Estadística

Grau en Matemàtiques

Tercer Curs

Pedro Delicado
Guadalupe Gómez
Jan Graffelman

Departament d'Estadística i Investigació Operativa

Universitat Politècnica de Catalunya

9 de febrer de 2015

Índex

Pròleg

Aquests apunts presenten una part dels continguts de l'assignatura d'*Estadística* del Grau en Matemàtiques de la FME. Els inicis d'aquests apunts es troben en la assignatura Estadística Matemàtica II de la antiga Diplomatura d'Estadística, i en l'assignatura Inferència Estadística de l'antiga Llicenciatura en Matemàtiques Volem agrair a Montserrat Amblàs la feina feta en l'edició d'aquests apunts. Els apunts es troben en procés d'actualització per l'assignatura d'*Estadística*.

Barcelona, 1 de febrer de 2014

Capítol 1

ESTADÍSTICA DESCRIPTIVA

Capítol 2

ESTIMACIÓ PUNTUAL

2.1 Mètodes per a trobar estimadors

2.1.1 El mètode dels moments

Definició 2.1 *Considerem la variable aleatòria X amb funció de distribució F. Prenem també una mostra aleatòria simple de mida n de X, és a dir, $X_1, \dots, X_n$ v.a. i.i.d. amb distribució donada per F. Sigui $x_1, \dots, x_n$ una realització d'aquesta m.a.s. Anomenem* **funció de distribució empírica** *a la funció*

$$F_n(x) = \frac{1}{n} \# \{ x_i \le x : i = 1 \dots n \} = \frac{1}{n} \sum_{i=1}^{n} I_{(-\infty,x]}(x_i)$$

que a cada nombre real x li assigna la proporció de valors observats que són menors o iguals que x.

És immediat comprovar que la funció F_n és una funció de distribució:

1. $F_n(x) \in [0,1] \quad \forall x \in \mathbb{R}$.
2. F_n és contínua per la dreta.
3. F_n és no decreixent.
4. $\lim_{n \to -\infty} F_n(x) = 0$.
5. $\lim_{n \to \infty} F_n(x) = 1$.

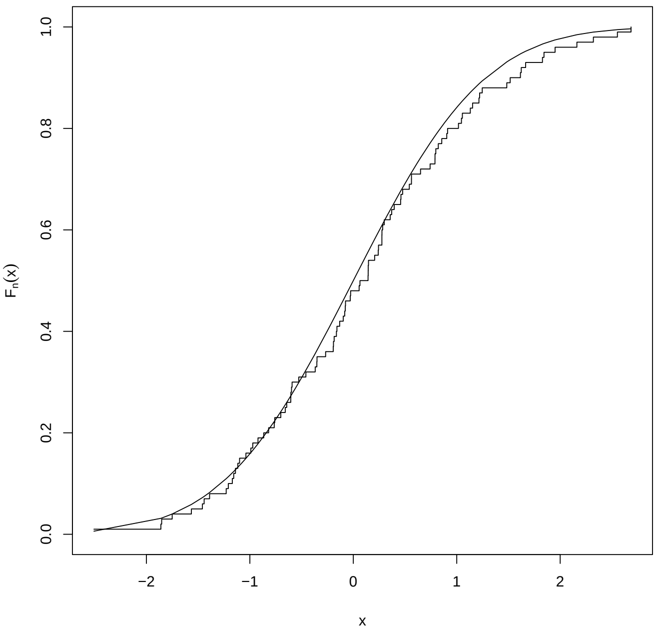

Figura 2.1: Funció de distribució empírica i teòrica per una mostra de $n = 100$ observacions d'una distribució $N(0, 1)$.

Proposició 2.1 *F_n és la funció de distribució d'una variable aleatòria discreta (que podem anomenar X_e) que posa massa $1/n$ a cadascun dels n punts x_i observats:*

x_i	x_1	x_2	$\ldots$	x_n
$p_i = \mathbf{P}(X_e = x_i)$	$1/n$	$1/n$	$\ldots$	$1/n$

A la distribució de X_e se l'anomena **distribució empírica** associada al conjunt de valors $\{x_1, \ldots, x_n\}$.

Teorema 2.1 Teorema de Glivenko-Cantelli

Sigui $\{X_n\}_{n\geq 1}$ una successió de v.a. i.i.d. definides a l'espai de probabilitat $(\Omega, \mathcal{A}, \mathbf{P})$, amb funció de distribució comú F. Denotem per F_n la funció de distribució empírica obtinguda de les n primeres variables aleatòries $X_1, \ldots, X_n$. Aleshores,

$$\sup_{x\in \mathbb{R}} |\, F_n(x) - F(x) \,| \longrightarrow 0 \ \text{ quasi segur}$$

Això garanteix la possibilitat de realitzar inferència estadística: els aspectes probabilístics d'una característica X, mesurada en una certa població, es resumeixen en una distribució de probabilitat F, la qual es pot aproximar per mitjà de les distribucions empíriques F_n (obtingudes per mostreig de la població). El teorema anterior afirma que aquestes aproximacions són uniformes en x. És per aquest motiu que de vegades el teorema de Glivenko-Cantelli també rep el nom de Teorema fonamental de l'Estadística Matemàtica, ja que dóna una fonamentació de la inferència estadística.

Tot seguit presentarem una conseqüència important de la convergència de F_n a F, la definició d'estimadors mitjançant el principi de substitució.

La convergència de F_n a F permet construir versions factibles de característiques poblacionals desconegudes. Suposem que volem estudiar una característica X en una població i que el resultat de l'observació de X pot ser modelat com una variable aleatòria amb distribució desconeguda, diguem-li F. Moltes de les preguntes rellevants sobre la característica X es podrien contestar si la seva funció de distribució F fos coneguda.

Per tal de fixar idees podem pensar que ens interessa conèixer quantitats numèriques (paràmetres) que depenen únicament de la funció de distribució desconeguda F:

$$\theta = \psi(F).$$

El teorema de Glivenko-Cantelli ens diu que F_n s'acosta a F a mesura que la mida de la mostra augmenta. Així, doncs, podem esperar que també es verifiqui que

$$\widehat{\theta}_n = \psi(F_n) \longrightarrow \theta = \psi(F),$$

és a dir, esperem que les quantitats numèriques calculades per la distribució empírica (estimadors) s'aproximin a les quantitats desconegudes a mesura que n creix. Es pot demostrar que aquest resultat és cert sota hipòtesis de regularitat força generals de les funcions ϕ que assignen nombres a funcions de distribució.

Definició 2.2 *Aquesta forma d'obtenir estimadors de paràmetres poblacionals desconeguts s'anomena* **principi de substitució**. *És un procediment molt general d'obtenció d'estimadors.*

Una aplicació molt important del principi de substitució és la definició dels estimadors basats en moments, que tractarem tot seguit.

Definició 2.3 *El* **moment no centrat d'ordre** k *d'una variable aleatòria* X *amb distribució* F *es defineix com*

$$\mu_k = E_F(X^k) = \int x^k \, dF(x)$$

Definició 2.4 *Si* X_e *és una variable aleatòria amb funció de distribució igual a* F_n*, la funció de distribució empírica d'una m.a.s. de mida* n *de* X*, es té que els seus moments no centrats (que notarem per* $m_{k,n}$*), són de la forma*

$$m_{k,n} = E_{F_n}(X_e^k) = \int x^k \, F_n(x) = \frac{1}{n}\sum_{i=1}^{n} X_i^k$$

i s'anomenen **moments mostrals no centrats d'ordre** k.

Si el paràmetre d'interès és funció dels moments (com ara $\sigma^2 = \mu_2 - \mu_1^2$) aleshores l'estimador pel mètode dels moments consisteix en substituir-los pels moments mostrals

$$\widehat{S_M}^2 = m_{2,n} - m_{1,n}^2$$

Exemple 2.1 ____

Sigui $X \sim N(\mu, \sigma^2)$. Llavors ja sabem que

$$\left.\begin{array}{lcl} \mu_1 & = & E(X) = \mu \\ \mu_2 & = & E(X^2) = \mu^2 + \sigma^2 \end{array}\right\} \Longrightarrow \begin{array}{lcl} \mu & = & \mu_1 \\ \sigma^2 & = & \mu_2 - \mu_1^2 \end{array}$$

Ja que $V(X) = E(X^2) - E(X)^2 = \sigma^2 \rightarrow E(X^2) = \sigma^2 + \mu^2$. Per tant,

$$m_{1,n} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X} \text{ és un estimador de } \mu = E(X)$$

$$m_{2,n} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 \text{ és un estimador de } E(X^2)$$

i per últim

$$m_{2,n} - (m_{1,n})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \overline{X}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2 \text{ és un estimador de } \sigma^2$$

____

Moltes característiques poblacionals d'interès es poden expressar com a funció dels moments no centrats d'ordres $1, \ldots, k$:

$$\theta = h(\mu_1, \ldots, \mu_k)$$

Per exemple, acabem de veure que la variància de X es pot expressar com $\sigma^2 = h(\mu_1, \mu_2) = \mu_2 - \mu_1^2$.

Definició 2.5 *L'estimador del paràmetre θ basat en el principi de substitució es coneix com* **l'estimador dels moments** *de θ i serà*

$$\widehat{\theta}_n = h(m_{1,n}, \ldots, m_{k,n})$$

Notem que l'estimador dels moments de θ no té perque ser únic, ja que diferents funcions h ens poden portar al mateix valor de θ.

Exemple 2.2

Sigui $X \sim U(0, \theta)$. Es pren una m.a.s. de X de mida n per tal d'estimar el paràmetre θ. L'estimador de moments $\widehat{\theta}_M$ de θ ve donat per la relació que segueix:

$$\mu_1 = E(X) = \frac{\theta}{2} \Longrightarrow m_{1,n} = \frac{\widehat{\theta}_M}{2} \Longrightarrow \widehat{\theta}_M = 2\, m_{1,n} = 2\, \overline{X}_n$$

Un altre estimador raonable per θ és el màxim de les observacions, que és un estadístic minimal suficient per a θ:

$$\widehat{\theta}_2 = \max_i X_i$$

Es tracta d'un estimador minimal basat en el principi de substitució. En efecte,

$$\theta = \sup \{\, x \in \mathbb{R} : F(x) < 1 \,\},$$

$$\widehat{\theta}_2 = \max_i X_i = \sup \{\, x \in \mathbb{R} : F_n(x) < 1 \,\}.$$

Propietats immediates

1. Els moments mostrals d'ordre k són estimadors no esbiaixats de $E(X^k)$, és a dir, $E(m_{k,n}) = \mu_k$.

   Definició 2.6 *Una successió d'estimadors* $\{T_n(X_1, \dots, X_n)\}_{n \geq 1}$ *del paràmetre* θ *es diu que és* **consistent** *si i només si*

$$T_n(X_1, \dots, X_n) \xrightarrow[n \to \infty]{P} \theta$$

2. El moment mostral d'ordre k és un estimador consistent de $E(X^k)$.

   En efecte, volem veure que

$$m_{k,n} = \frac{1}{n} \sum_{i=1}^{n} X_i^k \xrightarrow[n \to \infty]{P} \mu_k = E(X^k)$$

   Definim $Y_i = X_i^k$. Aleshores tenim que $Y_1, \dots, Y_n$ són i.i.d. amb $Var(Y_i) = Var(X_i^k)$ i per tant, per la llei dels grans nombres,

$$\frac{1}{n} \sum_{i=1}^{n} Y_i \xrightarrow[n \to \infty]{P} E(Y_i) = E(X_i^k)$$

   és a dir,

$$m_{k,n} \xrightarrow{P} \mu_k$$

   tal com volíem veure.

Exemple 2.3

Sigui $X \sim \Gamma(\alpha, \lambda)$. Es pren com abans una m.a.s. de X de mida n. Aleshores tenim

$$f(x\,;\,\alpha, \lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, e^{-\lambda x}$$

Per tant

$$\left.\begin{array}{rcl} \mu_1 &=& E(X) = \alpha/\lambda \\ \mu_2 &=& E(X^2) = \alpha\,(\alpha+1)/\lambda^2 \end{array}\right\} \Longrightarrow \begin{array}{rcl} \alpha &=& \lambda\,\mu_1 \\ \mu_2 &=& \lambda\,\mu_1\,(\lambda\,\mu_1 + 1)/\lambda^2 = \mu_1^2 + \frac{\mu_1}{\lambda} \end{array}$$

Així, doncs, tenim que

$$\lambda = \frac{\mu_1}{\mu_2 - \mu_1^2} \quad \text{i per tant} \quad \alpha = \frac{\mu_1^2}{\mu_2 - \mu_1^2}$$

Els estimadors de α i λ pel mètode dels moments són

$$\widehat{\alpha} = \frac{m_{1,n}^2}{m_{2,n} - m_{1,n}^2} = \frac{\overline{X}_n^2}{\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2}$$

$$\widehat{\lambda} = \frac{\overline{X}_n}{\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X}_n^2}$$

Construcció d'un estimador pel mètode dels moments

Cal seguir el següent esquema:

1. Identificar els paràmetres que volen ser estimats.
2. Calcular els primers moments.
3. Relacionar i expressar els paràmetres d'interès en funció dels primers moments.
4. Utilitzar la relació trobada en el pas anterior i substituir els moments poblacionals pels moments mostrals.

És a dir, si ens interessa estimar el parell (θ_1, θ_2), i trobem que

$$\begin{aligned} \theta_1 &= h_1(\mu_1, \mu_2) \\ \theta_2 &= h_2(\mu_1, \mu_2) \end{aligned}$$

Aleshores estimem

$$\begin{aligned} \widehat{\theta}_1 &= h_1(\widehat{\mu}_1, \widehat{\mu}_2) \\ \widehat{\theta}_2 &= h_2(\widehat{\mu}_1, \widehat{\mu}_2) \end{aligned}$$

Anem tot seguit a veure un exemple que posa de manifest que l'estimador obtingut seguint aquest mètode no té perque ser únic.

Exemple 2.4

Sigui $X \sim lognormal(\mu, \sigma^2)$, i prenem $X_1, \ldots, X_n$ una m.a.s. de X. Sabem que la funció de densitat de X ve donada per

$$f(x\,;\,\mu, \sigma^2) = \frac{1}{x\,\sigma\,\sqrt{2\,\pi}}\ exp\left\{-\frac{(\ln x - \mu)^2}{2\,\sigma^2}\right\}$$

Sabem també que

$$E(X) = exp\left\{\mu + \frac{\sigma^2}{2}\right\} \quad \text{i} \quad Var(X) = exp\left\{\,2\,\mu + \sigma^2\,\right\}\left(e^{\sigma^2} - 1\right)$$

<u>Resolució tipus 1</u>

Considerem

$$\mu_1 = e^{\mu + \frac{\sigma^2}{2}}$$

$$\mu_2 = e^{2\,\mu + \sigma^2}\left(e^{\sigma^2} - 1\right) + e^{2\,\mu + \sigma^2} = e^{2\,\mu + \sigma^2}\, e^{\sigma^2} = e^{2(\mu + \sigma^2)}$$

Aleshores, fent els logaritmes respectius tenim

$$\left.\begin{array}{rcl} log\,\mu_1 & = & \mu + \sigma^2/2 \\ log\,\mu_2 & = & 2\,\mu + \sigma^2 + \sigma^2 \end{array}\right\} \Longrightarrow \left.\begin{array}{rcl} log\,\mu_2 - 2\,log\,\mu_1 & = & \sigma^2 \\ log\,\mu_2 - 4\,log\,\mu_1 & = & -2\,\mu \end{array}\right\} \Longrightarrow$$

I amb tot això ja podem escriure

$$\sigma^2 = log\left(\frac{\mu_2}{\mu_1^2}\right) \quad \text{i} \quad \mu = \frac{1}{2}\,log\left(\frac{\mu_1^4}{\mu_2}\right)$$

Si ara definim

$$\widehat{\mu}_1 = m_{1,n} = \overline{X}_n \quad \text{i} \quad \widehat{\mu}_2 = m_{2,n} = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

ja tenim el següent:

$$\widehat{\mu} = \frac{1}{2}\,log\left(\frac{\widehat{\mu}_1^4}{\widehat{\mu}_2}\right) \quad \text{i} \quad \widehat{\sigma}^2 = log\left(\frac{\widehat{\mu}_2}{\widehat{\mu}_1^2}\right)$$

Resolució tipus 2

Com que $X \sim lognormal(\mu, \sigma^2)$, tenim que $log\, X \sim N(\mu, \sigma^2)$ i això ens permet definir

$$\tilde{\mu} = \frac{1}{n} \sum_{i=1}^{n} log\, X_i \quad \text{i} \quad \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} \big(log\, X_i - \tilde{\mu} \big)^2$$

2.1.2 La funció de versemblança. El mètode de la màxima versemblança

Definició 2.7 *Donada una mostra $X_1, \ldots, X_n$ amb funció de densitat (o de probabilitat) $f(x \,|\, \boldsymbol{\theta})$, on $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ definim la* **funció de versemblança** *per $\boldsymbol{x}$ fixat com*

$$\begin{array}{cccl} L_{\boldsymbol{x}} & : & \Theta & \longrightarrow \quad \mathbb{R}^+ \\ & & \boldsymbol{\theta} & \longrightarrow \quad L_{\boldsymbol{x}}(\boldsymbol{\theta}) = L(\theta_1, \ldots, \theta_k \,|\, x_1, \ldots, x_n) = \prod_{i=1}^{n} f(x_i \,|\, \boldsymbol{\theta}) \end{array}$$

El que donem a continuació són tot un seguit d'exemples de funcions de versemblança.

Exemple 2.5

Sigui $X_1, \ldots, X_n$ una m.a.s. tal que $X \sim Bern(p)$, aleshores sabem que

$$\mathbf{P}(X_1 = x_1 \,|\, p) = p^{x_1} \, (1-p)^{1-x_1}$$

i amb això ja podem escriure

$$\mathbf{P}(X_1 = x_1, \ldots, X_n = x_n \,|\, p) = \prod_{i=1}^{n} p^{x_i} \, (1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i} \, (1-p)^{n - \sum_{i=1}^{n} x_i}$$

amb la qual cosa la funció de versemblança és

$$L_1(p \,|\, x_1, \ldots, x_n) = p^{\sum_{i=1}^{n} x_i} \, (1-p)^{n - \sum_{i=1}^{n} x_i}$$

Exemple 2.6

Si ara prenem una m.a.s. $X_1', \ldots, X_m'$ de manera que $X_i' \sim Bin(k,p)$, amb k conegut, tenim en aquest cas que

$$\mathbf{P}(X_1' = x_1' \,|\, p) = \binom{k}{x_1'} p^{x_1'} (1-p)^{k-x_1'}$$

és a dir,

$$\mathbf{P}(X_1' = x_1', \ldots, X_m' = x_m' \,|\, p) = \prod_{i=1}^{m} \binom{k}{x_i'} p^{x_i'} (1-p)^{k-x_i'}$$

$$L_2(p \,|\, x_1', \ldots, x_m') = \left(\prod_{i=1}^{m} \binom{k}{x_i'} \right) p^{\sum_{i=1}^{m} x_i'} (1-p)^{m\,k - \sum_{i=1}^{m} x_i'}$$

Anem tot seguit a comparar aquestes dues funcions de versemblança. Suposem que quan realitzem l'exemple 2.5 amb una m.a.s. de $n = 30$, i observem $\sum_{i=1}^{30} x_i = S_1$. Aleshores

$$L_1(p \,|\, x_1, \ldots, x_{30}) = p^{S_1} (1-p)^{30-S_1}$$

Per altra banda, suposem també que en el segon experiment d'aquest exemple tenim $k = 3$ i realitzem $m = 10$ vegades, i observem també en aquest cas $\sum_{i=1}^{10} x_i' = S_2$. Aleshores

$$L_2(p \,|\, x_1', \ldots, x_{10}') = \left(\prod_{i=1}^{10} \binom{3}{x_i'} \right) p^{S_2} (1-p)^{30-S_2}$$

Si observem $S_1 = S_2$ tindrem que

$$\frac{L_1(p \,|\, x_1, \ldots, x_{30})}{L_2(p \,|\, x_1', \ldots, x_{10}')} = \frac{1}{\prod_{i=1}^{10} \binom{3}{x_i'}}$$

i per tant la inferència sobre p haurà de ser la mateixa.

Exemple 2.7

Prenem ara una m.a.s. $X_1, \ldots, X_n$ de tal manera que $X \sim exp(\,\mu\,)$. Sabem que la funció de densitat és

$$f(x_i \,|\, \mu) = \frac{1}{\mu}\,\exp\!\left(\frac{-x_i}{\mu}\right)$$

$$f(x_1, \ldots, x_n \,|\, \mu) = \frac{1}{\mu^n}\,\exp\!\left(-\frac{\sum_{i=1}^n x_i}{\mu}\right)$$

i la funció de versemblança serà la següent:

$$L(\mu \,|\, x_1, \ldots, x_n) = \frac{1}{\mu^n}\,\exp\!\left(\frac{-\sum_{i=1}^n x_i}{\mu}\right)$$

Exemple 2.8

Sigui $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma^2)$. La seva funció de densitat conjunta és

$$f(x_1, \ldots, x_n \,|\, \mu, \sigma^2) = \left(\frac{1}{\sqrt{2\,\pi\,\sigma^2}}\right)^n \exp\!\left\{-\frac{1}{2\,\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right\}$$

amb la qual la funció de versemblança serà en aquest cas

$$L(\mu, \sigma^2 \,|\, x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\,\pi\,\sigma^2}}\right)^n \exp\!\left\{-\frac{1}{2\,\sigma^2}\sum_{i=1}^n (x_i - \mu)^2\right\}$$

Si prenem una mostra de $n = 5$ alçades de noies de l'assignatura EM2 (en cm), suposem que $x_1 = 160, x_2 = 161, x_3 = 155, x_4 = 168, x_5 = 169$, aleshores

$$L(\mu,\sigma^2\,|\,160,161,155,168,169)=\left(\frac{1}{\sqrt{2\,\pi\sigma^2}}\right)^5\exp\Big\{-\frac{1}{2\sigma^2}\sum_{i=1}^{5}(x_i-\mu)^2\Big\}$$

Per a cada mostra $\boldsymbol{x}\in\mathcal{X}$ l'estimador de màxima versemblança de θ el denotem per $\widehat{\theta}(\boldsymbol{x})$ i correspon a un valor que atrapa el màxim de $L(\boldsymbol{\theta}\,|\,\boldsymbol{x})$ com a funció de $\boldsymbol{\theta}$.

El mètode de la màxima versemblança és un mètode intuïtivament apropiat perqu es tracta de seleccionar el valor de θ pel qual el nostre conjunt d'observacions té probabilitat màxima. Molt sovint els estimadors de màxima versemblança (MLE) són els que el sentit comú escolliria. A més, les propietats d'aquests estimadors són força bones.

Trobar el màxim de versemblança és un problema d'optimització.

Quan és possible val la pena dibuixar el logaritme de la funció de versemblança, en particular quan la dimensió de l'espai de paràmetres és 2 o 3. Aquest gràfic ens pot donar informació sobre el comportament de θ.

Exemple 2.9

Volem estimar $\theta=\mathbf{P}(X=1)$ a partir de n llançaments independents d'una moneda. Suposem que se sap que $\theta\in\Theta=\big\{1/4,1/2,3/4\big\}$ i basemnos en $n=4$. Sabem que

$$f_X(x\,;\,\theta)=\theta^x\,(1-\theta)^{1-x}\,1_{\{0,1\}}(x)$$

Donada una mostra x_1,x_2,x_3,x_4,

$$g(x_1x_2x_3x_4\,|\,\theta)=\prod_{i=1}^{n}f_X(x_i\,;\,\theta)=\theta^{\sum x_i}\,(1-\theta)^{4-\sum x_i}$$

Volem trobar θ. Veiem en Taula 2.1 què passaria per diferents valors del paràmetre:

θ	$1/4$	$1/2$	$3/4$	$\sum x_i$
$g(x_1x_2x_3x_4 \mid \theta)$	$81/256$	$16/256$	$1/256$	0
$g(x_1x_2x_3x_4 \mid \theta)$	$27/256$	$16/256$	$3/256$	1
$g(x_1x_2x_3x_4 \mid \theta)$	$9/256$	$16/256$	$9/256$	2
$g(x_1x_2x_3x_4 \mid \theta)$	$3/256$	$16/256$	$27/256$	3
$g(x_1x_2x_3x_4 \mid \theta)$	$1/256$	$16/256$	$81/256$	4

Taula 2.1: Valors de la funció de versemblança per diferents valors de $\sum x_i$ i θ.

Per tant, podem veure que

$$\widehat{\theta} = \frac{1}{4} \qquad \text{si } \sum x_i = 0 \text{ o } \sum x_i = 1$$

$$\widehat{\theta} = \frac{1}{2} \qquad \text{si } \sum x_i = 2$$

$$\widehat{\theta} = \frac{3}{4} \qquad \text{si } \sum x_i = 3 \text{ o } \sum x_i = 4$$

Per a cada $(x_1x_2x_3x_4)$, $\widehat{\theta}$ és aquell valor de Θ tal que

$$g(x_1x_2x_3x_4 \mid \widehat{\theta}\,) = \max_{\theta \in \Theta} g(x_1x_2x_3x_4 \mid \theta)$$

Ens podem trobar també amb alguns problemes associats amb l'estimador de la màxima versemblança.

Per exemple, el problema de trobar el màxim global i verificar que ho és. Una altra qüestió és la sensitivitat numèrica, és a dir, quan de sensible és el nostre estimador a petits canvis en les dades.

Anem a veure ara el que s'ha de fer per trobar l'estimador de la màxima versemblança (MLE):

a) Si $L(\boldsymbol{\theta} \mid \boldsymbol{x})$ és diferenciable en θ, les solucions de $\frac{\partial}{\partial \theta_i} L(\boldsymbol{\theta} \mid \boldsymbol{x}) = 0$ $i = 1, \ldots, k$ són possibles candidats per ser MLE. No obstant això, els zeros de la primera derivada no garanteixen el màxim, ja que poden ser màxims globals o locals, mínims globals o locals o també punts d'inflexió. D'altra banda, aquest procediment només localitza extrems a l'interior de Θ, amb la qual cosa cal verificar el valor de $L(\boldsymbol{\theta} \,|\, \boldsymbol{x})$ a la frontera de Θ.

Exemple 2.10

Considerem una variable aleatòria $X \sim exp(\lambda)$. Aleshores

$$f(x \,|\, \lambda) = \lambda\, e^{-\lambda\, x}$$

Anem a observar $x = 3$. Tenim que

$$L(\lambda \,|\, x = 3) = \lambda\, e^{-3\,\lambda}$$

funció que hem representat a la figura 2.2.

Volem el seu màxim, per tant, busquem la seva derivada:

$$L'(\lambda \,|\, x = 3) = e^{-3\,\lambda}\,(1 - 3\,\lambda)$$

funció que s'anul·la per $\lambda = \frac{1}{3}$, valor que efectivament és un màxim, ja que $L''(\lambda = 1/3 \,|\, x = 3) < 0$.

Anem ara a verificar el valor de la funció a la frontera de la regió:

$$\lim_{\lambda \to 0} \lambda\, e^{-\lambda\, x} = 0$$

$$\lim_{\lambda \to \infty} \lambda\, e^{-\lambda\, x} = 0$$

Així veiem que el màxim no s'assoleix a la frontera.

Exemple 2.11

Sigui $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, 1)$. Per tant, podem escriure

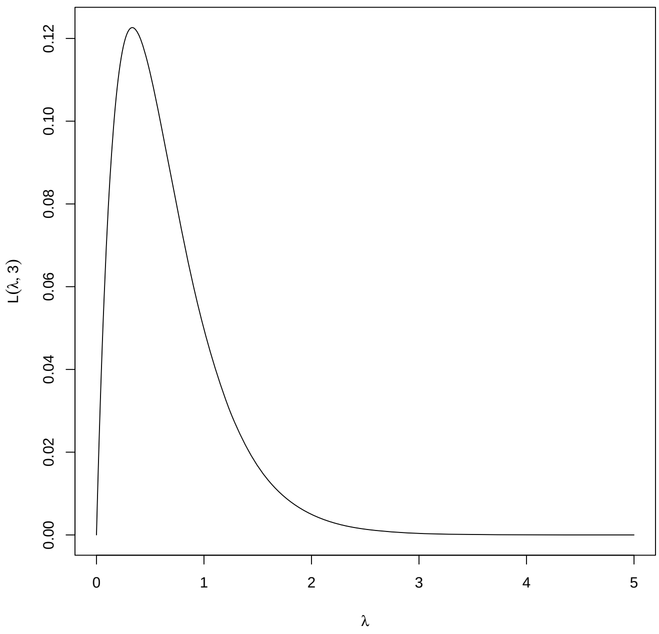

Figura 2.2: Funció de versemblança de la funció exponencial, per $x = 3$

$$L(\mu\,|\,\boldsymbol{x}) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x_i-\mu)^2} = \frac{1}{(2\,\pi)^{n/2}}\, e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$l(\mu, \mathbf{x}) = -\frac{n}{2}\ln(2\pi) - \frac{1}{2}\sum (x_i-\mu)^2$$

Si derivem, obtenim

$$\frac{d}{d\mu}\, l(\mu\,|\,x) = 0 \Longrightarrow \sum_{i=1}^{n}(x_i-\mu) = 0 \Longrightarrow \widehat{\mu} = \overline{x}$$

Aquest és, doncs, un candidat a ser l'estimador de la màxima versemblança, el que nosaltres hem anomenat MLE.

Per comprovar que és un màxim global (el nostre objectiu realment) veiem que $\widehat{\mu} = \overline{x}$ és l'única solució de $\frac{d}{d\mu}\, L(\mu\,|\,x) = 0$.

Notem que

$$\frac{d^2}{d\,\mu^2}\, L(\mu\,|\,x)|_{\mu=\overline{x}} < 0$$

i per tant $\overline{x}$ és l'únic punt extrem a l'interior de l'espai de paràmetres i és un màxim.

Per altra banda, hem de veure què passa a la frontera, en aquest cas $+\infty$ i $-\infty$.

Veiem que

$$\lim_{\mu\to+\infty} L(\mu\,|\,x) = \lim_{\mu\to-\infty} L(\mu\,|\,x) = 0,$$

per tant, l'estimador buscat és $\widehat{\mu} = \overline{x}$.

Aquest últim punt ens el podríem haver estalviat, ja que si $\overline{x}$ és l'únic extrem a l'interior i és un màxim no pot haver-hi màxim a l'infinit perquè si n'hi hagués, hi hauria d'haver un mínim a l'interior, cosa que aquí no passa.

---

b) També es pot trobar el màxim directament. Aquest procediment pot ser senzill quan la derivada tendeix a ser complicada però requereix més intuició, ja que no hi ha una metodologia a seguir. Una manera és trobar una fita superior de $L(\theta\,|\,x)$ i provar que només hi ha un punt que l'atrapi.

Exemple 2.11, pàgina 17. Continuació.

Com que

$$\sum_{i=1}^{n}(x_i-a)^2 \geq \sum_{i=1}^{n}(x_i-\overline{x})^2 \qquad \forall \ a$$

i la igualtat es dóna si i només si $a=\overline{x}$, això implica que

$$\forall \ \mu \quad L(\mu\,|\,x) \leq L(\overline{x}\,|\,x) \ \text{ amb igualtat si i només si } \ \mu=\overline{x}$$

Així, $\widehat{\mu}=\overline{x}$ és el MLE.

---

c) Molt sovint és més senzill treballar amb el logaritme de L, el $logL(\,\theta\,|\,x)$, ja que al ser la funció logarítmica estrictament creixent en $(0,\infty)$, els extrems de $L(\theta\,|\,x)$ i de $logL(\,\theta\,|\,x)$ han de coincidir.

Exemple 2.12

Considerem en aquest una m.a.s. $X_1,\ldots,X_n$ de $X\sim Bern(p)$, amb $\Theta=[0,1]$. La funció de versemblança és

$$L(p\,|\,x)=p^{\sum_{i=1}^{n}x_i}\,(1-p)^{n-\sum_{i=1}^{n}x_i}$$

$$log\,L(p\,|\,x)=\left(\sum_{i=1}^{n}x_i\right)log\,p\,+\left(n-\sum_{i=1}^{n}x_i\right)log(1-p)$$

Busquem ara els extrems d'aquesta funció

$$\frac{d}{d\,p}\,log\,L(p\,|\,x)=\frac{\sum_{i=1}^{n}x_i}{p}-\frac{n-\sum_{i=1}^{n}x_i}{1-p}=0 \Longrightarrow \widehat{p}=\frac{\sum_{i=1}^{n}x_i}{n}$$

ja que $0<\sum\limits_{i=1}^{n}x_i<n$.

Per tant, $\frac{\sum_{i=1}^{n}x_i}{n}$ és l'única solució a l'interior de $\Theta=[0,1]$.

Veiem ara si es tracta d'un màxim o d'un mínim:

$$\frac{d^2}{d\,p^2}\,log\,L(p\,|\,x)\bigg|_{p=\widehat{p}}=-\frac{\sum_{i=1}^{n}x_i}{p^2}+\frac{\sum_{i=1}^{n}x_i-n}{(1-p)^2}\bigg|_{p=\widehat{p}}=$$

$$-\frac{n\,\widehat{p}}{\widehat{p}^2}-\frac{n\,(1-\widehat{p})}{(1-\widehat{p})^2}=-\Big(\frac{n}{\widehat{p}}+\frac{n}{1-\widehat{p}}\Big)<0$$

amb la qual cosa és un màxim.

A les fronteres tenim que $L(0\,|\,x) = 0$ i $L(1\,|\,x) = 0$

Tractarem ara el cas multivariànciat.

Definició 2.8 *Si $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_k)$ i $X \sim f(x\,|\,\boldsymbol{\theta})$, la funció de versemblança es defineix anàlogament com*

$$L(\boldsymbol{\theta}\,|\,\boldsymbol{x}) = \prod_{i=1}^{n} f(x_i\,|\,\boldsymbol{\theta})$$

i el màxim de versemblança $\widehat{\boldsymbol{\theta}}(\boldsymbol{x})$ és un vector de components $\widehat{\boldsymbol{\theta}}(\boldsymbol{x}) = \big(\widehat{\theta}_1(\boldsymbol{x}), \ldots, \widehat{\theta}_k(\boldsymbol{x})\big)$, que maximitza $L(\boldsymbol{\theta}\,|\,\boldsymbol{x})$.

Exemple 2.13 ---

Sigui una m.a.s. $X_1, \ldots, X_n$ tal que $X \sim N(\theta, \sigma^2)$. La seva funció de versemblança serà

$$L(\,\theta, \sigma^2\,|\,\boldsymbol{x}) = \frac{1}{(2\,\pi\,\sigma^2)^{n/2}}\, e^{-\frac{1}{2\,\sigma^2}\,\sum_{i=1}^{n}(x_i-\theta)^2}$$

$$log\,L(\,\theta, \sigma^2\,|\,\boldsymbol{x}) = -\frac{n}{2}\,log\,2\,\pi - \frac{n}{2}\,log\,\sigma^2 - \frac{1}{2\,\sigma^2}\,\sum_{i=1}^{n}(x_i-\theta)^2$$

Busquem els extrems d'aquesta funció:

$$\frac{\partial}{\partial\theta}\,log\,L(\theta, \sigma^2\,|\,\boldsymbol{x}) = \frac{1}{\sigma^2}\,\sum_{i=1}^{n}(x_i-\theta) = 0$$

$$\frac{\partial}{\partial(\sigma^2)}\,log\,L(\theta, \sigma^2\,|\,\boldsymbol{x}) = -\frac{n}{2\,\sigma^2} + \frac{1}{2\,\sigma^4}\,\sum_{i=1}^{n}(x_i-\theta)^2 = 0$$

la qual cosa implica

$$\widehat{\theta} = \overline{x} \quad \text{i} \quad \widehat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2$$

Per provar que la solució és un màxim ho podem fer de dues maneres:

A) Com que per $\theta \neq \overline{x}$,

$$\sum_{i=1}^{n} (x_i - \theta)^2 > \sum_{i=1}^{n} (x_i - \overline{x})^2 \Longrightarrow L(\overline{x}, \sigma^2 \,|\, x) \geq L(\theta, \sigma^2 \,|\, x) \quad \forall \, \sigma^2$$

i ara cal provar que $L(\overline{x}, \sigma^2 \,|\, \boldsymbol{x})$ atrapa el màxim a $\widehat{\sigma}^2$ com a problema univariànciat.

B) Mitjançant càlcul a $\mathbb{R}^2$ haurem de verificar que

- $$\frac{\partial}{\partial \theta_j} L(\theta_1, \theta_2)\Big|_{\theta_1=\widehat{\theta}_1, \theta_2=\widehat{\theta}_2} = 0 \qquad j = 1, 2$$

- $$\frac{\partial^2}{\partial \theta_j^2} L(\theta_1, \theta_2)\Big|_{\theta_1=\widehat{\theta}_1, \theta_2=\widehat{\theta}_2} < 0 \qquad \text{almenys per } j = 1 \text{ o } j = 2$$

- $$\boldsymbol{H} = \left| \begin{array}{cc} \frac{\partial^2}{\partial \theta_1^2} L(\theta_1, \theta_2) & \frac{\partial^2}{\partial \theta_1 \, \partial \theta_2} L(\theta_1, \theta_2) \\ \frac{\partial^2}{\partial \theta_1 \, \partial \theta_2} L(\theta_1, \theta_2) & \frac{\partial^2}{\partial \theta_2^2} L(\theta_1, \theta_2) \end{array} \right|_{\theta_1=\widehat{\theta}_1, \theta_2=\widehat{\theta}_2} > 0$$

per tal d'afirmar que $(\widehat{\theta}_1, \widehat{\theta}_2)$ és un màxim.

En l'exemple,

$$\frac{\partial^2}{\partial \theta^2} \, log \, L(\theta, \sigma^2) = -\frac{n}{\sigma^2}$$

$$\frac{\partial^2}{\partial (\sigma^2)^2} \, log \, L(\theta, \sigma^2) = \frac{n}{2\,\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^{n} (x_i - \theta)^2$$

$$\frac{\partial^2}{\partial \theta \, \partial \sigma^2} \, log \, L(\theta, \sigma^2) = -\frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - \theta)$$

i veiem que en efecte els dos primers punts es verifiquen.

Veiem que també es compleix el tercer:

$$\boldsymbol{H} = \left| \begin{array}{cc} -n/\sigma^2 & -1/\sigma^4 \sum_{i=1}^{n} (x_i - \theta) \\ -1/\sigma^4 \sum_{i=1}^{n} (x_i - \theta) & n/2\,\sigma^4 - 1/\sigma^6 \sum_{i=1}^{n} (x_i - \theta)^2 \end{array} \right|_{\theta=\overline{x}, \sigma^2=\widehat{\sigma}^2} =$$

$$-\frac{n}{\sigma^2} \left(\frac{n}{2\,\sigma^4} - \frac{1}{\sigma^6} \sum_{i=1}^{n} (x_i - \theta)^2 \right) - \left(\frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - \theta) \right)^2 \Bigg|_{\theta=\overline{x}, \sigma^2=\widehat{\sigma}^2} =$$

$$\frac{1}{\widehat{\sigma}^6} \left[-\frac{n^2}{2} + \frac{n^2}{\widehat{\sigma}^2} \widehat{\sigma}^2 - \frac{1}{\widehat{\sigma}^2} \left(\sum_{i=1}^{n} (x_i - \overline{x}) \right)^2 \right] = \frac{n^2}{2\,\widehat{\sigma}^6} > 0$$

Exemple 2.14

Prenem una mostra aleatòria simple $X_1, \dots, X_n$ tal que segueixin la distribució d'una $X \sim \Gamma(\alpha, \lambda)$. Aleshores la seva funció de densitat és

$$f(x \mid \alpha, \lambda) = \frac{1}{\Gamma(\alpha)} \lambda^{\alpha} x^{\alpha-1} e^{-\lambda x} \quad \text{per } 0 \leq x < \infty$$

amb la qual cosa la funció de versemblança es pot escriure com

$$L(\theta) = L(\alpha, \lambda) = \left(\frac{1}{\Gamma^n(\alpha)} \right) \lambda^{n\,\alpha} \, (\prod_{i=1}^{n} x_i)^{\alpha-1} \, e^{-\lambda \sum_{i=1}^{n} x_i}$$

Anem a trobar el logaritme d'aquesta funció

$$l(\theta) = log\, L(\theta) = -n\, log\, \Gamma(\alpha) + n\, \alpha\, log\, \lambda + (\alpha - 1) \sum_{i=1}^{n} log\, x_i - \lambda \sum_{i=1}^{n} x_i$$

Si ara derivem aquesta darrera expressió respecte a cadascun dels dos paràmetres, α i λ, obtenim

$$\frac{\partial l}{\partial \alpha} = -n \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} + n\, log\, \lambda + \sum_{i=1}^{n} log\, x_i$$

$$\frac{\partial l}{\partial \lambda} = n\, \alpha \frac{1}{\lambda} - \sum_{i=1}^{n} x_i$$

Com que busquem els extrems la funció $l(\theta)$, anem a igualar aquestes últimes expressions trobades a 0:

$$n\, \alpha \frac{1}{\lambda} - \sum_{i=1}^{n} x_i = 0 \Longrightarrow \widehat{\lambda} = \frac{n\, \widehat{\alpha}}{\sum_{i=1}^{n} x_i} = \frac{\widehat{\alpha}}{\overline{x}_n}$$

i si ara substituïm aquest valor del paràmetre λ en l'altra de les igualtats obtenim la següent equació:

$$-n \frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} + n\, log\, \frac{\hat{\alpha}}{\overline{x}_n} + \sum_{i=1}^{n} log\, x_i = 0$$

la qual no admet una solució explícita, sinó que cal un mètode iteratiu per tal de trobar $\hat{\alpha}$.

Exemple 2.15

Càlcul de $\hat{\alpha}$ amb un mètode iteratiu (Newton-Raphson). Continuació de l'exemple anterior.

Les arrels d'una funció $f(\alpha) = 0$ es poden trobar mitjançant la iteració:

$$\hat{\alpha}_{n+1} = \hat{\alpha}_n + h_n \qquad h_n = -\frac{f(\hat{\alpha}_n)}{f'(\hat{\alpha}_n)}$$

En el nostre cas:

$$f(\hat{\alpha}) = -n\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} + n\ln\frac{\hat{\alpha}}{\overline{x}} + \sum_{i=1}^{n} x_i = -n\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})} + n\ln\hat{\alpha} - n\ln\overline{x} + \sum_{i=1}^{n} x_i$$

i

$$f'(\hat{\alpha}) = -n\frac{d}{d\hat{\alpha}}\left(\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})}\right) + \frac{n}{\hat{\alpha}},$$

on el quocient $\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})}$ és conegut com a la **funció digamma**, i la seva derivada, $\frac{d}{d\hat{\alpha}}\left(\frac{\Gamma'(\hat{\alpha})}{\Gamma(\hat{\alpha})}\right)$ és coneguda com a la **funció trigamma**.

Per començar la iteració, és necessari un valor inicial α_0. Podem posar un valor inicial arbitrari, per exemple $\alpha_0 = 1$, o bé utilitzar el valor de l'estimador obtingut per mètode de moments com a valor inicial.

Considerem una mostra de 10.000 observacions d'una distribució $\Gamma(\alpha = 2, \beta = 3)$. La mitjana d'aquesta mostra és 0.659269. Trobem el valor de l'estimador de màxima versemblança de manera iterativa en Taula 2.2:

i	α	$f(\alpha)$	$f'(\alpha)$	h
0	1.000000	3.102326e+03	-6449.341	4.810300e-01
1	1.481030	1.071487e+03	-2755.763	3.888170e-01
2	1.869847	2.364494e+02	-1672.638	1.413632e-01
3	2.011210	1.764700e+01	-1432.216	1.232146e-02
4	2.023532	1.141968e-01	-1413.740	8.077639e-05
5	2.023612	4.844877e-06	-1413.620	3.427285e-09
6	2.023612	-1.818989e-12	-1413.620	-1.286760e-15

Taula 2.2: Mètode iteratiu per obtenir el valor de l'estimador de màxima versemblança

i trobem $\hat{\alpha} = 2.023612$. Fent servir la relació $\hat{\lambda} = \frac{\hat{\alpha}}{\overline{x}}$ trobem $\hat{\lambda} = \frac{2.023612}{0.659269} = 3.069479$.

El valor de l'estimador obtingut per mètode de moments és

$$\hat{\alpha}_{MM} = \frac{\overline{x}^2}{\frac{1}{n}(\sum_{i=1}^{n} y_i^2) - \overline{x}^2} = 2.045233$$

i forma un millor punt inicial, a partir del qual la convergència a l'òptim és més ràpida (Taula 2.3):

i	α	$f(\alpha)$	$f'(\alpha)$	h
0	2.045233	-3.022058e+01	-1382.049	-2.186650e-02
1	2.023367	3.471806e-01	-1413.984	2.455335e-04
2	2.023612	4.477239e-05	-1413.620	3.167216e-08
3	2.023612	-1.818989e-12	-1413.620	-1.286760e-15

Taula 2.3: Mètode iteratiu per obtenir el valor de l'estimador de màxima versemblança

2.1.3 Propietat d'invariància del màxim de versemblança

Sigui $X_1, \ldots, X_n$ mostra aleatòria simple de $X \sim f(x \mid \theta)$ i sigui $\widehat{\theta}$ l'estimador de la màxima versemblança de θ. Si estem interessats en estimar una funció $\tau(\theta)$ del paràmetre, ho podem fer mitjançant $\tau(\widehat{\theta})$. Aquest és el resultat que garanteix el següent teorema i que es coneix com el principi d'invariància.

Teorema 2.2 principi d'invariància

Si $\widehat{\theta}$ és l'estimador de la màxima versemblança de θ, llavors per a tota funció $\tau(\theta)$ l'estimador de la màxima versemblança de $\tau(\theta)$ és $\tau(\widehat{\theta})$.

I ja per acabar aquesta part del tema, anem a donar un exemple d'aplicació d'aquest teorema que acabem d'enunciar.

Exemple 2.16

Sigui $X \sim N(\mu, \sigma^2)$. Hem trobat els estimadors MLE de (μ, σ^2), que són, respectivament,

$$\widehat{\mu} = \overline{X}_n \quad \text{i} \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2$$

Ara ens preguntem quin serà l'estimador de la màxima versemblança de σ. Doncs bé, com

$$\sigma = \sqrt{\sigma^2} = g(\sigma^2) \quad \text{essent g bijectiva}$$

tenim que l'estimador MLE de σ serà, pel teorema anterior,

$$\widehat{\sigma} = \sqrt{\widehat{\sigma}^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2}$$

Ara ens fem la pregunta de quin és l'estimador de la màxima versemblança de $E(X^2)$.

Com que

$$E(X^2) = \mu^2 + \sigma^2 = g(\mu, \sigma^2), \quad \text{essent g bijectiva}$$

l'estimador MLE de $E(X^2)$ és

$$\widehat{\mu}^2 + \widehat{\sigma}^2 = \overline{X}_n^2 + \frac{1}{n} \sum_{i=1}^{n} \left(X_i - \overline{X}_n \right)^2$$

Si $\tau(\theta_1, \ldots, \theta_k)$ és una funció de $\boldsymbol{\theta}$, $\tau(\widehat{\theta}_1, \ldots, \widehat{\theta}_k)$ és el MLE de $\tau(\boldsymbol{\theta})$. És a dir, la propietat d'invariància també se satisfà en el cas multivariànciat.

2.1.4 El mètode de Bayes

Distribucions a priori i a posteriori

Definició 2.9 *En un problema d'inferència estadística en el que les observacions es prenen d'una variable aleatòria X que segueix una distribució amb funció de densitat $f(x \mid \theta)$ on $\theta \in \Theta$ de vegades es disposa d'informació sobre el paràmetre θ abans de recollir les dades.*

És a dir, en algunes ocasions disposem de prou informació, ja sigui històrica (si s'han realitzat experiments similars amb anterioritat) o subjectiva (l'investigador pot creure que certs valors de θ són més plausibles que altres) que ens permeten assignar una distribució de probabilitat per θ. Aquesta llei s'anomena **llei a priori** *o distribució a priori per θ, perquè representa la versemblança que el veritable valor de θ sigui a les diferents regions de l'espai de paràmetres Θ abans d'haver observat cap valor de $f(x \mid \theta)$.*

El concepte de distribució a priori és força controvertit. Alguns estadístics creuen que en qualsevol problema estadístic es pot escollir una llei a priori per θ. Aquesta llei representa la informació subjectiva de l'experimentador i s'ha de treballar amb ella seguint les regles bàsiques de la probabilitat. Aquests són els estadístics que s'adhereixen plenament a la **filosofia bayesiana**.

Altres estadístics creuen que en certes ocasions no és apropiat el parlar d'una distribució de probabilitat sobre Θ, ja que θ és una quantitat fixa desconeguda per l'observador.

Tot i així tothom està d'acord, però, que quan hi ha informació extensa sobre les freqüències relatives de cada possible valor de θ és raonable definir la distribució a priori sobre θ basant-se en la llei derivada a partir de les freqüències relatives.

Denotem la llei a priori de θ per $\pi(\theta)$.

Prenem una mostra $X_1, ..., X_n$ d'una distribució $f(x \mid \theta)$. L'objectiu és actualitzar la llei a priori $\pi(\theta)$ amb l'ajut de $\boldsymbol{x}$ i $f(x \mid \theta)$. Suposem que el valor de θ és desconegut i que $\Theta = \mathbb{R}$ o $\Theta = [a, b]$.

La llei conjunta de $X_1, ..., X_n$ s'anomena distribució mostral i ve donada per $f(\boldsymbol{x} \mid \theta) = f(x_1 \mid \theta) \cdot \ldots \cdot f(x_n \mid \theta)$, i la llei conjunta de $\boldsymbol{x}$ i θ és $f(\boldsymbol{x}, \theta) = f(\boldsymbol{x} \mid \theta)\, \pi(\theta)$. La llei marginal de $\boldsymbol{x}$ és aleshores

$$m(\boldsymbol{x}) = \int_\Theta f(\boldsymbol{x} \mid \theta)\, \pi(\theta)\, d\theta$$

i podem obtenir la distribució a posteriori com a

$$\pi(\theta|\mathbf{x}) = \frac{f(\boldsymbol{x},\theta)}{m(\boldsymbol{x})} = \frac{f(\boldsymbol{x}|\theta)\pi(\theta)}{m(\boldsymbol{x})}$$

Definició 2.10 *Definim la* **llei a posteriori** *de θ com la llei condicional de θ donades les observacions $\boldsymbol{x}$ de $\boldsymbol{x}$, és a dir*

$$\pi(\theta\,|\,\boldsymbol{x}) = \frac{1}{m(\boldsymbol{x})}\, f(\boldsymbol{x}\,|\,\theta)\,\pi(\theta), \qquad \theta \in \Theta$$

i en inferència bayesiana s'utilitza aquesta llei per inferir sobre θ. Per exemple, un estimador puntual de θ seria $E(\theta\,|\,\boldsymbol{x})$.

La llei a posteriori ens informa sobre la versemblança relativa que el veritable valor de θ sigui a les diferents regions de l'espai de paràmetres Θ després d'haver observat $X_1, ..., X_n$. Fixem-nos que $\pi(\theta\,|\,\boldsymbol{x})$ és proporcional a la funció de versemblança i a la llei a priori de θ:

$$\pi(\theta\,|\,\boldsymbol{x}) \propto f(\boldsymbol{x}\,|\,\theta)\pi(\theta)$$

Observacions seqüencials

En molts experiments, i en particular en molts assajos clínics, les observacions $X_1, ..., X_n$ s'obtenen de forma seqüencial. Podem aleshores usar aquesta seqüencialització per anar actualitzant de forma progressiva la informació que es té sobre el paràmetre θ.

És a dir, disposem de la llei a priori $\pi(\theta)$. Observem $X_1 = x_1$. Aleshores

$$\pi(\theta \,|\, x_1) \propto f(x_1 \,|\, \theta)\, \pi(\theta)$$

recull a partir d'aquest instant la informació acumulada sobre θ. Així doncs, la nostra informació sobre θ queda representada per $\pi(\theta \,|\, x_1)$ i per tant aquesta actua a priori.

Quan s'observa $X_2 = x_2$,

$$\pi(\theta \,|\, x_1, x_2) \propto f(x_2 \,|\, \theta)\, \pi(\theta \,|\, x_1) \propto f(x_2 \,|\, \theta)\, f(x_1 \,|\, \theta)\, \pi(\theta)$$

I després d'haver observat $X_3 = x_3, \ldots, X_n = x_n$,

$$\begin{array}{rcl} \pi(\theta \,|\, x_1, \ldots, x_n) & \propto & f(x_n \,|\, \theta)\, \pi(\theta \,|\, x_1, \ldots, x_n) \\ & \propto & f(x_n \,|\, \theta)\, f(x_{n-1} \,|\, \theta) \cdot \ldots \cdot f(x_1 \,|\, \theta)\, \pi(\theta) \\ & = & f(\boldsymbol{x} \,|\, \theta)\, \pi(\theta) \end{array}$$

S'observa així, que la distribució a posteriori després d'haver pres n observacions de X és la mateixa tant si aquestes es prenen de forma seqüencial com si es prenen simultàniament.

Exemple 2.17 ______________________________

S'ha observat que la durada en hores d'un fluorescent es pot modelar per mitjà d'una variable aleatòria X que segueix un model exponencial de mitjana desconeguda, és a dir, $X \sim exp(\theta)$ amb $E(X) = 1/\theta$. La informació històrica en què ens podem basar pel disseny d'aquest tipus de llums ens diu que θ té una mitjana aproximada de 500 hores. De fet, modelant θ s'ha comprovat que

$$\theta \sim \Gamma(\alpha_0, \beta_0),$$

amb $E(\theta) = 1/5000 = 0.0002$ i $\sqrt{Var(\theta)} = 0.0001$.

Com que $E(\theta) = \frac{\alpha_0}{\beta_0}$ i $Var(\theta) = \frac{\alpha_0}{{\beta_0}^2}$ deduïm que

$$\left.\begin{array}{rl} \alpha_0/\beta_0 = & 0.0002 \\ \sqrt{\alpha_0}/\beta_0 = & 0.0001 \end{array}\right\} \Longrightarrow \alpha_0 = 4 \;\text{ i }\; \beta_0 = 20.000$$

Així, hem obtingut que $\theta \sim \Gamma(4, 20.000)$, i la seva funció de densitat és

$$\pi(\theta) = \frac{20.000^4}{(4-1)!}\,\theta^3 e^{-20.000\cdot\theta}, \quad \theta > 0$$

Fem a continuació una prova de vida en la qual es posen a funcionar 30 fluorescents del nou disseny fins que es fonen. Els resultats són aquests:

$$X_1 = x_1, \ldots, X_{30} = x_{30}, \; \sum_{i=1}^{30} x_i = 200.000$$

D'aquesta manera, la versemblança és:

$$f(x_1, \ldots, x_{30} \,|\, \theta) = \theta^{30} e^{-\theta\cdot\sum_{i=1}^{30} x_i},$$

i la densitat a posteriori de θ és

$$\pi(\theta \,|\, x_1, \ldots, x_{30}) = \frac{f(x_1, \ldots, x_{30}, \theta)}{m(x_1, \ldots, x_{30})} = \frac{f(x_1, \ldots, x_{30} \,|\, \theta)\pi(\theta)}{\int_0^\infty f(x_1, \ldots, x_{30} \,|\, \theta)\pi(\theta) d\theta}$$

El numerador és

$$f(x_1, \ldots, x_{30} \,|\, \theta)\pi(\theta) = \theta^{30} \;\; e^{-\theta\sum_{i=1}^{30} x_i} \;\; \frac{20.000^4}{3!} \;\; \theta^3 \;\; e^{-20.000\theta} =$$

$$\frac{20.000^4}{6} \;\; \theta^{33} \;\; e^{-\theta(\sum_{i=1}^{30} x_i + 20.000)}$$

i el denominador

$$m(x_1, \ldots, x_n) = \int_0^\infty \frac{20000^4}{6} \;\; \theta^{33} \;\; e^{-\theta(\sum_{i=1}^{30} x_i + 20000)} d\theta =$$

$$\int_0^{\infty} \frac{20000^4}{6} \frac{(\sum x_i + 20000)^{34}}{(\sum x_i + 20000)^{34}} \frac{\Gamma(34)}{\Gamma(34)} \; \theta^{33} \; e^{-\theta(\sum_{i=1}^{30} x_i + 20000)} d\theta$$

Si definim $\alpha = 34$ i $\beta = \sum_{i=1}^{30} x_i + 20000$ podem observar que aquesta integral és de fet la integral de la funció de densitat d'una variable aleatòria $\Gamma(\alpha, \beta)$, concretament

$$m(x_1, \ldots, x_{30}) = \frac{20000^4}{6} \frac{33!}{(\sum_{i=1}^{30} x_i + 20000)^{34}} \int_0^{\infty} f_{\Gamma(\alpha,\beta)}(\theta) d\theta$$

Així, doncs, tenim el següent:

$$\pi(\theta \,|\, x_1, \ldots, x_{30}) = \frac{(\sum_{i=1}^{30} x_i + 20000)^{34}}{33!} \; \theta^{33} \; e^{-\theta\,(\sum_{i=1}^{30} x_i + 20000)}$$

d'on es dedueix:

$$\theta \,|\, x_1, \ldots, x_{30} \sim \Gamma\big(34, \sum_{i=1}^{30} x_i + 20000\big)$$

Per tant un estimador de θ podria ser:

$$\hat{\theta} = E(\theta \,|\, x_1, \ldots, x_n) = \frac{34}{\sum_{i=1}^{30} x_i + 20000}$$

amb la qual cosa

$$\hat{\mu} = \frac{1}{\hat{\theta}} = \frac{\sum_{i=1}^{30} x_i + 20000}{34} = \frac{200000 + 20000}{34} = 6471 \text{ hores}$$

En canvi, si féssim servir $\bar{x}$ com a estimador de μ obtindríem

$$\hat{\mu} = \bar{x}_n = \frac{200000}{30} = 6667 \text{ hores}$$

Per tant, la informació a priori ens diu que no hem de ser tan optimistes. Malgrat que el nostre disseny ha millorat el fluorescent, (la mitjana és superior a 5000 hores), la informació a priori ens rebaixa la nova mitjana de 6667 hores a 6471 hores, és a dir, en 196 hores menys.

Distribucions a priori conjugades

Definició 2.11 *Sigui $\mathcal{F}$ la classe de totes les funcions de densitat (o de probabilitat) indexades per θ.*

$$\mathcal{F} = \{\, f(x \,|\, \theta) \,:\, \theta \in \Theta \,\}$$

Sigui Π una classe de distribucions sobre Θ. Direm que Π és una **família de distribucions conjugada** *per $\mathcal{F}$ si la distribució a posteriori de θ donada per la mostra $\boldsymbol{x}$ pertany a la classe Π per a tota mostra $\boldsymbol{x} \in \mathcal{X}$, per a tota a priori $\pi \in \Pi$ i per a tota versemblança $f \in \mathcal{F}$.*

Teorema 2.3 *Sigui $X_1, \ldots, X_n$ una m.a.s. de $X \sim Bern(\theta)$, $\quad 0 < \theta < 1$. Sigui $\theta \sim Beta(\alpha, \beta)$, on $\alpha > 0$, $\beta > 0$. La llei a posteriori de θ donada per les observacions $X_1 = x_1, \ldots, X_n = x_n$ és una*

$$Beta\Big(\alpha + \sum_{i=1}^{n} x_i \,,\, \beta + n - \sum_{i=1}^{n} x_i\Big)$$

Recordem aquí que $U\,[\,0, 1\,] = Beta\,(1, 1)$.

Demostració:

$$\pi(\theta) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\ \theta^{\alpha - 1}\ (1 - \theta)^{\beta - 1}$$

$$f(x_i \,|\, \theta) = \theta^{x_i}\ (1 - \theta)^{1 - x_i} \Longrightarrow f(\boldsymbol{x} \,|\, \theta) = \theta^{\sum_{i=1}^{n} x_i}\ (1 - \theta)^{n - \sum_{i=1}^{n} x_i} \ = L(\theta \,|\, x_1, \ldots, x_n)$$

Així, tenim el següent:

$$f(\boldsymbol{x}, \theta) = \theta^{\sum_{i=1}^{n} x_i}\ (1 - \theta)^{n - \sum_{i=1}^{n} x_i}\ \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\ \theta^{\alpha - 1}\ (1 - \theta)^{\beta - 1}$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\ \theta^{\sum_{i=1}^n x_i+\alpha-1}\ (1-\theta)^{n-\sum_{i=1}^n x_i+\beta-1}$$

i

$$m(\boldsymbol{x}) = \int_0^1 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\ \theta^{\sum_{i=1}^n x_i+\alpha-1}\ (1-\theta)^{n-\sum_{i=1}^n x_i+\beta-1} d\theta$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\ \frac{\Gamma\big(\alpha+\sum_{i=1}^n x_i\big)\,\Gamma\big(n-\sum_{i=1}^n x_i+\beta\big)}{\Gamma(\alpha+\beta+n)}\times$$

$$\underbrace{\int_0^1 \frac{\Gamma(\alpha+\beta+n)}{\Gamma\big(\alpha+\sum_{i=1}^n x_i\big)\,\Gamma\big(n-\sum_{i=1}^n x_i+\beta\big)}\ \theta^{\sum_{i=1}^n x_i+\alpha-1}\ (1-\theta)^{n-\sum_{i=1}^n x_i+\beta-1} d\theta}_{=1}$$

i per tant tenim que

$$m(\boldsymbol{x}) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\ \frac{\Gamma\big(\alpha+\sum_{i=1}^n x_i\big)\ \Gamma\big(n-\sum_{i=1}^n x_i+\beta\big)}{\Gamma(\alpha+\beta+n)}$$

Calculem ara la densitat a posteriori

$$\pi(\theta\,|\,\boldsymbol{x}) = \frac{\Gamma(\alpha)\,\Gamma(\beta)}{\Gamma(\alpha+\beta)}\ \frac{\Gamma(\alpha+\beta+n)}{\Gamma\big(\alpha+\sum_{i=1}^n x_i\big)\ \Gamma\big(n-\sum_{i=1}^n x_i+\beta\big)}\times$$

$$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\,\Gamma(\beta)}\ \theta^{\sum_{i=1}^n x_i+\alpha-1}\ (1-\theta)^{n-\sum_{i=1}^n x_i+\beta-1} =$$

$$\frac{\Gamma(\alpha+\beta+n)}{\Gamma\big(\alpha+\sum_{i=1}^n x_i\big)\,\Gamma\big(n-\sum_{i=1}^n x_i+\beta\big)}\ \theta^{\sum_{i=1}^n x_i+\alpha-1}\ (1-\theta)^{n-\sum_{i=1}^n x_i+\beta-1}$$

i per tant $\theta\,|\,\boldsymbol{x}$ té distribució $Beta\big(\alpha+\sum_{i=1}^n x_i\,,\ \beta+n-\sum_{i=1}^n x_i\big)$

$\square$

Teorema 2.4 *Sigui $X_1,\ldots,X_n$ una m.a.s. de $X \sim Poisson(\theta)$, $\theta>0$. Sigui $\theta\sim\Gamma(\alpha,\beta)$ amb $\alpha>0$ i $\beta>0$. Aleshores la llei a posteriori de θ donades les observacions $X_1=x_1,\ldots,X_n=x_n$ és una*

$$\Gamma\left(\alpha+\sum_{i=1}^n x_i, \beta+n\right)$$

Demostració:

Per a certes constants K_1, K_2 i K_3 tenim que

$$\begin{aligned} \pi(\theta) &= K_1\,\theta^{\alpha-1}\,e^{-\beta\,\theta} \\ L(\theta\,|\,\boldsymbol{x}) &= K_2\,e^{-n\,\theta}\,\theta^{\sum_{i=1}^n x_i} \end{aligned}$$

amb la qual cosa

$$\pi(\theta\,|\,\boldsymbol{x}) = K_3\,\theta^{\alpha-1}\,e^{-\beta\,\theta}\,e^{-n\,\theta}\,\theta^{\sum_{i=1}^n x_i} = K_3\,\theta^{\alpha+\sum_{i=1}^n x_i-1}\,e^{-(\beta+n)\,\theta}$$

que precisament es correspon amb la densitat d'una

$$\Gamma\left(\alpha+\sum_{i=1}^n x_i, \beta+n\right)$$

$\square$

Teorema 2.5 *Sigui $X_1, \ldots, X_n$ una mostra aleatòria simple d'una $X \sim N(\theta, \sigma^2)$, on $-\infty < \theta < +\infty$ i $\sigma^2 > 0$ és conegut. Sigui $\theta \sim N(\mu, \nu^2)$. Aleshores la llei a posteriori de θ donada per les observacions $X_1 = x_1, \ldots, X_n = x_n$ és una $N(\mu_1, {\nu_1}^2)$ amb*

$$\mu_1 = \frac{\sigma^2\,\mu + n\,\nu^2\,\overline{x}}{\sigma^2+n\nu^2} = \frac{\sigma^2}{\sigma^2+n\,\nu^2}\;\mu + \frac{n\,\nu^2}{\sigma^2+n\,\nu^2}\;\overline{x}$$

$${\nu_1}^2 = \frac{\sigma^2\,\nu^2}{\sigma^2+n\,\nu^2}$$

Demostració:

La llei a priori té per funció de densitat

$$\pi(\theta) = \frac{1}{\sqrt{2\,\pi\,\nu^2}}\,e^{-\frac{1}{2\nu^2}\,(\theta-\mu)^2}$$

La funció de versemblança donada $\boldsymbol{x} = \boldsymbol{x}$ és

$$L(\theta\,|\,x_1, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n\,e^{-\frac{1}{2\sigma^2}\,\sum_{i=1}^n (x_i-\theta)^2}$$

La llei a posteriori té una densitat que és proporcional a

$$\pi(\theta \,|\, x_1, \ldots, x_n) = \frac{f(\boldsymbol{x} \,|\, \theta)\, \pi(\theta)}{m(\boldsymbol{x})} = k\, f(\boldsymbol{x} \,|\, \theta)\, \pi(\theta). \tag{1}$$

essent

$$m(\boldsymbol{x}) = \int_{-\infty}^{+\infty} f(\boldsymbol{x} \,|\, \theta)\, \pi(\theta)\, d\theta. \tag{2}$$

i $f(\boldsymbol{x} \,|\, \theta)$ la llei conjunta de $\boldsymbol{x}$ donat θ, on de vegades escrivim $L(\theta \,|\, \boldsymbol{x})$ per referir-nos a la versemblança.

Comencem treballant el producte $f(\boldsymbol{x} \,|\, \theta)\, \pi(\theta)$:

$$f(\boldsymbol{x} \,|\, \theta)\, \pi(\theta) = \left(\frac{1}{\sqrt{2\, \pi\, \sigma^2}} \right)^n e^{-\frac{1}{2\,\sigma^2}\, \sum_{i=1}^n (x_i - \theta)^2}\, \frac{1}{\sqrt{2\, \pi\, \nu^2}}\, e^{-\frac{1}{2\,\nu^2}\, (\theta - \mu)^2} =$$

$$\frac{1}{(\sqrt{2\, \pi \sigma^2})^n\, \sqrt{2\, \pi\, \nu^2}}\, exp \underbrace{\left\{ -\frac{1}{2\, \sigma^2}\, \sum_{i=1}^n (x_i - \theta)^2 - \frac{1}{2\, \nu^2}\, (\theta - \mu)^2 \right\}}_{=E}. \tag{3}$$

Treballem ara l'exponent E:

$$E = -\frac{1}{2\, \sigma^2\, \nu^2} \left[\sum_{i=1}^n (x_i - \theta)^2\, \nu^2 + (\theta - \mu)^2\, \sigma^2 \right] = \ldots =$$

$$-\frac{1}{2\, \sigma^2\, \nu^2} \left[(n\, \nu^2 + \sigma^2)\, \theta^2 - 2\, (n\, \overline{x}\, \nu^2 + \mu\, \sigma^2)\, \theta + \nu^2\, \sum_{i=1}^n x_i^2 + \mu^2\, \sigma^2 \right] =$$

$$-\frac{1}{2\, \sigma^2\, \nu^2}\, (n\, \nu^2 + \sigma^2) \left[\theta^2 - 2\, \frac{n\, \overline{x}\, \nu^2 + \mu\, \sigma^2}{n\, \nu^2 + \sigma^2}\, \theta + \left(\frac{n\, \overline{x}\, \nu^2 + \mu\, \sigma^2}{n\, \nu^2 + \sigma^2} \right)^2 \right]$$

$$-\frac{1}{2\, \sigma^2 \nu^2} \left(\nu^2\, \sum_{i=1}^n {x_i}^2\, + \mu^2 \sigma^2 \right) + \frac{1}{2\, \sigma^2\, \nu^2}\, (n\, \nu^2 + \sigma^2)\, \frac{(n\, \overline{x}\, \nu^2 + \mu\, \sigma^2)^2}{(n\, \nu^2 + \sigma^2)^2} =$$

$$-\frac{1}{2\, \nu_1^2}\, (\theta - \mu_1)^2 \underbrace{- \frac{1}{2\, \sigma^2\, \nu^2}\, \left(\nu^2\, \sum_{i=1}^n x_i^2 + \mu^2\, \sigma^2 \right) + \frac{(n\, \overline{x}\, \nu^2 + \mu\, \sigma^2)^2}{2\, \sigma^2 \nu^2\, (n\, \nu^2 + \sigma^2)}}_{=A} \tag{4}$$

Calculem ara $m(\boldsymbol{x})$ a partir de (2) + (3) + (4):

$$m(\boldsymbol{x}) = \int_{-\infty}^{+\infty} \frac{1}{(\sqrt{2\,\pi\,\sigma^2})^n\,\sqrt{2\,\pi\,\nu^2}}\, exp\Big\{ -\frac{1}{2\,\nu_1^2}\,(\theta-\mu_1)^2 \Big\}\, exp\{\, A \,\}\, d\theta$$

Com que A és independent de θ surt fora la integral, i obtenim el següent:

$$m(\boldsymbol{x}) = exp\{\, A \,\}\, \frac{\sqrt{\nu_1^2}}{(\sqrt{2\,\pi\,\sigma^2})^n\,\sqrt{\nu^2}}\, \underbrace{\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\,\pi\,\nu_1^2}}\, exp\left\{ -\frac{1}{2\,\nu_1^2}\,(\theta-\mu_1)^2 \right\} d\theta}_{=1 \;\; \text{ja que} \;\; \theta \;\; \sim N(\mu_1, \nu_1^2)}$$

i per tant

$$m(\boldsymbol{x}) = exp\{\, A \,\}\, \frac{\sqrt{\nu_1^2}}{(\sqrt{2\,\pi\,\sigma^2})^n\,\sqrt{\nu^2}} \tag{5}$$

Així, doncs, tornant a (1) tenim que:

$$\pi(\theta \,|\, x_1, \ldots, x_n) = \frac{f(\boldsymbol{x}\,|\,\theta)\,\pi(\theta)}{m(\boldsymbol{x})} = \frac{\frac{1}{(\sqrt{2\,\pi\,\sigma^2})^n}\,\frac{1}{\sqrt{2\,\pi\,\nu^2}}\, exp\big\{ -\frac{1}{2\,\nu_1^2}\,(\theta-\mu_1)^2 \big\}\, exp\{\, A \,\}}{exp\{\, A \,\}\, \frac{\sqrt{\nu_1^2}}{(\sqrt{2\,\pi\,\sigma^2})^n\,\sqrt{\nu^2}}}$$

$$= \frac{1}{\sqrt{2\,\pi\,\nu_1^2}}\, exp\Big\{ -\frac{1}{2\,\nu_1^2}\,(\theta-\mu_1)^2 \Big\}$$

corresponent a la funció de densitat d'una $N(\mu_1, \nu_1^2)$.

$\square$

Teorema 2.6 *Sigui $X_1, \ldots, X_n$ una mostra aleatòria simple de $X \sim \exp(\theta)$, $\theta = \big(E(X)\big)^{-1} > 0$. Sigui $\theta \sim \Gamma(\alpha, \beta)$, amb $\alpha > 0$ i $\beta > 0$. La llei a posteriori de θ donades les observacions $X_1 = x_1, \ldots, X_n = x_n$ és una*

$$\Gamma\left(\alpha + n, \beta + \sum_{i=1}^{n} x_i\right)$$

Funcions de pèrdua

El principal requisit que voldríem per un estimador del paràmetre θ és que estigués a prop del veritable valor de θ.

Un bon estimador $\widehat{\theta}$ serà aquell pel qual sigui molt probable que l'error $\widehat{\theta} - \theta$ es faci 0.

Definició 2.12 *Suposem que per a cada valor possible $\theta \in \Theta$ i per a cada estimador possible $a \in \Theta$ definim una* **funció de cost o pèrdua**

$$\begin{array}{rccc} L: & \Theta \times \Theta & \longrightarrow & \mathbb{R}^+ \\ & (\theta, a) & \longrightarrow & L(\theta, a) \end{array}$$

on $L(\theta, a)$ és el cost per l'estadístic d'estimar el veritable valor θ mitjançant a.

Donat que θ és desconegut i pot prendre els valors de Θ segons indica la distribució π, per tal de tenir una idea global del cost de a es considera la funció de pèrdua esperada:

$$E\{\, L(\theta, a) \,\} = \int_\theta L(\theta, a)\, \pi(\theta) d\theta$$

En un problema d'inferència estadística una funció L per la qual $E\{\, L(\theta, a) \,\}$ es minimitzi s'anomena una **funció de pèrdua**.

L'elecció de la funció de pèrdua és molt sovint arbitrària. Les més usuals són les següents:

Funció de pèrdua de l'error absolut

$$L(\theta, \widehat{\theta}) = |\, \widehat{\theta} - \theta \,|$$

Funció de pèrdua quadràtica

$$L(\theta, \widehat{\theta}) = (\widehat{\theta} - \theta)^2$$

Estimadors de Bayes

Suposem que tenim informació a priori $\pi(\theta)$ sobre $\theta \in \Theta$, i que observem una mostra $X_1, \ldots, X_n$ de $X \sim f(x \,|\, \theta)$. La llei a posteriori ve donada per $\pi(\theta \,|\, x) \propto L(\theta \,|\, x)\, \pi(\theta)$. Suposem que el cost d'estimar θ mitjançant a ve modelat per $C(\theta, a)$. Aleshores podem calcular el cost (o pèrdua) esperat d'estimar θ mitjançant a després d'observar $\boldsymbol{x}$, i aquest serà:

$$E\{C(\theta, a) \,|\, \boldsymbol{x}\} = \int_\Theta C(\theta, a)\, \pi(\theta \,|\, x)\, d\theta$$

A aquesta quantitat se l'anomena **pèrdua a posteriori**.

Definició 2.13 *Un* **estimador de Bayes** $\theta^*(\boldsymbol{x})$ *de* θ *es defineix com aquell valor de* $a \in \Theta$ *que minimitza la pèrdua a posteriori, és a dir,*

$$E\big\{\, C\big(\theta, \theta^*(\boldsymbol{x})\big) \,|\, \boldsymbol{x} \,\big\} = min_{a \in \Theta} E\{\, C(\theta, a) \,|\, \boldsymbol{x} \,\}$$

Teorema 2.7 *Si la funció de pèrdua és quadràtica, és a dir, si* $L(\theta, a) = (\theta - a)^2$*, aleshores l'estimador de Bayes és l'esperança a posteriori:*

$$\theta^*(\boldsymbol{x}) = E(\theta \,|\, \boldsymbol{x})$$

Demostració:

En efecte,

$$E\big\{\, L(\theta, a) \,|\, \boldsymbol{x} \,\big\} = E\big\{\, (\theta - a)^2 \,|\, \boldsymbol{x} \,\big\} = E\big\{\, \theta^2 \,|\, \boldsymbol{x} \,\big\} - 2\, a\, E\big\{\, \theta \,|\, \boldsymbol{x} \,\big\} + a^2$$

Derivant respecte a a i igualant a zero:

$$-2\, E\big\{\, \theta \,|\, \boldsymbol{x} \,\big\} + 2\, a = 0 \Longrightarrow a = E(\theta \,|\, \boldsymbol{x}),$$

que és un mínim, ja que la segona derivada és $2 > 0$.

□

Teorema 2.8 *Si la funció de pèrdua és l'error absolut, és a dir, si* $L(\theta, a) = |\,\theta - a\,|$*, aleshores l'estimador de Bayes és la mediana (Me) de la distribució a priori:*

$$\theta^*(\boldsymbol{x}) = Me(\theta \,|\, \boldsymbol{x})$$

Demostració: Veure (**?**)

Exemple 2.18 ____

Sigui $X \sim Bern(\theta)$. S'observa una mostra aleatòria simple de X: $X_1 = x_1, \ldots, X_n = x_n$. Suposem que la distribució a priori de θ és

$$\theta \sim Beta(\alpha, \beta) \Longrightarrow E(\theta) = \frac{\alpha}{\alpha + \beta}$$

Tal com hem vist al teorema 2.1,

$$\theta \,|\, x_1, \ldots, x_n \sim Beta\left(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i\right)$$

Per tant, l'estimador de Bayes basat en la funció de pèrdua quadràtica serà:

$$\widehat{\theta}_1(\boldsymbol{x}) = E(\theta \,|\, x_1, \ldots, x_n) = \frac{\alpha + \sum_{i=1}^{n} x_i}{\alpha + \beta + n} = \frac{n}{\alpha + \beta + n}\,\overline{x} + \frac{\alpha + \beta}{\alpha + \beta + n}\,\frac{\alpha}{\alpha + \beta}$$

Observem que $\widehat{\theta}_1(\boldsymbol{x})$ és una combinació lineal entre la mitjana a priori $\alpha/\alpha + \beta$ i la mitjana mostral $\overline{X}_n$. Fixem-nos a més a més que a mesura que la grandària de la mostra creix $(n \longrightarrow \infty)$, tenim que $\widehat{\theta}_1(\boldsymbol{x}) \underset{n\to\infty}{\to} \overline{x}$.

Si volguéssim basar-nos en la pèrdua de l'error absolut, hauríem de determinar numèricament la mitjana d'una $Beta\big(\alpha + \sum_{i=1}^{n} x_i, \beta + n - \sum_{i=1}^{n} x_i\big)$.

Exemple 2.19

S'ha observat una m.a.s. de $X \sim N(\theta, \sigma^2) : X_1 = x_1, \ldots, X_n = x_n$. Partim d'una a priori $N(\mu, \nu^2)$ per θ. Al teorema 2.3 hem vist que la distribució a posteriori de $\theta \,|\, \boldsymbol{x}$ és $N(\mu_1, {\nu^2}_1)$, amb

$$\mu_1 = \frac{\sigma^2\,\mu + n\,\nu^2\,\overline{x}_n}{\sigma^2 + n\,\nu^2}, \qquad {\nu^2}_1 = \frac{\sigma^2\,\nu^2}{\sigma^2 + n\,\nu^2}.$$

L'estimador de Bayes suposant una pèrdua quadràtica és

$$\widehat{\theta}_2(\boldsymbol{x}) = E(\theta \,|\, \boldsymbol{x}) = \mu_1$$

Per tant

$$\widehat{\theta}_2(\boldsymbol{x}) = \frac{n\,\nu^2\,\overline{x}_n}{\sigma^2 + n\,\nu^2} + \frac{\sigma^2}{\sigma^2 + n\,\nu^2}\,\mu.$$

Altra vegada l'estimador de Bayes és una combinació lineal de la mitjana a priori i de la mitjana mostral. Si per exemple posem una a priori "poc informativa", és a dir, amb una variància mostral ν^2 molt gran, trobarem que $\widehat{\theta}_2(\boldsymbol{x}) - \overline{x}_n \underset{n\to\infty}{\to} 0$, és a dir, quan la informació a priori es fa menys precisa li donem més pes a la informació mostral.

En aquest cas l'elecció de la funció de pèrdua de l'error absolut condueix a $\widehat{\theta}_2(\boldsymbol{x})$ perquè la mediana és igual a la mitjana en un model normal.

2.2 Mètodes per avaluar els estimadors

Un cop hem mostrat diferents mètodes per trobar estimadors, la següent qüestió és desenvolupar criteris per avaluar-los i estudiar alguns dels estimadors que hem desenvolupat sota el marc d'aquests criteris. En aquesta secció estudiarem mesures de la qualitat d'un estimador, primer per a mostres finites i després en proposarem algunes d'asimptòtiques.

2.2.1 Error quadràtic mitjà

Recordem que hem definit l'error quadràtic mitjà (MSE) d'un estimador W d'un paràmetre θ com:

$$E_\theta(W-\theta)^2$$

Aquesta és una mesura intuïtiva del comportament d'un estimador: quant més petit sigui, millor serà l'estadístic W. De fet, per a qualsevol funcio ϕ creixent, $E_\theta(\phi\,|\,W-\theta\,|\,)$ ens pot servir per tal de mesurar el comportament de W, i per veure com d' allunyades estaran, en mitjana, les estimacions de θ que proporciona W.

L'error quadràtic mitjà es prefereix en general a altres mesures -basades per exemple en el valor absolut- perquè analíticament és més tractable. A més, com ja vam veure al tema anterior, el MSE es pot interpretar com:

$$\begin{aligned} E_\theta(W-\theta)^2 &= E_\theta\big(W-E_\theta(W)\big)^2+\big(E_\theta(W)-\theta\big)^2 \\ &= Var_\theta(W)+\big(B_\theta(W)\big)^2 \end{aligned}$$

Definició 2.14 *Es defineix el* **biaix** *d'un estimador puntual* W *respecte d'un paràmetre* θ *per:*

$$B_\theta(W)=E_\theta(W)-\theta$$

Els estimadors tals que $B_\theta(W)=0$ *per a tot* θ *s'anomenen* **estimadors no esbiaixats**.

Així, doncs, l'error quadràtic mitjà d'un estimador és la suma de la seva variància (una mesura de la seva dispersió) més el quadrat del seu biaix (mesura de l'exactitud de l'estimador). És per tant una mesura conjunta de precisió i exactitud de l' estimador. És per aquest motiu que ens agradarà trobar estimadors que tinguin error quadràtic mitjà petit, perquè d'aquesta manera controlarem tant la dispersió com l'exactitud de les estimacions.

Notem que si W és un estimador sense biaix, per la descomposició anterior, tenim que:

$$E_\theta(W-\theta)^2 = Var_\theta(W)$$

Exemple 2.20

Sigui $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma^2)$, amb tots dos paràmetres desconeguts: $\theta = (\mu, \sigma^2)$. Hem estudiat, entre altres, els següents estimadors per a μ i σ^2:

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \;\; S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2$$

També hem calculat les seves esperances:

$$E(\overline{X}_n) = \mu, \;\; E(S^2) = \sigma^2$$

El MSE d'aquests estimadors coincideix amb la variància:

$$E_\theta(\overline{X}_n - \mu)^2 = Var_\theta(\overline{X}_n) = \sigma^2/n$$
$$E_\theta(S^2 - \sigma^2)^2 = Var_\theta(S^2) = 2\,\sigma^4/n-1$$

Podem concloure, doncs, que $\overline{X}_n$ i S^2 són estimadors sense biaix de μ i σ^2, respectivament.

Exemple 2.21

De vegades val la pena permetre una mica de biaix en un estimador per tal d'obtenir una reducció important en la variància i per tant en el MSE. El MLE és un d'aquests casos:

Suposem que ens trobem en les mateixes condicions que a l'exemple anterior. Sigui
$$\widehat{\sigma}^2_{MLE} = \frac{n-1}{n}\, S^2$$
l'estimador de la màxima versemblança per σ^2. Calculem el seu error quadràtic mitjà. Per una banda, tenim que:
$$E(\widehat{\sigma}^2) = \frac{n-1}{n}\,\sigma^2 = \left(1-\frac{1}{n}\right)\sigma^2 \longrightarrow B_\theta(\widehat{\sigma}^2) = -\frac{1}{n}\,\sigma^2;$$
i per l'altra,
$$Var_\theta(\widehat{\sigma}^2) = \left(\frac{n-1}{n}\right)^2 Var(S^2) = \frac{2\,(n-1)}{n^2}\,\sigma^4$$
D'aquesta manera, l'error quadràtic mitjà és:
$$E_\theta(\widehat{\sigma}^2-\sigma^2)^2 = \frac{2(n-1)}{n^2}\,\sigma^4 + \frac{1}{n^2}\,\sigma^4 = \frac{2n-1}{n^2}\,\sigma^4 \; < \; \frac{2}{n-1}\,\sigma^4$$
ja que $n \geq 2$.

Per tant, l'estimador $\widehat{\sigma}^2$ té MSE més petit que S^2.

2.2.2 Millor estimador sense biaix. La informació de Fisher i la fita de Cramer-Rao

Hem discutit fins ara que les comparacions basades en el criteri del MSE poden portar a conclusions poc favorables. Per exemple, no podem desestimar l'estimador constant $W = 3$, ja que aquest és el millor en el cas que el paràmetre $\mu = 3$.

Si W_1 i W_2 són dos estimadors sense biaix de θ, és a dir, $E_\theta(W_1) = E_\theta(W_2) = \theta$, aleshores l'error quadràtic mitjà coincidirà amb la variància. Es tracta de quedar-se, doncs, amb l'estimador de variància més petita. L'objectiu serà, a més, caracteritzar el millor estimador sense biaix, és a dir, aquell que tingui la variància més petita.

Definició 2.15 *Un estimador W^* és el* **millor estimador sense biaix** *de $\tau(\theta)$ o* **UMVUE** *(també es coneix com estimador no esbiaixat uniformement de mínima variància) si*

$$E_\theta(W^*) = \tau(\theta) \qquad \forall\ \theta$$

i per qualsevol altre estimador W amb $E_\theta(W) = \tau(\theta)$, tenim que

$$Var_\theta(W^*) \leq Var_\theta(W) \qquad \forall\ \theta$$

Exemple 2.22 ____

Sigui $X_1, \dots, X_n$ una m.a.s. de $X \sim Poisson(\lambda)$. Ja sabem que $E(X_i) = Var(X_i) = \lambda$. Per tant, $\overline{X}_n$ i S^2 són ambdós estimadors no esbiaixats de λ. En aquest exemple el que farem serà determinar quin dels dos és millor, en el sentit de tenir mínima variància.

La variància de $\overline{X}_n$ és:

$$Var(\overline{X}_n) = \frac{Var(X)}{n} = \frac{\lambda}{n}$$

Calculem ara $Var(S^2)$. Es tracta d'un càlcul més pesat que es pot alleugerir si es té en compte el resultat següent:

En general, si posem $\theta_1 = E(X_i)$ i $\theta_j = E\big((X_i - \theta_1)^j\big)$, per a $j = 2, 3, 4$, aleshores es pot veure que

$$Var(S^2) = \frac{1}{n}\left(\theta_4 - \frac{n-3}{n-1}\,\theta_2^2\right)$$

Anem a concretar aquest resultat pel cas de la distribució de Poisson. En aquest cas, tenim que $\theta_1 = \theta_2 = \lambda$. Calcularem tot seguit θ_3 i θ_4.

$$E(X^3) = \sum_{k=0}^{\infty} k^3\, e^{-\lambda}\, \frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k^2\, e^{-\lambda}\, \frac{\lambda^k}{(k-1)!} = \sum_{z=0}^{\infty} (z+1)^2\, e^{-\lambda}\, \frac{\lambda^{z+1}}{z!} =$$

$$\lambda \sum_{z=0}^{\infty} (z^2 + 2\,z + 1)\, e^{-\lambda}\, \frac{\lambda^z}{z!} = \lambda(E(X^2) + 2\,E(X) + 1) =$$

$$\lambda(e^{-\lambda} + \lambda + \lambda^2 + 2\,\lambda + 1 - e^{-\lambda}) = \lambda^3 + 3\,\lambda^2 + \lambda$$

Així, tenim que:

$$\begin{aligned} \theta_3 &= E\big((X-\lambda)^3\big) = E(X^3 - 3\,X^2\,\lambda + 3\,X\,\lambda^2 - \lambda^3) \\ &= E(X^3) - 3\,\lambda\,E(X^2) + 3\,\lambda^2\,E(X) - \lambda^3 \\ &= \lambda^3 + 3\,\lambda^2 + \lambda - 3\,\lambda\,(\lambda + \lambda^2) + 3\,\lambda^3 - \lambda^3 = \lambda \end{aligned}$$

Calculem ara θ_4

$$E(X^4) = \sum_{k=0}^{\infty} k^4\,e^{-\lambda}\,\frac{\lambda^k}{k!} = \sum_{k=1}^{\infty} k^3\,e^{-\lambda}\,\frac{\lambda^k}{(k-1)!} = \sum_{z=0}^{\infty} (z+1)^3\,e^{-\lambda}\,\frac{\lambda^{z+1}}{z!}$$

$$= \lambda\left[\sum_{z=0}^{\infty} z^3\,e^{-\lambda}\,\frac{\lambda^z}{z!} + \sum_{z=0}^{\infty} 3\,z^2\,e^{-\lambda}\,\frac{\lambda^z}{z!} + \sum_{z=0}^{\infty} 3\,z\,e^{-\lambda}\,\frac{\lambda^z}{z!} + \sum_{z=0}^{\infty} e^{-\lambda}\,\frac{\lambda^z}{z!}\right]$$

$$= \lambda(\lambda^3 + 3\,\lambda^2 + \lambda + 3\,(\lambda + \lambda^2) + 3\,\lambda + 1) = \lambda^4 + 6\,\lambda^3 + 7\,\lambda^2 + \lambda$$

Així, obtenim:

$$\begin{aligned} \theta_4 &= E\big((X-\lambda)^4\big) = E(X^4) - 4\,\lambda\,E(X^3) + 6\,\lambda^2\,E(X^2) - 4\,\lambda^3\,E(X) + \lambda^4 \\ &= \lambda^4 + 6\,\lambda^3 + 7\,\lambda^2 + \lambda - 4\,\lambda\,(\lambda^3 + 3\,\lambda^2 + \lambda) + 6\,\lambda^2\,(\lambda + \lambda^2) - 3\,\lambda^4 \\ &= 3\,\lambda^2 + \lambda \end{aligned}$$

Per tant,

$$Var(S^2) = \frac{1}{n}\,\big(\theta_4 - \frac{n-3}{n-1}\,\theta_2^2\big) = \frac{1}{n}\,\big(3\,\lambda^2 + \lambda - \frac{n-3}{n-1}\,\lambda^2\big)$$

$$= \frac{\lambda}{n} + \frac{2\,\lambda^2}{n-1} \;>\; \frac{\lambda}{n} = Var(\overline{X}_n) \qquad \text{si} \quad n \geq 2$$

Podem concloure, doncs, que S^2 no pot ser el UMVUE per a λ, ja que $\overline{X}_n$ és preferible a S^2.

Si considerem la classe dels estimadors $W_a(\overline{X}, S^2) = a\,\overline{X} + (1-a)\,S^2$, tenim que, $\forall\ a$, $E_\lambda\big(W_a(\overline{X}, S^2)\big) = a\,\lambda + (1-a)\,\lambda = \lambda$, i per tant $W_a(\overline{X}, S^2)$ són tots estimadors no esbiaixats. Podem afirmar que $\overline{X}$ és millor que qualsevol altre estimador del tipus $W_a(\overline{X}, S^2)$? Haurem de calcular la variància en tots aquests casos?

Doncs bé, aquest exemple posa de manifest els problemes que hem trobat en la cerca del millor estimador sense biaix, i apunta cap a un enfocament diferent que no consisteixi a repassar tots els estimadors possibles.

Definició 2.16 *S'anomena* **informació de Fisher** *que sobre* θ *conté la variable* X *a*

$$I_X(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\,log\,f_X(X\,|\,\theta)\right)^2\right]$$

Exemple 2.23 ______________________________

Sigui $X \sim exp\,(1/\theta)$. Aleshores sabem que la seva funció de densitat ve donada per la següent expressió:

$$f(x\,|\,\theta) = \frac{1}{\theta}\,e^{-x/\theta}$$

Per tant

$$\begin{aligned}
log f(x\,|\,\theta) &= -log\theta - \frac{x}{\theta} \\
\frac{\partial}{\partial\theta}\,log f(x\,|\,\theta) &= -\frac{1}{\theta} + \frac{x}{\theta^2}, \\
\big(\frac{\partial}{\partial\theta}\,log f(x\,|\,\theta)\big)^2 &= \frac{1}{\theta^2} + \frac{x^2}{\theta^4} - 2\,\frac{x}{\theta^3}
\end{aligned}$$

Així, doncs, la informació de Fisher serà en aquest cas:

$$\begin{aligned}
I_X(\theta) &= E\Big[\big(\frac{\partial}{\partial\theta}\,log f(X\,|\,\theta)\big)^2\Big] \\
&= \frac{1}{\theta^2} + \frac{E(X^2)}{\theta^4} - \frac{2\,E(X)}{\theta^3}
\end{aligned}$$

Anem a escriure-la d'una altra manera. Sabem que

$$\begin{aligned} E(X) &= \theta \\ Var(X) &= \theta^2 \\ E(X^2) = Var(X) + \big(E(X)\big)^2 &= \theta^2 + \theta^2 = 2\,\theta^2 \end{aligned}$$

I amb això podem escriure el següent:

$$\begin{aligned} I_X(\theta) &= \frac{1}{\theta^2} + \frac{2\,\theta^2}{\theta^4} - \frac{2\,\theta}{\theta^3} = \frac{1}{\theta^2} + \frac{2}{\theta^2} - \frac{2}{\theta^2} \\ &= \frac{1}{\theta^2} = \frac{1}{Var(X)} \end{aligned}$$

És a dir, hem obtingut que la informació de Fisher és l'invers de la variància.

Anem tot seguit a generalitzar la definició anterior però pel cas d'una mostra aleatòria simple.

Definició 2.17 *Sigui $X_1, \ldots, X_n$ una m.a.s. amb funció de densitat conjunta*

$$f(\boldsymbol{x}\,|\,\theta) = \prod_{i=1}^{n} f(x_i\,|\,\theta)$$

Aleshores definim la **informació de Fisher** *que sobre θ conté el vector $\boldsymbol{x}$ com*

$$I_{\boldsymbol{x}}(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\, log f(\boldsymbol{x}\,|\,\theta)\right)^2\right]$$

Anem ara a fer un exemple on calcularem la informació de Fisher però pel cas d'una mostra d'una exponencial.

Exemple 2.24

Sigui $X_1, \ldots, X_n$ una m.a.s. d'una $X \sim exp\,(\theta)$. Sabem que

$$f(x_1, \ldots, x_n \,|\, \theta) = \prod_{i=1}^{n} \Big\{ \frac{1}{\theta}\, e^{-x_i/\theta} \Big\} = \frac{1}{\theta^n}\, e^{-\sum_{i=1}^{n} x_i/\theta}$$

Així, fent logaritmes, obtenim:

$$\ln f(x_1, \ldots, x_n \,|\, \theta) = -n\, \ln(\theta) - \frac{1}{\theta} \sum_{i=1}^{n} x_i$$

$$\frac{\partial}{\partial\theta} \ln f(x_1, \ldots, x_n \,|\, \theta) = -\frac{n}{\theta} + \frac{\sum_{i=1}^{n} x_i}{\theta^2}$$

$$\Big(\frac{\partial}{\partial\theta} \ln f(x_1, \ldots, x_n \,|\, \theta)\Big)^2 = \frac{n^2}{\theta^2} + \frac{\big(\sum_{i=1}^{n} x_i\big)^2}{\theta^4} - \frac{2\,n}{\theta^3} \Big(\sum_{i=1}^{n} x_i\Big)$$

amb això ja podem escriure la informació de Fisher:

$$I_{(X_1,\ldots,X_n)}(\theta) = \frac{n^2}{\theta^2} + \frac{E\big(\sum_{i=1}^{n} X_i\big)^2}{\theta^4} - \frac{2\,n}{\theta^3}\, E\Big(\sum_{i=1}^{n} X_i\Big)$$

Suposem ara que $X_i \sim exp(\theta) = \Gamma(1, \frac{1}{\theta})$. Aleshores

$$X_1 + \ldots + X_n \sim \Gamma(n, \frac{1}{\theta})$$

$$E(X_1 + \ldots + X_n) = \frac{n}{1/\theta} = n\,\theta$$

$$Var(X_1 + \ldots + X_n) = \frac{n}{1/\theta^2} = n\,\theta^2$$

i per tant

$$E\big[\,(X_1 + \ldots + X_n)^2\,\big] = n\,\theta^2 + n^2\,\theta^2 = \theta^2\, n\,(1+n)$$

$$I_{(X_1,\ldots,X_n)}(\theta) = \frac{n^2}{\theta^2} + \frac{\theta^2\, n\,(1+n)}{\theta^4} - \frac{2\,n}{\theta^3}\, n\,\theta$$

$$= \frac{1}{\theta^2}\,[n^2 + n + n^2 - 2\,n^2] = \frac{n}{\theta^2} = n\, I_X(\theta)$$

Lema 2.1 *Sigui $X_1, \ldots, X_n$ una m.a.s. Aleshores es compleix:*

$$I_{\boldsymbol{x}}(\theta) = n\, I_{X_1}(\theta)$$

Demostració: En efecte,

$$\frac{\partial}{\partial\theta} \ln f(\boldsymbol{x} \,|\, \theta) = \frac{\partial}{\partial\theta} \ln \prod_{i=1}^{n} f(x_i \,|\, \theta) = \frac{\partial}{\partial\theta} \sum_{i=1}^{n} \ln f(x_i \,|\, \theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f(x_i \,|\, \theta)$$

i per tant

$$I_{\boldsymbol{x}}(\theta) = E\left\{ \left(\frac{\partial}{\partial\theta} log f(\boldsymbol{x} \,|\theta) \right)^2 \right\} = E\left\{ \left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} log f(X_i \,|\, \theta) \right)^2 \right\}$$

$$= E\left\{ \sum_{i=1}^{n} \left(\frac{\partial}{\partial\theta} log f(X_i \,|\, \theta) \right)^2 + \sum_{i \neq j} \left(\frac{\partial}{\partial\theta} log f(X_i \,|\, \theta) \, \frac{\partial}{\partial\theta} log f(X_j \,|\, \theta) \right) \right\}$$

$$= n\, E\left\{ \left(\frac{\partial}{\partial\theta} log f(X_1 \,|\theta) \right)^2 \right\} + \sum_{i \neq j} E\left\{ \frac{\partial}{\partial\theta} log f(X_i \,|\, \theta) \right\} E\left\{ \frac{\partial}{\partial\theta} log f(X_j \,|\, \theta) \right\}$$

i com que

$$E\left\{ \frac{\partial}{\partial\theta} log f(X_1 \,|\, \theta) \right\} = \int \left(\frac{\partial}{\partial\theta} log f(x_1 \,|\, \theta) \right) f(x_1 \,|\, \theta)\, dx_1$$

$$= \int \frac{\frac{\partial}{\partial\theta} f(x_1 \,|\, \theta)}{f(x_1 \,|\, \theta)} f(x_1 \,|\, \theta) dx_1$$

$$= \frac{\partial}{\partial\theta} \int f(x_1 \,|\, \theta) dx_1 = 0$$

ja tenim que

$$I_{\boldsymbol{x}}(\theta) = E\left\{ \sum_{i=1}^{n} \left(\frac{\partial}{\partial\theta} log f(X_i \,|\, \theta) \right)^2 \right\} = n I_x(\theta)$$

□

Exemple 2.24, pàgina 50. Continuació.

Notem que en el cas que $X \sim exp\,(1/\theta)$

$$\frac{\partial}{\partial\theta}\, log f(x\,|\,\theta) = \frac{-1}{\theta} + \frac{x}{\theta^2}$$

$$\frac{\partial^2}{\partial\theta^2}\, log f(x\,|\,\theta) = \frac{1}{\theta^2} - \frac{2\,x}{\theta^3}$$

$$E\Big(\frac{\partial^2}{\partial\theta^2}\, log f(X\,|\,\theta)\Big) = \frac{1}{\theta^2} - \frac{2}{\theta^3}\,E(X) = \frac{1}{\theta^2} - \frac{2}{\theta^3}\,\theta$$

$$= \frac{1}{\theta^2} - \frac{2}{\theta^2} = \frac{-1}{\theta^2} = -I_X(\theta)$$

Lema 2.2 *Si la funció de densitat (o de probabilitat) satisfà*

$$\frac{d}{d\theta}\,E_\theta\Big\{\frac{\partial}{\partial\theta}\, log f(X\,|\,\theta)\Big\} = \int \frac{\partial}{\partial\theta}\Big[\Big(\frac{\partial}{\partial\theta}\, log f(x\,|\,\theta)\Big)\, f(x\,|\,\theta)\Big]\,dx$$

(fixem-nos que aquesta és una propietat que verifica la família exponencial) aleshores:

$$E_\theta\Big\{\Big(\frac{\partial}{\partial\theta}\, log f(X\,|\,\theta)\Big)^2\Big\} = -E_\theta\Big\{\frac{\partial^2}{\partial\theta^2}\, log f(X\,|\,\theta)\Big\} = I_X(\theta)$$

Notem que quan s'aplica a $\boldsymbol{x} = (X_1,\ldots,X_n)$ m.a.s. de $f(x\,|\,\theta)$ aleshores tenim que

$$I_{\boldsymbol{x}}(\theta) = -E_\theta\Big\{\frac{\partial^2}{\partial\theta^2}\, log f(\boldsymbol{x}\,|\,\theta)\Big\} = -n\,E_\theta\Big\{\frac{\partial^2}{\partial\theta^2}\, log f(X_1\,|\,\theta)\Big\}$$

Demostració:

$$I(\theta)+E\left[\frac{\partial^2}{\partial\theta^2}\,log f(X;\theta)\right]=\int\left[\left(\frac{\partial}{\partial\theta}\,log f(x;\theta)\right)^2+\left(\frac{\partial^2}{\partial\theta^2}\,log f(x;\theta)\right)\right]f(x;\theta)\,dx$$

$$=\int\left(\frac{\partial}{\partial\theta}\,log f(x;\theta)\right)\frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}\,f(x;\theta)dx+\int\left(\frac{\partial^2}{\partial\theta^2}\,log f(x;\theta)\right)f(x;\theta)\,dx$$

$$=\int\left(\frac{\partial}{\partial\theta}\,log f(x;\theta)\right)\left(\frac{\partial}{\partial\theta}\,f(x;\theta)\right)dx+\int\frac{\partial}{\partial\theta}\underbrace{\left(\frac{\partial}{\partial\theta}\,log f(x;\theta)\right)}_{=u}\,\underbrace{f(x;\theta)}_{=v}\,dx$$

$$=\int(u\,v'+u'\,v)=\int\frac{\partial}{\partial\theta}(u\,v)=\int\frac{\partial}{\partial\theta}\left[\frac{\frac{\partial}{\partial\theta}f(x;\theta)}{f(x;\theta)}\,f(x;\theta)\right]dx$$

$$=\frac{\partial^2}{\partial\theta^2}\underbrace{\int f(x;\theta)dx}_{=1}=0$$

$\square$

Teorema 2.9 Teorema de Cramer-Rao

Sigui $\boldsymbol{x}=(X_1,\ldots,X_n)$ una m.a.s. amb funció de densitat conjunta $f(\boldsymbol{x}\,|\,\theta)$ i $W(\boldsymbol{x})=W(X_1,\ldots,X_n)$ un estimador tal que $E_\theta(W(\boldsymbol{x}))=\tau(\theta)$ per a tot θ, on τ és una funció de θ que verifica:

H1: $\tau(\theta)$ és diferenciable en θ.

Suposem a més que la funció de densitat conjunta compleix

H2: Per a qualsevol funció $h(\boldsymbol{x})$ tal que $E_\theta|\,h(\boldsymbol{x})|<\infty$ es té el següent:

$$\frac{d}{d\theta}\int\ldots\int h(\boldsymbol{x})\,f(\boldsymbol{x}\,|\,\theta)\,dx_1\,\ldots\,dx_n$$
$$=\int\ldots\int h(\boldsymbol{x})\left[\frac{\partial}{\partial\theta}\,f(\boldsymbol{x}\,|\,\theta)\right]dx_1\ldots dx_n$$

Aleshores:

$$Var_\theta\big(W(\boldsymbol{x})\big)\geq\frac{\left(\frac{d}{d\theta}\,\tau(\theta)\right)^2}{E_\theta\left[\left(\frac{\partial}{\partial\theta}\,log f(\boldsymbol{x}\,|\,\theta)\right)^2\right]}$$

A la quantitat de la dreta d'aquesta desigualtat se l'anomena **Fita de Cramer-Rao**.

Nota 1: Aquest teorema és igualment vàlid en el cas discret. En aquest cas, la hipòtesi $H2$ afirma que es poden intercanviar el sumatori i la diferenciació.

Nota 2: Totes les lleis de la família exponencial compleixen les condicions per aplicar el teorema de Cramer-Rao.

Corol.lari 2.1 *Si un estimador no esbiaixat atrapa la fita de Cramer-Rao aleshores és UMVUE.*

Corol.lari 2.2 *Sota les mateixes condicions del teorema de Cramer-Rao, si* $\boldsymbol{x} = (X_1, \ldots, X_n)$ *és una m.a.s. de* X *amb llei* $f(x \,|\, \theta)$ *aleshores*

$$E_\theta\left[\left(\frac{\partial}{\partial\theta} logf(\boldsymbol{x} \,|\, \theta)\right)^2\right] = n\, E_\theta\left[\left(\frac{\partial}{\partial\theta} logf(X \,|\, \theta)\right)^2\right]$$

Demostració: Per la independència de les variables aleatòries, la versemblança de $\boldsymbol{x}$ és el producte de les versemblances,

$$\frac{\partial}{\partial\theta} logf(\boldsymbol{x} \,|\, \theta) = \frac{\partial}{\partial\theta} \sum_{i=1}^{n} logf(x_i \,|\, \theta) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta} logf(x_i \,|\, \theta)$$

Per tant,

$$E_\theta\left[\left(\frac{\partial}{\partial\theta} logf(\boldsymbol{x} \,|\, \theta)\right)^2\right] = E_\theta\left[\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} logf(X_i \,|\, \theta)\right)^2\right] =$$

$$\sum_{i=1}^{n} E_\theta\left[\left(\frac{\partial}{\partial\theta} logf(X_i \,|\, \theta)\right)^2\right] + \sum_{i \neq j} E_\theta\left[\frac{\partial}{\partial\theta} logf(X_i \,|\, \theta)\, \frac{\partial}{\partial\theta} logf(X_j \,|\, \theta)\right]$$

Ara bé, el segon dels sumatoris és igual a zero (mireu el lema anterior), degut a la independència entre X_i i X_j, amb la qual cosa ja tenim el que volíem.

$\square$

La quantitat $I_{X_i}(\theta) = E_\theta\Big[\left(\frac{\partial}{\partial\theta} log f(X_i \,|\, \theta)\right)^2\Big]$ és la informació de Fisher continguda en X_i.

La quantitat $I_{\boldsymbol{x}}(\theta) = E_\theta\Big[\left(\frac{\partial}{\partial\theta} log f(\boldsymbol{x} \,|\, \theta)\right)^2\Big]$ és la informació de Fisher continguda en $\boldsymbol{x}$.

Quan $X_1, \ldots, X_n$ són i.i.d. es verifica que la informació de Fisher continguda en la mostra és la suma de les informacions contingudes en cadascuna de les observacions, i com que aquestes són idènticament distribuides, tenim que

$$I_{\boldsymbol{x}}(\theta) = n\, I_{X_1}(\theta)$$

La informació de Fisher, doncs, ens proporciona una fita inferior per la variància del millor estimador sense biaix de θ.

$$Var_\theta\big(W(\boldsymbol{x})\big) \geq \frac{1}{I_{\boldsymbol{x}}} = \frac{1}{n\, I_X(\theta)}$$

Exemple 2.25

Sigui $X_1, \ldots, X_n$ una m.a.s. de $X \sim Poisson(\lambda)$. Per tant, tenim $E(X) = \lambda$ i considerem, d'acord amb això, $\tau(\lambda) = \lambda$ i $\tau'(\lambda) = 1$. Obtenim doncs,

$$I_{\boldsymbol{x}}(\lambda) = E_\lambda\left[\left(\frac{\partial}{\partial\lambda} log \prod_{i=1}^{n} f(X_i \,|\, \lambda)\right)^2\right] = -n\, E_\lambda\left[\frac{\partial^2}{\partial\lambda^2} log f(X \,|\, \lambda)\right] =$$

$$-n\, E_\lambda\left[\frac{\partial^2}{\partial\lambda^2} log\left(\frac{e^{-\lambda}\,\lambda^X}{X!}\right)\right] = -n\, E_\lambda\left[\frac{\partial^2}{\partial\lambda^2}\left(-\lambda + X\, log\lambda - log X!\,\right)\right]$$

$$= -n\, E_\lambda\left[-\frac{X}{\lambda^2}\right] = \frac{n}{\lambda}$$

Per tant, per a qualsevol estimador W no esbiaixat de λ es tindrà:

$$Var_\lambda(W) \geq \frac{1}{n/\lambda} = \frac{\lambda}{n}$$

Sabem d'altra banda que $\overline{X}$ és un estimador sense biaix de λ i que $Var_\lambda(\overline{X}) = \lambda/n$, amb la qual cosa podem deduïr que la mitjana mostral és el millor estimador (UMVUE) sense biaix de λ.

Exemple 2.26

Sigui $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma^2)$, amb tots dos paràmetres desconeguts.

a) Considerem en primer lloc el problema d'estimar el paràmetre μ.

Sabem que

$$f(x \,|\, \mu, \sigma^2) = \frac{1}{\sqrt{2\,\pi\,\sigma^2}}\, e^{-\frac{1}{2\,\sigma^2}\,(x-\mu)^2}$$

i per tant

$$log f(x \,|\, \mu, \sigma^2) = log\left(\frac{1}{\sqrt{2\,\pi\sigma^2}}\right) - \frac{1}{2\,\sigma^2}\,(x-\mu)^2$$

Anem a derivar aquesta expressió respecte al paràmetre que volem estimar:

$$\frac{\partial}{\partial\mu}\, log f(x \,|\, \mu, \sigma^2) = \frac{1}{2\,\sigma^2}\, 2\,(x-\mu) = \frac{x-\mu}{\sigma^2}$$

Així, doncs, hem obtingut que

$$I_X(\mu) = E_\mu\left[\left(\frac{\partial}{\partial\mu}\, log f(X \,|\, \mu, \sigma^2)\right)^2\right] = E_\mu\left[\left(\frac{X-\mu}{\sigma^2}\right)^2\right] =$$

$$\frac{1}{\sigma^4}\, Var(X) = \frac{\sigma^2}{\sigma^4} = \frac{1}{\sigma^2}$$

$$I_{\boldsymbol{X}}(\mu) = \frac{n}{\sigma^2}$$

$$V(\overline{X}_n) = \frac{\sigma^2}{n} = \frac{1}{I_{\boldsymbol{X}}(\mu)}$$

amb la qual cosa podem afirmar que $\overline{X}_n$ és UMVUE.

b) Considerem ara el problema d'estimar σ^2. Com que estem dins la família exponencial es compleixen les condicions del teorema de Cramer-Rao i per tant

$$I_X(\sigma^2) = -E_{\sigma^2}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\, log f(X\,|\,\theta)\right] =$$

$$-E_{\sigma^2}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\, log\left(\frac{1}{(2\,\pi\,\sigma^2)^{1/2}}\, e^{-\frac{(X-\mu)^2}{2\,\sigma^2}}\right)\right] =$$

$$-E_{\sigma^2}\left[\frac{\partial^2}{\partial(\sigma^2)^2}\left(log\, K - \frac{1}{2}\, log\, \sigma^2 - \frac{(X-\mu)^2}{2\,\sigma^2}\right)\right] =$$

$$-E_{\sigma^2}\left[\frac{\partial}{\partial\sigma^2}\left(-\frac{1}{2\,\sigma^2} + \frac{(X-\mu)^2}{2\,\sigma^4}\right)\right] =$$

$$-E_{\sigma^2}\left[\frac{1}{2\,\sigma^4} - \frac{(X-\mu)^2}{\sigma^6}\right] = \frac{-1}{2\,\sigma^4} + \frac{\sigma^2}{\sigma^6} = \frac{1}{2\,\sigma^4}$$

Qualsevol estimador no esbiaixat de σ^2, com ara

$$W(\boldsymbol{X}) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

ha de verificar

$$Var\big(W(\boldsymbol{X})\big) \geq \frac{1}{n\, I_X(\sigma^2)} = \frac{2\,\sigma^4}{n}$$

Hem vist que

$$Var(S^2) = \frac{2\,\sigma^4}{n-1} > \frac{2\,\sigma^4}{n}$$

i per tant S^2 no atrapa la fita de Cramer-Rao.

Definició 2.18 *Un estimador no esbiaixat per a $\tau(\theta)$ es diu que és* **eficient** *si la seva variància és la mínima possible, és a dir, si és igual a la fita de Cramer-Rao.* **L'eficiència** *d'un estimador no esbiaixat es defineix com el quocient entre aquesta fita i la seva variància. És un valor menor o igual que 1 en cas que es donin les hipòtesis del teorema de Cramer-Rao.*

2.2.3 Estadístics suficients. Teorema de factorització

El principi de suficiència

Un estadístic suficient per a un paràmetre θ és un estadístic que captura tota la informació sobre θ que conté la mostra. Qualsevol informació addicional que la mostra pot proporcionar, suplementària a la de l'estadístic suficient, no aporta informació rellevant sobre θ.

Principi: Si $T(X)$ és un estadístic suficient per a θ, qualsevol inferència sobre θ ha de dependre de la mostra $X = (X_1, \ldots, X_n)$ només mitjançant el valor $T(X)$. És a dir, si $\mathbf{x} = (x_1, \ldots, x_n)$, $\mathbf{y} = (y_1, \ldots, y_n)$ són tals que $T(\mathbf{x}) = T(\mathbf{y})$ aleshores la inferència que es farà sobre θ serà la mateixa tant si hem observat $\mathbf{x}$ com si hem observat $\mathbf{y}$.

Definició 2.19 *Un estadístic $T(X)$ és* **suficient** *per a θ si la distribució condicional de X donat el valor de $T(X)$ no depèn de θ, és a dir, si*

$$P_\theta\big(\boldsymbol{X} = \boldsymbol{x} \,|\, T(\boldsymbol{x}) = t(\boldsymbol{x})\big) \qquad \textit{no depèn de } \theta$$

Teorema 2.10 *Si $f(x \,|\, \theta)$ és la funció de densitat conjunta de X i $q(t \,|\, \theta)$ és la funció de densitat de $T(X)$, aleshores l'estadístic és suficient si i només si el quocient*

$$\frac{f(\mathbf{x} \,|\, \theta)}{q(T(\mathbf{x}) \,|\, \theta)}$$

és independent de θ.

Exemple 2.27

Sigui $X = (X_1, \ldots, X_n)$ mostra aleatòria simple de $X \sim Bern(\theta)$, on $0 < \theta < 1$. L'estadístic $T(X) = X_1 + \ldots + X_n \sim Bin(n, \theta)$ és suficient per a θ.

En efecte:

$$\frac{f(\mathbf{x}\,|\,\theta)}{q(T(X)\,|\,\theta)} = \frac{\prod_{i=1}^{n}\theta^{x_i}\,(1-\theta)^{1-x_i}}{\binom{n}{t}\,\theta^t\,(1-\theta)^{n-t}} = \frac{\theta^{\sum x_i}(1-\theta)^{n-\sum x_i}}{\binom{n}{t}\,\theta^t\,(1-\theta)^{n-t}} = \frac{\theta^t\,(1-\theta)^{n-t}}{\binom{n}{t}\,\theta^t\,(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}}$$

on hem definit $t = \sum_{i=1}^{n} x_i$, i podem observar que aquest valor no depèn de θ, amb la qual cosa ja tenim el que volíem.

Anem ara a veure quina forma té la distribució condicional de X, i veurem que tampoc dependrà del paràmetre θ.

$$\begin{aligned} & P_\theta\big(\,\boldsymbol{x} = \boldsymbol{x}\,|\,T(\boldsymbol{x}) = t\,\big) \qquad \text{on } t = T(\boldsymbol{x}) \\ &= \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x}, T(\boldsymbol{X}) = t)}{P(T(\boldsymbol{x}) = t)} = \frac{P_\theta(\boldsymbol{X} = \boldsymbol{x})}{P\big(\sum_{i=1}^{n} X_i = t\big)} \\ &= \frac{\prod_{i=1}^{n}\theta^{x_i}\,(1-\theta)^{1-x_i}}{\binom{n}{t}\,\theta^t\,(1-\theta)^{n-t}} = \frac{\theta^{\sum_{i=1}^{n} x_i}\,(1-\theta)^{n-\sum_{i=1}^{n} x_i}}{\binom{n}{t}\,\theta^t\,(1-\theta)^{n-t}} = \frac{1}{\binom{n}{t}} \end{aligned}$$

ja que $t = \sum_{i=1}^{n} x_i$.

El teorema que enunciem a continuació, provat per Halmos i Savage, ens ajuda a trobar un estadístic suficient mitjançant la inspecció de la funció de densitat.

Teorema 2.11 Teorema de Factorització

Sigui $f(\mathbf{x}\,|\,\theta)$ la funció de densitat conjunta de X. $T(X)$ és suficient per a θ si i només si existeix una funció $g(t\,|\,\theta)$ i una altra funció $h(x)$ tals que per qualsevol vector $\mathbf{x}$ de l'espai mostral i qualsevol valor del paràmetre θ podem factoritzar

$$f(\mathbf{x}\,|\,\theta) = g\big(T(x)\,|\,\theta\big)\,h(x)$$

Demostració:

Ens situem en el cas discret. Suposem que $T(X)$ és suficient. Escollim $g(t\,|\,\theta) = P_\theta(T(X) = t)$ i $h(x) = P\big(X = x\,|\,T(X) = T(x)\big)$, que no depèn de θ, per ser T suficient. Amb aquesta tria,

$$
\begin{aligned}
f(x \,|\, \theta) &= P_\theta\big(X = x \text{ i } T(X) = T(x)\big) \\
&= P_\theta\big(X = x \,|\, T(X) = T(x)\big)\, P_\theta\big(T(X) = T(x)\big) \\
&= h(x)\, g\big(T(x) \,|\, \theta\big)
\end{aligned}
$$

amb la qual cosa ja tenim la factorització.

Anem a veure ara el recíproc.

Prenem $t = T(\boldsymbol{x})$ i calculem la distribució condicional de X.

$$
P\big(\boldsymbol{X} = \boldsymbol{x} \,|\, T(\boldsymbol{x}) = t\big) = \frac{P(\boldsymbol{X} = \boldsymbol{x})}{P(T(\boldsymbol{x}) = t)}
$$

$$
= \frac{h(x)\, g(t \,|\, \theta)}{\sum\limits_{y:T(y)=t} P(\boldsymbol{X} = y)} = \frac{h(\boldsymbol{x})\, g(t \,|\, \theta)}{\sum\limits_{\tilde{y}:T(\tilde{y})=t} h(\tilde{y})\, g(t \,|\, \theta)}
$$

i aquest valor és independent de θ.

$\square$

Exemple 2.28

Sigui $X = (X_1, \ldots, X_n)$ una m.a.s. amb $X \sim N(\mu, \sigma^2)$ i σ^2 coneguda. Aleshores tenim:

$$
f(\mathbf{x} \,|\, \mu) = \underbrace{(2\,\pi\,\sigma^2)^{-n/2}\, exp\left\{ -\frac{1}{2\,\sigma^2} \sum_{i=1}^{n} (x_i - \overline{x})^2 \right\}}_{h(x)} \underbrace{exp\left\{ -\frac{n}{2\,\sigma^2}\, (\overline{x} - \mu)^2 \right\}}_{g(t \,|\, \mu)}
$$

essent $t = \overline{x}$.

Nota: $\sum\limits_{i=1}^{n} x_i$ també és suficient. Podríem definir $h(x)$ igual i $g(t \,|\, \mu) = exp\left\{ -\frac{n}{2\,\sigma^2}\, (\frac{t}{n} - \mu)^2 \right\}$.

Exemple 2.29

Sigui $X_1, \ldots, X_n \sim U\{1, \ldots, \theta\}$. Aleshores sabem que

$$f(x \mid \theta) = \begin{cases} 1/\theta & \text{si } x = 1, \ldots, \theta \\ 0 & \text{altrament} \end{cases}$$

Per tant,

$$\begin{aligned} f(x_1, \ldots, x_n \mid \theta) &= \begin{cases} \frac{1}{\theta^n} & \text{si } x_i \in \{1, \ldots, \theta\} \;\; i = 1, \ldots, n \\ 0 & \text{altrament} \end{cases} \\ &= \begin{cases} \frac{1}{\theta^n} & \text{si } x \in \{1, 2, \ldots\} \text{ i } max\, x_i < \theta \\ 0 & \text{altrament} \end{cases} \\ &= \tfrac{1}{\theta^n} 1_{\{1,\ldots,\theta\}} \left(max_{1 \leq i \leq n} x_i\right) \qquad x_i \in \{1, 2, \ldots\} \\ &= \underbrace{\frac{1}{\theta^n} 1_{\{1,\ldots,\theta\}} \left(max_{1 \leq i \leq n} x_i\right)}_{g(t \mid \theta)} \underbrace{\prod_{i=1}^{n} 1_N(x_i)}_{h(x)} \end{aligned}$$

on $t = max_{1 \leq i \leq n} x_i$.

Estadístics suficients r-dimensionals

Fins ara hem treballat amb estadístics suficients a valors reals però quan apliquem el teorema de factorització ens podem trobar que la funció $g(t \mid \theta)$ depengui de la mostra en més d'una funció $T(X)$.

En aquest cas $T(X) = \big(T_1(x), T_2(x), \ldots, T_r(x)\big)$ és un estadístic r-dimensional i les mateixes definicions i teoremes referents a la suficiència es poden aplicar.

Usualment quan el paràmetre sigui un vector de dimensió més gran que l'estadístic suficient també serà un vector de dimensió més gran que 1.

Exemple 2.30

Sigui $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, amb ambdós paràmetres desconeguts, $\theta = (\mu, \sigma^2)$. Aleshores

$$f(\mathbf{x}\,|\,\mu,\sigma^2) = (2\,\pi\,\sigma^2)^{-n/2}\, exp\Big\{ -\frac{1}{2\,\sigma^2}\, \sum_{i=1}^{n}(x_i - \overline{x})^2 \Big\}\, exp\Big\{ -\frac{n}{2\,\sigma^2}\,(\overline{x} - \mu)^2 \Big\}$$

Definint $T_1(x) = \overline{x}$ i $T_2(x) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$ tenim que:

$$f(\mathbf{x}\,|\,\mu,\sigma^2) = (2\,\pi\,\sigma^2)^{-n/2}\, exp\Big\{ -\frac{n-1}{2\,\sigma^2}\, t_2 \Big\}\, exp\Big\{ -\frac{n}{2\,\sigma^2}\,(t_1 - \mu)^2 \Big\} =$$

$$g(t_1, t_2\,|\,\mu, \sigma^2)$$

i prenent $h(x) = 1$.

Obtenim doncs que $T(X) = (T_1(x), T_2(x)) = (\overline{X}, S^2)$ és un estadístic suficient per a (μ, σ^2) en un model normal.

Notem a més a més que la definició, i per tant l'obtenció, dels estadístics suficients, depèn del model.

2.2.4 Estimadors consistents i eficiència relativa

Definició 2.20 *Una successió d'estimadors $W_n = W_n(X_1, \ldots, X_n)$ és una* **successió d'estimadors consistent** *pel paràmetre θ si per cada $\varepsilon > 0$ i per a cada $\theta \in \Theta$,*

$$\lim_{n\to\infty} \mathbf{P}_\theta\,(\,|\,W_n - \theta\,| < \varepsilon\,) = 1$$

Equivalentment, si

$$\forall\ \varepsilon > 0,\ \ \forall\theta \in \Theta,\ \ \lim_{n\to\infty} \mathbf{P}_\theta\,(\,|\,W_n - \theta\,| \geq \varepsilon) = 0$$

Hem d'adonar-nos aquí que estem permetent que la família de probabilitats varïi amb θ.

El següent resultat estableix una condició suficient perquè una successió d'estimadors sigui consistent. Denotem per $B_\theta(W)$ el biaix d'un estimador W de θ és a dir, $B_\theta(W) = E(W) - \theta$.

Teorema 2.12 *Si W_n és una successió d'estimadors de θ que satisfà:*

a) $\lim_{n\to\infty} Var_\theta(W_n) = 0$,

b) $\lim_{n\to\infty} B_\theta(W_n) = 0$.

Aleshores W_n és una successió consistent de θ.

Demostració: Observem que

$$E_\theta\big(\,(W_n - \theta)^2\,\big) = Var_\theta(W_n) + B_\theta^2(W_n)$$

Per altra banda, i usant la desigualtat de Chebyshev, es té que

$$\mathbf{P}_\theta\big(\,|\,W_n - \theta\,|\,\geq \varepsilon\big) = \mathbf{P}_\theta\big(\,(W_n - \theta)^2\,\geq \varepsilon^2\,\big) \leq \frac{E_\theta\big((W_n - \theta)^2\big)}{\varepsilon^2} =$$

$$\frac{1}{\varepsilon^2}\big[\,Var_\theta(W_n) + B_\theta^2(W_n)\,\big] \qquad \forall\ \varepsilon > 0 \ \text{ i } \ \forall\ \theta \in \Theta$$

i per hipòtesi del teorema ja podem afirmar que

$$\lim_{n\to\infty} \mathbf{P}_\theta\big(\,|\,W_n - \theta\,| < \varepsilon\,\big) = \lim_{n\to\infty} 1 - \mathbf{P}_\theta\big(\,|\,W_n - \theta\,| \geq \varepsilon\,\big) = 1 - 0 = 1$$

□

El requeriment de consistència és mínim, és a dir, hauríem d'esperar que a mesura que més dades siguin disponibles, un bon estimador hauria d'aproximar-se al verdader valor del paràmetre.

Definició 2.21 *Si W_1 i W_2 són dos estimadors no esbiaixats de θ, aleshores es defineix* **l'eficiència relativa** *de W_1 respecte de W_2 com*

$$RE(\theta, W_1, W_2) = \frac{Var_\theta(W_2)}{Var_\theta(W_1)}$$

D'aquesta manera, $RE(\theta, W_1, W_2) > 1$ si i només si $Var_\theta(W_2) > Var_\theta(W_1)$, si i només si W_1 utilitza les dades de manera més eficient que no pas W_2.

Exemple 2.31

Calculem l'eficiència relativa de la mediana mostral M_n respecte a la mitjana mostral $\overline{X}$ per tal d'estimar el paràmetre de centralitat μ de quatre distribucions simètriques (el paràmetre de centralitat és la mitjana poblacional i coincideix amb l'esperança quan aquesta existeix).

S'ofereixen els resultats en funció de la mida mostral n. Els valors de la variància de la mediana són aproximats i s'han calculat per mitjà de la fórmula

$$Var(M_n) \approx \frac{1}{4\, f^2(\mu)}$$

	$Var(\overline{X})$	$Var(M_n)$	$RE(\theta, M_n, \overline{X})$
Normal estàndar	$1/n$	$\pi/2\,n$	$2/\pi = 0.64$
Logística($\beta = 1$)	$\pi^2/3\,n$	$4/n$	$\pi^2/12 = 0.82$
Doble exponencial($\lambda = 1$)	$2/n$	$1/n$	2
Cauchy estandar	∞	$\pi^2/4\,n$	∞

Es pot concloure, doncs, que la mitjana mostral és més eficient que la mediana en les lleis normal i logística, i ho és menys en la doble exponencial i en la Cauchy.

Definició 2.22 *Sigui $T_n(\boldsymbol{X}) = T_n(X_1, \ldots, X_n)$ un estimador d'una funció $q(\theta)$ tal que*

$$\sqrt{n}\left[\,T_n(\boldsymbol{X}) - q(\theta)\,\right] \xrightarrow{\mathcal{D}} N\big(\,b(\theta), \sigma^2(\theta)\,\big)$$

Si $b(\theta) = 0$ direm que $T_n(\boldsymbol{X})$ és **asimptòticament no esbiaixat.** *Altrament, si $b(\theta) \neq 0$ direm que $T_n(\boldsymbol{X})$ és* **asimptòticament esbiaixat***.*

Exemple 2.32 ______________________________

Si $X_1, \ldots, X_n, \ldots$ són i.i.d. amb mitjana μ, variància 1 i $T_n(\boldsymbol{X}) = \overline{X}_n + a/\sqrt{n}$ per una constant $a \neq 0$, tenim que

$$\sqrt{n}\left[\,\overline{X}_n - \mu\,\right] \xrightarrow{\mathcal{D}} N(0,1)$$

i amb això

$$\sqrt{n}\left[\,T_n(\boldsymbol{X}) - \mu\,\right] = \sqrt{n}\left[\,\overline{X}_n + \frac{a}{\sqrt{n}} - \mu\,\right]$$

$$\sqrt{n}\left[\,\overline{X}_n - \mu\,\right] + a \xrightarrow{\mathcal{D}} N(a,1)$$

per tant T_n és asimptòticament esbiaixat: la diferència entre l'esperança de l'estimador i el paràmetre estimat, multiplicada per $\sqrt{n}$, no tendeix a 0.

Notem també que $T_n(\boldsymbol{X}) = \overline{X}_n + \frac{a}{\sqrt{n}} \xrightarrow{P} \mu$ i per tant T_n és un estimador consistent de μ.

.______________________________.

Definició 2.23 *Si*

a) $\sqrt{n}\left[\,T_n(X) - q(\theta)\,\right] \xrightarrow{\mathcal{D}} N(0, \sigma_T^2)$

b) $\sqrt{n}\left[\,S_n(X) - q(\theta)\,\right] \xrightarrow{\mathcal{D}} N(0, \sigma_S^2)$

aleshores **l'eficiència relativa asimptòtica** *de S_n respecte a T_n es defineix com*

$$ARE(\theta, S_n, T_n) = \frac{\sigma_T^2}{\sigma_S^2}$$

El valor de l'eficiència relativa asimptòtica es pot interpretar com el quocient de les mides de mostra necessàries per obtenir la mateixa precisió asimptòtica (o la mateixa variància asimptòtica) mitjançant els dos estimadors en l'estimació de $\tau(\theta)$. En efecte, si triem mida mostral m per a T i n per a S, les variànciaces asimptòtiques són, respectivament, $\sigma_T^2(\theta)/m$ i $\sigma_S^2(\theta)/n$. Si forcem a que siguin iguals, llavors tenim

$$\frac{\sigma_T^2(\theta)}{m} = \frac{\sigma_S^2(\theta)}{n} \Longleftrightarrow \frac{m}{n} = \frac{\sigma_T^2(\theta)}{\sigma_S^2(\theta)} = ARE(\theta, S_n, T_n)$$

És a dir, si $ARE(\theta, S_n, T_n) = 0.5$, aleshores S_n és menys eficient que T asimptòticament: per tenir la mateixa precisió amb l'estimador S fa falta una mostra el doble de gran que si féssim servir T ($ARE = 0.5 = m/n \Longrightarrow n = 2\,m$).

2.2.5 Propietats asimptòtiques dels estimadors de màxima versemblança

Sigui $X_1, \ldots, X_n$ m.a.s. amb funció de densitat $f(x \mid \theta)$. Sigui

$$L(\theta \mid \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i \mid \theta)$$

la funció de versemblança. Considerem $\widehat{\theta}$ el MLE de θ i $\tau(\theta)$ una funció contínua de θ. Sota certes condicions de regularitat sobre $f(x \mid \theta)$ tenim que

$$\tau(\,\widehat{\theta}\,) \xrightarrow[n\to\infty]{P} \tau(\theta)$$

és a dir, que $\tau(\,\widehat{\theta}\,)$ és un estimador consistent de $\tau(\theta)$. En particular, tenim que $\widehat{\theta} \xrightarrow{P} \theta$.

Teorema 2.13 *Sota condicions de regularitat per la funció de densitat o de probabilitat, si $\widehat{\theta}_n$ indica l'estimador màxim versemblant de θ i si denotem per θ_0 el veritable valor (desconegut) de θ, tenim que*

$$\sqrt{n\, I(\theta_0)}\, (\widehat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\mathcal{D}} Z \sim N(0,1)$$

i per tant $\widehat{\theta}_n$ és asimptòticament no esbiaixat i $\frac{1}{n\, I(\theta_0)}$ és la variància asimptòtica de $\widehat{\theta}_n$.

Aquest resultat ens permet estimar la precisió de l'estimador $\widehat{\theta}_n$ i basant-nos en aquest fet trobar intervals de confiança aproximats. Hem demostrat que $\widehat{\theta}_n \xrightarrow[n\to\infty]{P} \theta_0$, i per tant, a menys que $I(\theta)$ no sigui contínua en θ, tindrem que

$I(\,\widehat{\theta}_n\,) \overset{P}{\to} I(\theta_0)$, amb la qual cosa la variància asimptòtica $1/(n\,I(\theta_0))$ de $\widehat{\theta}_n$ pot estimar-se mitjançant $1/(n\,I(\,\widehat{\theta}_n\,))$.

D'altra banda, si

$$\left.\begin{array}{ll} \sqrt{n\,I(\theta_0)}\,(\widehat{\theta}_n-\theta_0) & \approx N(0,1) \\ \frac{\sqrt{n\,I(\theta_0)}}{\sqrt{n\,I(\widehat{\theta}_n)}} & \overset{P}{\to} 1 \end{array}\right\} \Longrightarrow \sqrt{n\,I(\widehat{\theta}_n)}\,(\widehat{\theta}_n-\theta_0) \approx N(0,1)$$

Per tant, si volem estimar el paràmetre θ d'una llei, calcularem primer el MLE $\widehat{\theta}_n$, tot seguit calcularem la informació observada, és a dir,

$$-\frac{\partial^2}{\partial\theta^2}\,logL(\theta\,|\,\boldsymbol{x})|_{\theta=\widehat{\theta}_n} = I\,Obs(\widehat{\theta}_n)$$

i es pot comprovar que

$$Var(\widehat{\theta}_n) = \frac{1}{I\,Obs(\widehat{\theta}_n)}$$

Si el que volem estimar és $\tau(\theta)$, usarem $\tau(\,\widehat{\theta}_n\,)$, amb

$$Var\big(\tau(\widehat{\theta}_n)\big) \approx \frac{\big(\tau'(\theta)\big)^2|_{\theta=\widehat{\theta}_n}}{I\,Obs(\,\widehat{\theta}_n\,)}$$

Per tant, per fer inferència, usarem el fet que

$$\sqrt{I\,Obs(\,\widehat{\theta}_n\,)}\,(\widehat{\theta}_n-\theta) \approx N(0,1)$$

o bé que

$$\sqrt{\frac{I\,Obs(\,\widehat{\theta}_n\,)}{\tau'(\theta)^2|_{\theta=\theta_n}}}\,\big(\,\tau(\,\widehat{\theta}_n\,)-\tau(\theta)\,\big) \approx N(0,1)$$

Exemple 2.33

Volem estimar els anomenats odds = avantatge = $\frac{p}{1-p}$, on $X_1, \ldots, X_n \sim Bern(p)$.

Sabem que $\widehat{p} = \overline{X}_n$ és el MLE.

$$Var_p(\widehat{p}) = \frac{p\,(1-p)}{n}$$

$$L(p) = \prod_{i=1}^{n} p^{x_i}\,(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}\,(1-p)^{n-\sum_{i=1}^{n} x_i}$$

Si ara li calculem el logaritme i tot seguit derivem respecte al paràmetre, obtenim

$$logL(p) = \left(\sum_{i=1}^{n} x_i\right) log(p) + \left(n - \sum_{i=1}^{n} x_i\right) log(1-p)$$

$$\frac{\partial}{\partial p}\, logL(p) = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{n - \sum_{i=1}^{n} x_i}{1-p}$$

$$\frac{\partial^2}{\partial p^2}\, logL(p) = \frac{-\sum_{i=1}^{n} x_i}{p^2} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{(1-p)^2}$$

i un estimador intuïtiu fóra $\widehat{Var}(\widehat{p}) = \frac{\widehat{p}\,(1-\widehat{p})}{n}$.

Si ara calculem

$$\widehat{Var}(\widehat{p}) \approx \frac{1}{-\frac{\partial^2}{\partial p^2}\, logL(p\,|\,x)|_{p=\widehat{p}}}$$

$$logL(p\,|\,x) = n\,\widehat{p}\; log\, p + n\;(1-\widehat{p})\; log(1-p)$$

$$\frac{\partial^2}{\partial p^2}\, logL(p\,|\,x) = -\frac{n\,\widehat{p}}{p^2} - \frac{n\,(1-\widehat{p})}{(1-p)^2}$$

$$I\,Obs\,(\widehat{p}_n) = -\left(-\frac{n\,\widehat{p}}{(\widehat{p})^2} - \frac{n\,(1-\widehat{p})}{(1-\widehat{p})^2}\right) = \frac{n}{\widehat{p}\,(1-\widehat{p})}$$

i obtenim un resultat idèntic. Per tant

$$\sqrt{\frac{n}{\widehat{p}\,(1-\widehat{p}\,)}}\,(\widehat{p}-p) \approx N(0,1)$$

Si volem estimar els odds $\tau(p) = \frac{p}{1-p}$ podem proposar

$$\tau(\,\widehat{p}\,) = \frac{\widehat{p}}{1-\widehat{p}} = \frac{\frac{\sum_{i=1}^n x_i}{n}}{1-\frac{\sum_{i=1}^n x_i}{n}} = \frac{\sum_{i=1}^n x_i}{n-\sum_{i=1}^n x_i}$$

Com que

$$\frac{\partial}{\partial p}\left(\frac{p}{1-p}\right) = \frac{1-p-(-1)\,p}{(1-p)^2} = \frac{1}{(1-p)^2} = \tau'(p)$$

amb la qual cosa

$$\widehat{Var}\left(\frac{\widehat{p}}{1-\widehat{p}}\right) = \frac{\left(\frac{1}{(1-p)^2}\right)^2\Big|_{p=\widehat{p}}}{I\,Obs\,(\widehat{p}_n\,)} = \frac{\frac{1}{(1-\widehat{p}\,)^4}}{\frac{n}{\widehat{p}\,(1-\widehat{p}\,)}} = \frac{\widehat{p}}{n\,(1-\widehat{p}\,)^3}$$

i amb això

$$\sqrt{\frac{n\,(1-\widehat{p}\,)^3}{\widehat{p}}}\,\big(\tau(\widehat{p}\,)-\tau(p)\big) = \sqrt{\frac{n\,(1-\widehat{p}\,)^3}{\widehat{p}}}\left(\frac{\widehat{p}}{1-\widehat{p}}-\frac{p}{1-p}\right) \approx N(0,1)$$

Anem a considerar els següents resultats ($n = 15$):

0 0 1 0 0 0 0 0 1 1 0 0 0 0 1 1

En aquest cas, doncs $\widehat{p} = \frac{5}{15} = 0.33$ i per tant

$$\widehat{Var}(\,\widehat{p}\,) = \frac{\widehat{p}\,(1-\widehat{p}\,)}{n} = \frac{0.222}{15} = 0.0148$$

Per altra banda,

$$I(p) = \frac{\sum_{i=1}^n x_i}{p^2} + \frac{n-\sum_{i=1}^n x_i}{(1-p)^2}$$

$$I(p)|_{p=\widehat{p}} = \frac{5}{(0.33)^2} + \frac{10}{(0.66)^2} = 45.001 + 22.5 = 67.5$$

$$\frac{1}{I(p)|_{p=\widehat{p}}} = \frac{1}{67.5} = 0.0148$$

$$\sqrt{67.5}\,(\widehat{p} - p) = \sqrt{67.5}\left(\frac{\sum_{i=1}^n x_i}{15} - p\right)$$

$$\tau(p) = \frac{p}{1-p} \text{ i } \tau(\widehat{p}) = \frac{0.33}{1-0.33} = 0.5 = \frac{5/15}{10/15}$$

i per tant

$$\sqrt{\frac{15\left(1-\frac{5}{15}\right)^3}{5/15}}\big(\tau(\widehat{p}) - \tau(p)\big) = \sqrt{\frac{15^2\,\frac{10^3}{15^3}}{5}}\left(0.5 - \frac{p}{1-p}\right) =$$

$$\sqrt{\frac{10^3}{5\,15}}\left(0.5 - \frac{p}{1-p}\right) = 10\sqrt{\frac{2}{15}}\left(0.5 - \frac{p}{1-p}\right) = 3.65\left(0.5 - \frac{p}{1-p}\right) \approx N(0,1)$$

2.2.6 Estimació per intervals de confiança

En aquest tema, fins ara hem estudiat els estimadors puntuals. Aquests estimadors ofereixen un únic valor com a estimació del paràmetre desconegut θ. A partir d'ara tractarem el problema de l'estimació per conjunts, on s'estudien estimadors que proporcionen un conjunt com a estimació de θ. El resultat d'una estimació per conjunts és una afirmació del tipus "$\theta \in C$", on $C = C(\boldsymbol{x})$ és un subconjunt de l'espai paramètric Θ que depèn de les dades observades $\boldsymbol{x}$. En el cas que $\Theta \subset \mathbb{R}$, els conjunts que s'acostumen a utilitzar per fer inferència de θ són els intervals.

Definició 2.24 *Un* **estimador per intervals** *d'un paràmetre $\theta \in \Theta \subset \mathbb{R}$ és qualsevol parell de funcions reals $L(x_1, \ldots, x_n)$ i $U(x_1, \ldots, x_n)$, tals que $L(\boldsymbol{x}) \leq U(\boldsymbol{x})$. Si s'observa el valor $X = \boldsymbol{x}$, mitjançant aquest estimador es fa la inferència "$L(\boldsymbol{x}) \leq \theta \leq U(\boldsymbol{x})$". A l'interval aleatori $[\,L(\boldsymbol{X}), U(\boldsymbol{X})\,]$ se l'anomena estimador per intervals de θ (o* **interval estimador** *de θ), mentre que al valor que ha pres en la mostra observada $[\,L(\boldsymbol{x}), U(\boldsymbol{x})\,]$ se l'anomena estimació per intervals de θ (o interval estimació de θ).*

Exemple 2.34 ____

Sigui X_1, X_2, X_3, X_4 una mostra de mida 4 de $X \sim N(\mu, 1)$. Un estimador per intervals de μ és $[\,\overline{X}-1, \overline{X}+1\,]$. Per a cada mostra observada x_1, x_2, x_3, x_4, l'estimació per intervals de μ és $[\,\overline{x}-1, \overline{x}+1\,]$.

Observem que si s'estima un paràmetre θ per mitjà d'un interval, la inferència que estem fent és menys precisa que no pas si s'estima amb un estimador puntual: ara ens limitem a afirmar que el paràmetre està en un cert conjunt, mentre que abans donàvem un valor concret com a estimació seva. Així, doncs, donat que es perd en precisió, és normal preguntar-se en aquest punt què es guanya en estimar un paràmetre θ per mitjà d'un interval. La resposta és que es guanya en CONFIANÇA: en general, la probabilitat que un estimador sigui exactament igual al paràmetre que es vol estimar és 0, mentre que la probabilitat que un estimador per intervals cobreixi al paràmetre serà positiva.

Tornant a l'exemple anterior, si s'estima μ per $\overline{X}$, es té que

$$\mathbf{P}(\overline{X} = \mu) = 0, \text{ ja que } \overline{X} \sim N\big(\mu, \frac{1}{4}\big)$$

No obstant això,

$$\mathbf{P}\big(\mu \in [\,\overline{X}-1, \overline{X}+1\,]\big) = \mathbf{P}(\,\overline{X}-1 \le \mu \le \overline{X}+1\,) = \mathbf{P}(\,-1 \le \overline{X}-\mu \le 1\,) =$$
$$\mathbf{P}\left(\,-2 \le \frac{\overline{X}-\mu}{1/\sqrt{4}} \le 2\,\right) = 0.9544$$

Així, doncs, podem afirmar que sacrificant precisió hem pogut augmentar la confiança que tenim que sigui correcta l'afirmació feta en la inferència.

____

Definició 2.25 *Donat un estimador per intervals* $[\,L(\boldsymbol{X}), U(\boldsymbol{X})\,]$ *per un paràmetre* θ*, la* **probabilitat de cobertura** *d'aquest estimador per intervals es defineix com la probabilitat que l'interval aleatori cobreixi al veritable valor del paràmetre* θ*, és a dir,*

$$\mathbf{P}_\theta\big(\,\theta \in [\,L(\boldsymbol{X}), U(\boldsymbol{X})\,]\,\big)$$

Observem que aquesta probabilitat pot variar amb θ.

Definició 2.26 *S'anomena* **coeficient de confiança** *de l'interval* $[\,L(\boldsymbol{X}), U(\boldsymbol{X})\,]$ *a l'ínfim de les probabilitats de cobertura:*

$$\inf_{\theta\in\Theta} \mathbf{P}_\theta\big(\,\theta \in [\,L(\boldsymbol{X}), U(\boldsymbol{X})\,]\,\big)$$

Definició 2.27 L'interval de confiança *és el nom que rep usualment un estimador per intervals junt amb el seu coeficient de confiança. A més de $C(\boldsymbol{x})$, es fa servir també la notació $IC_{1-\alpha}(\theta)$ per referir-se a un interval de confiança $1-\alpha$ per θ.*

Notem que en les expressions del tipus $\mathbf{P}_\theta(\,\theta \in [L(\boldsymbol{X}), U(\boldsymbol{X})]\,)$, el valor del paràmetre està fixat, i el que són variables aleatòries són els extrems de l'interval:

$$\mathbf{P}_\theta(\,\theta \in [\,L(\boldsymbol{X}), U(\boldsymbol{X})\,]\,) = \mathbf{P}_\theta(\,\{\,L(\boldsymbol{X}) \le \theta\,\} \cap \{\,U(\boldsymbol{X}) \ge \theta\,\}\,)$$

Exemple 2.35

Siguin $X_1, \ldots, X_n$ una m.a.s. d'una $X \sim U(0, \theta), Y = max\{ X_1, \ldots, X_n \}$. Considerem els intervals

$$\begin{array}{rcll} I_1 & = & [\, a\,Y, b\,Y\,] & 1 \leq a < b \\ I_2 & = & [\, Y + c, Y + d\,] & 0 \leq c < d \end{array}$$

Aleshores tenim que

$$\mathbf{P}(\, \theta \in I_1\,) = \mathbf{P}\left(\frac{1}{b} \leq \frac{Y}{\theta} \leq \frac{1}{a} \right) = \left(\frac{1}{a}\right)^n - \left(\frac{1}{b}\right)^n$$

i per tant la probabilitat de cobertura és independent de θ. Per altra banda, també podem observar el següent:

$$\mathbf{P}(\, \theta \in I_2\,) = \mathbf{P}\left(1 - \frac{d}{\theta} \leq \frac{Y}{\theta} \leq 1 - \frac{c}{\theta} \right) = \left(1 - \frac{c}{\theta}\right)^n - \left(1 - \frac{d}{\theta}\right)^n$$

que sí depèn de θ.

Exemple 2.36

Sigui $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma^2)$, amb σ conegut. Volem un interval de confiança pel paràmetre μ.

Donat que

$$Z = \frac{\overline{X}_n - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

densitat que hem representat a la figura 2.3.

Podem determinar el valor $z_{\alpha/2}$ de manera que

$$\mathbf{P}\Big(-z_{\alpha/2} < \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\Big) = p = 1-\alpha$$

per una probabilitat p prèviament establerta.

Anem a manipular ara l'expressió anterior:

$$\mathbf{P}\Big(-z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} < \overline{X}_n - \mu < z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\Big) = p = 1-\alpha$$

és a dir

$$\mathbf{P}\Big(\overline{X}_n - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\Big) = p = 1-\alpha$$

cosa que permet assegurar que, amb una probabilitat p, l'interval

$$\Big(\overline{x}_n - z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\,,\,\overline{x}_n + z_{\alpha/2}\,\frac{\sigma}{\sqrt{n}}\Big)$$

obtingut a partir d'un valor qualsevol de $\overline{x}_n$, conté a la mitjana μ.

També s'escriu:

$$IC_{1-\alpha}(\mu) = \overline{x}_n \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

Si per exemple sabem que $\sigma = 1.5$, $n = 12$ i que $\overline{x}_n = 3.6$, aleshores l'interval de confiança al 95% pel paràmetre μ serà:

$$\Big(3.6 - 1.96\,\frac{1.5}{\sqrt{12}}\,,\,3.6 + 1.96\,\frac{1.5}{\sqrt{12}}\Big) = (2.751\,,\,4.449)$$

donat que $z_{\alpha/2} = z_{0.05/2} = 1.96$.

Quantitats pivotals

Un dels mètodes més comuns de construcció d'intervals de confiança és l'ús de les quantitats pivotals.

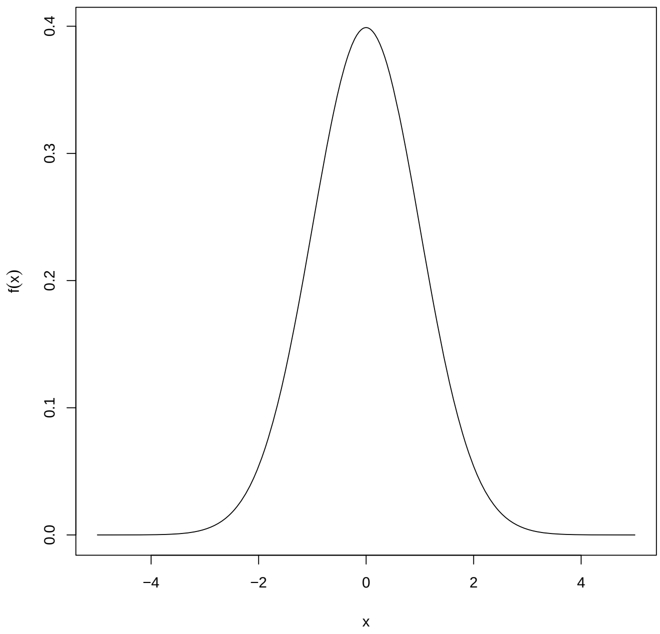

Figura 2.3: Distribució normal tipificada

Definició 2.28 *Sigui* $\boldsymbol{X} = (X_1, \ldots, X_n)$ *una m.a.s. de* $X \sim F(x;\theta)$. *Una funció* $Q(\boldsymbol{X}, \theta)$ *de la mostra i del paràmetre és una* **quantitat pivotal** *si la distribució de probabilitat de* $Q(\boldsymbol{X}, \theta)$ *no depèn del paràmetre* θ, *és a dir,* $Q(\boldsymbol{X}, \theta)$ *té la mateixa distribució per a qualsevol valor de* θ.

Exemples:

- Si $X \sim U(0, \theta)$ i $Y = \frac{X}{\theta}$, llavors $Y \sim U(0,1)$ és una quantitat pivotal.
- Si $X \sim N(\mu, \sigma^2)$, llavors $Z = \frac{X-\mu}{\sigma} \sim N(0,1)$ és una quantitat pivolat.

Donada una quantitat pivotal $Q(\boldsymbol{X}, \theta)$, per a qualsevol conjunt A de l'espai imatge de Q es té que $\mathbf{P}_\theta(\, Q(\boldsymbol{X}, \theta) \in A\,)$ no depèn de θ. Per tant, si es tria un conjunt A_α tal que

$$\mathbf{P}_\theta(\, Q(\boldsymbol{X}, \theta) \in A\,) = 1 - \alpha, \quad \text{per a tot } \theta$$

i s'observa la mostra $\boldsymbol{X} = \boldsymbol{x}$, aleshores el conjunt

$$C(\boldsymbol{x}) = \{\, \theta : Q(\boldsymbol{x}, \theta) \in A\,\}$$

és un conjunt de confiança $1 - \alpha$ per a θ.

Exemple 2.37

Sigui $\boldsymbol{X} \sim exp(\lambda)$, $E(X) = \frac{1}{\lambda}$, i que el que és vol és donar un interval de confiança per aquest paràmetre. Prenem una mostra de mida n de X. Es pot construir un interval de confiança basat en

$$V = \frac{\sum_{i=1}^{n} X_i}{\lambda}$$

que té una distribució $\Gamma(n, 1)$ per a qualsevol valor del paràmetre λ, amb la qual cosa podem afirmar que V és una quantitat pivotal i l'interval de confiança allà construït és un exemple d'interval basat en una quantitat pivotal.

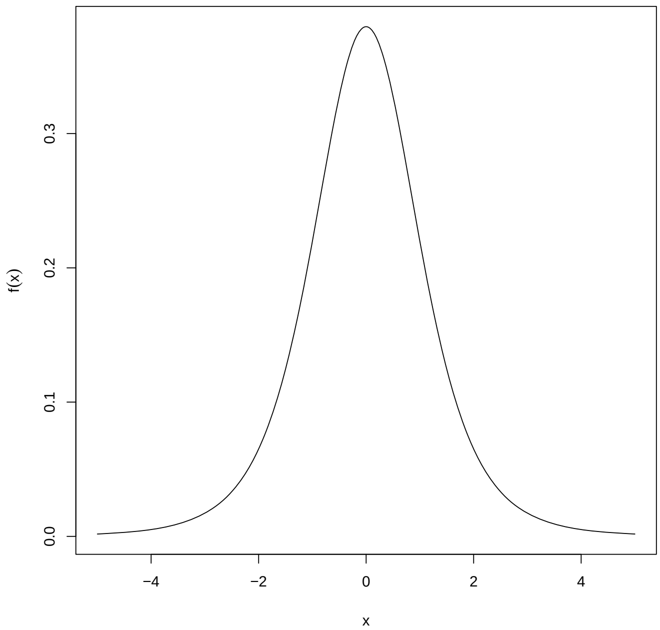

Figura 2.4: Densitat de distribució de la t d'Student amb 5 graus de llibertat

Exemple 2.38

Interval de confiança per la mitjana d'una població normal de variància desconeguda.

Prenem una m.a.s. $X_1, \ldots, X_n$ on $X_i \sim N(\mu, \sigma^2)$.

Per un resultat del tema 1 sabem que

$$T = \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} \sim t_{n-1}$$

A la figura 2.4 hem representat la funció de densitat de la distribució que estem tractant en aquest exemple.

Degut a la simetria de la distribució podem escriure

$$\mathbf{P}\big(-t_{n-1;1-\alpha/2} < T < t_{n-1;1-\alpha/2} \big) = 1 - \alpha = p$$

també per una probabilitat $p` establerta d'un principi.

Utilitzant el resultat que abans hem destacat, tenim

$$\mathbf{P}\left(-t_{n-1;1-\alpha/2} < \frac{\overline{X}_n - \mu}{S_n/\sqrt{n}} < t_{n-1;1-\alpha/2} \right) = 1 - \alpha = p$$

i manipulant aquestes desigualtats,

$$\mathbf{P}\left(-t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n}} < \overline{X}_n - \mu < t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n}} \right) = 1 - \alpha = p$$

$$\mathbf{P}\left(\overline{X}_n - t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n}} < \mu < \overline{X}_n + t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n}} \right) = 1 - \alpha = p$$

Així, doncs, l'interval pel paràmetre μ és

$$\left(\overline{x}_n - t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n}} \, , \, \overline{x}_n + t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n}} \right)$$

Anem a continuació a tractar un cas concret: Per determinar el contingut de bismut en un aliatge s'analitzen 10 mostres, amb els següents resultats (en les unitats adequades):

$$19.7, 21.4, 12.5, 13.8, 13.5, 20.4, 18.2, 16.6, 15.2, 16.3$$

Suposem que volem un interval de confiança al 95% pel contingut mitjà de bismut de l'aliatge. Per aquestes dades es té que

$$\overline{X}_n = \frac{\sum_{i=1}^{n} X_i}{n} = \frac{\sum\limits_{i=1}^{10}}{10} = \frac{167.6}{10} = 16.76$$

Calculem ara S^2. Sabem també per un resultat del tema anterior

$$S^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n \, \overline{X}_n^2 \right)$$

Aplicat al nostre cas particular

$$S^2 = \frac{1}{9} \Big(\sum_{i=1}^{10} X_i^2 - 10 \, (16.76)^2 \Big) = \frac{85.704}{9} = 9.52$$

i per tant tenim que $S = 3.086$.

Si volem un interval del 95%, això vol dir $p = 0.95$ i per tant $1 - \alpha = 0.95 \Longrightarrow \alpha = 0.05$.

Hem vist que l'interval de confiança en aquest cas era:

$$\left(\overline{x}_n - t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n}} \, , \, \overline{x}_n + t_{n-1;1-\alpha/2} \frac{S_n}{\sqrt{n}} \right)$$

que traduït al nostre exemple serà:

$$\left(16.76 - t_{9;1-0.05/2} \frac{3.086}{\sqrt{10}} \, , \, 16.76 + t_{9;1-0.05/2} \frac{3.086}{\sqrt{10}} \right)$$

i si mirem la taula de la distribució de la t d'Student podem afirmar que

$$t_{9;0.975} = 2.2622$$

amb la qual cosa tenim que l'interval de confiança al 95% serà:

$$\big(16.76 - 2.208 \, , \, 16.76 + 2.208 \big) = \big(14.552 \, , \, 18.968 \big)$$

Si ara en volguéssim un al 99% en aquest cas $\alpha = 1 - p = 1 - 0'99 = 0'01$ i l'interval quedaria de la forma:

$$\left(16.76 - t_{9;1-0.01/2} \frac{3.086}{\sqrt{10}} \, , \, 16.76 + t_{9;1-0.01/2} \frac{3.086}{\sqrt{10}} \right) =$$

$$\big(16.76 - t_{9;0.995} \, 0.976 \, , \, 16.76 + t_{9;0.995} \, 0.976 \big)$$

i per les taules de la distribució $t_{9;0.995} = 3.2498$ i per tant s'obté l'interval

$$\big(13.588 \, , \, 19.932 \big)$$

Notem que l'interval s'ha fet més gran, ja que estem exigint més precisió al resultat.

Exemple 2.39

Interval de confiança per la variància d'una població normal.

Si considerem com a l'exemple anterior una m.a.s. $X_1, \ldots, X_n$ tal que $X_i \sim N(\mu, \sigma^2)$, sabem que

$$W = \frac{(n-1)\,S_n^2}{\sigma^2} \sim \mathcal{X}^2_{n-1}$$

A la figura 2.5 hem dibuixat la funció de distribució de la χ^2 en aquest cas amb 10 graus de llibertat, que serà el cas concret pel qual trobarem l'interval de confiança.

Siguin $\mathcal{X}^2_{n-1;\alpha/2}$ i $\mathcal{X}^2_{n-1;1-\alpha/2}$ tals que

$$\mathbf{P}\big(W < \mathcal{X}^2_{n-1;\alpha/2}\big) = \frac{\alpha}{2}$$

$$\mathbf{P}\big(W < \mathcal{X}^2_{n-1;1-\alpha/2}\big) = 1 - \frac{\alpha}{2}$$

Aleshores la probabilitat en el rang de valors centrals

$$\big(\mathcal{X}^2_{n-1;\alpha/2}\,,\,\mathcal{X}^2_{n-1;1-\alpha/2}\big)$$

vindrà donada per:

$$\mathbf{P}\Big(\mathcal{X}^2_{n-1;\alpha/2} < \frac{(n-1)\,S_n^2}{\sigma^2} < \mathcal{X}^2_{n-1;1-\alpha/2}\Big) = 1-\alpha$$

és a dir,

$$\mathbf{P}\Big(\frac{1}{\mathcal{X}^2_{n-1;1-\alpha/2}} < \frac{\sigma^2}{(n-1)\,S_n^2} < \frac{1}{\mathcal{X}^2_{n-1;\alpha/2}}\Big) = 1-\alpha$$

o, el que és el mateix

$$\mathbf{P}\Big(\frac{(n-1)\,S_n^2}{\mathcal{X}^2_{n-1;1-\alpha/2}} < \sigma^2 < \frac{(n-1)\,S_n^2}{\mathcal{X}^2_{n-1;\alpha/2}}\Big) = 1-\alpha$$

Obtenim, doncs, l'interval de confiança per σ^2:

$$IC_{1-\alpha}(\sigma^2) = \left(\frac{(n-1)\,S_n^2}{\mathcal{X}^2_{n-1;1-\alpha/2}}\,,\,\frac{(n-1)\,S_n^2}{\mathcal{X}^2_{n-1;\alpha/2}}\right)$$

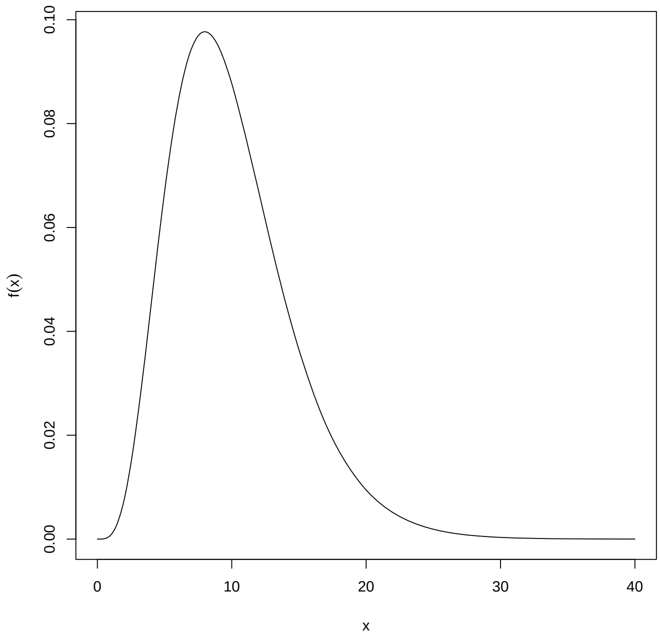

Figura 2.5: Densitat de distribució d'una χ^2_{10} amb 10 graus de llibertat

Anem a fer ara un cas concret. Si ens posem en la mateixa situació que quan volíem un interval pel paràmetre μ, teníem que $S_n^2 = 9.52$, $n = 10$. Si volem un interval de confiança al 95%, llavors $1 - \alpha = 0.95 \Longrightarrow \alpha = 0.05$.

Si ens mirem la taula de la distribució, obtenim els valors:

$$\mathcal{X}^2_{n-1;\alpha/2} = \mathcal{X}^2_{9;0.025} = 2.70$$

$$\mathcal{X}^2_{n-1;1-\alpha/2} = \mathcal{X}^2_{9;0.975} = 19.02$$

i per tant l'interval serà:

$$\left(\frac{9 \cdot 9.52}{19.02} \, , \, \frac{9 \cdot 9.52}{2.7} \right) = \left(4.505 \, , \, 31.73 \right)$$

Nota: l'interval no és simètric respecte a s^2.

Capítol 3

PROVES D'HIPÒTESIS

Una **hipòtesi estadística** és una conjectura o una afirmació sobre la distribució d'una o més variables aleatòries. Una **prova d'hipòtesi** (o un **test d'hipòtesi**) és un procediment per decidir si s'accepta o es rebutja una hipòtesi.

3.1 Introducció. Hipòtesis simples i compostes

Definició 3.1 *Considerem un problema estadístic on el valor d'un paràmetre* θ *és desconegut però sabem que* $\theta \in \Omega$.

Suposem que $\Omega = \Omega_0 \cup \Omega_1$ *amb* $\Omega_0 \cap \Omega_1 = \emptyset$ *i volem decidir si* $\theta \in \Omega_0$ *o bé si* $\theta \in \Omega_1$.

Denotem per $H_0 : \theta \in \Omega_0$ *i per* $H_1 : \theta \in \Omega_1$. *Com que* $\theta \in \Omega$, $\Omega = \Omega_0 \cup \Omega_1$ *i* $\Omega_0 \cap \Omega_1 = \emptyset$, *alguna de les dues hipòtesis ha de ser la verdadera.*

Un problema d'aquesta mena on hi ha dues possibles decisions a prendre s'anomena **problema de proves d'hipòtesi.**

Si escollim la hipòtesi errònia, haurem de patir una pèrdua o pagar un cost. En general, però, podrem prendre observacions i segons aquests valors observats decidirem. Aquests valors ens haurien de donar informació sobre el paràmetre θ en la qual basar la nostra decisió. Un procediment per a decidir si acceptem H_0 o H_1 s'anomena un **procediment estadístic.**

Definició 3.2 *En general les dues hipòtesis* H_0 *i* H_1 *es tracten i es consideren de forma diferent. Anomenarem a* H_0 **hipòtesi nul·la** *i a* H_1 **hipòtesi alternativa**.

Usualment disposem d'una mostra $X_1, \ldots, X_n$ d'una variable aleatòria amb distribució F i funció de densitat f. Sobre la distribució de X es realitzen dues afirmacions sobre les quals hem de decidir: la hipòtesi nul·la H_0 és més conservadora en el sentit que no es rebutjarà a menys que l'evidència per a fer-ho sigui molt clara. Aquesta hipòtesi sol establir un model senzill per la distribució de X (per exemple, si F pertany a una família paramètrica, H_0 fixa el valor del paràmetre) o bé proposa com a distribució de X aquella que de manera comuna és acceptada com una bona descripció del fenomen que modelitza X. Per altra banda, la hipòtesi alternativa H_1 especifica el tipus de llunyania de la hipòtesi nul·la que podria presentar la distribució de X. Si un investigador considera que un fenomen aleatori no ha estat adequadament modelitzat fins aquell moment i creu que té una explicació més satisfactòria, proposarà aquesta com a hipòtesi alternativa i el model vigent com a hipòtesi nul·la. Només si hi ha una evidència mostral suficient per a rebutjar la hipòtesi nul·la, serà acceptada la hipòtesi alternativa.

Es poden distingir tres tipus de proves d'hipòtesis:

A. Suposem que F i f pertanyen a una certa família indexada per un paràmetre $\theta \in \Theta$ i ens plantegem la prova:

$$\begin{cases} H_0 : \theta \in \Omega_0 \\ H_1 : \theta \in \Omega_1 \end{cases}$$

B. Proves d'ajustament (*goodness-of-fit tests*, en anglès):

$$\begin{cases} H_0 : f = f_0 \\ H_1 : f \neq f_0 \end{cases}$$

C. Per a dues distribucions f_0 i f_1 que no necessàriament pertanyen a la mateixa família paramètrica, es planteja la prova:

$$\begin{cases} H_0 : f = f_0 \\ H_1 : f = f_1 \end{cases}$$

Com que només hi ha dues decisions a prendre és equivalent

$$\text{acceptar } H_0 \Longleftrightarrow \text{rebutjar } H_1$$
$$\text{acceptar } H_1 \Longleftrightarrow \text{rebutjar } H_0$$

Definició 3.3 *Al realitzar una prova d'hipòtesi es poden cometre errors de dos tipus:*

error tipus I: *Rebutjar la hipòtesi nul·la H_0 si és certa.*

error tipus II: *Acceptar H_0 si aquesta és falsa.*

L'error de tipus I es considera més greu que no pas el de tipus II, donat que la hipòtesi nul·la és sempre la més conservadora.

El següent esquema il·lustra les diverses situacions:

		Decisió	
		Acceptar H_0	Rebutjar H_0
Realitat	H_0 certa	Decisió correcta	Error de tipus I (α)
	H_0 falsa	Error de tipus II (β)	Decisió correcta

Sigui $X_1, \ldots, X_n$ una mostra d'una distribució que depèn de θ. Sigui S l'espai mostral associat a $\underset{\sim}{X} = (X_1, \ldots, X_n)$.

Tenim com abans:

$$\begin{cases} H_0 : \theta \in \Omega_0 \\ H_1 : \theta \in \Omega_1 \end{cases}$$

Es tracta d'especificar un procediment mitjançant una partició de l'espai mostral en dos subconjunts. Un subconjunt de S conté els valors de $\underset{\sim}{X} = (X_1, \ldots, X_n)$ pels quals acceptarem H_0 i l'altre conté els valors de $\underset{\sim}{X}$ pels quals rebutjarem H_0 i per tant acceptarem H_1. És a dir, $S = C \cup A$ i $C \cap A = \emptyset$.

Definició 3.4 *El subconjunt de S pel qual H_0 és rebutjat s'anomena la* **regió crítica** *o la* **regió de rebuig** *C. El complementari d'aquesta dóna la* **regió d'acceptació** *A, o subconjunt de S pel qual acceptarem H_0.*

Les característiques d'una prova es poden descriure especificant per a cada valor $\theta \in \Omega$:

$$\begin{aligned} \Pi(\theta) &= \textit{Probabilitat que el procediment condueixi al rebutjament de } H_0 \\ &= \textit{Probabilitat } \{\textit{ Rebutjar } H_0 \} = \textit{Probabilitat } \{ \underset{\sim}{X} \in \textit{Regió crítica } C \} \end{aligned}$$

que s'anomena **funció de potència**.

Ens preguntem ara quina hauria de ser la funció de potència ideal. Doncs hauria de ser aquella tal que:

$$\Pi(\theta) = 0 \ \forall \ \theta \in \Omega_0$$

$$\Pi(\theta) = 1 \ \forall \ \theta \in \Omega_1$$

Si aquesta funció existís, aleshores prendríem sempre la decisió correcta amb probabilitat 1. Això, però, no passa gairebé mai.

Amb tot això, doncs, les probabilitats de rebutjar H_0 són:

$$\Pi(\theta) = \mathbf{P}_\theta\big(\underset{\sim}{X} \in C\big) = \begin{cases} \text{Probabilitat (Error de tipus I)} & \text{si} \quad \theta \in \Omega_0 \\ 1 - \text{Probabilitat (Error de tipus II)} & \text{si} \quad \theta \in \Omega_1 \end{cases}$$

Per una grandària mostral prefixada és impossible controlar en el sentit de minimitzar tots dos tipus d'error. Fixem-nos que si volem fer petita la probabilitat de l'error de tipus I, $\alpha = \mathbf{P}_\theta(\underset{\sim}{X} \in C)$, haurem de reduir els punts de la regió crítica C, però aleshores el conjunt $\overline{C}$, complementari de C, serà més gran, de manera que $\beta = \mathbf{P}_\theta(\underset{\sim}{X} \in \overline{C})$ augmentarà.

Donat que l'error de tipus I s'ha considerat més greu que no pas l'error de tipus II, els procediments clàssics a l'hora de fer una prova d'hipòtesi, el que fan és considerar únicament proves que garanteixin que la probabilitat de cometre un error de tipus I sigui inferior a un valor α donat (per exemple, $\alpha = 0.1, 0.05$ o 0.01) i buscar entre totes aquella que fa mínima la probabilitat de cometre un error de tipus II.

Si el valor obtingut per l'error de tipus II és inacceptablement gran, es poden prendre dues mesures per reduir-lo:

- Augmentar $\mathbf{P}_\theta(\underset{\sim}{X} \in C) = \alpha$.
- Augmentar la grandària de la mostra.

Definició 3.5 *Definim la* **grandària** *(o* **talla***)* α *d'una prova com:*

$$\alpha = \sup_{\theta \in \Omega_0} \Pi(\theta)$$

és a dir, és la màxima probabilitat, entre tots els valors de θ *que satisfan* H_0*, de fer una decisió incorrecta.*

Definició 3.6 *Per a qualsevol $\theta \in \Omega_0$ si es decideix rebutjar la hipòtesi nul·la H_0 estem prenent una decisió incorrecta. Aleshores, si $\theta \in \Omega_0$, $\Pi(\theta)$ és la probabilitat de prendre una decisió incorrecta. En molts problemes l'estadístic especificarà una fita superior α_0 $(0 < \alpha_0 < 1)$. A aquesta fita superior se l'anomena* **nivell de significació de la prova**. *És, doncs, aquell valor α_0 tal que*

$$\sup_{\theta \in \Omega_0} \Pi(\theta) \leq \alpha_0$$

El conjunt de proves de nivell de significació α conté a les proves amb talla α.

Fixat un nivell de significació α_0 només considerarem proves amb grandària $\alpha \leq \alpha_0$.

Definició 3.7 *Una prova que minimitzi $\beta = \mathbf{P}_\theta(\, \underset{\sim}{X} \in \overline{C}\,)$ per $\alpha = \mathbf{P}_\theta(\, \underset{\sim}{X} \in C\,)$ fixat s'anomena una* **prova més potent de mida** *α.*

Exemple 3.1 ____

Sigui $X_1, \ldots, X_n$ una mostra aleatòria simple de $X \sim U[0, \theta]$, on $\theta > 0$. Per tant, $\theta \in \Omega = \mathbb{R}^+$. Sigui $\underset{\sim}{X} = (X_1, \ldots, X_n) \in S = (0, \theta)^n$. Volem contrastar

$$\begin{cases} H_0 : 3 \leq \theta \leq 4 \\ H_1 : \theta < 3 \;\text{ o }\; \theta > 4 \end{cases}$$

Com sabem, l'estimador de la màxima versemblança de θ és $\widehat{\theta} = Y_n = \max\{\, X_1, \ldots, X_n \,\}$. Tot i que $Y_n < \theta$, si n és suficientment gran, hi ha una probabilitat força gran que Y_n estigui prop de θ i per tant acceptarem H_0 si $2.9 \leq Y_n \leq 4$. Així, doncs, tenim que en aquest exemple la regió crítica serà:

$$C = \{\, (x_1, \ldots, x_n) \in (0, \theta)^n : Y_n < 2.9 \;\text{ o }\; Y_n > 4 \,\}$$

és a dir,

$$\Pi(\theta) = \mathbf{P}\{Y_n < 2.9 \,|\, \theta\,\} + \mathbf{P}\{\, Y_n > 4 \,|\, \theta\,\}$$

Si

$$\begin{array}{ccl} \theta \leq 2.9 & \Longrightarrow & \Pi(\theta) = 1 + 0 = 1 \\ 2.9 < \theta < 4 & \Longrightarrow & \Pi(\theta) = (\,2.9/\theta\,)^n + 0 = (\,2.9/\theta\,)^n \\ \theta > 4 & \Longrightarrow & \Pi(\theta) = (\,2.9/\theta\,)^n + (\,1 - (\,4/\theta\,)^n\,) \end{array}$$

Recordem que $\mathbf{P}(Y_n < t) = (t/\theta)^n$.

La grandària de la prova és

$$\alpha = \sup_{3\leq\theta\leq 4} \Pi(\theta) = \Pi(3) = \left(\frac{2.9}{3}\right)^n = 0.100 \text{ en el cas } n = 68$$

Definició 3.8 *Si l'espai Ω_0 (o Ω_1) només conté un valor de θ, direm que tenim una* **hipòtesi simple**. *Si conté més d'un valor, direm que tenim una* **hipòtesi composta**.

Sota una hipòtesi simple la distribució de les dades queda completament especificada mentre que sota una hipòtesi composta aquesta distribució pertany a una certa classe. Quan la hipòtesi nul·la és simple la grandària de la prova és $\alpha = \Pi(\theta_0)$ on $H_0 : \theta = \theta_0$.

Exemple 3.2

Considerem una m.a.s. $X_1, \ldots, X_n$ d'una variable aleatòria $X \sim N(\mu, \sigma^2)$ i suposem que σ^2 és conegut.

Prenem $\Omega = \{3, 4\}$

a) Si considerem la prova

$$\begin{cases} H_0 : \mu = \mu_0 = 3 \\ H_1 : \mu = \mu_1 = 4 \end{cases}$$

tenim dues hipòtesis simples, amb $\Omega_0 = \{3\}$ i $\Omega_1 = \{4\}$.

b) Si ara considerem

$$\begin{cases} H_0 : \mu \leq 3 \\ H_1 : \mu > 3 \end{cases}$$

es tracta en aquest cas de dues hipòtesis compostes, amb $\Omega_0 = (-\infty, 3]$ i $\Omega_1 = (3, \infty)$.

a) Sota H_0 tenim $X_1, \ldots, X_n \sim N(3, \sigma^2)$. Sota H_1 $X_1, \ldots, X_n \sim N(4, \sigma^2)$.

b) Sota $H_0 : X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ amb $\mu \leq 3$, és a dir, la distribució no queda especificada.

Exemple 3.3

El Z-test.

Sigui $X_1, \ldots, X_n$ una mostra aleatòria simple de $X \sim N(\mu, \sigma^2)$ amb σ^2 conegut. Es vol contrastar

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu = \mu_1 \end{cases} \quad \text{on } \mu_1 > \mu_0$$

Veiem que la nostra intuïció ens duria a rebutjar H_0 per a valors grans de la mitjana mostral x_n.

Ens cal trobar B tal que

$$\mathbf{P}(\,\overline{X} \geq B \,|\, H_0\,) = \alpha$$

Tindrem

$$P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \geq \frac{B-\mu}{\sigma/\sqrt{n}}\right) = P\left(Z \geq \frac{B-\mu}{\sigma/\sqrt{n}}\right) = \alpha,$$

i $z_\alpha = \frac{B-\mu}{\sigma/\sqrt{n}}$. Com que sota la hipòtesi H_0, $\overline{X} \sim N(\mu_0, \sigma^2/n)$ tindrem que $B = \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$.

En aquest exemple suposem $\mu_0 = 5$, $\mu_1 = 6$, $\sigma^2 = 1$ i $\alpha = 0.05$. Observem la següent mostra:

$$\underset{\sim}{x} = (\,6.1\,,\,5.5\,,\,5.9\,,\,6.3\,)$$

Per tant, tenim que $\overline{x}_n = \overline{x}_4 = 5.95$ i com que $\alpha = 0.05$, $z_\alpha = 1.645$.

Rebutjarem $H_0 : \mu = 5$ en favor de $H_1 : \mu = 6$ si

$$\overline{X}_n \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

o, equivalentment, si

$$\frac{\overline{X}_n - \mu_0}{\sigma/\sqrt{n}} \geq 1.645$$

En el nostre cas

$$\frac{5.95-5}{1/\sqrt{4}} = 1.9 \geq 1.645$$

i per tant rebutgem H_0.

Lema 3.1 Lema de Neyman-Pearson

Sigui $X_1, \ldots, X_n$ una mostra aleatòria simple d'una distribució amb funció de densitat $f(x\,;\theta)$. Es desitja contrastar $H_0 : \theta = \theta_0$ enfront de $H_1 : \theta = \theta_1$. Si $L(\theta \,|\, \underset{\sim}{x})$ és la funció de versemblança, el millor test de mida α té una regió crítica de la forma

$$C = \left\{ \underset{\sim}{x} \in \mathcal{X}^n \; : \; \frac{L(\theta_1 \,|\, \underset{\sim}{x})}{L(\theta_0 \,|\, \underset{\sim}{x})} \geq A \right\}$$

per $A > 0$.

Exemple 3.4 ---

Suposem que la proporció p d'unitats defectuoses en un lot d'un producte determinat és desconeguda. Suposem que escollim a l'atzar una mostra aleatòria simple $Y_1, \ldots, Y_n$ de $Y \sim Bern(p)$, i el que es vol contrastar és:

$$\begin{cases} H_0 : p = p_0 \\ H_1 : p = p_1 \end{cases}$$

on $p_1 > p_0$. Ja hem vist que la funció de probabilitat conjunta té una raó de versemblança monòtona en l'estadístic $X = \sum_{i=1}^{n} Y_i \sim B(n, p)$. La versemblança de la mostra és

$$L(p \,|\, x) = \binom{n}{x} p^x \, (1-p)^{n-x}$$

Calculem el logaritme del quocient de versemblances:

$$log\left(\frac{L(p_1 \,|\, x)}{L(p_0 \,|\, x)} \right) = log\left(\frac{p_1^x \, (1-p_1)^{n-x}}{p_0^x \, (1-p_0)^{n-x}} \right) =$$

$$x \, log\left(\frac{p_1}{p_0} \right) + (n-x) \, log\left(\frac{1-p_1}{1-p_0} \right)$$

Així, doncs, la regió crítica del test de Neyman-Pearson serà:

$$C = \left\{ x : x\, log\left(\frac{p_1}{p_0}\right) + (n-x)\, log\left(\frac{1-p_1}{1-p_0}\right) \geq logA \right\} =$$

$$\left\{ x : x \left[log\left(\frac{p_1}{p_0}\right) - log\left(\frac{1-p_1}{1-p_0}\right)\right] \geq -n\, log\left(\frac{1-p_1}{1-p_0}\right) + logA \right\} =$$

$$\left\{ x : x \geq B = \frac{log(A) - n\, log(1-p_1/1-p_0)}{log\left(\frac{p_1\,(1-p_0)}{p_0\,(1-p_1)}\right)} \right\}$$

Per tal de determinar el valor de B, usarem el fet que la distribució de X és coneguda sota H_0 i que es vol definir una prova de mida α.

Suposem, per exemple, que $n = 10$, $p_0 = 0.5$ i $p_1 = 0.8$. Per a diferents valors de B s'obtenen proves de diferents mides α i potències β.

B	α	β
0	1.0000	0.0000
1	0.9990	0.0000
2	0.9893	0.0001
3	0.9453	0.0009
4	0.8281	0.0064
5	0.6230	0.0328
6	0.3770	0.1209
7	0.1719	0.3222
8	0.0547	0.6242
9	0.0107	0.8926
10	0.0010	1.0000

Veiem que no és possible construir una prova de mida α per a tots els valors de $\alpha \in [0, 1]$. Si per exemple volem $\alpha = 0.05$ tenim tres opcions:

- Acceptar $\alpha = 0.055$ com prou proper a 0.05 i rebutjar H_0 si $X \geq 8$.
- Pensar que 0.05 és la màxima probabilitat d'error de tipus I acceptable, i per tant, rebutjar H_0 si $X \geq 9$, donant lloc a $\alpha = 0.011$.
- Rebutjar H_0 si $X \geq 9$, acceptar-la si $X \leq 7$, i en cas que $X = 8$ usar un mètode aleatori per assegurar un 0.05.

Tant en l'exemple anterior com també en el del Z-test, hem observat que el millor test depèn de les observacions només mitjançant l'estadístic minimal suficient del paràmetre d'interès. El següent corol·lari estableix això que acabem de dir com a resultat general:

Corol.lari 3.1 *En les hipòtesis del lema de Neyman-Pearson, si T és un estadístic suficient per θ amb funció de densitat $g(t\,|\,\theta)$, el millor test per a la prova*

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

té regió crítica de la forma:

$$C = \left\{ t = T(\underset{\sim}{x}) : \frac{g(t\,|\,\theta_1)}{g(t\,|\,\theta_0)} \geq A \right\}$$

per algun $A \geq 0$.

3.1.1 Establint les conclusions: Regla de decisió i el p-valor

Una manera d'informar dels resultats d'una prova d'hipòtesi és a través de la mida α de la prova usada i la decisió presa sobre si es rebutja o no H_0. Si α és petit la decisió de rebutjar H_0 és força convincent, però si α és gran, la probabilitat de cometre un error de tipus I és gran, la qual cosa resta força al test si la decisió que es pren és la de rebutjar H_0.

Definició 3.9 *Una forma alternativa de presentar els resultats consisteix a donar el* **p-valor** *o* **valor de probabilitat***, definit com el suprem dels valors α pels quals es rebutjaria la hipòtesi nul·la si aquesta es provés a nivell α. El p-valor depèn de les dades mostrals, i es pot interpretar com un nivell de significació, com la probabilitat d'observar una altra mostra que sigui almenys tan poc favorable a la hipòtesi nul·la com la que s'ha observat. A partir del p-valor la* **regla de decisió** *serà rebutjar H_0 si el p-valor és petit i acceptar-la si és gran.*

Per exemple, si apliquem el lema de Neyman-Pearson al p-valor, obtenim que és

$$p = \sup_{\theta \in \Omega_0} \mathbf{P}_\theta \left\{ \frac{L(\theta_1 \,|\, \underset{\sim}{X})}{L(\theta_0 \,|\, \underset{\sim}{X})} \geq \frac{L(\theta_1 \,|\, \underset{\sim}{x})}{L(\theta_0 \,|\, \underset{\sim}{x})} \right\} = \mathbf{P}_{\theta_0} \left\{ \frac{L(\theta_1 \,|\, \underset{\sim}{X})}{L(\theta_0 \,|\, \underset{\sim}{X})} \geq \frac{L(\theta_1 \,|\, \underset{\sim}{x})}{L(\theta_0 \,|\, \underset{\sim}{x})} \right\}$$

Exemple 3.3, pàgina 89. Continuació.

En l'exemple del test Z el p-valor seria

$$\mathbf{P}(\,\overline{X}_n \geq \overline{x}_n \,|\, \mu = \mu_0\,) = \mathbf{P}\left(Z \geq \frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}} \right) = 1 - \Phi\left(\frac{\overline{x}_n - \mu_0}{\sigma/\sqrt{n}} \right)$$

on Φ és la funció de distribució de $Z \sim N(0,1)$.

En el cas concret que $\sigma = 1$, $\mu_0 = 5$, $n = 4$ i $\overline{x} = 5.95$, es té que el p-valor és:

$$p(\overline{x}) = \mathbf{P}\left(\frac{\overline{X}_4 - 5}{1/\sqrt{4}} \geq \frac{5.95 - 5}{1/\sqrt{4}} \right) = \mathbf{P}(Z \geq 1.9) = 1 - 0.9713 = 0.0287$$

i per tant hi ha suficient evidència en contra de H_0 per a rebutjar aquesta hipòtesi.

En general, pels casos en que la regió de rebuig d'una prova de mida α és tal que rebutgem H_0 si i només si $W(\underset{\sim}{X}) \geq c_\alpha$, on $W(\underset{\sim}{X})$ és l'estadístic apropiat pel problema i la constant c_α s'ha escollit per tal que la prova tingués mida α, aleshores, si observem $\underset{\sim}{x}$, el p-valor és

$$p(\underset{\sim}{x}) = \sup_{\theta \in \Omega_0} \mathbf{P}_\theta\,(\,W(\underset{\sim}{X}) \geq W(\underset{\sim}{x})\,)$$

A continuació veurem que hi ha una relació molt estreta entre l'estimació per intervals i les proves d'hipòtesi. En general es pot dir que cada mètode de construcció d'un interval de confiança es correspon a un mètode de prova d'hipòtesi, i viceversa.

La correspondència entre intervals de confiança i proves d'hipòtesis es deu a què ambdós procediments persegueixen la consistència entre els valors observats en la mostra i els paràmetres de la distribució poblacional, encara que des de perspectives diferents. El test d'hipòtesi fixa el paràmetre i es pregunta per quins valors mostrals són consistents amb aquest valor fixat (això és, es busca la regió d'acceptació), mentre que en l'estimació per intervals de confiança es pren com a fixa la mostra observada i es busquen els valors del paràmetre fan aquesta mostra més plausible (és a dir, es busca l'interval de confiança).

El resultat que exposem a continuació prova la relació existent entre proves i intervals.

Teorema 3.1 *Per a cada valor $\theta_0 \in \Theta$ es denota per $A(\theta_0)$ la regió d'acceptació d'una prova de nivell α d'un test que contrasta $H_0 : \theta = \theta_0$. Per a cada $\underset{\sim}{x} \in \mathcal{X}$ definim el conjunt $C(\underset{\sim}{x}) \subseteq \Theta$ com*

$$C(\underset{\sim}{x}) = \{\, \theta_0 \in \Theta : \underset{\sim}{x} \in A(\theta_0) \,\}$$

Aleshores el conjunt aleatori $C(\underset{\sim}{X})$ és una regió de confiança $1-\alpha$ pel paràmetre θ.

Recíprocament, sigui $C(\underset{\sim}{X})$ un estimador per regions de confiança $1-\alpha$ per θ. Per a cada $\theta_0 \in \Theta$ es defineix

$$A(\theta_0) = \{\, \underset{\sim}{x} \in \mathcal{X} : \theta_0 \in C(\underset{\sim}{x}) \,\}$$

Aleshores $A(\theta_0)$ és la regió d'acceptació d'una prova de nivell α d'un test que contrasta $H_0 : \theta = \theta_0$.

Exemple 3.5

Sigui $\underset{\sim}{X} \sim exp(\lambda)$, $\lambda = E(X)$, i que el que és vol és donar un interval de confiança per aquest paràmetre mitjançant la inversió d'un fet estadístic d'un test de mida n. Prenem una mostra de mida n de X, i comencem fent el test

$$\begin{cases} H_0 : \lambda = \lambda_0 \\ H_1 : \lambda \neq \lambda_0 \end{cases}$$

que té per estadístic

$$\lambda(\underset{\sim}{x}) = \frac{\frac{1}{\lambda_0^n}\, e^{-\sum_{i=1}^n x_i/\lambda_0}}{\sup_\lambda \frac{1}{\lambda^n}\, e^{-\sum_{i=1}^n x_i/\lambda}} = \frac{\frac{1}{\lambda_0^n}\, e^{-\sum_{i=1}^n x_i/\lambda_0}}{\frac{1}{(\sum_{i=1}^n x_i/n)^n\, e^{-n}}} =$$

$$\left(\frac{\sum_{i=1}^n x_i}{n\,\lambda_0}\right)^n e^n\, e^{-\sum_{i=1}^n x_i/\lambda_0}$$

Per a un valor fixat λ_0, la regió d'acceptació del test és

$$A(\lambda_0) = \left\{\, \underset{\sim}{x} : \left(\frac{\sum_{i=1}^n x_i}{\lambda_0}\right)^n e^{-\sum_{i=1}^n x_i/\lambda_0} \geq k^* \,\right\},$$

on la constant k^* es tria per tal que el test tingui mida α, o, el que és el mateix, per tal que

$$\mathbf{P}_{\lambda_0}\big(\, \underset{\sim}{X} \in A(\lambda_0)\,\big) = 1 - \alpha$$

Invertint la regió d'acceptació s'obté l'interval de confiança $1-\alpha$:

$$C(\underset{\sim}{x}) = \left\{\, \lambda : \left(\frac{\sum_{i=1}^n x_i}{\lambda}\right)^n e^{-\sum_{i=1}^n x_i/\lambda} \geq k^* \,\right\}$$

La funció $g(v) = v^n\, e^{-v}$ és positiva en tot $\mathbb{R}^+$, val 0 en $v=0$ i tendeix cap a 0 si v tendeix a infinit. A més, té un únic punt crític en $v=n$. Se segueix de tot això que té un únic màxim en $v=n$ i que els conjunts de la forma $\{\, v \geq 0 : g(v) \leq k^* \,\}$, amb $k^* \leq g(n) = n^n\, e^{-n}$, són intervals de la forma $[l,u]$, amb $l \geq n \geq u$ i $g(l) = g(u) = k^*$.

D'aquí es dedueix que $A(\lambda_0)$ és un interval per a qualsevol valor de λ_0 i que els conjunts de confiança $C(\underset{\sim}{x})$ també són intervals per a qualsevol valor de $\sum_{i=1}^n x_i$.

A la pràctica, la manera com es construeix un interval de confiança a partir d'una quantitat pivotal és la que donem a continuació. Suposarem que $Q(\underset{\sim}{x},\theta) \in \mathbb{R}$ i $\theta \in \mathbb{R}$. Per a un valor de α donat, es busquen nombres a i b tals que

$$\mathbf{P}_\theta(\, a \leq Q(\underset{\sim}{X},\theta) \leq b\,) = 1 - \alpha$$

Notem que a i b no depenen de θ per ser Q una quantitat pivotal, i que la tria d'aquests dos nombres no serà única en general.

Per a cada θ_0, el conjunt

$$A(\theta_0) = \{\, \underset{\sim}{x} : a \leq Q(\underset{\sim}{x}, \theta) \leq b \,\}$$

és la regió d'acceptació d'un test de mida α per contrastar $H_0 : \theta = \theta_0$ basat en l'estadístic $T(\underset{\sim}{X}) = Q(\underset{\sim}{X}, \theta_0)$. Si invertim aquest constrast obtindrem el conjunt de confiança $1 - \alpha$ per θ:

$$C(\underset{\sim}{x}) = \{\, \theta : a \leq Q(\underset{\sim}{x}, \theta) \leq b \,\}$$

Si $g_{\underset{\sim}{x}}(\theta) = Q(\underset{\sim}{x}, \theta)$ és una funció monòtona de θ per a cada $\underset{\sim}{x}$ fixat, llavors es té la garantia que $C(\underset{\sim}{x})$ serà un interval. En cas que $g_{\underset{\sim}{x}}(\theta)$ sigui creixent, llavors

$$C(\underset{\sim}{x}) = [\, L(\underset{\sim}{x}, a), U(\underset{\sim}{x}, b) \,]$$

mentre que si es tracta d'una funció decreixent

$$C(\underset{\sim}{x}) = [\, L(\underset{\sim}{x}, b), U(\underset{\sim}{x}, a) \,]$$

Si $g_{\underset{\sim}{x}}(\theta)$ és invertible,

$$C(\underset{\sim}{x}) = \left[\, \min\{\, g^{-1}_{\underset{\sim}{x}}(a), g^{-1}_{\underset{\sim}{x}}(b) \,\}\,,\, \max\{ g^{-1}_{\underset{\sim}{x}}(a), g^{-1}_{\underset{\sim}{x}}(b) \,\} \,\right]$$

Exemple 3.5, pàgina 94. Continuació.

En aquest exemple, $Q(\underset{\sim}{X}, \lambda) = \sum_{i=1}^{n} X_i/\lambda \sim \mathcal{X}^2_n$. Així, podem triar

$$a = \mathcal{X}^2_{n,1-\alpha/2} \quad \text{i} \quad b = \mathcal{X}^2_{n,\alpha/2}$$

on $\mathbf{P}(Y \geq \mathcal{X}^2_{n,p}) = p$ per $p \in (0, 1)$ i la variable Y és una $\mathcal{X}^2_{2n}$. En aquest cas,

$$g_{\underset{\sim}{x}}(\lambda) = Q(\underset{\sim}{x}, \lambda) = \frac{\sum_{i=1}^{n} x_i}{\lambda},$$

és a dir, $g_{\underset{\sim}{x}}$ és invertible i decreixent, per tant l'interval de confiança $1 - \alpha$ per λ serà

$$C(\underset{\sim}{x}) = \left[\, g^{-1}_{\underset{\sim}{x}}(b), g^{-1}_{\underset{\sim}{x}}(a) \,\right] = \left[\frac{\sum_{i=1}^n x_i}{\mathcal{X}^2_{n,1-\alpha/2}}, \frac{\sum_{i=1}^n x_i}{\mathcal{X}^2_{n,\alpha/2}} \right]$$

3.2 Contrastos uniformement més potents

Ens ocuparem ara de les proves d'hipòtesi amb hipòtesi alternativa composta.

Definició 3.10 *Suposem que volem contrastar:*

$$\begin{cases} H_0 : \theta \in \Omega_0 \\ H_1 : \theta \in \Omega_1 \end{cases}$$

on $\Omega_0 \cup \Omega_1 = \Omega$, $\Omega_0 \cap \Omega_1 = \emptyset$.

Per exemple, si $\Omega = [\theta_0, \infty)$, *podem contrastar*

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta > \theta_0 \end{cases}$$

i direm que es tracta d'una prova **unilateral**.

Si $\Omega = \mathbb{R}$,

$$\begin{cases} H_0 : \theta \leq \theta_0 \\ H_1 : \theta > \theta_0 \end{cases} \qquad i \qquad \begin{cases} H_0 : \theta \geq \theta_0 \\ H_1 : \theta < \theta_0 \end{cases}$$

són també proves unilaterals.

En canvi, la prova

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

es diu que és una prova **bilateral**.

Definició 3.11 *Direm que una prova d'hipòtesi és* **uniformement més potent (UMP)** *de mida* α *per contrastar* $H_0 : \theta \in \Omega_0$ *enfront de* $H_1 : \theta \in \Omega_1$, *si la seva funció de potència* $\Pi(\theta)$ *verifica que*

$$\sup_{\theta \in \Omega_0} \Pi(\theta) = \alpha$$

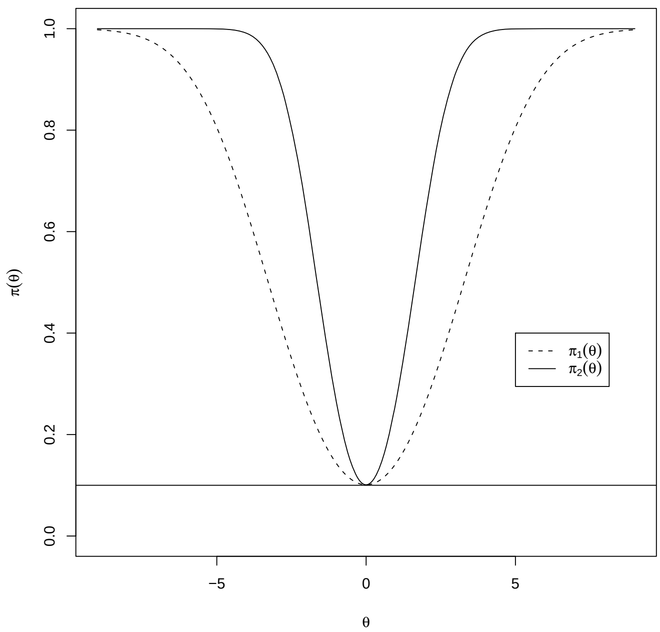

Figura 3.1: Representació gràfica de les funcions de potències $\Pi(\theta)$ i $\Pi^*(\theta)$.

i per a qualsevol altra prova amb funció de potència Π^* *que també sigui de mida* α*, és a dir, que compleixi*

$$\sup_{\theta \in \Omega_0} \Pi^*(\theta) = \alpha$$

es té que

$$\Pi(\theta) \geq \Pi^*(\theta), \qquad \textit{per a tot } \theta \in \Omega_1$$

La situació es representa gràficament a la Figura 3.1.

3.2.1 Lema de Neyman-Pearson per alternatives compostes

El següent resultat és una extensió del Lema de Neyman-Pearson pel cas d'hipòtesis alternatives compostes.

Teorema 3.2 *Es vol contrastar*

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

Per a cada $\theta \in \Omega_1$, es consideren els conjunts

$$C(\theta_1) = \left\{ \underset{\sim}{x} : \frac{L(\theta_1 \mid \underset{\sim}{x})}{L(\theta_0 \mid \underset{\sim}{x})} \geq A(\theta_1) \right\},$$

les regions crítiques de les proves més potents de mida α per a contrastar

$$\begin{cases} H_0 : \theta = \theta_0 \\ H_1 : \theta = \theta_1 \end{cases}$$

donades pel Lema de Neyman-Pearson.

Si aquestes regions crítiques no depenen de θ_1, llavors la prova estadística que té C per regió crítica és UMP de mida α.

Exemple 3.6

Volem contrastar

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

basant-nos en una m.a.s. $X_1, \ldots, X_n$ de $X \sim N(\mu, \sigma^2)$, amb σ^2 coneguda.

Veiem què passa si prenem qualsevol $\mu_1 > \mu_0$.

La nostra intuïció ens duria a rebutjar H_0 per a valors grans de la mitjana mostral $\overline{X}_n$. L'aplicació del Lema de Neyman-Pearson ens conduirà al mateix lloc. En efecte,

La funció de versemblança d'una mostra és

$$L(\mu \mid \underset{\sim}{x}) = (2\,\pi\sigma^2)^{-n/2}\, exp\left\{ -\frac{1}{2\,\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \right\}$$

i el quocient de versemblances

$$\frac{L(\mu_1 \mid \underset{\sim}{x})}{L(\mu_0 \mid \underset{\sim}{x})} = \frac{(2\,\pi\sigma^2)^{-n/2}\, exp\,\{\, -\frac{1}{2\,\sigma^2}\, \sum_{i=1}^{n} (x_i - \mu_1)^2\,\}}{(2\,\pi\sigma^2)^{-n/2}\, exp\,\{\, -\frac{1}{2\,\sigma^2}\, \sum_{i=1}^{n} (x_i - \mu_0)^2\,\}} =$$

$$exp\left\{ \frac{1}{2\,\sigma^2}\, \sum_{i=1}^{n} \Big((x_i - \mu_0)^2 - (x_i - \mu_1)^2 \Big) \right\} =$$

$$exp\left\{ \frac{n}{2\,\sigma^2}\, \big(2\,\overline{x}\,(\mu_1 - \mu_0)^2 + (\mu_0^2 - \mu_1^2)\,\big) \right\}$$

Així, la regió crítica del test de Neyman-Pearson té la forma

$$C = \left\{ \underset{\sim}{x} : exp\Big\{ \frac{n\,\big(2\,\overline{x}\,(\mu_1 - \mu_0)^2 + (\mu_0^2 - \mu_1^2)\,\big)}{2\,\sigma^2} \Big\} \geq A \right\}$$

Veiem que el quocient de les versemblances és funció creixent de l'estadístic suficient minimal $\overline{x}$, ja que $\mu_1 - \mu_0 > 0$, i per tant

$$C = \{\, \underset{\sim}{x} : \overline{x} \geq B\,\}$$

En aquest cas, les constants A i B es relacionen de la forma

$$B = \frac{\sigma^2\, ln(A)}{n\,(\mu_1 - \mu_0)} + \frac{\mu_1 + \mu_0}{2}$$

Ens cal trobar B per tal que $\mathbf{P}(\overline{X} \geq B \mid H_0) = \alpha$, i com que sota la hipòtesi H_0,

$$\overline{X} \sim N(\mu_0, \sigma^2/n)$$

tindrem que $B = \mu_0 + z_\alpha\, \sigma/\sqrt{n}$.

Per tant, la regió crítica és la mateixa per tots els valors possibles $\mu_1 \in \Omega_1 = (\mu_0, \infty)$. Se segueix, doncs, que la prova amb regió crítica C és UMP de mida α per contrastar

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

La funció de potència del test UMP és

$$\Pi(\mu) = \mathbf{P}(\, \underset{\sim}{X} \in C \mid \mu) = \mathbf{P}\Big(\overline{X}_n \geq \mu_0 + z_\alpha\, \frac{\sigma}{\sqrt{n}} \mid \mu \Big) =$$

$$\mathbf{P}\left(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \geq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_\alpha \mid \mu \right) = \mathbf{P}\left(Z \geq \frac{\mu_0 - \mu}{\sigma/\sqrt{n}} + z_\alpha \right),$$

essent $Z \sim N(0,1)$.

Comencem suposant, com a l'exemple del Z-test, que $\mu_0 = 5$, $\mu_1 = 6$, $\sigma^2 = 1$ i $\alpha = 0.05$. Si observem la mateixa mostra, que recordem era

$$\underset{\sim}{x} = (\,6.1\,,\,5.5\,,\,5.9\,,\,6.3\,)$$

tenim que $\overline{x} = 5.95$. Aleshores havíem vist que rebutjàvem H_0 en favor de H_1 si

$$\overline{X}_n \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

que amb les nostres dades això es traduia en el rebuig de la hipòtesi nul·la, ja que es complia

$$\frac{5.95 - 5}{1/\sqrt{4}} = 1.9 \geq 1.645$$

Anem a calcular ara la funció de potència per aquests valors de μ_0 i μ_1:

$$\Pi(\mu_1) = \mathbf{P}(\,\underset{\sim}{X} \in C \mid \mu = \mu_1\,) = \mathbf{P}\left(\,\overline{X}_n \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \mid \mu = \mu_1\,\right) =$$

$$\mathbf{P}\left(\,\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \geq \frac{\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} - \mu_1}{\sigma/\sqrt{n}}\,\right) = \mathbf{P}\left(\,Z \geq \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_\alpha\,\right) =$$

$$\mathbf{P}\left(\,Z \geq \frac{5-6}{1/2} + 1.645\,\right) = \mathbf{P}\big(\,Z \geq 1.645 - 2\,\big) =$$

$$\mathbf{P}\big(\,Z \geq -0.355\,\big) = \mathbf{P}\big(\,Z \leq 0.355\,\big) = 0.64$$

i el p-valor seria en aquest cas:

$$\text{p-valor} = \mathbf{P}\Big(\,\overline{X}_n \geq 5.9\,\Big) = \mathbf{P}\left(\,Z \geq \frac{5.95 - 5}{1/\sqrt{4}}\,\right) = \mathbf{P}\big(\,Z \geq 1.9\,\big) =$$

$$\mathbf{P}\big(\,Z \leq 1.9\,\big) = 1 - 0.9713 = 0.0287$$

En el cas que $\mu_1 = 8$ aleshores la funció de potència seria:

$$\Pi(8) = \mathbf{P}\left(\,Z > \frac{5-8}{1/2} + 1.645\,\right) = \mathbf{P}\big(\,Z > 1.645 - 6\,\big) =$$

$$\mathbf{P}\big(\, Z \leq 4.355 \,\big) = 1$$

i per últim, si suposéssim $\mu_1 = 6.5$,

$$\Pi(6.5) = \mathbf{P}\Big(\, Z > 1.645 - \frac{5 - 6.5}{1/2} \,\Big) = \mathbf{P}\big(\, Z > -1.355 \,\big) = 0.91$$

Exemple 3.7

Sigui $X_1, \ldots, X_n$ una m.a.s. de $X \sim N(\mu, \sigma^2)$ amb σ^2 conegut. Volem fer la prova

$$\begin{cases} H_0 : \mu \leq \mu_0 \\ H_1 : \mu > \mu_0 \end{cases}$$

Per a qualsevol $0 < \alpha_0 < 1$ existeix una prova UMP. També determinarem la funció de potència d'aquest procediment. La funció de densitat conjunta de $X_1, \ldots, X_n$ té una raó de versemblança monòtona en l'estadístic $\overline{X}_n$.

Un procediment δ_1 que rebutja H_0 si $\overline{X}_n \geq c$ és UMP per a la nostra prova i el seu nivell de significació és

$$\alpha_0 = P\big(\, \overline{X}_n \geq c \,|\, \mu = \mu_0 \,\big)$$

Sigui $Z \sim N(0,1)$ i $z_{\alpha_0} : P(Z \geq z_{\alpha_0}) = \alpha_0$. Sabem que per $\alpha_0 = 0.05 \Longrightarrow z_{\alpha_0} = 1.645$, mentre que si $\alpha_0 = 0.025 \Longrightarrow z_{\alpha_0} = 1.96$. Aleshores

$$Z = \frac{\sqrt{n}\,(\overline{X}_n - \mu_0)}{\sigma} \sim N(0,1) \qquad \text{si } H_0 \text{ és certa}$$

Si volem que

$$P\big(\, \overline{X}_n \geq c \,|\, \mu = \mu_0 \,\big) = P\Big(\, Z \geq \frac{\sqrt{n}\,(c - \mu_0)}{\sigma} \,\Big) = \alpha_0 \Longrightarrow$$

$$\frac{\sqrt{n}\,(c - \mu_0)}{\sigma} = z_{\alpha_0} \Longrightarrow c = \frac{\mu_0 + z_{\alpha_0}\,\sigma}{\sqrt{n}}$$

i la funció de potència serà

$$Pi(\mu \,|\, \delta_1) = P\,\{\,\text{rebutjar}\, H_0 \,|\, \mu\,\} = P\{\,\overline{X}_n \geq c \,|\, \mu\,\} =$$

$$P\left\{\,\overline{X}_n \geq \mu_0 + z_{\alpha_0}\,\frac{\sigma}{\sqrt{n}} \,|\, \mu\,\right\}$$

Per a qualsevol valor de μ, $Z' = \frac{\sqrt{n}\left(\overline{X}_n - \mu\right)}{\sigma} \sim N(0,1)$.

$$\Pi(\mu \,|\, \delta_1) = P\left\{\,\frac{\sqrt{n}\,(\,\overline{X}_n - \mu\,)}{\sigma} \geq \frac{\sqrt{n}\,(\,\mu_0 + z_{\alpha_0}\,\sigma/\sqrt{n}\, - \mu\,)}{\sigma}\,\right\}$$

$$P\left\{\,Z' \geq \frac{\sqrt{n}\,(\mu_0 - \mu)}{\sigma} + z_{\alpha_0}\,\right\} = 1 - \Phi\left(\,z_{\alpha_0} + \frac{\sqrt{n}\,(\mu_0 - \mu)}{\sigma}\,\right) =$$

$$\Phi\left(\,\frac{\sqrt{n}\,(\mu - \mu_0)}{\sigma} - z_{\alpha_0}\,\right)$$

Per altra banda, i seguint amb la mateixa prova que abans, tenim que la prova UMP δ_2 rebutja H_0 si $\overline{X}_n \leq c$ on

$$\alpha_0 = P\big(\,\overline{X}_n \leq c \,|\, \mu = \mu_0\,\big) = P\left(\,Z \leq \frac{\sqrt{n}\,(c - \mu_0)}{\sigma}\,\right) \Longrightarrow$$

$$\frac{\sqrt{n}\,(c - \mu_0)}{\sigma} = -z_{\alpha_0} \Longrightarrow c = \mu_0 - z_{\alpha_0}\,\frac{\sigma}{\sqrt{n}}$$

amb la qual cosa la funció de potència per aquest procediment serà:

$$\Pi(\mu \,|\, \delta_2) = P\big(\,\overline{X}_n \leq c \,|\, \mu\,\big) = \Phi\left(\,\frac{\sqrt{n}\,(\mu_0 - \mu)}{\sigma} - z_{\alpha_0}\,\right)$$

Exemple 3.8

En les mateixes hipòtesis de l'exemple anterior, suposem que volem fer la prova

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

Si contrastéssim H_0 enfront de $H_1' : \mu < \mu_1$ la regió crítica del test UMP seria $C_1 = \{ \underset{\sim}{x} : \overline{X}_n \leq B_1 \}$, mentre que si la prova fos de H_o vs $H_1' : \mu > \mu_1$ llavors la regió crítica seria $C_2 = \{ \underset{\sim}{x} : \overline{X}_n \geq B_2 \}$, on B_1 i B_2 són dues constants tals que:

$$\mathbf{P}\big(\underset{\sim}{X} \in C_1 \,|\, \mu = \mu_0 \big) + \mathbf{P}\big(\underset{\sim}{X} \in C_2 \,|\, \mu = \mu_0 \big) = \alpha$$

o, de forma equivalent,

$$\mathbf{P}\big(\overline{X}_n \leq B_1 \,|\, \mu = \mu_0 \big) + \mathbf{P}\big(\overline{X}_n \geq B_2 \,|\, \mu = \mu_0 \big) = \alpha$$

Una possibilitat, però n'hi ha moltes més, és escollir B_1 tal que

$$\mathbf{P}\big(\overline{X}_n \leq B_1 \,|\, \mu = \mu_0 \big) = \frac{\alpha}{2} \ \text{ és a dir } \ B_1 = \mu_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

i escollir B_2 tal que

$$\mathbf{P}\big(\overline{X}_n \geq B_2 \,|\, \mu = \mu_0 \big) = \frac{\alpha}{2} \ \text{ és a dir } \ B_2 = \mu_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

Aleshores rebutgem H_0 si $\overline{X}_n \leq \mu_0 - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ o $\overline{X}_n \geq \mu_0 + z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ és a dir, rebutgem H_0 si

$$|\, \overline{X}_n - \mu_0 \,| \geq z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

però no es tracta d'una prova UMP, ja que per exemple si rebutgem H_0 en els casos que $\overline{X}_n \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$ i $\mu > \mu_0$ aleshores aquesta prova té potència superior.

3.3 Prova de la raó de versemblança

Sigui $X_1, \ldots, X_n$ una mostra aleatòria simple de X, variable aleatòria amb funció de densitat (o de probabilitat) $f(x \,|\, \theta)$ per algun $\theta \in \Omega$. Es vol fer la següent prova:

$$\begin{cases} H_0 : \theta \in \Omega_0 \\ H_1 : \theta \in \Omega_1 \end{cases}$$

on $\Theta = \Omega_0 \cup \Omega_1$ i $\Omega_0 \cap \Omega_1 = \emptyset$.

Definició 3.12 *Es defineix* **l'estadístic de la raó de versemblances** *com*

$$\lambda = \lambda(\underset{\sim}{x}) = \frac{\max_{\theta\in\Omega_0} L(\theta \mid \underset{\sim}{x})}{\max_{\theta\in\Omega} L(\theta \mid \underset{\sim}{x})}$$

La **prova de la raó de versemblança**, *també anomenada prova de la raó de versemblança generalitzada -per distingir-la del test de Neyman-Pearson- estableix una regió crítica de la forma*

$$C = \left\{ \underset{\sim}{x} : \lambda(\underset{\sim}{x}) \leq A \right\}$$

per alguna constant A que es determinarà per tal que el test tingui mida α.

La idea intuïtiva que hi ha al darrere d'aquest mètode és ben simple: Observem que $0 \leq \lambda \leq 1$ i que com més a prop estigui λ de 1, més versemblant serà que $\theta \in \Omega_0$, mentre que com més lluny estigui del valor 1 més creible serà la hipòtesi alternativa $\theta \in \Omega_1$.

3.3.1 Relació amb el Lema de Neyman-Pearson

Quan H_0 i H_1 són hipòtesis simples, l'estadístic $\lambda(\underset{\sim}{x})$ val el següent:

$$\lambda = \lambda(\underset{\sim}{x}) = \frac{L(\theta_0 \mid \underset{\sim}{x})}{\max\left\{ L(\theta_0 \mid \underset{\sim}{x}), L(\theta_1 \mid \underset{\sim}{x}) \right\}} = \min\left\{ 1, \frac{L(\theta_0 \mid \underset{\sim}{x})}{L(\theta_1 \mid \underset{\sim}{x})} \right\}$$

El test que rebutja H_0 si $\lambda \leq A$ té la mateixa regió crítica (i per tant és la mateixa prova) que el que la rebutja si $L(\theta_1 \mid \underset{\sim}{x})/L(\theta_0 \mid \underset{\sim}{x}) \geq 1/A$, que precisament és la regió crítica donada pel Lema de Neyman-Pearson.

L'estadístic $\lambda(\underset{\sim}{x})$ de la prova de la raó de versemblança depèn de $\underset{\sim}{x}$ només a través de l'estadístic minimal suficient per θ. Les propietats d'aquestes proves per a mostres petites depenen de la modelització paramètrica concreta de què es tracti. La següent és una propietat comuna a totes les proves de raó de versemblances:

- Si H_0 és simple i existeix una prova UMP per contrastar aquesta hipòtesi enfront d'una altra H_1, aleshores la prova de la raó de versemblances coincideix amb el test UMP.

Moltes de les propietats asimptòtiques de les proves de la raó de versemblances són comuns a tots ells. Citem les següents:

- Sota les condicions de regularitat que garanteixen que l'estimador de la màxima versemblança és consistent, es té que la prova de la raó de versemblances és un test consistent.
- El test de la raó de versemblances és asimptòticament la prova no esbiaixada més potent.
- El test anterior és asimptòticament eficient, en el sentit de l'eficiència relativa asimptòtica.

Exemple 3.9

Sigui $X_1, \ldots, X_n$ una mostra aleatòria simple de $X \sim N(\mu, \sigma^2)$ amb μ, σ^2 desconeguts. El paràmetre és doncs, $\theta = (\mu, \sigma^2)$. Es desitja contrastar:

$$\begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu \neq \mu_0 \end{cases}$$

La versemblança té la següent forma:

$$L(\theta \,|\, \underset{\sim}{x}) = (2\,\pi\sigma^2)^{-n/2}\, exp\Big\{ -\frac{1}{2\,\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \Big\}$$

i l'estimador de la màxima versemblança (MLE) sota la hipòtesi H_0 és $\widetilde{\theta} = (\mu_0, \widetilde{\sigma}^2)$, on

$$\widetilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \mu_0)^2$$

i el valor de la versemblança en aquest punt és

$$\max_{\theta \in \Omega_0} L(\theta \,|\, \underset{\sim}{x}) = (2\,\pi\widetilde{\sigma}^2)^{-n/2}\, exp\Big\{ -\frac{n}{2} \Big\}$$

L'estimador MLE en general seria $\widehat{\theta} = (\overline{x}, \widehat{\sigma}^2)$, on

$$\widehat{\sigma}^2 = \frac{n-1}{n} S^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \overline{x})^2,$$

i el màxim de la versemblança:

$$\max_{\theta \in \Omega} L(\theta \,|\, \underset{\sim}{x}) = (2\,\pi\,\widehat{\sigma}^2)^{-n/2}\, exp\Big\{ -\frac{n}{2} \Big\}$$

Per tant, ara ja podem construir el nostre estadístic λ:

$$\lambda = \lambda(\underset{\sim}{x}) = \left(\frac{\widetilde{\sigma}^2}{\widehat{\sigma}^2} \right)^{-n/2} = \left(\frac{\sum_{i=1}^n (x_i - \mu_0)^2}{\sum_{i=1}^n (x_i - \overline{x})^2} \right)^{-n/2} =$$

$$\left(\frac{\sum_{i=1}^n (x_i - \overline{x})^2 + n\,(\overline{x} - \mu_0)^2}{\sum_{i=1}^n (x_i - \overline{x})^2} \right)^{-n/2} = \left(1 + \frac{t^2}{n-1} \right)^{-n/2}$$

on $t = \sqrt{n}\,(\overline{x} - \mu_0)/S$. Per tant, l'estadístic és decreixent en $|t|$.

La prova de la raó de versemblances rebutja H_0 si $\lambda < A$ per algun A, és a dir, si

$$\left(1 + \frac{t^2}{n-1} \right)^{-n/2} < A$$

o equivalentment si

$$\left(1 + \frac{t^2}{n-1} \right)^{n/2} > \frac{1}{A} \iff t^2 > B \qquad \text{per algun } B$$

Per tant, la prova de la raó de versemblances rebutja H_0 si

$$\frac{|\,\overline{x} - \mu_0\,|}{S/\sqrt{n}} > C$$

i la constant C es tria per tal que la prova tingui mida α. Així, doncs, aquest test coincideix amb el test T bilateral clàssic en el mostreig de la normal.

3.3.2 Distribució asimptòtica de la raó de versemblança

Teorema 3.3 *Sigui $X_1, \ldots, X_n$ una m.a.s. d'una població X amb funció de densitat (o de probabilitat) $f(x\,|\,\theta)$ per algun $\theta \in \Omega$. Es vol contrastar:*

$$\begin{cases} H_0 : \theta \in \Omega_0 \\ H_1 : \theta \in \Omega_1 \end{cases}$$

on $\Omega = \Omega_0 \cup \Omega_1$ i $\Omega_0 \cap \Omega_1 = \emptyset$.

Sota les hipòtesis del teorema de Cramer-Rao i suposant que també es verifiquen les condicions següents:

1. *El paràmetre θ és identificable, en el sentit que diferents valors de θ donen lloc a diferents distribucions de probabilitat per a X.*

2. *El conjunt $\{ x : f(x \,|\, \theta) > 0 \}$ és el mateix per a tot $\theta \in \Omega$.*

3. *Per últim, la quantitat*

$$e(\theta_0, \theta) = E_{\theta_0}\left[log\left(\frac{f(X \,|\, \theta)}{f(X \,|\, \theta_0)} \right) \right]$$

existeix per a tot parell $\theta, \theta_0 \in \Omega$.

Aquestes condicions asseguren l'existència i la continuïtat de les derivades respecte al paràmetre de la funció de versemblança, i a més a més també ens permeten afirmar que el suport de la distribució no depèn del paràmetre.

Doncs el teorema ens assegura que sota aquestes condicions, l'estadístic

$$Q_n = -2\,log\lambda(\mathcal{X}_n) \longrightarrow \mathcal{X}_d^2 \qquad sota\ \ H_0 : \theta \in \Omega_0$$

on $d = dim(\Omega) - dim(\Omega_0)$.

3.3.3 Aplicació a la distribució multinomial

En una prova sobre la validesa d'un model la distribució multinomial pot ser útil.

Sigui $X \sim \mathcal{F}(x;\theta)$ i volem contrastar:

$$\begin{cases} H_0 : \mathcal{F}(x;\theta) = \mathcal{F}_0(x;\theta) \\ H_1 : \mathcal{F}(x;\theta) \neq \mathcal{F}_0(x;\theta) \end{cases}$$

Podem pensar, per fixar idees, que $\mathcal{F}_0$ és una Weibull.

Disposem de valors $X_1, \ldots, X_n$. Els agrupem en classes disjuntes $C_1, C_2, \ldots, C_m$, on $C_1 = (-\infty, c_1], C_2 = (c_1, c_2], \ldots, C_m = (c_{m-1}, c_m], C_{m+1} = (c_m, \infty)$ i definim

$$p_j = \mathbf{P}(X \in C_j) = p_j(\theta) \qquad \text{per } \ j = 1, \ldots, m+1$$

També definim

$$p_j^0 = \mathbf{P}\big(X \in C_j \,|\, X \sim \mathcal{F}_0(x;\theta)\big) = p_j^0(\theta)$$

Així, doncs, el nostre problema s'ha convertit en contrastar el següent:

$$\begin{cases} H_0 : p_1 = p_1^0(\theta), \ldots, p_{m+1} = p_{m+1}^0(\theta) \\ H_1 : \text{algun } \ p_j \neq p_j^0(\theta) \end{cases}$$

Sigui $Y_j = \sum\limits_{i=1}^{n} 1_{X_i \in C_j}$. Aleshores

$$(Y_1, \ldots, Y_{m+1}) \sim M(n, p_1, \ldots, p_{m+1})$$

i la funció de versemblança té la forma:

$$L(\theta) = \frac{n!}{y_1! \ldots y_{m+1}!}\, p_1(\theta)^{y_1} \ldots p_{m+1}(\theta)^{y_{m+1}}$$

i el seu màxim s'assoleix per a $\widehat{p}_j = \frac{y_j}{n}$.

Sigui $\widehat{\theta}$ l'estimador de la màxima versemblança (MLE) de θ calculat a partir de la mostra $X_1, \ldots, X_n$. Aleshores el màxim de versemblança sota la hipòtesi H_0 ve donat per $p_j(\widehat{\theta})$.

Per tant, hem obtingut el següent:

$$\Lambda(y_1, \ldots, y_{m+1}) = \frac{p_1(\widehat{\theta})^{y_1} \ldots p_{m+1}(\widehat{\theta})^{y_{m+1}}}{\widehat{p}_1^{\,y_1} \ldots \widehat{p}_{m+1}^{\,y_{m+1}}} = \prod_{j=1}^{m+1} \left(\frac{p_j(\widehat{\theta})}{\widehat{p}_j} \right)^{y_j}$$

Volem aplicar el teorema anterior de la distribució asimptòtica, per tant, anem a aplicar logaritmes a l'expressió que acabem d'obtenir:

$$-2\, log \Lambda = -2 \sum_{j=1}^{m+1} y_j\, log \left(\frac{p_j(\widehat{\theta})}{\widehat{p}_j} \right) = -2\, n \sum_{j=1}^{m+1} \widehat{p}_j\, log \left(\frac{p_j(\widehat{\theta})}{\widehat{p}_j} \right)$$

Per tant, i pel teorema anterior, $-2\, log\Lambda$ segueix una distribució $\mathcal{X}^2_{m+1-k-1} = \mathcal{X}^2_{m-k}$ sota la hipòtesi H_0.

De fet, pot comprovar-se, mitjançant una extensió en sèrie de Taylor, que $-2\, log\Lambda$ és asimptòticament equivalent a

$$\sum_{j=1}^{m+1} \frac{\left(y_j - n\, p_j(\widehat{\theta}) \right)^2}{n\, p_j(\widehat{\theta})}$$

on trobem la forma usual d'observats esperats.

3.4 La prova de la t d'Student

Sigui $X \sim N(\mu, \sigma^2)$.

Prenem el conjunt $\Omega = \left\{ (\mu, \sigma^2) \,:\, -\infty < \mu < \infty, 0 < \sigma^2 < \infty \right\}$.

1. Considerem el problema de contrastar

$$\begin{cases} H_0 : \mu = 0\,,\ \sigma^2 > 0 \\ H_1 : \mu \neq 0\,,\ \sigma^2 > 0 \end{cases}$$

Prenem el subconjunt de Ω següent:

$$\Omega_0 = \left\{ (\mu, \sigma^2)\ :\ \mu = 0, 0 < \sigma^2 < \infty \right\}$$

Suposem a més a més que disposem d'una m.a.s. $X_1, \ldots, X_n$ de $X \sim N(\mu, \sigma^2)$, amb $n > 1$. Sabem que la funció de versemblança conjunta té l'aspecte que segueix:

$$L(\mu, \sigma^2, x_1, \ldots, x_n) = \left(\frac{1}{2\,\pi\,\sigma^2} \right)^{n/2} exp\left[-\frac{\sum_{i=1}^n (x_i - \mu)^2}{2\,\sigma^2} \right] = L(\,\Omega\,)$$

però sota la hipòtesi nul·la

$$L(0, \sigma^2) = \left(\frac{1}{2\,\pi\,\sigma^2} \right)^{n/2} exp\left[-\frac{\sum_{i=1}^n x_i^2}{2\,\sigma^2} \right] = L(w)$$

Cap d'aquestes dues funcions, ni tampoc el seu quocient, poden proporcionar la base per una prova de H_0 vs. H_1.

Suposem però, que calculem el màxim de L en Ω_0 i també el màxim de L en Ω, i que fem el quocient de les versemblances obtingudes.

- Obtenció del MLE de L en Ω_0

$$\frac{d\,lnL(w)}{d\,\sigma^2} = -\frac{n}{2\,\sigma^2} + \frac{\sum_{i=1}^n x_i^2}{2\,\sigma^4} = 0 \Longrightarrow \widehat{\sigma}^2 = \frac{\sum_{i=1}^n x_i^2}{n}$$

i aquest nombre maximitza $L(w)$. El màxim és

$$L(\widehat{w}) = \left(\frac{1}{2\,\pi\,\frac{\sum_{i=1}^n x_i^2}{n}} \right)^{n/2} exp\left\{ -\left(\frac{\sum_{i=1}^n x_i^2}{2\sum_{i=1}^n x_i^2/n} \right) \right\} = \left(\frac{n\,e^{-1}}{2\,\pi\,\sum_{i=1}^n x_i^2} \right)^{n/2}$$

- Obtenció del MLE de L en Ω

Com ja sabem

$$\widehat{\mu} = \frac{\sum_{i=1}^n x_i}{n} \quad \text{i} \quad \widehat{\sigma}^2 = \frac{\sum_{i=1}^n (x_i - \overline{x})^2}{n}$$

maximitzen L, i aquest màxim és precisament

$$L(\widehat{\Omega}) = \left[\frac{1}{2\,\pi\,\sum_{i=1}^{n}(x_i-\overline{x})^2/n} \right]^{n/2} exp\left[-\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{2\,\sum_{i=1}^{n}(x_i-\overline{x})^2/n} \right] =$$

$$\left[\frac{n\,e^{-1}}{2\,\pi\,\sum_{i=1}^{n}(x_i-\overline{x})^2} \right]^{n/2}$$

Anem tot seguit a calcular el quocient d'aquests dos màxims:

$$\Lambda(x_1,\ldots,x_n) = \Lambda = \frac{L(\widehat{w})}{L(\widehat{\Omega})} = \frac{\left(n\,e^{-1}/\,2\,\pi\,\sum_{i=1}^{n}x_i^2\right)^{n/2}}{\left(n\,e^{-1}/\,2\,\pi\,\sum_{i=1}^{n}(x_i-\overline{x})^2\right)^{n/2}} =$$

$$\left(\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sum_{i=1}^{n}x_i^2}\right)^{n/2} = \left(\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2+n\,\overline{x}^2}\right)^{n/2} = \left(\frac{1}{1+\frac{n\,\overline{x}^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}}\right)^{n/2}$$

Fixem-nos que sempre es verifica que $0 \le \Lambda \le 1$.

Sota $H_0 : \mu = 0$ i $\sigma^2 > 0$

- Si obtinguéssim $\overline{x} = 0$ i $\sum_{i=1}^{n} x_i^2 > 0$, l'experiment tendiria a confirmar H_0 i en aquest cas $\Lambda = 1$.
- Si tant $\overline{x}$ com $n\,\overline{x}^2/\sum_{i=1}^{n}(x_i-\overline{x})^2$ es desvien considerablement de 0, l'experiment tendiria a negar H_0; com més gran es desviï més petit és Λ.

Rebutgem H_0 si $0 \le \Lambda \le \lambda_0$ i aquesta és una prova de la raó de versemblança.

Fixem-nos en el següent:

$$\Lambda = \lambda_0 \Longleftrightarrow \left(\frac{\sum_{i=1}^{n}(x_i-\overline{x})^2}{\sum_{i=1}^{n}X_i^2}\right)^{n/2} \le \lambda_0$$

$$\Longleftrightarrow \left(1+\frac{n\,\overline{x}^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2}\right)^{n/2} \ge \frac{1}{\lambda_0}$$

$$\Longleftrightarrow 1+\frac{n\,\overline{x}^2}{\sum_{i=1}^{n}(x_i-\overline{x})^2} \ge \frac{1}{\lambda_0^{2/n}} - 1$$

$$\iff \frac{n\,\overline{x}^2}{\sum_{i=1}^n (x_i-\overline{x})^2} \geq \frac{1}{\lambda_0^{2/n}} - 1$$

$$\iff \frac{n\,\overline{x}^2}{\sum_{i=1}^n (x_i-\overline{x})^2/n-1} \geq (n-1)\left(\lambda_0^{-2/n} - 1\right)$$

$$\iff \frac{\sqrt{n}\,|\,\overline{x}\,|}{\sqrt{\sum_{i=1}^n (x_i-\overline{x})^2/n-1}} \geq \sqrt{(n-1)\left(\lambda_0^{-2/n} - 1\right)} = c$$

Sota H_0 ja sabem que

$$t(x_1,\ldots,x_n) = \frac{\sqrt{n}\,(\overline{x}-0)}{\sqrt{\sum_{i=1}^n (x_i-\overline{x})^2/\,(n-1)}} \sim t_{n-1}$$

O sigui que el test de la raó de versemblança de H_0 versus H_1 pot ser basat en la T. Per tant, si obtenim valors $x_1,\ldots,x_n$ rebutgem H_0 si $|\,t(x_1,\ldots,x_n)\,| \geq c$ essent c determinat per

$$\alpha = \mathbf{P}\left[\ |\,t(x_1,\ldots,x_n)\,| \geq c\,|\,H_0\ \right]$$

2. Si ara volem contrasar

$$\begin{cases} H_0 : \mu \leq \mu_0\,,\ \sigma^2 \\ H_1 : \mu > \mu_0\,,\ \sigma^2 \end{cases}$$

i considerem el següent subconjunt de Ω:

$$\Omega_0 = \left\{\ (\mu,\sigma^2)\ :\ \mu \leq \mu_0, \sigma^2 > 0\ \right\}$$

Els estimadors de la màxima versemblança restringits a Ω_0 són

$$\widehat{\mu} = \mu_0 \quad \text{i} \quad \widehat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\mu_0)^2$$

i per tant

$$\Lambda^{-1} = \left(1 + \frac{n\,(\overline{x}-\mu_0)^2}{\sum_{i=1}^n (x_i-\overline{x})^2}\right)^{n/2}$$

igual que abans.

3.5 La prova t per mostres aparellades

En molts casos, les dades que tenim en un experiment concret, són aparellades. Per exemple, en un experiment mèdic, els individus estan agrupats per edats o per pesos, i llavors un membre de cadascuna de les parelles està assignat de manera aleatòria al grup "treatment" i l'altre al grup de control. En algunes aplicacions la parella consisteix en una mesura abans i després d'un mateix objecte. Per exemple, la mesura de la pressió arterial abans i després d'haver fet un esforç físic.

Notem les parelles per (X_i, Y_i), on $i = 1, \ldots, n$, i suposem que X i Y tenen mitjanes μ_X i μ_Y i variàncies σ_X^2 i σ_Y^2. Suposarem també que les diferents parelles estan distribuides de forma independent i que $Cov(X_i, Y_i) = \sigma_{XY}$. Treballarem amb les diferències

$$D_i = X_i - Y_i$$

que són independents amb

$$E(D_i) = \mu_X - \mu_Y \quad \text{i} \quad Var(D_i) = \sigma_X^2 + \sigma_Y^2 - 2\,\sigma_{XY}$$

Una estimació natural per $\mu_X - \mu_Y$ seria $\overline{D}$, la mitjana de la diferència. De les propietats de D_i podem afirmar que

$$E(\overline{D}) = \mu_X - \mu_Y \quad \text{i} \quad Var(\overline{D}) = \frac{1}{n}\,(\sigma_X^2 + \sigma_Y^2 - 2\,\sigma_{XY})$$

Suposem que fem un experiment prenent una mostra simple de grandària n de X i una mostra independent i també simple de mida n de Y. Llavors $\mu_X - \mu_Y$ serà estimat per $\overline{X} - \overline{Y}$ i

$$E(\overline{X} - \overline{Y}) = \mu_X - \mu_Y$$

$$Var(\overline{X} - \overline{Y}) = \frac{1}{n}\,(\sigma_X^2 + \sigma_Y^2)$$

Comparant les variàncies de les dues estimacions veiem que la variànciaça de $\overline{D}$ és més petita si el terme de la covariànciaça és positiu. En aquesta circumstància, aparellar les dades és la forma més efectiva de procedir. En el cas simple en que $\sigma_X = \sigma_Y = \sigma$, les dues variànciaces es poden expressar de manera més simple com

$$Var(\overline{D}) = \frac{2\,\sigma^2\,(1-\rho)}{n}$$

$$Var(\overline{X} - \overline{Y}) = \frac{2\,\sigma^2}{n}$$

i el seu quocient és

$$\frac{Var(\overline{D})}{Var(\overline{X} - \overline{Y})} = 1 - \rho$$

Si el coeficient de correlació és 0.5, per exemple, un conjunt de dades aparellades amb n parells ens dóna la mateixa precisió que un conjunt de $2\,n$ dades sense aparellar.

3.5.1 Mètodes basats en la distribució normal

En aquest subapartat suposarem que les diferències D_i provenen d'una mostra normal amb distribució

$$E(D_i) = \mu_X - \mu_Y = \mu_D \quad \text{i} \quad Var(D_i) = \sigma_D^2$$

Generalment, σ_D serà desconegut, i les inferències estan basades en

$$t = \frac{\overline{D} - \mu_D}{s_{\overline{D}}}$$

que segueix una distribució t amb $n-1$ graus de llibertat. Un interval de confiança per μ_D del $100\,(1-\alpha)\%$ seria

$$\left(\overline{D} - t_{n-1,\alpha/2} s_{\overline{D}}\ ,\ \overline{D} + t_{n-1,\alpha/2} s_{\overline{D}}\right)$$

Una prova d'hipòtesis compostes per $H_0 : \mu_D = 0$ de nivell α tindria regió de rebuig

$$|\,\overline{D}\,| \geq t_{n-1,\alpha/2} s_{\overline{D}}$$

Si la mostra de mida n és prou gran, la validesa aproximada de l'interval de confiança i del test d'hipòtesi se segueixen del teorema central del límit. Si la mostra és petita i la distribució de les diferències és lluny de la normal, les probabilitats obtingudes seran força errònies.

Exemple 3.10

Per tal d'estudiar l'efecte que causa el tabac sobre la incorporació de les plaquetes, Levine (1973), va dibuixar mostres de sang d'11 individus abans i després de fumar, i va mesurar el grau al qual les plaquetes es van agregar. Les plaquetes estan implicades en la formació de coàguls de sang, i se sap que els fumadors pateixen més sovint de desordres que impliquen la seva formació que no pas les persones no fumadores. Les dades es mostren a la taula següent, que dóna el percentatge màxim de totes les plaquetes que es van agregar després de l'exposició a un estímul.

Abans	Després	Diferència
25	27	2
25	29	4
27	37	10
44	56	12
30	46	16
67	82	15
53	57	4
53	80	27
52	61	9
60	59	-1
28	43	15

De la columna de les diferències, tenim que $\overline{D} = 10.27$ i $s_{\overline{D}} = 2.40$. L'estadístic T_{10} és 4.28, i el p-valor per a un test amb hipòtesis compostes és 0.0016.

3.6 La prova t per a dues mostres independents

En molts experiments, les dues mostres es poden considerar com independents una de l'altra. En un estudi mèdic, per exemple, una mostra d'individus pot ser assignada a un tractament particular, i una altra mostra independent de la primera pot ser assignada a un tractament de control. Molts experiments són tals que si es repetissin, les mesures no serien exactament les mateixes. Per tractar aquest problema, molt sovint utilitzem un model estadístic: les observacions del grup de control són modelades com a variables aleatòries independents amb una distribució comú F, i les observacions del grup del tractament són modelades com a independents.

Suposem que tenim dues mostres $X_1, \ldots, X_m$ de $X \sim N(\mu_1, \sigma^2)$ i $Y_1, \ldots, Y_n$ de $Y \sim N(\mu_2, \sigma^2)$ independents. Podem pensar que X és el grup que ha rebut un determinat tractament, i que Y és el grup de control. Aleshores l'efecte del tractament es caracteritza per la diferència $\mu_1 - \mu_2$.

Volem fer la següent prova:

$$\left\{ \begin{array}{l} H_0 : \mu_1 \leq \mu_2 \\ H_1 : \mu_1 > \mu_2 \end{array} \right. \qquad \Longleftrightarrow \qquad \left\{ \begin{array}{l} H_0 : \mu_1 - \mu_2 \leq 0 \\ H_1 : \mu_1 - \mu_2 > 0 \end{array} \right.$$

En el cas que σ^2 fos conegut, tenim que

$$\lambda(\underset{\sim}{x}, \underset{\sim}{y}) = \frac{\sup_{(\mu_1,\mu_2,\sigma^2)\in\Omega_0} L(\mu_1, \mu_2, \sigma^2 \mid \underset{\sim}{x}, \underset{\sim}{y})}{\sup_{(\mu_1,\mu_2,\sigma^2)\in\Omega} L(\mu_1, \mu_2, \sigma^2 \mid \underset{\sim}{x}, \underset{\sim}{y})}$$

Rebutjarem H_0 si es compleix que $\lambda(\underset{\sim}{x}, \underset{\sim}{y}) \leq c$, on c és una certa constant, la que fa que el nostre test tingui mida α.

Una estimaciço natural per $\mu_1 - \mu_2$ és $\overline{X}_m - \overline{Y}_n$. Com que $\overline{X}_m - \overline{Y}_n$ es pot expressar com a combinació lineal de variables normals independents, es distribueix també com una normal:

$$\overline{X}_m - \overline{Y}_n \sim N\left[\mu_1 - \mu_2, \sigma^2 \left(\frac{1}{m} + \frac{1}{n}\right)\right]$$

Suposant, doncs que σ^2 fos conegut, un interval de confiança per $\mu_1 - \mu_2$ estaria basat en

$$Z = \frac{\overline{X}_m - \overline{Y}_n - (\mu_1 - \mu_2)}{\sigma \sqrt{\frac{1}{m} + \frac{1}{n}}}$$

que segueix una distribució normal $N(0, 1)$. L'interval de confiança seria de la forma:

$$\left(\left(\overline{X}_m - \overline{Y}_n\right) - z_{\alpha/2}\, \sigma \sqrt{\frac{1}{m} + \frac{1}{n}}\ ,\ \left(\overline{X}_m - \overline{Y}_n\right) + z_{\alpha/2}\, \sigma \sqrt{\frac{1}{m} + \frac{1}{n}} \right)$$

Normalment, però, σ^2 és desconegut, i aquesta s'haurà d'estimar a partir de les dades per mitjà del càlcul de la mitjana ponderada de S_X^2 i S_Y^2, que es defineix com:

$$s_p^2 = \frac{(m-1)\,S_X^2 + (n-1)\,S_Y^2}{n+m-2}$$

Aquest valor és molt útil en el cas que una mostra sigui més gran que l'altre. En aquest cas l'estimació de σ^2 de la primera mostra és molt més fiable i hauria de rebre el pes més gran. El següent teorema ens proporciona la distribució d'un estadístic que s'utilitza per a obtenir intervals de confiança i realitzar proves d'hipòtesis.

Teorema 3.4 *En les condicions que fins ara estem comentant, l'estadístic*

$$t = \frac{\overline{X}_m - \overline{Y}_n - (\mu_1 - \mu_2)}{s_p\sqrt{\frac{1}{m} + \frac{1}{n}}}$$

segueix una distribució t d'Student amb $n+m-2$ graus de llibertat.

I això és degut la fet que t és el quocient de dues variables U i V, tals que

$$U = \frac{\overline{X}_m - \overline{Y}_n}{\sigma\sqrt{\frac{1}{m} + \frac{1}{n}}} \sim N(0,1) \quad \text{i} \quad V = \sqrt{\frac{\frac{(m-1)\,S_X^2}{\sigma^2} + \frac{(n-1)\,S_Y^2}{\sigma^2}}{m+n-2}}$$

i en aquest cas es verifica

$$\frac{(m-1)\,S_X^2}{\sigma^2} \sim \mathcal{X}_{m-1}^2 \text{ i } \frac{(n-1)\,S_Y^2}{\sigma^2} \sim \mathcal{X}_{n-1}^2$$

la qual cosa fa que $V \sim \mathcal{X}_{m+n-2}^2$. Per tant, $t \sim t_{m+n-2}$.

Per acabar de veure més clarament tot això, anem a fer tot seguit un exemple.

Exemple 3.11

Suposem que s'utilitzen dos mètodes A i B per a determinar l'energia calorífica necessària per aconseguir la fusió del gel. Del que es tracta és de veure on difereixen els dos mètodes. La següent taula dóna el canvi en calor

Mètode A	Mètode B
79.98	80.02
80.04	79.94
80.02	79.98
80.04	79.97
80.03	79.97
80.03	80.03
80.04	79.95
79.97	79.97
80.05	
80.03	
80.02	
80.00	
80.02	

total per passar de gel a -72^oC a aigua a 0^oC. Remarquem que les dades són en calories per gram.

És obvi a partir d'aquesta taula que hi ha força diferència entre els dos mètodes. Si suposem que es compleixen les hipòtesis del teorema anterior, podem formar un interval del 95% de confiança per estimar la magnitud de la diferència de mitjanes entre els dos mètodes. A partir, doncs, de la taula, podem afirmar que

$$\begin{array}{lll} n_A = 13 & \overline{X}_A = 80.02 \quad \text{i} & s_A = 0.024 \\ n_B = 8 & \overline{X}_B = 79.98 \quad \text{i} & s_B = 0.031 \end{array}$$

i per tant

$$s_p^2 = \frac{12 \times s_A^2 + 7 \times s_B^2}{19} = 0.0007178 \Longrightarrow s_p = 0.027$$

amb la qual cosa, l'interval de confiança per $\mu_A - \mu_B$ al 95% serà:

$$\left(\big(\overline{X}_A - \overline{X}_B\big) - t_{19,0.025}\, s_p \sqrt{\frac{1}{13} + \frac{1}{8}}\, , \, \big(\overline{X}_A - \overline{X}_B\big) + t_{19,0.025}\, s_p \sqrt{\frac{1}{13} + \frac{1}{8}} \right)$$

$$= \big(0.02, 0.07\big)$$

Capítol 4

PROVES PER A LA VALIDESA D'UN MODEL I MÈTODES NO PARAMÈTRICS

4.1 Proves per a la validesa d'un model

4.1.1 Gràfics de probabilitat

Els gràfics de probabilitat són una eina molt útil per ajustar unes dades a una distribució teòrica.

En general els gràfics de probabilitat exploten alguna característica del model teòric amb l'objectiu de trobar una relació lineal entre les variables observades i aquesta característica.

Distingirem entre els procediments basats en l'ús de la distribució empírica i els basats en els moments de la distribució uniforme.

Els aventatges dels gràfics de probabilitat són:

- Són ràpids i senzills d'utilitzar, al contrari dels mètodes numèrics, que poden resultar feixucs, i a més requereixen un coneixement rigorós de la metodologia estadística o un estadístic per resoldre'l.
- Ajuden a presentar les dades d'una forma visualment comprensible. Això ha de permetre no solament el fet d'extreure conclusions, sinó

també la seva exposició a altres persones.

- Proporcionen els valors dels paràmetres de la distribució, així com els seus percentils, la taxa d'error, el percentatge que falla durant la garantia, i moltes altres quantitats.
- Serveixen per establir com de bé la distribució teòrica ajusta les dades.
- Ajuden a descobrir dades inusuals.
- Permeten verificar algunes de les hipòtesis en les quals es basen molts mètodes analítics.
- Es poden utilitzar també per a dades censurades.

Els gràfics, però, també presenten alguns inconvenients, que enunciem a continuació.

Inconvenients dels gràfics de probabilitat:

- No són absolutament objectius. Dues persones diferents podrien arribar a dos valors lleugerament diferents per a un mateix paràmetre. Això no obstant no implica que s'arribin a conclusions diferents sobre el mateix conjunt de dades.
- No proporcionen intervals de confiança pels paràmetres.

Construcció d'un gràfic de probabilitat mitjançant un paper de probabilitat

Per tal de construir un gràfic de probabilitat seguirem els passos següents:

1. Ordenar les n observacions del més petit al més gran.
2. Assignar un rang a cadascun dels temps anteriors. L'observació més petita li correspon rang 1, al segon error rang 2, ..., a l'última observació li correspon rang n.
3. Calcular la posició i-èssima del gràfic (i^{th} *plotting position*), F_i, de la següent manera:

   $$F_i = \frac{100\,(i-0.5)}{n} \qquad i = 1, \ldots, n$$

   F_i representa la proporció d'observacions que han tingut lloc abans del temps d'error corresponent al rang i.

4. Dibuixar la variable en paper de probabilitat. Començar etiquetant l'eix de les abscisses (eix horitzontal) de manera que inclogui tots els valors observats. Per cada temps d'error (per exemple, pel que té rang i), situar un punt en el paper que tingui com a abscissa el valor del temps corresponent i com a ordenada la posició F_i.

5. Valorar si la distribució escollida ajusta correctament les dades. Si el núvol de punts que resulta en el gràfic està situat al voltant d'una recta, llavors això voldrà dir que la distribució escollida és l'adequada. En aquest cas es pot procedir a estimar els paràmetres. Un núvol de punts curvilini indica que la distribució escollida no ajusta bé les dades; en aquest cas s'ha d'utilitzar un paper corresponent a una altra distribució. Si després de provar amb diferents papers no se'n troba cap tal que l'ajustament produeixi una recta, s'utilitza aquell que sembli més apropiat, s'ajusta una corba i s'interpreta el gràfic de la mateixa manera que si s'hagués ajustat una recta. En aquests casos el que no és en absolut aconsellable és l'extrapolació dels resultats fora de l'interval de valors en el qual s'està treballant. De vegades és recomanable la comprovació de l'ajustament mitjançant les anomenades proves d'ajustament (*goodness-of-fit tests*).

   Si en el traçat d'aquests gràfics apareixen punts molt allunyats de la resta, en principi s'ha de prescindir d'ells però s'ha d'esbrinar la causa que els ha produït; això pot contribuir a una millora del producte o de la forma en com s'han recollit les dades.

6. Determinar visualment una recta de forma que les desviacions entre les dades i la recta siguin tant petites com sigui possibles. Aquesta recta ens està estimant la funció de distribució.

7. Obtenir la informació desitjada.

<u>Ajustaments gràfics basats en la distribució uniforme</u>

Sigui $X_1, \ldots, X_n$ una mostra d'una certa població X. Suposem que volem fer la següent prova:

$$\begin{cases} H_0 : X \sim F_0 \text{ amb } \theta \text{ conegut} \\ H_1 : X \text{ no es distribueix com } F_0 \end{cases}$$

A partir de les dades definim l'estadístic d'ordre:

$$X_{(1)}, \ldots, X_{(n)}$$

i fem el següent plot:

$$\Big(X_{(i)}, F_\theta^{-1}(i/n+1) \Big) \quad 1 \leq i \leq n$$

Si el núvol de punts és rectilini aleshores acceptem la hipòtesi H_0.

Motivació:

La variable aleatòria $Y = F_\theta(\, X\,) \sim U[\,0,1\,]$.

Per tant $F_\theta(\, X_1\,), \ldots, F_\theta(\, X_n\,) \sim U[\,0,1\,]$.

Si ordenem les dades tenim que

$$F_\theta(\, X_{(1)}\,) \leq \ldots \leq F_\theta(\, X_{(n)}\,) \quad \text{i} \quad E\Big\{ F_\theta(\, X_{(i)}\,) \Big\} = \frac{i}{n+1}$$

Per tant un plot de $\Big(F_\theta(\, X_{(i)}\,), i/n+1 \Big)$ indica que

$F_\theta(\, X_i\,) \sim U[\,0,1\,] \;\; i = 1, \ldots, n$ o equivalentment $\Big(X_{(i)}, F_\theta^{-1}(i/n{+}1) \Big) \;\; 1 = 1, \ldots, n$

Un cas particular es presenta si θ és desconegut, però

$$F_\theta(\, x\,) = G\Big(\frac{x-\mu}{\sigma} \Big)$$

com és el cas normal.

Un plot de $\Big(X_{(i)}, G^{-1}\Big(\frac{i}{n+1} \Big) \Big)$ serà rectilini amb pendent σ i intersecció μ si G és efectivament la distribució de X.

Aquí Exemple en R

Dades simulades d'una Weibull amb $\alpha = 10$ i $\beta = 2$.

4.1.2 Prova de Kolmogorov-Smirnov

Sigui $X_1, \ldots, X_n$ una mostra d'una població que té per funció de distribució la funció F. Ens plantegem l'estimació d'aquesta funció.

Suposem que hem observat els n valors: $x_1, \ldots, x_n$ i construïm la funció de distribució empírica F_n.

$$F_n(x) = \begin{cases} 0 & x < x_1 \\ 1/n & x_1 < x < x_2 \\ 2/n & x_2 < x < x_3 \\ \cdot & \\ \cdot & \\ \cdot & \\ n-1/n & x_{n-1} < x < x_n \\ 1 & x \geq x_n \end{cases}$$

Aquesta funció es pot considerar com la funció de distribució d'una variable aleatòria discreta que assigna probabilitat $1/n$ a cadascun dels valors $x_1, \ldots, x_n$.

El que ens preguntem ara és si aquest estimador és bo.

Fixem-nos que:

$$F_n(x) = \frac{1}{n} \#\{\, i : x_i \leq x \,\} \underset{n\to\infty}{\overset{P}{\to}} P\{\, X_i \leq x \,\} = P\{\, X \leq x \,\} = F(x)$$

és a dir, per la llei dels grans nombres,

$$\lim_{n\to\infty} F_n(x) = F(x) \quad \text{per a cada} \quad -\infty < x < +\infty$$

Tenim convergència en probabilitat per a cada punt.

Definició 4.1 *Considerem una variable aleatòria X amb una funció de distribució F i construïm la funció de distribució empírica F_n a partir dels n valors observats $x_1, \ldots, x_n$.*

Definim

$$D_n = \sup_{-\infty < x < +\infty} |\, F_n(x) - F(x) \,|$$

D_n es pot prendre com una variable aleatòria que depèn de les variables aleatòries $X_1, \ldots, X_n$, i sabem que el teorema de Glivenko-Cantelli ens assegura que

$$\lim_{n\to\infty} D_n = 0$$

Suposem ara que volem fer una prova per saber si la funció de distribució d'una certa variable aleatòria X és F. Ens plantegem el següent test d'hipòtesi:

$$\begin{cases} H_0 : F(x) = F^*(x) \quad -\infty < x < \infty \\ H_1 : F^* \text{ no és la distribució de } X \end{cases}$$

La prova que a continuació descriurem, a diferència de la prova $\mathcal{X}^2$, no requereix agrupació de les dades.

Considerem com abans $F_n(x)$ la funció de distribució empírica construïda a partir dels valors observats $x_1, \ldots, x_n$.

Definim
$$D_n^* = \sup_{-\infty < x < \infty} |\, F_n(x) - F^*(x)\, |$$
La distribució de D_n^* és independent de $F^*(x)$ quan H_0 és cert i la seva llei asimptòtica és com segueix:

$$\forall \; t > 0 \;\; \lim_{n\to\infty} P\Big\{ \sqrt{n}\, D_n^* \leq t \Big\} = 1 - 2 \sum_{i=1}^{\infty} (-1)^{i-1}\, e^{-2\, i^2\, t^2} = H(t)$$

Els valors asimptòtics $H(t)$ es troben a la taula 4.1.

t	$H(t)$	t	$H(t)$
0.30	0.0000	1.20	0.8878
0.35	0.0003	1.25	0.9121
0.40	0.0028	1.30	0.9319
0.45	0.0126	1.35	0.9478
0.50	0.0361	1.40	0.9603
0.55	0.0772	1.45	0.9702
0.60	0.1357	1.50	0.9778
0.65	0.2080	1.60	0.9880
0.70	0.2888	1.70	0.9938
0.75	0.3728	1.80	0.9969
0.80	0.4559	1.90	0.9985
0.85	0.5347	2.00	0.9993
0.90	0.6073	2.10	0.9997
0.95	0.6725	2.20	0.9999
1.00	0.7300	2.30	0.9999
1.05	0.7798	2.40	1.0000
1.10	0.8223	2.50	1.0000
1.15	0.8580		

Taula 4.1: Valors de $H(t)$

Per exemple, $H(t) = 0.95 \longrightarrow \left\{ \begin{array}{lcl} 0.9603 & = & H(1.40) \\ 0.9478 & = & H(1.35) \end{array} \right.$

i per tant $1.35 < t < 1.40$.

Interpolant,

$$\frac{0.95 - 0.9478}{t - 1.35} = \frac{0.9603 - 0.9478}{1.40 - 1.35} \implies \frac{0.0022}{t - 1.35} = \frac{0.0125}{0.05}$$

$$0.00011 = 0.0125\,t - 0.016875$$

$$0.016985 = 0.0125\,t$$

$$t = 1.3588$$

Aquest resultat es deu als professors Kolmogorov i Smirnov i és de l'any 1930.

Definició 4.2 *El procediment que rebutja H_0 quan $\sqrt{n}\,D_n^* > c$ essent c tal que $H(c) = 1 - \alpha$ rep el nom de* **prova de Kolmogorov-Smirnov**.

Adonem-nos que és intuïtivament correcte, ja que pel teorema de Glivenko-Cantelli si H_0 és cert l'estadístic D_n^* prendrà valors petits.

Exemple 4.1

Ajustament d'una normal

Considerem el test següent:

$$\begin{cases} H_0 : X \sim N(0,1) \\ H_1 : X \ \text{ no és } \ N(0,1) \end{cases}$$

o equivalentment

$$\begin{cases} H_0 : F(x) = \phi(x) \\ H_1 : F(x) \neq \phi(x) \end{cases}$$

Disposem d'una mostra de grandària 25: $x_1, \ldots, x_{25}$. Denotem per $y_1, \ldots, y_{25}$ els 25 valors observats (veure la taula 4.2).

Comencem calculant la funció de distribució empírica F_n.

$$F_n(y_i) = \frac{1}{n}\,\#\{\, j : y_j \leq y_i \,\} = \frac{i}{n} = \frac{i}{25} = 0.04\,i$$

i calculem per a cada $y_i, \phi(y_i) = \mathbf{P}\{\, Z \leq y_i \,\}$.

$$D_n^* = \sup_{-\infty < x < \infty} |\, F_n(x) - \phi(x) \,|$$

Com que $\phi(x)$ és contínua $\phi(x) = \phi(x^-)$ però com que F_n és discreta $F_n(y_i) \neq F_n(y_i^-)$, de fet el que es compleix és que $F_n(y_i^-) = F_n(y_{i-1})$.

Hem de calcular totes les diferències $F_n(x) - \phi(x)$, i també $F_n(x^-) - \phi(x)$.

En aquest exemple, la màxima diferència es dóna en el punt -0.42.

$$\begin{array}{rcl} \widehat{F}_n(-0.42) & = & 0.20 \\ \widehat{F}_n(-0.42^-) & = & 0.16 \\ \phi(-0.42) & = & 0.3372 \end{array}$$

Per tant, tenim que

$$\begin{array}{rcl} \phi(-0.42) - \widehat{F}_n(-0.42) & = & 0.1372 \\ \phi(-0.42) - \widehat{F}_n(-0.42^-) & = & 0.1772 \end{array}$$

amb la qual cosa podem afirmar que

$$D_n^* = 0.1772 \quad \Longrightarrow \quad \sqrt{n}\, D_n^* = 5\, D_{25}^* = 0.886$$

i el p-valor correspon a la probabilitat

$$\mathbf{P}\{\, \sqrt{n}\, D_n^* \geq 0.886 \,|\, H_0 \,\} \simeq 1 - H(0.886)$$

Interpolant

$$\frac{0.90 - 0.85}{0.90 - 0.886} = \frac{0.6073 - 0.5347}{0.6073 - H(0.886)} \quad \Longrightarrow \quad H(0.886) = 0.587$$

i per tant el p-valor és $1 - 0.587 = 0.41$, cosa que vol dir que acceptem H_0.

i	y_i	$\widehat{F}_n(y_i)$	$\Phi(y_i)$
1	-2.46	0.04	0.0069
2	-2.11	0.08	0.0174
3	-1.23	0.12	0.1093
4	-0.99	0.16	0.1611
5	-0.42	0.20	0.3372
6	-0.39	0.24	0.3483
7	-0.21	0.28	0.4168
8	-0.15	0.32	0.4404
9	-0.10	0.36	0.4602
10	-0.07	0.40	0.4721
11	-0.02	0.44	0.4920
12	0.27	0.48	0.6064
13	0.40	0.52	0.6554
14	0.42	0.56	0.6628
15	0.44	0.60	0.6700
16	0.70	0.64	0.7580
17	0.81	0.68	0.7910
18	0.88	0.72	0.8106
19	1.07	0.76	0.8577
20	1.39	0.80	0.9177
21	1.40	0.84	0.9192
22	1.47	0.88	0.9292
23	1.62	0.92	0.9474
24	1.64	0.96	0.9495
25	1.76	1.00	0.9608

Taula 4.2: Valors observats

.——————————————————————————————————.

Prova de Kolmogorov-Smirnov per a dues mostres

Es tracta ara de decidir si les distribucions subjacents de dues mostres independents són o no les mateixes.

Sigui $X_1, \ldots, X_m$ una mostra d'una distribució contínua a$F(x)$ i $Y_1, \ldots, Y_n$ una mostra d'una altra distribució contínua $G(x)$.

Prenem el següent test d'hipòtesis:

$$\begin{cases} H_0 : F(x) = G(x) \text{ per a tot } x \\ H_1 : F(x) \neq G(x) \text{ per algun } x \end{cases}$$

Comencem construïnt les funcions de distribució empíriques per cada mostra: $F_m(x)$ i $G_n(x)$.

Es proposa l'estadístic:

$$D_{m\,n} = \sup_{-\infty<x<\infty} |\, F_m(x) - G_n(x)\, |$$

que rebutja H_0 si

$$\sqrt{\frac{m\,n}{m+n}}\, D_{m\,n} > c$$

on c és tal que

$$\lim_{\substack{n \to \infty \\ m \to \infty}} \mathbf{P}\left[\sqrt{\frac{m\,n}{m+n}}\, D_{m\,n} \leq c\right] = H(c) = 1 - \alpha$$

Com abans, si H_0 és cert, pel teorema de Glivenko-Cantelli, $D_{m\,n}$ tendirà a prendre valors petits.

4.1.3 Prova $\mathcal{X}^2$

Aquesta prova intenta determinar si podem acceptar la hipòtesi d'ajust d'una mostra a una distribució paramètrica determinada. Originalment, la prova correspon a un test de bondat d'ajust a dades categòriques amb distribució multinomial, de forma que per a cada categoria tenim una probabilitat teòrica associada. Aquesta prova s'estén de manera natural per a qualsevol distribució teòrica plantejada sempre que s'estableixin les categories (partició del suport de la variable) i es puguin calcular les probabilitats associades a cada categoria. Anem a veure'n alguns exemples:

- Disposem d'una mostra del nombre de cotxes aturats en una benzinera en diferents hores del dia. Voldríem determinar si aquesta mostra s'ajusta a un model de Poisson.
- A partir de les mesures de la longitud de les ales de certa espècie de papallones corresponent a una mostra de mida 30, volem determinar si la distribució de referència pot ser normal.

En el primer cas podem considerar com a categories cada valor de la variable, fins a un màxim raonable. En el segon cas, la variable és contínua i cal establir uns intervals i comptar el nombre d'ocurrències dins de cada interval.

Un cop categoritzada la mostra, calculem la probabilitat de cada categoria sota la hipòtesi nul·la del model teòric proposat i això ens permetria calcular els efectius esperats, sota la hipòtesi nul·la, de cada categoria.

En definitiva, doncs, el que nosaltres tenim és una mostra $X_1, \ldots, X_n$ d'una població X. Ens plantegem el següent test d'hipòtesis:

$$\begin{cases} H_0 : X \sim F_\theta \\ H_1 : X \text{ no es distribueix com } F_\theta \end{cases}$$

Notem que el paràmetre θ pot ser conegut però també pot ser desconegut.

El procediment que acabem de comentar el podem resumir en els següents punts:

1. Fem una partició de $\mathbb{R}$, o del rang de valors on X està definida, en k subintervals disjunts $I_1, I_2, \ldots, I_k$.
2. Calculem $p_j^0 = P\{\, X \in I_j \,|\, X \sim F_\theta \,\} \quad 1 \leq j \leq k$ i calculem també $n\, p_j^0$.
3. Sigui N_j el nombre d'observacions dins de la mostra amb valors dins de I_j, és a dir
$$N_j = \sum_{i=1}^{n} \mathbf{1}\,\{\, X_i \in I_j \,\}$$

4. Calculem
$$X^2 = \sum_{j=1}^{k} \frac{(N_j - n\,p_j^0)^2}{n\,p_j^0}$$
ja que $E(N_j \mid H_0) = n\,p_j^0$.

Hem de distingir dues situacions:

- La distribució de contrast està totalment especificada, és a dir, coneixem els valors dels paràmetres de la distribució de la hipòtesi nul·la.
- La distribució de contrast pertany a una classe coneguda, però els valors dels paràmetres són desconeguts, i caldrà doncs fer una estimació a partir de la mostra.

En ambdós casos, la distribució de l'estadístic X^2 (que fou proposat per Pearson l'any 1900), és asimptòticament una $\mathcal{X}^2$ però si bé en el primer cas els graus de llibertat són $k-1$ (nombre de classes que tenim menys 1), en el segon cas li hem de restar un més per cada paràmetre estimat.

Comentaris:

1. Els intervals I_j s'han d'escollir de forma que $n\,p_j^0$ siguin aproximadament iguals i assegurant que $n\,p_j^0 \geq 5$ (de vegades pot ser més petit).
2. Què fer quan el paràmetre θ és desconegut?

   En aquest cas el pas 1 anterior no es podrà realitzar i per tant la probabilitat serà del tipus $p_j^0(\theta)$.

Conèixer la distribució de l'estadístic X^2 sota la hipòtesi nul.la ens permet construir la regió crítica per decidir el test:

$$C = \{\, \underset{\sim}{X} : X^2 \geq \mathcal{X}^2_\alpha \,\} \quad \text{on} \;\; \alpha = \mathbf{P}(\underset{\sim}{X} \in C \mid H_0)$$

Exemple 4.2

Durant 3 mesos es compten el nombre de clients que utilitzen un caixer a primera hora del dia. Suposant independència entre els diferents dies, volem contrastar la hipòtesi que la distribució de la variable: nombre de clients que fan servir el caixer a primera hora del dia, segueix una distribució de Poisson de paràmetre 2.7.

Ens plantegem, doncs, el test següent:

$$\begin{cases} H_0 : X \sim & \text{Poisson}(2.7) \\ H_1 : X \not\sim & \text{Poisson}(2.7) \end{cases}$$

Disposem d'una mostra: $X_1, X_2, \ldots, X_{92}$, on $X_i \in \{0, 1, 2, \ldots\}$, els resultats de la qual representem a la taula 4.3.

k	N_j
0	5
1	10
2	25
3	27
4	14
5	8
més de 6	3

Taula 4.3: Valors de $N(j)$

Sota H_0 la taula d'efectius esperats es calcula en base a la probabilitat teòrica de la Poisson:

$$P(X_j = k) = e^{-2.7} \frac{2.7^k}{k!}$$

amb la qual cosa les diferents probabilitats i esperances són les que es mostren a la taula 4.4:

Amb totes aquestes dades ja podem calcular l'estadístic de Pearson:

$$X^2 = \sum_{j=1}^{7} \frac{(N_j - E_j)^2}{E_j} = \frac{(5 - 6.18)^2}{6.18} + \ldots + \frac{(3 - 5.22)^2}{5.22} = 6.40$$

Sota H_0 la distribució de l'estadístic serà una $\mathcal{X}^2$ amb $7 - 1 = 6$ graus de llibertat, i amb la següent regió crítica:

$$C = \{\, \underset{\sim}{X} : X^2 \geq \mathcal{X}^2_\alpha \,\} \quad \text{on } \mathcal{X}^2_\alpha \text{ és tal que } \mathbf{P}(\mathcal{X}^2_6 \geq \mathcal{X}^2_\alpha) = \alpha$$

Treballant amb $\alpha = 0.05$, tenim que

$$P\left(\mathcal{X}^2_6 \geq 12.59\right) = 0.05$$

k	$\mathbf{P}(X_j = k)$	$E_j = n\,\mathbf{P}(X_j = k)$
0	0.0672	6.18
1	0.1815	16.69
2	0.2450	22.54
3	0.2205	20.28
4	0.1488	13.69
5	0.0804	7.39
més de 6	0.0567	5.22

Taula 4.4: Valors esperats

i com que $X^2 = 6.40 < 12.59 = \mathcal{X}_\alpha^2$, acceptem H_0, la distribució proposada pot ser acceptada. Podem calcular el valor p de la prova:

$$\text{p-value } = P\left(\chi_6^2 \geq 6.40\right) = 1 - P\left(\chi_6^2 \leq 6.40\right) = 1 - 0.62 = 0.38$$

Exemple 4.3

Amb una mostra de 100 bombetes, mesurem el temps de funcionament ininterromput en mesos. Disposem de la mostra agrupada en forma de taula de freqüències (veure taula 4.5), i ens donen la mitjana de la mostra, que en aquest cas és $\overline{X} = 2.069$.

Sigui X_j la variable: Durada de la bombeta j-èssima, per tant tenim que $X_j \in \mathbb{R}^+$.

Ens estem plantejant el test d'hipòtesi:

$$\begin{cases} H_0 : X_j \sim exp\,(\lambda) \\ H_1 : X_j \nsim exp\,(\lambda) \end{cases}$$

X_j	N_j
$[0,1)$	31
$[1,2)$	30
$[2,3)$	13
$[3,4)$	10
$[4,5)$	6
$[5,\infty)$	10

Taula 4.5: Valors per $N(j)$

Com que no coneixem el valor del paràmetre λ, fem estimació pel mètode de la màxima versemblança. En el cas de l'exponencial,

$$\widehat{\lambda} = \overline{X} = 2.069$$

Calculem tot seguit la taula d'efectius esperats tenint en compte que si $X \sim exp\,(\lambda)$ llavors

$$P(X \leq x) = 1 - e^{-x/\lambda} \quad \text{i}$$

$$P(a \leq X \leq b) = \left(1 - e^{-b/\lambda}\right) - \left(1 - e^{-a/\lambda}\right)$$

valors, que, en el nostre cas concret, representem a la taula 4.6.

X_j	$\mathbf{P}(a \leq X \leq b)$	E_j
$[0,1)$	$0.3833 - 0.0000 = 0.3833$	38.3
$[1,2)$	$0.6196 - 0.3833 = 0.2364$	23.6
$[2,3)$	$0.7654 - 0.6196 = 0.1458$	14.6
$[3,4)$	$0.8553 - 0.7654 = 0.0899$	9
$[4,5)$	$0.9108 - 0.8553 = 0.0554$	5.5
$[5,\infty)$	$1.000 - 0.9108 = 0.0892$	8.9

Taula 4.6: Valors esperats

Per tant, tenim que l'estadístic de Pearson serà:

$$X^2 = \frac{(31 - 38.3)^2}{38.3} + \ldots + \frac{(10 - 8.9)^2}{8.9} = 3.5650$$

La distribució de referència és una $\mathcal{X}^2$ amb $k-1-1$ graus de llibertat (ja que hem estimat un paràmetre), és a dir, amb 4 graus de llibertat.

La regió crítica és la següent:

$$RC = \{\, \underset{\sim}{X} : X^2 \geq \mathcal{X}^2_\alpha \,\} \quad \text{on } \mathcal{X}^2_\alpha \text{ és tal que } \mathbf{P}(\, \mathcal{X}^2_4 \geq \mathcal{X}^2_\alpha \,) = \alpha$$

treballant, igual que abans, amb $\alpha = 0.05$, tenim que

$$P(\, \mathcal{X}^2_4 \geq 9.487 \,) = 0.05$$

i com que nosaltres hem obtingut $X^2 = 3.565 < 9.487 = \mathcal{X}^2_\alpha$, acceptem H_0, és a dir, no trobem evidències estadístiques significatives que indiquin que la distribució de referència no sigui una exponencial. Calculem també el valor p:

$$\text{p-value } = P\left(\chi^2_4 \geq 3.56\right) = 1 - P\left(\chi^2_4 \leq 3.56\right) = 1 - 0.53 = 0.47$$

.———————————————————————————————.

<u>Observacions</u>

En aquest test la distribució de l'estadístic és asimptòtica, amb la qual cosa requereix mides mostrals grans.

A més, la categorització de variables contínues en classes introdueix una certa arbitrarietat en el mètode. És important que els valors esperats no siguin massa petits (a la pràctica, haurien d'estar per sobre de 5), perquè la distribució de referència es veuria afectada per aquest fet. En aquests casos, convé agrupar classes.

4.2 Proves per dades categòriques

Són proves que permeten relacionar dues variables de tipus categòric. Aquesta situació es pot presentar en els següents casos:

Definició 4.3 *En una població es poden mesurar dues variables de tipus categòric A i B. Se selecciona una mostra de la població i per a cada individu* *de la mostra es mesuren les dues variables. Volem establir si hi ha relació entre aquestes dues variables, és a dir, si el coneixement del valor d'una d'elles fa canviar les probabilitats de l'altre per aquell individu. Això suposa un* **test d'independència** *de les dues variables en aquella població.*

Definició 4.4 *Existeixen k poblacions, en les quals es pot mesurar una variable categòrica. Es pretén establir si la distribució d'aquestes categories pot ser la mateixa per a totes les poblacions, en definitiva, el que es fa és un* **contrast d'homogeneïtat** *de les poblacions respecte a la característica que mesurem. Per això, caldrà seleccionar una mostra per a cada població i mesurar la variable en cada individu.*

Un exemple del primer cas seria que es vol analitzar en un institut si hi ha relació entre el fet de fer activitats extraescolars (cap, esportives, artístiques, complements) i la qualificació final (suspès, aprovat, notable, excel·lent). Se seleccionen a l'atzar 60 alumnes.

Per altra banda, un exemple del segon cas seria establir si la distribució del grup sanguini (O, A , B , AB) és homogènia en les poblacions de tres països (un europeu, un asiàtic i un africà).

4.2.1 Proves d'independència $\mathcal{X}^2$

Es considera la mostra representada per una taula de contingència amb R files i C columnes, on R i C són, respectivament, el nombre de classes de les variables A i B:

$$A \in \{ A_1, \ldots, A_R \} \quad B \in \{ B_1, \ldots, B_C \}$$

Es representa per $P_{ij} = \mathbf{P}\big(A = A_i \cap B = B_j \big)$ i les marginals per $P_{i+} = \mathbf{P}(A = A_i)$ i $P_{+j} = \mathbf{P}(B = B_j)$.

Aleshores es compleix que

$$P_{i+} = \sum_{j=1}^{C} P_{ij}, \quad P_{+j} = \sum_{i=1}^{R} P_{ij}$$

$$\sum_{i=1}^{C} \sum_{j=1}^{R} P_{ij} = 1$$

A la taula de contingència es comptabilitzen els efectius observats per a cada combinació de les dues classes:

	B_1	$\dots$	B_C	
A_1	O_{11}	$\dots$	O_{1C}	O_{1+}
.	.		.	.
.	.		.	.
.	.		.	.
A_R	O_{R1}	$\dots$	O_{RC}	O_{R+}
	O_{+1}	$\dots$	O_{+C}	

on $O_{ij} =$ nombre d'individus amb $A = A_i$ i $B = B_j$.

El planteig teòric del test d'independència seria:

$$\begin{cases} H_0 : P_{ij} = P_{i+} \cdot P_{+j} & \text{Les variables són independents} \\ H_1 : \text{No és cert que siguin independents} \end{cases}$$

Per a construir la taula d'efectius esperats, es pot considerar que el número corresponent seria $E_{ij} = N \cdot P_{ij}$ i que sota H_0 aquest valor seria $E_{ij} = N\big(P_{i+} \cdot P_{+j}\big)$. D'acord amb la mostra, els estimadors de P_{i+} i de P_{+j} són, respectivament:

$$\widehat{P}_{i+} = \frac{\sum\limits_{j=1}^{C} O_{ij}}{N} \qquad \widehat{P}_{+j} = \frac{\sum\limits_{i=1}^{R} O_{ij}}{N}$$

que són els valors de les marginals de les taules de contingència, dividits per la grandària de la mostra. La taula d'efectius esperats serà, doncs, la següent:

	B_1	$\dots$	B_C
A_1	$(O_{1+} \cdot O_{+1})/N$	$\dots$	$(O_{1+} \cdot O_{+C})/N$
.	.		.
.	.		.
.	.		.
A_R	$(O_{R+} \cdot O_{+1})/N$	$\dots$	$(O_{R+} \cdot O_{+C})/N$

L'estadístic que permet decidir el test és l'estadístic de Pearson:

$$X^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{\left(O_{ij} - \frac{O_{i+} \cdot O_{+j}}{N}\right)^2}{\frac{O_{i+} \cdot O_{+j}}{N}}$$

Si H_0 és cert, la distribució asimptòtica de l'estadístic X^2 de Pearson és una $\mathcal{X}^2$ amb $(R-1)\times(C-1)$ graus de llibertat, ja que el nombre de paràmetres estimats són $R-1$ per les probabilitats de A i $C-1$ per les probabilitats associades a B.

Rebutgem la hipòtesi nul·la si l'estadístic X^2 és prou gran:

$$C = \{\, \underset{\sim}{X} : X^2 > \mathcal{X}^2_\alpha \,\}$$

seria la regió crítica, on $\mathcal{X}^2_\alpha$ és tal que $P(\, \underset{\sim}{X} \in C \mid H_0\,) = \alpha$.

Exemple 4.4

Volem establir si hi ha relació entre l'hàbit de fumar (A) i el fet de patir trastorns de la son (B). Amb una mostra de 150 individus la taula de contingència resultant és:

		B		
		no	si	total
	molt	7	30	37
A	moderadament	12	36	48
	gens	29	36	65
		48	102	150

El test d'hipòtesi seria en aquest cas:

$$\begin{cases} H_0 : A \text{ i } B \text{ són independents} \\ H_1 : A \text{ i } B \text{ no són independents} \end{cases}$$

La taula esperada sota $H_0 : E_{ij} = \frac{O_{i+}\,O_{+j}}{N}$ seria:

	B		
A	*no*	*si*	*total*
molt	$(37\times 48)/150 = 11.84$	$(37\times 102)/150 = 25.16$	37
moderad.	$(48\times 48)/150 = 15.36$	$(48\times 102)/150 = 32.6$	48
gens	$(65\times 48)/150 = 20.8$	$(65\times 102)/150 = 4.42$	65
	48	102	

L'estadístic serà:

$$X^2 = \frac{(7-11.84)^2}{11.84} + \ldots + \frac{(36-44.2)^2}{44.2} = 8.744$$

que segueix una distribució $\mathcal{X}^2_{(3-1)\times(2-1)} = \mathcal{X}^2_2$, i amb una regió crítica de la forma:

$$C = \{\, \underset{\sim}{X} : X^2 > \mathcal{X}^2_\alpha \,\}$$

on $P(\,\mathcal{X}^2_2 > \mathcal{X}^2_\alpha\,) = \alpha$.

Si prenem $\alpha = 0.05$, llavors, $\mathcal{X}^2_\alpha = 5.99$ i per tant $X^2 = 8.744 > 5.99 = \mathcal{X}^2_{2,\alpha}$.

Calculem el valor p:

$$\text{p-value} = P\left(\chi^2_2 \geq 8.744\right) = 1 - P\left(\chi^2_2 \leq 8.744\right) = 0.0126$$

Rebutgem H_0 i podem concloure que existeix algun grau de dependència entre els dos factors amb $\alpha = 0.05$.

4.2.2 Proves d'homogeneïtat

Considerem R poblacions $\{\,\Omega_1, \ldots, \Omega_R\,\}$ i de cadascuna d'elles seleccionem una m.a.s. i a cada individu mesurem el valor d'una certa variable categòrica $\{\,B_1, \ldots, B_C\,\}$ que té C nivells. En aquest cas, cada fila de la taula de contingència correspon a la taula de freqüències de la variable per a cada població. El que s'intenta aquí és establir si les probabilitats són homogènies per a cada població, és a dir, ens estem plantejant el següent test d'hipòtesi:

$$\begin{cases} & P(B_1\,|\,\Omega_1) \quad = P(B_1\,|\,\Omega_2) \quad = \ldots \quad = P(B_1\,|\,\Omega_R) \quad = P(B_1) \\ H_0 : & \quad\vdots \qquad\qquad\qquad \vdots \qquad\qquad\qquad\qquad\quad \vdots \\ & P(B_C\,|\,\Omega_1) \quad = P(B_C\,|\,\Omega_2) \quad = \ldots \quad = P(B_C\,|\,\Omega_R) \quad = P(B_C) \\ H_1 : \text{No són homogènies} & \end{cases}$$

Per tal de construir la taula d'efectius esperats s'utilitza la següent estimació per a les probabilitats que sota H_0 són comunes a totes les poblacions:

$$\widehat{p}_{ij} = \frac{O_{i+}}{N}$$

	B_1	$\ldots$	B_C	
Ω_1	$(O_{1+} \times O_{+1})/N$	$\ldots$	$(O_{1+} \times O_{+C})/N$	O_{1+}
$\vdots$	$\vdots$		$\vdots$	$\vdots$
Ω_R	$(O_{R+} \times O_{+1})/N$	$\ldots$	$(O_{R+} \times O_{+C})/N$	O_{R+}
	O_{+1}	$\ldots$	O_{+C}	N

L'estadístic coincideix amb el corresponent a les proves d'independència:

$$X^2 = \sum_{i=1}^{R} \sum_{j=1}^{C} \frac{\left(O_{ij} - \frac{O_{i+} \cdot O_{+j}}{N} \right)^2}{\frac{O_{i+} \cdot O_{+j}}{N}}$$

amb la mateixa distribució asimptòtica de referència: $\mathcal{X}^2_{(R-1)\times(C-1)}$.

Exemple 4.5

Volem establir si la utilització de diferents sistemes operatius té una distribució diferent en dues facultats de la Universitat. Es seleccionen 50 alumnes de cada centre i se'ls demana l'entorn informàtic que fan servir preferentment. Els resultats els presentem a la taula següent:

	windows	unix	mac	
informàtica	27	19	4	50
telecomunicacions	20	22	8	50
	47	41	12	

Ens plantegem el següent test d'hipòtesi:

$$\begin{cases} H_0 : \text{Les distribucions d'ús són homogènies en els dos centres} \\ H_1 : \text{Les distribucions no són homogènies} \end{cases}$$

i la taula esperada sota la hipòtesi nul·la H_0 seria:

i l'estadístic seria:

$$X^2 = \frac{(27 - 23.5)^2}{23.5} + \ldots + \frac{(8-6)^2}{6} = 2.595$$

	windows	unix	mac
informàtica	$(47 \times 50)/100 = 23.5$	$(41 \times 50)/100 = 20.5$	$(12 \times 50)/100 = 6$
telecomunicacions	$(47 \times 50)/100 = 23.5$	$(41 \times 50)/100 = 20.5$	$(12 \times 50)/100 = 6$

estadístic que segueix una distribució: $\mathcal{X}^2_{(3-1)\times(2-1)} = \mathcal{X}^2_2$, amb una regió crítica

$$C = \{ \underset{\sim}{X} : X^2 > \mathcal{X}^2_\alpha \}$$

on $\mathbf{P}(\mathcal{X}^2_2 > \mathcal{X}^2_\alpha) = \alpha$.

Si prenem com abans, $\alpha = 0.05$, aleshores $\mathcal{X}^2_{2,\alpha} = 5.99$, i llavors,

$$X^2 = 2.595 < 5.99 = \mathcal{X}^2_\alpha$$

i per tant acceptem H_0 i podem concloure que les distribucions als dos centres són homogènies amb $\alpha = 0.05$, ja que

$$\text{p-value } = P\left(\chi^2_2 \geq 2.595\right) = 1 - P\left(\chi^2_2 \leq 2.595\right) = 0.2732$$

4.3 Mètodes no paramètrics

4.3.1 Prova dels signes

Del que es tractarà serà de comparar dos mètodes amb l'objectiu de saber quin dels dos és més eficient.

Suposem que disposem de dues drogues A i B, que s'administren a pacients per tal de saber quina de les dues estimula una activitat enzimàtica més gran.

Escollirem 2 conjunts de pacients i els aparellarem d'acord amb llurs característiques, fent aquests parells tan homogenis com sigui possible.

És factible de pensar que la distribució de la resposta de cada pacient és diferent, és a dir, que no tenim unes dades idènticament distribuïdes.

L'el·lecció de quin pacient pren la droga A i quin pren la droga B, ha de fer-se aleatòriament. Si les dues drogues són igual d'efectives, a cada parell el pacient que rep la droga A té la mateixa probabilitat que el pacient que rep la droga B d'exhibir una resposta més gran. D'altra banda, si la droga A és més efectiva que la droga B, a cada parell el pacient que rep la droga A té una probabilitat més gran que $1/2$ d'exhibir un valor més gran que no pas l'altre pacient.

També podríem tenir experiments amb només n pacients de forma tal que cada pacient rep els dos medicaments. És important aleshores aleatoritzar l'ordre dels medicaments.

Prova dels signes

Per $i = 1, 2, \ldots, n$, sigui (X_i, Y_i) un parell de respostes del parell i-èssim, on

X_i és la resposta del pacient que rep la droga A.

Y_i és la resposta del pacient que rep la droga B.

Sigui $p_i = \mathbf{P}(X_i > Y_i)$ i suposarem que $\mathbf{P}(X_i = Y_i) = 0$.

Suposem també que $p_1 = p_2 = \ldots = p_m = p$.

Ens plantegem el següent test d'hipòtesi:

$$\begin{cases} H_0 : p \leq \frac{1}{2} & \text{La droga A no és més efectiva que la droga B} \\ H_1 : p > \frac{1}{2} & \text{La droga A és més efectiva que la droga B} \end{cases}$$

Els n parells (X_i, Y_i) representen n assajos de Bernoulli amb

$$p = \mathbf{P}(X > Y)$$

Denotem per S

$$S = \#\{i : X_i > Y_i\} = \sum_{i=1}^{n} \mathbf{1}_{(X_i > Y_i)} \sim Bin(n, p)$$

$$= \#\{i : X_i - Y_i > 0\}$$

Si H_0 és cert, aleshores $S \sim Bin(n, 1/2)$, i rebutgem H_0 si $S > c$.

En una prova amb nivell α, escollim c tal que

$$\mathbf{P}\{Bin(n, 1/2) > c\} \geq \alpha$$

Per exemple, si tenim 15 parells de pacients i observem que en 11 d'ells la droga A és més efectiva que no pas la B, llavors tenim que

$$S = 11 \quad \text{i} \quad n = 15$$

i per tant

$$\mathbf{P}_{H_0}\{\, S \geq 11 \,\} = 0.0417 + 0.0139 + 0.0032 + 0.0005 + 0.0000 = 0.0593$$

Podem rebutjar, doncs la hipòtesi nul·la H_0 amb un risc de $\alpha > 0.0593$. La importància d'aquest test es posa de manifest en aquelles aplicacions on no podem quantificar numèricament la resposta d'un medicament.

4.3.2 Prova dels rangs signats de Wilcoxon

Continuem en aquesta secció amb dues drogues administrades a $2n$ pacients que han estat aparellats i suposem que hem mesurat en la mateixa escala i amb unitats apropiades la resposta de cada pacient.

Per $i = 1, 2, \ldots, n$, siguin

X_i la resposta del pacient del parell i que rep la droga A.

Y_i la resposta del pacient del parell i que rep la droga B.

Aleshores definim $D_i = X_i - Y_i$, amb la qual cosa tenim que $D_1, D_2, \ldots, D_n$ són variables aleatòries independents i suposem que també són idènticament distribuïdes (amb distribució contínua).

Per tant, tenim que $\mathcal{D} \sim \mathcal{F}$ i a més $f(x)$ és simètrica al voltant de la mediana θ, és a dir, $f(\theta - x\,) = f(\theta + x\,)$.

Sigui H_0 la següent hipòtesi: La droga A no és més efectiva que la droga $B \Longleftrightarrow H_0 : \mathbf{P}(\, \mathcal{D} \leq 0\,) \geq \frac{1}{2} \Longleftrightarrow H_0 : \theta \leq 0$.

Així, doncs, tenim el test d'hipòtesi:

$$\begin{cases} H_0 : \theta \leq 0 \Longleftrightarrow \mathbf{P}(\, \mathcal{D} \leq 0\,) \geq \frac{1}{2} \\ H_1 : \theta > 0 \Longleftrightarrow \mathbf{P}(\, \mathcal{D} \leq 0\,) < \frac{1}{2} \Longleftrightarrow \quad X \text{ tendeix a ser més gran que } Y \end{cases}$$

Prova de Wilcoxon (1945)

Calculem en primer lloc $D_1, D_2, \ldots, D_n$ i ordenarem de més petit a més gran les quantitats $|\, D_1\,|, |\, D_2\,|, \ldots, |\, D_n\,|$.

Suposem que $|\, D_i\,| \neq |\, D_j\,|$ per a tot $i \neq j$ i que $|\, D_i\,| \neq 0$.

Assignem rangs $1, 2, \ldots, n$ de la forma següent: A cada $|\, D_i \,|$ li correspon un rang i i R_i el signe de la diferència.

Tot seguit definim

$$\begin{aligned} S_n &= \text{ la suma dels rangs amb signe } + = \sum_{i=1}^{n} R_i \, \mathbf{1}\{\, D_i > 0 \,\} \\ &= \sum_{i=1}^{n} i \, \mathbf{1}\{\, D_j > 0 \,\} \end{aligned}$$

on i és tal que $j : R_j = i$.

La prova que utilitzarem es basarà en S_n.

Exemple 4.6

Suposem que tenim 6 parells:

i	X_i	Y_i	D_i	$\|\, D_i \,\|$	R_i	R_i+
1	3.84	3.03	0.81	0.81	2	+2
2	6.27	4.91	1.36	1.36	4	+4
3	8.75	7.65	1.10	1.10	3	+3
4	4.39	5.00	−0.61	0.61	1	−1
5	9.24	7.42	1.82	1.82	5	+5
6	5.61	7.59	−1.98	1.98	6	−6

i per tant tenim que

$$S_n = 2 + 4 + 3 + 5 = 14$$

Distribució de S_n

Sota $H_0 : \theta \leq 0$, tenim el següent:

Si $\theta = 0 \iff A$ i B són igual d'efectives i els rangs tenen la mateixa probabilitat de tenir signe + o -, i a més les n assignacions de rangs són independents.

Si $\theta < 0$, la droga A és menys efectiva que no pas la B, i per tant tenim més probabilitat d'assignar signes - que +, i S_n tendirà a prendre valors més petits.

Sota $H_1 : \theta > 0$, A és més efectiva que B, veurem, doncs, més signes + que - i S_n tendirà a prendre valors grans.

Rebutjarem H_0 quan $S_n \geq c$.

<u>Mitjana i variança de S_n quan $\theta = 0$</u>

Per a cada $i = 1, 2, \ldots, n$, es defineix

$$W_i = \begin{cases} 1 & \text{si} \quad R_i \ \text{rep signe } + \\ 0 & \text{si} \quad R_i \ \text{rep signe } - \end{cases} = \begin{cases} 1 & \text{si} \quad R_j = i \ \text{ i } \ D_j > 0 \\ 0 & \quad \text{altrament} \end{cases}$$

Aleshores

$$S_n = \sum_{i=1}^{n} i\, W_i = \sum_{i=1}^{n} i\, \mathbf{1}\{ D_j > 0 \}$$

i sota $\theta = 0$, $\mathbf{P}(W_i = 0) = \mathbf{P}(W_i = 1) = 1/2 \longrightarrow \begin{cases} E(W_i) & = & 1/2 \\ Var(W_i) & = & 1/4 \end{cases}$

A més, també tenim que $W_1, \ldots, W_n$ són independents, amb la qual cosa

$$E(S_n) = \sum_{i=1}^{n} i\, E(W_i) = \frac{1}{2} \sum_{i=1}^{n} i = \frac{1}{2}\frac{1}{2}\, n\,(n + 1) = \frac{n\,(n+1)}{4}$$

$$Var(S_n) = \sum_{i=1}^{n} i^2\, Var(W_i) = \frac{1}{4} \sum_{i=1}^{n} i^2 = \frac{1}{4}\frac{1}{6}\, n\,(n+1)\,(2\,n+1) = \frac{n\,(n+1)\,(2\,n+1)}{24}$$

Per a valors de n grans (és a dir, quan $n \geq 20$) tenim a més que

$$\frac{S_n - E(S_n)}{\sqrt{Var(S_n)}} = \frac{S_n - \frac{n\,(n+1)}{4}}{\sqrt{\frac{n\,(n+1)\,(2\,n+1)}{24}}} \quad \overset{n\to\infty}{\longrightarrow} \quad N(0, 1)$$

Per exemple, si prenem $n = 6$,

$$\frac{S_n - \frac{6\times 7}{4}}{\sqrt{\frac{6\times 7\times 13}{24}}} = \frac{S_n - 10.5}{\sqrt{22.75}} = \frac{S_n - 10.5}{4.77} = 0.73$$

per ser en aquest cas $S_n = 14$.

Podem calcular el valor p de la prova com

$$P(S_n \geq 14) \approx P\left(\frac{S_n - 10.5}{\sqrt{22.75}} \geq \frac{13.5 - 10.5}{\sqrt{22.75}}\right) = P(Z \geq 0.6290) =$$

$$= 1 - P(Z \leq 0.6290) = 0.2646$$

Per altra banda, per a valors de n petits, tenim el següent:

Sota $\theta = 0$, el rang k-èssim tant pot ser + com -, i qualsevol distribució de + i - entre els n valors és igualment probable. Hi ha 2^n formes d'assignar aquests signes.

Tenim, per tant, 2^n possibles valors de S_n (no tots diferents), i cadascun d'ells és igualment probable.

Suposem, per exemple, que $n = 3$, per tant, $2^3 = 8$.

Tenim la següent taula de possibles resultats:

rang								
1	+	+	+	−	+	−	−	−
2	+	+	−	+	−	+	−	−
3	+	−	+	+	−	−	+	−
S_n	6	3	4	5	1	2	3	0
P	1/8	1/8	1/8	1/8	1/8	1/8	1/8	1/8

amb la qual cosa tenim la següent taula de probabilitats:

S_n	$\mathbf{P}(S_n = x)$
0	1/8
1	1/8
2	1/8
3	1/4
4	1/8
5	1/8
6	1/8

(veure taula 9 de la pàgina 574 de (**?**))

4.3.3 Prova de Wilcoxon, Mann i Whitney

Considerem ara el problema en que una mostra aleatòria de m observacions $X_1, \ldots, X_m$ es pren d'una distribució contínua amb funció de densitat $f(x)$ i una altra mostra $Y_1, \ldots, Y_n$ es pren d'una altra distribució amb funció de densitat $g(x)$.

Suposem que la distribució de X_i coincideix amb la de Y_i, o bé que $\exists\ \theta$: la distribució de $Y_i + \theta$ coincideix amb la de X_i.

Així, doncs, tenim el següent test d'hipòtesi:

$$\begin{cases} H_0 : f(x) = g(x) & -\infty < x < \infty \\ H_1 : \exists\ \theta : f(x) = g(x - \theta) & -\infty < x < \infty \end{cases}$$

Per a resoldre aquestes hipòtesis podríem utilitzar una prova $\mathcal{X}^2$, una prova de Kolmogorov-Smirnov o bé mètodes basats en la hipòtesi de distribució normal de X i Y.

No obstant tot això, ara presentarem un mètode que fou proposat per Wilcoxon i per Mann-Whitney l'any 1940.

Procediment

Ordenem les $m+n$ observacions de les dues mostres de la més petita a la més gran. Suposem que no hi ha empats. De fet, el que podem suposar és que $\mathbf{P}(X=Y)=0$ i $\mathbf{P}(X_i=Y_j)=0$.

D'aquesta manera, doncs, obtenim $m+n$ valors ordenats, i assignem un rang.

Si H_0 és certa, és a dir, si $X_1,\ldots,X_m$ i $Y_1,\ldots,Y_n$ estan tretes de la mateixa distribució, els rangs corresponents a $X_1,\ldots,X_m$ estaran distribuïts (dispersos) entre els $m+n$ valors, i no estaran concentrats ni en els valors més petits ni els més grans.

Per tant, sota la hipòtesi nul·la H_0, els rangs assignats a $X_1,\ldots,X_m$ serien els mateixos que si haguéssim tret m rangs sense reemplaçament d'una caixa que contingui els rangs $1,2,\ldots,m+n$.

Sigui S la suma dels rangs assignats a les m observacions $X_1,\ldots,X_m$. Aleshores

$$E(S)=E\Big(\sum_{i=1}^{m}R(X_i)\Big)=m\,E\{R(X_i)\}=m\,\frac{1}{m+n}\,\Big[\frac{1}{2}\,(m+n)\,(m+n+1)\Big]=$$

$$\frac{m\,(m+n+1)}{2}$$

i per altra banda,

$$Var(S)=\frac{m\,n\,(m+n+1)}{12}$$

tenint en compte que estem sota la hipòtesi H_0.

Tenim ara $m+n$ unitats experimentals per assignar a un grup de tractament i a un grup de control.

L'assignació es fa a l'atzar: m unitats s'assignen al tractament i n al control.

Estem interessats en un test de la hipòtesi nul·la del qual és en aquest cas H_0 i diu que el tractament no té cap efecte.

Si H_0 és cert, qualsevol diferència que s'observi es deu a l'aleatorització.

Un estadístic per aquest mètode es calcula de la forma que segueix: Suposant, com abans, que no es produeixen empats, ordenem les $n+m$ observacions de menor a major i assignem el corresponent rang (un valor entre 1 i $m+n$). Tot seguit calculem la suma dels rangs R de les observacions del grup de control. Si aquesta suma és o bé molt petita o bé molt gran, rebutjarem H_0.

Per tal de trobar els valors crítics que defineixen la regió de rebuig, necessitem calcular la distribució del nostre estadístic en cas que la hipòtesi nul·la sigui certa. En aquest cas, si H_0 és certa, qualsevol assignació de rangs a les $m+n$ observacions és igualment probable. Hi ha $(m+n)!$ assignacions diferents, i per a cadascuna d'elles l'estadístic (la suma de rangs del grup de control) es pot calcular. Això dóna una llista de $(m+n)!$ possibles valors (no tots diferents) de l'estadístic, i cadascun amb probabilitat $1/(m+n)!$.

Exemple 4.7

Prenem $m=2$, $n=2$, i per tant, $m+n=4$ i $4!=24$. Llavors, la taula de rangs i dels possibles valors per S són:

R												
1	C_1	C_1	C_2	C_2	C_1	C_1	C_2	C_2	C_1	C_1	T_1	T_1
2	C_2	C_2	C_1	C_1	T_1	T_2	T_1	T_2	T_1	T_2	C_1	C_2
3	T_1	T_2	T_1	T_2	C_2	C_2	C_1	C_1	T_2	T_1	C_2	C_1
4	T_2	T_1	T_2	T_1	T_2	T_1	T_2	T_1	C_2	C_2	T_2	T_2
S	3	3	3	3	4	4	4	4	5	5	5	5

R												
1	C_2	C_2	T_2	T_2	T_1	T_1	T_2	T_2	T_1	T_1	T_2	T_2
2	T_1	T_2	C_1	C_2	C_1	C_2	C_1	C_2	T_2	T_2	T_1	T_1
3	T_2	T_1	C_2	C_1	T_2	T_2	T_1	T_1	C_1	C_2	C_1	C_2
4	C_1	C_1	T_1	T_1	C_2	C_1	C_2	C_1	C_2	C_1	C_2	C_1
S	5	5	5	5	6	6	6	6	7	7	7	7

D'aquesta manera, doncs, tenim la següent taula de probabilitats:

S	**P**
3	4/ 24
4	4/ 24
5	8/ 24
6	4/ 24
7	4/ 24

Les taules de la distribució nul.la que mostren la suma dels rangs estan extensament disponibles i varien en format. Notem que donat que la suma de les dues sumes de rangs és la suma dels enters entre 1 i $m+n$, que precisament és

$$\sum_{i=1}^{m+n} i = \frac{1}{2}(m+n)(m+n+1)$$

coneixent una suma de rangs automàticament coneixem l'altra.

Algunes taules es donen en termes de la suma de rangs del grup més petit, i d'altres en termes de la més petita de les sumes de rangs. Així, per exemple, la taula 8 de l'apèndix del llibre (**?**) (pag. 571) aprofita a més a més simetries addicionals.

Sigui n_1 la grandària més petita, és a dir,

$$n_1 = \min(n, m)$$

També definim n_2 com

$$n_2 = \max(n, m)$$

Per últim, considerem R com la suma dels rangs de la mostra de grandària n_1.

Aleshores definim

$$\begin{aligned} R' &= n_1(m+n+1) - R \\ R^* &= \min(R, R') \end{aligned}$$

Doncs la taula de la qual estem parlant ens dóna els valors crítics per R^*.

Rebutgem H_0 si l'estadístic està en la cua de la distribució nul·la, és a dir, si

$$R \geq c_1 \quad \text{ó} \quad R \leq c_2$$

on $\mathbf{P}_{H_0}\{ R \geq c_1 \text{ ó } R \leq c_2 \} = \alpha$.

Exemple 4.8

Anem a il·lustrar aquesta prova referint-nos a les dades sobre la fusió del gel. Les mides de les mostres són bastant petites ($m = 13, n = 8$), així que en absència de qualsevol coneixement anterior que faci referència a la suficiència de l'ús d'una distribució normal, el que semblaria més segur seria utilitzar un mètode no paramètric. La següent taula dóna a conèixer els rangs donats per les mesures i per a cadascun dels dos mètodes.

Mètode A	Mètode B
7.5	11.5
19.0	1.0
11.5	7.5
19.0	4.5
15.5	4.5
15.5	15.5
19.0	2.0
4.5	4.5
21.0	
15.5	
11.5	
9.0	
11.5	

El test d'hipòtesi que ens estem plantejant és el següent:

$$\begin{cases} H_0 : \text{Mètode } A = \text{Mètode } B \\ H_1 : \text{Mètode } A \text{ superior al Mètode } B \end{cases}$$

Rebutjarem H_0 si $R = \#$ rangs del Mètode B $\leq c_2$.

En aquest exemple concret, tenim que

$$\left.\begin{array}{rcl} R & = & 51 \\ R' & = & 8 \times (8 + 13 + 1) - 51 = 125 \end{array}\right\} \longrightarrow R^* = 51$$

De la taula 8 de l'apèndix B, també del llibre (**?**) podem extreure que, per $n_1 = 8$ i $n_2 = 13$,

Per un test bilateral	Per un test unilateral	Valor crític per R^*
0.20	0.100	69
0.10	0.050	64
0.05	0.025	60
0.01	0.005	53

Veiem, doncs, que el valor crític per un test bilateral amb $\alpha = 0.01$ és 53, mentre per $\alpha = 0.05$ és 60. Com que $51 \leq 53$, rebutgem H_0 amb nivell 1%.

Capítol 5

Distribucions de Probabilitat

Distribucions continues

Beta(α, β)

$f(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, \quad 0 \leq x \leq 1, \quad \alpha, \beta > 0$

$E(X) = \frac{\alpha}{\alpha+\beta}, \quad V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$M(t) = 1 + \sum_{k=1}^{\infty} (\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r}) \frac{t^k}{k!}$

Cauchy(θ, σ)

$f(x) = \frac{1}{\pi\sigma} \frac{1}{1+(\frac{x-\theta}{\sigma})^2}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty, \quad \sigma > 0.$

$E(X)$ **n.a.** , $\quad V(X)$ **n.a.**

$M(t)$ **n.a.**

$\chi^2(n)$

$f(x) = \frac{1}{\Gamma(n/2)2^{n/2}} x^{n/2-1} e^{-x/2}, \quad 0 \leq x < \infty, \quad n = 1, 2, \ldots$

$E(X) = n, \quad V(X) = 2 \cdot n$

$M(t) = \left(\frac{1}{1-2t}\right)^{n/2}$

Double exponential(μ, σ)

$f(x) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma} \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$

$E(X) = \mu, \quad V(X) = 2\sigma^2$

$M(t) = \frac{e^{\mu t}}{1-(\sigma t)^2}$

Exponential(λ)

$f(x) = \lambda e^{-\lambda x}, \quad x > 0, \quad \lambda > 0$

$E(X) = \frac{1}{\lambda}, \quad V(X) = \frac{1}{\lambda^2}$

$M(t) = \frac{1}{1-t/\lambda}$

$\varphi(t) = \frac{\lambda}{\lambda - it}$

Exponential (alternative parametrization)

$f(x) = \frac{1}{\beta} e^{-x/\beta}, \quad x \geq 0, \quad \beta > 0$

$E(X) = \beta, \quad V(X) = \beta^2$

$M(t) = \frac{1}{1-\beta t}, \quad t < \frac{1}{\beta}$

F(ν_1, ν_2)

$f(x) = \frac{\Gamma(\frac{\nu_1+\nu_2}{2})}{\Gamma(\frac{\nu_1}{2})\Gamma(\frac{\nu_2}{2})} \cdot \left(\frac{\nu_1}{\nu_2}\right)^{\nu_1/2} \cdot \frac{x^{(\nu_1-2)/2}}{(1+\frac{\nu_1}{\nu_2}x)^{(\nu_1+\nu_2)/2}}, \quad x \geq 0, \quad \nu_1, \nu_2 = 1, 2, \ldots$

$E(X) = \frac{\nu_2}{\nu_2-2}, \quad \nu_2 > 2, \quad V(X) = 2 \cdot \left(\frac{\nu_2}{\nu_2-2}\right)^2 \cdot \frac{\nu_1+\nu_2-2}{\nu_1(\nu_2-4)}, \quad \nu_2 > 4$

$M(t)$ **n.a.**

Gamma(α, β)

$f(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} \quad x > 0, \quad \alpha, \beta > 0$

$E(X) = \alpha\beta \quad V(X) = \alpha\beta^2$

$M(t) = \left(\frac{1}{1-\beta t}\right)^\alpha$

Gamma (alternative parametrization)

$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$

$E(X) = \frac{\alpha}{\beta} \quad V(X) = \frac{\alpha}{\beta^2}$

$M(t) = \left(\frac{1}{1-t/\beta}\right)^\alpha$

Inverse Gamma(α, β)

$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{-\alpha-1} e^{-\beta/x} \quad x > 0, \quad \alpha, \beta > 0$

$E(X) = \frac{\beta}{\alpha-1} \quad V(X) = \frac{\beta^2}{(\alpha-1)^2(\alpha-2)}$

Logistic(μ, β)

$f(x) = \frac{1}{\beta}\frac{e^{-(x-\mu)/\beta}}{(1+e^{-(x-\mu)/\beta})^2} \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \beta > 0$

$E(X) = \mu, \quad V(X) = \frac{\pi^2\beta^2}{3}$

$M(t) = e^{\mu t}\Gamma(1-\beta t)\Gamma(1+\beta t), \quad |t| < \frac{1}{\beta}$

Lognormal(μ, σ^2)

$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\frac{e^{-(\log x-\mu)^2/(2\sigma^2)}}{x}, \quad 0 \le x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$

$E(X) = e^{\mu + \frac{\sigma^2}{2}}, \quad V(X) = e^{2(\mu+\sigma^2)} - e^{2\mu+\sigma^2}$

$M(t) =$ **n.a.**

Normal(μ, σ^2)

$f(x) = \frac{1}{\sqrt{2\pi}\sigma}e^{\frac{-(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty, \quad -\infty < \mu < \infty, \quad \sigma > 0$

$E(X) = \mu, \quad V(X) = \sigma^2$

$M(t) = e^{\mu t + \sigma^2 t^2/2}$

$\varphi(t) = e^{it\mu - \frac{1}{2}\sigma^2 t^2}$

Pareto(α, β)

$f(x) = \frac{\beta\alpha^\beta}{x^{\beta+1}}, \quad \alpha < x < \infty, \quad \alpha > 0, \quad \beta > 0$

$E(X) = \frac{\beta\alpha}{\beta-1}, \quad \beta > 1. \quad V(X) = \frac{\beta\alpha^2}{(\beta-1)^2(\beta-2)}, \quad \beta > 2$

$M(t)$ **n.a**

---

t(ν)

$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \frac{1}{(1+\frac{x^2}{\nu})^{(\nu+1)/2}} \quad -\infty < x < \infty, \quad \nu = 1, \ldots$

$E(X) = 0, \quad \nu > 1, \quad V(X) = \frac{\nu}{\nu-2}, \quad \nu > 2$

$M(t)$ **n.a.**

Uniform(a, b)

$f(x) = \frac{1}{b-a}, \quad a \leq x \leq b$

$E(X) = \frac{b+a}{2}, \quad V(X) = \frac{(b-a)^2}{12}$

$M(t) = \frac{e^{bt}-e^{at}}{(b-a)t}$

Weibull(γ, β)

$f(x) = \frac{\gamma}{\beta} x^{\gamma-1} e^{-x^{\gamma}/\beta}, \quad 0 < x < \infty, \quad \gamma > 0, \quad \beta > 0$

$E(X) = \beta^{1/\gamma}\Gamma(1 + \frac{1}{\gamma})$**,** $\quad V(X) = \beta^{2/\gamma}\left(\Gamma(1 + \frac{2}{\gamma}) - \Gamma^2(1 + \frac{1}{\gamma})\right)$

Distribucions discretes

Bernoulli(p)

$P(X = x) = p^x(1-p)^{1-x}, \quad x = 0, 1; \quad 0 \leq p \leq 1$

$E(X) = p, \quad V(X) = p(1-p)$

$M(t) = (1-p) + pe^t$

Binomial(n, p)

$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}; \quad x = 0, 1, \ldots, n; \quad 0 \leq p \leq 1$

$E(X) = np, \quad V(X) = np(1-p)$

$M(t) = (pe^t + (1-p))^n$

$\varphi(t) = \left(1 - p + pe^{it}\right)^n$

Discrete Uniform

$P(X = x \mid N) = \frac{1}{N} \quad x = 1, \ldots, N$

$E(X) = \frac{N+1}{2}, \quad V(X) = \frac{(N+1)(N-1)}{12}$

$M(t) = \frac{1}{N} \sum_{i=1}^{N} e^{it}$

Geometric(p)

$P(X = x) = p(1-p)^{x-1}, \quad 0 \leq p \leq 1, \quad x = 1, \ldots \quad 0 \leq p \leq 1$

$E(X) = \frac{1}{p}, \quad V(X) = \frac{1-p}{p^2}$

$M(t) = \frac{pe^t}{1-(1-p)e^t}$

$\varphi(t) = \frac{pe^{it}}{1-(1-p)e^{it}}$

Geometric(p) (alternative formulation)

$P(X = x) = p(1-p)^x, \quad 0 \leq p \leq 1, \quad x = 0, \ldots \quad 0 \leq p \leq 1$

$E(X) = \frac{1-p}{p}, \quad V(X) = \frac{1-p}{p^2}$

$M(t) = \frac{p}{1-(1-p)e^t}$

$\varphi(t) = \frac{p}{1-(1-p)e^{it}}$

Hypergeometric

$P(X = x) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$

$E(X) = \frac{KM}{N}, \quad V(X) = \frac{KM}{N}\frac{(N-M)(N-K))}{N(N-1)}$

Multinomial

$P(\mathbf{X} = \mathbf{x}) = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k} \quad x_i = 0, 1, \ldots \quad x_i + \cdots x_k = n$

$E(X_i) = np_i, \quad V(X_i) = np_i(1-p_i), \quad Cov(X_i, X_j) = -np_i p_j$

Negative Binomial

$P(X = x) = \binom{r+x-1}{x} p^r (1-p)^x; \quad x = 0, 1, \ldots; \quad 0 \leq p \leq 1$

$E(X) = \frac{r(1-p)}{p}, \quad V(X) = \frac{r(1-p)}{p^2}$

$M(t) = \left(\frac{p}{1-(1-p)e^t}\right)^r$

Poisson

$P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!} \quad x = 0, 1, \ldots \quad 0 \leq \lambda < \infty$

$E(X) = \lambda, \quad V(X) = \lambda$

$M(t) = e^{\lambda(e^t - 1)}$

$\varphi(t) = e^{\lambda(e^{it} - 1)}$
