6.Estimació pel mètode de màxima versemblança

Estadística
Grau en Matemàtiques

Josep A. Sanchez
Dept. Estadística i I.O.(UPC)

UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH

Funció de versemblança

Sigui $\underset{\sim}{X} = \{X_1, \ldots, X_n\}$ una m.a.s. de una distribució $f(x|\theta_1, \ldots, \theta_k)$

La funció de versemblança $L(\theta|\underset{\sim}{X})$ es defineix com la funció de densitat conjunta de la mostra, avaluada en els valors obtinguts:

$$L(\theta|\underset{\sim}{X}) = f(x_1, \ldots, x_n|\theta_1, \ldots, \theta_k)$$

Degut a que les variables aleatòries que formen la mostra són independents, la funció de densitat conjunta factoritza de forma natural. A més, com són identicament distribuides, la densitat de cada component és la mateixa:

$$L(\theta|\underset{\sim}{X}) = \prod_{i=1}^{n} f(x_i|\theta_1, \ldots, \theta_k)$$

Funció de versemblança

En el cas discret, la funció de versemblança correspon a la funció de massa de probabilitat conjunta:

$$L(\theta|\underset{\sim}{X}) = P(X_1 = x_1, \ldots, X_n = x_n|\theta_1, \ldots, \theta_k)$$

Aquesta funció també factoritza pel fet de que la mostra són variables aleatòries independents amb la mateixa distribució:

$$L(\theta|\underset{\sim}{X}) = \prod_{i=1}^{n} P(X = x_i, |\theta_1, \ldots, \theta_k)$$

Exemple: Distribució de Bernouilli

Sigui $X_1, \ldots, X_n$ una m.a.s. amb $X_i \sim Bern(p)$

$$P(X = x|p) = p^x(1-p)^{1-x}$$

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

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

- Suport de la variable: $x \in \{0, 1\}$
- Espai de paràmetres: $p \in [0, 1]$

Exemple: Distribució exponencial (paràmetre d'escala)

Sigui $X_1, \ldots, X_n$ una m.a.s. amb $X_i \sim exp(\mu)$

$$f(x|\mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \mathbb{I}_{(0,+\infty)}(x)$$

$$f(x_1, \ldots, x_n|\mu) = \frac{1}{\mu^n} e^{-\frac{\sum_{i=1}^n x_i}{\mu}} \prod_{i=1}^n \mathbb{I}_{(0,+\infty)}(x_i)$$

$$L(\mu|x_1, \ldots, x_n) = \frac{1}{\mu^n} e^{-\frac{\sum_{i=1}^n x_i}{\mu}} \prod_{i=1}^n \mathbb{I}_{(0,+\infty)}(x_i)$$

- Suport de la variable: $x \in \mathbb{R}^+$
- Espai de paràmetres: $\mu \in \mathbb{R}^+$

Exemple: Distribució exponencial (paràmetre de taxa)

Sigui $X_1, \ldots, X_n$ una m.a.s. amb $X_i \sim exp(\lambda)$

$$f(x|\lambda) = \lambda e^{-\lambda x} \mathbb{I}_{(0,+\infty)}(x)$$

$$f(x_1, \ldots, x_n|\lambda) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \mathbb{I}_{(0,+\infty)}(x)$$

$$L(\lambda|x_1, \ldots, x_n) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i} \mathbb{I}_{(0,+\infty)}(x)$$

- Suport de la variable: $x \in \mathbb{R}^+$
- Espai de paràmetres: $\lambda \in \mathbb{R}^+$

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

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

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

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

- Suport de la variable: $x \in \mathbb{R}$
- Espai de paràmetres: $(\mu, \sigma^2) \in \mathbb{R} * \mathbb{R}^+$

L'**Estimador de Màxima Versemblança** (*Maximum likelihood estimator (MLE)*) és el valor que maximitza la funció de versemblança:

$$\hat{\theta}_{ML} = \underset{\theta \in \Theta}{\operatorname{argmax}}\, L(\theta | \underset{\sim}{X})$$

- L'estimador correspon al valor del paràmetre més creible (versemblant) d'acord als valors obtinguts de la mostra
- El càlcul de l'estimador s'obté com a resultat d'un problema d'optimització
- Els estimadors de màxima versemblança tenen molt bones propietats

Alguns problemes en l'estimació màxim-versemblant

- Com trobar el màxim global? Si trobem un màxim, podem garantir que és global?
- Sensibilitat numèrica: com de sensible es la estimación obtenida si los datos se perturban ligeramente.

Candidatos para estimadores ML

- Punts interiors on:

$$\frac{\partial}{\partial\theta} L(\theta|\underset{\sim}{X}), \qquad \text{i} \qquad \frac{\partial^2}{\partial\theta^2} L(\theta|\underset{\sim}{X})|_{\theta=\hat{\theta}} < 0$$

- Punts frontera

Exemples (unidimensionals)

- Bernouilli: $X_i \sim Bern(p)$

$$\hat{p}_{ML} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- Poisson: $X_i \sim Pois(\lambda)$

$$\hat{\lambda}_{ML} = \frac{\sum_{i=1}^{n} X_i}{n}$$

- Normal (amb σ^2 coneguda): $X_i \sim N(\mu, 1)$

$$\hat{\mu}_{ML} = \frac{\sum_{i=1}^{n} X_i}{n}$$

Principi d'invariància del MLE

Sigui $\hat{\theta}$ l'estimador de màxima versemblança (MLE) de θ. Llavors per qualsevol funció $\tau(\theta)$, el seu estimador MLE serà $\tau(\hat{\theta})$

Exemple:

- Sigui $X_1, \ldots, X_n$ una m.a.s. d'una variable $X \sim Bern(p)$
- Volem estimar el paràmetre $\theta = \log \frac{p}{1-p}$ (*log-odds*)
- Sabem que $\hat{p}_{ML} = \bar{X}$
- Llavors, $\hat{\theta}_{ML} = \log \frac{\bar{X}}{1-\bar{X}}$

Principi d'invariància del MLE

Demostració (1/2):

- Si $\eta = \tau(\theta)$ és bijectiva, reparametritzem la versemblança:

$$L^*(\eta|\underset{\sim}{X}) = \prod_{i=1}^{n} f(x_i|\tau^{-1}(\eta)) = L(\tau^{-1}(\eta)|\underset{\sim}{X})$$

  Llavors,

$$L^*(\hat{\eta}|\underset{\sim}{X}) = \sup_{\eta} L^*(\eta|\underset{\sim}{X}) = \sup_{\eta} L(\tau^{-1}(\eta)|\underset{\sim}{X}) =$$

$$= \sup_{\theta} L(\theta|\underset{\sim}{X}) = L(\hat{\theta}|\underset{\sim}{X}) = L^*(\tau(\hat{\theta})|\underset{\sim}{X})$$

Principi d'invariància del MLE

Demostració (2/2):

- Si $\eta = \tau(\theta)$ no és bijectiva, definim la versemblança inducida per τ:

$$L^*(\eta|\underset{\sim}{X}) = \sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta|\underset{\sim}{X})$$

Llavors,

$$L^*(\hat{\eta}|\underset{\sim}{X}) = \sup_{\eta} L^*(\eta|\underset{\sim}{X}) = \sup_{\eta} \sup_{\{\theta:\tau(\theta)=\eta\}} L(\theta|\underset{\sim}{X}) =$$

$$= \sup_{\theta} L(\theta|\underset{\sim}{X}) = L(\hat{\theta}|\underset{\sim}{X}) = \sup_{\{\theta:\tau(\theta)=\tau(\hat{\theta})\}} L(\theta|\underset{\sim}{X}) = L^*(\tau(\hat{\theta})|\underset{\sim}{X})$$

Exemples (multidimensionals)

- Normal: $X_i \sim N(\mu, \sigma^2)$

$$\hat{\mu}_{ML} = \frac{\sum_{i=1}^{n} X_i}{n} = \bar{X}$$

$$\hat{\sigma^2}_{ML} = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$$

- Gamma: $X_i \sim \Gamma(\alpha, \beta)$