Chapter 9

9.1 Intro:

Read! θ = parameter (could be a vector)

Ω = parameter space = set of all possible values for θ.

DEF: A statistic $T = t(\underbrace{X_1, X_2, \cdots, X_n}_{\text{Large } X\text{'s}})$ that is used to estimate the value of $\tau(\theta)$ is called an estimator of $\tau(\theta)$ and an observed value of the statistic

$$t = t(\underbrace{x_1, x_2, \cdots, x_n}_{\text{small } x\text{'s}})$$ is

called an estimate of $\tau(\theta)$

T = statistic

t = observed value of T

t = function applied to the RS.

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

9.2 Some methods of Estimation

Ex: Use $\bar{X} = \frac{1}{n}\sum X_i$ as an estimator for μ_1'

Recall $\mu_j' = E(X^j)$ (j^{th} moment about the orgin)

Typically, the μ_j''s depend on parameters

$$\Rightarrow \mu_j'(\theta_1, \cdots, \theta_k) = E(X^j) \quad j = 1, \ldots, k.$$

The method of moments uses this approach of using the sample moments to approximate the population moments (μ_k')

DEF: If $X_1, \cdots, X_n$ is a RS from pdf $f(x; \theta_1, \cdots, \theta_k)$, the first k sample moments are given by

$$M_j' = \frac{1}{n}\sum_{i=1}^{n} X_i^j, \quad j = 1, 2, \cdots, k$$

Ex: The first sample moment is $M_1' = \frac{1}{n}\sum X_i = \bar{X}$

and " " population " " $\mu_1' = E(X) = \mu$

The method of moments principle is to choose as estimators of the parameters $\theta_1, \cdots, \theta_k$ the values $\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_k$ that render the pop. moments equal to the sample moments.

i.e. $\hat{\theta}_1, \cdots, \hat{\theta}_k$ are solutions of the equations:

$$M_j' = \mu_j'(\hat{\theta}_1, \cdots, \hat{\theta}_k) \quad j = 1, \cdots, k$$

<u>EX:</u> X_i RS from $f(x; \mu, \sigma^2)$ $\quad E(X) = \mu$, $V(X) = \sigma^2$

Note: $\mu_1' = E(X) = \mu$

$$\sigma^2 = E(X^2) - \mu^2 = \mu_2' - (\mu_1')^2$$

So the MME (Method of Moments estimators) are solutions of

$$M_1' = \hat{\mu}$$

$$M_2' = \hat{\sigma}^2 + (\hat{\mu})^2$$

So $\boxed{\hat{\mu} = \frac{1}{n} \sum X_i = \bar{X}}$

$$\hat{\sigma}^2 = M_2' - \hat{\mu}^2 = \frac{1}{n} \sum X_i^2 - \bar{X}^2$$

$$\boxed{\hat{\sigma}^2 = \frac{1}{n} \sum_1^n (X_i - \bar{X})^2}$$

Ex: $X_i \sim EXP(1, \eta)$

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

So the MME would satisfy:

$$M_1' = \hat{\eta} + 1$$

$$\hat{\eta} = M_1' - 1 = \frac{1}{n}\sum x_i - 1$$

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

Ex: $X_i \sim GAM(\theta, K)$

$$\mu_1' = K\theta$$

$$\mu_2' = V(X) + \mu^2 = K\theta^2 + K^2\theta^2 = (K + K^2)\theta^2$$

So

$$\hat{\mu}_1' = \hat{K}\hat{\theta} = \frac{1}{n}\sum x_i = \bar{x} = M_1'$$

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

So $\hat{\theta} = \frac{\bar{X}}{\hat{K}} \Rightarrow \frac{1}{n}\sum x_i^2 = (\hat{K} + \hat{K}^2)\left[\frac{\bar{X}}{\hat{K}}\right]^2$

$$= \left(\frac{1}{\hat{K}} + 1\right)\bar{X}^2$$

$$\frac{1}{n}\sum x_i^2 = (\hat{k}+\hat{k}^2)\left[\frac{\bar{x}}{\hat{k}}\right]^2 = \left(\frac{1}{\hat{k}}+1\right)\bar{x}^2$$

$$\frac{1}{n\bar{x}^2}\sum x_i^2 = \frac{1}{\hat{k}}+1$$

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

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

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

Method of Maximum Likelihood:

DEF: Likelihood function

joint pdf $= f(x_1, x_2, \ldots, x_n; \theta)$

$\Rightarrow$ Likelihood function is: $L(\theta; x_1, \ldots, x_n) = f(x_1, \ldots, x_n; \theta)$

If $x_1, \ldots, x_n$ represents a RS from $f(x_i;\theta)$

then
$$L(\theta) = \prod_{i=1}^{n} f(x_i;\theta)$$

DEF: Maximum Likelihood Estimator

Let $L(\theta)$ = likelihood function, $\theta \in \Omega$ ← parameter space.

For a given set of observations $(X_1, X_2, \ldots, X_n)$ a value $\hat{\theta}$ in Ω at which $L(\theta)$ is a maximum is called the Maximum Likelihood estimate (MLE) of θ. That is $\hat{\theta}$ is a value of $\hat{\theta}$ that satisfies

$$f(X_1, \ldots, X_n; \hat{\theta}) = \max_{\theta \in \Omega} f(X_1, X_2, \ldots, X_n; \theta)$$

Note: It doesn't have to be unique!

Ex: Find the MLE for θ, where we have a RS from POI(θ).

Note: Use calculus to find MLE.
(You can max $L(\theta)$ or $\ln L(\theta)$)

$$L(\theta) = \prod_1^n f(X_i; \theta) = \frac{e^{-n\theta}\,\theta^{\sum_1^n x_i}}{\prod_1^n (x_i)!}$$

$$\ln L(\theta) = -n\theta + \left(\sum_1^n x_i\right)\ln\theta - \ln \prod_1^n x_i$$

$$\frac{d \ln L(\theta)}{d\theta} = -n + \frac{\sum_1^n x_i}{\theta} \overset{\text{set}}{=} 0$$

$$\hat{\theta} = \frac{\sum_1^n x_i}{n} = \bar{X}$$

$\boxed{9.2}$ Cont.

$\boxed{\text{Thm 9.2.1}}$ Invariance Property

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

Ex: RS from two-param exponential

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

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

$$= e^{-(x_i - \eta)} I_{(\eta,\infty)}(x)$$

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

$$= e^{-\sum (x_i - \eta)} \prod_{i=1}^{\infty} I_{(\eta,\infty)}(x_i)$$

$$\sum (x_i - \eta) = n \frac{\sum x_i}{n} - \sum \eta = n\bar{x} - n\eta$$

$$= n(\bar{x} - \eta)$$

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

$\eta < x_1 < \infty$ and $\eta < x_2 < \infty$, and .... $\cdots \eta < x_n < \infty$

Note if we order the data, we have to use the order stats

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

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

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

Thus,

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

Note that $L(\eta) > 0$ for all $\eta < X_{1:n}$

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

Since $L'(\eta) = n e^{-n(x-\eta)} > 0$, then $L(\eta)$ is always increasing for $\eta < X_{1:n}$. Hence, the mle for η is the largest value for which $L(\eta) \neq 0$.

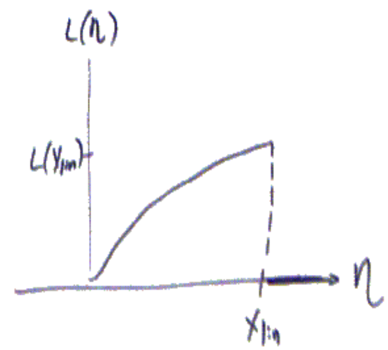
Thus, $\hat{\eta} = X_{1:n}$

Compare to MME $\hat{\eta}_{MME} = \bar{x} - 1$.

So if, suppose, the data is 1, 2, 3, 3.1

So the MME for $\hat{\eta} = \frac{1+2+3+3.1}{4} - 1 = 1.275$

and the MLE for $\hat{\eta} = 1$ (the smallest value)



The mle of a vector of parameters can also be found. For most cases (e.g. when Ω is cartesian product of k intervals). When Ω is in this form, and if the partial derivs of $L(\theta_1, \theta_2, \cdots, \theta_k)$ exist, and the MLEs don't occur on the boundary of Ω, then the MLEs will be the solutions of the simultaneous equations

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

Thm: Invariance Prop.

If $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_k)$ denotes the MLE of $\theta = (\theta_1, \ , \theta_k)$ then the MLE of $\tau = (\tau_1(\theta), \cdot \ , \tau_r(\theta))$ is $\hat{\tau} = (\hat{\tau}_1, \cdot, \hat{\tau}_r) = (\tau_1(\hat{\theta}), \tau_2(\hat{\theta}), \ \cdot, \tau_r(\hat{\theta}))$ for $1 \le r \le k$

Note: Multiparameter estimators often are not the same as the individual estimators when the other parameters are assumed to be known.

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

Find the MLEs of μ and $\theta = \sigma^2$

$$f(x_i; \mu, \theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sqrt{\theta}}\right)^2}$$

$$= \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{1}{2} \frac{(x_i - \mu)^2}{\theta}\right)$$

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

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

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

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

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

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

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

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

Next,

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

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

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

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

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

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

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 μ 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 θ is $\hat{\theta} = \frac{\sum(X_i-\mu_0)^2}{n}$

Sometimes the MLE is not easily solved for (explictly or you're just lazy).

In cases like these, numerical techniques can be used to solve for the MLEs. (e.g. Newton's method).

Ex: RS from $GAM(\theta, K)$

Sometime the mle is not easily solved for (explicitly or you're just lazy). In cases like these, numerical techniques can be used to solve for the mles.

EX: RS from $GAM(\theta, k)$

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

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

$$= -nk\ln\theta - n\ln\Gamma(k) + n(k-1)\underbrace{\frac{1}{n}\sum \ln(x_i)}_{\overline{\ln x}} - \frac{1}{\theta}\underbrace{\sum x_i}_{n\bar{x}}$$

$$= -nk\ln\theta - n\ln\Gamma(k) + n(k-1)\overline{\ln x} - \frac{n\bar{x}}{\theta}$$

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

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

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

the digamma function (in R → digamma(k))

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

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

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

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

Newton's Method

To solve $f(x)=0$, iterate $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$

You must have a good guess or it may not converge!.

One option would be to use the MME as the initial guess!

So define: $g(\hat{k}) = \ln \hat{k} - \psi(\hat{k}) + \overline{\ln x} - \ln \bar{x}$

then $g'(\hat{k}) = \frac{1}{\hat{k}} - \psi'(\hat{k})$ ← trigamma function

and $$k_{n+1} = k_n - \frac{g(k_n)}{g'(k_n)}$$

$$= k_n - \frac{\ln k_n - \psi(k_n) + \overline{\ln x} - \ln \bar{x}}{\frac{1}{k_n} - \psi'(k_n)}$$

R code:

```
mle = function(x) {
  n= length(x)
  eps= 1e-6
  c= mean(log(x)) - log(mean(x))
  f= function(k) { log(k) - digamma(k)+c }
  fp= function(k) { 1/k - trigamma(k) }
  k= n * mean(x)^2 / ((n-1) * var(x))  # MME.
  diff = 2
  while (abs(diff)>eps) {
    diff = f(k) / fp(k)
    k = k - diff
  }
  c(scale= mean(x)/k, shape= k)
}
x = rgamma(1000, shape= 5, scale= 9) # random data
mle(x)
```