Method of Moments

$$\mathbb{E}(X) = \int x f(x;\theta)dx = g_1(\theta) \quad \Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i = g_1(\theta)$$

$$\mathbb{E}(X^2) = \int x^2 f(x;\theta)dx = g_2(\theta) \quad \Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = g_2(\theta)$$

$$\vdots \qquad\qquad\qquad\qquad \vdots$$

solve for θ from RHS system of equations

Two Basic Examples

Bernoulli(p)

$$\mathbb{E}(X) = p \quad \Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i = p \quad \Rightarrow \quad \widehat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

Normal(μ, σ^2)

$$\mathbb{E}(X) = \mu \quad \Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i = \mu \quad \Rightarrow \quad \widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$$

$$\mathbb{E}(X^2) \overset{\text{why}}{\overset{\downarrow}{=}} \mu^2 + \sigma^2 \quad \Rightarrow \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 = \mu^2 + \sigma^2$$

$$\Rightarrow \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2$$

Maximum Likelihood Method

Likelihood Function

$$L(\theta) = f(x_1, ..., x_n; \theta) \overset{\text{if i.i.d.}}{\overset{\downarrow}{=}} \prod_{i=1}^{n} f(x_i; \theta)$$

Maximum Likelihood Estimator (MLE)

$$\widehat{\theta} = \underset{\theta}{\arg\max} \quad L(\theta)$$

Log-Likelihood Function

$$\ell(\theta) = \underset{\substack{\uparrow \\ \text{base } e}}{\log} \; L(\theta) \overset{\text{if i.i.d.}}{\overset{\downarrow}{=}} \sum_{i=1}^{n} \log f(x_i; \theta)$$

Two Basic Examples

Bernoulli(p) [Exercise 5.1 (p. 71)]

$$L(p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$$

$$\Rightarrow \quad \ell(p) = \sum_{i=1}^{n} x_i \log p + (1-x_i)\log(1-p) \quad \Rightarrow \quad ...$$

Normal(μ, σ^2) [Exercise 5.2 (p. 71)]

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right]$$

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

Multivariate Normal

Univariate $X_1, ..., X_n \overset{iid}{\sim} \mathrm{N}(\mu, \sigma^2)$

$$\widehat{\mu} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}, \quad \widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \widehat{\mu})^2$$

Multivariate $\boldsymbol{x}_1, ..., \boldsymbol{x}_n \overset{iid}{\sim} \mathrm{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $\boldsymbol{x}_i \in \mathbb{R}^d$

$$\widehat{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i = \bar{\boldsymbol{x}}, \quad \widehat{\boldsymbol{\Sigma}} = \frac{1}{n}\sum_{i=1}^{n} (\underbrace{\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}}_{d\times 1})(\underbrace{\boldsymbol{x}_i - \widehat{\boldsymbol{\mu}}}_{1\times d})^\top$$

Some "Trickier" Examples

Uniform$(0, \theta)$ [**Exercise 5.8a,b (p. 78)**]

$$\texttt{MOM}: \quad \mathbb{E}(X) = \theta/2 \quad \Rightarrow \quad \widehat{\theta} = 2\bar{X}$$

$$\texttt{MLE}: \quad L(\theta) = \prod_{i=1}^{n} (1/\theta) = (1/\theta)^n \Rightarrow ... \Rightarrow \widehat{\theta} = \max\{X_1, ..., X_n\}.$$

Multinomial$(n; p_1, ..., p_K)$

$$\texttt{MOM}: \quad \mathbb{E}(X_k) = np_k \quad \Rightarrow \quad \widehat{p}_k = X_k/n, \quad k = 1, 2, ..., K$$

$$\texttt{MLE}: \quad L(...) = \frac{n!}{x_1!...x_K!} p_1^{x_1} ... p_K^{x_K} \Rightarrow \ell(...) = C + \sum_{k=1}^{K} x_k \log p_k$$

$$\Rightarrow ...$$

Remarks Constrained optimization (box, equality, ...).

Example 5.2 (p. 72–73)

Suppose $X_1, X_2, ..., X_n$ are independent, each distributed as

$$X_i \sim \text{Poisson}(\theta v_i),$$

e.g., $X_i = \#$ of disasters and $v_i = \#$ of activities from team i.

MLE

$$L(\theta) = \prod_{i=1}^{n} \frac{e^{-\theta v_i}(\theta v_i)^{x_i}}{x_i!} \Rightarrow \ell(\theta) = \sum_{i=1}^{n} -\theta v_i + x_i \log(\theta v_i) + C$$

$$\ell'(\theta) = 0 \quad \Rightarrow \quad \widehat{\theta} = \frac{\sum_{i=1}^{n} X_i}{\sum_{i=1}^{n} v_i}$$

Remark Quantities "like" v_i called covariates (i.e., predictors).

A "Twist"

As before, $X_1, X_2, ..., X_n$ are independent, each distributed as

$$X_i \sim \text{Poisson}(\theta v_i).$$

But they are not observable, e.g., teams don't want to be overly embarrassed. Instead, observe only $Y_i = \mathbf{1}(X_i > 0)$.

A "Twist"

As before, $X_1, X_2, ..., X_n$ are independent, each distributed as

$$X_i \sim \text{Poisson}(\theta v_i).$$

But they are not observable, e.g., teams don't want to be overly embarrassed. Instead, observe only $Y_i = \mathbf{1}(X_i > 0)$.

Log-Likelihood $Y_i \sim \text{Bernoulli}(p_i)$, where $p_i = \mathbb{P}(X_i > 0) = 1 - e^{-\theta v_i}$, so

$$L(\theta) = \prod_{i=1}^{n} \left(1 - e^{-\theta v_i}\right)^{y_i} \left(e^{-\theta v_i}\right)^{1-y_i}$$

$$\Rightarrow \ell(\theta) = \sum_{i=1}^{n} y_i \log\left(1 - e^{-\theta v_i}\right) - (1 - y_i)(\theta v_i)$$

Remark Numeric optimization. [Example 5.5 (pp. 79–80)]

Newton's Method

Solving for $f(x) = 0$

$$x_{new} = x_{old} - \frac{f(x_{old})}{f'(x_{old})}$$

Solving for $\ell'(\theta) = 0$

$$\theta_{new} = \theta_{old} - \frac{\ell'(\theta_{old})}{\ell''(\theta_{old})}$$

Remark Of course, curvature $[\ell''(\theta) > 0 \text{ or } < 0]$ matters whether solution is local maximum or minimum.

Example 5.5 + Exercise 5.9 (pp. 79–80)

$$\ell'(\theta) = \sum_{i=1}^{n} y_i \left[\frac{v_i e^{-\theta v_i}}{1 - e^{-\theta v_i}} \right] - (1 - y_i) v_i$$

$$\ell''(\theta) = \sum_{i=1}^{n} y_i v_i \left[\frac{(-v_i) e^{-\theta v_i} (1 - e^{-\theta v_i}) - (e^{-\theta v_i})(v_i e^{-\theta v_i})}{(1 - e^{-\theta v_i})^2} \right]$$

$$= \sum_{i=1}^{n} y_i v_i \left[\frac{(-v_i) e^{-\theta v_i}}{(1 - e^{-\theta v_i})^2} \right]$$