

# **Theoretical Statistics: Topics for a Core Course**

## **Problem Solutions**

*Robert W. Keener*

HYUNSUNG KIM  
Department of Statistics  
Chung-Ang University

Update: December 1, 2021

# Contents

<b>1</b>	<b>Probability and Measure</b>	<b>3</b>
<b>2</b>	<b>Exponential Families</b>	<b>19</b>
<b>3</b>	<b>Risk, Sufficiency, Completeness, and Ancillarity</b>	<b>31</b>
<b>4</b>	<b>Unbiased Estimation</b>	<b>40</b>
<b>5</b>	<b>Curved Exponential Families</b>	<b>59</b>
<b>6</b>	<b>Conditional Distributions</b>	<b>67</b>
<b>7</b>	<b>Bayesian Estimation</b>	<b>70</b>
<b>8</b>	<b>Large-Sample Theory</b>	<b>72</b>
<b>9</b>	<b>Estimating Equations and Maximum Likelihood</b>	<b>76</b>

# Chapter 1

## Probability and Measure

### Problem 1.1

Prove (1.1). If measurable sets  $B_n$ ,  $n \geq 1$ , are increasing, with  $B = \bigcup_{n=1}^{\infty} B_n$ , called the limit of the sequence, then

$$\mu(B) = \lim_{n \rightarrow \infty} \mu(B_n).$$

**Solution.**

First, we showed that  $A_n$ 's are disjoint. If  $j < k$ , then  $B_j \subseteq B_{k-1}$ .

Since  $A_j \subset B_j \subseteq B_{k-1}$  and  $A_k \subset B_{k-1}^c$ ,  $A_j$  and  $A_k$  are disjoint.

Also  $B_n = \bigcup_{j=1}^n A_j$  and  $\bigcup_{n=1}^{\infty} A_n = B$ ,

$$\mu(B) = \sum_{i=1}^{\infty} \mu(A_i) = \lim_{n \rightarrow \infty} \sum_{i=1}^n \mu(A_i) = \lim_{n \rightarrow \infty} \mu\left(\bigcup_{i=1}^n A_i\right) = \lim_{n \rightarrow \infty} \mu(B_n).$$

□

### Problem 1.8

Prove *Boole's inequality*: For any events  $B_1, B_2, \dots$ ,

$$P\left(\bigcup_{i \geq 1} B_i\right) \leq \sum_{i \geq 1} P(B_i).$$

**Solution.**

Let  $B = \bigcup_{i=1}^{\infty} B_i$ , then  $1_B \leq \sum 1_{B_i}$ . Then by Fubini's theorem,

$$P(B) = \int 1_B dP \leq \int \sum 1_{B_i} dP = \sum \int 1_{B_i} dP = \sum P(B_i).$$

□

### Problem 1.10

Let  $\mu$  and  $\nu$  be measures on  $(\mathcal{E}, \mathcal{B})$ .

a) Show that the sum  $\eta$  defined by  $\eta(B) = \mu(B) + \nu(B)$  is also a measure.

***Solution.***

We should check following 2 conditions.

- (i) For arbitrary set  $A \in \mathcal{B}$ ,  $\mu(A) \geq 0$  and  $\nu(A) \geq 0$ .  $\Rightarrow \eta(A) = \mu(A) + \nu(A) \geq 0$ .  
 $\therefore \eta : \mathcal{B} \rightarrow [0, \infty]$ .
- (ii) For disjoint set  $B_1, B_2, \dots \subset \mathcal{B}$ , then

$$\begin{aligned} \eta\left(\bigcup_{i=1}^{\infty} B_i\right) &= \mu\left(\bigcup_{i=1}^{\infty} B_i\right) + \nu\left(\bigcup_{i=1}^{\infty} B_i\right) \\ &= \sum \mu(B_i) + \sum \nu(B_i) \\ &= \sum \{\mu(B_i) + \nu(B_i)\} \\ &= \sum \eta(B_i) \end{aligned}$$

$\therefore \eta$  is a measure. □

b) If  $f$  is a non-negative measurable function, show that

$$\int f d\eta = \int f d\mu + \int f d\nu.$$

***Solution.***

We show it by 2 stage.

- (i) Let  $f = \sum_{i=1}^n a_i 1_{A_i}$ , the non-negative simple function. Then,

$$\begin{aligned} \int f d\eta &= \int \sum_{i=1}^n a_i 1_{A_i} d\eta = \sum a_i \eta(A_i) \\ &= \sum a_i \{\mu(A_i) + \nu(A_i)\} = \int f d\mu + \int f d\nu. \end{aligned}$$

- (ii) For general case, let  $f_n$  is the sequence of non-negative simple functions increasing to  $f$ . (i.e.  $f_1 \leq f_2 \leq \dots \leq f$ )

$$\begin{aligned} \int f d\eta &= \lim_{n \rightarrow \infty} \int f_n d\eta = \lim_{n \rightarrow \infty} \left( \int f_n d\mu + \int f_n d\nu \right) \\ &= \lim_{n \rightarrow \infty} \int f_n d\mu + \lim_{n \rightarrow \infty} \int f_n d\nu = \int f d\mu + \int f d\nu. \end{aligned}$$

□

## Problem 1.11

Suppose  $f$  is the simple function  $1_{(1/2, \pi]} + 21_{(1, 2]}$ , and let  $\mu$  be a measure on  $\mathbb{R}$  with  $\mu\{(0, a^2]\} = a$ ,  $a > 0$ . Evaluate  $\int f d\mu$ .

**Solution.**

By the integral of simple function and the finite additivity, it can be simply computed as

$$\begin{aligned}
 \int f d\mu &= \mu\{(1/2, \pi]\} + 2\mu\{(1, 2]\} \\
 &= [\mu\{(0, \pi]\} - \mu\{(1/2, \pi]\}] + 2[\mu\{(0, 2]\} - \mu\{(1, 2]\}] \quad (\text{finite additivity}) \\
 &= (\sqrt{\pi} - 1/\sqrt{2}) + 2(\sqrt{2} - 1)
 \end{aligned}$$

□

**Problem 1.12**

Suppose that  $\mu\{(0, a)\} = a^2$  for  $a > 0$  and that  $f$  is defined by

$$f(x) = \begin{cases} 0, & x \leq 0, \\ 1, & 0 < x < 2, \\ \pi, & 2 \leq x < 5, \\ 0, & x \geq 5. \end{cases}$$

Compute  $\int f d\mu$ .

**Solution.**

Since  $f = 1_{(0,2)} + \pi 1_{[2,5)}$  is simple, the integral can be computed as

$$\begin{aligned}
 \int f d\mu &= \mu\{(0, 2)\} + \pi\mu\{[2, 5)\} \\
 &= \mu\{(0, 2)\} + \pi[\mu\{(0, 5)\} - \mu\{(0, 2)\}] \\
 &= 4 - \pi(25 - 4)
 \end{aligned}$$

□

**Problem 1.13**

Define the function  $f$  by

$$f(x) = \begin{cases} x, & 0 \leq x \leq 1, \\ 0, & \text{otherwise.} \end{cases}$$

Find simple functions  $f_1 \leq f_2 \leq \dots$  increasing to  $f$  (i.e.  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$  for all  $x \in \mathbb{R}$ ). Let  $\mu$  be Lebesgue measure on  $\mathbb{R}$ . Using our formal definition of an integral and the fact that  $\mu((a, b]) = b - a$  whenever  $b > a$  (this might be used to formally define Lebesgue measure), show that  $\int f d\mu = 1/2$ .

**Solution.**

Let  $f_n = \lfloor 2^n x \rfloor / 2^n$  for  $0 < x \leq 1$  and 0 otherwise. ( $\lfloor y \rfloor$  is a floor function.) Then,

$$f_1(x) = \lfloor 2x \rfloor / 2 = \begin{cases} 0, & x < 1/2, \\ 1/2, & 1/2 \leq x < 1, \\ 1, & x = 1 \end{cases}$$

$$\begin{aligned}
f_2(x) &= \lfloor 2^2 x \rfloor / 2^2 = \begin{cases} 0, & x < 1/2^2, \\ 1/2^2, & 1/2^2 \leq x < 2/2^2, \\ 2/2^2, & 2/2^2 \leq x < 3/2^2, \\ 3/2^2, & 3/2^2 \leq x < 1, \\ 1, & x = 1 \end{cases} \\
&\vdots \\
f_n(x) &= \lfloor 2^n x \rfloor / 2^n = \begin{cases} 0, & x < 1/2^n, \\ 1/2^n, & 1/2^n \leq x < 2/2^n, \\ \vdots \\ (2^n - 1)/2^n, & (2^n - 1)/2^n \leq x < 1, \\ 1, & x = 1 \end{cases}
\end{aligned}$$

$$\begin{aligned}
\therefore \int f_n d\mu &= \frac{1}{2^n} \left( \frac{1}{2^n} + \frac{2}{2^n} + \cdots + \frac{2^n - 1}{2^n} \right) \\
&= \frac{1 + 2 + \cdots + (2^n - 1)}{4^n} \\
&= \frac{2^n(2^n - 1)}{2 \cdot 4^n} \\
&\longrightarrow \frac{1}{2}.
\end{aligned}$$

My solution

Let the simple function  $f_n = \sum_{i=1}^n \frac{i}{n} 1_{(\frac{i-1}{n}, \frac{i}{n}]}$ . Then,

$$\int f_n d\mu = \sum_{i=1}^n \frac{i}{n} \frac{1}{n} = \frac{1}{n^2} \frac{n(n+1)}{2} \longrightarrow \frac{1}{2}.$$

□

## Problem 1.16

Define  $F(a-) = \lim_{x \uparrow a} F(x)$ . Then, if  $F$  is non-decreasing,  $F(a-) = \lim_{n \rightarrow \infty} F(a - 1/n)$ . Use (1.1)[Continuity of measure] to show that if a random variable  $X$  has cumulative distribution function  $F_X$ ,

$$P(X < a) = F_X(a-).$$

Also, show that

$$P(X = a) = F_X(a) - F_X(a-).$$

**Solution.**

(i) Let  $B_n = \{X \leq a - 1/n\}$  and  $\bigcup_{n=1}^{\infty} B_n = \{X < a\}$ . By continuity of measure,

$$P(B) = P(X < a) = \lim_{n \rightarrow \infty} P(X \leq a - 1/n) = \lim_{n \rightarrow \infty} F_X(a - 1/n) = F_X(a-).$$

(ii) Since  $\{X < a\}$  and  $\{X = a\}$  are disjoint with union  $\{X \leq a\}$ ,

$$P(X < a) + P(X = a) = P(X \leq a).$$

$$\therefore P(X = a) = F_X(a) - F_X(a-).$$

□

## Problem 1.17

Suppose  $X$  is a geometric random variable with mass function

$$p(x) = P(X = x) = \theta(1 - \theta)^x, \quad x = 0, 1, \dots,$$

where  $\theta \in (0, 1)$  is a constant. Find the probability that  $X$  is even.

**Solution.**

$$\begin{aligned} P(X \text{ is even}) &= P(X = 0) + P(X = 2) + P(X = 4) \cdots \\ &= \theta + \theta(1 - \theta)^2 + \theta(1 - \theta)^4 + \cdots \\ &= \frac{\theta}{1 - (1 - \theta)^2} \\ &= \frac{1}{2 - \theta} \end{aligned}$$

□

## Problem 1.18

Let  $X$  be a function mapping  $\mathcal{E}$  into  $\mathbb{R}$ . Recall that if  $B$  is a subset of  $\mathbb{R}$ , then  $X^{-1}(B) = \{e \in \mathcal{E} : X(e) \in B\}$ . Use this definition to prove that

$$X^{-1}(A \cap B) = X^{-1}(A) \cap X^{-1}(B),$$

$$X^{-1}(A \cup B) = X^{-1}(A) \cup X^{-1}(B),$$

and

$$X^{-1}\left(\bigcup_{i=0}^{\infty} A_i\right) = \bigcup_{i=0}^{\infty} X^{-1}(A_i).$$

**Solution.**

(i)

$$\begin{aligned} e \in X^{-1}(A \cap B) &\Leftrightarrow X(e) \in A \cap B \\ &\Leftrightarrow X(e) \in A \text{ and } X(e) \in B \\ &\Leftrightarrow e \in X^{-1}(A) \text{ and } e \in X^{-1}(B) \\ &\Leftrightarrow e \in X^{-1}(A) \cap X^{-1}(B). \end{aligned}$$

(ii) Similarly,

$$\begin{aligned}
 e \in X^{-1}(A \cup B) &\Leftrightarrow X(e) \in A \cup B \\
 &\Leftrightarrow X(e) \in A \text{ or } X(e) \in B \\
 &\Leftrightarrow e \in X^{-1}(A) \text{ or } e \in X^{-1}(B) \\
 &\Leftrightarrow e \in X^{-1}(A) \cup X^{-1}(B).
 \end{aligned}$$

(iii)

$$\begin{aligned}
 e \in X^{-1}\left(\bigcup_{i=0}^{\infty} A_i\right) &\Leftrightarrow X(e) \in \bigcup_{i=0}^{\infty} A_i \\
 &\Leftrightarrow X(e) \in A_i \text{ for some } i \\
 &\Leftrightarrow e \in X^{-1}(A_i) \text{ for some } i \\
 &\Leftrightarrow e \in \bigcup_{i=0}^{\infty} X^{-1}(A_i).
 \end{aligned}$$

□

## Problem 1.19

Let  $P$  be a probability measure on  $(\mathcal{E}, \mathcal{B})$ , and let  $X$  be a random variable. Show that the distribution  $P_X$  of  $X$  defined by  $P_X(B) = P(X \in B) = P(X^{-1}(B))$  is a measure (on the Borel sets of  $\mathbb{R}$ ).

### Solution.

To show that  $P_X$  is a measure, we should check 2 conditions.

(i)  $P_X(B) = P(X^{-1}(B)) \geq 0$ .

(ii) From Problem 1.18,  $X^{-1}(\bigcup_i A_i) = \bigcup_i X^{-1}(A_i)$  is hold.

Also, if  $B_i$  and  $B_j$  are disjoint, then  $X^{-1}(B_i)$  and  $X^{-1}(B_j)$  are disjoint. ( $\because$  If  $e \in \mathcal{E}$  lie in both  $X^{-1}(B_1)$  and  $X^{-1}(B_2)$ ,  $X(e)$  lies in  $B_1$  and  $B_2$ .)

Therefore, for arbitrary disjoint set  $B_1, B_2, \dots$  with union  $B = \bigcup_i B_i$ ,  $X^{-1}(B_1), X^{-1}(B_2), \dots$  are also disjoint with union  $X^{-1}(B)$ .

$$\therefore \sum_i P_X(B_i) = \sum_i P(X^{-1}(B_i)) = P(X^{-1}(B)) = P_X(B).$$

The above 2 conditions are hold,  $P_X$  is a measure.

□

## Problem 1.21

Let  $X$  have a uniform distribution on  $(0, 1)$ ; that is,  $X$  is absolutely continuous with density  $p$  defined by

$$p(x) = \begin{cases} 1, & x \in (0, 1), \\ 0, & \text{otherwise.} \end{cases}$$

Let  $Y_1$  and  $Y_2$  denote the first two digits of  $X$  when  $X$  is written as a binary decimal (so  $Y_1 = 0$  if  $X \in (0, 1/2)$  for instance). Find  $P(Y_1 = i, Y_2 = j)$ ,  $i = 0$  or  $1, j = 0$  or  $1$ .



**Solution.**

For all cases, the probability are same. In other words,

$$P(Y_1 = 0, Y_2 = 0) = P(Y_1 = 1, Y_2 = 0) = P(Y_1 = 0, Y_2 = 1) = P(Y_1 = 1, Y_2 = 1) = 1/4.$$

□

**Problem 1.22**

Let  $\mathcal{E} = (0, 1)$ , let  $\mathcal{B}$  be the Borel subsets of  $\mathcal{E}$ , and let  $P(A)$  be the length of  $A$  for  $A \in \mathcal{B}$ . ( $P$  would be called the *uniform probability measure* on  $(0, 1)$ .) Define the random variable  $X$  by

$$X(e) = \min\{e, 1/2\}.$$

Let  $\mu$  be the sum of Lebesgue measure on  $\mathbb{R}$  and counting measure on  $\mathcal{X}_0 = \{1/2\}$ . Show that the distribution  $P_X$  of  $X$  is absolutely continuous with respect to  $\mu$  and find the density of  $P_X$ .

**Solution.**

(i) Claim: For any  $B \in \mathbb{R}$ ,  $\mu(B) = 0 \implies P_X(B) = 0$ .

$$\begin{aligned} \{X \in B\} &= \{y \in (0, 1) : X(y) \in B\} \\ &= \begin{cases} B \cap (0, 1/2), & 1/2 \notin B, \\ (B \cap (0, 1/2)) \cup [1/2, 1), & 1/2 \in B. \end{cases} \end{aligned}$$

Let  $\lambda$  be a Lebesgue measure and  $\nu$  be a counting measure on  $\{1/2\}$ . Then,

$$\begin{aligned} \mu(B) = 0 &\iff \lambda(B) = 0 \text{ and } 1/2 \notin B. \quad (\mu(B) = \nu(B) = 0) \\ \implies P_X(B) &= P(X \in B) = P(B \cap (0, 1/2)) = \lambda(B \cap (0, 1/2)) = 0. \\ \therefore P_X &\text{ is absolutely continuous w.r.t. } \lambda + \nu \stackrel{\text{def}}{=} \mu. \end{aligned}$$

(ii) Find the density of  $P_X$ .

From the above equation,

$$P(X \in B) = \lambda(B \cap (0, 1/2)) + \frac{1}{2}1_B(1/2). \quad (1.1)$$

If  $f$  is the density, then (1.1) should be equal to  $\int_B f d(\lambda + \nu)$ .

$$\begin{aligned} \int_B f d(\lambda + \nu) &= \int f 1_B d(\lambda + \nu) \\ &= \int \left( 1_{B \cap (0, 1/2)} + \frac{1}{2}1_{\{1/2\} \cap B} \right) d(\lambda + \nu) \\ &= \lambda(B \cap (0, 1/2)) + \nu(B \cap (0, 1/2)) + \frac{1}{2}\lambda(\{1/2\} \cap B) + \frac{1}{2}\nu(\{1/2\} \cap B) \\ &= \lambda(B \cap (0, 1/2)) + \frac{1}{2}1_B(1/2). \end{aligned}$$

$$\therefore f = 1_{(0, 1/2)} + \frac{1}{2}1_{\{1/2\}}.$$

□

### Problem 1.23

The standard normal distribution  $N(0, 1)$  has density  $\phi$  given by

$$\phi(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}, \quad x \in \mathbb{R},$$

with respect to Lebesgue measure  $\lambda$  on  $\mathbb{R}$ . The corresponding cumulative distribution function is  $\Phi$ , so

$$\Phi(x) = \int_{-\infty}^x \phi(z) dz$$

for  $x \in \mathbb{R}$ . Suppose that  $X \sim N(0, 1)$  and that the random variable  $Y$  equals  $X$  when  $|X| < 1$  and is 0 otherwise. Let  $P_Y$  denote the distribution of  $Y$  and let  $\mu$  be counting measure on  $\{0\}$ . Find the density of  $P_Y$  with respect to  $\lambda + \mu$ .

**Solution.**

Let  $g(x) = x1_{|x|<1}$ , then  $Y = g(X)$ .

For any integrable function  $f$ , the expectation of  $f(Y)$  against the density of  $X$  is

$$Ef(Y) = Ef(g(X)) = \int f(g(x))\phi(x)dx = \int f(g(x))\phi(x)1_{(-1,1)}(x)dx + cf(0), \quad (1.2)$$

where  $c = \int \phi(x)1_{|x|>1}(x)dx = 2\Phi(-1)$ .

Similarly, the expectation of  $f(Y)$  against the density  $p$  of  $Y$  is

$$Ef(Y) = \int f p d(\lambda + \mu) = \int f p d\lambda + \int f p d\mu = \int f(x)p(x)dx + f(0)p(0). \quad (1.3)$$

Then, (1.2) and (1.3) should be same. Therefore,

$$\begin{aligned} f = 1_{\{0\}} &\implies p(0) = 2\Phi(-1), \\ f(0) = 0 &\implies \int f(x)\phi(x)1_{(-1,1)}(x)dx = \int f(x)p(x)dx \\ &\implies p(x) = \phi(x) \text{ for } 0 < |x| < 1. \end{aligned}$$

$\therefore p(x) = 2\Phi(-1)1_{\{0\}}(x) + \phi(x)1_{(0,1)}(|x|)$  is the density of  $P_Y$ .

□

### Problem 1.24

Let  $\mu$  be a  $\sigma$ -finite measure on a measurable space  $(X, \mathcal{B})$ . Show that  $\mu$  is absolutely continuous with respect to some probability measure  $P$ .

**Hint:** You can use the fact that if  $\mu_1, \mu_2, \dots$  are probability measures and  $c_1, c_2, \dots$  are non-negative constants, then  $\sum c_i \mu_i$  is a measure. (The proof for Problem 1.10 extends easily to this case.) The measure  $\mu_i$ , you will want to consider, are truncations of  $\mu$  to sets  $A_i$  covering  $X$  with  $\mu(A_i) < \infty$ , given by  $\mu_i(B) = \mu(B \cap A_i)$ . With the constants  $c_i$  chosen properly,  $\sum c_i \mu_i$  will be a probability measure.

**Solution.**

(i)  $\mu$  is finite  $\implies$  Trivial.

$\therefore$  Since  $\mu(X) < \infty$ ,  $P = \mu/\mu(X)$  be a probability measure. Therefore,  $P(N) = 0$  implies  $\mu(N) = 0$ .

(ii)  $\mu$  is infinite but  $\sigma$ -finite.

$\implies \exists A_1, A_2, \dots \in \mathcal{B}$  s.t.  $\bigcup A_i = X$  and  $0 < \mu(A_i) < \infty \forall i$ .

From Hint,  $\mu_i(B) = \mu(B \cap A_i)$  and also  $\mu_i(X) = \mu(A_i) \implies \mu_i$  is a finite measure.

Let  $b_i = 1/2^i$  (or any other seq of positive constants s.t.  $\sum b_i = 1$ ) and define  $c_i = b_i/\mu(A_i)$ .

Then  $P = \sum_i c_i \mu_i$  is a probability measure, since  $P(X) = \sum_i c_i \mu_i(X) = \sum_i \left( \frac{b_i}{\mu(A_i)} \right) \mu(A_i) = 1$ .

Suppose  $P(N) = \sum_i c_i \mu_i(N) = 0$ . Then  $\mu_i(N) = \mu(N \cap A_i) = 0 \forall i$ .

By Boole's inequality,

$$\mu(N) = \mu\left(\bigcup_i (N \cap A_i)\right) \leq \sum_i \mu(N \cap A_i) = 0.$$

For any null set for  $P$  is a null set for  $\mu$ .

$\therefore \mu \ll P$ .

□

## Problem 1.25

The monotone convergence theorem states that if  $0 \leq f_1 \leq f_2 \leq \dots$  are measurable functions and  $f = \lim_{n \rightarrow \infty} f_n$ , then  $\int f d\mu = \lim_{n \rightarrow \infty} \int f_n d\mu$ . Use this result to prove the following assertions.

- a) Show that if  $X \sim P_X$  is a random variable on  $(\mathcal{E}, \mathcal{B}, P)$  and  $f$  is a non-negative measurable function, then

$$\int f(X(e)) dP(e) = \int f(x) dP_X(x).$$

Hint: Try it first with  $f$  an indicator function. For the general case, let  $f_n$  be a sequence of simple functions increasing to  $f$ .

### Solution.

We will show the equation by 3 steps.

- (i) Suppose  $f = 1_A$  (indicator function).

Then,  $f(X(e)) = 1$  for  $X(e) \in A$ .

$\implies f \circ X = 1_B$  where  $B = \{e : X(e) \in A\}$ .

By definition,  $P_X(A) = P(B)$ .

$$\begin{aligned} \int f(X(e)) dP(e) &= \int 1_B(e) dP(e) = P(B), \\ \int f(x) dP_X(x) &= \int 1_A(x) dP_X(x) = P_X(A). \end{aligned}$$

$\therefore$  The statement is holds.

- (ii) Suppose  $f = \sum_{i=1}^n c_i 1_{A_i}$  (simple function).

$$\begin{aligned} \therefore \int f(X(e)) dP(e) &= \int \sum_i c_i 1_{A_i}(X(e)) dP(e) \\ &= \sum_i c_i \int 1_{A_i}(X(e)) dP(e) \\ &= \sum_i c_i \int 1_{A_i}(x) dP_X(x) \quad (\text{by (i)}) \end{aligned}$$

$$\begin{aligned}
 &= \int \sum_i c_i 1_{A_i}(x) dP_X(x) \\
 &= \int f(x) dP_X(x).
 \end{aligned}$$

- (iii) Suppose  $f$  is a non-negative measurable function (general case), and  $f_n$  is non-negative simple functions increasing to  $f$ .

$$\Rightarrow f_n \circ X \nearrow f \circ X.$$

$$\begin{aligned}
 \therefore \int f(X(e)) dP(e) &= \lim_{n \rightarrow \infty} \int f_n(X(e)) dP(e) \quad (\text{by M.C.T.}) \\
 &= \lim_{n \rightarrow \infty} \int f_n(x) dP_X(x) \quad (\text{by (ii)}) \\
 &= \int f(x) dP_X(x) \quad (\text{by M.C.T.})
 \end{aligned}$$

□

- b) Suppose that  $P_X$  has density  $p$  with respect to  $\mu$ , and let  $f$  be a non-negative measurable function. Show that

$$\int f dP_X = \int f p d\mu.$$

**Solution.**

It can be also showed by 3 steps.

- (i) Let  $f = 1_A$  (indicator function).

$$\int f dP_X = \int 1_A dP_X = P_X(A) = \int_A p d\mu = \int 1_A p d\mu = \int f p d\mu.$$

- (ii) Suppose  $f = \sum_{i=1}^n c_i 1_{A_i}$  (simple function).

$$\int f dP_X = \int \sum_i c_i 1_{A_i} dP_X = \sum_i c_i \int 1_{A_i} dP_X \stackrel{(i)}{=} \sum_i c_i \int 1_{A_i} p d\mu = \int \sum_i c_i 1_{A_i} p d\mu = \int f p d\mu.$$

- (iii) For general case, let a non-negative measurable function  $f$  and non-negative simple functions  $f_n$  s.t.  $f_n \nearrow f$ . Then by M.C.T.,

$$\int f dP_X \stackrel{\text{M.C.T.}}{=} \lim_{n \rightarrow \infty} \int f_n dP_X \stackrel{(ii)}{=} \lim_{n \rightarrow \infty} \int f_n p d\mu \stackrel{\text{M.C.T.}}{=} \int f p d\mu.$$

□

## Problem 1.26

*The gamma distribution.*

- a) The gamma function is defined for  $\alpha > 0$  by

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx.$$

Use integration by parts to show that  $\Gamma(x+1) = x\Gamma(x)$ . Show that  $\Gamma(x+1) = x!$  for  $x = 0, 1, \dots$

**Solution.**

By integration by parts,

$$\Gamma(x+1) = \int_0^\infty t^x e^{-t} dt = -t^x e^{-t} \Big|_0^\infty + (\alpha-1) \int_0^\infty t^{x-1} e^{-t} dt = x\Gamma(x).$$

Since  $\Gamma(1) = \int_0^\infty e^{-x} dx = 1$ ,

$$\Gamma(x+1) = x\Gamma(x) = x(x-1)\Gamma(x-1) = \cdots = x(x-1)\cdots 2 \cdot 1 \cdot \Gamma(1) = x!.$$

□

b) Show that the function

$$p(x) = \begin{cases} \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, & x > 0, \\ 0, & \text{otherwise,} \end{cases}$$

is a (Lebesgue) probability density when  $\alpha > 0$  and  $\beta > 0$ . This density is called the gamma density with parameters  $\alpha$  and  $\beta$ . The corresponding probability distribution is denoted  $\Gamma(\alpha, \beta)$ .

**Solution.**

Using change of variable,  $y = x/\beta \Rightarrow dx = \beta dy$ . Therefore,

$$\begin{aligned} \int p(x) dx &= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta} dx \\ &= \int \frac{1}{\Gamma(\alpha)\beta^\alpha} (\beta y)^{\alpha-1} e^{-y} \beta dy \\ &= \frac{1}{\Gamma(\alpha)} \int y^{\alpha-1} e^{-y} dy \\ &= 1. \end{aligned}$$

□

c) Show that if  $X \sim \Gamma(\alpha, \beta)$ , then  $EX^r = \beta^r \Gamma(\alpha + r)/\Gamma(\alpha)$ . Use this formula to find the mean and variance of  $X$ .

**Solution.**

$$\begin{aligned} EX^r &= \int_0^\infty x^r p(x) dx \\ &= \int_0^\infty \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha+r-1} e^{-x/\beta} dx \\ &= \int_0^\infty \frac{1}{\Gamma(\alpha+r)\beta^{\alpha+r}} x^{\alpha+r-1} e^{-x/\beta} dx \frac{\Gamma(\alpha+r)}{\Gamma(\alpha)} \beta^r \\ &= \beta^r \Gamma(\alpha+r)/\Gamma(\alpha). \end{aligned}$$

By using this,

$$\begin{aligned} EX &= \beta \Gamma(\alpha+1)/\Gamma(\alpha) = \alpha\beta, \\ EX^2 &= \beta^2 \Gamma(\alpha+2)/\Gamma(\alpha) = (\alpha+1)\alpha\beta^2, \\ Var(X) &= EX^2 - \{EX\}^2 = \alpha\beta^2. \end{aligned}$$

□

**Problem 1.27**

Suppose  $X$  has a uniform distribution on  $(0, 1)$ . Find the mean and covariance matrix of the random vector  $\begin{pmatrix} X \\ X^2 \end{pmatrix}$ .

**Solution.**

Since  $X \sim \text{Uniform}(0, 1)$ , the density of  $X$  is  $p(x) = 1$  for  $x \in (0, 1)$ . Thus,

$$EX^r = \int x^r dx = \frac{1}{r+1}.$$

Therefore,

$$E\left(\begin{pmatrix} X \\ X^2 \end{pmatrix}\right) = \begin{pmatrix} EX \\ EX^2 \end{pmatrix},$$

$$\text{Cov}\left(\begin{pmatrix} X \\ X^2 \end{pmatrix}\right) = \begin{pmatrix} \text{Var}(X) & \text{Cov}(X, X^2) \\ \text{Cov}(X^2, X) & \text{Var}(X^2) \end{pmatrix} = \begin{pmatrix} EX^2 - \{EX\}^2 & EX^3 - EXEX^2 \\ EX^3 - EXEX^2 & EX^4 - \{EX^2\}^2 \end{pmatrix}.$$

□

**Problem 1.28**

If  $X \sim N(0, 1)$ , find the mean and covariance matrix of the random vector  $\begin{pmatrix} X \\ I_{\{X > c\}} \end{pmatrix}$ .

**Solution.**

Let  $\Phi(x)$  is a cumulative distribution function of  $X$ , and  $\phi$  is a density of  $X$ . Then,

$$E\left(\begin{pmatrix} X \\ I_{\{X > c\}} \end{pmatrix}\right) = \begin{pmatrix} EX \\ EI_{\{X > c\}} \end{pmatrix} = \begin{pmatrix} 0 \\ 1 - \Phi(c) \end{pmatrix},$$

$$\text{Cov}\left(\begin{pmatrix} X \\ I_{\{X > c\}} \end{pmatrix}\right) = \begin{pmatrix} \text{Var}(X) & EX1_{X>c} - EXEI_{X>c} \\ EX1_{X>c} - EXEI_{X>c} & E1_{X>c}^2 - \{EI_{X>c}\}^2 \end{pmatrix} = \begin{pmatrix} 1 & \phi(c) \\ \phi(c) & \Phi(c)(1 - \Phi(c)) \end{pmatrix}.$$

$$\because EX1_{X>c} = \int_c^\infty \frac{1}{\sqrt{2\pi}} xe^{-x^2/2} dx = \phi(c) \text{ and } 1_{X>c}^2 = 1_{X>c}.$$

□

**Problem 1.32**

Suppose  $E|X| < \infty$  and let

$$h(t) = \frac{1 - E \cos(tX)}{t^2}.$$

Use Fubini's theorem to find  $\int_0^\infty h(t) dt$ . Hint:

$$\int_0^\infty (1 - \cos(u)) u^{-2} du = \frac{\pi}{2}.$$

**Solution.**

By Fubini's theorem,

$$\int_0^\infty h(t) dt = \int_0^\infty \left( \frac{1 - E \cos(tX)}{t^2} \right) dt$$

$$\begin{aligned}
&= \int_0^\infty E\left(\frac{1 - \cos(tX)}{t^2}\right) dt \\
&= \int_0^\infty \int \frac{1 - \cos(tx)}{t^2} dP_X(x) dt \\
&= \int \int_0^\infty \frac{1 - \cos(tx)}{t^2} dt dP_X(x) \quad (\text{by Fubini's theorem}) \\
&\quad u = tx \Rightarrow du = x dt, t = x/u \\
&= \int \int_0^\infty |x| \frac{1 - \cos(u)}{u^2} du dP_X(x) \\
&= \int \frac{\pi}{2} |x| dP_X(x) \\
&= \frac{\pi}{2} E|X|.
\end{aligned}$$

□

### Problem 1.33

Suppose  $X$  is absolutely continuous with density  $p_X(x) = xe^{-x}$ ,  $x > 0$  and  $p_X(x) = 0$ ,  $x \leq 0$ . Define  $c_n = E(1 + X)^{-n}$ . Use Fubini's theorem to evaluate  $\sum_{n=1}^\infty c_n$ .

**Solution.**

By Fubini's theorem,

$$\begin{aligned}
\sum_{n=1}^\infty c_n &= \sum_n E(1 + X)^{-n} \\
&= \sum_n \int_0^\infty (1 + x)^{-n} x e^{-x} dx \\
&= \int_0^\infty \sum_n (1 + x)^{-n} x e^{-x} dx \\
&\text{Since } \left| \frac{1}{1+x} \right| < 1, \quad \sum_{n=1}^\infty \left( \frac{1}{1+x} \right)^n = \frac{1/(1+x)}{1 - 1/(1+x)} = \frac{1}{x}. \\
&= \int_0^\infty \frac{1}{x} x e^{-x} dx \\
&= \int_0^\infty e^{-x} dx \\
&= 1.
\end{aligned}$$

□

### Problem 1.36

Suppose  $X$  and  $Y$  are independent random variables, and let  $F_X$  and  $F_Y$  denote their cumulative distribution functions.

- a) Use smoothing to show that the cumulative distribution function of  $S = X + Y$  is

$$F_S(s) = P(X + Y \leq s) = EF_X(s - Y). \quad (1.4)$$

**Solution.**

Since  $X$  and  $Y$  are independent,  $P(X + Y \leq s | Y = y) = P(X \leq s - y) = F_X(s - y)$ .

Thus,  $P(X + Y \leq s | Y) = F_X(s - Y)$ .

$$\therefore F_S(s) = P(X + Y \leq s) = EP(X + Y \leq s | Y) = EF_X(s - Y).$$

□

- b) If  $X$  and  $Y$  are independent and  $Y$  is almost surely positive, use smoothing to show that the cumulative distribution function of  $W = XY$  is  $F_W(w) = EF_X(w/Y)$  for  $w > 0$ .

**Solution.**

Since  $X$  and  $Y$  are independent,  $P(XY \leq w | Y = y) = P(X \leq w/y | Y = y) = P(X \leq w/y) = F_X(w/y)$ , for  $y > 0$ .

Thus,  $P(XY \leq w | Y) = F_X(w/Y)$  a.s.

$$\therefore F_W(w) = P(XY \leq w) = EP(XY \leq w | Y) = EF_X(w/Y).$$

□

**Problem 1.37**

Differentiating (1.4) with respect to  $s$  one can show that if  $X$  is absolutely continuous with density  $p_X$ , then  $S = X + Y$  is absolutely continuous with density

$$p_S(s) = Ep_X(s - Y)$$

for  $s \in \mathbb{R}$ . Use this formula to show that if  $X$  and  $Y$  are independent with  $X \sim \Gamma(\alpha, 1)$  and  $Y \sim \Gamma(\beta, 1)$ , then  $X + Y \sim \Gamma(\alpha + \beta, 1)$ .

**Solution.**

Since  $X \sim \Gamma(\alpha, 1)$ ,  $P_X(x - Y) = 0$  for  $Y \geq x$  ( $\because p_X(x) = 0, \forall x < 0$ ). Then,

$$\begin{aligned} P_S(x) &= EP_X(x - Y) \\ &= E \left[ \frac{1}{\Gamma(\alpha)} (x - Y)^{\alpha-1} e^{-(x-Y)} 1_{(0,x)}(Y) \right] \\ &= \int_0^x \frac{1}{\Gamma(\alpha)} (x - y)^{\alpha-1} e^{-(x-y)} \frac{1}{\Gamma(\beta)} y^{\beta-1} e^{-y} dy \\ &= \int_0^x \frac{1}{\Gamma(\alpha)\Gamma(\beta)} (x - y)^{\alpha-1} y^{\beta-1} e^{-x} dy \end{aligned}$$

Change the variable by  $u = y/x \Rightarrow du = 1/x dy$ .

$$\begin{aligned} &= \int_0^1 \frac{1}{\Gamma(\alpha)\Gamma(\beta)} (x - xu)^{\alpha-1} (xu)^{\beta-1} e^{-x} x du \\ &= \int_0^1 \frac{1}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha+\beta-1} (1 - u)^{\alpha-1} u^{\beta-1} e^{-x} du \\ &= \frac{1}{\Gamma(\alpha + \beta)} x^{\alpha+\beta-1} e^{-x}, \quad \text{for } x > 0. \end{aligned}$$

$\therefore S = X + Y \sim \Gamma(\alpha + \beta, 1)$ .

□



### Problem 1.38

Let  $Q_\lambda$  denote the exponential distribution with failure rate  $\lambda$ , given in Problem 1.30. Let  $X$  be a discrete random variable taking values in  $\{1, \dots, n\}$  with mass function

$$P(X = k) = \frac{2k}{n(n+1)}, \quad k = 1, \dots, n,$$

and assume that the conditional distribution of  $Y$  given  $X = x$  is exponential with failure rate  $x$ ,

$$Y|X = x \sim Q_x.$$

- a) Find  $E[Y|X]$ .

**Solution.**

From Problem 1.30,  $q_\lambda(x) = \lambda e^{-\lambda x}$ ,  $\lambda > 0$ . Then, the mean of  $Q_\lambda$  is

$$\begin{aligned} \int x dQ_\lambda(x) &= \int x q_\lambda(x) dx = \int_0^\infty x \lambda e^{-\lambda x} dx = \frac{1}{\lambda}. \\ \therefore E(Y|X = x) &= \frac{1}{x} \Rightarrow EY|X = \frac{1}{X}. \end{aligned}$$

□

- b) Use smoothing to compute  $EY$ .

**Solution.**

$$EY = E[E(Y|X)] = E\frac{1}{X} = \sum_{x=1}^n \frac{1}{x} \frac{2x}{n(n+1)} = \sum_{x=1}^n \frac{2}{n(n+1)} = \frac{2}{n+1}.$$

□

### Problem 1.39

Let  $X$  be a discrete random variable uniformly distributed on  $\{1, \dots, n\}$ , so  $P(X = k) = 1/n$ ,  $k = 1, \dots, n$ , and assume that the conditional distribution of  $Y$  given  $X = x$  is exponential with failure rate  $x$ .

- a) For  $y > 0$  find  $P[Y > y|X]$ .

**Solution.**

$$\begin{aligned} P(Y > y|X = x) &= \int_y^\infty x e^{-xt} dt = -e^{-xt} \Big|_y^\infty = e^{-xy}. \\ \therefore P(Y > y|X) &= e^{-XY}. \end{aligned}$$

□

- b) Use smoothing to compute  $P(Y > y)$ .

**Solution.**

Using the sum of the geometric sequence (등비급수의 합),

$$P(Y > y) = EP(Y > y|X) = Ee^{-XY} = \sum_{x=1}^n e^{-xy} \frac{1}{n} = \frac{1}{n} \frac{e^{-y}(1 - e^{-ny})}{1 - e^{-y}} = \frac{1 - e^{-ny}}{n(e^y - 1)}.$$

□

c) Determine the density of  $Y$ .

**Solution.**

Let the density of  $Y$  is  $p_Y$ . Then,

$$\begin{aligned}
 p_Y(y) &= \frac{d}{dy} (1 - P(Y > y)) \\
 &= \frac{d}{dy} \left( \frac{1 - e^{-ny}}{n(e^y - 1)} \right) \\
 &= \frac{-ne^{-ny}(ne^y - n) + ne^y(1 - e^{-ny})}{n^2(e^y - 1)^2} \\
 &= \frac{-ne^ye^{-ny} + ne^{-ny} + e^y - e^ye^{-ny}}{n(e^y - 1)^2} \\
 &= \frac{e^y(1 - e^{-ny})}{n(e^y - 1)^2} - \frac{e^{-ny}}{e^y - 1}.
 \end{aligned}$$

□

## Chapter 2

# Exponential Families

### Problem 2.1

Consider independent Bernoulli trials with success probability  $p$  and let  $X$  be the number of failures before the first success. Then  $P(X = x) = p(1 - p)^x$ , for  $x = 0, 1, \dots$ , and  $X$  has the geometric distribution with parameter  $p$ , introduced in Problem 1.17.

- a) Show that the geometric distributions form an exponential family.

**Solution.**

$$P(X = x) = p(1 - p)^x = \exp \{x \log(1 - p) - (-\log p)\}.$$

$\therefore$  the geometric distribution is an exponential family. □

- b) Write the densities for the family in canonical form, identifying the canonical parameter  $\eta$ , and the function  $A(\eta)$ .

**Solution.**

Let  $\eta(p) = \log(1 - p)$ . Then,  $P(X = x) = \exp \{ \eta x - (-\log(1 - e^\eta)) \}$ .

$\therefore A(\eta) = -\log(1 - e^\eta)$  with  $T(x) = x$ . □

- c) Find the mean of the geometric distribution using a differential identity.

**Solution.**

$$E_\eta(T) = A'(\eta) = \frac{e^\eta}{1 - e^\eta} = \frac{1 - p}{p}.$$

□

- d) Suppose  $X_1, \dots, X_n$  are i.i.d. from a geometric distribution. Show that the joint distributions form an exponential family, and find the mean and variance of  $T$ .

**Solution.**

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i) = p^n (1 - p)^{\sum_{i=1}^n x_i} = \exp \left\{ \sum_i x_i \log(1 - p) - (-n \log p) \right\}.$$

Therefore, the joint distribution is also exponential family.

Now, let  $\eta = \log(1 - p)$ . Then, the canonical exponential family is obtained with  $A(\eta) = -n \log(1 - e^\eta)$  and  $T = \sum_i x_i$ .

$$ET = \kappa_1 = A'(\eta) = \frac{ne^\eta}{1 - e^\eta} = \frac{n(1 - p)}{p},$$

$$Var(T) = \kappa_2 = A''(\eta) = \frac{ne^\eta(1 - e^\eta) + ne^{2\eta}}{(1 - e^\eta)^2} = \frac{np(1 - p) + n(1 - p)^2}{p^2} = \frac{n(1 - p)}{p^2}.$$

□

## Problem 2.2

Determine the canonical parameter space  $\Xi$ , and find densities for the one-parameter exponential family with  $\mu$  Lebesgue measure on  $\mathbb{R}^2$ ,  $h(x, y) = \exp[-(x^2 + y^2)/2]/(2\pi)$ , and  $T(x, y) = xy$ .

**Solution.**

By definition of the canonical exponential family, the pdf be represented as

$$p(x, y) = \exp(\eta T(x, y) - A(\eta))h(x, y) = \exp\{\eta xy - A(\eta) - (x^2 + y^2)/2\}/(2\pi).$$

Thus,

$$\begin{aligned} \int p(x, y) d\mu(x, y) = 1 &\iff \int \int \exp\{\eta xy - A(\eta) - (x^2 + y^2)/2\}/(2\pi) dx dy = 1 \\ &\iff e^{A(\eta)} = \int \int \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x - \eta y)^2\right\} dx \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(y^2 - \eta^2 y^2)} dy \\ &\iff e^{A(\eta)} = \frac{1}{\sqrt{2\pi}} \int e^{-\frac{1}{2}y^2(1 - \eta^2)} dy \quad \quad \quad \begin{matrix} \uparrow \\ \text{the integral is finite} \iff |\eta| < 1 \end{matrix} \quad (1 - \eta^2)^{-1/2}. \end{aligned}$$

$$\therefore A(\eta) = -\frac{1}{2} \log(1 - \eta^2) \text{ for } \eta \in \Xi = (-1, 1).$$

$\therefore$  The canonical exponential density is

$$\exp\left\{\eta xy + \frac{1}{2} \log(1 - \eta^2) - (x^2 + y^2)/2\right\}/(2\pi).$$

□

## Problem 2.4

Find the natural parameter space  $\Xi$  and densities  $p_\eta$  for a canonical one-parameter exponential family with  $\mu$  Lebesgue measure on  $\mathbb{R}$ ,  $T_1(x) = \log x$ , and  $h(x) = (1 - x)^2$ ,  $x \in (0, 1)$ , and  $h(x) = 0$ ,  $x \notin (0, 1)$ .

**Solution.**

$$\begin{aligned} e^{A(\eta)} &= \int_0^1 e^{\eta \log x} (1 - x)^2 dx \\ \text{Let } t = \log x &\Rightarrow dt = 1/x dx = e^{-t} dx, \quad -\infty < t < 0 \\ &= \int_{-\infty}^0 e^{\eta t} (1 - e^t)^2 e^t dt \end{aligned}$$

$$\begin{aligned}
&= \int_{-\infty}^0 \left( e^{t(\eta+1)} - 2e^{t(\eta+2)} + e^{t(\eta+3)} \right) dt \\
&= \left( \frac{1}{\eta+1} e^{t(\eta+1)} - \frac{2}{\eta+2} e^{t(\eta+2)} + \frac{1}{\eta+3} e^{t(\eta+3)} \right) \Big|_{-\infty}^0 \\
&= \frac{1}{\eta+1} - \frac{2}{\eta+2} + \frac{1}{\eta+3} \\
&= \frac{2}{(\eta+1)(\eta+2)(\eta+3)}, \quad \eta \in \Xi = (-1, \infty).
\end{aligned}$$

( $\because$  For a finite integral,  $\eta+1 > 0$ ,  $\eta+2 > 0$ , and  $\eta+3 > 0 \Rightarrow \eta > -1$ .)

$$\therefore p_\eta(x) = \exp(\eta T(x) - A(\eta)) h(x) = \frac{1}{2}(\eta+1)(\eta+2)(\eta+3)(1-x)^2 x^\eta, \quad x \in (0, 1).$$

□

## Problem 2.5

Find the natural parameter space  $\Xi$  and densities  $p_\eta$  for a canonical one-parameter exponential family with  $\mu$  Lebesgue measure on  $\mathbb{R}$ ,  $T_1(x) = -x$ , and  $h(x) = e^{-2\sqrt{x}}/\sqrt{x}$ ,  $x > 0$ , and  $h(x) = 0$ ,  $x \leq 0$ . (Hint: After a change of variables, relevant integrals will look like integrals against a normal density. You should be able to express the answer using  $\Phi$ , the standard normal cumulative distribution function.) Also, determine the mean and variance for a variable  $X$  with this density.

**Solution.**

$$e^{A(\eta)} = \int_0^\infty e^{-\eta x - 2\sqrt{x}} \frac{1}{\sqrt{x}} dx$$

To the integral is finite,  $\eta \in \Xi = [0, \infty)$ ,

Let  $t = \sqrt{x} \Rightarrow t^2 = x$ ,  $2t dt = dx$

$$\begin{aligned}
&= \int_0^\infty \frac{1}{t} e^{-\eta t^2 - 2t} 2t dt \\
&= 2 \int_0^\infty e^{-\eta(t^2 + \frac{2}{\eta}t)} dt \\
&= 2 \int_0^\infty e^{-\frac{2\eta}{2}(t + \frac{1}{\eta})^2} dt \cdot e^{-\eta(-\frac{1}{\eta^2})}
\end{aligned}$$

Since inner term of integration  $\sim N(-1/\eta, 1/2\eta)$ ,

$$\begin{aligned}
&= 2\sqrt{2\pi}/\sqrt{2\eta} \cdot e^{\frac{1}{\eta}} \Phi\left(\frac{-1/\eta}{\sqrt{1/(2\eta)}}\right) \\
&= \frac{2\sqrt{\pi}}{\sqrt{\eta}} e^{1/\eta} \Phi(-\sqrt{2/\eta})
\end{aligned}$$

$$\Rightarrow A(\eta) = \log 2\sqrt{\pi} - \frac{1}{2} \log \eta + \frac{1}{\eta} + \log \Phi\left(-\sqrt{\frac{2}{\eta}}\right).$$

$$\therefore p_\eta(x) = \exp\left\{-\eta x - 2\sqrt{x} - A(\eta)\right\} \frac{1}{\sqrt{x}}, \quad x > 0.$$

The mean of  $X$  is

$$E_\eta(X) = -E_\eta T = -A'(\eta) = \frac{1}{2\eta} + \frac{1}{\eta^2} - \frac{\frac{\sqrt{2}}{2}\eta^{-3/2}\phi\left(-\sqrt{\frac{2}{\eta}}\right)}{\Phi\left(-\sqrt{\frac{2}{\eta}}\right)}.$$

□

## Problem 2.6

Find the natural parameter space  $\Xi$  and densities  $p_\eta$  for a canonical two-parameter exponential family with  $\mu$  counting measure on  $\{0, 1, 2\}$ ,  $T_1(x) = x$ ,  $T_2(x) = x^2$ , and  $h(x) = 1$  for  $x \in \{0, 1, 2\}$ .

**Solution.**

$$\begin{aligned} e^{A(\eta)} &= \sum_{x=0}^2 e^{\eta_1 x + \eta_2 x^2} \\ &= 1 + e^{\eta_1 + \eta_2} + e^{2\eta_1 + 4\eta_2}. \end{aligned}$$

To the sum is finite,  $\eta_1 + \eta_2 < \infty$  &  $2\eta_1 + 4\eta_2 < \infty \Rightarrow \eta \in \Xi = \mathbb{R}^2$ .

$$\therefore p_\eta(x) = \exp \left[ \eta_1 x + \eta_2 x^2 - \left( 1 + e^{\eta_1 + \eta_2} + e^{2\eta_1 + 4\eta_2} \right) \right].$$

□

## Problem 2.7

Suppose  $X_1, \dots, X_n$  are independent geometric variables with  $p_i$  the success probability for  $X_i$ . Suppose these success probabilities are related to a sequence of "independent" variables  $t_1, \dots, t_n$ , viewed as known constants, through

$$p_i = 1 - \exp(\alpha + \beta t_i), \quad i = 1, \dots, n.$$

Show that the joint densities for  $X_1, \dots, X_n$  form a two-parameter exponential family, and identify the statistics  $T_1$  and  $T_2$ .

**Solution.**

Since  $X_i \stackrel{\text{ind}}{\sim} \text{Geo}(p_i)$ , then the joint density is represented by

$$\begin{aligned} \prod_{i=1}^n p(x_i) &= \prod_{i=1}^n (1 - p_i)^{x_i} p_i = \exp \left[ \sum_{i=1}^n x_i \log(1 - p_i) + \sum_{i=1}^n \log p_i \right] \\ &= \exp \left[ \sum_{i=1}^n x_i (\alpha + \beta t_i) + \sum_{i=1}^n \log(1 - \exp(\alpha + \beta t_i)) \right]. \end{aligned}$$

$$\therefore T_1 = \sum_{i=1}^n X_i, \quad T_2 = \sum_{i=1}^n t_i X_i.$$

□

## Problem 2.8

Assume that  $X_1, \dots, X_n$  are independent random variables with  $X_i \sim N(\alpha + \beta t_i, 1)$ , where  $t_1, \dots, t_n$  are observed constants and  $\alpha$  and  $\beta$  are unknown parameters. Show that the joint densities for  $X_1, \dots, X_n$  form a two-parameter exponential family, and identify the statistics  $T_1$  and  $T_2$ .

**Solution.**

The joint density of  $X_1, \dots, X_n$  is

$$\prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2} (x_i - \alpha - \beta t_i)^2 \right] = (2\pi)^{-\frac{n}{2}} \exp \left[ \sum_i x_i (\alpha + \beta t_i) - \frac{1}{2} \sum_i (\alpha + \beta t_i) \right] \exp \left( -\sum_i \frac{x_i^2}{2} \right).$$

$$\therefore T_1 = \sum_{i=1}^n X_i, \quad T_2 = \sum_{i=1}^n t_i X_i.$$

□

## Problem 2.9

Suppose that  $X_1, \dots, X_n$  are independent Bernoulli variables (a random variable is Bernoulli if it only takes on values 0 and 1) with

$$P(X_i = 1) = \frac{\exp(\alpha + \beta t_i)}{1 + \exp(\alpha + \beta t_i)}.$$

Show that the joint distribution for  $X_1, \dots, X_n$  form a two-parameter exponential family, and identify the statistics  $T_1$  and  $T_2$ .

### Solution.

The joint density of  $X_1, \dots, X_n$  is

$$\begin{aligned} \prod_{i=1}^n p_i^{x_i} (1 - p_i)^{1-x_i} &= \exp \left[ \sum_i x_i \log p_i + \sum_i (1 - x_i) \log(1 - p_i) \right] \\ &= \exp \left[ \sum_i x_i (\alpha + \beta t_i - \log(1 + e^{\alpha + \beta t_i})) + \sum_i (1 - x_i) (1 - \log(1 + e^{\alpha + \beta t_i})) \right] \\ &= \exp \left[ \sum_i x_i (\alpha + \beta t_i) - \sum_i \log(1 + e^{\alpha + \beta t_i}) \right]. \end{aligned}$$

$$\therefore T_1 = \sum_{i=1}^n X_i, \quad T_2 = \sum_{i=1}^n t_i X_i. \quad \square$$

## Problem 2.15

For an exponential family in canonical form,  $ET_j = \partial A(\eta)/\partial \eta_j$ . This can be written in vector form as  $ET = \nabla A(\eta)$ . Derive an analogous differential formula for  $E_\theta T$  for an  $s$ -parameter exponential family that is not in canonical form. Assume that  $\Omega$  has dimension  $s$ .

Hint: Differentiation under the integral sign should give a system of linear equations. Write these equations in matrix form.

### Solution.

Note that the exponential family form is

$$\begin{aligned} p_\eta(x) &= \exp \left[ \sum_{i=1}^s \eta_i T_i(x) - A(\eta) \right] h(x) \\ &= \exp \left[ \sum_{i=1}^s \eta_i(\theta) T_i(x) - B(\theta) \right] h(x) \\ &= p_\theta(x). \end{aligned}$$

Thus,

$$e^{B(\theta)} = \int \exp \left[ \sum_{i=1}^s \eta_i(\theta) T_i(x) \right] h(x) d\mu(x).$$

By differentiating with respect to  $\theta_i$ ,

$$e^{B(\theta)} \frac{\partial B(\theta)}{\partial \theta_i} = \int \left( \sum_{j=1}^s \frac{\partial \eta_j(\theta)}{\partial \theta_i} T_j(x) \right) \exp \left[ \sum_{j=1}^s \eta_j(\theta) T_j(x) \right] h(x) d\mu(x).$$

$$\begin{aligned}\frac{\partial B(\theta)}{\partial \theta_i} &= \int \left( \sum_{j=1}^s \frac{\partial \eta_j(\theta)}{\partial \theta_i} T_j(x) \right) p_\theta(x) d\mu(x) \\ &= \sum_{j=1}^s \frac{\partial \eta_j(\theta)}{\partial \theta_i} E_\theta T_j, \quad i = 1, \dots, s.\end{aligned}$$

□

## Problem 2.17

Let  $\mu$  denote counting measure on  $\{1, 2, \dots\}$ . One common definition for  $\sum_{k=1}^{\infty} f(k)$  is  $\lim_{n \rightarrow \infty} \sum_{k=1}^n f(k)$ , and another definition is  $\int f d\mu$ .

- a) Use the dominated convergence theorem to show that the two definitions give the same answer when  $\int |f| d\mu < \infty$ .

Hint: Find functions  $f_n$ ,  $n = 1, 2, \dots$ , so that  $\sum_{k=1}^n f(k) = \int f_n d\mu$ .

**Solution.**

Let  $f_n(k) = f(k)$ ,  $\forall k \leq n$ ,  $f_n(k) = 0$ ,  $\forall k > n$ . Then,  $f_n \rightarrow f$  pointwise and  $|f_n| \leq |f|$ . Also  $f_n$  is simple and  $\int f_n d\mu = \sum_{k=1}^n f(k)$ .

By D.C.T.,  $\int f_n d\mu \rightarrow \int f d\mu$ .

□

- b) Use the monotone convergence theorem, give in Problem 1.25, to show the definitions agree if  $f(k) \geq 0$  for all  $1, 2, \dots$ .

**Solution.**

Define  $f_n$  and  $f$  like part (a). Then,  $0 \leq f_1 \leq f_2 \leq \dots$ .

By M.C.T.,  $\sum_{k=1}^n f(k) = \int f_n d\mu \rightarrow \int f d\mu$ .

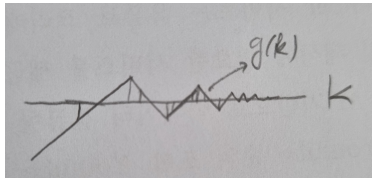
□

- c) Suppose  $\lim_{n \rightarrow \infty} f(n) = 0$  and that  $\int f^+ d\mu = \int f^- d\mu = \infty$  (so that  $\int f d\mu$  is undefined.) Let  $K$  be an arbitrary constant. Show that the list  $f(1), f(2), \dots$  can be rearranged to form a new list  $g(1), g(2), \dots$  so that

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n g(k) = K.$$

**Solution.**

Take positive parts of the sequence  $f$  until the sum is exceed  $K$ . And then, take negative parts of the sequence  $f$  until the sum is below  $K$ . Repeat this procedure like the figure below. (In the figure,  $g(k)$  is typo.  $\sum g(k)$  is correct.) Then, the sum of the sequence goes to 0 for  $k > k'$  where  $k'$  is



the number when the sum of the rearranged sequence exceed  $K$ , first. Let this sequence as  $g$ . Then,  $\lim_{n \rightarrow \infty} \sum_{k=1}^n g(k) \rightarrow K$ .

□



## Problem 2.19

Let  $p_n, n = 1, 2, \dots$ , and  $p$  be probability densities with respect to a measure  $\mu$ , and let  $P_n, n = 1, 2, \dots$ , and  $P$  be the corresponding probability measures.

- a) Show that if  $p_n(x) \rightarrow p(x)$  as  $n \rightarrow \infty$ , then  $\int |p_n - p| d\mu \rightarrow 0$ .

Hint: First use the fact that  $\int (p_n - p) d\mu = 0$  to argue that  $\int |p_n - p| d\mu = 2 \int (p - p_n)^+ d\mu$ . Then use dominated convergence.

**Solution.**

Since  $p_n$  and  $p$  be probability densities,  $\int (p_n - p) d\mu = \int p_n d\mu - \int p d\mu = 1 - 1 = 0$ . Since  $p_n - p = (p - p_n)^+ - (p - p_n)^-$ ,

$$\begin{aligned} \int (p_n - p) d\mu &= \int (p - p_n)^+ d\mu - \int (p - p_n)^- d\mu = 0, \\ \therefore \int (p - p_n)^+ d\mu &= \int (p - p_n)^- d\mu. \end{aligned}$$

Also  $|p_n - p| = (p - p_n)^+ + (p - p_n)^-$ ,

$$\int |p_n - p| d\mu = \int (p - p_n)^+ d\mu + \int (p - p_n)^- d\mu = 2 \int (p - p_n)^+ d\mu.$$

Also note that  $|(p - p_n)^+| \leq p \Rightarrow |(p - p_n)^+|$  is integrable. ( $\because p$  is a probability density  $\Rightarrow p$  is integrable.) From the condition,  $(p_n \rightarrow p) \Rightarrow (p(x) - p_n(x))^+ \rightarrow 0$ .

By D.C.T.,

$$\int |p_n - p| d\mu = 2 \int (p - p_n)^+ d\mu \rightarrow 0.$$

□

- b) Show that  $|P_n(A) - P(A)| \leq \int |p_n - p| d\mu$ .

Hint: Use indicators and the bound  $|\int f d\mu| \leq \int |f| d\mu$ .

**Solution.**

$$\begin{aligned} LHS &= \left| \int 1_A dP_n - \int 1_A dP \right| \\ &= \left| \int 1_A (p_n - p) d\mu \right| \\ &\leq \int |1_A (p_n - p)| d\mu \\ &\leq \int |p_n - p| d\mu. \end{aligned}$$

□

Remark: Distributions  $P_n, n \geq 1$ , are said to *converge strongly* to  $P$  if  $\sup_A |P_n(A) - P(A)| \rightarrow 0$ . The two parts above show that pointwise convergence of  $p_n$  to  $p$  implies strong convergence. This was discovered by Scheffé.

## Problem 2.22

Suppose  $X$  is absolutely continuous with density

$$p_\theta(x) = \begin{cases} \frac{e^{-(x-\theta)^2/2}}{\sqrt{2\pi}\Phi(\theta)}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

Find the moment generating function of  $X$ . Compute the mean and variance of  $X$ .

**Solution.**

$$p_\theta(x) = \exp \left[ -\frac{1}{2}(x^2 - 2x\theta + \theta^2) - \log \Phi(\theta) \right] \frac{1}{\sqrt{2\pi}}, \quad x > 0$$

is the exponential family with  $T = X$ ,  $A(\theta) = \frac{\theta^2}{2} + \log \Phi(\theta)$ . Therefore, the moment generating function of  $T = X$  is

$$\begin{aligned} M_X(u) &= \exp [A(\theta + u) - A(\theta)] \\ &= \exp \left[ \frac{1}{2}(\theta + u)^2 - \frac{1}{2}\theta^2 \right] \Phi(\theta + u)/\Phi(\theta) \\ &= \exp \left( \theta u + \frac{1}{2}u^2 \right) \Phi(\theta + u)/\Phi(\theta). \end{aligned}$$

Meanwhile, the *c.g.f.* of  $T = X$  for exponential family is  $K_X(u) = A(\theta + u) - A(\theta)$ . Thus,

$$\begin{aligned} EX &= K'_X(u) = A'(\theta) = \theta + \frac{\phi(\theta)}{\Phi(\theta)}, \\ \text{Var}(X) &= K''_X(u) = A''(\theta) = 1 + \frac{\phi'(\theta)\Phi(\theta) - \phi(\theta)^2}{\Phi(\theta)^2}. \\ &\left( \because \phi(x) = e^{-x^2/2}/\sqrt{2\pi} \Rightarrow \phi'(x) = -x\phi(x). \right) \end{aligned}$$

□

## Problem 2.23

Suppose  $Z \sim N(0, 1)$ . Find the first four cumulants of  $Z^2$ .

**Hint:** Consider the exponential family  $N(0, \sigma^2)$ .

**Solution.**

Let  $X \sim N(0, \sigma^2)$ . Then, its density is

$$p_\sigma(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi}} \exp \left[ -\frac{1}{2\sigma^2} x^2 - \log \sigma \right],$$

and also be the exponential family.

Reparameterize  $\eta = -\frac{1}{2\sigma^2}$ , then  $T = X^2$  and  $A(\eta) = \frac{1}{2} \log -\frac{1}{2\eta} = -\frac{1}{2} \log(-2\eta)$ . Thus, the cumulative generating function of  $T = X^2$  is

$$\begin{aligned} K_{X^2}(u) &= A(\eta + u) - A(\eta) \\ &= -\frac{1}{2} [\log(-2(\eta + u)) + \log(-2\eta)]. \end{aligned}$$

If  $\sigma = 1$ , then  $\eta = -\frac{1}{2}$  and  $X^2 = Z^2$ . Therefore,

$$K_{Z^2}(u) = -\frac{1}{2} [\log(-2u + 1)].$$

$$K'_{Z^2}(u) = \frac{1}{1-2u}, \quad K''_{Z^2}(u) = \frac{2}{(1-2u)^2},$$

$$K^{(3)}_{Z^2}(u) = 8(1-2u)^{-3}, \quad K^{(4)}_{Z^2}(u) = 48(1-2u)^{-4}.$$

$\therefore$  the cumulants are 1, 2, 8, and 48, respectively. □

## Problem 2.24

Find the first four cumulants of  $T = XY$  when  $X$  and  $Y$  are independent standard normal variates.

### Solution.

Since  $X$  and  $Y$  are independent, the function of these r.v.s are also independent. (i.e.  $g(X)$  and  $g(Y)$  are independent for any function  $g$ .) Then, first 2 cumulants are

$$\kappa_1 = ET = EXY = EXEY = 0,$$

$$\kappa_2 = E(T - ET)^2 = EX^2Y^2 = EX^2EY^2 = 1.$$

For third cumulants, we need to obtain  $EX^3$ .

$$\begin{aligned} EX^3 &= \int x^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ \text{Let } t &= x^2/2 \Rightarrow x dx = dt \\ &= \frac{2}{\sqrt{2\pi}} \int_0^\infty t e^{-t} dt \\ &= \frac{2}{\sqrt{2\pi}} \left[ -te^{-t} \Big|_0^\infty + \int_0^\infty e^{-t} dt \right] \\ &= 0. \end{aligned}$$

Thus, the third cumulants is

$$\kappa_3 = E(T - ET)^3 = EX^3Y^3 = EX^3EY^3 = 0.$$

Similarly, we need to obtain  $EX^4$ . [자꾸 적분이 틀림... 다시 해보기](#) Gamma dist 형태 사용

$$\begin{aligned} EX^4 &= \int x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \\ \text{Let } t &= x^2/2 \Rightarrow x dx = dt \end{aligned}$$

$$\kappa_4 = E(T - ET)^4 - 3\text{Var}(T)^2 = EX^4EY^4 - 3 \cdot 1^2 = 4.$$

□

## Problem 2.25

Find the third and fourth cumulants of the geometric distribution.

**Solution.**

$$f_p(x) = (1-p)^x p = \exp[x \log(1-p) + \log p]$$

is the exponential family. Reparameterize  $\eta = -\log(1-p)$  ( $\Rightarrow p = 1 - e^{-\eta}$ ), then the above form be the canonical exponential family with  $T = X$  and  $A(\eta) = -\log(1 - e^{-\eta})$ .

Thus, the cumulative generating function of  $T = X$  is

$$K_X(u) = A(\eta + u) - A(\eta) = -\log(1 - e^{\eta+u}) + \log(1 - e^{-\eta}).$$

Then, the first cumulants is

$$K'_X(0) = A'(\eta) = \frac{e^\eta}{1 - e^{-\eta}} = \frac{1-p}{p}.$$

The higher order cumulants are easily obtained by using the chain rule. By using the  $p' = \frac{dp}{d\eta} = -e^{-\eta} = p - 1$ ,

$$\begin{aligned} K''_X(0) &= A''(\eta) = \frac{d}{dp} A'(\eta) \frac{dp}{d\eta} = \frac{-p - (1-p)}{p^2} (p-1) = \frac{1}{p^2} - \frac{1}{p}, \\ K^{(3)}_X(0) &= A^{(3)}(\eta) = \frac{d}{dp} A''(\eta) p' = -\frac{2p'}{p^3} + \frac{p'}{p^2} = -\frac{2-2p}{p^3} + \frac{p-1}{p^2} = \frac{2}{p^3} - \frac{3}{p^2} + \frac{1}{p}, \\ K^{(4)}_X(0) &= A^{(4)}(\eta) = \frac{d}{dp} A^{(3)}(\eta) p' = -\frac{6p'}{p^4} + \frac{6p'}{p^3} - \frac{p'}{p^2} = \frac{6}{p^4} - \frac{12}{p^3} + \frac{7}{p^2} - \frac{1}{p}. \end{aligned}$$

□

## Problem 2.26

Find the third cumulant and third moment of the binomial distribution with  $n$  trials and success probability  $p$ .

**Solution.**

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} = \binom{n}{x} \left( \frac{p}{1-p} \right)^x (1-p)^n = \exp \left[ x \log \frac{p}{1-p} + n \log(1-p) \right] \binom{n}{x}$$

is the exponential family for  $\eta = \log \frac{p}{1-p}$  with  $T = X$  and  $A(\eta) = -n \log(1-p)$ . To use the chain rule,

$$p = \frac{e^\eta}{1 + e^\eta} \Rightarrow \frac{dp}{d\eta} = \frac{e^\eta(1 + e^\eta) - (e^\eta)^2}{(1 + e^\eta)^2} = p(1-p).$$

By using the above fact,

$$\begin{aligned} \kappa_1 &= A'(\eta) = \frac{d}{dp} A(\eta) \frac{dp}{d\eta} = \frac{n}{1-p} p(1-p) = np, \\ \kappa_2 &= A''(\eta) = \frac{d}{dp} A'(\eta) \frac{dp}{d\eta} = np(1-p), \\ \kappa_3 &= A'''(\eta) = \frac{d}{dp} A''(\eta) \frac{dp}{d\eta} = (n - 2np)p(1-p) = np(1-p)(1-2p). \end{aligned}$$

The third moment for  $T = X$  can be obtained by

$$EX^3 = \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3.$$

□

## Problem 2.27

Let  $T$  be a random vector in  $\mathbb{R}^2$ .

- a) Express  $\kappa_{2,1}$  as a function of the moments of  $T$ .

**Solution.**

To make simple derivation, we denote  $f_{ij}(u) = \frac{\partial^{i+j} f(u)}{\partial u_1^i \partial u_2^j}$ . Note that  $K(u) = \log M(u)$ , by taking derivative,

$$K_{10} = \frac{M_{10}}{M}, \quad K_{20} = \frac{M_{20}M - M_{10}^2}{M^2},$$

$$\begin{aligned} K_{21} &= \frac{1}{M^4} [(M_{21}M + M_{20}M_{01} - 2M_{10}M_{11})M^2 - 2MM_{01}(M_{20}M - M_{10}^2)] \\ &= \frac{1}{M^3} [M_{21}M^2 + M_{20}M_{01}M - 2M_{10}M_{11}M - 2M_{01}M_{20}M - 2M_{01}M_{10}^2] \\ &= \frac{1}{M^3} [M_{21}M^2 - M_{20}M_{01}M - 2M_{10}M_{11}M - 2M_{01}M_{10}^2]. \end{aligned}$$

Taking  $u = 0$ , then

$$\kappa_{2,1} = K_{2,1}|_{u=0} = ET_1^2T_2 - (ET_2)(ET_1^2) - 2(ET_1)(ET_1T_2) - 2(ET_2)(ET_1)^2.$$

□

- b) Assume  $ET_1 = ET_2 = 0$  and give an expression for  $\kappa_{2,2}$  in terms of moments of  $T$ .

**Solution.**

$\kappa_{2,2}$  can be obtain by one more derivative for  $K_{21}$ . (Too complicated. See Keener page 462.)

□

## Problem 2.28

Suppose  $X \sim \Gamma(\alpha, 1/\lambda)$ , with density

$$\frac{\lambda^\alpha x^{\alpha-1} e^{-\lambda x}}{\Gamma(\alpha)}, \quad x > 0.$$

Find the cumulants of  $T = (X, \log X)$  of order 3 or less. The answer will involve  $\psi(\alpha) = d \log \Gamma(\alpha) / d\alpha = \Gamma'(\alpha) / \Gamma(\alpha)$ .

**Solution.**

Since Gamma distribution is the exponential family,

$$\begin{aligned} p(x) &= \exp [\alpha \log \lambda + (\alpha - 1) \log x - \lambda x - \log \Gamma(\alpha)] \\ &= \exp [-\lambda x + \alpha \log x - (\log \Gamma(\alpha) - \alpha \log \lambda) - \log x]. \end{aligned}$$

Let  $\eta = (-\lambda, \alpha)$ , and  $A(\eta) = \log \Gamma(\eta_2) - \eta_2 \log(-\eta_1)$ .

Then, the cumulants are obtained by taking derivatives:

$$\begin{aligned} \kappa_{1,0} &= \frac{\partial A(\eta)}{\partial \eta_1} = -\frac{\eta_2}{\eta_1} = \frac{\alpha}{\lambda}, \\ \kappa_{0,1} &= \frac{\partial A(\eta)}{\partial \eta_2} = \psi(\eta_2) - \log(-\eta_1) = \psi(\alpha) - \log \lambda, \end{aligned}$$

$$\kappa_{2,0} = \frac{\partial^2 A(\eta)}{\partial \eta_1^2} = \frac{\eta_2}{\eta_1^2} = \frac{\alpha}{\lambda^2},$$

$$\kappa_{1,1} = \frac{\partial^2 A(\eta)}{\partial \eta_1 \partial \eta_2} = -\frac{1}{\eta_1} = \frac{1}{\lambda},$$

$$\kappa_{0,2} = \frac{\partial^2 A(\eta)}{\partial \eta_2^2} = \psi'(\eta_2) = \psi'(\alpha),$$

$$\kappa_{3,0} = \frac{\partial^3 A(\eta)}{\partial \eta_1^3} = -\frac{2\eta_2}{\eta_1^3} = \frac{2\alpha}{\lambda^3},$$

$$\kappa_{2,1} = \frac{\partial^3 A(\eta)}{\partial \eta_1^2 \partial \eta_2} = \frac{1}{\eta_1^2} = \frac{1}{\lambda^2},$$

$$\kappa_{1,2} = \frac{\partial^3 A(\eta)}{\partial \eta_1 \partial \eta_2^2} = 0,$$

$$\kappa_{0,3} = \frac{\partial^3 A(\eta)}{\partial \eta_2^3} = \psi''(\eta_2) = \psi''(\alpha).$$

□

## Chapter 3

# Risk, Sufficiency, Completeness, and Ancillarity

### Problem 3.2

Suppose data  $X_1, \dots, X_n$  are independent with

$$P_\theta(X_i \leq x) = x^{t_i \theta}, \quad x \in (0, 1).$$

where  $\theta > 0$  is the unknown parameter, and  $t_1, \dots, t_n$  are known positive constants. Find a one-dimensional sufficient statistic  $T$ .

**Solution.**

Take a first derivative for the above cdf, then the pdf of  $X$  be  $p_\theta(x) = t_i \theta x_i^{t_i \theta - 1}$ . Then, the joint density is

$$p_\theta(x_1, \dots, x_n) = \prod_{i=1}^n t_i \theta x_i^{t_i \theta - 1} = \theta^n \left( \prod_{i=1}^n x_i^{t_i} \right)^\theta \frac{\prod_{i=1}^n t_i}{\prod_{i=1}^n x_i}.$$

By factorization theorem, the sufficient statistic for  $\theta$  is

$$T = \prod_{i=1}^n x_i^{t_i}.$$

□

### Problem 3.3

An object with weight  $\theta$  is weighted on scales with different precision. The data  $X_1, \dots, X_n$  are independent, with  $X_i \sim N(\theta, \sigma_i^2)$ ,  $i = 1, \dots, n$ , with the standard deviations  $\sigma_1, \dots, \sigma_n$  known constants. Use sufficiency to suggest a weighted average of  $X_1, \dots, X_n$  to estimate  $\theta$ . (A weighted average would have form  $\sum_{i=1}^n w_i X_i$ , where the  $w_i$  are positive and sum to one.)

**Solution.**

The pdf of  $X_i$  is

$$p_\mu(x_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2\sigma_i^2}(x_i - \theta)^2} = \frac{1}{\sqrt{2\pi}\sigma_i} \exp \left[ \frac{\theta x_i}{\sigma_i^2} - \frac{x_i^2}{2\sigma_i^2} - \frac{\theta^2}{2\sigma_i^2} \right].$$

Then, the joint density is

$$p_{\mu}(x_1, \dots, x_n) = \frac{1}{(2\pi)^{n/2} \prod_i \sigma_i} \exp \left[ \sum_i \frac{\theta x_i}{\sigma_i^2} - \sum_i \frac{x_i^2}{2\sigma_i^2} - \sum_i \frac{\theta^2}{2\sigma_i^2} \right].$$

By factorization theorem,

$$T = \sum_{i=1}^n \frac{x_i}{\sigma_i^2}.$$

To estimate  $\theta$  using a weighted average form, let

$$w_i = \frac{T}{\sum_{i=1}^n \sigma_i^{-2}},$$

then the weighted average of  $X_1, \dots, X_n$  be an estimator for  $\theta$  satisfied the sufficiency.  $\square$

### Problem 3.4

Let  $X_1, \dots, X_n$  be a random sample from an arbitrary discrete distribution  $P$  on  $\{1, 2, 3\}$ . Find a two-dimensional sufficient statistic.

**Solution.**

Let  $p_i = P(\{i\})$  and  $n_i(x) = \sum_{j=1}^n 1_{\{x_j=i\}}(x)$  for  $i = 1, 2, 3$ . Then,  $n_1(x) + n_2(x) + n_3(x) = n$ . Since  $X_i$ s are i.i.d., the joint pdf is

$$\begin{aligned} P(X_1 = x_1, \dots, X_n = x_n) &= p_1^{n_1(x)} p_2^{n_2(x)} p_3^{n_3(x)} \\ &= p_1^{n_1(x)} p_2^{n_2(x)} p_3^{n - n_1(x) - n_2(x)}. \end{aligned}$$

By factorization theorem, the sufficient statistic  $T = (n_1, n_2)$ .  $\square$

### Problem 3.6

The beta distribution with parameters  $\alpha > 0$  and  $\beta > 0$  has density

$$f_{\alpha, \beta}(x) = \begin{cases} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1}, & x \in (0, 1), \\ 0, & \text{otherwise.} \end{cases}$$

Suppose  $X_1, \dots, X_n$  are i.i.d. from a beta distribution.

- a) Determine a minimal sufficient statistic (for the family of joint distributions) if  $\alpha$  and  $\beta$  vary freely.

**Solution.**

The joint density

$$\begin{aligned} f_{\alpha, \beta}(x_1, \dots, x_n) &= \exp \left[ \sum_i (\alpha - 1) \log x_i + \sum_i (\beta - 1) \log(1 - x_i) + n \log \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} \right] \\ &= \exp \left[ \sum_i \alpha \log x_i + \sum_i \beta \log(1 - x_i) + n \log \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} - \sum_i \log x_i (1 - x_i) \right], \end{aligned}$$



is the exponential family and full-rank. Then,

$$T = \left( \sum_{i=1}^n \log X_i, \sum_{i=1}^n \log(1 - X_i) \right)$$

is complete.

By factorization theorem,  $T$  is also sufficient.

Therefore,  $T$  is the minimal sufficient statistic for  $(\alpha, \beta)$ . (See Theorem 3.19)  $\square$

- b) Determine a minimal sufficient statistic if  $\alpha = 2\beta$ .

**Solution.**

Plug into the joint density,

$$\begin{aligned} f_{\beta}(x_1, \dots, x_n) &= \exp \left[ \sum_i 2\beta \log x_i + \sum_i \beta \log(1 - x_i) + n \log \frac{\Gamma(2\beta + \beta)}{\Gamma(2\beta)\Gamma(\beta)} - \sum_i \log x_i(1 - x_i) \right] \\ &= \exp \left[ \beta \sum_i (\log x_i^2 + \log(1 - x_i)) + n \log \frac{\Gamma(3\beta)}{\Gamma(2\beta)\Gamma(\beta)} - \sum_i \log x_i(1 - x_i) \right] \end{aligned}$$

is the one-parameter exponential family and full-rank. Thus,

$$T = \sum_i (\log x_i^2 + \log(1 - x_i)) = 2T_1 + T_2$$

is complete, and also sufficient by factorization theorem.

Therefore,  $T$  is the minimal sufficient statistic for  $\beta$ .  $\square$

- c) Determine a minimal sufficient statistic if  $\alpha = \beta^2$ .

**Solution.**

Now, the joint density is

$$\begin{aligned} p_{\beta}(x_1, \dots, x_n) &= \exp \left[ \sum_i (\beta^2 - 1) \log x_i + \sum_i (\beta - 1) \log(1 - x_i) + n \log \frac{\Gamma(2\beta + \beta)}{\Gamma(2\beta)\Gamma(\beta)} \right] \\ &= \exp \left[ (\beta^2 - 1)T_1(x) + (\beta - 1)T_2(x) + n \log \frac{\Gamma(\beta^2 + \beta)}{\Gamma(\beta^2)\Gamma(\beta)} \right]. \end{aligned}$$

Suppose  $p_{\theta}(x) \propto_{\theta} p_{\theta}(y)$ . Then W.L.O.G. we consider 2 cases as follows:

$$\begin{aligned} \frac{p_1(y)}{p_1(x)} = \frac{p_2(y)}{p_2(x)} &\implies \frac{p_2(x)}{p_1(x)} = \frac{p_2(y)}{p_1(y)}, \\ \frac{p_1(y)}{p_1(x)} = \frac{p_3(y)}{p_3(x)} &\implies \frac{p_3(x)}{p_1(x)} = \frac{p_3(y)}{p_1(y)}. \end{aligned}$$

Since  $p_1(x) = p_1(y) = 1$ , from the above equation,

$$\begin{aligned} 3T_1(x) + T_2(x) + n \log \frac{\Gamma(2^2 + 2)}{\Gamma(2^2)\Gamma(2)} &= 3T_1(y) + T_2(y) + n \log \frac{\Gamma(2^2 + 2)}{\Gamma(2^2)\Gamma(2)}, \\ 8T_1(x) + 2T_2(x) + n \log \frac{\Gamma(3^2 + 3)}{\Gamma(3^2)\Gamma(3)} &= 8T_1(y) + 2T_2(y) + n \log \frac{\Gamma(3^2 + 3)}{\Gamma(3^2)\Gamma(3)}. \end{aligned}$$

These 2 equations imply  $T(x) = T(y)$ .

Therefore,  $T$  is the minimal sufficient statistic by Theorem 3.11.  $\square$

### Problem 3.7

*Logistic regression.* Let  $X_1, \dots, X_n$  be independent Bernoulli variables, with  $p_i = P(X_i = 1)$ ,  $i = 1, \dots, n$ . Let  $t_1, \dots, t_n$  be a sequence of known constants that are related to the  $p_i$  via

$$\log \frac{p_i}{1 - p_i} = \alpha + \beta t_i,$$

where  $\alpha$  and  $\beta$  are unknown parameters. Determine a minimal sufficient statistic for the family of joint distributions.

**Solution.**

From the above equation,

$$p_i = \frac{e^{\alpha + \beta t_i}}{1 + e^{\alpha + \beta t_i}}.$$

Then, the density of  $X_i$  is

$$\begin{aligned} p_{\alpha, \beta}(x_i) &= p_i^{x_i} (1 - p_i)^{1 - x_i} \\ &= \left( \frac{e^{\alpha + \beta t_i}}{1 + e^{\alpha + \beta t_i}} \right)^{x_i} \left( \frac{1}{1 + e^{\alpha + \beta t_i}} \right)^{1 - x_i} \\ &= \exp [x_i(\alpha + \beta t_i) - x_i \log(1 + e^{\alpha + \beta t_i}) - (1 - x_i) \log(1 + e^{\alpha + \beta t_i})] \\ &= \exp [\alpha x_i + \beta x_i t_i - \log(1 + e^{\alpha + \beta t_i})]. \end{aligned}$$

Therefore, we can easily obtain the following sufficient statistic by factorization theorem for the joint density,

$$T = \left( \sum_{i=1}^n X_i, \sum_{i=1}^n X_i t_i \right).$$

And also, the joint density is exponential family and full-rank. Thus  $T$  is complete.

$\therefore T$  is minimal sufficient. □

### Problem 3.8

The multinomial distribution, derived later in Section 5.3, is a discrete distribution with mass function

$$\frac{n!}{x_1! \times \dots \times x_s!} p_1^{x_1} \times \dots \times p_s^{x_s},$$

where  $x_1, \dots, x_s$  are non-negative integers summing to  $n$ , where  $p_1, \dots, p_s$  are non-negative probabilities summing to one, and  $n$  is the sample size. Let  $N_{11}, N_{21}, N_{22}$  have a multinomial distribution with  $n$  trials and success probabilities  $p_{11}, p_{12}, p_{21}, p_{22}$ . (A common model for a two-by-two contingency table.)

- a) Give a minimal sufficient statistic if the success probabilities vary freely over the unit simplex in  $\mathbb{R}^4$ . (The unit simplex in  $\mathbb{R}^p$  is the set of all vectors with non-negative entries summing to one.)

**Solution.**

The density be the exponential family as follows:

$$\begin{aligned} p(\mathbf{N}) &= \frac{n!}{N_{11}! N_{12}! N_{21}! N_{22}!} p_{11}^{N_{11}} p_{12}^{N_{12}} p_{21}^{N_{21}} p_{22}^{N_{22}} \\ &= \frac{n!}{N_{11}! N_{12}! N_{21}! N_{22}!} \exp [N_{11} \log p_{11} + N_{12} \log p_{12} + N_{21} \log p_{21} + N_{22} \log p_{22}] \end{aligned} \quad (*)$$

$$\begin{aligned}
&= \frac{n!}{N_{11}!N_{12}!N_{21}!N_{22}!} \exp [N_{11} \log p_{11} + N_{12} \log p_{12} + N_{21} \log p_{21} \\
&\quad + (n - N_{11} - N_{12} - N_{21}) \log(1 - p_{11} - p_{12} - p_{21})] \\
&= \frac{n!}{N_{11}!N_{12}!N_{21}!N_{22}!} \exp \left[ N_{11} \log \left( \frac{p_{11}}{1 - p_{11} - p_{12} - p_{21}} \right) + N_{12} \log \left( \frac{p_{12}}{1 - p_{11} - p_{12} - p_{21}} \right) \right. \\
&\quad \left. + N_{21} \log \left( \frac{p_{21}}{1 - p_{11} - p_{12} - p_{21}} \right) + n \log(1 - p_{11} - p_{12} - p_{21}) \right].
\end{aligned}$$

Since it is the exponential family of full-rank,  $T = (N_{11}, N_{12}, N_{21})$  is complete, and also sufficient by factorization theorem. Thus  $T = (N_{11}, N_{12}, N_{21})$  is minimal sufficient.

If we use the equation (\*), the minimal sufficient statistic is  $T = (N_{11}, N_{12}, N_{21}, N_{22})$ .  $\square$

- b) Give a minimal sufficient statistic if the success probabilities are constrained so that  $p_{11}p_{22} = p_{12}p_{21}$ .

**Solution.**

확실하지 않음...

From the constraint, plug in  $p_{11} = (p_{12}p_{21})/p_{22}$  and denote  $C$  is the term which is not depend on parameters,

$$\begin{aligned}
p(\mathbf{N}) &= Cp_{11}^{N_{11}} p_{12}^{N_{12}} p_{21}^{N_{21}} p_{22}^{N_{22}} \\
&= C \left( \frac{p_{12}p_{21}}{p_{22}} \right)^{N_{11}} p_{12}^{N_{12}} p_{21}^{N_{21}} p_{22}^{N_{22}} \\
&= Cp_{12}^{N_{11}+N_{12}} p_{21}^{N_{11}+N_{21}} p_{22}^{N_{22}-N_{11}} \\
&= Cp_{12}^{N_{11}+N_{12}} p_{21}^{N_{11}+N_{21}} (1 - p_{11} - p_{12} - p_{21})^{n - N_{12} - N_{21} - 2N_{11}} \\
&= C \exp \left[ (N_{11} + N_{12}) \log \left( \frac{p_{12}}{1 - p_{11} - p_{12} - p_{21}} \right) \right. \\
&\quad \left. + (N_{11} + N_{21}) \log \left( \frac{p_{21}}{1 - p_{11} - p_{12} - p_{21}} \right) + n \log(1 - p_{11} - p_{12} - p_{21}) \right].
\end{aligned}$$

Since it is the exponential family of full-rank,  $T = (N_{11} + N_{12}, N_{11} + N_{21})$  is complete, and also sufficient by factorization theorem. Thus  $T = (N_{11} + N_{12}, N_{11} + N_{21})$  is minimal sufficient.  $\square$

## Problem 3.9

Let  $f$  be a positive integrable function on  $(0, \infty)$ . Define

$$c(\theta) = 1 / \int_{\theta}^{\infty} f(x) dx,$$

and take  $p_{\theta}(x) = c(\theta)f(x)$  for  $x > \theta$ , and  $p_{\theta}(x) = 0$  for  $x \leq \theta$ . Let  $X_1, \dots, X_n$  be i.i.d. with common density  $p_{\theta}$ .

- a) Show that  $M = \min\{X_1, \dots, X_n\}$  is sufficient.

**Solution.**

The joint density is

$$p_{\theta}(x_1, \dots, x_n) = c(\theta)^n \prod_i f(x_i)$$

$$= c(\theta)^n \prod_i f(x_i) \cdot 1_{\min\{x_1, \dots, x_n\} > \theta}.$$

By factorization theorem,  $M = \min\{X_1, \dots, X_n\}$  is sufficient.  $\square$

b) Show that  $M$  is minimal sufficient.

**Solution.**

Suppose  $p_\theta(x) \propto_\theta p_\theta(y)$ . Then, the region of 0 value should be equal. Thus  $T(x) = T(y)$  should be satisfied, therefore  $T = M$  is minimal sufficient.  $\square$

### Problem 3.10

Suppose  $X_1, \dots, X_n$  are i.i.d. with common density  $f_\theta(x) = (1 + \theta x)/2$ ,  $|x| < 1$ ;  $f_\theta(x) = 0$ , otherwise, where  $\theta \in [-1, 1]$  is an unknown parameter. Show that the order statistics are minimal sufficient. (Hint: A polynomial of degree  $n$  is uniquely determined by its value on a grid of  $n + 1$  points.)

**Solution.**

The joint density is

$$\begin{aligned} f_\theta(x_1, \dots, x_n) &= 2^{-n} (1 + \theta x_1) \cdots (1 + \theta x_n) \\ &= 2^{-n} (1 + \theta x_{(1)}) \cdots (1 + \theta x_{(n)}) 1_{\{-1 < x_{(1)} < \cdots < x_{(n)} < 1\}}. \end{aligned}$$

By factorization theorem, the order statistic  $T = (X_{(1)}, \dots, X_{(n)})$  is sufficient.

Now, suppose  $p_\theta(x) \propto_\theta p_\theta(y)$ . Since the joint density is the  $n$ -th order polynomials (다항식) with root( $\sqrt[n]{\cdot}$ )  $-1/x_{(i)}$ , the roots of  $p_\theta(x)$  and  $p_\theta(y)$  should be equal which implies  $T(x) = T(y)$ . Therefore,  $T = (X_{(1)}, \dots, X_{(n)})$  is minimal sufficient.  $\square$

### Problem 3.16

Let  $X_1, \dots, X_n$  be a random sample from an absolutely continuous distribution with density

$$f_\theta(x) = \begin{cases} 2x/\theta^2, & x \in (0, \theta), \\ 0, & \text{otherwise.} \end{cases}$$

a) Find a one-dimensional sufficient statistic  $T$ .

**Solution.**

The joint density is factorized as follows:

$$f_\theta(x_1, \dots, x_n) = 2^n \frac{x_1 \cdots x_n}{\theta^{2n}} = \frac{2^n}{\theta^{2n}} x_{(1)} \cdots x_{(n)} 1_{x_{(1)} > 0} 1_{x_{(n)} < \theta}.$$

By factorization theorem,  $T = \max\{X_1, \dots, X_n\}$  is the sufficient statistic for  $\theta$ .  $\square$

b) Determine the density of  $T$ .

**Solution.**

Since  $T = \max\{X_1, \dots, X_n\}$ ,

$$P(T \leq t) = \prod_{i=1}^n P(X_i \leq t)$$

$$\begin{aligned} \text{Note that } P(X_i \leq t) &= \int_0^t 2x/\theta^2 dx = t^2/\theta^2. \\ &= \frac{t^{2n}}{\theta^{2n}} \end{aligned}$$

Taking a derivative with respect to  $t$ , the density of  $T$  is

$$p_\theta(t) = \frac{2n}{\theta^{2n}} t^{2n-1}, \quad t \in (0, \theta).$$

□

c) Show directly that  $T$  is complete.

**Solution.**

Suppose  $E_\theta(T) = c$ ,  $\forall \theta > 0$ . Then,

$$\begin{aligned} E_\theta [f(T) - c] &= \frac{2n}{\theta^{2n}} \int_0^\theta [f(t) - c] t^{2n-1} dt = 0 \\ \Rightarrow [f(t) - c] t^{2n-1} &= 0 \quad \text{a.e. } 0 < t < \theta. \\ \Rightarrow f(T) &= c. \end{aligned}$$

Therefore,  $T$  is complete.

Other solution

$$\begin{aligned} E_\theta f(T) &= \frac{2n}{\theta^{2n}} \int_0^\theta f(t) t^{2n-1} dt = c \\ \Rightarrow \int_0^\theta f(t) t^{2n-1} dt &= c \theta^{2n} / 2n, \quad \forall \theta > 0 \\ \Rightarrow f(\theta) \theta^{2n-1} &= c \theta^{2n-1} \quad \text{a.e. } \theta \\ \Rightarrow f(t) &= c \quad \text{a.e. } t. \end{aligned}$$

□

**Problem 3.17**

Let  $X, X_1, X_2, \dots$  be i.i.d. from an exponential distribution with failure rate  $\lambda$  (introduced in Problem 1.30).

a) Find the density of  $Y = \lambda X$ .

**Solution.**

Note that the density of  $X$  is  $p_\lambda(x) = \lambda e^{-\lambda x}$ ,  $\lambda > 0$ , and its cdf is easily obtained by integration,  $P(X \leq x) = 1 - e^{-\lambda x}$ . Thus the cdf of  $Y$  is

$$P(Y \leq y) = P(X \leq y/\lambda) = 1 - e^{-y}, \quad y > 0,$$

and the density is

$$p(y) = e^{-y}, \quad y > 0.$$

□

b) Let  $\bar{X} = (X_1 + \cdots + X_n)/n$ . Show that  $\bar{X}$  and  $(X_1^2 + \cdots + X_n^2)/\bar{X}^2$  are independent.

**Solution.**

The joint density of  $X_1, \dots, X_n$  is

$$p_\lambda(x_1, \dots, x_n) = \lambda^n e^{-\lambda \sum_i x_i} = \lambda^n e^{-n\lambda \bar{x}},$$

is the full rank exponential family, and by factorization theorem,  $T = \bar{X}$  is the complete sufficient statistic.

Now, let  $Y = \lambda X$ , then

$$V = \frac{(Y_1^2 + \cdots + Y_n^2)}{\bar{Y}^2} = \frac{(X_1^2 + \cdots + X_n^2)}{\bar{X}^2}.$$

Since  $Y$  does not depend on parameter  $\lambda$ ,  $V$  is ancillary for  $P_\lambda$ .

$\therefore$  By Basu's theorem,  $T$  and  $V$  are independent.

□

### Problem 3.29

Find a function on  $(0, \infty)$  that is bounded and strictly convex.

**Solution.**

$f(x) = e^{-x}$  is bounded on range  $(0, 1)$  and strictly convex.

Or  $f(x) = 1/(x+1)$  is bounded on range  $(0, 1)$  and strictly convex.

□

### Problem 3.30

Use convexity to show that the canonical parameter space  $\Xi$  of a one-parameter exponential family must be an interval. Specifically, show that if  $\eta_0 < \eta < \eta_1$ , and if  $\eta_0$  and  $\eta_1$  both lie in  $\Xi$ , then  $\eta$  must lie in  $\Xi$ .

**Solution.**

Note that if a random variable  $X$  is a canonical exponential family, then it has the form,

$$p_\eta(x) = \exp[\eta T(x) - A(\eta)]h(x), \quad \eta \in \Xi.$$

Since  $\eta_0 < \eta < \eta_1$ , we define  $\eta = \gamma\eta_0 + (1-\gamma)\eta_1$  for some  $\gamma \in (0, 1)$ . Since the exponential function is convex,

$$\exp[(\gamma\eta_0 + (1-\gamma)\eta_1)T(x)] = e^{\eta T(x)} < \gamma e^{\eta_0 T(x)} + (1-\gamma)e^{\eta_1 T(x)}.$$

From the definition of exponential family,  $h(x) \geq 0$ , and for the positive function, the inequality sign does not change by integration against  $\mu$ . Thus, following inequality is satisfied:

$$\int e^{\eta T(x)} h(x) d\mu(x) < \gamma \int e^{\eta_0 T(x)} h(x) d\mu(x) + (1-\gamma) \int e^{\eta_1 T(x)} h(x) d\mu(x).$$

Because it should be satisfied the exponential family, RHS is finite. ( $\because$  RHS =  $\gamma e^{A(\eta_0)} + (1-\gamma)e^{A(\eta_1)} < \infty$ .)

Therefore,  $\eta$  must lie in  $\Xi$ .

□

**Problem 3.31**

Let  $f$  and  $g$  be positive probability densities on  $\mathbb{R}$ . Use Jensen's inequality to show that

$$\int \log \left( \frac{f(x)}{g(x)} \right) f(x) dx > 0,$$

unless  $f = g$  a.e. (If  $f = g$ , the integral equals zero.) This integral is called the *Kullback-Liebler information*.

**Solution.**

Suppose  $X$  be a absolutely continuous random variable with density  $f$ . Now, define  $Y = \frac{g(X)}{f(X)}$ , then

$$EY = \int \frac{g(x)}{f(x)} f(x) dx = \int g(x) dx = 1.$$

Next, let  $h(y) = \log \frac{1}{y} = -\log y$ , then  $h$  is strictly convex on  $(0, \infty)$ . Then, by Jensen's inequality,

$$Eh(Y) = \int \log \left( \frac{f(x)}{g(x)} \right) f(x) dx \geq h(EY) = \log 1 = 0.$$

The equality holds when  $Y$  is a constant. (then  $Y = EY = 1 \Rightarrow f(x) = g(x)$  a.e.)

□

## Chapter 4

# Unbiased Estimation

### Problem 4.1

Let  $X_1, \dots, X_m$  and  $Y_1, \dots, Y_n$  be independent variables with the  $X_i$  a random sample from an exponential distribution with failure rate  $\lambda_x$ , and the  $Y_j$  a random sample from an exponential distribution with failure rate  $\lambda_y$ .

- a) Determine the UMVU estimator of  $\lambda_x/\lambda_y$ .

**Solution.**

Note that if  $X \sim \text{Exp}(\lambda)$  with rate parameter  $\lambda$ , the density  $p_\lambda(x) = \lambda e^{-\lambda x}$ .

Since  $X_i, \forall i$  and  $Y_j, \forall j$  are independent, the joint density is

$$p_{\lambda_x, \lambda_y}(x, y) = \lambda_x^n \lambda_y^n e^{-\lambda_x \sum_{i=1}^m x_i - \lambda_y \sum_{i=1}^n y_i},$$

which is the 2-parameter full-rank exponential family. Thus, by factorization theorem,  $T = (\sum_{i=1}^m X_i, \sum_{i=1}^n Y_i)$  is the complete sufficient.

Since  $X_1, \dots, X_m$  are i.i.d.  $\text{Exp}(\lambda_x) = \Gamma(1, \lambda_x)$ ,  $T_1 = \sum_{i=1}^m X_i \sim \Gamma(m, \lambda_x)$ . Then,

$$E \frac{1}{T_1} = \int \frac{1}{t} \frac{\lambda_x^m}{\Gamma(m)} t^{m-1} e^{-\lambda_x t} dt = \frac{\Gamma(m-1) \lambda_x^m}{\Gamma(m) \lambda_x^{m-1}} = \frac{\lambda_x}{m-1}.$$

Similarly, we can obtain  $ET_2 = n/\lambda_y$ . ( $\because EY_i = 1/\lambda_y$ )

From the fact of the expectation of independent random variables,

$$E \frac{T_2}{T_1} = E \frac{1}{T_1} \cdot ET_2 = \frac{n \lambda_x}{(m-1) \lambda_y}.$$

Therefore,

$$\delta(x, y) = \frac{(m-1)T_2}{nT_1}$$

is the unbiased estimator of  $\lambda_x/\lambda_y$ , and also UMVU since it is the function of complete sufficient statistic  $T$ . (By Theorem 4.4 with Rao-Blackwellization)  $\square$

- b) Under squared error loss, find the best estimator of  $\lambda_x/\lambda_y$  of form  $\delta = c\bar{Y}/\bar{X}$ .



**Solution.**

Let  $\delta = c\bar{Y}/\bar{X} = dT_2/T_1$ . Then, the risk of  $\delta = c\bar{Y}/\bar{X}$  and  $\lambda_x/\lambda_y$  is

$$R\left(\frac{\lambda_y}{\lambda_x}, \delta\right) = E\left[d\frac{T_2}{T_1} - \frac{\lambda_x}{\lambda_y}\right]^2 = d^2 E\frac{T_2^2}{T_1^2} - 2\frac{\lambda_x}{\lambda_y} dE\frac{T_2}{T_1} + \frac{\lambda_x^2}{\lambda_y^2}. \quad (4.1)$$

Since  $T_1 \sim \Gamma(m, \lambda_x)$  and  $T_2 \sim \Gamma(n, \lambda_y)$ ,

$$ET_2^2 = \text{Var}(T_2) + \{ET_2\}^2 = n/\lambda_y^2 + n^2/\lambda_y^2,$$

$$E\frac{1}{T_1^2} = \int \frac{1}{t^2} \frac{\lambda_x^m}{\Gamma(m)} t^{m-1} e^{-\lambda_x t} dt = \frac{\Gamma(m-2)\lambda_x^2}{\Gamma(m)} = \frac{\lambda_x^2}{(m-1)(m-2)}.$$

Thus, using the fact that  $T_1$  and  $T_2$  are independent,

$$E\frac{T_2^2}{T_1^2} = E\frac{1}{T_1^2} \cdot ET_2^2 = \frac{n(n+1)\lambda_x^2}{(m-1)(m-2)\lambda_y^2}.$$

By using the result from a), we know  $ET_2/T_1 = (n\lambda_x)/((m-1)\lambda_y)$ , then the risk (4.1) be

$$\begin{aligned} R\left(\frac{\lambda_y}{\lambda_x}, \delta\right) &= d^2 E\frac{T_2^2}{T_1^2} - 2\frac{\lambda_x}{\lambda_y} dE\frac{T_2}{T_1} + \frac{\lambda_x^2}{\lambda_y^2} \\ &= d^2 \frac{n(n+1)\lambda_x^2}{(m-1)(m-2)\lambda_y^2} - 2\frac{\lambda_x}{\lambda_y} d \frac{n\lambda_x}{(m-1)\lambda_y} + \frac{\lambda_x^2}{\lambda_y^2} \\ &= \frac{\lambda_x^2}{\lambda_y^2} \left( \frac{n(n+1)}{(m-1)(m-2)} d^2 - 2\frac{n}{(m-1)} d + 1 \right) \\ &= \frac{\lambda_x^2}{\lambda_y^2} \frac{n(n+1)}{(m-1)(m-2)} \left[ \left( d - \frac{m-2}{n+1} \right)^2 - \left( \frac{m-2}{n+1} \right)^2 + 1 \right]. \end{aligned}$$

Thus, it has the minimum when  $d = \frac{m-1}{n+1}$ .

Therefore, the best estimator of  $\lambda_x/\lambda_y$  is  $c\bar{Y}/\bar{X} = dT_2/T_1 = [(m-2)n\bar{Y}] / [(n+1)m\bar{X}]$ .

$$\therefore c = \frac{(m-2)n}{(n+1)m}.$$

□

c) Find the UMVU estimator of  $e^{-\lambda_x} = P(X_1 > 1)$ .

**Solution.**

Let  $\delta$  is the unbiased estimator of  $P(X_1 > 1)$ . Then,

$$E\delta = P(X_1 > 1) \Rightarrow \delta = 1_{\{X_1 > 1\}}.$$

From the previous steps, we know  $T_1$  is C.S.S., thus  $\eta(T_1) = E(\delta|T_1) = P(X_1 > 1|T_1)$  is UMVU by Theorem 4.4. We follow by 2 steps.

(i) Find the joint density of  $X$  and  $T_1$ , where  $X = 1_{\{X_1 > 1\}}$  and  $T_1 = \sum_{i=1}^m X_i$ .

First, we should find the joint density of  $X$  and  $S$  where  $S = \sum_{i=2}^m X_i$ . Since  $X$  and  $S$  are independent, the joint density is easily obtained as

$$f(x, s) = f(x)f(s) = \lambda_x e^{-\lambda_x x} \frac{\lambda_x^{m-1}}{\Gamma(m-1)} s^{m-2} e^{-\lambda_x s} = \frac{\lambda_x^m}{\Gamma(m-1)} s^{m-2} e^{-\lambda_x(x+2)}, \quad x > 0, s > 0.$$

By transformation of variables  $X = X, T_1 = S + X$ , the jacobian  $|J| = \begin{vmatrix} 1 & 0 \\ -1 & 1 \end{vmatrix} = 1$ . Then, the joint density of  $X$  and  $T_1$  is

$$f(x, t) = \frac{\lambda_x^m}{\Gamma(m-2)} (t-x)^{m-2} e^{-\lambda_x t}, \quad 0 < x < t.$$

(ii) Derive  $E(\delta|T_1)$ .

To obtain the expectation, we derive the conditional density of  $X$  given  $T_1$ . Since  $T_1 \sim \Gamma(m, \lambda_x)$ , the conditional density be

$$f(x|t) = \frac{f(x, t)}{f(t)} = \frac{\Gamma(m)}{\Gamma(m-2)} (t-x)^{m-2} t^{-(m-1)} = (m-1) \frac{1}{t} \left(1 - \frac{x}{t}\right)^{m-2}, \quad 0 < x < t.$$

Then,

$$\begin{aligned} E(\delta|T_1) &= P(X_1 > 1|T_1) = P(X|T_1) \\ &= \int_1^t f(x|t) dx \\ &= \int_1^t (m-1) \frac{1}{t} \left(1 - \frac{x}{t}\right)^{m-2} dx \\ &= -\left(1 - \frac{x}{t}\right)^{m-1} \Big|_1^t \\ &= \left(1 - \frac{1}{t}\right)^{m-1} 1_{\{T_1 > 1\}}. \end{aligned}$$

□

## Problem 4.2

Let  $X_1, \dots, X_n$  be a random sample from  $N(\mu_x, \sigma^2)$ , and let  $Y_1, \dots, Y_m$  be an independent random sample from  $N(\mu_y, 2\sigma^2)$ , with  $\mu_x$ ,  $\mu_y$ , and  $\sigma^2$  all unknown parameters..

a) Find a complete sufficient statistic.

**Solution.**

Since  $X_1, \dots, X_n, Y_1, \dots, Y_m$  are mutually independent, the joint density,

$$\begin{aligned} f(x, y) &= f(x)f(y) \\ &= (2\pi\sigma^2)^{-n/2} (4\pi\sigma^2)^{-m/2} \exp \left[ -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_x)^2 - \frac{1}{4\sigma^2} \sum_{j=1}^m (y_j - \mu_y)^2 \right] \\ &= \exp \left[ -\frac{\sum_{i=1}^n x_i^2}{2\sigma^2} - \frac{\sum_{j=1}^m y_j^2}{4\sigma^2} + \frac{\mu_x \sum_{i=1}^n x_i}{\sigma^2} + \frac{\mu_y \sum_{j=1}^m y_j}{2\sigma^2} - \frac{n\mu_x^2}{2\sigma^2} - \frac{m\mu_y^2}{4\sigma^2} \right] (2\pi\sigma^2)^{-n/2} (4\pi\sigma^2)^{-m/2} \\ &= \exp \left[ \frac{\mu_x \sum_{i=1}^n x_i}{\sigma^2} + \frac{\mu_y \sum_{j=1}^m y_j}{2\sigma^2} - \frac{2 \sum_{i=1}^n x_i^2 + \sum_{j=1}^m y_j^2}{4\sigma^2} - \frac{n\mu_x^2}{2\sigma^2} - \frac{m\mu_y^2}{4\sigma^2} \right] (2\pi\sigma^2)^{-n/2} (4\pi\sigma^2)^{-m/2}, \end{aligned}$$

is the full-rank exponential family. Therefore,

$$T = \left( \sum_{i=1}^n X_i, \sum_{j=1}^m Y_j, 2 \sum_{i=1}^n X_i^2 + \sum_{j=1}^m Y_j^2 \right)$$

is the complete sufficient statistic for  $(\mu_x, \mu_y, \sigma^2)$ .  $\square$

b) Determine the UMVU estimator of  $\sigma^2$ .

Hint: Find a linear combination  $L$  of  $S_x^2 = \sum_{i=1}^n (X_i - \bar{X})^2 / (n-1)$  and  $S_y^2 = \sum_{j=1}^m (Y_j - \bar{Y})^2 / (m-1)$  so that  $(\bar{X}, \bar{Y}, L)$  is complete sufficient.

**Solution.**

To use the C.S.S. obtained from a), let

$$2(n-1)S_x^2 + (m-1)S_y^2 = 2 \left( \sum_i X_i^2 - n\bar{X}^2 \right) + \left( \sum_j Y_j^2 - m\bar{Y}^2 \right)$$

is the linear combination of  $T$ . Since the sample variance is unbiased, then its expectation is

$$E[2(n-1)S_x^2 + (m-1)S_y^2] = (2n + 2m - 4)\sigma^2.$$

Now, we let

$$S_p^2 = \frac{1}{2n + 2m - 4} [2(n-1)S_x^2 + (m-1)S_y^2],$$

then it is unbiased for  $\sigma^2$  based on complete sufficient  $T$ .

By theorem 4.4, it is UMVU of  $\sigma^2$ .  $\square$

c) Find a UMVU estimator of  $(\mu_x - \mu_y)^2$ .

**Solution.**

$$\begin{aligned} E(\bar{X} - \bar{Y})^2 &= \text{Var}(\bar{X} - \bar{Y}) + [E(\bar{X} - \bar{Y})]^2 \\ &= \text{Var}(\bar{X}) + \text{Var}(\bar{Y}) + [E\bar{X} - E\bar{Y}]^2 \\ &= \frac{\sigma^2}{n} + \frac{2\sigma^2}{m} + (\mu_x - \mu_y)^2. \end{aligned}$$

We obtained the unbiased estimator of  $\sigma^2$  in b). Thus, we let  $\delta$  as

$$\delta = (\bar{X} - \bar{Y})^2 - S_p^2 \left( \frac{1}{n} - \frac{2}{m} \right),$$

then it is the unbiased estimator of  $(\mu_x - \mu_y)^2$  based on complete sufficient  $T$ .

By Theorem 4.4,  $\delta$  is UMVU of  $(\mu_x - \mu_y)^2$ .  $\square$

d) Suppose we know the  $\mu_y = 3\mu_x$ . What is the UMVU estimator of  $\mu_x$ ?

**Solution.**

With restriction  $\mu_y = 3\mu_x$ , the joint density of exponential family is changed as

$$f(x, y) = \cdots = \exp \left[ \frac{\mu_x (2 \sum_i x_i + 3 \sum_j y_j)}{2\sigma^2} - \cdots \right] \cdots$$

Thus,  $T = \left( 2 \sum_i X_i + 3 \sum_j Y_j, 2 \sum_{i=1}^n X_i^2 + \sum_{j=1}^m Y_j^2 \right)$  is complete sufficient.

Take expectation for  $T_1$ , then

$$ET_1 = 2n\mu_x + 3m\mu_y = (2n + 9m)\mu_x,$$

Now we let  $\delta$  as

$$\delta = \frac{2 \sum_i x_i + 3 \sum_j y_j}{2n + 9m},$$

then, it is unbiased estimator based on  $T$ .

By Theorem 4.4,  $\delta$  is UMVU of  $\mu_x$ . □

### Problem 4.3

Let  $X_1, \dots, X_n$  be a random sample from the Poisson distribution with mean  $\lambda$ . Find the UMVU for  $\cos \lambda$ . (Hint: For Taylor expansion, the identity  $\cos \lambda = (e^{i\lambda} + e^{-i\lambda})/2$  may be useful.)

#### Solution.

The joint density is

$$p_\lambda(x) = \frac{\lambda^{\sum_i x_i} e^{-n\lambda}}{x_1! \cdots x_n!},$$

with  $T = \sum_i X_i$ , that is complete sufficient. ( $\because$  full-rank exponential family, and by factorization theorem) Note that  $T \sim \text{Poisson}(n\lambda)$ . Let  $\delta(t)$  is an unbiased estimator of  $\cos \lambda$  based on  $T$ . Then, its expectation should satisfy follows:

$$\begin{aligned} E\delta(T) &= \sum_{t=0}^{\infty} \delta(t) \frac{(n\lambda)^t e^{-n\lambda}}{t!} = \cos \lambda \\ \Leftrightarrow \sum_{t=0}^{\infty} \delta(t) \frac{(n\lambda)^t}{t!} &= e^{n\lambda} \cos \lambda \end{aligned}$$

Apply Talyor expansion for RHS,

$$\begin{aligned} e^{n\lambda} \cos \lambda &= e^{n\lambda} \frac{1}{2} (e^{i\lambda} + e^{-i\lambda}) \\ &= \frac{1}{2} (e^{(n+i)\lambda} + e^{(n-i)\lambda}) \\ &= \frac{1}{2} \sum_{t=0}^{\infty} \frac{((n+i)\lambda)^t + ((n-i)\lambda)^t}{t!} \\ &= \sum_{t=0}^{\infty} \frac{(n\lambda)^t}{t!} \left[ \frac{(1 + \frac{i}{n})^t + (1 - \frac{i}{n})^t}{2} \right]. \end{aligned}$$

Therefore,

$$\delta(t) = \frac{1}{2} \left\{ \left(1 + \frac{i}{n}\right)^t + \left(1 - \frac{i}{n}\right)^t \right\}$$

is unbiased estimator of  $\cos \lambda$  based on  $T$ .

By Theorem 4.4,  $\delta(t)$  is UMVU of  $\cos \lambda$ . □

### Problem 4.4

Let  $X_1, \dots, X_n$  be independent normal variables, each with unit variance, and with  $EX_i = \alpha t_i + \beta t_i^2$ ,  $i = 1, \dots, n$ , where  $\alpha$  and  $\beta$  are unknown parameters and  $t_1, \dots, t_n$  are known constants. Find UMVU estimators of  $\alpha$  and  $\beta$ .

**Solution.**

Since  $X_i \sim N(\alpha t_i + \beta t_i^2, 1)$ , the joint density is

$$\begin{aligned} f(x) &= (2\pi)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2} \sum_i (x_i - \alpha t_i - \beta t_i^2)^2 \right] \\ &= (2\pi)^{-\frac{n}{2}} \exp \left[ -\frac{1}{2} \sum_i (x_i^2 + \alpha^2 t_i^2 + \beta^2 t_i^4 - 2\alpha x_i t_i - 2\beta x_i t_i^2 + \alpha\beta t_i^3) \right]. \end{aligned}$$

Since these density is full-rank exponential family,

$$T = \left( \sum_i X_i t_i, \sum_i X_i t_i^2 \right)$$

is complete sufficient and its expected value is

$$ET_1 = \alpha \sum t_i^2 + \beta \sum t_i^3, \quad ET_2 = \alpha \sum t_i^3 + \beta \sum t_i^4.$$

By using this,

$$\hat{\alpha} = \frac{T_1 \sum t_i^4 - T_2 \sum t_i^3}{\sum t_i^2 \sum t_i^4 - (\sum t_i^3)^2}, \quad \hat{\beta} = \frac{T_1 \sum t_i^3 - T_2 \sum t_i^2}{(\sum t_i^3)^2 - \sum t_i^4 \sum t_i^2}$$

are unbiased estimator of  $\alpha$  and  $\beta$ . Since these are the function of  $T$ , also UMVU.  $\square$

**Problem 4.5**

Let  $X_1, \dots, X_n$  be i.i.d. from some distribution  $Q_\theta$ , and let  $\bar{X} = (X_1 + \dots + X_n)/n$  be the sample average.

a) Show that  $S^2 = \sum (X_i - \bar{X})^2 / (n-1)$  is unbiased for  $\sigma^2 = \sigma^2(\theta) = \text{Var}_\theta(X_i)$ .

**Solution.**

Denote  $\mu = EX_i$ . Also note that  $(n-1)S^2 = \sum_i X_i^2 - n\bar{X}^2$ . Then,

$$\begin{aligned} E((n-1)S^2) &= \sum_i EX_i^2 - nE\bar{X}^2 \\ &= \sum_i [\text{Var}X_i + (EX_i)^2] - n[\text{Var}\bar{X} + (E\bar{X})^2] \\ &= n(\sigma^2 + \mu^2) - n(\sigma^2/n + \mu^2) \\ &= (n-1)\sigma^2. \end{aligned}$$

Therefore  $S^2$  is unbiased for  $\sigma^2$ .  $\square$

b) If  $Q_\theta$  is the Bernoulli distribution with success probability  $\theta$ , show that  $S^2$  from (a) is UMVU.

**Solution.**

Note that the density of  $X_i$  is  $\theta^{x_i}(1-\theta)^{1-x_i} = \left(\frac{\theta}{1-\theta}\right)^{x_i} (1-\theta)$ .

The joint density,

$$p(x) = \exp \left[ \sum_i x_i \log \frac{\theta}{1-\theta} + n \log(1-\theta) \right],$$

is the full-rank exponential family. Thus,  $T = \bar{X}$  (or  $\sum_i X_i$ ) is complete sufficient. Meanwhile,

$$\begin{aligned} S^2 &= \frac{1}{n-1} \left( \sum X_i^2 - n\bar{X}^2 \right) \\ &= \frac{1}{n-1} \left( \sum X_i - n\bar{X}^2 \right) \quad (\because X_i^2 = X_i) \\ &= \frac{n\bar{X}(1 - \bar{X})}{n-1} \end{aligned}$$

is the function of  $T = \bar{X}$ . Therefore,  $S^2$  is UMVU of  $\sigma^2$  by Theorem 4.4.  $\square$

- c) If  $Q_\theta$  is the exponential distribution with failure rate  $\theta$ , find the UMVU estimator of  $\sigma^2 = 1/\theta^2$ . Give a formula for  $E_\theta[X_i^2 | \bar{X} = c]$  in this case.

***Solution.***

Note that the density of  $X_i$  is  $\theta e^{-\theta x}$  with mean  $1/\theta$  and variance  $1/\theta^2$ .

The joint density,

$$p(x) = \exp \left[ -\theta \sum x_i + n \log \theta \right],$$

is the full-rank exponential family. Thus,  $T = \bar{X}$  (or  $\sum_i X_i$ ) is complete sufficient with  $ET = 1/\theta$  and  $\text{Var}T = 1/(n\theta^2)$ .

Meanwhile,

$$E\bar{X}^2 = \text{Var}\bar{X} + (E\bar{X})^2 = \frac{1}{n\theta^2} + \frac{1}{\theta^2} = \frac{n+1}{n} \frac{1}{\theta^2} = \frac{n+1}{n} \sigma^2.$$

Thus,  $\delta = \frac{n}{n+1} \bar{X}^2$  is the unbiased estimator of  $\sigma^2$ , and also function of  $T = \bar{X}$ .

By Theorem 4.4,  $\delta$  is UMVU for  $\sigma^2$ .

Next, since  $X_i$ 's are i.i.d.,  $E_\theta[X_1^2 | \bar{X} = c] = \cdots = E_\theta[X_n^2 | \bar{X} = c]$ . Then,

$$\begin{aligned} E_\theta[S^2 | \bar{X} = c] &= \frac{nc^2}{n+1} \\ &= E_\theta \left[ \frac{\sum_i X_i^2 - n\bar{X}^2}{n-1} \middle| \bar{X} = c \right] \\ &= E_\theta \left[ \frac{\sum_i X_i^2}{n-1} - \frac{nc^2}{n-1} \middle| \bar{X} = c \right] \\ &= \frac{n}{n-1} E_\theta[X_1^2 | \bar{X} = c] - \frac{nc^2}{n-1}. \end{aligned}$$

$$\therefore E_\theta[X_i^2 | \bar{X} = c] = nc^2 \left( \frac{1}{n+1} + \frac{1}{n-1} \right) \frac{n-1}{n} = c^2 \left( \frac{n-1}{n+1} + 1 \right) = \frac{2nc^2}{n+1}.$$

$\square$

## Problem 4.6

Suppose  $\delta$  is a UMVU estimator of  $g(\theta)$ ;  $U$  is an unbiased estimator of zero,  $E_\theta U = 0$ ,  $\theta \in \Omega$ ; and that  $\delta$  and  $U$  both have finite variances for all  $\theta \in \Omega$ . Show that  $U$  and  $\delta$  are uncorrelated,  $E_\theta U \delta = 0$ ,  $\theta \in \Omega$ .

**Solution.**

For the convenience, we drop  $\theta$  in the notation of  $E$ ,  $\text{Var}$  and  $\text{Cov}$ .

Let  $\delta^* = \delta + cU$  where  $c$  is a constant. Then,  $\delta^*$  is an unbiased estimator for  $g(\theta)$  ( $\because E\delta^* = E(\delta + cU) = E\delta + cEU = g(\theta)$ ).

Since  $\delta$  is UMVU,

$$\begin{aligned}\text{Var}(\delta + cU) &= \text{Var}(\delta) + c^2\text{Var}(U) + 2c\text{Cov}(\delta, U) \geq \text{Var}(\delta), \\ &\Rightarrow c^2\text{Var}(U) + 2c\text{Cov}(\delta, U) \geq 0.\end{aligned}$$

Denote LHS as  $h(c)$ . Since  $h$  is convex and has minimum  $h(0) = 0$ ,  $h$  should have a local minimum at  $c = 0$ . Since  $h$  is differentiable, we can obtain the following result by taking a first derivative at  $c = 0$ :

$$h'(0) = 2\text{Cov}(\delta, U) = 0.$$

Since  $EU = 0$ , we finally obtain  $E_\theta U \delta = 0$  for all  $\theta \in \Omega$ . □

**Problem 4.7**

Suppose  $\delta_1$  is a UMVU estimator of  $g_1(\theta)$ ,  $\delta_2$  is UMVU estimator of  $g_2(\theta)$ , and that  $\delta_1$  and  $\delta_2$  both have finite variance for all  $\theta$ . Show that  $\delta_1 + \delta_2$  is UMVU for  $g_1(\theta) + g_2(\theta)$ .

Hint: Use the result in the previous problem.

**Solution.**

We can easily show that  $\delta_1 + \delta_2$  is unbiased for  $g_1(\theta) + g_2(\theta)$ .

Now, suppose  $\delta$  is any unbiased estimator of  $g_1(\theta) + g_2(\theta)$  with  $\text{Var}(\delta) < \infty$ . Let  $U = \delta - \delta_1 - \delta_2$ , then  $U$  is an unbiased estimator of 0 ( $\because EU = E\delta - E\delta_1 - E\delta_2 = 0$ ).

By Problem 4.6,

$$\text{Cov}(U, \delta_1 + \delta_2) = \text{Cov}(U, \delta_1) + \text{Cov}(U, \delta_2) = 0.$$

Since  $\delta = U + \delta_1 + \delta_2$ , we obtain the following result:

$$\begin{aligned}\text{Var}(\delta) &= \text{Var}(U) + \text{Var}(\delta_1 + \delta_2) \\ &\geq \text{Var}(\delta_1 + \delta_2).\end{aligned}$$

Because  $\delta$  is arbitrary,  $\delta_1 + \delta_2$  is UMVU. □

**Problem 4.8**

Let  $X_1, \dots, X_n$  be i.i.d. absolutely continuous variables with common density  $f_\theta$ ,  $\theta > 0$ , given by

$$f_\theta(x) = \begin{cases} \theta/x^2, & x > \theta, \\ 0, & x \leq \theta. \end{cases}$$

Find the UMVU estimator for  $g(\theta)$  if  $g(\theta)/\theta^n \rightarrow 0$  as  $\theta \rightarrow \infty$  and  $g$  is differentiable.

**Solution.**

The joint density is

$$\theta^n / (x_1 \cdots x_n)^2 1_{\{\min x_i\}}.$$

By factorization theorem,  $T = \min X_i$  is sufficient.

To obtain the density of  $T$ , the cumulative distribution function is

$$P(T \leq t) = 1 - P(\min X_i > t) = 1 - \prod_i P(X_i > t) = 1 - \left(\frac{\theta}{t}\right)^n, \quad t > \theta.$$

Take a derivative, we can obtain the density,

$$f(t) = \frac{n\theta^n}{t^{n+1}}, \quad t > \theta.$$

Now, suppose  $\delta(T)$  is the unbiased estimator of  $g(\theta)$ , then

$$\int_{\theta}^{\infty} \delta(t) \frac{n\theta^n}{t^{n+1}} dt = g(\theta) \iff n \int_{\theta}^{\infty} \delta(t) t^{-n-1} dt = \frac{g(\theta)}{\theta^n}$$

Taking a derivative w.r.t.  $\theta$ ,

$$\iff -n\delta(t)t^{-n-1} = g'(t)t^{-n} - ng(t)t^{-n-1}.$$

$$\therefore \delta(t) = g(t) - \frac{g'(t)}{nt}.$$

From the above result,  $g(t) = c$  for any constant  $c$  implies  $\delta(t) = c$ . Therefore  $T$  is complete.

In summary,  $T$  is complete sufficient, and  $\delta(T)$  is the function of  $T$ . Therefore, by Theorem 4.4,  $\delta(T)$  is UMVU.  $\square$

## Problem 4.10

Suppose  $X$  is an exponential variable with density  $p_{\theta}(x) = \theta e^{-\theta x}$ ,  $x > 0$ ;  $p_{\theta}(x) = 0$ , otherwise. Find the UMVU estimator for  $1/(1 + \theta)$ .

### Solution.

Since  $X$  is a full-rank exponential family,  $T = X$  is complete sufficient.

Let  $\delta(X)$  be an unbiased estimator of  $1/(1 + \theta)$ , and represent as  $\delta(x) = \sum_{n=0}^{\infty} c_n x^n$  by power series. Then,

$$E\delta(X) = \int_0^{\infty} \left( \sum_{n=0}^{\infty} c_n x^n \right) \theta e^{-\theta x} dx$$

By Fubini theorem,

$$= \sum_{n=0}^{\infty} \int_0^{\infty} c_n \theta x^n e^{-\theta x} dx$$

The inner integration follows  $\Gamma(n + 1, 1/\theta)$  with rate parameter.

$$= \sum_{n=0}^{\infty} \frac{c_n \Gamma(n + 1)}{\theta^n} \\ = \sum_{n=0}^{\infty} \frac{c_n n!}{\theta^n}.$$

(\*)

Meanwhile,  $1/(1 + \theta)$  can be written as the geometric series,

$$\frac{1}{1 + \theta} = \frac{1/\theta}{1 + 1/\theta} = - \sum_{n=1}^{\infty} \left(-\frac{1}{\theta}\right)^n = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{\theta^n} \quad \text{for } \theta > 1. \quad (**)$$



Since  $\delta$  is unbiased,  $(*)$  and  $(**)$  should be identical. Thus,

$$c_0 = 0, \quad c_n = \frac{(-1)^{n+1}}{n!}, \quad n = 1, 2, \dots$$

Therefore,

$$\delta(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n+1}}{n!} x^n = - \sum_{n=1}^{\infty} \frac{(-x)^n}{n!} = 1 - e^{-x} \quad \text{for } \theta > 1.$$

Since  $\delta$  is the function of the complete sufficient statistic  $X$ ,  $\delta(X)$  is UMVU by Theorem 4.4.

If  $\theta \leq 1$ , direct하게 unbiased임을 보일 수 있다?? 어떻게?? □

## Problem 4.11

Let  $X_1, \dots, X_3$  be i.i.d. geometric variables with common mass function  $f_\theta(x) = P_\theta(X_i = x) = \theta(1-\theta)^x$ ,  $x = 0, 1, \dots$ . Find the UMVU estimator of  $\theta^2$ .

**Solution.**

Since the joint mass,  $\theta^3(1-\theta)^{x_1+x_2+x_3}$ , is the full-rank exponential family,  $T = X_1 + X_2 + X_3$  is complete sufficient.

Denote  $g(\theta) = \theta^2$ , and let  $\delta(X) = 1_{\{X_1=0, X_2=0\}}$ , then  $\delta(X)$  be an unbiased estimator of  $g(\theta)$  ( $\because E\delta(X) = P(X_1=0, X_2=0) = \theta^2$ ).

Define  $\eta(T) = E[\delta(X)|T] = P(X_1=0, X_2=0|T)$ . Then it is UMVU by Theorem 4.4. To obtain  $\eta(t)$ , we first find the density of  $T$  by using the transformation  $T = X_1 + X_2 + X_3$  from the joint density,

$$P(T=t) = \binom{t+2}{2} \theta^3(1-\theta)^t = \binom{t+2}{2} \theta^2 [\theta(1-\theta)^t].$$

Because we already have 2 successes in first  $t+2$  trials of total  $t+3$  trials, we should add the combination term. (See RHS!!)

Next, the joint mass of  $X_1, X_2, T$  is

$$P(X_1=0, X_2=0, T=t) = P(X_1=0, X_2=0, X_3=t) = \theta^3(1-\theta)^t,$$

from the joint mass of  $X_1, X_2, X_3$ .

Therefore,

$$\begin{aligned} \eta(t) &= \frac{P(X_1=0, X_2=0, T=t)}{P(T=t)} = 1 / \binom{t+2}{2} = \frac{2}{(t+2)(t+1)}, \\ \therefore \eta(T) &= \frac{2}{(T+2)(T+1)} \end{aligned}$$

is the UMVU estimator of  $\theta^2$ . □

## Problem 4.12

Let  $X$  be a single observation, absolutely continuous with density

$$p_\theta(x) = \begin{cases} \frac{1}{2}(1+\theta x), & |x| < 1, \\ 0, & |x| \geq 1. \end{cases}$$

Here  $\theta \in [-1, 1]$  is an unknown parameter.

- a) Find a constant  $a$  so that  $aX$  is unbiased for  $\theta$ . The expected value is

**Solution.**

$$EX = \int_{-1}^1 \frac{x}{2}(1 + \theta x)dx = \frac{1}{4}x^2 + \frac{\theta}{6}x^3 \Big|_{-1}^1 = \frac{\theta}{3}.$$

Therefore, when  $a = 3$ ,  $aX$  be unbiased. □

- b) Show that  $b = E_\theta|X|$  is independent of  $\theta$ .

**Solution.**

By using the procedure in a),

$$E|X| = \int_{-1}^1 \frac{|x|}{2}(1 + \theta x)dx = \frac{1}{2}.$$

Thus,  $b = E_\theta|X|$  does not depend on  $\theta$ . □

- c) Let  $\theta_0$  be a fixed parameter value in  $[-1, 1]$ . Determine the constant  $c = c_{\theta_0}$  that minimizes the variance of the unbiased estimator  $aX + c(|X| - b)$  when  $\theta = \theta_0$ . Is  $aX$  uniformly minimum variance unbiased?

**Solution.**

To obtain the variance of  $aX + c(|X| - b)$ , we first derive the second moments.

$$EX^2 = E|X|^2 = \int_{-1}^1 \frac{x^2}{2}(1 + \theta x)dx = \frac{x^3}{6} + \frac{\theta x^4}{8} \Big|_{-1}^1 = \frac{1}{3},$$

$$EX|X| = \int_{-1}^1 \frac{x|x|}{2}(1 + \theta x)dx = \frac{\theta}{4}.$$

From the previous steps, we obtain  $a = 3$  and  $b = E|X|$ . Under  $\theta = \theta_0$ , the variance be

$$\begin{aligned} \text{Var}_{\theta_0}(3X + c(|X| - E|X|)) &= 9\text{Var}_{\theta_0}(X) + 6c\text{Cov}_{\theta_0}(X, |X| - E|X|) + c^2\text{Var}_{\theta_0}(|X| - E|X|) \\ &= 9\text{Var}_{\theta_0}(X) + 6c\text{Cov}_{\theta_0}(X, |X| - E|X|) + c^2\text{Var}_{\theta_0}(|X|) \\ &= 9\left(\frac{1}{3} - \frac{\theta_0^2}{9}\right) + 6c\left(\frac{\theta_0}{4} - \frac{1}{2} \cdot \frac{\theta_0}{3}\right) + c^2\left(\frac{1}{3} - \frac{1}{4}\right) \\ &= 3 - \theta_0^2 + \frac{\theta_0}{2}c + \frac{1}{12}c^2. \end{aligned} \tag{*}$$

Since (\*) is convex with respect to  $c$ , take a derivative,

$$\frac{\theta_0}{2} + \frac{c}{6} = 0$$

Therefore, (\*) has minimum when  $c = -3\theta_0$ .

Meanwhile,  $\text{Var}_{\theta_0}(3X) = 3 - \theta_0^2$  which is greater than (\*). Therefore,  $3X$  is not UMVU. □

## Problem 4.24

In the normal one-sample problem, the statistic  $t = \sqrt{n}\bar{X}/S$  has the non-central  $t$ -distribution on  $n - 1$  degrees of freedom and non-centrality parameter  $\delta = \sqrt{n}\mu/\sigma$ . Use our results on distribution theory for the one-sample problem to find the mean and variance of  $t$ .

### Solution.

Note that  $E\bar{X} = \mu = \sigma\delta/\sqrt{n}$ ,  $\text{Var}(\bar{X}) = \sigma^2/n$ , and the formula  $ES^r$  in (4.10) page 70.

Since  $\bar{X}$  and  $S^2$  are independent,

$$Et = E\left[\sqrt{n}\frac{\bar{X}}{S}\right] = \sqrt{n}(E\bar{X})(ES^{-1}) = \sqrt{n}\frac{\sigma\delta}{\sqrt{n}}\frac{\sigma^{-1}2^{-\frac{1}{2}}\Gamma[(-1+n-1)/2]}{(n-1)^{-\frac{1}{2}}\Gamma[(n-1)/2]} = \delta\frac{\sqrt{n-1}\Gamma[(n-2)/2]}{\sqrt{2}\Gamma[(n-1)/2]},$$

$$\begin{aligned} Et^2 &= n(E\bar{X}^2)(ES^2) = n\left(\frac{\sigma^2}{n} + \frac{\sigma^2\delta^2}{n}\right)\frac{\sigma^{-2}2^{-1}\Gamma[(-2+n-1)/2]}{(n-1)^{-1}\Gamma[(n-1)/2]} \\ &= (1+\delta^2)\frac{n-1}{2}\frac{\Gamma[(n-3)/2]}{\Gamma[(n-3)/2+1]} = (1+\delta^2)\frac{n-1}{n-3}. \end{aligned}$$

$$\text{Var}(t) = Et^2 - (Et)^2 = \frac{n-1}{n-3} + \delta^2\left[\frac{n-1}{n-3} - \frac{(n-1)(\Gamma[(n-2)/2])^2}{2(\Gamma[(n-1)/2])^2}\right].$$

□

## Problem 4.28

Let  $X_1, \dots, X_n$  be i.i.d. from the uniform distribution on  $(0, \theta)$ .

- a) Use the Hammersley-Chapman-Robbins inequality to find a lower bound for the variance of an unbiased estimator of  $\theta$ . This bound will depend on  $\Delta$ . Note that  $\Delta$  cannot vary freely but must lie in a suitable set.

### Solution.

Since  $\delta$  is the unbiased estimator of  $\theta$ ,  $g(\theta) = \theta$  and  $g(\theta + \Delta) - g(\theta) = \Delta$ . By Hammersley-Chapman-Robbins inequality, the lower bound for the variance of  $\delta$  be

$$LB = \frac{\Delta^2}{E_\theta \left[ \frac{p_{\theta+\Delta}(X)}{p_\theta(X)} - 1 \right]^2}.$$

Since  $X_i \sim \text{Uniform}(0, \theta)$ ,  $p_\theta(X) = 0 \Rightarrow p_{\theta+\Delta}(X) = 0$ . Thus  $\Delta$  should be negative and  $\theta + \Delta > 0$ , therefore  $\Delta \in (-\theta, 0)$ .

$$\frac{p_{\theta+\Delta}(X)}{p_\theta(X)} = \begin{cases} \frac{\theta^n}{(\theta+\Delta)^n}, & \max\{X_1, \dots, X_n\} < \theta + \Delta \text{ with probability } \frac{(\theta+\Delta)^n}{\theta^n} \text{ under } P_\theta, \\ 0, & \text{otherwise.} \end{cases}$$

For simplicity, denote  $Y = p_{\theta+\Delta}(X)/p_\theta(X)$ , then  $EY = 1$  and  $EY^2 = \frac{\theta^n}{(\theta+\Delta)^n}$ . Thus, the expectation in the denominator is

$$E(Y-1)^2 = \text{Var}(Y) = EY^2 - [EY]^2 = \frac{\theta^n}{(\theta+\Delta)^n} - 1.$$

Therefore, the lower bound be

$$LB = \frac{\Delta^2}{\frac{\theta^n}{(\theta+\Delta)^n} - 1}. \quad (*)$$

□

- b) In principle, the best lower bound can be found taking the supremum over  $\Delta$ . This calculation cannot be done explicitly, but an approximation is possible. Suppose  $\Delta = -c\theta/n$ . Show that the lower bound for the variance can be written as  $\theta^2 g_n(c)/n^2$ . Determine  $g(c) = \lim_{n \rightarrow \infty} g_n(c)$ .

**Solution.**

Plug in  $\Delta = -c\theta/n$  to (\*), then

$$LB = \frac{c^2 \theta^2 / n^2}{\left(1 - \frac{c}{n}\right)^{-n} - 1}.$$

Therefore,

$$g_n(c) = \frac{c^2}{\left(1 - \frac{c}{n}\right)^{-n} - 1},$$

$$g(c) = \lim_{n \rightarrow \infty} g_n(c) = \frac{c^2}{e^c - 1}.$$

□

- c) Find the value  $c_0$  that maximizes  $g(c)$  over  $c \in (0, 1)$  and give an approximate lower bound for the variance of  $\delta$ . (The value  $c_0$  cannot be found explicitly, but you should be able to come up with a numerical value.)

**Solution.**

Taking a derivative,

$$\frac{\partial g(c)}{\partial c} = \frac{\partial}{\partial c} \left( \frac{c^2}{e^c - 1} \right) = \frac{2c(e^c - 1) - c^2 e^c}{(e^c - 1)^2} = 0$$

$$\iff 2c(e^c - 1) = c^2 e^c$$

$$\iff 2e^c - 2 = ce^c$$

$$\iff 1 - e^c = \frac{c}{2}$$

$$\iff e^{-c} = 1 - \frac{c}{2}.$$

By numerical method,  $c_0 = 1.59362$ . Therefore, an approximated lower bound of  $g(c_0)\theta^2/n^2 = 0.64761\theta^2/n^2$ .

□

### Problem 4.30

Suppose  $X_1, \dots, X_n$  are independent with  $X_i \sim N(\alpha + \beta t_i, 1)$ ,  $i = 1, \dots, n$ , where  $t_1, \dots, t_n$  are known constants and  $\alpha, \beta$  are unknown parameters.

- a) Find the Fisher information matrix  $I(\alpha, \beta)$ .

**Solution.**

The log likelihood of  $\theta = (\alpha, \beta)$  is

$$l(\beta) = \frac{n}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^n (x_i - \alpha - \beta t_i)^2.$$

Take derivatives,

$$\begin{aligned} \frac{\partial l}{\partial \alpha} &= \sum_i (x_i - \alpha - \beta t_i), & \frac{\partial l}{\partial \beta} &= \sum_i (x_i - \alpha - \beta t_i) t_i, \\ \frac{\partial^2 l}{\partial \alpha^2} &= -n, & \frac{\partial^2 l}{\partial \beta^2} &= -\sum_i t_i^2, \\ \frac{\partial^2 l}{\partial \alpha \beta} &= -\sum_i t_i \end{aligned}$$

Therefore, the Fisher information matrix be

$$I(\alpha, \beta) = -E [\nabla l(\alpha, \beta)]^2 = \begin{pmatrix} n & \sum_i t_i \\ \sum_i t_i & \sum_i t_i^2 \end{pmatrix}.$$

□

- b) Give a lower bound for the variance of an unbiased estimator of  $\alpha$ .

**Solution.**

Since  $g(\theta) = \alpha$ , by Cramer-Rao lower bound,

$$\begin{aligned} LB &= \nabla g(\theta)^T I^{-1}(\alpha, \beta) \nabla g(\theta) \\ &= (1, 0) \frac{\begin{pmatrix} \sum_i t_i^2 & -\sum_i t_i \\ -\sum_i t_i & n \end{pmatrix}}{n \sum_i t_i^2 - (\sum_i t_i)^2} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{1}{n \sum_i t_i^2 - (\sum_i t_i)^2} \begin{pmatrix} \sum_i t_i^2 & -\sum_i t_i \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \\ &= \frac{\sum_i t_i^2}{n \sum_i t_i^2 - (\sum_i t_i)^2} \\ &= \frac{1/n}{1 - (\sum_i t_i)^2 / \sum_i t_i^2}. \end{aligned}$$

□

- c) Suppose we know the value of  $\beta$ . Give a lower bound for the variance of an unbiased estimator of  $\alpha$  in this case.

**Solution.**

Since  $\beta$  is known,  $\theta = \alpha$ . Then the lower bound be

$$LB = \frac{g'(\theta)}{I(\theta)} = \frac{1}{n}.$$

□

- d) Compare the estimators in parts (b) and (c). When are the bounds the same? If the bounds are different, which is larger?

**Solution.**

If  $\sum_i t_i = 0$ , two lower bounds be same. And by comparing the lower bounds,

$$LB_b) = \frac{1/n}{1 - (\sum_i t_i)^2 / \sum_i t_i^2} > \frac{1}{n} = LB_c).$$

□

e) Give a lower bound for the variance of an unbiased estimator of the product  $\alpha\beta$ .

**Solution.**

Since  $g(\theta) = \alpha\beta$ ,  $\nabla g(\theta) = \begin{pmatrix} \beta \\ \alpha \end{pmatrix}$ . Then the lower bound be

$$\begin{aligned} LB &= \nabla g(\theta)^T I^{-1}(\alpha, \beta) \nabla g(\theta) \\ &= (\beta, \alpha) \frac{\begin{pmatrix} \sum_i t_i^2 & -\sum_i t_i \\ -\sum_i t_i & n \end{pmatrix}}{n \sum_i t_i^2 - (\sum_i t_i)^2} \begin{pmatrix} \beta \\ \alpha \end{pmatrix} \\ &= \frac{1}{n \sum_i t_i^2 - (\sum_i t_i)^2} \begin{pmatrix} \beta \sum_i t_i^2 - \alpha \sum_i t_i, & -\beta \sum_i t_i + n\alpha \end{pmatrix} \begin{pmatrix} \beta \\ \alpha \end{pmatrix} \\ &= \frac{\beta^2 \sum_i t_i^2 - 2\alpha\beta \sum_i t_i + n\alpha^2}{n \sum_i t_i^2 - (\sum_i t_i)^2}. \end{aligned}$$

□

**Problem 4.31**

Find the Fisher information for the Cauchy location family with densities  $p_\theta$  given by

$$p_\theta(x) = \frac{1}{\pi[(x - \theta)^2 + 1]}.$$

Also, what is the Fisher information for  $\theta^3$ ?

**Solution.**

Since  $p_\theta$  is the location family, by using Example 4.11 in page 75,

$$\begin{aligned} I(\theta) &= \int \frac{[f'(x)]^2}{f(x)} dx = \int \frac{\left[-\frac{2\pi x}{\pi(x^2+1)}\right]^2}{1/[\pi(x^2+1)]} = \int \frac{4x^2}{\pi(x^2+1)^3} dx \\ \text{Change of variable } x &= \tan t \Rightarrow dx = \frac{\cos^2 t + \sin^2 t}{\cos^2 t} = \frac{1}{\cos^2 t} dt \\ &= \frac{4}{\pi} \int_{-\pi/2}^{\pi/2} \frac{\tan^2 t}{(\tan^2 t + 1)^3} \frac{1}{\cos^2 t} dt \\ &= \frac{4}{\pi} \int_{-\pi/2}^{\pi/2} \frac{\sin^2 t}{\cos^2 t} \cos^6 t \frac{1}{\cos^2 t} dt \quad (\because \tan^2 t + 1 = 1/\cos^2 t) \\ &= \frac{4}{\pi} \int_{-\pi/2}^{\pi/2} \sin^2 t \cos^2 t dt \\ &= \frac{4}{\pi} \int_{-\pi/2}^{\pi/2} \frac{1 - \cos 2x}{2} \frac{1 + \cos 2x}{2} dt \end{aligned}$$

$$\begin{aligned}
&= \frac{4}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1 - \cos^2 2x}{4} dt \\
&= \frac{1}{\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \frac{1 - \cos 4x}{2} dt \\
&= \frac{1}{2\pi} \left[ t - \frac{1}{4} \sin 4t \right]_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \\
&= \frac{1}{2}.
\end{aligned}$$

Let  $\xi = \theta^3$ , then  $\theta = \xi^{1/3} := h(\xi)$ .

$$\therefore I(\xi) = [h'(\xi)]^2 I(h(\xi)) = \left( \frac{1}{3} \xi^{-2/3} \right)^2 \frac{1}{2} = \frac{1}{18 \xi^{4/3}}.$$

□

### Problem 4.32

Suppose  $X$  has a Poisson distribution with mean  $\theta^2$ , so the parameter  $\theta$  is the square root of the usual parameter  $\lambda = EX$ . Show that the Fisher information  $I(\theta)$  is constant.

**Solution.**

$p_\theta(X) = (\theta^2)^x e^{-\theta^2} / x!$  and  $l(\theta) = \log 1/x! + 2x \log \theta - \theta^2$ .

$$\frac{\partial l(\theta)}{\partial \theta} = \frac{2x}{\theta} - 2\theta, \quad \frac{\partial^2 l(\theta)}{\partial \theta^2} = -\frac{2x}{\theta^2} - 2.$$

$$\therefore I(\theta) = E \left( -\frac{\partial^2 l(\theta)}{\partial \theta^2} \right) = \frac{2}{\theta^2} EX + 2 = 4.$$

□

### Problem 4.33

Consider the exponential distribution with failure rate  $\lambda$ . Find a function  $h$  defining a new parameter  $\theta = h(\lambda)$  so that Fisher information  $I(\theta)$  is constant.

**Solution.**

Let the original parameter is  $\theta$ , and reparameterized parameter is  $\lambda$ .

$p_\lambda(x) = \lambda e^{-\lambda x}$ ,  $l(\lambda) = \log \lambda - \lambda x$ .

$$\frac{\partial l(\lambda)}{\partial \lambda} = \frac{1}{\lambda} - x, \quad \frac{\partial^2 l(\lambda)}{\partial \lambda^2} = -\frac{1}{\lambda^2}.$$

$$\tilde{I}(\lambda) = \frac{1}{\lambda^2}.$$

From (4.18) in page 74,

$$I(\theta) = \frac{\tilde{I}(\lambda)}{[h'(\lambda)]^2} = \frac{1}{[\lambda h'(\lambda)]^2}.$$

Thus, if  $h(\lambda) = \log \lambda$ ,  $I(\theta)$  be constant.

□

### Problem 4.34

Consider an autoregressive model in which  $X_1 \sim N(\theta, \sigma^2/(1-\rho^2))$  and the conditional distribution of  $X_{j+1}$  given  $X_1 = x_1, \dots, X_j = x_j$ , is  $N(\theta + \rho(x_j - \theta), \sigma^2)$ ,  $j = 1, \dots, n-1$ .

a) Find the Fisher information matrix,  $I(\theta, \sigma)$ .

**Solution.**

The joint density is

$$p(x_1, \dots, x_n) = p(x_1)p(x_2|x_1) \cdots p(x_n|x_1, \dots, x_{n-1}),$$

thus the likelihood function be

$$\begin{aligned} l(\theta, \sigma) &= l(x_1) + \sum_{j=1}^{n-1} l(x_{j+1}|x_1, \dots, x_j) \\ &= \log \frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}} - \log \sigma - \frac{1-\rho^2}{2\sigma^2}(x_1 - \theta)^2 \\ &\quad - \frac{1}{\sigma^2} \sum_{j=1}^{n-1} [x_{j+1} - \theta - \rho(x_j - \theta)]^2 + \log \left( \frac{1}{\sqrt{2\pi}\sigma} \right)^{n-1} \\ &= -\frac{1-\rho^2}{2\sigma^2}(x_1 - \theta)^2 - \frac{1}{2\sigma^2} \sum_{j=1}^{n-1} [x_{j+1} - \rho x_j - \theta(1-\rho)]^2 \\ &\quad - \frac{n}{2} \log(2\pi) - n \log \sigma + \log \sqrt{1-\rho^2}. \end{aligned}$$

Take derivatives,

$$\begin{aligned} \frac{\partial l(\theta, \sigma)}{\partial \theta} &= \frac{1-\rho^2}{\sigma^2}(x_1 - \theta) + \frac{1-\rho}{\sigma^2} \sum_{j=1}^{n-1} [x_{j+1} - \rho x_j - \theta(1-\rho)], \\ \frac{\partial^2 l(\theta, \sigma)}{\partial \theta^2} &= -\frac{1-\rho^2}{\sigma^2} - \frac{(n-1)(1-\rho)^2}{\sigma^2}, \\ \frac{\partial^2 l(\theta, \sigma)}{\partial \theta \partial \sigma} &= -\frac{2(1-\rho^2)\epsilon_1}{\sigma^3} + \frac{2(1-\rho)}{\sigma^3} \sum_{j=1}^{n-1} \eta_{j+1}, \\ \frac{\partial l(\theta, \sigma)}{\partial \sigma} &= \frac{(1-\rho^2)(x_1 - \theta)^2}{\sigma^3} + \frac{1}{\sigma^3} \sum_{j=1}^{n-1} \eta_{j+1}^2 - \frac{n}{\sigma}, \\ \frac{\partial^2 l(\theta, \sigma)}{\partial \sigma^2} &= -\frac{3(1-\rho^2)(x_1 - \theta)^2}{\sigma^4} - \frac{3}{\sigma^4} \sum_{j=1}^{n-1} \eta_{j+1}^2 + \frac{n}{\sigma^2}, \end{aligned}$$

where  $\epsilon_j = x_j - \theta$ , and  $\eta_{j+1} = \epsilon_{j+1} - \rho\epsilon_j$ . (See [Hint](#) in part c))

We can easily infer that  $\epsilon_1 \sim N(0, \sigma^2/(1-\rho^2))$  and  $\eta_{j+1}|X_1=x_1, \dots, X_j=x_j = \eta_{j+1}|\epsilon_1, \eta_2, \dots, \eta_j \sim N(0, \sigma^2)$ . From this,  $\eta_2, \dots, \eta_n \sim \text{i.i.d. } N(0, \sigma^2)$  and these are independent of  $\epsilon_1$ . (???)

Therefore, the information matrix is given by taking expectation,

$$I(\theta, \sigma) = \begin{pmatrix} \frac{1-\rho^2+(n-1)(1-\rho)^2}{\sigma^2} & 0 \\ 0 & \frac{2n}{\sigma^2} \end{pmatrix}.$$

□



- b) Give a lower bound for the variance of an unbiased estimator of  $\theta$ .

**Solution.**

Since  $g(\theta, \sigma) = \theta$ ,

$$LB = \nabla g(\theta, \sigma)^T I^{-1}(\theta, \sigma) \nabla g(\theta, \sigma) = (1, 0) I^{-1}(\theta, \sigma) \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \frac{\sigma^2}{1 - \rho^2 + (n-1)(1-\rho)^2}.$$

□

- c) Show that the sample average  $\bar{X} = (X_1 + \cdots + X_n)/n$  is an unbiased estimator of  $\theta$ , compute its variance, and compare its variance with the lower bound.

Hint: Define  $\epsilon_j = X_j - \theta$  and  $\eta_{j+1} = \epsilon_{j+1} - \epsilon_j$ . Use smoothing to argue that  $\eta_2, \dots, \eta_n$  are i.i.d.  $N(0, \sigma^2)$  and are independent of  $\epsilon_1$ . Similarly,  $X_i$  is independent of  $\eta_{i+1}, \eta_{i+2}, \dots$ . Use these facts to find first  $\text{Var}(X_2) = \text{Var}(\epsilon_2)$ , then  $\text{Var}(X_3), \text{Var}(X_4), \dots$ . Finally find  $\text{Cov}(X_{i+1}, X_i), n\text{Cov}(X_{i+2}, X_i)$ , and so on.

**Solution.**

Let  $\bar{\epsilon} = \frac{1}{n} \sum_j \epsilon_j = \bar{X} - \theta$ . Since  $E\epsilon_j = 0$ ,  $E(\bar{X} - \theta) = E(\bar{\epsilon}) = 0$ . Thus  $\bar{X}$  is unbiased estimator of  $\theta$ .

Since  $\epsilon_2 = \rho\epsilon_1 + \eta_2$ ,  $\text{Var}(\epsilon_2) = \rho^2 \text{Var}(\epsilon_1) + \text{Var}(\eta_2) = \sigma^2 + \frac{\rho^2 \sigma^2}{1 - \rho^2} = \frac{\sigma^2}{1 - \rho^2}$ . In general,  $\epsilon_i = \rho\epsilon_{i-1} + \eta_i$  and we can obtain  $\text{Var}(X_i) = \text{Var}(\epsilon_i) = \frac{\sigma^2}{1 - \rho^2}$ .

If  $i > j$ ,

$$\begin{aligned} \epsilon_i &= \rho\epsilon_{i-1} + \eta_i \\ &= \rho^2(\epsilon_{i-2} + \eta_{i-1}) + \eta_i \\ &= \cdots \\ &= \rho^{i-j}\epsilon_j + \rho^{i-j+1}\eta_{j+1} + \cdots + \rho\eta_{i-1} + \eta_i. \end{aligned}$$

From this,

$$\text{Cov}(X_i, X_j) = \text{Cov}(\epsilon_i, \epsilon_j) = \rho^{|i-j|} \frac{\sigma^2}{1 - \rho^2}.$$

Then, the variance of  $\bar{X}$  is

$$\begin{aligned} \text{Var}(\bar{X}) &= \text{Var}(\bar{\epsilon}) \\ &= \frac{1}{n^2} \text{Var} \left( \sum_{i=1}^n \epsilon_i \right) = \frac{1}{n^2} \text{Cov} \left( \sum_{i=1}^n \epsilon_i, \sum_{i=1}^n \epsilon_i \right) \\ &= \frac{1}{n^2} \left[ \sum_{i=1}^n \text{Cov}(\epsilon_i, \epsilon_i) + 2 \sum_{i>j} \text{Cov}(\epsilon_i, \epsilon_j) \right] \\ &= \frac{\sigma^2}{n^2(1 - \rho^2)} [n + 2 \{ (n-1)\rho + (n-2)\rho^2 + \cdots + \rho^{n-1} \}] \\ &= \frac{\sigma^2}{n^2(1 - \rho^2)} \left[ n + 2 \sum_{j=1}^{n-1} (n-j)\rho^j \right] \\ &= \frac{\sigma^2}{n^2(1 - \rho^2)} \left[ n + 2 \left( n \frac{\rho - \rho^n}{1 - \rho} - \frac{\rho - \rho^{n+1}}{(1 - \rho)^2} + \frac{n\rho^n}{1 - \rho} \right) \right] \quad (*) \\ &= \frac{\sigma^2}{n^2(1 - \rho^2)} \left[ n + 2 \frac{n\rho - n\rho^2 - \rho + \rho^{n+1}}{(1 - \rho)^2} \right] \end{aligned}$$

$$\begin{aligned}
&= \frac{\sigma^2}{n^2(1-\rho^2)} \left[ \frac{n - 2n\rho + n\rho^2 + 2n\rho - 2n\rho^2 - 2\rho + 2\rho^{n+1}}{(1-\rho)^2} \right] \\
&= \frac{\sigma^2}{n^2(1-\rho^2)} \left[ \frac{-n\rho^2 - 2\rho + n + 2\rho^{n+1}}{(1-\rho)^2} \right] \\
&= \frac{\sigma^2}{n^2(1-\rho^2)} \left[ \frac{n(1-\rho^2) - 2(\rho - \rho^{n+1})}{(1-\rho)^2} \right] \\
&= \frac{\sigma^2}{n(1-\rho)^2} - \frac{2\sigma^2\rho(1-\rho^n)}{n^2(1-\rho^2)(1-\rho)^2}.
\end{aligned}$$

Note. (\*) is obtained as follows:

$$\begin{aligned}
\sum_{j=1}^{n-1} \rho^j &= \frac{\rho(1-\rho^{n-1})}{1-\rho}, \\
\sum_{j=1}^{n-1} j\rho^j &= \sum_j \rho \frac{d}{d\rho} \rho^j = \rho \frac{d}{d\rho} \sum_{j=1}^{n-1} \rho^j \\
&= \rho \frac{d}{d\rho} \left( \frac{\rho - \rho^n}{1-\rho} \right) \\
&= \rho \frac{(1-n\rho^{n-1})(1-\rho) + (\rho - \rho^n)}{(1-\rho)^2} \\
&= \rho \frac{1-\rho - n\rho^{n-1} + n\rho^n + \rho - \rho^n}{(1-\rho)^2} \\
&= \rho \frac{1-\rho^n - n\rho^{n-1}(1-\rho)}{(1-\rho)^2} \\
&= \frac{\rho - \rho^{n+1}}{(1-\rho)^2} - \frac{n\rho^n}{1-\rho}.
\end{aligned}$$

□

## Chapter 5

# Curved Exponential Families

### Problem 5.1

Suppose  $X$  has a binomial distribution with  $m$  trials and success probability  $\theta$ ,  $Y$  has a binomial distribution with  $n$  trials and success probability  $\theta^2$ , and  $X$  and  $Y$  are independent.

- a) Find a minimal sufficient statistic  $T$ .

**Solution.**

Note that  $f(x) = \binom{m}{x}\theta^x(1-\theta)^{m-x}$  and  $f(y) = \binom{n}{y}(\theta^2)^y(1-\theta^2)^{n-y}$ . Then the joint density is

$$f(x, y) = \binom{m}{x} \binom{n}{y} \exp \left[ x \log \frac{\theta}{1-\theta} + y \log \frac{\theta^2}{1-\theta^2} + m \log(1-\theta) + n \log(1-\theta^2) \right].$$

Since the canonical parameter  $\eta = \left( \log \frac{\theta}{1-\theta}, \log \frac{\theta^2}{1-\theta^2} \right)$  does not satisfy a linear constraint, the joint density is the curved exponential family. Therefore,  $T = (X, Y)$  is minimal sufficient.  $\square$

- b) Show that  $T$  is not complete, providing a non-trivial function  $f$  with  $E_\theta f(T) = 0$ .

**Solution.**

Note that  $ET_1 = EX = m\theta$ ,  $ET_1^2 = EX^2 = m\theta(1-\theta) + m^2\theta^2$  and  $ET_2 = EY = n\theta^2$ .

Now we let

$$f(T) = T_1^2 - T_1 - \frac{m(m-1)}{n}T_2,$$

then  $E_\theta f(T) = 0$  for all  $\theta \in (0, 1)$ . But  $f(T)$  is not always 0. Thus  $T$  is not complete.  $\square$

### Problem 5.2

Let  $X$  and  $Y$  be independent Bernoulli variables with  $P(X = 1) = p$  and  $P(Y = 1) = h(p)$  for some known function  $h$ .

- a) Show that the family of joint distributions is a curved exponential family unless

$$h(p) = \frac{1}{1 + \exp \left\{ a + b \log \frac{p}{1-p} \right\}}$$

for some constants  $a$  and  $b$ .

**Solution.**

The joint density is

$$f(x, y) = p^x(1-p)^{1-x}h(p)^y(1-h(p))^{1-y} = \exp \left[ x \log \left( \frac{p}{1-p} \right) + y \log \left( \frac{h(p)}{1-h(p)} \right) \right] (1-p)(1-h(p)).$$

If  $h(p)$  defined as the above for some constants  $a$  and  $b$ , then

$$\log \left( \frac{h(p)}{1-h(p)} \right) = - \left( a + b \log \left( \frac{p}{1-p} \right) \right)$$

which is the linear constraint. Therefore, unless the above case, the family of joint densities is a curved exponential family.  $\square$

- b) Give two functions  $h$ , one where  $(X, Y)$  is minimal but not complete, and one where  $(X, Y)$  is minimal and complete.

**Solution.**

- (i) Minimal but not complete

From a), the joint density is a curved exponential family except on the above  $h$  form.

If we let a function  $h(p) = \frac{p}{2}$ ,  $E_p(X - 2Y) = 0$  for all  $p \in (0, 1)$ . But  $X - 2Y$  is not always 0. Therefore  $T$  is not complete.

- (ii) Minimal and complete

For some function  $g$ ,

$$E_p g(X, Y) = g(1, 1)ph(p) + g(1, 0)p(1-h(p)) + g(0, 1)(1-p)h(p) + g(0, 0)(1-p)(1-h(p)).$$

To show  $T$  is complete, we need to find a function  $h$  that  $ph(p), p(1-h(p)), (1-p)h(p)$  are linearly independent.

If we let  $h(p) = p^2$ , then

$$\begin{aligned} E_p g(X, Y) &= p^3 [g(1, 1) - g(1, 0) - g(0, 1) + g(0, 0)] \\ &\quad + p^2 [g(0, 1) - g(0, 0)] + p [g(1, 0) - g(0, 0)] + g(0, 0). \end{aligned}$$

If this is 0 for all  $p \in (0, 1)$ , it must be  $g(0, 0) = g(1, 0) = g(0, 1) = g(1, 1) = 0$ . Thus  $g(X, Y) = 0$  and  $T$  is complete.  $\square$

## Problem 5.6

Two teams A and B play a series of games, stopping as soon as one of the team has 4 wins. Assume that game outcomes are independent and that on any given game team A has a fixed chance  $\theta$  of winning. Let  $X$  and  $Y$  denote the number of games won by the first and second team, respectively.

- a) Find the joint mass function for  $X$  and  $Y$ . Show that as  $\theta$  varies these mass functions form a curved exponential family.

***Solution.***

First, suppose team A has 4 wins which is the event  $X = 4$  and  $Y = y$ . Then, the first  $(3 + y)$  trials are random and last trial should be the winning of team A, thus the probability be

$$P(X = 4, Y = y) = \binom{3+y}{3} \theta^4 (1-\theta)^y, \quad y = 0, 1, 2, 3.$$

Similarly,

$$P(X = x, Y = 4) = \binom{3+x}{3} \theta^x (1-\theta)^4, \quad x = 0, 1, 2, 3.$$

Therefore, the joint mass have the form  $h(x, y) \exp [x \log \theta + y \log(1-\theta)]$ . Since the canonical parameters  $\log \theta$  and  $\log(1-\theta)$  have non-linear relationship, the family of the joint mass is a curved exponential family.  $\square$

b) Show that  $T = (X, Y)$  is complete.

***Solution.***

Let  $g$  be an arbitrary function of  $X$  and  $Y$ , and suppose

$$E_{\theta} g(X, Y) = \sum_{x=0}^3 g(x, 4) \binom{3+x}{3} \theta^x (1-\theta)^4 + \sum_{y=0}^3 g(4, y) \binom{3+y}{3} \theta^4 (1-\theta)^y = 0, \quad \theta \in (0, 1).$$

Since it is the polynomial in  $\theta$ , if  $\theta \rightarrow 0$ , then the only constant term  $g(0, 4)$  is remain (other terms are tend to 0). Thus  $g(0, 4)$  should be 0.

If  $g(0, 4) = 0$ , then the linear term  $g(1, 4)$  also should be 0 (dividing by  $\theta$  and letting  $\theta \rightarrow 0$ ). Similarly  $g(2, 4) = g(3, 4) = 0$ .

Then, now remain only

$$\frac{E_{\theta} g(X, Y)}{\theta^4} = \sum_{y=0}^3 g(4, y) \binom{3+y}{3} (1-\theta)^y = 0,$$

and this is the polynomial in  $1-\theta$ . Similarly, we can obtain  $g(4, 0) = \dots = g(4, 3) = 0$ . Therefore,  $g(X, Y) = 0$  almost surely.

By definition of completeness,  $T$  is complete.  $\square$

c) Find a UMVU estimator of  $\theta$ .

***Solution.***

Let  $\delta = 1_{\{\text{team A wins the first game}\}}$ . Then  $\delta$  is unbiased because  $E\delta = P(\text{team A wins the first game}) = \theta$ . Thus the UMVU estimator of  $\theta$  be  $E[\delta|T] = P[\delta = 1|X, Y]$ . By using the joint mass function,

$$\begin{aligned} E[\delta = 1|X = 4, Y = y] &= P(\delta = 1|X = 4, Y = y) \\ &= \frac{P(\delta = 1, X = 4, Y = y)}{P(X = 4, Y = y)} \\ &= \frac{\theta^2 \binom{2+y}{2} \theta^2 (1-\theta)^y}{\binom{3+y}{3} \theta^4 (1-\theta)^y} \\ &= \frac{3}{3+y}, \end{aligned}$$

and

$$P(\delta = 1|X = x, Y = 4) = \frac{\theta(1 - \theta) \binom{2+x}{x-1} \theta^{x-1} (1 - \theta)^3}{\binom{3+x}{3} \theta^x (1 - \theta)^4} = \frac{x}{3 + x}.$$

By combining 2 results, the UMVU estimator of  $\theta$  is

$$\frac{X - 1_{\{X=4\}}}{X + Y - 1}.$$

□

## Problem 5.7

Consider a sequential experiment in which observations are i.i.d. from a Poisson distribution with mean  $\lambda$ . If the first observation  $X$  is zero, the experiment stops, and if  $X > 0$ , a second observation  $Y$  is observed. Let  $T = 0$  if  $X = 0$ , and let  $T = 1 + X + Y$  if  $X > 0$ .

- a) Find the mass function for  $T$ .

**Solution.**

$$P(T = 0) = P(X = 0) = e^{-\lambda}.$$

For  $k = 0, 1, \dots$ ,

$$\begin{aligned} P(T = k + 1) &= P(X + Y = k, X > 0) \\ &= P(X + Y = k) - P(X + Y = k, X = 0) \\ &= \frac{(2\lambda)^k e^{-2\lambda}}{k!} - \frac{\lambda^k e^{-2\lambda}}{k!}. \end{aligned}$$

이게 왜 아니지??

$$\begin{aligned} P(T = k + 1) &= P(X + Y = k, X > 0) \\ &= P(X + Y = k) - P(X + Y = k, X = 0) \\ &= P(X + Y = k) - P(Y = k) \\ &= \frac{(2\lambda)^k e^{-2\lambda}}{k!} - \frac{\lambda^k e^{-\lambda}}{k!}. \end{aligned}$$

□

- b) Show that  $T$  is minimal sufficient.

**Solution.**

Part c) first!!

From c),  $(W, N)$  is minimal sufficient. Therefore  $T$  that is one-to-one function of  $(W, N)$  is also minimal sufficient.

□

- c) Does this experiment give a curved two-parameter exponential family or full rank one-parameter exponential family?

**Solution.**

Let  $N$  is the sample size(or # of experiments), and  $W$  as follows:

$$N = \begin{cases} 1, & \text{if } X = 0, \\ 2, & \text{if } X > 0, \end{cases} \quad W = \begin{cases} 0, & \text{if } X = 0, \\ X + Y, & \text{if } X > 0. \end{cases}$$

Note that  $X$  and  $Y$  are from i.i.d.  $\text{Poisson}(\lambda)$  and it satisfies the condition in Theorem 5.4.

By Theorem 5.4, the joint density is an exponential family with canonical parameter  $\eta = (\log \lambda, -\lambda)$  and sufficient statistic  $(X + Y, N) = (W, N)$ .

Since  $\eta$  does not satisfy a linear constraint, the density of  $W$  and  $N$  is a curved exponential family. Thus,  $(W, N)$  is minimal sufficient.  $\square$

d) Is  $T$  a complete sufficient statistic?

Hint: Write  $e^\lambda E_\lambda g(T)$  as a power series in  $\lambda$  and derive equations for  $g$  setting coefficients for  $\lambda^x$  to zero.

**Solution.**

Suppose  $E_\lambda g(T) = 0$  for all  $\lambda > 0$  and any function  $g$ . Then

$$E_\lambda g(T) = g(0)e^{-\lambda} + \sum_{k=0}^{\infty} g(k+1) \frac{(2\lambda)^k e^{-2\lambda} - \lambda^k e^{-\lambda}}{k!} = 0.$$

Multiplying  $e^{2\lambda}$ ,

$$e^{2\lambda} E_\lambda g(T) = g(0)e^\lambda + \sum_{k=0}^{\infty} \frac{g(k+1)(2^k - 1)}{k!} \lambda^k = 0.$$

This is the power series in  $\lambda$ , so the coefficients of the power should be 0. ( $g(0) = 0$  and  $g(k+1) = 0$  for  $k = 1, 2, \dots$ )

Note that  $T \neq 1$  (· If  $X > 0$ , then  $Y \geq 0$  and  $T > 1$ ). Therefore,  $g(T) = 0$  thus  $T$  is complete.  $\square$

## Problem 5.13

Consider a single two-way contingency table and define  $R = N_{11} + N_{12}$  (the first row sum),  $C = N_{11} + N_{21}$  (the first column sum), and  $D = N_{11} + N_{22}$  (the sum of the diagonal entries).

a) Show that the joint mass function can be written as a full rank three-parameter exponential family with  $T = (R, C, D)$  as the canonical sufficient statistic.

**Solution.**

The contingency table be as follows:

$N_{11}$	$N_{12}$	$R$
$N_{21}$	$N_{22}$	
$C$		$n$

By representing the elements with  $R, C, D$ , and  $n$ ,

$$N_{11} = (R + C + D - n)/2, \quad N_{12} = (n - D - C + R)/2,$$

$$N_{21} = (n - D - R + C)/2, \quad N_{22} = (n + D - R - C)/2.$$

Thus the joint mass function is

$$\begin{aligned} \binom{n}{n_{11}, \dots, n_{22}} p_{11}^{n_{11}} \cdots p_{22}^{n_{22}} &= \binom{n}{n_{11}, \dots, n_{22}} (\sqrt{p_{11}})^{r+c+d-n} (\sqrt{p_{12}})^{n-d-c+r} (\sqrt{p_{21}})^{n-d-r+c} (\sqrt{p_{22}})^{n+d-r-c} \\ &= \binom{n}{n_{11}, \dots, n_{22}} \sqrt{\frac{p_{11}p_{12}}{p_{21}p_{22}}}^r \sqrt{\frac{p_{11}p_{21}}{p_{12}p_{22}}}^c \sqrt{\frac{p_{11}p_{22}}{p_{12}p_{21}}}^d \sqrt{\frac{p_{12}p_{21}p_{22}}{p_{11}}}^n \\ &= \binom{n}{n_{11}, \dots, n_{22}} \exp \left[ r \log \sqrt{\frac{p_{11}p_{12}}{p_{21}p_{22}}} + c \log \sqrt{\frac{p_{11}p_{21}}{p_{12}p_{22}}} \right. \\ &\quad \left. + d \log \sqrt{\frac{p_{11}p_{22}}{p_{12}p_{21}}} + n \log \sqrt{\frac{p_{12}p_{21}p_{22}}{p_{11}}} \right], \end{aligned}$$

where  $r, c$ , and  $d$  are defined as before.

The density is a full-rank(3-parameter) exponential family, and the sufficient statistic  $(R, C, D)$  do not satisfy a linear constraint because there is a one-to-one linear association between it and  $(N_{11}, N_{12}, N_{21})$ , and the 3 canonical parameters  $\eta_r, \eta_c$ , and  $\eta_d$  can vary freely over  $\mathbb{R}^3$ . **마지막 부분 잘 이해 안됨,,,**  $\square$

- b) Relate the canonical parameter associated with  $D$  to the "cross-product ratio"  $\alpha$  defined as  $\alpha = p_{11}p_{22}/(p_{12}p_{21})$ .

**Solution.**

$$\eta_d = \log \sqrt{\alpha} \quad \square$$

- c) Suppose we observe  $m$  independent two-by-two contingency tables. Let  $n_i$ ,  $i = 1, \dots, m$ , denote the trials for table  $i$ . Assume that cell probabilities for the tables may differ, but that the cross-product ratios for all  $m$  tables are all the same. Show that the joint mass functions form a full rank exponential family. Express the sufficient statistic as a function of the variables  $R_1, \dots, R_m, C_1, \dots, C_m$ , and  $D_1, \dots, D_m$ .

**Solution.**

The joint mass can be obtained as

$$\exp \left[ \sum_{i=1}^m r_i \eta_{r,i} + \sum_{i=1}^m c_i \eta_{c,i} + \eta_d \sum_{i=1}^m d_i - \sum_{i=1}^m A_i(\eta_i) \right] h(x),$$

which is a full-rank( $2m + 1$  parameter) exponential family. Therefore, the complete sufficient statistic is

$$T = \left( R_1, \dots, R_m, C_1, \dots, C_m, \sum_{i=1}^m D_i \right),$$

which is a function of  $R_1, \dots, R_m, C_1, \dots, C_m$ .  $\square$

## Problem 5.16

For an  $I \times J$  contingency table with independence, the UMVU estimator of  $p_{ij}$  is  $\hat{p}_{i+}\hat{p}_{+j} = N_{i+}N_{+j}/n^2$ .

- a) Determine the variance of this estimator,  $\text{Var}(\hat{p}_{i+}\hat{p}_{+j})$ .



**Solution.**

Since  $N_{i+}$  and  $N_{+j}$  are independent with

$$N_{i+} \sim \text{Binomial}(n, p_{i+}), \quad N_{+j} \sim \text{Binomial}(n, p_{+j}).$$

$$E\hat{p}_{i+} = \frac{1}{n}EN_{i+} = p_{i+} \text{ and similarly, } E\hat{p}_{+j} = p_{+j}.$$

$$\begin{aligned} E\hat{p}_{i+}^2\hat{p}_{+j}^2 &= \frac{1}{n^4}EN_{i+}^2N_{+j}^2 \\ &= \frac{1}{n^4} [np_{i+}(1-p_{i+}) + n^2p_{i+}^2] [np_{+j}(1-p_{+j}) + n^2p_{+j}^2] \\ &= \frac{1}{n^2} [p_{i+}(1-p_{i+}) + np_{i+}^2] [p_{+j}(1-p_{+j}) + np_{+j}^2]. \end{aligned}$$

$$\begin{aligned} \therefore \text{Var}(\hat{p}_{i+}\hat{p}_{+j}) &= E\hat{p}_{i+}^2\hat{p}_{+j}^2 - [E\hat{p}_{i+}\hat{p}_{+j}]^2 \\ &= \frac{1}{n^2} [p_{i+}(1-p_{i+}) + np_{i+}^2] [p_{+j}(1-p_{+j}) + np_{+j}^2] - p_{i+}^2p_{+j}^2 \\ &= \frac{1}{n^2} [p_{i+}(1-p_{i+})p_{+j}(1-p_{+j}) + np_{i+}(1-p_{i+})p_{+j}^2 + np_{i+}^2p_{+j}(1-p_{+j})] \\ &= \frac{p_{i+}(1-p_{i+})p_{+j}^2 + p_{+j}(1-p_{+j})p_{i+}^2}{n} + \frac{p_{i+}(1-p_{i+})p_{+j}(1-p_{+j})}{n^2}. \end{aligned}$$

□

b) Find the UMVU estimator of the variance in (a).

**Solution.**

First, we should find the unbiased estimators in the numerator term.

$$\begin{aligned} EN_{i+}^2 &= np_{i+}(1-p_{i+}) + n^2p_{i+}^2 \\ &= np_{i+} - np_{i+}^2 + n^2p_{i+}^2 \\ &= EN_{i+} + (n^2 - n)p_{i+}^2 \\ \Rightarrow E\left(\frac{N_{i+}(N_{i+} - 1)}{n^2 - n}\right) &= p_{i+}^2, \end{aligned}$$

$$\begin{aligned} p_{i+}(1-p_{i+}) &= p_{i+} - p_{i+}^2 \\ &= E\frac{N_{i+}}{n} - E\left(\frac{N_{i+}(N_{i+} - 1)}{n^2 - n}\right) \\ &= E\left(\frac{N_{i+}(n - N_{i+})}{n^2 - n}\right), \end{aligned}$$

From the above, we obtain the unbiased estimators as follows:

$$\begin{aligned} E\left(\frac{N_{i+}(N_{i+} - 1)}{n^2 - n}\right) &= p_{i+}^2, \\ E\left(\frac{N_{i+}(n - N_{i+})}{n^2 - n}\right) &= p_{i+}(1-p_{i+}), \\ E\left(\frac{N_{+j}(N_{+j} - 1)}{n^2 - n}\right) &= p_{+j}^2, \end{aligned}$$

$$E \left( \frac{N_{+j}(n - N_{+j})}{n^2 - n} \right) = p_{+j}(1 - p_{+j}).$$

By using these, the expectation of  $\text{Var}(\hat{p}_{i+}\hat{p}_{+j})$  is

$$\begin{aligned} E[\text{Var}(\hat{p}_{i+}\hat{p}_{+j})] &= E \left[ \frac{p_{i+}(1 - p_{i+})p_{+j}^2 + p_{+j}(1 - p_{+j})p_{i+}^2}{n} + \frac{p_{i+}(1 - p_{i+})p_{+j}(1 - p_{+j})}{n^2} \right] \\ &= \frac{N_{i+}(n - N_{i+})N_{+j}(N_{+j} - 1) + N_{+j}(n - N_{+j})N_{i+}(N_{i+} - 1)}{n(n^2 - n)^2} \\ &\quad + \frac{N_{i+}(n - N_{i+})N_{+j}(n - N_{+j})}{n^2(n^2 - n)^2}. \end{aligned} \quad (*)$$

Therefore, (\*) is UMVU estimator of  $\text{Var}(\hat{p}_{i+}\hat{p}_{+j})$ . □

## Chapter 6

# Conditional Distributions

### Problem 6.3

Suppose that  $X$  and  $Y$  are independent and positive. Use a smoothing argument to show that if  $x \in (0, 1)$ , then

$$P\left(\frac{X}{X+Y} \leq x\right) = EF_X\left(\frac{xY}{1-x}\right), \quad (6.8)$$

where  $F_X$  is the cumulative distribution function of  $X$ .

***Solution.***

Since  $X$  and  $Y$  are positive and  $x \in (0, 1)$ ,

$$\frac{X}{X+Y} \leq x \Leftrightarrow X \leq \frac{xY}{1-x}.$$

Then

$$\begin{aligned} P\left(\frac{X}{X+Y} \leq x \middle| Y = y\right) &= P\left(X \leq \frac{xY}{1-x} \middle| Y = y\right) \\ &= E\left[I\left(X \leq \frac{xY}{1-x}\right) \middle| Y = y\right] \\ &= F_X\left(X \leq \frac{xy}{1-x}\right). \end{aligned}$$

Thus we obtain  $P\left(\frac{X}{X+Y} \leq x \middle| Y\right) = F_X\left(X \leq \frac{xY}{1-x}\right)$ , and taking the expectation for the both sides,

$$EP\left(\frac{X}{X+Y} \leq x \middle| Y\right) = P\left(\frac{X}{X+Y} \leq x\right) = EF_X\left(\frac{xY}{1-x}\right).$$

□

### Problem 6.11

Suppose  $X$  and  $Y$  are independent, both absolutely continuous with common density  $f$ . Let  $M = \max\{X, Y\}$  and  $Z = \min\{X, Y\}$ . Determine the conditional distribution for the pair  $(X, Y)$  given  $(M, Z)$ .

**Solution.**

Given  $M = m, Z = z$ , the natural solution can be

$$P(X = m, Y = z | M = m, Z = z) = P(X = z, Y = m | M = m, Z = z) = \frac{1}{2}.$$

To see that the above is true, we check by following procedure.

Let  $g(x, y) = h(x, y, x \vee y, x \wedge y)$  so that  $h(X, Y, M, Z) = g(X, Y)$ . Then

$$\begin{aligned} Eh(X, Y, M, Z | M, Z) &= E[g(X, Y) | M, Z] \\ &= \frac{1}{2}g(M, Z) + \frac{1}{2}g(Z, M). \end{aligned}$$

To show the last equality holds, we should check the smoothing works.

Taking the expectations on the both sides,

$$\begin{aligned} \int \int g(x, y) f(x) f(y) dx dy &= \int \int \frac{1}{2} [g(x \vee y, x \wedge y) + g(x \wedge y, x \vee y)] f(x) f(y) dx dy \\ &= \int \int \frac{1}{2} [g(x, y) + g(y, x)] f(x) f(y) dx dy. \end{aligned}$$

Since  $\int \int g(x, y) dx dy = \int \int g(y, x) dx dy$ , the above equality holds and the smoothing works. Therefore, the conditional distribution is correct.  $\square$

**Problem 6.14**

Let  $X$  and  $Y$  be absolutely continuous with density  $p(x, y) = e^{-x}$ , if  $0 < y < x$ ;  $p(x, y) = 0$ , otherwise.

- a) Find the marginal densities of  $X$  and  $Y$ .

**Solution.**

Take the integration with respect to  $x$  and  $y$ , respectively,

$$\begin{aligned} f(x) &= \int_0^x p(x, y) dy = e^{-x} y \Big|_0^x = x e^{-x}, \quad 0 < x < \infty, \\ f(y) &= \int_0^x p(x, y) dx = -e^{-x} \Big|_0^y = e^{-y}, \quad 0 < y < \infty. \end{aligned}$$

$\square$

- b) Compute  $EY$  and  $EY^2$  integrating against the marginal density of  $Y$ .

**Solution.**

$$\begin{aligned} EY &= \int_0^\infty y e^{-y} dy = 1 \quad (\because \text{the density of } \Gamma(2, 1)) \\ EY^2 &= \int_0^\infty y \cdot y e^{-y} dy = 2 \cdot 1 = 2 \end{aligned}$$

$\square$

- c) Find the conditional density of  $Y$  given  $X = x$ , and use it to compute  $E[Y|X]$  and  $E[Y^2|X]$ .

**Solution.**

The conditional density is  $p(y|x) = p(y, x)/p(x) = 1/x$ ,  $0 < y < x$ . Then

$$E[Y|X = x] = \int_0^x \frac{1}{x} y dy = \frac{1}{x} \cdot \frac{1}{2} y^2 \Big|_0^x = \frac{x}{2},$$

$$E[Y^2|X = x] = \int_0^x \frac{1}{x} y^2 dy = \frac{x^2}{3}.$$

Therefore,  $E[Y|X] = \frac{X}{2}$  and  $E[Y^2|X] = \frac{X^2}{3}$ . □

- d) Find the expectations of  $E[Y|X]$  and  $E[Y^2|X]$  integrating against the marginal density of  $X$ .

**Solution.**

By using the fact of  $\Gamma(2, 1)$ ,

$$EE[Y|X] = E\left[\frac{X}{2}\right] = \frac{1}{2} \int_0^\infty x \cdot x e^{-x} dx = \frac{1}{2} \cdot 2 = 1,$$

$$EE[Y^2|X] = E\left[\frac{X^2}{3}\right] = \frac{1}{3} \int_0^\infty x^2 \cdot x e^{-x} dx = \frac{1}{3} (2 + 2^2) = 2.$$

□

## Chapter 7

# Bayesian Estimation

### Problem 7.6

Consider a Bayesian model in which the random parameter  $\Theta$  has a Bernoulli prior distribution with success probability  $1/2$ , so  $P(\Theta = 0) = P(\Theta = 1) = 1/2$ . Given  $\Theta = 0$ , data  $X$  has density  $f_0$ , and given  $\Theta = 1$ ,  $X$  has density  $f_1$ .

- a) Find the Bayes estimate of  $\Theta$  under squared error loss.

**Solution.**

First, we should find the posterior distribution. Since  $\Theta \sim \text{Bernoulli}(1/2)$  and  $X|\Theta = \theta \sim f_\theta$ , the joint density is

$$p(\theta, x) = \lambda(\theta)p_\theta(x) = \frac{1}{2}f_\theta(x),$$

the marginal density is

$$q(x) = \sum_{\theta} p(\theta, x) = \frac{f_0(x) + f_1(x)}{2},$$

and the posterior density is

$$\begin{aligned}\lambda(\theta|x) &= \frac{\lambda(\theta)p_\theta(x)}{q(x)} = \frac{f_\theta(x)}{f_0(x) + f_1(x)}, \quad \theta = 0, 1 \\ &\sim \text{Bernoulli}\left(\frac{f_1(x)}{f_0(x) + f_1(x)}\right).\end{aligned}$$

Note that the Bayes estimator under the squared loss is the conditional mean (i.e. posterior mean) is

$$\delta_\Lambda = E[\Theta|X] = \frac{f_1(X)}{f_0(X) + f_1(X)}$$

□

- b) Find the Bayes estimate of  $\Theta$  if  $L(\theta, d) = I\{\theta \neq d\}$  (called *zero-one loss*).

**Solution.**

Define the success probability of posterior distribution  $p = \frac{f_1(x)}{f_0(x) + f_1(x)}$ . Under the zero-one loss, the estimator minimizing the loss be as

$$\begin{cases} 1, & p > \frac{1}{2}, \\ 0, & p < \frac{1}{2}, \end{cases}$$

and when  $p = 1/2$ , either is optimal.

Therefore, the Bayes estimator under the zero-one loss is

$$\delta_\Lambda = \begin{cases} 1, & f_1(X) > f_0(X), \\ 0, & f_1(X) < f_0(X). \end{cases}$$

□

## Problem 7.7

Consider Bayesian estimation in which the parameter  $\Theta$  has a standard exponential distribution, so  $\lambda(\theta) = e^{-\theta}$ ,  $\theta > 0$ , and given  $\Theta = \theta$ ,  $X_1, \dots, X_n$  are i.i.d. from an exponential distribution with failure rate  $\theta$ . Determine the Bayes estimator of  $\Theta$  if the loss function is  $L(\theta, d) = (d - \theta)^2/d$ .

### Solution.

By Theorem 7.1, the Bayes estimator should be minimizer of the conditional expected loss as

$$E \left[ \frac{(d - \Theta)^2}{d} \middle| X = x \right] = d - 2E[\Theta | X = x] + \frac{E[\Theta^2 | X = x]}{d}.$$

Taking the derivative,

$$1 - \frac{E[\Theta^2 | X = x]}{d^2} = 0 \Rightarrow \delta_\Lambda = \sqrt{E[\Theta^2 | X]}.$$

Since  $X_i | \Theta = \theta$  are i.i.d.  $\text{Exp}(\theta)$ , the posterior density is

$$\lambda(\theta | x) \propto \lambda(\theta) p(x_1, \dots, x_n | \theta) \propto \theta^n e^{-(\sum_i x_i + 1)\theta},$$

which is the density of  $\Gamma \left( n + 1, \frac{1}{\sum_i x_i + 1} \right)$ .

Using the expectation and variance formulas of Gamma distribution, the 2nd moment of the posterior distribution can be obtained as

$$E[\Theta^2 | X] = \frac{n + 1}{(\sum_i X_i + 1)^2} + \left( \frac{n + 1}{\sum_i X_i + 1} \right)^2 = \frac{(n + 1)(n + 2)}{(\sum_i X_i + 1)^2}.$$

Thus, the Bayes estimator is

$$\delta_\Lambda = \sqrt{\frac{(n + 1)(n + 2)}{(\sum_i X_i + 1)^2}}.$$

□

## Chapter 8

# Large-Sample Theory

### Problem 8.1

Random variables  $X_1, X_2, \dots$  are called "m-dependent" if  $X_i$  and  $X_j$  are independent whenever  $|i - j| \geq m$ . Suppose  $X_1, X_2, \dots$  are m-dependent with  $EX_1 = EX_2 = \dots = \xi$  and  $\text{Var}(X_1) = \text{Var}(X_2) = \dots = \sigma^2 < \infty$ . Let  $\bar{X}_n = (X_1 + \dots + X_n)/n$ . Show that  $\bar{X}_n \xrightarrow{P} \xi$  as  $n \rightarrow \infty$ .

Hint: You should be able to bound  $\text{Cov}(X_i, X_j)$  and  $\text{Var}(\bar{X}_n)$ .

#### Solution.

By Chebyshev's inequality,

$$P(|\bar{X}_n - \xi| > \epsilon) \leq \frac{E|\bar{X}_n - \xi|^2}{\epsilon^2},$$

for every  $\epsilon > 0$ . Thus, our claim is  $E|\bar{X}_n - \xi|^2 \rightarrow 0$  as  $n \rightarrow \infty$ .

By the covariance inequality (See (4.11)),  $|\text{Cov}(X_i, X_j)| \leq \sqrt{\text{Var}(X_i)}\sqrt{\text{Var}(X_j)} = \sigma^2$ . Since  $X_1, X_2, \dots$  are m-dependent, for any  $i$ ,

$$\sum_{j=1}^n \text{Cov}(X_i, X_j) = \sum_{j: |i-j| < m} \text{Cov}(X_i, X_j) \leq (2m-1)\sigma^2.$$

By using this, we can obtain

$$\begin{aligned} E|\bar{X}_n - \xi|^2 &= \text{Var}(\bar{X}_n) = \frac{1}{n^2} \text{Var}(X_1 + \dots + X_n) = \frac{1}{n^2} \sum_i \sum_j \text{Cov}(X_i, X_j) \\ &\leq \frac{1}{n^2} \sum_i (2m-1)\sigma^2 \\ &= \frac{(2m-1)\sigma^2}{n} \\ &\rightarrow 0 \text{ as } n \rightarrow \infty. \end{aligned}$$

Therefore, the bound of Chebyshev's inequality goes to 0 as  $n \rightarrow \infty$ , and we obtain  $\bar{X}_n \xrightarrow{P} \xi$ . □

### Problem 8.5

Let  $X_1, \dots, X_n$  be i.i.d. with common density  $f_\theta(x) = (x - \theta)^+ e^{\theta - x}$ . Show that  $M_n = \min\{X_1, \dots, X_n\}$  is a consistent estimator of  $\theta$ , and determine the limiting distribution for  $\sqrt{n}(M_n - \theta)$ .



**Solution.**(i) ConsistencyFor  $m > \theta$ ,

$$\begin{aligned}
P(X_i > m) &= \int_m^\infty (x - \theta)e^{\theta - x} dx = \int_{m-\theta}^\infty te^{-t} dt \\
&= -te^{-t} \Big|_{m-\theta}^\infty + \int_{m-\theta}^\infty e^{-t} dt \\
&= (m - \theta)e^{-(m-\theta)} + e^{-(m-\theta)} \\
&= (m - \theta + 1)e^{-(m-\theta)}.
\end{aligned}$$

Then,

$$P(M_n > m) = P(\min\{X_1, \dots, X_n\} > m) = [P(X_1 > m)]^n = [(m - \theta + 1)e^{-(m-\theta)}]^n.$$

Now, plug in  $m = \theta + \epsilon > \theta$ ,

$$P(M_n - \theta > \epsilon) = [(1 + \epsilon)e^{-\epsilon}]^n \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Therefore  $M_n$  is a consistent estimator of  $\theta$ .(ii) Limiting distributionFrom (i), let  $\epsilon = x/\sqrt{n}$ ,

$$P\left(M_n - \theta > \frac{x}{\sqrt{n}}\right) = \left[\left(1 + \frac{x}{\sqrt{n}}\right)e^{-\frac{x}{\sqrt{n}}}\right]^n.$$

To derive the limit of RHS, by using the fact that  $[(1+u)^2 e^{-u}]/u^2 \xrightarrow{n \rightarrow \infty} -1/2$  (It can be evaluated by L'Hôpital's rule.), then

$$n \left[ \left(1 + \frac{x}{\sqrt{n}}\right)e^{-\frac{x}{\sqrt{n}}} - 1 \right] = x^2 \frac{(1 + x/\sqrt{n})e^{-x/\sqrt{n}} - 1}{(x/\sqrt{n})^2} \rightarrow -\frac{x^2}{2} \text{ as } n \rightarrow \infty.$$

By using this,

$$P(\sqrt{n}(M_n - \theta) > x) = \left[\left(1 + \frac{x}{\sqrt{n}}\right)e^{-\frac{x}{\sqrt{n}}}\right]^n \approx \left[1 + \frac{-x^2/2}{n}\right]^n \rightarrow e^{-x^2/2} \text{ as } n \rightarrow \infty.$$

Therefore  $\sqrt{n}(M_n - \theta) \Rightarrow Y$  where  $Y$  has the cumulative distribution function  $P(Y \leq y) = 1 - e^{-y^2/2}$ ,  $y > 0$ . (Weibull distribution)

□

**Problem 8.6**Prove that if  $A_n \xrightarrow{P} 1$  and  $Y_n \Rightarrow Y$ , then  $A_n Y_n \Rightarrow Y$ . (This is a special case of Theorem 8.13.)**Solution.**Denote  $F$  be the *cdf* of  $Y$ . Let  $y$  be continuity point of  $F$ , and we only prove when  $y > 0$ . If  $y + \epsilon$  also continuity point of  $F$ , then

$$\{A_n Y_n \leq y\} \subset \{Y_n \leq y + \epsilon\} \cup \{A_n \leq y/(y + \epsilon)\}$$

$$P(A_n Y_n \leq y) \leq P(Y_n \leq y + \epsilon) + P(A_n \leq y/(y + \epsilon)) \rightarrow F(y + \epsilon) \text{ as } n \rightarrow \infty.$$

Then, the limsup of LHS have the following bound,

$$\limsup_n P(A_n Y_n \leq y) \leq F(y + \epsilon).$$

Since  $F$  is continuous at  $y$  and  $\epsilon$  is arbitrary, by letting  $\epsilon \rightarrow 0$ , we can obtain

$$\limsup_n P(A_n Y_n \leq y) \leq F(y).$$

Next, if  $y - \epsilon$  be positive and continuous point of  $F$ , then

$$P(A_n Y_n \leq y) \geq P(Y_n \leq y - \epsilon) - P(A_n < 0 \text{ or } A_n \geq y/(y - \epsilon)) \rightarrow F(y - \epsilon) \text{ as } n \rightarrow \infty.$$

(RHS를 천천히 생각해보면 이해할 수 있음!!! 전체 경우에서 특정 케이스를 제외하는 거라고 생각하자!!)

Then, the liminf of LHS have the following bound,

$$\liminf_n P(A_n Y_n \leq y) \geq F(y - \epsilon).$$

And then letting  $\epsilon \rightarrow 0$ , we can obtain

$$\liminf_n P(A_n Y_n \leq y) \geq F(y).$$

Finally, we obtain

$$F(y) \leq \liminf_n P(A_n Y_n \leq y) \leq \limsup_n P(A_n Y_n \leq y) \leq F(y),$$

thus  $\lim_n P(A_n Y_n \leq y) = F(y)$  which implies  $A_n Y_n \Rightarrow Y$  when  $y > 0$ .

When  $y < 0$  and  $y = 0$ , it can show using similar arguments. □

## Problem 8.24

Suppose  $X_1, \dots, X_n$  are i.i.d.  $N(0, \sigma^2)$ , and let  $M$  be the median of  $|X_1|, \dots, |X_n|$ .

- a) Find  $c \in \mathbb{R}$  so that  $\tilde{\sigma} = cM$  is a consistent estimator of  $\sigma$ .

### ***Solution.***

The cumulative distribution function of  $|X_i|$  is

$$P(|X_i| \leq x) = \Phi\left(\frac{x}{\sigma}\right) - \Phi\left(-\frac{x}{\sigma}\right) = 2\Phi\left(\frac{x}{\sigma}\right) - 1,$$

and then its density is  $\frac{2}{\sigma}\phi(x/\sigma)$ .

From the above *cdf*, the median of the random variable  $|X_i|$  is obtained as

$$2\Phi\left(\frac{x}{\sigma}\right) - 1 = \frac{1}{2} \Leftrightarrow \Phi\left(\frac{x}{\sigma}\right) = \frac{3}{4}$$

$$\therefore \text{ the median of } |X_i| = \Phi^{-1}\left(\frac{3}{4}\right) \cdot \sigma = 0.6745\sigma.$$

Since  $\tilde{\sigma} = cM$  ( $M$  is the sample median) is consistent for  $\sigma$ , it should be satisfy

$$\tilde{\sigma} = cM \xrightarrow{P} c \cdot 0.6745\sigma.$$

Therefore  $c = 1/0.6745 = 1.4826$ . □

- b) Determine the limiting distribution for  $\sqrt{n}(\tilde{\sigma} - \sigma)$ .

**Solution.**

By (8.5),

$$\begin{aligned}\sqrt{n}(M - 0.6745\sigma) &\Rightarrow N\left(0, \frac{1}{4 \cdot \left[\frac{2}{\sigma}\phi(0.6745\sigma)\right]^2}\right) = N(0, 0.6189\sigma^2). \\ \therefore \sqrt{n}(\tilde{\sigma} - \sigma) &\Rightarrow N(0, 1.4826^2 \cdot 0.6189\sigma^2) = N(0, 1.3604\sigma^2).\end{aligned}$$

□

- c) Find the maximum likelihood estimator  $\hat{\sigma}$  of  $\sigma$  and determine the limiting distribution for  $\sqrt{n}(\hat{\sigma} - \sigma)$ .

**Solution.**

The likelihood and log-likelihood function with respect to  $\sigma$  is

$$L(\sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\sum_i x_i^2}{2\sigma^2}\right), \quad l(\sigma) = -\frac{1}{2\sigma^2} \sum_i x_i^2 - n \log(\sqrt{2\pi}\sigma).$$

Taking a derivative to 0,

$$\frac{\partial l(\sigma)}{\partial \sigma} = \sigma^{-3} \sum_i x_i^2 - \frac{n}{\sigma} = 0.$$

Therefore, the MLE of  $\sigma$  is  $\sqrt{\sum_i x_i^2/n}$ . (Also, the MLE of  $\sigma^2$  is  $\sum_i x_i^2/n$  by the similar way.)

To derive the limiting distribution, we first evaluate the variance of  $\hat{\sigma}^2$ . By using the fact that  $EX_i^4 = 3\sigma^4$  (It can be evaluated integrating by parts),

$$\text{Var}(\hat{\sigma}^2) = \frac{1}{n^2} n \text{Var}(X_i^2) = \frac{1}{n} \left( EX_i^4 - [EX_i^2]^2 \right) = \frac{1}{n} [3\sigma^4 - (\sigma^2)^2] = \frac{2\sigma^4}{n}.$$

Thus

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \Rightarrow N(0, 2\sigma^4).$$

By the delta method with  $g(\theta) = \sqrt{\theta}$ ,

$$\sqrt{n}(\hat{\sigma} - \sigma) \Rightarrow N\left(0, \left(\frac{1}{2\sqrt{\sigma^2}}\right)^2 \cdot 2\sigma^4\right) = N\left(0, \frac{\sigma^2}{2}\right).$$

□

- d) Determine the asymptotic relative efficiency of  $\tilde{\sigma}$  with respect to  $\hat{\sigma}$ .

**Solution.**

The asymptotic relative efficiency is

$$ARE = \frac{\text{Var}(\hat{\sigma})}{\text{Var}(\tilde{\sigma})} = \frac{\sigma^2/2}{1.3604\sigma^2} = 0.3675.$$

□

## Chapter 9

# Estimating Equations and Maximum Likelihood