1.5.3 Un altro esempio: stima di distribuzioni di Bernoulli

Una variabile *binaria* $Y \in \{0,1\}$ può essere modellata con un unico parametro $\theta \in [0,1]$, assumendo una distribuzione (detta di Bernoulli)

$$\Pr(Y=1)=\theta, \quad \Pr(Y=0)=1-\theta,$$

che può anche essere scritta

$$p(y)=\theta^{y}(1-\theta)^{1-y}, \qquad y \in \{0,1\}.$$

Teorema 1.5.3. *Per un campione $y^{(1)},\dots,y^{(m)}$ di m dati generati da una distribuzione di Bernoulli, lo stimatore a massima verosimiglianza è:*

$$\hat{\theta} \stackrel{\text{def}}{=} \frac{1}{m}\sum_{i=1}^{m} y^{(i)}.$$

Dimostrazione. La funzione di verosimiglianza associata al campione è

$$\begin{aligned}\mathcal{L}(\theta|y) &= \prod_{i=1}^{m} p(y^{(i)})\\ &= \prod_{i=1}^{m} \theta^{y^{(i)}}(1-\theta)^{1-y^{(i)}}\end{aligned}$$

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

$$g(\theta) \stackrel{\text{def}}{=} -\log\mathcal{L}(\theta|y) = \sum_{i=1}^{m}\left(-y^{(i)}\log\theta-(1-y^{(i)})\log(1-\theta)\right).$$

La funzione $g(\theta)$ è convessa in θ (essendo somma di funzioni convesse) e quindi il suo minimo si ottiene quando la sua derivata prima si annulla:

$$0=g'(\theta)=\sum_{i=1}^{m}\left(-\frac{y^{(i)}}{\theta}+\frac{1-y^{(i)}}{1-\theta}\right).$$

Questo accade quando

$$\sum_{i=1}^{m}\frac{y^{(i)}}{\theta}=\sum_{i=1}^{m}\frac{1-y^{(i)}}{1-\theta},$$

che è equivalente a

$$\frac{1-\theta}{\theta}=\frac{m-\sum_i y^{(i)}}{\sum_i y^{(i)}}=\frac{m}{\sum_i y^{(i)}}-1.$$

Di conseguenza, il minimo si ottiene quando

$$\frac{1}{\theta}=\frac{m}{\sum_i y^{(i)}}.$$

□