Notes

MA4107: Statistical Inference

Sabarno Saha

ss22ms037@iiserkol.ac.in

Date: 09 September 2025

Contents

1. Data and Models

Definition 1.1 (Data): Let G generate a vector in $\mathbb{R}^n$ according to some model.

$$G : \Omega \longrightarrow \mathbb{R}^n \tag{1.1}$$

Then $\underset{\sim}{X}$ is called a realization of the data $G(\omega)$ for some $\omega \in \Omega$

$\underset{\sim}{X}$ comes from some distribution or model F. This course is all about parametric models.

Definition 1.2 (Parametric Model): Let F be some parameter model. Then

$$F \in \mathfrak{F} = \left\{F_\theta : \theta \in \Theta \subset \mathbb{R}^k \text{ for some } k \in \mathbb{N}\right\} \tag{1.2}$$

Θ is called the parameter space.

We define g to be the map that maps a set of parameters to a model. g is onto. For every F_θ, there always exists θ, such that

$$\theta \underset{g}{\mapsto} F_\theta \tag{1.3}$$

g also needs to be injective, otherwise we run into the problem of identifiability. The problem of identifiability arises when we cannot infer the parameters from the model F_θ.

Unlees explicitly mentioned otherwise, the data is represented by $\underset{\sim}{X}$. For all models in this course, we assume the following,

1. The parametrization is bijective.
2. The models $F_\theta \in \mathfrak{F}$ are all either discrete or continuous, not a mixture of both.

2. Statistic

Definition 2.1 (Statistic): Let the data be $\underset{\sim}{X} \in \mathbb{R}^n$. A statistic $T = T(X)$ is a measureable function of data and data only.

2.1. Ancillary Statistic

Definition 2.1.1 (Ancillary Statistic): Suppose the distribution of the test statistic $T(X)$, $\mathfrak{T}$ is independent of the parameter vector θ. Then $T(X)$ is called an Ancillary Statistic.

Theorem 2.1.1:

Let $f(x)$ be a pdf. Let μ and $\sigma > 0$ be any constants. Then

$$g(x|\mu, \sigma) = \frac{1}{\sigma} f\left(\frac{x - \mu}{\sigma}\right) \tag{2.1}$$

is a valid pdf.

The proof is in Casella and burger theorem 3.5.1. One just needs to show non-negativity and normalization.

2.2. Location and Scale families of distributions

Definition 2.2.1 (Location Family of distributions):

Let a family of distributions be,

$$\mathfrak{F} = \{f_\theta : \theta \in \Theta \text{ where } f_\theta(x) = g(x - \theta) \text{ for some known function } g \text{ on } \mathbb{R}^n\} \tag{2.2}$$

Then $\mathfrak{F}$ is a location family of distrbutions with the standard pdf $f(x)$ and the location parameter θ for the family.

Generally we will talk about these families on $\mathbb{R}$ rather than $\mathbb{R}^n$.

Definition 2.2.2 (Scale family of distributions):

Let a family of distributions be,

$$\mathfrak{F} = \left\{f_\theta : \theta \in \Theta \text{ where } f_\theta(x) = \frac{1}{\theta} g\left(\frac{x}{\theta}\right) \text{ for some known function } g \text{ on } \mathbb{R}^n\right\} \tag{2.3}$$

Then $\mathfrak{F}$ is a scale family of distrbutions with the standard pdf $f(x)$ and the scale parameter θ for the family.

Definition 2.2.3 (Location-Scale family of distribution):

Let a family of distributions be,

$$\begin{aligned}\mathfrak{F} = \Bigg\{f_{\mu,\sigma} : (\mu, \sigma) \in \mathbb{R} \times \mathbb{R}^+ \text{ where } f_{\mu,\sigma}(x) = \frac{1}{\sigma} g\left(\frac{x - \mu}{\sigma}\right) \\ \text{for some known function } g \text{ on } \mathbb{R}^n\Bigg\}\end{aligned} \tag{2.4}$$

Then $\mathfrak{F}$ is a location family of distrbutions with the standard pdf $f(x)$ and the location parameter μ and scale parameter σ for the family.

Now comes to the choice of test statistics. Note that ancillary statistics are not useful in inferring parameters from data. So we shall not use them. The test statistic just compresses the data, from $\underset{\sim}{X}$ to $T\left(\underset{\sim}{X}\right)$. We want to see when this data compression is lossless.

We can define the level sets of the test statistic $T(X)$. Let D be the set on which the data can lie in. Let us define the relation,

$$x \sim_T y \text{ if and only if } T(x) = T(y) \tag{2.5}$$

We can easily see that this is an equivalence relation. Then D can be partitioned into sets D_t s.t.,

$$D_t = \{X \in D : T(X) = t\} \tag{2.6}$$

These sets D_t are called level sets of the test Statistic T. If we are in some level set of the test statistic, the inferred parameters should not change if we stay in the same level set.

2.3. Sufficient Statistic

Definition 2.3.1 (Sufficient Statistic):

The test statistic $T = T\left(\underset{\sim}{X}\right)$ is said to be sufficient iff the distribution of $\left(\underset{\sim}{X} | T = t\right)$ is free of θ for all values of t.

Theorem 2.3.1 (Fisher Neyman Factorization theorem):

Suppose $\underset{\sim}{X}$ is an iid sample from $f_\theta(\cdot)$, which might be either a pmf or a pdf. A statistic $T = T(X)$ is sufficient for θ if and only if,

$$f_\theta\left(\underset{\sim}{X}\right) = g\left(T\left(\underset{\sim}{X}\right), \theta\right) h(x) \tag{2.7}$$

where $\theta \in \Theta \subset \mathbb{R}^k$, and the functions $g(\cdot)$ and $h(\cdot)$ are non negative functions and $h(\cdot)$ is independent of θ.

In principle we can get multiple sufficient statistics, with finer and finer partitions of D. We try to find the statistic which is sufficient and has the coarsest partitions in D.

2.4. Minimal Sufficient Statistic

Definition 2.4.1 (Minimal Sufficient Statistic):

A statistic T is said to be minimal sufficient for θ iff it is sufficient for θ and for any other sufficient statistic S $\exists$ a function g such that,

$$T = g(S) \tag{2.8}$$

Lemma 2.4.1: If T_1 and T_2 are minimal sufficient statistics, there exist injective functions g_1 and g_2 such that,

$$T_1 = g_2(T_2) \quad T_2 = g_1(T_1) \tag{2.9}$$

INSERT PROOF.

Theorem 2.4.2 (Characterization of minimal sufficiency):

Let $\underset{\sim}{X}$ be a joint pmf/pdf $f_\theta(x)$ and $T = T\left(\underset{\sim}{X}\right)$ be a statistic. Suppose the following property holds,

$$\frac{f_\theta(x)}{f_\theta(y)} \text{ is free of } \theta \text{ iff } T(x) = T(y) \ \forall\ x, y \ \text{ s.t. } f_\theta(y) \neq 0 \tag{2.10}$$

Then T is minimal sufficient for θ.

Proof: Assume for simplicity that $f_\theta(x) > 0\ \forall\ x \in \mathbb{R}^n$ and $\ \forall\ \theta \in \Theta$. We need to show two things,

1. T is sufficient for θ
2. For any other sufficient statistic S, $\exists$ a function g such that $T = g(S)$.

1: Sufficiency of T:

Let $\mathfrak{T} = \{T(y) : y \in \mathbb{R}^n\}$ be the image of T. Let $\{A_t : t \in \mathfrak{T}\}$ be the level sets of T. We can pick an appropriate y^* such that,

$$f_\theta(x) = f_\theta(y^*)\frac{f_\theta(x)}{f_\theta(y^*)} = g(T(x), \theta)h(x) \tag{2.11}$$

We choose y^* such that $T(y^*) = T(x)$, i.e. we choose y^* from the level set $A_{T(x)}$. Then by Fisher Neyman factorization theorem Theorem 2.3.1, T is sufficient for θ.

2: Minimality of T: Let S be any other sufficient statistic for θ. Then by Theorem 2.3.1,

$$f_\theta(x) = g_1(S(x), \theta)h_1(x) \tag{2.12}$$

Note that,

$$\frac{f_\theta(x)}{f_\theta(y)} = \frac{g_1(S(x), \theta)h_1(x)}{g_1(S(y), \theta)h_1(y)} \tag{2.13}$$

We need to show that there exists a function g such that $T(x) = g(S(x))$. We just need to show the single valuedness of g. To show that we just need to show if $S(x) = S(y)$ then $T(x) = T(y)$. Suppose $S(x) = S(y)$, then,

$$\frac{f_\theta(x)}{f_\theta(y)} = \frac{h_1(x)}{h_1(y)} \text{ which is free of } \ \theta \tag{2.14}$$

We can use the property in Theorem 2.4.2 to conclude that $T(x) = T(y)$. Thus g is single valued and T is minimal sufficient for θ.

∎

Note a sufficient statistic can also contain garbage information. A sufficient statistic can be paired with an ancillary statistic and the combined statistic would still be sufficient.

2.5. Complete Statistic

Definition 2.5.1 (Complete Statistic):

Let T be a statistic. Let $\{g_\theta(T = t) : \theta \in \Theta\}$ be a family of pdf/pmf of T. The statistic T is complete if given any measureable function h the following holds for all $\theta \in \Theta$,

$$\mathbb{E}_\theta[h(T)] = 0 \Longrightarrow P_\theta(h(T) = 0) = 1 \tag{2.15}$$

Unfortunately we have no easy characterization of complete statistic, like Theorem 2.4.2 for minimal sufficiency.

Theorem 2.5.1 (Basu's Theorem):

A complete sufficient statistic is independent of any ancillary statistic.

Proof: (Only the discrete case is shown here. The continuous case is similar.) Let T be a complete sufficient statistic and S be an ancillary statistic. It is enough to show that,

$$P(S = s \mid T = t) = P(S = s)\ \forall\ s, t \tag{2.16}$$

Note that here none of the probabilities depend on θ as S is ancillary and T is sufficient. Let us fix some s. Define the function,

$$h(t) = P(S = s|T = t) - P(S = s) \tag{2.17}$$

Note that $h(T)$ is a statistic as it does not depend on θ.

$$\begin{aligned}\mathbb{E}_\theta[h(T)] &= \sum_t h(t) P_{\theta(T=t)} \\ &= \sum_t [P(S = s|T = t) - P(S = s)] P_{\theta(T=t)} \\ &= \sum_t P(S = s, T = t) - P(S = s) \sum_t P_{\theta(T=t)} \\ &= P(S = s) - P(S = s) = 0\end{aligned} \tag{2.18}$$

Since T is complete, $P(h(T) = 0) = 1$. Thus $h(T) = 0$. This proves the independence of S and T.

∎

Theorem 2.5.2 (Lehmann Scheffe):

Let $\underset{\sim}{X}$ have a joint pdf/pmf. If T is a complete sufficient statistic for θ, then T is minimal sufficient.

Proof: This proof will be sketched and not completely written out.

We will construct a minimal statistic S and then show that T and S are related by one to one functions. This will prove that T is also minimal sufficient.

Define an equivalence relation on $\mathbb{R}^n$ as follows,

$$x \sim_S y \text{ if and only if } \frac{f_\theta(x)}{f_\theta(y)} \text{ is free of } \theta \ \forall\ \theta \in \Theta \tag{2.19}$$

Construct partitions of $\mathbb{R}^n$ using the equivalence relation. Define a statistic S such that $S(x) = S(y)$ if and only if $x \sim_S y$.

> **Result 1** (Lehman Scheffe sufficiency): The statistic S is minimal sufficient for θ.

A proof of this is not provided here.

Since T is sufficient for θ, we have $S = g_1(T)$, since S is minimal sufficient. Define,

$$g_2(S) = \mathbb{E}[T|S] \tag{2.20}$$

This is indeed a valid statistic. Then define,

$$g(T) = T - g_2(S) \tag{2.21}$$

Thus we can write,

$$\mathbb{E}_\theta[g(T)] = \mathbb{E}_\theta[T] - \mathbb{E}_\theta[g_2(S)] = \mathbb{E}_\theta[T] - \mathbb{E}_\theta[\mathbb{E}_\theta[T|S]] = 0 \tag{2.22}$$

Since T is complete, $P_\theta(g(T) = 0) = 1\ \forall\ \theta \in \Theta$. Thus,

$$P_\theta(T = g_2(S)) = P_\theta(T = g_2(g_1(T))) = 1\ \forall\ \theta \in \Theta \tag{2.23}$$

g_2 and g_1 can be shown to be inverses of each other. Thus T is minimal sufficient for θ. ∎

The reverse implication does not hold and an example shall be provided below.

3. Exponential Family of Distributions

Definition 3.1 (Exponential family of distribution):

Let $\underset{\sim}{X}$ have a joint pdf/pmf $f_\theta(\cdot)$ with $\theta \in \Theta \subset \mathbb{R}^p$. We say that the $f_\theta(\cdot)$ belongs to the k-parameter exponential family if $f_\theta(\cdot)$ admits the functional form,

$$f_\theta\left(\underset{\sim}{X}\right) = \exp\left[\left\{\sum_{j=1}^{k} c_j(\theta) T_j\left(\underset{\sim}{X}\right)\right\} - d(\theta)\right] S\left(\underset{\sim}{X}\right) \tag{3.1}$$

for all $x \in \mathfrak{X}, \forall \theta \in \Theta$. $\mathfrak{X}$ is the space of all values taken by $\underset{\sim}{X}$.

Equivalently the last restriction can also be rewritten as that the support of f,

$$\operatorname{supp}(f) = \{x : f_\theta(x) > 0\} \tag{3.2}$$

is free of θ.

$c_j's$ are called Natural or canonical parameters

$T_j's$ are called Natural or canonical statistics

The expression(3.1) needs to have some restrictions placed on it, otherwise we can do simple algebraic manipulations such that the upper limit in the sum k increases, which causes a pdf $f_\theta(\cdot)$ to belong to multiple parameter exponential families.

The expression is said to be minimal in the sense that the expression cannot be reduced further without breaking the functional form (3.1).

1. $c_j(\theta)$ must explicitly depend on theta. If some c_m is constant and free of theta, the corresponding $T_m\left(\underset{\sim}{X}\right)$ can be absorbed into $S\left(\underset{\sim}{X}\right)$.
2. If $i \neq j$, $c_i \neq c_j$. This means that no two c_j's can be the same. If $c_i = c_j$ for some $i \neq j$, we can define a combined statistic $\tilde{T} = T_i + T_j$.

Note that $\theta \in \Theta \subset \mathbb{R}^p$. To avoid the problem of identifiability, $k \geq p$. If $k < p$ we will lose information and not have an injective map. Generally $k = p$ for most cases. Later we will see an explicit example where $k > p$.

Result 2 (Sufficiency of Canonical statistic): The statistic $T = (T_1, ..., T_2)$,where $T_j\ \forall\ j \in \{1, ..., k\}$ are the canonical statistics defined in Definition 3.1, is sufficient for θ.

Proof: The proof of this is obvious. Using Theorem 2.3.1 and (3.1), we can see that

$$\begin{aligned} f_\theta\left(\underset{\sim}{X}\right) &= \exp\left[\left\{\sum_{j=1}^{k} c_j(\theta) T_j\left(\underset{\sim}{X}\right)\right\} - d(\theta)\right] S\left(\underset{\sim}{X}\right) \\ &= g\left(T\left(\underset{\sim}{X}\right), \theta\right) h\left(\underset{\sim}{X}\right) \end{aligned} \tag{3.3}$$

Note that the pdf $f_\theta(x)$ and the function $g\left(T\left(\underset{\sim}{X}\right), \theta\right)$ are non-negative. So $h\left(\underset{\sim}{X}\right) = S\left(\underset{\sim}{X}\right)$ must also be non negative. Thus using Theorem 2.3.1, $T = (T_1, ..., T_k)$ is a sufficient statistic for θ. ∎

Theorem 3.1 (Completeness of Canonical Statistics for the Exponential family):

Consider a k-parameter exponential family. Define the natural parameter space to be ,

$$C := \{(c_1(\theta), ..., c_k(\theta)) : \theta \in \Theta\} \tag{3.4}$$

where Θ is the parameter space. If C contains an open set in $\mathbb{R}^k$, then the statistic $T\left(\underset{\sim}{X}\right)$ is complete. Hence T is also minimal sufficient.

TRY TO SHOW.

We cannot drop the open set restriction on the parameter space due to the following counter example. Moreover this counter example also serves as an example for the $k > p$ case outlined above. The canonical statistic vector here is also minimal sufficient but not complete.

Example (Curved Normal distribution):

Let $\underset{\sim}{X} = (X_1, ..., X_n) \overset{iid}{\sim} \mathcal{N}\left(\theta, \theta^2\right), \theta > 0$. The pdf is then given by,

$$f_\theta\left(\underset{\sim}{X}\right) = \frac{e^{-\frac{n}{2}}}{(2\pi)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\theta^2}\sum_{i=1}^{n} x_i^2 + \frac{1}{\theta}\sum_{i=1}^{n} x_i - n\ln(\theta)\right] \tag{3.5}$$

The canonical parameters and canonical statistics are given by,

$$\begin{aligned} c_1(\theta) &= -\frac{1}{2\theta^2} & T_1\left(\underset{\sim}{X}\right) &= \sum_{i=1}^{n} x_i^2 \\ c_2(\theta) &= \frac{1}{\theta} & T_2\left(\underset{\sim}{X}\right) &= \sum_{i=1}^{n} x_i \end{aligned} \tag{3.6}$$

The natural parameter space is then defined as,

$$C = \left\{\left(-\frac{1}{2\theta^2}, \frac{1}{\theta}\right) : \theta > 0\right\} \tag{3.7}$$

This forms a graph in $\mathbb{R}^2$ and thus this does not contain an open set in $\mathbb{R}^2$. We will now show that

$$T\left(\underset{\sim}{X}\right) = \left(T_1\left(\underset{\sim}{X}\right), T_2\left(\underset{\sim}{X}\right)\right) = \left(\sum_{i=1}^{n} x_i^2, \sum_{i=1}^{n} x_i\right) \tag{3.8}$$

is not complete. To do this we explicitly construct a function,

$$h(T) = h(t_1, t_2) = \frac{t_2^2}{n + n^2} - \frac{t_1}{2n} \tag{3.9}$$

Note that in general $h(T) \neq 0$. Then we get ,

$$\begin{aligned} \mathbb{E}_\theta(h(T)) &= \frac{1}{n^2 + n}\mathbb{E}_\theta\left(\left(\sum_{i=1}^{n} x_i\right)^2\right) - \frac{1}{2n}\mathbb{E}_\theta\left(\sum_{i=1}^{n} x_i^2\right) \\ &= \left[\frac{1}{n^2 + n} - \frac{1}{2n}\right]\mathbb{E}_\theta\left(\sum_{i=1}^{n} x_i^2\right) + \frac{1}{n^2 + n}\mathbb{E}_\theta\left(\sum_{i,j;i\neq j} x_i x_j\right) \\ &= \theta^2\left[\frac{1-n}{1+n}\right] + \frac{n(n-1)}{n^2 + n}\theta^2 = 0 \end{aligned} \tag{3.10}$$

Thus T is not complete. Also we can very easily show using the functional form of the curved normal distribution, that T satisfies Theorem 2.4.2. Thus T is a minimal statistic. This serves as an example why the reverse implication in Theorem 2.5.2 does not hold.

This also serves as an example where $k > p$. The dimension of the parameter space $p = 1$ and this belongs to a $k = 2$ parameter exponential family.

4. Optimal Estimation

We now need to quantify the "error of approximation" to select among potential candidates for statistics.

Definition 4.1 (Mean Squared Error):

Let T be a test statistic. Let $\theta \in \Theta$ be a parameter. Then the mean squared error, function of the parameter θ, is given by,

$$\mathrm{MSE}_\theta(T) = \mathbb{E}_\theta\left[(T - \theta)^2\right] \tag{4.1}$$

Theorem 4.1 (Bias Variance Decomposition): The Mean squared error can be decomposed into,

$$\mathrm{MSE}_\theta(T) = \mathbb{E}_\theta\left[(T-\theta)^2\right] = \left[\underbrace{\mathbb{E}_\theta(T) - \theta}_{\mathrm{Bias}_\theta(T)}\right]^2 + \mathrm{Var}_\theta(T) \tag{4.2}$$

This is called the Bias-Variance decomposition.

Proof: This is trivial as,

$$\mathbb{E}_\theta\left[(T-\theta)^2\right] = \mathrm{Var}_\theta(T-\theta) + [\mathbb{E}_\theta(T) - \mathbb{E}_\theta(\theta)]^2 = \mathrm{Var}_\theta(T) + [\mathbb{E}_\theta(T) - \theta]^2 \tag{4.3}$$

∎

To find an optimal statistic, we need to find one such that the MSE is minimized uniformly in θ over the parameter space Θ. Unfortunately this is not possible due to one simple fact. Just fix some $\theta_a \in \Theta$. Then define the statistic to be $T = \theta_a$. This is a valid statistic. Any statistic chosen will be worse than this in some neighbourhood of θ_a.(Is MSE a continuous function of θ ?)

We have some possible alternatives to error quantifiers, and look for one dimensional summaries of the curve $\theta \mapsto \mathrm{MSE}_\theta$

We can have the

1. Bayes Approach :

$$R_1(T) = \int_\Theta \mathrm{MSE}_\theta(T)\omega(\theta)\,\mathrm{d}\theta \tag{4.4}$$

   $\omega(\theta)$ is called the prior.

2. Minimax Approach:

$$R_2(T) = \sup_{\theta\in\Theta} \mathrm{MSE}_\theta(T) \tag{4.5}$$

   This quantifies the worst possible scenario.

We will be heading in a different direction and only consider the case of parameters with $\mathrm{Bias}_\theta(T) = 0\ \forall\ \theta \in \Theta$.

4.1. UMVUE

Definition 4.1.1 (Unbiased Estimator): The statistic T is said to unbiased for θ if

$$\mathbb{E}_\theta[T] = \theta\ \forall\ \theta \in \Theta \tag{4.6}$$

Our aim is to find T^* among the class of unbiased estimators such that the mean squared error is minimized, which means the following property holds,

$$\underbrace{\mathrm{MSE}_\theta(T^*)}_{\mathrm{Var}_\theta(T^*)} \le \underbrace{\mathrm{MSE}_\theta(T)}_{\mathrm{Var}_\theta(T)}\ \forall\ \theta \in \Theta, \forall\ T \quad \text{unbiased for} \quad \theta \tag{4.7}$$

If such a T^* exists, it is called the Uniform Minimum Variance Unbiased Estimator (UMVUE).

Some comments on unbiased estimators,

1. An unbiased estimator need not exist.
2. For a given class of estimators, the MSE error need not be the lowest for an unbiased estimator.

Unbiased estimators might not very good for certain purposes. In general this approach is taken so that the MSE can be minimized over all values of θ in Θ, provided we restrict ourselves to unbiased estimators.

Theorem 4.1.1 (Rao-Blackwell theorem):

Let $\underset{\sim}{X}$ have a joint pdf/pmf $f_\theta(\cdot)$ with $\theta \in \Theta \subset \mathbb{R}^p$. Let $T = T\left(\underset{\sim}{X}\right)$ be a sufficient statistic for θ. Let $S = S\left(\underset{\sim}{X}\right)$ be a statistic such that

1. $\mathbb{E}_\theta[S] = \theta \ \forall \ \theta \in \Theta$
2. $\mathrm{Var}_\theta(S) < \infty \ \forall \ \theta \in \Theta$

Define the new statistic,

$$S^* = \mathbb{E}_\theta[S \mid T] \tag{4.8}$$

This is called the Rao-Blackwellization of S with respect to T.

Then the following holds,

1. S^* is unbiased for θ
2. $\mathrm{Var}_\theta(S^*) \leq \mathrm{Var}_\theta(S) \ \forall \ \theta \in \Theta$

with equality if and only if $P_\theta(S^* = S) = 1 \ \forall \ \theta \in \Theta$.

Proof: Note that S^* is a statistic as it is a conditional expectation on the sufficient statistic T. To show unbiasedness,

$$\mathbb{E}_\theta[S^*] = \mathbb{E}_\theta[\mathbb{E}_\theta[S \mid T]] = \mathbb{E}_\theta[S] = \theta \tag{4.9}$$

using the law of total expectation.

To show the variance reduction, we use a formula for the conditional variance.

Let X, Y be random variables, with $\mathbb{E}\left[Y^2\right] < \infty$. Then,

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y|X)] + \mathrm{Var}(\mathbb{E}[Y|X]) \geq \mathrm{Var}(\mathbb{E}[Y|X]) \tag{4.10}$$

Since $\mathrm{Var}(Y|X) \geq 0$. (WILL WRITE EQUALITY PROOF LATER) ∎

The above theorem leaves us with two questions,

- What unbiased estimator S should we start with to minimize MSE ?
- What sufficient statistic T should we use to minimize MSE ?

Theorem 4.1.2 (Minimal Statistic for unbiased estimation):

Let S be an unbiased estimator for θ and let T_1 and T_2 be two sufficient statistics for θ. Define the Rao-Blackwellizations,

$$\begin{aligned} S_1 &= \mathbb{E}_\theta[S \mid T_1] \\ S_2 &= \mathbb{E}_\theta[S \mid T_2] \end{aligned} \tag{4.11}$$

If $T_1 = h(T_2)$, for some function h, then

$$\operatorname{Var}_\theta(S_1) \leq \operatorname{Var}_\theta(S_2) \ \forall \ \theta \in \Theta \tag{4.12}$$

Thus if a minimal sufficient statistic T exists, the Rao-Blackwellization with respect to T is the best among all Rao-Blackwellizations with respect to any other sufficient statistic.

Proof: To show this we use the tower property of conditional expectation. If $Y = f(X)$, then

$$\mathbb{E}(Z|Y) = \mathbb{E}(\mathbb{E}(Z|X)|Y) \tag{4.13}$$

Since $T_1 = h(T_2)$,

$$\begin{aligned} S_1^* &= \mathbb{E}(S|T_1) \\ &= \mathbb{E}(\mathbb{E}(S|T_2)|T_1) \\ &= \mathbb{E}(S_2^*|T_1) \end{aligned} \tag{4.14}$$

Using Theorem 4.1.1,

$$\operatorname{Var}_\theta(S_1^*) = \operatorname{Var}_\theta(\mathbb{E}(S_2^*|T_1)) \leq \operatorname{Var}_\theta(S_2^*) \ \forall \ \theta \in \Theta \tag{4.15}$$

∎

Theorem 4.1.3 (Choice of Unbiased estimator): Assume that T is complete sufficient for θ. Let S_1 and S_2 be two unbiased estimators for θ. Define the Rao-Blackwellizations,

$$\begin{aligned} S_1^* &= \mathbb{E}_\theta[S_1 \mid T] \\ S_2^* &= \mathbb{E}_\theta[S_2 \mid T] \end{aligned} \tag{4.16}$$

Then

$$P_\theta(S_1^* = S_2^*) = 1 \ \forall \ \theta \in \Theta \tag{4.17}$$

Proof: Note that both S_1^* and S_2^* are unbiased for θ. Thus

$$\mathbb{E}_\theta[S_1^* - S_2^*] = 0 \ \forall \ \theta \in \Theta \tag{4.18}$$

We have acheived first order ancillarity. S_1^* and S_2^* are functions of the complete sufficient statistic T. Thus using the definition of completeness,

$$P_\theta(S_1^* - S_2^* = 0) = 1 \ \forall \ \theta \in \Theta \tag{4.19}$$

∎

This theorem shows that the choice of unbiased estimator does not matter if we have a complete sufficient statistic. The Rao-Blackwellization will always be the same.

Theorem 4.1.4 (Lehmann-Scheffe for UMVUE): Let T be a complete sufficient statistic for θ. Let S be any unbiased estimator for θ, such that $\mathrm{Var}_\theta(S) < \infty \ \forall \ \theta \in \Theta$. Then the Rao-Blackwellization,

$$S^* = \mathbb{E}_\theta[S \mid T] \tag{4.20}$$

Then,

1. $\mathrm{Var}_\theta(S^*) \le \mathrm{Var}_\theta(U) \ \forall \ \theta \in \Theta, \forall \ U \quad \text{unbiased for} \quad \theta$
2. If for some unbiased estimator U, $\mathrm{Var}_\theta(S^*) = \mathrm{Var}_\theta(U) \ \forall \ \theta \in \Theta$, then

$$P_\theta(U = S^*) = 1 \ \forall \ \theta \in \Theta \tag{4.21}$$

Thus S^* is the unique UMVUE for θ.

This theorem is the culmination of the previous theorems in this section. If a complete sufficient statistic exists, we can find the UMVUE for θ by Rao-Blackwellizing any unbiased estimator with respect to the complete sufficient statistic.

Remark:

1. If a complete sufficient statistic T exists, the UMVUE is a function of T.
2. If the UMVUE exists, it is unique.
3. This theorem can be used to check whether unbiased estimators exist or not. If we can find a complete sufficient statistic T and show that no function of T is unbiased for θ, then no unbiased estimator for θ exists. This means that the UMVUE exists iff an unbiased estimator exists.

In general, we have two approaches to find the UMVUE for θ. First find a complete sufficient statistic T for θ.

1. Find some unbiased estimator S for θ and Rao-Blackwellize it with respect to T.
2. Find a function of T which is unbiased for θ. We also need check that the function has finite variance.

5. Problems

5.1. Minimal Sufficient Statistic

Find minimal sufficient statistic for the following distributions:

1. $\mathcal{N}(\theta, 1), \theta \in \mathbb{R}$
2. $\mathcal{N}\left(0, \sigma^2\right), \sigma^2 > 0$
3. $\mathcal{N}\left(\mu, \sigma^2\right), \mu \in \mathbb{R}, \sigma^2 > 0$
4. $\mathrm{Exp}(\lambda), \lambda > 0$
5. $\mathrm{Poi}(\lambda), \lambda > 0$
6. $\mathrm{Beta}(\alpha, \beta), \alpha > 0, \beta > 0$
7. $\mathrm{Unif}(0, \theta), \theta > 0$
8. $\mathrm{Unif}(\theta, \theta + 1), \theta \in \mathbb{R}$
9. $\mathrm{Unif}(-\theta, \theta), \theta > 0$

5.2. Complete Sufficient Statistic

Check whether the following statistics are complete sufficient statistics:

1. $\mathcal{N}(\theta, 1), \theta \in \mathbb{R}, T(X) = \sum_{i=1}^n x_i$
2. $\mathcal{N}\left(0, \sigma^2\right), \sigma^2 > 0, T(X) = \sum_{i=1}^n x_i^2$
3. $\mathrm{Unif}(0, \theta), \theta > 0, T(X) = X_{(1)}$
4. $\mathrm{Exp}(\lambda), \lambda > 0, T(X) = \sum_{i=1}^n x_i$
5. $\mathrm{Poi}(\lambda), \lambda > 0, T(X) = \sum_{i=1}^n x_i$

5.3. Exponential Family

Check whether the following distributions belong to exponential family.

1. $\text{Gamma}(\alpha, \beta), \alpha > 0, \beta > 0$
2. $\text{Beta}(\alpha, \beta), \alpha > 0, \beta > 0$
3. $\text{Exp}(\lambda)$
4. $\text{Ber}(p)$
5. $\text{Binom}(n, p)$
6. $\text{Geo}(p)$

5.4. Exercise Set 1

Exercise 1: Let $X_1, ..., X_n$ be a sample from the uniform distribution on $(-\theta, \theta)$ ($\theta > 0$). Show that $S = \left(X_{(1)}, X_{(n)}\right)$ is a sufficient statistic for θ and it is not complete. Show that $\max\left(-X_{(1)}, X_{(n)}\right)$ is minimal sufficient. Is it complete? Notice the dimension of the minimal sufficient statistic and the dimension of the parameter.

Hint: Use Fisher Neyman to show sufficiency. S is not complete. An easy proof would be to check by taking different $X_{(1)}$ with same $X_{(n)}$ such that $X_{(n)} > X_{(1)}$.

Exercise 2: Let $X_1, ..., X_n$ be a sample from the distribution with density $f(x; \mu) = e^{-(x-\mu)} 1_{(\mu,\infty)}(x)$ with a parameter $\mu \in \mathbb{R}$. Show that $X_{(1)}$ is a complete sufficient statistic for μ. Use Basu's theorem to show that $X_{(1)}$ and $\left(\frac{1}{n}\right) \sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2$ are independent.

Hint: This is a location family of the exponential distribution.

Exercise 3: Let $Y_1, ..., Y_n$ follow the normal linear regression model, that is, they are independent with distribution $N\left(\beta_0 + \beta_1 x_i, \sigma^2\right)$, where $\sigma > 0$ is known and x_i are some fixed known constants. Find a minimal sufficient statistic for the parameter $(\beta_0, \beta_1)^T$ (find a sufficient statistic using the factorization lemma and then prove its minimality).

Hint: Exponential family.

Exercise 4: Let $X_1, X_2, ..., X_n$ be an i.i.d. sample from the pdf $f(x) = \exp\left[\left(\frac{x-\mu}{\sigma}\right)^4 - \xi(\theta)\right]$, $x \in \mathbb{R}$, where $\theta = (\mu, \sigma) \in \Theta = \mathbb{R} \times \mathbb{R}^+$. Show that a minimal sufficient statistics for θ is $T = \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2, \sum_{i=1}^{n} X_i^3, \sum_{i=1}^{n} X_i^4\right)$.

Hint: Exponential family.

Exercise 5: Based on an i.i.d. sample $X_1, X_2, ..., X_n$, find a minimal sufficient statistic for $\theta = (\mu, \sigma) \in \Theta = \mathbb{R} \times \mathbb{R}^+$ in the family of Cauchy distribution given by the density $f(x) = \left(\frac{\sigma}{\pi}\right)\left(\sigma^2 + (x - \mu)^2\right)^{-1}$, $x \in \mathbb{R}$.

Hint: Stack Exchange Link

Exercise 6: Show that if T is an unbiased estimator of θ (i.e., $E_\theta T = \theta$ for all $\theta \in \Theta$), then T^2 cannot be an unbiased estimator of θ^2.

Hint: Compute $\mathrm{Var}_\theta T$.

Exercise 7: Suppose that $X_1, ..., X_n$ is a random sample from $\mathrm{Poi}(\lambda)$ and $n > 2$. Show that $\left(1 - \frac{2}{n}\right)^{\overline{X}_n}$ is an unbiased estimator of $e^{-2\lambda}$.

Hint: $\sum_{i=1}^n X_i \sim \mathrm{Poi}(n\lambda)$.

Exercise 8: Let T be an unbiased estimator of $\theta \in \Theta$ such that $0 \neq E\left[T^2\right] < \infty$ for all $\theta \in \Theta$. Show that the mean squared error (MSE) $E\left[(aT - \theta)^2\right]$ is minimal for $a = \theta^2 / \left(E\left[T^2\right]\right)$. Consequently, if a does not depend on θ, aT is an estimator, and it is optimal (among estimators of the form aT) in terms of the mean squared error. Apply this result to find the MSE-optimal estimator of σ^2 of the form $b_n \sum_{i=1}^n \left(X_i - \overline{X}\right)^2$ for a sample $X_1, ..., X_n$ from a normal distribution with both parameters unknown.

Exercise 9: Let $X_1, ..., X_n$ be a sample from the uniform distribution on $(0, \theta)$ where $\theta > 0$ is unknown. Consider estimators of θ of the form $\hat{\theta}_b = bX_{(n)}$, $b > 0$. Find the estimator T of this form that has the smallest value of $E_\theta(T - \theta)^2$ and $E_\theta|T - \theta|$ for all values of $\theta > 0$ (if such an estimator exists).

Exercise 10: Consider a sample $X_1, ..., X_n$ from the $\mathrm{Poi}(\lambda)$ distribution with unknown $\lambda > 0$. (a) Consider the parameter $p_0 = P(X = 0) = e^{-\lambda}$. An unbiased estimator of p_0 is $\tilde{p}_0 = 1_{[X_1=0]}$. Find the uniformly minimum variance unbiased estimator (UMVUE) $\hat{p}_0$ of p_0 using Rao-Blackwellization of $\tilde{p}_0$ with respect to the complete sufficient statistic

(b) Consider the parameter $p_1 = \lambda e^{-\lambda}$ and its unbiased estimator $\tilde{p}_1 = 1_{[X_1=1]}$. Find the UMVUE by Rao-Blackwellization.

(c) Consider the parameter $r = e^{-2\lambda}$ and its unbiased estimator $\tilde{r} = 1_{[X_1=0, X_2=0]}$. Find the UMVUE by Rao-Blackwellization.

Exercise 11: Let $X_1, X_2, ..., X_n$ be an i.i.d. sample from $N(\mu, 1)$, where $\mu \in \mathbb{R}$ is unknown. (a) Starting from the computations of $E(\overline{x} - \mu)^k$ for $k = 2, 3, 4$, find the UMVUEs of μ^2, μ^3 and μ^4.

(b) Find a simple estimator of $g(t) = \Pr(X_1 \leq t)$ for a fixed $t \in \mathbb{R}$. Use it to find the UMVUE of $g(t)$. Also, find the UMVUE of $(d/dt)g(t)$.

[Hint: For the first problem in part (b), use the fact that $(X_1, \overline{x})$ has a bivariate Normal distribution. For the second problem in part (b), use the Dominated Convergence Theorem to find an unbiased estimator of $d_t g(t)$ that is based on $\overline{x}$.]

Exercise 12: Let $X_1, X_2, \ldots, X_n$ be an i.i.d sample from Unif$[0, \theta]$, where $\theta > 0$ is unknown. Show that $X_{(n)}$ is complete sufficient for θ. Find the UMVUE of θ^2.

Exercise 13: Let $X_1, \ldots, X_n$ be a sample from the gamma distribution parametrized as

$$f(x; a, p) = \frac{a^p}{\Gamma(p)} x^{p-1} \exp(-ax), x > 0 \tag{5.1}$$

where $p > 0$ is known and $a > 0$ is an unknown parameter. Assume that $n > \frac{2}{p}$. Find the UMVUE of a.

Hint: You may use the fact that the complete sufficient statistic $S = X_1 + \cdots + X_n$ has the gamma distribution with parameters a and np.

Exercise 14: Let $X_1, \ldots, X_n$ be a sample from the distribution with density

$f(x; \theta) = \frac{mx^{m-1}}{\theta^m} 1_{(0,\theta)}(x)$

where $\theta > 0$ is unknown. Find the UMVUE of θ.

Hint: Start with the sufficient statistic for θ.

Exercise 15: Consider a regular parametric problem where we observe a sample $X = (X_1, \ldots, X_n)$ following a distribution with density/frequency $f(x; \theta)$, with $\theta \in \Theta \subseteq \mathbb{R}$. Suppose that we wish to estimate $g(\theta)$ for a measurable function $g : \Theta \to \mathbb{R}$.

(a) Let $T = T(X)$ be an unbiased estimator of $g(\theta)$. Show that $\delta_a := T + aU$ is an unbiased estimator of $g(\theta)$ for any $a \in \mathbb{R}$ if $E_\theta(U) = 0$ for all $\theta \in \Theta$.

(b) Suppose that for a fixed U as in (a) and for some fixed $\theta_0 \in \Theta$, we have $\mathrm{Cov}_{\theta_0}(T, U) \neq 0$. Then, show that there exists $a_0 \neq 0$ such that $\mathrm{Var}_{\theta_0}(\delta_{a_0}) < \mathrm{Var}_{\theta_0}(T)$.

(c) Using (b), prove that T has uniformly (over θ) minimum variance among all unbiased estimators of $g(\theta)$ only if $\mathrm{Cov}_\theta(T, U) = 0$ for all $\theta \in \Theta$ and U satisfying $E_\theta(U) = 0$ for all $\theta \in \Theta$.

(d) Let V be any other unbiased estimator of $g(\theta)$. Use the identity $V = T + (V - T)$ to establish that T has uniformly (over θ) minimum variance among all unbiased estimators of $g(\theta)$ if $\mathrm{Cov}_\theta(T, U) = 0$ for all $\theta \in \Theta$ and for all U satisfying $E_\theta(U) = 0$ for all $\theta \in \Theta$.

(e) Now, assume that T is an uniformly minimum variance unbiased estimator (UMVUE) of $g(\theta)$. Let, if possible, S be another UMVUE of $g(\theta)$ different from T. Define $W = \frac{T+S}{2}$

Use the Cauchy-Schwarz inequality to prove that $\mathrm{Var}_\theta(W) \leq \mathrm{Var}_\theta(T)$ for all $\theta \in \Theta$.

(f) Show that the above inequality cannot be strict for any $\theta \in \Theta$, and then prove that $P_\theta(T = S) = 1$ for all $\theta \in \Theta$, i.e., UMVUE is unique (upto measure zero sets).

[Hint: Use the equality criterion in the Cauchy-Schwarz inequality]

(g) Show that if T is complete sufficient for θ, then $\mathrm{Cov}_\theta(T, U) = 0$ for all $\theta \in \Theta$ and for all U satisfying $E_\theta(U) = 0$ for all $\theta \in \Theta$.

[Hint: Consider the Rao-Blackwellisation of U with respect to T.]