

Esadística

Aniol Garcia i Serrano

Febrer 2021

1 Model estadístic

1.1 Definició

Sigui $(\Omega, \mathcal{F})$ un espai mesurable. Un **model estadístic** és una terna $(\Omega, \mathcal{F}, \mathcal{P})$ en què $(\Omega, \mathcal{F})$ és un espai mesurable i $\mathcal{P}$ una família de probabilitats sobre $(\Omega, \mathcal{F})$ ($\mathcal{P}$ és un conjunt de probabilitats, que són les possibles que podem tractar. Per exemple, si estudiem la llei d'un dau, $\mathcal{P}$ serien totes les variables aleatòries discretes que prenen valors $\{1, 2, 3, 4, 5, 6\}$).

El model s'anomena paramètric si $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$, amb $\Theta \subseteq \mathbb{R}^d$ (espai de paràmetres). (És a dir, que coneixem com és la llei però no coneixem els paràmetres d'aquesta llei).

Observació 1

Una família finita $\mathcal{P}$ sempre és parametritzable.

1.2 Conceptes associats a un model estadístic

Sigui $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$ un model estadístic paramètric i un vector aleatori $X = (X_1, \dots, X_n): \tilde{\Omega} \rightarrow \Omega \subseteq \mathbb{R}^n$ amb llei desconeguda P_θ sobre $(\Omega, \mathcal{F})$. La **funció de versenblança** és

$$L: \Omega \times \Theta \rightarrow \mathbb{R}_+$$
$$(x, \theta) \mapsto L(x, \theta) = \begin{cases} P_\theta(\{x\}) & P_\theta \text{ discreta} \\ f_\theta(x) & P_\theta \text{ abs. contínua amb densitat } f_\theta \end{cases}$$

Observació 2

1. Per donar la definició general ens falten nocions de teoria de la mesura.
2. Si la mostra és aleatòria simple (variables aleatòries independents identi-

cament distribuïdes)

$$L(x, \theta) = \begin{cases} \prod_{i=1}^n P_{\theta}(\{x_i\}) \\ \prod_{i=1}^n f_{\theta}(x_i) \end{cases}$$

3. La notació no reflecteix la dimensió.

4. Fixat $\theta \in \Theta$, $L(x, \theta)$ caracteritza la llei del vector.

Un model $(\Omega, \mathcal{F}, \{P_{\theta}, \theta \in \Theta\})$ és **exponencial** si existeixen funcions reals mesurables $\Psi, \Phi_1, \dots, \Phi_r: \Omega \rightarrow \mathbb{R}$ i funcions reals $D, C_1, \dots, C_r: \Theta \rightarrow \mathbb{R}$ tal que

$$L(x, \theta) = \exp \left\{ \sum_{k=1}^r C_k(\theta) \Phi_k(x) + D(\theta) + \Psi(x) \right\}$$

o bé

$$L(x, \theta) = \tilde{D}(\theta) \tilde{\Psi}(x) \exp \left\{ \sum_{k=1}^r C_k(\theta) \Phi_k(x) \right\}$$

per a $\tilde{D}(\theta), \tilde{\Psi} > 0$.

Observació 3

1. Aquestes funcions no són úniques.

2. En el cas que $\dim(\Theta) = 1$,

$$L(x, \theta) = \exp \{ C(\theta) \Phi(x) + D(\theta) + \Psi(x) \}$$

1.3 Suficiència

Una de les idees més atractives consisteix en resumir o reduir les observacions o dades que tenim sense perdre informació.

Tenim $(\Omega, \mathcal{F}, \{P_{\theta}, \theta \in \Theta\})$ un model estadístic paramètric i un vector $X = (X_1, \dots, X_n): \tilde{\Omega} \rightarrow \Omega \subseteq \mathbb{R}^n$. Estudiarem transformacions de la mostra que conservin tota la informació. Amb aquesta idea en ment, anomenarem **estadístic** a tota aplicació mesurable

$$T: (\Omega, \mathcal{F}) \rightarrow (\mathbb{R}^m, \mathcal{B}(\mathbb{R}^m))$$

amb $m \geq 1$ enter.

Definició 1

Un estadístic T és **suficient** si la llei del vector X condicionada per l'estadístic T no depèn de θ (És a dir, en el cas discret, T és suficient si $P_{\theta}(X = x | T = t)$ no depèn de θ).

Observació 4

Intuitivament un estadístic serà suficient si la coneixença de la mostra no aporta cap informació addicional de θ que la donada per T .

Teorema 1 (Teorema de factorització de Neyman-Fisher)

Sigui $T: \Omega \rightarrow \mathbb{R}^m$ un estadístic. T és suficient $\iff L(x, \theta) = \Psi(T(x), \theta) \cdot h(x)$, $\forall (x, \theta)$ amb $\Psi: \mathbb{R}^m \times \Theta \rightarrow \mathbb{R}_+$ i $h: \Omega \rightarrow \mathbb{R}_+$.

1.4 Informació

Considerem $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$ un model estadístic paramètric, Θ un interval obert, $X = (X_1, \dots, X_n): \tilde{\Omega} \rightarrow \Omega$ un vector i $L(x, \theta)$ la funció de versemblança associada.

Volem donar una eina que mesuri la informació que tenim del paràmetre. Sembla coherent utilitzar la funció de versemblança per saber la informació que dona la mostra del model. Una bona mesura serà la sensibilitat que la mostra reflecteix en front les variacions del paràmetre. Així tindrem en compte el "score", $sc(\theta) = \partial_\theta \ln L(x, \theta)$ per estudiar la seva variància.

Definició 2

*Un model estadístic és **regular** si:*

1. $\{x \in \Omega, L(x, \theta) > 0\} \subseteq \Omega$ no depèn de θ . (Sovint això es complirà si la funció de versemblança no conté indicadors respecte θ , en particular sempre funcionarà per les exponencials)
2. Existeixen $\partial_\theta \ln(L(x, \theta))$ i $\partial_\theta^2 \ln(L(x, \theta))$, i per a cada $\tilde{\theta} \in \Theta$, existeix $h: \Omega \rightarrow \mathbb{R}_+$ amb $\sum_{x \in \Omega} h(x) < +\infty$ (en el cas de les variables aleatòries discretes) o $\int_\Omega h(x) dx < +\infty$ (en el cas de les absolutament contínues) tal que $\forall \theta \in \text{Ent}(\tilde{\theta})$ (θ en un entorn de $\tilde{\theta}$), $|\partial_\theta L(x, \theta)| + |\partial_\theta^2 L(x, \theta)| < h(x)$. (El que fa això és permetre, en certa manera, poder intercanviar la integral de $\int f(x, y) dy$ per ∂_x quan ens interressi i assegurar l'existència del moment de segon ordre)
3. $\forall \theta \in \Theta$, $0 < E_\theta\{|\partial_\theta \ln(L(x, \theta))|^2\} < +\infty$

Observació 5

1. Per a alguns resultats es pot afeblir (2).
2. Les condicions (2) i (3) són tècniques (és a dir, que en els problemes vindran més o menys donades i.e. no les demostrarem i només ens haurem de preocupar de controlar (1)).

Definició 3

En un model estadístic regular, la informació de Fisher és

$$I(\theta) = E_\theta\{|\partial_\theta \ln L(x, \theta)|^2\} = E_\theta\{sc(\theta)^2\} = \begin{cases} \sum_{x \in \Omega} |\partial_\theta \ln(L(x, \theta))|^2 L(x, \theta) & P_\theta \text{ discreta} \\ \int_\Omega |\partial_\theta \ln(L(x, \theta))|^2 L(x, \theta) dx & P_\theta \text{ absolutament contínua} \end{cases}$$

Proposició 1

Si tenim un model estadístic regular,

$$I(\theta) = 0 \iff sc(\theta) = 0, \forall \theta \in \Theta \text{ quasi segurament}$$

Demostració. Una variable aleatòria no negativa té esperança zero si i només si és zero quasi segurament. $\square$

Proposició 2

Si un model estadístic regular. Aleshores $E_\theta[\partial_\theta \ln L(x, \theta)] = 0$.

Demostració. Farem la demostració en el cas absolutament continu:

$$E_\theta[\partial_\theta \ln L(x, \theta)] = \int_{\Omega} \partial_\theta \ln L(x, \theta) \cdot L(x, \theta) dx = \int_{\Omega} \frac{1}{L(x, \theta)} \cdot \partial_\theta L(x, \theta) \cdot L(x, \theta) dx \stackrel{(b)}{=} \partial_\theta \int_{\Omega} L(x, \theta) dx = \partial_\theta 1 = 0$$

El cas discret funciona igual canviant integrals per sumatoris sobre Ω . $\square$

Corol·lari 1

Si un model estadístic és regular, aleshores

$$I(\theta) = \text{Var}_\theta(\partial_\theta \ln L(x, \theta)) = \text{Var}_\theta(\ln L(x, \theta))$$

Proposició 3

Si un model estadístic és regular, aleshores

$$I(\theta) = -E_\theta[\partial_\theta^2 \ln L(x, \theta)]$$

Demostració. Ho farem en el cas discret:

$$\begin{aligned} E_\theta[\partial_\theta^2 \ln L(x, \theta)] &= \sum_{x \in \Omega} \partial_\theta^2 \ln L(x, \theta) \cdot L(x, \theta) = \sum_{x \in \Omega} \partial_\theta \left(\frac{1}{L(x, \theta)} \partial_\theta L(x, \theta) \right) \cdot L(x, \theta) = \\ &= - \sum_{x \in \Omega} \frac{1}{L(x, \theta)^2} (\partial_\theta L(x, \theta))^2 L(x, \theta) + \sum_{x \in \Omega} \frac{1}{L(x, \theta)} \partial_\theta^2 L(x, \theta) = \\ &= - \sum_{x \in \Omega} \frac{1}{L(x, \theta)^2} (\partial_\theta L(x, \theta))^2 L(x, \theta) + \sum_{x \in \Omega} \partial_\theta^2 L(x, \theta) = \\ &= -E_\theta[|\partial_\theta \ln L(x, \theta)|^2] + \partial_\theta \sum_{x \in \Omega} L(x, \theta) \end{aligned}$$

Aquest últim sumand sabem que és 0 per la proposició anterior. $\square$

1.5 Relació entre exponencial i regular

Proposició 4

Suposem $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$ un model estadístic uniparamètric amb $L(x, \theta) = \exp\{C(\theta)\phi(x) + D(\theta) + \Psi(x)\}$. Si Θ és un interval de $\mathbb{R}$ i C, D dues funcions dues vegades contínuament diferenciables amb $C'(\theta) \neq 0, \forall \theta$ i $E_\theta(\phi(x)^2) < +\infty, \forall \theta$. Aleshores el model és regular.

2 Estimació puntual

Considerem un model estadístic paramètric $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$ amb $\theta \in \mathbb{R}^d$ i n observacions independents d'una variable aleatòria X amb llei P_θ .

L'objectiu d'aquesta secció és, a partir de les observacions $x_1, \dots, x_n$, donar un valor aproximat del paràmetre desconegut o d'una funció del paràmetre $g: \Theta \rightarrow \mathbb{R}^k$. Un **estimador** de $g(\theta)$ és un estadístic $T: \Omega \rightarrow \mathbb{R}^k$ utilitzat amb aquesta finalitat.

2.1 Introducció

Introduïrem un cas general i després parlarem del cas que tractem durant aquest curs.

Considerem una funció mesurable

$$\begin{aligned} W: \mathbb{R}^k \times \Omega &\rightarrow \mathbb{R}_+ \\ (t, \theta) &\mapsto W(t, \theta) \end{aligned}$$

que representa l'error que estem cometent en proposar t com a valor de θ o $g(\theta)$. S'anomena **funció de pèrdua** associada a t (ha de complir diverses condicions que no veurem en profunditat, com per exemple que $W(t, \theta) = 0$ si $t = g(\theta)$).

Definició 4

Donada una funció de pèrdua W i un estadístic T , definirem una **funció de risc** associada a T com

$$\begin{aligned} R_T: \Theta &\rightarrow \mathbb{R} \\ \theta &\mapsto E_\theta\{W(T, \theta)\} \end{aligned}$$

Només està definida per a aquells $\theta \in \Theta$ tal que $R_T(\theta) < +\infty$. Si tenim una funció de pèrdua W i dos estadístics T, S amb funció de risc finita per a qualsevol $\theta \in \Theta$, direm que T és uniformement millor que S si $R_T(\theta) \leq R_S(\theta)$, $\forall \theta \in \Theta$.

Definició 5

Tenim W i ξ una classe d'estadístics que tenen la funció de risc finita per a qualsevol $\theta \in \Theta$. Direm que T és **òptim** dins la classe de ξ si

$$R_T(\theta) \leq R_S(\theta), \forall \theta \in \Theta, \forall S \in \xi$$

Trobar estimadors òptims dins una classe ξ és un problema complex i només el resolldrem quan $W(\theta, t) = |g(\theta) - t|^2$.

Així, en el nostre cas, la funció de risc associada a T és

$$R_T(\theta) = E_\theta[|g(\theta) - T|^2]$$

En el cas uniparamètric tenim

$$\begin{aligned} R_T(\theta) &= E_\theta [|T - g(\theta)|^2] = \\ &= E_\theta [|T - E_\theta T|^2] + E_\theta [|E_\theta T - g(\theta)|^2] + 2E_\theta [(T - E_\theta T)(E_\theta T - g(\theta))] = \\ &= \text{Var}_\theta(T) + (E_\theta T - g(\theta))^2 \end{aligned}$$

2.2 Estimadors sense biaix

Definició 6

Un estadístic T és **integrable** si $E_\theta(|T|) < +\infty$, $\forall \theta \in \theta$.

Definició 7

El **Biaix** d'un estimador integrable T respecte a $g(\theta)$ és $b_T(\theta) = E_\theta T - g(\theta)$. Si $b_T(\theta) = 0$, l'estimador s'anomena sense biaix.

Observació 6

1. Si T_1, T_2 són estimadors sense biaix de $g(\theta)$, aleshores $\lambda T_1 + (1 - \lambda)T_2$, $\forall \lambda \in [0, 1]$ són estimadors sense biaix de $g(\theta)$.
2. No necessàriament existeixen.
3. Siguin $X_1, \dots, X_n$ variables aleatòries idènticament distribuïdes (no cal que siguin independents) amb $E_\theta(|X_1|^r) < +\infty$, $\forall \theta, r \geq 1$. Aleshores $T = \frac{1}{n} \sum_{i=1}^n X_i^r$ és un estimador sense biaix de $g(\theta) = E_\theta(X_1^r)$.
4. Com a cas particular, $\bar{X}_n$ (la mitjana) és un estimador sense biaix de $E_\theta(\bar{X}_n) = E_\theta(X_1) = g(\theta)$.
5. Si $X_1, \dots, X_n$ són variables aleatòries idènticament distribuïdes amb moment de segon ordre, aleshores la variància mostral és un estimador esbiaixat de $g(\theta) = \text{Var}_\theta(X_1)$:

$$\begin{aligned} E_\theta S_n^2 &= \frac{1}{n} \sum_{i=1}^n E_\theta [(X_i - \bar{X}_n)^2] = \\ &= \frac{1}{n} \left[\sum_i [(X_i - E_\theta X_1)^2] \right] - 2 \sum_i E_\theta [(X_i - E_\theta X_1)(\bar{X}_n - E_\theta X_1)] + \sum_i E_\theta [(\bar{X}_n - E_\theta X_1)^2] = \\ &= \frac{1}{n} [n \text{Var}_\theta(X_1) - n E_\theta [(\bar{X}_n - E_\theta X_1)^2]] = \frac{1}{n} [n \text{Var}_\theta(X_1) - \text{Var}_\theta(X_1)] = \frac{n-1}{n} \text{Var}_\theta(X_1) \end{aligned}$$

2.3 Estimadors UMV (uniformement de mínima variància)

Dins la classe de ξ_g dels estimadors de $g(\theta)$ sense biaix i de quadrat integrable tenim que $R_T(\theta) = \text{Var}_\theta(T)$.

Definició 8

Un estimador òptim (que minimitzi la variància) dins ξ_g és un estimador UMV.

Proposició 5

Siguin T_1, T_2 dos estimadors UMV de $g(\theta)$, aleshores $P_\theta(x \in \Omega, T_1(x) = T_2(x)) = 1, \forall \theta \in \Theta$.

Demostració. Sabem que $R_{T_1}(\theta) = R_{T_2}(\theta)$. Agafem $T = \frac{1}{2}(T_1 + T_2)$, que té biaix 0. Com T és UMV, tenim que

$$R_{T_1} \leq R_T(\theta) = \text{Var}_\theta(T) = \frac{1}{4} \text{Var}_\theta(T_1 + T_2) = \quad (1)$$

$$= \frac{1}{4} [\text{Var}_\theta(T_1) + \text{Var}_\theta(T_2) + 2\text{Cov}_\theta(T_1, T_2)] \quad (2)$$

$$\leq R_{T_1} \quad (3)$$

ja que $\text{Cov}_\theta(T_1, T_2) = E_\theta[(T_1 - g(\theta))(T_2 - g(\theta))] \leq \sqrt{\text{Var}_\theta(T_1)\text{Var}_\theta(T_2)} = R_{T_1}(\theta)$.

Així,

$$0 \leq \text{Var}_\theta(T_1 - T_2) = \text{Var}_\theta(T_1) + \text{Var}_\theta(T_2) - 2\text{Cov}_\theta(T_1, T_2) = 0$$

i això implica que $T_1 = T_2 + \mathcal{C}$ quasi segurament. Com que cap dels dos estimadors té biaix, tenim que $T_1 = T_2$ quasi segurament. $\square$

2.4 Estimadors eficients

Considerem un model estadístic paramètric regular $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$ amb Θ un interval, un vector aleatori $X = (X_1, \dots, X_n): \tilde{\Omega} \rightarrow \Omega$ amb llei P_θ i funció de versemblança associada $L(x, \theta)$.

Definim ξ_F com el conjunt d'estimadors reals T de quadrat integrable tal que per a cada $\theta_0 \in \Theta$ podem trobar una funció positiva sumable $h(x)$ tal que $|T(x)\partial_\theta L(x, \theta)| \leq h(x), \forall \theta \in \text{Ent}(\theta_0)$.

Teorema 2 (Cramer-Rao)

$$T \in \xi_F, E_\theta T = g(\theta) \implies \text{Var}_\theta(T) \geq \frac{g'(\theta)^2}{I(\theta)}$$

Observació 7

1. Obtenim una cota inferior de la variància dels estimadors sense biaix de $g(\theta)$.
2. Es pot obtenir en condicions molt més febles de regularitat.
3. Si la igualtat es compleix, l'estimador que la compleix és UMV. Si la igualtat no es compleix, no vol dir que no puguem trobar un estimador UMV.
4. Aquesta cota no sempre és assolible.

Definició 9

Considerem un estimador T sense biaix de $g(\theta)$ que pertany a ξ_F . Diem que T és **eficient** si $\forall \theta \in \Theta$ es compleix

$$\text{Var}_\theta(T) = \frac{g'(\theta)^2}{I(\theta)}$$

Observació 8

1. El model ha de ser regular.
2. No necessàriament existeix.
3. T eficient $\implies T$ UVM (i per això serà únic).
4. El recíproc no és cert ni en models regulars.

Cramer-Rao. La desigualtat de Cauchy-Schwartz implica que

$$|E_\theta [(\partial_\theta \ln L(x, \theta))(T(x) - g(\theta))]| \leq \sqrt{I(\theta) \cdot \text{Var}_\theta(T)}$$

D'altra banda

$$\begin{aligned} E_\theta [(\partial_\theta \ln L(x, \theta))(T(x) - g(\theta))] &= E_\theta [T(x) \partial_\theta \ln L(x, \theta)] - E_\theta [g(\theta) \partial_\theta \ln L(x, \theta)] = \\ &= \int_\Omega T(x) \partial_\theta \ln L(x, \theta) \cdot L(x, \theta) dx = \int_\Omega T(x) \partial_\theta L(x, \theta) dx = \\ &= \partial_\theta \int_\Omega T(x) L(x, \theta) dx = \partial_\theta (E_\theta T) = \partial_\theta g(\theta) \end{aligned}$$

□

Proposició 6

Si $T \in \xi_F$. T és eficient $\iff \exists \lambda(\theta) \forall \theta \in \Theta, \lambda(\theta) \partial_\theta \ln L(x, \theta) = T(x) - g(\theta)$ quasi segurament.

Demostració. La desigualtat de Cauchy-Schwartz és una igualtat quan les variables són dependents. □

Observació 9

Suposant que tinguem la segona part de la proposició o $\lambda(\theta)$ és derivable. Derivant a les dues bandes de la igualtat obtenim

$$\lambda'(\theta) \partial_\theta \ln L(x, \theta) + \lambda(\theta) \partial_\theta^2 \ln L(x, \theta) = -g'(\theta)$$

apliquem esperances

$$E_\theta(\lambda'(\theta) \partial_\theta \ln L(x, \theta)) + \lambda(\theta) E_\theta(\partial_\theta^2 \ln L(x, \theta)) = -g'(\theta)$$

. A més,

$$\text{Var}_\theta(T) = \frac{g'(\theta)^2}{I(\theta)} = \frac{g'(\theta)}{g'(\theta)} \lambda(\theta) g'(\theta)$$

Proposició 7

Considerem un model estadístic uniparamètric exponencial $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$ amb Θ un interval i

$$L(x, \theta) = \exp C(\theta)\Phi(x) + \psi(x) + D(\theta)$$

complint C, D de classe $\mathcal{C}^2$ amb $C' \neq 0, \forall \theta \in \Theta$ i $E_\theta \Phi^2(x) < +\infty, \forall \theta \in \Theta$. Aleshores $T = \Phi(x)$ és un estimador sense biaix i eficient de $g(\theta) = \frac{-D'(\theta)}{C'(\theta)}$

Demostració. Sabem que el model és regular. Aleshores

$$\partial_\theta \ln L(x, \theta) = C'(\theta)\Phi(x) + D'(\theta) = C'(\theta) \left(\phi(x) - \left(-\frac{D'(\theta)}{C'(\theta)} \right) \right)$$

□

Observació 10

Assumint prou condicions de regularitat pot demostrar-se que la cota de Cramer-Rao només s'assoleix en models exponencial i, per tant, els estimadors eficients només existeixen en aquests tipus de models.

2.5 Mètode de la màxima versemblança

En principi pot semblar que no es basa en una idea d'optimalitat però té molt bones propietats asimptòtiques. Ho van treballar al segle XVIII Lambert, Lagrange i Bernoulli. El segle XIX Gauss va dir: "si tenim n observacions d'una variable amb llei desconeguda P_θ , les observacions $(x_1, \dots, x_n)$ haurien de ser les més probables i, per tant farem una estimació del paràmetre θ de manera que els resultats obtinguts tinguin probabilitat màxima.

Tenim $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$, $\Theta \subseteq \mathbb{R}^d$, n observacions $x_1, \dots, x_n$ d'una variable aleatòria $X = (X_1, \dots, X_n): \Omega \rightarrow \Omega$ amb llei P_θ i funció de versemblança $L(x, \theta)$ associada.

Definició 10

Un estadístic $\hat{\theta}(x)$ a valors de Θ s'anomena **estimador de màxima versemblança** del paràmetre θ si es compleix

$$L(x, \hat{\theta}(x)) = \sup_{\theta \in \Theta} L(x, \theta), \forall x \notin N, P_\theta(N) = 0, \forall \theta \in \Theta$$

Observació 11

1. No necessàriament existeixen però si existeix, és únic.
2. Si $\Theta \subseteq \mathbb{R}^d$ és obert, $L(x, \theta) > 0, \forall x \in \Omega, \forall \theta \in \Theta, \mathcal{C}^2$ respecte θ per a tot $x \in \Omega$. L'estimador de màxima versemblança $EMV(\hat{\theta})$ ha de complir les anomenades **equacions de versemblança**:

$$\partial_i \ln L(x, \theta)|_{\theta=\hat{\theta}(x)} = 0, i = 1, \dots, d$$

i la matriu

$$\left(-\partial_{\theta_i \theta_j}^2 \ln L(x, \theta) \Big|_{\theta = \hat{\theta}(x)} \right)_{i,j=1,\dots,d}$$

és definida positiva.

Proposició 8 (Propietat d'invariància funcional)

Si $g: \Omega \longrightarrow \tilde{\Omega}$ mesurable i bijectiva entre dos oberts de $\mathbb{R}^d$. Si $\hat{\theta}(x)$ és un EMV de θ en el model $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$, aleshores $g(\hat{\theta})$ és un EMV de $\tilde{\theta} = g(\theta)$ en el model $(\tilde{\Omega}, \tilde{\mathcal{F}}, \{\tilde{P}_{g^{-1}(\tilde{\theta})}, \tilde{\theta} \in \tilde{\Theta}\})$.

Demostració. La funció de versemblança del segon model és

$$\tilde{L}(x, \tilde{\theta}) = L(x, g^{-1}(\tilde{\theta}))$$

. Per tant

$$\begin{aligned} \tilde{L}(x, g(\hat{\theta}(x))) &= L(x, \hat{\theta}(x)) = \\ &= \sup_{\theta \in \Theta} L(x, \theta) = \sup_{\tilde{\theta} \in \tilde{\Theta}} L(x, g^{-1}(\tilde{\theta})) = \sup_{\tilde{\theta} \in \tilde{\Theta}} \tilde{L}(x, \tilde{\theta}) \end{aligned}$$

□

Proposició 9

1. $\exists$ estadístic suficient T i $\hat{\theta}EMV \implies \hat{\theta}$ és funció de T .
2. Si existeix estimador eficient serà obtingut per aquest mètode.

2.6 Mètode dels moments

És un mètode basat en la llei forta dels grans nombres i una idea de substitució.

Tenim $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$, $\theta = (\theta_1, \dots, \theta_d) \in \Theta \subseteq \mathbb{R}^d$, n observacions $x_1, \dots, x_n$ d'una variable aleatòria $X = (X_1, \dots, X_n): \tilde{\omega} \longrightarrow \Omega$ amb llei P_θ . Assumim que la llei té moments d'ordre $k \in \mathbb{N}$, $E_\theta(|x_1|^k) < +\infty$, $\forall \theta \in \Theta$ i suposem també que

$$\left. \begin{aligned} \theta_1 &= h_1(m_1, \dots, m_k) \\ &\vdots \\ \theta_d &= h_d(m_1, \dots, m_k) \end{aligned} \right\}$$

amb $k \geq d$, $m_j = E_\theta(x_1^j)$, $j = 1, \dots, k$ i $h_1, \dots, h_d: \mathbb{R}^k \longrightarrow \mathbb{R}$ contínues.

La llei forta dels grans nombres implica que

$$\hat{m}_j = \frac{1}{n} \sum_{l=1}^n x_l^j \xrightarrow[n \nearrow +\infty]{P_\theta} m_j = E_\theta(x_1^j), j = 1, \dots, k$$

i tenim que

$$l(\hat{m}_1, \dots, \hat{m}_k) \xrightarrow[P_\theta]{n \nearrow +\infty} h_l(m_1, \dots, m_k) = \theta_l, l = 1, \dots, d$$

Aleshores considerem

$$\hat{\theta}_l(x_1, \dots, x_d) = h_l(\hat{m}_1, \dots, \hat{m}_k), l = 1, \dots, d$$

i l'anomenem **estimador pel mètode dels moments**.

Proposició 10

1. $k \geq d$ però pot passar que $k > d$.
2. No te més justificació teòrica,

2.7 Propietats asimptòtiques dels estimadors

$(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$, $\Theta \subseteq \mathbb{R}^d$, $X_1, \dots, X_n$ són variables aleatòries independents idènticament distribuïdes amb llei desconeguda P_θ i $(x_1, \dots, x_n)$ són observacions de $X = (X_1, \dots, X_n): \Omega \rightarrow \Omega$.

Considerem un estadístic T_n on observarem la dependència de n perquè volem estudiar comportaments al límit, és a dir $T_n, n \geq 1$,

$$T_n: \Omega \rightarrow g(\theta)$$

(funció de n).

2.7.1 Propietats

Consistència $T_n, n \geq 1$ és un estimador **consistent** si $\forall \epsilon > 0, \forall \theta \in \Theta$,

$$\lim_{n \rightarrow \infty} P_\theta(|T_n - g(\theta)| > \epsilon) = 0 \iff T_n \xrightarrow[P_\theta]{n \nearrow \infty} g(\theta)$$

L'estimador és **fortament consistent** si

$$T_n \xrightarrow[P_\theta, q.s.]{n \nearrow \infty} g(\theta)$$

Observació 12

1. Si $\{T_n, n \geq 1\}$ té moments de segon ordre amb $Var_\theta(T_n) \xrightarrow{n \nearrow \infty} 0$ i $b_{T_n}(\theta) \xrightarrow{n \nearrow \infty} 0 \implies T_n$ consistent.
2. Amb les mateixes hipòtesis però sense biaix en lloc del comportament asimptòtic del biaix $\implies T_n$ consistent.

3. Si $\{T_n, n \geq 1\}$ són integrables, direm que són **asimptòticament sense biaix** de $g(\theta)$ si

$$b_{T_n}(\theta) = E_\theta T_n - g(\theta) \xrightarrow{n \nearrow \infty} 0$$

(Si $R_{T_n}(\theta) \xrightarrow{n \nearrow \infty} 0 \implies$ asimptòticament sense biaix).

4. Siguin $\{T_n, n \geq 1\}$ de quadrat integrable sense biaix de $g(\theta)$ amb Θ unidimensional diem que $\{T_n, n \geq 1\}$ **asimptòticament normal** si

$$\sqrt{n}(T_n - g(\theta)) \xrightarrow[n \nearrow \infty]{\mathcal{L}} N(0, \sigma^2(\theta)), \sigma^2(\theta) > 0$$

Si el model és regular, una successió asimptòticament normal és **asimptòticament eficient** si $\sigma^2(\theta) = \frac{g'(\theta)}{I(\theta)}$

2.8 Propietats dels EMV i estimador del mètode dels moments

2.8.1 EMV

1. En general tenen biaix però asimptòticament no.
2. Són consistents.
3. Tot EMV no és eficient; no obstant, si existeix l'estimador eficient serà l'EMV.
4. Solen ser asimptòticament normals i eficients.

2.8.2 Moments

1. Consistents i fortament consistents.
2. Certa normalitat asimptòtica.

3 Intervals de confiança

L'estimació puntual ens proporciona un estimador T que interpretem com una aproximació al valor de la funció $g(\theta)$. En aquest capítol desenvoluparem un mètode que consisteix en trobat un subconjunt de $g(\theta)$ depenent de les observacions x i que conté el valor del paràmetre amb una probabilitat fixada a priori.

Considerem un model estadístic $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$, amb $\Theta \subseteq \mathbb{R}^d$, n observacions $x_1, \dots, x_n$ d'una variable aleatòria $X = (X_1, \dots, X_n): \tilde{\omega} \rightarrow \Omega$ amb llei P_θ i una funció $g: \Theta \rightarrow \mathbb{R}^k$. Fixem un valor $\gamma \in (0, 1)$.

Definició 11

Una regió de confiança de $g(\theta)$ amb coeficient de confiança γ és un a família de

parts de $\mathbb{R}^k$, $\{S(x) \subseteq \mathbb{R}^k, x \in \Omega\}$ ($S(x)$ serà un interval de $\mathbb{R}^k$ on voldrem que el paràmetre estigui contingut) que satisfà:

1. $\forall \theta \in \Theta, \{x \in \Omega, g(\theta) \in S(x)\} \in \mathcal{F}$
2. $P_\theta(x \in \Omega, g(\theta) \in S(x)) > \gamma, \forall \theta \in$

Si $d = 1$, s'anomenen intervals de confiança.

La idea és que repetint moltes vegades l'experiència que dona lloc a les observacions, la regió $S(x)$ contindrà el valor real del paràmetre en una proporció γ dels casos.

3.1 Mètode del pivot

A continuació veiem un mètode general per construir regions de confiança.

Definició 12

Una funció $\pi: \Omega \times g(\Theta) \rightarrow \mathbb{R}^k$ s'anomena funció pivotant si, per a tot $\theta \in \Theta$, $\pi(\cdot, g(\theta)): \Omega \rightarrow \mathbb{R}^k$ és un vector aleatori i la seva llei no depèn de θ .

Vegem com amb aquesta definició construïm regions: Sigui v la llei de probabilitat en $\mathbb{R}^k$ del vector aleatori $\pi(\cdot, g(\theta))$, $\theta \in \Theta$ i sigui B un borelià de $\mathbb{R}^k$ que compleix $v(B) \geq \gamma$. Definim

$$S(x) = \{y \in (\Theta), \pi(x, y) \in B\} = \pi(x, \cdot)^{-1}(B)$$

Aleshores $\{S(x) \in \Omega\}$ és una regió de confiança per a $g(\theta)$ al nivell de confiança γ . Efectivament, per a tot $\theta \in \Theta$, $\{x \in \Omega, g(\theta) \in S(x)\} = \{x \in \Omega, \pi(x, g(\theta)) \in B\} \in \mathcal{F}$, i $P_\theta(x \in \Omega, g(\theta) \in S(x)) = P_\theta(\pi(x, g(\theta)) \in B) = v(B) \geq \gamma$.

Hem de tenir en compte que aquest mètode ens permet construir regions de confiança però es construeix d'una manera implícita i no explícitament (no tenim cap manera de trobar funcions pivotants, per exemple).

3.2 Teorema de Fisher

En el cas de normalitat, per trobar funcions pivotants serà molt útil fer servir el següent teorema:

Teorema 3 (Teorema de Fisher)

Siguin $X_1, \dots, X_n$ variables aleatòries independents idènticament distribuïdes amb llei $N(0, 1)$. Aleshores $\bar{X}_n$ i $Q_n = nS_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2$ són independents i $\bar{X}_n \sim N(0, \frac{1}{n})$ i $Q_n \sim \chi^2(n-1)$.

Demostració. Que $\bar{X}_n \sim N(0, \frac{1}{n})$ és obvi per les propietats de normalitat. D'altra banda, com que $X_1, \dots, X_n$ són variables aleatòries independents i idènticament distribuïdes amb llei $N(0, 1)$, denim que

$$f_{X_1, \dots, X_n}(x) = \prod_{i=1}^n f_{X_i}(x_i) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\sum_{i=1}^n \frac{x_i^2}{2}} = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n x_i^2}$$

amb $x = (x_1, \dots, x_n) \in \mathbb{R}^n$. Ara considerem una matriu ortogonal T de dimensió nn amb una primera columna $\left(\frac{1}{\sqrt{n}}, \dots, \frac{1}{\sqrt{n}}\right)$ i definim $Y = (y_1, \dots, y_n) = X \times T$. Aleshores, com a conseqüència obtenim

$$Y_1 = x \cdot \begin{pmatrix} \frac{1}{\sqrt{n}} \\ \vdots \\ \frac{1}{\sqrt{n}} \end{pmatrix} = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i = \sqrt{n} \bar{X}_n \quad (4)$$

$$\|Y\|^2 = YY^T = xT \cdot T^t x^T = x \cdot Id \cdot x^T = \|x\|^2 \quad (5)$$

$Y = g(x) = x \cdot T$ és absolutament continu amb densitat

$$\begin{aligned} f_Y(g) &= |J_{g^{-1}(y)}| \cdot f_X(g^{-1}(y)) = 1 \cdot f_X(yT^{-1}) = \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{\|yT^{-1}\|^2}{2}\right) = \\ &= \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{yT^{-1}Ty^T}{2}\right) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} \exp\left(-\frac{\|y\|^2}{2}\right) \end{aligned}$$

Veiem que

$$f_Y(y) = \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{\|y\|^2}{2}} = \prod_{i=1}^n \left(\frac{1}{2\pi}\right)^{\frac{n}{2}} e^{-\frac{y_i^2}{2}}$$

amb $y = (y_1, \dots, y_n)$ és el producte d' n densitats $N(0, 1)$, de manera que $Y = (Y_1, \dots, Y_n)$ són variables aleatòries independents idènticament distribuïdes amb llei $N(0, 1)$ Aleshores,

$$\begin{aligned} Q_n &= nS_n^2 = \sum_{i=1}^n (X_i - \bar{X}_n)^2 = \sum_{i=1}^n X_i^2 - n\bar{X}_n^2 = \\ &= \|x\|^2 - n\bar{X}_n^2 = \|Y\|^2 - Y_1^2 = \sum_{i=2}^n Y_i^2 \end{aligned}$$

Per tant, $\bar{X}_n$ i nS_n^2 són independents i $\bar{X}_n \sim N(0, \frac{1}{n})$ i $nS_n^2 \sim \psi^2(n-1)$ (suma de normals $N(0, 1)$ al quadrat independents) $\square$

Corol·lari 2

Siguin $X_1, \dots, X_n$ variables aleatòries independents idènticament distribuïdes amb distribució $N(\mu, \sigma^2)$. Aleshores

- $\bar{X}_n \sim N(\mu, \frac{\sigma^2}{n})$ i $\frac{nS_n^2}{\sigma^2} \sim \psi^2(n-1)$ independent
- $\sqrt{n-1} \frac{\bar{X}_n - \mu}{S_n} \sim t_{(n-1)}$

Corol·lari 3

Siguin $X_1, \dots, X_n$ variables aleatòries independents idènticament distribuïdes amb distribució $N(\mu_1, \sigma_1^2)$ i $Y = (Y_1, \dots, Y_n)$ variables aleatòries independents idènticament distribuïdes amb distribució $N(\mu_2, \sigma_2^2)$ independents. Aleshores

$$\frac{n_1 S_{n_1}^2}{(n_1 - 1) \sigma_1^2} : \frac{n_2 S_{n_2}^2}{(n_2 - 1) \sigma_2^2} = \frac{n_1 S_{n_1}^2 (n_2 - 1) \sigma_2^2}{n_2 S_{n_2}^2 (n_1 - 1) \sigma_1^2} \sim F_{n_1 - 1, n_2 - 1}$$

3.3 Altres mètodes

3.3.1 Mètode general per construir funcions pivotants

$X = (X_1, \dots, X_n)$ variables aleatòries independents idènticament distribuïdes amb llei P_θ i funció de distribució F_θ . Suposem que F_θ és contínua i estrictament creixent (per assegurar la bijectivitat i facilitar els càlculs). No és difícil comprovar que $F_\theta(X_i) \sim U(0, 1)$:

Demostració. Sigui $Y = F_\theta(X_i)$, $y \in (0, 1)$.

$$P_\theta(Y \leq y) = P_\theta(X_i \leq F_\theta^{-1}(y)) = F_\theta(F_\theta^{-1}(y)) = y$$

així

$$F_Y(y) = \begin{cases} 0 & \text{si } y < 0 \\ y & \text{si } 0 \leq y \leq 1 \\ 1 & \text{si } y > 1 \end{cases}$$

que és la funció de distribució d'una $U(0, 1)$

□

Ara fent un canvi de variable obtenim que

$$-\ln F_\theta(X_i) \sim \text{Exp}(1) = G(1, 1)$$

En definitiva

$$\Pi(X, \theta) = -\sum_{i=1}^n \ln F_\theta(X_i) \sim G(n, 1)$$

que és una funció pivotant.

Torna a aparèixer, però, el mateix problema. Resolem el problema teòricament i de manera implícita, però encara no tenim una manera explícita de trobar l'interval de confiança...

3.3.2 Interval de confiança mitjançant Txebeixev

Aquest és un mètode molt general i poc precís, de manera que dóna intervals molt grans.

Considerem un model estadístic paramètric $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$, $\Theta \subseteq \mathbb{R}$ i suposem que tenim un estimador T sense biaix de θ , de quadrat integrable i amb

funció de risc $R_T(\theta)$. Per a $\delta > 0$, la desigualtat de Txebychev aplicada a la funció x^2 ens implica que

$$P_\theta \left(\left| \frac{T - \theta}{\sqrt{R_T(\theta)}} \right| > \delta \right) \leq \frac{1}{\delta^2} E_\theta \left(\left| \frac{T - \theta}{\sqrt{R_T(\theta)}} \right|^2 \right) = \frac{1}{\delta^2}$$

Aleshores, si volem un interval de confiança al nivell γ tenim que per a $\delta_\gamma = \frac{1}{\sqrt{1-\gamma}}$

$$P_\theta \left(T - \delta_\gamma \sqrt{R_T(\theta)} \leq \theta \leq T + \delta_\gamma \sqrt{R_T(\theta)} \right) \geq \gamma$$

Aquí ens sortiran inequacions amb θ d'on l'hauríem d'aïllar. Com que això pot ser complicat, quan no sigui possible podem aproximar $R_T(\theta)$ per $R_T(T)$, obtenint l'interval

$$\left[T - \delta_\gamma \sqrt{R_T(T)}, T + \delta_\gamma \sqrt{R_T(T)} \right]$$

4 Contrast d'hipòtesis

4.1 Definicions i notacions bàsiques

Considerem un model estadístic paramètric $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$. Qüestionar alguna cosa del paràmetre és equivalent a fer una partició de l'espai de paràmetres

$$\Theta = \Theta_0 \cup \Theta_1, \quad \Theta_0 \cap \Theta_1 = \emptyset$$

Habitualment la qüestió que plantegem l'anomenarem **hipòtesi nul·la** (és curiós que es digui hipòtesi nul·la precisament a la hipòtesi que volem comprovar que sigui certa, però es diu que això prové del resultat primer contrast d'hipòtesi documentat: la nul·la possibilitat de diferenciar si s'ha posat primer el té o la llet) $H_0: \theta \in \Theta_0$ i la complementària serà l'**hipòtesi alternativa** $H_1: \theta \in \Theta_1 = \Theta \setminus \Theta_0$. Direm que Θ_0 és simple si té un sol element i composta si en té més d'un.

Aquestes dues hipòtesis defineixen un problema de test. Per resoldre'l, ens haurem de basar en les observacions, que faran acceptar o rebutjar la hipòtesi nul·la. Així, un test d'hipòtesi φ ve donat per una partició d' Ω (per tant una partició d'observacions

$$\Omega = A_0 \cup A_1, \quad A_0 \cap A_1 = \emptyset, \quad A_0 \in \mathcal{F}$$

(fixem-nos que no ens cal demanar també que $A_1 \in \mathcal{F}$, ja que en ser una σ -àlgebra el seu complementari també hi és) La regla de decisió és:

- Si les observacions $x_1, \dots, x$ pertanyen a A_0 , acceptarem la hipòtesi H_0
- Si les observacions $x_1, \dots, x$ pertanyen a A_1 , rebutjarem H_0 (alerta, no acceptem H_1 , només rebutgem H_0 !)

D'acord amb aquesta regla, anomenarem A_0 **regió d'acceptació** i a A_1 **regió de rebuig** o **regió crítica**. Per trobar-les treballarem a amb les probabilitats de prendre decisions incorrectes. Podem, doncs, fer dos errors:

Error de 1a espècie : Rebutjar la hipòtesi nul·la essent certa

$$\alpha_1 = P_\theta(x \in A_1), \forall \theta \in \Theta_0$$

Error de 2a espècie : Acceptar la hipòtesis nul·la essent falsa

$$\alpha_2(\theta) = P_\theta(x \in A_0), \forall \theta \in \Theta_1$$

Podem expressar els dos errors en termes d'una funció anomenada **funció de potència** definida com

$$\begin{aligned} \phi: \Theta &\longrightarrow [0, 1] \\ \theta &\longmapsto \phi(\theta) = P_\theta(A_1) \end{aligned}$$

Aleshores

$$\begin{aligned} \alpha_1(\theta) &= \phi(\theta), \forall \theta \in \Theta \\ \alpha_2(\theta) &= 1 - \phi(\theta), \forall \theta \in \Theta \end{aligned}$$

Definició 13 (Potència d'un test)

Definim la **potència d'un test** com la restricció de ϕ a Θ_1 , que indica la probabilitat de rebutjar H_0 quan és falsa.

Ara volem trobar A_0 i A_1 que minimitzi $\alpha_1(\theta)$ i $\alpha_2(\theta)$ (o minimitzar α_1 i maximitzar la potència), però això és minimitzar dues funcions, de manera que en general no serà possible.

El que farem, doncs, serà acotar l'error que tinguin pitjors conseqüències (ho farem per α_1) per una quantitat fixada a priori, i d'entre els tests que ho compleixin, triarem el que tinguin α_2 més petit (o, alternativament, la potència més gran). Fem-ho de manera formal:

Definició 14

Diem que un test A_0, A_1 té **nivell de significació** $\alpha \in (0, 1)$ (sovint molt petita) si $\sup_{\theta \in \Theta_0} \alpha_1(\theta) \leq \alpha$.

Definició 15

Anomenarem $\mathcal{D}_\alpha$ al conjunt dels tests (particions d' Ω en A_0 i A_1) amb nivell de significació α . Si $\varphi_1, \varphi_2 \in \mathcal{D}_\alpha$ direm que φ_1 és **més potent** que φ_2 si

$$\phi_{\varphi_2}(\theta) \leq \phi_{\varphi_1}(\theta), \forall \theta \in \Theta_1$$

Definició 16

Fixat $\alpha \in (0, 1)$ un nivell de dignificació, un test φ de $\mathcal{D}_\alpha$ és un test **UMP** (**uniformement de màxima potència**) si és més potent que qualsevol altre test de $\mathcal{D}$, $\phi_{\varphi'}(\theta) \leq \phi_\varphi(\theta), \forall \varphi' \in \mathcal{D}_\alpha, \forall \theta \in \Theta_1$

Aleshores, de tots els tests de $\mathcal{D}_\alpha$ agafarem el que tingui potència més gran.

Observació 13 1. *Pot no existir o no ser únic*

2. *Les dues hipòtesis no són intercanviables, no són simètriques*

3. *Els tests solen ser conservadors a favor de la hipòtesi nul·la. És poc freqüent rebutjar una hipòtesi nul·la correcta. Quan acceptem, és perquè les observacions no ens han donat prou evidències com per rebutjar-la.*

4.2 Contrast entre hipòtesis simples

Teorema 4 (Teorema fonamental de Neyman-Pearson)

Fixem $\alpha \in (0, 1)$ i assumim que $\exists c_\alpha > 0$ tal que $A_0 = \{x \in \Omega, L(x, \theta_1) \leq c_\alpha L(x, \theta_0)\}$ satisfà que $P_{\theta_0}(A_0^C) = \alpha$. Aleshores A_0 és un test UMP (una regió d'acceptació) amb nivell de significació α per a contrastar $H_0: \theta = \theta_0$ contra $H_1: \theta = \theta_1$.

4.3 Extensió a contrastos d'hipòtesis unilaterals

Suposem que tenim un model estadístic paramètric $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$ i que mitjançant el lema de Neyman-Pearson hem obtingut un test UMP de nivell de significació α amb regió A_0 per a contrastar $H_0: \theta = \theta_0$ contra $H_1: \theta = \theta_1$ assumint $\theta_0 < \theta_1$. Aleshores, si la regió A_0 no depèn de θ_1 i la funció de potència $\phi(\theta) = P_\theta(A_0^C)$ és creixent (és a dir, que la funció de versemblança creix respecte el paràmetre), la regió A_0 també serà la regió d'acceptació d'un test UMP de nivell de significació α per a contrastar $H_0: \theta \leq \theta_0$ contra $H_1: \theta > \theta_0$.

4.4 Test de la raó de versemblança

Tenim $(\Omega, \mathcal{F}, \{P_\theta, \theta \in \Theta\})$, $\Theta \subseteq \mathbb{R}^d$, $x_1, \dots, x_n$ observacions de $X = (X_1, \dots, X_n): \tilde{\Omega} \rightarrow \Omega$ i $L(x, \theta)$ la funció de versemblança de P_θ . Donada una partició $\Theta = \Theta_0 \cup \Theta_1$, amb $\Theta_0 \cap \Theta_1 = \emptyset$, anomenarem raó de versemblança al quocient

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_0} L(x, \theta)}{\sup_{\theta \in \Theta} L(x, \theta)}$$

Fixem-nos que si $\hat{\theta}(x)$ és l'estimador de màxima versemblança, $\sup_{\theta \in \Theta} L(x, \theta) = L(x, \hat{\theta}(x))$.

Si fixem $\alpha \in (0, 1)$, el test de la raó de versemblança a nivell de significació α es defineix com $A_0 = \{x \in \Omega, \lambda(x) \geq c_\alpha\}$, on $c_\alpha \leq 1$ es determina de manera que $P_\theta(A_0^C) \leq \alpha$, $\forall \theta \in \Theta_0$ i contrasta $H_0: \theta \in \Theta_0$ contra $H_1: \theta \in \Theta_1$.

(La idea intuïtiva aquí és que per acceptar la hipòtesi nul·la θ és certa, θ ha de ser molt propera a l'estimador de màxima versemblança. Aleshores el quocient $\lambda(x)$ serà molt proper a 1.)

Problemes:

1. El test no ens diu res de la potència.
2. Pot ser molt complicat trobar la llei de $\lambda(x)$. Si no la podem trobar, podríem trobar el comportament asimptòtic.

Proposició 11

Si els tests de N - P per contrastar hipòtesis simples i la raó de versemblança existeixen, aleshores els dos tests coincideixen

Demostració. $A_0^{RV} = \{x \in \Omega, \lambda(x) \geq c_\alpha\} = \{x \in \Omega, \frac{L(x, \theta_0)}{L(x, \theta_0) \vee L(x, \theta_1)} \geq c_\alpha\}$ ¹
Aleshores, operant obtenim

$$A_0^{RV} = \{x, \frac{L(x, \theta_0)}{L(x, \theta_1)} \geq c_\alpha, L(x, \theta_1) \geq L(x, \theta_0)\} \sqcup \{x, \frac{L(x, \theta_0)}{L(x, \theta_0)} \geq c_\alpha, L(x, \theta_1) < L(x, \theta_0)\}$$

Ara bé, a aquest últim conjunt $\frac{L(x, \theta_0)}{L(x, \theta_0)} = 1$ i $1 \geq c_\alpha$ sempre, de manera que simplificant els conjunts $(L(x, \theta_1) \leq L(x, \theta_0) \implies \frac{L(x, \theta_1)}{L(x, \theta_0)} \leq 1 \leq \frac{1}{c_\alpha}$, ja que $c_\alpha \geq 1$) $A_0^{RV} = \{x, \frac{L(x, \theta_1)}{L(x, \theta_0)} \geq \frac{1}{c_\alpha}\}$ que és una regió d'acceptació amb la forma del lema de Neyman-Pearson. (Això últim no ho tinc molt clar, cal revisar-ho!) $\square$

Per tant, en el cas d'hipòtesis simples sí que podem conèixer la potència mitjançant Neyman-Pearson. Ara bé, continuem tenint el problema de que no necessàriament coneixem la llei de $\lambda(x)$. El següent resultat adreça aquest problema:

Teorema 5

Suposem $\Omega \subseteq \mathbb{R}$ i que Θ és un interval obert de $\mathbb{R}$. Fixat un valor $\theta_0 \in \Theta$, assumim que el model estadístic associat a una observació satisfà les condicions clàssiques de regularitat que hem vist anteriorment. Suposem, també, que per a cada $n \geq 1$ existeix una única solució de l'equació de versemblança $(\partial_\theta \ln L(x, \theta) = 0)$. Aleshores, quan contrastem $H_0: \theta = \theta_0$ contra $H_1: \theta \neq \theta_0$ tenim que sota la hipòtesi nul·la

$$-2\ln\lambda(x) \xrightarrow[n \nearrow +\infty]{\mathcal{L}} \chi_{(1)}^2$$

Això permetria fer, per exemple

$$\alpha = P_{\theta_0}(\lambda(x) \leq c_\alpha) = P_{\theta_0}(-2\ln\lambda(x) \geq -2\ln c_\alpha) \approx P(\chi_{(1)}^2 \geq -2\ln c_\alpha)$$

Observació 14

1. Si tenim que Θ és un obert de $\mathbb{R}^k$, assumint les mateixes hipòtesis, quan contrastem $H_0: \theta = \theta_0$ contra $H_1: \theta \neq \theta_0$ obtenim que

$$-2\ln\lambda(x) \xrightarrow[n \nearrow +\infty]{\mathcal{L}} \chi_{(k)}^2$$

¹on $A \vee B$ vol dir el suprem d' A i B

2. També en condicions suficients de regularitat, si contrastem $H_0: \theta \in \Theta_0$ contra $H_1: \theta \notin \Theta_0$ podem utilitzar que

$$-2\ln\lambda(x) \xrightarrow[n \nearrow +\infty]{\mathcal{L}} \chi_{(r)}^2$$

amb r la diferència entre el nombre de paràmetres lliures de l'espai de paràmetres i el nombre de paràmetres lliures de Θ_0 .

5 Tests d'ajustament i independència

En aquest apartat estudiarem, si una mostra prové d'una llei concreta (tests d'ajustament), si dues o més mostres provenen d'una mateixa llei (tests d'homogeneïtat) o si hi ha relacions d'independència o associació entre agrupacions d'observacions (tests d'independència).

5.1 Teorema de Glivenko-Cantelli

També anomenat **teorema finamental de l'estadística**, és una extensió de la llei forta dels grans nombres. Intuitivament ens diu que com més gran sigui la mostra, més informació podrem obtenir de la distribució.

Definició 17

Siguin $X_1, \dots, X_n$ variables aleatòries definides en $(\Omega, \mathcal{F}, P)$ que prenen valors a $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Donat $w \in \Omega$, definim la **distribució empírica** associada a $X_1, \dots, X_n$ com la llei de probabilitat discreta en $\mathbb{R}$ que assigna una probabilitat $\frac{1}{n}$ a cadascun dels punts $X_1(w), \dots, X_n(w)$, és a dir que $\forall \mathcal{B} \in \mathcal{B}(\mathbb{R})$,

$$P_n^w(\mathcal{B}) = \frac{1}{n} \times \#\{i \in \{1, \dots, n\}, X_i(w) \in \mathcal{B}\} = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\mathcal{B}}(X_i(w))$$

(A cada element X_i li estem donant una massa $\frac{1}{n}$) En el cas particular d'una successió de variables aleatòries independents idènticament distribuïdes $\{X_k, k \geq 1\}$ (infinita) definides en $(\Omega, \mathcal{F}, P)$ podem definir com P_n^w la **n -èssima probabilitat empírica** associada a $X_1, \dots, X_n$.

Definició 18

La **n -èssima funció de distribució empírica** associada a la successió és la funció de distribució de la n -èssima probabilitat empírica:

$$F_n^w(t) = P_n^w((-\infty, t]) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, t]}(X_i(w)) = \begin{cases} 0 & t < X_{(1)} \\ \frac{i}{n} & X_{(1)} \leq t < X_{(i+1)}, 1 \leq i \leq n \\ 1 & t \geq X_{(n)} \end{cases}$$

$\forall t \in \mathbb{R}$. $X_{(i)}$ fa referència a l'estadístic d'ordre (el que ocupa la posició i -èssima quan els ordenem de manera creixent).

Observació 15

P_n^w i F_n^w són variables aleatòries (en podem calcular la seva esperança). Tal com en les variables aleatòries, per simplificar la notació no indicarem la dependència de w .

Vegem ara perquè diem que aquest teorema és una generalització de la llei forta dels grans nombres;

Observació 16

Fixat un borelià $\mathcal{B} \in \mathcal{B}(\mathbb{R})$ tenim que

$$nP_n(\mathcal{B}) = \#\{i \in \{1, \dots, n\}, X_i(w) \in \mathcal{B}\} = \sum_{i=1}^n \mathbb{1}_{\mathcal{B}}(X_i(w)) \sim \text{Binom}(n, \tilde{P}(\mathcal{B}))$$

amb $\tilde{P}(\mathcal{B}) = P(X_i \in \mathcal{B})$.

Per la llei dels grans nombres tenim que

$$P_n(\mathcal{B}) \xrightarrow[n \nearrow \infty]{q.s.} E[\mathbb{1}_{\mathcal{B}}(X)] = \tilde{P}(\mathcal{B})$$

Ara la pregunta que plantegem és si podem determinar algun tipus de classe ξ de conjunts tal que

$$\sup_{\mathcal{B} \in \xi} |P_n(\mathcal{B}) - P(\mathcal{B})| \xrightarrow[n \nearrow \infty]{q.s.} 0$$

Un problema que ens trobem és que aquesta convergència quasi segura depèn del conjunt $\mathcal{B}$, és a dir, que el conjunt $A_{\mathcal{B}}$ de mesura 0 on falla la convergència pot dependre de $\mathcal{B}$. El teorema de Gilvenko-cantelli ens dirà que això es compleix per a $\xi = \{(-\infty, t], t \in \mathbb{R}\}$ (les semirectes "negatives").

Teorema 6 (Gilvenko-Cantelli)

Sigui $\{X_k, k \geq 1\}$ una successió de variables aleatòries independents idènticament distribuïdes en $(\Omega, \mathcal{F}, P)$ amb funció de distribució F . Aleshores

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \xrightarrow[n \nearrow \infty]{q.s.} 0$$

amb F_n la funció de distribució empírica.

Demostració. (Pendent d'escriure)

□

5.2 Tests d'ajustament

En aquest apartat donarem eines per comprovar la possibilitat de que una certa mostra $X_1, \dots, X$ provingui d'una llei concreta.

5.2.1 Test de la χ^2

Multinomial Tindrem n realitzacions independents d'un experiment amb m resultats excloents $\Omega = A_1 \sqcup \dots \sqcup A_m$. Aquestes categories poden ser de tipus quantitatiu o qualitatiu. La mostra $(x_1, \dots, x_m)$ pot resumir-se en un vector de freqüències absolutes $(n_1, \dots, n_m)$ amb $n_1 + \dots + n_m = n$, amb n_i representa el nombre de vegades que ha ocorregut A_i (si estem estudiant n tirades d'un dau, no ens interessa el valor de les tirades, sinó quantes vegades ha sortit cada nombre).

Volem contrastar si la probabilitat de cada categoria A_i és $P_i^0 \in [0, 1]$ amb $\sum_{i=1}^m P_i^0 = 1$, és a dir, per a $P_i = P(A_i)$, contrastar $H_0: P_i = P_i^0 \forall i = 1, \dots, m$ contra $H_1: P_i \neq P_i^0$ per algun i . Inicialment aquest no era un problema paramètric, però ara l'hem "transformat" en un problema paramètric (o millor dit, parametritzat).

Considerem l'estadístic de chi-quadrat de pearson:

$$D_n(x) = \sum_{i=1}^m \frac{(n_i - nP_i^0)^2}{nP_i^0} = \sum_{j=1}^m \frac{n}{P_j^0} \left(\frac{n_j}{n} - P_j^0 \right)^2$$

Tenint en compte les consideracions següents, plantegem com a regió d'acceptació $A_0 = \{x \in \Omega, D_n(x) \leq c_\alpha\}$

Observació 17

1. Aquest resultat és una generalització del teorema central del límit i se'n pot trobar una prova al llibre de l'Olga Julià i David Màrquez.
2. Pel teorema de Bernoulli tenim que

$$\frac{n_i}{n} \longrightarrow P_i$$

Per tant, sota H_1 , $\exists i \in \{1, \dots, m\}$ tal que $P_i \neq P_i^0$ i aleshores $D_n \longrightarrow +\infty$.

3. Es pot demostrar que

$$D_n \xrightarrow[n \nearrow +\infty]{\mathcal{L}} \chi_{(m-1)}^2 \quad (6)$$

i així, sota H_0 ,

$$P_{H_0}(A_0^C) = P(\chi_{m-1}^2 > c_\alpha) = \alpha$$

4. L'aproximació (6) es considera vàlida si $nP_i^0 \geq 5$ i $n_i \geq 5$, $\forall i = 1, \dots, m$. Si els valors són inferiors no se sap com afecta a l'aproximació.
5. Si utilitzem el test de la raó de versemblança sortirà una regió d'acceptació similar a la del test multinomial.

Exemple 1

Tenim una taula amb els 500 primers decimals del nombre π . Una pregunta que un es pot fer és si aquests 500 primers decimals són aleatoris o segueixen alguna tendència. Agafem A_i l'esdeveniment "ha sortit el nombre i " n_i indica quantes vegades s'esdevé A_i , és a dir, d'entre aquests 500 decimals quantes vegades apareix i . Sigui $P_i = P(A_i)$. Plantegem la següent hipòtesi $H_0: P_i = \frac{1}{10}, \forall i = 0, \dots, 9$ i $H_1: P_i \neq \frac{1}{10}$ per a algun i . Hem de trobar la regió d'acceptació $A_0 = \{D_n \leq c_\alpha\}$. En aquest cas $D_n \rightarrow \chi^2_{(9)}$ (9 graus de llibertat perquè hi ha 10 categories, una per cada nombre). Aleshores imposant $P_{H_0}(A_0^C) = P_{H_0}(D_n > c_\alpha) = P(\chi^2_{(9)} > c_\alpha) = 0.05$ obtenim $c_\alpha = 16.92$. Si tenim una taula de freqüències absolutes n_i , podem calcular $D_n(x)$ empíricament:

$$D_n(x) = \sum_{i=1}^m \frac{(n_i - 500 * 1/10)^2}{500 * 1/10} = 6.56$$

Aquest nombre és menor que c_α , de manera que podem afirmar que els primers 500 decimals de π són aleatoris.

5.2.2 Test de Kolmogorov

Considerem $(\Omega, \mathcal{F}, \mathcal{P})$ un model estadístic on $\mathcal{P}$ és la família de totes les probabilitats amb lleis contínues. Sigui x una mostra aleatòria simple d'una llei $P \in \mathcal{P}$ amb funció de distribució F . Recordem la funció de distribució empírica:

$$\hat{F}_n(y) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{(-\infty, y]}(x_i), \quad y \in \mathbb{R}$$

Ara, si F_0 n'és la funció de distribució que volem contrastar (és a dir, que pensem que les observacions segueixen la llei F_0), plantejarem

$$H_0: F = F_0 \text{ contra } H_1: F \neq F_0$$

. Per fer-ho, utilitzarem l'estadístic de Kolmogorov:

$$\Delta_n = \sup_{y \in \mathbb{R}} |F_0(y) - \hat{F}_n(y)|$$

Sota la hipòtesi nul·la H_0 , el teorema de Glivenko-Cantelli ens diu que

$$\Delta_n \xrightarrow[n \nearrow \infty]{q.s.} 0$$

i, per tant, sembla lògic pensar com a regió d'acceptació $A_0 = \{\Delta_n \leq c_{\alpha,n}\}$, amb $c_{\alpha,n} > 0$ tal que $P_0(A_0^C) \leq \alpha$. El problema ara és saber la llei de Δ_n sota H_0 , que en general és complicat.

Teorema 7

La distribució d' Δ_n sota la hipòtesi nul·la és la mateixa per a tota funció contínua F_0 , és a dir, que no depèn de P_0 .

Demostració. Sigui F_0 la funció de distribució de $X_1, \dots, X$ i $(X_{(1)}, X_{(2)}, \dots, X_{(n)})$ l'estadístic d'ordre. Com que F_0 és contínua i creixent i $\hat{F}_n$ creixent i esglaonada, les diferències més grans s'obtenen en els salts o just abans:

$$\begin{aligned}\Delta_n(x) &= \sup_{1 \leq n} \left\{ \left| F_0(X_{(1)}) - \hat{F}_n(X_{(i)}) \right|, \left| F_0(X_{(i)}^-) - \hat{F}_n(X_{(i)}^-) \right| \right\} \\ &= \sup_{1 \leq n} \left\{ \left| F_0(X_{(1)}) - \frac{i}{n} \right|, \left| F_0(X_{(i)}^-) - \frac{i-1}{n} \right| \right\}\end{aligned}$$

Per tant, la llei d' Δ_n només depèn de $(F_0(X_{(1)}), \dots, F_0(X_{(n)}))$. Si F_0 és estrictament creixent tenim que $F_0(X_i) \sim U(0, 1)$. El cas més general de F_0 només contínua és bastant més complex i es pot trobar al llibre de Julià-MC-Rovira-Sarrà. Com que $X_1, \dots, X$ són variables aleatòries independents idènticament distribuïdes, $F_0(X_1), \dots, F_0(X_n)$ també ho són. Així, $(F_0(X_1), \dots, F_0(X_n)) \sim U(0, 1)^n$ (n uniformes independents), que és una llei independent d' F_0 . Ara només cal observar que $(F_0(X_{(1)}), \dots, F_0(X_{(n)}))$ és un estadístic d'ordre d'una $U(0, 1)^n$ (i, per tant, independent d' F_0). $\square$

Fins ara sabem que Δ_n no depèn de la llei d' F_0 , i a amb això en tenim prou, però no sabem ni com és la llei ni el valor de la constant $c_{\alpha, n}$.

Observació 18

Es podria provar que

$$(F_0(X_1), \dots, F_0(X_n)) \underset{\mathcal{L}}{\sim} f(x) = \frac{1}{n!} (x)$$

amb $D = \{x = (x_1, \dots, x_n) \in \mathbb{R}, x_1 < x_2 < \dots < x_n\}$.

Observació 19

La unció de distribució F_{Δ_n} de Δ_n està tabulada per a valors petits d' n . Per a valors grans d' n ,

$$P(\Delta_n > c_{\alpha, n}) = \alpha \iff c_{\alpha, n} = \sqrt{\frac{-\ln(\alpha/2)}{2n}}$$

5.3 Tests d'homogeneïtat

Suposem que tenim dues mostres independents $(x_1, \dots, X_{n_1})$ i $(y_1, \dots, y_{n_2})$ que provenen de lleis P_1 i P_2 respectivament. Amb aquests tests comprovarem si les lleis d'aquestes observacions són la mateixa llei (i per tant si podem agrupar les observacions en un sol grup), sense necessàriament trobar quines són aquestes lleis.

5.3.1 Test de la χ^2 d'homogeneïtat

Suposem que tenim les dades agrupades o que les variables prenen un nombre finit de valors. Sigui $A_1, \dots, A_m$ una partició d' Ω i siguin $(P_1^1, P_2^1, \dots, P_m^1)$ i

Classe	A_1	A_2	$\cdots$	A_m	Total
Mostra 1	$n_{1,1}$	$n_{1,2}$	$\cdots$	$n_{1,m}$	n_1
Mostra 2	$n_{2,1}$	$n_{2,2}$	$\cdots$	$n_{2,m}$	n_2
Total	$n_{\bullet,1}$	$n_{\bullet,2}$	$\cdots$	$n_{\bullet,m}$	$n_1 + n_2 = n$

$(P_1^2, P_2^2, \dots, P_m^2)$ les probabilitats d' $A_1, \dots, A_m$ respecte P_1 i P_2 . Definim a la taula següent les notacions que farem servir per a les cardinalitats:

La hipòtesi nul·la serà que les dades provenen de la mateixa llei. Sota aquesta hipòtesi, ajuntant les dues mostres tenim una mostra de mida n d'una multinomial, l'estimador de màxima versemblança de $P_i = P_1(A_i) = P_2(A_i)$ és

$$\hat{P}_i = \frac{n_{1,i} + n_{2,i}}{n_1 + n_2} = \frac{n_{\bullet,i}}{n}, \quad \forall i = 1, \dots, m$$

Es pot demostrar que l'estadístic de la chi-quadrat de Pearson convergeix asimptòticament:

$$D_{n_1, n_2} = \sum_{i=1}^m \frac{(n_{1,i} - n_1 \hat{P}_i)^2}{n_1 \hat{P}_i} + \sum_{i=1}^m \frac{(n_{2,i} - n_2 \hat{P}_i)^2}{n_2 \hat{P}_i} \xrightarrow{\mathcal{L}} \chi_{(2m-2-(m-1))}^2 = \chi_{(m-1)}^2$$

i $A_0 = \{D_{n_1, n_2} \leq C_{\alpha, m}\}$ amb $P(\chi_{(m-1)}^2 > c_{\alpha, m}) = \alpha$.

(Per què els graus de llibertat de la χ^2 són aquests? Per a cada mostra tenim m categories, per tant diem que té $m - 1$ graus de llibertat, ja que el nombre d'ocurrències de la última categoria ve donat pel total menys la suma de les anteriors. Com que tenim 2 mostres, multipliquem aquest nombre per 2, obtenint $2m - 2$ graus de llibertat. Ara bé, en fer l'estadístic D_{n_1, n_2} fem servir l'estimador $\hat{P}_i$, i cada cop que fem servir un estimador també perdem un grau de llibertat. En aquest cas en fem servir m : $\hat{P}_i$, $i = 1, \dots, m$. El que també passa aquí és que com que $\sum_{i=1}^m \hat{P}_i = 1$, l'últim queda determinat per la resta, de manera que ens queda aquest $m - 1$)

Observació 20

Es pot generalitzar a k mostres.

5.3.2 Test de Kolmogorov-Smirnov

Aquest següent test el podem aplicar quan les dues mostres provenen de distribucions contínues. Considerarem dues mostres independents $(x_1, \dots, X_{n_1})$ i $(y_1, \dots, y_{n_2})$ que provenen de lleis P_1 i P_2 respectivament amb funcions de distribució F_1 i F_2 respectivament. Contrastarem

$$H_0: P_1 = P_2 (\iff F_1 = F_2) \text{ contra } H_1: P_1 \neq P_2 (\iff F_1 \neq F_2)$$

Definirem l'estadístic de Kolmogorov-Smirnov com

$$\Delta_{n_1, n_2} = \sup_{y \in \mathbb{R}} |F_n^1(y) - F_n^2(y)|$$

amb F_n^1, F_n^2 les funcions de distribució empíriques associades a les mostres:

$$F_n^1(y) = \frac{1}{n_1} \sum_{i=1}^n 1_{(-\infty, y]}(x_i),$$

$$F_n^2(y) = \frac{1}{n_2} \sum_{i=1}^n 2_{(-\infty, y]}(y_i)$$

Es pot provar que Δ_{n_1, n_2} té una llei que no depèn ni de P_1 ni de P_2 , només de n_1 i n_2 . La regió d'acceptació serà:

$$A_0 = \{\Delta_{n_1, n_2} \leq c_{\alpha, n_1, n_2}\} \text{ amb } P(A_0^c) = \alpha$$

on P és la llei de A_{n_1, n_2} .

En aquest cas també existeixen taules pels valors de les constants c_{α, n_1, n_2} per a n_1 i n_2 petits i resultats asimptòtics per a valors més grans, però aquí no els veurem.

5.4 Test d'independència

Quan en un mateix individu es pot mesurar més d'una característica es poden plantejar si estan relacionades o no. Observem dues característiques (quantitatives o qualitatives) per a cada individu $\{(x_1, y_1), \dots, (x_n, y_n)\}$ (mostra aleatòria simple) d'un vector (X, Y) i que cadascuna d'aquestes dues característiques poden adquirir un nombre finit de valors $\{u_1, \dots, u_r\}$ i $\{v_1, \dots, v_s\}$ respectivament. Així, $\Omega = (\{u_1, \dots, u_r\} \times \{v_1, \dots, v_s\})^n$ i la família de probabilitats conjuntes $\mathcal{P} = \{\theta = (\theta_{i,j}), i = 1, \dots, r; j = 1, \dots, s; 0 \leq \theta_{i,j} \leq 1; \sum_{i=1}^r \sum_{j=1}^s \theta_{i,j} = 1\}$. Volem contrastar si X i Y són independents:

$$H_0: \theta_{i,j} = p_i q_j, \forall (i, j) \in \{1, \dots, r\} \times \{1, \dots, s\}$$

contra

$$H_1: \theta_{i,j} \neq p_i q_j \text{ per a algun } (i, j) \in \{1, \dots, r\} \times \{1, \dots, s\}$$

Disseminem $n_{i,j} = (u_i, v_j)$ és a dir quantes parelles (u_i, v_j) hi ha, $n_{i,\bullet} = \sum_{j=1}^s n_{i,j} = u_i$ el nombre de parelles amb u_i a la primera component i $n_{\bullet,j} = \sum_{i=1}^r n_{i,j} = v_j$ el nombre de parelles que tenen v_j a la segona component. Utilitzarem l'estadístic chi-quadrat de Pearson considerant que tenim una mostra de mida n d'una llei multidimensional que té $r \times s$ valors possibles, ja que cada (x_l, y_l) es pot representar per un 1 a la casella (i, j) , on $x_l = u_i$ i $y_l = v_j$ i zero a la resta. Sota la hipòtesi nul·la tenim dues mostres independents $x_1, \dots, x_{n_1}$ i $y_1, \dots, y_{n_2}$, cadascuna amb llei multinomial de paràmetres $(n, p_1, \dots, p_r)$ i $(n, q_1, \dots, q_s)$. Els estimadors de màxima versemblança són

$$\hat{p}_i = \frac{n_{i,\bullet}}{n}, \forall i = 1, \dots, r;$$

$$\hat{q}_j = \frac{n_{\bullet,j}}{n}, \forall j = 1, \dots, s;$$

Com ja hem vist en apartats anteriors, hem fet $(r-1)(s-1)$ estimacions, ja que tenim les condicions $\sum_{i=1}^r \hat{p}_i = \sum_{j=1}^s \hat{q}_j = 1$. Utilitzarem l'estadístic chi-quadrat de Pearson

$$D_n(x, y) = \sum_{i=1}^r \sum_{j=1}^s \frac{(n_{i,j} - n\hat{p}_i\hat{q}_j)^2}{n\hat{p}_i\hat{q}_j}$$

que, sota la hipòtesi nul·la convergeix cap a $\chi^2_{(rs-(r-1)-(s-1)-1)} = \chi^2_{(rs-r-s+1)}$ i la regió d'acceptació serà $A_0 = \{D_n(x, y) \leq c_{\alpha, rs-r-s+1}\}$ amb $P(\chi^2_{(rs-r-s+1)} > c_{\alpha, rs-r-s+1}) = \alpha$

5.5 Altres tests no paramètrics

Esmentarem sense explicar altres tests que es podrien realitzar

Tests d'aleatorietat Són tests per garantir que la mostra sigui realment aleatòria.

- Test de ratxes per a variables dicotòmiques
- Test de ratxes per a variables quantitatives
- Test de Von Neumann

Altres tests d'ajustament A part dels dos tests d'ajustament estudiats també hi ha el test G^2 de la raó de versemblança i el test de normalitat de Shapiro-Wilks.

Tests de localització o posició Serveixen per contrastar si un conjunt d'observacions prové d'una variable amb una certa mesura de posició o localització, habitualment un percentil. Per una mostra:

- Test de signes de Fisher
- Test de rangs de Wilcoxon

Per a dades aparellades:

- Test de signes de Fisher
- Test de rangs de Wilcoxon

Per a mostres independents:

- Test de la mediana
- Test de Wald-Wolfowitz
- Test de Kruskal-Wallis
- Test de Mann-Whitney

Tests d'associació Utilitzats per comprovar si existeix o no associació entre dues variables d'una població bidimensional. Sota normalitat, són d'independència.

- Test de correlació de rangs de Spearman
- Test de Kendall
- Test de Theie (dispersió)

6 Models lineals

Considerem una mostra formada per n parelles $(x_1, y_1), \dots, (x_n, y_n)$.

Definició 19

Dues variables (X, Y) segueixen un model de regressió lineal simple si existeixen dues constants β_0, β_1 i una variable aleatòria ϵ centrada (esperança 0) i amb variància finita σ^2 tal que per a cada valor fixat x , el valor associat de y compleix

$$y = \beta_0 + \beta_1 x + \epsilon \quad (7)$$

Observació 21

La variable X (que no és aleatòria sinó determinista) s'anomena independent, predictòria o explicativa, no és aleatòria i els valors són habitualment fixats per l'experimentador (és coneguda). Donada X , la variable aleatòria Y s'anomena variable dependent o de resposta. A la variable ϵ se l'anomena error.

6.1 Estimació dels coeficients del model lineal

Volem trobar $\hat{\beta}_0, \hat{\beta}_1$ tal que facin que la recta de regressió tingui el mínim error possible. Suposem que tenim

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad 1 \leq i \leq n$$

amb ϵ_i variables aleatòries independents idènticament distribuïdes centrades i amb variància finita σ^2 . Els paràmetres del model β_0, β_1 es diuen coeficients de regressió. Podem calcular

$$\begin{aligned} EY_i &= E(\beta_0 + \beta_1 X_i + \epsilon_i) = \beta_0 + \beta_1 X_i + E\epsilon_i = \beta_0 + \beta_1 X_i \\ \text{Var}(Y_i) &= \text{Var}(\beta_0 + \beta_1 X_i + \epsilon_i) = \text{Var}(\epsilon_i) = \sigma^2 \\ \text{Cov}(Y_i, Y_j) &= \begin{cases} \sigma^2 & i = j \\ 0 & i \neq j \end{cases} \end{aligned}$$

De moment no sabem la llei de ϵ_i , de manera que no podem utilitzar, per exemple, el mètode de la màxima versemblança pels paràmetres.

6.1.1 Mètode dels mínims quadrats

Trobem $\hat{\beta}_0$ i $\hat{\beta}_1$ tal que la recta $y = \hat{\beta}_0 + \hat{\beta}_1 x$ faci mínima les distàncies verticals entre els punts observats (x_i, y_i) i els punts de la recta estimada (x_i, y'_i) , amb $y'_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$. Volem minimitzar

$$f(\hat{\beta}_0, \hat{\beta}_1) = \sum_{i=1}^m (y_i - y'_i)^2 = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2$$

Per minimitzar aquesta funció, fem el sistema de les derivades parcials de $\hat{\beta}_0$ i $\hat{\beta}_1$ i comprovem amb la Hessiana.

Teorema 8

Els valors que minimitzen la funció són $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$ i $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$, amb $\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$, $S_{xy} = \sum_{i=1}^n (x_i - \bar{X})(y_i - \bar{Y})$

Proposició 12

Anomenarem $\bar{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ al valor estimat i $\hat{\epsilon}_i = y_i - \hat{y}_i$ els errors.

- $\sum_i \hat{\epsilon}_i = 0$

Demostració. $\sum_i (y_i - \hat{y}_i) = \sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = \sum_i (y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i) = \sum_i (y_i - \bar{y}) + \hat{\beta}_1 \sum_i (\bar{x} - x_i) = 0$ $\square$

.

- $\sum_i y_i = \sum_i \hat{y}_i$.

- La recta de regressió passa pel punt $(\bar{x}, \bar{y})$, ja que $\hat{\beta}_0 + \hat{\beta}_1 \bar{x} = \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x} = \bar{y}$

- $\sum_i x_i \hat{\epsilon}_i = 0$ i $\sum_i x_i \hat{\epsilon}_i = 0$

Demostració.

$$\begin{aligned} \sum_i x_i \hat{\epsilon}_i &= \sum_i x_i (y_i - \hat{y}_i) \stackrel{(1)}{=} \sum_i x_i (y_i - \hat{y}_i) - \sum_i \bar{x}_i (y_i - \hat{y}_i) = \\ &= \sum_i (x_i - \bar{x}_i) (y_i - \bar{y}) \stackrel{(2)}{=} \sum_i (x_i - \bar{x}) (y_i - \bar{y}) + \sum_i (x_i - \bar{x}) (\bar{y} - \hat{y}_i) = \\ &\stackrel{(3)}{=} S_{xy} - \beta_1 \sum_i (x_i - \bar{x})^2 = S_{xy} - \frac{S_{xy}}{S_{xx}} S_{xx} = 0 \end{aligned}$$

(1) Com que $\sum_i (y_i - \hat{y}_i) = 0$, si li restem una altra cosa que també val 0 no estem canviant res

(2) Sumem i restem $\bar{y}$

(3) Tenim que $\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x}$ i $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

□

5. El valor mínim de la funció $\sum_i (y_i - \beta_0 - \beta_1 x_i)^2$ és $S_{yy}(1 - R_{xy}^2)$, amb $R_{xy} = \frac{S_{xy}}{\sqrt{S_x S_y}}$ el coeficient de correlació lineal.

Demostració.

$$\begin{aligned} \sum_i (y_i - \bar{y} + \hat{\beta}_i \bar{x} - \hat{\beta}_i x_i)^2 &= \sum_i (y_i - \bar{y} - \hat{\beta}_i (x_i - \bar{x}))^2 = \\ &= S_{yy} + \hat{\beta}_1^2 S_{xx} - 2\hat{\beta}_1 S_{xy} = S_{yy} + \frac{S_{xy}^2}{S_{xx}^2} S_{xx} - 2\frac{S_{xy}}{S_{xx}} S_{xy} = \\ &= S_{yy}(1 - R_{xy}^2) \end{aligned}$$

□

Com més proper sigui R_{xy}^2 a 1 més petita és la suma, que és un bon indicador de la "qualitat" de l'ajust.

6. R_{xy}^2 s'anomena coeficient de determinació i és el quocient entre la variabilitat de la recta de regressió i de les dades

$$R_{xy}^2 = \frac{\sum_i (\hat{y}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

6.1.2 Propietats dels estimadors dels coeficients de regressió

Podem escriure

$$\begin{aligned} \hat{\beta}_1 &= \frac{S_{xy}}{S_{xx}} = \frac{1}{S_{xx}} \left[\sum_i (x_i - \bar{x}) y_i - \bar{y} \sum_i (x_i - \bar{x}) \right] = \frac{1}{S_{xx}} \sum_i (x_i - \bar{x}) y_i \\ \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} = \sum_i \frac{y_i}{n} - \bar{x} \sum_i \left(\frac{x_i - \bar{x}}{S_{xx}} \right) y_i = \sum_i \left(\frac{S_{xx} - n\bar{x}(x_i - \bar{x})}{nS_{xx}} \right) y_i = \sum_i \left(\frac{\sum_j x_j^2 - n\bar{x}x_i}{nS_{xx}} \right) y_i \end{aligned}$$

En podem deduir les propietats següents:

1. $\hat{\beta}_0$ i $\hat{\beta}_1$ no són esbiaixats:

$$E(\hat{\beta}_1) = \sum_i \left(\frac{x_i - \bar{x}}{S_{xx}} \right) E(y_i) = \sum_i \left(\frac{x_i - \bar{x}}{S_{xx}} \right) (\beta_0 + \beta_1 x_i) = \beta_1 \frac{\sum_i x_i - n\bar{x}}{S_{xx}} = \beta_1$$

$$E(\hat{\beta}_0) = E(\bar{y} - \bar{x}\hat{\beta}_1) = \beta_0 + \beta_1 \bar{x} - \beta_1 \bar{x} = \beta_0$$

2. Tenen per variància:

$$Var(\hat{\beta}_1) = \sigma^2 \sum_i \left(\frac{x_i - \bar{x}}{S_{xx}} \right)^2 = \frac{\sigma^2}{S_{xx}^2} S_{xx} = \frac{\sigma^2}{S_{xx}}$$

$$Var(\hat{\beta}_0) = \frac{\sigma^2}{n^2 S_{xx}^2} \left[n \left(\sum_j x_j^2 \right)^2 - n^2 \bar{x}^2 \sum_j x_j^2 \right] = \frac{\sigma^2 \sum_j x_j^2}{n S_{xx}^2} \left[\sum_j x_j^2 - n \bar{x} \right] = \frac{\sigma^2 \sum_j x_j^2}{n S_{xx}} = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$$

3. Covariància entre $\hat{\beta}_0$ i $\hat{\beta}_1$:

$$Cov(\hat{\beta}_0, \hat{\beta}_1) = \frac{\sigma^2}{n S_{xx}^2} \sum_i (x_i - \bar{x}) \left(\sum_j x_j^2 - n \bar{x} x_i \right) = -\frac{\sigma^2}{n S_{xx}^2} n \bar{x} \sum_i (x_i - \bar{x})^2 = -\frac{\sigma^2 \bar{x}}{S_{xx}}$$

6.1.3 Teorema de Gauss-Markov

Els estimadors $\hat{\beta}_0$ i $\hat{\beta}_1$ són estimadors lineals en y i no esbiaixats de β_0 i β_1 respectivament amb variància mínima. És a dir, per a tots T, S estimadors lineals en y i no esbiaixats de β_0 i β_1 respectivament es compleix que $Var(T) \geq Var(\hat{\beta}_0)$ i $Var(S) \geq Var(\hat{\beta}_1)$. A més, tenim la igualtat si i només si coincideixen ($\hat{\beta}_0 = T$ o $\hat{\beta}_1 = S$).

6.1.4 Estimador de la variància σ^2

L'estimador proposat és

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum \hat{\epsilon}_i^2 = \frac{1}{n-2} \sum_i (y_i - \hat{y}_i)^2 = \frac{1}{n-2} s_{yy} (1 - R_{xy}^2)$$

Comprovem ara que no té biaix. Tenim

$$\sum_i (y_i - \hat{y}_i)^2 = \sum_i (y_i - \hat{y}_i) y_i = \sum_i y_i^2 - \sum_i y_i \hat{y}_i = \sum_i y_i^2 - \sum_i \hat{y}_i^2$$

Calculem l'esperança de cada suma parcial de l'últim terme:

$$E \left(\sum_i y_i^2 \right) = \sum_i [Var(y_i) + (E(Y_i))^2] = n\sigma^2 + \sum_i (E(y_i))^2$$

$$E \left(\sum_i \hat{y}_i^2 \right) = \sum_i [Var(\hat{y}_i) + (E(\hat{y}_i))^2] = \sum_i Var(\hat{y}_i) + \sum_i (E(\hat{y}_i))^2 = \sum_i Var(\hat{y}_i) + \sum_i (E(y_i))^2$$

Aleshores

$$E \left(\sum_i (y_i - \hat{y}_i)^2 \right) = n\sigma^2 - \sum_i Var(\hat{y}_i)$$

Finalment

$$\sum_i \text{Var}(\hat{y}_i) = n\text{Var}(\hat{\beta}_0) + \left(\sum_i x_i^2\right)\text{Var}(\hat{\beta}_1) + 2\left(\sum_i x_i\right)\text{Cov}(\hat{\beta}_0, \hat{\beta}_1) = 2\sigma^2$$

Aleshores

$$E\left(\sum_i (y_i - \hat{y}_i)^2\right) = n\sigma^2 - 2\sigma^2 = (n-2)\sigma^2$$

Substituint a la fórmula que hem donat de l'estimador proposat dona $\hat{\sigma}^2$.

6.2 Model normal de regressió simple

Per tal de fer test d'hipòtesis i buscar intervals de confiança cal conèixer la llei dels errors. La suposició més usual és prendre $\epsilon_i \sim N(0, \sigma^2)$ independents i idènticament distribuïdes. Aleshores $y_i \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$ i

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}, \begin{pmatrix} \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right) & -\sigma^2 \frac{\bar{x}}{S_{xx}} \\ -\sigma^2 \frac{\bar{x}}{S_{xx}} & \sigma^2 \frac{1}{S_{xx}} \end{pmatrix}\right)$$

També sabem que

$$L(y_1, \dots, y_m, \beta_0, \beta_1, \sigma^2) = \frac{1}{\sigma^m \sqrt{(2\pi)^m}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right\}$$

Observació 22

Com $\sigma^2 > 0$, la versemblança serà més gran com més petita sigui la suma $\sum_i (y_i - \beta_0 - \beta_1 x_i)^2$, i així els estimadors de màxima versemblança de β_0 i β_1 coincideixen amb els trobats pel mètode dels mínims quadrats. Recordem que $\sum_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = S_{yy}(1 - R_{xy}^2)$. Aleshores

$$\frac{-1}{\sigma^2} \ln L(y_1, \dots, y_m, \beta_0, \beta_1, \sigma^2) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^2} S_{yy}(1 - R_{xy}^2) = 0 \implies \tilde{\sigma}^2 = \frac{1}{n} S_{yy}(1 - R_{xy}^2)$$

que no és el mateix que $\hat{\sigma}^2$: $\tilde{\sigma}^2 = \frac{n-2}{n} \hat{\sigma}^2$.

Pot demostrar-se que $\frac{\tilde{\sigma}^2}{\sigma^2}$ independent de $(\hat{\beta}_0, \hat{\beta}_1)$ i $\frac{n\tilde{\sigma}^2}{\sigma^2} = \frac{(n-2)\hat{\sigma}^2}{\sigma^2} \sim \chi_{(n-2)}^2$.