ISSN 0103-2509

A simple likelihood theory for finite populations

Rodrigues, Josemar

nº 4

São Carlos (SP)
1990

A Simple Likelihood Theory For Finite Populations

Josemar Rodrigues
ICMSC-UNIVERSIDADE DE SÃO PAULO
13560-SÃO CARLOS- SP-BRASIL

1 Summary

The purpose of this paper is to introduce a simple conditional likelihood approach in finite populations, based on the sufficient statistic for the population total.This is a simple version of the likelihood approach considered by Royall (1976). A characterization of the sufficient statistic for the population total under the normal linear superpopulation model is given. An orthogonal likelihood model in the sense of Cox and Reid (1987) for the population total is also discussed.

2 Introduction

Statistical inference about the population total of a model based directly on the likelihood function was proposed by Royall(1976).His approach for the normal superpopulation model is not immediately clear,at least from his method suggested in the paper. Under a multivariate normal regression model,the purpose of this paper is to propose a simple conditional likelihood procedure to shown that the best linear unbiased predictor (BLUP) is the maximum likelihood predictor (MLP) of the population total. A characterization of the identifiability of the conditional likelihood function for the population total is considered. When the conditional model is not identifiable ,based on the work of Cox and Reid (1987),we suggest an orthogonal model in which the BLUP is the MLP for the population total.

3 Normal Regression Models

3.1 Characterization of the Conditional Likelihood Function

The population of interest consists of N units. Associated with the ith unit is a number y_i whose value is unknown. In order to gain information about $t = y_1 + \ldots + y_N = 1'y$, where $y' = (y_1, \ldots, y_N)$ and $1' = (1, \ldots, 1)$, a sample s is selected from the population units. As Pereira and Rodrigues (1983), $i_k = 1$ if $k \in s$ and $i_k = 0$ if $k \notin s$ Suppose that y' is is a realized value of the random vector $Y' = (Y_1, \ldots, Y_N)$ with a N-variate normal mean vector $X\beta$ and known diagonal covariance matrix V. Here X is a known $N \times p$ matrix and β is a $p \times 1$ vector of unknown regression coefficients.Given s, we consider the following notations:

$$\begin{aligned} Y_s &= I_s Y, \quad X_s = I_s X, \quad X_r = (I - I_s)X \\ V_s &= I_s V, \quad V_r = (I - I_s)V, \quad V_s^- = I_s V^{-1} \end{aligned}$$

, where I_s is a diagonal matrix of order N with its kth diagonal element being i_k $(k = 1, \ldots, N)$.

It has been shown (Pereira & Rodrigues 1983) that the best linear unbiased predictor (BLUP) of $T = 1'Y$ under the normal model is

$$\hat{t}_{BLUP} = 1'Y_s + 1'X_r\hat{\beta}, \tag{1}$$

where $\hat{\beta} = (X_s'V_s^- X_s)^- X_s'V_s^- Y_s = B_s Y_s$. The following condition has been appeared in many bayesian and nonbayesian contexts (Bolfarine and Rodrigues 1988;Royall 1976).

Condition L: $V1 = X\gamma$ for some p-vector γ.

From the normal superpopulation model, it is not difficult to see that the conditional distribution of $1'X\hat{\beta}$, given $T = t$ and β is $N(\mu; \sigma^2)$,where

$$\mu = (1 - \alpha)1'X\beta + \alpha t \tag{2}$$

$$\sigma^2 = \sigma_{11}^2(1 - \rho_s^2) \tag{3}$$

$$\sigma_{11}^2 = 1'X(X_s'V_s^- X_s)^- X'1 \tag{4}$$

$$\alpha = \frac{1'XB_sV1}{1'V1}, \tag{5}$$

$$\rho_s = \frac{1'XB_sV1}{\sqrt{1'X(X_s'V_s^- X_s)^- X'11'V1}}. \tag{6}$$

Next we give the main results of the paper:

THEOREM 3.1.1: If Y has $N(X\beta; V)$ probability distribution, then when $1'X\hat{\beta}$ is observed and $T = t$ is given, the conditional likelihood function for $T = t$ is proportional to $N(1'X\hat{\beta}; \sigma^2)$ if and only if the known diagonal covariance matrix satisfies Condition L.

Proof: The result follows in a straightforward manner by noting that:

1. The conditional distribution of $1'X\hat{\beta}$,given $T = t$, is proportional to a $N(1'X\hat{\beta}; \sigma^2)$ distribution function $\Longleftrightarrow \alpha = 1$.

2. $\alpha = 1 \Longleftrightarrow 1'XB_sV1 = 1'V1 \Longleftrightarrow 1'[I - XB_s]V1 = 0 \Longleftrightarrow 1'P_{M^\perp(X)}[V1] = 0 \Longleftrightarrow V1 \varepsilon M(X)$, where $P_{M^\perp(X)}[V1]$ is the projection of the vector $V1$ onto the orthogonal complement of the space generated by the columns of X which we denote by $M(X)$.

It is interesting to observe from (2) that the conditional distribution of $1'X\hat{\beta}$,given $T = t$ and β, is identifiable on t if and only if Condition L is satisfied. This characterization is not clear in Royalls's approach (1967). The following result gives an interesting interpretation for Conditional L and shows the equivalence between this simple likelihood viewpoint and the Royall's approach (1976).

THEOREM 3.1.2: Under the assumptions of Theorem 3.1.1, given $T = t$, the necessary and sufficient condition for $1'X\hat{\beta}$ to be sufficient for t is that Condition L must be satisfied.

Proof: The result follows from Theorem 3.1.1 and Royall's Theorem (p.611, 1976) by noting that:

$$\begin{aligned} 1'V_r^{-}1 + 1'X_r(X_s'V_s^{-}X_s)^{-}X_r'1 &= Var(1'X\hat{\beta} - T) \\ &= E\{Var[(1'X\hat{\beta} - T) \mid T]\} + \\ &+ Var\{E[(1'X\hat{\beta} - T) \mid T]\} \\ &= (1 - \rho_s^2)\sigma_{11}^2. \end{aligned} \tag{7}$$

Theorem 3.1.2 gives an interesting version for Royall's Theorem (p.611,1976) in terms of Condition L. If Condition L is not satisfied then the BLUP of T is the MLP under the augmented model by adding to X a column proportional to the variance vector, and adding a corresponding coefficient to the vector β.

3.2 Orthogonal Likelihood Function

The conditional model introduced in Section 3.1 is identifiable on $T = t$ only when Condition L is satisfied. So, in this section we present a model where the unbiased predictors of T are the maximum likelihood predictors(MLP). As Pereira and Rodrigues (1983), we denote the class of all predictors $h'Y_s$ satisfying $E[h'Y_s - T] = 0$ by U_l.

LEMMA 3.2.1: If Y has a $N(X\beta; V)$ probability distribution then

$$\hat{T} = h'Y_s \in U_l \Longleftrightarrow h'X_s = 1'X. \tag{8}$$

Proof: It follows readily from the definition of U_l.

LEMMA 3.2.2: If Y has a $N(X\beta; V)$ probability distribution then

$$\begin{aligned} h'X_s &= 1'X \Longleftrightarrow \hat{T} - T = (h'I_s - 1)e, \quad \text{where} \\ e &= Y - X\beta \end{aligned} \tag{9}$$

Proof: See Hajek's book (1978).

THEOREM 3.2.1: If Y has a $N(X\beta; V)$ distribution function and $\hat{T} \in U_l$, then

$$\begin{pmatrix} 1'X\hat{\beta} - 1'X\beta \\ \hat{T} - T \end{pmatrix} \sim N\left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} b_{11}^2 & b_{12} \\ b_{21} & b_{22}^2 \end{pmatrix} \right\} \tag{10}$$

$$\begin{aligned} b_{11}^2 &= 1'X(X_s'V_s^- X_s)^- X'1 \\ b_{22}^2 &= (h'I_s - 1')V(I_s h - 1) \\ b_{12} = b_{21} &= 1'XB_s V(I_s h - 1) \end{aligned} \tag{11}$$

Proof: The result follows from Lemma 3.2.1 ,Lemma 3.2.2 and the known properties of the multivariate normal distribution.

In the present section ,under model (10), our main purpose is to make inference about the parameter of interest $\varphi = T$ in the presence of the nuisance parameter $\phi = 1'X\beta$. The procedure can be simplified by working with an orthogonal parametrization (Cox and Reid, 1987) in which the parameter φ is orthogonal to the new nuisance parameter λ. Therefore, we transform the origin parameters (φ, ϕ) in model (10) to the orthogonal parameters (φ, λ). This new likelihood function we call the orthogonal likelihood function on (φ, λ). As discussed by Cox and Reid (1987), under this orthogonal model, the maximum likelihood predictor of φ, when λ is given, varies only slowly with λ.The orthogonal parameter λ is obtained by solving the differential equation

$$\frac{b_{11}\rho}{b_{22}} = \frac{\partial\phi}{\partial\varphi}. \tag{12}$$

This differential equation gives a one-to-one transformation

$$\lambda = \phi - \frac{b_{11}\rho\varphi}{b_{22}} \tag{13}$$

It is unexpected that differentiating the orthogonal likelihood model with respect to φ, given λ, we find that the maximum likelihood predictor $\hat{\varphi}_\lambda$ is

$$\hat{\varphi}_\lambda = \hat{T}, \quad \forall \quad \hat{T} \in U_l. \tag{14}$$

References

[1] Cox, D. & Reid, N. (1987). Parameter orthogonality and approximate conditional inference (with discussion). J.R.Statist.Soc.B 49,1-49.

[2] Bolfarine, H. & Rodrigues, J. (1988). A simple projection predictor in finite populations. Australian J. of Statist.(to appear).

[3] Hajek, J. (1978). Theory of sampling from finite populations. manuscript.

[4] Pereira, C.A.B. & Rodrigues, J. (1983). Robust linear prediction in finite populations. Int.Statist. Rev. 51, 293-300.

[5] Royall, R.M. (1976). Likelihood functions in finite population sampling theory. Biometrika 63, 3, 605-614.