Estadística y Procesos Estocásticos Apuntes de clase

Javier Rodrigo López ¹
31 de enero de 2021



 $^{^{1}\}hbox{E-mail: javiolonchelo@gmail.com}$





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 $\bf ETSIST$ de la $\bf UPM$ por Javier Rodrigo López, alumno de $\bf 1^0$ de Ingeniería de Sonido e Imagen. bruuh

Índice general

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..., mientras que los elementos se denotan con letras minúsculas, a, b, c...

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:

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

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:

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

ÍNDICE GENERAL

6

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

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

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

Propiedades de la frecuencia relativa

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

8 Probabilidad

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

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

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

Consecuencias de esta definición

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

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:

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

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 $\{\omega_j\}$, $j=1,2,\ldots,n$, 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 Laplace¹.

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

$$P(A) = \frac{N^{0} \text{ de elementos de } A}{N^{0} \text{ de elementos de } \Omega}$$

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 $n(n \le m)$:

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

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**.

lacktriangle 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:

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

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 $n(n \le m)$:

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

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:

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

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

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

Que se podría generalizar a tres sucesos $A, B \ y \ C$, resultando en la **regla de la multipli**cación:

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

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, \ldots, 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:

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

10 Probabilidad

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

$$P(A) = \sum_{i=1}^{n} P(A \mid C_1) P(C_i) \qquad \forall A \in \mathcal{S}$$

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

$$P(C_i | A) = \frac{P(A | C_1) P(C_i)}{\sum_{i=1}^{n} P(A | C_i) P(C_i)}$$

Ejemplo 6 del documento anexo (falta referencia).

$$P(B) = 3$$

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 $n=1,2,\ldots$

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 ${\mu_X}^2$ 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 Var(X).

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

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

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

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

 $^{^{1}}$ v.a. \equiv variable aleatoria

 $^{^2 \}mathrm{Es}$ común representar la media como μ a secas.

12 Variables aleatorias

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** (\overline{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:

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

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:

$$P(T=n) = (1-p)^{n-1} n$$
 $n = 0, 1, 2, ...$

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

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

• Media: $\mu = \frac{1}{p}$

• Varianza: $\sigma^2 = \frac{1-p}{p^2}$

<u>NOTA:</u> 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:

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

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

$$X \sim \mathrm{BN}\left(r,p\right)$$

• 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:

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

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:

$$P(a < X \le b) = \int_{a}^{b} f(x)dx \qquad f(x) \ge 0, \qquad \begin{cases} a, b \in \mathbb{R} \\ a \le b \end{cases}$$

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

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

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:

$$Var(X) = \sigma^2$$

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:

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

14 Variables aleatorias

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 intereseante haberlo definido así lo encontramos cuando intentamos calcular la probabilidad en un punto.

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

<u>NOTA:</u> 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:

$$X \sim \mathrm{U}(a,b)$$

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

$$P(X \in (c,d)) = \frac{d-c}{b-a}$$
 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:

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

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:

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

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 signfica que definiremos otra v.a. Z que modelice a X para que tenga distribución normal.

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

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:

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

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

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

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:

$$F_X(x) \ge p \qquad (0$$

En v.a. continua

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

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

O, en otras palabras:

$$F_T(q_p) = p$$

Variables aleatorias

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:

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

Baia baiiiia la cabaiiia

3.8

Vectores aleatorios

Teorema central del límite

3.1	Variable aleatoria bidimensional discreta. Funciones
	de distribución conjunta, marginales y condiciona-
	das. Cálculo de probabilidades
3.2	Variable aleatoria bidimensional continua. Función de distribución y función de densidad. Cálculo de probabilidades
3.3	Variable aleatoria multidimensional
3.4	Variables aleatorias independientes
3.5	Vector de medias. Matriz de covarianzas
3.6	Transformaciones lineales de vectores aleatorios
3.7	Vectores aleatorios normales

18 Vectores aleatorios

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

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:

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

Prácticas con software estadístico

6.1	Modelos de distribución de probabilidad más comu-
	nes
6.2	Estadística descriptiva
6.3	Muestreo. Estimación puntual
6.4	Estimación por intervalos de confianza
6.5	Contraste paramétrico

Bla bla bla