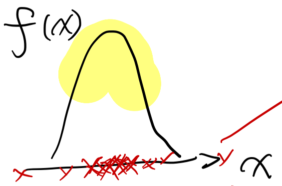
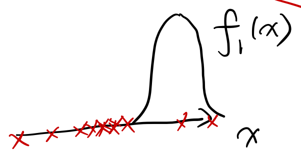
MLE (Maximum Likelihood Estimator)



$\{x_i\}_{i=1}^{n}$ $\quad \underbrace{\prod_{i=1}^{n} f(x_i)}_{\text{likelihood}} \leftarrow$ high



$\{x_i\}_{i=1}^{n}$ $\quad \prod_{i=1}^{n} f_1(x_i) \leftarrow$ low

The correct param maximizes the LH.

① Write the LH function of data $(x_1, \dots, x_n)$, or the joint dist, $f(\theta) = f_{X_1 X_2 \dots X_n}(x_1, x_2, \dots, x_n) = \prod_{i=1} f_{X_i}(x_i)$.

② Find $\hat{\theta}$ that maximizes the (log) likelihood.

Ex: $(X_1, \dots, X_n) \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, σ^2 known.

(i.i.d. = independently identically distributed)

i) $f(\theta) = \prod_{i=1}^{n} \left(\frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} \right)$, where $\theta \overset{\text{def}}{=} \mu$

ii) $\ln f(\theta) = \sum_{i=1}^{n} \left[\ln\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right) - \frac{(x_i-\mu)^2}{2\sigma^2} \right]$

$$\frac{\partial}{\partial \mu} f(\mu) = \sum_{i=1}^{n} \underbrace{-\frac{1}{2\sigma^2}}\, 2\underbrace{(x_i-\mu)}(-1) = \Big|_{\mu=\hat{\mu}} 0$$

$$\sum_{i=1}^{n} x_i = \sum \hat{\mu} = n\hat{\mu} \Rightarrow \hat{\mu}_{MLE} = \frac{1}{n}\sum X_i = \bar{X}. \quad \square$$

Ex: Find $\hat{p}_{MLE}$ for $(X_1, \dots, X_n) \overset{iid}{\sim} Ber(p)$

$$\left(\begin{array}{l} P[X=1] = p \\ P[X=0] = 1-p \end{array} \right)$$

$\Downarrow$ trick

$$P[X=x] = p^x (1-p)^{1-x}$$

$$L(\theta) = p_{X_1, \dots, X_n}(x_1, \dots, x_n)$$

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

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

$$\ln L(\theta) = \sum x_i \ln p + \left[\sum (1-x_i)\right] \cdot \ln(1-p) \quad , \quad \theta \overset{def}{=} p$$

$$\frac{\partial}{\partial \theta} L(\theta) = \sum x_i \frac{1}{p} + \sum (1-x_i) \frac{1}{1-p} (-1) \overset{set}{=} 0$$

$$(1-\hat{p}) \sum_{i=1}^{n} x_i = \hat{p} \sum_{i=1}^{n} (1-x_i)$$

$$\sum x_i - \underbrace{\hat{p} \sum x_i} = n\hat{p} - \underbrace{\hat{p} \sum x_i} \implies \hat{p}_{MLE} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

Invariance Principle of MLE:

Let $\hat{\theta}_1, \dots, \hat{\theta}_n$ be the MLEs for $\theta_1, \dots, \theta_n$, then $h(\hat{\theta}_1, \dots, \hat{\theta}_n)$ is the MLE for $h(\theta_1, \dots, \theta_n)$.

Ex: $\widehat{\sigma^2_{MLE}} = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2$, then

$$\hat{\sigma}_{MLE} = \sqrt{\widehat{\sigma^2_{MLE}}} = \left[\frac{1}{n} \sum (X_i - \bar{X})^2 \right]^{\frac{1}{2}} . \quad \text{(Devore)}$$