

Universidad Nacional de La Plata

FACULTAD DE CIENCIAS ASTRONÓMICAS Y GEOFÍSICAS

# ESTADÍSTICA APLICADA

*Resumen teórico*

Autor:  
Lorenzo Girotti

2024

## UNIDAD 1

---

### PROBABILIDAD

#### 1.1. Definiciones

Antes que nada, vamos a definir conceptos que utilizaremos a lo largo del resumen.

- **Variable aleatoria:** Es el resultado de un experimento aleatorio, que puede ser por ejemplo la acción de lanzar un dado y ver qué número queda en la cara superior. En notación, diremos que  $X$  es una variable aleatoria y la denotaremos por  $X$ .
- **Variable determinista:** Es un valor que podemos *determinar* (valga la redundancia) y usualmente asumiremos que se comportan como números reales o subconjuntos de los mismos. Los notamos como  $x$ .
- **Espacio muestral o universo:** Representa la totalidad de los posibles resultados de un experimento aleatorio, normalmente lo notaremos como  $E$  o  $U$ .
- **Eventos mutuamente excluyentes:** Un *evento* lo podemos pensar como un suceso o algo que puede ocurrir. Cuando tenemos dos o más eventos que tienen la característica de que no pueden coexistir o suceder al mismo tiempo decimos que son *mutuamente excluyentes*. Por ejemplo: sea el evento  $A$  que salga cruz; y el evento  $B$  que salga cara.  $A$  y  $B$  son eventos mutuamente excluyentes en el experimento de lanzar una moneda.
- **Probabilidad empírica:** Dado un evento que sea resultado de un experimento aleatorio que consiste en  $n$  pruebas, definimos la probabilidad de que ocurra ese evento como:

$$P(X) = \frac{\#X}{n} \quad (1.1)$$

siendo  $\#X$  la cantidad de veces que ocurre el evento aleatorio  $X$ .

## 1.2. Reglas de probabilidad y resultados importantes

➤ **Probabilidad aditiva:**

$$P(A) + P(B) - P(A \cap B). \quad (1.2)$$

en caso de que sean *mutuamente excluyentes*,  $P(A \cap B) = 0$

➤ **Probabilidad conjunta:**

$$P(A \cap B) = P(A|B)P(B) \quad (1.3)$$

si  $A$  y  $B$  son *estadísticamente independientes*, entonces  $P(A|B) = P(A)$  y  $P(A \cap B) = P(A)P(B)$ .

➤ **Probabilidad total:** Dado un espacio muestral compuesto de  $n$  subconjuntos:  $E = A_1 \cup \dots \cup A_n$ , la probabilidad de que ocurra un evento  $B$  dentro del espacio, es

$$P(B) = \sum_{i=1}^n P(B \cap A_i), \quad (1.4)$$

considerando que es la unión de todas las intersecciones entre  $B$  y los subconjuntos  $A_i$  y asumiendo que los mismos son *mutuamente excluyentes*. Si tenemos en cuenta la 1.3, la probabilidad total queda

$$P(B) = \sum_{i=1}^n P(B|A_i)P(A_i). \quad (1.5)$$

➤ **Teorema de Bayes:** Sean  $A_i$  como en el ítem anterior,

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{P(B)} \quad (1.6)$$

## UNIDAD 2

---

### VARIABLE ALEATORIA

#### 2.1. Función de distribución de una variable aleatoria

Si consideramos una variable aleatoria discreta, definimos a su *función de distribución* como

$$P(X < x) = \sum_{i=1}^{k(x_k < x)} P(X = x_i). \quad (2.1)$$

Si la variable aleatoria es continua, se define la función de densidad lineal de probabilidad:

$$f(x)dx = P(x \leq X < x + dx) \quad (2.2)$$

Notar que para una función continua no tiene sentido hablar de la probabilidad de que  $X$  sea igual a un valor determinado. Sí diremos que la variable aleatoria se encuentra en un intervalo infinitesimal de posibilidades.

Podemos relacionar la *densidad de probabilidad* con la *distribución* de la variable a través de:

$$f(x) = \frac{dF(x)}{dx} \quad (2.3)$$

. Tener en cuenta que

$$f(x) \geq 0; \quad \int_{-\infty}^{\infty} f(x)dx = 1$$

. De esta manera, podemos calcular la probabilidad acumulada de una variable aleatoria continua a través de su función de densidad de probabilidad.

$$F(x) = \int_{-\infty}^x f(t)dt. \quad (2.4)$$

$$P(x_1 < X < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x)dx. \quad (2.5)$$

Notar que

$$F(-\infty) = \int_{-\infty}^{-\infty} f(x)dx = 0; \quad F(\infty) = \int_{-\infty}^{\infty} f(x)dx = 1.$$

## 2.2. Operador Esperanza y parámetros de variables aleatorias

Definimos al *operador esperanza* como:

$$E[H(X)] = \sum_{i=1}^n H(x_i)P(X = x_i)$$

si  $X$  es discreta.

$$E[H(X)] = \int_{-\infty}^{\infty} H(x)f(x)dx$$

si  $X$  es continua.

El operador es *lineal*, es decir:

$$\begin{aligned} E[X + Y] &= E[X] + E[Y] \\ E[cX] &= cE[X], \end{aligned}$$

donde  $c$  es una constante.

### 2.2.1. Momentos de orden

Dado un  $H(X) = (X - c)^l$ , el valor esperado  $E[H(X)]$  se llama *momento de orden  $l$*  de una variable con respecto a  $c$ .

Si tomamos a  $c = \mu$  siendo  $\mu$  la media de una variable, definimos  $\mu_l = E[(X - \mu)^l]$  y tenemos:

$$\begin{aligned} \mu_0 &= 1 \\ \mu_1 &= 0 \\ \mu_2 &= E[(X - \mu)^2] = \sigma^2 && \text{Varianza} \\ \mu_3 &= E[(X - \mu)^3] && \text{Sesgo de X} \\ \mu_4 &= E[(X - \mu)^4] && \text{Curtosis de X} \\ &\vdots \end{aligned}$$

Nota:  $E[X] = \mu$

Cada *momento* define mejor el comportamiento de la variable.

### 2.2.2. Variable normalizada

Una variable normalizada es aquella que tiene media cero y varianza uno.

$$u = \frac{X - \mu}{\sigma_X} \quad (2.6)$$

### 2.2.3. Desigualdad de Chebychev

$$P(|x - \mu| > k\sigma) \leq \frac{1}{k^2} \quad (2.7)$$

siendo  $k \in \mathbb{R}$ .

La desigualdad nos da una idea probabilística de nuestra variable  $X$ , conociendo solo su media y varianza (sin importar cual es su función de probabilidad o probabilidad asociada). Por ejemplo: si elegimos  $k = 3$ , tenemos el *criterio de los tres sigmas* para descartar *outlayers*.

## 2.3. Transformación de variables

Si tenemos una transformación de variable  $Y(X)$  biyectiva, *la probabilidad en intervalos equivalentes se conserva*, esto implica

$$\begin{aligned} P(x \leq X < x + dx) &= P(y \leq Y < y + dy) \\ f(x)dx &= g(y)dy \end{aligned}$$

Entonces podemos definir la densidad de probabilidad de  $g$  como

$$g(y) = \left| \frac{dx}{dy} \right| f(x) \quad (2.8)$$

Gráficamente podemos decir que hay una *igualdad de áreas*.

## 2.4. Multivariabes

Todo el análisis que se realiza para una variable aleatoria, puede extenderse a una idea multivariable.

### 2.4.1. Dos variables

Siendo dos variables continuas, tenemos:

**Función de densidad de probabilidad conjunta**

$$f(x, y) = \frac{\partial}{\partial x} \frac{\partial}{\partial y} F(x, y), \quad (2.9)$$

y,

$$P(a \leq X < b, c \leq y < d) = \int_a^b \left[ \int_c^d f(x, y) dy \right] dx. \quad (2.10)$$

**Función de densidad de probabilidad marginal**

Se consideran todos los valores posibles para una variable, dejando una densidad de probabilidad con única dependencia de la otra.

$$g(x) = \int_{-\infty}^{\infty} f(x, y) dy \quad (2.11)$$

tenemos entonces,

$$P(a \leq X < b, -\infty < y < \infty) = \int_a^b \left[ \int_{-\infty}^{\infty} f(x, y) dy \right] dx = \int_a^b g(x) dx \quad (2.12)$$

**Independencia de variables**

Dos variables se dicen independientes *estadísticamente* si

$$f(x, y) = g(x)h(y) \quad (2.13)$$

donde  $g(x)$  y  $h(y)$  son las funciones de densidad marginales.

**Probabilidad condicional**

Partimos de la base de encontrar:

$$P(y \leq Y < y + dy \mid x \leq X < x + dx)$$

La función de densidad de probabilidad de Y si X:

$$f(y|x) = \frac{f(x, y)}{g(x)} \quad (2.14)$$

**Parámetros**

➤ **Valor esperado:**

$$E[H(x, y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(x, y) f(x, y) dx dy$$

➤ **Varianza conjunta:**

$$\sigma^2 [H(x, y)] = E \left[ (H(x, y) - E[H(x, y)])^2 \right]$$

➤ **Medias y varianzas para x (ídem para y):**

$$E[x] = \int_{-\infty}^{\infty} xg(x)dx = \mu_x$$

$$\sigma_x^2 = E[(x - \mu_x)^2] = \int_{-\infty}^{\infty} (x - \mu_x)^2 g(x)dx$$

➤ **Covarianza:**

$$\begin{aligned} \text{cov}(x, y) &= E[(x - \mu_x)(y - \mu_y)] = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y)f(x, y)dxdy \end{aligned} \quad (2.15)$$

➤ **Coefficiente de correlación:**

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma(x)\sigma(y)} \quad (2.16)$$

Como vemos, se hace una extensión de los parámetros que ya conocíamos; con la diferencia de que ahora la dependencia estadística entre las variables se vuelve de gran interés. La *covarianza* 2.15 es un parámetro ideal que permite analizar la dependencia estadística en conjunto con la *correlación* 2.16. Algunas de las características más importantes de la correlación son las siguientes:

1.  $-1 \leq \rho(x, y) \leq 1$
2. Si  $x$  e  $y$  son variables estadísticamente independientes, entonces  $\rho(x, y) = 0$
3. Si  $y$  es una función determinista de  $x$ ,  $y = H(x)$ , entonces  $\rho(x, y) = \pm 1$

#### 2.4.2. Transformación de variables con dos variables

Sean las variables  $u$  y  $v$  funciones deterministas de las variables aleatorias  $x$  e  $y$ , de forma que  $u = u(x, y)$  y  $v = v(x, y)$ . Para realizar la transformación, seguimos el mismo criterio de *igualdad de áreas* solo que ahora tenemos una dependencia de dos variables. Por lo tanto, de ahora en adelante tendremos que utilizar el *Jacobiano*.

Conociendo  $f(x, y)$  determinamos  $g(u, v)$  a través de:

$$g(u, v) = f(x(u, v), y(u, v)) \left| J \begin{pmatrix} x, y \\ u, v \end{pmatrix} \right| \quad (2.17)$$



### 2.4.3. Caso de n variables

Definimos un vector de variables aleatorias  $\vec{x} = (x_1, \dots, x_n) = (x_1 \cdots x_n)^T$  y entonces quedan:

**Función de distribución conjunta:**

$$F(x_1, \dots, x_n) = P(x_1 < x_1, \dots, x_n < x_n)$$

**Función de densidad de probabilidad:**

$$f(x_1, \dots, x_n) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(x_1, \dots, x_n)$$

**Función de densidad de probabilidad marginal:**

$$g_r(x_r) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_{r-1} dx_{r+1} \cdots dx_n$$

**Operador Esperanza:**

$$E[H(\vec{x})] = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_n) dx_1 \cdots dx_n$$

**Media de una variable en particular:**

$$\begin{aligned} E[x_r] &= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} x_r f(x_1, \dots, x_n) dx_1 \cdots dx_n \\ E[x_r] &= \int_{-\infty}^{\infty} x_r g_r(x_r) dx_r \end{aligned}$$

**Independencia de variables:**

Si las variables son independientes, tenemos

$$f(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n)$$

#### 2.4.4. Vector de media, varianza y covarianza

Si coleccionamos todas las medias de las variables  $(\mu_{x_r})$  obtenemos un vector de medias dado por:

$$\vec{\mu}_{\vec{x}} = (\mu_{x_1}, \dots, \mu_{x_n}) = (E[x_1] \cdots E[x_n])^T \quad (2.18)$$

Si aplicamos lo mismo para intentar calcular la varianza; obtenemos la *matriz de varianza-covarianza*

$$E[(\vec{x} - \vec{\mu}_{\vec{x}})(\vec{x} - \vec{\mu}_{\vec{x}})^T] = \begin{bmatrix} E[(x_1 - \mu_1)^2] & \cdots & E[(x_1 - \mu_1)(x_n - \mu_n)] \\ \vdots & \ddots & \vdots \\ E[(x_n - \mu_n)(x_1 - \mu_1)] & \cdots & E[(x_n - \mu_n)^2] \end{bmatrix}$$

que utilizando las definiciones de varianza y covarianza queda

$$E[(\vec{x} - \vec{\mu}_{\vec{x}})(\vec{x} - \vec{\mu}_{\vec{x}})^T] = \begin{bmatrix} \sigma_{x_1}^2 & \cdots & \text{cov}(x_1, x_n) \\ \vdots & \ddots & \vdots \\ \text{cov}(x_n, x_1) & \cdots & \sigma_{x_n}^2 \end{bmatrix} \quad (2.19)$$

### 2.5. Transformación de n variables

Siendo  $\vec{x} = (x_1, \dots, x_n)$  y le aplicamos una función determinista  $\vec{y}$  de modo tal que,

$$\begin{aligned} y_1 &= y_1(\vec{x}) \\ &\vdots \\ y_m &= y_m(\vec{x}) \end{aligned}$$

Notar que  $\vec{y} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ .

Entonces, la función de densidad de probabilidad de la nueva variable  $\vec{y}$  queda

$$g(\vec{y}) = \left| J \left( \frac{\vec{x}}{\vec{y}} \right) \right| f(\vec{x}(\vec{y})) \quad (2.20)$$

donde

$$\left| J \left( \frac{\vec{x}}{\vec{y}} \right) \right| = \left| \begin{array}{ccc} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_n}{\partial y_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_1}{\partial y_m} & \cdots & \frac{\partial x_n}{\partial y_m} \end{array} \right|$$

### 2.5.1. Transformación lineal y ortogonal

#### Regla de propagación de errores

Considerando  $m$  funciones lineales de  $n$  variables  $\vec{x} : \mathbb{R}^{n \times 1}$  con media y matriz de varianza-covarianza conocidas:

$$\begin{aligned} y_1 &= a_1 + t_{11}x_1 + \cdots + t_{1n}x_n \\ &\vdots \\ y_m &= a_m + t_{m1}x_1 + \cdots + t_{mn}x_n \end{aligned}$$

en forma matricial o vectorial:

$$\vec{y} = T\vec{x} + \vec{a}$$

la *Ley de Propagación de Errores* establece:

$$C_y = TC_x T^T \quad (2.21)$$

#### Matriz T para un caso no de transformación no lineal

En el caso de que la transformación no sea lineal, podemos *linealizarla* a través de un desarrollo de Taylor a primer orden, centrado en la media de  $x$

$$y_i = y_i(\vec{\mu}) + \left( \frac{\partial y_i}{\partial x_1} \right) (x_1 - \mu_1) + \cdots + \left( \frac{\partial y_i}{\partial x_n} \right) (x_n - \mu_n) + T.O.S$$

luego, la matriz  $T$  será:

$$T = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$$

**Todas las derivadas son evaluadas en  $\vec{\mu}$**

## UNIDAD 3

---

### FUNCIONES DE DISTRIBUCIÓN ESPECIALES

#### 3.1. Distribución Binomial

Esta función de distribución se puede asignar a experimentos aleatorios que tengan las siguientes características:

1. Sólo dos resultados posibles mutuamente excluyentes:  $A$  o  $\neg A$
2. Independencia de resultados en cada prueba.
3.  $P(A) = p = \text{cte.}$
4. Existe un número finito,  $n$ , de pruebas.

La variable aleatoria binomial cuenta la cantidad de *éxitos* en un total de  $n$  pruebas.

$$x = \sum_{i=1}^n x_i$$

con

$$x_i = \begin{cases} 1, & \text{si sucede } A \\ 0, & \text{si no sucede } A \end{cases}$$

.

$$P(k) = W_k^n = \binom{n}{k} p^k q^{n-k} \quad (3.1)$$

##### 3.1.1. Parámetros de la binomial

➤ *Media:*

$$E[x_i] = np$$

➤ *Varianza:*

$$\sigma^2(x_i) = npq$$

### 3.2. Multinomial

Dada una serie de eventos mutuamente excluyentes que terminan definiendo todo el espacio muestral:  $E = A_1 + \dots + A_l$ ; donde

$$P(A_j) = p_j, \quad \sum_{j=1}^l p_j = 1$$

La variable multinomial se expresa:

$$x_j = \sum_{i=1}^n x_{ij}$$

$$x_{ij} = \begin{cases} 1, & \text{si sucede } A_j \\ 0, & \text{en el resto.} \end{cases}$$

De esta manera, la probabilidad queda:

$$P(x_1 = k_1, \dots, x_l = k_l) = \frac{n!}{\prod_{j=1}^l k_j!} \prod_{j=1}^l p_j^{k_j}, \quad (3.2)$$

tener en cuenta que

$$\sum_{j=1}^l k_j = n$$

#### 3.2.1. Parámetros de la multinomial

➤ *Media:*

$$E[\vec{x}] = n\vec{p}, \quad \vec{p} = \begin{bmatrix} p_1 \\ \vdots \\ p_l \end{bmatrix}$$

➤ *Varianza:*

$$\sigma^2(x_j) = np_j(1 - p_j), \quad j = 1, \dots, l$$

➤ *Covarianza:*

$$\text{cov}(x_i, x_j) = -np_i p_j, \quad i \neq j$$

### 3.3. Poissoniana

Partiendo de una binomial, con las siguientes condiciones:  $n \rightarrow \infty$ ,  $p \rightarrow 0$  y  $np = \lambda$ ; tenemos

$$\begin{aligned}
P(k) &= \binom{n}{k} p^k q^{n-k} \\
&= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\
&= \frac{\lambda^k}{k!} \frac{n(n-1)(n-2)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^n \\
&= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \frac{\left(1 - \frac{1}{n}\right)\cdots\left(1 - \frac{k-1}{n}\right)}{\left(1 - \frac{\lambda}{n}\right)^k},
\end{aligned}$$

tomando  $\lim_{n \rightarrow \infty}$  queda:

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (3.3)$$

### 3.3.1. Parámetros de la poissoniana

➤ *Media:*

$$E[k] = \lambda$$

,

➤ *Varianza:*

$$\sigma^2 = \lambda$$

.

*Nota: Para la demostración utilizar la definición de media para variables discretas, teniendo en cuenta que la poissoniana va de 0 a  $\infty$ .*

## 3.4. Gaussiana

Es la llamada *distribución normal*, su función de densidad de probabilidad está dada por:

$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}b} \exp\left(-\frac{(x-a)^2}{2b^2}\right), \quad -\infty < x < \infty$$

### 3.4.1. Parámetros de la gaussiana

$$\mu = a, \quad \sigma^2 = b^2$$

*Nota: para la demostración proponer cambio de variable  $z = (x - \mu)/\sigma$  y tener en cuenta que la integral definida  $\int_{-\infty}^{\infty} \exp(-z^2/2) dz = \sqrt{2\pi}$*

De esta manera, la función de densidad de probabilidad queda:

$$f(x) = \phi(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty < x < \infty \quad (3.4)$$

### 3.4.2. Normalización de la gaussiana

Si utilizamos la variable normalizada  $x' = (x - \mu)/\sigma$

$$\varphi(x') = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x'^2}{2}\right) \quad (3.5)$$

Lo importante de normalizar la gaussiana es poder encontrar valores de manera rápida, ya que se encuentran tabulados.

### 3.4.3. Criterio de los 3 sigmas

Utilizando la desigualdad de Chebychev (2.7) con la gaussiana estandarizada, podemos notar lo siguiente:

$$\begin{aligned} P(|x - \mu| \leq \sigma) &= 68,3\%, \\ P(|x - \mu| \leq 2\sigma) &= 95,4\%, \\ P(|x - \mu| \leq 3\sigma) &= 99,8\%. \end{aligned}$$

De manera tal que cualquier valor aleatorio (que siga una distribución gaussiana) que se encuentre con un desvío superior a  $3\sigma$  es altamente improbable. Lo cual es un buen indicador para descartar *outlayers* o para encontrarlos y estudiar si son desechables o no.