ON A SELECTION PROCEDURE FOR SELECTING THE BEST LOGISTIC POPULATION COMPARED WITH A CONTROL*

by

Shanti S. Gupta
Purdue University

Zhengyan Lin**
Hangzhou University

and

Xun Lin
Purdue University

Technical Report #99-04

Department of Statistics
Purdue University
West Lafayette, IN USA

April 1999

*

99-64 Revised

ON A SELECTION PROCEDURE FOR SELECTING THE BEST LOGISTIC POPULATION COMPARED WITH A CONTROL*

Shanti S. Gupta
Department of Statistics
Purdue University
West Lafayette, IN 47907

and

Zhengyan Lin
Department of Mathematics
Hangzhou University
Hangzhou, China 310028

Xun Lin
Purdue University
Department of Statistics
West Lafayette, IN 47907

Abstract

In this paper we investigate the problem of selecting the best logistic population from $k(\geq 2)$ possible candidates. The selected population must also be better than a given control. We employ the empirical approach and develop a selection procedure. The performance (rate of convergence) of the proposed selection rule is also analyzed. We also carry out a simulation study to investigate the rate of convergence of the proposed empirical selection procedure. The results of the simulation study are provided in the paper.

AMS Classification: primary 62F07; secondary 62C12.

Keywords: Asymptotically optimal; empirical approach; selection procedure; logistic population; rate of convergence.

*This research was supported in part by US Army Research Office, Grant DAAH04-95-1-0165 at Purdue University.

ON A SELECTION PROCEDURE FOR SELECTING THE BEST LOGISTIC POPULATION COMPARED WITH A CONTROL*

by

Shanti S. Gupta
Purdue University

Zhengyan Lin**
Hangzhou University

and

Xun Lin
Purdue University

Abstract

In this paper we investigate the problem of selecting the best logistic population from $k(\geq 2)$ possible candidates. The selected population must also be better than a given control. We employ the empirical approach and develop a selection procedure. The performance (rate of convergence) of the proposed selection rule is also analyzed. We also carry out a simulation study to investigate the rate of convergence of the proposed empirical selection procedure. The results of the simulation study are provided in the paper.

AMS Classification: primary 62F07, secondary 62C12.

Keywords: Asymptotically optimal; empirical approach; selection procedure; logistic population; rate of convergence.

---

* This research was supported by the US Army Research Office, Grant DAAH04-95-1-0165 at Purdue University.

** Visiting Professor from Department of Mathematics, Hangzhou University, China, 310028.

1 Introduction

Logistic distributions have been widely used in studies that are related with growth processes. Berkson (1957) used the logistic distribution as a model to analyze quantal response. Plackett (1958) considered the use of the logistic distribution with life test data. The importance of the logistic distribution has resulted in numerous investigations involving the statistical aspects of the distribution. For example, Talacko (1956) showed that it could be a limiting distribution in variaous situations. Birnbaum and Dudman (1963), and Gupta and Shah (1965) studied its order statistics and their limiting properties. Gupta and Gnanadesikan (1966), and Gupta, Qureishi and Shah (1967) have considered the estimation of parameters of the logistic distribution, Gupta, Qureishi and Shah have constructed the best linear unbiased estimators of both location and scale parameters using order statistics.

It is now well recognized that the classical techniques for testing homogeneity hypotheses are inadequate to serve, in many practical situations, the experimenter's real purpose, which is to rank several competing populations or to select the best among them. Such realistic goals and formulations set the stage for the development of the ranking and selection theory. An important part of this development is the study of ranking and selection problems for specific parametric families of distributions including, of course, logistic distributions. Gupta and Han (1991) proposed an elimination type procedure based on the estimated sample means for selecting the best logistic population. In addition, Gupta and Han (1992) proposed another selection rule for selecting the best logistic population using the indifference zone approach. A very nice paper on ranking and selection procedures for the logistic populations is Panchapakesan (1992) which is published in "the Handbook of the Logistic Distribution", edited by Balakrishnan (1992). In this book one can find a good deal of recent developments related to the logistic distribution.

In this paper, we investigate the problem of selecting the best logistic population by using the observed sample medians. Assume that there are k independent logistic populations whose location parameters follow a prior normal distribution and the parameters of the prior normal distribution are unknown. Motivated by the empirical methodology, we propose an empirical selection procedure that is based on the past observed data.

2 Formulation of the Selection Problem and the Selection Rule

Let $\Pi_1, \ldots, \Pi_k$ be k independent logistic populations with unknown means $\theta_1, \ldots, \theta_k$. Let $\theta_{[1]} \leq \ldots \leq \theta_{[k]}$ denote the ordered values of the parameters $\theta_1, \ldots, \theta_k$. It is assumed that the exact pairing between the ordered and the unordered parameters is unknown. A population Π_i with $\theta_i = \theta_{[k]}$ is considered as the best population. For a given fixed control θ_0, population Π_i is defined to be good if the corresponding $\theta_i > \theta_0$, and bad otherwise. Our goal is to select the one which is the best among the k logistic populations and also good compared with the given standard θ_0. If there is no such population, we select none.

Let $\Omega = \{\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)\}$ be the parameter space and $\underset{\sim}{a} = (a_0, \ldots, a_k)$ be an action, where $a_i = 0$, or 1, for $i = 0, 1, \ldots, k$, and $\sum_{i=0}^{k} a_i = 1$. For each $i = 1, \ldots, k$, $a_i = 1$ means population Π_i is selected as the best among the k candidates and also good compared with θ_0, while $a_i = 0$ means population Π_i is not selected either because it is not the best among the k candidates or because it is bad compared with the control. $a_0 = 1$ means that all the k populations are excluded as bad and none of these k logistic populations is selected. The following loss function will be considered:

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

For each $i = 1, \ldots, k$, let $X_{i1}, \ldots, X_{iM}$ be a sample of size M from the i-th logistic population $\Pi_i = L(\theta_i, \sigma_i^2)$ which has the following conditional density distribution given θ_i and σ_i^2

$$\frac{1}{\sigma_i} \frac{e^{-(x_i - \theta_i)/\sigma_i}}{(1 + e^{-(x_i - \theta_i)/\sigma_i})^2}, \qquad -\infty < x_i < \infty. \tag{1}$$

For convenience, suppose (for now) M is an odd number, and we denote $M = 2s + 1$. Since logistic distribution is symmetric about its mean, the population mean and median are identical. We assume that for each $i = 1, \ldots, k$, the population median (and also the mean) θ_i is a realization of random variable Θ_i which follows a normal $N(\mu_i, \tau_i^2)$ prior distribution with parameters (μ_i, τ_i^2). The random variables $\Theta_1, \ldots, \Theta_k$ are mutually independent. σ_i^2, μ_i, τ_i^2 are unknown but fixed. In other words, σ_i^2, μ_i, τ_i^2 are fixed nuisance parameters. Let X_i be the median of $\{X_{i1}, \ldots, X_{iM}\}$, $i = 1, \ldots, k$, then the conditional distribution of X_i given (θ_i, σ_i^2) can be explicitly written out as follows:

$$f_i(x_i|\theta_i, \sigma_i^2) = \frac{(2s+1)!}{(s!)^2} \frac{1}{\sigma_i} \frac{(e^{-(x_i - \theta_i)/\sigma_i})^{s+1}}{(1 + e^{-(x_i - \theta_i)/\sigma_i})^{2s+2}}, \qquad -\infty < x_i < \infty. \tag{2}$$

From (2) we see that the density function $f_i(x_i|\theta_i, \sigma_i^2)$ is symmetric about $\Theta_i = \theta_i$, therefore,

$$EX_i = E(E(X_i|\Theta_i = \theta_i)) = E\Theta_i = \mu_i. \tag{3}$$

The posterior distribution density of Θ_i given $X_i = x_i$ is proportional to

$$\frac{(e^{-(x_i - \theta_i)/\sigma_i})^{s+1}}{(1 + e^{-(x_i - \theta_i)/\sigma_i})^{2s+2}} \cdot e^{-\frac{(\theta_i - \mu_i)^2}{2\tau_i^2}}, \qquad -\infty < \theta_i < \infty. \tag{4}$$

The selection procedure will be based on the sample medians X_i. An estimator of Θ_i given $X_i = x_i$ is the median of the posterior distribution of Θ_i. For $i = 1, \ldots, k$, denote $\varphi_i(x_i)$ to be the median of the posterior distribution of Θ_i given $X_i = x_i$.

Let $\underset{\sim}{X}=(X_1, \ldots, X_k)$ and $\mathcal{X}$ be the sample space generated by $\underset{\sim}{X}$. A selection procedure $\underset{\sim}{d}=(\underset{\sim}{d}_0, \ldots, \underset{\sim}{d}_k)$ is a mapping defined on the sample space $\mathcal{X}$. For every $\underset{\sim}{x} \in \mathcal{X}$, $\underset{\sim}{d}_i(\underset{\sim}{x}), i = 1, \ldots, k$, is the probability of selecting population Π_i as the best among the k populations and also good compared with the given control θ_0, $\underset{\sim}{d}_0(\underset{\sim}{x})$ is the probability of excluding all k populations as bad and selecting none. Also, $\sum_{i=0}^{k} d_i(\underset{\sim}{x}) = 1$, for all $\underset{\sim}{x} \in \mathcal{X}$.

We next derive a selection rule $d(\underset{\sim}{x})$ based on the posterior median $\varphi_i(x_i)$, $i = 1, \ldots, k$. For each $\underset{\sim}{x} \in \mathcal{X}$, let $I(\underset{\sim}{x}) = \{i | \varphi_i(x_i) = \max_{0 \leq j \leq k} \varphi_j(x_j), i = 0, 1, \ldots, k\}$, and $i^* = \min\{i | i \in I(\underset{\sim}{x})\}$. Then based on $\varphi_i(x_i)$, a selection procedure $d(\underset{\sim}{x}) = (d_0(\underset{\sim}{x}), \ldots, d_k(\underset{\sim}{x}))$ is constructed as follows:

$$\begin{cases} d_{i^*}(\underset{\sim}{x}) = 1, \\ d_j(\underset{\sim}{x}) = 0, \quad \text{for} \quad j \neq i^*. \end{cases} \tag{5}$$

Under the preceding statistical model, the expected risk of the selection procedure $d(\underset{\sim}{x})$ is denoted by $R(d(\underset{\sim}{x}))$. Denote $h_i(\theta_i | \mu_i, \tau_i^2)$ to be the prior density function of Θ_i given (μ_i, τ_i^2), we have

$$R(d(\underset{\sim}{x})) = -\int_{\mathcal{X}} [\sum_{i=0}^{k} d_i(\underset{\sim}{x}) \varphi_i(x_i)] f(\underset{\sim}{x}) d(\underset{\sim}{x}) + C, \tag{6}$$

where

$C = \int_{\Omega} \max(\theta_{[k]}, \theta_0) dH(\underset{\sim}{\theta})$,
$H(\underset{\sim}{\theta})$: the joint distribution of $\underset{\sim}{\theta} = (\theta_1, \ldots, \theta_k)$,
$f_i(x_i) = \int_R f_i(x_i | \theta_i, \sigma_i^2) h_i(\theta_i | \mu_i, \tau_i^2) d\theta_i$,
$f(\underset{\sim}{x}) = \Pi_{i=1}^{k} f_i(x_i)$,
$\varphi_0(x_0) = \theta_0$.

Note that sample median X_i is not a sufficient statistic for θ_i (the observation vector is a minimal sufficient statistic). So $d(\underset{\sim}{x})$ may not be a Bayes rule. Also, the selection procedure $d(\underset{\sim}{x})$ defined above depends on the unknown parameters (μ_i, τ_i^2), $i = 1, \ldots, k$ and the specific form of $\varphi_i(x_i)$. Since the parameters and the specific form of $\varphi_i(x_i)$ are both unknown, it is impossible to implement this selection procedure for the selection problem in practice.

To derive a practical selection rule, we assume there are past observations when the present selection is to be made. At time $l = 1, \ldots, n$, let X_{ijl} be the j-th observation from Π_i, that is, for each $i = 1, \ldots, k$, let

$$\Theta_{il} \sim N(\mu_i, \tau_i^2), \qquad l = 1, \ldots, n, \tag{7}$$

and

$$X_{ijl} \sim L(\theta_{il}, \sigma_i^2), \qquad j = 1, \ldots, M. \tag{8}$$

For $l = 1, \ldots, n$, denote $X_{i,l}$ to be the median of $(X_{i1l}, \ldots, X_{iMl})$, and

$$X_i(n) = \frac{1}{n}\sum_{l=1}^{n} X_{i,l}, \tag{9}$$

$$S_i^2(n) = \frac{1}{n-1}\sum_{l=1}^{n}(X_{i,l} - X_i(n))^2. \tag{10}$$

Then,

$$E(X_{i,l}) = E(E(X_{i,l}|\Theta_{il} = \theta_{il})) = E(\Theta_{il}) = \mu_i, \tag{11}$$

and

$$\begin{aligned} Var(X_{i,l}) &= Var(E(X_{i,l}|\theta_{il})) + E(Var(X_{i,l}|\theta_{il})) \\ &= Var(\Theta_{il}) + E(Var(X_{i,l}|\theta_{il})) \\ &= \tau_i^2 + E(Var(X_{i,l}|\theta_{il})) \\ &< \infty. \end{aligned} \tag{12}$$

Denote $\nu_i^2 = Var(X_{i,l})$. Since $(X_{i1}, \ldots, X_{in})$ are i.i.d., by the strong law of large numbers, we know that as $n \to \infty$,

$$\begin{cases} X_i(n) \longrightarrow \mu_i, & a.s. \\ S_i^2(n) \longrightarrow \nu_i^2, & a.s. \end{cases} \tag{13}$$

To derive an empirical selection procedure, we first consider the following lemmas.

Lemma 1 Let $\{Y_i, 1 \leq i \leq m\}$ be m i.i.d. random observations from continuous distribution function F; also let $\hat{\xi}$ and ξ be the sample median of $\{Y_i, 1 \leq i \leq m\}$ and population median of F, respectively. Then, for any $\epsilon > 0$,

$$P\{|\hat{\xi} - \xi| > \epsilon\} \leq 2e^{-2m\delta_\epsilon^2}, \tag{14}$$

where $\delta_\epsilon = \min\{F(\xi + \epsilon) - \frac{1}{2}, \frac{1}{2} - F(\xi - \epsilon)\}$.
This lemma is from Serfling (1980) and the proof can be found in it.

Back to our selection problem. Put

$$\sigma' = \min_{1 \leq i \leq k} \sigma_i, \quad \sigma^* = \max_{1 \leq i \leq k} \sigma_i.$$

$X_{i1}, \ldots, X_{iM}$ are i.i.d. from $L(\theta_i, \sigma_i^2)$, which has the following cumulative distribution function

$$F(t_i) = \frac{1}{1 + e^{-(t_i - \theta_i)/\sigma_i}} \qquad -\infty < t_i < \infty, \tag{15}$$

and for $0 < \epsilon \leq \sigma'$,

$$F(\theta_i + \epsilon) - \frac{1}{2} = \frac{1}{2} - F(\theta_i - \epsilon) = \frac{e^{\epsilon/\sigma_i} - 1}{2(e^{\epsilon/\sigma_i} + 1)} \geq \frac{\epsilon}{2(e+1)\sigma^*}. \tag{16}$$

Given $\Theta_i = \theta_i$, θ_i and X_i are the population median and sample median of $L(\theta_i, \sigma_i)$ respectively. We have, from Lemma 1,

$$P\{|X_i - \theta_i| > \epsilon\} \leq 2e^{\frac{-(2s+1)\epsilon^2}{2(e+1)^2\sigma^{*2}}}. \tag{17}$$

For any $0 < \epsilon \leq \sigma'$, denote $\mathcal{A}_i = \{\underset{\sim}{x} \in \mathcal{X} : \ |x_i - \theta_i| \leq \epsilon\}$. We show that the conditional density of X_i given θ_i and σ_i^2 is approximately $N(\theta_i, \frac{2}{s+1}\sigma_i^2)$ as $s \to \infty$.

From (2), the conditional density of X_i given θ_i and σ_i^2 is

$$\begin{aligned} f_i(x_i|\theta_i, \sigma_i^2) &= \frac{(2s+1)!}{(s!)^2} \frac{1}{\sigma_i} \frac{(e^{-(x_i-\theta_i)/\sigma_i})^{s+1}}{(1 + e^{-(x_i-\theta_i)/\sigma_i})^{2s+2}} \\ &= \frac{(2s+1)!}{(s!)^2} \frac{1}{\sigma_i} \frac{1}{(2 + e^{-(x_i-\theta_i)/\sigma_i} + e^{(x_i-\theta_i)/\sigma_i})^{s+1}}. \end{aligned} \tag{18}$$

By Stirling's formula, when s is large enough,

$$\frac{(2s+1)!}{(s!)^2} \approx \frac{2^{(2s+\frac{3}{2})}}{\sqrt{2\pi}} \sqrt{s+1}. \tag{19}$$

Also choosing $\epsilon = \epsilon_s \downarrow 0$ to be a sequence of fixed numbers which tend to 0 as $s \to \infty$, by Taylor's polynomial expansion, we have

$$\log(2 + e^{-(x_i-\theta_i)/\sigma_i} + e^{(x_i-\theta_i)/\sigma_i}) \approx \log 4 + \frac{1}{4} \frac{(x_i - \theta_i)^2}{\sigma_i^2} \tag{20}$$

on $\mathcal{A}_i$. When $s \to \infty$, from (17),

$$P\{\underset{\sim}{X} \notin \mathcal{A}_i\} \leq 2e^{\frac{-(2s+1)\epsilon^2}{2(e+1)^2\sigma^{*2}}} \longrightarrow 0. \tag{21}$$

Therefore, we see that as $s \to \infty$,

$$f_i(x_i|\theta_i, \sigma_i^2) \approx \frac{1}{\sqrt{2\pi}\sqrt{2/s+1}} \frac{1}{\sigma_i} e^{-\frac{s+1}{4}\frac{(x_i-\theta_i)^2}{\sigma_i^2}}, \tag{22}$$

that is, $f_i(x_i|\theta_i, \sigma_i^2)$ is approximately $N(\theta_i, \frac{2}{s+1}\sigma_i^2)$.

From above, we can see that for sufficiently large s, the conditional density of $X_{i,l}$ is approximately $N(\theta_i, \frac{2}{s+1}\sigma_i^2)$ given θ_i and σ_i. Since the prior distribution of θ_i is $N(\mu_i, \tau_i^2)$, the unconditional density of $X_{i,l}$ is approximately $N(\mu_i, \tau_i^2 + \frac{2}{s+1}\sigma_i^2)$.

For each population Π_i, let $W_i^2(n)$ be the measure of the overall sample variation for the past observations. That is,

$$\begin{cases} \bar{X}_{il} = \frac{1}{M}\sum_{j=1}^{M} X_{ijl}, \\ W_i^2(n) = \frac{1}{(M-1)n}\sum_{j=1}^{M}\sum_{l=1}^{n}(X_{ijl} - \bar{X}_{il})^2. \end{cases} \tag{23}$$

Then we define, for $i = 1, \ldots, k$,

$$\begin{cases} \hat{\mu}_i = X_i(n), \\ \hat{\sigma}_i^2 = \frac{3}{\pi^2}W_i^2(n), \\ \hat{\nu}_i^2 = S_i^2(n), \\ \hat{\tau}_i^2 = \max(\hat{\nu}_i^2 - \frac{2}{s+1}\hat{\sigma}_i^2, 0). \end{cases} \tag{24}$$

and

$$\begin{cases} \hat{\varphi}_i(x_i) = \begin{cases} (x_i\hat{\tau}_i^2 + \frac{2\hat{\sigma}_i^2}{s+1}\hat{\mu}_i)/\hat{\nu}_i^2, & \text{if } \hat{\nu}_i^2 - \frac{2}{s+1}\hat{\sigma}_i^2 > 0, \\ \hat{\mu}_i, & \text{if } \hat{\nu}_i^2 - \frac{2}{s+1}\hat{\sigma}_i^2 \le 0, \end{cases} \\ \hat{\varphi}_0(x_0) = \theta_0. \end{cases} \tag{25}$$

Then for each $\underline{x} \in \mathcal{X}$, let $\hat{I}(\underline{x}) = \{i|\hat{\varphi}_i(x_i) = \max_{0\le j\le k}\hat{\varphi}_j(x_j), i = 0, 1, \ldots, k\}$, and $\hat{i}^* = \min\{i|i \in \hat{I}(\underline{x})\}$. We propose the following selection procedure $d^{(n,s)}(\underline{x}) = (d_0^{(n,s)}(\underline{x}), \ldots, d_k^{(n,s)}(\underline{x}))$ as follows:

$$\begin{cases} d_{\hat{i}^*}^{(n,s)} = 1, \\ d_j^{(n,s)} = 0, \quad \text{for} \quad j \ne \hat{i}^*. \end{cases} \tag{26}$$

3 Asymptotic Optimality of the Proposed Selection Procedure

Consider the selection procedure $d^{(n,s)}(\underline{x})$ constructed in (26). $d^{(n,s)}(\underline{x})$ is similar to selection rule $d(\underline{x})$ except that normal approximation is used to estimate the unknown prior parameters and the specific form of $\varphi_i(x_i)$ for $d^{(n,s)}(\underline{x})$. A natural question to ask is: How good is the selection rule $d^{(n,s)}(\underline{x})$ compared with $d(\underline{x})$? Let $R(d^{(n,s)}(\underline{x}))$ be the conditional expected risk given the past observations $\{X_{ijl}, i = 1, \ldots, k, j = 1, \ldots, M, \text{ and } l = 1, \ldots, n\}$, then

$$R(d^{(n,s)}(\underline{x})) = -\int_{\mathcal{X}}[\sum_{i=0}^{k} d_i^{(n,s)}(\underline{x})\varphi_i(x_i)]f(\underline{x})d(\underline{x}) + C. \tag{27}$$

Since $d^{(n,s)}(\underline{x})$ is mimicking $d(\underline{x})$, $R(d^{(n,s)}(\underline{x})) - R(d(\underline{x}))$ should be close to 0 if the empirical selection rule works well. Note that $R(d^{(n,s)}(\underline{x})) - R(d(\underline{x}))$ can be negative because $d(\underline{x})$ is not a Bayes rule. Therefore, we use the overall integrated risk $E|R(d^{(n,s)}(\underline{x})) - R(d(\underline{x}))| \geq 0$ as a measure of the performance of the selection procedure $d^{(n,s)}(\underline{x})$, where E is the expectation taken with respect to the past observations $\{X_{ijl}\}$.

We first state some facts about $\varphi_i(x_i)$, the posterior median of θ_i given $X_i = x_i$ and μ_i. From the definition of $\varphi_i(x_i)$, we can see that $\varphi_i(x_i)$ is between x_i and μ_i. Besides,

Lemma 2 When s is large enough, for $1 \leq i \leq k$,

$$|\varphi_i(x_i) - x_i| \leq 2\sigma_i\sqrt{\frac{\log s}{s}} \tag{28}$$

Proof. We only prove $\varphi_i(x_i) \leq x_i + 2\sigma_i\sqrt{\frac{\log s}{s}}$ here. The proof of $\varphi_i(x_i) \geq x_i - 2\sigma_i\sqrt{\frac{\log s}{s}}$ is similar. To prove $\varphi_i(x_i) \leq x_i + 2\sigma_i\sqrt{\frac{\log s}{s}}$, it suffices to show that

$$\begin{aligned}
&\int_{x_i+2\sigma_i\sqrt{\frac{\log s}{s}}}^{\infty} f_i(x_i|\theta_i,\sigma_i^2)h_i(\theta_i|\mu_i,\tau_i^2)d\theta_i \\
= &\int_{x_i+2\sigma_i\sqrt{\frac{\log s}{s}}}^{\infty} \frac{(2s+1)!}{(s!)^2}\frac{1}{\sigma_i}\frac{1}{\sqrt{2\pi}\tau_i}\frac{(e^{(\theta_i-x_i)/\sigma_i})^{s+1}}{(1+e^{(\theta_i-x_i)/\sigma_i})^{2s+2}}\cdot e^{-\frac{(\theta_i-\mu_i)^2}{2\tau_i^2}}d\theta_i \\
= &\int_{2\sqrt{\frac{\log s}{s}}}^{\infty} \frac{1}{\sqrt{2\pi}\tau_i}\frac{(2s+1)!}{(s!)^2}\left(\frac{e^{\theta}}{(1+e^{\theta})^2}\right)^{s+1}\cdot e^{-\frac{(\sigma_i\theta+x_i-\mu_i)^2}{2\tau_i^2}}d\theta \longrightarrow 0,
\end{aligned} \tag{29}$$

as $s \to \infty$. We first show

$$t(\theta,s) := \frac{(2s+1)!}{(s!)^2}\left(\frac{e^{\theta}}{(1+e^{\theta})^2}\right)^{s+1} \longrightarrow 0 \qquad as\ s \to \infty, \tag{30}$$

uniformly for $\theta \geq 2\sqrt{\frac{\log s}{s}}$. Obviously it is enough to consider the case of $\theta = 2\sqrt{\frac{\log s}{s}}$ since $t(\theta,s)$ is decreasing on $\theta > 0$. When $\theta = 2\sqrt{\frac{\log s}{s}}$ and s is large enough, by Taylor's formula,

$$\log(1+e^{\theta}) = \log 2 + \frac{1}{2}\theta + \frac{1}{8}\theta^2 + o(\theta^2), \tag{31}$$

and by (19), when s is large enough,

$$\log\frac{(2s+1)!}{(s!)^2} \leq 2(s+1)\log 2 + \frac{1}{2}\log(s+1). \tag{32}$$

From (31) and (32), we obtain that

$$\log t(\theta,s) \;=\; (s+1)[\theta - 2\log(1+e^{\theta})] + \log\frac{(2s+1)!}{(s!)^2}$$

$$
\begin{aligned}
&\leq -2(s+1)\log 2 - \frac{s+1}{4}\theta^2 + 2(s+1)\log 2 + \frac{1}{2}\log s + o(s\theta^2) \\
&= -(\frac{s+1}{s} - \frac{1}{2})\log s + o(\log s) \longrightarrow -\infty, \qquad (33)
\end{aligned}
$$

as $s \to \infty$. Therefore, (30) is proved, from which we can immediately see that (29) holds true. It completes the proof of Lemma 2.

The next lemma is well known and can be found in Baum and Katz (1965).

Lemma 3 Let $X_1, \ldots, X_n$ be i.i.d. random variables with mean 0. Suppose for $\alpha > 1$, $E|X_i|^\alpha < \infty$, for $i = 1, \ldots, n$, then for any $\epsilon > 0$,

$$
P\{|\sum_{i=1}^{n} X_i/n| \geq \epsilon\} = o(n^{-(\alpha-1)}). \qquad (34)
$$

As a consequence of Lemma 3, we have

Lemma 4 Let $X_1, \ldots, X_n$ be independent random variables, with mean $EX_i = \mu$ and variance $\mathrm{Var}X_i = \sigma^2$, for $i = 1, \ldots, n$. Also let $\bar{X} = \frac{1}{n}\sum X_i$ and $S_n^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$. Suppose for $i = 1, \ldots, n$ and a fixed number $\alpha > 2$, $E|X_i|^\alpha < \infty$, then for any $\epsilon > 0$,

$$
P\{|S_n^2 - \sigma^2| \geq \epsilon\} = o(n^{-(\alpha/2-1)}). \qquad (35)
$$

Since $EX_{i.l}^4 < \infty$, for any $\epsilon > 0$, by Lemma 3,

$$
P\{|\hat{\mu}_i - \mu_i| \geq \epsilon\} = o(n^{-3}), \qquad (36)
$$

also by Lemma 4,

$$
P\{|\hat{\nu}_i^2 - \nu_i^2| \geq \epsilon\} = o(n^{-1}). \qquad (37)
$$

Similarly, we have for any $\epsilon > 0$,

$$
P\{|\hat{\sigma}_i^2 - \sigma_i^2| \geq \epsilon\} = o(n^{-1}). \qquad (38)
$$

When s is large enough, $\nu_i^2 - \frac{2}{s+1}\sigma_i^2 > 0$. Therefore, from (37) and (38), when s is sufficiently large,

$$
P\{\hat{\nu}_i^2 - \frac{2}{s+1}\hat{\sigma}_i^2 \leq 0\} = o(n^{-1}). \qquad (39)
$$

Besides, $\tau_i^2 = \nu_i^2 - E(Var(X_{i,l}|\theta_i))$ by (12) and

$$
\begin{aligned}
E(Var(X_{i,l}|\theta_i)) &= \int_{-\infty}^{\infty} (x_{il} - \theta_i)^2 \frac{(2s+1)!}{(s!)^2} \frac{1}{\sigma_i} \frac{(e^{-(x_{il}-\theta_i)/\sigma_i})^{s+1}}{(1 + e^{-(x_{il}-\theta_i)/\sigma_i})^{2s+2}} dx_{il} \\
&= \sigma_i \int_{-\infty}^{\infty} x^2 \frac{(2s+1)!}{(s!)^2} \Big(\frac{e^x}{(1+e^x)^2}\Big)^{s+1} dx. \qquad (40)
\end{aligned}
$$

We have

Lemma 5

$$\int_{-\infty}^{\infty} x^2 \frac{(2s+1)!}{(s!)^2}\Big(\frac{e^x}{(1+e^x)^2}\Big)^{s+1} dx = o\Big(\sqrt{\frac{\log s}{s}}\Big). \tag{41}$$

Proof.

$$\begin{aligned}
& \int_{-\infty}^{\infty} x^2 \frac{(2s+1)!}{(s!)^2}\Big(\frac{e^x}{(1+e^x)^2}\Big)^{s+1} dx \\
= & \ 2\int_{0}^{\infty} x^2 \frac{(2s+1)!}{(s!)^2}\Big(\frac{e^x}{(1+e^x)^2}\Big)^{s+1} dx \\
= & \ 2\Big(\int_0^{\sqrt{8\frac{\log s}{s}}} + \int_{\sqrt{8\frac{\log s}{s}}}^{3} + \int_3^{\infty}\Big) x^2 \frac{(2s+1)!}{(s!)^2}\Big(\frac{e^x}{(1+e^x)^2}\Big)^{s+1} dx \\
:= & \ T_1 + T_2 + T_3.
\end{aligned} \tag{42}$$

By Stirling's formula, when s is large enough,

$$\begin{aligned}
T_1 &= 2\frac{(2s+1)!}{(s!)^2}\int_0^{\sqrt{8\frac{\log s}{s}}} x^2\Big(\frac{e^x}{(1+e^x)^2}\Big)^{s+1} dx \\
&\leq 2\cdot 2^{2(s+1)}\sqrt{s+1}\cdot 2^{-2(s+1)}\int_0^{\sqrt{8\frac{\log s}{s}}} x^2 dx \\
&\leq \sqrt{s+1}\Big(8\frac{\log s}{s}\Big)^{3/2} \\
&= o\Big(\sqrt{\frac{\log s}{s}}\Big).
\end{aligned} \tag{43}$$

Using the same approach as in the proof of Lemma 2, we have

$$\begin{aligned}
T_2 &= 2\frac{(2s+1)!}{(s!)^2}\int_{\sqrt{8\frac{\log s}{s}}}^{3} x^2\Big(\frac{e^x}{(1+e^x)^2}\Big)^{s+1} dx \\
&\leq 2\frac{(2s+1)!}{(s!)^2}\Big(\frac{e^{\sqrt{8\frac{\log s}{s}}}}{(1+e^{\sqrt{8\frac{\log s}{s}}})^2}\Big)^{s+1}\int_{\sqrt{8\frac{\log s}{s}}}^{3} x^2 dx \\
&= o\Big(\sqrt{\frac{\log s}{s}}\Big).
\end{aligned} \tag{44}$$

Moreover,

$$\begin{aligned}
T_3 &= 2\frac{(2s+1)!}{(s!)^2}\int_3^{\infty} x^2\Big(\frac{e^x}{(1+e^x)^2}\Big)^{s+1} dx \\
&\leq 2\frac{(2s+1)!}{(s!)^2}\int_3^{\infty} x^2 e^{-(s+1)x} dx \\
&= o(\sqrt{\frac{\log s}{s}}).
\end{aligned} \tag{45}$$

This completes the proof of Lemma 5.

From Lemma 5, we observe that when s is sufficiently large,

$$E(Var(X_{i,l}|\theta_i)) = o\Big(\sqrt{\frac{\log s}{s}}\Big), \tag{46}$$

and therefore, by (37), (39) and the definition of $\hat{\tau}_i^2$, for $\epsilon \geq c\sqrt{\frac{\log s}{s}}$, where $c > 0$,

$$P\{|\hat{\tau}_i^2 - \tau_i^2| \geq \epsilon\} = o(n^{-1}), \tag{47}$$

and furthermore,

$$P\{\hat{\nu}_i^2/\hat{\tau}_i^2 \leq \nu_i^2/(2\tau_i^2)\} = o(n^{-1}). \tag{48}$$

Next we investigate the overall integrated risk $E|R(d^{(n,s)}(\underline{x})) - R(d(\underline{x}))|$. Let $P_{n,s}$ be the probability measure generated by the past observations $X_{ijl}, i = 1,\dots k$, $j = 1,\dots M$ and $l = 1,\dots,n$.

$$\begin{aligned}
& E|R(d^{(n,s)}(\underline{x})) - R(d(\underline{x}))| \\
\leq\ & \sum_{i=0}^{k}\sum_{j=0}^{k}\int_{\mathcal{X}} P_{n,s}\{i^* = i, \hat{i}^* = j\}|\varphi_i(x_i) - \varphi_j(x_j)|f(\underline{x})d\underline{x} \\
=\ & \sum_{i=1}^{k}\int_{\mathcal{X}} P_{n,s}\{i^* = i, \hat{i}^* = 0\}|\varphi_i(x_i) - \theta_0|f(\underline{x})d\underline{x} \\
& +\sum_{j=1}^{k}\int_{\mathcal{X}} P_{n,s}\{i^* = 0, \hat{i}^* = j\}|\theta_0 - \varphi_j(x_j)|f(\underline{x})d\underline{x} \\
& +\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{\mathcal{X}} P_{n,s}\{i^* = i, \hat{i}^* = j\}|\varphi_i(x_i) - \varphi_j(x_j)|f(\underline{x})d\underline{x} \\
\leq\ & 2\sum_{i=1}^{k}\int_{R} P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > |\varphi_i(x_i) - \theta_0|\}|\varphi_i(x_i) - \theta_0|f_i(x_i)dx_i \\
& +2\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2} P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > \frac{|\varphi_i(x_i) - \varphi_j(x_j)|}{2}\}|\varphi_i(x_i) - \varphi_j(x_j)| \\
& \qquad \times f_i(x_i)f_j(x_j)dx_idx_j \\
:=\ & I_1 + I_2.
\end{aligned} \tag{49}$$

For any $\epsilon > 0$, and $i, j = 1,\dots,k$, let

$$\begin{cases} \mathcal{X}_i = \{x_i : |\varphi_i(x_i) - \theta_0| \leq \epsilon\}, \\ \mathcal{X}_{ij} = \{(x_i, x_j) : |\varphi_i(x_i) - \varphi_j(x_j)| \leq \epsilon\}. \end{cases} \tag{50}$$

Then we have

$$I_1 = 2\sum_{i=1}^{k}\int_{\mathcal{X}_i} P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > |\varphi_i(x_i) - \theta_0|\}|\varphi_i(x_i) - \theta_0|f_i(x_i)dx_i$$

$$
\begin{aligned}
&+2\sum_{i=1}^{k}\int_{R-\mathcal{X}_i} P_{n,s}\{|\hat{\varphi}_i(x_i)-\varphi_i(x_i)|>|\varphi_i(x_i)-\theta_0|\}|\varphi_i(x_i)-\theta_0|f_i(x_i)dx_i \\
\leq\ & 2\sum_{i=1}^{k}\int_{\mathcal{X}_i}\epsilon f_i(x_i)dx_i \\
&+2\sum_{i=1}^{k}\int_{R} P_{n,s}\{|\hat{\varphi}_i(x_i)-\varphi_i(x_i)|>\epsilon\}|\varphi_i(x_i)-\theta_0|f_i(x_i)dx_i. \qquad (51)
\end{aligned}
$$

By Lemma 2, when s is large enough, $|\varphi_i(x_i)-x_i|\leq 2\sigma_i\sqrt{\frac{\log s}{s}}$. From now on, we always set $\epsilon = 16\sigma^*\sqrt{\frac{\log s}{s}}$. Therefore, for sufficiently large s,

$$
|x_i-\theta_0|\leq|\varphi_i(x_i)-x_i|+|\varphi_i(x_i)-\theta_0|\leq 2\epsilon \qquad (52)
$$

on $\mathcal{X}_i$ and

$$
\begin{aligned}
\int_{\mathcal{X}_i} f_i(x_i)dx_i &\leq \int_{\{|x_i-\theta_0|\leq 2\epsilon\}} f_i(x_i)dx_i \\
&\leq \int_{\{|x_i-\theta_0|\leq 2\epsilon\}} \frac{1}{\sqrt{2\pi}\tau_i}dx_i \\
&= \frac{4\epsilon}{\sqrt{2\pi}\tau_i}. \qquad (53)
\end{aligned}
$$

Thus,

$$
\begin{aligned}
&I_1 \leq \frac{8k}{\sqrt{2\pi}\tau_i}\epsilon^2 \\
&+2\sum_{i=1}^{k}\int_{R} P_{n,s}\{|\hat{\varphi}_i(x_i)-\varphi_i(x_i)|>\epsilon\}[|\varphi_i(x_i)-\mu_i|+|\mu_i-\theta_0|]f_i(x_i)dx_i. \quad (54)
\end{aligned}
$$

Moreover,

$$
\begin{aligned}
I_2 \ =\ & 2\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{\mathcal{X}_{ij}} P_{n,s}\{|\hat{\varphi}_i(x_i)-\varphi_i(x_i)|>\frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}|\varphi_i(x_i)-\varphi_j(x_j)| \\
&\qquad\qquad \times f_i(x_i)f_j(x_j)dx_idx_j \\
&+2\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2-\mathcal{X}_{ij}} P_{n,s}\{|\hat{\varphi}_i(x_i)-\varphi_i(x_i)|>\frac{|\varphi_i(x_i)-\varphi_j(x_j)|}{2}\}|\varphi_i(x_i)-\varphi_j(x_j)| \\
&\qquad\qquad \times f_i(x_i)f_j(x_j)dx_idx_j \\
\leq\ & 2\epsilon\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{\mathcal{X}_{ij}} f_i(x_i)f_j(x_j)dx_idx_j \\
&+2\sum_{i=1}^{k}\sum_{j=1}^{k}\int_{R^2} P_{n,s}\{|\hat{\varphi}_i(x_i)-\varphi_i(x_i)|>\frac{\epsilon}{2}\}|\varphi_i(x_i)-\varphi_j(x_j)| \\
&\qquad\qquad \times f_i(x_i)f_j(x_j)dx_idx_j. \qquad (55)
\end{aligned}
$$

From (28), when s is large enough, $|\varphi_i(x_i) - x_i| \le \epsilon$ and $|\varphi_j(x_j) - x_j| \le \epsilon$. Therefore, when s is sufficiently large,

$$\{(x_i, x_j) : |\varphi_i(x_i) - \varphi_j(x_j)| \le \epsilon\} \subset \{(x_i, x_j) : |x_i - x_j| \le 3\epsilon\}. \tag{56}$$

Thus, similar to (53),

$$\int_{\mathcal{X}_{ij}} f_i(x_i) f_j(x_j) dx_i dx_j \le \frac{6\epsilon}{\sqrt{2\pi}\min(\tau_i, \tau_j)}. \tag{57}$$

We observe that

$$\begin{aligned} I_2 \ \le \ & \sum_{i=1}^{k}\sum_{j=1}^{k} \frac{12\epsilon^2}{\sqrt{2\pi}\min(\tau_i, \tau_j)} \\ & +2\sum_{i=1}^{k}\sum_{j=1}^{k} \int_{R^2} P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > \frac{\epsilon}{2}\}[|\varphi_i(x_i) - \mu_i| + |\varphi_j(x_j) - \mu_j| \\ & \qquad + |\mu_i - \mu_j|] f_i(x_i) f_j(x_j) dx_i dx_j. \end{aligned} \tag{58}$$

From (54) and (58), it suffices to analyze the limiting behaviors of

$$\begin{aligned} & \textstyle\int_R P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > \frac{\epsilon}{2}\} f_i(x_i) dx_i, \\ & \textstyle\int_R P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > \frac{\epsilon}{2}\} |\varphi_i(x_i) - \mu_i| f_i(x_i) dx_i. \end{aligned} \tag{59}$$

We first analyze $\int_R P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > \frac{\epsilon}{2}\} f_i(x_i) dx_i$. Denote

$$\begin{aligned} \mathcal{Y}_i &= \{x_i : |\varphi_i(x_i) - \theta_i| \le \tfrac{\epsilon}{4}\}, \\ \mathcal{Z}_i &= \{x_i : |x_i - \theta_i| \le \tfrac{\epsilon}{8}\}. \end{aligned} \tag{60}$$

By Lemma 2, we know that when s is large enough, $|\varphi_i(x_i) - x_i| \le \frac{\epsilon}{8}$. Therefore, for sufficiently large s, we have

$$R - \mathcal{Y}_i \subset R - \mathcal{Z}_i \tag{61}$$

and

$$\begin{aligned} & \int_R P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > \frac{\epsilon}{2}\} f_i(x_i) dx_i \\ \le \ & \int_R \left(\int_{R-\mathcal{Z}_i} P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > \frac{\epsilon}{2}\} f_i(x_i|\theta_i, \sigma_i^2) h_i(\theta_i|\mu_i, \tau_i^2) dx_i \right) d\theta_i \\ & + \int_R \left(\int_{\mathcal{Y}_i} P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > \frac{\epsilon}{2}\} f_i(x_i|\theta_i, \sigma_i^2) h_i(\theta_i|\mu_i, \tau_i^2) dx_i \right) d\theta_i \\ \le \ & \int_R \left(\int_{R-\mathcal{Z}_i} f_i(x_i|\theta_i, \sigma_i^2) h_i(\theta_i|\mu_i, \tau_i^2) dx_i \right) d\theta_i \\ & + \int_R \left(\int_R P_{n,s}\{|\hat{\varphi}_i(x_i) - \theta_i| \ge \frac{\epsilon}{4}\} f_i(x_i|\theta_i, \sigma_i^2) h_i(\theta_i|\mu_i, \tau_i^2) dx_i \right) d\theta_i \\ \le \ & \int_R \left(\int_{|x_i - \theta_i| > \frac{\epsilon}{8}} f_i(x_i|\theta_i, \sigma_i^2) dx_i \right) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \end{aligned}$$

$$
\begin{aligned}
& + \int_R \Big(\int_R P_{n,s}\{|\hat{\varphi}_i(x_i) - \theta_i| \geq \frac{\epsilon}{4}, \hat{\nu}_i^2 - 2\hat{\sigma}_i^2/(s+1) > 0\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) \\
& \qquad \times h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + \int_R \Big(\int_R P_{n,s}\{\hat{\nu}_i^2 - 2\hat{\sigma}_i^2/(s+1) \leq 0\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
\leq \; & 2 \int_R e^{-\frac{(2s+1)\epsilon^2}{128(e+1)^2\sigma^{*2}}} h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + \int_R \Big(\int_R P_{n,s}\{|(x_i\hat{\tau}_i^2 + \frac{2\hat{\sigma}_i^2}{s+1}\hat{\mu}_i)/\hat{\nu}_i^2 - \theta_i| \geq \frac{\epsilon}{4}\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + o(n^{-1}) \\
\leq \; & 2e^{-\frac{(2s+1)\epsilon^2}{128(e+1)^2\sigma^{*2}}} \\
& + \int_R \Big(\int_R P_{n,s}\{|x_i - \theta_i| \geq \frac{\hat{\nu}_i^2}{\hat{\tau}_i^2}\frac{\epsilon}{8}\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + \int_R \Big(\int_R P_{n,s}\{|\hat{\mu}_i - \theta_i| \geq \frac{(s+1)\hat{\nu}_i^2\epsilon}{16\hat{\sigma}_i^2}\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + o(n^{-1}) \\
\leq \; & O(s^{-1}) \\
& + \int_R \Big(\int_R P_{n,s}\{|x_i - \theta_i| \geq \frac{\hat{\nu}_i^2}{\hat{\tau}_i^2}\frac{\epsilon}{8}, \frac{\hat{\nu}_i^2}{\hat{\tau}_i^2} \geq \nu_i^2/(2\tau_i^2)\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + \int_R \Big(\int_R P_{n,s}\{\frac{\hat{\nu}_i^2}{\hat{\tau}_i^2} \leq \nu_i^2/(2\tau_i^2)\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + \int_R \Big(\int_R P_{n,s}\{|\hat{\mu}_i - \theta_i| \geq \frac{(s+1)\hat{\nu}_i^2\epsilon}{16\hat{\sigma}_i^2}, \hat{\nu}_i^2/\hat{\sigma}_i^2 \geq \nu_i^2/(2\sigma_i^2)\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) \\
& \qquad \times h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + \int_R \Big(\int_R P_{n,s}\{\hat{\nu}_i^2/\hat{\sigma}_i^2 \leq \nu_i^2/(2\sigma_i^2)\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + o(n^{-1}) \\
\leq \; & O(s^{-1}) + \int_R \Big(\int_{|x_i - \theta_i| \geq \frac{\nu_i^2}{2\tau_i^2}\frac{\epsilon}{16}} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i + o(n^{-1}) \\
& + \int_R \Big(\int_R P_{n,s}\{|\theta_i - \mu_i| \geq \frac{(s+1)\nu_i^2\epsilon}{64\sigma_i^2}\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + \int_R \Big(\int_R P_{n,s}\{|\hat{\mu}_i - \mu_i| \geq \frac{(s+1)\nu_i^2\epsilon}{64\sigma_i^2}\} f_i(x_i|\theta_i, \sigma_i^2) dx_i \Big) h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + o(n^{-1}) \\
\leq \; & O(s^{-1}) + 2e^{-\frac{(2s+1)\nu_i^4\epsilon^2}{2\times 32^2(e+1)^2\sigma^{*2}\tau_i^4}} + \int_{|\theta_i - \mu_i| \geq \frac{(s+1)\nu_i^2\epsilon}{64\sigma_i^2}} h_i(\theta_i|\mu_i, \tau_i^2) d\theta_i \\
& + o(n^{-3}) + o(n^{-1}) \\
\leq \; & O(s^{-1}) + e^{-\frac{(s+1)^2\nu_i^4\epsilon^2}{2\times 64^2\sigma_i^4\tau_i^2}} + o(n^{-1}) \\
= \; & O(s^{-1}) + o(n^{-1}). \qquad (62)
\end{aligned}
$$

Using similar approach, we can obtain

$$\int_R P_{n,s}\{|\hat{\varphi}_i(x_i) - \varphi_i(x_i)| > \frac{\epsilon}{2}\}|\varphi_i(x_i) - \mu_i|f_i(x_i)dx_i = o(\frac{1}{n}) + O(\frac{1}{s}). \qquad (63)$$

At the beginning of this paper, M is assumed to be an odd number. However, from the proof we can see that this condition can be dropped. In other words, no matter M is even or odd, the asymptotic property will hold true. Combining (49), (54), (58), (59), (62) and (63), we finally obtain the asymptotic property of the derived selection procedure.

Theorem 1 The selection procedure $d^{(n,s)}(\underline{x})$ defined in (26) is asymptotically optimal with a convergence rate of order $o(\frac{1}{n}) + O(\frac{\log s}{s})$. That is,

$$E|R(d^{(n,s)}(\underline{x})) - R(d(\underline{x}))| = o(\frac{1}{n}) + O(\frac{\log s}{s}). \qquad (64)$$

Theorem 1 establishes the rate of convergence of $E|R(d^{(n,s)}(\underline{x})) - R(d(\underline{x}))|$ as both n and s go to infinity in an additive form. This implies that $E|R(d^{(n,s)}(\underline{x})) - R(d(\underline{x}))|$ will converge to 0 when both n and s go to infinity.

4 Simulations

We carried out a simulation study to investigate the performance of the selection procedure $d^{(n,s)}(\underline{x})$. The overall integrated risk $E|R(d^{(n,s)}(\underline{x})) - R(d(\underline{x}))|$ is used as measure of the performance of the selection rule.

We consider the following case in which $k = 3$, that is, we have 3 logistic populations Π_1, Π_2 and Π_3 and we would like to use the proposed selection procedure to select the best population compared with a control.

The simulation scheme is described as follows:

(1) For each i, generate past observations as follows:

$$\begin{cases} \text{for } l = 1, \ldots, n, \\ \text{(a) first generate } \Theta_{il} \text{ from normal distribution with density } N(\mu_i, \tau_i^2), \\ \text{(b) then generate } X_{ijl} \text{ from logistic distribution } L(\theta_{il}, \sigma_i). \end{cases} \qquad (65)$$

(2) For each i, generate current observations Θ_i from $N(\mu_i, \tau_i^2)$ and $(X_{i1}, \ldots, X_{iM})$ i.i.d. from $L(\theta_i, \sigma_i)$.

(3) Based on the past observations X_{ijl}, and the present observations, we construct $d(\underline{x})$ and $d^{(n,s)}(\underline{x})$. Then compute the losses $L(d(\underline{x}))$ and $L(d^{(n,s)}(\underline{x}))$.

(4) Repeat Steps (2) and (3) 1000 times. Calculate the averages of the conditional losses $L(d(\underline{x}))$ and $L(d^{(n,s)}(\underline{x}))$, respectively. Denote the averages to be $\hat{R}(d(\underline{x}))$ and $\hat{R}(d^{(n,s)}(\underline{x}))$. Then compute the absolute difference

$$D = |\hat{R}(d^{(n,s)}(\underline{x})) - \hat{R}(d(\underline{x}))|. \qquad (66)$$

(5) Repeat steps (1), (2), (3) and (4) 5000 times. The average of the Ds in (66), denoted by $D(n,s)$, is used as an estimator of the differences $E|R(d^{(n,s)}(\underline{x})) - R(d(\underline{x}))|$.

Tables 1, 2, and 3 give the results of simulation for the performance of the proposed empirical selection procedures. We choose $\theta_0 = 0.5$, $\mu_1 = 0.4$, $\mu_2 = 0.5$, $\mu_3 = 0.6$, $\tau_1 = \tau_2 = \tau_3 = 1$, and $\sigma_1 = \sigma_2 = \sigma_3 = 1$. The related figures are also attached.

References

Balakrishnan, N. (1992). *Handbook of the Logistic Distribution.* (N. Balakrishnan Ed.), Marcel Dekker, New York.

Baum, L. E. and Katz, M. (1965). Convergence Rates in the Law of Large Numbers, Transactions of the Amer. Math. Society, Vol. 120, 1, 108-123.

Berkson, J. (1957). Tables for the maximum likelihood estimate of the logistic function. *Biometrics*, Vol. 13, 28–34.

Birnbaum, A. and Dudman, J. (1963). Logistic order statistics. *Annals of Mathematical Statistics*, Vol. 34, 658–663.

Gupta S. S. and Gnanadesikan, M. (1966). Estimation of the parameters of the logistic distribution. *Biometrika*, Vol. 53, 3 and 4, 565–570.

Gupta S. S. and Han, SangHyun (1991). An elimination type two-stage procedure for selecting the population with the largest mean from k logistic populations. *American Journal of Mathematical and Management Sciences*, Vol. 11, 351–370.

Gupta S. S. and Han, SangHyun (1992). Selection and ranking procedures for logistic populations. *Order Statistics and Nonparametrics: Theory and Applications.* (P. K. Sen and I. A. Salama, eds.), Elsevier Science B. V., 377–404.

Gupta S. S. and Liang T. (1999). On empirical Bayes simultaneous selection procedures for comparing normal populations with a standard. *Journal of Planning and Inference*, Vol. 77, 73–88.

Gupta, S. S. and Panchapakesan S. (1996). Design of experiments with selection and ranking goals. *Handbook of Statistics.* (S. Ghosh and C. R. Rao, eds.), Elsevier Science B. V., Vol. 13, 555-584.

Gupta, S. S., Qureishi, A. S., and Shah, B. K. (1967). Best linear unbiased estimators of the parameters of the logistic distribution using order statistics. *Technometrics*, Vol. 9, No. 1, 43–56.

Gupta, S. S. and Shah, B. K. (1965). Exact moments and percentage points of the order statistics and the distribution of the range from the logistic distribution. *Annals of Mathematical Statistics*, Vol. 36, 907–920.

Panchapakesan, S. (1992). Ranking and selection procedures. Chapter 6 in *the Handbook of the Logistic Distribution.* (N. Balakrishnan Ed.), Marcel Dekker, New York, 145–167.

Plackett, R. L. (1959). The analysis of life test data. *Technometrics*, Vol. 1, 9–19.

Serfling, J. R. (1980). Approximation Theorems of Mathematical Statistics, New York, Wiley.

Talacko, J. (1956). Perk's distributions and their role in the theory of Wiener's stochastic variates. *Trabajos de Estadistica*, Vol. 7, 159–174.

Table 1

Performance of the selection rule

when $s = 5$

n	$D(n, s)$	$SE(D(n, s))$
5	0.08509841	0.027865659
10	0.04994583	0.017846225
15	0.02375784	0.009525936
20	0.01263961	0.006187404
30	0.00738123	0.003425718
40	0.00651246	0.003046143
50	0.00519491	0.002195022
60	0.00415648	0.001433615
70	0.00381372	0.001086059
80	0.00361584	0.000818471
90	0.00347570	0.000713619
100	0.00338605	0.000689365
125	0.00316974	0.000642861
150	0.00309849	0.000627380

Table 2

Performance of the selection rule

when $s = 10$

n	$D(n, s)$	$SE(D(n, s))$
5	0.07304765	0.019237152
10	0.03923180	0.012678379
15	0.01681043	0.008253716
20	0.01064427	0.004316348
30	0.00517343	0.002579812
40	0.00325936	0.001308793
50	0.00216874	0.000845426
60	0.00157431	0.000393872
70	0.00128317	0.000251801
80	0.00107369	0.000137618
90	0.00098735	0.000113625
100	0.00089170	0.000093612
125	0.00084682	0.000079186
150	0.00081364	0.000072564

Table 3

Performance of the selection rule

when $s = 50$

n	$D(n,s)$	$SE(D(n,s))$
5	0.05812975	0.010374924
10	0.01450924	0.007316987
15	0.00810957	0.003264550
20	0.00496331	0.001684319
30	0.00156783	0.000873342
40	0.00084367	0.000439320
50	0.00058916	0.000231186
60	0.00033976	0.000124923
70	0.00026253	0.000095978
80	0.00023609	0.000075646
90	0.00019546	0.000057012
100	0.00017925	0.000043645
125	0.00013687	0.000037202
150	0.00012795	0.000030269

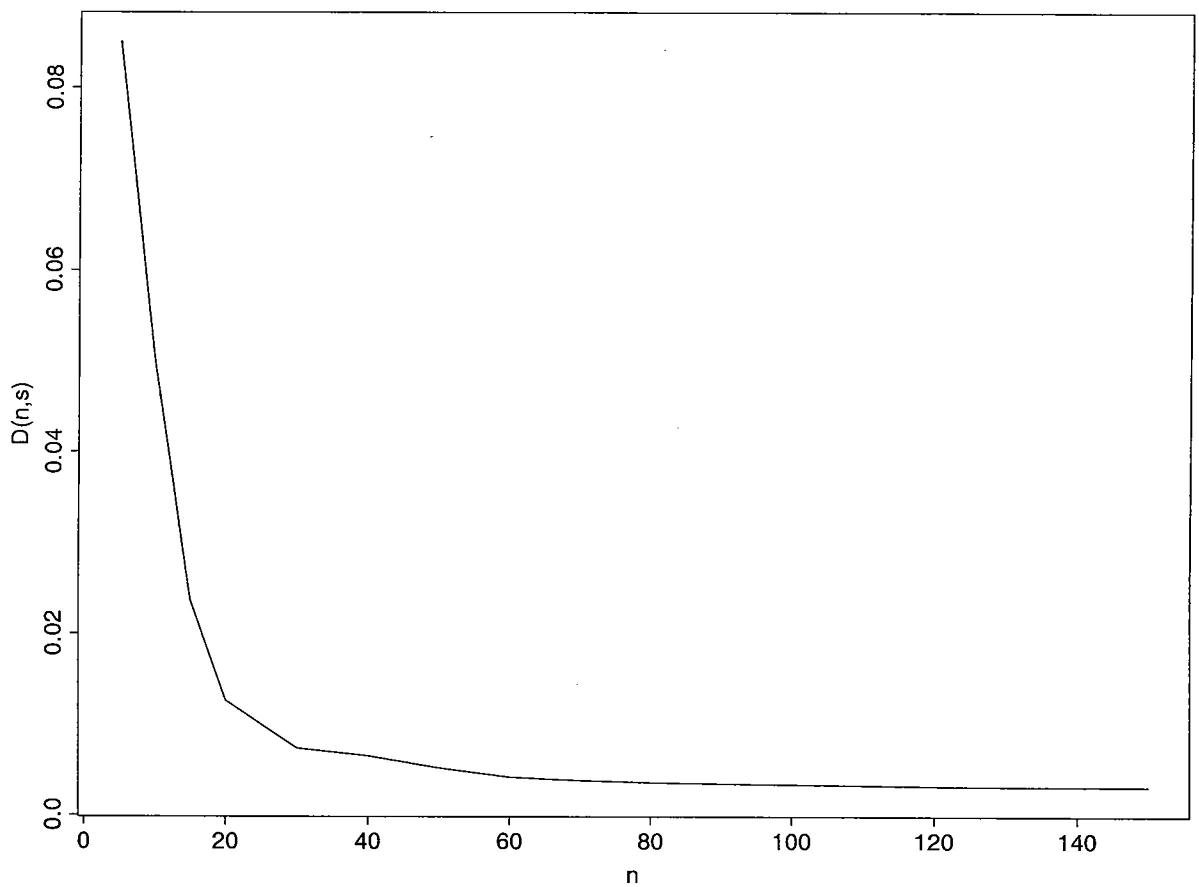

Figure 1: Graph for Table 1

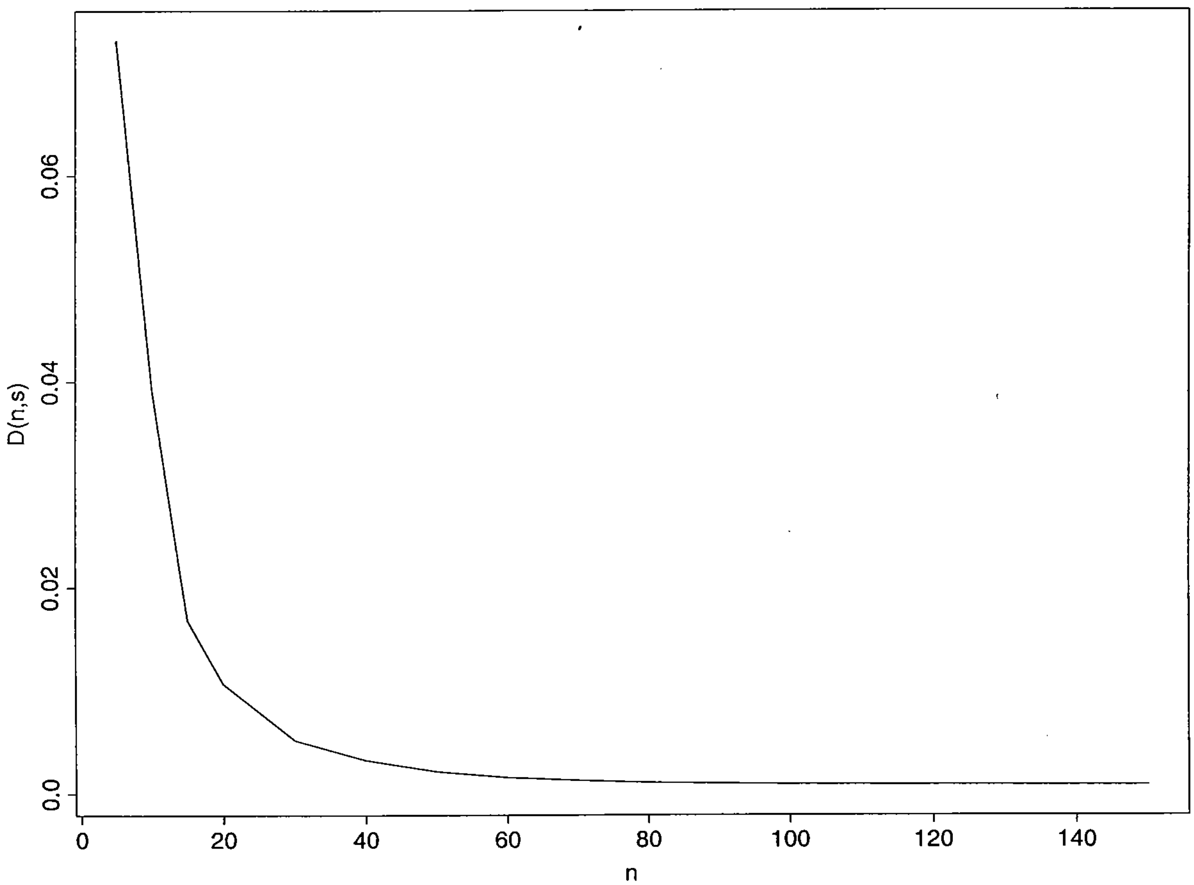

Figure 2: Graph for Table 2

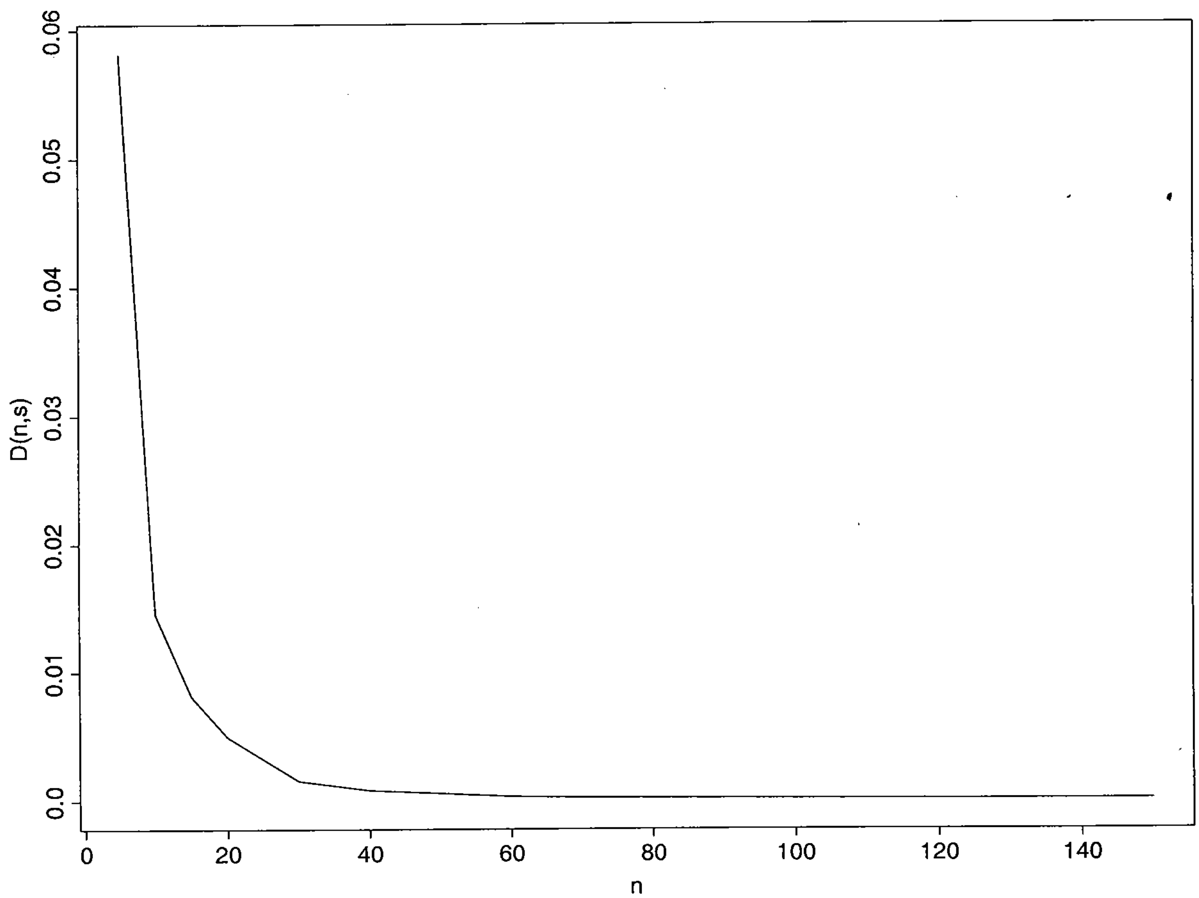

Figure 3: Graph for Table 3