Purposive Sampling as an Optimal Bayes Sampling Design

Jacqueline M. Guarte
Department of Mathematics, Physics, and Statistics,
Visayas State University

Purposive sampling takes place when the researcher's knowledge about the population is used to handpick the units to be included in the sample. This is hinged on the experienced researcher's belief that the handpicked sampling units will provide "enough" information to characterize the population.Bayesian analysis makes explicit use of prior information as part of the model to satisfy some optimality criteria. Hence, purposive rather than purely random locations of design points need to be chosen. This paper presents a proof that purposive sampling is an optimal Bayes sampling design. Purposive sampling satisfies the sufficient condition for an optimal Bayes sampling design set by Zacks (1969) for single-phase designs. It is shown that the posterior Bayes risk of the population parameter θ given the sample observations is independent of the observed values under purposive sampling. The parameter of interest is the population mean. The normal distribution is used for the sampling distribution and the prior distribution of the population mean due to its universal significance and mathematical maneuverability. The squared error loss function is used in determining the posterior Bayes risk associated with estimating the population mean, with the sample mean as estimator.The posterior Bayes risk under simple random sampling is also determined for comparison purposes. It is shown that the risk levels under purposive sampling are lower than those under simple random sampling when important model parameters are made to vary.

Keywords: purposive sampling,optimal Bayes sampling design, posterior Bayes risk

1. Introduction

Purposive sampling is a nonprobability sampling technique that involves selecting sampling units without replacement from the segment of the population with the most information on the characteristic of interest. Ghosh and Meeden (1997), in discussing finite population sampling, state that a Bayesian approach typically leads to purposeful designs. Also, the model-dependent approach (as opposed to the design-based approach) to carrying out statistical analysis in a sample survey problem involves purposive sampling (Shao and Tu, 1995). This is but natural since modeling would necessitate the design matrix to satisfy some optimality criteria. 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.

Both Basu and Godambe as cited by Zacks (1969) stated that the sampling plan is immaterial for a Bayesian analysis, and hence a Bayesian may as well choose the sample in a nonrandomized manner. Basu (1969) examined the role of the sufficiency principle and the likelihood principle in the analysis of survey data and arrived at the conclusion that, once the sample has been drawn, the inference should not depend in any way on the sampling design (i.e., model-based approach). The sufficiency principle states that if $T(X)$ is a sufficient statistic for θ then any inference about θ should depend on the sample X only through the value $T(X)$ (Casella and Berger, 1990). The likelihood principle states that once the data value x has been observed, the likelihood function $L(\theta|x)$ contains all relevant experimental information about the unknown parameter θ (Carlin and Louis, 1996). This poses the problem of designing a survey that will yield a good (representative) sample. Basu then examined the randomization principle from this viewpoint and noticed that there is very little, if any, use for it in survey designs. The randomization principle says that random sampling is the indispensable condition of finite population inference (Valliant et al., 2000). Godambe (1966) showed that the likelihood function in sample surveys is independent of the design. He emphasized that the likelihood principle implies that inference should be independent of the sampling design, in general.

Zacks (1969) explains that Basu and Godambe arrived at their conclusion by showing that the posterior distribution of the parameter subject to statistical estimation is independent of the particular manner in which the units were chosen. Zacks, however, finds this in itself inadequate. She agrees that the posterior distribution of the parameter being estimated, given the sample data, is independent of the sampling rule. However, she pointed out that different sets of data might lead to Bayesian inference of varying "precision." Citing that the posterior risk is a function of the data, she directs the sampling designer to adopt the principle of selecting a sampling rule that minimizes the expected Bayes risk. Zacks proves that an optimal Bayes sampling design is without replacement for any loss function that is nonnegative and bounded. She also proves that any nonrandomized single-phase sampling plan without replacement is optimal if the posterior risk is independent of the observed values.

2. The Bayesian Framework

In the Bayesian framework, a finite population of size N is a set $U=\{u_1,u_2,\ldots u_N\}$ of specified units. Each unit of U has an unknown value $x(u_i)=x_i(i=1,\ldots,\mathrm{N})$. A sample of a fixed size n is an ordered set $s_n=\langle u_{i1},u_{i2},\ldots,u_{in}\rangle$ of elements of U. The statistic representing the observed sample is denoted by $T_n=(s_n,(y_1,\ldots,y_n))$ where s_n is the ordered set of labels, and $y_i=x(u_{ij})$ for all $j=1,\ldots,n$. The posterior distribution of the unknown parameter θ, given $T_n=(s_n,(y_1,\ldots,y_n))$ is designated by H $(\theta|s_n,(y_1,\ldots y_n))$. For this study, we use the notation $\pi(\theta\,|\,\underset{\sim}{x})$ to denote the posterior distribution of θ given the sample observations $\underset{\sim}{x}$ to conform to the presently used notations.

In identifying an optimal Bayes sampling plan, Zacks (1969) proved the following theorem:

> *If the posterior Bayes risk $\rho_0\,(s_n,(y_1,\ldots,y_n))$ is independent of the observed sample values $(y_1,\ldots,y_n)$ then the optimal sampling plan is a nonrandomized single-phase one. If, in addition, $x_1,\ldots,x_n$ are priorly exchangeable random variables, then any single-phase plan is optimal.*

Zacks (1969) pointed out, however, that the condition $\rho_0\,(T_n)$ independent of $(y_1,\ldots,y_n)$ is sufficient but not necessary for the optimal Bayes sampling design to be a single-phase one. There are examples of optimal Bayes sampling designs that are single-phase but $\rho_0\,(T_n)$ depends on $(y_1,\ldots,y_n)$. We then verify whether purposive sampling meets the sufficient condition for an optimal Bayes sampling design using Zacks's theorem.

3. On 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 $\underset{\sim}{x}'$. We initially work with x', a sample observation, to facilitate the derivation.

3.1 The posterior distribution under purposive sampling

Under purposive sampling, the sampling distribution is

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

The prior distribution of θ is

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

The joint probability density function of X' and θ is

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

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

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

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

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

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

grouping with respect to θ

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

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

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

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

completing the square

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

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

$$=c-\frac{1}{2\rho}\left(\frac{\mu\sigma'^2+x'\tau^2}{\tau^2\sigma'^2}\right)^2+\frac{1}{2}\left(\frac{\mu^2\sigma'^2+x'^2\tau^2}{\tau^2\sigma'^2}\right),\ \text{letting the first term be } c \text{ and simplifying}$$

$$=c-\frac{1}{2}\frac{\tau^2\sigma'^2}{\sigma'^2+\tau^2}\frac{\left(\mu\sigma'^2+x'\tau^2\right)^2}{\left(\tau^2\sigma'^2\right)^2}+\frac{1}{2}\left(\frac{\mu^2\sigma'^2+x'^2\tau^2}{\tau^2\sigma'^2}\right),\ \text{since}\ \rho=\frac{\sigma'^2+\tau^2}{\tau^2\sigma'^2}$$

$$=c-\frac{1}{2}\left[\frac{1}{\sigma'^2+\tau^2}\frac{\left(\mu\sigma'^2+x'\tau^2\right)^2}{\tau^2\sigma'^2}-\frac{\left(\mu^2\sigma'^2+x'^2\tau^2\right)}{\tau^2\sigma'^2}\right],\ \text{factoring out}\ \frac{1}{2}$$

$$=c-\frac{1}{2\tau^2\sigma'^2}\left[\frac{\left(\mu\sigma'^2+x'\tau^2\right)^2}{\sigma'^2+\tau^2}-\left(\mu^2\sigma'^2+x'^2\tau^2\right)\right],\ \text{factoring out}\ \frac{1}{\tau^2\sigma'^2}$$

$$=c-\frac{1}{2\tau^2\sigma'^2}\left[\frac{\mu^2\sigma'^4+2\mu x'\sigma'^2\tau^2+x'^2\tau^4-\left(\sigma'^2+\tau^2\right)\left(\mu^2\sigma'^2+x'^2\tau^2\right)}{\sigma'^2+\tau^2}\right]$$

$$=c-\frac{1}{2\tau^2\sigma'^2}\left[\frac{2\mu x'\sigma'^2\tau^2-x'^2\sigma'^2\tau^2-\mu^2\sigma'^2\tau^2}{\sigma'^2+\tau^2}\right]$$

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

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

$$= c \ + \ \frac{1}{2}\frac{\left(x' - \mu\right)^2}{\left(\sigma'^2 + \tau^2\right)}$$

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

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

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

To find the marginal distribution of X', we have

$$m\left(x'\right) = \int_{-\infty}^{\infty} f\left(x',\theta\right)\, d\theta$$

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

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

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

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

$$= \frac{1}{2\pi\sigma'\tau}\exp\left\{-\frac{1}{2}\frac{\left(x'-\mu\right)^2}{\left(\sigma'^2+\tau^2\right)}\right\}\sqrt{2\pi}\,\frac{1}{\sqrt{\rho}}\cdot 1, \text{ since } \theta \sim N\left(\frac{1}{\rho}\left(\frac{\mu}{\tau^2}+\frac{x'}{\sigma'^2}\right),\frac{1}{\rho}\right)$$

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

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

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

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

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

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

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

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

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

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

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

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

If a sample $\underset{\sim}{X}'=(X_1',\ldots,X_n')$ from $N\left(\theta,\dfrac{\sigma^2}{a}\right)$ distribution is to be taken, then we need the posterior distribution of θ given $\underset{\sim}{x}'=(x_1',\ldots,x_n')$. Since $\overline{X}'$ is sufficient for θ, it follows from Lemma 1 (Berger, 1985) that $\pi(\theta\,|\,\underset{\sim}{x}')=\pi(\theta\,|\,\overline{x}')$. Noting that $\overline{X'}\sim N\left(\theta,\dfrac{\sigma'^2}{n}\right)$, it can be concluded from $\pi(\theta\,|\,x')$ that the posterior distribution of θ given $\underset{\sim}{x}'$ will be normal with the following parameters:

$$E(\theta\,|\,\underset{\sim}{x}')=E(\theta\,|\,\overline{x}')=\left(\frac{\frac{\sigma^2}{an}}{\frac{\sigma^2}{an}+\tau^2}\right)\mu+\left(\frac{\tau^2}{\frac{\sigma^2}{an}+\tau^2}\right)\overline{x}'$$

$$V(\theta\,|\,\underset{\sim}{x}')=V(\theta\,|\,\overline{x}')=\frac{\frac{\sigma^2}{an}\cdot\tau^2}{\frac{\sigma^2}{an}+\tau^2}=\frac{\sigma^2\tau^2}{\sigma^2+an\tau^2}=\frac{1}{\rho'}.$$

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

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

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

3.2 The posterior risk under purposive sampling

Given the target variable *X*, the squared error loss is $l\,(\theta,k) = (\theta\text{-}k)^2$ 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 $d\left(x'\right) = \overline{x}'$, we are actually determining the average loss of the estimator $\overline{x}'$ with respect to the posterior distribution of θ given the sample $\underset{\sim}{x}'$ 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 $g\left(\pi, \overline{x}'\right)$, as follows (Carlin and Louis, 1996):

$$g\left(\pi, \overline{x}'\right) = E_{\theta/\underset{\sim}{x}'}\left[l\left(\theta, \overline{x}'\right)\right]$$

$$= \int_{-\infty}^{\infty} l\left(\theta, \overline{x}'\right)\ \pi\left(\theta \mid \underset{\sim}{x}'\right)\ d\theta$$

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

To facilitate the evaluation of the posterior risk, the above function will be expressed in terms of matrix parameters of the normal distribution where θ is known to be an *n*-dimensional normal random vector. Thus, $\underset{\sim}{\theta} \sim N_n\left(\underset{\sim}{\mu}', \underset{\sim}{\Sigma}\right)$, where μ' is the $n\mathrm{x}1$ mean vector and $\underset{\sim}{\Sigma}$ is the $n\mathrm{x}n$ covariance matrix, $\underset{\sim}{\Sigma} > \underset{\sim}{0}$. Specifically,

$$E\left(\underset{\sim}{\theta}\right) = \underset{\sim}{\mu}' = \begin{bmatrix} \frac{1}{\rho'}\left(\frac{\mu}{\tau^2} + \frac{\Sigma x'}{\sigma'^2}\right) \\ \vdots \\ \frac{1}{\rho'}\left(\frac{\mu}{\tau^2} + \frac{\Sigma x'}{\sigma'^2}\right) \end{bmatrix} = \begin{bmatrix} E\left(\theta_1\right) \\ \vdots \\ E\left(\theta_n\right) \end{bmatrix}$$

$$Cov(\underset{\sim}{\theta}) = \underset{\sim}{\Sigma} = \begin{bmatrix} V(\theta_1) & Cov(\theta_1,\theta_2) & \ldots Cov(\theta_1,\theta_n) \\ Cov(\theta_2,\theta_1) & V(\theta_2) & \ldots Cov(\theta_2,\theta_n) \\ \vdots & \vdots & \vdots \\ Cov(\theta_n,\theta_1) & Cov(\theta_n,\theta_2) & \ldots V(\theta_n) \end{bmatrix}.$$

Then, we can write the posterior risk as an expression involving matrices:

$$g(\pi,\overline{x}') = \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}(\theta-\overline{x}')^2 \frac{1}{(2\pi)^{n/2}\,|\underset{\sim}{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}\left(\underset{\sim}{\theta}-\underset{\sim}{\mu}'\right)' \underset{\sim}{\Sigma}^{-1}\left(\underset{\sim}{\theta}-\underset{\sim}{\mu}'\right)\right] d\theta_1\ldots d\theta_n$$

$$= \int_{-\infty}^{\infty}\ldots\int_{-\infty}^{\infty}(\theta-\overline{x}')^2 \frac{1}{(2\pi)^{n/2}\,|\underset{\sim}{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2}\underset{\sim}{v}'\underset{\sim}{\Sigma}^{-1}\underset{\sim}{v}\right] d\theta_1\ldots d\theta_n,$$

where $\underset{\sim}{v} = \left(\underset{\sim}{\theta}-\underset{\sim}{\mu}'\right)$ and $\underset{\sim}{\Sigma} > \underset{\sim}{0}$.

Evaluating this expression at n=1, we have

$$g(\pi,\overline{x}') = \int_{-\infty}^{\infty}(\theta-\overline{x})^2 \frac{1}{(2\pi)^{1/2}\,\Sigma^{1/2}} \exp\left(-\frac{1}{2}v'\Sigma^{-1}v\right) d\theta,$$

where $v = \theta-\mu' = v'$, $\mu' = E(\theta\,|\,\underset{\sim}{x}') = \frac{1}{\rho'}\left(\frac{\mu}{\tau^2}+\frac{\Sigma x'}{\sigma'^2}\right)$,
and $\Sigma = V(\theta\,|\,\underset{\sim}{x}') = \frac{1}{\rho'}$.

$$= \frac{1}{(2\pi)^{1/2}\,\Sigma^{1/2}} \int_{-\infty}^{\infty}(\theta-\overline{x}')^2 \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta$$

$$= \frac{1}{(2\pi)^{1/2}\,\Sigma^{1/2}} \int_{-\infty}^{\infty}\left(\theta^2-2\overline{x}'\theta+\overline{x}'^2\right) \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta$$

$$= \frac{1}{(2\pi)^{1/2}\,\Sigma^{1/2}} \{\int_{-\infty}^{\infty}\theta^2 \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta$$

$$-2\overline{x}'\int_{-\infty}^{\infty}\theta \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta$$

$$+\,\overline{x}'^2\int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta\}$$

$$= \frac{1}{(2\pi)^{1/2}\,\Sigma^{1/2}}\left[T_1 - 2\overline{x}'T_2 + \overline{x}'^2 T_3\right].$$

Evaluating each of the three integrals:

$$T_1 = \int_{-\infty}^{\infty} \theta^2 \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta$$

$$= \int_{-\infty}^{\infty} \left[(\theta-\mu')'\,I_1(\theta-\mu') + 2\mu'\theta - \mu'^2\right] \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta,$$

since $(\theta-\mu')'\,I_1(\theta-\mu') = (\theta-\mu')^2, (\theta-\mu')$ is 1 x 1, $I_1 = [1]$

$$= \int_{-\infty}^{\infty} (\theta-\mu')'\,I_1(\theta-\mu') \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta$$

$$+\ 2\mu' \int_{-\infty}^{\infty} \theta \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta$$

$$-\ \mu'^2 \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta$$

$$= (2\pi)^{1/2}\,|\Sigma^{-1}|^{-1/2}\, tr\left[I_1\left(\Sigma^{-1}\right)^{-1}\right] \ +\ 2\mu' \cdot (2\pi)^{1/2}\,|\Sigma^{-1}|^{-1/2}\,\mu' \cdot 1$$

$-\ \mu'^2 (2\pi)^{1/2}\,|\Sigma^{-1}|^{-1/2}$, by Theorem 1.10.1 (Graybill, 1976) at n=1

$$= (2\pi)^{1/2}\,|\Sigma^{-1}|^{-1/2} \left\{ tr\left[\left(\frac{1}{\Sigma}\right)^{-1}\right] \ +\ 2\mu'^2 - \mu'^2 \right\},$$ Σ is a constant

$$= (2\pi)^{1/2}\,|\Sigma^{-1}|^{-1/2} \left[tr(\Sigma) \ +\ \mu'^2\right]$$

$= (2\pi)^{1/2}\,|\Sigma^{-1}|^{-1/2} \left(\Sigma \ +\ \mu'^2\right)$, trace of a constant is the constant

$$T_2 = \int_{-\infty}^{\infty} \theta \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta$$

$= \int_{-\infty}^{\infty} \theta_1 d_1 \exp\left[-\frac{1}{2}(\theta-\mu')'\,\Sigma^{-1}(\theta-\mu')\right] d\theta$, where $d_i = 1\ \forall\ i, i = 1, 2, \ldots, n$

$= (2\pi)^{1/2}\,|\Sigma^{-1}|^{-1/2}\,\mu'$, by Theorem 1.10.1 (Graybill, 1976) at n = 1

$$T_3 = \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}\left(\theta-\mu'\right)' \Sigma^{-1}\left(\theta-\mu'\right)\right] d\theta$$

$$= \left(2\pi\right)^{1/2} | \Sigma^{-1} |^{-1/2}, \text{ Aitken's integral (Graybill, 1976 and Searle, 1982) at } n = 1.$$

Thus, we have

$$g\left(\pi, \overline{x}'\right) = \frac{1}{\left(2\pi\right)^{1/2} \Sigma^{1/2}} [\left(2\pi\right)^{1/2} | \Sigma^{-1} |^{-1/2} \left(\Sigma + \mu'^2\right) - 2\overline{x}'\left(2\pi\right)^{1/2} | \Sigma^{-1} |^{-1/2} \mu'$$

$$+ \overline{x}'^2 \left(2\pi\right)^{1/2} | \Sigma^{-1} |^{-1/2}]$$

$$= \frac{1}{\left(2\pi\right)^{1/2} \Sigma^{1/2}} \cdot \left(2\pi\right)^{1/2} | \Sigma^{-1} |^{-1/2} \left(\Sigma + \mu'^2 - 2\mu'\overline{x}' + \overline{x}'^2\right)$$

$$= \frac{1}{\Sigma^{1/2}} \cdot \left(\frac{1}{\Sigma}\right)^{-1/2} \left(\Sigma + \mu'^2 - 2\mu'\overline{x}' + \overline{x}'^2\right)$$

$$= \frac{1}{\Sigma^{1/2}} \cdot \frac{1}{\frac{1}{\Sigma^{1/2}}} \left(\Sigma + \mu'^2 - 2\mu'\overline{x}' + \overline{x}'^2\right)$$

$$= \Sigma + \mu'^2 - 2\mu'\overline{x}' + \overline{x}'^2$$

$$= \frac{1}{\rho'} + \left[\frac{1}{\rho'}\left(\frac{\mu}{\tau^2} + \frac{\Sigma x'}{\sigma'^2}\right)\right]^2 - \frac{2\overline{x}'}{\rho'}\left(\frac{\mu}{\tau^2} + \frac{\Sigma x'}{\sigma'^2}\right) + \overline{x}'^2, \text{ by definition of } \Sigma \text{ and } \mu'$$

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

ρ' defined

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

σ'^2 defined

$$= \frac{\tau^2\sigma^2}{\sigma^2 + an\tau^2} + \left(\frac{\mu\sigma^2 + a\tau^2\Sigma x'}{\sigma^2 + an\tau^2}\right)^2 - \frac{2\overline{x}'\left(\mu\sigma^2 + a\tau^2\Sigma x'\right)}{\sigma^2 + an\tau^2} + \overline{x}'^2. \quad (3.1)$$

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 chapter is an optimal Bayesian sampling strategy.

4. Comparison of the Posterior Risk under Purposive Sampling and Simple Random Sampling

The posterior risk under simple random sampling is computed to allow further assessment of the optimality of purposive sampling. Comparisons are then made using the squared error loss function.

4.1 The posterior risk under simple random sampling

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 (μ, τ^2). Thus, the posterior distribution of θ given the sample $\underset{\sim}{x}$ will be normal with the following parameters:

$$E(\theta \mid \underset{\sim}{x}) = \left(\frac{\frac{\sigma^2}{n}}{\frac{\sigma^2}{n} + \tau^2} \right) \mu + \left(\frac{\tau^2}{\frac{\sigma^2}{n} + \tau^2} \right) \overline{x}$$

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

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

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

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

$$g(\pi, \overline{x}) = \int_{-\infty}^{\infty} l(\theta, \overline{x}) \; \pi(\theta \mid \underset{\sim}{x}) \; d\theta$$

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

$$=\frac{1}{\rho''}+\left[\frac{1}{\rho''}\left(\frac{\mu}{\tau^2}+\frac{\Sigma x}{\sigma^2}\right)\right]^2-\frac{2\bar{x}}{\rho''}\left(\frac{\mu}{\tau^2}+\frac{\Sigma x}{\sigma^2}\right)+\bar{x}^2, \text{ parallel to that in purposive sampling}$$

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

$$=\frac{\tau^2\sigma^2}{\sigma^2+n\tau^2}+\left(\frac{\mu\sigma^2+\tau^2\Sigma x}{\sigma^2+n\tau^2}\right)^2-\frac{2\bar{x}\left(\mu\sigma^2+\tau^2\Sigma x\right)}{\sigma^2+n\tau^2}+\bar{x}^2. \qquad (4.1)$$

Given the complexity of the expressions for the posterior risk of the sample mean as the estimator of θ using the squared error loss function, analytical comparisons are not feasible. Simulation of the posterior risk was done instead to analyze its behavior and the implications on the optimality of purposive sampling.

4.2 Effect of a on the posterior risk

Suppose we let X be $N(\theta, 25)$ and θ be $N(10, 15)$. That is, σ^2=25 and μ=10 and τ^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~$N(\theta$, 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 1 gives the behavior of the posterior risk under these situations.

Table 1 **Effect of *a* on the Posterior Risk using Squared Error Loss Function when σ^2=25, μ=10 and τ^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 (3.1).

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. Purposive selection is done since the sampler believes that observations far away in the tails of the distribution may not inform about the mean.

The relatively small average losses of the sample mean when the population became increasingly homogeneous show that its performance as an estimator under purposive sampling improves as the population becomes more compact. This is explained by the decreasing likelihood for extreme values to be selected. The relatively large reductions in the posterior risk as the population became more homogeneous are expected since the mean of the posterior distribution of θ given the sample minimizes the posterior risk.

4.3 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 let a=1 and a=2 to simulate simple random sampling and purposive sampling, respectively. The prior distribution of θ remains $N(10,15)$. We take a sample of size 1 each time and observe $x=8$, $x=10$, and $x=12$ respectively, as before. We again compute the posterior risk with the sample mean as the estimator of θ using the squared error loss function. Table 2 shows the results at selected values of σ^2.

Table 2 Effect of σ^2 on the Posterior Risk using Squared Error Loss Function when μ=10 and τ^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 (3.1).

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.

The amount of risk is shown to be consistently lower under purposive sampling at all observed values. This implies the superiority of the sample mean in estimating θ under purposive sampling when the sampled population is more compact. Due to the nature of the loss function, the average losses of the sample mean are equal at observed values having the same distance from the prior mean of θ.

4.4 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 again let a=1 and a=2 to simulate simple random sampling and purposive sampling, respectively. The sampling distribution remains $N(10,25)$. We take a sample of size 1 each time and observe $x=8$, $x=10$, and $x=12$, respectively, as before. We compute the posterior risk using the mean as the estimator of θ under the squared error loss function. Table 3 gives the results at selected values of τ^2.

Table 3 **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 (3.1).

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.

When the prior variance was made to decrease from 15 to 1, the average loss of the sample mean also decreased. Under both sampling designs, risk levels were lowest when the observed value was equal to the prior mean of θ. Again, the average losses of the sample mean were found to be lower under purposive sampling.

4.5 Effect of sample size on the posterior risk

Suppose we let the sample size vary when the sampling distribution is $N(10,25)$ 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 $N(10,12.50)$ when $a=2$ for our purposive samples. For each sample size used, three samples were drawn. Table 4 gives the results using the squared error loss function.

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.

Table 4. Effect of sample size (a) on the posterior risk using squared error loss function σ^2=25, μ=10 and τ^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
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

(a) Based on three samples.
(b) Using equation (3.1).

V. Summary and Conclusions

Purposive sampling is an optimal Bayesian sampling strategy. 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 based on 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. Also, the risk levels under purposive sampling are consistently lower than those under simple random sampling as the sample size increases.

References

BASU, D., 1969, Role of the sufficiency and likelihood principles in sample survey theory, *Sankhya*, A 31: 441-454.

BERGER, J.O., 1985, *Statistical Decision Theory and Bayesian Analysis*, 2nd edition, New York: Springer-Verlag, Inc.

CARLIN, B.P. and T.A. LOUIS, 1996, *Bayes and Empirical Bayes Methods for Data Analysis*, London: Chapman & Hall.

CASELLA, G. and R.L. BERGER, 1990, *Statistical Inference*, CA: Wadsworth, Inc.

GHOSH, M. and G. MEEDEN, 1997, *Bayesian Methods for Finite Population Sampling*, London: Chapman & Hall.

GODAMBE, V.P., 1966, A new approach to sampling from finite populations, *J. R. Statist. Soc.*, B 28: 310-328.

GRAYBILL, F.A., 1976, *Theory and Application of the Linear Model*, North Scituate, Mass.: Duxbury Press.

MOOD, A.M., F.A. GRAYBILL, and D.C. BOES, 1974, *Introduction to the Theory of Statistics*,3rd edition, Quezon City: JMC Press, Inc.

SEARLE, S.R., 1982, *Matrix Algebra Useful for Statistics*, New York: John Wiley & Sons, Inc.

SHAO, J. and D. TU, 1995, *The Jackknife and Bootstrap*, New York: Springer-Verlag, Inc.

VALLIANT, R., A.H. DORFMAN, and R.M. ROYALL, 2000, *Finite Population Sampling and Inference: A Prediction Approach*, New York: John Wiley & Sons, Inc.

ZACKS, S., 1969, Bayesian sequential designs for sampling finite populations, *Journal of American Statistics Association*, 64: 1342-1349.