1.5.2 Un semplice esempio: stima di gaussiane univariate

Sia $N(\mu, \sigma^2)$ una distribuzione gaussiana univariata con valore atteso μ e varianza σ^2 sconosciuti. Si ricordi che la sua densita di probabilità è data da

$$p(x) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).$$

Teorema 1.5.2. *Per un campione $x^{(1)}, \dots, x^{(m)}$ di m variabili aleatorie con distribuzione $N(\mu, \sigma^2)$, gli stimatori a massima verosimiglianza sono:*

$$\hat{\mu} \stackrel{\text{def}}{=} \frac{1}{m} \sum_{i=1}^{m} x^{(i)}$$

$$\hat{\sigma}^2 \stackrel{\text{def}}{=} \frac{1}{m} \sum_{i=1}^{m} (x^{(i)} - \hat{\mu})^2$$

Dimostrazione. La funzione di verosimiglianza è

$$\begin{aligned}\mathcal{L}(\mu, \sigma | x) &= \prod_{i=1}^{m} p(x^{(i)}) \\ &= \frac{1}{(2\pi\sigma^2)^{m/2}} \exp\left(-\sum_{i=1}^{m} \frac{(x^{(i)}-\mu)^2}{2\sigma^2}\right).\end{aligned}$$

Il massimo di $\mathcal{L}$ si ottiene in maniera equivalente minimizzando il logaritmo negativo di $\mathcal{L}$, che vale

$$-\log \mathcal{L}(\mu, \sigma | x) = \frac{m}{2} \log 2\pi + \frac{m}{2} \log \sigma^2 + \frac{1}{2} \sum_i \frac{(x^{(i)}-\mu)^2}{\sigma^2}$$

Per ogni valore di σ, la funzione $g(\mu) \stackrel{\text{def}}{=} -\log \mathcal{L}(\mu, \sigma | x)$ è convessa in μ e quindi il suo minimo (per σ fissato) si ottiene quando la derivata parziale rispetto a μ si annulla:

$$0 = \frac{\partial}{\partial \mu}(-\log \mathcal{L}(\mu, \sigma | x)) = \frac{1}{\sigma^2} \sum_{i=1}^{m} (x^{(i)} - \mu),$$

ovvero quando $\mu = (1/m) \sum_i x^{(i)} = \hat{\mu}$. Questo dimostra la prima parte del teorema. Fissando ora quindi $\mu = \hat{\mu}$ nella funzione $\mathcal{L}(\mu, \sigma | x)$, il massimo al variare di σ si ottiene quando è massimizzata l'espressione

$$(\sigma^2)^{-m/2} \exp\left(-\frac{\sum_i (x^{(i)} - \hat{\mu})^2}{2\sigma^2}\right).$$

Passando di nuovo al logaritmo negativo e introducendo $t \stackrel{\text{def}}{=} \sigma^2$, $v \stackrel{\text{def}}{=} \sum_i (x^{(i)} - \hat{\mu})^2$, il problema diventa quello di minimizzare l'espressione

$$G(t) \stackrel{\text{def}}{=} \frac{m}{2} \log t + \frac{v}{2t}$$

al variare di $t \in [0, \infty)$. Tale espressione non è convessa in t, ma il suo limite sia per $t \to 0^+$ che per $t \to \infty$ è $+\infty$, dunque il minimo si ottiene necessariamente quando $G'(t) = 0$, ovvero quando

$$\frac{m}{2t} - \frac{v}{2t^2} = 0,$$

il che equivale a prendere $t = v/m$, ovvero $\sigma^2 = (1/m) \sum_i (x^{(i)} - \hat{\mu})^2 = \hat{\sigma}^2$. □