

Unbiased Estimation and Search for the Best Estimator

Sujit Sandipan Chaugule^{1*}, Dr. Amiya Ranjan Bhowmick²

^{1*}Department of Pharmaceutical Sciences and Technology, Institute of Chemical Technology, Mumbai

²Department of Mathematics, Institute of Chemical Technology, Mumbai

Search for the Best Unbiased Estimator

Till now we have learnt the following: We can compute the risk function of an estimator under different loss functions. If the computation is not possible to execute analytically, we know how to obtain the risk function by computer simulation. We have seen several examples, but we should keep in mind that always the simulation from the population distribution may not be an easy task. This idea will be explained later

We can compute the maximum likelihood estimator of the parameters both analytically and numerically

Still we are not in a position to decide whether we have obtained the best estimator which has the uniformly minimum risk against all estimators of the parameter. This is definitely a difficult task and we can not even imagine the universe of all possible estimators to get hold of one best estimator. In addition, even if we get hold of the best estimator, we are not really sure whether this is the unique one or not. Let us try to address these questions in this document.

Concentrating on Unbiased Estimators

To obtain the best estimator, we reduce our universe of all possible estimators to the set of all unbiased estimators. Suppose that we have a random sample of size n from the population PDF (PMF) $f(x; \theta)$, $\theta \in \Theta \subseteq \mathbb{R}$. An obvious consequence of having the class of unbiased estimators is that if $\psi(X_1, \dots, X_n)$ be an unbiased estimator of θ , then $\mathbb{E}_\theta(\psi) = \theta$ for all $\theta \in \Theta$, which implies that under the squared error loss

$$\mathbb{E}_\theta(\psi_n - \theta)^2 = \mathbb{E}_\theta(\psi - \mathbb{E}_\theta(\psi_n))^2 = \text{Var}_\theta(\psi_n).$$

We keep the subscript n to emphasize the important role of sample size in studying the properties of the estimator. Therefore, for two unbiased estimators ψ'_n and ψ''_n , ψ'_n will be a better estimator of θ if

$$\text{Var}_\theta(\psi'_n) \leq \text{Var}_\theta(\psi''_n), \quad \text{for all } \theta.$$

Uniformly Minimum Variance Unbiased Estimator

Let X_1, X_2, \dots, X_n be a random sample of size n from the population PDF (PMF) $f(x|\theta)$. Suppose we are interested in estimating $g(\theta)$, a function of θ . An estimator $\psi_n^*(X_1, \dots, X_n)$ of $g(\theta)$ is defined to be a uniformly minimum variance unbiased estimator of $g(\theta)$ if $\mathbb{E}_\theta(\psi_n^*) = g(\theta)$ for all $\theta \in \Theta$, that is, ψ_n^* is unbiased and $\text{Var}_\theta(\psi_n^*) \leq \text{Var}_\theta(\psi_n)$ for any other estimator $\psi_n(X_1, \dots, X_n)$ of $g(\theta)$ which satisfies $\mathbb{E}_\theta(\psi_n) = g(\theta)$ for all $\theta \in \Theta$.

Search for the minimum bound

In general we may not be always interested in estimating the parameter θ , rather we may be interested in estimating some function of the parameter $g(\theta)$, say. For example, suppose a random sample of size n is available from the exponential distribution with rate parameter λ and we are interested in estimator $P(X > 1) = e^{-\lambda} = g(\lambda)$. The next discussion is generalized for estimating some function of the parameter $g(\theta)$.

In search of the best estimator for $g(\theta)$, some function of the parameter θ , we first try to find the least possible risk attained by any unbiased estimator of $g(\theta)$ under squared error loss function. In other words, we want to compute a lower bound (sharp) for the variance of any unbiased estimator ψ of $g(\theta)$.

Cramer-Rao Inequality

Suppose that X_1, X_2, \dots, X_n be a random sample (not necessarily IID) of size n from the population PDF (PMF) $f(x|\theta)$ and let $\psi(\mathbf{X})$ be any estimator satisfying

$$\frac{d}{d\theta} \mathbb{E}_\theta(\psi(\mathbf{X})) = \int_{\mathcal{X}} \frac{\partial}{\partial \theta} [\psi(\mathbf{X}) f(\mathbf{x}|\theta)] d\mathbf{x}$$

and

$$\text{Var}_\theta(\psi(\mathbf{X})) < \infty,$$

then

$$\text{Var}_\theta(\psi_n(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta} \mathbb{E}_\theta(\psi(\mathbf{X})) \right)^2}{\mathbb{E}_\theta \left(\left(\frac{\partial}{\partial \theta} \log f(\mathbf{X}|\theta) \right)^2 \right)}.$$

i IID Case

If the assumptions of the Cramer-Rao inequality are satisfied and, additionally, if X_1, X_2, \dots, X_n are IID with PDF $f(x|\theta)$, then

$$\text{Var}_\theta(\psi(\mathbf{X})) \geq \frac{\left(\frac{d}{d\theta} \mathbb{E}_\theta(\psi(\mathbf{X})) \right)^2}{n \mathbb{E}_\theta \left(\left(\frac{\partial}{\partial \theta} \log f(X|\theta) \right)^2 \right)}.$$

Consider estimating the mean parameter β from the exponential PDF given by $f(x) = \frac{1}{\beta} e^{-\frac{x}{\beta}}, 0 < x < \infty$ and $\beta \in (0, \infty)$. In this case, $g(\beta) = \beta$. Consider the MLE \bar{X}_n , which is an unbiased estimator of β and satisfies $\text{Var}_\beta(\bar{X}_n) = \frac{\beta^2}{n}$. Let us compute the Cramer-Rao (CR) bound for any unbiased estimator of β . The denominator in the expression for the C-R bound is computed as follows:

$$\begin{aligned}
n\mathbb{E}_\beta \left(\left(\frac{\partial}{\partial \beta} \log f(X|\beta) \right)^2 \right) &= n\mathbb{E}_\beta \left(\frac{\partial}{\partial \beta} \left(-\log \beta - \frac{X}{\beta} \right) \right)^2 \\
&= n\mathbb{E}_\beta \left(-\frac{1}{\beta} + \frac{X}{\beta^2} \right)^2 \\
&= \frac{n}{\beta^4} \mathbb{E}_\beta (X - \beta)^2 \\
&= \frac{n\beta^2}{\beta^4} \\
&= \frac{n}{\beta^2}.
\end{aligned}$$

Therefore, the C-R bound is $\frac{\beta^2}{n}$ and it is equal to $\text{Var}_\beta(\bar{X}_n)$. Thus, \bar{X}_n must be the UMVUE.

When C-R bound does not hold

Suppose that we have a random sample of size n from the following PDF

$$f(x|\theta) = \begin{cases} \frac{2x}{\theta^2}, & 0 < x < \theta \\ 0, & \text{otherwise.} \end{cases}$$

- Show that the MLE of θ is $Y_n = \max(X_1, \dots, X_n)$.
- The MLE is not unbiased as $\mathbb{E}_\theta(Y_n) = \frac{2n\theta}{2n+1} \neq \theta$ for all $\theta \in (0, \infty)$. However,

$$\text{Bias}_\theta(Y_n) = \mathbb{E}_\theta(Y_n) - \theta = -\frac{\theta}{2n+1} \rightarrow 0.$$

as $n \rightarrow \infty$. Therefore, MLE is asymptotically unbiased.

- However, we can obtain an unbiased estimator which is a function of Y_n as $\psi_n = \frac{2n+1}{2n} Y_n$, then

$$\mathbb{E}_\theta(\psi_n) = \frac{2n+1}{2n} \times \frac{2n}{2n+1} \theta = \theta.$$

The variance of ψ_n is obtained as follows:

- $\mathbb{E}_\theta(Y_n^2) = \frac{2n}{2n+2} \theta^2$.
- $\text{Var}_\theta(Y_n) = \mathbb{E}_\theta(Y_n^2) - (\mathbb{E}_\theta(Y_n))^2 = \frac{n\theta^2}{(n+1)(2n+1)^2}$.
- $\text{Var}_\theta(\psi_n) = \left(\frac{2n+1}{2n} \right)^2 \text{Var}_\theta(Y_n) = \frac{\theta^2}{4n(n+1)}$.

Compute the C-R bound to verify that it is equal to $\frac{\theta^2}{4n}$.

In the above problem, it is surprising to see that

$$\text{Var}_\theta(\psi_n) = \frac{\theta^2}{4n(n+1)} < \frac{\theta^2}{4n} = \text{C-R bound}.$$

- Check that the required condition which allows the interchange of the derivative with respect to parameter θ and the expectation does not satisfy for the given PDF.
- The result can be generalized further to observe that the support of the distribution depends on the unknown parameter θ , which is $(0, \theta)$. For such distributions, the criteria

involving the C-R lower bound are not utilized to identify the best unbiased estimator. In fact, when the support depends on the parameter, the C-R bound theorem does not work.

- Most of the book contains the $\text{Uniform}(0, \theta)$ example to demonstrate the above exception, but here, we have a different example, and you are encouraged to execute the above task. In line with this concept, one can actually test this using computer simulation as well.
- We can simulate the risk function of ψ_n^* using simulation and check that it completely lies below the C-R bound. One can use the probability integral transform to simulate the given PDF.

```
In [1]: using Distributions, Statistics, Plots, StatsBase
using Random, LaTeXStrings
```

```
In [2]: theta_vals = [2,4]
n = 100
plt = plot(layout=(1, 2), size=(800, 500))
for (idx, theta) in enumerate(theta_vals)
    u = rand(Uniform{0,1},n)
    x = theta.*sqrt.(u)
    histogram!(x, normalize = true, xlabel = "x", ylabel = "density",
    title = L"\theta = %$\theta", label = "", subplot = idx)
    plot!(x->2*x/(theta^2)*(0<x)*(x<theta), lw = 2, label = "",
    subplot = idx)
end
display(plt)
```

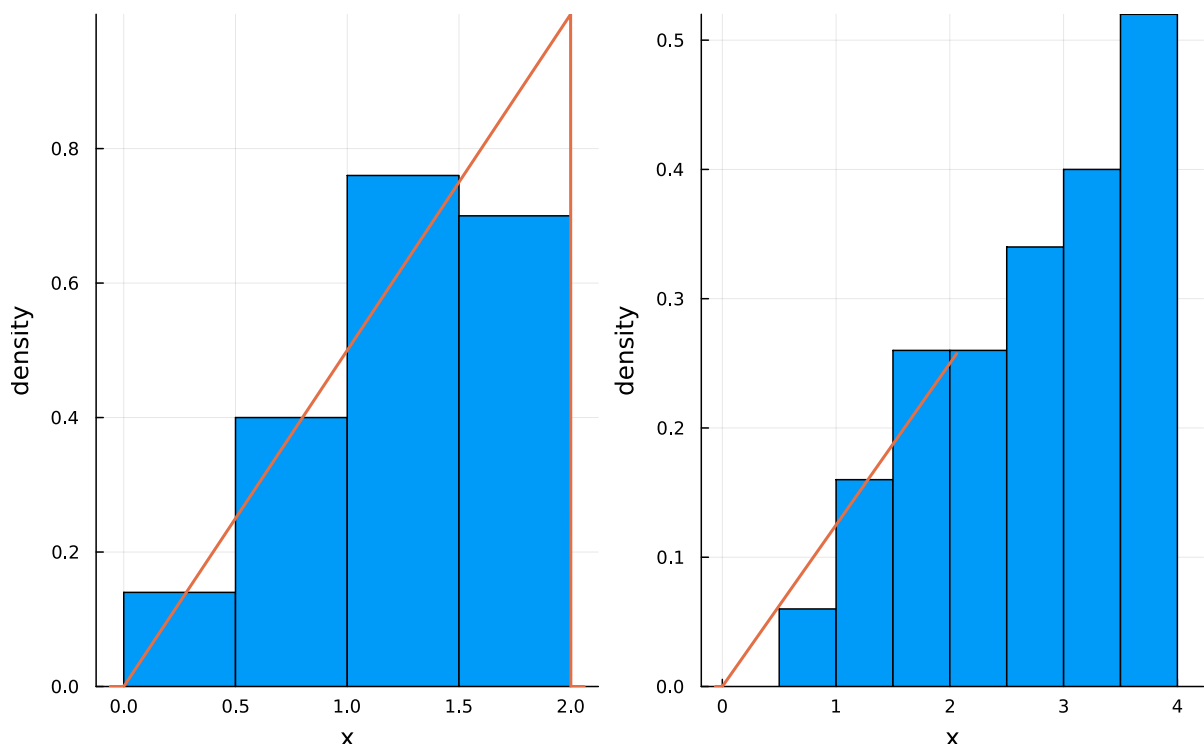


Figure 1 :The simulation of random samples from the given probability density function using probability integral transform. A sample size of $n = 100$ has been simulated for two values of θ . The overlaid exact PDF confirms the effective use of the PIT for simulation.

```
In [3]: ## simulation of risk function for n = 4
n = 4
```

```

theta_vals = 1:0.1:4
risk_vals = zeros(length(theta_vals))
M = 1000 # no of replications

for i in 1:length(theta_vals)
    theta = theta_vals[i]
    psi_n = zeros(M)
    loss_vals = zeros(M)
    for j in 1:M
        u = rand(Uniform(0,1), n)
        x = theta.*sqrt.(u)
        psi_n[j] = ((2*n+1)/(2*n))*maximum(x)
        loss_vals[j] = (psi_n[j] .- theta).^2
    end
    risk_vals[i] = mean(loss_vals)
end

```

```

In [4]: scatter(theta_vals, risk_vals, color = "red", markersize = 6,
               ylims = (0, maximum(theta_vals)^2/(4*n)), xlabel = L"\theta",
               ylabel = L"R(\psi_n, \theta)", label = "")

plot!(x->x^2/(4*n*(n+1)), color = "blue", lw = 2, label = "")
plot!(x->x^2/(4*n), color = "magenta", lw = 2, linestyle = :dash,
      label = "C-R bound")

```

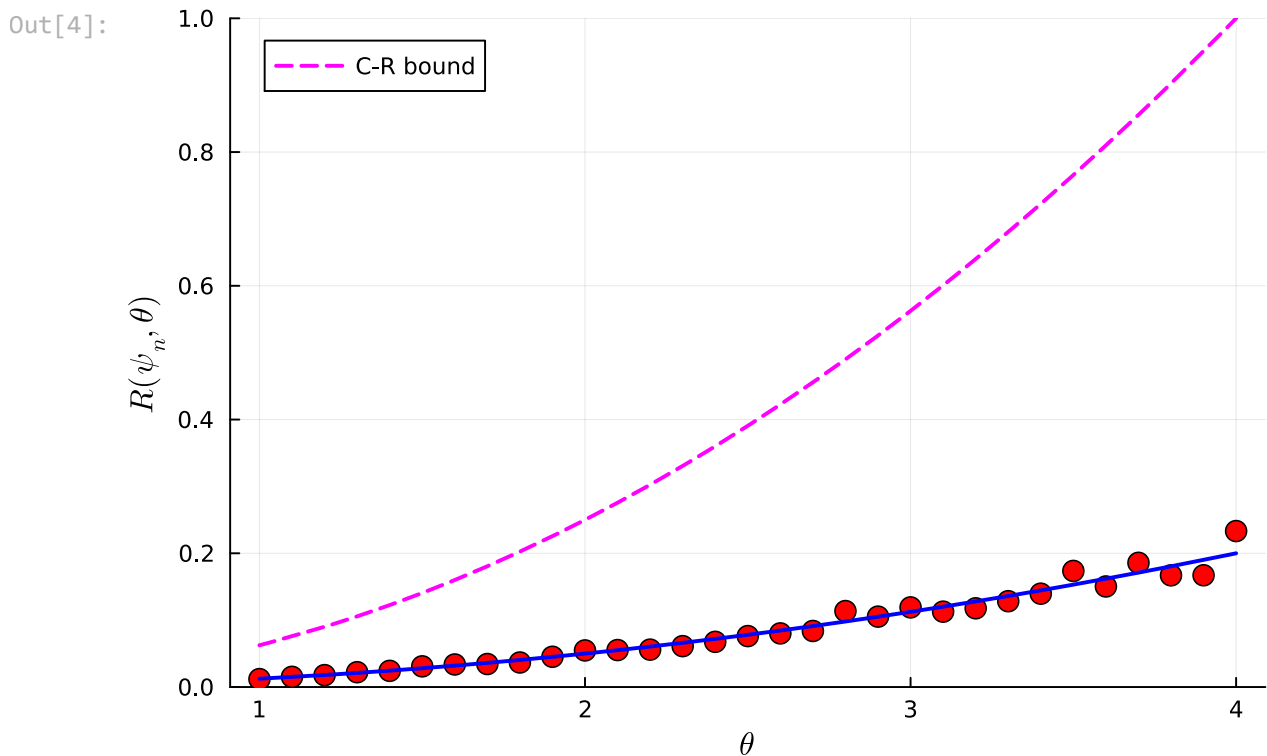


Figure 2 :The right most panel shows the risk function of ψ_n which is obtained based on $M = 1000$ replications (red dots).The true risk function (blue color) is added for reference. The C-R bound is shown using the dotted magenta colour. It is clear from the picture that the for the given PDF, characterization of the best estimator based on the C-R bound can not be established.

Therefore, we need to understand additional theories to characterize the UMVUE. We shall now introduce two important concepts, Sufficient Statistics and Complete Statistics which will help us address the gaps we encountered in characterizing the UMVUE

Fisher Information Number

The following quantity, which appears in the denominator of the lower bound for the variance of any unbiased estimator, is known as the Fisher Information:

$$\mathbb{E}_{\theta} \left(\left(\frac{\partial}{\partial \theta} \log f(\mathbf{X}|\theta) \right)^2 \right).$$

This quantity tells us as the information number increases, we have more information about the parameter θ . An alternative computation for the Fisher Information can be eased by the following theorem.

Computing Fisher Information

If $f(x|\theta)$ satisfies

$$\frac{d}{d\theta} \mathbb{E}_{\theta} \left(\frac{\partial}{\partial \theta} \log f(X|\theta) \right) = \int \frac{\partial}{\partial \theta} \left[\left(\frac{\partial}{\partial \theta} \log f(x|\theta) \right) f(x|\theta) \right] dx$$

then

$$\mathbb{E}_{\theta} \left(\left(\frac{\partial}{\partial \theta} \log f(X|\theta) \right)^2 \right) = -\mathbb{E}_{\theta} \left(\frac{\partial^2}{\partial \theta^2} \log f(X|\theta) \right)$$

When the CR bound will be attained

Let X_1, X_2, \dots, X_n be independent and identically distributed random variables following $f(x|\theta)$, and $f(x|\theta)$ satisfies the conditions of the Cramer-Rao theorem. Let

$$\mathcal{L}(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta)$$

denote the likelihood function. If $W(\mathbf{X}) = W(X_1, \dots, X_n)$ is any unbiased estimator of $g(\theta)$, then $W(\mathbf{X})$ attains the Cramer-Rao Lower Bound if and only if

$$a(\theta)[W(\mathbf{x}) - g(\theta)] = \frac{\partial}{\partial \theta} \log \mathcal{L}(\theta|\mathbf{x}),$$

for some function $a(\theta)$.

Sufficient Statistics

Let X_1, X_2, \dots, X_n be a random sample of size n from the PDF (PMF) $f(x|\theta)$, where θ may be a vector of parameters. A statistic $T(X_1, \dots, X_n)$ is said to be a sufficient statistic if the conditional distribution of (X_1, \dots, X_n) given $T = t$ does not depend on θ for any value t of T .

Sufficiency principle

If $T(\mathbf{X}) = T(X_1, \dots, X_n)$ is a sufficient statistic for θ , then any inference about θ should depend on the sample \mathbf{X} only through the value $T(\mathbf{X})$. That is, if \mathbf{x} and \mathbf{y} are two sample

points such that $T(\mathbf{x}) = T(\mathbf{y})$, then the inference about θ should be the same whether $\mathbf{X} = \mathbf{x}$ or $\mathbf{Y} = \mathbf{y}$ is observed.

Exponential Family and minimal sufficient statistics

Exponential family of densities

A one-parameter family (the parameter $\theta \in \Theta \subseteq \mathbb{R}$) of PDF (PMF) $f(x|\theta)$ that can be expressed as

$$f(x|\theta) = a(\theta)b(x)e^{c(\theta)d(x)}$$

for all $x \in (-\infty, \infty)$ and for all $\theta \in \Theta$, and for a suitable choice of functions $a(\cdot)$, $b(\cdot)$, $c(\cdot)$ and $d(\cdot)$, it is defined to belong to the exponential family or exponential class. If we have a random sample of size n from an exponential family, then the joint distribution of the data can be expressed as follows:

$$\prod_{i=1}^n f(x_i|\theta) = (a(\theta))^n \left[\prod_{i=1}^n b(x_i) \right] \exp \left[c(\theta) \sum_{i=1}^n d(x_i) \right].$$

Worked Examples

- **$\mathcal{G}(\alpha, \beta)$ family:**

$$f(x|\alpha, \beta) = \frac{x^{\alpha-1}e^{-\frac{x}{\beta}}}{\Gamma(\alpha)\beta^\alpha} = a(\alpha, \beta)b(x) \exp \left[(\alpha - 1) \log(x) + \left(-\frac{1}{\beta} \right) x \right],$$

where $a(\alpha, \beta) = \Gamma(\alpha)\beta^\alpha$, $b(x) = 1$, $c_1(\alpha, \beta) = \alpha - 1$, $c_2(\alpha, \beta) = -\frac{1}{\beta}$, $d_1(x) = \log(x)$ and $d_2(x) = x$. Therefore,

$$\left(\sum_{i=1}^n X_i, \sum_{i=1}^n \log(X_i) \right)$$

is jointly minimal sufficient for (α, β) .

- Show that the following PMFs belong to the exponential family and identify the corresponding minimal sufficient statistics
 - binomial(n, θ), n known.
 - Poisson(λ)
 - Geometric(p)
 - Discrete Uniform $\{1, 2, \dots, \theta\}$, $\theta \in \{1, 2, 3, \dots\}$.
- Identify which of the following PDFs belong to the exponential family and identify the corresponding minimal sufficient statistics
 - $\mathcal{N}(\mu, \sigma^2)$
 - $\mathcal{B}(a, b)$
 - $\mathcal{G}(\alpha, \beta)$
 - Cauchy(μ, σ)

- Uniform(0, θ)
- According to the factorization criteria,

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

is a sufficient statistic for θ .

- However, the above statement can be made much stronger in a sense that the above sufficient statistic

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

is in fact minimal sufficient.

k -parameter exponential family

A family of PDF (PMF) $f(x|\theta_1, \dots, \theta_k)$ that can be expressed as

$$f(x|\theta_1, \dots, \theta_k) = a(\theta_1, \dots, \theta_k)b(x) \exp \left[\sum_{j=1}^n c_j(\theta_1, \dots, \theta_k)d_j(x) \right]$$

for suitable choice of functions $a(\cdot, \dots, \cdot)$, $b(\cdot)$, $c_j(\cdot, \dots, \cdot)$ and $d_j(\cdot)$, $j = 1, 2, \dots, k$ is defined to belong to the exponential family.

Examples of distribution not belonging to the exponential family

- The family of Uniform(0, θ) does not belong to the exponential family
- The statement can be true in general. That is, any family of densities for which

the range of the values where the density is non-negative depends of the parameter θ does not belong to the exponential class

Sufficiency and Unbiasedness

Examples of distribution not belonging to the exponential family

Let W be any unbiased estimator of $g(\theta)$, and T be a sufficient statistic for θ . Define $\phi(T) = \mathbb{E}[W|T]$. Then,

- $\mathbb{E}_\theta \phi(T) = g(\theta)$.
- $\text{Var}_\theta \phi(T) \leq \text{Var}_\theta W$ for all θ . The above two conditions imply that $\phi(T)$ is uniformly better unbiased estimator of $g(\theta)$.

The Rao-Blackwell theorem states a very important fact that conditioning any unbiased estimator on a sufficient statistic will result in a uniform improvement, so we need to consider only statistics that are functions of a sufficient statistic in our search for the best estimators.

Uniqueness of the best estimator

If W is a best unbiased estimator of $g(\theta)$, then W is unique.

Complete Statistics

Uniqueness of the best estimator

Let $f(t|\theta)$ be a family of PDFs or PMFs for a statistic $T(\mathbf{X})$. The family of probability distributions is called complete if $\mathbb{E}_\theta g(T) = 0$ for all θ implies $P(g(X) = 0) = 1$ for all θ . Equivalently, $T(\mathbf{X})$ is a complete statistic.

Examples of complete family

- Let (X_1, X_2, \dots, X_n) be iid Uniform(0, θ) observations, $0 < \theta < \infty$. Then $T(\mathbf{X}) = \max(X_1, \dots, X_n) = X_{(n)}$ is a complete statistic.

- Let X_1, \dots, X_n be a random sample from a population with PDF

$$f(x|\theta) = \theta x^{\theta-1}, \quad 0 < x < 1, 0 < \theta < \infty.$$

Find a complete sufficient statistic for θ .

- Let X_1, X_2, \dots, X_n be IID with geometric distribution

$$\begin{aligned} P_\theta(X = x) \\ = \theta(1 - \theta)^{x-1}, \quad x = 1, 2, \dots, \quad 0 < \theta < 1. \end{aligned}$$

Show that $\sum_{i=1}^n X_i$ is sufficient for θ , and find the family of distributions of $\sum X_i$. Check whether the family is complete.

Complete statistics in the exponential family

Let X_1, X_2, \dots, X_n be iid observations from an exponential family with PDF or PMF of the form

$$f(x|\theta) = a(\theta)b(x) \exp \left(\sum_{j=1}^k c_j(\theta)d_j(x) \right),$$

where $\theta = (\theta_1, \dots, \theta_k)$. Then the statistic

$$\begin{aligned} T(\mathbf{X}) \\ = \left(\sum_{i=1}^n d_1(X_i), \sum_{i=1}^n d_2(X_i), \dots \right. \\ \left. \cdot, \sum_{i=1}^n d_k(X_i) \right) \end{aligned}$$

is complete if
 $\{(c_1(\theta), c_2(\theta), \dots, c_k(\theta)) : \theta \in \Theta\}$
contains an open set in \mathbb{R}^k .

- If $X_1, \dots, X_n \sim \mathcal{N}(\theta, \theta^2)$. Here the parameter space (θ, θ^2) does not contain a two-dimensional open set, as it consists of only the points on the parabola.

The following theorem delineates the connection between complete sufficiency of an unbiased estimator with the best unbiased estimator.

Best Unbiased Estimator

Important

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

Let us see an illustrative example. Suppose X_1, X_2, \dots, X_n be IID $f(x|\theta)$, where

$$f(x|\theta) = \theta x^{\theta-1} I_{(0,1)}(x), \\ 0 < \theta < \infty.$$

- First, we observe that this PDF belongs to the exponential family as

$$f(x|\theta) = \theta \times 1 \times \exp \\ [(\theta - 1) \log x] \\ = a(\theta)b(x) \exp[c(\theta)d(x)].$$

- Therefore, $\sum_{i=1}^n \log X_i$ is a complete sufficient (in fact minimal) statistic for θ .
- Compute the MLE of θ : the likelihood function $\mathcal{L}(\theta)$ is

$$\mathcal{L}(\theta) = \theta^n \left(\prod_{i=1}^n x_i^{\theta-1} \right),$$

and

$$l(\theta) = \log \mathcal{L}(\theta) = n \log \theta \\ + (\theta - 1) \sum_{i=1}^n \log X_i.$$

- Solving $l'(\theta) = 0$, we obtain the MLE of θ as

$$\hat{\theta}_n = T_n = -\frac{n}{\sum_{i=1}^n \log X_i}.$$

- First, let us check whether $\hat{\theta}_n$ is unbiased. We must note that the MLE is in fact a function T_n , which is a complete sufficient statistic $\sum_{i=1}^n \log X_i$. We obtain the sampling distribution of $\hat{\theta}_n$.

- **(Step - I)** Show that $Y_i = -\log X_i \sim \text{Exponential}(\text{rate} = \theta)$.

- **(Step - II)** Show that

$$Y = \sum_{i=1}^n Y_i = -\sum_{i=1}^n \log X_i \sim \text{Gamma}(\alpha = n, \beta = \frac{1}{\theta}).$$

- **(Step - III)** Show that the PDF of $\hat{\theta}_n$ is given by

$$f_{T_n}(z) = \frac{(n\theta)^n}{\Gamma(n)} z^{-(n+1)} e^{-\frac{n\theta}{z}},$$

$$0 < z < \infty.$$

- **(Step - IV)** Show that $\mathbb{E}_\theta(T_n) = \frac{n\theta}{n-1}$, where $n > 1$. Therefore, T_n is not an unbiased estimator of θ .
- **(Step - V)** We can get an unbiased estimator which is a function of T_n , which is a function of a complete sufficient statistic as

$$T_n^* = \frac{n-1}{n} T_n$$

$$= \frac{n-1}{n} \left(-\frac{n}{\sum_{i=1}^n \log X_i} \right).$$

Therefore, T_n^* is the best unbiased estimator.

- **(Step - VI)** Compute the variance of T_n^* .
- **(Step - VII)** Compute the Cramér-Rao bound for any unbiased estimator of θ and check whether

$$\text{Var}_\theta(T_n^*) = I_n(\theta)^{-1}$$

for all $\theta \in (0, \infty)$.