

Estadística y Procesos Estocásticos

Apuntes de clase

Javier Rodrigo López ¹

3 de febrero de 2021



¹E-mail: javiolonchelo@gmail.com



UNIVERSIDAD
POLITÉCNICA
DE MADRID



Introducción

Esta pequeña recopilación de fórmulas, teoremas y demás apuntes de teoría ha sido elaborada durante el primer semestre del curso 2019-2020, en la escuela [ETSIST](#) de la [UPM](#) por Javier Rodrigo López, alumno de 1º de Ingeniería de Sonido e Imagen. bruuh

Índice general

Introducción	2
1. Probabilidad	7
1.1. Espacio probabilístico	7
1.2. Combinatoria	8
1.3. Probabilidad condicionada. Independencia.	9
2. Variables aleatorias	11
2.1. Variable aleatoria discreta	11
2.1.1. Función de probabilidad	11
2.1.2. Media	11
2.1.3. Varianza y desviación típica	11
2.1.4. Función de distribución	11
2.1.5. Distribuciones	12
2.2. Variable aleatoria continua	13
2.2.1. Media y varianza	13
2.2.2. Función de distribución	13
2.2.3. Distribuciones	14
2.3. Desigualdad de Chebyshev	15
2.4. Cuantil	15
2.5. Percentil	15
3. Vectores aleatorios	17
3.1. Variable aleatoria bidimensional	17
3.2. Variable aleatoria bidimensional discreta	17
3.3. Variable aleatoria bidimensional continua	17
3.4. Independencia de variables aleatorias	18
3.5. Momentos	18
3.5.1. Esperanza	18
3.5.2. Covarianza	18
3.5.3. Notación matricial	19
3.6. Variable aleatoria multidimensional	19
3.7. Distribución multinomial	20
3.8. Distribución normal n-dimensional	20
3.9. Teorema central del límite	21
3.10. Chi cuadrado y la t de Student	21
4. Inferencia estadística	23
4.1. Estadística descriptiva de una variable: momentos, cuantiles, box-plot, histograma, función de distribución empírica y cálculo de proporciones	23
4.2. Muestra aleatoria. Media muestral y varianza muestral. Estimación paramétrica	23
4.3. Intervalos de confianza para la media y para proporciones poblacionales	23
4.4. Contraste de hipótesis. Nivel de significación y p-valor	23
5. Procesos estocásticos	25
5.1. Definición de proceso estocástico	25
5.2. Procesos estocásticos en tiempo continuo	25
5.3. Procesos estocásticos en tiempo discreto	26
5.4. Distribuciones de primer y segundo orden, media, autocorrelación y autocovarianza	26
5.5. Proceso de Bernoulli. Caminos aleatorios. Procesos normales. Proceso de Poisson	26
5.6. Procesos estacionarios. Densidad espectral	26

5.7. Sistemas lineales y procesos estocásticos	26
6. Prácticas con software estadístico	27
6.1. Modelos de distribución de probabilidad más comunes	27
6.2. Estadística descriptiva	27
6.3. Muestreo. Estimación puntual	27
6.4. Estimación por intervalos de confianza	27
6.5. Contraste paramétrico	27

Conceptos básicos de teoría de conjuntos

Notación básica

Definición 1. Un **conjunto** es una colección de elementos considerada como una sola entidad. Los conjuntos se denotan con letras mayúsculas, $A, B, C \dots$, mientras que los elementos se denotan con letras minúsculas, $a, b, c \dots$.

Utilizamos la notación $x \in A$ para indicar que x es un elemento de A . Si, por el contrario, x no pertenece a A , entonces se escribe como $x \notin A$.

Si todos los elementos de un conjunto A pertenecen a su vez a un conjunto B , entonces se dice que A es subconjunto de B y se denota como $A \subset B$.

Si A y B tienen exactamente los mismos elementos, entonces $A = B$.

Operaciones con conjuntos

Definición 2. La **unión** de A y B se denota como $A \cup B$ y contiene a todos los elementos de A y todos los elementos de B .

Definición 3. La **intersección** de A y B se denota como $A \cap B$ y contiene solo a los elementos que se encuentran en A y en B al mismo tiempo. Si no comparten elementos, se dice que son **disjuntos**. La intersección de estos es el conjunto vacío, denotado como \emptyset . De esta forma, $A \cap B = \emptyset$.

Definición 4. El **complementario** de un conjunto A contiene todos los elementos que no se encuentran en A . Se denota como \overline{A} y se cumple que:

$$A \cup \overline{A} = \Omega \quad A \cap \overline{A} = \emptyset$$

Definición 5. La **diferencia** de dos conjuntos A, B se define como el conjunto de los elementos de A que no pertenecen a B . Se denota como $A - B$, aunque también se puede escribir como $A \cap \overline{B}$.

Propiedades

Estas operaciones verifican las propiedades conmutativas, asociativas y distributivas, además de las siguientes propiedades:

$$A \cup A = A \cap A = A$$

$$A \cup \Omega = \Omega$$

$$A \cap \Omega = A$$

También podemos incluir las **Leyes de De Morgan**:

$$\overline{A \cup B} = \overline{A} \cap \overline{B}$$

$$\overline{A \cap B} = \overline{A} \cup \overline{B}$$

Estas leyes son muy importantes y tienen una propiedad interesante. No hace falta que sean una pareja de conjuntos, también puede suceder algo como lo siguiente:

$$\overline{A \cap B \cap C} = \overline{A} \cup \overline{B} \cup \overline{C}$$

Javier Rodrigo López

Capítulo 1

Probabilidad

1.1 Espacio probabilístico

Definición 6. Un experimento es **aleatorio** si se verifica que:

- Todos los resultados se conocen de antemano.
- Cualquier realización da lugar a un resultado que no se conoce previamente.
- Puede repetirse en condiciones idéntica.

Definición 7. El **espacio muestral** Ω asociado a un experimento es el conjunto de todos los posibles resultados del experimento.

Un suceso A es un subconjunto del espacio muestral Ω . Los sucesos que constan de un único elemento se denominan **sucesos elementales**. El espacio de sucesos S es el conjunto de los subconjuntos de Ω .

Definición 8. La **frecuencia relativa** de un suceso A se define como

$$f_r(A) = \frac{n_A}{n} \quad \begin{cases} n & \equiv \text{número de veces que se realiza el experimento} \\ n_A & \equiv \text{número de veces que sucede } A \end{cases}$$

La **ley de regularidad estadística** dice que la frecuencia relativa tiende a un número fijo según se incrementa el valor de n .

Propiedades de la frecuencia relativa

- $0 \leq f_r(A) \leq 1$
- $f_r(\Omega) = 1$
- $f_r(A \cup B) = f_r(A) + f_r(B)$ solo si A y B son incompatibles ($A \cap B = \emptyset$)

Definición 9. Una **probabilidad** P sobre (Ω, S) es una función $P : S \rightarrow [0, 1]$ que verifica:

- $P(\Omega) = 1$
- Para toda colección de sucesos A_1, A_2, \dots, A_n incompatibles dos a dos ($A_i \cap A_j = \emptyset \ \forall i \neq j$), se verifica que:

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

Consecuencias de esta definición

- $P(\bar{A}) = 1 - P(A)$
- $P(\emptyset) = 0$
- Si $A \subset B \implies P(A) \leq P(B)$
- $P(A - B) = P(A \cap \bar{B}) = P(A) - P(A \cap B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

NOTA: Esta última puede interpretarse de forma diferente para tres sucesos:

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

En general, para un número n de sucesos:

$$P(A_1 \cup \dots \cup A_n) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n+1} P(A_1 \cap \dots \cap A_n)$$

Definición 10. La terna (Ω, S, P) se llama **espacio probabilístico**.

Probabilidad en espacios muestrales finitos

Si Ω es finito, cada conjunto con un elemento $\{\omega_j\}$, $j = 1, 2, \dots, n$, es un suceso elemental y es suficiente asignar probabilidades a cada $\{\omega_j\}$. Si $A \in S$, se tiene:

$$P(A) = \sum_{\omega_j \in A} P(\{\omega_j\})$$

Si particularizamos podemos obtener la **Regla de Laplace**¹.

Si $P(\{\omega_j\}) = \frac{1}{n}$, $j = 1, 2, \dots, n$; tenemos que:

$$P(A) = \frac{\text{Nº de elementos de } A}{\text{Nº de elementos de } \Omega}$$

1.2 Combinatoria

Tenemos que tener en cuenta lo que significan los **elementos** y el **orden** en el que se encuentran dentro del conjunto.

Si no importa el orden, estamos hablando de **combinaciones**:

- Combinaciones de m elementos tomados de n en $n(n \leq m)$:

$$C_{m,n} \cdot P_n = V_{m,n} \quad \Rightarrow \quad C_{m,n} = \binom{m}{n} = \frac{m!}{(m-n)!n!}$$

¹Para poder utilizar la Regla de Laplace, necesitamos que los sucesos sean equiprobables.

Ahora bien, si el orden importa tendremos que distinguir otras dos situaciones. Si cogemos todos los elementos del conjunto, hablaremos de **permutaciones**.

- Permutaciones de m elementos:

$$P_m = m!$$

- Permutaciones con repetición de m elementos. Es decir, que hay un elemento que se repite r_1 veces, otro que se repite r_2 y demás:

$$\sum_{j=1}^i r_j = m \quad P_m^{r_1, r_2, \dots, r_i} = \frac{m!}{r_1! r_2! \dots r_i!}$$

Si, por el contrario, los juntamos en grupos más pequeños, estaremos hablando de **variaciones**.

- Variaciones con repetición de m elementos tomados de n en n :

$$VR_{m,n} = m^n$$

- Variaciones sin repetición de m elementos tomados de n en n ($n \leq m$):

$$V_{m,n} = \frac{m!}{(m-n)!}$$

1.3 Probabilidad condicionada. Independencia.

Definición 11. Sea A un suceso de probabilidad $P(A) > 0$, y B otro suceso de probabilidad $P(B)$. La probabilidad de B puede dejar de ser la misma, si se sabe que ha ocurrido el suceso A .

Sean A y B dos sucesos tal que $P(A) > 0$. Llamaremos **probabilidad condicionada** del suceso B respecto al suceso A , y la denotaremos por $P(B | A)$, al cociente:

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

La fórmula de la probabilidad condicionada también se puede expresar como:

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

Que se podría generalizar a tres sucesos A, B y C , resultando en la **regla de la multiplicación**:

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

olololo

Definición 12. Si $P(B | A) = P(B)$ siendo $P(A) > 0$, entonces que haya ocurrido A no influye en la probabilidad que tenía B de ocurrir, y se dice que A y B son independientes.

Despejando:

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

Definición 13. Diremos que los sucesos A_1, A_2, \dots, A_n son mutuamente independientes si la probabilidad conjunta de todos los subconjuntos que pueden formarse con ellos es el producto de las probabilidades individuales. Es decir:

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i) \quad \forall I[1, 2, \dots, n]$$

Teorema 1. *El teorema de la probabilidad total nos dice que, sean los sucesos C_1, \dots, C_n incompatibles dos a dos de modo que $\bigcup_{i=1}^n C_i = \Omega$, entonces:*

$$P(A) = \sum_{i=1}^n P(A | C_i) P(C_i) \quad \forall A \in \mathcal{S}$$

Teorema 2. *El teorema de Bayes nos dice, en las mismas hipótesis que el teorema anterior, que:*

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

Ejemplo 6 del documento anexo (falta referencia).

$$P(B) = 3$$

Capítulo 2

Variables aleatorias

2.1 Variable aleatoria discreta

Se dice que una variable aleatoria es **discreta** si solo puede tomar un número finito de valores, o bien un número infinito numerable (como pueden ser los números naturales \mathbb{N}).

2.1.1. Función de probabilidad

Si la v.a.¹ X puede tomar unos ciertos valores x_n con $n = 1, 2, \dots$

La **función de probabilidad** es una función que asigna a cada valor x_n una probabilidad.

2.1.2. Media

La **media** o **esperanza matemática** de una v.a. X nos proporciona información sobre la localización de X . Se representa como μ_X ² o como $E[X]$.

$$\mu_X = E[X] = \sum_{n=1}^{\infty} x_n \cdot P(X = x_n)$$

NOTA: Es importante mencionar que la esperanza es un operador lineal.

2.1.3. Varianza y desviación típica

La **varianza** de una v.a. X nos proporciona información sobre la dispersión de X . Se representa como $\text{Var}(X)$.

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

A veces es más útil representar la **desviación típica**. Se denota por la letra σ .

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

2.1.4. Función de distribución

Sea F la **función de distribución** (también conocida como función de distribución acumulativa) de la v.a. X . Las propiedades que caracterizan a la función de distribución son:

1. Es monótona decreciente, es decir: $\forall x, y \in \mathbb{R}, x < y \Rightarrow F(x) \leq F(y)$
2. $\lim_{x \rightarrow -\infty} F(x) = 0$
3. $\lim_{x \rightarrow \infty} F(x) = 1$
4. F es continua por la derecha en \mathbb{R} . Es decir: $\lim_{x \rightarrow a^+} F(x) = F(a), \quad \forall a \in \mathbb{R}$

¹v.a. \equiv variable aleatoria

²Es común representar la media como μ a secas.

2.1.5. Distribuciones

Distribución uniforme discreta

Una v.a. X tiene una distribución uniforme discreta si es una v.a. discreta en la que sus valores son todos **equiprobables**.

Pruebas de Bernoulli

Una prueba de Bernoulli es un experimento aleatorio con dos posibles resultados: **éxito** (E) y **fracaso** (\bar{E}). Si se realizan varias pruebas de Bernoulli, los resultados serán independientes entre sí.

La probabilidad del éxito se denota como p , de modo que la probabilidad del fracaso será $(1 - p)$.

Distribución binomial

Generalizando un poco, si N es el número de pruebas de Bernoulli que realizamos, la probabilidad de obtener n éxitos será:

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

Si una v.a. X tiene una distribución binomial, se denota como:

$$X \sim B(N, p)$$

- **Media:** $\mu = Np$
- **Varianza:** $\sigma^2 = Np(1 - p)$

Distribución geométrica

En esta distribución también se parte de las pruebas de Bernoulli. Sin embargo, y a diferencia de la distribución binomial, nos interesará conocer el **número de pruebas hasta el primer éxito**. Una v.a. T tiene distribución geométrica si:

$$P(T = n) = (1 - p)^{n-1} p \quad n = 1, 2, \dots$$

Si una v.a. T tiene una distribución geométrica, se denota como:

$$T \sim \text{Geo}(p)$$

- **Media:** $\mu = \frac{1}{p}$
- **Varianza:** $\sigma^2 = \frac{1 - p}{p^2}$

NOTA: Si lo reescribimos para la probabilidad de fracaso q (recordamos que $q = 1 - p$), entonces:

$$P(T > n) = q^n$$

Distribución binomial negativa

Este tipo de distribución nos interesa conocer el número de pruebas hasta obtener el r -ésimo éxito.

Se dice que una v.a. X tiene una distribución binomial negativa de parámetros r y p , si:

$$P(X = n) = \binom{n-1}{r-1} p^r (1 - p)^{n-r} \quad n = r, (r + 1), (r + 2), \dots$$

Si una v.a. X tiene una distribución binomial negativa, se denota como:

$$X \sim \text{BN}(r, p)$$

- **Media:** $\mu = \frac{r}{p}$
- **Varianza:** $\sigma^2 = \frac{r(1-p)}{p^2}$

Distribución de Poisson

Una v.a. X tiene distribución de Poisson si representa **el número de ciertos eventos que ocurren en un periodo fijo de tiempo** (siempre que los eventos se produzcan de forma aleatoria en el tiempo). Se cumple que:

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

Si una v.a. X tiene una distribución de Poisson, se denota como:

$$X \sim \text{Poisson}(\lambda)$$

- **Media:** $\mu = \lambda$
- **Varianza:** $\sigma^2 = \lambda$

NOTA:

Si N es muy grande, una v.a. que tenga distribución binomial **puede aproximarse como una distribución de Poisson**, cometiendo un error prácticamente despreciable.

La forma realizar esta aproximación es igualando las medias típicas de cada distribución:

$$Np = \lambda$$

2.2 Variable aleatoria continua

Una v.a. X es continua si existe una función $f_X(x)$ llamada **función de densidad** tal que:

$$P(a < X \leq b) = \int_a^b f(x) dx \quad f(x) \geq 0, \quad \begin{cases} a, b \in \mathbb{R} \\ a \leq b \end{cases}$$

NOTA: Como $P(-\infty < X < \infty) = 1$, la función de densidad verifica:

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

2.2.1. Media y varianza

La **media** (μ) de una v.a. X se define de forma análoga a cómo se definía en variables aleatorias discretas:

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

La varianza se define exactamente de la misma forma:

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

2.2.2. Función de distribución

La función de distribución con v.a. continua también se puede definir de forma análoga a como hicimos con v.a. discreta:

$$F(x) = P(X \leq x) = \int_{-\infty}^x f(\xi) d\xi$$

Esta es una función acotada entre 0 y 1 y, además, es monótona creciente.

La función de distribución nos permite calcular **probabilidades de intervalos** de la siguiente forma:

$$\begin{cases} P(a < X < b) = F(b) - F(a) \\ P(X > a) = 1 - F(a) \end{cases}$$

Un resultado interesante haberlo definido así lo encontramos cuando intentamos calcular la probabilidad en un punto.

$$P(X = a) = 0 \quad \forall a \in \mathbb{R}$$

NOTA: En variable continua, podemos relacionar la función de distribución con la función de densidad.

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

2.2.3. Distribuciones

Distribución uniforme

Una v.a. X tiene distribución uniforme en el intervalo (a, b) si su función de densidad es:

$$f(x) = \begin{cases} \frac{1}{b-a} & a < x < b \\ 0 & \text{en el resto de casos} \end{cases}$$



Figura 2.1. Representación de la función de densidad de una v.a. de distribución uniforme.

A veces, es común verla como $f(x) = \frac{1}{b-a}$ para $a < x < b$. Se da por supuesto que, fuera de ese rango, la función es nula.

Para indicar que X tiene una distribución uniforme, lo escribimos como:

$$X \sim U(a, b)$$

¿Cuándo la voy a usar? En algunas situaciones, y en la mayoría de los ejercicios que tendremos que resolver, es muy útil saber que:

$$P(X \in (c, d)) = \frac{d-c}{b-a} \quad \text{siempre y cuando } (c, d) \subset (a, b)$$

- **Media:** $\mu = \frac{a+b}{2}$
- **Varianza:** $\sigma^2 = \frac{(b-a)^2}{12}$

Distribución exponencial

En las v.a. con distribución exponencial, nos interesa el **tiempo que hay que esperar hasta que suceda cierto evento**. Una v.a. T tiene distribución exponencial si su función de densidad es:

$$f(t) = \begin{cases} \lambda e^{-\lambda t} & t > 0 \\ 0 & \text{resto} \end{cases} \quad (\lambda > 0)$$

Para indicar que T tiene una distribución exponencial, lo escribimos como:

$$T \sim \text{Exp}(\lambda)$$

- **Media:** $\mu = \frac{1}{\lambda}$ (tiempo medio de espera)
- **Varianza:** $\sigma^2 = \frac{1}{\lambda^2}$

Distribución normal

Una v.a. X tiene distribución normal si su función de densidad es la **campana de Gauss**:

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

Sin embargo, no es posible calcular su función de distribución en términos generales.

Para indicar que X tiene una distribución normal, caracterizada por su media y su desviación típica, lo escribimos como:

$$X \sim N(\mu, \sigma)$$

Además, se suele utilizar la **distribución normal estándar** $N(0, 1)$, ya que está tabulada. En otras palabras, para realizar los ejercicios utilizaremos una tabla con los valores más utilizados. Sin embargo, en la tabla no aparecen valores negativos de X . Eso se debe a que se puede calcular como $\Phi(-x) = 1 - \Phi(x)$.

Cuando nos encontremos con una v.a. X que se distribuya con una normal, lo que debemos hacer es **tipificarla**. Esto significa que definiremos otra v.a. Z que modelice a X para que tenga distribución normal.

$$X \sim N(\mu, \sigma) \rightarrow Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

2.3 Desigualdad de Chebyshev

Sea X una v.a. con media μ y varianza σ^2 , entonces para todo $k > 0$ se cumple lo siguiente:

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

NOTA: Es más fácil comprender la desigualdad de Chebyshev si la reescribimos de esta forma:

$$P[(\mu - k\sigma) < X < (\mu + k\sigma)] \geq 1 - \frac{1}{k^2}$$

2.4 Cuantil

En v.a. discreta

Sea X una v.a. discreta, llamaremos **cuantil de orden p** , denotado como q_p , al menor valor en el que se cumpla que:

$$F_X(x) \geq p \quad (0 < p < 1)$$

En v.a. continua

Sea T una v.a. continua, llamaremos **cuantil de orden p** , denotado como q_p , al número que verifica:

$$P(T \leq q_p) = \int_{-\infty}^{q_p} f(x)dx = p$$

O, en otras palabras:

$$F_T(q_p) = p$$

2.5 Percentil

El percentil es un tipo de cuantil que separa las muestras en 100 grupos diferentes. Para calcular el **percentil de orden n** de una v.a., basta con saber que:

$$P_n = q_{\frac{n}{100}}$$

Javier Rodrigo López

Capítulo 3

Vectores aleatorios

3.1 Variable aleatoria bidimensional

Una variable aleatoria bidimensional se puede definir como una transformación tal que:

$$(X, Y) : \Omega \longrightarrow \mathbb{R}^2$$

De modo que, teniendo el experimento $\omega \in \Omega$, la variable aleatoria bidimensional asigna a cada resultado de dicho experimento un elemento de \mathbb{R}^2 que es $(X(\omega), Y(\omega))$.

Esta es la **misma definición** que para variables aleatorias unidimensionales, pero extrapolada a dos dimensiones. Se podría decir que una variable aleatoria bidimensional está compuesta de dos variables aleatorias unidimensionales, como en un paquete.

3.2 Variable aleatoria bidimensional discreta

La **función de probabilidad conjunta** de una variable aleatoria bidimensional discreta es una función que proporciona la información esencial de las variables X e Y , y se define como:

$$P(X = x_i, Y = y_j) = P(X = x_i \cap Y = y_j), \quad i, j = 1, 2, \dots$$

A partir de esta función, se pueden calcular las funciones de probabilidad de las variables aleatorias X e Y . En este contexto, se les denomina **funciones de probabilidad marginales**:

$$P(X = x_i) = \sum_{j=1}^{\infty} P(X = x_i, Y = y_j), \quad i = 1, 2, \dots$$

$$P(Y = y_j) = \sum_{i=1}^{\infty} P(X = x_i, Y = y_j), \quad j = 1, 2, \dots$$

Es importante observar que la probabilidad del espacio muestral es 1, por lo que se verifica que:

$$\sum_{i=1}^{\infty} \sum_{j=1}^{\infty} P(X = x_i, Y = y_j) = 1$$

3.3 Variable aleatoria bidimensional continua

La **función de probabilidad conjunta** de una variable aleatoria bidimensional continua es una función que proporciona la información esencial de las variables X e Y , y se define como:

$$P((X, Y) \in D) = \iint_D f_{X,Y}(x, y) \, dy \, dx$$

O, en particular:

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \int_c^d f_{X,Y}(x, y) \, dy \, dx$$

Además:

$$\boxed{F_{X,Y}(x, y) = P(X \leq x, Y \leq y)} = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(\xi, \eta) d\xi d\eta$$

Al igual que antes, la probabilidad del espacio muestral es 1. Por tanto, se verifica que:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{X,Y}(x, y) dx dy = 1$$

Como consecuencia del teorema fundamental del cálculo, se obtiene la siguiente expresión:

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

Las variables aleatorias continuas X e Y tienen funciones de densidad, que en este caso denominaremos **funciones de densidad marginales**

3.4 Independencia de variables aleatorias

Sea (X, Y) una variable aleatoria bidimensional, se dice que X e Y son independientes si para cualquiera A y B :

$$P(X \in A, Y \in B) = P(X \in A) \cdot P(Y \in B)$$

Si X e Y son independientes, las variables aleatorias definidas a continuación también serán independientes:

$$U = g(X) \quad V = h(Y)$$

3.5 Momentos

3.5.1. Esperanza

Sea (X, Y) una variable aleatoria discreta, la **esperanza** de la variable aleatoria $g(X, Y)$ es:

$$E[g(X, Y)] = \sum_{x_i, y_j \in S} g(x_i, y_j) P(X = x_i, Y = y_j)$$

Sea (X, Y) una variable aleatoria continua, y análoga a la definición anterior, la esperanza de la variable aleatoria $g(X, Y)$ es:

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

La esperanza es un operador **lineal**.

3.5.2. Covarianza

La covarianza se usa para **medir el grado de correlación** entre las variables X e Y . Se define como:

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

Sin embargo, la **forma habitual** de calcular la covarianza es la siguiente:

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

Para dos números $a, b \in \mathbb{R}$, se cumple que:

$$\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$$

Esta propiedad puede dificultar algunas operaciones. Por eso, se define el **coeficiente de correlación lineal** mediante la normalización de la covarianza:

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

Es importante recalcar que $\rho(aX, bY) = \rho(X, Y)$. Además, se verifica que:

$$-1 \leq \rho(X, Y) \leq 1$$

Las variables aleatorias X e Y son **incorreladas** si $\rho(X, Y) = 0$.

Si las variables aleatorias X e Y son independientes, son incorreladas. Por lo tanto:

$$X \text{ e } Y \text{ son independientes} \implies \begin{cases} E[X \cdot Y] = E[X] E[Y] \\ \text{Cov}(X, Y) = 0 \\ \rho(X, Y) = 0 \end{cases}$$

3.5.3. Notación matricial

El **vector de medias** se define como:

$$\mathbf{m} = \begin{bmatrix} E[X] \\ E[Y] \end{bmatrix}$$

La **matriz de covarianza** se define como:

$$M = \begin{bmatrix} \text{Var}(X) & \text{Cov}(X, Y) \\ \text{Cov}(X, Y) & \text{Var}(Y) \end{bmatrix}$$

3.6 Variable aleatoria multidimensional

Prácticamente todas las definiciones mostradas para dos variables aleatorias se pueden generalizar para n variables aleatorias.

Una variable aleatoria n -dimensional se define como una aplicación:

$$(X_1, X_2, \dots, X_n) : \Omega \longrightarrow \mathbb{R}^n$$

Esta variable aleatoria asigna a cada resultado del experimento $\omega \in \Omega$ un elemento $(X_1(\omega), X_2(\omega) \dots X_n(\omega)) \in \mathbb{R}^n$

Para cualquiera de las operaciones lineales, resulta apropiada la notación por columnas:

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

Por lo tanto, escribiremos el vector de medias como:

$$\mathbf{m}_\mathbf{X} = E[\mathbf{X}] = \begin{bmatrix} E[X_1] \\ E[X_2] \\ \vdots \\ E[X_n] \end{bmatrix}$$

Y la matriz de covarianzas como:

$$M_\mathbf{X} = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_1, X_n) & \text{Cov}(X_2, X_n) & \cdots & \text{Var}(X_n) \end{bmatrix}$$

Linealidad

Si se realiza una combinación lineal con las variables aleatorias (X_1, X_2, \dots, X_n) se obtiene una variable aleatoria:

$$Y = a_1 X_1 + a_2 X_2 + \cdots + a_n X_n$$

Podemos expresar la variable aleatoria Y de la siguiente forma:

$$Y = A\mathbf{X}, \quad A = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \end{bmatrix}$$

Por la linealidad de la **esperanza**, tenemos que:

$$E[Y] = a_1 E[X_1] + a_2 E[X_2] + \cdots + a_n E[X_n]$$

La **varianza** de Y será:

$$\text{Var}(Y) = A \times M_{\mathbf{X}} \times A^{\top}$$

Donde A^{\top} es la matriz traspuesta de A . Se puede generalizar este resultado para m combinaciones lineales. Suponiendo que:

$$\mathbf{Y} = A\mathbf{X} + \mathbf{v}$$

Donde A es una matriz de dimensión $m \times n$ y \mathbf{v} es un vector columna, entonces la variable aleatoria m -dimensional \mathbf{Y} tiene un vector de medias $\mathbf{m}_{\mathbf{Y}}$ y una matriz de covarianza $M_{\mathbf{Y}}$ tales que:

$$\mathbf{m}_{\mathbf{Y}} = A\mathbf{m}_{\mathbf{X}} + \mathbf{v}; \quad \mathbf{m}_{\mathbf{Y}} = A\mathbf{m}_{\mathbf{X}}A^{\top}$$

3.7 Distribución multinomial

Si se realizan n pruebas de Bernoulli, la variable aleatoria bidimensional (X_1, X_2) definida por:

$X_1 \equiv$ número de éxitos

$X_2 \equiv$ número de fracasos

Esta variable solo podrá tomar los valores (n_1, n_2) con $n_1 = 0, 1, \dots, n$ y $n_2 = n - n_1$. Para estos valores:

$$P(X_1 = n_1, X_2 = n_2) = \frac{n!}{n_1!n_2!} p_1^{n_1} p_2^{n_2}$$

Donde $p_1 = p$ es la probabilidad de éxito y $p_2 = 1 - p$ es la probabilidad de fracaso.

Un caso interesante es aquel en el que en cada una de las n pruebas se puedan obtener k posibles resultados A_1, A_2, \dots, A_k con probabilidades p_1, p_2, \dots, p_k , donde necesariamente $p_1 + p_2 + \cdots + p_k = 1$. La variable n -dimensional (X_1, \dots, X_k) definida por:

$X_1 \equiv$ número de veces que se obtiene A_1

$X_2 \equiv$ número de veces que se obtiene A_2

\vdots

$X_k \equiv$ número de veces que se obtiene A_k

solo puede tomar valores (n_1, n_2, \dots, n_k) con $n_1, \dots, n_k = 0, 1, \dots, n$ y $n_1 + \cdots + n_k = n$. Para estos valores se tiene que:

$$P(X_1 = n_1, \dots, X_k = n_k) = \frac{n!}{n_1!n_2! \dots n_k!} p_1^{n_1} \dots p_k^{n_k}$$

Se dice que es (X_1, \dots, X_k) sigue una distribución multinomial de parámetros n, p_1, p_2, \dots, p_k , y se denota como:

$$(X_1, \dots, X_k) \sim \text{Mult}(n, p_1, p_2, \dots, p_k)$$

3.8 Distribución normal n-dimensional

Un vector aleatorio $\mathbf{X} = (X_1, \dots, X_n)$ sigue una distribución normal si la denominada función de densidad

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \dots \partial x_n} F_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

es igual a

$$\frac{1}{|M_{\mathbf{X}}|^{\frac{1}{2}} (2\pi)^{\frac{n}{2}}} e^{-\frac{1}{2}(\mathbf{X}-\mathbf{m}_{\mathbf{X}})^{\top} M_{\mathbf{X}}^{-1}(\mathbf{X}-\mathbf{m}_{\mathbf{X}})}$$

donde $M_{\mathbf{X}}$ es la matriz de covarianzas y $\mathbf{m}_{\mathbf{X}}$ es el vector de medias.

3.9 Teorema central del límite

Vamos a utilizar la definición de este teorema que nos será útil para resolver los ejercicios. La definición completa es [FALTA REFERENCIA].

Sean X_1, \dots, X_n variables aleatorias independientes e idénticamente distribuidas, cada una con media μ y varianza σ^2 .

El **teorema central del límite** afirma que si n es grande la suma S_n tiene aproximadamente distribución normal de media $n\mu$ y desviación típica $\sigma\sqrt{n}$ (varianza $n\sigma^2$), siendo esta suma:

$$S_n = X_1 + X_2 + \dots + X_n$$

Equivalentemente, la **variable normalizada** $\frac{S_n - n\mu}{\sigma\sqrt{n}}$ tiene aproximadamente distribución normal de media 0 y varianza 1.

3.10 Chi cuadrado y la t de Student

Sean Z_1, \dots, Z_n variables aleatorias independientes con distribución normal estándar, $Z_i \sim N(0, 1)$, con $i = 1, \dots, n$.

La distribución de probabilidad que sigue la variable aleatoria

$$X = Z_1^2 + \dots + Z_n^2$$

se denomina chi-cuadrado con n grados de libertad. Se denota $X \sim \chi_n^2$. La distribución χ_n^2 es positiva. Su media y su varianza son:

$$E(X) = n; \quad \text{Var}(X) = 2n$$

Sea $X \sim \chi_n^2$, y Z una variable aleatoria con distribución normal estándar, $Z \sim N(0, 1)$, independiente de X . La distribución de probabilidad que sigue la variable aleatoria

$$Y = \frac{Z}{\sqrt{X/n}}$$

se denomina t de Student con n grados de libertad. Se denota $Y \sim t_n$. La t de Student tiene una función de densidad simétrica muy similar, casi igual si n es grande, a la de una normal $N(0, 1)$. La principal diferencia es la varianza, que vale:

$$\text{Var}(Y) = \frac{n}{n-2}$$

Javier Rodrigo López

Capítulo 4

Inferencia estadística

- 4.1 Estadística descriptiva de una variable: momentos, cuantiles, box-plot, histograma, función de distribución empírica y cálculo de proporciones

- 4.2 Muestra aleatoria. Media muestral y varianza muestral. Estimación paramétrica

- 4.3 Intervalos de confianza para la media y para proporciones poblacionales

- 4.4 Contraste de hipótesis. Nivel de significación y p-valor

Javier Rodrigo López

Capítulo 5

Procesos estocásticos

5.1 Definición de proceso estocástico

Un proceso estocástico es una familia de variables aleatorias definidas sobre un espacio probabilístico.

Es decir, para cada $t \in T$, $X(t)$ es una variable aleatoria; el índice t representa a menudo el tiempo y así diremos, que un proceso estocástico es una familia de variables aleatorias que describe la evolución a lo largo del tiempo de algún proceso físico.

Otro punto de vista:

Sea E un experimento aleatorio cuyo espacio probabilístico asociado (Ω, S, P) . Un **proceso estocástico** o proceso aleatorio es un conjunto de funciones temporales $x(t, \omega)$, cada una correspondiente a un punto particular ω del espacio muestral Ω . Así, asociado a cada resultado específico ω , tenemos una función específica $x(t, \omega)$, a la que llamaremos una realización del proceso.

Denotaremos el proceso mediante: $X_t \equiv X(t) \equiv X(t, \omega)$, donde t es el tiempo y ω una variable que representa un resultado en el espacio muestral Ω .

Normalmente, un proceso estocástico será denotado como $x(t)$.

5.2 Procesos estocásticos en tiempo continuo

Primer ejemplo.

Tiramos un dado. Si se obtiene 1, 2 ó 3, entonces:

$$X(t) = t$$

Si se obtiene 4 ó 5, entonces:

$$X(t) = 2t$$

Si se obtiene un 6, entonces:

$$X(t) = 3t$$

Esto se puede modelizar de la siguiente manera:

$$X(t) = A \cdot t$$

De modo que la probabilidad de A será:

$$\begin{array}{c|c|c|c} A & 1 & 2 & 3 \\ \hline P(A) & 1/2 & 1/3 & 1/6 \end{array}$$

Segundo ejemplo.

La señal $X(t) = \cos(t + \phi)$, donde la fase ϕ es desconocida, se puede considerar como un proceso estocástico, donde ϕ es una variable aleatoria con distribución uniforme en el intervalo $[0, 2\pi)$

5.3 Procesos estocásticos en tiempo discreto

Simplemente son procesos estocásticos pero en tiempo discreto.

Por ejemplo, las pruebas de Bernoulli y el proceso binomial.

5.4 Distribuciones de primer y segundo orden, media, autocorrelación y autocovarianza

5.5 Proceso de Bernoulli. Caminos aleatorios. Procesos normales. Proceso de Poisson

5.6 Procesos estacionarios. Densidad espectral

5.7 Sistemas lineales y procesos estocásticos

Capítulo 6

Prácticas con software estadístico

6.1 Modelos de distribución de probabilidad más comunes

6.2 Estadística descriptiva

6.3 Muestreo. Estimación puntual

6.4 Estimación por intervalos de confianza

6.5 Contraste paramétrico
