

MLE converge para θ^*

①

Suponha que $Y_1, \dots, Y_n$ sejam variáveis aleatórias i.i.d. com densidade de probabilidade $f(y; \theta)$ indexada por um parâmetro θ .

A verossimilhança é

$$L(\theta) = f(y; \theta) = \prod_{i=1}^n f(y_i; \theta)$$

e a log-verossimilhança é

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^n \log f(y_i; \theta)$$

Suponha que exista um valor θ^* do parâmetro tal que os dados sejam gerados pela densidade $f(y; \theta^*)$

Vamos mostrar que o MLE $\hat{\theta}_n$ baseado numa amostra de tamanho n converge para θ^* à medida que n cresce. Isto é:

$$\hat{\theta}_n \longrightarrow \theta^* \text{ quando } n \text{ cresce}$$

Vamos modificar ligeiramente a log-verossimilhança ⁽²⁾
dividindo pelo número de elementos da amostra:

$$l_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(y_i; \theta)$$

Queremos transformar esta expressão numa
variável aleatória e para isto trocamos y_i
por Y_i :

$$l_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(Y_i; \theta)$$

As n variáveis aleatórias

$\log f(Y_1; \theta)$, $\log f(Y_2; \theta)$, ..., $\log f(Y_n; \theta)$

são i.i.d. e assim $l_n(\theta)$ é a média
aritmética de n v.a.s i.i.d.

Pela Lei dos Grandes Números, $l_n(\theta)$
converge para o valor esperado da v.a.

$\log f(Y; \theta)$.

Vamos obter este valor esperado.

A variável Y possui densidade $f(y; \theta^*)$ ③
com parâmetro θ^* .

A função de log-verossimilhança é avaliada
em todo valor θ , mas apenas em θ^* .

Assim, ^{em geral,} ~~o~~ valor θ em $\log f(Y; \theta)$ não
será igual a θ^* .

Portanto o valor esperado de $\log f(Y; \theta)$
é igual a

$$L(\theta) = E_{\theta^*} [\log f(Y; \theta)] = \int (\log f(y; \theta)) f(y; \theta^*) dy$$

Note que temos θ e θ^* na expressão acima.
Portanto, pela Lei dos Grandes Números,
temos que

$$(*) \quad l_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(Y_i; \theta) \longrightarrow \int \log f(y; \theta) f(y; \theta^*) dy \\ = L(\theta)$$

Note que, se avaliarmos $L(\theta)$ no ponto θ^*

temos

(4)

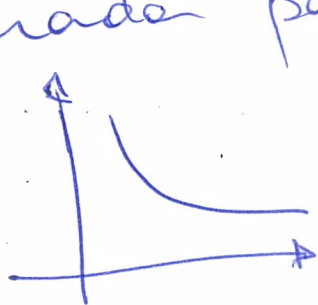
$$\begin{aligned} L_0(\theta^*) &= \int (\log f(y; \theta^*)) \cdot f(y; \theta^*) dy \\ &= E_{\theta^*} [\log f(Y; \theta^*)] \end{aligned}$$

Vamos mostrar que $L_0(\theta)$ é maximizado quando tomamos θ^* . Isto é,

$$L_0(\theta) \leq L_0(\theta^*) \quad \forall \theta$$

Para provar isto vamos usar a desigualdade de Jensen.

Uma função convexa tem a concavidade virada para cima. Por exemplo,

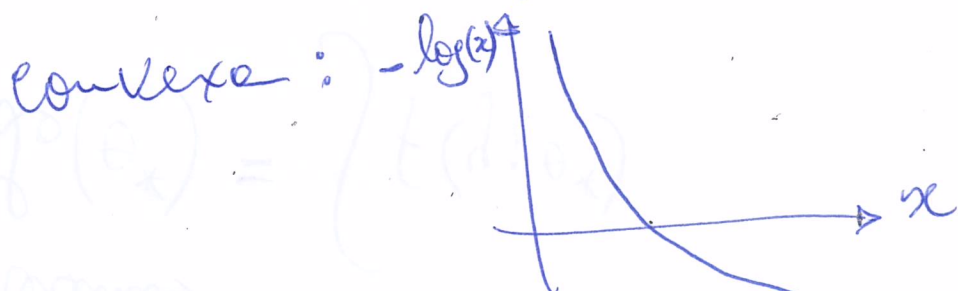


ou



são funções

convexas. A função $h(x) = -\log(x)$ é



A função $h(x) = x^2$ também é convexa

Desigualdade de Jensen Se h é ^{função} convexa ⑤

$$\text{então } E[h(Y)] \geq h[E(Y)].$$

Por exemplo, como $h(y) = y^2$ é convexa, temos

$$E[Y^2] \geq (E[Y])^2$$

e também, como $h(y) = -\log(y)$ é convexa,

$$E[-\log(Y)] \geq -\log(E[Y])$$

Vamos usar a desigualdade de Jensen para mostrar que $L_0(\theta^*) \geq L_0(\theta) \quad \forall \theta$

ou ainda, que $L_0(\theta^*) - L_0(\theta) \geq 0 \quad \forall \theta$.

De fato, vamos começar:

$$L_0(\theta^*) - L_0(\theta) = \int (\log f(y; \theta^*)) f(y; \theta^*) dy - \int (\log f(y; \theta)) \cdot f(y; \theta^*) dy$$

$$= \int [\log f(y; \theta^*) - \log f(y; \theta)] \cdot f(y; \theta^*) dy$$

$$= \int \log \left(\frac{f(y; \theta^*)}{f(y; \theta)} \right) \cdot f(y; \theta^*) dy$$

$$= E_{\theta^*} \left[\log \left(\frac{f(Y; \theta^*)}{f(Y; \theta)} \right) \right]$$

⑥

$$= E_{\theta^*} \left[-\log \left(\frac{f(Y; \theta)}{f(Y; \theta^*)} \right) \right]$$

Usando a desigualdade de Jensen com a função $h(x) = -\log(x)$, podemos escrever

$$E_{\theta^*} \left[-\log \left(\frac{f(Y; \theta)}{f(Y; \theta^*)} \right) \right] \geq -\log \left[E_{\theta^*} \left(\frac{f(Y; \theta)}{f(Y; \theta^*)} \right) \right]$$

$$= -\log \left[\int \frac{f(y; \theta)}{\cancel{f(y; \theta^*)}} \cancel{f(y; \theta^*)} dy \right]$$

$$= -\log \left[\int f(y; \theta) dy \right]$$

$$= -\log[1] \quad \text{pois } f(y; \theta) \text{ é uma densidade de probabilidade}$$

$$= 0$$

Em resumo, mostramos que $L(\theta^*) \geq L(\theta) \forall \theta$

(7)

Voltando a (*) e usando que $L_0(\theta^*) \geq L_0(\theta)$,

temos que

$$l_n(\theta) = \frac{1}{n} \sum_{i=1}^n \log f(Y_i; \theta) \longrightarrow L_0(\theta) \leq L_0(\theta^*)$$

A curva aleatória $l_n(\theta)$ converge para a curva fixa $L_0(\theta)$. Isto significa que o máximo e o argumento que maximiza $l_n(\theta)$ também converge para o máximo e o arg max de $L_0(\theta)$:

$$\hat{\theta}_n = \arg \max_{\theta} \frac{1}{n} \sum_{i=1}^n \log f(Y_i; \theta) \longrightarrow \arg \max_{\theta} L_0(\theta) = \theta^*$$

Isto é, provamos que $\hat{\theta}_n \longrightarrow \theta^*$ se n cresce.