THE UNIVERSITY OF HONG KONG
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE

STAT3602 Statistical Inference
(2020-2021 First Semester)

Example Class 5 [Midterm Exam Review]

Date: 19/10/2020

1 Midterm Exam Review: Frequentist and Bayesian Approach

1.1 Chapter 1: Frequentist Approach

Frequentist Approach: Formulation

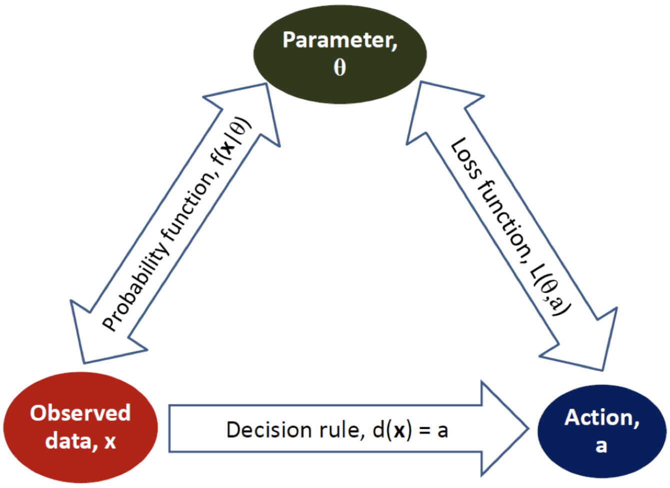

Figure 1: Frequentist Approach: Formulation

- Formulation and Risk Function

Criteria for a good decision rule

- Admissibility
- Minimaxity
- Unbiased rule

Randomised decision rules

1.2 Chapter 2: Bayesian Approach

Bayesian Decision

- Prior and Posterior Probability Function.
- Expected Posterior Loss
- Bayesian Decision

Bayesian Statistical Inference

- Point Estimation
- Hypothesis Testing about θ
- Interval Estimation of θ:

- Given fixed interval length, maximise posterior coverage probability.
- Given fixed posterior coverage probability, minimise interval length.
- Given fixed posterior coverage probability, require "equal-tailed" inteval.

- Predictive Distribution

Exercise 1

Papet, an old greedy man living in village A, wants to purchase a plot of land in his village. The land is currently owned by Jean, a hunchback recently arriving from Village B, who inherits the land from his deceased mother Florette, Papet's ex-lover.

The price of the land depends on the availability of water. Papet knows that there is a water source 1 mile from the land, and is considering to block it.

If Papet does not block the source (action a_0), then Jean will have access to water from this source and will sell his land at a price of 1 unit. If Papet blocks the source (action a_1), then Jean will need to search for an alternative water source, at an unknown distance of θ miles away, and will sell his land at a price of $e^{1-\theta}$ units. The following questions concern Papet's decision problem of whether to block the water source.

(a). Papet's loss function $L(\theta, a)$ can be defined as the price of the land. Write down the expressions for $L(\theta, a_0)$ and $L(\theta, a_1)$, for any $\theta > 0$

(b). Papet tries to make his decision based on an observation X, which is the growth rate of a plant found near the land plot. It is known that X is distributed under the density function

$$f(x \mid \theta) = \theta e^{-\theta x}, \quad x > 0.$$

Show that the risk function of Papet's decision rule d is given by

$$R(\theta, d) = \mathbb{P}_\theta(d(X) = a_1)\left(e^{1-\theta} - 1\right) + 1, \quad \theta > 0$$

(c). Find a minimax decision rule for Papet.

(d). Suppose that Papet has a prior belief about θ, represented by the prior density function

$$\pi(\theta) = ce^{-c\theta}, \quad \theta > 0,$$

for some fixed constant $c > 0$. Show that Papet's Bayes rule is to block the water source if and only if

$$c + X < \frac{1}{\sqrt{e} - 1}$$

[**You may find the following identity useful:** $\int_0^\infty \theta e^{-k\theta} d\theta = k^{-2}$ **for any** $k > 0$**.**]

Exercise 2

Each customer buys one of the K products in a shop, buying product i with probability θ_i. Suppose there are n customers. X_i of them buy product i. $\mathbf{X} = (X_1, \ldots, X_K)$ follows multinomial distribution with probability mass function

$$f(x \mid \theta) = \frac{n!}{x_1! \ldots x_K!}\theta_1^{x_1} \ldots \theta_K^{x_K}.$$

(Mean for X_i is $n\theta_i$).

Our prior distribution for $\theta = (\theta_1, \ldots, \theta_K)$ is $\theta \sim$ Dirichlet $(\alpha_1, \ldots, \alpha_K)$

$$\pi(\theta) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_K)}{\Gamma(\alpha_1) \ldots \Gamma(\alpha_K)}\theta_1^{\alpha_1 - 1} \ldots \theta_K^{\alpha_K - 1}.$$

(a). Find the posterior distribution of $\theta \mid x$

(b). Suppose there are m customers independent of the first n customers. Y_i of the m customers buy product i. Find the mean of the posterior predictive distribution of $Y_i \mid \mathbf{x}, i = 1, \ldots, K$

Exercise 3

Let θ be an unknown parameter in $[0, 1]$, representing the pollution index of a city. In a sample of air collected from the city, the content of a particular pollutant is observed to be $100X\%$, where $X \in [0, 1]$ is a random variable with the density function

$$f(x \mid \theta) = 1 + 2\theta \left(x - \frac{1}{2} \right), \quad x \in [0, 1]$$

Before the measurement of X, non-informative prior distribution is given to θ, so that θ is uniformly distribution over [0,1]

(a). Plot the functions $f(x \mid 0)$ and $f(x \mid 1)$ against $x \in [0, 1]$ on the same graph. Comment briefly on the likely magnitude of the observation X if the pollution index θ is high in the city.

(b). Show that the posterior distribution of θ given $X = x$ has the density function

$$\pi(\theta \mid x) = \left(x + \frac{1}{2} \right)^{-1} (1 + (2x - 1)\theta), \quad \theta \in [0, 1]$$

Describe briefly how the observation x affected our posterior belief about θ in the two cases where $x < 1/2$ and $x > 1/2$

(c). Based on the observation $X = x$, find the Bayes interval for θ which has length 1/3 and the maximum posterior coverage probability if

(i) $x < 1/2$

(ii) $x > 1/2$

(d). If a second sample of air is to be collected from the city independently of the first sample, and its pollutant content observed to be $100Y\%$, show that the posterior predictive distribution of Y given $X = x$ has the density function

$$g^*(y \mid x) = \frac{1}{3} \left(x + \frac{1}{2} \right)^{-1} (x + 1 + (4x + 1)y), \quad y \in [0, 1].$$

(e). If the first sample is not polluted, that is $X = 0$, predict the pollutant content Y of the second sample using the mean of the posterior predictive distribution g^* given in (d) above.