

Repaso VIP: Probabilidades y Estadísticas

Afshine AMIDI y Shervine AMIDI

6 de octubre de 2018

Traducido por Fermín Ordaz. Revisado por Fernando González-Herrera y Alonso Melgar López.

Introducción a la probabilidad y combinatoria

□ **Espacio muestral** – El conjunto de todos los posibles resultados de un experimento es conocido como el espacio muestral del experimento y se denota como S .

□ **Evento** – Cualquier subconjunto E del espacio muestral es conocido como un evento. Esto significa que un evento es un conjunto de posibles resultados de un experimento. Si el resultado de un experimento está contenido en E , entonces decimos que el evento E ha ocurrido.

□ **Axiomas de la probabilidad** – Para cada evento E , denota $P(E)$ como la probabilidad de que el evento E ocurra.

$$(1) \quad 0 \leq P(E) \leq 1 \quad (2) \quad P(S) = 1 \quad (3) \quad P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$$

□ **Permutación** – Una permutación es un arreglo de r objetos tomados de un grupo de n objetos, en un orden arbitrario. El número de estos arreglos es dado por $P(n, r)$, definido como:

$$P(n, r) = \frac{n!}{(n-r)!}$$

□ **Combinación** – Una combinación es un arreglo de r objetos tomados de un grupo de n objetos, donde el orden no importa. El número de estos arreglos es dado por $C(n, r)$, definido como:

$$C(n, r) = \frac{P(n, r)}{r!} = \frac{n!}{r!(n-r)!}$$

Observación: cabe resaltar que para $0 \leq r \leq n$, se tiene $P(n, r) \geq C(n, r)$.

Probabilidad condicional

□ **Regla de Bayes** – Para eventos A y B tal que $P(B) > 0$, se tiene:

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

Observación: Se tiene $P(A \cap B) = P(A)P(B|A) = P(A|B)P(B)$.

□ **Partición** – Sea $\{A_i, i \in \llbracket 1, n \rrbracket\}$ tal que para todo i , $A_i \neq \emptyset$. Se dice entonces que $\{A_i\}$ es una partición si se cumple:

$$\forall i \neq j, A_i \cap A_j = \emptyset \quad \text{y} \quad \bigcup_{i=1}^n A_i = S$$

Observación: Para cualquier evento B del espacio muestral, se cumple $P(B) = \sum_{i=1}^n P(B|A_i)P(A_i)$.

□ **Regla de Bayes extendida** – Sea $\{A_i, i \in \llbracket 1, n \rrbracket\}$ una partición del espacio muestral. Se cumple:

$$P(A_k|B) = \frac{P(B|A_k)P(A_k)}{\sum_{i=1}^n P(B|A_i)P(A_i)}$$

□ **Independencia** – Dos eventos A y B son independientes si y solo si se cumple:

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

Variables aleatorias

□ **Variable aleatoria** – Una variable aleatoria, generalmente denotada por X , es una función que asocia cada elemento de un espacio muestral a una línea real.

□ **Función de distribución acumulada (FDA)** – La función de distribución acumulada F (en inglés *CDF* - *Cumulative distribution function*), la cual es monótonamente creciente y es tal que:

$$\lim_{x \rightarrow -\infty} F(x) = 0 \quad \text{y} \quad \lim_{x \rightarrow +\infty} F(x) = 1$$

es definida como:

$$F(x) = P(X \leq x)$$

Observación: Se tiene $P(a < X \leq b) = F(b) - F(a)$.

□ **Función de densidad de probabilidad (FDP)** – La función de densidad de probabilidad f (en inglés *PDF* - *Probability density function*) es la probabilidad que X tome valores entre dos ocurrencias adyacentes de la variable aleatoria.

□ **Relaciones entre la FDA y FDP** – Estas son las propiedades más importantes para conocer en los casos discreto (D) y continuo (C).

Caso	FDA F	FDP f	Propiedades de PDF
(D)	$F(x) = \sum_{x_i \leq x} P(X = x_i)$	$f(x_j) = P(X = x_j)$	$0 \leq f(x_j) \leq 1$ and $\sum_j f(x_j) = 1$
(C)	$F(x) = \int_{-\infty}^x f(y)dy$	$f(x) = \frac{dF}{dx}$	$f(x) \geq 0$ and $\int_{-\infty}^{+\infty} f(x)dx = 1$

□ **Varianza** – La varianza de una variable aleatoria, frecuentemente denotada por $\text{Var}(X)$ o σ^2 , es la medida de dispersión de su función de distribución. Esta determinada de la siguiente manera:

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

□ **Desviación estándar** – La desviación estándar de una variable aleatoria, frecuentemente denotada por σ , es una medida de la dispersión de su función de distribución la cual es compatible con las unidades de la correspondiente variable aleatoria. Se determina de la siguiente manera:

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

□ **Valor esperado y momentos de la distribución** – Aquí están las expresiones del valor esperado $E[X]$, valor esperado generalizado $E[g(X)]$, $k^{\text{ésimo}}$ momento $E[X^k]$ y función característica $\psi(\omega)$ para los casos discreto y continuo:

Caso	$E[X]$	$E[g(X)]$	$E[X^k]$	$\psi(\omega)$
(D)	$\sum_{i=1}^n x_i f(x_i)$	$\sum_{i=1}^n g(x_i) f(x_i)$	$\sum_{i=1}^n x_i^k f(x_i)$	$\sum_{i=1}^n f(x_i) e^{i\omega x_i}$
(C)	$\int_{-\infty}^{+\infty} x f(x) dx$	$\int_{-\infty}^{+\infty} g(x) f(x) dx$	$\int_{-\infty}^{+\infty} x^k f(x) dx$	$\int_{-\infty}^{+\infty} f(x) e^{i\omega x} dx$

Observación: se tiene $e^{i\omega x} = \cos(\omega x) + i \sin(\omega x)$.

□ **Transformación de variables aleatorias** – Sean las variables X y Y asociadas por alguna función. Denotemos como f_X y f_Y la función de distribución de X y Y respectivamente, se tiene:

$$f_Y(y) = f_X(x) \left| \frac{dx}{dy} \right|$$

□ **Regla integral de Leibniz** – Sea g una función de x y posiblemente de c , y además sea a, b , un intervalo que puede depender de c . Se tiene:

$$\frac{\partial}{\partial c} \left(\int_a^b g(x) dx \right) = \frac{\partial b}{\partial c} \cdot g(b) - \frac{\partial a}{\partial c} \cdot g(a) + \int_a^b \frac{\partial g}{\partial c}(x) dx$$

□ **Desigualdad de Chebyshev** – Sea X una variable aleatoria con valor esperado μ . Para $k, \sigma > 0$, se tiene la siguiente desigualdad:

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

Variables aleatorias conjuntas

□ **Densidad condicional** – La densidad condicional de X con respecto a Y , frecuentemente denotada como $f_{X|Y}$, es definida como:

$$f_{X|Y}(x) = \frac{f_{XY}(x, y)}{f_Y(y)}$$

□ **Independencia** – Dos variables aleatorias X y Y son consideradas independientes si se tiene:

$$f_{XY}(x, y) = f_X(x) f_Y(y)$$

□ **Densidad marginal y distribución acumulada** – De la función conjunta de densidad de probabilidad f_{XY} , se tiene:

Caso	Densidad marginal	Función acumulativa
(D)	$f_X(x_i) = \sum_j f_{XY}(x_i, y_j)$	$F_{XY}(x, y) = \sum_{x_i \leq x} \sum_{y_j \leq y} f_{XY}(x_i, y_j)$
(C)	$f_X(x) = \int_{-\infty}^{+\infty} f_{XY}(x, y) dy$	$F_{XY}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{XY}(x', y') dx' dy'$

□ **Covarianza** – Definimos la covarianza de dos variables aleatorias X y Y , denotada como σ_{XY}^2 o comúnmente como $\text{Cov}(X, Y)$, de la siguiente manera:

$$\text{Cov}(X, Y) \triangleq \sigma_{XY}^2 = E[(X - \mu_X)(Y - \mu_Y)] = E[XY] - \mu_X \mu_Y$$

□ **Correlación** – Sean σ_X, σ_Y las desviaciones estándar de X y Y , definimos la correlación entre estas variables, denotada como ρ_{XY} , de la siguiente manera:

$$\rho_{XY} = \frac{\sigma_{XY}^2}{\sigma_X \sigma_Y}$$

Observaciones 1: cabe resaltar que para X, Y , variables aleatorias cualesquiera, se tiene que $\rho_{XY} \in [-1, 1]$. Si X y Y son independientes, entonces $\rho_{XY} = 0$.

□ **Distribuciones importantes** – Aquí están las distribuciones más importantes para tomar en cuenta:

Tipo	Distribución	FDP	$\psi(\omega)$	$E[X]$	$\text{Var}(X)$
(D)	$X \sim \mathcal{B}(n, p)$ Binomial	$P(X = x) = \binom{n}{x} p^x q^{n-x}$ $x \in \llbracket 0, n \rrbracket$	$(pe^{i\omega} + q)^n$	np	npq
	$X \sim \text{Po}(\mu)$ Poisson	$P(X = x) = \frac{\mu^x}{x!} e^{-\mu}$ $x \in \mathbb{N}$	$e^{\mu(e^{i\omega} - 1)}$	μ	μ
(C)	$X \sim \mathcal{U}(a, b)$ Uniform	$f(x) = \frac{1}{b-a}$ $x \in [a, b]$	$\frac{e^{i\omega b} - e^{i\omega a}}{(b-a)i\omega}$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$
	$X \sim \mathcal{N}(\mu, \sigma)$ Gaussian	$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$ $x \in \mathbb{R}$	$e^{i\omega\mu - \frac{1}{2}\omega^2\sigma^2}$	μ	σ^2
	$X \sim \text{Exp}(\lambda)$ Exponential	$f(x) = \lambda e^{-\lambda x}$ $x \in \mathbb{R}_+$	$\frac{1}{1 - \frac{i\omega}{\lambda}}$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

□ **Teorema del Límite Central** – Sea X_1, \dots, X_n una muestra aleatoria que sigue una distribución con media μ y varianza σ^2 , entonces se tiene:

$$\bar{X} \underset{n \rightarrow +\infty}{\sim} \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

Estimación de parámetros

□ **Muestra aleatoria** – Una muestra aleatoria es una colección de n variables aleatorias X_1, \dots, X_n que son independientes e idénticamente distribuidas a X .

□ **Estimador** – Un estimador es una función de los datos que es usada para inferir el valor de un parámetro desconocido en un modelo estadístico.

□ **Sesgo** – El sesgo de un estimador $\hat{\theta}$ se define como la diferencia entre el valor esperado de la distribución de $\hat{\theta}$ y el valor exacto, esto es:

$$\text{Sesgo}(\hat{\theta}) = E[\hat{\theta}] - \theta$$

Observación: se dice que un estimador es no sesgado cuando se tiene $E[\hat{\theta}] = \theta$.

□ **Media de la muestra** – La media de la muestra aleatoria se usa para estimar el valor exacto de la media μ de la distribución, se denota frecuentemente como \bar{X} y se define de la siguiente manera:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

Observación: la media de la muestra es no sesgada, esto es $E[\bar{X}] = \mu$.

□ **Media de la muestra** – La media de la muestra aleatoria se usa para estimar el valor exacto de la media μ de la distribución, se denota frecuentemente como \bar{X} y se define de la siguiente manera:

$$s^2 = \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Observación: la varianza de la muestra es no sesgada, esto es $E[s^2] = \sigma^2$.