Methods of Point Estimation

Goal: Introduce two methods → Method of moments
→ Maximum Likelihood Estimation.

Suppose

Population X has distribution → pmf/pdf of X is $f(x)$.

Let $\{X_1, X_2, \dots, X_n\}$ → random sample of size n from population.

Then for $k = 1, 2, 3, \dots$

① The k^{th} population moment or the k^{th} distribution moment → The expected value of the r.v. X^k ⇕ $E(X^k)$.

② The k^{th} sample moment for the random sample $\{X_1, X_2, \dots, X_n\}$ → $\dfrac{X_1^k + X_2^k + \dots + X_n^k}{n}$ ⇕ $\dfrac{\sum_{i=1}^{n} X_i^k}{n}$

The Method of Moments

If there are m parameters: $\theta_1, \theta_2, \theta_3, \dots, \theta_m$ that we want to estimate using $\{X_1, X_2, \dots, X_n\}$

⇕

get m-equations by "equating" the first m population moments to the first m sample moments

⇕

Pray that you can solve these equations to get estimators $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ for $\theta_1, \theta_2, \dots, \theta_m$ resp.

Example

① Sampling from exponential distribution

Suppose $\{X_1, X_2, X_3, \dots, X_n\}$ random sample from $\exp(\lambda)$.

⇕

only one parameter to estimate, ie λ.

⇕

only need to calculate first moments.

$\boxed{\text{first population moment} \to E(X) = \frac{1}{\lambda}}$

$\boxed{\text{first sample moment} \to \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right) = \bar{X} \text{ sample mean}}$

∴ Equating the population moments to sample moments:

$$\frac{1}{\lambda} = \bar{X} \Rightarrow \lambda = \frac{1}{\bar{X}}$$

∴ $\boxed{\text{Method of moments estimator for } \lambda \to \hat{\lambda} = \frac{1}{\bar{X}}}$

② Sampling from Normal distribution

ie $\{X_1, X_2, \dots, X_n\}$ coming from $N(\mu, \sigma^2)$

∴ want to estimate μ and σ^2 (ie $k=2$).

(i) Calculate k^{th} population moments:

$\boxed{k=1 \to E(X) = \mu}$

$k=2 \to E(X^2)$, recall awesome identity

⇕

$$\sigma^2 = V(X) = E(X^2) - (E(X))^2 = E(X^2) - \mu^2$$

∴ $\boxed{E(X^2) = \sigma^2 + \mu^2}$

Equating population moments to sample moments:

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

$$E(X^2) = \sigma^2 + \mu^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

∴ The moment estimators are

$\boxed{\hat{\mu} = \bar{X}}$ and $\boxed{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}$

③ Sampling from Gamma Distribution

$\{X_1, X_2, \dots, X_n\}$ from Gamma(α, β)

ie $\alpha \to$ shape

$\beta \to$ scale

Recall: $\boxed{E(X) = \alpha \cdot \beta}$ $\boxed{V(X) = \alpha\beta^2}$

∴ $E(X^2) - E(X)^2 = \alpha\beta^2$

⇒ $E(X^2) = \alpha\beta^2 + (E(X))^2$

∴ Population moments are

$\boxed{E(X) = \alpha\beta}$ $\boxed{E(X^2) = \alpha\beta^2 + (\alpha\beta)^2 = \alpha\beta^2(1+\alpha)}$

∴ Equating sample moments to pop'l moments we get

$\boxed{\bar{X} = \alpha \cdot \beta}$ and $\boxed{\frac{\sum_{i=1}^{n} X_i^2}{n} = \alpha\beta^2(1+\alpha)}$

Let $\boxed{A = \bar{X}}$ and $\boxed{B = \frac{1}{n}\sum_{i=1}^{n} X_i^2}$,

need to solve the following for α, β

$\boxed{A = \alpha \cdot \beta \quad \text{and} \quad B = \alpha\beta^2(1+\alpha)}$

⇒ $\boxed{\beta = \frac{A}{\alpha}}$, ∴ $B = \alpha \cdot \left(\frac{A}{\alpha}\right)^2 (1+\alpha)$

ie $B = \cancel{\alpha} \cdot \frac{A^2}{\alpha^{\cancel{2}}}(1+\alpha) = \frac{A^2(1+\alpha)}{\alpha}$

∴ $\alpha \cdot B = A^2 + A^2 \cdot \alpha \Rightarrow \alpha(B - A^2) = A^2$

⇒ $\boxed{\alpha = \frac{A^2}{(B - A^2)}}$

∴ The method of moment estimators are:

$\boxed{\hat{\alpha} = \frac{\bar{X}^2}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \bar{X}^2}}$, $\boxed{\hat{\beta} = \frac{\bar{X}}{\hat{\alpha}}}$

Maximum Likelihood Estimation

Suppose $\{X_1, X_2, \dots, X_n\}$ → random sample from pop'^{n} with pdf/pmf: $f(x)$.

Let $f(x_1, x_2, \dots, x_n)$ → "joint pdf/pmf" of $X_1, X_2, \dots, X_n$

⇕

X_i's are independent,

$\boxed{f(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i)}$

Now, given sample data → $\{x_1, x_2, \dots, x_n\}$

⇕

want to estimate parameters $\theta_1, \theta_2, \dots, \theta_m$

The Likelihood function for sample data $\{x_1, x_2, \dots, x_n\}$ → $f(\underbrace{x_1, x_2, \dots, x_n}_{\text{fixed sample data}}; \underbrace{\theta_1, \theta_2, \dots, \theta_m}_{\text{unknown parameters}})$

unknown parameters ⇕ vary over all possible choices of parameters.

⇕

"Likelihood function" is function of the parameters $\theta_1, \theta_2, \dots, \theta_n$ given sample data $\{x_1, x_2, \dots, x_n\}$

$\boxed{\text{Maximum likelihood Estimation}}$ → find choices of params $\hat{\theta}_1, \dots, \hat{\theta}_m$ which maximize the likelihood function for the observed sample data $\{x_1, x_2, x_3, \dots, x_n\}$

That is,

find $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ such that

$$f(x_1, x_2, \dots, x_n; \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m) \geq f(x_1, x_2, \dots, x_n; \theta_1, \theta_2, \dots, \theta_m)$$

for all possible values of $\theta_1, \theta_2, \dots, \theta_n$.

$\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ → maximum likelihood estimates (mle) for $\theta_1, \theta_2, \dots, \theta_m$ resp, using $\{x_1, x_2, \dots, x_n\}$

If we substitute X_i for x_i, get estimators

$\boxed{\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m \to \text{maximum likelihood estimators (mle) for } \theta_1, \theta_2, \dots, \theta_m \text{ resp.}}$

Examples

① Sampling from exponential distribution.

$\{X_1, X_2, \dots, X_n\}$ rand sample from $\exp(\lambda)$.

$\boxed{\text{pdf of } X_i \to f(x_i; \lambda) = \lambda e^{-\lambda x_i}}$

∴ The likelihood function is

$$f(x_1, x_2, \dots, x_n; \lambda) = \prod_{i=1}^{n} f(x_i; \lambda)$$
$$= \lambda e^{-\lambda x_1} \cdot \lambda e^{-\lambda x_2} \dots \lambda e^{-\lambda x_n}$$
$$= \lambda^n e^{-\lambda(x_1 + x_2 + \dots + x_n)}$$

∴ $\boxed{f(x_1, x_2, \dots, x_n; \lambda) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right)}$

⇕

want to maximize this as a function of λ's

If $\boxed{g(\lambda) = \lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right)}$,

$$\ln(g(\lambda)) = \ln\left(\lambda^n \exp\left(-\lambda \sum_{i=1}^{n} x_i\right)\right) = \ln(\lambda^n) + \ln\left(\exp\left(-\lambda \sum_{i=1}^{n} x_i\right)\right)$$

$= n\ln(\lambda) - \lambda \sum_{i=1}^{n} x_i$ → $\boxed{\text{log-likelihood function.}}$

⇕ easier to maximize.

Let $\boxed{h(\lambda) = \ln(g(\lambda))}$, then

$$h'(\lambda) = \frac{d}{d\lambda}\left(n\ln(\lambda) - \lambda \sum_{i=1}^{n} x_i\right) = n \cdot \frac{1}{\lambda} - \sum_{i=1}^{n} x_i$$

∴ $h'(\lambda) = 0 \Rightarrow n \cdot \frac{1}{\lambda} = \sum_{i=1}^{n} x_i \Rightarrow \boxed{\lambda = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i}}$

Also, $\boxed{h''(\lambda) = \frac{-n}{\lambda^2} < 0}$ ⇒ $\lambda = \frac{1}{\frac{1}{n}\sum_{i=1}^{n} x_i}$ is a pt of local maxima.

Since $\ln$ is an increasing function,

$\lambda = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^{-1}$ maximizes $h(\lambda)$ ⇒ $\lambda = \left(\frac{1}{n}\sum_{i=1}^{n} x_i\right)^{-1}$ maximizes the likelihood function $g(\lambda)$

∴ $\boxed{\text{mle for } \lambda \to \hat{\lambda} = \frac{1}{\bar{X}}}$

Note: this is the same as the moment estimator for λ.

② Sampling from Normal Distribution

$\{X_1, X_2, \dots, X_n\}$ → rand sample from $N(\mu, \sigma^2)$

$\boxed{X_i \text{ has pdf} \to f(x_i; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \cdot e^{-(x_i-\mu)^2/2\sigma^2}}$

The likelihood function is given as

$$f(x_1, x_2, \dots, x_n; \mu, \sigma^2) = \prod_{i=1}^{n} f(x_i; \mu, \sigma^2)$$
$$= \left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n \cdot e^{-\sum_{i=1}^{n}\left((x_i-\mu)^2/2\sigma^2\right)}$$

so that log-likelihood function is

$$\ln\left(f(x_1, x_2, \dots, x_n; \mu, \sigma^2)\right) = \ln\left(\frac{1}{\sqrt{2\pi}\,\sigma}\right)^n + \ln\left(e^{-\sum_{i=1}^{n}\left((x_i-\mu)^2/2\sigma^2\right)}\right)$$
$$= \frac{n}{2}\ln\left(\frac{1}{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2$$

∴ $\boxed{h(\mu, \sigma^2) = \frac{n}{2}\ln\left(\frac{1}{2\pi\sigma^2}\right) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2}$

Need to maximize $h(\mu, \sigma^2)$ → calculate critical values (ie $\frac{\partial h}{\partial \mu} = 0$ and $\frac{\partial h}{\partial \sigma^2} = 0$)

⇕

use Hessian matrix to show that these are max.

We can show that the mle's are

$\boxed{\hat{\mu} = \bar{X}}$ and $\boxed{\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$

Note: While $\hat{\mu}$ is unbiased, $\hat{\sigma}^2$ is biased!

Properties of MLE's

① Invariance Principle → If $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m$ are mle's for $\theta_1, \theta_2, \dots, \theta_m$, then $h(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_m)$ is an mle for $h(\theta_1, \theta_2, \dots, \theta_m)$.

② Large Sample Behavior → When sample size is large, the mle $\hat{\theta}$ for θ is atleast approx unbiased (ie $E(\hat{\theta}) \approx \theta$) and has variance that is either small as (or nearly as small as) can be achieved by any estimator.

$\boxed{\therefore \text{The mle } \hat{\theta} \text{ is exactly or atleast approx the MVUE for } \theta.}$