Resumen Estadística

Sea $\mathbb{X}$ el espacio muestral, un elemento $x = (x_1, \ldots, x_n) \in \mathbb{X}$ es una realización de un vector aleatorio $X = (X_1, \ldots, X_n)$ con valores en $\mathbb{X}$. X tiene una ley (distribución) desconocida que pertenece a una familia $\mathcal{P} = \{\mathrm{P}_\theta | \theta \in \Theta\}$ de probabilidad sobre $\mathbb{X}$, con θ los parámetros del modelo y Θ su espacio.

1. Estimación puntual

El problema de estimación puntual es aproximar una función $g(\theta)$ del parámetro θ, es decir

$$g : \Theta \longrightarrow \mathbb{R}^k$$

$$\theta \longrightarrow g(\theta).$$

Def 1. (Estimador) Se define el estimador puntual de $g(\theta)$ como

$$\hat{g} : \mathbb{X} \longrightarrow g(\Theta)$$

$$x \longrightarrow \hat{g}(x).$$

Def 2. (Estimador perfecto) Diremos que un estimador $\hat{g}(X)$ de $g(\theta)$ es perfecto si:

$$\mathrm{P}_\theta[\hat{g}(X) = g(\theta)] = 1, \quad \forall \theta \in \Theta.$$

Def 3. (Error cuadrático medio) Sea g una función a valores reales y $\hat{g}$ tal que $\mathrm{E}_\theta(\hat{g}(X)^2) < +\infty, \quad \forall \theta \in \Theta$. Se define el error cuadrático medio como:

$$\begin{aligned}\mathrm{ecm}_\theta(\hat{g}) &= \mathrm{E}_\theta[(\hat{g}(X) - g(\theta))^2] \\ &= \mathrm{Var}_\theta(\hat{g}(X)) + (\mathrm{E}_\theta(\hat{g}(X)) - g(\theta))^2.\end{aligned}$$

Entonces, decimos que $\hat{g}$ es mejor que $\tilde{g}$ en el sentido del error cuadrático medio ssi $\mathrm{ecm}_\theta(\hat{g}) \leq \mathrm{ecm}_\theta(\tilde{g})$.

Def 4. (Insesgado) Se dice que $\tilde{g}$ es un estimador insesgado para $g(\theta)$ si:

$$\mathrm{E}_\theta(\hat{g}(X)) = g(\theta), \quad \forall \theta \in \Theta.$$

En este caso $\mathrm{ecm}_\theta(\hat{g}) = \mathrm{Var}_\theta(\hat{g}(X))$.

Prop 0. Sea $\mathcal{P} = \{\mathrm{P}_\theta | \theta \in \Theta\}$ una familia paramétrica y $X = (X_1, X_2, \ldots, X_n)$ un vector aleatorio tal que $\mathbb{E}_\theta(X_i) = \mu(\theta), \mathrm{Var}_\theta(X_i) = \sigma^2(\theta) < \infty$. Entonces $\overline{X} = \hat{\mu}(x) = \frac{1}{n}\sum_{i=1}^n x_i$ es estimador insesgado de $\mu(\theta)$ y $\widehat{\sigma^2}(x) = \frac{1}{n-1}\sum_{i=1}^n (x_i - \hat{\mu}(x))^2$ es estimador insesgado de $\sigma^2(\theta)$.

Def 5. (EIVUM) Diremos que g^* es un EIVUM (estimador insesgado de varianza uniformemente mínima) para $g(\theta)$ si

$$\mathrm{Var}_\theta(g^*(X)) \leq \mathrm{Var}_\theta(\hat{g}(X)), \qquad \forall \theta \in \Theta, \forall \hat{g}$$

con $\hat{g}$ estimador insesgado de $g(\theta)$.

Def 6. (Estadístico) Una función $T(X)$ del vector aleatorio $X = (X_1, \ldots, X_n)$ es un estadístico, es decir, T es una función de $\mathbb{X}$ a $\mathbb{R}^k$.

Def 7. (Estadístico suficiente) Diremos que $T(X)$ es un estadístico suficiente para la familia $\mathcal{P} = \{\mathrm{P}_\theta | \theta \in \Theta\}$ si la ley de X condicional a $T(X)$ no depende de θ. Es decir, si existe una función H tal que

$$\mathrm{P}_\theta[X \in A \,|\, T(X) = t] = H(A, t),$$

donde $A \in \mathbb{X}$ y H no depende de θ.

Teo 1. (Factorización) Consideremos el modelo $\mathcal{P} = \{\mathrm{P}_\theta | \theta \in \Theta\}$ y $T(X)$ un estadístico. Entonces $T(X)$ es suficiente para $\mathcal{P}$ ssi

$$f_\theta(X) = g_\theta(T(X))\, h(X)$$

donde $f_\theta(X)$ representa la densidad del vector X, $g_\theta(\cdot)$ es una función que depende de θ y $h(\cdot)$ es una función que no depende de θ.

Teo 2. (Rao-Blackwell) Sea $\mathcal{P} = \{\mathrm{P}_\theta | \theta \in \Theta\}$ una familia paramétrica, $g(\theta)$ una función del parámetro θ, $\tilde{g}(X)$ un estimador insesgado de $g(\theta)$. Sea $T(X)$ un estadístico suficiente para $\mathcal{P}$. Definimos

$$\hat{g}(t) = \mathrm{E}_\theta[\tilde{g}(X) \,|\, T(X) = t].$$

Entonces $\hat{g}(T(X))$ es un estimador insesgado para $g(\theta)$ y

$$\mathrm{Var}_\theta(\hat{g}(T(X))) \leq \mathrm{Var}_\theta(\tilde{g}(X)), \quad \forall \theta \in \Theta.$$

Def 8. (Estadístico completo) Diremos que $T(X)$ es completo para la familia $\{\mathrm{P}_\theta | \theta \in \Theta\}$ si no existen estimadores insesgados de 0 no triviales que sean función de $T(X)$, es decir,

$$\mathrm{E}_\theta[f(T(X))] = 0 \ \ \forall \theta \in \Theta \Longrightarrow f = 0 \ \text{ ctp.}$$

Prop 1. Sea $T(X)$ estadístico suficiente y completo para $\{\mathrm{P}_\theta | \theta \in \Theta\}$. Sea $\tilde{g}(X)$ un estimador insesgado para $g(\theta)$, definimos

$$\hat{g}(t) = \mathrm{E}_\theta(\tilde{g}(X) \,|\, T(X)).$$

Entonces $\hat{g}(T(X))$ es el EIVUM de $g(\theta)$.

Def 9. (Familia exponencial) Consideremos una familia $\{\mathrm{P}_\theta | \theta \in \Theta\}$ sobre $\mathbb{X}$ y supongamos que P_θ tienen densidad f_θ. Si f_θ se puede escribir como

$$f_\theta(X) = C(\theta)\, e^{\sum_{i=1}^k Q_i(\theta) t_i(X)}\, h(X)$$

donde los $t_i(X)$ son funciones de X, no de θ, y $Q_i(\theta)$ son funciones solo de θ. Entonces diremos que esta familia es de tipo exponencial.

Prop 2. Si f_θ se descompone como

$$f_\theta(X) = C(\theta)\, e^{\sum_{i=1}^k \theta_i\, t_i(X)}\, h(X)$$

donde $\theta = (\theta_1, \ldots, \theta_k)$. Entonces $t(X) = (t_1(X), \ldots, t_k(X))$ es un estadístico suficiente completo si Θ contiene un rectángulo k-dimensional.

Teo 3. (Cota de Cramer-Rao) Supongamos que la familia $\{\mathrm{P}_\theta | \theta \in \Theta\}$ posee densidades $f_\theta(X)$ derivables. Consideremos una función $g(\theta)$ del parámetro $\theta \in \Theta \subseteq \mathbb{R}$ (g derivable). Si $\hat{g}(X)$ es un estimador insesgado de $g(\theta)$, entonces

$$\mathrm{Var}_\theta(\hat{g}(X)) \geq \frac{(g'(\theta))^2}{I_\theta}$$

con I_θ la informción de Fisher: $I_\theta = \mathrm{E}_\theta\left(\left(\frac{\partial \log f_\theta(X)}{\partial \theta}\right)^2\right)$. Bajo regularidad se tiene $I_\theta = -\mathrm{E}_\theta\left(\frac{\partial^2 \log f_\theta(X)}{\partial \theta^2}\right)$. Todo estimador de Cramer-Rao es EIVUM, pero no todo EIVUM alcanza la cota de Cramer-Rao.

Def 10. (Función de verosimilitud) Consideremos la familia $\{\mathrm{P}_\theta | \theta \in \Theta\}$ con densidades $f_\theta(X)$. La función de verosimilitud se define como

$$L : \Theta \longrightarrow \mathbb{R}^+$$

$$\theta \longrightarrow f_\theta(x) \quad \text{con } x \in \mathbb{X} \text{ fijo.}$$

Def 11. (EVM) Un estimador de máxima verosimilitud $\hat{\theta}$ es un elemento de Θ tal que

$$f_{\hat{\theta}(x)}(x) = \sup_{\theta \in \Theta} f_\theta(x) \quad \forall x.$$

Obs. Para maximizar es útil aplicar logaritmo y luego derivar, en caso que se pueda.

Algunas propiedades del EMV son:

(1) Los EMV no son insesgados en general, pero en algunos casos se puede corregir el sesgo.

(2) El EMV no es único.

(3) Bajo regularidad, el EMV es asintóticamente insesgado, es decir, si $\hat{\theta}_n$ es el EMV de θ, entonces $\mathrm{E}_\theta(\hat{\theta}_n) \longrightarrow \theta$.

(4) Bajo regularidad, la convergencia fuerte (débil) significa que la sucesión de v.a. $\hat{\theta}_n$ converge c.s. (y en probablidad) a θ.

(5) Bajo regularidad, en algunos casos es posible mostrar que $\hat{\theta}_n$ (EMV) es asintóticamente normal, es decir,

$$\frac{\hat{\theta}_n - \theta}{\sqrt{I_\theta^{-1}}} \longrightarrow \mathcal{N}(0,1).$$

(6) El EMV suele ser asintóticamente eficiente. La eficiencia se define como

$$\mathrm{ef}_\theta(\hat{\theta}_n) = \frac{I_\theta^{-1}}{\mathrm{Var}_\theta(\hat{\theta}_n)} \leq 1.$$

Se dice que $\hat{\theta}_n$ es asintóticamente eficiente si $\mathrm{ef}_\theta(\hat{\theta}_n) \longrightarrow 1$.

(7) El EMV es función de cualquier estadístico suficiente, pues cumple el teorema de factorización: $f_\theta(X) = g_\theta(T(X))\, h(X)$.

2. Intervalos de confianza

Se desea estimar una función del parámetro mediante un conjunto de valores plausibles (conjunto de confianza).

Def 12. (Intervalo de confianza) Sea $\{\mathrm{P}_\theta | \theta \in \Theta\}$ una familia paramétrica y $g(\theta)$ función del parámetro. Buscamos un conjunto $S(X)$ tal que $g(\theta) \in S(X)$ con cierta confianza, es decir

$$\mathrm{P}_\theta[g(\theta) \in S(X)] \geq 1 - \alpha, \quad \forall\, \theta \in \Theta.$$

Diremos que $S(X)$ es una región de confianza con nivel $1-\alpha$ para $g(\theta)$, con $\alpha \in [0,1]$. En el caso en que $g(\theta) \in \mathbb{R}$, $S(X)$ son intervalos acotados $S(X) = [L(X), R(X)]$ tales que

$$\mathrm{P}_\theta[L(X) \leq g(\theta) \leq R(X)] \geq 1 - \alpha, \quad \forall\, \theta \in \Theta.$$

Def 13. (Pivote) Un pivote es una función $T : \mathbb{X} \times \Theta \longrightarrow \mathbb{R}$ medible tal que

(1) La ley de $T(X,\theta)$ no depende de θ.

(2) $T(X,\theta)$ es función monótona de $g(\theta)$.

Pivotes importantes para intervalos de confianza

Si $(X_1, \ldots, X_n)$ son i.i.d. $\mathcal{N}(\mu, \sigma^2)$, entonces

- σ conocido: $T(X,\mu) = \dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim \mathcal{N}(0,1)$, pivote para $g(\mu) = \mu$.
- σ desconocido: $T(X,\theta) = \dfrac{\overline{X}_n - \mu}{\hat{\sigma}/\sqrt{n}} \sim t(n-1)$, pivote para $g(\theta) = \mu$.
- μ conocido: $T(X,\sigma) = \displaystyle\sum_{i=1}^n \frac{(X_i - \mu)^2}{\sigma^2} \sim \mathcal{X}^2(n)$, pivote para $g(\sigma) = \sigma^2$.
- μ desconocido: $T(X,\theta) = \displaystyle\sum_{i=1}^n \frac{(X_i - \overline{X}_n)^2}{\sigma^2} \sim \mathcal{X}^2(n-1)$, pivote para $g(\theta) = \sigma^2$.

Si $(X_1, \ldots, X_n)$ i.i.d. uniforme en $[0,\theta]$, entonces

$$T(X,\theta) = \frac{\max\{X_1, \ldots, X_n\}}{\theta}$$

es pivote para θ, pues $\dfrac{X_i}{\theta} \sim U[0,1]$.

3. Estadística Bayesiana

Tenemos un parámetro θ que vive en el espacio Θ y observamos un vector X con valores en $\mathbb{X}$. La familia de probabilidades está caracterizada por una familia de densidades condicionales $\{f(X|\theta) : \theta \in \Theta\}$.

Def 21. (Densidad a priori y posteriori) Definimos la densidad $\pi(\theta)$ como la densidad a priori, que corresponde a la opinión que se tiene del parámetro antes de observar X. Entonces se define una probabilidad en $\mathbb{X} \times \Theta$ a través de

$$f(X,\theta) = f(X|\theta)\, \pi(\theta).$$

Definimos la densidad a posteriori como

$$f(\theta|X) = \frac{f(X|\theta)\, \pi(\theta)}{f(X)} = \frac{f(X|\theta)\, \pi(\theta)}{\int_\Theta f(X|\theta)\, \pi(\theta)\, d\theta}$$

que corresponde a la opinión que se tiene sobre el parámetro después de observar X. Los parámetros de la ley a priori se llaman hiperparámetros.

Def 22. (Estimador MAP) En la estimación Bayesiana, se busca el estimador de $g(\theta)$ con mayor densidad a posteriori. Entonces, se define el estimador máximo a posteriori de θ como

$$\hat{\theta}_{MAP}(X) \in \mathrm{argmax}_{\theta \in \Theta} f(\theta|X).$$

También definimos la esperanza a posteriori como

$$\hat{\theta}_B(X) = \int \theta\, f(\theta|X)\, d\theta = \mathrm{E}(\theta|X).$$

Def 23. (Conjugadas) Sea $\{f(X|\theta) : \theta \in \Theta\}$ familia paramétrica y P familia de densidades a priori. Se dice que ambas familias están conjugadas si para todo $\pi \in P$ se tiene que $f(\theta|X) \in P$.

Def 24. (Densidad impropia) Sea $\pi(\theta)$ densidad a priori, decimos que es impropia si $\int_\Theta \pi(\theta)\, d\theta \neq 1$. Solo se puede suponer que $\pi(\theta) \geq 0$.

Def 25. (Densidad no informativa) Sea $\pi(\theta)$ densidad a priori, decimos que no es informativa si es igual para todo $\theta \in \Theta$.

Def 26. (Error cuadrático medio) Sea g una función a valores reales y $\tilde{g}$. Se redefine el error cuadrático medio como

$$\mathrm{ecm}_\theta(\tilde{g}) = \int_{\mathbb{X}} (\tilde{g}(x) - g(\theta))^2 f(x|\theta)\, dx.$$

Si $\hat{g}$ es óptimo, entonces tiene menor ecm_θ.

Cuadro 1: **Resumen distribuciones**

nombre	parámetros	notación	tipo	soporte	distribución	esperanza	varianza	f.g.m.
Bernoulli	$p \in (0,1)$	$X \sim \mathrm{Bernoulli}(p)$	discreta	$\{0,1\}$	$p_X(0) = 1-p$, $p_X(1) = p$	p	$p(1-p)$	$1-p+pe^t$
binomial	$n \in \mathbb{N}^*$, $p \in (0,1)$	$X \sim \mathrm{bin}(n,p)$	discreta	$\{0,1,\dots,n\}$	$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$	np	$np(1-p)$	$(1-p+pe^t)^n$
geométrica	$p \in (0,1)$	$X \sim \mathrm{geom}(p)$	discreta	$\{1,2,3,\dots\}$	$p_X(k) = (1-p)^{k-1} p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\frac{pe^t}{1-(1-p)e^t}$
binomial negativa	$r \in \mathbb{N}^*$, $p \in (0,1)$	$X \sim \mathrm{BN}(r,p)$	discreta	$\{r, r+1, \dots\}$	$p_X(k) = \binom{k-1}{r-1} (1-p)^{k-r} p^r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$\left(\frac{pe^t}{1-(1-p)e^t}\right)^r$
Poisson	$\lambda > 0$	$X \sim \mathrm{Poisson}(\lambda)$	discreta	$\{0,1,2,\dots\}$	$p_X(k) = e^{-\lambda} \frac{\lambda^k}{k!}$	λ	λ	$e^{\lambda(e^t-1)}$
uniforme	$a, b \in \mathbb{R}$, $a < b$	$X \sim \mathrm{unif}(a,b)$	continua	$[a,b]$	$f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{tb}-e^{ta}}{t(b-a)}$
exponencial	$\lambda > 0$	$X \sim \exp(\lambda)$	continua	$[0,\infty)$	$f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{[0,\infty)}(x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	$\frac{\lambda}{\lambda-t}\ \forall t < \lambda$
normal	$\mu \in \mathbb{R}$, $\sigma > 0$	$X \sim \mathcal{N}(\mu, \sigma^2)$	continua	$\mathbb{R}$	$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$	μ	σ^2	$e^{\mu t + \frac{1}{2}\sigma^2 t^2}$
gamma	$\theta > 0$, $\lambda > 0$	$X \sim \mathrm{gamma}(\theta, \lambda)$	continua	$[0,\infty)$	$f_X(x) = \frac{\lambda e^{-\lambda x} (\lambda x)^{\theta-1}}{\Gamma(\theta)} \mathbb{1}_{[0,\infty)}(x)$	$\frac{\theta}{\lambda}$	$\frac{\theta}{\lambda^2}$	$\left(\frac{\lambda}{\lambda-t}\right)^\theta\ \forall t < \lambda$