4 Classical estimation theory

4.1 Cramer-Rao inequality

Obtaining a point estimator of the parameter of interest is only the first step in inference.

Suppose $\boldsymbol{X} = (X_1, X_2, .., X_n)$ are i.i.d. from $f(x, \theta), \theta \in \mathbb{R}$ and we use a statistic $T_n(\boldsymbol{X})$ to estimate θ.

If $E_\theta(T_n) = \theta + b_n(\theta)$, then the quantity

$$b_n(\theta) = E_\theta(T_n) - \theta$$

is called bias. Here and further, the notation $E_\theta(.)$ is used to denote the expected value when θ is the parameter of the distribution in the sample.

$E_\theta(X) = \int x f(x, \theta)\, dx$

The bias generally depends on both θ and the sample size n although this dependence on the sample size may sometimes be suppressed in the notation.

We call an estimator unbiased if $b_n(\theta) \equiv 0$ for all $\theta \in \Theta$.

Interpretation of unbiasedness: when used repeatedly, an unbiased estimator, in the long run, will estimate the true value on average. This is why, in general, it is good to have estimators with zero bias for any $\theta \in \Theta$ and any sample size n.

Remark 4.8

Caution: for some families an unbiased estimators may not exist or, even when they exist, may not be very useful.

Example 4.29 (at lecture)

Geometric distribution

$$f(x,\theta) = \theta(1-\theta)^{x-1}, x = 1, 2, \ldots$$

an unbiased estimator of θ, say, $T(x)$ must satisfy

$$E_\theta[T(x)] = \sum_{x=1}^{\infty} T(x)\theta(1-\theta)^{x-1} = \theta$$

for all $\theta \in [0, 1]$. By a polynomial expansion, the estimator would satisfy

$$T(1) = 1, \qquad T(x) = 0 \quad \text{if} \quad x \geq 2.$$

Having in mind the interpretation of θ, such an estimator is neither reliable, nor useful.

Unbiasedness cannot be the only requirement for an estimator to be good. Note also that there may be many unbiased estimators for the same parameter and then we will be confronted with the problem to "select the best" amongst them.

A quantity to look at when comparing quality of estimators, is the mean squared error:

$$MSE_\theta(T_n) = E_\theta\Big[(T_n - \theta)^2\Big].$$

The following property holds:

$$MSE_\theta(T_n) = Var_\theta(T_n) + (b_n(\theta))^2$$

$b_n(\theta) = 0$

Proof:

$$\begin{aligned}
&MSE_\theta(T_n) \\
&\quad = \mathbb{E}_\theta\Big[(T_n - \mathbb{E}_\theta(T_n) + \mathbb{E}_\theta(T_n) - \theta)^2\Big] \\
&\quad = \mathbb{E}_\theta\Big[(T_n - \mathbb{E}_\theta(T_n))^2\Big] + \mathbb{E}_\theta\Big[(\mathbb{E}_\theta(T_n) - \theta)^2\Big] \\
&\qquad\quad - 2\mathbb{E}_\theta\Big[(T_n - \mathbb{E}_\theta(T_n))(\theta - \mathbb{E}_\theta(T_n))\Big] \\
&\quad = Var(T_n) + (b_n(\theta))^2
\end{aligned}$$

Keeping the MSE as small as possible is more important than just to ask for unbiasedness. We would like to find an estimator that minimises the MSE.

Unfortunately, in the class of *all* estimators, an estimator that minimizes the MSE simultaneously *for all* θ *values* does not exist!

To see why, lets take any estimator $\tilde{\theta}$. Since the parameter $\theta \in \Theta$ is unknown, there will be a certain value $\theta_0 \in \Theta$ for which $MSE_{\theta_0}(\tilde{\theta}) > 0$.

Then we can consider as a competitor to $\tilde{\theta}$ the estimator $\theta^* \equiv \theta_0$. Note that θ^* is not a very reasonable estimator (it does not even use the data!) but for the particular point θ_0 we have $MSE_{\theta_0}(\theta^*) = 0$ and hence,

$\theta_0 \in \Theta$

$$MSE_{\theta_0}(\tilde{\theta}) > MSE_{\theta_0}(\theta^*).$$

In other words, when considering the class of all estimators, there are so many estimators available to us that to find one that is uniformly better with respect to the MSE criterion, is just impossible. There are two ways out of this unpleasant situation:

- We can restrict the class of estimators considered.
- We could change the evaluation criterion when trying to define the best. Minimax, Bayes (defn of risk)

We deal with the first way out now (the other way was discussed in the Decision theory chapter: Bayes and minimax estimation).

The most common way to restrict the class of estimators is to impose the criterion of unbiasedness. This means to consider only the unbiased estimators when searching for the best one with the smallest MSE.

This simplifies, to a great extent, the task of minimising the MSE because then we only have to minimise the variance! In the smaller subset of unbiased estimators, one can very often find an estimator with the smallest MSE (variance) *for all θ values*.

It is called the uniformly minimum variance unbiased estimator (UMVUE). Let us first look at a well-known result that will help us in our search of the UMVUE (the Cramer-Rao theorem).

θ $\theta(1-\theta)$ $\log\theta$

$\frac{1}{\theta}$

Theorem 4.13

Let $\mathbf{X} = (X_1, X_2, \dots, X_n)$ have a distribution that depends on θ and $L(\mathbf{X}, \theta)$ be the joint density. Let $\tau(\theta)$ be a smooth (i.e. differentiable) function of θ that has to be estimated. Consider any unbiased estimator $W(\mathbf{X})$ of $\tau(\theta)$, i.e.$E_\theta\Big[W(\mathbf{X})\Big] = \tau(\theta)$. Suppose, in addition, that $L(\mathbf{X}, \theta)$ satisfies:

$$\frac{\partial}{\partial\theta}\int..\int h(X)L(X,\theta)\mathrm{d}X_1..\mathrm{d}X_n = \int..\int h(X)\frac{\partial}{\partial\theta}L(X,\theta)\mathrm{d}X_1..\mathrm{d}X_n \qquad (*)$$

for any function $h(\mathbf{X})$ with $E_\theta|h(\mathbf{X})| < \infty$. Then:

$$\mathrm{Var}_\theta(W(\mathbf{X})) \geq \frac{(\frac{\partial}{\partial\theta}\tau(\theta))^2}{I_{\mathbf{X}}(\theta)}$$

for all θ holds.

Proof.

The proof uses a clever application of the Cauchy-Schwartz inequality.

Cauchy-Schwartz Inequality:

If Z and Y are two random variables with finite variances $\mathsf{Var}(Z)$ and $\mathsf{Var}(Y)$ then

$$[\mathsf{Cov}(Z,Y)]^2 \overset{def}{=} \{\mathbb{E}[(Z-\mathbb{E}(Z))(Y-\mathbb{E}(Y))]\}^2 \leq \mathsf{Var}(Z)\mathsf{Var}(Y).$$

unbiased estimator $\tau(\theta)$

To prove the Cramer-Rao Theorem, we choose W to be the Z-variable, and the score V to be the Y-variable in the Cauchy-Schwartz Inequality. Since $\mathbb{E}_\theta\Big[V(\mathbf{X},\theta)\Big] = 0$ holds for the score, we have that

Cauchy-Schwartz

$$\begin{aligned}[\mathsf{Cov}_\theta(W,V)]^2 &= [\mathbb{E}_\theta(WV) - \mathbb{E}_\theta(W)\cancel{\mathbb{E}_\theta(V)}]^2 \\ &= [\mathbb{E}_\theta(WV)]^2 \\ [\mathrm{Cov}(W,V)]^2 &\leq \mathsf{Var}_\theta(W)\, Var_\theta(V) \qquad (6)\end{aligned}$$

0

Substituting the definition of the score, we get:

$$Cov_\theta(W,V) = \mathbb{E}_\theta(WV) = \int \ldots \int W(\mathbf{X}) \frac{\frac{\partial}{\partial\theta} L(\mathbf{X},\theta)}{\cancel{L(\mathbf{X},\theta)}} \cancel{L(\mathbf{X},\theta)} \mathrm{d}\mathbf{X}$$

where $\mathrm{d}\mathbf{X} = \mathrm{d}X_1 \mathrm{d}X_2 \ldots \mathrm{d}X_n$ is used as shorthand notation.

Now if we utilise condition (*), we can continue to get:

$$
\begin{aligned}
Cov_\theta(W, V) &= \frac{\partial}{\partial\theta} \underbrace{\int \dots \int W(\mathbf{X})L(\mathbf{X}, \theta)\mathrm{d}\mathbf{X}}_{E_\theta(W)} \\
&= \frac{\partial}{\partial\theta}\mathbb{E}_\theta(W) \\
&= \frac{\partial}{\partial\theta}\tau(\theta)
\end{aligned}
$$

W is unbiased for $\tau(\theta)$

Then, Inequality (6) implies that:

$$\left(\frac{\partial}{\partial\theta}\tau(\theta)\right)^2 \leq \mathrm{Var}_\theta(W) I_X(\theta)$$

and hence

$$\mathrm{Var}_\theta(W) \geq \frac{\left(\frac{\partial}{\partial\theta}\tau(\theta)\right)^2}{I_{\mathbf{X}}(\theta)}. \tag{7}$$

□

We will skip the discussion on the *multivariate version of this inequality* which is applicable for the case of estimating a multidimensional parameter.

Remark 4.9

The Cramer-Rao (CR) Inequality was stated for continuous random variables. By an obvious modification of condition (*) requiring the ability to interchange differentiation and summation (instead of differentiation and integration) one can formulate this for discrete random variables, too. In this case, even though $L(\mathbf{X}, \theta)$ may not be differentiable with respect to x, it has to be assumed to be differentiable with respect to θ.

4.1.1 Corollary for i.i.d. case.

If $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ are i.i.d. with $f(x,\theta)$ then

$$L(\mathbf{X},\theta) = \prod_{i=1}^{n} f(X_i,\theta) \quad \text{and} \quad I_{\mathbf{X}}(\theta) = nI_{X_1}(\theta)$$

and the CR Inequality becomes:

$$\text{Var}_\theta(W(\mathbf{X})) \geq \frac{(\frac{\partial}{\partial\theta}\tau(\theta))^2}{nI_{X_1}(\theta)}.$$

4.2 Comments on applying the CR Inequality in the search of the UMVUE

a) Since the Inequality gives a lower bound on the variance of an unbiased estimator of the parameter $\tau(\theta)$, it is obvious that, in the case where there exists an unbiased estimator of $\tau(\theta)$ whose variance is equal to the lower bound given by CR Inequality, this will be the UMVUE of $\tau(\theta)$.

Such a situation occurs often in the case of observations that arise from an exponential family.

$\tau(\theta) = \theta$

unbiased

Exercise 4.18 (at lecture)

Calculate the Cramér-Rao lower bound (CRLB) for the variance of an unbiased estimator of θ and find a statistic with variance equal to the bound when $X_1, X_2, \ldots, X_n$ are independent random variables each with a distribution from the exponential family of distributions.

i) Exponential (θ): $f(x,\theta) = \frac{1}{\theta}e^{-x/\theta}, \quad x > 0$ ✓

ii) Bernoulli (θ): $f(x,\theta) = \theta^x(1-\theta)^{1-x}$, $x \in \{0,1\}$, $\theta \in (0,1)$ ✓

iii) $N(\theta,1)$: $f(x,\theta) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-\theta)^2}$, $x \in R$, $\theta \in R$ ✓

iv) $N(0,\theta)$: $f(x,\theta) = \frac{1}{\sqrt{2\pi\theta}}e^{-\frac{x^2}{2\theta}}$, $x \in R$, $\theta > 0$

ii) Note the drawback related to the fact that condition (*) in the CR Theorem is a strong one and it often happens that it is not satisfied. A typical situation is when the range of the random variables $X_i, i = 1, 2, .., n$ depends on θ, for example, in the case of a random sample from uniform $[0, \theta)$ observations.

According to the general Leibnitz′ rule for differentiation of parameter-dependent integrals:

$$\frac{\partial}{\partial\theta}\int_{a(\theta)}^{b(\theta)} f(x,\theta)dx$$

$$= f(b(\theta),\theta)\frac{\partial}{\partial\theta}b(\theta) - f(a(\theta),\theta)\frac{\partial}{\partial\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)}\frac{\partial}{\partial\theta}f(x,\theta)dx$$

holds and we see that, if a and b were genuine functions of θ, on the RHS there would be some additional non-zero terms included and condition (*) would not hold.

Example 4.30 (at lecture)

Assume that the right-hand limit θ of the interval $[0, \theta)$ is to be estimated and n i.i.d. observations from the uniform distribution $[0, \theta)$ are given.

Solution:

We can show that the density of $Y = X_{(n)}$ is given by

$$f_Y(y, \theta) = \begin{cases} ny^{n-1}/\theta^n & \text{if } 0 < y < \theta \\ 0 & \text{else} \end{cases}$$

by noting that

$$F_{X_{(n)}}(y) = P(X_{(n)} < y) = P(X_1 < y \cap X_2 < y \cap \cdots \cap X_n < y) = \left(\frac{y}{\theta}\right)^n$$

for $0 < y < \theta$.

Then we can calculate:

$$\mathbb{E}(X_{(n)}) = \int_0^\theta y \frac{ny^{n-1}}{\theta^n} dy = \frac{n}{n+1}\theta \neq \theta$$

hence $X_{(n)}$ is biased for estimating θ. However,

$$\mathbb{E}\left[\frac{n+1}{n} X_{(n)}\right] = \theta$$

holds, that is, $T = \frac{n+1}{n} X_{(n)}$ is an unbiased estimator of $\tau(\theta) = \theta$. It can be shown that the variance is given by:

$$\text{Var}(T) = \mathbb{E}(T^2) - \theta^2 = \left(\frac{n+1}{n}\right)^2 \int_0^\theta y^2 \frac{ny^{n-1}}{\theta^n} dy - \theta^2 = \frac{\theta^2}{n(n+2)}.$$

The latter value is less than the value $\frac{\theta^2}{n}$ which we would get if we recklessly calculated the CR bound ignoring the fact that the regularity condition (*) is in fact violated in this example (since the support of the density depends on the unknown parameter).

iii) One more important observation is to be made in relation to the usefulness of the CR Inequality in the search of UMVUE. Even in cases where the CR Theorem is applicable, there is no guarantee that the lower bound on the variance is attainable!

There are two essentially different scenarios:

Scenario 1: bound is attainable:

Looking at the proof we see that the bound is achievable iff we have equality in Cauchy-Schwartz Inequality, which means that the score $V(\mathbf{X}, \theta)$ must have a representation of the form

$$V(\mathbf{X}, \theta) = k_n(\theta)[W(\mathbf{X}) - \tau(\theta)].$$

The factor $k_n(\theta)$ in this representation is a proportionality factor, so it does not involve the data. The quantity $W(\mathbf{X})$ is a statistic, i.e., it does not involve the parameter but just the data.

Scenario 2: bound is not attainable:

If $V(\mathbf{X}, \theta)$ cannot be written in the above form, then no unbiased estimator of $\tau(\theta)$ would have a variance equal to the one given by the CR bound and, in this case, the CR Inequality would be of no use when searching for UMVUE.

There still might be an UMVUE (with a variance slightly higher than the value given by the CR lower bound) but one would need to develop a specific method to find it in such situations. Next, we illustrate the two scenarios in an example.

4.3 CRLB attainability examples

$P(X=x) = \frac{e^{-\theta}\theta^{x}}{x!}$

Example 4.31 (at lecture)

i) Estimating the parameter θ in a Poisson(θ) distribution. In this case the CR bound is achievable and the unbiased estimator that achieves it, is $\hat{\theta} = \bar{X}$.

ii) Estimating the function $\tau(\theta) = \exp(-\theta)$ from a sample of Poisson(θ) distribution. In this case no unbiased estimator of $\tau(\theta)$ has variance equal to the bound. Nevertheless, UMVUE of $\tau(\theta)$ exists.

$n\bar{X} = n \times \frac{1}{n} \Sigma x_i = \Sigma x_i$

Solution:

Estimating the two parameters $\tau_1(\theta) = \theta$ and $\tau_2(\theta) = \exp(-\theta)$ from a sample of Poisson(θ) distribution.

The log-likelihood is:

$$\log L(X;\theta) = \log\left\{\frac{\theta^{\sum_{i=1}^n x_i} e^{(-n\theta)}}{\prod_{i=1}^n x_i!}\right\} = -n\theta + \left(\sum_{i=1}^n x_i\right)\log\theta - \sum_{i=1}^n \log(x_i!)$$

By taking the derivative with respect to θ we get for the score $V(X,\theta)$:

$$V(X,\theta) = \frac{\partial}{\partial\theta}\log L(X;\theta) = -n + \sum_{i=1}^n \frac{x_i}{\theta} = -n + \frac{n\bar{X}}{\theta}$$

$K_n(\theta)\,(W(X) - \tau(\theta))$

$\frac{n}{\theta}\,(\underset{\sim}{\bar{X}} - \theta)$

i) For $\tau_1(\theta) = \theta$ we can factorise the score:

$$V(X, \theta) = \frac{n}{\theta}(\bar{X} - \tau_1(\theta)). \quad \therefore \; K_n(\theta)(\bar{X} - \theta)$$

The factorization is possible and hence the bound is attainable. Here we have a statistic (i.e. a transformation of the data only, no parameter involved). Hence the statistic $\bar{X}$ attains the CR bound and therefore it must be the UMVUE for $\tau_1(\theta) = \theta$.

ii)

$V(x,\theta) = -n + \frac{n\bar{x}}{\theta} = -\frac{ne^{-\theta}}{e^{-\theta}} + \frac{n\bar{x}e^{-\theta}}{\theta e^{-\theta}} -$

However, for $\tau_2(\theta)$, we have:

$$V(X,\theta) = n\exp(\theta)\left[\frac{1}{\theta}\exp(-\theta)\bar{X} - \tau_2(\theta)\right]$$

$= \frac{n}{e^{-\theta}}\left[\underbrace{\frac{\bar{x}e^{-\theta}}{\theta}}_{\text{not a statistic}} - \underbrace{e^{-\theta}}_{\tau(\theta)}\right]$

and this can not be represented as

$$k(\theta,n)[\text{statistic} - \exp(-\theta)]$$

since $\frac{1}{\theta}\exp(-\theta)\bar{x}$ is NOT a statistic as it depends on the parameter θ.

Formal calculation of the Cramér–Rao bound gives $\frac{\theta}{n}e^{(-2\theta)}$ (check it!) but this bound is not attainable by any unbiased estimator of $\tau_2(\theta) = \exp(-\theta)$. Nevertheless we claim that UMVUE does exist and is given by

$$T = \left(1 - \frac{1}{n}\right)^{n\bar{X}}.$$

Another method (based on the Lehmann–Scheffé theorem) needs to be applied to finding this UMVUE. The justification that

$$T = \left(1 - \frac{1}{n}\right)^{n\bar{X}}$$

is UMVUE of $\tau_2(\theta) = \exp(-\theta)$ is given as an example after the formulation of the Lehmann–Scheffé theorem.

4.4 Which are the estimators that could attain the bound?

Theorem 4.14

If under the regularity conditions of CR Theorem there is an estimator of $\tau(\theta)$ which attains the lower bound, it should be the MLE of $\tau(\theta)$.

Proof: at lecture.

Conclusion:

As a consequence of this statement, when looking for UMVUE, it is a good idea to calculate the MLE first. If the MLE turns out to be unbiased and its variance equals the value given by the CR bound, the UMVUE has been constructed.

Otherwise, if either the MLE is biased or does not attain the bound, then it is sure that the bound is not attainable at all. In such situations, the UMVUE (which necessarily will have variance larger than the one given by the bound) usually turns out to be a bias-corrected MLE.

To outline a more specific way to construct UMVUE in such more delicate situations, let us first formulate the following famous theorem:

4.5 Rao-Blackwell theorem

θ $\hat{\theta}$

τ $\hat{\tau}(T)$

Theorem 4.15 (Rao-Blackwell Theorem)

Let W be *any* unbiased estimator of $\tau(\theta)$ and let T be a sufficient statistic for θ. Define

$$\hat{\tau}(T) = E_\theta(W \mid T).$$

Then $E_\theta[\hat{\tau}(T)] = \tau(\theta)$ and $\text{Var}_\theta[\hat{\tau}(T)] \leq \text{Var}_\theta(W)$ for all $\theta \in \Theta$, i.e. $\hat{\tau}(T)$ is uniformly better than W as an estimator of $\tau(\theta)$.

Proof: at lecture.

$\hat{\tau}(T) = E(W \mid T)$

Remark 4.10

Since $\hat{\tau}(T)$ is a conditional expected value, it is a function of the condition T, i.e., $\hat{\tau}(T)$ depends on the data only via the value of the sufficient statistic and has uniformly smaller variance than W. This theorem underlines the importance of sufficient statistics in the search of UMVUE.

It tells us that we only need to consider functions of sufficient statistics when looking for UMVUE, otherwise we can always improve the performance by conditioning on a sufficient statistic.

4.6 Uniqueness of UMVUE

Theorem 4.16

If an estimator W is UMVUE for $\tau(\theta)$, then W is unique.

Proof: at lecture.

Theorem 4.17

The estimator W is UMVUE if and only if W is uncorrelated with all unbiased estimators of zero (i.e., $Cov(W, U) = 0$ must hold for any random variable $U(\mathbf{X})$ with $E(U) = 0$).

Proof: at lecture - some details are given in the next few slides.

Proof.

An unbiased estimator of 0 is like a noise and it would be counter-intuitive to expect it to contribute somehow to further improvement of W.

If W has the property in the theorem, it must be the UMVUE already.

Indeed, assume that W with $E(W) = \tau(\theta)$ has the property and W' is a competitive unbiased estimator of $\tau(\theta)$, i.e. $E(W') = \tau(\theta)$ also holds.

Then $U = W' - W$ is an unbiased estimator of zero and then for any value of θ :

$$\mathrm{Var}(W') = \mathrm{Var}(W + U) = \mathrm{Var}(W) + \mathrm{Var}(U) + 2\mathrm{Cov}(W, U)$$

and, because we have the property $\mathrm{Cov}(W, U) = 0$, we get

$$\mathrm{Var}(W') \geq \mathrm{Var}(W),$$

i.e., no other unbiased estimator of $\tau(\theta)$ can be better than W in terms of variance. $\square$

Therefore, characterisation of the estimators that are uncorrelated with any unbiased estimator of zero turns out to be very important.

If it turned out that the family $f(T, \theta)$ of distributions does have the property that there are no unbiased estimators of zero (except the constant zero itself), then our search of UMVUE will be successfully finalised.

To be able to formulate such a characterisation, we need the notion of completeness.

4.7 Completeness of a family of distributions

$E_\theta[g(T)] = 0 \implies P_\theta(g(T) = 0) = 1$ for all θ

Definition 4.8

Let $\tilde{f}(t, \theta), \theta \in \Theta$ be a family of distributions for a statistic $T(\mathbf{X})$. The family is called complete if $E_\theta\big[g(T)\big] = 0$ for all $\theta \in \Theta$ implies $P_\theta(g(T) = 0) = 1$ for all $\theta \in \Theta$. Equivalently, $T(X)$ is called a complete statistic for θ.

Remark 4.11

Note that completeness is a property of the *whole* family of distributions and *not* only a property of a particular distribution.

Example 4.32

For $\mathbf{X} = (X_1, X_2, \dots, X_n)$ i.i.d. $N(0, \theta)$, $T = \bar{X}$ is not complete for θ. Note that θ denotes the variance in this example.

Solution:

It is suffice to find a counterexample. Take $g(t) = t \neq 0$. We have $E_\theta(g(T)) = E_\theta(\bar{X}) = 0$ for all $\theta > 0$ but $g(t) \neq 0$.

$\bar{X} \sim N(0, \frac{\theta}{n})$ $\quad$ $P(\bar{X} = 0) \neq 1$

Example 4.33

For $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. Bernoulli with $\theta \in (0,1)$ denoting the probability of success, $T(X) = \sum_{i=1}^{n} X_i$ is complete for θ.

Solution:

We know that if $T \sim Bin(n, \theta)$, then:

$$P_\theta(T = t) = \binom{n}{t} \theta^t (1-\theta)^{n-t} \qquad t = 0, 1, 2, \ldots, n$$

Take $E_\theta[g(T)] = 0$ for all $\theta \in (0,1)$, then

$$E[g(T)] = \sum_{t=0}^{n} g(t) \binom{n}{t} \theta^t (1-\theta)^{n-t} = 0.$$

Furthermore,

$$\underbrace{(1-\theta)^n}_{\neq 0} \sum_{t=0}^{n} g(t)\binom{n}{t}\eta^t = 0$$

for all $\eta = \frac{\theta}{1-\theta} \in (0, \infty)$ and $(1-\theta)^n \neq 0$.

Then $g(t)\binom{n}{t} = 0$ must hold and, since $\binom{n}{t} \neq 0$, then $g(t) = 0$ for $t = 0, 1, \ldots, n$ and $P_\theta(g(T) = 0) = 1$.

Hence, T is complete.

$g(t)\cancel{\binom{n}{t}}\cancel{\eta^t} = 0$

$g(t) = 0$

$t = 0, 1, \ldots, n$

Now we finally can formulate the theorem that allows us to find the UMVUE even in situations when the CR bound is not achievable.

4.8 Theorem of Lehmann-Scheffe

Theorem 4.18 (Theorem of Lehmann-Scheffe)

Let T be a complete sufficient statistic for a parameter θ and W be *any* unbiased estimator of $\tau(\theta)$. Then $\hat{\tau}(T) = E_\theta(W \mid T)$ is an *UMVUE* of $\tau(\theta)$. Furthermore, $\hat{\tau}(T)$ is the *unique* UMVUE in the sense that if T^* is any other UMVUE, then

$$P_\theta(\hat{\tau}(T) = T^*) = 1$$

for all θ.

Proof: at lecture.

Proof. (sketch)

This theorem is a direct consequence of the statements we made previously.

Indeed, according to the Rao–Blackwell Theorem, it suffices to consider as candidates for UMVUE only unbiased functions of the sufficient statistic T (otherwise by conditioning T we can improve any other unbiased competitor) and (again due to Rao-Blackwell) $\hat{\tau}(T)$ is such an unbiased estimator of $\tau(\theta)$. $E(\hat{\tau}(T)|T\} = \hat{\tau}(T)$

Because of the completeness, no further improvement of $\hat{\tau}(T)$ is possible, hence it is the unique UMVUE. □

The theorem has many useful applications. In many situations there will be no obvious candidate for the UMVUE of $\tau(\theta)$. But the theorem suggests that if we have found *any* (even if very poor) unbiased estimator of $\tau(\theta)$ and we know a statistic T that is a complete and sufficient statistic for θ then $E_\theta(W \mid T)$ is the uniformly best unbiased estimator of $\tau(\theta)$.

Remark 4.12

It is easy to work with exponential families because their structure directly helps us identify a **minimal sufficient and complete** statistic: once the function $d(x)$ in the definition of the exponential family has been identified, we know that

$$T(X) = \sum_{i=1}^{n} d(X_i)$$

is complete and minimal sufficient.

The Lehmann-Scheffe theorem can be used to construct UMVUE in such families. Hence, for such families, the Lehmann-Scheffé theorem can typically be used to construct UMVUE and you do not need further argumentation to claim that T is complete.

4.9 Examples finding UMVUE using Lehmann-Scheffe theorem

Exercise 4.19

i) For $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. Bernoulli with $\theta \in (0, 1)$ denoting the probability of success, $T(X) = \sum_{i=1}^{n} X_i$ is complete for θ and $\bar{X}$ is UMVUE for the expected value parameter θ. Hovewer, for the variance parameter $\theta(1-\theta)$, the UMVUE turns out to be $\bar{X}(1-\bar{X})\frac{n}{n-1}$.

ii) For $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. uniform in $[0, \theta)$, the statistic $T = X_{(n)}$ is complete and $\frac{n+1}{n}X_{(n)}$ is UMVUE for θ.

iii) For $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ i.i.d. Poisson (θ), the statistic $T(X) = \sum_{i=1}^{n} X_i$ is complete. For $\tau(\theta) = \exp(-\theta)$, the UMVUE turns out to be $(1-\frac{1}{n})^{n\bar{X}}$.

Note that parameters of interest in the above examples such as

$$\tau(\theta) = \theta(1-\theta)$$

for the variance of a Bernoulli distribution,

$$\tau(\theta) = e^{-\theta} = P(X_1 = 0)$$

for the Poisson distribution and

$$\tau(\theta) = \theta e^{-\theta} = P(X_1 = 1)$$

for the Poisson distribution, typically, have some probabilistic interpretation that can be exploited to suggest a (simple) unbiased estimator W, which then can be Rao–Blackwellised to obtain the UMVUE.

Note also that if we are dealing with a one-parameter exponential family

$$f(x,\theta) = a(\theta)b(x)\exp\{c(\theta)d(x)\}$$

then the statistic $T(X) = \sum_{i=1}^{n} d(X_i)$ is complete and minimal sufficient and we do not need to separately check completeness for densities that belong to this family.

Example 4.34 (try your hand at this one)

Find the UMVUE of θ^2 when $X_1, X_2, \dots, X_n$ are independent Bernoulli(θ) random variables. Check that your estimator does in fact have mean θ^2 and is unbiased.

Solution:

Step 1: $T(X) = \sum_{i=1}^n X_i$ is complete and sufficient for θ.

Step 2: If $W = X_1 X_2$ then $E_\theta(W) = \theta^2$, that is, W is unbiased for θ^2.

Step 3: $\hat{\tau}(T) = \mathbb{E}_\theta(W|T=t) = \cdots = \frac{t(t-1)}{n(n-1)}$ which is the UMVUE.

We can also check directly the unbiasedness of this estimator:

$$
\begin{aligned}
\mathbb{E}(\hat{\tau}(T)) &= \mathbb{E}\left[\bar{X}\left(\frac{n}{n-1}\bar{X} - \frac{1}{n-1}\right)\right] \\
&= \frac{n}{n-1}\mathbb{E}(\bar{X})^2 - \frac{\mathbb{E}(\bar{X})}{n-1} \\
&= \frac{n}{n-1}\left[\mathrm{Var}(\bar{X}) + (\mathbb{E}(\bar{X}))^2\right] - \frac{\theta}{n-1} \\
&= \frac{n}{n-1}\left(\frac{\theta(1-\theta)}{n} + \theta^2\right) - \frac{\theta}{n-1} \\
&= \theta^2
\end{aligned}
$$

Example 4.35 (here is another one to try)

Suppose $X_1, X_2, \ldots, X_n$ are independent Uniform$(0, \theta)$ random variables.

i) Find the UMVUE of $\tau(\theta) = \theta^2$ and calculate its variance.

ii) Find the UMVUE of $h(\theta) = \frac{1}{\theta}$.

Solution:

Step 1: $T = X_{(n)}$ is complete and sufficient for θ, with:

$$f_T(t) = \frac{nt^{n-1}}{\theta^n}, \quad 0 < t < \theta$$

Step 2: Hence

$$E(T^2) = \frac{n}{n+2}\theta^2$$

which leads to the unbaised estimator of $\tau(\theta)$:

$$T_1 = \frac{n+2}{n} T^2$$

Step 3: By Lehmann–Scheffé, $\hat{\tau}(T) = \frac{n+2}{n} T^2$ is the UMVUE with its variance calculated as:

$$\begin{aligned}
\mathbb{E}\Big(\frac{n+2}{n} T^2\Big)^2 - \theta^4 &= \Big(\frac{n+2}{n}\Big)^2 \mathbb{E}(T^4) - \theta^4 \\
&= \Big(\frac{n+2}{n}\Big)^2 n \int_0^\theta \frac{t^{n+3}}{\theta^n} dt - \theta^4 \\
&= \theta^4 \Big[\frac{(n+2)^2}{n} \frac{1}{n+4} - 1\Big] \\
&= \frac{4\theta^4}{n(n+4)}
\end{aligned}$$

The solution to ii) follows very similar lines. Try and solve this one yourself.