The objective of estimation:

* Find values of parameters of models/ distribution using data. These values of parameter are called parameter estimates.

* Estimation Techniques
  - Method of moments
  - maximum Likelihood Estimation
  - Bayesian Estimation

* Maximum Likelihood Estimation

consider a random variable $X \sim f(x;\theta)$, Probability distribution with parameters $\theta \in \mathbb{R}^m$; $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix}$;

Simplicity, Let us assume $m=1$.

* Data: $x_1, x_2, \ldots, x_N$ (Random samples)
* Objective: Estimate a value of parameter, θ.
* Likelihood function:

$$\cancel{L(\theta)} = L(x_1,..,x_n/\theta) = L(\theta / x_1 \ldots x_N)$$
$$= f(x_1;\theta) \cdot f(x_2;\theta), \ldots, f(x_N,\theta)$$
$$= \prod_{i=1}^{N} f(\underbrace{x_i}_{\text{Data}};\underbrace{\theta}_{\text{Unknown parameter}})$$

* Maximum Likelihood Estimation: finds a value of θ that maximizes $L(\theta)$

* Example: Data: $x_1, \ldots x_N$, Expontial distribution; $Exp(x_i;\lambda)$
  Estimate 'λ' using data.

$$f(x;\lambda) = \lambda e^{-\lambda x}, \quad x \geq 0$$
$$= 0 \quad , \quad x < 0$$

* Likelihood function

$$\mathcal{L}(\lambda) = \mathcal{L}(\lambda/\theta) = \prod_{i=1}^{N} f(x_i,\lambda) = \prod_{i=1}^{N} \lambda e^{-x_i \lambda}$$
$$= \lambda^n e^{-\lambda \sum_{i=1}^{N} x_i}$$

* $\ln(\mathcal{L}(\lambda))$: Log likelihood function

$$= n \ln \lambda - \lambda \sum_{i=1}^{N} x_i$$

$$\frac{\partial \ln(\mathcal{L}(\lambda))}{\partial \lambda} = 0 \iff \frac{n}{\lambda} - \sum_{i=1}^{n} x_i \iff \lambda = \frac{1}{\bar{x}}$$

$$\frac{\partial^2 \ln \mathcal{L}(\lambda)}{\partial \lambda^2} = -\frac{n}{\lambda^2} < 0$$

* Example 2 : Data: $x_1, \dots, x_n$

Distribution: Normal with μ, σ^2 parameters

Unknown Parameters: μ, σ^2

* Likelihood function:

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} f(x_i, \mu, \sigma^2)$$

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

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$$

* $\ln(L(\cdot)) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$

* $\frac{\partial \ln(L(\cdot))}{\partial \mu} = 0 \iff +\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0$

$$\iff \hat{\mu} = \frac{\sum_{i=1}^{n} x_i}{n} = \bar{x}$$

* $\frac{\partial \ln(L(\cdot))}{\partial \sigma^2} = 0 \iff -\frac{n \cancel{2\pi}}{2 \cancel{2\pi}\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0$

$$\iff +n\hat{\sigma}^2 = \sum_{i=1}^{n}(x_i-\mu)^2$$

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

* MLE for mean is the sample mean and variance is a biased version of the sample variance $= \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\mu)^2$

* Properties of MLE

→ How good are the estimates?

Minimum variance unbiased estimator (MVUE)

* MLE with large enough samples

$n \to \infty$

1) $\hat{\theta}$ (estimate of parameters) is approximately unbiased estimator

$E(\hat{\theta}) \approx \theta \to$ True value.

1) $\hat{\theta}$ (estimate of parameters) is approximately unbiased estimator

$$\boxed{E(\hat{\theta}) \simeq \theta} \rightarrow \text{True value}$$

2) $Var(\hat{\theta})$: Nearly as small as the variance that could be obtained with any other estimators

$\therefore$ $\boxed{\text{MLE} \approx \text{MVUE}}$

* Large sample property: $s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \mu)^2$

$$E[\hat{\sigma}^2] = \frac{n-1}{n} \sigma^2$$

$$\text{Bias} = E[\hat{\sigma}^2] - \sigma^2 = -\frac{\sigma^2}{n}$$

$n \to \infty$ (Large sample); Bias $\to 0$

MLE: Underestimate of σ^2

3) Invariance Property:

objective: Given sample, $x_1, \dots, x_n$, find an estimate for a function of θ, $h(\theta)$.

we can use the invariance property of MLE. as follows.

If $\hat{\theta}$ is an MLE of θ, then $h(\hat{\theta})$ is an MLE of $h(\theta)$ when $h(\cdot)$ is one-to-one function.

Note that $y = f(x)$ is a one-to-one function when $x = f^{-1}(x)$ or Existance of Inverse of f.

* Consider $h(\hat{\theta}) = \hat{\beta} \longrightarrow \hat{\theta} = h^{-1}(\hat{\beta})$
* construct Likelihood function for β

$$\max_{\beta} L^*(\beta) = \max_{\beta} \prod_{i=1}^{n} f(\beta, x_i)$$
$$= \max_{\beta} \prod_{i=1}^{n} f(h^{-1}(\beta), x_i)$$
$$= \max_{\theta} \prod_{i=1}^{n} f(\theta, x_i) = \max_{\theta} L(\theta)$$

* Then, the value of $\hat{\beta}$ that maximizes $L^*(\beta)$ will be called an MLE of $\beta = h(\theta)$.

* The invariance Property of MLE in

* The invariance Property of MLE in multivariate cases.

If $\hat{\theta}_1, \dots \hat{\theta}_k$ are MLEs of $\theta_1, \dots \theta_k$.

Then, $h(\hat{\theta}_1, \dots \hat{\theta}_k)$ is MLE of $h(\theta_1, \dots, \theta_k)$.