

Notes

# **MA4107: Statistical Inference**

Sabarno Saha

[ss22ms037@iiserkol.ac.in](mailto:ss22ms037@iiserkol.ac.in)

Date: 02 September 2025

# 1. Data and Models

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

$$G : \Omega \longrightarrow \mathbb{R}^n \quad (1.1)$$

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

$\tilde{X}$  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

$$F \in \mathfrak{F} = \{F_\theta : \theta \in \Theta \subset \mathbb{R}^k \text{ for some } k \in \mathbb{N}\} \quad (1.2)$$

$\Theta$  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  $F_\theta$ , there always exists  $\theta$ , such that

$$\theta \xrightarrow{g} F_\theta \quad (1.3)$$

$g$  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  $F_\theta$ .

Unless explicitly mentioned otherwise, the data is represented by  $\tilde{X}$ . 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  $\tilde{X} \in \mathbb{R}^n$ . A statistic  $T = T(\tilde{X})$  is a measureable function of data and data only.

Definition 2.2 (Ancillary Statistic): Suppose the distribution of the test statistic  $T(\tilde{X})$ ,  $\mathfrak{T}$  is independent of the parameter vector  $\theta$ . Then  $T(\tilde{X})$  is called an Ancillary Statistic.

Theorem 2.1:

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

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

is a valid pdf.

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

Definition 2.3 (Location Family of distributions):

Let a family of distributions be,

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

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

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

Definition 2.4 (Scale family of distributions):

Let a family of distributions be,

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

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

Definition 2.5 (Location-Scale family of distribution):

Let a family of distributions be,

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

Then  $\mathfrak{F}$  is a location family of distributions with the standard pdf  $f(x)$  and the location parameter  $\mu$  and scale parameter  $\sigma$  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  $\tilde{X}$  to  $T(\tilde{X})$ . 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,

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

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

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

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.

**Definition 2.6 (Sufficient Statistic):**

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

**Theorem 2.2 (Fisher Neyman Factorization theorem):**

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

$$f_\theta(\tilde{X}) = g(T(\tilde{X}), \theta) h(\tilde{X}) \quad (2.7)$$

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  $\theta$ .

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  $D$ .

**Definition 2.7 (Minimal Sufficient Statistic):**

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

$$T = g(S) \quad (2.8)$$

Lemma 2.3: If  $T_1$  and  $T_2$  are minimal sufficient statistics, there exist injective functions  $g_1$  and  $g_2$  such that,

$$T_1 = g_2(T_2) \quad T_2 = g_1(T_1) \quad (2.9)$$

INSERT PROOF.

Theorem 2.4 (Characterization of minimal sufficiency):

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

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

Then  $T$  is minimal sufficient for  $\theta$ .

Proof:

■

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.

Definition 2.8 (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  $\theta \in \Theta$ ,

$$\mathbb{E}_\theta[h(T)] = 0 \implies P_\theta(h(T) = 0) = 1 \quad (2.11)$$

Unfortunately we have no easy characterization of complete statistic, like [Theorem 2.4](#) for minimal sufficiency.

Theorem 2.5 (Basu's Theorem):

A complete sufficient statistic is independent of any ancillary statistic.

Theorem 2.6 (Lehmann Scheffe):

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

### 3. Families of distribution

Definition 3.1 (Exponential family of distribution):

Let  $\tilde{X}$  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,

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

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

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

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

is free of  $\theta$ .

$c_j$ 's are called Natural or canonical parameters

$T_j$ 's are called Natural or canonical statistics