7 POINT ESTIMATION

POINT ESTIMATOR - any function $W(X_1, X_2, \dots X_n)$ of a samples that is, any statistic is a point estimate

NOTE: 1) No correspondence between estimator and parameter
2) No mention of range

Estimate vs. Estimator

- Estimator - is a function of the sample
- Estimate - is the realized value of an estimator, a number, obtained when a sample is taken

§7.2

- Estimating a parameter with its sample analog is usually reasonable
  - The sample mean is a good estimate for the population mean

Method of moments Usually yields an estimate, however this estimate can usually be improved upon.

- Let $X_1, X_2, \dots X_n$ be a sample from a population w/ pdf or pmf $f(x|\theta_1, \theta_2, \dots \theta_k)$ Method of moments estimators are found by equating the first k sample moments to the coorresponding k population moments, and solving the resulting system of simultaneous equations.

$$m_1 = \frac{1}{n}\sum_i^n X_i^1 \qquad \mu_1' = E(X^1)$$
$$m_2 = \frac{1}{n}\sum_i^n X_i^2 \qquad \mu_2' = E(X^2)$$
$$\vdots$$
$$m_k = \frac{1}{n}\sum_i^n X_i^k \qquad \mu_k' = E(X^k)$$

NOTE: μ_j' will typically be a function of $\theta_1, \dots, \theta_k$ say $\mu_j'(\theta_1, \dots \theta_k)$ the estimator $(\tilde{\theta}_1, \tilde{\theta}_2, \dots \tilde{\theta}_k)$ of $(\theta_1, \theta_2, \dots \theta_k)$ is obtained by solving the following system of equations for $(\theta_1, \theta_2, \dots \theta_k)$ in terms of $(m_1, m_2, \dots)$

$$m_1 = \mu_1'(\theta_1, \theta_2, \dots \theta_k)$$
$$m_2 = \mu_2'(\theta_1, \theta_2, \dots \theta_k)$$
$$\vdots$$
$$m_k = \mu_k'(\theta_1, \theta_2, \dots \theta_k)$$

Example 7.2.1: $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ $\theta_1 = \theta$, $\theta_2 = \sigma^2$

$$m_1 = \bar{X} \qquad \mu_1 = \theta$$
$$m_2 = \frac{1}{n}\sum_1^n x_i^2 \qquad \mu_2 = \theta^2 + \sigma^2$$

Solve: $\bar{X} = \theta$

$$\frac{1}{n}\sum_1^n x_i^2 = \theta^2 + \sigma^2$$

To get: $\begin{cases} \tilde{\theta} = \bar{X} \\ \tilde{\sigma}^2 = \frac{1}{n}\sum_1^n X_i^2 - \bar{X}^2 = \frac{1}{n}\sum (X_i - \bar{X})^2 \end{cases}$ Note: this equals S_8

These are method of moment estimators

Cool crime rate application → Example 7.2.2: $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{binomial}(K, p)$, that is,

$$P(X_i = x \mid K, p) = \binom{K}{x} p^x (1-p)^{K-x}, \quad x = 0, 1, \ldots K$$

NOTE: here we assume K and p are unknown point estimators

$$m_1 = \bar{X} \qquad \mu_1 = Kp$$
$$m_2 = \frac{1}{n}\sum_1^n x_i^2 \qquad \mu_2 = Kp(1-p) + K^2p^2$$

Solve: $\bar{X} = Kp$

$$\frac{1}{n}\sum_1^n x_i^2 = Kp(1-p) + K^2p^2$$

To get: $\left.\begin{aligned} \tilde{K} &= \frac{\bar{X}^2}{\bar{X} - (1/n)\sum (X_i - \bar{X})^2} \\ \tilde{p} &= \bar{X}/\tilde{K} \end{aligned}\right\}$ Method of moments estimator

Example 7.2.3 · Y_i, $i=1,2,\dots k$ are independent $\chi^2_{(r_i)}$ random variables

· Note: $\sum_1^k Y_i \sim \chi^2_{(\sum r_i)}$

· Note: $\sum_1^k a_i Y_i \sim ?$ for a_i, known constants

→ we can assume $\chi^2_{(v)}$ will be a good approximation for some value of v

· This is almost Satterthwaite's problem of approximating the denominator of a t stat: $\sum_1^k a_i Y_i$ represents the square of the denominator of his stat.

· So he wanted to find $v \ni$

$$\sum_1^k a_i Y_i \sim \chi_{(v)}^2/v$$

· Since $E(\chi_{(v)}^2/v) = 1$ to match first moments we need

$$E\left(\sum_1^k a_i Y_i\right) = \sum_1^k a_i E(Y_i) = \sum_1^k a_i r_i = 1$$

Note: this gives us constraints on a but no information on how to estimate v

· second moments

$$E\left(\sum_1^k a_i Y_i\right)^2 = E\left(\frac{\chi_{(v)}^2}{v}\right)^2 = \frac{2}{v} + 1$$

· Applying the method of moments

$$\hat{v} = \frac{2}{\left(\sum_1^k a_i Y_i\right)^2 - 1}$$ Note: This can obtain negative values

· Satterthwaite's customization

$$E\left(\sum a_i Y_i\right)^2 = Var\left(\sum a_i Y_i\right) + \left(E \sum_1^k a_i Y_i\right)^2$$

$$= \left(E\left(\sum a_i Y_i\right)^2\right)\left[\frac{Var\left(\sum a_i Y_i\right)}{\left(E\left(\sum a_i Y_i\right)^2\right)} + 1\right]$$ NOTE: $E(\sum a_i Y_i) = 1$

Now, $$\hat{v} = \frac{\sum (a_i Y_i)^2}{\sum \frac{a_i^2}{r_i} Y_i^2}$$ ← Better

9.9.3

Maximum Likelihood Estimators – The most popular technique

Recall: $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta_1, \ldots \theta_k)$

$$L(\theta|\underset{\sim}{x}) = f_{\underset{\sim}{x}}(\underset{\sim}{x}|\theta) = \prod_1^n f(x_i|\theta_1, \ldots \theta_k)$$

Def 7.2.4 $\forall$ sample point $\underset{\sim}{x}$, let $\hat{\theta}(\underset{\sim}{x})$ be a parameter value at which $L(\theta|\underset{\sim}{x})$ attains its maximum as a function of θ with $\underset{\sim}{x}$ held fixed. A <u>maximum likelihood estimator</u> (MLE) of the parameter θ based on a sample $\underset{\sim}{X}$ is $\hat{\theta}(\underset{\sim}{X})$

NOTE: MLE coincides w/ range of the MLE

Note: <u>Pros</u>: usually a good estimate

<u>Cons</u>
- finding max of functions and ensuring they are global
- How sensitive the estimates are to small changes in data

Example 7.2.5 $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\theta, 1)$ then

$$L(\theta|\underset{\sim}{x}) = \prod_1^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\theta)^2}{2}} = \frac{1}{(2\pi)^{n/2}} e^{-\left(\sum_1^n (x_i-\theta)^2\right)/2}$$

Consider $\frac{d}{d\theta} L(\theta|\underset{\sim}{x}) = 0$

$\vdots$

$$\sum_1^n (x_i - \theta) = 0$$

Solution: $\hat{\theta} = \bar{x}$ Thus $\bar{x}$ is a candidate for the MLE

* To verify that $\bar{x}$ is the global max we use the following argument

1) $\hat{\theta} = \bar{x}$ is the only solution for $\frac{d}{d\theta} L(\theta|\underset{\sim}{x}) = 0$
2) Verify $\frac{d^2}{d\theta^2} L(\theta|\underset{\sim}{x})\big|_{\theta=\bar{x}} < 0$ (concave down $\to$ max)
3) Check boundaries at $\pm\infty$

Direct maximization

Example 7.2.6 $\forall$ number a $\sum_1^n (x_i - a)^2 \geq \sum_1^n (x_i - \bar{x})^2$ w/ equality at $a = \bar{x}$

this implies that $\forall \theta$ $e^{-(x_i-\theta)^2/2} \leq e^{-(x_i-\bar{x})^2/2}$ w/ equality iff $\theta = \bar{x}$ hence $\bar{x}$ is the MLE

* In most cases, especially when differentiating is to be used it is easier to work w/ the natural logarithm of $L(\theta|\underset{\sim}{x})$, $\ln(L(\theta|\underset{\sim}{x}))$ known as the log likelihood. This is possible because the log function is strictly increasing on $(0,\infty)$ which implies the extrema of $L(\theta|\underset{\sim}{x})$ and $\log(L(\theta|\underset{\sim}{x}))$

Example 7.2.7 $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Bernoulli}(p)$

- $L(p|\underset{\sim}{x}) = \prod_1^n p^{x_i}(1-p)^{1-x_i} = p^y(1-p)^{n-y}$ $\quad y = \sum x_i$
- Note this isn't hard to differentiate but the log is much easier

$$\ln(L(p|\underset{\sim}{x})) = y\log(p) + (n-y)\ln(1-p) \quad \text{if } 0<y<n$$

- Solve $\frac{d}{dp}\ln(L(p|\underset{\sim}{x})) = 0$
- to get $\hat{p} = y/n$
- It is easy to show this is the global max, thus the MLE of p is $\sum x_i/n$

NOTE: The maximization takes place only over the range of parameter values.

Example 7.2.8 $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\theta, 1)$ where it is known that θ must be nonnegative. Earlier, with no restrictions in θ, we found the MLE of θ is $\bar{x}$. With this restriction if $\bar{x}$ is negative we have big problems

If $\bar{x}$ is negative, the likelihood function $L(\theta|\underset{\sim}{x})$ is decreasing in θ for $\theta \geq 0$ and is maximized at $\hat{\theta} = 0$ thus the MLE is

$$\hat{\theta} = \begin{cases} \bar{x} & \bar{x} \geq 0 \\ 0 & \bar{x} < 0 \end{cases}$$

* An important feature of an MLE is that it is possible to use a computer to maximize $L(\theta|\underset{\sim}{x})$

Example 7.2.9 · $X_1, X_2, \dots X_n$ be a random sample from binomial(k,p) when p is known and k is unknown.

· For example, we flip a coin we know to be fair and observe x_i heads but we do not know how many times the coin was flipped

$$L(k|\underline{x},p) = \prod_i \binom{k}{x_i} p^{x_i}(1-p)^{k-x_i}$$

· Maximizing this by differentiation is difficult because of the factorials and because k must be an integer

· Let's try a different approach

$L(k|\underline{x},p) = 0$ if $k < \max_i x_i$ thus the MLE is an integer $k \geq \max_i x_i$ that satisfies

$$\frac{L(k,\underline{x},p)}{L(k-1|\underline{x},p)} \geq 1 \quad \text{and} \quad \frac{L(k+1|\underline{x},p)}{L(k|\underline{x},p)} < 1$$

· We will show that $\exists$ only one such value k

· The ratio of likelihoods is $\dfrac{L(k|\underline{x},p)}{L(k-1|\underline{x},p)} = \dfrac{(k(1-p))^n}{\prod_i (k-x_i)}$

· Thus the condition for a maximum is

$(k(1-p))^n \geq \prod_i (k-x_i)$ and $((k+1)(1-p))^n < \prod_i (k+1-x_i)$

· Let $z = 1/k$ and divide through by k^n

Solve $(1-p)^n = \prod_i (1-x_i z)$ for $0 \leq z \leq 1/\max_i x_i$

· As the RHS is a decreasing function of z with a value of 1 at $z=0$ and a value of 0 at $z = \frac{1}{\max_i x_i}$

· Thus $\exists$ a unique z, $\hat{z}$ that solves the equation is the MLE and the largest integer $\leq 1/\hat{z}$

$\therefore \exists$ a unique max for the loglikelihood

The invariance property of MLEs

Suppose that a distribution is indexed by a parameter θ but the interest is in finding an estimator for some function of θ say $\tau(\theta)$. The idea is that if $\hat{\theta}$ is the MLE of θ then $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$

For example, if θ is the mean of a normal dist the MLE of $\sin(\theta)$ is $\sin(\bar{x})$

*If the mapping $\theta \to \tau(\theta)$ is 1:1 then there is no problem

If we let $\eta = \tau(\theta)$ then the inverse function $\tau^{-1}(\eta) = \theta$ is well defined and

$$L^*(\eta|\underset{\sim}{x}) = \prod_{i} f(x_i|\tau^{-1}(\eta)) = L(\tau^{-1}(\eta)|\underset{\sim}{x})$$

and

$$\sup_{\eta} L^*(\eta|\underset{\sim}{x}) = \sup_{\eta} L(\tau^{-1}(\eta|\underset{\sim}{x})) = \sup_{\theta} L(\theta|\underset{\sim}{x})$$

- Thus the maximum of $L^*(\eta/\underset{\sim}{x})$ is attained at $\eta = \tau(\theta) = \tau(\hat{\theta})$ showing that the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$
- Note: there are many scenarios which we desire this for functions that aren't 1:1 such as θ^2 where there may be more than on value of $\theta \ni \tau(\theta) = \eta$
- Define for $\tau(\theta)$ the induced likelihood function L^* given by

$$L^*(\eta|\underset{\sim}{x}) = \sup_{\{\theta | \tau(\theta) = \eta\}} L(\theta|\underset{\sim}{x})$$

- The value $\hat{\eta}$ that maximizes $L^*(\eta|\underset{\sim}{x})$ will be the MLE of $\eta = \tau(\theta)$ and the maxima of L^* and L coincide

Theorem 7.2.10 If $\hat{\theta}$ is the MLE of θ, then $\forall$ function $\tau(\theta)$ the MLE of $\tau(\theta)$ is $\tau(\hat{\theta})$

Proof page 320

- Using this theorem, we now see that MLE of θ^2, the square of a normal mean, is $\bar{x}^2$,
- Another example: the MLE of $\sqrt{p(1-p)}$ where p is binomial is given by $\sqrt{\hat{p}(1-\hat{p})}$
- This also holds in the multivariate case, letting θ be a vector

Example 7.2.11 Let $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\theta, \sigma^2)$ with both θ and σ^2 unknown

$$L(\theta, \sigma^2|\underline{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2}\sum_i^n (x_i-\theta)^2/\sigma^2}$$

and

$$\log(L(\theta,\sigma^2|\underline{x})) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2}\sum(x_i-\theta)^2/\sigma^2$$

$$\partial/\partial\theta \log(L(\theta,\sigma^2|\underline{x})) = \frac{1}{\sigma^2}\sum_i(x_i-\theta)$$

and

$$\partial/\partial\sigma^2 \log(L(\theta,\sigma^2|\underline{x})) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_i(x_i-\theta)^2$$

Set equal to 0 and solve to get:

$$\hat{\theta} = \bar{x}$$
$$\hat{\sigma}^2 = n^{-1}\sum_i(x_i-\bar{x})^2$$

- Now if $\theta \neq \bar{x}$ then $\sum(x_i-\theta)^2 > \sum(x_i-\bar{x})^2$ $\therefore \forall\ \sigma^2$

$$\frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2}\sum_i (x_i-\bar{x})^2/\sigma^2} \geq \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-\frac{1}{2}\sum_i (x_i-\theta)^2/\sigma^2}$$

$\therefore$ We made this a one-dimensional problem, verifying $(\sigma^2)^{-n/2} e^{-1/2 \sum_i (x_i-\theta)^2/\sigma^2}$ achieves a max at $\sigma^2 = n^{-1}\sum(x_i-\bar{x})^2$

$(\bar{x}, n^{-1}\sum(x_i-\bar{x})^2)$ are the MLEs

Another example 7.2.12 ~~is~~ omitted

Bayes Estimators

- θ, the parameter, is considered to be a quantity who's variation can be described as a probability distribution, we call this the <u>prior distribution</u>

- A sample is then taken from the population so we can update the prior to get <u>the posterior</u> distribution

We donote: 1) $\pi(\theta)$ as the <u>prior</u>
2) $f(\underline{x}|\theta)$ as the <u>sampling distribution</u>
3) <u>Posterior</u>; $\dfrac{f(\underline{x}|\theta)\,\pi(\theta)}{\int f(\underline{x}|\theta)\,\pi(\theta)\,d\theta} = \pi(\theta|\underline{x})$

<u>Eg 7.2.14</u>
- $X_1, X_2, \ldots X_n \overset{iid}{\sim} \text{Bernoulli}(p)$
- $Y = \sum X_i \sim \text{bin}(n,p)$
- Assume the prior is $p \sim \text{beta}(\alpha,\beta)$
- Find $f_{Y,P}(y,p) = f(y|p) \times \pi(p) = \left[\binom{n}{y} p^y (1-p)^{n-y}\right] \times \left[\dfrac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1} (1-p)^{\beta-1}\right]$

$$= \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)(\beta)} p^{y+\alpha-1} (1-p)^{(n-y+\beta-1)}$$

- Find the distribution of y: $\displaystyle\int_0^1 \binom{n}{y} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)(\beta)} p^{y-\alpha-1} (1-p)^{n-y+\beta-1}\, dp$

$$= \binom{n}{y} \cdot \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(y+a)\,\Gamma(n-y+\beta)}{\Gamma(n+\alpha+\beta)}$$

$y \sim$ beta binomial

- Find the prior $\pi(p|y) = \dfrac{f_{Y,P}(y,p)}{f(y)} = \dfrac{\Gamma(n+\alpha+\beta)}{\Gamma(y+\alpha)\,\Gamma(n-y+\beta)} p^{y+\alpha-1} (1-p)^{n-y+\beta-1}$

$\sim \text{beta}(y+\alpha, n-y+\beta)$

- Bayes estimate $= \hat{p} = \dfrac{y+\alpha}{\alpha+\beta+n}$ as it is the mean of above

Def 7.2.15 Let $\mathcal{F}$ denote the class of pdfs or pmfs $f(x|\theta)$. A class Π of prior distributions is a <u>conjugate family</u> for $\mathcal{F}$ if the posterior distribution is in the class Π for all $f \in \mathcal{F}$, all priors in Π and all $x \in X$

ie The beta family is a conjugate of the binomial family. Thus if we start w/ a beta prior our posterior will also be beta

Example 7.2.16: Let $X \sim N(\theta, \sigma^2)$ and suppose that the prior distribution on θ is $N(\mu, \tau^2)$. Assume σ^2, μ, and τ^2 are known. The posterior distribution of θ is also normal, w/ mean and variance

$$E(\theta|x) = \frac{\tau^2}{\tau^2+\sigma^2} x + \frac{\sigma^2}{\sigma^2+\tau^2} \mu$$

$$var(\theta|x) = \frac{\sigma^2 \tau^2}{\sigma^2+\tau^2}$$

SKIPPED EM Algorithm

§7.3

Def 7.3.1 The <u>mean square error</u> (MSE) of an estimator, W, of a parameter θ is the function of θ defined by $E_\theta[(W-\theta)^2]$

- The MSE measures the average squared difference between the estimator W and parameter θ
- Notice: $E_\theta[(W-\theta)^2] = Var_\theta(W) + (E(W)-\theta)^2 = var_\theta(W) + (Bias_\theta(W))^2$

Def 7.3.2 The <u>bias</u> of a point estimator W of a parameter θ is the difference between the expected value of W and θ; That is,
$Bias_\theta(W) = E_\theta(W) - \theta$

- $Bias_\theta(W) = 0 \rightarrow$ <u>unbiased estimator</u>

$var_\theta(W) \longrightarrow$ precision
$E(W) - \theta \longrightarrow$ accuracy

Example 7.3.3 $X_1, X_2, \ldots, X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ the stats $\bar{X}$ and S^2 are both unbiased estimators since

$$E(\bar{X}) = \mu$$
$$E(S^2) = \sigma^2$$

- MSEs

$$E[(\bar{X}-\mu)^2] = var(\bar{X}) = \sigma^2/n$$
$$E[(S^2-\sigma^2)^2] = var(S^2) = 2\sigma^4/(n-1)$$ ← Requires normality

Example 7.3.4 $\hat{\sigma}^2 = \frac{1}{n}\sum_i^n (X_i - \bar{X})^2 = \frac{n-1}{n} S^2$

1) $E(\hat{\sigma}^2) = E\left(\frac{n-1}{n} S^2\right) = \frac{n-1}{n} E(S^2) = \frac{n-1}{n}\sigma^2$

∴ $\hat{\sigma}^2$ is a biased estimator

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

3) MSE: $E(\hat{\sigma}^2 - \sigma^2)^2 = \underbrace{\frac{2(n-1)\sigma^4}{n^2}}_{var(\hat{\sigma}^2)} + \underbrace{\left(\frac{n-1}{n}\sigma^2 - \sigma^2\right)^2}_{Bias_\theta(\hat{\sigma}^2)} = \left(\frac{2n-1}{n^2}\right)\sigma^4$

NOTE: $$\underbrace{\left(\frac{2n-1}{n^2}\right)\sigma^4}_{MSE(\hat{\sigma}^2)} < \underbrace{\left(\frac{2}{n-1}\right)\sigma^4}_{MSE(S^2}$$

$\hat{\sigma}^2$ is biased but has a better MSE than S^2

- Our allowing bias had a big decrease in Variance

Example 7.3.5 • $X_1, X_2, \ldots X_n \overset{iid}{\sim}$ Bernoulli(p). The MSE of $\hat{p}$, the MLE, as an estimator of p is

$$E_p(\hat{p}-p)^2 = Var_p(\bar{X}) = \frac{p(1-p)}{n}$$

- Let $Y = \sum X_i$, recall earlier $\hat{p}_B = \frac{Y+\alpha}{\alpha+\beta+n}$
- $MSE_\theta(\hat{p}_B) = E_p(\hat{p}_B - p)^2 = Var_p(\hat{p}_B) + (Bias_p\ \hat{p}_B)^2 = Var(\hat{p}_B) + (E(\hat{p}_B)-p)^2$

$$= \frac{np(1-p)}{(\alpha+\beta+n)^2} + \left(\frac{np+\alpha}{\alpha+\beta+n} - p\right)^2$$

- We might want to choose α, β to make MSE constant: $\alpha = \beta = \sqrt{\frac{n}{4}}$

$$\hat{p}_B = \frac{Y + \sqrt{n/4}}{n+\sqrt{n}} \qquad E(\hat{p}_B - p)^2 = \frac{n}{4(n+\sqrt{n})^2}$$

* Choosing between estimators! One may be better for smaller values than the other.

- For a fixed g in the group $\mathcal{G}$, denote the function that takes $\theta \to \theta'$ by $\bar{g}(\theta) = \theta'$. Then if $W(\underline{x})$ estimates θ we have

  Measurement Equivariance: $W(\underline{x})$ estimates $\theta \to \bar{g}(W(\underline{x}))$ estimates $\bar{g}(\theta) = \theta'$

  Formal Invariance: $W(\underline{x})$ estimates $\theta \to W(g(\underline{x}))$ estimates $\bar{g}(\theta) = \theta$

- Example 7.36 $X_1, X_2, \dots X_n$ be iid $f(x-\theta)$. For an estimator $W(\underline{x})$ to satisfy $W(g_a(\underline{x})) = g_a(W(\underline{x}))$ we must have

$$W(x_1, x_2, \dots x_n) + a = W(x_1+a, x_2+a, \dots x_n+a)$$

which specifies the equivariant estimators w/ respect to the group of transformations defined by $\mathcal{G} = \{g_a(\underline{x}) \mid -\infty < a < \infty\}$ where $g_a(x_1, x_2, \dots x_n) = (x_1+a, \dots x_n+a)$

- For these estimators we have

$$\begin{aligned}
&E(W(X_1, X_2, \dots X_n) - \theta)^2 \\
&= E_\theta((W(X_1+a, \dots X_n+a) - a - \theta)^2) \\
&= E_\theta((W(X_1-\theta, \dots X_n-\theta))^2) \qquad (a = -\theta) \\
&= \int\int\cdots\int ((W(x_1-\theta, \dots x_n-\theta))^2) \prod f(x_i-\theta)\, dx_i \\
&= \int\int\cdots\int ((W(u_1, u_2, \dots u_n))^2) \prod f(u_i)\, du_i
\end{aligned}$$

↑ this doesn't depend on θ ∴ the MSEs of these equivariant estimators are not functions of θ. The MSE can therefore be used to order the equivariant estimators and an equivariant estimator w/ the smallest MSE can be found

Definition 7.3.7 An estimator W^* is a best unbiased estimator of $\tau(\theta)$ if it satisfies $E_\theta(W^*) = \tau(\theta)\ \forall\ \theta$ and for any other estimator W' w/ $E_\theta(W) = \tau(\theta)$ we have $Var_\theta(W^*) \le Var_\theta(W')\ \forall\ \theta$

(UMVUE)

* Best Unbiased Estimator = Uniform minimum variance unbiased estimator

Example 7.3.8 Let $X_1, X_2, \dots X_n \overset{iid}{\sim}$ Poisson(λ)

Let $\bar{X}$ be the sample mean

Let S^2 be the sample variance

$E(\bar{X}) = \lambda$
$E(S^2) = \lambda$ } Both are unbiased estimators of λ, which is "best"?

Recall $var(\bar{X}) = \lambda/n$

$var(S^2) = ?$ Omitted

but we will find $Var_\lambda(\bar{X}) \le Var_\lambda(S^2)\ \forall\ \lambda$

thus $\bar{X}$ is a "better" estimator

Theorem 7.3.9 $X_1, X_2, \dots X_n$ be a sample w/ pdf $f(\underline{x}|\theta)$ and let $W(\underline{X}) = W(X_1, X_2, \dots, X_n)$ be an estimator satisfying

$$\frac{d}{d\theta} E_\theta(W(\underline{X})) = \int_x \frac{\partial}{\partial\theta}\left[W(\underline{x}) f(\underline{x}|\theta)\right] dx$$

and

$$Var_\theta(W(\underline{x})) < \infty$$

Then:

$$Var_\theta(W(\underline{x})) \ge \frac{\left(\frac{d}{d\theta} E_\theta(W(\underline{x}))\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta} \log f(\underline{X}|\theta)\right)^2\right)}$$

Proof on page 336 using Cauchy-Schwarz

Corollary 7.3.10 If $X_1, X_2, \dots X_n$ are iid theorem 7.3.9 becomes

$$Var_\theta(W(\underline{x})) \ge \frac{\left(\frac{d}{d\theta} E_\theta(W(\underline{x}))\right)^2}{n E_\theta\left(\left(\frac{\partial}{\partial\theta} \log f(x|\theta)\right)^2\right)}$$

Information number (Fisher information) $= E\left(\left(\frac{\partial}{\partial\theta}\log f(\underline{X}|\theta)\right)^2\right)$

* As the information number gets larger we have more information about θ and we have a smaller bound on the variance of the UMVUE

Lemma 7.3.11 If $f(x|\theta)$ satisfies $\frac{d}{d\theta} E_\theta\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right) = \int \frac{\partial}{\partial\theta}\left[\frac{\partial}{\partial\theta}\log(f(x|\theta))\, f(x|\theta)\right]$

Note: this is true for exponential families

then $E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f(x|\theta)\right)$

Example 7.3.12 · $X_1, X_2, \dots X_n \sim \text{Poisson}(\lambda)$

· Consider $\tau(\lambda)=\lambda \rightarrow \tau'(\lambda)=1$

· As the poisson is a exponential family Lemma 7.3.11

$$\rightarrow E_\lambda\left(\left(\frac{\partial}{\partial\lambda}\log\left(\prod_i f(x_i|\lambda)\right)\right)^2\right) = -n\,E_\lambda\left(\frac{\partial^2}{\partial\lambda^2}\log(f(x|\lambda))\right)$$

$$= -n\,E_\lambda\left(\frac{\partial^2}{\partial\lambda^2}\log\left(\frac{e^{-\lambda}\lambda^x}{x!}\right)\right)$$

$$= -n\,E_\lambda(-x/\lambda^2) = -n(-1/\lambda) = n/\lambda$$

$\therefore$ for any unbiased estimator W, of λ, $\text{var}(W) \geq \lambda/n$

Example 7.3.13 $X_1, X_2, \dots X_n$ iid $f(x|\theta) = 1/\theta \quad I(0<x<\theta)$

$\partial/\partial\theta \log(f(x|\theta)) = \partial/\partial\theta \log(1/\theta) = \frac{1}{1/\theta}\cdot\frac{-1}{\theta^2} = -1/\theta$

$E_x(1/\theta^2) = 1/\theta^2$

$E(W)=\theta$

$\left(\frac{d}{d\theta}\theta\right)^2 = 1$

$$\text{Var}(W(\underline{X})) \geq \frac{\left(\frac{d}{d\theta}E_\theta(W(\underline{X}))\right)^2}{E_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(\underline{X}|\theta)\right)^2\right)} = \frac{\left(\frac{d}{d\theta}\theta\right)^2}{nE_\theta\left(\left(\frac{\partial}{\partial\theta}\log f(x|\theta)\right)^2\right)} = \frac{1}{\frac{n}{\theta^2}} = \frac{\theta^2}{n}$$

We note that CR isn't applicable to this pdf as the interchange of integral and derivative operators doesn't hold.

Example 7.3.14, $X_1, X_2, \ldots X_n \overset{iid}{\sim} N(\mu, \sigma^2)$ Normal satisfys CR assumptions

Let μ be known and σ^2 be unknown $\quad \tau = \sigma^2$

$$var(W(\underline{X})) \geq \frac{\left(\frac{d}{d\sigma^2} E_{\sigma^2}(W(\underline{X})\right)^2}{E_{\sigma^2}\left(\left(\frac{\partial}{\partial\sigma^2}\log f(\underline{X}|\sigma^2)\right)^2\right)} \underset{7.3.10}{\uparrow}= \frac{\left(\frac{d}{d\sigma^2}\sigma^2\right)^2}{n E_{\sigma^2}\left(\left(\frac{d}{d\sigma^2}\log(f(x|\sigma^2))\right)^2\right)} = \frac{1}{n/2\sigma^4} = \frac{2\sigma^4}{n}$$

$\therefore$ Any unbiased estimator of σ^2 must have variance $\geq \frac{2\sigma^4}{n}$

*Recall, earlier, we found $Var(S^2|\mu,\sigma^2) = \frac{2\sigma^4}{n-1}$ so S^2 does not attain the Cramer Rao lower bound - the obvious question now is - can we find a statistic that attains the Cramer-Rao lower bound?

Corollary 7.3.15 - Let $X_1, X_2, \ldots X_n \overset{iid}{\sim} f(x|\theta)$, where $f(x|\theta)$ satisfies the assumptions of the Cramer Rao Theorem. Let $L(\theta|\underline{x}) = \prod_i f(x_i|\theta)$ denote the likelihood function.

If $W(\underline{X}) = W(X_1, X_2, \ldots X_n)$ is an unbiased estimator of $\tau(\theta)$ then $W(\underline{X})$ attains the Cramer Rao lower bound iff $a(\theta)[W(\underline{x}) - \tau(\theta)] = \frac{\partial}{\partial\theta}\log(L(\theta|\underline{x}))$ for some $a(\theta)$

Example 7.3.14 continued

$$L(\mu,\sigma^2|\underline{x}) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2)\sum_1^n \frac{(x_i-\mu)^2}{\sigma^2}}$$

$$\frac{\partial}{\partial\sigma^2}\log(L(\mu,\sigma^2|\underline{x})) = \frac{n}{2\sigma^4}\left(\sum_1^n \frac{(x_i-\mu)^2}{n} - \sigma^2\right)$$

if μ is known $a(\sigma^2) = \frac{n}{2\sigma^4}$, $W(\underline{x}) = \sum_1^n \left(\frac{(x_i-\mu)^2}{n}\right)$, $\tau(\sigma^2) = \sigma^2$

if μ is unknown we cannot attain the CRLB

Theorem 7.3.17 - Rao-Blackwell • Let W be any unbiased estimator of $\tau(\theta)$ and let T be a sufficient statistic for θ.

• Define $\phi(T) = E(W|T)$. Then $E_\theta(\phi(T)) = \tau(\theta)$ and $Var_\theta(\phi(T)) \leq Var_\theta(W)$ $\forall \theta \rightarrow \phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$

Proof 342

Theorem 7.3.19 If W is a UMVUE of $\tau(\theta)$ then W is unique

proof 344

Theorem 7.3.20 If $E_\theta(W) = \tau(\theta)$, W is the best unbiased estimator of $\tau(\theta)$ iff W is uncorrelated with all unbiased estimators of 0.

proof 345

Example 7.3.21 Let X be an observation from a uniform $(\theta, \theta+1)$ distribution. Then $E_\theta(X) = \int_\theta^{\theta+1} x\,dx = \theta + \frac{1}{2} \rightarrow X - \frac{1}{2}$ is an unbiased estimator of θ. w/ $var(X) = \frac{1}{12}$

find unbias estimates of 0

$\int_\theta^{\theta+1} h(x)dx = 0 \quad \forall \theta$

then $\frac{d}{d\theta}\int_\theta^{\theta+1} h(x)dx = h(\theta+1) - h(\theta) = 0 \quad \forall \theta$ *Note: we're looking for a periodic function period=1 ie $\sin(2\pi X)$

$$Cov_\theta\left(X - \tfrac{1}{2}, \sin(2\pi X)\right) = -\frac{\cos(2\pi\theta)}{2\pi}$$

$\therefore X - \frac{1}{2}$ is not UMVUE

Theorem 7.3.23 Let T be a complete sufficient statistic for a parameter θ and let $\phi(T)$ be any estimator based only on T. Then $\phi(T)$ is the UMVUE of its expected value.

*Skipped 7.3.4

<u>Example 7.3.24</u> $X_1, X_2, \ldots X_n \overset{iid}{\sim}$ binomial(k, θ)

$\tau(\theta) = P_\theta(X=1) = k\theta(1-\theta)^{k-1}$

Note: $\sum_1^n X_i \sim$ binomial(kn, θ) is a complete sufficient statistic

Consider $h(X_1) = \begin{cases} 1 & X_1 = 1 \\ 0 & \text{ow} \end{cases}$

$E(h(x_1)) = \sum_0^k h(x_1)\binom{k}{x_1}\theta^{x_1}(1-\theta)^{k-x_1} = k\theta(1-\theta)^{k-1}$

$\therefore\ \phi(\sum_1^n X_i) = E(h(x_1) \mid \sum_1^n X_i)$ is the UMVUE of $k\theta(1-\theta)^{k-1}$

$= k \dfrac{\binom{k(n-1)}{\sum X_i - 1}}{\binom{kn}{\sum X_i}}$

<u>Theorem 7.5.1</u> <u>Lehmann-Scheffé</u> Unbiased estimators based on complete sufficient statistics are unique