



# CURSO DE PROBABILIDAD

Dra. Begoña Vitoriano

[bvitoriano@mat.ucm.es](mailto:bvitoriano@mat.ucm.es)

Universidad Complutense de Madrid

Octubre 2021

---



# ÍNDICE

<b>INTRODUCCIÓN</b> .....	<b>1</b>
<b>I ESPACIO PROBABILIZABLE</b> .....	<b>3</b>
I.1 EXPERIMENTOS ALEATORIOS Y SUCESOS .....	3
I.2 ÁLGEBRAS Y $\sigma$ -ÁLGEBRAS.....	5
I.3 $\sigma$ -ÁLGEBRAS MÁS INTERESANTES.....	7
<b>II ESPACIO DE PROBABILIDAD</b> .....	<b>8</b>
II.1 FUNCIÓN DE PROBABILIDAD Y PROPIEDADES.....	9
II.2 PROBABILIDAD EN EL ESPACIO MUESTRAL DISCRETO.....	10
II.3 PROBABILIDAD EN EL ESPACIO PROBABILIDAD $(\mathbb{R}, \mathcal{B})$ .....	11
II.4 PROBABILIDAD CONDICIONADA .....	12
II.5 INDEPENDENCIA DE SUCESOS .....	14
II.6 FUNCIÓN DE DISTRIBUCIÓN EN $(\mathbb{R}, \mathcal{B})$ .....	15
II.6.1 <i>Clasificación de las funciones de distribución</i> .....	18
II.6.1.1 Funciones de distribución discretas .....	18
II.6.1.2 Funciones de distribución absolutamente continuas .....	18
II.6.1.3 Funciones de distribución mixtas.....	19
II.6.2 <i>Convergencia en Ley o en Distribución</i> .....	20
<b>III VARIABLES ALEATORIAS</b> .....	<b>21</b>
III.1 DEFINICIÓN DE VARIABLE ALEATORIA.....	21
III.2 OPERACIONES CON VARIABLES ALEATORIAS.....	24
III.3 INTEGRACIÓN DE VARIABLES ALEATORIAS.....	25
III.4 TRANSFORMACIONES DE VARIABLES ALEATORIAS.....	28

<b>IV CARACTERÍSTICAS DE LA VARIABLES ALEATORIAS.....</b>	<b>29</b>
<b>V FUNCIÓN CARACTERÍSTICA Y FUNCIÓN GENERATRIZ DE MOMENTOS....</b>	<b>32</b>
<b>VI DISTRIBUCIONES NOTABLES.....</b>	<b>35</b>
VI.1 DISTRIBUCIONES DISCRETAS.....	35
VI.1.1 Bernoulli y Binomial .....	36
VI.1.2 Geométrica.....	37
VI.1.3 Poisson .....	38
VI.1.4 Hipergeométrica.....	39
VI.1.5 Distribución Binomial Negativa.....	39
VI.2 DISTRIBUCIONES CONTINUAS.....	40
VI.2.1 Uniforme.....	41
VI.2.2 Exponencial.....	41
VI.2.3 Normal.....	42
VI.2.4 Log-normal.....	50
VI.2.5 Distribución Gamma .....	50
VI.2.6 Distribución Beta .....	51
VI.2.7 Distribución Weibull .....	52
VI.2.8 Distribución de Cauchy.....	52
<b>VII VARIABLES ALEATORIAS MULTIDIMENSIONALES.....</b>	<b>53</b>
VII.1 DISTRIBUCIONES Y VARIABLES ALEATORIAS N-DIMENSIONALES .....	53
VII.1.1 Variables y funciones de distribución.....	53
VII.1.2 Distribuciones marginales y condicionadas.....	55
VII.1.3 Función característica de una v.a. n-dimensional.....	57
VII.1.4 Características de variables n-dimensionales .....	58

VII.2 INDEPENDENCIA DE VARIABLES ALEATORIAS .....	59
VII.3 TRANSFORMACIONES DE VARIABLES ALEATORIAS .....	62
VII.4 REGRESIÓN Y CORRELACIÓN .....	63
<i>VII.4.1 Línea general de regresión</i> .....	63
<i>VII.4.2 Recta de regresión</i> .....	64
<b>VIII DISTRIBUCIONES NOTABLES MULTIDIMENSIONALES .....</b>	<b>67</b>
VIII.1 MULTINOMIAL.....	67
VIII.2 NORMAL MULTIVARIANTE .....	68
VIII.3 DISTRIBUCIONES ASOCIADAS AL MUESTREO EN POBLACIONES NORMALES .....	69
<i>VIII.3.1 Distribución Chi-cuadrado (<math>\chi^2</math>)</i> .....	69
<i>VIII.3.2 Distribución t de Student</i> .....	71
<i>VIII.3.3 F de Fisher-Snedecor (<math>F_{m,n}</math>)</i> : .....	72
<b>IX SUCESIONES DE VARIABLES ALEATORIAS .....</b>	<b>73</b>
IX.1 CONVERGENCIAS DE SUCESIONES DE VARIABLES ALEATORIAS .....	73
<i>IX.1.1 Definiciones de convergencias de sucesiones</i> .....	73
IX.1.1.1 Convergencia casi seguro .....	73
IX.1.1.2 Convergencia en media de orden $p$ ( $p \geq 1$ ) .....	73
IX.1.1.3 Convergencia en probabilidad .....	74
IX.1.1.4 Convergencia en Ley o débil.....	74
<i>IX.1.2 Propiedades y relaciones entre convergencias</i> .....	74
IX.2 LEYES DE LOS GRANDES NÚMEROS.....	74
<i>IX.2.1 Leyes débiles de los grandes números</i> .....	75
<i>IX.2.2 Leyes fuertes de los grandes números</i> .....	76
IX.3 PROBLEMA CENTRAL DEL LÍMITE .....	76



# Introducción

Desde muy antiguo (ya los egipcios la usaban) se viene utilizando la probabilidad y la estadística. Lo más extendido históricamente fueron los censos (de población y mortalidad) y los juegos de azar (Laplace y Bayes) que dieron lugar al estudio de la probabilidad.

A mediados del siglo XX, en 1933, Kolmogorov da la definición axiomática de la probabilidad. Previamente, se habían realizado intentos para definirla, como el de Laplace, que dio una definición frecuentista, según la cual  $P(A)=f/n$ , es decir, la frecuencia relativa de un suceso, y que en el caso de que todos los resultados sean equiprobables pasaba a definir la probabilidad de un suceso como

$$P(A) = \text{casos favorables a } A / \text{casos posibles como resultado}$$

La probabilidad hoy en día se estudia como la que proporciona los modelos teóricos que después se utilizan en estadística. Es decir, estaría en un nivel bajo los datos, en uno superior los modelos de probabilidad, y como una escalera que los une la inferencia estadística, de modo que a la vista de los datos se busca el modelo probabilístico apropiado y se utiliza para dar respuestas al problema real.





# I Espacio Probabilizable

## I.1 Experimentos aleatorios y sucesos

El objeto de estudio de la probabilidad son fenómenos para los que no se puede predecir el resultado por diversas causas como el azar, la falta de información, etc. No es relevante la causa.

Definición. Un *fenómeno aleatorio* es un fenómeno “físico”, “económico”... con las siguientes propiedades:

1. Todos los resultados posibles se conocen
2. Cualquier realización del fenómeno no se puede predecir
3. La realización del fenómeno puede repetirse bajo circunstancias idénticas

Definición. Un *experimento aleatorio* es cualquier realización de un fenómeno aleatorio.

Por tanto un experimento aleatorio es un proceso cuyo resultado no se conoce con exactitud.

Ejemplos: juegos de azar (monedas, dados, bolas...), ruleta, tiempo en llegar a la facultad, vida de un fluorescente, instante de llegada de un tren...

Definición. *Espacio muestral* asociado a un experimento aleatorio,  $S$ , es el conjunto de todos los resultados posibles del experimento.

Es importante tener en cuenta que ha de ser conocido y si está mal definido se hará un modelo que no corresponde al experimento.

Tipos de espacios muestrales:

- *Espacio muestral discreto:* el número de posibles resultados es numerable (finito o infinito numerable)
- *Espacio muestral no numerable:* en general, asociado a  $\mathbb{R}^k$ .

Definiciones.

*Suceso*: afirmación sobre el resultado de un experimento aleatorio. Se identifica con un subconjunto de  $S$ , el de los resultados que hacen que se verifique el suceso.

*Suceso elemental*: se cumple si y sólo si se da un único resultado.

*Suceso seguro*: siempre ocurre, se identifica con  $S$

*Suceso imposible*: nunca ocurre, se identifica con  $\emptyset$

*Suceso contrario* de  $A$ : suceso que ocurre cuando no ocurre  $A$ . Se denota  $A^c$  o  $\overline{A}$  y está formado por todos los resultados posibles que no son de  $A$ .

Dos sucesos  $A$  y  $B$  son *incompatibles o mutuamente excluyentes*, cuando no pueden ocurrir simultáneamente.

Ejemplos:

- Un ejemplo puede ser el resultado de tirar un dado con seis caras. Su espacio muestral sería  $S = \{1, 2, 3, 4, 5, 6\}$ . Y un suceso puede ser  $A = \text{par} = \{2, 4, 6\}$ , que sería un suceso compuesto, mientras que  $B = \text{sacar un } 3 = \{3\}$ , sería un suceso elemental. Un suceso imposible sería  $C = \text{sacar más de un } 7 = \emptyset$
- No todos los experimentos aleatorios tienen resultados numéricos; por ejemplo, al tirar una moneda el espacio muestral sería  $S = \{\text{cara}, \text{cruz}\}$ .
- Ni el espacio muestral discreto, por ejemplo, el tiempo que se puede tardar en llegar a la facultad desde casa,  $S = [30', \infty]$ .

Definición.  $A \Rightarrow B$  o  $A \subseteq B$ ,  $A$  implica  $B$  o  $A$  contenido en  $B$ , si siempre que se da  $A$  ocurre  $B$ .

Definiciones. Operaciones con sucesos: Sean  $A$  y  $B$  dos sucesos asociados a un experimento de espacio muestral  $S$

- Se llama disyunción o unión de sucesos ( $A \cup B$ ) al suceso que ocurre cuando ocurre al menos uno de los dos sucesos. Es decir, está formado por los resultados que pertenecen por lo menos a uno de los dos sucesos

- Se llama conjunción o intersección de sucesos ( $A \cap B$ ) al suceso que ocurre cuando tanto uno como otro suceso ocurren. Es decir, está formado por los resultados que pertenecen a ambos sucesos
- Se llama diferencia de sucesos ( $A - B$ ) al suceso que ocurre cuando ocurre  $A$  pero no ocurre  $B$ . Es decir, está formado por los resultados que son de  $A$  pero no son de  $B$ :  

$$A - B = A \cap \bar{B} = A - A \cap B$$
- Extensión directa de unión e intersección a una familia finita de sucesos:  $\bigcup_{i=1}^n A_i, \bigcap_{i=1}^n A_i$

Propiedades de las operaciones con sucesos:

- Idempotente:  $A \cup A = A, A \cap A = A$
- Complementariedad:  $(A^c)^c = A$
- Absorción:  $A \cup B = B \Leftrightarrow A \cap B = A \Leftrightarrow A \subseteq B$
- Conmutativa:  $A \cup B = B \cup A, A \cap B = B \cap A$
- Asociativa:  $A \cup (B \cap C) = (A \cup B) \cap C, A \cap (B \cup C) = (A \cap B) \cup C$
- Distributivas:  

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C), A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$
- Leyes de De Morgan:  $\overline{A \cup B} = \bar{A} \cap \bar{B}$  y  $\overline{A \cap B} = \bar{A} \cup \bar{B}$

## I.2 Álgebras y $\sigma$ -álgebras

Definición.  $Q \subseteq \wp(S), Q \neq \emptyset$ , es un álgebra si y solo si

1.  $A \in Q \Rightarrow A^c \in Q$
2. Es cerrado respecto a la unión finita:  $A_1, \dots, A_n \in Q \Rightarrow \bigcup_{i=1}^n A_i \in Q$

Definición.  $\mathcal{Q} \subseteq \wp(S)$ ,  $\mathcal{Q} \neq \emptyset$ , es una  $\sigma$ -álgebra si y solo si

1.  $A \in \mathcal{Q} \Rightarrow A^c \in \mathcal{Q}$
2. Es cerrada respecto a la unión numerable:  $A_1, \dots, A_n, \dots \in \mathcal{Q} \Rightarrow \bigcup_{i=1}^{\infty} A_i \in \mathcal{Q}$

Nótese que toda  $\sigma$ -álgebra es un álgebra.

Definición:  $(S, \mathcal{Q})$  donde  $S$  es un espacio muestral y  $\mathcal{Q}$  un álgebra o una  $\sigma$ -álgebra  $\mathcal{Q} \subseteq \wp(S)$ , forman un *espacio probabilizable* sobre el que se define la probabilidad.

Propiedades de las  $\sigma$ -álgebras

1.  $\emptyset, S \in \mathcal{Q}$
2.  $A_1, \dots, A_n, \dots \in \mathcal{Q} \Rightarrow \bigcap_{i=1}^n A_i \in \mathcal{Q}$
3.  $A, B \in \mathcal{Q} \Rightarrow A - B \in \mathcal{Q}$
4. La intersección de una familia de  $\sigma$ -álgebras es  $\sigma$ -álgebra

Demostración:

□

Ejemplos de  $\sigma$ -álgebras:

1.  $\wp(S)$  (todos los posibles subconjuntos de  $S$ )
2.  $\{\emptyset, S\}$
3.  $\{\emptyset, S, A, A^c\}$ , se llama  $\sigma$ -álgebra engendrada o generada por  $A$
4.  $\sigma$ -álgebra engendrada por  $A$  y  $B$ :  
 $\{\emptyset, S, A, A^c, B, B^c, A \cup B, A \cup B^c, A^c \cup B, A^c \cup B^c, A \cap B, A^c \cap B, A \cap B^c, A^c \cap B^c\}$

Definición: Dada  $T \subseteq \wp(S)$ , se llama  $\sigma$ -álgebra engendrada o generada por  $T$  a la mínima  $\sigma$ -álgebra que contiene a  $T$ ,  $\sigma(T)$ .

Nota:  $\sigma(T)$  siempre existe pues la intersección de todas las  $\sigma$ -álgebras que contienen a  $T$ , y siempre hay al menos una,  $\wp(S)$ .

### I.3 $\sigma$ -álgebras más interesantes

En un espacio muestral discreto, hay dos que son las más interesantes:

- $\wp(S)$
- Dado  $A \subseteq \wp(S)$ ,  $\sigma(A)$ .

En el espacio muestral no numerable  $\mathbb{R}$ ,  $\wp(\mathbb{R})$  no es operativo pues contiene demasiados conjuntos que no son interesantes y luego da problemas para definir la probabilidad. La  $\sigma$ -álgebra que se utiliza es la  $\sigma$ -álgebra de Borel.

Definición.  $\mathcal{B}$ ,  $\sigma$ -álgebra de Borel, es la  $\sigma$ -álgebra engendrada por los intervalos, es decir, la mínima  $\sigma$ -álgebra que contiene a todos los intervalos.

$$\mathcal{B} = \sigma(T), \quad \text{donde } T = \{(a, b], (-\infty, b], (a, +\infty) / a \leq b \in \mathbb{R}\}$$

Por las propiedades de las  $\sigma$ -álgebras, contiene a todos los intervalos abiertos, cerrados, semiabiertos, y así contiene a todos los abiertos (unión de intervalos abiertos), todos los cerrados (contrarios de abiertos), los elementos unitarios (por ser cerrados) y todos los conjuntos numerables (unión numerable de unitarios).

## II Espacio de probabilidad

El concepto de probabilidad, además de sus numerosas aplicaciones formales a través de la teoría de la probabilidad, está presente en la vida cotidiana y en las conversaciones que mantenemos habitualmente. No es infrecuente que se escuchen frases como "probablemente lloverá esta tarde", "es probable que este fin de semana vayamos al cine", "probablemente la bolsa alcanzará máximos históricos esta semana", todas ellas basadas en el concepto de probabilidad, o de verosimilitud, de que algún suceso determinado ocurra.

Uno de los primeros intentos de definir el concepto de probabilidad surgió a partir del concepto de frecuencia relativa. Aunque esta definición hoy día está superada, la presentamos a continuación, ya que nos permite ir afinando la intuición sobre lo que puede significar el concepto de probabilidad, y porque las propiedades de la probabilidad irán paralelas a las de la frecuencia relativa.

El comportamiento de la frecuencia relativa de un suceso se observa de forma experimental aumentando progresivamente el número de repeticiones del experimento (siempre que sea posible) y anotando los valores de la frecuencia relativa correspondiente.

A medida que el número de repeticiones del experimento aumenta, puede comprobarse que la frecuencia relativa de un suceso tiende a estabilizarse alrededor de un valor fijo que se puede considerar como una medida de la verosimilitud de ese suceso. Esto constituye un hecho empírico que se conoce como ley de regularidad estadística y que se puede enunciar como sigue: "la frecuencia relativa de un suceso se estabiliza cuando el número de repeticiones del experimento se hace indefinidamente grande".

Esta ley permite dar una definición frecuentista de la probabilidad de un suceso, como el límite de la frecuencia relativa de este cuando el número de repeticiones tiende a infinito:

$$P(A) = \lim_{n \rightarrow \infty} f(A)$$

Esta definición presenta algunos inconvenientes graves, que hacen que hoy día esté en desuso. En primer lugar, no es congruente con el concepto de límite, ya que no es posible repetir un experimento indefinidamente. Además, relaciona la probabilidad, una medida a priori, previa a la realización del experimento, con la frecuencia relativa, que es una medida a posteriori,

obtenida después de realizar el experimento. Aun así, consigue dar una idea intuitiva de lo que se pretende al definir la probabilidad de un suceso.

## II.1 Función de probabilidad y propiedades

En 1933 el matemático ruso A. N. Kolmogorov introdujo la definición axiomática de la probabilidad, estableciendo unas propiedades (axiomas) que aceptamos sin demostración. A partir de estos axiomas podemos construir toda la teoría de la Probabilidad.

Definición: Dado un espacio probabilizable  $(S, Q)$ , una función  $P : Q \rightarrow [0, 1]$  tal que a cada suceso le haga corresponder un número real  $P(A)$ , es función de probabilidad  $\Leftrightarrow$

$$1) P(S)=1$$

2)  $\sigma$ -aditividad:  $A_1, \dots, A_n, \dots$  disjuntos dos a dos, tales que

$$\bigcup_{n=1}^{\infty} A_n \in Q \Rightarrow P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

Como ya hemos indicado, es fácil observar que las propiedades que definen la probabilidad son análogas a las de la frecuencia relativa; pero, mientras que ésta es una medida empírica de la ocurrencia de un suceso, la probabilidad es una medida teórica de la citada ocurrencia. De modo que la probabilidad pretende evaluar la posibilidad de que un suceso ocurra antes de que el experimento se lleve a cabo.

Definición:  $(S, Q, P)$  se llama *espacio probabilístico o espacio de probabilidad*.

Propiedades de la probabilidad:

$$1) P(\emptyset) = 0$$

$$2) \text{ Aditividad finita: } A_1, \dots, A_n \text{ disjuntos dos a dos } \Rightarrow P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$$

$$3) P \text{ es monótona: } A, B \in Q, \text{ si } A \subseteq B \Rightarrow P(A) \leq P(B) \text{ y } P(B - A) = P(B) - P(A)$$

$$4) P(A^c) = 1 - P(A)$$

5)  $A, B \in \mathcal{Q} \Rightarrow P(A \cup B) = P(A) + P(B) - P(A \cap B)$

6)  $A_1, \dots, A_n \in \mathcal{Q} \Rightarrow P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$

7)  $P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n+1} P\left(\bigcap_{i=1}^n A_i\right)$

8) Desigualdad de Bonferroni:  $P\left(\bigcap_{i=1}^n A_i\right) \geq 1 - \sum_{i=1}^n P(A_i^c)$

Demostración. Las propiedades son consecuencia de la definición de probabilidad, y cada una de ellas puede demostrarse a partir de los axiomas y de las propiedades que la preceden.  $\square$

Teorema. Sea  $(S, \mathcal{Q}, P)$  espacio de probabilidad, y sea  $\{A_n\} \subseteq \mathcal{Q}$

1. Si  $\{A_n\}$  es creciente ( $A_n \subseteq A_{n+1} \forall n$ )  $\Rightarrow \lim P(A_n) = P(\lim A_n) = P\left(\bigcup_{n=1}^{\infty} A_n\right)$

2. Si  $\{A_n\}$  es decreciente ( $A_n \supseteq A_{n+1} \forall n$ )  $\Rightarrow \lim P(A_n) = P(\lim A_n) = P\left(\bigcap_{n=1}^{\infty} A_n\right)$

Demostración.  $\square$

Teorema. Sea  $(S, \mathcal{Q}, P)$  espacio de probabilidad, y una sucesión de sucesos  $\{A_n\} \subseteq \mathcal{Q}$ , si  $\exists \lim A_n \Rightarrow \lim P(A_n) = P(\lim A_n)$

La definición de probabilidad, junto con las propiedades deducidas a partir de ella, nos permiten asignar probabilidades a distintos sucesos asociados a un experimento aleatorio a partir de las probabilidades de otros sucesos.

## II.2 Probabilidad en el espacio muestral discreto

Si el experimento tiene un número finito o numerable de resultados posibles,  $S$  es discreto, se suele utilizar como  $\sigma$ -álgebra  $\wp(S)$ , y para definir una probabilidad basta con asignar



probabilidades a cada uno de los resultados posibles (sucesos elementales), de manera que sean mayores o iguales que cero y la suma sea 1.

Es decir, basta definir  $P(\{a_i\}) = p_i \quad \forall a_i \in S$ , con  $p_i \geq 0 \quad \forall i$  y  $\sum_{i/a_i \in S} p_i = 1$

Y se obtiene la probabilidad de cualquier suceso como la suma de las probabilidades de los sucesos elementales que lo forman, es decir, para todo  $A \in Q$ , se define:

$$P(A) = \sum_{a_i \in A} P(\{a_i\}) = \sum_{a_i \in A} p_i$$

Es inmediato comprobar que es función de probabilidad.

Si además el experimento es equiprobable, es decir, todos los resultados tienen la misma probabilidad de salir, se tiene que si hay  $n$  resultados posibles, la probabilidad de cada uno de ellos es  $P(\{a_i\}) = 1/n$ . Así, si  $A$  es un suceso que contiene  $k$  resultados elementales, su

probabilidad se puede calcular como  $P(A) = \sum_{a_i \in A} \frac{1}{n} = \frac{\text{número resultados favorables}}{\text{número resultados posibles}}$ , que es

la conocida regla de Laplace.

En ocasiones se ha utilizado esta regla como definición de la probabilidad de un suceso, pero en gran número de casos la regla de Laplace no es aplicable, ya que puede que los resultados no sean equiprobables o bien que no haya un número finito de resultados.

El recuento de casos, especialmente de los casos favorables a cada suceso, puede ser muy laborioso, por lo que se suele recurrir a técnicas de recuento para obtenerlos

## II.3 Probabilidad en el espacio probabilidad $(\mathbb{R}, \mathcal{B})$

En  $(\mathbb{R}, \mathcal{B})$  no se define la probabilidad para cada suceso ya que no verificaría los axiomas de Kolmogorov, razón también por la que no se usa  $\wp(\mathbb{R})$  sino  $\mathcal{B}$ .

El proceso para definir la probabilidad es el siguiente:

1. Definir la probabilidad sobre la clase que genera  $\mathcal{B}$ , es decir, para los intervalos de la clase  $\mathcal{T} = \{(a, b], (-\infty, b], (a, +\infty) / a \leq b \in \mathbb{R}\}$ , de modo que sea  $\sigma$ -aditiva.

Recuérdese que  $B = \sigma(T)$ . Con ello se tiene la terna  $(\mathbb{R}, T, P)$  con  $P$   $\sigma$ -aditiva sobre  $T$ .

2. Sea  $F = \left\{ \bigcup_{i=1}^n A_i / A_i \in T \text{ disjuntos 2 a 2} \right\}$ .  $F$  es álgebra, y definimos una

probabilidad:  $P_1(A) = \sum_{i=1}^n P(a_i, b_i]$   $\forall A = \bigcup_{i=1}^n (a_i, b_i] \in F$ . Esta medida cumple que

$P_1(A) = P(A) \quad \forall A \in T$ . A  $P_1$  se le llama  $P$ , y se dice que  $P$  se ha extendido de  $T$  a  $F$ .

3.  $P$  se puede extender de  $F$  a  $B$  de forma única por el siguiente teorema de extensión.

Teorema de Extensión de Caratheodory. Sea  $(S, Q, P)$  espacio de probabilidad, siendo  $Q$  álgebra. Entonces  $P$  se puede extender de  $Q$  a  $\sigma(Q)$  de forma única, es decir, existe una única medida de probabilidad sobre  $(S, \sigma(Q))$  tal que coincide con  $P$  en  $Q$ .

## II.4 Probabilidad condicionada

En el capítulo anterior hemos calculado probabilidades de algunos sucesos, suponiendo que no contábamos con más información que la que proporciona la descripción del experimento aleatorio. Sin embargo, en muchas situaciones, aún sin conocer el resultado del experimento, sí que disponemos de información adicional, y dicha información debe incorporarse al cálculo de la probabilidad de los sucesos.

El conocimiento de que un determinado suceso ha ocurrido, o no, puede influir en el cálculo de las probabilidades de otros sucesos. Consideremos un experimento cuyo espacio muestral es  $S$  y supongamos que se conocen las probabilidades de todos los sucesos. Supongamos, además, que se sabe que un suceso  $B$  ha ocurrido. ¿Cómo afecta esta información a la probabilidad de que otro suceso  $A$  ocurra? Esta nueva probabilidad se conoce como probabilidad del suceso  $A$  dado que sabemos que el suceso  $B$  ha ocurrido, o, más abreviadamente, probabilidad de  $A$  condicionada por  $B$ .

Definición. Sea  $(S, Q, P)$  espacio de probabilidad, y sea  $A, B$  sucesos de  $Q$ , tal que  $P(A) > 0$ . La probabilidad de  $B$  condicionada por  $A$  se define como  $P(B/A) = \frac{P(A \cap B)}{P(A)}$ .

Tiene sentido llamarla probabilidad pues es una medida de probabilidad, donde el espacio muestral se reduce a  $A$ .

Definición. Sea  $(S, Q, P)$  espacio de probabilidad, y sea  $A$  suceso de  $Q$ , tal que  $P(A) > 0$ . Sea  $Q_A = \{B \subseteq A / \exists C \in Q \ B = C \cap A\}$  y definamos  $\forall B \in Q_A, B = C \cap A \ C \in Q$   
 $P_A(B) = P(C/A) = \frac{P(C \cap A)}{P(A)} = \frac{P(B)}{P(A)} \Rightarrow (A, Q_A, P_A)$  es un espacio de probabilidad.

Demostración. La probabilidad condicionada es, efectivamente, una probabilidad, puesto que verifica los tres axiomas de la definición de Kolmogorov.  $\square$

El concepto de probabilidad condicionada nos permite incorporar información parcial sobre el resultado de un experimento.

De la definición de probabilidad condicionada, se obtiene directamente:

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

Teorema del producto. Sea  $(S, Q, P)$  espacio de probabilidad,  $A_1, \dots, A_n \in Q$ . Si  $P(A_1 \cap \dots \cap A_{n-1}) > 0 \Rightarrow P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2/A_1)P(A_3/A_1 \cap A_2) \dots P(A_n/A_1 \cap \dots \cap A_{n-1})$

Demostración.  $\square$

Teorema de la probabilidad total. Sea  $(S, Q, P)$  espacio de probabilidad,  $A \in Q, \{B_i\}_{i \in I} \in Q$  disjuntos dos a dos y  $A \subseteq \bigcup_{i \in I} B_i \Rightarrow P(A) = \sum_{i \in I / P(B_i) > 0} P(A/B_i)P(B_i)$ .

Demostración.

$$P(A) = P\left(A \cap \bigcup_{i \in I} B_i\right) = P\left(\bigcup_{i \in I} (A \cap B_i)\right) \stackrel{\text{disjuntos}}{=} \sum_{i \in I} P(A \cap B_i) = \sum_{i \in I / P(B_i) > 0} P(A/B_i)P(B_i) \quad \square$$

Teorema de Bayes. Sea  $(S, Q, P)$  espacio de probabilidad,  $A, B \in Q$ ,  $P(A), P(B) > 0$ ,

$$\{B_i\}_{i \in I \subset \mathbb{N}} \in Q \text{ disjuntos dos a dos, } A \subseteq \bigcup_{i \in I} B_i \Rightarrow P(B/A) = \frac{P(A/B)P(B)}{\sum_{i \in I / P(B_i) > 0} P(A/B_i)P(B_i)}.$$

Demostración.  $P(B/A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A/B)P(B)}{P(A)} \stackrel{\text{Prob.Total}}{=} \frac{P(A/B)P(B)}{\sum_{i \in I / P(B_i) > 0} P(A/B_i)P(B_i)} \quad \square$

Este resultado, conocido como teorema de Bayes, proporciona una regla simple para calcular las probabilidades condicionadas. En muchos casos, se entiende como una probabilidad a posteriori.

Ejemplo:  $P(\text{fallo mecánico}) = 0.05$ ,  $P(\text{mal carretera}) = 0.2$ ,  $P(\text{fallo humano}) = 0.2$  No más causas

$P(\text{accid./fallo mecánico}) = 0.1$ ,  $P(\text{accid./mal carretera}) = 0.08$ ,  $P(\text{accid./fallo humano}) = 0.15$

$P(\text{accidente}) = 0.1 \times 0.05 + 0.08 \times 0.2 + 0.15 \times 0.2 = 0.051$  (Probabilidad total)

$P(\text{fallo mecánico/accid.}) = P(\text{accid./fallo mec})P(\text{fallo mec.})/P(\text{accid.}) = 0.1 \times 0.05 / 0.051 = 0.098$

## II.5 Independencia de sucesos

Sea  $(S, Q, P)$  espacio de probabilidad, y sean  $A, B \in Q$

Definición. A y B son independientes  $\Leftrightarrow P(A \cap B) = P(A)P(B)$ .

Teorema.  $(S, Q, P)$  espacio de probabilidad,  $A, B \in Q$ ,  $P(A) > 0$ . A y B son independientes  $\Leftrightarrow P(B/A) = P(B)$ .

Demostración:  $\Rightarrow P(B/A) = \frac{P(A \cap B)}{P(A)} = \frac{P(A)P(B)}{P(A)} = P(B)$

$\Leftrightarrow P(B/A) = P(B) \Rightarrow P(A \cap B) = P(B/A)P(A) = P(B)P(A) \quad \square$

La idea es que dos sucesos A y B, asociados a un experimento aleatorio, son independientes si el hecho de que ocurra uno de ellos no influye en la probabilidad de que ocurra el otro; es decir, si  $P(A/B) = P(A)$  o  $P(B/A) = P(B)$

En otro caso se dice que los sucesos son dependientes.

Si dos sucesos son independientes, también lo son sus contrarios.

Ejemplo: Sacar una carta de una baraja.  $A =$  sale rey,  $B =$  es de copas,  $P(A) = 4/40 = 0.1$ ,  $P(B) = 10/40 = 0.25$ ,  $P(\text{rey de copas}) = 1/40 = 0.025 = P(A)P(B)$ . Son independientes.

Si extraemos la carta de rey de copas de la baraja:  $P(\text{rey de copas}) = 0 \neq P(A)P(B)$ . Son dependientes.

Definición.  $(S, Q, P)$  espacio de probabilidad,  $\{A_i\}_{i \in I} \in Q$  son independientes  
 $\Leftrightarrow \forall \{i_1, \dots, i_k\} \subseteq I \quad P(A_{i_1} \cap \dots \cap A_{i_k}) = P(A_{i_1}) \dots P(A_{i_k})$ .

También es extensible el concepto a independencia de sucesos de distintos experimentos aleatorios, mediante la construcción del espacio probabilidad producto que no vamos a ver en este curso.

## II.6 Función de distribución en $(\mathbb{R}, B)$

Las medidas de probabilidad definidas sobre experimentos cuyo resultado es numérico es posible caracterizarlas mediante funciones de punto, y no de conjunto, que son más fáciles de manejar.

Definición. Sea  $(\mathbb{R}, B, P)$  espacio de probabilidad. La *función de distribución* asociada a la medida de probabilidad  $P$  se define como

$$F : (-\infty, +\infty) \rightarrow [0, 1] \\ x \rightarrow F(x) = P((-\infty, x])$$

Se puede extender definiendo  $F(-\infty) = \lim_{x \rightarrow -\infty} F(x)$  y  $F(\infty) = \lim_{x \rightarrow \infty} F(x)$ .

Propiedades: Dada  $F$  función de distribución de una medida de probabilidad  $P$

1.  $F$  es monótona no decreciente

Dem.:

$$\forall x \in \mathbb{R}, \forall h > 0 \quad F(x+h) - F(x) = P((-\infty, x+h]) - P((-\infty, x]) = P(x, x+h] \geq 0$$

2.  $F(-\infty) = 0, \quad F(+\infty) = 1$

Dem:  $F(+\infty) = \lim_{x \rightarrow +\infty} F(x) = \lim_{n \rightarrow \infty} F(n) = \lim_{n \rightarrow \infty} P(-\infty, n] \underset{\text{suc.crec}}{=} P(\lim_{n \rightarrow \infty} (-\infty, n]) = P(\mathbb{R}) = 1$

$F(-\infty) = \lim_{x \rightarrow -\infty} F(x) = \lim_{n \rightarrow \infty} F(-n) = \lim_{n \rightarrow \infty} P(-\infty, -n] \underset{\text{suc.decrec}}{=} P(\lim_{n \rightarrow \infty} (-\infty, -n]) = P(\emptyset) = 0$

3. F continua por la derecha

Dem:  $\forall x \in \mathbb{R}$

$$\begin{aligned} \lim_{h \downarrow 0} F(x+h) &= \lim_{n \rightarrow \infty} F(x + \frac{1}{n}) = \lim_{n \rightarrow \infty} P(-\infty, x + \frac{1}{n}] = \lim_{n \rightarrow \infty} \left( P(-\infty, x] + P(x, x + \frac{1}{n}] \right) = \\ &= P(-\infty, x] + \lim_{n \rightarrow \infty} P(x, x + \frac{1}{n}] = F(x) + P\left(\lim_{n \rightarrow \infty} (x, x + \frac{1}{n}] \right) = F(x) + P(\emptyset) = F(x) \end{aligned}$$

Se podría haber definido  $F(x) = P(-\infty, x)$ , y sería igual excepto continua por la izquierda.

Teorema. Una función  $F : \mathbb{R} \rightarrow \mathbb{R}$  es función de distribución de alguna medida de probabilidad  $P : \mathcal{B} \rightarrow [0, 1]$  si y solo si se verifican las siguientes propiedades:

1.  $\lim_{x \rightarrow -\infty} F(x) = 0$  y  $\lim_{x \rightarrow \infty} F(x) = 1$
2. F es monótona no decreciente ( $F(x) \leq F(y) \quad \forall x \leq y$ )
3. F continua por la derecha

Demostración. Esquema de la demostración

$\Rightarrow$ ) Si es función de distribución, cumple las 3 propiedades

$\Leftarrow$ ) Sea  $\mathcal{F} : \{ \text{uniones finitas de intervalos disjuntos } (a_i, b_i] \}$  que es álgebra y

$$P : \mathcal{F} \rightarrow \mathbb{R}$$

$$P\left(\bigcup_{i=1}^n (a_i, b_i]\right) = \sum_{i=1}^n (F(b_i) - F(a_i))$$

A1) Se puede comprobar que es aplicación (no depende de los intervalos que se escojan)

A2) Valor  $\geq 0$

A3) Axioma 1 de Kolmogorov evidente

A4) Se comprueba la  $\sigma$ -aditividad, comprobando la aditividad finita y la continuidad en  $\emptyset$  ( $\{A_n\} \downarrow \emptyset \Rightarrow \lim_{n \rightarrow \infty} P(A_n) = 0$ )

Por último se hace la extensión de esa álgebra a la  $\sigma$ -álgebra engendrada que es B.  $\square$

Por lo tanto, ya no es necesaria la referencia a la medida de probabilidad.

Definición. Una *función de distribución* es toda función  $F : \mathbb{R} \rightarrow \mathbb{R}$  tal que

1.  $\lim_{x \rightarrow -\infty} F(x) = 0$  y  $\lim_{x \rightarrow \infty} F(x) = 1$
2. F es monótona no decreciente ( $F(x) \leq F(y) \quad \forall x \leq y$ )
3. F continua por la derecha

Teorema. Toda medida de probabilidad definida en la  $\sigma$ -álgebra de Borel tiene una única función de distribución asociada, y toda función de distribución tiene una única medida de probabilidad asociada.

Demostración. Sencilla a partir de lo anterior.  $\square$

Por lo tanto F caracteriza a P, es un resumen completo.

Propiedades:

1.  $P(a,b]=F(b)-F(a)$
2.  $P(a,b)=F(b^-)-F(a)$
3.  $P[a,b]=F(b)-F(a^-)$
4.  $P[a,b)=F(b^-)-F(a^-)$
5.  $P(\{a\})=F(a)-F(a^-)$

Demostración.  $\square$

Teorema. El conjunto  $D(F)$  de puntos de discontinuidad de una función de distribución F es numerable.

Demostración.  $D(F) = \{x \in \mathbb{R} / F(x) - F(x^-) > 0\} = \{x \in \mathbb{R} / P(x) > 0\}$

Sean  $E_n = \{x \in \mathbb{R} / P(x) \geq 1/n\}$ , que tienen a lo sumo  $n$  puntos.

$D(F) = \bigcup_{n=1}^{\infty} E_n$  que es una unión numerable de conjuntos finitos, y por lo tanto numerable.  $\square$

## II.6.1 Clasificación de las funciones de distribución

### II.6.1.1 Funciones de distribución discretas

Definición. Una función de distribución es discreta  $\dot{=} P_F$  es tal que  $\exists \{a_1, \dots, a_n, \dots\} \subseteq \mathbb{R}$  numerable con  $P_F(a_i) > 0$  y  $\sum_{i=1}^{\infty} P_F(a_i) = 1$

Definición. La colección de números  $\{P_F(a_1), \dots, P_F(a_n), \dots\}$  con numerable con  $P_F(a_i) > 0$  y  $\sum_{i=1}^{\infty} P_F(a_i) = 1$  recibe el nombre de *función de masa de probabilidad*.

Obsérvese que  $F(x) = P(-\infty, x] = \sum_{a_i \leq x} P_F(a_i) \Rightarrow F$  es escalonada.

Además,

$$P_F(a_i) = P_F(-\infty, a_i] - P_F(-\infty, a_i) = F(a_i) - \lim_{x \rightarrow a_i^-} F(x) = F(a_i) - F(a_i^-) \equiv \text{salto en } a_i$$

### II.6.1.2 Funciones de distribución absolutamente continuas

Definición. Una función de distribución  $F$  es *absolutamente continua*  $\dot{=}$  existe una función no negativa e integrable (Lebesgue) tal que  $\forall x \in \mathbb{R} \quad F(x) = \int_{-\infty}^x f(t) dt$

Definición. Llamaremos *función de densidad* de probabilidad a toda función  $f : \mathbb{R} \rightarrow \mathbb{R}$  tal que  $f(x) \geq 0 \forall x \in \mathbb{R}$  y  $\int_{-\infty}^{\infty} f(x) dx = 1$ .



Proposición. Sea  $f : \mathbb{R} \rightarrow [0, \infty)$  una función integrable tal que  $\int_{-\infty}^{\infty} f(x)dx = 1 \Rightarrow F(x) = \int_{-\infty}^x f(t)dt$  es función de distribución absolutamente continua que tiene a  $f$  como función de densidad.

Demostración.  $\square$

Proposición. Sea  $F$  una función de distribución absolutamente continua, entonces

- a)  $F$  es continua
- b) Si  $f$  es continua en  $x$ , entonces  $F$  es derivable en  $x$  y  $F'(x) = f(x)$
- c)  $P_F(x) = 0 \quad \forall x \in \mathbb{R}$
- d)  $\forall$  intervalo  $(a,b), (a,b], [a,b), [a,b] \quad (<a,b>) \quad P_F(<a,b>) = \int_a^b f(x)dx$
- e)  $P_F(B) = \int_B f(t)dt \quad \forall B \in \mathcal{B}$

Demostración.  $\square$

Existen las funciones de distribución singulares que son continuas, pero no son absolutamente continuas, sin embargo no se verán en este curso, y en general, tienen un interés meramente matemático.

### II.6.1.3 Funciones de distribución mixtas

Proposición. Sean  $F_1, F_2$  funciones de distribución, y  $\lambda \in [0,1] \Rightarrow$  la mixtura  $\lambda F_1 + (1-\lambda)F_2$  es función de distribución.

Demostración. Trivial  $\square$

Definición. Una función de distribución es *mixta*  $\stackrel{\cdot}{=} \exists F_1$  función de distribución discreta,  $\exists F_2$  función de distribución absolutamente continua y  $\exists \lambda \in [0,1]$  tal que  $F = \lambda F_1 + (1-\lambda)F_2$ .

En general, se trabaja con estas funciones separando la parte discreta (probabilidad acumulada en los saltos, aunque no es función de masa pues no suma 1) de la parte continua (derivada de la función de distribución donde es derivable, que no será función de densidad porque no integra 1). De modo que si  $\{a_1, \dots, a_n, \dots\}$  son los puntos de discontinuidad de  $F$  y  $P_F(a_i) = F(a_i) - F(a_i^-) \equiv$  salto en  $a_i$ , la probabilidad de un conjunto puede obtenerse como

$$P(B) = \sum_{a_i \in B} P_F(a_i) + \int_B F'(t) dt .$$

## II.6.2 Convergencia en Ley o en Distribución

Definición. Sea  $\{F_n\}_{n \in \mathbb{N}}$  una sucesión de funciones de distribución en  $\mathbb{R}$ . Dicha sucesión se dice que *converge en Ley* a la función de distribución  $F$  si para cada punto de continuidad de  $F$  se verifica la convergencia puntual  $\lim_{n \rightarrow \infty} F_n(x) = F(x)$ .

Aunque sólo se define para los puntos de continuidad, por las propiedades de las funciones de distribución, ésta quedaría completamente caracterizada, como se muestra a continuación.

Teorema. Sea  $\{F_n\}_{n \in \mathbb{N}}$  una sucesión de funciones de distribución que converge en Ley a la función de distribución  $F$ . Si  $\{F_n\}_{n \in \mathbb{N}}$  converge en Ley a  $G$ , entonces  $F$  y  $G$  son la misma función de distribución, es decir,  $F(x) = G(x) \forall x \in \mathbb{R}$ .

*Demostración.* Ambas coinciden en los puntos de continuidad de ambas funciones, que es un conjunto denso en  $\mathbb{R}$ , y por las propiedades de función de distribución se obtiene la igualdad.

$$F_n(x) = \begin{cases} 0 & x < 1/n \\ 1 - 1/n & 1/n \leq x < 1 \\ 1 & x \geq 1 \end{cases} \quad \lim_{n \rightarrow \infty} F_n(x) = \begin{cases} 0 & x < 0 \\ 1 & x \geq 0 \end{cases} \quad \text{función de distribución,}$$

degenerada en 0.

# III Variables aleatorias

## III.1 Definición de variable aleatoria

Una *variable aleatoria*  $X$  es una función que asigna un valor a cada resultado del experimento. Así por ejemplo, si al resultado del lanzamiento de una moneda le asignamos el valor 1 si es cara y  $-1$  si es cruz, tendremos una variable aleatoria cuyos posibles valores o *soporte* es el conjunto  $\{-1, 1\}$ . Las variables aleatorias son de gran importancia pues trasladan la probabilidad de un espacio no numérico a uno numérico con todas las ventajas que ello conlleva, entre otras las caracterizaciones y propiedades que se verán a continuación.

Definición Sea  $(S, Q, P)$  espacio de probabilidad. Una aplicación  $X: (S, P) \rightarrow (\mathbb{R}, B)$  se dice *variable aleatoria unidimensional* si  $X^{-1}(B) \in Q \quad \forall B \in B$ .

A la medida de probabilidad  $P_X$  definida en  $P_X: B \rightarrow [0, 1]$  tal que  $P_X(B) = P(X^{-1}(B)) = P(s \in S / X(s) \in B)$  se le llama *probabilidad inducida por la variable aleatoria*  $X$ .

Habitualmente se nota a las variables aleatorias con letras mayúsculas y a los valores muestrales de dicha variable con letras minúsculas. Así por ejemplo se habla de la variable  $X$  y de las muestras  $x_i$ .

La idea de variable aleatoria no es más que un caso particular de aplicación medible:

Definición. Sean  $(S_1, Q_1)$  y  $(S_2, Q_2)$  y una aplicación  $X: S_1 \rightarrow S_2$ .  $X$  es una aplicación medible respecto a  $Q_1$  y  $Q_2 \Leftrightarrow \forall A \in Q_2 \quad X^{-1}(A) = \{s_1 \in S_1 / X(s_1) \in A\} \in Q_1$ .

Teorema. Sean  $(S_1, Q_1)$  y  $(S_2, Q_2)$ ,  $X: S_1 \rightarrow S_2$ , y  $T \subset \wp(S_2) / \sigma(T) = S_2$ .  $X$  es medible  $\Leftrightarrow \forall A \in T \quad X^{-1}(A) \in Q_1$ .

Corolario.  $X: (S, Q) \rightarrow (\mathbb{R}, B)$  es variable aleatoria  $\Leftrightarrow X^{-1}(-\infty, a] \in Q \quad \forall a \in \mathbb{R}$

Demotración. Inmediata por el teorema anterior ya que los intervalos engendran la  $\sigma$ -álgebra de Borel.  $\square$

Dado que la probabilidad inducida está definida sobre la recta real, tiene función de distribución y todas las propiedades vistas para la probabilidad en la recta real.

La *función de distribución* de una variable aleatoria, o también conocida como función de distribución acumulada,  $F(x)$ , se define como la función que asigna a cada punto la probabilidad de que  $X$  tome un valor igual o inferior a  $x$ .

Definición. Función de distribución de la variable alatoria  $X$

$$F_x : \mathbb{R} \rightarrow [0,1]$$

$$x \rightarrow F_x(x) = P_x((-\infty, x]) = P(X^{-1}(-\infty, x]) = P(s \in S / X(s) \leq x)$$

Las propiedades que caracterizan a la función de distribución, como ya se vio en el tema anterior, son las siguientes:

- Su valor está comprendido entre 0 y 1, es decir,  $0 \leq F(x) \leq 1$
- Es no decreciente
- En los límites, se cumple  $\lim_{x \rightarrow +\infty} F(x) = 1$  y  $\lim_{x \rightarrow -\infty} F(x) = 0$

Las variables aleatorias que pueden ser de nuestro interés, se clasifican según es la función de distribución en: discretas, absolutamente continuas (en general, las llamaremos continuas) y mixtas. Las que no corresponden a ninguno de estos grupos exceden al objetivo de esta revisión básica.

Se dice que una variable aleatoria es *discreta* cuando puede tomar una cantidad numerable<sup>1</sup> de valores. Estas variables son caracterizadas, además de por la función de distribución, por la *función de masa de probabilidad*, que es una función que asigna a los posibles valores su probabilidad y al resto 0. Los valores de esta función han de estar entre 0 y 1 ya que son probabilidades y la suma de todos ha de ser 1. La función de distribución  $F(x)$  de este tipo de

---

<sup>1</sup> Es decir, una cantidad finita o infinita numerable (los números naturales, enteros, racionales,...)

variables se calcula como la suma de la función de probabilidad para valores iguales o inferiores al valor  $x$ , y resulta por lo tanto una función escalonada:

$$F(x) = \sum_{x_i \leq x} P(x = x_i) ; -\infty < x < +\infty$$

Se dice que una variable aleatoria es *absolutamente continua* cuando existe una función, denominada *función de densidad*,  $f(x)$ , tal que su integral para un conjunto  $I$  determina la probabilidad de ocurrencia de dicho conjunto:  $P(x \in I) = \int_I f(x) dx$

Se ha de tener en cuenta que la integral de la función de densidad en todo el espacio ha de ser la unidad.  $\int_{-\infty}^{+\infty} f(x) dx = 1$ . En este tipo de variables la probabilidad de ocurrencia de un valor concreto es cero ya que la integral en un punto es nula. Un concepto físico que se asemeja a lo que representa la función de densidad es precisamente la densidad de un cuerpo, en este caso referido a probabilidades.

La función de distribución en este caso se calcula como la integral de la función de densidad en el intervalo  $[-\infty, x]$ , y de aquí que su derivada constituye la función de densidad asociada a la variable  $x$ . La función de distribución en este caso es continua en todos los puntos, y es derivable en todos excepto a lo sumo en una cantidad numerable de ellos.

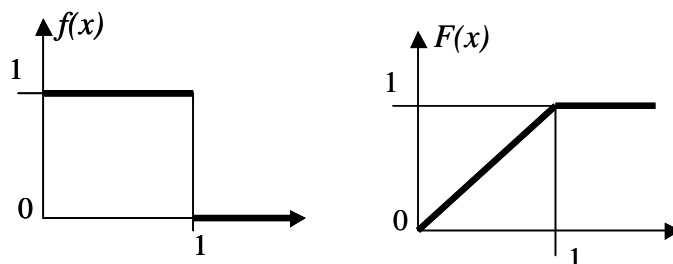
$$F(x) = P(x \in [-\infty, x]) = \int_{-\infty}^x f(y) dy ; -\infty \leq x \leq \infty$$

$$F'(x) = f(x)$$

Así por ejemplo para una variable aleatoria uniforme para el intervalo  $[0, 1]$  se tiene que:

$$f(x) = \begin{cases} 1 & 0 \leq x \leq 1 \\ 0 & \text{resto} \end{cases}$$

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



Las variables aleatorias *mixtas* son variables que en algunos intervalos se comportan como variables absolutamente continuas, pero que además concentran probabilidad en algún punto. Para ellas no existe otra caracterización teórica válida que no sea la función de distribución, que es continua en todos los puntos excepto en los que acumulan probabilidad que tiene un salto.

### III.2 Operaciones con variables aleatorias

Teorema.  $X : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$  continua, entonces  $X$  es variable aleatoria.

Demostración. Basta probarlo para los abiertos. Sea  $A$  un abierto de la recta real, entonces  $X^{-1}(A) \in \mathcal{B}$  ya que  $\mathcal{B}$  está generada por los abiertos, y  $X^{-1}(A)$  es abierto por ser  $X$  continua.  $\square$

Teorema. Sean  $X : (S, \mathcal{Q}) \rightarrow (\mathbb{R}, \mathcal{B})$  y  $f : (\mathbb{R}, \mathcal{B}) \rightarrow (\mathbb{R}, \mathcal{B})$  variables aleatorias, entonces  $f(X)$  es variable aleatoria.

Demostración.

$$A \in \mathcal{B} \Rightarrow f^{-1}(A) \in \mathcal{B} \Rightarrow X^{-1}(f^{-1}(A)) = (f \circ X)^{-1}(A) = (f(X))^{-1}(A) \in \mathcal{B} \square$$

Corolario. Sea  $X$  variable aleatoria. Entonces  $X^2, \text{sen}(X), -X, \dots$  son variables aleatorias.

Demostración. Toda transformación continua de una variable aleatoria será variable aleatoria.

$\square$

Teorema. Sean  $X, Y : (S, \mathcal{Q}) \rightarrow (\mathbb{R}, \mathcal{B})$  variables aleatorias. Entonces

- a)  $(X+Y)(s)=X(s)+Y(s)$  es variable aleatoria
- b)  $(X-Y)(s)=X(s)-Y(s)$  es variable aleatoria
- c)  $(XY)(s)=X(s)Y(s)$  es variable aleatoria

Demostración.  $\square$

Teorema. Sean  $X_1, \dots, X_n, \dots: (S, \mathcal{Q}) \rightarrow (\mathbb{R}, \mathcal{B})$  variables aleatorias. Entonces las siguientes funciones son variables aleatorias:

$$\max(X_1, \dots, X_n)(s) = \max(X_1(s), \dots, X_n(s)), \min(X_1, \dots, X_n)(s) = \min(X_1(s), \dots, X_n(s)), \\ \sup(X_n), \inf(X_n), \lim(X_n) \text{ (si existe)}$$

Demostración.  $\square$

### III.3 Integración de variables aleatorias

La definición formal de integral de una variable aleatoria no se va a incluir puesto que por una parte faltan conocimientos de la integral de Lebesgue, y porque a efectos más prácticos, casi carece de interés.

Basta decir que la integral de una variable aleatoria se define para las variables simples, de ahí para las variables no negativas, y por último para las variables arbitrarias, mediante la siguiente definición.

Definición. Sea  $X: (S, \mathcal{Q}) \rightarrow (\mathbb{R}, \mathcal{B})$  variable aleatoria. Se define las variable aleatorias *parte positiva de X* y *parte negativa de X* como  $X^+(s) = \max\{X(s), 0\}$ ,  $X^-(s) = -\min\{X(s), 0\}$ . Ambas son variables aleatorias no negativas y  $X = X^+ - X^-$ ,  $|X| = X^+ + X^-$

Definición. Sea  $X: (S, \mathcal{Q}) \rightarrow (\mathbb{R}, \mathcal{B})$  variable aleatoria. Se define  $\int_s X dP = \int_s X^+ dP - \int_s X^- dP$ , supuesto que alguna es finita. Además si  $\left| \int_s X dP \right| < \infty$  se dice que es *integrable*.

Proposición.  $X$  es integrable  $\Leftrightarrow |X|$  es integrable.

Demostración. Inmediata, pues  $|X| = X^+ + X^-$ .  $\square$

Definición. Una propiedad se cumple *casi seguro* (c.s.) en un espacio de probabilidad  $(S, Q, P)$  si y solo si el conjunto de puntos en que no se cumple la propiedad está incluido en un suceso que tiene probabilidad nula.

Propiedades de la integral. Sea  $A$  un suceso de  $Q$

a) Si  $P(A)=0 \Rightarrow \int_A X dP = 0$

b)  $A = \bigcup_{n=1}^{\infty} A_n$  disjuntos 2 a 2  $\Rightarrow \int_A X dP = \sum_{n=1}^{\infty} \int_{A_n} X dP$

c) Si  $X=Y$  casi seguro en  $A \Rightarrow \int_A X dP = \int_A Y dP$

d)  $\int_A (aX + bY) dP = a \int_A X dP + b \int_A Y dP$

e)  $\left| \int_A X dP \right| \leq \int_A |X| dP$

f) Si  $X \leq Y$  casi seguro en  $A \Rightarrow \int_A X dP \leq \int_A Y dP$

g)  $m \leq X \leq M$  c.s. en  $A \Rightarrow mP(A) \leq \int_A X dP \leq M \cdot P(A)$

h)  $\int_E X dP \leq \int_E Y dP \quad \forall E \subset A \Rightarrow X \leq Y$  c.s. en  $A$

i) Teorema de la convergencia dominada:  $X_1, \dots, X_n, \dots$  variables aleatorias

integrables tales que  $X_n \xrightarrow[n \rightarrow \infty]{c.s.} X$ , y tal que existe  $Y \geq 0$  variable aleatoria

integrable con  $|X| \leq Y \Rightarrow \exists \int_A X dP = \lim_{n \rightarrow \infty} \int_A X_n dP$

j)  $\{X_n\}_{n \in \mathbb{N}} \quad \sum_{n=1}^{\infty} \int_A X_n dP = \int_A \sum_{n=1}^{\infty} X_n dP$



$$k) \text{ Si } \exists \lim X_n \Rightarrow \int_A \lim X_n dP = \lim \int_A X_n dP$$

Estas propiedades están bien, y servirán para saber calcular de forma efectiva las integrales, sin embargo, para ello hay que trasladar el cálculo al espacio inducido por la variable aleatoria. Esto se realiza gracias al teorema de cambio de espacio de integración siguiente.

Teorema del Cambio de Espacio de Integración. Sea  $X : (S1, Q1, P) \rightarrow (S2, Q2)$  una función medible y sea  $P_X$  la medida de probabilidad inducida por X sobre  $(S2, Q2)$ . Sea  $Y : (S2, Q2, P_X) \rightarrow (\mathbb{R}, B)$  variable aleatoria  $\Rightarrow \int_{S1} Y(X) dP = \int_{S2} Y dP_X$ .

El caso más interesante para nosotros:

X variable aleatoria,  $g : (\mathbb{R}, B) \rightarrow (\mathbb{R}, B)$  entonces

$$\int_S (g \circ X) dP = \int_S g(X) dP = \int_{\mathbb{R}} g(x) dP_X = \int_{\mathbb{R}} g(x) dF_X(x)$$

Casos particulares:

- a) F discreta: La probabilidad se concentra en el *soporte de X*,  $D_X = \{a_1, \dots, a_n, \dots\}$  con probabilidades  $\{p(a_1), \dots, p(a_n), \dots\} \Rightarrow$

$$\begin{aligned} \int_{\mathbb{R}} g(x) dF_X(x) &= \int_{D_X} g(x) dF_X(x) + \int_{D_X^c} g(x) dF_X(x) = \int_{D_X} g(x) dF_X(x) \stackrel{D_X = \bigcup_n \{a_n\}}{=} \\ &= \sum_{n=1}^{\infty} \int_{\{a_n\}} g(x) dF_X(x) = \sum_{n=1}^{\infty} \int_{\mathbb{R}} g(x) I_{a_n}(x) dF_X(x) = \sum_{n=1}^{\infty} g(a_n) P(a_n) \end{aligned}$$

- b) F absolutamente continua: f función de densidad de modo que

$$P_X(B) = \int_B f(x) dx \quad \forall B \in B. \text{ Entonces } \int_{\mathbb{R}} g(x) dF_X(x) = \int_{\mathbb{R}} g(x) f_X(x)$$

- c) F mixta: en este caso se tiene que  $\sum_{a/P_X(a)>0} P_X(a) + \int_{\{x/\exists F'(x)\}} F'(x) dx = 1$ . Entonces

$$\begin{aligned} \int_{\mathbb{R}} g(x) dF_X(x) &= \int_{\{a/P_X(a)>0\}} g(x) dF_X(x) + \int_{\{x/\exists F'(x)\}} g(x) dF_X(x) = \\ &= \sum_{a/P_X(a)>0} g(a) P_X(a) + \int_{\mathbb{R}} g(x) F'(x) dx \end{aligned}$$

### III.4 Transformaciones de variables aleatorias

Teorema. Sea  $X$  variable aleatoria discreta con función de masa  $P_X(x)$  en el soporte  $D$  y sea  $g : \mathbb{R} \rightarrow \mathbb{R}$  medible Borel. Entonces la variable aleatoria  $Y=g(X)$  es discreta con función de masa

$$P(Y = y) = \sum_{x \in D/g(x)=y} P_X(x) \quad \forall y \in g(D)$$

Teorema. Sea  $X$  variable aleatoria absolutamente continua con función de densidad  $f(x)$  dentro del soporte  $C = \{x / f(x) > 0\}$  y sea  $g : \mathbb{R} \rightarrow \mathbb{R}$  tal que  $g(C)$  es discreto. Entonces la variable transformada  $Y=g(X)$  es discreta con función de masa

$$P(Y = y) = \int_{\{x \in C/g(x)=y\}} f(x) dx \quad \forall y \in g(D)$$

Teorema. Sea  $X$  variable aleatoria absolutamente continua con función de densidad  $f(x)$  dentro de  $C = \{x / f(x) > 0\}$ . Si  $C$  es un intervalo o unión de ellos, y  $g : \mathbb{R} \rightarrow \mathbb{R}$  una función medible estrictamente creciente o decreciente sobre  $C$  y tal que su inversa admite en  $g(C)$  derivada continua, entonces la variable aleatoria  $Y=g(X)$  es absolutamente continua con función de densidad  $h(y) = f(g^{-1}(y)) |(g^{-1})'(y)| \quad \forall y \in g(C)$ .

## IV Características de la variables aleatorias

La *esperanza* o *valor esperado* o *media* de una variable  $X$ , denotada por  $\mu$  o  $E(X)$ , es una medida central de comportamiento de la variable que representa su centro de masas.

Definición. Sea  $X : (S, \mathcal{Q}, P) \rightarrow (\mathbb{R}, \mathbb{B})$  una variable aleatoria. Se define la *esperanza* o *media* de  $X$  como su integral:  $\mu = E[X] = \int_S X dP = \int_{\mathbb{R}} x dF_X(x)$

Como tal integral, se calcula de manera distinta dependiendo de si se trata de una variable discreta o continua.

$$E[X] = \mu = \begin{cases} \sum_{x_i} x_i P(x_i) & \text{siendo } X_i \text{ variable discreta} \\ \int_{\mathbb{R}} x f(x) dx & \text{siendo } X_i \text{ variable continua} \end{cases}$$

La esperanza por su definición es un operador lineal, es decir,  $E(\sum_{i=1}^n c_i X_i) = \sum_{i=1}^n c_i E(X_i)$ .

Por otra parte, obsérvese que la esperanza de una variable puede ser un valor que no pertenezca al soporte, especialmente en variables aleatorias discretas, aunque siempre se encontrará entre el mínimo y el máximo de los posibles valores, si es que existe, ya que puede no existir.

La esperanza es un caso particular de momento respecto al origen.

Definición. El *momento de orden  $n$  de  $X$  respecto de  $a$*  se define como  $\alpha_{na} = E[(X - a)^n] = \int_{\mathbb{R}} (x - a)^n dF_X(x)$

Definición. El *momento de orden  $n$  respecto al origen* se define como  $\alpha_n = \alpha_{n0} = E[X^n]$

Definición. El *momento de orden  $n$  centrado o respecto a la media* se define como  $\mu_n = \alpha_{n\mu} = E[(X - \mu)^n]$

Todas estas esperanzas se calculan como integrales que son de forma distinta según la variable sea discreta, absolutamente continua o mixta.

Así como en los momentos respecto al origen destaca el de orden 1 que es la esperanza, en los momentos centrados o respecto a la media destaca el de orden 2: la *varianza*, definida como

$$V(X) = \mu_2 = \sigma^2 = \int_{\mathbb{R}} (x - \mu)^2 dF_X(x)$$

La *varianza* de una variable aleatoria es una medida para determinar el nivel de dispersión que tienen los valores de la variable respecto de su media. Por definición, la varianza es siempre no negativa y es un operador cuadrático, de modo que  $\text{var}(cX) = c^2 \text{var}(X)$

Obsérvese que la varianza va en unidades al cuadrado; para dar una medida relativa a las unidades que se están manejando se utiliza la *desviación típica o estándar*, que se define como la raíz cuadrada de la varianza de la variable y se denota por  $\sigma$ .

Una alternativa para el cálculo de la varianza, muy utilizada es la siguiente.

Corolario.  $\sigma^2 = V(X) = E[(X - \mu)^2] = E(X^2) - \mu^2$

Demostración.

$$\begin{aligned} \sigma^2 = V(X) &= E[(X - \mu)^2] = E[X^2 - 2\mu X + \mu^2] = E[X^2] - 2\mu E[X] + \mu^2 = \\ &= E[X^2] - 2\mu^2 + \mu^2 = E(X^2) - \mu^2 \end{aligned}$$

La media y la varianza de una variable aleatoria uniforme en el intervalo  $[0, 1]$ :

$$\mu = \int_0^1 x \cdot 1 dx = \left[ \frac{x^2}{2} \right]_0^1 = \frac{1}{2}$$

$$\sigma^2 = \int_0^1 x^2 \cdot 1 dx - \left( \frac{1}{2} \right)^2 = \left[ \frac{x^3}{3} \right]_0^1 - \frac{1}{4} = \frac{1}{12}$$

Proposición.  $\mu_n = \sum_{k=0}^n (-1)^{n-k} \binom{n}{k} \mu^{n-k} \alpha_k$

Proposición.  $\alpha_n = \sum_{k=0}^n \binom{n}{k} \mu^{n-k} \mu_k$

Teorema. Si existe  $\alpha_n = E[X^n]$  también existe  $\alpha_k \quad \forall k < n$

Teorema de Markov.

Sea  $X : (S, Q, P) \rightarrow (\mathbb{R}, B)$  variable aleatoria,  $g : (\mathbb{R}, B, P_X) \rightarrow (\mathbb{R}, B)$  variable aleatoria no negativa  $\Rightarrow P\{s \in S / g(X(s)) \geq k\} \leq \frac{E[g(X)]}{k}$

$$\begin{aligned} \text{Demostración. } E(g(X)) &= \int_S g(X) dP = \int_{\{s/g(X(s)) \geq k\}} g(X) dP + \int_{\{s/g(X(s)) < k\}} g(X) dP \geq \\ &\geq \int_{\{s/g(X(s)) \geq k\}} g(X) dP \geq kP\{s / g(X(s)) \geq k\} \end{aligned}$$

Teorema: Desigualdad de Tchebychev

$X$  variable aleatoria con  $\mu = E(X)$  y  $\sigma = \sqrt{V(X)} \Rightarrow$

$$P\{s \in S / |X(s) - \mu| \geq k\sigma\} \leq \frac{1}{k^2} \quad \forall k > 0$$

Demostración. Basta aplicar el teorema de Markov a  $g(x) = (x - \mu)^2$ :

$$P\{s \in S / |X(s) - \mu| \geq k\sigma\} = P\{s \in S / |X(s) - \mu|^2 \geq k^2\sigma^2\} \leq \frac{E[|X(s) - \mu|^2]}{k^2\sigma^2} = \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

Otra medida central de una variable aleatoria es la *mediana* de una función de distribución  $F$ : el valor o valores tales que  $F(x) \geq 1/2$  y  $F(x^-) \leq 1/2$ . Se nota por  $x_{0.5}$  y se calcula como el menor valor de la variable para el cual  $F_x(x) \geq 0.5$ . Cuando la variable tiene asimetría es habitual dar este valor de medida central, ya que la media se suele ver muy afectada por los valores más extremos.

Se define el *cuantil de orden  $p$*  de una función de distribución  $F$  como el valor o valores tales que  $F(x) \geq p$  y  $F(x^-) \leq p$ . Son relevantes los cuarteles, los deciles, los percentiles...

## V Función característica y función generatriz de momentos

Son funciones útiles para representar la distribución y sacar fácilmente los momentos. La ventaja de la función característica es que siempre existe, pero tiene como inconveniente que es una función compleja.

Definición. Sea  $F$  una función de distribución. Se define la *función característica de  $F$*  como  $\varphi: \mathbb{R} \rightarrow \mathbb{C}$  tal que  $\varphi(t) = \int_{\mathbb{R}} e^{itx} dF(x) \quad \forall t \in \mathbb{R}$ .

Recuérdese que  $e^{itx} = \cos(tx) + i\text{sen}(tx)$ , y puesto que ambas tienen valor absoluto menor que 1, esa integral existe  $\forall t \in \mathbb{R}$ . Recuérdese también que  $|e^{itx}| = 1$ .

Las propiedades de esa integral son análogas a las de la integral vista, y su cálculo es igual.

Definición. Sea  $X: (S, \mathcal{Q}, P) \rightarrow (\mathbb{R}, \mathcal{B}, P_X)$  variable aleatoria. La *función característica de  $X$*  se define como  $\varphi(t) = E[e^{itX}] = \int_S e^{itX} dP = \int_{\mathbb{R}} e^{itx} dF(x)$

Propiedades.

1)  $\varphi(0) = 1$

2)  $|\varphi(t)| \leq 1$

3)  $\varphi(-t) = \overline{\varphi(t)}$

4)  $\varphi$  es uniformemente continua

Demostración.  $\square$

Si una función es función característica tiene que cumplir estas propiedades, pero lamentablemente, si una función cumple estas propiedades, puede no ser función característica. Son propiedades necesarias pero no suficientes. Por lo tanto, si dada una función no cumple alguna de ellas, se puede asegurar que no es función característica. Pero si las cumple todas, a

pesar de ello, para saber si es función característica hay que obtener la distribución de la que proviene.

Teorema de inversión. Sea  $\varphi(t)$  la función característica correspondiente a la función de distribución F. Si a y b son dos puntos de continuidad de F entonces

$$F(b) - F(a) = \frac{1}{2\pi} \lim_{T \rightarrow \infty} \int_{-T}^T \frac{e^{itb} - e^{ita}}{it} \varphi(t) dt$$

Teorema. Si  $\varphi(t)$  es función característica de la función de distribución F, entonces no puede serlo de ninguna otra función de distribución.

Demostración.  $\square$

Teorema. Si  $\varphi(t)$  es integrable Lebesgue en  $\mathbb{R}$  ( $\int_{-\infty}^{\infty} |\varphi(t)| dt < \infty$ ) entonces  $\varphi(t)$  es la función característica de una variable aleatoria absolutamente continua con función de densidad

$$f(x) = F'(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{itx} \varphi(t) dt$$

Demostración.  $\square$

Ahora veamos la relación de la función característica y de los momentos, y otras propiedades.

Teorema. Sea X variable aleatoria con función característica  $\varphi(t)$ . Entonces la función característica de  $Y=aX+b$  es  $\varphi_Y(t) = e^{itb} \varphi_X(at)$

Demostración.

$$\varphi_Y(t) = E[e^{itY}] = \int_S e^{itY} dP = \int_S e^{it(aX+b)} dP = \int_S e^{itb} e^{itaX} dP = e^{itb} \int_S e^{i(at)X} dP = e^{itb} \varphi_X(at) \square$$

Teorema. Si existe  $\alpha_n = \int_{\mathbb{R}} x^n dF(x)$  finito, entonces

a)  $\exists \varphi^{(n)}(0)$  y  $\alpha_n = \frac{\varphi^{(n)}(0)}{i^n}$

b)  $\exists \varphi^{(n)}(t) = i^n \int_{\mathbb{R}} e^{itx} x^n dF(x)$

Demostración. Por inducción...□

Corolario. Si  $\exists \alpha_n = E[X^n] \quad \forall n \in \mathbb{N} \Rightarrow \varphi_X(t) = \sum_{n=0}^{\infty} E[X^n] \frac{(it)^n}{n!}$

Teorema. Sea  $\{F_n\}_{n \in \mathbb{N}}$  una sucesión de funciones de distribución y sea  $\{\varphi_n\}_{n \in \mathbb{N}}$  la sucesión de funciones características asociadas. Entonces

- a) Si la sucesión  $F_n \xrightarrow{Ley} F \Rightarrow \varphi_n(t) \rightarrow \varphi(t) \quad \forall t$ , siendo  $\varphi$  la función característica asociada a F
- b) Si  $\varphi_n(t) \rightarrow \varphi(t) \quad \forall t$ , y  $\varphi$  es continua en  $t=0$ , entonces  $\varphi$  es función característica y  $F_n \xrightarrow{Ley} F$  función de distribución asociada a  $\varphi$

Definición. La *función generatriz de momentos* se define como

$$M : \mathbb{R} \rightarrow \mathbb{R}$$
$$t \rightarrow M(t) = \int_{\mathbb{R}} e^{tx} dF(x)$$

supuesto que la integral existe y es finita en un entorno del 0.

Es una función real, pero tiene el inconveniente de que puede no existir ya que ahora no está acotada. Si existe, es única y caracteriza a la distribución.

Si  $M(t)$  existe para todo  $(-t_0, t_0)$  con  $t_0 > 0$  entonces existen las derivadas de todo orden de  $M(t)$  en  $t=0$  y además  $\alpha_k = M^{(k)}(0)$



## VI Distribuciones notables

### VI.1 Distribuciones discretas

Distribución	F. Masa	Media	Varianza	F. Característica
Degenerada (h)	$P(X=h)=1$	$h$	$0$	$e^{ith}$
Bernoulli (p)	$P(X=0)=1-p=q$ $P(X=1)=p$	$p$	$pq$	$q+pe^{it}$
Binomial(n,p)	$P(X=x)=\binom{n}{x}p^x(1-p)^{n-x} =$ $=\binom{n}{x}p^xq^{n-x} \quad x=0,\dots,n$	$np$	$npq$	$(q+pe^{it})^n$
Geométrica(p)	$P(X=x)=(1-p)^{x-1}p=q^{x-1}p$ $x=1,\dots$	$1/p$	$q/p^2$	$\frac{pe^{it}}{1-qe^{it}}$
Poisson( $\lambda$ )	$P(X=x)=\frac{e^{-\lambda}\lambda^x}{x!} \quad x=0,1,\dots$	$\lambda$	$\lambda$	$e^{\lambda(e^{it}-1)}$
Hipergeométrica (N,D,n)	$P(X=x)=\frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}$ $\max\{0, n+D-N\} \leq x \leq \min\{D, n\}$	$\frac{D}{N}n$	$\frac{D(N-D)n(N-n)}{N^2(N-1)}$	
Binomial Negativa(n,p)	$P(X=x)=\binom{n+x-1}{x}p^nq^x$ $x=0,1,\dots$	$nq/p$	$nq/p^2$	$\left(\frac{p}{1-qe^{it}}\right)^n$

## VI.1.1 Bernoulli y Binomial

Consideremos un experimento aleatorio en el que solo dos sucesos merecen nuestra atención: el suceso  $A$  y su contrario,  $\bar{A}$ . Si  $A$  ocurre, suele decirse que se ha obtenido un éxito, mientras que si no ocurre diremos que hubo fracaso (aunque bien podrían invertirse los términos). Supongamos que  $A$  ocurre con probabilidad  $p$  y, por tanto, que  $\bar{A}$  ocurre con probabilidad  $q=1-p$ .

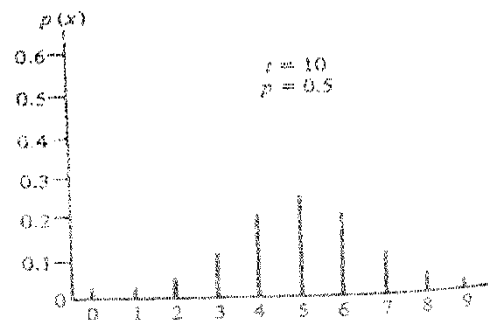
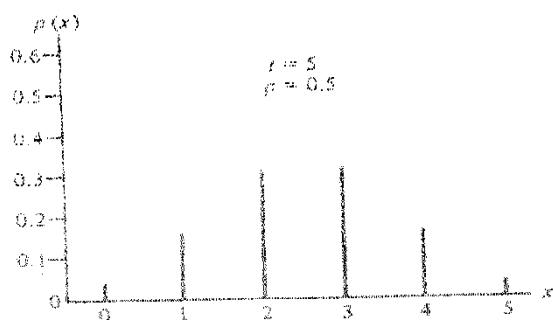
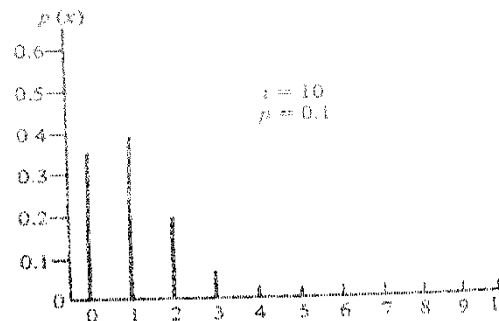
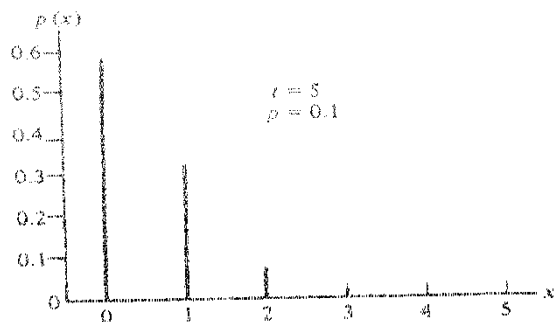
Un experimento de este tipo, en el que solo interesa conocer si un determinado suceso ha ocurrido o no, se denomina ensayo de Bernoulli. La variable aleatoria que asigna un 1 al éxito y un 0 al fracaso, se dice que tiene una distribución de Bernoulli. Es decir, su función de masa es tal que  $P(X=1) = p$  y  $P(X=0) = 1-p=q$ .

Para esta distribución  $E[X] = p$  y  $V[X] = p(1-p) = pq$

Supongamos que, en las mismas circunstancias descritas antes, el experimento se repite  $n$  veces de manera independiente; es decir, en las mismas condiciones iniciales. Supongamos que se define la variable aleatoria  $Y =$  número de éxitos obtenidos. La distribución de la variable aleatoria  $Y$  se conoce como distribución binomial de parámetros  $n$  y  $p$  y su función de masa es

$$P(Y = k) = \binom{n}{k} p^k (1-p)^{n-k} = \binom{n}{k} p^k q^{n-k} \quad k = 0, \dots, n$$

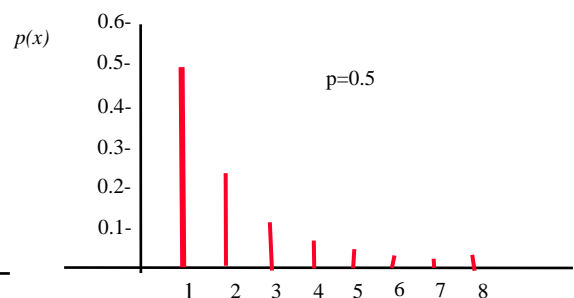
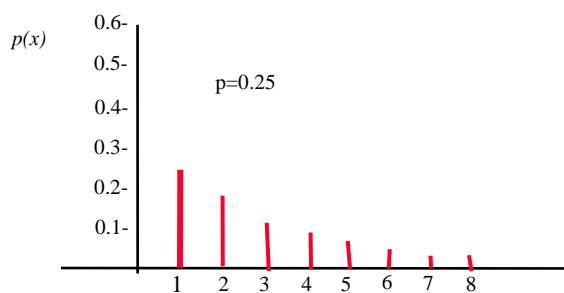
Su esperanza y varianza, respectivamente, son  $E[Y] = np$   $V[Y] = npq$



## VI.1.2 Geométrica

Supongamos que los ensayos de Bernoulli considerados anteriormente se repiten hasta que sale el primer éxito. Se llama distribución Geométrica a la de una variable  $X$  que expresa el instante en que ocurre el primer éxito. Sólo depende del parámetro  $p$  y su función de masa es

$$P(X = x) = (1 - p)^{x-1} p = q^{x-1} p \quad x = 1, \dots$$



La esperanza y varianza, respectivamente, son  $E[X] = 1/p$   $V[X] = q/p^2$

## VI.1.3 Poisson

Una variable aleatoria  $X$  con distribución de Poisson expresa el número de “sucesos” que ocurren en una porción fija de tiempo o espacio. La intensidad o número esperado de veces que ocurren esos sucesos por unidad de tiempo o espacio se representa mediante el parámetro positivo  $\lambda$ . Se usa, por ejemplo, para contar el número de llamadas a una centralita en un determinado periodo, el número de erratas en una página de un libro... Su función de masa es

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

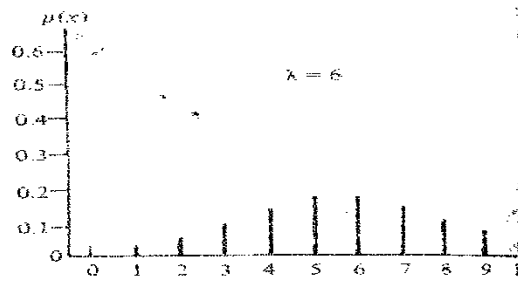
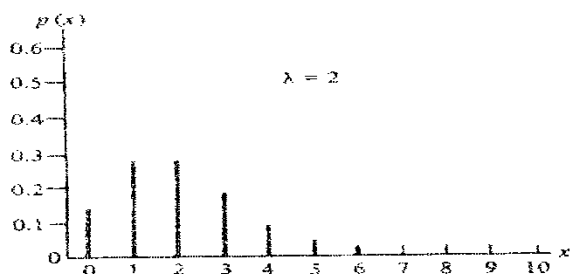
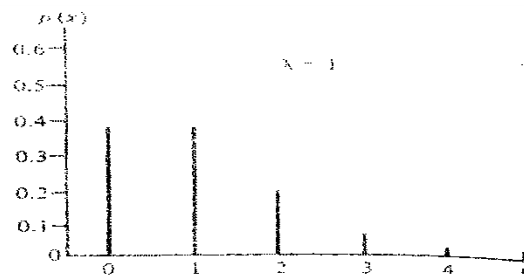
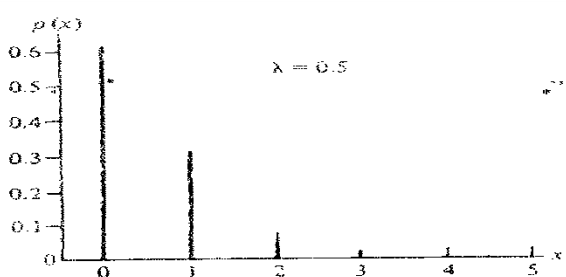
y su esperanza y varianza  $E[X] = \lambda \quad V[X] = \lambda$ .

La distribución de Poisson se conoce como la de los sucesos raros porque sirve para aproximar a la binomial cuando la probabilidad de éxito es muy pequeña.

Teorema. Sea  $Y_n \stackrel{d}{=} B(n, p_n)$ , si  $n \rightarrow \infty \quad np_n = \lambda \Rightarrow Y_n \xrightarrow{Ley} \wp(\lambda)$

Demostración.  $\square$

Observación. Si  $n$  aumenta y  $p$  disminuye,  $B(n, p) \approx \wp(\lambda = np)$ . Utilizar si  $n > 50$ ,  $p < 0,1$  y  $np < 5$ .



## VI.1.4 Hipergeométrica

Corresponde al muestreo sin reemplazamiento (con reemplazamiento sería la binomial).

Supóngase que en una población hay  $D$  individuos de un tipo y  $N-D$  de otro. Se extraen  $n$  individuos y se quiere observar la variable  $X$ : número de individuos tipo "D" en la muestra.

Si es con reemplazamiento sería  $P(X = x) = \binom{n}{x} \left(\frac{D}{N}\right)^x \left(\frac{N-D}{N}\right)^{n-x}$  que sería  $B(n, D/N)$

Si es sin reemplazamiento sería  $P(X = x) = \frac{\binom{D}{x} \binom{N-D}{n-x}}{\binom{N}{n}}$ , que es la Hipergeométrica

$H(N, D, n)$ .

La esperanza de esta distribución es  $nD/N$ , que es la misma que la de la correspondiente binomial. Y la varianza es  $V(X) = \frac{nD(N-D)(N-n)}{N^2(N-1)}$ . Obsérvese que la de la binomial

asociada sería  $\frac{nD(N-D)}{N^2}$ , es decir, la de la hipergeométrica sería  $(N-n)/(N-1)$  veces la de la

binomial. Cuando hay una gran diferencia entre  $n$  y  $N$ , ese término es prácticamente 1, por lo que se suele aproximar el muestreo sin reemplazamiento como si fuera con reemplazamiento, y utilizar la binomial que es mucho más sencilla.

## VI.1.5 Distribución Binomial Negativa

Es muy poco utilizada, y representa en la repetición de experimentos de Bernoulli el número de fracasos hasta el  $n$ -ésimo éxito.

$$P(X = x) = \binom{n+x-1}{x} p^n q^x \quad x = 0, 1, \dots$$

## VI.2 Distribuciones continuas

Distribución	F. de densidad	F. Distribución	Media	Varianza	F. Carácter.
Uniforme (a,b)	$f(x) = \frac{1}{b-a} I_{(a,b)}(x)$	$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{itb} - e^{ita}}{it(b-a)}$
Exponencial (λ)	$f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$	$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$	$1/\lambda$	$1/\lambda^2$	$\frac{1}{1-it/\lambda}$
Normal (μ,σ)	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	---	$\mu$	$\sigma^2$	$e^{it\mu - \frac{1}{2}t^2\sigma^2}$
Lognormal (μ,σ)	$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln x - \mu}{\sigma}\right)^2} I_{(0,\infty)}(x)$	--	$e^{\mu + \sigma^2/2}$	$e^{2\mu + \sigma^2} (e^{\sigma^2} - 1)$	
Gamma (p,a)	$f(x) = \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} I_{(0,\infty)}(x)$		$p/a$	$p/a^2$	$\frac{1}{(1-it/a)^p}$
Beta (α,β)	$f(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} I_{(0,1)}(x)$		$\frac{\alpha}{\alpha + \beta}$	$\frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}$	
Weibull (α,β)	$f(x) = \alpha\beta^\alpha x^{\alpha-1} e^{-(\beta x)^\alpha} I_{(0,\infty)}(x)$	$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-(\beta x)^\alpha} & x \geq 0 \end{cases}$	$\frac{1}{\alpha\beta} \Gamma\left(\frac{1}{\alpha}\right)$	$\frac{1}{\alpha\beta^2} \cdot \left(2\Gamma\left(\frac{2}{\alpha}\right) - \frac{1}{\alpha} \left(\Gamma\left(\frac{1}{\alpha}\right)\right)^2\right)$	
Cauchy (α,β)	$f(x) = \frac{\alpha/\pi}{\alpha^2 + (x-\beta)^2}$		$\cancel{\neq}$	$\cancel{\neq}$	

## VI.2.1 Uniforme

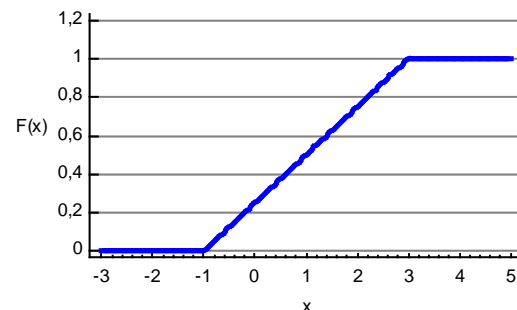
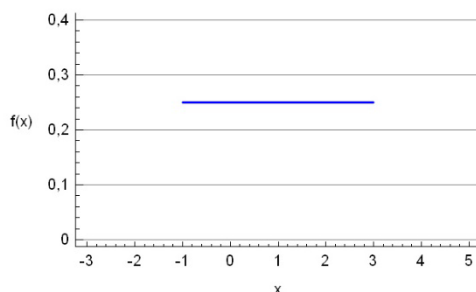
Nos encontramos con una variable aleatoria  $X$  que toma valores solo en el intervalo  $(a,b)$  y cuya función de densidad tendrá el mismo valor en todos los puntos entre  $a$  y  $b$ , es decir, se trata de una función constante en el intervalo  $(a,b)$  de altura  $1/(b-a)$ :

$$f(x) = \frac{1}{b-a} I_{(a,b)}(x)$$

La función de distribución es

$$F(x) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x < b \\ 1 & x \geq b \end{cases}$$

En la figura se pueden observar las gráficas de la función de densidad y de la función de distribución de una variable  $U(-1,3)$



La media coincide con el punto medio del intervalo  $(a,b)$ ,  $E[X] = \frac{a+b}{2}$ , y la varianza es

$$V[X] = \frac{(b-a)^2}{12}$$

## VI.2.2 Exponencial

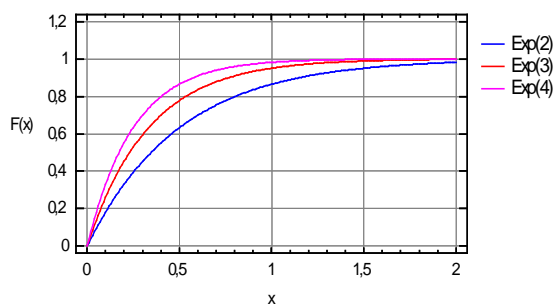
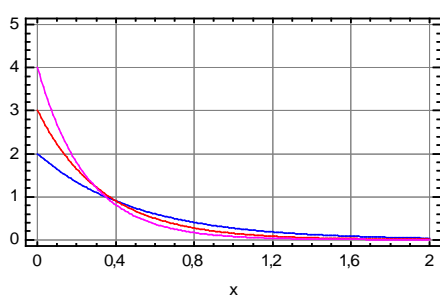
Dentro de las distribuciones continuas, otra distribución que aparece en numerosas ocasiones es la exponencial. Se utiliza, entre otras cosas, como modelo probabilístico para representar

tiempos de funcionamiento o de vida de algún dispositivo, tiempos de espera, o en general, tiempos entre dos sucesos.

Se dice que la variable aleatoria  $X$  tiene una distribución exponencial de parámetro  $\lambda > 0$ , si su función de densidad es de la forma  $f(x) = \lambda e^{-\lambda x} I_{(0,\infty)}(x)$

La función de distribución es

$$F(x) = \begin{cases} 0 & x < 0 \\ 1 - e^{-\lambda x} & x \geq 0 \end{cases}$$



Esperanza y varianza son:  $E[X] = 1/\lambda$   $V[X] = 1/\lambda^2$

Se dice que la distribución exponencial no tiene memoria. La distribución exponencial es la única distribución continua con esta propiedad, que puede interpretarse en el sentido de que la historia pasada solo influye en la historia futura a través del presente.

## VI.2.3 Normal

La distribución normal es, con mucho, la distribución de probabilidad más importante en estadística, tanto por razones de conveniencia matemática como experimental, ya que aproxima muy bien las distribuciones de variables relacionadas con un gran número de experimentos en áreas de conocimiento que van desde la medicina, la biología, la química hasta la economía, la sociología, etc.

Fue utilizada por primera vez por Abraham de Moivre en 1733, como la aproximación de la distribución de la suma de variables aleatorias binomiales. Utilizada también por Karl F. Gauss a



principios del siglo XIX en su estudio de los errores de medida, la normal es actualmente modelo de referencia para variables aleatorias en cuya distribución los valores más usuales se agrupan en torno a uno central, y los valores extremos son poco frecuentes.

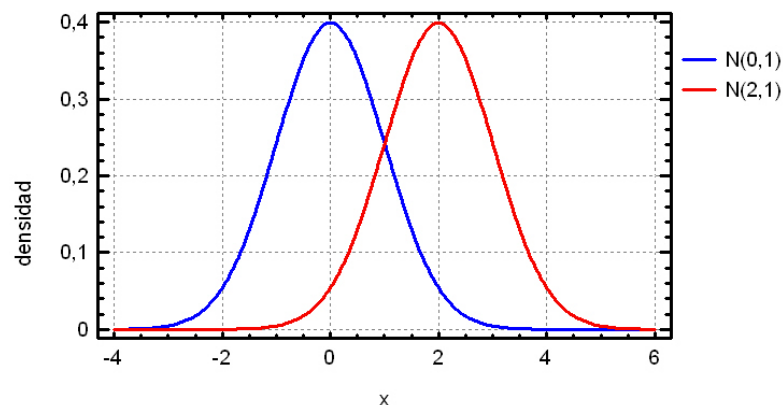
Se dice que una variable aleatoria  $X$  tiene una distribución normal de media  $\mu$  y desviación típica  $\sigma$ ,  $N(\mu, \sigma)$ , si su función de densidad es de la forma

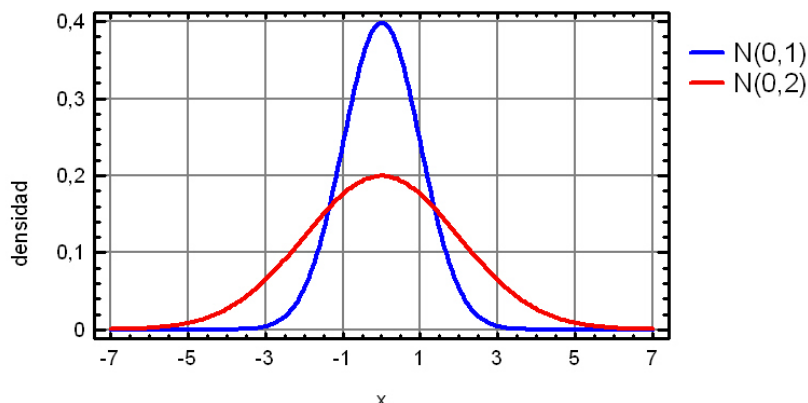
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

La función de densidad de la distribución normal es simétrica con respecto a su media, es decir,  $f(\mu - x) = f(\mu + x)$ . Además son iguales media, mediana y moda.

Esto significa que la función de densidad alcanza su máximo en el punto  $x = \mu$ .

En la figura se representan las funciones de densidad de algunas distribuciones normales para diferentes valores de los parámetros.





La media localiza la función; es decir, si cambia la media, la gráfica se desplaza, situación que se muestra en la primera figura. Por su parte, la desviación típica o la varianza tiene que ver con la estilización de la gráfica: cuando la varianza aumenta, la gráfica es más abierta (menos estilizada); mientras que si la varianza disminuye, la función es más cerrada (más estilizada), como puede verse en la figura.

Una propiedad de la distribución normal es que ante transformaciones lineales  $Y = aX + b$ , sigue siendo normal, y en particular  $N(a\mu + b, |a|\sigma)$ . La transformación lineal que vimos como más importante es la tipificación, que consiste en restar la media y dividir por la desviación típica, quedando entonces una distribución  $N(0,1)$ , también conocida como distribución Normal estándar. Su función de densidad es  $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$

La distribución normal estándar está tabulada, es decir, que hay tablas para calcular sus probabilidades. Por lo tanto, para calcular probabilidades de cualquier distribución normal, ésta se tipifica y se obtienen con la normal estándar, habitualmente denotada por  $Z$ .

Para buscar una probabilidad con la ayuda de la tabla, se procede de la siguiente manera: si se desea encontrar  $\Phi(z) = P(Z \leq z)$ , se busca  $z$  en la tabla, unidades y décimas en la columna de la izquierda y centésimas en la fila superior. El número que aparece en el interior de la tabla en la intersección de la fila y la columna correspondientes a  $z$  es la probabilidad buscada.

Por ejemplo, para encontrar  $\Phi(0.32) = P(Z \leq 0.32)$  buscamos 0.3 en la columna de la izquierda y 2 en la fila superior. La probabilidad buscada es 0.6255.

Recíprocamente, si conocemos el valor de la función de distribución en un punto  $z$ , las tablas nos proporcionan el valor de  $z$ .

Por ejemplo, si sabemos que  $\Phi(z) = 0.9778$ , buscando este número en el interior de la tabla podemos observar que se encuentra en la intersección de la fila correspondiente al 2.0 y de la columna del 0.01. Por tanto,  $z = 2.01$

Dada la tabla de valores y utilizando la simetría de la distribución respecto a su media 0, es posible obtener cualquier valor de la función de distribución.

- Si  $z \geq 0$ , la probabilidad buscada se obtiene directamente de la tabla.
- Si  $z < 0$ , basta tener en cuenta que, por simetría,  $P(Z \leq z) = P(Z \geq -z) = 1 - P(Z < -z)$ . En definitiva  $\Phi(z) = 1 - \Phi(-z)$  con  $-z > 0$ .

En el caso de que las probabilidades a calcular sean de una normal no estándar se hace primero la tipificación, como se muestra en el siguiente ejemplo:

Sea  $X$  una variable aleatoria que sigue una distribución  $N(2,3)$ . Veamos cómo se obtienen las probabilidades de algunos intervalos:

- $$P(X \leq 2.5) = P\left(\frac{X-2}{3} \leq \frac{2.5-2}{3}\right) = P(Z \leq 0.17) = \Phi(0.17) = 0.5675$$

$$\begin{aligned} P(1.2 < X \leq 2.3) &= P\left(\frac{1.2-2}{3} < Z \leq \frac{2.3-2}{3}\right) = P(-0.27 < Z \leq 0.1) = \\ &= \Phi(0.1) - \Phi(-0.27) = \Phi(0.1) - (1 - \Phi(0.27)) = 0.5398 - (1 - 0.6064) = \\ &= 0.1462 \end{aligned}$$

En ocasiones es útil calcular probabilidades de la distribución normal dadas en función de la desviación típica. Es decir, si  $X \approx N(\mu, \sigma)$ , se trata de calcular la probabilidad de que la variable se encuentre en el intervalo  $(\mu - k\sigma, \mu + k\sigma)$ . En definitiva:

$$P(\mu - k\sigma < X < \mu + k\sigma) = P(-k < Z < k) = \Phi(k) - \Phi(-k) = \Phi(k) - (1 - \Phi(k)) = 2\Phi(k) - 1$$

Así, se tienen las siguientes probabilidades:

$$P(\mu - \sigma < X < \mu + \sigma) = 2\Phi(1) - 1 = 0.6826$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 2\Phi(2) - 1 = 0.9544$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 2\Phi(3) - 1 = 0.9974$$

Probabilidades que, como se puede observar, no dependen ni de la media ni de la desviación típica de la distribución.

Una aplicación importante de la normal es la aproximación que de la distribución binomial se puede hacer mediante la normal.

En general, se puede comprobar que el polígono de frecuencias de una variable aleatoria con distribución  $B(n,p)$  se parece a la curva de densidad de una normal tanto más cuanto mayor sea el producto  $np$ , siempre que  $p$  o  $q$  no estén demasiado próximos a cero. En concreto, si se observa el polígono de frecuencias puede intuirse que, si  $p$  es fijo y  $n$  suficientemente grande, la distribución  $B(n,p)$  puede aproximarse por la distribución  $N(np, \sqrt{npq})$ . Este resultado fue demostrado por Abraham De Moivre de forma analítica.

Dado que la distribución binomial es simétrica si  $p=0.5$ , esta aproximación funcionará tanto mejor cuanto más cerca esté  $p$  de 0.5. Las condiciones para utilizar la aproximación normal dependerán de la precisión que se desee obtener en cada problema concreto.

En estas condiciones podemos utilizar la distribución normal para calcular probabilidades de la distribución binomial. Ello proporciona la ventaja de que las probabilidades de la normal están tabuladas, mientras que las de la binomial requieren cálculos tediosos cuando  $n$  es grande.

Sin embargo, hemos de tener en cuenta que la distribución binomial es discreta, mientras que la normal es continua, y debemos efectuar una corrección por continuidad en la aproximación.

Sea  $Y$  una variable aleatoria con distribución  $B(n,p)$ . Entonces, la probabilidad  $P(Y = k)$  tiene el mismo valor que el área de un rectángulo de base el intervalo  $[k-0.5, k+0.5]$  y altura  $P(Y=k)$ . El área de ese rectángulo se puede aproximar por el área bajo la función de densidad de la  $N(np, \sqrt{npq})$ . En general, si la variable aleatoria  $Y \approx B(n, p)$  con  $n$  grande, puede ser aproximada por la variable  $X \approx N(np, \sqrt{npq})$  realizando la corrección por continuidad, de modo que los cálculos de probabilidades relativos a la binomial se efectúan de la siguiente manera:

- $$P(Y \leq k) \approx P(X \leq k + 0.5) = \Phi\left(\frac{k + 0.5 - np}{\sqrt{npq}}\right)$$

- $P(Y < k) \approx P(X \leq k - 0.5) = \Phi\left(\frac{k - 0.5 - np}{\sqrt{npq}}\right)$
- $P(Y = k) \approx P(k - 0.5 \leq X \leq k + 0.5) = \Phi\left(\frac{k + 0.5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{k - 0.5 - np}{\sqrt{npq}}\right)$
- $P(j \leq Y \leq k) \approx P(j - 0.5 \leq X \leq k + 0.5) = \Phi\left(\frac{k + 0.5 - np}{\sqrt{npq}}\right) - \Phi\left(\frac{j - 0.5 - np}{\sqrt{npq}}\right)$

Cualquier otra posibilidad puede deducirse de las anteriores.

Supóngase que se tiene una moneda trucada, en la que la probabilidad de cara es  $p=0.4$ . Se lanza 100 veces dicha moneda. Sea  $Y$  la variable aleatoria número de caras obtenidas. Queremos conocer la probabilidad de obtener, como mucho, 45 caras.

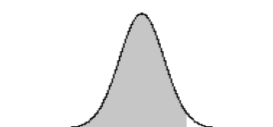
Como  $Y \approx B(100, 0.4)$ , si quisiéramos calcular la probabilidad exacta sería:

$$P(Y \leq 45) = \binom{100}{0} 0.4^0 0.6^{100} + \binom{100}{1} 0.4^1 0.6^{99} + \dots + \binom{100}{45} 0.4^{45} 0.6^{55}$$

Realizar estos cálculos puede ser muy tedioso. Sin embargo, puesto que podemos aproximar la distribución  $X \approx N(40, \sqrt{24})$  los cálculos son muy simples:

$$P(Y \leq 45) \approx P(X \leq 45 + 0.5) = P(Z \leq \frac{45.5 - 40}{4.9}) = \Phi(1.12) = 0.8686$$

Tabla 2: Distribución N(0,1)



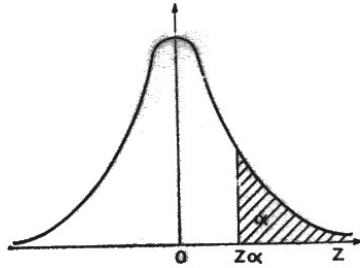
$P(Z < z)$

Segunda cifra decimal del valor de z										
z	0.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998

Otra tabla muy utilizada, especialmente en estadística es la de la Normal (0,1) pero a la derecha, que se incluye a continuación. Es especialmente utilizada en estadística.

Distribución normal  $N(0, 1)$

$$\int_{z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = \alpha$$



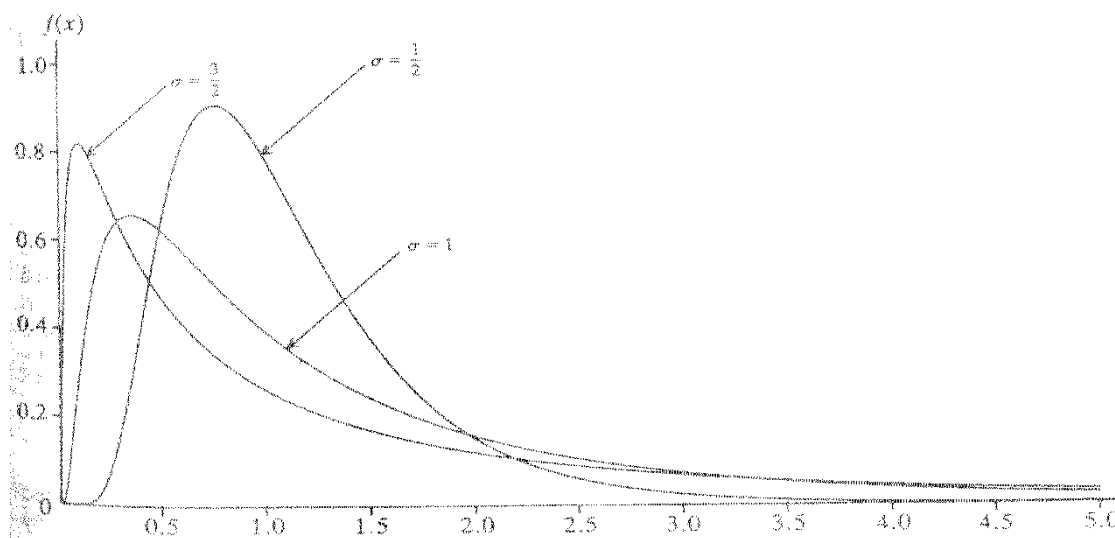
$z_\alpha$	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,4641
0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
0,3	0,3821	0,3783	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,1894	0,1867
0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1170
1,2	0,1151	0,1131	0,1112	0,1093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
2,0	0,0228	0,0222	0,0217	0,0212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
2,1	0,0179	0,0174	0,0170	0,0166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
2,3	0,0107	0,0104	0,0102	0,00990	0,00964	0,00939	0,00914	0,00889	0,00866	0,00842
2,4	0,00820	0,00798	0,00776	0,00755	0,00734	0,00714	0,00695	0,00676	0,00657	0,00639
2,5	0,00621	0,00604	0,00587	0,00570	0,00554	0,00539	0,00523	0,00508	0,00494	0,00480
2,6	0,00466	0,00453	0,00440	0,00427	0,00415	0,00402	0,00391	0,00379	0,00368	0,00357
2,7	0,00256	0,00336	0,00326	0,00317	0,00307	0,00298	0,00289	0,00280	0,00272	0,00264
2,8	0,00256	0,00248	0,00240	0,00233	0,00226	0,00219	0,00212	0,00205	0,00199	0,00193
2,9	0,00187	0,00181	0,00175	0,00169	0,00164	0,00159	0,00154	0,00149	0,00144	0,00139

$z_\alpha$	0,0	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
3	0,00135	0,00968	0,00687	0,00483	0,00337	0,00233	0,00159	0,00108	0,000723	0,000481
4	0,004317	0,00207	0,00133	0,000854	0,000541	0,000340	0,000211	0,000130	0,0000793	0,0000479
5	0,00287	0,00170	0,000996	0,000579	0,000333	0,000190	0,000107	0,0000599	0,0000332	0,0000182
6	0,00987	0,00530	0,00282	0,00149	0,000777	0,000402	0,000206	0,000104	0,0000523	0,0000260



## VI.2.4 Log-normal

Una distribución que se usa mucho en estudios socio-económicos para variables que tengan asimetría a la derecha es la log-normal o logarítmico-normal. Se obtiene de la normal haciendo su exponencial. Es decir, el logaritmo de una variable con distribución log-normal, sigue una distribución normal.



## VI.2.5 Distribución Gamma

Es una distribución, de la familia exponencial, muy utilizada en determinados contextos, como la fiabilidad y la supervivencia.

Su función de densidad es:  $f(x) = \frac{a^p}{\Gamma(p)} x^{p-1} e^{-ax} I_{(0,\infty)}(x)$

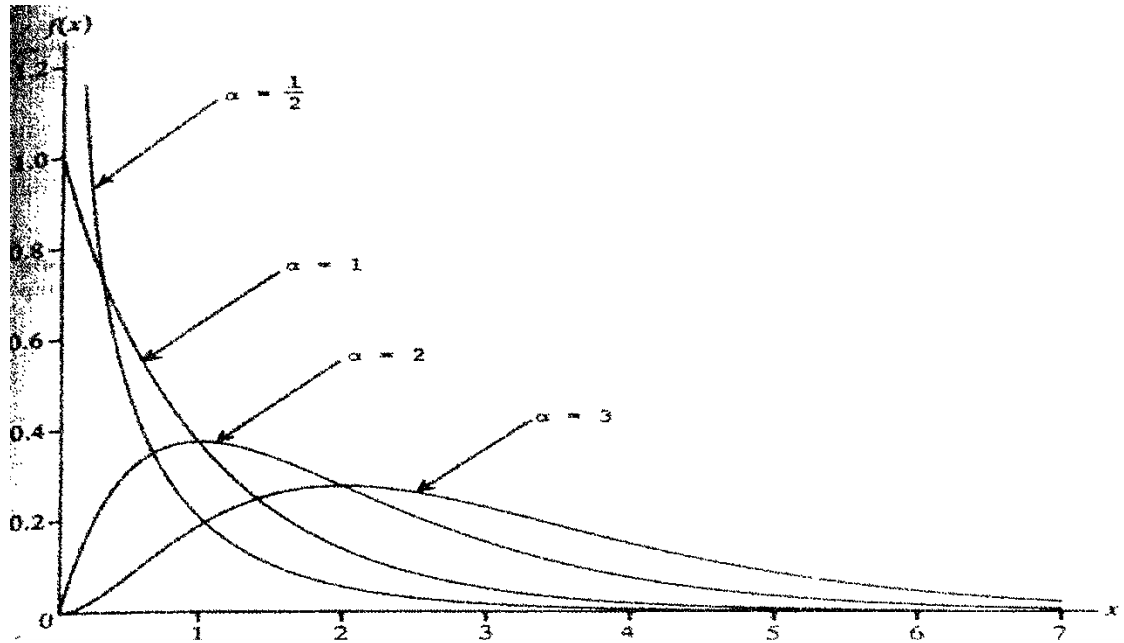
donde la función  $\Gamma(x)$  es la función matemática gamma, que hace que la integral de la función de densidad sea 1 ( $\Gamma(x) = \int_0^{\infty} t^{x-1} e^{-t} dt$ ). Esta función tiene las siguientes propiedades:

$$\Gamma(1) = 1 \quad \Gamma(1/2) = \sqrt{\pi} \quad \Gamma(p) = (p-1)\Gamma(p-1)$$



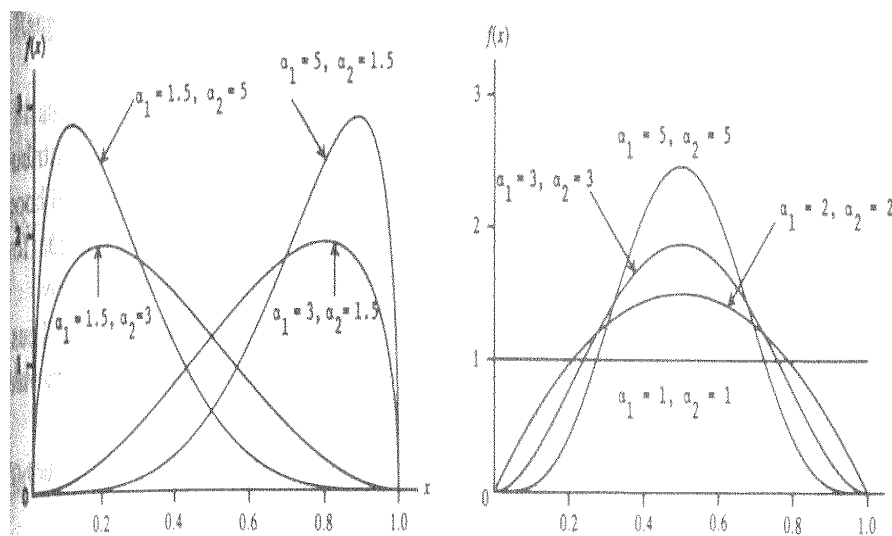
El caso más habitual es que  $p$  sea un número natural (en cuyo caso hablamos de distribución Erlang), y en ese caso  $\Gamma(p) = (p-1)!$ .

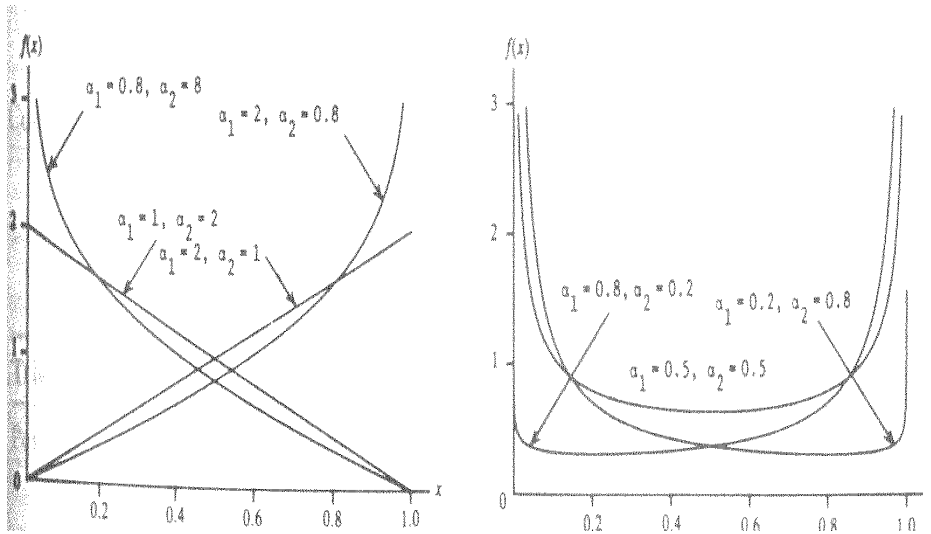
Para el caso  $p=1$  es una Exponencial.



## VI.2.6 Distribución Beta

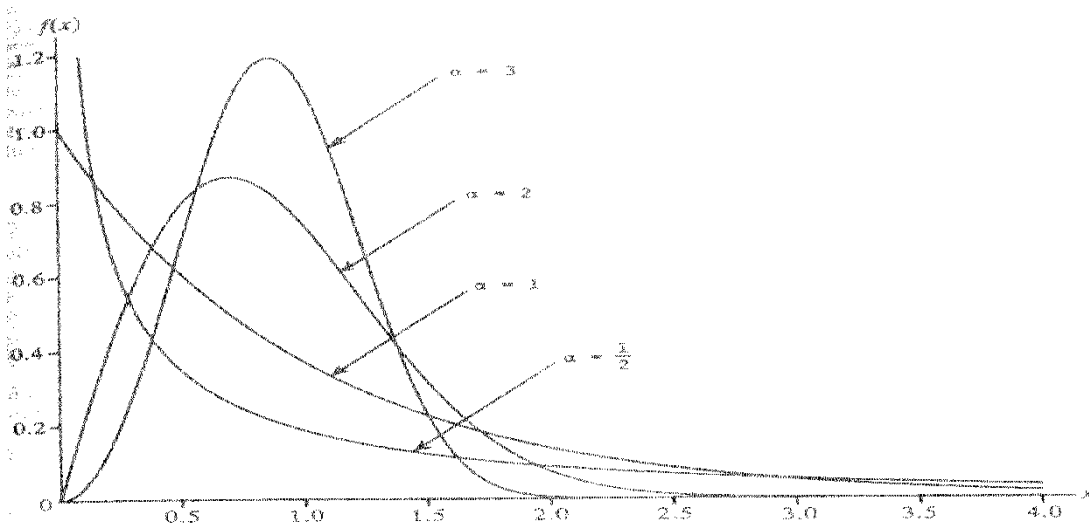
Muy utilizada en ingeniería, especialmente en el método PERT de gestión de proyectos, por sus variadas formas dependiendo del valor de los parámetros.





### VI.2.7 Distribución Weibull

Muy utilizada en ingeniería y simulación, para tiempos de reparación, etc.. También es de la familia exponencial, y para  $\alpha = 1$  es la propia Exponencial. Tiene como ventaja respecto a la Gamma que tiene una forma cerrada de función de distribución



### VI.2.8 Distribución de Cauchy

Parecida a la normal, pero con función de distribución cerrada. Por el contrario, tiene la desventaja de que no existe ninguno de sus momentos.

## VII Variables aleatorias multidimensionales

Aunque las definiciones que se han dado son válidas en general, hay algunas cuestiones específicas cuando las variables están definidas no en  $\mathbb{R}$ , sino en  $\mathbb{R}^n$ , siendo denominadas *variables multidimensionales*.

### VII.1 Distribuciones y variables aleatorias n-dimensionales

En esta sección introduciremos los conceptos básicos de variables aleatorias multidimensionales.

#### VII.1.1 Variables y funciones de distribución

Definición. Una variable aleatoria n-dimensional es toda aplicación medible  $X : (S, \mathcal{Q}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  es decir, toda aplicación  $X : S \rightarrow \mathbb{R}^n / X^{-1}(B) \in \mathcal{Q}, \forall B \in \mathcal{B}^n$

Teniendo en cuenta que la  $\sigma$ -álgebra de Borel en  $\mathbb{R}^n$  se define como:

- La mínima  $\sigma$ -álgebra que contiene a todos los “intervalos”  
 $(a, b] = \{(x_1, \dots, x_n) / a_k < x_k \leq b_k \forall k = 1, \dots, n\} \quad a, b \in \mathbb{R}^n$
- La mínima  $\sigma$ -álgebra que contiene a los “rectángulos medibles”  
 $A_1 \times \dots \times A_n$  con  $A_k \in \mathcal{B} \forall k = 1, \dots, n$
- $\sigma$ -álgebra engendrada por los “intervalos” de la forma  
 $(-\infty, b] = \{(x_1, \dots, x_n) / x_k \leq b_k \forall k = 1, \dots, n\} \quad b \in \mathbb{R}^n$

Teorema. Una aplicación  $X = (X_1, \dots, X_n) : (S, \mathcal{Q}) \rightarrow (\mathbb{R}^n, \mathcal{B}^n)$  es variable aleatoria n-dimensional si y solo si todas sus coordenadas definen sendas variables aleatorias unidimensionales.

Definición.  $X = (X_1, \dots, X_n) : (S, \mathcal{Q}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$  variable aleatoria, se define  $P_X : (\mathbb{R}^n, \mathcal{B}^n) \rightarrow [0, 1]$  la medida de probabilidad inducida como  $P_X(B) = P(X^{-1}(B)) \quad \forall B \in \mathcal{B}^n$ .

Definición. Sea  $(\mathbb{R}^n, \mathcal{B}^n, P)$  espacio de probabilidad. La función de distribución asociada a  $P$  se define como

$$F : \mathbb{R}^n \rightarrow [0, 1]$$

$$(x_1, \dots, x_n) \rightarrow F(x_1, \dots, x_n) = P((y_1, \dots, y_n) / y_k \leq x_k \forall k = 1, \dots, n)$$

Definición. Dada  $X = (X_1, \dots, X_n) : (S, \mathcal{Q}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$  variable aleatoria, la función de distribución de la variable aleatoria  $X$  se define como  $F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = P(s \in S / X_1(s) \leq x_1, \dots, X_n(s) \leq x_n) = P_X \left\{ (y_1, \dots, y_n) \in \mathbb{R}^n / y_k \leq x_k \quad \forall k \right\}$

Teorema. Una función  $F : \overline{\mathbb{R}}^n \rightarrow [0, 1]$  es función de distribución de una medida de probabilidad en  $\mathcal{B}^n$  si y solo si:

- 1.-  $F(+\infty, \dots, +\infty) = 1$
- 2.-  $F(-\infty, x_2, \dots, x_n) = \dots = F(x_1, \dots, -\infty) = 0$
- 3.-  $F$  continua por la derecha en cada variable
- 4.-  $\nabla_{a,b} F(x_1, \dots, x_n) \geq 0$  donde  $\nabla_{a,b} F(x_1, \dots, x_n) = \nabla_{a_n, b_n} \dots \nabla_{a_1, b_1} F(x_1, \dots, x_n)$  siendo  $\nabla_{a_i, b_i} G(y_1, \dots, y_n) = G(y_1, \dots, b_i, \dots, y_n) - G(y_1, \dots, a_i, \dots, y_n)$

Además esta probabilidad es única.

La propiedad 4 asegura que la probabilidad en los intervalos es positiva. En el caso de  $\mathbb{R}^2$ , la probabilidad de un rectángulo es:

$$P((x, y) / a_1 < x \leq b_1, a_2 < y \leq b_2) = F(b_1, b_2) - F(a_1, b_2) - F(b_1, a_2) + F(a_1, a_2)$$

Y la expresión de la propiedad 4 es:

$$\begin{aligned}\nabla_{a,b} F(x, y) &= \nabla_{a_2, b_2} \nabla_{a_1, b_1} F(x, y) = \nabla_{a_2, b_2} (F(b_1, y) - F(a_1, y)) = \\ &= F(b_1, b_2) - F(a_1, b_2) - (F(b_1, a_2) - F(a_1, a_2))\end{aligned}$$

Tipos de distribuciones:

- Una distribución se dice *discreta* si concentra su probabilidad en un conjunto finito o numerable de puntos, y queda determinada por la *función de masa de probabilidad conjunta* que son las probabilidades en esos puntos. De modo que

$$P_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \quad / \quad P_X(B) = \sum_{\vec{x} \in B \cap D} P_X(\vec{x}) \quad \forall B \in \mathcal{B}^n \quad \text{siendo el soporte } D \text{ los puntos donde la probabilidad es mayor que cero, y siendo } \sum_{(x_1, \dots, x_n) \in D} P_X(x_1, \dots, x_n) = 1.$$

- Una distribución se dice absolutamente continua si existe  $f(x_1, \dots, x_n) \geq 0$  tal que

$$F(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \dots \int_{-\infty}^{x_n} f(y_1, \dots, y_n) dy_1 \dots dy_n \quad \text{y entonces } P(B) = \int_B f(\vec{x}) d\vec{x} \quad \forall B \in \mathcal{B}^n.$$

En principio es integral de Lebesgue, pero se calculará normalmente de forma secuencial usando el teorema de Fubini.  $f$  se llama función de densidad y verifica

$$\int_{\mathbb{R}^n} f(y_1, \dots, y_n) d(y_1, \dots, dy_n) = 1.$$

- Hay distribuciones que no son ni discretas ni continuas porque acumulan la probabilidad en un conjunto de medida nula, pero que es no numerable (una recta...)

## VII.1.2 Distribuciones marginales y condicionadas

Esta sección se centra en el estudio de variables aleatorias de menor dimensión que la  $n$ -dimensional que llamaremos conjunta.

Definición. Sea la variable aleatoria  $X = (X_1, \dots, X_n) : (S, \mathcal{Q}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$  se llama *función de distribución marginal de  $X_k$*  a la función de distribución de  $X_k$  considerada como variable aleatoria unidimensional que es:

$$F_{X_k}(x_k) = F_{(X_1, \dots, X_n)}(+\infty, \dots, x_k, \dots, +\infty) = P(s \in S / X_k(s) \leq x_k) \quad \forall x_k \in \mathbb{R}$$

Procedimiento para obtener las distribuciones marginales:

1. Si X es discreta, sus marginales son discretas, y la función de masa de probabilidad de la marginal se obtiene sumando en todos los posibles valores de las otras, y el de la variable en cuestión fijo:

$$p_{X_i}(x_i) = \sum_{x_1} \dots \sum_{x_{i-1}} \sum_{x_{i+1}} \dots \sum_{x_n} p_{(X_1, \dots, X_n)}(x_1, \dots, x_n)$$

2. Si X es absolutamente continua con función de densidad f, sus marginales también son absolutamente continuas y la función de densidad marginal de una variable se obtiene fijando el valor de la variable en cuestión e integrando en las otras variables en toda la recta:

$$f_{X_i}(x_i) = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) dx_1 \dots dx_{i-1} dx_{i+1} \dots dx_n$$

Aplicando estas definiciones al caso de dos dimensiones, si (X,Y) es discreta la función de masa de X es  $p_X(x) = \sum_y p_{(X,Y)}(x, y)$  y análogamente para Y; si (X,Y) es absolutamente

continua, la función de densidad de X es  $f_X(x) = \int_{\mathbb{R}} f_{(X,Y)}(x, y) dy$ .

Definición. Dada la variable aleatoria bidimensional  $(X, Y) : (S, \mathcal{Q}, P) \rightarrow (\mathbb{R}^2, \mathcal{B}^2, P_X)$  para cada  $x \in \mathbb{R}$  fijo se define la función de distribución condicionada por  $X=x$  como

$$F_{Y/X=x}(y) = \lim_{h \rightarrow 0^+} \frac{F_{(X,Y)}(x, y) - F_{(X,Y)}(x-h, y)}{F_{(X,Y)}(x, \infty) - F_{(X,Y)}(x-h, \infty)} \text{ para todo } y \in \mathbb{R} \text{ para el que tal límite exista.}$$

1. Cuando (X,Y) es discreta, no sólo las marginales son discretas, también lo son las condicionadas y su función de masa de probabilidad se obtiene:

$$\forall x / P_X(x) > 0 \quad P_{Y/X=x}(y) = \frac{P_{(X,Y)}(x, y)}{P_X(x)}$$

2. Cuando (X,Y) es absolutamente continua, no sólo las marginales lo son sino también las condicionadas y su función de densidad se obtiene:

$$\forall x / f_X(x) > 0 \quad f_{Y/X=x}(y) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}$$

Ambos resultados se demuestran a partir de la función de distribución condicionada.

Observaciones:

- En ambos casos es “conjunta” entre “marginal”
- Si todas las marginales son absolutamente continuas, la conjunta no tiene por qué serlo.
- Si todas las marginales son discretas, la conjunta es discreta.
- Si se conoce la conjunta entonces se pueden conocer marginales y condicionadas
- Si se conocen las marginales solo, no se puede conocer la conjunta
- Si se conoce la marginal de una variable y las condicionadas de la otra a todos sus posibles valores, entonces se puede conocer la conjunta
- Para cada  $x$ , si existe, se tiene una distribución condicionada completa, y esta familia es la que permite conocer la conjunta. En muchos casos hablaremos de LA condicionada para referirnos a la familia de condicionadas.

### VII.1.3 Función característica de una v.a. n-dimensional

Definición. Sea  $(\mathbb{R}^n, \mathcal{B}^n, P)$  un espacio de probabilidad con  $F$  función de distribución asociada a  $P$ . Se define la *función característica asociada a  $P$*  como

$$\varphi: \mathbb{R}^n \rightarrow \mathbb{C}$$

$$(t_1, \dots, t_n) \rightarrow \varphi(t_1, \dots, t_n) = \int_{\mathbb{R}^n} e^{i(t_1, \dots, t_n) \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}} dP = \int_{\mathbb{R}^n} e^{i \sum_{j=1}^n t_j x_j} dF(x_1, \dots, x_n)$$

Definición. Dada una variable aleatoria  $(X_1, \dots, X_n): (S, \mathcal{Q}, P) \rightarrow (\mathbb{R}^n, \mathcal{B}^n, P_X)$  se define su *función característica asociada* como la función característica de su probabilidad inducida, es decir,

$$\varphi: \mathbb{R}^n \rightarrow \mathbb{C}$$

$$(t_1, \dots, t_n) \rightarrow \varphi(t_1, \dots, t_n) = E[e^{i t \cdot X}] = \int_S e^{i \sum_{j=1}^n t_j x_j} dP = \int_{\mathbb{R}^n} e^{i \sum_{j=1}^n t_j x_j} dF(x_1, \dots, x_n)$$

aplicando apropiadamente el teorema de cambio de espacio de integración.

Las propiedades son semejantes al caso unidimensional. En concreto, la de que caracteriza a la distribución, pudiéndose generalizar la fórmula del teorema de inversión.

La función característica de las marginales se obtienen a partir de la conjunta como

$$\varphi_{X_i}(t) = \varphi_{(X_1, \dots, X_n)}(0, \dots, t_i, \dots, 0)$$

## VII.1.4 Características de variables n-dimensionales

Son valores puntuales que no caracterizan la distribución pero ayudan a tomar decisiones.

Definición. Se llama momento respecto al origen de orden  $k_1, \dots, k_n$  de una variable aleatoria a la esperanza, caso de existir la integral,

$$\alpha_{k_1, \dots, k_n} = E[X_1^{k_1} \cdot \dots \cdot X_n^{k_n}] = \int_S X_1^{k_1} \cdot \dots \cdot X_n^{k_n} dP = \int_{\mathbb{R}^n} x_1^{k_1} \cdot \dots \cdot x_n^{k_n} dF(x_1, \dots, x_n)$$

Definición. Se llama momento centrado respecto a las medias de orden  $k_1, \dots, k_n$  de una variable aleatoria a la esperanza, caso de existir la integral,

$$\begin{aligned} \mu_{k_1, \dots, k_n} &= E[(X_1 - \mu_1)^{k_1} \cdot \dots \cdot (X_n - \mu_n)^{k_n}] = \int_S (X_1 - \mu_1)^{k_1} \cdot \dots \cdot (X_n - \mu_n)^{k_n} dP = \\ &= \int_{\mathbb{R}^n} (x_1 - \mu_1)^{k_1} \cdot \dots \cdot (x_n - \mu_n)^{k_n} dF(x_1, \dots, x_n) \end{aligned}$$

siendo  $\mu_j = E(X_j)$ .

Los momentos de las variables unidimensionales se obtienen poniendo orden 0 al resto.

$$\alpha_{0, \dots, k_j, \dots, 0} = E[X_j^{k_j}] = \int_S X_j^{k_j} dP = \int_{\mathbb{R}^n} x_j^{k_j} dF(x_1, \dots, x_n) = \int_{\mathbb{R}} x_j^{k_j} dF_{X_j}(x_j)$$

Con un 1 se obtiene la esperanza de la marginal, y el vector de medias o esperanzas es el vector con todas las esperanzas marginales.

Igualmente, se hace con la varianza, que sería  $V(X_j) = \sigma_j^2 = \sigma_{jj} = \mu_{0, \dots, 2, \dots, 0}$

Análogamente, se puede hablar del momento  $\mu_{0, \dots, 1, 0, \dots, 1, 0, \dots, 0}$  llamado covarianza y que se calcula como  $\text{cov}(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] = \sigma_{ij} = E(X_i X_j) - \mu_i \mu_j$



Nótese que es la esperanza o media de la multiplicación de las desviaciones de la variable  $X_i$  con respecto a su media  $\mu_i$  por las desviaciones de la variable  $X_j$  respecto a su media  $\mu_j$ .

Definición. La matriz de varianzas/covarianzas de la variable aleatoria n-dimensional se define como

$$\Sigma = (\sigma_{ij})_{ij} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}$$

Las covarianzas son medidas simétricas, es decir,  $C_{ij} = C_{ji}$ , y por lo tanto es una matriz simétrica. Además es una matriz semidefinida positiva ( $y^T \Sigma y \geq 0 \quad \forall y \in \mathbb{R}^n$ )

Si el valor de la covarianza es positivo, ello implica que  $X_i$  y  $X_j$  están correladas positivamente, es decir, si el valor que se tiene de  $X_i$  es mayor que su media  $\mu_i$ , entonces el valor que se espera para  $X_j$  será mayor que su media  $\mu_j$ . Por el contrario si el valor de la covarianza es negativo, ello implica que  $X_i$  y  $X_j$  están correladas negativamente, es decir, si el valor que se tiene de  $X_i$  es mayor que su media  $\mu_i$ , entonces el valor que se espera para  $X_j$  será menor que su media  $\mu_j$ .

Proposición. Si  $\exists \alpha_{k_1, \dots, k_n} \Rightarrow \alpha_{k_1, \dots, k_n} = \frac{\left[ \frac{\partial^{k_1 + \dots + k_n} \varphi(t_1, \dots, t_n)}{\partial t_1^{k_1} \dots \partial t_n^{k_n}} \right]_{t_1 = \dots = t_n = 0}}{i^{k_1 + \dots + k_n}}$

## VII.2 Independencia de variables aleatorias

Dentro del tratamiento conjunto de variables aleatorias, existe un concepto que es de máxima importancia: el concepto de *independencia*.

Definición. Se dice que una familia de variables aleatorias  $X_i : (S, \mathcal{Q}, P) \rightarrow (\mathbb{R}, \mathcal{B})$   $i = 1, \dots, k$  son independientes si

$$P\left(\bigcap_{i=1}^k X_i^{-1}(B_i)\right) = \prod_{i=1}^k P(X_i^{-1}(B_i)) \quad \forall B_1, \dots, B_k \in \mathcal{B}$$

O lo que es lo mismo, cuando la probabilidad de la variable aleatoria conjunta es la probabilidad producto:  $P_{(X_1, \dots, X_n)}(B_1, \dots, B_n) = \prod_{j=1}^n P_{X_j}(B_j) \quad \forall B_1, \dots, B_n \in \mathcal{B}$

Es decir, la conjunta es el producto de las marginales.

Teorema. Sean  $X_i : (S, \mathcal{Q}, P) \rightarrow (\mathbb{R}, \mathcal{B})$   $i = 1, \dots, n$  variables aleatorias. Son independientes si y solo si cualquiera de las siguientes caracterizaciones se verifica:

$$1. \quad F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{j=1}^n F_{X_j}(x_j) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

$$2. \quad \varphi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \prod_{j=1}^n \varphi_{X_j}(t_j) \quad \forall (t_1, \dots, t_n) \in \mathbb{R}^n$$

$$3. \quad P_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{j=1}^n P_{X_j}(x_j) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n \text{ en caso de conjunta discreta}$$

$$4. \quad f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{j=1}^n f_{X_j}(x_j) \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n \text{ salvo un conjunto de medida de}$$

Lebesgue 0, en caso de conjunta absolutamente continua.

Demostración.  $\square$

En caso de existir las condicionadas, las variables aleatorias X e Y son independientes si y solo si las condicionadas siempre coinciden con la marginal:  $F_{Y/X=x}(y) = F_Y(y) \quad \forall y, \forall x$ .

Obsérvese que el concepto de independencia estadística que se acaba de definir no implica que no haya relación alguna entre las variables, sino que el conocimiento de una no modifica la distribución de probabilidad de las otras. Por ejemplo, sea el experimento sacar una carta de una baraja española y observar el palo y número de la carta que ha salido. Así sea la variable X el palo

(numerando 1 a copas, 2 a oros, 3 a espadas y 4 a bastos), y la variable Y el número (sota es 8, caballo 9 y rey 10). La función de masa de ambas variables conjuntamente es  $p(x, y) = 1/40, \forall x \in \{1, 2, 3, 4\} \forall y \in \{1, \dots, 10\}$ . Las funciones de masa marginales son  $p(x) = 1/4, \forall x \in \{1, 2, 3, 4\}$  y  $p(y) = 1/10, \forall y \in \{1, \dots, 10\}$ . Evidentemente, la función conjunta es producto de las marginales, luego, se puede decir que ambas variables son estadísticamente independientes, es decir, si dada una carta se sabe cuál es el palo no modifica la probabilidad de cuál será su número.

**Teorema.** Dada una familia de variables independientes, entonces toda subfamilia de esas variables también son independientes.

Propiedades.

1.  $(X_1, \dots, X_n), (Y_1, \dots, Y_m)$  independientes, entonces  $g_1(X_1, \dots, X_n), g_2(Y_1, \dots, Y_m)$  son también independientes para todas funciones medibles.
2.  $(X_1, \dots, X_n)$  independientes, entonces  $E(X_1, \dots, X_n) = \prod_{j=1}^n E(X_j)$
3. X e Y independientes, entonces  $cov(X, Y) = 0$  (incorreladas)
4.  $(X_1, \dots, X_n)$  tales que  $cov(X_j, X_k) = 0 \quad \forall j \neq k \Rightarrow V(\sum_{j=1}^n a_j X_j) = \sum_{j=1}^n a_j^2 V(X_j)$
5.  $(X_1, \dots, X_n)$  independientes, entonces  $\varphi_{\sum_{j=1}^n X_j}(t) = \prod_{j=1}^n \varphi_{X_j}(t) \quad \forall t \in \mathbb{R}$

**Definición.** Una familia de distribuciones  $\{F_p\}_{p \in \Theta \subset \mathbb{R}^k}$  se dice reproductiva respecto del parámetro p cuando la suma de dos variables independientes dentro de dicha familia sea otra distribución de la familia con parámetro la suma de los parámetros.

Algunas familias reproductivas:

- B(n,p) respecto de n (p fijo)
- P(L) (Poisson) respecto de L
- Gamma (p,a) respecto de p (a fijo)

- $N(\mu, \sigma^2)$  respecto de los dos a la vez  $(\mu, \sigma^2)$  (ojo, no lo son para uno fijo, siempre es suma de medias y de varianzas)

## VII.3 Transformaciones de variables aleatorias

En muchísimas ocasiones es necesario transformar las variables con que trabajamos. Si solo se quiere la esperanza de la transformación basta hacer la integral de la correspondiente función. Por lo tanto, supongamos que nuestro interés va más allá, y queremos la distribución de una transformación. Vamos a suponer el caso continuo que es el más relevante y complejo a la vez.

Teorema. Sea  $(X, Y)$  variable aleatoria bidimensional absolutamente continua con función de densidad  $f(x, y)$  tal que existe  $C$ , el soporte, con  $P(C)=1$ , donde la transformación  $h: (\mathbb{R}^2, B^2) \rightarrow (\mathbb{R}^2, B^2)$  es inyectiva, y por lo tanto con inversa  $g = h^{-1}: x = h_1^{-1}(z, t) \quad y = h_2^{-1}(z, t) \quad \forall (z, t) \in h(C)$ . Entonces si las inversas tienen derivadas parciales continuas con jacobiano  $J$  no nulo en casi todo punto de  $h(C)$ , se tiene que la variable  $(Z, T)=h(X, Y)$  es absolutamente continua con función de densidad

$$f_{(Z, T)}(z, t) = f_{(X, Y)}(h_1^{-1}(z, t), h_2^{-1}(z, t)) \cdot |J| \quad \text{siendo} \quad J = \begin{vmatrix} \frac{\partial h_1^{-1}}{\partial z} & \frac{\partial h_1^{-1}}{\partial t} \\ \frac{\partial h_2^{-1}}{\partial z} & \frac{\partial h_2^{-1}}{\partial t} \end{vmatrix}$$

Una generalización interesante es si  $h$  permite fraccionar el soporte espacio muestral en una familia finita o numerable de transformaciones definidas en subrecintos de modo que en cada subrecinto es inyectiva y con inversas con derivada continua y jacobiano no nulo, entonces la función de densidad de la variable transformada es la suma de las funciones de densidad en cada original, ponderada cada una por el valor absoluto del correspondiente jacobiano.

Por último, una transformación interesante es la suma, ya que aparece muchas veces. En el caso de ser suma de dos variables independientes, se llama convolución y su función de densidad (que habría que hacer haciendo una transformación doble, y luego su marginal) resulta ser:

$$f_{S=X+Y}(s) = \int_{\mathbb{R}} f_X(s-y)f_Y(y)dy$$

## VII.4 Regresión y correlación

El objetivo es, conocido el valor de una variable dar un posible valor (esperado) para la otra variable.

### VII.4.1 Línea general de regresión

En este caso para dar el valor se busca una función  $g(x)$  que minimice de alguna manera el error que se espera cometer al estimar el valor de la otra variable con esta función. Luego, lo primero es dar una medida de ese error.

Definición. Se define el error cuadrático medio al estimar  $y$  con  $g(x)$  como

$$E.C.M.(g) = \int_{\mathbb{R}^2} (y - g(x))^2 dF(x, y)$$

Definición. Dada una variable aleatoria  $(X, Y)$  se llama línea general de regresión o curva de regresión de  $Y$  sobre  $X$  a la función  $y=g(x)$  que minimiza el error cuadrático medio de entre todas las posibles funciones medibles, que resulta ser la esperanza condicionada, caso de existir ésta:

$$m_{Y/X}(x) = \int_{\mathbb{R}} y dF_{Y/X=x}(y) \quad (c.s.)$$

En efecto:

$$\begin{aligned} \int_{\mathbb{R}^2} (y - g(x))^2 dF(x, y) &= \int_{\mathbb{R}^2} (y - m_{Y/X}(x))^2 dF(x, y) + \int_{\mathbb{R}^2} (m_{Y/X}(x) - g(x))^2 dF(x, y) + \\ &+ 2 \int_{\mathbb{R}^2} (y - m_{Y/X}(x))(m_{Y/X}(x) - g(x)) dF(x, y) \geq \\ &\geq \int_{\mathbb{R}^2} (y - m_{Y/X}(x))^2 dF(x, y) + \int_{\mathbb{R}} 2(m_{Y/X}(x) - g(x)) \underbrace{\int_{\mathbb{R}} (y - m_{Y/X}(x)) dF_{Y/X=x}}_{=0} dF_X = \\ &= \int_{\mathbb{R}^2} (y - m_{Y/X}(x))^2 dF(x, y) \end{aligned}$$

El procedimiento consiste en sacar la marginal de la conjunta, y con ella las condicionadas; posteriormente hacer la esperanza de de las condicionadas.

Observaciones:

1. Al ser una esperanza puede no existir o no estar definida
2. Si la conjunta está concentrada en una curva, es decir, para cada valor de  $x$ , la condicionada es una degenerada, entonces esa será la línea general de regresión y el error cuadrático medio será 0.
3. El caso de mayor error es el caso de independencia, en que se tiene  $m_{Y/X=x}(x) = E(Y)$  y  $ECM = V(Y)$
4. Entonces se puede poner  $ECM(m_{Y/X}) = V(Y)(1 - r_{Y/X}^2)$  con  $r_{Y/X}^2 \in [0,1]$  que mide el grado de dependencia funcional de  $Y$  sobre  $X$ . Siendo  $V(Y)$  la varianza total se puede ver

$$V(Y) = ECM(m_{Y/X}) + V(Y)r_{Y/X}^2$$

var. total = var. residual + var. explicada por la regresion

Así la línea general de regresión minimiza la varianza residual, y se puede obtener

$$r_{Y/X}^2 = \frac{\text{var explicada}}{\text{var total}} = \frac{V(Y) - ECM}{V(Y)}$$

Todo es aplicable a  $X$  sobre  $Y$ , pero no serán los mismos valores de error ni de bondad de la regresión.

## VII.4.2 Recta de regresión

La línea general de regresión no tiene por qué ser operativa funcionalmente, entonces se suele plantear minimizar el error cuadrático medio restringiéndonos a una familia de funciones, y la más normal es la más sencilla, las rectas.

Definición. Dada una variable aleatoria  $(X,Y)$  se llama recta de regresión de  $Y$  sobre  $X$  a la función  $y=ax+b$  que minimiza el error cuadrático medio de entre todas las rectas, y que resulta

ser  $y - E(Y) = \frac{\text{cov}(X, Y)}{V(X)}(x - E(X))$ , supuesto que existen todos los momentos y  $V(X) > 0$ .

En efecto, comprobemos que minimiza el error:

$$\begin{aligned}
ECM(y = ax + b) &= \int_{\mathbb{R}^2} (y - (ax + b))^2 dF(x, y) = \\
&= \int_{\mathbb{R}^2} (y^2 + a^2x^2 + b^2 - 2axy - 2by + 2abx) dF(x, y) = \\
&= E(Y^2) + a^2E(X^2) + b^2 - 2aE(XY) - 2bE(Y) + 2abE(X)
\end{aligned}$$

Derivando e igualando a cero:

$$\left. \begin{aligned}
\frac{\partial}{\partial a} ECM(y = ax + b) &= 2aE(X^2) - 2E(XY) + 2bE(X) = 0 \\
\frac{\partial}{\partial b} ECM(y = ax + b) &= 2b - 2E(Y) + 2aE(X) = 0
\end{aligned} \right\} \Rightarrow \begin{aligned}
\hat{a} &= \frac{\text{cov}(X, Y)}{V(X)} \\
\hat{b} &= E(Y) - \frac{\text{cov}(X, Y)}{V(X)} E(X)
\end{aligned}$$

Nota: Si  $V(X)=0$  entonces  $X$  es degenerada, y no da ninguna información sobre  $Y$

Definición. *Coficiente de correlación lineal*:  $\rho = \frac{\text{cov}(X, Y)}{\sqrt{V(X)V(Y)}}$ , mide el grado de relación

lineal.

Observaciones.

- La covarianza tiene el inconveniente de que es una medida dimensional que no nos permite cuantificar la correlación entre variables de una forma directa. Para obtener una medida adimensional de dicha correlación se utiliza el *coeficiente de correlación lineal*.
- $ECM(y = \hat{a}x + \hat{b}) = V(Y)(1 - \rho^2) \Rightarrow \rho^2 \in [0, 1]$  mide el grado de dependencia lineal entre ambas variables, y se llama *coeficiente de determinación*.

$$\text{Varianza total} = V(Y)$$

$$\text{Varianza residual} = ECM = V(Y)(1 - \rho^2)$$

$$\text{Varianza explicada} = V(Y) \rho^2$$

$\rho^2$  es simétrico respecto a  $X$  e  $Y$ , es decir, es único (vale para ambas rectas, la de  $Y$  sobre  $X$  y la de  $X$  sobre  $Y$ ).

- c)  $\rho \in [-1,1]$ , como puede observarse lleva el signo de la covarianza, y su interpretación es por lo tanto la misma. Es decir, si es positivo es correlación positiva (crecen simultáneamente), si es negativo la correlación es negativa (cuando una crece la otra decrece).
- d)  $\rho^2 = 1$  implica que la distribución se concentra en una recta y ambas rectas de regresión coinciden
- e) En general,  $0 \leq \rho^2 \leq r_{Y/X}^2$ , es decir, ajusta mejor la línea general que la recta, pudiendo incluso ser 0 el coeficiente de determinación lineal, y 1 el otro.
- f)  $\rho^2 = 1 \Rightarrow r_{Y/X}^2 = 1, r_{X/Y}^2 = 0 \Rightarrow \rho^2 = 0$
- g) Las dos rectas de regresión (la de Y sobre X y la de X sobre Y) se cortan en al menos un punto:  $(E(X), E(Y))$ .
- h) La recta de regresión está definida en todo punto (incluso fuera del rango de la variable independiente aunque no tenga sentido). La línea general de regresión no, sólo donde existe la condicionada
- i) Cada recta de regresión minimiza un error diferente (no se minimiza la distancia a la recta sino por variables), por lo tanto es absurdo usar una recta “al revés” (la de Y sobre X para obtener valores de X a partir de Y, por ejemplo).
- j) Es importante observar que NO es un ajuste a una nube de puntos, que minimizaría la distancia de un punto a una recta, ya que sólo se observan los errores en una variable.
- k) Dos variables se dicen *incorreladas* si el coeficiente de correlación lineal (o el de determinación o la covarianza) es cero. Si dos variables son independientes entonces son incorreladas, al revés no es siempre cierto, excepto en distribuciones normales.



# VIII Distribuciones notables multidimensionales

## VIII.1 Multinomial

La multinomial  $(n; p_1, \dots, p_k)$  es una extensión de la binomial cuando no hay sólo dos posibles resultado de cada experimento, sino que hay  $k+1$ .

Sean  $n$  pruebas independientes de un experimento con  $k+1$  posibles resultados  $(E_1, \dots, E_k, F)$  con probabilidades  $P(E_j) = p_j \geq 0$ , y  $\sum_{j=1}^k p_j \leq 1$ ,  $q = P(F) = 1 - \sum_{j=1}^k p_j$ .

Se define la distribución multinomial  $M(n; p_1, \dots, p_k)$ ,  $n \in \mathbb{N}$ ,  $p_j \geq 0$ ,  $\sum_{j=1}^k p_j \leq 1$  como la variable  $X = (X_1, \dots, X_k) : (S, \wp(S), P) \rightarrow (\mathbb{R}^k, B^k)$ , donde  $S$  son las cadenas de  $E_j, F$  como la variable en que cada coordenada mide  $X_j$ : numero de veces que ocurre el suceso tipo  $E_j$  siendo la función de masa de la variable aleatoria conjunta:

$$P(x_1, \dots, x_k) = \frac{n!}{x_1! \dots x_k! (n - \sum_{j=1}^k x_j)!} q^{n - \sum_{j=1}^k x_j} \prod_{j=1}^k p_j^{x_j} \quad \forall x_1, \dots, x_k \in \{0, 1, \dots, n\} \quad \sum_{j=1}^k x_j \leq n$$

- Función característica:  $\varphi(t_1, \dots, t_k) = \left( \sum_{j=1}^k p_j e^{it_j} + q \right)^n$
- Cada marginal unidimensional es una binomial,  $X_j \stackrel{d}{=} B(n, p_j)$ , y el resto de las marginales son multinomiales
- $E(X_j) = np_j \quad V(X_j) = np_j(1 - p_j) \quad \text{cov}(X_j, X_i) = -np_j p_i$

## VIII.2 Normal multivariante

Definición. Se define la distribución  $N(\mu, \Sigma)$  donde  $\mu \in \mathbb{R}^n$ ,  $\Sigma$  matriz real nxn simétrica y definida positiva, como la que tiene por función de densidad

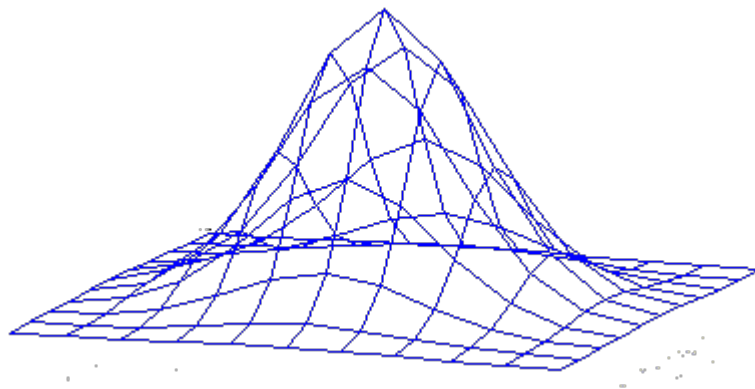
$$f_X(x_1, \dots, x_n) = \frac{1}{(\sqrt{2\pi})^n \sqrt{|\Sigma|}} e^{-\frac{1}{2}(x_1 - \mu_1, \dots, x_n - \mu_n) \Sigma^{-1} \begin{pmatrix} x_1 - \mu_1 \\ \vdots \\ x_n - \mu_n \end{pmatrix}} \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n$$

- $\varphi(t_1, \dots, t_n) = e^{i(t_1, \dots, t_n) \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} - \frac{1}{2}(t_1, \dots, t_n) \Sigma \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix}}$
- $E(X_j) = \mu_j$ , es decir, ese vector es el vector de medias
- $\Sigma$  es la matriz de varianzas-covarianzas
- Todas las marginales son normales multivariantes. Es fácil comprobar que el recíproco no es cierto en general (aunque todas las marginales sean normales, la conjunta puede no serlo), salvo que las variables sean independientes, que sí es cierto.
- Todas las condicionadas son normales multivariantes.
- Las marginales unidimensionales son independientes si y solo si son incorreladas dos a dos, en cuyo caso obsérvese que la matriz de varianzas-covarianzas es diagonal, y con ello se obtiene la independencia.
- Toda combinación lineal de las marginales unidimensionales es normal unidimensional:

$$Y = \sum_{j=1}^n a_j X_j + b = (a_1, \dots, a_n) \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix} + b = N((a_1, \dots, a_n) \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} + b, (a_1, \dots, a_n) \Sigma \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix})$$

- Una variable aleatoria n-dimensional es normal n-variante si y solo si toda combinación lineal de sus marginales es normal univariante.
- Para n=2, se puede expresar la función de densidad en función del coeficiente de correlación lineal:

$$f_{(x,y)}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]} \quad \forall (x, y) \in \mathbb{R}^2$$



## VIII.3 Distribuciones asociadas al muestreo en poblaciones normales

Son distribuciones que aparecen en inferencia estadística asociadas al muestreo, en general, en poblaciones normales. En estadística resultan fundamentales.

### VIII.3.1 Distribución Chi-cuadrado ( $\chi^2$ )

La distribución *Chi-cuadrado con  $n$  grados de libertad* surge como la suma de los cuadrados de  $n$  variables aleatorias normales estándar independientes. Es decir, sea  $Z_1, \dots, Z_n$  variables aleatorias independientes idénticamente distribuidas según una  $N(0,1)$ , entonces  $X = Z_1^2 + \dots + Z_n^2$  se distribuye según una Chi-cuadrado con  $n$  grados de libertad.

Aparece a menudo en inferencia estadística y su función de distribución está tabulada. La tabla se muestra al final de la sección.

A partir de la función característica del cuadrado de una normal estándar se deduce que ésta sigue una distribución Gamma de parámetros  $1/2$  y  $1/2$ , y dado que es reproductiva respecto al parámetro  $p$ , se tiene que  $\chi_n^2 \stackrel{d}{=} \gamma(p = n/2, 1/2)$ .

De ahí se deducen varias cosas:

$$E(\chi_n^2) = n, \quad V(\chi_n^2) = 2n, \quad \varphi(t) = \left(1 - \frac{it}{1/2}\right)^{-n/2}, \quad f(x) = \frac{(1/2)^{n/2}}{\Gamma(n/2)} x^{n/2-1} e^{-x/2} \quad \forall x > 0$$

**Función de distribución Chi-cuadrado:**  $x / P(\chi_n^2 < x) = \alpha$

<b>n/α</b>	<b>0,005</b>	<b>0,01</b>	<b>0,025</b>	<b>0,05</b>	<b>0,1</b>	<b>0,9</b>	<b>0,95</b>	<b>0,975</b>	<b>0,99</b>	<b>0,995</b>
1	0,000	0,000	0,001	0,004	0,016	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	0,211	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	0,584	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	1,064	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	1,610	9,236	11,070	12,833	15,086	16,750
6	0,676	0,872	1,237	1,635	2,204	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	2,833	12,017	14,067	16,013	18,475	20,278
8	1,344	1,646	2,180	2,733	3,490	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	4,168	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	4,865	15,987	18,307	20,483	23,209	25,188
11	2,603	3,053	3,816	4,575	5,578	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	6,304	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	7,042	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	7,790	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	8,547	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	9,312	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	10,085	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	10,865	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	11,651	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	12,443	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	13,240	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	14,041	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	14,848	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	15,659	33,196	36,415	39,364	42,980	45,559
25	10,520	11,524	13,120	14,611	16,473	34,382	37,652	40,646	44,314	46,928
26	11,160	12,198	13,844	15,379	17,292	35,563	38,885	41,923	45,642	48,290
27	11,808	12,879	14,573	16,151	18,114	36,741	40,113	43,195	46,963	49,645
28	12,461	13,565	15,308	16,928	18,939	37,916	41,337	44,461	48,278	50,993
29	13,121	14,256	16,047	17,708	19,768	39,087	42,557	45,722	49,588	52,336
30	13,787	14,953	16,791	18,493	20,599	40,256	43,773	46,979	50,892	53,672
40	20,707	22,164	24,433	26,509	29,051	51,805	55,758	59,342	63,691	66,766
50	27,991	29,707	32,357	34,764	37,689	63,167	67,505	71,420	76,154	79,490

60	35,534	37,485	40,482	43,188	46,459	74,397	79,082	83,298	88,379	91,952
70	43,275	45,442	48,758	51,739	55,329	85,527	90,531	95,023	100,425	104,215
80	51,172	53,540	57,153	60,391	64,278	96,578	101,879	106,629	112,329	116,321
90	59,196	61,754	65,647	69,126	73,291	107,565	113,145	118,136	124,116	128,299
100	67,328	70,065	74,222	77,929	82,358	118,498	124,342	129,561	135,807	140,169
150	109,142	112,668	117,985	122,692	128,275	172,581	179,581	185,800	193,208	198,360
200	152,241	156,432	162,728	168,279	174,835	226,021	233,994	241,058	249,445	255,264
300	240,663	245,972	253,912	260,878	269,068	331,789	341,395	349,874	359,906	366,844
400	330,903	337,155	346,482	354,641	364,207	436,649	447,632	457,305	468,724	476,606
500	422,303	429,388	439,936	449,147	459,926	540,930	553,127	563,852	576,493	585,207

### VIII.3.2 Distribución t de Student

Sea  $Z$  una variable normal estándar y  $X$  una variable independiente con distribución  $\chi_n^2$ .

$T = \frac{Z}{\sqrt{X/n}}$  sigue una distribución *t de Student con n grados de libertad*

Es simétrica respecto a 0 y de aspecto muy parecido a la normal estándar. Para valores grandes de  $n$  se puede sustituir por la normal estándar. Además está tabulada:

**Función de distribución t de Student:**  $x / P(t_n < x) = \alpha$

n / alfa	0,9	0,95	0,975	0,99	0,995
1	3,078	6,314	12,706	31,821	63,657
2	1,886	2,920	4,303	6,965	9,925
3	1,638	2,353	3,182	4,541	5,841
4	1,533	2,132	2,776	3,747	4,604
5	1,476	2,015	2,571	3,365	4,032
6	1,440	1,943	2,447	3,143	3,707
7	1,415	1,895	2,365	2,998	3,499
8	1,397	1,860	2,306	2,896	3,355
9	1,383	1,833	2,262	2,821	3,250
10	1,372	1,812	2,228	2,764	3,169
11	1,363	1,796	2,201	2,718	3,106
12	1,356	1,782	2,179	2,681	3,055
13	1,350	1,771	2,160	2,650	3,012
14	1,345	1,761	2,145	2,624	2,977
15	1,341	1,753	2,131	2,602	2,947
16	1,337	1,746	2,120	2,583	2,921

17	1,333	1,740	2,110	2,567	2,898
18	1,330	1,734	2,101	2,552	2,878
19	1,328	1,729	2,093	2,539	2,861
20	1,325	1,725	2,086	2,528	2,845
21	1,323	1,721	2,080	2,518	2,831
22	1,321	1,717	2,074	2,508	2,819
23	1,319	1,714	2,069	2,500	2,807
24	1,318	1,711	2,064	2,492	2,797
25	1,316	1,708	2,060	2,485	2,787
26	1,315	1,706	2,056	2,479	2,779
27	1,314	1,703	2,052	2,473	2,771
28	1,313	1,701	2,048	2,467	2,763
29	1,311	1,699	2,045	2,462	2,756
30	1,310	1,697	2,042	2,457	2,750
40	1,303	1,684	2,021	2,423	2,704
50	1,299	1,676	2,009	2,403	2,678
60	1,296	1,671	2,000	2,390	2,660
70	1,294	1,667	1,994	2,381	2,648
80	1,292	1,664	1,990	2,374	2,639
90	1,291	1,662	1,987	2,368	2,632
100	1,290	1,660	1,984	2,364	2,626
∞	1,280	1,645	1,960	2,330	2,575

### VIII.3.3 F de Fisher-Snedecor ( $F_{m,n}$ ):

Sean  $Z_1, \dots, Z_m, Y_1, \dots, Y_n$  variables aleatorias independientes normales  $N(0,1)$ , se define como

$$\frac{\sum_{i=1}^m Z_i^2 / m}{\sum_{j=1}^n Y_j^2 / n} = \frac{\chi_m^2 / m}{\chi_n^2 / n} = F_{m,n}$$

También está tabulada pero al ser dos parámetros se requieren varias tablas.

# IX Sucesiones de variables aleatorias

Las sucesiones de variables aleatorias surgen muy a menudo en estadística para estudiar propiedades asintóticas de los estimadores. Además hay un conjunto de resultados como las leyes de los grandes números y las del problema central del límite que son de gran relevancia en el uso práctico.

## IX.1 Convergencias de sucesiones de variables aleatorias

### IX.1.1 Definiciones de convergencias de sucesiones

#### IX.1.1.1 Convergencia casi seguro

Se dice que una sucesión de variables aleatorias definidas en un mismo espacio de probabilidad  $X_n : (S, Q, P) \rightarrow (\mathbb{R}, B)$  converge casi seguro (se denota c.s.) a la variable aleatoria  $X : (S, Q, P) \rightarrow (\mathbb{R}, B)$ , cuando  $P(s \in S / \lim_{n \rightarrow \infty} X_n(s) = X(s)) = 1$ .

Es decir, que el límite puntual existe y coincide con  $X(s)$ , casi seguro.

#### IX.1.1.2 Convergencia en media de orden $p$ ( $p \geq 1$ )

Se dice que una sucesión de variables aleatorias definidas en un mismo espacio de probabilidad  $X_n : (S, Q, P) \rightarrow (\mathbb{R}, B)$  converge en media de orden  $p$  (se denota por  $m^p$ ) a la variable aleatoria  $X : (S, Q, P) \rightarrow (\mathbb{R}, B)$ , cuando  $\lim_{n \rightarrow \infty} E \left[ |X_n - X|^p \right] = 0$ , supuesto que tanto  $|X_n|^p$  como  $|X|^p$  son integrables.

La más usual es la convergencia en media cuadrática, es decir,  $p=2$ .

### IX.1.1.3 Convergencia en probabilidad

Se dice que una sucesión de variables aleatorias definidas en un mismo espacio de probabilidad  $X_n : (S, Q, P) \rightarrow (\mathbb{R}, B)$  converge en Probabilidad a la variable aleatoria  $X : (S, Q, P) \rightarrow (\mathbb{R}, B)$ , cuando  $\lim_{n \rightarrow \infty} P(s \in S / |X_n(s) - X(s)| \geq \varepsilon) = 0 \quad \forall \varepsilon > 0$ .

Obsérvese que esta convergencia es el límite de una probabilidad, mientras que la casi seguro es la probabilidad de un límite.

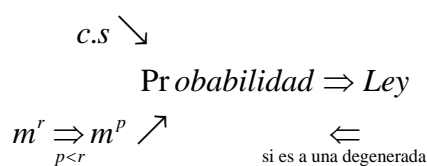
### IX.1.1.4 Convergencia en Ley o débil

Se dice que una sucesión de variables aleatorias definidas en un mismo espacio de probabilidad  $X_n : (S, Q, P) \rightarrow (\mathbb{R}, B)$  converge en Ley a la variable aleatoria  $X : (S, Q, P) \rightarrow (\mathbb{R}, B)$ , cuando  $\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x) \quad \forall x \in \mathbb{R}$  de continuidad de  $F_X$ .

Es la convergencia de las probabilidades inducidas, o de sus funciones de distribución, que ya se vio en el capítulo de espacio de probabilidad.

La convergencia en Ley no es única, también puede converger a otra variable, con tal de que la distribución sea la misma.

## IX.1.2 Propiedades y relaciones entre convergencias



## IX.2 Leyes de los grandes números

Son una serie de resultados que dan condiciones para que se dé la convergencia de las sucesiones de promedios a una degenerada. Las leyes débiles tratan la convergencia en probabilidad, y las leyes fuertes la convergencia casi seguro.



En concreto, dada una sucesión de variables aleatorias independientes definidas en un mismo espacio de probabilidad  $X_n : (S, Q, P) \rightarrow (\mathbb{R}, B)$ , se analiza la convergencia de

$$\frac{1}{n} \sum_{j=1}^n X_j - \frac{1}{n} \sum_{j=1}^n E(X_j) \rightarrow 0$$

## IX.2.1 Leyes débiles de los grandes números

Al ser una degenerada el límite, y la convergencia en probabilidad, es equivalente a estudiar la convergencia en Ley.

Teorema (condición suficiente). Sea  $(X_n)_n$  una sucesión de variables aleatorias independientes o al menos incorreladas dos a dos con  $V(X_n) = \sigma_n^2 < \infty \quad \forall n$  tal que

$$\lim_{n \rightarrow \infty} \frac{1}{n^2} \sum_{k=1}^n \sigma_k^2 = 0 \Rightarrow L.D.G.N$$

Corolarios del teorema anterior:

1.  $(X_n)_n$  variables aleatorias independientes con  $V(X_n) \leq M < \infty \quad \forall n \Rightarrow L.D.G.N.$
2.  $(X_n)_n$  variables aleatorias independientes con varianza finita, entonces  $L.D.G.N.$ , es decir, la sucesión de los promedios muestrales converge en probabilidad a la media poblacional. Por tanto, se puede estimar el valor de la media poblacional mediante el promedio muestral para  $n$  suficientemente grande.
3.  $(X_n)_n$  variables aleatorias independientes Bern(p), entonces.  $L.D.G.N.$  Éste es el Teorema de Bernoulli, históricamente la primera ley débil de los grandes números:

$$Y_n = \sum_{k=1}^n X_k \stackrel{d}{=} B(n, p), \text{ entonces } \hat{p} = \frac{Y_n}{n} \xrightarrow{P} p$$

Teorema de Khintchine. Sea  $(X_n)_n$  sucesión de variables aleatorias independientes idénticamente distribuidas con  $E(X_n) < \infty \quad \forall n \Rightarrow L.D.G.N.$

## IX.2.2 Leyes fuertes de los grandes números

Estudian la convergencia casi seguro.

Condición suficiente:  $(X_n)_n$  variables aleatorias con  $\sum_{n=1}^{\infty} \frac{\sigma_n^2}{n^2} < \infty \Rightarrow L.F.G.N.$

La primera ley fuerte de los grandes números en la historia fue el Teorema de Borel: Sean  $(X_n)_n$  variables aleatorias independientes idénticamente distribuidas Bern(p), entonces *L.F.G.N.*

Ese teorema viene a decir que las frecuencias muestrales convergen a la media.

Teorema de Kolmogorov. Sea  $(X_n)_n$  sucesión de variables aleatorias independientes idénticamente distribuidas con  $E(X_n) = \mu < \infty \Rightarrow L.F.G.N.$  a  $\mu$ .

## IX.3 Problema central del límite

El problema central del límite analiza la convergencia en Ley de las sumas de variables

aleatorias, en general, de la forma 
$$\frac{\sum_{j=1}^n X_j - E(\sum_{j=1}^n X_j)}{\sqrt{V(\sum_{j=1}^n X_j)}}$$

Teorema de Levy-Lindeberg. Sean  $(X_n)_n$  variables aleatorias independientes idénticamente

distribuidas con  $E(X_n) = \mu < \infty, V(X_n) = \sigma^2 < \infty, \sigma^2 \neq 0 \Rightarrow \frac{\sum_{j=1}^n X_j - n\mu}{\sigma\sqrt{n}} \xrightarrow{Ley} N(0,1)$

Teorema de Moivre (corolario del anterior). Sean  $(X_n)_n$  variables aleatorias independientes

con  $X_n \stackrel{d}{=} B(n, p) \Rightarrow \frac{X_n - np}{\sqrt{npq}} \xrightarrow{Ley} N(0,1)$

Teorema de Lindeberg. Sean  $(X_n)_n$  variables aleatorias independientes con  $E(X_n) = \mu_n < \infty, V(X_n) = \sigma_n^2 < \infty$ , si para  $S_n^2 = \sum_{k=1}^n V(X_k)$  se tiene

$$\lim_{n \rightarrow \infty} \frac{1}{S_n^2} \sum_{k=1}^n \int_{\{x \in \mathbb{R} / |x - \mu_k| \geq \varepsilon S_n\}} (x - E(X_k))^2 dF_{X_k}(x) = 0 \quad \forall \varepsilon > 0$$

entonces

$$\frac{\sum_{j=1}^n X_j - E\left(\sum_{j=1}^n X_j\right)}{\sqrt{V\left(\sum_{j=1}^n X_j\right)}} \xrightarrow{Ley} N(0,1)$$

Obsérvese que un dividiendo por n también hablaríamos de promedios, en las leyes de los grandes números es para permitirnos aproximar promedios por un número y aquí por una distribución no degenerada, lo que permitirá acotar la probabilidad de error en la aproximación, así como identificar el tamaño que debe tener una muestra.