

stats6255point 625.4 - Point estimation

Gunnar Stefánsson

26. nóvember 2019

Copyright This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

Efnisyfirlit

1	Methods of Point Estimation	4
1.1	Point Estimation	4
1.1.1	Handout	4
1.2	Maximum Likelihood Estimators	4
1.2.1	Handout	4
1.3	Method of Moments	5
1.3.1	Handout	5
1.4	Comparing estimators	6
1.4.1	Handout	6
1.5	Method of Least Squares	8
1.5.1	Handout	8
1.6	Linear Estimators	8
1.6.1	Handout	8
1.7	Minimum Chi-Squared	8
1.7.1	Handout	8
1.8	Induced Likelihood Function	9
1.8.1	Handout	9
2	The quality of estimators	10
2.1	Quality of estimators	10
2.1.1	Handout	10
2.2	Best estimators (UMVUE)	10
2.2.1	Handout	10
2.3	Best linear unbiased estimators (BLUE)	11
2.3.1	Handout	11
3	The Cramer-Rao inequality	12
3.1	The Cramer-Rao inequality	12
3.1.1	Handout	12
3.2	A version for i.i.d. random variables	13
3.2.1	Handout	13
3.3	Fisher information	14
3.3.1	Handout	14
3.4	Rewriting the Fisher information	14
3.4.1	Handout	14
3.5	The C-R inequality for i.i.d. random variables	14
3.5.1	Handout	14
3.6	When the assumptions fail	14
3.6.1	Handout	14
3.7	A corollary using the likelihood function	14
3.7.1	Handout	14
4	Sufficiency and unbiased estimators	16
4.1	Background	16
4.1.1	Handout	16
4.2	The Rao-Blackwell theorem	16
4.2.1	Handout	16
4.3	Lehmann–Scheffé	17
4.3.1	Handout	17

5 Overview of point estimation	19
5.1 Summary	19

1 Methods of Point Estimation

1.1 Point Estimation

A (point) estimator is a function of random variables, $T = T(X_1, \dots, X_n)$, which is itself also a random variable.

A point estimate is an outcome of the estimator $t = T(x_1, \dots, x_n) = T(\mathbf{X}(u))$.

1.1.1 Handout

A (point) estimator is a function of random variables, $T = T(X_1, \dots, X_n)$, which is itself also a random variable. A point estimate is an outcome of the estimator $t = T(x_1, \dots, x_n) = T(\vec{X}(u))$.

Many methods are used to derive estimators:

- Maximum likelihood
- Method of moment
- Minimum χ^2
- Least squares
- Best Linear Unbiased Estimators (BLUE)
- and any other method one can come up with

1.2 Maximum Likelihood Estimators

1.2.1 Handout

Consider a collection of random variables $\mathbf{X} = (X_1, \dots, X_n)$ which have a distribution with joint density f_g . For a given set of data \mathbf{x} the likelihood function is defined by

$$L_{\mathbf{x}}(\theta) := f_{\theta}(\mathbf{x})$$

and if we set

$$\hat{\theta}(\mathbf{x}) = \arg \max_{\theta \in \Theta} L_{\mathbf{x}}(\theta)$$

then the estimator

$$\hat{\theta} = \hat{\theta}(\mathbf{X})$$

the the maximum likelihood estimator (MLE) for θ .

Example 1.1. Let $X_1, \dots, X_n \sim U(0, \theta)$ be i.i.d. Then $f_{\theta}(x_i) = \begin{cases} 1/\theta & 0 \leq x \leq \theta \\ 0 & \text{otherwise} \end{cases}$, and the joint density is the product of these functions so the likelihood function is

$$L_{\mathbf{x}}(\theta) = \prod_{i=1}^n f_{\theta}(x_i) = \prod_{i=1}^n \frac{1}{\theta} I_{[0, \theta]}(x_i)$$

Order the values $x_{(1)} \leq \dots \leq x_{(n)}$ so that when they are all positive, i.e. $0 \leq x_{(1)} \leq x_{(n)}$ then

$$L_{\mathbf{x}}(\theta) = \frac{1}{\theta^n} I_{[x_{(n)}, \infty)}(\theta).$$

We therefore see that the function has a maximum at $x_{(n)}$ so the MLE is $\hat{\theta} = X_{(n)}$.

We can now investigate $\mathbb{E}\hat{\theta}, V\hat{\theta}$ etc.

1.3 Method of Moments

1.3.1 Handout

Let X_1, \dots, X_n be i.i.d. If one is to estimate a single parameter θ , then we can consider the relation

$$\bar{X} = \mathbb{E}_\theta(X_1) =: g(\theta)$$

as an equation where the parameter θ is the unknown. If the parameter is multivariate $\theta \in \Theta \subseteq \mathbb{R}^p$ then one can set up a system of equations

$$\frac{1}{n} \sum_{i=1}^n X_i^j = \mathbb{E}_\theta[X_1^j] \quad j = 1, \dots, p$$

and solve this for the elements of θ .

The resulting estimator is the method of moments estimators.

Example 1.2. Let $X_1, \dots, X_n \sim U(0, \theta)$ be i.i.d. Then $\mathbb{E}X_i = \frac{\theta}{2}$ and the method of moments estimator solves the equation $\bar{X} = \frac{\theta}{2}$ for θ as the unknown, i.e. $\hat{\theta} = 2\bar{X}$.

Example 1.3. Let $X_1, \dots, X_n \sim \text{Gamma}(\alpha, \beta)$ be i.i.d with density:

$$f(x|\alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\alpha x}, x \geq 0.$$

The first two moments then are:

$$\begin{aligned} \mu_1 &= \bar{X} = E(X^1) \\ &= \frac{\alpha}{\beta} \end{aligned}$$

$$\begin{aligned} \mu_2 &= \bar{X}^2 = E(X^2) \\ &= \text{Var}(X) + (E(X))^2 \\ &= \frac{\alpha}{\beta^2} + \left(\frac{\alpha}{\beta}\right)^2 \\ &= \frac{\alpha(\alpha+1)}{\beta^2} \end{aligned}$$

From the first equation we have:

$$\beta = \frac{\alpha}{\mu_1}$$

Substituting this into the second equation gives,

$$\begin{aligned}\mu_2 &= \frac{\alpha(\alpha+1)}{\left(\frac{\alpha}{\mu_1}\right)^2} \\ \mu_2 &= \frac{(\alpha+1)\mu_1^2}{\alpha} \\ \frac{\mu_2}{\mu_1^2} &= \frac{(\alpha+1)}{\alpha} \\ \alpha \frac{\mu_2}{\mu_1^2} - \alpha &= 1 \\ \alpha \left(\frac{\mu_2}{\mu_1^2} - 1 \right) &= 1 \\ \alpha &= \frac{\mu_1^2}{\mu_2 - \mu_1^2}\end{aligned}$$

Then

$$\begin{aligned}\beta &= \frac{\mu_1^2}{\mu_2 - \mu_1^2} \frac{1}{\mu_1} \\ &= \frac{\mu_1}{\mu_2 - \mu_1^2}\end{aligned}$$

So the method of moments estimators are,

$$\begin{aligned}\hat{\beta} &= \frac{\bar{X}}{\hat{\sigma}^2} \\ \hat{\alpha} &= \frac{\bar{X}^2}{\hat{\sigma}^2}\end{aligned}$$

1.4 Comparing estimators

1.4.1 Handout

Example 1.4. Compare the estimators $\hat{\theta}_1 = X_{(n)}$ and $\hat{\theta}_2 = 2\bar{X}$ for θ in $U(0, \theta)$.

- $\mathbb{E}[\hat{\theta}_1] = \mathbb{E}X_{(n)}$. The c.d.f. of X_i is

$$P[X_i \leq t] = \int_0^t \frac{1}{\theta} dt = \frac{t}{\theta} \quad \text{if } 0 \leq t \leq \theta$$

The c.d.f. of $X_{(n)}$ is

$$\begin{aligned}F(t) &= P[X_{(n)} \leq t] = P[X_1 \leq t, \dots, X_n \leq t] = P[X_1 \leq t] \cdots P[X_n \leq t] \\ &= \left(\frac{t}{\theta}\right)^n \quad \text{if } 0 \leq t \leq \theta\end{aligned}$$

and the p.d.f. is

$$f(t) = \begin{cases} \frac{nt^{n-1}}{\theta^n} & 0 \leq t \leq \theta \\ 0 & \text{e.w.} \end{cases}$$

so the expected value is

$$\mathbb{E}X_{(n)} = \int_0^\theta t \frac{n}{\theta^n} t^{n-1} dt = \frac{n}{\theta^n} \frac{1}{n+1} t^{n+1} \Big|_0^\theta = \frac{n}{n+1} \frac{\theta^{n+1}}{\theta^n} = \frac{n}{n+1} \theta.$$

Hence $\hat{\theta}_1 = X_{(n)}$ is biased, i.e. expected value $\neq \theta$.

- $\mathbb{E}\hat{\theta}_2 = \mathbb{E}[2\bar{X}] = 2\mathbb{E}\bar{X} = 2\frac{\theta}{2} = \theta.$

Note that

$$\mathbb{E}X_{(n)}^2 = \int_0^\theta t^2 \frac{n}{\theta^n} t^{n-1} dt = \frac{n}{\theta^n} \int_0^\theta t^{n+1} dt = \frac{n}{\theta^n(n+2)} t^{n+2} \Big|_0^\theta = \frac{n}{n+2} \theta^2$$

which gives

$$VX_{(n)} = \frac{n}{n+2} \theta^2 - \left(\frac{n}{n+1} \theta \right)^2 = \theta^2 \left\{ \frac{n}{n+2} - \frac{n^2}{(n+1)^2} \right\} = \theta^2 \frac{n}{(n+1)^2(n+2)}.$$

On the other hand

$$V\hat{\theta}_2 = V[2\bar{X}] = 4V[\bar{X}] = \frac{4}{n} V[X_1] = \frac{4}{n} \frac{1}{12} \theta^2 = \frac{\theta^2}{3n}.$$

So the unbiased estimator $\hat{\theta}_2$ is better i.e. has lower variance.

Example 1.5. In the $U(0, \theta)$ example one can consider $\hat{\theta}_3 = \frac{n+1}{n} X_{(n)}$ which satisfies $\mathbb{E}\hat{\theta}_3 = \theta$.

Let $X_n \sim n(\mu, \sigma^2)$ be independent.

Define

$$S^2 := \frac{1}{n-1} \sum (X_i - \bar{X})^2$$

so that

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

and thus $E[S^2] = \sigma^2$ i.e. $b_{\sigma^2}(S^2) = 0$ and

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

so that

$$MSE(S^2) = \frac{2\sigma^4}{n-1} + O^2$$

(see definition of MSE, mean squared error, below).

On the other hand, for

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

we obtain

$$E[\hat{\sigma}^2] = \frac{n-1}{n}\sigma^2$$

and

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

so that

$$b_{\sigma^2}(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2 - \sigma^2 = \frac{-\sigma^2}{n}$$

and therefore

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

so $MSE(\hat{\sigma}^2) < MSE(S^2)$.

Note: As a result of the above, it is of general interest to compare the existing estimators of variance, which only differ in multipliers, using $\frac{1}{n-1}$ or $\frac{1}{n+1}$.

1.5 Method of Least Squares

1.5.1 Handout

The method of least squares method gives the same result as maximum likelihood when the data are assumed to come from a normal distribution. Naturally, this is not generally the case.

The method of least squares can be used as a method of estimation even though the normal distribution is not applicable. The method is then just used to get an estimator, which may or may not be a good estimator compared to the MLE.

1.6 Linear Estimators

□

1.6.1 Handout

Linear estimators are estimators of the form $\sum_i a_i X_i$.

The coefficients a_i can be chosen to satisfy arbitrary requirements. Most commonly this is unbiasedness and minimum variance, leading to the Best Linear Unbiased Estimators (BLUE).

1.7 Minimum Chi-Squared

1.7.1 Handout

When the data (o_{ij}) are available as frequency tables it may be natural to look at a model of the expected frequencies e_{ij} as a function of parameters and then predict the measurements with the expectations e_{ij} . A common measure of quality is

$$X^2 = \sum_{i,j} \frac{(o_{ij} - e_{ij})^2}{e_{ij}}$$

and if there are unknown parameters in the model, they can be estimated by minimizing X^2 .

1.8 Induced Likelihood Function

1.8.1 Handout

Suppose that L_x is a likelihood function. We are interested in evaluating a function of parameter, i.e. evaluate $\tau(\theta)$ but not necessarily θ .

Induced likelihood function for $y = \tau(\theta)$ is the function L_x^* with

$$L_x^*(\eta) := \sup_{\{\theta: \tau(\theta)=\eta\}} L_x(\theta)$$

Theorem 1.1 If $\hat{\theta}$ is the MLE for θ then $\tau(\hat{\theta})$ is the MLE for $\eta = \tau(\theta)$.

Proof. Let $\hat{\eta}$ denote the value that maximizes $L^*(\eta|\mathbf{x})$. We must show that $L^*(\hat{\eta}|\mathbf{x}) = L^*[\tau(\hat{\theta}|\mathbf{x})]$. Now, as stated above, the maxima of L and L^* coincide, so we have:

$$L^*(\hat{\eta}|\mathbf{x}) = \sup_{\eta} \sup_{\{\theta: \tau(\theta)=\eta\}} L(\theta|\mathbf{x}) = \sup_{\theta} L(\theta|\mathbf{x}) = L(\hat{\theta}|\mathbf{x})$$

where the second equality follows because the iterated maximization is equal to the unconditional maximization over θ , which is attained at $\hat{\theta}$. Furthermore

$$L(\hat{\theta}|\mathbf{x}) = \sup_{\{\theta: \tau(\theta)=\tau(\hat{\theta})\}} L(\theta|\mathbf{x}) = L^*[\tau(\hat{\theta})|\mathbf{x}].$$

Hence, the string of equalities shows that $L^*(\hat{\eta}|\mathbf{x}) = L^*(\tau(\hat{\theta})|\mathbf{x})$ and that $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$. \square

Copyright 2019, Gunnar Stefánsson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

2 The quality of estimators

2.1 Quality of estimators

2.1.1 Handout

Let W be an estimator for a parameter θ .
We define the **mean squared error** with

$$MSE := \mathbb{E}[(W - \theta)^2].$$

The **bias** is

$$b_{\theta}(W) := \mathbb{E}[W] - \theta$$

and we note that

$$\begin{aligned} MSE &= E[(W - \theta)^2] = E[(W - EW + EW - \theta)^2] \\ &= E[(W - EW)^2] + E[\underbrace{(EW - \theta)^2}_{b_{\theta}(W)}] + 2E[(W - EW)(EW - \theta)] \\ &= V[W] + b_{\theta}(w)^2 \end{aligned}$$

2.2 Best estimators (UMVUE)

2.2.1 Handout

Definition 2.1. W is the *best unbiased estimator* or the *minimum variance unbiased estimator* (MINVUE) or the *uniformly minimum variance unbiased estimator* (UMVUE) for $\tau(\theta)$ if $E_{\theta}[W] = \tau(\theta)$ and for all other estimators W^* with $E_{\theta}[W^*] = \tau(\theta)$ we have $V_{\theta}[W] \leq V_{\theta}[W^*]$.

Example 2.1. $X_1, \dots, X_n \sim p(\lambda)$. Then we know

$$\mu : EX_i = \lambda$$

and

$$\sigma^2 : VX_i = \lambda$$

therefore

$$E_{\lambda} \bar{X} = \mu = \lambda$$

so that

$$E_{\lambda} S^2 = \sigma^2 = \lambda$$

and we thus have two unbiased estimators.

The question is, which one should be used and obviously one should compare $V_{\lambda} \bar{X}$ vs $V_{\lambda} S^2$? Or can one find $a, 0 \leq a \leq 1$ s.t.

$$V_{\lambda}[a\bar{X} + (1-a)S^2]$$

improves both?

Note: $V_\lambda[a\bar{X} + (1-a)S^2] = a^2V_\lambda\bar{X} + (1-a)^2V_\lambda S^2 + 2a(1-a)\text{Cov}_\lambda(\bar{X}, S^2)$

2.3 Best linear unbiased estimators (BLUE)

□

2.3.1 Handout

Certain estimators can be derived from scratch using a definition of optimality.

If Y_1, \dots, Y_n as independent random variables one can consider estimators of the form

$$W = \sum_{i=1}^n a_i Y_i$$

and choose the coefficients $(a_1^*, \dots, a_n^*) =: \underline{a}^*$ so that

$$E \sum a_i^* Y_i = \tau(\theta)$$

$$V \sum a_i^* Y_i = \min_{\underline{a}} V \sum a_i Y_i$$

Example 2.2. $Y_1, \dots, Y_n \sim n(\mu, \sigma^2)$ iid $\tau(\theta) = \mu$

$$W = \sum a_i Y_i$$

$$EW = \mu = E \sum a_i \bar{Y}_i = \mu$$

$$\Rightarrow \sum a_i \mu = \mu$$

$$\stackrel{(***)}{\Rightarrow} \sum a_i = 1 \quad (*)$$

$$VW \stackrel{(**)}{=} \sum a_i \sigma^2$$

We thus want

$$\min_{a_1, \dots, a_n} \sum a_i^2$$

$$m.t.t \sum a_i = 1$$

$$L = \sum a_i^2 + \lambda(\sum a_i - 1)$$

$$0 = \frac{\partial}{\partial a_i} L = 2a_i + \lambda \Rightarrow a_i = \frac{-\lambda}{2}$$

i.e. all the a_i are the same and $(*)$ implies $a_i = \frac{1}{n}$ and hence \bar{Y} is the BLUE for $n(\mu, \sigma^2)$.

Note: We assumed independence in $(**)$, and identical distributions in $(***)$ but not normality, and hence \bar{Y} is BLUE for μ if Y_1, \dots, Y_n are i.i.d. with expected value μ and a common finite variance σ^2 .

Copyright 2019, Gunnar Stefánsson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

3 The Cramer-Rao inequality

3.1 The Cramer-Rao inequality

3.1.1 Handout

Theorem 3.1 (The Cramer-Rao inequality) Assume that $\underline{X} = (X_1, \dots, X_n)' \sim f_\theta$ where f_θ is a density function and that the random variable $W(\underline{X})$ is such that

$$\frac{d}{d\theta} E_\theta[W(\underline{X})] = \int_{\mathcal{X}(\Omega)} \frac{\partial}{\partial \theta} W(\underline{x}) f_\theta(\underline{x}) d\underline{x} \quad (*)$$

and that $V_\theta[W(\underline{X})] < \infty$.

Then

$$V_\theta[W(\underline{X})] \geq \frac{\left(\frac{d}{d\theta} E_\theta[W(\underline{X})]\right)^2}{E_\theta \left[\left\{ \frac{\partial}{\partial \theta} \ln f_\theta(\underline{X}) \right\}^2 \right]}$$

Note: It is worth noting that

- (1) the condition (*) is quite useless but can be shown to hold for very many distributions, including the exponential family
- (2) The denominator contains the phenomenon

$$\ln f_\theta(\underline{X})$$

which is a function of parameters and random variables.

- (3) If $W(\underline{X})$ is an unbiased estimator for θ then $E_\theta[W(\underline{X})] = \theta$ and the numerator is then the constant 1.
- (4) If $W(\underline{X})$ unbiased and achieves these lower bounds, then W is UMVUE.
- (5) $E_\theta \frac{\partial}{\partial \theta} \ln f_\theta(\underline{X}) = 0$ since

$$\begin{aligned} E_\theta \frac{\partial}{\partial \theta} \ln f_\theta(\underline{X}) &= \int \left(\frac{\partial}{\partial \theta} \ln f_\theta(\underline{x}) \right) f_\theta(\underline{x}) d\underline{x} \\ &= \int \frac{\frac{\partial}{\partial \theta} f_\theta(\underline{x})}{f_\theta(\underline{x})} f_\theta(\underline{x}) d\underline{x} \\ &= \int \frac{\partial}{\partial \theta} f_\theta(\underline{x}) d\underline{x} = \frac{\partial}{\partial \theta} \underbrace{\int f_\theta(\underline{x}) d\underline{x}}_{=1} = 0, \end{aligned}$$

where the second to last step is only valid if the condition (*) is fulfilled.

Proof.

$$\begin{aligned}
\frac{d}{d\theta} E_{\theta}[W(\underline{X})] &= \int W(\underline{x}) \frac{\partial}{\partial \theta} f_{\theta}(\underline{x}) d\underline{x} \\
&= \int W(\underline{x}) \frac{\frac{\partial}{\partial \theta} f_{\theta}(\underline{x})}{f_{\theta}(\underline{x})} f_{\theta}(\underline{x}) d\underline{x} \\
&= \int W(\underline{x}) \left[\frac{\partial}{\partial \theta} \ln f_{\theta}(\underline{x}) \right] f_{\theta}(\underline{x}) d\underline{x} \\
&= E_{\theta} \left[\underbrace{W(\underline{X})}_W \underbrace{\frac{\partial}{\partial \theta} \ln f_{\theta}(\underline{X})}_{U_{\theta}} \right] = E_{\theta}[WU_{\theta}] \\
&= E_{\theta}[WU_{\theta}] - E_{\theta}W \underbrace{E_{\theta}U_{\theta}}_0 = \text{Cov}_{\theta}(W, U_{\theta})
\end{aligned}$$

We also have $V_{\theta}[U_{\theta}] = E[U_{\theta}^2] - \underbrace{(EU_{\theta})^2}_0$ and thus

$$\begin{aligned}
1 \geq \rho_{W, U_{\theta}}^2 &= \frac{\text{Cov}_{\theta}(W, U_{\theta})}{V_{\theta}[W] \cdot V_{\theta}[U_{\theta}]} \\
&= \frac{(E_{\theta}[WU_{\theta}])^2}{V_{\theta}[W] \cdot E[U_{\theta}^2]}
\end{aligned}$$

$$\begin{aligned}
\Rightarrow V_{\theta}[W] &\geq \frac{(E_{\theta}[WU_{\theta}])^2}{E[U_{\theta}^2]} = \frac{(\frac{d}{d\theta} E_{\theta}[W])^2}{E[U_{\theta}^2]} \\
\Rightarrow V_{\theta}[W] &\geq \frac{(\frac{d}{d\theta} E_{\theta}[W])^2}{E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \ln f_{\theta}(\underline{X}) \right)^2 \right]}
\end{aligned}$$

□

3.2 A version for i.i.d. random variables

3.2.1 Handout

Note: If $X_1, \dots, X_n \sim f_{\theta}$ are iid then the C-R theorem becomes

$$V_{\theta}[W(X)] \geq \frac{\left(\frac{d}{d\theta} E_{\theta}[W(X)] \right)^2}{n E_{\theta} \left[\left(\frac{d}{d\theta} \ln f_{\theta}(X) \right)^2 \right]}$$

Since:

$$\tilde{f}_{\theta}(x_1, \dots, x_n) = \prod_{i=1}^n f_{\theta}(x_i)$$

$$\ln \tilde{f}_{\theta}(\underline{x}) = \sum_{i=1}^n \ln f_{\theta}(x_i)$$

$$\ln \tilde{f}_{\theta}(\underline{X}) = \sum_{i=1}^n \ln f_{\theta}(X_i)$$

and

$$E_{\theta} \left[\left(\frac{d}{d\theta} \ln \tilde{f}_{\theta}(X) \right)^2 \right] = \sum_{i=1}^n E_{\theta} \left[\left(\frac{d}{d\theta} \ln f_{\theta}(X_i) \right)^2 \right]$$

3.3 Fisher information

3.3.1 Handout

Note: $E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \ln f_{\theta}(\underline{X}) \right)^2 \right]$

This is called the *Fisher information*. It is a way of measuring how much information an observable random variable, \underline{X} , carries about an unknown parameter, θ .

3.4 Rewriting the Fisher information

3.4.1 Handout

Note: If f_{θ} is the joint pdf of \underline{X} and is such that the order of differentiation and integration can be interchanged, that is

$$\frac{\partial}{\partial \theta} E_{\theta} \left[\frac{\partial}{\partial \theta} \ln f_{\theta}(\underline{X}) \right] = \int_{\theta \in \Theta} \frac{\partial}{\partial \theta} \left[\left(\frac{\partial}{\partial \theta} \ln f_{\theta}(\underline{x}) \right) f_{\theta}(\underline{x}) \right] dx$$

then

$$E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \ln f_{\theta}(\underline{X}) \right)^2 \right] = -E_{\theta} \left[\frac{\partial^2}{\partial \theta^2} \ln f_{\theta}(\underline{X}) \right]$$

3.5 The C-R inequality for i.i.d. random variables

3.5.1 Handout

Corollary 3.1 If X_1, \dots, X_n iid each with pdf f_{θ} . Then under the same assumptions,

$$V_{\theta}[W] \geq \frac{\left(\frac{d}{d\theta} E_{\theta}[W] \right)^2}{n E_{\theta} \left[\left(\frac{\partial}{\partial \theta} \ln f_{\theta}(\underline{X}) \right)^2 \right]}$$

Note: If X_1, \dots, X_n are iid, each with pdf f_{θ} and W is unbiased for θ , then we obtain

$$V_{\theta} W \geq \frac{1}{-n E_{\theta} \left[\frac{\partial^2}{\partial \theta^2} f_{\theta}(\underline{X}) \right]}$$

if the corresponding assumptions hold.

3.6 When the assumptions fail

3.6.1 Handout

Note: If $A_{\theta} = \{ \underline{x} : f_{\theta}(\underline{x}) > 0 \}$ then one usually requires $A_{\theta} = A_{\theta'}$, for all $\theta, \theta' \in \Theta$. I.e. this does not work for $U(0, \theta)$.

3.7 A corollary using the likelihood function

3.7.1 Handout

Corollary 3.2 Let $X_1, \dots, X_n \sim f_\theta$ iid where f_θ satisfies the condition of the C-R theorem and write $L_{\underline{x}}(\theta) = \prod_{i=1}^n f_\theta(x_i)$. An unbiased estimator $W(X)$ of $\tau(\theta)$ attains the C-R lower bound if and only if there is a function a such that

$$a(\underline{x}) = \frac{\partial}{\partial \theta} \ln L_{\underline{x}}(\theta)$$

Proof. Easy application of the Cauchy-Schwartz inequality. □

Example: Write up from 2015-10-01 14.05.54.jpg and 2015-10-01 14.09.23.jpg
This is an old version...

Example 3.1. Let $X_n \sim n(\mu, \sigma^2)$ be independently independent and $\theta = (\mu, \sigma^2)$. We then have:

$$L_{\mathbf{x}}(\theta) = L_{\mathbf{x}}(\mu, \sigma^2) = \prod_{i=1}^n f(x_i; \mu, \sigma^2) = \sigma^{2 \cdot (-n/2)} (2\pi)^{-n/2} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2}$$

So computing the logarithm and differentiating gives

$$\ln L_{\mathbf{x}}(\theta) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2$$

$$\frac{\delta}{\delta \sigma^2} \ln L_{\mathbf{x}}(\theta) = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum (x_i - \mu)^2$$

$$\frac{\delta^2}{\delta (\sigma^2)^2} \ln L_{\mathbf{x}}(\mu, \sigma^2) = \frac{n}{2\sigma^4} - \frac{1}{\sigma^6} \sum (x_i - \mu)^2$$

so the Fisher information is

$$-E\left[\frac{\delta^2}{\delta \theta_2^2} \ln f_\theta(\mathbf{x})\right] = -\frac{n}{2\sigma^4} + \frac{1}{\sigma^6} E\left[\sum_{i=1}^n (x_i - \mu)^2\right] = -\frac{n}{2\sigma^4} + \frac{n\sigma^2}{\sigma^6} = \frac{n}{2\sigma^4}$$

So if W is such that

$$EW = \sigma^2 \Rightarrow V[W] \geq \frac{2\sigma^4}{n}$$

Then to obtain the lower bound of the Cramer-Rao inequality we need

$$a(\theta)[W(\mathbf{x}) - \tau(\theta)] = \frac{\delta}{\delta \theta_2} l_{\mathbf{x}}(\theta)$$

which means $W(\mathbf{x})$ would have to be a function of $\sum (x_i - \mu)$ which is not possible since μ not known.

Copyright 2019, Gunnar Stefánsson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

4 Sufficiency and unbiased estimators

4.1 Background

4.1.1 Handout

Recall that if f is the joint pdf of X and Y , then

$$f_{X|Y}(x|y) := \frac{f(x,y)}{f_Y(y)}$$

where $f_Y(y) = \int f(x,y)dx$. And if

$$t(y) := E[X|Y = y] = \int xf_{X|Y}(x|y)dx$$

and we define $t(Y) := E[X|Y]$, then

$$\begin{aligned} E[E[X|Y]] &= \int_y t(y)f_Y(y)dy \\ &= \int_y \left(\int_x f_{X|Y}(x|y)dx \right) f_Y(y)dy \\ &= \iint xf(x,y)dxdy = E[X] \end{aligned}$$

Similarly we can show that

$$V[X] = V[E[X|Y]] + E[V[X|Y]]$$

4.2 The Rao-Blackwell theorem

4.2.1 Handout

Theorem 4.1 (Rao-Blackwell) Let W be any unbiased estimator of $\tau(\theta)$ and T be a sufficient statistic. Define $\phi(T) := E[W|T]$.

Then we have

$$E_{\theta}[\phi(T)] = \tau(\theta)$$

and

$$V_{\theta}[\phi(T)] \leq V_{\theta}[W], \forall \theta$$

Proof. From

$$E[X] = E[E(X|Y)], \quad \text{and} \quad V[X] = V[E(X|Y)] + E[V(X|Y)]$$

we have that

$$\tau(\theta) = E_{\theta}[W] = E_{\theta}[E(W|T)] = E_{\theta}[\phi(T)]$$

and so $\phi(T)$ is unbiased for $\tau(\theta)$. Furthermore, we have that

$$\begin{aligned} V_{\theta}[W] &= V_{\theta}[E(W|T)] + E_{\theta}[V(W|T)] \\ &= V_{\theta}[\phi(T)] + E_{\theta}[V(W|T)] \\ &\geq V_{\theta}[\phi(T)], \end{aligned}$$

where the last inequality comes from $V(W|T) \geq 0$.

Hence $\phi(T)$ is uniformly better than W and all that remains is to show that $\phi(T)$ is an estimator. That is, to show that $\phi(T) = E(W|T)$ is a function of only the sample and it is independent of θ . From the definition of sufficiency, and the fact that W is a function of only the sample, we get that the distribution of $W|T$ is independent of θ .

Therefore, $\phi(T)$ is a uniformly better unbiased estimator of $\tau(\theta)$. \square

4.3 Lehmann–Scheffé

4.3.1 Handout

add words and change this to a note...

Note 4.1. $V[U] = V[W]$ if and only if $P[U = W] = 1, \forall \theta$ Sometimes $E_\theta[T] = a + b\theta$ and then we get $U := \frac{t-a}{b}$

Note 4.2. $E[Y] = \theta, S := E_\theta[Y|U], U$ is not necessarily adequate.

Theorem 4.2 (Lehmann–Scheffé) *Let T be a complete and sufficient statistic for a parameter θ and $\phi(T)$ be any estimator based only on T . Then $\phi(T)$ is the unique best unbiased estimator of its expected value $\tau(\theta)$.*

Sönnun. By **Rao-Blackwell**: If R is any unbiased estimator of the parameter θ then:

$$\phi(T) = E[R|T]$$

is an unbiased estimate of θ such that:

$$\text{Var}[E(R|T)] \leq \text{Var}[R]$$

Then let S be any other unbiased estimator and

$$\psi(T) = E[S|T]$$

then

$$E_\theta[\phi(T) - \psi(T)] = 0$$

and because T is complete it follows that

$$P_\theta(\phi(T) = \psi(T)) = 1$$

\square

So $\phi(T)$ is the unique best unbiased estimator.

Example 4.1. Let $X_1 \dots X_n \sim \text{bin}(k, \theta)$ be iid, $Y := \sum_1^n X_i \sim \text{bin}(kn, \theta)$

1. Since X_i are binomial we get:

$$\begin{aligned} P[X_1 = x_1, \dots, X_n = x_n] &= \binom{k}{x_1} \dots \binom{k}{x_n} \theta^{x_1} (1-\theta)^{k-x_1} \dots \theta^{x_n} (1-\theta)^{k-x_n} \\ &= \binom{k}{x_1} \dots \binom{k}{x_n} \theta^{\sum x_i} (1-\theta)^{kn - \sum x_i} \end{aligned}$$

so we see that Y is sufficient for θ .

2. Now let $E_\theta[g(Y)] = 0$ for all θ and show that Y is complete.

$$E_\theta[g(Y)] = \sum_{y=0}^k ng(y) \binom{kn}{y} \theta^y (1-\theta)^{kn-y} = (1-\theta)^{kn} \sum_{y=0}^k ng(y) \binom{kn}{y} \left(\frac{\theta}{1-\theta}\right)^y$$

We observe that if $\theta \in \{0, 1\}$ the expected value of $g(Y)$ is trivially 0. Now if $0 < \theta < 1$ then $E_\theta[g(Y)] = 0$ if and only if

$$\sum_{y=0}^k ng(y) \binom{kn}{y} \left(\frac{\theta}{1-\theta}\right)^y = 0$$

But since this is a polynomial it is 0 only if every coefficient is 0, that is only if $g(y) = 0$ for all y . Therefore we conclude that Y is a complete and sufficient statistic.

3. Note that: $P[X_1 = 1] = k\theta(1-\theta)^{k-1} =: \tau(\theta)$. And therefore

$$W := \begin{cases} 1 & X_1 = 1 \\ 0 & \text{annars} \end{cases}$$

is an unbiased estimator for $\tau(\theta)$ since

$$E_\theta[W] = \sum_{w=0}^1 wP_\theta(W = w) = P_\theta(X_1 = 1) = \tau(\theta)$$

4. Finally, since Y is a complete and sufficient statistic and W is an unbiased estimator, we simply define $\phi(Y) := E[W|Y]$ to get the unique best unbiased estimator.

Copyright 2019, Gunnar Stefánsson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.

5 Overview of point estimation

5.1 Summary

Main points

Methods: MLE, m.o.m., BLUE, $\min \chi^2, \dots$

Quality criteria: bias, variance, MSE

Special attention: MINVUE (UMVUE)

Cramer-Rao (lower bd on variance)

Fisher information

Rao-Blackwell (condition on suff. statistic)

Copyright 2019, Gunnar Stefánsson

This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.