Theory of Inference

Definition:

Every statistical enquiry has, as its goal, the acquisition of knowledge about the characteristics of a group of individuals called population. The enquirer will generally have on hand information concerning a part of the population only, ie about only those individuals of the population which are included in the sample. The central problem of statistics is, then, to devise methods of inferring the unknown nature of the population from the known nature of the sample. This central problem is known as the problem of statistical inference.

Statistical Inference

- Estimation
  - Point Estimation
  - Interval estimation
- Testing of hypothesis.

Two types of inference

Statistical inference may be of two ~~different~~ main types —

The characteristics of the population in which we are interested, may be completely unknown & we may depend entirely on the sample to form an idea about (or to estimate) the characteristic. This type of inference is called estimation. Estimation may again be point estimation & interval estimation. In the first case, we compute a single value on the basis of the sample observations & put forward this value as the likely value (or estimate) of the population characteristic. In the latter case, we compute two values on the basis of sample observations & put forward the interval between these values as ~~the~~ likely to contain the true value of the characteristic. These two types are called point estimation & interval estimation respectively.

In the 2^{nd} type of statistical inference, we have to start with some tentative idea about the unknown population characteristic & we shall use the sample observations to judge how tenable this idea is. The tentative idea is called a hypothesis & the process of judging its reality is called a test of the hypothesis.

Estimation :

In statistical inference, sometimes the characteristics of the population in which we are interested, may be completely unknown & we may depend entirely on the sample to form an idea about (or to estimate) the characteristic. This type of inference is called estimation.

Estimation may again be point estimation & interval estimation. In the first case, we compute a single value on the basis of the sample obs. & put forward this value as the likely value (or estimate) of the population characteristic. In the latter case, we compute two values on the basis of sample observations & put forward the interval between these values as likely to contain the true value of the characteristic. These two types are called point estimation & interval estimation respectively.

Let us now make a distinction between an estimator & an estimate. If the sample observations, regarded as random variables, are $X_1, X_2, \ldots, X_n$, then any function of these random variables is itself a random variable & is called a statistic. On the other hand, any numerical characteristic of a population like its mean, variance etc. is called a parameter, generally denoted by θ.

If T be a statistic used in estimating θ, then we call 'T' as an estimator, while a particular value assumed by T in any particular case (i.e. for a given set of sample observations) is called an estimate.

Point Estimation

On the basis of sample observations, if one puts forward a single value, i.e a point, as an estimate, then, this category of estimation is known as point estimation. The estimator, used for this purpose, is known as a point estimator & its value for a given set of sample observations is known as a point estimate.

For example, in order to guess the unknown population proportion (p) one may put forward the sample ~~p~~ proportion as a point estimator. In that case, the value of the sample proportion for a given set of sample observations is a point estimate of 'p'.

Point estimation admits two problems:

1) To devise some means of obtaining a statistic which can be used as an estimator;

2) To find the "best" estimator among many possible estimators.

Methods of Estimation: There are a no. of methods for estimating an unknown parameter θ involved in a distribution. Among them are some popular & important methods are,

i) Method of moment estimation,

ii) Method of maximum likelihood estimation,

iii) Method of least square estimation & so on.

(ii) Method of Moment estimation [MME]:

Let, $x_1, x_2, \dots, x_n$ be 'n' sample observations, where, $X_i \overset{iid}{\sim} f_{\underset{\sim}{\theta}}(x_i)$, $\forall i = 1(1)n$, where, $\underset{\sim}{\theta} = (\theta_1, \theta_2, \dots, \theta_s)'$ is the unknown parametric vector, $s \geq 1$.

Let us define, $\mu_r' = E(X^r) = r^{th}$ order population raw moment, provided it exists, $\forall\ r = 1, 2, \dots$ & $m_r' = \frac{1}{n}\sum_{i=1}^{n} x_i^r = r^{th}$ order sample raw moment, $r \geq 1$.

Then, obviously, $\mu_r' = g_r(\underset{\sim}{\theta})$, $r \geq 1$, i.e.

$$\underset{\sim}{\mu}' = \begin{pmatrix} \mu_1' \\ \mu_2' \\ \vdots \\ \mu_s' \end{pmatrix} = \begin{pmatrix} g_1(\underset{\sim}{\theta}) \\ g_2(\underset{\sim}{\theta}) \\ \vdots \\ g_s(\underset{\sim}{\theta}) \end{pmatrix} = \underset{\sim}{g}(\underset{\sim}{\theta}) \cdots\cdots (*)$$

Now, by MME, we can write, $\mu_r' = m_r'$, $r = 1(1)s$.

$\therefore$, By MME, we have from $(*)$ that, $\underset{\sim}{g}(\underset{\sim}{\theta}) = \underset{\sim}{m}' = \begin{pmatrix} m_1' \\ m_2' \\ \vdots \\ m_s' \end{pmatrix}$.

If the function g^{-1} exists ~~exit~~ uniquely, then, we have,

$\underset{\sim}{\theta} = \underset{\sim}{g}^{-1}(\underset{\sim}{m}') = \underset{\sim}{h}(\underset{\sim}{m}')$, i.e,

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_s \end{pmatrix} = \begin{pmatrix} h_1(\underset{\sim}{m}') \\ h_2(\underset{\sim}{m}') \\ \vdots \\ h_s(\underset{\sim}{m}') \end{pmatrix} \Leftrightarrow \hat{\theta}_r = h_r(\underset{\sim}{m}') = h_r(m_1', \dots, m_s'), \ \forall\ r = 1(1)s.$$

Note:

① Provided $\underset{\sim}{g}^{-1}$, which is equivalent to $\underset{\sim}{h}$, exists uniquely, we define $\hat{\theta}_r$ as the unique MME of the parameter θ_r, $\forall\, r = 1(1)s$.

② If $\underset{\sim}{g}^{-1}$ does not exist, then, any solution to the equation $\underset{\sim}{g}(\underset{\sim}{\theta}) = \underset{\sim}{m}'$ will have the MME of $(\theta_1, \theta_2, \ldots, \theta_s)$ & then it will not be unique.

③ Generally, MME of θ is denoted by $\hat{\theta}_{MME}$.

④ Generally, MME is not an efficient estimator of the corresponding parameter.

⑤ Generally, MME provides as the initial or rough estimation of the corresponding unknown parameter.

Example: Let $X_1, X_2, \dots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$. Then, for $r=1$, $E(X_i^r) = E(X_1) = \mu_1' = \mu$.

& for $r=2$, $E(X_i^r) = E(X_i^2) = V(X_i) + E^2(X_i) = \sigma^2 + \mu^2 = \mu_2'$

Now, let us define, $m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{x}$ & $m_2' = \frac{1}{n}\sum_{i=1}^{n} x_i^2$.

$\therefore$ By MME, we have, $\mu_1' = m_1'$ ie, $\mu = \bar{x}$ ie $\hat{\mu}_{MME} = \bar{x}$ &

$\mu_2' = m_2'$ ie, $\sigma^2 + \mu^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ $\therefore$ $\hat{\sigma}^2_{MME} = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

$\therefore$ $\hat{\sigma}^2_{MME} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

Following situations may arise for MME.

(1) MME may not be unique:

Example: Let $X_1, X_2, \dots, X_n \overset{iid}{\sim} P(\lambda)$. Then, $E(X_1) = \mu_1' = \lambda$ & $V(X_1) = \lambda = \mu_2$.

Define, $m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ & $m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2$.

Now, by MME, we have, $\mu_1' = m_1'$ ie $\lambda = \bar{X}$ & $\lambda = S^2$.

However, generally $\bar{X} \neq S^2$. $\therefore$ $\hat{\lambda}_{MME}$ may not be unique.

(2) Modified MME:

Example: Let $X_1, X_2, \dots, X_n \overset{iid}{\sim} \frac{1}{2\theta} e^{-\frac{|x|}{\theta}}$, $-\infty < x < \infty$. [Laplace or double exponential distn]

Then, $E(X_1) = \mu_1' = 0$, but $m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$.

So, it is not possible frequently that $\bar{X} = 0$ (ie, it is meaningless to equate $\bar{X}$ with 0). So we need some ~~modifica~~ modifications.

Modification (a): $E(X_1^2) = \mu_2' = \frac{2}{2\theta}\int_0^\infty x^2 e^{-x/\theta}\,dx = \frac{2}{2\theta}\int_0^\infty e^{-\left(\frac{1}{\theta}\right)x} \cdot x^{3-1}\,dx$

$= \frac{1}{\theta} \cdot \frac{\Gamma 3}{\left(\frac{1}{\theta}\right)^3} = 2\theta^2$. Define $m_2' = \frac{1}{n}\sum_{i=1}^{n} X_i^2$.

$\therefore$ By MME, we have, $2\theta^2 = m_2'$ ie $\theta^2 = \frac{m_2'}{2}$. $\therefore \theta = \sqrt{\frac{m_2'}{2}}$, $\because \theta > 0$.

modification (b) : Let us make a transformation that $Y = |X|$.
Then, it is easy to observe that $Y \sim f_\theta(y) = \frac{1}{\theta} \cdot e^{-y/\theta}$.

$\therefore E(Y) = \theta$ & $m_1' = \frac{1}{n}\sum_{i=1}^{n} Y_i = \bar{Y}$. $\therefore \hat{\theta}_{MME} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{1}{n}\sum_{i=1}^{n} |X_i|$

(3) MME may be outside the parametric space.

Suppose, $X_1, X_2, \ldots, X_n \overset{iid}{\sim} Bin(m, p)$, ie $X_i \overset{iid}{\sim} Bin(m, p)$, $\forall\, i = 1(1)n$,
where 'm' & 'p' are both unknown.

Now, $E(X_1) = mp$; $V(X_1) = mp(1-p)$, $p \in [0,1]$.

Define, $m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X}$ & $m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = S^2$.

$\therefore$ By MME, we have, $mp = \bar{X}$ ...... (1) & $mp(1-p) = S^2$ ...... (2).

$\therefore$ (2) $\div$ (1) $\Rightarrow 1 - p = \frac{S^2}{\bar{X}}$ $\therefore p = 1 - \frac{S^2}{\bar{X}}$.

Realization: Suppose, $X_1, X_2, \ldots, X_{20} \overset{iid}{\sim} Bin(m, p)$, $m \geq 20$.

Let, $X_1 = 0, X_2 = 0, \ldots, X_{19} = 0$ & $X_{20} = 20$.

$\therefore$ $S^2 = \frac{1}{20}\sum_{i=1}^{20}(X_i - \bar{X})^2$, where $\bar{X} = \frac{20}{20} = 1$.

$\therefore S^2 = \frac{1}{20}\sum_i x_i^2 - \bar{x}^2 = \frac{20^2}{20} - 1^2 = 20 - 1 = 19$.

$\therefore p = 1 - \frac{S^2}{\bar{X}} = 1 - \frac{19}{1} = 1 - 19 = -18$.

$\therefore \hat{p}_{MME} = -18 \notin [0,1]$ $\therefore$ Here, MME is outside the parametric space.

(4) MME may not be a function of sufficient statistic.

Example: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} R(\theta_1, \theta_2)$.

Hence, $E(X_1) = \frac{\theta_1+\theta_2}{2} = \mu_1'$; $V(X_1) = \frac{(\theta_2-\theta_1)^2}{12} = \mu_2$.

Hence, by MME, we have, $\mu_1' = m_1' = \bar{x}$ & $\mu_2 = m_2 = s^2$.

$\therefore \frac{\theta_1+\theta_2}{2} = \bar{x}$ & $\frac{(\theta_2-\theta_1)^2}{12} = s^2$.

$\Rightarrow \theta_1+\theta_2 = 2\bar{x}$; $\quad \theta_2-\theta_1 = \sqrt{12}\,s = 2\sqrt{3}\,s$

$\therefore \hat{\theta}_{2\,MME} = \frac{2\bar{x}+2\sqrt{3}s}{2} = \bar{x}+\sqrt{3}s$. & $\hat{\theta}_{1\,MME} = 2\bar{x} - \hat{\theta}_{2\,MME}$

$= 2\bar{x} - \bar{x} - 2\sqrt{3}s = \bar{x} - 2\sqrt{3}s$.

But we know that $X_{(1)}$ & $X_{(n)}$ are jointly sufficient for θ_1 & θ_2.

Hence, $\hat{\theta}_{1\,MME}$ & $\hat{\theta}_{2\,MME}$ are not functions of sufficient statistics.

(5) MME may be worthless:

Example: Let $X_1, X_2, \ldots, X_n \overset{iid}{\sim} R(1, \theta), \theta > 1$. Then, $E(X_1) = \frac{\theta+1}{2}$.

Hence, by MME, we have, $\mu_1' = m_1' = \bar{x}$ ie $\frac{\theta+1}{2} = \bar{x}$ $\therefore \theta = 2\bar{x}-1$

Realization: Let, $\theta = 20$ & $n = 5$. Also, let, $x_1 = 1, x_2 = 1, \ldots, x_4 = 1, x_5 = 20$

$\therefore \bar{x} = \frac{24}{5} = 4.8$ $\therefore \hat{\theta}_{MME} = 2\bar{x} - 1 = 2 \times 4.8 - 1 = 9.6 - 1 = 8.6 < x_5 = 20$.

But, $x_{(5)} \leq \theta$, for $R(1, \theta)$.

Hence, here MME is proved to be an worthless estimator.

Remark:

1) The method of moments consists in ~~es~~ estimating θ by the statistic $T(x_1, x_2, \ldots, x_n) = h\left(\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} x_i^2, \ldots, \frac{1}{n}\sum_{i=1}^{n} x_i^k\right)$, where '$h$' is some known function of x_i's.

2) From the WLLN, one can say that, $\frac{1}{n}\sum_{i=1}^{n} x_i^j \xrightarrow{P} E(x^j)$. Thus, if one is ~~interest~~ interested in estimating the population moments, the method of moments leads to consistent & unbiased estimations. Moreover, the method of moment estimators in this case are asymptotically normally distributed.

② Let $X_1, X_2, \dots, X_n$ be a sample from $G(\alpha,\beta)$. Find the MME for (α,β).

A/ Since, $X_1, X_2, \dots, X_n$ is a sample from $G(\alpha,\beta)$ so their common p.d.f. is,

$$f_{\alpha,\beta}(x) = \begin{cases} \dfrac{\alpha^{\beta}}{\Gamma\beta} \cdot e^{-\alpha x} \cdot x^{\beta-1}, & 0 < x < \infty. \\ & \alpha > 0, \beta > 0. \\ 0, & \text{o.w.} \end{cases}$$

∴ we have, $E(X_1) = \dfrac{\beta}{\alpha} = \mu_1'$; $V(X_1) = \dfrac{\beta}{\alpha^2} = \mu_2$

Let us define, $m_1' = \frac{1}{n} \sum_{i=1}^{n} X_i = \bar{X}$ & $m_2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2 = S^2$.

By MME, we have, $\mu_1' = m_1' \Rightarrow \dfrac{\beta}{\alpha} = \bar{X} \dots ①$

& $\mu_2 = m_2 \Rightarrow \dfrac{\beta}{\alpha^2} = S^2 \dots ②$.

∴ $① \div ② \Rightarrow \dfrac{\beta}{\alpha} \cdot \dfrac{\alpha^2}{\beta} = \dfrac{\bar{X}}{S^2} \Rightarrow \alpha = \dfrac{\bar{X}}{S^2}$. ∴ $\hat{\alpha}_{MME} = \dfrac{\bar{X}}{S^2}$.

Also from ①, $\beta = \bar{X}\alpha = \bar{X} \cdot \dfrac{\bar{X}}{S^2} = \dfrac{\bar{X}^2}{S^2}$. ∴ $\hat{\beta}_{MME} = \dfrac{\bar{X}^2}{S^2}$.

(3) Let, $X_1, X_2, \dots, X_n$ be a sample from $N(\mu, \sigma^2)$. Find the MME for (μ, σ^2).

Ans/ Since $X_1, X_2, \dots, X_n$ is a sample from $N(\mu, \sigma^2)$, so their common p.d.f. is given by,

$$f_{\mu, \sigma^2}(x_i) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}, & -\infty < x < \infty \\ & -\infty < \mu < \infty \\ & \sigma^2 > 0. \\ 0, \text{ o.w} \end{cases}$$

$\therefore E(X_1) = \mu = \mu_1'$; $V(X_1) = \sigma^2 = \mu_2$

Let us now define $m_1' = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$ & $m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s^2$.

$\therefore$ By MME, we have, $\mu_1' = m_1' \Rightarrow \mu = \bar{x}$ ie, $\hat{\mu}_{MME} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$.

$\mu_2 = m_2 \Rightarrow \sigma^2 = s^2$ ie $\widehat{\sigma^2}_{MME} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$.

---

(4) Let, $X_1, X_2, \ldots, X_n$ be a sample from $B(\alpha, \beta)$. Find the MME for (α, β).

Ans/ Since, $X_1, X_2, \ldots, X_n$ is a sample from $B(\alpha, \beta)$, so their common p.d.f. is given by

$$f_{\alpha,\beta}(x_i) = \begin{cases} \frac{1}{B(\alpha,\beta)} \cdot x^{\alpha-1}(1-x)^{\beta-1}, & 0 < x < 1 \\ 0, \text{ o w} & \alpha > 0, \beta > 0. \end{cases}$$

$\therefore E(X_1) = \frac{\alpha}{\alpha+\beta}$, $V(X_1) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$.

Let us now define $m_1' = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{x}$ & $m_2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{x})^2 = s^2$.

$\therefore$ By ~~defn~~ MME, we have,

$\mu_1' = m_1' \Rightarrow \frac{\alpha}{\alpha+\beta} = \bar{x} \ldots (1)$; $\mu_2 = m_2 \Rightarrow \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} = s^2 \ldots (2)$.

$\therefore \frac{②}{①} \Rightarrow \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \cdot \frac{\alpha+\beta}{\alpha} = \frac{s^2}{\bar{x}}$;

or, $\frac{\alpha\beta}{(\alpha+\beta)^3+(\alpha+\beta)^2} \cdot \frac{\alpha+\beta}{\alpha} = \frac{s^2}{\bar{x}}$ or, $\frac{\beta}{(\alpha+\beta)^2+(\alpha+\beta)} = \frac{s^2}{\bar{x}} \cdots ③$

Now, from ①, we have, $\frac{\alpha+\beta}{\alpha} = \frac{1}{\bar{x}}$. $\therefore \alpha+\beta = \frac{\alpha}{\bar{x}}$

$\therefore$ from ③ or, $1+\frac{\beta}{\alpha} = \frac{1}{\bar{x}}$ $\therefore \frac{\beta}{\alpha} = \frac{1}{\bar{x}} - 1 = \frac{1-\bar{x}}{\bar{x}}$ $\therefore \beta = \frac{(1-\bar{x})\alpha}{\bar{x}} \cdots ④$

Also, $\alpha+\beta = \frac{\alpha}{\bar{x}} \cdots ⑤$

$\therefore$ From ⑤, ③ & ④, we have,

$$\frac{\frac{(1-\bar{x})\alpha}{\bar{x}}}{\frac{\alpha^2}{\bar{x}^2}+\frac{\alpha}{\bar{x}}} = \frac{s^2}{\bar{x}} \quad \text{or,} \quad \frac{\frac{(1-\bar{x})\alpha}{\bar{x}} \cdot \bar{x}^2}{\alpha^2+\alpha\bar{x}} = \frac{s^2}{\bar{x}}$$

$$\text{or, } \frac{(1-\bar{x})\bar{x}}{\alpha+\bar{x}} = \frac{s^2}{\bar{x}} \quad \text{or, } \frac{\alpha+\bar{x}}{(1-\bar{x})\bar{x}} = \frac{\bar{x}}{s^2} \quad \text{or, } \alpha = \frac{\bar{x}^2(1-\bar{x})}{s^2} - \bar{x}$$

$$\therefore \boxed{\hat{\alpha}_{MME} = \frac{\bar{x}^2(1-\bar{x})}{s^2} - \bar{x} = \bar{x}\left[\frac{\bar{x}(1-\bar{x})}{s^2} - 1\right]}$$

$$\therefore \text{From ④, } \beta = \frac{(1-\bar{x})}{\cancel{\bar{x}}} \cdot \cancel{\bar{x}}\left[\frac{\bar{x}(1-\bar{x})}{s^2} - 1\right] = \frac{\bar{x}(1-\bar{x})^2}{s^2} - (1-\bar{x})$$

$$\therefore \boxed{\hat{\beta}_{MME} = \bar{x} + \frac{\bar{x}(1-\bar{x})^2}{s^2} - 1}$$

5) A random sample of size n is taken from the lognormal p.d.f. $f(x;\mu,\sigma) = \begin{cases} \frac{1}{\sigma\sqrt{2\pi}} \cdot \frac{1}{x} e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}, & x>0, \ -\infty<\mu<\infty, \sigma>0 \\ 0, & \text{o.w.} \end{cases}$

Find the MME ~~of~~ for μ & σ^2.

A/ Since, $x_1, x_2, \ldots, x_n$ is a sample from $\Delta(\mu, \sigma^2)$, so, their common p.d.f. is given by,

$$f_{\mu\sigma}(x) = \begin{cases} \frac{1}{x\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2}(\ln x - \mu)^2}, & 0<x<\infty, \ -\infty<\mu<\infty, \sigma>0 \\ 0, & \text{o.w.} \end{cases}$$

$\therefore E(X) = e^{\mu + \sigma^2/2} = \mu_1'$; $V(X) = e^{\sigma^2}(e^{\sigma^2}-1)e^{2\mu} = \mu_2$.

~~Again, By MME, we have, m_1~~

Again, let us define, $m_1' = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}$; $m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = s^2$.

$\therefore$ By MME, we have, $\mu_1' = m_1 \Rightarrow e^{\mu + \sigma^2} = \bar{x} \ldots (1)$

& $m_2 = \mu_2 \Rightarrow e^{\sigma^2}(e^{\sigma^2}-1)e^{2\mu} = s^2 \Rightarrow \alpha(\alpha-1)e^{2\mu} = s^2 \ldots (2)$, where, $\alpha = e^{\sigma^2}$.

$\therefore \frac{(2)}{(1)} \Rightarrow \frac{\cancel{e^{\sigma^2}}(e^{\sigma^2}-1)e^{2\mu}}{e^{\mu}\cdot \cancel{e^{\sigma^2}}} = \frac{s^2}{\bar{x}}$ or, $e^{\mu}(e^{\sigma^2}-1) = \frac{s^2}{\bar{x}} \ldots (3)$.

Again, from (1), $e^{\mu} = \frac{\bar{x}}{e^{\sigma^2}} \ldots (4)$.

$\therefore$ From (3) & (4), we have, $\frac{\bar{x}}{e^{\sigma^2}}(e^{\sigma^2}-1) = \frac{s^2}{\bar{x}}$.

or, $\left(1 - \frac{1}{e^{\sigma^2}}\right) = \frac{s^2}{\bar{x}^2}$ or, $\frac{1}{e^{\sigma^2}} = 1 - \frac{s^2}{\bar{x}^2} = \frac{\bar{x}^2 - s^2}{\bar{x}^2}$ $\therefore e^{\sigma^2} = \frac{\bar{x}^2}{\bar{x}^2 - s^2}$.

$$\therefore \boxed{\hat{\sigma}^2_{MME} = \ln \frac{\bar{x}^2}{\bar{x}^2 - s^2} = 2\ln \bar{x} - \ln(\bar{x}^2 - s^2)}$$

$\therefore$ From (4) we have, $e^{\mu} = \frac{\bar{x}}{e^{\sigma^2}} = \bar{x}\cdot\frac{\bar{x}^2 - s^2}{\bar{x}^2} = \frac{\bar{x}^2 - s^2}{\bar{x}}$

$$\therefore \boxed{\hat{\mu}_{MME} = \ln \frac{\bar{x}^2 - s^2}{\bar{x}} = \ln(\bar{x}^2 - s^2) - \ln \bar{x}.}$$

Law of large numbers states that (as the sample size increases) the sample mean of a set of iid random variables approaches their theoretical mean.

Method of moments estimation is based solely on the law of large numbers, which we repeat here:

Let $M_1, M_2, \ldots$ be independent random variables having a common distribution possessing a mean μ_M. Then the sample means converge to the distributional mean as the number of observations increase.

$$\bar{M}_n = \frac{1}{n}\sum_{i=1}^{n} M_i \to \mu_M \quad \text{as } n \to \infty.$$

To show how the method of moments determines an estimator, we first consider the case of one parameter. We start with independent random variables $X_1, X_2, \ldots$ chosen according to the probability density $f_X(x|\theta)$ associated to an unknown parameter value θ. The common mean of the X_i, μ_X, is a function $k(\theta)$ of θ. For example, if the X_i are continuous random variables, then

$$\mu_X = \int_{-\infty}^{\infty} x f_X(x|\theta)\, dx = k(\theta).$$

The law of large numbers states that

$$\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \to \mu_X \quad \text{as } n \to \infty.$$

Thus, if the number of observations n is large, the distributional mean, $\mu = k(\theta)$, should be well approximated by the sample mean, i.e.,

$$\bar{X} \approx k(\theta).$$

This can be turned into an estimator $\hat{\theta}$ by setting

$$\bar{X} = k(\hat{\theta}).$$

and solving for $\hat{\theta}$.

We shall