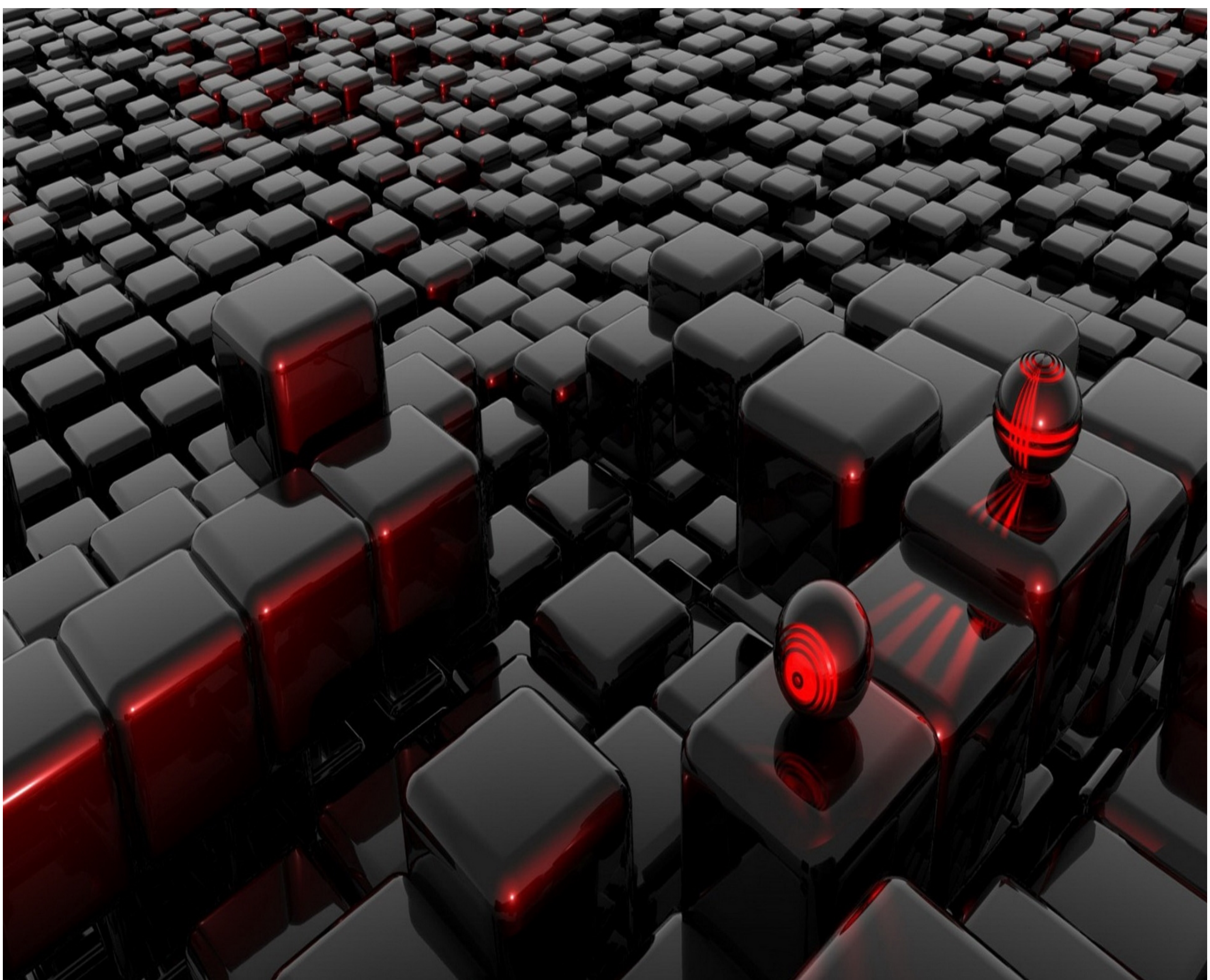
Probabilidad, procesos aleatorios e inferencia.

CIC-IPN



He ofendido a Dios y a la humanidad
porque mi trabajo no tuvo la calidad
que debía haber tenido.
Leonardo da Vinci (1452-1519)

...

índice

índice de tablas

Índice de figuras

Capítulo 1

Introducción

Tanto el científico como el ciudadano están sometidos a una permanente repetición de informaciones, observaciones, que se pueden presentar bajo modos muy dispares: los resultados de un experimento, percepciones sensoriales, datos númericos sobre las más variadas magnitudes, los comportamientos de los demás humanos, hechos históricos, opiniones personales, etc.

Es, y ha sido siempre una constante el que las personas hayan querido interpretar sus observaciones, experiencias, para aprender el mundo que les rodea. Y siempre han dirigido parte de sus esfuerzos a encontrar técnicas, métodos, procedimientos, o como quiera que se les denomine, para poder interpretar sus observaciones.

El conjunto de técnicas estadísticas se configura hoy día como el cuerpo más potente, objetivo y capaz en esta misión: la de interpretar lo que se pone a nuestros ojos permitiendo, incluso, hacer previsiones, a veces ciertamente fiables. (Pliego and Pérez, 2006)

Capítulo 2

Definición de probabilidad y técnicas de conteo

Dios no juega a los dados

Albert Einstein.
1879–1955

2.1. Modelos determinísticos y probabilísticos

Definición 2.1 (Modelo determinístico) *Un modelo determinístico podemos definirlo como aquel modelo matemático que cuando se le repite bajo las mismas condiciones, produce el mismo resultado. (Rincón, 2013)*

Como ejemplos de modelos determinísticos podemos citar:

1. Dejar caer un objeto desde cierta altura y registrar el tiempo en el que cae al suelo.
2. Las leyes de Kepler que nos indican el comportamiento de los planetas.
3. Supongamos que deseamos medir la fuerza aplicada a un objeto, observando la aceleración del cuerpo, de peso conocido.

4. Al extraer una canica roja de una caja que contiene sólo canicas rojas.
5. Colocar una batería en un circuito simple, y medir la intensidad de corriente.
6. Cuando mezclamos proporciones adecuadas de hidrógeno y oxígeno, sabemos que tendremos por resultado agua.
7. Sea la distancia recorrida de un objeto que sigue la función $s = -16t^2$.
8. Estirar un resorte de metal, un metro, repetidas veces y anotar su elasticidad.
9. Registrar la cantidad de energia producida por un grama de hidrógeno.
10. Medir la velocidad de la luz a distintas velocidades.

Definición 2.2 (Modelo probabilístico) *Un modelo probabilístico podemos definirlo como aquel modelo matemático que cuando se le repite bajo las mismas condiciones no produce el mismo resultado y tampoco es predecible, (Rincón, 2013).*

Como ejemplos de modelos probabilísticos podemos citar:

1. Lanzar un dado y observar el número de *soles* que cae.
2. Dejar una molécula de oxígeno en libre movimiento en un contenedor durante cierto tiempo y después registrar su posición.
3. Se fabrica una bombilla. Luego se prueba la duración poniéndola en un porta-lámparas y se registra el tiempo transcurrido hasta que se quema.
4. Contar el número de artículos defectuosos cada hora.
5. Registrar la temperatura cada 24 horas.
6. Registrar el precio diario de una acción.
7. Medir la resistencia a la tensión de una barra de acero
8. Tomar una muestra de 20 personas al azar para conocer su opinión, favorable o no, respecto de un candidato.
9. Examinar la posición y velocidad de una partícula subatómica.
10. Registrar el movimiento en conjunto de moleculas de un gas bajo cierta preción.

2.2. Espacios muestrales

Definición 2.3 (Especio muestral) *El espacio muestral de un experimento, denotado por Ω, es el conjunto de todos los posibles resultados de dicho experimento.*

Ejemplos:

1. Se examinan tres fusibles en secuencia y se anota el resultado de cada examen, entonces un resultado del experimento es cualquier secuencia de letras N y D de longitud 3, donde N representa no derectuoso y D defectuoso, por lo que el espacio muestral es

$$\Omega_1 = \{NNN, NND, NDN, NDD, DNN, DND, DDN, DDD\}$$

2. Se lanza un dado y se observa el número que aparece en la cara superior. En este caso el espacio muestral es

$$\Omega_2 = \{1, 2, 3, 4, 5, 6\}$$

3. Se lanza una moneda y se cuenta el número total de *soles*. En este caso el espacio muestral es

$$\Omega_3 = \{0, 1, 2, 3, 4\}$$

4. Medir la resistencia a la tensión de una barra de acero. Sea S la resistencia, entonces el espacio muestral es

$$\Omega_4 = \Omega = \{S|S \geq 0\}$$

5. Un experimento consiste en lanzar una moneda y después lanzarla una segunda vez si sale *sol*. Si sale *águila* en el primer lanzamiento, entonces se lanza un dado una vez. Para este caso el espacio muestral es

$$\Omega = \{SS, SA, A1, A2, A3, A4, A5, A6\}$$

Definición 2.4 (Evento) *Un evento A es un subconjunto del espacio muestral Ω_5.*

Se dice que es simple si consiste de un elemnto y compuesto si consiste de más de un elemento.

Son ejemplos de eventos:

1. En un lanzamiento de un dado, un número par ocurre

$$A_1 = \{2, 4, 6\}$$

2. Que ocurran dos *soles* al lanzar una moneda 4 veces.

$$A_2 = \{2\}$$

3. Que al probar una bombilla esta se queme en menos de 3 horas.

$$A_3 = \{t|t < 3\}$$

2.3. Definición de probabilidad

Definición 2.5 (Probabilidad clásica) *Sea ξ un experimento aleatorio con espacio muestral Ω de cardinalidad finita, y tal que todos los elementos de Ω tienen las mismas posibilidades de ocurrir. La probabilidad clásica (a priori o de Laplace) de un evento A se define en Eq (2.1), (Rincón, 2013):*

$$P(A) = \frac{\#A}{\#\Omega} \tag{2.1}$$

Ejemplo

Supongamos que se lanzan dos dados distinguibles y se desea calcular la probabilidad de que ambos muestren el mismo número.

Solución: Sea el espacio el experimento ξ = "Lanzar dos dados", con espacio muestral Ω = {(x,y): x,y=1,2,3,4,5,6} y el evento A = {(1,1), (2,2),...,(6,6) }. Entonces, dado que la cardinalidad de $\Omega = 36$ y la de A=6,

$$P(A) = \frac{\#A}{\#\Omega} = \frac{6}{36} = 0.1667$$

Definición 2.6 (Probabilidad geométrica) *Similar al caso de la probabilidad clásica, la probabilidad geométrica de A se define en Eq (2.2), (Rincón, 2013):*

$$P(A) = \frac{\text{\emph{área de} } A}{\text{\emph{área de} } \Omega} \tag{2.2}$$

Ejemplo

Se escoge un número al azar dentro del intervalo $(-1{,}1)$. ¿Cuál es la probabilidad de que la ecuación cuadrática

$$ax^2 + x + 1 = 0$$

tenga dos raíces reales?.

Solución: Sea $\Omega = (-1{,}1)$, $\text{long}(\Omega) = 2$, entonces, el evento A ocurre si el discriminante $b^2 - 4ac \geq 0$, es decir cuando $a \leq 1/4$. Por tanto

$$P(A) = \frac{Long(A)}{Long(\Omega)} = \frac{5/4}{2} = 0.625$$

Definición 2.7 (Probabilidad frecuentista) *Sea ξ un experimento aleatorio con espacio muestral Ω y sea A un evento. Se realizan n repeticiones del experimento aleatorio y se define $n(A)$ como el número de veces que ocurre el evento A en los n ensayos del experimento. La probabilidad frecuentista (o a posteriori) se define en Eq (2.3), (Rincón, 2013):*

$$P(A) = \frac{n(A)}{n} \tag{2.3}$$

Ejemplo

Supongamos que se desea obtener la probabilidad de obtener un número par al lanzar un dado 10 veces.

Solución:
Supongamos que se observaron los siguientes resultados de los 10 lanzamientos $\Omega = \{5,3,2,3,6,2,5,3,6,1\}$. Entonces

$$P(A) = \frac{n(A)}{n} = \frac{4}{10} = 0.4$$

Definición 2.8 (Probabilidad axiomática) *Esta definición considera la probabilidad como una función P, donde P representa una colección de eventos en el intervalo $[0,1]$, que satisface:*

1. $P(A) \geq 0$
2. $P(\Omega) = 1$
3. $P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k)$ *cuando* $A_1, \ldots, A_k$ *son ajenos.*

A estos tres axiomas se les conoce como axiomas de Kolmogorov, (Rincón, 2013).

Definición 2.9 (Probabilidad subjetiva) *Se refiere a la probabilidad de ocurrencia de un suceso basado en la experiencia previa, la opinión personal o la intuición del individuo. En este caso después de estudiar la información disponible, se asigna un valor de probabilidad a los sucesos basado en el grado de creencia de que el suceso pueda ocurrir, (Rincón, 2013).*

2.4. Técnicas de conteo

Definición 2.10 (Principio de multiplicación) *Si un procedimiento A puede efectuarse de n formas distintas y un segundo procedimiento B puede efectuarse de m formas distintas. Entonces el total de formas en que puede efectuarse el primer procedimiento seguido del segundo viene dado por la ecuación (2.4)*

$$n(AxB) = mn \tag{2.4}$$

Ejemplo Supongamos que un restaurante ofrece tres tipos de tamaños de pizza y 10 tipos de ingredientes, entonces, se desea saber cuantas pizzas diferentes se puede ofrecer?.

Solución:
El número de pizzas que el restaurante puede ofrecer es de

$$mn = 3 * 10 = 30$$

Por lo que, considerando el tamaño y los ingredientes, se pueden ofrecer 30 diferentes tipos de pizza.

Definición 2.11 (Ordenaciones con repeticiones) *Las ordenaciones de n objetos en muestras de tamaño k , considerando repeticiones, puede obtenerse mediante la ecuación (2.5).*

$$O_n^k = n^k \tag{2.5}$$

Ejemplo:
Supongamos que se tiene una urna con 10 bolas numeradas del 1 al 10, y que se extraen 3 bolas, se registran los números y se vuelven a depositar en la urna, para después extraer nuevamente otras 3 bolas y así sucesivamente, entonces, se desea saber cuantas muestras pueden obtenerse?

Solución:
Sea
$n = 10$ (número de bolas), y
$k = 3$ (número de extracciones)

entonces, el número de ordenaciones con repeticiones que se pueden obtener es

$$O_n^k = n^k = 10^3 = 1000$$

Por lo que el número de ordenaciones posibles es de $1,000$

Definición 2.12 (Ordenaciones sin repeticiones) *El número de ordenaciones sin repetición (permutaciónes) de n objetos en muestras de tamaño k viene dada por la ecuación (2.6).*

$$P_n^k = P_{n,k} = \frac{n!}{(n-k)!} \tag{2.6}$$

Ejemplo:
En una carrera participan 5 corredores, entonces, se desea saber de cuantas formas pueden registrarse los tres primeros lugares.

Solución:
Sea
n= 5 (total de corredores), y
k= 3 (tres primeros lugares).

se tiene que

$$P_n^k = P_{n,k} = \frac{5!}{(5-2)!} = 5 * 4 = 20$$

Por lo tanto, hay 20 formas de registrar los tres primeros lugares.

Definición 2.13 (Combinaciones) *Una combinación de n objetos tomados de k en k puede obtenerse mediante ecuación (2.7).*

$$C_n^k = C_{n,k} = \frac{n!}{k!(n-k)!} \tag{2.7}$$

Ejemplo:
Supóngase que se desea formar un equipo de 3 personas, sabiendo que se cuenta con 5 personas. Entonces, cuántos equipos distintos pueden formarse?.

Solución:
Sea
$n = 5$ (número de personas con que se cuenta), y

$k = 3$ (número de elementos del equipo).

Entonces, se tiene que

$$C_{n,k} = \frac{n!}{k!(n-k)!} = \frac{5!}{3!(5-3)!} = 10$$

Por tanto, se pueden formar 10 equipos distintos.

2.5. Algunos resultados

2.5.1. Teorema del Binomio

El teorema del binomio puede definirse como

$$(x+y)^n = \sum_{k=0}^{n} x^{n-k} y^k \tag{2.8}$$

Propiedades

1. $\binom{n}{k} = \binom{n}{n-k}$
   Demostración:
   Sea

$$\begin{aligned}\binom{n}{k} &= \frac{n!}{k!(n-k)!} \\ &= \frac{n!}{(n-k)!(n-(n-k))!} \\ &= \binom{n}{k}\end{aligned}$$

2. $\binom{n+1}{k} = \frac{n+1}{n-(k-1)}\binom{n}{k}$

Demostración:

$$\begin{aligned}
\binom{n+1}{k} &= \frac{(n+1)!}{k!(n+1-k)!} \\
&= \frac{n!(n+1)!}{k!(n+1-k)!} \\
&= \frac{n!}{k!}\frac{n+1}{(n-k+1)!} \\
&= \frac{n!}{k!}\frac{n+1}{(n-k)!(n-k+1)} \\
&= \frac{n!}{(n-k)!}\frac{n+1}{n-(k+1)} \\
&= \binom{n}{k}\frac{n+1}{n-(k-1)}
\end{aligned}$$

3. $\binom{n+1}{k+1} = \frac{n+1}{k+1}\binom{n}{k}$

Demostración:

$$\begin{aligned}
\binom{n+1}{k+1} &= \frac{(n+1)!}{(k+1)!(n-k)!} \\
&= \frac{n!(n+1)!}{(k+1)!(n-k)!} \\
&= \frac{n!(n+1)}{k!(k+1)(n-k)!} \\
&= \frac{n!}{k!(n-k)!}\frac{n+1}{k+1} \\
&= \binom{n}{k}\frac{n+1}{k+1|}
\end{aligned}$$

4. $\binom{n}{k+1} = \frac{n-k}{k+1}\binom{n}{k}$

Demostración:

$$\begin{aligned}
\binom{n}{k+1} &= \frac{n!}{(k+1)!(n-k-1)!} \\
&= \frac{n!}{k!(k+1)(n-k-1)!} \\
&= \frac{n!}{k!}\frac{1}{(k+1)(n-k-1)!} \\
&= \frac{n!}{k!}\frac{n-k}{(k+1)(n-k)!} \\
&= \frac{n!}{k!(n-k)!}\frac{n-k}{k+1} \\
&= \frac{n}{k}\frac{n-k}{k+1}
\end{aligned}$$

5. $\binom{n}{k} + \binom{n}{k+1} = \binom{n+1}{k+1}$

   Demostración:

$$\begin{aligned}
\binom{n}{k} + \binom{n}{k+1} &= \frac{n!}{k!(n-k)!} + \frac{n!}{(k+1)!(n-k-1)!} \\
&= \frac{n!(k+1) + n!(n-k)}{(n-k)!(k+1)!} \\
&= \frac{n!(k+1+n-k)}{(n-k)!(k+1)!} \\
&= \frac{n!(n+1)}{(n-k)!(k+1)!} \\
&= \frac{(n+1)!}{(n-k)!(k+1)!} \\
&= \binom{n+1}{k+1}
\end{aligned}$$

Por ejemplo, supongamos que deseamos obtener $(a+b)^{10}$. **Solución**

$$\begin{aligned}
(x+y)^{10} &= a^{10} + 10a^9b + 45a^8b^2 + 120a^7b^3 + 210a^6b^4 + \\
&= 252a^5b^5 + 210a^4b^6 + 120a^3b^7 + 45a^2b^8 + \\
&= 10ab^9 + b^{10}
\end{aligned}$$

2.5.2. Aproximación factorial

Para números no enteros

La función factorial ha sido extendida para valores no enteros empleando la función gamma definida por la integral compleja:

$$\Gamma(\alpha+1) = \int_0^{infty} e^{-x} x^{\alpha} dx \text{ para } \alpha > -1$$

donde α puede ser complejo, (Ross, 1985).

Mediante integración por partes se puede ver que

$$\Gamma(\alpha+1) \quad = \quad \alpha\Gamma(\alpha) \tag{2.9}$$

la cual es la fórmula de recursión de la función gamma, donde $\Gamma(1) = 1$, y

$$\Gamma(n+1) = n! \text{ para } n = 1, 2, 3, ... \tag{2.10}$$

Una aproximación asintomática es conocida como la fórmula de Stirling.

Aproximación de Stirling

El cálculo de límites de sucesiones es un tema interesante en donde existen bastantes métodos para realizar ese cálculo dependiendo de la estructura de la propia sucesión. En este caso emplearemos una apreximación para obtener n! mediante la fórmula de Stirling, (Ross, 1985).

La fórmula de Stirling nos dice lo siguiente:

$$\lim_{x\to\infty} \frac{n!}{n^n e^{-n}\sqrt{2\pi n}} = 1 \tag{2.11}$$

O también puede expresarse como

$$n! \approx n^n e^{-n}\sqrt{2\pi n} \tag{2.12}$$

Cuando tratamos con valores grandes de n, esta fórmula es de gran ayuda.

2.6. Ejercicios resueltos

2.6.1. Permutaciones

1. En una carrera de campo traviesa participan 15 personas. De cuántas formas pueden ser otorgados el primero, segundo y tercer premios, si el empate no se considera?

   Solución:

   $$P_3^{15} = \frac{15!}{(15-3)!} = 2,730$$

2. De cuántas maneras se puede acomodar una reunión de 7 personas, en una fila de siete sillas?

   Solución:

   $$7! = 5,040$$

3. Cuántas señales diferentes, cada una de 6 banderas colgadas en una línea vertical, pueden formarse con 4 banderas rojas idénticas y 2 azules idénticas?

   Solución:

   $$P_4^6 = \frac{6!}{(6-4)!} = 15$$

4. Cuántas permutaciones distintas pueden formarse con todas la letras de la palabra campana?

Solución:

$$P_4^7 = \frac{7!}{(7-4)!} = 840$$

5. Hallar n si $P(n,2)$=72.

Solución:

$$P_2^n = \frac{n!}{(n-2)!} = n(n-1) = n^2 - n = 72$$

Puesto que n debe ser positivo, la única respuesta es 9.

6. Se van a elegir a un presidente y a un tesorero de un club estudiantil compuesto por 50 personas. Cuántas opciones diferentes de funcionarios son posibles si no hay restricciones?

Solución:

$$P_2^{50} = \frac{50!}{(50-2)!} = 2450$$

7. En un año se otorgarán seis premios en un grupo de 25 estudiantes de posgrado del departamento de estadística. Si cada estudiante puede recibir un premio como máximo, cuántas selecciones posibles habrá?

Solución:

$$P_3^{25} = \frac{25!}{(25-3)!} = 13,800$$

8. De una urna que contiene los nombres de 30 empleados de una pequeña empresa se van a elegir aleatoriamente, sin reemplazo, los nombres de 3. El individuo cuyo nombre sale primero recibe $3,000$ pesos, el siguiente en salir su nombre recibe $2,000$ pesos y el tercero recibe $2,500$ pesos. ¿Cuántos puntos muestrales se asocian con este experimento?
Solución:

$$P_3^{30} = \frac{30!}{(30-3)!} = 24,360$$

9. Una operación de ensamblaje en línea de una fábrica incluye cuatro etapas que pueden ejecutarse en cualquier orden de sucesión. Si el fabricante desea comparar el tiempo que toma ensamblar en cada una de las sucesiones. Cuántas sucesiones distintas incluye el experimento?
   Solución:
$$P_4^4 = \frac{4!}{(4-4)!} = 24$$

10. Cuántas cantidades de tres cifras se pueden formar con los tres dígitos 0, 1, 2, 3 y 4 si no se permite la repetición?
    Solución:
$$P_3^5 = \frac{5!}{(5-3)!} = 60$$

2.6.2. Combinaciones

1. Un estudiante debe seleccionar 8 preguntas para contestar un examen que consta de 12 preguntas. De cuántas maneras puede hacer esta selección?
   Solución:
$$C_8^{12} = \frac{12!}{8!(12-8)!} = 495$$

2. De un grupo de 9 estudiantes se van a elegir 5 para un viaje de prácticas. De cuántas formas distintas es posible esto?
   Solución:
$$C_5^9 = \frac{9!}{5!(9-5)!} = 126$$

3. Arturo tiene en total 4 monedas de las siguientes denominaciones \$10, \$5, \$2 y \$1. Cuántas cantidades diferentes puede formar con estas monedas?
   Solución:
$$\binom{4}{4} + \binom{4}{3} + \binom{4}{2} + \binom{4}{1} = 15$$

4. El departamento de calidad tiene 3 vacantes para el puesto de coordinador, pre-

sentándose 10 candidatos. De cuántas maneras diferentes se pueden seleccionar a tres de estos para cubrir las vacantes?
Solución:

$$\binom{10}{3} = \frac{10!}{3!(10-3)!} = 120$$

5. De cuántas maneras se pueden escoger un comité, compuesto de 3 hombres y 2 mujeres de un grupo de 7 hombres y 5 mujeres?
Solución: De los 7 hombres se pueden escoger 3 de $\binom{7}{3}$ maneras, y de las cinco mujeres se pueden escoger 2 de $\binom{5}{2}$ maneras, por consiguiente:

$$\binom{7}{3}\binom{5}{2} = \frac{7!}{3!(7-3)!}\frac{5!}{2!(5-2)!} = 350$$

6. Un estudiante tiene que contestar 8 de 10 preguntas en un examen. Cuántas maneras, si las tres primeras preguntas son obligatorias?
Solución: Si contesta las 3 primeras preguntas, entonces puede escoger las otras cinco de las 7 restantes, es decir

$$\binom{7}{5} = \frac{7!}{5!(7-5)!} = 21$$

7. De cuántas maneras puede un profesor escoger uno o más estudiantes de seis elegibles?
Solución:

$$\binom{6}{1} + \binom{6}{2} + \binom{6}{3} + \binom{6}{4} + \binom{6}{5} + \binom{6}{6} = 6 + 15 + 20 + 15 + 6 + 1 = 63$$

8. En una clase hay 12 estudiantes.De cuántas maneras los 12 estudiantes pueden presentar 3 pruebas diferentes si a cada prueba le corresponden cuatro estudiantes?
Solución: Hay $\binom{12}{4}$ maneras de escoger 4 estudiantes que tomen la primera prueba, a continuación hay $\binom{8}{4}$ maneras de escoger 4 estudiantes que tomen la segunda prueba. Es decir, que por todo hay

$$\binom{12}{4} + \binom{8}{4} = 495 * 70 = 34,650$$

9. Una persona tiene en el bolsillo cinco billetes de diferentes denominaciones. De cuántas formas distintas puede sacar dos billetes simultáneamente?
   Solución:
   $$\binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$$

10. El director técnico de la selección de fútbol, decidió en casa llegar a tiros penales en un juego de selecciones, elegir a los cinco primeros tiradores en forma aleatoria, si solo pueden tirar los once jugadores que concluyen el partido. Cuántas quintetas, son posibles si el orden no importa?
    Solución:
    $$\binom{11}{5} = \frac{11!}{5!(11-5)!} = 462$$

¿Saben los animales contar?

Hormigas cuentakilómetros

Las hormigas son insectos sociales que se desarrollaron en el período Cretácico medio.La hormiga del desierto del Sahara viajan inmensas distancias sobre terreno arenoso en búsqueda de alimento, a menudo completamente desprovistas de puntos de referencia. Dichas hormigas no solo utilizan la luz del sol para orientarse sino que poseen un contador en sus patas, lo cual les permite calcular las distancias y regresar de nueva cuenta a su nido, eso fue lo que Investigadores Alemanes y Suizos descubrieron al experimentar con el tamaño de sus patas.

Primates cuentan

Hace alrededor de 60 mil millones de años los primates pequeños evolucionaron en muchas áreas del mundo. H. Kalmus suguiere, como muchos otros, que los animales tienen algún sentido de número. Por otro lado, Michael Beran, un científico de investigación en la Universidad Estatal de Georgia en Atlanta, Georgia, entreno a chimpaces para aprender a contar, con el resultado que un chimpancé aprendió a contar del 1 al 7 mientras que otro logró contar hasta 6, tres años más tarde experimento con los chimpancés y con el resultado de que lo hicieron, pero con el doble de fallas.

Cicadas generan números primos

Las cigarras son insectos que evolucionaron hace más de 1.8 millones de años. Lo sorprendente de estos animales es que conciben un comportamiento extraño, su aparición coincide con periodos de años que son por lo general los números primos 13 y 17. Investigadores han especulado que los ciclos de vida-números primos se produjeron de forma tal que aumentaran las posibilidades de subsistencia. Por su parte Mario Markus del Instituto de Fisiología Modecular del Instituto Max Planck, junto con sus colaboradores descubrieron que el número primo es el resultados de un modelo matemático que incluye presa y depredador.

Quipu
Los antiguos incas usaban los quipus, hechos de cuerda y nodos, para almacenar números. En los quipus, los tipos de nodos, posiciones, niveles de cordón y el color representaban los números asignados a los objetos. Se especula que los quipus pueden haber contenido más información como planes de construcción, patrones de baile e incluso historia de la civilización Inca. Hoy en día existen sistemas informáticos cuyos administradores de archivos son llamados quipus.

Para mayor referencia puede consultarse Pickover (2009).

2.7. Número de caminos

Se puede usar el principio del triángulo de Pascal para encontrar cuantos posibles caminos existen de un punto a otro.

2.7.1. Cuadrado

Por ejemplo para un cuadrado de $3x3$ tenemos la siguiente representación

$$\begin{array}{cccc} 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 \\ 1 & 3 & 6 & 10 \\ 1 & 4 & 10 & 20 \end{array}$$

También podemos ver el problema como una combinación, es decir, podemos ir 3 bloques hacia la derecha y 3 hacia la izquierda, o sea DDDIII, bo es difícil ver que se trata de una combinación. El número de formas lo calculamos mediante

$$\frac{6!}{3!(6-3)!} = 20$$

Es decir que hay 20 formas para una una malla con 3 bloques.

2.7.2. Cubo

Para el caso más general de un cubo, de $3x3x3$

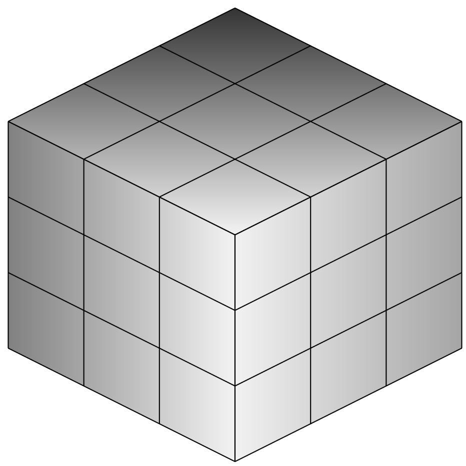


Figura 2.1: Formas de llegar en un cubo

podemos movernos en tres sentidos, 3 al este, 3 derecho y 3 abajo, suponiendo se está en la esquina superior izquierda, es decir EEEDDDBBBB, nuevamente, el problema se reduce a la siguiente expresión

$$\frac{9!}{3!3!3!} = 1,680$$

Es decir, existen $1,680$ formas de llegar.

2.8. Histograma

Gráfica que muestra la frecuencia de los datos, en la que el eje horizontal representa unidades discretas, ciertos rangos, o intervalos, en tanto que el eje vertical representa la frecuencia. Frecuentemente, se dibujan barras rectangulares con sus áreas proporcionales a las frecuencias dentro de los rangos o de los intervalos, para mayor referencia sobre los métodos descritos en las siguientes líneas consultar Rizzo (2007).

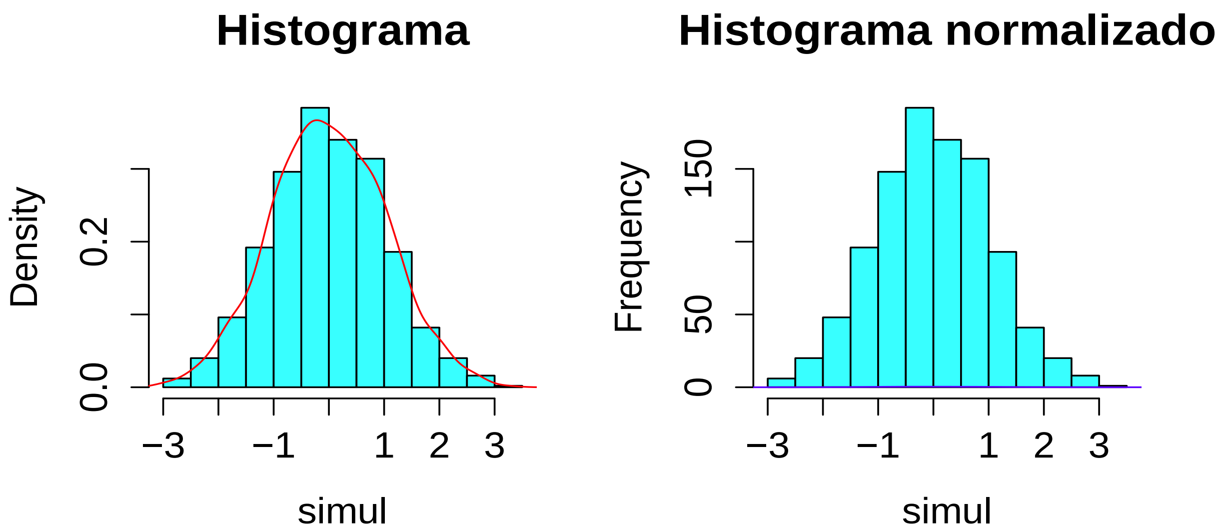


Figura 2.2: Histograma

Determinación del tamaño de clase

El histograma representa una aproximación de la densidad de los datos. Debido a que los datos en general, están contaminados por el ruido, resulta de crucial importancia el determinar el ancho de clase del histograma. Un ancho de intervalo estrecho puede suavizar los datos, presentando también tanto detalle, mientras más amplio el intervalo puede suavizar de más los datos, oscureciendo características importantes. Varias reglas se aplican comúnmente para la elección óptima de ancho de clase. Estas normas se discuten a continuación. La elección del parámetro de suavizado es un problema difícil que sigue atrayendo mucha atención en la investigación.

Regla de Sturges

Herbert Sturges (1926) considera un histograma idealizado con k clases, donde la i-ésima clase es el coeficiente binomial $\binom{k-1}{i}$ para $i = 0, 1, ..., k-1$. A medida que k incrementa, este histogram de frecuencias se aproxima a la curva de la densidad normal. El total del tamaño de muestra es

$$n = \sum_{i=0}^{k-1} \binom{k-1}{i} = (1+1)^{k-1} = 2^{k-1} \tag{2.13}$$

por la expansión binomial. Así que el número de clases a elegir viene dado por

$$k = 1 + log_2 n \tag{2.14}$$

Por ejemplo para n=100 aplicando la ecuación anterior se obtiene un valor de $k = 6$ y para $n = 100$ se obtiene un valor de $k = 8$.

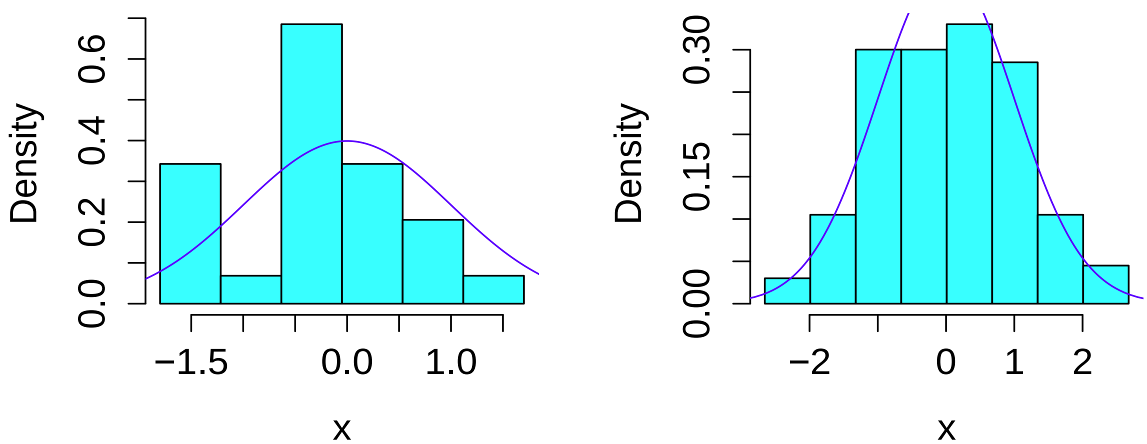


Figura 2.3: Regla de Sturges para n=25 (izq) y n=100(der)

Regla de Scott

Para seleccionar un óptimo parámetro de suavizamiento para estimación de la densidad, es necesario establecer un criterio para comparar el suavizamiento de los parámetros. Una opción es utilizar el cuadrado medio del error, cuyo valor asintótico es

$$AMISE = \frac{9}{16} \int f'(x)^2 n^{-2/3} \tag{2.15}$$

dado que la densidad f es desconocida, el valor óptimo de h no puede ser calculado exactamente, así que calibrado hacia la distribución normal con varianza σ^2, el tamaño de clase es

$$\hat{h} = 3.49\hat{\sigma}n^{-1/3} \tag{2.16}$$

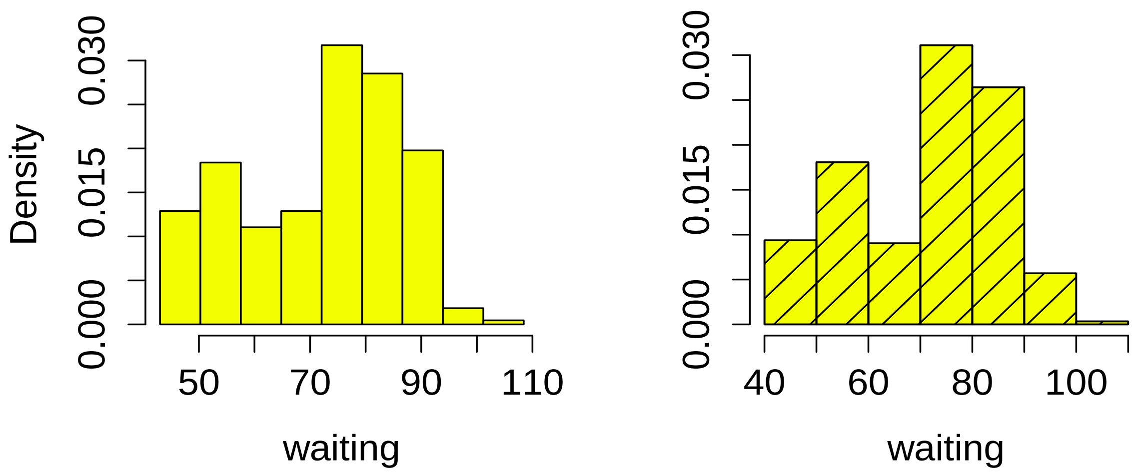


Figura 2.4: Histograma para n=299: Regla Scott (izq) y default n=100(der)

Regla de Freedman-Diaconis

La regla de FD es otro tipo de método para calcular el tamaño de la clase. La regla de FD se obtiene de la siguiente manera

$$\hat{h} = 2(RI)n^{-1/3} \tag{2.17}$$

n	Sturges	Scott	FD
10	5	2	3
20	6	3	5
30	6	4	4
50	7	5	7
100	8	7	9
200	9	9	11
500	10	14	20
1000	11	19	25
5000	14	40	52
10000	15	46	60

Tabla 2.1: Tabla comparativa

Cuál es la probabilidad de sacar un 6 con un dado?

Suponemos, lógicamente, un dado de 6 caras, con las caras numeradas del 1 al 6 y en el que todas los posibles resultados (sucesos elementales) son igualmente probables (equiprobables). Todos diríamos 1 de 6, es decir, un 16,67 % de probabilidad de sacar un 6, o cualquiera de los posibles resultados. Acabamos de aplicar la Regla de Laplace para el cálculo de probabilidades: simplemente dividiendo el número de casos favorables (1, porque solo hay un 6) entre el número de casos posibles (6, porque hay 6 posibles valores), obtenemos dicha probabilidad.

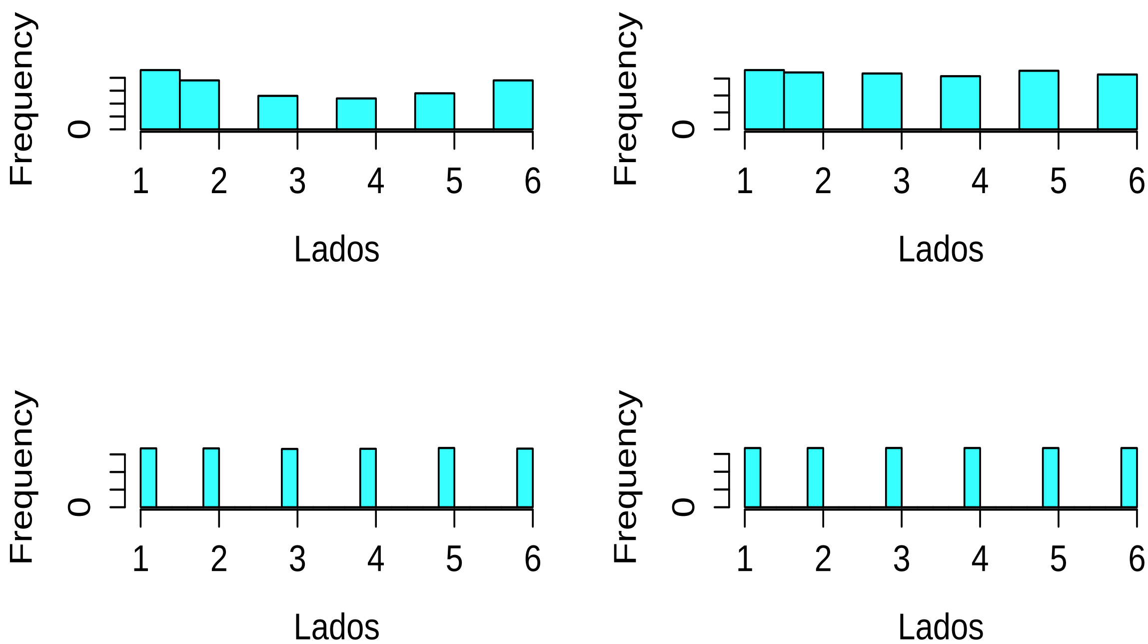


Figura 2.5: Histogramas para n=100,100,100000 y 10000000

Ahora bien, que pasa cuando repetimos el experimento no 10, 100, sino 10000, pues bien, aquí entra en juego la ley de los grandes números, la cual se enuncia del siguiente modo: "La frecuencia relativa de un suceso tiende a estabilizarse hacia una constante a medida que se repite el experimento".

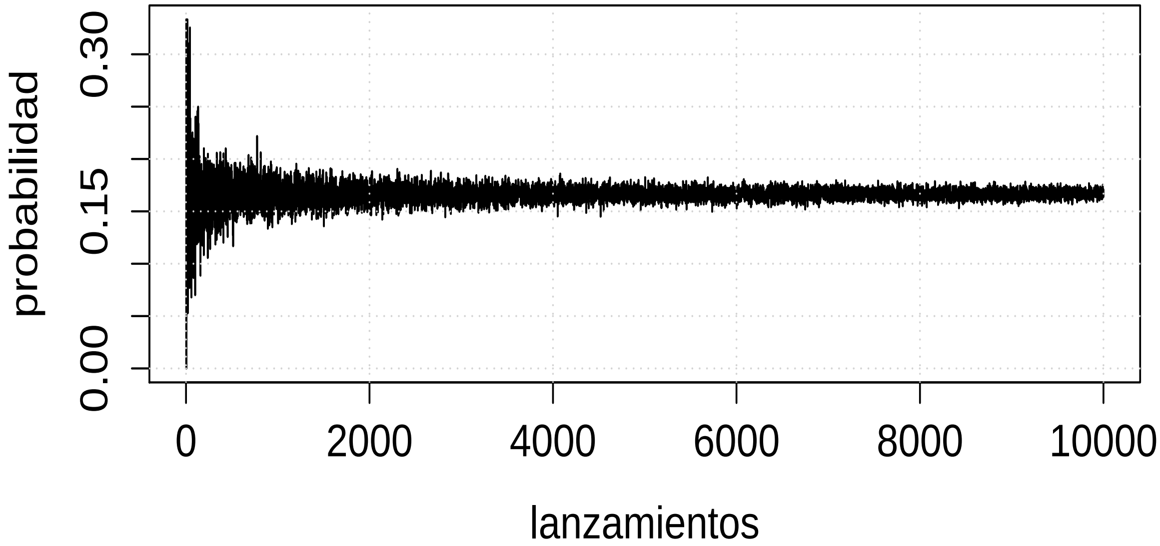


Figura 2.6: Gráfico para n=1000 lanzamientos

Código en R

```
#Experimento 1 con 100 valores
# n<-lanzamientos

n<-10000
prob<-rep(NA,n)

for(i in 1:n){
  eventos<-sample(1:6,i,replace=TRUE)
  esp.muestral<-length(eventos)
  cuenta<-length(eventos[eventos==6])
  prob[i]<-cuenta/esp.muestral
}
plot(prob,type="l",col="blue",xlab="lanzamientos",
     ylab="probabilidad",main="Experimento dado");grid()
```

Capítulo 3

Axiomas de probabilidad

En la sección anterior se mencionó los axiomas de probabilidad a saber:

Definición 3.1 (Axiomas de probabilidad)

$$Axiomas = \begin{cases} P(A) \geq 0 \\ P(\Omega) = 1 \\ P(\cup_{k=1}^{\infty} A_k) = \sum_{k=1}^{\infty} P(A_k) \textit{ cuando } A_1, ..., A_k \textit{ son ajenos.} \end{cases}$$

donde P(A) denota la probabilidad del evento y P(Ω) la probabilidad del espacio muestral. (Degroot, 1988)

Teorema 3.1 $P(A) = 1 - P(A^c)$
Para un evento $A \subset \Omega$*, se tiene que* $P(A) = 1 - P(A^c)$

Demostración:

Se tiene que $\Omega = A \cup A^c$ y $A \cap A = 0$ por la definición anterior, se tiene que

$$1 = P(A) + P(A^c) \tag{3.1}$$

Teorema 3.2 *La probabilidad del conjunto vacio es cero,* $P(0) = 0$.

Demostración:
De acuerdo al teorema anterior, si tomamos $A = 0$, entonces $P(A^c) = 1$, de ahí que

$$P(0) = 1 - P(A^c) = 1 - 1 = 0 \tag{3.2}$$

Teorema 3.3 *Si A_1, A_2 son subconjuntos de Ω tal que $A_1 \subset A_2$, entonces $P(A_1) \leq P(A_2)$.*

Demostración:
Sea $A_2 = A_1 \cup (A_1^c \cap A_2)$ y $A_1 \cap (A_1^c \cap A_2) = 0$, de ahí que

$$P(A_2) = P(A_1) + P(A_1^c \cap A_2)$$

como $P(A_1^c \cap A_2) \geq 0$, entonces $P(A_2) \leq P(A_1)$

Teorema 3.4 *Para $A \subset \Omega$, $0 \leq P(A) \leq 1$*

Demostración:
Sea $0 \subset A \subset \Omega$, por el Teorema 3, se tiene que

$$P(0) \leq P(A) \leq P(\Omega) \text{ } o \text{ } 0 \leq P(C) \leq 1$$

Teorema 3.5 *Si A_1 y A_2 son subconjuntos de Ω, entonces*

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2) \tag{3.3}$$

Demostración:
Los conjuntos $A_1 \cup A_2$ y A_2 pueden ser representados como una unión

$$A_1 \cap A_2 = A_1 \cup (A_1^c \cap A_2) \text{ y } A_2 = (A_1 \cap A_2)(A_1^c \cap A_2)$$

entonces

$$P(A_1 \cup A_2) = P(A_1) + P(A_1^c \cap A_2)$$

Y

$$P(A_2) = P(A_1 \cap A_2) + P(A_1^c \cap A_2)$$

si la segunda equación es resuelta para $A_1^c \cap A_2$, entonces

$$P(A_1 \cup A_2) = P(A_1) + P(A_2) - P(A_1 \cap A_2)$$

3.1. Probabilidad Condicional

Dados los eventos A Y B, estamos interesados en definir la probabilidad condicional de A dado que el evento B ha ocurrido.

Definición 3.2 (Probabilidad condicional) *Sea A y B dos eventos, de un espacio de probabilidad, la probabilidad condicional del evento A dado el evento B, denotado por $P[A|B]$, es definida como*

$$P[A|B] = \frac{P(A \cap B)}{P(B)} \text{ si } P(B) > 0 \tag{3.4}$$

y no está definida cuando $P(B) = 0$. (Degroot, 1988)

Se puede mostrar que también las probabilidades condicionales cumplen los tres axiomas, es decir:

1. $P[A|B] = \frac{P(A \cap B)}{P(B)} \geq 0$
2. $P[\Omega|B] = \frac{P(\Omega \cap B)}{P(B)} = 1$
3. Si $A_1, A_2, ...,$ es una secuencia de eventos mutuamente excluyentes, entonces: $\sum_{i=1}^{\infty} P[A_i|B]$

3.2. Ejercicios resueltos:

1. Muestre el valor de $P(A \cup B \cup C)$

   Solución:

$$\begin{aligned}
P(A \cup (B \cup C)) &= P(A) + P(B \cup C) - P(A \cap (B \cup C)) \\
&= P(A) + P(B) + P(C) - P(B \cap C) - P((A \cup B) \cap (A \cup C)) \\
&= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) \\
&\quad + P(A \cup B \cap A \cap C) \\
&= P(A) + P(B) + P(C) - P(B \cap C) - P(A \cap B) - P(A \cap C) \\
&\quad + P(A \cup B \cap C)
\end{aligned}$$

2. Muestre que $A = (A \cap B) \cup (A \cap B^{'})$

   Solución:
   Sea

$$\begin{aligned}
A &= A \cap (B \cap B^{'}) \text{propiedad distributiva} \\
&= A
\end{aligned}$$

3. Muestre que $(A \cap B)^{'} = A^{'} \cup B^{'}$

   Solución:

   Considere:

$$\begin{aligned}
x \in (A \cap B)^{'} \\
(x \in A)^{'}(x \in B)^{'} \\
(x \notin A)(x \notin B) \\
(x \in A^{'})(x \in B^{'}) \\
\text{por lo tanto} \\
x \in A^{'} \cup B^{'}
\end{aligned}$$

4. Muestre que $(A \cup B)^{'} = A^{'} \cap B^{'}$

   Solución:

   Considere:

$$\begin{aligned} x \in (A \cup B)^{'} \\ (x \in A)^{'}(x \in B)^{'} \\ (x \notin A)(x \notin B) \\ (x \in A^{'})(x \in B^{'}) \\ \text{por lo tanto} \\ x \in A^{'} \cap B^{'} \end{aligned}$$

5. Un cinco porciento de la población tiene una enfermedad particular. Una prueba es desarrollado para la enfermedad. La prueba da un falso positivo 3 % de las veces y 2 % de las veces falso negativo. Determine: (a) Cuál es la probabilidad que Joe (elegida aleatoriamente) de positivo en el test? y Supongamos que Joe dio positivo en el test, ¡cuál es la probabilidad de que Joe este enfermo?

   Solución:

   Sea A el evento que Joe tiene la enfermedad y T el evento que Joe dio positivo en el test. Entonces, tenemos los siguientes datos: $P(A) = 0.005$, de ahí que 1/2 % de la población tiene la enfermedad. Además sabemos que $P(T|A) = 0.98$. Sabemos que $P(T|A^c) = 0.03$

   a) Queremos calcular $P(T)$, entonces utilizando la regla de Bayes

$$\begin{aligned} P(T) &= P(T|A)P(A) + P(T|A^c)P(A^c) \\ &= (0.98)(0.005) + (0.03)(0.995) \\ &= 0.07885 \end{aligned}$$

b) Queremos calcular

$$\begin{aligned}
P(D|T) &= \frac{P(D\cap T)}{P(T)} \\
&= \frac{P(T|D)P(D)}{P(T|D)P(D)+P(T|D^c)P(D^c)} \\
&= \frac{(0.98)(0.05)}{(0.98)(0.05)} \\
&\approx 0.14
\end{aligned}$$

6. Dados los siguientes datos

- Uno porciento de las mujeres mayores de 50 % cáncer de mama.
- 90 % de las mujeres que tiene cáncer de mama dan positivo en el test de mamografía.
- 80 % de las mujeres dieron falso positivo

determine la probabilidad de que una mujer tenga cáncer si dio falso positivo en el examen de mamografía?.

Solución:
Sea A el evento de que de las mujeres mayores a 50 tengan cáncer. $P(B|A)$ denota el evento de las mujeres que tiene cáncer y dieron positivo en el examen de mamografía. Usando la regla de Bayes

$$\begin{aligned}
P(B|A) &= \frac{P(B\cap A)}{P(A)} \\
&= \frac{P(A|B)P(B)}{P(A|B)P(B)+P(A|B^c)P(B^c)} \\
&= \frac{(0.9)(0.01)}{(0.9)(0.001)+(0.08)(0.99))} \\
&= 0.10
\end{aligned}$$

7. Tres máquinas A, B y C, producen el 45 %, 30 % y 25 %, respectivamente del total de las piezas producidas en una gráfica. Los porcentajes de producción defectuosa de estas máquinas son del 3 %, 4 % y 5 %.

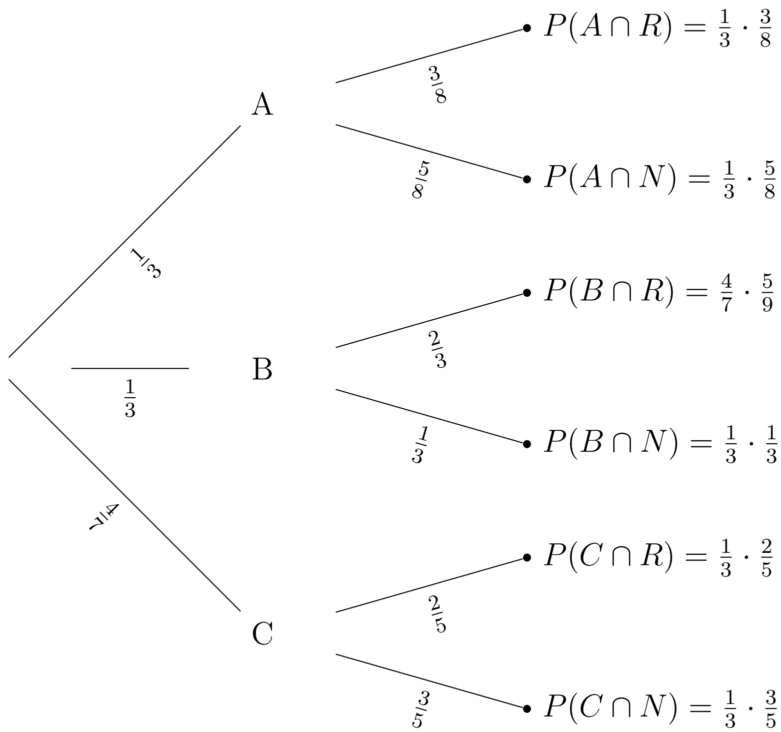
   a) Seleccionamos una pieza al azar,calcula la probabilidad de que sea defectuosa.

b) Tomamos al azar una pieza y resulta ser defectuosa, acalcula la probabilidad de haber sido producida por la máquina B.

c) ¿Qué máquina tiene la mayor probabilidad de haber producido la citada pieza defectuosa?

Solución:

Sea D la pieza defectuosa y N la pieza no es defectuosa



a) Para calcular la probabilidad de que la pieza elegida sea defectuosa, $P(D)$, por la propiedad de la probabilidad total

$$\begin{aligned} P(D) &= P(D|A)P(A) + P(D|A^c)P(A^c) \\ &= (0.45)(0.03) + (0.30)(0.04) + (0.25)(0.05) \\ &= 0.038 \end{aligned}$$

b) Debemos calcular $P(B|D)$, por el teorema de Bayes

$$\begin{aligned}
P(B|D) &= \frac{P(B\cap D)}{P(D)} \\
&= \frac{P(D|B)P(B)}{P(D|A)P(A)+P(D|B)P(B)+P(D|C)P(C)} \\
&= \frac{(0.3)(0.04)}{(0.45)(0.03)+(0.03)(0.04)+(0.25)(0.05))} \\
&= 0.316
\end{aligned}$$

c) Ahora calculamos $P(A|D)$ Y $P(C|D)$, comparándolas con el valor de $P(B|D)$ ya calculado.

$$\begin{aligned}
P(A|D) &= \frac{P(A\cap D)}{P(D)} \\
&= \frac{P(D|A)P(A)}{P(D|A)P(A)+P(D|B)P(B)+P(D|C)P(C)} \\
&= \frac{(0.45)(0.03)}{(0.45)(0.03)+(0.03)(0.04)+(0.25)(0.05))} \\
&= 0.355
\end{aligned}$$

$$\begin{aligned}
P(C|D) &= \frac{P(C\cap D)}{P(D)} \\
&= \frac{P(D|C)P(C)}{P(D|A)P(A)+P(D|B)P(B)+P(D|C)P(C)} \\
&= \frac{(0.25)(0.05)}{(0.45)(0.03)+(0.03)(0.04)+(0.25)(0.05))} \\
&= 0.329
\end{aligned}$$

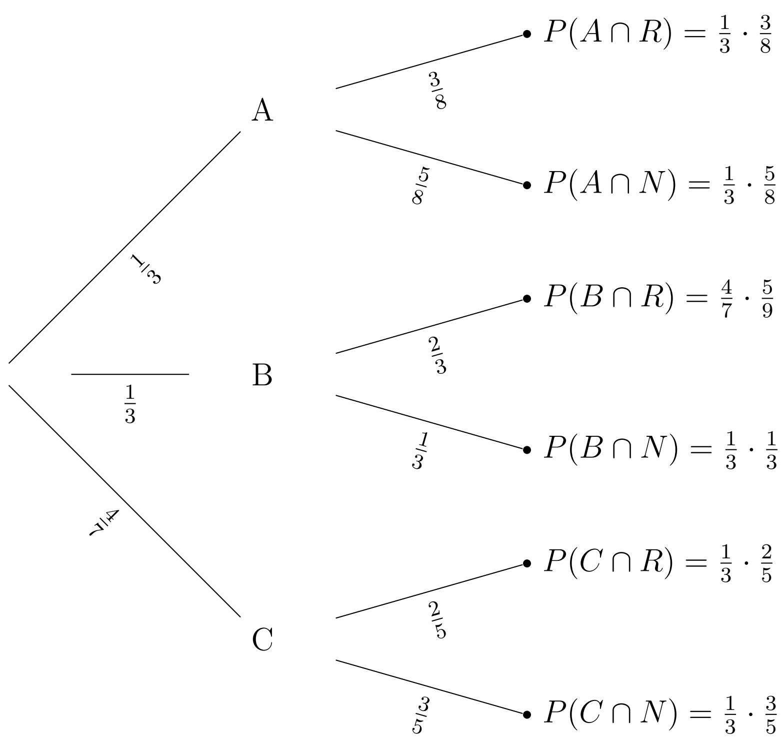
Entonces la máquina con mayor probabilidad de haber producido la pieza defectuosa es A.

8. Tenemos tres urnas: A con 3 bolas rojas y 5 negras, B con 2 bolas rojas y 1 negra y C con 2 bolas rojas y 3 negras. Escogemos una urna al azar y extraemos una bola. Sila bola ha sido roja, ¿cuál es la probabilidad de haber sido extraída de la urna A?

Solución:

Sea R el evento que denota sacar bola roja y N el evento de sacar bola negra. La probabilidad pedida es $P(A|R)$. Utilizando el teorema de Bayes,

$$\begin{aligned} P(A|R) &= \frac{P(A \cap R)}{P(R)} \\ &= \frac{P(R|A)P(A)}{P(R|A)P(A) + P(R|B)P(B) + P(R|C)P(C)} \\ &= \frac{(1/3)(3/8)}{(1/3)(3/8) + (1/3)(2/3) + (1/3)(2/5))} \\ &= 0.260 \end{aligned}$$



9. La probabilidad de que haya un accidente en una fábrica que dispone de alarma es 0.1. La probabilidad de que suene esta si se ha producido algún incidente es de 0.97 y la probabilidad de que suene si no se ha producido ningún incidente es de 0.02.

   En el supuesto de que haya funcionado la alarma, ¿cuál es la probabilidad de que no haya habido ningún incidente?

   Solución:

   Sea I el evento de producir incidentes y A el evento de sonar la alarma. Usando la regla de Bayes:

$$\begin{aligned} P(A^c|I) &= \frac{P(A^c \cap I)}{P(I)} \\ &= \frac{P(I|A^c)P(A^c)}{P(I|A)P(A)} \\ &= \frac{(0.9)(0.02)}{(0.9)(0.02) + (0.1)(0.97)} \\ &= 0.157 \end{aligned}$$

3.3. Ejemplo de los dos dados

Determínese la probabilidad de que al lanzar n veces dos dados se obtenga al menos un seis doble. ¿Cuántos lanzamientos habría que realizar para tener una probabilidad igual a 1/2 de obtener un seis doble?

Solución:

P[obtención de al menos un seis doble en n lanzamientos] $= 1 - \left(\frac{35}{36}\right)^n$

El número de lanzamientos que habría que realizar para tener una probabilidad igual a 1/2 de obtener al menos un seis doble se calcula de la ecuación

$$\begin{aligned} \frac{1}{2} &= 1 - \left(\frac{35}{36}\right)^n \\ \frac{1}{2} &= \left(\frac{35}{36}\right)^n \\ ln\left(\frac{1}{2}\right) &= nln\left(\frac{35}{36}\right) \\ -0.6931 &= -0.0282n \\ n &\approx 25 \end{aligned}$$

Simulación:

Para simular el problema se recurrió a fijar el número de lanzamientos posteriormente hallar la frecuencia de los pares ordenados $(1,1)$, $(1,2)$,...,$(6,6)$, y repetir el experimento muchas veces para poder calcular la probabilidad de que en cada par ordenado se haya obtenido por lo menos un doble seis.

Tabla 3.1: Sumulación de 25 lanzamientos y 100,000 repeticiones

dado1	dado2	exitos	probabilidad	dado1	dado2	exitos	probabilidad
1	1	50338	0.50338	4	1	50270	0.5027
1	2	50762	0.50762	4	2	50705	0.50705
1	3	50581	0.50581	4	3	50765	0.50765
1	4	50777	0.50777	4	4	50448	0.50448
1	5	50581	0.50581	4	5	50510	0.5051
1	6	50307	0.50307	4	6	50519	0.50519
2	1	50640	0.5064	5	1	50565	0.50565
2	2	50682	0.50682	5	2	50353	0.50353
2	3	50539	0.50539	5	3	50373	0.50373
2	4	50578	0.50578	5	4	50545	0.50545
2	5	50575	0.50575	5	5	50482	0.50482
2	6	50488	0.50488	5	6	50687	0.50687
3	1	50635	0.50635	6	1	50379	0.50379
3	2	50372	0.50372	6	2	50689	0.50689
3	3	50688	0.50688	6	3	50488	0.50488
3	4	50414	0.50414	6	4	50562	0.50562
3	5	50571	0.50571	6	5	50643	0.50643
3	6	50510	0.5051	6	6	50255	0.50255

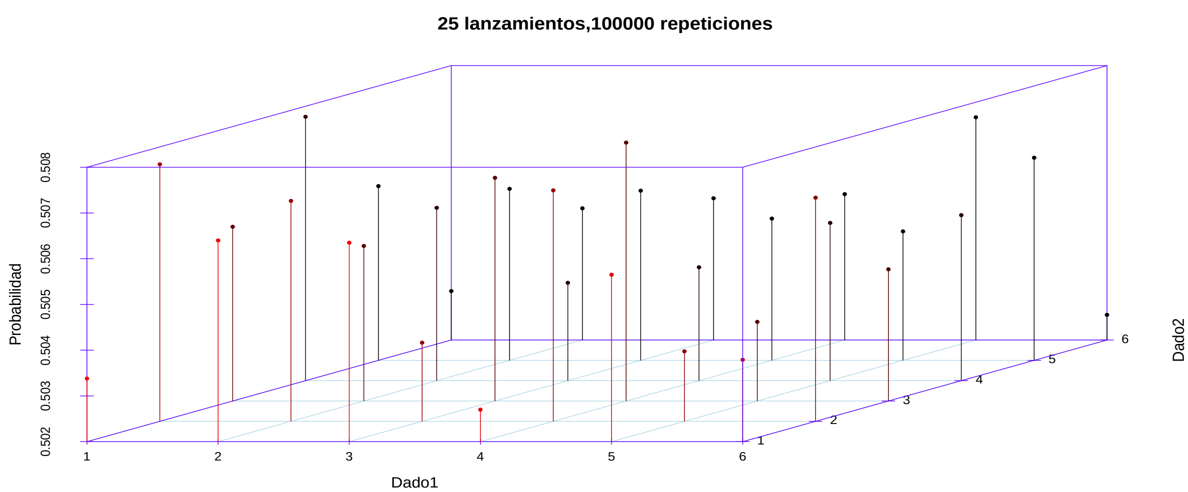


Figura 3.1: Gráfico de frecuencias

Si esto es correcto, entonces, al simular 50 lanzamientos se debería obtener la siguiente probabilidad

$$\begin{aligned} P &= 1 - \left(\frac{35}{36}\right)^{50} \\ &= 1 - 0.2445 \\ &= 0.7555 \end{aligned}$$

Resultado de la Simulación

Tabla 3.2: Simulación 50 lanzamientos, 10,000 repeticiones

dado1	dado2	exitos	probabilidad	dado1	dado2	exitos	probabilidad
1	1	7598	0.7598	4	1	7575	0.7575
1	2	7589	0.7589	4	2	7536	0.7536
1	3	7586	0.7586	4	3	7540	0.754
1	4	7526	0.7526	4	4	7604	0.7604
1	5	7557	0.7557	4	5	7567	0.7567
1	6	7590	0.759	4	6	7606	0.7606
2	1	7520	0.752	5	1	7569	0.7569
2	2	7591	0.7591	5	2	7561	0.7561
2	3	7589	0.7589	5	3	7572	0.7572
2	4	7573	0.7573	5	4	7655	0.7655
2	5	7482	0.7482	5	5	7547	0.7547
2	6	7588	0.7588	5	6	7480	0.748
3	1	7556	0.7556	6	1	7524	0.7524
3	2	7619	0.7619	6	2	7561	0.7561
3	3	7483	0.7483	6	3	7547	0.7547
3	4	7523	0.7523	6	4	7535	0.7535
3	5	7611	0.7611	6	5	7536	0.7536
3	6	7520	0.752	6	6	7529	0.7529

Podemos apreciar que la probabilidad de que caiga al menos un doble seis es, de acuerdo a los valores simulados, de 0.7529, lo cual concuerda con el resultado de la fórmula dada anteriormente.

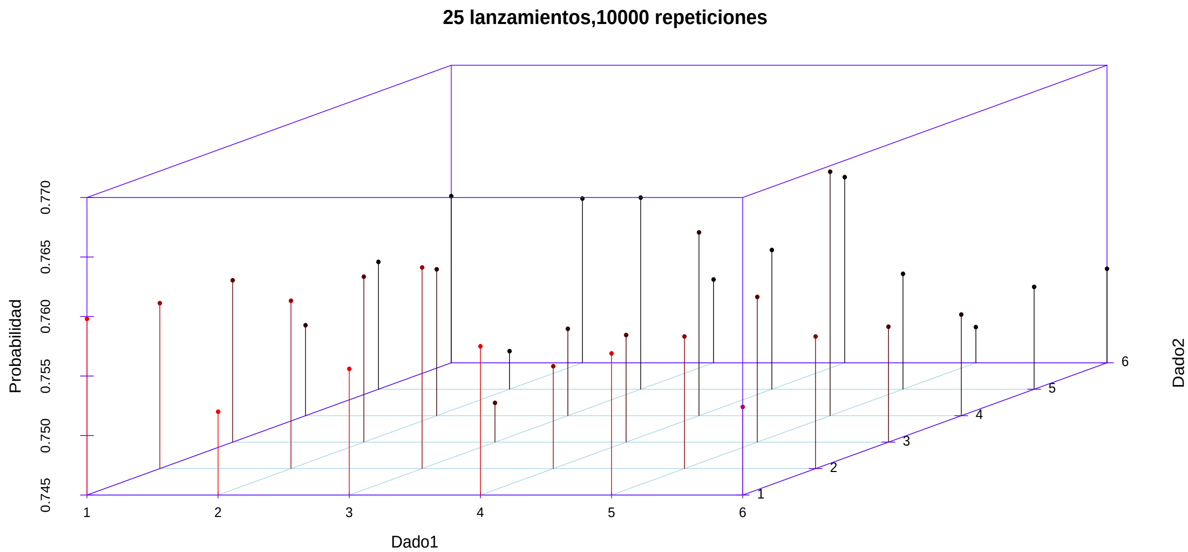


Figura 3.2: Gráfico de frecuencias

3.4. Teorema del límite central

Definición 3.3 (Teorema del límite central) *Supongamos que X_i son variables aleatorias independientes e idénticamente distribuidas con media $\mu = 0$ y varianza $\sigma^2 = 1$. Entonces*

$$\frac{X_1 + ... + X_n}{\sqrt{n}} \longrightarrow N(0, \sigma^2) \tag{3.5}$$

es decir que la función de distribución de le expresión anterior sigue una $N(0,1)$. (Pliego and Pérez, 2006)

El teorema nos dice que bajo condiciones general si sumamos variables aleatorias independientes y se normalizan, entonces en el límite (cuando sumamos muchas v.a.'s) obtendremos una distribución normal.

Ejemplo:

Para probar el teorema del límite central consideraremos la distribución lognormal$(3, 1)$, para $n = 10{,}100$ y 1000. En las figuras puede notarse que a medida que aumentamos el tamaño de muestra la distribución tiende a la normal. Así también, podemos hacerlo para el caso de más distribuciones y comprobar la veracidad del teorema del límite central.

Código en R

```
mul=3;
sigmal=1;
sigma2l=sigmal^2
mu=exp(mul+sigma2l/2);mu
sigma2=mu^2*(exp(sigma2l)-1);sigma2

for (j in c(10,100,1000)){
    res=c()
    for (i in 1:500){
      res[i]=mean(rlnorm(j,mul,sigmal))
    }
    par(mfrow=c(1,2))
    hist(res,prob=TRUE,main=paste("Lognormal(3,1) Xbar, n=",j),col=gray(.8))
    curve(dnorm(x,mu,sqrt(sigma2/j)),add=TRUE,col=j/10+1)
    qqnorm(res)
    qqline(res,col=j/10+1)
}
```

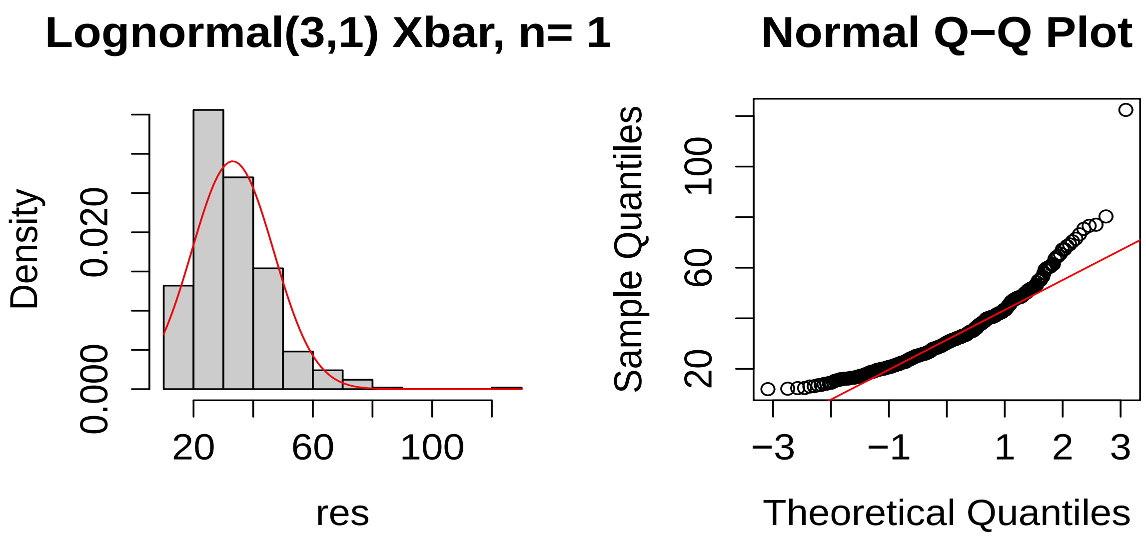


Figura 3.3: Gráfico de dispersión

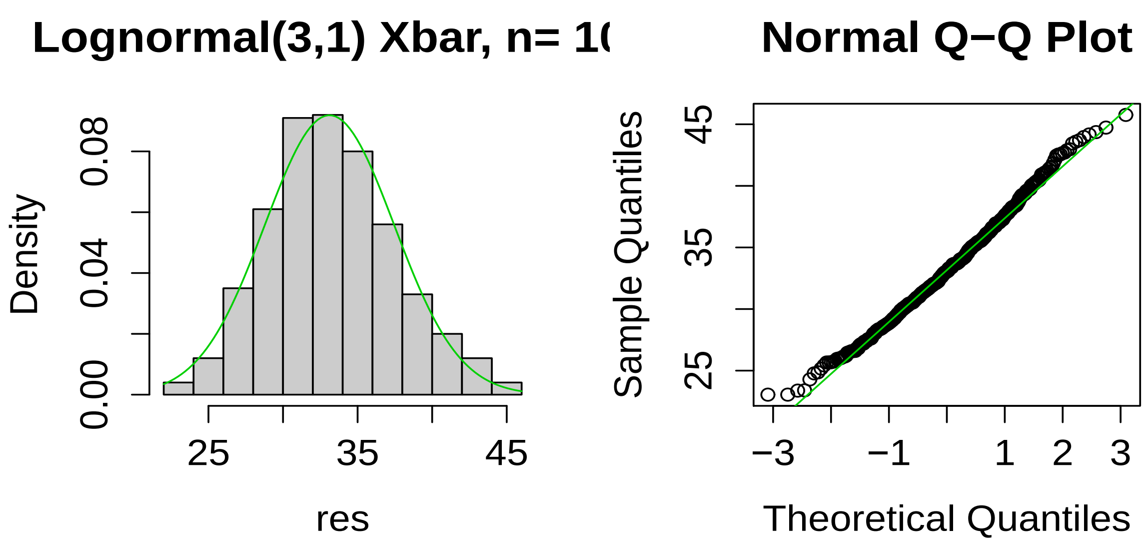


Figura 3.4: Gráfico de dispersión

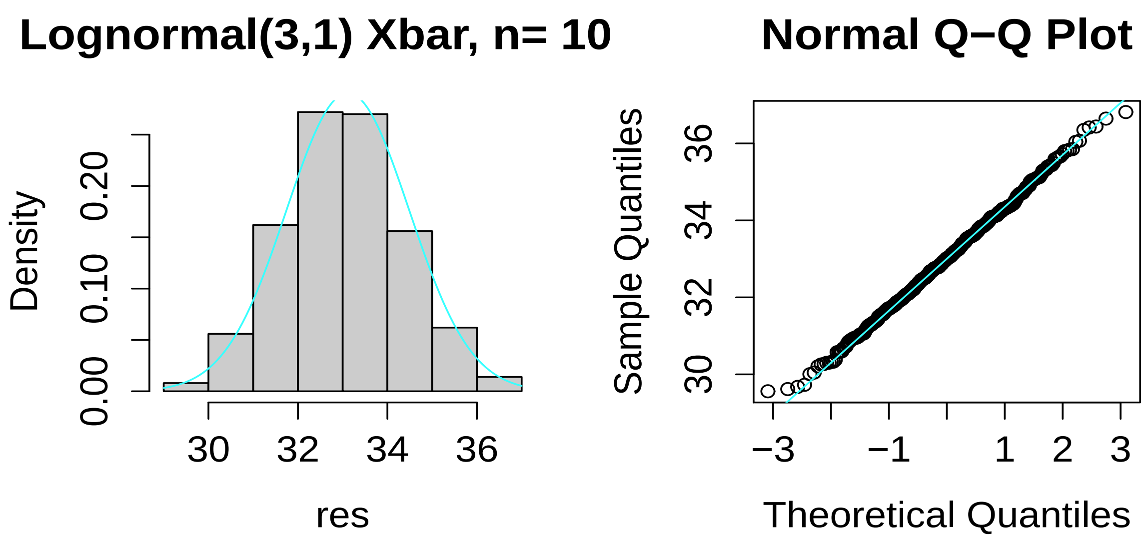


Figura 3.5: Gráfico de dispersión

Capítulo 4

Variable aleatoria

Definición 4.1 (Variable aleatoria) *Una variable aleatoria X en un espacio muestral S es una función que asocia un número real R a cada elemento del espacio muestral S. (Walpole et al., 1993)*

Generalmente se utiliza una letra mayúscula, digamos X, para denotar una variable aleatoria, y su correspondiente minúscula, x en este caso, para uno de sus valores.

Ejemplo

Se sacan dos bolas de manera sucesiva sin reemplazo de una urna que contiene cuatro bolas rojas y tres negras. Los posibles resultados y valores y de la variable aleatoria Y, donde Y es el número de bolas rojas, son:

Tabla 4.1: Ejemplo de variable aleatoria

Espacio muestral	y
RR	2
RB	1
BR	1
BB	0

4.1. Disttribuciones discretas de probabilidad

Una variable aleatoria discreta toma cada uno de sus valores con cierta probabilidad. Al lanzar una moneda tres veces, la variable X, que representa el número de caras, toma el valor 2 con probabilidad 3/8, pues 3 de los 8 puntos muestrales igualmente probables tienen como resultado dos caras y una cruz.

Con frecuencia es conveniente representar todas las probabilidades de una variable aleatoria X usando una fórmula, la cual necesariamente sería una función de los valores numéricos x que denotaremos con f(x), g(x), r(x), y así sucesivamente. Por lo tanto, escribimos f(x) = P(X = x); es decir, f(3) = P(X = 3). El conjunto de pares ordenados (x, f(x)) se llama función de probabilidad o distribución de probabilidad de la variable aleatoria discreta X.

Definición 4.2 (Función de probabilidad) *El conjunto de pares ordenados (x, f(x)) es una función de probabilidades, una función de masa de probabilidad o una distribución de probabilidad de la variable aleatoria discreta X si, para cada resultado posible x,*

1. $f(x) \geq 0$,
2. $\sum_x f(x) = 1$,
3. $P(X = x) = f(x)$

Ejemplo:

Sea X una variable aleatoria cuyos valores x son los números posibles de computadoras defectuosas que la escuela compra. Entonces, x puede ser cualquiera de los números 0, 1 y 2. Así,

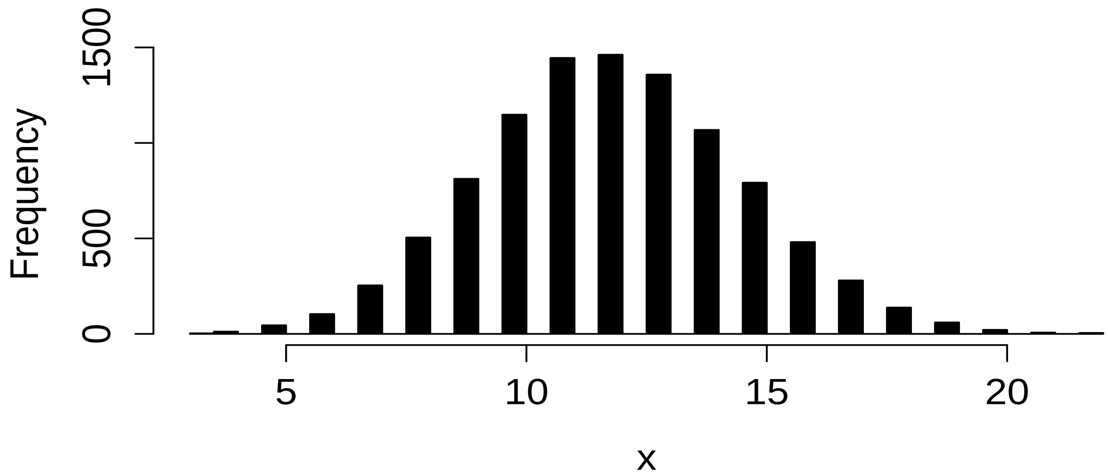


Figura 4.1: Función de densidad

$$
\begin{aligned}
f(0) &= P(X=0) = \frac{\binom{3}{0}\binom{5}{2}}{\binom{8}{2}} = \frac{10}{28} \\
f(1) &= P(X=1) = \frac{\binom{3}{1}\binom{5}{1}}{\binom{8}{2}} = \frac{15}{28} \\
f(2) &= P(X=2) = \frac{\binom{3}{2}\binom{5}{0}}{\binom{8}{2}} = \frac{3}{28}
\end{aligned}
$$

De manera que la densidad de probabilidad de X es

$$
f(x) = \begin{cases} \frac{10}{28} & \text{, si x es igual a 0} \\ \frac{15}{28} & \text{, si x es igual a 1} \\ \frac{3}{28} & \text{, si x es igual a 2} \end{cases}
$$

Definición 4.3 (Función de distribución) *La función de la distribución acumulada F(x) de una variable aleatoria discreta X con distribución de probabilidad f(x) es*

$$F(X) = P(X \leq x) = \sum_{t \leq x} f(t), \ \textit{para} \ -\infty < x < \infty$$

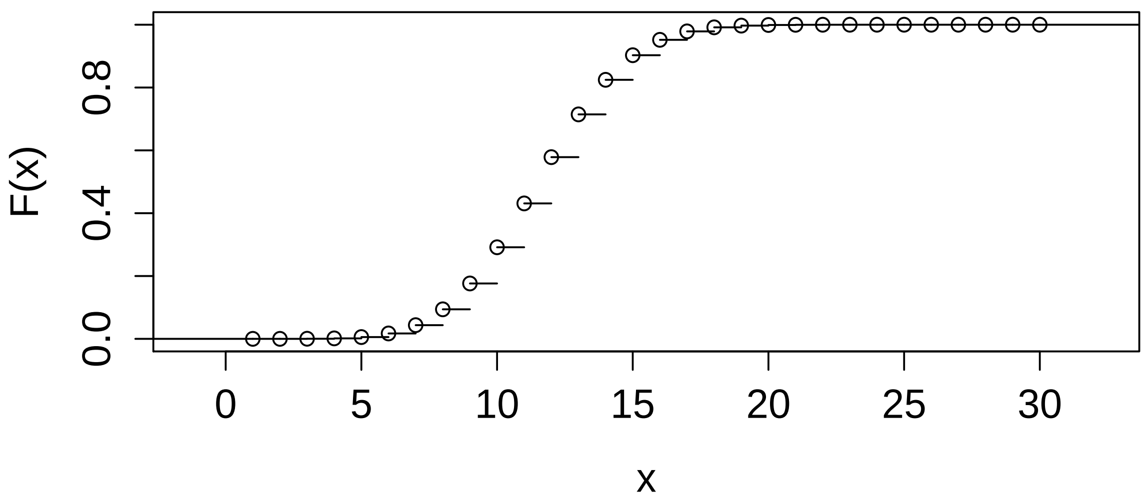


Figura 4.2: Función de distribución

Ejemplo:

Para los datos del ejemplo anterior, se tiene

$$\begin{aligned} F(0) &= P(X \leq 0) = \frac{10}{28}, \\ F(1) &= P(X \leq 1) = f(0) + f(1) = \frac{25}{28}, \\ F(2) &= P(X \leq 2) = f(0) + f(1) + f(2) = 1 \end{aligned}$$

De manera que la distribución de probabilidad de X es

$$F(x) = \begin{cases} 0, & \text{para } m < 0 \\ \frac{10}{28}, & \text{para } 0 \leq m < 1 \\ \frac{25}{28}, & \text{para } 1 \leq m < 2 \\ 1, & \text{para } 2 \leq m \end{cases}$$

4.2. Distribuciones continuas

Una variable aleatoria continua tiene exactamente una probabilidad cero de tomar exactamente cualquiera de sus valores. Aunque la distribución de probabilidad de una variable aleatoria continua no se puede representar de forma tabular, se puede establecer como una fórmula.

Definición 4.4 (Función de densidad) *La función $f(x)$ es una función de densidad de probabilidad para la variable aleatoria continua X, definida en el conjunto de número reales R, si*

1. $f(x) \geq 0$ *para* $x \in R$
2. $\int_{-\infty}^{\infty} f(x)dx = 1$
3. $P(a < X < b) = \int_a^b f(x)dx$

Ejemplo:

Suponga que en error en la temperatura de reacción, en grados centígrados, para un experimento de laboratorio controlado, es una variable aleatoria continua X, definida en el conjunto de los números reales R, si

$$\begin{aligned} f(x) &= \tfrac{x^2}{3}, \quad -1 < x < 2 \\ &= 0, \quad \text{de otro modo} \end{aligned}$$

1. verifique la condición 2.
2. Encuentre $P(0 < x < 2)$

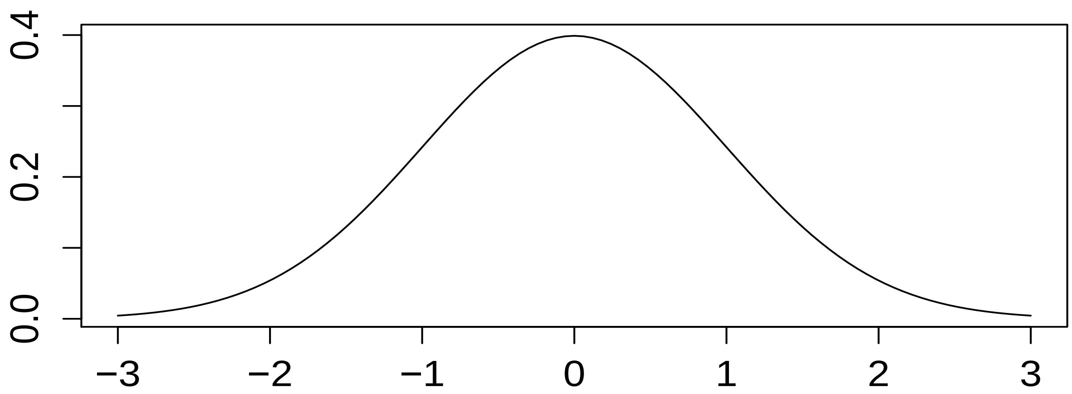


Figura 4.3: Función de densidad

Solución

1. $\int_{-\infty}^{\infty} f(x)dx = \int_{-1}^{2} \frac{x^2}{3}dx = \frac{8}{9} + \frac{1}{9} = 1$
2. $P(0 < x < 2) = \int_{0}^{1} \frac{x^2}{3}dx = \frac{1}{9}$

Definición 4.5 (Función de distribución) *La función de distribución o distribución acumulada F(x) de una variable aleatoria continua X con función de densidad f(x) es*

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

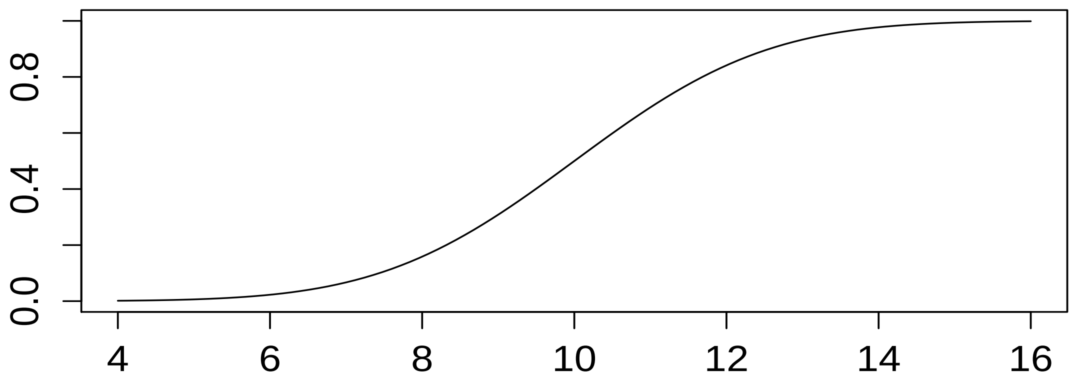


Figura 4.4: Función de distribución

4.2.1. Medidas de asimetría y curtosis

Las medidas de distribución nos permiten identificar la forma en que se separan o aglomeran los valores de acuerdo a su representación gráfica. Estas medidas describen la manera como los datos tienden a reunirse de acuerdo con la frecuencia con que se hallen dentro de la información. Su utilidad radica en la posibilidad de identificar las características de la distribución sin necesidad de generar el gráfico. Sus principales medidas son la Asimetría y la Curtosis.

Asimetría

Esta medida nos permite identificar si los datos se distribuyen de forma uniforme alrededor del punto central (Media aritmética). La asimetría presenta tres estados diferentes, cada uno de los cuales define de forma concisa como están distribuidos los datos respecto al eje de asimetría. Se dice que la asimetría es positiva cuando la mayoría de los datos se encuentran por encima del valor de la media aritmética, la curva es Simétrica cuando se distribuyen aproximadamente la misma cantidad de valores en ambos lados de la media y se conoce como asimetría negativa cuando la mayor cantidad de datos se aglomeran en los valores menores que la media.

La asimetría corresponde al tercer momento de una distribución X

$$Asimetría(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \tag{4.1}$$

Se dice que la distribución de X tiene asimetría positiva, negativa o nula dependiendo de si la curtosis es positiva, negativa o 0.

Curtosis

Esta medida determina el grado de concentración que presentan los valores en la región central de la distribución. Por medio del Coeficiente de Curtosis, podemos identificar si existe una gran concentración de valores (Leptocúrtica), una concentración normal (Mesocúrtica) ó una baja concentración (Platicúrtica).

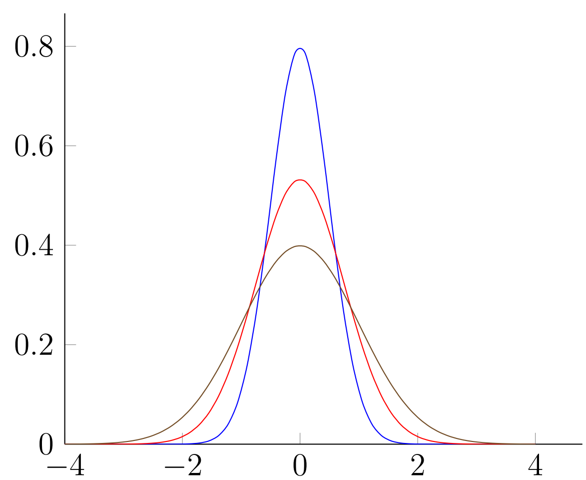


Figura 4.5: Distribución normal con curtosis diferentes

La curtosis de X corresponde al cuarto momento, es decir:

$$Curtosis(X) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] \tag{4.2}$$

Capítulo 5

Esperanza matemática

Se puede afirmar que el primer estudio sistemático del valor esperado se debe a Huygens (en su obra Libellus de Ratiotiniis in Ludo Aleae, de 1657), que calcula el valor justo de un juego a partir de una respuesta obvia en ciertas situaciones simétricas, y generalizando el valor esperado obtenido a cualquier situación. Comienza suponiendo que:

Los nombre de esperanza matemática y valor esperado tienen su origen en los juegos de azar y hacen referencia a la ganancia promedio esperada por un jugador cuando hace un gran número de apuestas.

Si la esperanza matemática es cero, $E(x) = 0$, el juego es equitativo, es decir, no existe ventaja ni para el jugador ni para la banca.

Definición 5.1 (Esperanza matemática) *Sea X una variable aleatoria con distribución de probabilidad $f(x)$. La media o valor esperado de X es*

$$\mu = E(x) = \sum_x x f(x) \text{ si } x \text{ es discreta} \tag{5.1}$$

$$\mu = E(x) = \int_{-\infty}^{\infty} x f(x) dx \text{ si } x \text{ es continua} \tag{5.2}$$

5.0.2. Propiedades

Tal como hemos mencionado anteriormente, con la definiciónn general el valor esperado sigue teniendo todas las propiedades enunciadas, acontinuación se enuncian las más importantes:

1. $E(X+Y) = E(X) + E(Y)$
2. Si X y Y son variables aleatorias independientes, entonces

$$E(XY) = E(X)E(Y)$$

Capítulo 6

Distribuciones Univariadas

6.1. Distribuciones discretas

6.1.1. Distribución uniforme discreta

La más simple de todas las distribuciones de probabilidad discreta es aquella donde la variable aleatoria toma cada uno de sus valores con una probabilidad idéntica. Tal distribución de probabilidad se denomina distribución uniforme discreta.

Definición 6.1 (Distribución uniforme discreta) *Si la variable aleatoria* X *toma los valores* x_1 , x_2 , ... , x_k , *con idénticas probabilidades, de la forma*

$$f(x) = \begin{cases} \frac{1}{k}, & para\ x = x_1, ..., x_k \\ 0, & d.o.m \end{cases}$$

entonces, se le llama distribución uniforme discreta. (*Walpole et al., 1993*)

6.1.2. Distribución Bernoulli

Definición 6.2 (Distribución Bernoulli) *Una variable aleatoria X tiene distribución Bernoulli con parámetro* $p(0 \leq X \leq 1)$, *con función de densidad*

$$f(x) = \begin{cases} p^x(1-p)^{1-x}, & para\ x = 0, 1 \\ 0, & d.o.m \end{cases}$$

Si X tiene distribución Bernoulli con parametro p, entonces X^2 y X son la misma variable aleatoria. (Mood et al., 1974)

Propiedades:

- Función generadora de momentos

$$M(t) \quad = \quad E\left[e^{tX}\right] = (1-p) + p^t \text{ para } -\infty < t < \infty$$

- Función característica

$$\phi(t) \quad = \quad E\left[e^{itX}\right] = (1-p) + p^{it} \text{ para } -\infty < t < \infty$$

- Esperanza

$$E(X) \quad = \quad p$$

- Varianza

$$Var(X) = p(1-p)$$

- Asimetría

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{1-2p}{\sqrt{p(1-p)}}$$

- Curtosis

$$E\left[(-u\sigma)^4\right] = \frac{3p^2-3p+1}{p(1-p)}$$

6.1.3. Distribución Binomial

Definición 6.3 (Distribución Binomial) *Una variable aleatoria X con parámetros n y p, cuya distribución de X es*

$$f(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{1-x}, & para\ x = 0, 1 \\ 0, & d.o.m \end{cases}$$

se le denomina diustribución binimial. (Degroot, 1988)

En esta distribución, n debe ser un entero positivo, y p un valor tal que 0 ≤p≤ 1.

La distribución binomial deriva su nombre del hecho de que los $n+1$ términos en la expansión binomial de $(q+p)$ n corresponden a los diversos valores de $b(x;n,p)$ para $x = 0, 1, 2, ..., n$. Es decir,

$$\begin{aligned} (p+q)^n &= \binom{n}{0} q^n + \binom{n}{1} pq^{n-1} + ... + \binom{n}{1} pq^{n-1} \\ &= b(0,n,p) + b(1,n,p) + b(2,n,p) + ... + b(n,n,p) \end{aligned}$$

como $p+q = 1$, vemos que

$$\sum_{0}^{n} b(x,n,p) = 1$$

Propiedades:

- Función generadora de momentos

$$M(t) = E\left[e^{tX}\right] = (1-p+pe^t)^n \text{ para } -\infty < t < \infty$$

- Función característica

$$\phi(t) = E\left[e^{itX}\right] = E\left[e^{itX}\right] = (1-p+pe^{it})^n \text{ para } -\infty < t < \infty$$

- Esperanza

$$E(X) = np$$

- Varianza

$$Var(X) = np(1-p)$$

- Asimetría

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{1-2p}{\sqrt{np(1-p)}}$$

- Curtosis

$$E\left[(-u\sigma)^4\right] = 3 + \frac{1-6p(1-p)}{np(1-p)}$$

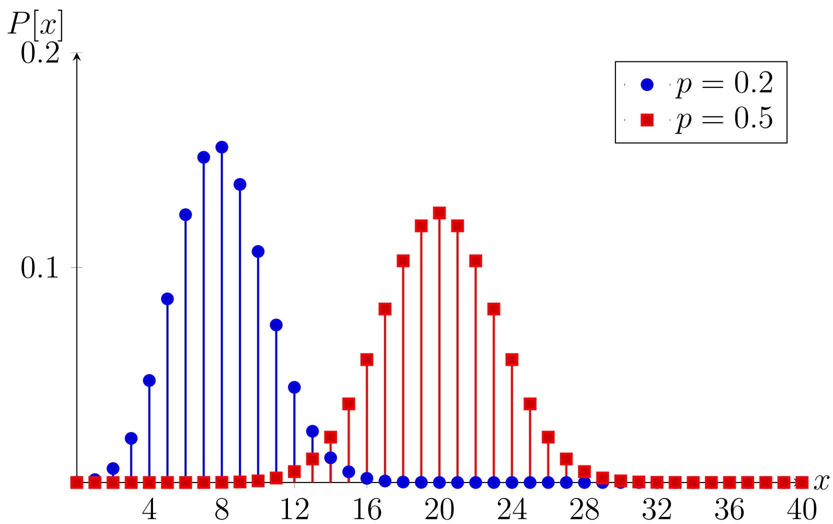


Figura 6.1: Función de densidad binomial

6.1.4. Distribución Binomial negativa

Definición 6.4 (Distribución Binomial negativa) *Supongamos que X representa el número de fallas hasta que ocurre el r-ésimo éxito, tiene la siguiente función de densidad:*

$$f(x) = \begin{cases} \binom{r+x-1}{x} p^r (1-p)^x, & para\ x = 0, 1, ..., n \\ 0, & d.o.m \end{cases}$$

la cual se dice tiene una distribución binomial negativa. (Degroot, 1988)

Propiedades:

- Función generadora de momentos

$$M(t) \quad = \quad E\left[e^{tX}\right] = \left[\frac{p}{1-(1-p)e^t}\right]^n \text{ para } |(1-p)e^t| < 1ot < -ln(1-p)$$

- Función característica

$$\phi(t) \quad = \quad E\left[e^{itX}\right] = \left[\frac{p}{1-(1-p)e^{it}}\right]^n \text{ para } |(1-p)e^{it}| < 1ot < -ln(1-p)$$

- Esperanza

$$E(X) \quad = \quad \frac{n(1-p)}{p}$$

- Varianza

$$Var(X) = \frac{n(1-p)}{p^2}$$

- Asimetría

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{2-p}{\sqrt{n(1-p)}}$$

- Curtosis

$$E\left[(-u\sigma)^4\right] = \frac{p^2-6p-3np+3n+6}{n(1-p)}$$

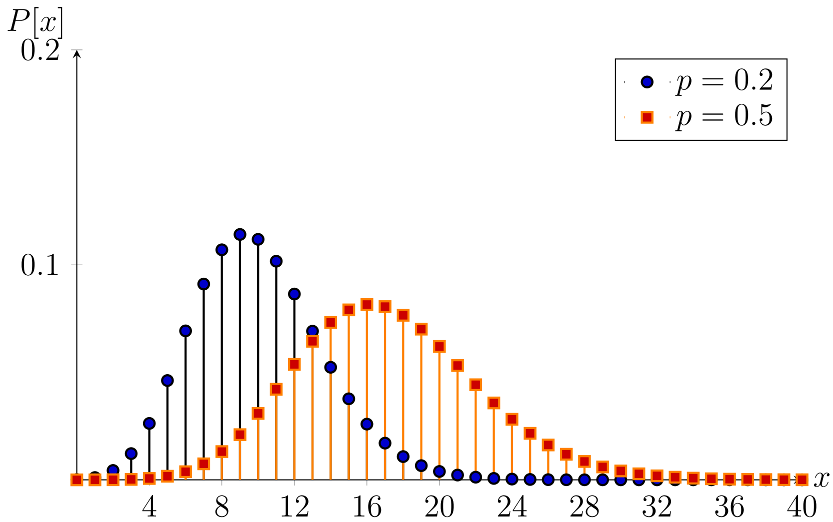


Figura 6.2: Función de densidad binomial negativa

6.1.5. Distribución geométrica

Definición 6.5 (Distribución geométrica) *Una variable aleatoria X sigue una distribución geométrica, con parámetro p, si*

$$f(x) = \begin{cases} p(1-p)^x, & para\ x = 0, 1, ..., n \\ 0, & d.o.m \end{cases}$$

Es el un caso especial de una variable aleatoria con distribución binomial negativa cuando $r = 1$. Esto sería, el número de fallas hasta obtener el primer éxito. (Degroot, 1988)

Propiedades:

- Función generadora de momentos

$$M(t) = E\left[e^{tX}\right] = \frac{p}{1-(1-p)e^t} \text{ para } t < -ln(1-p)$$

- Función característica

$$\phi(t) = E\left[e^{itX}\right] = \frac{p}{1-(1-p)e^{it}} \text{ para } t < -ln(1-p)$$

- Esperanza

$$E(X) = \frac{(1-p)}{p}$$

- Varianza

$$Var(X) = \frac{(1-p)}{p^2}$$

- Asimetría

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{2-p}{\sqrt{(1-p)}}$$

- Curtosis

$$E\left[(-u\sigma)^4\right] = 9 + \frac{p^2}{1-p}$$

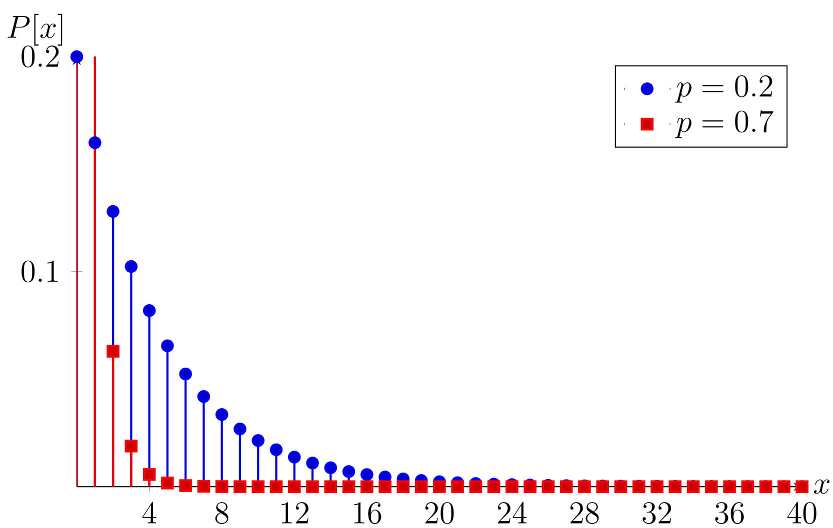


Figura 6.3: Función de densidad geométrica

6.1.6. Distribución hipergeométrica

Definición 6.6 (Distribución hipergeométrica) *Sea X una variable aleatoria que sigue la siguiente función de densidad dada por:*

$$f(x) = \begin{cases} \frac{\binom{A}{x}\binom{B}{n-x}}{\binom{A+B}{n}}, & para\ max(0, n-B) \leq x \leq min(n, A) \\ 0, & d.o.m \end{cases}$$

se dice que tiene una distribución hipergeométrica. (Degroot, 1988)

Propiedades

- Función generadora de momentos
- Esperanza

$$E(X) = \frac{n_2 n_1}{n_3}$$

- Varianza

$$Var(X) = \frac{n_2 n_3 (1-n_1)(n_3-n_2)}{n_3^2 (n_3-1)}$$

- Simetría

$$E\left(\frac{X-\mu}{\sigma}\right)^3 = \frac{(n_3-2n_1)(n_3-1)^{1/2}(n_3-n_2)}{[n_2 n_1 (n_3-n_1)(n_3-n_2)]^{1/2}(n_3-2)}$$

6.1.7. Distribución Poisson

Definición 6.7 (Distribución de Poisson) *Sea $\lambda > 0$. Una variable aleatoria X con media λ si X sigue la siguiente función de densidad*

$$f(x) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & para\ x = 0, 1, ... \\ 0, & d.o.m \end{cases}$$

entonces, se dice que tiene distribución de Poisson. (Degroot, 1988)

Propiedades:

- Función generadora de momentos

$$M(t) = E\left[e^{tX}\right] = e^{\lambda(e^t-1)}$$

- Función característica

$$\phi(t) = E\left[e^{itX}\right] = e^{\lambda(e^{it}-1)}$$

- Esperanza

$$E(X) = \lambda$$

- Varianza

$$Var(X) = \lambda$$

- Asimetría

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \lambda^{-1/2}$$

- Curtosis

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = 3 + \lambda^{-1}$$

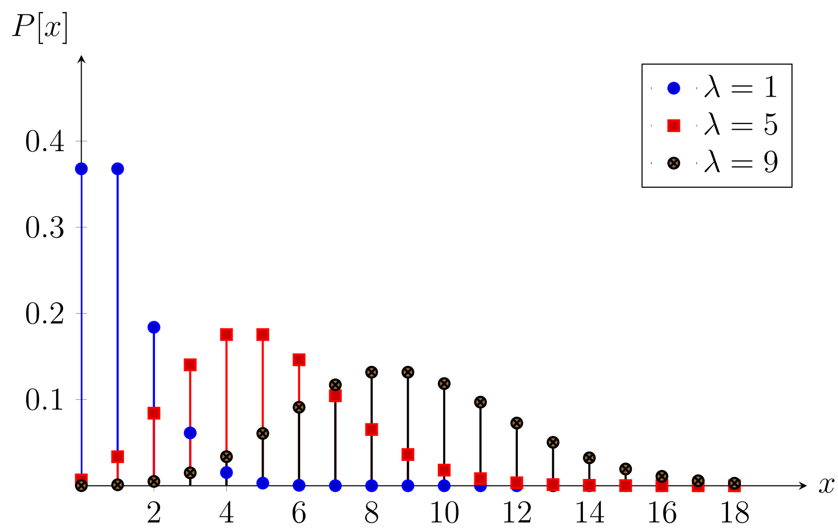


Figura 6.4: Función de densidad Poisson

6.2. Distribuciones continuas

6.2.1. Distribución exponencial

Definición 6.8 (Distribución exponencial) *Si una variable aleatoria X tiene tiene una función de densidad dada por*

$$f(x) = \begin{cases} \theta e^{-\theta x}. & \textit{para } x = 0, 1, \ldots \\ 0, & d.o.m \end{cases}$$

donde $\theta > 0$, entonces X es definida como una distribución exponencial. (Hogg and Craig, 1978)

Propiedades:

- Función generadora de momentos

$$M(t) = \frac{\theta}{\theta - t} \text{ para } t < \theta$$

- Función característica

$$\phi(t) = \frac{\theta}{\theta - it} \text{ para } t < \theta$$

- Esperanza

$$E(X) = \frac{1}{\theta}$$

- Varianza

$$Var(X) = \frac{1}{\theta^2}$$

- Asimetría

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = 1/2$$

- Curtosis

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = 1/9$$

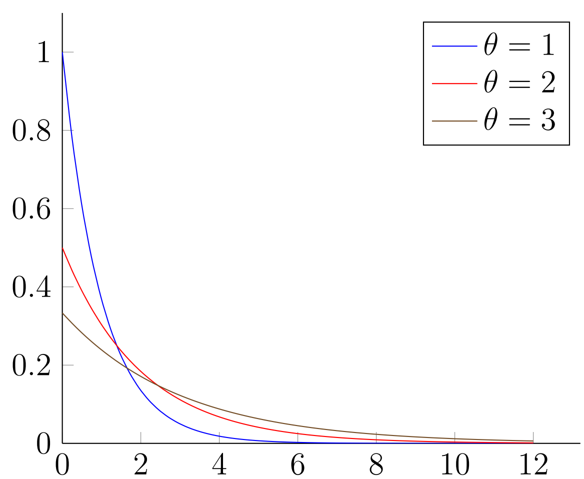


Figura 6.5: Distribución exponencial

6.2.2. Distribución normal

Definición 6.9 (Distribución normal) *Si una variable aleatoria X tiene tiene una función de densidad dada por*

$$f(x) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, & para \; -\infty < x < \infty \\ 0, & d.o.m \end{cases}$$

donde $-\infty < \mu < \infty$ y $\sigma > 0$ entonces X es definida como una distribución normal. (Casella and Berger, 2002)

Propiedades:

- Función generadora de momentos

$$M(t) \;=\; E[e^{tX}] = e^{t(t\sigma^2+2\mu)/2} \text{ para } -\infty < t < \infty$$

- Función característica

$$\phi(t) \;=\; E[e^{tX}] = e^{it(it\sigma^2+2\mu)/2} \text{ para } -\infty < t < \infty$$

- Esperanza

$$E(X) \;=\; \mu$$

- Varianza

$$Var(X) = \sigma^2$$

- Asimetría

$$E\left[\left(\frac{-\mu}{\sigma}\right)^3\right] = 0$$

- Curtosis

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = 3$$

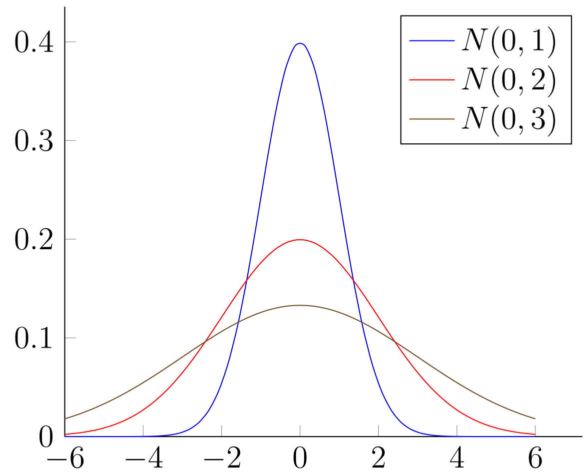


Figura 6.6: Distribución normal

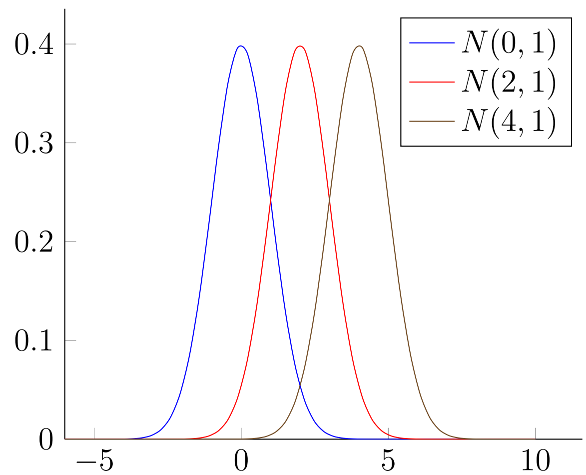


Figura 6.7: Distribución normal

6.2.3. Distribución gamma

La función gamma La función gamma fue introducida por primera vez por el matemático suizo Leonhard Euler (1707-1783), con el objetivo de generalizar la función factorial a valores no enteros.Más tarde por su gran importacia fue estudiada por grandes matemáticos como Adrien-Marie Legendre (1752-1833), Carl Friedrich Gauss (1777-1855), Christoph Gudermann (1798-1852), entre otros.

La función gamma pertenece a una categoría de funciones transcendentes especiales, y esta función ocurre en algunas constantes matemáticas especiales. Esta aparece en varias areas de estudio, como en las series asintóticas, integrales definidas, series hipergeométricas, la función Zeta de Riemann, teoría de números, otras.

Definición 6.10 (Función gamma) *La función gamma puede definirse como*

$$\Gamma(x) \quad = \quad \int_0^\infty t^x e^{-t} dt \text{ para } x \in (0, \infty)$$

Es posible demostrar que $\Gamma(x+1) = n!$. Para hacerlo considérese lo siguiente:

$$\Gamma(x+1) \quad = \quad \int_0^\infty t^x e^{-t} dt$$

Sea $u = t^x$ entonces $du = xt^{x-1}dx$ y $dv = e^{-t}dt$ y $v = -e^{-t}$. Empleando integración por partes

$$uv - \int vdu$$

Luego,

$$\begin{aligned}
\Gamma(x+1) &= \int_0^\infty t^x e^{-t} dt \\
&= -t^x e^{-t}|_0^\infty - \int_0^\infty -e^{-t} x t^{x-1} dt \\
&= x \int_0^\infty t^{x-1} e^{-t} dt \\
&= x\Gamma(x)
\end{aligned}$$

Para $\Gamma(x)$, se tiene

$$\Gamma(x) \quad = \quad \int_0^\infty t^{x-1} e^{-t} dt$$

Sea $u = t^{x-1}$ entonces $du = (x-1)t^{x-2}dx$ y $dv = e^{-t}dt$ y $v = -e^{-t}$. Empleando integración por partes

$$uv - \int vdu$$

Luego,

$$\begin{aligned}
\Gamma(x) &= \int_0^\infty t^{x-1}e^{-t}dt \\
&= -t^{x-1}e^{-t}|_0^\infty - \int_0^\infty -e^{-t}(x-1)t^{x-2}dt \\
&= (x-1)\int_0^\infty t^{x-2}e^{-t}dt \\
&= (x-1)\Gamma(x-1)
\end{aligned}$$

Podemos ver que si continuamos de forma iterativa, llegamos a expresiones como:

$$\begin{aligned}
\Gamma(x-1) &= (x-2)\Gamma(x-2) \\
\Gamma(x-2) &= (x-3)\Gamma(x-3) \\
\Gamma(x-3) &= (x-4)\Gamma(x-4)
\end{aligned}$$

y así sucesivamente por lo que

$$\Gamma(x+1) = x(x-1)(x-2)\Gamma(x-2)... = n!$$

Distribución Gamma

Definición 6.11 (Distribución gamma) *Si una variable aleatoria X tiene tiene una función de densidad dada por*

$$f(x) = \begin{cases} \frac{x^{\beta-1}e^{-x/\alpha}}{\alpha^\beta \Gamma x}, & para\ x\ = 0, 1, ... \\ 0 & d.o.m \end{cases}$$

donde $x > 0$ y $\alpha > 0$, entonces X es definida como una distribución Gamma. (Casella and Berger, 2002)

Propiedades:

- Función generadora de momentos

$$M(t) \;=\; (1-\alpha t)^{-\beta} \text{ para } t < 1/2$$

- Función característica

$$\phi(t) \;=\; (1-\alpha i t)^{-\beta} \text{ para } t < 1/2$$

- Esperanza

$$E(X) \;=\; \alpha\beta$$

- Varianza

$$Var(X) = \alpha^2\beta$$

- Asimetría

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{2}{\sqrt{\beta}}$$

- Curtosis

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = 3 + \frac{6}{\beta}$$

6.2.4. Distribución Beta

Definición 6.12 *Distribución Beta Si una variable aleatoria X tiene tiene una función de densidad dada por*

$$f(x) = \begin{cases} \frac{x^{\alpha-1}(1-x)^{\beta-1}}{B(\alpha,\beta)}, & para\ 0 < x < 1 \\ 0, & d.o.m \end{cases}$$

entonces X es definida como una distribución Beta. (Casella and Berger, 2002)

Propiedades:

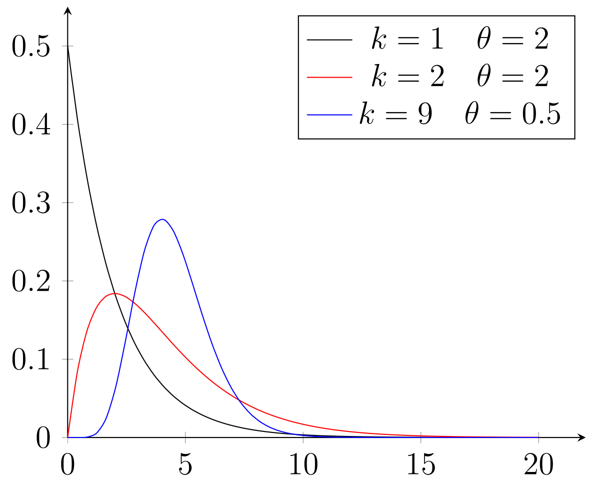


Figura 6.8: Distribución gamma

- Función generadora de momentos

$$M(t) \quad =$$

- Función característica

$$\phi(t) \quad =$$

- Esperanza

$$E(X) \quad = \quad \frac{\alpha}{\alpha + \beta}$$

- Varianza

$$Var(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$$

- Asimetría

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{2(\beta - \alpha)\sqrt{1 + \alpha + \beta}}{\sqrt{\alpha\beta}(\alpha + \beta + 2)}$$

- Curtosis

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{(\alpha^2\beta + \alpha\beta^2 - 2\alpha\beta + 2\beta^2 + 2\beta^2)(a + \alpha + \beta)}{\alpha\beta(\alpha + \beta + 2)(\alpha + \beta + 3)}$$

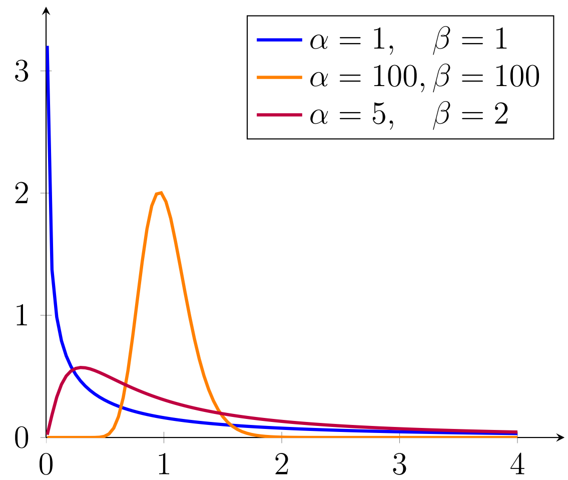


Figura 6.9: Distribución beta

6.2.5. Distribución Uniforme continua

Definición 6.13 (Distribución uniforme continua) *Si una variable aleatoria X tiene tiene una función de densidad dada por*

$$f(x) = \begin{cases} \frac{1}{b-a}, & para\ a < x < b \\ 0, & d.o.m \end{cases}$$

entonces X es definida como una distribución Uniforme. (Degroot, 1988)

Propiedades:

- Función generadora de momentos

$$M(t) = E[e^{tX}] = \begin{cases} 1 & \text{para } t = 0 \\ \frac{e^{bt}-e^{at}}{t(b-a)} & t \neq 0 \end{cases}$$

- Función característica

$$\phi(t) = E[e^{tX}] = \begin{cases} 1 & \text{para } t = 0 \\ \frac{e^{bt}-e^{at}}{t(b-a)} & t \neq 0 \end{cases}$$

- Esperanza

$$E(X) = \frac{1}{2}(a+b)$$

- Varianza

$$Var(X) = \frac{1}{12}(b-a)^2$$

- Asimetría

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = 0$$

- Curtosis

$$E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{9}{5}$$

6.2.6. Distribución Weibull

Definición 6.14 (Distribución Weibull) *Una variable aleatoria X con la siguiente función de densidad*

$$f(x) = \begin{cases} \frac{c}{\alpha}\left(\frac{x-k}{\alpha}\right)^{c-1} e^{-((x-k)/\alpha)^c}, & para\ x < k \\ 0, & d.o.m \end{cases}$$

se dice que sigue distribución Weibull. (Johnson et al., 1994)

6.2.7. Distrtibución Cauchy

Definición 6.15 (Distribución Cauchy) *Una variable aleatoria X con la siguiente función de densidad*

$$f(x) = \begin{cases} (\pi c)^{-1}\left[1+\left(\frac{x-\theta}{c}\right)^2\right]^{-1} & para\ c<0 \\ 0, & d.o.m \end{cases}$$

se dice que sigue distribución Cauchy. (Johnson et al., 1994)

6.2.8. Distribución log-normal

Definición 6.16 (Distribución log-normal) *Si una variable aleatoria X tiene tiene una función de densidad dada por*

$$f(x) = \begin{cases} \frac{1}{x\sqrt{2\pi\sigma^2}}e^{-\frac{(log(x)-\mu)^2}{2\sigma^2}}, & para\ 0<x<\infty \\ 0, & d.o.m \end{cases}$$

donde $< \mu < \infty$ y $\sigma > 0$ entonces, se dice que X tiene una distribución log-normal. (Johnson et al., 1994)

6.2.9. Distribución Pareto

Definición 6.17 (Distribución Pareto) *Una variable aleatoria X con la siguiente función de densidad*

$$f(x) = \begin{cases} \frac{\theta k^\theta}{x^{\theta+1}} & para\ k>0, \theta>0 x \leq k \\ 0, & d.o.m \end{cases}$$

se dice que sigue una distribución Pareto. (Johnson et al., 1995)

6.2.10. Distribución Rayleigh

Definición 6.18 (Distribución Rayleigh) *Una variable aleatoria X con la siguiente función de densidad*

$$f(x) = \begin{cases} \frac{x}{\sigma^2} e^{-\frac{x^2}{2\sigma^2}} & para\ x < \infty \\ 0, & d.o.m \end{cases}$$

se dice que sigue una distribución Raylegh. (Johnson et al., 1995)

6.2.11. Función delta de Dirac

La delta de Dirac o función delta de Dirac es una distribución o función generalizada introducida por primera vez por el físico inglés Paul Dirac y, como distribución, define un funcional en forma de integral sobre un cierto espacio de funciones. Se escribe como:

$$\delta_a(x) = \delta(x - a) \tag{6.1}$$

siendo $\delta(x)$, la función tiende a infinito cuando $x = 0$, y para cualquier otro valor de x es 0.

Definición 6.19 *La delta de Dirac es una función generalizada que viene definida por la siguiente fórmula integral:*

$$\int_{-\infty}^{\infty} \delta(x - a) f(x) dx = f(a) \tag{6.2}$$

La delta de Dirac no es una función estrictamente hablando, puesto que se puede ver que requeriría tomar valores infinitos. A veces, informalmente, se define la delta de Dirac como el límite de una sucesión de funciones que tiende a cero en todo punto del espacio excepto en un punto para el cual divergería hacia infinito; de ahí la "definición convencional"dada por la también convencional fórmula aplicada a las funciones definidas a trozos:

$$\delta(x) = \begin{cases} \infty, & x = 0 \\ 0, & x \neq 0 \end{cases}$$

Propiedades

$$\begin{aligned}
\delta(x) &= \delta(-x) \\
f(x)\delta^{'}(x) &= -f^{'}(x)\delta(x) \\
\delta^{'}(x) &= -\delta^{'}(-x) \\
x^n\delta(x) &= 0 \qquad \forall n > 0, x \in \mathbb{R} \\
(x-a)^n\delta(x-a) &= 0 \qquad \forall n > 0 \\
\delta(ax-b) &= |a|^{-1}\delta(x-(b/a)) \qquad \forall a > 0 \\
h(x)\delta(x-a) &= h(a)\delta(x-a) \\
h(x)\delta^{'}(x-a) &= h(a)\delta^{'}(x-a) - h^{'}(a)\delta(x-a)
\end{aligned}$$

Las funciones de la sucesión anterior no son continuas, pero es posible encontrar una sucesión de funciones continuas, como por ejemplo las dadas por

$$\delta_n(t) = \frac{n}{\pi(1+n^2t^2)} \tag{6.3}$$

las cuales tienen tambien integral 1 y parece además no haber inconveniente en considerar que convergen a la Delta de Dirac.

Es posible incluso encontrar sucesiones de funciones analíticas con integral 1 y con la propiedad de tener a la Delta de Dirac como "límite", tal es el caso de

$$\delta_n(t) = \frac{n}{\sqrt{(\pi)}} e^{n^2t^2} \tag{6.4}$$

Capítulo 7

Análisis de regresión

7.1. Mínimos cuadrados

El método de mínimos cuadrados tiene una larga historia que se remonta a los principios del siglo XIX. En Junio de 1801, Zach, un astrónomo que Gauss había conocido dos años antes, publicaba las posiciones orbitales del cuerpo celeste Ceres, un nuevo "pequeño planeta" descubierto por el astrónomo italiano G. Piazzi en ese mismo año. Desafortunadamente, Piazzi sólo había podido observar 9 grados de su órbita antes de que este cuerpo desapareciese tras de el sol. Zach publicó varias predicciones de su posición incluyendo una de Gauss que difería notablemente de las demás. Cuando Ceres fue redescubierto por Zach en Diciembre de 1801 estaba casi exactamente en donde Gauss había predicho. Aunque todavía no había revelado su método, Gauss había descubierto el método de mínimos cuadrados. En un trabajo brillante logró calcular la órbita de Ceres a partir de un número reducido de observaciones, de hecho, el método de Gauss requiere sólo un mínimo de 3 observaciones y todavía es, en esencia, el utilizado en la actualidad para calcular las órbitas.

Mínimos cuadrados es una técnica de análisis numérico enmarcada dentro de la optimización matemática, en la que, dados un conjunto de pares ordenados (variable independiente, variable dependiente) y una familia de funciones, se intenta encontrar la función continua, dentro de dicha familia, que mejor se aproxime a los datos (un "mejor ajuste"), de acuerdo con el criterio de mínimo error cuadrático.

Ejemplo:

Se cree que la cantidad de libras de vapor usadas en una planta por mes está relacionada con la temperatura ambiente promedio. A continuación se presenta las consultas y temperaturas del último año. (Montgomery et al., 2015)

Mes	Temperatura	Uso
Ene	21	185.79
Feb	24	214.47
Mar	32	288.03
Abr	47	424.84
May	50	454.68
Jun	59	539.03
Jul	68	621.55
Ago	74	675.06
Sep	62	562.03
Oct	50	452.93
Nov	41	369.95
Dic	30	273.98

Tabla 7.1: Datos de Temperatura-libras de vapor usado

Entonces, se desea obtener por mínimos cuadrados un modelo lineal que mejor represente los datos.

Solución:

Primero, graficamos los datos. Podemos apreciar que el modelo que el modelo sugerido por los datos es leneal es decir de la forma:

$$y = \beta_o + \beta_1 x + \varepsilon_i \text{ para i= } 1, 2, ..., 12$$

El criterio de mínimos cuadrados es minimizar la suma de los cuadrados de los errores, es decir:

$$S(\beta_o, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_o - \beta_1 x_i)$$

Los estimadores por mínimos cuadrados se designaran por $\hat{\beta}_o$ y $\hat{\beta}_1$. Derivando par-

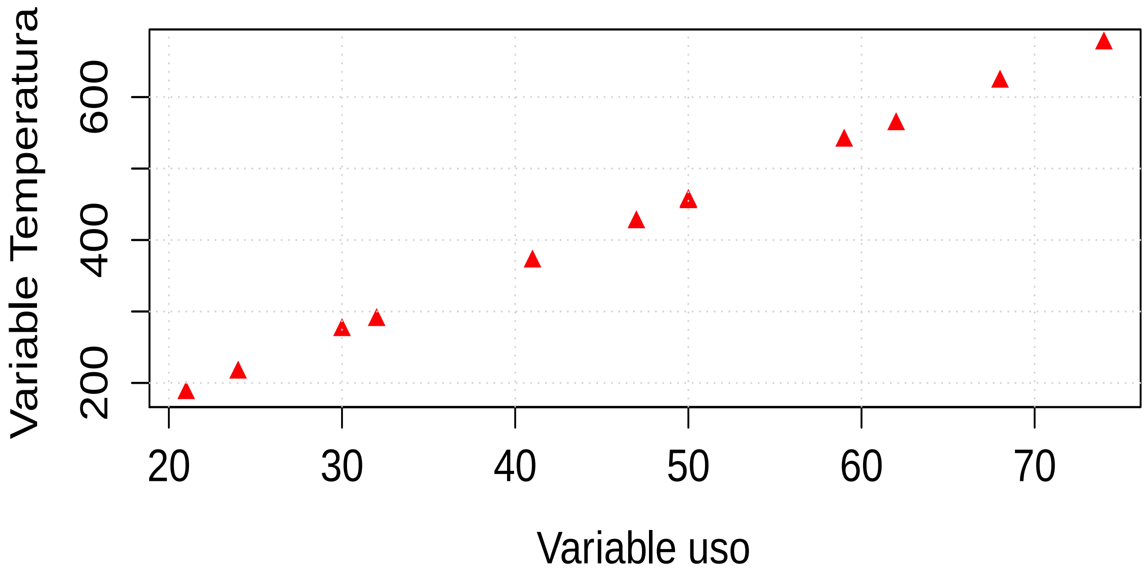


Figura 7.1: Gráfico de dispersión

cialmente con respecto a $\hat{\beta}_o$ y $\hat{\beta}_1$

$$\begin{aligned}\frac{\partial S}{\partial \beta_o}(\beta_o, \beta_1) &= 2\sum_{i=1}^{n}(y_i - \beta_o - \beta_1 x_i)\\ \frac{\partial S}{\partial \beta_1}(\beta_o, \beta_1)x_i &= 2\sum_{i=1}^{n}(y_i - \beta_o - \beta_1 x_i)\end{aligned}$$

Las ecuaciones a solucionar son:

$$\sum(y_i - \hat{\beta}_o - \hat{\beta}_1 x_i) = 0 \tag{7.1}$$

$$\sum(y_i - \hat{\beta}_o - \hat{\beta}_1 x_i)x_i = 0 \tag{7.2}$$

De Eq(7.1) se tiene

$$\begin{aligned}
\sum(y_i - \hat{\beta}_o - \hat{\beta}_1 x_i) &= 0 \\
\sum y_i - n\hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{n} x_i &= 0 \\
n\hat{\beta}_0 &= \sum y_i - \hat{\beta}_1 \sum_{i=1}^{n} x_i \\
\beta_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \qquad (7.3)
\end{aligned}$$

donde $\bar{x} = \frac{1}{n}\sum x_i$ y $\bar{y} = \frac{1}{n}\sum y_i$.

De Eq(7.2) se tiene

$$\sum(y_i - \hat{\beta}_o - \hat{\beta}_1 x_i) = 0$$

sustituyendo Eq(7.3) en la expresión anterior

$$\begin{aligned}
\sum(y_i - \bar{y} + \hat{\beta}_1 \hat{x} - \hat{\beta}_1 x_i) &= 0 \\
\sum y_i - \bar{y} &= \hat{\beta}_1 \sum(x_i - \bar{x}_i) \\
\hat{\beta}_1 &= \frac{\sum(y_i - \bar{y})x_i}{\sum(x_i - \bar{x})x_i} \\
&= \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} \qquad (7.4)
\end{aligned}$$

Realizando las sustituciones correspondientes, el modelo obtenido por mínimos cuadrados es:

$$y_i = 0.6938 + 0.1085 x_i + \epsilon_i \text{ para } i = 1:12 \qquad (7.5)$$

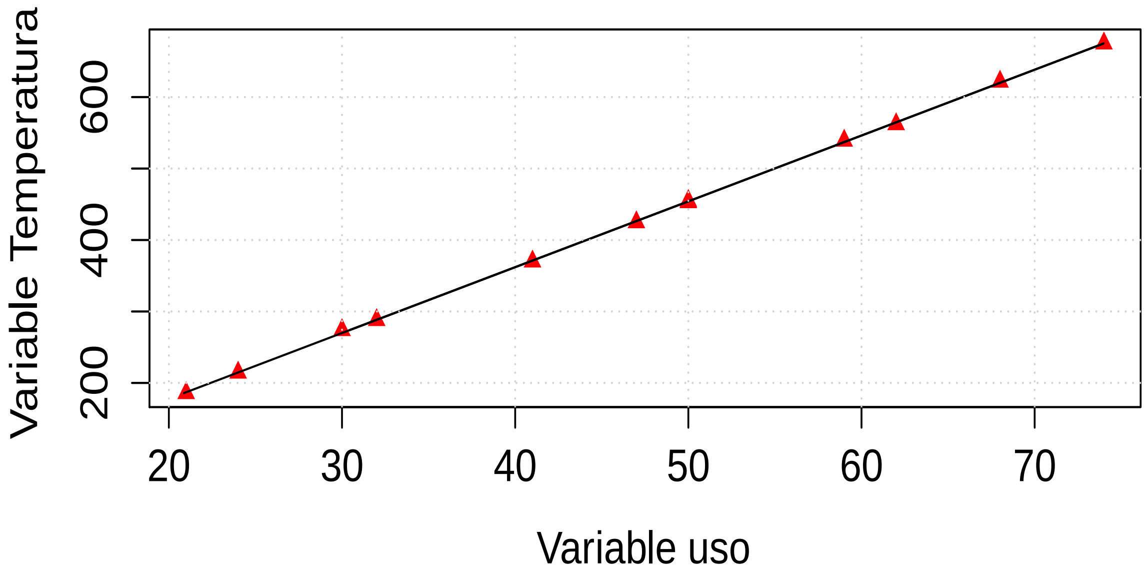


Figura 7.2: Gráfico de dispersión

Código en R

```
data

plot(data[,-1], col="red",main="",
     xlab="Variable uso",ylab="Variable Temperatura",pch=16)

y<-data[,2]
x<-data[,3]
n<-length(y)

# Función para estimar por mínimos cuadrados
model.mc<-function(y,x){
  b1<- (sum(x*y) - (sum(y)*sum(x))/n)/(sum(x^2)- sum(x)^2/n)
  bo<-mean(y)-b1*mean(x)
```

```
  resultado<-data.frame("b1"=b1,"bo"=bo)
  return(resultado)
}

mod1<-model.mc(y,x)

yhat<-mod1$bo+mod1$b1*x

plot(data[,-1], col="red",main="",
    xlab="Variable uso",ylab="Variable Temperatura",pch=17);grid()
lines(yhat,x,col="black");grid()
legend(20,650,c("valor","ajustado"),col=c("red","black"),
       lty=c(NA,1),pch=c(17,NA))
text(55,500,"Recta ajustada")
text(50,200,"Mínimos cuadrados")
```

7.1.1. Caso Challenger

El Challenger despega por última vez a las 11 : 38 del 28 de enero de 1986.El hielo no fue la causa del accidente del challenger, pero si las bajas temperaturas. Los anillos de la junta inferior del SRB derecho del Challenger se habían vuelto duros y quebradizos por culpa de las bajísimas temperaturas de la noche anterior. Como resultado, cuando los gases alcanzaron la junta en el momento de la ignición vaporizaron una sección de casi 70º de ambos anillos, eliminando la última barrera de defensa. El transbordador podría haber explotado en la misma rampa de despegue, pero milagrosamente los óxidos de aluminio resultantes de la combustión del propelente sólido sellaron temporalmente la junta.

Acausa de lo anterior 73 segundos después del despegue el challenger comienza a desintegrarse. En menos de un segundo, miles de toneladas de hidrógeno y oxígeno líquidos se mezclan mientras el transbordador se hace añicos bajo el efecto de las

La gráfica de los datos se muestra en la Figura

	Temp	Fail
1	66.00	0.00
2	67.00	0.00
3	68.00	0.00
4	70.00	0.00
5	72.00	0.00
6	75.00	0.00
7	76.00	0.00
8	79.00	0.00
9	53.00	1.00
10	58.00	1.00
11	70.00	1.00
12	75.00	1.00
13	67.00	0.00
14	67.00	0.00
15	69.00	0.00
16	70.00	0.00
17	73.00	0.00
18	76.00	0.00
19	78.00	0.00
20	81.00	0.00
21	57.00	1.00
22	63.00	1.00
23	70.00	1.00

Tabla 7.2: Datos de la catastrofe del Challenger

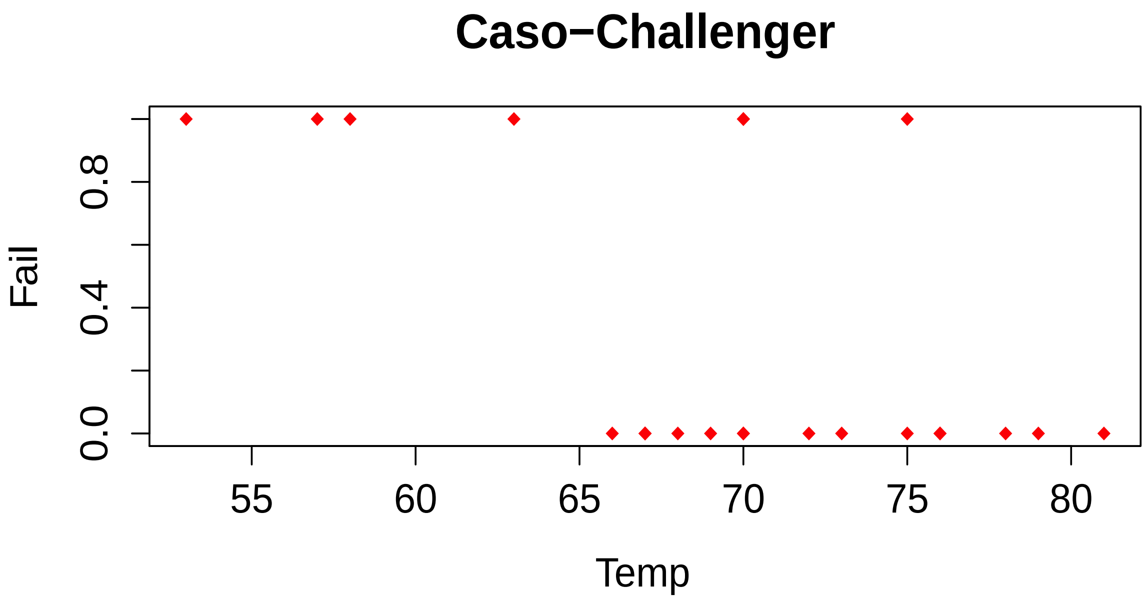


Figura 7.3: Gráfico de dispersión

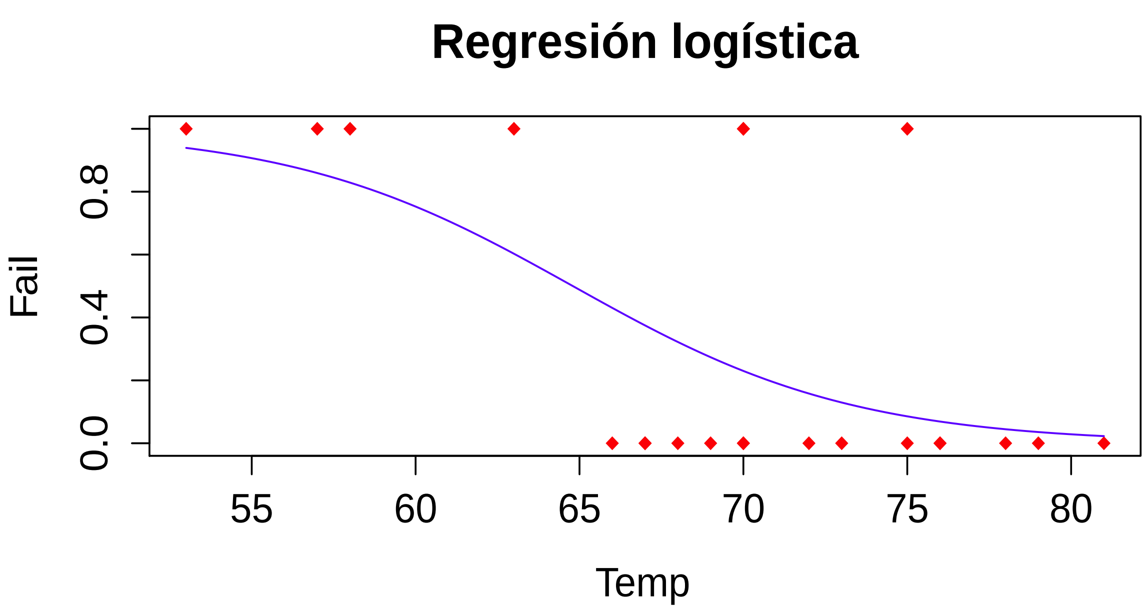


Figura 7.4: Gráfico de ajuste del modelo logístico

Capítulo 8

Análisis de componentes principales

8.1. Eigenvalores y eigenvectores

Los vectores propios, autovectores o eigenvectores de un operador lineal son los vectores no nulos que, cuando son transformados por el operador, dan lugar a un múltiplo escalar de sí mismos, con lo que no cambian su dirección. Este escalar λ recibe el nombre valor propio, autovalor, valor característico o eigenvalor. A menudo, una transformación queda completamente determinada por sus vectores propios y valores propios.

La palabra alemana eigen, que se traduce en español como propio, se usó por primera vez en este contexto por David Hilbert en 1904.Eigen se traducíó también como inherente, característico o el prefijo auto-, donde se aprecia el énfasis en la importancia de los valores propios para definir la naturaleza única de una determinada transformación lineal.

8.1.1. Diagonalización

Supongamos que queremos calcular la potencia de A^n de una matriz. Esto puede llevar una gran cantidad de operaociones. Pero si la matriz A es diagonalizable, el cálculo se fácilita notablemente. En ese caso diagonalizando la matriz tendremos $P^{-1}AP = D$ por $A = PDP^{-1}$, es decir que

$$A^n = PD^N P^{-1} \tag{8.1}$$

lo cual es muy fácil de calcular teniendo en cuenta que para hallar la potencia D^n basta elevar a la potencia n cada elemento diagonal.

Ejemplo: Dada la matriz

$$\begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix}$$

Calcular A^9.

Solución:

Primero planteamos la ecuación

$$Av = \lambda v$$

que es

$$\begin{bmatrix} 3 & 2 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$$

solucionando el sistema

$$\begin{bmatrix} 3-\lambda & 2 \\ 0 & 1-\lambda \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Los valores de λ que han de ser el sistema compatible determinado, son los valores propios. Las soluciones v de este sistema para cada λ, son los vectores propios. Entonces el sistema será compatible determinado cuando ocurra que

$$det \begin{bmatrix} 3-\lambda & 2 \\ 0 & 1-\lambda \end{bmatrix} = 0$$

es decir $(3-\lambda)(a-\lambda)=0$ lo cual se cumple para valores de $\lambda=1$, $\lambda=3$. Estos son los valores propios, para hallar los vectores propios

$$\begin{bmatrix} 3-\lambda & 2 \\ 0 & 1-\lambda \end{bmatrix}\begin{bmatrix} x \\ y \end{bmatrix}=\lambda\begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

resolvemos el sistema para $\lambda=1$, $\lambda=3$. cuyas soluciones son $V_1=(1,-1)$ y $V_2=(1,0)$.

Ahora el paso siguiente es expresar los valores propios y vectores propios de la siguiente manera

$$A^9=\begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}^9\begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}=\begin{bmatrix} 19683 & 19682 \\ 0 & 1 \end{bmatrix}$$

8.1.2. Matrices semejantes

Dos matrices cuadradas de orden n, A y B, son semejantes si existe una matriz cuadrada, P, con determinante distinto de cero, que satisfaga $A=PDP^{-1}$. A la matriz A se le llama transformada de A mediante la matriz de paso P.

Propiedades:

1. Si A y B son semejantes, entonces son equivalentes.
2. Tienen el mismo rango.
3. Poseen la misma traza
4. Los mismos valores propios
5. Poseen el mismo polinomio característico
6. Poseen el mismo determinante

Definición 8.1 (Eigenvectores) *Llamaremos vectores propios (eigenvectores) de una matriz cuadrada de orden n a aquellos vectores cuya dirección no se modifica al transformarlos mediante la matriz. Por tanto* $\boldsymbol{e}$ *es un vector propio de la matriz A si se verifica que*

$$A\boldsymbol{e} = \lambda \boldsymbol{e} \tag{8.2}$$

donde λ *es un escalar, que se denomina valor propio de la matriz.*

Ejemplo:

Calcular los valores y vectores propios para la matriz

$$\mathbf{A} = \begin{pmatrix} 4 & -5 \\ 2 & -3 \end{pmatrix}$$

Primero calculamos los valores propios

$$det(A - \lambda I) = \begin{vmatrix} 4 & -5 \\ 2 & -3 \end{vmatrix} = (4 - \lambda)(-3 - \lambda) + 10 = \lambda^2 - \lambda - 2$$

Con lo cual obtenemos los valores propios $\lambda_1 = -1$, $\lambda_2 = 2$ Buscamos los correspondientes vectores propios. Para $\lambda_1 = -1$:

$$det(A - (-1)I)\mathbf{e} = 0 \rightarrow \begin{pmatrix} 5 & -5 \\ 2 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

cuya solución es $x = y$ lo cual implíca múltiplos

$$e_1 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

$\lambda_1 = 2$ se obtiene siguiendo el mismo procedimiento.

8.1.3. Componentes principales

Supongamos que tenemos una matriz $\mathbf{X}$ de datos

$$\mathbf{X} = \begin{pmatrix} X_{1,1} & X_{1,2} & ... & X_{1,p} \\ X_{2,1} & X_{2,2} & ... & X_{2,p} \\ \vdots & \vdots & \vdots & \vdots \\ X_{n,1} & X_{n,2} & ... & X_{n,p} \end{pmatrix}$$

En donde cada renglón reprensenta a cada individuo, es decir, $X_{i,1}, ..., X_{i,p}$. Con una matriz de varianzas y covarianzas de

$$Var(\mathbf{X}) = \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{21}^2 & \sigma_2^2 & \dots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_{n1} & \sigma_{2n} & \dots & \sigma_n^2 \end{pmatrix}$$

Consideremos la siguientes combinaciones lineales

$$\begin{aligned} Y_1 &= e_{11}X_1 + \ e_{12}X_2 + \ \dots + \ e_{1p}X_p \\ Y_2 &= e_{21}X_1 + \ e_{22}X_2 + \ \dots + \ e_{2p}X_p \\ &\vdots \\ Y_n &= e_{p1}X_1 + \ e_{p2}X_2 + \ \dots + \ e_{pp}X_p \end{aligned}$$

Con varianza poblacional

$$Var(Y_i) = \sum_{k=1}^{p} \sum_{p}^{p} e_{ik}e_{il}\sigma_{kl} = \mathbf{e}^{'}\Sigma\mathbf{e} \tag{8.3}$$

Entonces el problema que se tiene es máximizar

$$Var(Y_i) \quad = \quad \mathbf{e}^{'}\Sigma\mathbf{e}$$

La solución puede obtenerse mediante el método de multiplicadores de Lagrange. El procedimiento consite en lo siguiente:

1. Se introduce la restricción mediante el multiplicador de Lagrange:

$$L = \mathbf{e}^{'}S\mathbf{e} - \lambda(\mathbf{e}^{'}\mathbf{e} - 1)$$

2. Se maximiza la expresión derivando parcialmente los componentes e igualando a cero

$$\begin{aligned} \frac{\partial L}{\partial \mathbf{e}_i} &= 2S\mathbf{e} - 2\lambda\mathbf{e} \\ 0 &= (S - \lambda)\mathbf{e} \end{aligned}$$

Por argumento similares el valor de lambda puede obtenerse mediante el cálculo de los eigenvalores

$$(S - \lambda I)\mathbf{e} \;\; = \;\; \mathbf{0}$$

8.1.4. Ejemplo componentes principales

Supongamos que se tiene los datos de 15 estudiantes con sus respectivas calificaciones en cada asignatura(8 asignaturas). Supongamos que se desea analizar todas las variables en una escala de dimensión menor, para mayor referencia Peña (2002).

Alumno	Lengua	Matematicas	Fisica	Ingles	Filosofia	Historia	Quimica	Educ..Fisica
1	5	5	5	5	5	5	5	5
2	7	4	3	8	4	7	3	8
3	5	8	7	6	5	6	7	5
4	7	2	4	8	7	7	3	6
5	8	9	10	8	8	7	9	4
6	4	9	8	4	3	4	7	5
7	6	4	4	6	5	5	3	7
8	4	7	8	3	3	2	8	3
9	5	5	4	5	6	5	5	1
10	7	4	5	7	8	8	4	6
11	7	8	8	7	7	6	7	9
12	4	3	3	4	3	2	1	4
13	7	4	4	7	8	7	4	5
14	3	5	5	2	3	3	5	7
15	5	6	6	5	5	5	6	6

Tabla 8.1: Datos de 15 estudiantes

	Lengua	Matematicas	Fisica	Ingles	Filosofia	Historia	Quimica	Educ..Fisica
1	-0.60	-0.53	-0.60	-0.67	-0.33	-0.27	-0.13	-0.40
2	1.40	-1.53	-2.60	2.33	-1.33	1.73	-2.13	2.60
3	-0.60	2.47	1.40	0.33	-0.33	0.73	1.87	-0.40
4	1.40	-3.53	-1.60	2.33	1.67	1.73	-2.13	0.60
5	2.40	3.47	4.40	2.33	2.67	1.73	3.87	-1.40
6	-1.60	3.47	2.40	-1.67	-2.33	-1.27	1.87	-0.40
7	0.40	-1.53	-1.60	0.33	-0.33	-0.27	-2.13	1.60
8	-1.60	1.47	2.40	-2.67	-2.33	-3.27	2.87	-2.40
9	-0.60	-0.53	-1.60	-0.67	0.67	-0.27	-0.13	-4.40
10	1.40	-1.53	-0.60	1.33	2.67	2.73	-1.13	0.60
11	1.40	2.47	2.40	1.33	1.67	0.73	1.87	3.60
12	-1.60	-2.53	-2.60	-1.67	-2.33	-3.27	-4.13	-1.40
13	1.40	-1.53	-1.60	1.33	2.67	1.73	-1.13	-0.40
14	-2.60	-0.53	-0.60	-3.67	-2.33	-2.27	-0.13	1.60
15	-0.60	0.47	0.40	-0.67	-0.33	-0.27	0.87	0.60

Tabla 8.2: Datos centrados

	Lengua	Matematicas	Fisica	Ingles	Filosofia	Historia	Quimica	Educ..Fisica
Lengua	2.26	-0.27	0.19	2.71	2.43	2.47	-0.01	0.89
Matematicas	-0.27	4.84	4.30	-0.52	-0.26	-0.37	4.42	-0.37
Fisica	0.19	4.30	4.54	-0.14	0.43	-0.10	4.41	-0.33
Ingles	2.71	-0.52	-0.14	3.52	2.76	3.17	-0.38	1.14
Filosofia	2.43	-0.26	0.43	2.76	3.67	2.98	0.38	0.21
Historia	2.47	-0.37	-0.10	3.17	2.98	3.50	-0.04	1.17
Quimica	-0.01	4.42	4.41	-0.38	0.38	-0.04	4.84	-0.77
Educ..Fisica	0.89	-0.37	-0.33	1.14	0.21	1.17	-0.77	3.97

Tabla 8.3: Varianzas y covarianzas

Los eigenvalores correspondientes son:

$$\lambda = (13.7480, 11.8272, 3.7729, 0.7966, 0.5311, 0.2883, 0.1279, 0.0410)$$

Los eigenvectores correspondientes:

8.1.5. Número de componentes a retener

- Retener suficientes componentes para garantizar un porcentaje predefinido de la varianza total, por ejemplo el 80 %.
- Retener aquellas componentes cuyos autovalores superen el promedio de todos

1	2	3	4	5	6	7	8
-0.11	0.41	-0.03	-0.15	-0.39	-0.10	-0.35	0.72
0.56	0.12	0.13	-0.35	0.23	0.66	-0.20	0.06
0.53	0.19	0.07	0.13	-0.47	-0.07	0.65	0.05
-0.17	0.49	-0.02	-0.55	-0.26	-0.09	-0.04	-0.60
-0.09	0.49	-0.31	0.67	-0.05	0.39	-0.15	-0.21
-0.15	0.50	-0.03	-0.10	0.67	-0.13	0.45	0.24
0.56	0.17	-0.04	0.16	0.23	-0.61	-0.43	-0.15
-0.13	0.17	0.94	0.24	-0.01	0.00	-0.08	-0.07

Tabla 8.4: Eigenvectores

los autovalores, $\sum_{i=1}^{p} \lambda_i/p$. Para la matriz de correlación este promedio es 1.0.

- Utilizar una representacián gráfica de λ_i frente a i, y determinar el punto en el que se produce la transición entre los autovalores grandes y los pequeños.

- Utilizar tests de significación.
  Para probar H_{0k} se calcula el promedio de los últimos k autovalores

$$\bar{\lambda} = \sum_{i=p-k+1}^{p} \frac{\lambda_i}{k}$$

y se calcula el estadístico

$$u = \left(n - \frac{2p+11}{6}\right)\left(kln(\bar{\lambda}) - \sum_{i=p-k+1}^{p} ln(\lambda_i)\right)$$

que sigue aproximadamente una distribución χ_v^2. Por lo que se rechaza H_0 si $u \leq \chi_v^2$

Para el ejemplo de las calificaciones, considerando los siguientes datos, sobre la proporción de la varianza explicada por las componentes, se considerá 2 componentes como apropiado, ya que representan el 82 % de la variabilidad.

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8
Sd	3.7078	3.4391	1.9424	0.8926	0.7288	0.5370	0.3576	0.2026
Prop of Var	0.4416	0.3799	0.1212	0.0256	0.0171	0.0093	0.0041	0.0013
Cumm Prop	0.4416	0.8215	0.9427	0.9683	0.9853	0.9946	0.9987	1.0000

Tabla 8.5: Proporción de varianzas

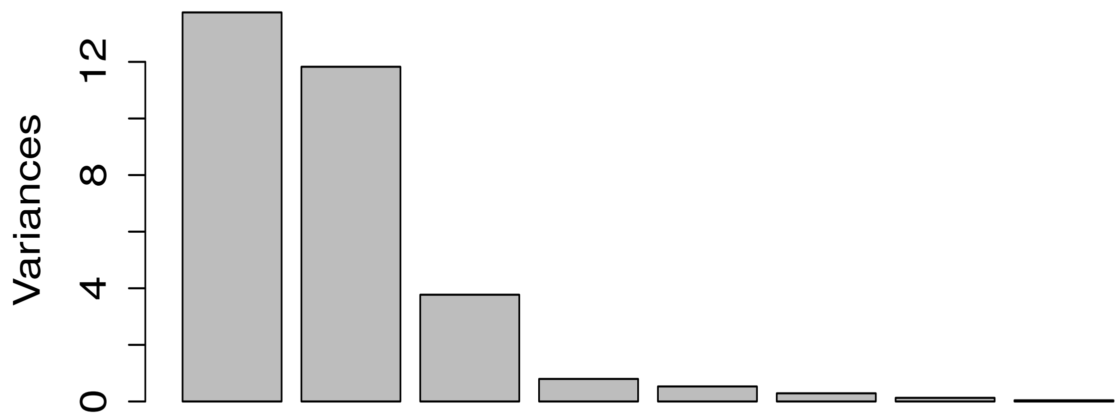


Figura 8.1: Gráfico de proporción de varianzas

Las componentes resultantes son:

$$
\begin{aligned}
Y_1 &= -0.1053X_1 + 0.5624X_2 + 0.5336X_3 - 0.1687X_4 - 0.0882X_5 - 0.1456X_6 + 0.5593X_7 \\
&\quad -0.1320X_8 \\
Y_2 &= 0.4071X_1 + 0.1190X_2 + 0.1891X_3 + 0.4900X_4 + 0.4870X_5 + 0.4971X_6 + 0.1718X_7 \\
&\quad +0.1781X_8 \\
&\ldots \\
Y_8 &= 0.7155X_1 + 0.0644X_2 + 0.0539X_3 + 0.5959X_4 - 0.2050X_5 + 0.2404X_6 - 0.1468X_7 \\
&= -0.0666X_8
\end{aligned}
$$

donde: X_1 = Lengua, X_2 = Matemáticas, X_3 = Física, X_4 = Inglés, X_5 = Filosofía, X_6 = Historia, X_7 = Química, X_8 = Educación física

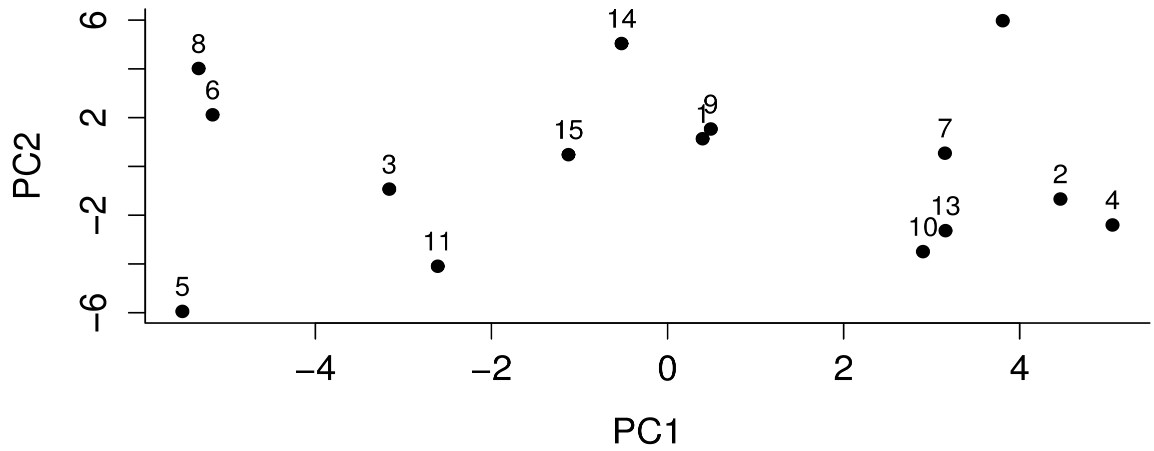


Figura 8.2: Gráfico de las dos primeras CP

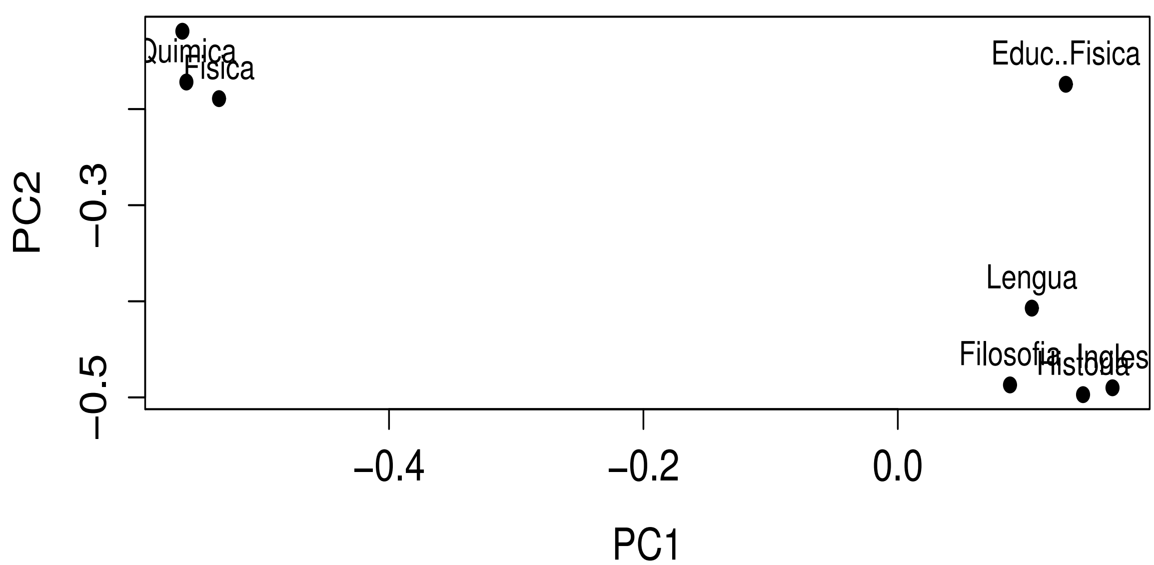


Figura 8.3: Gráfico de correlación de las dos primeras CP

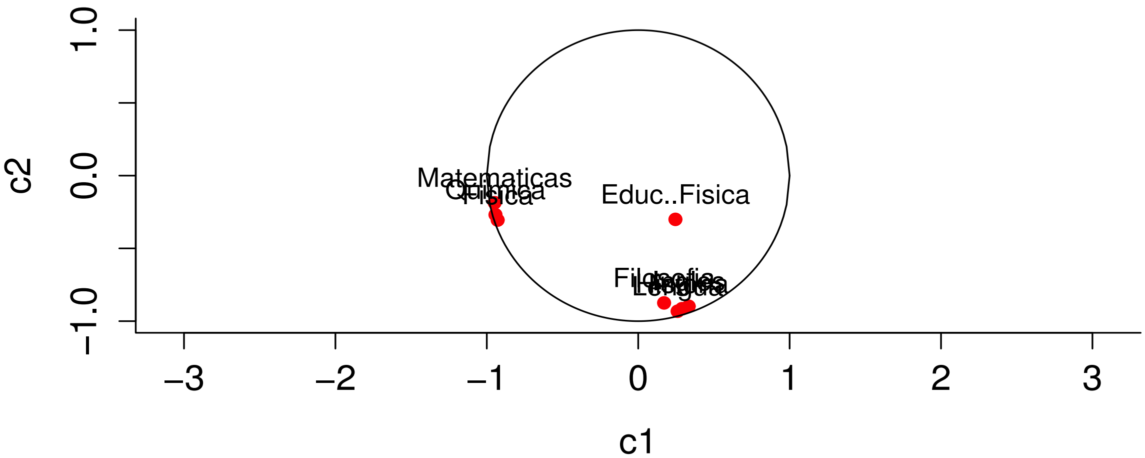


Figura 8.4: Gráfico de correlación de las dos CP con los datos originales

Capítulo 9

Cadenas de Markov

9.1. Ejemplos resueltos

1. Una cadena de Markov con espacio de estados $1, 2, 3$ tiene una matriz de transición

$$P = \begin{vmatrix} 1/3 & 1/3 & 1/3 \\ 0 & 1/2 & 1/2 \\ 0 & 0 & 1 \end{vmatrix}$$

Muestre que el estado 3 es absorvente y, que empezando del estado 1 encuentre el tiempo esperado hasta que la absorción ocurra.

Solución:

Sea $\psi(i)$ el tiempo esperado para llegar al estado 3 empezando del estado i, donde $i = 1, 2, 3$. Se tiene que

$$\begin{aligned} \psi(3) &= 0 \\ \psi(2) &= 1 + \frac{1}{2}\psi(2) + \frac{1}{2}\psi(2) \\ \psi(1) &= 1 + \frac{1}{3}\psi(1) + \frac{1}{3}\psi(2) + \frac{1}{3}\psi(3) \end{aligned}$$

Solucionando el sistema se tiene que

$$\psi(3) = 0, \psi(2) = 2, \psi(1) = 5/2$$

2. Considere la cadena de Markov con matriz de transición

$$P = \begin{vmatrix} 1/2 & 1/3 & 1/6 \\ 3/4 & 1/2 & 1/4 \\ 0 & 1 & 0 \end{vmatrix}$$

   Muestre que es irreducible y aperiódica.

   Solución:

   Dibujando el diagrama de transición correspondiente y observar que hay un camino de cada estado a cualquier otro.Por lo que implica que es irreducible.

3. Una compania̅ de seguros clasifica a sus clientes en tres categorias: insatisfecho, satisfecho, y preferido. Nadie se mueve de insatisfecho a preferido o viceversa in un anõ. 40 % de los clientes el la categoría de insatisfechos se convierten en satisfechos, 30 % de ellos en la categoría de satisfechos se mueven a preferidos, mientras que 10 % se pasan a la categoría de insatisfechos; 20 % de ellos en la categoría de preferidos bajan a la categoría de satisfechos. Dibuje la matriz de transiciones del modelo.

   Solución:

   La matriz de transiciones para esta cadena de Markov con tres estados es como sigue:

$$P = \begin{vmatrix} 0.6 & 0.4 & 0 \\ 0.1 & 0.6 & 0.3 \\ 0 & 0.2 & 0.8 \end{vmatrix}$$

4. Una companñía sin fines de lucro a veces suspende los dividendos. Supongamos que después de que un dividendo ha sido pagado el siguiente será pagado con una probabilidad de 0.9, mientras que despúes de que un dividendo ha sido suspendido el siguiente será suspendido con probabilidad 0.6. Para un recorrido largo cuál es la fracción de dividendos que serán pagados?

   Solución: Se tiene una cadena de Markov con dos estados: 1) Dividendo pagado y 2) Dividendo suspendido. Se dan las siguientes probabilidades de transición:

$$p_{1,1} = 0.9 \quad p_{2,2} = 0.6$$
$$p_{1,2} = 0.1 \quad p_{2,2} = 0.4$$

Sea π la distribución estacionaria. En el largo tiempo la fracción de dividendos que serán pagados es igual a $\pi(1)$. Pero

$$\pi(1) \times 0.1 \quad = \quad \pi(2) \times 0.4$$

y

$$\pi(1) \times +\pi(2) \quad = \quad 1$$

donde $\pi(1) = 4/5$. Así que para un tiempo largo 80 % de los dividendos serán pagados.

9.2. Caminata aleatoria y movimiento Browniano

La caminata aleatoria o paseo aleatorio o camino aleatorio, abreviado en inglés como RW (Random Walks), es una formalización matemática de la trayectoria que resulta de hacer sucesivos pasos aleatorios. Por ejemplo, la ruta trazada por una molécula mientras viaja por un líquido o un gas, el camino que sigue un animal en su búsqueda de comida, el precio de una acción fluctuante y la situación financiera de un jugador pueden tratarse como una caminata aleatoria. El término caminata aleatoria fue introducido por Karl Pearson en 1905.

Código en R

```
# generar n.pasos aleatorios (angulos)
n.pasos <- 10000
theta <- runif(n.pasos, 0, 2*pi) # radianes
salto <- 1
# obtención del número de pasos x y y
dx <- salto * cos(theta)
```

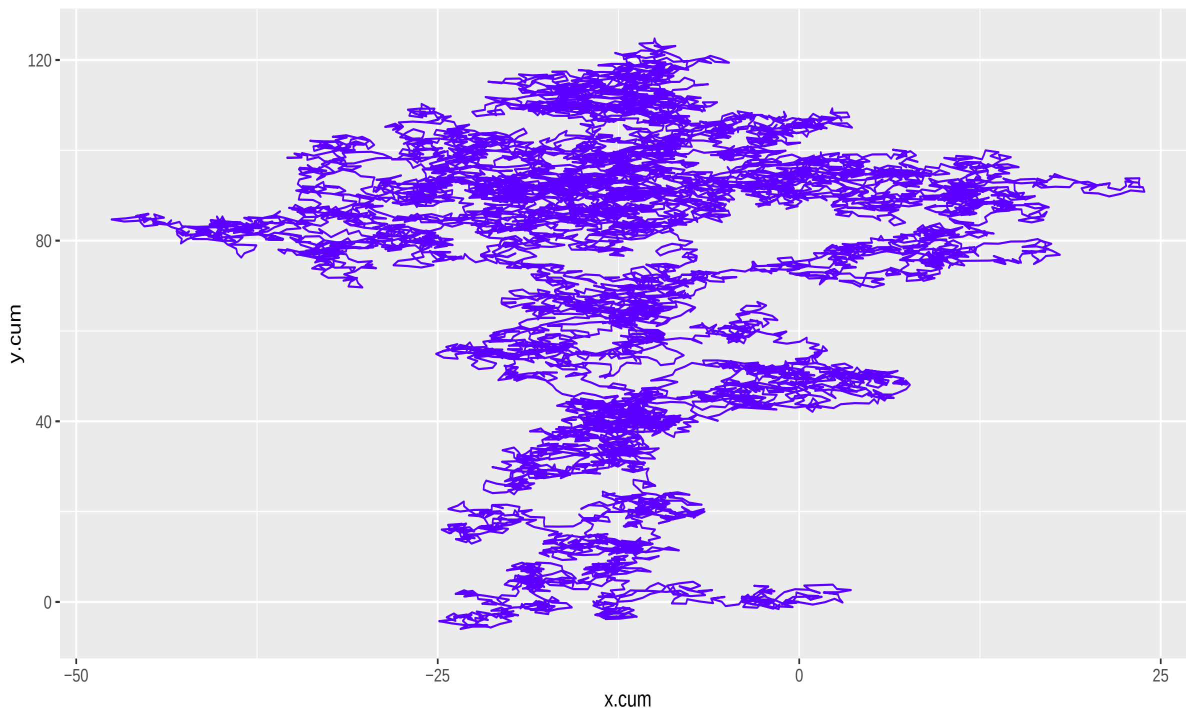


Figura 9.1: Caminata aleatoria

```
dy <- salto * sin(theta)
# calculo de el acumulado
x.cum <- c(0, cumsum(dx))
y.cum <- c(0, cumsum(dy))
# grafico del camino
qplot(x.cum, y.cum, geom="path", main="", col=I("blue"))
```

El movimiento browniano es el movimiento aleatorio que se observa en algunas partículas microscópicas que se hallan en un medio fluido (por ejemplo, polen en una gota de agua). Recibe su nombre en honor al escocés Robert Brown, biólogo y botánico que descubrió este fenómeno en 1827 y observó que pequeñas partículas de polen se desplazaban en movimientos aleatorios sin razón aparente. En 1785, el mismo fenómeno había sido descrito por Jan Ingenhousz sobre partículas de carbón en alcohol.

El segundo artículo de Einstein en 1905, "Sobre el movimiento de partículas pequeñas suspendidas en un líquido estacionario", predecía un fenómeno que había sido detectado hacía setenta años: el movimiento browniano. Este trabajo podría no ser considerado como genial, ya que después de todo, casi todo el mundo podía predecir la existencia del movimiento browniano. La diferencia es que la predicción de Einstein estaba fundamentada matemáticamente en un modelo estadístico que Einstein había desarrollado para analizar el movimiento de las moléculas suspendidas en un líquido. Cualquier otra predicción podría ser la consecuencia de leer un libro sobre Einstein.

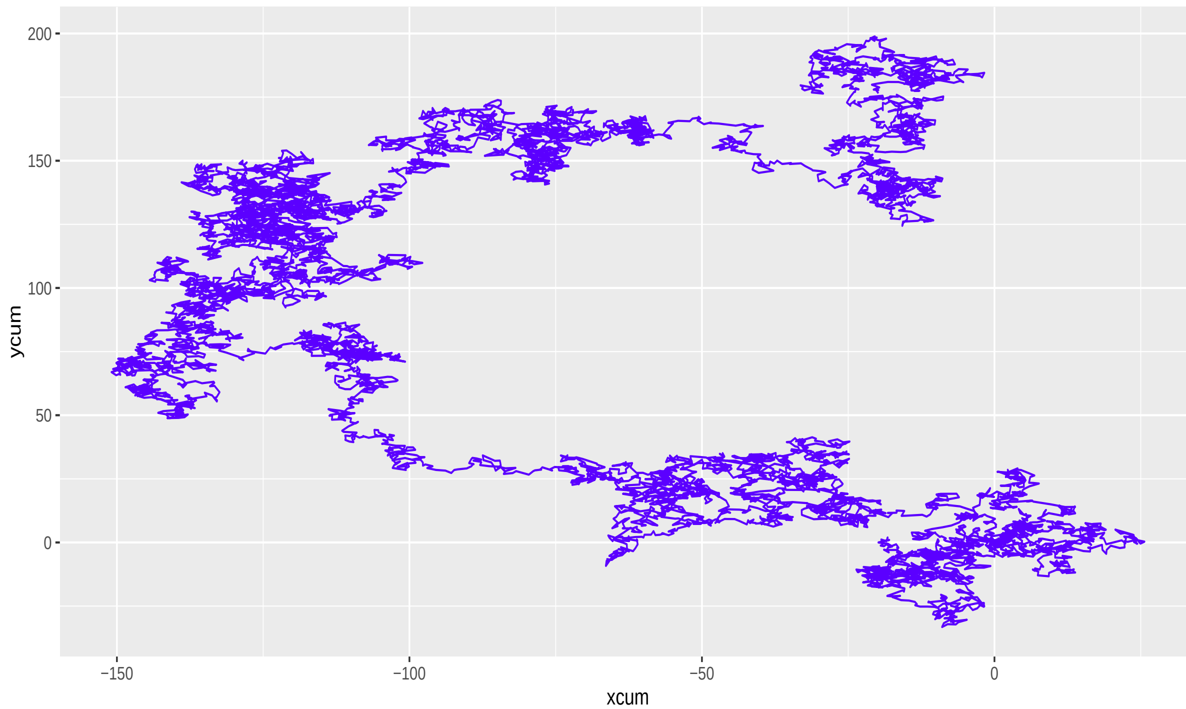


Figura 9.2: Mivimiento Browniano

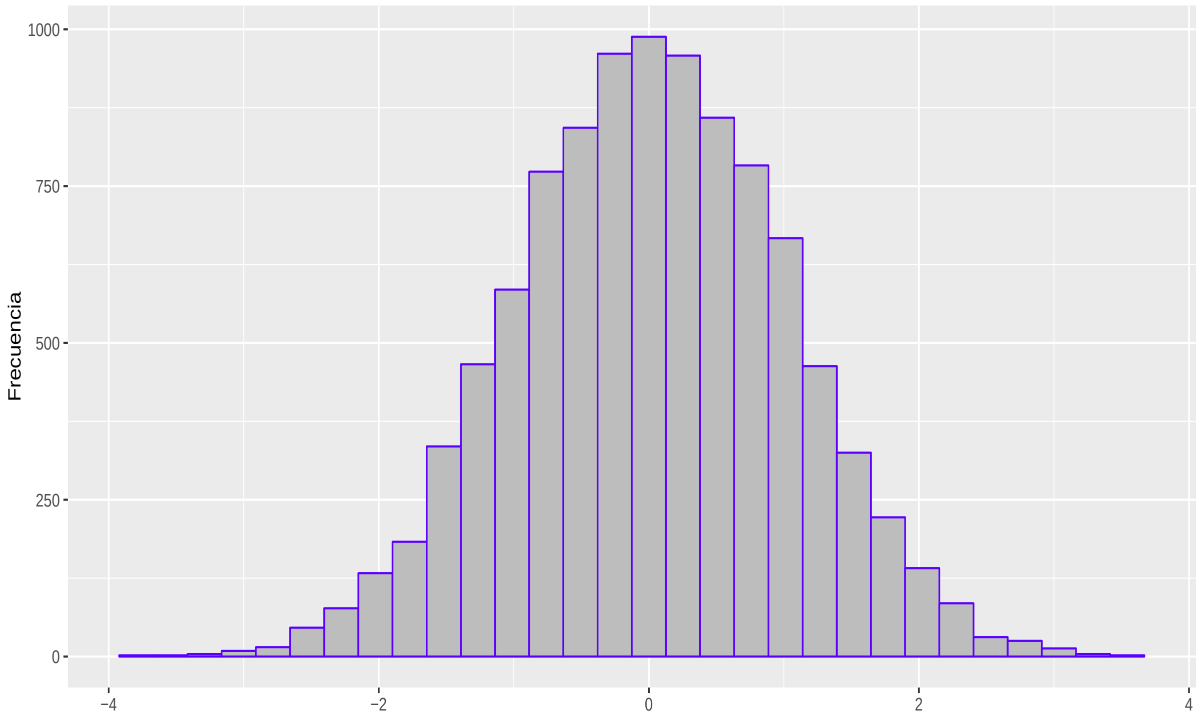


Figura 9.3: Histograma movimiento Browniano

Código en R

```
N <- 10000
xdis = rnorm(N, 0 ,1);
ydis = rnorm(N, 0 ,1);
xcum = cumsum(xdis);
ycum = cumsum(ydis);
qplot(xcum, ycum,geom="path",main="",col=I("blue"));
qplot(ydis,geom="histogram",fill=I("grey"),
      col=I("blue"),xlab="",ylab="Frecuencia")
```

Referencias

Casella, G. and Berger, R. L. (2002). *Statistical inference*, volume 2. Duxbury Pacific Grove, CA.

Degroot, M. H. M. H. (1988). *Probabilidad y estadística.* Number 310/D32pE.

Hogg, R. V. and Craig, A. T. (1978). *introduction to. mathematical stati sties EDITION.*

Johnson, N. L., Kemp, A. W., and Kotz, S. (1994). *Continuous Univariate distributions*, volume 444. John Wiley & Sons.

Johnson, N. L., Kemp, A. W., and Kotz, S. (1995). *Continuous Univariate distributions*, volume 444. John Wiley & Sons.

Montgomery, D. C., Peck, E. A., and Vining, G. G. (2015). *Introduction to linear regression analysis.* John Wiley & Sons.

Mood, A. M., Graybill, F., and Boes, D. C. (1974). *Introduction to theTheory of Statistics.* McGraw-Hill, New York.

Paulos, J. A. and Llosa, J. M. (1990). *El hombre anumérico: el analfabetismo matemático y sus consecuencias.* Tusquets editores.

Peña, D. (2002). *Análisis de datos multivariantes*, volume 24. McGraw-Hill Madrid.

Pickover, C. A. (2009). *The math book: from Pythagoras to the 57th dimension, 250 milestones in the history of mathematics.* Sterling Publishing Company, Inc.

Pliego, F. J. M. and Pérez, L. R.-M. (2006). *Fundamentos de probabilidad.* Editorial Paraninfo.

Rincón, L. (2013). *Introducción a la probabilidad.* Facultad de ciencias UNAM.

Rizzo, M. L. (2007). *Statistical computing with R.* CRC Press.

Ross, S. M. (1985). *Introduction to probability models.* Number 519.2 R826 1985. Academic Press.

Ross, S. M. S. M. (1976). *A first course in probability.*

Walpole, R. E., Myers, R. H., Myers, S. L., and Ye, K. (1993). *Probability and statistics for engineers and scientists*, volume 5. Macmillan New York.

Apéndice A

Matemáticas

A.1. Derivadas

1. Sea $f(x) = e^x - x$ obtenga $f^{'}(x)$
   Solución:

$$
\begin{aligned}
f^{'}(x) &= \frac{d}{dx}e^x - \frac{d}{dx}x \\
&= e^x - 1
\end{aligned}
$$

2. Sea $f(x) = xe^x$, encuentre $f^{'}(x)$
   Solución:

$$
\begin{aligned}
f^{'}(x) &= \frac{d}{dx}e^x - \frac{d}{dx}x \\
&= e^x - 1
\end{aligned}
$$

3. $f(x) = \frac{x^2+x-2}{x^3+6}$
   Solución:
   Aplicando la regla del cociente:

$$\begin{aligned} f^{'}(x) &= \frac{(x^3+6)\frac{d}{dx}(x^2+x-2)-(x^2+x-2)\frac{d}{dx}(x^3-6)}{(x^3+6)^2} \\ &= \frac{(x^3+6)(2x+1)-(x^2+x-2)(3x^2)}{(x^3+6)^2} \\ &= \frac{2x^4+12x+x^3+6-3x^4-3x^3+2}{(x^3+6)^2} \\ &= \frac{-x^4-2x^3+12x+8}{(x^3+6)^2} \end{aligned}$$

4. Sea $f(x) = \frac{e^x}{1+x^2}$
   Solución:

$$\begin{aligned} f^{'}(x) &= \frac{(1+x^2)\frac{d}{dx}(e^x)-(e^x)\frac{d}{dx}(1+x^2)}{(1+x^2)^2} \\ &= \frac{e^x(1+x^2)-e^x(2x)}{(1+x^2)^2} \\ &= \frac{e^x(1-2x+x^2)}{(1+x^2)^2} \\ &= \frac{e^x(1-x)^2}{(1+x^2)^2} \end{aligned}$$

5. Sea $f(x) = \frac{secx}{1+tanx}$
   Solución:

$$\begin{aligned} f^{'}(x) &= \frac{(1+tanx)\frac{d}{dx}(secx)-(secx)\frac{d}{dx}(1+tanx)}{(1+tanx)^2} \\ &= \frac{(1+tanx)(secxtanx)-secxsec^2x}{(1+tanx)^2} \\ &= \frac{secx(tan^2x+tanx-sec^2x)}{(1+tanx)^2} \end{aligned}$$

Soluciones en Mathematica:

1. $f(x) = e^x - x$

```
In[10]:= D[Exp[x] - x, x]
Out[10]= -1 + E^x
```

2. $f(x) = xe^x$

```
In[11]:= D[x*Exp[x], x]
Out[11]= E^x + E^x x
```

3. $f(x) = \frac{x^2+x-2}{x^3+6}$

```
In[12]:= D[(x^2 + x - 2)/(x^3 + 6), x]
Out[12]= -((3 x^2 (-2 + x + x^2))/(6 + x^3)^2) +
(1 + 2 x)/(6 + x^3)
```

4. $f(x) = \frac{e^x}{1+x^2}$

```
In[26]:= D[Exp[x]/(1 + x^2), x]
Out[26]= -((2 E^x x)/(1 + x^2)^2) + E^x/(1 + x^2)
```

5. $f(x) = \frac{secx}{1+tanx}$

```
In[27]:= D[Sec[x]/(1 + Tan[x]), x]
Out[27]= -(Sec[x]^3/(1 + Tan[x])^2) +
(Sec[x] Tan[x])/(1 + Tan[x])
```

A.2. Integrales

A.2 Integrales

1. $\int e^5 x dx$

   Solución:

   Sea $u = 5x$, $du = 5dx$ y $dx = \frac{1}{5}du$

$$\begin{aligned} \int e^{5x} dx &= \frac{1}{5} e^u du \\ &= \frac{1}{5} e^u + C \end{aligned}$$

2. $\int e^{senx} coxdx$

   Solución:

   Sea $u = senx$ entonces $du = cosxdx$, y $dx = \frac{1}{cosx}du$

$$\begin{aligned} \int e^{senx} cosxdx &= \int e^u cosx \left(\frac{1}{5}\right) du \\ &= e^u + C \end{aligned}$$

3. $\int 3xe^{-x^2} dx$

   Solución:

   Sea $u = -x^2$ entonces $du = -2xdx$. Luego

$$\begin{aligned} \int 3xe^{-x^2} dx &= \int -\frac{3}{2} e^u du \\ &= -\frac{3}{2} e^u + C \end{aligned}$$

4. $x^2 e^x dx$

   Solución:

   Sea $u = -x^2$ y $v = \int e^x dx$. Luego aplicando la regla de integración por partes

$$\int x^2 e^x dx = x^2 e^x - \int e^x 2x dx$$

   Sea $u = 2x$ y $v = \int 2x$. Aplicando integración por partes nuevamente

$$\begin{aligned} \int x^2 e^x dx &= x^2 e^x - \left(2xe^x - \int e^x dx\right) \\ &= e^x [x^2 - 2x - 2] + C \end{aligned}$$

5. $\int xsenx$
   Solución:
   Sea $u = x$ y $v = \int senxdx = -cosx$. Luego aplicando la regla de integración por partes

$$\begin{aligned}\int xsenxdx &= -xcosx - \int cosxdx \\ &= -xcosx + senx \\ &= senx - xcosx\end{aligned}$$

Soluciones en Mathematica:

1. $\int e^5 xdx$

```
In[23]:= Integrate[Exp[(5 x)], x]
Out[23]= E^(5 x)/5
```

2. $\int e^{senx} coxdx$

```
In[24]:= Integrate[Exp[Sin[x]]*Cos[x], x]
Out[24]= E^Sin[x]
```

3. $\int 3xe^{-x^2} dx$

```
In[25]:= Integrate[3 x*Exp[-x^2], x]
Out[25]= -((3 E^-x^2)/2)
```

4. $\int x^2 e^x dx$

```
In[26]:= Integrate[x^2*Exp[x], x]
Out[26]= E^x (2 - 2 x + x^2)
```

5. $\int xsenx$

```
In[27]:= Integrate[x*Sin[x], x]
Out[27]= -x Cos[x] + Sin[x]
```

Apéndice B

B.1. Examen 1

Nombre:

Resuelva los siguientes problemas de manera clara y legible, teniendo en cuenta el desarrollo y redacción de la solución. En el caso de las demostraciones que tengan que ver con conjuntos, usar un diagrama de Venn para demostrar algo no es válido.

1. En una promoción de una conocida marca de detergentes, se tiene que juntar en total 5 estampas diferentes que se pueden obtener al comprar una caja de dicho producto. En cada caja, se puede encontrar un sobre con 2 estampas distintas, es decir, la empresa asegura que no hay estampas repetidas en ningún sobre. Si alguien logra juntar las 5 estampas se hace acreedor a un premio. Cuál es la probabilidad de que, al comprar 3 cajas, una persona se haga acreedora del premio que anuncia la compañía de detergentes?

   Solución

   Sea Z el evento de interés: ganar el premio. Como se asegura que en cada paquete de estampas no hay repetidas, se puede concluir que en la primera caja siempre se obtendrán dos estampas diferentes y que no se tenían. En total, existen $\binom{5}{2} = 10$ combinaciones distintas de estampas:

$$(1,2)(2,3)(3,4)(4,5)$$

$$(1,3)(2,4)(3,5)$$

$$(1,4)(2,5)$$

$$(1, 5)$$

Para la segunda caja se pueden obtener tres resultados:

a) 2 estampas nuevas

b) 1 estampa nueva

c) 0 estampas nuevas

Para el caso a), supongamos que tenemos la primera y la segunda estampa. Existen 3 sobres en los que podemos obtener alguna de las dos nuevas, es decir

$$P(A) = \frac{3}{10}$$

Usando el mismo razonamiento, para b),

$$P(B) = \frac{6}{10}$$

Y para c),

$$P(C) = \frac{1}{10}$$

De aquí se puede concluir que si resulta el caso c) ya no puede ocurrir el evento Z.

Si ocurre A, solo se necesita una última estampa con probabilidad $P(D|A) = \frac{4}{10}$. Si resulta B, para ganar el premio se necesita forzosamente el sobre con las dos estampas faltantes con probabilidad $P(E|B) = \frac{1}{10}$.

Finalmente, usando el teorema de probabilidad total (dado que el evento de interés esta en una partición):

$$P(Z) = P(A) \cdot P(D|A) + P(B) \cdot P(E|B) = \frac{3}{10} \cdot \frac{4}{10} + \frac{6}{10} \cdot \frac{1}{10} \approx 0.18$$

2. Una población de electores contiene 40 % de republicanos y 60 % de demócratas. Se publica que 30 % de los republicanos y 70 % de los demócratas están a favor de un tema de elección. Se encuentra que una persona seleccionada al azar de esta población está a favor del tema en cuestión. Encuentre la probabilidad condicional de que esta persona sea un demócrata.

   Solución

   Para este ejercicio nos preguntan la probabilidad condicional $P(\text{Demócrata}|\text{Favor})$,es decir, que sea un demócrata dado que la personas que se eligió esta a favor. Las probabilidades se resumen en el árbol de la Fig. (2.4) obteniendo:

$$P(\text{Demócrata}|\text{Favor}) = \frac{(0.7) \cdot (0.6)}{(0.7) \cdot (0.6) + (0.4) \cdot (0.3)} \approx 0.77$$

Se puede utilizar sloped para que aparezca la probabilidad inclinada. De manera global como en este caso o individual como lo hice en tres ocaciones en el arbol anterior

3. Una prueba de diagnóstico para una enfermedad es tal que (correctamente) detecta la enfermedad en 90 % de los individuos que en realidad tienen la enfermedad. También, si una persona no tiene la enfermedad, la prueba reportará que él o ella no la tiene con probabilidad 0.9. Sólo 1 % de la población tiene la enfermedad en cuestión. Si una persona es seleccionada al azar de la población y la prueba de diagnóstico indica que tiene la enfemerdad, ¿cuál es la probabilidad condicional de que tenga, en realidad, la enfermedad?

   Solución

   Se debe proceder con cautela para resolver este problema (y en general, cualquier otro que se plantee en la vida). El evento que sucede o del que se depende es que la prueba resulta positivo, y por consiguiente, el evento que esta sujeto a que resultó positivo el examen es que este realmente enferma esta persona. Lo antes dicho se puede representar de la siguiente forma:

$$P(\text{Enfermo}|\text{Positivo})$$

   Del árbol de la Fig. (2.5) se obtiene que

$$P(\text{Enfermo}|\text{Positivo}) = \frac{(0.9)\cdot(0.01)}{(0.9)\cdot(0.01)+(0.1)\cdot(0.99)} \approx 0.083 \approx 8.33\,\%$$

4. Un espacio muestral esta formado por cinco eventos simples: E_1, E_2, E_3, E_4 y E_5.

   a) Si $P(E_1) = P(E_2) = 0.15$, $P(E_3) = 0.4$ y $P(E_4) = 2P(E_5)$, halle las probabilidades de E_4 y E_5.

   b) Si $P(E_1) = 3P(E_2) = 0.3$, encuentre las probabilidades de los eventos simples restantes si se sabe que son igualmente probables.

   Solución

   a) Como son eventos simples y forman un espacio muestral se sigue que

$$\sum_{i=1}^{5} P(E_i) = 1$$

Por lo tanto,

$$\begin{aligned}3P(E_5) &= 1 - (P(E_1) + P(E_2) + P(E_3))\\ &= 1 - (0.15 + 0.15 + 0.4)\\ &= 1 - 0.7 = 0.3\end{aligned}$$

Por lo tanto, $P(E_5) = 0.1$ y $P(E_4) = 0.2$.

b) Del mismo modo, se sigue que $P(E_1) = 0.3$ y $P(E_2) = 0.1$ y

$$3P(E_3) = 1 - P(E_1) + P(E_2) = 0.6$$

Por lo tano, $P(E_3) = P(E_4) = P(E_5) = 0.2$.

5. Si A y B son eventos y $B \subset A$:

a) Demuestre que $A = B \cup (A \cap \bar{B})$

b) Demuestre que $P(A) = P(B) + P(A \cap \bar{B})$

Solución

a) Sea x un elemento arbitrario de U. Si $x \in B \cup (A \cap \bar{B})$

$$\begin{aligned}&\iff x \in B \lor x \in (A \cap \bar{B})\\ &\iff x \in B \lor (x \in A \land x \notin b)\end{aligned}$$

Sean

$$A = x \in A$$

$$B = x \in B$$

Sustituyendo,

$$\begin{aligned}&\iff B \lor (A \land \neg B)\\ &\iff (B \lor A) \land (B \lor \neg B)\end{aligned}$$

Transformando otra vez a su forma original

$$\begin{aligned}&\iff (x \in B \lor x \in A) \land (x \in B \lor x \notin B)\\ &\iff x \in A \cap U\\ &\iff x \in A\end{aligned}$$

ya que $B \subset A$. Por lo tanto,

$$A = B \cup (A \cap \bar{B})$$

b) Usando la demostración anterior,

$$\begin{aligned} P(A) &= P(B \cup (A \cap \bar{B})) \\ &= P(B) + P(A \cap \bar{B})) - P(B \cap (A \cap \bar{B})) \\ &= P(B) + P(A \cap \bar{B})) - P((B \cap \bar{B}) \cap A) \\ &= P(B) + P(A \cap \bar{B})) - P(\emptyset \cap A) \\ &= P(B) + P(A \cap \bar{B})) - P(\emptyset) \end{aligned}$$

Por vacuidad, se llega a que

$$P(A) = P(B) + P(A \cap \bar{B})$$

6. Si dos eventos, A y B, son tales que $P(A) = 0.5$, $P(B) = 0.3$ y $P(A \cap B) = 0.1$ encuentre lo siguiente:

   a) $P(A|B)$
   b) $P(B|A)$
   c) $P(A|A \cup B)$
   d) $P(A|A \cap B)$
   e) $P(A \cap B|A \cup B)$

Solución

a) Por definición,

$$P(A|B) = \frac{P(A \cap B}{P(B)}$$

Sustituyendo valores se obtiene que

$$P(A|B) = \frac{0.1}{0.3} = \frac{1}{3}$$

b) Procediendo del mismo modo que en el ejercicio anterior

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

sustituyendo los valores correspondientes

$$P(B|A) = \frac{0.1}{0.5} = \frac{1}{5}$$

c) Volviendo a usar la definición de probabilidad condicional

$$P(A|A \cup B) = \frac{P(A \cap (A \cup B))}{P(A \cup B)}$$

pero sabemos que

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.5 + 0.3 - 0.1 \\ &= 0.7 \end{aligned}$$

Por otro lado,

$$\begin{aligned} P(A \cap (A \cup B)) &= P((A \cap A) \cup (A \cap B)) \\ &= P(A \cup (A \cap B)) \\ &= P(A) + P(A \cap B) - P(A \cap (A \cap B)) \\ &= P(A) + P(A \cap B) - P(A \cap B) \\ &= P(A) \end{aligned}$$

Sustituyendo valores se obtiene que

$$\begin{aligned} P(A|A \cup B) &= \frac{P(A \cap (A \cup B))}{P(A \cup B)} \\ &= \frac{0.5}{0.7} = \frac{5}{7} \end{aligned}$$

d) Usando la definición de probabilidad condicional

$$\begin{aligned} P(A|A \cap B) &= \frac{P(A \cap (A \cap B))}{P(A \cap B)} \\ &= \frac{P(A \cap B)}{P(A \cap B)} \\ &= 1 \end{aligned}$$

e) Por definición

$$\begin{aligned} P(A\cap B|A\cup B) &= \frac{P[(A\cap B)\cap(A\cup B)]}{P(A\cup B)} \\ &= \frac{P(A\cap B)}{P(A\cup B)} \\ &= \frac{0.1}{0.7} \\ &= \frac{1}{7} \end{aligned}$$

7. Si A y B son eventos mutuamente excluyentes y $P(B) > 0$, demuestre que

$$P(A|A\cup B) = \frac{P(A)}{P(A)+P(B)}$$

Solución

Por definición

$$P(A|A\cup B) = \frac{P[A\cap(A\cup B)]}{P(A\cup B)}$$

por otro lado, por hipótesis, se sigue que A y B son eventos mutuamente excluyentes, es decir, se cumple que

$$A\cap B = \emptyset$$

por lo tanto

$$P(A\cup B) = P(A) + P(B)$$

además

$$A\cap(A\cup B) = A$$

como $P(B) > 0$, se sigue que

$$P(A|A\cup B) = \frac{P(A)}{P(A)+P(B)}$$

8. Suponga que dos dados balanceados se lanzan repetidamente y la suma de las dos caras superiores se determina en cada tiro. ¿Cuál es la probabilidad de obtener

   a) una suma de 3 antes de obtener una suma de 7?

   b) una suma de 4 antes de obtener una suma de 7?

Solución

Se tiene que componer el evento mediante eventos simples considerando algunas particularidades. Existen solo dos casos en que puede ocurrir el evento de interés (sacar un 3), por lo que se puede describir el evento como:

$$P(Z) = \frac{2}{36} + \frac{28}{36} \cdot \frac{2}{36} + \frac{28}{36} \cdot \frac{28}{36} \cdot \frac{2}{36} + ...$$

Lo cual se puede interpretar como la probabilidad de que salga una suma de 3 en el primer tiro o bien, que no salga una suma de 3 ni 7 en el primer tiro y salga 3 en el segundo tiro o bien, que no salga una suma de 3 ni 7 en el primer y segundo tiro y salga una suma de 3 en el tercero, de aquí al infinito.

Notemos que la suma anterior es una serie geométrica con $a = \frac{2}{36}$ y $r = 28/36$, es decir

$$S = a\frac{1 - r^n}{1 - r}$$

como $r < 1$, la suma se reduce a

$$S = a\frac{1}{1 - r}$$

sustituyendo valores

$$S = \frac{2}{36} \cdot \frac{1}{1 - 28/36} = \frac{1}{4}$$

por lo que $P(Z) = \frac{1}{4}$

Para el caso b) se realiza el mismo procedimiento con $a = \frac{3}{26}$ y $r = \frac{27}{36}$ para obtener $P(Z) = \frac{1}{3}$.

9. Una fraternidad local está realizando una rifa en la que se han de vender 50 boletos, uno por cliente. Hay tres premios para ser concedidos. Si los cuatro organizadores de la rifa compran un boleto cada uno, ¿cuál es la probabilidad de que los cuatro organizadores ganen

   a) todos los premios?
   b) exactamente dos de los premios?
   c) exactamente uno de los premios?
   d) ninguno de los premios?

   Solución

   a) Para todos los puntos, el espacio consta de elegir 3 boletos de entre 50 posibles, es decir $\binom{50}{3} = 19600$ posibles combinaciones. Para el primero inciso, se tiene que elegir 3 personas organizadoras de entre las 4, es decir, $\binom{4}{3} = 4$ por lo que

$$P(A) = \frac{4}{19600}$$

b) En este caso solo se tienen que elegir 2 personas de las 4 en total, es decir $\binom{4}{2} = 6$ y para el otro premio restante existen $\binom{46}{1} = 46$ maneras de elegir. Por lo tanto

$$P(B) = \frac{6 * 46}{19600} = \frac{276}{19600}$$

c) Del mismo modo que en el ejercicio anterior, existe $\binom{4}{1} = 4$ formas que un integrante gane un premio y $\binom{46}{2} = 1035$ para que 2 de los 46 restantes ganen, por lo tanto

$$P(C) = \frac{4140}{19600}$$

d) Finalmente, se puede utilizar la suma de los eventos anteriores para el complemento o calcular $\binom{46}{3} = 15180$ formas de elegir 3 personas ajenas a los organizadores, por lo tanto

$$P(D) = \frac{15180}{19600}$$

10. En la enfermera del doctor Segura no se puede confiar, pues durante la ausencia del médico la probabilidad de que no le inyecte un suero a un enfermo es de 0.6. Se sabe que si a un enfermo grave se le inyecta el suero tiene igual probabilidad de mejorar que de empeorar, pero si no se le inyecta entonces la probabilidad de que mejore es de 0.25. A su regreso, el Dr. Segura se encuentra con que un enfermo ha empeorado. ¿Cuál es la probabilidad de que la enfermera olvidara inyectar el suero a este paciente?

    Solución

    Utilizando la regla de Bayes, buscamos la probabilidad de que la enfermera no inyectara el suero dado que el enfermo empeoró. Sean I y M los eventos "la enfermera inyectó el sueroz .el paciente mejoró", respectivamente. Se busca calcular $P(\bar{I}|\bar{M})$. Utilizando la regla de Bayes se tiene que

$$\begin{aligned} P(\bar{I}|\bar{M} &= \frac{P(\bar{M}|\bar{I})}{P(\bar{M}|\bar{I}) + P(\bar{M}|I)} \\ &= \frac{0.75 \cdot 0.6}{(0.75 \cdot 0.6) + (0.5 \cdot 0.4)} \\ &\approx .685 \end{aligned}$$

B.2. Paradoja del falso positivo

E n gran número de problemas prácticos, los eventos de mayor interés son aquellos cuya ocurrencia está condicionada a la ocurrencia de otro evento. De aquí que interese introducir el concepto de probabilidad condicional, esto es, la probabilidad condicionada a que haya ocurrido o pudiese ocurrir cierto evento.

La magnitud de este problema es la mejor entendida en términos de probabilidades condicionales.Supongamos un grupo de personas de las que el 1 % sufre una cierta enfermedad, y el resto está bien. Escogiendo un individuo al azar:

$$P(enfermo) = 0.01 \text{ y } P(sano) = 0.99$$

Supongamos que aplicando una prueba a una persona que no tiene la enfermedad, hay una posibilidad del 1 % de conseguir un falso positivo, esto es:

$$P(positivo|sano) = 0.01 \text{ y } P(negativo|sano) = 0.99$$

Finalmente, supongamos que aplicando la prueba a una persona que tiene la enfermedad, hay una posibilidad del 1 % de un falso negativo, esto es:

$$P(negativo|enfermo) = 0.01 \text{ y } P(positivo|enfermo) = 0.99$$

Ahora, uno puede calcular lo siguiente: La fracción de individuos en el grupo que están sanos y dan negativo:

$$\begin{aligned} P(sano \cap negativo) &= P(sano)P(begativo|sano) \\ &= 0.99 * 0.99 = 0.9801 \end{aligned}$$

La fracción de individuos en el grupo que están enfermos y dan positivo:

$$P(enfermo \cap positivo) = P(enfermo)P(positivo|enfermo) = 0.01 * 0.99 = 0.0099$$

La fracción de individuos en el grupo que dan falso positivo:

$$\begin{aligned} P(sano \cap positivo) &= P(sano)P(positivo|sano \\ &= 0.99 * 0.01 = 0.099 \end{aligned}$$

La fracción de individuos en el grupo que dan falso negativo:

$$\begin{aligned} P(enfermo \cap negativo) &= P(enfermo)P(negativo|enfermo) \\ &= 0.01 * 0.01 = 0.001 \end{aligned}$$

Además, la fracción de individuos en el grupo que dan positivo:

$$P(positivo) = P(sano \cap positivo) + P(enfermo \cap positivo) =$$

Finalmente, la probabilidad de que un individuo realmente tenga

$$P(enfermo|positivo) = \frac{P(enfermo|positivo)}{P(positivo)} =$$

Máquina enigma

La máquina Enigma era un dispositivo electromecánico, es decir, tenía una parte eléctrica y otra mecánica. El mecanismo consistía en una serie de teclas, con las letras del alfabeto, al igual que una máquina de escribir, que en realidad eran interruptores que accionaban los dispositivos eléctricos y hacían mover unos cilindros rotatorios.

El funcionamiento, cara al usuario, era bastante sencillo. El operador tenía que teclear las letras de su mensaje y anotar las letras que devolvía la máquina (a través de un alfabeto que se iba iluminando). El código a usar se fijaba con las posiciones de los cilindros que constaban, cada uno, de 26 cables que se conectaban al teclado pero, con la particularidad, que el primer cilindro giraba un veintiseisavo de vuelta después de cada pulsación, de tal manera que la posición de las conexiones iba cambiando con cada entrada del teclado, obteniendo un cifrado polialfabético. Además, para dar mayor robustez, el segundo cilindro sólo daba un giro cuando el primero había completado 26 giros y el tercero cuando el segundo había dado sus correspondientes 26 y añadió la posibilidad de que los rodillos pudiesen ser intercambiados de posición, de manera que el número de posibilidades aumentase hasta tener 105.456 alfabetos.

B.3. Ventaja (Odds)

Los odds (o ventaja) representa que equipo , caballo o atleta tiene una alta probabilidad de ganar. Aunque existen diversas formas de calcularlos, en general es como sigue: Un Odds de $3-5$ indica que tu ganancia será de $3/5$ de un dolar, en otras palabras, por cada 5 de apuesta habrá 3 de ganancia, por ejemplo:

$$Odds(caballo) = \frac{P(ganar)}{P(perder)} = 3:5 \tag{B.1}$$

y probabilidad de ganar es de

$$P(ganeelcaballo) = \frac{3}{3+5} \tag{B.2}$$

Momios Momio, o en inglés Ódd', es el cociente entre la probabilidad de que un evento suceda y la probabilidad de que no suceda. Para efectos prácticos, el momio es la cuota que ofrecemos pagarte por apostar y acertar a un determinado resultado.

Para realizar una apuesta siempre es importante conocer quién es el favorito y quién es el no favorito. El equipo favorito siempre tendrá un momio negativo, aunque existen casos en que ambos equipos tienen un momio negativo (para este caso, el favorito es el que tenga el momio más negativo), o ambos equipos tienen un momio positivo (para este caso, el favorito es el que tenga el momio menos positivo).

1. ¿América -160 (Favorito)
2. Empate $+230$
3. Guadalajara $+380$ (No favorito)

B.4. Cálculo del valor de pi

A lo largo de la historia han sido muchas las formas utilizadas por el ser humano para calcular aproximaciones cada vez más exactas de este número Pi, cociente entre la longitud de una circunferencia cualquiera y el diámetro de la misma: se han usado las áreas de polígonos inscritos y circunscritos a una circunferencia, se han encontrado interesantes aproximaciones numéricas con algunas fracciones sencillas, se han desarrollado series infinitas y productos infinitos de todas las formas que uno pueda imaginar...Vamos, de todo. Pero de entre todos estos métodos hay varios que destacan sobre el resto, y uno de los que más lo hacen es el denominado algoritmo de Chudnovsky.

El algoritmo de Chudnovsky es un algoritmo creado por David Volfovich Chudnovsky y Gregory Volfovich Chudnovsky, hermanos y matemáticos ucranianos nacionalizados estadounidenses, mediante el cual podemos obtener muy buenas aproximaciones del número Pi. Se basa en la siguiente expresión relacionada con el número Pi que encontró Ramanujan:

$$\frac{1}{\pi} = \frac{2\sqrt(2)}{9801}\sum_{k=0}^{\infty}\frac{1103+26390k}{(k!)^4 396^{4k}} \tag{B.3}$$

La expresión del algoritmo de Chudnovsky es la siguiente:

$$\frac{1}{\pi} = 12\sum_{k=0}^{\infty}\frac{(-1)^k(6k)!(13591409+545140134k)}{(3k)!(k!)^3 640320^{3k+3/2}} \tag{B.4}$$

Código en R

```
#  Calculo del valor de pi
```

```
options(digits=20)

# Algoritmo de Ramanujan
Ramanujan <- function(k){
  cero <- 1103
  for(i in 1:k){
    uno <- factorial(4*i)*(1103 + 26390*i)/((factorial(i)^4)*396^(4*i))

  }
  dos <- sum(uno) + cero
  da <- (2*sqrt(2)/9801)*dos
  return(1/da)
}
Ramanujan(20)

# Algoritmo de Chudnovsky
Chudnovsky <- function(k){
  cero <- 1*(13591409)/(640320^(3/2))
  for(i in 1:k){
    uno <- ((-1)^i*factorial(6*i)*(13591409 + 545140134*i))/
    ((factorial(3*i)*(factorial(i)^3)*640320^(3*i+3/2)))
  }
  dos <- 12*(sum(uno)+cero)
  return(1/dos)
}
Chudnovsky(20)
```

Mathematica muestra una aproximación del número Pi con 50 decimales. Evidentemente, podemos aumentar el número de decimales para conseguir aproximaciones cada vez más exactas. Mathematica emplea el algoritmo de Chudnovsky para la aproximación de Pi.

En las siguientes lineas describimos las fórmulas que a lo largo de la historia se emplearon para aproximar el valor de Pi

1. Por Arquímedes

$$3\frac{10}{71} < \pi < \frac{10}{70}$$

2. Por Francois Viete

$$\frac{2}{\pi} = \sqrt{\frac{1}{2}}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{2}}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{\frac{1}{2}+\frac{1}{2}\sqrt{2}}}...$$

3. Por Jhon Wallis

$$\frac{\pi}{2} = \frac{2}{1}\frac{2}{3}\frac{4}{3}\frac{4}{3}\frac{6}{5}\frac{6}{5}\frac{8}{7}\frac{8}{7}...$$

4. Por Isaac Newton

$$\frac{\pi}{6} = \frac{1}{2} + \frac{1}{2*3}\left(\frac{1}{2}\right)^3 + \frac{1*3}{2*4*5}\left(\frac{1}{2}\right)^5 + \frac{1*3*5}{2*4*6*7}\left(\frac{1}{2}\right)^7 + ...$$

5. Por Gottfried Leibnitz

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \frac{1}{9} - \frac{1}{11} + ...$$

6. Por Leonard Euler

$$\frac{\pi^2}{6} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \frac{1}{5^2} + ...$$

7. Por Machin

$$\pi = 16\ arctan(1/5) - 4\arctan(1/239)$$

B.5. Econofísica

La econofísica es un novedoso campo de investigación científica que aplica teorías y métodos, originalmente desarrollados por físicos, para entender y resolver problemas en la economía y, especialmente, aquellos que involucran aspectos estocásticos y de dinámica no lineal.

Ejemplos de econofísica incluyen el uso de la teoría de la percolación para explicar fluctuaciones en los mercados, el uso de modelos de infarto cardíaco, criticalidad autorganizada y dinámica de placas tectónicas para explicar las caídas en las bolsas de valores. La Econofísica se preocupa por explicar fenómenos de escalamiento y autosimilares como las leyes de potencias en la distribución de la riqueza. Otro problema de la Econofísica, es el estudio de la existencia de caos determinista en los patrones de transacciones económicas y sus horizontes de predicción temporal.

La econofísica surgió en los años 1990, principalmente en el entorno del prestigiado Instituto Santa Fe de Nuevo México, que se especializa al estudio de los sistemas complejos. Uno de los principales exponentes de la Econofísica es Brian Arthur, quien acuñó el término economía adaptativa para denominar sistemas económicos formados por un número grande de agentes que realizan transacciones de tipo económico. El mejor ejemplo se conoce como el problema del bar .El Farol". Aparentemente, fue el profesor de física de la Universidad de Boston Eugene Stanley, el primero en llamar así a esta disciplina.

Habiendo discutido a muy grandes rasgos de lo que se trata la economía, en qué se basa y cuáles son sus puntos débiles, tal vez sea más fácil darnos cuenta en qué pueden contribuir otras áreas de la ciencia, en especial la física.

El interés de los físicos en economía en general, y sobre todo en tratar de entender y describir los mercados financieros, despega con la disponibilidad de enormes cantidades de datos financieros en los $80's$; cuando muchos físicos y matemáticos fueron contratados en Wall Street para analizar el mercado.

Por ejemplo fue Daniel Bernoulli, el descubridor del principio de Bernoulli en hidrodinámica, quien introduce el concepto de utilidad para describir las preferencias de la gente. De manera contrapuesta, a los economistas les gusta hacer notar que en el contexto de finanzas, Louis Bachelier resolvió el problema matemático del movimiento Browniano antes que Einstein. Específicamente, Bachelier, desarrolló un modelo de caminata aleatoria para describir la dinámica de los precios. Si bien el trabajo de Bachelier parece haber pasado desapercibido hasta muchos añoos más tarde que resurgió el modelo de caminata aleatoria en relación con la hipótesis de mercados eficientes, esta línea de modelación desemboca en la fórmula de Black y Scholes, que como habíamos mencionado, permite asignar un precio, y por lo tanto abrir el mercado de las opciones. La contraparte física, el movimiento browniano, y su corroboración experimental, que también mereció un Premio Nobel, terminaron de disipar las dudas respecto a la naturaleza molecular de la materia.

B.6. Análisis ROC

La curva ROC se comenzó a utilizar durante la Segunda Guerra Mundial para el análisis de señales de radar, a partir de lo cual se desarrolló la Teoría de Detección de Señales. Después del ataque a Pearl Harbor en 1941, el ejército de los Estados Unidos comenzó un programa de investigación para detectar correctamente los aparatos japoneses a partir de sus señales de radar.

En los años $50's$, las curvas ROC se utilizaron en Psicofísica para evaluar la capacidad de detección de humanos (y también de no humanos) en señales débiles. En medicina el análisis ROC se ha utilizado de forma muy extensa en epidemiología e investigación médica, de tal modo que se encuentra muy relacionado con la Medicina basada en la evidencia. En Radiología, el análisis ROC es la técnica de preferencia para evaluar nuevas técnicas de diagnóstico por imagen.

Más recientemente, las curvas ROC se han mostrado muy útiles para la evaluación de técnicas de aprendizaje automático. La primera aplicación de las ROC en esta área fue por Spackman, quien demostró el valor de las curvas ROC para la comparación de diferentes algoritmos de clasificación.

En la Teoría de detección de señales una curva ROC (acrónimo de Receiver Operating Characteristic, o Característica Operativa del Receptor) es una representación gráfica de la sensibilidad frente a (1 ˘ especificidad) para un sistema clasificador binario según se varía el umbral de discriminación. Otra interpretación de este gráfico es la representación de la razón o ratio de verdaderos positivos (VPR = Razón de Verdaderos Positivos) frente a la razón o ratio de falsos positivos (FPR = Razón de Falsos Positivos) también según se varía el umbral de discriminación (valor a partir del cual decidimos que un caso es un positivo). ROC también puede significar Relative Operating Characteristic (Característica Operativa Relativa) porque es una comparación de dos características operativas (VPR y FPR) según cambiamos el umbral para la decisión.

El análisis de la curva ROC, o simplemente análisis ROC, proporciona herramientas para seleccionar los modelos posiblemente óptimos y descartar modelos subóptimos independientemente de (y antes de especificar) el coste de la distribución de las dos clases sobre las que se decide. La curva ROC es también independiente de la distribución de las clases en la población (en diagnóstico, la prevalencia de una enfermedad en la población). El análisis ROC se relaciona de forma directa y natural con el análisis de coste/beneficio en toma de decisiones diagnósticas.

La curva ROC se desarrolló por ingenieros eléctricos para medir la eficacia en la detección de objetos enemigos en campos de batalla mediante pantallas de radar, a partir de lo cual se desarrolla la Teoría de Detección de Señales (TDS). El análisis ROC se aplicó posteriormente en medicina, radiología, psicología y otras áreas durante varias décadas. Sólo recientemente ha encontrado aplicación en áreas como aprendizaje automático (o machine learning en inglés), y minería de datos (data mining en inglés).

La tabla de contingencia puede proporcionar varias medidas de evaluación (ver caja de terminología). Para dibujar una curva ROC sólo son necesarias las razones de Verdaderos

Positivos (VPR) y de falsos positivos (FPR). La VPR mide hasta qué punto un clasificador o prueba diagnóstica es capaz de detectar o clasificar los casos positivos correctamente, de entre todos los casos positivos disponibles durante la prueba. La FPR define cuántos resultados positivos son incorrectos de entre todos los casos negativos disponibles durante la prueba.

Un espacio ROC se define por FPR y VPR como ejes x e y respectivamente, y representa los intercambios entre verdaderos positivos (en principio, beneficios) y falsos positivos (en principio, costes). Dado que VPR es equivalente a sensibilidad y FPR es igual a 1-especificidad, el gráfico ROC también es conocido como la representación de sensibilidad frente a (1-especificidad). Cada resultado de predicción o instancia de la matriz de confusión representa un punto en el espacio ROC.

B.7. Probabilidad y coincidencia

Una de las principales características de las personas anuméricas es la tendencia a sobrestimar la frecuencia de las coincidencias. Generalmente dan mucha importancia a todo tipo de correspondencias, y, en cambio, dan muy poca a evidencias estadísticas menos relumbrantes, pero absolutamente concluyentes. El primer problema que se plantea es el halla el número de personas que se debe tener para que al seleccionar al azar dos de ellas, la probabilidad de que hallan nacido el mismo día sea el 0.5.

El segundo problema que se plantea es el del cartero. El ejemplo del cartero que ha de distribuir veintiuna cartas entre veinte buzones nos permitirá ilustrar un principio numérico que a veces sirve para explicar la certeza de un determinado tipo de coincidencias. Como 21 es mayor que 20, puede estar seguro, sin necesidad de mirar previamente las direcciones, que por lo menos uno de los buzones tendrá más de una carta. Este principio de sentido común, que se conoce a veces como principio del casillero o de los cajones de Dirichlet, puede servir a veces para llegar a conclusiones que no son tan obvias.

El tercer problema que se plantea es el del asesor bursátil. Si durante seis semanas seguidas recibieras por correo las predicciones de un asesor de bolsa acerca de cierto índice del mercado de valores y las seis fueran acertadas, ¿estarías dispuesto a pagar por recibir la séptima predicción?. Supongamos uno que se hace pasar por asesor financiero imprime un logotipo en papel de lujo y envía $32,000$ cartas a otros tantos inversores potenciales en un cierto valor de la bolsa. Las cartas hablan del elaborado sistema informático de su compañía, de su experiencia financiera y de sus contactos. En $16,000$ de las cartas predice que las acciones subirán y, en las otras $16,000$, que bajarán. Tanto si suben las acciones como si bajan, envía una segunda carta pero sólo a las $16,000$ personas que recibieron la «predicción» correcta. En $8,000$ de ellas, se predice un alza para la semana siguiente, y en las $8,000$ restantes, una caída. Ocurra lo que ocurra, $8,000$ personas habrán recibido ya dos predicciones acertadas. Manda una tercera tanda de cartas, ahora sólo a estas $8,000$ personas, con una nueva predicción de la evolución del valor para la semana siguiente: $4,000$ predicen un alza y $4,000$ una caída. Pase lo que pase, $4,000$ personas habrán recibido tres predicciones acertadas seguidas. Sigue así unas cuantas veces más, hasta que 500 personas han recibido seis predicciones correctas seguidas. En la siguiente carta se les recuerda esto y se les dice que para seguir recibiendo una información tan valiosa por séptima vez habrán de aportar 500 dólares. Por lo anterior, puede concluirse que El puro azar siempre deja lugar a una cantidad suficiente de aciertos que permiten justificar casi cualquier cosa a alguien predispuesto a creer.

El cuarto planteamiento que se hace es el referente al valor promedio. Por ejemplo, el valor esperado de una cantidad es la media de los valores que toma, pesados según sus probabilidades respectivas. Por ejemplo, si $1/4$ de las veces la cantidad vale 2, $1/3$ vale 6, otro $1/3$ de las veces vale 15 y el $1/12$ restante vale 54, el valor esperado de dicha magnitud es 12. En efecto,

$$2(1/4) + 6(1/3) + 15(1/3) + 54(1/12) = 12$$

Un caso interesante es que la mayoría de la gente no se da cuenta de que los sucesos aleatorios pueden presentar una apariencia completamente ordenada. Los grupos, series y pautas que presentan las sucesiones aleatorias son hasta cierto punto predecibles. Las sucesiones de caras y cruces de una longitud dada, pongamos veinte tiradas, tienen generalmente cierto número de series de caras consecutivas.

Este último caso que acabamos de considerar no tiene nada de sorprendente, pero el mismo tipo de descripción probabilística (empleando unas matemáticas ligeramente más difíciles que la distribución binomial) se puede aplicar a los acontecimientos muy raros. El número de accidentes anuales en un cruce concreto, el número de aguaceros anuales que caen en un desierto determinado, el número de casos de leucemia en una comarca dada, el número de muertes anuales por coz de caballo en ciertos regimientos de caballería del ejército prusiano, etcétera, todos estos casos han sido descritos con gran precisión usando la distribución de probabilidad de Poisson.

B.8. Diferencia entre antinomia y paradoja

Una paradoja es una declaración en apariencia verdadera que conlleva a una auto-contradicción lógica o a una situación que contradice el sentido común. En palabras simples, una paradoja es lo opuesto a lo que uno considera cierto. Del latín paradoxum, pero es encontrada también en textos griegos como paradoxa. Se encuentra compuesta por el prefijo para, que significa çontrario a.º .ªlterado", en conjunción con el sufijo doxa, que significa .ºpinión".

El término antinomia (anti=contra; nomos= norma, ley) , se refiere a todo conflicto entre ideas, proposiciones, actitudes. Son paradojas que alcanzan un resultado que se autocontradice, aplicando correctamente modos aceptados de razonamiento. Muestran fallos en un modo de razón, axioma o definición previamente aceptados.

Ejemplos:

1. Paradoja del hotel infinito: un hotel de infinitas habitaciones puede aceptar más huéspedes, incluso si está lleno.
2. “Es una persona tan pobre que lo único que tiene es mucho dinero”.
3. La paradoja de Ruseel ¿Existe un conjunto de todos los conjuntos que no se contienen a sí mismos?.
4. La paradoja de Zenón, sobre auiles y la tortuga.
5. La paradoja de los gemelos.

B.9. Estadística Bayesiana

La estadística bayesiana es una alternativa a la estadística clásica para la solución de problemas típicos estadísticos como son: estimación, contraste de hipótesis y predicción. Ha generado un enorme interés en los últimos 20 años y ha tenido una gran aceptación en muchas áreas de la investigación científica. La estadística bayesiana, parte del hecho de que toda forma de incertidumbre debe describirse por medio de modelos de probabilidad, y además, la probabilidad es el único lenguaje posible para describir una lógica que trata con todos los niveles de incertidumbre, y no sólo con los extremos de verdad o falsedad.

La teoría bayesiana plantea la solución a un problema estadístico desde el punto de vista subjetivo de la probabilidad, según el cual, la probabilidad de que un estadístico asigne a uno de los posibles resultados de un proceso, representa su propio juicio sobre la verosimilitud de que se tenga el resultado. Este juicio estará basado en opiniones e información acerca del proceso. Esta última también es una desventaja, pues algunos investigadores rechazan que la información inicial se incluya en un proceso de inferencia científica. Pero esta situación se puede evitar estableciendo una distribución a priori no informativa o de referencia, la cual se introduce cuando no se posee mucha información previa acerca del problema.

Cuando un investigador tiene conocimiento previo a un problema, éste conocimiento previo puede cuantificarse en un modelo de probabilidad. Si los juicios de una persona sobre la verosimilitud relativa a ciertas combinaciones de resultados satisfacen ciertas condiciones de consistencia, se puede decir que sus probabilidades subjetivas se determinan de manera única.

Una desventaja, es que algunos investigadores rechazan que la información inicial se incluya en un proceso de inferencia científica. Pero esta situación se puede evitar estableciendo una distribución a priori no informativa o de referencia, la cual se introduce cuando no se posee mucha información previa acerca del problema. A un problema específico se le puede asignar cualquier tipo de distribución a priori, ya que finalmente al actualizar la información a priori que se tenga acerca del parámetro, mediante el teorema de Bayes y obtener la distribución a posteriori del parámetro, es con esta con las que se hacen las inferencias del mismo.

B.10. Frecuencia absoluta, relativa y acumulada

Definición B.1 (Frecuencia absoluta) *La frecuencia absoluta puede definirse como el número de veces que aparece ese valor en el estudio. Se suele denotar por F_i a la frecuencia absoluta del valor $X = x_i$ de la variable X. Dada una muestra de N elementos, la suma de todas las frecuencias absolutas debe dar el total de la muestra estudiada N.*

Definición B.2 (Frecuencia relativa) *Frecuencia relativa f_i, puede definirse como el cociente entre la frecuencia absoluta y el tamaño de la muestra N. Es decir,*

$$f_i = \frac{n_i}{N} \tag{B.5}$$

Definición B.3 (Frecuencia acumulada) *La frecuencia acumulada es la suma de las* *frecuencias absolutas de todos los valores inferiores o iguales al valor considerado.*

$$F_i = \frac{N_i}{N} \tag{B.6}$$

Definición B.4 (Riesgo) *De acuero al diccionario riesgo significa posbilidad de que suceda un peligro o contingencia. La definición varia un poco según el contexto, por ejemplo en finanzas el riesgo significa la posibilidad de posibles perdidas al invertir*

Definición B.5 (peligrosidad) *De acuerodo al diccionario peligrosidad se define como algo que es arriesgasdo o inseguro.*

Definición B.6 (amenaza) *De a cuerdo al diccionario, amenaza es sinónimo de peligro, acción de amenazar.*

Definición B.7 (vulnerabilidad) *Según el diccionario vulnerabilidad es carácter de lo que atacable, o suceptible.*

La palabra vulnerable es de origen latín *vulnerabilis*, una palabra formada por *vulnus* que significa *herida* y el sufijo *abilis* que expresa *posibilidad*, por lo tanto, es la posibilidad de ser herido.

B.11. El hombre anumérico (reseña)

El libro está redactado como si fuera una pequeña recopilación de artículos independientes con un fondo común. Esto es, se puede leer indistintamente un apartado de cada capítulo sin miedo a perdernos por no haber leído los anteriores, ya que, en general, es totalmente independiente.

El autor hace enfásis especialmente y mediante varios ejemplos en la facilidad de manejar grandes números mediante el uso de varias técnicas numéricas, extrañándose de la falta de rigor en la utilización de estos números por parte de los medios de comunicación o su uso para buscar un efecto más efectivo. También nos inicia en el campo de las probabilidades con las técnicas necesarias para calcular la posibilidad de que algo pueda suceder en un conjunto de acciones.

El autor nos da a conocer aquí algunas situaciones en las cuáles la falta de cultura matemática pueden afectar a nuestras relaciones. Esto puede ser visto en varios aspectos, como son la toma de decisiones ante un negocio (¿Merece la pena correr el riesgo, quizás se debería hacer un estudio de mercado?), utiliza la paradoja de los presos para hacernos ver como a pesar de que ante algunas decisiones con respecto a algún negocio con otra persona las matemáticas nos demuestran que nos interesa ser "legal" y permanecer fiel a nuestro compromiso, buscamos razones para asegurarnos un mínimo y al final salimos perdiendo por ser individualistas.

Todas esas encuestas y estudios sobre un conjunto de personas deben seguir unas pautas y ser consideradas en su justa medida, con las medidas correctoras pertinentes; si no lo hacemos así corremos el riesgo de llegar a conclusiones erróneas, que pueden llevarnos a tomar decisiones también erróneas. Además, no solamente hay que medir las probabilidades de que haya una mejora tomando una decisión, sino que también ha de medirse la situación que se produce si no se toma esa decisión. Para mayor referencia consultar Paulos and Llosa (1990).

B.12. Mendel y el fraude científico

Posiblemente, el más sorprendente ejemplo de fraude científico hace referencia al monje astriaco Gregor Mendel (1822 – 1884), al que se le considera el fundador de la moderna teoría genética. En 1865, Mendel publicó un artículo en donde presentaba una serie de experimentos llevados a cabo con guisantes de jardín. Uno de dichos experimentos afectaba al color (verde o amarillo) de las semillas de dichos guisantes. Mendel comenzó su experimento produciendo guisantes amarillos de raza pura. Finalmente, Mendel cruzó guisantes amarillos puros con guisantes verdes puros. Las semillas resultantes de este cruce, conocidas como semillas híbridas de primera generación, resultaron ser siempre amarillas. Es decir no resultó ninguna semilla verde en ese generación.

Tras ello Mendel crizó semillas de esa primera generación, para obtener semillas de la segunda generación. Sorprendentemente, las semillas verdes reaparecieron aquí. De hecho aproximadamente un 25 % de las semillas de la segunda generación fueron verdes y un 75 % restantes, amarillas.

En 1936 Ronald A. Fisher publicó un artículo en el que se analizaban los datos de Mendel y se concluía que dichos datos se ajustaban demasiado bien a su teoría como para que pudieran ser debidos al azar. Fisher uso el contraste de bondad de ajuste que había desarrollado Karl Pearson, para demostrar que la probabilidad de que los datos completos del experimento de Mendel se ajustarán a su teoría, al menos como tan exactamente Mendel había indicado, era igual a 0.00004, Ross (1976).

B.12.1. Proceso estocástico

Un proceso estocástico es una colección o familia de variables aleatorias $\{X_t, t\varepsilon T\}$ ordenadas según el subíndice t, que en general se suelen identificar con el tiempo.

Por tanto, para cada instante t tendremos una variable aleatoria distinta representada por Xt, con lo que un proceso estocástico puede interpretarse como una sucesión de variables aleatorias cuyas características pueden variar a lo largo del tiempo.

A los posibles valores que puede tomar la variable aleatoria se le denominaran estados, por lo que se puede tener un espacio de estados discreto y un espacio de estados continuo. Por otro lado, la variable tiempo puede ser de tipo discreto o de tipo continuo. En el caso del tiempo discreto se podría tomar como ejemplo que los cambios de estado ocurran cada día, cada mes, cada año, etc.. En el caso del tiempo continuo, los cambios de estado se podrían realizar en cualquier instante.

Por tanto, dependiendo de cómo sea el conjunto de subíndices T y el tipo de variable aleatoria dado por Xt se puede establecer la siguiente clasificación de los procesos estocásticos:

- Si el conjunto T es continuo, diremos que X_t es un proceso estocástico de parámetro continuo.
- Si por el contrario T es discreto diremos que nos encontramos frente a un proceso estocástico de parámetro discreto.
- Si para cada instante t la variable aleatoria X_t es de tipo continuo, diremos que el proceso estocástico es de estado continuo.
- Si para cada instante t la variable aleatoria X_t es de tipo discreto, diremos que el proceso estocástico es de estado discreto.