

---

◁ **Apunte de Probabilidades** ▷  
Por Sebastián Soto Rojas (spsoto@uc.cl)

---

Estos apuntes se encuentran basados en el texto *Mathematical Statistics and Data Analysis* de John Rice (3a edición, Duxbury Advanced Series) y se dejan a disposición de todos aquellos quienes deseen repasar estos tópicos para su preparación en el curso IEE2513 Comunicaciones e ICS2123 Modelos Estocásticos.

## Índice

<b>1. Probabilidad</b>	<b>3</b>
1.1. Introducción . . . . .	3
1.2. Espacios muestrales . . . . .	3
1.3. Medidas de probabilidad . . . . .	4
1.4. Cómputo de probabilidades: métodos de conteo . . . . .	5
1.5. Probabilidad condicional . . . . .	7
1.6. Independencia estadística . . . . .	8
<b>2. Variables aleatorias</b>	<b>9</b>
2.1. Variables aleatorias discretas . . . . .	9
2.2. Variables aleatorias continuas . . . . .	12
2.3. Funciones de variables aleatorias . . . . .	16
<b>3. Distribuciones conjuntas</b>	<b>18</b>
3.1. Introducción . . . . .	18
3.2. Variables aleatorias discretas . . . . .	19
3.3. Variables aleatorias continuas . . . . .	19
3.4. Variables aleatorias independientes . . . . .	20
3.5. Distribuciones condicionales . . . . .	21
3.6. Funciones de variables aleatorias distribuidas conjuntamente . . . . .	22
3.7. Valores extremos y estadísticos de orden . . . . .	24

<b>4. Valores esperados</b>	<b>26</b>
4.1. El valor esperado de una variable aleatoria . . . . .	26
4.2. Varianza y desviación estándar . . . . .	30
4.3. Covarianza y correlación . . . . .	33
4.4. Esperanza condicional y predicción . . . . .	37
4.5. La función generadora de momentos . . . . .	39
<b>5. Teoremas de los límites</b>	<b>43</b>
5.1. La ley de números grandes . . . . .	43
5.2. Convergencia en la distribución y el teorema del límite central . . . . .	45
<b>6. Distribuciones derivadas de la distribución normal</b>	<b>51</b>
6.1. Distribución $\chi^2$ . . . . .	51
6.2. Distribución $t$ -student . . . . .	51
6.3. Distribución $F$ . . . . .	52

# 1. Probabilidad

## 1.1. Introducción

La idea de probabilidad es antigua, pero su formulación matemática ocurrió de forma relativamente reciente, con sus principales ideas originadas en los juegos de azar.

Durante este siglo la probabilidad ha sido aplicada a una gran variedad de fenómenos, con ejemplos de ellos:

- Genética, para estudiar mutaciones y asegurar variabilidad natural.
- La teoría cinética de gases.
- Análisis de colas en los sistemas informáticos.
- Teorías para analizar el ruido en sistemas electrónicos y dispositivos de comunicaciones.
- En investigación de operaciones, para modelar las demandas de inventarios de bienes.
- Análisis financieros y de compañías de seguros.
- Para analizar sistemas complejos y mejorar su desempeño y confiabilidad, tal como en aeronaves comerciales y militares.

Primero se analizará la teoría de probabilidades como modelos matemáticos para fenómenos aleatorios. Luego se analizará la estadística, preocupada de procedimientos para analizar datos, particularmente aquellos con un carácter aleatorio, por lo que es un requisito comprender antes probabilidades.

## 1.2. Espacios muestrales

**Definición:**

- Para una situación que ocurre de forma aleatoria, una realización se conoce como *experimento*.
- El conjunto de todos los posibles resultados del experimento se conoce como *espacio muestral* y se denota por  $\Omega$ , así como un elemento se denota por  $\omega$ .
- Cuando nos interesamos por subconjuntos en particular de  $\Omega$ , estos se conocen como *eventos*.
- Se pueden aplicar las definiciones de teoría de conjuntos en el espacio muestral así como sus eventos, incluyendo unión, intersección, complemento y resta. También puede considerarse el evento vacío.
- Dos eventos se dicen *disjuntos* si  $A_1 \cap A_2 = \emptyset$ .
- Un conjunto  $\{A_1, A_2, \dots, A_n\}$  se dice mutuamente excluyente si para cada  $i, k$  se tiene que:

$$A_i \cap A_k = \emptyset \longleftrightarrow i \neq k$$

### 1.3. Medidas de probabilidad

**Definición:** Una *medida de probabilidad* de  $\Omega$  es una función  $P : A \subseteq \Omega \rightarrow \mathbb{R}$  que satisface los siguientes axiomas:

1. Como  $\Omega$  consiste en todos los elementos, entonces  $P(\Omega) = 1$ .
2. La probabilidad no es negativa. Si  $A \subset \Omega$ , entonces  $P(A) \geq 0$ .
3. Si  $A_1$  y  $A_2$  son eventos disjuntos, entonces

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

En general, si  $A_1, A_2, \dots, A_n$  son mutuamente excluyentes, entonces:

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

Como consecuencia de los axiomas se deducen las siguientes propiedades:

**Proposición:**

1.  $P(A^c) = 1 - P(A)$ . Se deduce pues  $A \cap A^c = \emptyset$  y  $A \cup A^c = \Omega$ .
2.  $P(\emptyset) = 0$ . Se deduce pues  $\emptyset = \Omega^c$ .
3. Si  $A \subset B$ , entonces  $P(A) \leq P(B)$ . Se prueba notando que:

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

Luego, del tercer axioma:

$$P(B) = P(A) + P(B \cap A^c) \rightarrow P(A) = P(B) - P(B \cap A^c) \leq P(B)$$

4. Ley de adición:  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ . Para demostrarlo, notamos que:

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

Como estos tres conjuntos son disjuntos:

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

Pero  $A \setminus B = A \setminus (A \cap B) \rightarrow P(A \setminus B) = P(A) - P(A \cap B)$  y de forma análoga para  $B \setminus A$ . De esta forma,

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

donde se descuenta una vez la intersección pues esta se considera dos veces al sumar ambas probabilidades.

## 1.4. Cómputo de probabilidades: métodos de conteo

Para espacios muestrales finitos las probabilidades son especialmente fáciles de calcular, enumerando y contando los eventos favorables, de modo que:

$$P(A) = \frac{\text{números de formas que } A \text{ puede ocurrir}}{\text{número total de ocurrencias}}$$

En algunos ejemplos es fácil contar el número de ocurrencias y calcular las probabilidades, pero para situaciones más complejas se deben desarrollar formas de contar las ocurrencias.

### 1.4.1. El principio de multiplicación

**Teorema:** Si un experimento tiene  $m$  realizaciones favorables y otro experimento tiene  $n$  realizaciones favorables, entonces hay  $mn$  realizaciones favorables para ambos experimentos.

*Demostración:*

Sean  $a_1, \dots, a_m$  las realizaciones del primer experimento y  $b_1, \dots, b_n$  las realizaciones del segundo experimento. Las realizaciones de ambos experimentos son pares de la forma  $(a_i, b_j)$ , que por lo tanto son entradas de un arreglo cuadrangular de  $m \times n$ , en la cual el par  $(a_i, b_j)$  se encuentra en la fila  $i$ -ésima y columna  $j$ -ésima. Por lo tanto hay  $mn$  entradas en el arreglo. ■

El principio puede extenderse a  $k$  experimentos, donde cada experimento tiene  $n_i$  realizaciones. Por lo tanto, existirá un total de:

$$\prod_{i=1}^k n_i$$

### 1.4.2. Permutaciones

**Definición:** Se define una *permutación* como un arreglo ordenado de objetos. Si no se permite el duplicado de objetos, se conoce como muestreo sin reemplazo. Si se permite duplicado de objetos, se conoce como muestreo con reemplazo.

Se puede utilizar el principio de multiplicación para contar el número de diferentes muestras en un conjunto de  $n$  elementos. Si suponemos que el muestreo está realizado sin reemplazo, el primer elemento puede ser escogido de  $n$  formas, el segundo de  $n-1$  formas y así sucesivamente hasta la  $r$ -ésima iteración con  $n-r+1$  formas, probando así la siguiente proposición:

$$\underbrace{n \times n \times \cdots \times n}_r = n^r$$

Supongamos ahora que el muestreo se realiza con reemplazo. La primera muestra puede ser escogida de  $n$  formas, la segunda de  $n$  formas, la tercera de  $n$  formas y así sucesivamente hasta la  $r$ -ésima iteración con  $n$  formas, probando así la siguiente proposición:

**Proposición:** Para un conjunto de tamaño  $n$  y una muestra de tamaño  $r$ , existen  $n^r$  muestras ordenadas con reemplazo y  $n(n-1)(n-2)\cdots(n-r+1)$  muestras ordenadas sin reemplazo.

**Corolario:** Existen  $n!$  formas de ordenar  $n$  elementos distintos.

### 1.4.3. Combinaciones

Si ahora no estamos interesados en obtener muestras de forma ordenada, si no que más bien de los elementos obtenidos sin importar el orden en que se obtuvieron, estamos interesados en calcular el número de combinaciones. Si  $r$  objetos son tomados de un conjunto de  $n$  objetos sin reemplazo y sin importar el orden, ¿cuántas combinaciones son posibles?

De acuerdo al principio de multiplicación, el número de muestras ordenadas obtenidas equivale a  $n(n-1)\cdots(n-r+1)$  y como cada muestra de tamaño  $r$  puede ser ordenada de  $r!$  formas, entonces el número de muestras no ordenadas es:

$$\frac{n(n-1)\cdots(n-r+1)}{r!} = \frac{n!}{(n-r)!r!} \triangleq \binom{n}{r}$$

**Proposición:** El número de muestras ordenadas de  $r$  objetos seleccionadas de un conjunto de  $n$  objetos sin reemplazo es  $\binom{n}{r}$ .

**Definición:** Los números  $\binom{n}{k}$ , conocidos como *coeficientes binomiales*, ocurren en la expansión:

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

**Proposición:** En particular, puede notarse que:

$$2^n = \sum_{k=0}^n \binom{n}{k}.$$

lo cual puede interpretarse como la suma total de conjuntos posibles, de todos los tamaños, sin importar el orden, en un conjunto de  $n$  elementos: es la suma de combinaciones de 0 elementos, de 1 elemento, de 2 elementos y así hasta los  $n$  elementos.

Puede extenderse este resultado como sigue:

**Proposición:** El número de formas en que  $n$  objetos pueden ser agrupadas en  $r$  clases de  $n_i$  elementos la clase  $i$ -ésima con  $\sum_{i=1}^r n_i = n$  es:

$$\binom{n}{n_1 n_2 \cdots n_r} = \frac{n!}{n_1! n_2! \cdots n_r!}$$

*Demostración:*

Existen  $\binom{n}{n_1}$  formas de escoger los objetos de la primera clase,  $\binom{n-n_1}{n_2}$  formas de escoger la segunda,  $\binom{n-n_1-n_2}{n_3}$  formas de escoger la tercera. De acuerdo al principio de multiplicación, el total vendrá dado por:

$$\binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \cdots \binom{n-n_1-\cdots-n_{r-1}}{n_r} = \frac{n!}{\cancel{(n-n_1)!} n_1! \cancel{(n-n_1-n_2)!} n_2! \cdots \cancel{(n-n_1-\cdots-n_{r-1})!} 0! n_r!}$$

Es decir,

$$\binom{n}{n_1} \binom{n-n_1}{n_2} \binom{n-n_1-n_2}{n_3} \cdots \binom{n-n_1-\cdots-n_{r-1}}{n_r} = \frac{n!}{n_1! n_2! \cdots n_r!} \quad \blacksquare$$

**Definición:** Los números  $\binom{n}{n_1 \cdots n_r}$  se conocen como *coeficientes multinomiales* y ocurren en la expansión:

$$(x_1 + x_2 + \cdots + x_r)^n = \sum \binom{n}{n_1 n_2 \cdots n_r} x_1^{n_1} x_2^{n_2} \cdots x_r^{n_r}$$

donde  $n_1, \dots, n_r$  son tales que  $n_1 + \cdots + n_r = n$ .

## 1.5. Probabilidad condicional

**Definición:** Sean  $A$  y  $B$  dos eventos tales que  $P(B) \neq 0$ , se define la *probabilidad condicional de  $A$  dado  $B$*  como:

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

La idea de esta definición es que, dado que ocurrió el evento  $B$ , el espacio relevante es en realidad  $B$  en vez de  $\Omega$ , y la probabilidad condicional es una medida de probabilidad con respecto al espacio en  $B$ .

Dado esto, se deduce la *ley de multiplicación*, que nos permite calcular  $P(A \cap B)$  dado  $P(A|B)$  y  $P(B)$ . Si  $A$  y  $B$  son eventos y  $P(B) \neq 0$ , entonces:

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

**Teorema:** (*Regla de Bayes*) Sea  $A$  y  $B_1, \dots, B_n$  eventos donde  $B_i$  es disjunto,  $\bigcup_{i=1}^n B_i = \Omega$  y  $P(B_i) > 0$  para todo  $i$ . Entonces,

$$P(B_j|A) = \frac{P(A|B_j) P(B_j)}{\sum_{i=1}^n P(A|B_i) P(B_i)}$$

## 1.6. Independencia estadística

**Definición:** Dos eventos se dicen estadísticamente independientes si la ocurrencia de uno no entrega información sobre la ocurrencia del otro. En otras palabras,  $A$  y  $B$  son *independientes* si y solo si:

$$P(A|B) = P(A) \quad \text{y} \quad P(B|A) = P(B)$$

Luego, si

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

entonces

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

**Definición:** Se dice que una colección de eventos  $A_1, A_2, \dots, A_n$  son *mutuamente independientes* si para cualquier colección  $A_{i_1}, \dots, A_{i_m}$

$$P(A_{i_1} \cap \dots \cap A_{i_m}) = P(A_{i_1}) \dots P(A_{i_m})$$



## 2. Variables aleatorias

### 2.1. Variables aleatorias discretas

#### Definición:

- Una *variable aleatoria* es básicamente un número aleatorio. En general, una variable aleatoria es una función de  $\Omega$  a los números reales. Como el resultado del experimento en  $\Omega$  es aleatorio, el número producido por la función es asimismo aleatorio.
- Usaremos letras itálicas en mayúsculas para denotar las variables aleatorias.
- Una variable aleatoria *discreta* es una variable aleatoria que puede tomar un número finito o a lo más contable de valores.
- En general, se mide la probabilidad de que la variable aleatoria cumpla cierta condición lógica. A modo de ejemplo,

$$P(X = a) \quad ; \quad P(X \leq a) \quad ; \quad P(X > a)$$

**Definición:** En general, la medida de probabilidad en el espacio muestral determina las probabilidades de los distintos valores de  $X$ . Si estos valores son denotados por  $x_1, \dots, x_n$ , entonces

1. Existe una función  $p(x_i) = P(X = x_i)$  y  $\sum p(x_i) = 1$ . Esta función se conoce como la *función de probabilidad de masa* o la *función frecuencia* de la variable aleatoria  $X$ .
2. Se define la *función de distribución acumulada (cdf)* de una variable aleatoria como

$$F(x) = P(X \leq x), \quad -\infty < x < \infty$$

3. Dos variables  $X$  e  $Y$  se dicen *independientes* si y solo si

$$P(X = x_i \wedge Y = y_j) = P(X = x_i) P(Y = y_j)$$

#### 2.1.1. Variables Bernoulli

Una *variable Bernoulli* toma exclusivamente dos valores: 0 y 1, con probabilidades  $1 - p$  y  $p$  respectivamente. La función de frecuencia es por lo tanto,

$$X \sim \text{Bernoulli}(p) \longleftrightarrow p(x) = \begin{cases} p^x (1 - p)^{1-x}, & \text{si } x = 0 \text{ y } x = 1, \\ 0, & \text{en otro caso.} \end{cases}$$

### 2.1.2. La distribución binomial

Supongamos que se realizan  $n$  experimentos o pruebas donde cada experimento resulta en éxito con probabilidad  $p$  y en falla con probabilidad  $1 - p$ . El número total de éxitos,  $X$ , es una variable aleatoria *binomial* con parámetros  $n$  y  $p$ .

Se tiene que:

$$X \sim \text{Binomial}(n, p) \longleftrightarrow p(k) = P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$$

donde  $p^k (1 - p)^{n-k}$  es la probabilidad de que ocurran  $k$  éxitos y  $n - k$  fallas, tal como se espera y  $\binom{n}{k}$  cuenta las  $\binom{n}{k}$  formas o secuencias sobre las cuales se puede obtener los  $k$  éxitos.

Se observa entonces que si  $X_1, \dots, X_n$  son variables Bernoulli independientes entre ellas, todas con probabilidad de éxito  $p$ , entonces

$$Y = X_1 + X_2 + \dots + X_n \sim \text{Binomial}(n, p)$$

### 2.1.3. Las distribuciones geométrica y binomial negativa

- La *distribución geométrica* también se construye a partir de intentos Bernoulli, pero considerando ahora una sucesión infinita. La variable aleatoria  $X$  consiste en el número de intentos hasta obtener el primer éxito. Es decir, deben ocurrir  $k - 1$  intentos fallidos y 1 intento exitoso, con lo cual

$$X \sim \text{Geométrica}(p) \longleftrightarrow P(X = k) = (1 - p)^{k-1} p$$

Puede notarse que estas probabilidades suman uno cuando se suman todos los casos posibles hasta infinito:

$$\sum_{k=1}^{\infty} (1 - p)^{k-1} p = p \sum_{k=0}^{\infty} (1 - p)^k = \frac{p}{1 - (1 - p)} = 1$$

- La *distribución binomial negativa* consiste en una generalización de la distribución geométrica. Sea  $X$  ahora el número total de intentos para obtener  $r$  éxitos en total. Entonces, cada intento en particular tiene probabilidad  $p^r (1 - p)^{k-r}$ , pero para los  $k - 1$  intentos restantes con  $r - 1$  probabilidades de éxito existen  $\binom{k-1}{r-1}$  formas. Luego,

$$X \sim \text{Bin. Neg.}(p, r) \longleftrightarrow P(X = k) = \binom{k-1}{r-1} p^r (1 - p)^{k-r}$$

Se observa que la distribución binomial negativa puede expresarse como la suma de  $r$  variables aleatorias geométricas.

### 2.1.4. La distribución hipergeométrica

Suponiendo un conjunto de  $n$  elementos de dos tipos exclusivamente, habiendo  $r$  del primer tipo y  $n - r$  del segundo tipo. Si  $X$  denota el número de intentos del primer tipo  $m$ , sin reemplazo, entonces  $X$  es una variable aleatoria hipergeométrica:

$$X \sim \text{H.G.}(r, n, m) \longleftrightarrow P(X = k) = \frac{\binom{r}{k} \binom{n-r}{m-k}}{\binom{n}{m}}$$

donde  $\binom{r}{k}$  denota las formas de obtener los  $k$  elementos del primer tipo en el grupo,  $\binom{n-r}{m-k}$  las formas de obtener los  $m - k$  elementos del segundo tipo y  $\binom{n}{m}$ , el denominador, toma en consideración las  $\binom{n}{m}$  formas de obtener un grupo de  $m$  elementos sobre el total de  $n$  elementos.

### 2.1.5. La distribución Poisson

La función de frecuencia Poisson se obtiene de tomar el límite de la distribución binomial en infinito para el número de intentos  $n$  y  $p$  para la probabilidad de éxito de la forma  $np = \lambda$ . Es decir,

$$\begin{aligned} p(k) &= \lim_{\substack{n \rightarrow \infty \\ p \rightarrow \infty \\ np = \lambda}} \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} = \lim_{\substack{n \rightarrow \infty \\ p \rightarrow \infty \\ np = \lambda}} \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \lim_{\substack{n \rightarrow \infty \\ p \rightarrow \infty \\ np = \lambda}} \frac{\lambda^k}{k!} \frac{n!}{(n-k)! n^k} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \end{aligned}$$

De esta forma,

$$X \sim \text{Poisson}(\lambda) \longleftrightarrow P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$$

Nuevamente, puede notarse que las probabilidades suman 1:

$$F(\infty) = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda} = e^{-\lambda} e^{\lambda} = 1$$

La distribución Poisson puede ser utilizada para aproximar distribuciones binomiales en que  $n$  es grande y  $p$  pequeño, lo cual a su vez sugiere cómo aparecen estas distribuciones en la práctica. Si  $X$  es una variable aleatoria que equivale al número de veces que un evento ocurre en un intervalo de tiempo dado.

Heurísticamente, si dividimos el intervalo en un número grande de subintervalos de igual longitud, y asumimos que los subintervalos son tan pequeños que la probabilidad de que más de un evento en el subintervalos es despreciable relativo a la probabilidad de un evento, que por si misma es muy pequeña. Asumamos que la probabilidad de un evento es la misma en cada subintervalo y

que cuando ocurre un evento en el subintervalo es independiente de lo que ocurre en los próximos subintervalos, entonces la variable  $X$  es prácticamente una variable binomial, con los subintervalos consistiendo en los intentos, y por el resultado anterior  $X$  tiene casi una distribución Poisson.

Los supuestos importantes son:

1. La distribución no tiene memoria: lo que ocurre en un subintervalo es independiente de lo que ocurre en otro subintervalo.
2. La probabilidad de un evento es la misma en cada subintervalo.
3. Los eventos no ocurren simultáneamente.

La distribución Poisson tiene importancia teórica y práctica fundamental. Puede ser utilizada en muchas áreas, incluyendo las siguientes:

- **Análisis de sistemas telefónicos.** El número de llamadas llegando a una unidad durante una unidad de tiempo puede modelarse como una variable Poisson si la unidad entrega servicio a un gran número de clientes que actúan más o menos independientemente.
- **Radioactividad.** Para modelar el número de partículas alfa emitidas por una fuente radioactiva durante cierto período de tiempo.
- **Compañías de seguros.** La distribución Poisson ha sido como modelo en las compañías de seguros. Por ejemplo, el número accidentes extraños, como caídas en la ducha, para grandes grupos de población.
- **Ingeniería de tránsito.** La distribución Poisson ha sido utilizada como un modelo para el tráfico liviano. El número de vehículos que pasa un marcador en una autopista durante una unidad de tiempo puede ser contada. Si el tráfico es liviano, los vehículos actúan de forma independiente entre ellos. En tráfico alto, sin embargo, esto no sucede porque el movimiento de un auto puede influir en los otros.

## 2.2. Variables aleatorias continuas

En muchas aplicaciones es de interés estudiar variables aleatorias que pueden tomar un continuo de valores en vez de un conjunto contable de estos.

### Definición:

- Una *variable aleatoria continua* es una variable aleatoria en la cual  $X$  puede tomar el valor de cualquier número real.
- En analogía a la función de frecuencia, se dice que  $f(x)$  es la función de *densidad de la variable aleatoria* (pdf) que tiene la propiedad de ser no negativa, continua a tramos y tal que

$$\int_{-\infty}^{\infty} f(x) dx = 1$$

- Si  $X$  es una variable aleatoria con función de densidad  $f$ , entonces para todo  $a < b$  la probabilidad de que  $X$  esté en el intervalo  $(a, b)$  viene dada por:

$$P(a < X < b) = \int_a^b f(x) dx$$

- Si  $F$  es la cdf de una variable continua y es estrictamente creciente en el intervalo  $I$  y que  $F = 0$  a la izquierda y  $F = 1$  a la derecha de  $I$ , entonces  $I$  puede ser no acotado. Bajo este supuesto, la función  $F^{-1}$  está bien definida, i.e.  $x = F^{-1}(y)$  si  $y = F(x)$ , entonces el *cuantil*  $p$ -ésimo se define como el valor  $x_p$  está definido únicamente como  $x_p = F^{-1}(p)$ .
- El caso  $p = \frac{1}{2}$  se conoce como la *mediana* de  $F$ , y  $p = \frac{1}{4}$  y  $p = \frac{3}{4}$ , los cuales corresponden a los *cuartiles inferior* y *superior* de  $F$ .

Una consecuencia de la definición anterior es que la probabilidad de que una variable continua tome un valor en particular es cero, i.e.

$$P(X = c) = \int_c^c f(x) dx = 0$$

Esto implica que:

$$P(a < x < b) = P(a \leq x \leq b)$$

Para un pequeño  $\epsilon$ , si  $f$  es continua en  $x$ , entonces de acuerdo al teorema del valor medio integral se tiene que:

$$P\left(x - \frac{\epsilon}{2} \leq X \leq x + \frac{\epsilon}{2}\right) = \int_{x-\epsilon/2}^{x+\epsilon/2} f(t) dt \approx \epsilon f(x)$$

De aquí se sigue que la probabilidad de un intervalo pequeño cerca de  $x$  es proporcional a  $f(x)$ . Se dice entonces la notación:

$$P(a \leq X \leq b) = \int_a^b f(x) dx = F(b) - F(a)$$

### 2.2.1. La distribución uniforme

Respondiendo a la pregunta “escoge un número al azar entre 0 y 1”, se tiene una variable aleatoria cuya distribución es la uniforme entre  $[0, 1]$ . La densidad viene dada por:

$$f(x) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1, \\ 0, & \text{en otro caso.} \end{cases}$$

De este modo la integral en todo  $\mathbb{R}$  es 1 y asigna una densidad de probabilidad igual para cada número. En general, para el intervalo  $[a, b]$  la función de densidad es:

$$X \sim \text{Uniforme}(a, b) \longleftrightarrow f(x) = \begin{cases} 1/(b-a), & \text{si } a \leq x \leq b, \\ 0, & \text{en otro caso.} \end{cases}$$

La cdf en este caso viene dada por:

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

### 2.2.2. La distribución exponencial

La distribución exponencial tiene como función densidad:

$$X \sim \text{Exponencial}(\lambda) \longleftrightarrow f(x) = \begin{cases} \lambda e^{-\lambda x}, & \text{si } x \geq 0, \\ 0, & \text{si } x < 0. \end{cases}$$

donde  $\lambda > 0$  es el único parámetro al igual que en la distribución Poisson. La distribución acumulada viene dada por:

$$F(x) = \int_{-\infty}^x f(x)dx = \begin{cases} 1 - e^{-\lambda x}, & \text{si } x \geq 0, \\ 0, & \text{si } x < 0. \end{cases}$$

El despeje de la mediana es directo:

$$1 - e^{-\lambda \eta} = \frac{1}{2} \longrightarrow \eta = \frac{\log 2}{\lambda}$$

La distribución exponencial se usa para modelar tiempos de vida o tiempos de espera, reemplazando  $x$  por  $t$  en dicho contexto.

Se dice que la distribución exponencial no tiene memoria pues:

$$P(X > t + \alpha | X > \alpha) = \frac{P(X > t + \alpha \wedge X > \alpha)}{P(X > \alpha)} = \frac{e^{-(t+\alpha)}}{e^{-\alpha}} = e^{-t}$$

Es decir, no depende de  $\alpha$  la probabilidad condicional, y por lo tanto no “recuerda” el tiempo que ha transcurrido. La ausencia de memoria es de hecho una característica de la distribución.

**Relación con el proceso Poisson.** Supongamos que un evento ocurre en el tiempo como un proceso Poisson con parámetro  $\lambda$  en un instante  $t_0$ . Sea  $T$  el tiempo para la próxima ocurrencia del evento. Entonces,

$$P(T > t) = P(\text{sin eventos en } [t_0, t_0 + t])$$

Como el número de eventos en el intervalo  $(t_0, t_0 + t)$ , de largo  $t$ , sigue una distribución Poisson con parámetro  $\lambda t$ , la probabilidad es  $e^{-\lambda t}$  y por lo tanto  $T$  sigue una distribución de parámetro  $\lambda$ . Si el siguiente evento ocurre en un tiempo  $t_1$ , la distribución para la siguiente ocurrencia sigue siendo exponencial e independiente de la distancia de tiempo entre los eventos.

### 2.2.3. La distribución Gamma

La distribución Gamma tiene como densidad a una función que depende de dos parámetros,  $\alpha$  y  $\lambda$ :

$$T \sim \text{Gamma}(\alpha, \lambda) \longleftrightarrow g(t) = \frac{\lambda^\alpha}{\Gamma(\alpha)} t^{\alpha-1} e^{-\lambda t}, \quad t \geq 0$$

Para  $t < 0$ ,  $g(t) = 0$ . Se define la función Gamma como:

$$\Gamma(x) \triangleq \int_0^\infty u^{x-1} e^{-u} du, \quad x > 0$$

Si  $\alpha = 1$  se recupera la densidad exponencial. Este parámetro se denomina *parámetro de forma* y  $\lambda$  se denomina *parámetro de escala*.

### 2.2.4. La distribución normal

La distribución normal juega un rol central en probabilidades y estadística. También se conoce como distribución Gaussiana en honor al matemático Carl Friedrich Gauss, quien la propuso como modelo para los errores de medición. Ha sido utilizada para modelar un sin fin de situaciones reales.

La densidad depende de dos parámetros,  $\mu$  (media) y  $\sigma$  (desviación estándar) y su expresión es:

$$X \sim N(\mu, \sigma) \longleftrightarrow f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad x \in \mathbb{R}$$

La cdf no puede ser evaluada en forma cerrada para esta función de densidad. No tiene una fórmula explícita pero se computa numéricamente una tabla para la densidad normal estándar, donde  $\mu = 0$  y  $\sigma = 1$ . Se utiliza la variable  $\Phi(x)$  para referirse a esta función de distribución acumulada (y  $\phi$  para la densidad) y puede ser encontrada una tabla con diversos valores en el anexo.

### 2.2.5. La distribución Beta

La distribución beta es útil para modelar variables que se encuentran restringidas al intervalo  $[0, 1]$ :

$$X \sim B(a, b) \longleftrightarrow f(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} u^{a-1} (1-u)^{b-1}; \quad 0 \leq u \leq 1$$

Si  $a = b = 1$  se recupera la distribución uniforme.

## 2.3. Funciones de variables aleatorias

Supongamos que una variable aleatoria  $X$  tiene una función densidad  $f(x)$ . Surge la pregunta: ¿cuál es la función de densidad de  $Y = g(X)$  para una función  $g$  dada? A modo de ejemplo,  $X$  puede ser la velocidad de una partícula de masa  $m$  y estamos interesados en obtener la función de densidad probabilística para su energía cinética,  $Y = \frac{1}{2}mX^2$ .

Utilizaremos la notación  $f_X$  y  $F_X$  para referirnos a las distribuciones de  $X$  y  $f_Y$  y  $F_Y$  para referirnos a las de  $Y$ .

### 2.3.1. La distribución normal

Si  $X \sim N(\mu, \sigma^2)$  y que  $Y = aX + b$  donde  $a > 0$ , la probabilidad de densidad acumulada de  $Y$  es:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(aX + b \leq y) \\ &= P\left(X \leq \frac{y-b}{a}\right) \\ &= F_X\left(\frac{y-b}{a}\right) \end{aligned}$$

Luego,

$$f_Y(y) = \frac{d}{dy} F_X\left(\frac{y-b}{a}\right) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right)$$

Es decir,

$$f_Y(y) = \frac{1}{a\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-b-a\mu}{a\sigma}\right)^2\right]$$

Se sigue que  $X \sim N(\mu, \sigma^2) \longrightarrow Y \sim N(a\mu + b, a^2\sigma^2)$ .

Esta proposición es útil para encontrar probabilidades de la distribución normal. Supongamos que  $X \sim N(\mu, \sigma^2)$  y se desea buscar  $P(x_0 < X < x_1)$  para algunos números  $x_0$  y  $x_1$ . Se tiene que:

$$P(x_0 < X < x_1) = P\left(\frac{x_0 - \mu}{\sigma} < \frac{X - \mu}{\sigma} < \frac{x_1 - \mu}{\sigma}\right)$$

Se sigue que  $Z = (X - \mu) / \sigma \sim N(0, 1)$ , para la cual está calculada numéricamente en la tabla su distribución de probabilidad. Luego,

$$P(x_0 < X < x_1) = \Phi\left(\frac{x_1 - \mu}{\sigma}\right) - \Phi\left(\frac{x_0 - \mu}{\sigma}\right)$$



### 2.3.2. La distribución $\chi^2$

Es común que dado  $Z \sim N(0, 1)$  se desee encontrar la densidad de  $X = Z^2$ . En este caso, se tiene que:

$$\begin{aligned}F_X(x) &= P(X \leq x) \\&= P(-\sqrt{x} \leq Z \leq \sqrt{x}) \\&= \Phi(\sqrt{x}) - \Phi(-\sqrt{x})\end{aligned}$$

Derivando, y utilizando la regla de la cadena:

$$f_X(x) = \frac{1}{2\sqrt{x}}\phi(\sqrt{x}) + \frac{1}{2\sqrt{x}}\phi(-\sqrt{x}) = \frac{1}{\sqrt{x}}\phi(\sqrt{x})$$

donde se usó el hecho de que la distribución estándar es simétrica con respecto al origen. De esta forma,

$$X \sim \chi^2 \longleftrightarrow f_X(x) = \frac{1}{\sqrt{2\pi x}}e^{-x/2}, \quad x \geq 0$$

Se puede observar que esta es una función gamma con  $\alpha = \lambda = \frac{1}{2}$  y se conoce como *distribución  $\chi^2$*  con un grado de libertad.

### 2.3.3. Caso general

Observando estos ejemplos, se observan ciertas etapas básicas en cada caso: calcular la cdf de la variable transformada y luego diferenciar con respecto a la variable para encontrar la densidad.

**Proposición:** Sea  $X$  una variable aleatoria continua con densidad  $f(x)$  y sea  $Y = g(X)$  donde  $g$  es una función diferenciable y estrictamente monótona en un intervalo  $I$ . Supongamos que  $f(x)$  si  $x$  no está en  $I$ . Luego,  $Y$  tiene la función de densidad

$$f_Y(y) = f_X(g^{-1}(y)) \frac{dg^{-1}}{dy}(y)$$

Para las siguientes proposiciones, considérese una variable aleatoria cualquiera  $X$  con densidad  $f$  y distribución acumulada  $F$ .

**Proposición:**

- Sea  $Z = F(X)$ , entonces  $Z$  tiene una distribución uniforme en  $[0, 1]$ .
- Sea  $U$  una variable aleatoria uniforme en  $[0, 1]$  y sea  $X = F^{-1}(U)$ . Luego la cdf de  $X$  es  $F$ .

Estas proposiciones son útiles para generar números pseudoaleatorios dada una cdf. Los computadores suelen traer rutinas para generar números pseudoaleatorios con distribución uniforme en  $[0, 1]$ . Se dicen pseudoaleatorios porque son generados de acuerdo a cierta regla o algoritmo que no son en realidad aleatorios pero para los propósitos de modelación logran serlo. La segunda proposición plantea que basta aplicar  $F^{-1}$  para a los números uniformemente aleatorios para obtener los resultados.

### 3. Distribuciones conjuntas

#### 3.1. Introducción

En este capítulo nos preocupamos de la estructura de distribuciones conjuntas de dos o más variables aleatorias definidas en el mismo espacio muestral. Estas surgen de forma natural en diversas aplicaciones, entre ellas:

- Estudios ecológicos: diversas especies y sus interacciones.
- Climatología: cada componente de velocidad del viento.
- Fisiología: distribución conjunta de diversas variables en una población.
- Estudio de poblaciones: distribuciones de edad.

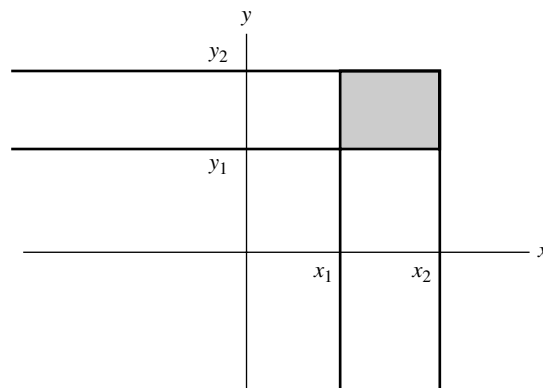
El comportamiento conjunto de dos variables aleatorias,  $X$  e  $Y$  se determina por la función de distribución acumulada:

$$F(x, y) = P(X \leq x, Y \leq y)$$

Luego, la probabilidad de pertenencia a determinado rectángulo viene dada por:

$$P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F(x_2, y_2) - F(x_2, y_1) - F(x_1, y_2) + F(x_1, y_1)$$

Esto puede comprenderse observando la siguiente figura, donde se sustraen ambos recuadros de los extremos y luego se le suma el externo para recuperar lo perdido por la sustracción de los recuadros de los extremos:



**Figura 3.1:** Probabilidad buscada marcada en gris.

Este mismo razonamiento puede extenderse a  $n$  variables aleatorias:

$$F(x_1, \dots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

### 3.2. Variables aleatorias discretas

**Definición:** Supongamos que  $X$  e  $Y$  son variables aleatorias discretas definidas en el mismo espacio muestral que toman valores  $x_1, x_2, \dots$  y  $y_1, y_2, \dots$  respectivamente. Su *función de frecuencia conjunta* se define como

$$p(x_i, y_j) = P(X = x_i, Y = y_j)$$

Para encontrar la función de frecuencia de  $Y$  se suma sobre todos los  $x$ , por lo cual  $p_y$  se conoce como la función de frecuencia marginal de  $Y$ . De forma análoga,

$$p_X(x) \triangleq \sum_i p(x, y_i)$$

El caso de múltiples variables es análogo. Si  $X_1, \dots, X_m$  son variables aleatorias discretas en el mismo espacio muestral, su función de frecuencia conjunta es:

$$p(x_1, \dots, x_m) \triangleq P(X_1 = x_1, \dots, X_m = x_m)$$

La *función de frecuencia marginal* de  $X_k$  es, por ejemplo:

$$p_{X_k}(x_k) \triangleq \sum_{x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_m} p(x_1, x_2, \dots, x_m)$$

La *función de frecuencia bidimensional* de  $X_1$  y  $X_2$  es, por ejemplo:

$$p_{X_1, X_2}(x_1, x_2) \triangleq \sum_{x_3, \dots, x_m} p(x_1, x_2, \dots, x_m)$$

### 3.3. Variables aleatorias continuas

**Definición:** Supongamos que  $X$  e  $Y$  son variables aleatorias con una cdf conjunta  $F(x, y)$ , se define su función de densidad conjunta como la función continua a tramos  $f(x, y)$  que es no negativa y tal que

$$\iint_{\mathbb{R}^2} f(x, y) dx dy = 1$$

Para cualquier conjunto bidimensional integrable  $A$  se tiene que:

$$P((X, Y) \in A) \triangleq \iint_A f(x, y) dx dy$$

En particular, se tiene que:

$$F(x, y) = P(X \leq x \wedge Y \leq y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) dv du$$

Se sigue del Teorema Fundamental del Cálculo que:

$$f(x, y) = \frac{\partial^2 F}{\partial x \partial y}(x, y) = \frac{\partial^2 F}{\partial y \partial x}(x, y)$$

De forma análoga al caso univariable, se sigue que:

$$P(x \leq X \leq x + dx \wedge y \leq Y \leq y + dy) = f(x, y) dx dy$$

Definición: Se define la cdf marginal de  $X$ ,  $F_x$ , como:

$$\begin{aligned} F_X(x) &= P(X \leq x) = \lim_{y \rightarrow \infty} F(x, y) \\ &= \int_{-\infty}^x \int_{-\infty}^{\infty} f(u, y) dy du \end{aligned}$$

Se sigue que la densidad marginal de  $X$  es:

$$f_X(x) = F'_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

Para el caso de  $\mathbb{R}^3$  y en general  $\mathbb{R}^n$  se puede hacer la generalización respectiva de forma análoga al caso discreto.

### 3.4. Variables aleatorias independientes

**Definición:** El conjunto de variables aleatorias  $X_1, X_2, \dots, X_n$  se dicen independientes si su cdf conjunta se factoriza en el producto de sus cdfs marginales:

$$F(x_1, x_2, \dots, x_n) = F_{X_1}(x_1) F_{X_2}(x_2) \cdots F_{X_n}(x_n)$$

En otras palabras, para dos variables  $X$  e  $Y$  independientes se tendrá que:

$$\begin{aligned} F(x, y) &= F_X(x) F_Y(y) = \left( \int_{-\infty}^x f_X(u) du \right) \left( \int_{-\infty}^y f_Y(v) dv \right) \\ \int_{-\infty}^x \int_{-\infty}^y f(x, y) dv du &= \int_{-\infty}^x \int_{-\infty}^y f_X(u) f_Y(v) dv du \\ &\longrightarrow f(x, y) = f_X(u) f_Y(v) \end{aligned}$$

De lo anterior se concluye que:

$$P(X \in A \wedge Y \in B) = P(X \in A) P(Y \in B)$$

Se puede demostrar asimismo que si  $g$  y  $h$  son funciones, entonces  $g(X)$  y  $h(Y)$  son independientes también.

### 3.5. Distribuciones condicionales

#### 3.5.1. Caso discreto

**Definición:** Si  $X$  e  $Y$  son variables aleatorias distribuidas conjuntamente, la probabilidad condicional de  $X = x_i$  dado  $Y = y_j$  es, si  $p_Y(y_j) > 0$

$$P(X = x_i | Y = y_j) \triangleq \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)} = \frac{p_{XY}(x_i, y_j)}{p_Y(y_j)} \triangleq p_{Y|X}(y|x)$$

Se define como cero si  $p_Y(y_j) = 0$ .

Se nota entonces que si  $X$  e  $Y$  son independientes, entonces  $p_{Y|X}(y|x) = p_Y(y)$ .

La definición de función de frecuencia condicional puede ser reexpresada como:

$$p_{XY}(x, y) = p_{X|Y}(x|y) p_Y(y)$$

Sumando a ambos lados obtenemos la ley de probabilidad total:

$$p_X(x) = \sum_y p_{X|Y}(x|y) p_Y(y)$$

#### Ejemplo

Supongamos que el contador de partículas es imperfecto e independientemente detecta cada partícula entrante con probabilidad  $p$ . Si la distribución del número entrante de partículas en una unidad de tiempo distribuye Poisson con parámetro  $\lambda$ , ¿cuál es la distribución del número de partículas contadas?

#### Solución:

Sea  $N$  el número real de partículas y  $X$  el número contado. Se sabe que la distribución condicional de  $X$  dado  $N = n$  es binomial con probabilidad  $p$ . Por la ley de probabilidad total:

$$\begin{aligned} P(X = k) &= \sum_{n=0}^{\infty} \underbrace{P(N = n)}_{\sim \text{poisson}} \underbrace{P(X = k | N = n)}_{\sim \text{binomial}} \\ &= \sum_{n=k}^{\infty} \frac{\lambda^n e^{-\lambda}}{n!} \binom{n}{k} p^k (1-p)^{n-k} \\ &= \sum_{n=k}^{\infty} \frac{\lambda^n e^{-\lambda}}{(n-k)! k!} p^k (1-p)^{n-k} \\ &= \frac{(\lambda p)^k e^{-\lambda}}{k!} \sum_{n=k}^{\infty} \frac{(1-p)^{n-k}}{(n-k)!} \\ &= \frac{(\lambda p)^k}{k!} e^{-\lambda p} \end{aligned}$$

Es decir,  $X \sim \text{Poisson}(\lambda p)$ .

### 3.5.2. El caso continuo

**Definición:** Análogamente, si  $X$  e  $Y$  son variables aleatorias continuas y conjuntas, entonces la densidad condicional de  $Y$  dado  $X$  se define como:

$$f_{Y|X}(y|x) \triangleq \frac{f_{XY}(x, y)}{f_X(x)}$$

si  $f_X(x) > 0$  y 0 de otra forma.

Se obtiene de forma análoga la ley de probabilidad total en este caso integrando:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{Y|X}(y|x) f_X(x) dx$$

## 3.6. Funciones de variables aleatorias distribuidas conjuntamente

Primero consideramos algunos casos particulares.

### 3.6.1. Sumas

Sean  $X$  e  $y$  variables aleatorias discretas tomando valores en los enteros y teniendo una función de frecuencia conjunta  $p(x, y)$  y sea  $Z = X + Y$ . Para encontrar la función frecuencia de  $Z$ , notamos que  $Z = z$  cuando  $X = x$  e  $Y = z - x$ , donde  $x$  es un entero. La probabilidad de que  $Z = z$  es por lo tanto la suma sobre todos los  $x$  de estas probabilidades conjuntas, es decir,

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p(x, z - x)$$

Si  $X$  e  $Y$  son independientes, entonces:

$$p_Z(z) = \sum_{x=-\infty}^{\infty} p_X(x) p_Y(z - x),$$

suma conocida como convolución de las sucesiones  $p_X$  y  $p_Y$ .

El caso continuo puede deducirse de forma similar, obteniendo así que:

$$f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x) dx$$

Si las variables son continuas e independientes, entonces:

$$f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x) dx,$$

la cual también se conoce como la convolución de las funciones  $f_X$  y  $f_Y$ .

### 3.6.2. Cocientes

Consideremos ahora el caso del cociente entre dos variables continuas aleatorias. La derivación es similar a la deducida para la suma de dichas variables. Supongamos que  $X$  e  $Y$  son continuas con una densidad conjunta  $f$  y sea  $Z = Y/X$ , entonces  $F_Z(z) = P(Z \leq z)$ . Si  $X > 0$ , entonces es equivalente a buscar  $y \leq xz$ , si  $X < 0$ , a buscar  $y \geq xz$ . Luego,

$$F_Z(z) = \int_{-\infty}^0 \int_{xz}^{\infty} f(x, y) dy dx + \int_0^{\infty} \int_{-\infty}^{xz} f(x, y) dy dx$$

Para remover la independencia de  $x$ , se hace  $y = xv$  para obtener:

$$\begin{aligned} F_Z(z) &= \int_{-\infty}^0 \int_z^{\infty} x f(x, xv) dv dx + \int_0^{\infty} \int_{-\infty}^z x f(x, xv) dv dx \\ &= \int_{-\infty}^z \int_{-\infty}^{\infty} |x| f(x, xv) dx dv \end{aligned}$$

Derivando con respecto a  $z$ :

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f(x, xz) dx$$

Si son independientes las variables:

$$f_Z(z) = \int_{-\infty}^{\infty} |x| f_X(x) f_Y(xz) dx$$

#### Ejemplo

Suponiendo que  $X$  e  $Y$  son variables independientes que distribuyen normal estándar. Calculemos la distribución de  $Z = Y/X$ .

#### Solución:

De acuerdo al resultado anterior se tiene que:

$$f_Z(z) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |x| e^{-x^2/2} e^{-x^2 z^2/2} dx = \frac{1}{\pi} \int_0^{\infty} x e^{-x^2(1+z^2)/2} dx$$

Haciendo  $u = x^2 \rightarrow du = 2x dx$ :

$$f_Z(z) = \frac{1}{2\pi} \int_0^{\infty} e^{-u(1+z^2)/2} du = \frac{1}{\pi(1+z^2)}$$

Esta función de densidad se conoce como densidad de Cauchy y está definida para todo  $z$ .

### 3.6.3. Caso general

Para el caso general, suponemos que  $X$  e  $Y$  son variables aleatorias continuas distribuidas conjuntamente y que  $X$  e  $Y$  son mapeados en  $U$  y  $V$  bajo la transformación:

$$\begin{aligned} u &= g_1(x, y) \\ v &= g_2(x, y) \end{aligned}$$

De acuerdo al Teorema de la Función Inversa, de existir la inversa, deberá cumplirse que:

$$\frac{\partial(g_1, g_2)}{\partial(x, y)} = \frac{\partial g_1}{\partial x} \frac{\partial g_2}{\partial y} - \frac{\partial g_2}{\partial x} \frac{\partial g_1}{\partial y} \neq 0$$

Esto lleva directamente al siguiente resultado, que puede ser demostrado mediante lo aprendido del teorema de sustitución de integrales dobles en el curso de cálculo en varias variables:

**Proposición:** Bajo los supuestos mencionados anteriormente, la distribución conjunta de  $U$  y  $V$  es:

$$f_{UV}(u, v) = f_{XY}(h_1(u, v), h_2(u, v)) \left[ \frac{\partial(g_1, g_2)}{\partial(x, y)} \right]^{-1}$$

para  $(u, v)$  tal que  $u = g_1(x, y)$  y  $v = g_2(x, y)$  para algunos  $(x, y)$  y cero en otro caso.

El resultado es extendible a  $\mathbb{R}^n$  de forma similar, obteniendo así que:

$$f_{Y_1 \dots Y_n}(y_1, \dots, y_n) = f_{X_1 \dots X_n}(x_1, \dots, x_n) \left[ \frac{\partial(g_1, \dots, g_n)}{\partial(x_1, \dots, x_n)} \right]^{-1}$$

## 3.7. Valores extremos y estadísticos de orden

Sean  $X_1, X_2, \dots, X_n$  variables independientes y aleatorias con una cdf común  $F$  y densidad  $f$  (iid., independientes e idénticamente distribuidas). Sea  $U$  el máximo de  $X_i$  y  $V$  el mínimo. Las cdfs de  $U$  y  $V$ , y por lo tanto sus densidades, serán calculadas.

- Notamos que  $U \leq u \iff X_i \leq u$  para todo  $i$ . Luego,

$$F_U(u) = P(U \leq u) = P(\forall i : X_i \leq u) \stackrel{\text{iid.}}{=} \prod_{i=1}^n P(X_i \leq u) = F^n(u)$$

Derivando con respecto a  $u$ :

$$f_U(u) = nF^{n-1}(u) \frac{dF}{du} = nF^{n-1}(u) f(u)$$

- De forma análoga,  $V \geq v \iff X_i \geq v$  para todo  $i$ . Luego, se puede deducir mediante lógica complementaria que:

$$1 - F_V(v) = [1 - F(v)]^n \implies F_V(v) = 1 - [1 - F(v)]^n$$

Finalmente, derivando con respecto a  $v$ :

$$f_V(v) = n[1 - F(v)]^{n-1} \frac{dF}{dv} = n[1 - F(v)]^{n-1} f(v)$$



Ahora asumamos los mismos supuestos, ordenamos los  $X_i$  y denotamos esta lista por:

$$X_{(1)} < X_{(2)} < \cdots < X_{(n)}$$

Estos resultados se conocen como estadísticos de orden. Debe notarse de inmediato que no necesariamente  $X_1 = X_{(1)}$ . Se nota aquí que  $X_{(1)}$  es el mínimo,  $X_{(n)}$  el máximo y si  $n = 2m + 1$ , entonces  $X_{(m+1)}$  es la mediana de  $X_i$ .

**Teorema:** La densidad de  $X_{(k)}$ , el  $k$ -ésimo estadístico de orden es:

$$f_k(x) = \frac{n!}{(k-1)!(n-k)!} f(x) F^{k-1}(x) [1-F(x)]^{n-k}$$

## 4. Valores esperados

### 4.1. El valor esperado de una variable aleatoria

El concepto de valor esperado de una variable aleatoria recuerda la noción de promedio ponderado. Los posibles valores de una variable aleatoria son ponderados por sus probabilidades, tal como se especifica en la siguiente definición:

**Definición:** Sea  $X$  una variable aleatoria discreta con función de frecuencia  $p(x)$ , el *valor esperado* o *promedio* de  $X$ , denotado por  $\mathbb{E}(X)$ , es:

$$\mathbb{E}(x) \triangleq \sum_i x_i p(x_i)$$

También suele denotarse por  $\mu$  ó  $\mu_X$  y puede entenderse como el centro de masa de la función frecuencia.

Bajo esta definición, calculemos el promedio de cada una de las distribuciones discretas enunciadas en el capítulo 2.

- **Bernoulli.** Usando la definición:

$$\mathbb{E}(X) = p \cdot 1 + (1 - p) \cdot 0 = p$$

- **Binomial.** Usando la misma definición:

$$\begin{aligned} \mathbb{E}(X) &= \sum_{k=0}^n k \binom{n}{k} p^k (1-p)^{n-k} = \sum_{k=1}^n k \frac{n!}{(n-k)! k!} p^k (1-p)^{n-k} \\ &= \sum_{k=1}^n \frac{n \cdot (n-1)!}{(n-k)! (k-1)!} p^k (1-p)^{n-k} \\ &= n \sum_{k=1}^n \binom{n-1}{k-1} p^k (1-p)^{n-k} \\ &= n \sum_{k=0}^{n-1} \binom{n-1}{k} p^{k+1} (1-p)^{n-1-k} \\ &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k} \end{aligned}$$

$\nearrow 1^{n-1}$

Finalmente,

$$\mathbb{E}(X) = np$$

- **Geométrica.** Se tiene que:

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=1}^{\infty} k(1-p)^{k-1} p = p \sum_{k=1}^{\infty} k(1-p)^{k-1} \\ &= -p \frac{d}{dp} \sum_{k=0}^{\infty} (1-p)^k \\ &= -p \frac{d}{dp} \frac{1}{1-(1-p)} = -p \frac{d}{dp} \frac{1}{p}\end{aligned}$$

Finalmente,

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

Es decir, si la probabilidad es 10% de ocurrencia de un error, entonces tras 10 inspecciones se espera que aparezca un artículo defectuoso.

- **Binomial negativa.** Se tiene que

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=r}^{\infty} k \binom{k-1}{r-1} p^r (1-p)^{k-r} = \sum_{k=r}^{\infty} k \frac{(k-1)!}{(r-1)!(k-r)!} p^r (1-p)^{k-r} \\ &= \frac{p^r}{(r-1)!} \sum_{k=r}^{\infty} (k-r+1)(k-r+2) \cdots (k-1) k (1-p)^{k-r} \\ &= \frac{p^r}{(r-1)!} (-1)^r \frac{d^r}{dp^r} \sum_{k=0}^{\infty} (1-p)^k = \frac{p^r}{(r-1)!} \frac{d^r}{dp^r} \frac{1}{p} \\ &= \frac{p^r}{(r-1)!} (-1)^{2r} \frac{r!}{p^{r+1}}\end{aligned}$$

Finalmente,

$$\boxed{\mathbb{E}(X) = \frac{r}{p}}$$

- **Hipergeométrica.** El cálculo es complejo y no se mencionará acá. Se puede demostrar que:

$$\boxed{\mathbb{E}(X) = \frac{mr}{n}}$$

- **Poisson.** En este caso se tiene que:

$$\begin{aligned}\mathbb{E}(X) &= \sum_{k=0}^{\infty} \frac{k\lambda^k}{k!} e^{-\lambda} = \lambda e^{-\lambda} \sum_{k=1}^{\infty} \frac{\lambda^{k-1}}{(k-1)!} \\ &= \lambda e^{-\lambda} \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = \lambda e^{-\lambda} e^{\lambda}\end{aligned}$$

Finalmente,

$$\boxed{\mathbb{E}(X) = \lambda}$$

Entonces el parámetro  $\lambda$  de la distribución Poisson puede interpretarse como la cuenta promedio de ocurrencias.

En el caso discreto, es razonable e intuitivo que se reemplaza la suma por la integración. Se tiene entonces la siguiente definición:

**Definición:** Si  $X$  es una variable aleatoria continua con densidad  $f(x)$ , entonces:

$$\mathbb{E}(X) \triangleq \int_{-\infty}^{\infty} x f(x) dx$$

considerando que  $\int |x|f(x) dx < \infty$ . Si no existe o diverge, la esperanza no está definida.

Bajo esta definición se puede calcular la esperanza para diversas distribuciones continuas:

1. **Uniforme.** Mediante esta definición, podemos realizar los mismos cálculos para las diversas densidades:

$$\mathbb{E}(X) = \int_a^b \frac{x}{(b-a)} dx = \frac{1}{b-a} \frac{b^2 - a^2}{2} = \frac{a+b}{2}$$

Lo cual es un resultado razonable considerando que es el centro de masa de la función densidad.

2. **Exponencial.** Se tiene que:

$$\begin{aligned} \mathbb{E}(X) &= \int_0^{\infty} \lambda x e^{-\lambda x} dx = -\lambda \int_0^{\infty} \frac{d}{d\lambda} e^{-\lambda x} dx \\ &= -\lambda \frac{d}{d\lambda} \frac{1}{\lambda} \end{aligned}$$

Es decir,

$$\mathbb{E}(X) = \frac{1}{\lambda}$$

3. **Gamma.** Por definición:

$$\mathbb{E}(X) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \underbrace{\int_0^{\infty} t^\alpha e^{-\lambda t} dt}_{\Gamma(\alpha+1)/\lambda^{\alpha+1}} = \frac{\alpha}{\lambda}$$

donde se usó la relación  $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ .

4. **Normal.** Reemplazando mediante la definición:

$$\begin{aligned} \mathbb{E}(X) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (x-\mu + \mu) e^{-(x-\mu)^2/2\sigma^2} dx \\ &= \frac{1}{\sigma\sqrt{2\pi}} \left( \int_{-\infty}^{\infty} (x-\mu) e^{-(x-\mu)^2/2\sigma^2} dx + \mu \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx \right) \\ &= \frac{1}{\sigma\sqrt{2\pi}} \mu \int_{-\infty}^{\infty} e^{-(x-\mu)^2/2\sigma^2} dx \end{aligned}$$

Es decir,

$$\boxed{\mathbb{E}(x) = \mu}$$

donde se canceló la integral pues esta simétrica en torno al punto  $(\mu, 0)$  (de forma análoga a una función impar). La segunda integral vale  $\sigma\sqrt{2\pi}$ , por lo que el resultado total es  $\mu$ . De aquí que este parámetro se conoce también como el promedio de la distribución normal.

5. **Beta.** En este caso, se tendrá que:

$$\mathbb{E}(X) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \underbrace{\int_0^1 u^a (1-u)^{b-1} du}_{\Gamma(a+1)\Gamma(b)/\Gamma(a+b+1)} = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(a+1)\Gamma(b)}{\Gamma(a+b+1)}$$

Se tiene entonces que:

$$\boxed{\mathbb{E}(X) = \frac{a}{a+b}}$$

donde se uso la relación  $\Gamma(\star + 1) = \star\Gamma(\star)$  para simplificar los términos.

**Teorema:** (*Desigualdad de Markov*) Si  $X$  es una variable aleatoria con  $P(X \geq 0) = 1$  y para la cual  $\mathbb{E}(X)$  existe, entonces  $P(X \geq t) \leq \mathbb{E}(X)/t$ .

*Demostración:*

Se probará para el caso continuo, considerando que es análogo para el caso con sumatoria. Separando integrales, se tiene que:

$$\mathbb{E}(X) = \int_{-\infty}^t xp(x) dx + \int_t^{\infty} xp(x) dx \geq \int_t^{\infty} xp(x) dx \geq t \int_t^{\infty} p(x) dx = tP(X \geq t)$$

De aquí se deduce que:

$$\boxed{P(X \geq t) \leq \frac{\mathbb{E}(X)}{t}} \quad \blacksquare$$

#### 4.1.1. Esperanza de funciones de variables aleatorias

**Teorema:** Suponiendo que  $Y = g(X)$  donde  $X$  es una variable aleatoria, entonces:

1. Si  $X$  es discreta con función de frecuencia  $p(x)$ , entonces:

$$\mathbb{E}(Y) = \sum_x g(x)p(x)$$

siempre que  $\sum |g(x)|p(x) < \infty$ .

2. Si  $X$  es continua con función de densidad  $f(x)$ , entonces:

$$\mathbb{E}(Y) = \int_{-\infty}^{\infty} g(x) f(x) dx$$

siempre que  $\int |g(x)|f(x)dx < \infty$ .

Se puede demostrar exactamente el mismo resultado para funciones de varias variables. De aquí se nota de inmediato que  $\mathbb{E}[g(x)] \neq g[\mathbb{E}(X)]$ .

**Corolario:** Si  $X$  e  $Y$  son variables independientes y  $g$  y  $h$  son funciones, entonces

$$\mathbb{E}[g(X)h(Y)] = \mathbb{E}[g(X)]\mathbb{E}[h(Y)]$$

siempre que los miembros del lado derecho de la ecuación existan.

#### 4.1.2. Esperanza de combinaciones lineales de variables aleatorias

**Teorema:** Si  $X_1, \dots, X_n$  son variables aleatorias conjuntas con esperanzas  $\mathbb{E}(X_i)$  e  $Y$  es la función  $Y = a + \sum_{i=1}^n b_i X_i$ , entonces:

$$\mathbb{E}(Y) = a + \sum_{i=1}^n b_i \mathbb{E}(X_i)$$

## 4.2. Varianza y desviación estándar

Se introducirá el concepto de desviación estándar de una variable aleatoria, la cual indica cuán dispersa está la distribución de probabilidad respecto a su esperanza. Primero se definirá la varianza de una variable aleatoria y luego la desviación estándar en términos de la varianza.

**Definición:** Si  $X$  es una variable aleatoria con valor esperado  $\mathbb{E}(x) = \mu$ , entonces la *varianza* de  $X$  es

$$\text{Var}(X) = \sigma^2(X) \triangleq \mathbb{E}[(X - \mu)^2]$$

Asimismo, se define la *desviación estándar* como:

$$\sigma(X) = \sqrt{\text{Var}(X)}$$

donde se usa esta medida para hacer coincidir las unidades de medición de la variable.

Se puede notar entonces que en el caso discreto la fórmula de la varianza es:

$$\text{Var}(X) = \sum_{x_i} (x_i - \mu)^2 p(x_i)$$

y en el caso continuo:

$$\text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

**Teorema:** Si  $\text{Var}(X)$  existe e  $Y = a + bX$ , entonces  $\text{Var}(Y) = b^2 \text{Var}(X)$  y por lo tanto  $\sigma_Y = |b| \sigma_X$ .

*Demostración:*

Se tiene que:

$$\begin{aligned} \text{Var}(Y) &= \mathbb{E}(Y - \mu_Y)^2 \\ &= \mathbb{E}(a + bX - a - b\mu_X)^2 \\ &= \mathbb{E}[b^2(X - \mu_X)^2] \\ &= b^2 \mathbb{E}(X - \mu_X)^2 \end{aligned}$$

Finalmente,

$$\text{Var}(Y) = b^2 \text{Var}(X) \quad \blacksquare$$

El siguiente teorema entrega una forma alternativa para calcular la varianza:

**Teorema:** La varianza de  $X$ , de existir, también puede ser calculada como:

$$\text{Var}(X) = \mathbb{E}(X^2) - \mu_X^2$$

*Demostración:*

Se tiene que:

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X - \mu)^2 = \mathbb{E}(X^2 - 2\mu X + \mu^2) \\ &= \mathbb{E}(X^2) - 2\mu \underbrace{\mathbb{E}(X)}_{\mu} + \mu^2 \end{aligned}$$

Finalmente,

$$\boxed{\text{Var}(X) = \mathbb{E}(X^2) - \mu^2}$$

Considerando la descripción dada de la varianza como indicación de qué tan dispersos están los valores de una variable aleatoria, existe una famosa desigualdad para entregar un aspecto cuantitativo de este indicador:

**Teorema:** (*Desigualdad de Chebyshev*) Sea  $X$  una variable aleatoria con media  $\mu$  y varianza  $\sigma^2$ , entonces para cualquier  $t > 0$ :

$$P(|X - \mu| > t) \leq \frac{\sigma^2}{t^2}$$

Este teorema plantea que si  $\sigma^2$  es muy pequeño, entonces hay una alta probabilidad de  $X$  no se desvíe mucho de  $\mu$ . Otra interpretación puede hacerse haciendo  $t = k\sigma$ :

$$P(|X - \mu| > k\sigma) \leq \frac{1}{k^2}$$

Es decir, la probabilidad de que  $X$  se aleje una distancia  $k\sigma$  (o  $k$  veces la desviación) de la media es menor a  $1/k^2$ , donde a medida que se aumenta  $k$  disminuye la probabilidad.

**Corolario:** Si  $\text{Var}(X) = 0$ , entonces  $P(X = \mu) = 1$ .

*Demostración:*

Por contradicción, supongamos que  $P(X = \mu) < 1$ . Luego, se tendrá que

$$P(|X - \mu| \geq \epsilon) > 0$$

lo cual contradice la desigualdad de Chebyshev, pues:

$$P(|X - \mu| \geq \epsilon) \leq 0 \quad \blacksquare$$

#### 4.2.1. Un modelo para el error de medición

Los valores de las constantes físicas no son conocidas con precisión pero deben ser determinadas mediante procedimientos experimentales. Operaciones tan sencillas como pesar un objeto, determinar un voltaje o medir un intervalo de tiempo son en realidad bastante complicadas cuando se toman en cuenta todos los detalles y posibles fuentes de error.

Una distinción realizada usualmente es aquella entre error aleatorio y error sistemático, donde el primero puede producir errores en las mediciones sin cambiar deliberadamente variables en el experimento, mientras que el segundo puede producirse en cada medición, como consecuencia de la mala calibración de instrumentos o errores asociados con la teoría detrás del método de medición.

Si el valor real que se desea medir se denota por  $x_0$ , entonces la medida  $X$  se modela como:

$$X = x_0 + \beta + \epsilon$$

donde  $\beta$  es la constante de error sistemático y  $\epsilon$  es la componente aleatoria del error, donde  $\epsilon$  es una variable aleatoria tal que  $\mathbb{E}(\epsilon) = 0$  y  $\text{Var}(\epsilon) = \sigma^2$ . Se tiene entonces que:

$$\mathbb{E}(X) = x_0 + \beta$$

y que

$$\text{Var}(X) = \sigma^2$$

donde  $\beta$  se llama comúnmente el sesgo del procedimiento de medición. Los dos factores afectando el tamaño del error son el sesgo y el tamaño de la varianza,  $\sigma^2$ . Una medición perfecta (e ideal) debiese entregar  $\beta = 0$  y  $\sigma^2 = 0$ .

**Definición:** Una medida del tamaño del error total de medición que se usa comúnmente es el **error cuadrático medio** (*MSE*, por sus términos en inglés mean squared error), el cual se define como:

$$\text{MSE}(X, x_0) = \mathbb{E}(X - x_0)^2$$



El error cuadrático medio, que es la desviación cuadrática esperada de  $X$  con respecto a  $x_0$ , puede ser descompuesta en el sesgo y la varianza.

**Teorema:**  $\text{MSE} = \beta^2 + \sigma^2$ .

*Demostración:*

$$\begin{aligned}\mathbb{E}(X - x_0)^2 &= \text{Var}(X - x_0) + [\mathbb{E}(X - x_0)]^2 \\ &= \text{Var}(X) + \beta^2 \\ &= \sigma^2 + \beta^2\end{aligned}$$

### 4.3. Covarianza y correlación

La varianza de una variable aleatoria es una medida de su variabilidad, y la covarianza de dos variables aleatorias es una medida de su variabilidad conjunta, o de su grado de asociación. Después de definir covarianza, se revisarán algunas propiedades y se discutirá una medida de asociación llamada correlación, definida en términos de la covarianza.

**Definición:** Si  $X$  e  $Y$  son variables distribuidas conjuntamente con esperanzas  $\mu_X$  y  $\mu_Y$  respectivamente, la *covarianza* de  $X$  e  $Y$  es:

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mu_X)(Y - \mu_Y)]$$

siempre que esta esperanza exista. De aquí se sigue que  $\text{Var}(X) = \text{Cov}(X, X)$ .

La covarianza es el valor promedio del producto de la desviación de  $X$  de su media y de la desviación de  $Y$  de su media. Si las variables aleatorias están positivamente asociadas –si  $X$  es más grande que su media, entonces  $Y$  también–, la covarianza será positiva. Si la asociación es negativa –esto es, cuando  $X$  es más grande que su media,  $Y$  tiende a ser más pequeño que su media– la covarianza es negativa.

Realizando una expansión de términos, se puede encontrar una expresión equivalente para la covarianza:

$$\text{Cov}(X, Y) = \mathbb{E}(XY) - \mu_X \mu_Y$$

De aquí se puede notar que si  $X$  e  $Y$  son independientes, entonces  $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) = \mu_X \mu_Y$  y por lo tanto la covarianza es cero (pero la expresión recíproca no es verdadera).

**Proposición:** Se verifican las siguientes identidades:

1.  $\text{Cov}(a + X, y) = \text{Cov}(X, Y)$ .
2.  $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$ .

$$3. \text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z).$$

$$4. \text{CoV}(aW + bX, cY + dZ) = ac \text{Cov}(W, Y) + ad \text{Cov}(W, Z) + bc \text{Cov}(X, Y) + bd \text{Cov}(X, Z).$$

**Teorema:** Si  $U = a + \sum_{i=1}^n b_i X_i$  y  $V = c + \sum_{j=1}^m d_j Y_j$ , entonces:

$$\text{Cov}(U, V) = \sum_{i=1}^n \sum_{j=1}^m b_i d_j \text{Cov}(X_i, Y_j)$$

**Corolario:**

$$1. \text{Var}\left(a + \sum_{i=1}^n b_i X_i\right) = \sum_{i=1}^n \sum_{j=1}^n b_i b_j \text{Cov}(X_i, X_j).$$

$$2. \text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i) \text{ si } X_i \text{ son independientes.}$$

El cociente de correlación se define en términos de la covarianza:

**Definición:** Si  $X$  e  $Y$  son variables aleatorias distribuidas conjuntamente y las varianzas y covarianzas de ambas  $X$  e  $Y$  existen y las varianzas de ambas  $X$  e  $Y$  son no nulas, entonces el *cociente de correlación* de  $X$  e  $Y$ , denotado por  $\rho$  es:

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

Puede notarse que el cociente será adimensional. Se sigue de forma sencilla que ante transformaciones lineales de  $X$  e  $Y$  el cociente de correlación no cambia. Como no depende de las unidades de medición y no discrimina las transformaciones lineales que se le puedan efectuar a las variables, el cociente de correlación es en muchos casos una medida útil de asociación, más que la covarianza.

**Teorema:**  $-1 \leq \rho \leq 1$ . Adicionalmente,  $\rho = \pm 1$  si y solo si  $P(Y = a + bX) = 1$  para algunas constantes  $a$  y  $b$ .

*Demostración:*

Como la varianza de una variable aleatoria es no negativa:

$$\begin{aligned} 0 &\leq \text{Var}\left(\frac{X}{\sigma_X} + \frac{Y}{\sigma_Y}\right) \\ &= \text{Var}\left(\frac{X}{\sigma_X}\right) + \text{Var}\left(\frac{Y}{\sigma_Y}\right) + 2 \text{Cov}\left(\frac{X}{\sigma_X}, \frac{Y}{\sigma_Y}\right) \\ &= \frac{\text{Var}(X)}{\sigma_X^2} + \frac{\text{Var}(Y)}{\sigma_Y^2} + \frac{2 \text{Cov}(X, Y)}{\sigma_X \sigma_Y} \\ &= 2(1 + \rho) \end{aligned}$$

de donde se observa que  $\rho \geq -1$ . Similarmente, se puede demostrar que:

$$0 \leq \text{Var} \left( \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} \right) = 2(1 - \rho) \longrightarrow \rho \leq 1$$

Supongamos que  $\rho = 1$ , entonces:

$$\text{Var} \left( \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} \right) = 0 \longrightarrow P \left( \frac{X}{\sigma_X} - \frac{Y}{\sigma_Y} = c \right) = 1$$

para alguna constante  $c$ . Por lo tanto,  $P(Y = a + bX) = 1$  para algunos  $a$  y  $b$ , y de forma análoga para  $\rho = -1$ . ■

**Ejemplo**

La distribución normal bivariada, de parámetros  $\mu_X$ ,  $\mu_Y$ ,  $\sigma_X$ ,  $\sigma_Y$  y  $\rho$  tiene como función de densidad:

$$f(x, y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} \exp \left( -\frac{1}{2(1-\rho^2)} \left[ \frac{(x-\mu_X)^2}{\sigma_X^2} + \frac{(y-\mu_Y)^2}{\sigma_Y^2} - \frac{2\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y} \right] \right)$$

Demostraremos que la covarianza de la distribución viene dada por  $\rho\sigma_X\sigma_Y$ .

**Solución:**

Por definición, la covarianza es:

$$\text{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_X)(y - \mu_Y) f(x, y) dx dy$$

Haciendo el cambio de variables  $u = (x - \mu_X)/\sigma_X$  y  $v = (y - \mu_Y)/\sigma_Y$  se tiene que  $dx dy = \sigma_X\sigma_Y du dv$  y por lo tanto:

$$\text{Cov}(X, Y) = \frac{\sigma_X\sigma_Y}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} uv \exp \left[ -\frac{u^2 + v^2 - 2\rho uv}{2(1-\rho^2)} \right] du dv$$

Aplicando la técnica de completación de cuadrados se tendrá que toda esta expresión es equivalente a:

$$\frac{\sigma_X\sigma_Y}{2\pi\sqrt{1-\rho^2}} \int_{-\infty}^{\infty} v \exp(-v^2/2) \left\{ \int_{-\infty}^{\infty} u \exp \left[ \frac{(u - \rho v)^2}{2(1-\rho^2)} \right] du \right\} dv$$

La integral interior es la media de una variable aleatoria que distribuye normal de parámetros  $\mu = \rho v$  y  $\sigma = (1 - \rho^2)$  salvo la constante  $1/\sqrt{2\pi(1-\rho^2)}$ , por lo que toda la expresión es igual

a:

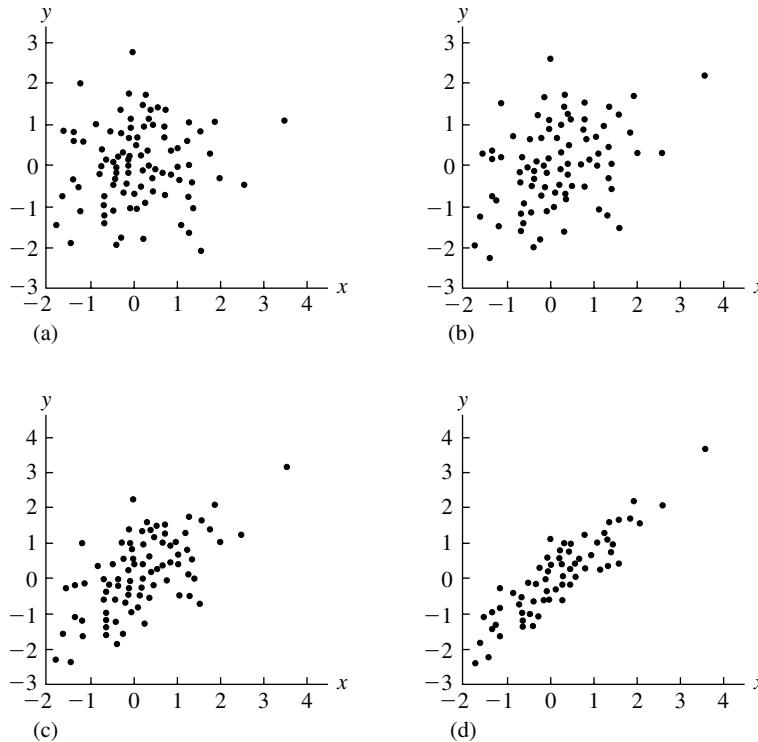
$$\begin{aligned}
 \text{Cov}(X, Y) &= \frac{\rho\sigma_X\sigma_Y}{\sqrt{2\pi}} \int_{-\infty}^{\infty} v^2 e^{-v^2/2} dv = 2 \frac{\rho\sigma_X\sigma_Y}{\sqrt{2\pi}} \int_0^{\infty} v^2 e^{-v^2/2} dv \\
 &= 2 \frac{\rho\sigma_X\sigma_Y}{\sqrt{2\pi}} \int_0^{\infty} \sqrt{2t} e^{-t} dt = \frac{2}{\sqrt{\pi}} \rho\sigma_X\sigma_Y \Gamma\left(\frac{3}{2}\right) \\
 &= \frac{2}{\sqrt{\pi}} \rho\sigma_X\sigma_Y \Gamma\left(\frac{3}{2}\right) = \frac{2}{\sqrt{\pi}} \rho\sigma_X\sigma_Y \frac{1}{2} \Gamma\left(\frac{1}{2}\right)
 \end{aligned}$$

Finalmente,

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

Esto implica que el coeficiente de correlación es  $\rho$  (coincidente con el parámetro de la distribución) para la distribución normal bivariada.  $\square$

El coeficiente de correlación mide la fuerza de una relación lineal entre  $X$  e  $Y$ . La correlación también afecta a la apariencia del gráfico de muestras en el plano coordenado, construido al generar pares de variables independientes  $(X_i, Y_i)$ . En la siguiente figura se puede observar el coeficiente de correlación para distintos tipos de variables aleatorias con ciertos grados de independencia:



**Figura 4.1:** (a)  $\rho = 0$ , (b)  $\rho = 0,3$ , (c)  $\rho = 0,6$ , (d)  $\rho = 0,9$ .

## 4.4. Esperanza condicional y predicción

### 4.4.1. Definiciones

Siempre asociada a una distribución condicional hay asociada una esperanza condicional.

**Definición:** Supongamos que  $Y$  y  $X$  son variables discretas aleatorias y que la función de frecuencia condicional de  $Y$  es  $p_{Y|X}(y|x)$ . La *esperanza condicional* de  $Y$  dado  $X = x$  es:

$$\mathbb{E}(Y|X = x) \triangleq \sum_y y p_{Y|X}(y|x)$$

Para el caso continuo, se tiene que:

$$\mathbb{E}(Y|X = x) \triangleq \int y f_{Y|X}(y|x) dy$$

Y en el caso más general, la esperanza condicional de la función  $h(Y)$  es:

$$\mathbb{E}[h(Y)|X = x] \triangleq \int_{-\infty}^{\infty} h(y) f_{Y|X}(y|x) dy$$

Asumiendo que la esperanza condicional de  $Y|X$  existe para cada  $x$  en el rango de  $X$ , es una función bien definida de  $X$  y por lo tanto una variable aleatoria, que escribimos como  $\mathbb{E}(Y|X)$  y por lo tanto, puede existir su esperanza,  $\mathbb{E}[\mathbb{E}(Y|X)]$ .

**Teorema:**  $\mathbb{E}(Y) = \mathbb{E}[\mathbb{E}(Y|X)]$ .

*Demostración:*

Se probará para el caso continuo. Tenemos que demostrar que:

$$\begin{aligned} \mathbb{E}(Y) &= \int_{-\infty}^{\infty} \mathbb{E}(Y|X = x) p_X(x) dx \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y p_{Y|X}(y|x) p_X(x) dy dx \\ &= \int_{-\infty}^{\infty} y \underbrace{\left( \int_{-\infty}^{\infty} p_{Y|X}(y|x) p_X(x) dx \right)}_{p_Y(y)} dy \\ &= \int_{-\infty}^{\infty} y p_Y(y) dy \end{aligned}$$

lo cual es cierto pues es la definición de esperanza. El tercer paso se dio utilizando la ley de probabilidades totales. ■

**Ejemplo** Consideremos las sumas del tipo

$$T = \sum_{i=1}^N X_i$$

donde  $N$  es una variable aleatoria con esperanza finita y  $X_i$  son variables aleatorias que son independientes de  $N$  y tienen esperanza común  $\mu_X$ . Calcular  $\mathbb{E}(T)$ .

**Solución:**

Se tiene que:

$$\mathbb{E}(T) = \mathbb{E}[\mathbb{E}(T|N)]$$

Se tiene que  $\mathbb{E}(T|N = n) = n\mu_X \rightarrow \mathbb{E}(T|N) = N\mu_X$ . Entonces,

$$\mathbb{E}(T) = \mathbb{E}(N\mu_X) = \mu_N\mu_X$$

**Teorema:**  $\text{Var}(Y) = \text{Var}[\mathbb{E}(Y|X)] + \mathbb{E}[\text{Var}(Y|X)]$ . □

**Ejemplo** Continuando el ejemplo anterior, asumamos también que  $X_i$  son variables independientes con la misma media  $\mu_X$  y varianza  $\sigma_X^2$ . Calcular  $\text{Var}(T)$ .

**Solución:**

Se tiene que:

$$\text{Var}(T) = \text{Var}[\mathbb{E}(T|N)] + \mathbb{E}[\text{Var}(T|N)]$$

Se tiene que  $\mathbb{E}(T|N) = N\mu_X$  y

$$\text{Var}(T|N) = \text{Var}\left(\sum_{i=1}^N X_i\right)$$

Como las variables son independientes, desaparecen los términos asociados a las covarianzas, de modo que:

$$\text{Var}(T|N) = \text{Var}\left(\sum_{i=1}^N X_i\right) = \sum_{i=1}^N \sigma_X^2 = N\sigma_X^2$$

Luego, reemplazando,

$$\begin{aligned}\text{Var}(T) &= \text{Var}(N\mu_X) + \mathbb{E}(N\sigma_X^2) \\ &= \mu_X^2 \text{Var}(N) + \sigma_X^2 \mu_N\end{aligned}$$

#### 4.4.2. Predicción

En esa sección se trata el problema de predecir el valor de una variable aleatoria a partir de otra, tal como se podría desear, por ejemplo, al medir una variable física utilizando un instrumento. Asumamos que las mediciones deben medirse a través de un instrumento que produce una respuesta  $X$ , relacionada con  $Y$  de cierta forma pero corrompida por ruido aleatorio.  $Y$  y  $X$  tienen una distribución conjunta, y deseamos predecir la variable  $Y$  a partir de la respuesta del instrumento  $X$ .

Partamos de la situación trivial: deseamos predecir  $Y$  por una constante  $c$ , donde el mejor valor de  $c$  debe ser obtenido a partir de la medición de la efectividad de la predicción. Una medición ampliamente utilizada es el error cuadrático medio:

$$\text{MSE} = \mathbb{E} [(Y - c)^2]$$

El problema consiste en encontrar  $c$  que minimice el MSE. Denotemos  $\mathbb{E}(Y)$  por  $\mu$  y observemos que:

$$\begin{aligned} \mathbb{E}(Y - c)^2 &= \text{Var}(Y - c) + \mathbb{E}^2(Y - c) \\ &= \text{Var}(Y) + (\mu - c)^2 \end{aligned}$$

El primer término es constante y el segundo se minimiza tomando  $\mu = c$ .

Consideremos ahora el problema de predecir  $Y$  por una función  $h(X)$  de modo de minimizar  $\text{MSE} = \mathbb{E}^2[Y - h(X)]$ . Se tiene entonces que:

$$\mathbb{E}^2[Y - h(X)] = \mathbb{E} \{ \mathbb{E} [(Y - h(X))^2 | X] \}$$

La esperanza de adentro se minimiza haciendo  $h(x) = \mathbb{E}(Y|X = x)$  de acuerdo al resultado obtenido en la parte anterior. Entonces,

$$h(X) = \mathbb{E}(Y|X)$$

#### 4.5. La función generadora de momentos

**Definición:** La *función generadora de momentos* de una variable aleatoria  $X$  es  $M(t) = \mathbb{E}(e^{tX})$  si esta está definida. En el caso discreto,

$$M(t) = \sum_x e^{tx} p(X),$$

y en el caso continuo

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

**Proposición:** Si la función generadora de momentos existe en un intervalo abierto conteniendo al cero, entonces

$$M^{(r)}(0) = \mathbb{E}(X^r)$$

donde  $(r)$  representa  $r$  veces la derivación con respecto a  $t$ .

*Demostración:*

Asumiendo alternancia de operadores, se tiene que:

$$\begin{aligned} \frac{d^r}{dt^r} \int_{-\infty}^{\infty} e^{tx} f(x) dx &= \int_{-\infty}^{\infty} f(x) \left( \frac{d^r}{dt^r} e^{tx} \right) dx \\ &= \int_{-\infty}^{\infty} f(x) x^r e^{tx} dx \end{aligned}$$

Haciendo  $t = 0$ :

$$M^{(r)}(0) = \int_{-\infty}^{\infty} x^r f(x) dx \triangleq \mathbb{E}(X^r) \quad \blacksquare$$

Teniendo calculada explícitamente la función generadora de momentos, puede calcularse de forma rápida cualquier momento central. A modo de ejemplo,

$$\text{Var}(X) = \mathbb{E}(X^2) - \mu_X^2 = M^{(2)}(0) - [M'(0)]^2$$

Esta es la utilidad práctica de la función generadora de momentos.

**Proposición:** Si  $X$  tiene la función generadora de momentos  $M_X(t)$  e  $Y = a + bX$ , entonces  $Y$  tiene la función generadora de momentos  $M_Y(t) = e^{at} M_X(bt)$ .

*Demostración:*

Se tiene que:

$$\begin{aligned} M_Y(t) &= \mathbb{E}(e^{tY}) \\ &= \mathbb{E}(e^{at+btX}) \\ &= e^{at} \mathbb{E}(e^{btX}) \\ &= e^{at} M_X(bt) \quad \blacksquare \end{aligned}$$

**Proposición:** Si  $X$  e  $Y$  son variables aleatorias independientes con funciones generadoras de momentos  $M_X$  y  $M_Y$  y  $Z = X + Y$ , entonces  $M_Z(t) = M_X(t) M_Y(t)$  en el intervalo común donde ambas funciones generadoras de momentos existen.  $\square$

Por inducción esta propiedad puede extenderse a la suma de una cantidad arbitraria de variables aleatorias independientes. Esta propiedad es muy útil en las funciones generadoras de momentos.



### 4.5.1. Cálculo de funciones generadoras de momentos

A continuación calcularemos la función generadora de momentos para algunas distribuciones de interés:

- **Distribución Poisson.** En la distribución Poisson, se tiene por definición que:

$$\begin{aligned} M(t) &= \sum_{k=0}^{\infty} e^{tk} \frac{\lambda^k}{k!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{k=0}^{\infty} \frac{(\lambda e^t)^k}{k!} \\ &= e^{\lambda(e^t-1)} \end{aligned}$$

Se tiene entonces que:

$$\begin{aligned} M(0) &= 1 \longrightarrow \text{¡lo esperable!} \\ M'(t) &= \lambda e^t e^{\lambda(e^t-1)} \longrightarrow M'(0) = \mathbb{E}(X) = \lambda \\ M''(t) &= \lambda e^t e^{\lambda(e^t-1)} + \lambda^2 e^{2t} e^{\lambda(e^t-1)} \longrightarrow M''(0) = \mathbb{E}(X^2) = \lambda^2 + \lambda \end{aligned}$$

Luego, se sigue que:

$$\boxed{\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \lambda}$$

- **Distribución normal.** Se tiene que:

$$M(t) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{tx} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

Notamos que:

$$\begin{aligned} tx - \frac{(x-\mu)^2}{2\sigma^2} &= -\frac{x^2 - 2\mu x + \mu^2 - 2\sigma^2 tx}{2\sigma^2} \\ &= -\frac{x^2 - 2(\mu + \sigma^2 t)x + (\mu^2 + 2\mu\sigma^2 t + \sigma^4 t^2) - (2\mu\sigma^2 t + \sigma^4 t^2)}{2\sigma^2} \\ &= -\frac{(x - \mu - \sigma^2 t)^2}{2\sigma^2} + \frac{t(2\mu + t\sigma^2)}{2} \end{aligned}$$

Es decir,

$$M(t) = e^{\frac{t(2\mu + t\sigma^2)}{2}} \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-\mu-\sigma^2 t)^2}{2\sigma^2}} dx = e^{\frac{t(2\mu + t\sigma^2)}{2}}$$

De aquí:

$$\begin{aligned} M(0) &= 1 \\ M'(0) &= \mathbb{E}(X) = \mu \\ M''(0) &= \mu + \sigma^2 \longrightarrow \text{Var}(X) = \sigma^2 \end{aligned}$$

Se puede haber obtenido el mismo resultado notando que  $Y = \mu + \sigma X$  y calculando el resultado para la distribución normal estándar, mucho más sencilla de calcular.

- **Distribución Gamma.** Por definición:

$$M(t) = \int_0^{\infty} e^{tx} \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} dx = \frac{\lambda^\alpha}{\Gamma(\alpha)} \underbrace{\int_0^{\infty} x^{\alpha-1} e^{-x(\lambda-t)} dx}_{\Gamma(\alpha)/(\lambda-t)^\alpha}$$

Luego,

$$M(t) = \left( \frac{\lambda}{\lambda-t} \right)^\alpha$$

Se puede demostrar así que:

$$M'(0) = \boxed{\mathbb{E}(X) = \frac{\alpha}{\lambda}}$$

$$M''(0) = \mathbb{E}(X^2) = \frac{\alpha(\alpha+1)}{\lambda^2} \rightarrow \boxed{\text{Var}(X) = \frac{\alpha}{\lambda^2}}$$

- **Suma de variables Poisson.** Si  $X \sim \text{Poisson}(\lambda)$  e  $Y \sim \text{Poisson}(\mu)$ , entonces:

$$M_{X+Y}(t) = M_X(t) M_Y(t) = e^{\lambda(e^t-1)} e^{\mu(e^t-1)} = e^{(\lambda+\mu)(e^t-1)}$$

Es decir,  $\boxed{X+Y \sim \text{Poisson}(\lambda+\mu)}$ .

- **Suma de variables normales.** Si  $X \sim N(\mu, \sigma^2)$  y  $Y \sim N(\nu, \tau^2)$ , entonces:

$$M_{X+Y}(t) = e^{\mu t + t^2 \sigma^2 / 2} e^{\nu t + t^2 \tau^2 / 2} = e^{(\mu+\nu)t + t^2(\sigma^2 + \tau^2) / 2}$$

Es decir,  $\boxed{X+Y \sim N(\mu+\nu, \sigma^2 + \tau^2)}$

- **Suma de variables Gamma.** Si  $X \sim \Gamma(\lambda, \alpha_1)$  e  $Y \sim \Gamma(\lambda, \alpha_2)$ , entonces:

$$M_{X+Y}(t) = \left( \frac{\lambda}{\lambda-t} \right)^{\alpha_1} \left( \frac{\lambda}{\lambda-t} \right)^{\alpha_2} = \left( \frac{\lambda}{\lambda-t} \right)^{\alpha_1 + \alpha_2}$$

donde  $t < \lambda$ . Luego,  $\boxed{X+Y \sim \Gamma(\lambda, \alpha_1 + \alpha_2)}$ . Si  $\alpha_1 = \alpha_2 = 1$  (en cuyo caso las distribuciones gamma también son exponenciales), entonces se deduce que la suma de  $n$  exponenciales de parámetro  $\lambda$  resulta en una distribución Gamma de parámetros  $\lambda$  y  $n$ .

## 5. Teoremas de los límites

En este capítulo se estudia el comportamiento en el límite de la suma de variables aleatorias independientes cuando el número de sumandos se vuelve grande. Los resultados presentados aquí son tanto interesante como útiles en estadística, debido a que la mayoría de cantidades estadísticas comúnmente calculadas, tales como promedios, pueden ser representadas como sumas.

### 5.1. La ley de números grandes

Es habitual pensar que si una moneda se lanza muchas veces y se calcula la proporción entre caras y sellos, la proporción será cercana a  $1/2$ . John Kerrich, un matemático sudafricano probó esta creencia empíricamente mientras fue un prisionero de la Segunda Guerra mundial. Lanzó la moneda 10.000 veces y observó 5.067 caras.

La ley de números grandes es una formulación matemática de esta creencia. Los lanzamientos sucesivos de monedas son modelados como intentos aleatorios independientes. La variable aleatoria  $X_i$  toma el valor 0 o 1 de acuerdo a los resultados del intento  $i$ -ésimo en una cara o un sello, y la proporción de caras y sellos en  $n$  intentos es:

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$$

La ley de números grandes plantea que  $\bar{X}_n$  tiende a  $1/2$  tal como se plantea en el siguiente teorema:

**Teorema:** (*Ley de números grandes*) Sean  $X_1, X_2, \dots, X_i$  una secuencia de variable aleatorias independientes con  $\mathbb{E}(X_i) = \mu$  y  $\text{Var}(X_i) = \sigma^2$ . Dado  $\bar{X}_n$  definido anteriormente, se tiene que para todo  $\varepsilon > 0$ :

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0$$

*Demostración:*

Primero encontramos  $\mathbb{E}(\bar{X}_n)$  y  $\text{Var}(\bar{X}_n)$ :

$$\mathbb{E}(\bar{X}_n) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \mu$$

Como  $X_i$  son independientes:

$$\text{Var}(\bar{X}_n) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) = \frac{\sigma^2}{n}$$

De acuerdo a la desigualdad de Chebyhshev:

$$0 \leq P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

Haciendo  $n \rightarrow \infty$  se cumple el teorema del sándwich y por lo tanto:

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0 \quad \blacksquare$$

Si una secuencia de variables aleatorias  $Z_n$  es tal que

$$\lim_{n \rightarrow \infty} P(|Z_n - \alpha| > \varepsilon) = 0$$

para todo  $\varepsilon > 0$  y donde  $\alpha$  es un escalar, entonces se dice que  $Z_n$  converge en probabilidad a  $\alpha$ .

**Ejemplo**

**Integración Monte Carlo.** Supongamos que se desea calcular

$$I(f) = \int_0^1 f(x) dx$$

donde no se puede determinar la primitiva o utilizar una tabla de integrales. Introduzcamos el método de integración Monte Carlo.

**Solución:**

El método más común es aproximar la integral mediante una suma y otro método es el método de Monte Carlo. Se generan variables aleatorias uniformes en  $[0, 1]$   $X_1, X_2, \dots, X_n$  para calcular

$$\hat{I}(f) = \frac{1}{n} \sum_{i=1}^n f(X_i)$$

Por la ley de números grandes, esto debe ser cercano a  $\mathbb{E}[f(X)]$  que en este caso no es más que

$$\mathbb{E}[f(X)] = \int_0^1 f(x) dx = I(f)$$

**Ejemplo**

**Mediciones repetitivas.** Supongamos que repetidas mediciones no sesgadas  $X_1, X_2, \dots, X_n$  se hacen sobre una cantidad. Si  $n$  es grande, entonces de acuerdo a la ley de números grandes se tendrá que  $\bar{X}$  será cercano a  $\mu$ , pero ¿qué tan cercano depende del error de medición  $\sigma^2$ , tal como se pudo observar en la medición del teorema?

**Solución:**

Afortunadamente,  $\sigma^2$  puede ser estimado y por lo tanto

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

puede ser estimado de los datos para estimar la precisión de  $\bar{X}$ . Primero, notamos que

$$\frac{1}{n} \sum_{i=1}^n X_i^2 \longrightarrow \mathbb{E}(X^2)$$

de acuerdo a la ley de números grandes. Se puede demostrar asimismo que si  $Z_n$  converge en probabilidad a  $\alpha$  y  $g$  es una función continua, entonces

$$g(Z_n) \longrightarrow g(\alpha)$$

lo cual implica que:

$$\bar{X}^2 \longrightarrow \mathbb{E}^2(X)$$

Se puede demostrar entonces que:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^2 - \bar{X}^2 = \mathbb{E}(X^2) - \mathbb{E}^2(X) = \text{Var}(X)$$

Se puede demostrar a modo general que:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n X_i^r = \mathbb{E}(X^r)$$

## 5.2. Convergencia en la distribución y el teorema del límite central

En muchas ocasiones deseamos calcular  $P(a < X < b)$  pero no sabemos la cdf  $X$  precisamente y se puede hacer en ocasiones aproximando  $F_X$  mediante un argumento asintótico. El teorema del límite más famoso en teoría de probabilidades es el teorema del límite central. Antes de examinarlo, se desarrollará terminología introductoria, teoría y ejemplos.

**Definición:** Sean  $X_1, X_2, \dots, X_n$  una secuencia de variables aleatorias con funciones de distribución acumulada  $F_1, \dots, F_n$  y sea  $X$  una variable aleatoria con función de distribución  $F$ . Decimos que  $X_n$  converge en distribución a  $X$  si:

$$\lim_{n \rightarrow \infty} F_n(x) = F(x)$$

en cada punto en que  $F$  es continua.

**Teorema. Teorema de continuidad.** Sea  $F_n$  una sucesión de funciones de distribución acumuladas con la correspondiente función generadora de momentos  $M_n$ , de modo que:

$$M_n(t) = \mathbb{E}(e^{tX_n})$$

Sea  $F$  la función de distribución acumulada con la función generadora de momentos  $M$ . Si  $M_n(t) \longrightarrow M(t)$  para todo  $t$  en un intervalo abierto conteniendo al cero, entonces  $F_n(x) \longrightarrow F(x)$ .

Nos enfocaremos ahora en el teorema del límite central, el cual se preocupa de la propiedad asintótica de sumas de variables aleatorias. Si  $X_1, X_2, \dots, X_n$  es una sucesión de variables aleatorias independientes con media  $\mu$  y varianza  $\sigma^2$  y si

$$S_n = \sum_{i=1}^n X_i$$

Sabemos por la ley de números grandes que  $S_n/n$  converge en probabilidad a  $\mu$ . Esto se sigue por el hecho de que:

$$\text{Var} \left( \frac{S_n}{n} \right) = \frac{1}{n^2} \text{Var} (S_n) = \frac{\sigma^2}{n} \rightarrow 0$$

El teorema del límite central se preocupa no del hecho de que  $S_n/n$  converja a este cociente si no cómo  $S_n/n$  fluctúa cerca de  $\mu$ . Para analizar estas situaciones estandarizaremos:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

donde  $Z_n$  tiene esperanza 0 y varianza 1, como es fácil de notar:

$$\mathbb{E} (Z_n) = \frac{1}{\sigma\sqrt{n}} [\mathbb{E} (S_n) - n\mu] = \frac{n\mu - n\mu}{\sigma\sqrt{n}} = 0$$

$$\text{Var} (Z_n) = \frac{\text{Var} (S_n)}{n\sigma^2} = 1$$

El teorema del límite central plantea que la distribución de  $Z_n$  converge a la distribución normal estándar.

**Teorema.** (*Teorema del límite central*) Sean  $X_1, X_2, \dots, X_n$  una secuencia de variables aleatorias independientes teniendo cada una de ellas media 0 y varianza  $\sigma^2$  y la función de distribución común  $F$  y la función generadora de momentos  $M$  definida en una vecindad de cero. Sea

$$S_n = \sum_{i=1}^n X_i$$

Entonces,

$$\lim_{n \rightarrow \infty} P \left( \frac{S_n}{\sigma\sqrt{n}} \leq x \right) = \Phi(x) \quad ; \quad -\infty < x < \infty$$

Se suele usar la notación  $\sim$  para indicar la tendencia asintótica. En otras palabras, el teorema enuncia que:

$$\frac{S_n}{\sigma\sqrt{n}} \sim N(0, 1)$$

*Demostración:*

Sea  $Z_n = S_n/(\sigma\sqrt{n})$ . Demostraremos que la función generadora de momentos de  $Z_n$  tiende a la función generadora de momentos de la distribución normal. Con ello, por unicidad de la transformada inversa de Laplace, habremos demostrado que ambas distribuciones en el límite son iguales.

Se tiene por propiedades de la función generadora de momentos que:

$$M_{S_n}(t) = M^n(t) \longrightarrow M_{Z_n}(t) = M^n\left(\frac{t}{\sigma\sqrt{n}}\right)$$

Luego,  $M(s)$  puede expandirse en Taylor en torno al origen con un polinomio de orden 2. Este es el hecho central de la demostración:

$$M(s) \approx M(0) + sM'(0) + \frac{s^2}{2}M''(0)$$

Como  $M(0) = \int f(x)dx = 1$ ,  $M'(0) = 0$  y  $M''(0) = \sigma^2$ , entonces cuando  $n \rightarrow \infty$   $t/(\sigma\sqrt{n}) \rightarrow 0$  y por lo tanto:

$$\begin{aligned} M\left(\frac{t}{\sigma\sqrt{n}}\right) &= 1 + \frac{1}{2}\sigma^2\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \varepsilon_n \\ &= 1 + \frac{t^2}{2n} + \varepsilon_n \end{aligned}$$

donde  $\frac{\varepsilon_n}{t^2/n\sigma^2} \rightarrow 0$  cuando  $n \rightarrow \infty$ . Reemplazando, se tendrá que:

$$M_{Z_n}(t) \approx \left(1 + \frac{t^2}{2n}\right)^n$$

Luego, cuando  $n \rightarrow \infty$  se tiene que este límite converge a la exponencial:

$$\lim_{n \rightarrow \infty} M_{Z_n}(t) = e^{t^2/2}$$

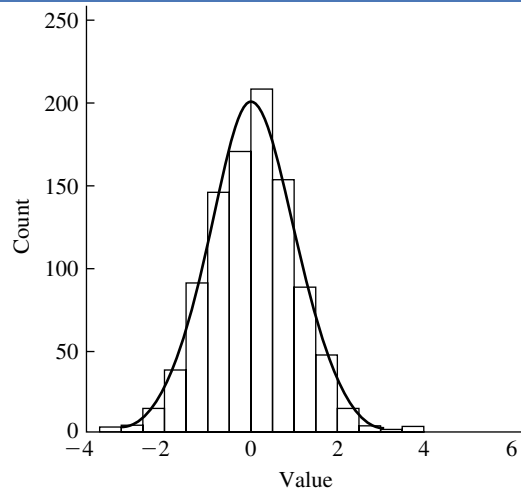
donde  $e^{t^2/2}$  es la función generadora de momentos de la distribución normal estándar, justo como que sería demostrar y concluyendo así la demostración del teorema. Notar que este teorema exige que exista como mínimo la segunda derivada de la función generadora de momentos.

**Ejemplo**

Como la distribución uniforme en  $[0, 1]$  tiene media  $1/2$  y varianza  $1/12$ , la suma de 12 variables aleatorias uniformes, menos 6, tiene media 0 y varianza 1. La distribución de esta suma es muy cercana a la normal, tan cercana que fue uno de los algoritmos iniciales para generar números aleatorios normales antes de que existieran algoritmos más sofisticados.

**Solución:**

La siguiente figura muestra un histograma con 1000 de dichas sumas de 12 números aleatorios entre 0 y 1. El ajuste a la distribución normal es sorprendentemente correcto, especialmente considerando que 12 es un número pequeño para considerar que  $n$  tiende a infinito:



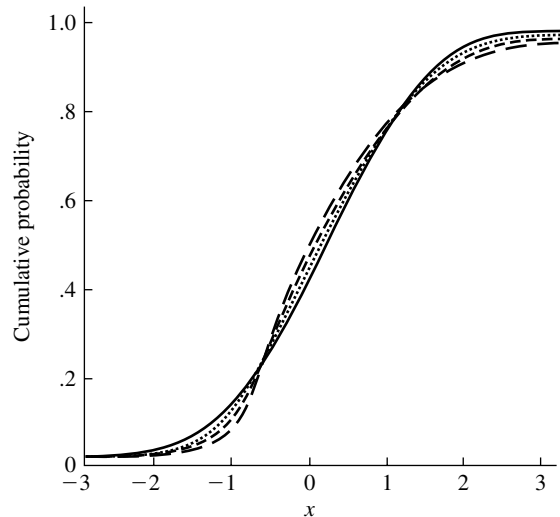
**Figura 5.1:** Histograma de 1000 valores, en que cada uno es la suma de 12 variables aleatorias entre 0 y 1 menos 6, comparado con la densidad normal estándar.

**Ejemplo**

La suma de  $n$  variables exponenciales aleatorias con parámetro  $\lambda = 1$  sigue una distribución gamma con  $\lambda = 1$  y  $\alpha = n$ . La siguiente figura muestra la cdf de la función gamma estandarizada y la cdf de la función normal estandarizada a medida que  $n$  aumenta. Se comprueba cómo a medida que se aumenta  $n$  (y por lo tanto  $\alpha$ ) las distribuciones comienzan a parecerse.

**Solución:**





**Figura 5.2:** Línea sólida: distribución normal estándar. Líneas punteadas largas:  $\alpha = 5$ . Líneas punteadas cortas:  $\alpha = 10$ . Puntos:  $\alpha = 30$ .

**Ejemplo**

**Error de medición.** Supongamos que  $X_1, \dots, X_n$  son mediciones repetitivas e independientes de una cantidad  $\mu$  y que  $\mathbb{E}(X_i) = \mu$  y  $\text{Var}(X_i) = \sigma^2$ . Se usa el promedio de las mediciones,  $\bar{X}$ , se usa como estimador de  $\mu$ . La ley de números grandes nos dice que  $\bar{X}$  converge en probabilidad a  $\mu$ , por lo que  $\bar{X}$  debe ser cercano a  $\mu$  si  $n$  es grande.

La desigualdad de Chebyshev permite acotar la probabilidad a un determinado tamaño, pero el teorema central del límite nos entrega una aproximación mucho más fina del error real.

**Solución:**

Supongamos que deseamos encontrar  $P(|\bar{X} - \mu| < c)$  para alguna constante  $c$  arbitraria. Se tiene que:

$$\begin{aligned}
 P(|\bar{X} - \mu| < c) &= P(-c < \bar{X} - \mu < c) \\
 &= P\left(-\frac{c}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{c}{\sigma/\sqrt{n}}\right)
 \end{aligned}$$

donde como  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  entonces  $n(\bar{X} - \mu) / (\sigma\sqrt{n}) \sim N(0, 1)$ . De esta forma,

$$P(|\bar{X} - \mu| < c) = P\left(-\frac{c}{\sigma/\sqrt{n}} < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < \frac{c}{\sigma/\sqrt{n}}\right) \approx \Phi\left(\frac{c\sqrt{n}}{\sigma}\right) - \Phi\left(-\frac{c\sqrt{n}}{\sigma}\right)$$

## 6. Distribuciones derivadas de la distribución normal

En este capítulo se estudian tres distribuciones derivadas de la distribución normal. Estas distribuciones ocurren en diversos problemas estadísticos.

### 6.1. Distribución $\chi^2$

**Definición:** Si  $Z$  es una distribución normal estándar, entonces la distribución  $U = Z^2$  se denomina *distribución chi-cuadrado* (o  $\chi_1^2$ ) con un grado de libertad. Tal como vimos en la sección 2.3, es el caso especial de la distribución gamma con parámetros  $\alpha = 1/2$  y  $\lambda = 1/2$ . Entonces,

$$X \sim N(0, 1) \longrightarrow X^2 \sim \chi_1^2$$

Se sigue entonces que como:

$$X \sim N(\mu, \sigma^2) \longrightarrow \frac{X - \mu}{\sigma} \sim N(0, 1) \longrightarrow \boxed{\frac{(X - \mu)^2}{\sigma^2} \sim \chi_1^2}$$

**Definición:** Si  $U_1, U_2, \dots, U_n$  son variables aleatorias independientes tales que  $U_i \sim \chi_1^2$  para todo  $i = 1, \dots, n$ . Entonces,  $V = U_1 + \dots + U_n$  se llama la distribución *chi-cuadrado* con  $n$  grados de libertad y se nota como  $\chi_n^2$ .

Tal como dedujimos en el capítulo 4, la suma de variables Gamma con mismo parámetro de escala distribuye Gamma con el mismo parámetro de escala y la suma de sus parámetros de forma. Entonces,  $\chi_n^2 = \Gamma(\lambda = 1/2, \alpha = n/2)$  y por lo tanto la función de densidad es:

$$\boxed{X \sim \chi_n^2 \iff f(v) = \frac{1}{2^{n/2} \Gamma(n/2)} v^{(n/2)-1} e^{-v/2} \quad ; \quad v \geq 0}$$

Su función generadora de momentos es:

$$\boxed{M(t) = (1 - 2t)^{-n/2}}$$

Se sigue asimismo que si  $U \sim \chi_n^2$  y  $V \sim \chi_m^2$  entonces  $\boxed{U + V \sim \chi_{n+m}^2}$ .

### 6.2. Distribución $t$ -student

**Definición:** Si  $Z \sim N(0, 1)$  y  $U \sim \chi_n^2$  y  $Z$  y  $U$  son independientes, entonces la distribución de  $Z/\sqrt{U/n}$  se denomina la *distribución  $t$ -student* con  $n$  grados de libertad. En otras palabras,

$$\frac{Z}{\sqrt{U/n}} \sim t(n)$$

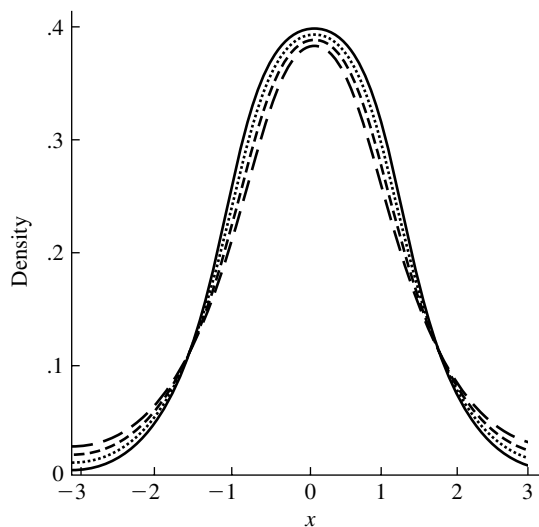
Se puede demostrar con los procedimientos ya demostrados en el capítulo 3 que la función de densidad de esta distribución es:

$$X \sim t(n) \longleftrightarrow f(t) = \frac{\Gamma[(n+1)/2]}{\sqrt{n\pi} \Gamma(n/2)} \left(1 + \frac{t^2}{n}\right)^{-(n+1)/2} ; \quad -\infty < t < \infty$$

Se observa con facilidad que la distribución es simétrica en torno a cero. Asimismo, se puede notar que cuando  $n \rightarrow \infty$  la distribución tiende a ser la distribución normal estándar. En otras palabras,

$$\frac{Z}{\sqrt{U/n}} \sim N(0, 1)$$

En efecto, para 20 o 30 grados de libertad las distribuciones ya son muy similares. En la siguiente figura se muestran algunas distribuciones  $t$ , donde puede notarse como las colas de la distribución a los extremos tienden a *alivianarse* (hacerse más planas e inclinadas a cero) a medida que los grados de libertad aumentan.



**Figura 6.1:** Línea sólida: distribución normal estándar. Líneas punteadas largas:  $n = 5$ . Líneas punteadas cortas:  $n = 10$ . Puntos:  $n = 30$ .

### 6.3. Distribución $F$

**Definición:** Sean  $U$  y  $V$  variables  $\chi^2$  de  $m$  y  $n$  grados de libertad respectivamente. La distribución de

$$W = \frac{U/m}{V/n}$$

se denomina *distribución  $F$*  con  $m$  y  $n$  grados de libertad y se denota  $F_{m,n}$ .

La función de densidad viene dada por:

$$W \sim F_{m,n} \longleftrightarrow f(w) = \frac{\Gamma[(m+n)/2]}{\Gamma(m/2)\Gamma(n/2)} \left(\frac{m}{n}\right)^{m/2} w^{m/2-1} \left(1 + \frac{m}{n}w\right)^{-(m+n)/2} ; \quad w \geq 0$$

Se puede observar que el cuadrado de una variable  $t_n$  distribuye  $F_{1,n}$ . La esperanza existe para  $n > 2$  y es igual a  $\mathbb{E}(W) = n/(n-2)$ .

# Anexo 1: Distribuciones

Distribución	Densidad de Probabilidad	$\Theta_X$	Parámetros	Esperanza y Varianza
Binomial	$\binom{n}{x} p^x (1-p)^{n-x}$	$x = 0, \dots, n$	$n, p$	$\mu_X = np$ $\sigma_X^2 = np(1-p)$
Geométrica	$p(1-p)^{x-1}$	$x = 1, 2, \dots$	$p$	$\mu_X = 1/p$ $\sigma_X^2 = (1-p)/p^2$
Binomial-Negativa	$\binom{x-1}{r-1} p^r (1-p)^{x-r}$	$x = r, r+1, \dots$	$r, p$	$\mu_X = r/p$ $\sigma_X^2 = r(1-p)/p^2$
Poisson	$\frac{(\nu t)^x e^{-\nu t}}{x!}$	$x = 0, 1, \dots$	$\nu$	$\mu_X = \nu t$ $\sigma_X^2 = \nu t$
Exponencial	$\nu e^{-\nu x}$	$x \geq 0$	$\nu$	$\mu_X = 1/\nu$ $\sigma_X^2 = 1/\nu^2$
Gamma	$\frac{\nu^k}{\Gamma(k)} x^{k-1} e^{-\nu x}$	$x \geq 0$	$k, \nu$	$\mu_X = k/\nu$ $\sigma_X^2 = k/\nu^2$
Gamma Traslada	$\frac{\nu^k}{\Gamma(k)} (x-\gamma)^{k-1} e^{-\nu(x-\gamma)}$	$x \geq \gamma$	$k, \nu, \gamma$	$\mu_X = k/\nu + \gamma$ $\sigma_X^2 = k/\nu^2$
Normal	$\frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$	$-\infty < x < \infty$	$\mu, \sigma$	$\mu_X = \mu$ $\sigma_X^2 = \sigma^2$
Log-Normal	$\frac{1}{\sqrt{2\pi}(\zeta x)} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\zeta}\right)^2\right]$	$x \geq 0$	$\lambda, \zeta$	$\mu_X = \exp\left(\lambda + \frac{1}{2}\zeta^2\right)$ $\sigma_X^2 = \mu_X^2 (e^{\zeta^2} - 1)$
Uniforme	$\frac{1}{(b-a)}$	$a \leq x \leq b$	$a, b$	$\mu_X = (a+b)/2$ $\sigma_X^2 = (b-a)^2/12$
Beta	$\frac{1}{B(q,r)} \frac{(x-a)^{q-1} (b-x)^{r-1}}{(b-a)^{q+r-1}}$	$a \leq x \leq b$	$q, r$	$\mu_X = a + \frac{q}{q+r}(b-a)$ $\sigma_X^2 = \frac{qr(b-a)^2}{(q+r)^2(q+r+1)}$
Hipergeométrica	$\frac{\binom{m}{x} \binom{N-m}{n-x}}{\binom{N}{n}}$	$\max\{0, n+m-N\} \leq x \leq \min\{n, m\}$	$N, m, n$	$\mu_X = n \frac{m}{N}$ $\sigma_X^2 = \left(\frac{N-n}{N-1}\right) n \frac{m}{N} \left(1 - \frac{m}{N}\right)$

- Propiedades función  $\Gamma(\cdot)$ :

$$(1) \quad \Gamma(k) = \int_0^\infty u^{k-1} e^{-u} du; \quad (2) \quad \Gamma(a+1) = a\Gamma(a);$$

$$(3) \quad \Gamma(n+1) = n!, \quad \text{si } n \in \mathbb{N}_0; \quad (4) \quad \Gamma(1/2) = \sqrt{\pi}$$

- Propiedades función  $B(\cdot, \cdot)$ :

$$(1) \quad B(q, r) = \int_0^1 x^{q-1} (1-x)^{r-1} dx; \quad (2) \quad B(q, r) = \frac{\Gamma(q)\Gamma(r)}{\Gamma(q+r)}$$

- Propiedad distribución Gamma:

$$\text{Si } T \sim \text{Gamma}(k, \nu) \Rightarrow F_T(t) = 1 - \sum_{x=0}^{k-1} \frac{(\nu t)^x e^{-\nu t}}{x!}, \quad \text{si } k \in \mathbb{N}$$

## Anexo 2: Tablas de percentiles y valores

Distribución Normal Estándar $k_p$											Distribución t-student $t_p(\nu)$				
$k_p$	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09	$\nu$	$t_{0.90}$	$t_{0.95}$	$t_{0.975}$	$t_{0.99}$
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359	1	3.078	6.314	12.706	31.821
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753	2	1.886	2.920	4.303	6.965
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141	3	1.638	2.353	3.182	4.541
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517	4	1.533	2.132	2.776	3.747
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879	5	1.476	2.015	2.571	3.365
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224	6	1.440	1.943	2.447	3.143
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549	7	1.415	1.895	2.365	2.998
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852	8	1.397	1.860	2.306	2.896
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133	9	1.383	1.833	2.262	2.821
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389	10	1.372	1.812	2.228	2.764
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621	11	1.363	1.796	2.201	2.718
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830	12	1.356	1.782	2.179	2.681
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015	13	1.350	1.771	2.160	2.650
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177	14	1.345	1.761	2.145	2.624
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319	15	1.341	1.753	2.131	2.602
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441	16	1.337	1.746	2.120	2.583
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545	17	1.333	1.740	2.110	2.567
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633	18	1.330	1.734	2.101	2.552
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706	19	1.328	1.729	2.093	2.539
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767	20	1.325	1.725	2.086	2.528
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817	21	1.323	1.721	2.080	2.518
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857	22	1.321	1.717	2.074	2.508
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890	23	1.319	1.714	2.069	2.500
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916	24	1.318	1.711	2.064	2.492
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936	25	1.316	1.708	2.060	2.485
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952	26	1.315	1.706	2.056	2.479
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964	27	1.314	1.703	2.052	2.473
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974	28	1.313	1.701	2.048	2.467
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981	29	1.311	1.699	2.045	2.462
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986	30	1.310	1.697	2.042	2.457
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990	$\infty$	1.282	1.645	1.960	2.326

Distribución Chi-Cuadrado $c_p(\nu)$								
$\nu$	$c_{0.025}$	$c_{0.05}$	$c_{0.10}$	$c_{0.90}$	$c_{0.95}$	$c_{0.975}$	$c_{0.99}$	$c_{0.995}$
1	0.00	0.00	0.02	2.71	3.84	5.02	6.63	7.88
2	0.05	0.10	0.21	4.61	5.99	7.38	9.21	10.60
3	0.22	0.35	0.58	6.25	7.81	9.35	11.34	12.84
4	0.48	0.71	1.06	7.78	9.49	11.14	13.28	14.86
5	0.83	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	3.82	4.57	5.58	17.28	19.68	21.92	24.72	26.76
12	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93