

Breve introducción a la Teoría de la Probabilidad

Asignatura de Física Estadística
3º Física, UGR

14 de febrero de 2017

Índice

1. Introducción	2
2. El concepto de probabilidad	2
2.1. Definiciones	2
2.2. Axiomas de Kolmogórov	3
2.2.1. Teoremas importantes sobre probabilidad	3
2.3. Probabilidad condicional	4
2.4. Eventos independientes	4
2.5. Combinatoria	4
2.5.1. Permutaciones	5
2.5.2. Combinaciones	6
2.6. Aproximación de Stirling	6
3. Variables aleatorias y distribución de probabilidad	7
3.1. Distribuciones discretas	7
3.1.1. Función de distribución	8
3.2. Distribuciones continuas	9
3.3. Cambio de variables	10
4. Momentos y cumulantes	11
4.1. Función generadora de momentos	12
4.2. Función generadora de cumulantes	13
5. Distribuciones especiales	14
5.1. Distribución binomial	14
5.2. Distribución de Poisson	15
5.3. Distribución normal	16
5.4. Ley de los grandes números	16
5.5. Teorema del Límite Central	17

1. Introducción

Introduciremos en primer lugar algunos conceptos de teoría de probabilidad básica. A un conjunto S que contenga todos los resultados posibles de un experimento aleatorio, lo denominaremos *espacio muestral*. Si un espacio muestral tiene un número finito de puntos, se llama *espacio muestral finito*. Si tiene tantos puntos como los números naturales, *espacio muestral infinito contable* y si pertenece a la recta real, por ejemplo, *espacio muestral infinito no contable*. En general, un espacio muestral finito o infinito contable se denomina *espacio muestral discreto*, mientras que uno infinito no contable se llama *espacio muestral no discreto*.

Un evento es un subconjunto A del espacio muestral S , es decir, un conjunto de resultados posibles. Como eventos particulares tenemos al mismo S , el cual es el evento seguro, y el conjunto vacío \emptyset , que se denomina evento imposible porque un elemento de \emptyset no puede ocurrir.

Empleando operaciones de conjuntos sobre eventos en S , podemos obtener eventos nuevos. Por ejemplo, consideremos A y B eventos, entonces,

1. $A \cup B$ es el evento “A o B”. Se denomina unión.
2. $A \cap B$ es el evento “A y B”. Se denomina intersección.
3. A' es el evento “no A”. Se denomina complemento o negación de A .

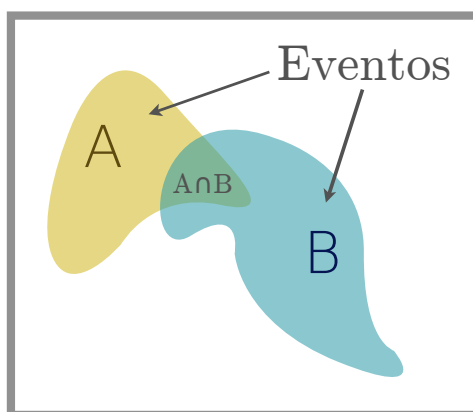


Figura 1: Conjuntos

2. El concepto de probabilidad

El primer concepto que debemos introducir es el de probabilidad, que podemos definir como una medida cuantitativa de la posibilidad, u oportunidad, con que esperamos que ocurra cierto evento. Vamos a ver cómo podemos llevar estos conceptos a una teoría matemática. Para ello, hay dos procedimientos importantes mediante los cuales podemos calcular la probabilidad de un evento.

2.1. Definiciones

Definición 2.1. Clásica: Si un evento puede ocurrir de h maneras diferentes de un número total de n maneras posibles, todos ellos son igualmente posibles. Entonces, la probabilidad del evento es $\frac{h}{n}$.

Definición 2.2. Frecuentista: Si después de n repeticiones de un experimento, donde n es muy grande, se observa que un evento ocurre h veces, entonces la probabilidad de dicho evento es $\frac{h}{n}$. Esto también se denomina la *probabilidad empírica* de un evento.

Ambos enfoques presentan inconvenientes, el primero porque “igualmente probable” es una definición vaga y el segundo porque un “número grande” también es un concepto ambiguo. Debido a estas dificultades, se propone el *enfoque axiomático* de la probabilidad.

2.2. Axiomas de Kolmogórov

Axioma 1. La probabilidad S es no negativa y menor o igual que 1.

$$0 \leq p(S) \leq 1$$

Axioma 2. La probabilidad del evento seguro, S , es igual a 1, denotado simbólicamente,

$$P(S) = 1$$

Axioma 3. Si A_1, A_2, \dots son eventos mutuamente excluyentes (incompatibles dos a dos), entonces:

$$P(A_1 \cup A_2 \cup \dots) = \sum P(A_i)$$

y en particular para el caso de dos eventos,

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

2.2.1. Teoremas importantes sobre probabilidad

Teorema 2.1. Para todo evento A ,

$$0 \leq P(A) \leq 1$$

es decir, la probabilidad está entre 0 y 1.

Teorema 2.2. El evento imposible tiene probabilidad cero, $P(\emptyset) = 0$.

Teorema 2.3. Si A' es el complemento de A , entonces

$$P(A') = 1 - P(A)$$

Teorema 2.4. Si A y B son dos eventos cualesquiera, entonces,

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

Teorema 2.5. Para cualesquiera eventos A y B ,

$$P(A) = P(A \cap B) + P(A \cap B')$$

2.3. Probabilidad condicional

Sean A y B dos eventos tales que $P(A) > 0$. Consideremos $P(A|B)$, la probabilidad de que ocurra un evento A , sabiendo que también sucede otro evento B . La probabilidad condicional se escribe $P(A|B)$, y se lee “la probabilidad de A dado B ”.

No tiene por qué haber una relación causal o temporal entre A y B . A puede preceder en el tiempo a B , sucederlo o pueden ocurrir simultáneamente. A puede causar B , viceversa o pueden no tener relación causal. Las relaciones causales o temporales son nociones que no pertenecen al ámbito de la probabilidad. Pueden desempeñar un papel o no dependiendo de la interpretación que se le dé a los eventos.

Un ejemplo clásico es el lanzamiento de una moneda para luego lanzar un dado. ¿Cuál es la probabilidad que en el dado salga un 6 dado que ya haya salido una cara en la moneda? Esta probabilidad se denota de esta manera: $P(6|C)$.

Así, Dado un espacio de probabilidad y dos eventos (o sucesos) A, B con $P(B) > 0$, la probabilidad condicional de A dado B está definida como:

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

o

$$P(A \cap B) = P(A) P(B|A)$$

2.4. Eventos independientes

Si $P(A|B) = P(A)$, es decir, que la probabilidad de que ocurra A no depende de la ocurrencia o no de B , entonces decimos que A y B son eventos independientes. Esto equivale a:

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

Inversamente, si se cumple esta relación, entonces A y B son independientes.

De la misma manera, se establece que tres eventos A_1, A_2, A_3 son *independientes* si son independientes por parejas, es decir,

$$P(A_j \cap A_k) = P(A_j) P(A_k) \quad j \neq k \quad \text{donde} \quad j, k = 1, 2, 3$$

2.5. Combinatoria

En muchos casos el número de puntos muestrales en un espacio muestral no es muy grande, de manera que no es difícil la enumeración directa o el conteo de los puntos muestrales necesario para obtener las probabilidades. Sin embargo, cuando es imposible realizar este proceso, se hace uso del análisis combinatorio.

Ejemplo 2.1. Imaginemos que lanzamos una moneda tres veces. Es inmediato ver las diferentes maneras de obtener caras y cruces al repetir el experimento.

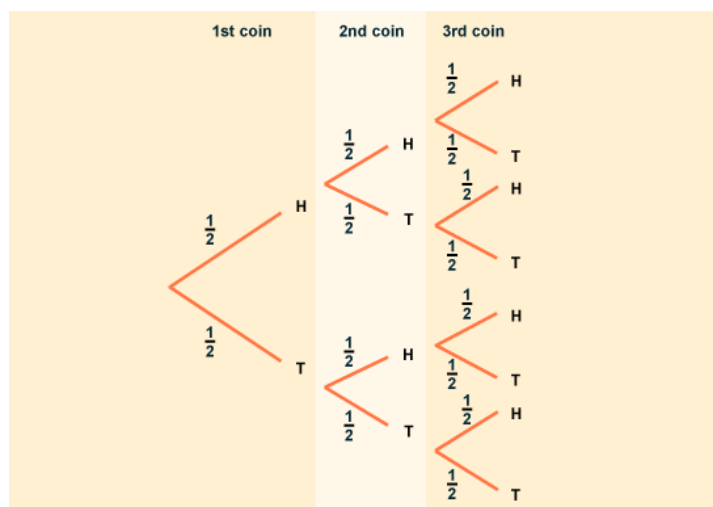


Figura 2: Diagrama de árbol para las posibilidades de obtener sucesivamente cara ('head') y cruz ('tail') lanzando una moneda.

La combinatoria trata del número de diferentes maneras que existen de considerar conjuntos formados a partir de elementos de un conjunto dado, respetando ciertas reglas, como el tamaño, el orden, la repetición, la partición. Así un problema combinatorio consiste usualmente en establecer una regla sobre cómo deben ser las agrupaciones y determinar cuántas existen que cumplan dicha regla.

2.5.1. Permutaciones

Un tipo importante de esas agrupaciones son las llamadas permutaciones. Supongamos que tenemos n objetos diferentes y queremos ordenar estos objetos en línea. Dado que hay n formas de escoger el primer elemento y, una vez escogido éste, sólo tenemos $(n - 1)$ formas de escoger el segundo elemento, y así sucesivamente, vemos que cuando llegamos al elemento k -ésimo sólo tenemos $[n - (k - 1)]$ posibles elementos para escoger, lo que nos lleva a que tenemos $n(n - 1)(n - 2) \dots 2 \cdot 1$ formas de ordenar el conjunto.

Por tanto, dado un conjunto finito A de n elementos, el número de todas sus permutaciones es igual a factorial de n :

$$n! = n(n - 1)(n - 2) \dots 1$$

Supongamos ahora que tenemos n objetos diferentes y queremos ordenar k de estos objetos en línea. Como hay n maneras de escoger el primer objeto, $(n - 1)$ maneras de escoger el segundo, y $(n - k + 1)$ maneras de escoger el k -ésimo, es fácil deducir que el número de permutaciones diferentes que podemos hacer viene dado por,

$$n(n - 1)(n - 2) \dots (n - k + 1) = \frac{n!}{(n - k)!}$$

2.5.2. Combinaciones

En una permutación nos interesa el orden de los objetos. Por ejemplo, abc es una permutación diferente de bca . Sin embargo, en muchas situaciones solo nos interesa la selección de los objetos sin tener en cuenta su orden. Este tipo de selecciones se denominan *combinaciones*. Por ejemplo, abc y bca son la misma combinación.

El número de formas de escoger k elementos a partir de un conjunto de n (o combinaciones de n tomadas en k) se denota por $C_{(n,k)}$, C_k^n o $\binom{n}{k}$ y es,

$$\binom{n}{k} = C_{(n,k)} = \frac{n!}{k!(n-k)!}$$

2.6. Aproximación de Stirling

Por último, cuando n es grande, no es práctica la evaluación directa de $n!$. En tal caso se puede emplear la siguiente fórmula de aproximación,

$$n! \sim \sqrt{2\pi n} n^n e^{-n}$$

O bien,

$$\ln n! \approx n \ln n - n$$

La fórmula resulta útil en diversas áreas como la mecánica estadística, donde aparecen ecuaciones que contienen factoriales del número de partículas. Puesto que en la materia ordinaria los sistemas macroscópicos típicos tienen en torno a $N \approx 10^{23}$ partículas, la fórmula de Stirling resulta muy buena aproximación. Además, la fórmula de Stirling es diferenciable lo que permite el cálculo aproximado de máximos y mínimos en expresiones con factoriales.

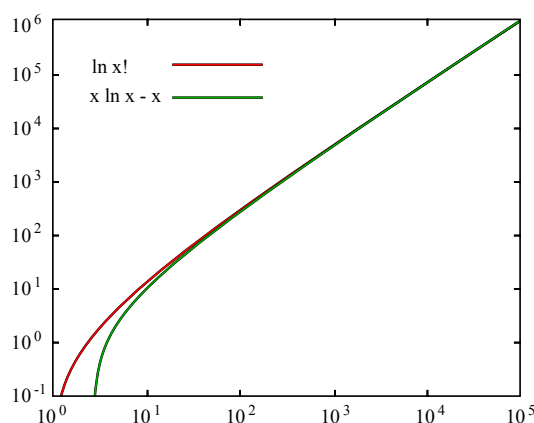


Figura 3: Comparación de la aproximación de Stirling con $x!$. Escala doble logarítmica.

3. Variables aleatorias y distribución de probabilidad

Supongamos que a cada punto del espacio muestral le asignamos un número. Tenemos entonces una función definida en el espacio muestral. Esta función recibe el nombre de *variable aleatoria* (o *variable estocástica*). Usualmente se denota con una letra mayúscula, como X o Y . En general, una variable aleatoria tiene alguna significación física, geométrica o de otro tipo.

Una variable aleatoria que toma un número finito o contable infinito de valores se llama una *variable aleatoria discreta*, mientras que una que toma un número de valores infinito no contable recibe el nombre de *variable aleatoria no discreta*.

3.1. Distribuciones discretas

Sea X una variable aleatoria discreta y asumamos que los posibles valores que puede asumir están dados por x_1, x_2, x_3, \dots con determinado orden. Supongamos también que estos valores se asumen con probabilidades dadas por

$$P(X = x_k) = f(x_k) \quad k = 1, 2, \dots$$

Entonces, $f(x)$ se define como la *función o distribución de probabilidad* y viene dada por,

$$P(X = x) = f(x)$$

En general, diremos que $f(x)$ es una función de probabilidad si,

1. $f(x) \geq 0$
2. $\sum_k f(x_k) = 1$

Ejemplo 3.1. Calcula la función de probabilidad correspondiente a la variable aleatoria número de caras (X) que podremos obtener al lanzar una moneda dos veces.

Suponiendo que es equiprobable obtener cara (C) y cruz (⊕), tendremos,

$$P(CC) = \frac{1}{4} \quad P(C\oplus) = \frac{1}{4} \quad P(\oplus C) = \frac{1}{4} \quad P(\oplus\oplus) = \frac{1}{4}$$

Entonces,

$$\begin{cases} P(X = 0) = P(CC) = \frac{1}{4} \\ P(X = 1) = P(C\oplus \cup \oplus C) = P(C\oplus) + P(\oplus C) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2} \\ P(X = 2) = P(\oplus\oplus) = \frac{1}{4} \end{cases}$$

Así, la distribución de probabilidad viene dada en la siguiente tabla,

x	0	1	2
$f(x)$	1/4	1/2	1/4

3.1.1. Función de distribución

La distribución de probabilidad no debe confundirse con la *función de distribución acumulada* o *función de distribución*, que para una variable aleatoria X está definida por,

$$F(x) = P(X \leq x)$$

donde x puede ser un número real cualquiera, $x \in (-\infty, \infty)$.

La función de distribución acumulada $F(x)$ tiene las siguientes propiedades,

1. $F(x)$ es creciente
2. $\lim_{x \rightarrow -\infty} F(x) = 0$; $\lim_{x \rightarrow \infty} F(x) = 1$
3. $F(x)$ es continua desde la derecha. Es decir $\lim_{h \rightarrow 0^+} F(x+h) = F(x) \forall x$

Variables aleatorias discretas En el caso de una variable aleatoria discreta. La función de distribución acumulada para una variable aleatoria X , puede obtenerse a partir de su función de probabilidad de la siguiente manera,

$$F(x) = \sum_{u \leq x} f(u)$$

donde la suma sustituye todos los valores u tomados por X para $u \leq x$.

Ejemplo 3.2. Consideremos la variable aleatoria discreta del ejemplo anterior (el número de caras (X) que podremos obtener al lanzar una moneda dos veces). La función de distribución acumulada viene definida por,

$$F(x) = \begin{cases} 0 & -\infty < x < 0 \\ \frac{1}{4} & 0 \leq x < 1 \\ \frac{3}{4} & 1 \leq x < 2 \\ 1 & 2 \leq x < \infty \end{cases}$$

Y su gráfica es la siguiente,

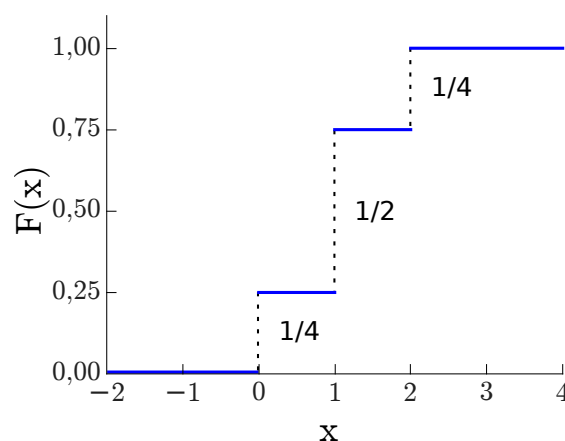


Figura 4: Función de distribución para $F(x)$.

3.2. Distribuciones continuas

Para una variable continua, X , no tiene sentido definir una función de probabilidad, ya que $P(X = x) = 0$ para todo x . No obstante, es razonable definir la función de distribución, $F(x)$ como anteriormente, con la única diferencia que esta función es ya una función continua en lugar de una función escalera.

Por tanto, se dice que una variable aleatoria no discreta X es simplemente continua, si su función de distribución se puede representar como

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(u) du \quad (-\infty < x < \infty)$$

donde la función $f(x)$ es la función de *densidad de probabilidad* de x , definida como,

$$f(x) = \frac{dF}{dx}$$

y tiene las siguientes propiedades,

1. $f(x) \geq 0 \quad \forall x$
2. $\int_{-\infty}^{\infty} f(x) dx = 1$
3. $\int_{-\infty}^x f(u) du = F(x)$
4. $P(a < X < b) = \int_a^b f(x) dx$

Medidas de centralidad Se puede utilizar la función de distribución para calcular algunas medidas de localización y de dispersión de una variable.

En primer lugar definimos la *moda* como el valor más probable, o el que ocurre con más frecuencia, en nuestra distribución de probabilidad. En el caso de tener máximos equiprobables, una distribución puede ser *bimodal*, *trimodal* o *multimodal*.

Por otro lado, definimos la *mediana* de una variable continua, X , como el punto $x_{0,5}$ tal que $F(x_{0,5}) = 0,5$. Es decir, aquel valor de x para el que se alcanza el 50% de la probabilidad acumulada.

Con más generalidad, puede definirse el percentil como el punto x^p para el cual $F(x^p) = p$. En particular, el primer cuartil es $Q_1 = x_{0,25}$ y el tercer cuartil es el punto $Q_3 = x_{0,75}$.

Pero, ¿y el caso de las variables discretas?

En el caso de variables discretas, esta definición de la mediana no es adecuada, ya que puede ser posible de que no exista ningún valor x tal que $F(x) = 0,5$. Por ello se emplea la siguiente definición, para cualquier percentil, que debe cumplir,

$$P(X \leq x) \geq p \quad P(X \geq x) \geq p$$

y en particular para $p = 0,5$,

$$P(X \leq x) \geq 0,5 \quad P(X \geq x) \geq 0,5$$

Interpretación gráfica Si $f(x)$ es la función de densidad para una variable aleatoria X , entonces podemos representarla gráficamente. Como $f(x) \geq 0$, la curva no puede caer por debajo del eje x . Además, el área completa limitada por la curva debe ser igual a 1, para que la distribución de probabilidad esté correctamente normalizada, y es fácil ver, que la probabilidad de que X se encuentre entre a y b , se representa por la integral bajo la curva.

Por otro lado, la función $F(x)$ es monótonamente creciente y crece de 0 a 1.

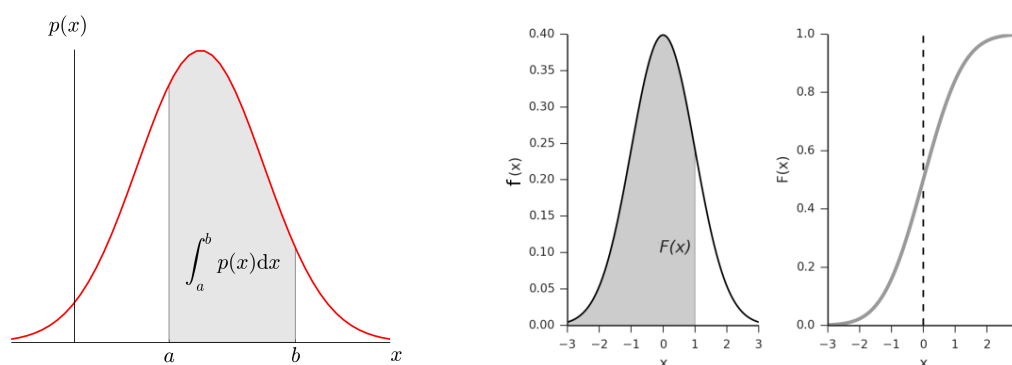


Figura 5: Probabilidad en un intervalo $[a, b]$. Figura 6: Cálculo de la función acumulada.

3.3. Cambio de variables

Dada las distribuciones de probabilidad de una o más variables aleatorias, puede interesarnos encontrar las distribuciones de otras variables aleatorias que dependen de ellas.

Variables discretas

Teorema 3.1. Sea X una variable aleatoria discreta con función de probabilidad $f(x)$. Supongamos que la variable aleatoria discreta U se define en términos de X por $U = \phi(X)$, donde para cada valor de X corresponde uno y sólo uno de los valores de U e inversamente, de manera que $X = \psi(U)$. Entonces la función de probabilidad de U está dada por

$$g(u) = f[\psi(u)]$$

Variables continuas

Teorema 3.2. Sea X una variable aleatoria continua con distribución de probabilidad $f(x)$. Definamos $U = \phi(X)$ donde $X = \psi(U)$. Entonces la densidad de probabilidad de U está dada por $g(u)$ donde

$$g(u) |du| = f(x) |dx|$$

o,

$$g(u) = f(x) \left| \frac{dx}{du} \right| = f[\psi(u)] |\psi'(u)|$$

Teorema 3.3. Sean X e Y variables aleatorias continuas con función de densidad conjunta $f(x, y)$. Definamos $U = \phi_1(X, Y)$, $V = \phi_2(X, Y)$ donde $X = \psi_1(U, V)$, $Y = \psi_2(U, V)$. Entonces la función de densidad conjunta de U y V está dada por $g(u, v)$ donde

$$g(u, v) |du dv| = f(x, y) |dx dy|$$

o

$$g(u, v) = f(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right| = f[\psi_1(u, v), \psi_2(u, v)] |J|$$

donde J es el determinante jacobiano y está dado por,

$$J = \frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{vmatrix}$$

Nota. El cambio de variables en distribuciones de probabilidad es importante, y en el caso de variables continuas, implica al jacobiano de la transformación.

4. Momentos y cumulantes

Una vez introducidas las distribuciones de probabilidad, el siguiente paso es caracterizarlas. Hasta ahora hemos visto para ello las medidas de centralidad de una variable (moda y mediana). Otra medida de centralidad es la media.

Supongamos que generamos una muestra, de una variable discreta con función de probabilidad $P(\cdot)$. En este caso, la media muestral es,

$$\bar{x} = \sum_i x_i P(X = x_i)$$

donde $P(X = x_i)$ es la probabilidad de que nuestra variable estocástica tome cierto valor x_i .

Es fácil ver que si nuestra distribución es continua basta sustituir la suma por la integral y tendremos,

$$\bar{x} = \int x f(x) dx$$

Es común denominar la media como $E[X]$ o μ .

De la misma forma que hemos definido la media podemos definir esperanzas de orden más alto. Formalmente, para $k \in \mathbb{N}$, el momento de orden k es $E[X^k]$. De forma análoga a la definición anterior, se calcula como,

$$\langle x^k \rangle = \sum_i x_i^k P(X = x_i)$$

para el caso y discreto y,

$$\langle x^k \rangle = \int x^k f(x) dx$$

para el caso continuo.

Notación. Indistintamente, emplearemos de ahora en adelante tanto E como $\langle \cdot \rangle$ para referirnos a cualquier momento de una distribución de probabilidad. Durante el curso, adoptaremos $\langle \cdot \rangle$ para no confundirlo con la energía de un sistema.

Algunos momentos tienen nombres especiales. El primer momento es la media, el segundo está relacionado con la varianza (σ^2), el tercero con el sesgo (indica la asimetría) y el cuarto con la kurtosis (analiza la concentración de valores en torno a la media). Por último, el momento central de orden k se define como $E[(X - \mu)^k]$.

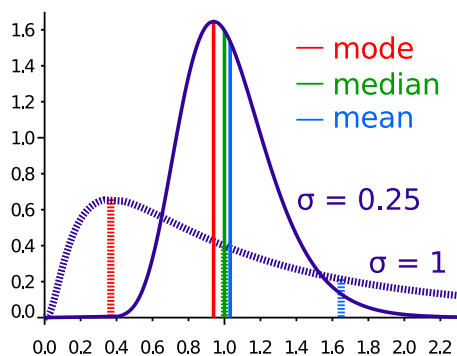


Figura 7: Comparación de la media, mediana y moda para dos distribuciones log-normal con diferente varianza.

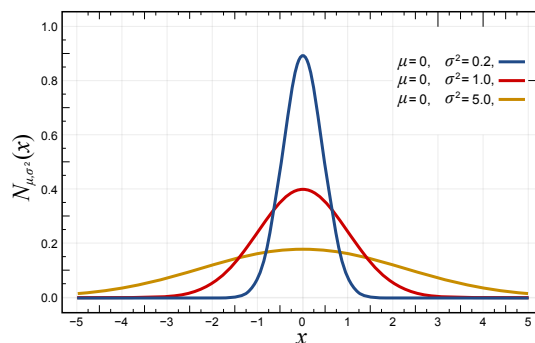


Figura 8: Distribución normal con distinto valor de la varianza.

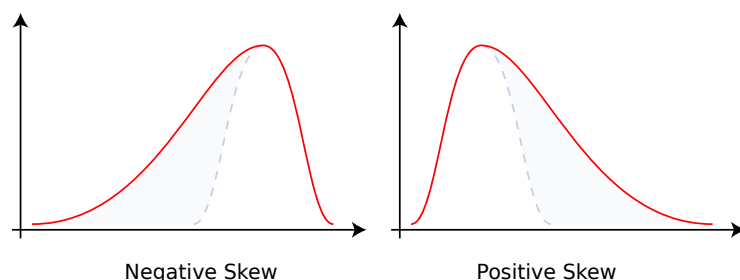


Figura 9: Sesgo negativo y positivo.

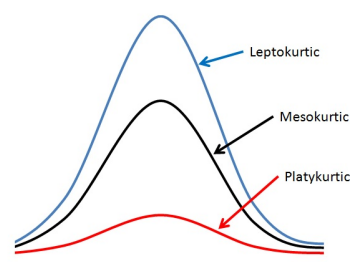


Figura 10: Kurtosis.

4.1. Función generadora de momentos

La función generadora de momentos o función generatriz de momentos de una variable aleatoria X es

$$M_X(t) := \mathbb{E}(e^{tX}), \quad t \in \mathbb{R},$$

siempre que esta esperanza exista.

Viene dada por la siguiente suma en el caso discreto,

$$M_X(t) = \sum_{i=1}^{\infty} e^{tx_i} P(X = x_i)$$

Viene dada por la integral de Riemann-Stieltjes en el caso continuo,

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

La función generadora de momentos se llama así porque, si existe en un entorno de $t = 0$, permite generar los momentos de la distribución de probabilidad:

$$\mathbb{E}(X^n) = M_X^{(n)}(0) = \left. \frac{d^n M_X}{dt^n} \right|_{t=0}$$

Si la función generadora de momentos está definida en tal intervalo, entonces determina unívocamente a la distribución de probabilidad.

4.2. Función generadora de cumulantes

Un problema clave con las funciones generadoras de momentos es que los momentos y la propia función generadora no siempre existen, porque las integrales que los definen no son siempre convergentes. Por el contrario, la función característica siempre existe y puede usarse en su lugar.

La función característica de una variable aleatoria o de su distribución de probabilidad es una función de variable real que toma valores complejos, que permite la aplicación de métodos analíticos (es decir, de análisis funcional) en el estudio de la probabilidad.

Dada una variable aleatoria continua X ,

$$\varphi_X: \mathbb{R} \rightarrow \mathbb{C}; \quad \varphi_X(t) = \mathbb{E}[e^{itX}] = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx$$

De la misma manera, si nuestra variable es discreta,

$$\varphi_X(t) = \sum e^{itx_i} P(X = x_i)$$

Cuando los momentos de una variable aleatoria existen, se pueden calcular mediante las derivadas de la función característica.

$$\varphi_X^{(n)}(0) = i^n \mathbb{E}[X^n]$$

A partir de esta definición de función característica se pueden definir los cumulantes (κ_n) como el logaritmo natural de la función característica, esto es¹,

$$H(t) = \log \mathbb{E}[e^{itX}] = \sum_{n=1}^{\infty} \kappa_n \frac{(it)^n}{n!} = \mu it - \sigma^2 \frac{t^2}{2} + \dots$$

En física estadística, muchas cantidades extensivas (proporcionales al tamaño o el volumen del sistema) están relacionadas con cumulantes de una variable aleatoria. Como se verá durante el curso, la función de partición canónica viene determinada por $Z(\beta) = \langle e^{-\beta E} \rangle$, donde E representa a la energía del sistema y $\beta = \frac{1}{kT}$, donde k es la constante de Boltzmann. La energía libre de Helmholtz viene dada por $F = -\frac{1}{\beta} \log Z$, y está claramente relacionada con la función generadora de los cumulantes para la energía. A partir de esta expresión, podrán obtenerse multitud de propiedades termodinámicas del sistema, como la energía interna, la entropía o el calor específico.

¹Observa que también pueden definirse a partir de la función generadora de momentos, $K(t) = \log \mathbb{E}[e^{tX}] = \sum_{n=1}^{\infty} \kappa_n \frac{t^n}{n!}$, pero pueden no estar bien definidos para todos los valores reales de t si $E(e^{tX})$ no está bien definida.

5. Distribuciones especiales

5.1. Distribución binomial

Imaginemos que estamos realizando un experimento como lanzar repetidamente una moneda, o un dado, o sacar repetidamente una canica de una urna (donde podremos buscar éxito en un evento particular. Por ejemplo, obtener cara o sacar un 4).

Consideremos p la probabilidad de que tengamos éxito en un experimento de este tipo. Entonces $q = 1 - p$ es la probabilidad de fracaso. La probabilidad de que el evento ocurra exactamente x veces en n pruebas (es decir, n éxitos y $n - x$ fracasos), está dada por la función de probabilidad,

$$f(x) = P(X = x) = \binom{n}{x} p^x q^{n-x} = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

donde la variable aleatoria X denota el número de éxitos en n pruebas y $x = 0, 1, \dots, n$.

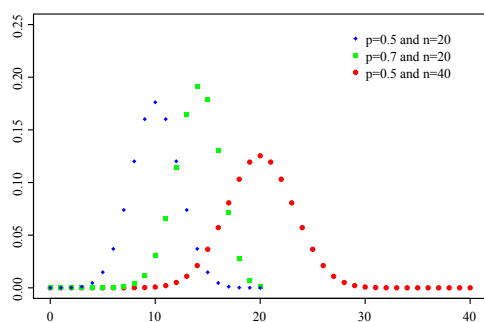


Figura 11: Función de distribución para distintos valores de p y n .

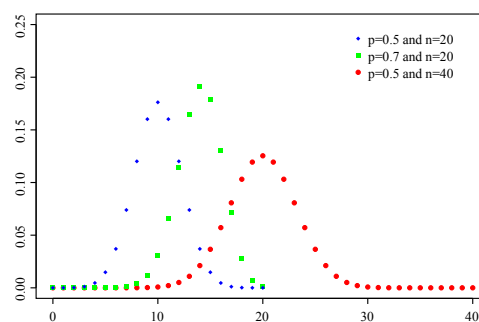


Figura 12: Función de distribución de probabilidad para distintos valores de p y n .

Algunas propiedades importantes de esta distribución se presentan en la siguiente tabla,

Media	$\mu = np$
Varianza	$\sigma^2 = npq$
Sesgo	$\frac{q-p}{\sqrt{npq}}$
Kurtosis	$3 + \frac{1-6pq}{npq}$
Función generadora de momentos	$M(t) = (q + pe^t)^n$
Función característica	$\varphi(t) = (q + pe^{it})^n$

5.2. Distribución de Poisson

Sea X una variable discreta que puede tomar los valores $0, 1, 2, \dots$ tal que la función de probabilidad de X está dada por,

$$f(x) = P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad x = 0, 1, 2, \dots$$

donde λ es una constante positiva. Esta distribución se denomina de Poisson.

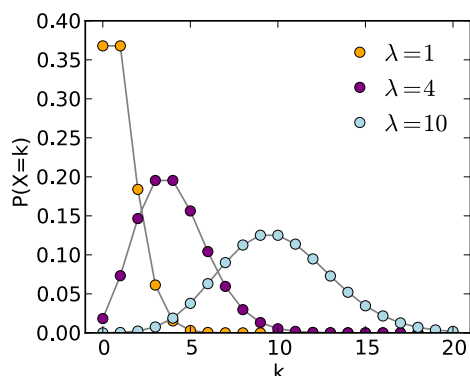


Figura 13: Función de distribución para distintos valores de λ .

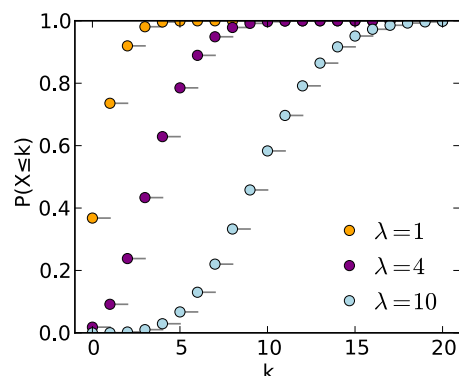


Figura 14: Función de distribución de probabilidad para distintos valores de λ .

Algunas propiedades importantes de esta distribución se presentan en la siguiente tabla,

Media	$\mu = \lambda$
Varianza	$\sigma^2 = \lambda$
Sesgo	$\frac{1}{\sqrt{\lambda}}$
Kurtosis	$3 + \frac{1}{\lambda}$
Función generadora de momentos	$M(t) = e^{\lambda(e^t - 1)}$
Función característica	$\varphi(t) = e^{\lambda(e^{it} - 1)}$

Si consideramos una distribución binomial, con n grande, y si la ocurrencia de un evento p es muy pequeña, de manera que $q = 1 - p$ se acerca a 1, el evento se denomina evento raro. Para tales casos, la distribución binomial se aproxima bastante a la distribución de Poisson con $\lambda = np$.

5.3. Distribución normal

Uno de los ejemplos más importantes de una distribución de probabilidad continua es la distribución normal o gaussiana. La función de densidad para esta distribución está dada por,

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad -\infty < x < \infty$$

donde μ es la media y σ^2 la varianza.

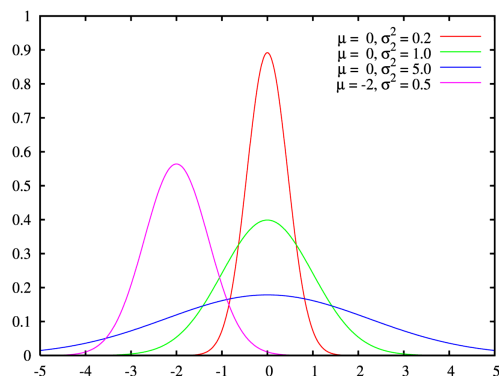


Figura 15: Función de distribución para distintos valores de μ y σ^2 .

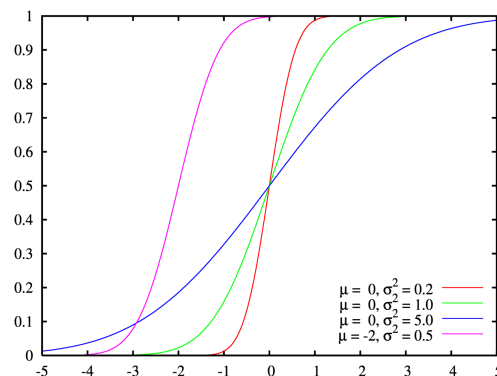


Figura 16: Función de distribución de probabilidad para distintos valores de μ y σ^2 .

Algunas propiedades de la distribución normal se muestran en la siguiente tabla,

Media	μ
Varianza	σ^2
Sesgo	0
Kurtosis	0
Función generadora de momentos	$M(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$
Función característica	$\varphi(t) = e^{i\mu t - \frac{\sigma^2 t^2}{2}}$

5.4. Ley de los grandes números

La ley fuerte de los grandes números establece que si X_1, X_2, X_3, \dots es una sucesión infinita de variables aleatorias independientes que tienen el mismo valor esperado $\mu < \infty$, entonces el promedio

$$\bar{X}_n = \frac{(X_1 + \dots + X_n)}{n}$$

converge en probabilidad a μ . En otras palabras,

$$P\left(\lim_{n \rightarrow \infty} \bar{X}_n = \mu\right) = 1$$

5.5. Teorema del Límite Central

Teorema 5.1. Teorema del Límite Central

Sea $S_n := \frac{X_1 + \dots + X_n}{n}$ y supongamos que $\{X_1, \dots, X_n\}$ es una secuencia de variables estocásticas independientes e idénticas con media μ y varianza $\sigma^2 < \infty$. Cuando se cumpla la condición $n \rightarrow \infty$, la variable estocástica $\sqrt{n}(S_n - \mu)$ converge a una distribución normal $N(0, \sigma^2)$:

$$\sqrt{n} \left(\left(\frac{1}{n} \sum_{i=1}^n X_i \right) - \mu \right) \xrightarrow{d} N(0, \sigma^2).$$

Demostración. Consideremos $\{X_1, \dots, X_n\}$ variables estocásticas idénticamente distribuidas, cada una con media μ y varianza σ^2 . Es inmediato ver que la suma $X_1 + \dots + X_n$ tiene media $n\mu$ y varianza $n\sigma^2$. Consideremos por tanto la siguiente variable estandarizada,

$$Z_n = \frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma^2}} = \sum_{i=1}^n \frac{X_i - \mu}{\sqrt{n\sigma^2}} = \sum_{i=1}^n \frac{1}{\sqrt{n}} Y_i,$$

donde hemos definido la variable $Y_i = \frac{X_i - \mu}{\sigma}$, que tendrá media $\mu = 0$ y varianza unidad $\sigma^2 = 1$. La función característica de Y_n viene dada por,

$$\varphi_{Z_n} = \varphi_{\sum_{i=1}^n \frac{1}{\sqrt{n}} Y_i}(t) = E \left[e^{it \sum_{i=1}^n \frac{1}{\sqrt{n}} Y_i} \right] = \varphi_{Y_1} \left(\frac{t}{\sqrt{n}} \right) \cdot \varphi_{Y_2} \left(\frac{t}{\sqrt{n}} \right) \cdots \varphi_{Y_n} \left(\frac{t}{\sqrt{n}} \right) = \left[\varphi_{Y_1} \left(\frac{t}{\sqrt{n}} \right) \right]^n$$

donde hemos empleado en los dos últimos pasos el hecho de que son variables independientes e idénticas, respectivamente. Por último, desarrollando en series de Taylor,

$$\varphi_{Y_1} \left(\frac{t}{\sqrt{n}} \right) = E \left[e^{t \frac{Y_1}{\sqrt{n}}} \right] = \cancel{E(1)} + i \frac{t}{\sqrt{n}} \cancel{E[Y_1]} + \frac{t^2}{2n} \cancel{E[Y_1^2]} - \frac{i}{6} \frac{t^3}{n^{\frac{3}{2}}} E[Y_1^3] + o \left(\frac{t^3}{n^{\frac{3}{2}}} \right), \quad t \rightarrow 0$$

donde $o(t^3)$ implica que la función va a cero más rápido que t^3 . Tomando el límite de la función exponencial, $e^x = \lim_{n \rightarrow \infty} \left(1 + \frac{x}{n} \right)^n$, y tomando el desarrollo en serie hasta orden n (cierto en la medida en que $n \rightarrow \infty$), tendremos,

$$\varphi_{Y_1} \left(\frac{t}{\sqrt{n}} \right) = 1 - \frac{t^2}{2n}$$

Esto no es otra cosa que la función característica de una distribución normal $N(0, 1)$. No obstante, la distribución

$$S_n = \frac{X_1 + \dots + X_n}{n}$$

converge a una distribución normal $N\left(\mu, \frac{\sigma^2}{n}\right)$, y la suma $X_1 + \dots + X_n$ a una normal $N(n\mu, n\sigma^2)$. \square

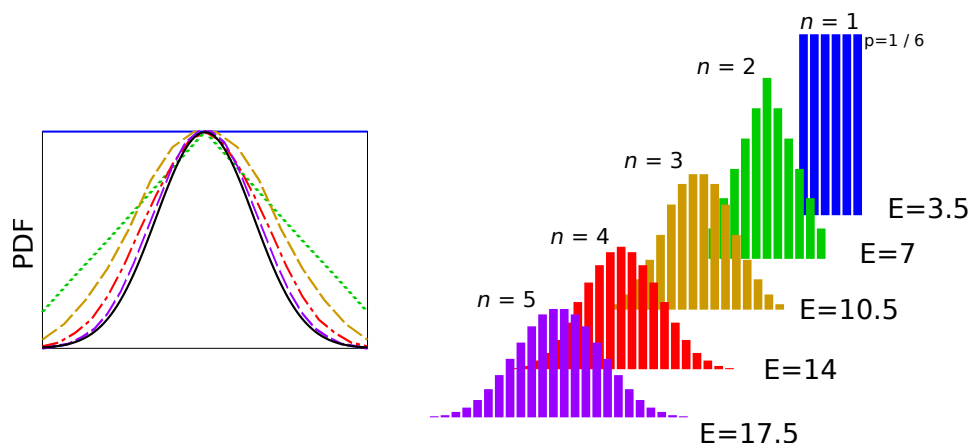


Figura 17: Evolución de la distribución de probabilidad suma de dados mostrando su convergencia a una normal, de acuerdo al teorema del límite central. En la figura izquierda se muestra la comparación con una normal (curva negra).

Referencias

- [1] MR Spiegel, J Schiller & RA Srinivasan (2003). Probabilidad y Estadística. Editorial Schaum. McGraw-Hill.
- [2] JJ Brey & J de la Rubia Pacheco (2001). Mecánica estadística. Universidad Nacional de Educación a Distancia, UNED.