



UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

FACULTAD DE CIENCIAS

MODELO MICROANALITICO DE SIMULACION PARA INDICADORES DE EMPLEO EN MEXICO EN EL PERIODO 2000 - 2010

T E S I S

QUE PARA OBTENER EL TITULO DE:

A C T U A R I O

P R E S E N T A :

OMAR DE LA RIVA TORRES



**DIRECTOR DE TESIS:
ACT. TANIA CHAVEZ RAZO**



2038/8

**FACULTAD DE CIENCIAS
SECCION ESCOLAR**



Universidad Nacional
Autónoma de México

Dirección General de Bibliotecas de la UNAM

Biblioteca Central



UNAM – Dirección General de Bibliotecas
Tesis Digitales
Restricciones de uso

DERECHOS RESERVADOS ©
PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL

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

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



UNIVERSIDAD NACIONAL
AUTÓNOMA DE
MÉXICO

MAT. MARGARITA ELVIRA CHÁVEZ CANO
Jefa de la División de Estudios Profesionales de la
Facultad de Ciencias
Presente

Comunicamos a usted que hemos revisado el trabajo de Tesis: "MODELO MICROANALITICO DE SIMULACION PARA INDICADORES DE EMPLEO EN MEXICO EN EL PERIODO 2000-2010"

realizado por DE LA RIVA TORRES OMAR

con número de cuenta 8916372-4 , pasante de la carrera de ACTUARIO

Dicho trabajo cuenta con nuestro voto aprobatorio.

Atentamente

Director de Tesis

Propietario ACT. TANIA CHAVEZ RAZO

Propietario M. EN D. ALEJANDRO MINA VALDES

Propietario DR. EN E. JOAQUIN FRANCISCO CURIEL CANEDO

Suplente ACT. JAIME VAZQUEZ ALAMILLA

Suplente MAT. HUGO VILLASENOR HERNANDEZ

Tania Chavez Razo
J. Curiel

[Signature]
[Signature]

Consejo Departamental de MATEMATICAS

[Signature]
M. EN C. JOSE ANTONIO FLORES DIAZ

Prólogo

Desde las civilizaciones más antiguas todas las actividades de hombre han girado en torno al trabajo. Las primeras ocupaciones del hombre estuvieron orientadas a la sobrevivencia después, cuando se inicia el comercio, la acumulación de riquezas, la división por clases sociales y muchos hombres que ya no pueden satisfacer sus necesidades básicas con sus propias labores, se ven obligados a comerciar con su fuerza trabajo. Esta condición se acentúa con la Revolución Industrial, las nacientes fábricas necesitaron grandes cantidades de mano de obra que sólo puede ser obtenida al atraer campesinos a los centros fabriles en las ciudades. En la actualidad la mayor parte de la población económicamente activa en el mundo se concentra en centros urbanos, empleada principalmente en la producción de bienes y servicios. De la misma forma que en la Revolución Industrial, el uso de la tecnología crea nuevas formas de empleo pero en un corto periodo de tiempo elimina muchos más de los que crea.

La importancia del trabajo para el hombre radica en hacerlo sentirse útil además de ser un medio en la generación de ingresos para la solvencia de sus necesidades. La pérdida o carencia del empleo implica serios problemas tanto para el individuo como para la sociedad. En casos de periodos prolongados de desempleo puede resultar en la disminución de la autoestima. Para la sociedad, reducción de productividad y en algunos casos actitudes antisociales y criminales.

Con esta investigación se pretende hacer un ejercicio del comportamiento probable del empleo en México con el uso de técnicas y métodos de estimación tanto para las variables demográficas como económicas con la utilización de recursos computacionales, empleando el análisis microanalítico.

Los dos primeros capítulos comprenden una descripción del desarrollo del empleo en México y cualidades de la simulación por computadora.

La parte central la constituyen el capítulo tres, donde se presentan la metodología, fuentes de información y justificación de la variables utilizadas y el capítulo cuatro que muestra los resultados de la simulación probando tres hipótesis referentes al empleo.

En los Apéndice A y B se concluye este trabajo exponiendo una recopilación ecléctica para la generación de distribuciones de probabilidad y un prueba estadística de aleatoriedad para el generador de números pseudoaleatorios, de uso elemental para toda la investigación.

Índice

	Prólogo	1
Capítulo 1.	<i>El Empleo en México</i>	5
	1.1. Factores de cambio en la población económicamente activa	5
	1.2. Descripción del empleo de 1900 a 1970	6
	1.3. Cambio en el periodo 1970-1990	10
	1.4. El mercado laboral a nivel nacional	12
Capítulo 2.	<i>Aspectos Generales de la Simulación</i>	16
	2.1. Características principales	16
	2.2. Elementos en la modelación de sistemas	18
	2.3. Clasificación de los modelos de simulación	19
	2.4. Desventajas de la simulación	20
	2.5. Ventajas de la simulación	20
	2.6. Diseño de un modelo de simulación	21
	2.7. Microsimulación	23
	2.8. Tipología de la microsimulación	24
Capítulo 3.	<i>Diseño del Modelo de Simulación</i>	25
	3.1. La utilización de la microsimulación	25
	3.2. Objetivos del modelo	30
	3.3. La PEA como un criterio de la actividad económica	30
	3.4. Movilidad de la fuerza laboral	30
	3.5. Resultados producidos	31

	<i>Índice</i>
3.6. Posibles fuentes de información	35
3.6.1. Educación	35
3.6.2. Salarios	35
3.6.3. Desempleo	36
3.6.4. Limitación de la información	36
3.7. Elección del lenguaje de computación	36
3.8. La muestra inicial	37
3.9. Rutinas demográficas	44
3.10. Rutinas económicas	52
Capítulo 4. Hipótesis de simulación	55
4.1. Variables demográficas	55
4.2. Variables económicas	68
4.3. Hipótesis I: tasa de desempleo abierto para la generación de probabilidades de empleo y desempleo.	71
4.4. Hipótesis II: tasa de desempleo abierto y ocupación informal para la generación de probabilidades de empleo y desempleo.	74
4.5. Hipótesis III: tasa de ingresos insuficientes y desocupación para la generación de probabilidades de empleo y desempleo.	75
Conclusiones	77
Apéndice A	80
A.1. Métodos de generación de números pseudoaleatorios	80
A.1.2. Métodos congruenciales	81
A.1.3. Método de multipasos digitales	85
A.2. Pruebas estadísticas para la validación de números pseudoaleatorios	86
A.2.1. Prueba χ^2 de bondad de ajuste	86
A.2.2. Prueba de bondad de ajuste de Cramer-von Mises	87
A.2.3. Prueba de series	88
A.2.4. Prueba de corridas por arriba y por abajo	88
A.2.5. Prueba de diferencias	89
A.2.6. Prueba de correlación	89
A.3. Generación de distribuciones de probabilidad	90
A.3.1. Método de la transformación inversa	91
A.3.2. Método de rechazo	91
A.3.3. Método de composición o mezclas	92
A.4. Descripción y algoritmos para la generación de distribuciones de probabilidad (distribuciones continuas)	93
A.5. Descripción y algoritmos para la generación de distribuciones de probabilidad (distribuciones discretas)	105
Apéndice B	112
B.1. El generador de números pseudoaleatorios de <i>Mathematica</i>	112
Bibliografía	115

1. El Empleo en México

1.1. Factores de cambio en la población económicamente activa

Fuerza laboral, fuerza de trabajo o Población Económicamente Activa (PEA) se considera al conjunto de personas que participan en la producción de bienes o servicios.

Para el caso de México se considera como límite de edad inferior los doce años. La Población Económicamente Activa ocupada (PEA_o), es aquella que realiza una actividad económica en un instante dado y por la cual recibe a cambio un sueldo, salario u otro tipo de pago en especie; y la Población Económicamente Activa desocupada (PEA_d) es la que no realiza una actividad económica en el mismo instante conforman el total de la PEA.

El estudio de los dos conjuntos anteriores se lleva a cabo a través de los indicadores del empleo.

Los factores económicos son quienes tienen mayor incidencia sobre el comportamiento de la PEA a corto plazo, esta última depende de factores como las relaciones de producción existentes en los distintos sectores, los niveles de salario, la movilidad del factor empleo, avances tecnológico, las relaciones de intercambio que

afectan a los diversos sectores cuya producción esta orientada al mercado externo además de la contracción o expansión de la demanda efectiva. Las decisiones en materia política que se manifiestan en estrategias fiscales, arancelarias o de industrialización así como la injerencia de las organizaciones sindicales afectan la PEA.

A largo plazo los factores demográficos determinan el tamaño, crecimiento y distribución geográfica de la PEA. Las influencias socioculturales y psicológicas señalan las actitudes, valores, habilidades que junto con los factores económicos explican el grado de utilización de la PEA en el sistema productivo.

1.2. Descripción del empleo de 1900 a 1970

Durante el periodo 1900-1960, México experimentó grandes cambios en los aspectos económico, sociocultural, político y demográfico y que impactaron la composición de la PEA en su asignación por sector de actividad. En 1900, más de 60% de su población estaba dedicada a actividades relacionadas con la agricultura, ya en 1960 paso a un 54%. Como resultado de la industrialización y la ampliación del sector de servicios la población no agrícola creció de 1.4 a 5.2 millones, es decir, un 274.2%.

DISTRIBUCIÓN DE LA POBLACIÓN OCUPADA POR SECTORES ECONÓMICOS 1/
1900-1990

Cuadro 1.1

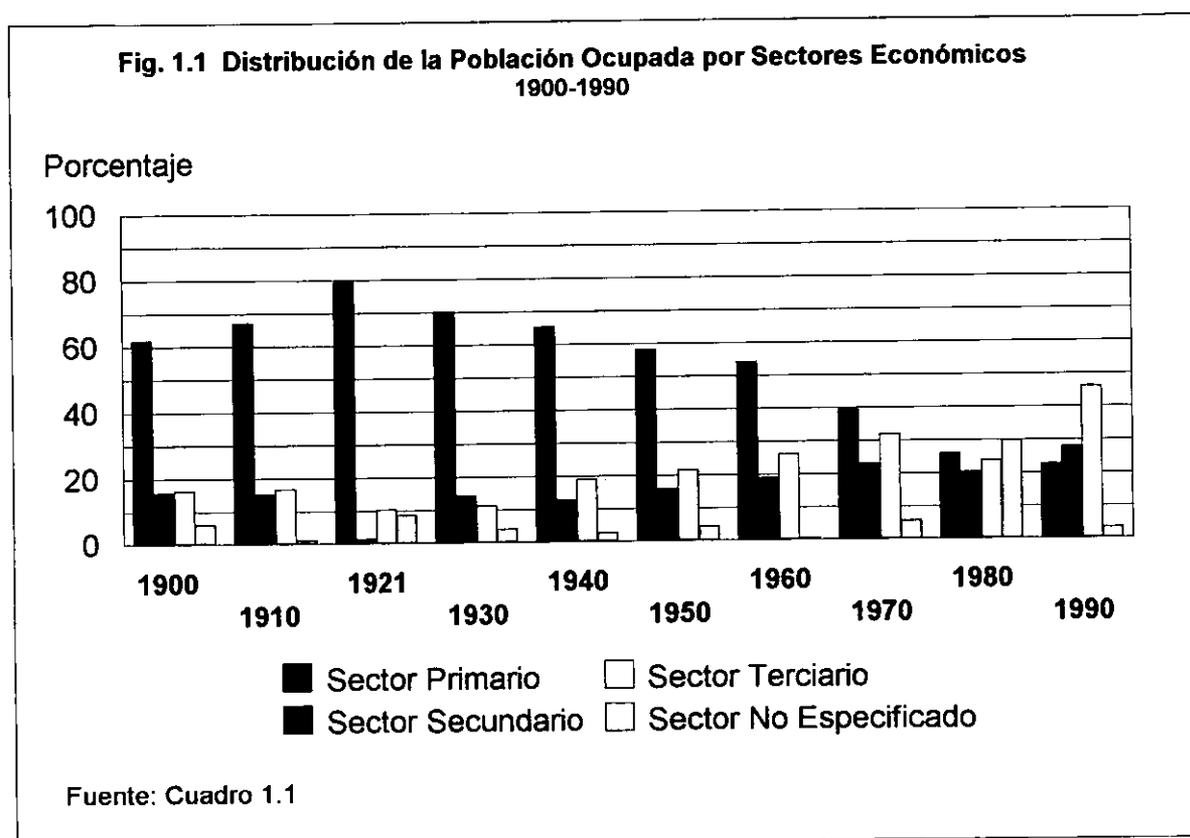
Año	Total	Sec. Primario	Sec. Secundario	Sec. Terciario	Sec. No Especificado
1900 2/	5,131,051	3,177,840	803,294	838,091	311,826
1910 2/	5,337,889	3,584,191	803,262	884,589	65,847
1921	4,383,561	3,488,102	61,318	454,293	379,848
1930	5,165,803	3,626,278	743,407	586,930	209,188
1940	5,858,116	3,830,871	746,313	1,117,274	163,658
1950	8,272,093	4,823,901	1,319,163	1,774,063	354,966
1960	11,332,016	6,143,540	2,147,343	2,959,342	81,791
1970	12,955,057	5,103,519	2,973,540	4,130,473	747,525
1980	21,941,693	5,699,971	4,464,410	5,225,275	6,552,037
1990	23,403,413	5,300,114	6,503,224	10,796,203	803,872

1/ Las agrupaciones se hicieron de la siguiente manera: sector primario: agricultura, ganadería, silvicultura, pesca y caza; sector secundario: industria del petróleo, industria extractiva, industria de la transformación, construcción, generación de energía eléctrica; sector terciario: comercio, transporte, servicios y gobierno.

2/ 1895, 1900 y 1910 se refieren a la población ocupada, concepto equivalente a la PEA.

Fuentes:

- 1900-1921: Estadísticas Históricas de Porfiriato. Fuerza de Trabajo y Actividad Económica por Sectores. El Colegio
1921-1990: INEGI, DGE. Censos Generales de Población y Vivienda.



Los cambios económicos y socioculturales, los patrones de consumo y la movilidad ocupacional, han modificado la composición ocupacional de la PEA, aunque de manera horizontal. La reducción de la mortalidad y la estabilidad de la natalidad determinaron que la tasa de crecimiento de la población, a partir de los 50, fuera mayor a 30%, originando su rejuvenecimiento.

En 1930 la proporción de menores de 15 años era de 39.2% y aumentó a 45% en 1960. Dicho rasgo, define las características de ingreso, salidas de la actividad y edades activas de la población. En este período, menos del 50% de la población tenía la función de productora y consumidora.

La Tasa Bruta de Actividad ($TBA = \frac{PEA}{Población\ Total} \times 100$), utilizada como un indicador de la proporción de la población empleada en la producción de bienes y servicios y como medida de la relación de dependencia, de 1900 a 1950 descendió de 33.4% a 32.4%. El rejuvenecimiento de la población explica el aumento de las relaciones de

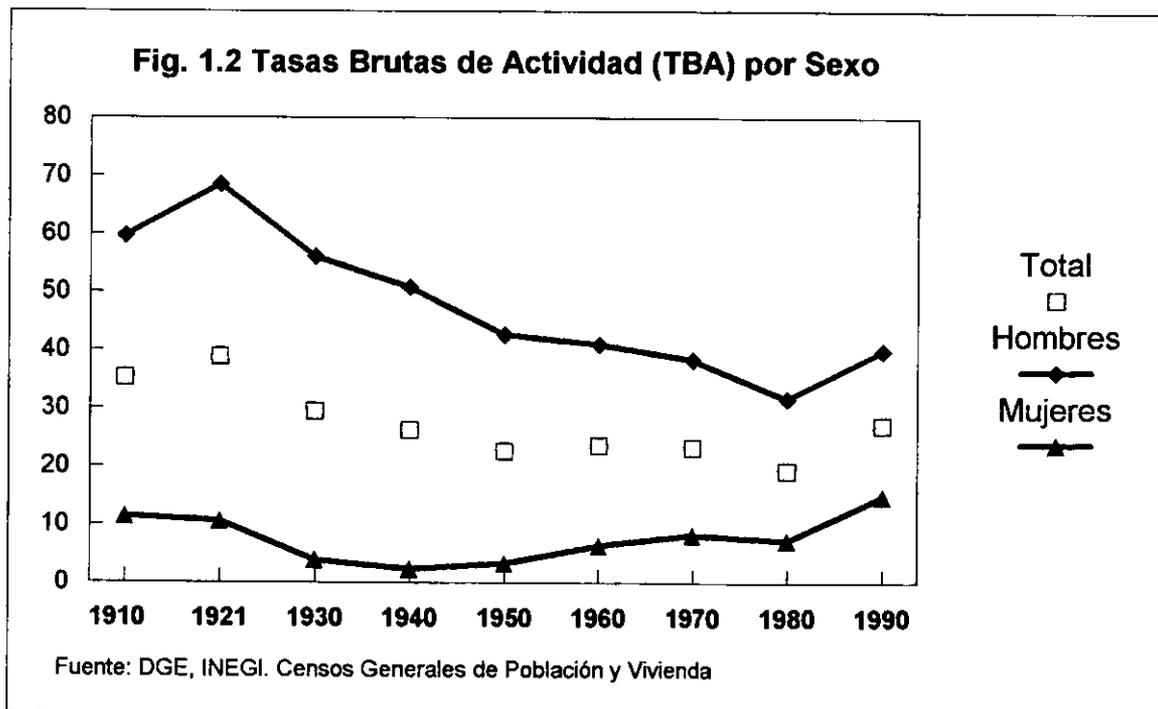
dependencia. En 1900, por cada trabajador dependían 1.9 personas; en 1960, la cifra de dependientes fue de 2.3 personas.

El nivel de participación de la población en la actividad económica de México se ha visto disminuido notablemente a partir de 1950. De acuerdo con información censal, en 1950 la TBA fue 32.3% y en 1970 sólo el 26.8% de los mexicanos que eran considerados como activos.

POBLACIÓN ECONÓMICAMENTE ACTIVA POR SEXO			
Año	Total	Hombres	Mujeres
1900	5,359,764	4,486,786	872,978
1910	5,581,293	4,802,734	778,559
1921	4,883,561	4,554,178	329,383
1930	5,165,803	4,926,228	239,575
1940	5,858,116	5,425,659	432,457
1950	8,272,093	7,144,872	1,127,221
1960	11,253,297	9,235,022	2,018,275
1970	12,955,057	10,488,800	2,466,257
1980	22,066,084	15,924,806	6,141,278
1990	55,913,847	27,084,182	28,829,665

Cuadro 1.2

Fuente: DGE, INEGI. Censos Generales de Población y Vivienda



La disminución de las tasas de participación y particularmente las masculinas, se considera como un indicador del desarrollo económico ya que en estudios comparativos en países desarrollados y subdesarrollados, sobre niveles de actividad, se ha encontrado una relación inversa con los niveles de desarrollo. El descenso se justifica como una respuesta a las demandas de calificación y eficiencia creadas por la sociedad industrial moderna. Se argumenta que los jóvenes necesitan mayor tiempo de entrenamiento y se posterga su entrada a la actividad; por otra parte, a medida que se consolidan los sistemas de seguridad social públicos y privados los trabajadores de mayor edad se retiran cada vez más pronto.

Es necesario tomar precauciones con respecto a las irregularidades de la información censal; En el censo de 1960 y 1980 se sobrestimó la población activa y por el contrario, en el censo de 1970 existió una subestimación del contingente de activos.

La población activa ubicada en el sector agropecuario, fue la de menor crecimiento en el periodo 1950-1970. Según información censal, sólo el 5.5% del total de personas que se incorporaron a la actividad en ese lapso fue absorbido por el sector primario de la economía, pero todavía en 1970 el 39.4% de los activos pertenecía a dicho grupo.

En lo que respecta a la participación femenina en el sector primario de la economía, se confirmó, al menos en 1970, que el mercado de trabajo agrícola no es de mano de obra femenina. La mujer mexicana participó más activamente en los procesos que se vinculaban al sector terciario de la economía, y en especial a los servicios.

Los sectores secundario y terciario participaron activamente en la absorción del crecimiento de la población económicamente activa femenina (PEAf) durante 1950-1970. Al final del periodo, el 22.9% de la PEAf se ubicaba en el sector secundario y el 31.9% en el terciario. Por su dinamismo en la incorporación de la PEAf destacan la industria de la transformación con el 30.9% y los servicios y gobierno con el 30.1%; dos ramas no agrícolas que absorbieron más del 60% del incremento de la PEAf en 1950-1970. El proceso de industrialización y sus servicios correspondientes, dan muestras evidentes de avance. Sin embargo, existen procesos paralelos que ponen de manifiesto la insuficiencia del crecimiento para absorber la PEAf (por ejemplo, el abultamiento de los servicios de consumo, esparcimiento, preparación y

venta de alimentos y bebidas, servicios domésticos, etc., donde es probable esperar la prevalencia del estrato artesanal y no del fabril).

Todavía en 1970 existían en México enormes desigualdades sectoriales, a las que habría que añadirse las regionales, y las repercusiones de todos estos procesos sobre la estructura social del país. Pero no es posible simplificar las transformaciones sucedidas y plantear la ocurrencia de una sobreterciarización de la economía mexicana.

1.3. Cambios en el periodo 1970-1990

La política económica adoptada después de la Segunda Guerra Mundial estuvo marcada por la fuerte inversión del gobierno federal en la industrialización de México además de la amplia protección de dicho sector, el desarrollo del mercado interno y la sustitución de la importación de insumos. Estas estrategias permitieron un rápido crecimiento económico que culminó en la década de los 70, para que en el periodo 1980-1990 con la apertura de mercados, desarrollo de la capacidad exportadora para una economía globalizada, reestructuración del aparato productivo y la reducción de la participación del gobierno mexicano, iniciaran los periodos de crisis y recesión económica; el combate a la inflación se vuelve la parte central de las políticas económicas. Los salarios reales son afectados de manera negativa además de la disminución de la distribución salarial. La crisis económica de 1981 causada por el aumento del endeudamiento interno y externo de los sectores públicos y privados marca un cambio radical en la estrategia de desarrollo de México. La inestabilidad de la balanza de pagos motiva constantes devaluaciones del peso mexicano por lo que se opta por un deslizamiento constante del tipo de cambio.

Debido a las altas tasas de crecimiento de la población mexicana en el periodo 1950-1970, la población de 12 años y más experimento un aumento considerable y por consiguiente, una mayor oferta de la fuerza de trabajo en contraposición con el sector productivo, incapaz de absorberla en ocupaciones formales protegidas por la seguridad social.

El sector agrícola, ha experimentado un estancamiento en su desarrollo a partir de mediados 60 y el cual los subsecuentes gobiernos no han podido reactivar. La expulsión de la fuerza de trabajo proveniente del campo disminuye en los 80 ocasionada por la crisis de los mercados urbanos de empleo. La presión de la demanda de trabajo tiene como vía alternativa de escape la migración a Estados Unidos, que anteriormente se componía principalmente por población de áreas agrícolas, pero en últimos años ha sido mayor la presencia de la población de centros urbanos.

La mayor aceptación cultural de la participación de la mujer en actividades fuera del hogar; la elevación del nivel educativo de la población femenina; el aumento de la demanda de la fuerza de trabajo femenina para el desempeño de actividades como la maquila, la agroexportación y los servicios en general y el deterioro de los niveles de vida de un gran número de familias como resultado de las inadecuadas políticas económicas aplicadas por el gobierno mexicano dieron como resultado el aumento de la participación de la fuerza de trabajo femenina.

Por otro lado, se observa un racionamiento en la función de generador de empleos ya sea directa o indirectamente por parte del gobierno. La caída de los salarios ocurrida a partir de 1976; el deterioro de la condiciones de trabajo en las renegociaciones de los contratos colectivos, en varias ramas de la actividad económica en la década de los 80's y la imposición de topes al aumento de los salarios contribuyeron a la reducción de los ingresos reales de la familias cuyos miembros se desenvolvían en el sector formal de la economía. La conjunción de estos factores obligaron al rápido crecimiento de las actividades en el sector informal como mecanismo de supervivencia de la población, esto se traduce en el incremento de los trabajos por cuenta propia y relaciones laborales no remuneradas de trabajos familiares principalmente en jóvenes, que no pudieron ser absorbidos por el aparato productivo o educativo a pesar del aumento del presupuesto federal aplicado a este último rubro; mujeres, pensionados y jubilados que contemplaron un aumento sostenido por las políticas de retiro voluntario y temprano aplicadas en los sectores público y privado.

La tendencias derivadas de la combinación de los factores antes descritos se manifiesta de la siguientes forma:

- El aumento de la tasa de participación femenina en las edades jóvenes y medias implica un aumento en la población femenina asalariada pero ocupada en las actividades formales con las retribuciones más bajas.
- El incremento en las actividades en el sector informal contiene la disminución de las tasas de participación en las edades extremas.
- La mayor importancia de las estructuras de empleo de las características de la ocupación urbana, y en especial en el sector terciario, por el crecimiento de la población económicamente activa urbana.
- Un aumento del nivel medio de escolaridad como de la población ocupada con niveles técnicos y profesionales de educación.
- El comercio y los servicios toman una mayor importancia en el aspecto ocupacional, un estancamiento en el sector industrial, principalmente en el sector manufacturero, una reducción en el empleo agropecuario y en el sector gubernamental.
- Incremento en el número de trabajadores pobres y en el promedio de trabajadores por hogar.

1.4. El mercado laboral a nivel nacional

El desarrollo económico en México nunca ha sido heterogéneo. La industrialización se ha estado acompañado de una urbanización acelerada, incorporación de servicios modernos, mejoramiento de los niveles de bienestar para su fuerza laboral y una mayor inversión en la educación y salud. Por otro lado, en las regiones donde el trabajo es predominantemente agrícola existen altos niveles de marginación. La participación en la trabajo agropecuario no sería un indicador negativo para el desarrollo del mercado laboral si la producción agropecuaria fuera lucrativa y eficiente como en los países desarrollados. Estos argumentos dan respuesta al cuestionamiento de los contrastantes niveles de bienestar y desarrollo económico en los estados altamente industrializados y en los estados donde la mayoría de su población se

ocupa en actividades agropecuarias y sus progresos en la industria y el comercio han sido casi nulos.

Jusidman y Eternod¹ proponen una clasificación en seis grupos para las 32 entidades de la República Mexicana de acuerdo a 17 indicadores del mercado laboral. Los primeros diez indicadores tienen una relación directamente proporcional al desarrollo del mercado laboral:

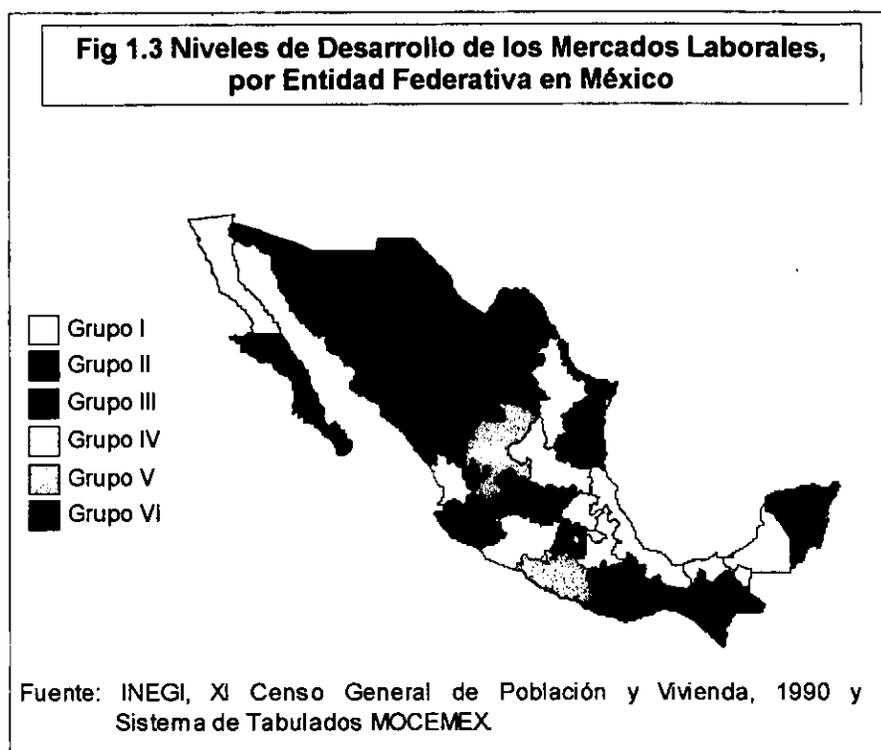
1. Tasa de Participación Bruta Total.
2. Tasa de Participación Bruta Femenina.
3. Tasa de Población Ocupada con Educación Profesional Terminada, Sumando a la que tiene Posgrado.
4. Tasa de Población Ocupada como Profesionales y Técnicos.
5. Tasa de Población Ocupada en la Industria Manufacturera.
6. Tasa de Población Ocupada en el Comercio.
7. Tasa de Población Ocupada en Servicios Financieros y en Servicios Profesionales y Técnicos.
8. Tasa de Población Ocupada como Obreros y Empleados.
9. Número de Años de Escolaridad Promedio de la Población Económicamente Activa.
10. Tasa de Población Ocupada con más de Cinco Salarios Mínimos.

Los restantes siete indicadores mantienen una relación inversamente proporcional al desarrollo del mercado laboral: Tasa de Participación de la Población de 12 a 14 años.

1. Tasa de Participación de la Población de 55 años y más.
2. Tasa de Población Ocupada sin Instrucción.
3. Tasa de Población Ocupada en Actividades Agropecuarias.
4. Tasa de Población Ocupada con menos de un Salario Mínimo.
5. Tasa de Población Ocupada como Trabajadores Familiares no Remunerados.
6. Tasa de Población Ocupada como Trabajadores por cuenta Propia.

La selección de estos indicadores se basa en las tendencias del empleo en México y posiblemente estos indicadores no mostrarían el desarrollo laboral en otros países.

¹ Jusidman, Clara; M. Eternod. *La Participación de la Población en la Actividad Económica en México*.



Grupo I: Distrito Federal, Nuevo León y Baja California

Estos tres estados presentan las condiciones más favorables para la inserción de la fuerza laboral además de ostentar los más altos niveles en los indicadores de participación bruta total y femenina; tasas de población ocupada en actividades profesionales, técnicas, servicios financieros y la mayor participación de obreros y empleados. Los niveles medios de escolaridad son los más altos y cuentan con las mayores tasa de ocupados con educación profesional y con posgrado.

Grupo II: Coahuila, México, Sonora, Chihuahua, Tamaulipas, Jalisco, Baja California Sur y Aguascalientes

Varios de estos estados son fronterizos o con áreas metropolitanas; en estas entidades se ha manifestado un desarrollo industrial, en gran medida, relacionado con

la maquila; actividades comerciales o de turismo importante o agricultura comercial moderna.

Grupo III: Querétaro, Morelos, Colima, Quintana Roo, Sinaloa, Yucatán, Guanajuato y Durango

El rasgo común de este conjunto de estados es la heterogeneidad de sus estructuras económicas, se localizan unidades agrícolas de muy pequeña escala y baja productividad en contraposición de desarrollos modernos en agricultura comercial y servicios turísticos de alta calidad.

Grupo IV: Tlaxcala, San Luis Potosí, Nayarit, Tabasco, Campeche, Puebla, Michoacán, Hidalgo y Veracruz

Son estados preponderantemente agrícolas, en varios casos con agricultura y ganadería de trópico, con extensos sectores de población campesina empobrecida. A pesar de observarse un importante desarrollo petrolero, turístico, pesquero, industrial y minero los beneficios no han cubierto al total de la población de estas regiones.

Grupo V: Zacatecas y Guerrero

Dos estados empobrecidos por la baja productividad de sus recursos naturales que limitan su desarrollo y la presencia de algunas actividades empresariales modernas no es suficiente para detener el flujo migratorio de sus habitantes hacia otros estados de la República Mexicana y a Estados Unidos.

Grupo VI: Oaxaca y Chiapas

La mayor parte de la población de los estados de este grupo se dedica a la agricultura, principalmente de pequeños predios agrícolas, con algún desarrollo de la agricultura de plantación para exportación. El comercio y la industria carecen del apoyo de servicios modernos que les permita superar sus bajos niveles de desarrollo. Su fuerza laboral indica los niveles más bajos de escolaridad e ingreso.

2. Aspectos Generales en la Simulación

2.1. Características principales

El inicio de los métodos de simulación por computadora es atribuido a los trabajos realizados por Von Neumann y Ulam, durante la Segunda Guerra Mundial, cuando utilizaron el término de métodos de Monte Carlo como una palabra clave para designar una técnica matemática que utilizaban en las investigaciones secretas realizadas en Los Alamos, Nuevo Mexico. El nombre proviene de los casinos de juego de la ciudad de Monte Carlo en Mónaco. Sus investigaciones estaban relacionadas con una simulación directa del comportamiento aleatorio de la difusión de neutrones en materiales fisionables.

Tiempo después los métodos de Monte Carlo fueron utilizados para evaluar integrales multidimensionales y para resolver ecuaciones integrales. Los métodos se utilizaron para resolver problemas que por su naturaleza eran muy costosos para ser tratados de manera experimental o sumamente complejos para ser resueltos analíticamente.

Con el acelerado desarrollo de las computadoras a fines de la década de los 50 fue posible implementar nuevas aplicaciones que no se limitaban a problemas de física o ingeniería sino también se logró experimentar con modelos matemáticos simulados en ciencias biológicas, medicas y disciplinas humanísticas como la sociología y la economía.

R.E. Shannon¹ define a la simulación como: "El proceso de diseñar un modelo de un sistema real y realizar un experimento con él para entender el comportamiento del sistema o evaluar varias estrategias (dentro de los límites impuestos por un conjunto de criterios) para la operación del sistema".

La definición de simulación de F. Hillier² es: "Una técnica de muestreo estadístico controlado para estimar el desempeño de sistemas estocásticos complejos cuando los modelos analíticos no son suficientes. El modelo de simulación describe la operación de un sistema en términos de eventos individuales de cada uno de los componentes que lo conforman. La simulación es la técnica de la realización de experimentos de muestreo sobre el modelo del sistema. Los experimentos se llevan cabo en el modelo en lugar de realizarlos en propio sistema, porque esto último sería inconveniente, costoso y muy tardado. Por esta razón los experimentos de simulación deben considerarse iguales a los experimentos estadísticos comunes y se basan en los fundamentos formales de la Teoría Estadística."

Una definición propia de simulación en sistemas sociales es: "Un proceso basado en información estadística, teorías de probabilidad y el uso de sistemas computacionales, en el cual se pueden crear varios escenarios para observar el comportamiento del sistema o modelo y probar hipótesis de investigación."

La predicción del futuro ha sido la causa principal para el uso de la simulación. La modelación de sistemas por simulación es incapaz de generar soluciones por si misma como los modelos analíticos; sólo puede servir como una herramienta para el análisis del comportamiento de un sistema en condiciones especificados por el diseño del sistema.

¹ Shannon, Robert E. *System Simulation: The Art and Science*.

² Hillier, Frederick J., G. J. Lieberman. *Introduction to Operation Research*.

Para usar la simulación es necesario verificar si existe una o más de las siguientes condiciones:

- Si aún no se han desarrollado métodos analíticos para resolver el modelo matemático o no existe una formulación matemática completa del problema.
- Existen soluciones analíticas y son posibles pero no están al alcance de la habilidad matemática del personal disponible. Los costos del diseño, prueba y corrida de la simulación se deben comparar con el costo de una asesoría externa.
- Se desea observar una simulación histórica de un proceso en un periodo de tiempo además de la estimación de ciertos parámetros.
- La simulación puede ser la única opción posible por la dificultad en el manejo experimental y observación del fenómeno en su medio ambiente actual, como los estudios de vehículos espaciales y vuelos interplanetarios.
- Si se busca reducir el tiempo requerido por sistemas o procesos en periodos de larga duración, la simulación proporciona un completo control sobre el tiempo, puesto que se puede aumentar o reducir la velocidad del comportamiento de un fenómeno.

2.2. Elementos de la modelación de sistemas

La simpleza de su funcionamiento y un comportamiento que se asemeje a la realidad son las principales características deseables para modelos científicos simuladores de sistemas reales. Los modelos matemáticos de sistemas económicos constan de los tres elementos: componentes, variables y parámetros y relaciones funcionales.

Los componentes de los modelos económicos tienen la tendencia a variar ampliamente, pues depende del sistema económico simulado. Se pueden citar a los sectores gubernamentales, industrias de calzado, automotriz, transporte, etc. como ejemplo de componentes.

Las variables que aparecen en los modelos económicos se emplean para relacionar un componente con otro y se dividen en variables exógenas, variables de estado y variables endógenas. Las variables exógenas son las independientes o de entrada del modelo y bajo el supuesto de que han sido predeterminadas y son independientes del

sistema modelado, estas variables actúan sobre el sistema pero no son afectadas por él. Las variables exógenas se dividen en controlables y no controlables.

Las variables de estado describen la situación del sistema o de alguna de sus partes en un determinado instante en un periodo. Las variables endógenas son las independientes o de salida del sistema y son generadas por la interacción de las variables exógenas con las de estudio.

Las relaciones funcionales son las identidades y las características de operación. Las identidades son las definiciones formales relacionadas con las componentes, es decir una definición de utilidad puede ser diferente para una compañía sin fines de lucro que para una compañía comercial. Las características de operación son ecuaciones matemáticas que relacionan a las variables endógenas y variables de estado del sistema con sus variables exógenas. Al contrario de los componentes y las variables observables directamente en un sistema real, los parámetros de las características de operación se derivan sobre la base de la inferencia estadística.

2.3. Clasificación de los modelos de simulación

Modelos determinísticos. En este tipo de modelos las variables endógenas y exógenas no son variables estocásticas y las características de operación están regidas por relaciones exactas, en la cuales se asumen situaciones de completa certeza. Los procesos en serie, tareas en un taller y ciertos tipos inventarios son ejemplos de modelos determinísticos.

Modelos estocásticos. Se presentan cuando alguna de las características de operación esta determinada por una función de probabilidad. Los modelos estocásticos son de funcionamiento complejo y es muy difícil analizarlos por medio de técnicas analíticas, de ahí la importancia para la simulación de estos modelos.

Modelos estáticos. La mayoría de los problemas que se resuelven con técnicas de investigación de operaciones como la programación lineal, análisis de redes y teoría de juegos no se toma en cuenta la variable tiempo y este es el rasgo característico de los modelo estáticos.

Modelos dinámicos. Los modelos matemáticos que tratan de las interacciones varían con el tiempo, se denominan modelos dinámicos.

2.4. Desventajas de la simulación

1. Puede interferir en el funcionamiento de un sistema. Si las personas son una parte integral del sistema, por el hecho de estar siendo observada, pueden modificar su comportamiento habitual.
2. Puede ser difícil mantener las mismas condiciones de operación cada corrida o replicación del modelo. El tiempo consumido y el costo puede ser mayor para obtener para obtener el tamaño de muestra y por lo tanto la significancia estadística.
3. No permite explorar todas las posibles alternativas de experimentación en la realidad.
4. El desarrollo de una buena simulación es a menudo costoso e implica un gran consumo de tiempo además de requerir personal altamente calificado con el que muchas veces no se cuenta.
5. Una simulación puede aparecer como el reflejo preciso del mundo real cuando en verdad no lo es. Varios problemas intrínsecos de la simulación hacen caer en resultados erróneos si no son corregidos correctamente.
6. La simulación es imprecisa y no se puede medir el grado de imprecisión. El análisis de sensibilidad del modelo para cambiar algunos valores parametrales pueden resolver parcialmente esta dificultad. Los resultados de la simulación son generalmente numéricos y dados con la exactitud que el experimentador decida, por esta razón surge el peligro de atribuirle a los números un alto grado de validez.

2.5. Ventajas de la simulación

1. Permite visualizar cambios ocurridos en un sistema por modificaciones operativas, ambientales, organizacionales o informativas.
2. Posibilita la experimentación de situaciones en las cuales no se tiene suficiente información y preparase para cuando ocurran eventualidades de este tipo.

3. En ciertos problemas de tipo estocástico, la simulación puede sustituir la información de los valores esperados y momentos, ante la insuficiencia de datos para describir un proceso.
4. Las soluciones analíticas puede verificarse utilizando la simulación.

2.6. Diseño de un modelo de simulación

Formulación del problema. Se deben definir claramente los objetivos de la investigación antes de la planeación de un experimento de simulación. Los objetivos de la investigación pueden ser el rechazo o aceptación de una o varias hipótesis relativas al comportamiento del sistema bajo estudio; la estimación de los efectos en cambios en los parámetros, en las características de operación o en las variables exógenas del sistema.

Recolección y procesamiento de datos. La información descriptiva la constituyen los datos, referentes al sistema a simular, son una parte integral para la formulación previa de un problema. La reducción significativa de los datos pueden sugerir hipótesis de cierta validez, las cuales se utilizaran en el diseño de modelos matemáticos para conformar el sistema a simular. Los datos permiten sugerir refinamientos o mejoras en los modelos matemáticos en la simulación. Los datos, reducidos a una forma final, son utilizados para estimar los parámetros de las características operativas relacionadas con las variables exógenas, endógenas y de estado del sistema. Sin la existencia de los datos sería imposible probar la validez de un modelo para su simulación.

La recolección de datos es el proceso de captación de los hechos disponibles para hacerlos alcanzables cuando su uso sea necesario. El procedimiento de recolección y almacenamiento de datos ocurre simultáneamente, pues el primero implica que sean o hayan sido almacenados. Las fuentes de datos comprenden censos, encuestas, mapas, estudios realizados con anterioridad relativos al tema a investigar. La recolección y almacenamiento de datos implica la identificación, edición, codificación, transcripción y verificación de ellos. La conversión de los datos consiste en la transformación de medios generalmente escritos a formas que faciliten su manejo de acuerdo con los fines que se les tenga destinados.

Estimación de los parámetros de las características operacionales a partir de datos reales. Después de la recolección de datos el paso siguiente es la estimación de los valores de los parámetros del modelo. En los modelos económicos dicha estimación se hacen con técnicas econométricas.

Evaluación del modelo y de los parámetros estimado. La verificación de la validez del modelo que se ha desarrollado es un paso necesario para continuar con el experimento. Esto se logra con la comparación del comportamiento del modelo con datos reales históricos para probar su capacidad predictiva. De acuerdo con las características que se desean evaluar, se utilizan pruebas estadísticas para las medias, varianzas y conteo de datos.

En este punto se debe hacer un análisis detallado de las variables que pudieran omitirse o la inclusión de nuevas variables además de la correcta definición de las variables funcionales.

Formulación de un programa por computadora. La computadora depende de la definición de variables, relaciones funcionales y cálculo de parámetros. Si existe deficiencias al realizar estas tareas, ni la computadora más potente que exista podrá corregir el funcionamiento del modelo. Existe diferentes tipos de lenguajes de programación tales como DYNAMO, SIMSCRIPT, GPSS, SIMULA que facilitan la labor de programación además de paquetes de computo para cálculos econométricos, que permiten realizar simulaciones de manera muy sencilla pero con estos paquetes se corre el peligro del uso indiscriminado de simulaciones antes de evaluar otras opciones e inclusive sea utilizado por personal sin el conocimiento necesario en el tema que quiera analizar. La elección de un lenguaje de programación o paquete de computo depende de las propias características del modelo.

Validación. La evaluación de la eficacia del funcionamiento del modelo se hace con la comparación de datos históricos, si es que se cuenta con ellos, con los datos simulados por las variables endógenas. La revisión de estudios relacionados con las predicciones del modelo simulado son de gran ayuda para lograr una validación aceptable.

Análisis de los datos simulados. El último paso en la simulación consiste de cuatro aspectos:

- Recolección y procesamiento de datos.
- Cálculo de estadísticas de prueba.
- Interpretación de datos.
- Conclusiones generales.

2.7. Microsimulación

En un modelo demográfico cada individuo en cierto grupo tiene las opciones de morir, casarse, divorciarse, procrear un hijo, incapacitarse para un empleo. Las probabilidades de que ocurran estos eventos se obtienen a partir de la información a través de los censos de población y otras investigaciones. Las probabilidades difieren entre cada grupo; por ejemplo los individuos de ocho años tienen mayor riesgo de morir que los de trece años .

Un modelo de microsimulación, en la cual la unidad de estudio es el individuo, la familia o el hogar y no sus agregados, tiene pocas posibilidades de producir inconsistencias internas en las variables demográficas usadas para describirlo. Mediante la microsimulación es posible introducir distribuciones de probabilidad entre las componentes demográficas del modelo de simulación, lo cual es una característica difícilmente factible en la simulación de los agregados demográficos.

Una breve descripción de un proceso de la microsimulación es la siguiente:

Se genera un número aleatorio U_{ij} para verificar si un evento del tipo i (por ejemplo la muerte) sucede al individuo j ; p_i denota la probabilidad del evento i para la subpoblación a la cual pertenece este individuo (p_i denota la probabilidad de morir en el grupo de edad i). Cuando $U_{ij} < p_i$, el evento ocurre (el individuo j muere). Es claro que la simulación debe permanecer consistente: un individuo que muere debe permanecer muerto; un individuo que no está casado no puede divorciarse, etcétera.

Los modelos completos de simulación pueden consistir de submodelos o subsistemas que representen el desarrollo demográfico, oferta de mano de obra, impuestos y pagos de la seguridad social, ingresos y gastos de las familias. En la mayoría de los casos

estos submodelos un sistema de bloque recursivo, es decir un submodelo esta influenciado por submodelos previos pero no por submodelos sucesivos.

2.8. Tipología de la microsimulación

Existen distintos tipos de microsimulaciones. Una primera clasificación depende del tipo de datos. Si los datos disponibles son empíricos se denomina microsimulación empírica. En la microsimulación existen hipotéticamente microunidades (por ejemplo, un hogar representativo), y para esa microunidad se utilizan datos estándares. En la microsimulación estática, solamente los efectos inmediatos y de primer orden de una medición se calculan y no es posible simular en el tiempo la retroalimentación. La microsimulación dinámica se divide en longitudinal y transversal. En la microsimulación longitudinal, la dinámica es tomada en cuenta (vida histórica de cada microunidad es simulada en cada corrida) pero las interacciones entre las microunidades no son consideradas. Debido a que las interacciones entre las microunidades y los efectos del desarrollo económico son totalmente ignorados los resultados de este tipo de microsimulación son totalmente tentativos. En la microsimulación transversal, se consideran las interacciones entre las microunidades. Dentro de cada unidad de tiempo, todas las microunidades son procesadas y al final de cada período las interacciones han ocurrido. Esto produce el desarrollo por periodos y una muestra hipotética que se utilizará en el siguiente periodo.

3. Diseño del Modelo de Simulación

3.1. La utilización de la microsimulación

Una de las características más importantes de los sistemas socioeconómicos, es su total independencia en la toma de decisiones. Individuos, hogares (familias) toman decisiones importantes relacionadas con la educación, selección de la ocupación y lugar de residencia, tenencia de la vivienda, gastos y ahorro, inversión y deudas, matrimonio y paternidad. Las empresas deciden la composición y nivel de producción, expansión, inversión, locaciones, investigación y recursos humanos. A las unidades gubernamentales federales, estatales y locales les corresponden decidir a largo plazo en los asuntos de educación, defensa nacional y seguridad pública, aprovechamiento de los recursos naturales, ingresos públicos e impuestos, emisión de dinero, propiedades y deudas gubernamentales.

El tomador de decisiones debe contar con las mejores herramientas que le permitan conocer las posibles consecuencias, derivadas de los cursos de acción tomados o no,

predicciones serían más exitosas si estuvieran basadas en el conocimiento de las unidades elementales de decisión, su comportamiento y su respuesta a los cambios en su entorno.

Los sistemas económicos son sistemas bastante complicados con millones de componentes interactuantes. Dichas componentes; individuos, hogares, empresas y unidades gubernamentales, llevan a cabo acciones influyentes a otras unidades que de igual manera modifican el comportamiento de otras componentes.

En un principio, el microanálisis consistía sólo en la resolución de implicaciones teóricas de hipótesis sin comprobación acerca del comportamiento de unidades económicas. Era imposible la prueba de esas hipótesis porque los únicos datos disponibles estaban estimados de agregados económicos recolectados de varias fuentes de tal forma que se perdían todas las conexiones identificables entre hogares e individuos. La estimación de agregados de ingreso no pueden ser usados para probar las hipótesis de las relaciones entre consumo de los hogares a menos que se piense que todos los hogares tienen una respuesta homogénea a cualquier nivel de ingresos. Cuando se trata de probar hipótesis más complicadas, como la influencia del gasto de los hogares en el ahorro, deudas, propiedad de la vivienda, número de hijos, ocupación, retiro laboral, es imposible apoyarse en series agregadas. Algo que puede afirmarse de las series agregadas es que mientras no conduzcan a nuevas investigaciones de componentes básicas en sistemas económicas como los hogares pueden ser utilizadas para obtener información del gran agregado del sector de los hogares.

Para investigar el comportamiento e interacción de los componentes de un sistema económico es necesario medir un substancial número de variables en un periodo sucesivo de tiempo para el gran número de hogares y sus miembros y las empresas. Además debe ser posible el rápido acopio de datos¹ en respuesta a los cambios ocurridos en la investigación.

¹ La recolección de datos implica costos y la única técnica conocida para obtener la información necesaria a un costo razonable es el muestreo por encuesta diseñado y ejecutado científicamente.

para poder llegar a decisiones racionales ejecutadas eficientemente tanto en los sectores gubernamentales como en los privados.

Las decisiones que buscan la racionalidad siempre están basadas en predicciones. El objeto de una decisión es influenciar de cierta manera el futuro, cualquier decisión no puramente aleatoria debe estar basada en un juicio del curso probable de los eventos en ausencia de acción y el probable efecto de las acciones que pueden ser tomadas. Las predicciones sólo se pueden basar en las relaciones entre los elementos y variables involucrados, conocidas a través de las experiencias del pasado y presente. Este conocimiento de las relaciones se adquiere por la estrategia seguida de acuerdo con el método científico, consistente en la acumulación de experiencia, formulación de hipótesis de las relaciones entre variables, y después confrontar esas hipótesis con la nueva experiencia (prueba). Si se pudiera experimentar en la economía y entonces hacer cambios intencionales en ciertas variables para observar que sucede con las otras, sería de gran ayuda pero esto es casi imposible.

Tradicionalmente, la mayoría de las predicciones en la economía han estado basadas en las relaciones estimadas entre agregados económicos y su cambios en el tiempo. Las predicciones basadas en las relaciones estimadas por series de tiempo económicas no han sido muy exitosas, son conocidas las dificultades estadísticas para probar las hipótesis de las relaciones obtenidas por medio de las series de tiempo económicas. El problema surge de la naturaleza propia de las series generadas en la economía. La carencia de independencia entre las observaciones sucesivas, multicolinealidad, relaciones simultáneas y de retroalimentación entre las variables, errores autocorrelacionados, errores de medición, omisión de datos y observaciones escasas; todos estos factores combinados hacen difícil el manejo con las herramientas convencionales de la inferencia estadística.

Algunas de esas dificultades estadísticas pueden ser disminuidas con una mejor fuente de datos o nuevas herramientas de inferencia estadística, pero siguen existiendo problemas esenciales inherentes a la naturaleza propia de la economía. Los sistemas socioeconómicos son una complicada estructura con millones de unidades interactivas como individuos, hogares, compañías. Es razonable pensar que las

Si se desea la dirección de las consecuencias de los cursos de acción alternativos, entonces las predicciones condicionales son necesarias y los modelos de autocontenidos tendrán poca relevancia.

La mayoría de los modelos no son autocontenidos o sistemas cerrados; las salidas generadas por ellos dependen de insumos exógenos. Con dichos modelos sólo se puede producir predicciones condicionales.

Los tomadores de decisiones especialmente los que se desenvuelven en sectores gubernamentales y empresariales, por necesidad deben restringir mucha de su atención a la predicción de información agregada. Los agregados que se necesitan, varían de acuerdo con el tipo de unidad tomadora de decisiones (utd) que las utilice así como el tipo de decisión que se este seleccionando. Sin embargo, solamente un pequeño número de factores pueden ser considerados en relación con cualquier decisión particular.

El resultado de esta restricción es que las utd's con varios propósitos deben basar sus decisiones en la predicción de fenómenos altamente agregados como los ingresos gubernamentales federales, estatales, ventas de varios artículos, inversiones, niveles de precios, nacimientos, muertes, matrimonios; o en predicciones de elementos ligeramente menos agregados como la distribución conjunta de la población clasificada por sexo, edad, ingresos, estado civil, localidad o condición laboral.

Existe una dificultad inherente pero, no una imposibilidad práctica en la agregación de cualquier factor en la más simple relación de las utd's en comprensibles relaciones entre grandes unidades agregadas como las empresas, el sector de los hogares y el gubernamental.

Una mayor dificultad en tal agregación no puede ser hecha correctamente sin un razonable modelo proveniente del mismo sistema económico en términos de la conducta e interacción de las utd's. Es entonces, cuando se posibilitan las formas en que podrían encontrarse las relaciones de agregación sin una pérdida importante de la precisión de la representación.

La agregación de las relaciones entre las unidades tomadoras de decisión es bastante fácil si las relaciones a agregar son lineales. Bajo estas circunstancias, la agregación

Para usar efectivamente el conocimiento de las unidades tomadoras de decisiones, deben estar integradas en un sistema o modelo. Los modelos de comportamiento son sistemas integrados por hipótesis, teorías o postulados de la conducta de organismos o sistemas de organismos, además instrumentos de predicción y utilizados para simular respuestas y otras conductas de las unidades o sistemas de unidades en investigación.

Los elementos individuales de un modelo, las hipótesis, teorías o postulados incorporados en él, pueden ser elementos predictivos por su propio derecho. Sin embargo, el rango útil de predicción de las hipótesis individuales es a menudo restringido porque las aplicaciones individuales apropiadas probablemente requieran que los valores de ciertas variables sean disponibles y que los valores de otras ciertas variables se mantengan constantes. Pero en la realidad no se puede contar con tales condiciones deseables.

Desde el punto de vista de la investigación, los modelos son de gran importancia porque incrementan el rango de posibles predicciones y así sugieren y hacen posibles pruebas más extensivas de las hipótesis y teorías que involucran.

Desde el punto de vista de una política a seguir, los modelos son esenciales porque el diseñador de las políticas usualmente requiere de predicciones de la combinación de efectos de varios factores actuando sobre una gran cantidad de unidades.

Además de la propia utilidad como elementos predictivos, los modelos pueden tener un valor explicativo. Tal valor explicativo puede ser logrado uniendo el conocimiento del comportamiento en un nivel de observación de las más elementales hipótesis, teorías o postulados.

Algunos modelos están autocontenidos en el sentido de que una vez iniciados generan salidas o predicen comportamientos sin la necesidad de insumos adicionales de información. Si se desean pronósticos incondicionales, tal modelo es el ideal si se hace una simulación exacta de la manera en que sus componentes pudieran modelar su funcionamiento.

de la actividad económica sumado a la acumulación de datos estadísticos conducirá a la estimación de características operacionales además de la subsecuente alineación de la simulación con las distribuciones observadas y cambios socioeconómicos experimentados de una muestra de la población total que conforma a la República Mexicana.

3.4. Movilidad de la fuerza laboral

El interés de simular el comportamiento de los individuos en el tiempo con respecto a su participación en la PEA, debe tomar en cuenta los movimientos de entrada y de salida de la actividad laboral. Tales movimientos están asociados con procesos demográficos y en el entorno, como el envejecimiento, matrimonio, natalidad, morbilidad y mortalidad.

La entrada de jóvenes trabajadores a la PEA_o y los retiros por enfermedad o edad avanzada son las mayores fuentes de cambio en la composición de la PEA. Este cambio de personal intergeneracional cuenta con millones de ingresos y retiros a la fuerza laboral anualmente. Para el caso de la PEA femenina, los retiros y reingresos dependen de la condición de actividad de se modifica de acuerdo con sus responsabilidades familiares. Como resultado de estos movimiento, el número total de personas pertenecientes a la PEA_o pueden ser varios miles mayor en cierto periodo de tiempo durante un año. El cambio de los niveles de salarios inducen cambio en la cantidad próxima de trabajadores de una población.

3.5. Resultados producidos

Este modelo generará cuatro posibles resultados: Primero, individuos dentro PEA_d que pueden entrar a la PEA_o. La probabilidad de ocurrencia se designará por E_{it} donde el subíndice i indica una combinación particular de las características de un individuo y circunstancias ambientales; el subíndice t se refiere a un periodo en particular. Segundo, individuos dentro de la PEA_d y que pueden permanecer en este estado. La probabilidad asociada es $1-E_{it}$. Tercero, individuos dentro de la PEA_o y pueden

puede ser útil si un limitado número de variables aparece una y otra vez. Pero si existen relaciones no lineales, entonces las relaciones estables en un micronivel son totalmente consistentes con la ausencia de las relaciones estables en niveles agregados.

Las predicciones de los agregados son necesarias pero deben obtenerse por la agregación del comportamiento de las unidades elementales en lugar de intentar agregar las relaciones de comportamiento de esas unidades elementales. Esto es, los agregados deben obtenerse de las evaluaciones de sistemas reales de una manera análoga a la forma en que los censos de población o encuestas obtiene los agregados asociados a los sistemas socioeconómicos reales.

3.2. Objetivos del modelo

1. Generar información desagregada a niveles que permitan conocer la evolución de las características demográficas y económicas de cada individuo dentro de la población durante el proceso de simulación.
2. Implementar el modelo de ocupación para la PEA para hacer posible la prueba de tres hipótesis utilizando como insumo principal de información la tasa de desempleo abierto (TDA), tasa de desempleo abierto y ocupación informal (TDAOI), tasa de ingresos insuficientes y desocupación (TIID).

3.3. PEA como un criterio de la actividad económica

La población ocupada dentro de la PEA no es un indicador refinado pero si un indicador operacional de la actividad económica individual de la actividad económica durante un periodo de tiempo². Pertenecer a la PEA significa estar empleado o desempleado. Los dos posibles estados de la actividad laboral varían continuamente en varias dimensiones. Estar dentro o no de la PEA_t es una distinción operacional que se ha utilizado por distintas instituciones gubernamentales como privadas interesadas en conocer las características laborales de la población. La aceptación de este criterio

² El Instituto Nacional de Estadística Geografía e Informática (INEGI), organismo gubernamental encargado, entre otras actividades, de la realización de censos, encuestas y diversas investigaciones relacionadas con la dinámica demográfica y laboral de México, considera a la semana anterior a la enumeración como intervalo de tiempo de referencia.

sugeridas por las flechas que van del primer al segundo conjunto de celdas y etiquetadas por E_{it} y S_{it} .

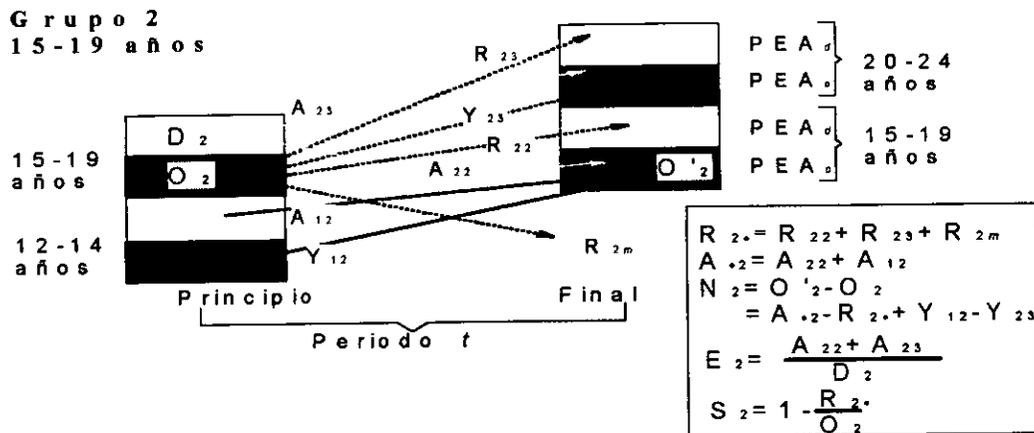
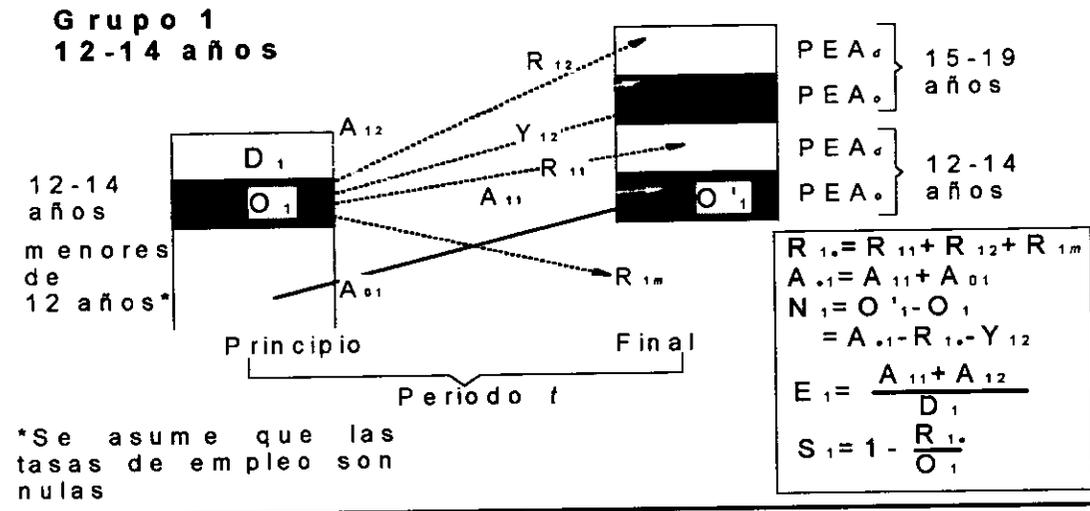
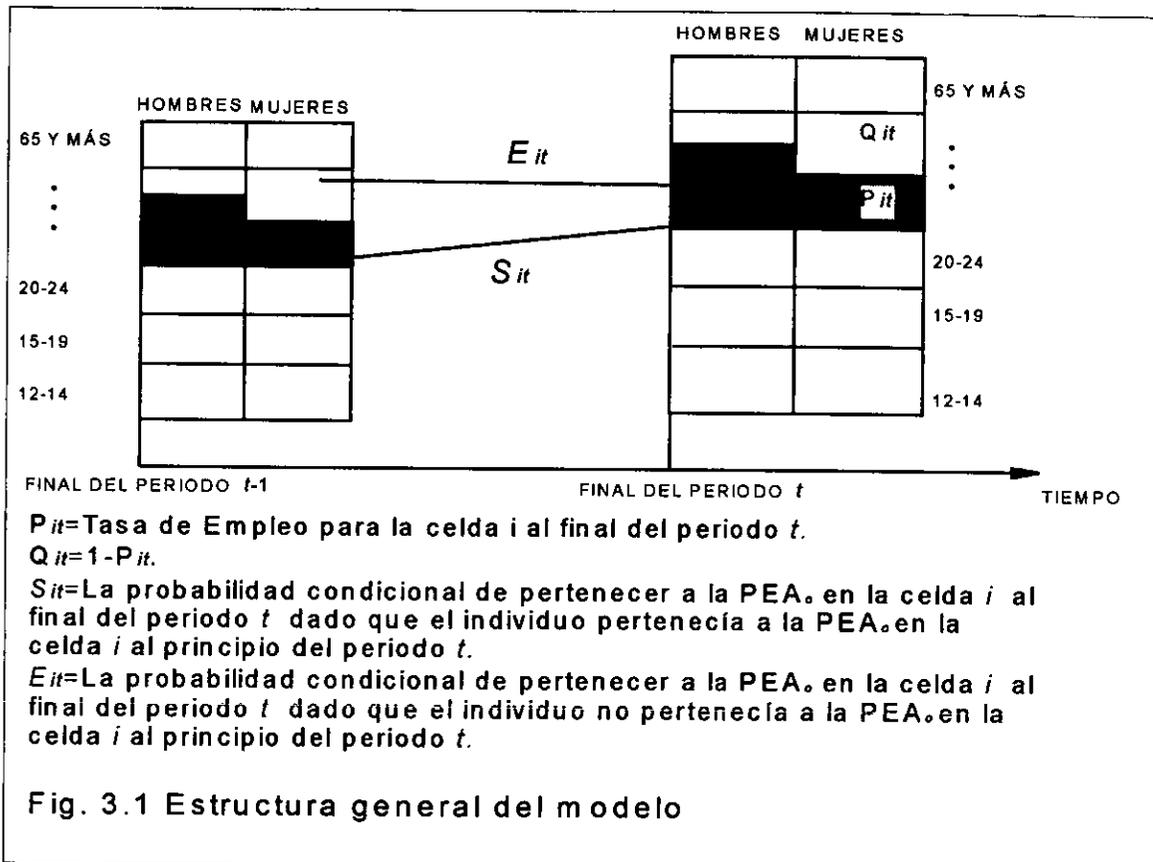


Fig. 3.2 Flujos laborales hacia la PEA_o y la PEA_d.

En la figura 3.2, en el grupo 1 de 12-14 años el total de la PEA_o al final del periodo t difiere del total al principio debido a:

1. D_1 es el total del grupo de 12-14 años dentro de la PEA_d.
2. O_1 es el total del grupo de 12-14 años dentro de la PEA_o.
3. Entradas a la PEA_o provenientes de la PEA_d del mismo grupo de edad, A_{11} .
4. Entradas a la PEA_o de los menores de 12 años al grupo 12-14, A_{01} ; ($A_{11} + A_{01}$ es el total de los ingresos, $A_{.o1}$).

permanecer en ella y su probabilidad se designa por S_{it} . Finalmente, individuos dentro de la PEA_o y pueden dejarla; su probabilidad es $1-S_{it}$



En el modelo representado en la figura 3.1, el conjunto de celdas en el tiempo $t-1$ sobre el eje del tiempo se refiere a las clases de individuos definido de acuerdo con el sexo y la edad. Dentro de cada celda, los individuos están clasificados dentro de la PEA_o y PEA_d . Esto se indica por dos de las celdas con áreas sombreadas y sin sombrear. El segundo conjunto de celdas en el tiempo t en el cual, el tamaño de las celdas y sus fracciones sombreadas son diferentes sugiriendo los posibles efectos de los cambios de la composición y tamaño de la población y las cambiantes tasas de participación. Las probabilidades de entrar a la PEA_o y permanecer en ella son

Las estimaciones de estas probabilidades deben ser consistentes tanto de los flujos hacia dentro y fuera de la PEA_o y con las proporciones observadas de individuos dentro de ella en instantes sucesivos de tiempo; o las llamadas Tasas Refinada de Actividad ($TRA = \frac{PEA}{PT(12 \text{ años y más})} \times 100$). Los flujos observados consisten de los cambios mensuales totales o ingresos y retiros de la PEA_o. Las tasas de participación laboral observadas están dadas en grandes detalles por los censos de población y en mayor detalle en encuestas como las Encuestas Nacionales de Empleo (ENE) y Encuestas Nacionales de Educación, Capacitación y Empleo (ENECE).

3.6. Posibles fuentes de información

Las tasas de participación laboral revelan que las características demográficas; sexo, edad, estado civil, número y edad de los hijos; son importantes determinantes de la probabilidad de pertenecer a la PEA_o. Las tasas de participación también han sido utilizadas en la investigación de las determinantes económicas de la propensión de estar dentro de la PEA_o, especialmente tales determinadas como salarios, desempleo, etc.

3.6.1. Educación

Las probabilidades de participación de la PEA se encuentran estrechamente relacionadas con el grado máximo de estudio de cada persona dentro de la PEA. Esta afirmación es aún más cierta para el caso de la PEA femenina.

3.6.2. Salarios

Para los adultos masculinos, existe una estrecha relación entre salarios ofrecidos y tasas de participación. Las costumbres de la sociedad mexicana asigna el rol al adulto masculino como el agente económico que percibe ingresos. Los diferenciales en los salarios pueden influenciar la vocación y selección de empleo en el adulto masculino. La cuestión de la selección de trabajar o no depende de su salud, responsabilidades familiares y de los ingresos no monetarios o de los ingresos que percibiría por trabajar. Para las mujeres, jóvenes y ancianos se puede esperar niveles salariales que

5. El total de los que permanecen ocupados pero entran en siguiente grupo de edad, Y_{12} .
6. El total de los que salen de la PEA_0 para ingresar a la PEA_d en el mismo grupo de edad, R_{11} .
7. Salidas de la PEA_0 para ingresar al siguiente grupo de edad de la PEA_d , R_{12} .
8. Salidas definitivas por muerte, R_{1d} ($R_{11}+R_{12}+R_{1d}$ es el total de salidas).

N_1 es el cambio neto total en la PEA_0 . Para el grupo 2 de 15-19 años el cambio neto total de la PEA_0 se obtiene de la siguiente relación:

$$N_2 = A_{22} - R_{22} + Y_{12} - Y_{23}.$$

Las relaciones descritas anteriormente son válidas también para los restantes grupos de edad y ambos sexos. Para el último grupo de edad en la actividad, sin un límite superior, de cada sexo se omite el paso al siguiente grupo. Esto significa que en el último grupo de actividad de tamaño constante con una tasa constante de participación para que $D_\alpha = 0$ implique $R_{\alpha\alpha} > A_{\alpha\alpha}$ para compensar el influjo de $Y_{\alpha\alpha}$, ($\alpha =$ último grupo de edad en la actividad.). De manera contraria, para el primer grupo de edad masculino en la actividad cuando las tasas de participación son constantes, se observan valores promedio de $A_{uu} > R_{uu}$. Esto es necesario para mantener constantes las tasas de participación para tales grupos de edad puesto que la pérdida de participantes de la PEA_0 hacia el siguiente grupo de edad es generalmente mayor que el influjo de los participantes de la PEA_0 provenientes de grupos anteriores, es decir, $Y_{u, u+1} > Y_{u-1, u}$. Para el grupo de edad 20-24 femenino, $Y_{u, u+1}$ es generalmente menor que $Y_{u-1, u}$, y las salidas son generalmente mayores a las entradas cuando las tasas de participación son constantes. Esto es efecto de la alta incidencia del matrimonio y la maternidad en este grupo de edad.

Las probabilidades de entrada a la PEA_0 durante el mes t para individuos en el grupo de edad u es $\frac{(A_{uu} + A_{u, u+1})}{D_u}$ y $1 - \frac{R_{uu}}{O_u}$ es la probabilidad de que un individuo en la PEA_0 en el grupo u salga de la actividad.

Mathematica, cubrió tales requisitos. *Mathematica* es un software computacional y lenguaje proyectado para cálculos matemáticos y otras aplicaciones a la física e ingeniería. *Mathematica* puede ser utilizado como una calculadora numérica y simbólica; permite visualizar funciones y datos; actúa como medio de modelación y análisis de datos; es un lenguaje de programación de alto nivel que permite crear grandes y pequeños programas. Además de tener un gran número de funciones integradas, *Mathematica* incluye un completo lenguaje de programación, que permite agregar nuevas extensiones al software. La naturaleza interactuante de *Mathematica* se traduce en el hecho de que en el momento de finalizar la escritura un programa inmediatamente se puede correr. Los programas escritos en el lenguaje de *Mathematica* pueden hacer uso de aspectos simbólicos. Ellos pueden crear y manipular estructuras de datos simbólicos. *Mathematica* soporta varios estilos de programación como la programación de procedimientos, con estructuras en bloques, condicional, iteración y recursión y programación funcional, con funciones puras, operadores funcionales y operaciones de programación estructurada.

3.8. La muestra inicial

De acuerdo con los resultados del XI Censo General de Población y Vivienda 1990, relacionados con los indicadores de empleo, presentados en el capítulo 1, el comportamiento demográfico y económico de la población difiere en grandes proporciones entre cada agrupación de estados con condiciones similares de mercado laboral e incluso en las áreas más y menos urbanizadas de un mismo estado. La información necesaria para realizar una simulación por cada una de las 32 entidades de la República Mexicana precisaría de los datos correspondientes, por otra parte de un tratamiento especial para las áreas según su nivel de urbanización. Además, los estudios que han medido los fenómenos demográficos a utilizar en el modelo, difícilmente son disponibles a nivel estatal. Estas son las razones por lo que esta investigación estuvo limitada a nivel nacional.

El inicio de la simulación consistió en obtener una muestra representativa de la población mexicana. Ante el alto grado de agregación de la información fue necesario

influyen sus tasas de participación tanto los salarios ofrecidos y los salarios ganados por el jefe de la familia son probablemente influencia de la decisión individual de trabajar.

Sin embargo, por la constante inestabilidad de la economía nacional mexicana y políticas de bajos salarios, la tendencia lógica de un miembro de un hogar es la búsqueda de un ingreso monetario adicional sin importar el nivel de ingreso del jefe del hogar.

3.6.3. Desempleo

Los efectos del desempleo sobre la participación de la PEA_o han sido objeto de muchas controversias. La pérdida del empleo del jefe de la familia resulta no sólo en la búsqueda de un nuevo empleo sino también conduce a otro miembro de su familia a ingresar a la PEA y posteriormente pertenecer a la PEA_o.

3.6.4. Limitación de la información

Las variables particulares que se han discutido no agotan todos los factores que influyen las posibilidades de estar dentro de la PEA_o o no. Estas variables representan simplemente algunos de los más obvios que han sido sujetos de investigación. Sin embargo, no hay datos disponibles que pudieran permitir estimar los efectos separados de esas variables sobre la probabilidad de estar en la PEA_o. La información disponible es la que se encuentra contemplada por los censos de población y encuestas de empleo, ingresos y gastos de los hogares. A menudo, es necesaria información que no se publica habitualmente por dependencias responsables; en cuyo caso es necesario recurrir a los cuestionarios fuentes y reprocesar los datos para obtener los tabulados de la información requerida, muchas ocasiones imposible de llevar a cabo por su dificultad o elevado costo.

3.7. Elección del lenguaje de computación

Era necesario un software que interactuará eficientemente con hojas de cálculo y procesadores de texto además de características deseables para la simulación;

Mathematica, cubrió tales requisitos. *Mathematica* es un software computacional y lenguaje proyectado para cálculos matemáticos y otras aplicaciones a la física e ingeniería. *Mathematica* puede ser utilizado como una calculadora numérica y simbólica; permite visualizar funciones y datos; actúa como medio de modelación y análisis de datos; es un lenguaje de programación de alto nivel que permite crear grandes y pequeños programas. Además de tener un gran número de funciones integradas, *Mathematica* incluye un completo lenguaje de programación, que permite agregar nuevas extensiones al software. La naturaleza interactuante de *Mathematica* se traduce en el hecho de que en el momento de finalizar la escritura un programa inmediatamente se puede correr. Los programas escritos en el lenguaje de *Mathematica* pueden hacer uso de aspectos simbólicos. Ellos pueden crear y manipular estructuras de datos simbólicos. *Mathematica* soporta varios estilos de programación como la programación de procedimientos, con estructuras en bloques, condicional, iteración y recursión y programación funcional, con funciones puras, operadores funcionales y operaciones de programación estructurada.

3.8. La muestra inicial

De acuerdo con los resultados del XI Censo General de Población y Vivienda 1990, relacionados con los indicadores de empleo, presentados en el capítulo 1, el comportamiento demográfico y económico de la población difiere en grandes proporciones entre cada agrupación de estados con condiciones similares de mercado laboral e incluso en las áreas más y menos urbanizadas de un mismo estado. La información necesaria para realizar una simulación por cada una de las 32 entidades de la República Mexicana precisaría de los datos correspondientes, por otra parte de un tratamiento especial para las áreas según su nivel de urbanización. Además, los estudios que han medido los fenómenos demográficos a utilizar en el modelo, difícilmente son disponibles a nivel estatal. Estas son las razones por lo que esta investigación estuvo limitada a nivel nacional.

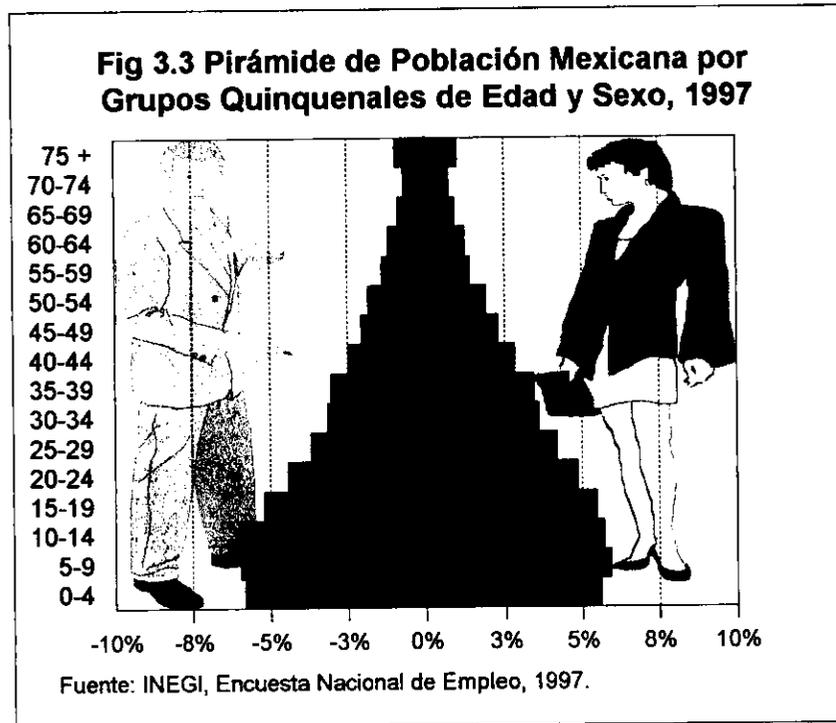
El inicio de la simulación consistió en obtener una muestra representativa de la población mexicana. Ante el alto grado de agregación de la información fue necesario

diseñar un programa de computación para la desagregación de los datos. La esencia del programa consistió en obtener probabilidades empíricas acorde con la composición deseada de la muestra. Las características de cada individuo se obtuvieron con diferentes rutinas del programa principal *MUESTRA1*.

La posibilidad de que una persona tuviera un rasgo específico obedeció a probabilidades incondicionales pero en mayor medida a las probabilidades condicionales relacionados con su condición de actividad o inactividad. La función de distribución uniforme en el rango (0, 1) se usó para que cada rutina generará los valores de su distribución empírica correspondiente.

Las características iniciales fueron: Sexo, Edad, Estado Civil, Escolaridad y Condición Escolar para la población de 6 años y más; para el caso de mujeres de 12 años y más, Paridez; para la población de 12 años y más tanto femenina como masculina, Condición de Actividad; para la PEA, Condición de Ocupación; finalmente para la PEA_o; Rama de Actividad (Cuadro 3.1). La Encuesta Nacional de Empleo 1997 (ENE-1997) y su módulo especial de la Encuesta Nacional de Educación, Capacitación y Empleo 1997 (ENECE-1997) fueron las principales fuentes de información.

A cada rutina se le asignó un ciclo de números aleatorios (Random[]) para obtener una independencia probabilística entre las características de un individuo. La primera rutina se basó en la composición de la pirámide de población mexicana por grupos de edad y sexo, para obtener una cota inferior de edad de acuerdo con los grupos quinquenales y posteriormente con $\text{IntA} []$ sumarle un entero generado aleatoriamente en el rango [0, 4]; con excepción del grupo quinquenal 65 y más con el cual se utilizó $\text{Int65} []$ para sumar a la cota inferior del grupo, enteros del rango [0, 34], para los restantes grupos se usó IntA . El motivo de utilizar este procedimiento se debió a que las probabilidades se obtuvieron de la ENE-1997 y por lo tanto implicaba una corrección por la mala declaración de la edad.



Este problema se reduce en cierta forma con la agregación en grupos quinquenales, pero para los objetivos de el modelo era necesario contar las edades desplegadas individualmente, y gracias a la uniformidad para generar enteros de $IntA[]$ e $Int65[]$ la corrección fue innecesaria. $Genero[]$ asignó el sexo de la persona con base de la proporción de hombres y mujeres por cada grupo quinquenal.

Para obtener el estado civil de las personas se clasificó como Menores a la población menor de 12 años y para las personas de 12 años o más, Solter a los solteros, con $Nupcia[]$ se agrupó en Unid a las personas casadas o en unión libre; $Separad[]$ consideró las personas con ruptura de unión, es decir, divorciadas, viudas y separadas.

Cuadro 3.1 Ejemplos de los Resultados Producidos por Muestra 1.

Sexo	Edad	Estado Civil	Años de Escolaridad (mayores de 6 años)	Condición de Escolaridad	Condición de Empleo	Condición de Ocupación	Condición de Actividad
Mujer	26	Solter	9.5 Sest	0	PEA	Ocupad	SerVar
Hombre	40	Unid	6 Sest	1/NA	PEA	Ocupad	Transp
Hombre	40	Unid	7.5 Sest	NA	PEA	Ocupad	Comerc
Hombre	38	Unid	9.5 Sest	NA	PEA	Ocupad	HotRest
Mujer	14	Solter	7.5 Est	0	PEI	NA	NA
Hombre	13	Solter	2 Est	NA	PEI	NA	NA
Hombre	53	Unid	17 Sest	NA	PEA	Ocupad	Comerc
Mujer	28	Solter	9.5 Sest	4	PEI	NA	NA
Hombre	3	Menor	0 NEst	NA	NA	NA	NA
Hombre	52	Unid	9 Sest	NA	PEA	Ocupad	Transp
Mujer	9	Menor	2 Est	NA	NA	NA	NA
Mujer	17	Unid	6 Sest	0	PEA	Ocupad	Comerc
Hombre	60	Unid	6 Sest	NA	PEA	Ocupad	Agropec
Mujer	16	Solter	9 Est	0	PEI	NA	NA
Mujer	85	Separad	0 Sest	1	PEA	Ocupad	SerVar
Hombre	4	Menor	0 NEst	NA	NA	NA	NA
Hombre	11	Menor	0 Est	NA	NA	NA	NA
Mujer	61	Separad	2 Sest	9	PEA	Desocupad	NA
Mujer	39	Solter	9.5 Sest	4	PEI	NA	NA
Mujer	28	Separad	9.5 Sest	2	PEA	Ocupad	SerVar
Hombre	23	Unid	4.5 Sest	NA	PEA	Ocupad	Agropec
Hombre	19	Solter	4.5 Sest	NA	PEA	Ocupad	Agropec
Mujer	18	Unid	9.5 Sest	0	PEA	Ocupad	Comerc

1/ No Aplica.

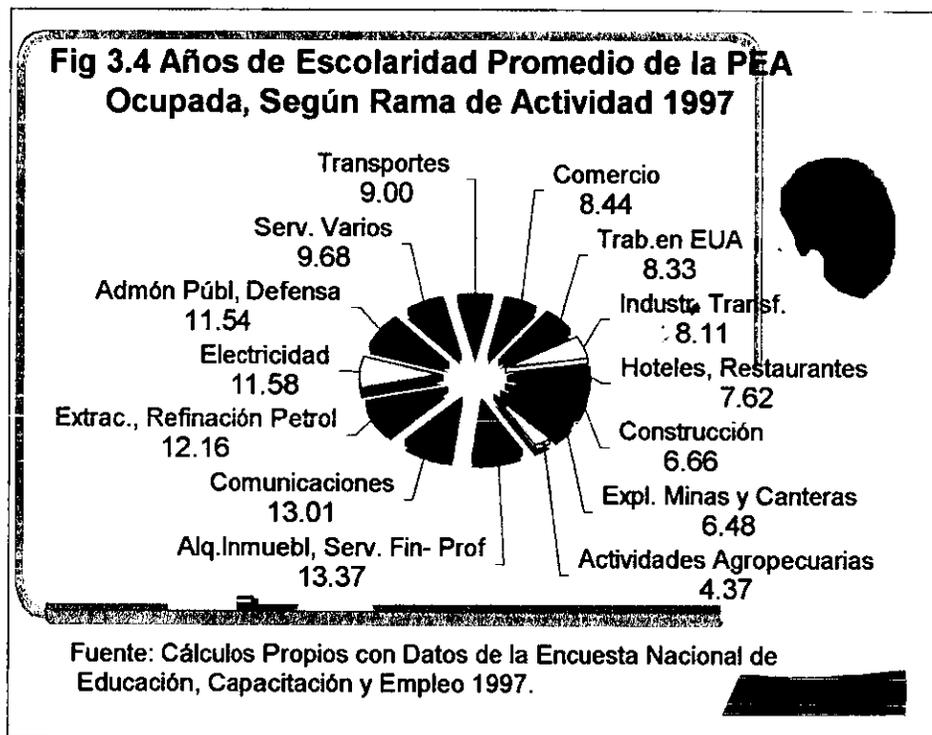
De la distribución de probabilidad empírica se derivó del estado civil de la población de 12 años y más registrada por el Censo de Población y Vivienda 1995, fuente más reciente desagregada en grupos quinquenales de edad. Tal distribución se utilizó tanto para la población femenina como masculina. La importancia del estado civil en el modelo, radica en ser un determinante de la fecundidad de las mujeres de 12 años y más y cuya consecuencia final es el crecimiento total de la población.

Población de 12 años y más por Grupos Quinquenales de Edad, Según Estado Civil.						
Cuadro 3.2						
Grupos de Edades	Hombres			Mujeres		
	Solteros	Unidos	Separados	Solteras	Unidas	Separadas
12-14	3,214,979	1,608	0	3,154,911	11,081	0
15-19	4,717,378	254,320	7,480	4,293,414	760,845	48,510
20-24	2,896,164	1,536,511	53,417	2,402,635	2,270,313	169,159
25-29	1,157,834	2,393,508	68,427	909,458	2,822,441	219,361
30-34	480,041	2,572,744	77,658	464,266	2,668,424	273,519
35-39	261,916	2,433,113	91,768	280,329	2,439,555	290,567
40-44	146,690	1,929,684	87,192	179,264	1,749,518	328,952
45-49	98,748	1,575,223	83,110	119,615	1,420,984	304,206
50-54	72,535	1,252,747	88,654	91,875	1,097,045	284,916
55-59	52,166	929,484	98,176	68,764	764,314	312,763
60-64	29,849	799,440	97,427	55,432	589,351	364,091
65 años y más	67,157	1,398,077	413,103	115,401	809,293	1,199,700

Fuente: INEGI, Censo de Población y Vivienda, 1995.

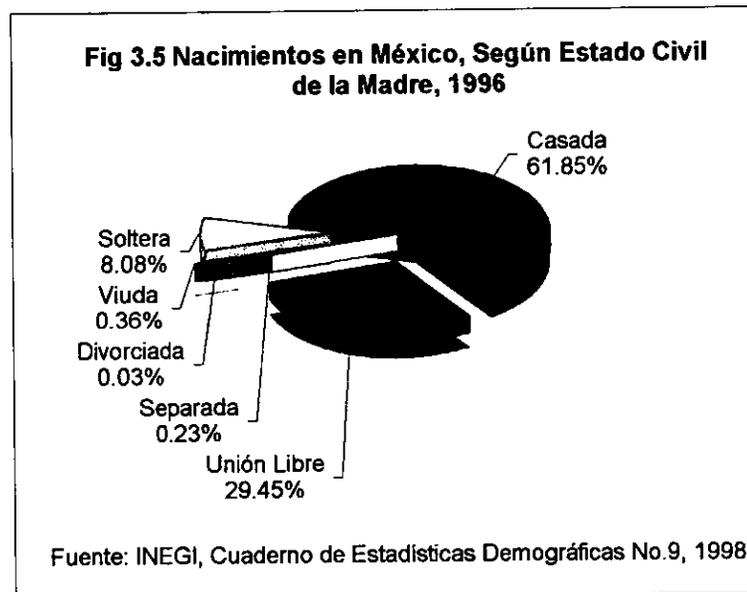
La rutina Escolaridad[] indicó la escolaridad por sexo, utilizando una distribución empírica a nivel nacional de las personas de 6 años y más. La agrupación de las edades obtenida de la ENECE-97 inició con el grupo de 6-11 años, para continuar con grupos trienales hasta el grupo de 22-24 años, los restantes grupos de edades estaban clasificados de manera quinquenal y el grupo final de 65 años y más. Al no contar con las edades individuales y grado de escolaridad se fijó una edad mínima en relación con las distribuciones observadas por grupo de edad y el máximo grado de estudios. En este sentido, una persona menor de 21 años no podría obtener los años de estudio equivalente a estudios superiores, una persona menor de 15 años tampoco obtendría una escolaridad equivalente a 12 años (preparatoria completa). Cuando los

años de estudio se localizaron agregados se utilizó la escolaridad promedio para asignar un valor concreto de instrucción y no un rango. La capacitación para el trabajo y los estudios subprofesionales, no son considerados como estudios requisito para aspirar al siguiente grado por lo que se les asignó un valor 0.5 años más el último grado de estudios obtenido (6 y 9 respectivamente). Esta distinción obedece a que la escolaridad se utilizó como un factor para la ocupación de la PEA según la rama de actividad.



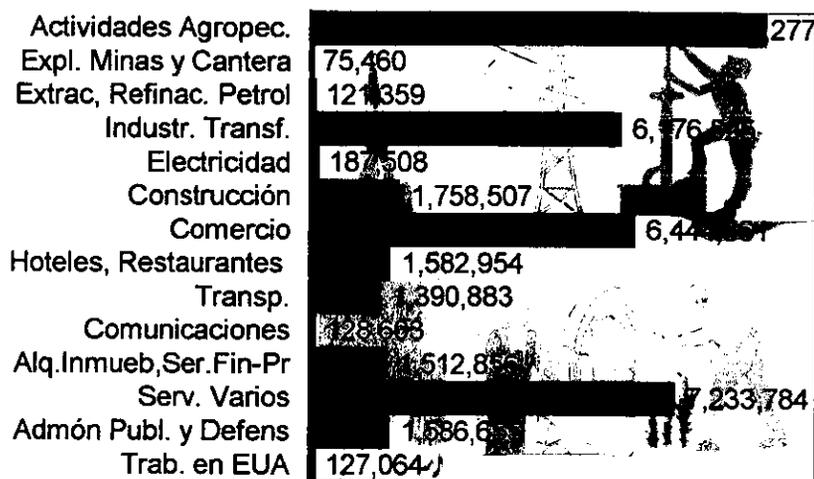
El número de hijo nacidos vivos, en un principio se limitaría sólo para mujeres casadas y sería considerada sólo la fecundidad de mujeres unidas, pero de acuerdo con los nacimientos registrados en 1996, los nacimientos de madres solteras se colocaron en el tercer lugar del total de la mujeres con algún hijo registrado vivo, este hecho es

explicado por la mayor aceptación por parte de la sociedad de la fecundidad de mujeres solteras además de que en muchas ocasiones la madre en unión libre se registra como soltera. Entonces fue necesario tomar en cuenta la paridez de las mujeres solteras. En sentido opuesto la paridez registrada, de las mujeres con ruptura de unión no alcanzó más del 0.6% del total, por tal razón el modelo no consideraría elegible una mujer con ruptura de unión, para tener un hijo mientras se mantuviera en esta condición.



Las proporciones de la PEA y de la PEI de cada grupo quinquenales de edad a partir del grupo 15-19 años hasta el de 65 años y más, incorporando el grupo de 12-14 años, sirvieron para fijar la condición de actividad con CondAct[]. Para el caso de la PEA, se clasificó a este efectivo en ocupados y desocupados usando la rutina CondOcup[].

Ramactividad[] sólo tomó en cuenta la PEA_o y conforme a el máximo nivel de instrucción de cada individuo, se le asignó la rama de actividad en que se encontraría ocupado. En esta clasificación se excluyó a los Trabajadores en Estados Unidos cuya condición de ocupación obedece a la situación de empleo en ese país.

Fig. 3.6 Población Ocupada Según Rama de Actividad

Fuente: INEGI, Encuesta Nacional de Empleo 1997.

Para la generación de las distribuciones empíricas del modelo, se descontó a todos los no especificados de cada una de las características de la población de la muestra, por no existir una certeza de su distribución y poder realizar un prorrateo. Los datos usados en el modelo corresponden a la ENE-1997 y ENECE-1997 realizadas en el trimestre Abril-Junio. El desarrollo de la muestra sería evaluado anualmente, en ausencia de movimientos migratorios.

3.9. Rutinas demográficas

Las tablas demográficas constituyen una de las formas de presentar el comportamiento de la población frente a un determinado fenómeno demográfico en estado puro, sin interferencia de otros fenómenos demográficos. La distribución por edades del evento estudiado y la proporción final de las personas a las cuales les llega a ocurrirle son dos de las características que se pueden apreciar mediante una tabla.

El desarrollo de las características demográficas de la muestra inicial se apoyó en tablas demográficas, algunas construidas previamente y otras diseñadas

especialmente para el modelo. El periodo de referencia del modelo fue anual, entonces se realizó la equivalencia para las tablas con periodos distintos:

1. Mortalidad[]. La principal rutina demográfica. Al aplicarse a cada persona de la muestra decidía si sobrevivía o no, si ocurría lo primero entonces el proceso continuaba para las demás rutinas. Las probabilidades de sobrevivencia fueron distintas para hombres y mujeres de acuerdo con las tablas de mortalidad elaboradas por S. Camposortega.

Tabla de Probabilidades de Muerte por Grupos de Edad y Sexo. México, Nivel Nacional 1995-2010.

Cuadro 3.3

Grupos de Edades	1995		2000		2005		2010	
	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres	Hombres	Mujeres
0	0.04664	0.03145	0.04297	0.02826	0.03984	0.02564	0.03714	0.02347
1	0.01969	0.01731	0.01830	0.01570	0.01710	0.01138	0.01606	0.01327
5	0.00586	0.00465	0.00545	0.00423	0.00510	0.00388	0.00480	0.00359
10	0.00369	0.00252	0.00344	0.00229	0.00322	0.00210	0.00303	0.00195
15	0.00609	0.00355	0.00568	0.00323	0.00532	0.00297	0.00501	0.00275
20	0.00992	0.00550	0.00926	0.00501	0.00869	0.00461	0.00819	0.00427
25	0.01231	0.00717	0.01151	0.00655	0.01081	0.00602	0.01020	0.00559
30	0.01534	0.00919	0.01437	0.00840	0.01352	0.00774	0.01278	0.00719
35	0.01952	0.01160	0.01832	0.01062	0.01728	0.00981	0.01636	0.00912
40	0.02426	0.01420	0.02282	0.01304	0.02156	0.01203	0.02046	0.01123
45	0.03176	0.01879	0.02997	0.01729	0.02840	0.01602	0.02700	0.01496
50	0.04205	0.02643	0.03982	0.02440	0.03785	0.02269	0.03610	0.02123
55	0.05728	0.03768	0.05449	0.03495	0.05200	0.03262	0.04977	0.30620
60	0.08132	0.05602	0.07778	0.05225	0.07459	0.04901	0.07172	0.04623
65	0.11979	0.08712	0.11538	0.08191	0.11136	0.07738	0.10770	0.07343
70	0.18335	0.14242	0.17815	0.13537	0.17335	0.12913	0.16891	0.12362
75	0.29111	0.24514	0.28586	0.23654	0.28091	0.22873	0.27626	0.22167
80	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000	1.00000

Fuente: Camposortega, S., *Proyecciones de la Población Mexicana 1970-2040*. Tesis de Maestría, El Colegio de México, 1980.

2. Actualiza[]. Esta rutina envejece en un año a la población sobreviviente después del proceso de Mortalidad[].
3. Natalidad[]. Decide el nacimiento de un nuevo miembro de la muestra de una mujer soltera ó unida de 15 a 49 años, antes revisa su paridez y si aún esta dentro del

limite, el proceso de nacimiento continua y aumenta su paridez por uno, en caso contrario se ignora el proceso; asigna el sexo femenino al bebé con una probabilidad de 0.488 y 1-0.488 el masculino. Las probabilidades de nacimiento de un hijo estan asociadas con las tasas de fecundidad calculadas por S. Camposortega, de una hipótesis baja. Se decidió tomar esta hipótesis, porque las tendencias de crecimiento de la población observadas de 1990 a 1997 han demostrado sobreestimación incluso tomando hipótesis de fecundidad baja.

Tasas Anuales de Fecundidad, Tasa Global de Fecundidad y Tasa Bruta de Reproducción, Según Hipótesis Baja de S. Camposortega, 1995-2010.

Cuadro 3.4

	1995	2000	2005	2010
Tasa Global de Fecundidad	2.6770	2.5480	2.4950	2.4740
Tasa Bruta de Reproducción	1.3060	1.2430	1.2170	1.2070
Grupos de Edades	Tasas de Fecundidad			
15-19	0.0612	0.0585	0.0575	0.0570
20-24	0.1841	0.1771	0.1742	0.1731
25-29	0.1427	0.1362	0.1336	0.1325
30-34	0.0825	0.0778	0.0759	0.0752
35-39	0.0449	0.0418	0.0450	0.0400
40-44	0.0172	0.0156	0.0150	0.0148
45-49	0.0029	0.0025	0.0023	0.0022

Fuente: Camposortega, S., *Proyecciones de la Población Mexicana 1970-2040*. Tesis de Maestría, El Colegio de México, 1980.

4. Unión[]: Usando la técnica propuesta por J. Quilodrán³, se desarrolló esta tabla de probabilidades de unión. El comportamiento de las uniones de la muestra es un factor primordial de las probabilidades de procreación. Sin embargo, al contar con una tabla de fecundidad general, el papel del rutina Unión[] sería sólo para establecer la restricción a las mujeres separadas para no procrear un hijo si aún poseía las características para hacerlo. Por esta misma razón el modelo de uniones es por demás rudimentario; por otra parte no fue posible elaborar tablas quinquenales proyectadas y el modelo utilizó las mismas probabilidades para todo el periodo de simulación. Los datos observados de las uniones en México sugerían

³ Quilodrán, Julieta. *Tablas de Nupcialidad para México*.

tablas con incrementos en las edades medias a la primera unión y en el número de célibes definitivos, pero los fines de la investigación son otros, entonces la tabla construida con datos del Censo de Población y Vivienda 1995 es suficiente para la simulación.

Proporciones de Solteros por Sexo y Grupos Quinquenales de Edades, 1995
(Por Cada 1,000 Habitantes)

Cuadro 3.5

Grupos de Edades	Mujeres	Hombres
12-14	996	1,000
15-19	841	947
20-24	496	646
25-29	230	320
30-34	136	153
35-39	93	94
40-44	79	68
45-49	65	56
50-54	62	51

Fuente: INEGI, Censo de Población y Vivienda, 1995.

Proporciones de Solteros para cada Edad Exacta, 1995.
(Por Cada 1,000 Habitantes)

Cuadro 3.6

Grupos de Edades	Mujeres	Hombres
12-14	1,000	1,000
15-19	700	862
20-24	380	510
25-29	188	246
30-34	112	126
35-39	88	82
40-44	72	60
45-49	63	54
50-54	59	48

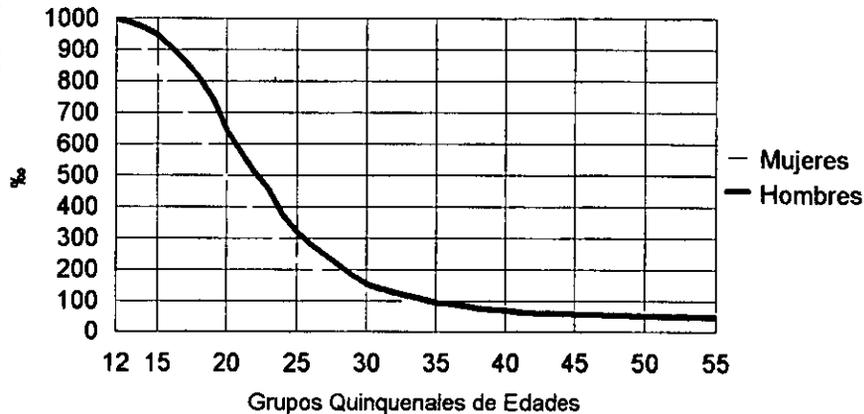
Fuente: INEGI, Censo de Población y Vivienda, 1995.

Tabla de Nupcialidad de Solteros, México Nivel Nacional 1995.

Cuadro 3.7

Edad X	Mujeres			Hombres		
	C_x	$m(x, x + a)$	$a\pi_x$	C_x	$m(x, x + a)$	$a\pi_x$
12	1,000	138	0.30000	1,000	300	0.13800
15	862	352	0.45714	700	320	0.40835
20	510	264	0.50526	380	192	0.51765
25	246	120	0.40426	188	76	0.48780
30	126	44	0.21429	112	24	0.34921
35	82	22	0.18182	88	16	0.26829
40	60	6	0.12500	72	9	0.10000
45	54	6	0.06000	63	4	0.11111
50	48	-	-	59	-	-

Fig 3.7 Ajuste Gráfico de las Proporciones de Solteros de 15 años y más, 1995.



Fuente: INEGI, Censo de Población y Vivienda, 1995.

5. Uniones[]. Con el mismo procedimiento de Unión1[] sólo que utilizando al efectivo de separados, se decidió sobre las uniones de orden dos o mayor. Ante la falta de datos acerca del número de uniones y tipo de ruptura de unión (separación, divorcio o viudez), tampoco se llevó un registro de las rupturas de uniones de la muestra y se asumieron las mismas probabilidades de nuevas uniones sin importar el número de uniones anteriores. Hubo una modificación en el orden y número de grupos quinquenales por ser una función creciente

Proporciones de Separados por Sexo y Grupos Quinquenales de Edades, 1995
(Por Cada 1,000 Habitantes)

Cuadro 3.8

Grupos de Edades	Mujeres	Hombres
12-14	0	0
15-19	10	2
20-24	35	12
25-29	56	19
30-34	80	25
35-39	97	33
40-44	146	40
45-49	165	47
50-54	193	63
55-59	273	91
60-64	361	105

Fuente: INEGI, Censo de Población y Vivienda, 1995.

Proporciones de Separados para cada Edad Exacta, 1995.
(Por Cada 1,000 Habitantes)

Cuadro 3.9

Grupos de Edades	Mujeres	Hombres
12-14	3	1
15-19	20	6
20-24	43	15
25-29	66	21
30-34	87	28
35-39	117	36
40-44	154	43
45-49	176	53
50-54	225	74
55-59	324	99
60-64	400	110

Fuente: INEGI, Censo de Población y Vivienda, 1995.

Tabla de Nupcialidad de Separados, México Nivel Nacional 1995.

Cuadro
3.10

Edad X	Mujeres			Hombres		
	S _x	m _s (x, x + a)	a n' x	S _x	m _s (x, x + a)	a n' x
60	400	.76	0.19000	110	11	0.10000
55	324	.99	0.30556	99	25	0.25253
50	225	.49	0.21778	74	21	0.28378
45	176	.22	0.12500	53	10	0.18868
40	154	.37	0.24026	43	7	0.16279
35	117	.30	0.25641	36	8	0.22222
30	87	.21	0.24000	28	7	0.25000
25	66	.23	0.35000	21	6	0.28571
20	43	.23	0.53000	15	9	0.60000
15	20	.17	0.85000	6	5	0.83333
12	3	-	-	1	-	-

Fig 3.8 Ajuste Gráfico de las Proporciones de Mujeres Separadas de 15 años y más, 1995.

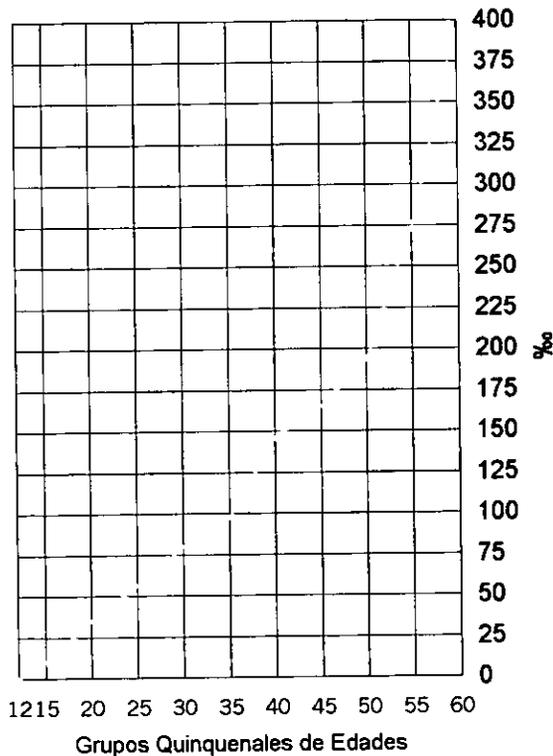
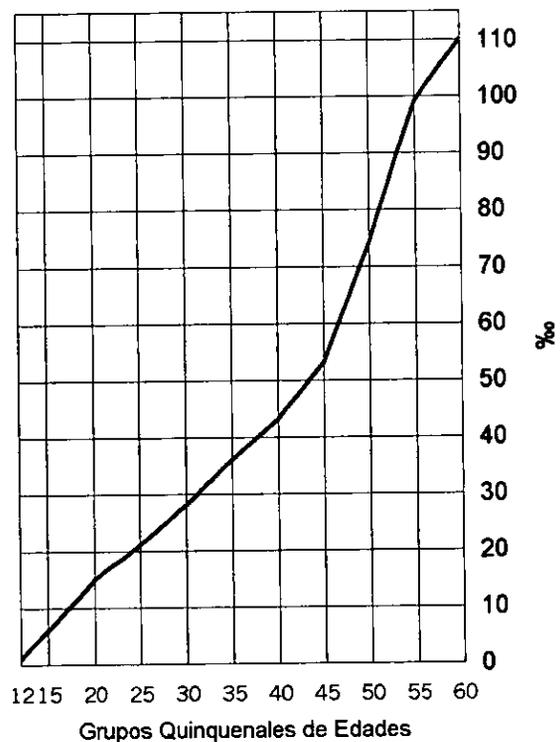
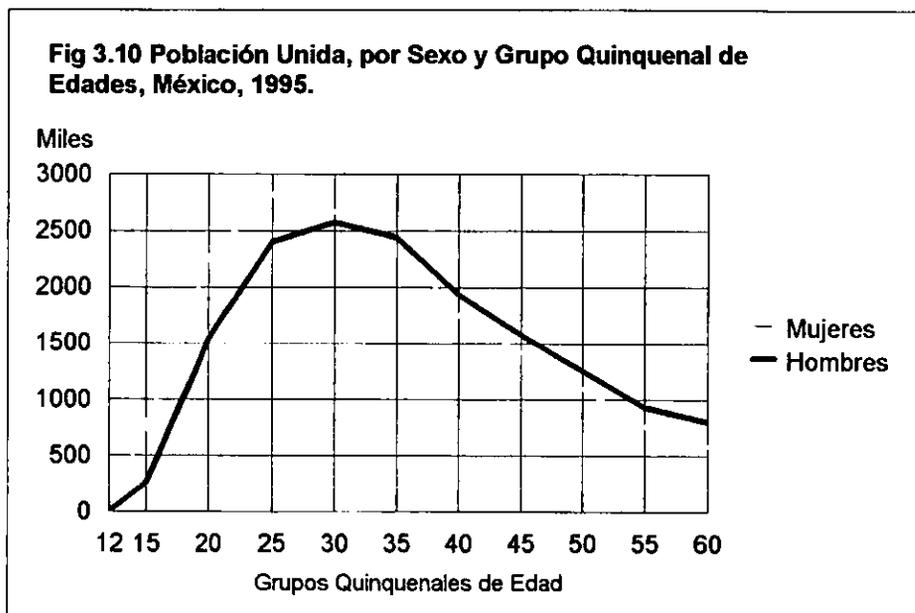


Fig 3.9 Ajuste Gráfico de las Proporciones de Hombres Separados de 15 años y más, 1995.



6. Separación[]: La elaboración de una tabla de probabilidades de ruptura de una unión es muy complicada por la distribución de la población unida, quien es la que corre el riesgo de ruptura. La proporción de unidos es creciente a partir de los 12 años para llegar a su máximo entre los 20 y 30 años y después decrecer en las edades posteriores (fig. 3.9). Estas tendencias se explican por la incidencia de separaciones, divorcios y en gran medida en los grupos de edades mayores, por la viudez. Las proporciones de separados por grupos quinquenales de edades sin ajustes, se usó como las probabilidades de separación de la población unida.



Proporciones de Separados por Sexo y Grupos Quinquenales de Edades, 1995 (Por Cada 100 Habitantes)

Cuadro 3.11

Grupos de Edades	Mujeres	Hombres
12-14	0.00000	0.00000
15-19	0.00950	0.00150
20-24	0.03493	0.01191
25-29	0.05551	0.01890
30-34	0.08030	0.02481
35-39	0.09651	0.03293
40-44	0.14570	0.04030
45-49	0.16490	0.04730
50-54	0.19332	0.06270
55-59	0.27295	0.09092
60-64	0.36089	0.10513
65 años y más	0.56473	0.21993

Fuente: INEGI, Censo de Población y Vivienda, 1995.

7. Educación[]: Esta rutina se compuso por tres tablas de escolaridad para cada sexo con algunas modificaciones a los procedimientos descritos por A. Mina⁴. La misma tabla determinó las probabilidades de deserción escolar. De no ocurrir la deserción, se asumió que se aprobaba el curso. La primer tabla se diseñó para la población de 6 a 11 años de edad con el supuesto que la población sin estudiar para el grupo es el efectivo de 12 a 14 años sin instrucción, de haber seguido exactamente la técnica para la elaboración de las tablas de escolaridad, se hubiera caído en contradicciones como probabilidades negativas. Parte del problema puede ser explicado por ser un análisis de tipo transversal en el cual se estudian a varias generaciones donde las más antiguas poseen los menores niveles de escolaridad descontrolando al modelo de escolaridad, diseñado para una población que al incrementarse la edad se incrementan los años de instrucción. Además la información suministrada por la ENECE-1997 hace mención al número de años-escolaridad aprobados, subestimando al conjunto de estudiantes en el primer grado de educación primaria donde un elevado porcentaje de este grupo aprobará el primer curso. La segunda tabla se hizo para la PEA de 12 a 39 años que no presentó complicación alguna y la tercera para la PEI de 12 a 39 años. Estas probabilidades de deserción escolar se utilizaron para todo el periodo de simulación, por la complejidad para elaborar tablas proyectadas.

Probabilidades de Deserción Escolar 1997					
					Cuadro 3.12
Grupo de Edades	Estudiando	Sin Estudiar	Total	A _x	a _x
6-11	6,533,165	72,280	6,605,445	0.98906	0.10460
12-14	2,934,524	379,073	3,313,597	0.88560	-
Mujeres					
6-11	6,488,236	78,905	6,567,141	0.98798	0.08848
12-14	2,927,100	323,190	3,250,290	0.90057	-

⁴ Mina Valdés, Alejandro. *Curso Básico de Demografía*.

Grupo de Edades					
	Estudiando	Sin Estudiar	Total	A _x	a _x
Probabilidades de Deserción Escolar 1997. PEA					
Hombres					
12-14	553,601	222,741	776,342	0.71309	0.66856
15-19	670,252	2,165,602	2,835,854	0.23635	0.57966
20-24	344,808	3,125,972	3,470,780	0.09935	0.58398
25-29	135,382	3,140,229	3,275,611	0.04133	0.53091
30-34	53,571	2,709,555	2,763,126	0.01939	0.18130
35-39	41,509	2,573,587	2,615,096	0.01587	-
Mujeres					
12-14	228,724	87,813	316,537	0.72258	0.66190
15-19	390,242	1,207,115	1,597,357	0.24430	0.43397
20-24	259,966	1,619,984	1,879,950	0.13828	0.64280
25-29	91,206	1,755,246	1,846,452	0.04939	0.31910
30-34	49,906	1,433,924	1,483,830	0.03363	0.02122
35-39	50,293	1,477,464	1,527,757	0.03292	-
Probabilidades de Deserción Escolar 1997. PEI					
Hombres					
12-14	2,380,923	84,052	2,464,975	0.96590	0.07188
15-19	1,750,046	202,115	1,952,161	0.89647	0.11191
20-24	445,269	114,016	559,285	0.79614	0.28573
25-29	53,855	40,850	94,705	0.56866	0.79488
30-34	4,291	32,497	36,788	0.11664	0.15045
35-39	3,758	34,166	37,924	0.09909	-
Mujeres					
12-14	2,619,120	235,377	2,854,497	0.91754	0.34158
15-19	2,083,431	1,365,220	3,448,651	0.60413	0.66934
20-24	507,838	2,034,371	2,542,209	0.19976	0.81831
25-29	70,519	1,872,395	1,942,914	0.03629	0.55803
30-34	26,941	1,652,520	1,679,461	0.01604	0.66742
35-39	7,508	1,399,782	1,407,290	0.00533	-
Fuente: Cálculos Propios con datos de la Encuesta Nacional de Educación, Capacitación y Empleo 1997.					

3.10. Rutinas económicas.

1. Actividad []: Para determinar las entradas y salidas de la actividad se elaboró una tabla de vida activa con el método utilizado por M. Cervera⁵ para la elaboración de este tipo de tabla. Se obtuvieron 12 series de vida activa para cada sexo con la serie l_x obtenida por S. Camposortega y proyecciones de PEA hechas por C. Rodríguez⁶. Para los fines de la simulación, las probabilidades de entrada y salida

⁵ Cervera Flores, Miguel. *Tablas de vida activa para la República Mexicana*.

⁶ Rodríguez Hernández, Constanza. *Proyecciones de P.E.A. Nacional 1980-2010*.

de la actividad se calcularon con las series ${}_n h_x^{ia}$ (número de personas que entran a la actividad entre las edades x y $x+n$ en un año), ${}_n h_x^{ai}$ (número de personas que salen de la actividad entre las edades x y $x+n$ en un año) y con l_x^a (número de sobrevivientes activos entre las edades x y $x+n$) y l_x^i (número de sobrevivientes activos entre las edades x y $x+n$). Las edades marcadas en negritas en el cuadro 3.13 corresponden a las cotas superiores para entrar a la actividad.

Probabilidades de entrada y salida de la actividad económica.		
Cuadro 3.13		
Grupo de edades	1997	
Hombres	$P({}_n h_x^{ia})$	$P({}_n h_x^{ai})$
12-14	0.15599	-
15-19	0.28384	-
20-24	0.23950	-
25-29	0.10420	-
30-34	0.01386	-
35-39	-	0.00191
40-44	-	0.01329
45-49	-	0.02637
50-54	-	0.04225
55-59	-	0.07316
60-64	-	0.16185
65-69	-	0.26858
70-74	-	0.36338
75-79	-	0.56954
80 y más	-	1.00000
Mujeres	$P({}_n h_x^{ia})$	$P({}_n h_x^{ai})$
12-14	0.08304	-
15-19	0.12266	-
20-24	0.07068	-
25-29	0.02559	-
30-34	0.01511	-
35-39	-	0.01517
40-44	-	0.06159
45-49	-	0.10313
50-54	-	0.15261
55-59	-	0.12114
60-64	-	0.18164
65-69	-	0.27327
70-74	-	0.37162
75-79	-	0.58498
80 y más	-	1.00000

Fuente: Cálculos propios con datos de la Encuesta Nacional de Empleo, 1997 y *Proyecciones de la Población Mexicana 1970-2040*.

Probabilidades de entrada y salida de la actividad económica.						
Cuadro 3.13						
Grupo de edades	2000		2005		2010	
	$P(nh_x^{ia})$	$P(nh_x^{ai})$	$P(nh_x^{ia})$	$P(nh_x^{ai})$	$P(nh_x^{ia})$	$P(nh_x^{ai})$
Hombres						
12-14	0.14271	-	0.14111	-	0.13432	-
15-19	0.32592	-	0.21104	-	0.32627	-
20-29	0.30186	-	0.26465	-	0.31910	-
30-39	0.08814	-	0.27650	-	0.09679	-
40-49	-	0.05111	-	0.04954	-	0.04793
50-59	-	0.33789	-	0.33628	-	0.33522
60 y más	-	0.83428	-	0.83598	-	0.83754
Mujeres						
12-14	0.08575	-	0.08785	-	0.09096	-
15-19	0.16109	-	0.28238	-	0.17144	-
20-29	0.00000	-	0.08562	-	0.06302	-
30-39	-	0.14123	-	0.36285	-	0.19594
40-49	-	0.24338	-	0.24972	-	0.24390
50-59	-	0.47970	-	0.48447	-	0.48950
60 y más	-	0.84939	-	0.85158	-	0.85351

Fuente: Cálculos propios con datos de la Encuesta Nacional de Empleo 1997, *Proyecciones de la Población Mexicana 1970-2040* y *Proyecciones de P.E.A. Nivel Nacional 1980-2010*.

2. Ocupación []: La inexistencia de registros de la condición de ocupación anterior obligó a asumir como condiciones iniciales a la Tasa de Empleo como la probabilidad de que un individuo dentro de la PEA_d saliera para pertenecer a la PEA_o y a la Tasa de Desempleo como la probabilidad de que un individuo dentro de la PEA_o pasará a la PEA_d . Las probabilidades subsecuente se obtuvieron de la interacción del programa agregando la Condición de Ocupación Anterior a quienes ingresaron a la PEA.
3. Ramactividad []: Esta rutina fue la misma usada en *MUESTRA1*.

4. Hipótesis de Simulación

4.1. Variables demográficas

La prueba de esta hipótesis tuvo dos propósitos. Evaluar el funcionamiento del modelo para los dos posibles estados de empleo de la PEA con las mismas tasas de empleo y desempleo de la ENE 1997 para todo el periodo de simulación; y conocer el comportamiento de las variables demográficas y económicas comunes a las tres hipótesis.

La carencia de información de tipo longitudinal con respecto a datos de empleo, obligó a considerar a la Tasas de Desempleo Abierto para cada grupo quinquenal y sexo, como las probabilidades iniciales para salir del estado de Ocupado y la Tasas de Empleo correspondieron a las probabilidades para entrar a la ocupación. Solamente en el primer año se utilizó este supuesto y los precedentes años de simulación el propio modelo produjo las probabilidades correspondientes para los cuatro posibles estados.

La primera variable a considerar fue el volumen de población para cada año así como los nacimientos y defunciones. La población inicial fue 93,926,566 habitantes, sin considerar a los no especificados (cuadro 4.1).

Volúmenes Totales de Población, México

Cuadro 4.1

Año	Población Total	Hombres	Mujeres	Nacimientos	Muertes
1998	95,103,779	46,279,498	48,824,281	1,887,298	710,085
1999	96,343,610	46,935,731	49,407,878	1,936,140	696,309
2000	97,575,926	47,531,852	50,044,074	1,896,064	663,748
2001	98,820,766	48,120,458	50,700,308	1,871,017	626,177
2002	100,213,384	48,813,010	51,400,374	1,996,253	603,635
2003	101,517,085	49,415,393	52,101,692	1,939,897	636,196
2004	102,912,207	50,071,626	52,840,581	1,956,177	561,055
2005	104,223,422	50,677,766	53,545,657	1,901,074	589,859
2006	105,577,217	51,385,346	54,191,872	1,952,420	598,625
2007	106,874,656	52,024,046	54,850,610	1,902,326	604,887
2008	108,145,796	52,684,037	55,461,759	1,863,503	592,364
2009	109,421,945	53,307,709	56,114,235	2,063,880	787,731
2010	110,753,197	54,014,037	56,739,160	2,180,349	849,096

Las tendencias demográficas no son visibles en periodos cortos de tiempo. Por esta razón sólo se presentan las pirámides de población para los años 2000, 2005 y 2010 (figs. 4.1, 4.2 y 4.3).

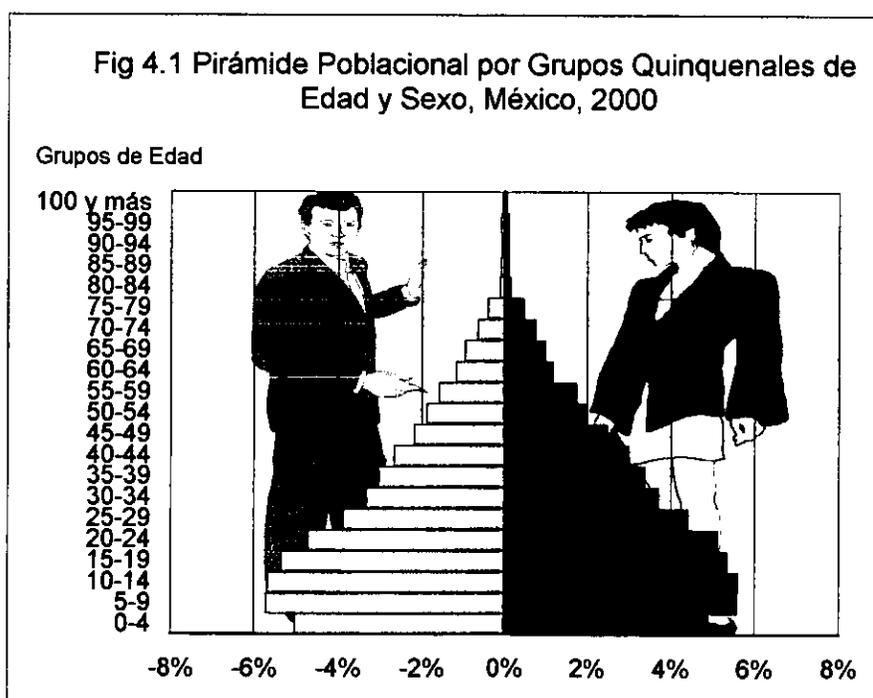


Fig 4.2 Pirámide Poblacional por Grupos Quinquenales de Edad y Sexo, México, 2005

Grupos de Edad

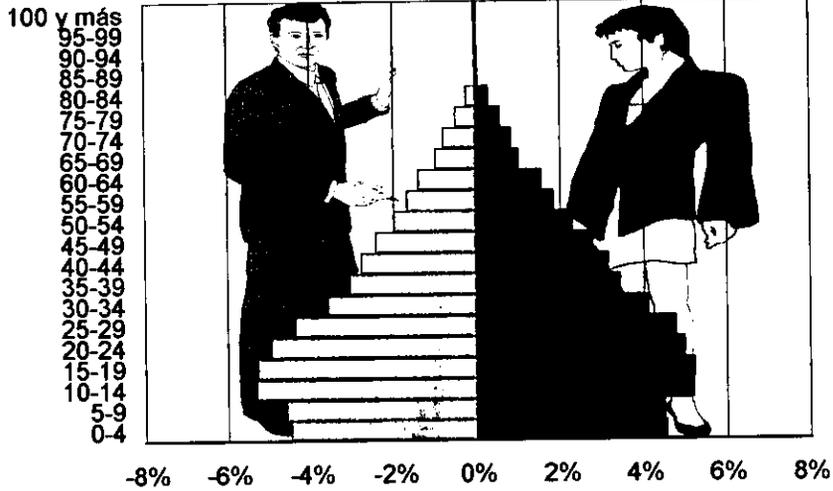
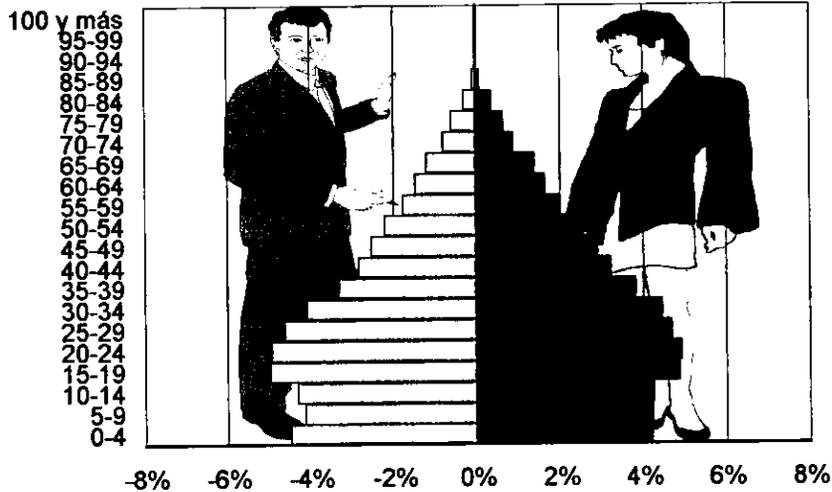
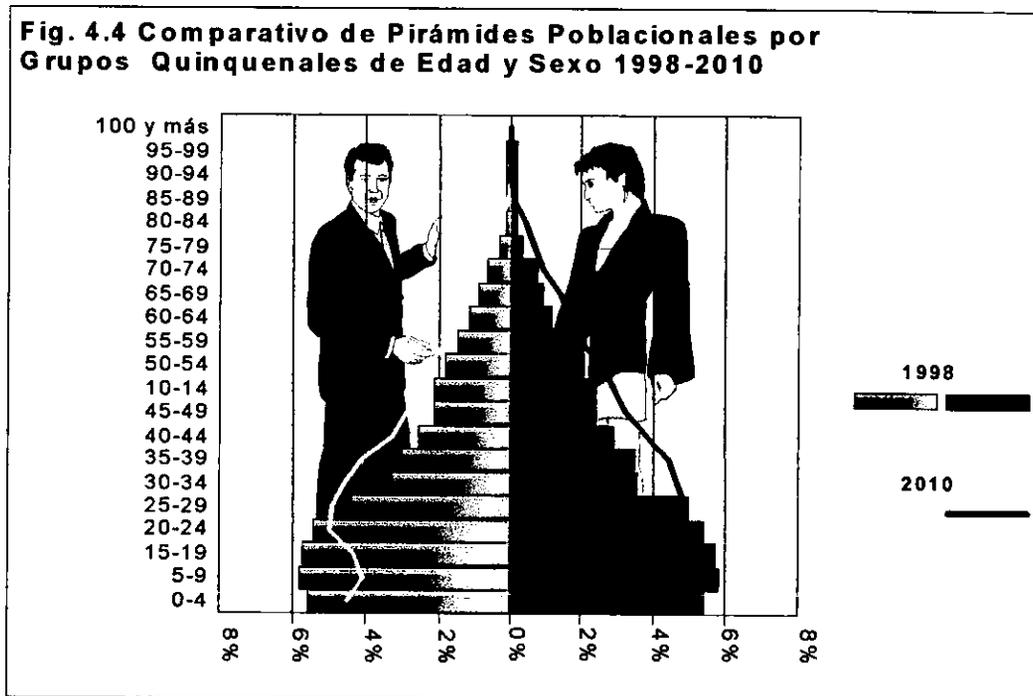


Fig. 4.3 Pirámide Poblacional por Grupos Quinquenales de Edad y Sexo, México, 2010

Grupos de Edad

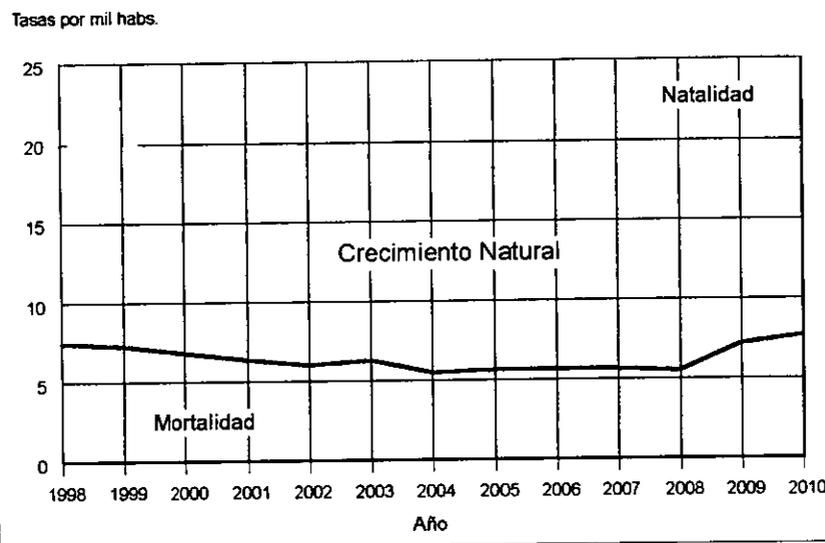


En la figura 4.4 se presentan la pirámide de población del primer periodo de simulación 1998, y la silueta de la pirámide del año 2010. De manera gráfica se corrobora la tendencia esperada del estrechamiento de la base de la pirámide y la transferencia de efectivos de los grupos quinquenales de edades activas y de 65 ó más años.



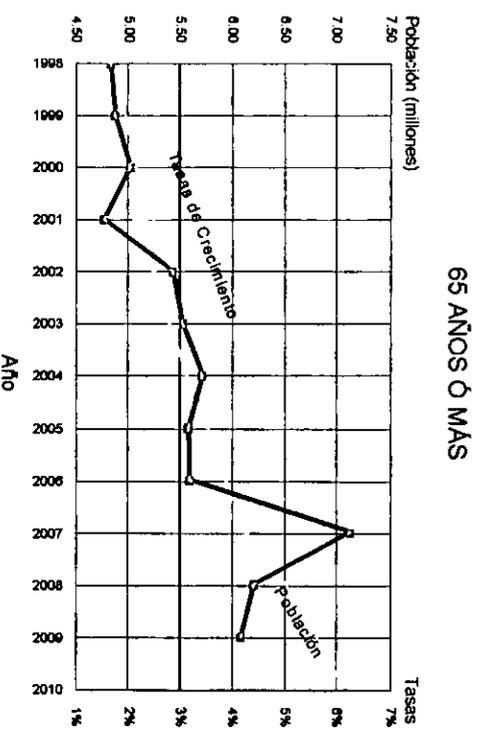
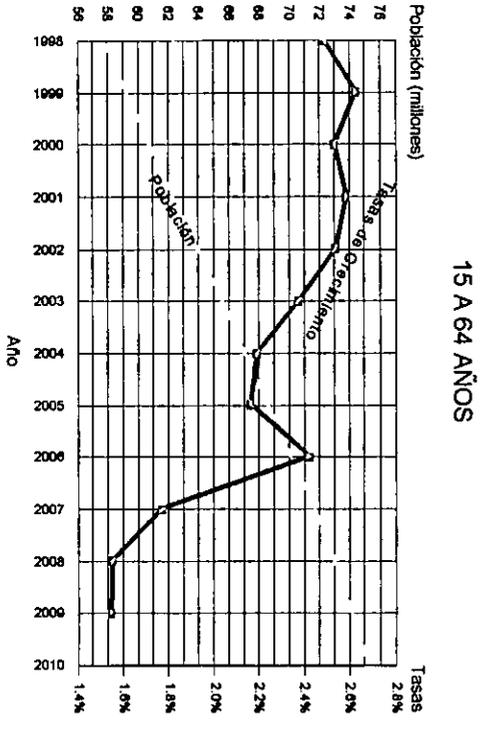
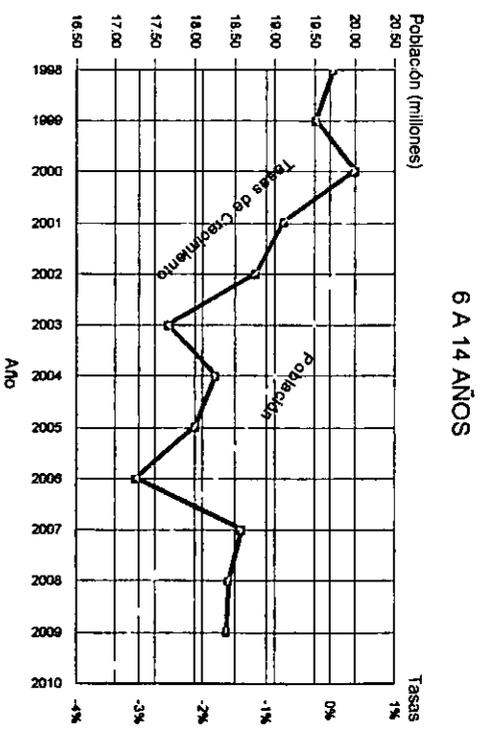
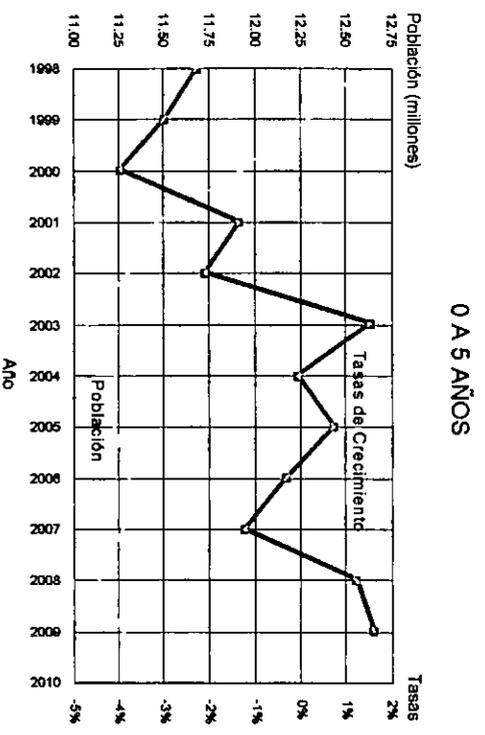
Unos de los factores más importantes para el crecimiento de la población, la natalidad, demostró una tendencia decreciente en sus tasas por mil habitantes, salvo en el año 2002 donde hubo un pequeño incremento para continuar disminuyendo hasta el año 2008, a partir de dicho año se ubicó la tasa de natalidad en su nivel más bajo para llegar en el 2010 a los mismos niveles de 1998. Con respecto a la mortalidad, sus tasas reflejaron trayectorias más o menos simétricas a las tasa de natalidad. Un hecho que podría explicar el incremento de las tasas de natalidad y mortalidad, es el aumento del volumen de población y en especial en las edades fértiles para las mujeres y en las edades de 65 o más años para ambos sexo (fig. 4.5).

Fig. 4.5 Tasas Brutas de Natalidad y de Mortalidad 1998-2010



La figura 4.8 y el cuadro 4.2 muestran los cambios observados en los grupos de edad por tamaño y tasas de crecimiento. En el grupo de 0 a 5 años señala un marcada tendencia a la disminución, las tasas de crecimiento pasan de negativas hasta alcanzar una tasa cercana al 1.5%. Para el grupo de 6 a 14 años, que podría interpretarse como la población donde se deben alcanzar los niveles mayores en cuanto a la absorción escolar, la tasas de crecimiento decrecen de manera irregular pero es clara la reducción del total de la población en este grupo. De manera contraria, en el grupo de 15-64 años, donde se concentra la mayor proporción de la fuerza laboral se verifica un incremento casi exponencial, a pesar de disminución de las tasas de crecimiento. En la población de 65 ó más años se observó una incremento tanto en su volumen de población como en las tasa porcentuales de crecimiento. Las perspectivas mostradas en la simulación, aunque en cierto grado tentativas pero no por eso alejadas de la realidad por las bases teóricas con las que se diseñó, sugieren un reorientación en cuanto a la satisfacción de las necesidades de la población. En la materia educativa la prioridad hacia la educación primaria debe ser cambiada hacia la consecución de mayores espacios y recursos para la educación media superior y superior. En el sector de la salud, problema central debe fijarse hacia la población

4.6 Tamaño y Crecimiento de la Población por Grupos de Edad, 1998-2010



Cuadro 4.2

Indicador	1998	1999	2000	2001	2002	2003	2004
Población Total	95,103,779	96,343,610	97,575,926	98,820,766	100,213,384	101,517,085	102,912,207
Porcentaje de Hombres	46,279,498	46,935,731	47,531,852	48,120,458	48,813,010	49,415,393	50,071,626
Porcentaje de Mujeres	48,824,281	49,407,878	50,044,074	50,700,308	51,400,374	52,101,692	52,840,581
Porcentaje de Hombres	48.66%	48.72%	48.71%	48.69%	48.71%	48.68%	48.65%
Porcentaje de Mujeres	51.34%	51.28%	51.29%	51.31%	51.29%	51.32%	51.35%
Población menor de un año (infantil)	1,887,298	1,936,140	1,896,064	1,871,017	1,996,253	1,939,897	2,018,795
Porcentaje de población menor de un año	1.98%	2.01%	1.94%	1.89%	1.99%	1.91%	1.96%
Población de 0 a 4 años	10,482,205	10,164,107	9,690,717	9,392,657	9,252,393	9,294,973	9,417,704
Porcentaje de población de 0 a 4 años	11.02%	10.55%	9.93%	9.50%	9.23%	9.16%	9.15%
Población de 0 a 14 años (niños y niñas)	32,432,217	32,154,194	31,742,170	31,343,921	31,044,609	30,574,976	30,248,111
Porcentaje de población de 0 a 14 años	34.10%	33.37%	32.53%	31.72%	30.98%	30.12%	29.39%
Población de 6 a 14 años (en edad escolar)	19,844,805	19,853,571	19,808,467	19,884,880	19,739,607	19,507,922	19,011,989
Porcentaje de población de 6 a 14 años	20.87%	20.61%	20.30%	20.12%	19.70%	19.22%	18.47%
Población de 12 años y más (en edad de trabajar)	69,086,120	70,704,162	72,444,934	74,185,707	75,831,300	77,538,259	79,327,873
Porcentaje de población de 12 años y más	72.64%	73.39%	74.24%	75.07%	75.67%	76.38%	77.08%
Población de 15 años y más	62,671,562	64,189,415	65,833,756	67,476,845	69,168,776	70,942,109	72,664,096
Porcentaje de población de 15 años y más	65.90%	66.63%	67.47%	68.28%	69.02%	69.88%	70.61%
Población de 18 años y más (en edad de votar)	56,448,614	57,844,989	59,357,833	61,033,483	62,671,562	64,344,707	65,964,001
Porcentaje de población de 18 años y más	59.35%	60.04%	60.83%	61.76%	62.54%	63.38%	64.10%
Población de 15 a 29 años (joven)	27,060,870	27,560,559	28,191,746	28,785,362	29,436,586	29,923,752	30,561,200
Porcentaje de población de 15 a 29 años	28.45%	28.61%	28.89%	29.13%	29.37%	29.48%	29.70%
Población de 15 a 64 años (en edad productiva)	57,772,352	59,208,802	60,765,479	62,304,622	63,916,402	65,539,453	67,096,129
Porcentaje de población de 15 a 64 años	60.75%	61.46%	62.28%	63.05%	63.78%	64.56%	65.20%
Población de 65 años y más (de la tercera edad)	4,899,210	4,980,613	5,068,278	5,172,223	5,252,374	5,402,656	5,567,967
Porcentaje de población de 65 años y más	5.15%	5.17%	5.19%	5.23%	5.24%	5.32%	5.41%

Cuadro 4.2 (continuación)

Indicador	2005	2006	2007	2008	2009	2010
Población Total	104,223,422	105,577,217	106,874,656	108,145,796	109,421,945	110,753,197
Porcentaje de Hombres	50,677,766	51,385,346	52,024,046	52,684,037	53,307,709	54,014,037
Porcentaje de Mujeres	53,545,657	54,191,872	54,850,610	55,461,759	56,114,235	56,739,160
Porcentaje de Hombres	48.62%	48.67%	48.68%	48.72%	48.72%	48.77%
Porcentaje de Mujeres	51.38%	51.33%	51.32%	51.28%	51.28%	51.23%
Población menor de un año (infantil)	1,901,074	1,952,420	1,902,326	1,863,503	2,063,880	1,792,119
Porcentaje de población menor de un año	1.82%	1.85%	1.78%	1.72%	1.89%	1.62%
Población de 0 a 4 años	9,426,470	9,512,883	9,386,395	9,294,973	9,358,843	9,630,604
Porcentaje de población de 0 a 4 años	9.04%	9.01%	8.78%	8.59%	8.55%	8.70%
Población de 0 a 14 años (niños y niñas)	29,899,957	29,588,121	28,997,009	28,612,537	28,471,021	28,372,085
Porcentaje de población de 0 a 14 años	28.69%	28.03%	27.13%	26.46%	26.02%	25.62%
Población de 6 a 14 años (en edad escolar)	18,671,349	18,276,857	17,720,812	17,474,098	17,196,076	16,915,548
Porcentaje de población de 6 a 14 años	17.91%	17.31%	16.58%	16.16%	15.72%	15.27%
Población de 12 años y más (en edad de trabajar)	80,945,915	82,537,657	84,295,962	85,924,023	87,236,490	88,262,168
Porcentaje de población de 12 años y más	77.67%	78.18%	78.87%	79.45%	79.72%	79.69%
Población de 15 años y más	74,323,465	82,537,657	77,843,833	79,533,259	80,950,924	82,381,113
Porcentaje de población de 15 años y más	71.31%	78.18%	72.84%	73.54%	73.98%	74.38%
Población de 18 años y más (en edad de votar)	67,668,455	69,400,461	71,202,599	72,913,315	74,412,383	75,974,068
Porcentaje de población de 18 años y más	64.93%	65.73%	66.62%	67.42%	68.00%	68.60%
Población de 15 a 29 años (joven)	30,909,354	31,322,631	31,923,761	31,912,490	31,981,370	32,184,251
Porcentaje de población de 15 a 29 años	29.66%	29.67%	29.87%	29.51%	29.23%	29.06%
Población de 15 a 64 años (en edad productiva)	68,565,141	70,049,181	71,748,625	73,019,765	74,150,641	75,297,797
Porcentaje de población de 15 a 64 años	65.79%	66.35%	67.13%	67.52%	67.77%	67.99%
Población de 65 años y más (de la tercera edad)	5,758,325	5,939,916	6,129,022	6,513,494	6,800,283	7,083,315
Porcentaje de población de 65 años y más	5.52%	5.63%	5.73%	6.02%	6.21%	6.40%

envejecida donde se observará un crecimiento sostenido por el aumento de las esperanza de vida. El problema de mayor magnitud será la oferta de empleo para la creciente fuerza laboral.

Los índices de masculinidad del cuadro 4.2 se calcularon para validar la distribución por sexo en agrupaciones de edad, de acuerdo a los conceptos de población joven, en edad de votar, etcétera.

	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Índice de masculinidad de la población menor de un año	106.7	100.3	107.4	101.1	108.1	98.1	101.2	94.9	110.4	103.3	112.3	109.1	102.7
Índice de masculinidad de la Población de 0 a 4 años	103.9	103.6	102.7	101.7	102.3	101.1	101.7	98.8	101.2	100.0	104.0	105.7	106.6
Índice de masculinidad de la población de 0 a 14 años	108.9	101.8	102.1	102.2	101.9	101.8	101.4	101.2	101.3	101.6	102.3	102.8	102.2
Índice de masculinidad de la población de 6 a 14 años	100.5	101.4	101.2	102.1	101.8	102.2	101.8	101.9	101.7	101.8	102.4	101.8	100.7
Índice de masculinidad de la población de 12 años y más	92.1	92.7	92.6	92.7	92.6	92.9	92.8	92.9	92.9	93.2	93.3	93.2	93.5
Índice de masculinidad de la población de 15 años y más	91.4	91.7	91.7	91.7	92.0	92.0	92.1	92.1	79.3	92.5	92.5	92.4	92.9
Índice de masculinidad de la población de 18 años y más	90.3	90.6	90.6	91.0	91.0	91.1	91.1	91.4	91.5	91.7	91.7	91.9	92.1
Índice de masculinidad de la población de 15 a 29 años	92.5	93.2	93.7	94.6	96.1	96.4	97.7	97.5	98.9	99.3	99.8	99.8	100.5
Índice de masculinidad de la población de 15 a 64 años	91.5	91.8	91.8	91.8	92.0	92.1	92.3	92.3	92.5	92.5	92.7	92.8	93.4
Índice de masculinidad de la población de 65 años y más	90.5	90.9	90.8	89.9	91.2	91.1	89.9	90.4	90.7	91.3	89.8	87.9	87.3

La paridez fue la característica determinante para el crecimiento de la población. Aunque existe una relación entre la condición de ocupación y el número de hijo nacidos vivos de la población femenina, es decir históricamente la PEI femenina tiende a procrear un mayor número de hijos, la simulación utilizó Tasas de Fecundidad sin diferenciar la condición de ocupación. En los cuadros siguientes se muestran algunos indicadores relacionados con la fecundidad.

Población Femenina de 12 años y más, según número de hijos nacidos vivos, 2000
Cuadro 4.4

Grupos Quinquenales de Edad	0 hijos	1 a 2 hijos	3 a 5 hijos	6 y más hijos
12-14	3,278,663	0	0	0
15-19	4,348,174	845,339	33,814	0
20-24	2,394,501	2,151,545	463,371	2,505
25-29	1,263,625	2,052,609	975,584	25,047
30-34	648,719	1,566,695	1,298,691	123,983
35-39	350,659	1,072,015	1,560,433	329,369
40-44	210,396	740,141	1,423,927	572,326
45-49	121,478	542,269	1,135,885	634,944
50 años y más	444,586	1,083,286	2,308,089	3,615,547

Población Femenina de 12 años y más, según número de hijos nacidos vivos, 2005
Cuadro 4.5

Grupos Quinquenales de Edad	0 hijos	1 a 2 hijos	3 a 5 hijos	6 y más hijos
12-14	3,276,159	0	0	0
15-19	4,551,055	910,462	3,757	0
20-24	2,276,780	2,637,458	291,799	0
25-29	1,089,548	2,829,068	1,055,735	1,252
30-34	856,610	2,022,552	1,381,347	23,795
35-39	508,456	1,529,124	1,447,721	122,731
40-44	316,846	1,038,202	1,599,256	325,612
45-49	206,638	711,337	1,420,170	563,559
50 años y más	512,213	1,505,330	3,203,522	3,743,287

Población Femenina de 12 años y más, según número de hijos nacidos vivos, 2010
Cuadro 4.6

Grupos Quinquenales de Edad	0 hijos	1 a 2 hijos	3 a 5 hijos	6 y más hijos
12-14	2,909,219	0	0	0
15-19	4,509,728	880,405	6,262	0
20-24	2,448,352	2,778,974	230,433	0
25-29	1,112,091	3,059,501	1,026,930	0
30-34	710,085	2,671,272	1,552,919	2,505
35-39	691,300	1,972,458	1,561,686	23,795
40-44	469,633	1,480,283	1,514,096	116,469
45-49	308,079	1,016,912	1,582,976	313,089
50 años y más	648,719	2,063,880	4,244,228	3,711,978

Cuadro 4.7	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007
Tasa Bruta de Natalidad (por mil habitantes)	19.84	20.10	19.43	18.93	19.92	19.11	19.62	18.24	18.49	17.80

	2008	2009	2010
Tasa Bruta de Natalidad (por mil habitantes)	17.23	18.86	19.69

Cuadro 4.8	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Promedio de hijos nacidos vivos de las mujeres de 12 años y más	2.56	2.52	2.48	2.44	2.40	2.38	2.34	2.31	2.29	2.25	2.23	2.21	2.19

En todo el periodo de simulación se utilizó probabilidades constantes para determinar el estado civil de la población. La población soltera no tuvo variaciones importantes. En el caso de la población unida, las edades donde alcanzó sus máximos inició en el grupo de 35 a 39 años en el año 2000 con casi un 70% del total del citado grupo para disminuir hasta un poco más del 60% del total de grupo de 25 a 29 años. Para la población separada, la curva que describió sus trayectorias tendió a hacerse mas convexa en respuesta a la disminución de la población unida y el aumento de la esperanza de vida

Fig 4.7 Distribución Porcentual por Estado Civil, 2000

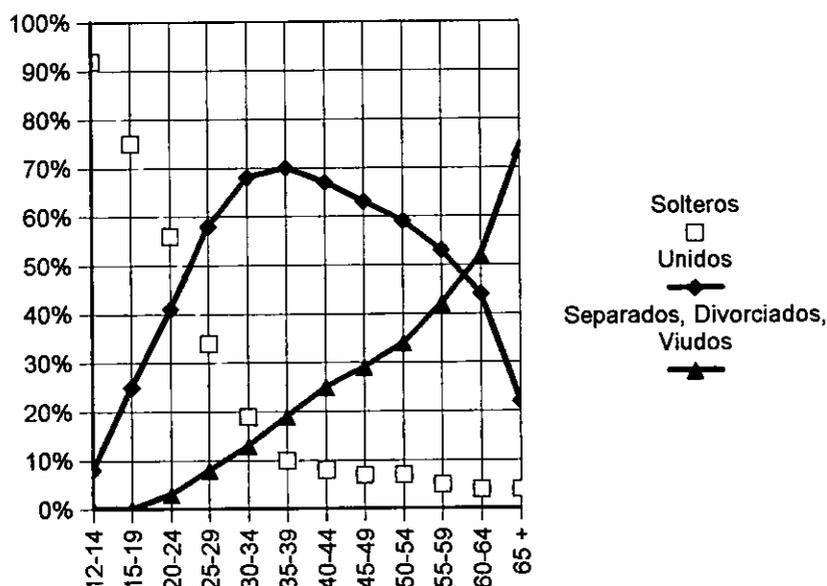


Fig 4.8 Distribución Porcentual por Estado Civil, 2005

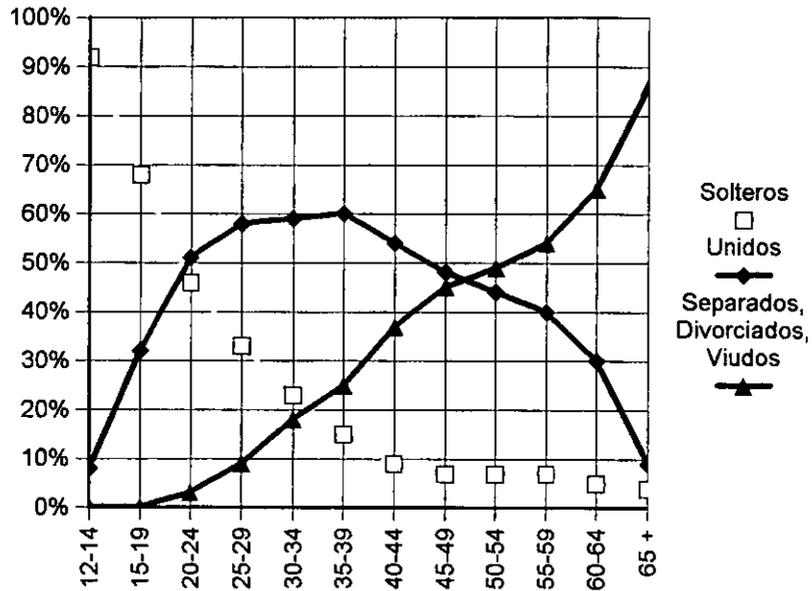
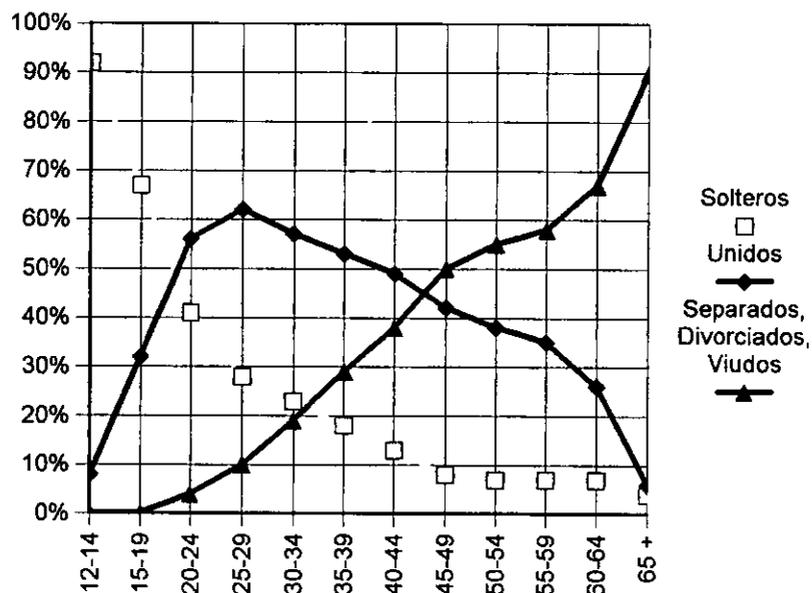


Fig 4.9 Distribución Porcentual por Estado Civil, 2010



Indicadores de Educación	Cuadro 4.7												
	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
Porcentaje de la población													
De 6 a 14 años con uno o más grados escolares aprobados	98.78%	98.56%	98.32%	98.32%	98.26%	98.09%	98.02%	97.93%	97.90%	97.94%	97.97%	98.00%	97.94%
De 15 años y más con uno o más grados escolares aprobados	89.32%	89.86%	90.37%	90.76%	91.16%	91.54%	91.89%	92.20%	92.53%	92.83%	93.13%	93.36%	93.65%
De 6 a 14 años que asiste a la escuela	94.26%	93.25%	92.28%	91.38%	90.44%	89.70%	89.47%	88.99%	86.32%	88.61%	88.70%	88.75%	88.60%
De 15 años y más que asiste a la escuela	12.82%	13.73%	14.28%	14.84%	15.36%	15.69%	15.11%	14.97%	14.41%	13.64%	13.74%	12.99%	12.32%
De 15 años y más sin instrucción	10.68%	10.14%	9.63%	9.24%	8.84%	8.46%	8.11%	7.80%	7.47%	7.17%	6.87%	6.64%	6.35%
De 15 años y más con primaria incompleta	19.37%	19.06%	18.82%	18.70%	19.17%	17.40%	17.22%	16.93%	16.64%	16.31%	16.04%	15.78%	15.49%
De 15 años y más con primaria completa	18.38%	16.87%	16.51%	18.00%	15.85%	17.65%	15.67%	15.05%	14.70%	14.32%	14.00%	13.71%	13.41%
De 15 años y más con secundaria incompleta	8.24%	10.92%	8.89%	6.68%	9.55%	10.71%	9.95%	11.21%	7.98%	8.02%	8.10%	8.15%	8.19%
De 15 años y más con secundaria completa	14.78%	12.39%	15.60%	14.61%	12.01%	11.70%	17.73%	15.95%	11.89%	11.60%	11.49%	11.41%	11.31%
De 15 años y más con rezago educativo	30.04%	29.20%	28.45%	27.94%	28.01%	25.86%	25.33%	24.73%	24.11%	23.48%	22.91%	22.43%	21.84%
De 15 años y más con educación postprimaria	51.58%	53.93%	55.04%	54.06%	56.14%	56.49%	59.00%	60.22%	61.19%	62.20%	63.09%	63.88%	64.75%
De 15 años y más con educación postsecundaria	28.56%	30.62%	30.55%	32.78%	34.58%	34.08%	31.33%	37.41%	37.61%	42.57%	43.50%	44.30%	45.26%
De 20 años y más con algún grado aprobado en educación superior o posgrado	11.78%	11.50%	12.24%	12.46%	13.99%	14.96%	15.04%	17.11%	18.11%	17.93%	18.04%	19.88%	20.42%
Promedio de escolaridad de la población de 15 años y más	8.27	8.46	8.64	8.80	8.96	9.11	9.19	9.38	9.50	9.59	9.71	9.79	9.88

El cuadro 4.7 señala indicadores de educación en el periodo 1998-2010. Dichos indicadores se produjeron a partir de una tabla de aprobación-deserción escolar con información de la ENECE-1997. Las expectativas en la educación son el abatimiento de los niveles de analfabetismo, cubrir las demandas de educación postsecundaria entre muchas otras, pero esta tabla sólo contempla cambios demográficos y ante probabilidades altas de aprobación en los niveles de educación primaria y secundaria aunado a el deceso de personas con bajos niveles educativos. Tradicionalmente el desempeño de los programas educativos nacionales se ha medido con los grados aprobados y nunca en la calidad de la misma. Puede ser subjetivo medir la educación de esta manera pero para los fines de esta investigación, el nivel educativo de los individuos fue el determinante para asignar la rama de actividad a quienes conseguían ser empleados.

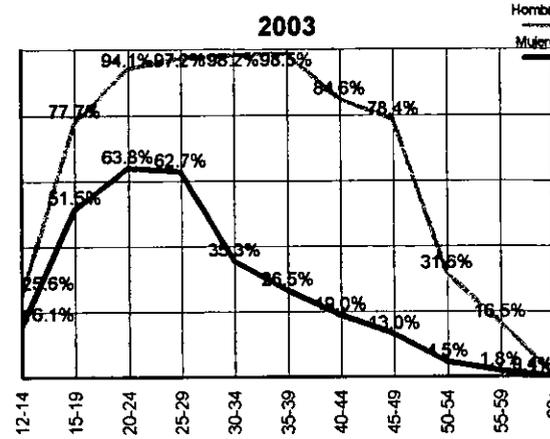
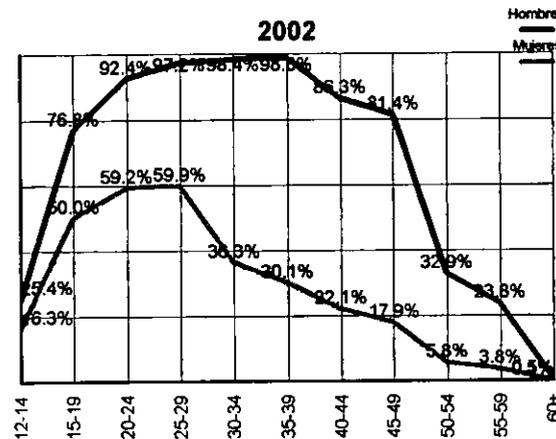
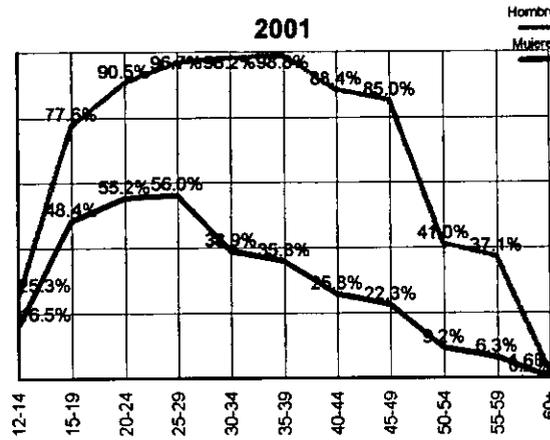
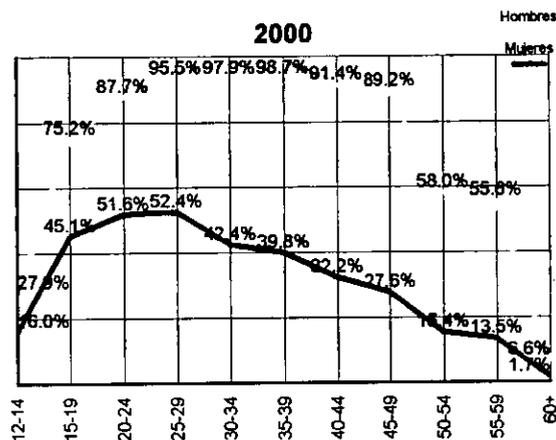
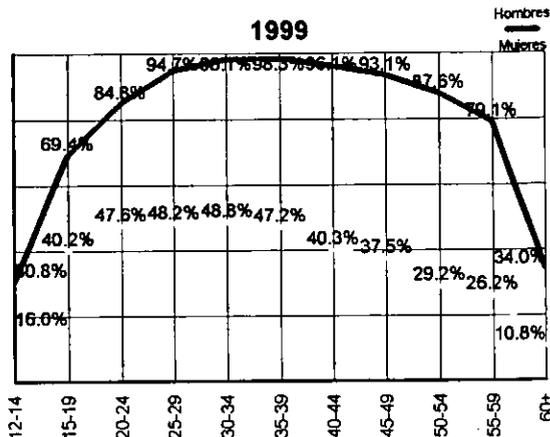
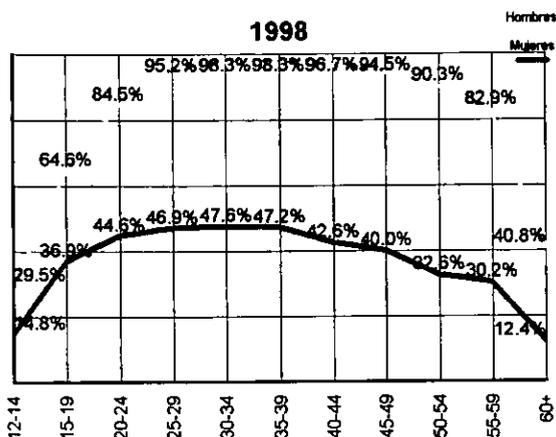
4.2 Variables económicas

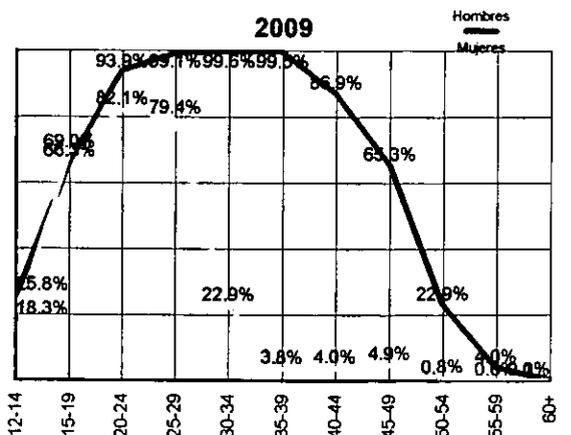
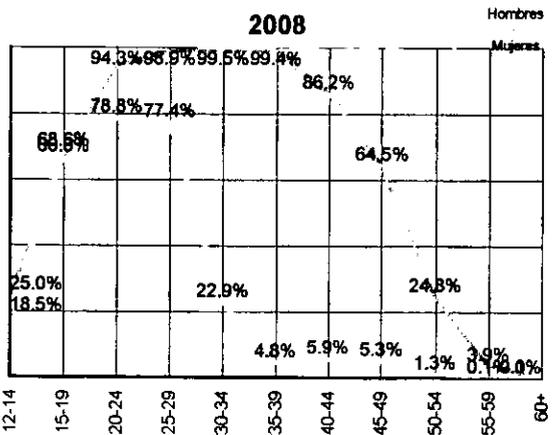
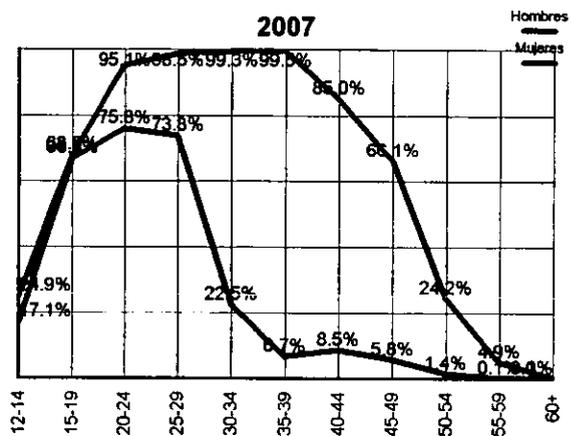
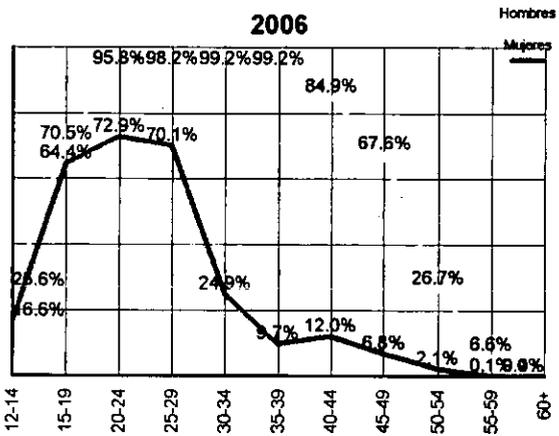
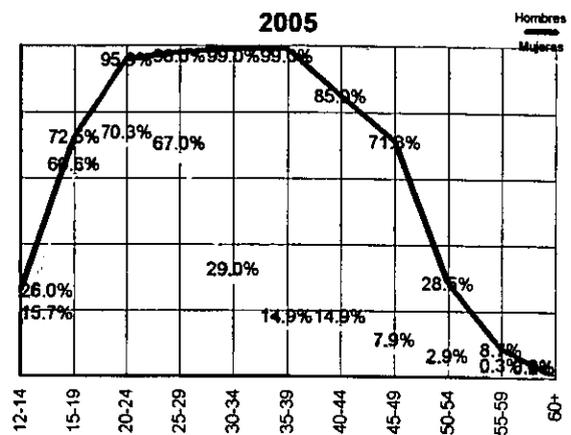
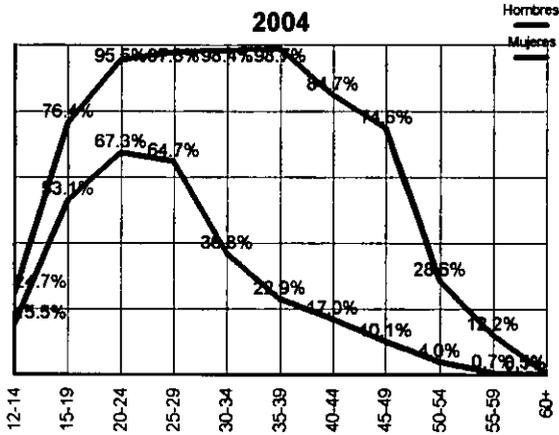
De manera opuesta a lo que sucede con las variables demográficas de la población, las variables económicas responden rápidamente a políticas, recesiones, crisis económicas que las afectan negativamente.

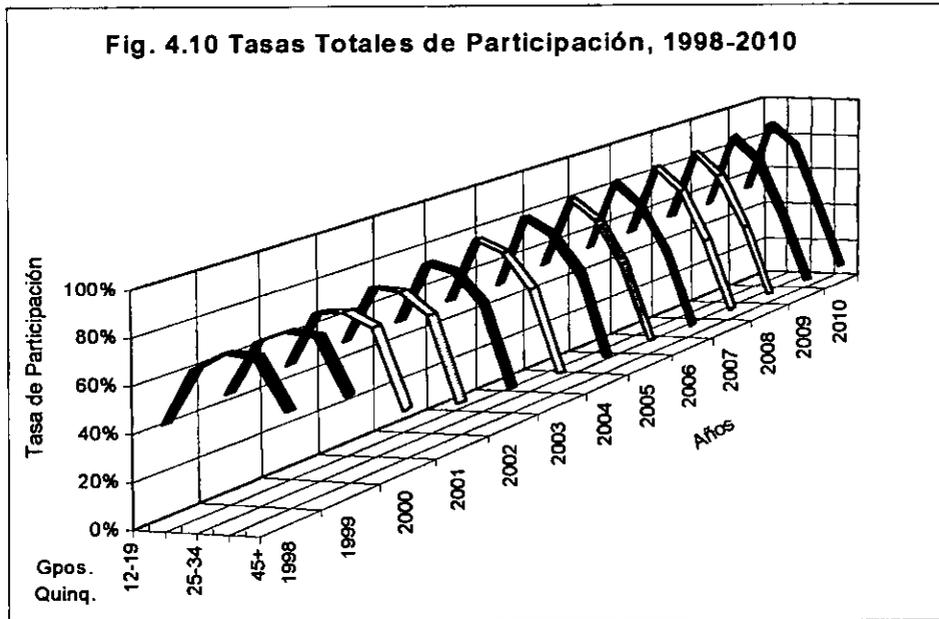
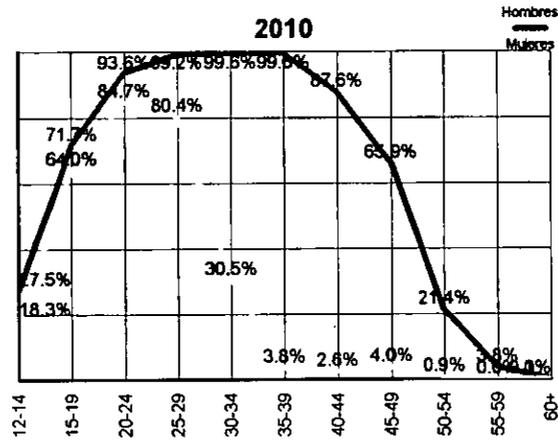
Las tasas de participación de la figura 4.9 están orientadas a el descenso de la edades para el retiro de la actividad. Durante todo el proceso de simulación la curva de participación masculina siempre se mantuvo por arriba de la femenina. El cambio más importantes en la PEA masculina fue la reducción en las edades de 40 años o más, los grupos restantes mantuvieron una tendencia relativamente uniforme.

Los cambio más notorios aparecen en la PEA femenina, en 1998 las mayores tasas de participación se localizaron entre las edades de 20 a 44 años, para después disminuir rápidamente y encontrar los máximos niveles entre las edades de 15 a 29 años no sólo en la edad de actividad máxima sino también en los niveles de participación de la población total en tales edades. En contraste, a partir del 2003 las edades de 40 años o más, la PEA femenina no rebasó el 20% y la edad para la salida de la actividad inició en los 50 años.

Fig 4.9 Tasas de Participación por Grupos Quinquenales de Edad y Sexo, 1998-2010







4.3. Hipótesis I: tasa de desempleo abierto para la generación de probabilidades de empleo y desempleo

La Tasa de Desempleo Abierta (TDA) es el indicador oficial para medir la evolución del mercado laboral. Los indicadores de empleo mostrados en el cuadro 4.8 resumen en comportamiento de las variables económicas con relación a las variables sociodemográficas. La escolaridad, el factor determinante para el ingreso a la rama de actividad, derivado del crecimiento de los niveles de escolaridad provocó que la

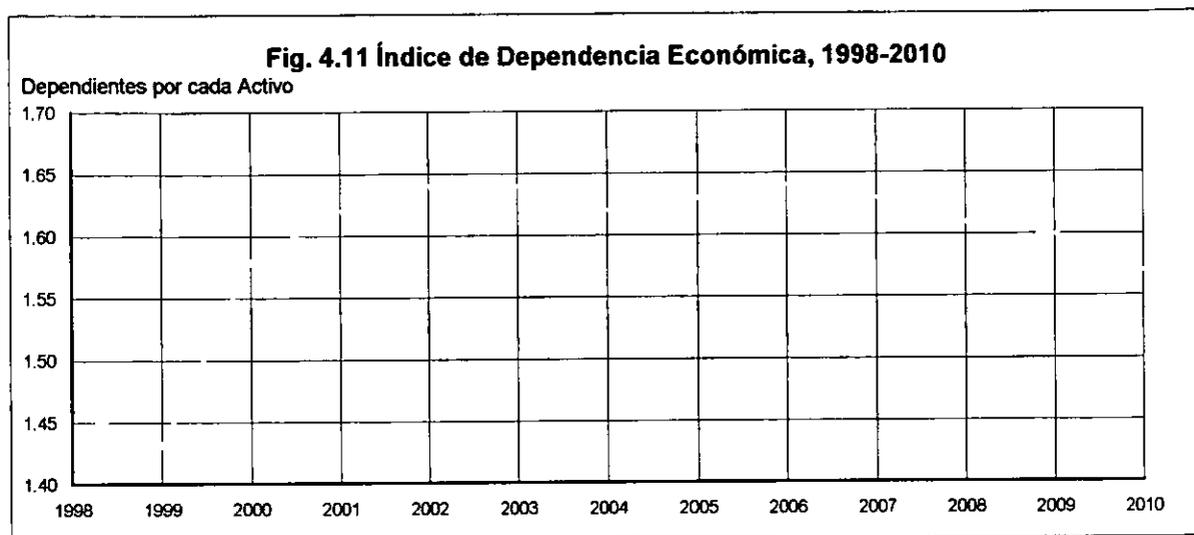
participación en el sector primario fue reduciéndose y los sectores secundario y terciario, donde los niveles de escolaridad requeridos fueron superiores a 8 grados escolares.

Indicadores de Empleo	1998	1999	2000	2001	2002	2003	2004
Población de 12 años y más	72.64%	73.39%	74.24%	75.07%	75.67%	76.38%	77.08%
Población Económicamente Activa	40.68%	41.21%	38.80%	38.11%	37.80%	37.94%	38.13%
Índice de Dependencia Económica	1.46	1.43	1.58	1.62	1.64	1.64	1.62
PEA en el Sector Primario	20.26%	20.44%	19.60%	19.24%	19.03%	18.57%	18.18%
PEA en el Sector Secundario	20.35%	20.30%	20.58%	20.51%	20.37%	20.67%	20.67%
PEA en el Sector Terciario	59.39%	59.26%	59.82%	60.25%	60.60%	60.77%	61.15%
Tasa de Desempleo Abierto Hombres	1.89%	2.59%	2.90%	3.07%	3.33%	3.74%	4.24%
Tasa de Desempleo Abierto Mujeres	3.86%	4.05%	4.33%	4.62%	5.64%	5.32%	5.13%
Tasa de Desempleo Abierto General	2.56%	3.09%	3.38%	3.59%	4.10%	4.27%	4.54%
Ocupados en el Comercio	16.60%	16.69%	16.97%	17.19%	17.16%	17.63%	17.25%
Población Económicamente Inactiva	44.01%	43.85%	47.74%	49.24%	50.05%	50.33%	50.54%
Mujeres Activas Unidas	54.07%	50.38%	47.98%	46.80%	46.16%	46.26%	47.12%
Mujeres Activas con menos de 3 hijos	66.16%	68.52%	74.75%	78.32%	80.89%	82.99%	84.04%

Indicadores de Empleo	2005	2006	2007	2008	2009	2010
Población de 12 años y más	77.67%	78.18%	78.87%	79.45%	79.72%	79.69%
Población Económicamente Activa	37.99%	37.92%	38.06%	38.33%	38.44%	38.92%
Índice de Dependencia Económica	1.63	1.64	1.63	1.61	1.60	1.57
PEA en el Sector Primario	17.83%	17.46%	16.79%	16.08%	15.87%	15.57%
PEA en el Sector Secundario	20.60%	20.45%	20.61%	20.47%	20.41%	20.34%
PEA en el Sector Terciario	61.57%	62.09%	62.60%	63.45%	63.72%	64.09%
Tasa de Desempleo Abierto Hombres	4.69%	4.74%	5.14%	5.33%	5.26%	5.31%
Tasa de Desempleo Abierto Mujeres	5.26%	4.55%	4.03%	3.95%	3.99%	3.91%
Tasa de Desempleo Abierto General	4.88%	4.68%	4.76%	4.85%	4.82%	4.83%
Ocupados en el Comercio	17.58%	17.99%	18.22%	18.31%	18.43%	18.74%
Población Económicamente Inactiva	51.08%	51.50%	51.74%	51.76%	51.79%	51.16%
Mujeres Activas Unidas	46.95%	47.08%	47.68%	48.38%	49.01%	49.97%
Mujeres Activas con menos de 3 hijos	85.70%	87.18%	88.55%	88.99%	89.14%	88.57%

El Índice de Dependencia Económica (IDE) mide la cantidad de dependientes que soporta económicamente cada persona activa. Esta dependencia económica es un sentido indirecto, al ser la PEA quien sostiene la economía y la generadora de los recursos vía impuestos y aportaciones para fondos de pensiones y no con relación a los menores, amas de casa, incapacitados o personas dentro de la clasificación de la PEI que cada activo debe sostener económicamente y dedicarle algún cuidado especial. La siguiente figura (4.11) muestra el crecimiento del índice por los factores

ya mencionados de la reducción de las edades de retiro y la postergación de las edades para la entrada a la actividad.



La figura 4.12 muestra las trayectorias de las tasas de desempleo, donde de 1998 a 2005 la tasa de desempleo femenina fue mayor a la masculina, situación invertida a partir de 2006 cuando la reducción de las tasas de participación femeninas redujo las tasa de desempleo.

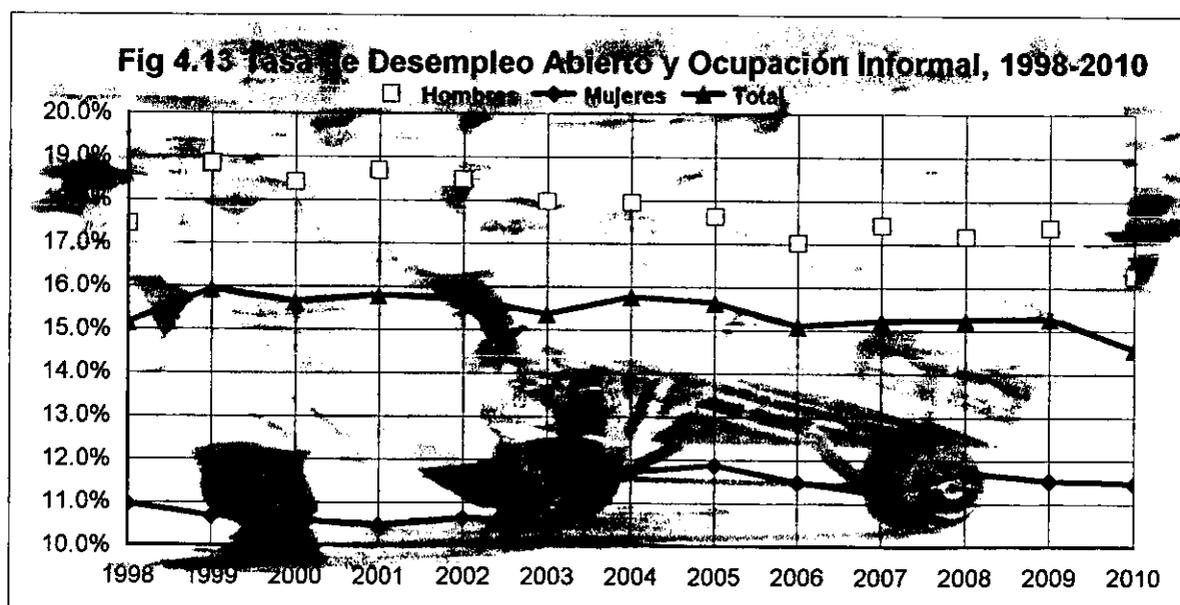


4.4. Hipótesis II: tasa de desempleo abierto y ocupación informal para la generación de probabilidades de empleo y desempleo

El uso de la Tasa de Desempleo Abierto y Ocupación Informal (TDAOI) como condiciones iniciales para las probabilidades de empleo y desempleo esta basada en la declaración de J. Rifkin¹, indicando que en 1994 cerca de 50% de la PEA mexicana se encontraba desempleada o subempleada. La TDAOI no la calcula el INEGI, las tasas respectivas se obtuvieron al sumar la TDA y la proporción de la población ocupada sin local. La desagregación por grupos quinquenales se logro al multiplicar la TDAOI por las tasas de participación de 1997. En cuadro 4.9 se resumen las TDAOI anuales producidas por la simulación.

	1998	1999	2000	2001	2002	2003	2004	2005	2006
Hombres	17.45%	18.86%	18.42%	18.69%	18.48%	17.98%	17.95%	17.64%	17.03%
Mujeres	10.95%	10.67%	10.59%	10.47%	10.63%	10.52%	11.71%	11.87%	11.49%
Total	15.16%	15.92%	15.65%	15.79%	15.70%	15.36%	15.76%	15.63%	15.09%

	2007	2008	2009	2010
Hombres	17.43%	17.18%	17.39%	16.28%
Mujeres	11.21%	11.71%	11.51%	11.45%
Total	15.22%	15.24%	15.28%	14.57%



¹ Rifkin, Jeremy. *El Fin del Trabajo*.

Las TDAOI's no mostraron variaciones de $\pm 2\%$ con los valores al inicio de la simulación, motivado por las bajas tasas de participación. Las TDAOI masculinas tuvieron una tendencia a decrecer mientras que en las TDAOI femeninas se mantuvo decreciendo sólo hasta el año 2003 para mantener un crecimiento continuo para finalizar en un 11.45%.

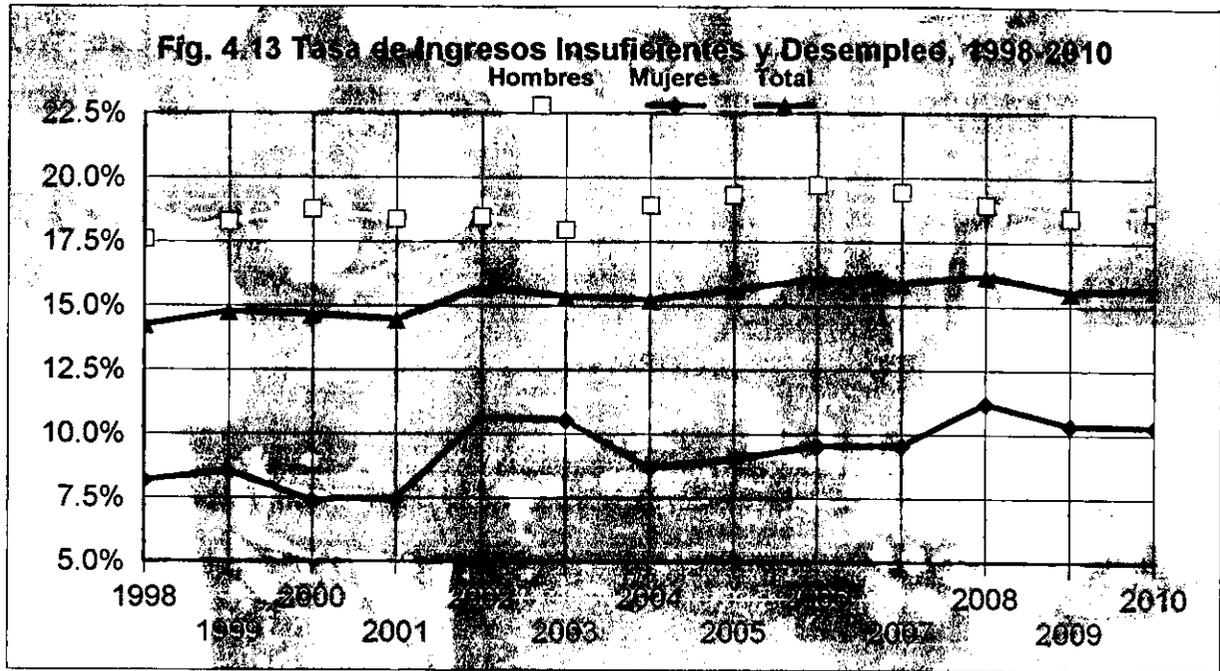
4.5. Hipótesis III: tasa de ingresos insuficientes y desocupación para la generación de probabilidades de empleo y desempleo

Debido a que las Tasas de Desempleo Abierto e Ingresos Insuficientes (TIID) son calculadas de manera global por el INEGI, se utilizó el mismo procedimiento utilizado para la desagregación de las TDAOI. La razón para considerar la tasa anterior se justifica en el hecho de que las políticas gubernamentales sólo prestan atención a los indicadores de ocupación sin tomar en cuenta las condiciones de empleo como la carencia de prestaciones laborales, bajos niveles de ingresos, etcétera.

Las TIID's masculinas casi duplicaron a las femeninas durante todo el periodo, su tendencia fue hacia un crecimiento moderado pero nunca rebasando el 20%. Las TIID's femeninas tuvieron un comportamiento más inestable, donde sus niveles máximo y mínimo fueron del 11.17% y 7.40%.

	1998	1999	2000	2001	2002	2003	2004	2005	2006
Hombres	17.60%	18.27%	18.76%	18.37%	18.48%	17.98%	18.93%	19.33%	19.71%
Mujeres	8.16%	8.54%	7.40%	7.46%	10.63%	10.52%	8.65%	9.03%	9.53%
Total	14.21%	14.74%	14.66%	14.46%	15.70%	15.36%	15.23%	15.63%	16.05%

	2007	2008	2009	2010
Hombres	19.43%	18.93%	18.44%	18.70%
Mujeres	9.55%	11.17%	10.33%	10.29%
Total	15.87%	16.16%	15.53%	15.68%



Conclusiones

Los resultados producidos por la simulación, son un prototipo del comportamiento de los indicadores de empleo basado en los supuestos ya establecidos. Las variables utilizadas fueron las necesarias para que el trabajo estuviera fundamentado, sin embargo, los factores determinantes para la evolución del empleo son los económicos debiéndose agregar las políticas gubernamentales que consideren al desempleo como un problema menor que mejorará en torno al progreso de las variables macroeconómicas. Cuestiones metodológicas de las instituciones a cargo de la medición del desempleo minimizan sus proporciones reales. La condición de empleado no es una garantía verdadera de un nivel digno de vida para la familia del trabajador. Entre 1991 y 1997, el 70% de los empleos creados han sido ocupaciones no asalariadas o en establecimientos con menos de 5 empleados. En 1997 el 65% de los asalariados recibían menos de dos salarios mínimos y el 57% de esos asalariados no contaban con ninguna prestación social.¹

¹ García Guzmán, Brígida. *Población, Trabajo y Desarrollo*.

La creación de empleos suficientes para la población que se incorpora a la actividad económica, protegidos por la seguridad social y con una distribución más justa de la riqueza, son los retos a los que tradicionalmente se ha enfrentado el país. La globalización de la economía y el desarrollo de la tecnología han motivado la pérdida de miles de empleos que jamás serán recuperados. Los sectores de servicios e industrial han venido absorbiendo a la creciente fuerza laboral urbana pero la automatización de la industria prevé que esta división de la economía nacional será de las primeras en utilizar sólo una pequeña parte de mano de obra humana. El desarrollo del sector del conocimiento (científicos, profesores, etc.) requerirá de personal con una alta preparación académica al que no podrá acceder los trabajadores desplazados por el desempleo tecnológico. La economía globalizada favorecerá a las empresas transnacionales por contar con una mayor cantidad de recursos tanto económicos como tecnológicos provocando la pérdida de empleos por la desaparición de pequeñas empresas. Las decisiones correspondientes al gobierno estarán condicionadas a los intereses de las transnacionales.

El aumento de la población 65 años o más y el incremento sostenido de la esperanza de vida causará una presión mayor para los fondos de pensión y la población en edades productivas. La creación de empleos no será el único dilema al que se enfrentará a mediano plazo, dichos empleos deberán ser suficientemente remunerados para sostener a la PEI y la formación de fondos de pensión necesarios para el mantenimiento económico y atención médica para el momento de retiro de la actividad de la propia PEA.

El medio ambiente es un aspecto al que no se le da la importancia apropiada. El incremento de la población implicará el crecimiento de centros urbanos que agrupan la mayor proporción de habitantes y la consolidación de nuevas ciudades. Los nuevos habitantes demandarán viviendas y servicios públicos, la satisfacción de estas necesidades motivará la deforestación de enormes proporciones de tierra y los consiguientes efectos de degradación del medio ambiente. La carencia de planeación de los asentamientos humanos no sólo dificultará la prestación de servicios públicos también pondrá en peligro la vida de los pobladores establecidos en zonas riesgosas.

**ESTA TESIS NO DEBE
SALIR DE LA BIBLIOTECA**

Conclusiones

Un problema adicional a los ya descritos, es el deficiente manejo de los recursos públicos destinados al mejoramiento de los niveles de vida de la población marginada y la corrupción en general que podría echar por tierra todos los esfuerzos por lograr un país con menores desigualdades y contrastes.

Apéndice A

A.1. Métodos de generación de números pseudoaleatorios

La parte central de la simulación de funciones de probabilidad radica en la generación de números aleatorios. Se pensaría que la naturaleza de los números pseudoaleatorios, del todo determinística, podría estar en contradicción del objetivo para lo que se crearon, para la modelación de funciones de probabilidad; sin embargo, con base a técnicas de análisis numérico se pudo generar sucesiones de números pseudoaleatorios que logran superar pruebas estadísticas respecto a su aleatoridad.

En los principios de la simulación se utilizaron métodos de generación de números pseudoaleatorios que comprendían técnicas manuales, procesos físicos, uso de tablas de números previamente calculados, como la elaborada por la corporación RAND en 1955 (A Million Random Digits with 100 000 Normal Deviates) y después otros métodos utilizando computadoras analógicas y en la actualidad con el uso de computadoras digitales.

A.1.1. Método de los cuadrados medios

En 1946, Von Neumann desarrollo una técnica que consistía en iniciar con un numero de 8 dígitos, tomar los cuatro dígitos centrales y elevarlos al cuadrado. Este método se llamó de cuadrados centrales.

Cuadro A.1 Método de los cuadrados centrales de Von Neumann

(No. Aleatorio) ²	No. Central	No. Aleatorio
8653 ²	74874409	8744
8744 ²	76457536	4575
4575 ²	20930625	9306
9306 ²	86601636	6016
6016 ²	36192256	1922

Con este método rápidamente se caía en ciclos repetitivos de la misma secuencia de números e incluso un mismo número.

A.1.2. Métodos congruenciales

Los métodos congruenciales de Lehmer permitieron generar números pseudoaleatorios que cumplieran con características que son deseables para cualquier generador de números pseudoaleatorios:

- Uniformemente distribuidos.
- Estadísticamente independientes.
- Facilidad para reproducirlos.
- Sin repetición dentro de una longitud determinada de sucesiones.
- Posibilidad de generar números pseudoaleatorios a gran velocidad.
- Requerimiento de un mínimo de espacio en la memoria de una computadora.

Los métodos congruenciales se basan en una relación congruencial que puede expresarse de la siguiente manera:

$$n_{i+1} = an_i + c \pmod{m} \quad (\text{A.1})$$

donde $n_i, a, c, m \in \mathbf{Z}^+$

Desarrollando (A.1)

$$\begin{aligned} n_1 &= an_0 + c \pmod{m} \\ n_2 &= an_1 + c \pmod{m} = a^2n_0 + c(a+1) \pmod{m} \\ n_3 &= a^3n_0 + c(a^2+a+1) \pmod{m} = a^3n_0 + \frac{c(a^3-1)}{(a-1)} \pmod{m} \\ &\vdots \\ &\vdots \\ &\vdots \\ n_i &= a^i n_0 + \frac{c(a^i-1)}{(a-1)} \pmod{m} \end{aligned} \quad (\text{A.2})$$

Una vez dado el valor inicial n_0 , un factor constante a y una constante aditiva c , la ecuación (A.2) conducen a una relación de congruencia módulo m para todo valor de i para toda sucesión $\{n_1, n_2, \dots, n_i, \dots\}$. Los términos siguientes de $\{n_i\}$ son enteros tal como lo determina la ecuación (A.2) y forman una sucesión de residuos módulo m , además $n_i < m \forall n_i$; a partir de los enteros en la sucesión $\{n_i\}$, se pueden obtener números racionales en el intervalo $(0, 1)$ con sólo formar la sucesión $\{r_i\} = \{\frac{n_i}{m}\}$.

Con los métodos congruenciales es imposible obtener una sucesión que no se repita; sin embargo, el periodo de una sucesión se puede fijar con un valor muy alto con sólo elegir un módulo suficiente grande.

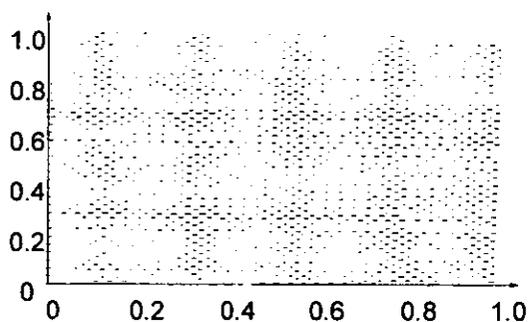


Fig. A.1 (r_i, r_{i+1}) con $a = 65$, $n_0 = 0$, $c = 1$, $m = 2048$, con 2048 puntos.

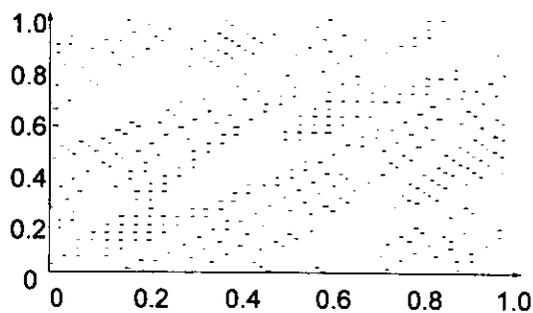


Fig. A.2 (r_i, r_{i+1}) con $a=65$, $n_0=0$, $c=1$, $m=2048$, con 512 puntos.

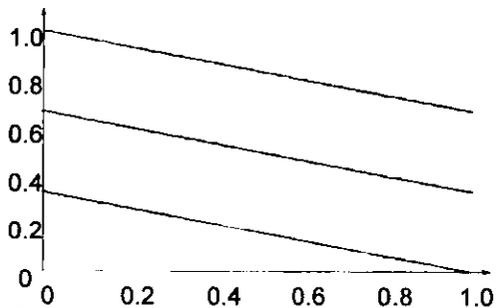


Fig. A.3 (r_i, r_{i+1}) con $a=1365$, $n_0=0$, $c=1$, $m=2048$.

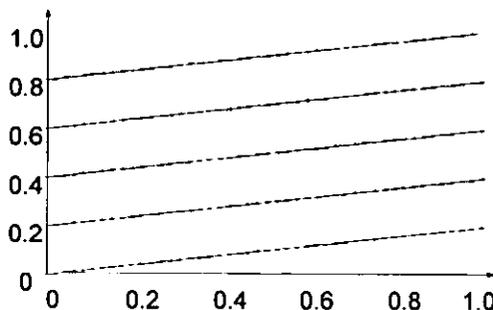


Fig. A.4 (r_i, r_{i+1}) con $a=1229$, $n_0=0$, $c=1$, $m=2048$

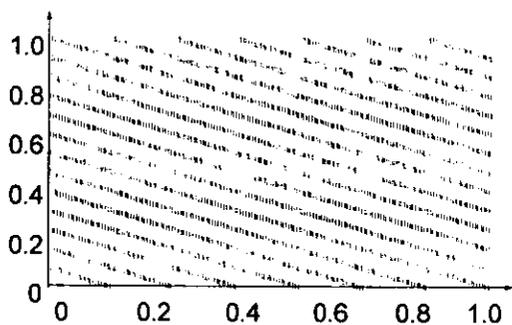


Fig. A.5 (r_i, r_{i+1}) con $a=157$, $n_0=0$, $c=1$, $m=2048$

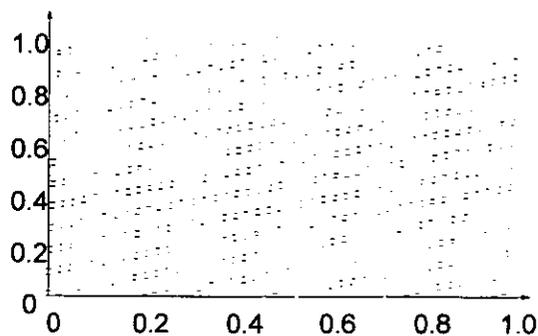


Fig. A.6 (r_i, r_{i+1}) con $a=43$, $n_0=0$, $c=1$, $m=2048$

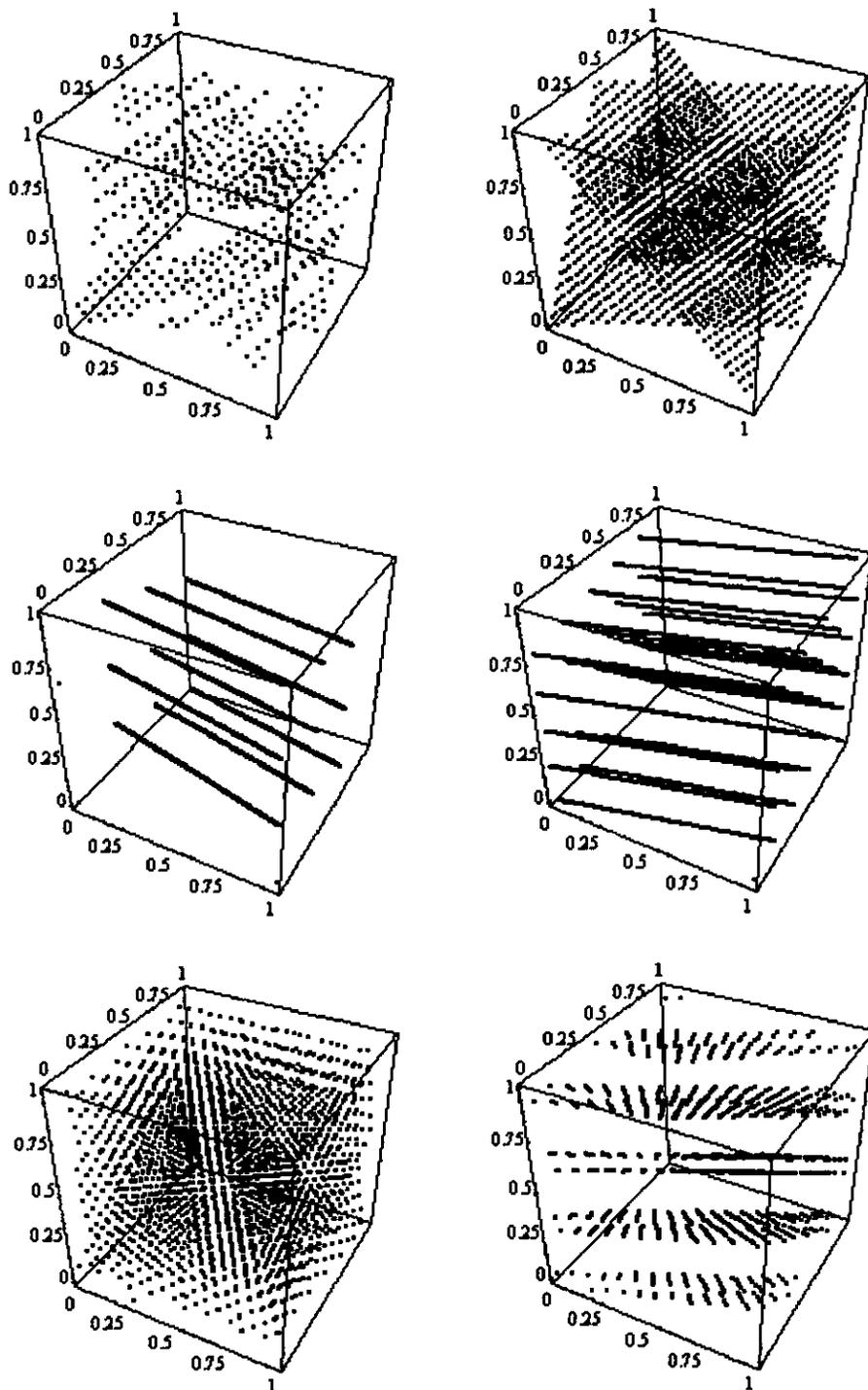


Fig. A.7 (r_i, r_{i+1}, r_{i+2}) con los parámetros a, n_0, m, c de las figuras A.1 a A.6 respectivamente.

El método aditivo de congruencias es una variación del primer método llamado congruencial compuesto. El aditivo de congruencias presupone k valores iniciales y $k \in \mathbb{Z}^+$ y calcula una sucesión de números mediante la siguiente relación congruencial:

$$n_{i+1} = n_i + n_{i+k} \pmod{m} \tag{A.3}$$

Si $k = 1$, la ecuación (A.3) se convierte en la sucesión de Fibonacci; las propiedades estadísticas de la sucesión tienden a mejorar cuando k se incrementa. Este es el único método que produce periodos mayores a m .

El método multiplicativo de congruencias calcula la sucesión $\{n_i\}$ de enteros no negativos, cada uno de los cuales siempre es menor que m , por medio de la relación de congruencia:

$$n_{i+1} = a n_i \pmod{m}$$

El método multiplicativo se comporta de manera muy satisfactoria en lo relativo a su estadística, tanto las pruebas de frecuencias, series y otras relativas a su aleatoridad indican que los números generados están uniformemente distribuidos y no correlacionados.

A.1.3. Método de multipasos digitales

Sea p un número primo pequeño (usualmente $p=2$), sea $k \in \mathbb{Z}$ y $k \geq 2$ y genera una sucesión lineal recurrente de k -ésimo orden $y_0, y_1, \dots \in \mathbb{Z}_p$

$$y_{n+k} \equiv \sum_{l=0}^{k-1} a_l y_{n+l} \pmod{p} \quad \text{para } n = 0, 1, \dots$$

(A.4)

Donde y_0, \dots, y_{k-1} son valores iniciales no todos cero. Los coeficientes enteros a_0, \dots, a_{k-1} en (A.4) son seleccionados de tal forma que si son vistos como elementos de un campo finito F_p , entonces su polinomio característico es:

$$f(x) = x^k - \sum_{l=0}^{k-1} a_l x^l \in F_p[X]$$

De la recursión (A.4) esta un polinomio primitivo de F_p además se tiene que las $per(y_n)=p^k-1$.

En el método multipasos digitales, atribuido a Tausworthe, la secuencia y_0, y_1, \dots se transforma en una secuencia x_0, x_1, \dots de números pseudoaleatorios uniformemente distribuidos, de la siguiente manera: seleccione un entero m con $2 \leq m \leq k$ y sea

$$x_n = \sum_{j=1}^m y_{mn+j-1} p^{-j} \quad \text{para } n=0, 1, \dots,$$

En otras palabras, los números x_n son obtenidos por la partición de la secuencia y_0, y_1, \dots en bloques consecutivos de una longitud m y después interpretar cada bloque como un dígito de la expansión en base p de un número en $I=[0,1)$. Los números x_n son llamados números pseudoaleatorios de multipasos digitales o número pseudoaleatorios de k -pasos digitales si es que se quiere hacer énfasis en el valor de k .

A.2. Pruebas estadísticas para la validación de números pseudoaleatorios

En la medida que los números pseudoaleatorios logran pasar pruebas estadísticas, pueden tratarse como verdaderos números aleatorios. Estas son algunas pruebas que se utilizan para probar la aleatoridad de los números generados.

A.2.1 Prueba χ^2 de bondad de ajuste

Se quiere probar la hipótesis nula:

$$H_0 : F(x)=F_0(x) \quad \forall x$$

Donde $F_0(x)$ es una función de distribución completamente especificada, contra la alternativa:

$$H_1 : F(x) \neq F_0(x) \quad \text{para alguna } x$$

Se asume que las N observaciones han sido agrupadas en K categorías mutuamente excluyentes y se denotan por N_j y Np_j^0 el número esperado y el número observado de ensayos para la j -ésima categoría, $j=1, \dots, k$ respectivamente, cuando H_0 es verdadera.

El criterio de la prueba sugerida por Pearson usa la siguiente estadística:

$$Y = \sum_{j=1}^k \frac{(N_j - Np_j^0)^2}{Np_j^0}, \quad \sum_{j=1}^k N_j = N$$

Que tiende a ser pequeña cuando H_0 es verdadera y grande cuando H_0 es falsa. La distribución exacta de la variable aleatoria Y es bastante complicada, pero para muestras grandes su distribución es aproximadamente χ^2 con $k-1$ grados de libertad.

Bajo la hipótesis H_0 se espera:

$$P(Y > \chi^2_{1-a}) = a \quad (\text{A.5})$$

Donde a es el nivel de significancia, es decir 0.05 ó 0.1; el cuantil χ^2_{1-a} que corresponde a la probabilidad $1-a$ esta dado en las tablas de valores para la distribución χ^2 . Cuando se utiliza la prueba para verificar uniformidad, simplemente se divide el intervalo $[0,1]$ en k subintervalos de tamaño $\frac{1}{k}$ sin traslaparse para que $Np_j = \frac{N}{k}$, en este caso se tiene $Y = \frac{k}{N} \left(\sum_{j=1}^k (N_j - \frac{N}{k})^2 \right)$ y (A.5) se puede aplicar para probar un generador de números pseudoaleatorios.

A.2.2. Prueba de bondad de ajuste de Cramer-von Mises

El objetivo de esta prueba es dada una muestra X_1, \dots, X_N de alguna función de distribución acumulativa desconocida, se desea probar la hipótesis nula:

$$H_0 : F(x) = F_0(x),$$

donde $F_0(x)$ es una función de distribución completamente especificada, contra la alternativa

$$H_1 : F(x) \neq F_0(x)$$

para al menos un valor de x . Denotando por $X_{(1)}, \dots, X_{(N)}$ la estadística de orden y considerando la siguiente estadística de orden:

$$Y = \frac{1}{12N} + \sum_{i=1}^N \left[F_0(X_{(i)}) - \frac{2i-1}{2N} \right]^2.$$

En otras palabras, la ordenada de $F_0(x)$ se encuentra en cada valor de la muestra aleatoria $X_{(i)}$, y de esta se subtrae la cantidad $\frac{(2i-1)}{2N}$, que es el promedio justo antes y después del brinco en $X_{(i)}$; esto es, el promedio de $\frac{(i-1)}{N}$ y $\frac{i}{N}$. La diferencia se eleva al cuadrado, para que la diferencias positivas no se cancelen con las negativas y los resultados se agreguen conjuntamente.

Las cantidades de Y están tabuladas para usarlas como una función distribución asintótica de Y . La prueba de bondad de ajuste de Cramer-von Mises, con nivel de significancia α , se rechaza H_0 si y sólo si $Y > y_{1-\alpha}$, donde la cantidad $y_{1-\alpha}$ puede encontrarse en las tablas correspondientes a la prueba.

A.2.3. Prueba de series

La prueba de series es usada para verificar el grado de aleatoridad entre dos números sucesivos en una secuencia y representa una extensión de la prueba χ^2 de bondad de ajuste.

Sea $X_1 = (U_1, \dots, U_k)$, $X_2 = (U_{k+1}, \dots, U_{2k})$, \dots , $X_N = (U_{(N-1)k+1}, \dots, U_{Nk})$, una secuencia de N k -tuple. Se quiere probar la hipótesis de que los vectores aleatorios X_1, X_2, \dots, X_N están uniforme e independientemente distribuidos sobre un hipercubo unitario de k -dimensiones.

Dividiendo este hipercubo en r^k hipercubos elementales, cada uno con volumen $\frac{1}{r^k}$, y denotando por V_{j_1, \dots, j_k} el número de k -tuples que caen dentro del elemento

$$\left(\frac{j_1-1}{r}, \frac{j_1}{r}\right), i = 1, \dots, k; \quad j_i = 1, \dots, r,$$

se tiene que la estadística

$$Y = \frac{r^k}{N} \sum_{j_1, \dots, j_k = 1}^r \left(V_{j_1, \dots, j_k} - \frac{N}{r^k} \right)^2,$$

tiene una distribución asintótica χ^2 con r^k-1 grados de libertad. Puesto que hay r^k hipercubos dentro de los que puede caer X_i , la cuestión del espacio surge. Si $k = 3$ y $r = 1000$, la prueba de series requiere $1000^3=10^7$ contadores, un requerimiento problemático en términos tanto de almacenaje como de búsqueda. En tales circunstancias la prueba es raramente usada para $k > 2$.

A.2.4. Prueba de corridas por arriba y por abajo

Para esta prueba la magnitud de cada elemento es comparada con la de su predecesor inmediato en una secuencia dada. Si el elemento siguiente es mas grande, se tiene una corrida por arriba; si es más pequeño, una corrida por abajo. Así se observa que tanto la secuencia decrece o se incrementa. Una decisión concerniente a

un generador de número pseudoaleatorios puede estar basado en el número y extensión de las corridas.

Supongase que hay N términos, es decir $X_1 < X_2 < \dots < X_N$ cuando están arreglados en orden de secuencia; la secuencia de observaciones ordenadas representa una permutación de los N términos. Hay $N!$ permutaciones, cada una de ellas representa una posible muestra de observaciones. Bajo la hipótesis nula cada una de las hipótesis alternativas es igualmente probable de ocurrir. La prueba de aleatoridad, usando corridas por arriba y por abajo para la secuencia X_1, X_2, \dots, X_N de dimensión N , esta basada en la secuencia derivada de dimensión $N-1$, cuyo i -ésimo elemento es 0 ó 1 dependiendo en si $X_{i+1} - X_i$, $i = 1, \dots, N-1$, es negativa o positiva. Un gran número de largas corridas puede no ocurrir en una verdadera muestra aleatoria. La prueba rechaza la hipótesis nula si hay al menos r corridas de tamaño t o mayor, donde tanto r y t están determinados por nivel de significancia deseado.

Los números esperados de ocurrencia de corridas en una verdadera muestra aleatoria son

$$\frac{2N-1}{3} \text{ para el total de corridas}$$

$$\frac{N+1}{12} \text{ para corridas de tamaño 1}$$

$$\frac{11N-14}{12} \text{ para corridas de tamaño 2}$$

⋮

$$\frac{2[(k^2+3k+1)N-(k^3+3k^2-k-4)]}{(k+3)!} \text{ para corridas de tamaño } k,$$

$$\frac{2}{N!} \text{ para } k < N-1$$

A.2.5. Prueba de diferencias

La prueba de diferencias es concerniente a la aleatoridad de los dígitos en una secuencia de números. Sea U_1, \dots, U_N tal secuencia y se dice que la subsecuencia $U_j, U_{j+1}, \dots, U_{j+r}$ de $r+1$ números, representa una diferencia de longitud r si U_j y U_{j+r} cae entre a y β ($0 \leq a \leq \beta \leq 1$) pero U_{j+i} , $i=1, \dots, r-1$ no es. Para una verdadera secuencia de números aleatorios la probabilidad de obtener una diferencia de tamaño r es igual a

$$P(r)=(0.9)^r(0.1)$$

Una prueba χ^2 de bondad de ajuste basada en la comparación del número esperado y el número actual de diferencias de longitud r puede ser usada.

A.2.6. Prueba de correlación

El coeficiente de correlación para dos variables estocásticas X y Y se denota generalmente por ρ , la relación entre el coeficiente de correlación y la covarianza es:

$$\rho = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}$$

donde $\text{Cov}(X, Y) = E[(X - E(X))(Y - E(Y))]$ y $\sigma_X^2 = E[X - E(X)]^2$. Se utilizará el símbolo γ para la covarianza. Si el i -ésimo número pseudoaleatorio en la secuencia (r_i) esta distribuido independientemente entonces:

$$\gamma_j = E[(r_i - 0.5)(r_{i+j} - 0.5)] = 0 \text{ para } ij > 0 \quad (\text{A.6})$$

Donde $E(r_i) = 0.5$ y $E(r_{i+j}) = 0.5$ porque r_i y r_{i+j} se supone estan uniformemente distribuidos en $[0, 1)$. La covarianza rezagada γ_j se puede estimar a través de

$$\hat{\gamma}_j = \frac{1}{n-j} \sum_{i=1}^{n-j} [(r_i - 0.5)(r_{i+j} - 0.5)]$$

Nótese que para $j=0$ esta ecuación reduce el estimador de máxima verosimilitud de la varianza. Se tiene de (A.6) que $E(\hat{\gamma}_j)$ para $j > 0$. Puede ser probable que $\hat{\gamma}_j$ sea asintóticamente normal con $\text{Var}(\hat{\gamma}_j) = \frac{1}{144(n-j)}$. Así que la distribución normal puede ser utilizada para probar $H_0 : \rho_j = 0$ con $j = 1, 2, \dots, J$

A.3. Generación de distribuciones de probabilidad

Al intentar resolver problemas de tipo estocástico, se impone la necesidad de genera valores de variables aleatorias a partir de funciones de probabilidad conocidas. Al considerar los procesos estocásticos que involucran a variables continuas o discretas, se define la función $F(x)$, llamada función de distribución acumulativa de x . Si la variable aleatoria es discreta, entonces x tendrá valores enteros no negativos y $F(x)$ será una función escalonada.

Si $F(x)$ es continua en el dominio de x implica $F(x)$ sea diferenciable y $\frac{dF(x)}{dx} = f(x)$, la función $f(x)$ recibe el nombre de función de densidad.

La función de distribución acumulativa se define como:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(t) dt,$$

donde $F(x) \in [0,1]$ y $f(t)$ representa el valor de la función de densidad de la variable aleatoria X , cuando $X=t$.

Los métodos de generación de variables distribuidas uniformemente en el intervalo $(0,1)$, $U_1 \sim U_{(0,1)}$, a partir de las técnicas de números pseudoaleatorios se utilizan para modelar funciones de probabilidad más complejas. A continuación se dan a conocer cuatro métodos para simular funciones de probabilidad.

A.3.1. Método de transformación inversa

Para genera los valores x_i de las variables aleatorias a partir de cierta estadística de población cuya función de densidad esta dada por $f(x)$, es necesario contar con la función de distribución acumulativa $F(x)$. Puesto que $0 \leq F(x) \leq 1$, se generan números aleatorios U distribuidos uniformemente en $(0,1)$, i.e. $U \rightarrow U_{[0,1]}$, y hacer $F(x) = U$, x queda determinado suprayectivamente por esta relación.

Por consiguiente $\forall U_0 \in [0,1] \exists x_0$ tal que por la función inversa conocida entonces $x_0 = F^{-1}(U_0)$, donde $F^{-1}(U_0)$ es la transformación inversa o mapeo de r sobre el intervalo unitario del dominio de x .

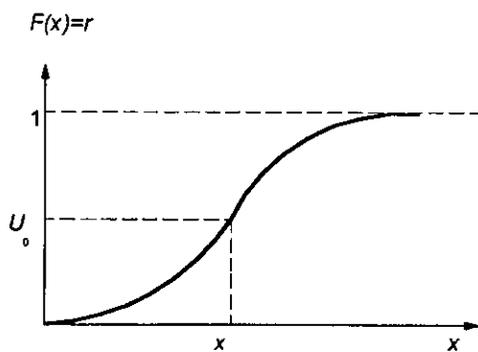


Fig A.1 función de distribución acumulativa

Este método se resume de la siguiente forma:

$U = F(x) = \int_{-\infty}^x f(t)dt$ entonces $P[X \leq x] = F(x) = P[U \leq F(X)] = P[F^{-1}(U) \leq x]$ y $F^{-1}(U)$ es una variable que tiene a $f(x)$ como función de densidad.

En muchas ocasiones no es tan fácil hallar la transformación inversa $F^{-1}(x)$ y por ello sólo son viables otros métodos o técnicas numéricas.

A.3.2. Método de rechazo

Supongase que F tiene una función de densidad f , i.e. $F(u) = \int_{-\infty}^u f(t)dt$ para $u \in \mathbb{R}$. Se busca una función de distribución G con densidad g tal que $f(t) \leq c \cdot g(t) \forall t \in \mathbb{R}$ con una constante $c \geq 1$. Ahora se generan dos muestras independientes, $x \rightarrow U_{[0,1]}$ y z de la distribución G . Si $c \cdot x > \frac{f(z)}{g(z)}$, entonces se rechaza z y se intenta con otra pareja x y z . Esto se repite hasta que finalmente $c \cdot x \leq \frac{f(z)}{g(z)}$ y entonces la variable aleatoria es $y=z$.

Nótese que

$$P(z \leq u \text{ y } z \text{ es aceptado}) = \int_{-\infty}^u \frac{f(t)}{c \cdot g(t)} dG(t) = \int_{-\infty}^u \frac{f(t)}{c \cdot g(t)} g(t) dt = \frac{1}{c} \int_{-\infty}^u f(t) dt = \frac{1}{c} F(u)$$

y tomando $u \rightarrow \infty \Rightarrow P(z \text{ sea aceptada}) = \frac{1}{c}$. Usando la probabilidad condicional se tiene

$$P(z \leq u | z \text{ es aceptada}) = \frac{\frac{1}{c} F(u)}{\frac{1}{c}} = F(u)$$

En consecuencia, la variable aleatoria y tiene la función de distribución F . La constante c deber tan cercana a 1 como sea posible para que la probabilidad de ser aceptada sea alta.

A.3.3. Método de composición o mezclas

Se refiere a F como una mezcla tal que $F = \sum_{j=1}^r p_j F_j$ con funciones de distribución sobre

\mathbb{R} y una distribución discreta de probabilidad (p_1, \dots, p_r) con $\sum_{j=1}^r p_j = 1$ y $p_j \geq 0$ para $1 \leq j \leq r$.

Primero se toma una muestra j de la distribución discreta y entonces y de la distribución F_j correspondiente. La variable aleatoria y tiene la función de distribución F . A menudo, este método es aplicado de la siguiente manera: Se hace una partición en \mathbb{R} , de un número finito de intervalos I_1, \dots, I_r y sea F_j la función de densidad con apoyo de I_j inducido por F .

A.3.4. Método de proporción uniforme

Supongase que F tiene una función de densidad f y sea $A = \{(u, v) \in \mathbb{R}^2 : 0 < u \leq \sqrt{f(u/v)}\}$ y se asume que A está contenido en algún intervalo compacto $[0, a] \times [b_1, b_2]$. Ahora se generan dos números aleatorios x_1, x_2 y se hace $z_1 = ax_1$ y $z_2 = b_1 + (b_2 - b_1)x_2$, si $(z_1, z_2) \notin A$ entonces se reintenta con otra pareja z_1, z_2 hasta que finalmente $(z_1, z_2) \in A$ y la variable aleatoria será $y = \frac{z_2}{z_1}$.

A.4. Descripción y algoritmos para la generación de distribuciones de probabilidad (distribuciones continuas)

Uniforme	$U(a, b)$
Aplicaciones posibles	Usada como un primer modelo para una cantidad que se intuye esta variando aleatoriamente entre a y b sin embargo se conoce que tan pequeña es. La distribución $U_{(0,1)}$ es esencial en la generación de variables aleatorias de otras distribuciones más complejas.
Densidad (fig. A.2)	$f(x) = \begin{cases} \frac{1}{b-a} & \text{si } a \leq x \leq b \\ 0 & \text{e.o.c.} \end{cases}$
Distribución (fig. A.3)	$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leq x \leq b \\ 1 & \text{si } b < x \end{cases}$
Parámetros	a y b números reales con $a < b$; a es un parámetro de locación, $b-a$ es un parámetro de escala.
Rango	$[a, b]$
Media	$\int_a^b \frac{1}{b-a} x dx = \frac{a+b}{2}$
Varianza	$\int_a^b \frac{[x-E(x)]^2}{b-a} dx = \frac{(b-a)^2}{12}$
Moda	No existe una única
Estimador de Máxima Verosimilitud (E.M.V.)	$\hat{a} = \min_{1 \leq i \leq n} X_i, \hat{b} = \max_{1 \leq i \leq n} X_i$
Comentarios	<ol style="list-style-type: none"> 1. La distribución $U_{(0,1)}$ es un caso especial de la distribución beta (cuando $a_1 = a_2 = 1$). 2. Si $X \rightarrow U_{(0,1)}$ y $[x, x+\Delta x]$ es un subintervalo de $[0, 1]$ con $\Delta x \geq 0$, $P\{X \in [x, x+\Delta x]\} = \int_x^{x+\Delta x} 1 dx = (x+\Delta x) - x = \Delta x$
Generación de las variables	La distribución se genera utilizando el método de la transformación inversa y se obtiene $x = a + (b-a)U_i, U_i \sim U_{(0,1)}$

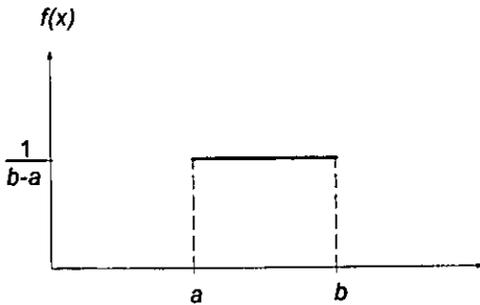


Fig. A.2 función de densidad *uniforme(a, b)*

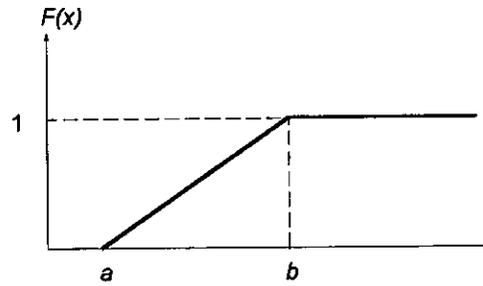


Fig. A.3 función de distribución *uniforme(a, b)*

Exponencial

$$\exp(\beta)$$

Aplicaciones posibles

Tiempo entre eventos independientes, por ejemplo llegadas a un servidor, que ocurre con tasas constantes; tiempos de vida de mecanismos con tasas constantes de riesgo.

Densidad (fig. A.4)

$$f(x) = \begin{cases} \left(\frac{1}{\beta}\right)e^{-\frac{x}{\beta}} & \text{si } x \geq 0 \\ 0 & \text{e.o.c.} \end{cases}$$

Distribución (fig. A.5)

$$F(x) = \begin{cases} 1 - e^{-\frac{x}{\beta}} & \text{si } x \geq 0 \\ 0 & \text{e.o.c.} \end{cases}$$

Parámetros

Parámetro de escala $\beta > 0$

Rango

$[0, \infty)$

Media

$$\int_0^{\infty} x \frac{1}{\beta} e^{-\frac{x}{\beta}} dx = \beta$$

Varianza

$$\int_0^{\infty} (x - \beta)^2 \frac{1}{\beta} e^{-\frac{x}{\beta}} dx = \beta^2$$

Moda

0

E.M.V.

$$\hat{\beta} = \bar{X}_{(n)}$$

Comentarios

1. La distribución $\exp(\beta)$ es un caso especial tanto de la distribución *gamma* como de la *Weibull* (para condiciones con parámetro $a=1$ y parámetro escalar β en ambos casos).

2. Si X_1, X_2, \dots, X_m son variables aleatorias independientes $\exp(\beta)$, entonces $X_1 + X_2 + \dots + X_m \rightarrow \text{gamma}(m, \beta)$, también llamada la distribución *m-Erlang*.

3. La distribución *exponencial* es la única distribución continua sin memoria.

La distribución se genera utilizando el método de la transformación inversa.

$$U = F(x) \Rightarrow U = 1 - e^{-\frac{x}{\beta}}, \quad x = -\beta \log(1 - U_i), \quad U_i \sim U_{(0,1)}$$

Si X es v.a. *m-Erlang* con media β , se puede escribir como $X = Y_1 + Y_2 + \dots$

$+ Y_m \rightarrow \exp\left(\frac{\beta}{m}\right)$, y se usa el método de la transformación inversa para generar

las Y_i s, $[Y_i = \left(\frac{\beta}{m}\right) \ln U_i]$ donde $U_1, U_2, \dots, U_m \rightarrow U_{(0,1)}$, entonces

Generación de las variables

$$X = \sum_{i=1}^m Y_i = \sum_{i=1}^m \left(\frac{\beta}{m}\right) \ln U_i = \frac{\beta}{m} \ln \left(\prod_{i=1}^m U_i \right)$$

para que sólo se calcule un logaritmo en lugar de m . El algoritmo es entonces, el siguiente:

1. Generar $U_1, U_2, \dots, U_m \rightarrow U_{(0,1)}$.

2. Sea $X = \frac{\beta}{m} \ln \left(\prod_{i=1}^m U_i \right)$

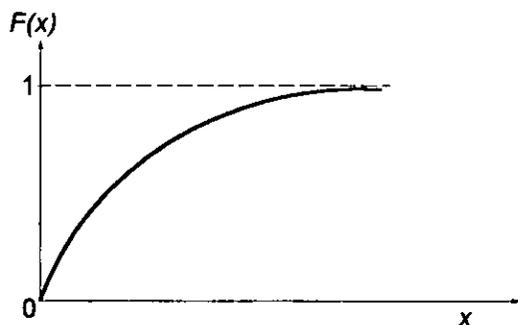


Fig. A.5 función de distribución $\exp(\beta)$

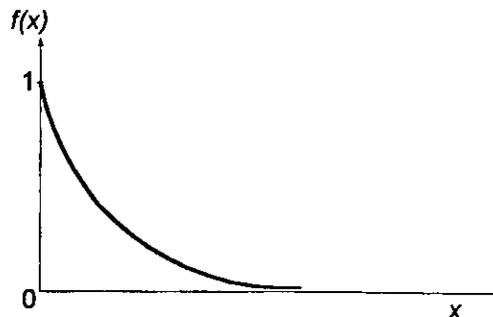


Fig. A.4 función de densidad $\exp(\beta)$

Gamma	<i>gamma</i> (a, β)
Aplicaciones posibles	Tiempo para completar algunas tareas, por ejemplo servicios a clientes o reparaciones de maquinarias.
Densidad (fig. A.7)	$f(x) = \begin{cases} \frac{\beta^a x^{a-1} e^{-\frac{x}{\beta}}}{\Gamma(a)} & \text{si } x > 0 \\ 0 & \text{e.o.c.} \end{cases}$ <p>donde $\Gamma(a)$ es la función <i>gamma</i> definida por $\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt$ para cualquier número real $z > 0$. Algunas propiedades de la función <i>gamma</i>: $\Gamma(z+1) = z\Gamma(z)$ para cualquier $z > 0$, $\Gamma(k+1) = k!$ para cualquier entero no negativo k. $\Gamma(k+\frac{1}{2}) = \sqrt{\pi} \cdot 1 \cdot 3 \cdot 5 \dots \frac{(2k-1)}{2^k}$ para cualquier entero positivo k, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$</p> <p>Si a no es un entero, no hay una forma aproximada. Si a es un entero positivo entonces:</p>
Distribución	$F(x) = \begin{cases} 1 - e^{-\frac{x}{\beta}} \sum_{j=0}^{a-1} \frac{(\frac{x}{\beta})^j}{j!} & \text{si } x > 0 \\ 0 & \text{e.o.c.} \end{cases}$
Parámetros	Parámetro de forma $a > 0, \beta > 0$.
Rango	$[0, \infty)$
Media	$a\beta$
Varianza	$a\beta^2$
Moda	$\beta(a-1)$ si $a \geq 1$, 0 si $a < 1$
E.M.V.	<p>Las dos siguientes ecuaciones deben cumplirse:</p> $\ln \hat{\beta} + \Psi(\hat{a}) = \frac{\sum_{i=1}^n \ln X_i}{n}, \quad \hat{a}\hat{\beta} = \bar{X}_{(n)}$ <p>que puede ser resuelta numéricamente. $\Psi(\hat{a}) = \frac{\Gamma'(\hat{a})}{\Gamma(\hat{a})}$ es llamada la función <i>bigamma</i>]. Aproximaciones a \hat{a} y $\hat{\beta}$ pueden obtenerse definiendo $T = \left[\ln \bar{X}_{(n)} - \sum_{i=1}^n \ln \frac{X_i}{\bar{a}} \right]^{-1}$, para obtener a \hat{a} como una función de T, y sea $\hat{\beta} = \frac{\bar{X}_{(n)}}{\hat{a}}$</p>
Comentarios	<ol style="list-style-type: none"> 1. La distribución $\exp(\beta)$ es una <i>gamma</i>(1, β). 2. Para un entero positivo m, la distribución <i>gamma</i>(m, β) se denomina la distribución <i>m-Erlang</i>(β). 3. La distribución χ^2 con k grados de libertad es la misma que una distribución <i>gamma</i>($\frac{k}{2}, 2$). 4. Si X_1, X_2, \dots, X_m son variables aleatorias con $X_i \rightarrow \text{gamma}(a_i, \beta)$, entonces $X_1 + X_2 + \dots + X_m \rightarrow \text{gamma}(a_1 + a_2 + \dots + a_m, \beta)$.

5. Si X_1 y X_2 son variables aleatorias independientes con $X_i \rightarrow \text{gamma}(a_i, \beta)$ entonces $\frac{X_1}{X_1+X_2} \rightarrow \text{beta}(a_1, a_2)$.

$$6. \lim_{x \rightarrow 0} f(x) = \begin{cases} \infty & \text{si } a < 1 \\ \frac{1}{\beta} & \text{si } a = 1 \\ 0 & \text{si } a > 1 \end{cases}$$

La distribución $\text{gamma}(1, 1)$ es la distribución *exponencial* con parámetro 1, así que sólo es necesario considerar los casos $0 < a < 1$ y $a > 1$.

Caso 1. $0 < a < 1$: Utilizando el método de rechazo, entonces

$$f(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{x^{a-1}}{\Gamma(a)} & \text{si } 0 < x \leq 1 \\ \frac{e^{-x}}{\Gamma(a)} & \text{si } 1 < x \end{cases}$$

$$c = \int_0^{\infty} f(x) dx = \frac{b}{a\Gamma(a)} \text{ donde } b = \frac{(e+a)}{e} > 1, \text{ que produce la densidad } s(x) = \frac{f(x)}{c}$$

$$s(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \frac{ax^{a-1}}{b} & \text{si } 0 < x \leq 1 \\ \frac{ae^{-x}}{b} & \text{si } 1 < x \end{cases}$$

Generar una variable Y con densidad $s(x)$ puede hacerse con el método de la transformación inversa; la función de distribución correspondiente a s es

$$S(x) = \int_0^x s(y) dy = \begin{cases} \frac{x^a}{b} & \text{si } 0 \leq x \leq 1 \\ 1 - \frac{ae^{-x}}{b} & \text{si } 1 < x \end{cases}$$

que puede ser invertido para obtener

$$S^{-1}(u) = \begin{cases} (bu)^{\frac{1}{a}} & \text{si } u \leq \frac{1}{b} \\ -\ln \frac{b(1-u)}{a} & \text{e.o.c} \end{cases}$$

Para generar Y con densidad s , primero se genera $U_1 \sim U_{(0, 1)}$, Si $U_1 \leq \frac{1}{b}$, entonces $Y = (bU_1)^{\frac{1}{a}}$, (en este caso, $Y \leq 1$). En otro caso, si $U_1 > \frac{1}{b}$, entonces $Y = -\ln \left[b \left(\frac{1-U_1}{a} \right) \right]$, que será mayor que 1. Nótese que

$$\frac{f(Y)}{f(X)} = \begin{cases} e^{-Y} & \text{si } 0 < x \leq 1 \\ Y^{a-1} & \text{si } 1 < x \end{cases}$$

Así se obtiene el algoritmo final [$b = \frac{(e+a)}{e}$ que debe ser calculado de antemano]:

1. Generar $U_1 \sim U_{(0, 1)}$ y sea $P = bU_1$. Si $P > 1$, ir al paso 3. En otro caso, continuar en el siguiente paso.
2. Sea $Y = P^{\frac{1}{a}}$ y generar $U_2 \sim U_{(0, 1)}$. Si $U_2 \leq e^{-Y}$, sea $X = Y$, en otro caso, volver al paso 1.
3. Sea $Y = -\ln \left[\frac{(b-P)}{a} \right]$ y generar $U_2 \sim U_{(0, 1)}$. Si $U_2 \leq Y^{a-1}$, sea $X = Y$. En otro caso volver al paso 1.

Caso 2. $a > 1$: Primero $\lambda = (2a-1)^{\frac{1}{2}}$, $\mu = a^{\lambda}$ y $c = \frac{4a^{\lambda} e^{-\mu}}{\Gamma(a)}$ Entonces se define $f(x) = c \cdot r(x)$, donde

$$s(x) = \begin{cases} \frac{\lambda \mu x^{\lambda-1}}{(\mu+x^{\lambda})^2} & \text{si } x > 0 \\ 0 & \text{e.o.c} \end{cases}$$

La función de distribución correspondiente a la densidad $s(x)$ es

Generación de las variables

$$S(x) = \begin{cases} \frac{x^\lambda}{\mu+x^\lambda} & \text{si } x \geq 0 \\ 0 & \text{e.o.c} \end{cases}$$

que fácilmente se invierte para obtener $S(x) = \left(\frac{\mu u}{1-u}\right)^{\frac{1}{\lambda}}$ para $0 < u < 1$. Usando el método de la transformación inversa para generar Y con densidad s . Después se agrega una ventajosa preprueba para eficiencia computacional con las siguientes constantes: $a=(2a-1)^{-\frac{1}{2}}$, $b=a \ln(4)$, $q=a+\frac{1}{a}$, $\theta=4.5$ y $d=1+\ln \theta$.

1. Generar $U_1, U_2 \sim U_{(0,1)}$.
2. Sea $V = a \ln \left[\frac{U_1}{(1-U_1)} \right]$, $Y = ae^V$, $Z = U_1^2 U_2$, $W = b + qV - Y$.
3. Si $W + d - \theta Z \geq 0$, sea $X=Y$. En otro caso proceder con el paso 4.
4. Si $W \geq \ln Z$, sea $X=Y$. En otro caso, volver al paso 1.

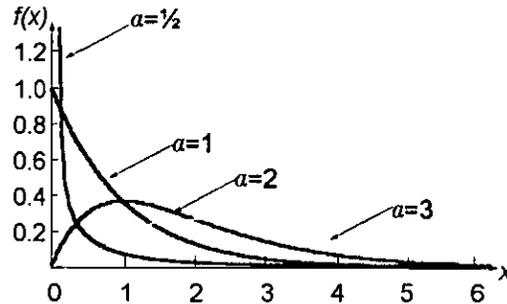


Fig A.6 funciones de densidad $gamma(a, 1)$

Weibull

Weibull(a,β)

Aplicaciones posibles

Usada extensamente en modelos de confiabilidad para mecanismos de vida, algunas veces para completar tareas (la densidad toma similares formas en densidades $gamma$).

Densidad (fig A.7)

$$f(x) = \begin{cases} a\beta^{-a}x^{a-1}e^{-\left(\frac{x}{\beta}\right)^a} & \text{si } x > 0 \\ 0 & \text{e.o.c} \end{cases}$$

Distribución

$$F(x) = \begin{cases} 1 - e^{-\left(\frac{x}{\beta}\right)^a} & \text{si } x > 0 \\ 0 & \text{e.o.c} \end{cases}$$

Parámetros

Parámetro de forma $a > 0$, parámetro de escala $\beta > 0$.

Rango

$[0, \infty)$

Media

$$\frac{\beta}{a} \Gamma\left(\frac{1}{a}\right)$$

Varianza

$$\frac{\beta^2}{a} \{2\Gamma\left(\frac{2}{a}\right) - \frac{1}{a}[\Gamma\left(\frac{1}{a}\right)]^2\}$$

Moda

$$\begin{cases} \beta \left(\frac{a-1}{a}\right)^{\frac{1}{a}} & \text{si } a \geq 1 \\ 0 & \text{si } a < 1 \end{cases}$$

Las dos siguientes ecuaciones deben cumplirse:

E.M.V.

$$\frac{\sum_{i=1}^n x_i^a \ln x_i}{\sum_{i=1}^n x_i^a} - \frac{1}{a} = \frac{\sum_{i=1}^n \ln x_i}{n}, \quad \hat{\beta} = \left(\frac{\sum_{i=1}^n x_i^a}{n} \right)^{\frac{1}{a}}$$

La primera ecuación puede ser resuelta numéricamente por el método de Newton, la segunda ecuación da directamente el valor de $\hat{\beta}$ estimada. El paso recursivo general para las iteraciones de Newton es:

$$\hat{a}_{k+1} = \hat{a}_k + \frac{A + \frac{1}{\hat{a}_k} - \frac{C_k}{\hat{B}_k}}{\frac{1}{\hat{a}_k^2} + \frac{(B_k H_k - C_k^2)}{\hat{B}_k^2}}$$

donde

$$A = \frac{\sum_{i=1}^n \ln X_i}{n}, \quad B_k = \sum_{i=1}^n X_i^{\hat{a}_k}, \quad C_k = \sum_{i=1}^n X_i^{\hat{a}_k} \ln X_i, \quad H_k = \sum_{i=1}^n X_i^{\hat{a}_k} (\ln X_i)^2$$

como valor inicial se utiliza la siguiente estimación :

$$\hat{a}_0 = \left\{ \frac{\frac{6}{\pi^2} \left[\sum_{i=1}^n (\ln X_i)^2 - \frac{\left(\sum_{i=1}^n \ln X_i \right)^2}{n} \right]}{n-1} \right\}^{-\frac{1}{2}}$$

1. Las distribuciones $\exp(\beta)$ y la *Weibull*(1, β) se comportan igual.
2. $X \sim \text{Weibull}(a, \beta)$ si y sólo si $X^a \sim \exp(\beta^a)$.
3. El logaritmo natural de una variable aleatoria *Weibull* tiene una distribución conocida como de *Valor Extremo* o distribución *Gumbel*.
4. Cuando $a \rightarrow \infty$, la distribución *Weibull* llega a degenerarse en una distribución β . La densidad *Weibull* para grandes valores de a tiene una cresta puntiaguda en la moda.

Comentarios

$$5. \lim_{x \rightarrow 0} f(x) = \begin{cases} \infty & \text{si } a < 1 \\ \frac{1}{\beta} & \text{si } a = 1 \\ 0 & \text{si } a > 1 \end{cases}$$

La función de distribución *Weibull* es fácilmente invertible para obtener $F^{-1}(u) = \beta[-\ln(1-u)]^{\frac{1}{a}}$ que deriva en el siguiente algoritmo de transformación inversa:

Generación de las variables

1. Generar $U_1 \sim U_{(0,1)}$
2. Sea $X = \beta(-\ln U)^{\frac{1}{a}}$

Utilizando el hecho que U y $1-U$ tienen la misma distribución $U_{(0,1)}$.

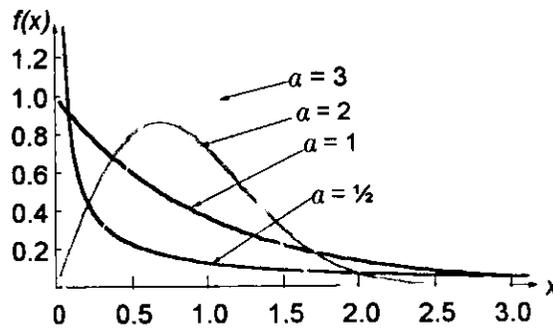


Fig. A.7 función de densidad *Weibull*($a, 1$)

Normal	$N(\mu, \sigma^2)$
Aplicaciones posibles	Representa cantidades, por ejemplo, medición de errores; que es la suma de un gran número de cantidades, en virtud del <i>Teorema del Límite Central</i> .
Densidad (fig. A.8)	$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ para todo número real x
Distribución	No tiene una forma aproximada.
Parámetros	Parámetro de locación $\mu \in (-\infty, \infty)$, parámetro de escala $\sigma > 0$
Rango	$(-\infty, \infty)$
Media	μ
Varianza	σ^2
Moda	μ
E.M.V.	$\hat{\mu} = \bar{X}_{(n)}, \quad \hat{\sigma} = \left[\frac{n-1}{n} s_{(n)}^2 \right]^{\frac{1}{2}}$
Comentarios	<ol style="list-style-type: none"> 1. Si dos variables aleatorias <i>normales</i> no están correlacionadas, entonces ellas son independientes. Para otras distribuciones esta implicación no es cierta en general. 2. Supongase que la distribución conjunta X_1, X_2, \dots, X_m es una <i>normal multivariada</i>, y sea $\mu_i = E(X_i)$ y $C_{ij} = \text{Cov}(X_i, X_j)$. Entonces para cualquier número real a, b_1, b_2, \dots, b_m, la variable aleatoria $a + b_1X_1 + b_2X_2 + \dots + b_mX_m$ tiene una distribución <i>normal</i> con media $\mu = a + \sum_{i=1}^m b_i\mu_i$ y varianza $\sigma^2 = \sum_{i=1}^m \sum_{j=1}^m b_i b_j C_{ij}$. Nótese que no es necesario asumir la independencia de las X_i's. Si las X_i's son independientes, entonces $\sigma^2 = \sum_{i=1}^m b_i^2 \text{Var}(X_i)$. 3. La distribución $N(0,1)$ es llamada a menudo la distribución <i>estándar</i> o <i>normal unitaria</i>. 4. Si X_1, X_2, \dots, X_m son variables aleatorias independientes <i>normales estándar</i>, entonces $X_1^2 + X_2^2 + \dots + X_m^2$ tiene una distribución χ^2 con k grados de libertad, que también es la distribución <i>gamma</i> $(\frac{k}{2}, 2)$. 5. Si $X \rightarrow N(\mu, \sigma^2)$, entonces e^X tiene una distribución <i>lognormal</i> con parámetros μ y σ, denotada por $LN(\mu, \sigma^2)$. 6. Si $X \rightarrow N(0,1)$, y Y tiene una distribución χ^2 con k grados de libertad y además X y Y son independientes entonces $\frac{X}{\sqrt{Y}}$ tiene una distribución <i>t</i> de <i>Student</i> con k grados de libertad. 7. Cuando $\sigma \rightarrow 0$, la distribución <i>normal</i> llega a degenerarse en μ. Dado $X \sim N(0, 1)$, se puede obtener $X' \sim N(\mu, \sigma^2)$ definiendo $X' = \mu + \sigma X$, por consiguiente se debe centrar la atención en la generación de v.a.'s <i>normales estándar</i>. Uno de los primeros métodos para generar v.a.'s $N(0, 1)$, se debe a Box y Müller, que es aún de uso extenso en relación de la disponibilidad actual de los algoritmos más rápidos. Tiene la ventaja de mantener una correspondencia inyectiva entre las v.a.'s $U_{(0,1)}$ y las v.a.'s $N(0, 1)$ producidas; el método consiste en generar las variables U_1 y $U_2 \sim U_{(0,1)}$ y entonces definir $X_1 = (-2 \ln U_2)^{\frac{1}{2}} \cos 2\pi U_1$ y $X_2 = (-2 \ln U_2)^{\frac{1}{2}} \sin 2\pi U_1$. Entonces X_1 y $X_2 \sim N(0, 1)$. Este método utiliza dos v.a.'s $U_{(0,1)}$ para producir dos v.a.'s $N(0, 1)$. Una mejora a este método se atribuye a Marsaglia y Bray y se conoce como el método polar, que elimina los cálculos trigonométricos, el método también genera pares de v.a.'s $N(0, 1)$: <ol style="list-style-type: none"> 1. Generar U_1 y $U_2 \sim U_{(0,1)}$, y sea $V_i = 2U_i - 1$ para $i=1,2$ y $W = V_1^2 + V_2^2$. 2. Si $W > 1$, volver al paso 1. En otro caso, $Y = \left[\frac{-2 \ln W}{W} \right]^{\frac{1}{2}}$, $X_1 = V_1 Y$, y $X_2 = V_2 Y$.
Generación de las variables	

Puesto que el rechazo puede ocurrir en el paso 2 con probabilidad $1 - \frac{\pi}{4}$, el método polar requerirá una v.a. $U_{(0,1)}$, para genera cada par de v.a.'s $N(0,1)$

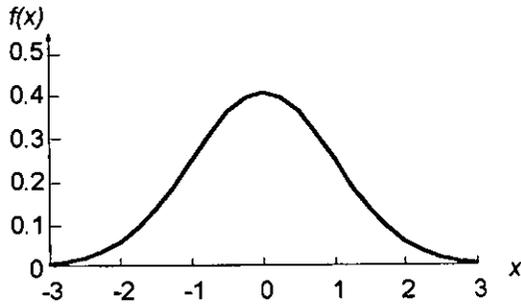


Fig. A.8 función de densidad $Normal(0,1)$

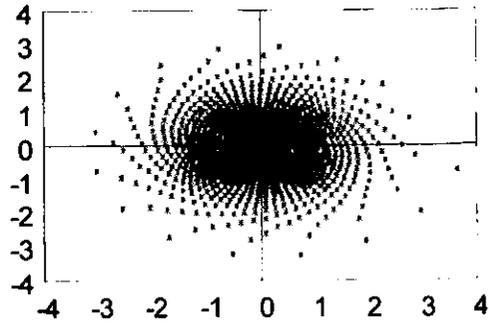


Fig A.9 $U_{\mu} = (65U + 1) \bmod 2048$, utilizando el algoritmo de Box-Müller

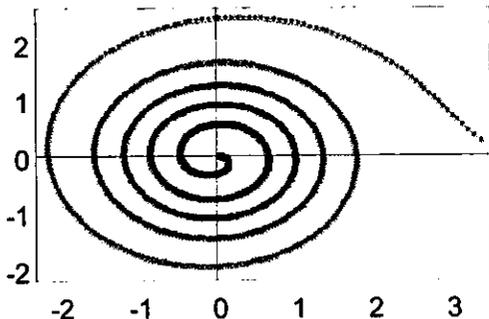


Fig A.10 $U_{\mu} = (1229U + 1) \bmod 2048$, utilizando el algoritmo de Box-Müller

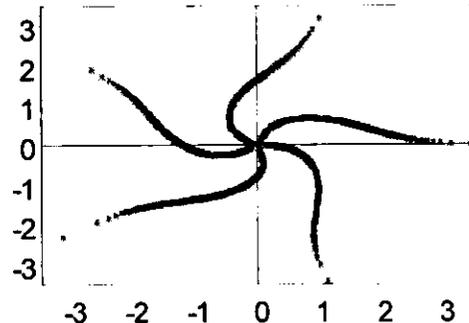


Fig A.11 Igual que A.10 intercambiando U_i por U_{μ} , con el algoritmo de Box-Müller.

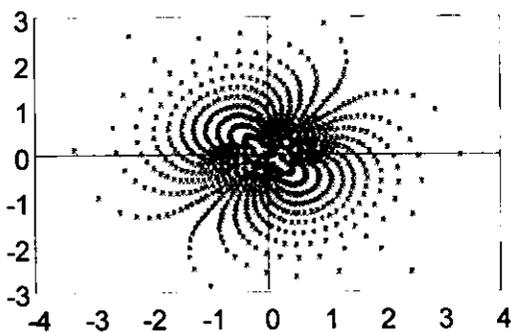


Fig. A.12 Como en A.9, utilizando el método polar.

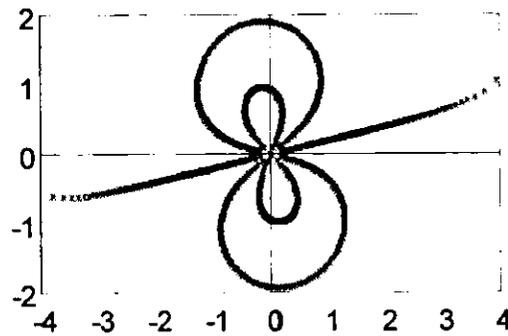


Fig. A.13 Como en A.10, utilizando el método polar.

Lognormal

$LN(\mu, \sigma^2)$

Aplicaciones posibles

Representa cantidades, por ejemplo, medición de errores, que es el producto de una gran número de otras cantidades; a veces para completar alguna tarea [La densidad toma en similares formas para densidades $gamma(u, \beta)$ con $u > 1$]

Densidad (fig. A.9)

$$f(x) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} & \text{si } x > 0 \\ 0 & \text{e.o.c} \end{cases}$$

Distribución

No hay una forma aproximada.

Parámetros

Parámetro de forma $\sigma > 0$, parámetro escalar $\mu \in (-\infty, \infty)$

Rango

$[0, \infty)$

Media

$e^{\mu + \frac{\sigma^2}{2}}$

Varianza

$e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$

Moda

$e^{\mu - \sigma^2}$

E.M.V.

$$\hat{\mu} = \frac{\sum_{i=1}^n \ln X_i}{n}, \quad \hat{\sigma} = \left[\frac{\sum_{i=1}^n (\ln X_i - \hat{\mu})^2}{n} \right]^{\frac{1}{2}}$$

Comentarios

1. $X \rightarrow LN(\mu, \sigma^2)$ si y sólo si $\ln X \rightarrow N(\mu, \sigma^2)$. Si se tienen datos X_1, X_2, \dots, X_n que se piensan pueden ser lognormales, los logaritmo naturales de los datos $\ln X_1, \ln X_2, \dots, \ln X_n$, pueden ser tratados como si se distribuyeran normalmente para probar hipótesis de distribución, estimación de parámetros y bondad de ajuste.
2. Cuando $\sigma \rightarrow 0$, la distribución *lognormal* llega a degenerarse en e^μ . Para valores pequeños de σ tiene una cresta puntiaguda en la moda.
3. $\lim_{x \rightarrow 0} f(x) = 0$, a pesar de los parámetros.

Una propiedad especial de la distribución *Lognormal*, es que si $Y \rightarrow N(\mu, \sigma^2)$ entonces $e^Y \rightarrow LN(\mu, \sigma^2)$, esto es utilizado para obtener el siguiente algoritmo:

1. Generar $Y \rightarrow N(\mu, \sigma^2)$.
2. Sea $X = e^Y$.

Generación de las variables

Nótese que μ y σ^2 no son la media y varianza de la distribución $LN(\mu, \sigma^2)$. De hecho, si $X \rightarrow LN(\mu, \sigma^2)$ y $\mu_1 = E(X)$ y $\sigma_1^2 = Var(X)$, entonces $\mu_1 = e^{\mu + \frac{\sigma^2}{2}}$ y $\sigma_1^2 = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$. Entonces, si se quiere generar una v.a. *lognormal* con media μ_1 y varianza σ_1^2 dadas, se debe resolver para μ y σ^2 en término de μ_1 y σ_1^2 , antes de generar las v.a.'s. Las fórmulas se obtienen fácilmente como $\mu = \ln \left[\frac{\mu_1^2}{(\sigma_1^2 + \mu_1^2)^{\frac{1}{2}}} \right]$ y $\sigma^2 = \ln \left[\frac{(\sigma_1^2 + \mu_1^2)^{\frac{1}{2}}}{\mu_1^2} \right]$.

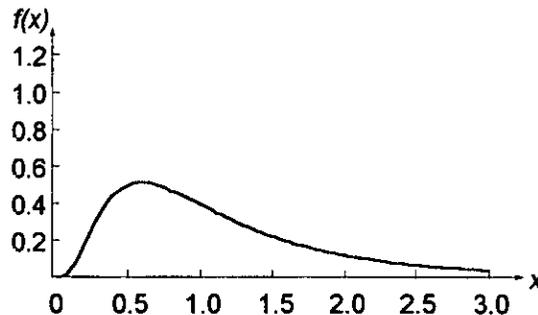


Fig. A.9 función de densidad $Lognormal(0,1)$

Beta	$beta(a_1, a_2)$
Aplicaciones posibles	Usada como un modelo aproximado en ausencia de datos; distribución de una proporción aleatoria, como la proporción de artículos defectuosos en un embarque; a veces a completar tareas, por ejemplo en una red PERT.
	$f(x) = \begin{cases} \frac{x^{a_1-1}(1-x)^{a_2-1}}{B(a_1, a_2)} & \text{si } 0 < x < 1 \\ 0 & \text{e.o.c} \end{cases}$
Densidad (fig. A.10)	<p>donde $B(a_1, a_2)$ es la función <i>beta</i> definida por:</p> $B(z_1, z_2) = \int_0^1 t^{z_1-1}(1-t)^{z_2-1} dt$ <p>para cualquier número real $z_1 > 0$ y $z_2 > 0$. Algunas propiedades de la función <i>beta</i>:</p> $B(z_1, z_2) = B(z_2, z_1), \quad B(z_1, z_2) = \frac{\Gamma(z_1)\Gamma(z_2)}{\Gamma(z_1+z_2)}$
Distribución	No hay una forma aproximada, en general. Si a_1 ó a_2 es un entero positivo, una expansión binomial puede ser utilizada para obtener $F(x)$, que será un polinomio de x y las potencias de x serán, en general, un número positivo real dentro del rango $[0, a_1+a_2-1]$.
Parámetros	Parámetros de forma $a_1 > 0$ y $a_2 > 0$.
Rango	$[0, 1]$
Media	$\frac{a_1}{a_1+a_2}$
Varianza	$\frac{a_1 a_2}{(a_1+a_2)^2(a_1+a_2+1)}$
Moda	$\left\{ \begin{array}{ll} \frac{a_1-1}{a_1+a_2-2} & \text{si } a_1 > 1, a_2 > 1 \\ 0 \text{ y } 1 & \text{si } a_1 < 1, a_2 < 1 \\ 0 & \text{si } a_1 < 1, a_2 \geq 1, \text{ si } a_1 = 1, a_2 > 1 \\ 1 & \text{si } a_1 \geq 1, a_2 < 1, \text{ si } a_1 > 1, a_2 = 1 \\ \text{No existe una única} & \text{si } a_1 = a_2 = 1 \end{array} \right.$
E.M.V.	<p>Las dos siguientes ecuaciones deben cumplirse:</p> $\Psi(\hat{a}_1) - \Psi(\hat{a}_1 + \hat{a}_2) = \ln G_1, \quad \Psi(\hat{a}_2) - \Psi(\hat{a}_1 + \hat{a}_2) = \ln G_2$ <p>donde Ψ es la función <i>bigamma</i>, $G_1 = \left[\prod_{i=1}^n X_i \right]^{\frac{1}{n}}$, $G_2 = \left[\prod_{i=1}^n (1-X_i) \right]^{\frac{1}{n}}$, nótese que $G_1+G_2 \leq 1$.</p> <ol style="list-style-type: none"> 1. La distribución $U_{(0,1)}$ y la <i>beta</i>(1,1) se comportan igual. 2. Si X_1 y X_2 son variables aleatoria independientes con $X_i \rightarrow \text{gamma}(a_i, \beta)$, entonces $\frac{X_1}{X_2+X_1} \rightarrow \text{beta}(a_2, a_1)$. 3. Una variable aleatoria <i>beta</i> X en $[0,1]$ puede ser reparametrizada para obtener una variable aleatoria <i>beta</i> en $[a,b]$ de la mismo forma con la transformación $a+(b-a)X$. 4. $X \rightarrow \text{beta}(a_1, a_2)$ si y sólo si $1-X \rightarrow \text{beta}(a_2, a_1)$. 5. La densidad <i>beta</i>(1,2) es una densidad triangular izquierda y la <i>beta</i>(2,1) es una triangular derecha. 6. $\lim_{x \rightarrow 0} f(x) \begin{cases} \infty & \text{si } a_1 < 1 \\ a_2 & \text{si } a_1 = 1 \\ 0 & \text{si } a_1 > 1 \end{cases}, \quad \lim_{x \rightarrow 1} f(x) \begin{cases} \infty & \text{si } a_2 < 1 \\ a_1 & \text{si } a_2 = 1 \\ 0 & \text{si } a_2 > 1 \end{cases}$ 7. La densidad es simétrica alrededor de $x=1/2$ si y sólo si $a_1=a_2$. La media y la moda son iguales si y sólo si $a_1=a_2$.
Comentarios	
Generación de las variables	Se puede obtener $X \rightarrow \text{beta}(a_1, a_2)$ sobre el intervalo $[a, b]$ para $a < b$ para que $X' = a+(b-a)X$, donde $X \rightarrow \text{beta}(a_1, a_2)$ en el intervalo $[0, 1]$, entonces se debe considerar sólo el último caso para referirse como la distribución <i>beta</i> (a_1, a_2).

Algunas propiedades de la distribución $beta(a_1, a_2)$ para ciertas combinaciones de (a_1, a_2) facilita la generación de v.a.'s $beta$. Si $X \rightarrow beta(a_1, a_2)$, entonces $1-X \rightarrow beta(a_1, a_2)$ para que se genere de una vez la v.a. $beta(a_1, a_2)$ si se puede obtener fácilmente una v.a. $beta(a_1, a_2)$. Tal situación ocurre cuando ya sea a_1 ó a_2 es igual a uno. Si $a_1=1$, entonces para $0 \leq x \leq 1$ se tiene que $f(x) = a_1 x^{a_1-1}$; la función de distribución es $F(x) = x^{a_1}$, entonces se puede generar fácilmente $X \rightarrow beta(a_1, 1)$, por el método de la transformación inversa, i.e., $X = U^{\frac{1}{a_1}}$ para $U \sim U_{(0,1)}$. Finalmente la distribución $beta(1, 1)$ es simplemente $U_{(0,1)}$.

Un método general para producir v.a.'s $beta(a_1, a_2)$ para cualquier $a_1 > 0$ y $a_2 > 0$ es un resultado del hecho que si $Y_1 \rightarrow gamma(a_1, 1)$, $Y_2 \rightarrow gamma(a_2, 1)$, Y_1 y Y_2 independientes, entonces $\frac{Y_1}{Y_1+Y_2} \rightarrow beta(a_1, a_2)$. Esto implica el siguiente algoritmo:

1. Generar $Y_1 \rightarrow gamma(a_1, 1)$ y $Y_2 \rightarrow gamma(a_2, 1)$ independiente de Y_1 .
2. Sea $X = \frac{Y_1}{Y_1+Y_2}$.

Generar las dos v.a.'s $gamma$ Y_1 y Y_2 puede hacerse con cualquier método para la generación $gamma$, pero se debe tener cuidado de revisar si a_1 y a_2 son menor o mayor a 1.

Este método es totalmente conveniente, puesto que esencialmente surge de tener generadores $gamma(a_1, 1)$ para todo $a > 0$; su eficiencia dependerá de la velocidad para seleccionar generadores $gamma$.

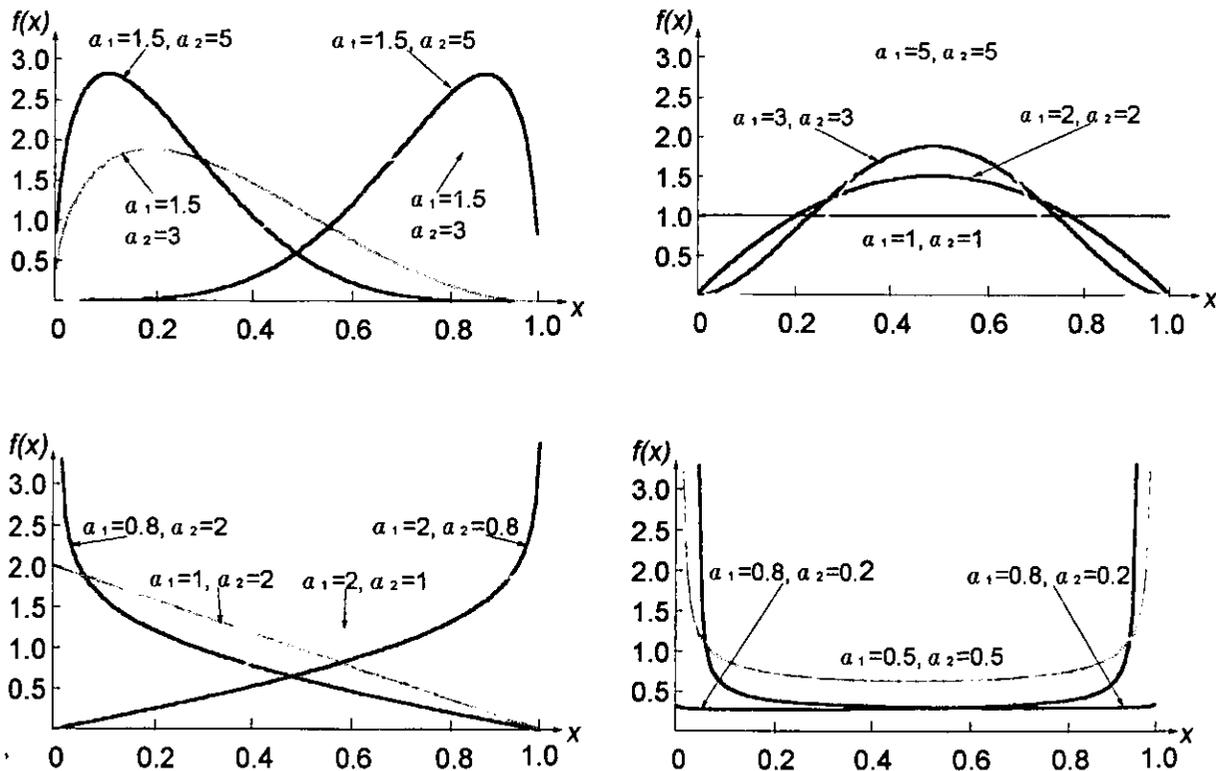


Fig A.10 función de densidad $beta(a_1, a_2)$

Triangular	$triang(a, b, c)$
Aplicaciones posibles	Usada como un modelo aproximado en ausencia de datos.
Densidad (fig. A.11)	$f(x) = \begin{cases} \frac{2(x-a)}{(b-a)(c-a)} & \text{si } a \leq x \leq c \\ \frac{2(b-x)}{(b-a)(b-c)} & \text{si } c < x \leq b \\ 0 & \text{e.o.c} \end{cases}$
Distribución	$F(x) = \begin{cases} 0 & \text{si } x < a \\ \frac{(x-a)^2}{(b-a)(c-a)} & \text{si } a \leq x \leq c \\ 1 - \frac{(b-x)^2}{(b-a)(b-c)} & \text{si } c < x \leq b \\ 1 & \text{si } b < x \end{cases}$
Parámetros	a, b y c números reales con $a < b < c$, a es un parámetro de locación, $b-a$ es un parámetro de escala, c es un parámetro de forma.
Rango	$[a, b]$
Media	$\frac{a+b+c}{3}$
Varianza	$\frac{a^2+b^2+c^2-ab-ac-bc}{18}$
Moda	c
E.M.V.	El uso de la distribución triangular, es de un modelo aproximado en ausencia de datos, por lo tanto sus E.M.V.'s no son relevantes.
Comentarios	Los casos extremos como cuando $c \rightarrow b$ y $c \rightarrow a$ son llamadas las distribuciones <i>triangulares derecha</i> e <i>izquierda</i> respectivamente. Para los casos cuando $a=0$ $b=1$, tanto la distribución <i>triangular derecha</i> e <i>izquierda</i> con casos especiales de la distribución beta. Si se tiene que $X \rightarrow triang(0, 1, \frac{c-a}{b-a})$, entonces $X' = a + (b-a)X \rightarrow triang(a, b, c)$, entonces la atención se centra en la generación de v.a.'s $triang(0, 1, c)$, donde $0 < c < 1$. La función de distribución se invierte fácilmente obteniendo, para $0 \leq u \leq 1$,
Generación de las variables	$F^{-1}(u) = \begin{cases} (cu)^{\frac{1}{2}} & \text{si } 0 \leq u \leq c \\ 1 - [(1-c)(1-u)]^{\frac{1}{2}} & \text{si } c < u \leq 1 \end{cases}$ <p>Por lo tanto, se puede ya definir el siguiente algoritmo con el método de la transformación inversa para generar $X \rightarrow triang(0, 1, c)$:</p> <ol style="list-style-type: none"> 1. Generar $U \sim U_{(0, 1)}$. 2. Si $U \leq c$, sea $X = (cU)^{\frac{1}{2}}$. En otro caso, sea $X = 1 - [(1-c)(1-U)]^{\frac{1}{2}}$.

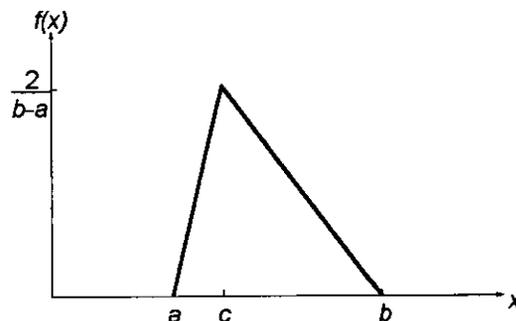


Fig A.11 función de densidad $triang(a, b, c)$

A.5. Descripción y algoritmos para la generación de distribuciones de probabilidad (distribuciones discretas)

<i>Bernoulli</i>	<i>Bernoulli(p)</i>
Aplicaciones posibles	Indicador de las salidas en un evento con dos posibles salidas; usada para generar otras variables aleatorias discretas, por ejemplo <i>binomial</i> , <i>geométrica</i> y <i>binomial negativa</i> .
Masa (fig. A.12)	$p(x) = \begin{cases} 1-p & \text{si } x=0 \\ p & \text{si } x=1 \\ 0 & \text{e.o.c} \end{cases}$
Distribución	$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ 1-p & \text{si } 0 \leq x < 1 \\ 1 & \text{si } 1 \leq x \end{cases}$
Parámetros	$p \in (0, 1)$
Rango	$\{0, 1\}$
Media	p
Varianza	$p \cdot (1-p)$
Moda	$\begin{cases} 0 & \text{si } p < \frac{1}{2} \\ 0 \text{ y } 1 & \text{si } p = \frac{1}{2} \\ 1 & \text{si } p > \frac{1}{2} \end{cases}$
E.M.V.	$\hat{p} = \bar{X}_{(n)}$
Comentarios	<ol style="list-style-type: none"> Una variable aleatoria <i>Bernoulli</i> X puede ser pensada como la salida de un experimento en el que sucede un "éxito" o "fracaso". Si la probabilidad de éxito es p, y si se obtiene éxito $X=1$ y $X=0$ si se obtiene un fracaso. Tal experimento es llamado a menudo ensayos <i>Bernoulli</i>, que provee una forma conveniente de relacionar varias distribuciones discretas con la distribución <i>Bernoulli</i>. Si t es un entero positivo y X_1, X_2, \dots, X_t son variables aleatorias independientes <i>Bernoulli(p)</i>, $X_1 + X_2 + \dots + X_t$ tiene una distribución <i>binomial</i> con parámetros t y p. Una variable aleatoria <i>binomial</i> puede ser pensada como un número de éxitos en un número fijo de ensayos independientes <i>Bernoulli</i>. Supongase que se inicia una serie de ensayos <i>Bernoulli</i> con probabilidad p de éxito en cada ensayo. Entonces el número de fracasos antes de tener el primer éxito tiene una distribución <i>geométrica</i> con parámetro p. Para un entero positivo s, el número de fracasos antes de observar el s-ésimo éxito tiene una distribución <i>binomial negativa</i> con parámetros s y p. La distribución <i>Bernoulli(p)</i> es un caso especial de la distribución <i>binomial</i> (con $t = p = 1$).
Generación de las variables	<p>El siguiente algoritmo es totalmente intuitivo y es equivalente al método de la transformación inversa (los roles de U y $1-U$ estan invertidos):</p> <ol style="list-style-type: none"> Generar $U \sim U_{(0,1)}$. Si $U \leq p$, sea $X=1$. En otro caso, $X=0$.

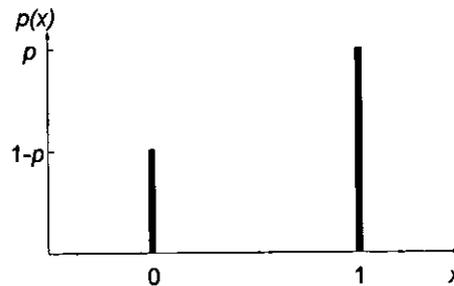


Fig. A.12 función de masa *Bernoulli*(p), con $p > 0.5$

Uniforme Discreta	$UD(i, j)$
Aplicaciones posibles	Indicador de las resultado de un evento con varias posibles resultados; cada uno equivalentemente igual de ocurrir; usado como primer modelo para cuantificar como esta variando una cantidad entre los enteros i, j pero se sabe que tanto varía.
Masa (fig. A.13)	$p(x) = \begin{cases} \frac{1}{j-i+1} & \text{si } x \in \{i, i+1, \dots, j\} \\ 0 & \text{e.o.c} \end{cases}$
Distribución	$F(x) = \begin{cases} 0 & \text{si } x < i \\ \frac{[x]-i+1}{j-i+1} & \text{si } i \leq x \leq j \\ 1 & \text{si } i < x \end{cases}$
Parámetros	donde $[x]$ denota el mayor entero $\leq x$.
Rango	i y j enteros con $i \leq j$; i es un parámetro de locación.
Media	$\{i, i+1, \dots, j\}$
Varianza	$\frac{j-i+1}{12}$
Moda	No existe una única
E.M.V.	$\hat{i} = \min_{1 \leq k \leq n} X_k, \hat{j} = \max_{1 \leq k \leq n} X_k$
Comentarios	La distribución $DU_{(0,1)}$ y la <i>Bernoulli</i> ($1/2$) se comportan igual. El algoritmo siguiente es exactamente el método de la transformación
Generación de las variables	inversa: 1. $U \sim U_{(0,1)}$ 2. Sea $X = i + [(j-i+1)U]$

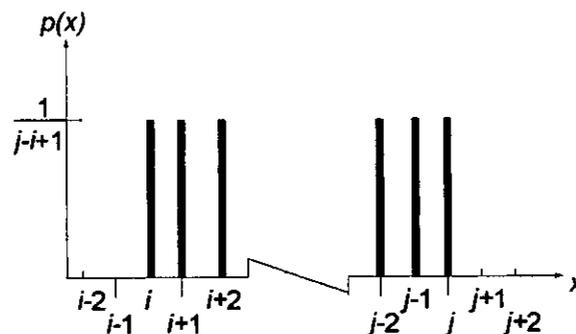


Fig A.13 función de masa $DU(i, j)$

Binomial	$bin(t, p)$ Número de éxitos en t ensayos independientes <i>Bernoulli</i> con probabilidad p de éxito en cada ensayo; número de artículos defectuosos en una serie de tamaño t ; número de artículos en una serie (por ejemplo, un grupo de personas) de tamaño aleatorio; número de artículos demandados en un inventario.
Aplicaciones posibles	
Masa (fig. A.14)	$p(x) = \begin{cases} \binom{t}{x} p^x (1-p)^{t-x} & \text{si } x \in \{i, i+1, \dots, j\} \\ 0 & \text{e.o.c.} \end{cases}$ donde $\binom{t}{x}$ es el <i>coeficiente binomial</i> definido por $\binom{t}{x} = \frac{t!}{x!(t-x)!}$
Distribución	$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ \sum_{i=0}^{[x]} \binom{t}{i} p^i (1-p)^{t-i} & \text{si } 0 \leq x \leq t \\ 1 & \text{si } t < x \end{cases}$
Parámetros	t es un entero positivo, $p \in (0, 1)$
Rango	$\{0, 1, \dots, t\}$
Media	tp
Varianza	$tp(1-p)$
Moda	$\begin{cases} p \cdot (t+1) - 1 \text{ y } p \cdot (t+1) & \text{si } p \cdot (t+1) \text{ es un entero} \\ [p \cdot (t+1)] & \text{e.o.c.} \end{cases}$
E.M.V.	Si t es conocida entonces $\hat{p} = \frac{\bar{X}_{(n)}}{t}$. Si tanto t como p son desconocidas, es más complicado hallar \hat{t} y \hat{p} . Se puede tomar la siguiente aproximación: Sea $M = \max_{1 \leq i \leq n} X_i$ para $k = 0, 1, \dots, M$; f_k el número de X_i 's $\geq k$. Entonces se puede demostrar que \hat{t} y \hat{p} son los valores de t y p que maximizan la función: $g(t, p) = \sum_{k=1}^M f_k \cdot \ln(t-k+1) + nt \cdot \ln(1-p) + n\bar{X}_{(n)} \ln\left(\frac{p}{1-p}\right)$ sujeta a las restricciones $t \in \{M, M+1, \dots\}$ y $0 < p < 1$. Esto es fácil de ver tomando un valor fijo de t , sea t_0 , el valor de p que maximiza $g(t_0, p)$ es $\frac{\bar{X}_{(n)}}{t_0}$ así \hat{t} y \hat{p} son los valores de t y $\frac{\bar{X}_{(n)}}{t}$ que sigue del mayor valor de $g(t, \frac{\bar{X}_{(n)}}{t})$ para $t \in \{M, M+1, \dots\}$. Puesto que esto puede envolver una infinidad de evaluaciones de g , un procedimiento operacional práctico puede evaluar $g(M, \frac{\bar{X}_{(n)}}{M})$, $g(M+1, \frac{\bar{X}_{(n)}}{M+1})$, $g(M+2, \frac{\bar{X}_{(n)}}{M+2})$, etc., hasta que aparezca ese lejano incremento en t en que g solamente decrece, \hat{t} y \hat{p} son seleccionados para ser los valores de t y $\frac{\bar{X}_{(n)}}{t}$ que sigue a el mayor valor observado de $g(t, \frac{\bar{X}_{(n)}}{t})$. Sin embargo, hablando estrictamente, este procedimiento no garantiza absolutamente que los verdaderos E.M.V.'s serán obtenidos.
Comentarios	<ol style="list-style-type: none"> 1. Si Y_1, Y_2, \dots, Y_t son variables aleatorias <i>Bernoulli</i>(p), entonces $Y_1 + Y_2 + \dots + Y_t \rightarrow bin(t, p)$. 2. Si X_1, X_2, \dots, X_m son variables aleatorias independientes y $X_i \rightarrow bin(t_i, p)$, entonces $X_1 + X_2 + \dots + X_m \rightarrow bin(t_1 + t_2 + \dots + t_m, p)$. 3. La distribución $bin(t, p)$ es simétrica si y sólo si $p = 1/2$. 4. $X \rightarrow bin(t, p)$ si y sólo si $t-X \rightarrow bin(t, 1-p)$. 5. La distribución $bin(1, p)$ y la <i>Bernoulli</i>(p) se comportan igual.
Generación de las variables	Para generar una v.a. $bin(t, p)$, se tiene que tener en cuenta que la suma de t v.a. <i>Bernoulli</i> (p) tiene una distribución $bin(t, p)$. De esta relación se deriva el siguiente algoritmo:

1. Genera Y_1, Y_2, \dots, Y_t como v.a.'s *Bernoulli*(p).
2. Sea $X=Y_1+Y_2+\dots+Y_m$.

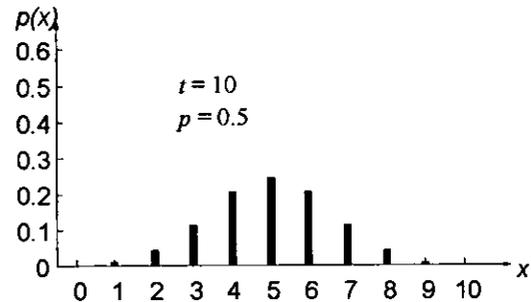
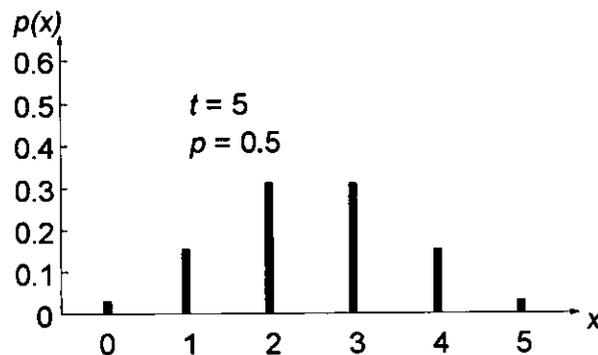
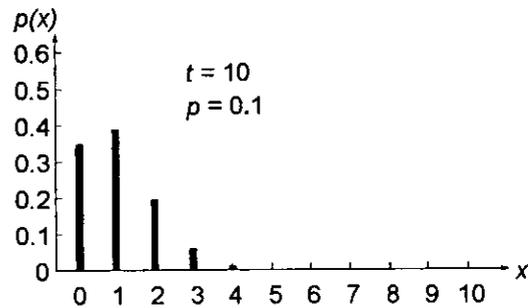
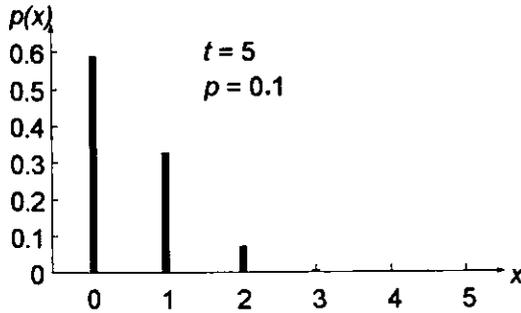


Fig. A.14 función de masa $bin(t, p)$

Geométrica

$geom(p)$

Número de fracasos antes del primer éxito en una secuencia de ensayos *Bernoulli* con probabilidad p en cada ensayo; número de artículos inspeccionados antes de encontrar el primer artículo defectuoso; número de artículos en un grupo de tamaño aleatorio; número de artículos demandados en un inventario.

Aplicaciones posibles

Masa
(fig. A.15)

$$p(x) = \begin{cases} p^x \cdot (1-p)^{1-x} & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{e.o.c.} \end{cases}$$

Distribución

$$F(x) = \begin{cases} 1 - (1-p)^{[x]+1} & \text{si } x \geq 0 \\ 0 & \text{e.o.c.} \end{cases}$$

Parámetros

$p \in (0,1)$

Rango

$\{0, 1, \dots\}$

Media

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

Varianza

$\frac{1-p}{p^2}$

Moda

E.M.V.

$\hat{p} = \frac{1}{\bar{x}_{(n)+1}}$

Comentarios

1. Si Y_1, Y_2, \dots es una secuencia de variables aleatorias independiente *Bernoulli*(p), y $X = \min \{i: Y_i=1\}-1$, entonces $X \rightarrow geom(p)$.
2. Si X_1, X_2, \dots, X_s son variables aleatorias independientes $geom(p)$, entonces $X_1 + X_2 + \dots + X_s$ tiene una distribución *binomial negativa* con parámetros s y p .

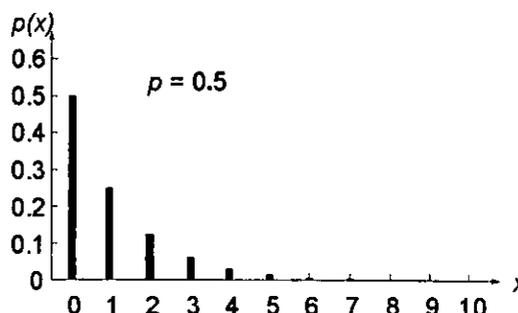
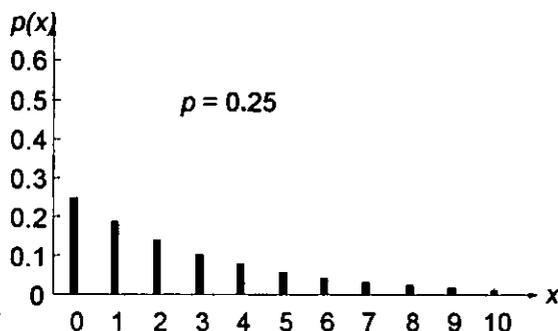
3. La distribución *geométrica* es la análoga discreta a la distribución *exponencial*, en el sentido de que la distribución *geométrica* es la única discreta sin memoria.

4. La distribución *binneg*(1, p) y la *geom*(p) se comportan igual.

El siguiente algoritmo es equivalente al método de la transformación inversa si se reemplaza U por $1-U$ en el paso 2.

Generación de las variables

1. Generar $U \sim U_{(0, 1)}$.
2. Sea $X = \left\lceil \frac{\ln U}{\ln(1-p)} \right\rceil$.



A.15 función de masa *geom*(p)

Binomial Negativa

binneg(s, p)

Número de fracasos antes del s -ésimo éxito en una secuencia de ensayos *Bernoulli* con probabilidad p en cada ensayo; número de artículos buenos inspeccionados antes de encontrar el s -ésimo artículo defectuoso; número de artículos en un grupo de tamaño aleatorio; número de artículos demandados en un inventario.

Aplicaciones posibles

Masa (fig. A.16)

$$p(x) = \begin{cases} \binom{s+x-1}{x} p^x \cdot (1-p)^x & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{e.o.c} \end{cases}$$

Distribución

$$F(x) = \begin{cases} \sum_{i=0}^{\lfloor x \rfloor} \binom{s+i-1}{i} p^s \cdot (1-p)^i & \text{si } x \geq 0 \\ 0 & \text{e.o.c.} \end{cases}$$

Parámetros

s un entero positivo, $p \in (0, 1)$

Rango

$\{0, 1, \dots\}$

Media

$$\frac{s(1-p)}{p}$$

Varianza

$$\frac{s(1-p)}{p^2}$$

Moda

$$\text{Sea } y = \frac{\lfloor s(1-p) - 1 \rfloor}{p}; \text{ entonces Moda} = \begin{cases} y \text{ y } y+1 & \text{si } y \text{ es un entero} \\ \lfloor y \rfloor + 1 & \text{e.o.c} \end{cases}$$

Si s es conocida, entonces $\hat{p} = \frac{s}{\bar{X}_{(n)} + s}$. Si tanto s como p son desconocidos, se parece a el problema de encontrar los E.M.V.'s para hallar los parámetros *binomiales*. Sea $M = \max_{1 \leq k \leq n} X_k$ para $k = 0, 1, \dots, M$; f_k el número de X_i 's $\geq k$. Se puede demostrar que \hat{s} y \hat{p} son los valores para los cuales s y p maximizan la función:

E.M.V.

$$h(s, p) = \sum_{k=1}^M f_k \cdot \ln(s+k-1) + ns \cdot \ln p + n\bar{X}_{(n)} \ln(1-p)$$

sujeta a las restricciones $s \in \{1, 2, \dots\}$ y $0 < p < 1$. Se toma un valor fijo de s , sea s_0 , el valor de p que maximiza $h(s_0, p)$ es $\frac{s_0}{\bar{X}_{(n)} + s_0}$ para que pueda ser

examinado $h(1, \cdot)$, $h(2, \frac{2}{\bar{X}_{(n)}+2})$, etc., hasta que aparezca ese lejano incremento en s en que h solamente decrece, entonces \hat{s} y \hat{p} son los valores seleccionados para ser los valores de s y $\frac{s}{\bar{X}_{(n)}+s}$ que sigue a el mayor valor observado de $h(s, \frac{s}{\bar{X}_{(n)}+s})$. Igualmente, este procedimiento operacional no garantiza absolutamente que los verdaderos E.M.V.'s serán obtenidos.

1. Si Y_1, Y_2, \dots, Y_s son variables aleatorias independiente $geom(p)$, entonces $Y_1 + Y_2 + \dots + Y_s \rightarrow binneg(s, p)$.
2. Si Y_1, Y_2, \dots , es una secuencia de variables aleatorias independientes $Bernoulli(p)$, y $X = \min \{i: \sum_{j=1}^i Y_j = s\} - s$, entonces $X \rightarrow binneg(s, p)$.
3. Si X_1, X_2, \dots, X_m son variables aleatorias independientes y $X_i \rightarrow binneg(s_i, p)$, entonces $X_1 + X_2 + \dots + X_m | binneg(s_1 + s_2 + \dots + s_m, p)$.
4. La distribución $geom(p)$ es un caso especial de la distribución *binomial negativa* (con $s = p = 1$).

Comentarios

La relación entre las funciones de distribución $Binneg(s, p)$ y $geom(p)$ da como resultado el siguiente algoritmo:

Generación de las variables

1. Generar Y_1, Y_2, \dots, Y_s como v.a.'s $geom(p)$.
2. Sea $X = Y_1 + Y_2 + \dots + Y_s$.

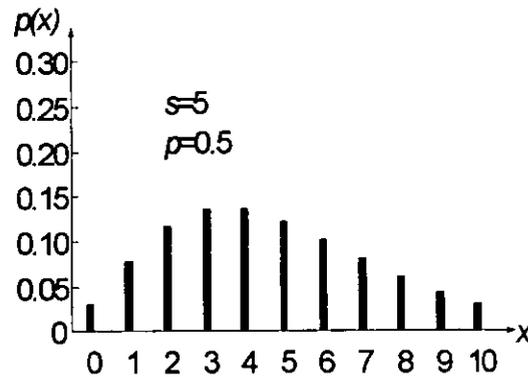
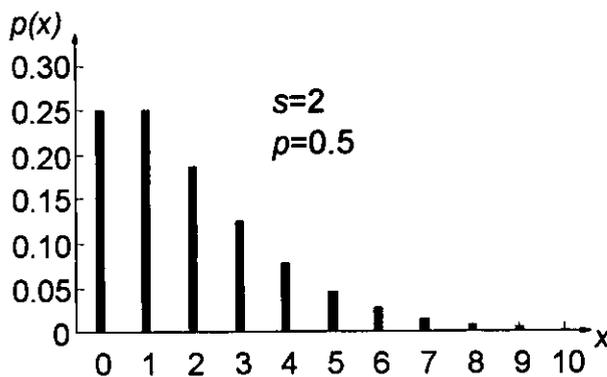


Fig A.16 función de masa *negbin*(s, p)

Poisson

$Poisson(\lambda)$

Aplicaciones posibles

Número de eventos que ocurren en un intervalo de tiempo cuando los eventos ocurren independientemente uno del otro; número de artículos en un grupo de tamaño aleatorio; número de artículos demandados en un inventario.

Masa (fig. A.17)

$$p(x) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x!} & \text{si } x \in \{0, 1, \dots\} \\ 0 & \text{e.o.c} \end{cases}$$

Distribución

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ e^{-\lambda} \sum_{i=0}^{\lfloor x \rfloor} \frac{\lambda^i}{i!} & \text{si } 0 \leq x \end{cases}$$

Parámetros

$\lambda > 0$

Rango

$\{0, 1, \dots\}$

Media

λ

Varianza

λ

Moda $\begin{cases} \lambda - 1 \text{ y } \lambda & \text{si } \lambda \text{ es un entero} \\ [\lambda] & \text{e.o.c} \end{cases}$

E.M.V. $\hat{\lambda} = \bar{X}_{(n)}$

1. Si Y_1, Y_2, \dots , es una secuencia de variables aleatorias independientes, idénticamente distribuidas $X = \max\{i: \sum_{j=1}^i Y_j \leq 1\}$. Entonces la distribución de

Comentarios las Y_i 's es $\exp(-\frac{1}{\lambda})$ si y sólo si $X \rightarrow \text{Poisson}(\lambda)$. También, si $X' = \max\{i: \sum_{j=1}^i Y_j \leq \lambda\}$, entonces la distribución de las Y_i 's es $\exp(-1)$ si y sólo si $X' \rightarrow \text{Poisson}(\lambda)$.

2. Si X_1, X_2, \dots, X_m son variables aleatorias independientes y $X_i \rightarrow \text{Poisson}(\lambda_i)$, entonces $X_1 + X_2 + \dots + X_m \rightarrow \text{Poisson}(\lambda_1 + \lambda_2 + \dots + \lambda_m)$.

El algoritmo para generar las v.a.'s es basado esencialmente, en la relación entre las distribuciones $\text{Poisson}(\lambda)$ y $\exp(-\frac{1}{\lambda})$:

1. Sea $a = e^{-\lambda}$, $b = 1$, $i = 0$.
2. Generar $U_{i+1} \sim U_{(0,1)}$ y reemplazar b por $bU_{i+1} \sim U_{(0,1)}$. Si $b < a$, y sea $X = i$. En otro caso, ir al paso 3.
3. Reemplazar i por $i+1$ y regresar al paso dos.

El algoritmo justifica que si $X = i$ si y sólo si

Generación de las variables $\sum_{j=1}^i Y_j \leq 1 < \sum_{j=1}^{i+1} Y_j$ donde $Y_j = (-\frac{1}{\lambda}) \ln U_j \sim \exp(-\frac{1}{\lambda})$ y las Y_j 's son independientes.

Esto es $X = \max\{i: \sum_{j=1}^i Y_j \leq 1\}$ para que $X \rightarrow \text{Poisson}(\lambda)$.

Desafortunadamente, este algoritmo tiende a ser lento a medida que λ , puesto que una λ grande implica $a = e^{-\lambda}$ pequeña, requiriendo más ejecuciones en el paso 2 para lograr que el producto acumulativo U_{i+1} 's quede bajo de a . [De hecho, puesto que X es 1 menor que el número de las U_{i+1} 's requeridas, el número esperado de ejecuciones en el paso 2 es $E(X) = \lambda + 1$.]

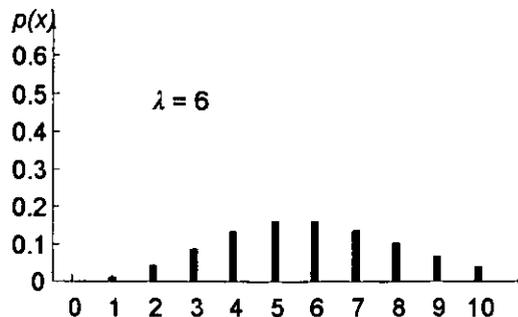
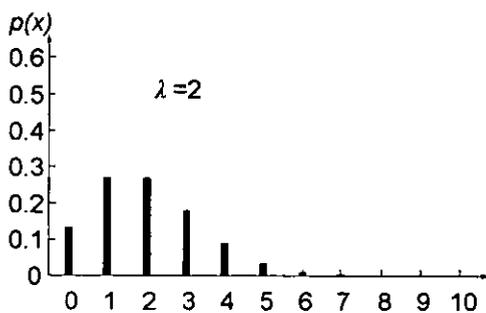
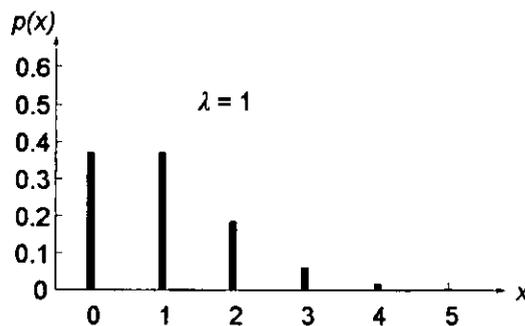
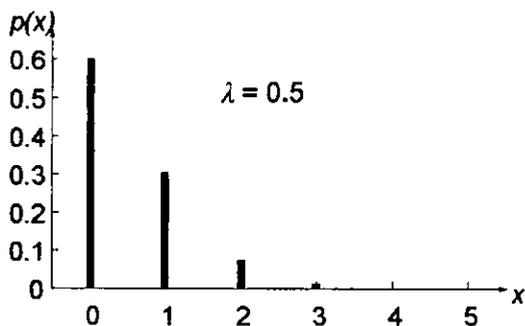
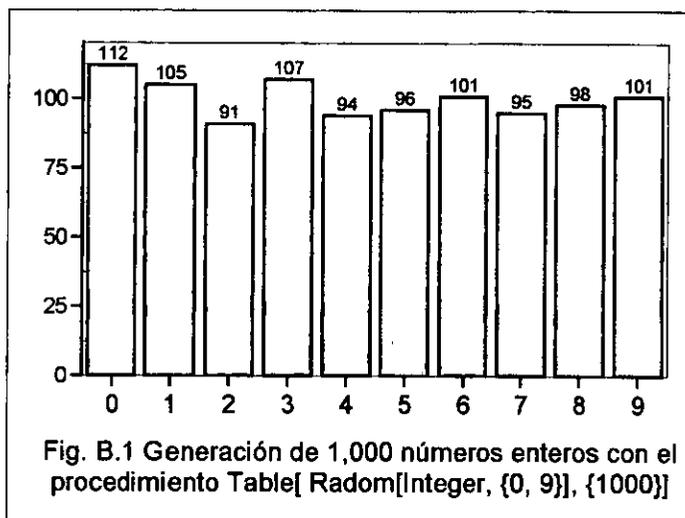


Fig. A.17 función de masa $\text{Poisson}(\lambda)$

Apéndice B

B.1. El generador de números pseudoaleatorios de *Mathematica*

La función interna de *Mathematica* `Random []`, ha sido seleccionada como de una nueva clase de generadores de números pseudoaleatorios y supera la mayoría de las pruebas estadísticas de aleatoridad y posee grandes periodos de generación. El procedimiento del generador `Random []` es llamado de sustracción con préstamo. El resultado es que este generador puede producir cualquiera de 210 ciclos separados de números pseudoaleatorios (dependiendo de la semilla utilizada), cada ciclo tiene un periodo cercano a 10^{445} . Cuando se inicializa *Mathematica*, almacena en la memoria la hora del día y lo usa como semilla para el generador. Una prueba de la uniformidad de la función `Random []` es la generación de 1,000 números enteros.



La figura B.1 muestra una distribución de números enteros sin variaciones extremas. Las barras de frecuencias indican que la sucesión de números pseudoaleatorios no presenta una proporción exacta de 100 apariciones por cada número entero en el rango [0, 1], de ocurrir así, el generador no sería el adecuado por tener una naturaleza más determinística. Este procedimiento es equivalente a definir una función suprayectiva $\mathbb{R}_{(0, 1)} \mapsto \mathbb{Z}_{[0, 9]}$.

El siguiente paso fue la realización de la prueba χ^2 de bondad de ajuste para probar la hipótesis nula sobre la distribución uniforme del generador Random[], para una sucesión de 1,000,000 números pseudoaleatorios:

$$H_0 : F(x) = F_U(x) \quad \forall x$$

Donde $F_U(x)$ es la función de distribución uniforme, contra la alternativa:

$$H_1 : F(x) \neq F_U(x) \text{ para alguna } x$$

Con $N=1,000,000$ de repeticiones iteradas de Random[], agrupadas en $k=5$ categorías mutuamente excluyentes.

Bajo la hipótesis H_0 se espera:

$$P(Y > \chi^2_{1-\alpha}) = \alpha$$

Donde α es el nivel de significancia, es decir 0.05 el cuantil $\chi^2_{1-\alpha}$ que corresponde a la probabilidad $1-\alpha$ esta dado en las tablas de valores para la distribución χ^2 . Para

verificar uniformidad, simplemente se divide el intervalo $[0, 1]$ en k subintervalos de tamaño $\frac{1}{k}$ sin traslaparse en este caso se tiene $Y = \frac{k}{N} \left(\sum_{j=1}^k (O_j - \frac{N}{k})^2 \right)$ O_j son la frecuencias observadas para el k -ésimos subintervalo.

Subintervalo	O_j	$\frac{N}{k}$
(0.0, 0.2]	199,921	200,000
(0.2, 0.4]	199,940	200,000
(0.4, 0.6]	199,706	200,000
(0.6, 0.8]	200,042	200,000
(0.8, 1.0]	200,391	200,000

Para este caso específico, el valor de la estadística Y es 1.25461 y el valor en tablas de $\chi^2_{0.95} = 9.49$, entonces la hipótesis H_0 no se rechaza con un nivel de significancia de 0.05.

Bibliografía

1. **Camposortega C., Sergio.** *Proyecciones de la Población Mexicana 1970-2040.* Tesis Maestría en Demografía, El Colegio de México, 1980, 570 pp.
2. **Conteo de Población y Vivienda, 1995.** Resultados Definitivos. Tabulados Básicos. Instituto Nacional de Estadística, Geografía e Informática. INEGI. México. 1996, 569 pp.
3. **García Guzmán, Brígida.** *Población, Trabajo y Desarrollo.* En: Demos, Carta Demográfica sobre México, Instituto de Investigaciones Sociales UNAM, 1999.
4. **Gaylord, Richard J., S.M. Kamin, P.R., Welling.** *An Introduction to Programming with Mathematica.* Springer-Verlang New York, Santa Clara, California, 1996, 452 pp.
5. **Hillier, Frederick J., G.J. Lieberman.** *Introduction to Operation Research.* Holden-Day, San Francisco, California, 1974, 639 pp.
6. **Instituto Nacional de Estadística, Geografía e Informática. Secretaría del Trabajo y Previsión Social.** Encuesta Nacional de Empleo, 1997. INEGI: STyPS. México. 1998, 376 pp.
7. **Instituto Nacional de Estadística, Geografía e Informática. Secretaría del Trabajo y Previsión Social.** Encuesta Nacional de Educación, Capacitación y Empleo, 1997. INEGI: STyPS. México. 1998, 318 pp.

8. **Instituto Nacional de Estadística, Geografía e Informática.** *Estadísticas Históricas de México*, Tomo I, INEGI, México., 1994, 596 pp.
9. **Jusidman, Clara; M. Eternod.** *La Participación de la Población en la Actividad Económica en México*. Tomo VI. Instituto Nacional de Estadística Geografía e Informática, México D.F., 1995, 214 pp.
10. **Kleijnen, Jack P.C., W. van Groenendaal.** *Simulation: A Statistical Perspective*. John Wiley & Sons, New York, New York, 1992, 241 pp.
11. **Law, Averill M.** *Simulation Modeling and Analysis*. McGraw-Hill, New York, New York, 1971, 502 pp.
12. **Mina Valdés, Alejandro.** *Curso Básico de Demografía*, Comunicaciones Internas Facultad de Ciencias UNAM, México D.F., 1996, 155 pp.
13. **Naylor, Thomas H.** *Computer Simulation Experiments with Models of Economics Systems*. John Wiley & Sons, New York, New York, 1971, 502 pp.
14. **Naylor, Thomas H, J.L. Balintfy, D.S. Burdick and K. Chu.** *Computer Simulation Techniques*. John Wiley & Sons, New York, New York, 1966, 390 pp.
15. **Niederreiter, Harald.** *Random Number Generation and Cuasi-Monte Carlo Methods*. Society for Industrial and Applied Mathematics. Philadelphia, Pennsylvania, 1992, 241 pp.
16. **Orcutt, Guy H.** *Microanalysis of Socioeconomics Systems: A Simulation Study*. Harper & Brothers, New York, New York, 1961, 425 pp.
17. **Quilodrán, Julieta.** *Tablas de Nupcialidad para México*. En: *Demografía y Economía*, Vol. 14, No. 1, El Colegio de México, 1980.
18. **Rifkin, Jeremy.** *El Fin del Trabajo*. Paidós, México D.F., 1996, 360 pp.
19. **Ripley, Brian D.** *Stochastic Simulation*. John Wiley & Sons, New York, New York, 1987, 237 pp.
20. **Rodriguez Hernández, Constanza.** *Proyecciones de Población Económicamente Activa Nivel Nacional 1980-2010*. Consejo Nacional de Población, México D.F., 1988, 107 pp.
21. **Rubinstein, Reuven Y.** *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York, New York, 1981, 278 pp.
22. **Shannon, Robert E.** *System Simulation: The Art and Science*. Prentice-Hall, Englenwood Cliffs, New Jersey, 1975, 387 pp.
23. **Varian, Hal R.** *Computacional Economics and Finance: Modeling and Analysis with Mathematica*. Springer-Verlang New York, Santa Clara, California, 1996, 468 pp.
24. **Wickham-Jones, Tom.** *Computer Graphics With Mathematica*. Springer-Verlang New York, Santa Clara, California, 1994, 721 pp.
25. **Wolfram, Stephen.** *Mathematica: A System for Doing Mathematics by Computer*. Addison-Wesley Publishing, Redwood City, California, 1991, 961 pp.