

Computación para Estudiantes de Física

Apunte Número 3

Nociones sobre Teoría de Probabilidades

Sergio J. Sciutto
Departamento de Física
Facultad de Ciencias Exactas
Universidad Nacional de La Plata
C. C. 67 - 1900 La Plata, Argentina.

AÑO 2011

1. Introducción

Un *fenómeno aleatorio* es un fenómeno empírico que se caracteriza por la propiedad de que, al observarlo bajo determinado conjunto de condiciones, no siempre se obtiene el mismo resultado (de manera que no existe regularidad determinista) sino que los diferentes resultados ocurren con regularidad estadística, es decir, existen números entre 0 y 1, bien determinados, que indican la frecuencia relativa con la que se observan los diferentes resultados en una serie de repeticiones independientes del fenómeno.

Dos conceptos estrechamente ligados al fenómeno aleatorio son los de *evento aleatorio* y de la *probabilidad* de un evento aleatorio. Un evento aleatorio es cada uno de los resultados posibles de un fenómeno aleatorio. La frecuencia relativa con la que un evento aleatorio dado aparece en una sucesión larga de observaciones independientes y al azar, define, en el límite en que el número de observaciones tiende a infinito, la probabilidad del evento correspondiente.

Consideremos, por ejemplo, un bolillero que contiene 10 bolillas, numeradas del 1 al 10. Se extrae una bolilla, se observa su número, se la introduce nuevamente en el bolillero, y así sucesivamente se continúa extrayendo bolillas adicionales.

Ejemplos de eventos son: Extraer la bolilla 1, extraer una bolilla par mayor que tres, etc. Las probabilidades de estos eventos dependen de las características del experimento. Bajo ciertas circunstancias es posible realizar un modelo matemático de la realidad y decir algo acerca de las frecuencias con que pueden darse los distintos eventos. Por ejemplo, si las 10 bolillas son idénticas –salvo por el número que las identifica– y el método de extracción asegura que cualquiera de ellas puede ser extraída en cualquier momento, entonces podemos intuir que la frecuencia relativa de aparición de la bolilla i será $1/10$, es decir, el cociente entre el número de eventos “similares” o “igualmente probables” y el número total de eventos posibles. Del mismo modo, la probabilidad de extraer una bolilla par mayor que 3 debe ser $4/10 = 2/5$, etc.

El conjunto de todos los resultados posibles de un experimento se denomina *espacio muestral* del experimento aleatorio. En nuestro ejemplo, el espacio muestral es el conjunto $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$, el cual contiene todos los resultados posibles del experimento aleatorio “extraer una bolilla”. En general utilizaremos la letra S para referirnos al espacio muestral de un experimento.

Se define un *evento* E como un subconjunto cualquiera de S , $E \subset S$. El evento “extraer una bolilla par mayor que 3” se representa por el subconjunto $\{4, 6, 8, 10\}$ de S .

En todo experimento aleatorio siempre es posible definir dos eventos particulares: El evento seguro $E = S$, y el evento imposible $E = \emptyset$.

2. Axiomas de la Teoría de Probabilidades

Dada una situación aleatoria, que queda descrita por un espacio muestral S , una función P definida en el conjunto de todos los subconjuntos de S (conjunto de eventos), que asigna a cada evento un número real, es decir $P(E) \in \mathbb{R}$, para todo $E \subset S$, es una **función de probabilidades** si y solo si satisface los siguientes tres axiomas:

1. $P(E) \geq 0$ para todo $E \subset S$.
2. $P(S) = 1$ (S es el evento seguro).
3. Dados $E_1 \subset S$ y $E_2 \subset S$, con $E_1 \cap E_2 = \emptyset$ (eventos mutuamente excluyentes), entonces

$$P(E_1 \cup E_2) = P(E_1) + P(E_2) \quad (1)$$

Es de notar que toda la teoría de probabilidades se construye sobre estos tres axiomas.

2.1. Resultados inmediatos

Se listan a continuación una serie de resultados básicos que surgen en forma directa de los axiomas de la Teoría de Probabilidades:

- El evento imposible tiene probabilidad cero, es decir,

$$P(\emptyset) = 0. \quad (2)$$

Demostración. $S \cap \emptyset = \emptyset$, luego, por el axioma 3 resulta $P(S \cup \emptyset) = P(S) + P(\emptyset)$; pero $S \cup \emptyset = S$, y entonces $P(S \cup \emptyset) = P(S)$, y luego reemplazando en la expresión anterior queda $P(\emptyset) = 0$ **c.q.d.**

- Sea E un evento cualquiera, y sea E_c el evento complementario, el cual verifica: $E \cup E_c = S$ con $E \cap E_c = \emptyset$. Sea F otro evento cualquiera, entonces

$$P(F \cap E_c) = P(F) - P(E \cap F). \quad (3)$$

Demostración. Se tiene que $(F \cap E) \cap (F \cap E_c) = \emptyset$ y $(F \cap E) \cup (F \cap E_c) = F$, entonces por el axioma 3 resulta $P(F) = P(F \cap E) + P(F \cap E_c)$, que es equivalente a la ecuación (3) **c.q.d.**

- (Corolario del teorema anterior) Tomando $F = S$ en (3), resulta

$$P(E_c) = 1 - P(E). \quad (4)$$

Demostración. Trivial a partir de las propiedades de S , E y E_c .

- Sean E y F eventos cualesquiera, entonces

$$P(E \cup F) = P(E) + P(F) - P(E \cap F). \quad (5)$$

Demostración. Se tiene que $(E \cup F) = E \cup (F \cap E_c)$; pero $E \cap (F \cap E_c) = \emptyset$, entonces $P(E \cup F) = P(E) + P(F \cap E_c)$. Aplicando (3) al último término del segundo miembro se obtiene la ecuación (5) **c.q.d.**

3. Probabilidad Condicional

En un ensayo aleatorio con espacio muestral S se toman dos eventos A y B , subconjuntos de S . Por probabilidad condicional del evento B , dado el evento A , denotada $P(B|A)$, se entiende la probabilidad de que B ocurra bajo la suposición de que A ha ocurrido.

Consideremos la siguiente situación: En una caja hay 20 bolillas, 15 blancas y 5 verdes. Si las bolillas son iguales salvo por su color, la probabilidad de sacar cualquier bolilla es $1/20$, y la probabilidad de sacar una bolilla blanca (verde) es $15/20$ ($5/20$). Se procede a extraer una bolilla la que resulta ser verde. En estas condiciones nos preguntamos cuál es la probabilidad de volver a extraer una bolilla verde.

En el presente ejemplo, esta probabilidad es fácil de calcular: $4/19$, que resulta diferente a $5/20$. Este sencillo ejemplo nos muestra que la información disponible *a priori* ha modificado la probabilidad de un evento.

Sea $N = 20 \times 19 = 380$ el número de pares de bolillas que pueden ser extraídos en nuestro experimento. Sean los eventos $A = \text{“La primera bolilla es verde”}$, y $B = \text{“La segunda bolilla es verde”}$. La multiplicidad de estos eventos es:

$$N_A = 5 \times 19 = 95, \quad N_B = 15 \times 5 + 5 \times 4 = 95. \quad (6)$$

Por otra parte, la multiplicidad de $A \cap B$, $N_{A \cap B} = 5 \times 4 = 20$.

La probabilidad condicional $P(B|A)$ corresponde a la fracción de experimentos en los que ocurrió B luego de haber ocurrido A . En términos de cocientes de multiplicidades, esto corresponde a:

$$P(B|A) = \frac{N_{A \cap B}}{N_A} = \frac{20}{95} = \frac{4}{19} \quad (7)$$

que, como corresponde, es igual a la fracción calculada en forma directa.

Los cocientes de multiplicidades coinciden con los cocientes de las respectivas probabilidades (esto es evidente en la ecuación (7), luego de dividir numerador y denominador por N); y esta propiedad es de carácter general, y nos lleva a la definición rigurosa de probabilidad condicional:

La **Probabilidad Condicional** del evento B dado el evento A (no imposible), se define como

$$P(B|A) = \frac{P(A \cap B)}{P(A)}, \quad P(A) > 0. \quad (8)$$

Si $P(A)=0$ entonces $P(B|A)$ está indefinida.

Corolario de la definición: Si ambos eventos son no imposibles, es decir $P(A) > 0$ y $P(B) > 0$, entonces tanto $P(A|B)$ como $P(B|A)$ están definidas, y se verifica

$$P(A|B) = \frac{P(B \cap A)}{P(B)} = \frac{P(A)}{P(B)} P(B|A), \quad (9)$$

en donde se ha aplicado la definición (8). Luego resulta

$$P(B)P(A|B) = P(A)P(B|A), \quad (10)$$

ecuación en donde la relación entre probabilidades absolutas y condicionales se presenta en forma simétrica.

3.1. Teorema de Bayes

Sean $C_1, C_2, \dots, C_n, n > 1$, eventos mutuamente excluyentes y exhaustivos, es decir, $C_i \cap C_j = \emptyset \forall i \neq j$, y $C_1 \cup C_2 \cup \dots \cup C_n = S$. Entonces

$$P(C_j|B) = \frac{P(B|C_j)P(C_j)}{\sum_{i=1}^n P(B|C_i)P(C_i)} \quad (11)$$

para todo evento $B \subset S$.

Demostración. Por la definición de probabilidad condicional es

$$P(C_j|B) = \frac{P(C_j \cap B)}{P(B)}, \quad (12)$$

pero por el carácter excluyente y exhaustivo de los C_i se verifica que $B = (B \cap C_1) \cup (B \cap C_2) \cup \dots \cup (B \cap C_n)$. Luego, aplicando reiteradamente el axioma 3 de la Teoría de Probabilidades, se obtiene

$$P(B) = \sum_{i=1}^n P(B \cap C_i). \quad (13)$$

Reemplazando $P(B \cap C_i) = P(B|C_i)P(C_i)$ en esta última ecuación, aplicando luego el resultado en la ecuación (12), y teniendo en cuenta que $P(C_j \cap B) = P(B|C_j)P(C_j)$, se obtiene la ecuación (11) **c.q.d.**

El Teorema de Bayes es un importante resultado de la teoría básica de probabilidades, que encuentra aplicaciones en numerosos problemas relacionados con análisis probabilísticos y estadísticos.

3.2. Independencia entre eventos

Se dice que dos eventos A y B son independientes si y solo si

$$P(A|B) = P(A). \quad (14)$$

En otras palabras, decir que un evento A es independiente de otro B significa que el tener información sobre B no modifica el estado de conocimiento acerca de A , concepto precisado cuantitativamente por la ecuación (14).

Observese que si $P(A|B) = P(A)$, entonces por (10) resulta $P(B|A) = P(B)$, y por lo tanto la relación de independencia es recíproca (Si A es independiente de B , entonces B es independiente de A).

Si A y B son independientes, se tiene

$$\frac{P(A \cap B)}{P(A)} = P(B|A) = P(B) \quad (15)$$

y entonces,

$$P(A \cap B) = P(A)P(B). \quad (16)$$

En otras palabras, la probabilidad de que ocurran conjuntamente dos eventos independientes es el producto de las respectivas probabilidades.

4. Ensayos independientes de Bernoulli

En muchas aplicaciones nos encontramos con la siguiente situación: Un experimento aleatorio se repite un determinado número de veces en forma independiente. En cada instancia estamos interesados en observar la ocurrencia (“éxito”) o no (“fracaso”) de un dado evento. Sabiendo que la probabilidad de “éxito” es p , la de “fracaso” es $q = 1 - p$.

Si el experimento se repite n veces, podemos preguntarnos ¿Cuál es la probabilidad de obtener k éxitos?

Por ejemplo, si $n = 3$ y $k = 2$, las instancias posibles son:

$$\begin{array}{ll} \text{E E F} & \text{con probabilidad } ppq = p^2q \\ \text{E F E} & \text{con probabilidad } pqp = p^2q \\ \text{F E E} & \text{con probabilidad } qpp = p^2q \end{array}$$

Las probabilidades de cada instancia pueden evaluarse fácilmente dado que hemos asumido que los ensayos son independientes. Como las instancias son mutuamente excluyentes, la probabilidad buscada será la suma de las probabilidades de las instancias, es decir, $3p^2q$.

En el caso general, la probabilidad de cada instancia (k éxitos y $n - k$ fracasos) es $p^k q^{n-k}$, y el número de instancias es el número de combinaciones de n elementos

tomados de a k . Por lo tanto, si $b(k, n, p)$ representa a la probabilidad de obtener k éxitos en n ensayos de Bernoulli con probabilidad de éxito p , entonces

$$b(k, n, p) = \binom{n}{k} p^k q^{n-k}. \quad (17)$$

A esta expresión se le suele llamar *ley de probabilidades binomial*.

5. Variables aleatorias y funciones de distribución

En muchos casos el análisis de los fenómenos aleatorios se realiza más convenientemente si se “mapea” el espacio muestral sobre el conjunto de los números reales. Las funciones definidas a tal efecto se llaman *variables aleatorias*.

Una variable aleatoria X es, entonces una función $X : S \rightarrow \mathbb{R}$.

El conjunto de variables aleatorias suele dividirse en dos grandes grupos: Variables aleatorias discretas y variables aleatorias continuas, según el tipo de argumento $s \in S$ que defina a $X(s)$. Consideremos los siguientes ejemplos:

1. Sea $S = \{\text{blanco, rojo}\}$ el espacio muestral del experimento “extraer una bolilla de color de una urna”. Definimos la variable *discreta*

$$X(s) = \begin{cases} 0 & \text{si } s = \text{blanco} \\ 1 & \text{si } s = \text{rojo} \end{cases} \quad (18)$$

2. Se arrojan tres dardos sobre un gran tablero plano. Definimos la variable *continua*

$$X = \text{Área del triángulo determinado por los 3 dardos} \quad (19)$$

que toma valores en el semieje real positivo.

5.1. Función de distribución acumulativa

Para toda variable aleatoria X , se define la función de distribución acumulativa F_X (o simplemente F), $F_X : \mathbb{R} \rightarrow \mathbb{R}$ dada por

$$F_X(x) = \text{Prob}\{X \leq x\}. \quad (20)$$

Claramente $F_X(-\infty) = 0$ y $F_X(+\infty) = 1$; y además $F_X(x_1) \leq F_X(x_2) \forall x_1 < x_2$ (F_X es monótonamente creciente, y la imagen de F_X es el intervalo $[0, 1]$).

Análogamente se define la *distribución acumulativa complementaria*, $G_X(x)$, $G_X : \mathbb{R} \rightarrow \mathbb{R}$ dada por

$$G_X(x) = \text{Prob}\{X > x\} = 1 - F_X(x). \quad (21)$$

Para variables aleatorias continuas F es, en general, una función suave, mientras que para variables discretas F es una función escalonada. Por ejemplo, la función acumulativa asociada a la variable aleatoria del ejemplo 1 precedente es

$$F(x) = \begin{cases} 0 & \text{si } x < 0 \\ p_b & \text{si } 0 \leq x < 1 \\ 1 & \text{si } x > 1 \end{cases} \quad (22)$$

donde p_b es la probabilidad de extraer una bolilla blanca.

En los casos en donde F es continua y derivable, se define la *función de densidad de probabilidades*, o simplemente *distribución de probabilidades*

$$f_X(x) = \frac{dF_X(x)}{dx}. \quad (23)$$

Si f_X existe, entonces

$$F_X(x) = \int_{-\infty}^x f_X(x') dx'. \quad (24)$$

Como $F(\infty) = 1$, las distribuciones de probabilidades deben cumplir con la siguiente normalización:

$$\int_{-\infty}^{+\infty} f_X(x) dx = 1. \quad (25)$$

En los casos de variables discretas es posible definir densidades de probabilidad utilizando adecuadamente la distribución δ de Dirac.

5.2. Funciones de distribución bivariadas

Sean X_1 y X_2 dos variables aleatorias. Se define la distribución conjunta de probabilidades como

$$F_{X_1 X_2} = \text{Prob}\{(X_1 \leq x_1) \text{ y } (X_2 \leq x_2)\}. \quad (26)$$

Si $F_{X_1 X_2}$ es una función derivable, entonces se puede definir la distribución conjunta de probabilidades, $h(x_1, x_2)$, la cual verifica

$$F_{X_1 X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} h(x'_1, x'_2) dx'_2 dx'_1 \quad (27)$$

$$h(x_1, x_2) = \frac{\partial^2}{\partial x_1 \partial x_2} F_{X_1 X_2}(x_1, x_2) \quad (28)$$

Las funciones de distribución individuales f_{X_1} y f_{X_2} se relacionan con h vía

$$f_{X_1}(x_1) = \int_{-\infty}^{+\infty} h(x_1, x_2) dx_2 \quad (29)$$

$$f_{X_2}(x_2) = \int_{-\infty}^{+\infty} h(x_1, x_2) dx_1 \quad (30)$$

Estas distribuciones son frecuentemente denominadas *distribuciones marginales*.

5.3. Variables aleatorias independientes

Las distribuciones h , f_{X_1} y f_{X_2} adoptan una relación muy particular cuando X_1 y X_2 son independientes.

Notemos ante todo que si este es el caso, entonces se verifica:

$$\text{Prob}\{(X_1 \leq x_1)|(X_2 \leq x_2)\} = \text{Prob}\{X_1 \leq x_1\}, \quad \forall x_1, x_2 \in \mathbb{R}. \quad (31)$$

Por lo tanto resulta:

$$\text{Prob}\{(X_1 \leq x_1) \text{ y } (X_2 \leq x_2)\} = \text{Prob}\{X_1 \leq x_1\} \text{Prob}\{X_2 \leq x_2\}, \quad (32)$$

es decir,

$$\begin{aligned} \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} h(x'_1, x'_2) dx'_2 dx'_1 &= \left[\int_{-\infty}^{x_1} f_{X_1}(x'_1) dx'_1 \right] \left[\int_{-\infty}^{x_2} f_{X_2}(x'_2) dx'_2 \right] \\ &= \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1}(x'_1) f_{X_2}(x'_2) dx'_2 dx'_1. \end{aligned} \quad (33)$$

Como x_1 y x_2 son arbitrarios, esta última ecuación nos lleva a

$$h(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2), \quad (34)$$

es decir, la distribución conjunta de probabilidades de dos variables aleatorias independientes es el producto de las distribuciones individuales.

5.4. Composición de variables aleatorias con funciones

Sea X una variable aleatoria y g una función real. $Y = g(X)$ es una nueva variable aleatoria. ¿Cuál es su función de distribución?

Se verifica:

$$F_Y(y) = \text{Prob}\{Y \leq y\} = \text{Prob}\{g(X) \leq y\}. \quad (35)$$

Para una función g general, el cálculo de la probabilidad indicada en (35) debe ser realizado teniendo en cuenta las particularidades de cada caso. Si g es tal que existe una inversa g^{-1} (unívocamente definida), entonces es

$$\text{Prob}\{g(X) \leq y\} = \begin{cases} \text{Prob}\{X \leq g^{-1}(y)\} & \text{si } g' > 0 \\ \text{Prob}\{X \geq g^{-1}(y)\} & \text{si } g' < 0, \end{cases} \quad (36)$$

o sea,

$$F_Y(y) = \begin{cases} F_X(g^{-1}(y)) & \text{si } g' > 0 \\ 1 - F_X(g^{-1}(y)) & \text{si } g' < 0. \end{cases} \quad (37)$$

De donde obtenemos, al aplicar la definición (23):

$$f_Y(y) = \begin{cases} \frac{1}{g'(y)} f_X(g^{-1}(y)) & \text{si } g' > 0 \\ -\frac{1}{g'(y)} f_X(g^{-1}(y)) & \text{si } g' < 0, \end{cases} \quad (38)$$

en donde se ha utilizado la identidad $(g^{-1})' = (g')^{-1}$. Observese que la expresión de f_Y dada por (38) puede reducirse a una única fórmula:

$$f_Y(y) = \frac{1}{|g'(y)|} f_X(g^{-1}(y)) \quad (39)$$

Ejemplo. Consideremos el caso de la variable $Y = 1/X$, es decir $g(x) = 1/x$. Se tiene $|g'(y)| = y^{-2}$, y aplicando (39) resulta:

$$f_{1/X}(y) = y^2 f_X(1/y). \quad (40)$$

En modo análogo es posible hallar la distribución de probabilidades de la variable aleatoria que resulta de componer dos variables X_1 y X_2 a través de una función real g , es decir, $Y = g(X_1, X_2)$. En este caso se verifica:

$$F_Y(y) = \text{Prob}\{Y \leq y\} = \text{Prob}\{g(X_1, X_2) \leq y\}, \quad (41)$$

es decir:

$$F_Y(y) = \iint_{D_y} h(x_1, x_2) dx_1 dx_2, \quad (42)$$

donde D_y es la región del plano $x_1 x_2$ definida por $D_y = \{(x_1, x_2)/g(x_1, x_2) \leq y\}$.

Un caso particular de mucha importancia es el de la suma de dos variables aleatorias independientes. En este caso es $g(x_1, x_2) = x_1 + x_2$, $h(x_1, x_2) = f_1(x_1)f_2(x_2)$, y D_y es el lugar geométrico de los puntos (x_1, x_2) tales que $x_1 + x_2 \leq y$. La integral (42) puede entonces expresarse de la siguiente manera:

$$F_{X_1+X_2}(y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{y-x_1} f_1(x_1)f_2(x_2) dx_2 dx_1. \quad (43)$$

Luego derivando respecto de y resulta

$$f_{X_1+X_2}(y) = \int_{-\infty}^{+\infty} f_1(x)f_2(y-x) dx, \quad (44)$$

es decir, la densidad de probabilidades de la suma de dos variables aleatorias independientes se obtiene realizando la *convolución* de las densidades de los sumandos.

6. Funciones de distribución usuales

El propósito de esta sección es listar, para referencia futura, las funciones de distribución más usuales en teoría de probabilidades y estadística.

Distribución uniforme. (a, b , reales, $b > a$.)

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{en otro caso} \end{cases} \quad (45)$$

Distribución normal. (m, σ , reales, $\sigma > 0$.)

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2\right\}, \quad -\infty < x < \infty. \quad (46)$$

Distribución de Cauchy. (α, β , reales, $\beta > 0$.)

$$f(x) = \frac{1}{\pi\beta} \frac{1}{\left[1 + \left(\frac{x-\alpha}{\beta}\right)^2\right]}, \quad -\infty < x < \infty. \quad (47)$$

Distribución exponencial. (β real, $\beta > 0$.)

$$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad 0 \leq x < \infty. \quad (48)$$

Distribución χ^2 . (ν real, $\nu > 0$.)

$$f(\chi^2) = \frac{1}{2^{\nu/2}\Gamma(\frac{\nu}{2})} (\chi^2)^{\frac{\nu}{2}-1} e^{-\frac{\chi^2}{2}}, \quad 0 \leq \chi^2 < \infty. \quad (49)$$

Distribución F . (ν_1, ν_2 , reales positivos.)

$$f(F) = \frac{\nu_1^{\nu_1/2} \nu_2^{\nu_2/2}}{B\left(\frac{\nu_1}{2}, \frac{\nu_2}{2}\right)} F^{\frac{\nu_1-2}{2}} (\nu_1 + \nu_2 F)^{-\frac{\nu_1+\nu_2}{2}}, \quad 0 \leq F < \infty. \quad (50)$$

Distribución de Student. (ν real, $\nu > 0$.)

$$F(t) = \frac{1}{\sqrt{\nu}B\left(\frac{1}{2}, \frac{\nu}{2}\right)} \int_{-t}^t \left(1 + \frac{x^2}{\nu}\right)^{-\frac{\nu+1}{2}} dx, \quad 0 \leq t < \infty, \quad (51)$$

con $f(t) = F'(t)$.

7. Esperanza, momentos y magnitudes asociadas

Se define la esperanza de una variable aleatoria como

$$E(X) = \langle X \rangle = \int_{-\infty}^{+\infty} x f(x) dx. \quad (52)$$

Esta definición sólo tiene sentido cuando la integral del segundo miembro es convergente.

$E(X)$ es también llamada la *media* de X , y se la suele simbolizar con la letra μ , $\mu = \langle X \rangle$.

Si $Y = g(X)$, entonces es:

$$\langle Y \rangle = \int_{-\infty}^{+\infty} y f_Y(y) dy = \int_{-\infty}^{+\infty} y dF_Y. \quad (53)$$

Si existe $g'(x)$, entonces (ver sección 5.4) será

$$y dF_Y = g(x) dF_X, \quad -\infty < x < \infty, \quad (54)$$

por lo que resulta

$$\langle Y \rangle = \langle g(X) \rangle = \int_{-\infty}^{+\infty} g(x) f_X(x) dx. \quad (55)$$

Esta expresión nos indica que el valor medio de cualquier función de una variable aleatoria puede ser calculado utilizando directamente la distribución de la variable original, no siendo necesario calcular la distribución de la variable compuesta.

7.1. Momentos de orden superior

Se define el momento de orden n de una variable aleatoria via

$$\mu'_n = \langle x^n \rangle = \int_{-\infty}^{+\infty} x^n f(x) dx. \quad (56)$$

El momento n -simo no está definido en los casos en donde la integral del segundo miembro es divergente.

Para $n = 0$ se tiene $\mu'_0 = 1$ (normalización de la distribución); mientras que para $n = 1$ resulta $\mu'_1 = \mu$.

Es conveniente definir además los *momentos centrados en la media*, μ_n , via:

$$\mu_n = \langle (x - \mu)^n \rangle = \int_{-\infty}^{+\infty} (x - \mu)^n f(x) dx. \quad (57)$$

Notar que $\mu_0 = \mu'_0 = 1$, y que $\mu_1 = 0$.

Los momentos de orden 1, 2, 3 y 4 suelen combinarse convenientemente para definir diversos parámetros asociados, tal como se describe en las próximas secciones.

7.2. Varianza y desviación estandar

La varianza, V , y desviación estandar, σ , de una distribución son dos parámetros frecuentemente utilizados que se definen via

$$V = \mu_2 = \langle (X - \mu)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2. \quad (58)$$

$$\sigma = \sqrt{V} \quad (59)$$

La media define el “centro de gravedad” de una distribución, mientras que la desviación estandar cuantifica el grado de dispersión alrededor de la media. En la figura 1 se ilustra este concepto mostrando dos distribuciones con diferente desviación estandar.

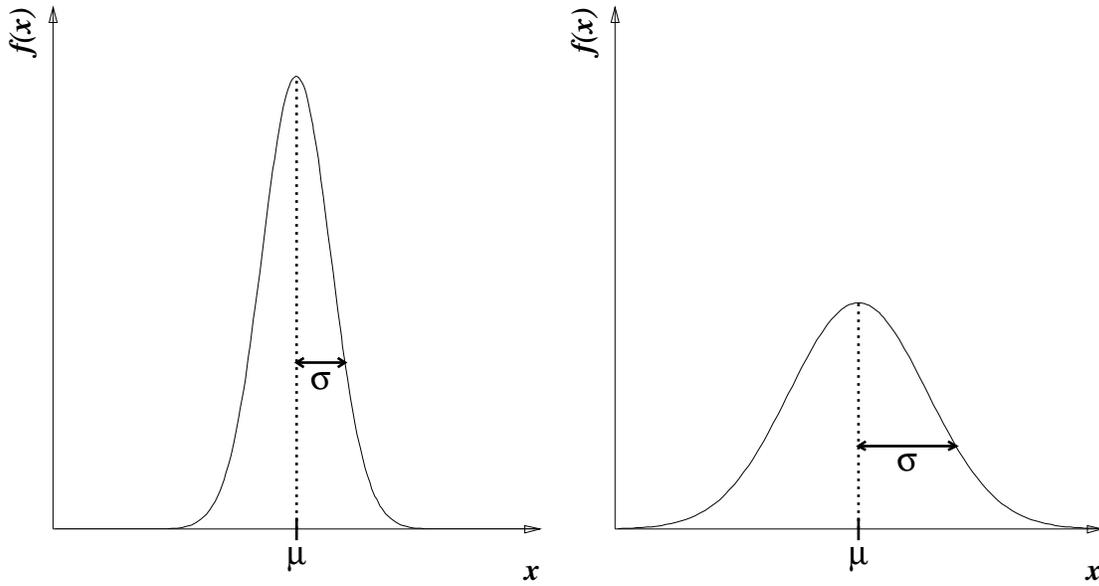


Figura 1: *Media y desviación estandar de distribuciones continuas. Se utiliza una única escala para ambos gráficos.*

7.3. Sezgo y curtosis

Se definen los coeficientes de sezgo (skewness), γ_1 y curtosis, γ_2 , via

$$\gamma_1 = \frac{\mu_3}{\sigma^3} = \frac{\langle (x - \mu)^3 \rangle}{\langle (x - \mu)^2 \rangle^{3/2}} \quad (60)$$

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 = \frac{\langle (x - \mu)^4 \rangle}{\langle (x - \mu)^2 \rangle^2} - 3. \quad (61)$$

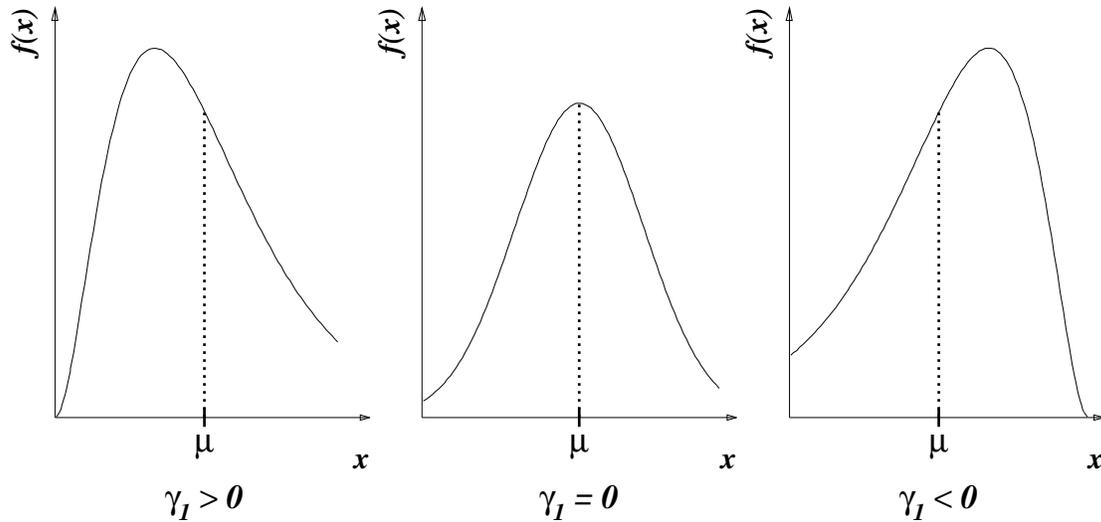


Figura 2: El sesgo en tres casos representativos. Se utiliza una única escala para todos los gráficos, y todas las distribuciones poseen igual media y desviación estandard.

El coeficiente de sesgo es un parámetro que da una medida cuantitativa de la simetría de la distribución alrededor de su media, tal como se ilustra en la figura 2.

En estos gráficos puede verse que un sesgo nulo corresponde a una distribución totalmente simétrica alrededor de la media, mientras que si el sesgo es positivo (negativo) la distribución está desbalanceada con predominancia en la región de x menor (mayor) que la media μ .

La curtosis de una distribución se relaciona con el grado de “achatación” de una distribución. Tal como se ilustra en la figura 3, las distribuciones que son “aplanadas” (especialmente alrededor de la media) se corresponden con parámetros γ_2 positivos y se las llama distribuciones platicúrticas, mientras que las que se concentran cerca de la media (“en punta”) tienen curtosis negativa y son distribuciones leptocúrticas. El caso intermedio, $\gamma_2 = 0$, corresponde a las distribuciones mesocúrticas, como, por ejemplo, la distribución normal¹.

7.4. Otros parámetros de centralización

7.4.1. Moda

Es el punto donde $f(x)$ es máxima. Si $f(x)$ tiene más de un máximo, la distribución se denomina multimodal. En la figura 4 se ilustra este concepto.

¹El término substractivo 3 en la ecuación (61) se coloca deliberadamente para que γ_2 sea cero en el caso de la distribución normal.

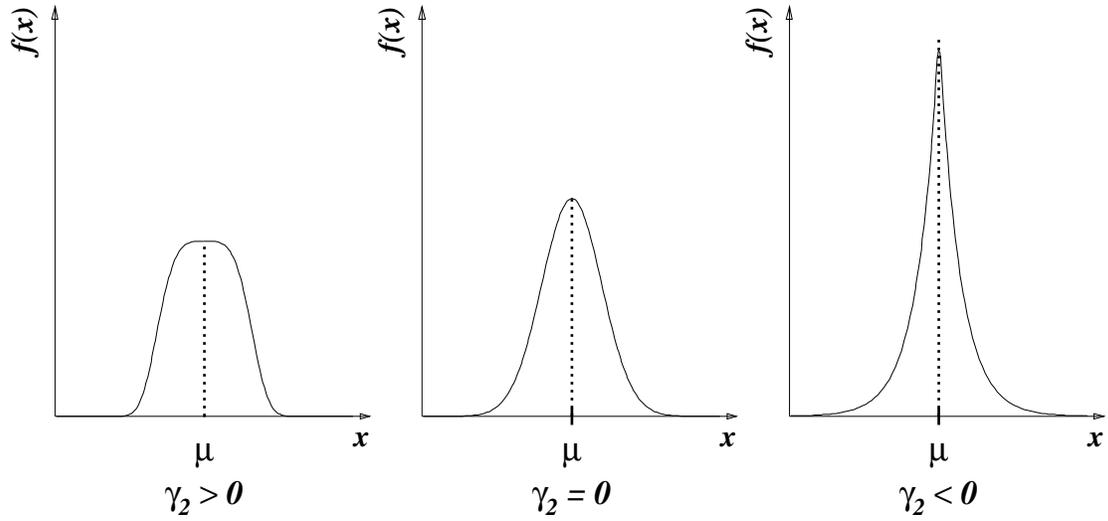


Figura 3: La curtosis en tres casos representativos. Cuando γ_2 es positivo, cero o negativo, la distribución es platicúrtica, mesocúrtica, o leptocúrtica, respectivamente. Se utiliza una única escala para todos los gráficos, y todas las distribuciones poseen igual media y desviación estandar.

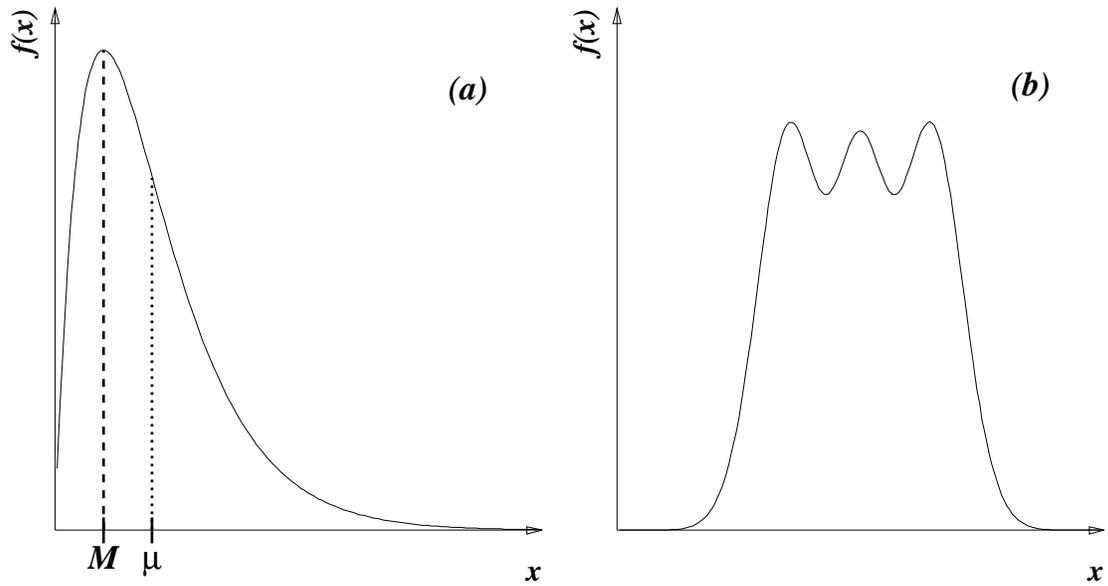


Figura 4: (a) La moda, M , y la media μ de una distribución de probabilidades. (b) Distribución multimodal.

7.4.2. n -tiles

Sean n entero positivo, y sean $y_j = j/n$ ($j = 0, 1, \dots, n$), $n+1$ puntos que subdividen el intervalo $[0, 1]$ en n subintervalos de igual longitud. Sea X una variable aleatoria, y $F(X)$ su función acumulativa de probabilidades.

Los puntos x_j definidos por la relación

$$F(x_j) = y_j, \quad j = 0, 1, \dots, n \quad (62)$$

se denominan n -tiles de la distribución. Claramente $x_0 = -\infty$ y $x_n = +\infty$.

Los puntos x_j tienen la propiedad básica de dividir el eje real en n intervalos igualmente probables. En efecto, sea $0 \leq j \leq n - 1$, entonces

$$\text{Prob}\{x_j \leq X \leq x_{j+1}\} = F(x_{j+1}) - F(x_j) = y_{j+1} - y_j = \frac{j+1}{n} - \frac{j}{n} = \frac{1}{n}; \quad (63)$$

es decir, todos los subintervalos son igualmente probables con probabilidad $1/n$.

Los n -tiles son ampliamente utilizados para la descripción de distribuciones de probabilidad, especialmente para ciertos valores particulares de n –en donde reciben nombres especiales– tal como se detalla a continuación.

Mediana. Corresponde al caso $n = 2$. En este caso $y_1 = 0,5$, y x_1 es la denominada *mediana* de la distribución, es decir, el punto donde la distribución acumulativa vale $1/2$, y que, por lo tanto, divide al eje x en dos semiejes igualmente probables. Ver representación en la figura 5.

Cuartiles. Corresponden al caso $n = 4$. x_1, x_2 y x_3 definen el primer, segundo y tercer cuartil, respectivamente. Ver representación en la figura 5.

Percentiles. Corresponden al caso $n = 100$. x_1, x_2, \dots, x_{99} , respectivamente llamados “percentilo 1”, “percentilo 2”, \dots , “percentilo 99”, definen el conjunto de abscisas que dividen al eje x en cien subintervalos igualmente probables.

7.5. Esperanzas y momentos de distribuciones conjuntas

Sea una distribución $h(x, y)$ la densidad de probabilidades conjunta de dos variables aleatorias X e Y . Sea una función $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, entonces se define

$$E[g(X, Y)] = \langle g(X, Y) \rangle = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(x, y) h(x, y) dx dy. \quad (64)$$

Consideremos algunos casos particulares:

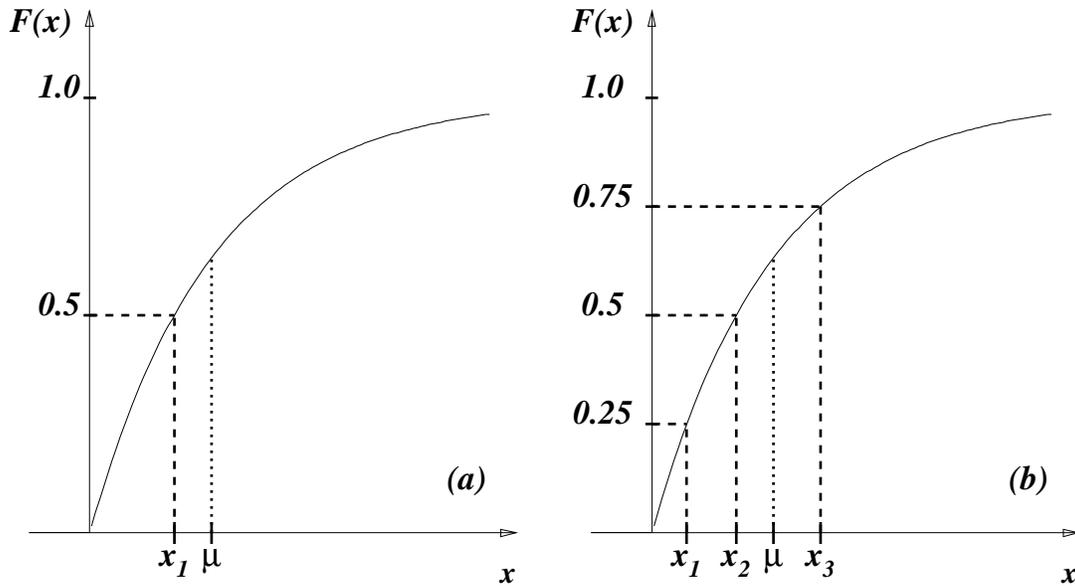


Figura 5: (a) Función de distribución acumulativa en donde se señalan la mediana x_1 y la media μ . (b) Lo mismo que (a), pero para los cuartiles x_1 , x_2 y x_3 .

1. $g(x, y) = x$. En este caso resulta:

$$\begin{aligned}
 E(X) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xh(x, y) dx dy \\
 &= \int_{-\infty}^{+\infty} x \left[\int_{-\infty}^{+\infty} h(x, y) dy \right] dx
 \end{aligned} \tag{65}$$

$$= \int_{-\infty}^{+\infty} x f_X(x) dx. \tag{66}$$

2. $g(x, y) = x + y$. Por las propiedades del cálculo integral resulta:

$$E(X + Y) = E(X) + E(Y). \tag{67}$$

3. $g(x, y) = (x + y)^2$. En este caso se verifica:

$$E[(X + Y)^2] = E(X^2) + E(Y^2) + 2E(XY). \tag{68}$$

Definiendo $V_{X+Y} = E[(X + Y)^2] - E^2[(X + Y)]$, se obtiene

$$V_{X+Y} = V_X + V_Y + 2 \text{Cov}(X, Y), \tag{69}$$

en donde la **covarianza** de dos variables aleatorias, $\text{Cov}(X, Y)$ se define via

$$\text{Cov}(X, Y) = E(XY) - E(X)E(Y). \tag{70}$$

4. Si X e Y son independientes y $g(x, y) = xy$, se tiene $h(x, y) = f_X(x)f_Y(y)$, y entonces

$$E(XY) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} xyf_X(x)f_Y(y) dydx \quad (71)$$

$$= \left[\int_{-\infty}^{+\infty} xf_X(x)dx \right] \left[\int_{-\infty}^{+\infty} yf_Y(y)dy \right], \quad (72)$$

es decir,

$$E(XY) = E(X)E(Y). \quad (73)$$

Esta expresión nos dice que la esperanza del producto de dos variables aleatorias *independientes* es igual al producto de las esperanzas de los factores.

5. (Corolario del punto anterior) Si X e Y son variables aleatorias independientes, entonces $\text{Cov}(X, Y) = 0$, y entonces (69) se reduce a

$$V_{X+Y} = V_X + V_Y. \quad (74)$$

A partir de la covarianza, definida en la ecuación (70), puede definirse el **coeficiente de correlación**

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sigma_X \sigma_Y}. \quad (75)$$

Puede demostrarse (hacerlo!) que $-1 \leq \rho \leq 1$.

El coeficiente ρ da una medida del grado de independencia (o, alternativamente, de correlación) existente entre dos variables aleatorias dadas: Cuando ρ es cercano a cero, las variables son prácticamente independientes, mientras que $|\rho|$ cercano a uno indica un alto grado de correlación.

8. Función característica de una distribución

La transformada de Fourier de una distribución de probabilidades, f_X , es frecuentemente utilizada, y se la refiere con el nombre de **función característica**

$$\begin{cases} \phi_X(u) = \int_{-\infty}^{+\infty} e^{iux} f_X(x) dx \\ f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-iux} \phi_X(u) du \end{cases} \quad (76)$$

Ejemplo. La distribución normal $N(0, 1)$ (distribución (46) con $m = 0$ y $\sigma = 1$). En este caso la integral para calcular ϕ puede llevarse a la forma

$$\phi_{N(0,1)}(u) = \frac{e^{-u^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}(x-iu)^2} dx \quad (77)$$

que nos da

$$\phi_{N(0,1)} = e^{-u^2/2}. \quad (78)$$

8.1. Algunas propiedades de las funciones características

Se listan a continuación algunas propiedades básicas de las funciones características. Estas propiedades pueden demostrarse fácilmente (hacerlo!) utilizando la definición de función característica y aplicando algunas características ya descritas de las variables aleatorias que intervienen en cada caso.

1. Escalado y traslación:

$$\phi_{aX+b}(u) = e^{iub}\phi_X(au). \quad (79)$$

2. Suma de variables aleatorias:

$$\phi_{X+Y}(u) = \phi_X(u)\phi_Y(u). \quad (80)$$

3. Si existe el momento k -simo de X , entonces

$$E(X^k) = \frac{1}{i^k} \frac{d^k \phi_X}{du^k}(0) \quad (81)$$

Esta última propiedad permite escribir el desarrollo de Taylor de $\phi_X(u)$ alrededor de $u = 0$ de la siguiente manera:

$$\phi_X(u) = 1 + i \frac{E(X)}{1!} u - \frac{E(X^2)}{2!} u^2 - i \frac{E(X^3)}{3!} u^3 + \dots \quad (82)$$

9. Ley de los Grandes Números y Teorema Central del Límite

Sean X_1, X_2, \dots, X_n una serie de variables aleatorias independientes, y sea

$$S_n = \frac{1}{n} \sum_{j=1}^n X_j. \quad (83)$$

S_n suele ser llamada “media muestral”.

Dado que las variables son independientes, podemos aplicar la ecuación (74) reiteradamente para obtener

$$V(S_n) = \frac{1}{n^2} \sum_{j=1}^n V(X_j). \quad (84)$$

Por otra parte, aplicando (67), podemos ver que

$$E(S_n) = \frac{1}{n} \sum_{j=1}^n E(X_j). \quad (85)$$

En el caso particular en que todas las variables aleatorias X_j tienen la misma media y desviación estandar, es decir $E(X_j) = \mu$, y $V(X_j) = \sigma^2$ para todo j (con μ y σ finitos), entonces resulta:

$$E(S_n) = \frac{n\mu}{n} = \mu, \quad V(S_n) = \frac{n\sigma^2}{n^2} = \frac{\sigma^2}{n}, \quad (86)$$

de donde surge que S_n tiene esperanza μ (la misma que las X_j), y desviación estandar σ/\sqrt{n} , que es más reducida que la desviación estandar de las variables X_j , y tiende a cero con n tendiendo a infinito.

Esto significa que para n tendiendo a infinito, la media muestral se distribuye justo sobre μ con desviación estandar nula, independientemente de la naturaleza de las variables X_j con tal que todas ellas tengan media y desviación estandar finita.

Este hecho es conocido bajo el nombre de **ley de los grandes números**, y fue descubierto por primera vez por Bernouilli. La ley de los grandes números tiene validez general, excepto para algunas distribuciones con momentos no definidos (por ejemplo la distribución de Cauchy).

Es posible ir aún más allá y demostrar que S_n tiene distribución normal en el límite $n \rightarrow \infty$, y este es, aproximadamente, el enunciado del *Teorema Central del Límite*.

Para demostrarlo, procedamos primero a definir

$$S_n^* = \frac{1}{\sqrt{n}} \sum_{j=1}^n X_j^* \quad (87)$$

donde

$$X_j^* = \frac{X_j - \mu_j}{\sigma_j}, \quad j = 1, \dots, n. \quad (88)$$

μ_j y σ_j son, respectivamente, la media y la desviación estandar de X_j , ambas finitas. X_j^* y S_j^* suelen llamarse las estandarizaciones de X_j y S_n , respectivamente. Es evidente que ambas variables estandarizadas tienen media 0 y varianza 1.

Para demostrar el Teorema Central del Límite bastará demostrar que para $n \rightarrow \infty$ la distribución de probabilidades de S_n^* es normal de media 0 y desviación estandar 1, $N(0, 1)$.

Aplicando la definición de la función característica, y algunas de las propiedades listadas en la sección 8.1, se puede ver que

$$\phi_{S_n^*} = \phi \left[\frac{1}{\sqrt{n}} \sum_{j=1}^n X_j^* \right] = \phi \left[\sum_{j=1}^n \frac{X_j^*}{\sqrt{n}} \right], \quad (89)$$

y entonces, aplicando (79) y (80)

$$\phi_{S_n^*}(u) = \prod_{j=1}^n \phi_{X_j^*} \left(\frac{u}{\sqrt{n}} \right). \quad (90)$$

y por lo tanto

$$\ln \phi_{S_n^*}(u) = \sum_{j=1}^n \ln \phi_{X_j^*} \left(\frac{u}{\sqrt{n}} \right). \quad (91)$$

Para u fijo, u/\sqrt{n} tiende a cero con n tendiendo a infinito, luego podemos utilizar convenientemente el desarrollo de Taylor (82):

$$\phi_{X_j^*} \left(\frac{u}{\sqrt{n}} \right) = 1 - \frac{1}{2!} \frac{u^2}{n} + A_j \left(\frac{u}{\sqrt{n}} \right)^3 \quad (92)$$

(notar que se ha utilizado el hecho que $E(X_j^*) = 0$ y que $E[(X_j^*)^2] = 1$). El factor A_j engloba a todos los términos de la serie no considerados (el de menor grado es cúbico).

Entonces se verifica:

$$\ln \phi_{S_n^*}(u) = \sum_{j=1}^n \ln \left[1 - \frac{u^2}{2n} + \frac{A_j u^3}{n^{3/2}} \right] \quad (93)$$

En el límite para $n \rightarrow \infty$ los argumentos de los logaritmos del segundo miembro tienden a 1, y por lo tanto se justifica reemplazar estos logaritmos por su desarrollo de Taylor alrededor de 1, resultando

$$\ln \phi_{S_n^*}(u) = \sum_{j=1}^n \left[-\frac{u^2}{2n} + \frac{A_j u^3}{n^{3/2}} \right] = -\frac{u^2}{2} + \frac{u^3}{n^{3/2}} \sum_{j=1}^n A_j. \quad (94)$$

Sea $A = \frac{1}{n} \sum_{j=1}^n A_j$, es decir $\sum_{j=1}^n A_j = nA$. Cuando n tiende a infinito, todos los A_j se tornan virtualmente iguales a $-i\mu_j^3/3!$, y por lo tanto A resulta ser un número finito perfectamente determinado. Al reemplazar en (94) obtenemos

$$\ln \phi_{S_n^*}(u) = -\frac{u^2}{2} + \frac{u^3 A}{\sqrt{n}}, \quad (95)$$

es decir

$$\lim_{n \rightarrow \infty} \ln \phi_{S_n^*}(u) = -\frac{u^2}{2}, \quad (96)$$

que al exponenciar resulta

$$\lim_{n \rightarrow \infty} \phi_{S_n^*}(u) = e^{-\frac{u^2}{2}}. \quad (97)$$

Comparando con el ejemplo desarrollado en la sección 8 podemos ver fácilmente que el segundo miembro de la ecuación (97) es la función característica de la distribución $N(0, 1)$, lo que implica entonces, al antitransformar, que la distribución de probabilidades de S_n^* es la normal de media cero y desviación estandar 1, que es el resultado al que se quería arribar. Esto completa la demostración del Teorema Central del Límite.

Observese que se ha hecho uso de la hipótesis de existencia y finitud de la desviación estandar de las variables X_j . Si esto no se cumple, entonces ya no se puede garantizar la validez del Teorema.