# Statistical Inference I

MAT-3102-1
Spring 2024

Kahaan Shah

May 30, 2024

# Contents

# Chapter 1

# Introduction

## Lecture 1

23 Jan. 08:30

**Exercise.** Two unbiased dice are thrown and with $X_1$ is the number observed on the first die and $X_2$ as the number observed on the 2nd die. Write down the joint probability distribution of $(X_1 \wedge X_2, X_1, \vee X_2)$. Write down the marginal pmf and the conditional pmf $X_1 \wedge X_2 \mid X_1 \vee X_2 = 5$.

| | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| 1 | 1/36 | 2/36 | 2/36 | 2/36 | 2/36 | 2/36 |
| 2 | 0 | 1/36 | 2/36 | 2/36 | 2/36 | 2/36 |
| 3 | 0 | 0 | 1/36 | 2/36 | 2/36 | 2/36 |
| 4 | 0 | 0 | 0 | 1/36 | 2/36 | 2/36 |
| 5 | 0 | 0 | 0 | 0 | 1/36 | 2/36 |
| 6 | 0 | 0 | 0 | 0 | 0 | 1/36 |

Table 1.1: Bivariate Table

**Answer.** To find the marginal distribution we just sum along a row or a column. Now write down the conditional pmf of $X_1 \wedge X_2 \mid X_1 \vee X_2 = 5$. This is given by $\frac{P(X_1 \wedge X_2 = i, X_1 \vee X_2 = 5)}{P(X_1 \vee X_2 = 5)}$. So this $\frac{2}{36}$ except when $i = 5$ then it is $\frac{1}{36}$. We have $X_1, X_2$ as independent and identical distributed (iid) but $X_1 \wedge X_2$ and $X_1 \vee X_2$ are not independent. ⊛

# Chapter 2

# Estimators

We assume that there is a population with a known distribution family but with unknown parameters characterizing this distribution. Our job is to estimate this unknown parameter based on the random samples selected from the population. We will denote $F_\theta$ where $F$ is the unknown distribution characterized by $\theta$.

## 2.1 Some definitions

**Definition 2.1.1** (Parameter). The parameter is a function of all the population observations. Some examples are population mean and variance.

**Definition 2.1.2** (Parameter Space). This is the set of admissible values for the parameter. We denote this by $\Theta$.

Consider $(X_1, \dots X_n)$ selected from $F_\theta$. If $\theta$ is the mean then the natural guess is that the sample mean is the same as the population mean, but this is just a guess.

Suppose now that our job is to estimate some function of $\theta$, given by $g(\theta)$. We are now proposing $T(X_1, \dots X_n)$ as an estimator of $g(\theta)$. We want that $T(X_1, \dots X_n)$ is very close to $g(\theta)$. So we want $\mathbb{E}((T(X_1, \dots X_n) - g(\theta))^2)$ which is the mean squared error of $T$ for estimating $g(\theta)$.

Suppose we have 2 estimators $T_1, T_2$ to estimate $g(\theta)$. We will say $T_1$ is better than $T_2$ for estimating $g(\theta)$ if the MSE of $T_1 \leq$ MSE of $T_2$.

## Lecture 2

## 2.2 Evaluating Estimators

25 Jan. 08:30

**Definition 2.2.1** (Minimum MSE estimator). We say $T_1$ is an estimator for $g(\theta)$ with minimum MSE if $MSE_\theta(T_1) \leq MSE_\theta(T_2)$ where $T_2$ is any other estimator of $g(\theta)$, $\forall \theta \in \Theta$.

**Remark.** Recall that for a random variable $X$ with $\mathbb{E}(X^2) < \infty \Rightarrow V(X) < \infty$. We know $V(X) = 0 \Leftrightarrow X = c$ with probability 1. Suppose we have distribution $N(\theta, 1)$. Suppose $T_1(X_1, \dots X_n)$ is some good estimator and $T_2(X_1, \dots X_n) = 1$. We have $MSE_1(T_2) = \mathbb{E}_1(1-\theta)^2 = (1-1)^2 = 0$. We also have $MSE_\theta(T_1) = \mathbb{E}_\theta(T_1 - \theta)^2$. So we see that no matter how good $T_1$ is, we see that $T_2$ beats $T_1$ at $\theta = 1$. So we cannot hope to achieve an estimator with minimum MSE.

**Definition 2.2.2** (Unbiased Estimator). Suppose $T(X_1, \dots X_n) = T$ is an estimator for $g(\theta)$. We say that $T$ is an unbiased estimator for $g(\theta)$ if $\mathbb{E}_\theta(T) = g(\theta), \forall \theta \in \Theta$. We write $\mathcal{U}(g(\theta))$ to denote the class of all unbiased estimators of $g(\theta)$.

**Example.** Consider $X_1 \dots X_4 \sim \text{Bin}(10, p)$. We have $\frac{X_1}{10}$ as an unbiased estimator ($\mathbb{E}(X_1) = 10p$). Similarly $\frac{X_1+X_2}{20}$ is also unbiased by linearity of expectation. We also have $\frac{\lambda(X_1)+(1-\lambda)X_2}{10}$ as an unbiased estimator.

**Lemma 2.2.1.** Suppose we have $T_1, T_2$ as unbiased estimators of $g(\theta)$, then we have $\lambda(T_1)+(1-\lambda)T_2$ where $\lambda \in \mathbb{R}$ is an unbiased estimator of $g(\theta)$. So we have that if there exists $T_1, T_2 \in \mathcal{U}$ then cardinality of $\mathcal{U}$ is $\infty$.

**Proof.** $\mathbb{E}(\lambda(T_1) + (1-\lambda)T_2) = \lambda g(\theta) + (1-\lambda)g(\theta) = g(\theta)$. ■

**Lemma 2.2.2.** The MSE of an unbiased estimator is its variance. So we are trying to find an estimator with minimum variance.

**Proof.** Consider $MSE_\theta(T_1) = \mathbb{E}_\theta(T_1 - g(\theta))^2 = \mathbb{E}_\theta(T_1 - \mathbb{E}_\theta(T_1))^2 = \text{Var}_\theta(T_1)$. ■

**Definition 2.2.3** (Minimum variance unbiased estimator (MVUE)). $T_1$ is a MVUE if

1. $T_1 \in \mathcal{U}$ (or $\mathbb{E}_\theta(T_1) = g(\theta), \forall \theta \in \Theta$ ).
2. $\text{Var}_\theta(T_1) \leq \text{Var}_\theta(T_2)$ where $T_2$ is any other unbiased estimator of $g(\theta)$.

**Remark.** Suppose we have $(X_1 \dots X_4) \sim \text{Bin}(10, p)$. Now consider $g(p) = 2p + 3$. We can obtain an unbiased estimator for it by obtaining an unbiased estimator of $p$ and applying the same linear transformation. Now suppose we have a non-linear function $g(p) = p(1-p)$. Consider $\frac{X_1(10-X_2)}{100}$. Since they are independent we get $p(1-p)$.

**Definition 2.2.4** (Indicator random variable). An indicator random variable $1_A$ is 1 if $\omega = A$ otherwise it is 0. We have $\mathbb{E}(1_A) = P(A)$.

**Example.** Consider $X_1 \dots X_4 \sim \text{Ber}(p)$ with $g(p) = p(1-p)$. This is given by the event $X_1 = 1, X_2 = 0$. We have that $1_{X_1=1, X_2=0} \in \mathcal{U}(p(1-p))$.

**Definition 2.2.5** (Bias). Suppose $T$ is any estimator of $g(\theta)$. The difference between its expectation and $g(\theta)$ given by $|E_\theta(T) - g(\theta)|$ is called the bias of $T$ and denoted by $B_\theta(T)$.

**Lemma 2.2.3.** We have $MSE_\theta(T) = \text{Var}_\theta(T) + B_\theta^2(T)$

**Proof.** Consider

$$\begin{aligned} MSE_\theta(T) &= \mathbb{E}_\theta(T - g(\theta))^2 \\ &= \mathbb{E}_\theta(T - \mathbb{E}_\theta(T) + \mathbb{E}_\theta(T) - g(\theta))^2 \\ &= \mathbb{E}_\theta(T - \mathbb{E}_\theta(T))^2 + B_\theta^2(T) + 2(\mathbb{E}_\theta(T) - g(\theta)\mathbb{E}_\theta(T - \mathbb{E}_\theta(t))) \\ &= \text{Var}_\theta(T) + B_\theta^2(T) \end{aligned}$$

■

# Lecture 3

6 Feb. 08:30

**Exercise.** Consider $X_1 \dots X_n$ iid $\sim P(\lambda)$. Find an unbiased estimator of $g(\lambda) = e^{-\lambda}$.

**Answer.** We can use the indicator function $\mathbf{1}_{(X_1=0)}$. ⊛

**Note.** Note that $F_{X(n)}(x) = \mathbb{P}(X_{(n)} \leq x) = \mathbb{P}(\cap_{i=n}^{n}(X_i \leq x))$

**Exercise.** Consider $X_1 \dots X_n \sim U(0, \theta)$ and give an unbiased estimator of $\theta$. Also give an unbiased estimator as a function of $X_{\max} = X_{(n)} = \prod_{i=1}^{n} \mathbb{P}(X_i \leq x)$.

**Answer.** $2X_1$ since $\mathbb{E}(X_1) = \frac{\theta}{2}$. From the formula above we have $F_{X(n)}(x) = (\frac{x}{\theta})^n$. Now to obtain the distribution function we differentiate to obtain $\frac{n}{\theta}(\frac{x}{\theta})^{n-1}$ between 0 to $\theta$. So the expectation is given by $\frac{n}{\theta^n}\int_0^\theta x^n = \theta\frac{n}{n+1}$. So an unbiased estimator of $\theta$ is given by $\frac{n+1}{n}X_{(n)}$. ⊛

**Definition 2.2.6** (Consistent Estimators)**.** $T(X_1 \dots X_n)$ is a consistent estimator of $g(\theta)$ if it is

1. Asymptotically unbiased: $\mathbb{E}_\theta(T(X_1, \dots X_n)) \to g(\theta)$
2. $\mathrm{Var}_\theta(X_1, \dots X_n) \to 0$

**Example.** $X_{(n)}$ is a consistent estimator of $\theta$.

**Exercise.** Consider $X_1 \dots X_n \sim P(\lambda)$. Check whether $\overline{X}$ is a consistent estimator as $n \to \infty$ or not.

**Exercise.** Consider $X_1 \dots X_n \sim U(0, \theta)$. Find the pdf of $X_{(1)} = X_{\min}$.

**Answer.** We have $1 - F_{X_{(1)}}(x) = P(X_{(1)} > x)$. ⊛

**Exercise.** Consider $X_1, X_2 \sim \mathrm{Ber}(p)$ iid. Give an estimator for $g(p) = p^3$.

**Answer.** Consider $\sum_{x_1,x_2} T(x_1, x_2)P(X_1 = x_1, X_2 = x_2) = \sum_{x_1,x_2} T(x_1, x_2)p^{x_1}(1-p)^{1-x_1}p^{x_2}(1-p)^{1-x_2} = p^3$. But in this setup we can never get anything over $p^2$ so there is no unbiased estimator for $p^3$. ⊛

## 2.3 Moment Generating Functions

**Remark.** The $r$th moment is given by $\mathbb{E}(X^r)$ if $\mathbb{E}(|X|^r) < \infty$. Note that if $r \leq s$ and the $s$th moment is finite then the $r$th moment is also finite. On the other hand if the $r$th moment does not exist then $s$th moment also does not exist.

**Definition 2.3.1** (Moment generating function)**.** The MGF of a random variable $X$ is defined as $M_X(t) = \mathbb{E}(e^{tX})$ if $\mathbb{E}(e^{tX}) < \infty$ for $|t| \leq a$ for some $a > 0$.

**Remark.** We have

$$\begin{aligned}\mathbb{E}(e^{tX}) &= \sum_x e^{tx}\mathbb{P}(X=x) \\ &= \sum_x (1+\frac{tx}{1!}+\frac{(tx)^2}{2!}\dots)\mathbb{P}(X=x) \\ &= 1+\sum_x \frac{tx\mathbb{P}(X=x)}{1}+\sum_x \frac{t^2x^2\mathbb{P}(X=x)}{2!}+\dots \\ &= 1+\frac{t\mathbb{E}(x)}{1!}+\frac{t^2\mathbb{E}(x^2)}{2!}+\dots\end{aligned}$$

So we see that:

1. If MGF exists then it implies that $X$ has moments of all orders.
2. If $X$ does not have $r$-th order moment for some $r \geq 1$ then MGF$(X)$ does not exist.

**Lemma 2.3.1.** $\frac{\mathrm{d}^r}{\mathrm{d}t^r}(\mathrm{MGF}_X(t))\mid_{t=0}=\mathbb{E}(X^r)$.

**Remark.** If the MGF exists it uniquely specifies moments of $X$. If MGF$(X)$ = MGF$(Y)$ then for practical purposes it implies that $X$ and $Y$ have the same distribution (this is not true for the most general case).

**Lemma 2.3.2.** $\mathrm{MGF}_{X+Y}(t) = \mathrm{MGF}_X(t)\mathrm{MGF}_Y(t)$

**Proof.**

$$\begin{aligned}\mathrm{MGF}_{X+Y}(t) &= \mathbb{E}(e^{t(X+y)}) \\ &= \mathbb{E}(e^{tX}e^{tY}) \\ &= \mathbb{E}(e^{tX})\mathbb{E}(e^{tY}) \\ &= RHS\end{aligned}$$

■

**Lemma 2.3.3.** Consider $X_1 \sim \mathrm{Bin}(n,p)$ and $X_2 \sim \mathrm{Bin}(m,p)$. Then $X_1+X_2 \sim \mathrm{Bin}(n+m,p)$.

**Proof.** We have $\mathbb{E}(e^{tX}) = \sum_{x=0}^n e^{tx}\binom{n}{x}p^x(1-p)^{n-x} = \sum_{x=0}^n \binom{n}{x}(e^tp)^x(1-p)^{n-x} = (e^tp+(1-p))^n$. So we have $\mathrm{MGF}_{X_1+X_2}(t) = (e^tp+(1-p))^n(e^tp+(1-p))^m = (e^tp+(1-p))^{n+m}$. This is the MGF of $\mathrm{Bin}(n+m,p)$. ■

# Lecture 4

8 Feb. 08:30

**Exercise.** Consider $X_1, X_n$ iid $\sim P(\lambda)$. Use MGF to show that $X_1+X_2 \sim P(2\lambda)$.

**Answer.** We have $M_X(t) = \mathbb{E}(e^{tX}) = \sum_{x=0}^{\inf} e^{tx}\frac{\lambda^x}{x!}e^{-\lambda} = e^{-\lambda}\sum_{x\geq 0}\frac{(\lambda e^t)^x}{x!} = e^{-\lambda}e^{\lambda e^t} = e^{\lambda(e^t-1)}$. Now we have $M_{X_1+X_2}(t) = m_{X_1}(t)M_{X_2}(t) = e^{2\lambda(e^t-1)}$. ⊛

**Exercise.** Find $M_X(t)$ where:

- $X \sim N(0,1)$
- $X \sim (\mu,\theta)$

**Exercise.** Use MGF to show that a linear combination of normals are normally distributed such that $\sum_i l_i X_i \sim N(\sum_i l_i \mu_i, \sum_i l_i^2 \sigma_i^2)$

## 2.4 Order Statistics

Consider $X_{(1)} \dots X_{(n)}$ where $X_{(1)}$ is the min and $X_{(n)}$ is the max and $X_{(1)} < X_{(2)} \dots X_{(n)}$. Here $X_i$ are continuous random variables. We know that the joint density function is given by $\prod_{i=1}^n f_{X_i}(x_i, \theta)$. So there are $n!$ possible sample vectors that will be mapped to a single order statistic realization. So $f_{X_{(1)} \dots X_{(n)}}(x_1, \dots x_n) = n! \prod_{i=1}^n f_{X_i}(x_i) = n! f_{(X_1 \dots X_n)}(x_1 \dots x_n)$ if $x_1 < x_2 < \dots x_n$, 0 otherwise.

**Lemma 2.4.1.** If $X_1 \dots X_n$ iid $\sim F_\theta$ then the joint density of the corresponding order statistic is given by $f_{X(1), \dots X(n)}(x_1 \dots x_n) = n! f_{X_1 \dots X_n}(x_1 \dots x_n)$ where $x_1 < x_2 \dots < x_n$ and 0 otherwise.

# Chapter 3

# Sufficiency

## 3.1 Sufficient Statistics

**Definition 3.1.1** (Sufficient Statistic)**.** A statistic $T(X_1 \dots X_n)$ is said to be sufficient if $T(X_1 \dots X_n)$ captures all the information in the sample about the unknown parameter $\theta$.

Consider some statistic $T_1$, and the conditional distribution of $T_1|T = t$. If this distribution depends on $\theta$ then studying $T_1$ may give additional information about $\theta$ which means that $T$ could not capture all the information about $\theta$ so it is not sufficient. So $T$ is sufficient for $\theta$ when the conditional distribution of $T_1|T = t$ is independent of $\theta$ for all estimators $T_1$ and all $t$.

**Example.** Consider $T(X_1 \dots X_n) = (X_1 \dots X_n)$. We can see this is sufficient since $\mathbb{P}(X_1 = y|T = (x_1 \dots x_n)) = 1$ if $y = x_1$ and 0 otherwise. Here $X_1$ is th new statistic. More generally we see that $\mathbb{P}(T_1 = y|T = (x_1 \dots x_n)) = 1$ if $y = T_1(x_1 \dots x_n)$ and 0 otherwise.

**Example.** Consider the order statistic $T(X_1 \dots X_n) = (X_{(1)} \dots X_{(n)})$. Similar to the previous example this is also sufficient.

**Example.** Consider $X_1, X_2, X_3 \sim \text{Ber}(p)$. Consider $T(X_1 \dots X_3) = \sum_i X_i$. Now consider $X_1|T = t$. For $t = 0$ we have $\mathbb{P}(X_1 = y|\sum_i X_i = 0) = 1$ if $y = 0$ and 0 otherwise. But for $t \geq 1$ we have $\mathbb{P}(X_1 = y|\sum_i X_i = t)$. This has 2 probabilities:

$$\mathbb{P}(X_1 = 0|T = t) = \frac{\mathbb{P}(\sum_{i=2}^3 X_i = t)\mathbb{P}(X_1 = 0)}{\mathbb{P}(T = t)} = \frac{\binom{2}{t}p^t(1-p)^{2-t}(1-p)}{\binom{3}{t}p^t(1-p)^{3-t}} = \frac{\binom{2}{t}}{\binom{3}{t}}$$

. Similarly we have

$$\mathbb{P}(X_1 = 1|T = t) = \frac{\mathbb{P}(\sum_{i=2}^3 X_i = t-1)\mathbb{P}(X_1 = 1)}{\mathbb{P}(T = t)} = \frac{\binom{2}{t-1}p^{t-1}(1-p)^{2-(t-1)}(p)}{\binom{3}{t}p^t(1-p)^{3-t}} = \frac{\binom{2}{t-1}}{\binom{3}{t}}$$

.

## Lecture 5

We now want a condition to check whether a statistic is sufficient or not. 13 Feb. 08:30

**Theorem 3.1.1** (Sufficient Statistics)**.** $T$ is sufficient for $\theta$ if and only if $(X_1, \dots X_n)|T = t$ is independent for $\theta$ for all $t$.

**Proof.** If $T$ is sufficient then we know $(X_1 \dots X_n)|T = t$ is independent from the definition of sufficiency. Now we want to show that reverse direction. Consider $T_1(X_1 \dots X_n)$ and $T_1|T = t$. We have $P(T_1 = t_1|T = t) = \sum_{x_1 \dots x_n \in \Omega, T_1 = t_1} \mathbb{P}(X_1 = x_1 \dots X_n = x_n|T = t)$. From the given condition the RHS is independent of $\theta$ and the set over which we are summing also does not depend on $\theta$. So the final expression will not depend on $\theta$ and so we have proved sufficiency of $T$. ■

**Example.** Consider $X_1 \dots X_4 \sim \text{Ber}(p)$. Consider $T(X_1 \dots X_4) = \sum_i X_i$. Now consider $\mathbb{P}(X_1 = x_1, \dots X_4 = x_4|T = t)$. It is 0 if $\sum_i X_i \neq t$. This probability is given by

$$\begin{aligned}
\frac{\mathbb{P}(X_1 = x_1, \dots X_4 = t - \sum_{i=1}^3 X_i)}{\mathbb{P}(T = t)} &= \frac{\mathbb{P}(X_1 = x_1)\mathbb{P}(X_2 = x_2)\mathbb{P}(X_3 = x_3)\mathbb{P}(X_4 = t - \sum_{i=1}^3 X_i)}{\mathbb{P}(T = t)} \\
&= \frac{p^{x_1}(1-p)^{1-x_1}p^{x_2}(1-p)^{1-x_2}p^{x_3}(1-p)^{1-x_3}p^{t-\sum_i x_i}(1-p)^{t-\sum_i x_i}}{\binom{4}{t}p^t(1-p)^{4-t}} \\
&= \frac{1}{\binom{4}{t}}
\end{aligned}$$

So this shows that $T$ is a sufficient statistic for $p$.

**Example.** Consider $X_1 \dots X_n \sim P(\lambda)$ where $T = \sum_i X_i$. We have

$$\begin{aligned}
\mathbb{P}(X_1 = x_1 \dots X_n = x_n|T = t) &= \frac{\mathbb{P}(X_1 = x_1)\mathbb{P}\mathbb{P}(X_n = t - \sum_{i=1}^{n-1})}{\mathbb{P}(T = t)} \\
&= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\frac{e^{-\lambda}\lambda^{x_2}}{x_2!}\cdots\frac{e^{-\lambda}\lambda^{t-\sum_{i=1}^{n-1} x_i}}{(t - \sum_{i=1}^{n-1} x_i)!}\frac{1}{\frac{e^{-n\lambda}(n\lambda)^t}{t!}} \\
&= \frac{t!}{n^t}
\end{aligned}$$

So $T$ is sufficient for $\lambda$.

**Exercise.** Suppose $T$ is sufficient for $\theta$. Let $T_1$ be any other estimator such that $T = g(T_1)$. Then $T_1$ is also sufficient for $\theta$.

**Theorem 3.1.2** (Factorization theorem). Consider $X_1 \dots X_n \sim F_\theta$ where $f_{X_1,\dots X_n,\theta}(x_1 \dots x_n)$ denotes the joint pmf/pdf of the observed random vector. If the following factorization holds for $f$: $f_{X_1 \dots X_n}(x_1, \dots, x_n, \theta) = g(T(x_1, \dots x_n), \theta)h(x_1, \dots, x_n)$ and $h(\widetilde{x} \geq 0)$ then $T$ is sufficient for $\theta$.

**Proof.** If $T$ is sufficient then the factorization holds. This is because if $T$ is sufficient then we can take $\mathbb{P}(X_1 = x_1 \dots X_n = x_n|T = t) = h(x_1 \dots x_n). = h(\widetilde{x})$. So the probability $\mathbb{P}_\theta(\widetilde{X} = \widetilde{x}) = \mathbb{P}_\theta(T = t)h(\widetilde{x})$ where the first term can be taken as $g(t, \theta)$. Now to prove the opposite direction. We have $\mathbb{P}_\theta(X = \widetilde{x}) = g(t, \theta)h(\widetilde{x})$. Since $h \geq 0$ we have $g \geq 0$. Furthermore the sum over the function must be 1 when considering the conditional probability distribution $\frac{\mathbb{P}_\theta(\widetilde{X} = \widetilde{x})}{\mathbb{P}_\theta(T = t(\widetilde{x}))}$ when summing over support set of $t(x)$. Consider $A_t = \{t(\widetilde{x}) = t : x(\Omega)\}$. So we can rewrite the conditional distribution as $\frac{\mathbb{P}_\theta(\widetilde{X} = \widetilde{x})}{\sum_{\widetilde{y} \in A_t} \mathbb{P}_\theta(\widetilde{X} = \widetilde{y})}$. Since the factorization principle holds we can write this as $\frac{g(t(\widetilde{x}), \theta)h(\widetilde{x})}{\sum_{\widetilde{y} \in A_t} g(t(\widetilde{y}), \theta)h(\widetilde{y})}$. But since $t(\widetilde{x}) = t(\widetilde{y})$ we finally have $\frac{h(\widetilde{x})}{\sum_{\widetilde{y} \in A_t} h(\widetilde{y})}$ which is independent of $\theta$ so $T$ is sufficient. ■

**Example.** Consider $X_1 \dots X_n \sim \text{Ber}(p)$. We have $f_{\widetilde{X}}(\widetilde{x}, p) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} = p^{\sum_i x_i}(1-p)^{n-\sum_i x_i} = g(t(\widetilde{x}), p)h(x)$ where $t(\widetilde{x}) = \sum_i x_i$ and $h(\widetilde{x} = 1)$. So $t$ is sufficient.

**Example.** Consider $X_1 \dots X_n \sim P(\lambda)$. So we have $f_{\widetilde{X}(\widetilde{x},\lambda)} = \prod_{i=1}^n \frac{e^{-\lambda}\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\lambda^{\sum_i x_i}}{x_i!}$. So we have again $t(\widetilde{x}) = \sum_i x_i$, $g(t(\widetilde{x}), \lambda) = e^{-\lambda}\lambda^{\sum_i x_i}$ and $h(\widetilde{x})$ as the denominator. So $t(\widetilde{x})$ is sufficient.

**Example.** Consider $X_1 \dots X_n \sim N(\mu, 1)$. We have $f_{\widetilde{X}}(\widetilde{x}, \mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2}} = \frac{1}{\sqrt{2\pi}^n} e^{-\frac{1}{2}\sum_i (x_i-\mu)^2}$. We can again choose $t = \sum_i x_i$ and factorize.

# Lecture 6

## 3.2 Minimal Sufficient Statistic

15 Feb. 08:30

**Definition 3.2.1** (Minimal Sufficient Statistic)**.** Sufficient statistic $T$ is said to be minimal sufficient if for any other sufficient statistic $T_1$, $T$ can be expressed as a function of $T_1$. So $T = g(T_1)$ for some $g$.

We want a condition to check minimal sufficiency.

**Theorem 3.2.1** (Minimal Sufficiency)**.** $T(\widetilde{X})$ is minimal sufficient if take any two sample points $\widetilde{x}, \widetilde{y} \in \Omega$. Consider the ratio of pmf at these two sample points: $\frac{p_{\widetilde{X}}(\widetilde{x},\theta)}{p_{\widetilde{X}}(\widetilde{y},\theta)}$. If ratio is independent of $\theta$ if and only if $T(\widetilde{x}) = T(\widetilde{y})$ then it implies $T$ is minimal sufficient.

**Proof.** First we will show that $T$ is sufficient. Then we will consider $T_1$ and show that $\exists g$ such that $T = g(T_1)$ completing the proof. Consider $\mathcal{T} = \{T(\widetilde{x}) = \widetilde{x} \in \Omega\}$ which is the support set of $T$. Consider $t \in \mathcal{T}$. We have $A_t = \{\widetilde{x} \in \Omega | T(\widetilde{x}) = t\}$. Now $\mathbb{P}(T = t) = \mathbb{P}(\widetilde{X} \in A_t)$. This probability is given by $\sum_{y \in A_t} p_X(\widetilde{y})$.

Now consider the joint pmf $p_{(\widetilde{X})}(\widetilde{x})$:

$$\begin{aligned} p_{(\widetilde{X})}(\widetilde{x}) &= \frac{p_{(\widetilde{X})}(\widetilde{x})}{\mathbb{P}(T = t)} \mathbb{P}(T = T(\widetilde{x})) \\ &= \frac{p_{(\widetilde{X})}(\widetilde{x})}{\sum_{\widetilde{y} \in A_{T(x)}} p_{\widetilde{X}}(\widetilde{y})} \mathbb{P}(T = T(\widetilde{x})) \end{aligned}$$

By assumption since $T(\widetilde{y}) = T(\widetilde{x})$ the ratio is independent of $\theta$. So the total summation is independent of $\theta$ and the ratio above is independent of $\theta$. Now, by the factorization theorem we have $g = \mathbb{P}(T = T(\widetilde{x}))$ and the ratio as $h$. So by factorization theorem we have $g$ as sufficient.

Take another sufficient $T_1$. The values taken by $T_1$ induces a partition on $\Omega$. For $T$ this is is given by $A_t | t \in \mathcal{T}$. We will show that the partition induced by $T_1$ is finer than the partition induced by $T$. We can write the ratio as

$$\frac{p_{\widetilde{X}}(\widetilde{x})}{p_{\widetilde{X}}(\widetilde{y})} = \frac{g_{T_1}(T_1(\widetilde{x}), \theta) h(\widetilde{x})}{g_{T_1}(T_1(\widetilde{y}), \theta) h(\widetilde{y})}$$

Consider $\widetilde{x}, \widetilde{y}$ such that $T_1(\widetilde{x}) = T_1(\widetilde{y})$. So the $g$ parts cancel and the ratio is independent of $\theta$. But since the ratio is independent then $T(\widetilde{x}) = T(\widetilde{y})$. ■

**Example.** Consider $X_1 \dots X_n \sim \text{Ber}(p)$. We have $\widetilde{x}, \widetilde{y} \in \Omega$ so we have $\frac{p_{\widetilde{X}}(\widetilde{x})}{p_{\widetilde{X}}(\widetilde{y})} = \frac{p^{\sum_i x_i}(1-p)^{n-\sum_i x_i}}{p^{\sum_i y_i}(1-p)^{n-\sum_i y_i}} = p^{(\sum_i x_i + \sum_i y_i)}(1-p)^{\sum_i y_i - \sum_i x_i}$. We can see that this would be independent of $p$ if and only if $\sum_i x_i = T(\widetilde{x}) = T(\widetilde{y}) = \sum_i y_i$. So this implies that $T$ is a minimal sufficient statistic.

**Exercise.** If $T$ is sufficient then for any bijective $g$ show that $g(T)$ is also sufficient.

**Exercise.** If $T$ is minimally sufficient then for any bijective $g$ show that $g(T)$ is minimal sufficient too.

**Example.** Consider $X_1 \dots X_n \sim P(\lambda)$. We have the ratio as $\frac{e^{n\lambda} \frac{\lambda^{\sum_i x_i}}{\prod_i x_i!}}{e^{n\lambda} \frac{\lambda^{\sum_i y_i}}{\prod_i y_i!}} = \lambda^{\sum_i x_i - \sum_i y_i} \prod_i (\frac{y_i!}{x_i!})$. So $T = \sum_i X_i$ is sufficient.

## Lecture 7

**Example.** Consider $X_1, \dots, X_n \sim U(0, \theta)$. We have $f_X(\tilde{x}) = \frac{1}{\theta^n} \mathbb{1}_{X_{(n)} \leq \theta} \mathbb{1}_{X_{(1)} > 0}$. This is automatically a factorization into $g$ and $h$, so the factorization theorem implies that $X_{(n)}$ is sufficient. 20 Feb. 08:30

**Exercise.** Given $X_1, \dots, X_n \sim U(0, \theta)$ check if $\overline{X}$ is sufficient for $\theta$.

**Example.** Consider $X_1, \dots X_n \sim U(\alpha, 1)$ where $\alpha < 1$. We have $f_X(x) = \frac{1}{(1-\theta)^n} \mathbb{1}_{X_{(1)} \geq \alpha} \mathbb{1}_{X_{(n)} \leq 1}$. So by factorization theorem we have $X_{(1)}$ as a sufficient statistic.

**Example.** Consider $X_1, \dots, X_n \sim N(\mu, \sigma^2)$ so both $\mu, \sigma$ are unknown. We want to obtain the minimal sufficient statistic. We have:

$$
\begin{aligned}
\frac{f_{\tilde{X}}(\tilde{x})}{f_{\tilde{X}}(\tilde{y})} &= \frac{\frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{1}{2} \sum_{i=1}^n (\frac{x_i - \mu}{\sigma})^2\right)}{\frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left(-\frac{1}{2} \sum_{i=1}^n (\frac{y_i - \mu}{\sigma})^2\right)} \\
&= \exp\left(-\frac{1}{2} \sum_{i=1}^n ((\frac{x_i - \mu}{\sigma})^2 - (\frac{y_i - \mu}{\sigma})^2)\right) \\
&= \exp\left(-\frac{\sum_i x_i^2 - \sum_i y_i^2 - 2\mu(\sum_i x_i - \sum_i y_i)}{2\sigma^2}\right)
\end{aligned}
$$

So the expression is independent of $\mu$ and $\sigma$ when $(\sum_i x_i, \sum_i x_i^2) = (\sum_i y_i, \sum_i y_i^2)$. So the minimal sufficient statistic is given by $(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i^2)$

**Example.** Consider the previous example of $U(\alpha, 1)$. We have the ratio:

$$
\frac{f_{\tilde{X}}(\tilde{x})}{f_{\tilde{X}}(\tilde{y})} = \frac{\frac{1}{(1-\alpha)^n} \mathbb{1}_{x_{(1)} \geq \alpha} \mathbb{1}_{x_{(n)} \leq 1}}{\frac{1}{(1-\alpha)^n} \mathbb{1}_{y_{(1)} \geq \alpha} \mathbb{1}_{y_{(n)} \leq 1}}
$$

So $X_{(1)}$ is a minimal sufficient statistic.

**Example.** Consider $X_1, \dots, X_n \sim U(\alpha, \beta)$ where $\alpha < \beta$. The minimal sufficient statistic is given by $(X_{(1)}, X_{(n)})$

# Chapter 4

# Minimum Variance Unbiased Estimator

## Lecture 8

22 Feb. 08:30

### 4.1 Fisher's Information Statistic

**Definition 4.1.1 (Fisher's Information).** Assumptions

1. $S_\theta = \{x | f(x,\theta) > 0\}$ where $S = \cup_{\theta \in \Theta} S_\theta$. Here $S_\theta$ does nor depend on $\theta$.
2. $\frac{\partial}{\partial x}(\int f(x,\theta)\mathrm{d}x) = \int \frac{\partial}{\partial \theta}(f(x,\theta))$.

Consider $\int_S f(x,\theta)\mathrm{d}x = 1$. This means $\int S(\frac{\frac{\partial}{\partial \theta} f(x,\theta)}{f(x,\theta)}) f(x,\theta) = \int_S \frac{\partial}{\partial \theta}(\log f(x,\theta)) f(x,\theta)$ $= \mathbb{E}\left[\frac{\partial}{\partial \theta}\log(f(x,\theta))\right] = 0$. This means that the variance $\mathrm{Var}\left[\frac{\partial}{\partial \theta}(\log(f(x,\theta)))\right] = \mathbb{E}\left[(\frac{\partial}{\partial \theta}\log(f(x,\theta)))^2\right]$. This value gives the information in the sample $X$ about the parameter $\theta$ denoted by $I_X(\theta)$.

**Lemma 4.1.1.** Consider $X_1, \ldots X_n$ iid. We have $\mathrm{Var}\left[\frac{\partial}{\partial \theta}\log\left(f(\widetilde{X},\theta)\right)\right] = nI_{X_1}(\theta)$.

**Proof.** Inside the log it becomes a sum and the independence means you can add the variances. ■

**Definition 4.1.2 (Information of a statistic).** Consider a statistic $T = T(X_1, \ldots X_n)$. The information of the statistic is given by $\mathrm{Var}\left[\frac{\partial}{\partial \theta}\log g(T,\theta)\right]$. In this case $g_T(T,\theta)$ is the marginal density/pmf of $\theta$. This is denoted by $I_T(\theta)$.

**Example.** Consider $X_1, \ldots, X_n \sim \mathrm{Ber}(p)$. We want $I_{\widetilde{X}}(p)$. The joint pmf $p_{\widetilde{X}}(\widetilde{x}) = p^{\sum_i x_i (1-p)^{n-\sum_i x_i}}$. So we have $\frac{\partial}{\partial p}(\log(p_{\widetilde{X}}(\widetilde{x}))) = \frac{\partial}{\partial p}(\sum_i x_i \log p + (n - \sum_i x_i)\log(1-p)) = \frac{\sum_i x_i}{p} - \frac{n - \sum_i x_i}{1-p} = \frac{\sum_i x_i - np}{p(1-p)}$. So now we know that $\mathbb{E}\left[\frac{\sum_i x_i - np}{p(1-p)}\right] = 0$. Now we want $\frac{1}{p^2(1-p)^2}\mathbb{E}\left[\sum_i (x_i - np)^2\right]$ $= \frac{1}{p^2(1-p)^2}\mathbb{E}\left[(\sum_i x_i - \mathbb{E}\left[\sum_{x_i}\right])^2\right] = \frac{1}{p^2(1-p)^2}\mathrm{Var}\left[\sum_i x_i\right] = \frac{np(1-p)}{p^2(1-p)^2} = \frac{n}{p(1-p)}$.

Now consider $T(X_1, \ldots X_n) = \sum_i x_i \sim \mathrm{Bin}(n,p)$. So we have $\log(p_T(t)) = \binom{n}{2} + t\log p + (n - t)\log(1-p)$. Now taking the derivative we get $\frac{t}{p} - \frac{n-t}{1-p} = \frac{t-np}{p(1-p)}$. Since $t = \sum_i x_i$. So this means $I_T(p) = I_{\widetilde{X}}(p)$.

**Exercise.** Consider $X_1, \ldots, X_n \sim F_\theta$ which satisfies the assumptions above with $T(X_1, \ldots X_n)$. $T$ is sufficient $\Leftrightarrow I_T(\theta) = I_{\widetilde{X}}(\theta)$. (Use factorization principle)

**Exercise.** Consider $X_1, \ldots, X_n \sim N(\mu, 1)$ and $T = \sum_i x_i$. Then show that $I_T(\mu) = I_{\widetilde{X}}(\mu)$. Similarly consider $X_1, \ldots, X_n \sim N(0, \sigma^2)$ and $T = \sum_i X_i^2$ then $I_T(\sigma) = I_{\widetilde{X}}(\sigma)$.

**Exercise.** Consider $X_1, \dots X_n \sim F_\theta$. Consider a bijective transformation $(X_1, \dots X_n) \to (Y_1, \dots, Y_n)$. Then show that $I_{\underset{\sim}{X}}(\theta) = I_{\underset{\sim}{Y}}(\theta)$. (Use Jacobian)

## 4.2 Minimum Variance Unbiased Variable

**Theorem 4.2.1** (Cramer Rao Lower Bound (CRLB)). Suppose we have $X_1, \dots, X_n \sim F_\theta$ such that it satisfies the two assumptions for Fisher's information. We want to estimate $g(\theta)$. We also assume that $\mathcal{U}_g \neq \varnothing$. Consider $T \in \mathcal{U}_g$ We then have $\mathrm{Var}_\theta[T] \geq \frac{|\frac{\mathrm{d}g}{\mathrm{d}\theta}|^2}{I_{\overline{X}}(\theta)}$.

**Proof.** Consider:

$$\begin{aligned}
\int T(\overline{x}) f_{\overline{X}}(x,\theta)\mathrm{d}x &= g(\theta) \\
\frac{\mathrm{d}}{\mathrm{d}\theta}(\int T(\overline{x}) f_{\overline{X}}(x,\theta)\mathrm{d}x) &= \frac{\mathrm{d}}{\mathrm{d}\theta} g(\theta) \\
\int \frac{\mathrm{d}}{\mathrm{d}\theta}(T(\overline{x}) f_{\overline{X}}(x,\theta))\mathrm{d}x &= g'(\theta) \\
\int T(\overline{X}) \frac{\mathrm{d}}{\mathrm{d}\theta} \log f_{\overline{X}}(\overline{x},\theta)\mathrm{d}x &= g'(\theta) \\
\mathbb{E}_\theta(T(\overline{X}) \frac{\mathrm{d}}{\mathrm{d}\theta} \log f(\overline{X}),\theta) &= g'(\theta) \\
\mathrm{Cov}_\theta\left[T(\overline{X}), \frac{\mathrm{d}}{\mathrm{d}\theta} \log f(\overline{X},\theta)\right] &= g'(\theta).
\end{aligned}$$

Consider all random variables defined on probability space $\mathcal{R} = (\Omega, \mathcal{F}.\mathbb{P})$ with finite second moment. Now consider two random variables $X, Y \in \mathcal{R}$. For any such $X, Y$ and $\alpha, \beta \in \mathbb{R}$ we can define $\alpha X + \beta Y \in \mathcal{R}$. We also have $(\alpha X + \beta Y)(\omega) = \alpha X(\omega) + \beta Y(\omega)$. We can define an inner product space $\langle X, Y\rangle = \mathbb{E}(XY)$. From the Cauchy-Schwarz inequality we have $(\mathbb{E}(XY))^2 \leq \mathbb{E}(X^2) + \mathbb{E}(Y^2)$. With a little more work we can show that for all $X, Y \in \mathcal{R}$ we have $(\mathrm{Cov}[X,Y])^2 \leq \mathrm{Var}[X]\,\mathrm{Var}[Y]$. So applying this to the above we get $\mathrm{Cov}^2_\theta\left[T(\overline{X}) FI_{\overline{X}}(\theta)\right] \leq \mathrm{Var}_\theta[T]\, I_{\overline{X}}(\theta)$. So finally we have $\frac{|g'(\theta)|^2}{I_{\overline{X}}(\theta)} \leq \mathrm{Var}_\theta[T]$. ■

**Example.** Consider $X_1, \dots X_n \sim \mathrm{Ber}(p)$. Consider unbiased estimators $T_1 = 2X_1 + 3$ and $t_2 = 2\overline{X} + 3$. We have $\mathrm{Var}_p[T_1] = 4p(1-p)$. And $\mathrm{Var}_p[T_2] = \frac{4}{n}p(1-p)$. We have the lower bound as $\frac{4p(1-p)}{n} = \mathrm{Var}_p[T_2]$.

**Remark.** Under a restricted setup the CRLB can be used to justify that a estimator is MVUE if there exists an MVUE that attains this lower bound.

## Lecture 9

27 Feb. 08:30

**Lemma 4.2.1.** $I_{\overline{X}}(\theta) = -\mathbb{E}(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log f_{\overline{X}}(\overline{x},\theta))$

**Proof.** Consider $X_1, \dots X_n \sim F_\theta$. We have $\mathbb{E}(FI_{\overline{X}}(\theta)) = 0$.
So we have $\int(\frac{\mathrm{d}}{\mathrm{d}\theta} \log f_{\overline{X}}(\overline{x},\theta)) f_{\overline{X}}(\overline{x},\theta)\mathrm{d}x = 0$. We can take the derivative of this expression, and then swap it with the integral again to get $\int \frac{\mathrm{d}}{\mathrm{d}\theta}(\frac{\mathrm{d}}{\mathrm{d}\theta} \log f_{\overline{X}}(\overline{x},\theta)) f_{\overline{X}}(\overline{x},\theta)\mathrm{d}x = 0$. So now we have $\int \frac{\mathrm{d}^2}{\mathrm{d}\theta^2}(\log f_{\overline{X}}(\overline{x},\theta)) f_{\overline{X}}(x,\theta)\mathrm{d}x + \int \frac{\mathrm{d}}{\mathrm{d}\theta} \log f_{\overline{X}}(x,\theta) \frac{\mathrm{d}}{\mathrm{d}\theta} f_{\overline{X}}(\overline{x},\theta)\mathrm{d}x = 0$. We then see that we can write the first term as $\mathbb{E}(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}(\log f(\overline{X},\theta)))$. For the second term we have $\int(\frac{\mathrm{d}}{\mathrm{d}\theta} \log f_{\overline{X}}(\overline{x},\theta))^2 f_{\overline{X}}(\overline{x},\theta)\mathrm{d}x$ by multiplying and dividing the second term. This is the same as $I_{\overline{X}}(\theta)$. This tells us that $I_{\overline{X}}(\theta) = -\mathbb{E}(\frac{\mathrm{d}^2}{\mathrm{d}\theta^2} \log f_{\overline{X}}(\overline{x},\theta))$. This is a useful expression to use sometimes. ■

**Corollary 4.2.1.** If $T \in \mathcal{U}_\pi$ such that $\text{Var}_\theta[T] = \frac{\phi'(\theta)^2}{I_{\overline{X}}(\theta)} \forall \theta \in \Theta$ then $T$ is MVUE.

**Remark.** RCLB gives a valid lower bound only if the required assumptions are satisfied. If $\mathcal{U}_\phi$ is empty then this lower bound is not helpful.

**Remark.** In order to attain the equality for the RCLB we need equality in the Caucchy-Schwarz inequality. This happens when $T$ and $FI_{\overline{X}}(\theta)$ are perfectly linearly related.

**Remark.** If $X, Y$ are perfectly linearly related then their standardized versions are also perfectly linearly related. So we have $\frac{T-\phi(\theta)}{\pm\frac{|\phi'(t)|}{\sqrt{I_{\overline{X}}(\theta)}}}$ is perfectly linearly related with $\frac{FI_{\overline{X}}(\theta)}{\sqrt{I_{\overline{X}}(\theta)}}$. So $T(\overline{X})$ is perfectly linearly related with $\frac{FI_{\overline{X}}(\theta)}{I_{\overline{X}}(\theta)} \times (\pm|\psi'(\theta)|) + \psi(\theta)$

## Lecture 10

29 Feb. 08:30

**Definition 4.2.1** (Perfectly Linearly Related)**.** $X$ and $Y$ are perfectly linear related if there exists $a, b \in \mathbb{R}$ such that $\mathbb{P}(Y = aX + b) = 1$.

**Example.** Consider $X_1, \dots, X_n \sim \text{Ber}(p)$ with $\psi(p) = p$. So we have $\frac{FI_{\overline{X}}(\theta)}{I_{\overline{X}(\theta)}} = \frac{\frac{\sum_i X_i}{p} \frac{n-\sum_i X_i}{1-p}}{\frac{n}{p(1-p)}} = \frac{\sum_i X_i - np}{n}$. So we get that $T(X)$ is perfectly linearly related to $\pm\frac{\sum_i X_i - np}{n} + p = \overline{X}$. So $\overline{X}$ is an MVUE.

**Exercise.** Show that neither positive or negative square root of the final expression is never free of $p$.

**Exercise.** Consider $X_1, \dots, X_n \sim P(\lambda)$. Show that there exists no unbiased estimator that attains RCLB for $\psi(\lambda) = e^{-\lambda}$.

**Remark.** Consider $X, Y$ on $(\Omega, \mathcal{F}, \mathbb{P})$ with $\mathbb{E}(|X|), \mathbb{E}(|Y|) < \infty$. We have $\mathbb{E}(X|Y = y) = \sum_x x\mathbb{P}(X = x|Y = y) = \phi(y)$. Now consider $\mathbb{E}(X|Y)$ which is no longer a number but itself a random variable. It takes the value $\phi(y)$ with probability $\mathbb{P}(Y = y)$. So we have $\mathbb{E}(\mathbb{E}(X|Y)) = \sum_y \phi(y)\mathbb{P}(Y = y) = \sum_y(\mathbb{E}(X|Y = y))\mathbb{P}(Y = y) = \sum_y(\sum_x x\frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)})\mathbb{P}(Y = y) = \sum_x x(\sum_y \mathbb{P}(X = x, Y = y))$. Since we are summing the joint distribution over all values of $Y$ this is the same as $\sum_x x\mathbb{P}(X = x) = \mathbb{E}(X)$. So we have $\mathbb{E}(\mathbb{E}(X|Y)) = \mathbb{E}(X)$

**Proposition 4.2.1.** $\text{Var}(\mathbb{E}(X|Y)) \leq \text{Var}(X)$ for $X, Y$ defined on same probability space with finite second moments.

**Proof.** We have

$$\begin{aligned}
\mathrm{Var}\,[X] &= \mathbb{E}(X-\mathbb{E}(X))^2 \\
&= \sum_x (x-\mu)^2 \mathbb{P}(X=x) \\
&= \sum_x \sum_y (x-\mu)^2 \mathbb{P}(X=x, Y=y) \\
&= \sum_x \sum_y (x-\mathbb{E}(X|Y=y)+\mathbb{E}(X|Y=y)-\mu)^2 \mathbb{P}(X=x, Y=y)
\end{aligned}$$

We now have 3 terms as follows:

$$\begin{aligned}
T_1 &= \sum_x \sum_y (x-\mathbb{E}(X|Y=y))^2 \mathbb{P}(X=x, Y=y) \\
&= \sum_y \sum_x (x-\mathbb{E}(X|Y=y))^2 \mathbb{P}(X=x|Y=y)\mathbb{P}(Y=y) \\
&= \sum_y \mathrm{Var}\,[X|Y=y]\,\mathbb{P}(Y=y) \\
&= \mathbb{E}(\mathrm{Var}\,[X|Y]) \geq 0 \\
T_1 &= \sum_x \sum_y (\phi(y)-\mu)^2 \mathbb{P}(X=x, Y=y) \\
&= \sum_x \sum_y (\phi(y)-\mathbb{E}(\phi(y)))^2 \mathbb{P}(X=x, Y=y) \\
&= \sum_y (\phi(y)-\mathbb{E}(\phi(y)))^2 \mathbb{P}(Y=y) \\
&= \mathrm{Var}\,[\phi(y)] \\
&= \mathrm{Var}\,[\mathbb{E}(X|Y)] \\
T_3 &= 2\sum_x \sum_y (x-\phi(y))(\phi(y)-\mu)\mathbb{P}(X=x|Y=y)\mathbb{P}(Y=y) \\
&= 2\sum_y (\phi(y)-m)(\sum_x (x-\phi(y))\mathbb{P}(X|Y=y))\mathbb{P}(Y=y) \\
&= 0
\end{aligned}$$

So ultimately we have

$$\begin{aligned}
\mathrm{Var}\,[X] &= \mathbb{E}(\mathrm{Var}\,[X|Y]) + \mathrm{Var}\,[\mathbb{E}(X|Y)] \\
\mathrm{Var}\,[X] &\geq \mathrm{Var}\,[\mathbb{E}(X)]
\end{aligned}$$

■

**Theorem 4.2.2** (RBL Theorem). Consider $X_1, \ldots, X_n \sim F_\theta$ where we want to estimate $\psi(\theta)$. We have $T_1 \in \mathcal{U}_\psi$ and $T$ is sufficient. Then $\mathbb{E}(T_1|T)\ in \mathcal{U}_\psi$ and $\mathrm{Var}_\theta\,[\mathbb{E}(T_1|T)] \leq \mathrm{Var}_\theta\,[T_1]\,, \forall\theta \in \Theta$.

**Proof.** We have $\mathbb{E}(T_1|T)$ is a valid estimator as it does not depend on $\theta$ since $T$ is sufficient and it is also unbiased. We also have $\mathrm{Var}_\theta\,[\mathbb{E}(T_1|T)] \leq \mathrm{Var}_\theta\,[T_1]$. So we have $\mathbb{E}(T_1|T)$ is an unbiased estimator of $\psi(\theta)$ with a lower variance than $T_1$. ■

# Lecture 11

## 4.3 Complete Sufficient Statistic

5 Mar. 08:30

**Definition 4.3.1** (Complete Sufficient Statistic)**.** Consider a family $F_\theta$ where $T$ is a sufficient statistic. For some $g$ if we have $\mathbb{E}_\theta\left[g(T)\right] = 0 \forall\theta \in \Theta \Rightarrow \mathbb{P}_\theta(g(t) = 0) = 1 \forall\theta \in \Theta$ then $T$ is called complete sufficient statistic.

**Example.** Consider $X_1, \ldots, X_n \sim \text{Ber}(p)$. Consider $T = \sum_i X_i$ and $g(T) = T$. So $\mathbb{E}_p(T) = \sum_{t=0}^n tp^t(1-p)^{n-t}\binom{n}{t} = 0, \forall p \in [0, 1]$. Since this is a polynomial we can see that all its terms must have 0 as a coefficient for it to equal so, so we have that $\mathbb{P}_p(g(T) = 0) = 1$.

**Example.** We have $\sum_i X_i$ is complete for:

- $X_1, \ldots, X_n \sim \text{Bin}(10, p)$
- $X_1, \ldots, X_n \sim P(\lambda)$
- $X_1, \ldots, X_n \sim N(\mu, 9)$

For $X_1, \ldots X_n \sim N(0, \sigma^2)$ we have $\sum_i X_i^2$ as complete. For $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$ we have $(\sum_i X, \sum_i X_i^2)$ as complete. For $X_1, \ldots, X_n \sim U(0, \theta)$ we have $X_{(n)}$ as complete. For $X_1, \ldots, X_n \sim U(a, b)$ we have $(X_{(1)}, X_{(n)})$ as complete.

**Theorem 4.3.1** (RBLS Theorem)**.** Consider $X_1, \ldots, X_n \sim F_\theta$ where $F_\theta$ is complete (meaning the minimal sufficient family is complete) with $T_1 \in \mathcal{U}_\psi$ and $T'$ is complete minimal sufficient. Then $\mathbb{E}(T_1|T')$ is the MVUE.

*Proof.* Consider $\mathbb{E}_\theta\left[\phi_1(t')\right] = \mathbb{E}_\theta\left[\phi_2(t')\right] = \psi(\theta)\forall\theta \in \Theta$. So we have $\mathbb{E}_\theta\left[\phi_1(t') - \phi_2(t')\right] = 0, \forall\theta \in \Theta \Rightarrow \mathbb{P}_\theta(\phi_1(t') = \phi_2(t')) = 1, \forall\theta \in \Theta$ ∎

**Theorem 4.3.2** (RBLS v2)**.** Consider $X_1, \ldots, X_n \sim F_\theta$ where $T'$ is a complete minimal sufficient statistic. Then $g(t)$ is a (the) MVUE of $\mathbb{E}_\theta\left[g(t)\right]$

**Example.** Consider $X_1, \ldots, X_n \text{Ber}(p)$ where $\psi(p) = p$ where we have $T_1 = X_1 \in \mathcal{U}_\psi$. We also have $T = \sum_i X_i$ as a complete minimal sufficient statistic. So we have $\mathbb{E}(X_1|T)$ as an MVUE for $p$. This takes the value $\mathbb{E}(X_1|T = t)$ with probability $\mathbb{P}(T = t) = \binom{n}{t}p^t(1-p)^{n-t}$. We have $\mathbb{E}(X_1|T = t) = \mathbb{P}(X_1 = 1|T = t) = \frac{\mathbb{P}(X_1=1|\sum_i X_i=t)}{\mathbb{P}(\sum_i X_i=t)} = \frac{\mathbb{P}(X_1=1)\binom{n-1}{t-1}p^{t-1}(1-p)^{n-t}}{\binom{n}{t}p^t(1-p)^{n-t}} = \frac{t}{n} = \overline{X}$

**Example.** Consider $X_1, \ldots, X_n \sim U(0, \theta)$. Construct an MVUE for $\psi(\theta) = \theta$. We have $X_{(n)}$ as a complete minimal sufficient statistic. We know that $\mathbb{E}(X_{(n)}) = \frac{n\theta}{n+1}$. So we have $\mathbb{E}(\frac{n+1}{n}X_{(n)}) = \theta$ so $\frac{n+1}{n}X_{(n)}$ is MVUE for $\theta$.

# Lecture 12

## 4.4 Properties of MVUE

7 Mar. 08:30

**Proposition 4.4.1.** Let $\mathcal{U}_0$ be the class of unbiased estimators of 0. So if $T_! \in \mathcal{U}_0$ then $\mathbb{E}_\theta\left[T_1\right] = 0, \forall\theta \in \Theta$. $T$ is MVUE of $g(\theta)$ if and only if $\text{Cov}_\theta\left[T, T_1\right] = 0, \forall\theta \in \Theta, \forall T_1 \in \mathcal{U}_0$.

**Proof.** Assume $T$ is an unbiased estimator of $g(\theta)$ such that $\operatorname{Cov}_\theta[T, T_1] = 0, \forall T_1 \in \mathcal{U}_0$. Let $T^\star$ some unbiased estimator of $g(\theta)$. We have $T - T^\star \in \mathcal{U}_0$. So we know that $\operatorname{Cov}_\theta[T, T - T^\star] = 0$. This means that $\operatorname{Var}_\theta[T] = \operatorname{Cov}_\theta[T, T^\star]$. We know that $\operatorname{Var}_\theta[T - T^\star] = \operatorname{Var}_\theta[T] + \operatorname{Var}_\theta[T^\star] - 2\operatorname{Cov}_\theta[T, T^\star] = \operatorname{Var}_\theta[T^\star] - \operatorname{Var}_\theta[T] \geq 0$. So the variance of $T$ is minimum.

Now consider $T$ is MVUE of $g(\theta)$ and $T_0 \in \mathcal{U}_0$. This means that $T + \lambda T_0$ is also an unbiased an estimator. So we have $\operatorname{Var}_\theta[T + \lambda T_0] = \operatorname{Var}_\theta[T] + \lambda^2 \operatorname{Var}_\theta[T_0] + 2\lambda \operatorname{Cov}_\theta[T, T_0]$. Let $\lambda = -\frac{\operatorname{Cov}_\theta[T_1,T_0]}{\operatorname{Var}_\theta[T_0]}$. So we get $\operatorname{Var}_\theta[T] + \frac{\operatorname{Cov}^2_\theta[T_1,T_0]}{\operatorname{Var}_\theta[T_0]} - \frac{2\operatorname{Cov}^2_\theta[T,T_0]}{\operatorname{Var}_\theta[T_0]} = \operatorname{Var}_\theta[T] - \frac{\operatorname{Cov}^2_\theta[T,T_0]}{\operatorname{Var}_\theta[T_0]}$. This is lesser than the variance of $T$ if the covariance is greater than 0. ■

**Remark.** Consider $T \in \mathcal{U}_g$ then $T + \mathcal{U}_0 = \{T + T_0 | T_0 \in \mathcal{U}_0\} = \mathcal{U}_g$.

**Proposition 4.4.2.** MVUE (if exist) is unique.

**Proof.** Consider $T_1, T_2$ are MVUE. This means $T_1 - T_2 \in \mathcal{U}_0$. So $\operatorname{Cov}_\theta[T_1, T_1 - T_2] = 0$. So we have $\operatorname{Var}_\theta[T_1] = \operatorname{Cov}_\theta[T_1, T_2] = \operatorname{Var}_\theta[T_2]$. So we have $\operatorname{Var}_\theta[T_1 - T_2] = \operatorname{Var}_\theta[T_1] + \operatorname{Var}_\theta[T_2] - 2\operatorname{Cov}_\theta[T_1, T_2] = 0$. This means that $T_1 - T_2 = 0$ so $T_1 = T_2$. ■

**Proposition 4.4.3.** Consider $X_1, X_2, \ldots, X_n \sim F_\theta$. We may be interested in estimating several functions of $\theta$, $g_1(\theta), \ldots, g_k(\theta)$. Consider $T_i$ as the MVUE of $g_i(\theta)$. Then $\sum_{i=1}^k a_i T_i$ is the MVUE for $\sum_{i=1}^k a_i g_i(\theta)$.

**Proof.** Let $T_0 \in \mathcal{U}_0$. Now consider $\operatorname{Cov}_\theta[\sum_i a_i T_i, T_0] = \sum_i a_i \operatorname{Cov}_\theta[T_i, T_0] = 0$. ■

**Exercise.** Consider $X_1, \ldots, X_n \sim P(\lambda)$. Construct the MVUE of $e^{-\lambda}$ and compare variance with CRLB.

**Answer.** Let $Y = 1$ if $X_1 = 0$ and 0 otherwise. So $\mathbb{E}(Y) = \mathbb{P}(X_1 = 0) = e^\lambda$. Now consider $\mathbb{E}(Y | \sum_i X_i)$. We see that $\mathbb{E}(Y|T = t) = \mathbb{P}(Y = 1|T = t) = \frac{\mathbb{P}(Y=1,T=t)}{\mathbb{P}(T=t)} = \frac{P(X_1=0,\sum_{i=2} X_i=t)}{\mathbb{P}(\sum_i X_1=t)} = \frac{e^{-\lambda}e^{-(n-1)\lambda}\frac{(n-1)\lambda^t}{t!}}{\frac{e^{-n\lambda}(n\lambda)^t}{t!}} = (\frac{n-1}{n})^t = (\frac{n-1}{n})^{\sum_i X_i}$. ⊛

**Exercise.** Let $X_1, \ldots X_n \sim N(\mu, 1)$. Start with $T_1 = \frac{X_1+X_2}{2}$ and construct $\mathbb{E}(T_1|T = \overline{X})$. Justify that this is the MVUE by comparing its variance with CRLB.

**Answer.** $(T_1, T) = (\frac{X_1+X_2}{2}, \overline{X})$ follows a bivariate normal distribution with mean vector $(\mu, \mu)$ and covariance matrix $\begin{pmatrix} \frac{1}{2} & \frac{1}{n} \\ \frac{1}{n} & \frac{1}{n} \end{pmatrix}$. We want $\mathbb{E}(\frac{X_1+X_2}{2}|\overline{X} = x)$. ⊛

**Exercise.** Consider $X_1, X_2, X_3 \sim N(\mu, 1)$. Construct MVUE of $\mu^2$ and compare its variance with CRLB.

# Chapter 5

# Hypothesis Testing

## Lecture 13

## 5.1 Motivation

21 Mar. 08:30

Consider $X_1, \ldots, X_n \sim F_\theta$ where $\theta \in \Theta$. We have $\Theta_0, \Theta_1 \subseteq \Theta$ such that $\Theta_0 \cap \Theta_1 = \varnothing$. We have the belief $H_0 : \theta \in \Theta_0$ (null hypothesis) against $H_1 : \theta \in \Theta_1$ (alternate hypothesis). Consider $\Omega$ that is the sample space (all possible values of $(X_1, \ldots, X_n)$). Given $\widetilde{x}$ (sample realisation) one is to decide whether to accept or reject $H_0$. So we want to partition $\Omega$ into an acceptance region and a rejection region.

**Definition 5.1.1** (Rejection/Critical Region)**.** $\mathcal{C}$ such that if $\widetilde{x} \in \mathcal{C}$ then reject $H_0$.

**Definition 5.1.2** (Acceptance region)**.** $\mathcal{C}^c$ such that if $\widetilde{x} \in \mathcal{C}^c$ then accept $H_0$.

After our sample vector is realised the decision is non-random, therefore this test is called a non-random test.

**Definition 5.1.3** (Test Function)**.** The test function for non-randomised tests is given by $\phi(\widetilde{x}) = \begin{cases} 1, & \text{if } \widetilde{x} \in \mathcal{C} \\ 0, & \text{if } \widetilde{x} \in \mathcal{C}^c \end{cases}$. For randomised tests this is given by $\phi(\widetilde{x}) = \begin{cases} 1, & \text{if } \widetilde{x} \in \mathcal{C}_0; \\ \gamma, & \text{if } \widetilde{x} \in \mathcal{C}_1; . \\ 0 & \text{if} \widetilde{x} \in \mathcal{C}_2. \end{cases}$

## 5.2 Error

**Definition 5.2.1** (Simple Null/Simple Alternative)**.** Testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$.

**Definition 5.2.2** (Errors)**.** We have two types of errors:

1. Type I error: Rejecting $H_0$ when $H_0$ is true. The probability is given by $\mathbb{P}_\theta(\widetilde{X} \in \mathcal{C} | \theta = \theta_0) = \mathbb{P}_{\theta_0}(\widetilde{x} \in \mathcal{C})$.

2. Type II error: Accepting $H_0$ when $H_1$ is true. The probability is given by $\mathbb{P}(\widetilde{x} \in \mathcal{C}^c | \theta = \theta_1) = 1 - \mathbb{P}_{\theta_1}(\widetilde{x} \in \mathcal{C})$.

**Example.** Consider $X_1, \ldots, X_{20} \sim \text{Ber}(p)$ with test $\mathcal{C} = \{(x_1, \ldots, x_{20}) : \sum_i x_i \geq 18\}$. So probability of type I is given by $\mathbb{P}(\text{Bin}(20, \frac{1}{2}) \geq 18)$ and probability of type II is given by $1 - \mathbb{P}(\text{Bin}(20, \frac{3}{4}) \geq 18)$.

## 5.3 Best Test

Ideally we want to minimise the probability of both type I and type II errors simultaneously. But this is not possible for a fixed sample size testing problem since to reduce type I error we want to reduce the size of $\mathcal{C}$ but this automatically increases the size of $\mathcal{C}^c$ which in turn increases type II error and vice-versa. We usually try to reduce type I error more than type II error for philosophical reasons.

**Definition 5.3.1** (Level-$\alpha$ test)**.** A test of level-$\alpha$ is a test where $\mathbb{P}(\text{Type I}) \leq \alpha$. In the simple null situation this is given by $\mathbb{P}_{\theta_0}(\widetilde{x} \in \mathcal{C}) \leq \alpha$. In other situations this is given by $\mathbb{P}_\theta(\widetilde{x} \in \mathcal{C}) \leq \alpha, \forall \theta \in \Theta_0$.

Among level-$\alpha$ tests we choose the test with minimum type II error probability. Suppose $\phi$ is the test function corresponding to the partition $\mathcal{C}, \mathcal{C}^c$. Then probability of type I error is $\mathbb{P}_{\theta_0}(\widetilde{X} \in \mathcal{C}) = \mathbb{E}_{\theta_0}(\phi)$. Similarly we have probability of type II error as $1 - \mathbb{E}_{\theta_1}(\phi)$.

**Definition 5.3.2** (Power)**.** The power of the test is given by $\mathbb{P}_{\theta_1}(\widetilde{X} \in \mathcal{C}) = \mathbb{E}_{\theta_1}(\phi)$.

**Definition 5.3.3** (MP-$\alpha$ )**.** For a simple null test, we say $\phi$ is MP level-$\alpha$ test if:

1. $\phi$ is level-$\alpha$ so $\mathbb{E}_{\theta_0}(\phi) \leq \alpha$
2. $\mathbb{E}_{\theta_1}(\phi) \geq \mathbb{E}_{\theta_1}(\phi')$ where $\phi'$ is level-$\alpha$.

**Definition 5.3.4** (UMP-$\alpha$ )**.** We say $\phi$ is UMP level-$\alpha$ test if:

1. $\phi$ is level-$\alpha$ so $\mathbb{E}_\theta(\phi) \leq \alpha, \forall \theta \in \Theta_0$
2. $\mathbb{E}_\theta(\phi) \geq \mathbb{E}_\theta(\phi'), \forall \theta \in \Theta_1$ where $\phi'$ is level-$\alpha$.

## Lecture 14

26 Mar. 08:30

**Lemma 5.3.1** (Neyman-Pearson Lemma)**.** Consider $X_1, \ldots X_n = \widetilde{X}$. We have $f_{\widetilde{X}}(\widetilde{x}|\theta_1) = f_1(\widetilde{x})$ and $f_{\widetilde{X}}(\widetilde{x}|\theta_0) = f_0(\widetilde{x})$. Fix any $\alpha \in (0,1)$. There exists most powerful level-$\alpha$ test $\phi_K$ of the form $\phi_K(\widetilde{x}) = \begin{cases} 1, & \text{if } f_1(\widetilde{x}) > K f_0(\widetilde{x}); \\ \gamma, & \text{if } f_1(\widetilde{x}) = K f_0(\widetilde{x}); \\ 0, & \text{if } f_1(\widetilde{x}) < K f_0(\widetilde{x}); \end{cases}$ for some $K \in (0,\infty)$ and $\gamma \in [0,1]$ such that $\mathbb{E}_{\theta_0}[\phi_K] = \alpha$.

**Proof.** Say $\phi'$ is any other test of level-$\alpha$. So $\mathbb{E}_{\theta_0}[\phi'] \leq \alpha$. Now to show $\mathbb{E}_{\theta_1}[\phi_K] \geq \mathbb{E}_{\theta_1}[\phi']$. Consider $Q(\widetilde{x}) = [\phi_K(\widetilde{x}) - \phi'(\widetilde{x})][f_1(\widetilde{x}) - K f_0(\widetilde{x})]$. We have three regions when the ratio is greater, lesser and equal to $K$ given by $E_1, E_2, E_3$ respectively. We have $\int_\Omega Q(\widetilde{x}) dx = \int_{E_1} + \int_{E_2} + \int_{E_3} \geq 0$ since over each reason the terms are positive, positive and 0 respectively. So we get $\int_\Omega \phi_K(\widetilde{x}) f_(\widetilde{x}) d\widetilde{x} - \int_\Omega \phi'(\widetilde{x}) f_1(\widetilde{x}) d\widetilde{x} \geq k(\int_\Omega \phi_K(\widetilde{x}) f_0(\widetilde{x}) d\widetilde{x} - \int_\Omega \phi'(\widetilde{x}) f_0(\widetilde{c}) d\widetilde{x})$. Since the RHS is $\geq 0$ we have $\mathbb{E}_{\theta_1}[\phi_K] \geq \mathbb{E}_{\theta_1}[\phi']$.

Now to prove existence. ■

**Example.** Consider $X_1, \ldots, X_n \sim \text{Ber}(p)$. We have $\lambda((\widetilde{x})) \frac{f_1(\widetilde{x})}{f_0(\widetilde{x})} = \frac{(\frac{3}{4})^{\sum_i x_i} (\frac{1}{4}^{10 - \sum_i x_i})}{(\frac{1}{4})^{\sum_i x_i} (\frac{3}{4}^{10 - \sum_i x_i})} = 3^{2 \sum_i x_i - 10}$ which is called the ratio statistic. This is an increasing function of $\sum_i x_i$ Now consider $\alpha = 0.1$. We have $\lambda(\widetilde{x} > K) \Leftrightarrow \sum_i x_i > k_0$ and so on. So we have $\phi_K = \begin{cases} 1, & \text{if } \displaystyle\sum_i x_i > k_0; \\ \gamma, & \text{if } \displaystyle\sum_i x_i = k_0; \\ 0, & \text{otherwise.} \end{cases}$ where $k_0 \in (0,\infty)$ and $\gamma \in [0,1]$ such that $\mathbb{E}_{\frac{1}{4}}[\phi_K] = 0.1$.

**Example.** Consider $X_1, \dots X_n \sim P(\lambda)$. Let $H_0 : \lambda = \lambda_0$ and $H_1 : \lambda = \lambda_1$ with $\lambda_1 < \lambda_0$. We have $\lambda(\tilde{x}) = e^{(\lambda_0 - \lambda_1)n} (\frac{\lambda_1}{\lambda_0})^{\sum_i x_i}$. So $\lambda(\tilde{x})$ is monotonically decreasingly related to $\sum_i x_i$. So the MP level-$\alpha$ is given by $\phi_K = \begin{cases} 1, & \text{if } \sum\limits_i x_i < k_0; \\ \gamma, & \text{if } \sum\limits_i x_i = k_0; \\ 0, & \text{otherwise.} \end{cases}$. We then choose $k_0$ and $\gamma$. So we have $k_0 = \min\{m \geq 0 : \mathbb{P}(\sum_i x_i \leq m) > \alpha\}$ and $\gamma = \frac{\alpha - \mathbb{P}(\sum_i X_i < k_0)}{\mathbb{P}(\sum_i X_i = k_0)}$.

## Lecture 15

2 Apr. 08:30

**Example.** Consider $X_1, \dots, X_n \sim N(\mu, 9)$. We are testing $\mu = \mu_0$ against $\mu = \mu_1$ with $\mu_1 > \mu_0$ . We want to attain MP-($\alpha$) test. We have $\lambda(\tilde{x}) = \frac{(\frac{1}{\sqrt{2\pi 9}})^n - \exp(\frac{1}{2}\sum_i(\frac{x_i - \mu_1}{3})^2)}{(\frac{1}{\sqrt{2\pi 9}})^n - \exp(\frac{1}{2}\sum_i(\frac{x_i - \mu_0}{3})^2)}$ $= \exp\left(-\frac{1}{18}(-\sum_i 2x_i(\mu_1 - \mu_0)) + n(\mu_1^2 - \mu_0^2)\right) = \exp\left(\frac{n(\mu_0^2 - \mu_1^2)}{18}\right) \exp\left(\frac{n\overline{x}(\mu_1 - \mu_0)}{9}\right)$. So it is monotonically increasing with $\overline{x}$. So we have $\phi(\tilde{x}) = \begin{cases} 1, & \text{if } \lambda(\tilde{x}) > k; \\ 0, & \text{if } \lambda(\tilde{x}) < k. \end{cases}$. This is equivalent to saying $\phi(\tilde{x}) = \begin{cases} 1, & \text{if } \overline{x} > k'; \\ 0, & \text{if } \overline{x} < k'. \end{cases}$ where $k'$ is such that the test of of size $\alpha$ meaning $\mathbb{E}_{\mu_0}(\phi) = \alpha$. So this means that $\mathbb{P}(\overline{X} > k' | \overline{X} \sim N(\mu_0, \frac{9}{n})) = \alpha$. Now consider $Z = \frac{\overline{X} - \mu_0}{\sqrt{\frac{9}{n}}}$. So saying $\overline{X} > k' \Leftrightarrow Z > \frac{k' - \mu_0}{\sqrt{\frac{9}{n}}}$. So we have MP-$\alpha$ test as reject $H_0$ if $\overline{x} > \mu_0 + \Phi_\alpha \sqrt{\frac{9}{n}}$. If we say $\mu_1 < \mu_0$ then we see that $;(\tilde{x})$ is decreasing with respect to $\overline{x}$. So we will reject $H_0$ if $\overline{x} < k'$. So we want $Z < \frac{k' - \mu_0}{\sqrt{\frac{9}{n}}}$ so we reject $H_0$ if $\overline{x} < \mu_0 - \Phi_\alpha \sqrt{\frac{9}{n}}$.

**Remark.** Notice that for all $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ the critical region of the MP$-\alpha$ test is always given by reject $H_0$ if $\overline{x} > \mu_0 + \Phi_\alpha \sqrt{\frac{\sigma^2}{9}}$ for $N \sim (\mu, \sigma_0^2)$ where $\sigma_0$ is known. So this is a UMP-$\alpha$ test.

**Corollary 5.3.1.** The ratio statistic must be a function of the minimal sufficient statistic.

**Proof.** We have $\lambda(\tilde{x}) = \frac{f(\tilde{x}, \theta_1)}{f(\tilde{x}, \theta_0)} = \frac{g(t, \theta_1) h(\tilde{x})}{g(t, \theta_0), h(\tilde{x})}$. So $T$ can be a minimal sufficient statistic and any test can be expressed in terms of it. ■

**Corollary 5.3.2.** Any test can be expressed in terms of a sufficient statistic $T$.

**Proof.** Consider $\mathbb{E}(\phi | T)$ is a valid estimator or test function, a random variable and ust be a function of $T$. We also have $\mathbb{E}_\theta(\mathbb{E}(\phi | T)) = \mathbb{E}_\theta(\phi), \forall \theta \in \Theta$. So it has the same level or power. ■

**Remark.** Given $H_0 : \mu = \mu_0$ and $H_1 : \mu \neq \mu_0$. We don't have the existence of UMP in this case of a both sided problem.

## Lecture 15

4 Apr. 08:30

**Lemma 5.3.2.** The MP-$\alpha$ test is essentially unique.

**Proof.** We had early considered $E_1, E_2, E_3$ where $\lambda(\widetilde{x})$ is $> k_0, = k_0, < k_0$ respectively where $\mathbb{E}_{\theta_0} = (\mathbb{P}_{\theta_0}(\widetilde{X} \in E_1)) + \gamma\mathbb{P}_{\theta_0}(\widetilde{X} \in E_2) = \gamma$. Now consider any other test $\phi'$ of size-$\alpha$. We have $Q(\widetilde{x}) = (\phi(\widetilde{x}) - \phi'(\widetilde{x}))(f_1(\widetilde{x}) - k_0 f_0(\widetilde{x}))$. So we have $\int Q(\widetilde{x})d\widetilde{x} = \int_{E_1} + \int_{E_2} + \int_{E_3} \geq 0$. This is equal to $\mathbb{E}_{\theta_1}(\phi) - \mathbb{E}_{\theta_1}(\phi') - k_0(\mathbb{E}_{\theta_0}(\phi) - \mathbb{E}_{\theta_0}(\phi'))$. Now consider if $\phi'$ is also MP-$\alpha$ we have $\mathbb{E}_{\theta_1}(\phi') = \mathbb{E}_{\theta_1}(\phi)$ as $\phi$ is also MP-$\alpha$. This takes the integrals to 0, which means that $\phi = \phi'$ on $E_1$ and $E_3$. ∎

## 5.4 Gamma Distribution

**Example.** Consider $X_1, \ldots X_n \sim N(0, \sigma^2)$ with $H_0 : \sigma = \sigma_0$ against $H_1 : \sigma = \sigma_1$. For this we would need to consider the distribution of $\sum_i (X_i^2)$.

**Definition 5.4.1** (Gamma Function)**.** $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ for $\alpha > 0$.

**Exercise.** Apply integration by parts to show that for $\alpha \geq 1$, $\Gamma(\alpha) = \alpha\Gamma(\alpha - 1)$. Furthermore we can show that $\Gamma(n) = n!$.

**Remark.** We have $\Gamma(1) = 1$ and $\Gamma(\frac{1}{2}) = \pi$.

**Definition 5.4.2** (Density function of gamma distribution)**.** Consider the integral $\int_0^\infty \frac{y^{\alpha-1}e^{-\frac{y}{b}}}{\beta^\alpha\Gamma(\alpha)} = 1$. This is the density function of the gamma distribution with paramaters $\alpha, \beta$, written as $\Gamma(\alpha, \beta)$.

**Remark.** MGF of the gamma distribution. This is given by $\mathbb{E}(e^{tX}) = \int_0^\infty \frac{e^{tX}x^{\alpha-1}e^{\frac{x}{\beta}}}{\beta^\alpha\Gamma(\alpha)}dx = \frac{(1-\beta t)^\alpha}{\beta^2\alpha}$.

**Remark.** Consider $X_1, \ldots X_n \sim \Gamma(\alpha_i, \beta)$. We have the MGF of the joint distribution as $\frac{(1-\beta t)^{\sum_i \alpha_i}}{\beta^{\sum_i 2\alpha_i}}$ which is the MGF of $\Gamma(\sum_i \alpha_i, \beta)$.

**Exercise.** Find the distribution of $X^2$ when $X \sim N(0, 1)$.

**Answer.** This is given by $\mathbb{P}(X^2 \leq x) = \mathbb{P}(-\sqrt{x} \leq X \leq \sqrt{x}) = \mathbb{P}(x \leq \sqrt{x}) - \mathbb{P}(X \leq -\sqrt{x})$. This can be written as $F_X(\sqrt{(x)}) - F_X(-\sqrt{x})$. Now differentiating we get $f_X(\sqrt{x})(\frac{1}{w\sqrt{x}}) + f_X(\sqrt{-x})(\frac{1}{2\sqrt{x}}) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x}{2}}x^{-\frac{1}{2}} = \frac{x^{-\frac{1}{2}}e^{-\frac{x}{2}}}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})}$. This implies that $X^2 \sim \Gamma(\frac{1}{2}, 2)$. ⊛

**Remark.** Let us say $X \sim N(\mu, \sigma^2)$. The distribution of $(\frac{x-\mu}{\sigma})^2$ follows $\Gamma(\frac{1}{2}, 2)$.

**Remark.** Consider $X_1, \ldots, X_n \sim N(0, 1)$. We have $\sum_i (X_i^2) \sim \Gamma(\frac{n}{2}, 2)$. This is called the chi-squared distribution and denoted by $\chi_n^2$

# Lecture 17

## 5.5 T-test

9 Apr. 08:30

**Definition 5.5.1** (Chi-squared distribution)**.** We have $\Gamma(\frac{m}{2}, 2) \equiv \chi_m^2$.

**Corollary 5.5.1** (Properties of $\chi^2$).
- Additivity: $\chi^2_m + \chi^2_n = \chi^2_{m+n}$
- If $X \sim N(0,1)$ then $X^2 \sim \chi^2_1$ so if $X_i \sim N(\mu_1, \sigma_i^2)$ then $\sum_i (\frac{X_i - \mu_i}{\sigma_i})^2 \sim \chi^2_n$.

**Exercise.** Find the MP test of $X_1, \ldots, X_n \sim N(\mu_0, \sigma^2)$ where $H_0 : \sigma = \sigma_0$ and $H_1 : \sigma = \sigma_1$ with $\sigma_1 > \sigma_0$.

**Answer.** We have $\lambda(\widetilde{x})$. The exponent term is given by $\exp\Big(-\frac{1}{2}(\sum_i (x_i - \mu_0))^2 (\frac{1}{\sigma_1^2}) - \frac{1}{\sigma_0^2}\Big)$. Sp we clearly have $\lambda(\widetilde{x})$ as monotonically increasing with $\sum_i (x_i - \mu_0)^2$ and we can use some $c'$ with it for the decision boundary. We know that $\sum_i (\frac{X_i - \mu_0}{\sigma_0})^2 \sim \chi^2_n$ under $H_0$. So if $\sum_i (\frac{X_i - \mu_0}{\sigma_0})^2 > c''(= \frac{c'}{\sigma_0^2})$. So $\chi^2_{\alpha;n}$ is the point such that $\mathbb{P}(X \geq \chi^2_{\alpha;n}) = \alpha$ where $X \sim \chi^2_n$. So we reject $H_0$ if $\sum_i (X_i - \mu_0)^2 > \sigma_0^2 \chi^2_{\alpha;n}$. Since this is independent of the choice of $\sigma_1$ this is a UMP-$\alpha$ test. ⊛

**Corollary 5.5.2.** If $X \sim N(0,1)$ and $Y \sim \chi^2_n$ and both are independent, then $\frac{X}{\sqrt{\frac{Y}{n}}} \sim t_n$.

**Corollary 5.5.3.** Consider $X_1, \ldots, X_n \sim N(\mu_x \sigma_x^2)$ and $Y_1, \ldots, Y_m \sim N(\mu_y, \sigma_y^2)$. Consider $Z_1^X = \sum_i a_{1i} X_i, \ldots, Z_p^X = \sum_i a_{pi} X_i$. Similarly consider $Z_1^Y = \sum_j b_{1j} Y_j, \ldots, Z_q^Y = \sum_j b_{qj} Y_j$. So $Z^X$ is a vector of dimension $p \times 1$ and $Z^Y$ is a vector of dimension $q \times 1$. These are pairwise independent.

**Theorem 5.5.1** (Normal distribution and $\chi^2$ ). Consider $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. We have that:
- $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$
- $S^2 = \frac{1}{n-1} \sum_i (X_i - \overline{X})^2$ and $(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$.
- $\overline{X} and S^2$ are independent distributed

**Proof.** Consider $(n-1)S^2 = \sum_i (X_1 - \overline{X})^2 = [(X_1 - \overline{X})^2 + \sum_{i=2}(X_i - \overline{X})^2] = [(\sum_{i=2} X_i - \overline{X})^2 + \sum_{i=2}(X_i - \overline{X})^2] = f(X_2 - \overline{X}, \ldots, X_n - \overline{X})$. So we need to show that $\overline{X}$ and $\begin{pmatrix} X_2 - \overline{X} \\ \vdots \\ X_n - \overline{X} \end{pmatrix}$ are independent. Now consider $\text{Cov}\left[\overline{X}, X_2 - \overline{X}\right] = \text{Cov}\left[\frac{\sum_i X_i}{n}, X_2\right] - \text{Var}\left[\overline{X}\right] = \frac{\sigma^2}{n} - \frac{\sigma^2}{n} = 0$. So they are independent. ■

**Example.** Consider $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$. We have $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$ and $\mu_1 > \mu_0$. Since we do not know $\sigma$ we cannot use the statistic we used earlier for the critical region. Consider $\frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$ and $\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$ and they are independent. So we have $\frac{\overline{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} \cdot \frac{1}{\sqrt{(n-1)\frac{S^2}{\sigma^2(n-1)}}} \sim t_{n-1}$. So ultimately we have the test as $\frac{\overline{x} - \mu_0}{\sqrt{\frac{S^2}{n}}} > t_{\alpha;n-1}$ where $t_{\alpha;n-1}$ is such that $\mathbb{P}(X > t_{\alpha;n-1}) = \alpha$ for $X \sim t_{n-1}$. From the other direction this is given by $< t_{1-\alpha;n-1} = -t_{\alpha;n-1}$

**Definition 5.5.2** ($f$-distribution). Consider $X \sim \chi^2_m$ and $Y \sim \chi^2_n$ independent. Then $F = \frac{\frac{X}{m}}{\frac{Y}{n}} = \frac{nX}{mY} \sim F_{m,n}$.

# Lecture 18

11 Apr. 08:30

**Exercise.** Write down the density of $t$.

**Answer.** We have $X \sim N(0,1)$ and $Y \sim \chi^2_n$. Now consider $T = \frac{X}{\sqrt{\frac{Y}{n}}} \sim T_n$. Now consider $(X,Y) \to (T,W)$ where $W = Y$. We have $f_{X,Y}(X,Y) = f_X(x)f_Y(y) = \frac{1}{\sqrt{2\pi e^{-\frac{x^2}{2}}}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} y^{\frac{n}{2}-1}e^{-\frac{y}{2}}$. We have $X = (T,W)$ and $Y = W$. So we have the density function as $\frac{1}{\sqrt{2\pi}} e^{\frac{-(t\sqrt{\frac{w}{n}})^2}{2}} \frac{1}{\Gamma(\frac{n}{2})2^{\frac{n}{2}}} w^{\frac{n}{2}-1}e^{-\frac{w}{2}}\sqrt{\frac{w}{n}}$. Now we integrate with respect to $w$ to get $f_T(t) = \frac{1}{\sqrt{2\pi}\Gamma(\frac{n}{2})2^{\frac{n}{2}}\sqrt{n}}\Gamma(\frac{n+1}{2})(\frac{2}{1+\frac{t^2}{n}})^{\frac{n+1}{2}} = \frac{1}{\sqrt{n\pi}\Gamma(\frac{n}{2})}\Gamma(\frac{n+1}{2})(1+\frac{t^2}{2})^{-\frac{n+1}{2}}$ ⊛

# Lecture 19

## 5.6 Likelihood Ratio Test

16 Apr. 08:30

**Definition 5.6.1** (Likelihood Ratio Test)**.** Consider $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$. Now consider $\lambda(\widetilde{x}) = \frac{\sup_{\theta \in \Theta_0} f(\widetilde{x},\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} f(\widetilde{x},\theta)}$. We see that $0 \leq \lambda(\widetilde{x}) \leq 1$ and values of $\lambda(\widetilde{x})$ indicate rejection of $H_0$. So we reject $H_0$ when $\lambda(\widetilde{x}) < c$ and accept $H_0$ otherwise where $c$ is such that the test is of size $\alpha$.

**Corollary 5.6.1** (Monotone Ratio Property)**.** Consider $\lambda(\widetilde{x})$ when $f(\widetilde{x},\theta)$ is the minimal sufficient statistic $T = g(t,\theta)h(x_1,\dots,x_n)$. So we have $\lambda(\widetilde{x}) = \frac{\sup_{\theta \in \Theta_0} g(t,\theta)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} g(t,\theta)}$. The family satisfies the monotone likelihood ratio property with the minimal sufficient statistic. So $\lambda(\widetilde{x}) < c_1 \Leftrightarrow t < c'$.

**Example.** Consider $X_1,\dots,X_n \sim N(\mu,\sigma_0^2)$ with $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. Consider $\lambda(\widetilde{x}) = \frac{\sup_{\theta \in \Theta_0} g(\overline{x},\mu)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} g(\overline{x}.\mu)} = \frac{g(\overline{x},\mu_0)}{g(\overline{x},\overline{x})} = \exp\Big(-\frac{n}{2}(\frac{\overline{x}-\mu_0}{\sigma_0})^2\Big)$. Since we want $\lambda(\widetilde{x}) < c$, this is equivalent to $|\overline{x} > \mu_0| > c'$. So $\mathbb{P}_{H_0}(|\overline{X} - \mu_0| > c') = \mathbb{P}_{H_0}(|Z| > c'\frac{\sqrt{n}}{\sigma_0})$. So $c' = \Phi_{\frac{\alpha}{2}}\frac{\sigma_0}{\sqrt{n}}$.

**Example.** Consider $X_1,\dots,X_n \sim N(\mu,\sigma_0^2)$ with $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$. We have $\lambda(\widetilde{x}) = \frac{g(\overline{x},\mu_0)}{\sup_{\theta \in \Theta_0 \cup \Theta_1} g(\overline{x},\mu)}$. We have $\sup_{\theta \in \Theta_0 \cup \Theta_1} g(\overline{x},\mu) = g(\overline{x},\overline{x})$ if $\overline{x} \geq \mu_0$ and $g(\overline{x},\mu_0)$ otherwise. So the ratio would be $\frac{g(\widetilde{x},\mu_0)}{g(\overline{x},\widetilde{x})}$ or 1 for the cases above respectively. So overall the function is monotonically decreasing with $\overline{x}$. So now we choose $\overline{x} > c'$ such that the test if of size $\alpha$ using $N(0,1)$. This gives the same test as the Neyman-Pearson UMP test.

# Lecture 20

18 Apr. 08:30

**Example.** Consider $X_1,\dots,X_n \sim N(\mu,\sigma^2)$ with $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. So we have $\lambda(\widetilde{x}) = \frac{\sup_{\mu,\sigma \in \Theta_0} f(\widetilde{x},\mu,\sigma)}{\sup_{\mu,\sigma \in \Theta_0 \cup \Theta_1} f(\widetilde{x},\mu,\sigma)}$. We have $(\hat{\mu},\hat{\sigma}) = (0,\frac{1}{n}\sum_i x_i^2)$ for the numerator and $(\hat{\mu},\hat{\sigma}) = (\overline{x},\frac{1}{n})\sum_i(x_i - \overline{x})^2$. Now consider the numerator after substituting which is $\frac{1}{(2\pi\frac{\sum_i x_i^2}{n})^{\frac{n}{2}}}\exp\big(-\frac{n}{2}\big)$. Similarly for the denominator we have $\frac{1}{(2\pi\frac{\sum_i(x_i-\overline{x})^2}{2})^{\frac{n}{2}}}\exp\big(-\frac{n}{2}\big)$. So we have $\lambda^{\frac{2}{n}}(\widetilde{x}) = \frac{\sum_i(x_i-\overline{x})^2}{\sum_i x_i^2} = \frac{\sum_i(x_i-\overline{x})^2}{\sum_i(x_i-\overline{x})^2+n\overline{x}^2} = \frac{1}{1+\frac{n\overline{x}^2}{\sum_i(x_i-\overline{x})^2}}$. So $\lambda(\widetilde{x})$ is monotonically decreasing with $\frac{n\overline{x}^2}{\sum_i(x_i-\overline{x})^2}$. So we have $\lambda^{\frac{2}{n}} < c' \Leftrightarrow \frac{n\overline{x}^2}{\sum_i(x_i-\overline{x})^2} > c'' \Leftrightarrow \frac{\sqrt{n}|\overline{x}|}{\sqrt{\sum_i(x_i-\overline{x})^2\frac{1}{n-1}}} > c'''$. We see that the final distribution is a T distribution with so $c''' = t_{\frac{\alpha}{2};n-1}$.

**Exercise.** Take $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$.

**Example.** Consider $X_1, \ldots, X_m \sim N(\mu, \sigma^2)$ and $Y_1, \ldots, Y_n \sim N(\mu_2, \sigma^2)$ with $H_0 : \mu_1 = \mu_2$ against $H_2 : \mu_1 \neq \mu_2$. We have $\sigma^2$ is unknown, but have the same variance. This is called a homoscedastic assumption. So we have $\lambda(\widetilde{x}, \widetilde{y}) = \frac{\sup_{\mu_1,\mu_2,\sigma\in\Theta_0} f(\widetilde{x},\widetilde{y},\mu_1,\mu_2,\sigma)}{\sup_{\mu_1,\mu_2,\sigma\in\Theta_0\cup\Theta_1} f(\widetilde{x},\widetilde{y},\mu_1,\mu_2,\sigma)}$. For the numerator we have $\hat{(\mu_1)} = \hat{\mu_2} = \frac{\sum_i x_i + \sum_i y_i}{m+n} = \frac{m\overline{x}+n\overline{y}}{m+n} = M$ and $\hat{\sigma^2} = \frac{1}{n+m}(\sum_i (x_i - \frac{m\overline{x}+n\overline{y}}{n+m})^2 + \sum_j (y_j - \frac{m\overline{x}+n\overline{y}}{m+n})^2) = Q$. Similarly under $H_0 \cup H_1$ we have $\hat{\mu_1} = \overline{x} = M_x$, $\hat{\mu_2} = \overline{y} = M_y$ and $\hat{\sigma^2} = \frac{\sum_i (x_i - \overline{x})^2 + \sum_j (y_j - \overline{y})^2}{m+n} = Q'$. So for the numerator we have $\frac{1}{(2\pi)^{\frac{n+m}{2}}} \exp\Big(\frac{1}{2Q}(\sum_i (x_i - M)^2 + \sum_j (y_j - M)^2)\Big) = \frac{1}{(2\pi Q)^{\frac{n+m}{2}}} \exp(-\frac{n+m}{2})$. Similarly the denominator will be $\frac{1}{(2\pi Q')^{\frac{n+m}{2}}} \exp(-\frac{n+m}{2})$. So we have
$\lambda^{\frac{2}{m+n}}(\widetilde{x}, \widetilde{y}) = \frac{Q'}{Q} = \frac{\sum_i (x_i = \overline{x})^2 + \sum_j (y_j - \overline{y})^2}{\sum_i (x_i - M)^2 + \sum_j (y_j - M)^2}$. We can now show that $\sum_i (x_i - M)^2 + \sum_j (y_j - M)^2 = \sum_i (x_i - \overline{x})^2 + \sum_j (y_j - \overline{y})^2 + \frac{nm}{n+m}(\overline{x} - \overline{y})^2$. So $\lambda(\widetilde{x}, \widetilde{y}) = \frac{Q'}{Q' + \frac{nm}{n+m}(\overline{x} - \overline{y})^2} = \frac{1}{1 + \frac{\frac{nm}{n+m}(\overline{x}-\overline{y})^2}{Q'}}$. So we have $\lambda^{\frac{2}{n+m}} < c_1 \Leftrightarrow \frac{\frac{nm}{n+m}(\overline{x}-\overline{y})^2}{Q'} > c_2 \Leftrightarrow \frac{\sqrt{\frac{nm}{n+m}}|\overline{x} - |}{\sqrt{\frac{Q'}{n+m-2}}} > c_3$. Since the last distribution is $T$ we have $c_3 = t_{\frac{\alpha}{2};n+m-2}$.

# Lecture 21

## 5.7 Analysis of Variance

20 Apr. 08:30

**Example.** Consider many groups $X_{11}, \ldots, X_{1a} \sim N(\mu_1, \sigma^2) \ldots X_{b1}, \ldots X_{bn} \sim N(\mu_b, \sigma^2)$. We have $\sigma$ unknown but same across groups (homoscedastic setup). Now consider $H_0 : \mu_1 = \mu_b$ against $H_1 : \mu_i \neq \mu_j$ for some $1 \leq i \leq j \leq b$. Now consider $\lambda(\widetilde{x}) = \frac{\sup_{\Theta_0} f(\widetilde{x},\mu_1,\ldots,\mu_b,\sigma^2)}{\sup_{\Theta_0\cup\Theta_1} f(\widetilde{x},\mu_1,\ldots,\mu_b,\sigma^2)}$. For the numerator we have $\hat{\mu} = \frac{\sum_{i=1}^b \sum_{j=1}^a x_{ij}}{ab} = \overline{x}_{00}$ and $\hat{\sigma^2} = \frac{1}{ab} \sum_{i=1}^b \sum_{j=1}^a (x_{ij} - \overline{x}_{00})^2 = Q_1$. For the denominator we have $\hat{\mu_i} = \frac{\sum_{j=1}^a x_{ij}}{a} = \overline{x}_{i0}$ and $\hat{\sigma^2} = \frac{1}{ab} \sum_{i=1}^b \sum_{j=1}^a (x_{ij} - \overline{x}_{i0})^2 = Q_2$. So the numerator is given by $\frac{1}{(2\pi Q_1)^{\frac{ab}{2}}} \exp\Big(-\frac{1}{2Q_1^2} \sum_i \sum_j (x_{ij} - \overline{x}_{00})^2\Big) = \frac{1}{(2\pi Q_1)^{\frac{ab}{2}}} \exp(-\frac{ab}{2})$. Similarly for the denominator we have $\frac{1}{(2\pi Q_2)^{\frac{ab}{2}}} \exp(-\frac{ab}{2})$. So we have $\lambda(\widetilde{x})^{\frac{2}{ab}} = \frac{Q_2}{Q_1}$ so $\lambda(\widetilde{x}) < c_1 \Leftrightarrow \frac{Q_2}{Q_1} < c_2$. We have $Q_1 = Q_2 + Q_3$ where $Q_3 = \sum_{i=1}^b (\overline{x}_{i0} - \overline{x}_{00})^2$. So we have $\frac{Q_3}{Q_2} > c_3$. Here we have $Q_2$ as the variance due to error which is unexplained by the model and $Q_3$ as the variance due to groups which is explained by the model. $Q_3$ SSB which is between sum of squares and $Q_2$ is SSE. We have $\frac{abQ_2}{\sigma^2} \sim \chi^2_{b(a-1)}$ under $H_0$. We also have $\frac{abQ_3}{\sigma^2} \sim \chi^2_{b-1}$. So we have $\frac{\frac{Q_3}{b-1(\sigma^2)}}{\frac{Q_2}{b(a-1)\sigma^2}} > c_4 \sim F_{b-1,b(a-1)}$. So set $c_4 = F_{\alpha;b-1,b(a-1)}$.

# Lecture 22

From the above example we know that we can take the SSB which has $b-1$ degrees of freedom, with corresponding $\text{MSB} = \frac{\text{SSB}}{b-1}$. Similarly we have SSE with $ab - b$ degrees of freedom and corresponding $\text{MSE} = \frac{\text{SSE}}{ab-b}$. So we have $\frac{\text{MSB}}{\text{MSE}} \sim F_{b-1,ab-b}$. Lastly the SST has degrees of freedom $ab - b + b - 1 = ab - 1$. 23 Apr. 08:30

## 5.8 Maximum Likelihood Estimation

Consider $X_1, \ldots, X_n \sim f_\theta$. Consider $f(\widetilde{x}, \theta) = f(x_1, \ldots, x_n, \theta)$ which is the joint pmf/pdf when observed as a function of $\widetilde{x}$. Now consider $L(x_1, \ldots, x_n, \theta)$ where $\widetilde{x}$ being observed is fixed and $L$ is considered as a function of $\theta$. Here $L$ is known as the likelihood function. We now want to maximise the value of $L$ over $\theta$. So $\hat{\theta}(x_1, \ldots, x_n)$ which is the MLE of $\theta$ us given by the value which maximises $L(\widetilde{x}, \theta)$.

**Example.** Consider $X_1, \dots, X_n \sim N(\mu, \sigma_0^2)$. We have $L(\widetilde{x}, \mu) = \frac{1}{(\sqrt{2\pi\sigma_0})^n} \exp\Big(\frac{-1}{2\sigma_0^2} \sum_i (x_i - \mu)^2\Big)$. Maximising this function is equivalent to minimising the summed term in the exponent. We know that the deviations are at a minium if and only if we have $\mu = \overline{x}$. Therefore the MLE of $\mu$ is given by $\hat{\mu}(\widetilde{x}) = \overline{x}$.

**Example.** Now consider $X_1, \dots, X_n \sim N(\mu, \sigma)$. The likelihood function is then $L(\widetilde{x}, \mu, \sigma) = \frac{1}{(\sqrt{2\pi\sigma})^n} \exp\big(-\frac{1}{2\sigma^2} \sum_i (x_i - \mu)^2\big)$. We first take the log likelihood. We then take the partial derivatives with respect to $\mu$ and $\sigma$ and equate them to 0. We solve this system of equations to get $\hat{\mu} = \overline{x}$ and $\hat{\sigma^2} = \frac{1}{n} \sum_i (x_i - \overline{X})^2$ which is the sample variance.

**Exercise.** Compute $E(\hat{\sigma^2})$ and show that it is not equal to $\sigma^2$. Show that $\mathbb{E}(\frac{1}{n-1} \sum_i (X_i - \overline{X})^2) = \sigma^2$.

## Lecture 22

25 Apr. 08:30

**Example.** Consider $X_1, \dots, X_n \sim U(0, \theta)$. We want the MLE for $\theta$ where $\Theta = (0, \infty)$. We have $L(\widetilde{x}, \theta) = \frac{1}{\theta^n} \mathbb{1}_{X_{(n)} \leq \theta}$. This function is not differentiable but it is obvious that $L(\widetilde{x}, \theta)$ is maximised at $X_{(n)}$ so this is the MLE.

**Example.** Consider $X_1, \dots, X_n \sim P(\lambda)$. We have $L(\widetilde{x}, \theta) = \frac{e^{-n\lambda} \lambda^{\sum_i x_i}}{\prod_{i=1}^n x_i!}$. We have $\log(L(\widetilde{x}, \theta)) = -n\lambda + \sum_i x_i \log(\lambda) - C$. We now take $\frac{\mathrm{d}L}{\mathrm{d}\lambda} = -n + \frac{\sum_i x_i}{\lambda}$. Equating this to 0 we get $\hat{\lambda} = \frac{\sum_i x_i}{n} = \overline{X}$.

**Example.** Consider $X_1, \dots, X_n \sim \mathrm{Ber}(P)$. Let $H_0 : p = \frac{1}{2}$ and $H_1 : p = \frac{3}{4}$. Obtain the MP-$\alpha$ test.