STAT3602 Statistical Inference

Example Class 2

LIU Chen

Department of Statistics and Actuarial Science, HKU

07/09/2020

Outline

1. Review: Chapter 1 - Decision Problem: Frequentist Approach
2. Exercise 1: When Harry met Sally – *a story about fate* (緣份)
3. Exercise 2: Zixia (紫霞) and Joker (至尊寶)

Parameter space Θ

- collection of k unknown parameters, usually a subset of $\mathbb{R}^k$.
- The true parameter is some **unknown** $\theta \in \Theta$

Sample space S

collection of all possible realisations x of a random vector X.

Action space $\mathcal{A}$

collection of all possible actions under consideration

Review: Decision Problem: Frequentist Approach

Statistical Model

- a link between Θ and S
- a family of probability functions $\{f(\cdot|\theta) : \theta \in \Theta\}$

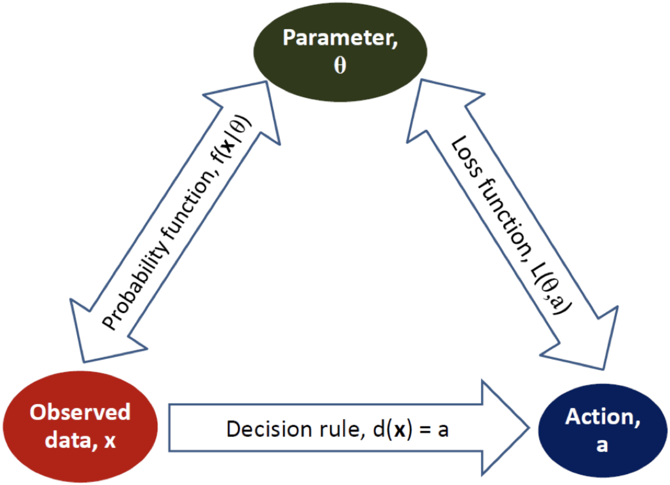
Loss function $L : \Theta \times \mathcal{A} \to \mathbb{R}$

$L(\theta, a)$ is the *loss* incurred by taking action a when θ is the true parameter

Decision rule $d : S \to A$

an action, $d(x)$, to be taken when X is observed to be $x \in S$

Review: Formulation



Review: Frequentist Approach: Risk Function

The risk function of the decision rule $d(\cdot)$ is the expected loss incurred by adopting decision rule d under each possible $\theta \in \Theta$, i.e.

$$R(\theta, d) = \mathbb{E}_\theta [L(\theta, d(x))] \tag{1}$$

- if X is continuous,

$$R(\theta, d) = \int_S L(\theta, d(x)) f(x|\theta) \, dx; \tag{2}$$

- if X is discrete,

$$R(\theta, d) = \sum_{x \in S} L(\theta, d(x)) f(x|\theta) . \tag{3}$$

Review: Admissibility

Definition. A rule d *strictly dominates* another rule d^* if

$$R(\theta, d) \leq R(\theta, d^*) \text{ for all } \theta \in \Theta \quad \text{and} \quad R(\theta', d) < R(\theta', d^*) \text{ for some } \theta' \in \Theta.$$

If d strictly dominates d^*, then obviously d is the better choice.

Definition. A rule strictly dominated by another rule is *inadmissible*.

Definition. If d is not inadmissible, then it is said to be *admissible*.

Review: Minimaxity

Definition. A rule d is *minimax* if, for all possible rules d',

$$\sup\{R(\theta, d') : \theta \in \Theta\} \geq \sup\{R(\theta, d) : \theta \in \Theta\}.$$

Review: Unbiased rule

Definition. A rule d is *unbiased* if

$$\boxed{\underbrace{\mathbb{E}_\theta L(\theta', d(\boldsymbol{X}))}_{\text{(expected loss w.r.t. false }\theta)} \geq \underbrace{R(\theta, d)}_{\text{(expected loss w.r.t. true }\theta)} \quad \forall\, \theta, \theta' \in \Theta}$$

Outline

1. Review: Chapter 1 - Decision Problem: Frequentist Approach
2. Exercise 1: When Harry met Sally – *a story about fate* (緣份)
3. Exercise 2: Zixia (紫霞) and Joker (至尊寶)

When Harry met Sally: Background I

The year is 2019. You are Sally, a flashy schoolgirl and a frequent moviegoer. One day you are standing in a long queue, waiting to buy a ticket for a popular film, which is 120 minutes long and will start immediately. Your experience tells you that you need to wait $X \in (0, 120)$ minutes before you can successfully buy a ticket, where X is a random variable with the density function

$$f(x|\theta) = \frac{1}{120\,\theta}\left(1 - \frac{x}{120}\right)^{(1-\theta)/\theta}, \quad x \in (0, 120),$$

for some unknown parameter $\theta \in (0, 1)$. Thus, you will miss the first X minutes of the film if you stay in the queue.

At this juncture, your schoolmate Harry suddenly shows up with a ticket in his hands.

When Harry met Sally: Background II

"Take my ticket and go watch the film," said Harry. "It's only 80 dollars."

The normal price of the ticket is 60 dollars.

To help yourself decide whether to buy Harry's ticket, you calculate a loss function as follows.

- Suppose that each minute of the film is worth 0.5 dollar. If you stay in the queue, you will lose $60 - 0.5(120 - X) = 0.5X$ dollars.
- If you buy Harry's ticket, you will lose $80 - 0.5(120) = 20$ dollars.

Your loss function $L(\theta, a)$ can be defined as $\mathbb{E}_\theta[0.5X]$ if you take the action $a = a_0$, staying in the queue, and the constant 20 if you take the action $a = a_1$, buy Harry's ticket.

Let $X_1, \ldots, X_n$ be independent observations following the common density function $f(x|\theta)$. You may rely on the data $(X_1, \ldots, X_n)$ for your decision.

When Harry met Sally: Answer the Following Question I.

Write down an expression for the risk function $R(\theta, d)$ of a decision rule d. Hence show that a minimax rule is to always buy Harry's ticket.

When Harry met Sally: Answer the Following Question I.

Write down an expression for the risk function $R(\theta, d)$ of a decision rule d. Hence show that a minimax rule is to always buy Harry's ticket.

Note that $L(\theta, a_1) \equiv 20$ and

$$
\begin{aligned}
L(\theta, a_0) &= 0.5 \int_0^{120} x\, f(x|\theta)\, dx \\
&= \frac{0.5}{120\,\theta} \int_0^{120} x \left(1 - \frac{x}{120}\right)^{(1-\theta)/\theta} dx \\
&= \frac{60\,\theta}{\theta + 1}.
\end{aligned}
$$

When Harry met Sally: Answer the Following Question II.

Write down an expression for the risk function $R(\theta, d)$ of a decision rule d. Hence show that a minimax rule is to always buy Harry's ticket.

Risk function:

$$
\begin{aligned}
R(\theta, d) &= L(\theta, a_0)\,\mathbb{P}_\theta\big(d(X_1, \ldots, X_n) = a_0\big) + L(\theta, a_1)\,\mathbb{P}_\theta\big(d(X_1, \ldots, X_n) = a_1\big) \\
&= \frac{20(2\,\theta - 1)}{\theta + 1}\,\mathbb{P}_\theta\big(d(X_1, \ldots, X_n) = a_0\big) + 20.
\end{aligned}
$$

Write down an expression for the risk function $R(\theta, d)$ of a decision rule d. Hence show that a minimax rule is to always buy Harry's ticket.

Risk function:

$$
\begin{aligned}
R(\theta, d) &= L(\theta, a_0)\, \mathbb{P}_\theta\big(d(X_1, \dots, X_n) = a_0\big) + L(\theta, a_1)\, \mathbb{P}_\theta\big(d(X_1, \dots, X_n) = a_1\big) \\
&= \frac{20(2\,\theta - 1)}{\theta + 1}\, \mathbb{P}_\theta\big(d(X_1, \dots, X_n) = a_0\big) + 20.
\end{aligned}
$$

Let $d_1(x_1, \dots, x_n) \equiv a_1$ be the rule that always buys Harry's ticket, so that $R(\theta, d_1) \equiv 20$.
Then for any rule d,

$$
\sup_{\theta \in (0,1)} R(\theta, d) \geq \sup_{\theta > 1/2} R(\theta, d) \geq 20 = \sup_{\theta \in (0,1)} R(\theta, d_1),
$$

which implies d_1 is minimax.

Outline

1. Review: Chapter 1 - Decision Problem: Frequentist Approach
2. Exercise 1: When Harry met Sally – *a story about fate* (緣份)
3. Exercise 2: Zixia (紫霞) and Joker (至尊寶)

Zixia (紫霞) and Joker (至尊寶): Background

Figure: Zixia and Joker in A Chinese Odyssey Part 1-Pandora's Box(西遊記第壹佰零壹回之月光寶盒).

Zixia (紫霞) and Joker (至尊寶): Background I

Zixia (紫霞) suspects that Joker (至尊寶), a man whom she loves, is having an affair with another woman. She has got so furious that she draws her sword against Joker's throat at a speed of X cm per second, threatening to kill him.

Zixia's love for Joker is indexed by an unknown parameter $\theta \in (0, \infty)$. The greater the value of θ, the deeper is her love for him. It is known that X is a random variable with the density function

$$f(x|\theta) = \theta (1 + x)^{-\theta-1}, x \in (0, \infty). \tag{4}$$

Based on the observed value of X, Joker needs to decide quickly whether to tell Zixia a lie or not. If he keeps his mouth shut, he will be hurt by the sword, so that his body will suffer a loss equal to $1/\theta$, but no harm will be done to his soul. If he tells a lie, he can reduce his bodily loss by one half but his soul will suffer a loss of one unit, so that the total loss becomes $1/(2\theta) + 1$.

Zixia (紫霞) and Joker (至尊寶): Answer the Questions!

1. Define the loss function for Joker's decision problem.
2. Show that the risk function of a decision rule d is given by

$$R(\theta, d) = \frac{1}{\theta} + \left(1 - \frac{1}{2\theta}\right) \mathbb{P}_\theta(d(X) = a), \tag{5}$$

   for a particular action a. Specify the action rule a.
3. Deduce from b that $\sup_{\theta>0} R(\theta, d) = \infty$ for any decision rule d.
4. For $\forall c > 0$, define the decision rule d_c by

$$d_c(x) = \begin{cases} \text{“to lie”}, & x > c; \\ \text{“not to lie”}, & x \leq c. \end{cases}$$

   1. let c_1, c_2 be **distinct** positive constants. Show that d_{c_1} does not strictly dominate d_{c_2}.
   2. Is d_c an unbiased decision rule?

Zixia (紫霞) and Joker (至尊寶): Answer the Questions!

1. Define the loss function for Joker's decision problem.

Action space $\mathcal{A}$

$$\mathcal{A} = \{\text{"to lie"}, \text{"not to lie"}\} \tag{6}$$

Loss function $L : \Theta \times \mathcal{A} \to \mathbb{R}$

$$L(\theta, \text{"to lie"}) = \frac{1}{2\theta} + 1;$$

$$L(\theta, \text{"not to lie"}) = \frac{1}{\theta}.$$

2. Show that the risk function of a decision rule d is given by

$$R(\theta, d) = \frac{1}{\theta} + \left(1 - \frac{1}{2\theta}\right) \mathbb{P}_\theta\left(d(X) = a\right), \tag{7}$$

for a particular action a. Specify the action rule a.

$$\begin{aligned}
R(\theta, d) &= \mathbb{E}_\theta\left[L(\theta, d(X))\right] \\
&= \mathbb{P}_\theta\left(d(X) = \text{“to lie”}\right)\left[\frac{1}{2\theta} + 1\right] + \mathbb{P}_\theta\left(d(X) = \text{“not to lie”}\right)\left[\frac{1}{\theta}\right] \\
&= \mathbb{P}_\theta\left(d(X) = \text{“to lie”}\right)\left[\frac{1}{2\theta} + 1\right] + \left[1 - \mathbb{P}_\theta\left(d(X) = \text{“to lie”}\right)\right]\left[\frac{1}{\theta}\right] \\
&= \frac{1}{\theta} + \left(1 - \frac{1}{2\theta}\right) \mathbb{P}_\theta\left(d(X) = a\right).
\end{aligned}$$

Thus, the action a is “to lie”.

3. Deduce from b that $\sup_{\theta>0} R(\theta, d) = \infty$ for any decision rule d.

$$R(\theta, d) = \sum_{x \in S} L(\theta, d(x)) f(x|\theta). \tag{8}$$

$$\begin{aligned} R(\theta, d) &= \frac{1}{\theta}\left[1 - \frac{1}{2}\mathbb{P}_\theta(d(X) = \text{“to lie”})\right] + \mathbb{P}_\theta(d(X) = \text{“to lie”}) \\ &\geq \frac{1}{2\theta} \to \infty \text{ as } \theta \to 0 \\ \Longrightarrow & \sup_{\theta>0} R(\theta, d) = \infty \end{aligned}$$

There is no unique minimax rule, because all decision rules share the same maximum risk, that is infinity, so taht they are all minimax.

Zixia (紫霞) and Joker (至尊寶): Answer the Questions!

4. For $\forall c > 0$, define the decision rule d_c by

$$d_c(x) = \begin{cases} \text{“to lie”}, & x > c; \\ \text{“not to lie”}, & x \leq c. \end{cases}$$

let c_1, c_2 be **distinct** positive constants. Show that d_{c_1} does not strictly dominate d_{c_2}.

$$\begin{aligned} R(\theta, d_{c_1}) - R(\theta, d_{c_2}) &= \left(1 - \frac{1}{2\theta}\right)\{\mathbb{P}_\theta(X > c_1) - \mathbb{P}_\theta(X > c_2)\} \\ &= \left(1 - \frac{1}{2\theta}\right)\left\{\int_{c_1}^{\infty} \theta(1+x)^{-\theta-1}dx - \int_{c_2}^{\infty} \theta(1+x)^{-\theta-1}dx\right\} \\ &= \left(1 - \frac{1}{2\theta}\right)\left\{(1+c_1)^{-\theta} - (1+c_2)^{-\theta}\right\} \end{aligned}$$

which has different signs under $\theta < 1/2$ and $\theta > 1/2$ respectively. Thus, d_{c_1} does not strictly dominate d_{c_2}

4. For $\forall c > 0$, define the decision rule d_c by

$$d_c(x) = \begin{cases} \text{“to lie”}, & x > c; \\ \text{“not to lie”}, & x \leq c. \end{cases}$$

Is d_c an unbiased decision rule?

$$\mathbb{E}_\theta \left[L\left(\theta', d(X)\right) \right] - R(\theta, d)$$

$$= \frac{1}{\theta'} + \left(1 - \frac{1}{2\theta'}\right)(1+c)^{-\theta} - \frac{1}{\theta} - \left(1 - \frac{1}{2\theta}\right)(1+c)^{-\theta}$$

$$= \left(\frac{1}{\theta'} - \frac{1}{\theta}\right)\left[1 - \frac{(1+c)^{-\theta}}{2}\right] \begin{cases} > 0, \theta' < \theta, \\ < 0, \theta' > \theta; \end{cases}$$

so that d_c is not unbiased.

Practice more!

Past paper:

- 2014. Question 1, (c) (just use the result of (b))
- 2013. Question 1, (a) – (c)