ESTADÍSTICA ESPAÑOLA
Vol. 37, Núm. 140, 1995, págs. 309 a 320

Inferencia Bayesiana en Poblaciones Finitas: Un análisis comparativo

por
BEGOÑA FONT BELAIRE
Departamento de Economía Financiera y Matemática
Universidad de Valencia
Edificio Departamental Oriental
Av. Tarongers s/n
Tel: 34-6-382-8366
Fax:34-382-8370
E-mail: Maria.B.Font@uv.es

RESUMEN

En este artículo, se intentará realizar un análisis comparativo sobre algunos resultados de predictores Bayes de la media poblacional bajo función de pérdida cuadrática. Se introducirán para ello dos modelos de regresión bajo la hipótesis de normalidad: un modelo básico y un modelo en dos etapas en dos supuestos: varianzas conocidas y desconocidas, y se aplicará un análisis bayesiano. Para los dos modelos se proporcionarán los estimadores Bayes de la media poblacional y su pérdida cuadrática media asociada.

Palabras clave: Inferencia Bayesiana; poblaciones finitas; predicción; regresión múltiple.

Clasificación AMS: 62F15

1. INTRODUCCIÓN

El objetivo principal de este artículo es la recopilación y estudio comparativo de los modelos aplicados en Inferencia Bayesiana para estimar la media de una característica continua de una Población Finita.

Son muy numerosos los trabajos publicados en esta línea, nos podemos referir a Ericson (1969, 70, 83, 88), Smouse (1982, 84), Murgui (1982), Ghosh y Lahiri (1987), Bolfarine y Zacks (1991, 92), Scott y Smith (1969), Malec y Sedransk (1985), Rodrigues (1988), Royall y Pfeffermann (1982), Bernardo y Girón (1991) entre otros; las hipótesis y criterios usuales propuestos han sido los siguientes:

– Se asume la Población Finita de interés como realización de un determinado modelo de superpoblación sobre el que se asume la hipótesis de normalidad dados los parámetros y una distribución sobre éstos normal si las varianzas son conocidas y normal-gamma si las varianzas no son conocidas.

– Se establecen dos modelizaciones de la población en estudio fundamentalmente, que atienden a una estructura de población monoetápica (la población se supone dividida en elementos) y una estructura de población bietápica (la población se supone dividida en unidades y éstas a su vez en elementos).

– El criterio para seleccionar el estimador, considerando como estimador óptimo áquel que minimice la pérdida cuadrática esperada dada una muestra de la población en estudio. Aunque algunos autores (por ejemplo: Smouse (1982, 84), Ericson (1983, 88), Scott y Smith (1969), Malec y Sedransk (1985)) presentan estimadores óptimos bajo el criterio de minimizar la pérdida cuadrática esperada dada una muestra entre los estimadores lineales de la media debilitando las hipótesis de normalidad que señalábamos antes.

En este artículo presentaré en la sección 2 dos modelos de regresión múltiple bajo la hipótesis de normalidad: un modelo básico (que atiende a una estructura monoetápica de la población) y un modelo de dos etapas (que atiende a una estructura bietápica), en los casos de varianzas conocidas y desconocidas completando los estudios realizados sobre estos modelos, presentando: la distribución predictiva del vector no observado (de la característica de interés de la población) dada la muestra, el estimador de la media poblacional por elemento y su pérdida cuadrática asociada, todo ello en expresiones 'fáciles' de interpretar. En la sección 3 se discutirán brevemente algunas propiedades de los modelos analizados y se realizarán unos breves comentarios sobre las posibles ventajas y dificultades de su aplicabilidad práctica.

2. MODELOS, ESTIMADORES Y ERRORES COMETIDOS AL USAR ESTOS ESTIMADORES

Estudiaremos en esta sección dos modelos Bayesianos de Población Finita.

2.1 Modelo Básico

Supongamos una población de N elementosde modo que asociado a un elemento i tenemos una cantidad de interés y_i que asumiremos realización de una variable aleatoria Y_i y un p-vector $\boldsymbol{X}_i=(x_{i1},\ldots,x_{ip})'$ conocido. Nos planteamos dos situaciones:

1. Modelo Básico con Varianzas Conocidas

$$\begin{aligned} Y|X,\beta &\sim N_N(\mathbf{X}\beta, \mathbf{V}), \\ \beta &\sim N_p(\mathbf{b}, \mathbf{B}), \end{aligned}$$

donde $\boldsymbol{Y}=(Y_1,\ldots,Y_N)'$ es el vector de características de la población, $\boldsymbol{X}=(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N)'$ es una matriz Nxp conocida, **V** es una matriz NxN definida positiva conocida, **b** es un p-vector conocido y **B** es una matriz pxp definida positiva conocida.

2. Modelo Básico con Varianzas Desconocidas

$$\begin{aligned} Y|X,\beta,\sigma^2 &\sim N_N(\mathbf{X}\beta, \mathbf{W}\sigma^2), \\ \beta \mid \sigma^2 &\sim N_p(\mathbf{b}, \mathbf{C}\sigma^2), \\ \sigma^2 &\sim Ga^{-1}(a_0, b_0), \end{aligned}$$

donde $\boldsymbol{Y}=(Y_1,\ldots,Y_N)'$ es el vector de características de la población, $\boldsymbol{X}=(\boldsymbol{X}_1,\ldots,\boldsymbol{X}_N)'$ es una matriz Nxp conocida, **W** es una matriz NxN definida positiva conocida, **b** es un p-vector conocido, **C** es una matriz pxp definida positiva conocida y a_0 y b_0 son enteros positivos conocidos.

Supongamos ahora, que observamos los valores del vector $\boldsymbol{Y}$ para una muestra s de n elementos y deseamos estimar la media poblacional:

$$\bar{Y} = \frac{\sum_{i=1}^{N} Y_i}{N} = \frac{n}{N}\overline{y_s} + \frac{N-n}{N}\overline{Y_r},$$

donde $\overline{y_s}$ es la media muestral observada y $\overline{Y_r}$ es la media de los elementos no muestreados. Entonces si denotamos por(1):

– $\boldsymbol{Y}=(\boldsymbol{y_s}',\boldsymbol{Y_r}')'$ con $\boldsymbol{y_s}$ la parte observada dada una muestra de $\boldsymbol{Y}$ e $\boldsymbol{Y_r}$ la parte no observada de $\boldsymbol{Y}$.

– $\boldsymbol{X}=(\boldsymbol{X_s}',\boldsymbol{X_r}')'$ con $\boldsymbol{X_s}$ la parte de $\boldsymbol{X}$ correspondiente a los elementos observados y $\boldsymbol{X_r}$ la parte correspondiente a elementos no observados.

– Y particionamos las matrices:

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_s & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_r \end{pmatrix}, \quad \mathbf{W} = \begin{pmatrix} \mathbf{W}_s & \mathbf{W}_{sr} \\ \mathbf{W}_{rs} & \mathbf{W}_r \end{pmatrix},$$

denotando con s la parte correspondiente a los elementos observados, r la de los elementos no observados y sr las correspondiente a los elementos observados y no observados.

– $\boldsymbol{1_r}$ un $(N\text{-}n)$-vector de unos y $\boldsymbol{1_s}$ un n-vector de unos.

Tenemos realizando los cálculos oportunos que las distribuciones predictivas del vector $\boldsymbol{Y_r}/\boldsymbol{y_s}$, serían las siguientes:

– *para el modelo básico con varianzas conocidas,* $\mathbf{Y}_r/\boldsymbol{y_s}$ se distribuye Normal $(N\text{-}n)$-variante con media y matriz de varianzas-covarianzas dadas respectivamente por:

$$\begin{aligned} E(\boldsymbol{Y}_r|\boldsymbol{y}_s) &= \boldsymbol{X}_r\hat{\boldsymbol{\beta}}_B + \mathbf{V}_{rs}\mathbf{V}_s^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_s\hat{\boldsymbol{\beta}}_B), \\ Var(\boldsymbol{Y}_r|\boldsymbol{y}_s) &= \mathbf{V}_r - \mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{V}_{sr} + \\ &\quad + (\boldsymbol{X}_r - \mathbf{V}_{rs}\mathbf{V}_s^{-1}\mathbf{X}_s)(\boldsymbol{X}_s'\mathbf{V}_s^{-1}\boldsymbol{X}_s + \mathbf{B}^{-1})^{-1}\,(\boldsymbol{X}_r - \mathbf{V}_{rs}\mathbf{V}_s^{-1}\boldsymbol{X}_s)', \end{aligned}$$

donde

$$\hat{\boldsymbol{\beta}}_B = (\boldsymbol{X}_s'\mathbf{V}_s^{-1}\boldsymbol{X}_s + \mathbf{B}^{-1})^{-1}\,(\boldsymbol{X}_s'\mathbf{V}_s^{-1}\boldsymbol{y}_s + \mathbf{B}^{-1}\mathbf{b}),$$

es el estimador Bayes para función de pérdida cuadrática del coeficiente de regresión.

– *para el modelo básico con varianzas desconocidas,* $\boldsymbol{Y_r}/\boldsymbol{y_s}$ se distribuye $(N\text{-}n)$-variante t con $n+2a_0$ grados de libertad con media y matriz de varianzas-covarianzas para $n+2a_0>2$ dadas por:

(1) Suponemos sin pérdida de generalidad el vector $\boldsymbol{Y}$ ordenado de modo que los n primeros valores corresponden a elementos muestreados y los $N\text{-}n$ restantes a los no observados.

$$\begin{aligned}
E(Y_r|y_s) &= X_r\hat{\beta}'_B + W_{rs}W_s^{-1}(y_s - X_s\hat{\beta}'_B),\\
Var(Y_r|y_s) &= \left[\frac{1}{2}(y_s - X_s b)'(W_s + X_s C X_s')^{-1}(y_s - X_s b) + b_0\right]\frac{2}{n+2a_0-2}\times\\
&\times\left[W_r - W_{rs}W_s^{-1}W_{sr} + \right.\\
&\left.+ (X_r - W_{rs}W_s^{-1}X_s)(X_s'W_s^{-1}X_s + C^{-1})^{-1}(X_r - W_{rs}W_s^{-1}X_s)'\right]
\end{aligned}$$

donde:

$$\hat{\beta}'_B = (X_s'W_s^{-1}X_s + C^{-1})^{-1}(X_s'W_s^{-1}y_s + C^{-1}b),$$

es el estimador Bayes para función de pérdida cuadrática del coeficiente de regresión.

Y los estimadores Bayes que minimizan la pérdida cuadrática esperada dada la muestra para la media poblacional por elemento y su error vendrán dados:

– para el modelo básico con varianzas conocidas por:

$$\begin{aligned}
\hat{\bar{Y}}_{B1} &= \frac{n}{N}\bar{y}_s + \frac{1}{N}1_r'\left[X_r\hat{\beta}_B + V_{rs}V_s^{-1}(y_s - X_s\hat{\beta}_B)\right]1_r,\\
Var(\bar{Y}|y_s) &= \frac{1}{N^2}1_r'\,Var(Y_r|y_s)1_r.
\end{aligned}$$

– para el modelo básico con varianzas desconocidas (con $n+2a_0>2$) por:

$$\begin{aligned}
\hat{\bar{Y}}_{B2} &= \frac{n}{N}\bar{y}_s + \frac{1}{N}1_r'\left[X_r\hat{\beta}_B' + W_{rs}W_s^{-1}(y_s - X_s\hat{\beta}_B')\right]1_r,\\
Var(\bar{Y}|y_s) &= \frac{1}{N^2}1_r'\,Var(Y_r|y_s)1_r.
\end{aligned}$$

2.2 Modelo en Dos Etapas

Supongamos una población de N elementos estructurada en K unidades de modo que cada unidad i contiene M_i elementos, y supongamos que asociado al elemento j-ésimo de la unidad i-ésima tenemos una cantidad de interés y_{ij} que asumiremos como realización de una variable aleatoria Y_{ij} y un p-vector conocido $X_{ij}=(x_{ij1},\ldots,x_{ijp})'$. Nos planteamos dos situaciones:

1. Modelo de Dos Etapas con Varianzas Conocidas

$$\begin{aligned}
Y_{ij}|X_{ij},\beta_i &\sim N(X_{ij}'\beta_i,\sigma^2),\quad j = 1,2,\ldots,M_i,\ i = 1,2,\ldots,K, \text{ independientes},\\
\beta_i \mid \beta &\sim N_p(\beta, B_\beta),\quad i = 1,2,\ldots,K, \text{ iid},\\
\beta &\sim N_p(\beta_0, B_{\beta 0}),
\end{aligned}$$

donde con iid indicamos "independientes e igualmente distribuidas", β_0 es un escalar conocido, $\mathbf{B}_\beta$ y $\mathbf{B}_{\beta 0}$ son matrices *pxp* definidas positivas conocidas y σ^2 es un escalar positivo conocido.

2. Modelo de Dos Etapas con Varianzas Desconocidas

$$
\begin{aligned}
Y_{ij}|X_{ij},\beta_i,\sigma^2 &\sim N(X_{ij}'\beta_i,\sigma^2), \quad j = 1,2,\dots,M_i,\ i = 1,2,\dots,K,\\
\beta_i \mid \beta &\sim N_p(\beta, C_\beta\sigma^2), \quad i = 1,2,\dots,K, \quad \text{iid},\\
\beta|\sigma^2 &\sim N_p(\beta_0, C_{\beta 0}\sigma^2),\\
\sigma^2 &\sim Ga^{-1}(a_0,b_0),
\end{aligned}
$$

donde con iid indicamos 'independientes e igualmente distribuidas', β_0 es un escalar conocido, $\mathbf{C}_\beta$ y $\mathbf{C}_{\beta 0}$ son matrices *pxp* definidas positivas conocidas y a_0 y b_0 son escalares positivos conocidos.

Supongamos ahora, que obtenemos una muestra de *n* elementos de la población por el siguiente procedimiento:

– Seleccionamos *k* de las *K* unidades en las que está estructurada la población.

– Si *i* es una de las unidades seleccionadas, obtenemos una muestra s_i de m_i elementos, con $\sum_{i=1}^{k} m_i = n$.

y para los elementos de la muestra observamos los valores de ***Y***; nuestra intención es estimar la media poblacional, es decir:

$$
\overline{Y} = \frac{\sum_{i=1}^{K}\sum_{j=1}^{M_i} Y_{ij}}{N} = \frac{n}{N}\overline{y_s} + \frac{\sum_{i=1}^{k} M_i - n}{N}\overline{Y_{sr}} + \frac{\sum_{i=k+1}^{K} M_i}{N}\overline{Y_{rr}},
$$

donde $\overline{y_s}$ es la media muestral por elemento, $\overline{Y_{sr}}$ es la media de los elementos no muestreados de unidades muestreadas y $\overline{Y_{rr}}$ es la media de los elementos no muestreados de unidades no muestreadas. Entonces si denotamos por(2):

– $\boldsymbol{Y}=(\boldsymbol{y}_{s1}',\boldsymbol{Y}_{r1}',\dots,\boldsymbol{y}_{sk}',\boldsymbol{Y}_{rk}',\boldsymbol{Y}_{rr}')'$ con $\boldsymbol{y}_{si}$ el vector observado para la unidad muestreada *i*-ésima,

(2) Suponemos sin pérdida de generalidad el vector ***Y*** ordenado de modo que los *k* primeros vectores Y_i corresponden a las unidades observadas y los *K-k* restantes a las unidades no observadas, ordenándose además los vectores Y_i correspondientes a unidades observadas de modo que los m_i primeros lugares corresponden a los valores de **Y** de elementos muestreados de la unidad *i* muestreada.

– $\boldsymbol{Y}_{ri}$ el vector no observado para la unidad muestreada i-ésima e $\boldsymbol{Y}_{rr}$ el vector no observado para las unidades no muestreadas.

– $\mathbf{y}_s=(\mathbf{y}_{s1}',\ldots,\mathbf{y}_{sk}')'$

– $\mathbf{Y}_r=(\mathbf{Y}_{r1}',\ldots,\mathbf{Y}_{rk}',\mathbf{Y}_{rr}')'$

– $\boldsymbol{X}=(\boldsymbol{X}_{s1}',\boldsymbol{X}_{r1}',\ldots,\boldsymbol{X}_{sk}',\boldsymbol{X}_{rk}',\boldsymbol{X}_{rr}')'$ con $\boldsymbol{X}_{si}$ la parte de $\boldsymbol{X}$ correspondiente a los elementos observados de la unidad muestreada i-ésima, $\boldsymbol{X}_{ri}$ la parte de $\boldsymbol{X}$ correspondiente a los elementos no observados de la unidad muestreada i-ésima y $\boldsymbol{X}_{rr}$ la parte de $\boldsymbol{X}$ correspondiente a los elementos no observados de las unidades no muestreadas.

– $\mathbf{1}_r=(\mathbf{1}_{M1-m1}',\ldots,\mathbf{1}_{Mk-mk}',\mathbf{1}_{Mk+1}',\ldots,\mathbf{1}_{MK}')'$ un (N-n)-vector de unos.

– $\boldsymbol{1}_s=(\boldsymbol{1}_{m1}',\ldots,\boldsymbol{1}_{mk}')'$ un n-vector de unos

tenemos realizando los cálculos oportunos que las distribuciones predictivas del vector $Y_r|y_s$, serían las siguientes:

– *para el modelo en dos etapas con varianzas conocidas,* $\boldsymbol{Y}_r|\boldsymbol{y}_s$ se distribuye Normal (N-n)-variante con media y matriz de varianzas-covarianzas dadas por:

$$
E(Y_r|y_s) = \begin{pmatrix} X_{r1}\left[\hat{\beta}_{B_{s1}} - (X_{s1}'X_{s1}\sigma^{-2}+\mathbf{B}_\beta^{-1})^{-1}\hat{\gamma}_B\right] \\ \vdots \\ X_{rk}\left[\hat{\beta}_{B_{sk}} - (X_{sk}'X_{sk}\sigma^{-2}+\mathbf{B}_\beta^{-1})^{-1}\hat{\gamma}_B\right] \\ X_{rr}\left[\hat{\beta}_{B_r} - \mathbf{B}_\beta\hat{\gamma}_B\right] \end{pmatrix},
$$

$$
\begin{aligned}
Var(Y_r|y_s) = &\begin{pmatrix} \mathrm{diag}_{i=1}^{k}\left\{X_{ri}(X_{si}'X_{si}\sigma^{-2}+\mathbf{B}_\beta^{-1})^{-1}X_{ri}'\right\} & 0 \\ 0 & \mathrm{diag}_{i=k+1}^{K}\left\{X_{rri}\mathbf{B}_\beta X_{rri}'\right\} \end{pmatrix} \\
&- \begin{pmatrix} Z_{sr}\mathbf{D}^{-1}Z_{sr}' & Z_{sr}\mathbf{D}^{-1}X_{rr}' \\ X_{rr}\mathbf{D}^{-1}Z_{sr}' & X_{rr}\mathbf{D}^{-1}X_{rr}' \end{pmatrix} + I_{N-n}\sigma^2,
\end{aligned}
$$

donde

$$
\begin{aligned}
\hat{\beta}_{B_{si}} &= (X_{si}'X_{si}\sigma^{-2}+\mathbf{B}_\beta^{-1})^{-1}\left[X_{si}'y_{si}\sigma^{-2}+\mathbf{B}_\beta^{-1}\left[I_p - K\left(K\mathbf{B}_\beta^{-1}+I_p\mathbf{B}_{\beta 0}^{-1}\right)^{-1}\mathbf{B}_\beta^{-1}\right]\beta_0\right], \\
\hat{\beta}_{Br} &= \left[I_p - K\left(K\mathbf{B}_\beta^{-1}+I_p\mathbf{B}_{\beta 0}^{-1}\right)^{-1}\mathbf{B}_\beta^{-1}\right]\beta_0, \\
\hat{\gamma}_B &= \mathbf{B}_\beta^{-1}\mathbf{D}^{-1}\left[\mathbf{B}_\beta^{-1}\sum_{i=1}^{k}\hat{\beta}_{B_{si}} + (K-k)\hat{\beta}_{Br}\right], \\
Z_{sr} &= \left(\mathbf{B}_\beta^{-1}(X_{s1}'X_{s1}\sigma^{-2}+\mathbf{B}_\beta^{-1})^{-1}X_{r1}',\ldots,\mathbf{B}_\beta^{-1}(X_{sk}'X_{sk}\sigma^{-2}+\mathbf{B}_\beta^{-1})^{-1}X_{rk}'\right)', \\
\mathbf{D} &= \sum_{i=1}^{k}\mathbf{B}_\beta^{-1}(X_{si}'X_{si}\sigma^{-2}+\mathbf{B}_\beta^{-1})^{-1}\mathbf{B}_\beta^{-1} - k\mathbf{B}_\beta^{-1} - I_p\mathbf{B}_{\beta 0}^{-1},
\end{aligned}
$$

siendo $\hat{\beta}_{B_{si}} - (X_{si}'X_{si}\sigma^{-2}+\mathbf{B}_\beta^{-1})^{-1}\hat{\gamma}_B$ el estimador Bayes para pérdida cuadrática del coeficiente de regresión para la unidad i observada y $\hat{\beta}_{B_r} - \mathbf{B}_\beta\hat{\gamma}_B$ el estimador Bayes del coeficiente de regresión para las unidades no observadas.

– *para el modelo en dos etapas con varianzas desconocidas,* $\mathbf{Y}_r|\mathbf{y}_s$ se distribuye (N-n)-variante t con $n+2a_0$ grados de libertad con media y matriz de varianzas-covarianzas para $n+2a_0>2$ dadas por:

$$E(Y_r|y_s) = \begin{pmatrix} X_{r1}\left[\hat{\beta}^t_{B_{s1}} - (X_{s1}'X_{s1} + C_\beta^{-1})^{-1}\hat{\gamma}^t_B\right] \\ \vdots \\ X_{rk}\left[\hat{\beta}^t_{B_{sk}} - (X_{sk}'X_{sk} + C_\beta^{-1})^{-1}\hat{\gamma}^t_B\right] \\ X_{rr}\left[\hat{\beta}^t_{B_r} - C_\beta\hat{\gamma}^t_B\right] \end{pmatrix},$$

$$\mathrm{Var}(Y_r|y_s) = M_G\left\{\begin{pmatrix} \mathrm{diag}_{i=1}^{k}\left\{X_{ri}(X_{si}'X_{si} + C_\beta^{-1})^{-1}X_{ri}'\right\} & 0 \\ 0 & \mathrm{diag}_{i=k+1}^{K}\left\{X_{ri}C_\beta X_{ri}'\right\} \end{pmatrix} - \right.$$

$$\left. - \begin{pmatrix} Z_{sr}{}^t\mathbf{D}_t{}^{-1}Z_{sr}{}^{t\prime} & Z_{sr}{}^t\mathbf{D}_t{}^{-1}X_{rr}' \\ X_{rr}\mathbf{D}_t{}^{-1}Z_{sr}{}^{t\prime} & X_{rr}\mathbf{D}_t{}^{-1}X_{rr}' \end{pmatrix} + I_{N-n}\right\},$$

donde:

$$\hat{\beta}^t_{B_{si}} = (X_{si}'X_{si} + C_\beta{}^{-1})^{-1}\left[X_{si}'y_{si} + C_\beta{}^{-1}\left[I_p - K\left(KC_\beta^{-1} + I_pC_{\beta 0}^{-1}\right)^{-1}C_\beta^{-1}\right]\beta_0\right].$$

$$\hat{\beta}^t_{B_r} = \left[I_p - K\left(KC_\beta^{-1} + I_pC_{\beta 0}^{-1}\right)^{-1}C_\beta^{-1}\right]\beta_0.$$

$$\hat{\gamma}_B{}^t = C_\beta^{-1}D_t{}^{-1}\left[C_\beta^{-1}\sum_{i=1}^{k}\hat{\beta}^t_{B_{si}} + (K-k)\hat{\beta}^t_{B_r}\right].$$

$$Z_{sr}{}^t = \left(C_\beta^{-1}(X_{s1}'X_{s1} + C_\beta{}^{-1})^{-1}X_{r1}', \ldots, C_\beta^{-1}(X_{sk}'X_{sk} + C_\beta{}^{-1})^{-1}X_{rk}'\right).$$

$$D_t = \sum_{i=1}^{k}C_\beta^{-1}(X_{si}'X_{si} + C_\beta{}^{-1})^{-1}C_\beta{}^{-1} - kC_\beta{}^{-1} - I_pC_{\beta 0}{}^{-1}.$$

$$M_G = \frac{2}{n+2a_0-2}\left\{\frac{1}{2}\sum_{i=1}^{k}(y_{si} - X_{si}\beta_0)'\left[I_{m_i} - X_{si}(X_{si}'X_{si} + C_\beta^{-1})^{-1}X_{si}'\right](y_{si} - X_{si}\beta_0) - \right.$$

$$\left. - \frac{1}{2}V_G + b_0\right\}.$$

$$V_G = \left\{\sum_{i=1}^{k}(y_{si} - X_{si}\beta_0)'\left[I_{m_i} - X_{si}(X_{si}'X_{si} + C_\beta^{-1})^{-1}X_{si}'\right]X_{si}\right\} \times$$

$$\times D_G^{-1}\left\{\sum_{i=1}^{k}X_{si}'\left[I_{m_i} - X_{si}(X_{si}'X_{si} + C_\beta^{-1})^{-1}X_{si}'\right](y_{si} - X_{si}\beta_0)\right\}.$$

$$D_G = \sum_{i=1}^{k}X_{si}'\left[I_{m_i} - X_{si}(X_{si}'X_{si} + C_\beta^{-1})^{-1}X_{si}'\right]X_{si} + C_{\beta 0}^{-1}.$$

siendo $\hat{\beta}_{B_{si}}{}^t - (\mathbf{X}_{si}'\mathbf{X}_{si} + C_\beta^{-1})^{-1}\hat{\gamma}_B{}^t$ el estimador Bayes para pérdida cuadrática del coeficiente de regresión para la unidad *i* observada y $\beta_{B_r}{}^t - C_\beta\hat{\gamma}_B{}^t$ el estimador Bayes del coeficiente de regresión para las unidades no observadas.

Y los estimadores Bayes que minimizan la pérdida cuadrática esperada dada la muestra para la media poblacional por elemento y su error vendrán dados por:

– para el modelo en dos etapas con varianzas conocidas:

$$
\begin{aligned}
\hat{\hat{Y}}_{2E1} &= \frac{n}{N}\overline{y_s} + \frac{1}{N}\sum_{i=1}^{k} \mathbf{1}_{M_i-m_i}{}'X_{ri}\left[\hat{\beta}_{B_{si}} - (X_{si}'X_{si}\sigma^{-2} + B_\beta^{-1})^{-1}\hat{\gamma}_B\right] + \\
&+ \frac{1}{N}\sum_{i=k+1}^{K} \mathbf{1}_{M_i}{}'X_{rri}\left[\hat{\beta}_{B_{ri}} - B_\beta\hat{\gamma}_B\right\} \\
\mathrm{Var}(\overline{Y}|y_s) &= \frac{N-n}{N^2}\sigma^2 + \frac{1}{N^2}\sum_{i=1}^{k} \mathbf{1}_{M_i-m_i}{}'X_{ri}(X_{si}'X_{si}\sigma^{-2} + B_\beta^{-1})^{-1}\left(X_{ri}'\mathbf{1}_{M_i-m_i} - \hat{\Gamma}_B\right) + \\
&+ \frac{1}{N^2}\sum_{i=k+1}^{K} \mathbf{1}_{M_i}{}'X_{rri}B_\beta\left(X_{rri}'\mathbf{1}_{M_i} - \hat{\Gamma}_B\right\}
\end{aligned}
$$

donde:

$$
\hat{\Gamma}_B = B_\beta^{-1}D^{-1}\left[\sum_{i=1}^{k} B_\beta^{-1}(X_{si}'X_{si}\sigma^{-2} + B_\beta^{-1})^{-1}X_{ri}'\mathbf{1}_{M_i-m_i} + \sum_{i=k+1}^{K} X_{rri}'\mathbf{1}_{M_i}\right],
$$

– para el modelo en dos etapas con varianzas desconocidas (con $n+2a_0>2$):

$$
\begin{aligned}
\hat{\hat{Y}}_{2E2} &= \frac{n}{N}\overline{y_s} + \frac{1}{N}\sum_{i=1}^{k} \mathbf{1}_{M_i-m_i}{}'X_{ri}\left[\hat{\beta}^t_{Bsi} - (X_{si}'X_{si}\sigma^{-2} + C_\beta^{-1})^{-1}\hat{\gamma}^t_B\right] + \\
&+ \frac{1}{N}\sum_{i=k+1}^{K} \mathbf{1}_{M_i}{}'X_{rri}\left[\hat{\beta}^t_{Br} - C_\beta\hat{\gamma}^t_B\right\} \\
\mathrm{Var}(\overline{Y}|y_s) &= M_G\left\{\frac{N-n}{N^2} + \frac{1}{N^2}\sum_{i=1}^{k} \mathbf{1}_{M_i-m_i}{}'X_{ri}(X_{si}'X_{si} + C_\beta^{-1})^{-1}\left(X_{ri}'\mathbf{1}_{M_i-m_i} - \hat{\Gamma}^t_B\right) + \right. \\
&\left. + \frac{1}{N^2}\sum_{i=k+1}^{K} \mathbf{1}_{M_i}{}'X_{rri}C_\beta\left(X_{rri}'\mathbf{1}_{M_i} - \hat{\Gamma}^t_B\right)\right\},
\end{aligned}
$$

donde:

$$
\hat{\Gamma}^t_B = C_\beta^{-1}D_t^{-1}\left[\sum_{i=1}^{k} C_\beta^{-1}(X_{si}'X_{si} + C_\beta^{-1})^{-1}X_{ri}'\mathbf{1}_{M_i-m_i} + \sum_{i=k+1}^{K} X_{rri}'\mathbf{1}_{M_i}\right]
$$

Los modelos en dos etapas presentados son casos particulares de los modelos multietápicos de Malec y Sedransk(1985) y Bernardo y Girón(1991) que presentan expresiones matriciales para las distribuciones predictivas y estimadores, la tarea en esta parte del artículo ha sido obtener expresiones manejables para estos dos modelos (que son los modelos usuales a emplear en una población que responda a una estructura bietápica respecto a la característica de interés) y que sean comparables con los resultados obtenidos para los modelos básicos anteriores. Y casos particulares de los modelos estudiados en este artículo han sido estudiados en Scott y Smith(1969) y Murgui(1982).

3. PROPIEDADES DE LOS MODELOS Y ESTIMADORES OBTENIDOS Y COMENTARIOS FINALES

Los modelos presentados y analizados en la sección anterior presentan las siguientes propiedades inmediatas:

1. La recuperación de los estimadores clásicos en los modelos básicos cuando se asume una distribución no informativa de los parámetros.

Efectivamente, en el caso del modelo básico con varianzas conocidas se recuperan para $B^{-1} \to 0$, los resultados de Royall (1976) sobre estimador y error cuadrático medio asociado para la media poblacional obtenidos desde un punto de vista de superpoblación clásica y los estimadores de regresión, de razón y estimador media muestral presentados por Cochran (1977) para estimar la media poblacional desde una prespectiva de población fija si además asumimos: a) $\mathbf{V} = I_N\sigma^2$, b) $p = 1$, $\mathbf{V} = \mathrm{diag}_{i=1}^{N}\left\{\sigma^2 X_i\right\}$ y c) $p = 1$, $\mathbf{V} = I_N\sigma^2$, $X = I_N$, respectivamente. Y en el caso del modelo básico con varianzas desconocidas se recuperan asimismo todos los estimadores anteriores para la distribución inicial $\pi\ (\beta,\sigma^2) \propto \frac{1}{\sigma^2}$, sustituyendo en este caso **V** por **W**.

2. La recuperación de los modelos básicos (con unidades independientes) a partir de los modelos de dos etapas cuando asumimos que la varianza entre unidades es muy pequeña.

3. La recuperación de los resultados clásicos en los modelos de dos etapas cuando asumimos que la varianza entre unidades es muy grande y bajo la hipótesis adicional de muestrear en todas las unidades.

Notemos que las propiedades 2 y 3 confirman un buen comportamiento de los modelos anteriores para la descripción de una población finita. La dificultad de aplicación de estos modelos nace de la necesidad de tomar decisiones iniciales acerca de la población como son:

– La elección entre modelos monoetápicos y bietápicos. En este sentido, usaremos un modelo monoetápico cuando nuestras creencias previas de la población nos señalen que la población responde a esta estructura, esto es, no podamos distinguir unidades de elementos o bien las unidades sean muy similares entre sí, y usaremos un modelo bietápico cuando nuestras creencias previas nos señalen una población estructurada según unidades (en principio diferentes entre sí), que se dividen a su vez en elementos.

– La determinación de las distribuciones iniciales. Respecto a este punto y salvo que tengamos unas creencias muy fuertes acerca de la población en estudio es aconsejable la elección de distribuciones iniciales mínimo informativas o con varianzas altas que representen nuestros conocimientos iniciales acerca de la población. Bolfarine y Zacks (1992) detectaron que el modelo básico era bastante sensible a la distribución inicial asignada y los modelos bietápicos en la práctica suelen manifestarse bastante robustos.

Otra cuestión a discutir es la relación entre los modelos planteados y el método aplicado para la obtención de la muestra de la población, respecto a este punto, el trabajo de Bayarri y Font(1994) apoyándose en los modelos con varianzas conocidas antes descritos, realiza un análisis del funcionamiento del muestreo aleatorio simple (que tiene asociado un modelo monoetápico) frente a un muestreo aleatorio estratificado y cluster (que tienen asociado un modelo bietápico).

REFERENCIAS

BAYARRI, M.J. y FONT, B. (1994). «A (Bayesian) Note on Non-Random Samples from Finite Populations», *Technical Report 94-32, Depart. of Statist., Purdue University.*

BERNARDO, J.M. y GIRÓN, J. (1991). «Robust Sequential Prediction from Non-random Samples: The Election Night Case», *Bayesian Statistics*, 4, 61-77.

BOLFARINE, H. y ZACKS, S. (1991). «Bayes and Minimax Prediction in Finite Populations», *J. Statist. Planning and Inference*, 28(3), 139-151.

BOLFARINE, H. y ZACKS, S. (1992). «Prediction Theory for Finite Populations,» *Springer-Verlag.*

COCHRAN, W.G. (1977). «Sampling Techniques. Third edition», *John Wiley & Sons,* New York.

ERICSON, W.A. (1969). «Subjetive Bayesian Models in Sampling Finite Population», *J. Roy. Statist. Soc. (Ser. B)*,31, 195-233.

ERICSON, W.A. (1970). «On the Posterior Mean and Variance of a Population Mean», *J. Americ. Statist. Assoc*, 65, 649-652.

ERICSON, W.A. (1983). «A Bayesian approach to regression estimation in finite populations», *Technical Report N.120, Depart. of Statist., University of Michigan*, 1-21.

ERICSON, W.A. (1988). «Bayesian Inference in Finite Populations», *Handbook of Statistics*, 6, 213-246.

GHOSH, M. y LAHIRI, P. (1987). «Robust Empirical Bayes Estimation of Means From Stratified Samples», *J. Americ.Statist. Assoc.*, 82, 1153-1162.

MALEC, D. y SEDRANSK, J. (1985). «Bayesian Inference for Finite Population Parameters in Multistage Cluster Sampling», *J. Amer. Statist. Association*, 80, 897-902.

MURGUI, J.S. (1982). «Diseño e Inferencia en Poblaciones Finitas: Modelos de Superpoblación», *Tesis Doctoral, Universidad de Valencia.*

RODRIGUES, J. (1988). «Some results on restricted Bayes least squares predictors for finite populations», *South African Statistics Journal*, 22, 45-53.

ROYALL, R.M. (1976). «The Linear Least-Squares Prediction Approach to Two-Stage Sampling» *J. Amer. Statist. Association*, 71, 657-664.

ROYALL, R.M. y PFEFFERMANN, D. (1982). «Balanced samples and robust Bayesian inference in finite population sampling», *Biometrika*, 69, 401-409.

SCOTT, A. & SMITH, T.M.F. (1969). «Estimation in Multi-stage Surveys», *J. Amer. Statist. Association*, 64, 830-840.

SMOUSE. E.P. (1982). «Bayesian Least Squares Inference for Finite Population Models», Artículo presentado en *Institute for Mathematical Statistic meeting in Cincinnati, Ohio, August*

SMOUSE, E.P. (1984). «A Note on Bayesian Least Squares Inference for Finite Population Models» *J. Amer. Statist. Association*, 79, 390-392.

BAYESIAN INFERENCE IN FINITE POPULATIONS: A REVIEW

SUMMARY

In this communication, I shall try to provide a review about some results of Bayes predictors of the population mean under squared error loss function. I shall introduce two regression models under normality: a basic model and a two stage model in two cases: known and unknown variances, and apply a Bayes analysis. For the two models I shall done the Bayes predictive distributions, the Bayes estimators of the population mean and their associated expected loss.

Keywords: Bayesian inference; finite population; multiple regression; prediction.