1. Estadística descriptiva

X, Y variables con muestras $x_1, \ldots, x_n$, e $y_1, \ldots, y_n$, respectivamente.

- La **media** $\overline{x} = \frac{1}{n}\sum_{i=1}^n x_i$
- La **varianza** $V_x = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})^2 = \overline{x^2} - \overline{x}^2$ y la desviación típica $\sqrt{V_x}$
- La **asimetría** $\text{asim}_x = \frac{\frac{1}{n}\sum (x_i - \overline{x})^3}{V_x^{3/2}}$
- La **covarianza**, cumple $|\text{cov}_{x,y}| < \sqrt{V_x}\sqrt{V_y}$

$$\text{cov}_{x,y} = \frac{1}{n}\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y}) = \overline{xy} - \overline{xy}$$

- El **coeficiente de correlación** (adimensional, tipificado)

$$\rho_{x,y} = \frac{\text{cov}_{x,y}}{\sqrt{V_x}\sqrt{V_y}}$$

- La **recta de regresión** de Y sobre X, $y = \hat{b}x + \hat{a}$, es

$$\hat{b} = \frac{\text{cov}_{x,y}}{V_x}, \quad \hat{a} = -\hat{b}\overline{x} + \overline{y}$$
$$\frac{y - \overline{y}}{\sqrt{V_y}} = \rho_{x,y}\frac{x - \overline{x}}{\sqrt{V_x}}$$

  - La **bondad del ajuste** $\sqrt{E(a,b)} = \sqrt{V_y}\sqrt{1 - \rho_{x,y}^2}$ donde $\rho_{x,y}^2 = R^2, \; E(a,b) = \frac{1}{n}\sum_{i=1}^n (y_i - (\hat{a} + \hat{b}x_i))^2$
  - El ajuste logarítmico $y = B\ln(x) + A$: cambio de variable $Z = \ln(X)$.
  - El ajuste exponencial $y = Ce^{Dx}$: cambio de variable $W = \ln(Y)$ y tomar $C = e^{\hat{a}}, \; D = \hat{b}$.
  - El ajuste potencial $y = CX^H$: cambios de variable $W = \ln(Y)$ y $Z = \ln(X)$.
- Ajuste por norma euclídea (sin asumir $Y = f(X)$): tomar $E(a,b) = \frac{1}{n}\sum_{i=1}^n \|y_i, (\hat{a} + \hat{b}x_i)\|^2$

2. Variables aleatorias

- X **variable aleatoria** es una lista de valores $x_1, \ldots, x_n$ con sus probabilidades $p_1, \ldots, p_n$ con $p_i = P(X = x_i)$.
- **Función de masa o de densidad** $f_X(x)$ da las probabilidades de cada dato. Cumple $f_X \geq 0$, $\sum_{i=1}^n f_X(x_i) = \int_{\mathbb{R}} f_X = 1$ en el caso discreto y continuo, resp.
- **Función de distribución** $F_X(x)$ cumple $0 \leq F_X \leq 1$, F_X creciente, $F_X' = f_X$

$$F_X(x) = P(X \leq x) = \sum_{j|x_j<x} p_j = \int_{-\infty}^x f_X(t)dt$$

- La **esperanza** $E(X) = \sum_{i=1}^n x_i P(X = x_i) = \int_{\mathbb{R}} x f_X(x)dx$
  - es lineal: $E(aX + b) = aE(X) + b, \; E(X + Y) = E(X) + E(Y)$
  - Cauchy-Schwarz: $|E(XY)|^2 \leq E(X^2)E(Y^2)$
- La **covarianza** $\text{cov}(X,Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y)$ es un producto escalar (**bilineal!**):
  - $cov(X_1 + X_2, X_3) = cov(X_1, X_3) + cov(X_2, X_3)$
  - $cov(\lambda X_1, X_2) = \lambda cov(X_1, X_2)$
- el **coeficiente de correlación** $\rho_{X,Y} = \frac{\text{cov}_{X,Y}}{\sqrt{V(X)}\sqrt{V(Y)}}$

$$X \perp\!\!\!\perp Y \implies \rho_{X,Y} = 0 \iff \text{cov}_{X,Y} = 0$$
$$\iff E(XY) = E(X)E(Y)$$

- La **varianza** $V(X) = E[(X - E(X))^2] = E(X^2) - E(X)^2$
  - $V(\lambda X) = \text{cov}(\lambda X, \lambda X) = \lambda^2 V(X)$
  - $V(X + Y) = V(X) + V(Y) + 2\text{cov}_{X,Y}$

2.1. Modelos de distribución

- **Bernouilli:** $X \sim Ber(p)$: $x_k \in \{0,1\}, \; p = P(X = 1)$

$$E(X) = p \qquad V(X) = p(1-p) \qquad P(X = k) = p^k(1-p)^{1-k}$$

- **Binomial:** $X \sim Binom(n,p)$. Repetir una Bernouilli n veces y contar los aciertos. $x_i = k = 0, \ldots, n$

$$P(X = k) = \binom{n}{k}p^k(1-p)^{n-k}$$
$$E(X) = np \qquad V(X) = np(1-p)$$

- **Poisson:** $X \sim Poisson(\lambda = np)$. Binomial con $n \to \infty, p \to 0$.

$$E(X) = \lambda \qquad V(X) = \lambda \qquad P(X = k) = \frac{e^{-k}\lambda^k}{k!}$$

- **Exponencial:** $X \sim Exp(\lambda), \; E(X) = \frac{1}{\lambda}, \; V(X) = \frac{1}{\lambda^2}$

$$f_X(x) = \lambda e^{-\lambda x}\mathbf{1}_{[0,\infty)} \qquad F_X(x) = 1 - e^{-\lambda x}, \; x \in [0, \infty]$$

- **Gamma:** $X \sim Gamma(\lambda, t)$

$$\text{función gamma } \Gamma(s) = (s-1)\Gamma(s-1) = \int_0^\infty t^{s-1}e^{-t}dt$$
$$f_X(x) = \frac{1}{\Gamma(t)}\lambda^t x^{t-1}e^{-\lambda x}, \qquad E(X^k) = \frac{(t+k-1)!}{\lambda^k(t-1)!}$$
$$\text{y es de utilidad } \int_0^\infty x^{t-1}e^{-bx}dx = \frac{\Gamma(t)}{b^t}$$

Propiedad: $U \sim Gamma(a,b) \perp\!\!\!\perp V \sim Gamma(a,c) \implies U + V \sim Gamma(a, b+c)$

- **Normal**: $X \sim \mathcal{N}(\mu, \sigma^2)$

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}, \quad E(X) = \mu, \; V(X) = \sigma^2$$

2.2. Cambio de variable

Si $Y = H(X)$ entonces

$$f_Y(H(x))|H'(x)| = f_X(x)$$
$$f_Y(y) = f_X(H^{-1}(y))|(H^{-1})'(y)|$$

Si $Z = X + Y$ entonces $f_Z(z) = \int_{-\infty}^\infty f_{(X,Y)}(x, z-x)dx$

3. Vectores aleatorios

$X_1, \dots, X_n$ vv. aa., $\mathbb{X} = (X_1, \dots, X_n)^t$ vector aleatorio.

- La **función de densidad conjunta** $f_{(X_1,\dots,X_n)}(x_1, \dots x_n)$
- $X_1, \dots, X_n$ **indep.** $\iff$ $P(X_1 \in A_1, \dots, X_n \in A_n) = P(X_1 \in A_1) \cdot \dots \cdot P(X_n \in A_n)$
- El **vector de medias** $E(\mathbb{X}) = (E(X_1), \dots, E(X_n))$
- La **matriz de covarianzas**

$$V(\mathbb{X}) = \operatorname{cov}(\mathbb{X}) = \begin{pmatrix} V(X_1) & \dots & \operatorname{cov}_{X_1,X_n} \\ \vdots & \ddots & \vdots \\ \operatorname{cov}X_n, X_1 & \dots & V(X_n) \end{pmatrix}$$

es simétrica y semidefinida positiva.

3.1. Vectores normales

Sea $\vec{m} \in \mathbb{R}^n, V$ matriz simétrica def. positiva de $n \times n$. Entonces $\mathbb{X} \sim \mathcal{N}(\vec{m}, V) \iff$

$$f_{\mathbb{X}}(\vec{x}) = \frac{1}{(\sqrt{2\pi})^n} \frac{1}{\sqrt{\det V}} \exp\left\{-\frac{1}{2}(\vec{x} - \vec{m})^t V^{-1} (\vec{x} - \vec{m})\right\}$$

V es def. pos $\implies V = UU^t,\ \det U \neq 0$ escribimos

$$\mathbb{X} \sim \mathcal{N}(\vec{m}, V) \iff \mathbb{X} = \vec{m} + U\mathbb{Y} \iff$$
$$f_{\mathbb{X}}(\vec{x}) = \frac{1}{(\sqrt{2\pi})^n} \frac{1}{|\det U|} \exp\left\{-\frac{1}{2}\|U^{-1}(\vec{x} - \vec{m})\|^2\right\}$$

Además se tiene

$$E(\mathbb{X}) = \vec{m} \qquad \operatorname{cov}(\mathbb{X}) = V \qquad X_j \sim \mathcal{N}(m_j, V_{jj})$$

- Si $\vec{h} \in \mathbb{R}^n, B \in \mathbb{R}^{n \times n},\ \det B \neq 0,\ \mathbb{X} \sim \mathcal{N}(\vec{m}, V)$ vector normal, entonces $\mathbb{Z} = h + B\mathbb{X} \sim \mathcal{N}(\vec{h} + B\vec{m}, BVB^t)$.
- Si $\vec{a} \in \mathbb{R}^n$ entonces la combinación lineal $\sum_{i=1}^n a_i X_i = \vec{a}^t \mathbb{X}$ es v.a. y $\sum_{i=1}^n a_i X_i \sim \mathcal{N}(\vec{a}^t \vec{m}, \vec{a}^t V \vec{a})$.

3.2. Distr. asociadas a vectores normales

- **Chi-cuadrado** $Z \sim \chi^2_n$ con n grados de libertad. $Z = \|\mathbb{X}\|^2 = \sum X_i^2,\ X_i \sim \mathcal{N}(0,1),\ X_1, \dots, X_n$ indep.

$$E(Z) = n, \quad E(Z^\alpha) = 2^\alpha \frac{\Gamma(n/2 + \alpha)}{\Gamma(n/2)}, \quad V(Z) = 2n$$
$$X_i^2 \sim \mathit{Gamma}(1/2, 1/2), \qquad Z \sim \mathit{Gamma}(1/2, n/2)$$

- **F de Fischer** $Z \sim F_{n,m}$. Si $U \sim \chi^2_n,\ V \sim \chi^2_m,\ U \perp\!\!\!\perp V$ entonces

$$Z = \frac{U/n}{V/m} \sim F_{n,m}$$
$$E(Z) = \frac{m}{m-2}, \quad V(Z) = \frac{2m^2(m+n-2)}{n(m-4)(m-2)^2}$$

- **t de Student** $Z \sim t_n$. Si $Y \sim \mathcal{N}(0,1),\ U_n \sim \chi^2_n,\ Y \perp\!\!\!\perp U_n$ entonces

$$Z = \frac{Y}{\sqrt{\frac{U_n}{n}}} \sim t_n, \qquad E(Z) = 0, \quad V(Z) = \frac{n}{n-2}$$

4. Modelo de muestreo aleatorio

Sea X una v.a. Sean $X_1, \dots, X_n$ clones independientes de X que cumplen

- X_i, X_j independientes si $i \neq j$
- $X_i \overset{\mathrm{d}}{=} X,\ i = 1, \dots, n$ (X_i es igual a X en distribución)

Un **estadístico** T es una función $T = H(X_1, \dots, X_n)$ donde $H : \mathbb{R}^n \to \mathbb{R}$.

4.1. Estadístico media muestral

$$\overline{X} = \frac{1}{n}\sum_{i=1}^n X_i \qquad E(\overline{X}) = E(X) \qquad V(\overline{X}) = \frac{1}{n}V(X)$$

- Si $X \sim \mathit{Ber}(p)$ entonces $n\overline{X} \sim \mathit{Binom}(n,p)$:

$$P(\overline{X} = \frac{k}{n}) = \binom{n}{k} p^k (1-p)^{n-k}$$

- Si $X \sim \mathit{Poisson}(\lambda)$

$$P(\overline{X} = \frac{k}{n}) = P(n\overline{X} = k) = e^{-n\lambda}\frac{(n\lambda)^k}{k!}$$

- Si $X \sim \mathcal{N}(\mu, \sigma^2)$ entonces $\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n})$

$$P(|\overline{X} - E(X)| \geq \lambda) \leq \frac{V(X)}{n\lambda^2}$$
$$\sqrt{n}(\overline{X} - E(X)) \overset{\mathrm{d}}{=}_{n\to\infty} \mathcal{N}(0, V(X))$$

4.2. Estadístico cuasivarianza muestral

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2$$
$$n(n-1)S^2 = (n-1)\sum_{i=1}^n X_j^2 - \sum_{i \neq j} x_i x_j$$
$$E(S^2) = V(X)$$
$$V(S^2) = \frac{1}{n}E((X - E(X))^4) - \frac{n-3}{n(n-1)}V(X)^2$$

Estimación de la dispersión de S^2 con Chebyshev:

$$P(|S^2 - V(X)| > \lambda) \leq \frac{V(S^2)}{\lambda^2} \leq \frac{1}{\lambda n^2}E((X - E(X))^4)$$

Teorema (de Fischer-Cochran)**.** *Si $\mathbf{X} \sim \mathcal{N}(\mu, \boldsymbol{\sigma^2})$, $X_1, \dots, X_n$ clones independientes de X y $\overline{X}$, S^2 son los estadísticos habituales entonces $\overline{X}$ **y S^2 son independientes**. Además*

$$\overline{X} \sim \mathcal{N}(\mu, \frac{\sigma^2}{n}), \quad \frac{(n-1)}{\sigma^2}S^2 \sim \chi^2_{n-1}, \quad \frac{\overline{X}}{S/\sqrt{n}} \sim t_{n-1}$$

4.3. Estadísticos máximo y mínimo

$$Mn = \max\{X_1, \dots X_n\} \qquad m_n = \min\{X_1, \dots, X_n\}$$
$$F_{M_n}(t) = F_X(t)^n \qquad F_{m_n}(t) = 1 - (1 - F_X(t))^n$$

Para α **mínimo esencial** ($F_X(x < \alpha) = 0 \wedge F_X(x > \alpha) > 0$ y β **máximo esencial** ($F_X(x < \beta) < 1 \wedge F_X(x > \beta) = 1$) de X

$$P(\alpha \leq x \leq r) \xrightarrow{n\to\infty} 1,\ r > \alpha$$
$$P(r \leq x \leq \beta) \xrightarrow{n\to\infty} 1,\ r < \beta$$

5. Estimación de parámetros

Sea X v.a. con modelo descrito por $f(x;\theta), \theta \in \Theta$ desconocido. Sea $x_1, \dots, x_n$ una muestra de X.

- El **soporte** $\text{sop}_\theta = \{x \in \mathbb{R} \mid f(x;\theta) > 0\}$
- Un **estimador** es un estadístico $T = h(x_1, \dots, x_n) \mapsto \widehat{\theta}$ que dados unos datos da una estimación $\widehat{\theta}$

5.1. Método de momentos

El **estimador por momentos de orden** n se obtiene de

$$\overline{x^n} = E_\theta(X^n)$$

X es	**orden 1**	**orden 2**
Ber(p)	$M_p = \overline{x}$	-
Unif$([0,a])$	$M_a = 2\overline{x}$	-
Exp(λ)	$M_\lambda = \frac{1}{\overline{x}}$	-
Tri$([-\theta,\theta])$	-	$M_\theta = \sqrt{6\overline{x^2}}$
$\mathcal{N}(\mu,\sigma^2)$	-	$M_\mu = \overline{x},\ M_{\sigma^2} = V_x$

- En las distribuciones simétricas, se utilizan momentos de orden par.

5.2. Método de máxima verosimilitud

Sea X una v.a. con distribución $f(x;\theta)$ discreta y finita, $x_1, \dots, x_n$ muestra de X

- La **función de verosimilitud** $\text{VERO} : \Theta \to \mathbb{R}$

$$\text{VERO}(\theta; x_1, \dots, x_n) = \prod_{j=1}^{n} f(x_j;\theta)$$

- La **estimación de máxima verosimilitud**, si existe, es el máximo global $\widehat{\theta}$ de VERO en Θ
  - Para minimizar, es útil tomar $\log \text{VERO}$

X es	$\text{VERO}(\theta; x_1, \dots, x_n)$	máx VERO en
Ber(p)	$p^{n\overline{x}}(1-p)^{n(1-\overline{x})}$	$\widehat{p} = \overline{x}$
Unif$([0,a])$	$\begin{cases} \left(\frac{1}{a}\right)^n & \text{si } a \geq \text{máx}\, x_j \\ 0 & \text{si } a < \text{máx}\, x_j \end{cases}$	$\widehat{a} = \text{máx}\, x_j$
Exp(λ)	$\lambda^n e^{-n\overline{x}\lambda}$	$\widehat{\lambda} = \frac{1}{\overline{x}}$

Sean T, T' estimadores

- El **sesgo** $\text{sesgo}(T) = |E_\theta(T) - \theta|$
  - T se dice **insesgado** $\iff \text{sesgo}(T) = 0 \iff \theta \in \Theta,\ E_\theta(T) = \theta$
  - Conocido el sesgo, se puede eliminar: para $X \sim \text{Unif}([0,a])$ se obtiene un esitmador insesgado con $T = \frac{n+1}{n} \text{máx}\, x_j$
- T es más eficiente que $T' \iff V_\theta(T) < V_\theta(T')$

5.3. Límites de calidad de estimadores

- La **variable de información**

$$Y = \frac{\partial}{\partial\theta} \log f(x;\theta) = \frac{\partial_\theta f(x;\theta)}{f(x;\theta)}$$

- La **cantidad de información** $I_X(\theta) = V_\theta(Y)$

X es	Y	$I_X(\theta)$
Ber(p)	$\frac{x-p}{p(1-p)}$	$\frac{1}{p(1-p)}$
$\mathcal{N}(\mu,\sigma^2)$	$\frac{(x-\mu_0)^2-\sigma^2}{\sigma^4}$ para μ_0 fijo	$\frac{1}{2\sigma^4}$
Exp(λ)	$\lambda^2 x - \lambda$	λ^2

Lema (de Cramer-Rao). *Sea X una v.a. con $\theta \in \Theta = (a,b)$, $sop_\theta = \{a_1, \dots, a_M\}$, $M < \infty$, $\theta \mapsto f(x;\theta)$ derivable en Θ. Entonces $E_\theta(Y) = 0, \forall\theta \in \Theta$.*