INTRODUCTION TO STATISTICAL INFERENCE

Sangita Das, notes by Ramdas Singh

Third Semester

List of Symbols

Placeholder

Contents

Chapter 1

SUFFICIENCY

1.1 Introduction to Sufficient Statistics

We start by defining terms for the sake of completion, whilst assuming the most basic definitions.

Definition 1.1. An *estimator* is any function of the random sample which is used to estimate the unknown value of the given paramteric function $g(\theta)$.

If $\underline{X} = (X_1, \ldots, X_n)$ is a random sample from a population with a probability distribution P_θ, a function $d(X)$ used for estimating $g(\theta)$ is known as an estimator. Let $\underline{x} = (x_1, \ldots, x_n)$ be a realization of $\underline{X} = (X_1, \ldots, X_n)$. Then $d(\underline{x})$ is called an *estimate*.

Definition 1.2. The *parameter space* is the set of all possible values of a parameter.

For example, the normal distribution $N(\mu, \sigma^2)$ has the parameter space $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Similarly, the binomial distribution $\mathrm{Bin}(n, p)$ has the constraints $n \in \mathbb{N}$ and $p \in [0, 1]$.

Throughout this course, we will assume any data, otherwise stated, will be *independent and identically distributed*; the are separate datapoints that follow the same probability distribution and are indepedent.

Definition 1.3. Let $X_1, \ldots, X_n$ be a random sample from a population P_θ, where $\theta \in \Theta$. A statistic $T = T(X_1, \ldots, X_n) = T(\underline{X})$ is said to be a *sufficient statistic* for the family $\mathcal{P} = \{P_\theta : \theta \in \Theta\}$ if the conditional distribution of $X_1, \ldots, X_n$ given $T = t$ is independent of θ.

We shall look at some examples.

Example 1.4. Let $X_1, \ldots, X_n$ be a random sample from the Bernoulli distribution with parameter $p \in (0, 1)$. We claim that $T = \sum_{i=1}^n X_i$ is sufficient for $\{\mathrm{Ber}(p) \mid 0 < p < 1\}$. To show this, we simply have

$$
\begin{aligned}
P(X_i = x_i \text{ for all } i | T = t) &= \frac{P(X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = t - \sum_{i=1}^{n-1} x_i)}{P(\sum_{i=1}^n X_i = t)} && (1.1) \\
&= \frac{P(X_1 = x_1) \cdots P(X_{n-1} = x_{n-1}) \cdot P(X_n = t - \sum_{i=1}^{n-1} x_i)}{\binom{n}{t} p^t (1-p)^{n-t}} \\
&= \frac{p^{x_1}(1-p)^{1-x_1} \cdots p^{x_{n-1}}(1-p)^{1-x_{n-1}} p^{t - \sum x_i}(1-p)^{1-t+\sum x_i}}{\binom{n}{t} p^t (1-p)^{n-t}} \\
&= \frac{1}{\binom{n}{t}}. && (1.2)
\end{aligned}
$$

Thus, the statistic T is sufficient. The above expression is valid when $\sum_{i=1}^n x_i = t$, and the probability evaluates to 0 if $\sum_{i=1}^n x_i \neq t$.

Example 1.5. Let $X_1, \ldots, X_n$ be a random sample from Poisson(λ) for $\lambda > 0$. We claim that the statistic $T = \sum_{i=1}^n X_i$ is sufficient. Recall that the probability mass function is $f(x, \lambda) = \frac{e^{-\lambda}\lambda^x}{x!}$ where x is a non-negative integer, and $\lambda > 0$. We have

$$
\begin{aligned}
P(X_i = x_i \mid T = t) &= \frac{P(X_1 = x_1)\cdots P(X_{n-1} = x_{n-1}) \cdot P(X_n = t - \sum_{i=1}^{n-1} x_i)}{P\left(\sum_{i=1}^n X_i = t\right)} \qquad (1.3) \\
&= \frac{\frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\cdots\frac{e^{-\lambda}\lambda^{x_{n-1}}}{x_{n-1}!}\cdot\frac{e^{-\lambda}\lambda^{t-\sum x_i}}{(t-\sum x_i)!}}{\frac{e^{-n\lambda}(n\lambda)^t}{t!}} \\
&= \frac{e^{-n\lambda}\lambda^t}{x_1!\cdots x_{n-1}!(t-\sum_{i=1}^{n-1} x_i)!}\cdot\frac{t!}{e^{-n\lambda}(n\lambda)^t} \\
&= \frac{t!}{x_1!\cdots x_{n-1}!\left(t-\sum_{i=1}^{n-1} x_i\right)!}\cdot\frac{1}{n^t} \\
&= \binom{t}{x_1, x_2, \ldots, x_n}\cdot\frac{1}{n^t}. \qquad (1.4)
\end{aligned}
$$

This shows that the conditional distribution of $(X_1, \ldots, X_n)$ given $T = t$ does not depend on λ, so by the definition of sufficiency, T is a sufficient statistic for λ.

Definition 1.6. A *regular model* may be one of two things.

1. All P_θ are continuous with probability density function $f(x \mid \theta)$.
2. All P_θ are discrete with prbability mass function $p(x \mid \theta)$, and there exists a countable set $S = \{x_1, x_2, \ldots\}$ independent of θ such that $\sum_{i=1}^\alpha p(x_i|\theta) = 1$.

1.2 Factorization Theorems

The following theorem proves to be useful for finding sufficiency.

Theorem 1.7 (The *Neyman-Fisher factorization theorem*). *Let $f(\underline{x} \mid \theta)$ be the density of $\underline{X}$ under the probability model P_θ for $\theta \in \Theta$. Then if the model is regular, a statistic $T(\underline{X})$ is sufficient for θ if and only if there exist functions g and h such that*

$$f(\underline{x} \mid \theta) = g(T(\underline{x}), \theta)h(\underline{x}). \qquad (1.5)$$

Note that the functions are defined with $T : \mathbb{R}^n \to I \subseteq \mathbb{R}^k$ (for $k \leq n$), $g : I \times \Theta \to \mathbb{R}_{\geq 0}$, and $h : \mathbb{R}^n \to R_{\geq 0}$. The functions g and h need not be unique.

A little less formally, the theorem basically states this: let X be a random variable with probability mass/density function $f(x, \theta)$ for $\theta \in \Theta$. Then $T(X)$ is sufficient if and only if $f(x, \theta) = g(T(x), \theta)h(x)$ for all $\theta \in \Theta$.

Example 1.8. Let $X_1, \ldots, X_n$ be independent and identically distributed $N(\mu, \sigma^2)$ random variables, with $\mu \in \mathbb{R}$ and $\sigma^2 > 0$. Let us find a sufficient test statistic. We look at cases; the first case being when σ^2 is known ($\sigma^2 = 1$). Since these are independent, we have the joint probability density funciton of these random variables as

$$f(x_1,\ldots,x_n \mid \mu,\sigma^2) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma} e^{-\dfrac{(x_i-\mu)^2}{2\sigma^2}} = \frac{1}{(2\pi)^{n/2}} e^{-\sum_{i=1}^n (x_i-\mu)^2} \tag{1.6}$$

$$= \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\left(\sum_{i=1}^n x_i^2 - 2\mu\sum_{i=1}^n x_i + n\mu^2\right)\right)$$

$$= \frac{1}{(2\pi)^{n/2}} e^{-\dfrac{1}{2}\sum_{i=1}^n x_i^2} \times e^{-\dfrac{1}{2}(-2\mu\sum_{i=1}^n x_i + n\mu^2)}. \tag{1.7}$$

Make the former term $h(x)$ and the latter term $g(\sum_{i=1}^n x_i, \mu)$ with $T(x) = \sum_{i=1}^n x_i$. The second case now involves μ being known, and we set it to $\mu = 0$ to get $T(x) = \sum_{i=1}^n x_i^2$, $h(x) = 1/(2\pi)^{n/2}$, and $g(T(x), \sigma^2) = \sigma^{-n} e^{-T(x)/2\sigma^2}$.

We move on to another factorization theorem.

Definition 1.9. The family of distributions $\{P_\theta \mid \theta \in \Theta\}$ is said ot be a *single parameter exponential family* if there eixst real valued functions $c(\theta), d(\theta)$ on Θ and $T(x), S(x)$ on $\mathbb{R}^n$ and a set $A \subset \mathbb{R}^n$ such that

$$f(\underline{x} \mid \theta) = \exp(c(\theta)T(\underline{x}) + d(\theta) + S(x))\mathbf{1}_A(x) \tag{1.8}$$

where A must not depend on θ.

Example 1.10. Suppose $X \sim \text{Poisson}(\lambda)$ for $\lambda > 0$. With $A = \{0, 1, 2, \ldots\}$, we have

$$f(x \mid \lambda) = \exp(x\log(\lambda) - \lambda - \log(x!))\mathbf{1}_A(x) \tag{1.9}$$

with $T(x) = x$, $c(\lambda) = \log(\lambda)$, $d(\lambda) = -\lambda$, and $S(x) = -\log(x!)$.

Consider $X_1, \ldots, X_n$ independent and identically distributed random variables following the distribution P_θ, and suppose that $\{P_\theta \mid \theta \in \Theta\}$ is an exponential family, that is, $f(x \mid \theta) = \exp(c(\theta)T(x_i) + d(\theta) + S(x))\mathbf{1}_A(x)$. Then,

$$f_{x_1,\ldots,x_n}(x_1,\ldots,x_n \mid \theta) = \prod_{i=1}^n \exp(c(\theta)T(x_i) + d(\theta) + S(x_i))\mathbf{1}_A(x_i) \tag{1.10}$$

$$= \exp(c(\theta)\sum_{i=1}^n T(x_i) + md(\theta) + \sum_{i=1}^n S(x_i))\mathbf{1}_{A^n}(x_1,\ldots,x_n). \tag{1.11}$$

$(x_1, \ldots, x_n)$ has distribution belonging to a single parameter exponential family. Thus, if $\{P_\theta \mid \theta \in \Theta\}$ is a single parameter family with density $f(x,\theta) = \exp(c(\theta)T(x) + d(\theta) + S(x))\mathbf{1}_A(x)$, then $T(x)$ is sufficient for θ.

Corollary 1.11. *If $x_1, \ldots, x_n$ are independent and indentically distributed random variables following the distribution P_θ with density $f(x \mid \theta) = \exp(c(\theta)T(x) + d(\theta) + S(x))\mathbf{1}_A(x)$, then $\sum_{i=1}^n T(X_i)$ is sufficient for θ.*

The exponential family is expanded.

Definition 1.12. A family of distributions $\{P_\theta : \theta \in \Theta\}$ with density $f(x \mid \theta)$ is called a *k-parameter exponential family* if there exists real valued functions $c_1(\theta), \ldots, c_k(\theta), d(\theta)$ on Θ and

$T_1(\underline{x}), \ldots, T_k(\underline{x}), S(x)$ on $\mathbb{R}^n$, and a set $A \subset \mathbb{R}^n$ such that

$$f(\underline{x} \mid \theta) = \left(\exp(\sum_{j=1}^{n} c_j(\theta) T_j(\underline{x}) + d(\theta) + S(\underline{x})) \right) \mathbf{1}_A(\underline{x}). \tag{1.12}$$

Here, $(T_1, \ldots, T_k)$ is a k-dimensional sufficient statistic for θ. Note that the parameter here is θ and not $(c_1(\theta), \ldots, c_k(\theta))$.

We look at more examples.

Example 1.13. For a normal distribution with $\sigma^2 = 1$, we have

$$f(x \mid \theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\theta)^2}{2}} \mathbf{1}_A(x) = \exp\left(-\frac{1}{2}\log(2\pi) - \frac{x^2}{2} + x\theta - \frac{\theta^2}{2} \right) \mathbf{1}_A(x). \tag{1.13}$$

Here, $c(\theta) = \theta$, $T(x) = x$, $S(x) = -\frac{x^2}{2} - \frac{1}{2}\log(2\pi)$, and $d(\theta) = -\frac{\theta^2}{2}$.

Remark 1.14.

1. The Neyman-Fisher factorization theorem holds if $\underline{\theta}$ and $\underline{T}$ are vectors. Their dimensions need not be equal.
2. If T is sufficient and T is a function of U, then U is also sufficient.
3. If V is a function of T, then V need not be sufficient. But if V is one-to-one with T, then V is also sufficient. $V = B(T)$ and $T = B^{-1}(V)$ shows that $g(T, \theta) = g(B^{-1}(V), \theta) = g^*(V, \theta)$.

1.3 Minimal Sufficiency

Again, we being with a few definitions.

Definition 1.15. A *partition* of a space $\mathcal{H}$ is a collection $\{E_i\}$ of subsets of $\mathcal{H}$ such that

$$\bigcup_{n \geq 1} E_i = \mathcal{H} \text{ and } E_i \cap E_j = \emptyset \text{ for } i \neq j. \tag{1.14}$$

The E_i's are called *partition sets*. Let $T : \mathcal{H} \to \mathcal{J}$. The partition of $\mathcal{H}$ induced by the function T is the collection of the sets $T_y = \{x \mid T(x) = y\}$ for $y \in \mathcal{J}$.

We say that $\mathcal{P}_2$ is a *reduction* of $\mathcal{P}_1$ if each partition set of $\mathcal{P}_2$ is the union of the same members of $\mathcal{P}_1$.

Index