# Bayes Linear Estimation for Finite Population with Emphasis to Categorical Data

Gonçalves, Kelly
*Instituto de Matemática, Departamento de Estatística*
*Av. Athos da Silveira Ramos 149, Centro de Tecnologia - Bloco C*
*21941-909 Rio de Janeiro, Brasil*
*E-mail: kelly@dme.ufrj.br*

Moura, Fernando
*Instituto de Matemática, Departamento de Estatística*
*Av. Athos da Silveira Ramos 149, Centro de Tecnologia - Bloco C*
*21941-909 Rio de Janeiro, Brasil*
*E-mail: fmoura@im.ufrj.br*

Migon, Helio
*Instituto de Matemática, Departamento de Estatística*
*Av. Athos da Silveira Ramos 149, Centro de Tecnologia - Bloco C*
*21941-909 Rio de Janeiro, Brasil*
*E-mail: migon@im.ufrj.br*

## Introduction

The designed-based sample survey theory has been very appealed to official statistics agencies around all over the world. As pointed out by Skinner et al. (1989), the main reason for this is that it is essentially distribution-free. In some specific situations, such as in small domains estimation and in the presence of the non-response, the designed-based approach proved to be inefficient, providing inadequate predictors. Supporters of the designed-based approach argue that model-based inference much depends on the model assumptions which might not be true. On the other hand, interval inference for target population parameters relies on the Central Limit theorem, which could not be applied in many practical situations, where the sample size is not large enough and/or independent assumptions of the random variables involved are not realistic.

Basu (1971) did not accept estimates of population quantities that depend on the sampling rule, like the inclusion probabilities. He argued that this estimation procedure does not satisfy the likelihood principle, to which he was adept to. The question that poses is: is it possible to conciliate both approaches? In the superpopulation model context, Zacks (2002) shows that some designed-based estimators could be recovered by using a general regression model approach.

In the Bayesian context, another appealing proposal of conciliating designed-based and model-based approach was presented by O'Hagan (1985) in an unpublished report. O'Hagan (1985)'s approach is based upon the Bayes Linear Estimator (BLE), therefore distribution-free. This methodology is an alternative to the methods of randomization and appears midway between two extreme views: on the one hand the procedures based on randomization and on the other based on superpopulation models. His model formulation assumes only second order exchangeability, which in practice means the need of stating first and second moments only, describing a prior knowledge about the structures present in the population. He dealt with several population structures, such as stratification and clustering by assuming suitable hypotheses about the first and second moments and showed how some common designed-based estimators could be obtained.

In this work we apply the Bayes linear method to a general regression framework, and so all the estimators presented in O'Hagan (1985) can be obtained as particular cases of it. We also present a Bayes linear approach for obtaining estimation of proportions for categorical data, including discussions about the prior elicitation. We illustrate it with a simple numeric example.

## Bayes Linear Estimation for Finite Population

Bayes methods have been receiving much attention and support from practical and theoretical points of view. However, Goldstein and Wooff (2007) argues that as the complexity of the problem increases, our actual ability to fully specify the prior or the sampling model is impaired. He concludes that in such situations, there is a need to develop methods based on partial belief specification. One of this methodologies, termed Bayes Linear, is exploit in this article and is described in this section.

Let $\mathbf{y}_s$ be the vector of observations and $\boldsymbol{\theta}$ the parameter to be estimated. Let us suppose that the joint distribution of $\boldsymbol{\theta}$ and $\mathbf{y}_s$ is partially specified by only their first two moments:

$$(1) \qquad \begin{pmatrix} \boldsymbol{\theta} \\ \mathbf{y}_s \end{pmatrix} \sim \left[ \begin{pmatrix} \mathbf{a} \\ \mathbf{f} \end{pmatrix}, \begin{pmatrix} \mathbf{R} & \mathbf{AQ} \\ \mathbf{QA}' & \mathbf{Q} \end{pmatrix} \right],$$

where $\mathbf{a}$ and $\mathbf{f}$ denote mean vectors and $\mathbf{R}$, $\mathbf{AQ}$, $\mathbf{QA}'$ and $\mathbf{Q}$ the covariance matrix elements, respectively, of $\boldsymbol{\theta}$ and $\mathbf{y}_s$.

For each value of $\boldsymbol{\theta}$ and each possible estimate $\mathbf{d}$, we associate a quadratic loss function $L(\boldsymbol{\theta}, \mathbf{d}) = (\boldsymbol{\theta} - \mathbf{d})'(\boldsymbol{\theta} - \mathbf{d}) = tr(\boldsymbol{\theta} - \mathbf{d})(\boldsymbol{\theta} - \mathbf{d})'$. The BLE of $\boldsymbol{\theta}$ is the value of $\mathbf{d}$ which minimizes $L(\boldsymbol{\theta}, \mathbf{d})$ within the class of linear estimates of the form $\mathbf{d} = \mathbf{d}(\mathbf{y}_s) = \mathbf{h} + \mathbf{H}\mathbf{y}_s$, for some vector $\mathbf{h}$ and matrix $\mathbf{H}$. Thus, the BLE of $\boldsymbol{\theta}$, $\mathbf{d}^*$, and its associated risk matrix, $V(\mathbf{d}^*)$, are respectively given by:

$$(2) \qquad \begin{aligned} \mathbf{d}^* &= \mathbf{a} + \mathbf{A}(\mathbf{y}_s - \mathbf{f}), \\ V(\mathbf{d}^*) &= \mathbf{R} - \mathbf{AQA}'. \end{aligned}$$

In the finite population context, the vector values of a characteristic of interest $\mathbf{y} = (y_1, ..., y_N)'$ can be split up into two parts: $\mathbf{y} = (\mathbf{y}_s, \mathbf{y}_{\bar{s}})$; where $\mathbf{y}_s$ is the known part provided by the sample data and $\mathbf{y}_{\bar{s}}$ is the unknown part. A general problem is to predict a function of the vector $\mathbf{y}$, such as the total $T = \sum_{i=1}^{N} y_i = \mathbf{1}'_s \mathbf{y}_s + \mathbf{1}'_{\bar{s}} \mathbf{y}_{\bar{s}}$, given the sample data $\mathbf{y}_{\bar{s}}$. In many practical situations the relation between the response variable and a set of auxiliary variables can be represented by a regression model. In the next section, several estimators suitable for applications in many sampling contexts are obtained as particular cases.

## General Linear Regression Model

We consider the following robust regression model for finite population prediction purposes:

$$(3) \qquad \begin{aligned} \mathbf{y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \ \boldsymbol{\epsilon} \sim [\mathbf{0}, \mathbf{V}], \\ \boldsymbol{\beta} &\sim [\mathbf{a}, \mathbf{R}], \end{aligned}$$

where $\boldsymbol{\beta}$ is a p-vector of parameters and $\boldsymbol{\epsilon}$ is a random N-vector with mean $\mathbf{0}$ and covariance $\mathbf{V}$.

As the response vector $\mathbf{y}$ is partitioned into $\mathbf{y}_s$ and $\mathbf{y}_{\bar{s}}$, the matrix of auxiliary variables, $\mathbf{X}$, is analogously partitioned into $\mathbf{X}_s$ and $\mathbf{X}_{\bar{s}}$, as well as $\mathbf{V}$, which is respectively partitioned into $\mathbf{V}_s$ and $\mathbf{V}_{\bar{s}}$. The first aim is to predict $\mathbf{y}_{\bar{s}}$ and then the total $T$. We do it in two steps detailed below.

After we have observed $\mathbf{y}_s$, we obtain $\boldsymbol{\beta}^*$ via Bayes Linear method from the population model (3) applied to the sample data, adapting the structure (1) and using the results in (2). The BLE of $\boldsymbol{\beta}$ and its measure of dispersion are, respectively, given by:

$$(4) \qquad \begin{aligned} \boldsymbol{\beta}^* &= \mathbf{a} + \mathbf{RX}'_s (\mathbf{X}_s \mathbf{RX}'_s + \mathbf{V}_s)^{-1} (\mathbf{y}_s - \mathbf{X}_s \mathbf{a}), \\ V(\boldsymbol{\beta}^*) &= \mathbf{C} = \mathbf{R} - \mathbf{RX}'_s (\mathbf{X}_s \mathbf{RX}'_s + \mathbf{V}_s)^{-1} \mathbf{X}_s \mathbf{R}. \end{aligned}$$

From the following equations: $\mathbf{C}^{-1} = \mathbf{R}^{-1} + \mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s$ and $\mathbf{A} = \mathbf{R}\mathbf{X}_s'\mathbf{Q}^{-1} = \mathbf{C}\mathbf{X}_s'\mathbf{V}_s^{-1}$, where $\mathbf{Q} = \mathbf{X}_s\mathbf{R}\mathbf{X}_s' + \mathbf{V}_s$, we can rewrite (4) as:

$$\boldsymbol{\beta}^* = \mathbf{C}\left(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s + \mathbf{R}^{-1}\mathbf{a}\right).$$

If we place a vague prior distribution on $\boldsymbol{\beta}$, taking $\mathbf{R}^{-1} \to 0$, we obtain minimum least square estimator of $\boldsymbol{\beta}$, given by $\hat{\boldsymbol{\beta}} = \left(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s\right)^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s$.

The second step is carried out by using some properties of conditional expectations to obtain:

$$\begin{aligned}\hat{E}\left[\mathbf{y}_{\bar{s}} \mid \mathbf{y}_s\right] &= E\left[\mathbf{X}_{\bar{s}}\boldsymbol{\beta} \mid \mathbf{y}_s\right] = \mathbf{X}_{\bar{s}}\boldsymbol{\beta}^* \\ \hat{V}\left[\mathbf{y}_{\bar{s}} \mid \mathbf{y}_s\right] &= \mathbf{V}_{\bar{s}} + V\left[\mathbf{X}_{\bar{s}}\boldsymbol{\beta} \mid \mathbf{y}_s\right] = \mathbf{V}_{\bar{s}} + \mathbf{X}_{\bar{s}}\mathbf{C}\mathbf{X}_{\bar{s}}'.\end{aligned}$$

Thus, the BLE for the total $T$ and its associate variance are respectively given by:

$$\begin{aligned}T^* &= \mathbf{1}_s'\mathbf{y}_s + \mathbf{1}_{\bar{s}}'\mathbf{X}_{\bar{s}}\boldsymbol{\beta}^*, \\ V(T^*) &= \mathbf{1}_{\bar{s}}'\left[\mathbf{V}_{\bar{s}} + \mathbf{X}_{\bar{s}}\mathbf{C}\mathbf{X}_{\bar{s}}'\right]\mathbf{1}_{\bar{s}}.\end{aligned} \tag{5}$$

It is possible to obtain as particular cases , several estimators for different population structures, including stratification, clustering and binary data, which was obtained by O'Hagan (1985). The examples below ilustrates it for the case analogous to the simple random sampling. The first is described in O'Hagan (1985), but the second is a original contribution.

**Example 1: Fully Exchangeable**

The case where the population has no relevant structure can be done by imposing:

$$E(y_i) = m,\ V(y_i) = v \text{ and } Cov(y_i, y_j) = c,\ i, j = 1, \ldots, N,\ \forall i \neq j. \tag{6}$$

Applying the general result established in (5) to (6) with $\mathbf{X} = \mathbf{1}$, $\mathbf{a} = m$, $\mathbf{R} = c$ and $\mathbf{V} = \sigma^2\mathbf{I}$, where $\sigma^2 = v - c$, we obtain the BLE of $T$ and its associated variance:

$$\begin{aligned}T_{srs}^* &= n\bar{y}_s + (N-n)\hat{\mu}, \\ V(T_{srs}^*) &= (N-n)\sigma^2 + (N-n)^2 c\,\sigma^2(\sigma^2 + nc)^{-1},\end{aligned} \tag{7}$$

where

$$\begin{aligned}\bar{y}_s &= n^{-1}\mathbf{1}_s'\mathbf{y}_s \text{ is the sample mean,} \\ \hat{\mu} &= \omega\bar{y}_s + (1-\omega)m \text{ is the expected value of } \mathbf{y}_{\bar{s}}, \\ \omega &= \frac{n\sigma^{-2}}{c^{-1} + n\sigma^{-2}}, \text{where } \ \sigma^2 = v - c.\end{aligned}$$

It should be noted that $\hat{\mu}$ is a weighted average of the prior mean $m$ and the sample mean $\bar{y}_s$, where $\omega$ is the ratio between two population quantities, which their interpretations are given bellow.

The mean $m$ can be ultimately viewed as the investigator's prior of the true population mean $\bar{y}$. The uncertainty about $y_i$ is split up into two components: the uncertainty about the overall level of the $y_i$'s (between variation) and the one with respect to how much each $y_i$ may vary from that overall level (within variation). A useful measure of variability of units within the population is given by $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2$. It is not difficult to show that $E(S^2) = v - c = \sigma^2$. Therefore, $\sigma^2$ can be interpreted as a prior estimate of variability within the population. We also obtain $V(\bar{y}) = c + N^{-1}\sigma^2$. In many applications $N$ is large and thus the constant $c$ could be viewed as the between variation.

Letting $v \to \infty$ and keeping $\sigma^2$ fixed, that is, assuming prior ignorance, the estimates in (7) yield to the total estimates in the designed-based context for the simple random sampling case: $T_{srs}^* = N\bar{y}_s$ and $V(T_{srs}^*) = N^2\left(1 - \frac{n}{N}\right)\frac{\sigma^2}{n}$.

**Example 2: Ratio Estimator**

When there is an auxiliary variable and it is reasonable to assume that the relationship between the response variable and the explanatory variable can be modeled by a linear regression through the origin, we use the ratio type estimator.

The hypothesis of exchangeability is applied to the rate $y_i/x_i$ as follows:

$$(8) \qquad E\left(\frac{y_i}{x_i}\right) = m,\ Var\left(\frac{y_i}{x_i}\right) = v \text{ and } Cov\left(\frac{y_i}{x_i}, \frac{y_j}{x_j}\right) = c,\ i,j = 1,\ldots,N,\ \forall i \neq j.$$

Applying the general result established in (5) to (8) with $\mathbf{X} = (x_1,\ldots,x_N)'$, $\mathbf{a} = m$, $\mathbf{R} = c$ and $\mathbf{V} = \sigma^2 diag(x_1,\ldots,x_N)$, where $\sigma^2 = v - c$, we obtain:

$$T^*_{ra} = n\bar{y}_s + (N-n)\hat{\mu}\bar{x}_{\bar{s}},$$

where

$$\hat{\mu} = \omega\frac{\bar{y}_s}{\bar{x}_s} + (1-\omega)m, \text{ is the expected values of } y_{\bar{s}},$$

$$\omega = \frac{\sigma^{-2}n\bar{x}_s}{(c^{-1} + \sigma^{-2}n\bar{x}_s)}.$$

Letting $v \to \infty$ and $n \to \infty$, but keeping $\sigma^2$ fixed, we recover the ratio type estimator, found in designed-based approach: $T^*_{ra} = N\bar{x}(\bar{y}_s/\bar{x}_s)$.

**Bayes Linear for Categorical Data**

Often we may be interested in cases where the observed characteristic is wether or not the population unit possesses some attribute. The estimation of a proportion is a special case of estimating a population mean. BLE for binary data was briefly introduced by O'Hagan (1985).

However, we developed, more generally, a model for when we are interested in analyzing more than one attribute in a population. Let $y_{ij}$ be the attribute variable such that $y_{ij} = 1$, if $i$-th unit has $j$-th attribute; and $y_{ij} = 0$, otherwise.

The interest is to estimate $\mathbf{p} = (p_1,\ldots,p_k)'$, where $p_j = N^{-1}\sum_{i=1}^{N} y_{ij}$, $j = 1,\ldots,k$, given $\mathbf{y}_s = (y_{11},\cdots,y_{n1},\cdots,y_{1k},\cdots,y_{nk})'$. But in such cases we usually do not have all the data, but some statistics, such as the sample proportion. Thus, the estimator and its variance are obtained as:

$$(9) \qquad \begin{aligned} \mathbf{p}^* &= \tfrac{n\bar{\mathbf{y}}_s + (N-n)E(\bar{\mathbf{y}}_{\bar{s}}|\bar{\mathbf{y}}_s)}{N}, \\ V(\mathbf{p}^*) &= \tfrac{(N-n)^2 V(\bar{\mathbf{y}}_{\bar{s}}|\bar{\mathbf{y}}_s)}{N^2}, \end{aligned}$$

where $\bar{\mathbf{y}}_s$ is a $k$-vector whose $j$-th position is given by the sample mean for category $j$, analogously we have in $\bar{\mathbf{y}}_{\bar{s}}$.

As we are dealing with situations in which for each unit is only possible to associate an attribute, so we have $\sum_{j=1}^{k} p_j = 1$. Thus, it follows that $p^*_k = 1 - \sum_{j=1}^{k-1} p^*_j$.

In the absence of any other structural information we suppose that the units in any given category are second-order exchangeable, but we do not assume any exchangeability between units of different categories. Our prior beliefs are represented, for $i = 1,\ldots,N$, $j = 1,\ldots,k-1$, as follows:

$$(10) \qquad \begin{aligned} m_j &= E(y_{ij}) = P(y_{ij} = 1), v_j = Var(y_{ij}) = m_j(1-m_j), \\ c_j &= Cov(y_{ij}, y_{i'j}) = m_j(m_{jj} - m_j) \text{ and } \sigma^2_j = v_j - c_j = m_j(1-m_{jj}), \end{aligned}$$

where $m_{jj} = P(y_{i'j} = 1 \mid y_{ij} = 1)$, for all $i \neq i'$.

For $j \neq j'$ we define the covariance between these categories as

$$(11) \qquad d_{jj'} = Cov(y_{ij}, y_{i'j'}) = \begin{cases} m_j(m_{j'j} - m_{j'}), & \text{if } i \neq i', \\ -m_j m_{j'}, & \text{if } i = i'. \end{cases}$$

Using the general notation in (3), let $\boldsymbol{\theta}$ be a vector of dimension $k-1$, $\mathbf{X}_s = \mathbf{I}_s$, and the equations in (10) and (11) can be compactly written as

$$\begin{aligned} E(\bar{\mathbf{y}}_s) &= \mathbf{X}_s\mathbf{a} = \mathbf{X}_s(m_1, \ldots, m_{k-1})', \\ Var(\bar{\mathbf{y}}_s) &= \begin{pmatrix} Q_{11} & \ldots & Q_{1\,k-1} \\ \vdots & \ddots & \vdots \\ Q_{k-1\,1} & \ldots & Q_{k-1\,k-1} \end{pmatrix} = \mathbf{X}_s\mathbf{R}\mathbf{X}_s' + \mathbf{V}_s, \end{aligned}$$

where $Q_{jj} = c_j + \sigma_j^2/n$ and $Q_{jj'} = d_{jj'} - m_j m_{j'j}/n$, for $d_{jj'}$ when $i \neq i'$.

Then, for $i \neq i'$ we conclude that

$$\mathbf{R} = \begin{pmatrix} c_1 & d_{12} & \ldots & d_{1\,k-1} \\ d_{21} & c_2 & \ldots & d_{2\,k-1} \\ \vdots & \vdots & \ddots & \vdots \\ d_{k-1\,1} & d_{k-1\,2} & \ldots & c_{k-1} \end{pmatrix}, \quad \mathbf{V}_s = \frac{1}{n}\begin{pmatrix} \sigma_1^2 & -m_1 m_{21} & \ldots & -m_1 m_{k-1\,1} \\ -m_2 m_{12} & \sigma_2^2 & \ldots & -m_2 m_{k-1\,2} \\ \vdots & \vdots & \ddots & \vdots \\ -m_{k-1} m_{1\,k-1} & -m_{k-1} m_{2\,k-1} & \ldots & \sigma_{k-1}^2 \end{pmatrix}.$$

So, the estimator is a $k-1$-vector described in (9) with the quantities defined above.

**Prior elicitation**

As $m_j = P(y_{ij} = 1)$ and $m_{jj'} = P(y_{i'j} = 1 \mid y_{ij'} = 1)$, we have $0 < m_j < 1$, $0 < m_{jj'} < 1$, $m_{jj'}m_{j'} = m_{j'j}m_j$, for all $j \neq j'$ and $m_{jj} + m_{j'j} \leq 1$. And hence we get $m_{j'j} = m_{jj'}m_{j'}/m_j$, for all $j < j'$.

But, if an expert have difficulties in specifying some of these conditional probabilities, the prior correlation may help this task. Define $\rho_{jj'}$ as the prior correlation between two units in categories $j$ and $j'$, $j = 1, \ldots, k-1$. Note that, as $-1 < \rho_{jj} < 1$, then $2m_j - 1 < m_{jj} < 1$, and thus using the same argument we get other restrictions for the cases where we have two distinct categories.

Therefore, an alternative is to elicit $\rho_{jj'}$, for all $j, j' = 1, \ldots, k-1$ and deterministically we get

$$\begin{aligned} m_{jj} &= m_j + \rho_{jj}(1 - m_j), \\ m_{jj'} &= \frac{m_j m_{j'} + \rho_{j'j}\sqrt{m_j(1-m_j)m_{j'}(1-m_{j'})}}{m_j'}, \\ m_{j'j} &= m_{jj'}m_{j'}/m_j. \end{aligned}$$

It should be noted that we have to specify $\rho_{jj'}$, for $j, j' = 1, \ldots, k-1$, in order to find $m_{jj}$ and $m_{j'j}$ such that $m_{jj} + m_{j'j} \leq 1$.

If we do not have much prior information on these parameters, we use non-informative priors. To elicit $m_j$, we take $m_j = 0.5$, which results in a large $v_j$, for all $j = 1, \ldots, k-1$. To elicit $m_{jj'}$, we use the fact that $\mathbf{R}$ is the prior covariance matrix of the regression parameters and its diagonal is composed by $c_j$. Therefore, we chose $m_{jj}$ that makes $c_j$ large, for $j = 1, \ldots, k-1$. Thus, we have to take $m_{jj} \approx 1$, and hence we can arbitrarily chose $m_{j'j}$ that satisfies $m_{jj} + m_{j'j} \leq 1$.

**An Illustrative Example**

We considered the estimator for three categories and evaluated the effect of the estimator for the following values of $N$ and $n$: ($N = 10000$, $n = 50$); ($N = 10000$, $n = 200$); ($N = 1000$, $n = 50$); ($N = 1000$, $n = 200$). We fixed $\bar{\mathbf{y}}_s = (0.06, 0.30, 0.64)$ and we used a non-informative priors assuming that: $m_1 = m_2 = 0.5$, $m_{11} = m_{22} = 0.9$ and so $m_{12} = m_{21} = 0.05$.

To construct a credible region we assume that $\mathbf{p}^*$ follows approximately a multivariate normal distribution. However, as we have $p_1 + p_2 + p_3 = 1$ the parametric space is a simplex. So we did the

95% confidence ellipse for each of the two population parameters using the chi-squared statistics with 2 degrees of freedom. The third remain parameter is determined by the restriction.

Figure 1 shows that the point estimates are similar, but the confidence regions shows that when we increase the value of $n$ the precision of the estimates increases too. Because, we assume non-informative priors, the estimates obtained are closer to the sample proportion.

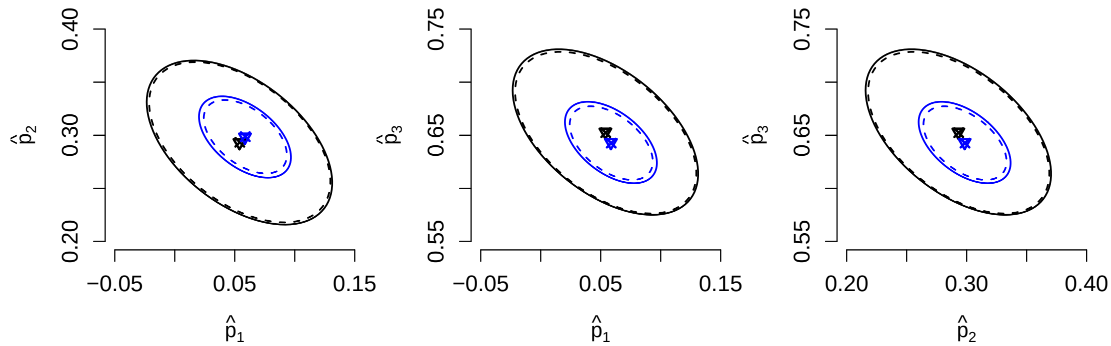


***Figure 1: BLE and the** 95% **simultaneous credible regions for each of the two proportion parameters:** ($N = 10000$, $n = 50$) **- black solid line,** ($N = 10000$, $n = 200$) **- blue solid line,** ($N = 1000$, $n = 50$) **- black dashed line,** ($N = 1000$, $n = 200$) **- blue dashed line.***

## Conclusions

To elicit a complete joint prior distribution in so many dimensions would be an enormous task. The Bayes linear methods only require the elicitation of prior means, variances and covariances for the parameters. This could facilitate the work of a person without statistical background conducting this elicitation. An example of a successful elicitation using this estimators is in O'Hagan (1998).

We derived the designed-based estimators using the structure of the BLE applied to general regression model approach. We extended the estimator to multiple categorical data, and we concluded that even if the estimator has many quantities to elicitate, it is possible to reparametrize them or work with non-informative priors to carry it out.

## REFERENCES


Basu, D. (1971) An essay on the logical foundations of survey sampling, Part 1 (with discussion). In Foundations of Statistical Inference (eds. Godambe and Sprott), 203-242. Holt, Reinhart and Wilnston, Toronto.

Goldstein, M. and Woof, D. (2007) Bayes Linear Statistics: Theory and Methods. Durhan University, UK: Wiley series in probability and statistics.

O'Hagan, A. (1985) Bayes linear estimators for finite populations. Tech. Rep. 58, Department of Statistics - University of Warwick.

O'Hagan, A. (1998) Eliciting expert beliefs in substantial practical applications. The Statistician, 47, 21-35.

Skinner, C., Holt, D. and Smith, T. (1989) Analysis of complex surveys. John Wiley & Sons.

Zacks, S. (2002) In the Footsteps of Basu: The Predictive Modelling Approach to Sampling from Finite Population. Sankhyā: The Indian Journal of Statistics, Series A, 64, 532-544.


## ABSTRACT


*In this work we set up a general regression framework, where all Bayes linear estimators presented in O'Hagan (1985) are obtained as particular cases of it. We also presented a Bayes linear approach for obtaining estimation of proportions for categorical data. We illustrate it with a numeric example.*