# Formulario Estadística

Abel Doñate Muñoz

## Contents

**Tabla de distribuciones discretas**

| $Modelo$ | $p(X=k)$ | $E[X]$ | $Var[X]$ | $G_X(z)$ |
|---|---|---|---|---|
| **Bernoulli** $\sim Be(p)$ | $\begin{cases} p(X=1)=p \\ p(X=0)=1-p \end{cases}$ | $p$ | $p(1-p)$ | $(1-p)+pz$ |
| **Binomial** $\sim Bin(N,p)$ | $\binom{N}{k} p^i(1-p)^{N-k}$ | $Np$ | $Np(1-p)$ | $((1-p)+pz)^N$ |
| **Uniforme** $\sim U(1,N)$ | $\frac{1}{N}$ | $\frac{N+1}{2}$ | $\frac{N^2-1}{12}$ | $\frac{1}{N}\frac{z(z^N-1)}{z-1}$ |
| **Poisson** $\sim Po(\lambda)$ | $\frac{\lambda^k}{k!}e^{-\lambda}$ | $\lambda$ | $\lambda$ | $e^{\lambda(z-1)}$ |
| **Geométrica** $\sim Geom(p)$ | $p(1-p)^{k-1}$ | $\frac{1}{p}$ | $\frac{1-p}{p^2}$ | $\frac{pz}{1-(1-p)z}$ |
| **Binomial negativa** $\sim BinN(r,p)$ | $\begin{cases} 0 & \text{si } k<r \\ \binom{k-1}{r-1}p^r(1-p)^{k-r} & \text{si } k\geq r \end{cases}$ | $\frac{r}{p}$ | $r\frac{1-p}{p^2}$ | $\left(\frac{pz}{1-(1-p)z}\right)^r$ |

**Tabla de distribuciones continuas**

| $Modelo$ | $f_X(x)$ | $E[X]$ | $Var[X]$ | $G_X(z)$ |
|---|---|---|---|---|
| **Uniforme** $\sim U(a,b)$ | $\frac{1}{b-a}\mathbb{I}_{[a,b]}$ | $\frac{b+a}{2}$ | $\frac{(b-a)^2}{12}$ | 1 en $t=0$, $\frac{e^{ibt}-e^{iat}}{it(b-a)}$ |
| **Exponencial** $\sim Exp(\lambda)$ | $\lambda e^{-\lambda x}, x\geq 0, \lambda>0$ | $\frac{1}{\lambda}$ | $\frac{1}{\lambda^2}$ | $\frac{\lambda}{\lambda-it}$ |
| **Normal** $\sim N(\mu,\sigma^2)$ | $\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$ | $\mu$ | $\sigma^2$ | $e^{i\mu t-\frac{\sigma^2t^2}{2}}$ |
| **Gamma** $\sim Gamma(\lambda,\tau)$ | $\frac{\lambda^\tau}{\Gamma(\tau)}x^{\tau-1}e^{-\lambda x}, x>0, \lambda,\tau>0$ | $\frac{\tau}{\lambda}$ | $\frac{\tau}{\lambda^2}$ | $\left(1-\frac{it}{\lambda}\right)^{-\tau}$ |
| **Beta** $\sim Beta(\alpha,\beta)$ | $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{\alpha-1}(1-x)^{\beta-1}, x\in[0,1]$ | $\frac{\alpha}{\alpha+\beta}$ | $\frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$ | Sin forma sencilla |
| **Weibull** $\sim Weibull(\alpha,\beta)$ | $\frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}e^{-(x/\beta)^\alpha}, x,\alpha,\beta>0$ | $\beta\Gamma\left(1+\frac{1}{\alpha}\right)$ | $\beta^2\left[\Gamma(1+\frac{2}{\alpha})-\Gamma^2(1+\frac{1}{\alpha})\right]$ | $\sum_{k\geq 0}\frac{(it)^k\beta^k}{k!}\Gamma(1+\frac{k}{\alpha})$ |
| **Cauchy** $\sim Cauchy(\theta,\gamma)$ | $\frac{1}{\pi\gamma}\frac{1}{1+(\frac{x-\theta}{\gamma})^2}, \gamma>0$ | No definida | No definida | $e^{\theta it-\gamma\|t\|}$ |
| $\chi^2_p$ | $\frac{1}{\Gamma(p/2)2^{p/2}}x^{\frac{p}{2}-1}e^{-\frac{x}{2}}, x>0, p\in\mathbb{N}$ | $p$ | $2p$ | $(1-2it)^{-\frac{p}{2}}$ |
| **Doble expon** $\sim DobExp(\mu,\gamma)$ | $\frac{1}{2\gamma}e^{-\frac{\|x-\mu\|}{\gamma}}, \gamma>0$ | $\mu$ | $2\gamma^2$ | $\frac{e^{\mu it}}{1+\gamma^2t^2}$ |
| **Lognormal** $\sim LogN(\mu,\sigma^2)$ | $\frac{1}{\sigma\sqrt{2\pi}}\frac{1}{x}e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}}, x,\sigma>0$ | $e^{\mu+\frac{\sigma^2}{2}}$ | $e^{2(\mu+\sigma^2)}-e^{2\mu+\sigma^2}$ | Sin forma sencilla |

# 1 Notación

$\tilde{X}$ representa un vector, mientras que $\bar{X}$ representa la media.

Para los intervalos de confianza tomamos $t_\beta$ como el $t$ tal que tiene una probabilidad de $\beta$ a la derecha.

# 2 Intro

**Proposition.** *Sea $\tilde{X}$ absolutamente continua con $f_{\tilde{X}}(x_1, \ldots, x_n)$. Sea $g : \mathbb{R}^n \to \mathbb{R}^n \in \mathcal{C}^1$ biyectiva. Si $g(\tilde{X}) = \tilde{Y}$ encontramos $f_{\tilde{Y}}$ de la siguiente forma:*

*Sea $h := g^{-1}$:*

$$f_{\tilde{Y}}(y_1, \ldots, y_n) = f_{\tilde{X}}(h(y_1, \ldots, y_n))|J_h(y_1, \ldots, y_n)|$$

**Definition** (Varianza, covarianza y correlación).

$$\begin{aligned} s_x^2 &= \frac{1}{N}\sum(x_i - \overline{x}_N)^2 \\ s_{xy} &= \frac{1}{N}\sum(x_i - \overline{x}_N)(y_i - \overline{y}_N) \\ r_{xy} &= \frac{s_{xy}}{s_x s_y} \end{aligned}$$

**Definition** (Parámetros a estimar).

$$X \sim F \Rightarrow \theta = \Phi(F)$$

**Definition** (Estimador). $X_1, \ldots, X_n$

$$\hat{\theta} = T(X_1, \ldots, X_n)$$

**Definition** (Función de distribución empírica).

$$F_n(x) = \frac{1}{n}\sum I_{(\infty,x]}(x_i)$$

**Theorem.** $X_1, \ldots, X_n, x \in \mathbb{R}$

1. $E(F_n(x)) = F(x), \quad Var(F_n(x)) = \frac{1}{n}F(x)(1-F(x))$
2. $F_n(x) \to F(x)$ *casi seguro*
3. $\frac{\sqrt{n}(F_n(x)-F(x))}{\sqrt{F(x)(1-F(x))}} \to N(0,1)$ *en distribución*
4. $\frac{\sum(x_i-\mu)}{\sigma\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n-\mu)}{\sigma} \to N(0,1), \quad TCL$

**Theorem** (Glivenko-Cantelli). *$\{X_n\}$ vaiid*

$$\sup_{x\in\mathbb{R}}|F_n(x) - F(x)| \to 0$$

**Theorem.** *Sean $x_1, \ldots, x_n$, $S^2 = \frac{1}{n-1}\sum(x_i - \overline{x})^2$ la varianza muestral*

1. $\min_a \sum(x_i - a)^2 = \sum(x_i - \overline{x})$
2. $(n-1)S^2 = \sum(x_i - \overline{x})^2 = \sum x_i^2 - n\overline{x}^2$

**Definition** (Media muestral). $\overline{X} = \frac{1}{n}\sum X_i$

**Definition** (Varianza muestral). $S^2 = \frac{1}{n-1}\sum(X_i - \overline{X})^2$

**Theorem.** *$X$ con media $\mu$ y varianza $\sigma^2$*

1. $E(\overline{X}) = \mu$
2. $V(\overline{X}) = \frac{\sigma^2}{n}$
3. $E(S^2) = \sigma^2$

**Proposition.**

$$N(\mu_1, \sigma_1^2) + N(\mu_2, \sigma_2^2) = N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

**Theorem.**

$$\psi_{\overline{X}}(t) = (\psi_X\left(\frac{t}{n}\right))^n$$

**Definition** (Distribución $\chi_k^2$). *Sea $X_i \sim N(0,1)$*

$$\chi_k^2 = \sum X_i^2 = \Gamma\left(\frac{n}{2}, \frac{1}{2}\right)$$

**Proposition.** *Sea $\tilde{X} \sim N_p(\tilde{\mu}, \Sigma)$*

$$\tilde{Y} = A\tilde{X} \sim N_p(A\tilde{\mu}, A\Sigma A^t)$$

**Proposition.** *Sea $\tilde{X} \sim N_p(\tilde{\mu}, \Sigma)$*

$$(\tilde{X} - \tilde{\mu})^t\Sigma^{-1}(\tilde{X} - \tilde{\mu}) \sim \chi_p^2$$

**Theorem** (Fisher). *$X_i \sim N(\mu, \sigma^2)$ vaiid entonces*

$$\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right), \quad \frac{(n-1)}{\sigma^2}S^2 \sim \chi_{n-1}^2 \quad \text{indep}$$

**Definition** (t de Student). *Sean $X \sim N(0,1), Y \sim \chi_r^2$ indep*

$$T = \frac{X}{\sqrt{\frac{Y}{r}}} \sim t_r, \quad f_T(t) = \frac{\Gamma(\frac{r+1}{2})}{\sqrt{\pi r}\Gamma\left(\frac{r}{2}\right)}\left(1 + \frac{t^2}{r}\right)^{-\frac{r+1}{2}}$$

*Si $r > 1 \Rightarrow E(T) = 0$, si $r > 2 \Rightarrow Var(T) = \frac{r}{r-1}$*

**Proposition.** *Sean $X_i \sim N(\mu, \sigma)$ vaiid*

$$\frac{\sqrt{n}(\overline{X} - \mu)}{S} \sim t_{n-1}$$

**Definition** (Distribución $F$). *Sean $X \sim \chi_r^2, Y \sim \chi_s^2$ indep*

$$F = \frac{X/r}{Y/s} \sim F_{r,s}, \quad f_F(t) = \frac{r\Gamma((r+s)/2)}{s\Gamma(r/2)\Gamma(s/2)}\frac{(xr/s)^{r/2-1}}{(1 + xr/s)^{(r+s)/2}}$$

$s > 2 \Rightarrow E(F) = \frac{s}{s-2}, \quad s > 4 \Rightarrow Var(F) = \frac{2s^2(r+s-2)}{r(s-2)^2(s-4)}$

**Proposition.** $F^{-1} \sim F_{s,r}, \quad T^2 \sim F_{1,r}$

*Sean $X_1, ..., X_n \sim N(\mu_1, \sigma_1), Y_1, ..., Y_m \sim N(\mu_2, \sigma_2)$ vaiid*

$$\frac{S_X^2/\sigma_1^2}{S_Y^2/\sigma_2^2} \sim F_{n-1,m-1}$$

**Definition** (Localización y escala).

$$X = \mu + \sigma Z, \quad f(x|\mu, \sigma) = \frac{1}{\sigma}f\left(\frac{x-\mu}{\sigma}\right)$$

- $Z \sim f(x) \iff X = \mu + \sigma Z \sim f(x|\mu, \sigma)$
- $X \sim f(x|\mu, \sigma) \iff \frac{X-\mu}{\sigma} \sim f(x)$

**Definition** (Estimador plug-in).

$$F_n(x) = \frac{1}{n}\sum I_{(-\infty,x]}(X_i), \Rightarrow \hat{\theta}_n = \Phi(F_n)$$

**Definition** (Método de los momentos). *Sea $\mu_k = E(X^k)$ y que existe una biyección*

$$\mu_i = g_i(\theta_1, \dots, \theta_k) \iff \theta_i = h_i(\mu_1, \dots, \mu_k)$$

*entonces* $\hat{\theta}_i = h_i(m_1, \dots, m_k)$ *con* $m_j = \frac{1}{n}\sum X^j$

**Definition** (Estimador máximo verosímil).

$$l(\theta, \tilde{x}) = \ln L(\theta, \tilde{x}) = \ln f_{\tilde{x}}(\theta) \Rightarrow \hat{\theta} = \arg\max_{\theta} l(\theta, \tilde{x})$$

**Theorem** (Principio de invariancia). $\psi = \Psi(\theta)$ *con* $\Psi$ *biyectiva* $\Rightarrow \hat{\psi}_{ML} = \Psi(\hat{\theta}_{ML})$.

**Definition** (Kullback-Leiber).

$$D_{K,L}(f\|g) = \int_S \ln\left(\frac{f(x)}{g(x)}\right) f(x)dx \geq 0$$
$$E_f(\ln(f(X))) \geq E_f(\ln(g(X)))$$

**Theorem** (Bayes).

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

# 3 Tema 3

**Definition.** *$T_n(X_1, \dots, X_n)$ estadístico. Llamamos **distribución de muestreo** a la distribución de $T_n$ y **error estándar** a la desviación estándar de $T_n$.*

**Definition** (Sesgo). $B(\hat{\theta}) = E(\hat{\theta}) - \theta$

**Definition** (ECM). $ECM(\hat{\theta}) = E((\hat{\theta}-\theta)^2) = (B(\hat{\theta}))^2 + Var(\hat{\theta})$

**Definition.** *Un estimador $W(\tilde{X})$ de $\theta_F$ es **inadmisible** si* $\exists V(\tilde{X}) : ECM_F(W, \theta_F) \geq ECM_F(V, \theta_F)\ \forall F \in \mathcal{F}, \quad \exists F_0 : ECM_{F_0}(W, \theta_{F_0}) > ECM_{F_0}(V, \theta_{F_0})$.

**Definition** (Eficiencia relativa de W respecto V).

$$ER(\theta_F, W, V) = \frac{Var_F(V)}{Var_F(W)} = \frac{1/Var_F(W)}{1/Var_F(V)} = \frac{Precision(W)}{Precision(V)}$$

**Theorem** (Cramer-Rao). *Sea $W(\tilde{X})$ un estimador insesgado para $\tau(\theta)$*

$$Var(W(\tilde{X})) \geq \frac{\left(\frac{d\tau(\theta)}{d\theta}\right)^2}{E\left[\left(\frac{\partial}{\partial\theta}\log f(\tilde{X}|\theta)\right)^2\right]}$$

**Definition** (Información de Fisher).

$$I_{\tilde{x}}(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\ln f_{\tilde{x}}\right)^2\right] = Var_\theta\left(\frac{\partial}{\partial\theta}\ln f_{\tilde{x}}\right)$$
$$I_{\tilde{x}}(\theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\ln f_{\tilde{x}}\right]$$

**Proposition.** $I_{\tilde{x}}(\theta) = nI_{x_i}(\theta)$

**Theorem** (Alcanzar CR). *$W(\tilde{X})$ estimador de $\tau(\theta)$ alcanza CR* $\iff$

$$\frac{\partial}{\partial\theta}\ln f_{\tilde{x}} = a(\theta)(W(\tilde{X}) - \tau(\theta))$$
$$\iff f_{\tilde{x}}(\tilde{X}) = u(\tilde{X})h(\theta)\exp(W(\tilde{X})k(\theta))$$

**Definition** (Estadístico suficiente). $\frac{f(\tilde{x}|\theta)}{q(T(\tilde{x})|\theta)}$ *no dep. de* $\theta$

**Theorem** (Factorización). *$T(\tilde{X})$ suficiente* $\iff$

$$f(\tilde{x};\theta) = g(T(\tilde{x});\theta)h(\tilde{x})$$

**Definition** (Estadístico suficiente minimal). *$T(X)$ suficiente es minimal $\iff$ para cualquier otro $S(X)$ suficiente, $T(X)$ es función de $S(X)$*

**Theorem.** *$f(\tilde{x}|\theta)$ la verosimilitud. Si*

$$\frac{f(\tilde{x}|\theta)}{f(\tilde{y}|\theta)} = g(\tilde{x}, \tilde{y}, \cancel{\theta}) \iff T(\tilde{x}) = T(\tilde{y})$$

*Entonces $T(x)$ es suficiente minimal*

**Definition** (Completitud). *Si*

$$E_\theta[g(T)] = 0\ \forall\theta \Rightarrow P(g(T) = 0) = 1\ \forall\theta$$

*entonces decimos que $T$ es completo*

**Proposition.** *suficiente completo $\Rightarrow$ suficiente minimal (el recíproco no es cierto)*

**Theorem.** *$X$ familia exponencial*

$$f(x|\theta) = h(x)c(\theta)\exp(\sum g_j(\theta)t_j(x))$$

*entonces el estadístico $T(\tilde{X}) = (T_1(\tilde{x}), \dots, T_k(\tilde{X}))$, donde $T_j(\tilde{X}) = \sum t_j(X_i)$ es suficiente completo*

**Theorem** (Pitman-Koopman-Darmois). *Si el soporte de las distribuciones no depende de $\theta$ y $\exists$ un estadístico suficiente, el modelo es de una familia exponencial.*

**Theorem** (Rao-Blackwell). *$T(\tilde{X})$ suficiente $W(\tilde{X})$ insesgado de $\tau(\theta)$. Definimos $W_T = E_\theta(W|T)$*

1. *$W_T$ función únicamente de $T(X)$*
2. $E_\theta(W_T) = \tau(\theta)\ \forall\theta$
3. $Var_\theta(W_T) \leq Var_\theta(W)\ \forall\theta$

**Proposition.** *Si $W$ es el mejor estimador insesgado de $t(\theta)$ y su varianza es finita, entonces $W$ es único*

**Definition** (Ruido blanco). *Estadístico $U$ tal que* $E_\theta(U) = 0\ \forall\theta$

**Theorem.** *$W$ estimador insesgado de $\tau(\theta)$ $\forall\theta$. $W$ UMVUE $\iff$ $W$ incorrelado con todos los ruidos blancos.*

**Proposition.** *$W$ insesgado de $\tau(\theta)$ UMVUE $\iff$ es función del estadístico minimal suficiente $T$ para $\theta$ y está incorrelado con los estimadores insesgados de $\theta$ que sean función del estadístico $T$ minimal suficiente para $\theta$*

**Theorem** (Lehmann-Scheffé). *$T(\tilde{X})$ suficiente y completo para $\theta$, $W(\tilde{X})$ estimador insesgado cualquiera de $\tau(\theta)$, entonces*

$$W_T(\tilde{X}) = E_\theta(W|T(\tilde{X}))$$

*es UMVUE de $\tau(\theta)$. Si la variancia es finita $\forall\theta \Rightarrow W_T$ es único*

**Proposition.** *$T(\tilde{X})$ suficiente y completo*

1. *$W$ insesgado de $\tau(\theta)$ y función de $T$, entonces $W$ es UMVUE*

2. *Cualquier función de $T$ que tenga esperanza finita es el UMVUE de su esperanza.*

**Definition** (Consistencia en probabilidad). *La sucesión $\hat{\theta}_n = T(\tilde{X})$ es consistente si*

$$\lim P(|\hat{\theta}_n - \theta| < \varepsilon) = 1 \ (\iff \lim \hat{\theta}_n = \theta \text{ en probabilidad})$$

**Theorem.** *Si $\hat{\theta}_n$ verifica*

- $\lim Var(\hat{\theta}_n) = 0 \ \forall\theta$
- $\lim B_\theta(\hat{\theta}_n; \theta) = 0 \ \forall\theta$

*entonces $\hat{\theta}_n$ es consistente en media cuadrática y en probabilidad*

**Theorem.** *$\hat{\theta}_n$ consistente en probabilidad*

1. $a_n \to 1, \ b_n \to 0 \Rightarrow (a_n\hat{\theta}_n + b_n)$ *consistente*
2. *$g$ continua $\Rightarrow g(\hat{\theta}_n)$ consistente para $g(\theta)$*
3. *$\hat{\delta}_n$ estimadores consistentes para $\delta$, $g(\theta, \delta)$ continua $\Rightarrow g(\hat{\theta}_n, \hat{\delta}_n)$ consistente para $g(\theta, \delta)$*

**Definition** (Normalidad asintótica). *$\hat{\theta}_n$ la presenta si $\exists v_n(\theta) \to 0$ no negativos tal que*

$$\frac{(\hat{\theta}_n - \theta)}{\sqrt{v_n(\theta)}} \to_D N(0,1) \Rightarrow \hat{\theta}_n \sim AN(\theta, v_n(\theta))$$

**Definition** (Eficiencia relativa asintótica). *$T_n(\tilde{X}), S_n(\tilde{X})$ estimadores de $\theta$ asintoticamente normales varianza del orden $1/n$*

$$\begin{aligned} &\sqrt{n}(T_n(\tilde{X}) - \theta) \to_D N(0, \sigma_T^2(\theta)) \\ &\sqrt{n}(S_n(\tilde{X}) - \theta) \to_D N(0, \sigma_S^2(\theta)) \\ &ARE(\theta, S_n, T_n) = \frac{\sigma_T^2(\theta)}{\sigma_S^2(\theta)} \end{aligned}$$

**Theorem** (Slutzky). *$X_n \to_D X, Y_n \to_P a$. Entonces*

1. $X_n + Y_n \to_D X + a$
2. $X_n Y_n \to_D aX$
3. *$g(x, y)$ continua $\Rightarrow g(X_n, Y_n) \to_D g(x, y)$*

**Theorem** (Método delta). $a_n \to \infty$

$$\begin{aligned} &a_n(\hat{\theta}_n - \theta) \to_D N(0, \sigma^2(\theta)) \Rightarrow \\ &\Rightarrow a_n(g(\hat{\theta}_n) - g(\theta)) \to_D N(0, (g'(\theta))^2\sigma^2(\theta)) \end{aligned}$$

*Estimador de momentos: $\hat{\theta}_n = \bar{x}_n, \theta = E[X] \Rightarrow$
$a_n = \sqrt{n}, \sigma^2(\theta) = Var(X)$*

**Theorem.** *Si se verifican*

1. *Distinta $\theta \to$ distinta distribución ($\theta$ identificable)*
2. *$\{x : f(x|\theta) > 0\}$ es el mismo $\ \forall\theta$*
3. *$E_{\theta_0}(\log(\frac{f(X|\theta)}{f(X|\theta_0)}))$ existe $\ \forall\theta, \theta_0$*
4. *$\Theta$ es un abierto*
5. *$\frac{\partial f(x|\theta)}{\partial\theta}$ es continua en $\theta$*

*entonces si 1, 2, y 3:*

$$E_{\theta_0}\left[\log\left(\frac{L(\theta|X_n)}{L(\theta_0|X_n)}\right)\right] < 0, \lim P_{\theta_0}\{L(\theta_0|X_n) > L(\theta|X_n)\} = 1$$

*Si además 4 y 5:* $\quad \dfrac{\partial}{\partial\theta}\log L(\theta|X_n) = 0$

**Theorem.** *Si además $\exists\frac{\partial^3}{\partial\theta^3}$ acotada por $K(x)$ tal que $E_\theta(K(X)) \leq k$, $\theta_n$ raíces del score: $\hat{\theta}_n \to_P \theta_0$, entonces*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to_D N(0, \frac{1}{I_X(\theta_0)}), \quad X \sim f(x; \theta_0)$$

**Theorem.** *Los estimadores*

$$O_n = -\frac{\partial^2 \log L(\theta|X_n)}{\partial\theta^2}|_{\theta=\hat{\theta}_n}, \quad E_n = I_{X_n}(\hat{\theta}_n)$$

*divididos por $n$ son estimadores consistentes de $I_X(\theta_0)$*

# 4 Estimación por intervalos

**Definition** (Estimador por intervalos). $[L(\tilde{X}), U(\tilde{X})]$

**Definition** (Probabilidad de cobertura). $P_\theta(\theta \in [L(\tilde{X}), U(\tilde{X})])$

**Definition** (Coeficiente de cobertura (confianza)). $\inf_\theta P_\theta(\theta \in [L(\tilde{X}), U(\tilde{X})])$

**Definition** (Intervalo de confianza). $IC_{1-\alpha}(\theta)$

**Proposition.** *$IC_{1-\alpha}$ de $\theta$ es $[L(X), U(X)] \Rightarrow IC_{1-\alpha}$ de $\tau(\theta)$ es $[\tau(L(X)), \tau(U(X))]$ si es creciente (al revés si decreciente)*

**Definition** (Cantidad pivotal). *$Q(\tilde{X}, \theta)$ es cantidad pivotal si $f_Q$ no depende de $\theta$.*

*e.g. $(\overline{x} - \mu)/\sigma$ en familias de localización y escala.*

**Proposition.** *Sea $A_\alpha : P(Q(\tilde{X}, \theta) \in A_\alpha) = 1-\alpha$ entonces $C_\alpha(\tilde{x}) = \{\theta : Q(\tilde{x}, \theta) \in A_\alpha\}$ es conjunto de confianza $1-\alpha$*

**Proposition.** $X_1, ..., X_n \sim N(\mu, \sigma)$
$IC_{1-\alpha}(\mu) = [\overline{x}_n \mp z_{\alpha/2}\frac{\sigma}{\sqrt{n}}]$ *si $\sigma$ conocida*
$IC_{1-\alpha}(\mu) = [\overline{x}_n \mp t_{n-1,\alpha/2}\frac{S_n}{\sqrt{n}}]$ *si $\sigma$ desconocida*
$IC_{1-\alpha}(\sigma^2) = \left[\frac{(n-1)S_n^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)S_n^2}{\chi^2_{n-1,1-\alpha/2}}\right]$

**Definition** (Cant. pivotal asintótica). $Q_n(\tilde{X}_n, \theta) \to_D Q$
*e.g.*

$$Q_n = \frac{T(\tilde{X}_n) - E^A_\theta(T(\tilde{X}_n))}{\sqrt{V^A_\theta(T(\tilde{X}_n))}} \to_D N(0,1)$$

**Proposition.** $\frac{\hat{\theta}_n - \theta}{\hat{se}(\hat{\theta}_n)} \to_D N(0,1) \Rightarrow [\hat{\theta}_n \mp z_{\alpha/2}\hat{se}(\hat{\theta}_n)]$ *intervalo de confianza $(1-\alpha)$ asintótico para $\theta$*

**Definition** (Intervalo EMV). $\left[\hat{\theta}_n \mp z_{\alpha/2}\sqrt{(\hat{I}_n(\theta))^{-1}}\right]$

$$Q_n^{EMV} = \frac{\hat{\theta}_n - \theta}{\sqrt{(I_n(\theta))^{-1}}} \to_D Z \sim N(0,1)$$

**Definition** (Función score). $S_n = \frac{\partial}{\partial\theta}\log L$

$$Q_n^S = \frac{S_n}{\sqrt{I_n(\theta)}} \to_D Z \sim N(0,1)$$

**Definition** (Precisión). *$\Delta$ es la mitad del intervalo*

# 5 Contrastes de hipótesis

**Definition.** *$H_0$ hipótesis nula, $H_1$ hipótesis alternativa*

**Definition** (Contraste de hipótesis). $\begin{cases} H_0 : F \in \mathcal{F}_0 \\ H_1 : F \in \mathcal{F}_1 \end{cases}$ *es una función de decisión* $D : \mathcal{X}^n \to \{0,1\}, \tilde{x} \mapsto D(x)$

**Definition.** *$R$ región crítica tal que $D(R) = 1$ (rechaza)*

$\alpha = P(\text{error tipo I}) = P(X \in R | H_0 \text{ cierta})$ *(significación)*
$\eta = P(\text{error tipo II}) = P(X \notin R | H_0 \text{ falsa})$

**Definition** (Potencia del test). *$\beta = 1 - \eta$ probabilidad de decidir de forma acertada rechazar $H_0$*

**Definition** (p-valor). *ínfimo de los $\alpha$ para los cuales se rechazaría la hipótesis nula. Probabilidad de que, siendo $H_0$ cierta, se observe otra muestra que sea al menos tan poco favorable a la hipótesis nula como la que se ha observado.*

**Proposition.** $p(\tilde{x}) = \sup_{\theta \in \Theta_0} P_\theta(T(\tilde{X}) \geq T(\tilde{x}))$

**Definition** (Función de potencia).

$$\beta(\theta) = P(X \in R) = \begin{cases} P(\text{Error tipo I}) & \text{si } \theta \in \Theta_1 \\ 1 - P(\text{Error tipo II}) & \text{si } \theta \in \Theta_2 \end{cases}$$

**Definition** (Nivel). $\sup \beta(\theta) \leq \alpha$

**Theorem** (Neyman-Pearson). *$H_0 : \theta = \theta_0$ vs $H_0 : \theta = \theta_1$. Región crítica del contraste más potente de tamaño $\alpha$:*

$$R = \left\{ \tilde{x} : \frac{L(\theta_1; \tilde{x})}{L(\theta_0; \tilde{x})} \geq A_\alpha \right\}$$
$$p - valor = P\left( \frac{L(\theta_1; \tilde{X})}{L(\theta_0; \tilde{X})} \geq \frac{L(\theta_1; \tilde{x})}{L(\theta_0; \tilde{x})} \right)$$

**Definition** (Uniformemente más potente de tamaño $\alpha$).

$$\beta(\theta) \leq \alpha \ \forall \theta \in \Theta_0, \quad \beta(\theta) \text{ máximo } \forall \theta \in \Theta_1$$

**Theorem** (Neyman-Pearson). *$H_0 : \theta = \theta_0$ vs $H_0 : \theta \in \Theta_1$. Si $R(\theta_1)$ (del teorema anterior) no depende de $\theta_1$, la prueba estadística de $R$ es UMP de tamaño $\alpha$.*

**Proposition.** $X \sim N(\mu, \sigma^2), \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$

$$R = \{x : \overline{x} \geq B\}, \text{ con } B = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

**Theorem.** *$f(x;\theta) = c(\theta)h(x)e^{\eta(\theta)t(x)}$ con $\eta(\theta)$ creciente. El test de NP es UMP si $H_1 : \theta > \theta_0$.*

**Definition** (Razón de verosimilitudes).

$$\lambda = \lambda(x) = \frac{\max_{\theta \in \Theta_0} L(\theta; x)}{\max_{\theta \in \Theta} L(\theta; x)} \Rightarrow R = \{x : \lambda(x) \leq A\}$$

**Theorem.** $\begin{cases} H_0 : \theta \in \Theta_0 \\ H_1 : \theta \in \Theta_1 \end{cases}$

$$Q_n = -2\log(\lambda(X_n)) \to_n \chi^2_d, \quad d = \dim(\Theta) - \dim(\Theta_0)$$

*bajo la hipótesis nula*

Dos tests para $\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta \neq \theta_0 \end{cases}$

**Definition** (Test del Score). $S_n(\theta; x_n) = \frac{\partial \log L}{\partial \theta}$

$$T_n^s(X_n) = \frac{S_n(\theta_0; X_n)^2}{I_n(\theta_0)}, \quad S_n(\theta_0; X_n) \approx N(0, I_n(\theta_0))$$

*Rechaza $H_0$ si $T_n^S(x_n) > \chi^2_{1,\alpha}$*

**Definition** (Test de Wald). $\hat{\theta}_n \approx N(\theta_0, (I_n(\theta_0))^{-1})$

$$W_n = (\hat{\theta}_n - \theta_0)^2 I_n(\theta_0) \approx_{en\ H_0} \chi^2_1$$

*Rechaza $H_0$ si $W_n^S(x_n) > \chi^2_{1,\alpha}$*

# 6 Modelo lineal normal

**Definition** (Modelo teórico). *$y_i | x_i \sim N(\beta_0 + \beta_1 x_{1i} + \cdots + \beta_{p-1i}, \sigma^2)$, $e_i = y_i - \hat{y}_i$, $e_i, e_j$ son independientes $\forall i, j$.*

Formulación matricial $Y = X\beta + \epsilon$:

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \epsilon = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{pmatrix} X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{p-11} \\ \vdots & & & \vdots \\ 1 & x_{1i} & \cdots & x_{p-1i} \end{pmatrix}$$

**Theorem** (Minimos cuadrados). $b = (X^T X)^{-1} X^T Y$

Propiedades: $\overline{e} = 0$ y pasa por $(\hat{x}_1, ..., \hat{x}_{p-1}, \hat{y})$

**Proposition.** *Una variable explicativa*

$$b_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}, \quad b_0 = \overline{y} - b_1 \overline{x}$$

**Proposition.** *Con diferente varianza y $V$ matriz diagonal de varianza:*

$$b = (X^T V^{-1} X)^{-1} X^T V^{-1} Y$$

**Definition** (ANOVA decomposition). $SS_T = SS_R + SS_E$

$$SS_T = \sum (y_i - \overline{y})^2, SS_R = \sum (y_i - \hat{y}_i)^2, SS_E = \sum (\hat{y}_i - \overline{y})^2$$
$$s_E^2 = \frac{SS_E}{p-1}, \quad s_R^2 = \frac{SS_R}{n-p}, \quad F = \frac{s_E^2}{s_R^2}$$

$\hat{\sigma^2} = s_R^2$

**Definition** (Coeficiente determinación).

$$R^2 = 100 \frac{SS_E}{SS_T} = 100(1 - \frac{SS_R}{SS_T}), R^2_{adj} = 100(1 - \frac{s_R^2}{s_R^2})$$

**Proposition.** $b | X \sim N(\beta, \sigma^2 (X^T X)^{-1})$

**Proposition.** *Una variable* $Cov(b_0, b_1 | x) = \frac{-\sigma^2 \overline{x}}{\sum (x_i - \overline{x})^2}$

$$b_0 | x \sim N(\beta_0, \frac{\sigma^2}{n} \frac{\sum x_i^2}{\sum (x_i - \overline{x})^2}), \quad b_1 | x \sim N(\beta_1, \frac{\sigma^2}{\sum (x_i - \overline{x})^2})$$

**Proposition.** $s_R^2 | X \sim \frac{\sigma^2}{n-p} \chi^2_{n-p}$

**Proposition.** $Var(b_j|x) = s^2_{b_j}$, $t_j = b_j/s_{b_j}$

$$(b_j \mp t^{\alpha/2}_{n-p} s_{b_j}) \simeq (b_j \mp 2 s_{b_j})$$

*Si 0 en el intervalo, se puede quitar la variable del modelo*

**Definition.** $SS_{R0}$ *suma sin las q variables a retirar*

$$F = \frac{(SS_{R0} - SS_R)/q}{(SS_R)/(n-p)}, \quad \begin{cases} +F \Rightarrow \textit{no retirar} \\ -F \Rightarrow \textit{retirar} \end{cases}$$

*Si se pueden retirar,* $F \sim F_{q,n-p}$.
$F = s_e^2/s_R^2$ *para comparar vs el modelo nulo,* $N(\overline{y}, \sigma^2)$.

**Proposition.** *Intervalo de confianza* $1-\alpha$ *de* $E(y|x_0)$

$$\left(\hat{y}(x_0) \pm t^{\alpha/2}_{n-p} s_R \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum (x_i - \overline{x})^2}}\right)$$

**Proposition.** *Intervalo de predicción* $1-\alpha$ *de* $y(x_0)$

$$\left(\hat{y}(x_0) \pm t^{\alpha/2}_{n-p} s_R \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{\sum (x_i - \overline{x})^2}}\right)$$

El intervalo de 95% podemos tener en cuenta $t^{0.25}_{n-p} \approx 2$

Para más de una variable, $\hat{y}(x_0) \sim N(E(y|x_0), \sigma^2 x_0'(X'X)^{-1}x_0)$ son respectivamente

$$\left(\hat{y}(x_0) \pm t^{\alpha/2}_{n-p} s_R \sqrt{x_0^T (X^T X)^{-1} x_0}\right)$$
$$\left(\hat{y}(x_0) \pm t^{\alpha/2}_{n-p} s_R \sqrt{1 + x_0^T (X^T X)^{-1} x_0}\right)$$

Tipos de errores:

1. *Regular*: $e_i = y_i - \hat{y}_i$, $e = Y - \hat{Y} \sim N(0, \sigma^2(I - H))$ $H = X(X'X)^{-1}X'$ es la projection/hat matrix
2. *Standarized*: $e_i^s = e_i/s_{e_i}$. ($s_{e_i}$ aproxima se) Si el modelo es correcto, 95% deberían estar en $(-2, 2)$.
3. *Deleted*: $e_i^{sd} = (y_i - \hat{y}_{(i)})/s_{e_{(i)}}$. Como los standarized, pero para la $i$-ésima predicción se hace un modelo con todos los datos excepto el $i$. No hace "trampas".

**Definition** (Distance in X-space)**.**

$$H = X(X^TX)^{-1}X^T, \quad h_{ii} = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum (x_i - \bar{x})^2}$$

*If* $h_{ii} > 3p/n$, *the observation i is unusually far. (No indication by itself that model is wrong)*

**Definition** (Degree of outlierness)**.** $|e_i^s|$

**Definition** (Influence on the fitted model)**.** *Se mide con la distancia de Cook:*

$$DC_i = \frac{(\hat{Y} - \hat{Y}_{(i)})'(\hat{Y} - \hat{Y}_{(i)})}{s_R^2 p} = (e_i^s)^2 \frac{h_{ii}}{1 - h_{ii}} \frac{1}{p}$$

*o con* $DFFIT_i = \hat{y}_i - \hat{y}_{(i)}$

Criterios para la selección de modelos:

1. Maximizar $R^2_{adj}$
2. Minimizar $C_p = \frac{SS_R}{s_R^2} + 2p - n$ donde $s_r^2$ es la varianza residual del modelo con todas las variables.
3. Minimizar $AIC = Const + n \log(SS_R) + 2p$.
4. Minimizar $BIC = Const + n \log(SS_R) + p \log(n)$.

Tipos de cross validation:

Sirven para medir el rendimiento del modelo más "honestamente"

1. *One-shot*: Separar los datos en dos, uno para entrenar y el otro para medir. Barato pero poco fiable.
2. *Leave-one-out*: Para cada dato, predecirlo con un modelo entrenado con todos los otros datos. $PRESS = SS_R^{l.one.out} = \sum_i^n (e_i^d)^2$. Muy costoso, pero fiable.
3. *Leave-p-out*: Para cada subconjunto de $p$ datos, predecirlo con un modelo entrenado con los otros. Aún más costoso y fiable.
4. *k-fold*: Separar en $k$ grupos del mismo tamaño y predecir cada uno con un modelo entrenado con los otros. $SS_R^{kfold} = \sum_{i=1}^n (y_i - \hat{y}_{(i)}^{kfold})^2$, $R^2_{kfold} = 100(1 - SS_R^{kfold}/SS_T)$. Es el bueno.

Si hay alguna categórica, con $k$ posibles categorías, se añaden $k-1$ variables indicadoras (una es la *baseline*) y en el modelo se ponen estas indicadoras y el producto de las indicadoras por las variables continuas. Después se intenta simplificar el modelo con los estadísticos $t$ y $F$.

# 7 Modelo de respuesta categórica

Queremos hacer un modelo para una variable binaria $y$ que dependa de variables continuas $x_i$.

En nuestros datos, tenemos $n_i$ experimentos con las explicatorias $x_i$, de los cuales $y_i$ han sido *si*, y $n_i - y_i$ *no*.

**Definition** (Modelo teórico)**.** $y_i|x_i \sim Bin(n_i, \pi(x_i))$

$$\pi(x_i) = \frac{e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{p-1} x_{(p-1)i}}}{1 + e^{\beta_0 + \beta_1 x_{1i} + \ldots + \beta_{p-1} x_{(p-1)i}}}$$

$$\log \frac{\pi(x_i)}{1 - \pi(x_i)} = \log Odds(x_i) = \beta_0 + \beta_1 x_{1i} + \ldots + \beta_{p-1} x_{(p-1)i}$$

Se entrena minimizando el *Pearson statistic* (que da menos importancias a los cercanos a 0 o 1):

$$X^2(Y, \hat{Y}) = \sum_{i=1}^n \frac{(y_i - n_i \hat{\pi}(x_i))^2}{n_i \hat{\pi}(x_i)(1 - \hat{\pi}(x_i))} = \sum_{i=1}^n \left(e_i^P\right)^2$$

o la *deviance* ($\approx SS_R$ en modelos lineales): $ResDev =$

$$2 \sum_{i=1}^n \left(y_i \log \frac{y_i}{n \hat{\pi}(x_i)} + (n_i - y_i) \log \frac{n_i - y_i}{n_i(1 - \hat{\pi}(x_i))}\right) = \sum_{i=1}^n \left(e_i^D\right)^2$$

Si el modelo es correcto, $ResDev \sim \chi^2_{n-p}$ y el 95% de los $e_i^D$ deberían estar en $(-2, 2)$.

# 8 Miscelánea

**Propiedades de la distribución Gamma**
<u>Parametrización con scale</u> $\theta$

$$f(x) = \frac{1}{\Gamma(\alpha)\theta^{\alpha}} x^{\alpha-1} e^{-x/\theta}, \quad E(X) = \alpha\theta, \quad Var(X) = \alpha\theta^2$$

$$\Gamma(\alpha_1, \theta) + \Gamma(\alpha_2, \theta) \sim \Gamma(\alpha_1 + \alpha_2, \theta), \quad c\Gamma(\alpha, \theta) \sim \Gamma(\alpha, c\theta)$$

$$X \sim \Gamma(\alpha, scale = \theta) \Rightarrow \frac{2}{\theta} X \sim \chi^2_{2\alpha}$$

<u>Parametrización con rate</u> $\beta$

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad E(X) = \frac{\alpha}{\beta}, \quad Var(X) = \frac{\alpha}{\beta^2}$$

$$\Gamma(\alpha_1, \beta) + \Gamma(\alpha_2, \beta) \sim \Gamma(\alpha_1 + \alpha_2, \beta), \quad c\Gamma(\alpha, \beta) \sim \Gamma(\alpha, \frac{\beta}{c})$$

$$X \sim \Gamma(\alpha, rate = \beta) \Rightarrow 2\beta X \sim \chi^2_{2\alpha}$$

$$\sum_0^{\infty} n x^{n-1} = \frac{1}{(1-x)^2}$$

$$\Gamma(n) = (n-1)!, \quad B(a, b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$aN(\mu, \sigma^2) + b \sim N(a\mu + b, (a\sigma)^2)$$

$$aN(\mu_1, \sigma_1^2) + bN(\mu_2, \sigma_2^2) = N(a\mu_1 + b\mu_2, (a\sigma_1)^2 + (b\sigma_2)^2)$$

$$X \sim Weibull(\alpha, \beta) \Rightarrow X^{\alpha}/\beta \sim Exp(1)$$