MAT-206: Método de máxima verosimilitud

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Método de máxima verosimilitud

Definición 1 (Estimador máximo verosímil):

Un estimador $\widehat{\boldsymbol{\theta}}_{\mathsf{ML}}$ es llamado estimador máximo verosímil (MLE) de $\boldsymbol{\theta}$, si

$$L(\widehat{\boldsymbol{\theta}}_{\mathsf{ML}}; \boldsymbol{x}) \geq L(\boldsymbol{\theta}; \boldsymbol{x}), \qquad \forall\, \boldsymbol{\theta} \in \Theta.$$

Es decir, $\widehat{\boldsymbol{\theta}}_{\mathsf{ML}}$ debe ser solución del siguiente problema de optimización

$$\max_{\theta \in \Theta} L(\boldsymbol{\theta}; \boldsymbol{x})$$

o equivalentemente,

$$\max_{\theta \in \Theta} \ell(\boldsymbol{\theta}; \boldsymbol{x}).$$

Además, en ocasiones escribimos

$$\widehat{\boldsymbol{\theta}}_{\mathsf{ML}} := \arg\max_{\theta \in \Theta} \ell(\boldsymbol{\theta}; \boldsymbol{x}).$$

Resultado 1 (Invarianza del MLE):

Si $\boldsymbol{\gamma} = \boldsymbol{g}(\boldsymbol{\theta})$ y $\boldsymbol{g}$ es biyectiva. Entonces $\widehat{\boldsymbol{\theta}}$ es el MLE para $\boldsymbol{\theta}$ si y solo si $\widehat{\boldsymbol{\gamma}} = \boldsymbol{g}(\widehat{\boldsymbol{\theta}})$ es el MLE para $\boldsymbol{\gamma}$.

Demostración:

Considere $L(\boldsymbol{\theta}; \boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{\theta})$ y como $\boldsymbol{g}$ es biyectiva tenemos que

$$\widetilde{L}(\boldsymbol{\gamma}; \boldsymbol{x}) = f(\boldsymbol{x}; \boldsymbol{g}^{-1}(\boldsymbol{\gamma})).$$

Además,

$$\begin{aligned} & \widetilde{L}(\widehat{\boldsymbol{\gamma}}; \boldsymbol{x}) \geq \widetilde{L}(\boldsymbol{\gamma}; \boldsymbol{x}), \quad \forall\, \boldsymbol{\gamma} \quad \Longleftrightarrow \quad f(\boldsymbol{x}; \boldsymbol{g}^{-1}(\widehat{\boldsymbol{\gamma}})) \geq f(\boldsymbol{x}; \boldsymbol{g}^{-1}(\boldsymbol{\gamma})), \quad \forall\, \boldsymbol{\gamma} \\ \Longleftrightarrow \quad & f(\boldsymbol{x}; \widehat{\boldsymbol{\theta}}) \geq f(\boldsymbol{x}; \boldsymbol{\theta}), \quad \forall\, \boldsymbol{\theta} \quad \Longleftrightarrow \quad L(\widehat{\boldsymbol{\theta}}; \boldsymbol{x}) \geq L(\boldsymbol{\theta}; \boldsymbol{x}), \quad \forall\, \boldsymbol{\theta}. \end{aligned}$$

Definición 2:

Si $\widehat{\boldsymbol{\theta}}_{\text{ML}}$ es el MLE de $\boldsymbol{\theta}$ y $\boldsymbol{\gamma} = \boldsymbol{g}(\boldsymbol{\theta})$. Entonces el MLE de $\boldsymbol{\gamma}$ es definido como:

$$\widehat{\boldsymbol{\gamma}}_{\text{ML}} = \boldsymbol{g}(\widehat{\boldsymbol{\theta}}_{\text{ML}}).$$

Observación:

Si $\ell(\boldsymbol{\theta})$ es continuamente diferenciable, el estimador máximo verosímil $\widehat{\boldsymbol{\theta}}_{\text{ML}}$ es dada como una solución de las ecuaciónes de verosimiltud:

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0},$$

donde $\boldsymbol{U}(\boldsymbol{\theta}) = \partial \ell(\boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ corresponde al vector score.

Ejemplo:

Considere $X_1, \ldots, X_n$ muestra aleatoria desde $\mathrm{N}(\mu, \sigma^2)$. De este modo

$$L(\mu, \sigma^2; \boldsymbol{x}) = (2\pi\sigma^2)^{-n/2} \exp\Big\{ -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2 \Big\}.$$

Lo que permite obtener la función de log-verosimilitud

$$\ell(\mu, \sigma^2) = -\frac{n}{2} \log 2\pi\sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2,$$

diferenciando con respecto a μ y σ^2 lleva a las ecuaciones de verosimilitud

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0,$$

$$\frac{\partial \ell(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0.$$

Método de máxima verosimilitud

resolviendo las ecuaciones anteriores para μ y σ^2, sigue que

$$\widehat{\mu}_{\mathsf{ML}} = \overline{x}, \qquad \widehat{\sigma}^2_{\mathsf{ML}} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \widehat{\mu}_{\mathsf{ML}})^2.$$

La matriz de información de Fiher de $\boldsymbol{\theta} = (\mu, \sigma^2)^\top$ y luego de evaluar en $\boldsymbol{\theta} = \widehat{\boldsymbol{\theta}}_{\mathsf{ML}}$, adopta la forma:

$$\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}) = \frac{n}{\sigma^2}\begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{2\sigma^2} \end{pmatrix}$$

Ejemplo:

Sea $X_1, \ldots, X_n$ una muestra de variables aleatorias IID desde U$[0, \theta]$. Esto es

$$f(x;\theta) = \begin{cases} \dfrac{1}{\theta}, & 0 \le x \le \theta, \\ 0, & \text{en otro caso.} \end{cases} = \frac{1}{\theta}\, I_{[0,\theta]}(x), \qquad \theta > 0.$$

De este modo,[1]

$$L(\theta; \boldsymbol{x}) = \frac{1}{\theta} \prod_{i=1}^{n} I_{[0,\theta]}(x_i) = \frac{1}{\theta^n} \prod_{i=1}^{n} I_{[x_i,\infty)}(\theta).$$

Tenemos que

$$\prod_{i=1}^{n} I_{[x_i,\infty)}(\theta) = 1 \iff I_{[x_i,\infty)}(\theta) = 1,\ \forall\, i \iff x_i \le \theta,\ \forall\, i$$

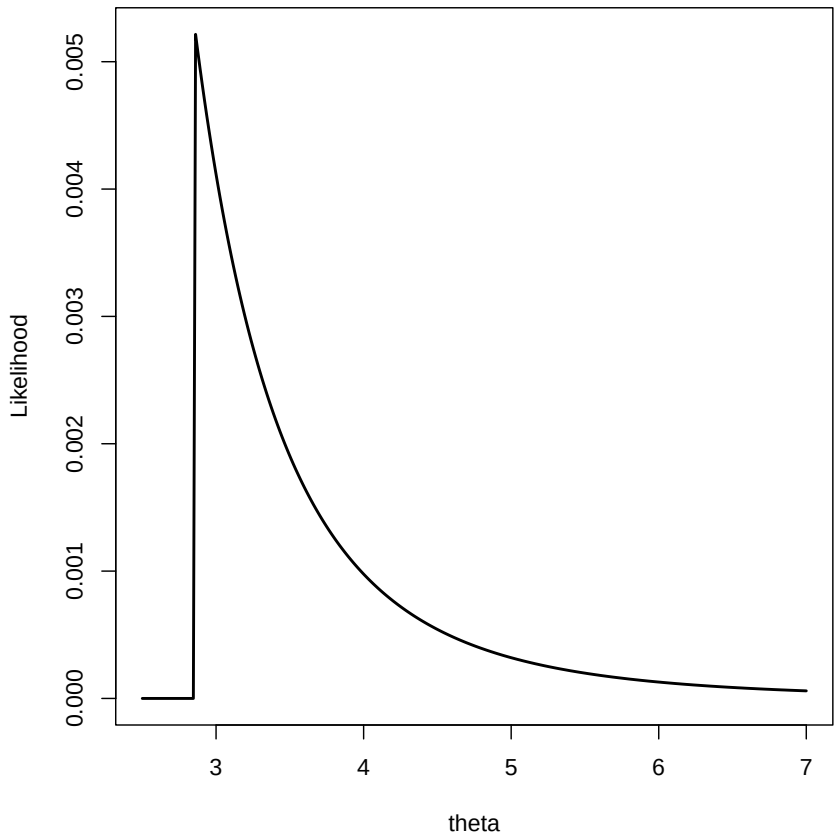
$$\iff \max_i\{x_i\} \le \theta \iff I_{[x_{(n)},\infty)}(\theta) = 1,$$

donde $x_{(n)} = \max\{x_1, \ldots, x_n\}$.

[1] Basta notar que $0 \le x_i \le \theta \Rightarrow x_i \le \theta < \infty$.

Método de máxima verosimilitud

Considere una muestra $\boldsymbol{x} = (2.85, 1.51, 0.69, 0.57, 2.29)^\top$ desde la distribución $\mathrm{U}(0, \theta)$. La función de verosimilitud es dada por:



Método de máxima verosimilitud

De este modo, la función

$$L(\theta; \boldsymbol{x}) = \frac{1}{\theta^n} \, I_{[x_{(n)}, \infty)}(\theta),$$

es monótona decreciente, de ahí que

$$L(\theta; \boldsymbol{x}) \leq \frac{1}{(x_{(n)})^n},$$

y por tanto sigue que $\widehat{\theta}_{\mathsf{ML}} = x_{(n)}$.

Método de máxima verosimilitud

Ejemplo:

Suponga $X_1, \ldots, X_n$ variables aleatorias IID con densidad

$$f(x; a, b) = \frac{b}{2} \exp\{-b|x - a|\}, \qquad x \in \mathbb{R},$$

con $a \in \mathbb{R}$ y $b > 0$. De este modo, para b conocido, tenemos

$$\ell(a; \boldsymbol{x}) = \log(b/2) - b \sum_{i=1}^{n} |x_i - a|.$$

Es decir, podemos obtener $\widehat{a}_{\mathsf{ML}}$, equivalentemente, como la solución de

$$\min_a \sum_{i=1}^{n} |x_i - a|,$$

y es bien sabido que $\widehat{a}_{\mathsf{ML}} = \mathsf{median}\{x_1, \ldots, x_n\}$

Ejemplo:

Sea $X_1, \ldots, X_n$ variables IID con distribución común en la FE 1-paramétrica y $\theta \in \Theta$. Considere $\phi = \eta(\theta)$ y sea $\gamma(\phi) = \gamma(\eta(\theta)) = b(\theta)$. Sabemos que

$$L(\phi) = \exp\left[\phi \sum_{i=1}^{n} T(x_i) - n\gamma(\phi)\right] \prod_{i=1}^{n} h(x_i).$$

De este modo, la función de log-verosimilitud adopta la forma:

$$\ell(\phi) = \phi \sum_{i=1}^{n} T(x_i) - n\gamma(\phi) + \sum_{i=1}^{n} \log h(x_i),$$

lo que lleva a

$$\frac{\mathrm{d}\ell(\phi)}{\mathrm{d}\phi} = \sum_{i=1}^{n} T(x_i) - n\gamma'(\phi).$$

Método de máxima verosimilitud

Resolviendo la condición de primer orden $\mathrm{d}\ell(\phi)/\,\mathrm{d}\phi = 0$, tenemos que $\widehat{\phi}_{\mathsf{ML}}$ es solución de la ecuación:

$$\gamma'(\phi) = \frac{1}{n}\sum_{i=1}^{n} T(x_i).$$

Finalmente por la propiedad de invarianza del MLE sigue que $\widehat{\theta}_{\mathsf{ML}} = \eta^{-1}(\widehat{\phi}_{\mathsf{ML}})$. Por otro lado, es fácil notar que

$$\frac{\mathrm{d}^2\,\ell(\phi)}{\mathrm{d}\phi^2} = -n\gamma''(\phi) = -nb''(\theta) = -\,\mathsf{var}\left(\sum_{i=1}^{n} T(x_i)\right) \leq 0.$$

De lo anterior, sigue que $\ell(\phi)$ es cóncava y por tanto su máximo en $\Phi = \eta(\Theta)$ debe ser único.

Ejemplo:

Suponga $X_1, \ldots, X_n$ muestra aleatoria con distribución Weibull, en cuyo caso,

$$f(x;\boldsymbol{\theta}) = \frac{\alpha}{\beta}\left(\frac{x}{\beta}\right)^{\alpha-1}\exp\left\{-\left(\frac{x}{\beta}\right)^{\alpha}\right\}, \qquad x > 0,$$

con $\boldsymbol{\theta} = (\alpha, \beta)^\top \in \mathbb{R}_+ \times \mathbb{R}_+$. La función de log-verosimilitud es dada por

$$\ell(\boldsymbol{\theta}) = n(\log\alpha - \log\beta) + (\alpha-1)\sum_{i=1}^{n}\log\left(\frac{x_i}{\beta}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\beta}\right)^{\alpha}.$$

Diferenciando obtenemos las ecuaciones:

$$\frac{n}{\alpha} + \sum_{i=1}^{n}\log\left(\frac{x_i}{\beta}\right) - \sum_{i=1}^{n}\left(\frac{x_i}{\beta}\right)^{\alpha}\log\left(\frac{x_i}{\beta}\right) = 0$$

$$-\frac{n\alpha}{\beta} + \frac{\alpha}{\beta}\sum_{i=1}^{n}\left(\frac{x_i}{\beta}\right)^{\alpha} = 0.$$

Ejemplo (Datos de falla de resortes):

En un experimento industrial se desea determinar la confiabilidad de cierto tipo de resortes cuando son sometidos a repetidos ciclos de esfuerzo hasta que fallen.

Los tiempos de falla, en unidades de 10^3 ciclos de esfuerzo para 60 resortes fueron divididos en grupos de 10 para 6 diferentes niveles de presión.

Note que conforme la presión decrece existe un rápido aumento de número promedio de ciclos hasta la falla.

Además existe un patrón lineal entre $\log \overline{y}$ y $\log s^2$, sugiriendo que la varianza es proporcional al promedio al cuadrado.

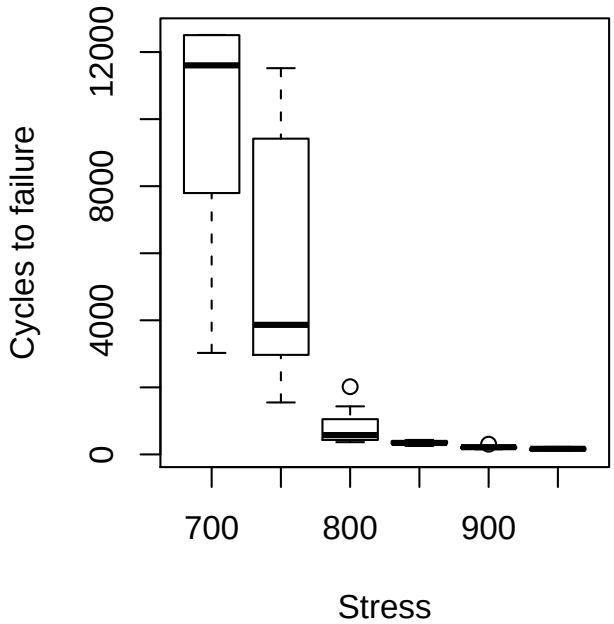
Método de máxima verosimilitud

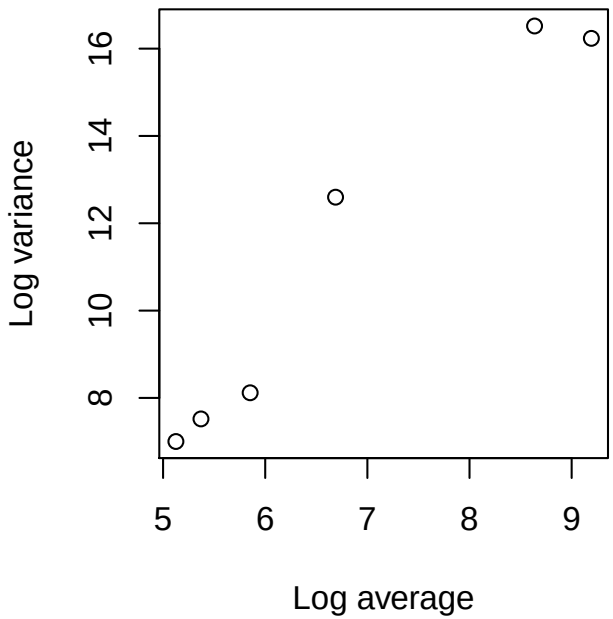
Tiempos de falla (en unidades de 10^3 ciclos) de resortes sometidos a repetidos ciclos de presión bajo un esfuerzo dado.

	Stress (N/mm^2)					
	950	900	850	800	750	700
	225	216	324	627	3402	12510
	171	162	321	1051	9417	12505
	198	153	432	1434	1802	3027
	189	216	252	2020	4326	12505
	189	225	279	525	11520	6253
	135	216	414	402	7152	8011
	162	306	396	463	2969	7795
	135	225	379	431	3012	11604
	117	243	351	365	1550	11604
	162	189	333	715	11211	12470
$\overline{y}$	168.3	215.1	348.1	803.3	5636.1	9828.4
s	33.1	42.9	57.9	544.0	3864.3	3354.7

Método de máxima verosimilitud

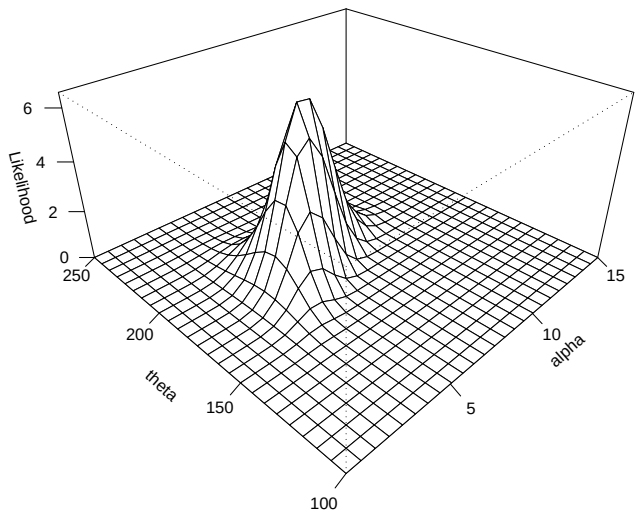
Tiempos de falla (en unidades de 10^3 ciclos) de resortes sometidos a repetidos ciclos de presión bajo un esfuerzo dado.

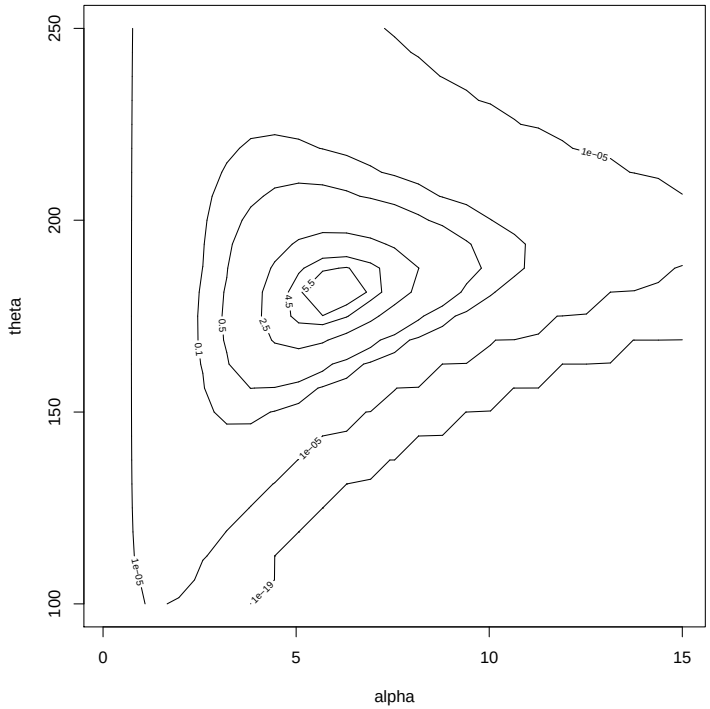




Método de máxima verosimilitud

Verosimiltud para los datos de fallas de resortes a una presión de 950 N/mm^2





Método de máxima verosimilitud

Con el objetivo de resolver el problema

$$\max_{\theta \in \Theta} \ell(\boldsymbol{\theta}),$$

suponga la expansión de Taylor en torno de $\boldsymbol{\theta}^*$, como:

$$\ell(\boldsymbol{\theta}^* + \boldsymbol{p}) = \ell(\boldsymbol{\theta}^*) + \Big(\frac{\partial \ell(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta}}\Big)^\top \boldsymbol{p} + \frac{1}{2}\,\boldsymbol{p}^\top \Big(\frac{\partial^2 \ell(\boldsymbol{\theta}^*)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top}\Big)\boldsymbol{p} + o(\|\boldsymbol{p}\|^2),$$

donde $o(u)$ es un término de error de orden menor que u, conforme $u \to 0$, es decir,

$$\lim_{u \to 0} \frac{o(u)}{u} = 0.$$

Defina la función cuadrática,

$$q_k(\boldsymbol{p}) = \ell(\boldsymbol{\theta}^{(k)}) + \boldsymbol{U}^\top(\boldsymbol{\theta}^{(k)})\boldsymbol{p} + \frac{1}{2}\,\boldsymbol{p}^\top \boldsymbol{H}(\boldsymbol{\theta}^{(k)})\boldsymbol{p},$$

donde $\boldsymbol{U}(\boldsymbol{\theta}) = \partial \ell(\boldsymbol{\theta})/\partial \boldsymbol{\theta}$ y $\boldsymbol{H}(\boldsymbol{\theta}) = \partial^2 \ell(\boldsymbol{\theta})/\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^\top$.

Método de máxima verosimilitud

Minimizando $q_k(\boldsymbol{p})$ con relación a $\boldsymbol{p}$, lleva al sistema de ecuaciones $\partial q_k(\boldsymbol{p})/\partial \boldsymbol{p} = \mathbf{0}$.
Es decir, obtenemos:

$$\boldsymbol{H}(\boldsymbol{\theta}^{(k)})\boldsymbol{p} = -\boldsymbol{U}(\boldsymbol{\theta}^{(k)}). \tag{1}$$

Métodos tipo-Newton adoptan la forma:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \lambda_k \boldsymbol{p}_k, \qquad k = 0, 1, \ldots,$$

donde $\boldsymbol{p}_k$ es la dirección de búsqueda dada por la solución del sistema dado en (1), mientras que λ_k es un largo de paso que debe ser escogido para garantizar que

$$\ell(\boldsymbol{\theta}^{(k)} + \lambda_k \boldsymbol{p}_k) \geq \ell(\boldsymbol{\theta}^{(k)}).$$

Es fácil notar que la dirección dada por (1), satisface

$$\boldsymbol{U}^\top(\boldsymbol{\theta}^{(k)})\boldsymbol{p}_k = \boldsymbol{U}^\top(\boldsymbol{\theta}^{(k)})\{-\boldsymbol{H}(\boldsymbol{\theta}^{(k)})\}^{-1}\boldsymbol{U}(\boldsymbol{\theta}^{(k)}) > 0,$$

es decir, corresponde a una dirección de ascenso.

Método de máxima verosimilitud

Sea $\boldsymbol{g}_k = \boldsymbol{U}(\boldsymbol{\theta}^{(k+1)}) - \boldsymbol{U}(\boldsymbol{\theta}^{(k)})$, esta clase de métodos actualiza una estimación de $\{\boldsymbol{H}(\boldsymbol{\theta}^{(k)})\}^{-1}$ usando un método secante, tal como el método de Davidon-Fletcher-Powell (DFP)

$$\boldsymbol{B}_{k+1}^{-1} = \boldsymbol{B}_k^{-1} + \frac{\boldsymbol{s}_k \boldsymbol{s}_k^\top}{\boldsymbol{s}_k^\top \boldsymbol{g}_k} - \frac{\boldsymbol{B}_k^{-1} \boldsymbol{g}_k \boldsymbol{g}_k^\top \boldsymbol{B}_k^{-1}}{\boldsymbol{g}_k^\top \boldsymbol{B}_k^{-1} \boldsymbol{g}_k},$$

o bien, el método de Broyden-Fletcher-Goldfarb-Shanno (BFGS)

$$\boldsymbol{B}_{k+1}^{-1} = \boldsymbol{B}_k^{-1} + \frac{\boldsymbol{s}_k \boldsymbol{g}_k^\top \boldsymbol{B}_k^{-1}}{\boldsymbol{s}_k^\top \boldsymbol{g}_k} - \frac{\boldsymbol{B}_k^{-1} \boldsymbol{g}_k \boldsymbol{s}_k^\top}{\boldsymbol{s}_k^\top \boldsymbol{g}_k} + \left\{1 + \frac{\boldsymbol{g}_k^\top \boldsymbol{B}_k^{-1} \boldsymbol{g}_k}{\boldsymbol{s}_k^\top \boldsymbol{g}_k}\right\} \frac{\boldsymbol{s}_k \boldsymbol{s}_k^\top}{\boldsymbol{s}_k^\top \boldsymbol{g}_k},$$

donde $\boldsymbol{s}_k = \boldsymbol{\theta}^{(k+1)} - \boldsymbol{\theta}^{(k)}$.

Método de máxima verosimilitud

El método Fisher-scoring, corresponde a un algoritmo en la clase quasi-Newton donde $-\boldsymbol{H}(\boldsymbol{\theta})$ es aproximada mediante la matriz de información de Fisher.

De este modo, obtenemos el siguiente esquema iterativo:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \lambda_k \, \boldsymbol{\mathcal{F}}^{-1}(\boldsymbol{\theta}^{(k)}) \boldsymbol{U}(\boldsymbol{\theta}^{(k)}), \qquad k = 0, 1, \ldots,$$

donde $\boldsymbol{\mathcal{F}}(\boldsymbol{\theta}) = \mathrm{E}\{-\boldsymbol{H}(\boldsymbol{\theta})\}$.

Método de máxima verosimilitud

Cuando tenemos $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ observaciones (vectores) independientes, tenemos que

$$\ell(\boldsymbol{\theta}) = \sum_{i=1}^{n} \log f(\boldsymbol{x}_i; \boldsymbol{\theta}).$$

De este modo, $\boldsymbol{U}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \boldsymbol{U}_i(\boldsymbol{\theta})$, con $\boldsymbol{U}_i(\boldsymbol{\theta}) = \frac{\partial}{\partial \theta} \log f(\boldsymbol{x}_i; \boldsymbol{\theta})$. Notando que

$$\frac{1}{n} \sum_{i=1}^{n} \boldsymbol{U}_i(\boldsymbol{\theta}) \boldsymbol{U}_i^{\top}(\boldsymbol{\theta}) \xrightarrow{\mathsf{P}} \mathsf{E}\{\boldsymbol{U}_n(\boldsymbol{\theta}) \boldsymbol{U}_n^{\top}(\boldsymbol{\theta})\},$$

conforme $n \to \infty$, lleva al Algoritmo BHHH (Berndt et al., 1974),[2] definido como:

$$\boldsymbol{\theta}^{(k+1)} = \boldsymbol{\theta}^{(k)} + \lambda_k \left\{ \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{U}_i(\boldsymbol{\theta}^{(k)}) \boldsymbol{U}_i^{\top}(\boldsymbol{\theta}^{(k)}) \right\}^{-1} \boldsymbol{U}_n(\boldsymbol{\theta}^{(k)}), \qquad k = 0, 1, \ldots.$$

[2] Annals of Economic and Social Measurements **3**, 653-665.

Ejemplo:

Suponga $X_1, \ldots, X_n$ variables aleatorias desde la distribución Cauchy$(\theta, 1)$, con densidad

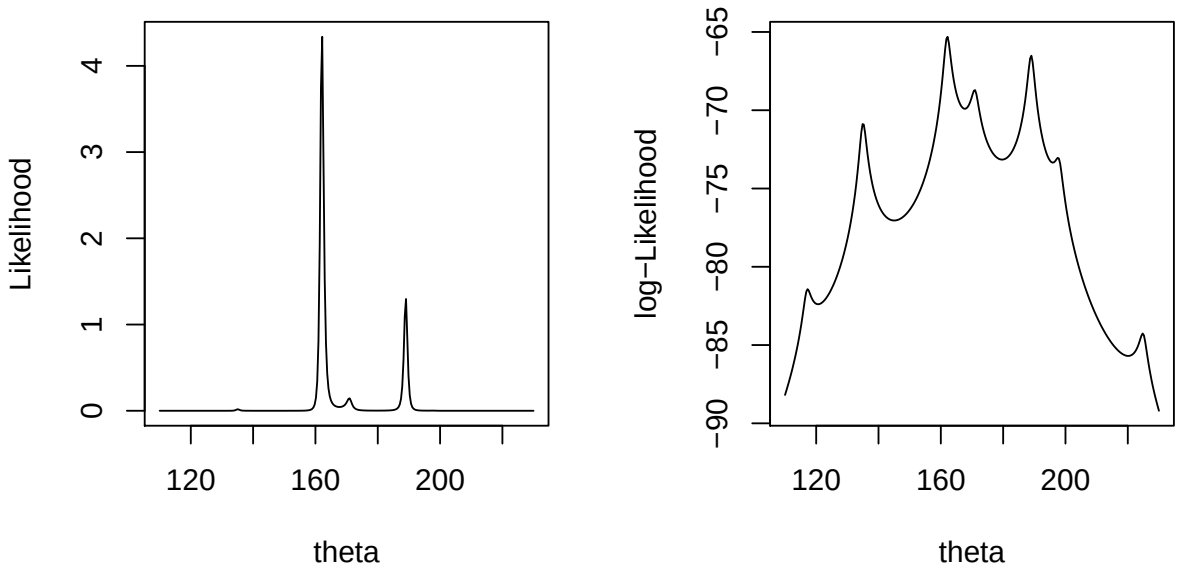
$$f(x;\theta) = \frac{1}{\pi\{1 + (x-\theta)^2\}}, \qquad x \in \mathbb{R}, \theta \in \mathbb{R}.$$

Así, la función de log-verosimilitud adopta la forma:

$$\begin{aligned}\ell(\theta; \boldsymbol{x}) &= -n \log \pi - \log \prod_{i=1}^{n} \{1 + (x_i - \theta)^2\} \\ &= -n \log \pi - \sum_{i=1}^{n} \log(1 + (x_i - \theta)^2).\end{aligned}$$

Método de máxima verosimilitud

Para los datos de fallas de resortes a una presión de 950 N/mm^2 bajo un modelo Cauchy$(\theta, 1)$, tenemos



Calculando la primera derivada, obtenemos

$$U(\theta; \boldsymbol{x}) = \frac{\partial \ell(\theta; \boldsymbol{x})}{\partial \theta} = \sum_{i=1}^{n} \frac{2(x_i - \theta)}{1 + (x_i - \theta)^2}.$$

Por tanto, el estimador máximo verosímil debe satifacer la condición de primer-orden:

$$U(\theta; \boldsymbol{x}) = \sum_{i=1}^{n} \frac{2}{1 + (x_i - \theta)^2} (x_i - \theta) = \sum_{i=1}^{n} \omega_i(\theta)(x_i - \theta) = 0, \qquad (2)$$

donde $\omega_i(\theta) = 2/(1 + (x_i - \theta)^2)$.

Método de máxima verosimilitud

Es fácil notar que la Ecuación (2) no tiene solución explícita. Para aplicar el algoritmo Newton-Raphson, calculamos

$$\frac{\partial}{\partial\theta}\, U(\theta;\boldsymbol{x}) = \sum_{i=1}^{n} \frac{\partial}{\partial\theta}\, \omega_i(\theta)(x_i-\theta) + \sum_{i=1}^{n} \omega_i(\theta)\frac{\partial}{\partial\theta}(x_i-\theta),$$

como

$$\frac{\partial}{\partial\theta}\omega_i(\theta) = \frac{4}{\{1+(x_i-\theta)^2\}^2}\,(x_i-\theta) = \omega_i^2(\theta)(x_i-\theta),$$

esto lleva a

$$\frac{\partial}{\partial\theta}\, U(\theta;\boldsymbol{x}) = \sum_{i=1}^{n} \omega_i^2(\theta)(x_i-\theta)^2 - \sum_{i=1}^{n} \omega_i(\theta).$$

De este modo,

$$-H(\theta;\boldsymbol{x}) = -\frac{\partial}{\partial\theta}\,U(\theta;\boldsymbol{x}) = \sum_{i=1}^{n}\omega_i(\theta)\{1-\omega_i(\theta)(x_i-\theta)^2\}.$$

Finalmente el método Newton-Raphson, adopta la forma:

$$\begin{aligned}\theta^{(k+1)} &= \theta^{(k)} - \frac{U(\theta^{(k)};\boldsymbol{x})}{H(\theta^{(k)};\boldsymbol{x})}\\ &= \theta^{(k)} + \frac{\sum_{i=1}^{n}\omega_i(\theta^{(k)})(x_i-\theta^{(k)})}{\sum_{i=1}^{n}\omega_i(\theta^{(k)})\{1-\omega_i(\theta^{(k)})(x_i-\theta^{(k)})^2\}}.\end{aligned} \tag{3}$$

Observación:

En la biblioteca heavy se encuentra una alternativa al esquema iterativo en (3) que utiliza un Algoritmo EM.

Ejemplo:

Suponga $X_1, \ldots, X_n$ muestra aleatoria desde Poi(λ). En este caso,

$$\ell(\lambda; \boldsymbol{x}) = \sum_{i=1}^{n} (x_i \log \lambda - \lambda - \log x_i!), \qquad \lambda > 0,$$

despreciando aquellos términos que no dependen de λ, tenemos

$$\ell(\lambda; \boldsymbol{x}) = \sum_{i=1}^{n} (x_i \log \lambda - \lambda).$$

Considere $\phi = \log \lambda$, es decir $\lambda = e^{\phi}$ y note que $\phi \in \mathbb{R}$. Así,

$$\ell(\phi; \boldsymbol{x}) = \sum_{i=1}^{n} (x_i \phi - e^{\phi}).$$

Luego, estimamos ϕ y hacemos $\widehat{\lambda}_{\mathsf{ML}} = e^{\widehat{\phi}_{\mathsf{ML}}}$.