Bayesian Look Ahead Sampling Methods to Allocate

Up to M Observations Among k Populations

BY

Yanmin Liu
B.S., Hebei Normal University, 2002
M.S., Beijing Institute of Technology, 2005

THESIS

Submitted as partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Mathematics
in the Graduate College of the
University of Illinois at Chicago, 2011

Chicago, Illinois

Defense Committee:

Klaus J. Miescke, Chair and Advisor
Jie Yang
Jing Wang
Sally Freels, Epidemiology and Biostatistics
Stanley L. Sclove, Information and Decision Sciences

This thesis is dedicated to my family for their love and support.

ACKNOWLEDGMENTS

I would like to thank everyone who has supported my study and research during my stay at University of Illinois at Chicago.

I would like to express my sincere gratitude to my thesis advisor, Prof. Klaus Miescke, for giving me tremendous guidance in my research. Without his inspiration and support, I wouldn't have made it this far with my PhD studies.

I would like to thank the other committee members, Prof. Jie Yang, Prof. Jing Wang, Prof. Sally Freels and Prof. Stanley Sclove for taking the time to review my thesis and for their invaluable insight and comments. I am also grateful to Prof. Sam Hedayat, Prof. Dibyen Majumdar, Prof. Huahun Chen, Prof. Hui Xie, Prof. Robert Anderson and all the other professors I took courses from for their enlightenment and encouragement for me.

I would like to thank the Department of Mathematics, Statistics, and Computer Science at UIC for supporting me during my years as a graduate student. I appreciate all the administrative staff, especially Kari Dueball, for helping me dealing with the paperwork and providing invaluable information and suggestions.

ACKNOWLEDGMENTS (Continued)

I would also like to express my deepest gratitude to my family for their unconditional love and support. I am also thankful to my friends and fellow students for their help and encouragement.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

CHAPTER **PAGE**

LIST OF TABLES

LIST OF TABLES (Continued)

LIST OF MATHEMATICS SYMBOLS

The symbols used in the thesis with their explanations are listed below. Different meanings for the same symbol occur when there is no confusion from context.

M	Maximum Number of Additional Observations
k	Number of Populations
c	The cost of sampling one more observation
$\mathbf{x}$	Vector $(x_1, x_2, \ldots, x_n)$
$\mathbb{R}$	The set of all real numbers
$\mathsf{r}(m_1, \ldots, m_k)$	Bayesian look ahead risk if drawing m_i more observations from population i
r_i	Bayesian look ahead risk of optimal allocation of i more observations determined by fixed sample-size sampling algorithm
$\tilde{\mathsf{r}}_i$	Bayesian look ahead risk of optimal allocation of up to i more observations determined by i-truncated sampling algorithm.

SUMMARY

In this paper, we study the Bayesian look ahead sampling methods for allocating up to M observations among k populations to select the best population(s).

First, we investigated the properties of fixed sample-size sampling algorithm proposed by Professor Klaus J. Miescke, which always draws fixed number of observations at the next step. Then we proposed and studied a m-truncated sampling algorithm, which draws up to m observations sequentially.

Based on these two algorithms, respectively, two Bayesian look-ahead sampling methods for allocating up to M observations among k populations are developed. To investigate the properties of and compare these two methods, we implement them to allocate up to M observations among k normal distributions with the same variance or k binomial populations to select the best population.

For given values of M, the Bayes risks of these two methods are calculated or estimated. The smaller the Bayes risk, the better the method. It turns out that when the sampling cost is large compared with the decision loss, the second method is better than the first. When the sampling cost is not very large, then in the normal case the two methods are comparable, with one method occasionally better than the other. On the other hand, in the binomial case, the second method dominates most of the time.

These two methods are then applied in various other situations. All we need to do is to calculate the look-ahead Bayesian risk of the Bayes rule if we are to draw m_i observations from population P_i, for $i = 1, \ldots, k$, where $m_1 + \ldots + m_k = m$ and $0 \leq m \leq M$.

CHAPTER 1

INTRODUCTION

1.1 Selection models

In the real life full of complexities, we often face the problem of selecting the best one or more populations among several populations. These are usually the populations of the responses to certain "treatments", which might be, for example, different training methods for the new employee, different newly developed drugs for a certain disease, or different varieties of wheat in an agricultural experiment. There are various ways selection problem have been formulated for k competing populations. For example, the goal might be to find a best population, t best populations or a best population compared with a control.

Among the early contributors to the literature of *selection rules* are Paulson(1949, 1952), Bahadur (1950), Bahadur and Robbins (1950), Bahadur and Goodman (1952), Bechhofer (1954), Bechhofer, Dunnett, and Sobel (1954), Dunnett(1955, 1960), Gupta (1956, 1965), Sobel (1956), Lehmann (1957a,b, 1961, 1963, 1966), Hall (1959), and Eaton (1967a,b). The first research monograph was written by Bechhofer, Kiefer, and Sobel (1968) with the focus on a sequential approach for exponential (Koopman–Darmois) families. The dramatic developments that would follow in the field motivated Gupta and Panchapakesan (1979) to write their classical monograph that provides an up to the time complete overview of the entire related literature. Soon after, an extension of this overview followed with Gupta and Huang (1981). A categorized guide to selection and ranking procedures was provided by Dudewicz and Koo (1982). Collections of research papers on selection rules are included in Gupta and Yackel (1971), Gupta and Moore (1977), Gupta (1977), Dudewicz(1982), Santner and Tamhane(1984), Gupta and Berger(1982, 1988), Hoppe (1993), Miescke and Herrendörfer (1993, 1994), Miescke and Rasch (1996a,b), Panchapakesanand Balakrishnan (1997), and Balakrishnan and Miescke (2006). Several books emphasizing the selecting methodologies are by Dudewicz (1976), Gibbons, Olkin, and Sobel (1977), Büringer, Martin, and Schriever (1980), Mukhopadhyay and Solanky (1994), Bechhofer, Santner, and Goldsman (1995), Rasch (1995), Horn and Volland (1995) and Liese and Miescke(2008).

Selection problems in various settings are not only statistically highly relevant, but also theoretically challenging, with techniques quite different from those of estimation and testing problems. In this paper, we use Bayesian method to find the optimal selection rules.

Let $X = (X_1, ..., X_k)$ be the vector of observations from the k populations that take on values in $(\mathcal{X}_i, \mathfrak{A}_i)$ and have the distribution P_{i,θ_i}, $i = 1, ..., k$, where the parameters $\theta_1, ..., \theta_k$ belong to the same parameter set Δ. The general selection model is

$$\mathcal{M}_s = (\times_{i=1}^{k} \mathcal{X}_i, \bigotimes_{i=1}^{k} \mathfrak{A}_i, (P_\theta)_{\theta \in \Delta^k}), \tag{1.1}$$

where it is assumed that $\mathsf{P} = (P_\theta)_{\theta \in \Delta^k}$ is a stochastic kernel, which allows us to use Bayes techniques to find optimal selection rules.

When the sampling design is unbalanced, we have to deal with an *unbalanced selection model.* More specifically, let $X_{i,1}, ..., X_{i,n_i}$ be observations from population P_{θ_i}, $i = 1, ..., k$, where all the observations are independent. Then we have the selection model

$$\mathcal{M}_{us} = (\mathsf{X}_{i=1}^{k} \mathcal{X}^{n_i}, \bigotimes_{i=1}^{k} \mathfrak{A}^{\otimes n_i}, (\bigotimes_{i=1}^{k} P_{\theta_i}^{\otimes n_i})_{\theta \in \Delta^k}). \tag{1.2}$$

It is of the form (Equation 1.1) if we identify $\mathcal{X}^{n_i}$ with $\mathcal{X}_i$, $\mathfrak{A}^{\otimes n_i}$ with $\mathfrak{A}_i$, and $\bigotimes_{i=1}^{k} P_{\theta_i}^{\otimes n_i}$ with P_θ. If $n_1 = \cdots = n_k$, $\mathcal{M}_{us}$ is balanced.

Often we reduce the model $\mathcal{M}_s$ by means of a statistic $V : \mathsf{X}_{i=1}^{k} \mathcal{X}_i \to_m \mathbb{R}^k$ and the reduced model is

$$\mathcal{M}_{ss} = (\mathbb{R}^k, \mathfrak{B}_k, (Q_\theta)_{\theta \in \Delta^k}), \tag{1.3}$$

where $Q_\theta = P_\theta \circ V^{-1}$. Usually the statistic V is sufficient for θ, and therefore $\mathcal{M}_{ss}$ and $\mathcal{M}_s$ are equivalent. We call $\mathcal{M}_{ss}$ the *standard selection model.*

The typical goal in selection theory is to find a best population. To specify what is a best population, we choose a functional $\kappa : \Delta \to \mathbb{R}$ according to the purpose of the experiment where a population i_0 is considered to be best if $i_0 \in \mathsf{M}_\kappa(\theta)$ with

$$\mathsf{M}_\kappa(\theta) = \underset{i \in \{1,...,k\}}{\arg\max}\ \kappa(\theta_i) = \{i : \kappa(\theta_i) = \max_{1 \le l \le k} \kappa(\theta_l)\}. \tag{1.4}$$

Although there may be more than one populations, a point selection rule selects exactly one population and therefore the decision space is $\mathcal{D}_{pt} = \{1, ..., k\}$. Given the model $\mathcal{M}_s$

from (Equation 1.1), a point selection rule D is a stochastic kernel $\mathsf{D}(A|x)$, $A \in \mathfrak{P}(\{1, ..., k\})$, $x = (x_1, ..., x_k) \in \times_{i=1}^{k} \mathcal{X}_i$. Let

$$\begin{aligned} \varphi_i(x) &= \mathsf{D}(\{i\}|x), \quad x \in \times_{i=1}^{k} \mathcal{X}_i,\, i = 1, ..., k, \\ \mathsf{D}(A|x) &= \sum\nolimits_{i=1}^{k} \varphi_i(x)\delta_i(A), \quad A \subseteq \{1, ..., k\},\, x \in \times_{i=1}^{k} \mathcal{X}_i, \end{aligned}$$

we may identify the stochastic kernel D with $\varphi = (\varphi_1, ..., \varphi_k)$, where

$$\varphi_i : \times_{i=1}^{k} \mathcal{X}_i \to_m [0, 1], \quad \sum\nolimits_{i=1}^{k} \varphi_i(x) = 1. \tag{1.5}$$

For brevity, φ is also called a selection rule or a selection.

Let $L : \Delta^k \times \mathcal{D}_{pt} \to \mathbb{R}$ be any loss function. The risk of a selection rule φ under L is

$$\mathsf{R}(\theta, \varphi) = \sum\nolimits_{i=1}^{k} L(\theta, i) \int \varphi_i(x) P_\theta(dx) = \sum\nolimits_{i=1}^{k} L(\theta, i)\mathsf{E}_\theta \varphi_i,\ \theta \in \Delta^k. \tag{1.6}$$

Subset selection rules are decisions on subsets of the set of k populations, and a selected subset should contain the best population(s) in some specified way. For example, we might want to select a subset of random size containing the best population. Sometimes, the experimenters are interested in selecting the subset of t $(1 < t < k)$ best populations. In this case, the decision space is

$$\mathcal{D}_{su} = \{A : A \subseteq \{1, ..., k\}, |A| = t\},$$

Given the model $\mathcal{M}_s$ in (Equation 1.1), we call every stochastic kernel $\mathsf{K} : \mathfrak{P}(\mathcal{D}_{su}) \times \mathsf{X}_{i=1}^{k}\mathcal{X}_i \to_k [0,1]$ a subset selection rule. Let $\varphi_A(x) = \mathsf{K}(\{A\}|x)$, every subset selection rule can be represented by

$$\varphi_A : \mathsf{X}_{i=1}^{k}\mathcal{X}_i \to_m [0,1], \quad A \in \mathcal{D}_{su}, \ \sum\nolimits_{A\in\mathcal{D}_{su}} \varphi_A(x) = 1, \ x \in \mathsf{X}_{i=1}^{k}\mathcal{X}_i.$$

If the experimenters have other objectives, we need to accordingly modify the decision spaces, the selection rules and the loss functions, and, therefore, solve various other selection problems in their corresponding formulations.

CHAPTER 2

TWO BAYESIAN SAMPLING METHODS

2.1 Introduction

Let $P_1, \ldots, P_k$ be k normal populations with common given variance σ^2, but their means, denoted by $\theta_1, \ldots, \theta_k$, respectively, are unknown. Our objective is to find the population with the largest mean based on the independent random samples of respective sizes $n_1, \ldots, n_k$. In the decision theoretic approach, let $L(\theta, i)$ be the given loss for selecting population i at any $\theta = (\theta_1, \ldots, \theta_k) \in R^k$. In this section, let $L(\theta, i, n) = \theta_{[k]} - \theta_i + nc$, where $\theta_{[k]} = \max_{i=1,\ldots,k}\{\theta_1, \ldots, \theta_k\}$, $\theta_{[k]} - \theta_i$ is the decision loss due to selecting population i as the best population, and nc is the sampling cost with c being the cost of sampling one more observation. It is assumed that $\theta = (\theta_1, ..., \theta_k)$ is a realization of a random vector $\Theta = (\Theta_1, ..., \Theta_k)$, where $\Theta_i \sim \mathrm{N}(\mu_i, {v_i}^{-1})$, $i = 1, ..., k$, are independent.

Since sample mean is sufficient for the distribution mean, we base our selection rule on the sample means. Suppose X_i is the i-th sample mean, then we have

$$X_i|\Theta = \theta \sim N(\theta_i, {p_i}^{-1}),\ \Theta_i \sim N(\mu_i, {v_i}^{-1}),$$

$$\Theta_i|X = x \sim N(\frac{p_i x_i + v_i \mu_i}{p_i + v_i}, \frac{1}{p_i + v_i}),\ X_i \sim N(\mu_i, p_i^{-1} + v_i^{-1})$$

where X_i is the sample mean of the i-th population and $p_i = n_i \sigma^{-2}$ is its precision.

We also consider selecting the population with the largest probability of success from k Bernoulli populations. Suppose the probability of success of the ith population is θ_i, where θ_i is a realization of the random variable $\Theta_i \sim$Beta(α_i, β_i) with $\alpha_i > 0$, $\beta_i > 0$, $i = 1, \ldots, k$. $\Theta_1, \ldots, \Theta_k$ are independent.

Because the sample total is sufficient for θ_i, we base our selection rule on these k sample totals. We have

$$X_i|\Theta = \theta \sim B(n_i, \theta_i),\ \Theta_i \sim Beta(\alpha_i, \beta_i)$$

$$\Theta_i|X = x \sim Beta(\alpha_i + x_i, \beta_i + n_i - x_i),\ X_i \sim PE(n_i, \alpha_i, \beta_i, 1),$$

where X_i is the ith sample sum. Here the unconditional marginal distribution of X_i, $i = 1, \ldots, k$ is a Pólya-Eggenberger-type distribution, sometimes called beta-binomial distribution with the following probability mass function

$$P\{X_i = x_i\} = \binom{n_i}{x_i} \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} \frac{\Gamma(\alpha_i + x_i)\Gamma(\beta_i + n_i - x_i)}{\Gamma(\alpha_i + \beta_i + n_i)},\ x_i = 0, 1, \ldots, n_i.$$

2.2 Fixed sample-size sampling algorithm

Suppose previously, we have drawn n_i observations from the i-th population for $i = 1, \ldots, k$. To select the best population, we want to draw m more observations from among k populations. To solve the problem of allocating these m observations among k populations, Professor Klaus Miescke proposed a fixed sample size sampling method without considering sampling cost. That is, the loss function is $L(\theta, i) = \theta_{[k]} - \theta_i$.

Suppose we are to draw m_i observations from the ith population. In the following, we will calculate the look ahead Bayes risk of the Bayes selection rule corresponding to this allocation for the normal and the binomial case, respectively.

a) Normal Case

Let x=$(x_1, \ldots, x_k)$ be the vector of means of the samples drawn previously. Let Y=$(Y_1, \ldots, Y_k)$ be the vector of means of samples that will be drawn. Here, $Y_i = \sum_{j=1}^{m_i} Y_{ij}/m_i$. Then it is easy to derive that

$$\Theta_i|X = x, Y = y \sim N(\frac{\alpha_i\mu_i(x) + q_i y_i}{\alpha_i + q_i}, \frac{1}{\alpha_i + q_i}),\ i = 1, \ldots, k, independent,$$

$$Y_i|X = x \sim N(\mu_i(x), \frac{\alpha_i + q_i}{\alpha_i q_i}),\ i = 1, \ldots, k, independent,$$

where $\alpha_i = p_i + \nu_i$ and $\mu_i(x) = \frac{\nu_i\mu_i + p_i x_i}{\nu_i + p_i}$.

Then the look ahead Bayes risk is

$$\begin{aligned}
& E\{\min_{i=1,\ldots,k} E\{L(\Theta, i)|X = x, Y\}|X = x\} \\
= \ & E\{\min_{i=1,\ldots,k} E(\Theta_{[k]} - \Theta_i|X = x, Y)|X = x\} \\
= \ & E_x\{E(\Theta_{[k]}|X = x, Y) - \max_{i=1,\ldots,k} E(\Theta_i|X = x, Y)\} \\
= \ & E_x(\Theta_{[k]}) - E_x\{\max_{i=1,\ldots,k} E(\Theta_i|X = x, Y)\} \\
= \ & E_x(\Theta_{[k]}) - E_x\{\max_{i=1,\ldots,k} \frac{\alpha_i\mu_i(x) + q_i Y_i}{\alpha_i + q_i}\}
\end{aligned}$$

b) Bernoulli Case

Let x=$(x_1, \ldots, x_k)$ be the vector of sums of the samples drawn previously. Let Y=$(Y_1, \ldots, Y_k)$ be the vector of sums of samples that will be drawn. Here, $Y_i = \sum_{j=1}^{m_i} Y_{ij}$. Then we have

$$\Theta_i | X = x, Y = y \sim Beta(a_i + y_i, b_i + m_i - y_i), i = 1, \ldots, k, \, independent$$

$$P(Y_i = y_i | X = x) = \binom{m_i}{y_i} \frac{\Gamma(a_i + b_i)\Gamma(a_i + y_i)\Gamma(b_i + m_i - y_i)}{\Gamma(a_i)\Gamma(b_i)\Gamma(a_i + b_i + m_i)},$$

where $y_i = 0, 1, \ldots, m_i$, $a_i = \alpha_i + x_i$, $b_i = \beta_i + n_i - x_i, i = 1, \ldots, k$, and $Y_1, \ldots, Y_k$ are independent.

The look ahead Bayes risk is

$$\begin{aligned} & E\{\min_{i=1,\ldots,k} E(L(\Theta, i)|X = x, Y)|X = x\} \\ = \; & E_x(\Theta_{[k]}) - E_x\{\max_{i=1,\ldots,k} \frac{a_i + Y_i}{a_i + b_i + m_i}\} \end{aligned}$$

Denote the Bayes look ahead risk corresponding to the allocation $(m_1, \ldots, m_k)$ by $\mathsf{r}(m_1, \ldots, m_k)$, the fixed sample-size sampling algorithm is as follows:

If there exists an allocation $(m_1^*, \ldots, m_k^*)$ such that $m_1^* + \ldots + m_k^* = m$ and

$$\mathsf{r}(m_1^*, \ldots, m_k^*) = \min_{m_1 + \ldots + m_k = m} \{\mathsf{r}(m_1, \ldots, m_k)\},$$

then $(m_1^*, \ldots, m_k^*)$ is the optimal allocation and $\mathsf{r}(m_1^*, \ldots, m_k^*)$ is the Bayes risk of our final decision. If there are more than one optimal allocations, assign them equal probability and randomly choose one as our final allocation.

It is easy to see that the optimal allocation of m more observations at the next step maximizes $E_x\{\max_{i=1,\ldots,k} \frac{\alpha_i\mu_i(x)+q_iY_i}{\alpha_i+q_i}\}$ in the normal case and $E_x\{\max_{i=1,\ldots,k} \frac{a_i+Y_i}{a_i+b_i+m_i}\}$ in the binomial case.

In the following, we will take the sampling cost into consideration and suppose that the cost of sampling one more observation is c. Let the loss function be

$$L(\theta, i, n+m) = \theta_{[k]} - \theta_i + nc + mc,$$

where $\theta_{[k]} - \theta_i$ is the loss from selecting ith population, that is, the decision loss, and $nc+mc$ is the cost of sampling n+m observations.

Then the look ahead Bayes risk corresponding to the allocation $(m_1, \ldots, m_k)$ in the normal case is

$$E_x(\Theta_{[k]}) - E_x\{\max_{i=1,\ldots,k} \frac{\alpha_i\mu_i(x) + q_iY_i}{\alpha_i + q_i}\} + nc + mc,$$

while the risk in the Bernoulli case is

$$E_x(\Theta_{[k]}) - E_x\{\max_{i=1,\ldots,k} \frac{a_i + Y_i}{a_i + b_i + m_i}\} + nc + mc.$$

We can see that in both cases, the optimal allocation does't change, only the final Bayes risk increases by nc+mc.

2.3 Properties of fixed sample-size algorithm

Theorem 2.3.1 Without sampling cost, that is, c=0, the Bayesian risk of the optimal allocation of $m^{'}$ more observations at the second stage is no more than the Bayesian risk of the optimal allocation of m more observations if $m < m^{'}$.

Proof. For any allocation $(m_1^{'}, \cdots, m_k^{'})$ of $m^{'}$ more observations, we can find one allocation $(m_1, \cdots, m_k)$ of m more observations such that $m_i \leq m_i^{'}$, $i = 1, \cdots, k$. Then the Bayesian risk of $(m_1^{'}, \cdots, m_k^{'})$ is no more than that of $(m_1, \cdots, m_k)$ because the Bayesian rule using m_i observations from population i, $i = 1, \cdots, k$, is one of rules using $m_i^{'}$ observations(it just ignores $m_i^{'} - m_i$ observations) from population i, $i = 1, \cdots, k$, which have no less risk than the Bayesian rule of the allocation $(m_1^{'}, \cdots, m_k^{'})$.

Therefore, the Bayesian risk of the optimal allocation of $m^{'}$ more observations, that is, the minimum of Bayesian risks of all allocations of $m^{'}$, is no more than the Bayesian risk of the allocation $(m_1, \cdots, m_k)$ of m more observations for which there exists $(m_1^{'}, \cdots, m_k^{'})$ such that $m_i \leq m_i^{'}$, $i = 1, \cdots, k$.

But, for any allocation $(m_1, \cdots, m_k)$ of m more observations, we can find one allocation $(m_1^{'}, \cdots, m_k^{'})$ of $m^{'}$ more observations such that $m_i \leq m_i^{'}$, $i = 1, \cdots, k$. Therefore, the Bayesian risk of the optimal allocation of $m^{'}$ more observations is no more than the Bayesian risk of any allocation of m observations. Thus, the Bayesian risk of the optimal allocation of $m^{'}$ more observations is no more than the Bayesian risk of the optimal allocation of m more observations.

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
-0.3	0.2	0.6	0	0.0570	0	0	0
-0.3	0.2	0.6	1	0.0570	0	1	0
-0.3	0.2	0.6	2	0.0568	0	2	0
-0.3	0.2	0.6	3	0.0563	0	3	0
-0.3	0.2	0.6	5	0.0547	0	5	0
-0.3	0.2	0.6	10	0.0496	0	7	3
-0.3	0.2	0.6	20	0.0411	0	12	8
-0.3	0.2	0.6	30	0.0354	0	17	13

TABLE I

BAYES RISK AND OPTIMAL ALLOCATION WHEN $X = (-0.3, 0.2, 0.6)$

Example 2.3.2 Let $k = 3, n = (4, 8, 12)$, $\sigma^2 = 1$, $\mu_i = 0$, $\nu_i = 1$, for $i = 1, 2, 3$. Given various observations at the first stage, calculate the Bayes risk of optimal allocation at the second stage for m=0, 1, 2, 3, 5, 10, 20, 30, respectively.

From the computation result (Table 1-6), we can see that the Bayes risk of optimal allocation decreases as the sample size at the second stage increases. We also can see that the optimal allocation at the second stage tends to draw more observations from the population from which fewer observations have been drawn or larger sample mean has been obtained at the first stage.

Example 2.3.3 n=(5, 12, 17), k=3, $\alpha = (1, 1, 1), \beta = (1, 1, 1)$. Given various observations at the first stage, calculate the Bayes risk of optimal allocation at the second stage for m=0, 1, 2, 3, 5, 10, 20, 30, respectively.

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
-0.3	0.6	0.2	0	0.0650	0	0	0
-0.3	0.6	0.2	1	0.0649	0	1	0
-0.3	0.6	0.2	2	0.0646	0	2	0
-0.3	0.6	0.2	3	0.0639	0	3	0
-0.3	0.6	0.2	5	0.0617	0	4	1
-0.3	0.6	0.2	10	0.0555	0	7	3
-0.3	0.6	0.2	20	0.0458	0	12	8
-0.3	0.6	0.2	30	0.0395	0	17	13

TABLE II

BAYES RISK AND OPTIMAL ALLOCATION WHEN $X = (-0.3, 0.6, 0.2)$

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
0.6	-0.3	0.2	0	0.1009	0	0	0
0.6	-0.3	0.2	1	0.0968	1	0	0
0.6	-0.3	0.2	2	0.0884	2	0	0
0.6	-0.3	0.2	3	0.0813	3	0	0
0.6	-0.3	0.2	5	0.0711	5	0	0
0.6	-0.3	0.2	10	0.0574	9	0	1
0.6	-0.3	0.2	20	0.0430	14	0	6
0.6	-0.3	0.2	30	0.0349	19	0	11

TABLE III

BAYES RISK AND OPTIMAL ALLOCATION WHEN $X = (0.6, -0.3, 0.2)$

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
0.6	0.2	-0.3	0	0.1072	0	0	0
0.6	0.2	-0.3	1	0.1034	1	0	0
0.6	0.2	-0.3	2	0.0955	2	0	0
0.6	0.2	-0.3	3	0.0885	3	0	0
0.6	0.2	-0.3	5	0.0785	4	1	0
0.6	0.2	-0.3	10	0.0611	7	3	0
0.6	0.2	-0.3	20	0.0429	12	8	0
0.6	0.2	-0.3	30	0.0335	17	13	0

TABLE IV

BAYES RISK AND OPTIMAL ALLOCATION WHEN $X = (0.6, 0.2, -0.3)$

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
0.2	-0.3	0.6	0	0.0720	0	0	0
0.2	-0.3	0.6	1	0.0710	1	0	0
0.2	-0.3	0.6	2	0.0670	2	0	0
0.2	-0.3	0.6	3	0.0628	3	0	0
0.2	-0.3	0.6	5	0.0558	5	0	0
0.2	-0.3	0.6	10	0.0456	9	0	1
0.2	-0.3	0.6	20	0.0341	14	0	6
0.2	-0.3	0.6	30	0.0274	19	0	11

TABLE V

BAYES RISK AND OPTIMAL ALLOCATION WHEN $X = (0.2, -0.3, 0.6)$

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
0.2	0.6	-0.3	0	0.0860	0	0	0
0.2	0.6	-0.3	1	0.0846	1	0	0
0.2	0.6	-0.3	2	0.0799	2	0	0
0.2	0.6	-0.3	3	0.0751	3	0	0
0.2	0.6	-0.3	5	0.0675	5	0	0
0.2	0.6	-0.3	10	0.0532	7	3	0
0.2	0.6	-0.3	20	0.0375	12	8	0
0.2	0.6	-0.3	30	0.0291	17	13	0

TABLE VI

BAYES RISK AND OPTIMAL ALLOCATION WHEN $X = (0.2, 0.6, -0.3)$

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
3	7	9	0	0.1074	0	0	0
3	7	9	1	0.0768	1	0	0
3	7	9	2	0.0734	2	0	0
3	7	9	3	0.0666	3	0	0
3	7	9	5	0.0611	5	0	0
3	7	9	10	0.0498	8	2	0
3	7	9	20	0.0387	12	7	1
3	7	9	30	0.0322	16	11	3

TABLE VII

BAYES RISK AND OPTIMAL ALLOCATION WHEN X=(3,7,9)

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
3	7	8	0	0.0986	0	0	0
3	7	8	1	0.0680	1	0	0
3	7	8	2	0.0646	2	0	0
3	7	8	3	0.0578	3	0	0
3	7	8	5	0.0523	5	0	0
3	7	8	10	0.0426	9	1	0
3	7	8	20	0.0330	13	7	0
3	7	8	30	0.0275	18	12	0

TABLE VIII

BAYES RISK AND OPTIMAL ALLOCATION WHEN X=(3,7,8)

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
3	5	9	0	0.0746	0	0	0
3	5	9	1	0.0634	1	0	0
3	5	9	2	0.0571	2	0	0
3	5	9	3	0.0521	3	0	0
3	5	9	5	0.0459	5	0	0
3	5	9	10	0.0390	10	0	0
3	5	9	20	0.0316	15	1	4
3	5	9	30	0.0271	20	1	9

TABLE IX

BAYES RISK AND OPTIMAL ALLOCATION WHEN X=(3,5,9)

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
3	5	8	0	0.0596	0	0	0
3	5	8	1	0.0596	0	0	1
3	5	8	2	0.0534	2	0	0
3	5	8	3	0.0509	3	0	0
3	5	8	5	0.0459	5	0	0
3	5	8	10	0.0400	10	0	0
3	5	8	20	0.0322	13	3	4
3	5	8	30	0.0270	17	7	6

TABLE X

BAYES RISK AND OPTIMAL ALLOCATION WHEN X=(3,5,8)

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
2	7	9	0	0.0678	0	0	0
2	7	9	1	0.0678	0	0	1
2	7	9	2	0.0625	0	2	0
2	7	9	3	0.0622	0	3	0
2	7	9	5	0.0585	1	4	0
2	7	9	10	0.0504	4	6	0
2	7	9	20	0.0396	8	9	3
2	7	9	30	0.0329	10	13	7

TABLE XI

BAYES RISK AND OPTIMAL ALLOCATION WHEN X=(2,7,9)

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
2	7	8	0	0.0543	0	0	0
2	7	8	1	0.0543	1	0	0
2	7	8	2	0.0543	2	0	0
2	7	8	3	0.0509	3	0	0
2	7	8	5	0.0484	5	0	0
2	7	8	10	0.0429	6	4	0
2	7	8	20	0.0346	10	9	1
2	7	8	30	0.0293	14	13	3

TABLE XII

BAYES RISK AND OPTIMAL ALLOCATION WHEN X=(2,7,8)

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
2	5	9	0	0.0640	0	0	0
2	5	9	1	0.0640	1	0	0
2	5	9	2	0.0577	2	0	0
2	5	9	3	0.0552	3	0	0
2	5	9	5	0.0502	5	0	0
2	5	9	10	0.0444	10	0	0
2	5	9	20	0.0375	12	3	5
2	5	9	30	0.0315	14	8	8

TABLE XIII

BAYES RISK AND OPTIMAL ALLOCATION WHEN X=(2,5,9)

x_1	x_2	x_3	m	Bayes Risk	m_1^*	m_2^*	m_3^*
2	5	8	0	0.0920	0	0	0
2	5	8	1	0.0807	1	0	0
2	5	8	2	0.0744	2	0	0
2	5	8	3	0.0694	3	0	0
2	5	8	5	0.0633	5	0	0
2	5	8	10	0.0548	6	3	1
2	5	8	20	0.0426	10	6	4
2	5	8	30	0.0351	14	10	6

TABLE XIV

BAYES RISK AND OPTIMAL ALLOCATION WHEN X=(2,5,8)

From the computation result (Table 7-14), we can observe that as m increases, the Bayes risk of optimal allocation decreases, and that the optimal allocation at the second stage tends to draw more observations from the population which has the larger sample proportion at the first stage or from which fewer observations have been drawn at the first stage. The same result has been observed for the normal case.

If the cost is not zero, drawing more observations will decrease the decision risk, but increase the sampling cost at the same time. The optimal sample size at the second stage m^* belongs to

$$\arg\min_{m \in N} \{E_x\{\Theta_{[k]}\} - \max_{m_1+\ldots+m_k=m} E_x\{\max_{i=1,\cdots,k} E(\Theta_i|x,Y)\} + nc + mc\}$$

or equivalently,

$$\arg\max_{m\in N}\{\max_{m_1+\ldots+m_k=m} E_x\{\max_{i=1,\cdots,k} E(\Theta_i|x,Y)\} - mc\},$$

where $N = \{0, 1, 2, \ldots\}$.

2.4 m-truncated sampling algorithm

Even we are allowed to allocate m more observations, it is not necessary that we draw exactly m additional observations when the sampling cost is taken into account, maybe drawing fewer observations will lead to smaller Bayes risk.

For Professor Klaus Miescke's sampling algorithm, no matter what the optimal allocation is, m more observations are always drawn at the next step. This is why it is called a fixed-sample size sampling algorithm. If we compare the Bayes risk without additional observations with the minimum of the Bayes risks respectively corresponding to drawing $1, \ldots,$ m more observations according to the optimal allocation determined by fixed sample-size algorithm, when the former is no more than the latter, stop sampling, otherwise, draw one more observation from the population favored by the fixed sample-size 1 sampling algorithm, proceed this way until the former is no more than the latter or m more observations are drawn, then maybe we can end up with smaller Bayes risk with smaller sample size. It is based on this idea that I propose and study the m-truncated sampling algorithm, which proceeds as follows:

Step 1. Calculate the Bayes risk without additional observation. Then calculate the look-ahead Bayes risk of the optimal allocation of i more observations at the second stage, which is determined by the fixed sample size i sampling procedure, $i = 1, \ldots, m$, and find the minimum look ahead risk.

Step 2. Compare the Bayes risk without additional observation with the minimum look-ahead Bayes risk found in Step 1. If the former is not greater than the latter, then stop sampling more observations and make a decision according to the Bayes decision rule based on whatever we have, *i.e.*, the former is the Bayes risk of this decision. Otherwise, draw one more observation from the population favored by the optimal allocation of one more observation, which is determined by the fixed sample size 1 sampling procedure, update the prior with the new observation, set m to $m-1$. If $m > 0$, return to Step 1, otherwise, go to Step 3.

Step 3. If m more observations have been drawn, stop sampling more observations and the Bayes risk of our final decision is the Bayes risk without additional observations under the updated prior.

Obviously, at most m additional observations can be drawn using this procedure.

2.5 Two Bayesian sampling methods

Because of budget restriction, we can't always have as many observations as we want. Usually, there is a limit to the number of observations that can be drawn in the future. Suppose we can draw up to M observations. To allocate up to M observations among k populations in an optimal way, I proposed two Bayesian methods based on the two previously mentioned algorithms, respectively.

The *first method*, based on the fixed sample size sampling algorithm, is as follows. Calculate the Bayes risk without additional observations, then calculate the look-ahead Bayes risk of the optimal allocation of m observations at the next step, determined by the fixed sample size sampling algorithm, for $m = 1, 2, \ldots, M$. Denote those risks by r_0, $\mathsf{r}_1, \ldots, \mathsf{r}_M$, respectively. We call m^* the optimal sample size if

$$\mathsf{r}_{m^*} = \min_{i=0,\ldots,M} \{\mathsf{r}_i\}.$$

If $m^* = 0$, then we make a decision without further sampling. Otherwise, we adopt the optimal allocation of m^* observations determined by the fixed sample size m^* sampling algorithm as our optimal allocation of up to M observations and the Bayes risk of our decision is r_{m^*}.

The *second method* is based on the m-truncated sampling algorithm, and its process, similar to that of the first method, is as follows. Calculate the Bayes risk without additional observations, then find the look-ahead Bayes risk of the allocation of up to m observations determined by the m-truncated sampling algorithm, for $m = 1, 2, \ldots, M$(We estimate the Bayes risk of this allocation by averaging the 10,000 risks obtained by independently running the m-truncated sampling procedure 10,000 times). Denote these risks by $\tilde{\mathsf{r}}_0$, $\tilde{\mathsf{r}}_1, \ldots, \tilde{\mathsf{r}}_M$, respectively. Obviously, $\tilde{\mathsf{r}}_0 = \mathsf{r}_0$. Find m^{**} such that

$$\tilde{\mathsf{r}}_{m^{**}} = \min_{i=0,\ldots,M} \{\tilde{\mathsf{r}}_i\}.$$

If $m^{**} = 0$, then we just make a decision without further sampling. Otherwise, we use the allocation determined by the m^{**}-truncated sampling algorithm and $\tilde{\mathsf{r}}_{m^{**}}$ is the Bayes risk of our final decision.

2.6 Comparison of two sampling methods

Example 2.6.1 Let $k = 3, n = (6, 9, 15)$, $\sigma^2 = 1$, $\mu_i = 1$, $\nu_i = 0.5$, for $i = 1, 2, 3$. The observation at the first stage is (0.5, 1.1, 1.6). Calculate the Bayes risks of the two algorithms for $m = 0, 1, \cdots, 12$ with different sampling costs.

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.00000001	0	0.0256	0.0256	30	30	30	30
0.00000001	1	0.0256	0.0256	31	31	31	31
0.00000001	2	0.0256	0.0256	32	31.3527	31	32
0.00000001	3	0.0256	0.0257	33	32.1723	31	33
0.00000001	4	0.0255	0.0255	34	32.9766	32	34
0.00000001	5	0.0253	0.0252	35	33.8569	32	35
0.00000001	6	0.0251	0.0249	36	34.7278	32	36
0.00000001	7	0.0249	0.0247	37	35.6825	32	37
0.00000001	8	0.0246	0.0246	38	36.61	33	38
0.00000001	9	0.0243	0.0243	39	37.5064	33	39
0.00000001	10	0.0239	0.0238	40	38.4195	34	40
0.00000001	11	0.0236	0.0240	41	39.3409	34	41
0.00000001	12	0.0233	0.0235	42	40.2694	34	42

TABLE XV

METHOD 1 VS METHOD 2 WHEN COST=10^{-8}

Where Bayes Risk 1 is the Bayes risk of fixed sample-size sampling algorithm, while Bayes Risk 2 is the estimated Bayes risk of the m-truncated sampling algorithm based on 10,000 runs. SS is the final sample size of the first algorithm, while ESS is the estimate of the expected

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.000001	0	0.0257	0.0257	30	30	30	30
0.000001	1	0.0257	0.0257	31	30	30	30
0.000001	2	0.0257	0.0255	32	31.0937	31	32
0.000001	3	0.0256	0.0257	33	31.5534	31	33
0.000001	4	0.0255	0.0256	34	32.1679	31	34
0.000001	5	0.0253	0.0255	35	32.7965	31	35
0.000001	6	0.0251	0.0252	36	33.4409	31	36
0.000001	7	0.0249	0.0249	37	34.0786	31	37
0.000001	8	0.0246	0.0244	38	34.7808	32	38
0.000001	9	0.0243	0.0246	39	35.5863	32	39
0.000001	10	0.0240	0.0243	40	36.3512	32	40
0.000001	11	0.0237	0.0237	41	37.1061	32	41
0.000001	12	0.0233	0.0231	42	37.8648	32	42

TABLE XVI

METHOD 1 VS METHOD 2 WHEN COST=10^{-6}

sample size of the second algorithm, and Min(Max) SS is the minimal(maximal) sample size of the second algorithm in 10,000 runs.

From the computation result (Table 15-20), we can see that for fixed sampling cost, as m increases from 1 to 12, the expected sample size of the m-truncated sampling algorithm increases most of the time, while the Bayes risk decreases most of the time.

When m is fixed, as sampling cost increases, the expected sample size of the m-truncated sampling algorithm decreases. When the sampling cost is very small (for example, when cost=10^{-8}), the expected sample size of the m-truncated sampling algorithm is close to m.

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.00001	0	0.0259	0.0259	30	30	30	30
0.00001	1	0.0259	0.0259	31	30	30	30
0.00001	2	0.0259	0.0259	32	30	30	30
0.00001	3	0.0259	0.0260	33	31.2415	31	33
0.00001	4	0.0258	0.0257	34	31.6134	31	34
0.00001	5	0.0257	0.0258	35	32.0752	31	35
0.00001	6	0.0255	0.0252	36	32.5411	31	36
0.00001	7	0.0252	0.0255	37	33.0595	31	37
0.00001	8	0.0249	0.0252	38	33.5806	31	38
0.00001	9	0.0247	0.0247	39	34.0888	31	39
0.00001	10	0.0243	0.0245	40	34.6536	31	40
0.00001	11	0.0240	0.0236	41	35.153	31	41
0.00001	12	0.0237	0.0235	42	35.7396	31	42

TABLE XVII

METHOD 1 VS METHOD 2 WHEN COST=10^{-5}

When sampling cost gets larger, the difference between m and the expected sample size gets larger for large value of m.

We can also see that when sampling cost is large, (for example, when cost=$8 * 10^{-5}$), the second method is better than the first method for M=1,. . ., 12. When the sampling cost is not very large, the two methods are comparable. Sometimes, the first is better; Sometimes, the second prevails. The difference of their risks is not large.

Example 2.6.2 Let $k = 3, n = (5, 8, 13), \alpha = (1, 1, 1), \beta = (0.5, 0.5, 0.5)$. The observation at the first stage is (2, 4, 7). Calculate the Bayes risks of two algorithms for M=0, 1, $\cdots$, 12 with different costs.

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.00003	0	0.0265	0.0265	30	30	30	30
0.00003	1	0.0266	0.0265	31	30	30	30
0.00003	2	0.0266	0.0265	32	30	30	30
0.00003	3	0.0266	0.0265	33	30	30	30
0.00003	4	0.0265	0.0265	34	31.3731	31	34
0.00003	5	0.0264	0.0263	35	31.6946	31	35
0.00003	6	0.0262	0.0261	36	32.0508	31	36
0.00003	7	0.0260	0.0262	37	32.4168	31	37
0.00003	8	0.0257	0.0259	38	32.8575	31	38
0.00003	9	0.0254	0.0253	39	33.2473	31	39
0.00003	10	0.0251	0.0252	40	33.6791	31	40
0.00003	11	0.0248	0.0249	41	34.1505	31	41
0.00003	12	0.0245	0.0245	42	34.5753	31	42

TABLE XVIII

METHOD 1 VS METHOD 2 WHEN COST=$3 * 10^{-5}$

From the simulation result (Table 21-26), we can see that most of the time, the second method is better than the first. The superiority of the former becomes more obvious when the sampling cost gets larger.

For both methods, we need to know the look-ahead risk of the Bayesian rule if we are to draw m_i observations from population Π_i, $i = 1, ..., k$, where $m_1 + \ldots + m_k = m$ and $1 \leq m \leq M$. Denote the look-ahead Bayes risk by $\mathsf{r}(m_1, \ldots, m_k)$. After these risks have been calculated, the remaining work is straightforward.

Therefore, if we want to apply these methods in another situation, we only need to know how to calculate $\mathsf{r}(m_1, \ldots, m_k)$ in that situation.

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.00005	0	0.0271	0.0271	30	30	30	30
0.00005	1	0.0272	0.0271	31	30	30	30
0.00005	2	0.0272	0.0271	32	30	30	30
0.00005	3	0.0272	0.0271	33	30	30	30
0.00005	4	0.0272	0.0271	34	30	30	30
0.00005	5	0.0271	0.0267	35	31.5152	31	35
0.00005	6	0.0269	0.0266	36	31.7884	31	36
0.00005	7	0.0267	0.0267	37	32.1134	31	37
0.00005	8	0.0265	0.0265	38	32.4776	31	38
0.00005	9	0.0262	0.0264	39	32.8295	31	39
0.00005	10	0.0259	0.0257	40	33.1881	31	40
0.00005	11	0.0257	0.0256	41	33.5547	31	41
0.00005	12	0.0254	0.0264	42	33.353	31	42

TABLE XIX

METHOD 1 VS METHOD 2 WHEN COST=$5 * 10^{-5}$

In the following two chapters, I will consider other selection problems. I will derive the formula for $\mathsf{r}(m_1, \ldots, m_k)$ under various conditions. In particular, I will set up the formula for $\mathsf{r}(m_1, \cdots, m_k)$ where $m_1 + \ldots + m_k = 1$, because for the m-truncated sampling algorithm, we need to know the optimal allocation of the next observation. I also prove a theorem that helps easily find the optimal allocation of the next observation in the subset selection of b best normal populations case.

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.00008	0	0.0280	0.0280	30	30	30	30
0.00008	1	0.0281	0.0280	31	30	30	30
0.00008	2	0.0282	0.0280	32	30	30	30
0.00008	3	0.0282	0.0280	33	30	30	30
0.00008	4	0.0282	0.0280	34	30	30	30
0.00008	5	0.0281	0.0280	35	30	30	30
0.00008	6	0.0280	0.0276	36	31.5947	31	36
0.00008	7	0.0278	0.0277	37	31.8536	31	37
0.00008	8	0.0276	0.0269	38	32.0801	31	38
0.00008	9	0.0274	0.0273	39	32.4111	31	39
0.00008	10	0.0271	0.0268	40	32.7209	31	40
0.00008	11	0.0269	0.0266	41	33.0005	31	41
0.00008	12	0.0266	0.0264	42	33.353	31	42

TABLE XX

METHOD 1 VS METHOD 2 WHEN COST=$8 * 10^{-5}$

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.000001	0	0.0979	0.0979	26	26	26	26
0.000001	1	0.0875	0.0875	27	27	27	27
0.000001	2	0.0808	0.0809	28	27.5238	27	28
0.000001	3	0.0788	0.0789	29	28.4525	28	29
0.000001	4	0.0736	0.0729	30	29.5068	29	30
0.000001	5	0.0717	0.0710	31	30.2441	29	31
0.000001	6	0.0681	0.0662	32	31.2638	30	32
0.000001	7	0.0650	0.0634	33	31.8918	30	33
0.000001	8	0.0628	0.0603	34	32.9406	31	34
0.000001	9	0.0603	0.0582	35	33.7853	32	35
0.000001	10	0.0584	0.0559	36	34.7522	33	36
0.000001	11	0.0563	0.0541	37	35.7087	34	37
0.000001	12	0.0546	0.0528	38	36.6829	34	38

TABLE XXI

METHOD 1 VS METHOD 2 WHEN COST=10^{-6}

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.00001	0	0.0981	0.0981	26	26	26	26
0.00001	1	0.0878	0.0878	27	27	27	27
0.00001	2	0.0810	0.0809	28	27.5275	27	28
0.00001	3	0.0791	0.0790	29	28.452	28	29
0.00001	4	0.0739	0.0734	30	29.5104	29	30
0.00001	5	0.0720	0.0715	31	30.2499	29	31
0.00001	6	0.0684	0.0663	32	31.2566	30	32
0.00001	7	0.0653	0.0639	33	31.8912	30	33
0.00001	8	0.0631	0.0603	34	32.9441	31	34
0.00001	9	0.0606	0.0583	35	33.7722	32	35
0.00001	10	0.0587	0.0565	36	34.7581	33	36
0.00001	11	0.0566	0.0542	37	35.7235	34	37
0.00001	12	0.0549	0.0531	38	36.6783	34	38

TABLE XXII

METHOD 1 VS METHOD 2 WHEN COST=10^{-5}

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.0001	0	0.1005	0.1005	26	26	26	26
0.0001	1	0.0902	0.0902	27	27	27	27
0.0001	2	0.0835	0.0835	28	27.5205	27	28
0.0001	3	0.0817	0.0817	29	28.4533	28	29
0.0001	4	0.0766	0.0759	30	29.5059	29	30
0.0001	5	0.0747	0.0744	31	30.1897	29	31
0.0001	6	0.0713	0.0691	32	31.2576	30	32
0.0001	7	0.0683	0.0671	33	31.8998	30	33
0.0001	8	0.0661	0.0632	34	32.9361	31	34
0.0001	9	0.0638	0.0614	35	33.7088	31	35
0.0001	10	0.0619	0.0593	36	34.7174	32	36
0.0001	11	0.0600	0.0575	37	35.6357	33	37
0.0001	12	0.0583	0.0565	38	36.5106	34	38

TABLE XXIII

METHOD 1 VS METHOD 2 WHEN COST=10^{-4}

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.0005	0	0.1109	0.1109	26	26	26	26
0.0005	1	0.1010	0.1010	27	27	27	27
0.0005	2	0.0947	0.0945	28	27.538	27	28
0.0005	3	0.0933	0.0929	29	28.4478	28	29
0.0005	4	0.0886	0.0879	30	29.5158	29	30
0.0005	5	0.0871	0.0860	31	30.2477	29	31
0.0005	6	0.0841	0.0826	32	31.1979	30	32
0.0005	7	0.0815	0.0797	33	31.8627	30	33
0.0005	8	0.0797	0.0766	34	32.7963	30	34
0.0005	9	0.0778	0.0750	35	33.548	30	35
0.0005	10	0.0763	0.0734	36	34.2986	30	36
0.0005	11	0.0748	0.0721	37	35.3848	31	37
0.0005	12	0.0735	0.0710	38	36.0193	31	38

TABLE XXIV

METHOD 1 VS METHOD 2 WHEN COST=$5 * 10^{-4}$

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.001	0	0.1239	0.1239	26	26	26	26
0.001	1	0.1145	0.1144	27	27	27	27
0.001	2	0.1087	0.1082	28	27.5215	27	28
0.001	3	0.1078	0.1073	29	28.4541	28	29
0.001	4	0.1036	0.1027	30	29.5142	29	30
0.001	5	0.1026	0.1012	31	30.047	29	31
0.001	6	0.1001	0.0977	32	31.1832	30	32
0.001	7	0.0980	0.0956	33	31.4871	30	33
0.001	8	0.0967	0.0925	34	32.3444	30	34
0.001	9	0.0953	0.0914	35	32.8898	30	35
0.001	10	0.0943	0.0900	36	33.5455	30	36
0.001	11	0.0933	0.0890	37	34.3923	30	37
0.001	12	0.0925	0.0887	38	34.9046	30	38

TABLE XXV

METHOD 1 VS METHOD 2 WHEN COST=10^{-3}

Cost	m	Bayes Risk 1	Bayes Risk 2	SS	ESS	Min SS	Max SS
0.01	0	0.3579	0.3579	26	26	26	26
0.01	1	0.3575	0.3574	27	27	27	27
0.01	2	0.3607	0.3561	28	27.5237	27	28
0.01	3	0.3688	0.3561	29	27.5259	27	28
0.01	4	0.3736	0.3560	30	27.529	27	28
0.01	5	0.3816	0.3560	31	27.5285	27	28
0.01	6	0.3881	0.3558	32	27.5253	27	28
0.01	7	0.3950	0.3560	33	27.5299	27	28
0.01	8	0.4027	0.3558	34	27.5225	27	28
0.01	9	0.4103	0.3559	35	27.5255	27	28
0.01	10	0.4183	0.3560	36	27.5224	27	28
0.01	11	0.4263	0.3560	37	27.5212	27	28
0.01	12	0.4345	0.3561	38	27.53	27	28

TABLE XXVI

METHOD 1 VS METHOD 2 WHEN COST=10^{-2}

CHAPTER 3

SELECTION OF THE BEST POPULATION(S)

3.1 Selection of the best population

Point selection rules are decisions on which of the k populations is the best. Whether a population is the best or not is based on certain criteria. For example, the population with the largest mean is considered best among k normal populations with the same known variance. Although there may be more than one best populations, a point selection rule selects exactly one population.

In the following sections, we will find the optimal allocation of m more observations to select the best normal, Poisson, or Gamma population under various conditions.

3.1.1 Selection of the smallest normal variance

There are k normal populations with the same mean. Our objective is to choose the population associated with the smallest variance.

Suppose $X_1, ..., X_n$, are i.i.d random variables from $X \sim$N(μ, ϕ), where μ is known.

Given ϕ, the pdf of $X = (X_1, ..., X_n)$ is

$$\begin{aligned} p(x|\phi) &= \prod_{i=1}^{n}(2\pi\phi)^{-1/2}e^{-\frac{1}{2}(x_i-\mu)^2/\phi} \\ &= (2\pi\phi)^{-n/2}e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2/\phi} \\ &= (2\pi\phi)^{-n/2}e^{-\frac{s}{2\phi}}. \end{aligned}$$

where $s = \sum_{i=1}^{n}(x_i - \mu)^2$.

By Fisher-Neyman Factorization Theorem, s is sufficient for ϕ.

Given ϕ, let $\frac{s}{\phi} = X$, then $X \sim X^2(n)$, the pdf of X is

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

we can get the pdf of S, which is

$$\begin{aligned} f(s|\phi) &= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}s^{\frac{n}{2}-1}\phi^{1-\frac{n}{2}}e^{-\frac{s}{2\phi}}\phi^{-1} \\ &= \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}s^{\frac{n}{2}-1}\phi^{-\frac{n}{2}}e^{-\frac{s}{2\phi}}, \quad s > 0 \end{aligned}$$

Suppose $\Phi \sim lg(\alpha, \beta)$, $\alpha > 0$, $\beta > 0$, then the pdf of Φ is

$$\pi(\phi) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\phi^{-\alpha-1}e^{-\frac{\beta}{\phi}},\ \phi > 0$$

Given $x=(x_1,...,x_n)$, that is, given s, $\Phi \sim lg(\alpha+\frac{n}{2},\beta+\frac{s}{2})$, therefore, given X=x, the pdf of Φ is

$$\pi(\phi|s)=\frac{(\beta+\frac{s}{2})^{\alpha+\frac{n}{2}}}{\Gamma(\alpha+\frac{n}{2})}\phi^{-(\alpha+\frac{n}{2})-1}e^{-\frac{\beta+\frac{s}{2}}{\phi}},\ \phi>0$$

Because $f(s|\phi)\pi(\phi)=\pi(\phi|s)m(s)$, we have

$$\begin{aligned}
m(s) &= \frac{f(s|\phi)\pi(\phi)}{\pi(\phi|s)} \\
&= \frac{\frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})}s^{\frac{n}{2}-1}\phi^{-\frac{n}{2}}e^{-\frac{s}{2\phi}}\frac{\beta^\alpha}{\Gamma(\alpha)}\phi^{-\alpha-1}e^{-\frac{\beta}{\phi}}}{\frac{(\beta+\frac{s}{2})^{\alpha+\frac{n}{2}}}{\Gamma(\alpha+\frac{n}{2})}\phi^{-(\alpha+\frac{n}{2})-1}e^{-\frac{\beta+\frac{s}{2}}{\phi}}} \\
&= \frac{\beta^\alpha s^{\frac{n}{2}-1}}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})\Gamma(\alpha)}\cdot\frac{\Gamma(\alpha+\frac{n}{2})}{(\beta+\frac{s}{2})^{\alpha+\frac{n}{2}}} \\
&= \frac{\Gamma(\alpha+\frac{n}{2})}{2^{\frac{n}{2}}\Gamma(\alpha)\Gamma(\frac{n}{2})}\cdot\frac{\beta^\alpha s^{\frac{n}{2}-1}}{(\beta+\frac{s}{2})^{\alpha+\frac{n}{2}}},\quad s>0.
\end{aligned}$$

It is easy to calculate the expectation of S.

$$\begin{aligned}
E(S) &= \int_0^\infty \frac{\Gamma(\alpha+\frac{n}{2})}{2^{\frac{n}{2}}\Gamma(\alpha)\Gamma(\frac{n}{2})}\frac{\beta^\alpha s^{\frac{n}{2}}}{(\beta+\frac{s}{2})^{\alpha+\frac{n}{2}}}ds \\
&= \frac{\Gamma(\alpha+\frac{n}{2})\beta^\alpha}{2^{\frac{n}{2}}\Gamma(\alpha)\Gamma(\frac{n}{2})}\int_0^\infty \frac{s^{\frac{n}{2}+1-1}}{(\beta+\frac{s}{2})^{\alpha-1+\frac{n}{2}+1}}ds \\
&= \frac{\Gamma(\alpha+\frac{n}{2})\beta^\alpha}{2^{\frac{n}{2}}\Gamma(\alpha)\Gamma(\frac{n}{2})}\cdot\frac{2^{\frac{n+2}{2}}\Gamma(\alpha-1)\Gamma(\frac{n+2}{2})}{\Gamma(\alpha+\frac{n}{2})\beta^{\alpha-1}} \\
&= \frac{n\beta}{\alpha-1}.
\end{aligned}$$

At the end of the first stage, n_i observations have been drawn from the i-th population. Let $s_i = \sum_{j=1}^{n_i}(x_{ij} - \mu)^2$ and s=$(s_1, \ldots, s_k)$, then the updated prior is

$\Phi_i \sim lg(\alpha_i + \frac{n_i}{2}, \beta_i + \frac{s_i}{2})$, $i = 1, ..., k$, and Φ_i's are independent.

Suppose at the second stage, m_i observations, $Y_{i1}, \ldots, Y_{im_i}$, are to be drawn from the i-th population. Let $W_i = \sum_{j=1}^{m_i}(Y_{ij} - \mu)^2$ and $W = (W_1, ..., W_k)$. The posterior distribution of Φ_i, given $S = s, W = w$, is

$$\Phi_i \sim lg(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2}, \beta_i + \frac{s_i}{2} + \frac{w_i}{2}),\ i = 1, ..., k,$$

and Φ_i's are independent.

The marginal pdf of W_i, given $S = s$, is

$$f(W_i = w_i) = \frac{\Gamma(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2})}{2^{\frac{m_i}{2}}\Gamma(\alpha_i + \frac{n_i}{2})\Gamma(\frac{m_i}{2})} \cdot \frac{(\beta_i + \frac{s_i}{2})^{\alpha_i + \frac{n_i}{2}} w_i^{\frac{m_i}{2} - 1}}{(\beta_i + \frac{s_i}{2} + \frac{w_i}{2})^{\alpha_i + \frac{n_i}{2} + \frac{m_i}{2}}},\ w_i > 0$$

$i = 1, ..., k$, and W_i's are independent.

Because we want to choose the population associated with the smallest variance $\phi_{[1]} = \min\{\phi_1, ..., \phi_k\}$, the loss function is

$L(\phi, i, n) = \phi_i - \phi_{[1]} + nc.$

Suppose $\alpha_i > 1$, $i = 1, ..., k$. To determine the optimum allocation of m more observations at stage 2 using the fixed sample-size sampling algorithm, one has to calculate $r(m_1, ..., m_k)$, that is, the look ahead Bayes risk corresponding to the allocation $(m_1, ..., m_k)$.

$$\begin{aligned} r(m_1, ..., m_k) &= E\{\min_{i=1,...,k} E\{L(\Phi, i, n+n)|S=s, W\}|S=s\} \\ &= E_s\{\min_{i=1,...,k} \frac{\beta_i + \frac{s_i}{2} + \frac{W_i}{2}}{\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1} - E\{\Phi_{[1]}|S=s, W\}\} + nc + mc \\ &= E_s\{\min_{i=1,...,k} \frac{\beta_i + \frac{s_i}{2} + \frac{W_i}{2}}{\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1}\} - E_s\{\Phi_{[1]}\} + nc + mc \end{aligned}$$

Therefore, the optimum allocation minimizes

$$E_s\{\min_{i=1,...,k} \frac{\beta_i + \frac{s_i}{2} + \frac{W_i}{2}}{\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1}\}$$

subject to $m_1 + ... + m_k = m$.

Let's consider the special case where m=1. Suppose $m_i = 1$, $m_j = 0, j \neq i$, $i \in \{1, ..., k\}$, let

$$l_i = E_s\{\min\{\min_{j\neq i} \frac{\beta_j + \frac{s_j}{2}}{\alpha_j + \frac{n_j}{2} - 1}, \frac{\beta_i + \frac{s_i}{2} + \frac{W_i}{2}}{\alpha_i + \frac{n_i}{2} + \frac{1}{2} - 1}\}\},$$

then the best allocation of the next observation is to draw one observation from the i^*th population, where $l_{i^*} = \min_{i=1,...,k}\{l_i\}$.

3.1.2 Selection of the largest normal mean with random mean and variance

In the section, we will consider the case where both mean and variance of each population are unknown.

Suppose $X_i \sim \mathrm{N}(\theta_i, \phi_i), i = 1, ..., k$, and X_i's are independent.

The prior density of Θ_i and Φ_i is $\pi(\theta_i,\phi_i)=\pi_1(\theta_i|\phi_i)\pi_2(\phi_i)$,

where $\pi_1(\theta_i|\phi_i)$ is a $N(\mu_i,\tau_i\phi_i)$ density and $\pi_2(\phi_i)$ is an $lg(\alpha_i,\beta_i)$ density.

Suppose at the first stage, $X_{i1}=x_{i1},...,X_{in_i}=x_{in_i}, i=1,...,k$, have been observed, then the updated prior density of Θ_i and Φ_i is:

$\pi(\theta_i,\phi_i|x)=\pi_1(\theta_i|\phi_i,x)\pi_2(\phi_i|x)$,

where $\pi_1(\theta_i|\phi_i,x)$ is a normal density with mean $\mu_i(x)=\frac{\mu_i+n_i\tau_i\bar{x}_i}{n_i\tau_i+1}$, ($\bar{x}_i=\frac{1}{n_i}\sum_{j=1}^{n_i}x_{ij}$) and variance $({\tau_i}^{-1}+n_i)^{-1}\phi_i$ and $\pi_2(\phi_i|x)$ is an inverted gamma density with parameters $\alpha_i+\frac{n_i}{2}$ and $\beta_i^{'}$ where

$\beta_i^{'}=\{\beta_i^{-1}+\frac{1}{2}\sum_{j=1}^{n_i}(x_{ij}-\hat{x}_i)^2+\frac{n_i(\bar{x}_i-\mu_i)^2}{2(1+n_i\tau_i)}\}^{-1}$.

Suppose at the second stage, $Y_{i1}=y_{i1},...,Y_{im_i}=y_{im_i}, i=1,...,k$, have been observed, then the posterior distribution of Θ_i and Φ_i, is

$\pi(\theta_i,\phi_i|x,y)=\pi_1(\theta_i|\phi_i,x,y)\pi_2(\phi_i|x,y)$,

where $\pi_1(\theta_i|\phi_i,x,y)$ is a normal density with mean $\gamma_i(x,y)=\frac{\mu_i(x)+m_i({\tau_i}^{-1}+n_i)^{-1}\bar{y}_i}{m_i({\tau_i}^{-1}+n_i)^{-1}+1}$, ($\bar{y}_i=\frac{1}{m_i}\sum_{j=1}^{m_i}y_{ij}$) and variance $({\tau_i}^{-1}+n_i+m_i)^{-1}\phi_i$ and $\pi_2(\phi_i|x,y)$ is an inverted gamma density with parameters $\alpha_i+\frac{n_i}{2}+\frac{m_i}{2}$ and $\beta_i^{''}$, where

$\beta_i^{''}=\{{\beta_i^{'}}^{-1}+\frac{1}{2}\sum_{j=1}^{m_i}(y_{ij}-\hat{y}_i)^2+\frac{m_i(\bar{y}_i-\mu_i(x))^2}{2(1+m_i({\tau_i}^{-1}+n_i)^{-1})}\}^{-1}$.

$(\Theta_i,\Phi_i)^T$'s, given $X=x$, $Y=y$, are independent.

According to James O. Berger's book, we know that the marginal posterior density of Θ_i, given $X=x,Y=y$, is a

$$T(2(\alpha_i+\frac{n_i}{2})+m_i,\gamma_i(x,y),(({\tau_i}^{-1}+n_i+m_i)(\alpha_i+\frac{n_i}{2}+\frac{m_i}{2})\beta_i^{''})^{-1})$$

and Θ_i's, $i = 1, ..., k$, are independent.

The marginal density of $Y_i = (Y_{i1}, ..., Y_{im_i})$ given $X = x$, $m(y_i|x)$, is

$$(2\pi)^{-\frac{m_i}{2}}(1+n_i\tau_i)^{\frac{1}{2}}(1+n_i\tau_i+m_i\tau_i)^{-\frac{1}{2}}(\Gamma(\alpha_i+\frac{n_i}{2})\beta_i'^{\alpha_i+\frac{n_i}{2}})^{-1}\Gamma(\alpha_i+\frac{n_i+m_i}{2})\beta_i^{*\alpha_i+\frac{n_i+m_i}{2}}$$

where

$$\begin{aligned}
\beta_i' &= \{\beta_i^{-1}+\frac{1}{2}\sum_{j=1}^{n_i}(x_{ij}-\hat{x}_i)^2+\frac{n_i(\bar{x}_i-\mu_i)^2}{2(1+n_i\tau_i)}\}^{-1}, \\
\beta_i^* &= \{\beta_i^{-1}+\frac{1}{2}\sum_{j=1}^{n_i}(x_{ij}-\hat{z}_i)^2+\frac{1}{2}\sum_{j=1}^{m_i}(y_{ij}-\hat{z}_i)^2+\frac{(m_i+n_i)(\bar{z}_i-\mu_i)^2}{2(1+n_i\tau_i+m_i\tau_i)}\}^{-1}, \\
\bar{z}_i &= \frac{1}{m_i+n_i}(\sum_{j=1}^{n_i}x_{ij}+\sum_{j=1}^{m_i}y_{ij}),
\end{aligned}$$

and Y_i's, given $X = x$, are independent.

Our objective is to choose the population with the largest mean.

Let the loss function be $L(\theta, i, n) = \theta_{[k]} - \theta_i + nc$, then

$$\begin{aligned}
r(m_1,\ldots,m_k) &= E_x\{\min_{i=1,\ldots,k} E\{L(\Theta, i, n+m)|X=x, Y\}\} \\
&= E_x\{\min_{i=1,\ldots,k} E\{\Theta_{[k]}-\Theta_i|X=x,Y\}\}+nc+mc \\
&= E_x\{\Theta_{[k]}\} - E_x\{\max_{i=1,\ldots,k} E\{\Theta_i|X=x,Y\}\}+nc+mc
\end{aligned}$$

Suppose $\alpha_i > \frac{1}{2}, i = 1, ..., k$, then

$$E\{\Theta_i | X = x, Y\} = \frac{\mu_i(x) + m_i({\tau_i}^{-1} + n_i)^{-1}\bar{Y_i}}{m_i({\tau_i}^{-1} + n_i)^{-1} + 1}$$

Therefore, the optimum allocation $(m_1^*, ..., m_k^*)$ maximizes, subject to $m_1 + ... + m_k = m$,

$$E_x\{\max_{i=1,...,k} \frac{\mu_i(x) + m_i({\tau_i}^{-1} + n_i)^{-1}\bar{Y_i}}{m_i({\tau_i}^{-1} + n_i)^{-1} + 1}\}$$

If we allocate the next observation to the ith population, then the expected posterior gain

$$g_i = E_x\{\max\{\max_{j \neq i} \mu_j(x), \frac{\mu_i(x) + ({\tau_i}^{-1} + n_i)^{-1}\bar{Y_i}}{({\tau_i}^{-1} + n_i)^{-1} + 1}\}\}$$

Therefore,the optimum allocation is to allocate the next observation to the population with $g_{[k]}$, where $g_{[k]} = \max_{i=1,...,k}\{g_i\}$.

3.1.3 Selection of the smallest normal variance with random mean and variance

In this section, we will consider selecting the population with the smallest variance.

According to page 288 of James O. Berger's book, the marginal posterior distribution of Φ_i, given $X = x, Y = y$, is a $lg(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2}, \beta_i^{''})$ and Φ_i's are independent .

Let the loss function be

$$L(\phi, i) = \phi_i - \phi_{[1]} + nc + mc,$$

where $\phi = (\phi_1, ..., \phi_k)$, and $\phi_{[1]} = \min\{\phi_1, ..., \phi_k\}$, then

$$\begin{aligned} r(m_1, \ldots, m_k) &= E_x\{\min_{i=1,\ldots,k} E\{L(\Phi, i, n+m)|X=x, Y\}\} \\ &= E_x\{\min_{i=1,\ldots,k} E\{\Phi_i - \Phi_{[1]} + nc + mc|X=x, Y\}\} \\ &= E_x\{\min_{i=1,\ldots,k} E\{\Phi_i|X=x, Y\}\} - E_x\{\Phi_{[1]}\} + nc + mc \end{aligned}$$

Suppose $\alpha_i > 1$, $i = 1, ..., k$, then

$$E\{\Phi_i|X=x, Y\} = \frac{1}{(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1)\beta_i''}.$$

Therefore, the optimum allocation $(m_1^*, ..., m_k^*)$ minimizes, subject to $m_1 + ... + m_k = m$,

$$\begin{aligned} & E_x\{\min_{i=1,\ldots,k} \frac{1}{(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1)\beta_i''}\} \\ = \; & E_x\{\min_{i=1,\ldots,k} \frac{\beta_i'^{-1} + \frac{1}{2}\sum_{j=1}^{m_i}(Y_{ij} - \bar{Y}_i)^2 + \frac{m_i(\bar{Y}_i - \mu_i(x))^2}{2(1+m_i(\tau_i{}^{-1}+n_i)^{-1})}}{\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1}\} \end{aligned}$$

Suppose m=1. If we allocate the next observation to the ith population, then

$$\begin{aligned} l_i &= E_x\{\min_{i=1,\ldots,k} E\{\Phi_i|X=x, Y\}\} \\ &= E_x\{\min\{\min_{j\neq i} \frac{1}{(\alpha_j + \frac{n_j}{2} - 1)\beta_j'}, \frac{\beta_i'^{-1} + \frac{(Y_{i1} - \mu_i(x))^2}{2(1+(\tau_i{}^{-1}+n_i)^{-1})}}{\alpha_i + \frac{n_i}{2} - \frac{1}{2}}\}\} \end{aligned}$$

Therefore, the optimum allocation is to allocate the next observation to the population associated with $l_{[1]} = \min\{1_1, ..., l_k\}$.

3.1.4 Selection of the normal population with the largest absolute value of mean

Suppose there are k normal populations, where population Π_i has mean θ_i and variance σ^2. θ_i is a realization of Θ_i, which follows normal distribution with mean μ_i and variance v_i^{-1}, $i = 1, ..., k$, respectively. In this section, our objective is to choose the population with the largest absolute value of mean.

The loss function is

$$L(\theta, i, n) = |\theta|_{[k]} - |\theta_i| + nc,$$

where $|\theta|_{[k]} = \max\{|\theta_1|, \cdots, |\theta_k|\}$.

Suppose at the second stage, m_i observations are to be drawn from population Π_i, $i = 1, \cdots, k$, then the look ahead Bayes risk $r(m_1, \ldots, m_k)$ is

$$
\begin{aligned}
& E_x\{\min_{i=1,\cdots,k} E\{|\Theta|_{[k]} - |\Theta_i||X=x,Y\}\} \\
= \; & E_x\{|\Theta|_{[k]}\} - E_x\{\max_{i=1,\cdots,k} E\{|\Theta_i||X=x,Y\}\} + nc + mc \\
= \; & E_x\{|\Theta|_{[k]}\} - E_x\{\max_{i=1,\cdots,k}\{\frac{\sqrt{2}}{\sqrt{\pi(\alpha_i+q_i)}} e^{-\frac{(\alpha_i\mu_i(x)+q_iY_i)^2}{2(\alpha_i+q_i)}} + \frac{\alpha_i\mu_i(x)+q_iY_i}{\alpha_i+q_i} \\
& [1-2\Phi(-\frac{\alpha_i\mu_i(x)+q_iY_i}{\sqrt{\alpha_i+q_i}})]\}\} + nc + mc
\end{aligned}
$$

Therefore, the optimal allocation maximizes

$$
E_x\{\max_{i=1,\cdots,k}\{\frac{\sqrt{2}}{\sqrt{\pi(\alpha_i+q_i)}} e^{-\frac{(\alpha_i\mu_i(x)+q_iY_i)^2}{2(\alpha_i+q_i)}} + \frac{\alpha_i\mu_i(x)+q_iY_i}{\alpha_i+q_i}[1-2\Phi(-\frac{\alpha_i\mu_i(x)+q_iY_i}{\sqrt{\alpha_i+q_i}})]\}\}
$$

subject to $m_i + \cdots + m_k = m$.

Suppose m=1. If we allocate the next observation to the ith population, then the expected gain

$$
\begin{aligned}
g_i \; = \; & E_x\{\max\{\max_{j\neq i}\{\frac{\sqrt{2}}{\sqrt{\pi\alpha_j}} e^{-\frac{\alpha_j\mu_j^2(x)}{2}} + \mu_j(x)[1-2\Phi(-\mu_j(x)\sqrt{\alpha_j})]\}, \\
& \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i+q_i)}} e^{-\frac{(\alpha_i\mu_i(x)+q_iY_i)^2}{2(\alpha_i+q_i)}} + \frac{\alpha_i\mu_i(x)+q_iY_i}{\alpha_i+q_i}[1-2\Phi(-\frac{\alpha_i\mu_i(x)+q_iY_i}{\sqrt{\alpha_i+q_i}})]\}\},
\end{aligned}
$$

$i = 1, \cdots, k$, therefore, the optimal allocation of the next observation is to draw one observation from the population corresponding to $g_{[k]}$, where $g_{[k]} = \max\{g_1, \cdots, g_k\}$.

3.1.5 Selection of the Poisson population with the smallest mean

Suppose populations Π_i can be characterized with Poisson distribution with mean λ_i, where λ_i is a realization of Λ_i, $i = 1, ..., k$.

Suppose Λ_i follows a Gamma distribution with parameters k_i and θ_i, $i = 1, ..., k$, then the pdf of Λ_i is

$$\pi(\lambda; k_i, \theta_i) = \frac{1}{{\theta_i}^k \Gamma(k_i)} \lambda^{k_i - 1} e^{-\frac{\lambda}{\theta_i}}$$

for $\lambda > 0$, where k_i, $\theta_i > 0$, $i = 1, ..., k$.

Suppose at the first stage, $n_1, ..., n_k$ observations have been drawn from population $\Pi_1, ..., \Pi_k$, respectively.

Let $x_i = \sum_{j=1}^{n_i} x_{ij}$, $i = 1, ..., k$, and $x^T = (x_1, ..., x_k)$, then the updated prior $\Lambda_i | x \sim$Gamma$(k_i + x_i, \frac{\theta_i}{n_i \theta_i + 1})$, $i = 1, ..., k$.

At the second stage, m more observations need to be drawn from these k populations. Our objective is to find the optimal allocation of these m observations among k populations.

The loss function is

$$L(\lambda, i, n) = \lambda_i - \lambda_{[1]} + nc,$$

where $\lambda = (\lambda_1, ..., \lambda_k)$, and $\lambda_{[1]} = \min\{\lambda_1, ..., \lambda_k\}$.

Suppose at the second stage, m_i observations are to be drawn from population Π_i, $i = 1, ..., k$. Let $y_i = \sum_{j=1}^{m_i} y_{ij}$, $i = 1, ..., k$, and $y^T = (y_1, ..., y_k)$, then $\Lambda_i | x, y \sim$Gamma$(k_i + x_i + y_i, \frac{\theta_i}{n_i \theta_i + m_i \theta_i + 1})$, $i = 1, ..., k$.

The marginal probability mass function of Y_i given x is

$$
\begin{aligned}
P(Y_i = y|x) &= \binom{k_i + x_i + y_i - 1}{k_i + x_i - 1} (\frac{\frac{m_i\theta_i}{n_i\theta_i+1}}{\frac{m_i\theta_i}{n_i\theta_i+1} + 1})^y (\frac{1}{\frac{m_i\theta_i}{n_i\theta_i+1} + 1})^{k_i+x_i} \\
&= \binom{k_i + x_i + y_i - 1}{k_i + x_i - 1} (\frac{m_i\theta_i}{n_i\theta_i + m_i\theta_i + 1})^y (\frac{n_i\theta_i + 1}{n_i\theta_i + m_i\theta_i + 1})^{k_i+x_i}
\end{aligned}
$$

for $y > 0$, $i = 1, ..., k$, and given $X = x$, Y_i's are independent.

The look ahead Bayes risk $r(m_1, \ldots, m_k)$ is

$$
\begin{aligned}
& E_x\{\min_{i=1,\ldots,k} E(L(\Lambda, i, n+m)|X = x, Y) \\
= \; & E_x\{\min_{i=1,\ldots,k} E(\Lambda_i|X = x, Y)\} - E_x\{\Lambda_{[1]}\} + nc + mc \\
= \; & E_x\{\min_{i=1,\ldots,k} \frac{\theta_i(k_i + x_i + Y_i)}{n_i\theta_i + m_i\theta_i + 1}\} - E_x\{\Lambda_{[1]}\} + nc + mc
\end{aligned}
$$

Therefore, the optimal allocation minimizes

$$
E_x\{\min_{i=1,\ldots,n} \frac{\theta_i(k_i + x_i + Y_i)}{n_i\theta_i + m_i\theta_i + 1}\}
$$

subject to $m_i + ... + m_k = m$.

Suppose m=1. If we allocate the next observation to the ith population, then we get

$$\begin{aligned} l_i &= E_x\{\min_{i=1,\ldots,n} \frac{\theta_i(k_i + x_i + Y_i)}{n_i\theta_i + m_i\theta_i + 1}\} \\ &= E_x\{\min\{\min_{j\neq i} \frac{\theta_j(k_j + x_j)}{n_j\theta_j + 1}, \frac{\theta_i(k_i + x_i + Y_i)}{n_i\theta_i + \theta_i + 1}\}\}. \end{aligned}$$

Therefore, the optimal allocation of the next observation will draw one observation from the population with $l_{[1]}$, where $l_{[1]} = \min\{l_1, ..., l_k\}$

3.1.6 Selection of the best Gamma population

Suppose population Π_i can be characterized with the Gamma distribution with the common shape parameter a and the inverse scale parameter θ_i, where θ_i is a realization of Θ_i and $\Theta_i \sim$Gamma (α_i, β_i) with $\alpha_i > 0$ and $\beta_i > 0$, $i = 1, ..., k$.

At the first stage, n_i observations have been drawn from population Π_i, $i = 1, ..., k$. Let $x_i = \sum_{j=1}^{n_i} x_{ij}$ (if $n_i = 0$, then $x_i = 0$),$i = 1, ..., k$, and $x^T = (x_1, ..., x_k)$, then $\Theta_i|x \sim$Gamma$(\alpha_i + n_i a, \beta_i + x_i)$, $i = 1, ..., k$.

Suppose at the second stage, m_i observations are to be drawn from population Π_i, $i = 1, ..., k$. Let $y_i = \sum_{j=1}^{m_i} y_{ij}$, $i = 1, ..., k$, and $y^T = (y_1, ..., y_k)$, then $\Theta_i|X = x, Y = y \sim$ Gamma$(\alpha_i + n_i a + m_i a, \beta_i + x_i + y_i)$.

The marginal probability density function of Y_i given $X = x$ is

$$f(Y_i = y|X = x) = \frac{\Gamma(\alpha_i + n_i a + m_i a)(\beta_i + x_i)^{\alpha_i + n_i a} y^{m_i a - 1}}{\Gamma(\alpha_i + n_i a)\Gamma(m_i a)(\beta_i + x_i + y)^{\alpha_i + n_i a + m_i a}},$$

for $y > 0$, $i = 1, ..., k$, and given x, Y_i's are independent.

Let the loss function be

$$L(\theta, i, n) = \theta_i - \theta_{[1]} + nc,$$

then the look ahead Bayes risk $r(m_1, \ldots, m_k)$ is

$$\begin{aligned}
& E_x\{\min_{i=1,\ldots,k} E(L(\Theta, i, n+m)|X = x, Y)\} \\
= \;& E_x\{\min_{i=1,\ldots,k} E(\Theta_i|X = x, Y)\} - E_x\{\Theta_{[1]}\} + nc + mc \\
= \;& E_x\{\min_{i=1,\ldots,k} \frac{\alpha_i + n_i a + m_i a}{\beta_i + x_i + Y_i}\} - E_x\{\Theta_{[1]}\} + nc + mc
\end{aligned}$$

Therefore, the optimal allocation minimizes

$$E_x\{\min_{i=1,\ldots,k} \frac{\alpha_i + n_i a + m_i a}{\beta_i + x_i + Y_i}\}$$

subject to $m_1 + ... + m_k = m$.

Suppose m=1, that is, we want to find the optimal allocation of the next observation among k populations. Let

$$l_i = E_x\{\min\{\min_{j \neq i} \frac{\alpha_j + n_j a}{\beta_j + x_j}, \frac{\alpha_i + n_i a + a}{\beta_i + x_i + Y_i}\}\}$$

then the best allocation will draw one observation from the population with $l_{[1]}$, where $l_{[1]} = \min\{l_1, ..., l_k\}$.

Especially, when a=1, that is, these k gamma populations are also exponential populations, the optimal allocation minimizes

$$E_x\{\min_{i=1,...,k} \frac{\alpha_i + n_i + m_i}{\beta_i + x_i + Y_i}\}$$

subject to $m_1 + ... + m_k = m$.

Let

$$l_i = E_x\{\min\{\min_{j \neq i} \frac{\alpha_j + n_j}{\beta_j + x_j}, \frac{\alpha_i + n_i + 1}{\beta_i + x_i + Y_i}\}\},$$

then the optimal allocation of the next observation is to allocate the next observation to the population associated with $l_{[1]}$, where $l_{[1]} = \min\{l_1, ..., l_k\}$.

3.2 Subset selection of best populations

In many situations, people need to select a subset of populations where the selected subset should contain one or more best populations based on the given criteria. If the subsets are restricted to have a fixed size t, then usually it is desired that it contain t best populations. For example, experimenters would like to select 3 best treatments at the first round of screening to reduce the total number of observations needed to make their terminal point selection. Another example where the selection of a fixed-size subset is needed is to admit the 10 best applicants into a PhD program at a University. This type of problem can be treated through moderately extending the framework of the point selection problem.

In the following two sections, we will find the optimal allocation of m observations to select b $(1 < b < k)$ normal populations with b largest means or b $(1 < b < k)$ Bernoulli populations with b largest probabilities of success among k normal or Bernoulli populations, respectively.

3.2.1 Subset selection of b largest normal means

We consider the following two-stage selection model where $X = (X_1, ..., X_k)$ can be observed at stage 1, and $Y = (Y_1, ..., Y_k)$ at stage 2. More specifically, for $\theta = (\theta_1, ..., \theta_k) \in \mathbb{R}^k$, let $X_i \sim N(\theta_i, {p_i}^{-1})$ with ${p_i}^{-1} = \sigma^2/n_i$, and $Y_i \sim N(\theta_i, {q_i}^{-1})$ with ${q_i}^{-1} = \sigma^2/m_i$, $i = 1, ..., k$, which are altogether independent. Apparently, X and Y play the role of summary statistics: X_i as the sample mean based on n_i, and Y_i as the sample mean based on m_i, observations from $N(\theta_i, \sigma^2)$, $i = 1, ..., k$, that are altogether independent. In the Bayes approach, let the means parameter $\theta = (\theta_1, ..., \theta_k)$ be the outcome of a random parameter $\Theta = (\Theta_1, ..., \Theta_k)$, where $\Theta_i \sim N(\mu_i, {v_i}^{-1})$, $i = 1, ..., k$, and they are independent.

Our objective is to choose b best populations with $1 < b < k$, from k populations. The decision space D is the set of all subsets of size b of $\{1, 2, ..., k\}$, and there are $\binom{k}{b}$ elements in that decision space. The loss function is assumed to be $L(\theta, A, n) = \sum_{i \in A}(\theta_{[k]} - \theta_i) + nc$, where $\theta_{[k]} = \max(\theta_1, ..., \theta_k)$. Let us consider the fixed total sample size allocation problem where the total number of observations n is fixed. Then the optimal allocation, i.e. the one that achieves the minimum Bayes risk, is determined by

$$\min_{n_1 + ... + n_k = n} E[\min_{A \in D} E(L(\Theta, A, n)|X)]. \tag{3.1}$$

At the end of stage 1, the optimal allocation for the second stage, with a total number of m observations allowed at stage 2, is the one that achieves

$$\min_{m_1+...+m_k=m} E\{\min_{A\in D} E(L(\Theta, A, n+m)|X=x, Y)|X=x\}. \tag{3.2}$$

To evaluate the inner conditional expectation in (Equation 3.1), we need the conditional distribution of Θ, given $X = x$ and $Y = y$, which is as follows.

$$\Theta_i \sim N(\frac{\alpha_i \mu_i(x) + q_i y_i}{\alpha_i + q_i}, \frac{1}{\alpha_i + q_i}), \tag{3.3}$$

where $\alpha_i = p_i + v_i$ and $\mu_i(x) = \frac{v_i\mu_i + p_i x_i}{v_i + p_i}$, $i = 1, ..., k$, and they are independent. The outer conditional expectation in (Equation 3.1) is w.r.t. the conditional distribution of Y, given $X = x$, which is as follows.

$$Y_i \sim N(\mu_i(x), \frac{\alpha_i + q_i}{\alpha_i q_i}), \tag{3.4}$$

$i = 1, ..., k$, and they are independent.

$$\begin{aligned}
r(m_1,\ldots,m_k) &= E\{\min_{A\in D} E\{L(\Theta,A,n+m)|X=x,Y\}|X=x\} \\
&= E\{\min_{A\in D} E\{\sum_{i\in A}(\Theta_{[k]}-\Theta_i)|X=x,Y\}|X=x\}+nc+mc \\
&= E_x\{E\{b\Theta_{[k]}|X=x,Y\}\}-E_x\{\max_{A\in D} E\{\sum_{i\in A}\Theta_i|X=x,Y\}\}+nc+mc \\
&= bE_x\{\Theta_{[k]}\}-E_x\{\max_{A\in D} E\{\sum_{i\in A}\Theta_i|X=x,Y\}\}+nc+mc
\end{aligned}$$

Therefore, the optimal allocation, $r(m_1^*,\ldots,m_k^*)$, maximizes, subject to $m_1+...+m_k=m$, the following quantity

$$\begin{aligned}
& E_x\{\max_{A\in D} E\{\sum_{i\in A}\Theta_i|X=x,Y\}\} \\
= \;& E_x\{\max_{A\in D}[\frac{\alpha_{i1}\mu_{i1}(x)+q_{i1}Y_{i1}}{\alpha_{i1}+q_{i1}}+...+\frac{\alpha_{ib}\mu_{ib}(x)+q_{ib}Y_{ib}}{\alpha_{ib}+q_{ib}}]\} \\
= \;& E\{\max_{A\in D}[\mu_{i1}(x)+(\frac{q_{i1}}{\alpha_{i1}(\alpha_{i1}+q_{i1})})^{1/2}N_{i1}+...+\mu_{ib}(x)+(\frac{q_{ib}}{\alpha_{ib}(\alpha_{ib}+q_{ib})})^{1/2}N_{ib}]\}.
\end{aligned}$$

Let $q=1/\sigma^2$. For $q_i=q$, $q_j=0$, and $j\neq i$ where $1\leq i\leq k$, we have

$$\begin{aligned}
& E\{\max_{A\in D}[\mu_{i1}(x)+(\frac{q_{i1}}{\alpha_{i1}(\alpha_{i1}+q_{i1})})^{1/2}N_{i1}+...+\mu_{ib}(x)+(\frac{q_{ib}}{\alpha_{ib}(\alpha_{ib}+q_{ib})})^{1/2}N_{ib}]\} \\
= \;& E\{\max[\max_{\{A:A\in D,\,i\notin A\}}\sum_{k\in A}\mu_k(x),\max_{\{A:A\in D,\,i\in A\}}\sum_{j\in A,\,j\neq i}\mu_j(x)+\mu_i(x)+\sigma_iN_i]\}
\end{aligned}$$

Let $\mu_{[1]}(x) < \mu_{[2]}(x) < ... < \mu_{[k]}(x)$. If $q_{(i)} = 1$ and $q_{(j)} = 0$, for $j \neq i$, then we have the following for $i \in \{1, ..., k-b\}$.

$$\begin{aligned}
g_{(i)} &= E\{\max[\max_{\{A: A \in D,\, i \notin A\}} \sum_{k \in A} \mu_{[k]}(x), \\
&\qquad \max_{\{A: A \in D,\, i \in A\}} \sum_{j \in A,\ j \neq i} \mu_{[j]}(x) + \mu_{[i]}(x) + \sigma_{(i)} N_{(i)}]\} \\
&= E\{\max[\mu_{[k-b+1]}(x) + \mu_{[k-b+2]}(x) + ... + \mu_{[k]}(x), \\
&\qquad \mu_{[k-b+2]}(x) + ... + \mu_{[k]}(x) + \mu_{[i]}(x) + \sigma_{(i)} N_{(i)}]\} \\
&= E\{\mu_{[k-b+2]}(x) + \mu_{[k-b+3]}(x) + ... + \mu_{[k]}(x) \\
&\qquad + \max[\mu_{[k-b+1]}(x), \mu_{[i]}(x) + \sigma_{(i)}(x) N_{(i)}]\} \\
&= \mu_{[k-b+2]}(x) + \mu_{[k-b+3]}(x) + ... + \mu_{[k]}(x) + \mu_{[i]}(x) \\
&\qquad + E\{\max[\mu_{[k-b+1]}(x) - \mu_{[i]}(x), \sigma_{(i)} N_{(i)}]\} \\
&= \mu_{[k-b+2]}(x) + \mu_{[k-b+3]}(x) + ... + \mu_{[k]}(x) + \mu_{[i]}(x) \\
&\qquad + \sigma_{(i)} E\{\max[\frac{\mu_{[k-b+1]}(x) - \mu_{[i]}(x)}{\sigma(i)}, N_{(i)}]\} \\
&= \mu_{[k-b+2]}(x) + \mu_{[k-b+3]}(x) + ... + \mu_{[k]}(x) + \mu_{[i]}(x) + \sigma_{(i)} T(\frac{\mu_{[k-b+1]}(x) - \mu_{[i]}(x)}{\sigma(i)})
\end{aligned}$$

On the other hand, for $i \in \{k-b+1, k-b+2, ..., k\}$,

$$
\begin{aligned}
g_{(i)} &= E\{\max[\max_{\{A:A\in D,\, i\notin A\}} \sum_{k\in A} \mu_{[k]}(x), \\
&\qquad \max_{\{A:A\in D,\, i\in A\}} \sum_{j\in A\,,\, j\neq i} \mu_{[j]}(x) + \mu_{[i]}(x) + \sigma_{(i)} N_{(i)}]\} \\
&= E\{\max[\mu_{[k-b]}(x) + \mu_{[k-b+1]}(x) + ... + \mu_{[k]}(x) - \mu_{[i]}(x), \\
&\qquad \mu_{[k-b+1]}(x) + ... + \mu_{[k]}(x) + \sigma_{(i)} N_{(i)}]\} \\
&= \mu_{[k-b+1]}(x) + ... + \mu_{[k]}(x) + E\{\max[\mu_{[k-b]}(x) - \mu_{[i]}(x), \sigma_{(i)} N_{(i)}]\} \\
&= \mu_{[k-b+1]}(x) + ... + \mu_{[k]}(x) + \sigma_{(i)} E\{\max[\frac{\mu_{[k-b]}(x) - \mu_{[i]}(x)}{\sigma_{(i)}}, N_{(i)}]\} \\
&= \mu_{[k-b+1]}(x) + ... + \mu_{[k]}(x) + \sigma_{(i)} T(\frac{\mu_{[k-b]}(x) - \mu_{[i]}(x)}{\sigma_{(i)}})
\end{aligned}
$$

Let us consider the two populations $P_{(k-b)}$ and $P_{(k-b+1)}$, since they turn out to play a special role in this situation: these are the only two populations between which a preference in terms of order relation " $<$ " can be established that does not depend on $\mu_{(1)}(x), ..., \mu_{(k)}(x)$. In fact, the following theorem shows that the next allocation is not assigned to that one of the two populations for which more prior plus sampling information has been gathered so far.

Theorem 3.2.1 *At every $X = x$, the following holds.* $\alpha_{(k-b)} > (=,<) \alpha_{(k-b+1)}$ *if and only if* $R^{(k-1)}(x) < (=,>) R^{(k-b+1)}(x)$.

$$
\begin{aligned}
g_{(k-b)}(x) &= \mu_{[k-b+2]}(x) + ... + \mu_{(k)}(x) + \mu_{[k-b]}(x) \\
&\quad + \sigma_{(k-b)} T(\frac{\mu_{[k-b+1]}(x) - \mu_{[k-b]}(x)}{\sigma_{(k-b)}})
\end{aligned}
$$

$$
\begin{aligned}
g_{(k-b+1)}(x) &= \mu_{[k-b+2]}(x) + ... + \mu_{(k)}(x) + \mu_{[k-b+1]}(x) \\
&\quad + \sigma_{(k-b+1)} T(\frac{\mu_{[k-b]}(x) - \mu_{[k-b+1]}(x)}{\sigma_{(k-b+1)}}).
\end{aligned}
$$

Because $T(w) = T(-w) + w$, we have

$$
\begin{aligned}
g_{(k-b+1)}(x) &= \mu_{[k-b+2]}(x) + ... + \mu_{(k)}(x) + \mu_{[k-b+1]}(x) \\
&\quad + \sigma_{(k-b+1)} [T(\frac{\mu_{[k-b+1]}(x) - \mu_{[k-b]}(x)}{\sigma_{(k-b+1)}}) + \frac{\mu_{[k-b]}(x) - \mu_{[k-b+1]}(x)}{\sigma_{(k-b+1)}}] \\
&= \mu_{[k-b+2]}(x) + ... + \mu_{(k)}(x) + \mu_{[k-b+1]}(x) \\
&\quad + \sigma_{(k-b+1)} T(\frac{\mu_{[k-b+1]}(x) - \mu_{[k-b]}(x)}{\sigma_{(k-b+1)}}) + \mu_{[k-b]}(x) - \mu_{[k-b+1]}(x) \\
&= \mu_{[k-b+2]}(x) + ... + \mu_{(k)}(x) + \mu_{[k-b]}(x) \\
&\quad + \sigma_{(k-b+1)} T(\frac{\mu_{[k-b+1]}(x) - \mu_{[k-b]}(x)}{\sigma_{(k-b+1)}}).
\end{aligned}
$$

The rest follows from the fact that $\gamma T(x/\gamma)$ is strictly increasing in γ for every $x \in \mathbb{R}$.

3.2.2 Subset selection of b greatest probabilities of success

In this section, our objective is to choose b best Binomial populations, that is, b populations with largest probabilities of success, where $1 < b < k$. Let the loss function be

$$L(\theta, A, n) = \sum_{i \in A} (\theta_{[k]} - \theta_i) + nc \tag{3.5}$$

where $|A| = b$, and $A \subset \{1, ..., k\}$.

Suppose at the first stage, n_i observations, $x_{i1}, \ldots, x_{in_i}$, have been drawn from population Π_i, and at the second stage, m_i observations, $Y_{i1}, \ldots, Y_{im_i}$, are to be drawn from population Π_i, $i = 1, \ldots, k$. Let $\Theta = (\Theta_1, ..., \Theta_k)$, $x_i = \sum_{i=1}^{n_i} x_{ij}$, $Y_i = \sum_{i=1}^{m_i} Y_{ij}$, $x^T = (x_1, \ldots, x_k)$ and $Y^T = (Y_1, \ldots, Y_k)$, then the look ahead Bayes risk, $r(m_1, \ldots, m_k)$, is

$$\begin{aligned}
& E_x\{\min_{A \in D} E_x\{L(\Theta, A, n+m)|Y\}\} \\
= \; & bE_x\{\Theta_{[k]}\} - E_x\{\max_{A \in D} E_x\{\sum_{i \in A} \Theta_i |Y\}\} + nc + mc \\
= \; & bE_x\{\Theta_{[k]}\} - E_x\{\max_{A \in D} \sum_{i \in A} \frac{a_i + Y_i}{a_i + b_i + m_i}\} + nc + mc
\end{aligned}$$

Therefore, the optimal allocation maximizes

$$E_x\{\max_{A \in D} \sum_{i \in A} \frac{a_i + Y_i}{a_i + b_i + m_i}\}$$

subject to $m_1 + \ldots + m_k = m$.

We know that given $X = x$, $\Theta_i \sim$Beta(a_i, b_i), where $a_i = \alpha_i + x_i$, $b_i = \beta_i + n_i - x_i$; Given $X = x$, and $Y = y$, $\Theta_i \sim$Beta$(a_i + y_i, b_i + m_i - y_i)$, $i = 1, ..., k$, and $\Theta_1, ..., \Theta_k$ are independent.

Let $\mu_{[1]} \leq \mu_{[2]} \leq ... \leq \mu_{[k]}$ be the ordered sequence of $\mu_1, ..., \mu_k$, $P_{(t)}$ be the population associated with $\mu_{[t]}$, and $a_{(t)}$, $b_{(t)}$, $m_{(t)}$, $g_{(t)}$ and $\varepsilon_{(t)}$ be associated with population $P_{(t)}$, where $\mu_t = \frac{a_t}{a_t+b_t}$, $\varepsilon_{(t)} = \frac{1}{a_{(t)}+b_{(t)}}$, $t = 1, ..., k$.

Let $m_{(i)} = 1$, and $m_{(j)} = 0$, for $j \neq i$, denote the posterior gain corresponding to this allocation by $g_{(i)}$, then

(1) for $1 \leq i \leq k - b$,

$$\begin{aligned}
g_{(i)} &= E_x\{\max_{A \in D} E_x(\sum_{i \in A} \Theta_i | Y)\} \\
&= E_x\{\max[\mu_{[k-b+1]} + ... + \mu_{[k]}, \mu_{[k-b+2]} + ... + \mu_{[k]} + \frac{a_{(i)} + Y_{(i)}}{a_{(i)} + b_{(i)} + 1}]\} \\
&= E_x\{\mu_{[k-b+2]} + ... + \mu_{[k]} + \max[\mu_{[k-b+1]}, \mu_{[k]} + \frac{a_{(i)} + Y_{(i)}}{a_{(i)} + b_{(i)} + 1}]\} \\
&= \mu_{[k-b+2]} + ... + \mu_{[k]} + \max[\mu_{[k-b+1]}, \frac{a_{(i)} + 1}{a_{(i)} + b_{(i)} + 1}]\mu_{(i)} \\
&\quad + \max[\mu_{[k-b+1]}, \frac{a_{(i)}}{a_{(i)} + b_{(i)} + 1}](1 - \mu_{(i)})
\end{aligned}$$

(2) for $k - b + 1 \leq i \leq k$

$$
\begin{aligned}
g_{(i)} &= E_x\{\max_{A\in D} E_x(\sum_{i\in A}\Theta_i|Y)\} \\
&= E_x\{\max[\mu_{[k-b]}+\mu_{[k-b+1]}+...+\mu_{[k]}-\mu_{[i]},\mu_{[k-b+1]}+...+\mu_{[k]}-\mu_{[i]} \\
&\quad +\frac{a_{(i)}+Y_{(i)}}{a_{(i)}+b_{(i)}+1}]\} \\
&= \mu_{[k-b+1]}+...+\mu_{[k]}-\mu_{[i]}+E_x\{\max[\mu_{[k-b]},\frac{a_{(i)}+Y_{(i)}}{a_i+b_i+1}]\} \\
&= \mu_{[k-b+1]}+...+\mu_{[k]}-\mu_{[i]}+\max(\mu_{[k-b]},\frac{a_{(i)}+1}{a_{(i)}+b_{(i)}+1})\mu_{[i]} \\
&\quad +\max(\mu_{[k-b]},\frac{a_{(i)}}{a_{(i)}+b_{(i)}+1})(1-\mu_{[i]})
\end{aligned}
$$

Therefore, the optimal allocation of the next observation draws one observation, with equal probabilities, from one of those populations $P_{(t)}$ with $g_{(t)}=\max\{g_{(1)},...,g_{(k)}\}$, $t=1,...,k$.

Finding all populations which are tied for maximum value of the expected posterior gains can be done through paired comparisons. One of these is seen to be different from all others: the comparison of $g_{(k-b+1)}$ and $g_{(k-b)}$ is made only through the respective fractions. A similar phenomenon is in the normal case, as is shown in the previous theorem.

$$
\begin{aligned}
g_{(k-b)} &= \mu_{[k-b+2]} + ... + \mu_{[k]} + \max[\mu_{[k-b+1]}, \frac{a_{(k-b)}+1}{a_{(k-b)}+b_{(k-b)}+1}]\mu_{[k-b]} \\
&\quad + \max[\mu_{[k-b+1]}, \frac{a_{(k-b)}}{a_{(k-b)}+b_{(k-b)}+1}](1-\mu_{[k-b]}) \\
&= \mu_{[k-b+2]} + ... + \mu_{[k]} + \max[\mu_{[k-b+1]}, \frac{\mu_{[k-b]}+\varepsilon_{(k-b)}}{1+\varepsilon_{(k-b)}}]\mu_{[k-b]} \\
&\quad + \max[\mu_{[k-b+1]}, \frac{\mu_{[k-b]}}{1+\varepsilon_{(k-b)}}](1-\mu_{[k-b]}) \\
&= \mu_{[k-b+2]} + ... + \mu_{[k]} + \max[\mu_{[k-b+1]}, \frac{\mu_{[k-b]}+\varepsilon_{(k-b)}}{1+\varepsilon_{(k-b)}}]\mu_{[k-b]} + \mu_{[k-b+1]}(1-\mu_{[k-b]}) \\
&= \mu_{[k-b+2]} + ... + \mu_{[k]} + \mu_{[k-b+1]} + \max\{0, [\frac{\mu_{[k-b]}+\varepsilon_{(k-b)}}{1+\varepsilon_{(k-b)}} - \mu_{[k-b+1]}]\mu_{[k-b]}\} \\
&= \mu_{[k-b+1]} + ... + \mu_{[k]} + \max\{0, (1-\mu_{[k-b+1]})\mu_{[k-b]} - [\frac{1-\mu_{[k-b]}}{1+\varepsilon_{(k-b)}}]\mu_{[k-b]}\}
\end{aligned}
$$

$$
\begin{aligned}
g_{(k-b+1)} &= \mu_{[k-b+1]} + ... + \mu_{[k]} - \mu_{[k-b+1]} + \max[\mu_{[k-b]}, \frac{\mu_{[k-b+1]}+\varepsilon_{(k-b+1)}}{1+\varepsilon_{(k-b+1)}}]\mu_{[k-b+1]} \\
&\quad + \max[\mu_{[k-b]}, \frac{\mu_{[k-b+1]}}{1+\varepsilon_{(k-b+1)}}](1-\mu_{[k-b+1]}) \\
&= \mu_{[k-b+2]} + ... + \mu_{[k]} + [\frac{\mu_{[k-b+1]}+\varepsilon_{(k-b+1)}}{1+\varepsilon_{(k-b+1)}}]\mu_{[k-b+1]} \\
&\quad + \max[\mu_{[k-b]}, \frac{\mu_{[k-b+1]}}{1+\varepsilon_{(k-b+1)}}](1-\mu_{[k-b+1]}) \\
&= \mu_{[k-b+2]} + ... + \mu_{[k]} + \mu_{[k-b+1]} + \frac{\mu_{[k-b+1]}-1}{1+\varepsilon_{(k-b+1)}}\,\mu_{[k-b+1]} \\
&\quad + \max\{\frac{\mu_{[k-b+1]}}{1+\varepsilon_{(k-b+1)}}\,(1-\mu_{[k-b+1]}),\,(1-\mu_{[k-b+1]})\mu_{[k-b]}\} \\
&= \mu_{[k-b+1]} + ... + \mu_{[k]} + \max\{0, (1-\mu_{[k-b+1]})\mu_{[k-b]} - [\frac{1-\mu_{[k-b+1]}}{1+\varepsilon_{(k-b+1)}}]\mu_{[k-b+1]}\}
\end{aligned}
$$

Therefore, to compare $g_{(k-b)}$ and $g_{(k-b+1)}$, we only need to compare two fractions $\frac{(1-\mu_{[k-b]})\mu_{[k-b]}}{1+\varepsilon_{(k-b)}}$ and $\frac{(1-\mu_{[k-b+1]})\mu_{[k-b+1]}}{1+\varepsilon_{(k-b+1)}}$.

3.3 Simultaneous selection and estimation

After a population has been selected, a natural follow-up question may arise. The question is how large the parameter of the selected population is. That is, we need to estimate the parameter of the selected population. Most research works have dealt with either estimation or selection problem, except those by Cohen and Sackrowitz (1988), Gupta and Miescke (1990, 1993), Bansal and Miescke (2002, 2005), and Misra, van der Meulen, and Branden (2006). All their works have been done with Bayesian approach. By incorporating loss due to selection with that due to estimation in one loss function and then letting both types of decision, selection and estimation, be subject to risk evaluation, the decision theoretic treatment leads to 'selecting after estimation' instead of 'estimating after selection', which has been pointed out by Cohen and Sackrowitz (1988).

In the following sections, we will find the optimal allocation of m more observation to solve the problem of simultaneous estimation and selection of the best parameter for both normal and Bernoulli distributions.

3.3.1 Selection and estimation of the largest normal mean

Given k normal populations $\Pi_1, ..., \Pi_k$ with a common variance σ^2 and mean $\theta_1, .., \theta_k$, respectively, we want to choose the population with the largest mean and, at the same time, estimate the selected mean. That is, our objective is to select the population which is associated with $\theta_{[k]} = \max\{\theta_1, .., \theta_k\}$ and simultaneously estimate $\theta_{[k]}$.

The loss function is assumed to be additive:

$$L(\theta, d) = A(\theta, s) + B(\theta_s, l_s)$$

where A represents the loss of selecting population Π_s at θ and B the loss of estimating θ_s by l_s.

Supposed the observed data are k independent samples of sizes $n_1, .., n_k$ from $\Pi_1, ..., \Pi_k$ with sample means $x_1, ..., x_k$, respectively. Let x be $(x_1, ..., x_k)^T$.

Since Bayes rules are used, only nonrandomized decision rules need to be considered here, which are represented as follows:

$$d(x) = (s(x), l_{s(x)}(x)),\ x \in R^k$$

where $s(x) \in \{1, ..., k\}$ is the selection rule at x and $l_i(x) \in \Omega$, $i = 1, ..., k$, is a collection of k estimates based on x for θ_i, $i = 1, ..., k$, respectively, available at selection.

Let the vector of the k unknown means be random and denoted by Θ. Under a prior distribution of it, the posterior risk at $X = x$ is

$$r(d(x)|X = x) = r_A(s(x)|x) + r_B(s(x), l_{s(x)}(x)|x),$$

where

$$r_A(s(x)|x) = E\{A(\Theta, s(x))|X = x\},$$

and

$$r_B(s(x), l_{s(x)}(x)|x) = E\{B(\Theta_{s(x)}, l_{s(x)}(x))|X = x\}.$$

Formula (Equation 3.1) and Theorem 3.2.1.

Lemma 3.3.1 Let $l_i^*(x)$ minimize $r_B(i, l_i(x)|x), i = 1, ..., k$. Furthermore, let $s^*(x)$ minimize $r_A(s(x)|x) + r_B(s(x), l^*_{s(x)}(x)|x)$. Then the Bayes decision rule, at $X = x$, is $d^*(x) = (s^*(x), l^*_{s^*(x)}(x))$.

Let's consider the following loss function

$$L_1(\theta, d, n) = a(\theta_{[k]} - \theta_s) + |\theta_s - l_s| + nc,$$

where c is the cost of sampling one observation and $a > 0$ gives relative weights to the two types of losses.

At $\theta = (\theta_1, ..., \theta_k) \in R^k$, $X_i \sim \mathrm{N}(\theta_i, {p_i}^{-1})$ and $Y_i \sim \mathrm{N}(\theta_i, {q_i}^{-1})$, are independent sample means of the samples from population Π_i at stage 1 and stage 2, respectively, $i = 1, ..., k$, which altogether are assumed to be independent, where where ${p_i}^{-1} = \frac{\sigma^2}{n_i}$ and ${q_i}^{-1} = \frac{\sigma^2}{m_i}$.

$\Theta = (\Theta_1, ..., \Theta_k)$ are random and follow a distribution where $\Theta_i \sim \mathrm{N}(\mu_i, {v_i}^{-1})$, $i = 1, ..., k$, are independent.

Given $X = x$, $Y = y$, $\Theta_i \sim \mathrm{N}(\frac{\alpha_i\mu_i(x)+q_iy_i}{\alpha_i+q_i}, \frac{1}{\alpha_i+q_i})$, $i = 1, ..., k$,

where $\alpha_i = p_i + v_i$, $\mu_i(x) = \frac{v_i\mu_i+p_ix_i}{v_i+p_i}$, $i = 1, ..., k$, and Θ_i's are independent.

The conditional distribution of Y_i, given $X = x$, is $Y_i \sim \mathrm{N}(\mu_i(x), \frac{\alpha_i+q_i}{\alpha_iq_i})$, $i = 1, ..., k$, and Y_i's are independent.

By Lemma 1, the Bayes rule employs the estimator $l_i^*(x,y) = \frac{\alpha_i \mu_i(x) + q_i y_i}{\alpha_i + q_i}$ for θ_i, $i = 1, ..., k$, and it remains to find $s^*(x)$. For any decision rule $d = (s, l_s^*)$, the posterior risk at $X = x, Y = y$, turns out to be the following for selection $s(x) = i \in \{1, ..., k\}$.

$$
\begin{aligned}
& a(E\{\Theta_{[k]}|X = x, Y = y\} - \frac{\alpha_i \mu_i(x) + q_i y_i}{\alpha_i + q_i}) + \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i + q_i)}} + nc + mc \\
= \quad & aE\{\Theta_{[k]}|X = x, Y = y\} - (al_i^*(x,y) - \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i + q_i)}}) + nc + mc
\end{aligned}
$$

Then we have the following theorem.

Theorem 3.3.2 *1 Under loss function L_1 and the normal prior considered above, the Bayes rule $d^*(x,y) = (s^*(x,y), l^*_{s^*(x,y)}(x,y))$ employs $l_i^*(x,y) = \frac{\alpha_i \mu_i(x) + q_i y_i}{\alpha_i + q_i}$, $i = 1, ..., k$, and $s^*(x,y)$ maximizes $al_i^*(x,y) - \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i + q_i)}}$, $i = 1, ..., k$.*

Let us now consider fixed total sample size allocation problems. At the end of stage 1, we have drawn n observations from among the k populations and $x = (x_1, \ldots, x_k)$ has been observed. We want to draw m more observations at the second stage. If we draw m_i observations from population Π_i, $i = 1, \ldots, k$, where $m_i \geq 0$, $i = 1, \ldots, k$ and $m_1 + \ldots + m_k = m$, then $r(m_1, \ldots, m_k)$, the corresponding look ahead Bayes risk, is the following:

$$
\begin{aligned}
& E_x\{aE\{\Theta_{[k]}|X=x,Y\} - \max_{i=1,\ldots,k}(al_i^*(x,Y) - \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i+q_i)}})\} + nc + mc \\
= \;& aE_x\{\Theta_{[k]}\} - E_x\{\max_{i=1,\ldots,k}(al_i^*(x,Y) - \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i+q_i)}})\} + nc + mc
\end{aligned}
$$

Therefore, the optimal allocation maximizes, subject to $m_1 + ... + m_k = m$,

$$
\begin{aligned}
& E_x\{\max_{i=1,\ldots,k}(al_i^*(x,Y) - \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i+q_i)}})\} \\
= \;& E\{\max_{i=1,\ldots,k}(\frac{a\alpha_i\mu_i(x) + aq_iY_i}{\alpha_i+q_i} - \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i+q_i)}})\},
\end{aligned}
$$

where $Y_i \sim N(\mu_i(x), \frac{\alpha_i+q_i}{\alpha_i q_i}), i = 1, ..., k$, independent.

Because $Y_i \sim N(\mu_i(x), \frac{\alpha_i+q_i}{\alpha_i q_i})$, we have $N_i = \frac{Y_i - \mu_i(x)}{\sqrt{\frac{\alpha_i+q_i}{\alpha_i q_i}}} \sim N(0,1)$, $i = 1, ..., k$, and they are independent.

Therefore,

$$
\begin{aligned}
& E\{\max_{i=1,\ldots,k}(\frac{a\alpha_i\mu_i(x) + aq_iY_i}{\alpha_i+q_i} - \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i+q_i)}})\} \\
= \;& E\{\max_{i=1,\ldots,k}(\frac{a\alpha_i\mu_i(x) + aq_i[\sqrt{\frac{\alpha_i+q_i}{\alpha_i q_i}}N_i + \mu_i(x)]}{\alpha_i+q_i} - \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i+q_i)}})\} \\
= \;& E\{\max_{i=1,\ldots,k}(a\mu_i(x) - \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i+q_i)}} + a\frac{\sqrt{q_i}}{\sqrt{\alpha_i(\alpha_i+q_i)}}N_i)\}
\end{aligned}
$$

Suppose $m_i = 1$, $m_j = 0$, $j \neq i$, let

$$g_i = E\{\max[\max_{j\neq i}(a\mu_j(x) - \frac{\sqrt{2}}{\sqrt{\pi\alpha_j}}), a\mu_i(x) - \frac{\sqrt{2}}{\sqrt{\pi(\alpha_i + q)}} + a\frac{\sqrt{q}}{\sqrt{\alpha_i(\alpha_i + q)}}N_i]\},$$

where $N_i \sim N(0, 1)$, and $q = \frac{1}{\sigma^2}$, for $i = 1, ..., k$, then the optimal allocation for the next observation draws one observation with equal probabilities from one of the populations Π_i for which $g_i = \max\{g_1, \ldots, g_k\}$.

3.3.2 Selection and estimation of the smallest normal variance

In this section, our objective is to choose the population with the smallest variance and, at the same time, estimate its variance.

Suppose population Π_i can be described by $X_i \sim N(\mu, \phi_i)$, $i = 1, ..., k$, and $X_i's$ are independent, where μ is known and ϕ_i is a realization of $\Phi_i \sim lg(\alpha_i, \beta_i)$, $i = 1, ..., k$.

At the first stage, n_i observations, $x_{i1}, ..., x_{in_i}$, have been drawn from population Π_i, $i = 1, ..., k$. Let $s_i = \sum_{j=1}^{n_i}(x_{ij} - \mu)^2$, $i = 1, ..., k$, then the updated prior is $\Phi_i \sim lg(\alpha_i + \frac{n_i}{2}, \beta_i + \frac{s_i}{2})$, $i = 1, ..., k$, and Φ_i's are independent.

At the second stage, suppose m_i observations, $Y_{i1}, ..., Y_{im_i}$, are to be drawn from population Π_i, i=1,...,k. Let $w_i = \sum_{j=1}^{m_i}(y_{ij} - \mu)^2$, $i = 1, ..., k$, s=$(s_1, \ldots, s_k)^T$ and w=$(w_1, \ldots, w_k)^T$. Given $S = s$, $W = w$, $\Phi_i \sim \text{lg}(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2}, \beta_i + \frac{s_i}{2} + \frac{w_i}{2})$, $i = 1, ..., k$, and Φ_i's are independent.

The marginal pdf of W_i, given $S = s$, is

$$f(W_i = w_i) = \frac{\Gamma(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2})}{2^{\frac{m_i}{2}}\Gamma(\alpha_i + \frac{n_i}{2})\Gamma(\frac{m_i}{2})} \cdot \frac{(\beta_i + \frac{s_i}{2})^{\alpha_i + \frac{n_i}{2}} w_i^{\frac{m_i}{2}-1}}{(\beta_i + \frac{s_i}{2} + \frac{w_i}{2})^{\alpha_i + \frac{n_i}{2} + \frac{m_i}{2}}},$$

for $w_i > 0$, $i = 1, ..., k$, and W_i's are independent.

Let the loss function be

$$L(\phi, (h, l_h), n) = \phi_h - \phi_{[1]} + a(\phi_h - l_h)^2 + nc,$$

where $\phi_h - \phi_{[1]}$ is the loss of selecting population Π_h at $\phi = (\phi_1, \ldots, \phi_k)$ and $(\phi_h - l_h)^2$ the loss of estimating ϕ_h by l_h, a is a positive constant giving relative weights to the two types of losses, and c is the cost of sampling one observation.

The Bayesian rule $(h^*, l^*_{h^*})$ at $S = s$ and $W = s$ minimizes $E\{\Phi_h - \Phi_{[1]} + a(\Phi_h - l_h)^2 | S = s, W = s\}$.

Suppose $\alpha_i > 2$, it is easy to see that

$$l^*_i = E\{\Phi_i | S = s, W = w\} = \frac{\beta_i + \frac{s_i}{2} + \frac{w_i}{2}}{\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1},$$

$i = 1, ..., k$, and h^* minimizes, for $i = 1, ..., k$,

$$\frac{\beta_i + \frac{s_i}{2} + \frac{w_i}{2}}{\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1} - E\{\Phi_{[1]} | S = s, W = w\} + a\frac{(\beta_i + \frac{s_i}{2} + \frac{w_i}{2})^2}{(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1)^2(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 2)}.$$

Then the Bayesian risk $r(m_1, \ldots, m_k)$ for this allocation is

$$
\begin{aligned}
& E_s\{\min_{i\in\{1,\ldots,k\}}(\frac{\beta_i+\frac{s_i}{2}+\frac{W_i}{2}}{\alpha_i+\frac{n_i}{2}+\frac{m_i}{2}-1}-E\{\Phi_{[1]}|W,S=s\} \\
& +a\frac{(\beta_i+\frac{s_i}{2}+\frac{W_i}{2})^2}{(\alpha_i+\frac{n_i}{2}+\frac{m_i}{2}-1)^2(\alpha_i+\frac{n_i}{2}+\frac{m_i}{2}-2)})\}+nc+mc \\
= \; & E_s\{\min_{i\in\{1,\ldots,k\}}(\frac{\beta_i+\frac{s_i}{2}+\frac{W_i}{2}}{\alpha_i+\frac{n_i}{2}+\frac{m_i}{2}-1}+a\frac{(\beta_i+\frac{s_i}{2}+\frac{W_i}{2})^2}{(\alpha_i+\frac{n_i}{2}+\frac{m_i}{2}-1)^2(\alpha_i+\frac{n_i}{2}+\frac{m_i}{2}-2)})\} \\
& -E_s\{\Phi_{[1]}\}+nc+mc.
\end{aligned}
$$

The optimum allocation $(m_1^*,...,m_k^*)$ minimizes

$$
E_s\{\min_{i\in\{1,\ldots,k\}}(\frac{\beta_i+\frac{s_i}{2}+\frac{W_i}{2}}{\alpha_i+\frac{n_i}{2}+\frac{m_i}{2}-1}+a\frac{(\beta_i+\frac{s_i}{2}+\frac{W_i}{2})^2}{(\alpha_i+\frac{n_i}{2}+\frac{m_i}{2}-1)^2(\alpha_i+\frac{n_i}{2}+\frac{m_i}{2}-2)})\}
$$

for any allocation $(m_1,...,m_k)$ subject to $m_1+...+m_k=m$.

For $i=1,\ldots,k$, let

$$
\begin{aligned}
g_i = E_s\{\min[\min_{j\neq i}(\frac{\beta_j+\frac{s_j}{2}}{\alpha_j+\frac{n_j}{2}-1}+a\frac{(\beta_j+\frac{s_j}{2})^2}{(\alpha_j+\frac{n_j}{2}-1)^2(\alpha_j+\frac{n_j}{2}-2)}), \\
\frac{\beta_i+\frac{s_i}{2}+\frac{W_i}{2}}{\alpha_i+\frac{n_i}{2}+\frac{1}{2}-1}+a\frac{(\beta_i+\frac{s_i}{2}+\frac{W_i}{2})^2}{(\alpha_i+\frac{n_i}{2}+\frac{1}{2}-1)^2(\alpha_i+\frac{n_i}{2}+\frac{1}{2}-2)}]\}
\end{aligned}
$$

then the optimum allocation is to draw the next observation, with equal probabilities, from one of the populations Π_i for which $g_i=\min\{g_1,\ldots,g_k\}$.

3.3.3 Selection and estimation of the largest probability of success

In this section, our objective is to choose, among k independent populations, the one with the largest probability of success and at the same time, estimate its probability of success.

Suppose population Π_i follows a Bernoulli distribution with probability of success θ_i, where θ_i is a realization of $\Theta_i \sim \text{Be}(\alpha_i, \beta_i)$ with $\alpha_i > 0$ and $\beta_i > 0$, for $i = 1, ..., k$, and Θ_i's are independent.

Suppose at the first stage, n_i observations, $x_{i1}, \ldots, x_{in_i}$, have been drawn from population Π_i for $i = 1, \ldots, k$, where $n_1 + \ldots + n_k = n$. Let $x_i = \sum_{i=1}^{n_i} x_{ij}$,and $x = (x_1, \ldots, x_k)^T$. At the second stage, we are to draw m_i observations, $Y_{i1}, \ldots, Y_{im_i}$, from population Π_i for $i = 1, \ldots, k$, where $m_1 + \ldots + m_k = m$. Let $Y_i = \sum_{i=1}^{m_i} Y_{ij}$, and $Y = (Y_1, \ldots, Y_k)^T$.

It is easy to see that, given $X = x, Y = y$, $\Theta_i \sim \text{Be}(a_i + y_i, b_i + m_i - y_i)$, for $i = 1, \ldots, k$ and Θ_i's are independent, where $a_i = \alpha_i + x_i$, and $b_i = \beta_i + n_i - x_i$, for $i = 1, ..., k$.

The conditional pmf of Y given $X = x$ is

$$P(Y_i = y_i) = \frac{\Gamma(a_i + b_i)}{\Gamma(a_i)\Gamma(b_i)} \cdot \frac{\Gamma(a_i + y_i)\Gamma(b_i + m_i - y_i)}{\Gamma(a_i + b_i + m_i)} \binom{m_i}{y_i},$$

where $y_i = 0, ..., m_i$, for $i = 1, ..., k$, which is a Pólya-Eggenberger distribution with four parameters m_i, a_i, b_i and 1. Y_i's are independent.

Since our objective is to choose the population Π_i with $\theta_i = \max\{\theta_1, ..., \theta_k\}$ and to simultaneously estimate θ_i, let the loss function be

$$L(\theta, d, n) = \theta_{[k]} - \theta_s + a(\theta_s - l_s)^2 + nc,$$

where $\theta_{[k]} - \theta_s$ is the loss of selecting population Π_s at $\theta = (\theta_1, ..., \theta_k)$, and $(\theta_s - l_s)^2$ the loss of estimating θ_s by l_s, a is a positive constant giving relative weights to the two types of losses, and c is the cost of sampling one observation.

By Lemma 1, at X=x and Y=y, the Bayes rule employs the estimator $l_i^*(x, y) = \frac{a_i + y_i}{a_i + b_i + m_i}$ for θ_i, $i = 1, ..., k$, and $s^*(x)$ minimizes, for $i = 1, ..., k$,

$$E\{\Theta_{[k]}|X = x, Y = y\} - \frac{a_i + y_i}{a_i + b_i + m_i} + a\frac{(a_i + y_i)(b_i + m_i - y_i)}{(a_i + b_i + m_i)^2(a_i + b_i + m_i + 1)} + nc + mc$$

Therefore, $r(m_1, .., m_k)$, the look-ahead Bayes risk for allocation $(m_1, .., m_k)$, is the following:

$$\begin{aligned}
& E_x\{E\{\Theta_{[k]}|X = x, Y\} - \max_{i=1,...,k}(\frac{a_i + Y_i}{a_i + b_i + m_i} \\
& -a\frac{(a_i + Y_i)(b_i + m_i - Y_i)}{(a_i + b_i + m_i)^2(a_i + b_i + m_i + 1)})\} + nc + mc \\
= \quad & E_x\{\Theta_{[k]}\} - E_x\{\max_{i=1,...,k}(\frac{a_i + Y_i}{a_i + b_i + m_i} - a\frac{(a_i + Y_i)(b_i + m_i - Y_i)}{(a_i + b_i + m_i)^2(a_i + b_i + m_i + 1)})\} \\
& +nc + mc
\end{aligned}$$

Thus, the optimal allocation $(m_1^*, .., m_k^*)$ maximizes

$$E_x\{\max_{i=1,\dots,k}(\frac{a_i + Y_i}{a_i + b_i + m_i} - a\frac{(a_i + Y_i)(b_i + m_i - Y_i)}{(a_i + b_i + m_i)^2(a_i + b_i + m_i + 1)})\}$$

subject to $m_1 + ... + m_k = m$.

For $i = 1, ..., k$, let

$$g_i = E_x\{\max[\max_{j\neq i}(\frac{a_j}{a_j + b_j} - a\frac{a_j b_j}{(a_j + b_j)^2(a_j + b_j + 1)}),$$

$$\frac{a_i + Y_i}{a_i + b_i + 1} - a\frac{(a_i + Y_i)(b_i + 1 - Y_i)}{(a_i + b_i + 1)^2(a_i + b_i + 2)}]\}$$

$$= \max[\max_{j\neq i}(\frac{a_j}{a_j + b_j} - a\frac{a_j b_j}{(a_j + b_j)^2(a_j + b_j + 1)}),$$

$$\frac{a_i + 1}{a_i + b_i + 1} - a\frac{(a_i + 1)b_i}{(a_i + b_i + 1)^2(a_i + b_i + 2)}]\frac{a_i}{a_i + b_i}$$

$$+ \max[\max_{j\neq i}(\frac{a_j}{a_j + b_j} - a\frac{a_j b_j}{(a_j + b_j)^2(a_j + b_j + 1)}),$$

$$\frac{a_i}{a_i + b_i + 1} - a\frac{a_i(b_i + 1)}{(a_i + b_i + 1)^2(a_i + b_i + 2)}]\frac{b_i}{a_i + b_i},$$

then the optimal allocation of the next observation draws one observation, with equal probabilities, from one of the populations Π_i with $g_i = \max\{g_1, \ldots, g_k\}$.

CHAPTER 4

SELECTION OF THE BEST POPULATION(S) COMPARED WITH CONTROL

4.1 Introduction

Although the experimenter is generally interested in selecting the best population(s) of the competing ones, in certain conditions, even the best population may not be good enough to warrant the experimenter's selecting it. For example, if we want to choose the most effective one from among the k competing new treatments, the best treatment will not be worth considering unless its mean effect reaches a specified level or it is better than the mean effect of the treatment currently used. Therefore, the problem of simultaneous comparison of k given experimental populations among themselves and with a standard is of practical interest. This problem has been studied by many researchers under different types of formulations with different loss functions. When the true value of the parameter of the standard population is not known, it is necessary to take a random sample from it and this population is called the control. We can differentiate these two situations by referring to them as the "specified standard" and "variable control" cases. However, it is convenient to refer to the population in either case as the control although at the expense of some precision.

4.2 Selection of the best population compared with a control

Suppose there are k independent populations and one control population, we are interested in selecting the best population compared with the control. In the following sections, we will try to find the optimal allocation of m observations at the second stage to select the best normal, Bernoulli, Poisson or Gamma population compared with a control.

4.2.1 Selection of the best normal population

Suppose there are k normal populations, where population Π_i has mean θ_i and variance σ^2 for $i = 1, \ldots, k$. There is also a control normal distribution Π_0 with mean θ_0 and variance σ^2, where θ_0 is known. If $\theta_i > \theta_0$, then population Π_i is considered to be better than Π_0. We want to choose the best normal population compared with the control. If $\theta_{[k]} \leq \theta_0(\theta_{[k]} = \max\{\theta_1, \ldots, \theta_k\})$, that is, there is no population better than control, we just choose Π_0.

The selection rule a is a measurable mapping from the sample space to $[0,1]^{k+1}$, where a_i is the probability of selecting population Π_i as the best population compared with the control and $\sum_{i=0}^{k} a_i = 1$. Let D be the decision space, that is, the set consisting of all selection rules.

The loss function is

$$L(\theta, a, n) = \max(\theta_{[k]}, \theta_0) - \sum_{i=0}^{k} a_i \theta_i + nc$$

Suppose at the second stage, m_i observations are to be drawn from population Π_i, $i = 1, ..., k$, where $m_1 + \cdots + m_k = m$, then the look ahead Bayes risk $r(m_1, \ldots, m_k)$ is

$$
\begin{aligned}
& E_x\{\min_{a\in D} E\{L(\Theta, a(x,Y))|X=x,Y\}\} + nc + mc \\
= \ & E_x\{\min_{a\in D} E\{\max(\Theta_{[k]},\theta_0) - a_0(Y)\theta_0 - \sum_{i=1}^{k} a_i\Theta_i|X=x,Y\}\} \\
& +nc + mc \\
= \ & E_x\{E\{\max(\Theta_{[k]},\theta_0)|X=x,Y\} - \max_{a\in D}[a_0(Y)\theta_0 + \sum_{i=1}^{k} a_i(Y)E(\Theta_i|X=x,Y)]\} \\
& +nc + mc \\
= \ & E_x\{\max(\Theta_{[k]},\theta_0)\} - E_x\{\max_{a\in D}[a_0(Y)\theta_0 + \sum_{i=1}^{k} a_i(Y)E(\Theta_i|X=x,Y)]\} + nc + mc \\
= \ & E_x\{\max(\Theta_{[k]},\theta_0)\} - E_x\{\max[\theta_0, \max_{i=1,\cdots,k} \frac{\alpha_i\mu_i(x) + q_iY_i}{\alpha_i + q_i}]\} + nc + mc
\end{aligned}
$$

Therefore, the optimal allocation maximizes

$$
E_x\{\max[\theta_0, \max_{i=1,\cdots,k} \frac{\alpha_i\mu_i(x) + q_iY_i}{\alpha_i + q_i}]\}
$$

subject to $m_1 + \cdots + m_k = m$.

For $i = 1, \ldots, k$, let

$$
g_i = E_x\{\max[\theta_0, \max_{j\neq i} \mu_j(x), \frac{\alpha_i\mu_i(x) + qY_i}{\alpha_i + q}]\},
$$

then the optimal allocation of the next observation draws one observation, with equal probabilities, from one of the populations with $g_{[k]}$, where $g_{[k]} = \max\{g_1, \cdots, g_k\}$.

4.2.2 Selection of the best normal population in terms of variance

Suppose there are k normal populations, where population Π_i has a common mean μ and variance ϕ_i for $i = 1, \ldots, k$. There also exists a control normal population Π_0 with mean μ and a known variance ϕ_0. ϕ_i is a realization of Φ_i, which follows the inverse gamma distribution with shape parameter α_i and scale parameter β_i, $i = 1, ..., k$.

If $\phi_i < \phi_0$, population Π_i is considered better than Π_0. Our objective is to select the population with the smallest variance among these k populations and better than the control. If $\min\{\phi_1, \cdots, \phi_k\} > \phi_0$, that is, there is no population better than the control, we will choose Π_0 as our best population.

Let $\phi = (\phi_1, \cdots, \phi_k)^T$, $a = (a_0, \cdots, a_k)^T$, and the loss function be

$$L(\phi, a, n) = \sum_{i=0}^{k} a_i \phi_i - \min\{\phi_0, \phi_{[1]}\} + nc$$

where a_i is the probability that Π_i is the best population compared with the control and $\sum_{i=0}^{k} a_i = 1$. Let D be the decision space, that is, the set consisting of all selection rules.

Suppose at the second stage, m_i observations are to be drawn from population Π_i, $i = 1, \cdots, k$, then $r(m_1, \ldots, m_k)$, the look ahead Bayes risk corresponding to this allocation, is

$$
\begin{aligned}
& E_x\{\min_{\delta\in D} E\{\delta_0(x,Y)\phi_0 + \sum_{i=1}^{k}\delta_i(x,Y)\Phi_i - \min(\Phi_{[1]},\phi_0)|X=x,Y\}\} + nc + mc \\
= \; & E_x\{\min_{\delta\in D} E\{\delta_0(Y)\phi_0 + \sum_{i=1}^{k}\delta_i(Y)\Phi_i|X=x,Y\} - E\{\min(\Phi_{[1]},\phi_0)|X=x,Y\}\} \\
& +nc+mc \\
= \; & E_x\{\min_{\delta\in D} E\{\delta_0(Y)\phi_0 + \sum_{i=1}^{k}\delta_i(Y)\Phi_i|X=x,Y\}\} - E_x\{\min(\Phi_{[1]},\phi_0)\} \\
& +nc+mc \\
= \; & E_x\{\min[\phi_0, \min_{i=1,\ldots,k} E(\Phi_i|X=x,Y)]\} - E_x\{\min(\Phi_{[1]},\phi_0)\} + nc + mc \\
= \; & E_x\{\min[\phi_0, \min_{i=1,\ldots,k} \frac{\beta_i + \frac{s_i}{2} + \frac{W_i}{2}}{\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1}]\} - E_x\{\min(\Phi_{[1]},\phi_0)\} + nc + mc.
\end{aligned}
$$

Therefore, the optimal allocation minimizes

$$E_x\{\min[\phi_0, \min_{i=1,\ldots,k} \frac{\beta_i + \frac{s_i}{2} + \frac{W_i}{2}}{\alpha_i + \frac{n_i}{2} + \frac{m_i}{2} - 1}]\}$$

subject to $m_1 + ... + m_k = m$.

For $i = 1, \ldots, k$, let

$$l_i = E_x\{\min[\phi_0, \min_{j\neq i} \frac{\beta_j + \frac{s_j}{2}}{\alpha_j + \frac{n_j}{2} - 1}, \frac{\beta_i + \frac{s_i}{2} + \frac{W_i}{2}}{\alpha_i + \frac{n_i}{2} - \frac{1}{2}}]\},$$

then the optimal allocation of the next observation draws one observation, with equal probabilities, from one of the populations with $l_{[1]}$, where $l_{[1]} = \min\{l_1, ..., l_k\}$.

4.2.3 Selection of the best Bernoulli population

There are k Bernoulli populations, where population Π_i has the probability of success θ_i. There is also a control Bernoulli distribution with probability of success θ_0, where θ_0 is known. If $\theta_i > \theta_0$, then population Π_i is considered to be better than Π_0. Here, our objective is to choose the best Bernoulli population compared with the control. If $\theta_{[k]} \leq \theta_0$, that is, there is no population better than control, we just choose Π_0.

The selection rule a is a measurable mapping from the sample space to $[0,1]^{k+1}$, where a_i is the probability of selecting population Π_i as the best population compared with the control and $\sum_{i=0}^{k} a_i = 1$. D is the decision space consisting of all selection rules.

The loss function is

$$L(\theta, a, n) = \max[\theta_{[k]}, \theta_0] - \sum_{i=0}^{k} a_i\theta_i + nc$$

Suppose at the second stage, m_i observations are to be drawn from population Π_i, for $i = 1, ..., k$, where $m_1 + \ldots + m_k = m$, then the look ahead Bayes risk for this allocation is

$$\begin{aligned}
& E_x\{\min_{a\in D} E\{\max[\Theta_{[k]}, \theta_0] - a_0(x,Y)\theta_0 - \sum_{i=1}^{k} a_i(x,Y)\Theta_i | X = x, Y\}\} + nc + mc \\
= \; & E_x\{\max[\Theta_{[k]}, \theta_0]\} - E_x\{\max_{a\in D}\{a_0(x,Y)\theta_0 + \sum_{i=1}^{k} a_i(x,Y)E\{\Theta_i | X = x, Y\}\}\} \\
& +nc + mc \\
= \; & E_x\{\max[\Theta_{[k]}, \theta_0]\} - E_x\{\max[\theta_0, \max_{i=1,\cdots,k} \frac{a_i + Y_i}{a_i + b_i + m_i}]\} + nc + mc
\end{aligned}$$

Therefore, the optimal allocation maximizes

$$E_x\{\max[\theta_0, \max_{i=1,\cdots,k} \frac{a_i + Y_i}{a_i + b_i + m_i}]\}$$

subject to $m_1 + \cdots + m_k = m$.

For $i = 1, \ldots, k$, let

$$g_i = E_x\{\max[\theta_0, \max_{j \neq i} \frac{a_j}{a_j + b_j}, \frac{a_i + Y_i}{a_i + b_i + 1}]\},$$

then the optimal allocation of the next observation draws one observation, with equal probabilities, from one of the populations with $g_{[k]}$, where $g_{[k]} = \max\{g_1, \cdots, g_k\}$.

4.2.4 Selection of the best Poisson population

Suppose population Π_i follows a Poisson distribution with mean λ_i, for $i = 1, ..., k$. Π_0 is the control population with know mean λ_0.

Our objective is to select the population Π_i with $\lambda_i = \min\{\lambda_1, ..., \lambda_k\}$ and $\lambda_i < \lambda_0$, where $1 \leq i \leq k$. If there is no such population existing, we just choose Π_0 as our best population.

Let the loss function be

$$L(\lambda, \delta, n) = \sum_{i=0}^{k} \delta_i \lambda_i - \min(\lambda_{[1]}, \lambda_0) + nc.$$

At the first stage, n_i observations, $x_{i1}, \ldots, x_{in_i}$, have been drawn from population Π_i, for $i = 1, \ldots, k$, where $n_1 + \ldots + n_k = n$. At the second stage, suppose we are to draw m_i observations, $Y_{i1}, \ldots, Y_{im_i}$, from population Π_i, $i = 1, \ldots, k$, where $m_1 + \ldots + m_k = m$.

Let $x_i = \sum_{j=1}^{n_i} x_{ij}$, $x = (x_1, \ldots, x_k)^T$, $Y_i = \sum_{j=1}^{m_i} Y_{ij}$, $Y = (Y_1, \ldots, Y_k)^T$, and D be the decision space consisting of all selection rules, then the look ahead Bayes risk corresponding to the allocation $(m_1, \ldots, m_k)$ is

$$
\begin{aligned}
& E_x\{\min_{\delta \in D} E\{\delta_0(x,Y)\lambda_0 + \sum_{i=1}^k \delta_i(x,Y)\Lambda_i - \min(\Lambda_{[1]}, \lambda_0)|X = x, Y\}\} + nc + mc \\
= \; & E_x\{\min_{\delta \in D} E\{\delta_0(x,Y)\lambda_0 + \sum_{i=1}^k \delta_i(x,Y)\Lambda_i|X = x, Y\} - E\{\min(\Lambda_{[1]}, \lambda_0)|X = x, Y\}\} \\
& + nc + mc \\
= \; & E_x\{\min_{\delta \in D} E\{\delta_0(Y)\lambda_0 + \sum_{i=1}^k \delta_i(Y)\Lambda_i|X = x, Y\}\} - E_x\{\min(\Lambda_{[1]}, \lambda_0)\} + nc + mc \\
= \; & E_x\{\min[\lambda_0, \min_{i=1,\ldots,k} E\{\Lambda_i|X = x, Y\}]\} - E_x\{\min(\Lambda_{[1]}, \lambda_0)\} + nc + mc \\
= \; & E_x\{\min[\lambda_0, \min_{i=1,\ldots,k} \frac{\theta_i(k_i + x_i + Y_i)}{n_i\theta_i + m_i\theta_i + 1}]\} - E_x\{\min(\Lambda_{[1]}, \lambda_0)\} + nc + mc
\end{aligned}
$$

where the probability density function of Y_i given X=x has been given previously for $i = 1, \ldots, k$, and Y_i's are independent.

Therefore, the optimal allocation minimizes

$$
E_x\{\min[\lambda_0, \min_{i=1,\ldots,k} \frac{\theta_i(k_i + x_i + Y_i)}{n_i\theta_i + m_i\theta_i + 1}]\}
$$

subject to $m_1 + ... + m_k = m$.

For $i = 1, ..., k$, let

$$l_i = E_x\{\min[\lambda_0, \min_{j \neq i} \frac{\theta_j(k_j + x_j)}{n_j\theta_j + 1}, \frac{\theta_i(k_i + x_i + Y_i)}{n_i\theta_i + \theta_i + 1}]\},$$

then the optimal allocation of the next observation draws one observation, with equal probabilities, from one of the populations with $l_{[1]}$, where $l_{[1]} = \min\{l_1, ..., l_k\}$.

4.2.5 Selection of the best Gamma population

Suppose there are k populations which can be characterized by k Gamma distributions with the same, respectively. Given $\Theta_i = \theta_i$, $X_i \sim$ Gamma (a, θ_i) and $\Theta_i \sim$Gamma (α_i, β_i), for $i = 1, ..., k$. The control population Π_0 follows Gamma(a, θ_0), where θ_0 is a constant.

Suppose at the first stage, n_i observations have been drawn from population Π_i, for $i = 1, \ldots, k$, where $n_1 + \ldots + n_k = n$. We want to find the optimal allocation of the m observations among the k populations at the second stage to select the best population comparing with the standard. For $1 \leq i \leq k$, population Π_i is said to be the best comparing with the standard if $\theta_i = \min\{\theta_i, ..., \theta_k\}$ and $\theta_i < \theta_0$. If there is no such population existing, we will choose Π_0 as our best population.

Let D be the decision space consisting of all selection rules and the loss function be

$$L(\theta, \delta, n) = \sum_{i=0}^{k} \delta_i\theta_i - \min(\theta_{[1]}, \theta_0) + nc,$$

where c is the cost of sampling one observation.

Suppose m_i observations are to be drawn from population Π_i, for $i = 1, \ldots, k$, with $m_1 + \ldots + m_k = m$, then $r(m_1, \ldots, m_k)$, the look ahead Bayes risk for this allocation, is

$$E_x\{\min_{\delta \in D} E\{\sum_{i=0}^{k} \delta_i(Y)\Theta_i - \min(\Theta_{[1]}, \Theta_0)|X = x, Y\}\} + nc + mc$$

$$= E_x\{\min[\theta_0, \min_{i=1,\ldots,k} E\{\Theta_i|X = x, Y\}]\} - E_x\{\min(\Theta_{[1]}, \Theta_0)\} + nc + mc$$

$$= E_x\{\min[\theta_0, \min_{i=1,\ldots,k} \frac{\alpha_i + n_i a + m_i a}{\beta_i + x_i + Y_i}]\} - E_x\{\min(\Theta_{[1]}, \Theta_0)\} + nc + mc,$$

where the probability density function of Y_i given X=x has been given previously, for $i = 1, \ldots, k$, and Y_i's are independent.

Therefore, the optimal allocation minimizes

$$E_x\{\min[\theta_0, \min_{i=1,\ldots,k} \frac{\alpha_i + n_i a + m_i a}{\beta_i + x_i + Y_i}]\}$$

subject to $m_1 + ... + m_k = m$.

For $i = 1, \ldots, k$, let

$$l_i = E_x\{\min[\theta_0, \min_{j \neq i} \frac{\alpha_j + n_j a}{\beta_j + x_j}, \frac{\alpha_i + n_i a + a}{\beta_i + x_i + Y_i}]\}$$

then the optimal allocation of the next observation draws one observation, with equal probabilities, from one of the populations with $l_{[1]}$, where $l_{[1]} = \min\{l_1, ..., l_k\}$.

Especially, if the populations are exponentially distributed, that is, $a = 1$, then the optimal allocation minimizes

$$E_x\{\min[\theta_0, \min_{i=1,\ldots,k} \frac{\alpha_i + n_i + m_i}{\beta_i + x_i + Y_i}]\}$$

subject to $m_1 + ... + m_k = m$.

For $i = 1, \ldots, k$, let

$$l_i^e = E_x\{\min[\theta_0, \min_{j \neq i} \frac{\alpha_j + n_j}{\beta_j + x_j}, \frac{\alpha_i + n_i + 1}{\beta_i + x_i + Y_i}]\}$$

then the optimal allocation of the next observation draws one observation, with equal probabilities, from one of the populations with $l_{[1]}^e$, where $l_{[1]}^e = \min\{l_1^e, ..., l_k^e\}$.

4.3 Selection of all good populations compared with a control while excluding bad populations

In certain practical situations, people may be interested in selecting all good populations while excluding bad ones compared with a control. For example, suppose there are k independent newly developed manufacturing processes, we want to find out all manufacturing processes whose performance is no worse than the specified standard and exclude those with worse performance. We can also make further selection based on the selection result. In the following sections, we will find out the optimal allocation of m observations at the second stage to select good normal populations while excluding bad ones under several conditions.

4.3.1 Selection of all good normal populations

In this section, our objective is to find the optimal allocation of m observations among k populations at the second stage to select all good normal populations and exclude all bad ones.

In the first part, we compare each population with its own control, while in the second part, we compare all populations with a common control. Therefore, we need to set two different scenarios and use different losses functions.

4.3.1.1 Comparing each population with its own control

Suppose there are k normal populations, where population Π_i has mean θ_i and variance σ^2 for $i = 1, ..., k$. For each population Π_i, there also exists a control normal population $\Pi_{0,i}$ with a known mean $\theta_{0,i}$ and variance σ^2, $i = 1, ..., k$. θ_i is a realization of Θ_i, which follows the normal distribution with mean μ_i and variance v_i^{-1}, for $i = 1, ..., k$, and Θ_i's are independent.

For $i = 1, ..., k$, if $\theta_i \geq \theta_{0,i} + d_i$, population Π_i is considered better than Π_0, if $\theta_i < \theta_{0,i}$, we say population Π_i worse than Π_0, where d_i is a nonnegative constant. Our objective is to find all populations better than their respective control and exclude all bad ones.

Let the loss function be

$$L(\theta, A, n) = \sum_{j \notin A} L_{1,j} I_{[\theta_{0,j}+d_j,\infty)}(\theta_j) + \sum_{j \in A} L_{2,j} I_{(-\infty,\theta_{0,j})}(\theta_j) + nc,$$

where A is a subset of {1,. . . , k} consisting of numbers corresponding to selected populations, $L_{1,j} > 0$ is the loss of not selecting population Π_j when it is better than its control, while $L_{2,j} > 0$ is the loss of selecting population Π_j when it is a bad population, for $j = 1, \ldots, k$, and c is the cost of sampling one observation.

Suppose, at the second stage, m_i observations are to be drawn from population Π_i, $i = 1, \cdots, k$, where $m_1 + \ldots + m_k = m$. Let D be the set consisting of all subsets of $\{1, \ldots, k\}$, then the Bayes risk is

$$
\begin{aligned}
& E_x\{\min_{A\in D} E\{\sum_{j\notin A} L_{1,j} I_{[\theta_{0,j}+d_j,\infty)}(\theta_j) + \sum_{j\in A} L_{2,j} I_{(-\infty,\theta_{0,j})}(\theta_j)|X=x,Y\}\} \\
& +nc+mc \\
= \; & E_x\{\min_{A\in D}\{\sum_{j\notin A} L_{1,j} P(\Theta_j \geq \theta_{0,j}+d_j|X=x,Y) + \sum_{j\in A} L_{2,j} P(\Theta_j < \theta_{0,j}|X=x,Y)\}\} \\
& +nc+mc \\
= \; & E_x\{\min_{A\in D}\{\sum_{j\notin A} L_{1,j}\Phi(\frac{\frac{\alpha_j\mu_j(x)+q_jY_j}{\alpha_j+q_j} - \theta_{0,j} - d_j}{\sqrt{\frac{1}{\alpha_j+q_j}}}) + \sum_{j\in A} L_{2,j}\Phi(\frac{\theta_{0,j} - \frac{\alpha_j\mu_j(x)+q_jY_j}{\alpha_j+q_j}}{\sqrt{\frac{1}{\alpha_j+q_j}}})\}\} \\
& +nc+mc \\
= \; & E_x\{\min_{A\in D}\{\sum_{j\notin A} L_{1,j}\Phi(\frac{\alpha_j\mu_j(x)+q_jY_j}{\sqrt{\alpha_j+q_j}} - (\theta_{0,j}+d_j)\sqrt{\alpha_j+q_j}) \\
& +\sum_{j\in A} L_{2,j}\Phi(\theta_{0,j}\sqrt{\alpha_j+q_j} - \frac{\alpha_j\mu_j(x)+q_jY_j}{\sqrt{\alpha_j+q_j}})\}\} + nc + mc
\end{aligned}
$$

where $\Phi(x)$ is the cumulative distribution function of standard normal distribution, and the probability density function of Y_i given X=x has been given previously, for $i = 1, \ldots, k$, and Y_i's are independent.

Therefore, the optimal allocation minimizes

$$
\begin{aligned}
& E_x\{\min_{A\in D}\{\sum_{j\notin A} L_{1,j}\Phi(\frac{\alpha_j\mu_j(x)+q_jY_j}{\sqrt{\alpha_j+q_j}} - (\theta_{0,j}+d_j)\sqrt{\alpha_j+q_j}) \\
& +\sum_{j\in A} L_{2,j}\Phi(\theta_{0,j}\sqrt{\alpha_j+q_j} - \frac{\alpha_j\mu_j(x)+q_jY_j}{\sqrt{\alpha_j+q_j}})\}\}
\end{aligned}
$$

subject to $m_1 + ... + m_k = m$.

4.3.1.2 Comparing all populations with a common control

Suppose populations Π_i follows a normal distribution with mean θ_i and a common known variance σ^2, for $i = 1, ..., k$. Population Π_i is considered good if $\theta_i \geq \theta_0$, where θ_0 is a constant, otherwise, Π_i is considered bad.

A selection rule δ is a measurable mapping from sample space Y to $[0,1]^k$, D is the set consisting of all such mappings, that is, D is the set of all possible selection rules.

Let the loss function be

$$L(\theta, \delta(y), n) = \sum_{i=1}^{k} l(\theta_i, \delta_i(y)) + nc$$

where

$$l(\theta_i, \delta_i(y)) = \delta_i(y)(\theta_0 - \theta_i)I_{[0,\infty)}(\theta_0 - \theta_i) + (1 - \delta_i(y))(\theta_i - \theta_0)I_{[0,\infty)}(\theta_i - \theta_0).$$

Suppose we are to draw m_i observations from population Π_i, $i = 1, \ldots, k$, where $m_1 + \ldots + m_k = m$, then the Bayes risk for this allocation is

$$
\begin{aligned}
& E_x\{\min_{\delta\in D} E\{L(\theta,\delta(x,Y),n+m)|X=x,Y\}\} \\
= \;& E_x\{\min_{\delta\in D}[\sum_{i=1}^{k}(\delta_i(x,Y)\int_0^{\theta_0}(\theta_0-\theta_i)f(\theta_i|X=x,Y)d\theta_i \\
& +(1-\delta_i(x,Y))\int_{\theta_0}^{\infty}(\theta_i-\theta_0)f(\theta_i|X=x,Y)d\theta_i)]\}+nc+mc \\
= \;& E_x\{\min_{\delta\in D}[\sum_{i=1}^{k}(\delta_i(x,Y)[\int_0^{\theta_0}(\theta_0-\theta_i)f(\theta_i|X=x,Y)d\theta_i \\
& +\int_{\theta_0}^{\infty}(\theta_0-\theta_i)f(\theta_i|X=x,Y)d\theta_i]+\int_{\theta_0}^{\infty}(\theta_i-\theta_0)f(\theta_i|X=x,Y)d\theta_i)]\}+nc+mc \\
= \;& E_x\{\min_{\delta\in D}[\sum_{i=1}^{k}\delta_i(x,Y)[\theta_0-E\{\Theta_i|X=x,Y\}]+\sum_{i=1}^{k}\int_{\theta_0}^{\infty}(\theta_i-\theta_0)f(\theta_i|X=x,Y)d\theta_i]\} \\
& +nc+mc \\
= \;& E_x\{\min_{\delta\in D}\sum_{i=1}^{k}\delta_i(x,Y)[\theta_0-E\{\Theta_i|X=x,Y\}]\}+E_x\{\sum_{i=1}^{k}\int_{\theta_0}^{\infty}(\theta_i-\theta_0)f(\theta_i|X=x,Y)d\theta_i\} \\
& +nc+mc \\
= \;& E_x\{\sum_{i\in A(x,Y)}[\theta_0-E\{\Theta_i|X=x,Y\}]\}+E_x\{\sum_{i=1}^{k}\int_{\theta_0}^{\infty}(\theta_i-\theta_0)f(\theta_i|X=x,Y)d\theta_i\}+nc+mc \\
= \;& E_x\{\sum_{i\in A(x,Y)}[\theta_0-\frac{\alpha_i\mu_i(x)+q_iY_i}{\alpha_i+q_i}]\}+E_x\{\sum_{i=1}^{k}\int_{\theta_0}^{\infty}(\theta_i-\theta_0)f(\theta_i|X=x,Y)d\theta_i\}+nc+mc
\end{aligned}
$$

where $A(x,Y)=\{i|1\le i\le k \text{ and } \theta_0\le E\{\Theta_i|X=x,Y\}\}=\{i|1\le i\le k \text{ and } \theta_0\le \frac{\alpha_i\mu_i(x)+q_iY_i}{\alpha_i+q_i}\}$.

Because

$$\begin{aligned}
& E_x\{\sum_{i=1}^{k}\int_{\theta_0}^{\infty}(\theta_i-\theta_0)f(\theta_i|X=x,Y)d\theta_i\} \\
= \quad & E_x\{\sum_{i=1}^{k}E\{I_{[\theta_0,\infty)}(\Theta_i)(\Theta_i-\theta_0)|X=x,Y\}\} \\
= \quad & \sum_{i=1}^{k}E_x\{E\{I_{[\theta_0,\infty)}(\Theta_i)(\Theta_i-\theta_0)|X=x,Y\}\} \\
= \quad & \sum_{i=1}^{k}E_x\{I_{[\theta_0,\infty)}(\Theta_i)(\Theta_i-\theta_0)\}
\end{aligned}$$

does not depend on the allocation of the m observations at the second stage, the optimal allocation minimizes

$$E_x\{\sum_{i\in A(x,Y)}[\theta_0-\frac{\alpha_i\mu_i(x)+q_iY_i}{\alpha_i+q_i}]\}$$

subject to $m_1+\ldots+m_k=m$.

For $i=1,\ldots,k$, let

$$\begin{aligned}
l_i \quad &= \quad E_x\{\sum_{j\neq i,j\in A(x,Y_i)}[\theta_0-\mu_j(x)]+\sum_{\{i\}\bigcap A(x,Y_i)}[\theta_0-\frac{\alpha_i\mu_i(x)+q_iY_i}{\alpha_i+q_i}]\} \\
&= \quad \sum_{j\neq i,j\in A(x,Y_i)}[\theta_0-\mu_j(x)]+E_x\{\sum_{\{i\}\bigcap A(x,Y_i)}[\theta_0-\frac{\alpha_i\mu_i(x)+q_iY_i}{\alpha_i+q_i}]\}
\end{aligned}$$

where if $\theta_0\leq\frac{\alpha_i\mu_i(x)+q_iY_i}{\alpha_i+q_i}$, $A(x,Y_i)=\{j|j\neq i,\theta_0\leq\mu_j(x)\}\bigcup\{i\}$, otherwise, $A(x,Y_i)=\{j|j\neq i,\theta_0\leq\mu_j(x)\}$.

then the optimal allocation of the next observation draws one observation, with equal probabilities, from one of the populations with $l_{[1]}=\min\{l_1,...,l_k\}$.

4.3.2 Selection of all good normal populations compared with unknown control

Suppose there are k normal populations, where population Π_i has mean θ_i and variance σ^2, for $i = 1, \ldots, k$. There also exists a control normal population Π_0 with mean θ_0 and variance σ^2. σ^2 is a known constant and θ_i is a realization of Θ_i, which follows the normal distribution with mean μ_i and variance v_i^{-1}, for $i = 0, 1, ..., k$.

For $i = 1, \ldots, k$, if $\theta_i > \theta_0 + d_i$, population Π_i is considered better than Π_0, if $\theta_i < \theta_0$, we say population Π_i worse than Π_0, where $d_i \geq 0$ is known. Our objective is to select all populations better than the control and exclude all bad ones.

Let the loss function be

$$L(\theta, A, n) = \sum_{j \notin A} L_{1,j} I_{[\theta_0 + d_j, \infty)}(\theta_j) + \sum_{j \in A} L_{2,j} I_{(-\infty, \theta_0)}(\theta_j) + nc,$$

where A is a subset of $\{1, \ldots, \text{k}\}$ consisting of numbers corresponding to selected populations, $L_{1,j} > 0$ is the loss of not selecting population Π_j when it is better than Π_0, while $L_{2,j} > 0$ is the loss of selecting population Π_j when it is a bad population, for $j = 1, \ldots, k$, and c is the cost of sampling one observation.

Suppose at the second stage, m_i observations are to be drawn from population Π_i, $i = 0, 1, \cdots, k$, where $m_0 + \ldots + m_k = m$, then the Bayes risk is

$$
\begin{aligned}
& E_x\{\min_{A\in D} E\{\sum_{j\notin A} L_{1,j} I_{[\Theta_0+d_j,\infty)}(\Theta_j) + \sum_{j\in A} L_{2,j} I_{(-\infty,\Theta_0)}(\theta_j)|X=x,Y\}\} + nc + mc \\
= \; & E_x\{\min_{A\in D}\{\sum_{j\notin A} L_{1,j} P(\Theta_j \geq \Theta_0 + d_j|X=x,Y) + \sum_{j\in A} L_{2,j} P(\Theta_j < \Theta_0|X=x,Y)\}\} \\
& +nc+mc \\
= \; & E_x\{\min_{A\in D}\{\sum_{j\notin A} L_{1,j}\Phi(\frac{\frac{\alpha_j\mu_j(x)+q_jY_j}{\alpha_j+q_j} - \frac{\alpha_0\mu_0(x)+q_0Y_0}{\alpha_0+q_0} - d_j}{\sqrt{\frac{1}{\alpha_j+q_j}+\frac{1}{\alpha_0+q_0}}}) \\
& + \sum_{j\in A} L_{2,j}\Phi(\frac{\frac{\alpha_0\mu_0(x)+q_0Y_0}{\alpha_0+q_0} - \frac{\alpha_j\mu_j(x)+q_jY_j}{\alpha_j+q_j}}{\sqrt{\frac{1}{\alpha_j+q_j}+\frac{1}{\alpha_0+q_0}}})\}\} + nc + mc
\end{aligned}
$$

where D consists of all subsets of $\{1,\ldots,k\}$. $\Phi(x)$ is the cumulative distribution function of standard normal distribution, and the probability density function of Y_i given X=x has been given previously, for $i=0,1,\ldots,k$, and Y_i's are independent.

Therefore, the optimal allocation minimizes

$$
\begin{aligned}
& E_x\{\min_{A\in D}\{\sum_{j\notin A} L_{1,j}\Phi(\frac{\frac{\alpha_j\mu_j(x)+q_jY_j}{\alpha_j+q_j} - \frac{\alpha_0\mu_0(x)+q_0Y_0}{\alpha_0+q_0} - d_j}{\sqrt{\frac{1}{\alpha_j+q_j}+\frac{1}{\alpha_0+q_0}}}) \\
& + \sum_{j\in A} L_{2,j}\Phi(\frac{\frac{\alpha_0\mu_0(x)+q_0Y_0}{\alpha_0+q_0} - \frac{\alpha_j\mu_j(x)+q_jY_j}{\alpha_j+q_j}}{\sqrt{\frac{1}{\alpha_j+q_j}+\frac{1}{\alpha_0+q_0}}})\}\}
\end{aligned}
$$

subject to $m_1 + ... + m_k = m$.

4.3.3 Selection of all good normal populations in terms of variance

Suppose there are k normal populations, where population Π_i has a common known mean μ and variance ϕ_i, for $i=1,\ldots,k$. There also exists a control normal population Π_0 with known mean μ and variance ϕ_0. ϕ_i is a realization of the inverse gamma distribution Φ_i with shape parameter α_i and scale parameter β_i, for $i = 1, ..., k$. Φ_i's are independent.

For $i = 1, ..., k$, if $\phi_i < \phi_0$, population Π_i is considered better than Π_0, if $\phi_i > \phi_0$, we say population Π_i worse than Π_0. Our objective is to find all populations better than or equal to the control and exclude all bad ones.

Let the loss function be

$$L(\phi, A, n) = \sum_{j \notin A} L_{1,j} I_{(-\infty, \phi_0]}(\phi_j) + \sum_{j \in A} L_{2,j} I_{(\phi_0, \infty)}(\phi_j) + nc$$

where A is a subset of $\{1, \ldots, \mathrm{k}\}$ consisting of numbers corresponding to selected populations, $L_{1,j} > 0$ is the loss of not selecting population Π_j when it is better than or equal to the control, while $L_{2,j} > 0$ is the loss of selecting population Π_j when it is worse than Π_0, for $j = 1, \ldots, k$, and c is the cost of sampling one observation.

Suppose, at the second stage, m_i observations are to be drawn from population Π_i, $i = 1, \cdots, k$, where $m_1 + \ldots + m_k = m$, then the Bayes risk for this allocation is

$$
\begin{aligned}
& E_x\{\min_{A\in D} E\{\sum_{j\notin A} L_{1,j} I_{(-\infty,\phi_0]}(\Phi_j) + \sum_{j\in A} L_{2,j} I_{(\phi_0,\infty)}(\Phi_j)|X=x,Y\}\} + nc + mc \\
= \; & E_x\{\min_{A\in D}\{\sum_{j\notin A} L_{1,j} P(\Phi_j \le \phi_0|X=x,Y) + \sum_{j\in A} L_{2,j} P(\Phi_j > \phi_0|X=x,Y)\}\} + nc + mc \\
= \; & E_x\{\min_{A\in D}\{\sum_{j\notin A} L_{1,j} \frac{\Gamma(\alpha_j + \frac{n_j}{2} + \frac{m_j}{2}, \frac{\beta_j + \frac{s_j}{2} + \frac{W_j}{2}}{\phi_0})}{\Gamma(\alpha_j + \frac{n_j}{2} + \frac{m_j}{2})} \\
& + \sum_{j\in A} L_{2,j}(1 - \frac{\Gamma(\alpha_j + \frac{n_j}{2} + \frac{m_j}{2}, \frac{\beta_j + \frac{s_j}{2} + \frac{W_j}{2}}{\phi_0})}{\Gamma(\alpha_j + \frac{n_j}{2} + \frac{m_j}{2})})\}\} + nc + mc
\end{aligned}
$$

where D consists of all subsets of $\{1,\ldots,k\}$, $\Gamma(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2}, \frac{\beta_i + \frac{s_i}{2} + \frac{W_i}{2}}{\phi_0})$ is the upper incomplete gamma function, and the probability density function of W_i given X=x has been given previously, for $i = 1, \cdots, k$, and W_i's are independent.

Therefore, the optimal allocation minimizes

$$
E_x\{\min_{A\in D}\{\sum_{i\notin A} L_{1,i} \frac{\Gamma(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2}, \frac{\beta_i + \frac{s_i}{2} + \frac{W_i}{2}}{\phi_0})}{\Gamma(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2})} + \sum_{i\in A} L_{2,i}(1 - \frac{\Gamma(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2}, \frac{\beta_i + \frac{s_i}{2} + \frac{W_i}{2}}{\phi_0})}{\Gamma(\alpha_i + \frac{n_i}{2} + \frac{m_i}{2})})\}\}
$$

subject to $m_1 + ... + m_k = m$.

4.4 Selection of all good normal populations close to a control

Suppose population Π_i can be characterized by normal distribution with mean θ_i and a common variance σ^2. There is also a control population Π_0 characterized by N(θ_0, σ^2), where θ_0 is a constant. For $i = 1, \ldots, k$, the distance between population Π_i and Π_0 is measured by

$\delta_i = \frac{(\theta_i-\theta_0)^2}{2\sigma^2}$; For a given constant $c > 0$, population Π_i is said close to the control population if $\delta_i \leq c$, and bad otherwise. we want to select all populations close to the control and excluding all bad populations.

A decision rule $d = (d_1, ..., d_k)$ is a mapping defined on the sample space Y into $[0, 1]^k$. Let the loss function be

$$L(\theta, d, n) = \sum_{i=1}^{k} L_i(\theta, d_i) + nc,$$

where $L_i(\theta, d_i) = d_i(\delta_i - c)I_{(c,\infty)}(\delta_i) + (1 - d_i)(c - \delta_i)I_{[0,c]}(\delta_i)$ and c is the cost of sampling one observation.

Suppose at the second stage, m_i observations are to be drawn from population Π_i, $i = 1, \ldots, k$, then the Bayes risk is

$$
\begin{aligned}
& E_x\{\min_{d\in D} E\{\sum_{i=1}^{k} L_i(\Theta, d_i)|X=x,Y\}\} + nc + mc \\
= \; & E_x\{\min_{d\in D}\sum_{i=1}^{k} E\{d_i(x,Y)(\frac{(\Theta_i-\theta_0)^2}{2\sigma^2} - c)I_{(c,\infty)}(\frac{(\Theta_i-\theta_0)^2}{2\sigma^2}) \\
& +(1-d_i(x,Y))(c-\frac{(\Theta_i-\theta_0)^2}{2\sigma^2})I_{[0,c]}(\frac{(\Theta_i-\theta_0)^2}{2\sigma^2})|X=x,Y\}\} + nc + mc \\
= \; & E_x\{\min_{d\in D}\sum_{i=1}^{k}\{d_i(x,Y)\int_{A_i}(\frac{(\theta_i-\theta_0)^2}{2\sigma^2} - c)f(\theta_i|X=x,Y)d\theta_i \\
& +(1-d_i(x,Y))\int_{\bar{A}_i}(c-\frac{(\theta_i-\theta_0)^2}{2\sigma^2})f(\theta_i|X=x,Y)d\theta_i\}\} + nc + mc \\
= \; & E_x\{\min_{d\in D}\sum_{i=1}^{k}\{d_i(x,Y)\int_{A_i}(\frac{(\theta_i-\theta_0)^2}{2\sigma^2} - c)f(\theta_i|X=x,Y)d\theta_i \\
& +d_i(x,Y)\int_{\bar{A}_i}(\frac{(\theta_i-\theta_0)^2}{2\sigma^2} - c)f(\theta_i|X=x,Y)d\theta_i \\
& +\int_{\bar{A}_i}(c-\frac{(\theta_i-\theta_0)^2}{2\sigma^2})f(\theta_i|X=x,Y)d\theta_i\}\} + nc + mc \\
= \; & E_x\{\min_{d\in D}\sum_{i=1}^{k}\{d_i(x,Y)E\{\frac{(\Theta_i-\theta_0)^2}{2\sigma^2} - c|X=x,Y\} \\
& +\int_{\bar{A}_i}(c-\frac{(\theta_i-\theta_0)^2}{2\sigma^2})f(\theta_i|X=x,Y)d\theta_i\}\} + nc + mc \\
= \; & E_x\{\min_{d\in D}\sum_{i=1}^{k} d_i(x,Y)E\{\frac{(\Theta_i-\theta_0)^2}{2\sigma^2} - c|X=x,Y\}\} \\
& +\sum_{i=1}^{k} E_x\{\int_{\bar{A}_i}(c-\frac{(\theta_i-\theta_0)^2}{2\sigma^2})f(\theta_i|X=x,Y)d\theta_i\} + nc + mc \\
= \; & E_x\{\min_{d\in D}\sum_{i=1}^{k} d_i(x,Y)E\{\frac{(\Theta_i-\theta_0)^2}{2\sigma^2} - c|X=x,Y\}\} \\
& +\sum_{i=1}^{k} E_x\{E\{(c-\frac{(\Theta_i-\theta_0)^2}{2\sigma^2})I_{[0,c]}(\frac{(\Theta_i-\theta_0)^2}{2\sigma^2})|X=x,Y\}\} + nc + mc \\
= \; & E_x\{\min_{d\in D}\sum_{i=1}^{k} d_i(x,Y)E\{\frac{(\Theta_i-\theta_0)^2}{2\sigma^2} - c|X=x,Y\}\} \\
& +\sum_{i=1}^{k} E_x\{(c-\frac{(\Theta_i-\theta_0)^2}{2\sigma^2})I_{[0,c]}(\frac{(\Theta_i-\theta_0)^2}{2\sigma^2})\} + nc + mc \\
= \; & E_x\{\sum_{i\in B(x,Y)} E\{\frac{(\Theta_i-\theta_0)^2}{2\sigma^2} - c|X=x,Y\}\} \\
& +\sum_{i=1}^{k} E_x\{(c-\frac{(\Theta_i-\theta_0)^2}{2\sigma^2})I_{[0,c]}(\frac{(\Theta_i-\theta_0)^2}{2\sigma^2})\} + nc + mc
\end{aligned}
$$

where $A_i = \{\theta_i | \frac{(\theta_i - \theta_0)^2}{2\sigma^2} > c\}$, $\bar{A}_i = \{\theta_i | \frac{(\theta_i - \theta_0)^2}{2\sigma^2} \leq c\}$ and $B(x, Y) = \{i | E\{\frac{(\Theta_i - \theta_0)^2}{2\sigma^2} | X = x, Y\} \leq c\}$.

Therefore, the optimal allocation minimizes

$$E_x\{\sum_{i \in B(x,Y)} E\{\frac{(\Theta_i - \theta_0)^2}{2\sigma^2} - c | X = x, Y\}\}$$

subject to $m_i + ... + m_k = m$.

For $i = 1, \ldots, k$, let

$$l_i = E_x\{\sum_{j \neq i, j \in B(x, Y_i)} E\{\frac{(\Theta_j - \theta_0)^2}{2\sigma^2} - c | X = x\} + \sum_{\{i\} \bigcap B(x, Y_i)} E\{\frac{(\Theta_i - \theta_0)^2}{2\sigma^2} - c | X = x, Y_i\}\}$$

where if $E\{\frac{(\Theta_i - \theta_0)^2}{2\sigma^2} | X = x, Y_i\} \leq c$, $B(x, Y_i) = \{j | j \neq i, E\{\frac{(\Theta_j - \theta_0)^2}{2\sigma^2} | X = x\} \leq c\} \bigcup \{i\}$, otherwise, $B(x, Y_i) = \{j | j \neq i, E\{\frac{(\Theta_j - \theta_0)^2}{2\sigma^2} | X = x\} \leq c\}$, then the optimal allocation of the next observation draws one observation, with equal probabilities, from one of the populations with $l_{[1]}$, where $l_{[1]} = \min\{l_1, ..., l_k\}$.

APPENDICES

Appendix A

DERIVATION OF CONDITIONAL DISTRIBUTIONS

A.1 Derivation of conditional distribution of Y

In the following, we will derive the conditional distribution of Y given X=x when both mean and variance of the normal distribution are random, where $x = (x_1, ..., x_n)^T$, and $Y = (Y_1, ..., Y_m)^T$, are vectors of observations at the first and the second step, respectively.

$$\begin{aligned}\prod_{i=1}^{m} \varphi_{\theta,\phi}(y_i)\pi(\theta,\phi|x) &= \prod_{i=1}^{m} \varphi_{\theta,\phi}(y_i)\pi_1(\theta|\phi,x)\pi_2(\phi|x) \\ &= m(y|x)\pi(\theta,\phi|x,y) \\ &= m(y|x)\pi_1(\theta|\phi,x,y)\pi_2(\phi|x,y)\end{aligned}$$

where $\pi_1(\theta|\phi, x, y)$ is the pdf of normal distribution with mean $\mu(x, y) = \frac{\mu+(m+n)\tau\tilde{x}}{(n+m)\tau+1}$, where $\tilde{x} = \frac{1}{m+n}(\sum_{i=1}^{n} x_i + \sum_{i=1}^{m} y_i)$, and variance $(\tau^{-1} + n + m)^{-1}\phi$ and $\pi_2(\phi|x, y)$ is an inverted gamma density with parameters $\alpha + \frac{n+m}{2}$ and β^*,where

$\beta^* = \{\beta^{-1} + \frac{1}{2}\sum_{i=1}^{n}(x_i - \tilde{x})^2 + \frac{1}{2}\sum_{i=1}^{m}(y_i - \tilde{x})^2 + \frac{(m+n)(\tilde{x}-\mu)^2}{2(1+n\tau+m\tau)}\}^{-1}$.

Appendix A (Continued)

Therefore,

$$
\begin{aligned}
m(y|x) &= \frac{f(y,\theta,\phi|x)}{\pi(\theta,\phi|x,y)} \\
&= \frac{\prod_{i=1}^{m}\varphi_{\theta,\phi}(y_i)\pi(\theta,\phi|x)}{\pi(\theta,\phi|x,y)} \\
&= \frac{(2\pi\phi)^{-\frac{m}{2}}e^{-\frac{1}{2\phi}\sum_{i=1}^{m}(y_i-\theta)^2}(2\pi\frac{\phi}{\tau^{-1}+n})^{-\frac{1}{2}}e^{-\frac{1}{2}\frac{\tau^{-1}+n}{\phi}(\theta-\mu(x))^2}}{(2\pi\frac{\phi}{\tau^{-1}+n+m})^{-\frac{1}{2}}e^{-\frac{1}{2}\frac{\tau^{-1}+n+m}{\phi}(\theta-\mu(x,y))^2}} \\
&\quad \cdot \frac{[\Gamma(\alpha+\frac{n}{2})\beta'^{\alpha+\frac{n}{2}}\phi^{\alpha+\frac{n}{2}+1}]^{-1}e^{-\frac{1}{\phi\beta'}}}{[\Gamma(\alpha+\frac{n+m}{2})\beta^{*\alpha+\frac{n+m}{2}}\phi^{\alpha+\frac{n+m}{2}+1}]^{-1}e^{-\frac{1}{\phi\beta^*}}} \\
&= \frac{(2\pi)^{-\frac{m}{2}}(\frac{\tau}{1+n\tau})^{-\frac{1}{2}}[\Gamma(\alpha+\frac{n}{2})\beta'^{\alpha+\frac{n}{2}}]^{-1}}{(\frac{\tau}{1+n\tau+m\tau})^{-\frac{1}{2}}[\Gamma(\alpha+\frac{n+m}{2})\beta^{*\alpha+\frac{n+m}{2}}]^{-1}} \\
&\quad \cdot e^{-\frac{1}{2\phi}[\sum_{i=1}^{m}(y_i-\theta)^2+\frac{1+n\tau}{\tau}(\theta-\mu(x))^2+\frac{2}{\beta'}-\frac{1+n\tau+m\tau}{\tau}(\theta-\mu(x,y))^2-\frac{2}{\beta^*}]}
\end{aligned}
$$

In the following, we will prove that

$$\sum_{i=1}^{m}(y_i-\theta)^2+\frac{1+n\tau}{\tau}(\theta-\mu(x))^2+\frac{2}{\beta'}-\frac{1+n\tau+m\tau}{\tau}(\theta-\mu(x,y))^2-\frac{2}{\beta^*}=0$$

The coefficient of θ^2

$$m+\frac{1+n\tau}{\tau}-\frac{1+n\tau+m\tau}{\tau}=0$$

The coefficient of θ

$$-2\sum_{i=1}^{m}y_i-2\frac{1+n\tau}{\tau}\mu(x)+2\frac{1+n\tau+m\tau}{\tau}\mu(x,y)=0$$

Appendix A (Continued)

After standard calculation, we have that the rest

$$\sum_{i=1}^{m} {y_i}^2 + \frac{1+n\tau}{\tau}\mu^2(x) + \frac{2}{\beta'} - \frac{1+n\tau+m\tau}{\tau}\mu^2(x,y) - \frac{2}{\beta^*}$$

is also equal to 0.

Therefore,

$$m(y|x) = (2\pi)^{-\frac{m}{2}}(1+n\tau)^{\frac{1}{2}}(1+n\tau+m\tau)^{-\frac{1}{2}}\Gamma(\alpha+\frac{n+m}{2})\Gamma(\alpha+\frac{n}{2})^{-1}\beta^{*\alpha+\frac{n+m}{2}}\beta'^{-\alpha-\frac{n}{2}}.$$

A.2 Derivation of pdf of the sum of Gamma distributions given Gamma prior

Suppose $X \sim \text{Gamma}(a,\theta)$ given θ, which is a realization of a random variable Θ, where $\Theta \sim \text{Gamma}(\alpha,\beta)$. n observations $x_1,...,x_n$ have been drawn from X. Let $y = \sum_{i=1}^{n} x_i$, then

$\Theta|y \sim \text{Gamma}(\alpha+na, \beta+y)$

Because $X_i|\theta \sim \text{Gamma}(a,\theta)$, for $i = 1,...,k$, we have $Y = X_1+...+X_n|\theta \sim \text{Gamma}(na,\theta)$.

In the following, we will derive the marginal distribution of Y.

$f(y|\theta)\pi(\theta) = m(y)\pi(\theta|y)$,

where

$$f(y|\theta) = \frac{\theta^{na}}{\Gamma(na)}y^{na-1}e^{-\theta y}, y > 0,\, \theta > 0,$$

$$\pi(\theta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta},\, \theta > 0,$$

Appendix A (Continued)

$$\pi(\theta|y) = \frac{(\beta+y)^{\alpha+na}}{\Gamma(\alpha+na)}\theta^{\alpha+na-1}e^{-(\beta+y)\theta},\ \theta > 0,\ y > 0,$$

therefore, the marginal probability density function of Y

$$\begin{aligned} m(y) &= \frac{f(y|\theta)\pi(\theta)}{\pi(\theta|y)} \\ &= \frac{\frac{\theta^{na}}{\Gamma(na)}y^{na-1}e^{-\theta y}\frac{\beta^{\alpha}}{\Gamma(\alpha)}\theta^{\alpha-1}e^{-\beta\theta}}{\frac{(\beta+y)^{\alpha+na}}{\Gamma(\alpha+na)}\theta^{\alpha+na-1}e^{-(\beta+y)\theta}} \\ &= \frac{\Gamma(\alpha+na)\beta^{\alpha}y^{na-1}}{\Gamma(\alpha)\Gamma(na)(\beta+y)^{\alpha+na}},\ \ y > 0. \end{aligned}$$

A.3 Derivation of pmf of the sum of Poisson distributions given Gamma prior

Suppose $X \sim$ Poisson(λ) given λ, which is a realization of a random variable Λ, where $\Lambda \sim$ Gamma(k, θ), with k being the shape parameter and θ being the scale parameter. n observations $x_1, ..., x_n$ have been drawn from X. Let $y = \sum_{i=1}^{n} x_i$, then $\Lambda|y \sim$ Gamma($k + y, \frac{\theta}{n\theta+1}$)

Because $X_i|\lambda \sim$ Poisson(λ), $i = 1, ..., k$, we have $Y = X_1 + ... + X_n|\lambda \sim$ Poisson($n\lambda$).

In the following, we will derive the marginal distribution of Y.

$f(y|\lambda)\pi(\lambda) = m(y)\pi(\lambda|y)$, where

$$f(y|\lambda) = \frac{(n\lambda)^{y}e^{-n\lambda}}{y!},\ y = 0, 1, ...,\ \ \lambda > 0,$$

$$\pi(\lambda) = \frac{1}{\theta^{k}\Gamma(k)}\lambda^{k-1}e^{-\frac{\lambda}{\theta}},\ \ \lambda > 0,$$

$$\pi(\lambda|y) = \frac{1}{(\frac{\theta}{n\theta+1})^{k+y}\Gamma(k+y)}\lambda^{k+y-1}e^{-\frac{\lambda(n\theta+1)}{\theta}},\ \ \lambda > 0,$$

Appendix A (Continued)

therefore, the marginal probability mass function of Y

$$
\begin{aligned}
m(y) &= \frac{f(y|\lambda)\pi(\lambda)}{\pi(\lambda|y)} \\
&= \frac{\frac{(n\lambda)^y e^{-n\lambda}}{y!}\frac{1}{\theta^k\Gamma(k)}\lambda^{k-1}e^{-\frac{\lambda}{\theta}}}{\frac{1}{(\frac{\theta}{n\theta+1})^{k+y}\Gamma(k+y)}\lambda^{k+y-1}e^{-\frac{\lambda(n\theta+1)}{\theta}}} \\
&= \frac{\Gamma(k+y)n^y\theta^y}{\Gamma(k)y!(n\theta+1)^{k+y}}, \quad y = 0, 1, ...
\end{aligned}
$$

CITED LITERATURE

Bahadur, R.R. (1950). On a problem in the theory of k populations. *Ann. Math. Statist.* **21**, 362–375.

Bahadur, R.R. and Goodman, L. (1952). Impartial decision rules and sufficient statistics. *Ann. Math. Statist.* **23**, 553–562.

Bahadur, R.R. and Robbins, H. (1950). The problem of the greater mean. *Ann. Math. Statist.* **21**, 469–487. Correction (1951) **22**, 310.

Balakrishnan, N. and Miescke, K.-J., eds. (2006). *In Memory of Dr. Shanti Swarup Gupta.* Special Issue. *J. Statist. Plann. Inference* **136**.

Bansal, N.K. and Miescke, K.-J. (2002). Simultaneous selection and estimation in general linear models. *J. Statist. Plann. Inference* **104**, 377–390.

Bansal, N.K. and Miescke, K.-J. (2005). Simultaneous selection and estimation of the best treatment with respect to a control in general linear models. *J. Statist. Plann. Inference* **129**, 387–404.

Bechhofer, R.E. (1954). A single-sample multiple decision procedure for ranking means of normal populations with known variances. *Ann. Math. Statist.* **25**, 16–39.

Bechhofer, R.E., Dunnett, C.W., and Sobel, M. (1954). A two-sample multiple decision procedure for ranking means of normal populations with a common unknown variance. *Biometrika* **41**, 170–176.

Bechhofer, R.E., Kiefer, J., and Sobel, M. (1968). *Sequential Identification and Ranking Procedures.* Univ. of Chicago Press, Chicago.

Bechhofer, R.E., Santner, T.J., and Goldsman, D.M. (1995). *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons.* Wiley, New York.

Bechhofer, R.E., J. Kiefer and M. Sobel (1968). *Sequential Identification and Ranking Procedures.* The University of Chicago Press, Chicago.

Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*, 2nd ed. Springer, New York.

Büringer, H., Martin, H., and Schriever, K.H. (1980). *Nonparametric Sequential Selection Procedures.* Birkhäuser, Boston.

Cohen, A. and Sackrowitz, H.B. (1988). A decision theory formulation for population selection followed by estimating the mean of the selected population. *Statistical Decision Theory and Related Topics IV.* J.O. Berger and S.S. Gupta, eds., Springer, New York, Vol. **2**, 33–36.

Dudewicz, E.J. (1976). *Introduction to Statistics and Probability.* Holt, Rinehart, and Winston, New York.

Dudewicz, E.J., ed. (1982). *The Frontiers of Modern Statistical Inference Procedures.* Proceedings and Discussions of The IPASRAS Conference 1982. American Sciences Press, Columbus, OH.

Dudewicz, E.J. and Koo, J.O. (1982). *The Complete Categorized Guide to Statistical Selection and Ranking Procedures.* American Sciences Press, Columbus, OH.

Dunnett, C.W. (1955). A multiple comparison procedure for comparing several treatments with a control. *J. Amer. Statist. Assoc.* **50**, 1096–1121.

Dunnett, C.W. (1960). On selecting the largest of k normal population means (with Discussion). *J. Roy. Statist. Soc.* **B-22**, 1–40.

Eaton, M.L (1967a). Some optimum properties of ranking procedures. *Ann. Math. Statist.* **38**, 124–137.

Eaton, M.L. (1967b). The generalized variance: testing and ranking problem. *Ann. Math. Statist.* **38**, 941–943.

Gibbons, J.D., Olkin, I., and Sobel, M. (1977). *Selecting and Ordering Populations: A New Statistical Methodology.* Wiley, New York. Second (corrected) printing 1999 by SIAM: *Classics in Appl. Math.* **26**.

Gupta, S.S. (1956). On a decision rule for a problem in ranking means. Ph.D. Thesis, *Mimeo. Ser. 150,* Institute of Statistics, Univ. of North Carolina, Chapel Hill.

Gupta, S.S. (1965). On some multiple decision (selection and ranking) rules. *Technometrics* **7**, 225–245.

Gupta, S.S., ed. (1977). Special Issue on Selection and Ranking Problems. *Commun. Statist. - Theor. Meth.* **A-6**, No. 1.

Gupta, S.S. and Berger, J.O., eds. (1982). *Statistical Decision Theory and Related Topics III*, Vol. **1** and **2**. Proceedings of the Third Purdue Symposium 1981. Academic Press, New York.

Gupta, S.S. and Berger, J.O., eds. (1988). *Statistical Decision Theory and Related Topics IV*, Vol. **1** and **2**. Proceedings of the Fourth Purdue Symposium 1986. Springer, New York.

Gupta, S.S. and Huang, D.-Y. (1981). *Multiple Statistical Decision Theory: Recent Developments. Lecture Notes in Statistics* **6**, Springer, New York.

Gupta, S.S. and Moore, D.S., eds. (1977). *Statistical Decision Theory and Related Topics II.* Proceedings of the Second Purdue Symposium 1976. Academic Press, New York.

Gupta, S.S. and K.J. Miescke (1993) On combining selection and estimation in the search for the largest binomial parameter. *J. Statist. Plann. Inference* **36**, 129-140.

Gupta, S.S. and K.J. Miescke (1994) Bayesian look ahead one stage sampling allocations for selecting the largest normal mean. *Statist. Papers* **35**, 169-177.

Gupta, S.S. and Panchapakesan, S. (1979). *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations.* Wiley, New York.

Gupta, S.S. and Yackel, J., eds. (1971). *Statistical Decision Theory and Related Topics.* Proceedings of the First Purdue Symposium 1971. Academic Press, New York.

Gupta, S.S. and K.J. Miescke (1990) On finding the largest normal mean and estimating the selected mean. *Sankhyā* B-52, 144-157.

Gupta, S.S. and K.J. Miescke (1996) Bayesian look ahead one-stage sampling allocations for selection of the best population. *J. Statist. Plann. Inference* **54**, 229-244.

Gupta, S.S. and T. Liang (2002) Selecting the most reliable Poisson population provided it is better than a control: a nonparametric empirical Bayes approach. *J. Statist. Plann. Inference* **103**, 191-203.

Gupta, S.S. and T. Liang (1999) Selecting good exponential populations compared with a control: a nonparametric empirical Bayes approach. *Sankhyā* B-61, 289-304.

Gupta, S.S. and K.J. Miescke (1996) Bayesian look-ahead sampling allocations for selecting the best Bernoulli population. *Madan Puri Festschrift* 353-369.

Gupta, S.S. (1962) On a selection and ranking procedure for gamma populations. *Ann. Instit. Statist. Math.* (Annals of the Institute of Statistical Mathematics) **14**, 199-212.

Gupta, S.S. and T. Liang (1999) On empirical Bayes simultaneous selection procedures for comparing normal populations with a standard. *J. Statist. Plann. Inference* **77**, 73-88.

Gupta, S.S. and K.J. Miescke (1984). Sequential selection procedures-a decision theoretic approach. *Ann. Statist.* **12**, 336-350.

Gupta, S.S. and S. Panchapakesan (1979). *Multiple Decision Procedures: Theory adn Methodology of Selecting and Ranking Populations*, Wiley, New York.

Hall, W.J. (1959). The most economical character of Bechhofer and Sobel decision rules. *Ann. Math. Statist.* **30**, 964–969.

Hoppe, F.M., ed. (1993). *Multiple Comparisons, Selection, and Applications in Biometry.* Festschrift in Honor of Charles W. Dunnett. Dekker, New York.

Horn, M. and Volland, R. (1995). *Multiple Tests und Auswahlverfahren.* Biometrie, Gustav Fischer, Stuttgart.

Lehmann, L.E. (1957a). A theory of some multiple decision problems, I. *Ann. Math. Statist.* **28**, 1–25.

Lehmann, L.E. (1957b). A theory of some multiple decision problems, II. *Ann. Math. Statist.* **28**, 547–572.

Lehmann, E.L. (1961). Some model I problems of selection. *Ann. Math. Statist.* **32**, 990–1012.

Lehmann, E.L. (1963). A class of selection procedures based on ranks. *Math. Ann.* **150**, 268–275.

Lehmann, E.L. (1966). On a theorem of Bahadur and Goodman. *Ann. Math. Statist.* **37**, 1–6.

Liang, T. (1997) Simultaneously selecting normal populations close to a control. *J. Statist. Plann. Inference* **61**, 297-316.

Liese, F. and K.J. Miescke (2008). *Statistical Decision Theory*, Springer, New York.

Miescke, K.-J. and Rasch, D., eds. (1996a). 40 Years of Statistical Selection Theory, Part I. Special Issue. *J. Statist. Plann. Inference* **54**, No. 2.

Miescke, K.-J. and Rasch, D., eds. (1996b). 40 Years of Statistical Selection Theory, Part II. Special Issue. *J. Statist. Plann. Inference* **54**, No. 3.

Misra, N., van der Meulen, E.C., and Branden, K.V. (2006). On some inadmissibility results for the scale parameters of selected gamma populations. *J. Statist. Plann. Inference* **136**, 2340–2351.

Mukhopadhyay, N. and Solanky, T.K.S. (1994). *Multistage Selection and Ranking Procedures: Second-Order Asymptotics.* Dekker, New York.

Panchapakesan, S. and Balakrishnan, N., eds. (1997). *Advances in Statistical Decision Theory and Applications.* In Honor of Shanti S. Gupta. Birkhäuser, Boston.

Paulson, E. (1949). A multiple decision procedure for certain problems in the analysis of variance. *Ann. Math. Statist.* **20**, 95–98.

Paulson, E. (1952). On the comparison of several experimental categories with a control. *Ann. Math. Statist.* **23**, 239–246.

Rasch, D. (1995). *Mathematische Statistik.* J.A. Barth, Heidelberg.

Santner, T.J. and Tamhane, A.C., eds. (1984). *Design of Experiments: Ranking and Selection.* Essays in honor of Robert E. Bechhofer. Dekker, New York.

VITA

Education

- **Ph.D.** in Statistics, University of Illinois at Chicago, Chicago, USA, 2011
- **M.S.** in Mathematics, Beijing Institute of Technology, Beijing, China, 2005
- **B.S.** in Mathematics, Hebei Normal University, Shijiazhuang, China, 2002

Certificates

- SAS Certified Advanced Programmer for SAS9, SAS Institute, 2011
- SAS Certified Base Programmer for SAS9, SAS Institute, 2011

Teaching Experience

- Math 090: Intermediate Algebra
- Math 121: Precalculus
- Math 180: Single Variable Calculus
- Stat 101: Introduction to Statistics

Publications

- S-connectivity and countable paracompactness of L-topological spaces

  Master Thesis, 2005

- S-connectivity of relative productive space of L-topological spaces

  Fuzzy Systems and Mathematics, Vol.19, 2005

Professional Membership

- American Statistical Association(AMS)
- Institute of Mathematical Statistics(IMS)