Probabilidad

Alejandro Zubiri

Índice

CHAPTER 1

Nociones Básicas

1. Experimentos Aleatorios

1.1. Espacio muestral. Son sucesos aleatorios repetidos cuyos resultados no se pueden determinar de antemano. El objetivo es describir el fenónemo desde un punto de vista aleatorio que describa el proceso.
El espacio muestral es el conjunto formado por todos los posibles resultados:

$$\Omega = \{\omega_1, \dots, \omega_n\}$$

Y tenemos diferentes tipos:

- Discreto: finito o numerable.
- Continuo.

Un suceso es cualquier subconjunto del espacio muestral. El proceso de codificación es el que consiste en pasar de una variable cualitativa a una cuantitativa.

1.2. Conjunto de partes. El conjunto de partes del espacio muestral es el conjunto de todos los posibles resultados de un experimento.

2. Probabilidad

Cualesquiera que sean A, se tiene que $0 \leq n(A) \leq n$, y que por tanto

$$\lim_{n\to\infty} \frac{n(A)}{n} = P(A)$$

Si A ocurre siempre, $n(A) = n \implies P(A) = 1$ Además, si A y B son excluyentes, se tiene que $P(A \cap B) = 0$, y que $P(A \cup B) = P(A) + P(B)$.

3. Axiomas de Kolmogorov

Las propiedades anteriores dotan a $\mathbb{A}, \cap, \cup$ de una estructura de álgebra de Boole (σ-álgebra). Una probabilidad P definida sobre un álgebra de sucesos A, de un espacio muestral finito Ω, es una función $P : A \to [0, 1]$.

- $P(\Omega) = 1$
- $P(A) \in [0, 1] \forall A$
- Si A y B son excluyentes, se tiene que $P(A \cap B) = 0$, y que $P(A \cup B) = P(A) + P(B)$.

También destacamos las leyes de De Morgan:

$$\overline{A \cup B} = \bar{A} \cap \bar{B}$$
$$\overline{A \cap B} = \bar{A} \cup \bar{B} \tag{1}$$

3.1. Propiedades.

- $P(\phi) = 0$
- $P(\bar{A}) = 1 - P(A)$
- $P(A - B) = P(A) - P(A \cap B)$, y además si $B \subset A$, $P(A - B) = P(A) - P(B)$
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- $P(A_1 \cup A_2 \cup A_3) = \sum_{i=1}^{3} P(A_i) - \sum_{i<j} P(A_i \cap A_j) + P(A_1 \cap A_2 \cap A_3)$

CHAPTER 2

Combinatoria

Supongamos que tenemos m elementos, y vamos a generar grupos. Cada grupo contiene n elementos.

1. Variaciones

1.1. Variaciones con repetición. Se llaman variaciones de repetición de m elementos, tomados de n en n, a los distintos grupos que se pueden formar de tal manera que cada grupo contiene n elementos distintos o iguales, y que un grupo se diferencia de los demás o bien en algún elemento o su posición. El número total es

$$VR(m,n) = m^n$$

Ejemplo: una subrutina genera 1 byte de forma aleatoria en cada ejecución. La cantidad del espacio muestral es:

- Como tenemos $m = 2$ elementos $(0,1)$, y los tomamos de 8 en 8, la cantidad de elementos en el espacio muestral es $V_8^2 = 8^2 = 64$.

1.2. Variaciones sin repetición. Se les llama, de m elementos, tomados de n en n, a los diferentes grupos que se pueden formar tal que

- Cada grupo contiene n elementos **distintos**.
- Cada grupo se diferencia de los demás en algún elemento o en su orden.

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

2. Permutaciones

2.1. Permutaciones ordinarias. Si se tienen permutaciones de n elementos, buscamos los distintos grupos que se pueden formar ordenando de diferentes formas n elementos:

$$P(n) = n!$$

2.2. Permutaciones con repetición. Donde el primer elemento se repite α veces, el segundo β veces, y así hasta el último que se repite γ veces. La suma $\alpha + \beta + \cdots + \gamma = n$. A los distintos grupos de n elementos que se pueden formar con los m elementos, donde cada grupo se compone de α veces el primer elemento, β veces el segundo, etc, el número total es

$$PR(n) = \frac{n!}{\alpha!\beta!\ldots\gamma!}$$

3. Combinaciones

3.1. Combinaciones sin repetición. De m elementos, tomados de n a n, a los diferentes grupos que se pueden formar de tal manera que cada grupo contenta n elementos distintos (el orden no importa). La cantidad total es

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

3.2. Combinaciones con repetición. De m elementos, los diferentes grupos que se pueden formar tal que cada grupo contiene n elementos distintos, en cada grupo se pueden repetir elementos, y cada grupo se diferencia al menos en un elemento. El número total es:

$$CR(m,n) = \binom{m+n-1}{n} = \frac{(m+n-1)!}{n!(m-1)!}$$

4. Diferencias entre ellas

Las diferencias entre estas operaciones son las siguientes:

- **Variaciones**: el orden SÍ importa.
- **Combinaciones**: el orden NO importa.
- **Permutaciones**: caso especial de variaciones donde ordenamos todos los elementos del conjunto $(n = m)$.

CHAPTER 3

Probabilidad condicionada

- Modelos dinámicos: experimentos que se realizan a lo largo del tiempo. Los experimentos son dependientes entre sí.
- Modelos estáticos: mismo período de tiempo.

La probabilidad de que ocurra B si ocurre A es

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

1. Regla del producto

Supongamos un modelo dinámico donde ocurren sucesos de forma consecutiva $\{A_1, \ldots, A_n\}$, se cumple que

$$\begin{aligned} &P(A_1 \cap \cdots \cap A_n) = \\ &P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \ldots P(A_n|A_1 \cap \cdots \cap A_{n-1}) \end{aligned} \tag{2}$$

Definición 1. *A y B son sucesos independientes si*

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

Definición 2. *Sea H una familia de sucesos que no incluyen al espacio vacío. Se dice que H está formada por sucesos completamente independientes, si para todo subconjunto de H se verifica que $H = \{\Delta_1, \ldots, \Delta_k\}$*

$$P(\Delta_1 \cap \cdots \cap \Delta_n) = \prod_{i=1}^{j} P(\Delta_i)$$

Teorema 1. *Dado un modelo matemático probabilístico representado por $(\Omega, \mathcal{A}, P)$, un sistema completo formado por $\{A_1, \ldots, A_n\} \in \mathcal{A}$ (con todos ellos excluyentes), y sabiéndose $\exists B \in \mathcal{A} P(B|A_i) \forall i$, se verifica que*

$$P(B) = \sum P(A_i)P(B|A_i)$$

Entonces se verifica que

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)}$$

2. Independencia condicional

Si se verifica que

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

Entonces A y B son condicionalmente independientes dado C.

CHAPTER 4

Tema Cuatro

1. Variable aleatoria

Existen varias formas de clasificar variables:

- Cualitativa: en función de una cualidad
- Cuantitativa: se puede medir y evaluar con un número.
- Cuasi-cuantitativa: se ordenan en función de una magnitud que puede pertenecer a diferentes intervalos.

Dado un modelo matemático $(\Omega, \mathcal{A}, P)$ formada por un espacio muestral Ω, un álgebra de sucesos $\mathcal{A}$, y una probabilidad P. Una variable discreta X definida sobre el modelo es una función definida como

$$X : \Omega \to \mathbb{R}$$

Que se denomina como función de masa X_i.
Se llama función de distribución de X a la función $F(X)$ que

$$F(X = x_j) = P(\{\omega \in \Omega : X(\omega) \leq x_j\}) = P(-\infty, x_j]$$

- $\lim_{x \to -\infty} F(X) = 0$
- $\lim_{x \to \infty} F(X) = 1$
- F es monótona decreciente.
- F es continua por la derecha.
- $P(x_1 < x \leq x_2) = F(x_2) - F(x_1)$

2. Estadística descriptiva

Dada una variable aleatoria discreta X, se define esperanza de X como

$$\mu = E(X) = \sum_i x_i P(X = x_i)$$

también la moda, que se define como el valor con mayor probabilidad.
La mediana se define como el valor que se encuentra en la mitad del resto de los valores:

$$P(-\infty, med] = F(X = med) \geq \frac{1}{2}$$

2.1. Distribución de una variable condicionada por un suceso A. Dada una función de masa $P(X)$, la probabilidad de que X tome x_i si ha ocurrido A viene dada por

$$P(X = x_i|A) = \frac{P(\{\omega \in \Omega | X(\omega) = x_i\} \cap A)}{P(A)}$$

y entonces

$$E(X|A) = \sum x_i P(x_i|A)$$

3. Momentos respecto al origen

Se denomina momento respecto al origen de orden $r \forall r \in \mathbb{N}$ y se denota como α_r:

$$\alpha_r = E[X^r] = \sum_{i=0}^{n} x_i^r P(X = x_i)$$

El momento de orden 1 es el valor medio. Con esto podemos definir la varianza como

$$\sigma^2 = E[X^2] - E[X]^2 = \alpha_2 - \alpha_1^2$$

4. Momentos respecto a la media

Dada la variable aleatoria X discreta con función de masa $P(X = x_i)$, se llama momento respecto a la media de orden r, denota por μ_r como

$$\mu_r = E[(X - \mu)^r] = \sum (x_i - \mu)^r P(X = x_i)$$

5. Medidas de forma de la distribución

5.1. Coeficiente de asimetría Fisher. El coeficiente de asimetría Fiser, denotado por γ_1, se define como

$$\gamma_1 = \frac{\mu_3}{\sigma^3} \tag{3}$$

En función de los valores de γ_1, tenemos:

- $\gamma_1 = 0$: distribución simétrica.
- $\gamma_1 > 0$: distribución asimétrica positiva.
- $\gamma_1 < 0$: distribución asimétrica negativa.

5.2. Curtosis. Se define como

$$\gamma_2 = \frac{\mu_4}{\sigma^4} - 3 \tag{4}$$

Este coeficiente mide el apuntamiento, es decir, como de ”plana” es la distribución.

- $\gamma_2 = 0$: mesocúrtica.
- $\gamma_2 > 0$: menos aplastada.
- $\gamma_2 < 0$: más aplastada.

6. Entropía

Definimos la entropía asociada a una variable aleatoria X como

$$H(X) = -\sum p(x_i) \log[p(x_i)]$$

7. Distribuciones de probabilidad más comunes

7.1. Distribución binomial. Se deben dar las siguientes condiciones:

- El experimento consiste en una secuencia de n ensayos, donde n se conoce con antelación.
- Cada ensayo tiene dos posibles resultados, que definimos como éxito (S), o fracaso (F).
- Cada ensayo es independiente del anterior.
- La probabilidad de éxito es constante y se define como p.

Si queremos, dada una probabilidad p de que ocurra A, que de n elementos, ocurra A x veces, la probabilidad es:

$$P(x) = b(x,n,p) = C_{x,n}p^x(1-p)^{n-x} = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}$$

Y el valor medio es

$$E[X] = n \cdot p$$

la varianza

$$\sigma^2 = np(1-p)$$

7.2. Distribución de Poisson. Si $n \to \infty$ y $x \in [0, \infty)$, entonces la función masa binomial se aproxima a una función llamada **masa de Poisson**:

$$\lim_{(n,p)\to(\infty,0)} b(x,n,p) = p(x,\lambda)$$

donde $\lambda = n \cdot p$, y la función masa es

$$p(x,\lambda) = \left\{ \begin{array}{c} \frac{e^{-\lambda}\lambda^x}{x!} \ si \ x = 0,1,\ldots,n \\ 0 \ else \end{array} \right\}$$

Para $X \sim P(X,\lambda)$, la función distibución de X se denota por

$$P(X;\lambda) = P\{\omega \in \Omega : X \leq x\} = \sum_{y=0}^{x} p(y,\lambda)$$

Donde se cumple que $\mu = E(X) = \lambda$, y que $V(X) = \sigma^2 = \lambda$

7.3. Distribución geométrica. Si tenemos dos posibles resultados, éxito (S) y fracaso (F), y queremos saber la probabilidad de obtener éxito en el ensayo k (es decir, obteniendo fracaso $k-1$ veces), la probabilidad de obtener éxito en el ensayo k es:

$$P(X = k) = (1-p)^{k-1}p \ si \ k = 1,2,3\ldots$$

y la función distribución como

$$F(X = k) = P(-\infty, X] = \sum_{j=1}^{k}(1-p)^{j-1}p \ si \ x = 0,1,\ldots,n$$

teniendo que $\mu = E(X) = \frac{1}{p}$, y que $V(X) = \sigma^2 = \frac{1-p}{p^2}$

8. Distribución binomial negativa

Condiciones:

- Cada ensayo tiene un fracaso o un éxito.
- Los ensayos son independientes.
- La probabilidad de éxito es constante.
- El experimento continua hasta obtener r éxitos observados.

Con esto, obtenemos la probabilidad de que se produzcan F fracasos antes de obtener E éxitos, dada una probabilidad p:

$$nb(F,E,p) = \binom{F+E-1}{F} p^E (1-p)^F \tag{5}$$

Además, se tiene que

- $\mu = r\frac{1-p}{p}$
- $\sigma^2 = \frac{\mu}{p}$

9. Distribución hipergeométrica

Esta distribución nos dice la probabilidad de obtener un número determinado de éxitos al extraer una muestra de una población finita formada por éxitos y fracasos.
Tenemos las siguientes suposiciones:

- La población es finita y consiste de N elementos.
- Cada individuo se caracteriza como un éxito S o fracaso F, y hay M éxitos.
- Se elige una muestra de n individuos tal que cada subconjunto tenga las mismas probabilidades de ser elegido.

Si X es el número de éxitos, entedremos M éxitos y $(N-M)$ fracasos:

$$P(X=x) = h(x;n,M,N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Además, se da que

- $E(X) = n\frac{M}{N}$
- $V(X) = n\frac{N-n}{N-1}\frac{M}{N}(1-\frac{M}{N})$

Ejemplo:
De 5 animales de una especie en extinción, se toman 25 animales y se liberan con otros individuos. Después de mezclarse, se tomaron 10 de esos 25 animales. Se define X como el número de animales marcados en la segunda muestra que contiene 10 animales. Indica la función masa de X y ciertas probabilidades.
Datos:

- $M = 5$
- $N = 25$
- $n = 10$
- x será el número de ejemplares marcados.

Ahora nos interesa que $x = 2$, por lo que aplicamos la ley de Laplace:

$$P(X=2) = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} \approx 0.38$$

Para $x \leq 2$, tenemos que

$$P(X \leq 2) = P(X=0) + P(X=1) + P(X=2) = 0.056 + 0.2569 + 0.385 = 0.6988$$

Y la esperanza es $E(X) = 2$.

9.1. Diferencias entre las distribuciones.

- **Binomial**: Cuando tienes un número fijo de ensayos independientes con probabilidad constante de éxito.
- **Poisson**: cuando n tiende a infinito y p tiende a 0.
- **Geométrica**: Cuando te interesa el número de ensayos hasta el primer éxito.
- **Binomial negativa**: Cuando te interesa el número de ensayos hasta el r-ésimo éxito.
- **Hipergeométrica**: Cuando extraes una muestra sin reemplazo de una población finita.

CHAPTER 5

Tema Cinco

1. Distribución conjunta de un vector aleatorio discreto

Supongamos un experimento caracterizado por un espacio muestral $\Omega = \{\omega_i\}$. Empezamos con variables aleatorias discretas, dadas por

$$X = \{x_1, \ldots, x_n\}$$

A cada elemento ω de Ω se le asigna un vector aleatorio discreto $(X(\omega))$.

1.1. Función distribución. Se define como

$$P(X \leq x_i, Y \leq y_i) = \sum_{k=1}^{i} \sum_{k=1}^{j} P(x_k, y_k)$$

Definición 3. *La distribución marginal de X es la probabilidad de que X tome el valor x_i a medida que Y cambia:*

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

2. Momentos y parámetros de una distribución

Definición 4. *El momento respeto al origen de orden (k, s) se define como:*

$$\alpha_{k,s} = E(X^k, y^S) = \sum \sum x_i^k y_j^s P(x_i, y_j)$$

El momento $k = 1, s = 0$ corresponde a la media de X, y el momento $k = 0, s = 1$ corresponde a la media de Y.

El conjunto de las medias se denomina **vector de medias**.

Definición 5. *El momento respecto al valor medio de orden k, s se define como:*

$$\mu_{k,s} = E[(X - \mu_x)^k (Y - \mu_y)^s] = \sum \sum (x_i - \mu_x)^k (y_j - \mu_y)^s P(x_i, y_j)$$

Con esto podemos definir la desviación típica como:

$$\sigma_y^2 = \mu_{0,2}$$

y

$$\sigma_x^2 = \mu_{2,0}$$

Definición 6. *Definimos la covarianza $\sigma_{x,y}$ como:*

$$\sigma_{x,y} = \mu_{1,1}$$

que se puede simplificar como:

$$\sigma_{x,y} = E[X, Y] - E[X]E[Y]$$

Se da que, si x e y son independientes, $E[XY] = E[X]E[Y]$, y por tanto $\sigma_{x,y} = 0$.

3. Distribuciones condicionadas

La probabilidad de que $X = x$ condicionada por $Y = y$, es decir

$$P(X = x|Y = y) = \frac{P(x, y)}{P(y)}$$

Con esto podemos generar una distribución de probabilidad para una variable, condicionada a que la segunda variable sea igual a un valor determinado y_0. De aquí sigue que:

$$E(X|Y = y_0) = \sum x_i P(x_i|Y = y_0) \tag{6}$$

DEFINICIÓN 7. *Definimos la matriz covarianza como:*

$$Cov(x, y) = \begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix} \tag{7}$$

Se da que el determinante de esta matriz es:

$$\det(Cov(x, y)) = \sigma_x^2\sigma_y^2 - \sigma_{xy}^2 = \sigma_x^2\sigma_y^2(1 - (\frac{\sigma_{xy}}{\sigma_x\sigma_y})^2) = \sigma_x^2\sigma_y^2(1 - \rho^2)$$

donde ρ es el coeficiente de correlación lineal.

4. Correlación lineal

La recta $y(x)$ que más aproxima los datos es:

$$y(x) = a + bx \tag{8}$$

donde

$$\begin{aligned} b &= \frac{\sigma_{xy}}{\sigma_x} \\ a &= E[y] - bE[x] \end{aligned} \tag{9}$$

CHAPTER 6

Variable Continua

Definida por variables que toman valores reales.

DEFINICIÓN 8. *Definimos la función densidad $\rho : \mathbb{R} \to \mathbb{R}$ como la que da la probabilidad de que x se halle en un intervalo $[a, b]$.*

- $\rho(x) > 0$.
- $\int_{-\infty}^{\infty} \rho(x)\,\mathrm{d}x = 1$

Sin embargo, ya no podemos encontrar la probabilidad de que x tome un valor específico. Solo podemos calcular:

$$P(a \leq x \leq b) = \int_a^b \rho(x)\,\mathrm{d}x \tag{10}$$

Ahora, definimos

$$E[x] = \int x\rho(x)\,\mathrm{d}x \tag{11}$$

ahora nuestra función distribución se define como

$$F(x_0) = P(x \leq x_0) = \int_{-\infty}^{x_0} \rho(x)\,\mathrm{d}x \tag{12}$$

Por tanto, tenemos que si $F(x)$ es derivable en x_0, entonces

$$\rho(x_0) = \frac{\mathrm{d}F}{\mathrm{d}x} \tag{13}$$

Propiedades:

- $P(a, b] = F(b) - F(a)$
- $P[a, b] = F(b) - F(a^-)$
- $P(a) = F(a^+) - F(a^-)$

1. Percentiles

Sea p un número definido entre 0 y 1. El percentil p de una distribución se denota como $\eta(100p)$ y está definido como:

$$p = F(\eta(100p)) = \int_{-\infty}^{x_M} \rho(t)\,\mathrm{d}t \tag{14}$$

En este caso, resolveríamos la ecuación de igualar nuestra función distribución a $\frac{1}{2}$ si quisiéramos obtener la mediana.

2. Probabilidad condicionada

Dada una variable continua X con función densidad $f(x)$, se define la función densidad de la variable aleatoria continua condicionada $X|A$ como

$$f(X|A) = \frac{f(x)}{P(A)}$$

3. Distribución exponencial

Se define como

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

y se da que la esperanza es

$$E[X] = \frac{1}{\lambda}$$

y su función distribución

$$F(x) = 1 - e^{-\lambda x}$$

además de $V(x) = \frac{1}{\lambda^2}$ y $med = \frac{\ln 2}{\lambda}$.
Cuando queremos ver la probabilidad de que x esté a una distancia de k desviaciones de la media, buscamos

$$P(-k \leq \frac{x-\mu}{\sigma} \leq k)$$

Cuando n tiende a 30, la distribución binomial se aproxima a una normal.

4. Desigualdad de Chebyshev

Supongamos una variable aleatoria X con media μ y desviación típica σ. Se verifica que

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