Part I. Methods to Derive Point Estimators

Example 1. Let $X_1, X_2, \dots, X_n$ be a random sample from (μ, σ^2) .

Please find a good point estimator for $1. \mu$ $2. \sigma^2$

Solutions.

$1. \hat{\mu} = \bar{X}$

$2. \widehat{\sigma^2} = S^2$

There are the typical estimators for μ and σ^2. Both are unbiased estimators.

Property of Point Estimators

Unbiased Estimators. $\hat{\theta}$ is said to be an unbiased estimator for θ if $E(\hat{\theta}) = \theta$.

$$E(\bar{X}) = E\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right)$$

$$= \frac{E(X_1) + E(X_2) + \cdots + E(X_n)}{n} = \frac{\mu + \mu + \cdots + \mu}{n}$$

$$= \mu$$

$E(S^2) = \sigma^2$ (*We have shown this to be true in class earlier.)

Unbiased estimator may not be unique.

Example 2. $E(\sum a_i X_i) = (\sum a_i)\mu$

$\tilde{\mu} = \frac{\sum_{i=1}^n a_i X_i}{\sum_{i=1}^n a_i}$ $\quad since\ E(\tilde{\mu}) = \mu$

Variance of the unbiased estimators

$$\left.\begin{array}{l} Var(\bar{X}) = \dfrac{\sigma^2}{n} \\ Var(X_i) = \sigma^2 \end{array}\right\} Var(\bar{X}) < Var(X_i)\ when\ n > 1$$

Methods for deriving point estimators

1. Maximum Likelihood Estimator (MLE)
2. Method Of Moment Estimator (MOME)

1. The Maximum Likelihood Estimators (MLE)

Approach: To estimate model parameters by maximizing the likelihood

By maximizing the likelihood, which is the joint probability density function of a random sample, the resulting point estimators found can be regarded as yielded by the most likely underlying distribution from which we have drawn the given sample.

Example 3. $X_i \sim N(\mu, \sigma^2)$, i.i.d., $i = 1,2,\ldots,n$; Derive the MLE for μ and σ^2.

Solution.

[i]

$$f(x_i) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}} = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right],$$

$$x_i \in R, i = 1, 2, \ldots, n$$

[ii] likelihood function= $L = f(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} f(x_i)$

$$= \prod_{i=1}^{n} \left\{ (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left[-\frac{(x_i-\mu)^2}{2\sigma^2}\right] \right\}$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left[-\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^2}\right]$$

[iii] log likelihood function

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

[iv]

$$\begin{cases} \dfrac{\partial l}{\partial \mu} = \dfrac{2\sum_{i=1}^{n}(x_i-\mu)}{2\sigma^2} = 0 \\ \dfrac{\partial l}{\partial \sigma^2} = -\dfrac{n}{2\sigma^2} + \dfrac{\sum_{i=1}^{n}(x_i-\mu)^2}{2\sigma^4} = 0 \end{cases} \Rightarrow \begin{cases} \hat{\mu} = \bar{X} \\ \widehat{\sigma^2} = \dfrac{\sum(X_i-\bar{X})^2}{n} \end{cases}$$

R. A. Fisher (http://en.wikipedia.org/wiki/Ronald_Fisher)

R A Fisher, with his sons George (18) and Harry (14), 1938

Mrs Fisher in 1938, with daughters, left to right, Elizabeth, Phyllis, Rose, June, Margaret, Joan

2. The Method of Moment Estimators (MOME)

Approach: To estimate model parameters by equating the population moments to the sample moments

Order	Population Moment	Sample Moment
1st	$E(X)$	$=\frac{X_1+X_2+\cdots+X_n}{n}$
2nd	$E(X^2)$	$=\frac{X_1{}^2+X_2{}^2+\cdots+X_n{}^2}{n}$
$\vdots$	$\vdots$	$\vdots$
kth	$E(X^k)$	$=\frac{X_1{}^k+X_2{}^k+\cdots+X_n{}^k}{n}$
$\vdots$	$\vdots$	$\vdots$

Example 3 (continued). $X_i \sim N(\mu, \sigma^2), \text{i.i.d.}, i = 1,2,\ldots,n$; Derive the MOME for μ and σ^2.

Solution.

$$E(X) = \mu = \bar{X}$$

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

$$\Rightarrow \begin{cases} \hat{\mu} = \bar{X} \\ \widehat{\sigma^2} = \frac{\sum_{i=1}^{n} X_i{}^2}{n} - (\bar{X})^2 \end{cases}$$

$$\frac{\sum_{i=1}^{n} X_i{}^2}{n} - (\bar{X})^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X} + \bar{X})^2}{n} - (\bar{X})^2$$

$$= \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2 + \sum_{i=1}^{n} 2\bar{X}(X_i - \bar{X}) + \sum_{i=1}^{n}(\bar{X})^2}{n} - (\bar{X})^2$$

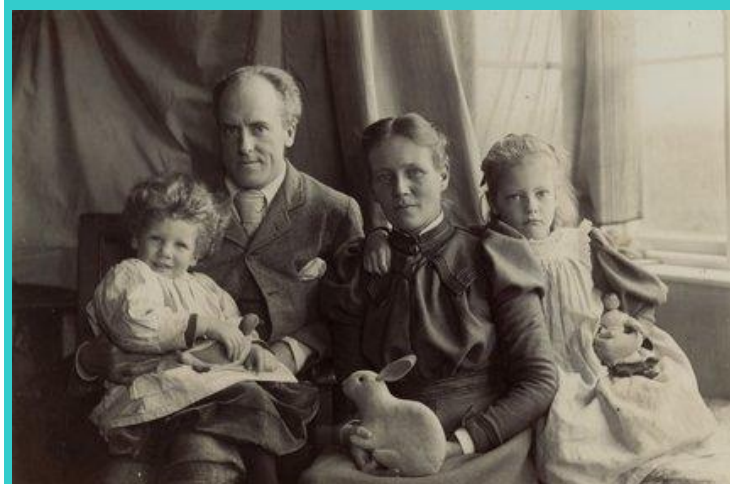
$$= \frac{\sum(X_i - \bar{X})^2}{n}$$

Therefore, the MLE and MOME for σ^2 are the same for the normal population.

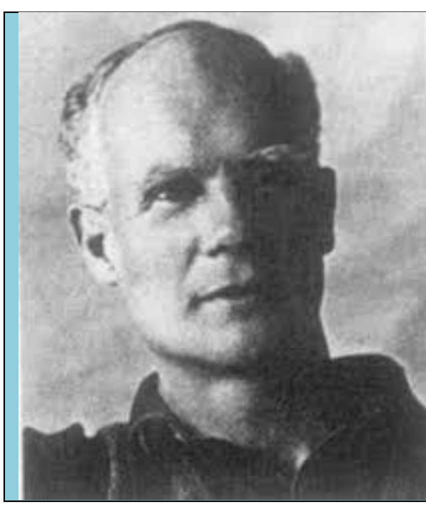
$$E\left(\widehat{\sigma^2}\right) = E\left[\frac{\sum(X_i-\bar{X})^2}{n}\right] = E\left[\frac{n-1}{n}\frac{\sum(X_i-\bar{X})^2}{n-1}\right] = \frac{n-1}{n}E(S^2) = \frac{n-1}{n}\sigma^2 \xrightarrow{n\to\infty} \sigma^2$$

(asymptotically unbiased)

Karl Pearson (http://en.wikipedia.org/wiki/Karl_pearson)



Left to Right: Egon Sharpe Pearson, Karl Pearson, Maria Pearson (née Sharpe), Sigrid Letitia Sharpe Pearson



Egon Sharpe Pearson, CBE FRS (11 August 1895 – 12 June 1980) was one of three children (Sigrid, Helga, and Egon) and the son of Karl Pearson and, like his father, a leading British statistician.

Comments on method of moments:

(1) Instead of using the first *d* moments, we could use higher order moments (or other functions of the data, for example, correlations) instead, leading to different estimating equations. But the method of moments estimator may be altered by which moments we choose.

Example 4: $X_1, \ldots, X_n$ iid Poisson(λ). The first moment is

$\mu_1(\lambda) = E_\lambda(X) = \lambda$. Thus, the method of moments estimator based on the first moment is $\hat{\lambda} = \bar{X}$.

We could also consider using the second moment to form a method of moments estimator.

$$\mu_2(\lambda) = E_\lambda(X^2) = \lambda + \lambda^2.$$

The method of moments estimator based on the second moment solves

$$\frac{1}{n}\sum_{i=1}^n X_i^2 = \hat{\lambda} + \hat{\lambda}^2$$

Solving this equation (by taking the positive root), we find that

$$\hat{\lambda} = -\frac{1}{2} + \left[\frac{1}{4} + \frac{1}{n}\sum_{i=1}^n X_i^2\right]^{1/2}.$$

The two method of moments estimators are different.

For example, for the data

> rpois(10,1)

[1] 2 3 0 1 2 1 3 1 2 1,

the method of moments estimator based on the first moment is 1.1 and the method of moments estimator based on the second moment is 1.096872.

(2) The method of moments does not use all the information that is available.

Example 5. $X_1, \ldots, X_n$ iid Uniform $(0, \theta)$.

The method of moments estimator based on the first moment is $\hat{\theta} = 2\bar{X}$. If $2\bar{X} < \max X_i$, we know that $\theta \geq \max X_i > \hat{\theta}$

<u>Now, we will see more examples for the MLE and MOME:</u>

Example 6. Let $X_i \sim Bernoulli(p)$, i.i.d. $i = 1, \ldots, n$.

. Please derive

1. The MLE of p

2. The MOME of p.

Solution.

1. MLE

   [i]

   $$f(x_i) = p^{x_i}(1-p)^{1-x_i}, i = 1, \ldots, n$$

   [ii]

   $$L = f(x_1, x_2, \ldots, x_n) = \prod f(x_i) = p^{\sum x_i}(1-p)^{\, n - \sum x_i}$$

   [iii]

   $$l = \ln L = \left(\sum x_i\right) \ln p + \left(n - \sum x_i\right) \ln(1-p)$$

   [iv]

   $$\frac{dl}{dp} = \frac{\sum x_i}{p} - \frac{n - \sum x_i}{1-p} = 0 \; \Rightarrow \hat{p} = \frac{\sum x_i}{n}$$

2. MOME

   $$E(X) = p = \frac{X_1 + X_2 + \cdots + X_n}{n} \Rightarrow \hat{p} = \frac{\sum x_i}{n}$$

Example 7. Let $X_1, X_2, \dots, X_n$ be a random sample from exp(λ)

Please derive

1. The MLE of λ

2. The MOME of λ.

Solution:

1. MLE:

$$f(x) = \lambda \exp(-\lambda x)$$

$$L = f(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \lambda \exp(-\lambda x_i)$$

$$= \lambda^n \exp(-\lambda \sum_{i=1}^{n} x_i)$$

$$l = lnL = nln(\lambda) - \lambda \sum_{i=1}^{n} x_i$$

$$\frac{\partial l}{\partial \lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0$$

Thus

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

2. MOME:

$$E(X) = \frac{1}{\lambda}$$

Thus setting:

$$E(X) = \frac{1}{\lambda} = \bar{X}$$

We have:

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

Example 8 (Post Office)

A post office has two clerks, Lucy and Ricky. It is known that their service times are two independent exponential random variables with the same parameter λ, and it is known that Lucy spends on the average 10 minutes with each customer. Suppose Lucy and Ricky are each serving one customer at the moment,

(a) What is the distribution of the waiting time for the next customer in line?
(b) What is the probability that the next customer in line will be the last customer (among the two customers being served and him/herself) to leave the post office?

Solution:

(a) Let X_1 and X_2 denote the service time for Lucy and Ricky, respectively. Then the waiting time for the next customer in line is:
$Y=\min(X_1,X_2)$, where $X_i \sim$ iid $\exp(\lambda)$, i=1,2.

Furthermore, $E(X_1)=1/\lambda=10$, $\lambda=1/10$. From the definition of exponential R.V.,

$f_{X_i}(x_i) = \lambda e^{-\lambda x_i}$, $x_i>0$, i=1,2 and $P(X_i>x)=\int_x^\infty f_{x_i}(u)du = e^{-\lambda x_i}$, $x_i>0$, i=1,2.

$F_Y(y)=P(Y\le y)=1-P(Y>y)=1-P(X_1>y,X_2>y)=1-P(X_1>y)\cdot P(X_2>y)=1-(e^{-\lambda y})^2=1-e^{-2\lambda y}$

Therefore $f_Y(y)=\frac{d}{dy}F_Y(y)=(2\lambda)\, e^{-(2\lambda)y}$, $y>0$. Thus $Y \sim \exp(2\lambda)$, where $\lambda=1/10$.

Note: Because of the memoryless property of the exponential distribution, it does not matter how long the current customers has been served by Lucy or Ricky.

(b) One of the two customers being served right now will leave first. Now between the customer who is still being served and the next customer in line, their service time would follow the same exponential distribution because of the memoryless property of the exponential random variable. Therefore, each would have the same chance to finish first or last. That is, the next customer in line will be the last customer to leave with probability 0.5.

Part II. Mean Squared Error (M.S.E.)

How to evaluate an estimator?

For unbiased estimators, all we need to do is to compare their variances, the smaller the variance, the better is the estimator. Now, what if the estimators are not all unbiased? How do we compare them?

Definition: Mean Squared Error (MSE)

Let $T = t(X_1, X_2, \ldots, X_n)$ be an estimator of $\tau(\theta)$, then the M.S.E. of the estimator T is defined as :

$\mathrm{MSE}_t(\tau(\theta)) = \mathrm{E}\left[(\mathrm{T} - \tau(\theta))^2\right]$: average squared distance from T to $\tau(\theta)$

$=\mathrm{E}\left[(\mathrm{T} - \mathrm{E}(\mathrm{T}) + \mathrm{E}(\mathrm{T}) - \tau(\theta))^2\right]$

$=\mathrm{E}\left[(\mathrm{T} - \mathrm{E}(\mathrm{T}))^2\right] + \mathrm{E}\left[(\mathrm{E}(\mathrm{T}) - \tau(\theta))^2\right] + 2\mathrm{E}\left[(\mathrm{T} - \mathrm{E}(\mathrm{T}))(\mathrm{E}(\mathrm{T}) - \tau(\theta))\right]$

$=\mathrm{E}\left[(\mathrm{T} - \mathrm{E}(\mathrm{T}))^2\right] + \mathrm{E}\left[(\mathrm{E}(\mathrm{T}) - \tau(\theta))^2\right] + 0$

$=\mathrm{Var}(\mathrm{T}) + (\mathrm{E}(\mathrm{T}) - \tau(\theta))^2$

Here $|\mathrm{E}(\mathrm{T}) - \tau(\theta)|$ is "the bias of T"

If unbiased, $(\mathrm{E}(\mathrm{T}) - \tau(\theta))^2 = 0$.

The estimator has smaller mean-squared error is better.

Example 1. Let $X_1, X_2, \ldots, X_n \overset{i.i.d}{\sim} N(\mu, \sigma^2)$

M.L.E. for μ is $\hat{\mu} = \bar{X}$; M.L.E. for σ^2 is $\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{n}$

1. M.S.E. of $\hat{\sigma}^2$?

2. M.S.E. of the sample variance S^2 as an estimator of σ^2

Solution.

1.

$\mathrm{MSE}_{\hat{\sigma}^2}(\sigma^2) = E\left[\left(\hat{\sigma}^2 - \sigma^2\right)^2\right] = \mathrm{Var}\left(\hat{\sigma}^2\right) + \left(E\left(\hat{\sigma}^2\right) - \sigma^2\right)^2$

To obtain $\mathrm{Var}\left(\hat{\sigma}^2\right)$, there are 2 approaches.

a. By the first definition of the Chi-square distribution.

Note $X_i \sim N(\mu, \sigma^2)$; $W = \frac{\sum_{i=1}^{n}(X_i - \bar{X})^2}{\sigma^2} \sim {\chi_{n-1}}^2$, Gamma$\left(\lambda = \frac{1}{2}, S = \frac{n-1}{2}\right)$

$$E(W) = \frac{S}{\lambda} = n - 1; \mathrm{Var}(W) = \frac{S}{\lambda^2} = 2(n-1)$$

$$\mathrm{Var}\left(\hat{\sigma}^2\right) = \mathrm{Var}\left(\frac{W}{n}\sigma^2\right) = \frac{\sigma^4}{n^2}\mathrm{Var}(W) = \frac{\sigma^4}{n^2} 2(n-1)$$

b. By the second definition of the Chi-square distribution.

For $Z \sim N(0,1)$, $W = \sum_{i=1}^{n} {Z_i}^2$

$$\mathrm{Var}\left(Z^2\right) = E\left[\left(Z^2 - E\left(Z^2\right)\right)^2\right]$$
$$= E\left[\left(Z^2 - \left(Var(Z) + [E(Z)]^2\right)\right)^2\right]$$
$$= E[(Z^2 - 1)^2]$$

Since $E(Z^2) = \mathrm{Var}(Z) + [E(Z)]^2 = 1$ from $Z \sim N(0,1)$,

$$E[(Z^2 - 1)^2 = E\left[Z^4 - 2Z^2 + 1\right]$$
$$= E\left(Z^4\right) - 2E(Z^2) + 1 = E(Z^4) - 1$$

Calculate the 4th moment of $Z \sim N(0,1)$ using the moment generating function (mgf) of Z;

$M_Z(t) = E(e^{tZ}) = e^{t^2/2}$

We know that the kth population moments can be obtained through the relations of:

$$E(Z^k) = M^{(k)}{}_Z(t=0)$$

https://en.wikipedia.org/wiki/Moment-generating_function

For the given problem, we have:

$M'_Z(t) = te^{t^2/2}$

$M''_Z(t) = te^{t^2/2} + t^2e^{t^2/2}$

$M^{(3)}{}_Z(t) = 3te^{t^2/2} + t^2e^{t^2/2}$

$M^{(4)}{}_Z(t) = 3e^{t^2/2} + 6t^2e^{t^2/2} + t^4e^{t^2/2}$

Set $t=0$, $M^{(4)}{}_Z(0) = 3 = E(Z^4)$

$$\mathrm{Var}(Z^2) = 3 - 1 = 2$$

$$\mathrm{Var}(W) = \sum_{i=1}^{n-1} \mathrm{Var}(Z_i{}^2) = 2(n-1)$$

$$\hat{\sigma}^2 = \frac{\sigma^2}{n}W, \mathrm{Var}(\hat{\sigma}^2) = \frac{\sigma^4}{n^2}2(n-1)$$

$$\mathrm{MSE}_{\hat{\sigma}^2}(\sigma^2) = \mathrm{Var}(\hat{\sigma}^2) + [E(\hat{\sigma}^2) - \sigma^2]^2$$

$$= \frac{2(n-1)}{n^2}\sigma^4 + [E\left(\frac{n-1}{n}S^2\right) - \sigma^2]^2$$

$$= \frac{2(n-1)}{n^2}\sigma^4 + \left[\frac{n-1}{n}\sigma^2 - \sigma^2\right]^2 \text{ (we know } E(S^2) = \sigma^2\text{)} = \frac{2n-1}{n^2}\sigma^4$$

The M.S.E. of $\hat{\sigma}^2$ is $\frac{2n-1}{n^2}\sigma^4$

We know S^2 is an unbiased estimator of σ^2

$$E[(S^2 - \sigma^2)^2] = \mathrm{Var}(S^2) + 0 = \mathrm{Var}\left(\frac{\sigma^2 W}{n-1}\right)$$

$$= \left(\frac{\sigma^2}{n-1}\right)^2 \mathrm{var}(W) = \frac{2\sigma^4}{n-1}$$

Exercise:

For a random sample from a normal distribution, please compare the MSE of $\widehat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n}$ and $\widehat{\sigma}^2 = S^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n-1}$.

Which one is a better estimator (in terms of the MSE)?

Solution.

$$MSE_{\widehat{\sigma}^2}(\sigma^2) = E\left[\left(\widehat{\sigma}^2 - \sigma^2\right)^2\right] = Var\left(\widehat{\sigma}^2\right) + (E\left(\widehat{\sigma}^2\right) - \sigma^2)^2$$

Then, we have M.S.E. of $\widehat{\sigma}^2$ is $\frac{2n-1}{n^2}\sigma^4$.

$$MSE_{S^2}(\sigma^2) = E[(S^2 - \sigma^2)^2] = Var(S^2) + (E(S^2) - \sigma^2)^2$$

$$MSE_{S^2}(\sigma^2) = Var(S^2) + 0 = Var\left(\frac{\sigma^2 W}{n-1}\right) =$$
$$\left(\frac{\sigma^2}{n-1}\right)^2 var(W) = \frac{2\sigma^4}{n-1}$$

$$\because\ \frac{2n-1}{n^2}\sigma^4 - \frac{2\sigma^4}{n-1} = \frac{1-3n}{n^2(n-1)}\sigma^4 < 0$$

$$\therefore\ MSE_{\widehat{\sigma}^2}(\sigma^2) = \frac{2n-1}{n^2}\sigma^4 < \frac{2\sigma^4}{n-1} = MSE_{S^2}(\sigma^2)$$

$$\therefore\ \widehat{\sigma}^2 = \frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{n} \text{ is the better estimator}$$

An extracurricular note:

Finally, here is the link to the British Doctor's Study where they established the link between smoking and lung cancer:

https://en.wikipedia.org/wiki/British_Doctors_Study

https://breathe.ersjournals.com/content/12/3/275?ctkey=shareline

There was an American study carried out in a similar time that helped determine this causal relationship as well:

https://pubmed.ncbi.nlm.nih.gov/15415260/