Lectures 1-5

1 Introduction and Definitions

- The basic inference problem: Population, Sample, Probability model, Parameters.
- Goal is to infer aspects of population from information in sample.
- Types of inference: Estimation, Hypothesis testing
- Sample space Ω. $\mathbf{X} = (X_1, \cdots, X_n)$ be a random vector defined on the sample space. The outcome of the experiment is a realization $\mathbf{x} = (x_1, \cdots, x_n)$ of the random vector $\mathbf{X}$. We call $\mathbf{x}$ the data.
- Typical model:$\mathbf{X}$ has distribution $f(x_1, \cdots, x_n|\theta)$. This distribution is known except for the parameter θ. Given the data $\mathbf{x}$, the goal is to infer the unknown parameter θ.
- $\mathcal{F}$ represents the set of all possible probability distributions for $\mathbf{X}$. We'll call $\mathcal{F}$ the model(or probability model) for the experiment.
- Often the elements of $\mathcal{F}$ are indexed by one or more parameters. We'll often denote a vector of parameters by θ and let Θ be the collection of all possible values of θ. Θ is called the parameter space.
- If $\mathcal{F}$ can be expressed as a collection of distributions indexed by finite dimensional vectors $\Theta = (\theta_1, \cdots, \theta_k)$, where Θ is a subset of $\mathbb{R}^k$, then $\mathcal{F}$ will be called a parametric family. If $\mathcal{F}$ cannot be so expressed, it will be called nonparametric.
- Suppose $\theta = (\theta_1, \theta_2)$. If θ_1 is the only parameter of interest, then θ_2 is called a nuisance parameter.
- A model is said to be identifiable if $F_{\theta_1} = F_{\theta_2}$ whenever $\theta_1 = \theta_2$.
- Let T be a real-valued or vector-valued function whose domain contains the range of $\mathbf{X}$. If T does not depend on the unknown parameter θ, then $T = T(\mathbf{X})$ is called a statistic. The probability distribution of T is called its sampling distribution.

Example 1 *Have a population of N items, possibly a shipment of manufactured goods. An unknown number M of the N items are defective. A random sample of size n is drawn without replacement and inspected. Let X be the number of defectives in the sample.*

Example 2 *There are unknown number N number of fish in a pond. You catch M of them, tag them and let them go. Allow them to mingle for a while. Then you catch n fish and note the number of tagged ones among them. Let X be the number of tagged fish in the recaptured sample.*

Example 3 *Experimenter makes n independent determinations of the value of a physical constant μ and measurements are subject to error. $X_1, \cdots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$.*

Example 4 *Let $\mathcal{F}$= family of all continuous distributions that are symmetric about 0. Then $\mathcal{F}$ is a nonparametric family.*

2 Sufficiency for data reduction

[CB6.2, BD 1.5]

2.1 Sufficiency

Definition 1 *A statistic $T(\mathbf{X})$ is a sufficient statistic if the conditional distribution of $\mathbf{X}$ given $T(\mathbf{X}) = t$ does not depend on θ, regardless of what t is.*

Example 5 *Suppose $X_1, \cdots, X_n$ are i.i.d. Poisson with mean θ. Then $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is a sufficient statistic.*

- Basic idea of sufficiency: Given data $\mathbf{X} = (X_1, \cdots, X_n)$, can we find a statistic $T(X)$ of smaller dimension than n that contains as much information about θ as $\mathbf{X}$ does? If a statistic exists, we can reduce (perhaps greatly) the amount of data without throwing away information. The search for good estimation and testing procedures can be narrowed.

- We can think of a partition of the sample space where each set A_t in the partition is such that $T(\mathbf{x}) = t$ for each $\mathbf{x} \in A_t$. All $\mathbf{x} \in A_t$ are equivalent in that each one contains the same information about θ as the others.

- If T_1 and T_2 are any two statistics such that $T_1(x) = T_1(y)$ if and only if $T_2(x) = T_2(y)$, then T_1 and T_2 are said to be equivalent.

- Sufficiency Principle: Consider sample $\mathbf{X}$ from model $\mathcal{F}$, and let $T(\mathbf{X})$ be a sufficient statistic. Suppose experimenter 1 observes $\mathbf{X} = x$ while experimenter 2 observes$\mathbf{X} = y$. If $T(x) = T(y)$, then experimenters 1 and 2 should make the same inference about θ.

Theorem 1 *(Fisher-Neyman Factorization Theorem): Let $f(\mathbf{x}|\theta)$ denote the joint pdf or pmf of the data $\mathbf{X}$. A statistic $T(\mathbf{X})$ is sufficient if and only if there exist functions $g(t|\theta)$ and $h(\mathbf{x})$(where h does not depend on θ) such that $f(\mathbf{x}|\theta) = g(T(\mathbf{x})|\theta)h(\mathbf{x})$ for all $\mathbf{x}$ and all parameter values θ.*

Example 6 *Estimating the Size of a Population: Consider a population with N members labeled consecutively from I to N. The population is sampled with replacement and n members of the population are observed and their labels $X_l, \cdots, X_n$ are recorded. Then $X_{(n)}$ is indeed sufficient.*

Example 3 (revisited): $X_1, \cdots, X_n$ are i.i.d. $\mathcal{N}(\mu, \sigma^2)$ and $\theta = (\mu, \sigma^2)$.
$T(X) = (\sum X_i, \sum X_i^2)$ is jointly sufficient for θ.
Qn: Is the dimension of a sufficient statistic the always same to the dimension of the parameters?
HW: Eg 1.5.5 of BD: Linear Regression
Let $f_X(x|\theta)$ be the joint pdf or pmf of $\mathbf{X}$ and $q(t|\theta)$ be the pdf or pmf of $T(\mathbf{X})$. Then T is a sufficient statistic for θ, iff, for every x, the ratio $f_X(x|\theta)/q(T(x)|\theta)$ is constant as a function of θ.

Example 7 *Suppose we observe* $\mathbf{X} = (X_1, \cdots, X_n)$*, where*

$$X_i = \rho X_{i-1} + Z_i, \quad i = 2, 3, \cdots, n.$$

The quantity ρ *is an unknown parameter such that* $|\rho| < 1$*.* $Z_2, \cdots, Z_n$ *are i.i.d.* $\mathcal{N}(0, \sigma^2)$*, where* σ^2 *is another unknown parameter.* $X_1 \sim \mathcal{N}(0, \sigma^2/(1-\rho^2))$ *and* $X_1, Z_2, \cdots, Z_n$ *are mutually independent.*
The parameter space is $\Theta = (\rho, \sigma^2) : |\rho| < 1, \sigma^2 > 0$*.*
This model is called an autoregressive model and is used in time series analysis.

$$f(\mathbf{x}|\rho, \sigma) = (2\pi\sigma^2)^{-n/2}\sqrt{1-\rho^2}e^{\left\{-\frac{1}{2\sigma^2}(x_1^2(1-\rho^2)+\sum_{i=2}^n(x_i-\rho x_{i-1})^2)\right\}}.$$

$T_1(\mathbf{X}) = \sum_{i=2}^{n-1} X_i^2, T_2(\mathbf{X}) = \sum_{i=2}^n X_i X_{i-1}$ *and* $T_3(\mathbf{X}) = X_1^2 + X_n^2$ *are jointly sufficient statistics.*

Proposition 1 *Let* $T(\mathbf{X}) = (T_1(\mathbf{X}), \cdots, T_k(\mathbf{X}))$ *be a sufficient statistic and* r *be a 1-1 function, not depending on* θ *and with domain equal to the range of* $T(\mathbf{X})$*. Then* $r(T(\mathbf{X}))$ *is a sufficient statistic.*

2.2 Minimal Sufficiency

Definition 2 *: A statistic* $T(\mathbf{X})$ *is a minimal sufficient statistic if it is a function of every other sufficient statistic.*

Theorem 2 *Let* $f(\mathbf{x}|\theta)$ *be the pdf or pmf of* $\mathbf{X}$*. Suppose there exists a statistic* $T(\mathbf{X})$ *such that, for any two points* $\mathbf{x}$ *and* $\mathbf{y}$*, the ratio* $f(\mathbf{x}|\theta)/f(\mathbf{y}|\theta)$ *is constant as a function of* θ *iff* $T(\mathbf{x}) = T(\mathbf{y})$*. Then* $T(\mathbf{X})$ *is a minimal sufficient statistic.*

Example 8 $X_1, \cdots, X_n$ *iid Unif(*$\theta, \theta+1$*). Then* $T(\mathbf{X}) = (X_{(1)}, X_{(n)})$ *is minimal sufficient.*

Proposition 2 *If* $T(X)$ *is a minimal sufficient statistic for* θ*, then its one-to-one function is also a minimal sufficient statistic for* θ*.*

Proposition 3 *There is always a one-to-one function between any two minimal sufficient statistics.*

Example 3 (revisited): $T_1(\mathbf{X})$=$(\bar{X}, S^2)$ is minimal sufficient.
HW: For $X_1, \cdots, X_n$ iid from cauchy distn, show that the minimal sufficient statistics is the order statistics. Does the order-statistics provide any data reduction?

2.3 Ancillarity

Definition 3 *A statistic $S(\mathbf{X})$ is an ancillary statistic if its distribution does not depend on θ.*

Example 8 continued: The range is ancillary.

Definition 4 *Let $f(x)$ be any pdf. Then for any $\mu \in \mathbb{R}$ and any $\sigma > 0$ the family of pdfs $g(x) = f((x-\mu)/\sigma)/\sigma$, indexed by the parameter (μ, σ) is called the location-scale family with standard pdf $f(x)$, and μ is called the location parameter and σ is called the scale parameter for the family.*

HW: In the above definition g is indeed a pdf.
HW: X is a random variable with pdf f if and only if there exists a random variable Z with pdf g and $X = \sigma Z + \mu$.
HW: Let $X_1, \cdots, X_n$ be iid from a location family. Show that the range is an ancillary statistic. Can you think of another ancillary statistic?
HW: Let $X_1, \cdots, X_n$ be iid from a scale family. Show that the following statistic $T(\mathbf{X})$ is ancillary. $T(\mathbf{X}) = (X_1/X_n, \cdots, X_{n-1}/X_n)$.

2.4 Completeness

Definition 5 *Let $f_T(t|\theta)$ be a family of pdfs or pmfs for a statistic $T(\mathbf{X})$. The family of probability distributions is called complete if $E[g(T)|\theta] = 0$ for all θ implies $Pr[g(T) = 0|\theta] = 1$ for all θ. equivalently T is a complete statistic.*

Example 5 revisited: In the Poisson eg, restrict $\Theta = \{1, 2\}$. Then $g(0) = 2, g(2) = 2, g(1) = -2$ and 0 otherwise is a function that has expectation zero for all θ. Thus the family is not complete. When $\Theta = \mathbb{R}^+$, then the family is complete.

Proposition 4 *For a statistic T(X), if a non-constant function of T, say r(T) is ancillary, then T(X) cannot be complete.*

Proposition 5 *If T(X) is a complete statistic, then a function of T, say $T^* = r(T)$ is also complete.*

Proposition 6 *If a complete sufficient statistic exists, then a minimal sufficient statistic is complete.*

Theorem 3 *(Basu 1955) If T(X) is complete and minimal sufficient statistic, then T(X) is independent of every ancillary statistic.*

HW: For exponential distribution, find $E(X_1/(X_1 + \cdots + X_n))$