Maximum Likelihood Estimation.

Likelihood.

Def'n. Given a random sample $Y_1, Y_2, \dots, Y_n$ from a discrete dist'n D w/ an unknown parameter θ, the likelihood f'tion is defined as

$$L(\theta; y_1, y_2, \dots, y_n) = p^{\theta}_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = p^{\theta}(y_1) p^{\theta}(y_2) \cdots p^{\theta}(y_n)$$

where p^{θ} is a pmf of D.

If $Y_1, \dots, Y_n$ come from a continuous dist'n D w/ pdf f^{θ}, we have this definition:

$$L(\theta; y_1, \dots, y_n) = f^{\theta}_{Y_1, \dots, Y_n}(y_1, \dots, y_n) = f^{\theta}(y_1) \cdot f^{\theta}(y_2) \cdots f^{\theta}(y_n)$$

Example.

Bernoulli. $Y_1, \dots, Y_n \sim B(p)$ $\quad\boxed{p \longleftrightarrow \theta}$

$$p(y) = \begin{cases} p & y = 1 \\ 1-p & y = 0 \end{cases}$$

$$= p^{y}(1-p)^{1-y} \quad \text{for } y = 0, 1$$

$$L(p; y_1, y_2, \dots, y_n) = p^{y_1}(1-p)^{1-y_1} \cdot p^{y_2}(1-p)^{1-y_2} \cdot \dots \cdot p^{y_n}(1-p)^{1-y_n}$$

$$= p^{\sum y_i}(1-p)^{n-\sum y_i}$$

11/15/2024.

For computational reasons, take the $\ln(\cdot)$; get the log-likelihood, i.e.,

$$\ell(p; y_1, \dots, y_n) = (\textstyle\sum y_i) \cdot \ln(p) + (n - \sum y_i) \cdot \ln(1-p)$$

Next, we differentiate with respect to p

$$\ell'(p; y_1, \dots, y_n) = (\textstyle\sum y_i) \cdot \frac{1}{p} + (n - \sum y_i) \cdot (-1) \cdot \frac{1}{1-p}$$

We equate the derivative to zero:

$$\left(\sum y_i\right)\cdot\frac{1}{p} + \left(n-\sum y_i\right)\cdot\frac{1}{p-1} = 0$$

$$\frac{1}{p(p-1)}\Big(\underbrace{\left(\sum y_i\right)(p-1) + \left(n-\sum y_i\right)\cdot p}_{=0}\Big) = 0$$

$$\cancel{\left(\sum y_i\right)}\cdot p - \sum y_i + n\cdot p - \cancel{\left(\sum y_i\right)}\cdot p = 0$$

$$\boxed{\hat{p}_{MLE} = \frac{\sum y_i}{n} = \bar{y}}$$

Example. $\boxed{\text{Normal}}$

$Y_1, Y_2, \dots, Y_n \sim N(\mu, \sigma)$

The density: $f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$ for all $y \in \mathbb{R}$

$$L(\mu, \sigma; y_1, \dots, y_n) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \prod_{i=1}^{n} e^{-\frac{(y_i-\mu)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2}$$

Example. $Y_1, Y_2, \dots, Y_n \sim U(0, \theta)$ $\boxed{\theta > 0 \text{ unknown}}$

The density: $f(y) = \frac{1}{\theta} 1_{(0,\theta)}(y) = \frac{1}{\theta} 1_{\{0 \le y \le \theta\}}$

$$L(\theta; y_1, \dots, y_n) = \frac{1}{\theta} 1_{\{0 \le y_1 \le \theta\}} \cdot \frac{1}{\theta} 1_{\{0 \le y_2 \le \theta\}} \cdots \frac{1}{\theta} 1_{\{0 \le y_n \le \theta\}}$$

$$= \left(\frac{1}{\theta}\right)^n 1_{\{0 \le y_1, \dots, y_n \le \theta\}}$$

$$L(\theta; y_1, \dots, y_n) = \left(\frac{1}{\theta}\right)^n 1_{\{0 \le \min(y_1, \dots, y_n)\}} 1_{\{\max(y_1, \dots, y_n) \le \theta\}}$$

Def'n. An estimator $\hat{\Theta} = \hat{\Theta}(y_1, \dots, y_n)$ is called the maximum likelihood estimator (MLE) if it has this property: for any $\hat{\Theta}' = \hat{\Theta}'(y_1, \dots, y_n)$ we have

$$\boxed{L(\hat{\Theta}; y_1, \dots, y_n) \geq L(\hat{\Theta}'; y_1, \dots, y_n) \quad \text{for all } y_1, \dots, y_n}$$

Example [cont'd]. $\boxed{\text{Normal w/ known } \sigma}$

$$L(\mu; y_1, \dots, y_n) = \boxed{\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n} \checkmark \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right)$$

$$\propto \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right)$$

$\Rightarrow$

$$\ell(\mu; y_1, \dots, y_n) = \ln\left(L(\mu; y_1, \dots, y_n)\right)$$

$$= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2$$

$\Rightarrow$

$$\ell'(\mu; y_1, \dots, y_n) = +\frac{1}{\cancel{2}\sigma^2}\underbrace{\sum_{i=1}^{n}\cancel{2}(y_i-\mu)(+1)}_{=0} = 0$$

$$\sum_{i=1}^{n}(y_i - \mu) = 0$$

$$\sum_{i=1}^{n} y_i - n \cdot \mu = 0$$

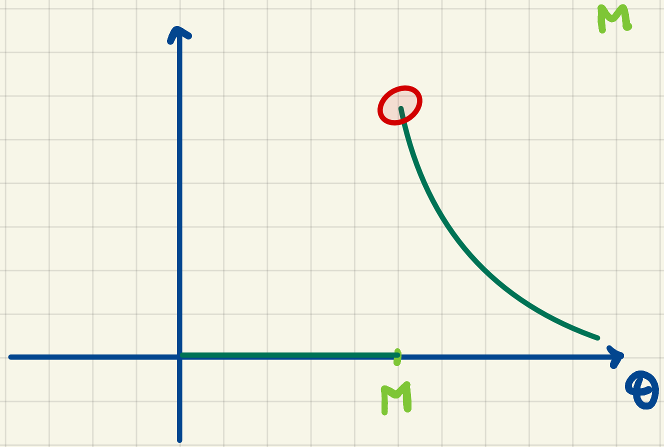
$$\boxed{\hat{\mu}_{MLE} = \bar{y}}$$

Example [cont'd]. $U(0,\theta)$

$$L(\theta; y_1, \dots, y_n) = \left(\frac{1}{\theta}\right)^n \mathbb{1}_{\{0 \le \min(y_1, \dots, y_n)\}} \cdot \mathbb{1}_{\{\max(y_1, \dots, y_n) \le \theta\}}$$

assume $y_1, \dots, y_n \ge 0$

$$= \theta^{-n} \mathbb{1}_{\{\underbrace{\max(y_1, \dots, y_n)}_{M} \le \theta\}}$$



$$\boxed{\hat{\theta}_{MLE} = \max(y_1, \dots, y_n)}$$

Q: Is this the same as the moment matching estimator?

$\longrightarrow$: Matching the theoretical to sample mean

$$\frac{\theta}{2} = \bar{Y} \implies \boxed{\hat{\theta}_{MM} = 2\bar{Y}}$$