicable también a la demostración del teorema que establece la convergencia de las funciones generadoras de momentos, mismo que se presenta en el libro después del *Teorema 1.10*.) Véase Casella y Berger [1990], *op. cit.*, pp. 65–66.

la generadora de momentos lo que hace válido recurrir a ellas para definir o determinar la distribución de una variable aleatoria.

Ahora sólo se requiere extender las nociones de esperanza y de función generadora de momentos a más de una variable para plantear formalmente la técnica de la función generadora de momentos

Definición 1.15
F.g.m. de un vector aleatorio

La función generadora de momentos de un vector aleatorio $(X_1,\ldots, X_k)$ se define como

$$m_{X_1,\ldots,X_k}(t_1,\ldots,t_n) = E\left[\exp\sum_{j=1}^{k} t_j X_j\right]$$

si la esperanza existe para todos los valores de $t_1,\ldots, t_k$ tales que $|t_j| < h$. para alguna $h > 0$, $j = 1,\ldots, k$

◈

Teorema 1.11 [32]
Esperanza de dos funciones de v.a.i.'s

Sean X y Y dos variables aleatorias independientes y $g_1(\cdot)$ y $g_2(\cdot)$ dos funciones univariadas, entonces

$$E[g_1(X)\, g_2(Y)] = E[g_1(X)]\cdot E[g_2(Y)].$$

◈

Teorema 1.12 [33]
Independencia y f.g.m. de dos v.a.'s conjuntas

Dos variables aleatorias conjuntamente distribuidas X y Y son independientes si y sólo si

$$m_{X,Y}(t_1, t_2) = m_X(t_1)\, m_Y(t_2),$$

para todo t_1 y t_2 para los cuales $|t_j| < h$, para alguna $h > 0, j = 1, 2.$

◈

[32] La demostración de este teorema puede consultarse en Mood *et al.* [1974], p 160.
[33] La demostración de este teorema puede consultarse en Mood *et al* [1974], p 161.

- *Nota*: Los *Teoremas 1.11* y *1.12* pueden generalizarse a k variables aleatorias.

☞ ***La técnica de la función generadora de momentos***

Nuevamente, dadas las variables aleatorias $X_1,\ldots, X_n$ con función de densidad [masa] conjunta $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ y funciones $g_1(\cdot,\ldots,\cdot),\ldots,g_k(\cdot,\ldots,\cdot)$, se busca la distribución conjunta de $Y_1 = g_1(X_1,\ldots,X_n),\ldots,Y_k = g_k(X_1,\ldots,X_n)$

La función generadora de momentos del vector aleatorio $(Y_1,\ldots, Y_n)$, si existe, es

$$
\begin{aligned}
m_{Y_1,\ldots,Y_k}(t_1,\ldots,t_k) &= E[e^{t_1Y_1+\cdots+t_kY_k}] \\
&= \sum\cdots\sum e^{t_1g_1(x_1,\ldots,x_n)+\cdots+t_kg_k(x_1,\ldots,x_n)} f_{X_1,\ldots,X_k}(x_1,\ldots,x_n)
\end{aligned}
$$

para el caso discreto

y

$$
\begin{aligned}
m_{Y_1,\ldots,Y_k}(t_1,\ldots,t_k) &= E[e^{t_1Y_1+\cdots+t_kY_k}] \\
&= \int\cdots\int e^{t_1g_1(x_1,\ldots,x_n)+\cdots+t_kg_k(x_1,\ldots,x_n)} f_{X_1,\ldots,X_k}(x_1,\ldots,x_n)dx_1\ldots dx_k
\end{aligned}
$$

para el caso continuo.

Si el resultado de la suma o de la integral, según sea el caso, es una función de $t_1,\ldots, t_k$ que puede ser reconocida como la función generadora de momentos conjunta de alguna distribución conjunta conocida, entonces $X_1,\ldots,X_n$ tiene *esa* distribución conjunta, puesto que como ya se dijo la función generadora de momentos, cuando existe, es única y permite determinar de manera única su función de distribución.

Para $k > 1$, este método tiene una utilidad limitada puesto que sólo es posible reconocer unas cuantas funciones generadoras de momentos. *El supuesto de independencia de las variables aleatorias $X_1,\ldots, X_n$ hace más factible la identificación de la función generadora de momentos.*

Asimismo, como para $k = 1$, la función generadora de momentos es una función univariada se tiene más oportunidad de identificar la distribución a la que corresponde [34]

La aplicación más útil de la técnica de la función generadora de momentos consiste en determinar la distribución de las sumas de variables aleatorias independientes, como podrá verificarse cabalmente en el Capítulo 3. Este resultado se expone en el siguiente teorema.

Resultado 1.18
Distribución de la suma de v.a.i.'s
(aplicación de la *Técnica de la f.g.m.*)

Si $X_1, \ldots, X_n$ son variables aleatorias independientes y la función generadora de momentos de cada una existe para toda $-h < t < h$, para alguna $h > 0$, sea $Y = \sum_{i=1}^{n} X_i$. Entonces

$$M_Y(t) = E\left[\exp\left(\sum_{i=1}^{n} X_i t\right)\right] = \prod_{i=1}^{n} M_{X_i}(t) \qquad \text{para} \quad -h < t < h \qquad [1.18].$$

Demostración.

$$\begin{aligned} M_Y(t) &= E\left[\exp\left(\sum_{i=1}^{n} X_i t\right)\right] && \text{(por definición de f.g.m.)} \\ &= E\left[e^{tX_1 + tX_2 + \cdots + tX_n}\right] = E\left[e^{tX_1} e^{tX_2} \ldots e^{tX_n}\right] \\ &= E\left[\prod_{i=1}^{n} e^{tX_i}\right] = \prod_{i=1}^{n} E\left[e^{tX_i}\right] && \text{(usando una generalización del Teorema 1.12)} \\ &= \prod_{i=1}^{n} M_{X_i}(t) && \text{(por definición de f.g.m.)} \end{aligned}$$

◊

[34] Mood *et al.* [1974] comentan tambien que esta técnica es muy poderosa en conexión con ciertas técnicas de matemáticas avanzadas (la teoría de las transformaciones) las cuales, en muchos casos, permiten determinar la distribución asociada con la función generadora de momentos generada.

Por lo tanto, dada la unicidad de la función generadora de momentos, si se reconoce $\prod_{i=1}^{n} M_{X_i}(t)$ como la función generadora de momentos de una distribución particular, entonces se ha encontrado la distribución de $\sum_{i=1}^{n} X_i$. En el ANEXO 7 se presentan las distintas funciones generadoras de momentos que se requerirán para aplicar este resultado.

Con el *Resultado 1.18* se cierra la "caja de herramientas" a utilizar en el Capítulo 3 para demostrar una buena parte de las relaciones del esquema de Leemis, así como aquellas que se adicionaron. En dicho capítulo se hará referencia explícita a cada una de estas herramientas con el número de *Resultado* que aquí se le asignó.

En el siguiente capítulo se presentan las distintas distribuciones de probabilidad con las que se *trabajará en el* Capítulo 3.

♦ ♦ ♦

2

PRESENTACIÓN DE LAS DISTRIBUCIONES

En este capítulo se presentan las 28 distribuciones que se incluyen en el diagrama de Leemis. 9 de ellas son discretas y 19 son continuas. Cada presentación se inicia definiendo la distribución en términos de su función de densidad [o de masa] de probabilidad Adicionalmente, para aquellas distribuciones en que fue posible, se incluyen algunos hechos históricos, algunos ejemplos de aplicación (ya sea clásicos o novedosos) y algunas propiedades interesantes asociadas.[1] Esto con el fin de hacer amena la lectura y al mismo tiempo dar cuenta de elementos que pudieran pasar inadvertidos cuando se estudia el tema. Al final de cada presentación se enumeran las relaciones que guarda la distribución en cuestión con otras; la verificación de dichas relaciones será el objetivo del siguiente capítulo.

El capítulo está dividido en dos grandes secciones. Primero se presentan las distribuciones discretas y después las continuas El orden de presentación atiende a dos criterios: es alfabético y antepone las distribuciones más conocidas a aquellas que lo son menos.[2] Las distribuciones no tan conocidas, o por lo menos no "clásicas" (la arcoseno, la triangular y la Weibull discreta, por ejemplo), se incluyen al final de cada sección bajo el rubro de "Otras distribuciones".

Se espera que los contenidos que se exponen en este capítulo diluyan en alguna medida la ruptura que pudiera haber entre teoría y práctica, haciendo que las variables aleatorias representen modelos "tangibles". Además, con los hechos históricos que se mencionan se busca sacar del anonimato a algunos de los personajes que han contribuido al desarrollo del conocimiento en el campo de la probabilidad y de la estadística.

- *Sobre la notación* Se utilizan los términos *función de densidad de probabilidad*, abreviándolo como *f.d.p.*, o simplemente *función de densidad* tanto en el caso continuo como en el discreto. Para describir mejor la función de densidad de cada variable aleatoria X se recurrirá a la notación $f_X(x;\Theta)$, donde Θ es el vector de parámetros. la mayoría de las distribuciones son uniparamétricas y a lo más tienen tres parámetros. Así, se dirá que X tiene distribución D con parámetro(s) θ [(θ_1,θ_2) o $(\theta_1,\theta_2,\theta_3)$], lo cual se abreviará como $D(\theta)$ [$D(\theta_1,\theta_2)$ o $D(\theta_1,\theta_2,\theta_3)$]. Además, se recurrirá en ocasiones a la notación de Wilks "$\sim$" para abreviar "se distribuye D" o "tiene distribución D". Con las siglas "e o.c" se abrevia "en otro caso".

---

[1] La mayor parte de las propiedades que se exponen aparecen en la generalidad de la bibliografía consultada Los ejemplos fueron adaptados a partir de distintas fuentes. Los datos históricos fueron tomados en su gran mayoría de los volúmenes de Johnson y Kotz (véanse Johnson y Kotz [1969] y [1970])

[2] "Más o menos conocidas" de acuerdo a la frecuencia con que se encontraron en la literatura revisada

DISTRIBUCIONES UNIVARIADAS DISCRETAS

La distribución Bernoulli

Ensayo Bernoulli.[3] Un ensayo Bernoulli es un experimento aleatorio con dos, *y sólo dos*, posibles resultados –digamos éxito o fracaso–, de tal manera que estos resultados son mutuamente ajenos.

Definición

Una v.a. X tiene *distribución Bernoulli* con parámetro p si:

$$X = \begin{cases} 1 & \text{con} \quad \text{probabilidad} \quad p \\ 0 & \text{con} \quad \text{probabilidad} \quad 1-p \end{cases}$$

donde $0 \leq p \leq 1$. Bajo estas condiciones, la f.d.p. de X está dada por:

$$f_X(x) = f_X(x;p) = \begin{cases} p^x(1-p)^{1-x} & \text{para} \quad x = 0,1 \\ 0 & \text{e.o.c.} \end{cases} \qquad [2.1]$$

Frecuentemente, el valor $X = 1$ está asociado a un "éxito", y a p se le refiere como la *probabilidad de éxito*, mientras que el valor $X = 0$ está asociado a un "fracaso" y por ende al valor $1-p$ se le reconoce como la probabilidad de fracaso, denotándola generalmente como q.

Ejemplos de aplicación

Una variable aleatoria X que se defina como 1 si un ensayo Bernoulli resulta en éxito y como 0 si ese mismo ensayo Bernoulli resulta en fracaso tiene distribución Bernoulli con parámetro $p = P[\text{éxito}]$. Muchos experimentos pueden ser modelados como ensayos Bernoulli. El más simple y clásico es el de lanzar una moneda (sea $X = 1$ si la moneda cae en sol). Otros ejemplos incluyen juegos de apuestas (sea $X = 1$ si sale rojo), las elecciones ($X = 1$ si el candidato "A" obtiene un voto) y la incidencia de una enfermedad (sea $X = 1$ si una persona al azar se infecta).

[3] Nombre otorgado en honor al matemático suizo James Bernoulli (1654-1705), considerado como uno de los fundadores de la teoría de la probabilidad.

Relaciones con otras distribuciones

1 La distribución Bernoulli es un caso especial de la distribución binomial [3 1]

2 La suma de v a i.i.d Bernoulli se distribuye binomial [3.2]

La distribución binomial

Definición

Se dice que una v.a X tiene *distribución binomial* con parámetros n y p si su f.d p está dada por

$$f_X(x) = f_X(x;n,p) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x} & \text{si} \quad x = 0, 1, \ldots, n \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.2]}^{4}$$

donde n es un entero positivo y $0 \le p \le 1$

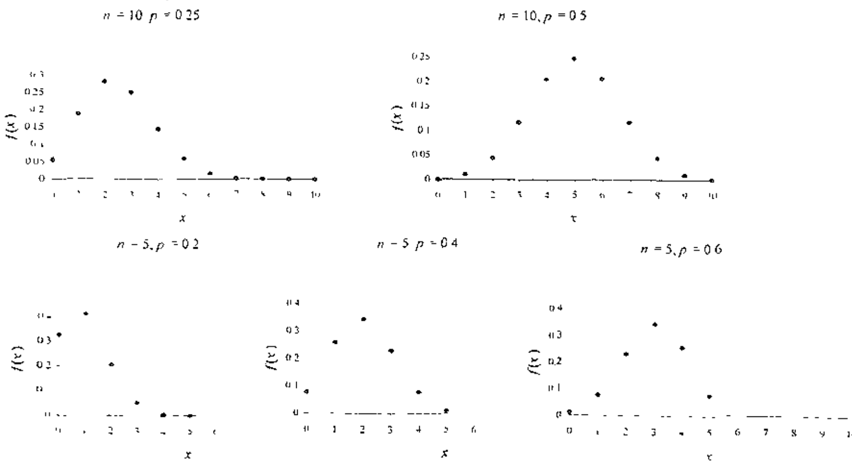


FIGURA 2.1. F d.p. binomial para distintos valores de n y p

[4] Johnson y Kotz [1969] mencionan que la distribucion binomial puede ser definida en terminos de la expansion binomial $(p + q)^n$, pues el $(k + 1)$-ésimo término de esta expansión es $\binom{n}{k} p^k q^{n-k}$. Advierten también que algunas veces se utiliza una forma más general en la cual (*sólo*) el lado derecho de la expresión [2 2] se sustituye por $a - bx$, donde a y b son números reales con b distinto de cero.

[5] Boyer, C.B. (1950), "Cardan and the Pascal triangle" en American Mathematical Monthly, 57, pp. 387-390

Algo de historia

La distribución binomial es una de las más antiguas en ser objeto de estudio. Fue derivada por James Bernoulli en su tratado *Ars Conjectandi* publicado en 1713. Mucho antes, los coeficientes del binomio pudieron encontrarse en algunos trabajos de Pascal. Referencias previas están dadas en un artículo de Boyer.[5]

Ejemplos de aplicación

La distribución binomial es aplicable a aquellos problemas que involucran *repetidos ensayos independientes* cuyos resultados pueden clasificarse en *dos categorías*, por ejemplo, éxito y fracaso, donde *la probabilidad de éxito es la misma* para cada ensayo. Es decir, la distribución binomial es aplicable para experimentos que consisten en realizar varios ensayos independientes Bernoulli y en los que interese el número de veces en que el experimento resulte en éxito.[6]

Ejemplo clásico. Considere un experimento aleatorio que consiste en la realización de n ensayos repetidos Bernoulli, donde p es la probabilidad de éxito en cada ensayo $(0 \le p \le 1)$. El espacio muestral de este experimento es el conjunto de n–adas de la forma $(z_1, z_2, \ldots, z_n)$, donde cada $z_i = 0$ ó 1, según haya sido el resultado del i–ésimo ensayo.[7] Como los ensayos son independientes, la probabilidad de cualquier n–ada específica, por ejemplo $(0, 0, 1, 0, 1, 1, \ldots, 0, 1)$ está dada por $qqpqpp\ldots qp$. Sea X una variable aleatoria que representa el número de éxitos que suceden en los n ensayos. Observe que X puede tomar los valores $0, 1, \ldots, n$. Ahora,

$$\begin{aligned} P[X = x] &= P[\text{exactamente } x \text{ éxitos y } n - x \text{ fracasos en } n \text{ ensayos}] \\ &= \binom{n}{x} p^x (1-p)^{n-x}, \quad \text{para} \quad x = 0, 1, \ldots, n \end{aligned}$$

---

[5] Cuando se generaliza de repetidos e independientes ensayos Bernoulli a repetidos ensayos independientes con *más de dos* posibles resultados surge la *distribución multinomial* (o *distribución binomial generalizada*, como la llama Tsokos [1972]), sobre la cual no se ahondará en este trabajo pues se trata de una distribución multivariada

[7] Hay que distinguir que 0 ó 1 es el resultado del i–ésimo ensayo, mientras que cada n–ada $(z_1, z_2, \ldots, z_n)$ es un resultado del experimento.

ya que *cada* resultado del experimento que tiene exactamente x éxitos tiene probabilidad $p^x(1-p)^{n-x}$ y existen $\binom{n}{x}$ resultados de esa forma. Por tanto, X tiene distribución binomial con parámetros n y p.

Con este ejemplo resulta claro que una variable aleatoria Bernoulli es una variable aleatoria binomial con parámetros $(1, p)$.

◈

Otro ejemplo.[8] Considere un muestreo *con* reemplazo de un lote con M artículos, de los cuales K se consideran "defectuosos". Sea X el número de artículos defectuosos en una muestra de tamaño n Las extracciones individuales son ensayos Bernoulli donde "defectuoso" corresponde a "éxito", y el experimento de tomar una muestra de tamaño n con reemplazo consiste en repetir n ensayos independientes Bernoulli donde $p = P[\text{éxito}] = K/M$ Entonces X tiene la distribución binomial con parámetros n y $p = K/M$

$$\binom{n}{x}\left[\frac{K}{M}\right]^x\left[1-\frac{K}{M}\right]^{n-x} \quad \text{para} \quad x = 0, 1, \ldots, n$$

◈

✂ *Propiedades*

- Como puede observarse en las gráficas de la *Figura 2.1*, los términos $f_X(x;\, n,\, p)$ parecen aumentar de manera monótona, para después decrecer también de manera monótona. El siguiente téorema verifica que en efecto esto sucede.

Teorema 2.2.1

Suponga que X tiene distribución Binomial (n, p), con $0 < p < 1$, entonces conforme x va de 0 a n, $f_X(x;\, n,\, p) = P[X = x]$ primero crece monótonamente y después decrece monótonamente, alcanzando su máximo valor cuando x es el mayor entero menor o igual que $(n+1)p$

[8] Este ejemplo resultará de especial interés en el capítulo 3, cuando se aborde la relación entre la distribución binomial y la distribución hipergeométrica.

Demostración. Considere el cociente $P[X=x]/P[X=x-1]$. Se busca determinar para qué valores de x este cociente es mayor (densidad decreciente), menor (densidad creciente) o igual (alcanza su máximo) que 1. Entonces:

$$\frac{P[X=x]}{P[X=x-1]}=\frac{(n-x+1)\,p}{x\,(1-p)},$$

Por lo tanto, el cociente es mayor o igual que 1 si y sólo si $P[X=x]\geq P[X=x-1]$, lo que ocurre si y sólo si $(n-x+1)p \geq x(1-p)$ o, equivalentemente, si y sólo si $(n+1)p \geq x$. En consecuencia, el cociente es menor que 1 si y sólo si $(n+1)p < x$. Como la f.d.p de X es positiva sólo sobre valores *enteros* ($\{0, 1, 2, \ldots, n\}$), el *máximo valor* de la f.d.p. se alcanza cuando $x=(n+1)p$ si el valor es entero; si no, se alcanza en el entero más grande que no exceda al valor $(n+1)p$.

◆

- El sesgo de la distribución es positivo si $p < ½$ y negativo si $p > ½$.. La distribución es simétrica si y sólo si $p = ½$.

Relaciones con otras distribuciones

La importancia de la distribución binomial en estadística radica en el hecho de que está relacionada con una amplia variedad de distribuciones. La distribución binomial puede ser considerada como una forma límite de la distribución hipergeométrica, así como de la distribución beta–binomial. A su vez las distribuciones Poisson y normal son formas límite de la distribución binomial. Además, como ya se había mencionado, la distribución Bernoulli es un caso especial de la binomial y esta última puede obtenerse a partir de una suma de variables aleatorias independientes Bernoulli. También está relacionada consigo misma si se considera la suma de variables aleatorias independientes binomiales.[9] Así:

[9] Johnson y Kotz [1969] mencionan que la diferencia de dos v.a.i. con distribución binomial tiene también distribución binomial de la forma general (comentada en la primera nota al pie de este apartado sobre la distribución binomial). Además, presentan una distribución especial asociada con la distribución binomial, que es la distribución condicional de la v.a. binomial X_1 con parámetros (n_1, p_1) dado que X_1+X_2, donde X_2 es otra variable binomial independiente de la primera, con parámetros (n_2, p_2).

1. La distribución Bernoulli es un caso especial de la distribución binomial [3 1]
2. La suma de v.a.i i d Bernoulli se distribuye binomial. [3 2]
3. La distribución binomial satisface la *propiedad reproductiva* [10] [3.3] y [3 3*bis*]
4. La distribución binomial es una aproximación a la distribución hipergeométrica [3 4]
5. La distribución binomial es una aproximación a la distribución beta-binomial. [3.5]
6. La distribución Poisson es una aproximación a la distribución binomial. [3 7]
7. La distribución normal es una aproximación a la distribución binomial. [3.21]

La distribución binomial negativa

Definición

Se dice que una v.a. X tiene *distribución binomial negativa* con parámetros r y p si su f d.p. está dada por

$$f_X(x) = f_X(x;r,p) = \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x & \text{si} \quad x = 0,1,2,\ldots \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.3]}$$

donde r es un entero no negativo y $0 < p \leq 1$ [11-12]

[10] La *propiedad reproductiva* se refiere al hecho de que la suma de v.a.i.i d conserve la misma distribucion que las variables originales, con las modificaciones pertinentes en los parámetros

[11] Debido a la relacion que existe entre la distribución geométrica y la binomial negativa, la expresión [2.3] presentada por Leemis coincide también con la considerada por Mood *et al* [1974], pero difiere de aquélla expuesta por Tsokos [1972] y Ross [1988], quienes definen la densidad binomial negativa considerando a r (en lugar del cero) como el punto de masa más pequeño, como sigue:

$$f_X(x) = f_X(x;r,p) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & \text{si} \quad x = r, r+1, r+2,\ldots \\ 0 & \text{e.o.c.} \end{cases} \qquad [2.3bis]$$

Casella y Berger [1990] presentan ambas expresiones. Posteriormente se ilustrará cuál es la diferencia práctica entre las expresiones [2 3] y [2 3*bis*].

[12] La distribución binomial negativa toma su nombre de la expresion $(Q-P)^{-N}$, donde $Q - P = 1$, que corresponde a la expansion de la expresión binomial negativa (que define los coeficientes binomiales con enteros negativos) El $(k+1)$-ésimo termino de la expansion $(Q-P)^{-N}$ es

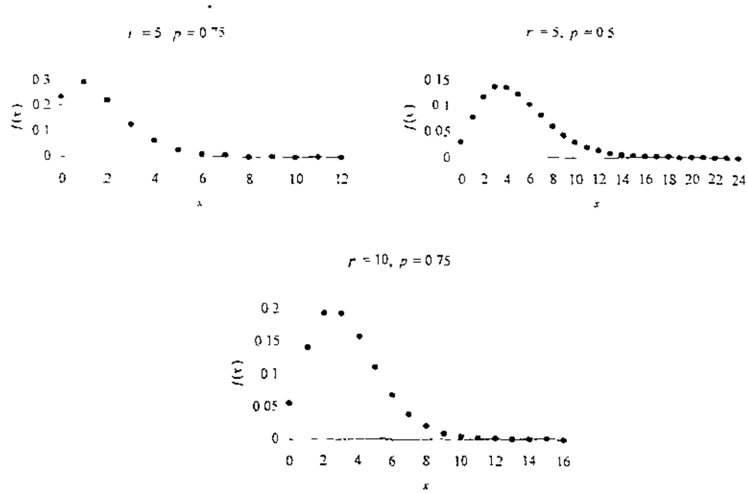


FIGURA 2.2. F.d.p binomial negativa para distintos valores de r y p.

Algo de historia

Pascal y Fermat discutieron formas especiales de la distribución binomial negativa. Montmort publicó una derivación de la misma en 1714. En 1907, 'Student' utilizó esta distribución como una alternativa para la distribución Poisson, para describir los conteos de las placas de un dispositivo de medición. En 1920, Greenwood y Yule obtuvieron la distribución como consecuencia de ciertas suposiciones simples en los modelos de propensión a sufrir accidentes, mientras que en 1923, Eggenberger y Pólya la obtuvieron como un caso límite de un 'esquema de urnas'. El número de aplicaciones de la distribución binomial negativa es grande. Una parte importante del desarrollo de técnicas estadísticas ha estado basado en esta distribución.

$$Q^{-N}\binom{N+x-1}{r-1}(P/Q)^x = \binom{N+x-1}{r-1}(P/Q)^x(1-P/Q)^N.$$

Sustituyendo $N = r$ y $P = (1-p)/p$ se obtiene el término en la expresión [2.3]. A diferencia de lo que ocurre con la distribución binomial r no necesariamente debe ser entero; cuando lo es, a la distribución se le conoce a veces como *distribución Pascal*.

Ejemplos de aplicación

Ejemplo clásico Considere una secuencia de ensayos independientes, repetidos Bernoulli con probabilidad p $(0 < p \leq 1)$ de éxito en cada ensayo Sea X una variable aleatoria que representa el número de fracasos antes del r–ésimo éxito Observe por un lado, que X puede tomar los valores 0, 1, 2,… y, por otro lado, que se requieren al menos r ensayos para obtener r éxitos Si $X = x$ entonces habrán de realizarse $(x + r)$ ensayos, de los cuales el último debe resultar en éxito, teniendo éste probabilidad p Por ende, entre los primeros $(x + r - 1)$ ensayos deberá haber $(r - 1)$ éxitos y, obviamente, x fracasos Por lo tanto la probabilidad del evento $\{X = x\}$ es

$$\binom{x+r-1}{r-1} p^{r-1} q^x \cdot p = \binom{x+r-1}{r-1} p^r q^x \quad , \qquad x = 0, 1, 2, \ldots$$

que corresponde a la densidad binomial negativa de la expresión [*2.3*].[13]

- *Nota*: La distribución binomial negativa es también una distribución discreta de "tiempo de espera". En este ejemplo, X representa cuánto hay que esperar (en términos del número de fracasos) por el r–ésimo éxito.

◈

Otro ejemplo. La distribución binomial negativa tiene importancia en la consideración del *muestreo binomial inverso*, que es una técnica útil en el muestreo de poblaciones en biología. Suponga que una proporción p de individuos en una población poseen cierta característica Si se extrae una muestra de los individuos de esta población hasta que exactamente r individuos con esa cierta característica son encontrados, entonces el número de individuos que excede a los r y que son observados o extraídos en la muestra tiene distribución binomial negativa [14]

◈

[13] Observe que esta expresión es equivalente a $\binom{x+r-1}{r-1} = \binom{x+r-1}{x}$, lo que significa que es equivalente pedir que ocurran los $(r - 1)$ fracasos dentro de los primeros $(x + r - 1)$ ensayos a pedir que sucedan x éxitos dentro de esos primeros $(x + r - 1)$ ensayos

Por otra parte, de haber considerado a X como el *número de ensayos realizados hasta obtener el r–ésimo éxito*, el recorrido de X tendría sentido desde r $(r, r + 1, r + 2, \ldots)$ y su f.d p. estaría dada por la expresión [*2.3bis*].

[14] Si el contexto hubiera sido muestrear hasta que se observen r individuos con tal característica, entonces el *número de individuos extraídos en la muestra* sería una variable aleatoria binomial negativa.

La distribución binomial negativa es utilizada frecuentemente en sustitución de la distribución Poisson cuando se tiene duda sobre si los requerimientos estrictos –especialmente el de independencia– de ésta última se satisfacen. Entre los campos específicos donde se ha encontrado que la distribución binomial negativa provee representaciones útiles pueden mencionarse las estadísticas de accidentes, los procesos de vida y muerte, los datos psicológicos, en la demanda (por parte de los hogares) de productos de amplio consumo y como pesos de (*distribuciones* LAG) de series de tiempo en economía. Se han encontrado también aplicaciones médicas y militares.

Relaciones con otras distribuciones

1. La distribución geométrica es un caso especial de la distribución binomial negativa. [3.9]
2. La suma de v.a.i. con distribución geométrica tiene distribución binomial negativa. [3.10]
3. La distribución binomial negativa satisface la propiedad reproductiva. [3.11] y [3.11*bis*]
4. La distribución Poisson es una aproximación para la distribución binomial negativa. [3.8]

■

La distribución geométrica

Definición

Se dice que una v.a. X tiene *distribución geométrica* (o *Pascal*) con parámetro p si su f.d.p. está dada por

$$f_X(x) = f_X(x;\, p) = \begin{cases} p(1-p)^x & \text{si} \quad x = 0, 1, \ldots \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.4]}$$

donde $0 < p \leq 1$.[15]

[15] La expresión [2.4] presentada por Leemis en su artículo coincide con la considerada por Mood *et al* [1974], pero no con aquélla expuesta por Tsokos [1972], Ross [1988] y Casella y Berger [1990]. Estos autores definen a la densidad geométrica considerando al 1 (en lugar del 0) como el punto de masa más pequeño, como sigue:

$$f_X(x;p) = \begin{cases} p(1-p)^{x-1} & \text{si} \quad x = 1, 2, \ldots \\ 0 & \text{e.o.c.} \end{cases} \qquad [2.4bis]$$

Más adelante se ilustrará cuál es la diferencia practica entre las expresiones [2.4] y [2.4*bis*]

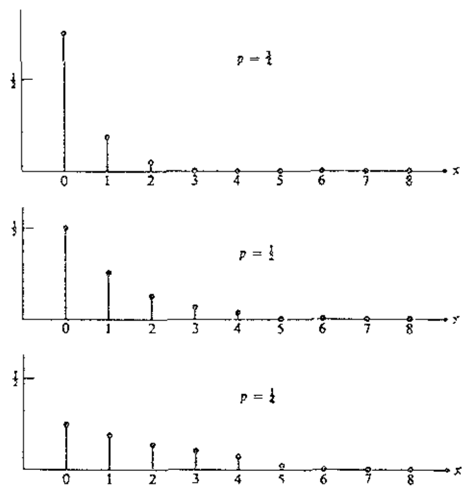


FIGURA 2 3. F d.p. geometrica para distintos valores de p

Esta distribución debe su nombre a que los términos sucesivos de la f.d.p. conforman una serie geométrica con parámetro p.

Ejemplos de aplicación

Ejemplo clásico. Suponga un experimento que consiste en repetir un ensayo Bernoulli, donde la probabilidad de éxito es p, hasta obtener el primer éxito. Observe que para cualquier número determinado k de ensayos existe una probabilidad positiva $(1-p)^k$ de que *no* ocurra un éxito. Sea X el número de ensayos ocurridos *antes* de obtener el primer éxito, siendo entonces fracasos tales ensayos. Así, X representa el número de *fracasos* ocurridos *antes* de obtener el *primer éxito*. No es posible limitarse a ningún número de ensayos–fracaso previamente determinado; es decir, X puede tomar los valores 0, 1, 2, ... Puede calcularse $P[X=x]$

$$P[X=x]=(1-p)^k p \qquad \text{para} \quad x=0,1,2,\ldots,$$

pues para que X sea igual a x es necesario que se realicen $x+1$ ensayos, de los cuales (a) los x *primeros* deberán ser *fracasos*, y (b) el *último* ensayo deberá resultar en *éxito*. Como se está suponiendo que los ensayos son independientes, la probabilidad de que ocurran x fracasos antes del primer éxito es el producto de las probabilidades de los dos eventos mencionados. Por lo tanto, X tiene distribución geométrica con parámetro p..[16]

❖ *Nota*: La distribución geométrica es la distribución de "tiempos de espera" más simple. En el ejemplo anterior se está en espera de un éxito. Así, X representa cuánto (en términos del *número de fracasos*) hay que esperar para obtener un éxito

◆

Otro ejemplo: El problema de las llaves (primera parte). A continuación se presenta un problema que no sólo dará pie para obtener la distribución geométrica, sino que más adelante será retomado para obtener la distribución rectangular o uniforme discreta, haciendo una modificación en la selección de las llaves.

Un persona con n llaves quiere abrir una puerta y prueba las llaves aleatoriamente Sólo *una* llave abre la puerta. Encontrar el número medio de llaves incorrectas probadas suponiendo que cada llave *incorrecta no se elimina* de selecciones subsecuentes (es decir, se está planteando un muestreo aleatorio *sin* reemplazo).

Sea X el número de llaves probadas *antes* de elegir la llave que abre la puerta. Como cada llave incorrecta no es descartada, siempre se tienen n llaves para elegir. Así, X puede tomar los valores $0, 1, 2, \ldots$ Puede pensarse la selección de cada llave en términos del contexto de una urna. Cada "extracción" de una llave es un ensayo independiente Bernoulli con probabilidad $p = 1/n$ de que la llave elegida sea la correcta (éxito). Bajo estas condiciones:

$$P[X=0] = P[\text{elegir la llave correcta en la primera selección}]$$
$$= \frac{1}{n}$$
$$P[X=1] = P[\text{elegir } \textit{una} \text{ llave incorrecta y a continuación elegir la llave correcta}]$$
$$= \left(\frac{n-1}{n}\right)\left(\frac{1}{n}\right) = \left(1-\frac{1}{n}\right)\left(\frac{1}{n}\right)$$

[16] De haber considerado a X como el número de *ensayos* realizados *hasta* obtener el *primer éxito* se habría dado lugar a la expresión [2 4*bis*], pues X es igual a x si y sólo si: (a) los $(x-1)$ *primeros* ensayos son *fracasos*, y (b) el *último* ensayo resulta en *éxito*.

$P[X=2] = P$[elegir primero *dos* llaves incorrectas y en la tercera selección elegir la llave correcta]

$$= \left(\frac{n-1}{n}\right)\left(\frac{n-1}{n}\right)\left(\frac{1}{n}\right) = \left(1-\frac{1}{n}\right)^2\left(\frac{1}{n}\right)$$

$P[X=k] = P$[elegir primero $k-1$ llaves incorrectas y en la k–ésima seleccion elegir la llave correcta]

$$= \underbrace{\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-1}{n}\right)\left(\frac{1}{n}\right)}_{(k-1)-\text{extracciones}} = \left(1-\frac{1}{n}\right)^{j-1}\left(\frac{1}{n}\right)$$

$\vdots$

Es decir, X tiene distribución geométrica con parámetro $p = 1/n$ Calculando ahora el número medio de llaves incorrectas probadas antes de abrir la puerta:

$$E[X] = \sum_{j=0}^{\infty} j \cdot \left[(1-p)^j p\right]$$

$$= p\sum_{j=0}^{\infty} j(1-p)^j$$

$$= p \cdot (1-p)\sum_{j=0}^{\infty} j(1-p)^{j-1} \qquad \text{(multiplicando por } \frac{1-p}{1-p}\text{)}$$

$$= p \cdot (1-p)\sum_{j=0}^{\infty} -\frac{d}{dp}(1-p)^j$$

$$= p \cdot (p-1)\frac{d}{dp}\sum_{j=0}^{\infty}(1-p)^j$$

$$= p \cdot (p-1)\frac{d}{dp}\left[\frac{1}{1-(1-p)}\right] \qquad \text{(aplicando que } \sum_{j=0}^{\infty} r^j = \frac{1}{1-r} \text{ para } |r|<1\text{)}$$

$$= p \cdot (p-1)\frac{d}{dp}\left[\frac{1}{p}\right]$$

$$= p \cdot (p-1)\left[-\frac{1}{p^2}\right]$$

$$= \frac{1-p}{p}$$

Esta es la media de una variable aleatoria con distribución geométrica con parámetro p Como en este caso $p = 1/n$. entonces $(1-p)/p = (1/p) - 1 = n - 1$, lo que significa que en promedio se requiere que la persona intente tantas veces como el número de llaves incorrectas de las que dispone antes de abrir la puerta [17]

Propiedades

- La distribución geométrica posee una propiedad interesante que puede denominarse como la *"propiedad de la pérdida de memoria"*, pues establece que la probabilidad de que una variable aleatoria geométrica sea mayor o igual que $i + j$ dado que se sabe que es mayor o igual que i es igual a la probabilidad incondicional de que ésta sea mayor o igual que j

Teorema 2.4.1

Si X tiene densidad geométrica con parámetro p, entonces

$$P[X \geq i + j \mid X \geq i] = P[X \geq j] \qquad \text{para} \qquad i, j = 0, 1, 2, \ldots$$

Demostración. Un elemento necesario para demostrar este teorema es el cálculo de $P[X \geq k]$ [18]:

$$\begin{aligned} P[X \geq k] &= \sum_{m=k}^{\infty} p(1-p)^m \\ &= p\left[\sum_{m=0}^{\infty}(1-p)^{k+m}\right] \\ &= p(1-p)^k \sum_{m=0}^{\infty}(1-p)^m \\ &= p(1-p)^k \frac{1}{1-(1-p)} \\ &= (1-p)^k \end{aligned}$$

---

[17] Si la f.d.p. de X está dada por la expresión [*2.4bis*], la media es $1/p = n$ que en términos prácticos tiene la misma interpretación

[18] En términos del ejemplo clásico presentado, este resultado significa que la probabilidad de que se requieran al menos k fracasos antes de obtener un éxito es igual a la probabilidad de obtener primero exactamente k fracasos.

De haber considerado la expresión [*2.4bis*], se habría obtenido que $P[X \geq k] = (1-p)^{k-1}$. Es decir, la probabilidad de que se requieran al menos k ensayos para obtener un éxito es igual a la probabilidad de que los primeros $k - 1$ ensayos sean fracasos.

Así,

$$
\begin{aligned}
P[X \geq i+j \mid X \geq i] &= \frac{P[X \geq i+j]}{P[X \geq i]} \\
&= \frac{(1-p)^{i+j}}{(1-p)^{i}} \\
&= (1-p)^{j} \\
&= P[X \geq j]
\end{aligned}
$$

◈

Relaciones con otras distribuciones

La relación más extendida con la distribución geométrica es que ésta es un caso particular de la distribución binomial negativa. Además, la distribución geométrica se vincula con ella misma a través de la mínima estadística de orden. El artículo de Leemis presenta la relación que guarda la distribución geométrica con la distribución Weibull discreta; ésta no se menciona en ninguna de las fuentes consultadas, pero refuerza la similitud entre las distribuciones geométrica y exponencial.

1. La distribución geométrica es un caso especial de la distribución binomial negativa. [3.9]
2. La suma de v.a.i. con distribución geométrica tiene distribución binomial negativa. [3.10]
3. La mínima estadística de orden de una distribución geométrica es también geométrica. [3.12]
4. La distribución geométrica es un caso especial de la distribución Weibull discreta. [3.13]

La distribución hipergeométrica

Definición

Se dice que X es una v.a. con *distribución hipergeométrica* con parámetros n_1, n_2 y n_3 si su f.d.p. está dada por

$$f_X(x;n_1,n_2,n_3)=\begin{cases}\dfrac{\binom{n_1}{x}\binom{n_3-n_1}{n_2-x}}{\binom{n_3}{n_2}} & \text{si} \quad x=0,1,\ldots,\text{mín}(n_1,n_2)\\ 0 & \text{e.o.c.}\end{cases} \qquad [2.5]$$

donde n_1 es un entero no negativo, n_3 es un entero positivo y n_2 es un entero positivo a lo más tan grande como n_3.[19]

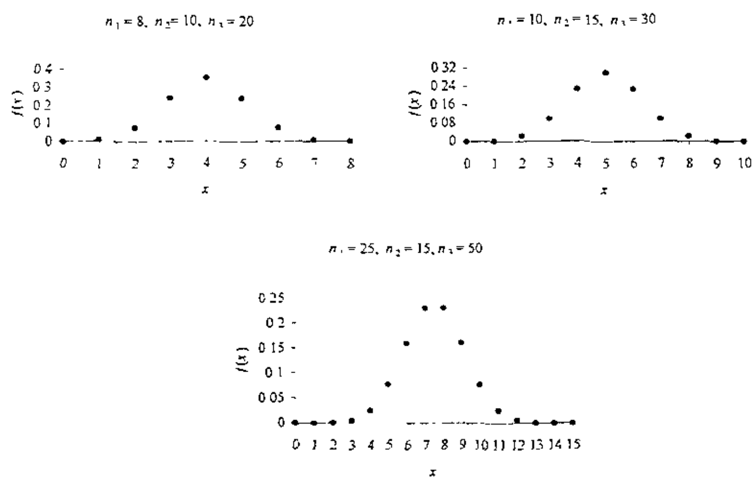


FIGURA 2.4. F.d.p. hipergeométrica para distintos valores de n_1, n_2 y n_3.

[19] Johnson y Kotz [1969] explican que el término "hipergeométrica" se debe a que las cantidades en el lado derecho de la expresión [2.5] son términos sucesivos en la expansión de

$$\frac{(N-n)!\,(N-X)!}{N!\,(N-X-n)!}F(-n,-X;N-X-n+1;1)$$

donde $F(\alpha,\beta,\gamma,z)=1+\frac{\alpha\beta}{\gamma}\frac{z}{1!}+\frac{\alpha(\beta+1)\beta(\beta+1)}{\gamma(\gamma+1)}\cdot\frac{z^2}{2!}+\cdots$ es una función hipergeométrica.

Ejemplos de aplicación

Ejemplo clásico. Procesos de control de calidad Suponga que se tiene un lote de N artículos de los cuales D tienen algún defecto de fabricación Se propone como parámetro de control de calidad que el lote sea rechazado en su totalidad si al extraer (*sin reemplazo*) una muestra de n artículos del lote, x de ellos (por lo menos) resultan defectuosos. Si se define X como una variable aleatoria que cuenta el número de artículos defectuosos en la muestra de tamaño n es claro que X a los menos puede ser cero y que no puede tomar valores mayores que la n fijada si $n \leq D$, o mayores que D si $n \geq D$. Es decir, X puede ser un valor entero x tal que $0 \leq x \leq \min(D,n)$. Ahora, siguiendo la regla de la probabilidad clásica (casos favorables entre casos totales):

$$P[X = x] = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} \quad \text{para } x = 0, 1, \ldots, \min(D,n)$$

Por lo tanto, X tiene distribución hipergeométrica con parámetros D, n y N.

◊

Otro ejemplo Estimación del tamaño de poblaciones animales a partir de datos de "captura-recaptura".[20] Un número desconocido, digamos N, de animales habita cierta región. Para obtener alguna información sobre el tamaño de la población, algunos ecologistas realizan frecuentemente el experimento siguiente: Primero capturan un número, digamos r, de estos animales, los marcan de alguna manera y los sueltan. Después de dejar que los animales marcados se dispersen dentro de la región, se realiza una nueva captura de tamaño n. Sea X el número de animales marcados en esta segunda captura. Suponiendo que la población de animales permaneció fija durante el tiempo transcurrido entre las dos capturas y que cada vez que se capturó un animal era igualmente probable que fuera uno de los animales aún no capturados, se sigue que X es una variable aleatoria hipergeométrica tal que

[20] Tomado de Ross [1988]. Johnson y Kotz [1969] hacen referencia también a una aplicación de este tipo, mencionando que data al menos de 1896 (véanse Ross [1988], pp 140–141, y Johnson y Kotz [1969], p 152).

$$P[X=i]=\frac{\binom{r}{i}\binom{N-r}{n-i}}{\binom{N}{n}}\equiv P_i(N)$$

Suponga ahora que la X observada es igual a i. Entonces, en tanto $P_i(N)$ representa la probabilidad del evento observado cuando hay en efecto N animales presentes en la región, pareciera que una estimación razonable de N sería el valor que maximizara $P_i(N)$. A tal estimador se le llama *estimador de máxima verosimilitud*.

La maximización de $P_i(N)$ puede hacerse de manera simple notando que

$$\frac{P_i(N)}{P_i(N-1)}=\frac{(N-r)(N-n)}{N(N-r-n+i)}$$

Ahora, este cociente es mayor o igual que 1 si y sólo si

$$(N-r)(N-n)\geq N(N-r-n+i)$$

o, equivalentemente, si y sólo si

$$N\leq\frac{rn}{i}$$

Por lo tanto, $P_i(N)$ primero es creciente y después decreciente, y alcanza su valor máximo en el entero más grande que no exceda a rn/i. Este valor es por tanto el estimador máximo verosímil de N. Por ejemplo, suponga que la captura inicial constó de $r = 50$ animales, los cuales son marcados y después puestos en libertad. Si una captura subsecuente consiste de $n = 40$ animales, de los cuales $i = 4$ están marcados, entonces se estimaría que hay alrededor de 500 animales en la región.

❖ Note que el estimador anterior podría haberse obtenido suponiendo que la proporción de animales marcados en la región, r/N, es aproximadamente igual a la proporción de animales marcados en la segunda captura, i/n [21]

Relaciones con otras distribuciones

Existe una gran variedad de aproximaciones a las probabilidades individuales (y a las sumas acumuladas de estas probabilidades) para las distribuciones hipergeométricas. Muchas de éstas están basadas en la aproximación de la distribución hipergeométrica por una distribución binomial con parámetros n y D/N, la cual se probará e ilustrará en el capítulo siguiente.[22]

- *La distribución binomial es una aproximación a la distribución hipergeométrica* [3 4]

La distribución Poisson

Definición

Se dice que X es una v.a. con *distribución Poisson* con parámetro μ si su f.d.p está dada por

$$f_X(x) = f_X(x;\mu) = \begin{cases} e^{-\mu}\dfrac{\mu^x}{x!} & \text{si} \quad x = 0, 1, 2, \ldots \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.6]}$$

donde $\mu > 0$ [23]

[21] Johnson y Kotz [1969] mencionan que otra forma natural en que surge la distribución hipergeométrica es en la *teoría de excedentes*. Considere dos muestras aleatorias de tamaños n_1 y n_2, extraídas de una población en la que una cierta característica que se mide tiene distribución continua El número de excedentes ${}^{m}\varepsilon_{2,1}$ se define como el número (de n_2) de valores observados en la segunda muestra que exceden en al menos (n_1-m+1) los valores en la primera muestra. La probabilidad $P[{}^{m}\varepsilon_{2,1} = k]$ puede calcularse a través de la f m p. de una distribución hipergeométrica con parámetros (n_1-1), (n_1+n_2-1) y $(k+m-1)$, dentro de límites apropiados para k

Por otra parte, exponen una aplicación de la distribución hipergeométrica a problemas lingüísticos sumamente interesante, pero que por restricciones de espacio no se presenta aquí. [Véase Johnson y Kotz [1969], pp. 152–155]

[22] Mood *et al.* [1974] exponen una propiedad que relaciona a estas dos distribuciones. Considere $D/N = p$ entonces la media de la distribución hipergeométrica coincide con la media de la distribución binomial, y la varianza de la distribucion hipergeométrica es $(N-n)/(N-1)$ veces la varianza de la distribución binomial

[23] El parámetro μ es precisamente el valor de la media (así como el de la varianza) de la variable aleatoria X y usualmente se determina de forma empírica

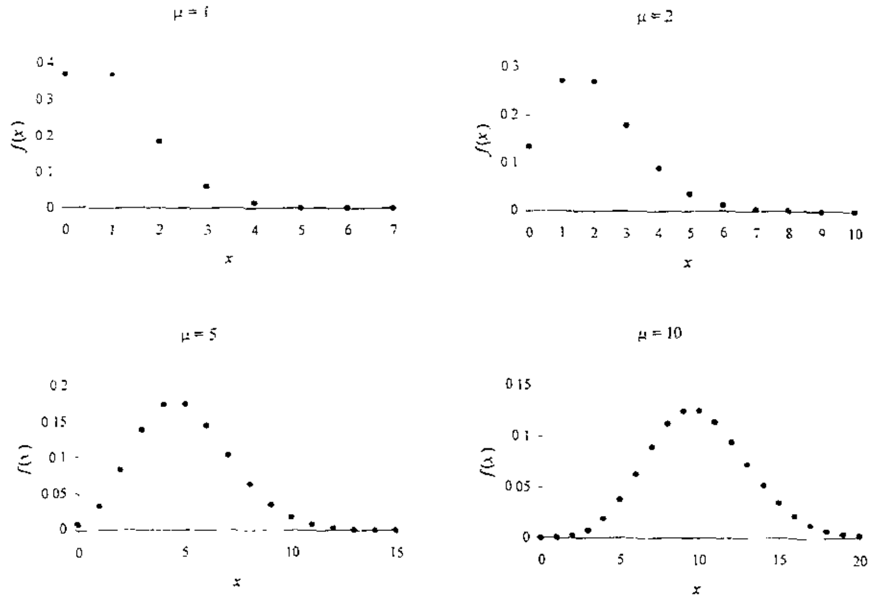


FIGURA 2.5. F.d.p. Poisson para distintos valores de μ.

Algo de historia

S.D. Poisson –matemático francés del siglo XVII– introdujo la distribución que lleva su nombre en su libro sobre la aplicación de la teoría de la probabilidad a los asuntos de litigios, juicios criminales, y temas similares (*Recherches sur la probabilité des jugements en matière criminelle et en matière civile*) publicado en 1837. Él llegó a esta distribución considerando formas límite de la distribución binomial. En 1898, Bortkiewicz expuso algunas circunstancias en las que la distribución Poisson puede surgir. Es curioso que uno de los resultados considerados fue el número de muertes a consecuencia de una patada de mula, por año, de integrantes de los cuerpos del ejército de Prusia.

En 1907, 'Student' utilizó la distribución Poisson para representar, como una primera aproximación, el número de partículas que caen en un área pequeña A cuando un número grande de tales áreas se esparce de manera aleatoria sobre una superficie grande en comparación con A.

Ejemplos de aplicación

El campo de aplicación de esta distribución es extenso El hecho de que los valores que una variable aleatoria Poisson puede tomar sean los enteros no negativos justifica que cualquier fenómeno aleatorio para el cual sea de interés llevar un conteo de algún tipo sea un candidato propicio a ser modelado mediante una distribución Poisson. Los ejemplos de estos conteos van desde el número de errores de dedo en una página mecanografiada hasta el número de partículas radioactivas emitidas por unidad de tiempo, pasando por el número de accidentes automovilísticos fatales ocurridos en un cruce particular y el número de personas que entran a un banco cierto día

Ejemplo. Suponga que el número de errores tipográficos en una página de algún libro tiene distribución Poisson con parámetro $\mu = ½$. Calcúlese la probabilidad de que haya al menos un error en esa página Sea X el número de errores en esa página. Entonces,

$$P[X \geq 1] = 1 - P[X = 0] = 1 - e^{-½} \approx 0.395$$

◊

Otro conjunto amplio de aplicaciones de la v.a Poisson se deriva de que su f.d.p es una aproximación a la f.d.p de una v a. binomial con parámetros (n, p) Las probabilidades Poisson se pueden utilizar para aproximar sus correspondientes probabilidades binomiales para valores grandes de n y valores pequeños de p de tal forma que $\mu = np$ adopte un tamaño moderado.[24] La relación binomial–Poisson se verificará en el capítulo siguiente.

Finalmente, la distribución Poisson proporciona un buen modelo para la distribución de probabilidad del número Y de *eventos raros* o eventos que ocurren de manera poco frecuente en alguna dimensión como el espacio, el tiempo o el volumen, en donde μ representa el valor de la media de Y[25]

---

[24] Mendenhall *et al* [1986] hacen explicito que esta relación opera satisfactoriamente cuando np es menor que 7 (otros autores señalan 5) Por su parte Johson y Kotz [1969] puntualizan que no es necesario que $\mu = np$ sea pequeño (y por ende no se requiere que los valores de las observaciones sean "pequeños", aunque Bortkiewicz llamó a esta relacion Poisson–binomial la "ley de los pequeños números"), sino que es lo grande de n y lo pequeño de p lo que importa

[25] Ross [1988] presenta una exposición clara en torno a este uso, específicamente sobre "eventos" que ocurren en ciertos puntos del tiempo, tales como la ocurrencia de un terremoto o la entrada de un individuo a algún establecimiento (banco, oficina postal, gasolinería, etc). Es dentro de estos contextos que se introduce el concepto de función $o(h)$ [Véase Ross [1988], pp. 133–136.]

Propiedades

- La función de densidad Poisson (al igual que la densidad binomial) posee una cierta monotonicidad como lo establece el siguiente teorema.

Teorema 2.6.1

Suponga que X *tiene distribución Poisson* (μ), entonces su f.d.p., $f_X(x;\ n, p) = P[X = x]$, primero crece monótonamente y después decrece monótonamente, alcanzando su máximo valor cuando x es el mayor entero menor o igual que μ.

Demostración. El *Teorema 2.6 1* se puede probar considerando el cociente

$$\frac{P[X = x]}{P[X = x-1]} = \frac{e^{-\mu}\mu^{x} / x!}{e^{-\mu}\mu^{x-1} / (x-1)!} = \frac{\mu}{x},$$

y siguiendo un razonamiento análogo al que se aplicó en la demostración del *Teorema 2.2.1* para la distribución binomial.[26]

◆

Relaciones con otras distribuciones[27]

Como ya se ha mencionado, la distribución Poisson puede ser utilizada como una aproximación a la distribución binomial.[28] La gran ventaja de esta relación radica en la simplificación del cálculo de probabilidades. La función de densidad Poisson sólo involucra un parámetro (en lugar de dos) y, además, evita el cálculo de combinatorias que en ocasiones puede resultar engorroso. Asimismo, la distribución Poisson puede ser utilizada como una aproximación a la distribución binomial negativa. *Al igual que la distribución binomial, la Poisson tiene la propiedad de que la suma de variables*

---

[26] Ross [1988] usa esta propiedad para exponer una forma sencilla de calcular probabilidades Poisson de manera recursiva Así, partiendo de que $P[X = 0] = e^{-\mu}$, se tiene que $P[X = k] = (\mu/k)\ P[X = k-1]$ para $k = 1, 2, \ldots$, que bien puede explotarse en un programa para computadora [véase Ross [1988], pp. 136-137].

[27] Johnson y Kotz [1969] sugieren un trabajo de Haight para la consulta de información sobre la Poisson y distribuciones relacionadas: Haight, F. (1967), *Handbook of the Poisson distribution*, Nueva York, John Wiley

[28] Johnson y Kotz [1969] mencionan otro vínculo entre las distribuciones Poisson y binomial Si X_1 y X_2 son v.a.i.i.d. Poisson con parámetros μ_1 y μ_2, respectivamente, la distribución *condicional* de X_1 dado $X_1 + X_2 = X$ es binomial con parametros X y $\mu_1/(\mu_1 + \mu_2)$.

aleatorias independientes Poisson vuelve a ser Poisson. Cuando el parámetro μ es suficientemente grande, la distribución Poisson tiende a la distribución normal.[29] Así.

1.	La distribución Poisson satisface la propiedad reproductiva.	[3.6] y [3.6bis]
2.	La distribución Poisson es una aproximación a la distribución binomial.	[3.7]
3.	La distribución Poisson es una aproximación a la distribución binomial negativa	[3.8]
4	La distribución normal es una aproximación a la distribución Poisson.	[3.22]

La distribución rectangular o uniforme discreta

Definición

Se dice que una v.a. X tiene *distribución rectangular* o *uniforme discreta* con parámetro n, si su f.d.p. está dada por

$$f_X(x) = f_X(x;n) = \begin{cases} \dfrac{1}{n} & \text{si} \quad x = 0, 1, \ldots, n-1 \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.7]}$$

donde n es un entero positivo.[30]

[29] Johnson y Kotz [1969] hacen referencia a las relaciones entre la Poisson y otras distribuciones gamma (además de la normal, la Ji-cuadrada por ejemplo). Aunque no se ahondará en ella posteriormente, se considera conveniente mostrar la relación siguiente señalada por ellos y por otros autores

$$\int_x^{\infty} \tfrac{1}{\Gamma(\alpha)} z^{\alpha-1} e^{-z}\, dz = \sum_{y=0}^{\alpha-1} \frac{x^y e^{-x}}{y!}, \quad \alpha = 1, 2, \ldots$$

donde el integrando del lado izquierdo (como se verá más adelante) es una f.d p., en la cual se apoya la construcción de la f.d.p. gamma, y $\Gamma(\alpha) = (\alpha-1)!$ ya que α es un entero positivo

[30] La expresión [2.7] presentada por Leemis en su artículo difiere ligeramente de aquélla considerada por Mood *et al* [1974] y Casella y Berger [1990] en cuanto al rango de valores que puede tomar la v a X. El rango definido por estos otros autores abarca también n valores, pero se inicia en 1 y culmina en n Así,

$$f_X(x) = f_X(x,n) = \begin{cases} \dfrac{1}{n} & \text{si} \quad x = 1, \ldots, n \\ 0 & \text{e o c} \end{cases}, \qquad [2\ 7bis]$$

define una distribución uniforme discreta sobre los enteros $1, 2, \ldots, n$.

FIGURA 2.6 F.d.p. rectangular

Ejemplos de aplicación

El problema de las llaves (*segunda parte*). Se retomará el problema de las llaves que se presentó en el apartado dedicado a la distribución geométrica, introduciendo una variante que permitirá obtener la distribución rectangular y contrastar el efecto de un muestreo sin reemplazo con aquél de un muestreo con reemplazo.

Una persona con n llaves quiere abrir una puerta y prueba las llaves aleatoriamente. Sólo *una* llave abre la puerta. Encontrar el número medio de llaves incorrectas probadas suponiendo que cada llave incorrecta se va *eliminando* de selecciones subsecuentes; es decir, (ahora) se está planteando un muestreo aleatorio *sin* reemplazo.

Sea X el número de llaves probadas *antes* de elegir la llave que abre la puerta. Observe que X puede tomar los valores $0, 1, 2, \ldots, n-1$. El caso extremo $X = 0$ se refiere a la situación en que la persona abre la puerta con la *primera* llave seleccionada. En contraste, el caso $X = n - 1$ significa que la persona abre la puerta después de haber intentado con las n llaves (es decir una vez descartadas las $n - 1$ llaves incorrectas). Describiendo cada evento en términos de la intersección de otros eventos y siguiendo el razonamiento de la probabilidad clásica, se tiene que:

$P[X = 0]$

$= P$[elegir la llave correcta en la primera selección (de entre n llaves)]

$$= \frac{1}{n}$$

$P[X = 1]$

$= P$[elegir *una* llave incorrecta (de n–1 posibles de entre n llaves)

y a continuación elegir la llave correcta (de entre n–1 llaves)]

$$= \left(\frac{n-1}{n}\right)\left(\frac{1}{n-1}\right) = \frac{1}{n}$$

$P[X = 2]$

$= P$[elegir *dos* llaves incorrectas (la primera de n–1 posibles de entre n llaves,

y la segunda de n–2 posibles de entre n–1 llaves) y a continuación elegir

la llave correcta (de entre n–2 llaves)]

$$= \left(\frac{n-1}{n}\right)\left(\frac{n-2}{n-1}\right)\left(\frac{1}{n-2}\right) = \frac{1}{n}$$

⋮

$P[X = k]$

$$= \left(\frac{n-1}{n}\right)\left(\frac{n-2}{n-1}\right)\cdots\left(\frac{n-k+1}{n-k+2}\right)\left(\frac{n-k}{n-k+1}\right)\left(\frac{1}{n-k}\right) = \frac{1}{n}$$

$$P[X = n-1] = \frac{(n-1)!}{n(n-1)!} = \frac{1}{n}$$

Entonces X tiene distribución rectangular sobre los enteros $0, 1, 2, \ldots, n-1$ [32] Por lo tanto, el número medio de llaves incorrectas probadas antes de abrir la puerta es

$$E[X]=\sum_{j=0}^{n-1} j\cdot\frac{1}{n}=\frac{1}{n}\sum_{j=1}^{n-1} j=\frac{1}{n}\left[\sum_{j=1}^{n} j-n\right]=\frac{1}{n}\left[\frac{n(n+1)}{2}-n\right]=\frac{n-1}{2}.$$

Es decir, en promedio se requiere haber probado la mitad del total de llaves disponibles.

Recuerde que anteriormente se obtuvo que cuando no se descartaban las llaves incorrectas (*muestreo aleatorio con reemplazo*) *la distribución de X era geométrica con parámetro* $p=1\,n$ Entonces se requería en promedio probar todas las llaves incorrectas, $n-1$, para abrir la puerta (véase *Ejemplos de aplicación* dentro del apartado sobre la distribución geométrica).

❖ *Nota*: En general, la media de una variable aleatoria con distribución rectangular, que asigna el *mismo* valor $(1\cdot n)$ a *todos y cada uno* de los puntos de masa es el punto medio del intervalo de puntos de masa [33]

♦

Relaciones con otras distribuciones

- La distribución rectangular es un caso especial de la distribución beta-binomial. [3.14]

■

[32] Observe que de haber considerado a X como el número de *intentos* realizados para encontrar la llave correcta (es decir, de haberse incluido la selección de la llave correcta en el conteo) se habría obtenido que X se distribuye rectangular sobre los enteros $1, 2, \ldots, n$ (véase expresión 2.7*bis*).

[33] Este resultado es muy ilustrativo para la comprensión del concepto de media de una variable aleatoria. Para una distribución rectangular sobre los enteros $1, 2, \ldots, n$ la media es $n(n+1)/2$.

Otras distribuciones univariadas discretas

La distribución beta-binomial

Definición

Se dice que una v a. X tiene *distribución beta-binomial* con parámetros n_1, n_2 y n_3 si su f d p. está dada por

$$f_X(x) = f_X(x, n_1, n_2, n_3) = \begin{cases} \dfrac{\dbinom{n_2 + x - 1}{x}\dbinom{n_1 + n_3 - x - 1}{n_1 - x}}{\dbinom{n_1 + n_2 + n_3 - 1}{n_1}} & \text{si} \quad x = 0, 1, \ldots, n_1 \\ 0 & \text{e.o.c} \end{cases} \qquad \textbf{[2.8]}$$

donde n_1, n_2 y n_3 son enteros no negativos [34]

Una expresión equivalente a [*2.8*] para la f.d.p. de la distribución beta-binomial es la siguiente:

$$f_X(x) = \begin{cases} \dbinom{n_1}{x} \cdot \dfrac{\Gamma(n_2 + n_3)}{\Gamma(n_2)\Gamma(n_3)} \cdot \dfrac{\Gamma(n_2 + x)\Gamma(n_1 + n_3 - x)}{\Gamma(n_1 + n_2 + n_3)} & \text{si} \quad x = 0, 1, \ldots, n_1 \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.8bis]}$$ [35]

donde $\Gamma(\cdot)$ es la notación de *la función gamma*, definida como $\Gamma(\theta) = \int_0^\infty y^{\theta - 1} e^{-x} dx$ para $\theta > 0$ Para el caso en que θ es un entero no negativo, no es difícil demostrar que $\Gamma(\theta) = (\theta - 1)!$. En el ANEXO 6 se aborda la función gamma con mayor detalle.

---

[34] Tanto la expresión de la f.d.p. como el rango de los parámetros corresponden con lo expuesto en el artículo de Leemis En la bibliografía consultada, solo Mood *et al.* [1974] hacen referencia a esta distribución, presentándola con la expresión [2 *8bis*].

[35] La expresión [*2 8bis*] coincide con aquélla presentada por Mood *et al.* [1974] Sin embargo, estos autores contemplan un rango más amplio para los parámetros En lugar de los tres enteros no negativos n_1, n_2 y n_3, consideran n, α y β, respectivamente, donde n es un entero no negativo, $\alpha > 0$ y $\beta > 0$ Es decir, la forma de la distribución beta-binomial presentada por Leemis es un caso particular de aquélla presentada por Mood *et al* [1974]. Sólo cuando los tres parámetros son enteros no negativos es que las expresiones [*2.8*] y [*2 8bis*] son equivalentes.

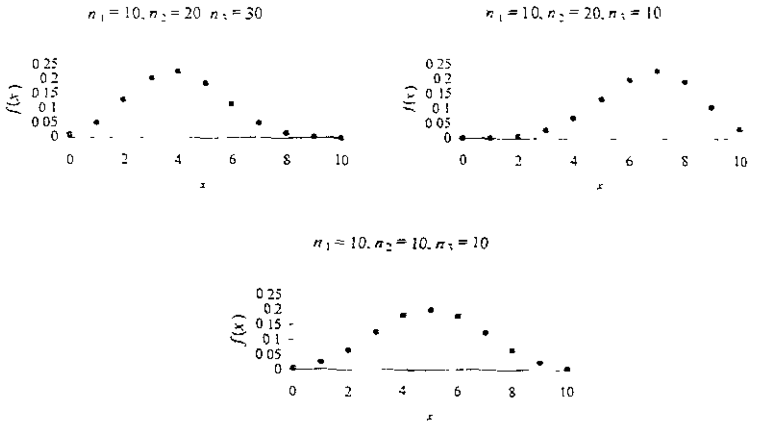


FIGURA 2.7 F.d.p. beta binomial para distintos valores de n_1, n_2 y n_3

Por otra parte, se define a la *función beta*, denotada como $B(\cdot,\cdot)$, por

$$B(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}\,dx \qquad \text{para } a > 0,\ b > 0.$$

Mediante un cambio de variable, se puede demostrar que $B(a,b) = B(b,a)$. La función beta está relacionada con la función gamma de acuerdo con la siguiente fórmula:

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}.$$

Por lo tanto, la distribución beta-binomial puede ser expresada también en términos de funciones beta. Se deduce que es precisamente de esta propiedad y del hecho de que la distribución binomial es una aproximación a la expresión [2.8] de donde la distribución beta-binomial deriva su nombre. Observe que ambas distribuciones tienen los mismos puntos de masa.

Relaciones con otras distribuciones

Como ya se mencionó, la distribución binomial es una aproximación a la distribución beta-binomial. Asimismo, la distribución rectangular es un caso particular de esta última.

1. La distribución binomial es una aproximación a la distribución beta-binomial. [3.5]
2. La distribución rectangular es un caso especial de la distribución beta-binomial. [3.14]

La distribución Weibull discreta

Definición

Se dice que X es una variable aleatoria con *distribución Weibull discreta* con parámetros p y β si su f.d.p. está dada por

$$f_X(x) = f_X(x; p, \beta) = \begin{cases} (1-p)^{x^\beta} - (1-p)^{(x+1)^\beta}, & \text{para} \quad x = 0, 1, .. \\ 0 & \text{e o.c.} \end{cases} \qquad \textbf{[2.9]}$$

donde $0 < p < 1$ y $\beta > 0$

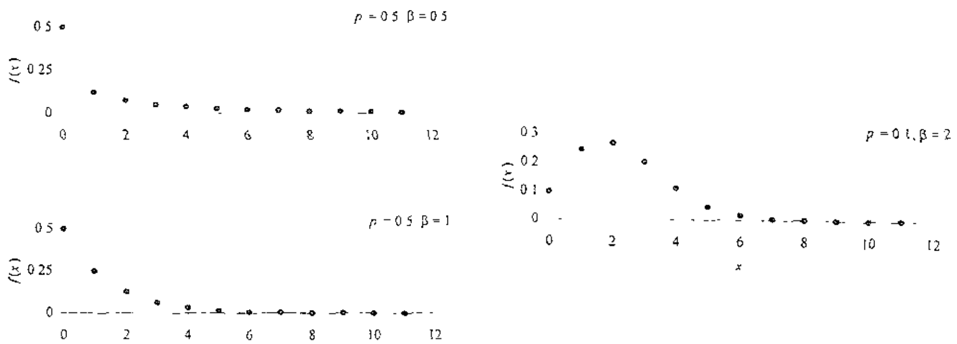


FIGURA 2 8 F d p Weibull discreta para distintos valores de p y β

Leemis presenta la distribución Weibull discreta en su artículo y la relaciona con la distribución geométrica, señalando que la segunda es un caso especial de la primera. Sin embargo, dentro de la bibliografía revisada no se encontró referencia alguna a esta distribución Es por ello que no se incluye información adicional al respecto. Baste decir que al parecer la distribución Weibull discreta está "inspirada" en la similitud (que podrá verificarse más adelante) entre la distribución geométrica y la exponencial, pues esta última es un caso especial de la *distribución Weibull continua*, conocida simplemente como *distribución Weibull.*

En la *Figura 2.8* puede observarse que las tendencias de las densidades Weibull discretas son similares a aquéllas de las densidades Weibull en la versión continua (véase *Figura 2.21*) Observe que por la forma cómo está definida la densidad Weibull discreta, el valor de ésta en cero ($P[X = 0]$) es precisamente el valor del parámetro p; de tal forma que conforme más cercano esté ese valor al 1, más rápidamente la función de densidad converge a cero.

Relaciones con otras distribuciones

- La distribución geométrica es un caso especial de la distribución Weibull discreta. [3.13]

■

En la siguiente sección se presentan las distribuciones continuas del esquema de Leemis. Cuatro de ellas son casos particulares: la distribución normal estándar, la distribución Erlang, la distribución Cauchy estándar y la distribución uniforme estándar. Éstas se incluyen dentro de los apartados dedicados a las versiones generales: la distribución normal, la distribución gamma, la distribución Cauchy y la distribución uniforme, respectivamente.

DISTRIBUCIONES UNIVARIADAS CONTINUAS

La distribución beta

Definición

Se dice que una v.a. X tiene *distribución beta* con parámetros β y γ si su f.d.p está dada por

$$f_X(x) = f_X(x;\beta,\gamma) = \begin{cases} \dfrac{\Gamma(\beta+\gamma)}{\Gamma(\beta)\Gamma(\gamma)} \cdot x^{\beta-1}(1-x)^{\gamma-1} & \text{si} \quad 0 < x < 1 \\ 0 & \text{e.o.c.} \end{cases} \qquad [2.10]$$

donde $\beta > 0$ y $\gamma > 0$.

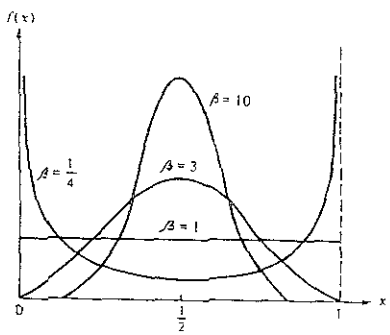


FIGURA 2 9. F d.p. beta para distintos valores de β y γ, con $\beta = \gamma$

Puesto que la función beta es $B(\beta,\gamma) = \dfrac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta+\gamma)}$, la expresión [2.10] puede escribirse en términos de esta función, justificando así el nombre de la distribución, como sigue:[36]

[36] Para mayor referencia sobre la funcion beta, vease sección *La distribución beta-binomial*

$$f_X(x) = f_X(x;\beta,\gamma) = \begin{cases} \dfrac{1}{B(\beta,\gamma)} \cdot x^{\beta-1}(1-x)^{\gamma-1} & \text{si} \quad 0 < x < 1 \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.10}\textbf{\textit{bis}}\textbf{]}.$$

La densidad beta puede tomar una amplia variedad de formas como se muestra en las *Figuras* 2.9 y 2.10

Cuando $\beta = \gamma$ la densidad beta *es simétrica alrededor de* ½, dando más y más peso a las regiones alrededor de ½, conforme se incrementa el valor común β (véase *Figura* 2.9). Cuando $\beta < \gamma$ la densidad está sesgada hacia la izquierda (es decir, los valores más pequeños se vuelven más parecidos, véase *Figura 2.10*); y está sesgada hacia la derecha cuando $\beta > \gamma$. La gráfica tiene forma de U cuando $(\beta - 1)$ y $(\gamma - 1)$ son negativos y tiene forma de J si sólo uno de ellos es negativo. La función de densidad alcanza su máximo en $x = (\beta - 1)/(\beta + \gamma - 2)$ cuando $(\beta - 1)$ y $(\gamma - 1)$ son positivos; y el mismo punto es su mínimo cuando estos son negativos.

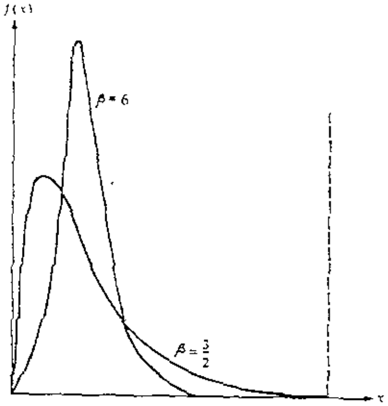


FIGURA 2.10. F.d p. beta cuando $\dfrac{\beta}{\beta+\gamma} = \dfrac{1}{20}$.

Ejemplos de aplicación

La diversidad de formas que puede tomar la densidad beta hace que esta distribución pueda ser utilizada para modelar un experimento para el cual una de las formas sea apropiada De hecho, puede ser utilizada para modelar un fenómeno aleatorio cuyo conjunto de posibles valores sea algún intervalo finito $[c,d]$, el cual puede ser transformado en el intervalo $[0,1]$ haciendo que c denote el origen y tomando $d-c$ como la unidad de medición

La distribución beta se emplea frecuentemente, por ejemplo, cuando una variable aleatoria toma valores que son porcentajes o proporciones, o cuando se trata de algún fenómeno físico del tipo continuo que toma valores entre 0 y 1. Existe una aplicación directa de la distribución beta al análisis de los procesos de Markov cuando las probabilidades de transición son "inciertas".[37]

Propiedades

- La función de distribución acumulativa de una variable aleatoria beta es

$$F_X(x;\beta,\gamma)=I_{(0,1)}(x)\int_0^x \frac{1}{B(\beta,\gamma)}\cdot u^{\beta-1}(1-u)^{\gamma-1}\,du+I_{[1,\infty)}(x);$$

  y se le conoce como la *función beta incompleta* Se han elaborado un buen número de tablas con sus valores, pues no puede ser calculada directamente cuando β y γ son grandes Tsokos [1972] menciona que la función beta incompleta y la función binomial acumulativa están relacionadas como sigue:

$$\sum_{x=\beta}^{n}\binom{n}{x}p^x(1-p)^{n-x}=\frac{\Gamma(n+1)}{\Gamma(\beta)\Gamma(n-\beta+1)}\cdot\int_0^x u^{\beta-1}(1-u)^{n-\beta}\,du,\qquad 0<x<n.$$

- Para la distribución beta es más sencillo calcular los momentos de manera directa mediante la definición, en lugar de recurrir a la función generadora de momentos cuyo cálculo es complejo

[37] Johnson y Kotz [1970] citan una referencia bibliográfica de Silver [1963] acerca de esto: Silver, E. A (1963), *Markovian decision processes with uncertain transition probabilities or rewards*, Operation Research Center, M.I.T., Technical Report, No. 1

Relaciones con otras distribuciones

1 La distribución normal es una aproximación a la distribución beta [3.24]

2. La transformación $-\log X$ de una v.a. beta se distribuye exponencial. [3.43]

3. La distribución arcoseno es un caso especial de la distribución beta. [3.58]

4. La distribución uniforme $(0,1)$ es un caso especial de la distribución beta. [3.59]

5. El cociente de una v.a. gamma entre la suma de dos v.a.i gamma se distribuye beta. [3.60]

■

La distribución Cauchy

Definición

Se dice que una v.a X tiene *distribución Cauchy* con parámetros a y α si su f.d.p. está dada por

$$f_X(x)=f_X(x;\mathrm{a},\alpha)=\frac{1}{\alpha\pi\left[1+\left(\frac{x-\mathrm{a}}{\alpha}\right)^2\right]} \qquad \text{para} \quad -\infty<x<\infty \qquad \textbf{[2.11]}$$

donde $-\infty<\mathrm{a}<\infty$ y $\alpha>0$

- Una *forma estándar* de la distribución Cauchy se obtiene fijando $\mathrm{a}=0$ y $\alpha=1$ La función de densidad de probabilidad estándar está dada por $\frac{1}{\pi(1+x^2)}$ para todo $-\infty<x<\infty$ (***distribución Cauchy estándar***).

Aunque a ojo la distribución Cauchy no parece muy distinta de la normal (véase *Figura 2.11*), en realidad hay una gran diferencia entre ellas. Si bien la densidad Cauchy es simétrica alrededor del parámetro α, su media y cualquier otro momento más elevado no existen, ($E[X]=\infty$ y por ende tampoco existe la varianza para la distribución Cauchy).

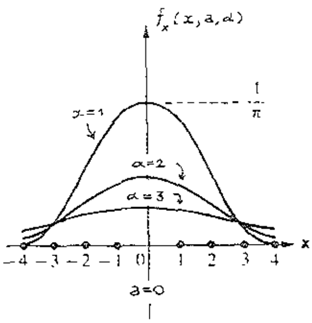


FIGURA 2.11 F.d.p. Cauchy para a = 0 y distintos valores de α

Los parámetros a y α son parámetros de localización y de escala, respectivamente; α es la mediana de la distribución y a $\pm\ \alpha$ son los cuartiles superior e inferior. La función de densidad tiene puntos de inflexión en a $\pm\ \alpha/\sqrt{3}$.

La función de distribución acumulativa Cauchy es

$$F_X(x) = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-a}{\alpha}\right)$$

y por ende $\frac{1}{2} + \frac{1}{\pi}\arctan(x)$ es la función de distribución acumulativa Cauchy estándar. Puede verificarse que los valores de $F_X(x)$ en los puntos de inflexión son 0.273 y 0.727, comparados con 0.159 y 0.841 para la distribución normal.

La diferencia más notable entre las distribuciones normal y Cauchy radica en las colas más largas y más pesadas de la segunda (véase *Figura* 2.12).

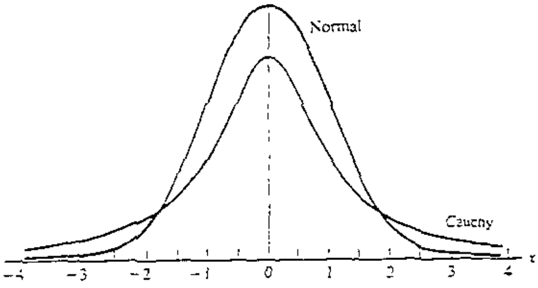


FIGURA 2.12 *Comparación entre las densidades* normal estandar y Cauchy estándar

Ejemplos de aplicación

Una de las situaciones en las que surge la densidad Cauchy es en la descripción de la distribución del punto de intersección, X, de una línea recta fija con otra línea recta variable, orientada aleatoriamente en dos dimensiones a través de un punto fijo A. *Imagine que una lámpara (localizada en el punto A) que emite* un delgado rayo de luz es girada alrededor de su centro, que está localizado a una unidad de distancia del eje x (véase *Figura 2.13*). Cuando la lámpara detiene su giro, considere el punto X en el cual el rayo intersecta el eje x (si el rayo no está apuntando hacia el eje x se repite el experimento).

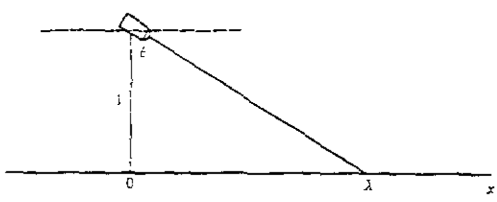


FIGURA 2.13 Aplicación de la distribución Cauchy estandar

Como se indica en la *Figura 2.13*, el punto X está determinado por el ángulo θ entre la lámpara y el eje y, el cual dada la situación física parece tener distribución uniforme entre $-\pi/2$ y $\pi/2$.[38] Entonces, la función de distribución de X está dada por

[38] La distribución uniforme se presenta mas adelante

$$\begin{aligned} F_X(x) &= P[X \le x] \\ &= P[\tan\theta \le x] \\ &= P[\theta \le \tan^{-1}(x)] \\ &= P[\theta \le \arctan(x)] \\ &= \tfrac{1}{2} + \tfrac{1}{\pi}\arctan(x) \end{aligned}$$

que corresponde a la función de distribución acumulativa de una variable aleatoria Cauchy estándar. La última igualdad se sigue puesto que θ, al ser uniforme en ($-\pi/2$ y $\pi/2$), conduce a que

$$P[\theta \le c] = \frac{c-(-\pi/2)}{\pi} = \frac{1}{2} + \frac{c}{\pi}, \qquad -\frac{\pi}{2} < c < \frac{\pi}{2}.$$

◈

Relaciones con otras distribuciones

La distribución Cauchy estándar es una distribución t de Student con un grado de libertad. Es también la distribución del cociente U/V, donde U y V son variables normal estándar independientes. Como U y V tienen la misma distribución, entonces la distribución de U/V y V/U debe ser también la misma; y, por lo tanto, si X es Cauchy estándar entonces lo es también $1/X$. En general, el recíproco de una variable aleatoria Cauchy es también Cauchy. Además, esta distribución cumple con la propiedad reproductiva. Finalmente, toda variable Cauchy (a,α) puede llevarse a la forma estándar y viceversa. Por lo tanto, en el Capítulo 3 se verificará que:

1. La distribución Cauchy estándar es un caso especial de la distribución t de Student. [3.49]
2. Toda v.a Cauchy estándar puede llevarse a Cauchy (a,α). [3.51]
3. Una variable aleatoria Cauchy (a,α) es Cauchy estándar cuando a $= 0$ y $\alpha = 1$. [3.52]
4. Una transformación lineal de una v a. Cauchy es también Cauchy [3.53]
5. El recíproco de una variable aleatoria Cauchy es también Cauchy. [3.54] y [3.55]
6. La distribución Cauchy (a,α) satisface la propiedad reproductiva. [3.56]
7. El cociente de dos v.a i.i.d normal estándar se distribuye Cauchy estándar [3 57]

▩

La distribución exponencial

Definición

Se dice que X es una v.a. con *distribución exponencial* con parámetro α si su f.d.p. está dada por

$$f_X(x) = f_X(x;\alpha) = \begin{cases} \frac{1}{\alpha} \cdot e^{-x/\alpha} & \text{si} \quad x > 0 \\ 0 & \text{e.o.c.} \end{cases} \qquad [2.12]$$

donde $\alpha > 0$.[39]

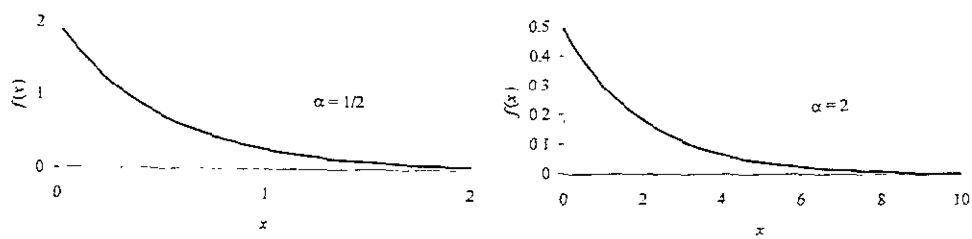


FIGURA 2.14. F.d.p exponencial para distintos valores de α

Ejemplos de aplicación

La distribución exponencial surge de manera frecuente como la distribución del *tiempo transcurrido* hasta que un evento específico ocurra. Por ejemplo, el tiempo, comenzando desde este momento, hasta que ocurra un temblor, hasta que estalle una nueva guerra, hasta que una llamada telefónica recibida corresponda a un número equivocado o la duración misma de una llamada telefónica son todas variables aleatorias que tienden a tener en la práctica distribuciones exponenciales. Otro

[39] Autores como Johnson y Kotz proporcionan una expresión biparamétrica de la f.d.p. exponencial, a saber $f_X(x) = (1/\alpha)^*\exp(-(x-\theta)/\alpha)$, donde $x > \theta$ y $\alpha > 0$, y consideran que un caso especial de ésta lo determina $\theta = 0$, mismo que se conoce como distribución exponencial *uni-paramétrica*, y el cual es mucho más frecuente de encontrar.

Por otro lado, es muy común encontrar que en la expresión de la densidad exponencial se maneje el parámetro α de manera directa, en lugar de hacerlo en términos de su inverso; es decir como $f_X(x) = \alpha\cdot\exp(-\alpha\cdot x)$, $x > 0$.

evento asociado con esta distribución es la *falla* en el funcionamiento de algún objeto, la distribución exponencial puede utilizarse como modelo para los *tiempos de vida* de diversos objetos

✂ *Propiedades*

- Al igual que la distribución geométrica (*Teorema 2.4 1*), la exponencial cumple la propiedad de la "ausencia de memoria"

Teorema 2.12.1
La distribución exponencial no tiene memoria

Sea X una v a. con distribución exponencial con parámetro α, entonces

$$P[X > s + t \mid X > s] = P[X > t], \qquad \text{para } s,\ t > 0$$

Demostración

Se requiere calcular primero $P[X > x]$:

$$\begin{aligned} P[X > x] &= 1 - P[X \le x] = 1 - F_X(x) \\ &= 1 - \int_0^x \frac{1}{\alpha} e^{-y/\alpha}\, dy = 1 + \left[e^{-y/\alpha} \Big|_0^x \right] \\ &= 1 + \left[e^{-x/\alpha} - 1 \right] = e^{-x/\alpha} \end{aligned}$$

Así,

$$\begin{aligned} P[X > s + t \mid X > t] &= \frac{P[X > s + t, X > t]}{P[X > t]} = \frac{P[X > s + t]}{P[X > t]} \\ &= \frac{e^{-(s+t)/\alpha}}{e^{-t/\alpha}} = e^{-s/\alpha} = P[X > s] \end{aligned}$$

◆

Una forma de interpretar este resultado es pensando a X como el tiempo de vida de alguna pieza o instrumento, entonces la probabilidad condicional de que la pieza durará $s + t$ unidades de tiempo, *dado* que ha durado ya t unidades de tiempo, es la misma que la probabilidad *inicial* de perdurar s unidades de tiempo. En otras palabras, la distribución del tiempo de vida de una pieza "vieja" es la misma que la de una pieza "nueva", lo que significa que la pieza no está sujeta al deterioro (*no "guarda memoria"* de que ya ha estado en uso por un tiempo t).

Más aún, Ross [1988] demuestra que la distribución *exponencial es la única* distribución *continua que posee esta propiedad.*[5]

Ejemplo Considere que la sucursal de un banco es atendida por dos cajeros. Suponga que al entrar la persona C a la sucursal hay dos personas dentro, A y B, siendo atendidas por los cajeros (cada uno atiende a una persona). Suponga además que la persona C será atendida por uno de los cajeros tan pronto como la persona A o B se vayan. Si el tiempo que le toma a un cajero atender a un cliente se distribuye exponencialmente con parámetro α, ¿cuál es la probabilidad de que, de los res clientes, la persona C sea la última en irse del banco?

La respuesta se obtiene razonando de la siguiente forma. Considere el tiempo en el que la persona C encuentra primero un cajero libre. En ese momento una de las otras personas se habrá ido apenas y la otra estará siendo atendida todavía Como la distribución exponencial no tiene memoria, se sigue que el tiempo adicional que esta persona (A o B) permanecerá en la *sucursal se distribuye* exponencialmente con parámetro α. Es decir, es como *si el servicio para esta persona apenas hubiera comenzado justo en ese momento. Por lo tanto, por simetría* la probabilidad de que la *persona restante termine primero* antes que la persona C es ½.

◆

Relaciones con otras distribuciones

La distribución exponencial está vinculada con la distribución gamma de manera recíproca, lo cual está ampliamente difundido en la literatura sobre variables aleatorias: por un lado, la exponencial es un caso especial de la gamma y, por otro lado, la gamma se obtiene de la suma de exponenciales. Atendiendo al esquema de Leemis, se introducirá la distribución Erlang como *intermediaria para* verificar estas relaciones. *Como la distribución Ji-cuadrada es también un caso especial de la distribución gamma*, la Ji-cuadrada y la exponencial están relacionadas también Asimismo, la distribución exponencial guarda relación unilateral con la distribución uniforme estándar y recíproca con las distribuciones Laplace, Weibull y Rayleigh siendo resultado de una transformación. Además, la distribución exponencial se relaciona con ella misma a través de la mínima estadística de orden de una muestra aleatoria.

[5] Véase Ross [1988], p. 175.

1. La distribución exponencial es un caso especial de la distribución gamma [3.28]
2. La mínima estadística de orden de una muestra aleatoria exponencial es también exponencial [3 31]
3. La distribución exponencial es un caso especial de la distribución Weibull [3 32]
4. La distribución exponencial es un caso especial de la distribución Erlang [3 33]
5. La suma de v a.i.i d. exponencial tiene distribución Erlang [3.34]
6. La distribución Ji-cuadrada es un caso especial de la distribución exponencial [3.35]
7. La distribución exponencial es un caso especial de la distribución Ji-cuadrada [3.45]
8. La raíz cuadrada de una v a. exponencial se distribuye Rayleigh [3.36]
9. El cuadrado de una v a Rayleigh se distribuye exponencial. [3 37]
10. La transformación $X^{1/\beta}$ de una v.a exponencial se distribuye Weibull [3.38]
11. La transformación X^{β} de una v.a. Weibull se distribuye exponencial. [3.39]
12. La diferencia de dos v.a i. exponenciales se distribuye Laplace. [3.40]
13. El valor absoluto de una v.a Laplace se distribuye exponencial [3 41]
14. La transformación $-\alpha \log X$ de una v.a. uniforme estándar se distribuye exponencial [3.42]
15. La transformación $-\log X$ de una v.a. beta se distribuye exponencial. [3.43]

❖ *Sobre la similitud entre la Geométrica* (p) *y la Exponencial* (α). La similitud entre las distribuciones exponencial y geométrica puede apreciarse desde distintos ángulos Por sí mismas, observe la similitud de las tendencias que siguen sus densidades (*Figuras 2.4* y *2.14*). Por otra parte, comparten la propiedad de la "ausencia de memoria" (*Teoremas 2 4.1* y *2 14 1*), así como la propiedad de que la mínima estadística de orden de una muestra aleatoria de estas distribuciones tiene la misma distribución de origen con un cambio en el parámetro. A nivel de interpretación, ambas pueden utilizarse para modelar *tiempos de espera*.

■

La distribución F

Definición

Se dice que una v.a X tiene *distribución* F con parámetros n_1 y n_2 si su f.d.p está dada por

$$f_X(x) = f_X(x; n_1, n_2) = \begin{cases} \dfrac{\Gamma\left(\frac{n_1+n_2}{2}\right)\cdot\left(\frac{n_1}{n_2}\right)^{n_1/2} x^{(n_1/2)-1}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)\cdot\left[1+\left(\frac{n_1}{n_2}\right)x\right]^{(n_1+n_2)/2}} & \text{si} \quad x > 0 \\ 0 & \text{e.o.c.} \end{cases} \qquad [2.13]$$

donde n_1 y n_2 son enteros positivos.[41]

❖ *Sobre la notación*: Se suele leer que X tiene *distribución* F (o se *distribuye* F) *con* n_1 y n_2 *grados de libertad*, lo cual se puede denotar como $X \sim F_{(n_1,n_2)}$.

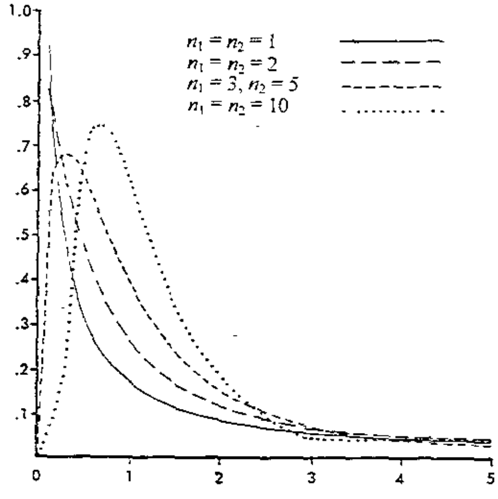


FIGURA 2.15. F.d.p. F para distintos valores de n_1 y n_2

El orden en que se dan los grados de libertad es importante puesto que la densidad de la distribución F no es simétrica respecto a n_1 y n_2. *Primero* se cita el número de grados de libertad en el *numerador* del cociente n_1/n_2 que aparece en la expresión [2.13].

[41] *La distribución F recibe su nombre en honor a Sir Ronald Fisher (1890-1862), considerado uno de los fundadores* de la estadística moderna por sus muchas importantes contribuciones

Ejemplos de aplicación

La aplicación más común de la distribución F en el trabajo estadístico se ubica en las pruebas estándar asociadas con el análisis de varianza pues, como se verá más adelante, esta distribución resulta del cociente de dos variables ji-cuadrada, $\chi^2_{(n)}$ y $\chi^2_{(m)}$, divididas a su vez entre sus grados de libertad correspondientes Asimismo, se aplica en la prueba de igualdad de varianzas de dos poblaciones, como se esboza a continuación.

Ejemplo Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una población normal $N(\mu_X, \sigma_X^2)$, y sea Y_1, $Y_2, \ldots, Y_m$ una muestra aleatoria de una población normal, independiente de la anterior, $N(\mu_Y, \sigma_Y^2)$. Suponga que se está interesado en comparar la variabilidad de las poblaciones. Entonces una cantidad de interés sería la razón σ_X^2 / σ_Y^2. Alguna información acerca de esta razón está contenida en S_X^2 / S_Y^2, la razón de las varianzas muestrales [42] La distribución F permite comparar estas cantidades al proporcionar la distribución de

$$\frac{S_X^2 / S_Y^2}{\sigma_X^2 / \sigma_Y^2} = \frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2} = W$$

Un examen de esta última expresión muestra cómo se deriva la distribución F. Las razones S_X^2 / σ_X^2 y S_Y^2 / σ_Y^2 son variables ji-cuadrada, $\chi^2_{(n)}$ y $\chi^2_{(m)}$, respectivamente, y son independientes. La cantidad $W = (S_X^2 / \sigma_X^2)/(S_Y^2 / \sigma_Y^2)$ tiene distribución $F_{(n-1,m-1)}$. Es posible calcular el valor esperado de W:

$$\begin{aligned}
E(W) &= E\left(\frac{\chi^2_{n-1}/(n-1)}{\chi^2_{m-1}/(m-1)}\right) && \text{(por definición)} \\
&= E\left(\frac{\chi^2_{n-1}}{n-1}\right)E\left(\frac{m-1}{\chi^2_{m-1}}\right) && \text{(independencia)} \\
&= \left(\frac{n-1}{n-1}\right)\left(\frac{m-1}{m-3}\right) && \text{(cálculos con ji - cuadradas)} \\
&= \frac{m-1}{m-3}.
\end{aligned}$$

[42] Para mayor referencia sobre S^2, véase el apartado sobre la distribución ji-cuadrada

Note que esta última expresión es finita y positiva sólo si $m > 3$ Se tiene entonces que

$$E\left(\frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2}\right) = \frac{m-1}{m-3}.$$

Acudiendo a que $\frac{m-1}{m-3}$ es el *valor esperado* de $\frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2}$ y considerando una m suficientemente grande, se tiene que:

$$\frac{S_X^2 / \sigma_X^2}{S_Y^2 / \sigma_Y^2} \approx \frac{m-1}{m-3} \approx 1,$$

como cabía esperar.

◆

Relaciones con otras distribuciones

La distribución F se vincula con la distribución Ji-cuadrada de manera bilateral, con la distribución t de Student de forma unilateral y con ella misma a través de transformaciones diversas.

1. El cociente de dos v.a.i. ji-cuadrada divididas entre sus respectivos grados de libertad tiene distribución F [3.47]
2. La distribución ji-cuadrada es una aproximación para la distribución de n_1X, cuando X es una v.a. F. [3.48]
3. El cuadrado de una v.a. t de Student tiene distribución F. [3.50]
4. El recíproco de una v.a con distribución F tiene también distribución F. [3.68]

■

La distribución gamma

Definición

Se dice que X es una v.a con *distribución gamma* con parámetros α y β si su f.d p. está dada por

$$f_X(x) = f_X(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta} & \text{si} \quad 0 < x < \infty \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.14]}$$ [43]

donde $\alpha, \beta > 0$ y $\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy$ es la *función gamma*.[44]

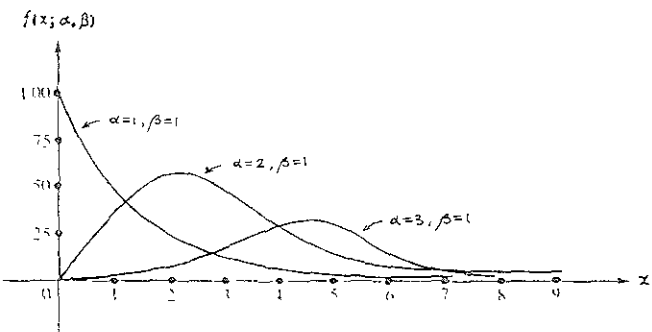


FIGURA 2 16 F d p. gamma para distintos valores de α y β.

---

[43] La expresión utilizada aquí para la densidad gamma difiere de la presentada en el artículo de Leemis en tanto que *intercambia* las posiciones de α y β $f_X(x;\alpha,\beta) = \dfrac{1}{\Gamma(\beta)\alpha^{\beta}} x^{\beta-1} e^{-x/\alpha}$, para $0 < x < \infty$. Las expresiones son indistintas siempre y cuando se aclare previamente el papel de los parámetros; se adoptó la expresión [2.14] para hacer homogénea la presentación con la literatura. Otra variante en la forma de presentar la densidad gamma, como lo hacen Mood *et al.* [1974] y Ross [1988], se justifica a partir de cómo se define la densidad exponencial; ya sea $f_X(x,\beta) = \beta e^{-\beta x}$, para $0 < x < \infty$, como lo hacen estos autores, o $f_X(x;\beta) = \dfrac{1}{\beta} e^{-x/\beta}$, para $0 < x < \infty$, como se hace en el artículo de Leemis y en este trabajo Así, en los textos de Mood *et al.* [1974] y Ross [1988] se presenta la densidad gamma como:

$$f_X(x) = f_X(x;\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} = \begin{cases} \dfrac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} & \text{si} \quad 0 < x < \infty \\ 0 & \text{e.o.c.} \end{cases}$$

[44] Para mayor referencia sobre la función gamma puede consultarse el ANEXO 6

Como puede apreciarse en la *Figura 2.16*, el parámetro α es un parámetro de forma (es el que más influye sobre el "pico" de la distribución), mientras que β es un parámetro de escala (es el que más influye sobre la dispersión de la distribución)

- Cuando α es un entero positivo n, la distribución gamma recibe el nombre de ***distribución Erlang***, que se define como

$$f_X(x) = f_X(x; n, \beta) = \begin{cases} \dfrac{1}{(n-1)!\,\beta^n} x^{n-1} e^{-x/\beta} & \text{si} \quad 0 < x < \infty \\ 0 & \text{e.o.c.} \end{cases}$$

ya que para un entero positivo $\Gamma(n) = (n-1)!$

Algo de historia

Laplace obtuvo la distribución gamma como la distribución de la "constante de precisión" (h = ½ σ^{-2}) dados los valores de n variables independientes normales con media cero y *desviación estándar σ (suponiendo una distribución uniforme previa para h).*[45] Posteriormente, la distribución gamma apareció otra vez en 1900 en un trabajo de Pearson como la distribución aproximada de las "estadísticas ji-cuadrada", utilizadas para diversas pruebas en tablas de contingencia.[46]

Ejemplos de aplicación

Como aplicaciones teóricas, la distribución gamma surgió de manera natural en la teoría asociada con las variables aleatorias distribuidas normalmente, como la suma de cuadrados variables independientes normal estándar. Además, en ciertas situaciones, la distribución gamma puede ser utilizada en lugar de la distribución normal, pues si se elige α suficientemente grande la distribución gamma imita de manera cercana a la normal.

[45] Laplace, P.S. (1836), *Théorie Analytique des Probabilités* (suplemento de la tercera edición) Esta es una referencia de Johnson y Kotz [1970].

[46] Pearson, K. (1900), *On a criterion that a given system of deviations from the probable in the case of a correlated system of variables in such that it can be reasonably supposed to have arisen from random sampling*, Philosophical Magazine, 5ª serie, 50, pp. 157-175. Esta es una referencia de Johnson y Kotz [1970].

Por lo que toca a las aplicaciones prácticas, las distribuciones gamma han sido utilizadas para ajustar distribuciones exponenciales en la representación de los tiempos de vida en situaciones de "prueba de vida (o de duración)", aunque fueron las distribuciones Weibull las que se popularizaron para este propósito. El hecho de que la suma de variables aleatorias independientes con distribución exponencial tenga distribución gamma condujo a que ésta apareciera en la teoría de contadores aleatorios y de otros temas vinculados con procesos aleatorios en el tiempo, en particular procesos de precipitación meteorológica

Relaciones con otras distribuciones

El vínculo de la distribución gamma con las distribuciones exponencial y ji-cuadrada sea tal vez la más extendida en la literatura. Sin embargo, la distribución gamma también está asciada con las distribuciones normal y beta.

1	La distribución gamma satisface la propiedad reproductiva.	[3.26] y [3.26 *bis*]
2.	La distribución Erlang es un caso especial de la distribución gamma.	[3.27]
3	La distribución exponencial es un caso especial de la distribución gamma.	[3.28]
4	La distribución ji-cuadrada es un caso especial de la distribución gamma.	[3.29]
5	El cuadrado de una v.a. normal estándar se distribuye gamma	[3 30]
6	La distribución normal es una aproximación de la distribución gamma.	[3.23]
7	El cociente de una v a gamma entre la suma de dos v.a.i. gamma se distribuye beta.	[3 60]

La distribución ji-cuadrada

Definición

Se dice que una v a. X tiene *distribución ji-cuadrada* con parámetro n si su f.d.p. está dada por

$$f_X(x) = f_X(x,n) = \begin{cases} \dfrac{1}{2^{n/2}\Gamma\left(\frac{n}{2}\right)} \cdot x^{n/2-1} e^{-x/2} & \text{si} \quad x > 0 \\ 0 & \text{e.o.c} \end{cases} \qquad \textbf{[2.15]}$$

donde n es un entero positivo

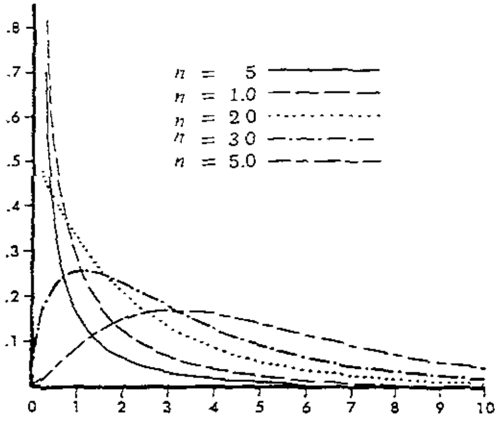


FIGURA 2 17 F.d.p. ji-cuadrada para distintos valores de n

- *Sobre la notación*: En el lenguaje estadístico, al parámetro n se le refiere como "grados de libertad"; es decir, se dice que X tiene distribución ji-cuadrada con n grados de libertad, lo cual se denota como $\chi^2_{(n)}$.

Ejemplos de aplicación

La distribución ji-cuadrada está vinculada fuertemente con la distribución gamma. Es uno de los dos casos especiales más importantes de la distribución gamma (el otro caso es el de la distribución exponencial). Además, mediante las distribuciones ji-cuadrada se obtienen las mejores aproximaciones para las integrales de probabilidad de las distribuciones gamma. Por otra parte, la distribución ji-cuadrada juega un papel importante en la inferencia estadística, especialmente en el muestreo de poblaciones con distribución normal: facilita la deducción de la distribución de S^2, que es uno de los estimadores del parámetro desconocido σ^2 de una normal. Esto se expone a continuación.

Ejemplo. Sea $X_1, X_2 \ldots, X_n$ una muestra aleatoria de una distribución normal (μ, σ^2). Entonces $\overline{X}$ es un estimador de la media desconocida y S^2 lo es de la varianza; donde

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 .$$

La expresión $U = (n-1)\, S^2/\sigma^2$ tiene distribución ji-cuadrada con $n-1$ grados de libertad Como S^2 es una función lineal de U, su densidad puede ser obtenida a partir de la densidad de U [12]

Relaciones con otras distribuciones

La distribución ji-cuadrada se vincula con las distribución normal y F a través de transformaciones multivariadas Con la segunda guarda una relación de ida y vuelta, siendo por un lado una distribución límite. Por otra parte, la distribución ji-cuadrada es un caso especial de la distribución gamma y, en consecuencia, de la distribución exponencial; con esta última la relación es recíproca. Asimismo, la distribución ji-cuadrada satisface la propiedad reproductiva. Siempre que sea pertinente. en la verificación de estas relaciones se hará uso de que la ji-cuadrada es también una gamma

1 La distribución ji-cuadrada es un caso especial de la distribución gamma [3 29]

2 La distribución ji-cuadrada satisface la propiedad reproductiva. [3.44] y [3.44*bis*]

3. La distribución ji-cuadrada es un caso especial de la distribución exponencial. [3.35]

4. La distribución exponencial es un caso especial de la distribución ji-cuadrada [3 45]

5 La suma de cuadrados de v.a i normal estándar se distribuye ji-cuadrada. [3.46]

5 El cociente de dos v a i. ji-cuadrada divididas entre sus respectivos grados de libertad tiene distribución F. [3 47]

6. La distribución ji-cuadrada es una aproximación para la distribución de n_1X, cuando X es una v.a F. [3 48]

[12] La densidad de S^2 es

$$f_{S^2}(y) = \left(\frac{n-1}{2\sigma^2}\right)^{(n-1)/2} \frac{1}{\Gamma[(n-1)/2]}\, y^{(n-3)/2} e^{-(n-1)y/2\sigma^2}, \qquad \text{para} \quad 0 < y < \infty$$

Una prueba de este resultado puede consultarse en Mood *et al* [1974], pp. 243–245 Por su parte, Casella y Berger [1990], pp 221–223 presentan una discusión al respecto

La distribución normal

Definición

Se dice que X es una v.a. con *distribución normal* con parámetros μ y σ^2 si su f.d p. está dada por

$$f_X(x) = f_X(x;\, \mu, \sigma^2) = \frac{1}{2\pi\sigma} \exp\left\{ -\frac{(x-\mu)^2}{2\sigma^2} \right\} \qquad \text{para} \quad -\infty < x < \infty \qquad \textbf{[2.16]}$$

donde $-\infty < \mu < \infty$ y $\sigma > 0$.

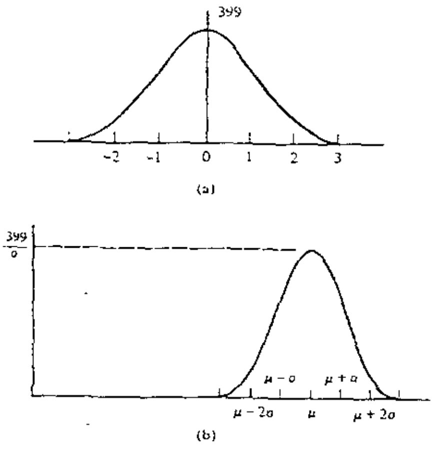


FIGURA 2.18. F.d.p. normal (a) con $\mu = 0$ y $\sigma^2 = 1$, y (b) μ y σ^2 arbitrarios.

La densidad normal es una curva con forma de campana simétrica alrededor de μ. Los valores μ y σ^2 corresponden a la media y a la varianza de X, respectivamente.

❖ *Sobre la notación.* Para denotar que una v.a. X se distribuye normalmente con media μ y varianza σ^2, se escribe $X \sim N(\mu,\sigma^2)$. En particular, si $Z \sim N(0,1)$ se dice que Z es una variable aleatoria *normal estándar* (véase *Figura 2.18a*). La densidad de la ***distribución normal estándar*** está dada por:

$$f_Z(x) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \qquad \text{para} \quad -\infty < z < \infty$$

Existe también una notación especial para denotar la función de distribución de una variable aleatoria normal estándar; a saber, se utiliza $\Phi(z)$ como símbolo para representar $P[Z \leq z]$, es decir

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-\frac{t^2}{2}\right) dt$$

Algo de historia

En 1733 el matemático francés Abraham de Moivre descubrió la distribución normal como resultado de la aproximación de las probabilidades asociadas con variables aleatorias binomiales cuando el parámetro binomial n era grande. Posteriormente, este resultado fue extendido por otros matemáticos como Pierre Laplace, quien en 1774 estudió las propiedades matemáticas de la densidad normal. Actualmente ese resultado es el *Teorema de Límite Central*, el cual se presenta dentro del apartado sobre las aplicaciones de la distribución normal.

Por un error histórico[48], el descubrimiento de la distribución normal le fue atribuido a Gauss (1777-1855) por haber sido el primero en referirse a ella en un documento escrito en 1809. Es por ello que en ocasiones se hace referencia a esta distribución como *distribución Gaussiana*. La función fue estudiada en el siglo XIX por científicos que observaron que los errores de las mediciones seguían un patrón que podía ser aproximado de manera cercana por una función que denominaron la "curva normal de error"

[48] Así lo llama Tsokos [1974].

Ejemplos de aplicación

Las propiedades matemáticas únicas de la densidad normal, así como el hecho de que pueda ser aplicada para modelar una muy amplia variedad de fenómenos aleatorios hacen que la distribución normal sea para muchos la función de probabilidad más importante. Cantidades tan disímiles como la altura de un humano, la velocidad en cualquier dirección de una molécula de gas o el error cometido al medir una cantidad física presentan un "comportamiento normal". Aunque son pocos los fenómenos que obedecen de manera precisa la ley de probabilidad normal, las leyes que en efecto obedecen estos fenómenos pueden, bajo ciertas condiciones, ser aproximadas cercanamente por la densidad normal. Como podrá verificarse en Capítulo 3, algunas de las densidades que pueden ser aproximadas por la densidad normal bajo ciertas condiciones son la binomial, la Poisson, la gamma y la beta.

Ejemplo[49]. A continuación se presenta un ejemplo práctico de la densidad normal relacionado con calificaciones escolares.

Se considera que un examen escolar es adecuado –en el sentido de que proporciona una calificación válida a todos y cada uno de los alumnos– si los puntajes obtenidos en la prueba realizada pueden ser aproximados por una densidad normal; es decir, si la gráfica de frecuencias de los puntajes tiene aproximadamente la forma de campana de la densidad normal. Esto se justifica considerando que se espera que en el centro se concentre la mayoría de las calificaciones, mientras que en las colas estén los extremos representados por aquéllos que obtuvieron calificaciones no aprobatorias (cola izquierda) y por aquéllos que desarrollaron una prueba perfecta (cola derecha).

El examinador utiliza los puntajes de la prueba para estimar los parámetros normales μ y σ^2 y entonces asigna *MB* (muy bien) a aquellos cuyo puntaje es mayor que $\mu + \sigma$; *B* (bien) a aquellos cuyo puntaje está entre $\mu - \sigma$ y $\mu + \sigma$; *S* (suficiente) a aquellos cuyo puntaje está entre $\mu - 2\sigma$ y $\mu - \sigma$, y *NA* (no aprobó) a aquellos que obtuvieron un puntaje por debajo de $\mu - 2\sigma$. Bajo estos criterios, interesa saber cuál es la distribución porcentual de los alumnos de acuerdo con sus calificaciones.

[49] Ejercicio adaptado de Ross [1988], p. 167.

$$P\{X > \mu + \sigma\} = P\left\{\frac{X-\mu}{\sigma} > 1\right\} = 1 - \Phi(1) = .1587$$

$$P\{\mu < X < \mu + \sigma\} = P\left\{0 < \frac{X-\mu}{\sigma} < 1\right\} = \Phi(1) - \Phi(0) = .3413$$

$$P\{\mu - \sigma < X < \mu\} = P\left\{-1 < \frac{X-\mu}{\sigma} < 0\right\}$$
$$= \Phi(0) - \Phi(-1) = .3413$$

$$P\{\mu - 2\sigma < X < \mu - \sigma\} = P\left\{-2 < \frac{X-\mu}{\sigma} < -1\right\}$$
$$= \Phi(2) - \Phi(1) = .1359$$

$$P\{X < \mu - 2\sigma\} = P\left\{\frac{X-\mu}{\sigma} < -2\right\} = \Phi(-2) = .0228$$

Por lo tanto, 16% de los aplicantes obtendrá una calificación *MB* en el examen, 68% una calificación *B* (34% caerá entre μ y $\mu + \sigma$, y 34% en el intervalo de $\mu - \sigma$ a μ); 14% una calificación *S*, y 2% no aprobarán.

◈

El Teorema del Límite Central. Por otra parte, el *Teorema del Límite Central* constituye uno de los dos resultados más importantes de la teoría de la probabilidad [50] En su forma más simple, este teorema proporciona una base teórica para la observación empírica que frecuentemente se encuentra en la práctica y que consiste en que la suma de un número grande de variables aleatorias independientes tiene una distribución *aproximadamente* normal.

Si bien el teorema proporciona un método sencillo para calcular probabilidades aproximadas para sumas de variables aleatorias independientes, el Teorema del Límite Central permite también explicar el hecho sobresaliente de que las frecuencias empíricas de muchas poblaciones naturales exhiben curvas con forma de campana, es decir, normales. En otras palabras, el Teorema del Límite Central da cuenta de lo frecuente que es la presencia de la distribución normal en la realidad

[50] Ross [1988] considera que el otro resultado fundamental es la *Ley fuerte de los grandes números*.

Teorema 2.16.1
Teorema del Límite Central [16]

Sea $X_1, X_2, \ldots$ una sucesión de *variables aleatorias independientes e idénticamente distribuidas* cada una con media μ y varianza σ^2. Entonces la distribución de

$$\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$$

tiende a la normal estándar conforme $n \to \infty$. Es decir,

$$P\left\{\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}} \leq a\right\} \to \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{a} e^{-x^2} dx, \qquad \text{conforme} \qquad n \to \infty.$$ [17]

◆

Ross [1988] relata que la primera versión del Teorema del Límite Central fue probada por DeMoivre alrededor de 1733, para el caso especial en que X_i son variables aleatorias independientes e idénticamente distribuidas Bernoulli con parámetro $p = ½$. Posteriormente, Laplace generalizó este resultado para cualquier p arbitraria y descubrió la forma general del Teorema del Límite Central presentada arriba. Sin embargo, su demostración no fue estrictamente rigurosa. Una demostración verdaderamente rigurosa del teorema fue presentada primero por el matemático ruso Liapounoff en el periodo 1901-1902.[18]

---

[16] Existen varias versiones equivalentes del teorema del límite central. La versión que aquí se presenta es la que expone Ross [1988]. Una demostración del teorema (que recurre a resultados intermedios de convergencia) puede consultarse en el mismo texto de Ross [1988] (véase Ross, [1988], pp. 341-342). El ejemplo de aplicación del teorema también fue adaptado de la misma fuente.

[17] Dada la suma de n v.a.i.i.d. X_i:

$$E\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} \mu = n\mu \quad \text{y} \quad Var\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} Var[X_i] = \sum_{i=1}^{n} \sigma^2 = n\sigma^2,$$

de donde $\frac{X_1 + X_2 + \ldots + X_n - n\mu}{\sigma\sqrt{n}}$ es una variable normal estándar (véase el apartado ✂ *Propiedades*).

[18] Existen también teoremas del límite central para cuando las X_i son variables aleatorias independientes pero no necesariamente idénticamente distribuidas. Una versión de ellos puede consultarse en Ross [1988], p. 346.

Ejemplo de aplicación del teorema del límite central. Un astrónomo está interesado en medir, en años luz, la distancia entre su observatorio y una estrella distante. Aunque el astrónomo dispone de una técnica de medición, él sabe que, debido al cambio en las condiciones atmosféricas y el error normal, cada vez que realiza una medición, no obtendrá la distancia exacta, sino una mera estimación. Ante ello, el astrónomo decide llevar a cabo una serie de mediciones y entonces utilizar el promedio de ellas como el valor estimado de la distancia real. Suponiendo que los valores de esas mediciones son variables aleatorias independientes e idénticamente distribuidas con media común d (la distancia real) y una varianza común de 4 años luz, ¿cuántas mediciones tendría que realizar el astrónomo para estar razonablemente seguro de que su distancia estimada sea precisa hasta por ± 0.5 años luz?

Suponga que el astrónomo decide realizar n observaciones. Si $X_1, X_2, \ldots, X_n$ son las n mediciones, entonces, del Teorema de Límite Central, se sigue que

$$Z_n = \frac{\sum_{i=1}^{n} X_i - nd}{2\sqrt{n}}$$

tiene aproximadamente distribución normal estándar. Entonces

$$P\left\{-.5 \le \frac{\sum_{i=1}^{n} X_i}{n} - d \le .5\right\} = P\left\{-.5\frac{\sqrt{n}}{2} \le Z_n \le .5\frac{\sqrt{n}}{2}\right\} \approx \Phi\left(\frac{\sqrt{n}}{4}\right) - \Phi\left(-\frac{\sqrt{n}}{4}\right) = 2\Phi\left(\frac{\sqrt{n}}{4}\right) - 1$$

Por tanto, si el astrónomo quisiera, por ejemplo, estar 95 por ciento seguro de que su valor estimado es preciso hasta por 0.5 años luz, él deberá realizar n^* mediciones, donde n^* es tal que

$$2\Phi\left(\frac{\sqrt{n^*}}{4}\right) - 1 = 0.95 \qquad \text{o} \qquad \Phi\left(\frac{\sqrt{n^*}}{4}\right) = 0.975$$

y así, de una tabla con probabilidades acumuladas para la distribución normal estándar, $\frac{\sqrt{n^*}}{4} = 1.96$, lo que equivale a decir que $n^* = 61.47$. Como n^* no es un valor entero, él deberá realizar 62 observaciones.

◊

Observe que el análisis anterior fue hecho bajo la suposición de que la aproximación normal es una buena aproximación cuando $n = 62$. *La respuesta a la pregunta sobre qué* tan grande debe ser n para que la aproximación sea adecuada depende de la distribución de las X_i.

✂ *Propiedades*

- Aplicando el criterio de la primera y segunda derivadas *respecto a* x *a la expresión [2.16]*, es posible verificar que la densidad normal alcanza su máximo en $x = \mu$, y que $\mu - \sigma$ y $\mu + \sigma$ son puntos de inflexión

- Toda variable aleatoria $X \sim N(\mu,\sigma^2)$ puede *estandarizarse* –transformarse en una variable con distribución *normal estándar*– mediante la transformación $Z = (X - \mu)/\sigma$ (lo cual se verificará en el Capítulo 3). *De aquí que los valores de la función de* distribución de X pueda expresarse en términos de $\Phi(z)$, como sigue:

$$\begin{aligned} F_X(x) &= P[X \leq x] \\ &= P[(X - \mu)/\sigma \leq (x - \mu)/\sigma] \\ &= \Phi[(x - \mu)/\sigma]. \end{aligned}$$

  Debido a que el integrando $\exp(-z^2)$ de $\Phi(z)$ no tiene una antiderivada que pueda expresarse de manera explícita en términos de funciones elementales, no es posible calcular la integral directamente. Para evaluarla es necesario recurrir a métodos de aproximación. Es por ello que existen tablas de valores de la función de distribución normal estándar.

- Otra propiedad útil de $\Phi(z)$ resulta de la simetría de la densidad normal:

$$\Phi(-z) = 1 - \Phi(z).$$

Relaciones con otras distribuciones

El Teorema del Límite Central muestra que, bajo condiciones relajadas, la distribución normal puede ser utilizada para *aproximar* una amplia variedad de distribuciones en *muestras grandes* En este sentido, dentro del esquema de Leemis, de todas y cada una de las distribuciones podría partir una línea punteada hasta la distribución normal (lo cual resultaría poco operativo y ocasionaría una saturación del espacio). En el Capítulo 3 se presta atención a las relaciones de la distribución normal con la Poisson, la binomial, la gamma y la beta, que son las que se muestran aparecen en el esquema. Además de ser una distribución límite, la normal se vincula de manera bilateral con las distribución lognormal Asimismo, la distribución normal estándar se relaciona con las distribuciones Cauchy estándar y Ji-cuadrada (gamma) mediante transformaciones particulares; y con la t de Student como distribución límite Por último. se verificará que la distribución normal satisface la propiedad reproductiva, que la diferencia de normales es normal y que puede pasarse de una variable normal (μ,σ^2) a una normal estándar y viceversa.

1.	Una transformación lineal de una v.a normal es también normal	[3.15]
2.	La distribución normal satisface la propiedad reproductiva.	[3.16]. [3.16*bis*] y [3.17]
3	La diferencia de dos v.a.i normal también se distribuye normal.	[3.18]
4.	Toda v.a. normal (μ,σ^2) puede llevarse a la forma estándar.	[3.19]
5.	Toda v a. estándar puede llevarse a normal (μ,σ^2).	[3.20]
6.	La distribución normal es una aproximación a la distribución binomial.	[3.21]
7.	La distribución normal es una aproximación a la distribución Poisson	[3 22]
8.	La distribución normal es una aproximación a la distribución gamma	[3 23]
9.	La distribución normal es una aproximación a la distribución beta	[3.24]
10	La distribución normal estándar es una aproximación a la distribución t	[3.25]
11	El cuadrado de una v.a. normal estándar se distribuye gamma.	[3.30]
12.	La suma de cuadrados de v.a i.i.d normal estándar se distribuye ji-cuadrada	[3 46]
13	El cociente de dos v.a i. normal estándar se distribuye Cauchy estándar.	[3.57]
14.	El logaritmo natural de una v.a. lognormal se distribuye normal.	[3 64]
15.	La exponencial de una v.a normal se distribuye lognormal.	[3 65]

La distribución t de Student

Definición

Se dice que una v.a. X tiene *distribución t de Student* o, simplemente, *distribución t* con parámetro n si su f.d.p. está dada por

$$f_X(x) = f_X(x;n) = \begin{cases} \dfrac{\Gamma\left(\frac{n+1}{2}\right)}{(n\pi)^{1/2}\Gamma\left(\frac{n}{2}\right)} \cdot \dfrac{1}{\left[1+(x^2/n)\right]^{(n+1)/2}} & \text{si} \quad -\infty < x < \infty \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.17]}$$

donde n es un número no negativo.

- *Sobre la notación:* Se suele leer que X tiene *distribución t de Student con n grados de libertad*, lo cual se puede denotar como $X \sim t_{(n)}$.

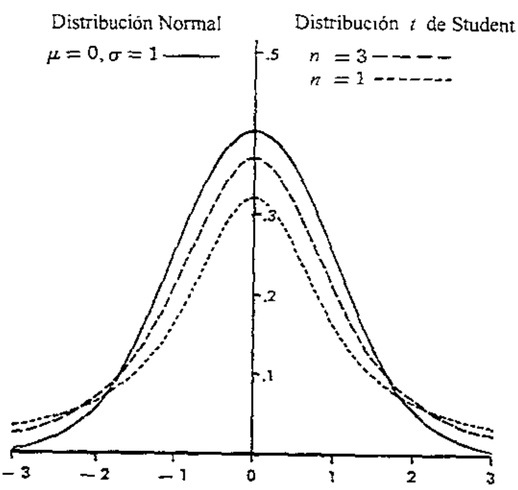


FIGURA 2.19. F.d.p. t de Student para distintos valores del parámetro n comparada con la f.d.p. normal estándar

Algo de historia

La derivación de distribuciones de probabilidad asociadas con $\bar{X}$ y S^2 para muestras aleatorias de poblaciones normales $N(\mu,\sigma^2)$ es en algún sentido el primer paso en el análisis estadístico En particular, en la mayoría de los casos prácticos la varianza σ^2 es desconocida. Así, para tener una idea de la variabilidad de $\bar{X}$ (como una estimación de μ), es necesario estimar la varianza. El primero en abordar esta cuestión fue W S Gosset, quien publicaba bajo el seudónimo de 'Student', a principios de 1900. El trabajo de Student permitió obtener lo que hoy se conoce como la distribución t.

Ejemplos de aplicación

La construcción de pruebas de hipótesis y de intervalos de confianza para los valores esperados de distribuciones normales son las dos aplicaciones más fuertes de la distribución t

Ejemplo. Sea $X_1, X_2, \ldots, X_n$ es una muestra aleatoria de una población normal $N(\mu,\sigma^2)$. Entonces la distribución de

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S} \qquad [2.17.1]$$

es t de Student con $(n-1)$ grados de libertad[54] Esto no es difícil de verificar, una vez hechas algunas manipulaciones Multiplicando la expresión [2.17 1] por un 1 de la forma $(1/\sigma)/(1/\sigma)$ y haciendo un ligero rearreglo se obtiene

$$\frac{\sqrt{n}(\bar{X}-\mu)}{S} = \frac{(\bar{X}-\mu)/(\sigma/\sqrt{n})}{\sqrt{S^2/\sigma^2}}$$

El numerador de la expresión anterior es una variable aleatoria normal estándar y el denominador es $\sqrt{\chi^2_{n-1}/(n-1)}$, *independiente* del numerador. Por lo tanto, la distribución de [2.17.1] puede encontrarse resolviendo el problema simplificado de encontrar la distribución de $U/\sqrt{V/p}$, donde

[54] Para mayor referencia sobre S^2, vease el apartado sobre la distribución ji-cuadrada Si la varianza fuera conocida, entonces en la expresión anterior puede sustituirse el valor de σ; y en ese caso $\sqrt{n}(\bar{X}-\mu)/\sigma$ tiene distribucion normal estándar (véase el apartado sobre la distribución normal) Precisamente, Student quien propuso la expresión [X.2] como alternativa para cuando la varianza es desconocida, como sucede generalmente en la práctica

$U \sim N(0,1)$, $V \sim \chi^2_{(p)}$, y U y V son independientes, que es precisamente la distribución t de Student.[20]

◆

✂ *Propiedades*

- La distribución t no tiene función generadora de momentos (al igual que la distribución Cauchy) porque no tiene momentos de todos lo órdenes. De hecho, si se tienen p grados de libertad, sólo existen $p-1$ momentos. Así, una variable aleatoria $t_{(1)}$ no tiene media, $t_{(2)}$ no tiene varianza, etcétera

- Si $n > 1$, la media de la distribución t ($E[X]$) es 0. Tal es la primera condición que permite establecer una similitud distribución normal estándar (véase *Figura 2.19*), la cual se verifica ceda vez mejor conforme el parámetro n crece

Relaciones con otras distribuciones

La distribución t de Student se vincula con la distribución F. Además, ante las dos propiedades previas, cabe esperar una relación de la distribución t con la normal estándar y con la Cauchy. En efecto:

1. La distribución normal estándar es una aproximación a la distribución t [3.25]
2. La distribución Cauchy estándar es un caso especial de la distribución t. [3.49]
3. El cuadrado de una v.a. t tiene distribución F. [3.50]

■

[20] No se abundará más en la derivación de la distribución t de Student. Sin embargo, este es un buen ejercicio de manejo de una transformación que involucra de *dos distribuciones distintas*. Así, es posible aplicar la técnica de la transformacion considerando las funciones $t = u / \sqrt{v/p}$ y $w = v$. Casella y Berger presentan una exposición clara de esta derivacion (véase Casella y Berger [1990], p. 226).

La distribución uniforme

Definición

Se dice que una v.a X tiene *distribución uniforme* con parámetros a y b si su f.d.p. está dada por

$$f_X(x) = f_X(x;a,b) = \begin{cases} \dfrac{1}{b-a} & \text{si} \quad a < x < b \\ 0 & \text{e.o.c.} \end{cases} \qquad [2.18]$$

donde $-\infty < a, b < \infty$.

- *Sobre la notación:* Una forma de abreviarlo es $X \sim U(a,b)$, y otra forma de leerlo es "X se distribuye *uniformemente* sobre el intervalo (a,b)".[56]

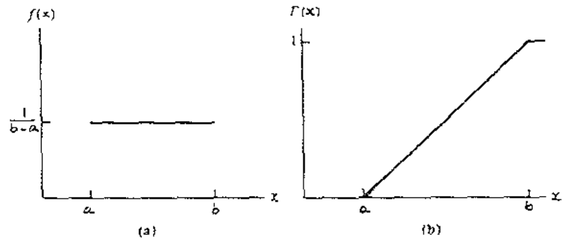


FIGURA 2 20 Para valores a y b arbitrarios: (a) f.d p uniforme, (b) f.d a. uniforme

Esta distribución debe su nombre al hecho de que su densidad es uniforme o constante sobre el intervalo (a,b).

- Cuando los valores de los parámetros son $a = 0$ y $b = 1$ se dice que la distribución es ***uniforme estándar***.

[56] Para algunos autores el rango de la distribución es el intervalo *cerrado* $[a,b]$. Por tratarse de una distribución continua, esto no altera en lo absoluto la función de densidad

Algo de historia

Johnson y Kotz [1970] consideran que, dado lo natural que resulta la distribución uniforme, probablemente ha sido utilizada por mucho más tiempo del que puede inferirse de los registros escritos, como los que proporcionan Bayes [1763] y Laplace [1812] acerca del uso de la distribución. *Un interés histórico particular se vincula con la distribución de la suma de variables aleatorias* independientes cada una con distribución uniforme

Ejemplos de aplicación

La distribución uniforme provee un modelo útil para algunos fenómenos aleatorios. Por ejemplo, si se conoce que los valores de alguna variable aleatoria X sólo pueden estar en un intervalo finito, digamos $[a,b]$, y si se supone que cualesquiera dos subintervalos de $[a,b]$ de igual longitud tienen *la misma probabilidad de contener a X, entonces X tiene una distribución uniforme. Cuando se* habla de un número aleatorio del intervalo $[0,1]$ se está pensando en el valor de una variable aleatoria uniformemente distribuida sobre el intervalo $[0,1]$. La *simulación de distribuciones* se realiza precisamente mediante la generación de números aleatorios

Propiedades

- Al ser $F_X(x) = P[X \leq x]$ en si misma una función de la variable aleatoria X, es también una variable aleatoria (véase Capítulo 1) para la cual puede encontrarse su distribución. Si X es una variable aleatoria continua, entonces $F_X(X)$ se distribuye uniforme estándar

Teorema 2.18.1
$F_X(x)$ tiene distribución uniforme estándar

Si X es una variable aleatoria continua, entonces $F_X(X)$ se distribuye uniforme estándar.

Demostración:

Considere $Y = F_X(X)$. Sea $G_Y(y) = P[Y \leq y]$ Entonces

$$G_Y(y) = P[Y \le y] = P[F_X(X) \le y]$$
$$= P[F_X^{-1}(F_X(X)) \le F_X^{-1}(y)] \qquad \text{(por ser } Y \text{ continua existe } Y^{-1}\text{)}$$
$$= P[X \le F_X^{-1}(y)]$$
$$= F_X[F_X^{-1}(y)]$$
$$= y$$

Derivando (es decir, aplicando el *Teorema 1 2*)

$$g_Y(y) = 1, \qquad y \in (0,1).$$

que corresponde a la f.d.p. de una variable aleatoria uniforme estándar

◈

Relaciones con otras distribuciones

Toda variable aleatoria uniforme (a, b) puede llevarse a la forma uniforme estándar y viceversa Por otro lado, la distribución uniforme se vincula con la distribuciones exponencial y triangular mediante transformaciones. Además, la distribución uniforme estándar es un caso especial de la distribución beta

1. La transformación $-\alpha \log X$ de una v a. uniforme estándar es exponencial. [3.42]
2. La distribución uniforme estándar es un caso especial de la distribución beta [3.59]
3. Toda v.a. uniforme estándar puede llevarse a uniforme (a,b). [3.61]
4. Una v a. uniforme (a,b) es uniforme estándar cuando $a = 0$ y $b = 1$. [3.62]
5. La diferencia de dos v.a.i i.d. uniforme estándar se distribuye triangular. [3.63]

La distribución Weibull

Definición

Se dice que X es una v.a. con *distribución Weibull* con parámetros α y β si su f.d.p. está dada por

$$f_X(x) = f_X(x;\alpha,\beta) = \begin{cases} (1/\alpha)^* \beta x^{\beta-1} \cdot \exp\left[-(1/\alpha)^* x^{\beta}\right] & \text{si } x > 0 \\ 0 & \text{e.o.c.} \end{cases} \qquad [2.19]$$

donde $\alpha,\beta > 0$.[57]

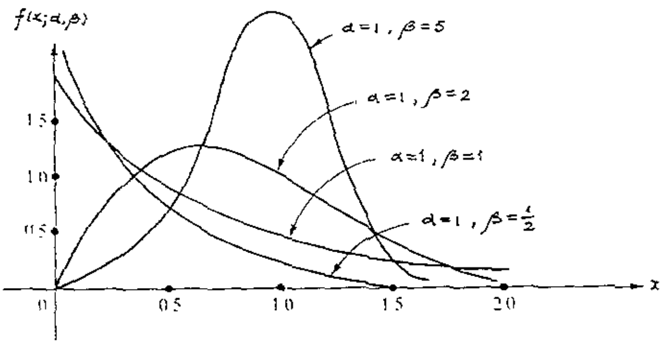


FIGURA 2.21. F.d.p. Weibull para $\alpha = 1$ y distintos valores de β

El parámetro α es un parámetro de escala y β es un parámetro de forma.[58] Como puede observarse en la *Figura 2.21*, aunque la densidad está sesgada hacia la derecha esta asimetría se reduce conforme el parámetro β se incrementa.

---

[57] Algunos autores definen esta distribución en términos de una densidad con tres parámetros: α, $\beta > 0$, y $\gamma \geq 0$ (véase Tsokos [1972] y Ross [1988]); a saber, $f(x;\alpha,\beta,\gamma) = (1/\alpha)^*\beta(x-\gamma)^{\beta-1}*\exp[-(1/\alpha)^*(x-\gamma)^{\beta}]$, donde $x > \gamma$. Así, el formato aquí presentado corresponde al caso especial para $\gamma = 0$. Seguramente, aquellos autores que utilizan la expresión biparamétrica no sólo lo hacen por simplicidad, sino también por ser recurrente en la práctica, especialmente en la representación de tiempos de vida. Cualquier otra diferencia en la presentación de la densidad Weibull depende del formato adoptado para la densidad exponencial.

[58] El parámetro γ es un parámetro de localización o de umbral.

Algo de historia

En 1951, el físico suizo Walodi Weibull introdujo la distribución que lleva su nombre en un artículo titulado "A statistical distribution function of wide applicability" (*Una función de distribución estadística de amplia aplicación*) como una función de densidad aplicable a varios fenómenos físicos Weibull la había usado desde 1939 para representar la distribución de la fuerza interruptora de algunos materiales. En la literatura rusa se le conoce como la distribución Weibull-Gnedenko

Ejemplos de aplicación

La distribución Weibull es muy útil en la "teoría de la confiabilidad" y el control de calidad como modelo de falla. Es común que resulte adecuada cuando las condiciones de "aleatoriedad estricta" de la distribución exponencial no se satisface. Su flexibilidad para ser aplicada en un número amplio de problemas se justifica por su relación con las densidades exponencial y Rayleigh, las cuales son casos especiales de la densidad Weibull, como se verificará en el Capítulo 3

Ejemplo: La densidad Weibull como un modelo de falla. La confiabilidad se entiende como la probabilidad de que un sistema o un componente funcione adecuadamente por un periodo de tiempo deseado si las condiciones de operación requeridas se satisfacen. Es decir, la confiabilidad es la probabilidad de que un sistema dado no fallará en su funcionamiento por un intervalo de tiempo definido La distribución Weibull es muy útil para modelar *funciones de riesgo*

La función de riesgo. Suponga que la variable aleatoria T es el tiempo de vida de un objeto. La *función de riesgo (hazard rate function)*, $h_T(t)$, asociada con la variable aleatoria T se define como

$$h_T(t) = \lim_{\delta \to 0} \frac{P[t \le T < t + \delta \mid T \ge t]}{\delta}$$

Por lo tanto, se puede interpretar a $h_T(t)$ como la razón de cambio de la probabilidad de que el objeto sobreviva un periodo de tiempo muy pequeño después del tiempo t, dado que objeto sobrevive hasta el tiempo t. Cuando T es una variable aleatoria continua, entonces

$$h_T(t) = \frac{f_T(t)}{1 - F_T(t)} = -\frac{d}{dt}\log(1 - F_T(t))$$

Se puede demostrar que la función de riesgo determina de manera *única* la distribución de T, en tanto que permite obtener la función de distribución F_T (y en consecuencia la densidad de T, $f_T(t)$)

La función de confiabilidad. Al denominador $1 - F_T(t)$ de la función de riesgo se le conoce como la *función de confiabilidad* (*reliability function*) de un sistema al tiempo t y puede ser denotada como $R(t)$.

Cuando se habla de un sistema, puede tratarse de un sistema muy complejo constituido por varios subsistemas que se distinguen como componentes Para calcular la *probabilidad de sobrevivencia de tal sistema se requiere conocer el comportamiento de la confiabilidad de cada uno de esos* componentes. Al iniciar un estudio confiabilidad, la primera pregunta que debe hacerse es: ¿cuál es la función de densidad de probabilidad que mejor describe el comportamiento de los tiempos aleatorios en los que falla el sistema? Precisamente, la densidad Weibull es aplicable como una función de falla cuando el sistema en estudio consiste de más de un componente y cuando la falla se debe al más serio defecto de una serie grande de defectos que pudieran presentarse en el sistema.

Mediante la elección apropiada de los parámetros de la densidad Weibull, es posible estudiar la confiabilidad de sistemas complejos cuyas funciones de falla sean crecientes, decrecientes o constantes. La función de confiabilidad para una razón de falla Weibull es

$$R(t) = \exp\left(-\frac{t^{\beta}}{\alpha}\right), \quad \text{para} \quad \alpha,\beta > 0,\ t \geq 0$$

y, por ende, la función de riesgo es

$$h_T(t) = \frac{\beta}{\alpha} t^{\beta-1}, \qquad \text{para} \quad \alpha,\beta > 0.$$[59]

◆

[59] Para la expresión tri-paramétrica de la densidad Weibull (α,β,γ) $R(t) = \exp\left(-\frac{(t-\gamma)^{\beta}}{\alpha}\right)$ para $\alpha, \beta > 0$ y $t \geq \gamma$ y $h_T(t) = \frac{\beta}{\alpha}(t-\gamma)^{\beta-1}$ para $\alpha, \beta > 0$, y $\gamma \geq 0$.

Por otra parte, esta distribución ofrece una aproximación muy cercana a la distribución normal para ciertos valores de los parámetros Por ejemplo, se ha encontrado que ambas son prácticamente idénticas para $\alpha = 1$, $\beta = 3.2589$ de la densidad Weibull, y para $\mu = 0\ 8964$ y $\sigma^2 = 0.0924$ de la densidad normal (véase *Figura 2.22*).

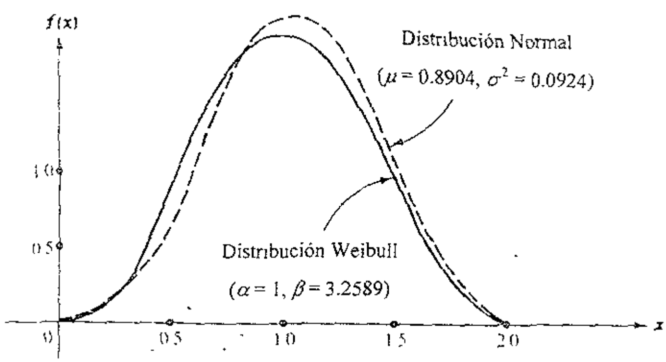


FIGURA 2.22 Comparación entre una densidad Weibull y una normal

☞☜ *Relaciones con otras distribuciones*

Como ya se mencionó, la flexibilidad de la densidad Weibull para ser aplicada a un gran número de problemas se justifica por su relación con las densidades exponencial y Rayleigh, las cuales son casos especiales ésta

1 La distribución exponencial es un caso especial de la distribución Weibull [3 32]

2 La transformación $X^{1/\beta}$ de una v.a. exponencial se distribuye Weibull [3.38]

3. La transformación X^{β} de una v.a. Weibull se distribuye exponencial. [3.39]

4 La distribución Rayleigh es un caso especial de la distribución Weibull. [3.67]

Otras distribuciones univariadas continuas

La distribución arcoseno

Definición

Se dice que una v.a. X tiene *distribución arcoseno* si su f.d.p está dada por

$$f_X(x) = \begin{cases} \dfrac{1}{\pi[x(1-x)]^{1/2}} & \text{si} \quad 0 < x < 1 \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.20]}$$

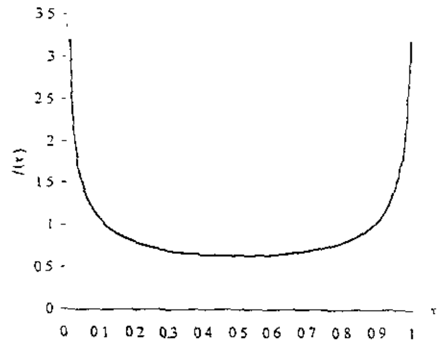


FIGURA 2.23 F.d.p. arcoseno (compare con la *Figura* 2.9 para $a = b = ¼$)

Esta distribución debe su nombre a que la función de distribución asociada a [2.20] es

$$F_X(x) = \mathrm{P}[X \le x] = (2/\pi)\ \mathrm{sen}^{-1}\sqrt{x}\,.$$

Ejemplos de aplicación

Ejemplo. La distribución arcoseno surge de manera interesante en la teoría de las "caminatas aleatorias". Suponga que una partícula se mueve a lo largo de la recta real con pasos de longitud unitaria, empezando desde el cero, siendo igualmente probable dar un paso hacia la derecha (creciente) o hacia la izquierda (decreciente). Sea T_{2n} la variable aleatoria que denota el número de veces en los primeros $2n$ pasos para las cuales el punto es en el intervalo 0 a $2n$ inclusive en al momento de concluir un paso. Entonces

$$P[T_{2n}=2k]=\binom{2k}{k}\binom{2n-2k}{n-k}2^{-2n} \quad ,k=0,1,\ldots,n$$

La razón $T_{2n}/(2n)$ puede ser interpretada como la "fracción de tiempo transcurrida en la parte positiva de la recta real" Conforme $n\to\infty$, la distribución límite de $T_{2n}/(2n)$ es la distribución arcoseno; es decir,

$$\lim_{n\to\infty}\left\{\sum_{k\le nx}P[T_{2n}=2k]\right\}=\frac{1}{\pi}\int_0^x t^{-1/2}(1-t)^{-1/2}dt=(2/\pi)\operatorname{sen}^{-1}\sqrt{x}\,.$$

Relaciones con otras distribuciones

- La distribución arcoseno es un caso especial de la distribución beta. [3 58]

La distribución Laplace

Definición

Se dice que una v.a. X tiene *distribución Laplace* o *doble exponencial* con parámetros α_1 y α_2 si su f.d.p. está dada por

$$f_X(x)=f_X(x;\alpha_1,\alpha_2)=\begin{cases}[1/(\alpha_1+\alpha_2)]e^{-x/\alpha_1} & \text{si} \quad x\ge 0\\ [1/(\alpha_1+\alpha_2)]e^{x/\alpha_2} & \text{si} \quad x<0\end{cases} \qquad \textbf{[2.21]}$$

donde α_1 , $\alpha_2>0$.

- *Para esta distribución se considera pertinente presentar en primer plano la comparación entre las expresiones que utilizan distintos autores, ya que, a primera vista, ninguno de los textos consultados coincide con la expresión [2.21] que presenta Leemis en su artículo·*

- Casella y Berger [1990] presentan también una expresión biparamétrica de la densidad doble exponencial $f_X(x;\mu,\sigma)$ como·

$$f_X(x;\mu,\sigma) = \frac{1}{2\sigma} \cdot e^{-\ x-\mu\ /\sigma} \qquad \text{para} \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0 \qquad [2.21']$$

Cabe recordar que Casella y Berger utilizan la misma expresión de Leemis para la densidad exponencial ($f(x;\ \lambda) = 1/\lambda^{*}\exp[-x/\lambda]$) *Sin embargo, aunque Mood et al.* [1974] *presentan la* expresión alternativa ($f(x;\ \alpha) = \lambda^{*}\exp[-\lambda x]$) también utilizan la expresión [2.21'] para definir la densidad Laplace o doble exponencial.

Si se toma $\alpha_1 = \alpha_2 = \alpha$ en la expresión [2.21] de Leemis se obtiene la densidad Laplace como la definen, Casella y Berger y Mood *et al* en [2.21'] con parámetros $\mu = 0$ y $\sigma = \alpha$. Como la condición es que $\sigma > 0$ se valida que α_1 y α_2 son mayores que cero, como se había establecido.[60]

◆

- Ross [1988] presenta una expresión uniparamétrica $f_Y(x;\lambda)$ para la densidad Laplaciana:

$$f_Y(x;\lambda) = \frac{1}{2}\lambda \cdot e^{-\lambda\ x} \qquad \text{para} \quad -\infty < x < \infty, \quad \lambda \geq 0 \qquad [2.21'']$$

Considerando que Ross utiliza la expresión $f(x;\ \lambda) = \lambda^{*}\exp[-\lambda x]$, para $x > 0$ y $\lambda > 0$, en lugar de $f(x;\ \lambda) = 1/\lambda^{*}\exp[-x/\lambda]$, también para $x > 0$ y $\lambda > 0$, como lo haría Leemis, puede *decirse que la expresión* [2.21] *se corresponde con la expresión* [2.21''] *tomando también* $\lambda = \alpha$.

◆

Por lo tanto, la expresión [2.21] define una densidad Laplace o doble exponencial con *media cero* (*y en ese sentido puede verse como un caso particular de* [2.21'] y [2.21'']) cuya construcción involucra *dos* variables aleatorias exponenciales con parámetros α_1 y α_2 (y en este otro sentido puede verse como un caso más general) Sobre la relación entre las distribuciones exponencial y doble exponencial se habla a continuación.

[60] En su artículo, Leemis no aclara del todo los rangos de los parametros.

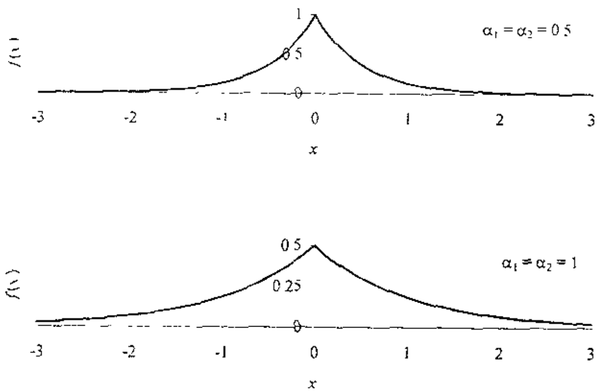


FIGURA 2 24 F d.p Laplace o doble exponencial para distintos valores de $\alpha = \alpha_1 = \alpha_2$

Propiedades

- Casella y Berger [1990] expresan que la distribución doble exponencial se obtiene reflejando la distribución exponencial con respecto a su media

- Bajo la expresión [2.21'], la doble exponencial provee una distribución simétrica con colas gruesas (mucho más gruesas que aquellas de la normal), pero retiene todavía todos sus momentos, de hecho $E[X] = \mu$ y $Var[X] = 2\sigma^2$.

- La distribución doble exponencial no tiene forma de campana; de hecho tiene un pico (o, más formalmemente, un punto en el que no es derivable) en $x = \mu$. Al tratar analíticamente con esta distribución es importante recordar este punto. El signo de valor absoluto puede ocasionar problemas al integrar; es mejor dividir la integran en regiones alrededor de $x = \mu$.

Relaciones con otras distribuciones

1. La diferencia de dos v.a i. exponencial se distribuye LaPlace. [3.40]
2. El valor absoluto de una v.a LaPlace se distribuye exponencial. [3 41]

La distribución lognormal

Definición

Se dice que X es una v.a. con *distribución lognormal* con parámetros μ y σ^2 si su f.d.p. está dada por

$$f_X(x) = f_X(x; \mu, \sigma^2) = \begin{cases} \dfrac{1}{x\sqrt{2\pi}\sigma} \exp\left\{-\dfrac{(\log_e x - \mu)^2}{2\sigma^2}\right\} & \text{si} \quad x > 0 \\ 0 & \text{e.o.c.} \end{cases} \qquad \textbf{[2.22]}$$ [61]

donde $-\infty < \mu < \infty$ y $\sigma > 0$.

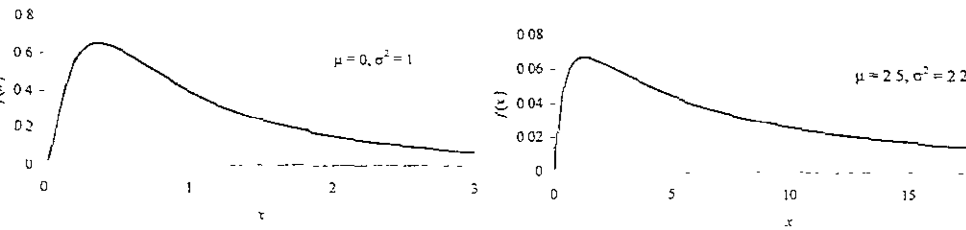


FIGURA 2.25. F.d.p. lognormal para distintos valores de μ y σ^2

Relaciones con otras distribuciones

La distribución lognormal recibe su nombre del hecho de que el logaritmo natural de una variable aleatoria con esta distribución se distribuye normal (*Nota: NO se está diciendo que la distribución lognormal se obtenga de aplicar el logaritmo natural a una variable aleatoria normal*). Por otro lado, como el logaritmo de un producto es la suma de los logaritmos, la distribución lognormal implícitamente satisface la propiedad reproductiva. Así, en el Capítulo 3 se verificará que:

---

[61] Esta expresión no es la que presenta Leemis en su artículo. Se optó por ella porque es la que se presenta en la literatura revisada y además porque es la que resulta conveniente para demostrar la relación que guarda con la distribución normal. Leemis define la distribución lognormal como:

$$f_X(x) = f_X(x; \alpha, \beta^2) = \begin{cases} \dfrac{1}{x\sqrt{2\pi}\beta} \exp\left\{-\dfrac{\log_e(x/\alpha)^2}{2\beta^2}\right\} & \text{si} \quad x > 0 \\ 0 & \text{e.o.c.} \end{cases}.$$

1. El logaritmo natural de una v.a. lognormal se distribuye normal. [3 64]
2. La exponencial de una v a normal se distribuye lognormal [3 65]
3. El producto de v a i i d con distribución lognormal es también lognormal [3 66]

La distribución Rayleigh

Definición

Se dice que X es una v.a. con *distribución Rayleigh* con parámetro α si su f.d p está dada por

$$f_X(x) = f_X(x;\alpha) = \begin{cases} \dfrac{2x}{\alpha} e^{-x^2/\alpha} & \text{si} \quad x > 0 \\ 0 & \text{e.o.c} \end{cases} \qquad \textbf{[2.23]}$$

para $\alpha > 0$.[62]

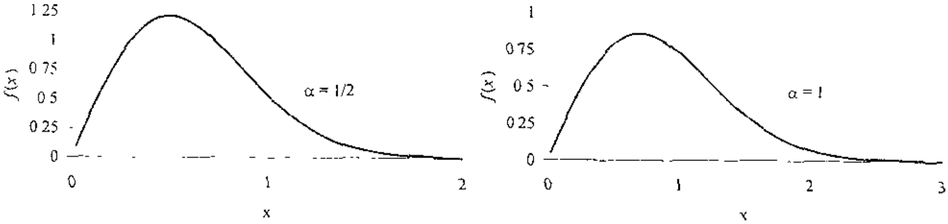


FIGURA 2.26 F.d.p Rayleigh para distintos valores de α

[62] La distribución Rayleigh es otra de las distribuciones cuya presentación difiere en la literatura En algunas fuentes el formato de la f.d.p Rayleigh depende del formato adoptado para la f.d p exponencial; de tal suerte que si el parámetro α se maneja en términos de su inverso ($1/\alpha$*exp[$-x/\alpha$]), la densidad Rayleigh esta dada por $f(x,\alpha) = x/\alpha$*exp[$-x^2/2\alpha$], de donde se infiere que Casella y Berger [1990] presentan el caso particular para $\alpha = 1$. En cambio, si el manejo del parámetro α se hace de manera directa (α*exp[$-\alpha x$]) entonces la densidad Rayleigh se presenta como $f(x;\alpha) = \alpha x$*exp[$-\alpha x^2/2$], como lo hace Ross [1988] Sin embargo, Mood *et al* –pese a considerar también este último formato de la densidad exponencial– la definen como $f(x;\alpha) = (1/\alpha^2)x$*exp[$-\frac{1}{2}(x/\alpha)^2$], $x > 0$ y $\alpha > 0$ Por su parte, Mendenhall *et al.* [1986], así como Tsokos [1972] coinciden con Leemis en cuanto al formato de la densidad Rayleigh (así como en el formato de la densidad exponencial). Como se verificará al momento de abordar la relación entre las distribuciones exponencial y Rayleigh, las variaciones dependen de la transformación considerada

Ejemplos de aplicación

Ejemplo. Un problema que permite derivar esta distribución y darle una interpretación es el siguiente. Para localizar un objeto en el plano xy, se determina su distancia al origen mediante la fórmula de Pitágoras $r = x^2 + y^2$, donde x y y son las coordenadas rectangulares (x,y) del objeto de interés. Si *se considera este objeto como un punto aleatorio en el plano* (X,Y), y si se supone que las coordenadas rectangulares X y Y son v.a normales con parámetros $\mu = 0$ y $\sigma^2 = \alpha/2$, entonces $R = \sqrt{x^2 + y^2}$ tiene distribución Rayleigh.

♦

Ejemplo. La distribución Rayleigh es importante en la teoría del sonido y sirve además como modelo del tiempo de vida de un artículo (algún componente electrónico, por ejemplo) Esta última aplicación se inserta dentro de la *teoría de la confiabilidad.* Así, la distribución Rayleigh, como caso especial que es de la distribución Weibull, al igual que la distribución exponencial, es útil para modelar *funciones de riesgo.*[63]

A continuación se ilustra cómo se obtiene la función de distribución F_T a partir de la función de riesgo $h_T(t)$. Considere que una variable aleatoria T tiene una *función de riesgo lineal.*

$$h_T(t) = a + bt,$$

entonces su función de distribución está dada por:

$$F_T(t) = 1 - \exp\,[-at - bt^2/2], \qquad t \geq 0$$

Calculando la derivada de esta expresión se obtiene que la densidad de T (véase *Teorema 1.2* del *Capítulo 1*):

$$f_T(t) = (a + bt) * \exp\,[-at - bt^2/2], \qquad t \geq 0$$

Cuando $a = 0$ *y* $b = 2/\alpha$, *la expresión anterior corresponde a la función de densidad Rayleigh.*

♦

[63] Para mayor referencia sobre la *teoría de la confiabilidad, funciones de riesgo* y *modelos de falla* puede consultar los *Ejemplos de aplicación* dentro del apartado de la distribucion Weibull

Relaciones con otras distribuciones

La distribución Rayleigh se vincula de manera bilateral con la distribución exponencial mediante transformaciones. Además, es un caso especial de la distribución Weibull

1. La raíz cuadrada de una v.a exponencial se distribuye Rayleigh [3.36]
2. El cuadrado de una v a. Rayleigh se distribuye exponencial. [3.37]
3. La distribución Rayleigh es un caso especial de la distribución Weibull [3.67]

La distribución triangular

Definición

Se dice que una v a. X tiene *distribución triangular* si su f.d.p. está dada por

$$f_X(x) = \begin{cases} 1+x & \text{si} \quad -1 < x < 0 \\ 1-x & \text{si} \quad 0 \le x < 1 \\ 0 & \text{e.o.c} \end{cases} \qquad \textbf{[2.24]}$$ [64]

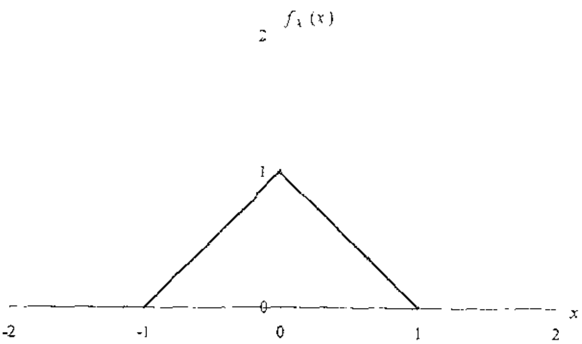


FIGURA 2 27. F d p triangular

[64] Esta es la expresión presentada por Leemis Es una distribucion nada común en la literatura Johnson y Kotz [1970] presentan otras densidades triangulares dentro de su apartado sobre la distribución uniforme [Véase Johnson y Kotz [1970]-2, p 64]

Como puede observarse, esta distribución debe su nombre a la forma de su función de densidad La expresión [2.24] es la que presenta Leemis en su artículo Sólo se encontró referencia a la distribución triangular en Johnson y Kotz [1970]. Estos autores presentan otras densidades triangulares dentro de su apartado sobre la distribución uniforme

Relaciones con otras distribuciones

- La diferencia de dos v.a.i.i.d. uniforme estándar se distribuye triangular. [3.63]

■

En el capítulo siguiente se verificarán todas y cada una de las relaciones entre las distribuciones expuestas aquí, de acuerdo al orden asignado por el número entre corchetes. Como se verá, cada una de las relaciones se describe con mayor detalle.

3

RELACIONES ENTRE LAS DISTRIBUCIONES

En este capítulo se presentan a detalle 68 relaciones entre las distribuciones introducidas en el capítulo previo, las cuales fueron enumeradas al finalizar la exposición de cada distribución; 12 de estas relaciones no se encontraban en el esquema original de Leemis, sino que fueron incorporadas a lo largo del desarrollo de este trabajo Este es un capítulo práctico y sintético se aboca a verificar todas y cada unas de las relaciones aplicando los resultados expuestos en el Capítulo 1, y haciendo uso de la información sobre las distribuciones que se presentó en el Capítulo 2 Así, se recogen aquí las tres técnicas expuestas en el Capítulo 1, así como el *Teorema del Límite Central* presentado en el Capítulo 2 (dentro de la sección de *La distribución normal*).

La *técnica de la transformación* abarca los *Resultados 1.1* a *1.8*, para el caso univariado, y el *Teorema 1.8* para el caso bivariado, a partir del cual puede obtenerse una densidad marginal. La *técnica de la función de distribución* está presente en los *Resultados 1.9* a *1.17* (y recuerde que está presente también como un insumo de la técnica de la transformación), pero aquí no se utilizan todos ellos. La *técnica de la función generadora de momentos* se aplica mediante el *Resultado 1.18* cuando se tiene que ver con la suma de variables aleatorias independientes e idénticamente distribuidas o bajo una misma distribución con parámetros distintos. El *Teorema de Límite Central* se utiliza para verificar las aproximaciones de distintas distribuciones a la distribución normal

Asimismo, para probar algunas de las relaciones se hace uso de relaciones verificadas previamente o a verificar más adelante, lo cual refleja la interrelación que existe entre las distribuciones Además, en una buena parte de los casos la aplicación de cada técnica va acompañada de "trucos" y desarrollos algebraicos que requieren de un poco de paciencia para seguirse. También es recurrente el uso de varias propiedades de la función gamma (véase ANEXO 1.6)

El orden de presentación de las 68 relaciones responde a un criterio de bloques Se verifican primero las relaciones entre las distribuciones discretas, para después pasar a los casos continuos. A su vez, dentro de los bloques discreto y continuo se eligieron distribuciones a manera de "centros" en torno a los cuales se vinculan las otras distribuciones. Así, se abrieron 14 secciones: 4 cuyo centro es una distribución discreta; 8 cuyo centro es una distribución continua; y 2 que concentran "otras relaciones" que por cuestiones de secuencia se consideró conveniente aislar. Puesto que toda relación consta de una distribución "origen", una distribución "destino" y un conjunto de condiciones que las vinculan, *en general*, se procuró que cada relación quedara dentro de la sección de la distribución "destino". Al final de cada apartado se especifica la ubicación de todas aquellas otras relaciones que involucran también a la distribución en cuestión, pero que están en otra sección.

RELACIONES CON LA DISTRIBUCIÓN BINOMIAL

[3.1]

$$\text{Binomial } (1, p) \equiv \text{Bernoulli } (p)$$

"Una v.a. con distribución binomial $(1, p)$ es una Bernoulli (p)"

("La distribución Bernoulli (p) es un caso especial de la distribución binomial (n, p) cuando $n = 1$")

Demostración:.

Sea $X \sim$ binomial (n, p). Basta sustituir directamente el valor $n = 1$ en la f.d.p binomial (n, p) y observar que $\binom{1}{x}$ sólo tiene sentido para $x = 0$ ó 1, y que en ambos casos $\binom{1}{x} = 1$. Así

$$\begin{aligned} f_X(x) = f_X(x; 1, p) &= \binom{1}{x} p^x (1-p)^{1-x} \quad , \qquad x = 0, 1 \\ &= p^x (1-p)^{1-x} \end{aligned}$$

que corresponde a la f.d.p. Bernoulli (p).

♦

Otra forma de obtener este resultado es mediante la función generadora de momentos, como sigue. La función generadora de momentos de una variable aleatoria X con distribución binomial (n, p) es:

$$M_X(t) = (q + pe^t)^n, \text{ donde } q = 1 - p.$$

Cuando $n = 1$ la expresión anterior queda como $M_X(t) = (q + pe^t)$ que es la función generadora de momentos de una variable aleatoria con distribución Bernoulli (p).

♦♦

[3.2]

$$\boxed{\sum_{1}^{n} \text{Bernoulli}(p) \sim \text{Binomial}\,(n, p)}$$

"La suma de n v.a.i.i.d. Bernoulli (p) se distribuye binomial (n, p)"

Demostración

Sean $X_1, X_2, \ldots, X_n$ v.a.i.i.d. Bernoulli (p). Defínase $Y = \sum_{i=1}^{n} X_i$. Como cada X_i sólo puede tomar valores 0 ó 1,[1] entonces la nueva variable aleatoria Y puede tomar los valores $y = 0, 1, \ldots, n$. Aplicando la *Técnica de la función generadora de momentos* vía el *Resultado1 18*:

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\ &= \prod_{i=1}^{n} (q + pe^t) \\ &= (q + pe^t)^n \qquad \text{(ya que la expresión } (q + pe^t) \text{ no depende de } i) \end{aligned}$$

que corresponde a la función generadora de momentos de una variable aleatoria binomial (n, p)

◊

No es difícil apreciar la valía de la aplicación del *Resultado1 18* en este caso. Basta calcular $P[Y = y]$ a pie para $n = 2$; es decir, $P[X_1 + X_2 = y]$. Para ello, observe que el evento $\{X_1 + X_2 = y\}$ puede escribirse como la *unión* de los *eventos ajenos* $\{X_1 = k, X_2 = y - k\}$,

[1] Interesan los valores de X_i para los cuales su f.d.p., $f_{X_i}(x)$, es positiva ya que con ellos se determinan los valores de la nueva variable aleatoria Y para los cuales $f_Y(x)$ también es positiva, es decir el rango de Y

para $0 \leq k \leq y$. Como cada X_i es Bernoulli (p), Y puede tomar los valores $y = 0, 1, 2$. Además, como X_1 y X_2 son *independientes* entonces:

Para $y = 0$

$$\begin{aligned} P[X_1 + X_2 = 0] &= P[X_1 = 0, X_2 = 0] \\ &= P[X_1 = 0] \cdot P[X_2 = 0] \\ &= (1-p) \cdot (1-p) \\ &= (1-p)^2 \qquad [3.2.1a] \end{aligned}$$

Para $y = 1$

$$\begin{aligned} P[X_1 + X_2 = 1] &= P[X_1 = 0, X_2 = 1] + P[X_1 = 1, X_2 = 0] \\ &= P[X_1 = 0] \cdot P[X_2 = 1] + P[X_1 = 1] \cdot P[X_2 = 0] \\ &= p \cdot (1-p) + p \cdot (1-p) \cdot p \\ &= 2 \cdot p(1-p) \qquad [3.2.1b] \end{aligned}$$

Para $y = 2$

$$\begin{aligned} P[X_1 + X_2 = 2] &= P[X_1 = 1, X_2 = 1] \\ &= P[X_1 = 1] \cdot P[X_2 = 1] \\ &= p \cdot p \qquad [3.2.1c] \\ &= p^2 \end{aligned}$$

Para $y \neq 0, 1, 2$

$$P[X_1 + X_2 = y] = 0 \qquad [3.2.1d]$$

Los resultados [3.2.1a] a [3.2.1d] pueden resumirse mediante la expresión:

$$\begin{aligned} P[X_1 + X_2 = y] &= \binom{2}{y} p^2 (1-p)^{2-y} \qquad \text{si} \qquad y = 0, 1, 2 \\ &= 0 \qquad\qquad\qquad\qquad\quad \text{e.o.c.} \end{aligned}$$

que corresponde a una f.d.p. binomial $(2, p)$

Así, tomando como base el desarrollo anterior y aplicando un proceso de inducción sobre n se podría probar también la relación [3.2]. Resulta más sencillo aplicar la *Técnica de la función generadora de momentos* (¿o no?).

♦♦

- *Relaciones de la distribución binomial con ella misma*

[3.3]

$$\boxed{\sum_{1}^{m} \text{Binomial}(n, p) \sim \text{Binomial}\,(mn, p)}$$

"La suma de m v.a.i.i.d. binomial (n, p) se distribuye también binomial (mn, p)"

(La distribución binomial satisface la propiedad reproductiva)

Demostración

Sean $X_1, X_2, \ldots, X_m$ v.a.i.i.d. binomial (n, p). Defínase $Y = \sum_{i=1}^{m} X_i$. Observe que la nueva variable aleatoria Y puede tomar los valores $y = 0, 1, \ldots, mn$. Aplicando la *Técnica de la función generadora de momentos* vía el *Resultado1.18*.

$$\begin{aligned}
M_Y(t) &= \prod_{i=1}^{m} M_{X_i}(t) \\
&= \prod_{i=1}^{m} (q + pe^t)^n \\
&= \left[(q + pe^t)^n\right]^m \qquad \text{(ya que la expresión } (q + pe^t)^n \text{ no depende de } i\text{)} \\
&= (q + pe^t)^{mn}
\end{aligned}$$

que corresponde a la función generadora de momentos de una variable aleatoria binomial (mn, p).

◊

La relación [3.3] puede generalizarse a m v.a.i.i.d. binomial (n_i, p), como se expresa en [3.3*bis*], que puede demostrarse también mediante la *Técnica de la función generadora de momentos*

[3.3 *bis*]

$$\sum_{1}^{m} \text{Binomial}(n_i, p) \sim \text{Binomial}\left(\sum_{i=1}^{m} n_i, p\right)$$

"La suma de m v.a.i.i.d. binomial con parámetro n_i, $i = 1, 2, \ldots, m$, y parámetro común p se distribuye también binomial con parámetro la suma de los n_i y parámetro p"

Demostración

$$M_Y(t) = \prod_{i=1}^{m} M_{X_i}(t)$$

$$= \prod_{i=1}^{m} (q + pe^t)^{n_i}$$

$$= (q + pe^t)^{n_1} \cdot (q + pe^t)^{n_2} \cdots (q + pe^t)^{n_i}$$

$$= (q + pe^t)^{\sum_{i=1}^{m} n_i}$$

que corresponde a la función generadora de momentos de una variable aleatoria binomial $\left(\sum_{i=1}^{m} n_i, p\right)$

♦

Una forma de interpretar este resultado es considerar a cada X_i como el número de éxitos obtenidos en n_i ensayos[2] Así, $X_1 + X_2 + \ldots + X_m$ será el número de éxitos obtenidos en $n_1 + n_2 + \ldots + n_m$ ensayos.

La demostración de [3.3*bis*] puede hacerse a pie siguiendo un razonamiento similar a aquel descrito en la segunda parte de la verificación de la relación [3.2], y usando la relación $\binom{n_1 + n_2}{y} = \sum_{k=0}^{y} \binom{n_1}{k}\binom{n_2}{y-k}$, para el caso de $m = 2$, en el procedimiento de inducción que requiere seguirse.

---

[2] Véase *Ejemplos de aplicación* dentro del apartado dedicado a la distribución binomial para mejor comprensión de este contexto.

- *La distribución binomial como distribución límite*

[3.4]

$$\text{Hipergeométrica }(n_1, n_2, n_3) \xrightarrow{p=\frac{n_1}{n_3},\ n_3\to\infty} \text{Binomial }(n_2, p)$$

"La distribución binomial (n_2, p) puede utilizarse como una aproximación de la distribución hipergeométrica (n_1, n_2, n_3) cuando $p = n_1/n_3$ y n_3 es suficientemente grande"

Demostración Con el fin de proporcionar después una interpretación de este resultado dentro del contexto del control de calidad (véase *Ejemplos de aplicación* de la sección 2.5), se considerará que $n_1 = D$, $n_2 = n$ y $n_3 = N$

Sea X una v.a. hipergeométrica (D, n, N). Defínase $p = D/N$ (proporción de artículos defectuosos en el lote), entonces $D = Np$ Así, la f.d.p. de X está dada por:

$$f_X(x) = f_X(x, D, n, N) = \frac{\binom{Np}{x}\binom{N-Np}{n-x}}{\binom{N}{n}}, \qquad x = 0, 1, \ldots, min(D, n)$$

$$= \frac{\dfrac{(Np)!}{(Np-x)!\,x!}\cdot\dfrac{(N-Np)!}{[(N-Np)-(n-x)]!\,(n-x)!}}{\dfrac{N!}{(N-n)!\,n!}}$$

$$= \frac{(N-n)!\,n!}{N!}\cdot\frac{(Np)!}{(Np-x)!\,x!}\cdot\frac{(N-Np)!}{[(N-Np)-(n-x)]!\,(n-x)!}$$

Reagrupando

$$f_X(x; D, n, N) = \frac{n!}{(n-x)!\,x!}\cdot\frac{(Np)!}{(Np-x)!}\cdot\frac{(N-Np)!}{[(N-Np)-(n-x)]!}\cdot\frac{(N-n)!}{N!} \qquad [3.4.1]$$

Con el fin de simplificar se procederá a desarrollar los últimos tres factores de la expresión [3.4.1]

Para el segundo factor observe que

$$\frac{(Np)!}{(Np-x)!}=\frac{\overbrace{Np(Np-1)(Np-2)\dots(Np-(x-1))}^{x-\text{términos}}(Np-x)!}{(Np-x)!}$$

$$=Np(Np-1)(Np-2)\dots(Np-(x-1))$$

$$=Np[N(p-\tfrac{1}{N})][N(p-\tfrac{2}{N})\dots N(p-\tfrac{x-1}{N})]=N^{x}\prod_{i=0}^{x-1}(p-\tfrac{i}{N}) \qquad [3.4.2]$$

Análogamente para el tercer factor:

$$\frac{(N-Np)!}{[(N-Np)-(n-x)]!}=\frac{\overbrace{(N-Np)(N-Np-1)(N-Np-2)\dots(N-Np-(n-x-1))}^{(n-x)-\text{términos}}(N-Np-(n-\cdots}{(N-Np-(n-x))!}$$

$$=(N-Np)(N-Np-1)(N-Np-2)\dots(N-Np-(n-x-1))$$

$$=\{N(1-p)\}\{N[(1-p)-\tfrac{1}{N}]\}\{N[(1-p)-\tfrac{2}{N}]\}\dots\{N[(1-p)-\tfrac{n-x-1}{N}]\}$$

$$=N^{n-x}\prod_{i=0}^{n-x-1}[(1-p)-\tfrac{i}{N}] \qquad [3.4.3]$$

De manera más inmediata para el cuarto factor:

$$\frac{(N-n)!}{N!}=\frac{(N-n)!}{N(N-1)(N-2)\dots(N-(n-1))(N-n)!}$$

$$=\frac{1}{N[N(1-\frac{1}{N})][N(1-\frac{2}{N})]\dots[N(1-\frac{n-1}{N})]}=\frac{1}{\prod_{i=0}^{n-1}N(1-\frac{i}{N})}=\frac{1}{N^{n}\prod_{i=0}^{n-1}(1-\frac{i}{N})} \qquad [3.4.4]$$

Recogiendo los resultados [3.4.2], [3.4.3] y [3.4.4] y reagrupando se tiene que:

$$f_X(x;D,n,N)=\binom{n}{x}\cdot\frac{N^{x}\cdot N^{n-x}}{N^{n}}\cdot\frac{\prod_{i=0}^{x-1}(p-\frac{i}{N})\cdot\prod_{i=0}^{n-x-1}[(1-p)-\frac{i}{N}]}{\prod_{i=0}^{n-1}[1-\frac{i}{N}]}$$

Se está ya en condiciones de obtener $\lim_{N\to\infty} f_X(x;D,n,N)$, lo cual se hará calculando $\lim_{N\to\infty}$ de cada uno de los cuatro factores de esta última expresión. Así,

- $\lim_{N\to\infty}\binom{n}{x}=\binom{n}{x}$ (ya que este factor no depende de N) [3.4.5]
- $\lim_{N\to\infty}\prod_{i=0}^{x-1}(p-\frac{i}{N})=\prod_{i=0}^{x-1}\lim_{N\to\infty}(p-\frac{i}{N})=\prod_{i=0}^{x-1}p=p^{x}$ [3.4.6]
- $\lim_{N\to\infty}\prod_{i=0}^{n-x-1}[(1-p)-\frac{i}{N}]=\prod_{i=0}^{n-x-1}\lim_{N\to\infty}[(1-p)-\frac{i}{N}]=\prod_{i=0}^{n-x-1}(1-p)=(1-p)^{n-x}$ [3.4.7]
- $\lim_{N\to\infty}\prod_{i=0}^{n-1}N(1-\frac{i}{N})=1$ [3.4.8]

Finalmente, reuniendo los resultados [3 4.5] a [3 4.8] se obtiene

$$\lim_{N \to \infty} f_X(x;D,n,N) = \binom{n}{x} p^x (1-p)^{n-x}$$

que corresponde a la f.d.p binomial (n, p) Por lo tanto, para N grande y $p = D/N$

$$f_X(x;D,n,N) \approx \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0,1,2,\ldots,n$$

◈

Una forma de interpretar el resultado anterior es la siguiente. Considere la aplicación vinculada con el control de calidad para la distribución hipergeométrica que se expuso en el capítulo 2 [3] Por la naturaleza del experimento, observe que la extracción de la muestra de tamaño n se efectúa *sin* reemplazo. Así, la probabilidad de seleccionar un artículo *defectuoso* varía (disminuye de hecho) de una extracción a otra, siendo esta probabilidad ${}^{D-i+1}/_{N-i+1}$ para el *i*-ésimo artículo *defectuoso* que se incorpora a la muestra, $1 \le i \le n$.

Por otra parte, se puede pensar el mismo contexto en términos de la distribución binomial. Para ello considere que *seleccionar un artículo defectuoso* representa un *éxito*. Al ir seleccionando los artículos del lote se irá integrando una muestra. Observe que (y he aquí la clave) que en el contexto de la distribución binomial la selección de la muestra es *con* reemplazo, porque la probabilidad de seleccionar un artículo *defectuoso* permanece *constante* de una extracción a otra siempre es p Más aún, en este caso, esa probabilidad es la proporción de artículos defectuosos en el lote ${}^{D}/_{N}$ [4]

La relación [3.4] expresa que cuando el tamaño de la muestra (n) es *suficientemente grande*, la probabilidad de seleccionar un artículo defectuoso dentro de la muestra de tamaño n es la *misma*. independientemente del esquema de muestreo que se aplique; es decir, ya sea *con* reemplazo (distribución binomial) o *sin* reemplazo (distribución hipergeométrica)

◈ ◈

[3] Véase *Ejemplos de aplicación* dentro de la sección dedicada a la distribución hipergeométrica en el Capítulo 2 (sección 2 5)

[4] Piense en términos de la probabilidad clásica casos favorables (número de artículos defectuosos, D) entre casos totales (número de artículos en el lote, N)

[3.5]

$$\text{Beta-binomial } (n_1, n_2, n_3) \xrightarrow{p = n_2/n_3,\ n_3 \to \infty} \text{Binomial } (n_1, p)$$

"La distribución binomial (n_1, p) puede utilizarse como una aproximación de la distribución beta-binomial (n_1, n_2, n_3) cuando $p = n_2/n_3$ y n_3 es suficientemente grande"

La demostración de esta relación está pendiente. Desarrollando cada una de las combinatorias en la expresión de la densidad beta-binomial se obtiene.

$$f_X(x) = f_X(x; n_1, n_2, n_3) = \frac{\binom{n_2 + x - 1}{x}\binom{n_1 + n_3 - x - 1}{n_1 - x}}{\binom{n_1 + n_2 + n_3 - 1}{n_1}}, \quad x = 0, 1, \ldots, n_1$$

$$= \frac{n_1!}{(n_1 - x)!\,x!} \cdot \frac{(n_1 + n_3 - x - 1)!}{(n_3 - 1)!} \cdot \frac{(n_3 p + x - 1)!}{(n_3 p - 1)!} \cdot \frac{(n_3(p+1) - 1)!}{(n_3(p+1) + n_1 - 1)!}$$

de donde puede identificarse $\binom{n_1}{x}$ como primer término. Es posible realizar un desarrollo similar al de la demostración de la relación [3.4], sin embargo la dificultad para llegar a la expresión de la densidad binomial radica en que se tiene el término $(1 + p)$, en lugar de $(1 - p)$.

◇

- ***Otras relaciones con la distribución binomial***

Véase también [3.7] en la sección *Relaciones con la distribución Poisson.*

Véase también [3.21] en la sección *Relaciones con la distribución normal.*

♠

RELACIONES CON LA DISTRIBUCIÓN POISSON

- ***Relaciones de la distribución Poisson con ella misma***

[3.6]

$$\sum_{1}^{n} \text{Poisson}(\mu) \sim \text{Poisson}(n\mu)$$

"La suma de n v.a.i.i.d. Poisson (μ) se distribuye también Poisson ($n\mu$)"

(La distribución Poisson satisface la propiedad reproductiva)

Demostración

Sean $X_1, X_2, \ldots, X_n$ v.a.i.i.d. Poisson (μ). Defínase $Y = \sum_{i=1}^{n} X_i$. Como la f.d.p. de cada X_i es positiva para los valores $x_i = 0, 1, 2, \ldots$, el rango de la nueva variable aleatoria Y está conformado también por los valores $y = 0, 1, 2, \ldots$. Aplicando la *Técnica de la función generadora de momentos* vía el *Resultado 1.18*

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_X(t) \\ &= \prod_{i=1}^{n} \exp[\mu(e^t - 1)] \\ &= \exp\left\{\sum_{i=1}^{n} \mu(e^t - 1)\right\} \\ &= \exp\{n\mu(e^t - 1)\} \qquad \text{(ya que la expresión } [\mu(e^t - 1)] \text{ no depende de } i) \end{aligned}$$

que corresponde a la función generadora de momentos de una variable aleatoria Poisson ($n\mu$)

◊

La relación [3 6] puede generalizarse al caso en que el parámetro de la distribución Poisson de las X_i no fuera el valor común μ, sino que $X_1, X_2, \ldots, X_n$ fueran v.a.i con distribución Poisson (μ_i). Esto se expresa en [3.6*bis*], la cual puede demostrarse también mediante la *Técnica de la función generadora de momentos*

[3.6 *bis*]

$$\boxed{\sum_{1}^{n} \text{Poisson}(\mu_i) \sim \text{Poisson}\left(\sum_{i=1}^{n} \mu_i\right)}$$

"La suma de n v.a.i.i.d. Poisson con parámetro μ_i, $i = 1, 2, \ldots, n$, se distribuye también Poisson con parámetro la suma de los μ_i"

Demostración:

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\ &= \prod_{i=1}^{n} \exp[\mu_i (e^t - 1)] \\ &= \exp\left\{\sum_{i=1}^{n} \mu_i (e^t - 1)\right\} \\ &= \exp\left\{(e^t - 1)\sum_{i=1}^{n} \mu_i\right\} \end{aligned}$$

que corresponde a la función generadora de momentos de una variable aleatoria Poisson $\left(\sum_{i=1}^{n} \mu_i\right)$.

◆

La demostración de la relación [3 6*bis*] puede hacerse a pie mediante un proceso de inducción, siguiendo un razonamiento similar al descrito en la segunda parte de la verificación de la relación [3 2] y recurriendo al Teorema del Binomio [5] A continuación se presenta la demostración de [3 6*bis*] para $n = 2$ con el fin de exhibir el nuevamente el valor de la *Técnica de la función generadora de momentos* Sea $Y = X_1 + X_2$, entonces la f.d.p. de Y está dada por

$$
\begin{aligned}
f_Y(y) = f_{X_1+X_2}(y) &= P[X_1 + X_2 = y] \qquad , \qquad \text{para } y = 0,1,2,\ldots \\
&= P\left[\bigcup_{k=0}^{y}\{X_1 = k, X_2 = y-k\}\right] \\
&= \sum_{k=0}^{y} P[X_1 = k, X_2 = y-k] \\
&= \sum_{k=0}^{y} P[X_1 = k]P[X_2 = y-k] \qquad \text{(aplicando que } X_1 \text{ y } X_2 \text{ son independientes)} \\
&= \sum_{k=0}^{y} \frac{\mu_1^k}{k!}e^{-\mu_1} \frac{\mu_2^{y-k}}{(y-k)!}e^{-\mu_2} \\
&= e^{-\mu_1+\mu_2}\sum_{k=0}^{y} \frac{\mu_1^k \mu_2^{y-k}}{k!\,(y-k)!} \\
&= \frac{e^{-(\mu_1+\mu_2)}}{y!}\sum_{k=0}^{y}\binom{y}{k}\mu_1^k\,\mu_2^{y-k} \qquad \text{(multiplicando y dividiendo por } y!) \\
&= \frac{e^{-(\mu_1+\mu_2)}}{y!}(\mu_1+\mu_2)^y \qquad \text{(por el Teorema del Binomio)}
\end{aligned}
$$

que corresponde a la f.d.p. de una variable aleatoria Poisson $(\mu_1 + \mu_2)$

◆ ◇

---

[5] El Teorema del Binomio establece que si n es un entero positivo y a y b son dos números reales, entonces $(a+b)^n = \sum_{k=0}^{n}\binom{n}{k}a^k b^{n-k}$

- *La distribución Poisson como distribución límite*

[3.7]

$$\text{Binomial }(n,p) \xrightarrow{\mu=np \quad n\to\infty} \text{Poisson }(\mu)$$

"La distribución Poisson (μ) puede utilizarse como una aproximación de la distribución binomial (n,p) si $\mu = np$ y n es suficientemente grande"

Demostración:

Sea X una v.a. binomial con parámetros n y p, tales que n es suficientemente grande y p suficientemente pequeña como para que np adopte un tamaño moderado. Defínase $\mu = np$. Entonces, la f.d.p. de la v.a. X está dada por:

$$f_X(x) = f_X(x;n,p) = \frac{n!}{(n-x)!\,x!}\,p^x(1-p)^{n-x} \quad , \qquad x = 0,1,\ldots,n$$

$$= \frac{n!}{(n-x)!\,x!}\cdot\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)^{n-x}$$

$$= \frac{n(n-1)\ldots(n-(x-1))\,(n-x)!}{(n-x)!\,x!}\cdot\left(\frac{\mu}{n}\right)^x\left(1-\frac{\mu}{n}\right)^{n-x}$$

$$= \frac{n(n-1)\ldots(n-(x-1))}{n^x}\cdot\frac{\mu^x}{x!}\cdot\frac{\left(1-\frac{\mu}{n}\right)^n}{\left(1-\frac{\mu}{n}\right)^x}$$

Como μ es un valor moderado se tiene que:

$$\lim_{n \to \infty} \frac{n(n-1)\ldots(n-x+1)}{n^x} = 1$$

$$\lim_{n \to \infty} \left(1 - \frac{\mu}{n}\right)^n = e^{-\mu} \qquad \lim_{n \to \infty} \left(1 - \frac{\mu}{n}\right)^x = 1$$

Por lo tanto, cuando n es grande y $\mu = np$ moderada

$$f_X(x;n,p) \approx e^{-\mu} \frac{\mu^x}{x!},$$

es decir, la f.d.p. de X se aproxima a la f.d.p de una variable aleatoria Poisson (μ).

◈

Una forma de interpretar el resultado anterior es considerar una situación en la que se realizan n ensayos independientes Bernoulli, cada uno de los cuales resulta en "éxito" con probabilidad p[6] Entonces, cuando n es suficientemente grande y p suficientemente pequeña como para hacer $\mu = np$ moderada, el número de éxitos que pueden ocurrir en los n ensayos es aproximadamente una variable aleatoria Poisson con parámetro $\mu = np$

Finalmente, observe que np es el valor de la media ($E[X] = np$) de una variable aleatoria con distribución binomial (n, p) y recuerde también que el parámetro μ es el valor de la media para una variable aleatoria con distribución Poisson (μ).

◈ ◈

[6] Vease *Ejemplos de aplicación* dentro de la sección dedicada a la distribución binomial en el Capítulo 2 (sección 2 2)

[3.8]

$$\text{Binomial negativa } (n,p) \xrightarrow{\mu=n(1-p),\ n\to\infty} \text{Poisson } (\mu)$$

"La distribución Poisson (μ) puede utilizarse como una aproximación de la distribución binomial negativa (n,p) si $\mu=n(1-p)$ y n es suficientemente grande"

Demostración: (se adoptará un procedimiento análogo al que se siguió para demostrar [3 7])

Sea X una v.a. binomial negativa con parámetros n y p, tales que n es suficientemente grande y p suficientemente pequeña como para que $n(1-p)$ adopte un tamaño moderado Defínase $\mu=n(1-p)$. Entonces, la f d.p. de X está dada por

$$f_X(x)=f_X(x;n,p)=\frac{(n+x-1)!}{(n-1)!x!}p^n(1-p)^x \quad , \qquad x=0,1,\ldots,n$$

$$=\frac{(n+x-1)!}{(n-1)!x!}\cdot\left(1-\frac{\mu}{n}\right)^n\left[1-\left(1-\frac{\mu}{n}\right)\right]^x$$

$$=\frac{(n+x-1)(n+x-2)\ldots(n+1)n(n-1)!}{(n-1)!x!}\cdot\left(1-\frac{\mu}{n}\right)^n\left[1-\left(1-\frac{\mu}{n}\right)\right]^x$$

$$=\frac{(n+x-1)(n+x-2)\ldots(n+1)n}{n^x}\cdot\frac{\mu^x}{x!}\cdot\left(1-\frac{\mu}{n}\right)^n$$

Como μ es moderada se tiene que:

$$\lim_{n\to\infty}\frac{(n+x-1)(n+x-2)\ldots(n+1)n}{n^x}=1 \qquad\qquad \lim_{n\to\infty}\left(1-\frac{\mu}{n}\right)^n=e^{-\mu}$$

Por lo tanto,

$$f_X(x;n,p)\approx e^{-\mu}\frac{\mu^x}{x!}, \qquad \text{para} \quad x=0,1,2,\ldots.$$

Es decir, la f.d.p de X se aproxima a la f.d.p de una variable aleatoria Poisson con parámetro $\mu = n(1-p)$ cuando moderada

◇

Como una forma de interpretar el resultado anterior considere que la variable aleatoria X cuenta el *número de fracasos* ocurridos antes de obtener el n–ésimo éxito, en una secuencia de ensayos independientes Bernoulli con probabilidad p de "éxito" en cada ensayo.[7] Entonces, cuando n es suficientemente grande y p suficientemente pequeña como para hacer que $\mu = n(1-p)$ sea moderada, el número de fracasos antes del n–ésimo éxito es aproximadamente una variable aleatoria Poisson con parámetro $\mu = n(1-p)$

Observe que para que la relación [3 8] se verifique es necesario definir la f.d.p de una variable aleatoria X binomial negativa, como lo hizo Leemis en su artículo, mediante la expresión [2 3], pues en tal caso los valores que X puede tomar, o mejor dicho los valores para los cuales $f_X(x)$ es positiva, son $x = 0, 1, 2, \ldots$; es decir, los *mismos* valores para los cuales la f.d.p. de una variable aleatoria Poisson es positiva también (Recuerde que la expresión alternativa [2.3 *bis*] considera que $x = n, n+1, n+2, \ldots,$ pues en tal caso la variable aleatoria X puede interpretarse como el *número de ensayos* realizados hasta obtener el n–ésimo éxito)

Finalmente, al igual que en la relación [3 7], observe también que $n(1-p)$ es el valor de la media ($E[X] = n(1-p)$) de una variable aleatoria con distribución binomial negativa (n, p) y que el parámetro μ es el valor de la media para una variable aleatoria con distribución Poisson (μ).

◈ ◈

- ***Otras relaciones con la distribución Poisson***

Véase también [3 22] en la sección *Relaciones con la distribución normal.*

---

[7] Véase *Ejemplos de aplicación* dentro de la sección dedicada a la distribución binomial negativa en el Capítulo 2 (sección 2.3)

RELACIONES CON LA DISTRIBUCIÓN BINOMIAL NEGATIVA

[3.9]

$$\boxed{\text{Binomial negativa } (1, p) \equiv \text{Geométrica } (p)}$$

"Una v.a. con distribución binomial negativa $(1, p)$ es una geométrica (p)"

(La distribución geométrica (p) es un caso especial de la distribución binomial negativa (n, p) cuando $n = 1$)

Demostración:

Sea X una v.a. binomial negativa $(1, p)$. Basta sustituir directamente el valor $n = 1$ en la f.d.p binomial negativa (n, p), $f_X(x) = f_X(x:n, p) = \binom{n+x-1}{x} p^n (1-p)^x$ si $x = 0,1,2,\ldots$ Así,

$$f_X(x) = f_X(x:1, p) = \binom{1+x-1}{x} p^1 (1-p)^x \quad , \quad x = 0,1,2,\ldots$$
$$= p\ (1-p)^x$$

que corresponde a la f.d.p. geométrica (p).

♦

Una forma alternativa de obtener este resultado es mediante la función generadora de momentos, como sigue. La función generadora de momentos de una variable aleatoria X con distribución binomial negativa (n, p) es:

$$M_X(t) = \left(\frac{p}{1-qe^t} \right)^n, \text{ donde } q = 1 - p.$$

Cuando $n = 1$ la expresión anterior queda como $M_X(t) = \frac{p}{1-qe^t}$, que es la función generadora de momentos de una variable aleatoria con distribución geométrica (p).

Recuerde que X puede interpretarse como el *número de fracasos* ocurridos antes de obtener el n-ésimo éxito, en una secuencia de ensayos independientes Bernoulli con probabilidad p de "éxito" en cada ensayo Si $n = 1$, X cuenta el *número de fracasos* ocurridos antes de obtener el *primer* éxito, que es precisamente una posible interpretación de una variable aleatoria geométrica (p)

◊ ◊

[3.10]

$$\boxed{\sum_{1}^{n} \text{Geométrica}(p) \sim \text{Binomial negativa } (n, p)}$$

"La suma de n v.a.i.i.d. geométrica (p) se distribuye binomial negativa (n, p)"

Demostración:

Sean $X_1, X_2, \ldots, X_n$ v.a.i.i.d geométrica (p). Defínase $Y = \sum_{i=1}^{n} X_i$. Como los valores para los cuales la f.d p. de cada X_i es positiva son $0, 1, 2, \ldots$, el rango de la nueva variable aleatoria Y es también el conjunto de valores $y = 0, 1, 2, \ldots$ Aplicando la *Técnica de la función generadora de momentos* vía el *Resultado 1.18.*

$$M_Y(t) = \prod_{i=1}^{n} M_{X_i}(t)$$

$$= \prod_{i=1}^{n} \left(\frac{p}{1 - qe^t} \right)$$

$$= \left(\frac{p}{1 - qe^t} \right)^n \qquad \text{(ya que la expresión} \left(\frac{p}{1 - qe^t} \right) \text{no depende de } i\text{)}$$

que corresponde a la función generadora de momentos de una variable aleatoria binomial negativa con parámetros n y p.

◊

- ***Relaciones de la distribución binomial negativa con ella misma***

[3.11]

$$\sum_{1}^{m} \text{Binomial negativa } (n,p) \sim \text{Binomial negativa } (mn,p)$$

"La suma de m v.a.i.i.d. binomial negativa (n,p)
se distribuye también binomial negativa (mn,p)"
(La distribución binomial negativa satisface la propiedad reproductiva)

Demostración:

Sean $X_1, X_2, \ldots, X_m$ v.a.i.i.d. binomial negativa (n,p). Defínase $Y = \sum_{i=1}^{m} X_i$. Observe que la nueva variable aleatoria Y puede tomar los valores $y = 0, 1, 2, \ldots$. Aplicando la *Técnica de la función generadora de momentos* vía el *Resultado 1.18*:

$$M_Y(t) = \prod_{i=1}^{m} M_{X_i}(t)$$

$$= \prod_{i=1}^{m} \left(\frac{p}{1-qe^t} \right)^n$$

$$= \left(\frac{p}{1-qe^t} \right)^{mn} \quad \text{(ya que la expresión } \left(\frac{p}{1-qe^t} \right)^n \text{ no depende de } i\text{)}$$

que corresponde a la función generadora de momentos de una variable aleatoria binomial negativa con parámetros mn y p.

◆

La relación [3.11] puede generalizarse a m v.a.i.i.d. binomial negativa (n_i, p), como se expresa en [3.11*bis*], que puede demostrarse fácilmente también a través de la *Técnica de la función generadora de momentos*.

[3.11 *bis*]

$$\sum_{1}^{m} \text{Binomial Negativa } (n_i, p) \sim \text{Binomial Negativa } \left(\sum_{i=1}^{m} n_i, p\right)$$

"La suma de m v.a.i.i.d. binomial negativa

con parámetro n_i, $i = 1, 2, \ldots, m$, y parámetro común p

se distribuye también binomial con parámetro la suma de los n_i y parámetro p"

Demostración:

$$M_Y(t) = \prod_{i=1}^{m} M_{X_i}(t)$$

$$= \prod_{i=1}^{m} \left(\frac{p}{1-qe^t}\right)^{n_i}$$

$$= \left(\frac{p}{1-qe^t}\right)^{n_1} \left(\frac{p}{1-qe^t}\right)^{n_2} \cdots \left(\frac{p}{1-qe^t}\right)^{n_m}$$

$$= \left(\frac{p}{1-qe^t}\right)^{\sum_{i=1}^{m} n_i}$$

que corresponde a la función generadora de momentos de una variable aleatoria con distribución binomial negativa $\left(\sum_{i=1}^{m} n_i, p\right)$.

◊ ◊

- **Otras relaciones con la distribución binomial negativa**

Véase [3.8] en la sección *Relaciones con la distribución Poisson.*

◊

RELACIONES CON LA DISTRIBUCIÓN GEOMÉTRICA

- *Relaciones de la distribución geométrica con ella misma*

[3.12]

$$\boxed{\min_{1\leq i\leq n} \{\text{Geométrica } (p)\} \sim \text{Geométrica } (1-(1-p)^n)}$$

"La mínima estadística de orden de una distribución geométrica (p) es también geométrica (p'), con $p' = 1-(1-p)^n$"

Demostración:

Sean $X_1, X_2, \ldots, X_n$ v.a.i.i.d. geométrica (p). Defínase $Y = \min\{X_1, X_2, \ldots, X_n\}$ Para determinar la distribución de Y se recurrirá primero a la *Técnica de la función de distribución* vía el *Resultado 1 17.*

$$F_Y(y) = 1-[1-F_X(y)]^n = 1-\{P[X>y]\}^n$$

Calculando $P[X>y]$:

$$\begin{aligned} P[X>y] &= 1-P[X\leq y] \\ &= 1-\textstyle\sum_{x=0}^{y} p(1-p)^x \\ &= 1-p\left[\dfrac{1-(1-p)^{y+1}}{1-(1-p)}\right] \\ &= (1-p)^{y+1} \end{aligned}$$

(ya que $0<(1-p)\leq 1$, puede aplicarse $\sum_{j=0}^{n} r^j = \dfrac{1-r^{n+1}}{1-r}$, si $r <$

Así

$$
\begin{aligned}
F_Y(y) &= P[Y \le y] \\
&= 1-[(1-p)^{y+1}]^n \\
&= 1-[(1-p)^n]^{y+1}
\end{aligned}
$$

Ahora, la f.d.p de está dada por Y:

$$
\begin{aligned}
f_Y(y) &= P[Y = y] \\
&= P[Y \le y] - P[Y \le y-1] \\
&= \{1-[(1-p)^n]^{y+1}\} - \{1-[(1-p)^n]^y\} \\
&= [(1-p)^n]^y - [(1-p)^n]^{y+1} \\
&= [(1-p)^n]^y [1-(1-p)^n]
\end{aligned}
$$

que puede identificarse como la f.d p de una v.a. geométrica con parámetro $1-(1-p)^n$

◈

La propiedad de que la mínima estadística de un conjunto de v.a.i.i.d. tenga –aunque con un parámetro alterno– la misma distribución que las variables "originales" es también aplicable a la distribución exponencial, como puede verificarse en la sección correspondiente, lo que exhibe (nuevamente) la similitud entre este par de distribuciones discreta y continua, respectivamente.[8] La relación [3.13] que se presenta enseguida da cuenta también de esa similitud

◈ ◈

[8] Recuerde que tanto la distribucion geométrica como la exponencial registran tiempos de espera y cumplen también la propiedad de la "perdida de memoria"

[3.13]

$$\boxed{\text{Weibull discreta } (p, 1) \equiv \text{Geométrica } (p)}$$

"La distribución geométrica (p) es un caso especial de la distribución Weibull discreta (p, β) cuando β = 1"

Demostración:

Sea X una v.a. Weibull discreta $(p, 1)$ Basta sustituir el valor $\beta = 1$ en la expresión de la f d.p Weibull discreta (p, β)

$$\begin{aligned} f_X(x) = f_X(x; p, 1) &= (1-p)^x - (1-p)^{x+1} \\ &= (1-p)^x[1-(1-p)] \\ &= (1-p)^x p \end{aligned}$$

que corresponde a la f.d.p. geométrica (p).

♦

En la sección correspondiente a las relaciones con la distribución exponencial puede consultarse la versión *continua de este resultado.*

♦♦

- ***Otras relaciones con la distribución geométrica***

Véanse [3.9] y [3.10] en la sección *Relaciones con la distribución binomial negativa*.

♠

OTRAS RELACIONES ENTRE DISTRIBUCIONES DE VARIABLES ALEATORIAS DISCRETAS

[3.14]

Beta-binomial $(n-1, 1, 1) \equiv$ Rectangular (n)

"La distribución rectangular (n) (o uniforme discreta) es un caso especial de la distribución beta-binomial (n_1, n_2, n_3) cuando $n_1 = n-1$, $n_2 = 1$ y $n_3 = 1$"

Demostración

Sea X una v.a. beta binomial $(n-1, 1, 1)$. Sustituyendo los valores correspondientes de los parámetros en la f.d.p. beta binomial (n_1, n_2, n_3): $f_X(x;n_1,n_2,n_3) = \binom{n_1}{x} \frac{\Gamma(n_2+n_3)}{\Gamma(n_2)\Gamma(n_3)} \cdot \frac{\Gamma(n_2+x)\Gamma(n_1+n_3-x)}{\Gamma(n_1+n_2+n_3)}$

$$
\begin{aligned}
f_X(x;n-1,1,1) &= \binom{n-1}{x} \cdot \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} \cdot \frac{\Gamma(1+x)\Gamma(n-1+1-x)}{\Gamma(n-1+1+1)} \qquad , \quad x = 0, 1, \ldots, n-1 \\
&= \binom{n-1}{x} \cdot \frac{\Gamma(2)}{\Gamma(1)\Gamma(1)} \cdot \frac{\Gamma(x+1)\Gamma(n-x)}{\Gamma(n+1)} \\
&= \binom{n-1}{x} \cdot 1 \cdot \frac{x!\,(n-x-1)!}{n!} \\
&= \frac{(n-1)!}{(n-x-1)!\,x!} \cdot \frac{x!\,(n-x-1)!}{n(n-1)!} \\
&= \frac{1}{n}
\end{aligned}
$$

que corresponde a la f.d.p. rectangular (n).

◆

RELACIONES CON LA DISTRIBUCIÓN NORMAL

- ***Relaciones de la distribución normal con ella misma***

[3.15]

$$a\,[\text{Normal}\,(\mu,\sigma^2)] + b \sim \text{Normal}\,(\,|a|\mu + b, (a\sigma)^2)$$

"Una transformación lineal de una v.a. X normal (μ,σ^2), $Y = aX + b$,
también se distribuye normal $(a\mu + b,(a\sigma)^2)$"

Demostración. Sea X una v.a. Normal (μ,σ^2) Defínase $Y = aX + b$ con $a,b \in \Re$, $a \neq 0$ Como $-\infty < x < \infty$, observe que $-\infty < y < \infty$. Aplicando la *Técnica de la transformación* mediante el *Resultado 1.6*.

$$f_Y(y) = \frac{1}{|a|} f_X\left(\frac{y-b}{a}\right) \quad , \qquad -\infty < y < \infty$$

$$= \frac{1}{|a|} \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ \frac{\left(\frac{y-b}{a} - \mu\right)^2}{2\sigma^2} \right\}$$

$$= \frac{1}{\sqrt{2\pi}\,(|a|\sigma)} \cdot \exp\left\{ \frac{\left(\frac{y-b-a\mu}{a}\right)^2}{2\sigma^2} \right\}$$

$$= \frac{1}{\sqrt{2\pi}\,(|a|\sigma)} \cdot \exp\left\{ \frac{[y-(a\mu+b)]^2}{2a^2\sigma^2} \right\} \quad , \qquad -\infty < y < \infty$$

◆

que corresponde a una densidad normal con media $|a|\mu + b$ y varianza $(a\sigma)^2$.

[3.16]

$$\boxed{\sum_{1}^{n} \text{Normal}(\mu,\sigma^2) \sim \text{Normal}\,(n\mu,\, n\sigma^2)}$$

"La suma de n v.a.i.i.d. normal (μ,σ^2) también se distribuye normal ($n\mu$, $n\sigma^2$)"

(La distribución normal satisface la propiedad reproductiva)

Demostración. Sean $X_1, X_2, \ldots, X_n$ v.a.i.i.d. normal (μ,σ^2). Defínase $Y = \sum_{i=1}^{n} X_i$. Observe que $-\infty < y < \infty$. Aplicando la *Técnica de la función generadora de momentos* vía el *Resultado 1.18*:

$$
\begin{aligned}
M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\
&= \prod_{i=1}^{n} \exp[\mu t + (\sigma^2 t^2/2)] \\
&= \exp\left\{\sum_{i=1}^{n} [\mu t + (\sigma^2 t^2/2)]\right\} \\
&= \exp\{n[\mu t + (\sigma^2 t^2/2)]\} \qquad \text{(ya que la expresión } [\mu t + (\sigma^2 t^2/2)] \text{ no depende de } i) \\
&= \exp\{(n\mu)t + (n\sigma^2)t^2/2)]\}, \quad -\infty < y < \infty
\end{aligned}
$$

que corresponde a la f.g.m. de una distribución normal con parámetros $n\mu$ y $n\sigma^2$.

◊

La relación [3.16] puede generalizarse al caso en que los *parámetros* de la distribución normal de las X_i *no* fueran los valores *comunes* μ y σ^2, sino que $X_1, X_2, \ldots, X_n$ fueran v.a.i.i.d. Normal (μ_i, σ_i^2). Esto se expresa en [3.16*bis*], la cual puede demostrarse también mediante la *Técnica de la función generadora de momentos*.

[3.16 *bis*]

$$\sum_{1}^{n} \text{Normal}(\mu_i, \sigma_i^2) \sim \text{Normal}\left(\sum_{1}^{n} \mu_i, \sum_{1}^{n} \sigma_i^2\right)$$

"La suma de n v.a.i.i.d. normal con media μ_i y varianza σ_i^2, $i = 1, 2, \ldots, n$, se distribuye también normal con media la suma de las μ_i y varianza la suma de las σ_i^2"

Demostración Sean $X_1, X_2, \ldots, X_n$ v.a.i.i.d. normal (μ_i, σ_i^2). Defínase $Y = \sum_{i=1}^{n} X_i$. Observe que $-\infty < y < \infty$. Acudiendo a la *Técnica de la función generadora de momentos* (*Resultado 1.18*):

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\ &= \prod_{i=1}^{n} \exp[\mu_i t + (\sigma_i^2 t^2/2)] \\ &= \exp\left\{\sum_{i=1}^{n} [\mu_i t + (\sigma_i^2 t^2/2)]\right\} \\ &= \exp\left\{t\sum_{i=1}^{n} \mu_i + t^2 \sum_{i=1}^{n} (\sigma_i^2/2)]\right\}, \qquad -\infty < y < \infty \end{aligned}$$

que corresponde a la función generadora de momentos de una variable aleatoria normal$\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$

◆

La siguiente relación se refiere a la distribución de *cualquier* combinación lineal de v.a.i.i.d. normales, siendo una generalización de los resultados [3.15] y [3.16*bis*].

[3.17]

$$\sum_{1}^{n}\{a_i * \text{Normal}(\mu_i, \sigma_i^2)\} \sim \text{Normal}\left(\sum_{1}^{n} a_i\mu_i, \sum_{1}^{n} a_i^2\sigma_i^2\right)$$

***"La combinación lineal de n v.a.i.i.d. normal** (μ_i, σ_i^2), $Y = \sum_{i=1}^{n} a_i X_i$,*

***también se distribuye normal** $\left(\sum_{1}^{n} a_i\mu_i, \sum_{1}^{n} a_i^2\sigma_i^2\right)$"*

Demostración Sean $X_1, X_2, \ldots, X_n$ v.a i.i.d normal (μ_i, σ_i^2). Defínase $Y = \sum_{i=1}^{n} a_i X_i$ Observe que $-\infty < y < \infty$. Se hará uso de las relaciones [3.15] y [3 16*bis*].

Por hipótesis se tiene que

$$X_i \sim N(\mu_i, \sigma_i^2), \qquad i = 1, 2, \ldots, n.$$

Entonces de [3.15]

$$a_iX_i \sim N(a_i\mu_i, a_i^2\sigma_i^2), \qquad i = 1, 2, \ldots, n,$$

y de [3 16*bis*]

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i\mu, \sum_{i=1}^{n} a_i^2\sigma^2\right).$$

◈

A su vez, la relación [3.17] permite verificar el resultado siguiente en torno a las distribución conjunta de la suma y la diferencia de dos v.a.i.i.d normal, que involucra además un dato curioso acerca de la distribución normal.

[3.18]

$$[\text{Normal}_1(\mu,\sigma^2)] - [\text{Normal}_2(\mu,\sigma^2)] \sim \text{Normal}(0,2\sigma^2)$$

"La diferencia de dos v.a.i. normal (μ,σ^2) también es normal $(0,2\sigma^2)$, y además las funciones X_1+X_2 y X_1-X_2 son independientes"

Demostración. Sean X_1 y X_2 v.a.i.i.d. normal con parámetros μ y σ^2. Defínanse $Y_1=X_1+X_2$ y $Y_2=X_1-X_2$ Observe que $-\infty<y_1,y_2<\infty$

Por el resultado [3.17] se sabe que:

$$Y_1=X_1+X_2\sim N(2\mu\,,2\sigma^2) \qquad (a_1=1,a_2=1)$$

y

$$Y_2=X_1-X_2\sim N(0,2\sigma^2) \qquad (a_1=1,a_2=-1)$$

Ahora, la f.d.p. conjunta de Y_1 y Y_2 puede calcularse recurriendo al *Teorema 1.8*:

$$f_{Y_1,Y_2}(y_1,y_2)=|J|\,f_{X_1}(g_1^{-1}(y_1,y_2))\,f_{X_2}(g_2^{-1}(y_1,y_2)) \quad , \quad -\infty<y_1,y_2<\infty$$

ya que X_1 y X_2 son independientes

Se tiene que

$$\left.\begin{aligned} y_1 &= x_1+x_2 \\ y_2 &= x_1-x_2 \end{aligned}\right\} \Rightarrow \quad \begin{aligned} x_1 &= \frac{y_1+y_2}{2} \\ x_2 &= \frac{y_1-y_2}{2} \end{aligned}$$

De donde:

$$J=\begin{vmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{vmatrix} = -\frac{1}{4}-\frac{1}{4}=-\frac{1}{2} \quad ; \qquad |J|=\frac{1}{2}$$

Así,
$$f_{Y_1,Y_2}(y_1,y_2)=\frac{1}{2}f_{X_1}\left(\frac{y_1+y_2}{2}\right)f_{X_2}\left(\frac{y_1-y_2}{2}\right), \qquad -\infty<y_1,y_2<\infty$$

$$=\frac{1}{2}\left\{\frac{1}{\sqrt{2\pi}\sigma}\exp-\left[\frac{\left(\frac{y_1+y_2}{2}-\mu\right)^2}{2\sigma^2}\right]\right\}\cdot\left\{\frac{1}{\sqrt{2\pi}\sigma}\cdot\exp-\left[\frac{\left(\frac{y_1-y_2}{2}-\mu\right)^2}{2\sigma^2}\right]\right\}$$

$$=\frac{1}{4\pi\sigma^2}\exp-\left\{\frac{[(y_1+y_2)-2\mu]^2}{8\sigma^2}+\frac{[(y_1-y_2)-2\mu]^2}{8\sigma^2}\right\}$$

Desarrollando el exponente:

$$[(y_1+y_2)-2\mu]^2+[(y_1-y_2)-2\mu]^2$$
$$=2[(y_1-4\mu y_1+4\mu^2)+y_2^2]$$
$$=2[(y_1-2\mu)^2+y_2^2]$$

Sustituyendo la expresión anterior en $f_{Y_1,Y_2}(y_1,y_2)$.

$$f_{Y_1,Y_2}(y_1,y_2)=\frac{1}{4\pi\sigma^2}\exp-\left\{\frac{2[(y_1-2\mu)^2+y_2^2]}{8\sigma^2}\right\}$$

$$=\frac{1}{4\pi\sigma^2}\exp-\left\{\frac{(y_1-2\mu)^2+y_2^2}{4\sigma^2}\right\}$$

$$=\underbrace{\frac{1}{\sqrt{2\pi}(2\sigma)}\exp-\left\{\frac{(y_1-2\mu)^2}{4\sigma^2}\right\}}_{Y_1\sim N(2\mu,2\sigma^2)}\cdot\underbrace{\frac{1}{\sqrt{2\pi}(2\sigma)}\exp-\left\{\frac{y_2^2}{4\sigma^2}\right\}}_{Y_2\sim N(0,2\sigma^2)}, \qquad -\infty<y_1,y_2<\infty$$

de donde no sólo se verifica que Y_1 es normal $(2\mu,2\sigma^2)$, (como ya se sabia por [3.17]) y que Y_2 es normal $(0.2\sigma^2)$, sino también se observa que son independientes

◊

Ross [1988] comenta que es posible demostrar que si X_1 y X_2 son v.a.i con función de distribución común F_X, entonces X_1+X_2 será independiente de X_1-X_2 si y sólo si F_X es una función de distribución normal.

◊ ◊

La siguiente relación establece el procedimiento a seguir para *estandarizar* cualquier variable aleatoria normal (μ,σ^2), es una relación recíproca que permite regresar al punto de partida.

[3.19]

$$\frac{[\text{Normal}(\mu,\sigma^2)]-\mu}{\sigma} \sim \text{Normal } (0,1)$$

"Una v.a. normal (μ,σ^2), X, puede estandarizarse mediante la transformación definida por $Z = (X - \mu)/\sigma$"

Demostración. Sea X una v.a. normal (μ, σ^2). Defínase $Z = \dfrac{X-\mu}{\sigma}$ Observe que $-\infty < z < \infty$. El resultado [3.19] se obtiene directamente aplicando la relación [3.15] con $a = 1/\sigma$ y $b = -\mu\ \sigma$

Como ya se mencionó en la presentación de la distribución normal, el cálculo de la integral $\int_{-\infty}^{t} \exp-\left[\frac{(x-\mu)^2}{2\sigma^2}\right]dx$ es complejo. Así, la importancia de la relación [3.19] radica en que permite calcular la probabilidad acumulada de *cualquier* variable aleatoria normal con media μ y varianza σ^2, ya que

$$F_X(x) = \text{P}[X \leq x] = \text{P}[(X-\mu)/\sigma \leq (x-\mu)/\sigma] = \Phi((x-\mu)/\sigma) = \Phi(z)$$

Los valores de $\Phi(z)$ pueden encontrarse en tablas dentro de la literatura estadística.

◆

De manera recíproca, si se aplica la transformación inversa a una variable aleatoria normal estándar, se obtendrá nuevamente una variable aleatoria normal con parámetros μ y σ^2, como se verificará en la relación [3.20].

[3.20]

$$\boxed{\mu\,[\text{Normal}\,(0,1)] + \sigma \sim \text{Normal}\,(\mu,\sigma^2)}$$

"Una v.a. normal estándar, Z, bajo la transformación definida por $X = \mu + \sigma Z$ se distribuye normal (μ,σ^2)"

Demostración. Sea Z una v.a normal (0,1). Defínase $X = \mu + \sigma Z$. Nuevamente, como X es una transformación lineal de Z, la conclusión se obtiene sin problema aplicando directamente la relación [3 15] con $a = \sigma$ y $b = \mu$.

◈

- ***La distribución normal como distribución límite***

[3.21]

$$\boxed{\text{Binomial}\,(n,p) \xrightarrow[n\to\infty]{\substack{\mu=np\\ \sigma^2=np(1-p)}} \text{Normal}\,(np, np(1-p))}$$

"La distribución binomial (n,p) puede ser aproximada por la distribución normal (μ,σ^2) cuando n es suficientemente grande, siendo μ y σ^2 la media y la varianza de la distribución binomial ($\mu = np$ y $\sigma^2 = np(1-p)$)"

Demostración:

Sean $X_1, X_2, \ldots, X_n$ v a i.i.d. Bernoulli (p) La media y la varianza de cada X_i son $\mu' = p$ y $\sigma'^2 = p(1-p)$, respectivamente

Defínase $Y = \sum_{i=1}^{n} X$. Entonces Y es una v.a. binomial (n, p) (véase relación [3.2]), $y = 0, 1, 2, \ldots, n$, cuya media y varianza están dadas por $\mu = n\mu' = np$ y $\sigma^2 = n\sigma'^2 = np(1-p)$.

Por el *Teorema de Límite Central* (véase *Teorema 2.16.1* en la sección sobre la distribución normal en el Capítulo 2) se sabe que

$$\frac{X_1 + X_2 + \ldots + X_n - n\mu'}{\sigma'\sqrt{n}} = \frac{Y - n\mu'}{\sigma'\sqrt{n}}$$

tiende a la normal estándar cuando $n \to \infty$; lo que equivale a decir (por la relación [3.19]) que Y se distribuye normal con media $n\mu' = np = \mu$ y varianza $\left(\sigma'\sqrt{n}\right)^2 = n\sigma'^2 = np(1-p) = \sigma^2$, cuando n es suficientemente grande.

◆

En la *Figura 3.1* se presenta una serie de ilustraciones de la aproximación binomial–normal

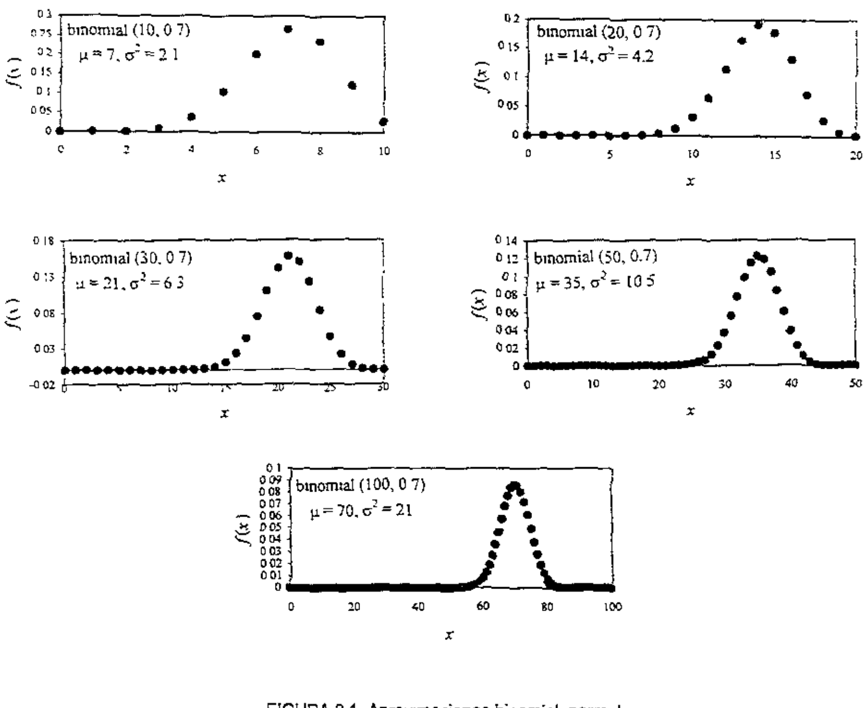


FIGURA 3.1 Aproximaciones binomial–normal

[3.22]

$$\text{Poisson}(\lambda) \xrightarrow[\lambda \to \infty]{\mu=\sigma^2=\lambda} \text{Normal}(\lambda,\lambda)$$

"La distribución Poisson (λ) puede ser aproximada por la distribución normal (μ,σ^2) cuando λ es suficientemente grande, siendo μ y σ^2 la media y la varianza de la distribución Poisson ($\mu = \lambda$ y $\sigma^2 = \lambda$)"

Demostración:

Sean $X_1, X_2, \ldots, X_\lambda$ v.a.i.i.d Poisson (1), para algún entero positivo λ. Entonces para cada X_i la media y la varianza son $\mu' = 1$ y $\sigma'^2 = 1$, respectivamente

Defínase $Y = \sum_{i=1}^{n} X_i$. Entonces Y es una v.a Poisson (λ) (véase relación [3.6]), $y = 0, 1, 2, \ldots$, cuya media y varianza están dadas por $\mu = \lambda\mu' = \lambda$ y $\sigma^2 = \lambda\sigma'^2 = \lambda$.

Por el *Teorema de Límite Central* (véase *Teorema 2.16.1* en la sección sobre la distribución normal en el Capítulo 2) se sabe que

$$\frac{X_1 + X_2 + \ldots + X_\lambda - \lambda\mu'}{\sigma'\sqrt{\lambda}} = \frac{Y - \lambda\mu'}{\sigma'\sqrt{\lambda}}$$

tiende a la normal estándar cuando $\lambda \to \infty$; lo que equivale a decir (por la relación [3.19]) que Y se distribuye normal con media $\lambda\mu' = \lambda = \mu$ y varianza $(\sigma'\sqrt{\lambda})^2 = \lambda\sigma'^2 = \lambda = \sigma^2$, cuando λ es suficientemente grande.

◈

En la *Figura 3.2* se presenta una serie de ilustraciones de la aproximación Poisson–normal.

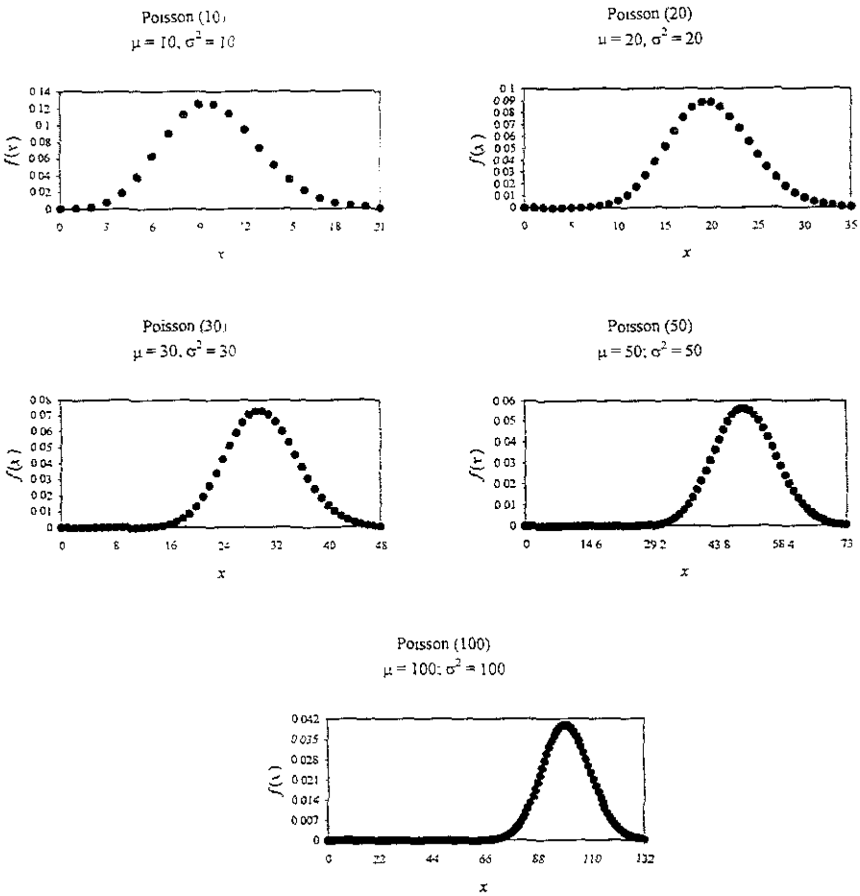


FIGURA 3.2. Aproximaciones Poisson–normal

[3.23]

$$\text{Gamma}\,(\alpha,\beta) \xrightarrow[\alpha\to\infty]{\substack{\mu=\alpha\beta \\ \sigma^2=\alpha\beta^2}} \text{Normal}\,(\alpha\beta,\alpha\beta^2)$$

"La distribución gamma (α,β) puede ser aproximada por la distribución normal (μ,σ^2) cuando α es suficientemente grande, siendo μ y σ^2 la media y la varianza de la distribución gamma ($\mu = \alpha\beta$ y $\sigma^2 = \alpha\beta^2$)"

Demostración

Sean $X_1, X_2, \ldots, X_\alpha$ v a i i.d. gamma (1,β), para algún entero positivo α. Así, para cada X_i la media y la varianza son $\mu' = \beta$ y $\sigma'^2 = \beta^2$, respectivamente.

Defínase $Y = \sum_{i=1}^{n} X_i$ Entonces Y es una v.a gamma (α,β) (véase relación [3.26]), $y > 0$, cuya media y varianza están dadas por $\mu = \alpha\mu' = \alpha\beta$ y $\sigma^2 = \alpha\sigma'^2 = \alpha\beta^2$.

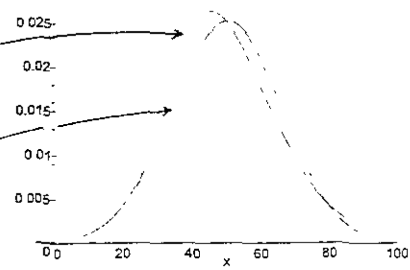
Por el *Teorema de Límite Central* (véase *Teorema 2.16.1* en la sección sobre la distribución normal en el Capítulo 2) se sabe que

$$\frac{X_1 + X_2 + \ldots + X_\alpha - \alpha\mu'}{\sigma'\sqrt{\alpha}} = \frac{Y - \alpha\mu'}{\sigma'\sqrt{\alpha}} =$$

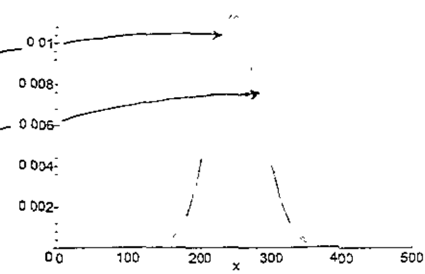
tiende a la normal estándar cuando $\alpha \to \infty$, lo que equivale a decir (por la relación [3.19]) que Y se distribuye normal con media $\alpha\mu' = \alpha\beta = \mu$ y varianza $\left(\sigma'\sqrt{\alpha}\right)^2 = \alpha\sigma'^2 = \alpha\beta^2 = \sigma^2$, cuando α es suficientemente grande

◊

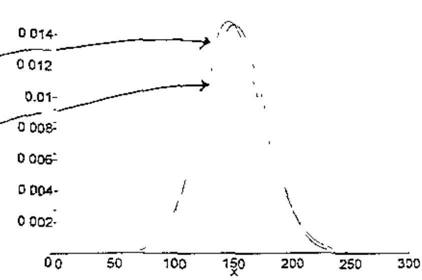
En la *Figura* 3.3 se presenta una serie de ilustraciones de la aproximación gamma–normal.



Gamma $(x;10.5)$, Normal $(x;50,250)$



Gamma $(x,50,5)$, Normal $(x;250.1250)$



Gamma $(x;30.5)$, Normal $(x;150,750)$

FIGURA 3 3. Aproximaciones gamma–normal

[3.24]

$$\text{Beta}\ (\beta,\gamma) \xrightarrow{\beta=\gamma\to\infty} \text{Normal}\left(\frac{1}{2},\frac{1}{8\beta+4}\right)$$

"La distribución beta (β,γ) puede ser aproximada por la distribución normal (μ,σ^2) cuando ambos parámetros (β y γ) son iguales y suficientemente grandes, siendo μ y σ^2 la media y la varianza de la distribución beta cuando $\beta=\gamma$"

Demostración

Nota: El *Teorema de Límite Central* avala la validez de la relación [3.24] Sin embargo, para la distribución beta resulta poco práctico detallar la aplicación de ese teorema (como se hizo para las relaciones [3.21] a [3.23]) debido a la complejidad de las expresiones para la media y la varianza de la distribución, así como de la función generadora de momentos que se requeriría para determinar la distribución de la suma de v.a i.i d. beta. Por ello, se prefirió presentar una breve discusión acerca de la aproximación beta–normal.

A diferencia de la relación [3 23], la aproximación beta–normal opera bajo condiciones sobre los *dos* parámetros de la distribución Esto se justifica porque sólo cuando $\beta=\gamma$ la densidad beta es simétrica (véanse *Propiedades* y *Figura 2.9* en la sección sobre la distribución beta en el Capítulo 2) y recuerde que la simetría es una característica de la densidad normal Esta simetría se da con respecto al valor ½, es decir, cuando $\beta=\gamma$ la media de la distribución es ½; esto se verifica fácilmente sustituyendo $\beta=\gamma$ en la expresión general de la media de la distribución beta:

$$\mu=\frac{\beta}{\beta+\gamma}=\frac{\beta}{2\beta}=\frac{1}{2}$$

Por otra parte, la varianza de la distribución beta (β,γ) es $\sigma^2=\dfrac{\beta\gamma}{(\beta+\gamma)^2(\beta+\gamma+1)}$, de tal manera que cuando $\beta=\gamma$ esta expresión se reduce a $\dfrac{1}{8\beta+4}$. Así, cuando $\beta\to\infty$ ocurre que $\sigma^2\to 0$, lo que indica que se trata de una normal muy estrecha, como puede apreciarse en la *Figura 3 4*

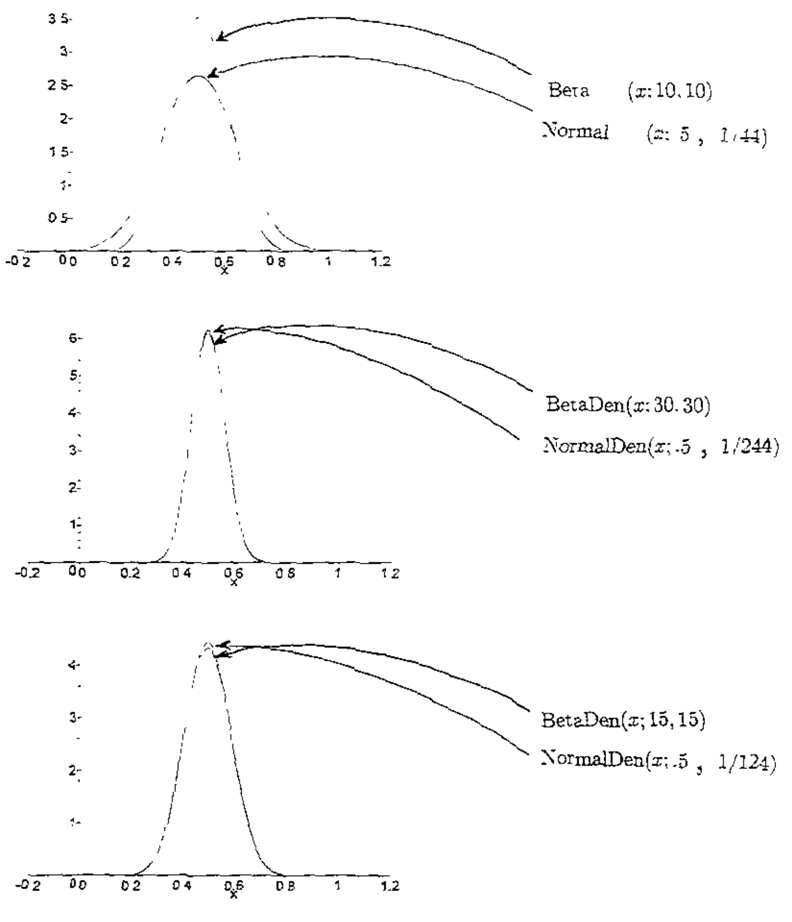


FIGURA 3.4 Aproximaciones beta-normal

[3.25]

$$t \text{ de Student } (n) \xrightarrow{n\to\infty} \text{Normal } (0,1)$$

"Una v.a. t de Student (n) puede ser aproximada por la distribución normal estándar cuando el parámetro n es suficientemente grande"

Demostración

Nuevamente, el *Teorema de Límite Central* permitirá verificar la validez de la relación [3.25]. Cabe señalar que para la distribución t es un poco más "delicado" detallar la aplicación del teorema, en el sentido de que se requiere acudir a la relación entre la distribución t y la Cauchy, la cual se verificará en la sección correspondiente a esta última distribución

Observe que en este caso, en el diagrama de Leemis, la relación se está estableciendo directamente entre la distribución t y la normal *estándar*. Una justificación para ello es que la media de una variable aleatoria $t_{(n)}$ es *cero* para toda $n > 1$. Además, el valor de la varianza está dado por el cociente $n/(n-2)$ (véanse *Propiedades* en la sección sobre la distribución t de Student en el Capítulo 2), y así cuando el parámetro n es suficientemente grande el valor de la varianza tiende a *uno* Con estos elementos en consideración se procederá a la aplicación del *Teorema de Límite Central.*

Sean $X_1, X_2, \ldots, X_n$ v a i.i.d. Cauchy $(0,1/n)$ Defínase $Y = \sum_{i=1}^{n} X_i$ Observe que como, para cada X_i, $-\infty < x_i < \infty$ entonces $-\infty < y < \infty$. Ahora, la distribución Cauchy (al igual que la t de Student) no tiene media ni varianza, ni, en general, función generadora de momentos. Por ello se recurrirá a la relación [3.56] que establece que la suma de v.a.i i.d. Cauchy (a, α) se distribuye Cauchy (na, $n\alpha$) Entonces Y es una v.a Cauchy (0,1)

Defínase $Y = \sum_{i=1}^{n} X_i$. Entonces Y es una v.a. binomial (n,p) (véase relación [3 2]), $y = 0,1,2,\ldots,n$, cuya media y varianza están dadas por $\mu = n\mu' = np$ y $\sigma^2 = n\sigma'^2 = np(1-p)$

Por el *Teorema del Límite Central* (véase *Teorema 2.16.1* en la sección sobre la distribución normal en el Capítulo 2) se sabe que

$$\frac{X_1 + X_2 + \dots + X_n - n\mu'}{\sigma'\sqrt{n}} = \frac{Y - n\mu'}{\sigma'\sqrt{n}}$$

tiende a la normal estándar cuando $n \to \infty$, lo que equivale a decir (por la relación [3.19]) que Y se distribuye normal con media $n\mu' = np = \mu$ y varianza $(\sigma'\sqrt{n})^2 = n\sigma'^2 = np(1-p) = \sigma^2$, cuando n es suficientemente grande.

En la *Figura 2.19* (Capítulo 2) se presenta una gráfica que compara densidades t de Student con la densidad normal estándar

- ***Otras relaciones con la distribución normal***

Véase [3.30] en la sección *Relaciones con la distribución gamma.*
Véase [3.46] en la sección *Relaciones con la distribución ji-cuadrada.*
Véase [3.57] en la sección *Relaciones con la distribución Cauchy.*
Véanse [3.64] y [3.65] en Otras relaciones entre distribuciones univariadas continuas

♠

RELACIONES CON LA DISTRIBUCIÓN GAMMA

- ***Relaciones de la distribución gamma con ella misma***

[3.26]

$$\sum_{1}^{n} \text{Gamma}(\alpha,\beta) \sim \text{Gamma}\,(n\alpha,\beta)$$

"La suma de n v.a.i.i.d. gamma (α,β) también se distribuye gamma ($n\alpha,\beta$)"

(La distribución gamma satisface la propiedad reproductiva)

Demostración Sean $X_1, X_2, \ldots, X_n$ v.a i.i d. gamma (α,β). Defínase $Y = \sum_{i=1}^{n} X_i$. Observe que $y > 0$. Aplicando la *Técnica de la función generadora de momentos* vía el *Resultado 1.18*:

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\ &= \prod_{i=1}^{n}\left(\frac{1}{1-\beta t}\right)^{\alpha} \qquad \text{para} \quad t < \frac{1}{\beta} \\ &= \left(\frac{1}{1-\beta t}\right)^{n\alpha} \qquad \text{(ya que la expresión } \left(\frac{1}{1-\beta t}\right)^{\alpha} \text{ no depende de } i\text{)} \end{aligned}$$

que corresponde a la f.g m de una distribución gamma con parámetros $n\alpha$ y β.

◇

La relación [3.26] puede generalizarse al caso en que uno de los *parámetros* de la distribución gamma de las X_i *no* fuera el valor *común* α, sino que $X_1, X_2, \ldots, X_n$ fueran v.a.i.i.d. Gamma (α_i,β) Esto se expresa en [3.26*bis*], la cual puede demostrarse también mediante la *Técnica de la función generadora de momentos*.

◇ ◇

[3.26 *bis*]

$$\sum_{1}^{n} \text{Gamma}(\alpha_i, \beta) \sim \text{Gamma}\left(\sum_{1}^{n} \alpha_i, \beta\right)$$

"La suma de n v.a.i.i.d. gamma (α_i, β), $i = 1, 2, \ldots, n$, se distribuye también gamma con parámetros $\left(\sum_{1}^{n} \alpha_i, \beta\right)$"

Demostración. Sean $X_1, X_2, \ldots, X_n$ v.a.i.i.d. gamma (α_i,β). Defínase $Y = \sum_{i=1}^{n} X_i$. Observe que $y > 0$. Acudiendo a la *Técnica de la función generadora de momentos* (*Resultado 1.18*):

$$
\begin{aligned}
M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\
&= \prod_{i=1}^{n} \left(\frac{1}{1-\beta t}\right)^{\alpha_i} \qquad \text{para} \quad t < \frac{1}{\beta} \\
&= \left(\frac{1}{1-\beta t}\right)^{\sum_{i=1}^{n} \alpha_i}
\end{aligned}
$$

que corresponde a la función generadora de momentos de una variable aleatoria gamma $\left(\sum_{1}^{n} \alpha_i, \beta\right)$

◆

A continuación se presentan los casos especiales de la distribución gamma; a saber: las distribuciones Erlang (relación [3.27]), exponencial ([3.28]) y Ji-cuadrada ([3.29]).

- *Casos especiales de la distribución gamma*

[3.27]

$$\boxed{\text{Gamma } (n,\beta) \equiv \text{Erlang } (n,\beta)}$$

"La distribución Erlang (n,β) es un caso especial de la distribución gamma (α,β) cuando α es un entero positivo"

Demostración. Sea X una v.a con distribución gamma (n,β), con n un entero positivo El resultado se obtiene fácilmente sustituyendo el valor $\alpha = n$ en la expresión de la densidad gamma $f_X(x,\alpha,\beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1}e^{-x/\beta}$ y considerando el hecho de que $\Gamma(n) = (n-1)!$

$$\begin{aligned} f_X(x;n,\beta) &= \frac{1}{\Gamma(n)\beta^n} x^{n-1}e^{-x/\beta} \quad , \qquad x > 0 \\ &= \frac{1}{(n-1)!\beta^n} x^{n-1}e^{-x/\beta} \end{aligned}$$

que corresponde a la f.d.p Erlang (n,β)

◈

La distribución Erlang es poco común en la literatura, es decir, cuando el parámetro α de la distribución gamma es un entero positivo no se enfatiza que se está tratando con la distribución Erlang. Es así que generalmente se plantea que la suma de v.a.i.i.d exponencial se distribuye gamma y no Erlang.

[3.28]

$$\text{Gamma }(1,\beta) \equiv \text{Exponencial }(\beta)$$

"La distribución exponencial (β) es un caso especial de la distribución gamma (α,β) cuando $\alpha = 1$"

Demostración. Sea X una v.a. con distribución gamma $(1,\beta)$. El resultado se obtiene sin problema sustituyendo el valor $\alpha = 1$ en la expresión de la densidad gamma $f_X(x;\alpha,\beta) = \dfrac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}$ y tomando en cuenta que $\Gamma(1) = 1$.

$$\begin{aligned} f_X(x;1,\beta) &= \frac{1}{\Gamma(1)\beta^{1}} x^{1-1} e^{-x/\beta} \quad , \qquad x > 0 \\ &= \frac{1}{\beta} e^{-x/\beta} \end{aligned}$$

que corresponde a la f.d.p. de una v.a. exponencial (β)

♦

[3.29]

$$\boxed{\text{Gamma } (n/2,2) \equiv \text{Ji-cuadrada } (n)}$$

"La distribución Ji-cuadrada (n) es un caso especial de la distribución gamma (α, β) cuando $\alpha = n/2$ y $\beta = 2$"

Demostración Sea X una v.a con distribución gamma $(n/2,2)$ El resultado se obtiene directamente sustituyendo los valores $\alpha = n/2$ y $\beta = 2$ en la expresión de la densidad gamma

$$f_X(x;\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}$$

$$f_X(x;n/2,2) = \frac{1}{\Gamma(n/2)2^{n/2}} \cdot x^{\frac{n}{2}-1} e^{-x/2} \quad , \qquad x > 0$$

◊

La relación [3.29] será fundamental para la verificación de algunas relaciones con la distribución Ji-cuadrada Si bien es posible justificarlas de manera directa, se prefirió hacerlo apelando al hecho de que la distribución Ji-cuadrada es un forma especial de la distribución gamma y como tal –en este caso– "hereda" las propiedades de ésta (véase *Relaciones con la distribución Ji-cuadrada*)

◊ ◊

- ***Relaciones de la distribución gamma con otras distribuciones mediante transformaciones***

[3.30]

$$\boxed{[\text{Normal }(0.1)]^2 \sim \text{Gamma }(½.2)}$$

***"El cuadrado de una v.a. normal estándar, Z, se distribuye gamma** (½.2)"*

Demostración. Sea Z una v.a. normal (0.1). Defínase $Y = Z^2$. Como $-\infty < z < \infty$, observe que $y > 0$. Aplicando la *Técnica de la transformación* mediante el *Resultado 1 1* y tomado en cuenta el hecho de que $\Gamma(½) = \sqrt{\pi}$:

$$\begin{aligned}
f_Y(y) &= \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})], \qquad y > 0 \\
&= \frac{1}{2\sqrt{y}}\left[\frac{1}{\sqrt{2\pi}}\cdot e^{-(\sqrt{y})^2/2} + \frac{1}{\sqrt{2\pi}}\cdot e^{-(-\sqrt{y})^2/2}\right] \\
&= \frac{1}{2\sqrt{y}}\left[\frac{1}{\sqrt{2\pi}}\cdot e^{-y/2} + \frac{1}{\sqrt{2\pi}}\cdot e^{-y/2}\right] \\
&= \frac{1}{2\sqrt{y}}\left[2\cdot\frac{e^{-y/2}}{\sqrt{2\pi}}\right] \\
&= \frac{1}{\sqrt{2}\sqrt{\pi}\sqrt{y}}\cdot e^{-y/2} \\
&= \frac{1}{2^{1/2}\Gamma\left(\frac{1}{2}\right)}\cdot y^{-1/2}\cdot e^{-y/2} \\
&= \frac{1}{2^{1/2}\Gamma\left(\frac{1}{2}\right)}\cdot y^{\frac{1}{2}-1}\cdot e^{-y/2}
\end{aligned}$$

que corresponde a la f.d.p Gamma (½.2) (que a su vez, apelando a [3.29], es una Ji-cuadrada (1)).

◆

- ***Otras relaciones con la distribución gamma***

Véase también [3.23] en la sección *Relaciones con la distribución normal.*

Véase también [3.60] en la sección *Relaciones con la distribución beta*

RELACIONES CON LA DISTRIBUCIÓN EXPONENCIAL

- ***Relaciones que exhiben el paralelismo entre las distribuciones exponencial y geométrica*** [9]

[3.31]

$$\boxed{\min_{1\le i\le n}\ \{\text{Exponencial}\ (\alpha)\} \sim \text{Exponencial}\ (n/\alpha)}$$

***"La mínima estadística de orden de una distribución exponencial** (α)*

es también exponencial** (α')**, con** $\alpha' = \frac{n}{\alpha}$ **"

Demostración

Sean $X_1, X_2, \ldots, X_n$ v.a.i.i.d exponencial (α). Defínase $Y = \text{mín}\{X_1, X_2, \ldots, X_n\}$. Observe que como $x_i > 0$, para cada X_i, entonces $y > 0$ Para determinar la distribución de Y se recurrirá a la *Técnica de la función de distribución* vía el *Resultado 1.17*:

- $$f_Y(y) = n\,[1 - F_X(y)]^{n-1} f_X(y) \quad , \qquad y > 0$$

Calculando $F_X(y)$

- $$F_X(x) = P[X \le x] = \int_0^x \frac{1}{\alpha} e^{-u/\alpha} du = -e^{-u/\alpha}\Big|_0^x = -e^{-x/\alpha} + 1 = 1 - e^{-x/\alpha}$$
-

Entonces

$$
\begin{aligned}
f_Y(y) &= n\,[1-F_X(y)]^{n-1} f_X(y) \quad , \qquad y>0 \\
&= n[1-(1-e^{-y/\alpha})]^{n-1} \cdot \frac{1}{\alpha} e^{-y/\alpha} \\
&= n[e^{-y/\alpha}]^{n-1} \cdot \frac{1}{\alpha} e^{-y/\alpha} \\
&= \frac{n}{\alpha} e^{-\left(\frac{n}{\alpha}\right)y} \cdot e^{y/\alpha} e^{-y/\alpha} \\
&= \frac{n}{\alpha} e^{-\left(\frac{n}{\alpha}\right)y} \quad , \quad y>0
\end{aligned}
$$

que puede identificarse como la f.d.p. de una v.a. exponencial con parámetro $\frac{n}{\alpha}$.

◆

Recuerde que la propiedad de que la mínima estadística de un conjunto de v.a.i.i.d. tenga –aunque con un parámetro alterno– la misma distribución que las variables "originales" es también aplicable a la distribución geométrica (véase [3.12]), como puede verificarse en la sección correspondiente, lo que exhibe el paralelismo entre este par de distribuciones discreta y continua, respectivamente. La relación [3.32] que se presenta enseguida da cuenta también de ese paralelismo (en la sección *Relaciones con la distribución geométrica* puede consultarse el símil discreto).

◆◆

[9] *Recuerde que tanto la distribución geométrica como la exponencial registran tiempos de espera y cumplen* también la propiedad de la "pérdida de memoria".

[3.32]

$$\boxed{\text{Weibull}\ (\alpha,1) \equiv \text{Exponencial}\ (\alpha)}$$

"Una v.a. con distribución Weibull $(\alpha, 1)$ es una exponencial (α)"

(La distribución exponencial (α) es un caso especial de la distribución Weibull (α, β) cuando $\beta = 1$)

Demostración.

Sea X una v.a. Weibull $(\alpha, 1)$ La relación se prueba sin problema con sólo sustituir el valor $\beta = 1$ en la expresión de la f.d.p Weibull $f_X(x,\alpha,\beta) = \frac{1}{\alpha}\cdot\beta\, x^{\beta-1}\ \exp-\left[\frac{1}{\alpha}\cdot x^{\beta}\right],\quad x>0$

$$\begin{aligned} f_X(x;\alpha,1) &= \frac{1}{\alpha}\cdot 1\cdot x^{1-1}\cdot\exp-\left[\frac{1}{\alpha}\cdot x^{1}\right], \qquad x>0 \\ &= \frac{1}{\alpha}\cdot e^{-x/\alpha} \end{aligned}$$

que corresponde a la f d p. exponencial (α).

◈

El resultado [3.32] es el análogo continuo de la relación [3.13] que involucra a las distribuciones geométrica y Weibull discreta (véase *Relaciones con la distribución geométrica*). Más adelante se presentarán otros vínculos entre las distribuciones exponencial y Weibull a través de transformaciones

◈ ◈

- ***Casos especiales de los que surge la distribución exponencial***

Recuerde que la distribución exponencial es un caso especial de la distribución gamma (véase [3.28]), así como de la distribución Weibull (véase [3 32]). Asimismo, en la sección de *Relaciones con la distribución ji-cuadrada*, se verificará que es también un caso especial de esa distribución (véase [3 45]) Resta por verificar entonces la relación [3.33]

[3.33]

$$\boxed{\text{Erlang } (1,\beta) \equiv \text{Exponencial } (\beta)}$$

"La distribución exponencial (β) es un caso especial de la distribución Erlang (n, β) cuando n = 1"

Demostración. Sea X una v.a. con distribución Erlang $(1,\beta)$. El resultado se obtiene fácilmente sustituyendo el valor $n = 1$ en la expresión de la densidad Erlang: $f_X(x;n,\beta) = \frac{1}{(n-1)!\beta^n} x^{n-1} e^{-x/\beta}$, x

$$\begin{aligned} f_X(x;1,\beta) &= \frac{1}{(1-1)!\beta^1} x^{1-1} e^{-x/\beta} \quad , \quad x > 0 \\ &= \frac{1}{\beta} \cdot e^{-x/\beta} \end{aligned}$$

que corresponde a la f.d.p. exponencial (β).

♦

Otra forma de verificar la relación [3.33] (que acude a otros resultados) es apoyándose en [3.27] y [3.28] (véase *Relaciones con la distribución gamma*) y aplicando la propiedad transitiva: al ser X una v.a Erlang $(1,\beta)$ es también una gamma $(1,\beta)$ y como una v.a. exponencial (β) es también gamma $(1,\beta)$ entonces las distribuciones Erlang $(1,\beta)$ y exponencial (β) son equivalentes.

♦♦

El vínculo entre las distribuciones Erlang y exponencial es bilateral, pues a partir de la segunda se puede obtener la primera como se expone a continuación.

[3.34]

$$\sum_{1}^{n} \text{Exponencial}(\beta) \sim \text{Erlang}\ (n,\beta)$$

"La suma de n v.a.i.i.d. exponencial (β) se distribuye Erlang (n,β)"

Demostración. Sean $X_1, X_2, \ldots, X_n$ v.a.i.i.d. exponencial (β). Defínase $Y = \sum_{i=1}^{n} X_i$ Observe que $y > 0$. Aplicando la *Técnica de la función generadora de momentos* vía el *Resultado 1 18.*

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{n} M_{X_i}(t) \\ &= \prod_{i=1}^{n} \left(\frac{1}{1-\beta t} \right) \qquad \text{para} \quad t < \frac{1}{\beta} \\ &= \left(\frac{1}{1-\beta t} \right)^{n} \qquad \text{(ya que la expresión} \left(\frac{1}{1-\beta t} \right) \text{no depende de } i\text{)} \end{aligned}$$

que corresponde a la f.g.m. de una distribución gamma con parámetros n y β, que (por la relación [3 27]) es lo mismo que decir una distribución Erlang con parámetros n y β.

◊

Como ya se ha advertido, la distribución Erlang es poco común en la literatura. Generalmente se plantea que la suma de v.a.i.i.d. exponencial tiene distribución gamma

◊ ◊

- *Casos especiales de la distribución exponencial*

[3.35]

$$\boxed{\text{Exponencial } (2) \equiv \text{Ji-cuadrada } (2)}$$

"La distribución ji-cuadrada (2) es un caso especial de la distribución exponencial (α) cuando $\alpha = 2$"

Demostración Sea X una v.a. exponencial (2). El resultado se puede obtener directamente sustituyendo el valor $\alpha = 2$ en la expresión de la densidad exponencial $f_X(x;\alpha) = \frac{1}{\alpha} e^{-x/\alpha}$, $x > 0$ y manipulando adecuadamente un 1 para completar la densidad ji-cuadrada con parámetro 2.

Por otra parte, el resultado también puede verificarse acudiendo a otras relaciones y a la propiedad transitiva De la relación [3.28] (véase *Relaciones con la distribución gamma*) se infiere que X, que es exponencial (2), puede verse también como una v.a. gamma (1,2), que es lo mismo que decir gamma ($\frac{2}{2}$,2). Ahora, de la relación [3.29] se obtiene que una gamma ($\frac{2}{2}$,2) es también una ji-cuadrada (2). Por lo tanto, X puede verse como una ji-cuadrada (2).

◆

El razonamiento anterior justificará a su vez la relación [3.45], que es otra forma de enunciar [3.35] (véase *Relaciones con la distribución ji-cuadrada*).

A continuación se verifican los vínculos entre la distribución exponencial con otras distribuciones a *través de transformaciones*.

- ***Relaciones de la distribución exponencial con otras distribuciones mediante transformaciones***

[3.36]

$$\sqrt{\text{Exponencial}(\alpha)} \sim \text{Rayleigh}\ (\alpha)$$

"La raíz cuadrada de una v.a. exponencial (α) ***se distribuye Rayleigh*** (α)***"***

Demostración Sea X una v.a. con distribución exponencial (α). Defínase

$$Y = \sqrt{X}.$$

Observe que la transformación tiene sentido en tanto que $x > 0$, siendo $y > 0$. Se recurrirá a la *Técnica de la transformación*, mediante el *Resultado 1.2*, para determinar la f.d.p. de Y.

$$\begin{aligned} f_Y(y) &= 2y\ f_X(y^2), \quad y > 0 \\ &= 2y \cdot \frac{1}{\alpha} e^{-y^2/2} \end{aligned}$$

que corresponde a la f.d.p. Rayleigh (α)

◊

El resultado siguiente establece que a través de la transformación inversa puede pasarse de la distribución Rayleigh a la distribución exponencial.

[3.37]

$$[\text{Rayleigh }(\alpha)]^2 \sim \text{Exponencial }(\alpha)$$

"El cuadrado de una v.a. Rayleigh (α) se distribuye exponencial (α)"

Demostración. Sea X una v.a. con distribución Rayleigh (α). Defínase

$$Y = X^2.$$

Observe que $y > 0$. Y es una función monótona creciente ya que X es positiva. Entonces, se recurrirá a la *Técnica de la transformación*, mediante el *Resultado 1.1.* para determinar la f.d.p de Y.

$$\begin{aligned} f_Y(y) &= \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})], \qquad y > 0 \\ &= \frac{1}{2\sqrt{y}}\left[\frac{2\sqrt{y}}{\alpha}e^{-(\sqrt{y})^2/\alpha} + 0\right], \quad \text{pues} \quad -\sqrt{y} < 0 \\ &= \frac{1}{\alpha}\cdot e^{-y/\alpha}, \qquad y > 0 \end{aligned}$$

que corresponde a la f.d.p. exponencial (α)

♦

Aunque no son fáciles de verificar a primera vista, las relaciones [3.36] y [3.37] son casos particulares de los resultados [3.38] y [3.39], respectivamente, los cuales se presentan a continuación. La justificación de esto es que *la distribución Rayleigh es un especial de de la distribución Weibull*, lo cual se verifica dentro de la sección *Otras relaciones entre distribuciones univariadas continuas*, al final del capítulo.

♦♦

[3.38]

$$\boxed{[\text{Exponencial }(\alpha)]^{1/\beta} \sim \text{Weibull }(\alpha,\beta)}$$

"La transformación de una v.a. exponencial (α), X, definida por $X^{1/\beta}$ se distribuye Weibull (α,β)"

Demostración Sea X una v.a. con distribución exponencial (α). Defínase

$$Y = X^{1/\beta}$$

Observe que $y > 0$ Se recurrirá a la *Técnica de la transformación*, mediante el *Resultado 1.4*, para determinar la f.d.p. de Y.

$$r = \frac{1}{\beta}$$

$$\begin{aligned} f_Y(y) &= \beta\, y^{\beta-1} f_X(y^\beta) \quad , \qquad & y > 0 \\ &= \beta\, y^{\beta-1} \left\{ \frac{1}{\alpha} e^{-y^\beta/\alpha} \right\} & \\ &= \frac{1}{\alpha}\, \beta\, y^{\beta-1} e^{-y^\beta/\alpha} \quad , \qquad & y > 0 \end{aligned}$$

que corresponde a la f.d.p. Weibull (α,β)

◊

El siguiente resultado muestra cómo a través de la transformación inversa puede pasarse de una distribución Weibull a una distribución exponencial.

[3.39]

$$[\text{Weibull }(\alpha,\beta)]^{\beta} \sim \text{Exponencial }(\alpha)$$

"La transformación de una v.a. Weibull (α,β), X, definida por X^{β} se distribuye exponencial (α)"

Demostración. Sea X una v.a. con distribución Weibull (α,β). Defínase

$$Y = X^{\beta}.$$

Observe que $y > 0$ Además, Y es una función monótona creciente ya que X y β sólo toman valores positivos. Entonces, se recurrirá a la *Técnica de la transformación*, nuevamente mediante el *Resultado1.4*, para determinar la f.d.p. de Y.

$$r = \beta > 0 \quad , \qquad f_Y(y) = \frac{1}{\beta}\, y^{\frac{1}{\beta}-1} f_X(y^{1/\beta}) \quad , \qquad y > 0$$

$$= \frac{1}{\beta}\, y^{\frac{1}{\beta}-1} \left\{ \frac{1}{\alpha}\, \beta \left(y^{1/\beta}\right)^{\beta-1} \cdot \exp-\left[\frac{1}{\alpha}\left(y^{1/\beta}\right)^{\beta}\right] \right\}$$

$$= \frac{1}{\alpha}\, \beta \cdot \frac{1}{\beta}\, y^{\frac{1}{\beta}-1} \cdot y^{1-\frac{1}{\beta}} \exp-\left[\frac{1}{\alpha}\, y\right]$$

$$= \frac{1}{\alpha}\, e^{-y/\alpha} \quad , \qquad y > 0$$

que corresponde a la f.d.p. exponencial (α).

♦

El resultado [3.39] es la relación recíproca de [3.38] y es además una forma alternativa de vincular la distribución Weibull con la distribución exponencial, pues recuerde que en [3.32] se había verificado que la segunda es un caso espacial de la pnmera.

♦♦

[3.40]

$$\{\text{Exponencial}_1\ (\alpha) - \text{Exponencial}_2\ (\alpha)\} \sim \text{Laplace}\ (\alpha,\alpha)$$

"La diferencia de dos v.a.i. exponencial (α) se distribuye Laplace (α,α)"

Demostración. Sean X_1 y X_2 v a.i. con distribución exponencial (α). Defínase

$$Y = X_1 - X_2$$

Observe que como $x_1, x_2 > 0$ entonces $-\infty < y < \infty$. Como Y es una transformación bivariada, se recurrirá a la *Técnica de la transformación*, vía el *Teorema1.8*, para determinar la f d.p. de Y Para ello se requiere elegir otra función conveniente para construir una densidad conjunta y a partir de ella calcular la densidad marginal de la transformación de interés. Así, considere

$$Y_1 = X_1 - X_2. \qquad , \quad -\infty < y_1 < \infty$$

y

$$Y_2 = X_1 \qquad , \quad y_2 > 0$$

Calculando las funciones inversas:

$$\left.\begin{array}{l} y_1 = x_1 - x_2 \\ y_2 = x_1 \end{array}\right\} \Rightarrow \begin{array}{l} x_1 = y_2 \qquad\quad = g_1^{-1}(y_1, y_2) \\ x_2 = y_2 - y_1 = g_2^{-1}(y_1, y_2) \end{array}$$

de donde el Jacobiano resulta como.

$$J = \begin{vmatrix} 0 & 1 \\ -1 & 1 \end{vmatrix} = 0 - (-1) = 1 = |J|$$

Entonces aplicando el *Teorema 1.8*: $f_{Y_1,Y_2}(y_1,y_2) = J\ f_{X_1}(g_1^{-1}(y_1,y_2))f_{X_2}(g_2^{-1}(y_1,y_2))$

$$
\begin{aligned}
f_{Y_1Y_2}(y_1,y_2) &= \frac{1}{\alpha}e^{-y_2/\alpha}\cdot\frac{1}{\alpha}e^{-(y_2-y_1)/\alpha} \quad ; \qquad -\infty < y_1 < \infty\ ,\ y_2 > 0 \\
&= \frac{1}{\alpha^2}e^{-\frac{2y_2}{\alpha}-\frac{y_1}{\alpha}} \\
&= \frac{1}{\alpha^2}e^{\frac{y_1-2y_2}{\alpha}}
\end{aligned}
$$

A partir de la densidad conjunta anterior, calculando la función de densidad marginal de Y_1

$$
\begin{aligned}
f_{Y_1}(y_1) &= \int_0^\infty f_{Y_1,Y_2}(y_1,y_2)dy_2 \quad , \qquad -\infty < y_1 < \infty \\
&= \int_0^\infty \frac{1}{\alpha^2}e^{\frac{y_1-2y_2}{\alpha}}dy_2 \\
&= \frac{1}{\alpha}e^{\frac{y_1}{\alpha}}\int_0^\infty \frac{1}{\alpha}e^{-\frac{2}{\alpha}y_2}dy_2 \\
&= \frac{1}{\alpha}e^{\frac{y_1}{\alpha}}\left[-\frac{1}{2}e^{-\frac{2}{\alpha}y_2}\Big|_0^\infty\right] \\
&= \frac{1}{\alpha}e^{\frac{y_1}{\alpha}}\left[-\frac{1}{2}(0+1)\right] \\
&= -\frac{1}{2\alpha}e^{\frac{y_1}{\alpha}} \quad , \qquad -\infty < y_1 < \infty
\end{aligned}
$$

que corresponde a la f.d.p. Laplace (α,α).

◆

[3.41]

$$|\text{Laplace }(\alpha,\alpha)| \sim \text{Exponencial }(\alpha)$$

"El valor absoluto de una v.a. Laplace (α,α) se distribuye exponencial (α)"

Demostración. Sea X una v.a. con distribución Laplace (α,α). Defínase

$$Y = |X|.$$

Observe que $y \geq 0$. Se recurrirá a la *Técnica de la transformación*, vía el *Resultado 1.5*, para determinar la f.d.p. de Y. Recuerde que este resultado aplica porque Y es una transformación monótona decreciente para $-\infty < x \leq 0$ y monótona creciente $x > 0$. Entonces:

$$f_Y(y) = \frac{1}{2\sqrt{y}}[f_X(\sqrt{y}) + f_X(-\sqrt{y})], \qquad y \geq 0$$

Puesto que

$$f_X(x) = \begin{cases} \frac{1}{2\alpha} e^{-x/\alpha} & , \quad x \geqslant 0 \\ \frac{1}{2\alpha} e^{x/\alpha} & , \quad x < 0 \end{cases}$$

Entonces

$$f_Y(y) = \frac{1}{2\alpha} e^{-y/\alpha} \cdot I_{[0,\infty)}(y) + \frac{1}{2\alpha} e^{y/\alpha} \cdot I_{(-\infty,0]}(y)$$

Ya que $-\infty < -y \leq 0 \quad \Leftrightarrow \qquad 0 \leq y < \infty$, entonces $I_{(-\infty,0]}(-y) = I_{[0,\infty)}(y)$

Por lo tanto,

$$\begin{aligned} f_Y(y) &= \frac{1}{2\alpha} e^{-y/\alpha} \cdot I_{[0,\infty)}(y) + \frac{1}{2\alpha} e^{y/\alpha} \cdot I_{(-\infty,0)}(y) \\ &= 2\left(\frac{1}{2\alpha} \cdot e^{-y/\alpha} \cdot I_{[0,\infty)}(y)\right) \\ &= \frac{1}{\alpha} e^{-y/\alpha} \quad , \quad y \geq 0 \end{aligned}$$

que corresponde a la f.d.p. exponencial (α)

◆

Esta última conclusión requiere de algunas precisiones. Recuerde que la definición de la densidad exponencial considera que $y > 0$ (que y sea mayor *estrictamente* que cero). Sin embargo, en este caso, al trabajar sobre la transformación $Y = |X|$ se tiene que $y \geq 0$, puesto que el rango de la v.a. X *incluye al cero. Mas observe que esto no afecta de manera alguna la densidad exponencial*, en tanto que $f_Y(y)$ continúa siendo función de densidad; su gráfica difiere de aquélla definida para $y > 0$ solamente en un punto (el punto que corresponde a $y = 0$).

De hecho, esto justifica que Casella y Berger [1990] hayan definido la densidad exponencial (al igual que la densidad gamma) de una v.a., digamos Y, para $y \geq 0$, pero que en su tabla recopilatoria que se presenta al final del libro a dicha densidad se le asocie la condición $y > 0$.

◆◆

Hasta aquí se han verificado relaciones bilaterales de la distribución exponencial con otras distribuciones. *Las últimas dos relaciones a presentar dentro de esta sección, si bien se dan a través* de una transformación, a diferencia de las anteriores no son bilaterales.

[3.42]

$$-\alpha \log [\text{Uniforme } (0,1)] \sim \text{Exponencial } (\alpha)$$

"La transformación de una v.a. uniforme estándar, X, definida por $-\alpha \log X$ se distribuye exponencial (α)" [10]

Demostración Sea X una v.a con distribución uniforme estándar Defínase

$$Y = -\alpha \log X$$

Observe que $0 < x < 1 \Rightarrow -\infty < \log x < 0 \Rightarrow y = -\alpha \log x > 0$, ya que $\alpha > 0$.

Para determinar la distribución de Y se recurrirá la *Técnica de la transformación* mediante *Resultado 1 8* con $c = -\alpha < 0$.

$$\begin{aligned}
f_Y(y) &= \frac{1}{|-\alpha|} \cdot e^{-\frac{y}{\alpha}} \cdot f_X\left(e^{-\frac{y}{\alpha}}\right), && y > 0 \\
&= \frac{1}{\alpha} \cdot e^{-\frac{y}{\alpha}} \cdot f_X\left(e^{-\frac{y}{\alpha}}\right) \\
&= \frac{1}{\alpha} \cdot e^{-\frac{y}{\alpha}} \cdot 1 && \text{ya que como } y > 0 \text{ entonces } 0 < e^{-\frac{y}{\alpha}} < 1 \\
&= \frac{1}{\alpha} \cdot e^{-\frac{y}{\alpha}}, && y > 0
\end{aligned}$$

que corresponde a la f d.p exponencial (α).

◊

[10] La transformación está definida para el *logaritmo natural*

[3.43]

$$-\log\,[\text{Beta}\,(\beta,1)] \sim \text{Exponencial}\,(\beta)$$

"El menos logaritmo natural de una v.a. beta $(\beta, 1)$ se distribuye exponencial (β)"

Demostración. Sea X una v.a. con distribución beta $(\beta, 1)$. Defínase

$$Y_0 = -\log X$$

Observe que $0 < x < 1 \;\Rightarrow\; -\infty < \log x < 0 \;\Rightarrow\; y_0 = -\log x > 0.$

Al igual que en la relación anterior, para determinar la distribución de Y se recurrirá la *Técnica de la transformación* a través del *Resultado 1.8* con $c = -1 < 0$.

$$\begin{aligned}
f_{Y_0}(y) &= \frac{1}{|-1|} e^{y/(-1)} \cdot f_X\left(e^{y/(-1)}\right), \qquad y > 0 \\
&= e^{-y} \cdot f_X\left(e^{-y}\right) \\
&= e^{-y} \cdot \left\{ \frac{\Gamma(\beta+1)}{\Gamma(\beta)\Gamma(1)} \left(e^{-y}\right)^{\beta-1} (1-e^{-y})^{1-1} \right\} \\
&= e^{-y} \cdot \left\{ \frac{\beta \cdot \Gamma(\beta)}{\Gamma(\beta)} \left(e^{-y}\right)^{\beta} \cdot e^{y} \right\} \\
&= \beta\, e^{-\beta \cdot y}, \qquad y > 0
\end{aligned}$$

que corresponde a la f.d.p. exponencial (β).

◆

Observe que la transformación Y_0 es un caso particular de la transformación Y definida en [3.42]. Como se verificará más adelante, la distribución uniforme estándar (utilizada en [3.42]) es un caso especial de la distribución beta (utilizada en [3 43]; véase *Relaciones con la distribución beta*)

◆◆

- ***Otras relaciones con la distribución exponencial***

Véase también [3.28] en la sección *Relaciones con la distribución gamma*

Véase también [3 45] en la sección *Relaciones con la distribución ji-cuadrada*

RELACIONES CON LA DISTRIBUCIÓN JI-CUADRADA

- ***Relaciones de la distribución ji-cuadrada con ella misma***

[3.44]

$$\boxed{\sum_{1}^{r}[\text{Ji}-\text{cuadrada}(n)] \sim \text{Ji-cuadrada}(rn)}$$

"La suma de r v.a.i.i.d. ji-cuadrada (n) se distribuye ji-cuadrada (r n)"

Demostración. Sean $X_1, X_2, \ldots, X_r$ v a.i.i.d. ji-cuadrada (n) Defínase $Y = \sum_{i=1}^{r} X_i$. Observe que $y > 0$ Al igual que se ha hecho anteriormente cuando se tiene la suma de v.a.i.i d., en este caso es posible aplicar la *Técnica de la función generadora de momentos* via el *Resultado 1 18* para demostrar la relación. Sin embargo, existe otra forma alternativa de hacerlo que hace uso de resultados ya verificados previamente y que permiten insistir en el vínculo que existe entre las distribuciones ji-cuadrada y gamma.

De la relación [3.29] se sabe que la distribución ji-cuadrada con parámetro n es un caso especial de la distribución gamma con parámetros $\frac{n}{2}$ y 2. Ahora, de la relación [3.26] se obtiene que la suma de v.a.i.i.d gamma también tiene distribución gamma, en este caso Y es una gamma con parámetros $r \cdot \frac{n}{2} = \frac{r\,n}{2}$ y 2 Por lo tanto, al mismo tiempo, Y es una v a. ji-cuadrada con parámetro $r\,n$.

◇

En otras palabras al ser la distribución ji-cuadrada (n) un caso especial de la distribución gamma ($\frac{n}{2}$, 2), la primera "hereda" todas las propiedades de la segunda. Así, es también válido plantear el resultado [3.44*bis*] que se desprende de [3.26*bis*]. Sólo hay que tener cuidado con la aseveración de la "herencia", pues esto no siempre sucede. El ejemplo más claro de esto es que la distribución exponencial es también un caso especial de la distribución gamma, sin embargo la suma de v.a.i.i.d. exponencial *no* se distribuye exponencial, sino Erlang (que a su vez es una gamma).

◆◆

[3.44*bis*]

$$\boxed{\sum_{1}^{r}[\text{Ji}-\text{cuadrada}(n_i)] \sim \text{Ji}-\text{cuadrada}\left(\sum n_i\right)}$$

"La suma de r v.a.i. ji-cuadrada (n_i) se distribuye ji-cuadrada $\left(\sum_{i=1}^{r} n_i\right)$"

Demostración. Basta aplicar la relaciones [3.29] y [3.26*bis*] (véase demostración de [3.44]).

◆

- ***Casos especiales de la distribución ji-cuadrada***

[3.45]

Ji-cuadrada (2) $\equiv$ Exponencial (2)

"La distribución exponencial (2) es un caso especial de la distribución ji-cuadrada (n) cuando n = 2"

Demostración. Sea X una v.a. ji-cuadrada (2). El resultado se puede obtener directamente sustituyendo el valor $n = 2$ en la expresión de la densidad ji-cuadrada $f_X(x;n) = \dfrac{x^{n/2-1} \cdot e^{-x/2}}{2^{n/2}\Gamma\left(\frac{n}{2}\right)}$, $x > 0$ y aplicar el hecho de que $\Gamma(1) = 1$, con lo que se obtiene la densidad exponencial con parámetro 2.

◈

Por otra parte, el resultado también puede verificarse acudiendo al hecho de que una v.a gamma (1,2) puede verse también ya sea como una v.a. exponencial (2), (véase [3.28] dentro de *Relaciones con la distribución gamma*) o como una v.a. ji-cuadrada (2) (ya que una v.a. gamma (1,2) es lo mismo que decir una v.a. gamma ($\frac{2}{2}$,2); véase [3.29] también dentro de *Relaciones con la distribución gamma*). Por ende, las distribución de una v.a exponencial (2) coincide con la de una v.a. ji-cuadrada (2).

Observe que [3.45] es otra forma de enunciar [3.35] (véase *Relaciones con la distribución exponencial*)

◈ ◈

- ***Relaciones de la distribución ji-cuadrada con otras distribuciones mediante transformaciones***

[3.46]

$$\boxed{\sum_{1}^{n}[\text{Normal}\,(0,1)]^2 \sim \text{Ji-cuadrada}\,(n)}$$

"La suma de los cuadrados de n v.a.'s normal estándar se distribuye ji-cuadrada (n)"

Demostración. Sean $Z_1, Z_2, \ldots, Z_n$ v.a.i.i.d. normal estándar. Defínase $Y = \sum_{i=1}^{n} Z_i^2$. Observe que $y \geq 0$. Hay que tener cuidado con que Y es la suma de los *cuadrados* de las variables aleatorias Z_i, por lo cual no es aplicable la *Técnica de la función generadora de momentos* de manera directa vía el *Resultado 1.18*.

La demostración se desprende fácilmente de las relaciones [3.29] y [3.30] verificadas dentro de la sección de *Relaciones con la distribución gamma*. De [3.30] se sabe que el cuadrado de una v.a. normal estándar se distribuye gamma ($\frac{1}{2}$, 2). De [3.29] se tiene que una v.a. gamma ($\frac{1}{2}$, 2) es a su vez una ji-cuadrada (1). Finalmente, de [3.44] se deduce que Y se distribuye ji-cuadrada (n).

◆

De no haber contado con los resultados [3.29], [3.30] y [3.44] una alternativa para demostrar la relación [3.46] sería utilizando la *Técnica función generadora de momentos*, previo cálculo de la función generadora de momentos de Y. Para hacerlo se requiere hacer un desarrollo algebraico que si bien no es complicado si requiere de cierto ingenio, pues en el camino se deberá identificar una densidad normal $(0, 1/\sqrt{1-2t})$.[11]

---

[11] El desarrollo por este camino puede consultarse en Mood *et al.* [1974], pp. 242-243.

Por otra parte, aunque aquí se ha supuesto directamente que las Z_i son v a. normal estándar, tiene sentido suponer que estas Z_i sean v a. normales estandarizadas ($Z_i = (X_i - \mu)/\sigma$) a partir de v a normal (μ,σ^2), digamos X_i, para $i = 1,\ldots,n$, respectivamente (o incluso, siendo el caso más general provenientes de v a. normales $(\mu_i,\sigma_i^{\,2})$) En dicho caso, el resultado quedaría enunciado como sigue:

Sea $X_1, X_2,\ldots, X_n$ una m a. de una distribución normal (μ,σ^2), entonces $Y = \sum_{i=1}^{n} (X_i - \mu)^2/\sigma^2$ tiene distribución ji-cuadrada con n grados de libertad. Este resultado se aplica ampliamente en estadística para determinar cantidades pivotales para el cálculo de intervalos de confianza

Observación final. Aplicando los resultados [3.18] y [3 46] puede verificarse que si Z_1 y Z_2 son v.a.i. normal estándar entonces la distribución de Z_i^2 es la misma que la de $(Z_1 - Z_2)^2/2$; a saber, ambas son ji-cuadrada (1), que es lo mismo que decir gamma $(\frac{1}{2}, 2)$.

◊ ◊

[3.47]

$$\boxed{\frac{\dfrac{\text{Ji}-\text{cuadrada}\,(n_1)}{n_1}}{\dfrac{\text{Ji}-\text{cuadrada}\,(n_2)}{n_2}} \sim F\,(n_1\,, n_2)}$$

"El cociente de dos v.a.i. con distribución ji-cuadrada divididas entre sus respectivos grados de libertad tiene distribución F"

Demostración. Sean X_1 y X_2 v.a i con distribución ji-cuadrada con n_1 y n_2 grados de libertad, respectivamente. Considere Y una función bivariada $g(X_1,X_2)$ definida como

$$Y = \frac{X_1 / n_1}{X_2 / n_2}$$

Aquí el orden que se da a los grados de libertad es importante. En este caso, se busca mostrar que Y tiene distribución F con parámetros (n_1,n_2); los grados de libertad de la variable que aparece en el numerador se citan en primer lugar.

Se recurrirá a la *Técnica de la transformación* para determinar la distribución de Y, mediante el *Teorema 1.8*:

$$f_{Y_1,Y_2}(y_1,y_2)=|J|\,f_{X_1}(g_1^{-1}(y_1,y_2))f_{X_2}(g_2^{-1}(y_1,y_2))\,I_\Psi(y_1,y_2)$$

donde $\Psi=\{(y_1,y_2)$ para los cuales existe $(x_1,x_2)\in\vartheta$, tal que $(y_1,y_2)=(g_1(x_1,x_2),g_2(x_1,x_2))\}$.

Para ello se elegirá X_2 como la segunda función. Así:

$$Y_1=g_1(x_1,x_2)=\frac{X_1/n_1}{X_2/n_2}\quad ,\quad y_1>0$$
$$Y_2=g_2(x_1,x_2)=X_2\quad ,\quad y_2>0$$

Observe que como $x_1, x_2>0$ y n_1, n_2 son enteros positivos el recorrido de Y_1 es $y_1>0$.

Calculando primero las funciones inversas (expresando a x_1, x_2 como funciones de y_1, y_2):

$$\left.\begin{aligned} y_1&=\frac{x_1/n_1}{x_2/n_2}\\ y_2&=x_2\end{aligned}\right\}\Rightarrow\quad \begin{aligned} x_1&=\frac{n_1}{n_2}(y_1y_2)\\ x_2&=y_2\end{aligned}$$

Calculando el jacobiano:

$$J=\begin{vmatrix}\frac{n_1}{n_2}y_2 & \frac{n_1}{n_2}y_1\\ 0 & 1\end{vmatrix}=\frac{n_1}{n_2}y_2=|J|$$

Entonces

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2)&=|J|\cdot f_{X_1}(x_1)f_{X_2}(x_2)\quad ,\quad y_1,y_2>0\\ &=\frac{n_1}{n_2}y_2\cdot\frac{1}{2^{n_1/2}\,\Gamma\left(\frac{n_1}{2}\right)}\left(\frac{n_1}{n_2}y_1y_2\right)^{n_1/2-1}\cdot\exp-\left\{\frac{\frac{n_1}{n_2}y_1y_2}{2}\right\}\\ &\quad\cdot\frac{1}{2^{n_2/2}\,\Gamma\left(\frac{n_2}{2}\right)}\,y_2^{n_2/2-1}\cdot\exp-\left\{\frac{y_2}{2}\right\}\end{aligned}$$

Calculando ahora la f.d.p marginal de Y_1.

$$f_{Y_1}(y_1)=\int_0^{\infty}\frac{\left(\frac{n_1}{n_2}\right)}{2^{(n_1+n_2)/2}\,\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\left(\frac{n_1}{n_2}\cdot y_1 y_2\right)^{n_1/2-1} y_2^{\,n_2/2}\exp-\left\{\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)y_2\right\}dy_2$$

$$=\frac{\left(\frac{n_1}{n_2}\right)^{n_1/2} y_1^{\,n_1/2-1}}{\Gamma\left(\frac{n_1}{n_2}\right)\Gamma\left(\frac{n_2}{2}\right)}\cdot\int_0^{\infty}\frac{1}{2^{(n_1+n_2)/2}}\cdot y_2^{(n_1+n_2)/2}\,y_2^{-1}\exp-\left\{\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)y_2\right\}dy_2 \quad ..\ (*)$$

Sean $u=\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)y_2$, $du=\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)dy_2$

Multiplicando y dividiendo el integrando de la última expresión de $f_{Y_1}(y_1)$ por $\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)y_2$ se obtiene.

$$\left[\frac{1}{\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)}\right]\frac{1}{2^{(n_1+n_2)/2}}\int_0^{\infty}y_2^{(n_1+n_2)/2}\,y_2^{-1}\left[\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)\right]\exp-\left\{\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)y_2\right\}dy_2$$

$$=\left[\frac{1}{\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)}\right]\frac{1}{2^{(n_1+n_2)/2}}\int_0^{\infty}y_2^{(n_1+n_2)/2}\,y_2^{-1}\cdot u e^{-u}\,du$$

Sustituyendo ahora el valor de $\dfrac{2u}{\left(\frac{n_1}{n_2}y_1+1\right)}$ que ha sido despejado de la expresión de u

$$\left[\frac{1}{\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)}\right]\frac{1}{2^{(n_1+n_2)/2}}\int_0^{\infty}\left\{\frac{2u}{\frac{n_1}{n_2}y_1+1}\right\}^{(n_1+n_2)/2}\cdot\left\{\frac{\frac{n_1}{n_2}y_1+1}{2u}\right\}u e^{-u}\,du$$

$$=\left[\frac{1}{\frac{1}{2}\left(\frac{n_1}{n_2}y_1+1\right)}\right]\frac{1}{2^{(n_1+n_2)/2}}\cdot\frac{2^{(n_1+n_2)/2-1}}{\left[\frac{n_1}{n_2}y_1+1\right]^{(n_1+n_2)/2}}\cdot\left\{\frac{n_1}{n_2}y_1+1\right\}\cdot\underbrace{\int_0^{\infty}u^{(n_1+n_2)/2-1}e^{-u}\,du}_{\Gamma\left(\frac{n_1+n_2}{2}\right)}$$

$$=\left[\frac{2}{\frac{n_1}{n_2}y_1+1}\right]\cdot\frac{2^{(n_1+n_2)/2-1}}{\left[\frac{n_1}{n_2}y_1+1\right]^{(n_1+n_2)/2}}\cdot\left\{\frac{n_1}{n_2}y_1+1\right\}\cdot 2^{-(n_1+n_2)/2}\cdot\Gamma\left(\frac{n_1+n_2}{2}\right)$$

$$=\left[\frac{1}{\frac{n_1}{n_2}y_1+1}\right]^{(n_1+n_2)/2}\cdot\Gamma\left(\frac{n_1+n_2}{2}\right)$$

Sustituyendo este último resultado en (*)

$$f_{Y_1}(y_1)=\frac{\left(\frac{n_1}{n_2}\right)^{n_1/2}y^{\,n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\cdot\left[\frac{1}{\frac{n_1}{n_2}y_1+1}\right]^{(n_1+n_2)/2}\cdot\Gamma\left(\frac{n_1+n_2}{2}\right)$$

$$=\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)\left(\frac{n_1}{n_2}\right)^{n_1/2}\cdot y_1^{\,n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)\left[\frac{n_1}{n_2}y_1+1\right]^{(n_1+n_2)/2}}$$

que es la f d.p. de una v.a. $F(n_1, n_2)$.

◈

- *La distribución ji-cuadrada como distribución límite*

[3.48]

$$n_1[\mathrm{F}(n_1, n_2)] \xrightarrow{n_2 \to \infty} \text{Ji-cuadrada } (n_1)$$

"La transformación de una v.a. $F(n_1, n_2)$, X, definida por $Y = n_1 X$ se distribuye ji-cuadrada (n_1) cuando n_2 es suficientemente grande"

Observación. La demostración que aquí se presenta requiere considerar el supuesto adicional de que n_1 y n_2 sean enteros positivos *pares*. Esto es particularmente importante para n_1, ya que *dicho parámetro permanece en la expresión final; en cambio, es posible hacer tender n_2 a infinito* por los enteros positivos pares, haciéndolo "desaparecer" finalmente en la expresión final.

Demostración. Sea X una v.a. aleatoria con distribución F con parámetros (n_1, n_2). Defínase

$$Y = n_1 X \quad , \quad y > 0$$

Aunque lo que interesa es la *distribución límite* de Y, es necesario determinar primero la densidad de Y. Para ello, se utilizará la *Técnica de la transformación*, vía el Resultado1.6 con $c = n_1$ y $d = 0$.

$$f_Y(y) = \left|\frac{1}{n_1}\right| f_X\left(\frac{y}{n_1}\right) = \frac{1}{n_1} f_X\left(\frac{y}{n_1}\right) \quad , \quad \text{pues } n_1 \text{ es un entero positivo}$$

$$= \frac{1}{n_1}\left\{ \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)\cdot\left(\frac{n_1}{n_2}\right)^{n_1/2}\left(\frac{y}{n_1}\right)^{n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)\left[\frac{n_1}{n_2}\left(\frac{y}{n_1}\right)+1\right]^{(n_1+n_2)/2}}\right.$$

$$= \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\cdot\left[n_1^{-1+n_1/2-n_1/2+1}\right]\left(\frac{1}{n_2}\right)^{n_1/2}\cdot\frac{y^{n_1/2-1}}{\left[\frac{n_1}{n_2}\left(\frac{y}{n_1}\right)+1\right]^{(n_1+n_2)/2}}$$

$$= \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right)\Gamma\left(\frac{n_2}{2}\right)}\cdot\frac{\left(\frac{1}{n_2}\right)^{n_1/2} y^{n_1/2-1}}{\left[\frac{y}{n_2}+1\right]^{(n_1+n_2)/2}}$$

$$f_Y(y) = \frac{y^{n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right)} \left\{ \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_2}{2}\right)} \cdot \left(\frac{\frac{1}{n_2}}{\frac{y}{n_2}+1} \right)^{n_1/2} \cdot \left(\frac{1}{\frac{y}{n_2}+1} \right)^{n_2/2} \right\}$$

$$= \frac{y^{n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right)} \cdot \left\{ \frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_2}{2}\right)} \cdot \left(\frac{1}{y+n_2} \right)^{n_1/2} \cdot \left(\frac{1}{\frac{y}{n_2}+1} \right)^{n_2/2} \right\} \qquad (3.48.1)$$

Desarrollando el primer factor del producto entre llaves, haciendo uso de que, como n_1 y n_2 son enteros positivos *pares* entonces $\frac{n_1+n_2}{2}$ y $\frac{n_2}{2}$ también lo son.

$$\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_2}{2}\right)} = \frac{\left(\frac{n_1+n_2}{2}-1\right)!}{\left(\frac{n_2}{2}-1\right)!} = \frac{\left[\left(\frac{n_2}{2}-1\right)+\frac{n_1}{2}\right]!}{\left(\frac{n_2}{2}-1\right)!} = \frac{\prod_{i=1}^{n_1/2}\left[\left(\frac{n_2}{2}-1\right)+i\right]\left(\frac{n_2}{2}-1\right)!}{\left(\frac{n_2}{2}-1\right)!}$$

$$= \prod_{i=1}^{n_1/2}\left[\frac{n_2}{2}+(i-1)\right] = \prod_{i=1}^{n_1/2}\left[n_2\left(\frac{1}{2}+\frac{(i-1)}{n_2}\right)\right] = n_2^{n_1/2}\prod_{i=1}^{n_1/2}\left(\frac{1}{2}+\frac{(i-1)}{n_2}\right)$$

Sustituyendo este último resultado en la expresión (3.48.1):

$$\circ \quad f_Y(y) = \frac{y^{n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right)} \left\{ \left[n_2^{n_1/2} \prod_{i=1}^{n_1/2}\left(\frac{1}{2}+\frac{(i-1)}{n_2}\right) \right] \left(\frac{1}{y+n_2} \right)^{n_1/2} \cdot \left(\frac{1}{\frac{y}{n_2}+1} \right)^{n_2/2} \right\}$$

Calculando ahora $\lim_{n_2\to\infty} f_Y(y)$.

$$\lim_{n_2\to\infty} \frac{y^{n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right)} = \frac{y^{n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right)} \qquad \ldots (3.48.2)$$

$$\lim_{n_2\to\infty} n_2^{n_1/2}\prod_{i=1}^{n_1/2}\left(\frac{1}{2}+\frac{(i-1)}{n_2}\right) = \lim_{n_2\to\infty} n_2^{n_1/2} \cdot \prod_{i=1}^{n_1/2}\lim_{n_2\to\infty}\left(\frac{1}{2}+\frac{(i-1)}{n_2}\right)$$

$$= \lim_{n_2\to\infty} n_2^{n_1/2} \cdot \prod_{i=1}^{n_1/2}\left(\frac{1}{2}\right)$$

$$= \left(\frac{1}{2}\right)^{n_1/2} \cdot \lim_{n_2\to\infty} n_2^{n_1/2} \qquad \ldots (3.48.3)$$

$$\lim_{n_2\to\infty}\left(\frac{1}{1+y_2/n_2}\right)^{n_2/2} = \frac{1}{\lim\limits_{n_2\to\infty}\left(1+\frac{y}{n_2}\right)^{n_2/2}} = \frac{1}{\lim\limits_{n_2\to\infty}\left(1+\frac{y/2}{n_2/2}\right)^{n_2/2}}$$

$$= \frac{1}{\lim\limits_{\frac{n_2}{2}\to\infty}\left(1+\frac{y/2}{n_2/2}\right)^{n_2/2}} = \frac{1}{e^{y/2}} = e^{-y/2} \qquad (3.48.4)$$

Reuniendo los resultados (3.48.2) y (3.48.4):

$$\lim_{n_2\to\infty} f_Y(y) = \frac{y^{n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right)}\cdot\left[\left(\frac{1}{2}\right)^{n_1/2}\lim_{n_2\to\infty} n_2^{n_1/2}\right]\cdot\left[\left(\frac{1}{y+n_2}\right)^{n_1/2}\right] e^{-y/2}$$

$$= \frac{y^{n_1/2-1}}{\Gamma\left(\frac{n_1}{2}\right)}\left(\frac{1}{2}\right)^{n_1/2}\cdot e^{-y/2}\left[\lim_{n_2\to\infty}\left(\frac{n_2}{y+n_2}\right)^{n_1/2}\right]$$

Como

$$\lim_{n_2\to\infty}\left(\frac{n_2}{y+n_2}\right)^{n_1/2} = \lim_{n_2\to\infty}\left(\frac{1}{\frac{y}{n_2}+1}\right)^{n_1/2}$$

$$= \lim_{n_2\to\infty}\frac{1}{\underbrace{\left(\frac{y}{n_2}+1\right)\cdots\left(\frac{y}{n_2}+1\right)}_{\frac{n_1}{2}\text{ - veces}}} = 1$$

Por lo tanto:

$$\lim_{n_2\to\infty} f_Y(y) = \frac{y^{n_1/2-1}\cdot e^{-y/2}}{2^{n_1/2}\,\Gamma\left(\frac{n_1}{2}\right)}$$

que corresponde a la f.d.p Ji-cuadrada (n_1).

♦

- ***Otras relaciones con la distribución ji-cuadrada***

Véase también [3.29] en la sección *Relaciones con la distribución gamma*.

Véase también [3.35] en la sección *Relaciones con la distribución exponencial*.

♠

RELACIONES CON LA DISTRIBUCIÓN t DE STUDENT

• *Casos especiales de la distribución t de Student*

[3.49]

$$\boxed{t \text{ de Student } (1) \equiv \text{Cauchy } (0,1)}$$

"La distribución Cauchy estándar es un caso especial de la distribución $t_{(n)}$ cuando $n = 1$"

Demostración. Sea X una v.a. con distribución $t_{(n)}$ Basta sustituir el valor del parámetro $n = 1$ en la f.d.p. de X y considerar el hecho de que $\Gamma(½) = \sqrt{\pi}$

$$f_X(x) = f_X(x,n) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{(n\pi)^{1/2}\Gamma\left(\frac{n}{2}\right)} \frac{1}{\left[1+(x^2/n)\right]^{(n+1)/2}}, \qquad -\infty < x < \infty$$

$$\begin{aligned} f_X(x;1) &= \frac{\Gamma\left(\frac{1+1}{2}\right)}{(\pi)^{1/2}\Gamma\left(\frac{1}{2}\right)} \cdot \frac{1}{\left[1+x^2\right]^{(1+1)/2}}, \qquad -\infty < x < \infty \\ &= \frac{1}{\pi^{1/2}\pi^{1/2}} \cdot \frac{1}{\left[1+x^2\right]} \\ &= \frac{1}{\left[1+x^2\right]}, \qquad -\infty < x < \infty \end{aligned}$$

◆

En el Capítulo 2 se mencionó que la función generadora de momentos correspondiente a la distribución Cauchy no existe. Precisamente, como consecuencia de su vínculo, la función generadora de momentos para la distribución t de Student tampoco existe, aunque ésta sí tiene media y varianza finitas y es posible calcular $E[X^r]$

◆ ◆

- ***Relaciones de la distribución t con otras distribuciones mediante transformaciones***

[3.50]

$$[t \text{ de Student } (n)]^2 \sim F(1,n)$$

"El cuadrado de una v.a. $t_{(n)}$ se distribuye $F_{(1,n)}$"

Demostración. Sea X una v a. con distribución $t_{(n)}$. Defínase

$$Y = X^2$$

Utilizando la *Técnica de la transformación* mediante el *Resultado 1.1* se encontrará la f.d.p de Y.

$$
\begin{aligned}
f_Y(y) &= \frac{1}{2\sqrt{y}}\left[f_x(\sqrt{y}) + f_x(-\sqrt{y})\right], \quad y > 0 \\
&= \frac{1}{2\sqrt{y}}\left\{\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left[1+\frac{(\sqrt{y})^2}{n}\right]^{(n+1)/2}} + \frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left[1+\frac{(-\sqrt{y})^2}{n}\right]^{(n+1)/2}}\right\} \\
&= \frac{1}{2\sqrt{y}}\left\{\frac{2\,\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\,\Gamma\left(\frac{n}{2}\right)\left[1+\frac{y}{n}\right]^{(n+1)/2}}\right\} \\
&= y^{-\frac{1}{2}}\left\{\frac{\Gamma\left(\frac{n+1}{2}\right)}{n^{1/2}\sqrt{\pi}\,\Gamma\left(\frac{n}{2}\right)\left[1+\frac{y}{n}\right]^{(n+1)/2}}\right\}
\end{aligned}
$$

Recordando que $\Gamma(½) = \sqrt{\pi}$ se obtiene:

$$
\begin{aligned}
f_Y(y) &= y^{-1/2}\left\{\frac{\Gamma\left(\frac{n+1}{2}\right)}{n^{1/2}\,\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)\left[1+\frac{y}{n}\right]^{(n+1)/2}}\right\} \\
&= \frac{\Gamma\left(\frac{n+1}{2}\right)\left(\frac{1}{n}\right)^{1/2} y^{-1/2}}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n}{2}\right)\left[1+\frac{y}{n}\right]^{(n+1)/2}}, \quad y > 0
\end{aligned}
$$

que corresponde a la f.d.p. de una v.a. F con parámetros 1 y n.

◆

- ***Otras relaciones con la distribución t de Student***

Véase también [3.25] en *Relaciones con la distribución normal.*

RELACIONES CON LA DISTRIBUCIÓN CAUCHY

- ***Relaciones de la distribución Cauchy con ella misma***

[3.51]

$$\boxed{\text{a} + \alpha\cdot[\text{Cauchy }(0,1)] \sim \text{Cauchy }(\text{a},\alpha)}$$

"Una v.a. Cauchy estándar, X, puede llevarse a la forma general mediante la transformación definida por $Y = \text{a} + \alpha X$"

Demostración. Sea X una v.a. Cauchy $(0,1)$. Defínase

$$Y = \text{a} + \alpha X.$$

Para determinar la distribución de Y se recurrirá a la *Técnica de la transformación* mediante el *Resultado 1.6* con $c = \alpha$ y $d = \text{a}$.

$$\begin{aligned} f_Y(y) &= \frac{1}{\alpha}\cdot f_X\left(\frac{y-\text{a}}{\alpha}\right), \qquad (\text{ya que } \alpha > 0) \qquad -\infty < y < \infty \\ &= \frac{1}{\alpha}\cdot\frac{1}{\pi\left[1+\left(\dfrac{y-\text{a}}{\alpha}\right)^2\right]} \\ &= \frac{1}{\alpha\pi\left[1+\left(\dfrac{y-\text{a}}{\alpha}\right)^2\right]} \end{aligned}$$

La relación inversa de [3 51] se presenta a continuación

[3.52]

$$\text{Cauchy }(a,\alpha) \xrightarrow[\alpha=1]{a=0} \text{Cauchy }(0,1)$$

"A la distribución Cauchy con parámetros $a = 0$ ***y*** $\alpha = 1$
se le llama distribución Cauchy estándar"

Demostración Esta relación se verifica prácticamente por definición de la distribución Cauchy estándar; basta sustituir los valores $a = 0$ y $\alpha = 1$ en la f.d.p. Cauchy (a,α).

◆

Una generalización de la relación [3.51] se presenta en seguida y alude a cualquier transformación lineal de una v.a. Cauchy.

[3.53]

$$c + d\,[\text{Cauchy}\,(a,\alpha)] \sim \text{Cauchy}\,(ac + d\,.\,\alpha|c|)$$

***"Una transformación lineal,* c*X* + d*, de una v.a. Cauchy* (a,α), *X*,**
***también se distribuye Cauchy con parámetros* ac + d *y* α|c|"**

Demostración Sea X una v.a. Cauchy (a,α) Defínase

$$Y = c + dX.$$

Para determinar la distribución de Y se recurrirá a la *Técnica de la transformación* mediante el *Resultado 1 6.*

$$f_Y(y) = \frac{1}{|c|}\cdot f_X\left(\frac{y-d}{c}\right) \quad , \qquad -\infty < y < \infty$$

$$= \frac{1}{|c|}\cdot\frac{1}{\alpha\pi\left[1+\dfrac{\left(\dfrac{y-d}{c}-a\right)^2}{\alpha^2}\right]}$$

$$= \frac{1}{\alpha\pi\left[1+\dfrac{(y-d-ac)^2}{\alpha^2c^2}\right]}$$

$$= \frac{1}{[\alpha|c|]\pi\left[1+\dfrac{(y-[ac+d])^2}{[\alpha\ c\ ]^2}\right]}$$

◇

[3.54]

$$\frac{1}{\text{Cauchy}(\mathrm{a},\alpha)} \sim \text{Cauchy}\,(\mathrm{a}(\alpha^2+\mathrm{a}^2)^{-1}, \alpha(\alpha^2+\mathrm{a}^2)^{-1})$$

***"El inverso multiplicativo de una v.a. Cauchy* (a,α)**
***también se distribuye Cauchy* (a',α') *con* $a' = a(\alpha^2+a^2)^{-1}$ *y* $\alpha' = \alpha(\alpha^2+a^2)^{-1}$"**

Demostración. Sea X una v.a Cauchy (a,α). Defínase

$$Y = \frac{1}{X}.$$

Puesto que Y *es una transformación de* X, *a primera vista parece conveniente recurrir a la Técnica de la transformación* vía el *Resultado1.3*. Sin embargo, no es así. Por un lado, la aplicación del *Resultado1.3* conlleva a expresiones algebraicas complicadas de simplificar para llegar a una distribución Cauchy (esto como consecuencia de la complejidad de los parámetros de la v.a. Y). Por otro lado, el aplicarlo de manera mecánica conduce a pasar por alto cuál es el comportamiento de la *transformación* (recuerde que la función $\frac{1}{x}$ define una hipérbola de dos hojas): no puede decirse que *es monótona decreciente para todo el rango de* X, $-\infty < x < \infty$, *es* monótona decreciente en los intervalos $(-\infty,0)$ y $(0,\infty)$, pero está indefinida para $x = 0$. Entonces, el resultado obtenido sólo sería válido para todos los valores de y, exceptuando cuando $x = 0$.

Por lo anterior, se recurrió a un procedimiento alternativo para demostrar la relación [3.54]. En dicho procedimiento se hace uso de otros resultados: [3.15] y [3.18], *verificados en la sección de Relaciones con la distribución normal*; [3.51] y [3.53] verificados arriba; y [3.57] por verificar al final dentro de esta misma sección. Así, se irán hilando propiedades de transformaciones de dos v.a.i.i.d. normal estándar *cuyo cociente da lugar a una v.a. Cauchy estándar* que bien puede llevarse a la forma general Cauchy (a,α).

Se tiene que

$$X \sim \text{Cauchy } (\text{a}, \alpha)$$

Considere

$$U \sim \text{Normal } (0,1) \qquad \text{y} \qquad V \sim \text{Normal } (0,1)$$

y así defínase

$$X' = \frac{U}{V}.$$

Entonces

$$X' \sim \text{Cauchy } (0,1) \qquad \text{...de [3 57]}$$

Y así X puede expresarse en términos de X'.

$$X = \text{a} + \alpha X' = \text{a} + \alpha \frac{U}{V} \qquad \text{.. de [3 51].}$$

Ahora, observe que

$$\left[\text{a} + \alpha \frac{U}{V}\right]^{-1} = \alpha(\alpha^2 + \text{a}^2)^{-1} \cdot \frac{(\alpha U - \text{a}V)}{(\text{a}U + \alpha V)} + \text{a}(\alpha^2 + \text{a}^2)^{-1} \qquad (*)$$

$$= \alpha(\alpha^2 + \text{a}^2)^{-1} \cdot \frac{(\alpha U - \text{a}V)(\alpha^2 + \text{a}^2)^{-1/2}}{(\text{a}U + \alpha V)(\alpha^2 + \text{a}^2)^{-1/2}} + \text{a}(\alpha^2 + \text{a}^2)^{-1}$$

donde

$$\begin{aligned} (\alpha U - \text{a}V)(\alpha^2 + \text{a}^2)^{-1/2} &\sim \text{Normal}(0,1) \\ (\text{a}U + \alpha V)(\alpha^2 + \text{a}^2)^{-1/2} &\sim \text{Normal}(0,1) \end{aligned} \qquad (**)$$

Es decir $\left[\text{a} + \alpha \frac{U}{V}\right]^{-1} = \frac{1}{X}$ puede expresarse como una transformación lineal de una v.a Cauchy estándar. Considerando

$$c = \text{a}(\alpha^2 + \text{a}^2)^{-1} \qquad \text{y} \qquad d = \alpha(\alpha^2 + \text{a}^2)^{-1}$$

y aplicando [3 53]

$$\frac{1}{X} \sim \text{Cauchy}\,(\text{a}(\alpha^2 + \text{a}^2)^{-1}, \alpha(\alpha^2 + \text{a}^2)^{-1})$$

◆

(*) Igualdad encontrada en Jonhson y Kotz [1970]-1, p 160, quienes refieren que fue utilizada por Savage[1966]
(**) Más adelante se verificarán estos resultados paso a paso.

Verificación de (**)

Como $U \sim \text{Normal}\,(0,1) \quad \Rightarrow \quad \alpha U \sim \text{Normal}\,(0,\alpha^2)$ ... de [3.15]

Como $V \sim \text{Normal}\,(0,1) \quad \Rightarrow \quad aV \sim \text{Normal}\,(0,a^2)$ ... de [3 15]

$\Rightarrow \quad (\alpha U - aV) \sim \text{Normal}\,(0,\, \alpha^2 + a^2)$ ... de [3.18]

$\Rightarrow \quad (\alpha^2 + a^2)^{-\frac{1}{2}}\,(\alpha U - aV) \sim \text{Normal}\,(0,1)$ ... de [3.15]

ya que $[(\alpha^2 + a^2)^{-\frac{1}{2}}]^2\,(\alpha^2 + a^2) = (\alpha^2 + a^2)^{-1}\,(\alpha^2 + a^2) = 1$

Análogamente,

$U \sim \text{Normal}\,(0,1) \quad \Rightarrow \quad aU \sim \text{Normal}\,(0,a^2)$

$V \sim \text{Normal}\,(0,1) \quad \Rightarrow \quad \alpha V \sim \text{Normal}\,(0,\alpha^2)$

$\Rightarrow \quad (aU - \alpha V) \sim \text{Normal}\,(0,\, \alpha^2 + a^2)$

$\Rightarrow \quad (\alpha^2 + a^2)^{-\frac{1}{2}}\,(aU + \alpha V) \sim \text{Normal}\,(0,1)$

◆◆

[3.55]

$$\frac{1}{\text{Cauchy}(0,1)} \sim \text{Cauchy}\ (0,1)$$

"El inverso multiplicativo de una v.a. Cauchy estándar también se distribuye Cauchy estándar"

Demostración. La relación [3.55] es un caso particular de [3.54] ya demostrada Lo que es interesante rescatar es que en el caso de una v a. Cauchy estándar la complejidad de los parámetros a' y α' de $\frac{1}{X}$ se diluye ya que a' = $0(1^2 + 0^2)^{-1} = 0$ y α' = $1(1^2 + 0^2)^{-1} = 1$, de tal suerte que se obtiene también una v.a Cauchy estándar.

◇

La relación [3.55] también genera sentido si se piensa que la distribución Cauchy estándar puede verse como la distribución del cociente U/V, donde U y V son variables normal estándar independientes Como U y V tienen la misma distribución, entonces la distribución de U/V y V/U debe ser también la misma; de aquí que si X es Cauchy estándar entonces $1/X$ también lo es.

◇ ◇

[3.56]

$$\boxed{\sum_{1}^{n} \text{Cauchy}(a, \alpha) \sim \text{Cauchy}\,(na\,, n\alpha)}$$

"La suma de n v.a.i.i.d. Cauchy (a,α) también se distribuye Cauchy (na , nα)"

(La distribución Cauchy satisface la propiedad reproductiva)

Demostración Sean $X_1, X_2, \ldots, X_n$ v a.i.i.d Cauchy (a , α). Defínase $Y = \sum_{j=1}^{n} X_j$ Observe que como cada $-\infty < x_j < \infty$ entonces $-\infty < y < \infty$. Ahora, dada la inexistencia de la función generadora de momentos no es posible recurrir a ella para determinar la distribución de Y (de lo contrario dicha distribución se podría determinar aplicando el *Resultado 1.18*) En casos como éste, la *función característica* resulta de gran utilidad. Sólo a manera de ejemplo se introduce aquí para demostrar [3 56].

La función característica de una distribución Cauchy (a , α) es

$$E\left(e^{itX}\right) = \exp[ita - |t|\alpha]$$

Entonces

$$E\left[e^{it\sum_{j=1}^{n} X_j}\right] = E\left[\prod_{j=1}^{n} e^{itX_j}\right] = \prod_{j=1}^{n} E\left[e^{itX_j}\right]$$

$$= \prod_{j=1}^{n} \exp[ita - |t|\alpha] = \exp\left[\sum_{j=1}^{n} [ita - |t|\alpha]\right]$$

$$= \exp[it(na) - |t|(n\alpha)]$$

que corresponde a la función característica de una distribución Cauchy ($na, n\alpha$).

◆

[3.57]

$$\frac{\text{Normal}_1(0,1)}{\text{Normal}_2(0,1)} \sim \text{Cauchy}\ (0,1)$$

"El cociente de dos v.a.i. normal estándar se distribuye Cauchy estándar"

Demostración. Sean Z_1 y Z_2 v.a.i.i.d. normal (0,1). Defínase $Y = Z_1/Z_2$. Como Y es función bivariada $g(Z_1,Z_2)$ y ya que se conoce la distribución de la suma de dos v.a.i. normales estándar (véase resultado [3.16]) se aplicará la *técnica de las transformación* para funciones bivariadas para determinar la distribución de Y, vía el *Teorema 1.8*, considerando.

$$Y_1 = Z_1 + Z_2$$

y

$$Y_2 = Z_1/Z_2$$

Observe que $-\infty < y_1, y_2 < \infty$. Entonces, del *Teorema 1.8*:

$$f_{Y_1,Y_2}(y_1,y_2) = J\, f_{Z_1}(z_1)\cdot f_{Z_2}(z_2), \qquad -\infty < y_1, y_2 < \infty$$

Haciendo los cálculos:

$$\left.\begin{aligned} y_1 &= z_1 + z_2 \\ y_2 &= z_1/z_2 \end{aligned}\right\} \Rightarrow \quad z_1 = \frac{y_1 y_2}{1+y_2} \qquad z_2 = \frac{y_1}{1+y_2}$$

Así,

$$J = \begin{vmatrix} \dfrac{y_2}{1+y_2} & \dfrac{y_1}{(1+y_2)^2} \\ \dfrac{1}{1+y_2} & -\dfrac{y_1}{(1+y_2)^2} \end{vmatrix} = -\frac{y_1 y_2}{(1+y_2)^3} - \frac{y_1}{(1+y_2)^3} = -\frac{y_1}{(1+y_2)^2}$$

Sustituyendo en el *Teorema 1.8*

$$\begin{aligned} f_{Y_1,Y_2}(y_1,y_2) &= f_{Z_1}(z_1) f_{Z_2}(z_2)\ |J| \\ &= \left|-\frac{y_1}{(1+y_2)^2}\right| \cdot \frac{1}{\sqrt{2\pi}} \exp-\left\{\frac{(y_1y_2)^2}{2(1+y_2)^2}\right\} \cdot \frac{1}{\sqrt{2\pi}} \exp-\left\{\frac{y_1^2}{2(1+y_2)^2}\right\} \end{aligned}$$

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{|y_1|}{(1+y_2)^2}\cdot\frac{1}{2\pi}\exp-\left\{\frac{(y_1y_2)^2}{2(1+y_2)^2}+\frac{y_1^2}{2(1+y_2)^2}\right\}$$

$$= \frac{|y_1|}{(1+y_2)^2}\cdot\frac{1}{2\pi}\cdot\exp-\left\{\frac{y_1^2(y_2^2+1)}{2(1+y_2)^2}\right\}$$

Por lo tanto:

$$f_{Y_1,Y_2}(y_1,y_2) = \frac{|y_1|}{(1+y_2)^2}\cdot\frac{1}{2\pi}\exp-\left\{\frac{y_1^2(y_2^2+1)}{2(1+y_2)^2}\right\} \quad , \quad -\infty < y_1, y_2 < \infty$$

Calculando ahora la f.d.p. marginal de Y_2:

$$f_{Y_2}(y_2) = \int_{-\infty}^{\infty}\frac{1}{2\pi}\cdot\frac{|y_1|}{(1+y_2)^2}\exp-\left\{\frac{(1+y_2^2)y_1^2}{2(1+y_2)^2}\right\}dy_1$$

$$= \frac{1}{2\pi(1+y_2)^2}\int_{-\infty}^{\infty}|y_1|\exp-\left\{\frac{(1+y_2^2)y_1^2}{2(1+y_2)^2}\right\}dy_1$$

$$= \frac{1}{2\pi(1+y_2)^2}\int_{-\infty}^{0}(-y_1)\exp-\left\{\frac{(1+y_2^2)y_1^2}{2(1+y_2)^2}\right\}dy_1 + \frac{1}{2\pi(1+y_2)^2}\int_0^{\infty}y_1\exp-\left\{\cdots\right.$$

Sea

$$u = \frac{(1+y_2^2)y_1^2}{2(1+y_2)^2} \quad , \quad du = \frac{(1+y_2^2)}{(1+y_2)^2}\,y_1\,dy_1$$

Multiplicando y dividiendo la última expresión de $f_{Y_2}(y_2)$ por $\frac{(1+y_2)^2}{(1+y_2^2)}$ para completar el integrando como se requiere para efectuar el cambio de variable $u\,du$.

$$f_{Y_2}(y_2) = \frac{1}{2\pi(1+y_2)^2}\cdot\frac{(1+y_2)^2}{(1+y_2^2)}\int_{-\infty}^{0}-\frac{(1+y_2^2)}{(1+y_2)^2}\,y_1\exp-\left\{\frac{(1+y_2^2)y_1^2}{2(1+y_2)^2}\right\}dy_1$$

$$+\frac{1}{2\pi(1+y_2)^2}\cdot\frac{(1+y_2)^2}{(1+y_2^2)}\int_0^{\infty}\frac{(1+y_2^2)}{(1+y_2)^2}\,y_1\cdot\exp-\left\{\frac{(1+y_2^2)y_1^2}{2(1+y_2)^2}\right\}dy_1$$

$$= -\frac{1}{2\pi(1+y_2^2)}\int_{-\infty}^{0}e^{-u}\,du + \frac{1}{2\pi(1+y_2^2)}\int_0^{\infty}e^{-u}\,du$$

$$= 2\left[\frac{1}{2\pi(1+y_2^2)}\int_{-\infty}^{\infty}e^{-u}\,du\right] = \frac{1}{\pi(1+y_2^2)}$$

que corresponde a la f.d.p. Cauchy estándar.

◆

- ***Otras relaciones con la distribución Cauchy***

Véase [3 49] en *Relaciones con la distribución t de Student.*

RELACIONES CON LA DISTRIBUCIÓN BETA

- ***Casos especiales de la distribución beta***

[3.58]

$$\text{Beta}(½,½) \equiv \text{Arcoseno}$$

"La distribución arcoseno es un caso especial de la distribución beta (β,γ) cuando $\beta = \gamma = ½$"

Demostración Sea X una v.a. con distribución beta (½,½). Basta sustituir los valores de los parámetros $\beta = ½$ y $\gamma = ½$ en la f.d.p. beta y considerar el hecho de que $\Gamma(½) = \sqrt{\pi}$

$$f_Y(x) = f_Y(x,\beta,\gamma) = \frac{\Gamma(\beta+\gamma)}{\Gamma(\beta)\Gamma(\gamma)} \cdot x^{\beta-1}(1-x)^{\gamma-1} \quad \text{si} \quad 0 < x < 1$$

$$\begin{aligned} f_X(x) = f_X(x;\tfrac{1}{2},\tfrac{1}{2}) &= \frac{\Gamma(\frac{1}{2}+\frac{1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})} \cdot x^{\frac{1}{2}-1}(1-x)^{\frac{1}{2}-1} \quad \text{si} \quad 0 < x < 1 \\ &= \frac{\Gamma(1)}{\Gamma(\frac{1}{2})\Gamma(\frac{1}{2})} \cdot x^{-1/2}(1-x)^{-1/2} \\ &= \frac{1}{\sqrt{\pi}\sqrt{\pi}} \cdot x^{-1/2}(1-x)^{-1/2} \\ &= \frac{1}{\pi[x(1-x)]^{1/2}} \end{aligned}$$

que corresponde a la f.d.p arcoseno.

◊

En la *Figura 2.9* presentada en el Capítulo 2 se puede apreciar la gráfica de la densidad arcoseno.

[3.59]

$$\text{Beta }(1,1) \equiv \text{Uniforme }(0,1)$$

"La distribución uniforme estándar es un caso especial de la distribución beta (β,γ) cuando $\beta = \gamma = 1$"

Demostración. Sea X una v.a. con distribución beta $(1,1)$. Nuevamente, basta sustituir los valores de los parámetros $\beta = 1$ y $\gamma = 1$ en la f.d.p beta y considerar el hecho de que $\Gamma(1) = 1$.

$$\begin{aligned} f_X(x;1,1) &= \frac{\Gamma(1+1)}{\Gamma(1)\Gamma(1)} \cdot x^{1-1}(1-x)^{1-1} \quad \text{si} \qquad 0 < x < 1 \\ &= \frac{1!}{0!0!} \cdot x^0 (1-x)^0 \\ &= 1 \end{aligned}$$

que corresponde a la f.d.p. uniforme estándar.

◆

También en la *Figura* 2.9 que se presentó en el Capítulo 2 se puede apreciar la gráfica de la densidad *uniforme* estándar.

- *Relaciones de la distribución beta con otras distribuciones mediante transformaciones*

[3.60]

$$\frac{\text{Gamma}(\alpha_1,\beta)}{\text{Gamma}(\alpha_1,\beta)+\text{Gamma}(\alpha_2,\beta)} \sim \text{Beta}\,(\alpha_1,\alpha_2)$$

"El cociente de una v.a. gamma (α_1,β) entre la suma de dos v.a.i. gamma, (α_1,β) y (α_2,β), se distribuye beta (α_1,α_2)"

Demostración. Sean X_1 y X_2 v.a.'s gamma (α_i,β), $i = 1, 2$. Defínase

$$Y = \frac{X_1}{X_1 + X_2}$$

Como Y es una función bivariada $g(X_1,X_2)$ puede aplicarse el método de la transformación para determinar su distribución. Para ello, es necesario elegir otra función conveniente para obtener una f.d.p. conjunta a partir de la cual se extraerá la densidad marginal de Y. Como ya se conoce la distribución de la suma de dos v.a.i gamma (α,β), (véase [3.26*bis*]) defínase

$$Y_1 = \frac{X_1}{X_1 + X_2}$$

y

$$Y_2 = X_1 + X_2.$$

En cuanto a los recorridos de Y_1 y Y_2, observe que:

$$x_1, x_2 > 0$$

$$\Rightarrow \qquad y_2 = x_1 + x_2 > 0$$

$$\Rightarrow \qquad 0 < x_1 < x_1 + x_2$$

$$\Rightarrow \qquad 0 < x_1 < x_1 + x_2$$

$$\Rightarrow \qquad 0 < \frac{x_1}{x_1 + x_2} = y_1 < 1$$

Calculando la funciones inversas se obtiene que:

$$\left.\begin{array}{l} y_1 = x_1/(x_1+x_2) \\ y_2 = x_1+x_2 \end{array}\right\} \Rightarrow \begin{array}{l} x_1 = y_1 y_2 = g_1^{-1}(y_1,y_2) \\ x_2 = y_2(1-y_1) = g_2^{-1}(y_1,y_2) \end{array}$$

De aquí que:

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} y_2 & y_1 \\ -y_2 & (1-y_1) \end{vmatrix} = y_2(1-y_1) + y_1 y_2 = y_2 - y_2 y_1 + y_1 y_2 = y_2$$

$$|J| = |y_2| = y_2 \quad , \quad \text{pues} \quad y_2 > 0$$

Así $\quad f_{Y_1,Y_2}(y_1,y_2) = |J| \cdot f_{X_1}(x_1) f_{X_2}(x_2) \quad , \quad 0 < y_1 < 1 \quad , \quad y_2 > 0$

$$= y_2 \cdot \frac{1}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} (y_1 y_2)^{\alpha_1-1} \left[y_2(1-y_1) \right]^{\alpha_2-1} \cdot e^{-y_2/\beta}$$

$$= \frac{1}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} (y_2)^{1+(\alpha_1-1)+(\alpha_2-1)} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1} e^{-y_2/\beta}$$

$$= \frac{1}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1)\Gamma(\alpha_2)} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1} \cdot y_2^{\alpha_1+\alpha_2-1} e^{-y_2/\beta}$$

$$= \left[\frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \cdot y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1} \right] \cdot \left[\frac{1}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1+\alpha_2)} y_2^{\alpha_1+\alpha_2-1} e^{-y_2/\beta} \right]$$

Considerando que

$$B(\alpha_1,\alpha_2) = \frac{\Gamma(\alpha_1)\Gamma(\alpha_2)}{\Gamma(\alpha_1+\alpha_2)} \Leftrightarrow \frac{\Gamma(\alpha_1+\alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} = \frac{1}{B(\alpha_1,\alpha_2)}$$

es posible separar la densidad de Y_2, que es gamma ($\alpha_1 + \alpha_2, \beta$), en la última expresión de la densidad conjunta $f_{Y_1,Y_2}(y_1,y_2)$:

$$f_{Y_1,Y_2}(y_1,y_2) = \left[\frac{1}{B(\alpha_1,\alpha_2)} y_1^{\alpha_1-1}(1-y_1)^{\alpha_2-1} \right] \cdot \left[\frac{1}{\beta^{\alpha_1+\alpha_2}\Gamma(\alpha_1+\alpha_2)} y_2^{\alpha_1+\alpha_2-1} e^{-y_2/\beta} \right]$$

De este producto no sólo se observa que tiene Y_1 distribución beta (α_1,α_2), sino también que ¡Y_1 y Y_2 son independientes!

♦

- ***Otras relaciones con la distribución beta***

Véase [3.24] en *Relaciones con la distribución normal.*

♤

RELACIONES CON LA DISTRIBUCIÓN UNIFORME

- ***Relaciones de la distribución uniforme con ella misma***

[3.61]

$$\boxed{a + (b - a)\cdot[\text{Uniforme }(0,1)] \sim \text{Uniforme }(a,b)}$$

"Una v.a. uniforme estándar, X, puede llevarse a la forma general mediante la transformación definida por $Y = a + (b - a)X$"

Demostración. Sea X una v.a uniforme $(0,1)$. Defínase

$$Y = a + (b - a)X.$$

Para determinar la distribución de Y se recurrirá a la *Técnica de la transformación* mediante el *Resultado 1.6* con $c = b - a$ y $d = a$, donde $c > 0$ ya que $b > a$

Respecto al recorrido de $y = a + (b - a)\,x$, observe que $a < y < b$ ya que:

$$\begin{aligned}
& 0 < x < 1 \\
\Rightarrow\quad & 0 < (b - a)\,x < (b - a) \\
\Rightarrow\quad & a < (b - a)\,x + a < (b - a) + a \\
\Rightarrow\quad & a < y < b.
\end{aligned}$$

Entonces.

- $$f_Y(y) = \frac{1}{b-a} f_X\left(\frac{y-a}{b-a}\right) \quad , \quad a < y < b$$
-
- $$= \frac{1}{b-a} \cdot 1 \quad , \text{ ya que } \quad 0 < \frac{y-a}{b-a} < 1$$

que corresponde a la densidad uniforme (a,b).

♦

La relación inversa de [3.61] se presenta a continuación.

[3.62]

$$\text{Uniforme } (a,b) \xrightarrow[b=1]{a=0} \text{Uniforme } (0,1)$$

"A la distribución uniforme con parámetros $a = 0$ y $b = 1$ se le llama distribución uniforme estándar"

Demostración. Esta relación se verifica prácticamente por definición de la distribución uniforme estándar; basta sustituir los valores $a = 0$ y $b = 1$ en la f.d.p. uniforme $(0,1)$.

♦

- ***Relaciones de la distribución uniforme con otras distribuciones mediante transformaciones***

[3.63]

$$\{\text{Uniforme}_1\,(0,1) - \text{Uniforme}_2\,(0,1)\} \sim \text{Triangular}$$

"La diferencia de dos v.a.i. uniforme estándar se distribuye triangular"

Demostración Sean X_1 y X_2 v.a.i i.d uniforme $(0,1)$. Defínase

$$Y = X_1 - X_2.$$

Observe que como $0 < x_1, x_2 < 1$ entonces $-1 < y = x_1 - x_2 < 1$.

Aunque Y es una transformación bivariada, en lugar de recurrir a la *Técnica de la transformación*, via el *Teorema1.8*, para determinar la f d p de Y se acudirá a la *Técnica de la función de distribución* mediante el *Resultado1.11*. Con la aplicación de la *Técnica de la transformación* se complican los cálculos pese a que la densidad original, uniforme $(0,1)$, parece sencilla

$$\begin{aligned}
f_Y(y) &= \int_{-\infty}^{\infty} f_{X_2}(x_1 - y) f_{X_1}(x_1)\, dx_1 \\
&= \int_{-\infty}^{\infty} I_{(0,1)}(x_1 - y) I_{(0,1)}(x_1)\, dx_1 \\
&\overset{(*)}{=} \int_{-\infty}^{\infty} \left\{ I_{(-1,0)}(y)\, I_{(0,y+1)}(x_1) + I_{[0,1)}(y) I_{[y,1)}(x_1) \right\} dx_1 \\
&= I_{(-1,0)}(y) \int_0^{y+1} dx_1 + I_{[0,1)}(y) \int_y^1 dx_1 \\
&= (1 + y) I_{(-1,0)}(y) + (1 - y) I_{[0,1)}(y)
\end{aligned}$$

La descomposición de las funciones indicadoras desarrollada de la segunda a la tercera igualdad (paso señalado con $(*)$) resulta de la visualización geométrica del conjunto

$$A = \{0 < x_1 < 1\} \times \{0 < x_1 - y < 1\}$$

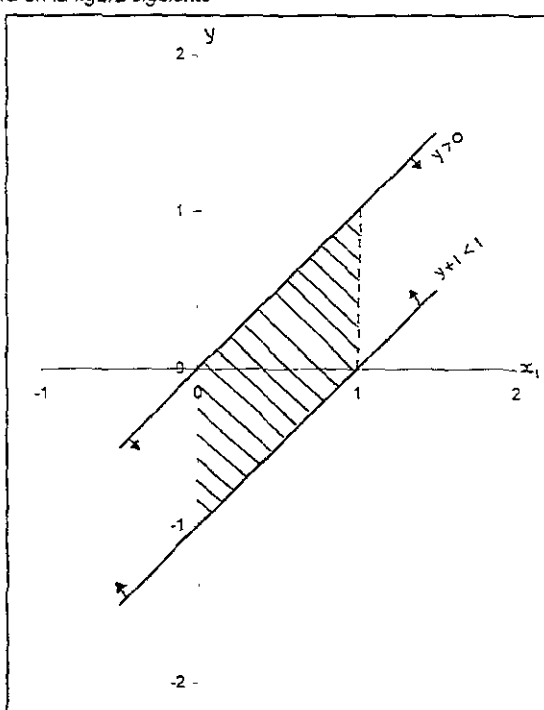
Como ya se ha observado $\quad x_1, x_2 \in (0,1) \;\Rightarrow\; -1 < x_1 - x_2 < 1$

Ahora $\quad y = x_1 - x_2 \;\Rightarrow\; x_2 = x_1 - y \;\Rightarrow\; 0 < x_1 - y < 1$

A es un subconjunto de R^2 en el sistema formado por los ejes coordenados X_1 y Y. El rango de las abscisas está dado de manera directa (valores entre 0 y 1). Ahora, para determinar el rango de las ordenadas basta observar qué es lo que pasa en las fronteras, es decir, en los límites inferior y superior del conjunto $\{0 < x_1 - y < 1\}$:

$$x_1 - y = 0 \;\Rightarrow\; y = x_1$$
$$x_1 - y = 1 \;\Rightarrow\; y = x_1 - 1$$

El conjunto A se ilustra en la figura siguiente



◆

- ***Otras relaciones con la distribución uniforme***

Véase también [3.42] en *Relaciones con la distribución exponencial*

Véase también [3.59] en *Relaciones con la distribución beta.*

♤

OTRAS RELACIONES ENTRE DISTRIBUCIONES UNIVARIADAS CONTINUAS

[3.64]

$$\boxed{\log[\text{Lognormal } (\mu,\sigma^2)] \sim \text{Normal } (\mu,\sigma^2)}$$

"El logaritmo natural de una v.a lognormal (μ,σ^2) se distribuye normal (μ,σ^2)"

Demostración Sea X una v.a. lognormal (μ,σ^2). Defínase la transformación

$$Y = \log X$$

Observe que la función logaritmo tiene sentido puesto que $x > 0$. Así, $-\infty < y < \infty$. Aplicando la *Técnica de la Transformación* mediante el *Resultado 1.8*, con $c = 1$, para determinar la distribución de Y

$$\begin{aligned} f_Y(y) &= \left|e^y\right| \cdot f_X(e^y) \qquad , \qquad -\infty < y < \infty \\ &= e^y \cdot \frac{1}{e^y\sqrt{2\pi}\sigma} \exp-\left\{\frac{[\log(e^y)-\mu]^2}{2\sigma^2}\right\} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \exp-\left\{\frac{(y-\mu)^2}{2\sigma^2}\right\} \end{aligned}$$

que corresponde a una densidad normal con media μ y varianza σ^2

♦

Como se expuso en el Capítulo 2, lo que caracteriza a una v.a. lognormal, y de ahí su nombre, es el hecho de que su logaritmo natural se distribuye normal. Aún más, *la relación* [3.64] *tiene su inverso* como se enuncia a continuación.

[3.65]

$$\exp[\text{Normal } (\mu,\sigma^2)] \sim \text{Lognormal } (\mu,\sigma^2)$$

"La exponencial de una v.a normal (μ,σ^2) se distribuye lognormal (μ,σ^2)"

Demostración. Sea X una v.a. normal (μ,σ^2). Defínase la transformación

$$Y = e^X$$

Observe que $y > 0$. Aplicando la *Técnica de la Transformación* mediante el *Resultado 1.7* para determinar la distribución de Y:

$$\begin{aligned} f_Y(y) &= \left|\frac{1}{y}\right| \cdot f_X(\log y) \qquad , \qquad y > 0 \\ &= \frac{1}{y} \cdot \frac{1}{\sqrt{2\pi}\sigma} \exp - \left\{\frac{[\log y - \mu]^2}{2\sigma^2}\right\} \\ &= \frac{1}{y\sqrt{2\pi}\sigma} \exp - \left\{\frac{(\log y - \mu)^2}{2\sigma^2}\right\} \end{aligned}$$

que corresponde a una densidad lognormal (μ,σ^2).

◆

Las relaciones [3.64] y [3.65] permitirán demostrar fácilmente la relación siguiente.

[3.66]

$$\prod_{1}^{n}[\text{Lognormal}\,(\mu,\sigma^2)] \sim \text{Lognormal}\,(n\mu,n\sigma^2)$$

"El producto de n v.a.i.i.d. lognormal (μ,σ^2) también se distribuye lognormal $(n\mu, n\sigma^2)$"

(Observe la analogía con la propiedad reproductiva)

Demostración. Sean $X_1, X_2, \ldots, X_n$ v.a.i.i.d lognormal (μ,σ^2). Defínase

$$Y = \prod_{i=1}^{n} X_i \,.$$

Observe que $y > 0$. La relación se demostrará haciendo uso de los resultados [3.16] (véase *Relaciones con la distribución normal*), [3.64] y [3.65] (recién demostrados) y de las propiedades siguientes:

- $\log Y = \log (X_1 \cdot X_2 \cdot \ldots \cdot X_n) = \sum_{i=1}^{n} \log X_i$
- $e^{\log Y} = Y$

Ahora,

	$X_i \sim \text{lognormal}\,(\mu,\sigma^2)$,	para $i = 1, 2, \ldots, n$	por hipótesis
$\Rightarrow$	$\log X_i \sim \text{normal}\,(\mu,\sigma^2)$,	para $i = 1, 2, \ldots, n$	aplicando [3.64]
$\Rightarrow$	$\sum_{i=1}^{n} \log X_i \sim \text{normal}\,(n\mu,n\sigma^2)$		aplicando [3.16]

Como $\sum_{i=1}^{n} \log X_i = \log Y$ entonces $\log Y \sim \text{normal}\,(n\mu,n\sigma^2)$.

$\Rightarrow$	$e^{\log Y} \sim \text{lognormal}\,(n\mu,n\sigma^2)$	aplicando [3.65]

Como $e^{\log Y} = Y$ entonces $Y \sim \text{lognormal}\,(n\mu,n\sigma^2)$.

◆

La relación [3.66] puede generalizarse al caso en que los *parámetros* de la distribución normal de las X_i no fueran los valores *comunes* μ y σ^2, sino que $X_1, X_2, \ldots, X_n$ fueran v.a.i.i.d. lognormal (μ_i, σ_i^2). Para demostrarlo se podría seguir el razonamiento anterior, haciendo uso de [3.16*bis*], en lugar de [3.16]

◆◆

Las dos relaciones siguientes son resultados 'sueltos" deducidos de lo expuesto por Leemis en su esquema.

[3.67]

$$\boxed{\text{Weibull }(\alpha,2) \equiv \text{Rayleigh }(\alpha)}$$

"La distribución Rayleigh (α) es un caso especial de la distribución Weibull (α, β) cuando $\beta = 2$"

Demostración. Sea X una v.a. Weibull (α, 2). La relación se prueba sin problema sustituyendo el valor $\beta = 2$ en la expresión de la f.d.p. Weibull $f_X(x) = f_X(x;\alpha,\beta) = \frac{1}{\alpha}\cdot\beta x^{\beta-1}\cdot\exp-\left[\frac{1}{\alpha}x^{\beta}\right]$ si $x>0$

$$\begin{aligned} f_X(x;\alpha,2) &= \frac{1}{\alpha}\cdot 2x^{2-1}\cdot\exp-\left[\frac{1}{\alpha}x^2\right] \qquad \text{si} \qquad x>0 \\ &= \frac{2}{\alpha}\cdot x e^{-x^2/\alpha} \end{aligned}$$

que corresponde a la f.d.p. Rayleigh (α).

◆

Como la distribución exponencial es también un caso especial de la distribución Weibull, con la relación [3.67] queda "cerrado" el vínculo Weibull-exponencial-Rayleigh conformado también por las relaciones [3.36] a [3.39], expuestas en la sección de *Relaciones con la distribución exponencial*.

◆◆

[3.68]

$$\frac{1}{F(n_1,n_2)} \sim F\,(n_1,n_2)$$

"El inverso multiplicativo de una v.a. F (n_1,n_2) también se distribuye F (n_2,n_1)"

Demostración. Sea X una v.a. con distribución $F\,(n_1,n_2)$ Defínase

$$Y = \frac{1}{X}.$$

Observe que la transformación está bien definida puesto que $x > 0$ Así, $y > 0$ Además, Y es una transformación monótona decreciente. Entonces puede aplicarse la *Técnica de la Transformación* mediante el *Resultado 1.3* para determinar la distribución de Y. Puesto que el manejo de una densidad F es laborioso no se presenta el desarrollo de la aplicación de la técnica. En su lugar, se acude a un procedimiento alternativo para demostrar la relación haciendo uso de un resultado verificado previamente.

Del resultado [3.47] se sabe que X *puede verse como el cociente de dos v.a.i. con distribución* ji-cuadrada, digamos X_1 y X_2, divididas entre sus respectivos grados de libertad n_1 y n_2, es decir, $X = (X_1/n_1)/(X_2/n_2)$ En consecuencia, el cociente $(X_2/n_2)/(X_1/n_1) = 1/X$ tiene distribución F (n_2,n_1)

◈

Con la relación [3.68] se cierra el conjunto de relaciones incluidas en el *Diagrama final* anexo, que es una modificación del diagrama original de Leemis. Al margen de ello, a continuación se presentan dos relaciones adicionales que se encontraron durante el desarrollo de este trabajo, pero cuya verificación requiere de algunas consideraciones adicionales.

- ***Otras relaciones propuestas***

[A.1]

"La transformación de una v.a., X, con distribución $F(n_1, n_2)$ definida por $Y = \dfrac{n_1 X_1 / n_2}{1 + [n_1 X_2 / n_2]}$ se distribuye beta $\left(\frac{n_1}{2}, \frac{n_2}{2}\right)$" [12]

Demostración. Esta demostración está pendiente. Sea X una v.a. con distribución $F(n_1,n_2)$. Defínase $Y = \dfrac{n_1 X_1 / n_2}{1 + [n_1 X_2 / n_2]} = \dfrac{a}{1+a}$. Recurriendo la Técnica de la transformación (previa aplicación de la Técnica de la función de distribución) se obtiene que:

$$f_Y(y) = \frac{n_2}{n_1}\left(\frac{1}{1-y} + \frac{y}{(1-y)^2}\right) \cdot f_X\left(\frac{n_2}{n_1} \cdot \frac{y}{1-y}\right), \qquad 0 < y < 1$$

donde el recorrido de Y está entre 0 y 1 porque $0 < a < 1 + a \quad \Rightarrow \quad 0 < \dfrac{a}{1+a} < 1$. Sustituyendo la f.d.p. de X evaluada en $[n_2 y / n_1 (1 - y)]$ y simplificando se obtiene

$$f_Y(y) = \frac{\Gamma\left(\frac{n_1+n_2}{2}\right) y^{\frac{n_1}{2}-1}}{\Gamma\left(\frac{n_1}{2}\right)\cdot\Gamma\left(\frac{n_2}{2}\right)} \cdot \frac{1}{(1-y)^{\frac{n_1}{2}-1}\left[\dfrac{y}{1-y}+1\right]^{\frac{n_1+n_2}{2}}}$$

de cuyo segundo factor no fue posible recuperar el término $(1 - y)^{(n_2/2)-1}$

◇

[A.2]

"La mínima estadística de orden de una distribución Weibull (α,β,γ) es también Weibull (α', β, γ), donde $\alpha' = \alpha n^{-1/\beta}$" [13]

Demostración. Esta relación es válida si se define a la densidad Weibull como $f_X(x) = (\beta/\alpha)\cdot[(x-\gamma)/\alpha]^{\beta-1}\cdot\exp-[(x-\gamma)/\alpha]^{\beta}$, $\gamma < x$, que difiere de la expresión [2.19] aun cuando $\gamma = 0$ (véase sección sobre La distribución Weibull en el Capítulo 2). Sin embargo, es este un resultado interesante en tanto que la distribución exponencial cumple una relación análoga a [A.2] (véase [3.31]) y es un caso especial de la distribución Weibull (véase [3.32]); bajo esta nueva expresión de la densidad Weibull esto se cumple cuando $\gamma = 0$ y $\beta = 1$.

◆

---

[12] Presentada por Casella y Berger [1990], p. 228.
[13] Presentada por Johnson y Kotz [1970]-1, p. 254.

Consideraciones finales y conclusiones

Aportaciones de este trabajo

El esquema que presenta Leemis en su artículo publicado en 1986 ha sido el eje para el desarrollo de este trabajo. Como lo explicó el propio Leemis, el propósito del artículo era presentar una figura que ilustrara algunas de las relaciones entre distribuciones univariadas comunes que generalmente se discuten por separado en los textos introductorios de probabilidad, lo cual dificulta a los estudiantes comprender las relaciones entres esas distribuciones. Este trabajo se adhiere al *objetivo integrador* de Leemis, al recoger una buena cantidad de información acerca de diferentes distribuciones de probabilidad en el Capítulo 2 e insistir en las relaciones que guardan unas con otras desde la última parte de la presentación de cada distribución, para después ahondar en ellas en el Capítulo 3

En este trabajo se incluyen las gráficas de todas y cada una de las densidades que participan en el diagrama de Leemis, para algunos parámetros, varias de ellas fueron encontradas en la literatura, mas elaborar aquellas que hacían falta enriqueció el conocimiento acerca del comportamiento de las distribuciones Además, es de llamar la atención que Leemis incluyó algunas distribuciones poco comunes, como la Weibull discreta, la arcoseno y la triangular. Se espera que este referente gráfico en su conjunto sea para el lector un medio de sensibilización con las distribuciones, pues siempre ver dibujos es más grato que tratar con "frías" formas algebraicas (pensando en las densidades).

Centrando ahora la atención en las relaciones entre las distribuciones, el *Diagrama Final* con el que se cierra esta sección representa una conclusión global de este trabajo. Este diagrama es una adaptación del diagrama de Leemis. En él están señaladas cada una de las relaciones con un número entre paréntesis, según el orden seguido en el Capítulo 3 (así, por ejemplo, el número (1) corresponde a la relación [3.1]). Asimismo, incluye las 12 relaciones adicionales encontradas a lo largo del desarrollo del trabajo, dando un total de 67 relaciones en el diagrama.

Es así que, por lo que toca a las relaciones entre las distribuciones, este trabajo contribuye de tres formas·

1. Presentando ejemplos de aplicación del Teorema del Límite Central, así como de tres técnicas para determinar las distribuciones de funciones de una o más variables aleatorias.[1] Así al verificar todas y cada una de la relaciones se adquiere al mismo tiempo soltura en el manejo de esos instrumentos, cuyo sustento teórico suele revisarse en los cursos de probabilidad y estadística y que de manera breve fue abordado en el Capítulo 1.

2. Determinando *explícitamente* los parámetros de todas y cada una de las distribuciones resultantes. Esto es importante sobre todo para aquellas relaciones en las que una distribución se vincula consigo misma, pues si bien en el esquema de Leemis se señala(n) con precisión el (los) parámetro(s) de la distribución "origen", en todos los casos el (los) parámetros de la distribución "destino" (distribución resultante) cambian.

   Por ejemplo, para aquellas distribuciones que satisfacen la llamada propiedad reproductiva, la suma de variables aleatorias independientes con distribución D se distribuye también D teniendo como parámetro(s) la suma de alguno(s) de los parámetros de las variables aleatorias originales (relaciones (3), (6), (11), (16), (17), (26), (44) y (56), que, salvo la última, tienen una versión "bis", como puede verse en el Capítulo 3). Cabe resaltar que fue particularmente interesante el hallazgo de los parámetros resultantes de las relaciones (12), (31) y (54)

3. Incorporando relaciones adicionales a las presentadas por Leemis, con lo cual se cuenta ahora con un esquema todavía más completo. Las relaciones agregadas fueron: (3), (11), (15), (17), (18), (26), (30), (39), (43), (53), (67) y (68). Salvo la (43) y la (67) que se encontraron en la literatura revisada, las demás surgieron de manera intuitiva una vez que se tuvo cierta familiaridad con las distintas distribuciones. Hay que aclarar que para (30) se entiende que ésta pudo no haber sido incluida por Leemis porque queda contemplada implícitamente a través de

[1] Cabe resaltar que en el caso de dos o mas variables aleatorias la independencia entre ellas fue una condicion sumamente valiosa.

(46) y de (29). En ese sentido, con (30) no se está agregando información nueva, pero se optó por incluirla a manera de referencia rápida De hecho, es muy posible que varias de las relaciones entre dos distribuciones no incluidas actualmente en el diagrama se puedan obtener a través de dos (o más) pasos con las relaciones que sí se registran

Bajo la visión integradora de las distribuciones que se propuso Leemis, para verificar varias de las relaciones mostradas se hizo uso a su vez de otras relaciones En particular, las relaciones (15), (17) y (18) relativas a la distribución normal y (29) relativa al vínculo gamma – ji-cuadrada representaron resultados de gran utilidad Precisamente, el que *una variable aleatoria ji-cuadrada sea al mismo tiempo gamma* es hecho que se pasa por alto con frecuencia y fue uno de los aspectos en los que se insistió recurrentemente en este trabajo.

Aclaraciones sobre otras modificaciones hechas al diagrama de Leemis. Es necesario puntualizar tres de las modificaciones hechas al diagrama de Leemis y vertidas en el *Diagrama Final* (i) los parámetros de la distribución lognormal son μ y σ^2, en lugar de α y β, con lo cual resulta más inmediata la verificación del vínculo de ésta con la distribución normal (que tiene también parámetros μ y σ^2); (ii) la línea continua que une a la distribución F con la ji-cuadrada (relación (48)) fue cambiada por una línea punteada, ya que esa relación se verifica cuando n es suficientemente grande ($n \to \infty$); y, (iii) al adoptar para la densidad gamma la expresión [2.*14*], en lugar de la que presenta Leemis (en la que se intercambian las posiciones de α y β) se modificaron las condiciones que avalan los vínculos (28) y (29), de tal manera que en lugar de pedir $\beta = 1$ (en (28)) o $\alpha = 2$ y $\beta = n/2$ (en (29)) se pide que $\alpha = 1$ y que $\alpha = n/2$ y $\beta = 2$, respectivamente Esto último tiene un costo sobre (28), ya que en lugar de que la distribución resultante sea una exponencial con parámetro α como lo indica el diagrama, se obtiene una exponencial con parámetro β. Sin embargo, se prefirió esto a cambio de homogeneizar este trabajo con la literatura consultada, pues en toda ella se define a la distribución gamma a través de la expresión [2.*14*].

Por ende, si bien se considera que la aportación práctica principal de este trabajo puede ser el *Diagrama Final*, se tiene confianza en que de cada uno de los capítulos de esta tesis pueden extraerse elementos valiosos para los cursos de probabilidad y estadística. elementos que van desde aspectos teóricos hasta el conocimiento y manejo de interrelaciones entre varias de las distribuciones de variables aleatorias comunes, pasando por información de muy diversa índole sobre tales distribuciones

Aspectos por mejorar

Todo trabajo de investigación es perfectible en función de la disponibilidad de tiempo y de información que se tenga para continuar su desarrollo. Ahondar más en las características de cada una de las distribuciones no sólo enriquecería el contenido del Capítulo 2, sino también es muy posible que contribuya al hallazgo de otras relaciones que pudieran incorporarse al diagrama. Por otro lado, valdría la pena abordar la *función característica* en el Capítulo 1 como una herramienta adicional para determinar la distribución de una función de una o más variables aleatorias (particularmente de su suma), pues mientras que la función generadora de momentos no existe para todas las distribuciones, la función característica sí. Esto permitiría comprender plenamente la verificación de la relación (56) y hubiera sido útil para corroborar la (57)

Horizontes futuros

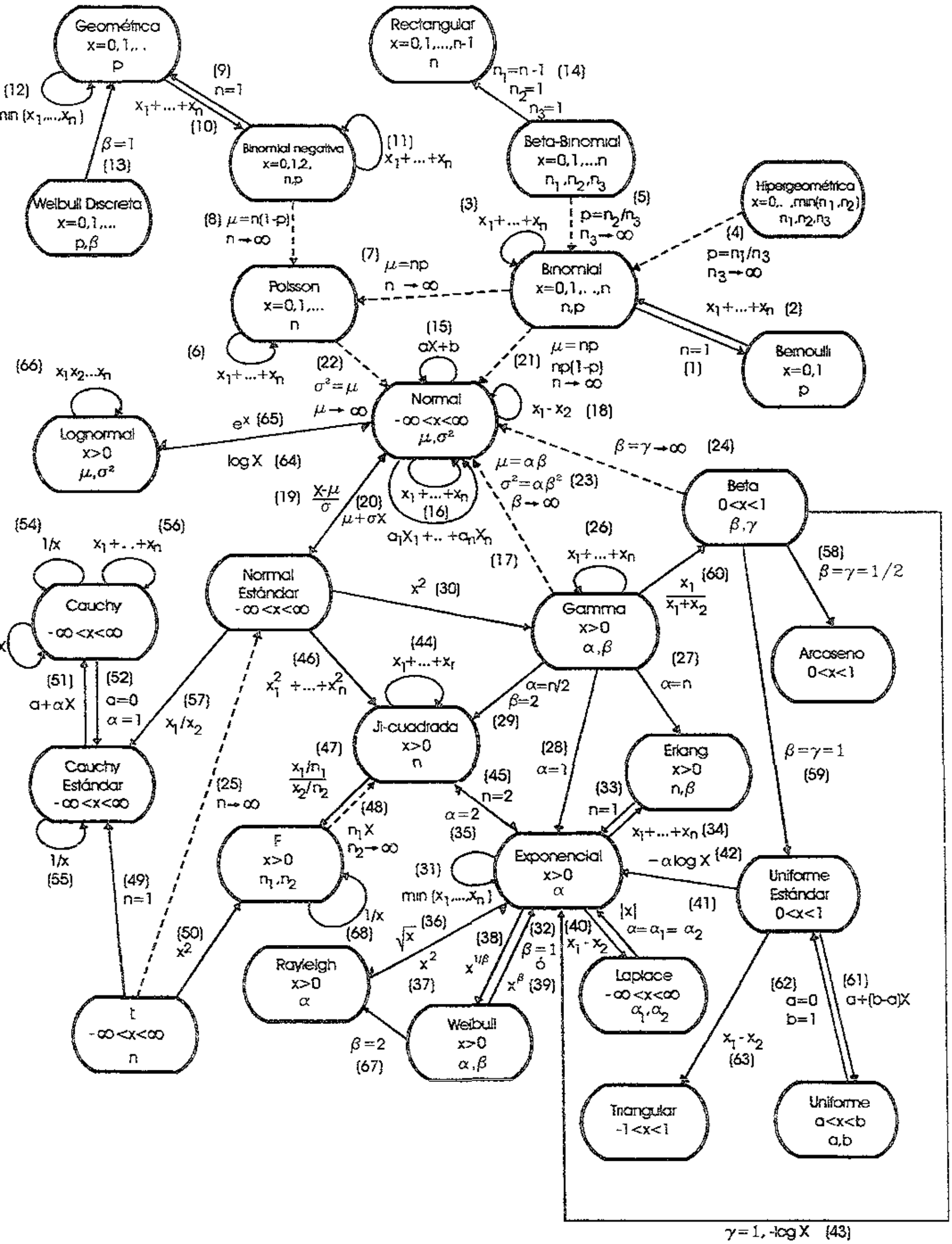
Como una forma de enriquecer este trabajo y darle un enfoque más práctico se vislumbra la posibilidad de incorporar *ejemplos de aplicación de las distribuciones* que resulten *más tangibles* para los lectores, así como *ejemplos de aplicación práctica de las distintas relaciones verificadas*.

Asimismo, las puertas para el hallazgo de nuevas relaciones entre las distintas distribuciones, que atiendan a criterios similares a los que aplican sobre aquéllas encontradas hasta ahora, quedan abiertas. Una posible veta en este sentido podría incluir las relaciones que aquí quedaron pendientes: [3.5], dentro de la sección dedicada a las relaciones entre distribuciones discretas en el tercer capítulo, y [A.1] y [A.2], propuestas al final del mismo capítulo

Así, toda colaboración que derive en la mejora de este trabajo será gratamente recibida, bajo la consigna de que el material que aquí se presenta busca ser de utilidad para quienes estén vinculados con la probabilidad y la estadística, empezando, por supuesto, por los alumnos y profesores de estos cursos en la Facultad de Ciencias.[2]

[2] Un forma para hacer llegar propuestas puede ser a través de la maestra Margarita Chavez Cano, directora de esta tesis.

Diagrama Final: *Relaciones entre las distribuciones**



* Adaptación del diagrama de Leemis (1986). Resultado del trabajo de tesis "Algunas relaciones entre distribuciones de variables aleatorias univariadas discretas y continuas" (2000).

BIBLIOGRAFÍA

1. Casella, George y Berger, Roger L. [1990], *Statistical inference*. Estados Unidos: Duxbury Press, pp. 1–246 (capítulos 1–5)

2. Johnson, Norman L y Kotz Samuel [1969], *Discrete univariate distributions*, Estados Unidos John Wiley, FALTAN PAGINAS.

3 Johnson, Norman L y Kotz Samuel [1970], *Continous univariate distributions-1*, Estados Unidos: John Wiley, pp 40–136, 154–232 y 250–271 (capítulos 13, 14, 16–18 y 20)

4. Johnson, Norman L. y Kotz Samuel [1970], *Continous univariate distributions-2*, Estados Unidos: John Wiley, pp. 22–129 (capítulos 23–27)

5. Leemis, Lawrence M. [1986], "Relations among common univariate distributions", en *The American Statistician*, Mayo, Vol. 40, No 2, pp 143–146.

6. Mendenhall, William, Scheaffer, Richard L y Wackerly Dennis, L. [1986], *Estadística Matemática con Aplicaciones*, México: Grupo Editorial Iberoamérica, pp. 73–296 (capítulos 3–7).

7 Mood, Alexander M , Graybill, Franklin A. y Boes, Duane C. [1974], *Introduction to the theory of statistics*, Singapur McGraw-Hill, pp. 1–270 (capítulos I–VI).

8. Ross, Sheldon [1988], *A first course in probability*, Singapur. Macmillan, pp 23–257 y 108–361 (capítulos 2 y 4–8).

9. Tsokos, Chris P. [1972], *Probability distributions. an introduction to probability theory with applications*, Estados Unidos: Duxbury Press, pp. 92–424 (capítulos 2–6).

ANEXOS

ANEXO 1

ESPACIO DE PROBABILIDAD

Definición A.1.1
Espacio muestral

Se llama *espacio muestral* de un experimento al conjunto, Ω, de todos los posibles resultados de ese experimento

♦

Una vez definido un espacio muestral, se está en condición de considerar conjuntos de los resultados posibles de un experimento.

Definición A.1.2
Evento y espacio de eventos

Un *evento* es cualquier colección, E, de posibles resultados de un experimento, es decir, cualquier subconjunto de Ω. La clase de todos los eventos asociados con un experimento determinado se define como el *espacio de eventos*.

♦

A cada evento E en el espacio muestral Ω se quiere asociar un número entre cero y uno, el cual se denominará la *probabilidad de ocurrencia de* E (o, simplemente, *probabilidad de* E) denotada como $P[E]$. Para ello, se requiere definir el dominio de la función $P[\cdot]$ como todos los subconjuntos de Ω Es decir, para cada $E \subset \Omega$ se define $P[E]$ como la probabilidad de que E ocurra. Pero esto no es tan simple Hay algunas dificultades técnicas que hay que superar, mismas que suelen ser más de interés para los probabilistas que para los estadísticos (Casella y Berger [1990], p. 6). Aquí se expondrán algunos de estos "tecnicismos", considerados como necesarios para cimentar teóricamente la definición de *variable aleatoria*.

Definición A.1.3
Sigma-álgebra

Una colección de subconjuntos de Ω se llama *sigma-álgebra* (o *campo de Borel*), denotada como $\mathcal{B}$, si satisface las tres propiedades siguientes:

1. $\phi \in \mathcal{B}$.
2. Si $A \in \mathcal{B}$ entonces $A^c \in \mathcal{B}$.
3. Si $A_1, A_2, \ldots \in \mathcal{B}$ entonces $\bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$.

Haciendo uso del concepto de sigma-álgebra se puede definir una función de probabilidad.

♦

Definición A.1.4
Función de probabilidad

Dado un espacio muestral Ω y una sigma-álgebra $\mathcal{B}$, una *función de probabilidad* es una función $P[\cdot]$ con dominio $\mathcal{B}$ que satisface.

1. $P[A] \geq 0$ para todo $A \in \mathcal{B}$.
2. $P[\Omega] = 1$
3. Si $A_1, A_2, \ldots \in \mathcal{B}$ y $A_1 \cap A_2 = \phi$ para todo $i \neq j$ entonces $P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$

◈

A las tres propiedades dadas en la *Definición A 1 4* se les llama usualmente *Axiomas de la Probabilidad* o *Axiomas de Kolmogorov* [1]

Para cualquier espacio muestral pueden definirse varias funciones de probabilidad diferentes. Cuál de ellas refleja lo que es factible de observar en un experimento particular es algo que debe discutirse. La realidad física del experimento puede dictar la asignación de probabilidad, y otras veces esa asignación puede ser más bien intuitiva

Las *Definiciones A.1.1, A 1.3* y *A 1 4* convergen en la definición de "espacio de probabilidad" que se presenta a continuación

Definición A.1.5
Espacio de probabilidad

Un *espacio de probabilidad* es la terna $(\Omega, \mathcal{B}, P[\cdot])$, donde Ω es un espacio muestral, $\mathcal{B}$ es una sigma-álgebra de eventos y $P[\cdot]$ es una función de probabilidad con dominio $\mathcal{B}$

◈

Definir un espacio de probabilidad *no* es una mera cuestión de formalidad. Es un término único que proporciona una forma expedita de suponer la existencia de los tres componentes en su notación Estos tres componentes están relacionados $\mathcal{B}$ es una colección de subconjuntos de Ω, y $P[\cdot]$ es una función que tiene a $\mathcal{B}$ como su dominio El uso principal del espacio de probabilidad es proveer un método conveniente para establecer supuestos previos para establecer definiciones y teoremas

[1] Puesto que más adelante se calcularán probabilidades, recuerde que $P[\cdot]$ cumple tres propiedades Suponiendo un espacio muestral, Ω, y una sigma–álgebra, $\mathcal{B}$, dados de tal manera que $P[\cdot]$ es una funcion de probabilidad con dominio $\mathcal{B}$

- *(i)* $P[\phi] = 0$.
- *(ii)* Si $A_1, \ldots, A_n$ son eventos mutuamente excluyentes en $\mathcal{B}$ entonces $P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$.
- *(iii)* Si A es un evento en $\mathcal{B}$ entonces $P[A^c] = 1 - P[A]$

ANEXO 2

CÁLCULO DE PROBABILIDADES PARA EL EJEMPLO 1.2

Como ya se había dicho en el Capítulo 1, el valor de una variable aleatoria puede estar determinado por el resultado de un experimento y por ende pueden asignarse probabilidades a los posibles valores de la variable aleatoria $P[X(\omega) = x]$ donde $P[X = x]$, donde $\omega \in \Omega$ y $x \in \Re_X$.[2] No sólo eso, pueden calcularse también probabilidades de eventos como $\{X(\omega) \leq x\} = \{X \leq x\}$ o $\{X = x\}$, para *cualquier* número real x. Así, retomando el *Ejemplo 1.2*,

$$
\begin{aligned}
P[X=0] &= P[(S,S,S)] = \tfrac{1}{8}\\
P[X=1] &= P[(A,S,S) \text{ o } (S,A,S) \text{ o } (A,S,A)] = \tfrac{3}{8}\\
P[X=2] &= P[(A,A,S) \text{ o } (A,S,A) \text{ o } (S,A,A)] = \tfrac{3}{8}\\
P[X=3] &= P[(A,A,A)] = \tfrac{1}{8},
\end{aligned}
$$

Como X puede tomar los valores 0 a 3, debe cumplirse que

$$1 = P\left(\bigcup_{x=0}^{3}\{X = x\}\right) = \sum_{x=0}^{3} P[X = x]$$

que, por supuesto, se corresponde con las probabilidades calculadas arriba.

Ahora, obsérvese también que:

$$
\begin{aligned}
&P[X=0.5] = 0,\\
&P[X=4] = 0,\\
&P[X=-38] = 0,
\end{aligned}
$$

y, en general, $\quad P[X=x] = 0 \quad \text{para todo } x \notin \Re_X.$

Asimismo,

$$
\begin{aligned}
&P[X \leq 6.4] = P[X \leq 6] = 1,\\
&P[X < 3] = P[X \leq 2] = \tfrac{7}{8},\\
&P[X > 2] = P[X \geq 3] = \tfrac{1}{8},\\
&P[X \leq 0] = \tfrac{1}{8},\\
&P[X \leq -2] = 0,
\end{aligned}
$$

y, en general, $\quad P[X \leq x] = 0 \quad \text{para todo } x < 0 \text{ (o } P[X < x] = 0, \text{ para todo } x < 0.$

así como, $\quad P[X > x] = 1 \quad \text{para todo } x > 3 \text{ (o } P[X \geq x] = 1, \text{ para todo } x > 3.$

♦

[2] De hecho, Tsokos [1972] en la definición que presenta de variable aleatoria establece como supuesto que pueden calcularse las probabilidades de $\{X(\omega) = x\}$ y de $\{X(\omega) \leq x\}$, donde x es cualquier número real en $\Re_X$.

ANEXO 3

PROPIEDADES DE LA FUNCIÓN DE DISTRIBUCIÓN ACUMULATIVA Y DE LA FUNCIÓN DE DENSIDAD [MASA] DE PROBABILIDAD

Caso univariado:

Teorema A.3.1
Propiedades de la f.d.a.

Una función $F_X(\cdot)$ es una función de distribución acumulativa si y sólo si cumple las tres condiciones siguientes:

1. $F_X(-\infty) \equiv \lim_{x \to -\infty} F_X(x) = 0$ y $F_X(+\infty) \equiv \lim_{x \to +\infty} F_X(x) = 1$.
2. $F_X(\cdot)$ es monótona no decreciente. Es decir, si $a \leq b$ entonces $F_X(a) \leq F_X(b)$.
3. $F_X(\cdot)$ es continua por la derecha. Es decir, $\lim_{h \to 0} F_X(x + h) = F_X(x)$, $h > 0$.

Teorema A.3.2
Propiedades de una f.m.p. y de una f.d.p.

Una función $f_X(x)$ es una función de densidad de probabilidad [o función de masa de probabilidad] de una variable aleatoria X si y sólo si

1. $f_X(x) \geq 0$ para todo x.
2. $\sum_x f_X(x) = 1$, en el caso discreto, o $\int_{-\infty}^{\infty} f_X(x)dx = 1$, en el caso continuo.

Caso multivariado:

Teorema A.3.3
Propiedades de $F_{X,Y}(\cdot,\cdot)$

(i) $F_{X,Y}(-\infty, y) \equiv \lim_{x \to -\infty} F_{X,Y}(x, y) = 0$ para toda y, $F_X(x,-\infty) \equiv \lim_{y \to -\infty} F_{X,Y}(x, y) = 0$ para toda x, y $\lim_{\substack{x \to \infty \\ y \to \infty}} F_{X,Y}(x, y) = F_{X,Y}(\infty,\infty) = 1$.

(ii) Si $x_1 < x_2$ y $y_1 < y_2$, entonces

$$P[x_1 \leq X \leq x_2; y_1 \leq Y \leq y_2]$$

$$= F_{X,Y}(x_2, y_2) - F_{X,Y}(x_2, y_1) - F_{X,Y}(x_1, y_2) + F_{X,Y}(x_1, y_1) \geq 0.$$

(iii) $F_X(\cdot)$ es continua por la derecha en cada argumento; es decir,

$$\lim_{0<h\to 0} F_X(x+h, y) = \lim_{0<h\to 0} F_X(x, y+h) = F_{X,Y}(x, y).$$

♦

Teorema A.3.4
Propiedades de una f.m.p. y de una f.d.p. conjunta

1. $f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) \geq 0$
2. Si $(X_1, \ldots, X_k)$ es un vector aleatorio discreto $\sum f_{X_1,\ldots,X_k}(x_1,\ldots,x_k) = 1$, donde la suma es sobre todos los posibles valores de $(X_1, \ldots, X_k)$; y si $(X_1, \ldots, X_k)$ es un vector aleatorio continuo $\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f_{X_1,\ldots,X_k}(x_1,\ldots,x_k)dx_1 \ldots dx_k = 1$.

♦

DEMOSTRACIÓN DEL TEOREMA 1.6

Teorema 1.6
Independencia de funciones de v.a.i.'s

Si $X_1,\ldots, X_k$ son variables aleatorias independientes y $g_1(\cdot),\ldots, g_k(\cdot)$ son k funciones tales que $Y_j = g_j(X_j)$, $j = 1,\ldots, k$ son variables aleatorias, entonces $Y_1,\ldots, Y_k$ también son independientes.

Demostración. Defínase $g_j^{-1}(B_j) = \{z : g_j(z) \in B_j\}$.

Entonces los eventos $\{Y_j \in B_j\}$ y $\{X_j \in g_j^{-1}(B_j)\}$ son equivalentes; y por lo tanto,

$$P[Y_1 \in B_1,\ldots, Y_k \in B_k] = P[X_1 \in g_1^{-1}(B_1),\ldots, X_k \in g_k^{-1}(B_k)] = \prod_{j=1}^{k} P[X_j \in g_j^{-1}(B_j)] = \prod_{j=1}^{k} P[Y_j \in$$

♦

ANEXO 4

MUESTRA ALEATORIA

De manera breve se explicará en qué consiste el procedimiento de muestreo aleatorio para la obtención de una *muestra aleatoria* de n observaciones. Sean N y n el número de elementos en la población y en la muestra, respectivamente Si se hace el muestreo de tal manera que cada una de las $\binom{N}{n}$ muestras tiene la misma probabilidad de ser elegida, el muestreo se denomina aleatorio y el resultado es una *muestra aleatoria*. A su vez, la selección de cada uno de los elementos de la muestra puede realizarse *con* o *sin reemplazo*. El método de selección con reemplazo consiste en extraer los n elementos de tal manera que todos los N elementos de la población tengan la misma probabilidad ($1/N$) de ser extraídos; es decir, una vez elegido el primer elemento éste "se devuelve" a la población para así elegir el segundo, y así sucesivamente. En el método de selección sin reemplazo la probabilidad de extracción del i-ésimo elemento es *mayor* que aquélla del elemento previo, $i = 1,2,\ldots,n$, pues cada elemento seleccionado se va descartando y con ello el denominador (casos totales) se reduce, siendo $1/N$ y $1/_{N-n-1}$ las probabilidades de selección del primer y del último elementos de la muestra

ESTADÍSTICAS DE ORDEN

Las observaciones más pequeña, más grande o media son valores de una muestra aleatoria que pueden proporcionar información sintética acerca de la muestra. Por ejemplo, la más baja temperatura invernal o el más alto nivel de precipitación registrados en los años recientes determinado pueden ser de interés para la el diseño de acciones para atender emergencias en el futuro. El precio mediano de las casas vendidas en el mes previo pueden ser de utilidad para estimar el costo de la vida Todos estos son ejemplos de *estadísticas de orden*

Definición A.4.1
Estadísticas de orden

Las *estadísticas de orden* de una muestra aleatoria $X_1, X_2,\ldots, X_n$ son los valores muestrales colocados en orden ascendente, y se denotan como $X_{(1)}, X_{(2)},\ldots, X_{(n)}$. Así, las estadísticas de orden son variables aleatorias que satisfacen $X_{(1)} \leq \ldots \leq X_{(n)}$, de tal forma que

$$\begin{aligned} X_{(1)} &= \text{mín}[X_1,\ldots,X_n] \\ X_{(2)} &= \text{segunda más pequeña } X_i \\ &\vdots \\ X_{(n)} &= \text{máx}[X_1,\ldots,X_n] \end{aligned}$$

A $X_{(1)}$ se le llama la *mínima estadística de orden* y a $X_{(n)}$ la *máxima estadística de orden*.

◆

Por ser variables aleatorias, tiene sentido discutir las probabilidades de que las estadísticas de orden tomen diversos valores. Para calcular estas probabilidades se requiere de sus funciones de densidad [o de masa] de probabilidad. Ahora bien, es importante notar que las estadísticas de orden son *funciones* de las variables aleatorias en la muestra, por lo cual sus probabilidades pueden ser calculadas en términos de las probabilidades de la muestra. Para el caso en que $X_1, X_2, \ldots, X_n$ sean variables aleatorias (independientes e idénticamente distribuidas) discretas el cálculo de *probabilidades para las estadísticas de orden consiste en esencia en una tarea de conteo.*

A continuación se presentan las funciones de distribución y de densidad [masa] de probabilidad de la j-ésima estadística de orden, tanto para el caso discreto como continuo. En el Capítulo 1 se particulariza el resultado continuo para la mínima y la máxima estadísticas de orden como ilustración de la técnica de la función de distribución.

Teorema A.4.1
Distribución de la *j*-ésima estadística de orden de una m.a. para el caso discreto

Sea $X_1, X_2, \ldots, X_n$ una muestra aleatoria de una distribución discreta con función de masa de probabilidad $f_X(x_i) = p_i$ donde $x_1 < x_2 < \ldots$ son los posibles valores de X en orden ascendente. Defínase

$$\begin{aligned} P_0 &= 0 \\ P_1 &= p_1 \\ P_2 &= p_1 + p_2 \\ &\vdots \\ P_i &= p_1 + p_2 + \ldots + p_i \\ &\vdots \end{aligned}$$

Sean $X_{(1)}, \ldots, X_{(n)}$ las estadísticas de orden de la muestra. Entonces

$$F_{X_{(j)}}(x_i) = P[X_{(j)} \le x_i] = \sum_{k=j}^{n} \binom{n}{k} P_i^k (1-P_i)^{n-k} \qquad [A.4.1.1]$$

y

$$f_{X_{(j)}}(x_i) = P[X_{(j)} = x_i] = \sum_{k=j}^{n} \binom{n}{k} [P_i^k (1-P_i)^{n-k} - P_{i-1}^k (1-P_{i-1})^{n-k}] \qquad [A.4.1.2].$$

♦

Teorema A.4.2
Distribución de la *j*-ésima estadística de orden de una m.a. para el caso continuo

Sean $X_{(1)},\ldots, X_{(n)}$ las estadísticas de orden de una muestra aleatoria, $X_1,\ldots, X_n$, de una población continua con función de distribución $F_X(x)$ y función de densidad de probabilidad $f_X(x)$. Entonces

$$F_{X_{(j)}}(x)=\sum_{k=j}^{n}\binom{n}{k}[F_X(x)]^k[1-F_X(x)]^{n-k} \qquad [A.4.2.1]$$

y

$$f_{X_{(j)}}(x)=\frac{n!}{(j-1)!\,(n-j)!}\,f_X(x)[F_X(x)]^{j-1}[1-F_X(x)]^{n-j} \qquad [A.4.2.2]$$

◆

- *Observación*: El resultado [*A.4.2.1*] no se restringe a distribuciones discretas; en ese sentido, [*A.4.1.1*] es un corolario de [*A.4.2.1*].

ANEXO 5

ESPERANZAS Y MOMENTOS

El concepto de esperanza de una variable aleatoria es uno de los más importantes en la teoría de la probabilidad. A la esperanza de una variable aleatoria X se le llama también *valor esperado* o *media* de X. Una de sus propiedades más importantes es la de ser una buena aproximación precisamente para un valor de X. Como se verá más adelante, la esperanza corresponde al primer momento de una variable aleatoria. Además, es posible determinar los diferentes momentos de algunas variables aleatorias mediante la "función generadora de momentos", misma que es una herramienta sumamente poderosa en la determinación de la distribución de una función de una (o varias) variable(s) aleatoria(s); en particular, de la suma de variables aleatorias.

Definición A.5.1
Esperanza de una v.a. discreta

Si X es una variable aleatoria discreta con función de masa de probabilidad $f_X(\cdot)$, entonces la *esperanza*, el *valor esperado* o la *media* de X, denotada por $E[X]$, se define como

$$E[X] = \sum_x x f_X(x)$$

♦

Así, para el caso discreto, el valor esperado de X puede entenderse como un promedio ponderado de los posibles valores que puede tomar X, donde el peso (ponderador) asignado a cada valor es justamente su probabilidad de ocurrencia.[3]

Definición A.5.2
Esperanza de una v.a. continua

Si X es una variable aleatoria continua con función de densidad de probabilidad $f_X(\cdot)$, entonces la *esperanza*, el *valor esperado* o la *media* de X, denotada por $E[X]$, se define como

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx$$

♦

[3] Ross menciona que otra motivación de la definición de esperanza está dada por la interpretación frecuentista de las probabilidades (justificada en parte por la ley fuerte de los grandes números). Suponga que X debe tomar valores $x_1, x_2, \ldots, x_n$ con probabilidades $f_X(x_1), f_X(x_2), \ldots, f_X(x_n)$, respectivamente; y considere a X como la ganancia en un solo juego de azar. Es decir, con probabilidad $f_X(x_i)$ se ganan x_i unidades, $i = 1,2,\ldots,n$. Ahora, por la interpretación frecuentista se sigue que de jugarse continuamente este juego, entonces la proporción de tiempo que se gana x_i será $f_X(x_i)$. Como esto es cierto para todo i, $i = 1,2,\ldots,n$, se sigue que la ganancia promedio por juego será $\sum_{i=1}^{n} x_i f_X(x_i) = E[X]$.

- *Nota:* Es posible que para algunas variables aleatorias la esperanza no exista, en el sentido de que no sea finita [4]

Puesto que para definir la media de una variable aleatoria X se recurre a la función de densidad de probabilidad de X se puede decir también que $E[X]$ es la media de la densidad de X. Además, $E[X]$ es el centro de gravedad de la unidad de masa determinada por la función de densidad de X. Así, la media de X es una medida de dónde están "centrados" los valores de X.

Frecuentemente se tiene una variable aleatoria dada y su distribución de probabilidad, pero se está interesado no en el valor esperado de X sino en el valor esperado de alguna función de X, digamos $g(X)$. Como $g(X)$ es también una variable aleatoria, una forma de obtener su valor esperado es calculando su distribución de probabilidad (a partir del conocimiento de la distribución de X) y con ella calcular $E[g(X)]$ vía la definición de esperanza. Pero –afortunadamente– existe una forma mucho más sencilla de hacerlo, que suprime la necesidad del cálculo de la distribución de probabilidad de $g(X)$ que en ocasiones puede ser engorroso o complejo. Esta forma se expone en el teorema siguiente.

Teorema A.5.1 [5]
Esperanza de una función de una v.a.

(a) Si X es una variable aleatoria discreta con función de masa de probabilidad $f_X(x)$, entonces para cualquier función g con valores reales

$$E[g(X)] = \sum_x g(x) f_X(x).$$

(b) Si X es una variable aleatoria continua con función de densidad de probabilidad $f_X(x)$, entonces para cualquier función g con valores reales

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) dx.$$

◊

Además, el operador esperanza cumple varias propiedades que facilitan el cálculo de los valores esperados.

A una clase importante de esperanzas se les llama "momentos". A continuación se introducen las definiciones de momentos y de función generadora de momentos.

---

[4] Una variable aleatoria con distribución Cauchy, que se presenta en el Capítulo 2, es un ejemplo de esta situación.

[5] Versión tomada de Ross [1988], *op. cit.*, p. 255. De manera muy elocuente, Ross llama a este resultado *Ley del estadístico inconsciente*, pensando en la idea de que un estadístico tiene un enfoque más práctico que un probabilista, y en ese sentido corre el "riesgo" de no darse cuenta de que –por definición– lo que requiere es $f_Y(y)$, si $Y = g(X)$, en lugar de $f_X(x)$. Lo sorprendente es que tal "omisión" no conduce a un resultado incorrecto.

Definición A.5.3
Momento y momento central

Si X es una variable aleatoria, el *r-ésimo momento* de X (o F_X), μ'_r, se define como

$$\mu'_r = E[X^r].$$

y el *r-ésimo momento central* de X, μ_r, se define como

$$\mu_r = E[(X - \mu)^r],$$

donde $\mu = \mu'_1 = E[X]$.

◆

Ésta es la definición del r-ésimo momento central alrededor de la media, pero que puede definirse el r-ésimo momento central alrededor de cualquier número real a como $E[(X - a)^r]$.

Un momento tan importante como la media de una variable aleatoria es el segundo momento central, más comúnmente conocido como la varianza.

ANEXO 6

PROPIEDADES DE LA FUNCIÓN GAMMA

La función gamma $\Gamma(\alpha)$ aparece no sólo en la expresión de la densidad del mismo nombre, sino también de las densidades *t de Student*, *F*, *ji-cuadrada* y *beta* (véase Capítulo 2). Está definida como:

$$\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy, \quad \alpha > 0.$$

El valor de esta integral es finito como consecuencia de que α es una constante positiva. Si además se considera que α es un entero positivo la integral puede expresarse en forma cerrada, es decir, en términos de funciones elementales (de otra forma no es posible).

A continuación se presentan algunas de las propiedades que satisface la función gamma, las cuales son muy útiles en la verificación de varias de las relaciones expuestas en el Capítulo 3.

(*i*) $\Gamma(\alpha + 1) = \alpha\,\Gamma(\alpha),$ si $\alpha > 0$

(*ii*) $\Gamma(n) = (n-1)!\,,$ si n es un entero positivo

(*iii*) $\Gamma\left(\dfrac{n}{2}\right) = \dfrac{(n-1)!\sqrt{\pi}}{2^{n-1}\left(\dfrac{n-1}{2}\right)!}\,,$ si n es un entero positivo *impar*

(*iv*) $\Gamma\left(n + \dfrac{1}{2}\right) = \dfrac{1\cdot 3\cdot 5\cdot \ldots \cdot(2n-1)}{2^n}\sqrt{\pi}$, si n es un entero

La propiedad (*i*) puede demostrarse mediante integración por partes, (*ii*) puede demostrarse combinando (*i*) con el hecho fácilmente verificable de que $\Gamma(1) = 1$; (*i*) y (*ii*) proporcionan relaciones recursivas que facilitan el cálculo complejo de valores de la función gamma. La demostración de las propiedades (*iii*) y (*iv*) requiere de un desarrollo algebraico cuidadoso.

Casos particulares de uso recurrente en el Capítulo 3 son:

- $\Gamma(1) = 1$ (aplicando la propiedad (*ii*) con $n = 1$)
- $\Gamma(½) = \sqrt{\pi}$ (aplicando la propiedad (*iii*) con $n = 1$)[6]

Finalmente, la función beta puede expresarse en términos de funciones gamma como $B(\beta,\gamma) = \dfrac{\Gamma(\beta)\Gamma(\gamma)}{\Gamma(\beta + \gamma)}$.

---

[6] Aplicando la propiedad (*iv*) para $n = 1$ se puede verificar que $\Gamma(1/2) = 2\Gamma(3/2) = \sqrt{\pi}$

ANEXO 7

ALGUNAS FUNCIONES GENERADORAS DE MOMENTOS

Distribución	Función generadora de Momentos
Discretas	
Bernoulli (p)	$q + pe^t$
Binomial (n,p)	$(q + pe^t)^n$
Binomial negativa (r,p)	$\left(\frac{p}{1-qe^t}\right)^r$
Geométrica (p)	$\left(\frac{p}{1-qe^t}\right)$
Hipergeométrica (n_1, n_2, n_3)	No es útil
Poisson (μ)	$\exp[\mu(e^t - 1)]$
Rectangular (uniforme discreta) (n)	$\sum_{j=0}^{n-1} \frac{1}{n} e^{jt}$

Distribución	Función generadora de Momentos
Continuas	
Beta (α,β)	No es útil
Cauchy (a,α)	No existe
Exponencial (α)	$\frac{1}{1-\alpha t}$, para $t < 1/\alpha$
$F\ (n_1, n_2)$	No existe
Gamma (α,β)	$\left(\frac{1}{1-\alpha t}\right)^{\beta}$, para $t < 1/\alpha$
Ji-cuadrada (n)	$\left(\frac{1}{1-2t}\right)^{n/2}$, para $t < ½$
Lognormal (μ,σ^2)	No es útil
Normal (μ,σ^2)	$\exp[\mu t + ½\, \sigma^2 t^2]$
t de Student (n)	No existe
Uniforme (a,b)	$\frac{e^{bt} - e^{at}}{(b-a)t}$

ANEXO 8

ARTÍCULO DE LAWRENCE LEEMIS

Relationships Among Common Univariate Distributions

LAWRENCE M. LEEMIS*

Common univariate distributions are usually discussed separately in introductory probability textbooks, which makes it difficult for students to understand the relationships among these distributions. The purpose of this article is to present a figure that illustrates some of these relationships.

KEY WORDS: Limiting distributions; Transformations of random variables.

1. INTRODUCTION

Students in a first course in probability usually study common univariate distributions. Most introductory textbooks discuss each of the distributions in separate sections. One of the drawbacks of this approach is that students often do not grasp all of the interrelationships among the distributions. The purpose of this article is to present and discuss a figure that overcomes this shortfall.

There are several excellent sources for studying univariate distributions. Hastings and Peacock's (1975) handbook shows graphs of densities and variate relationships for several distributions. Hirano, Kuboki, Aki, and Kuribayashi (1983) gave graphs of univariate distributions for many combinations of parameter values. For more detail, Johnson and Kotz (1970) have done a four-volume series covering univariate and multivariate distributions. Recently, Patil, Boswell, Joshi, and Ratnaparkhi (1985) and Patil, Boswell, and Ratnaparkhi (1985) have also completed volumes on discrete and continuous distributions. Other books on distributions and modeling include Ord (1972), Patel, Kapadia, and Owen (1976), and Shapiro and Gross (1981). Diagrams that relate these distributions to one another may be found in Nakagawa and Yoda (1977), Taha (1982), and Marshall and Olkin (1985).

2. DISCUSSION

The diagram in Figure 1 shows some relationships among common univariate distributions that might be presented in

*Lawrence M. Leemis is Assistant Professor, School of Industrial Engineering, University of Oklahoma, 202 West Boyd, Room 124, Norman, OK 73019 The author expresses gratitude to Bruce Schmeiser for his helpful advice on the diagram, and to the referees and editor for their careful reading and helpful suggestions.

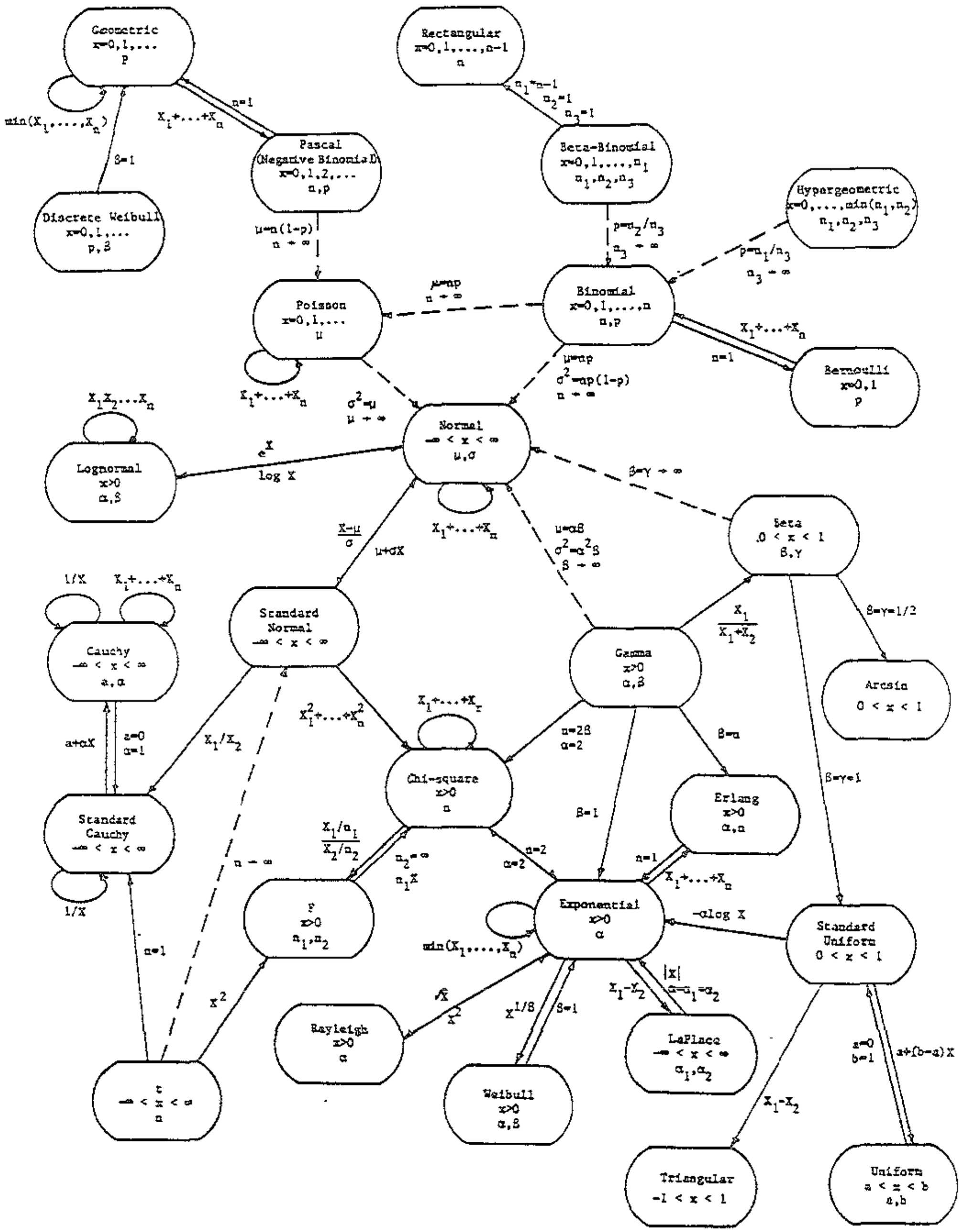


Figure 1. Relationships Among Distributions.

an introductory probability course. There are 9 discrete distributions, shown on the upper part of the diagram, and 19 continuous distributions. The first line of each entry is the name of the distribution. The next line contains the region of support for the distribution. The last line contains the distribution's parameters. The parameters must satisfy the following: n is an integer; $0 < p < 1$; α and σ are positive scale parameters; β and γ are positive shape parameters; and μ, a, and b are location parameters. When μ or σ are used as parameters, they denote the mean and standard deviation of the distribution, respectively.

There are three types of relationships among distributions: limiting distributions, transformations, and special cases. Limiting distributions are indicated with a dashed arrow. Transformations (which assume independent random variables) and special cases are indicated by a solid arrow. One-to-one transformations have an arrow pointing in both directions. The random variable X is used for all distributions. Thus the arrow from the t distribution to the F distribution indicates that the square of a t random variable has an F distribution.

The normal and exponential distributions play a central role in Figure 1. This fact is partially due to their natural genesis via the central limit theorem and superpositioning principle. In addition, some distributions generalize the exponential distribution (such as the Weibull), since it is often used in reliability to model component lifetimes.

The transformation relationships in Figure 1 can be combined to form other relationships. A path from the standard normal to the chi-square to the exponential to the Rayleigh, for example, indicates that the random variable $(X_1^2 + X_2^2)^{1/2}$ has the Rayleigh distribution if X_1 and X_2 are standard normal random variables.

The relationship between the uniform (0, 1) distribution and the exponential distribution is valid by the probability integral transformation. Since the probability integral transformation states that the cumulative distribution function for a random variable is uniformly distributed between 0 and 1, an arrow could be drawn from the uniform (0, 1) distribution to every other distribution shown in the figure, although not all of these relationships are closed form. This relationship is known in the simulation literature as the inverse-cdf technique for random variate generation. A survey article of general methods for random variate generation is given by Schmeiser (1980).

The probability mass functions and probability density functions for the distributions in Figure 1 are listed in the Appendix. All continuous lifetime distributions [i.e., those with support on $(0, \infty)$] may be generalized to have support on (a, ∞) by replacing x with $x - a$ in the density function. In addition, the parameterizations chosen for the distributions in Figure 1 are not unique.

There are many relationships that Figure 1 does not indicate. First, because of space constraints, there are relationships (e.g., between the exponential and Poisson distributions) that are not included. Second, combining two random variables with different distributions is not included. For example, the defining formula for the Student-t distribution, $Z(\chi^2/n)^{1/2}$, where Z is standard normal and χ^2 has the chi-square distribution with n df, is not shown. Third, analogies between discrete and continuous distributions (e.g., the geometric and exponential) are not shown. Finally, the distributions included are oriented toward the classical distributions, and families of distributions (e.g., the Pearson system) are not included.

3. CONCLUSIONS

There are two applications for this diagram. First, after presenting common univariate distributions in an introductory course in probability, it can be used to indicate how distributions relate to one another. Second, in an advanced course (e.g., simulation or reliability), Figure 1 provides a quick review of important univariate distributions.

APPENDIX: DISTRIBUTION PARAMETERIZATIONS

Discrete Distributions

Bernoulli:

$$f(x) = p^x(1-p)^{1-x}, \qquad x = 0, 1$$

Beta-Binomial:

$$f(x) = \binom{n_2 + x - 1}{x} \times \binom{n_1 + n_3 - x - 1}{n_1 - x} \Big/ \binom{n_1 + n_2 + n_3 - 1}{n_1}, \qquad x = 0, 1, \ldots, n_1$$

Binomial:

$$f(x) = \binom{n}{x} p^x(1-p)^{n-x}, \qquad x = 0, 1, \ldots, n$$

Discrete Weibull:

$$f(x) = (1-p)^{x^\beta} - (1-p)^{(x+1)^\beta}, \qquad x = 0, 1, \ldots$$

Geometric:

$$f(x) = p(1-p)^x, \qquad x = 0, 1, \ldots$$

Hypergeometric:

$$f(x) = \binom{n_1}{x}\binom{n_3 - n_1}{n_2 - x} \Big/ \binom{n_3}{n_2}, \qquad x = 0, 1, \ldots, \min(n_1, n_2)$$

Pascal (negative binomial):

$$f(x) = \binom{n - 1 + x}{x} p^n (1-p)^x, \qquad x = 0, 1, \ldots$$

Poisson:

$$f(x) = \mu^x e^{-\mu}/x!, \qquad x = 0, 1, \ldots$$

Rectangular:

$$f(x) = 1/n, \qquad x = 0, 1, \ldots, n - 1$$

Continuous Distributions

Arcsin:

$$f(x) = 1/\pi[x(1-x)]^{1/2}, \qquad 0 < x < 1$$

Beta:

$f(x) = [\Gamma(\beta + \gamma)/\Gamma(\beta)\Gamma(\gamma)] \times x^{\beta-1}(1 - x)^{\gamma-1}, \quad 0 < x < 1$

Cauchy:

$f(x) = 1/\alpha\pi[1 + ((x - a)/\alpha)^2], \quad -\infty < x < \infty$

Chi-square:

$f(x) = [1/2^{n/2}\Gamma(n/2)]\, x^{n/2-1}e^{-x/2}, \quad x > 0$

Erlang:

$f(x) = [1/\alpha^n(n - 1)!]\, x^{n-1}e^{-x/\alpha}, \quad x > 0$

Exponential:

$f(x) = (1/\alpha)\, e^{-x/\alpha}, \quad x > 0$

F:

$f(x) = \Gamma((n_1 + n_2)/2) \times (n_1/n_2)^{n_1/2}\, x^{n_1/2-1} / \Gamma(n_1/2)\,\Gamma(n_2/2) \times [(n_1/n_2)\, x + 1]^{((n_1 + n_2)/2)}, \quad x > 0$

Gamma:

$f(x) = [1/\alpha^\beta\Gamma(\beta)]\, x^{\beta-1}e^{-x/\alpha}, \quad x > 0$

LaPlace:

$f(x) = [1/(\alpha_1 + \alpha_2)]e^{-x/\alpha_1}, \quad x \geq 0$

$= [1/(\alpha_1 + \alpha_2)]e^{x/\alpha_2}, \quad x < 0$

Lognormal:

$f(x) = [1/(2\pi)^{1/2}\, x\beta]\exp[-\frac{1}{2}(\log(x/\alpha)/\beta)^2], \quad x > 0$

Normal:

$f(x) = [1/(2\pi)^{1/2}\sigma]\, \exp[-\frac{1}{2}((x - \mu)/\sigma)^2], \quad -\infty < x < \infty$

Rayleigh:

$f(x) = (2x/\alpha)\, e^{-x^2/\alpha}, \quad x > 0$

t:

$f(x) = \Gamma((n + 1)/2)/(n\pi)^{1/2}\Gamma(n/2) \times [x^2/n + 1]^{(n+1)/2}, \quad -\infty < x < \infty$

Triangular:

$f(x) = 1 + x, \quad -1 < x < 0$

$= 1 - x, \quad 0 \leq x < 1$

Uniform:

$f(x) = 1/(b - a), \quad a < x < b$

Weibull:

$f(x) = (1/\alpha)\, \beta x^{\beta-1} \exp[-(1/\alpha)x^\beta], \quad x > 0$

[*Received February 1985. Revised September 1985.*]

REFERENCES

Hastings, N. A. J., and Peacock, J. B. (1975), *Statistical Distributions*, London: Butterworth & Company.

Hirano, K., Kuboki, H., Aki, S., and Kuribayashi, A. (1983), *Figures of Probability Density Functions in Statistics* (Vol. 1 and 2), ed. M. Isida, Computer Science Monographs 19 and 20, The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo, Japan 106.

Johnson, N. L., and Kotz, S. (1970), *Distributions in Statistics* (Vol. 1–4), New York: John Wiley.

Marshall, A. W., and Olkin, I. (1985), "A Family of Bivariate Distributions Generated by the Bivariate Bernoulli Distribution," *Journal of the American Statistical Association*, 80, 332–338.

Nakagawa, T., and Yoda, H. (1977), "Relationships Among Distributions," *IEEE Transactions on Reliability*, 26, 5, 352–353.

Ord, J. K. (1972), *Families of Frequency Distributions*, New York: Hafner Publishing.

Patel, J. K., Kapadia, C. H., and Owen, D. B. (1976), *Handbook of Statistical Distributions*, New York: Marcel Dekker.

Patil, G. P., Boswell, M. T., Joshi, S. W., and Ratnaparkhi, M. V. (1985), *Discrete Models*, Burtonsville, MD: International Co-operative Publishing House.

Patil, G. P., Boswell, M. T., and Ratnaparkhi, M. V. (1985), *Univariate Continuous Models*, Burtonsville, MD: International Co-operative Publishing House.

Schmeiser, B. W. (1980), "Random Variate Generation: A Survey," in *Simulation With Discrete Models: A State-of-the-Art View*, eds. T. I. Oren, C. M. Shub, and P. F. Roth, New York: Institute of Electrical and Electronic Engineers, pp. 79–104.

Shapiro, S. S., and Gross, A. J. (1981), *Statistical Modeling Techniques*, New York: Marcel Dekker.

Taha, H. A. (1982), *Operations Research: An Introduction* (3rd ed.), New York: Macmillan.