Statistica Sinica Preprint No: SS-2018-0063

<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>A Unified Theory for Robust Bayesian Prediction Under a General Class of Regret Loss Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Manuscript ID</strong></td>
<td>SS-2018-0063</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://www.stat.sinica.edu.tw/statistica/">http://www.stat.sinica.edu.tw/statistica/</a></td>
</tr>
<tr>
<td><strong>DOI</strong></td>
<td>10.5705/ss.202018.0063</td>
</tr>
<tr>
<td><strong>Complete List of Authors</strong></td>
<td>Nader Nematollahi, Mohammad Jafari Jozani and Razieh Jafaraghaie</td>
</tr>
<tr>
<td><strong>Corresponding Author</strong></td>
<td>Nader Nematollahi</td>
</tr>
<tr>
<td><strong>E-mail</strong></td>
<td><a href="mailto:nematollahi@atu.ac.ir">nematollahi@atu.ac.ir</a></td>
</tr>
</tbody>
</table>

Notice: Accepted version subject to English editing.
A Unified Theory for Robust Bayesian Prediction Under a General Class of Regret Loss Functions

Nader Nematollahi\textsuperscript{a,∗}, Mohammad Jafari Jozani\textsuperscript{b} and Razieh Jafaraghaie\textsuperscript{c}

\textsuperscript{a} Department of Statistics, Allameh Tabataba’i University, Tehran, Iran.
\textsuperscript{b} Department of Statistics, University of Manitoba, Winnipeg, MB, CANADA, R3T 2N2
\textsuperscript{c} Department of Mathematics and Statistics, Shoushtar Branch, Islamic Azad University, Shoushtar, Iran.

December 28, 2019

Abstract

We study robust Bayesian prediction problems using the posterior regret Γ-minimax (PRGM) approach. We provide a unified theory for PRGM prediction under a very general class of regret loss functions that includes squared error (SE), linear-exponential (LINEX), Entropy and many other loss functions as its special cases. We apply our results to the problem of predicting unknown parameters of finite populations under different superpopulation models (normal and non-normal with or without auxiliary variables) and several classes of prior distributions including the commonly used $\epsilon$-contaminated class of priors. Results are augmented with real world applications and simulation studies.

Keywords: Bayes predictor; Finite population; Posterior regret Γ-minimax; Robust Bayesian analysis.

1 Introduction

Bayesian approach provides a very attractive methodology towards inference about population parameters and allows for prior information about the underlying problem to be incorporated in the analysis through the prior distribution. Perhaps the main barrier in using the Bayesian approach is due to the subjectivity involved in choosing a single and completely specified prior distribution for the parameter of interest. In other words, a practitioner who produces subjective Bayesian estimates might be vulnerable to public criticism just as the sampler using a purposive sampling plan (Little, 2004). In general, there is no single method for choosing a prior distribution and different users may produce different priors and therefore arrive at different posteriors and conclusions. In some situations, one might choose a family of prior distributions

\[ \text{Corresponding author: nematollahi@atu.ac.ir} \]
which depends on some unknown hyper parameters. However, as it was shown in Ghosh and Kim (1993),
even in simple examples, failure in specifying the correct values of one or more of hyper parameters might
have serious consequences from a Bayesian viewpoint and protection should be taken against such problems.
To address this issue, one solution is to use a robust Bayesian approach by choosing a class $\Gamma$ of plausible prior
distributions for the parameter of interest and obtaining Bayesian solutions that are relatively insensitive
to the uncertainty in prior distribution elicitation. This is a common practice when the underlying problem
must be solved by two or more decision-making agents and they do not necessarily agree on which prior
distribution to use. Also, as any selected prior distribution is just an approximation of the true and unknown
prior distribution, sometimes it is better to choose a wide class of prior distributions instead of choosing
a single one (Berger, 1984). By using robust Bayesian analysis, a range of Bayes estimators are obtained
and one needs to specify which of these estimators is most appropriate. In other words, it is interesting not
only to investigate the range of estimators but also to construct the optimal procedures. Several methods
can be used for this purpose including the $\Gamma$-minimax (GM) (e.g., Berger, 1984), the conditional $\Gamma$-minimax
(CGM) (e.g., Betro and Ruggeri, 1992), the most stable (MS) (e.g., Mezasaki and Zielinski, 1991) and the
posterior regret $\Gamma$-minimax (PRGM) approach that minimizes maximal posterior regret in specifying the
optimum robust Bayes estimator/predictor (e.g., Rios Insua et al., 1995; Boratynska, 2006; Jafari Jozani
and Parsian, 2008).

In this paper, we consider the problem of robust Bayesian prediction using the PRGM approach under
a very general class of loss functions and different classes of prior distributions. We geared our methodology
towards predicting unknown parameters of finite populations, however, results are equally valid for other
problems such as PRGM estimation in parametric inference based on infinite populations. To this end, we
consider a finite population and adopt a model-based approach that views the finite population as a sample
from a superpopulation (parametric) model, which depends on some unknown parameters. A sample of
size $n$ is selected from this population and the goal is to predict some characteristics of the underlying
population such as the population mean, total, variance, etc. The Bayesian prediction of finite population
total (or mean) within the class of unbiased and linear unbiased predictors are studied in Godambe and
Joshi (1965) and Godambe (1955), respectively. More information about the Bayesian inference for finite
populations can be found in Hill (1968), Ericson (1969), Bolfarine (1990), Hamner et al. (2001), Liu and
Rong (2007), Kim and Saleh (2008), Ghosh (2008), Pfeffermann and Rao (2009), Chen et al. (2012), Si
et al. (2015) and references there in. The need for robust Bayesian analysis in survey sampling has been
recognized by many authors, e.g., Godambe and Thompson (1971), Ghosh and Kim (1993), Ghosh (2008),
and Zangeneh and Little (2015). In the context of the robust Bayesian approach in finite population, Ghosh
and Kim (1993) obtained a robust Bayes predictor of the finite population mean based on the posterior
risk and robustness procedure suggested by Berger (1984) using ML-II priors and studied its performance
over a class of prior distributions under the SE loss function. According to our best knowledge, no previous work has been done on the PRGM prediction of the finite population parameters.

The paper is organized as follows. In Section 2, we give some definitions and preliminary results and define the general class of loss functions that will be used throughout the paper. In Section 3, we obtain our main result regarding the PRGM prediction of unknown parametric functions under general classes of loss functions and prior distributions. In Section 4, we provide some applications of our results to obtain PRGM predictors of the finite population mean and/or variance under the SE and LINEX loss functions and some classes of prior distributions for normal and non-normal superpopulation models. In Section 5, we provide real-life examples to predict finite population parameters such as the mean and variance under normal and non-normal models. Also, we compare the estimated risk and bias of PRGM and Bayes predictors under the SE loss function via simulation studies. Finally, concluding remarks are given in Section 6. Some of the proofs and the details of the derivations in a number of examples as well as numerical results for two real data applications are presented in a supplementary document.

2 Preliminaries

In this section, we give some notations and preliminary results which will be used throughout the paper. Consider a finite population of \( N \) units denoted by an index set \( \mathcal{U} = \{1, 2, \ldots, N\} \) and suppose \( \mathbf{y} = (y_1, \ldots, y_N)^\top \) is the vector of unknown values associated with a characteristic of interest \( y \). From \( \mathcal{U} \), a sample \( s = \{i_1, \ldots, i_n\} \) of size \( n(s) = n \) is selected by using the simple random sampling without replacement (SRSWOR) method. A typical sample point is then the set of labels of units contained in the observed sample along with \( \mathbf{y}(s) = (y_{i_1}, \ldots, y_{i_n})^\top \), where \( y_{i_j} \) is the observed value of the characteristic of interest for unit \( i_j \) selected in the sample. Using \( \mathbf{y}(s) \), we are interested in making inference about some unknown finite population quantities, \( \gamma(y) \), such as the population mean \( \gamma_1(\mathbf{y}) = \frac{1}{N} \sum_{i=1}^{N} y_i = \bar{Y} \) and population variance \( \gamma_2(\mathbf{y}) = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{Y})^2 \).

We adopt a model-based approach (e.g., Little, 2004), where the finite population \( \mathbf{y} \) is assumed to be a realization from a superpopulation model \( f(\mathbf{y}|\theta) \) that depends on some unknown parameters \( \theta \), and making inference about \( \gamma(y) \) reduces to the problem of predicting outcomes of the non-sampled units \( \mathbf{y}_\pi = \{y_i : i \notin s\} \). We follow a robust Bayesian methodology, where one specifies a class \( \Gamma \) of prior distributions for superpopulation parameters \( \theta \). The posterior predictive density, \( h(\gamma(\mathbf{y})|\mathbf{y}_s) \), the posterior distribution of \( \gamma(\mathbf{y}) \) given the observed data \( \mathbf{y}_s \), is the basis of inference about \( \gamma(\mathbf{y}) \). This is a general setting that covers many situations not necessarily restricted to a finite population context. For example, for predicting a future observation in its usual parametric setting, we have a sequence of random variables \( y_1, y_2, \ldots, y_n \) and we want to predict the future random variable \( y_{n+1} \). Here, \( \mathbf{y}_s = (y_1, \ldots, y_n)^\top, \mathbf{y}_s = y_{n+1} \).
and \( \gamma(y) = y_s = y_{n+1} \). So, in each case our goal is to predict some functions \( \gamma(y) \) using predictors \( \delta(y_s) \).

Let \( L(\gamma(y), \delta(y_s)) \) be the loss function for predicting \( \gamma(y) \) using \( \delta(y_s) \). To obtain a robust Bayesian predictor of \( \gamma(y) \), one needs to calculate the posterior risk of \( \delta = \delta(y_s) \) for a given prior \( \pi \in \Gamma \) as follows:

\[
\rho(\pi, \delta) = E[L(\gamma(y), \delta(y_s))|y_s].
\]

Then, the posterior risk (1) will be minimized over the class of all possible predictors \( D \) when the prior distribution is also changing in \( \Gamma \). One may also attempt to determine an optimal predictor minimizing some measures such as maximal posterior regret that is defined below (e.g., Berger, 1990).

**Definition 2.1.** \( \delta^{PRGM}_\Gamma \) is a PRGM predictor of \( \gamma(y) \) if, 
\[
\sup_{\pi \in \Gamma} R(\delta^{PRGM}_\Gamma, \delta^\pi) = \inf_{\delta \in D} \sup_{\pi \in \Gamma} R(\delta, \delta^\pi),
\]

where

\[
R(\delta, \delta^\pi) = \rho(\pi, \delta) - \rho(\pi, \delta^\pi),
\]

is the posterior regret due to loss of optimality by using \( \delta \) instead of the Bayes predictor \( \delta^\pi \).

In order to obtain robust Bayes predictors of \( \gamma(y) \) one needs to specify a loss function \( L(\gamma(y), \delta(y_s)) \) to measure the amount of error that is made in predicting \( \gamma(y) \) using \( \delta(y_s) \). In this paper, we introduce a general class of loss functions \( L(\gamma(y), \delta) \) where \( L \) is assumed to be strictly bowl shaped function of both \( \gamma(y) \) and \( \delta \) with unique minimum at \( \delta = \gamma(y) \), and satisfies some additional conditions. Note that \( f(t) \) is called a strictly bowl shaped (BS) function on its domain if, as a function of \( t \), it first decreases and then increases with a unique minimum at \( t_0 \). In other words, \( f'(t) < 0 \) for all \( t < t_0 \) and \( f'(t) > 0 \) for \( t > t_0 \). Obviously any convex loss function is a BS function.

**Definition 2.2.** Consider a class \( \Gamma \) of prior distributions on an unknown parameter \( \theta \). Suppose \( \pi \in \Gamma \) and let \( R(\delta, \delta^\pi) \) be the posterior regret as in (2), where \( \delta^\pi \) is the Bayes estimator of \( \gamma(y) \) with respect to \( \pi \). Suppose \( L \) is a class of loss functions \( L(\gamma(y), \delta) : \mathbb{R}^2 \to \mathbb{R}^+ \) such that \( L \) is a BS function of both \( \gamma(y) \) and \( \delta \). We call \( L \) to be the class of regret loss functions if

\[
R(\delta, \delta^\pi) = L(\delta^\pi, \delta).
\]

One can easily show that many commonly used loss functions satisfy (3). Table 1 provides a number of such loss functions. In the following result we obtain a necessary condition for a strictly BS loss function to satisfy (3). Assume that the predictive distribution of \( \gamma(y) \) given \( y_s \), say \( h(\gamma(y)|y_s) \), is not trivial, i.e., degenerate, as the result is always true for degenerate \( h(\gamma(y)|y_s) \). Now, we have the following result which is proved in the supplementary document:

**Lemma 2.3.** Suppose \( \delta^\pi(y_s) \) is the Bayes predictor of \( \gamma(y) \) under a strictly BS loss function \( L(\gamma(y), \delta) \) with respect to a prior distribution \( \pi \). Suppose the posterior predictive distribution \( h(\gamma(y)|y_s) \) is not degenerate. Then, (3) does not hold for any strictly BS loss function \( L(\gamma(y), \delta) \) that is bounded.
Table 1: Examples of regret loss functions with associated PRGM predictors, where $\delta = \inf_{\pi \in \Gamma} \delta^\pi$ and $\overline{\delta} = \sup_{\pi \in \Gamma} \delta^\pi$, respectively, with $\delta^\pi$ being the Bayes predictor w.r.t. the underlying loss function.

<table>
<thead>
<tr>
<th>Loss function</th>
<th>$L(\gamma(y), \delta)$</th>
<th>PRGM predictor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Squared Error Loss (SEL)</td>
<td>$(\delta - \gamma(y))^2$</td>
<td>$\frac{\delta + \overline{\delta}}{2}$ Zen and DasGupta (1993)</td>
</tr>
<tr>
<td>Linear Exponential Loss (LINEX)</td>
<td>$b{e^{c(\delta - \gamma(y))} - c(\delta - \gamma(y))} - 1$</td>
<td>$\frac{-1}{c} \log \frac{e^{-c(\delta - \gamma(y))} - c\delta}{-c(\delta - \gamma(y))}$ Boratynska and Drozdowicz (1999)</td>
</tr>
<tr>
<td>Entropy Loss (EL)</td>
<td>$\frac{\gamma(y)}{\overline{\delta}} - \ln \frac{\gamma(y)}{\overline{\delta}} - 1$</td>
<td>$\overline{\delta} - \delta$ Jafari Jozani and Parsian (2008)</td>
</tr>
<tr>
<td>Stein’s Loss (SL)</td>
<td>$\frac{\delta}{\gamma(y)} - \ln \frac{\delta}{\gamma(y)} - 1$</td>
<td>$\frac{\ln \frac{\delta}{\gamma(y)} - 1}{\delta - \gamma(y)}$ Jafari Jozani and Jafari Tabrizi (2013)</td>
</tr>
<tr>
<td>Square Log Error Loss (SLEL)</td>
<td>$(\ln \delta - \ln \gamma(y))^2$</td>
<td>$\sqrt{\frac{\delta}{\overline{\delta}}}$ Kiapur and Nematollahi (2011)</td>
</tr>
<tr>
<td>h-Loss (HL)</td>
<td>$(h(\delta) - h(\gamma(y)))^2$</td>
<td>$h^{-1}\left(h(\delta) + h(\overline{\delta})\right)$ Jafari Jozani et al. (2012)</td>
</tr>
<tr>
<td>Intrinsic Loss (IL)</td>
<td>$\ln \frac{\beta(\gamma(y))}{\beta(\delta)} + (\delta - \gamma(y))\frac{\beta'(\gamma(y))}{\beta'(\gamma(y))}$, $\beta(\cdot) &gt; 0$</td>
<td>$\frac{\delta H(\overline{\delta}) - \delta H(\overline{\delta}) - \ln \frac{\beta(\delta)}{\beta(\gamma(y))}}{H(\overline{\delta}) - H(\overline{\delta})}$ Jafari Jozani and Jafari Tabrizi (2013)</td>
</tr>
</tbody>
</table>

3 PRGM predictor of $\gamma(y)$ under the regret loss functions

In this section, we obtain PRGM predictors of $\gamma(y)$ under regret loss functions and a general class $\Gamma$ of priors. PRGM predictors are constructed to minimize the maximum posterior regret in predicting $\gamma(y)$. Posterior regret in the Bayesian analysis can essentially be considered as the difference between posterior risk associated with the best predictor that could have been used if we knew the true prior distribution in $\Gamma$ and the posterior risk of the predictor that was in fact used. Using the minimax strategy, we then choose a predictor that minimizes the maximum of this posterior regret within the class of all predictors when the prior distribution varies in $\Gamma$. This helps to protect against the effects of priors that are causing the worst posterior risks. The key result of this section is given in the following theorem which is proved in the supplementary document.

**Theorem 3.1.** Suppose $L(\gamma(y), \delta)$ is a regret loss function. Let $y$ be a random vector with pdf $f(y|\theta)$. Suppose $\theta$ has a prior distribution with a pdf $\pi(\cdot)$ which belongs to a class $\Gamma$ of priors and $\mathcal{D}$ is the class of all predictors. Let $\underline{\delta}(y_s) = \inf_{\pi \in \Gamma} \delta^\pi(y_s)$ and $\overline{\delta}(y_s) = \sup_{\pi \in \Gamma} \delta^\pi(y_s)$ are finite, where $\delta^\pi(y_s)$ is the Bayes predictor of $\gamma(y)$ with respect to $\pi \in \Gamma$ under the loss function $L(\gamma(y), \delta)$. The PRGM predictor of $\gamma(y)$,
denoted by $\delta^p(y_s)$, is given as a solution to the following equation
\begin{equation}
L(\delta(y_s), \delta^p(y_s)) = L(\delta(y_s), \delta^p(y_s)), \quad \text{for all } y_s \in \mathbb{R}^n.
\end{equation}

If the solution to (4) is not unique, the PRGM predictor is chosen to be a solution that results in the minimum $L(\delta(y_s), \delta^p(y_s))$.

Table 1 provides the PRGM predictors of $\gamma(y)$ under a general class of prior distributions $\Gamma$ for some commonly used loss functions in the literature.

**Remark 3.2.** In some cases, the posterior regret function (2) has the following form
\[
R(\delta(y_s), \delta^\pi(y_s)) = k(y_s, \alpha)L(\delta^\pi(y_s), \delta(y_s)),
\]
where $k(y_s, \alpha)$ is a function of $y_s$ and $\alpha$, with $\alpha$ being a hyper-parameter associated with the prior distribution $\pi_\alpha$ in the class $\Gamma$ of priors. When $k(y_s, \alpha)$ does not depend on $\alpha$, i.e., $k(y_s, \alpha) = k'(y_s)$ then, similar to the proof of Theorem 3.1, it can be shown that the PRGM predictor of $\gamma(y)$ is the solution of equation (4).

For example, suppose the loss function is given by
\[
L_w(\gamma(y), \delta) = \left[\left(\frac{\delta}{\gamma(y)}\right)^{w/2} - \left(\frac{\gamma(y)}{\delta}\right)^{w/2}\right]^2 = \left(\frac{\delta}{\gamma(y)}\right)^w + \left(\frac{\gamma(y)}{\delta}\right)^w - 2.
\]

Then, the Bayes predictor of $\gamma(y)$ is given by $\delta^\pi(y_s) = 2w\sqrt{E(\gamma^w(y)|y_s)}$ and one can easily show that
\[
R(\delta(y_s), \delta^\pi(y_s)) = \frac{E(\gamma^w(y)|y_s)}{E(\gamma^{-w}(y)|y_s)} L_w(\delta^\pi(y_s), \delta(y_s)) = K(y_s, \pi_\alpha) L_w(\delta^\pi(y_s), \delta(y_s)).
\]

Now, if $K(y_s, \pi_\alpha)$ does not depend on the hyper-parameter $\alpha$, then the PRGM predictor of $\gamma(y)$ is the solution of equation (4) and is given by $\delta^p(y_s) = \sqrt{\delta(y_s) \delta^\pi(y_s)}$.

### 4 PRGM prediction under various superpopulation models

In this section, we use Theorem 3.1 to find the PRGM predictors of some characteristics of finite populations such as the mean and the variance under some regret loss functions and various normal and non-normal superpopulation models with or without using auxiliary variables. First, we consider the prediction of the finite population mean and variance when the underlying population is assumed to be generated from a normally distributed superpopulation. We then study the problem when auxiliary variables are also used in the prediction process. Finally, we provide results for PRGM prediction for a non-normal superpopulation model. It is worth noting that results can be obtained for any population parameters, however, we only present those for predicting the population mean and variance as two main parameters of interest in many finite population studies.
4.1 PRGM prediction of mean (Normal model without auxiliary variables)

Consider the following superpopulation model (Ghosh and Kim, 1993), where

\[ y_i = \theta + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), \quad i = 1, \ldots, N, \quad (5) \]

with \( \varepsilon_i, \ i = 1, \ldots, N \), being independent and identically distributed. Suppose \( \sigma^2 \) is known and \( \theta \) is distributed according to a \( N(\mu, \tau^2) \) distribution. We find the PRGM predictors of the mean \( \gamma_1(y) \) under the SE and LINEX loss functions and three classes of prior distributions. Let \( M = \sigma^2/\tau^2, \ B = M/(M + n), \ \bar{y}_s = n^{-1} \sum_{i=1}^{n} y_i \) and \( y_{\bar{s}} = \{y_i : i \notin s\} \). It is easy to see that

\[ y_{\bar{s}} | y_s \sim MN \left( \{\bar{y}_s - B(\bar{y}_s - \mu)\}I_{N-n}, \sigma^2(I_{N-n} + (M + n)^{-1}J_{N-n}) \right), \quad (6) \]

where \( MN \) denotes a multivariate normal distribution, \( I_{N-n} \) is a vector of 1s, \( I_{N-n} \) is the identity matrix and \( J_{N-n} = I_{N-n}1_{N-n}^{\prime} \). Using (6), the conditional distribution of \( \gamma_1(y) \) given \( y_s \) is given by

\[ \gamma_1(y)|y_s \sim N \left( \bar{y}_s - (1 - f)B(\bar{y}_s - \mu), \frac{\sigma^2(1 - f)}{n}(1 - (1 - f)B) \right), \quad (7) \]

with \( f = n/N \). In general, choosing an appropriate class \( \Gamma \) of priors for posterior robustness is important. The ultimate goal is to choose a large \( \Gamma \) to ensure that nothing is left out and at the same time make sure \( \Gamma \) is not too large such that posterior robustness is not achievable. As suggested by Berger (1990), the process of robust Bayesian analysis should be regarded as a data-interactive process in which one starts with a perhaps very large \( \Gamma \), checks the robustness, and progressively refines \( \Gamma \) (if needed) until robustness is achieved. In this section, we consider the following classes of prior distributions for \( \theta \) in (5):

\[ \Gamma_{\mu} = \{\pi : \pi \ is \ N(\mu, \tau_0^2); \mu_1 \leq \mu \leq \mu_2, \ \tau_0^2 \ is \ a \ known \ constant\}, \]

\[ \Gamma_{\tau^2} = \{\pi : \pi \ is \ N(\mu_0, \tau^2); \tau_1^2 \leq \tau \leq \tau_2^2, \ \mu_0 \ is \ a \ known \ constant\}, \]

\[ \Gamma_\epsilon = \{\pi : \pi = (1 - \epsilon)\pi_0 + \epsilon q, q \in Q\}. \]

Classes \( \Gamma_{\mu} \) and \( \Gamma_{\tau^2} \) are appealing as they are very easy to work with (cf. Goldstein, 1980; Berger, 1985). However, they often fail to include many priors that are plausible. For example, they do not admit much variation in the prior tails, and hence may provide a false illusion that robustness is obtained. In such situations, one might consider \( \Gamma_\epsilon \) as a very rich and flexible alternative class of prior distributions (e.g., Berger, 1985) where \( \pi_0 \) is a base prior, \( q \) is a contamination, \( Q \) is a class of plausible distribution functions and \( 0 \leq \epsilon \leq 1 \) reflects the amount of contamination.

**Example 4.1.** (SE and LINEX loss functions) Under the SE loss function, using (7), and as we show in the supplementary document, the PRGM predictors of the population mean \( \gamma_1(y) = \frac{1}{N} \sum_{i=1}^{N} y_i \) under \( \Gamma_{\mu} \),
The Bayes predictor under the prior distribution \( \gamma \) is very confident in \( \epsilon \)-contaminated class of priors when \( \epsilon \) is close to one. Under the LINEX loss function, the PRGM predictors of the population parameters are given respectively by

\[
\delta_{\mu}^{\text{PRGM}}(y_s) = y_s - (1 - f) B_0(y_s - \mu_1 + \mu_2) / 2,
\]

(8)

\[
\delta_{\tau}^{\text{PRGM}}(y_s) = y_s - (1 - f)(y_s - \mu_0)(B_1 + B_2) / 2,
\]

(9)

\[
\delta_{\epsilon}^{\text{PRGM}}(y_s) = f y_s + (1 - f) \left( a \delta(y_s) + \theta_l f(y_s|\theta_l) / a + f(y_s|\theta_l) \right),
\]

(10)

where \( B_i = \frac{\sigma_i^2}{\sigma^2 + \nu_i^2}, i = 0, 1, 2 \), and \( \delta(y_s) = E_{\pi_0}(y_s|\theta) \). Also, \( a = \frac{1 + \epsilon}{\epsilon} m(y_s|\pi_0) \), with \( m(y_s|\pi_0) \) being the marginal (predictive) density of \( y_s \) under the prior distribution \( \pi_0 \), \( \theta_l = \frac{\sigma}{\sqrt{\nu}} \xi + y_s \), \( \xi \in \{l, u\} \), where \( \nu_l \) and \( \nu_u \) are solutions to the following equation in \( \nu \) for some specific values of \( c \) and \( b \) defined in the supplementary document for Example 4.1:

\[
e^{-\nu^2/2} - \nu^2 b - c = 0.
\]

(11)

One can see that \( \delta_{\mu}^{\text{PRGM}} \) is a Bayes predictor of \( \gamma_1(y) \) with respect to (w.r.t.) \( \pi_{\mu^*} \in \Gamma_\mu \) with a \( \mu^* = \frac{\mu_1 + \mu_2}{2} \in [\mu_1, \mu_2] \). Similarly, \( \delta_{\tau}^{\text{PRGM}} \) is a Bayes predictor of \( \gamma_1(y) \) w.r.t. \( \pi_{\tau^*} \in \Gamma_\tau \), with \( \tau^* = 2\nu_1^2 \nu_l^2 + \alpha^2 (\nu_l^2 + \nu_u^2) / 2 \alpha^2 + \nu_l^2 + \nu_u^2 \) (\( \nu_l^2 \leq \nu_u^2 \leq \nu^2_2 \)). Also \( \delta_{\epsilon}^{\text{PRGM}}(y_s) \) can be considered as a compromise between the Bayes predictor under the prior distribution \( \pi_0 \) (associated with \( \epsilon = 0 \) corresponding to the case where one is very confident in \( \pi_0 \)) and the predictor obtained as the mid-range of the class of Bayes predictors under the \( \epsilon \)-contaminated class of priors when \( \epsilon \) is close to one. Under the LINEX loss function, the PRGM predictors of the population mean \( \gamma_1(y) \) under \( \Gamma_\mu \), \( \Gamma_\tau \) and \( \Gamma_\epsilon \) are given respectively by

\[
\delta_{\mu}^{\text{PRGM}}(y_s) = y_s - \frac{c(1 - f) \sigma^2}{2n} (1 - (1 - f) B_0) - \frac{1}{c} \ln \frac{c B_0 (1 - f)(y_s - \mu_1) - c B_0 (1 - f)(y_s - \mu_2)}{c (\mu_2 - \mu_1)(1 - f) B_0},
\]

(12)

\[
\delta_{\tau}^{\text{PRGM}}(y_s) = y_s - \frac{c(1 - f) \sigma^2}{2n} - \frac{1}{c} \ln \frac{c B_2 (1 - f)(y_s - \mu_0) - c B_2 (1 - f) \sigma^2}{2n} - \frac{c B_1 (1 - f)(y_s - \mu_0) - c (1 - f) \sigma^2}{2n},
\]

(13)

\[
\delta_{\epsilon}^{\text{PRGM}}(y_s) = f y_s - \frac{c \sigma^2 (1 - f)^2}{2(N - n)} - \frac{1}{c} \ln \frac{a \sigma^2 + c (1 - f) \sigma^2}{a + f(y_s|\theta_l)} - \frac{a \sigma^2 + c (1 - f) \sigma^2}{a + f(y_s|\theta_l)} - \ln \frac{a \sigma^2 + c (1 - f) \sigma^2}{a + f(y_s|\theta_l)},
\]

(14)

where \( \nu_l \) and \( \nu_u \) are obtained numerically as solutions to a nonlinear equation in \( \nu \) that can be found in the supplementary document.

4.2 PRGM prediction of variance (Normal model without auxiliary variables)

Here we consider model (5) assuming that both \( \theta \) and \( \sigma^2 \) are unknown, and obtain the PRGM predictors of the population variance \( \gamma_2(y) = \frac{1}{N} \sum_{i=1}^{n} (y_i - \gamma_1(y))^2 \) under the SE loss function and two classes of prior distributions. Let the prior distribution for \( \sigma^2 \) be the inverse gamma distribution \( \Pi(\alpha, \beta) \) with known \( \alpha \) and \( \beta \) while we choose a noninformative prior for \( \theta \) with density \( \pi(\theta) = 1, \theta \in \mathbb{R} \), resulting in \( \pi(\theta, \sigma^2) \propto \sigma^{-2(\alpha + 1)} e^{-\frac{\sigma^2}{2}} \). Let \( A = \frac{1}{2} \sum_{i \in S} (y_i - \overline{y}_S)^2 + \beta, A_i = \frac{1}{2} \sum_{i \in S} (y_i - \overline{y}_S)^2 + \beta_i \), \( i = 0, 1, 2 \),

\[\sqrt{\text{S}} \]
\[\bar{y}_s = \frac{1}{N-n} \sum_{i \in s} y_i, \quad s^2 = \frac{1}{n-1} \sum_{i \in s} (y_i - \bar{y}_s)^2 \quad \text{and} \quad s^2_\alpha = \frac{1}{N-n} \sum_{i \in s} (y_i - \bar{y}_s)^2. \]

It can be shown that
\[
E(\theta | y_s) = \bar{y}_s, \quad V(\theta | y_s) = \frac{2A}{n(n+2\alpha-3)} \quad \text{and} \quad \pi(\theta, \sigma^2 | y_s) = c \left(\frac{1}{\sigma^2}\right)^{\frac{n+2\alpha+2}{2}} \exp\left\{ -\frac{1}{\sigma^2}(A + \frac{y}{2}(\bar{y}_s - \theta)^2) \right\},
\]

where \(c = \sqrt{2\pi \Gamma(\frac{n+2\alpha-1}{2})}\). Also,
\[
\sum_{i=1}^{N} (y_i - \bar{Y})^2 = (n-1)s^2 + (N-n-1)s^2_\alpha + f^2 + (1-f)(\bar{y}_s - \bar{y})^2,
\]

(15)

where \(\bar{y}_s, s^2\) and \(s^2_\alpha\) given \(\theta, \sigma^2\) are distributed according to \(N(\theta, \frac{\sigma^2}{n-1})\), \(\Gamma(\frac{n-1}{2}, \frac{2\sigma^2}{n})\) and \(\Gamma(\frac{N-n-1}{2}, \frac{2\sigma^2}{n})\), respectively. By first taking the expectation of (15), conditional on \(\theta, \sigma^2\) and \(y_s\), and then calculating the expectation w.r.t. the posterior distribution of \((\theta, \sigma^2)\) given \(y_s\), one can easily see that
\[
E \left[ \sum_{i=1}^{N} (y_i - \bar{Y})^2 | y_s \right] = \sum_{i \in s} (y_i - \bar{y}_s)^2 + \frac{2AN}{n + 2\alpha - 3} N_0,
\]

with \(N_0 = \frac{N-n-1}{N} + \frac{(f^2 + (1-f)^2)}{n(N-n)}\). Now, the Bayes predictor of \(\gamma_2(y)\) under the SE loss function is
\[
\delta^\pi(y_s) = \frac{1}{N} \sum_{i \in s} (y_i - \bar{y}_s)^2 + \frac{2A}{n + 2\alpha - 3} N_0.
\]

(16)

We consider the following classes of prior distributions for \(\sigma^2\):

\[
\Gamma_\alpha = \{ \pi : \pi \text{ is } \Gamma(\alpha, \beta_0); \alpha_1 \leq \alpha \leq \alpha_2, \beta_0 \text{ is a known constant}\},
\]

\[
\Gamma_\beta = \{ \pi : \pi \text{ is } \Gamma(\alpha_0, \beta); \beta_1 \leq \beta \leq \beta_2, \alpha_0 \text{ is a known constant}\}.
\]

Under \(\Gamma_\alpha\), the Bayes predictor of \(\gamma_2(y)\) is given by (16) when \(A\) is replaced with \(A_0\). Following Table 1, the PRGM predictor of \(\gamma_2(y)\) under the SEL loss function is obtained as follows
\[
\delta^{\text{PRGM}}(y_s) = \frac{1}{N} \sum_{i \in s} (y_i - \bar{y}_s)^2 + \frac{2A_0}{n + 2\alpha^* - 3} N_0,
\]

(17)

where \(\alpha^* = \frac{2\alpha_0 + (n-3)\alpha_1 + \alpha_2}{(n-3)+\alpha_1+\alpha_2}\). Note that \(\delta^{\text{PRGM}}(y_s)\) is a Bayes predictor of \(\gamma_2(y)\) w.r.t. \(\pi_{\alpha^*} \in \Gamma_\alpha\) with \(\alpha^* \in [\alpha_1, \alpha_2]\). Under \(\Gamma_\beta\), the Bayes predictor of \(\gamma_2(y)\) under the SE loss function is given by (16) where \(\alpha\) is replaced with \(\alpha_0\). Using Table 1, one can easily show that the PRGM predictor of \(\gamma_2(y)\) is
\[
\delta^{\text{PRGM}}(y_s) = \frac{1}{N} \sum_{i \in s} (y_i - \bar{y}_s)^2 + \frac{2A^*}{n + 2\alpha^* - 3} N_0,
\]

where \(A^* = \frac{1}{2} \sum_{i \in s} (y_i - \bar{y}_s)^2 + \beta^*\) and \(\beta^* = \frac{\beta_1 + \beta_2}{2}\). Here again, \(\delta^{\text{PRGM}}\) is a Bayes predictor of \(\gamma_2(y)\) w.r.t. \(\pi_{\beta^*} \in \Gamma_{\beta^*}\), \((\beta_1 \leq \beta^* \leq \beta_2)\).

4.3 PRGM prediction of mean (Normal model with auxiliary variables)

In many populations, particularly those that have been previously sampled or surveyed, a frame of units is available along with some auxiliary data on each unit. In other cases, a full frame of all units is not
available but can be constructed by sampling in stages and assembling a partial frame at each stage. In both single-stage and multistage sampling designs, auxiliary data may be used to construct efficient estimators of population parameters such as population total and mean. A superpopulation model is often used to formalize the relationship between a target variable and auxiliary data. To this end, quantities of interest are modeled as being realizations of random variables with a particular joint probability distribution. A frequently used model in survey sampling is

$$y_i = \beta x_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2), i = 1, \ldots, N,$$

(18)

where $x_i$’s are the values of the auxiliary variable and $\varepsilon_i$’s are independent and identically distributed.

To perform a Bayesian analysis, one can use prior information that often exists in survey sampling in the form of auxiliary variables through administrative records. The Bayesian approach utilizes this auxiliary information explicitly through prior distributions for finite population parameters, i.e., distributions which relate these parameters and the auxiliary variables. It is common to assume that in model (18), $x_i$’s (hence $\sum_{i=1}^{N} x_i$) and $\sigma^2$ are known and $\beta$ follows a $N(\mu, \tau^2)$ distribution. Let $x_s = \{x_i : i \in s\}$, $x_s' = \{x_i : i \notin s\}$, $b_s = \sum_{i \in s} x_i y_i$, $d_s = \sum_{i \in s} x_i^2$, $c_s = \sum_{i \notin s} x_i^2$ and $a_s = \sum_{i \notin s} x_i = \sum_{i=1}^{N} x_i - \sum_{i \in s} x_i$. Then, one can easily show that the posterior distribution of $\mathbf{y}_s$ given $\mathbf{y}_s$ and $\mathbf{x}_s$ is a multivariate normal distribution

$$MN \left( ((1 - B_s)\mu + \frac{b_s}{d_s} B_s) \mathbf{x}_s, \sigma^2 (I_{N-n} + \frac{B_s}{d_s} \mathbf{xx}_s) \right),$$

(19)

where $B_s = \frac{\sigma^2 d_s}{\sigma^2 + \tau^2 d_s}$. Using (19), we can show that

$$\gamma_1(\mathbf{y})|\mathbf{y}_s, \mathbf{x}_s \sim N \left( f\mathbf{y}_s + ((1 - B_s)\mu + \frac{b_s}{d_s} B_s)\frac{a_s}{N}, \frac{\sigma^2}{N^2}((N - n) + \frac{B_s}{d_s} a_s^2) \right).$$

(20)

Thus,

$$E(\gamma_1(\mathbf{y})|\mathbf{y}_s, \mathbf{x}_s) = f\mathbf{y}_s + ((1 - B_s)\mu + \frac{b_s}{d_s} B_s)\frac{a_s}{N}, \quad V(\gamma_1(\mathbf{y})|\mathbf{y}_s, \mathbf{x}_s) = \frac{\sigma^2}{N^2}((N - n) + \frac{B_s}{d_s} a_s^2),$$

(21)

and

$$-\frac{1}{c} \ln E(e^{-c\gamma_1(\mathbf{y})}|\mathbf{y}_s, \mathbf{x}_s) = f\mathbf{y}_s + ((1 - B_s)\mu + \frac{b_s}{d_s} B_s)\frac{a_s}{N} - \frac{1}{2} \frac{\sigma^2}{N^2}((N - n) + \frac{B_s}{d_s} a_s^2).$$

(22)

We consider the $\Gamma_{\mu}$, $\Gamma_{\tau^2}$ and $\Gamma_\epsilon$ classes of prior distributions for $\beta$ in (18). In the following examples, we obtain the PRGM predictors of $\gamma_1(\mathbf{y}) = \frac{1}{N} \sum_{i=1}^{N} y_i$ under the SE and LINEX loss functions.

**Example 4.2.** *(SE and LINEX loss functions)* Under the SE loss function, as we show in the supplementary document, the PRGM predictors of the population mean under $\Gamma_{\mu}$, $\Gamma_{\tau^2}$ and $\Gamma_\epsilon$ classes of priors are obtained
respectively by

\[
\delta^{PRGM}_\mu(y_s) = f\bar{y}_s + \left(1 - B_{s0}\right)\left(\frac{\mu_1 + \mu_2}{2}\right) + \frac{b_s}{d_s B_{s0}} a_s \frac{\mu_2 - \mu_1}{N},
\]

\[
\delta^{PRGM}_\tau(y_s) = f\bar{y}_s + \frac{a_s}{N} \mu_0 + \frac{a_s}{N} \frac{b_s}{d_s - \mu_0} \left(\frac{B_{s1} + B_{s2}}{2}\right),
\]

\[
\delta^{PRGM}_\gamma(y_s) = f\bar{y}_s + \frac{(1 - f)\bar{x}_s}{2} \left(\frac{a\delta^0(y_s) + \beta f(y_s|\beta_\gamma)}{a + f(y_s|\beta_\gamma)} + \frac{a\delta^0(y_s) + \beta_u f(y_s|\beta_u)}{a + f(y_s|\beta_u)}\right),
\]

where \( a \) is defined as in Example 4.1, \( B_i = \frac{\tau^2 d_s}{\sigma^2 + \tau^2 d_s} \), \( i = 0, 1, 2 \) and \( \delta^0(y_s) = E_{\gamma_0}(\gamma|y_s) \). Also \( \beta_\gamma = \frac{a}{\sqrt{d_s}} \nu_\gamma \frac{1}{d_s} \), \( \xi \in \{l, u\} \), where \( \nu_0 \) and \( \nu_u \) are obtained as solutions to a nonlinear equation in \( \nu \) that can be found in the supplementary document for Example 4.2. Here again we show that \( \delta^{PRGM}_\mu(y_s) \) and \( \delta^{PRGM}_\tau(y_s) \) are Bayes estimators with specific choices of prior distributions within \( \Gamma_\mu \) and \( \Gamma_\tau \), respectively.

Under the LINEX loss function, the PRGM predictors of the population mean under \( \Gamma_\mu \), \( \Gamma_\tau \) and \( \Gamma_\gamma \) classes of priors are obtained respectively by

\[
\delta^{PRGM}_\mu = f\bar{y}_s + a_s b_s \frac{B_{s0} N d_s}{N d_s} - \frac{c \sigma^2}{2 N^2} \left(N - n\right) + \frac{B_{s0} a_s}{d_s} - \frac{1}{c} \left[\ln \left(1 - B_{s0}\right)\mu_2 - \frac{1}{N} a_s \left(1 - B_{s0}\right)\mu_1\right],
\]

\[
\delta^{PRGM}_\tau = f\bar{y}_s + \frac{a_s}{N} \mu_0 - \frac{c \sigma^2}{2 N^2} \left(1 - f\right) - \frac{1}{c} \left[\ln \left(-B_{s1} + B_{s2}\right)\frac{a_s}{d_s} - \frac{1}{N} a_s \left(-B_{s1} + B_{s2}\right)\mu_0\right] - \frac{c \sigma^2}{2 N^2} \frac{a_s}{d_s} \mu_0 + \frac{1}{2 N^2} \frac{a_s}{d_s} \mu_0 + \frac{1}{2 N^2} \frac{a_s}{d_s} \mu_0,
\]

\[
\delta^{PRGM}_\gamma = f\bar{y}_s - \frac{c \sigma^2 (1 - f)^2}{2 (N - n)} - \frac{1}{c} \left[\ln \left(\frac{b_0 + e^{-c(1-f)\beta_\gamma} f(y_s|\beta_\gamma)}{a + f(y_s|\beta_\gamma)}\right) - \ln \left(\frac{b_0 + e^{-c(1-f)\beta_\gamma} f(y_s|\beta_\gamma)}{a + f(y_s|\beta_\gamma)}\right)\right],
\]

where \( t = E(e^{-c(1-f)\beta_\gamma}|y_s) \) and \( b_0 = a t \). Also, to calculate \( \beta_\tau \) and \( \beta_u \) one needs to numerically solve for \( \nu_0 \) and \( \nu_u \) as solutions to a nonlinear equation defined in the supplementary document.

### 4.4 PRGM prediction of mean (non-normal population)

In many applications such as Business surveys dealing with income data as well as some medical research, the underlying variable of interest is a positive and continuous random variable with a right-skewed distribution. In such cases, using normal models is not appropriate and one might decide to use other densities such as gamma model (e.g., Engelhardt and Bain, 1977, Glaser, 1973, and Gross and Clark, 1975). In this section, we assume that our sample is taken from a superpopulation that is distributed according to a Gamma model. In other words, given \( \theta \), suppose \( y_1, \ldots, y_N \) are conditionally independent with

\[
y_i|\theta \sim \Gamma(\alpha, \theta), \quad i = 1, \ldots, N,
\]

where \( \alpha \) is assumed to be known. So, we have

\[
f(y_i|\theta) = \frac{\theta^\alpha y_i^{\alpha-1} e^{-\theta y_i}}{\Gamma(\alpha)}, \quad i = 1, \ldots, N,
\]
and \( T = \sum_{i \in s} y_i \) given \( \theta \) has \( \Gamma((N-n)\alpha, \theta) \) distribution. We obtain the PRGM predictors of the population mean \( \gamma_1(y) = \frac{1}{N} \sum_{i=1}^{N} y_i \) under the SE loss function and two classes of prior distributions for \( \theta \). Let \( \theta \sim \Gamma(a, b) \) with known \( a \) and \( b \), \( C = \sum_{i \in s} y_i + b \) and \( C_i = \sum_{j \in s} y_j + b, \ i = 0, 1, 2 \). It is easily seen that
\[
f(T = t|y_s) = \frac{\Gamma((N-n)\alpha + n\alpha + a)}{\Gamma((N-n)\alpha)\Gamma(n\alpha + a)} C(1 + \frac{C}{T})^{(N-n)\alpha + n\alpha + a}.
\]
Using the change of variable \( U = (1 + \frac{C}{T})^{-1} \), we obtain the distribution of \( U \) given \( y_s \) as \( \text{Beta}(n\alpha + a, (N-n)\alpha) \), so, \( E(T|y_s) = CE(1-U|y_s) = C(\frac{N-n}{n\alpha + a - 1}) \). Hence the Bayes predictor of \( \gamma_1(y) \) under the SE loss function is as follow:
\[
\delta_\pi(y_s) = \frac{1}{N} \sum_{i \in s} y_i + \frac{1}{N} C(\frac{N-n}{n\alpha + a - 1}).
\]
We consider the following classes of prior distributions for \( \theta \) in (29):
\[
\begin{align*}
\Gamma_a &= \{ \pi : \pi = \Gamma(a, b_0); a_1 \leq a \leq a_2, \ b_0 \text{ is a known constant} \}, \\
\Gamma_b &= \{ \pi : \pi = \Gamma(a_0, b); b_1 \leq b \leq b_2, \ a_0 \text{ is a known constant} \}.
\end{align*}
\]
Over \( \Gamma_a \), the Bayes predictor of \( \gamma_1(y) \) under the SE loss function is given by (30) when \( C \) is replaced with \( C_0 = \sum_{j \in s} y_j + b_0 \). The PRGM predictor of \( \gamma_1(y) \) is then obtained as follows
\[
\delta_{a}^{\text{PRGM}}(y_s) = \frac{1}{N} \sum_{i \in s} y_i + \frac{1}{N} C_0(\frac{N-n}{n\alpha + a - 1}),
\]
where \( a^* = \frac{a_1a_2 + na_2^2}{n\alpha + a_2^2 - 1} \). Note that \( \delta_{a}^{\text{PRGM}}(y_s) \) is a Bayes predictor of \( \gamma_1(y) \) w.r.t. \( \pi_{a^*} \in \Gamma_a \). Under the class \( \Gamma_b \) of priors, the Bayes predictor of \( \gamma_1(y) \) is given by (30) with \( a \) being replaced with \( a_0 \). Also, the PRGM predictor of \( \gamma_1(y) \) is given by \( \delta_{b}^{\text{PRGM}}(y_s) = \frac{1}{N} \sum_{i \in s} y_i + \frac{1}{N} C^*(\frac{N-n}{n\alpha + a - 1}) \), where \( C^* = \sum_{i \in s} y_i + b^* \) and \( b^* = \frac{b_1+b_2}{2} \). Here again, \( \delta^{\text{PRGM}} \) is a Bayes predictor of \( \gamma_1(y) \) w.r.t. \( \pi_{b^*} \in \Gamma_{b^*} \).

5 Real Data Applications and Simulation Studies

In this section, we study the performance of PRGM predictors of population mean and/or variance with respect to several classes of prior distributions compared with their corresponding Bayes predictors under the commonly used SE loss function and different superpopulation models. To this end, we consider three data sets to predict

1. the average and the variance of weight loss of 579 participants in a special diet program in a clinical study in Iran,

2. the average weight of 224 seventh-month sheep at the Research Farm of Ataturk University, Erzurum, Turkey (Ozturk et al., 2005; Jafaraghaie and Nematollahi, 2018); and
(3) the average remission time (in months) of 128 patients with bladder cancer from a study conducted by the American Cancer Society (Lee and Wang, 2003; Lemonte and Cordeiro, 2013).

The first study deals with model (5) associated with a normal superpopulation model without auxiliary information, while the second study considers model (18) using an auxiliary variable. Finally, our third study deals with model (29) based on a Gamma distribution as an example of a non-normal superpopulation model. We present the results for the first application in the paper, while more details regarding the second and third applications are presented in the supplementary document. In our first application, we study a finite population consisting of the measurements on the weight loss of 579 participants that were enrolled in a special diet program in a clinical study in Isfahan city of Iran in 2006. The weight loss is computed as the difference of the weight of each person before starting the program and after finishing it. As the normality assumption for the weight loss was not rejected using the Kolmogrove-Smirnove test with a p-value = 0.514, we assume that our data set is a realization of a normal superpopulation model \( N(\theta, \sigma^2) \) with \( \theta = 3.82295 \) and \( \sigma^2 = 6.499586 \) that are obtained using the maximum likelihood approach.

We would like to predict the average and variance of the weight loss due to the special diet. Doctors who are involved with this study have previous information associated with similarly conducted researches in the clinic and they want to incorporate this information in the prediction process. Bayesian methodology can be used to express doctor’s previous experience as suitable classes of prior distributions in the analysis. We do this by incorporating their information in the prediction process through Bayesian and robust Bayesian approaches. For the PRGM prediction of the population mean and variance we choose reasonable classes of prior distributions for \( \theta \) and \( \sigma^2 \) instead of working with completely determined prior distributions, and obtain the PRGM predictors of the mean and variance of the weight loss. We also obtain the bias and variance associated with each prediction using simulation studies.

5.1 Predicting the average weight loss

To predict the finite population mean, we consider a single prior distribution \( N(\mu_0 = 6, \tau_0^2 = 0.5) \) as well as three classes of prior distributions for \( \theta \) denoted by \( \Gamma_\mu = \{N(\mu, \tau_0^2) : \mu \in [2, 8]\} \), \( \Gamma_{\tau^2} = \{N(\mu_0, \tau^2) : \tau^2 \in [0.1, 0.7] \subseteq \mathbb{R}^+\} \) and \( \Gamma_\epsilon = \{\pi = (1-\epsilon)\pi_0 + \epsilon q : \pi_0 \sim N(6, 0.5), q \sim N(8, 0.3)\}, \epsilon = 0.5\}. \) We obtain the usual Bayes predictor \((\delta^{\mu_0, \tau_0^2}_\mu)\), the PRGM predictor over the class \( \Gamma_\mu \) \((\delta^{PRGM}_\mu)\), the PRGM predictor over the class \( \Gamma_{\tau^2} \) \((\delta^{PRGM}_{\tau^2})\) and the Bayes and PRGM predictors over the class \( \Gamma_\epsilon \) \((\delta^{\pi}_{\Gamma_\epsilon}, \delta^{PRGM}_{\Gamma_\epsilon})\). Table 2 summarizes the predicted values under the SE loss function for fixed sample size \( n = 50 \). As we observe the PRGM predicted values are closer to the true mean weight loss, i.e., 3.82295 than their corresponding Bayes predictors. To obtain the bias and precision associated with each prediction we perform simulation studies. To this end, by considering the weight loss data as a realization
Table 2: The PRGM predicted values of the finite population mean over $\Gamma_{\mu}$, $\Gamma_{\tau^2}$ and $\Gamma_{\epsilon}$ under the SE loss function. Corresponding Bayes predictions are obtained under a $N(6, 0.5)$ prior distribution.

<table>
<thead>
<tr>
<th>$\delta_{\mu_0, \epsilon_{\mu}}^k$</th>
<th>$\delta_{\tau^2}^{PRGM}$</th>
<th>$\delta_{\epsilon}^{PRGM}$</th>
<th>$\delta_{\epsilon}^k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.753405</td>
<td>4.564885</td>
<td>4.68362</td>
<td>4.239414</td>
</tr>
</tbody>
</table>

of a superpopulation model, we extract samples from the underlying model in order to compute predictors, MSEs and biases. We repeat this process 10000 times and calculate the estimated MSE (EMSE) and absolute bias (EAB) of each predictor. To study the effect of the sample size we repeat our study with different sample sizes, $n = 20, 30$, and $50$, for comparison. Also, we study the effect of $\mu_0$ and $\tau_0^2$ on the performance of the Bayes predictors compared with their corresponding PRGM predictors under $\Gamma_{\mu}$, $\Gamma_{\tau^2}$ and $\Gamma_{\epsilon}$ classes of priors. To this end, we consider $\pi_0 \sim N(\mu_0, \tau_0^2)$ with $\mu_0 = 2, 4, 6, 8$ and $\tau_0^2 = 0.1, 0.3, 0.5$ and $0.7$. Also, we use the following four $\epsilon$-contaminated classes of prior distributions corresponding to different choices of $\pi_0$ and $q$ and study PRGM prediction with different values of contamination $\epsilon \in \{0, 0.2, \ldots, 0.8, 1\}$:

1. $\Gamma_{\epsilon}^1 = \{ \pi : \pi = (1 - \epsilon)N(6, 0.3) + \epsilon N(5, 0.2) \}$,
2. $\Gamma_{\epsilon}^2 = \{ \pi : \pi = (1 - \epsilon)N(6, 0.3) + \epsilon N(8, 0.3) \}$,
3. $\Gamma_{\epsilon}^3 = \{ \pi : \pi = (1 - \epsilon)N(6, 0.5) + \epsilon N(5, 0.2) \}$,
4. $\Gamma_{\epsilon}^4 = \{ \pi : \pi = (1 - \epsilon)N(6, 0.5) + \epsilon N(8, 0.3) \}$,

To calculate PRGM predictors we use the necessary expressions developed in Example 4.1, where $\Gamma_{\epsilon}^i$, $i = 1, \ldots, 4$, the necessary values of $\nu_l$ and $\nu_u$ are obtained numerically as solutions to corresponding nonlinear equations (11). We perform simulation studies using the following steps in order to calculate the precisions and biases associated with each prediction:

1. Generate $\epsilon_1^*, \epsilon_2^*, \ldots, \epsilon_n^*$ from a $N(0, 6.499586)$ distribution.
2. Create $y_i^* = 3.822954 + \epsilon_i^*$ and consider $y_i^*, i = 1, \ldots, n$ as generated samples from the underlying superpopulation model.
3. Calculate the Bayes and PRGM predictors.
4. Repeat steps 1–3 for $b = 10^4$ times and calculate the values of EMSE and EAB for each predictor using the following formula:

$$\text{EMSE} = \frac{1}{b} \sum_{i=1}^{b} (\delta_i^k - \bar{Y})^2, \quad \text{EAB} = \frac{1}{b} \sum_{i=1}^{b} |\delta_i^k - \bar{Y}|, \quad k = \text{Bayes, PRGM},$$

where, $\delta_i^k$ is the predictor in $i$–th repetition of sampling and $\bar{Y}$ is the population mean.
Table 3: Simulated MSE and absolute bias for the Bayes and PRGM predictors for \( \mu_0 = 2, 4, 6, 8, \tau_0^2 = 0.1, 0.3, 0.5, 0.7 \) and \( \mu \in [2, 8] \) over \( \Gamma_n \) (losing weight data).

<table>
<thead>
<tr>
<th>( n )</th>
<th>( \mu_0 = 2 )</th>
<th>( \mu_0 = 4 )</th>
<th>( \mu_0 = 6 )</th>
<th>( \mu_0 = 8 )</th>
<th>( n )</th>
<th>( \mu_0 = 2 )</th>
<th>( \mu_0 = 4 )</th>
<th>( \mu_0 = 6 )</th>
<th>( \mu_0 = 8 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \tau_0^2 = 0.1 )</td>
<td>20</td>
<td>1.8300</td>
<td>0.0394</td>
<td>2.6092</td>
<td>0.7793</td>
<td>20</td>
<td>0.5999</td>
<td>0.1282</td>
<td>2.6562</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>1.4227</td>
<td>0.0387</td>
<td>2.0217</td>
<td>0.6093</td>
<td>30</td>
<td>0.3765</td>
<td>0.1073</td>
<td>1.4955</td>
</tr>
<tr>
<td>EMSE</td>
<td>50</td>
<td>0.9089</td>
<td>0.0366</td>
<td>1.2975</td>
<td>0.4004</td>
<td>50</td>
<td>0.1933</td>
<td>0.0796</td>
<td>0.7053</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>1.3446</td>
<td>0.1619</td>
<td>1.6085</td>
<td>0.8702</td>
<td>20</td>
<td>0.6979</td>
<td>0.2853</td>
<td>1.5915</td>
</tr>
<tr>
<td>EAB</td>
<td>30</td>
<td>1.1821</td>
<td>0.1592</td>
<td>1.4129</td>
<td>0.7641</td>
<td>30</td>
<td>0.5356</td>
<td>0.2620</td>
<td>0.6313</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.9387</td>
<td>0.1529</td>
<td>1.1268</td>
<td>0.6104</td>
<td>50</td>
<td>0.3709</td>
<td>0.2248</td>
<td>0.4311</td>
</tr>
<tr>
<td>( \tau_0^2 = 0.3 )</td>
<td>20</td>
<td>0.9129</td>
<td>0.0879</td>
<td>1.2792</td>
<td>0.4315</td>
<td>20</td>
<td>0.4623</td>
<td>0.1579</td>
<td>0.6030</td>
</tr>
<tr>
<td></td>
<td>30</td>
<td>0.5986</td>
<td>0.0797</td>
<td>0.8255</td>
<td>0.2945</td>
<td>30</td>
<td>0.2897</td>
<td>0.1255</td>
<td>0.3630</td>
</tr>
<tr>
<td>EMSE</td>
<td>50</td>
<td>0.3118</td>
<td>0.0649</td>
<td>0.4285</td>
<td>0.1708</td>
<td>50</td>
<td>0.1528</td>
<td>0.0881</td>
<td>0.1872</td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>0.9130</td>
<td>0.2363</td>
<td>1.0953</td>
<td>0.5968</td>
<td>20</td>
<td>0.5844</td>
<td>0.3166</td>
<td>0.6832</td>
</tr>
<tr>
<td>EAB</td>
<td>30</td>
<td>0.7246</td>
<td>0.2259</td>
<td>0.8667</td>
<td>0.4782</td>
<td>30</td>
<td>0.4496</td>
<td>0.2833</td>
<td>0.5151</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.5041</td>
<td>0.2033</td>
<td>0.6063</td>
<td>0.3501</td>
<td>50</td>
<td>0.3201</td>
<td>0.2366</td>
<td>0.3581</td>
</tr>
</tbody>
</table>

Tables 3 and 4 present EMSEs and EABs of Bayes and corresponding PRGM predictors under \( \Gamma_\mu \) and \( \Gamma_{\tau^2} \), respectively. From Table 3 we observe that for small and large values of \( \mu_0 \) (\( \mu_0 = 2, 6, 8 \)) [moderate values of \( \mu_0 \) (\( \mu_0 = 4 \)] and all values of \( \tau_0^2 \) the PRGM predictors [the Bayes predictors] perform reasonably well compared with their corresponding Bayes predictors [the PRGM predictors] in terms of the EMSE and EAB. This is very useful observation as for the Bayesian prediction one needs to specify a prior distribution, which might be very difficult to do so in practice. However, the PRGM predictor works with a class of plausible priors and still performs as good as the Bayesian approach. From Table 4, we observe that for all values of \( \tau_0^2 \) and large values of \( \mu_0 \) (\( \mu_0 = 6, 8 \)) and for large values of \( \tau_0^2 \) and moderate values of \( \mu_0 \) (\( \tau_0^2 = 0.5, 0.7 \) and \( \mu_0 = 4 \)), the PRGM predictors perform reasonably well compared with the Bayes predictors in terms of the estimated MSE and absolute bias. We have opposite result for other values of \( \tau_0^2 \) and \( \mu_0 \). Note that the estimated values of the MSE and the bias decrease as the sample size increases. Tables 5 and 6 present EMSEs and EABs of the Bayes and PRGM predictors under the classes \( \Gamma_\epsilon^i, i = 1, \ldots, 4 \), of priors, for different values of \( \epsilon \). According to these tables, PRGM predictors perform reasonably well compared with their corresponding Bayes predictors in terms of EMSE and EAB under \( \Gamma_\epsilon^2 \) and \( \Gamma_\epsilon^4 \) classes of priors. Also, under \( \Gamma_\epsilon^1 \) and \( \Gamma_\epsilon^3 \), for large values of \( \epsilon \), the PRGM predictors are better than the Bayes predictors and for small values of \( \epsilon \) the Bayes predictors are better than the PRGM predictors. So, when the contaminated distribution is far from the \( \pi_0 \) distribution (\( \Gamma_\epsilon^2 \) and \( \Gamma_\epsilon^4 \)), the PRGM predictor performs reasonably well compared to the Bayes predictor, and when these two distributions are close (\( \Gamma_\epsilon^1 \) and \( \Gamma_\epsilon^3 \)), the PRGM predictor (Bayes predictor) is preferred for large (small) values of \( \epsilon \). Note that, in these tables for \( \epsilon = 0 \), the PRGM predictor is equal to the Bayes predictor.
Table 4: Simulated MSE and absolute bias for the Bayes and PRGM predictors for $\tau^2 = 0.1, 0.3, 0.5, 0.7, \mu_0 = 2, 4, 6, 8$ and $\tau^2 \in [0.1, 0.7]$ over $\Gamma_{\tau^2}$ (losing weight data).

<table>
<thead>
<tr>
<th>$\mu_0$</th>
<th>$n$</th>
<th>$\tau^2 = 0.1$</th>
<th>$0.3$</th>
<th>$0.5$</th>
<th>$0.7$</th>
<th>$\delta^\pi$</th>
<th>$\delta^{PRGM}$</th>
<th>$\mu_0$</th>
<th>$n$</th>
<th>$\tau^2 = 0.1$</th>
<th>$0.3$</th>
<th>$0.5$</th>
<th>$0.7$</th>
<th>$\delta^\pi$</th>
<th>$\delta^{PRGM}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>20</td>
<td>1.8300</td>
<td>0.9129</td>
<td>0.5999</td>
<td>0.4623</td>
<td>1.8302</td>
<td>2.6092</td>
<td>6</td>
<td>20</td>
<td>2.6092</td>
<td>1.2792</td>
<td>0.8136</td>
<td>0.6030</td>
<td>0.6029</td>
<td>0.6029</td>
</tr>
<tr>
<td>30</td>
<td>1.4227</td>
<td>0.5986</td>
<td>0.3765</td>
<td>0.2897</td>
<td>1.4227</td>
<td>2.0217</td>
<td>0.8255</td>
<td>0.4555</td>
<td>0.3630</td>
<td>0.3630</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMSE</td>
<td>50</td>
<td>0.9089</td>
<td>0.3118</td>
<td>0.1933</td>
<td>0.1528</td>
<td>0.9089</td>
<td>1.2975</td>
<td>0.4285</td>
<td>0.2503</td>
<td>0.1872</td>
<td>0.1872</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.3446</td>
<td>0.9130</td>
<td>0.698</td>
<td>0.5844</td>
<td>1.3447</td>
<td>1.6085</td>
<td>1.0953</td>
<td>0.8322</td>
<td>0.6832</td>
<td>0.6832</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EAB</td>
<td>30</td>
<td>1.1821</td>
<td>0.7246</td>
<td>0.5356</td>
<td>0.4496</td>
<td>1.1821</td>
<td>1.4129</td>
<td>0.8667</td>
<td>0.6313</td>
<td>0.5151</td>
<td>0.5151</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.9387</td>
<td>0.5041</td>
<td>0.3709</td>
<td>0.3201</td>
<td>0.9387</td>
<td>1.1268</td>
<td>0.6063</td>
<td>0.4311</td>
<td>0.3581</td>
<td>0.3581</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5: Simulated MSE and absolute bias for the Bayes and PRGM predictors over $\Gamma_{\epsilon}$ (losing weight data).

<table>
<thead>
<tr>
<th>$\Gamma_{\epsilon}$</th>
<th>$\pi_0 \sim N(6, 0.3)$</th>
<th>$q \sim N(5, 0.2)$</th>
<th>$n$</th>
<th>$\epsilon$</th>
<th>0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta^{PRGM}$</td>
<td>1.284</td>
<td>0.894</td>
<td>0.755</td>
<td>0.623</td>
<td>0.508</td>
<td>0.314</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^\pi$</td>
<td>1.284</td>
<td>0.741</td>
<td>0.645</td>
<td>0.592</td>
<td>0.567</td>
<td>0.545</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMSE</td>
<td>0.833</td>
<td>0.529</td>
<td>0.425</td>
<td>0.354</td>
<td>0.289</td>
<td>0.218</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^{PRGM}$</td>
<td>0.833</td>
<td>0.484</td>
<td>0.429</td>
<td>0.411</td>
<td>0.396</td>
<td>0.393</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^\pi$</td>
<td>0.432</td>
<td>0.244</td>
<td>0.195</td>
<td>0.175</td>
<td>0.149</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.097</td>
<td>0.83</td>
<td>0.738</td>
<td>0.655</td>
<td>0.578</td>
<td>0.448</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EAB</td>
<td>0.869</td>
<td>0.613</td>
<td>0.533</td>
<td>0.479</td>
<td>0.429</td>
<td>0.374</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.605</td>
<td>0.410</td>
<td>0.353</td>
<td>0.332</td>
<td>0.306</td>
<td>0.287</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^{PRGM}$</td>
<td>0.605</td>
<td>0.452</td>
<td>0.439</td>
<td>0.433</td>
<td>0.429</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^\pi$</td>
<td>0.256</td>
<td>0.223</td>
<td>0.194</td>
<td>0.179</td>
<td>0.153</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Gamma_{\epsilon}$</td>
<td>1.284</td>
<td>0.894</td>
<td>0.755</td>
<td>0.623</td>
<td>0.508</td>
<td>0.314</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^\pi$</td>
<td>1.284</td>
<td>0.741</td>
<td>0.645</td>
<td>0.592</td>
<td>0.567</td>
<td>0.545</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EMSE</td>
<td>0.833</td>
<td>0.529</td>
<td>0.425</td>
<td>0.354</td>
<td>0.289</td>
<td>0.218</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^{PRGM}$</td>
<td>0.833</td>
<td>0.484</td>
<td>0.429</td>
<td>0.411</td>
<td>0.396</td>
<td>0.393</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^\pi$</td>
<td>0.432</td>
<td>0.244</td>
<td>0.195</td>
<td>0.175</td>
<td>0.149</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.097</td>
<td>0.83</td>
<td>0.738</td>
<td>0.655</td>
<td>0.578</td>
<td>0.448</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EAB</td>
<td>0.869</td>
<td>0.613</td>
<td>0.533</td>
<td>0.479</td>
<td>0.429</td>
<td>0.374</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>0.605</td>
<td>0.410</td>
<td>0.353</td>
<td>0.332</td>
<td>0.306</td>
<td>0.287</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^{PRGM}$</td>
<td>0.605</td>
<td>0.452</td>
<td>0.439</td>
<td>0.433</td>
<td>0.429</td>
<td>0.43</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^\pi$</td>
<td>0.256</td>
<td>0.223</td>
<td>0.194</td>
<td>0.179</td>
<td>0.153</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 6: Simulated MSE and absolute bias for the Bayes and PRGM predictors over $\Gamma$ (losing weight data) continued.

<table>
<thead>
<tr>
<th>$\Gamma^2$</th>
<th>$\pi_0 \sim N(6,0.3)$</th>
<th>q $\sim N(8,0.3)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>$\epsilon$</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>$\delta^{PRGM}$</td>
<td>1.284</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>1.284</td>
</tr>
<tr>
<td>EMSE</td>
<td>30</td>
<td>$\delta^{PRGM}$</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>0.833</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$\delta^{PRGM}$</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>0.432</td>
</tr>
<tr>
<td>20</td>
<td>$\delta^{PRGM}$</td>
<td>1.097</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>1.097</td>
</tr>
<tr>
<td>EAB</td>
<td>30</td>
<td>$\delta^{PRGM}$</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>0.869</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$\delta^{PRGM}$</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>0.605</td>
</tr>
<tr>
<td>20</td>
<td>$\delta^{PRGM}$</td>
<td>0.819</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>0.819</td>
</tr>
<tr>
<td>EMSE</td>
<td>30</td>
<td>$\delta^{PRGM}$</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>0.504</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$\delta^{PRGM}$</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>0.256</td>
</tr>
<tr>
<td>20</td>
<td>$\delta^{PRGM}$</td>
<td>0.836</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>0.836</td>
</tr>
<tr>
<td>EAB</td>
<td>30</td>
<td>$\delta^{PRGM}$</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>0.634</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$\delta^{PRGM}$</td>
</tr>
<tr>
<td></td>
<td>$\epsilon$</td>
<td>0.434</td>
</tr>
</tbody>
</table>
5.2 Predicting the Variance of Weight Loss

For prediction of finite population variance, we consider a single prior \( \Gamma(\alpha_0 = 10, \beta_0 = 3) \) as well as two classes of prior distributions \( \Gamma_\beta = \{ \Gamma(\alpha, \beta) : \beta \in [1, 7] \subseteq \mathbb{R}^+ \} \) and \( \Gamma_\alpha = \{ \Gamma(\alpha, \beta_0) : \alpha \in [4, 10] \subseteq \mathbb{R}^+ \} \) for \( \sigma^2 \). We obtain the Bayes predictor (\( \delta_{\pi_{\alpha_0, \beta_0}} \) with \( \alpha_0 = 10 \) and \( \beta_0 = 3 \)), the PRGM predictor over the class \( \Gamma_\alpha (\delta_{\text{PRGM}}^{\Gamma_\alpha}) \), and the PRGM predictor over the class \( \Gamma_\beta (\delta_{\text{PRGM}}^{\Gamma_\beta}) \). Table 7 summarizes the predicted values under the SE loss function for fixed sample size \( n = 50 \). As we observe the PRGM predicted values are closer to the variance of the weight loss, i.e., 6.499586, than their corresponding Bayes predictions. To evaluate the performance of the Bayes and PRGM predictors of the population variance we performed a simulation study similar to what we presented for the mean (and by using variance instead of the mean) and calculated the EMSE and EAB of each PRGM predictor for different sample sizes \( n = 20, 30, \) and \( 50 \) over \( \Gamma_\alpha \) and \( \Gamma_\beta \) classes of priors with \( \alpha_0 = 4, 6, 8, 10, \) and \( \beta_0 = 1, 3, 5, 7 \). We compared the performance of the PRGM predictors of the population variance with their corresponding Bayes predictors with respect to associated inverse gamma prior distributions with \( \alpha_0 = 4, 6, 8, 10, \) and \( \beta_0 = 1, 3, 5, 7 \). The estimated MSE and the bias of each predictor under \( \Gamma_\beta \) and \( \Gamma_\alpha \), presented in Tables 8, and 9. From Table 8, we observe that the performance of the PRGM predictors with respect to the Bayes predictor are quite satisfactory in terms of the estimated MSE as well as the associated bias for small values of \( \beta_0 \) (\( \beta_0 = 1, 3 \)) and all values of \( \alpha \). But we have quite the opposite results for moderate to large values of \( \beta_0 \) (\( \beta_0 = 5, 7 \)). Note that the MSE and the bias decrease as the sample size increases. From Table 9, we observe that for all values of \( \beta_0 \) and large (small to moderate) values of \( \alpha_0 \), the PRGM predictors (the Bayes predictors) are preferred to the Bayes predictors (the PRGM predictors) in terms of the estimated MSE and bias. Also, the MSE and the bias decrease as the sample size increases.

6 Discussion

In this paper, we studied the PRGM prediction of population parameters under general classes of loss functions and prior distributions. In particular, we studied the PRGM prediction in finite populations and developed a unified approach to calculate these predictions under a very general setting. Under two different normal superpopulation models and different classes of prior distributions on the parameter of the
Table 8: Simulated MSE and absolute bias for the Bayes and PRGM predictors of finite population variance for $\alpha = 4, 6, 8, 10$, $\beta = 1, 3, 5, 7$ and $\beta \in [1,7]$ over $\Pi_\beta$ (losing weight data).

<table>
<thead>
<tr>
<th>$\alpha_0$</th>
<th>$\beta_0 = 1$</th>
<th>$\beta_0 = 3$</th>
<th>$\beta_0 = 5$</th>
<th>$\beta_0 = 7$</th>
<th>$\beta_0 = 8$</th>
<th>$\beta_0 = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 20$</td>
<td>4.496</td>
<td>4.074</td>
<td>3.699</td>
<td>3.372</td>
<td>3.881</td>
<td>4.074</td>
</tr>
<tr>
<td></td>
<td>3.699</td>
<td>4.074</td>
<td>4.496</td>
<td>5.219</td>
<td>5.850</td>
<td>6.573</td>
</tr>
<tr>
<td></td>
<td>1.611</td>
<td>1.529</td>
<td>1.455</td>
<td>1.391</td>
<td>1.491</td>
<td>1.529</td>
</tr>
<tr>
<td>$n = 30$</td>
<td>2.888</td>
<td>2.675</td>
<td>2.485</td>
<td>2.318</td>
<td>2.577</td>
<td>2.485</td>
</tr>
<tr>
<td>EMSE</td>
<td>5.219</td>
<td>4.875</td>
<td>4.546</td>
<td>4.232</td>
<td>4.708</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.048</td>
<td>1.969</td>
<td>1.891</td>
<td>1.814</td>
<td>1.929</td>
<td></td>
</tr>
<tr>
<td>$n = 50$</td>
<td>1.804</td>
<td>1.706</td>
<td>1.616</td>
<td>1.532</td>
<td>1.660</td>
<td>1.532</td>
</tr>
<tr>
<td>EAB</td>
<td>2.685</td>
<td>2.535</td>
<td>2.391</td>
<td>2.255</td>
<td>2.462</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.419</td>
<td>1.372</td>
<td>1.326</td>
<td>1.281</td>
<td>1.349</td>
<td></td>
</tr>
<tr>
<td>$\alpha_0 = 6$</td>
<td>6.284</td>
<td>5.738</td>
<td>5.229</td>
<td>4.754</td>
<td>5.479</td>
<td>4.754</td>
</tr>
<tr>
<td></td>
<td>3.933</td>
<td>3.637</td>
<td>3.361</td>
<td>3.103</td>
<td>3.497</td>
<td>3.103</td>
</tr>
<tr>
<td></td>
<td>2.067</td>
<td>1.946</td>
<td>1.832</td>
<td>1.726</td>
<td>1.888</td>
<td>1.726</td>
</tr>
<tr>
<td>$n = 30$</td>
<td>2.225</td>
<td>2.112</td>
<td>2.002</td>
<td>1.897</td>
<td>2.057</td>
<td>1.897</td>
</tr>
<tr>
<td>EMSE</td>
<td>6.584</td>
<td>6.210</td>
<td>5.850</td>
<td>5.503</td>
<td>6.028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.399</td>
<td>3.227</td>
<td>3.061</td>
<td>2.902</td>
<td>3.144</td>
<td></td>
</tr>
<tr>
<td>$n = 50$</td>
<td>2.225</td>
<td>2.112</td>
<td>2.002</td>
<td>1.897</td>
<td>2.057</td>
<td>1.897</td>
</tr>
<tr>
<td>EAB</td>
<td>3.026</td>
<td>2.923</td>
<td>2.821</td>
<td>2.720</td>
<td>2.872</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.644</td>
<td>1.595</td>
<td>1.547</td>
<td>1.499</td>
<td>1.571</td>
<td></td>
</tr>
</tbody>
</table>

Table 9: Simulated MSE and absolute bias for the Bayes and PRGM predictors of finite population variance for $\beta_0 = 1, 3, 5, 7$, $\alpha_0 = 4, 6, 8, 10$ and $\alpha \in [4,10]$ over $\Pi_\alpha$ (losing weight data).

<table>
<thead>
<tr>
<th>$\beta_0 = 1$</th>
<th>$\alpha_0 = 4$</th>
<th>$\alpha_0 = 6$</th>
<th>$\alpha_0 = 8$</th>
<th>$\alpha_0 = 10$</th>
<th>$\beta_0 = 1$</th>
<th>$\alpha_0 = 4$</th>
<th>$\alpha_0 = 6$</th>
<th>$\alpha_0 = 8$</th>
<th>$\alpha_0 = 10$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1.611</td>
<td>2.067</td>
<td>2.685</td>
<td>3.399</td>
<td>1.611</td>
<td>2.067</td>
<td>2.685</td>
<td>3.399</td>
<td>1.611</td>
</tr>
<tr>
<td>$n = 30$</td>
<td>2.888</td>
<td>3.933</td>
<td>5.219</td>
<td>6.584</td>
<td>2.888</td>
<td>3.933</td>
<td>5.219</td>
<td>6.584</td>
<td>2.888</td>
</tr>
<tr>
<td></td>
<td>1.418</td>
<td>1.715</td>
<td>2.048</td>
<td>2.370</td>
<td>1.418</td>
<td>1.715</td>
<td>2.048</td>
<td>2.370</td>
<td>1.418</td>
</tr>
<tr>
<td>$n = 50$</td>
<td>1.804</td>
<td>2.225</td>
<td>2.653</td>
<td>3.026</td>
<td>1.804</td>
<td>2.225</td>
<td>2.653</td>
<td>3.026</td>
<td>1.804</td>
</tr>
<tr>
<td>EAB</td>
<td>2.685</td>
<td>2.535</td>
<td>2.391</td>
<td>2.255</td>
<td>2.685</td>
<td>2.535</td>
<td>2.391</td>
<td>2.255</td>
<td>2.685</td>
</tr>
<tr>
<td></td>
<td>1.419</td>
<td>1.372</td>
<td>1.326</td>
<td>1.281</td>
<td>1.419</td>
<td>1.372</td>
<td>1.326</td>
<td>1.281</td>
<td>1.419</td>
</tr>
<tr>
<td>$\beta_0 = 3$</td>
<td>4.074</td>
<td>5.738</td>
<td>7.690</td>
<td>9.637</td>
<td>4.074</td>
<td>5.738</td>
<td>7.690</td>
<td>9.637</td>
<td>4.074</td>
</tr>
<tr>
<td></td>
<td>1.529</td>
<td>1.946</td>
<td>2.535</td>
<td>3.227</td>
<td>1.529</td>
<td>1.946</td>
<td>2.535</td>
<td>3.227</td>
<td>1.529</td>
</tr>
<tr>
<td>$n = 20$</td>
<td>2.048</td>
<td>2.675</td>
<td>3.637</td>
<td>4.875</td>
<td>2.048</td>
<td>2.675</td>
<td>3.637</td>
<td>4.875</td>
<td>2.048</td>
</tr>
<tr>
<td>EMSE</td>
<td>2.318</td>
<td>3.103</td>
<td>4.232</td>
<td>5.503</td>
<td>2.318</td>
<td>3.103</td>
<td>4.232</td>
<td>5.503</td>
<td>2.318</td>
</tr>
<tr>
<td></td>
<td>1.358</td>
<td>1.639</td>
<td>1.969</td>
<td>2.292</td>
<td>1.358</td>
<td>1.639</td>
<td>1.969</td>
<td>2.292</td>
<td>1.358</td>
</tr>
<tr>
<td>$n = 30$</td>
<td>1.706</td>
<td>2.112</td>
<td>2.542</td>
<td>2.923</td>
<td>1.706</td>
<td>2.112</td>
<td>2.542</td>
<td>2.923</td>
<td>1.706</td>
</tr>
<tr>
<td>EAB</td>
<td>1.254</td>
<td>1.497</td>
<td>1.814</td>
<td>2.138</td>
<td>1.254</td>
<td>1.497</td>
<td>1.814</td>
<td>2.138</td>
<td>1.254</td>
</tr>
<tr>
<td></td>
<td>0.962</td>
<td>1.091</td>
<td>1.281</td>
<td>1.499</td>
<td>0.962</td>
<td>1.091</td>
<td>1.281</td>
<td>1.499</td>
<td>0.962</td>
</tr>
</tbody>
</table>
underlying superpopulation model, we obtained the PRGM predictors of the finite population mean under the LINEX and SE loss functions. Furthermore, we obtained the PRGM predictor of the finite population variance under the SE loss function in a normal superpopulation model. We also considered a non-normal model and derived the Bayes and PRGM predictors of the finite population mean under the SE loss function. We also applied the results to different real data sets to illustrate practical utility of the Bayes and PRGM procedures. We provided real-life data to predict finite population means and variances under normal and non-normal models. We compared the estimated risk and bias of the obtained predictors under the SE loss function by some simulation studies. Although in some cases Bayes predictors have smaller risks and biased than robust Bayes predictors, but it is recommended to use the robust Bayes predictors due to lack of confidence in \( \delta^\pi \) under \( \pi = \pi(\mu_0, \tau_0^2) \), especially if it is hard to specify a single prior distribution for the parameters of the underlying superpopulation models.

The proposed methodology in this paper studied the impact of the prior distribution as an input to the Bayesian prediction process on the predicted values when prior distribution ranges in certain classes of priors. If the impact is considerable, there is sensitivity and one should use robust Bayesian prediction as a solution that is relatively insensitive to the uncertainty in the prior distribution elicitation. It is worth noting that Bayesian analysis depends on other subjective inputs such as the loss function and/or the model. A future research direction would be to study sensitivity of Bayesian prediction jointly with respect to the prior and the loss function. Another important research direction is to study the effect of model-misspecification or imprecise probability models on Bayesian prediction of the parameters of interest. One can also use other classes of prior distributions to obtain robust Bayesian analysis, such as classes of priors with given marginals when we are dealing with multi-parameter cases, classes of \( \epsilon \)-contaminated priors with shape constrains, or generalized moment classes of priors where one considers classes of prior distributions that satisfy some moment conditions with a priori specified moments (e.g., Berger (1990)).

Acknowledgement

Authors would like to thank Professor William Strawderman from the Rutgers University for useful discussion and his help regarding the proof of Lemma 1.5. We would also like to thank an Associate Editor and three anonymous referees for their constructive comments which improved the quality and presentation of our results. Mohammad Jafari Jozani gratefully acknowledges the research support of NSERC Canada. The research of Nader Nematollahi is supported by the Research Council of the Allameh Tabataba’i University and part of this work was done during his visit to the University of Manitoba, Department of Statistics.
References


