

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA



Universidad de El Salvador
Hacia la libertad por la cultura

**"ASPECTOS TEÓRICOS EN LA GENERACIÓN DE
VARIABLES ALEATORIAS"**

TRABAJO DE GRADUACIÓN

PRESENTADO POR:

NORMA YOLIBETH LÓPEZ DE BERMÚDEZ

PARA OPTAR AL TÍTULO DE:

LICENCIADA EN MATEMÁTICA

CIUDAD UNIVERSITARIA, MAYO DE 2016

UNIVERSIDAD DE EL SALVADOR
FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA
ESCUELA DE MATEMÁTICA

**"ASPECTOS TEÓRICOS EN LA GENERACIÓN DE
VARIABLES ALEATORIAS"**

TRABAJO DE GRADUACIÓN

PRESENTADO POR:

NORMA YOLIBETH LÓPEZ DE BERMÚDEZ

PARA OPTAR AL TÍTULO DE:

LICENCIADA EN MATEMÁTICA

ASESOR DE LA INVESTIGACIÓN:

M.Sc. CARLOS ERNESTO GÁMEZ RODRÍGUEZ

CIUDAD UNIVERSITARIA, MAYO DE 2016

AUTORIDADES UNIVERSITARIAS

ING. LUIS ARGUETA ANTILLÓN

RECTOR INTERINO

DRA. ANA LETICIA ZAVALA DE AMAYA

SECRETARIA GENERAL

LIC. FRANCISCO CRUZ LETONA

FISCAL

LIC. MAURICIO HERNÁN LOVO

DECANO FACULTAD DE CIENCIAS NATURALES Y MATEMÁTICA

DR. JOSÉ NERYS FUNES TORRES

DIRECTOR DE LA ESCUELA DE MATEMÁTICA

CIUDAD UNIVERSITARIA, MAYO DE 2016

Agradecimientos

A Dios, por darme la sabiduría y las fuerzas para terminar mis estudios superiores.

A mis padres, por su ayuda y apoyo para terminar mis estudios..

A mis hermanos Mario y Joel, por su apoyo incondicional y estar siempre conmigo.

A mi esposo Josué y mi hijo Kevin, por ser mi inspiración.

A mis suegros, por siempre animarme.

A mis familiares y amigos, que siempre me apoyaron y animaron a seguir adelante.

A mi asesor, MSc. Carlos Gámez, por su disponibilidad de tiempo, dedicación y apoyo en cada aspecto de este trabajo.

Al jurado, Lic. Alfredo Aguilar y Lic. Porfirio Rodríguez, por sus revisiones y correcciones de este trabajo.

Índice

1. Resumen.	1
2. Introducción.	2
3. Metodología	9
4. Análisis teórico y resultados	10
4.1. Introducción al Método Monte Carlo	10
4.2. ¿Cómo pueden los números aleatorios resolver problemas?	11
4.2.1. Historia del Método de Monte Carlo	11
4.2.2. Histograma resultados de la simulación	13
4.2.3. Muestra de trayectorias	14
4.3. Algo de probabilidad básica	15
4.3.1. Eventos y variables aleatorias	15
4.3.2. Variables aleatorias discretas y continuas	15
4.3.3. La Función de densidad de probabilidad	17
4.3.4. Valores esperados	18
4.3.5. Probabilidades condicionales	19
4.3.6. Varianza de una suma de variables aleatorias y probabilidades de densidad conjunta.	22
4.4. Generación de números aleatorios	24
4.4.1. Requisitos para un Generador de Números Aleatorios (RNG)	24
4.4.2. Técnica de media-cuadrada y otra técnicas de medio-dígito.	26
4.4.3. Generadores lineales de una congruencia de números aleatorios	27
4.5. Algunas aplicaciones	29
4.5.1. Ejemplos	31
5. Algunas distribuciones de probabilidad y sus usos	34
5.1. Caso discreto CDF - inversión : ensayos de Bernoulli	34
5.1.1. Inversión: dos resultados de CDF	35
5.1.2. Distribuciones múltiples resultados	36
5.2. Método de Alias Walker : selección rueda de ruleta.	38
5.3. Simulación de probabilidad : la distribución binomial	41
5.3.1. Muestreo de la binomial	42
5.4. Otra simulación: la distribución de Poisson	43

5.4.1.	El muestreo de la distribución de poisson por simulación	44
5.5.	Caso continuo inversión - CDF : distribución exponencial	45
5.5.1.	Inversión CDF - el método canónico para la exponencial	46
5.5.2.	Simulación de eventos discretos	48
5.5.3.	La transformación de variables aleatorias: la distribución de Cauchy	49
5.5.4.	El teorema del límite central y de la distribución normal	51
5.5.5.	El muestreo de la distribución normal.	52
5.5.6.	Muestreo aproximado a través del teorema límite central	55
5.6.	Ejemplo rechazo de muestreo: La distribución Beta	56
5.6.1.	El muestreo de la distribución Beta	59
6.	Repaso de probabilidad	62
6.1.	Distribuciones y variables aleatorias	62
6.2.	Comentarios sobre paquete de software.	64
7.	Método Monte Carlo y simulación	64
7.1.	Números aleatorios	64
7.2.	Estimaciones de áreas y volumen por la Técnica Monte Carlo	68
7.3.	Simulación	72
8.	Conclusiones	76
9.	Blibiografía	77

1. Resumen.

El Método de Monte Carlo es un método numérico para el manejo de números aleatorios. Estos permiten realizar simulaciones muy complejas de manera eficiente, además, es aplicable a cualquier tipo de problema, ya sea estocástico o determinístico.

Generalmente en estadística los modelos aleatorios se usan para simular fenómenos que poseen algún componente aleatorio. Una secuencia de número aleatorios es aquella en la cual es imposible predecir cuál será el siguiente número de la secuencia.

En computación las secuencias de números aleatorios que se usan son en realidad pseudo-aleatorios, puesto que son generados por un algoritmo que se encarga que dicha secuencia sea lo suficientemente impredecible y que no se repita en ciclos. Estos algoritmos utilizan una semilla o número inicial como punto de partida para la generación de la secuencia.

Dos secuencias serán iguales si son generadas con la misma semilla y por tanto es recomendable usar distintas semillas en cada simulación para variar la secuencia de números aleatorios que se utiliza.

Además, esta secuencia de números aleatorios se suele construir con distribución uniforme dentro del intervalo $(0,1)$, es decir, si se escoge una cantidad suficientemente grande de números de la secuencia se obtendrá la misma densidad de ellos en cada fracción de dicho intervalo.

La mayor parte de los lenguajes de programación incluyen su propio algoritmo de generación de secuencias de números aleatorios uniformemente distribuidos en el intervalo $(0,1)$, y es ésta la que se usa como partida para la generación de distribuciones más complejas mediante la utilización de métodos Monte Carlo.

En el Método Monte Carlo, por otro lado, el objeto de la investigación es el objeto en sí mismo, un suceso aleatorio o pseudo-aleatorio se usa para estudiar el modelo

2. Introducción.

En el trabajo que a continuación se presenta, veremos la importancia del Método Monte Carlo para la generación de variables aleatorias, además puede ser usado para resolver cualquier problema que tenga una interpretación probabilística.

El Método Monte Carlo es una clase amplia de algoritmos computacionales que confía en un muestreo aleatorio repetido para obtener resultados numéricos. Este método es principalmente usado en tres clases de problemas: **Optimización, Integración Numérica, y la Generación de una Distribución de Probabilidad.**

Cuando la distribución de probabilidad de una variable es también compleja, a menudo se usa en Cadena de Markov Monte Carlo (MCMC). La idea central es diseñar un modelo juicioso de una Cadena de Markov con una prescrita distribución de probabilidad estacionaria.

El Método Monte Carlo varía, pero tiende a seguir un patrón particular:

- Define un dominio de posibles entradas.
- Genera entradas aleatorias a partir de una distribución de probabilidad sobre el dominio.
- Realiza un cálculo determinista de las entradas.
- Agrega los resultados

Por ejemplo: Consideremos un círculo inscrito en un cuadrado unitario, dado que el círculo y el cuadrado tienen una proporción de área igual a $\frac{\pi}{4}$, el valor de π puede ser aproximado usando el Método Monte Carlo:

1. Dibujamos el cuadrado e inscribimos un círculo dentro de él.
2. Dispersamos uniformemente algunos objetos de tamaño uniforme (granos de arroz o arena)
3. Contamos el número de objetos dentro del círculo y el número total de estos objetos.
4. La razón de los dos conteos es una estimación de la proporción de las dos áreas, el cuál es $\frac{\pi}{4}$. Multiplica el resultado por 4 se estima π .

En este ejemplo, el dominio de las entradas es el cuadrado que esta circunscrito en el círculo. Generamos entradas aleatorias, lo que nos lleva a considerar dos puntos importantes:

1. Si los granos no están uniformemente distribuidos, nuestra aproximación será pobre.
2. Debemos generar un gran número de entradas.

La aproximación es generalmente pobre, si solo unos pocos granos caen aleatoriamente en todo el cuadrado. En promedio la aproximación mejora, si hay más granos.

Bosquejo histórico.

En la década de 1930, Enrico Fermi experimentó primero con el método Monte Carlo, mientras estudiaba la difusión de neutrones, pero no publicó nada. En 1940, se hizo una versión moderna del método como Cadenas de Markov Monte Carlo inventada por Stanislaw Ulam, mientras trabajaba en un proyecto de armas nucleares en el Laboratorio Nacional de Los Álamos.

En 1946 los físicos de este laboratorio estaban investigando el blindaje contra la radiación, ellos no pudieron resolver el problema con los métodos matemáticos convencionales deterministas. Stanislaw Ulam tuvo la idea de usar los experimentos aleatorios. En ese mismo año se describió la idea con John von Neumann y comenzaron a planear cálculos reales.

Un colega de von Neumann y Ulam, Nicolas Metrópolis, sugirió utilizar el nombre de Monte Carlo que se refiere al Casino de Monte Carlo en Mónaco. En la década de 1950 fueron utilizados en Los Álamos los primeros trabajos, en ese tiempo se comenzaron a encontrar una amplia aplicación del Método Monte Carlo en campos diferentes.

El uso secuencial de Monte Carlo en el procesamiento avanzado de señales y de la inferencia Bayesiana es más reciente. En 1993 se publicó la primera aplicación de un algoritmo de remuestreo de Monte Carlo en la inferencia estadística bayesiana. Estas metodologías secuenciales de Monte Carlo se pueden interpretar como un muestreador de aceptación-rechazo.

De 1950 a 1996 todas las publicaciones sobre las metodologías secuenciales de Monte Carlo incluyendo la poda y los métodos de remuestreo de Monte Carlo son aplicaciones actuales que se pueden introducir a diferentes situaciones en la física computacional, la química molecular y en algoritmos naturales.

Justificación

En este trabajo mostraremos como el Método Monte Carlo puede ser usado para resolver diversos problemas, donde aprenderemos y mejoraremos las habilidades en probabilidad, estadística y programación teniendo muy en cuenta que una buena fuente de números aleatorios es esencial para el método.

El Método Monte Carlo es una técnica para la simulación de fenómenos por medio de algoritmos que emplean la generación de números aleatorios. Aún así, muchos métodos surgieron y desplazaron este método, pero en muchas aplicaciones el Método Monte Carlo es insuperable.

Debido a la falta de estudio sobre la generación de variables aleatorias en el plan de estudios en la Licenciatura en Matemática y en la Licenciatura en Estadística, ha surgido la necesidad de realizar un trabajo en el cuál se investigue la generación de estas variables aleatorias por medio del Método Monte Carlo, así los estudiantes de la Escuela de Matemática puedan conocer este tema y este pueda ser utilizado en algunas materias de las carreras antes mencionadas.

Antecedentes

En la Universidad de El Salvador no se ha encontrado algún registro de trabajos de tesis sobre la generación de variables aleatorias usando el Método Monte Carlo y dentro de la Facultad de Ciencias Naturales y Matemática se habla muy superficialmente en la asignatura de Estadística II, no se profundiza en la generación de variables aleatorias, se utiliza sobre todo en la aplicación de funciones construidas en R de variables aleatorias.

Se trabajó en la generación de variables aleatorias en un curso de matemática financiera impartido por Dr. Begoña Vitoriano (Profesora de la Universidad Complutense de Madrid), pero el enfoque fue dado más a la generación de números aleatorios con una distribución normal para variables escalares y vectoriales.

En los cursos de muestreo impartido por el Dr. Domingo Gonzalez Morales (profesor de la Universidad Complutense de Madrid) se utilizan funciones predeterminadas en MATLAB para generación de números aleatorios.

Objetivo general

Aprender los aspectos teóricos para la Generación de Variables Aleatorias.

Objetivos específicos.

- Conocer las aplicaciones del Método Monte Carlo para calcular el valor de una integral.
- Analizar algunas distribuciones de probabilidad utilizando el Método Monte Carlo, en especial el Método de Aceptación - Rechazo.

Planteamiento del problema

Un problema interesante que puede ser resuelto mediante el uso de la simulación es el famoso Problema de Cumpleaños. Supongamos que en una habitación hay “ n ” personas, y que es igualmente probable que alguien cumpla años en cualquier día del año.

A partir de la teoría de probabilidades, se puede demostrar que, contrariamente a la intuición, sólo 23 personas tienen que estar presentes para que la probabilidad sea mejor que el cincuenta por ciento de que al menos dos personas tendrán la misma fecha de nacimiento (siempre es divertido probar este experimento en una gran parte o en la clase para ver que funciona en la práctica). Muchas personas sienten curiosidad por el razonamiento teórico detrás de este resultado, por lo que lo vamos a discutir brevemente antes de resolver el problema de la simulación.

Después de que a alguien se le pide su cumpleaños, las posibilidades de que la próxima persona no tendrá la misma fecha de nacimiento son $\frac{364}{365}$. Las posibilidades de que el cumpleaños de la tercera persona no coincida con los de las primeras dos personas son $\frac{363}{365}$.

Las posibilidades de que dos eventos independientes sucesivos ocurran es el producto de la probabilidad de los eventos por separados. En general, la probabilidad de que la persona n pregunte si tendrá un cumpleaños diferente de la de cualquier persona que ya hemos preguntado es

$$\left(\frac{364}{365}\right) \left(\frac{363}{365}\right) \left(\frac{362}{365}\right) \cdots \left(\frac{365 - (n - 1)}{365}\right)$$

La probabilidad de que la n -ésima persona tenga una fecha de cumpleaños diferente a la de cualquier persona en la habitación que ya se ha pedido es:

$$1 - \frac{(364)(363)\cdots[365 - (n - 1)]}{365^n}$$

esto muestra que con 23 personas las posibilidades que se repita una persona es de 50.7%, con 55 o más las posibilidades son de 98.6% o casi teóricamente cierto que al menos dos personas de cada 55 personas tendrán el mismo cumpleaños, veamos la tabla con las siguientes cantidades

n	Teórico	Simulación
5	0.027	0.028
10	0.117	0.110
15	0.253	0.225
20	0.411	0.412
22	0.476	0.462
23	0.507	0.520
25	0.569	0.553
30	0.706	0.962
35	0.814	0.819
40	0.891	0.885
45	0.941	0.936
50	0.970	0.997
55	0.986	0.987

Sin el uso de la teoría de probabilidades, podemos escribir una rutina que utiliza un generador de números aleatorios para calcular las probabilidades aproximadas para grupos de n personas. Claramente, todo lo que necesitamos es seleccionar n enteros al azar del conjunto $\{1, 2, 3, \dots, 365\}$ y examinarlos de alguna manera para determinar si hay una coincidencia.

Al repetir este experimento un gran número de veces, podemos calcular la probabilidad de que al menos encontremos a dos personas que cumplan años en cualquier reunión de n personas.

Una forma de escribir una rutina para simular el problema de cumpleaños es utilizando el enfoque de la comprobación debemos marcando los días en el calendario para saber si hay una coincidencia. Por supuesto, hay muchas otras maneras de abordar este problema.

3. Metodología

Se describe aquí los aspectos importantes de la metodología del presente trabajo de investigación:

1. Tipo de investigación. Este proyecto de investigación tiene las características siguientes: Bibliográfico, porque se ha hecho una extensa recopilación de libros impresos y de libros obtenidos por Internet para contar con el suficiente material que cubra las necesidades del estudio. El objetivo es compilar coherentemente la información más útil y destacada del tema. Descriptivo, ya que se pretende estudiar a detalle la teoría preliminar y del tema en sí.
2. Forma de Trabajo. Revisión de la bibliografía a utilizar. Se tendrán reuniones periódicas con el asesor del trabajo para tratar los diferentes aspectos de la investigación como estudiar y discutir la teoría y tratar los diferentes aspectos del trabajo escrito. Presentar mediante exposiciones los resultados.
3. Exposiciones.

Se tendrán dos exposiciones:

Primera exposición: presentación del perfil del trabajo de investigación.

Segunda exposición: presentación final del trabajo de investigación.

4. Análisis teórico y resultados

Exploraciones en el Método Monte Carlo

4.1. Introducción al Método Monte Carlo

Definiciones

No hay consenso sobre la forma de como debe ser definido Monte Carlo. Por ejemplo, Riley lo define como un modelo probabilístico de una simulación del método estocástico. El Método Monte Carlo está reservado para la integración y para pruebas estadísticas. Siwolowsky distingue entre la simulación, un Método Monte Carlo y una Simulación Monte Carlo:

- Una simulación es una representación ficticia de la realidad.
- El Método Monte Carlo es una técnica que puede ser usada para resolver problemas de matemática y estadística.
- Simulación Monte Carlo usa un muestreo repetido para determinar las propiedades de algún fenómeno o comportamiento.

Ejemplos:

■ **Simulación.**

Seleccionamos una variable uniforme pseudo-aleatoria en el intervalo de $[0, 1]$, utilizamos esta variable para simular el lanzamiento de una moneda: Si el valor es menor o igual a 0.5 designamos el resultado a Cara, si el valor es mayor que 0.50 designamos el valor a Corona. Esta es una simulación pero no una simulación Monte Carlo.

■ **Método Monte Carlo.**

Vertimos una caja de monedas en una mesa y luego calculamos la razón de las monedas que caen Cara versus Corona, es el método Monte Carlo el que determina el comportamiento de repetir el lanzamiento de la moneda, pero esta no es una simulación.

■ **Simulación Monte Carlo**

Seleccionamos un gran número de variables uniformes pseudo-aleatorias en el intervalo $[0, 1]$ y asignamos valores menores o iguales a 0.50 como Cara y mayores que 0.50 a Corona, es una simulación Monte Carlo por el comportamiento de lanzar repetidamente una moneda.

4.2. ¿Cómo pueden los números aleatorios resolver problemas?

El Método Monte Carlo es una técnica para el análisis de fenómenos que emplean la generación de números aleatorios. Este método se llama **Monte Carlo** por Stanislaw Ulam y John von Neumann, quienes lo inventaron para resolver problemas en el laboratorio de Los Álamos a mediados de 1940. El método es extensamente usado en matemática, ciencia, industria, comercio y entretenimiento. Los números aleatorios están en el corazón de los algoritmos usados para hacer predicciones acerca de los procesos estocásticos, los cuales tienen algún componente aleatorio. Es por ello, que la aleatoriedad es un ingrediente esencial en los juegos de todo tipo. Además, los números aleatorios son usados directamente en la transmisión y seguridad de datos.

Monte Carlo y los números aleatorios

El Método Monte Carlo no siempre requiere números verdaderamente aleatorios para ser útil. Lo único que esté necesita para tener una buena simulación es la secuencia pseudo-aleatoria que sea al parecer “suficientemente aleatoria” en cierto sentido y probando que los números están uniformemente distribuidos podemos buscar otra distribución deseada cuando se considere un número suficientemente grande de elementos de la secuencia, este se considera uno de los más comunes y sencillos.

Sawilowsky enumera las características de una Simulación Monte Carlo de Alta Calidad. Los números generados tienen ciertas características:

- Los números generados producen valores que pasan las pruebas de aleatoriedad.
- Hay suficientes para asegurar resultados acertados.
- Se utiliza la técnica de muestreo apropiada.
- El algoritmo usado es válido para lo que se está modelando.
- Es simulado el fenómeno en cuestión.

Los algoritmos de muestreo para números pseudo-aleatorios son usados para transformar números pseudo-aleatorios uniformemente distribuidos de acuerdo a la distribución de probabilidad dada.

4.2.1. Historia del Método de Monte Carlo

La primera aplicación de análisis que más adelante sería llamada **Método Monte Carlo**, fue empleada por el francés Comte de Buffon (1707-1788) en 1733. Él imaginaba un experimento para la estimación del valor de $\pi=3.1415926\dots$ lanzando una aguja sobre una superficie rayada como un piso de madera. Veamos la simulación con el Método Monte Carlo.

Problema de la Aguja de Buffon.

Consideremos una aguja de longitud L , lanzada al azar sobre una superficie plana con líneas paralelas cada una con una distancia d . Sea H_n una variable aleatoria que cuenta el número de veces que la aguja toca o cruza una línea en n lanzamientos, es decir, el número de aciertos.

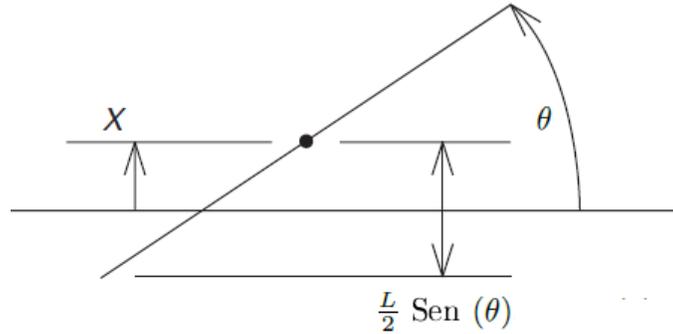


Figura 1: Aguja de Buffon entre grietas

Gráficamente vemos que la probabilidad de un éxito es igual al área bajo la curva, como lo muestra la Figura 2.

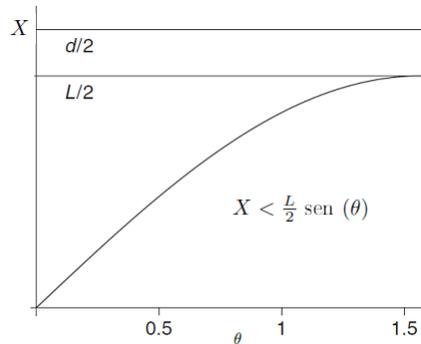


Figura 2: Área bajo la curva

Por ejemplo, tomemos $d = 2L$, la probabilidad de éxito es $\frac{1}{\pi}$, en la realización de este experimento el valor de π no se conoce, más bien emerge una estimación en el experimento, lo que es notable en la solución de Buffon, ya que nadie había soñado que π podía ser determinado por un experimento aleatorio.

Tal como lo previsto por Buffon, los científicos de Los Álamos buscaron resolver problemas como el problema de flujo de neutrones, pero ellos no podían resolverlo de manera real, y se dieron cuenta que podían producir el mismo resultado estadístico a través del uso de un modelo matemático. Este experimento real necesita la generación de verdaderos números aleatorios y un gran número de ellos. El modelo convierte los números aleatorios en una simulación.

Siguiendo con el ejemplo de la aguja de Buffon por medio de la simulación por computadora, nos daremos una elección conveniente para los parámetros $d = 2$ y $L = 1$. Para el programa en MATLAB generamos dos números aleatorios, uno es X donde $0 \leq X \leq \frac{d}{2}$, y otro es θ , donde $0 \leq \theta \leq \frac{\pi}{2}$. Calculamos $X \leq \frac{L}{2} \text{sen}(\theta)$, si se cumple se registra un éxito. Repitiendo el proceso n veces, π es

estimado como $\frac{n}{H_n}$. Por ahora, tomamos $n = 1000$ lanzamientos.

Terminal MATLAB

```
>Lanzamientos = 1000    % El porcentaje hace que el resto de la linea sea un comentario
> x = aleatorio(1,lanzamientos)    % Vector de 10 000 números pseudo aleatorios en un rango de[0, 1)
>  $\theta_1 = \text{rand}(1,\text{lanzamientos})$     % Suprime la impresión
>  $\theta = 0.5 * \pi * \theta_1$     %  $\theta_1$  ahora con 10 000 números aleatorios entre 0 y  $\frac{\pi}{2}$ 
> hist =  $x \leq 0.5 * \text{sen}(\theta_1)$     % hist es un vector con ceros donde es falso y unos donde es verdadero
> sum = (hist)    % sin punto y coma imprime el número de aciertos en los 10 000 lanzamientos
> piEst =  $\frac{\text{lanzamientos}}{\text{sum}(\text{hist})}$ 
```

4.2.2. Histograma resultados de la simulación

En la simulación de la aguja de Buffon, obtenemos un sólo número, una estimación para π . ¿Es demasiado alto, demasiado bajo, cercano al valor correcto, que tan cerca? No lo sabemos, pero hay una manera de responder a esta pregunta mediante la ejecución de múltiples experimentos y la histograma de los resultados.

Un histograma es un tipo especial de gráfica de barras para representar una serie de valores numerados $v_1, v_2, v_3, \dots, v_n$, n en total. Se crea una subdivisión del eje x , por ejemplo, tenemos puntos $x_1, x_2, x_3, \dots, x_n$ y establecemos intervalos o contenedores $(x_0, x_1], (x_1, x_2], (x_2, x_3], \dots, (x_{n-1}, x_n]$, se hace un recuento de cuántos valores se encuentran en cada contenedor, y se hace un barra de esa altura. Esto es llamado **Histograma de Frecuencias**. Dividiendo las frecuencias por n , el número total de valores, calculamos la **Frecuencia Relativa**. Haciendo las alturas iguales a esos valores genera un **Histograma de Densidad**. Un histograma de densidad es una aproximación a la distribución probabilística de los valores que se representan en el proceso.

Vamos a modificar nuestro problema de la aguja de Buffon para producir un histograma, esta vez haremos una estimación para $\frac{1}{\pi}$, queremos que la variable aleatoria esté en el numerador en lugar del denominador. Estaremos realizando 2000 simulaciones para el lanzamiento de la aguja de Buffon. El histograma resultante puede ser usado para proporcionar los estados de precisión a cerca de la estimación de π . Éste tiene forma de campana con una distribución normal. El promedio de los 2000 experimentos es una aproximación de la media de las distribución y una aproximación de $\frac{1}{\pi}$. El promedio es $\bar{p} = 0.3180$. El ancho de esta distribución normal es determinada por un parámetro llamado **Desviación Estándar**.

Una propiedad de la distribución normal es la probabilidad de que la muestra cae entre dos desviaciones estándar y en cada lado la media es 0.954. Por tanto, una muestra usualmente se encuentra entre $0.3180 - 2 * 0.0467$ y $0.3180 + 2 * 0.0467$.

Para el promedio de 2000 muestras, el intervalo disminuye $\frac{(2 * 0.0467)}{\sqrt{2000}} = 0.0021$ a cada lado de la media. Finalmente, desde un intervalo centrado en la media real contendrá una media muestral si y solo si, el mismo intervalo centrado en la media muestral contiene a la media real, la probabilidad es $0.3180 - 0.0021 \leq \frac{1}{\pi} \leq 0.3180 + 0.0021 = 0.954$.

4.2.3. Muestra de trayectorias

Otro método para mostrar los resultados de la ejecución de un experimento es el gráfico de la muestra de trayectoria. Supongamos que los resultados en el i -ésimo ensayo del experimento es R_i . El gráfico de la muestra de trayectoria de los puntos (i, R_i) para $i = 1, 2, 3, \dots, n$, donde n es el número de ensayos.

Para ilustrarlo, consideremos el experimento de un juego en el que un jugador apuesta \$1 contra la casa. Supongamos que las posibilidades de ganar son incluso, que el jugador tenga un 50 % de posibilidades de ganar cada apuesta versus un 50 % de oportunidades de perder, pero la casa tiene mucho más dinero que el jugador, es decir 20 veces más. El jugador empieza con \$100 y la casa con \$2000 y ellos juegan hasta que uno o el otro quede quebrado.

Podríamos hacer muchas preguntas acerca de este experimento tales como, ¿cuál es la probabilidad de que el jugador arruinara la casa? ¿Cuánto tiempo jugará la última hasta que uno o el otro este en la quiebra?, etc.

Si tenemos un código simulando este experimento. El jugador gana si el número aleatorio es menor que 0.5 sino el jugador pierde. El construido con herramientas de trazado se utiliza para crear una muestra de trayectoria de la fortuna del jugador en el primer ensayo (Figura 3a). Entonces 20 ensayos adicionales son ejecutados y calculado el tiempo de juego (Figura 3b).

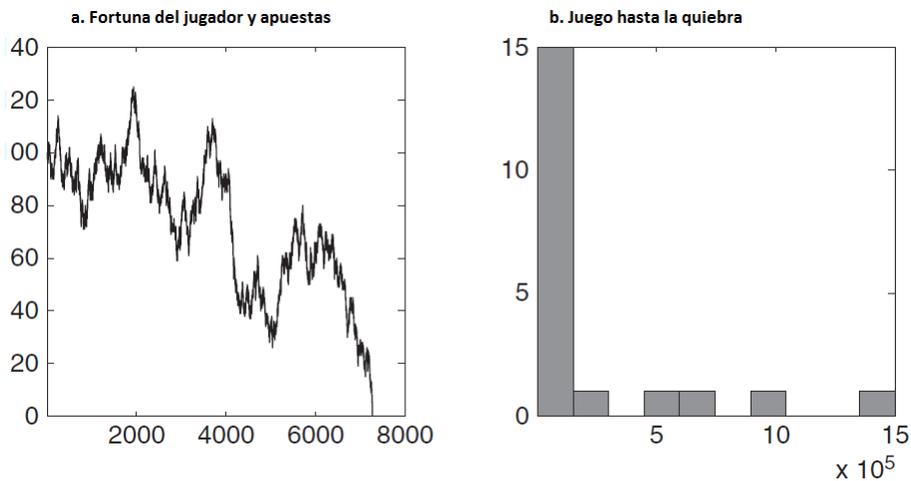


Figura 3: Resultados de la simulación

Nota: No hay límite superior para el tiempo de jugar, como vemos en el histograma, una ejecución puede tomar un tiempo muy largo, por ejemplo: 1,500,000 juegos.

4.3. Algo de probabilidad básica

4.3.1. Eventos y variables aleatorias

Por un **Evento E** en un experimento probabilístico nos referimos a algún conjunto de resultados designados. El número de resultados o cardinalidad de E, es denotado por $|E|$. El conjunto Ω de todos los posibles resultados de el experimento es el universo. La probabilidad de E, escrita por $Pr(E)$, es decir, la fracción de veces que ocurre E en un larga secuencia de ensayos del experimento. Además, $Pr(E)$ es el limite de esa fracción por un número finito de ensayos. Una variable aleatoria es el mapeo de los resultados del experimento, en el conjunto Ω .

Por ejemplo, veamos el experimento de lanzar una moneda dos veces. Los resultados de cada lanzamiento es **cara** (H) y **corona** (T). Los resultados del experimento son $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$. Algunos posibles eventos son “ambos resultados son cara”, “ocurre al menos una cara”, etc.

Una posible variable aleatoria podría ser el número de caras que ocurren en el ensayo del experimento. Sea X esta variable aleatoria que puede ser 0, 1 o 2, porque cada resultado en Ω es igualmente probable. Por lo tanto, $Pr(X = 0) = \frac{1}{4}$ porque solo ocurre (T, T) , $Pr(X = 1) = \frac{1}{2}$ porque solo ocurren dos resultados para este evento y $Pr(X = 2) = \frac{1}{4}$ porque solo ocurre (H, H) .

Supongamos que los resultados de E son divididos en dos subconjuntos tal que $E = E_1 \cup E_2$ y que E_1 y E_2 no tienen elementos en común, es decir, E_1 y E_2 son eventos disjuntos o $E_1 \cap E_2 = \emptyset$. La probabilidad de ocurrencia de algún evento E, es la suma de las probabilidades de que algún evento E_1 pueden producirse y la probabilidad que algún evento E_2 pueda producirse. Por lo tanto,

$$Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2). \quad (4.1)$$

si $E_1 \cap E_2 = \emptyset$.

En general, si tenemos dos eventos E_1 y E_2 (no necesariamente disjuntos), tenemos

$$Pr(E_1 \cup E_2) = Pr(E_1) + Pr(E_2) - Pr(E_1 \cap E_2). \quad (4.2)$$

La intersección debe ser restada porque esos eventos ya están contados en los primeros dos términos.

4.3.2. Variables aleatorias discretas y continuas

Las variables aleatorias pueden ser discretas y continuas (o mixtas):

Una **variable aleatoria discreta** solo asume un número de resultados finitos o infinitos contables. Por ejemplo, el experimento de el lanzamiento de una moneda tiene un número finito de resultados, ellos son 4. Por lo tanto, la variable aleatoria del número de caras es una variable discreta. Por ejemplo: Sea X el número de lanzamientos de una moneda hasta que aparezca la primera cara. En principio, podría formar cualquier número de lanzamientos de la moneda, para que esto suceda así, X es una variable aleatoria discreta pero no está finitamente valorada.

Una *variable aleatoria continua* X es uno de los que su $Pr(X = x) = 0$, para cada valor real x . Hay muchos fenómenos que pueden ser considerados continuos y esto es ventajoso hacerlo desde el punto de vista de la simplificación de la matemática. Eso incluye el paso del tiempo, la mayor parte de dimensiones biológicas, distancias, velocidades, etc. Incluso en la medición de una variable, ésta se limita necesariamente a las resoluciones finitas. Las mediciones son tomadas de ejemplo de variables aleatorias continuas.

Sea X una variable aleatoria, ya sea discreta o continua. La *función de distribución acumulada o CDF (Cumulative Distribution Function) de X* es definida como $CDF(x) = Pr(X \leq x)$. Por ejemplo, consideremos el experimento de el doble lanzamiento de una moneda y la variable aleatoria $x < 0$ entonces $CDF(x) = 0$, ya no hay resultados para los cuales $X < 0$. En $x = 0$ la CDF salta hasta $\frac{1}{4}$ entonces $Pr(X = 0) = \frac{1}{4}$. Luego en $x = 1$, la CDF otra vez hace otro salto a $\frac{1}{2}$. En total $CDF(1) = \frac{3}{4}$, ya que el evento $X \leq 1$ consiste en los resultados $\{(H, T), (T, H), (T, T)\}$, ver Figura 4.

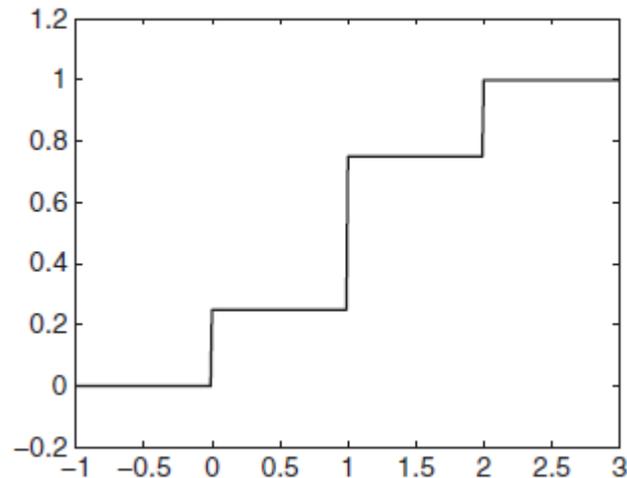


Figura 4: CDF para el número de caras en dos lanzamientos de moneda

Entre $x = 1$ y $x = 0$ la CDF está sin cambios y así permanecerá en el valor de $\frac{1}{4}$ durante este intervalo. Similarmente, entre $x = 1$ y $x = 2$ la CDF tiene un valor constante de $\frac{3}{4}$. Finalmente, para $x = 2$ la CDF salta para $\frac{1}{4}$ desde 1, $CDF(2) = 1$ y allí permanece, puesto que ahora todos los posibles resultados son incluidos.

En general, como x tiende a menos infinito, más y más resultados del experimento son omitidos, y así el CDF tiende a 0, es decir, $\lim_{x \rightarrow -\infty} CDF(x) = 0$.

Similarmente, como x tiende a más infinito, más y más resultados del experimento son incluidos, con el resultado que la CDF tiende a 1, es decir, $\lim_{x \rightarrow \infty} CDF(x) = 1$.

También mientras x aumenta la CDF permanecerá constante o aumenta más y más resultados que satisfacen la condición que $X \leq x$. Es decir, que la CDF satisface la condición que si $x_1 \leq x_2$ entonces $CDF(x_1) \leq CDF(x_2)$.

4.3.3. La Función de densidad de probabilidad

Para una variable aleatoria discreta, *la función de densidad de probabilidad o pdf (probability density function)* es definida como $\text{pdf}(x) = \text{Pr}(X = x)$. Equivalentemente

$$\text{pdf}(x) = \lim_{\varepsilon \rightarrow \infty} (\text{CDF}(x) - \text{CDF}(x - \varepsilon)).$$

Por lo tanto, la probabilidad de que un resultado sea (estrictamente) más grande que $x = a$ y menor o igual $x = b$, esta dada por

$$\text{Pr}(a < X \leq b) = \sum_{a < X \leq b} \text{pdf}(x) = \text{CDF}(b) - \text{CDF}(a).$$

Para una variable aleatoria continua la función de densidad de probabilidad es la derivada de la *CDF*, es decir,

$$\text{pdf}(x) = \frac{d}{dx} \text{CDF}(x). \quad (4.3)$$

En este caso la probabilidad de que un resultado se encuentre entre a y b esta dado por la integral

$$\text{Pr}(a < X \leq b) = \int_a^b \text{pdf}(x) dx = \text{CDF}(b) - \text{CDF}(a) \quad (4.4)$$

En particular, la *CDF* esta dada por

$$\text{CDF}(x) = \int_{-\infty}^x \text{pdf}(t) dt \quad (4.5)$$

La distribución uniforme.

La más importante distribución en el Método Monte Carlo es la distribución uniforme, porque a partir de esta muestra se derivan todos los demás. La distribución uniforme es una distribución continua y recibe el nombre porque el valor de la función de probabilidad de densidad es el mismo para cualquier punto de Ω . La *pdf* es constante en Ω .

Un ejemplo, es la variable aleatoria X uniformemente distribuida y definida en el intervalo $\Omega = [0, 6] = \{x : 0 \leq x \leq 6\}$. Para cada $x \in [0, 6]$, la densidad es $\frac{1}{6}$, es decir, $\text{pdf}(x) = \frac{1}{6}$, mientras que para $x \notin [0, 6]$, es decir, $\text{pdf}(x) = 0$. Esto es así porque $\text{Pr}(\Omega) = 1$. Aquí,

$$\text{cdf}(x) = \int_{-\infty}^x \text{pdf}(t) dt = \int_0^x \frac{1}{6} dt = \frac{x}{6} \{0 \leq x \leq 6\}.$$

Entonces el $\lim_{x \rightarrow \infty} \text{CDF}(x) = \text{CDF}(6) = 1$. Esta densidad (Figura 5a) y la distribución acumulativa (Figura 5b).

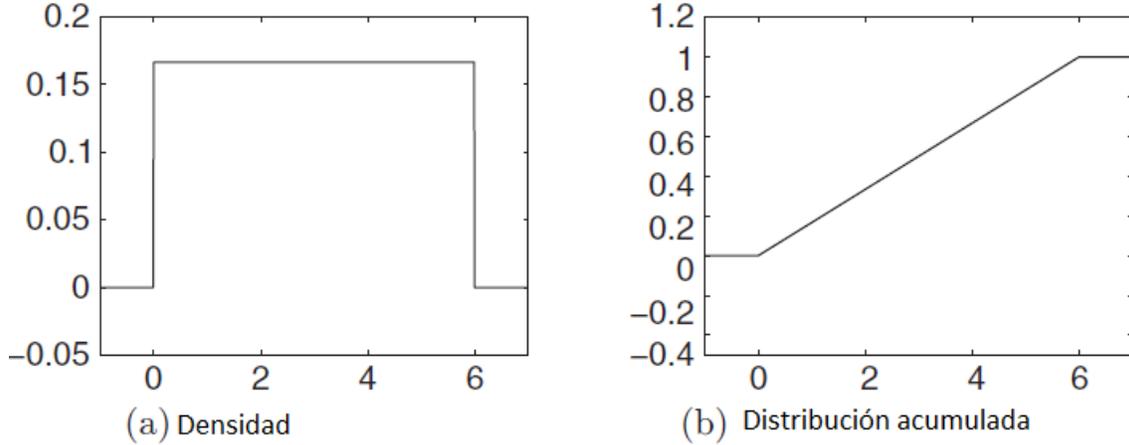


Figura 5: Distribución uniforme en $[0, 6]$

Por conveniencia, denotamos por $U(a, b)$ la distribución uniforme continua en el intervalo $[a, b]$. Esta función de densidad es

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x \leq b \\ 0 & \text{en otro caso} \end{cases} \quad (4.6)$$

y la función de distribución es

$$F(x) = \int_{-\infty}^x \frac{1}{b-a} dt = \int_a^x \frac{1}{b-a} dt = \frac{x-a}{b-a}. \quad (4.7)$$

4.3.4. Valores esperados

Sea r una función definida en una variable aleatoria X , eso es, en los resultados de un experimento. Si X es discreta, entonces el valor esperado o expectativa de r , es la suma de estos posibles valores $v = r(x)$ ponderados por la probabilidad de este valor $\sum_{v=r(x)} v * Pr(X = x)$. Alternativamente, esto puede ser escrito por

$$E(r(X)) = \sum_{x \in \Omega} r(x) \times pdf(x). \quad (4.8)$$

El valor esperado es denotado por $E(r)$ y tiene una muy simple interpretación. Si un experimento se ejecuta muchas veces, el valor esperado es el valor del promedio de los valores de r para todos los ensayos del experimento. Esto es, dibujar muchas muestras independientes para la variable aleatoria X , calcular los valores $r(X)$ y el valor que esperamos para el promedio es $E(r)$.

Este simple hecho es la base fundamental para todo el Método Monte Carlo. El valor esperado también es útil como medida de el “centro” de la distribución de los valores de r , el histograma de esos valores tiende a agruparse alrededor de $E(r)$.

Las expectativas que se producen con frecuencia son la media, por lo general denotado por μ y se define como la expectativa de la propia variable aleatoria,

$$\mu = E(X) = \begin{cases} \sum_{\Omega} x \text{ pdf}(x) & \text{si } X \text{ es discreta} \\ \int_{-\infty}^{\infty} x \text{ pdf}(x) dx & \text{si } X \text{ es continua} \end{cases}. \quad (4.9)$$

y la varianza, usualmente conocida por var y definida como la expectativa del cuadrado de la diferencia entre la variable aleatoria y la media,

$$\text{var}(X) = E((X - \mu)^2) \quad (4.10)$$

4.3.5. Probabilidades condicionales

En un experimento de probabilidad a menudo sucede que el conocimiento adicional hace que algunos resultados originales sean imposibles. Por ejemplo: se reparten 5 cartas al azar de una baraja de 52 cartas, pero suponemos que las primeras 4 se conoce que son de corazones. Esta información permite un cálculo improvisado de la probabilidad que las 5 cartas serán de la misma mano o palo (corazón), una mano conocida como *color(flush)*.

En general, sea A y B eventos en Ω y supongamos que queremos calcular la probabilidad de A dado que el resultado debe ser restringido por B. Esto se refiere a la **condición de probabilidad de A dado B** y se denota por

$$P(A|B) = \frac{Pr(A \wedge B)}{Pr(B)}. \quad (4.11)$$

Es decir, que aquí actuá B como el universo.

El ejemplo anterior de repartir las tarjetas (con respecto al orden es muy importante)

$$Pr(5 \text{ cartas de cada 5 son tratadas como corazones}) = \frac{13}{52} \times \frac{12}{51} \times \frac{11}{50} \times \frac{10}{49} \times \frac{9}{48}.$$

y

$$Pr(\text{las primeras 4 cartas son corazones y la quinta cualquier carta}) = \frac{13 \times 12 \times 11 \times 10 \times 48}{52 \times 51 \times 50 \times 49 \times 48}.$$

Por lo tanto,

$$Pr(\text{As de corazón}/4 \text{ primeras cartas son corazón}) = \frac{9}{48}.$$

En otro ejemplo, supongamos que conocemos la suma del lanzamiento de dos dados que es 8, ¿cuál es la probabilidad de que un dado sea 3? El evento B es el conjunto de 5 tiradas $\{6 + 2, 5 + 3, 4 + 4, 3 + 5, 2 + 6\}$. El evento de sacar un 3 en uno de los dos dados, puede suceder solo en 2 de esos lanzamientos, por lo tanto tenemos:

$$P(A|B) = \frac{Pr(A \wedge B)}{Pr(B)} = \frac{\frac{2}{36}}{\frac{5}{36}} = \frac{2}{5}.$$

Sino fuera por la información adicional, la probabilidad (al menos) de que al lanzar el dado, éste sea 3 sería $\frac{11}{36}$, visto directamente por el conteo entre los 36 posibles lanzamientos, no contando el 3, ya que este está dos veces.

Un uso muy importante del acondicionamiento está en el cálculo de la *descomposición* de probabilidades. Supongamos los eventos $B_1, B_2, B_3, B_4, \dots, B_n$ que forman la descomposición de el universo, eso es que ellos no tienen elementos en común y juntos incluyen todos los posibles resultados, simbólicamente,

$$\Omega = \bigcup_{i=1}^n B_i \text{ y } B_i \cap B_j = \emptyset.$$

Entonces para cualquier evento A, se sigue como

$$Pr(A) = Pr(A | B)Pr(B_1) + Pr(A | B_2)Pr(B_2) + \dots + Pr(A | B_n)Pr(B_n). \quad (4.12)$$

Eventos independientes y variables aleatorias independientes

Dos eventos son independientes si el resultado de cualquiera de ellos no tiene efecto en el otro. Por ejemplo, en el lanzamiento de dos monedas, el resultado de unos de los lanzamientos no tiene ningún efecto en el otro. Si los eventos A y B son independientes, entonces

$$Pr(A \wedge B) = Pr(A)Pr(B). \quad (4.13)$$

Esto es porque

$$Pr(A) = Pr(A | B) = \frac{Pr(A \wedge B)}{Pr(B)}.$$

Formula de Bayes.

A veces es fácil de confundir $Pr(A | B)$ con $Pr(B | A)$. Por ejemplo, supongamos que un experto testifica ante una corte donde solo hay una en dos millones de posibilidades que las pruebas de ADN a la izquierda de la escena del crimen coincidiera con el ADN de una persona escogida al azar de la población.

Ya que el ADN del acusado es emparejado, la fiscalía afirma que la probabilidad de que el acusado sea inocente es uno en dos millones. En otras palabras

$$Pr(\text{inocente} | \text{coincidencia de ADN}) = \frac{1}{2,000,000}.$$

pero las pruebas de ADN en realidad dicen que

$$Pr(\text{coincidencia de ADN} \mid \text{inocente}) = \frac{1}{2,000,000}.$$

¿Puede el orden de un condicional invertirse? La formula de Bayes responde a la pregunta. Ya que $Pr(A \wedge B)$ es simétrica en A y B, podemos escribir

$$Pr(A \mid B) Pr(B) = Pr(A \wedge B) = Pr(B \mid A) Pr(A).$$

Por lo tanto, una forma de la *formula de Bayes* es

$$Pr(B \mid A) = \frac{Pr(A \mid B) Pr(B)}{Pr(A)}. \quad (4.14)$$

Tomando un paso mas allá, el evento A puede ser descompuesto por el evento B y su complemento, que denotamos por B^c , por tanto,

$$Pr(A) = Pr(A \mid B) Pr(B) + Pr(A \mid B^c) Pr(B^c).$$

Desde, ahora la *formula de Bayes* toma la forma

$$Pr(B \mid A) = \frac{Pr(A \mid B) Pr(B)}{Pr(A \mid B) Pr(B) + Pr(A \mid B^c) Pr(B^c)}. \quad (4.15)$$

Volviendo al problema anterior de la prueba de ADN, si queremos revertir la condición, necesitamos la información adicional respecto a una estimación a priori de $Pr(B)$, la cual es la probabilidad de la inocencia de los acusados.

Veamos otro ejemplo, en un incidente al anochecer de atropello y fuga que involucra un taxi. Un testigo dice que la cabina era azul. En esta ciudad solo hay dos colores de taxi, el 15% son azules y el 85% son verdes. Bajo las condiciones de iluminación y distancia, se ha demostrado que un testigo puede hacer la distinción entre los dos colores solo el 80% de las veces. Entonces ¿cuál es la probabilidad que el taxi sea azul?

Dado que las pruebas muestran que el color puede ser identificado el 80% de las veces. Pero esto no tiene en cuenta la preponderancia de los taxis verdes. Podemos hacerlo usando la formula de Bayes. Sea el evento $A = \{\text{El testigo dice que la cabina era azul}\}$ y el suceso $B = \{\text{El taxi era en realidad azul}\}$. La alternativa B^c es $G = \{\text{la cabina es verde}\}$.

Entonces por Bayes,

$$Pr(B \mid A) = \frac{Pr(A \mid B) Pr(B)}{Pr(A \mid B) Pr(B) + Pr(A \mid B^c) Pr(B^c)} = \frac{(0.8)(0.15)}{(0.8)(0.15) + (0.2)(0.85)} = 0.4138.$$

es menor que el 50%.

4.3.6. Varianza de una suma de variables aleatorias y probabilidades de densidad conjunta.

Sea X_1 y X_2 dos variables aleatorias con media μ_1 y μ_2 entonces

$$\begin{aligned}\text{var}(X_1 + X_2) &= \mathbb{E}(X_1 + X_2 - \mu_1 - \mu_2)^2 = \mathbb{E}[(X_1 - \mu_1) + (X_2 - \mu_2)]^2, \\ &= \mathbb{E}(X_1 - \mu_1)^2 + \mathbb{E}(X_2 - \mu_2)^2 + 2\mathbb{E}((X_1 - \mu_1)(X_2 - \mu_2)).\end{aligned}$$

En general, el término $\mathbb{E}((X_1 - \mu_1)(X_2 - \mu_2))$ no tiene que ser igual a cero. En efecto, este término es llamado *covarianza* de dos variables aleatorias X_1 y X_2 , denotada por $\text{cov}(X_1, X_2)$.

Entonces

$$\text{cov}(X_1, X_2) = \mathbb{E}((X_1 - \mu_1)(X_2 - \mu_2)). \quad (4.16)$$

Si X_1 y X_2 , son independientes $\text{cov}(X_1, X_2) = 0$ y obtenemos el siguiente teorema:

Teorema 1.

Si X_1 y X_2 son variables aleatorias independientes, entonces

$$\text{var}(X_1 + X_2) = \text{var}(X_1) + \text{var}(X_2). \quad (4.17)$$

Distribución de probabilidad conjunta.

Para ver esto es necesario que primero discutamos la *distribución conjunta* de dos variables aleatorias X_1 y X_2 . Si X_1 y X_2 son discretas, su *pdf conjunta* es

$$f_{X_1, X_2}(X_1, X_2) = \text{Pr}(X_1 = x_1 \text{ y } X_2 = x_2).$$

Esta es una función definida en puntos discretos en el plano X_1, X_2 , esos puntos conforman Ω .

Ilustremos esa condición con dos ejemplos. Primero, una probabilidad de densidad conjunta se puede hacer a partir de dos densidades univariadas simplemente multiplicándose. Si X tiene una densidad de $f_X(x)$ y Y tiene una densidad de $f_Y(y)$ entonces

$$f_{XY}(x, y) = f_X(x) f_Y(y).$$

Si X e Y son independientes, es una densidad conjunta. Un ejemplo concreto de esto es la *pdf conjunta* del lanzamiento de un dado donde

$$\text{Pr}(\text{Dado}_1 = i \text{ y } \text{Dado}_2 = j) = \text{Pr}(\text{Dado}_1 = i)\text{Pr}(\text{Dado}_2 = j).$$

Para el otro ejemplo, si el experimento de probabilidad consiste en seleccionar dos cartas de una baraja de 8 cartas que se componen de dos de diamante, dos de corazones, dos de trébol, dos de espada. Sea X_1 la primera carta y X_2 la segunda carta y denotemos la densidad conjunta como $f(x_1, x_2)$.

Es fácil ver esto

$$f(x_1, x_2) = \begin{cases} \frac{2}{8} \times \frac{2}{7} = \frac{1}{14} & \text{Si } x_1 \neq x_2, \\ \frac{2}{8} \times \frac{1}{7} = \frac{1}{28} & \text{Si } x_1 = x_2. \end{cases}$$

Obteniendo en el *color(flush)* de este juego el evento $A = \{(C, C), (D, D), (H, H), (S, S)\}$ y

$$Pr(\text{flush}) = 4 \times \frac{1}{28} = \frac{1}{7}.$$

Distribución marginal

Sea $f(x_1, x_2)$ una función de densidad conjunta que define $f_1(x_1)$ y $f_2(x_2)$ tal que en el caso discreto tenemos

$$f_1(x_1) = \sum_{x_2} f(x_1, x_2), \quad f_2(x_2) = \sum_{x_1} f(x_1, x_2), \quad (4.18)$$

y en el caso continuo

$$f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_2, \quad f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2) dx_1. \quad (4.19)$$

Estas son llamadas *marginales*. Para obtener el marginal de la primera variable, sumamos o integramos el resultado con la segunda variable, hacemos lo mismo con la primera variable. Los marginales es una función de probabilidad de densidad univariante y representa la distribución de cada variable por separada, independiente del valor de la otra. Eso es si $f_1(x_1)$ representa la probabilidad (densidad) al dar valor al x_1 sobre todos los valores de x_2 .

Es fácil ver que si la densidad conjunta es el producto de las dos densidades univariantes como en el primer ejemplo $f_{X_1, X_2} = f_{X_1} f_{X_2}$, entonces f_{X_1} y f_{X_2} serán las marginales.

Continuando con el ejemplo anterior de las 8 cartas, las marginales son

$$f_1(x_1) = f(x_1, C) + f(x_1, D) + f(x_1, H) + f(x_1, S) = 3 \times \frac{1}{4} + \frac{1}{28} = \frac{1}{4}.$$

para cualquier x_1 . Esto nos dice que la probabilidad de cualquier primer carta que sea específica es $\frac{1}{4}$. Similarmente la marginal de x_2 ,

$$f_2(x_2) = f(C, x_2) + f(D, x_2) + f(H, x_2) + f(S, x_2) = 3 \times \frac{1}{4} + \frac{1}{28} = \frac{1}{4}.$$

para todo x_2 . Note que $f(H, H) = \frac{1}{28}$ y $f_1(H) f_2(H) = \frac{1}{16}$. Por tanto esta distribución conjunta no surge como el producto de las marginales.

4.4. Generación de números aleatorios

En esta sección damos un primer vistazo a los problemas y principios básicos de la generación de números aleatorios. Ya que una computadora determinista, podría parecer imposible que podría ser utilizado para generar números aleatorios. Esto es correcto: los números generados son algorítmicamente computarizadas y son bastante deterministas.

Sin embargo, aparecen ser al azar y deben pasar rigurosas pruebas diseñadas para garantizar que pueden proporcionar los mismos resultados que los números verdaderamente aleatorios haría producir para el problema dado. En consecuencia, los números generados por computadora al azar a veces se denominan "pseudo aleatorios".

4.4.1. Requisitos para un Generador de Números Aleatorios (RNG) .

Las cualidades deseadas de un generador de números aleatorios son los siguientes: Deberían

1. Ser rápidas las aplicaciones que utilizan millones de números aleatorios,
2. Ser reproducibles para hacer que la depuración sea posible,
3. Ser susceptibles al análisis para asegurar sus propiedades distributivas,
4. Tener un período muy largo, todos los generadores de números aleatorios con el tiempo se repiten,
5. Ser aparentemente aleatorio para la aplicación prevista, la calidad es de primordial importancia.

Una clase de técnica que se han tratado son híbridos análogos/digitales, por ejemplo, circuitos electrónicos generando "ruido blanco" o circuitos oyendo los clicks de un contador de Geiger una medición radioactiva en decadencia.

Sin embargo, estos métodos tienen muchos inconvenientes: son demasiado lentos, no repetibles, tienden a ser sesgada, y requieren competencias específicas de equipo. Hoy en día, los generadores de números aleatorios son prácticamente todos algorítmicos.

Veremos un ejemplo de un algoritmo intrincado con resultados predecibles o pobremente aleatorios. El algoritmo actúa en los números decimales de 10 dígitos, el cambio del actual número X en un nuevo número de la secuencia. El algoritmo está dado en los siguientes pasos:

1. Tomar N que es el dígito más significativo de X . Pasos 2-13 se repiten exactamente $n + 1$ veces.
2. Sea M el segundo dígito más significativo de X . Ir al paso 3 + M .
3. Si $X < 5 \times 10^9$, sea $X = X + 5 \times 10^9$.
4. Reemplazar X por $\lfloor X/10^5 \rfloor \bmod(10^{10})$. (La notación $\lfloor x \rfloor$ significa el entero mayor n con $n \leq x$).
5. Vuelva a colocar X por $(1001001001 \times X) \bmod(10^{10})$

6. Si $X < 10^8$ a continuación, establezca X siendo $X + 9814055677$; de lo contrario, establezca X sea $10^{10} - X$.
7. Intercambiando la orden bajamos a cinco dígitos de X , con el orden más altos de cinco dígitos de X .
8. Reemplazar X por $(1001001001 \times X) \bmod(10^{10})$.
9. Para cada dígito d de X , disminuir d por 1 si $d > 0$.
10. Si $X < 10^5$, establezca X como $X^2 + 99999$; de lo contrario, establezca X como $X - 99999$.
11. Si $X < 10^9$, establezca X como $10 \times X$ y repita este paso.
12. Sustituya X por 10 dígitos el centro de $X(X - 1)$.
13. Si $N > 0$, N disminuirlos por uno y volver al paso 2. Si $N = 0$, el algoritmo termina con el valor actual de X como el siguiente valor de la secuencia.

Desafortunadamente, la primera vez que Knuth corrió este algoritmo casi inmediatamente convergió con el valor 6065038420 (que sorprendentemente, se convierte a sí mismo un algoritmo complicado). Después de esto, cuando se encontró con un valor inicial diferente, convergió a un ciclo que tiene una longitud de 3178(!).

La lección que debemos aprender es que la complejidad no es un sustituto de aleatoriedad, y de hecho, a veces aparente la complejidad oculta un sencillo comportamiento.

Generadores usando un valor almacenado

Por último, un generador de números aleatorios será de la forma:

$$\text{ejemplo de salida} = f(\text{variables}, \text{parámetros})$$

para alguna función f . En cada petición de un número aleatorio, el algoritmo calcula la función prescrita utilizando los parámetros y los valores actuales de sus variables y la "muestra" de salida para algún proceso aleatorio hipotético.

Como una petición se hace una tras otra, el generador produce una secuencia de muestras, R_1, R_2, R_3, \dots . Los parámetros de la función son valores fijos sobre la ejecución de las muestras; los cambios de muestra a muestra son las variables.

Las variables podrían ser de tiempo o cantidades externas del algoritmo, o uno o más valores almacenados internamente que cambian de iteración a iteración. Las variables almacenadas internamente a veces se llaman las *semillas*.

En esta sección consideraremos generadores que tienen solamente una sola semilla. Si su valor fue asignado por algún proceso externo en cualquier momento durante la producción de muestras, entonces

sería imposible generar exactamente la misma secuencia de las muestras, por sí el programa debe ser depurado. (De lo contrario, el proceso externo sería parte de la función f). También sería difícil de garantizar las propiedades de distribución de la secuencia.

Por esta razón, la variable debe ser generada internamente por el propio generador de números aleatorios. Una excepción a esta regla es en la asignación del primer valor. Típicamente, un generador se "sembró" por el reloj de la computadora. La semilla inicial debe ser guardada por la aplicación con fines de depuración, si llega a ser necesario.

En cualquier caso, si en cualquier momento la semilla se repite durante la generación de la secuencia de R_1, R_2, R_3, \dots . Entonces, a partir de ese momento, la secuencia se repetirá también. De ahí que, como hemos dicho en la introducción, un generador de números aleatorios se repetirá, y es deseable tener un periodo tan largo como sea posible.

Como resultado de lo anterior, vemos que un generador de números es efectiva como un libro que contiene una larga lista de números. La semilla inicial selecciona el punto de partida para la lectura de los números de la lista, y a partir de entonces, los números siguen en orden predeterminado. Cuando se llega al final de la lista, el siguiente número vendrá de vuelta dentro del libro en algún momento.

4.4.2. Técnica de media-cuadrada y otra técnicas de medio-dígito.

La idea aquí es que después de los primeros dígitos en el cálculo de la mayoría de las funciones matemáticas, como $\sin(\)$ o $\log(\)$, los dígitos son entonces esencialmente "aleatorios". Para ilustrar, el código siguiente genera los dígitos de la función $\sin(\)$ después de la quinta como la secuencia pseudo aleatoria. La generación de una secuencia de cien mil de ellos podría tomar medio segundo.

Este es un largo tiempo. En comparación, el generador integrado de MATLAB construye el mismo número de muestras en unos 0.02 segundos. La razón es que el cálculo de la función $\sin(\)$ es una función que relativamente consume mucho tiempo.

Terminal MATLAB

```
> tenpow5 = 100000;
> x = 0.5;
> for i = 1 : 10
    x = tenpow5 * sin(x);
    w = floor(x);    % primeros 5 dígitos
    x = x - w    % dígitos después del 5
> end
% Ahora a calcular algo de tiempo
tic
> for i = 1 : 100000
    x = tenpow5 * sin(x);
```

```

    w = floor(x);
    x = x - w;
> end

toc
% comparar con el generador de Matlab
tic;
x = aleatorio(1,100000);
toc
\ % 0.0193

```

En los primeros días de la computación, calcular una función trascendental, como $\text{sen}(\)$, era mucho más lento que ahora, por ejemplo en relación con el momento de realizar una multiplicación. Por lo tanto, en lugar de $\text{sen}(x)$, von Neumann usó simple multiplicación de x^2 , la semilla al cuadrado. Aquí un código de ejemplo. Un número entero de 7 dígitos se eleva al cuadrado, lo que resulta en un resultado de 13 o 14 dígitos. Los 3 dígitos más significativos se caen entonces los próximos 7 dígitos se guardan y el resto de los dígitos más significativos se desechan. Este se convierte entonces en la siguiente semilla.

Terminal MATLAB

```

> tenpow7 = 1000000;
> x = 0.1234567; % semilla inicial
> for i = 1 : 10
    x = x * tenpow7; % x ahora es un entero
    w = x * x; % al cuadrado
    w = floor(w/1000); % drop últimos tres dígitos
    w = w/tenpow7; % 7 dígitos decimales
    x = w - floor(w) % drop los dígitos enteros
> end

```

Desafortunadamente, el generador de mediana cuadrada sufre de varios problemas, uno de los cuales es un corto período y un ciclo rápido para la mayoría de semillas iniciales. Buscamos para hallar un mejor generador.

4.4.3. Generadores lineales de una congruencia de números aleatorios

El generador de números aleatorios de congruencia lineal es uno de los más extensamente utilizados y simples. Eso depende de una sencilla recurrencia para calcular el próximo término en la secuencia del término anterior. Antes de describir este generador, es necesario revisar los fundamentos de la aritmética modular.

Aritmética modular

Todos los números en esta sección son enteros. Utilizamos la notación $m|x$ en el sentido de que m divide a x (exactamente, sin resto). Por ejemplo, $8|16$, siendo el cociente 2, o $8|-24$ con cociente -3. Se dice que a es congruente con b módulo m , y escribe

$$a \equiv b \pmod{m}, \text{ si } m|(a-b)$$

Por ejemplo $21 \equiv 5 \pmod{8}$ ya que $8|(21-5)$ o igualmente $-19 \equiv 5 \pmod{8}$ ya que $8|(-19-5)$.

Es fácil ver que la congruencia módulo m es:

Reflexiva: $a \equiv a \pmod{m}$,

Simétrica: si $a \equiv b \pmod{m}$ entonces $b \equiv a \pmod{m}$, y

Transitiva: si $a \equiv b \pmod{m}$ y $b \equiv c \pmod{m}$, entonces $a \equiv c \pmod{m}$.

Por transitividad, si $a \equiv b \pmod{m}$ entonces $a - b = pm$ para algún entero p .

Similarmente, si $b \equiv c \pmod{m}$ entonces $b - c = qm$ para algún entero q . Por lo tanto $a - c = b + pm - b + qm = (p + q)m$, y así $m|(a - c)$.

Un hecho importante acerca de los números enteros útiles en el trabajo con la aritmética modular es el algoritmo de la división (en realidad un teorema).

Teorema (Algoritmo de División)

Dado dos números enteros a y m , con $m > 0$, existen enteros únicos q y r , el cociente y el resto, de forma que $a = qm + r$ y $0 \leq r < m$

Por ejemplo, $(14 \pmod{11} + 31 \pmod{11}) \pmod{11} = (3 + 9) \pmod{11} = 1$, sino también $(14 + 31) \pmod{11} = 1$. Estos hechos simples hacen la aplicación de la aritmética modular bastante fácil.

Por otra parte, si a es primo relativo con m (es decir, a y m no tienen factores en común que no sea 1), entonces podemos incluso dividir por un módulo en la aritmética m . Es decir, siempre podemos resolver ecuaciones de la forma $ax \pmod{m} = b \pmod{m}$ para cualquier b . Esto resulta ser muy útil e importante.

4.5. Algunas aplicaciones

El método Monte Carlo es especialmente útil para simular fenómenos con significativa incertidumbre en las entradas y sistemas con un gran número de grados de libertad. Las áreas de aplicación son:

Ciencias físicas.

El método Monte Carlo es muy importante en la física computacional, física química y en la aplicación de campos relacionados. En la física estadística el modelado molecular de Monte Carlo es una alternativa a la dinámica molecular computacional, además es usado para calcular teorías de campos estadísticos de simples partículas y del sistema complejo de partículas. El método Monte Carlo en física es usado para diseñar detectores, la comprensión de su comportamiento y la comparación de los datos experimentales de la teoría.

Ingeniería.

El Método Monte Carlo es usado extensamente en ingeniería para el análisis cuantitativo probabilístico en el diseño de procesos, la necesidad surge de la interacción co-lineal y no lineal del comportamiento de las simulaciones de los procesos típicos.

- En la ingeniería microelectrónica, el Método Monte Carlo es aplicado para el análisis correlativo y no correlativo de variaciones en circuitos integrados análogos y digitales.
- En geostatística y en geometalurgia, el Método Monte Carlo sustenta el diseño de diagramas de flujo del procesamiento de minerales y contribuye al análisis cuantitativo de riesgo.
- En el análisis de producción de la energía eólica, la salida de la energía de una granja durante toda su vida se calcula dando diferentes niveles de incertidumbre.

Biología computacional.

El Método Monte Carlo es usado en varios campos de la biología computacional, por ejemplo, para la inferencia bayesiana en filogenia o para el estudio de sistemas biológicos tales como genomas, proteínas o membranas. En los casos donde esto no es factible en la conducta de un experimento físico, la consideración de los experimentos puede ser por el comportamiento.

Los gráficos en computadora.

El rastreo de caminos, se refiere a Monte Carlo Ray Tracing, hace en 3D una escena de muestras aleatoriamente trazadas de posibles caminos. Un muestreo repetido de cualquier pixel dado eventualmente hará el promedio de las muestras que convergen en la solución correcta, la fabricación de este método es uno de los más precisos en la interpretación de gráficos 3D que existen.

Aplicaciones estadísticas.

En la estadística aplicada, el Método Monte Carlo es generalmente usado para dos propósitos.

1. Para comparar las competencias estadísticas de pequeñas muestras bajo las condiciones de datos realistas. La potencia de las propiedades en estadística pueden ser calculadas para los datos de las distribuciones teóricas clásicas (Curva Normal, la Distribución de Cauchy), los datos reales a menudo no tienen tales distribuciones.
2. Para proporcionar las implementaciones en las pruebas de hipótesis que son más eficientes que las pruebas exactas tales como pruebas de permutaciones.

Una prueba de aleatoriedad aproximada está basada en un conjunto especificado de todas las permutaciones. El enfoque Monte Carlo está basado en un número específico de permutaciones aleatorias.

La inteligencia artificial para los juegos.

El Método Monte Carlo se ha desarrollado en una técnica llamada **Árbol de Búsqueda de Monte Carlo** que es útil para la búsqueda del mejor movimiento en un juego. Los posibles movimientos se organizan en un árbol de búsqueda y un gran número de simulaciones aleatorias se utilizan para estimar cada movimiento.

El árbol de búsqueda del Método Monte Carlo (MCTS) tiene cuatro pasos.

1. Coloca el nodo raíz del árbol, selecciona los nodos óptimos hasta que se alcanza un nodo hoja.
2. Expande el nodo hoja y elige uno de los hijos.
3. Juega un juego simulado a partir de ese nodo.
4. Utiliza los resultados de ese juego simulado para actualizar el nodo.

Diseño y modelos 3D.

El Método Monte Carlo también es eficiente en la solución de ecuaciones diferenciales integrales de los campos de radiación y el transporte de energía y por tanto estos métodos se han utilizado en los cálculos que produce imágenes foto-realistas de modelos 3D virtuales, con aplicaciones en los video juegos, el diseño generado por computadora.

Finanzas y negocios.

El Método Monte Carlo en finanzas a menudo se utiliza para evaluar inversiones en proyectos en una unidad de negocio o nivel corporativo o para evaluar los derivados financieros. Ellos pueden ser usados para modelar los programas del proyecto, donde en las simulaciones se agregan las estimaciones para el peor caso, el mejor caso probable para determinar los resultados del proyecto en general. El Método Monte Carlo se utiliza también en la valoración de opciones, por defecto el análisis de riesgo.

4.5.1. Ejemplos

Recolectando cupones

Las empresas minoristas con frecuencia llevan a cabo un juego en el que el objetivo es recoger todas las letras en una palabra, por ejemplo, su nombre. Una de las letras se proporciona en cada producto. La pregunta es, ¿Cuántos productos puede uno esperar a comprar con el fin de recoger todas las letras?

Por lo general, en la versión comercial del juego, una de las letras es rara; y el juego se degenera con suerte en conseguir por ejemplo una que tiene esa letra. Pero en nuestro juego vamos a suponer que cada letra tiene la misma probabilidad de ser adquirida. Con eso, tenemos muchos productos para comprar y deletrear, digamos, la palabra ANGELFISH de nueve letras.

El siguiente código hará el cálculo. Los resultados se muestran en la Figura 6. Esto demuestra que a pesar de que se producen probablemente igual las letras, la recogida de las 9 cartas podría tener 100 o más compras.

Terminal Matlab

```
> nLetras = 9; %ANGELFISH
> nEnsayos = 10000;
> for i=1: nEnsayos
> suceso = 0;
> nIntentos(i) = 0;
> for j=1: nLetras > ANGELFISH(j)=0; % reinicio de carta, no archivada
> end
> while suceso == 0 >
nIntentos(i) = nIntentos(i)+1; % inc. conteo
> compra = 1+floor(nLetras*rand); % obteniendo letras
> ANGELFISH(compra) = 1;
> if sum(ANGELFISH)==nLetras> suceso = 1;
> end
> end
> end
> hist(nIntentos)
```

Para el problema recolectando cupones, es posible encontrar una solución analítica. Por lo tanto, este es un ejemplo en el cual podemos usar Monte Carlo para hacer un cálculo de probabilidades y para sugerir el camino para una solución analítica.

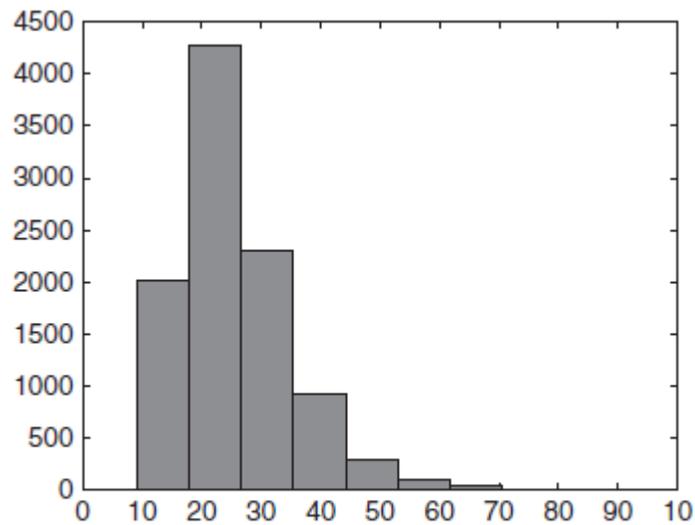


Figura 6: Recuento de las cartas de ANGELFISH

Barajar cartas

Un barajeo puede ser visto como una permutación de algo fijo "en casa" en el orden de las tarjetas. La baraja de 52 cartas estándar consta de 13 cartas dentro de un conjunto y 4 palos. Dentro de un conjunto en un orden canónico podría ser as, dos, tres, . . . , diez, jack, reina y rey. A continuación, los conjuntos se pueden pedir, por ejemplo, el puente de juego de cartas puede ser: trébol, diamantes, corazones y espadas.

Esto identifica cada tarjeta con un número de 1 a 52; en este caso 1 es el as de trébol, 14 el as de diamantes, 27 as de corazones, y el 40 el as de espadas. La generación de una permutación de los números enteros 1 a 52 produce igualmente una reproducción aleatoria de la baraja.

Una investigación sobre la naturaleza de un juego en particular requeriría una baraja uniformemente distribuida en el espacio de permutaciones, es decir, todas las permutaciones igualmente probables. Por supuesto, si después que un barajeo humano, las permutaciones que salen son probablemente no deseados, podemos simular el proceso mecánico de barajar en sí.

Aquí queremos encontrar un algoritmo para generar permutaciones igualmente probables . Para lograr esto, debe ser el caso de que después de haber elegido un conjunto inicial de cartas, la carta siguiente para ser seleccionado debe ser igualmente probable entre las cartas restantes.

Por supuesto, la permutación de identidad, a saber, el orden 1, 2, 3, . . . , 52, también debe generarse con la misma probabilidad para cualquier otra permutación. Y como siempre, el algoritmo debe ser eficiente, así se pueden hacer miles de barajadas de ser necesarios en un experimento.

Otra característica deseable de un barajador de tarjeta es que sea capaz de generar barajadas parciales si es necesario. Esto ahorra tiempo. Por ejemplo, puede que sólo sea necesario conocer los primeros m cartas de la baraja; después las cartas se barajan. El siguiente algoritmo cumple con todos los criterios anteriormente expuestos.

Además, que se lleva a cabo dentro de la matriz original; una segunda matriz no es necesaria. Una carta de la baraja restante es seleccionada para la posición k , k va de 1 a m . La tarjeta seleccionada se intercambia con la tarjeta ya en la posición k y la selección continúa.

Código

```
>n = 4; % una permutación de 4 objetos
> m = n; % hacer toda la baraja
> perm = 1:n %permutación identidad
> for i=1:m % Si m=n la última es superflua
> j = i+floor((1+n-i)*rand); % de las restantes
> cambiar = perm(i);
> perm(i) = perm(j);
> perm(j) = cambiar;
> end
> perm
```

5. Algunas distribuciones de probabilidad y sus usos

El rango de todos los posibles experimentos de probabilidad cubre mucho terreno, desde el resultado del lanzamiento de una moneda ponderada hasta el precio de una acción el próximo mes, desde el tiempo de la siguiente llamada de teléfono en el tablero de un conmutador hasta el peso en el nacimiento del primer hijo de una madre.

Para simular éstos y otros experimentos de probabilidad, necesitamos conocer como convertir números aleatorios distribuidos uniformemente en una muestra de cualquier distribución basadas en el experimento.

En este capítulo presentamos algunas de las principales distribuciones de probabilidad y mostramos como utilizar la computadora para sacar una muestra aleatoria de cada uno de ellos. Esto significa convertir las muestras de U sacadas del intervalo uniforme $[0, 1)$ con distribución $U(0, 1)$ en las muestras de la densidad requerida.

Presentaremos la densidad de probabilidad y la función de distribución acumulada a lo largo de cada distribución con sus medias y varianzas. Vamos a mostrar como las muestras aleatorias pueden ser usadas para el cálculo de probabilidades y resolver problemas.

Las técnicas principales utilizadas para el muestreo son **la inversión CDF, simulación, composición, cartografía y el rechazo**. Nuestro objetivo es introducir estas metodologías de muestreo e ilustrarlas. En muchos casos, es mejor pero más complicado, los métodos disponibles para el muestreo de esas distribuciones.

5.1. Caso discreto CDF - inversión : ensayos de Bernoulli

Supongamos un experimento que sólo tiene dos resultados digamos, "fracaso" y "éxito" (o de vez en cuando 0 y 1, etc). Tal experimento se llama *Ensayo de Bernoulli*. El lanzamiento de una moneda es un ejemplo de un ensayo de bernoulli y también lo es el girar una rueda de ruleta con solo dos sectores. Sea p la probabilidad de éxito y $p - 1$ la probabilidad de fracaso. Entonces la *pdf* de un ensayo de bernoulli es (utilizando el 0 para fracaso y el 1 para éxito)

$$f(x) = \begin{cases} 1 - p, & \text{si } x = 0 \\ p, & \text{si } x = 1 \end{cases}$$
$$= p^x(1 - p)^{1-x}, \quad x = 0, 1.$$

La CDF es

$$F(x) = \begin{cases} 0, & \text{si } x < 0, \\ 1 - p, & \text{si } 0 \leq x < 1, \\ 1, & \text{si } 1 \geq x. \end{cases}$$

Consecuentemente usando las ecuaciones anteriores, la media en el ensayo de Bernoulli es

$$\mu = 0 \cdot (1 - p) + 1 \cdot p = p.$$

y la varianza es

$$\text{var} = (0 - p)^2 (1 - p) + (1 - p)^2 p = p(1 - p)(p + 1 - p) = pq.$$

5.1.1. Inversión: dos resultados de CDF

Dibujando un ensayo de Bernoulli para la muestra T usando $U \sim U(0, 1)$ es muy fácil. Si $U < p$ entonces declaramos que ha sucedido un "éxito", en caso contrario, declaramos un "fracaso". Esto funciona porque $U(0, 1)$ es la densidad $pdf = 1$ para $0 \leq t \leq 1$ y también

$$Pr(T=\text{éxito}) = Pr(U < p) = \int_0^p pdf(t) dt = p.$$

Por ejemplo, supongamos que queremos simular el giro de una rueda de ruleta y ver si el resultado es de color rojo. Hay 18 rojos y 18 sectores negros en la rueda de ruleta de un casino y dos sectores verdes 0 y 00. Por lo tanto, la probabilidad de éxito es $p = \frac{18}{38} = 0.4736$.

Usando el generador de números aleatorios en U uniformemente al azar en $[0, 1)$ y declaramos que "rojo" ocurrió si $U < 0.4736 \dots$, de lo contrario "no rojo" ocurre. Alternativamente, ya que $q = 1 - p = 1 - 0.4736 = 0.5263 \dots$. En ese caso nuestra muestra es "no rojo" y de ser "rojo" si $U \geq 0.5263 \dots$.

Éste no es el mejor método que el de antes, pero es consecuente con la función de distribución acumulativa que se muestra en la Figura 7. La interpretación es que U esta entre 0 y 1 sobre el eje Y. Si U es menor que $q = 1 - p$ volvemos a 0.

Aquí significa "no rojo" y si U es igual o superior a $1 - p$, volvemos a 1, que significa "rojo".

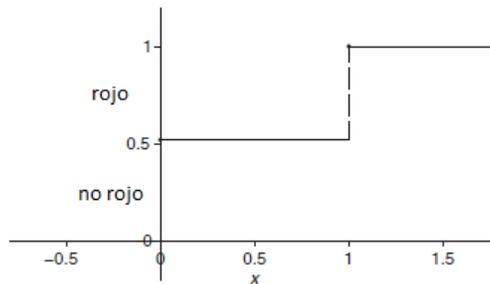


Figura 7: Muestras para el rojo

5.1.2. Distribuciones múltiples resultados

El ensayo de Bernoulli de resultados múltiples es una distribución discreta sencilla. Más generalmente, las distribuciones discretas tienen varios resultados. Si su número no es demasiado largo, podemos usar el gráfico de CDF para sacar las muestras con las frecuencias correctas indicadas anteriormente. Algunos ejemplos harán esto más claro.

Supongamos que queremos simular el número de huevos que pone una de las aves playeras durante la época de crianza. Estas aves tiene $X = 2, 3, 4$ y raramente 5 huevos por nidada. Además, supongamos que las frecuencias observadas de esos tamaños son $p_2 = 0.15$, $p_3 = 0.20$, $p_4 = 0.60$, y $p_5 = 0.05$ respectivamente. La CDF de X se muestra en la siguiente figura

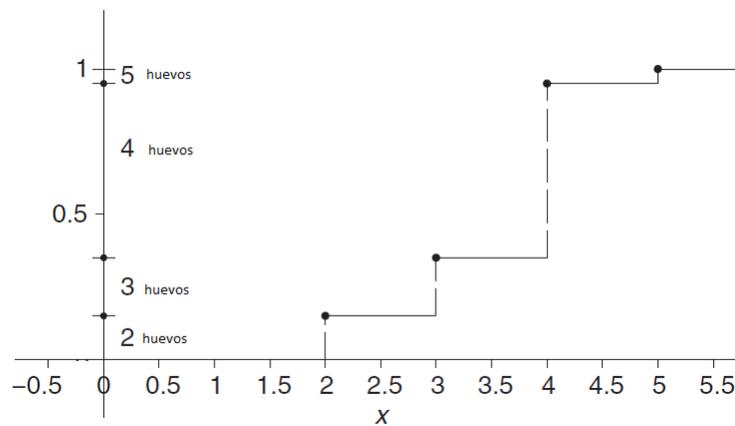


Figura 8: CDF para el tamaño de puesta de huevos de las aves playeras

Sea $U \sim U(0,1)$ una muestra uniforme en $[0,1)$. A partir del punto $(0,U)$ (en el eje Y) avanzar a la derecha hasta encontrar un salto en el CDF, un segmento de línea discontinua vertical. Ahora avanzamos abajo en el eje X y regresamos a este valor como la elección.

Como se indica a lo largo del eje Y, se selecciona a 2 con probabilidad 0.15, ya que el intervalo ocupa una fracción en el intervalo unitario. Del mismo modo, se selecciona a 3 con probabilidad 0.2, su intervalo ocupa una fracción en el intervalo unitario y así sucesivamente.

Esta técnica es llamada *invertir la CDF*. Pero una distribución discreta se puede implementar en la computadora de varias maneras.

El método mas general establece puntos de ruptura en el intervalo $[0,1]$ y luego determina en cual subintervalo de $U \sim U(0,1)$ hay caídas. El siguiente código generará una muestra de X .

Terminal MATLAB

```
> % Establece los resultados  
>oc(1)=2; oc(2)=3; oc(3)=4; oc(4)=5  
> % Establece los puntos de ruptura
```

```

>bp(1)=1.5; bp(2)=bp(1)+0.2; bp(3)=bp(2)+0.6; bp(4)=1
>Fin de lo establecido, ir aqui para conseguir las muestras
> U = aleatorio; k = 1
> while U >= bp(k)
    k = k + 1
    end % k es seleccionado
>selección =oc(k)

```

Una desventaja de este método es que el programa debe pasar por el bucle *while* para cada punto de regreso. La búsqueda puede ser acelerado por la reordenación de los resultados de acuerdo con la disminución de la probabilidad, esto reduce el tiempo previsto en el bucle *while*.

Por ejemplo, poner el primer punto de ruptura en 0.6 que asignamos a 4 huevos, entonces $0.6+0.2 = 0.8$ para 3 huevos y así sucesivamente. Primeramente, el valor esperado a través del bucle es

$$(1) (0.15) + (2) (0.2) + (3) (0.6) + (4) (0.05) = 2.55$$

mientras que el segundo número esperado de veces a través del bucle es

$$(1) (0.6) + (2) (0.2) + (3) (0.15) + (4) (0.05) = 1.65$$

Si bien, el arreglo de las probabilidades nos puede producir una diferencia poco significativa en la rapidez de la obtención de los resultados, en millones de selecciones si se nota la diferencia.

5.2. Método de Alias Walker : selección rueda de ruleta.

Las distribuciones con resultados múltiples con probabilidades asignadas arbitrariamente son como una rueda de ruleta (generalizado). No es pérdida de generalidad asumir que los resultados finitos son designados a $1, 2, 3, \dots, n$. Asumimos que sus correspondientes probabilidades son $p_1, p_2, p_3, \dots, p_n$. Seleccionar un resultado es como girar la ruleta con los sectores que tienen ángulo central $2\pi p_1$ para el resultado 1, $2\pi p_2$ como resultado 2, ... , y $2\pi p_n$ para el resultado n .

El método alias toma más tiempo para establecer la inversión CDF, pero corre más rápido porque no hay ningún bucle de búsqueda. El método utiliza dos matrices, una matriz que ajusta la probabilidad $Q(\cdot)$ y una matriz Alias $A(\cdot)$.

Una vez más demostraremos el método con un ejemplo. En algunas especies la hembra puede afectar a la fracción de los machos versus hembras entre su descendencia que depende de varios factores. Por ejemplo, la hembra Venga de color hurrumbre produce más descendencia masculina si se aparea con un macho grande, pero produce más descendencia femenina si se aparea con un macho pequeño.

La razón es que el macho grande engendrará mayor descendencia y se aparean con mayor frecuencia y extensamente que los machos pequeños. Así que si ella se aparea con un macho grande se pueden propagar más eficientemente los genes en tener machos. Ya que las hembras siempre serán capaces de aparearse si tiene descendencias pequeñas, es mejor tener hembras.

Entonces hay $M = 9$ posibles resultados, dejando $Y = \text{int}[UM]$, la parte entera de el producto UM , donde $U \sim U(0, 1)$, y seleccionamos $0, 1, 2, \dots, 8$, igualmente probable cerca del 11%. Por lo tanto, utilizamos la matriz Q para ajustar esas probabilidades que están más abajo. Mientras tanto, el exceso de probabilidad de selección sobrepasa el 11%.

Algoritmo Alias Walker.

El Método de Alias es una familia de algoritmos eficientes para el muestreo de una distribución de probabilidad discreta, debido a A.J. Walker. Es decir, devuelve valores enteros $1 \leq i \leq n$ de acuerdo con algún p_i con distribución de probabilidad arbitraria. Los algoritmos suelen utilizar $O(n \log(n))$ o $O(n)$ de tiempo de procesamiento previo, después de lo cual los valores aleatorios se pueden extraer de la distribución en el tiempo $O(1)$.

El método funciona así:

Operación

Internamente, el algoritmo consulta dos tablas, una tabla de probabilidad U_i y una tabla de alias K_i (para $1 \leq i \leq n$). Para generar un resultado aleatorio, tiramos un dado para determinar un índice en las dos tablas. Sobre la base de la probabilidad almacenada en ese índice, a continuación, se gira una moneda, y el resultado de la moneda, se utiliza para elegir entre un resultado de i y K_i .

Más concretamente, el algoritmo funciona como sigue:

1. Generar una variable aleatoria uniforme $0 \leq x < 1$.
2. Sea $i = \lfloor nx \rfloor + 1$ e $y = nx + 1 - i$.
3. Si $y < U_i$, volver a i. Este es el lanzamiento de la moneda.
4. De lo contrario, volver K_i .

Una formulación alternativa de la tabla de probabilidades, propuesto por Marsaglia es el método del "histograma cuadrado", utiliza la condición $x < V_i$ en el tercer paso (donde $V_i = (U_i + i - 1)/n$) en lugar de calcular y .

La generación de tablas

La distribución puede ser rellena con probabilidades adicionales $\pi = 0$ para aumentar n a un valor conveniente, tal como una potencia de dos.

Para generar la tabla, primero inicializa $U_i = n\pi$. Mientras hace esto, dividimos las entradas de la tabla en tres categorías:

- El grupo esta "demasiado lleno", donde $U_i > 1$,
- El grupo esta "medio lleno", eran $U_i < 1$ y K_i no se ha inicializado, y
- El grupo esta "exactamente lleno", donde $U_i = 1$ o K_i ha sido inicializado.

Si $U_i = 1$, el valor correspondiente a K_i nunca será consultado y no es importante, pero un valor de $K_i = i$ es razonable.

Mientras que las entradas de la tabla no son todas exactamente completas, repetimos los pasos siguientes:

1. Elige arbitrariamente una entrada "demasiada llena" $U_i > 1$ y una entrada "medio llena" $U_j < 1$. (Si uno de ellos existe, el otro también debe existir).
2. Asignar el espacio no utilizado en la entrada j para el resultado i , mediante el establecimiento de $K_j = i$.
3. Retire el espacio asignado a partir de la entrada i cambiando $U_i = U_i - (1 - U_j) = U_i + U_j - 1$.
4. Entrada de j es ahora exactamente completa.
5. Asignar entrada i en la categoría adecuada sobre la base del nuevo valor de U_i .

Cada iteración se mueve al menos una entrada para la categoría "exactamente completa" (y los últimos dos movimientos), por lo que el procedimiento se da por terminado después de un máximo de $n-1$ iteraciones. Cada iteración se puede hacer en el tiempo $O(1)$, por lo que la tabla se puede configurar en el tiempo $O(n)$.

Vose señala que el error del punto flotante de redondeo puede causar la garantía mencionada en el paso 1 y puede ser violada. Si una categoría vacía antes que la otra, las entradas restantes pueden tener un U_i ajustado a 1, con un error despreciable.

A medida que el procedimiento de búsqueda es un poco más rápido si $y < U_i$ (porque los K_i no tiene por qué ser consultados), una meta durante la generación de la tabla es maximizar la suma de la U_i . Hacer esto de manera óptima resulta ser muy duro, pero una heurística de "Robin Hood" viene razonablemente cerca de: robo de los más ricos y dar a los más pobres. Es decir, en cada paso elija el más grande U_i y el más pequeño U_j . Debido a que esto requiere la clasificación de los U_i , requiere un tiempo $O(n \log n)$.

Eficiencia.

Aunque el método de alias es muy eficiente y si la generación de una desviación uniforme es en sí rápida, hay casos en los que está muy lejos de ser óptimo en términos de uso de bits aleatorios. Esto es debido a que utiliza una variable aleatoria x de precisión completa cada vez, incluso cuando sólo se necesitan unos pocos bits aleatorios.

Un caso surge cuando las probabilidades son particularmente bien equilibradas, por lo que muchos $U_i = 1$ y el K_i no son necesarios. Generando y es una pérdida de tiempo. Por ejemplo, si $p_1 = p_2 = 1/2$, entonces un valor aleatorio x de 32-bits se podría utilizar para hacer 32 opciones, pero el método de alias sólo generará una.

Otro caso surge cuando las probabilidades están fuertemente desequilibrada, por lo que muchos $U_i \approx 0$. Por ejemplo, si $p_1 = 0.999$ y $p_2 = 0.001$, entonces, la gran mayoría de las veces, sólo se requieren unos pocos bits aleatorios para determinar que el caso 1 se aplica.

5.3. Simulación de probabilidad : la distribución binomial

A menudo toda serie de ensayos independientes de Bernoulli, son realizados y estamos interesados en los resultados de todos ellos. Sea la probabilidad de un suceso por p y sea $q = 1 - p$. Sea X el número que representa a los éxitos en una serie de n ensayos. Sea X uno de los valores de $\Omega = \{0, 1, \dots, n\}$, pero no son igualmente probables.

El número de formas en las que puede haber un suceso x en n ensayos esta dado por la ecuación

$$\frac{n(n-1)(n-2)\dots(n-x+1)}{x(x-1)\dots 2 \cdot 1}. \quad (5.1)$$

en el que hay x factores en ambas partes, en el numerador y el denominador.

Como una ilustración de esto, supongamos que los tres símbolos S, s, F son usados para hacer palabras de 4 letras (Usamos S mayúscula y s minúscula en lugar de S_1 y S_2). La regla es que una y solo una S debe ser usada, lo mismo para s. Debería ser $4(3) = 12$ palabras. Ellos son (S en el primer lugar) SsFF, SFsF, SFFs, (S en el segundo lugar) sSFF, FSsF, FFSs, (S en el tercer lugar) sFSF, FsSF, FFSS y (S en el cuarto lugar) sFFS, FsFS, FFSs.

Volviendo al caso general, en realidad el éxito no se distingue, así que si elegimos una de las posibles opciones (de distinta colocación). ¿Cuántas otras opciones no son más que los reordenamientos de los éxitos? Pero esto es solo contando los números con ordenes diferentes del sub índice de S, que es la permutación de el símbolo x . Usando la misma lógica anterior

$$\text{número de permutaciones del símbolo } x = x(x-1)(x-2)\dots(2)(1) = x!. \quad (5.2)$$

Ya que los ensayos son independientes, éste es el producto de $p^x q^{n-x}$ (recordar que p es la probabilidad de éxito y q es la probabilidad de fracaso). Ésta es también la probabilidad de cada una de las otras posibilidades. Por lo tanto, la probabilidad de que halla éxitos esta dada por

$$pdf(x) = Pr(X = x) = \binom{n}{x} p^x q^{n-x} \quad x = 0, 1, 2, \dots, n. \quad (5.3)$$

Para la variable aleatoria X , ésta también es su función de densidad, ésta es la «Distribución binomial», a menudo denotada por $b(n, p)$, n ensayos independientes de Bernoulli con probabilidad p de éxito en cada ensayo.

La ecuación (5.3) esta estrechamente relacionada con el teorema binomial para la expansión del binomio $(a + b)^n$.

$$(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^k b^{n-k}. \quad (5.4)$$

De hecho, sumando (5.3) sobre x , da

$$\sum_{x=0}^n \binom{n}{x} p^x q^{n-x} = (p+q)^n = 1.$$

Como se esperaba, entonces $Pr(\Omega) = 1$.

Media y varianza.

De acuerdo con 4.9, la media de la distribución binomial $b(n, p)$ es

$$\begin{aligned} \mu &= \sum_{x=0}^n x \binom{n}{x} p^x q^{n-x} = \sum_{x=1}^n \frac{n!}{(x-1)!(n-x)!} p^x q^{n-x} \\ &= np \sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x}. \end{aligned}$$

Cambiando las variables, poniendo $y = x - 1$ y $m = n - 1$. Entonces $n - x = m - y$, y entonces,

$$\sum_{x=1}^n \frac{(n-1)!}{(x-1)!(n-x)!} p^{x-1} q^{n-x} = \sum_{y=0}^m \binom{m}{y} p^y q^{m-y} = (p+q)^n = 1.$$

Por lo tanto,

$$\mu = np. \tag{5.5}$$

Como ya hemos visto, que la media es lineal. Entonces la variable aleatoria binomial es la suma de n ensayos independientes de Bernoulli, se deduce que la media de la distribución binomial es la suma de n veces la media de un ensayo de Bernoulli, como asegura (5.5).

La varianza de una variable binomial aleatoria es la suma de las varianzas de los n ensayos independientes de Bernoulli,

$$\text{var} = npq. \tag{5.6}$$

5.3.1. Muestreo de la binomial

Si n es «pequeño», podemos probar X mediante la inversión CDF. Alternativamente y mas sencillo de implementar una muestra de $b(n, p)$ puede ser generado por *Simulación*, que es por n repeticiones del ensayo de Bernoulli subyacente.

Por ejemplo, supongamos que estamos interesados en cuantas veces ocurre rojo en 20 veces consecutivas de la rueda de la ruleta del casino. Lo encontraremos repitiendo el ensayo de Bernoulli 20 veces: Tomamos una muestra de $U = U(0, 1)$, comprobar que si $U < 18/38$ y añadir 1 a la cuenta si es así.

Si n es bastante grande, la simulación es menos deseable. Tal es el caso del siguiente problema. Una mutación se produce a una de las células hijas durante la división celular con probabilidad $p = 10^{-7}$.

5.4. Otra simulación: la distribución de Poisson

Consideremos un problema en el que el número de ensayos de Bernoulli de hecho es muy grande. Éste es el problema de contar el número de éxitos de un fenómeno cuando se producen los éxitos a una velocidad conocida λ . Por ejemplo, supongamos que las llamadas telefónicas que entran al 911 están en el cuadro de distribución aleatorio, pero en promedio de 1 a 10 minutos. ¿Cuántas llamadas llegarán entre las 10:00 y las 10:30 a.m?

En cada segundo de tiempo hay una pequeña probabilidad de que habrá una llamada entrante, por lo que la llegada o no llegada de una llamada entrante, en cualquier segundo dado es un ensayo de Bernoulli. Así, en minutos hay 1800 de tales ensayos de Bernoulli.

Para analizar esto, hacemos cuatro supuestos sobre los eventos del fenómeno:

1. La velocidad del evento λ debe ser positivo y
2. El número de eventos que ocurren son independientes.
3. La probabilidad de exactamente un evento en un intervalo suficiente corto de longitud h es aproximadamente λh .
4. La probabilidad de que dos o más eventos en un intervalo suficientemente corto es despreciable.

Sea X_t la variable aleatoria que denota el número de eventos que ocurren en un intervalo de longitud t . Se divide el intervalo en n sub intervalos de igual longitud $\frac{t}{n}$, donde n es suficientemente grande y que las condiciones 1 y 3 se mantienen.

A continuación, si un evento ocurre en un intervalo dado es un ensayo de Bernoulli con una probabilidad de éxito de $\frac{\lambda t}{n}$, y por lo tanto la probabilidad de x eventos en n sub intervalos es:

$$\binom{n}{x} \left(\frac{\lambda t}{n}\right)^x \left(1 - \frac{\lambda t}{n}\right)^{n-x}.$$

Si $n \rightarrow \infty$ en esta expresión. El grupo de $x!$ en el denominador de $\binom{n}{x}$ con $(\lambda t)^x$, puesto que x es fijo.

Por otro lado, el grupo $1/n^x$ con los factores restantes de $\binom{n}{x}$ se obtiene,

$$\frac{n!}{n^x (n-x)!} = \frac{n(n-1)(n-2)\dots(n-x+1)}{n^x} \rightarrow_{n \rightarrow \infty} 1.$$

El factor restante, $\left(1 - \frac{\lambda t}{n}\right)^{n-x}$ tiende a $e^{-\lambda t}$ como sigue,

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^{n-x} = \lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^n \left(1 - \frac{\lambda t}{n}\right)^{-x} = e^{-\lambda t} 1.$$

Por lo tanto,

$$Pr(X_t = x) = \frac{(\lambda t)^x e^{-\lambda t}}{x!}. \quad (5.7)$$

Esta es la densidad de la *distribución de Poisson*. Note que λt es un número puro (sin dimensiones). Si la velocidad del evento de $\lambda = 1/10$ minutos y queremos respuestas sobre un periodo de 30 minutos, entonces $\lambda t = 3$.

La densidad de Poisson también se aplica a los "eventos" distribuidos en el espacio así como en el tiempo. Por ejemplo, supongamos que los nudos que se producen en un determinado lote de madera es un proceso de Poisson con una velocidad de 2 nudos y 8 clavos en promedio.

¿Cuál es la probabilidad de encontrar un clavo y no tener nudos? Por (5.7), ya que $\lambda t = \frac{2}{8} \times 8$, aquí es

$$Pr(X_8 = 0) = \frac{(2)^0 e^{-2}}{0!} = e^{-2} = 0.135.$$

Note que $0! = 1$ (por definición).

Ya que la media de la binomial $b(n, \frac{\lambda t}{n})$ es $n \frac{\lambda t}{n} = \lambda t$ y es independiente de n , es natural creer que la media del proceso de Poisson correspondiente es λt .

Recordemos el problema de mutación de la sección anterior. La probabilidad p del ensayo de Bernoulli para un proceso de Poisson es $\frac{\lambda t}{n}$, por eso, $\lambda t = np$. Por lo tanto, la aproximación de Poisson de esta binomial es

$$Pr(X = x) = \frac{(np)^x e^{-np}}{x!} = \frac{(0.1)^x e^{-0.1}}{x!}.$$

para $n = 10^6$ y $p = 10^{-7}$. Compare esto con (5.3). Para $x = 0$ la probabilidad exacta es $(1 - p)^n = 0.904837331$ para nueve lugares decimales mientras que la aproximación es $e^{-0.1} = 0.904837418$.

Aunque la distribución de Poisson se puede utilizar como una aproximación a la binomial, todavía tenemos el problema de dibujar las muestras de ella. Como veremos a continuación, la técnica de muestreo para la distribución de Poisson es exactamente por simulación. Pero ambas, la distribución de Poisson y binomial para n grande se puede mostrar una buena aproximación usando el teorema del límite central y la distribución binomial.

5.4.1. El muestreo de la distribución de poisson por simulación

El siguiente método se explicará en la sección 5.5. Es esencialmente una técnica de simulación y es exacta.

Sea $U_0 = 1$ y U_i sean muestras de $U(0, 1)$ para $i = 0, 1, 2, \dots$. Regresa el mayor entero de x tal que

$$\prod_{i=0}^x U_i > e^{-\lambda t}.$$

Está será una muestra de la distribución de Poisson con parámetro λ .

5.5. Caso continuo inversión - CDF : distribución exponencial

La distribución exponencial se utiliza para modelar el tiempo aleatorio para el tiempo de espera para eventos a ocurrir. Un ejemplo de lo que entendemos por eventos son los clientes que llegan a una fila, el fracaso de los componentes mecánicos y nacimientos o muertes biológicas. Los supuestos acerca de estos eventos son los mismos que los realizados para la distribución de Poisson. De hecho, existe una íntima conexión entre los dos.

La variable aleatoria exponencial es el tiempo de espera (o distancia) para el primer evento de Poisson que se produzca. Dado que la probabilidad de que no ocurra un evento de Poisson es el intervalo $[0, t)$ es $e^{-\lambda t}$, como se ve tomando $x = 0$ en los eventos (5.7), tenemos

$$\begin{aligned} Pr(\text{un evento ocurre en } [0, t)) &= 1 - Pr(\text{el evento no ocurre en } [0, t)) \\ &= 1 - e^{-\lambda t}. \end{aligned} \tag{5.8}$$

Ésta es la CDF de la distribución exponencial

$$CDF = F(t) = 1 - e^{-\lambda t}.$$

La distribución exponencial tiene un único parámetro λ , la velocidad de evento. Denotemos la distribución por $E(\lambda)$. Por su naturaleza, como en el período en el cual un evento ocurre, la exponencial es una distribución continua. Por lo tanto, la densidad exponencial es la derivada de su CDF

$$\text{pdf} = F'(t) = \lambda e^{-\lambda t}.$$

Del mismo modo, las probabilidades están dadas por integrales. Sea W el tiempo de espera aleatorio antes de que ocurra un evento exponencial. Entonces, la probabilidad de que W ocurra en algún subconjunto S de los números reales es

$$Pr(W \in S) = \int_S \lambda e^{-\lambda t} dt.$$

Propiedades de la distribución exponencial

La distribución exponencial es «sin memoria» (también llamado Markoviano). Por ejemplo, si un evento todavía no ha ocurrido por un tiempo, entonces la probabilidad de que este ocurra en el intervalo de "a" a $a + b$ es el mismo como que este ocurriendo en 0 a "b".

$$Pr(W < a + b | W \geq a) = Pr(W < b).$$

Esto es así porque en términos de la CDF, el lado izquierdo es exactamente

$$\frac{F(a+b) - F(a)}{1 - F(a)} = \frac{(1 - e^{-\lambda(a+b)}) - (1 - e^{-\lambda a})}{1 - (1 - e^{-\lambda a})} = 1 - e^{-\lambda b}.$$

La media de la exponencial es $\mu = 1/\lambda$, porque usando integración por partes con $u = t$, $du = \lambda e^{-\lambda t}$, tenemos

$$\mu = E(W) = \int_0^{\infty} t \lambda e^{-\lambda t} dt = \frac{1}{\lambda}. \quad (5.9)$$

Un cálculo similar muestra que la varianza es

$$\text{var} = 1/\lambda^2. \quad (5.10)$$

5.5.1. Inversión CDF - el método canónico para la exponencial

El método para obtener una muestra para $E(\lambda)$ es simple. Sea $U \sim U(0, 1)$ y resolvamos $U = 1 - e^{-\lambda W}$ para W . Esto dará

$$W = -\frac{1}{\lambda} \ln(1 - U).$$

y W es la muestra deseada. Notar que la mayoría de computadoras que generan números aleatorios devuelven números estrictamente menores que 1, en cuyo caso el argumento del algoritmo no será 0. Este es un método con un propósito general, y su explicación esta dada por el siguiente teorema.

Teorema 2.1

(Invirtiendo la función de distribución acumulada, caso continuo). Sea $F(x)$ la función de distribución de la variable aleatoria X , supongamos que F es continua, y sea $U = F(X)$. Entonces U es uniformemente distribuida en $[0, 1)$.

Prueba

Necesitamos demostrar que $Pr(U \leq u) = u$ para $0 \leq u \leq 1$. Ya que F es continua y creciente, F^{-1} existe. Si $x = F^{-1}(u)$, entonces

$$Pr(U \leq u) = Pr(F(X) \leq u) = Pr(X \leq F^{-1}(u)) = Pr(X \leq x) = F(x) = u.$$

Así si U es una muestra para $U(0,1)$, entonces $X = F^{-1}(U)$ es una muestra para la CDF de F .

Invirtiendo la CDF también puede trabajar para una distribución que es una mezcla de componentes discretos y continuos.

Caso general: invirtiendo la distribución acumulada.

Muestra $U \sim U(0, 1)$; regresando

$$X = \inf F^{-1}([U, 1]) = \inf \{\xi : F(\xi) \geq U\}$$

El principal inconveniente con el método de invertir el CDF es que este usualmente requiere una fórmula explícita para la CDF, que a menudo no están disponibles. Sin embargo en los casos en los que esta disponible, es una técnica muy poderosa.

El muestreo de la distribución de Poisson.

Puesto que la variable aleatoria de Poisson es el número de eventos exponenciales que se producen en el intervalo $[0, t)$, podemos simular la distribución de Poisson por muestreo de exponenciales hasta que se excede el tiempo t . Damos aquí tres métodos para hacer esto:

Método 1: Dejar que W_i sea $E(\lambda)$ y dejar que K sea el mayor número entero tal que $\sum_{i=1}^K W_i < t$.

Método 2: Dejar que Y_i sea $E(1)$ y dejar que K sea el mayor número entero tal que $\sum_{i=1}^K Y_i < \lambda t$. Esto trabaja porque si Y es $E(1)$, entonces $X = (\lambda Y)$ es $E(\lambda)$ y $\sum_{i=1}^k Y_i < \lambda t$ si y solo si $\sum_{i=1}^k X_i < t$. Este método es una mejora, ya que uno no tiene que dividir por λ en cada muestra.

Método 3: A partir del método 2 toma de registros. Sea $U_0 = 1$ y $U_i \sim U(0, 1)$ para $i = 1, 2, \dots$. Sea k el entero mayor tal que $\prod_{i=0}^k U_i > e^{-\lambda t}$.

Este trabajo es porque $-\sum_{i=1}^k Y_i < -\lambda t$ si y solo si $\prod_{i=0}^k U_i > e^{-\lambda t}$ y Y_i es $E(1)$ si y solo si $e^{-\lambda t}$ es $U(0, 1)$.

Tiempo de espera independientes múltiples.

Supongamos que un proceso A con tiempo de espera W_A dada por una distribución exponencial con parámetros λ_A , y un proceso B con tiempo de espera W_B y una distribución exponencial con parámetro λ_B . La pregunta es ¿cuál es la distribución para el proceso que consiste en que se produzca A o B? Suponemos que los procesos A y B son independientes, uno puede ver

$$Pr(A \text{ o } B \text{ ocurre en } [t, t + \Delta t] \mid \text{no antes}) \approx \lambda_A \Delta t + \lambda_B \Delta t.$$

Con la aproximación que se convierte en exacta cuando $\Delta t \rightarrow 0$. Por lo tanto, la hora del evento para el proceso combinado es exponencial con parámetro $\lambda_A + \lambda_B$. Alternativamente, podemos argumentar para la CDFs, así

$$\begin{aligned} Pr(W_A \text{ o } W_B < t) &= Pr(W_A < t) + Pr(W_B < t) - Pr(W_A \text{ y } W_B < t), \\ &= (1 - e^{-\lambda_A t}) + (1 - e^{-\lambda_B t}) - (1 - e^{-\lambda_A t})(1 - e^{-\lambda_B t}), \\ &= 1 - e^{-(\lambda_A + \lambda_B)t}. \end{aligned}$$

dando el mismo resultado.

5.5.2. Simulación de eventos discretos

Hay muchas simulaciones en los negocios y la industria en la que es necesario analizar los procesos estocásticos en tiempos de espera. Éste es un campo de la teoría de colas. Si uno desea tener un análisis exacto de tal proceso, es necesario que los tiempos de espera sean distribuidos exponencialmente. Ésto es así porque la exponencial es la única distribución que tiene la propiedad «sin memoria». Es decir, el futuro depende solo del presente y no como llegó el presente.

Aún así, cualquier cosa más allá de la simple organización es difícil analizarlo directamente, por ello se convierte en Simulación Monte Carlo, se conoce como simulación de eventos discretos en este campo. Entre sus muchos beneficios, la simulación permite que cualquier distribución para el tiempo de espera sea utilizado tan fácilmente como cualquier otro.

Debido a su importancia, hay varias soluciones de software dedicados a brindar respuestas para el más intrincado de las situaciones de cola. Mucho se ha escrito acerca de esta área, incluyendo el importante problema matemático de simulación de análisis de resultados. En esta sección damos una necesaria y breve descripción para uno de los principales enfoques la simulación de eventos discretos.

Ejecutor de simulación.

Un software de simulación es administrado por un "gestor de simulación y la invocación de otros módulos cuando sea necesario. Estos módulos incluyen el procesamiento de las entidades de la simulación, la aplicación de sus interdependencias y visualización de datos y registro de los módulos.

Se mantiene un calendario de eventos, siempre en orden cronológico y el evento producto de la simulación es el próximo evento en la programación. Cada evento tiene una etiqueta que describe cuando esto ocurre en tiempo absoluto y su naturaleza. Cuando un evento se maneja, puede generar nuevos eventos cada uno de los cuales es programado para el muestreo en la distribución adecuada en el tiempo de espera y la unión dentro del programa maestro en el tiempo fijado.

Consideremos el ejemplo de simulación del crecimiento de tejido celular, digamos que comenzamos con una célula en el tiempo 0 y llevando la simulación durante 8 horas. Asumimos que el tiempo de maduración celular es distribuida exponencialmente con parámetro $\lambda = 0.5$ por hora. Necesariamente el programa maestro contiene solamente el tiempo de los eventos, ya que todos los eventos son lo mismo, a saber la fisión binaria de células idénticas.

Una simulación más ambiciosa podría atraer el apiñamiento y otros efectos en las células adecuadas. Dibujando $E(0.5)$, podríamos decir que obtenemos $W = 3$ horas. Así que muestra célula original se divide en el tiempo $t = 3$ y nuestro programa maestro en $S[1] = 3$.

Ahora movemos el tiempo adelante de $t = 3$ y manejemos la división celular de la primera célula. Sacamos $E(0.5)$ de nuevo para obtener $W = 4$, es decir, para la primera célula hija, así ahora $S[2] = 7$ (al momento actual es $t = 3$ y el tiempo de la división celular es 4, así el tiempo del evento total es 7). Y de nuevo para la segunda célula hija, digamos que obtenemos $W = 1$. Ahora debemos cambiar el $S[2] = 7$ abajo para mantener el programa en el tiempo de forma ordenada, así $S[2] = 4$ y $S[3] = 7$.

El próximo evento en la programación es en el tiempo $t = 4$, la división de una célula hija de la primera generación. Muestra para $E(0.5)$ para obtener $W = 2$, es decir, su tiempo total será $4+2=6$ por lo que el programa maestro se convierte en $S[3]=6$ y $S[4]=7$.

En $W = 5$ el evento ocurre en un tiempo total de $t = 9$, que está más allá de nuestro estudio, así que no es necesario incluirlo en el programa maestro. (Pero se tendría que tener en consideración).

Continuando de esta manera, llegamos al punto en el que todos los eventos posteriores se producen más allá del tiempo $t = 8$ y la simulación ha terminado.

5.5.3. La transformación de variables aleatorias: la distribución de Cauchy

En lo anterior se dijo que si Y es $E(1)$, entonces $X = (1/\lambda)Y$ es $E(\lambda)$. Transformando o mapeando las variables aleatorias, esto ocurre frecuentemente, y es muy útil. En general, una transformación de Y sobre X esta dada por $X = g(Y)$ para alguna función g uno a uno. Para encontrar la distribución de X (o para verificar que tiene alguna distribución indicada) se puede examinar ya sea la pdf o CDF de X .

En nuestro ejemplo específico, sabemos que $CDF_Y(y) = Pr(Y \leq y) = 1 - e^{-y}$, ya que Y es $E(1)$. Por lo tanto, tenemos la siguiente cadena de igualdades

$$CDF_X(x) = Pr(X \leq x) = Pr(1/\lambda Y \leq x) = Pr(Y \leq \lambda x) = 1 - e^{-\lambda x}.$$

Así, la CDF de X es $1 - e^{-\lambda x}$ si X es $E(\lambda)$.

Alternativamente, la pdf puede ser igualada con el fin de llevar a cabo, la técnica de cambio de variables en la integración. Sea $t = g(\tau)$ y $A = [a, b]$ es un intervalo. Poner $B = g^{-1}(A)$ (Recordar que g es uno a uno). Entonces

$$\int_A f(t)dt = \int_B f(g(\tau))g^{-1}(\tau) d\tau. \tag{5.11}$$

La aplicación de las densidades es la siguiente donde $X = g(Y)$

$$\begin{aligned} Pr(X \in A) &= Pr(g(Y) \in A), \\ &= Pr(Y \in g^{-1}(A)) = \int_{g^{-1}(A)} f_Y(\tau)d\tau. \end{aligned}$$

donde f_Y es la pdf de Y . Pero por un cambio de variables

$$Pr(X \in A) = \int_A f_X(t) dt = \int_{g^{-1}(A)} f(x)(g(\tau))g'(\tau) d\tau.$$

donde f_X es la pdf de X . Igualando los 2 miembros finales

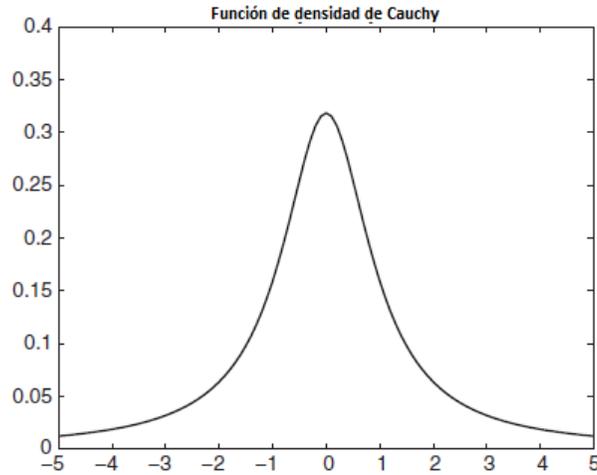


Figura 9: Probabilidad de la Función densidad de Cauchy

$$f_Y(\tau) = f_X(g(\tau))g'(\tau). \quad (5.12)$$

Para nuestro mapeo exponencial $f_Y(\tau) = e^{-\tau}$, $g(\tau) = \tau/\lambda$ y anticipamos $f_X(\tau) = \lambda e^{-\lambda\tau}$.

Sustituyendo en (5.12) da

$$e^{-\tau} = \lambda e^{-\lambda(\tau/\lambda)} (1/\lambda).$$

es una identidad y por lo tanto hemos verificado que el pdf es X .

Desafortunadamente (2.12) esta implícita en f_X . Para obtener una expresión explícita, tenemos que aplicar el cambio de variable para g^{-1} en lugar de g . Al hacerlo da

$$f_X(x) = f_Y(g^{-1}(x)) (g^{-1})'(x) \quad (5.13)$$

Como un ejemplo más, trabajamos la «Densidad de Cauchy $f_X(t)$ ».

Sea $Y \sim \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ y $X = \tan(Y)$. Entonces podemos calcular que

$$f_X = \frac{1}{\pi} \frac{1}{1+t^2}. \quad (5.14)$$

Para ver esto, nos damos cuenta de que $g(t) = \tan(t)$, así que $g^{-1}(t) = \tan^{-1}(t)$ y por lo tanto

$$f_X(t) = \frac{1}{\pi} \frac{d}{dt} (\tan^{-1}(t)) = \frac{1}{\pi} \frac{1}{1+t^2}.$$

como se desea.

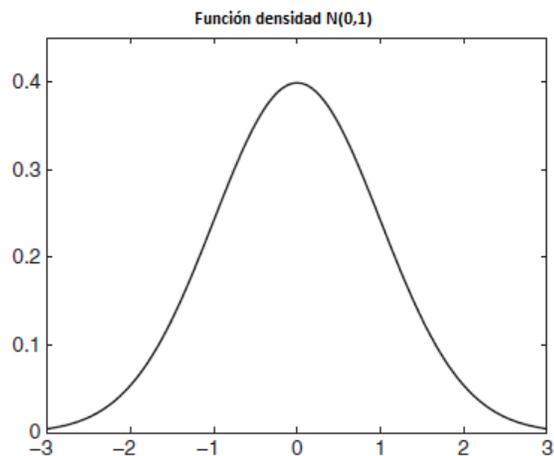


Figura 10: Función densidad normal estándar

La densidad de Cauchy es simétrica cerca de $t = 0$ y así su media μ_X es 0. Pero la varianza de la distribución de Cauchy es infinita, ya que la integral

$$\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{t^2}{1+t^2} dt. \quad \text{es ilimitada}$$

Antes del trabajo de Gauss, se pensaba que la distribución de Cauchy era candidato a la distribución normal, es decir, la distribución del teorema del límite central. Aunque la pdf es una de las dos distribuciones que tienen apariencia similar, las colas de la distribución de Cauchy es demasiado «gorda», como lo evidencia el hecho de que la distribución normal tiene varianza finita pero la distribución de Cauchy no.

5.5.4. El teorema del límite central y de la distribución normal

En muchos de los histogramas generados por los problemas que hemos trabajado, hasta ahora la figura que ha surgido es la de una curva en forma de campana. Esto no es por accidente esta es consecuencia predicha por el Teorema del Límite Central, que ahora consideraremos.

La gráfica de la función $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ (veáse Figura 10) es simétrica cerca del origen, en todas las partes positiva, en forma de campana y tiende a 0 cuando $x \rightarrow \pm\infty$. Además, la integral de f sobre todos los reales es $\sqrt{2\pi}$ y por lo tanto,

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

es la función de probabilidad de densidad.

Reemplazando x por x/σ , donde σ es un parámetro, tiene el efecto de reescalar el eje x en términos de σ , esto es, reescalar la función en $x = 2\sigma$, tiene el mismo valor como la función original en $x = 2$. Como un efecto secundario, para mantener la integral unidad, así como f debe ser dividida por σ .

Finalmente, reemplazando x por $x - \mu$, tiene el efecto de desplazar la función a $x = \mu$, tiene el mismo valor de la original en $x = 0$.

Con estos cambios, la «función de probabilidad normal de densidad» es

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}. \quad (5.15)$$

Denotemos esta distribución por $N(\mu, \sigma^2)$. Por el principio de la transformación para variables aleatorias, esto es fácil de ver, que si X es la distribución para $N(\mu, \sigma^2)$, entonces $Y = \frac{X - \mu}{\sigma}$ es la distribución para $N(0, 1)$.

La media y la varianza de la distribución es $N(\mu, \sigma^2)$ muy apropiadamente μ, σ^2 respectivamente.

5.5.5. El muestreo de la distribución normal.

Discutiremos tres métodos de muestreo para la distribución normal. Para una comparación de estos métodos.

Nuestro primer método aprovecha el teorema del limite central y es crudamente aproximado. Su ventaja es que éste es rápido y fácil para programar.

Algoritmo del límite central.

Muestra de n copias en $U(0, 1)$ de variables aleatorias U_1, U_2, \dots, U_n en $U(0, 1)$ y volvemos

$$X = \frac{\sum_{i=1}^n U_i - \frac{n}{2}}{\sqrt{n/12}}$$

Por el teorema del limite central X es aproximadamente normal con parámetros $\mu = 0$ y $\sigma = 1$, que es $N(0, 1)$.

Algoritmo Box-Muller.

Este ejemplo produce una muestra exacta. Muestra variables aleatorias $U_1, U_2 \sim U(0, 1)$ y ponemos

$$\begin{aligned} X_1 &= \cos(2\pi U_1) \sqrt{-2 \ln(U_2)} \\ X_2 &= \text{sen}(2\pi U_1) \sqrt{-2 \ln(U_2)} \end{aligned} \quad (5.16)$$

Entonces X_1 y X_2 son variables aleatorias independiente con distribuciones normalmente con media 0 y varianza 1.

Veamos la prueba de este algoritmo:

Sea $f_{X_1, X_2}(x_1, x_2)$ la densidad conjunta de X_1 y X_2 y sea $f_{U_1, U_2}(u_1, u_2)$ la densidad conjunta de U_1 y U_2 . Entonces

$$f_{U_1, U_2}(u_1, u_2) = f_{U_1}(u_1) f_{U_2}(u_2) = 1, \quad \text{si } 0 \leq u_1, u_2 < 1,$$

y 0 en otro caso. Sea τ el siguiente cambio de variables

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \tau(u_1, u_2) = \begin{pmatrix} \cos(2\pi u_1) \sqrt{-2 \ln(u_2)} \\ \text{sen}(2\pi u_1) \sqrt{-2 \ln(u_2)} \end{pmatrix}$$

Elevando al cuadrado a x_1 y x_2 nos da

$$x_1^2 + x_2^2 = -2(\ln(u_2)) \cos^2(2\pi u_1) - 2(\ln(u_2)) \text{sen}^2(2\pi u_1) = -2\ln(u_2).$$

Así,

$$u_2 = e^{-\frac{1}{2}(x_1^2 + x_2^2)}. \quad (5.17)$$

Esto demuestra que $0 < u_2 \leq 1$. La variable u_1 puede ser restringida de manera similar debido a la periodicidad del seno y el coseno. Por la técnica del cambio de variables, la relación entre las densidades puede escribirse como:

$$f_{X_1, X_2}(\tau(u_1, u_2)) \det(\tau') = f_{U_1, U_2}(u_1, u_2)$$

recordemos que $f_{X_1, X_2}(x_1, x_2) = 1/\det(\tau')$.

Ahora encontremos el determinante

$$\begin{aligned} \det(\tau') &= \det \begin{pmatrix} \frac{\partial x_1}{\partial u_1} & \frac{\partial x_1}{\partial u_2} \\ \frac{\partial x_2}{\partial u_1} & \frac{\partial x_2}{\partial u_2} \end{pmatrix} \\ &= \det \begin{pmatrix} -2\pi \text{sen}(2\pi u_1) \sqrt{-2 \ln(u_2)} & \frac{-(1/u_2)}{\sqrt{-2 \ln(u_2)}} \cos(2\pi u_1) \\ -2\pi \cos(2\pi u_1) \sqrt{-2 \ln(u_2)} & \frac{-(1/u_2)}{\sqrt{-2 \ln(u_2)}} \text{sen}(2\pi u_1) \end{pmatrix} \\ &= \frac{2\pi}{u_2} \text{sen}^2(2\pi u_1) + \frac{2\pi}{u_2} \cos^2(2\pi u_1) = \frac{2\pi}{u_2} \end{aligned}$$

Por la ecuación (5.17), tenemos que

$$[\det(\tau')]^{-1} = \frac{u_2}{2\pi} = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)}$$

y

$$f_{X_1, X_2}(x_1, x_2) = 1 \cdot \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_1^2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x_2^2}.$$

Con esto queda demostrado la Box-Muller.

Aunque podría parecer que conocer a X_1 imparte alguna información sobre X_2 , esto no es así, X_2 todavía puede ser cualquier valor real absoluto. Un ejemplo se muestra en la Figura 11. El gráfico marca $X_1 = 0.7$ representa gráficamente los puntos $U(X_1, X_2)$ en el plano satisfaciendo la primera ecuación de 5.16 para el valor dado de X_1 , en este caso 0.7.

El gráfico marca $X_2 = 2.0$ muestra los puntos que satisfacen la segunda ecuación de 5.16 dado para el valor de X_2 . Como se muestra, hay un único par (U_1, U_2) dado con estas condiciones

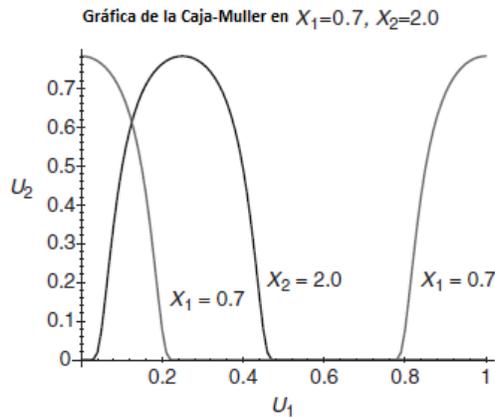


Figura 11: Resolviendo (U_1, U_2) para un (X_1, X_2) dado.

Algoritmo Marsaglia-Bray

Este método también tiene dos muestras exactas similares al algoritmo Box-Muller:

- Muestra de variables aleatorias $U_1, U_2 \sim U(0, 1)$ y ponemos $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$,
 $S = V_1^2 + V_2^2$
- Si $S \geq 1$ ir a 1, otra vuelta.

$$X_1 = V_1 \left(-\frac{2}{s} \ln(s) \right)^{1/2}$$

$$X_2 = V_2 \left(-\frac{2}{s} \ln(s) \right)^{1/2}$$

Entonces X_1 y X_2 son variables aleatorias independientes con distribución normal con media 0 y varianza 1.

El algoritmo de Marsaglia-Bray se derivó de una mejora en la velocidad sobre el algoritmo de Box-Muller, ya que ésta solo usa una en la evaluación de una función trascendental, mientras que la Box-Muller se usa en los tres.

Esto es compensado en cierta medida por el hecho de que una fracción en el paso 1 es rechazada y se debe repetir. ¿Que es está fracción? ahora la velocidad de la computadora y la eficiencia de la evaluación en la función trascendental ha hecho estas cuestiones sean discutibles en la mayoría de los casos.

La prueba del algoritmo de Marsaglia-Bray de muestras exactas para $N(0, 1)$ es similar para el algoritmo de la Box-Muller.

5.5.6. Muestreo aproximado a través del teorema límite central

Supongamos que queremos una muestra de los resultados de 1000 ensayos de Bernoulli con probabilidad de éxito $p = 0.995$. Esto podría corresponder al número de bombillas aceptables para salir de un proceso de fabricación en una hora.

Claro que podríamos simular los 1000 ensayos, o podríamos en su lugar explotar el TLC como sigue. Sea R_{1000} la variable aleatoria para el número de éxitos, y sea μ y σ^2 referidas a las variables aleatorias del ensayo de Bernoulli, $\mu = p$ y $\sigma^2 = pq$.

Entonces N se define por

$$N = \frac{R_{1000} - 1000\mu}{\sqrt{1000\sigma^2}} = \frac{R_{1000} - 995}{2.2305}$$

es aproximadamente $N(0, 1)$.

Ahora sea n una muestra para $N(0, 1)$, decimos $n = 0.68$. Entonces $r = 2.2305n + 995 = 996.52 \approx 997$ redondeo a un entero, esto es correspondiente a la muestra para R_{1000} .

5.6. Ejemplo rechazo de muestreo: La distribución Beta

El algoritmo de Marsaglia-Bray para la distribución normal introduce una nueva idea en el muestreo, es decir, una prueba impuesta a los valores que son candidatos en el orden que proceden. Esta idea puede ser expandida sobre la base para formar una potente y completo método de muestreo, en general llamado Muestreo de Rechazo.

Consideremos el problema de muestreo de un ensayo de Bernoulli con probabilidad de $\frac{1}{3}$ usando el lanzamiento de monedas. Veamos una solución: lanzamos la moneda dos veces, usaremos sus siglas en inglés (H=Cara y T=Cruz) si el resultado es HH (Cara Cara), devuelve «éxito», si tenemos HT o TH, devuelve «fracaso», y si es TT, rechazo de ensayo y volvemos otra vez.

Esto da las probabilidades correctas porque

$$\begin{aligned} Pr(\text{devuelve éxito}) &= Pr(\text{devuelve éxito} \mid \text{hay un regresar}) \\ &= \frac{Pr(\text{hay un regreso y es éxito})}{Pr(\text{hay un regreso})} \\ &= \frac{1/4}{3/4} \\ &= \frac{1}{3} \end{aligned}$$

Pero añadiendo *el criterio de aceptación/rechazo*, sólo los valores relativos de las probabilidades. Es decir, que podría haber sido $1/7$, $1/7$, $1/7$ y «éxito» será devuelto $1/3$ con el tiempo.

Consideremos, un segundo ejemplo, utilizaremos el lanzamiento de tres monedas, queremos seleccionar «1» con probabilidad $2/5$, «2», «3», y «4» con probabilidad $1/5$. Ya que el lanzamiento de una moneda dos veces tiene probabilidad de $1/4$ y esto es menos que $2/5$, el objetivo es dividir las probabilidades por $M = 8/5$ (traer 2 a 1), recordando que solo sus valores relativos importan.

El objetivo es modificar las probabilidades que son $1/4$, $1/8$, $1/8$, y $1/8$, para encontrar la disciplina mas aceptable. Razonamos de la siguiente manera, si el resultado que ocurre es HH, aceptamos eso todo el tiempo y regresa «1», porque $\frac{1/4}{1/4} = 1$. Si el resultado que ocurre es HT, aceptamos esto la mitad del tiempo y mapeamos su regreso para «2», dado esto la probabilidad correcta es $\frac{1/8}{1/4} = \frac{1}{2}$.

Similarmente, aceptamos TH y TT la mitad del tiempo y mapeamos el resultado para «3» y «4» respectivamente. Si un valor es rechazado, comenzamos otra vez y lanzamos dos monedas otras vez.

La probabilidad de aceptar un valor es éste: todo el tiempo para HH y $1/2$ para los otros, ya que

$$\frac{1}{4} \cdot 1 + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} + \frac{1}{4} \cdot \frac{1}{2} = \frac{5}{8}$$

o 62.5 % del tiempo, así 37.5 % del tiempo debe comenzar otra vez. Cuando el valor es regresado, la probabilidad que éste sea «1» esta dado por

$$Pr(\text{el valor regresado es "1"}) = \frac{Pr(\text{"1" es seleccionado y regresado})}{Pr(\text{algún valor es regresado})} = \frac{\frac{1}{4} \cdot 1}{5/8} = \frac{2}{5}$$

De la misma manera para «2», tenemos

$$Pr(\text{el valor regresado es "2"}) = \frac{Pr(\text{"2" es seleccionado y regresado})}{Pr(\text{algún valor es regresado})} = \frac{\frac{1}{4} \cdot \frac{1}{2}}{5/8} = \frac{1}{5}$$

y las otras probabilidades del objetivo son igualmente consideradas que tienen frecuencias adecuadas.

Para llevar a cabo actualmente este método, lanzamos una moneda dos veces, si el resultado es HH, entonces devuelve "1", si el resultado es HT, TH o TT, entonces tenemos que hacer el experimento de un lado y resolver si se aceptan los resultados y devolver su valor asignado o rechazar los resultados y empezar de nuevo.

Ya que el experimento es 50-50 en la decisión, podría ser en sí mismo un lanzamiento de moneda, es decir, que H significa aceptar y T significa rechazo. Por ejemplo, Si el tercer lanzamiento es H entonces en el caso HT devuelve "2", en el caso de TH devuelve "3" y en el caso de TT devuelve "4". Si este es T entonces empezamos otra vez.

La Figura 12 ilustra la situación del ejemplo. Para cada posible resultados de los dos lanzamientos de la moneda, hay que ver la probabilidad resultante para obtener el valor correcto.

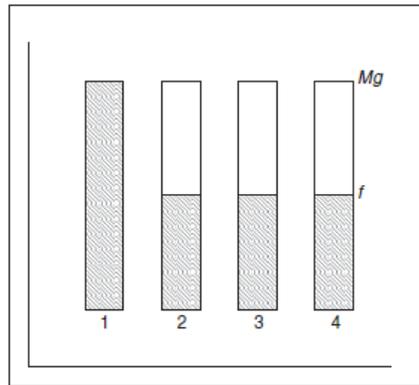


Figura 12: Densidad de f y propuesta de densidad Mg

La densidad de la muestra en realidad, los lanzamientos de la moneda en los ejemplos anteriores, se llama la **densidad propuesta g** . Puede ser cualquier densidad conveniente que es distinto de cero, donde la densidad objetivo f es distinto de cero. Sea $h(\cdot)$ es el criterio de aceptación.

El método de muestreo rechazo generalizado consta de dos pasos simples.

1. Sea Y una muestra para la pdf g .
2. Con probabilidad $h(Y)$ devolver $X = Y$, en otro caso vamos a (1).

Como en el ejemplo el paso 2 es un ensayo de Bernoulli con probabilidad de éxito $h(Y)$ y es llevado a cabo a través de una parte del experimento.

Refiriéndonos otra vez a la Figura 12, veamos que $M = 8/5$ y debemos escalar la propuesta g (en el ejemplo, una distribución uniforme dada por el lanzamiento de dos monedas) por $8/5$ para tener $f \leq Mg$.

Entonces para cada resultado dibujamos para g , la proporción de f que se encuentra bajo la parte de g dada la probabilidad de aceptación. Para el primer resultado de "1" vemos que $f = 2g$, así que es siempre aceptado. Para el resto $f = (1/2)g$ (la parte sombreada representa solo $1/2$ de la barra vertical en esa posición), y así aceptamos solo con probabilidad $1/2$.

Las muestras de rechazo es un método completamente general para muestras con distribución continua o discreta. Para ver que funciona en el caso continuo, calculemos la probabilidad infinitesimal que devuelve el valor de X igual a x como sigue:

$$\begin{aligned} Pr(X = x)dx &= \frac{Pr(Y = x \text{ y } x \text{ es devuelto})}{Pr(\text{algún valor es devuelto})} \\ &= \frac{g(x) dx h(x)}{\int_{-\infty}^{\infty} g(z) h(z) dz}. \end{aligned}$$

Por lo tanto la función de probabilidad de densidad del valor devuelto es

$$Pdf_X(x) = \frac{g(x) h(x)}{\int_{-\infty}^{\infty} g(z) h(z) dz}.$$

Para construir un algoritmo de muestreo sea f el objetivo de la pdf, sea g la propuesta de pdf y sea la constante M tal que $f(x) \leq Mg(x)$, $\forall x$. El criterio de aceptación/rechazo es

$$h(x) = f(x)/(Mg(x)). \tag{5.18}$$

Con esto, la pdf de los valores aceptados es

$$\frac{g(x) h(x)}{\int gh} = \frac{f(x)/M}{\int f(x)/M dz} = f(x).$$

Como anteriormente, la probabilidad de una aceptación es

$$Pr(Y \text{ es aceptado}) = \int_{-\infty}^{\infty} g(z) h(z) dz = 1/M.$$

Ésta es la eficiencia del generador, así mantenemos a M como el más pequeño posible. La Figura 13 ilustra la situación para f y g continuas.

Notemos que éste no es un problema si $g(x) = 0$ para algún x , ya que solo usamos (5.8) para calcular $h(x)$ para valores de x que son muestra para una distribución con pdf g . La probabilidad de obtener a x tal que $g(x) = 0$, y esa situación nunca ocurrirá.

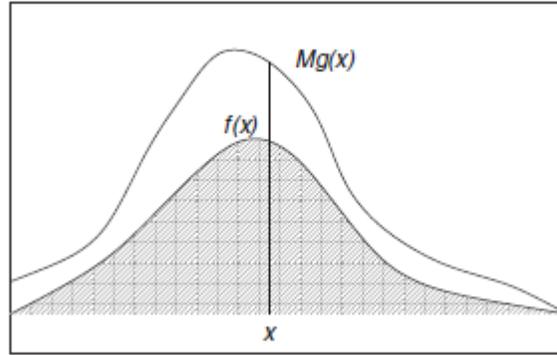


Figura 13: Método de rechazo para el caso continuo.

5.6.1. El muestreo de la distribución Beta

La distribución Beta, $Be(\alpha, \beta)$, es definida como la densidad dada por

$$f(x) = cx^{\alpha-1}(1-x)^{\beta-1}, \quad 0 \leq x \leq 1. \quad (5.19)$$

donde la constante c es elegida para hacer $\int_0^1 f(x) dx = 1$. Los parámetros α y β pueden ser cualquier valor real positivo. Si cualquiera α y β es menor que 1, la densidad es sin cota, pero la integral es finita. La media de la distribución beta es

$$\mu = \frac{\alpha}{\alpha + \beta}.$$

y la varianza es

$$\text{var} = \frac{\alpha\beta}{(\alpha + \beta + 1)(\alpha + \beta)^2}.$$

Supongamos que queremos dibujar las muestras para $Be(2, 3)$, cuya densidad es $f(x) = 12x(1-x)^2$, $0 \leq x \leq 1$, ver figura 2.13. Usaremos la uniforme $U(0, 1)$ como la propuesta, así $g(x)=1$, $0 \leq x \leq 1$, ya que $f(x) \leq 16/9$ (para la figura), podemos tomar $M = 16/9$. Entonces el criterio de aceptación queda así

$$h(x) = \frac{f}{Mg} = \frac{12x(1-x)^2}{16/9 * 1} = \frac{27}{4}x(1-x)^2. \quad (5.20)$$

En el algoritmo de muestreo de rechazo se convierte en (1) muestra $U \sim U(0, 1)$ y (2) con probabilidad $h(U)$ devuelve $X = U$, en otro caso vamos a (1).

Una corrida podría ir de la siguiente manera, $U \sim U(0, 1) = 0.88531386$, es decir, $h(U) = 0.0786$ como dado por (2.20). Ahora, el muestreo $R \sim U(0, 1)$ como parte del experimento en paralelo, digamos, $R = 0.616$. Éste no es menor que o igual a 0.0786 , así, $U = 0.88531386$ es rechazado. Volvemos a intentar esta vez $U \sim U(0, 1) = 0.694688$, es decir, $h(U) = 0.4371$. Otra llamada a $U(0, 1)$ produce

a $R = 0.00506$, que es menor que o igual a 0.4371 , por lo que el experimento tiene éxito. Esta vez $U = 0.384$ es aceptado y se devuelve.

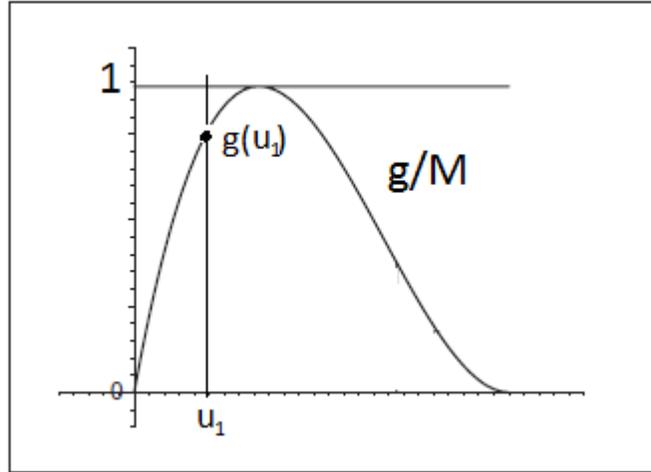


Figura 14: La distribución β para $12x(1-x)^2$

Este ejemplo muestra que uno puede ver el muestreo de rechazo como un ejemplo de dar- fallar. Notar que el valor escogido U es aceptado si $\frac{16}{9}R \leq f(U)$ y rechazado si $\frac{16}{9}R \geq f(U)$. Trazamos el punto $(U, \frac{16}{9}R)$ en la figura, y vemos que U es aceptado si se encuentra por debajo de la gráfica de f y rechazado si esta por encima (o en) la gráfica.

2.7.2 Muestreo de una distribución beta sin cota.

Tomamos un ejemplo de muestras de rechazo. Esta vez buscamos muestras para una distribución beta sin cota. Por simplicidad tomamos $\alpha = \beta$ con $\alpha < 1$ (así que la densidad es sin cota si $x=0,1$), pero trabajamos la misma idea más generalmente, la densidad es

$$f(x) = cx^{\alpha-1}(1-x)^{\alpha-1}, \quad 0 \leq x \leq 1.$$

donde c es el factor normalizador. Recordar que solo los valores relativos de f son necesarios, por lo que no es necesario encontrar a c . Esta es una importante característica del método de muestreo de rechazo.

Como la distribución propuesta introducimos una composición dividiendo el dominio en sub intervalos $[0, \frac{1}{2})$, $[\frac{1}{2}, 1)$. Tendremos más que decir acerca de la composición en la próxima sección. Por simetría de esta $f(x)$ necesitamos muestras de X del intervalo $[0, \frac{1}{2})$ y con probabilidad $\frac{1}{2}$, devolvemos ya sea X o $1 - X$.

Por lo tanto, como densidad propuesta usamos

$$g(x) = c'x^{\alpha-1}, \quad 0 \leq x \leq \frac{1}{2}.$$

La CDF es la integral

$$G(x) = \int_0^x c' t^{\alpha-1} dt = c' \frac{x^\alpha}{\alpha}.$$

ya que $G(1/2) = 1$, obtenemos $c' = \alpha / (\frac{1}{2})^\alpha$

Una muestra para g por la inversión CDF implica resolver

$$U = \frac{c'}{\alpha} Y^\alpha$$

para Y . Obtenemos

$$Y = \left(\frac{\alpha}{c'} U \right)^{1/\alpha} = \frac{1}{2} U^{1/\alpha}.$$

donde hemos sustituido por c' .

Ya que $\alpha < 1$, el valor más grande de $(1-x)^{\alpha-1}$ para $0 \leq x \leq \frac{1}{2}$, ocurre cuando $x = \frac{1}{2}$. Por lo tanto, podemos tomar $M = (\frac{1}{2})^{\alpha-1} c'$ y con esta elección que tenemos

$$\begin{aligned} Mg(x) &= \left(\frac{(\frac{1}{2})^{\alpha-1} c'}{c'} \right) (c' x^{\alpha-1}) = (\frac{1}{2})^{\alpha-1} c x^{\alpha-1}, \\ &\geq (1-x)^{\alpha-1} c x^{\alpha-1}, \\ &= f(x), \quad 0 \leq x \leq 1/2. \end{aligned}$$

Finalmente, el criterio de aceptación esta dada por

$$h(x) = \frac{f(x)}{Mg(x)} = \frac{c x^{\alpha-1} (1-x)^{\alpha-1}}{(\frac{1}{2})^{\alpha-1} c x^{\alpha-1}} = (2(1-x))^{\alpha-1}.$$

Como se predijo c cae fuera y no es necesariamente su valor.

El algoritmo para $Be(\alpha, \alpha)$ es:

1. Generar $U \sim U(0, 1)$ y poner $Y = (\frac{1}{2} U^{1/\alpha})$.
2. Como parte del experimento, si $R \sim U(0, 1) < h(Y)$ entonces acepta a Y y vamos al paso 3, si es otro volvemos a paso 1.
3. Muestra de nuevo $U \sim U(0, 1)$, si $U < \frac{1}{2}$ donde $X = Y$, en otro caso volver a $X = 1 - Y$.

6. Repaso de probabilidad

6.1. Distribuciones y variables aleatorias

Definición 6.1.1

Sea S el mismo espacio para el experimento. Una función de valor real que está definida en S es llamada una *Variable Aleatoria*.

Distribución de una variable Aleatoria

Definición 6.1.2

Sea X una variable aleatoria. La distribución de X es una colección de todas las probabilidades de la forma $\mathbb{P}r(X \in C)$, para todo conjunto C de números reales tal que $\{X \in C\}$ es un evento.

Distribuciones discretas

Definición 6.1.3

Distribución discreta/Variable aleatoria. Decimos que una variable aleatoria E tiene una distribución discreta o que \mathcal{X} es una variable aleatoria discreta si \mathcal{X} solo puede tomar un número finito k de diferentes valores x_1, x_2, \dots, x_n o a lo sumo, una secuencia infinita de diferentes valores x_1, x_2, \dots .

Definición 6.1.4

Función Probabilidad/p.f. Si X es una variable aleatoria que tiene una distribución discreta, la función probabilidad (p.f) de X , está definida como la función f tal que para todo número real x , $f(x) = \mathbb{P}r(X = x)$

El cierre de la secuencia $\{x : f(x) > 0\}$ se llama el apoyo de la distribución de X .

Teorema 6.1.5

Sea X una variable aleatoria discreta con p.f. Si x no es un posible valor de X , entonces $f(x) = 0$. También, si la secuencia x_1, x_2, \dots , incluye todos los posibles valores de X , entonces $\sum_{i=1}^{\infty} f(x_i) = 1$.

Teorema 6.1.6

Si X tiene una distribución discreta, la probabilidad de cada subconjunto de C de la recta real puede ser determinada por la relación $\mathbb{P}r(X \in C) = \sum_{x_i \in C} f(x_i)$.

Definición 6.1.7

Distribución de Bernulli/Variable aleatoria. Una variable aleatoria Z que toma solo dos valores 0 y 1 con $\mathbb{P}r(Z = 1) = p$, tiene una distribución de Bernulli con parámetro p . También, decimos que Z es una variable aleatoria Bernulli con parámetro p .

Distribución uniforme en números enteros

Definición 6.1.8

Sea $a \leq b$ enteros. Supongamos que el valor de una variable aleatoria X es la misma en cada uno de los enteros a, \dots, b . Entonces, se dice que X tiene una distribución uniforme en a, \dots, b .

Teorema 6.1.9

Si X tiene una distribución uniforme en los enteros a, \dots, b , la función de probabilidad de X es:

$$f(x) = \begin{cases} \frac{1}{b-a+1}, & \text{para } x = a, \dots, b \\ 0 & \text{otro caso} \end{cases}$$

Distribución Binomial

Definición 6.1.10

Distribución binomial/Variable Aleatoria. La distribución discreta representada por la función de probabilidad

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{para } x=0, 1, \dots, n \\ 0 & \text{otro caso} \end{cases}$$

Una variable aleatoria con esta distribución se dice que es una *distribución binomial con parámetro n y p* .

Distribución Continua

Definición 6.2.1

Distribución continua/Variable aleatoria. Decimos que una variable aleatoria X tiene una distribución continua o que X es una variable aleatoria continua si existe una función f no negativa, definida en la recta real, tal que para cada intervalo de números reales, la probabilidad que X toma un valor en el intervalo es la integral de f en este

$$\mathbb{P}r(a \leq X \leq b) = \int_a^b f(x) dx.$$

donde $\mathbb{P}r(X \geq a) = \int_a^\infty f(x) dx$ y $\mathbb{P}r(X \leq b) = \int_{-\infty}^b f(x) dx$.

Definición 6.2.2

Función de densidad de probabilidad/pdf. Si X tiene una distribución continua, la función f descrita en la definición 3.2.1 es llamada la función de densidad de probabilidad de X . La cerradura del conjunto $\{x : f(x) > 0\}$ es llamado el soporte de la distribución de X . Cada pdf de f debe satisfacer los dos requerimientos siguientes:

$f(x) \geq 0$, para todo x y $\int_a^b f(x) dx = 1$.

6.2. Comentarios sobre paquete de software.

Los sistemas de software matemáticos como MATLAB, GNU Octave, Mathematica, tienen colecciones de generadores de números aleatorios con varias distribuciones, por ejemplo, uno puede generar pseudo números aleatorios con distribución uniforme en el intervalo $(0, 1)$. Ellos son particularmente útiles para graficar y mostrar generadores de puntos aleatorios.

7. Método Monte Carlo y simulación

7.1. Números aleatorios

En este capítulo se da un énfasis particular a los problemas en la cuál la simulación envuelve un elemento de cambio ya que la aleatoriedad es parte de este método.

Comenzaremos dando el concepto de números aleatorios. Consideremos una sucesión de números reales x_1, x_2, \dots todos en el intervalo unitario $(0, 1)$. Informalmente, la secuencia es aleatoria, si los números al parecer son distribuidos a lo largo del intervalo, sino estos forman un patrón x_1, x_2, \dots .

Por ejemplo, si tenemos los números decimales que comienzan con el dígito 3, entonces los números son agrupados en el subintervalo $0.3 \leq x \leq 0.4$ y no son distribuidos aleatoriamente en el intervalo $(0, 1)$. Si los números son monótonamente decrecientes, ellos no son aleatorios. Una definición precisa de aleatoriedad es difícil formular.

Algoritmos y generadores de números

Los sistemas de computadora tienen «Generadores de Números Aleatorios», estos son procedimientos que producen un único número aleatorio o todo un conjunto de números aleatorios en cada llamada, a este procedimiento lo llamaremos «Aleatorio».

Por ejemplo, los generadores de números aleatorios están contenidos en un sistema de software matemático como MATLAB, Maple, Mathematica, entre otros.

Estos devuelven uno o un conjunto de pseudo números aleatorios «*uniformemente distribuidos*» en el intervalo unitario $(0, 1)$, dependiendo si el argumento es una variable, un escalar o un conjunto. El generador de números aleatorios puede producir cientos de miles de números pseudo aleatorios, antes de repetirlo, por lo menos teóricamente.

Una secuencia de números es «*uniformemente aleatorios*» en el intervalo $(0, 1)$, es si ningún subconjunto del intervalo contiene más que su parte de números que le corresponde.

El código de computadora que produce los números aleatorios no pueden ser verdaderamente aleatorios, la manera en que ellos son producidos es completamente «*determinista*», ningún elemento al azar esta realmente presente. Algunos autores prefieren enfatizar este punto llamando a dicha sucesión «*pseudo número aleatorio*».

Un programa generador de n números aleatorios x_1, x_2, \dots, x_n uniformemente distribuidos en el intervalo abierto $(0, 1)$ es el siguiente:

Octave

```
function y = aleatorio(n, l0)
```

```
% aleatorio.m genera un arreglo de enteros aleatorio a partir de l0 que pueden estar entre 1 y 2^31-1
```

```
% Ejemplo
```

```
% y=aleatorio(6,3830461261); y
```

```
% mayor número de dígitos de precisión
```

```
format long;
```

```
% en caso el usuario no den o l0 proveemos valores
```

```
if nargin == 1;
```

```
    l0 = 3537
```

```
    elseif nargin == 0;
```

```
        n = 1;
```

```
end
```

```
% dar espacio para las variables
```

```
l = zeros(1, n + 1);
```

```
x = zeros(1, n);
```

```
l(1) = l0;
```

```
for i = 1 : n
```

```
    l(i + 1) = rem(7^5 * l(i), 2^31 - 1);
```

```
    x(i) = l(i + 1)/(2^31 - 1);
```

```
end
```

```
% dar respuesta final
```

```
y=x;
```

allí todo l_i es entero en el rango de $1 < l_i < 2^{31} - 1$. El entero inicial l_0 es llamado la «semilla» de la sucesión y es seleccionado como cualquier entero ente 1 y el número primo $2^{31} - 1 = 2.1447483647$.

Una función externa produce o genera una nueva matriz de números pseudo aleatorios y podría basarse en el siguiente código:

Octave

```
function y=aleatorio1(n,semilla)
```

```
% aleatorio1.m genera números aleatorios a partir de la semilla
```

```
% ejemplo
```

```

% y=aleatorio1(6,265437)
% variables
k = 16807;
j = 2147483647;
x = zeros(1, n);
% seleccionar el valor de la semilla
if nargin==1
semilla=3537;
elseif nargin==0;
n=1;
end
for i=1:n
semilla= mod(k*semilla,j);
x(i)=semilla/j;
end
% respuesta
y=x;

```

Estas aplicaciones no podrían producir aleatoriedad de alta calidad y no podrían ser adecuado para aplicaciones que requieren una estadística exacta o criptografía.

En pocas palabras la precaución a cerca de los generadores de números aleatorios es necesario en algunas simulaciones, el fracaso de aleatoriedad puede llevar a una errónea conclusión.

Hay tres puntos que debemos recordar:

1. Los algoritmos de tipo ilustrado aquí por aleatorios y aquellos que producen secuencias periódicas, esto es, una secuencia que se repite eventualmente. El periodo de este es 2^{30} un número bastante grande.
2. Si un generador de números aleatorios es usado o produce puntos aleatorios en un espacio n-dimensional, estos puntos se encuentran en un número pequeño de planos.
3. Los dígitos individuales que hacen los números aleatorios generados por rutinas como el de Aleatorio no son, en general, dígitos aleatoriamente independientes. Por ejemplo, podría pasar que el dígito 3 siga al dígito 5 con mayor frecuencia de lo esperado.

Como una comprobación de un generador de números aleatorios, calculamos una larga serie de números aleatorios y determinaremos que la proporción de ellos se encuentra en el intervalo $\left(0, \frac{1}{2}\right)$. La respuesta de la computadora debe ser aproximadamente del 50%.

Este experimento descrito también puede ser interpretado como la simulación del lanzamiento de una moneda. Un solo lanzamiento corresponde a la selección de un número aleatorio x en el intervalo de

(0, 1). Asociamos a H a los eventos $0 < x < \frac{1}{2}$ y T a los eventos $\frac{1}{2} < x < 1$. Mil lanzamientos de una moneda corresponde a 1000 elecciones de un número aleatorio. Teóricamente si los números aleatorios fueron verdaderamente aleatorios, el valor del límite como el número de números aleatorios usados aumenta sin límite y sería exactamente el 50 %.

Usos del pseudo código aleatorio

Ilustraremos el uso de los procedimientos aleatorios correctos e incorrectos productores de puntos distribuidos uniformemente.

Consideremos el problema de generar 1000 puntos aleatorios uniformemente distribuidos dentro de la elipse $x^2 + 4y^2 = 4$. Una manera de hacerlo es generar puntos aleatorios en el rectángulo $-2 \leq x \leq 2$, $-1 \leq y \leq 1$ y descartar aquellos que no se encuentran en la elipse.

OCTAVE

```
%elipse.m
u=4*aleatorio1(2000,265437)-2;
v=2*aleatorio1(2000,374501)-1;
x=zeros(1,1000);
y=zeros(1,1000);
j=1;
for i=1:2000
if u(i)^2+4*v(i)^2<=4
x(j)=u(i);
y(j)=v(i);
j=j+1;
end
end
plot(x,y,'or')
```

Un generador de números aleatorios produce una secuencia de números que son aleatorios en el sentido de que están distribuidos uniformemente a lo largo de un cierto intervalo, tal como $[0, 1)$ y no es posible predecir el siguiente número en la secuencia de saber los anteriores. La esperanza es que la nueva secuencia es más aleatoria que la original. Por ejemplo, una reproducción aleatoria puede eliminar cualquier correlación de los sucesores cercanos de un número en una secuencia.

Existen pruebas estadísticas que se pueden realizar en una secuencia de números aleatorios. Si bien este tipo de pruebas no certifican la aleatoriedad de una secuencia, que son particularmente importantes en las aplicaciones. Por ejemplo, son útiles en la elección entre los diferentes generadores de números aleatorios, y es reconfortante saber que el generador de números aleatorios que se utiliza han pasado

estas pruebas. Existen situaciones en las que los generadores de números aleatorios son útiles a pesar de que no pasen las pruebas rígidas para una verdadera aleatoriedad.

Por lo tanto, si uno está produciendo matrices aleatorias para ensayos de un código de álgebra lineal, entonces una estricta aleatoriedad puede no ser importante. Por otra parte, la estricta aleatoriedad es esencial en la integración Monte Carlo y otras aplicaciones.

7.2. Estimaciones de áreas y volumen por la Técnica Monte Carlo

Integración numérica

Ahora veremos las integrales, siendo primero la aproximación de una integral definida por el Método Monte Carlo. Si seleccionamos los primeros n elementos x_1, x_2, \dots, x_n de una secuencia aleatoria en el intervalo de $(0, 1)$, entonces

$$\int_0^1 f(x)dx \approx \frac{1}{n} \sum_{i=1}^n f(x_i).$$

Aquí, la integral es aproximada por el promedio de los n números $f(x_1), f(x_2), \dots, f(x_n)$. Cuando se lleva a cabo eficientemente, el error es de orden $\frac{1}{\sqrt{n}}$, que no es competitivo con un buen algoritmo, como el Método de Romberg.

Sin embargo, en mayor dimensiones, el Método Monte Carlo es bastante atractivo,

$$\int_0^1 \int_0^1 \int_0^1 f(x, y, z) dx dy dz \approx \frac{1}{n} \sum_{i=1}^n f(x_i, y_i, z_i).$$

donde (x_i, y_i, z_i) es una secuencia aleatoria de n puntos en el cubo unitario $0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1$.

Para obtener puntos aleatorios en el cubo, tenemos una secuencia aleatoria en $(0, 1)$ denotada por $\xi_1, \xi_2, \xi_3, \xi_4, \dots$, para obtener el primer número aleatorio P_1 en el cubo tenemos $P_1 = (\xi_1, \xi_2, \xi_3)$, el segundo será $P_2 = (\xi_4, \xi_5, \xi_6)$, etc.

En el caso general del intervalo (a, b) , el promedio de f con los n puntos aleatorios de (a, b) simplemente no es una aproximación de la integral sino más bien para $\frac{1}{b-a} \int_a^b f(x)dx$.

Similarmente, en mayor dimensión, el promedio de f sobre la región es obtenida integrando y dividiendo por el área, volumen o medida de la región.

Por ejemplo:

$$\frac{1}{8} \int_1^3 \int_{-1}^1 \int_0^2 f(x, y, z) dx dy dz.$$

es el promedio de f sobre el paralelepípedo descrito por las tres desigualdades $0 \leq x \leq 2, -1 \leq y \leq 1, 1 \leq z \leq 3$. Si los (x_i, y_i) denotan un punto aleatorio apropiada con distribución uniforme, el siguiente ejemplo ilustra la Técnica Monte Carlo.

$$\int_0^5 f(x)dx \approx \frac{5}{n} \sum_{i=1}^n f(x_i).$$

$$\int_2^5 \int_1^6 f(x,y)dxdy \approx \frac{15}{n} \sum_{i=1}^n f(x_i, y_i).$$

En cada caso, los puntos aleatorios deben ser uniformemente distribuidos en la región involucrada. En general, tenemos que

$$\int_a^b f \approx (\text{medida de } A) \times (\text{promedio de } f \text{ sobre los } n \text{ puntos de } A).$$

Aquí, usamos el hecho que el promedio de f de una función en un conjunto es igual a la integral de la función sobre el conjunto dividido por el promedio del conjunto.

Ejemplo y pseudo código.

Consideremos el promedio para obtener el valor numérico de la integral

$$\int \int_{\Omega} \text{sen}(\sqrt{\ln(x+y+1)})dxdy = \iint_{\Omega} f(x,y)dxdy.$$

sobre el disco en el espacio xy , definimos la desigualdad $\Omega = \{(x,y) : (x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 \leq \frac{1}{4}\}$.

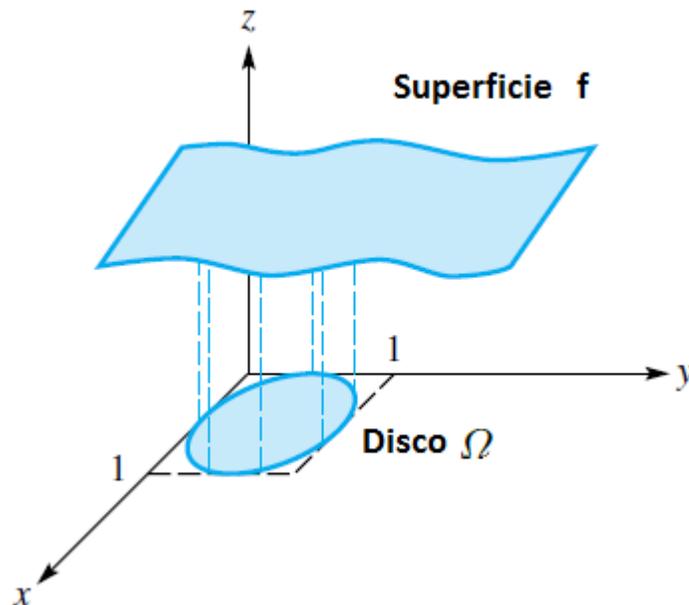


Figura 15: Superficie $f(x,y)$ por encima del disco

Vamos a generar puntos aleatorios en el cuadrado y descartamos los que no se encuentran en el disco.

Si los puntos son $P_i = (x_i, y_i)$, entonces la integral estimada es, donde Pro= Promedio de la longitud de los puntos aleatorios de f ,

$$\int_a^b f \approx (\text{área}\Omega) \times (\text{Pro}) = (\pi r^2) \left(\frac{1}{n} \sum_{i=1}^n f(P_i) \right) = \left(\pi \left(\frac{1}{2} \right)^2 \right) \left(\frac{1}{n} \sum_{i=1}^n f(P_i) \right) = \frac{\pi}{4} \frac{1}{n} \sum_{i=1}^n f(P_i) = \frac{\pi}{4n} \sum f(P_i)$$

La estimación intermedio para la integral está dada cuando n es un múltiplo de 1000.

Calculando volumen

El volumen de un región complicada en 3-dimensiones puede ser calculada por la técnica Monte Carlo. Tomando un caso, determinemos el volumen de la región cuyos puntos satisfacen las desigualdades

$$\begin{aligned} 0 \leq x \leq 1, 0 \leq y \leq 1, 0 \leq z \leq 1, \\ x^2 + \text{sen}(y) \leq z, \\ x - z + e^y \leq 1. \end{aligned}$$

La primera linea define un cubo cuyo volumen es 1. La región definida por todas las desigualdades dadas es un subconjunto de este cubo. Si generamos n puntos aleatorios en el cubo y determinemos que m de ellos que satisfacen las últimas dos desigualdades, entonces el volumen de la región deseada es aproximadamente $\frac{m}{n}$.

Programa volumen de una región.

Enteros i, m ;

Arreglo real $(r_{ij})_{1:n \times 1:3}$;

Reales vol, x, y, z

Enteros $n = 5000, iprt = 1000$

Llamando Aleatorio $((r_{ij}))$

for $i = 1$ a n hacer

$x = r_{i,1}$

$y = r_{i,2}$

$z = r_{i,3}$

If $x^2 + \text{sen}(y) \leq z, x - z + e^y \leq 1$ entonces $m = m + 1$

If $\text{mod}(i, iprt) = 0$ entonces $vol = \text{real}(i) / \text{real}(j)$

output i, vol

end if

end for

end program

Observe que las estimaciones intermedios se imprimen cuando llegemos a 1000, 2000, . . . , 5000 puntos. Un valor aproximado de 0,14 se determina por el volumen de la región.

Ejemplo del Cono

Consideremos el problema de encontrar el volumen arriba del cono $z^2 = x^2 + y^2$ y dentro de la esfera $x^2 + y^2 + (z - 1)^2 = 1$. Como se muestra en la figura

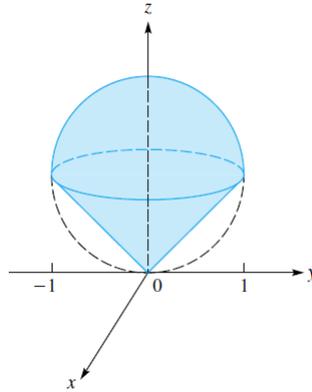


Figura 16: Región del cono de sorbete

Programa Cono

Enteros i, m ;

Reales vol, x, y, z ;

Arreglo real $(r_{ij})_{1:n \times 1:3}$

Enteros $n = 5000, iprt = 1000, m = 0$

Llamando Aleatorio $((r_{ij}))$

for $i = 1$ a n hacer

$x = 2r_{i,1} - 1, y = 2r_{i,2} - 1, z = 2r_{i,3}$

if $x^2 + y^2 \leq z^2, x^2 + y^2 \leq z(2 - z)$ entonces $m = m + 1$

if mod(i, iprt) = 0 entonces vol=8 real(m)/real(i)

output i, vol

end if

end for

end programa Cono

7.3. Simulación

Ilustraremos la idea de *Simulación*. Consideremos una situación en la cual un elemento aleatorio está presente y trataremos de imitar la situación en computadora. Se pueden extraer conclusiones estadísticas si el experimento se ha realizado muchas veces. Las aplicaciones de simulación incluyen servidores, clientes, y colas de como podrían ocurrir en los negocios.

Problema del dado injusto

En los problemas de simulación, debemos a menudo producir variables aleatorias con una distribución preescrita.

Supongamos, por ejemplo, que queremos simular el lanzamiento de un dado cargado y que las probabilidades de diversos resultados se han determinado como se muestra:

<i>Resultados</i>	1	2	3	4	5	6
<i>Probabilidad</i>	0.2	0.14	0.22	0.16	0.17	0.11

Si la variable aleatoria x se distribuye de manera uniforme en el intervalo $(0, 1)$, a continuación, al dividir este intervalo en seis sub-intervalos de longitudes dadas por la tabla, podemos simular el lanzamiento de este dado cargado. Por ejemplo, estamos de acuerdo en que si x está en $(0, 0.2)$, la matriz muestra 1; si x está en $[0.2, 0.34)$, la matriz muestra 2, y así sucesivamente.

Un pseudocódigo para contar el resultado de 5.000 lanzamientos del dado y calcular la probabilidad podría escribirse de la siguiente manera:

Programa Dado _ Cargado

Enteros i, j

Arreglo real $(y_i)_{1:6}, (m_i)_{1:6}, (r_i)_{1:n}$

Real $n = 5000$

$(y_i)_6 = (0.2, 0.34, 0.56, 0.72, 0.89, 1.0)$

$(m_i)_6 = (0.0, 0.0, 0.0, 0.0, 0.0, 0.0)$

Llamando Aleatorio $((r_{ij}))$

fora $i = 1$ a n hacer

 for $j = 1$ a 6 hacer

 if $r_i < y_i$ entonces

$m_j = m_j + 1$

 exit loop j

 end if

 end for

end for

```

output real (m_i) / real (n)
end programa Dado_Cargado

```

Los resultados son 0,2024, 0,1344, 0,2252, 0,1600, 0,1734, y 0,1046, que son aproximaciones razonables a las probabilidades de la tabla.

El Problema de la Aguja de Buffon

El siguiente ejemplo de una simulación es un problema muy antiguo conocido como *Problema de la Aguja de Buffon*. Imagine que una aguja de longitud 1, la unidad se deja caer sobre una hoja de papel rayado por líneas paralelas con una unidad de diferencia. ¿Cuál es la probabilidad de que la aguja se cruce con una de las líneas? Para hacer el problema preciso, asumimos que el centro de la aguja entre las líneas es un punto aleatorio.

Supóngase además que la orientación angular de la aguja es otra variable aleatoria. Por último, suponemos que nuestras variables aleatorias provienen de una distribución uniforme. La figura muestra la geometría de la situación

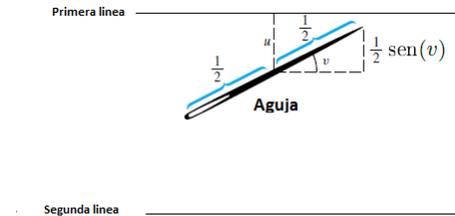


Figura 17: Problema de la Aguja de Buffon

Sea u la distancia del centro de la aguja a la línea más cercana de las dos y sea v el ángulo horizontal. Allí u y v son dos variables aleatorias. La aguja se cruza con una de las líneas si y sólo si $u \leq \frac{1}{2} \text{sen}(v)$. Realizamos el experimento muchas veces, digamos 5000 veces.

Debido a la simetría del problema, seleccionamos u con una distribución aleatoria uniforme en el intervalo $(0, \frac{1}{2})$ y v con una distribución aleatoria uniforme en el intervalo $(0, \frac{\pi}{2})$ y determinamos el número de veces por $2u \leq \text{sen}(v)$. Dejamos que $w = 2u$ y prueba que $w \leq \text{sen}(v)$, donde w es una variable aleatoria de $(0, 1)$.

En este programa, las respuestas intermedias se imprimen de modo que su progresión puede ser observada. Además las respuestas teóricas $t = \frac{2}{\pi} \approx 0.63662$.

Programa Aguja_de_Buffon

```

Entero i, m
Real pron, v, w
Arreglo Real (r_ij)_{1:n \times 1:2}

```

```

Entero  $n = 5000$ ,  $iprt = 1000$ 
 $m = 0$ 
Llamando Aleatorio ( $(r_{ij})$ )
for  $i = 1$  hasta  $n$  hacer
     $w = r_{i1}$ 
     $v = \left(\frac{\pi}{2}\right) r_{i2}$ 
    if  $w \leq \text{sen}(v)$  entonces  $m = m + 1$ 
    if  $\text{mod}(i, iprt) = 0$  entonces
         $prob = \frac{\text{real}(m)}{\text{real}(i)}$ 
        output  $i$ ,  $prob$ ,  $\left(\frac{2}{\pi}\right)$ 
    end if
end for
end programa Aguja_de_Buffon

```

Blindaje de neutrones

Nuestro último ejemplo se refiere blindaje de neutrones. Tomamos un modelo simple de neutrones que penetran en una pared de plomo.

Se supone que cada neutrón entra en la pared de plomo en un ángulo recto a la pared y recorre una distancia unidad. Entonces choca con un átomo de plomo y rebota en una dirección aleatoria.

Una vez más, recorre una distancia unidad antes de chocar con otro átomo de plomo, se rebota en una dirección aleatoria y así sucesivamente.

Supongamos que después de ocho colisiones, toda la energía del neutrón se gasta. Supóngase también que la pared de plomo es de 5 unidades de espesor en la dirección x y para todos los propósitos prácticos infinitamente de espesor en la dirección y .

La pregunta es: ¿Qué porcentaje de los neutrones se puede esperar a emerger desde el otro lado de la pared de plomo? Veamos la Figura 18

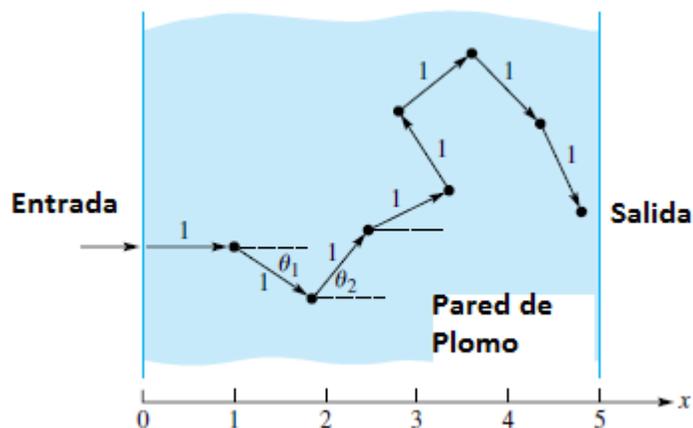


Figura 18: Experimento del blindaje de neutrones

Sea x la distancia medida desde la superficie inicial en la que el neutrón entra. De trigonometría, recordamos que en un triángulo rectángulo con hipotenusa 1, un lado es $\cos(\theta)$. También, tomemos en cuenta que $\cos(\theta) \leq 1$ cuando $\frac{\pi}{2} \leq \theta \leq \pi$.

La primera colisión ocurre en un punto donde $x = 1$. La segunda se produce en un punto donde $x = 1 + \cos(\theta_1)$. La tercera colisión ocurre en un punto donde $x = 1 + \cos(\theta_1) + \cos(\theta_2)$, y así sucesivamente. Si $x \geq 5$, el neutrón ha salido.

Si $x < 5$ para los ocho colisiones, el muro ha protegido el área a partir de un neutrón particular. Para una simulación de Monte Carlo, podemos usar ángulos aleatorios θ_i en el intervalo $(0, \pi)$ porque da simetría.

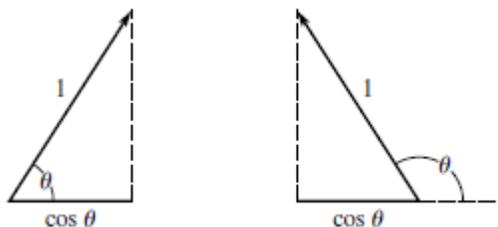


Figura 19: triángulos rectángulos con hipotenusa 1

8. Conclusiones

1. La importancia del método Monte Carlo se basa en la existencia de problemas de difícil solución por métodos analíticos o numéricos, pero que dependen de factores aleatorios y se asocian a un modelo determinístico.
2. El Método Monte Carlo permite obtener soluciones de problemas matemáticos o físicos por medio de pruebas aleatorias repetidas.
3. El Método de Alias Walker es muy eficiente y esta basado en la generación de variables aleatorias con probabilidades con distribuciones discretas.
4. El Método Monte Carlo nos ayuda a encontrar una evaluación aproximada de una integral definida, normalmente de integrales múltiples, donde se evalúa la función en un conjunto de puntos aleatorios que permiten calcular el valor de la integral.
5. En las aplicaciones del Método Monte Carlo generalmente utiliza una distribución uniforme $U(0, 1)$ en la generación de las variables aleatorias.
6. El Método de Aceptación - Rechazo es utilizado para muestras con distribución continua o discreta, que necesita una distribución conocida para generar números aleatorios que de otra forma sería difícil e imposibles de generar.

9. Bibliografía

Referencias

- [1] Cheney, E. (2007). Numerical Mathematics and Computing (6 edition). Belmont, CA: Brooks Cole.
- [2] Shonkwiler, Ronald W., and Franklin Mendivil. *Explorations in Monte Carlo Methods*. Dordrecht: Springer, 2009. Print.
- [3] James E.Gentle, Computational Statistic -Springer - 2009
- [4] DeGroot, Morris H. Probability and Statistic. 4th ed. USA: Pearson, 2012. Print.
- [5] Simulation and the Monte Carlo method. Reuven Y Rubinstein - John Wiley & Sons - 1981
- [6] Handbook of Monte Carlo methods Dirk P.Kroese - Thomas Taimre - Zdravko I.Botev - Wiley - 2011