

# Capítulo 2

## Ruido en los sistemas de comunicaciones

Cuando una señal se transmite a través de un canal de comunicaciones hay dos tipos de imperfecciones que hacen que la señal recibida sea diferente de la señal transmitida. Una clase de imperfecciones es de naturaleza determinista, como es la distorsión. La segunda clase es no determinista, como el ruido. Para hacer un estudio de este tipo de fenómenos, se caracterizan como procesos aleatorios.

Por otro lado, la información que se transmite también se va a modelar, debido a su naturaleza, mediante procesos aleatorios. Esto se debe a que cualquier señal de información debe tener un grado de incertidumbre. Si no es así, no contiene información.

Por esta razón, en este capítulo se va a estudiar la teoría de los procesos aleatorios. Antes se va a hacer un pequeño resumen de algunos conceptos básicos de teoría de la probabilidad para luego introducir el concepto de variable aleatoria, y finalmente estudiar los procesos aleatorios y cómo se utilizan para modelar el ruido en un sistema de comunicaciones.

### 2.1. Probabilidad

En este apartado se repasarán, de forma breve, algunos de los conceptos básicos de la teoría de la probabilidad. Nos centraremos en los aspectos que son necesarios para el tratamiento de procesos aleatorios.

La teoría de la probabilidad trabaja con fenómenos que se producen de forma masiva. Hay un sin número de ejemplos: juegos de azar, movimiento de electrones, tasas de nacimiento y muerte, etc. Y lo que la teoría de la probabilidad trata de hacer es

establecer promedios para esos fenómenos. En particular, su propósito es describir y predecir estos promedios en términos de probabilidades de sucesos o eventos.

### 2.1.1. Espacio de probabilidad

Antes de poder definir lo que es un espacio de probabilidad, es necesario hacer varias definiciones.

#### Experimento aleatorio

El concepto fundamental en el que se basa la teoría de la probabilidad es el **experimento aleatorio**. Un experimento aleatorio es un experimento cuya salida no puede ser predicha con exactitud. Por ejemplo, tirar una moneda, tirar un dado, sacar una carta de la baraja, etc.

#### Espacio muestral (Espacio de muestras)

Todo experimento aleatorio tiene ciertos valores de salida. En el caso del lanzamiento de una moneda, cara o cruz, en el caso del dado, 1, 2, 3, 4, 5 o 6. Se define **espacio muestral** como el conjunto de todas las posibles salidas de un experimento. Normalmente se denota con la letra griega omega  $\Omega$ .

En cuanto a su naturaleza, existen dos tipos de espacios muestrales:

- discretos,
- no discretos (continuos).

Ejemplos de los primeros son el dado o la moneda antes mencionados. En ese caso el espacio de muestras es para la moneda cara y cruz, en el caso del dado, 1, 2, 3, 4, 5 y 6. Un ejemplo de variable aleatoria con un espacio muestral continuo es el valor del voltaje en una resistencia, que puede tomar cualquier valor dentro de un rango de valores de voltaje. En este caso el espacio de muestras es todo ese conjunto continuo de posibles valores.

#### Sucesos (Eventos)

Un suceso, o un evento, es un subconjunto del espacio de muestras sobre el que se pueda definir una probabilidad. Para que esta medida de probabilidad tenga sentido, hay

que imponer una serie de restricciones. Vamos primero a ver qué es una probabilidad.

La probabilidad de un suceso  $E$  es un número,  $P(E)$ , no negativo, normalmente entre 0 y 1 ( $0 \leq P(E) \leq 1$ ), asignado a ese evento y que describe lo probable o improbable que es dicho suceso. Este número se puede interpretar de la forma siguiente:

*Si un determinado experimento se realiza un número  $N$  de veces (suponiendo que  $N$  es suficientemente largo) y el evento  $A$  ocurre  $N_A$  veces, entonces podemos decir que la probabilidad será bastante cercana a la relación  $N_A/N$ :*

$$P(A) \approx \frac{N_A}{N}$$

Esta puede ser una definición intuitiva de probabilidad, es decir, que es una medida que nos indica lo frecuentemente que se produce un suceso cuando se realiza un cierto experimento.

Para el caso de espacios discretos, la idea es simple. Pero para el caso de espacios continuos hay un cierto matiz. ¿Cuál es la probabilidad de sacar un 5 al tirar un dado? Si el dado no está trucado, esta probabilidad es  $1/6$ . Pero, ¿cuál es la probabilidad de que el voltaje en una resistencia valga 1 V? La respuesta es 0. Aunque esto puede parecer anti-intuitivo, la explicación está en que el conjunto de valores que puede tomar es infinito, así que la probabilidad de tener uno de ellos es nula. En resumen, no es posible definir una probabilidad para un valor concreto. Lo que sí es posible es definir la probabilidad de que el valor de tensión esté en un cierto intervalo, por ejemplo entre 0.99 y 1.01 voltios. Ese suceso sí tiene una probabilidad.

Así pues, los sucesos en experimentos con espacios muestrales discretos han de estar formados por un subconjunto del espacio muestral, incluidos sucesos de un único elemento. Y en el caso de sucesos continuos, cada suceso ha de tener una probabilidad, así que hay que coger “regiones” del espacio muestral (no un único valor). Normalmente se define el campo sigma, denotado por  $B$ , como la colección de los subconjuntos de  $\Omega$ , es decir, de los sucesos.

Algunas definiciones sobre sucesos son las siguientes:

- *Suceso trivial:* es el que ocurre en todo experimento, es decir, que su probabilidad es 1. Ejemplo,  $\Omega$ .
- *Conjunto nulo ( $\emptyset$ ):* El que no tiene ningún elemento.
- *Unión de sucesos ( $E_1 \cup E_2$ ):* es el suceso que ocurre cuando sucede  $E_1$ ,  $E_2$  o ambos.
- *Intersección de sucesos ( $E_1 \cap E_2$ ):* el evento que ocurre cuando los eventos  $E_1$  y  $E_2$  se producen al mismo tiempo.

- *Complemento de un suceso* ( $E^c$ ): es el espacio muestral menos el propio suceso.
- *Eventos exclusivos o disjuntos*: aquellos para los que  $E_1 \cap E_2 = \emptyset$ . Para ellos se cumple que  $P(E_1 \cup E_2) = P(E_1) + P(E_2)$

### Espacio de probabilidad

El espacio de probabilidad se define como el triplete  $(\Omega, \mathcal{B}, P)$ , es decir, el espacio muestral, el espacio con los distintos sucesos y la medida de probabilidad que nos dice la probabilidad de cada suceso. Algunas de las propiedades que tienen estas probabilidades sobre sucesos son las siguientes:

1.  $P(E^c) = 1 - P(E)$ .
2.  $P(\emptyset) = 0$ .
3.  $P(E_1 \cup E_2) = P(E_1) + P(E_2) - P(E_1 \cap E_2)$ .
4. Si  $E_1 \subset E_2$  entonces  $P(E_1) \leq P(E_2)$ .

#### 2.1.2. Probabilidad condicional

Suponemos que existen dos sucesos,  $E_1$  y  $E_2$ , definidos sobre el mismo espacio de probabilidad con sus correspondientes probabilidades  $P(E_1)$  y  $P(E_2)$ . Si sabemos que uno de los eventos se ha producido, por ejemplo  $E_2$ , esto nos puede proporcionar cierta información sobre el otro que cambia su probabilidad con respecto al caso en el que no conocemos que ha sucedido  $E_2$ . A esta nueva probabilidad se le denomina probabilidad condicional, o condicionada. La propiedad condicional del suceso  $E_1$  dado el suceso  $E_2$ , denotada como  $P(E_1|E_2)$  se define como:

$$P(E_1|E_2) = \begin{cases} \frac{P(E_1 \cap E_2)}{P(E_2)}, & P(E_2) \neq 0 \\ 0, & P(E_2) = 0 \end{cases} .$$

#### Ejemplo

Se lanza un dado no cargado  
 $E_1$ : resultado mayor que 3  
 $E_2$ : resultado par

$$P(E_1) = P(4) + P(5) + P(6) = \frac{1}{2}$$

$$P(E_2) = P(2) + P(4) + P(6) = \frac{1}{2}$$

$$P(E_1 \cap E_2) = P(4) + P(6) = \frac{1}{3}$$

La probabilidad de  $E_1|E_2$  es

$$P(E_1|E_2) = \frac{1/3}{1/2} = \frac{2}{3}$$

Se comprueba que el resultado obtenido coincide con la probabilidad de tener un 4 o un 6 cuando el espacio muestral es el suceso  $E_2$ .

### Sucesos estadísticamente independientes

De la probabilidad condicional se deriva una importante definición estadística. Si ocurre que  $P(E_1|E_2) = P(E_1)$  esto significa que el conocimiento de  $E_2$  no aporta información sobre  $E_1$  y por tanto no cambia su probabilidad con respecto a la probabilidad a priori (sin el conocimiento de que se ha producido  $E_2$ ). En este caso, se dice que los dos sucesos son estadísticamente independientes. Para este tipo de sucesos, se tiene que

$$P(E_1 \cap E_2) = P(E_1) \cdot P(E_2).$$

### Teorema de la probabilidad total

Si los sucesos  $E_i$ , con  $i = 1, \dots, N$  forman una partición del espacio muestral  $\Omega$ , lo que quiere decir que se cumplen las siguientes condiciones

- $\cup_{i=1}^N E_i = \Omega$
- $E_i \cap E_j = \emptyset$  para todo  $i \neq j$

entonces, si para un suceso  $A$  se dispone de las probabilidades condicionales  $P(A|E_i)$  para todos los eventos de la partición,  $i = 1, \dots, N$ , la probabilidad  $P(A)$  se obtiene mediante el *teorema de la probabilidad total*

$$P(A) = \sum_{i=1}^N P(A|E_i)P(E_i).$$

**Regla de Bayes**

Por otro lado, la *Regla de Bayes* (aunque su idea se debe a Bayes, finalmente la formuló Laplace) nos dice que las probabilidades condicionales de los sucesos de la partición dado  $A$ ,  $P(E_i|A)$ , se obtienen mediante la siguiente expresión

$$P(E_i|A) = \frac{P(A|E_i)P(E_i)}{P(A)} = \frac{P(A|E_i)P(E_i)}{\sum_{j=1}^N P(A|E_j)P(E_j)}.$$

## 2.2. Variable aleatoria

Una *variable aleatoria* (*v.a.*) (real) no es más que una función que asigna un número del conjunto de números reales a cada una de las posibles salidas de un experimento aleatorio, es decir, a cada uno de los elementos del espacio muestral.

$$\begin{aligned}\Omega &\rightarrow \mathbb{R} \\ \omega \in \Omega &\rightarrow X(\omega) \in \mathbb{R}\end{aligned}$$

Por tanto, una v.a. mapea los resultados de un experimento aleatorio en la recta real.

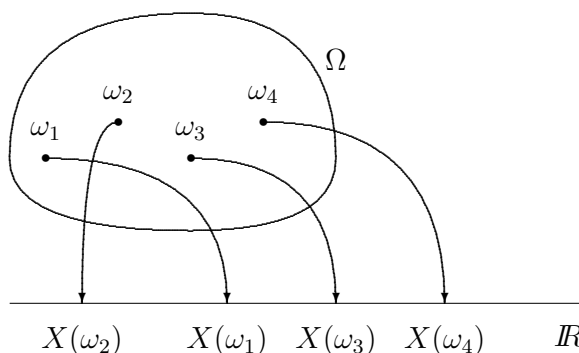


Figura 2.1: Variable aleatoria como un mapeo de  $\Omega$  a  $\mathbb{R}$ .

Por ejemplo, en el experimento lanzar un dado la asignación ya existe, pero en el caso de lanzar una moneda, es posible asignar un cero a la cara y un uno a la cruz, etc. Las variables aleatorias normalmente se denotan con mayúscula  $X$ ,  $Y$ , y no se suele expresar la dependencia implícita con  $\omega$ . De nuevo, al clasificar en cuanto al tipo de valores que puede tomar, vamos a tener principalmente dos categorías de variable aleatoria:

- Discreta: número finito de valores.
- Continua: rango continuo de valores (uno o varios intervalos).

*Rango* (o Recorrido) de una v.a. es el conjunto de números reales que tienen asociado un resultado del espacio muestral, es decir:

$$\text{Rango}_X = \{x \in \mathbb{R} \mid \exists \omega \in \Omega, X(\omega) = x\}.$$

Probabilísticamente, una variable aleatoria se caracteriza mediante dos funciones:

- Función de distribución,  $F_X(x)$ .
- Función densidad de probabilidad,  $f_X(x)$ .

### 2.2.1. Función de distribución

La *función de distribución (FD)* de una variable aleatoria se define como

$$F_X(x) = P(X \leq x),$$

es decir, como la probabilidad de que la variable aleatoria  $X$  tome un valor menor o igual que  $x$ . Las principales propiedades de la función de distribución son las siguientes:

1.  $0 \leq F_X(x) \leq 1$ .
2.  $x_1 < x_2 \rightarrow F_X(x_1) \leq F_X(x_2)$  ( $F_X(x)$  es no decreciente).
3.  $F_X(-\infty) = 0$  y  $F_X(\infty) = 1$  ( $\lim_{x \rightarrow -\infty} F_X(x) = 0$  y  $\lim_{x \rightarrow \infty} F_X(x) = 1$ ).
4.  $F_X(x^+) = F_X(x)$  ( $F_X(x)$  es continua por la derecha).
5.  $F_X(b) - F_X(a) = P(a < X \leq b)$ .

Para calcular otras probabilidades incluyendo o no los límites del intervalo

$$\begin{aligned} P(a \leq X \leq b) &= F_X(b) - F_X(a^-). \\ P(a < X < b) &= F_X(b^-) - F_X(a). \\ P(a \leq X < b) &= F_X(b^-) - F_X(a^-). \end{aligned}$$

6.  $P(X = a) = F_X(a) - F_X(a^-)$ .
7.  $P(X > x) = 1 - F_X(x)$ .

En las expresiones anteriores,

$$F_X(x^\pm) = \lim_{\varepsilon \rightarrow 0} F_X(x \pm \varepsilon).$$

Esta distinción  $F_X(x^\pm)$  se realiza para tener en cuenta el caso particular de funciones de distribución para v.a. discretas, para las que  $F_X(x_i^-) \neq F_X(x_i)$ , siendo  $\{x_i\}$  el conjunto discreto de valores que forman el rango de  $X$ . En general, para variables aleatorias continuas  $F_X(x) = F_X(x^-)$ , lo que implica que la probabilidad de tomar un valor concreto,  $P(X = a) = 0$ . (Y para ambas, discreta y continuas,  $F_X(x) = F_X(x^+)$ , ver propiedad 4).



Para variables aleatorias discretas  $F_X(x)$  es una función del tipo escalera, con discontinuidades en los valores discretos que forman el rango de la variable aleatoria. Para una variable continua tiene una variación continua. La Figura 2.2 muestra ejemplos de función de distribución discreta, en este caso el experimento lanzar un dado, y continua.

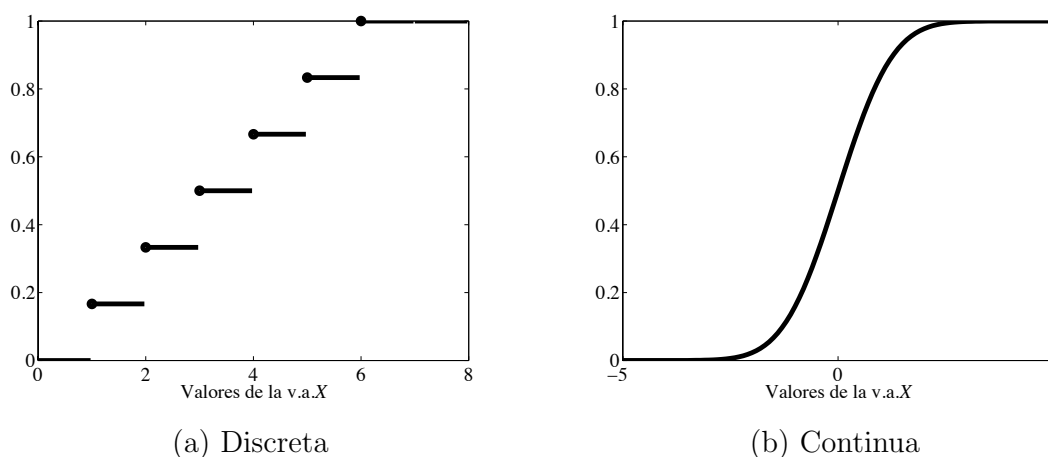


Figura 2.2: Ejemplos de v.a. discreta y v.a. continua

### Interpretación frecuencial (probabilística)

Para presentar una interpretación empírica, constructiva, de la función de distribución, podemos escribir:

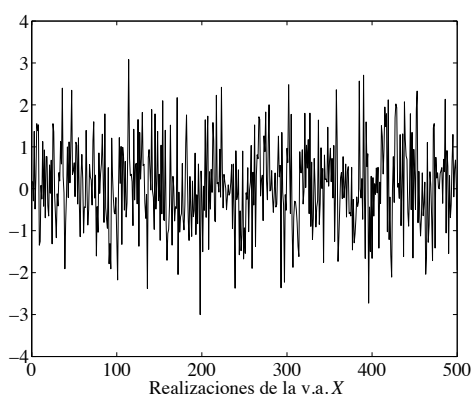
$$F_X(x) = P(X \leq x) = \lim_{n \rightarrow \infty} \frac{n_x}{n},$$

donde  $n$  es el número de realizaciones del experimento aleatorio, y  $n_x$  es el número de resultados para los cuales  $X \leq x$ . Obviamente no podremos nunca realizar un número infinito de experimentos, pero podemos realizar una estima a partir de un número limitado de los mismos. La Figura 2.3 muestra 500 realizaciones de un experimento y la estima realizada de este modo comparada con la función de distribución teórica.

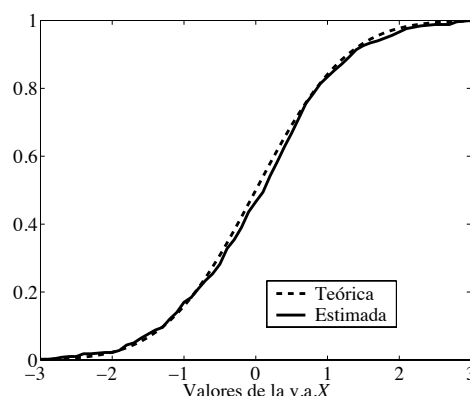
#### 2.2.2. Función de densidad de probabilidad

La otra función empleada para caracterizar una variable aleatoria es la *función densidad de probabilidad (f.d.p.)*, que se denota como  $f_X(x)$ . La función de densidad de probabilidad se define como la derivada de la función de distribución

$$f_X(x) = \frac{d}{dx} F_X(x).$$



(a) Realizaciones



(b) Estima de la función de distribución

Figura 2.3: Estima de la función de distribución mediante su interpretación frecuencial

Esta función indica como se distribuye la probabilidad de la variable aleatoria. Sus principales propiedades son las siguientes:

1.  $f_X(x) \geq 0$ .
2.  $\int_{-\infty}^{\infty} f_X(x) dx = 1$ .
3.  $\int_{a^+}^{b^+} f_X(x) dx = P(a < X \leq b)$ .
4. En general,  $P(X \in A) = \int_A f_X(x) dx$ .
5.  $F_X(x) = \int_{-\infty}^{x^+} f_X(u) du$ .

En el caso de variables continuas tiene una variación continua, y en el caso de variables discretas, la f.d.p. incluye impulsos situados en los valores discretos que puede tomar la variable (la derivada de una función con escalones). El valor en cada uno de esos valores discretos corresponde a la probabilidad de que la variable aleatoria tome dicho valor.

El matiz  $a^+$  sirve para tratar las señales discretas. En este caso, el impulso está situado en  $a$ , e integrar desde  $a^+$  no lo incluye. Para variables continuas podemos utilizar directamente  $a$ .

En el caso de variables discretas, en ocasiones en lugar de trabajar con la f.d.p., se trabaja con la **función masa de probabilidad**, o a veces los llamados *puntos de*

*masa*. En el caso de una variable discreta, sólo unos valores concretos  $\{x_i\}_{i=1, \dots, N}$  son posibles. En ese caso se define la función masa de probabilidad o puntos de masa como

$$\{p_i\} = P(X = x_i).$$

En este caso se cumple que

1.  $p_i \geq 0$ .
2.  $\sum_{i=1}^N p_i = 1$ .

La diferencia con la f.d.p. es que se suele representar en función de  $i$  en lugar de con respecto a  $x_i$ , pero conceptualmente es lo mismo.

En otras ocasiones, para variables aleatorias discretas, una vez conocido el espacio muestral  $\{x_i\}_{i=1, \dots, N}$ , las probabilidades de cada uno de los valores de dicho espacio se denotan como  $p_X(x_i)$ .

En este curso en general se trabajará con la f.d.p., pero cuando se trabaje con variables aleatorias discretas, en ocasiones en lugar de utilizar la notación  $f_X(x)$  se utilizará la notación  $p_X(x_i)$ .

### Interpretación frecuencial

Para dar una interpretación empírica de la f.d.p., podemos definir la función densidad de probabilidad como

$$f_X(x) = \lim_{\Delta x \rightarrow 0} \frac{P(x \leq X \leq x + \Delta x)}{\Delta x},$$

es decir

$$f_X(x) = \frac{\text{Probabilidad de un intervalo}}{\text{Longitud del intervalo}} = \text{Densidad de Probabilidad},$$

cuando la longitud del intervalo se lleva al límite infinitesimal. Utilizando la definición frecuencial de la probabilidad,

$$f_X(x) = \lim_{\Delta x \rightarrow 0} \left\{ \frac{1}{\Delta x} \lim_{n \rightarrow \infty} \frac{n_x}{n} \right\},$$

donde  $n$  es el número de realizaciones del experimento aleatorio, y  $n_x$  es el número de resultados para los cuales  $x \leq X \leq x + \Delta x$ .

Esto es equivalente a hacer un histograma, que consiste en dividir la recta real en intervalos de anchura  $\Delta x$  y levantar una barra vertical con la frecuencia relativa de cada intervalo. En este caso, se puede comprobar que un histograma tiende a la función densidad de probabilidad cuando el número de realizaciones crece y la longitud del intervalo disminuye. La Figura 2.4 muestra un histograma con un valor  $\Delta x = 0.2$  realizado a partir de 1000 realizaciones y lo compara con la función densidad de probabilidad teórica para una distribución gaussiana de media nula.

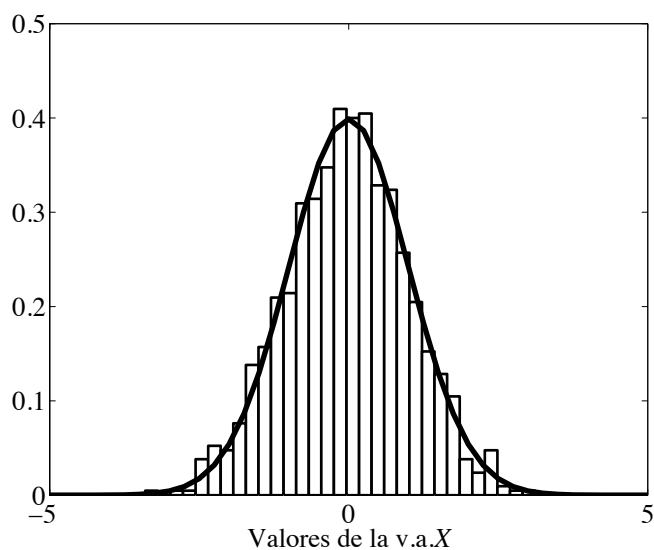


Figura 2.4: Aproximación de la f.d.p. mediante un histograma.

### 2.2.3. Variables aleatorias de interés

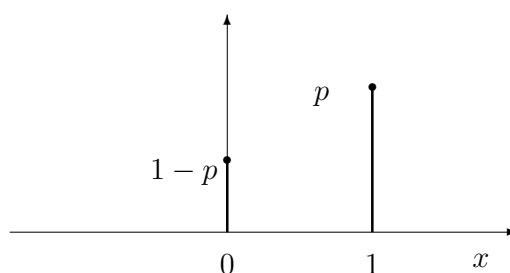
A continuación vamos a ver las variables aleatorias más frecuentemente utilizadas en comunicaciones.

#### Variable aleatoria de Bernoulli

Esta es una variable aleatoria discreta que toma dos valores, 1 y 0, con probabilidades

- $P(1) = p$ ,
- $P(0) = 1 - p$ ,

respectivamente.

Figura 2.5:  $f_X(x)$  de una v.a. de Bernoulli.

Se trata de una distribución con un parámetro, en este caso  $p$ . Su función densidad de probabilidad es, obviamente:

$$f_X(x) = \begin{cases} 1 - p, & x = 0 \\ p, & x = 1 \\ 0, & \text{en otro caso} \end{cases} .$$

Una variable aleatoria de Bernoulli es un buen modelo para

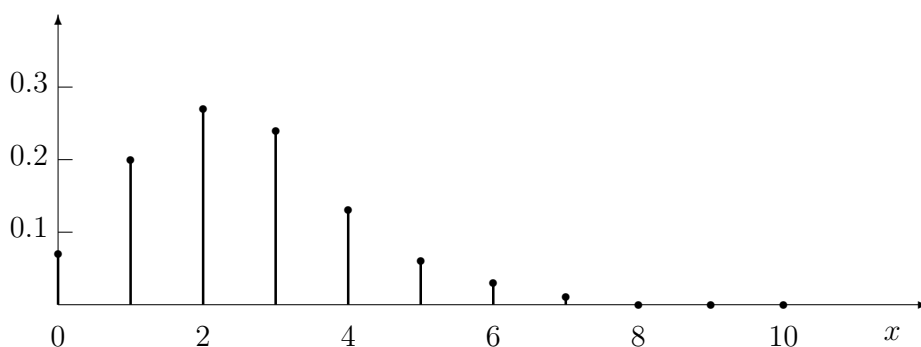
- *Generador de datos binario.* En este caso, lo normal es que el parámetro  $p$  valga  $1/2$ , es decir, que los 1's y los 0's sean equiprobables.
- *Modelo de errores.* Por otro lado, en cualquier transmisión sobre un canal de comunicaciones se van a producir errores. Un error se puede modelar como la suma módulo-2 (XOR) del bit de entrada con un 1. Por tanto, este tipo de variables también se pueden emplear para modelar errores. En este caso, el parámetro  $p$  es precisamente la tasa de errores.

### Variable aleatoria binomial

Es también una variable aleatoria discreta. Esta variable modela el número de 1's en una secuencia de  $n$  experimentos de Bernoulli independientes, con lo que tiene dos parámetros,  $n$  y  $p$ . Su función densidad de probabilidad es la siguiente:

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & 0 \leq x \leq n \text{ y } x \in \mathbb{Z} \\ 0, & \text{en otro caso} \end{cases} .$$

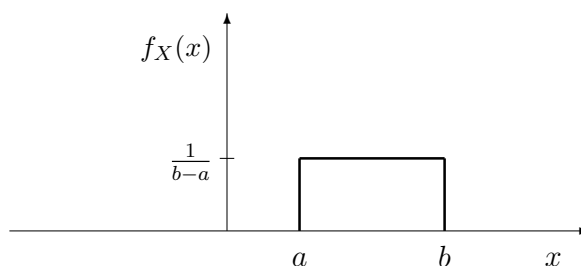
Esta variable se utiliza, por ejemplo, para modelar el *número total de bits recibidos con error* cuando una secuencia de  $n$  bits es transmitida a través de un canal con probabilidad de error de bit  $p$ .

Figura 2.6:  $f_X(x)$  de una v.a. binomial.

### Variable aleatoria uniforme

Esta es una variable aleatoria continua de dos parámetros,  $a$  y  $b$ , que toma valores en el intervalo  $(a,b)$  con la misma probabilidad para intervalos de igual longitud. Su función de densidad es

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{en otro caso} \end{cases} .$$

Figura 2.7:  $f_X(x)$  de una v.a. uniforme.

Este modelo se utiliza para variables continuas con rango conocido para las cuales nada más se conoce. Por ejemplo, para modelar una *fase aleatoria en una senoide*, se suele emplear una v.a. uniforme entre 0 y  $2\pi$ .

### Variable aleatoria gaussiana o normal

Se trata de una variable aleatoria continua con dos parámetros,  $\mu$  y  $\sigma$ . Su función densidad de probabilidad es una gaussiana de media  $\mu$  y varianza  $\sigma^2$  (o lo que es lo mismo, desviación típica  $\sigma$ ),

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} .$$

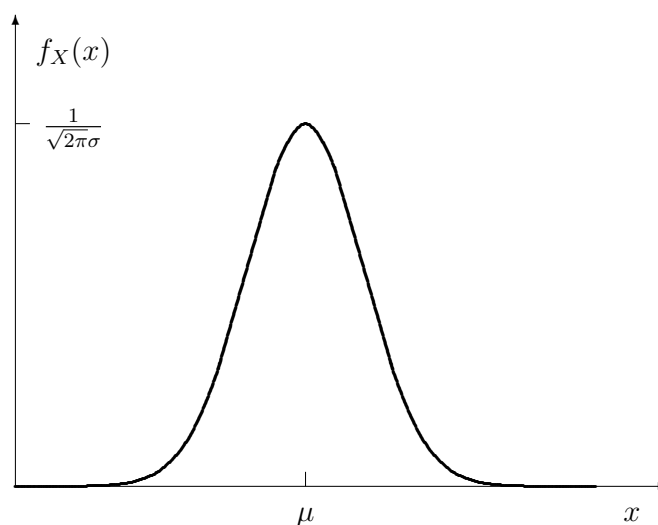


Figura 2.8: Función densidad de probabilidad para una v.a. gaussiana

En ocasiones se denota como  $\mathcal{N}(\mu, \sigma^2)$ . La gaussiana es la v.a. más importante y la más utilizada sin duda en comunicaciones. La principal razón es que el ruido térmico, que es la mayor fuente de ruido en los sistemas de comunicaciones, tiene una distribución gaussiana.

La función de distribución,  $F_X(x)$ , para una v.a. gaussiana de media nula y varianza unidad se denota comúnmente como  $\Phi(x)$

$$\Phi(x) = P(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

Una función relacionada con esta función de distribución, que se utiliza con mucha frecuencia es la función  $Q(x) = 1 - \Phi(x)$ , lo que proporciona  $P(X > x)$ , que es de interés, como ya veremos, para evaluar probabilidades de error. Algunas de sus propiedades son

1.  $Q(-x) = 1 - Q(x)$ .
2.  $Q(0) = \frac{1}{2}$ .
3.  $Q(\infty) = 0$ .

Esta función no tiene solución analítica ( $\Phi(x)$  no la tiene), pero es fácil calcularla de forma numérica y normalmente se presenta tabulada para sus valores positivos.

Para una distribución  $\mathcal{N}(\mu, \sigma^2)$ , un simple cambio de variable sirve para estimar

$P(X > x)$ ,

$$P(X > x) = Q\left(\frac{x - \mu}{\sigma}\right).$$

### 2.2.4. Funciones de una variable aleatoria

Una función de una variable aleatoria  $Y = g(X)$  es también ella misma un variable aleatoria. Para encontrar su función de distribución podemos partir de la definición de la función de distribución

$$F_Y(y) = P(Y \leq y) = P(g(X) \leq y)$$

Esta probabilidad se reduce a

$$F_Y(y) = P(x \in B_x),$$

donde  $B_x$  es

$$B_x = \{x \in \mathbb{R} \mid g(x) \leq y\}.$$

#### Ejemplo

Para la transformación  $Y = -2X$ , queremos calcular  $F_Y(y)$ .

En este caso, es sencillo calcular  $B_x$ ,

$$B_x = \{x \in \mathbb{R} \mid -2x \leq y\} = \{x \geq -y/2\},$$

de modo que

$$F_Y(y) = P(Y \leq y) = P(X \geq -y/2).$$

Esta probabilidad se puede calcular conocida  $F_X(x)$  o  $f_X(x)$ .

Por otro lado, la función densidad de probabilidad de la variable aleatoria  $Y$  se puede calcular, a partir de  $f_X(x)$  y de la transformación  $g(x)$ , como

$$f_Y(y) = \sum_i \frac{f_X(x_i)}{|g'(x_i)|},$$

donde  $\{x_i\}$  son las soluciones de la ecuación  $y = g(x)$  y  $g'(x)$  es la derivada de la misma. Para poder obtener esta expresión es preciso que la ecuación tenga un número finito de soluciones y que para todas estas soluciones exista  $g'(x_i)$  y no sea nula.

#### Ejemplo



Tenemos una variable aleatoria  $X$  gaussiana con media nula y varianza unidad, es decir  $\mu = 0$  y  $\sigma = 1$ . Queremos encontrar la función densidad de probabilidad de la variable aleatoria

$$Y = aX + b.$$

En este caso  $g(x) = ax + b$ , y por tanto  $g'(x) = a$ . La ecuación  $y = ax + b$  tiene una única solución

$$x_1 = \frac{y - b}{a}$$

Aplicando la expresión para calcular la f.d.p. tenemos

$$f_Y(y) = \frac{f_X\left(\frac{y-b}{a}\right)}{|a|} = \frac{1}{\sqrt{2\pi}|a|} e^{-\frac{(y-b)^2}{2a^2}}.$$

Se puede comprobar que es una gaussiana  $\mathcal{N}(b, a^2)$

De este ejemplo podemos sacar una conclusión importante: *una función lineal de una variable aleatoria gaussiana es también una variable aleatoria gaussiana.*

### 2.2.5. Momentos estadísticos

A continuación vamos a ver como se calculan algunos momentos estadísticos asociados a una variable aleatoria. No conviene olvidar que una variable aleatoria representa la salida de un experimento aleatorio. Si se conoce la f.d.p. es posible obtener algunos estadísticos de la misma, lo que equivale a decir estadísticos del experimento aleatorio.

#### Valor esperado (Media)

El valor esperado (esperanza matemática) de una variable aleatoria es equivalente a su media (aritmética), y a menudo se denota como  $m_X$ . El valor esperado mide el valor medio obtenido cuando el número de experimentos es suficientemente grande. Este valor esperado se define como

$$m_X = E(X) = \int_{-\infty}^{\infty} x \cdot f_X(x) dx.$$

#### Valor esperado de una función de $X$

El valor esperado de la variable aleatoria  $Y = g(X)$  se obtiene como

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) \cdot f_X(x) dx.$$

**Momento de orden  $n$** 

En general, el momento de orden  $n$  nos da el valor esperado (la media) de  $x^n$ , y se define como

$$m_X^n = \int_{-\infty}^{\infty} x^n \cdot f_X(x) dx.$$

En este caso, la función es  $g(x) = x^n$ . El valor esperado, la media, es por tanto el momento de orden 1.

**Varianza**

La varianza se puede ver como el valor esperado para el caso particular

$$g(x) = (x - m_X)^2.$$

Por tanto,

$$\sigma_X^2 = \int_{-\infty}^{\infty} (x - m_X)^2 \cdot f_X(x) dx.$$

$\sigma_X^2$  es la varianza de la v.a. y  $\sigma_X$  es por tanto la desviación típica. Estos parámetros nos dan idea de la variabilidad de la v.a. Como curiosidad, tenemos que

$$\sigma_X^2 = E((X - E(X))^2) = E(X^2) - [E(X)]^2.$$

$$\sigma_X^2 = E((X - m_X)^2) = E(X^2) - (m_X)^2.$$

**Propiedades**

A continuación se presentan algunas de las propiedades de estos estadísticos. Para una constante  $c$

1.  $E[X + Y] = E[X] + E[Y] = m_X + m_Y$  (Operador lineal)
2.  $E[c] = c$
3.  $E[c \cdot X] = c \cdot E[X]$
4.  $E[X + c] = E[X] + c$
5.  $\text{Var}(c) = 0$
6.  $\text{Var}(c \cdot X) = c^2 \cdot \text{Var}(X)$
7.  $\text{Var}(X + c) = \text{Var}(X)$

### 2.2.6. Variables aleatorias multidimensionales

Si dos variables aleatorias están definidas sobre el mismo espacio muestral  $\Omega$ , es posible trabajar con ellas de forma conjunta. Este caso podemos plantearlo como un problema multidimensional, o también como un problema de vectores de variables aleatorias. En este caso seguiremos la primera alternativa.

#### Funciones de distribución y densidad de probabilidad conjuntas

En este caso se define su *función de distribución conjunta* como

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y).$$

Y la *función densidad de probabilidad conjunta* como

$$f_{X,Y}(x, y) = \frac{\partial^2}{\partial x \partial y} F_{X,Y}(x, y).$$

Estas dos funciones tienen las siguientes propiedades (la mayoría extensión de las propiedades para el caso de una única variable aleatoria)

1.  $F_X(x) = F_{X,Y}(x, \infty)$ .
2.  $F_Y(y) = F_{X,Y}(\infty, y)$ .
3.  $f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dy$ .
4.  $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx$ .
5.  $\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$ .
6.  $P((X, Y) \in A) = \int \int_{(x,y) \in A} f_{X,Y}(x, y) dx dy$ .
7.  $F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(u, v) du dv$ .

### Función de densidad condicional

Como sucedía para el caso de sucesos, el hecho de conocer el resultado de una variable aleatoria condiciona el conocimiento que se tiene sobre la otra. La función de densidad de probabilidad de la variable  $Y$  condicionada por  $X = x$  se define como

$$f_{Y|X}(y|x) = \begin{cases} \frac{f_{X,Y}(x,y)}{f_X(x)}, & f_X(x) \neq 0 \\ 0, & \text{en otro caso} \end{cases}.$$

De aquí surge la definición de **variables aleatorias estadísticamente independientes**. Si el conocimiento de  $X$  no aporta nada sobre el conocimiento de  $Y$ , entonces

$$f_{Y|X}(y|x) = f_Y(y).$$

Para este tipo de variables aleatorias se cumple

$$f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y).$$

### Momentos estadísticos

El valor esperado de una función  $g(X, Y)$  de las variables aleatorias  $X$  e  $Y$  se obtiene como

$$E(g(X, Y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot f_{X,Y}(x, y) dx dy.$$

Es interesante resaltar los siguientes casos particulares:

- Si  $g(X, Y) = X \cdot Y$ , se tiene la esperanza del producto de las dos variables aleatorias, que se denomina la **correlación** entre  $X$  e  $Y$ .
- En el caso en que  $g(X, Y) = (X - m_X) \cdot (Y - m_Y)$  tenemos la denominada **covarianza**.

La versión normalizada de la covarianza es lo que se conoce como **coeficiente de correlación**,  $\rho_{X,Y}$ , que se define como

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

Su módulo está limitado entre 0 y 1, es decir  $0 \leq |\rho_{X,Y}| \leq 1$ . Algunos valores de esta variable nos aportan una información especial sobre las variables aleatorias implicadas.

- Cuando  $\rho_{X,Y} = 0$  se dice que las señales están **inacorreladas**. Si dos variables aleatorias son independientes, entonces es fácil comprobar que están incorreladas. Sin embargo, lo recíproco no es cierto: incorrelación no implica independencia.

- Por otro lado, un valor  $\rho_{X,Y} = \pm 1$  indica una relación lineal entre las variables aleatorias, es decir  $Y = aX + b$ . En este caso,  $\rho_{X,Y} = 1$  indica un valor positivo de  $a$ , mientras que  $\rho_{X,Y} = -1$  indica que  $a$  es negativo.

Es común utilizar la notación  $\rho$ , sin hacer referencia a las variables aleatorias implicadas que se sobreentienden.

De forma intuitiva, la correlación nos va a indicar el grado de relación estadística entre las dos variables aleatorias. En general, una correlación alta indica una relación alta, y una correlación baja suele indicar una relación baja.

### Funciones de variables aleatorias multidimensionales

Sobre variables aleatorias multidimensionales (o múltiples), al igual que para las unidimensionales, se pueden definir funciones sobre las variables  $X$  e  $Y$

$$\begin{cases} Z = g(X, Y) \\ W = h(X, Y) \end{cases} .$$

Para obtener  $F_{Z,W}(z, w)$  se procede como en el caso unidimensional.

$$F_{Z,W}(z, w) = P(Z \leq z, W \leq w) = P((x, y) \in B_{xy}),$$

donde en este caso

$$B_{xy} = \{(x, y) \in \mathbb{R}^2 \mid g(x, y) \leq z, h(x, y) \leq w\}.$$

Al igual que en el caso de una única v.a., si se conocen las raíces (soluciones)  $\{x_i, y_i\}$  del sistema de ecuaciones

$$\begin{cases} z = g(x, y) \\ w = h(x, y) \end{cases} ,$$

entonces la f.d.p. de las nuevas variables se obtiene mediante la expresión

$$f_{Z,W}(z, w) = \sum_i \frac{f_{X,Y}(x_i, y_i)}{|\det \mathbf{J}(x_i, y_i)|}.$$

donde  $\det \mathbf{J}$  denota el determinante de la matriz jacobiano  $\mathbf{J}$ . Se necesita que el número de soluciones sea finito y que el jacobiano sea no nulo. El jacobiano se define como

$$\mathbf{J}(x, y) = \begin{bmatrix} \frac{\partial z(x,y)}{\partial x} & \frac{\partial z(x,y)}{\partial y} \\ \frac{\partial w(x,y)}{\partial x} & \frac{\partial w(x,y)}{\partial y} \end{bmatrix}.$$

De nuevo es necesario que el número de raíces sea finito y que el determinante sea no nulo para todas ellas.

Todo lo que hemos estado viendo aplicado a dos variables aleatorias se puede extender de forma inmediata a un número mayor de variables aleatorias.

### Variables aleatorias conjuntamente gaussianas

O lo que es lo mismo gaussianas multidimensionales. Vamos a ver algunas de sus propiedades. Dos variables aleatorias  $X$  e  $Y$  conjuntamente gaussianas están caracterizadas por una función densidad de probabilidad conjunta

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{-\frac{1}{(1-\rho^2)}\left(\frac{(x-\mu_X)^2}{2\sigma_X^2} + \frac{(y-\mu_Y)^2}{2\sigma_Y^2} - \frac{\rho(x-\mu_X)(y-\mu_Y)}{\sigma_X\sigma_Y}\right)}$$

Cuando se tiene este tipo de distribución conjuntamente gaussiana en las variables  $X$  e  $Y$ , no sólo  $X$  e  $Y$  van a tener una distribución gaussiana (son v.a. gaussianas) sino que además las probabilidades condicionales también son gaussianas. Esta es la principal diferencia entre dos variables aleatorias que cada una tiene una distribución gaussiana y dos variables aleatorias con una distribución conjuntamente gaussiana. Con una distribución conjuntamente gaussiana, las variables aleatorias individuales son de la forma siguiente:  $X$  es gaussiana de media  $\mu_X$  y varianza  $\sigma_X^2$ ,  $Y$  es gaussiana de media  $\mu_Y$  y varianza  $\sigma_Y^2$ , y además su coeficiente de correlación es  $\rho$ .

Este concepto se puede extender a más variables aleatorias (más dimensiones) llegándose a una expresión de la forma

$$f_{\mathbf{X}}(x_1, x_2, \dots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det(\mathbf{C})}} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})\mathbf{C}^{-1}(\mathbf{x}-\boldsymbol{\mu})^T}.$$

donde la variable aleatoria  $\mathbf{X} = (X_1, X_2, \dots, X_n)$ ,  $\mathbf{x} = [x_1, x_2, \dots, x_n]^T$ , y el vector de medias es  $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]^T$ . Finalmente,  $\mathbf{C}$  es la matriz de covarianza, con

$$C_{i,j} = \text{Cov}(X_i, X_j) = \rho_{i,j}\sigma_i\sigma_j$$

es decir,

$$\mathbf{C} = \begin{bmatrix} \sigma_1^2 & \rho_{1,2}\sigma_1\sigma_2 & \cdots & \rho_{1,n}\sigma_1\sigma_n \\ \rho_{1,2}\sigma_1\sigma_2 & \sigma_2^2 & \cdots & \rho_{2,n}\sigma_2\sigma_n \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,n}\sigma_1\sigma_n & \rho_{2,n}\sigma_2\sigma_n & \cdots & \sigma_n^2 \end{bmatrix}.$$

Las propiedades de las variables aleatorias conjuntamente gaussianas son

1. Las variables aleatorias conjuntamente gaussianas están completamente caracterizadas por su vector de medias  $\boldsymbol{\mu}$  y su matriz de covarianza  $\mathbf{C}$ . A estos dos parámetros se les denomina propiedades de *segundo orden*, y describen completamente estas variables aleatorias.

2. Si  $n$  variables aleatorias son conjuntamente gaussianas, cualquier subconjunto también está distribuido de forma conjuntamente gaussiana. En particular, todas las variables individuales son gaussianas.
3. Cualquier subconjunto de v.a. conjuntamente gaussianas, condicionadas a otro subconjunto de las mismas v.a. conjuntamente gaussianas originales, tiene una distribución conjuntamente gaussiana, aunque los parámetros se modifican en este caso.
4. Cualquier conjunto de variables aleatorias obtenidas como combinaciones de lineales de  $(X_1, X_2, \dots, X_n)$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n,1} & a_{n,2} & \cdots & a_{n,n} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

es conjuntamente gaussiano. En particular, individualmente cualquier combinación lineal  $Y_i$  es gaussiana.

5. Dos variables aleatorias conjuntamente gaussianas incorreladas son independientes. Por tanto, *para v.a. conjuntamente gaussianas, independencia e incorrelación son equivalentes*. Esto no es cierto en general para otro tipo de v.a.
6. Si las señales están incorreladas,  $\rho_{i,j} = 0 \forall i \neq j$ ,  $\mathbf{C}$  es una matriz diagonal.

### Suma de variables aleatorias

Si tenemos una secuencia de variables aleatorias,  $(X_1, X_2, \dots, X_n)$ , que tienen básicamente las mismas propiedades, parece lógico pensar que el comportamiento del promedio de las mismas,

$$Y = \frac{1}{n} \sum_{i=1}^n X_i,$$

sea, por así decirlo, “menos aleatorio”. La *ley de los grandes números* y el *teorema del límite central* plantean de forma rigurosa esta intuición.

**Ley de los grandes números (débil)** Esta ley plantea que si las variables aleatorias  $(X_1, X_2, \dots, X_n)$  están *incorreladas* y todas tienen la misma media  $m_X$  y varianza  $\sigma_X^2 < \infty$ , independientemente de su distribución, para cualquier  $\varepsilon > 0$ ,

$$\lim_{n \rightarrow \infty} P(|Y - m_X| > \varepsilon) = 0.$$

Esto significa que el promedio ( $Y$ ) converge, en probabilidad, al valor esperado de las v.a.  $X_i$ . Es decir, que cuantas más variables sumemos, más se parece su combinación a la media (menor es su varianza).

**Teorema del límite central** Este teorema va un poco más allá. No sólo dice que el promedio de v.a. converge a la media sino que nos dice como es su distribución. En concreto, el teorema plantea que: si  $(X_1, X_2, \dots, X_n)$  son *independientes* con medias  $m_1, m_2, \dots, m_n$ , y varianzas  $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ , entonces la distribución de

$$Y = \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{X_i - m_i}{\sigma_i}$$

converge a una distribución gaussiana de media 0 y varianza 1,  $\mathcal{N}(0, 1)$ .

En el caso particular de que sean *independientes e idénticamente distribuidas (i.i.d)*, es decir, que todas tengan la misma distribución con la misma media  $m$  y la misma varianza  $\sigma^2$ , el promedio

$$Y = \frac{1}{n} \sum_{i=1}^n X_i,$$

converge a una distribución  $\mathcal{N}(m, \frac{\sigma^2}{n})$ . Esto es así aunque la distribución original no sea gaussiana.

**Nota:** Recordar que la ley de los grandes números es válida para señales incorreladas mientras que el teorema del límite central exige independencia.