



SCHOOL OF STATISTICS
UNIVERSITY OF THE PHILIPPINES DILIMAN



WORKING PAPER SERIES

PURPOSIVE SAMPLING AS AN OPTIMAL BAYES SAMPLING STRATEGY

by

Jacqueline M. Guarte¹
Erniel B. Barrios²

UPSS Working Paper No. 2008-03
September 2008

School of Statistics
Ramon Magsaysay Avenue
U.P. Diliman, Quezon City
Telefax: 928-08-81
Email: updstat@yahoo.com

¹Department of Mathematics, Physics and Statistics, Leyte State
University, Visca, Baybay, Leyte 6521-A, Philippines

²School of Statistics, University of the Philippines, Diliman, Quezon City
1101

ABSTRACT

Purposive sampling is described as a random selection of sampling units within the segment of the population with the most information on the characteristic of interest. The procedure effectively assigns zero probability of inclusion to units towards the tails of the distribution believed to have minimal information or none at all. This sampling strategy is shown to be optimal under the Bayesian framework. Simulation of the posterior risk indicate lower risk levels of the sample mean as estimator under purposive sampling relative to those under simple random sampling. This was consistently shown under the squared error loss function as effects of parameters of the posterior risk on its level were simulated.

Key Words: purposive sampling; posterior distribution; posterior risk; squared error loss function

1. INTRODUCTION

While probability sampling is the usual prescription in data generation for statistical inference in finite populations, nonprobability sampling is often the only recourse due to the nature of the population and the constraints faced in accessing the sampling units. Such populations include those that are difficult to define, locate, or recruit and those populations wherein some units may provide more information than the rest (relative to a characteristic of interest). Purposive sampling would suitably generate samples informative of the population constrained by peculiar behavior as stated above. In purposive sampling, the researcher's knowledge about the population is used to select the units to be included in the sample. In this study, purposive sampling is defined as random sampling from the most informative segment of the population.

Though bias is eminent in purposive samples, a certain degree of optimality is anticipated since the selection scheme would suitably fit in the Bayesian framework. Zacks (1969) shows that, generally, the optimal Bayes selection is a sequential one and

nonrandomized and proves that it is without replacement. In certain cases, the optimal Bayes sampling designs are nonrandomized and single-phase ones. A single-phase design is any design in which the units are drawn directly from the population. Zacks proves that any nonrandomized single-phase sampling plan without replacement is optimal if the posterior risk is independent of the observed values.

In this paper, we prove that purposive sampling satisfies the above condition and assess this optimality by simulating the posterior risk and comparing this to that under a design-unbiased sampling plan (i.e., simple random sampling). In the process, the important parameters that need to be considered when using purposive sampling are also examined.

2. PROVING THE OPTIMALITY OF PURPOSIVE SAMPLING

Consider a normal population. Let the variable of interest be X , $X \sim N(\theta, \sigma^2)$, σ^2 is known. Suppose we do purposive sampling. Say, we believe that the true mean of X is also normal with mean μ and some variance τ^2 . That is, $\theta \sim N(\mu, \tau^2)$. Consequently, we purposively select our sample in the vicinity of μ by intentionally excluding those in the tails of the distribution, i.e., assigning zero inclusion probability to the units towards the tails of the distribution. Then our sample is generated from a “new” population with a relatively small variance, $\frac{\sigma^2}{a}$, $a > 1$. The variable of interest in the “new” population, X' , is now distributed as $X' \sim N\left(\theta, \frac{\sigma^2}{a}\right)$. We are interested in estimating θ . To find the posterior Bayes risk (or posterior risk for short) associated with estimating θ , we need first to find the posterior

distribution of θ (or posterior distribution for short) given the sample observations x' . We initially work with x' , a sample observation, to facilitate the derivation.

2.1 THE POSTERIOR DISTRIBUTION

Under purposive sampling, the sampling distribution is

$$f(x' | \theta) = \frac{1}{\sqrt{2\pi\sigma^2/a}} \exp\left\{-\frac{1}{2} \frac{(x' - \theta)^2}{\sigma^2/a}\right\}.$$

The prior distribution of θ is

$$\pi(\theta) = \frac{1}{\sqrt{2\pi\tau}} \exp\left\{-\frac{1}{2} \frac{(\theta - \mu)^2}{\tau^2}\right\}.$$

The joint probability density function of X' and θ is

$$\begin{aligned} f(x', \theta) &= \pi(\theta) f(x' | \theta) \\ &= \frac{1}{\sqrt{2\pi\tau}} \exp\left\{-\frac{1}{2} \frac{(\theta - \mu)^2}{\tau^2}\right\} \frac{1}{\sqrt{2\pi\sigma^2/a}} \exp\left\{-\frac{1}{2} \frac{(x' - \theta)^2}{\sigma^2/a}\right\} \\ &= \frac{\sqrt{a}}{2\pi\sigma\tau} \exp\left\{-\frac{1}{2} \left[\frac{(\theta - \mu)^2}{\tau^2} + \frac{(x' - \theta)^2}{\sigma^2/a} \right]\right\}. \end{aligned}$$

We re-express the above formulation to make it easier to find the marginal density of X' , $m(x')$, as follows:

Working on the expression $\frac{1}{2} \left[\frac{(\theta - \mu)^2}{\tau^2} + \frac{(x' - \theta)^2}{\sigma^2/a} \right]$ (Berger, 1985) gives

$$\frac{1}{2} \left[\frac{(\theta - \mu)^2}{\tau^2} + \frac{(x' - \theta)^2}{\sigma^2/a} \right] = \frac{1}{2} \left[\frac{\theta^2 - 2\mu\theta + \mu^2}{\tau^2} + \frac{x'^2 - 2\theta x' + \theta^2}{\sigma'^2} \right], \text{ letting } \frac{\sigma^2}{a} = \sigma'^2$$

$$\begin{aligned}
&= \frac{1}{2} \left[\left(\frac{1}{\tau^2} + \frac{1}{\sigma'^2} \right) \theta^2 - 2 \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \theta + \left(\frac{\mu^2}{\tau^2} + \frac{x'^2}{\sigma'^2} \right) \right] \\
&= \frac{1}{2} \left[\rho \theta^2 - 2 \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \theta \right] + \frac{1}{2} \left(\frac{\mu^2}{\tau^2} + \frac{x'^2}{\sigma'^2} \right), \text{ letting } \frac{1}{\tau^2} + \frac{1}{\sigma'^2} = \rho \\
&= \frac{1}{2} \rho \left[\theta^2 - \frac{2}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \theta \right] + \frac{1}{2} \left(\frac{\mu^2}{\tau^2} + \frac{x'^2}{\sigma'^2} \right) \\
&= \frac{1}{2} \rho \left\{ \theta^2 - \frac{2}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \theta + \left[\frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2 - \left[\frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2 \right\} + \frac{1}{2} \left(\frac{\mu^2}{\tau^2} + \frac{x'^2}{\sigma'^2} \right),
\end{aligned}$$

completing the square

$$\begin{aligned}
&= \frac{1}{2} \rho \left\{ \theta^2 - \frac{2}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \theta + \left[\frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2 \right\} - \frac{1}{2} \rho \left[\frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2 + \frac{1}{2} \left(\frac{\mu^2}{\tau^2} + \frac{x'^2}{\sigma'^2} \right) \\
&= \frac{1}{2} \rho \left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2 - \frac{1}{2\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right)^2 + \frac{1}{2} \left(\frac{\mu^2}{\tau^2} + \frac{x'^2}{\sigma'^2} \right) \\
&= c - \frac{1}{2\rho} \left(\frac{\mu\sigma'^2 + x'\tau^2}{\tau^2\sigma'^2} \right)^2 + \frac{1}{2} \left(\frac{\mu^2\sigma'^2 + x'^2\tau^2}{\tau^2\sigma'^2} \right), \text{ letting the first term be } c \\
&= c - \frac{1}{2} \frac{\tau^2\sigma'^2}{\sigma'^2 + \tau^2} \frac{(\mu\sigma'^2 + x'\tau^2)^2}{(\tau^2\sigma'^2)^2} + \frac{1}{2} \left(\frac{\mu^2\sigma'^2 + x'^2\tau^2}{\tau^2\sigma'^2} \right), \text{ since } \rho = \frac{\sigma'^2 + \tau^2}{\tau^2\sigma'^2} \\
&= c - \frac{1}{2} \left[\frac{1}{\sigma'^2 + \tau^2} \frac{(\mu\sigma'^2 + x'\tau^2)^2}{\tau^2\sigma'^2} - \frac{(\mu^2\sigma'^2 + x'^2\tau^2)}{\tau^2\sigma'^2} \right] \\
&= c - \frac{1}{2\tau^2\sigma'^2} \left[\frac{(\mu\sigma'^2 + x'\tau^2)^2}{\sigma'^2 + \tau^2} - (\mu^2\sigma'^2 + x'^2\tau^2) \right] \\
&= c - \frac{1}{2\tau^2\sigma'^2} \left[\frac{2\mu x'\sigma'^2\tau^2 - x'^2\sigma'^2\tau^2 - \mu^2\sigma'^2\tau^2}{\sigma'^2 + \tau^2} \right] \\
&= c - \frac{1}{2} \left[\frac{-(x'^2 - 2\mu x' + \mu^2)}{\sigma'^2 + \tau^2} \right]
\end{aligned}$$

$$\begin{aligned}
&= c + \frac{1}{2} \frac{(x' - \mu)^2}{(\sigma'^2 + \tau^2)} \\
&= \frac{1}{2} \rho \left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2 + \frac{(x' - \mu)^2}{2(\sigma'^2 + \tau^2)}, \text{ by definition of } c.
\end{aligned}$$

Hence, our new expression for $f(x', \theta)$ is

$$f(x', \theta) = \frac{\sqrt{a}}{2\pi\sigma'\tau} \exp \left\{ -\frac{1}{2} \rho \left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2 \right\} \exp \left\{ -\frac{(x' - \mu)^2}{2(\sigma'^2 + \tau^2)} \right\}.$$

To find the marginal distribution of X' , we have

$$\begin{aligned}
m(x') &= \int_{-\infty}^{\infty} f(x', \theta) d\theta \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sigma'\tau} \exp \left\{ -\frac{1}{2} \rho \left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2 \right\} \exp \left\{ -\frac{(x' - \mu)^2}{2(\sigma'^2 + \tau^2)} \right\} d\theta \\
&= \frac{1}{2\pi\sigma'\tau} \exp \left\{ -\frac{(x' - \mu)^2}{2(\sigma'^2 + \tau^2)} \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \rho \left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2 \right\} d\theta \\
&= \frac{1}{2\pi\sigma'\tau} \exp \left\{ -\frac{(x' - \mu)^2}{2(\sigma'^2 + \tau^2)} \right\} \int_{-\infty}^{\infty} \exp \left\{ -\frac{1}{2} \frac{\left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2}{\frac{1}{\rho}} \right\} d\theta \\
&= \frac{1}{2\pi\sigma'\tau} \exp \left\{ -\frac{(x' - \mu)^2}{2(\sigma'^2 + \tau^2)} \right\} \sqrt{2\pi} \frac{1}{\sqrt{\rho}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi} \frac{1}{\sqrt{\rho}}} \exp \left\{ -\frac{1}{2} \frac{\left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \right]^2}{\frac{1}{\rho}} \right\} d\theta \\
&= \frac{1}{2\pi\sigma'\tau} \exp \left\{ -\frac{(x' - \mu)^2}{2(\sigma'^2 + \tau^2)} \right\} \sqrt{2\pi} \frac{1}{\sqrt{\rho}} \cdot 1, \text{ since } \theta \sim N \left(\frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right), \frac{1}{\rho} \right)
\end{aligned}$$

$$= \frac{1}{\sqrt{2\pi\rho\sigma'\tau}} \exp\left\{-\frac{1}{2} \frac{(x' - \mu)^2}{(\sigma'^2 + \tau^2)}\right\}.$$

Since $\rho = \frac{1}{\tau^2} + \frac{1}{\sigma'^2} = \frac{\sigma'^2 + \tau^2}{\tau^2\sigma'^2}$,

$$\begin{aligned} m(x') &= \frac{1}{\sqrt{2\pi \frac{(\sigma'^2 + \tau^2)}{\tau^2\sigma'^2} \cdot \sigma'^2\tau^2}} \exp\left\{-\frac{1}{2} \frac{(x' - \mu)^2}{(\sigma'^2 + \tau^2)}\right\} \\ &= \frac{1}{\sqrt{2\pi(\sigma'^2 + \tau^2)}} \exp\left\{-\frac{1}{2} \frac{(x' - \mu)^2}{(\sigma'^2 + \tau^2)}\right\}. \end{aligned}$$

Thus, $X' \sim N(\mu, \sigma'^2 + \tau^2)$ marginally.

The posterior distribution of θ given the sample x' will then be

$$\begin{aligned} \pi(\theta | x') &= \frac{f(x', \theta)}{m(x')} = \frac{\frac{1}{2\pi\sigma'\tau} \exp\left\{-\frac{1}{2} \rho \left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2}\right)\right]^2\right\} \exp\left\{-\frac{1}{2} \frac{(x' - \mu)^2}{(\sigma'^2 + \tau^2)}\right\}}{\frac{1}{\sqrt{2\pi\rho\sigma'\tau}} \exp\left\{-\frac{1}{2} \frac{(x' - \mu)^2}{(\sigma'^2 + \tau^2)}\right\}} \\ &= \frac{(2\pi)^{-1}}{(2\pi)^{-1/2} \rho^{-1/2}} \exp\left\{-\frac{1}{2} \rho \left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2}\right)\right]^2\right\} \\ &= \frac{1}{\sqrt{2\pi} \frac{1}{\sqrt{\rho}}} \exp\left\{-\frac{1}{2} \frac{\left[\theta - \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2}\right)\right]^2}{\frac{1}{\rho}}\right\}. \end{aligned}$$

Thus, the posterior distribution of θ given the sample x' is $N\left(\frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2}\right), \frac{1}{\rho}\right)$.

With $\rho = \frac{\sigma'^2 + \tau^2}{\tau^2 \sigma'^2}$, the mean and variance of this distribution can be simplified to

$$\begin{aligned} E(\theta | x') &= \frac{1}{\rho} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \\ &= \frac{\tau^2 \sigma'^2}{\sigma'^2 + \tau^2} \left(\frac{\mu}{\tau^2} + \frac{x'}{\sigma'^2} \right) \\ &= \left(\frac{\sigma^2}{\sigma^2 + a\tau^2} \right) \mu + \left(\frac{a\tau^2}{\sigma^2 + a\tau^2} \right) x' \\ V(\theta | x') &= \frac{1}{\rho} = \frac{\sigma^2 \tau^2}{\sigma^2 + a\tau^2}. \end{aligned}$$

If a sample $\underline{X}' = (X'_1, \dots, X'_n)$ from $N\left(\theta, \frac{\sigma^2}{a}\right)$ distribution is to be taken, then we need the posterior distribution of θ given $\underline{x}' = (x'_1, \dots, x'_n)$. Since \bar{X}' is sufficient for θ , it follows from Lemma 1 (Berger, 1985) that $\pi(\theta | \underline{x}') = \pi(\theta | \bar{x}')$. Noting that $\bar{X}' \sim N\left(\theta, \frac{\sigma'^2}{n}\right)$, it can be concluded from $\pi(\theta | \underline{x}')$ that the posterior distribution of θ given \underline{x}' will be normal with the following parameters:

$$\begin{aligned} E(\theta | \underline{x}') &= E(\theta | \bar{x}') = \left(\frac{\frac{\sigma^2}{an}}{\frac{\sigma^2}{an} + \tau^2} \right) \mu + \left(\frac{\tau^2}{\frac{\sigma^2}{an} + \tau^2} \right) \bar{x}' \\ V(\theta | \underline{x}') &= V(\theta | \bar{x}') = \frac{\frac{\sigma^2}{an} \cdot \tau^2}{\frac{\sigma^2}{an} + \tau^2} = \frac{\sigma^2 \tau^2}{\sigma^2 + an\tau^2} = \frac{1}{\rho'}. \end{aligned}$$

With $\rho' = \frac{\sigma^2 + an\tau^2}{\sigma^2 \tau^2}$, we can also express the posterior mean as

$$E(\theta | \underline{x}') = \frac{1}{\rho'} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x'}{\sigma'^2} \right)$$

where Σ is the summation notation, summing over all values in the sample.

2.2 THE POSTERIOR RISK

Given the target variable X , the squared error loss is $l(\theta, k) = (\theta - k)^2$, where $k = d(x)$ is the decision or action taken to estimate θ (Carlin and Louis, 1996). Under purposive sampling with target variable X' and $d(x') = \bar{x}'$, we are determining the average loss of the estimator \bar{x}' with respect to the posterior distribution of θ given the sample \underline{x}' when we find this posterior risk (Mood et al., 1974).

We now find the posterior risk under purposive sampling, the risk of estimating θ with the sample mean, denoted by $g(\pi, \bar{x}')$, as follows (Carlin and Louis, 1996):

$$\begin{aligned} g(\pi, \bar{x}') &= E_{\theta|\underline{x}'} [l(\theta, \bar{x}')] \\ &= \int_{-\infty}^{\infty} l(\theta, \bar{x}') \pi(\theta|\underline{x}') d\theta \\ &= \int_{-\infty}^{\infty} (\theta - \bar{x}')^2 \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{\rho'}} \exp \left\{ -\frac{1}{2} \frac{\left[\theta - \frac{1}{\rho'} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x'}{\sigma'^2} \right) \right]^2}{\frac{1}{\rho'}} \right\} d\theta \end{aligned}$$

To facilitate the evaluation of the posterior risk, the above function is expressed in terms of matrix parameters of the normal distribution where θ is known to be an n -dimensional normal random vector. Thus, $\underline{\theta} \sim N_n(\underline{\mu}', \underline{\Sigma})$, where $\underline{\mu}'$ is the $n \times 1$ mean vector and $\underline{\Sigma}$ is the $n \times n$ covariance matrix, $\underline{\Sigma} > \underline{0}$. Then, we can write the posterior risk as an expression involving matrices:

$$\begin{aligned}
g(\pi, \bar{x}') &= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\theta - \bar{x}')^2 \frac{1}{(2\pi)^{n/2} |\underline{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}(\theta - \underline{\mu}')' \underline{\Sigma}^{-1} (\theta - \underline{\mu}')\right] d\theta_1 \dots d\theta_n \\
&= \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} (\theta - \bar{x}')^2 \frac{1}{(2\pi)^{n/2} |\underline{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2} \underline{v}' \underline{\Sigma}^{-1} \underline{v}\right] d\theta_1 \dots d\theta_n, \text{ where} \\
&\quad \underline{v} = (\theta - \underline{\mu}') \text{ and } \underline{\Sigma} > 0.
\end{aligned}$$

Evaluating this expression at $n=1$, we have

$$\begin{aligned}
g(\pi, \bar{x}') &= \int_{-\infty}^{\infty} (\theta - \bar{x}')^2 \frac{1}{(2\pi)^{1/2} \Sigma^{1/2}} \exp\left(-\frac{1}{2} v' \Sigma^{-1} v\right) d\theta, \text{ where } v = \theta - \mu' = v', \\
&\quad \mu' = E(\theta | \bar{x}') = \frac{1}{\rho'} \left(\frac{\mu}{\tau^2} + \frac{\Sigma \bar{x}'}{\sigma'^2} \right), \text{ and } \Sigma = V(\theta | \bar{x}') = \frac{1}{\rho'}. \\
&= \frac{1}{(2\pi)^{1/2} \Sigma^{1/2}} \int_{-\infty}^{\infty} (\theta - \bar{x}')^2 \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \\
&= \frac{1}{(2\pi)^{1/2} \Sigma^{1/2}} \int_{-\infty}^{\infty} (\theta^2 - 2\bar{x}'\theta + \bar{x}'^2) \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \\
&= \frac{1}{(2\pi)^{1/2} \Sigma^{1/2}} \left\{ \int_{-\infty}^{\infty} \theta^2 \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \right. \\
&\quad \left. - 2\bar{x}' \int_{-\infty}^{\infty} \theta \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \right. \\
&\quad \left. + \bar{x}'^2 \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \right\} \\
&= \frac{1}{(2\pi)^{1/2} \Sigma^{1/2}} [T_1 - 2\bar{x}'T_2 + \bar{x}'^2T_3]
\end{aligned}$$

Evaluating each of the three integrals:

$$\begin{aligned}
T_1 &= \int_{-\infty}^{\infty} \theta^2 \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \\
&= \int_{-\infty}^{\infty} \left[(\theta - \mu')' I_1 (\theta - \mu') + 2\mu'\theta - \mu'^2 \right] \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta, \\
&\quad \text{since } (\theta - \mu')' I_1 (\theta - \mu') = (\theta - \mu')^2, (\theta - \mu') \text{ is } 1 \times 1, I_1 = [1] \\
&= \int_{-\infty}^{\infty} (\theta - \mu')' I_1 (\theta - \mu') \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \\
&\quad + 2\mu' \int_{-\infty}^{\infty} \theta \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \\
&\quad - \mu'^2 \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \\
&= (2\pi)^{1/2} |\Sigma^{-1}|^{-1/2} \text{tr}\left[I_1 (\Sigma^{-1})^{-1}\right] + 2\mu' \cdot (2\pi)^{1/2} |\Sigma^{-1}|^{-1/2} \mu' \cdot 1 \\
&\quad - \mu'^2 (2\pi)^{1/2} |\Sigma^{-1}|^{-1/2}, \text{ by Theorem 1.10.1 (Graybill, 1976) at } n = 1 \\
&= (2\pi)^{1/2} |\Sigma^{-1}|^{-1/2} \left\{ \text{tr}\left[\left(\frac{1}{\Sigma}\right)^{-1}\right] + 2\mu'^2 - \mu'^2 \right\} \\
&= (2\pi)^{1/2} |\Sigma^{-1}|^{-1/2} (\Sigma + \mu'^2) \\
T_2 &= \int_{-\infty}^{\infty} \theta \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \\
&= \int_{-\infty}^{\infty} \theta_1 d_1 \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta, \text{ where} \\
&\quad d_i = 1 \forall i, i = 1, 2, \dots, n \\
&= (2\pi)^{1/2} |\Sigma^{-1}|^{-1/2} \mu', \text{ by Theorem 1.10.1 (Graybill, 1976) at } n = 1 \\
T_3 &= \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\theta - \mu')' \Sigma^{-1} (\theta - \mu')\right] d\theta \\
&= (2\pi)^{1/2} |\Sigma^{-1}|^{-1/2}, \text{ Aitken's integral (Graybill, 1976 and Searle, 1982)} \\
&\quad \text{at } n = 1.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
g(\pi, \bar{x}') &= \frac{1}{(2\pi)^{1/2} \Sigma^{1/2}} [(2\pi)^{1/2} |\Sigma^{-1}|^{-1/2} (\Sigma + \mu'^2) - 2\bar{x}' (2\pi)^{1/2} |\Sigma^{-1}|^{-1/2} \mu' \\
&\quad + \bar{x}'^2 (2\pi)^{1/2} |\Sigma^{-1}|^{-1/2}] \\
&= \frac{1}{(2\pi)^{1/2} \Sigma^{1/2}} \cdot (2\pi)^{1/2} |\Sigma^{-1}|^{-1/2} (\Sigma + \mu'^2 - 2\mu'\bar{x}' + \bar{x}'^2) \\
&= \Sigma + \mu'^2 - 2\mu'\bar{x}' + \bar{x}'^2 \\
&= \frac{1}{\rho'} + \left[\frac{1}{\rho'} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x'}{\sigma'^2} \right) \right]^2 - \frac{2\bar{x}'}{\rho'} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x'}{\sigma'^2} \right) + \bar{x}'^2 \\
&= \frac{\tau^2 \sigma'^2}{\sigma'^2 + n\tau^2} + \left[\frac{\tau^2 \sigma'^2}{\sigma'^2 + n\tau^2} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x'}{\sigma'^2} \right) \right]^2 - \frac{2\bar{x}' \tau^2 \sigma'^2}{\sigma'^2 + n\tau^2} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x'}{\sigma'^2} \right) + \bar{x}'^2 \\
&= \frac{\tau^2 \frac{\sigma^2}{a}}{\frac{\sigma^2}{a} + n\tau^2} + \left[\frac{\tau^2 \frac{\sigma^2}{a}}{\frac{\sigma^2}{a} + n\tau^2} \left(\frac{\mu}{\tau^2} + \frac{a\Sigma x'}{\sigma^2} \right) \right]^2 - \frac{2\bar{x}' \tau^2 \frac{\sigma^2}{a}}{\frac{\sigma^2}{a} + n\tau^2} \left(\frac{\mu}{\tau^2} + \frac{a\Sigma x'}{\sigma^2} \right) + \bar{x}'^2 \\
&= \frac{\tau^2 \sigma^2}{\sigma^2 + an\tau^2} + \left(\frac{\mu\sigma^2 + a\tau^2 \Sigma x'}{\sigma^2 + an\tau^2} \right)^2 - \frac{2\bar{x}' (\mu\sigma^2 + a\tau^2 \Sigma x')}{\sigma^2 + an\tau^2} + \bar{x}'^2 \quad (2.1)
\end{aligned}$$

We note that the resulting expression does not depend on the observed sample values. Since the posterior risk satisfies the sufficient condition for optimality by Zacks (1969), the purposive sampling described in this study is an optimal Bayesian sampling plan.

Under simple random sampling, the posterior risk using squared error loss is parallel to that under purposive sampling since we have $X \sim N(\theta, \sigma^2)$ under simple random sampling while we have $X' \sim N\left(\theta, \frac{\sigma^2}{a}\right)$, $a > 1$ under purposive sampling, with the prior distribution of θ the same for both sampling designs, $\theta \sim N(\mu, \tau^2)$. Thus, the posterior distribution of θ given the sample x will be normal with the following parameters:

$$E(\theta | \underline{x}) = \left(\frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} + \tau^2} \right) \mu + \left(\frac{\tau^2}{\frac{\sigma^2}{n} + \tau^2} \right) \bar{x}$$

$$V(\theta | \underline{x}) = \frac{\frac{\sigma^2}{n} \cdot \tau^2}{\frac{\sigma^2}{n} + \tau^2} = \frac{\sigma^2 \tau^2}{\sigma^2 + n\tau^2} = \frac{1}{\rho''}.$$

With $\rho'' = \frac{\sigma^2 + n\tau^2}{\sigma^2 \tau^2}$, we can also express the posterior mean as follows:

$$E(\theta | \underline{x}) = \frac{1}{\rho''} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x}{\sigma^2} \right) = \mu''.$$

The posterior risk using squared error loss under simple random sampling is then determined similarly:

$$g(\pi, \bar{x}) = \int_{-\infty}^{\infty} l(\theta, \bar{x}) \pi(\theta | \underline{x}) d\theta$$

$$= \int_{-\infty}^{\infty} (\theta - \bar{x})^2 \frac{1}{\sqrt{2\pi} \frac{1}{\sqrt{\rho''}}} \exp \left\{ -\frac{1}{2} \frac{\left[\theta - \frac{1}{\rho''} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x}{\sigma^2} \right) \right]^2}{\frac{1}{\rho''}} \right\} d\theta$$

$$= \frac{1}{\rho''} + \left[\frac{1}{\rho''} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x}{\sigma^2} \right) \right]^2 - \frac{2\bar{x}}{\rho''} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x}{\sigma^2} \right) + \bar{x}^2$$

$$= \frac{\tau^2 \sigma^2}{\sigma^2 + n\tau^2} + \left[\frac{\tau^2 \sigma^2}{\sigma^2 + n\tau^2} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x}{\sigma^2} \right) \right]^2 - \frac{2\bar{x} \tau^2 \sigma^2}{\sigma^2 + n\tau^2} \left(\frac{\mu}{\tau^2} + \frac{\Sigma x}{\sigma^2} \right) + \bar{x}^2$$

$$= \frac{\tau^2 \sigma^2}{\sigma^2 + n\tau^2} + \left(\frac{\mu \sigma^2 + \tau^2 \Sigma x}{\sigma^2 + n\tau^2} \right)^2 - \frac{2\bar{x} (\mu \sigma^2 + \tau^2 \Sigma x)}{\sigma^2 + n\tau^2} + \bar{x}^2. \quad (2.2)$$

3. SIMULATION RESULTS

The optimality of purposive sampling is further assessed by comparing

its posterior risk with that of a design-unbiased sampling plan (simple random sampling). This is done as the effects on the posterior risk of its parameters are examined.

3.1 EFFECT OF a ON THE POSTERIOR RISK

Suppose we let X be $N(\theta, 25)$ and θ be $N(10, 15)$. That is, $\sigma^2 = 25$, $\mu = 10$, and $\tau^2 = 15$. With X' distributed as $N\left(\theta, \frac{25}{a}\right)$, we compute the posterior risk of the sample mean as the estimator of θ using the squared error loss function as we let a vary from 1 to 250. The value of a reflects the confidence of the sampler on the probable location of the true mean where sample selection may be focused on. Small a means selection from a wider range of values while large a implies selection in a shorter range of values. We take a sample of size 1 with observed values of 8 representing a number below the prior mean of θ , 10 exactly the same as the prior mean of θ , and 12 a number above the prior mean of θ .

Note that at $a = 1$, when $X \sim N(\theta, 25)$ is the sampled population, a simple random sample is drawn. At $a = 2$, the sample is drawn from $X' \sim N(\theta, 12.5)$, a purposive selection since the sampling population is constrained to be about "half" the size of the initial population. This effectively ignores the tails of the distribution since the sampler thinks that it will be less likely that the true mean is located in the tails. At $a > 2$, we are taking purposive samples from increasingly more homogeneous populations. TABLE I gives the behavior of the posterior risk under these situations.

TABLE I: Effect of a on the posterior risk using squared error loss function when $\sigma^2 = 25, \mu = 10$, and $\tau^2 = 15$ at $n = 1$.

Value of a	Posterior Risk (a)		
	$x = 8$	$x = 10$	$x = 12$
1	10.94	9.38	10.94
2	7.64	6.82	7.64
5	4.00	3.75	4.00
10	2.22	2.14	2.22
25	0.95	0.94	0.95
50	0.49	0.48	0.49
100	0.25	0.25	0.25
125	0.20	0.20	0.20
250	0.10	0.10	0.10

(a) Using equation (2.1).

All observed values yield a decreasing average risk or loss of the sample mean as the population becomes more homogeneous or compact. By the nature of the loss function, observed values of the same distance from the prior mean of θ have the same risk. Also, the risk level was lowest when the observed value was equal to the prior mean of θ , except at $a \geq 100$ when all risk levels became equal.

The average risk of the sample mean decreased substantially from $a = 1$ to $a = 2$. This implies that the average loss of the sample mean in estimating θ with respect to the posterior distribution of θ given the sample is reduced under purposive sampling when the sampled population is more homogeneous. The sampler will be in a less risky situation when samples are drawn as closely as possible to the prior mean.

3.2 EFFECT OF σ^2 ON THE POSTERIOR RISK

Suppose we allow the variance of the variable of interest, σ^2 , to vary from 1 to 250.

We again compute the posterior risk with the sample mean as the estimator of θ using the squared error loss function. TABLE II shows the results at selected values of σ^2 .

TABLE II: Effect of σ^2 on the posterior risk using squared error loss function when $\mu = 10$ and $\tau^2 = 15$ at $n = 1$.

Value of σ^2	Posterior Risk (a)					
	$a = 1$			$a = 2$		
	$x = 8$	$x = 10$	$x = 12$	$x = 8$	$x = 10$	$x = 12$
1	0.95	0.94	0.95	0.49	0.48	0.49
2	1.82	1.76	1.82	0.95	0.94	0.95
5	4.00	3.75	4.00	2.22	2.14	2.22
10	6.64	6.00	6.64	4.00	3.75	4.00
25	10.94	9.38	10.94	7.64	6.82	7.64
50	13.91	11.54	13.91	10.94	9.38	10.94
100	16.07	13.04	16.07	13.91	11.54	13.91
125	16.58	13.39	16.58	14.70	12.10	14.70
250	17.71	14.15	17.71	16.58	13.39	16.58

(a) Using equation (2.1).

The average loss of the sample mean decreased substantially as the population variance was reduced from 25 down to 1 at all observed values under the two sampling designs. When the population variance was increased from 25 up to ten times its initial level, the average loss of the sample mean increased substantially for both sampling designs. When the observed value was equal to the prior mean of θ , the average losses were the lowest, similar to the effect of a on the posterior risk.

3.3 EFFECT OF τ^2 ON THE POSTERIOR RISK

Suppose we allow the prior variance of θ to vary from 1 to 150 with the prior mean of θ fixed at 10. We compute the posterior risk using the mean as the estimator of θ under the squared error loss function. TABLE III gives the results at selected values of τ^2 .

TABLE III: Effect of τ^2 on the posterior risk using squared error loss function when $\mu = 10$ and $\sigma^2 = 25$ at $n = 1$.

Value of τ^2	Posterior Risk (a)					
	$a = 1$			$a = 2$		
	$x = 8$	$x = 10$	$x = 12$	$x = 8$	$x = 10$	$x = 12$
1	4.66	0.96	4.66	4.36	0.93	4.36
2	5.28	1.85	5.28	4.70	1.72	4.70
5	6.94	4.17	6.94	5.61	3.57	5.61
10	9.18	7.14	9.18	6.79	5.56	6.79
15	10.94	9.38	10.94	7.64	6.82	7.64
25	13.50	12.50	13.50	8.78	8.33	8.78
50	17.11	16.67	17.11	10.16	10.00	10.16
75	19.00	18.75	19.00	10.80	10.71	10.80
100	20.16	20.00	20.16	11.16	11.11	11.16
150	21.51	21.43	21.51	11.56	11.54	11.56

(a) Using equation (2.1).

The effect of τ^2 on the average loss of the sample mean is seen to be similar to that of σ^2 . As the prior variance of θ increased so did the average loss of the sample mean for the three observed values under both sampling designs.

3.4 EFFECT OF SAMPLE SIZE ON THE POSTERIOR RISK

Suppose we let the sample size vary when the sampling distribution is $N(\theta, 25)$ and θ is $N(10, 15)$. Then, we set $\theta = \mu$ for the mean of the sampling distribution. We generate simple random samples from $N(10, 25)$ when $a = 1$. We generate random samples from $N(10, 12.5)$ when $a = 2$ for our purposive samples. TABLE IV gives the results using the squared error loss function.

TABLE IV: Effect of sample size (a) on the posterior risk using squared error loss function when $\sigma^2 = 25$, $\mu = 10$, and $\tau^2 = 15$.

Sample Size	Posterior Risk (b)					
	$a = 1$			$a = 2$		
	Min	Max	Mean	Min	Max	Mean
5	3.7651	4.4123	3.9815	2.1430	2.2110	2.1860
10	2.1430	2.1608	2.1493	1.1541	1.1593	1.1565
15	1.5008	1.5169	1.5068	0.7897	0.7960	0.7920
30	0.7904	0.7916	0.7909	0.4054	0.4056	0.4055
50	0.4839	0.4845	0.4842	0.2459	0.2461	0.2460
(a) 75	0.3261	0.3264	0.3263	0.1648	0.1648	0.1648
100	0.2459	0.2459	0.2459	0.1240	0.1240	0.1240
150	0.1648	0.1648	0.1648	0.0829	0.0829	0.0829
200	0.1240	0.1240	0.1240	0.0622	0.0622	0.0622

Based on three samples.
 (b) Using equation (2.1).

We find that the average loss of the sample mean as an estimator of θ decreased and approached zero as the sample size became very large. The range of the expected loss of the sample mean as the estimator of θ is seen to decrease as the sample size increases. These observations hold for both simple random sampling and purposive sampling. Also, the risk levels under purposive sampling are consistently lower than those under simple random sampling as the sample size increases.

4. CONCLUSIONS

Purposive sampling as described in this study is an optimal Bayesian sampling plan. The posterior risk is shown to be independent of the observed sample values, a sufficient condition for optimality of any nonrandomized single-phase sampling plan without

replacement. Simulation of the posterior risk indicates that the performance of the sample mean as an estimator of the population mean under purposive sampling improves as the population becomes more compact using the squared error loss function. The average losses of the sample mean are lower under purposive sampling regardless of the prior belief on the variability of the population mean. The risk level under purposive sampling is reduced as the sample size is increased.

BIBLIOGRAPHY

- Berger, J.O. (1985). *Statistical Decision Theory and Bayesian Analysis*. 2nd Ed. New York: Springer-Verlag, Inc., 126-128.
- Carlin, B.P. and Louis, T.A. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. London: Chapman & Hall, 4, 7-9.
- Graybill, F.A. (1976). *Theory and Application of the Linear Model*. North Scituate, Mass.: Duxbury Press, 48-50.
- Mood, A.M., Graybill, F.A., and Boes, D.C. (1974). *Introduction to the Theory of Statistics*. 3rd Ed. Quezon City: JMC Press, Inc., 346.
- Searle, S.R. (1982). *Matrix Algebra Useful for Statistics*. New York: John Wiley & Sons, Inc., 340.
- Zacks, S. (1969). Bayesian sequential designs for sampling finite populations. *J. Am. Statist. Assoc.* **64**, 1342-1349.