Methods of Point Estimation

Goal: Introduce two methods $\longrightarrow$ Method of moments
$\qquad\qquad\qquad\qquad\qquad\quad\;\;\hookrightarrow$ Maximum Likelihood Estimation.

Suppose

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

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

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

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

② The k^{th} sample moment for the random sample $\{X_1, X_2, \ldots, X_n\}$ $\longrightarrow$ $\dfrac{X_1^k + X_2^k + \cdots + X_n^k}{n}$ $\Updownarrow$ $\dfrac{\sum_{i=1}^n X_i^k}{n}$

The Method of Moments

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

$\Updownarrow$

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

$\Updownarrow$

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

Example

① Sampling from exponential distribution

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

$\Updownarrow$

only one parameter to estimate, ie λ.

$\Updownarrow$

only need to calculate first moments.

first population moment $\longrightarrow E(X) = \dfrac{1}{\lambda}$

first sample moment $\longrightarrow \dfrac{1}{n}\left(\sum_{i=1}^n X_i\right) = \overline{X}$ sample mean

$\therefore$ Equating the population moments to sample moments:

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

$\therefore$ Method of moments estimator for λ $\longrightarrow \hat{\lambda} = \dfrac{1}{\overline{X}}$

② Sampling from Normal distribution

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

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

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

$k=1 \longrightarrow E(X) = \mu$

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

$\Updownarrow$

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

$\therefore \boxed{E(X^2) = \sigma^2 + \mu^2}$

Equating population moments to sample moments:

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

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

$\therefore$ The moment estimators are

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

③ Sampling from Gamma Distribution

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

ie $\alpha \longrightarrow$ shape
$\quad\;\beta \longrightarrow$ scale

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

$\therefore E(X^2) - E(X)^2 = \alpha\beta^2$

$\Rightarrow E(X^2) = \alpha\beta^2 + (E(X))^2$

$\therefore$ Population moments are

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

$\therefore$ Equating sample moments to popl'n moments we get

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

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

need to solve the following for α, β

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

$\Rightarrow \boxed{\beta = \dfrac{A}{\alpha}}$, $\therefore B = \alpha \cdot \left(\dfrac{A}{\alpha}\right)^2 (1+\alpha)$

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

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

$\Rightarrow \boxed{\alpha = \dfrac{A^2}{(B - A^2)}}$

$\therefore$ The method of moment estimators are:

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

Maximum Likelihood Estimation

Suppose $\{X_1, X_2, \ldots, X_n\} \longrightarrow$ random sample from popl'n with pdf/pmf: $f(x)$.

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

$\Updownarrow$

X_i's are independent,

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

Now, given sample data $\longrightarrow \{x_1, x_2, \ldots, x_n\}$

$\Updownarrow$

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

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

unknown parameters $\Updownarrow$ vary over all possible choices of parameters.

$\Updownarrow$

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

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

That is,

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

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

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

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

If we substitute X_i for x_i, get estimators

$\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_m \longrightarrow$ maximum likelihood estimators (mle) for $\theta_1, \theta_2, \ldots, \theta_m$ resp.

Examples

① Sampling from exponential distribution.

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

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

$\therefore$ The likelihood function is

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

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

$\Updownarrow$

want to maximize this as a function of λ's

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

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

$= n\ln(\lambda) - \lambda \sum_{i=1}^n x_i \longrightarrow$ $\boxed{\text{log-likelihood function.}}$ $\Updownarrow$ easier to maximize.

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

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

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

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

Since $\ln$ is an increasing function,

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

$\therefore \boxed{\text{mle for } \lambda \longrightarrow \hat{\lambda} = \dfrac{1}{\overline{X}}}$

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

② Sampling from Normal Distribution

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

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

The likelihood function is given as

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

So that log-likelihood function is

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

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

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

$\Updownarrow$

use Hessian matrix to show that these are max.

We can show that the mle's are

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

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

Properties of MLE's

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

② Large Sample Behavior $\longrightarrow$ 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.

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

Method of Moments.

Population moments. ⟺ quantities involving the parameters of interest

Sample moment ⟺ quantities involving the random sample.

Want. To repackage the problem of finding estimators.

⟺

problem of finding solutions to a system of equation.

---

n-linear equations in n-variable $\xrightarrow[\text{some condition}]{\text{under}}$ can guarantee solutions

---

$\{X_1, X_2, \dots, X_n\}$ random sample popl^n pdf $f(x)$.

If $k = 1, 2, 3, \dots$

k^{th} population moment $= E(X^k)$

For example $k=1$ gives $\boxed{E(X) = \mu}$ = mean of the populat.

$k=2$; Want $E(X^2)$

$$V(X) = E(X^2) - (E(X))^2$$
$$\Rightarrow E(X^2) = V(X) + (E(X))^2$$
$$\boxed{E(X^2) = \sigma^2 + \mu^2}$$

The k^{th} sample moment $=$ Avg of $\{X_i^k ;\ i = 1, 2, \dots, n\}$

$$= \frac{X_1^k + X_2^k + \dots + X_n^k}{n}$$

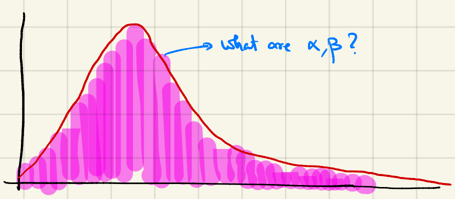
Method of moments: If estimating $\theta_1, \theta_2, \dots, \theta_m$

Calculate the m popl^n moments and the m- sample moments.

And equating these m-moments gives m-equations that involve $\theta_1, \theta_2, \dots, \theta_m$, and $X_1, X_2, \dots, X_n$

⟺

Solve for θ_i's.



data $= \{x_1, x_2, x_3, \dots, x_{150}\}$

⟺

use these 150 numbers to estimate α, β.

Suppose sampling from Gamma(α, β)

$\{X_1, X_2, \dots, X_n\}$ rand sample

$X_i \sim$ Gamma(α, β)

Want estimators for α, β.

Need to calculate the first two popl^n and sample moments.

1) $E(X) = \mu_X = \alpha\beta$

2) $E(X^2) = ?$ Note $V(X) = E(X^2) - (E(X))^2$

$$\Rightarrow E(X^2) = V(X) + (E(X))^2$$
$$= \alpha\beta^2 + (\alpha\beta)^2$$
$$E(X^2) = \alpha\beta^2(1+\alpha)$$

Equating population moments to sample moments.

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

$\alpha, \beta > 0$

$\alpha\beta = A$ and $\alpha\beta^2(1+\alpha) = B$

$$\boxed{\beta = \frac{A}{\alpha}}$$

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

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

$$A^2(1+\alpha) = \alpha B \Rightarrow A^2 + A^2\alpha = \alpha B$$
$$\Rightarrow \alpha(B - A^2) = A^2$$
$$\Rightarrow \boxed{\alpha = \frac{A^2}{B - A^2}}$$

∴ Method of moment estimators are

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

Use R to calculate empirical bias of $\hat{\alpha}$ and $\hat{\beta}$

Given a random sample $\{X_1, X_2, \ldots, X_n\}$ coming from $N(\mu, \sigma^2)$

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

$$S_{xx} = \underline{\underline{\sum_{i=1}^{n}(X_i - \bar{X})^2}} \rightsquigarrow \text{squared deviations of } X_i \text{ from the sample mean } \bar{X}.$$

$$\underline{\underline{S^2 = \frac{1}{(n-1)} S_{xx}}} \qquad \underline{\underline{\hat{\sigma}^2 = \frac{1}{n} S_{xx}}} \quad (\text{if } X \sim N(\mu, \sigma^2))$$

Want to study/understand S_{xx}.

<u>Goal</u>: Calculate $E(S_{xx})$?

$$\begin{aligned}
S_{xx} &= \sum_{i=1}^{n}(X_i - \bar{X})^2 \\
&= \sum_{i=1}^{n}(X_i - \mu + \mu - \bar{X})^2 \\
&= \sum_{i=1}^{n}\left((\underline{X_i - \mu}) - (\underline{\bar{X} - \mu})\right)^2 \\
&= \sum_{i=1}^{n}\left((X_i - \mu)^2 - 2(X_i - \mu)(\bar{X} - \mu) + (\bar{X} - \mu)^2\right) \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\bar{X} - \mu)\sum_{i=1}^{n}(X_i - \mu) + \sum_{i=1}^{n}(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - 2(\bar{X} - \mu)(n\bar{X} - n\mu) + n(\bar{X} - \mu)^2 \\
&= \sum_{i=1}^{n}(X_i - \mu)^2 - \cancel{2}n(\bar{X} - \mu)^2 + \cancel{n(\bar{X} - \mu)^2}
\end{aligned}$$

$$S_{xx} = \sum_{i=1}^{n}(X_i - \mu)^2 - \frac{(\bar{X} - \mu)^2}{1/n}$$

If $Z \sim N(0,1) \implies Z^2 \sim$ Chi sq dist with $df = 1$.

$$\frac{S_{xx}}{\sigma^2} = \sum_{i=1}^{n}\left(\frac{X_i - \mu}{\sigma}\right)^2 - \left(\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right)^2$$