# [Chap 9] Point Estimation.

## [9.1] Intro:

Read 9.1!

$\theta$ = a parameter (could be a vector)

$\Omega$ = parameter space

Large X

DEF: A statistic $T = t(X_1, \dots, X_n)$ that is used to estimate the value of $\tau(\theta)$ is called an <u>estimator</u> of $\tau(\theta)$ and an observed value of the stat $t = t(\underbrace{x_1, \dots, x_n}_{\text{small } x\text{'s}})$ is called an <u>estimate</u> of $\tau(\theta)$.

$T$ = statistic

$t$ = observed value of $T$

$t$ = function applied to RS

$\left.\begin{array}{l}\hat{\theta} = \text{same idea as } T \\ \tilde{\theta} = \text{same idea as } T\end{array}\right\}$ estimators of $\theta$

# [9.2] Some Methods of Estimation.

## Method of Moments

Use sample moments to estimate population moments

Sample moments: $M_j' = \frac{1}{n}\sum_{k=1}^{n} X_k^j$ (jth sample moments)

Population moments: $\mu_j' = E(X^j)$ (jth population moments)

When you have K parameters to estimate, find a system of equations and solve for the parameters (plug in "hats" to all parameters on the left):

$$\left.\begin{aligned}
\mu_1(\hat\theta_1, \hat\theta_2, \ldots, \hat\theta_n) &= M_1' \\
\mu_2(\hat\theta_1, \hat\theta_2, \ldots, \hat\theta_n) &= M_2' \\
&\vdots \\
\underbrace{\mu_K(\hat\theta_1, \hat\theta_2, \ldots, \hat\theta_n)}_{\text{Population moments with hats}} &= M_K'
\end{aligned}\right\}$$

Now solve the system of equations!

sample moments

k is the number of parameters in the distribution you are estimating for example, suppose you have 2 parameters to estimate.

The MMEs (Method of Moments Estimator) can be found

Two parameters to estimate, so use two sample moments

Note, we have

$$\left.\begin{aligned} \mu_1'(\hat{\theta}_1, \hat{\theta}_2) &= \frac{1}{n}\sum x_i = \bar{x} \\ \mu_2'(\hat{\theta}_1, \hat{\theta}_2) &= \frac{1}{n}\sum x_i^2 \end{aligned}\right\}$$

Then solve the system of equations.

Ex: Suppose $x_i$ is a RS from $f(x; \mu, \sigma^2)$

$E(X) = \mu$

$V(X) = \sigma^2$

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

let $\theta_1 = \mu$ and $\theta_2 = \sigma^2$

So $\hat{\theta}_1 = \hat{\mu} = \frac{1}{n}\sum x_i = \bar{x} \Rightarrow$ So $\boxed{\hat{\mu} = \bar{x}}$

$$\hat{\theta}_2 = \hat{\sigma}^2 = \frac{1}{n}\sum x_i^2 + \hat{\mu}^2$$

$$\boxed{\hat{\sigma}^2} = \frac{1}{n}\sum x_i^2 + \frac{1}{n^2}(\sum x_i)^2 = \boxed{\frac{1}{n}\sum_i^n (x_i - \bar{x})^2}$$

We can also write these in terms of $s^2$ (the sample variance)

This is because

$$s^2 = \frac{1}{n-1}\sum(x_i-\bar{x})^2 = \frac{1}{n-1}\sum(x_i^2 - 2\bar{x}x_i + \bar{x}^2)$$

$$(n-1)s^2 = \sum x_i^2 - 2\bar{x}\underbrace{\sum x_i}_{n\bar{x}} + n\bar{x}^2$$

$$\frac{(n-1)s^2}{n} = \frac{1}{n}\sum x_i^2 - \frac{n}{n}\bar{x}^2$$

So the second sample moment can be written as

$$\boxed{M_2' = \frac{1}{n}\sum x_i^2 = \frac{n-1}{n}s^2 + \bar{x}^2}$$

So $\boxed{\begin{array}{l}\hat{\mu} = \bar{x} \\ \hat{\sigma}^2 = \frac{n-1}{n}s^2\end{array}}$

From the previous slide, we can therefore write

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

In terms of sample moments, we have

$$\mu_1'(\hat{\theta}_1, \hat{\theta}_2) = M_1' = \frac{1}{n}\sum X_i = \bar{X}$$

and $$\mu_2'(\hat{\theta}_1, \hat{\theta}_2) = M_2' = \frac{1}{n}\sum X_i^2 = \frac{n-1}{n}S^2 + \bar{X}^2 \qquad (*)$$

This is equivalent to

$$\widehat{E(X)}(\hat{\theta}_1, \hat{\theta}_2) = \bar{X}$$

and $$\widehat{Var(X)}(\hat{\theta}_1, \hat{\theta}_2) = \frac{n-1}{n}S^2$$

either way results the same.

Now some examples!

Two parameter exponential

EX: So we have a RS from $X_i \sim EXP(1, \eta)$

The $E(X_i) = \eta + 1$

only one parameter to find so: $\hat{\eta} + 1 = M_1' = \bar{X}$

$$\boxed{\hat{\eta} = \bar{X} - 1}$$

Ex: RS from $X_i \sim GAM(\theta, \kappa)$

Remember

$E(X^2) = Var(X) + \mu^2$

$\mu_1' = E(X_i) = \kappa\theta$

$\hat{\mu}_2 = \sigma^2 + \mu^2 = \kappa\theta^2 + \kappa^2\theta^2$

Two parameters to estimate, so use two sample moments

So $\hat{\mu}_1' = \hat{\kappa}\hat{\theta} = \frac{1}{n}\sum X_i = \bar{X}$ $\quad (\hat{\mu}_1' = M_1')$

$\hat{\mu}_2' = (\hat{\kappa} + \hat{\kappa}^2)\hat{\theta}^2 = M_2' = \frac{1}{n}\sum X_i^2$

First equation gives $\boxed{\hat{\theta} = \frac{\bar{X}}{\hat{\kappa}}}$

Now solve the second equation for $\hat{\kappa}$ (plug in $\hat{\theta}$ to it)

So $\frac{1}{n}\sum X_i^2 = (\hat{\kappa} + \hat{\kappa}^2)\hat{\theta}^2 = (\hat{\kappa} + \hat{\kappa}^2)\left(\frac{\bar{X}}{\hat{\kappa}}\right)^2 = \left(\frac{1}{\hat{\kappa}} + 1\right)\bar{X}^2$

$$\frac{1}{n\bar{X}^2}\sum X_i^2 = \frac{1}{\hat{\kappa}} + 1$$

$$\underbrace{\frac{1}{n\bar{x}^2}\sum x_i^2 - 1} = \frac{1}{\hat{K}}$$

$$\frac{\sum x_i^2 - n\bar{x}^2}{n\bar{x}^2} = \frac{1}{\hat{K}}$$

So $\boxed{\hat{K} = \frac{n\bar{x}^2}{\sum x_i^2 - n\bar{x}^2} \quad \text{and} \quad \hat{\theta} = \frac{\bar{X}}{\hat{K}}}$

EX: Redo with $\widehat{Var}$ so $X_i \sim GAM(\theta, K)$

and $E(X) = K\theta$ $\quad\quad \hat{K}\hat{\theta} = \bar{X} \Rightarrow \hat{\theta} = \frac{\bar{X}}{\hat{K}}$

$Var(X) = K\theta^2 \Rightarrow \quad \hat{K}\hat{\theta}^2 = \frac{n-1}{n}s^2$

So $\hat{k}\hat{\theta}^2 = \hat{k}\left(\frac{\bar{X}}{\hat{k}}\right)^2 = \frac{n-1}{n} S^2$

$$\Rightarrow \frac{\bar{X}^2}{\hat{k}} = \frac{(n-1)S^2}{n} \Rightarrow \boxed{\hat{k} = \frac{n\bar{X}^2}{(n-1)S^2}}$$

So

$$\hat{\theta} = \frac{\bar{X}}{\hat{k}} = \bar{X}\Big/\left(\frac{(n-1)S^2}{n\bar{X}^2}\right) = \boxed{\frac{(n-1)S^2}{n\bar{X}} = \hat{\theta}}$$

Which is better?

you decide!

Put a subscript on them to distinguish with MLEs - our next topic

$$\boxed{\begin{aligned} \hat{\theta}_{MME} &= \frac{(n-1)S^2}{n\bar{X}} \\ \hat{k}_{MME} &= \frac{n\bar{X}^2}{(n-1)S^2} \end{aligned}}$$

# Method of Maximum Likelihood

DEF: Likelihood function

$X_i \sim f(x_i;\theta)$ RS of $X_1, \cdots, X_n$

The joint dist of this Random sample is

$$f(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

The likelihood function is simply consider it a function of $\theta$, the parameters!

In particular, $L(\theta; x_1, \cdots, x_n) = f(x_1, x_2, \cdots, x_n; \theta)$

DEF: Maximum Likelihood Estimator

The value of $\theta$ which maximizes the likelihood function is an MLE (doesn't have to be unique). We call it $\hat{\theta}$.

$\hat{\theta}$ satisfies $f(x_1, \cdots, x_n; \hat{\theta}) = \max_{\theta \in \Omega} f(x_1, \cdots, x_n; \theta)$

Ex! Find the MLE for $\theta$, where we have a RS from POI($\theta$)

Note: Use calculus to find MLE & we maximize $L(\theta)$ (or $\ln L(\theta)$).

The pdf of POI($\theta$) is $f(x_i) = \dfrac{e^{-\theta}\theta^{x_i}}{x_i!}$

So the joint pdf of RS is

$$f(x_1, \dots, x_n; \theta) = \prod_{i=1}^{n} f(x_i;\theta) = \prod_{i=1}^{n} \frac{e^{-\theta}\theta^{x_i}}{(x_i)!} = \frac{e^{-n\theta}\theta^{\Sigma x_i}}{\Pi (x_i)!}$$

So $L(\theta) = \dfrac{e^{-n\theta}\theta^{\Sigma x_i}}{\prod_1^n x_i!}$. The log of this simplifies deriv.

$$\ln L(\theta) = -n\theta + \Sigma x_i \ln\theta - \ln \prod_1^n x_i!$$

Note: $\sum x_i = n\bar{x}$, so

$$\ln L(\theta) = -n\theta + n\bar{x}\ln\theta - \underbrace{\ln \prod (x_i)!}_{\text{constant wrt } \theta}$$

$$\frac{d \ln L(\theta)}{d\theta} = -n + n\bar{x}\frac{1}{\theta} + 0 \overset{set}{=} 0$$

$$\frac{n\bar{x}}{\theta} = n \Rightarrow \boxed{\hat{\theta} = \bar{x}}$$

<u>Thm 9.2.1</u> Invariance Property

If $\hat{\theta}$ is the MLE of $\theta$, and if $u(\theta)$ is a function of $\theta$, then $u(\hat{\theta})$ is <u>an</u> MLE of $u(\theta)$.

It might not be unique. (MLEs don't have to be unique)

Ex (9.22) RS from two-param Exponential.

So $X_i \sim EXP(1, \eta)$

So $f(x_i) = \frac{1}{1} e^{-\frac{(x_i-\eta)}{1}} I_{(\eta,\infty)}(x_i) = e^{-(x_i-\eta)} I_{(\eta,\infty)}(x_i)$

So $L(\eta) = \prod_1^n f(x_i) = \prod_1^n e^{-(x_i-\eta)} I_{(\eta,\infty)}(x_i) = e^{-\Sigma(x_i-\eta)} \prod I_{(\eta,\infty)}(x_i)$

simplify the exponent:

$$-\sum_1^n (x_i - \eta) = -\left[\Sigma x_i - n\eta\right] = -\left[n\bar{x} - n\eta\right] = -n(\bar{x}-\eta)$$

So $L(\eta) = e^{-n(\bar{x}-\eta)} \prod_{i=1}^n I_{(\eta,\infty)}(x_i)$

If we order the data, then we have the order stats

We can combine all of the Indicator functions together if we switch over to the order statistics:

So

$$\eta < X_{1:n} < X_{2:n} < \cdots < X_{n:n} < \infty$$

Hence,

$$\underbrace{-\infty < \eta < X_{1:n}}_{I_{(-\infty, X_{1:n})}(\eta)} < \infty$$

So

$$L(\eta) = e^{-n(\bar{x}-\eta)} I_{(-\infty, X_{1:n})}(\eta)$$

Note taking the deriv of the exponential part is

$L'(\eta) = n e^{-(\bar{x}-\eta)} > 0$, so $L(\eta)$ is always increasing

However, the indicator function is the clue to maximizing the Likelihood.

So

$$L(\eta) > 0 \quad \text{for all } \eta \leq X_{1:n}$$

$$L(\eta) = 0 \quad \text{for all } \eta > X_{1:n}$$


The plot of the Likelihood function is

So it follows the MLE is the largest value of for which $L(\eta) \neq 0$.

$$\boxed{\hat{\eta}_{MLE} = X_{1:n}}$$

Compare that to the MME $\rightarrow$ $\boxed{\hat{\eta}_{MME} = \bar{X} - 1}$

Suppose the data is $X_1 = 1,\ X_2 = 2,\ X_3 = 3,\ X_4 = 3.1$

Then the MLE is $\boxed{\hat{\eta}_{MLE} = X_{1:n} = 1}$ and MME is $\boxed{\hat{\eta}_{MME} = \frac{1+2+3+3.1}{4} - 1 = 1.275}$

## MLEs for a vector of parameters

MLEs are the

If MLEs don't exist on a boundary, then the solutions of the simultaneous equations

$$\frac{\partial}{\partial \theta_j} \ln L(\theta_1, \cdots, \theta_k) = 0 \quad \text{for } j = 1, \cdots, k$$

## Invariance Prop of MLEs

If $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_k)$ denotes the MLE of $\theta = (\theta_1, \cdots, \theta_k)$ then the MLE of $\tau = (\tau_1(\theta), \cdots, \tau_r(\theta))$ is

$$\hat{\tau} = (\hat{\tau}_1, \hat{\tau}_2, \cdots, \hat{\tau}_r) = (\tau_1(\hat{\theta}), \cdots, \tau_r(\hat{\theta}))$$

for $1 \leq r \leq k$

Note: $\tau : \mathbb{R}^k \longrightarrow \mathbb{R}^r$

Note: Multiparameters often are not the same as the individual parameters are assumed to be Known.

EX: RS from $X_i \sim N(\mu, \sigma^2)$

Find the MLE of $\mu$ and $\theta = \sigma^2$

$$L(\theta, \mu) = \prod_{i=1}^{n} f(x_i) = \frac{1}{(\sqrt{2\pi\theta})^n} \exp\left\{-\frac{1}{2} \frac{\Sigma (x_i - \mu)^2}{\theta}\right\}$$

$$\ln L(\theta, \mu) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\theta - \frac{1}{2}\frac{\Sigma(x_i - \mu)^2}{\theta}$$

constants ⇒ deriv is zero.

const wrt $\mu$

So the partial derivatives are

$$\frac{\partial \ln L(\theta,\mu)}{\partial\theta} = 0 - \frac{n}{2\theta} + \frac{1}{2}\frac{\sum(x_i-\mu)^2}{\theta^2}$$

$$\frac{\partial \ln L(\theta,\mu)}{\partial\mu} = 0 + 0 + -\frac{1}{2\theta}\sum 2(x_i-\mu)(-1)$$

$$= \frac{1}{\theta}\sum(x_i-\mu)$$

$$0 = \frac{1}{\hat\theta}\sum(x_i-\hat\mu) = \sum x_i - n\hat\mu = 0$$

$$\sum x_i = n\hat\mu$$

$$\boxed{\hat\mu = \frac{1}{n}\sum x_i = \bar{x}}$$

Next,

$$-\frac{n}{2\hat\theta} + \frac{1}{2}\frac{\sum(x_i-\hat\mu)^2}{\hat\theta^2} = 0$$

$$\frac{1}{2}\frac{\sum(x_i-\hat\mu)^2}{\hat\theta^2} = \frac{n}{2\hat\theta}$$

$$\hat\theta = \frac{\sum(x_i-\hat\mu)^2}{n}$$

$$\hat\theta = \frac{\sum(x_i-\bar{x})^2}{n}$$

$$= \frac{n-1}{n}\frac{\sum(x_i-\bar{x})^2}{n-1}$$

$$\boxed{\hat\theta = \frac{n-1}{n}s^2}$$

Note the MLE is <u>not</u> $s^2$!

We use $s^2$ because $E(s^2) = \sigma^2$

Note that $E(\hat\theta) = \frac{n-1}{n}\sigma^2$ (<u>Not</u> unbiased)

Now suppose $\theta = \theta_0$ is known.

Then $L(\mu) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln\theta_0 - \frac{1}{2\theta_0}\sum(x_i-\mu)^2$

The MLE of $\mu$ is $\hat{\mu} = \bar{x}$.

Now suppose $\mu = \mu_0$ is known.

$L(\theta) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln\theta - \frac{1}{2\theta}\sum(x_i-\mu_0)^2$

The MLE of $\theta$ is $\hat{\theta} = \frac{\sum(x_i-\mu_0)^2}{n}$

Sometimes the mle is not easily solved for (either explicitly, you are lazy, or you want to avoid taking derivatives.)

In cases like this, you need to use iterative methods for solving for the mle. There are several options, but I will only illustrate and teach you one: The Newton Raphson Method.

**Newton-Raphson Method:**

A root finding algorithm which solves for the zeros (or roots) of a function $g(x) = 0$ (e.g. you can solve for $x$).

To use, iterate the formula, where $x_0$ is an initial guess for the root:

$$x_{n+1} = x_n - \frac{g(x_n)}{g'(x_n)}, \quad n = 0, 1, 2, \ldots$$

You **MUST** have a good guess or it won't converge!! The MME is a good choice for the guess for the MLE

**EX:** a random sample from $X_i \sim \text{GAM}(\theta, \kappa)$

The Likelihood function is the product of all of the marginals:

$$L(\theta,\kappa)=\prod_{i=1}^{n} f(x_i)=\frac{1}{\theta^{n\kappa}[\Gamma(k)]^n}\left(\prod_{i=1}^{n} x_i\right)^{\kappa-1}\exp\left[-\sum_{i=1}^{n}\frac{x_i}{\theta}\right]$$

The log likelihood function will be easier to work with:

$$\ln L(\theta,\kappa)=-n\kappa\ln\theta-n\ln\Gamma(k)+(k-1)\ln\left(\prod_{i=1}^{n} x_i\right)-\sum_{i=1}^{n}\frac{x_i}{\theta}$$

$$\ln L(\theta,\kappa)=-n\kappa\ln\theta-n\ln\Gamma(k)+(k-1)\sum_{i=1}^{n}\ln x_i-\sum_{i=1}^{n}\frac{x_i}{\theta}$$

Let $\overline{\ln X}=\frac{1}{n}\sum_{i=1}^{n}\ln x_i$. Similarly, $\bar{x}=\frac{1}{n}\sum_{i=1}^{n} x_i$

$$\ln L(\theta,\kappa)=-n\kappa\ln\theta-n\ln\Gamma(k)+n(k-1)\overline{\ln X}-\frac{n\bar{x}}{\theta}$$

So $\frac{\partial \ln L(\theta,k)}{\partial\theta} = -\frac{nk}{\theta}+\frac{n\bar{x}}{\theta^2} \overset{set}{=} 0$

$\overset{\text{find est}}{\Longrightarrow}$ $\frac{\hat{k}}{\hat{\theta}}=\frac{\bar{x}}{\hat{\theta}^2} \Rightarrow \boxed{\hat{\theta}=\frac{\bar{x}}{\hat{k}}}$

and $\frac{\partial \ln L(\theta,k)}{\partial k} = -n\ln\theta - n\underbrace{\frac{d\ln\Gamma(k)}{dk}}_{\Psi(k)} + n\overline{\ln x} \overset{set}{=} 0$

the digamma function

(in R → digamma(k))

find est $\Rightarrow$ $-n\ln\frac{\bar{x}}{\hat{k}} - n\Psi(\hat{k}) + n\overline{\ln x} = 0$

$= -n\ln\bar{x} + n\ln\hat{k} - n\Psi(\hat{k}) + n\overline{\ln x} = 0$

$(*) \Rightarrow \ln\hat{k} - \Psi(\hat{k}) + \overline{\ln x} - \ln\bar{x} = 0$

To solve $(*)$, a numerical technique must be used (Like Newton's Method!)

So define: $g(\hat{\kappa}) = \ln \hat{\kappa} - \varphi(\hat{\kappa}) + \overline{\ln X} - \ln \overline{x}$

then: $g'(\hat{\kappa}) = \frac{1}{\hat{\kappa}} - \varphi'(\hat{\kappa})$

So we iterate the following formula:

$$k_{n+1} = k_n - \frac{g(k_n)}{g'(k_n)} \quad \text{which gives } \hat{\kappa} = \lim_{n\to\infty} k_n$$

It also converges very fast - it has quadratic convergence, which means that when it is close enough, the number of digits that are correct DOUBLE each iteration

First load the mle function into R. As an example, next generate random data from a gamma distribution

Then run mle(x) to find the MLE. It will return MLEs to both the scale and shape parameters

R Code:

```
mle = function(x) {
  n=length(x)
  eps=0.000000001
  c = mean(log(x)) - log(mean(x))
  g = function(k) = { log(k) - digamma(k) + c }
  gp = function(k) = { 1/k - trigamma(k) }
  # Use the MME for kappa as the initial guess:
  k = n/(n-1)*mean(x)^2/var(x)
  diff = 2 #any number larger than eps
  while ( abs(diff) > eps ) {
    diff = g(k)/gp(k)
    k = k - diff
  }
  c(scale=mean(x)/k,shape=k)
}

## Now an example of how to use the above function:
# Generate some random data from a GAM(9,5) dist:
x = rgamma(1000,shape=5,scale=9)

# This should find the mles to the shape and scale parameters:
# The scale parameter should be close to 9
# and the shape close to 5
mle(x)
# Run this to see what it gives you!  You should be able to copy
and paste this text into any editor.
```