RT-MAE 9406

PREDICTIVISTIC STATISTICAL INFERENCE IN FINITE POPULATIONS

by

Pilar L. Iglesias Zuazola, Carlos A.B. Pereira and Sergio Wechsler

Palavras Chaves: Bayes, Sampling, Finite Population, Exchangeability.
(Key words)

Classificação AMS: 62A15, 62D05
(AMS Classification)

- março de 1994 -

Predictivistic statistical inference in finite populations

Pilar L. Iglesias-Zuazola, Carlos A.B. Pereira, Sérgio Wechsler

Universidade de São Paulo
São Paulo, Brazil

Abstract

The familiar finite population model - where information provided by a subset of units is used in order to reduce the uncertainty about a function of the complete population list of values - is explored from a predictivistic point of view. Under this approach, only operationally meaningful quantities are considered and therefore no superpopulation parameters are involved. Actually, the only parameter present is the list of population values - a function of it being estimated. All probabilistic statements follow from uncertainty about the values of the unsampled units and from the structure of the population relative to the values. Sampling plans play no rôle, as they do in classical sampling theory. Superpopulation models - which can be understood by invariant properties of infinite sequences - are also discarded. Any hierarchical assignment of probabilities that corresponds to the marginal distribution of the complete vector of numerical values is allowed. This approach is strict Bayesian: it does not violate the Likelihood Principle, it is not built on the consideration of a family of labelled distributions and it does not need assumptions of extendibility to infinite populations.

1 Introduction

The structure being considered in this paper is given by a finite population of clearly labelled units, $P = \{1, 2, \ldots, N\}$, having its size, N, known. To each unit there is associated a real-valued number or vector, so that we have the set of unknowns $X = (X_1, X_2, \ldots, X_N) \in \mathcal{X}^N$. The sample space set $\mathcal{X}$ is known. Each X_i is unknown as long as unit i is not available for inspection. One is interested on making statistical inference about some parameter T with incorporation of information provided by a sample y. The *parameter* T is a function $T = T(X_1, X_2, \ldots, X_N)$ and a *sample* of $n < N$ units becomes available, i.e., n units are inspected and have, therefore, their x_i values revealed. We will write $y = (y_1, y_2, \ldots, y_n) = (x_{i_1}, x_{i_2}, \ldots, x_{i_n})$, for some labels $1 \leq i_1 \leq i_2 \leq \cdots \leq i_n \leq N$.

The point of the authors is that such a population is not only finite in its size, but also closed, i.e., it should be protected against introduction of models possessing unobservable, and therefore lacking operational meaning, quantities. The sole randomness for You is due to Yours ignorance of (some) of the x_i and the probability statements of You should refer exclusively to X. As it is very well-known, most approaches do not keep the population closed. For example,

a) Frequentist, Cochran-style sampling models [Cochran (1977)]:

This approach is first of all non-Bayesian: it does not allow (true) probability for X and it violates the Likelihood Principle. Even among frequentist statisticians, it makes some trouble, as the sole source of randomness is induced by the sampling plan. This, in turn, generates likelihood functions which are always flat on their support, yielding the somewhat awkward situation where there is no non-trivial reduction of uncertainty about X [Basu (1969)]. This is overcome by

b) Frequentist superpopulation models:

Essentially there is now a probability distribution $F_X(\cdot|\theta)$ on $\mathcal{X}^N$ and inference concentrates on $\theta \in \Theta$. This is again non-Bayesian and we shall not discuss it unless it is changed to

c) Bayesian superpopulation models ("light" Bayesian models):

There is a formal Bayesian update of densities for θ, allowing computation of predictive densities:

$$f_T(t|y) = \int_\Theta f(\theta|y) f_T(t|y,\theta)\, d\theta \,.$$

We propose a return to basic subjectivistic reasoning, by considering

d) De facto, strict (predictivistic) Bayesian models:

We suggest that the problem must be dealt by Ockham's razor and reduce to assessment of a prior distribution $f_T(t)$ for T, the parameter, to be updated by Bayes' formula, yielding

$$f_T(t|y) \propto f_T(t) f(y|T=t) .$$

This can also be implemented by assessing first $f_X(x_1, x_2, \ldots, x_N)$, as such marginal distribution decomposes on

$$f_X(x_1, x_2, \ldots, x_N) = \int_{T(X)} f_T(t) f_X(x|T=t)\, dt .$$

[There is no distinction really between prior and likelihood – both are equally important components of $f_X(x)$]. In other words, the hierarchy by successive assessment of $f_X(x)$, $f_T(t)$, $f_X(x|T=t)$, $f(y|T=t)$, $f_T(t|y)$ is satisfactory as well. It should be noticed the *predictivistic* approach where neither unobservable quantities nor infinite sequences of variates are being considered.

2 Examples

1. Estimation of population proportion

Consider the situation where $\mathcal{X} = \{0, 1\}$ and $T = T(X) = \sum_{i=1}^{N} X_i$. Suppose a sample $y = (y_1, \ldots, y_n)$ is available. Let us denote the prior distribution for T by

$$\text{Prob}\,\{T = t\} = a_t \, , \quad \text{for } t = 0, 1, 2, \ldots, N \, .$$

The specification of the distribution of X, given $T = t$, yields the posterior distribution for T, given y, as

$$\text{Prob}\,\{T = t|y\} \propto a_t \text{Prob}\,\{y|T = t\} = a_t \sum_{x: x_{i_j} = y_j} \text{Prob}\,\{X = x|T = t\} \, .$$

We emphasize that the sample $y = (y_1, \ldots, y_n)$ is "available", i.e., it was not necessarily obtained by a lottery - or, if it was, its sampling plan is irrelevant. All the randomness present is due to uncertainty about the x_i-values. The use of Bayes' formula is the coherent way of recalculating probabilities derived from the original prior distribution $f_X(x_1, \ldots, x_N)$. Notice that the specification of this prior is given by a_t and $\text{Prob}_X\{x_1, \ldots, x_N | T = t\}$.

In the example, when $a_t \equiv (N+1)^{-1}$ and X, given $T = t$, is also uniform on the appropriate set, we obtain the posterior

$$P(T = t|y) \propto \binom{N-n}{t - \sum_{i=1}^{n} y_i} \binom{N}{t}^{-1} I\left(t \geq \sum_{i=1}^{n} y_i\right) .$$

One should notice that the "hypergeometric" term on the posterior above results exclusively from the two uniform distributions assumed. There is no assumption of a "draw without replacement" whatsoever. Sampling plans play no rôle.

It should be noticed, however, that the posterior is exactly the same obtained by assuming $X_1, X_2, \ldots, X_N$ extendible to a sequence of exchangeable random quantities with de Finetti measure uniform on $(0,1)$.

Proof: In fact, if $X_1, X_2, \ldots, X_N$ are exchangeable, the conditional distribution of X, given $T = t$, is uniform. On the other hand, if $X_1, \ldots, X_N$ is extendible, de Finetti's Representation Theorem (de Finetti (1937)) yields

$$\text{Prob}\{T = t\} = \int_0^1 \binom{N}{t} \theta^t (1-\theta)^{N-t} d\mu(\theta) ,$$

for each $t = 0, 1, \ldots, N$. Integrating with μ uniform on $\Theta = (0,1)$, we obtain $\text{Prob}\{T = t\} = (N+1)^{-1}$. □

In conclusion, we go back to a superpopulation (Bernoulli) model when using the two assumptions above (with, specifically, uniform "prior" for θ on $(0,1)$). However, suppose now that Prob $\{T = 0\} = 0$, and Prob $\{T = t\} > 0$, for $t \neq 0$, , i.e., 0 is removed from the prior support of T. Now there is no superpopulation model yielding the same distribution Prob $\{T = t|y\}$.

Proof: Suppose, by contradiction, that there is a superpopulation model yielding the same posterior distribution. In other words, assume the existence of a non-degenerate distribution G on $[0, 1]$ such that

$$P(T = t) = \int_0^1 \binom{N}{t} \theta^t (1-\theta)^{N-t} dG(\theta) ,$$

for each $t = 0, 1, \ldots, N$. Then

$$0 = P(T = 0) = \int_0^1 (1-\theta)^N dG(\theta) ,$$

implying that G is degenerate on 1 and $P(T = N) = 1$. □

In conclusion, there are prior distributions Prob $\{T = t\}$ which do not correspond to any (Bayesian) superpopulation model, in the sense that the posterior does not correspond to any attainable from such a model. This also justifies attention to predictivistic Bayesian models.

In general the problem of asking what priors do correspond to a superpopulation model is stated as: what prior values a_t satisfy

$$a_t = \int_0^1 \binom{N}{t} \theta^t (1-\theta)^{N-t} d\mu(\theta), \quad \text{for all } t = 0, 1, \ldots, N ,$$

for some non-degenerate measure μ on $[0, 1]$. This is actually Hausdorff's reduced moment problem whose solution characterizes extendible exchangeable sequences (see Iglesias-Zuazola (1993) and also De Finetti's proof of the representation theorem (1937)).

2. Estimation of population total (discrete case)

Here $\mathcal{X} = Z_+$ (the nonnegative integers) and $T = T(X) = \sum_{i=1}^{N} X_i$. Suppose again that a sample $y = (y_1, \ldots, y_n)$ is available and let

$$\text{Prob}\{T = t\} = a_t \,, \quad \text{for } t \in Z_+ \,.$$

We have again

$$\text{Prob}\{T = t|y\} \propto a_t \sum_{x : x_{i_j} = y_j} \text{Prob}\{X = x|T = t\} \,.$$

Let us also suppose that the distribution of X given $T = t$ is uniform on the set $\{(x_1, \ldots, x_N) \in Z_+^N : \sum_{i=1}^{N} x_i = t\}$. By using standard results [Feller (1968)], we obtain

$$\text{Prob}\{T = t|y\} \propto a_t \frac{\binom{N-n-t-\sum_{i=1}^{n} y_i - 1}{t - \sum_{i=1}^{n} y_i}}{\binom{N-t-1}{t}} I\left(t \geq \sum_{i=1}^{n} y_i\right) \,.$$

It is easy to see that the likelihood is the same when we use the geometric Bayesian superpopulation model. And also, if there exists a non-degenerate probability measure μ on $[0, 1]$ such that

$$a_t = \int_0^1 \binom{N + t - 1}{t} (1 - \theta)^N \theta^t d\mu(\theta) \,, \quad \text{for each } t \in Z_+ \,,$$

then the posterior distribution is exactly the same obtained by assuming $X_1, X_2, \ldots, X_N$ conditionally independent and geometrically distributed with parameter θ with de Finetti measure μ.

This is a condition of extendibility for a class of finite exchangeable sequences. The existence of a non-degenerate measure μ satisfying the equations above can be verified by Hausdorff's Theorems. This can be seen by noting that the equations above have a solution if, and only if, the problem

$$c_t = \int_0^1 \theta^t \, d\nu(t) \,, \quad \text{for } t \in Z_+$$

has a solution for ν, a probability measure on $[0,1]$, with c_t satisfying the relation

$$\binom{N+t-1}{t}^{-1} a_t = \sum_{r=0}^{N} \binom{N}{r} (-1)^r c_{r+t} = \Delta^N c_t \ .$$

Hausdorff's theorem shows (Shohat and Tamarkin (1943)) that the sequence is extendible (in the sense above) if, and only if,

$$\Delta^k a_t = \sum_{r=0}^{k} \binom{k}{r} (-1)^r a_{t+r} \geq 0 \ , \quad \text{for } k = 1, 2, \ldots,; \ t = 0, 1, \ldots$$

It is easy to find prior distributions for T which do satisfy these conditions. On the other hand, it is not simple to find the de Finetti measure once these inequalities are satisfied. In this manner, it is not clear what are the advantages of extendibility.

The applications are not restricted to judgements of uniform distributions for X, given $T = t$ (see, for example, Wechsler (1993)).

3. Estimation of population total (continuous case)

We now consider $\mathcal{X} = \mathbb{R}_+$ and $T = T(X) = \sum_{i=1}^{N} X_i$. Let us denote by $f(t)$ the prior density for T and suppose that the distribution of X, given $T = t$, is uniform on the set $\{(x_1, \ldots, x_N) \in \mathbb{R}_+^N : \sum_{i=1}^{N} x_i = t\}$. It is known that if $X_1, \ldots, X_N$ are independent random quantities with common distribution exponential (1) then the distribution of $X_1, \ldots, X_N$, given $T = t$, is uniform on the appropriate set. Using this fact, we find that the posterior density of T given y is given by

$$f(t|y) \propto \left(1 - \frac{\sum_{i=1}^{n} y_i}{t}\right)^{N-n-1} \left(\frac{\sum_{i=1}^{n} y_i}{t}\right)^{n-1} \frac{1}{t} f(t) I\left(t \geq \sum_{i=1}^{n} y_i\right) .$$

If $f(t) = \frac{a m_0^a}{t^{a+1}} I(t \geq m_0)$, i.e., if T has a Pareto prior distribution (a, m_0), and when $\sum_{i=1}^{n} y_i \geq m_0$, we obtain the posterior density

$$f\left(t \Big| \sum_{i=1}^{n} y_i\right) \propto \left(1 - \frac{\sum^n y_i}{t}\right)^{N-n-1} \left(\frac{\sum^n y_i}{t}\right)^{n+a-1} \frac{1}{t} I\left(t \geq \sum^{n} y_i\right) .$$

Note that this posterior does not correspond to any density obtained from an exponential superpopulation model.

A probability interval for T, given y, can be constructed by using tables of the Beta $(n+a, N-n)$ distribution (with abcissas B_α):

$$\text{Prob}\left\{\frac{\sum^n y_i}{B_{1-\alpha/2}} < T < \frac{\sum^n y_i}{B_{\alpha/2}} | y\right\} = 1-\gamma \ .$$

The mode of the posterior density above is

$$\hat{T}_1 = \left(\frac{N+a-1}{n+a}\right)\sum_{i=1}^n y_i$$

and its posterior mean is

$$\hat{T}_2 = \frac{(N+a-1)}{(n+a-1)}\sum^n y_i \ .$$

Notice that for $a = 1$, $\hat{T}_2$ coincides with the usual total predictor of simple random sampling.

The examples above refer to estimation of population totals for finite sequences of exchangeable variates. Under a further judgement of linearity of posterior mean, i.e.,

$$E\left(T|\sum_{i=1}^n y_i\right) = a + b\sum_{i=1}^n y_i \ ,$$

Ericson's theorem (Ericson (1969), p.323) can be used to determine a and b, which depend on the first and second prior moments of T only. In particular, for a and b given on Example 1, the prior distribution of T is determined. This can be proved by Ericson's theorem and the rule of succession for finite sequences (Zabell (1989), de Finetti (1937)). Diaconis and Ylvisaker (1979, pp.279–280) obtain a similar result for infinite sequences.

4. Estimation of population maximum

Consider $\mathcal{X} = \mathbb{Z}_+$ and $T = T(X) = X_{(N)}$, the population maximum. A sample $y = (y_1, \ldots, y_n)$ is available and let a_t denote the prior distribution for T, with the distribution of X, given $T = t$, uniform on the set $\{(x_1, \ldots, x_N) \in Z_+^N : x_{(N)} = t\}$.

The posterior distribution is given by

$$\text{Prob}\{T=t|y\} = \begin{cases} \dfrac{1}{(t+1)^n}\left\{\dfrac{1}{1+(\frac{t}{t+1})^N}\right\}a_t & \text{if } t = y_{(n)} \\ \dfrac{1}{(t+1)^n}\left\{\dfrac{1-(\frac{t}{t+1})^{N-n}}{1-(\frac{t}{t+1})^N}\right\}a_t & \text{if } t > y_{(n)} \end{cases} .$$

The mode of the posterior above is $\widehat{T}_1 = y_{(n)}$ whenever a_t decreases in t. Extendibility for a mixture of discrete uniform distributions obtains if, and only if, there exists a non-degenerate probability measure μ on Z_+ such that

$$a_t = \int_{\theta \geq t} \frac{(t+1)^N - t^N}{(t+1)^N} \, d\mu(\theta) \ , \quad \text{for each } t \in Z_+ \ .$$

Such a measure μ exist if, and only if, the function defined by

$$h(t) = (t+1)^N \left\{ \frac{a_t}{(t+1)^N - t^N} - \frac{a_{t+1}}{(t+2)^N - (t+1)^N} \right\} \ , \quad t \in Z_+$$

defines a probability measure on Z_+. Using this fact, it can be shown that if $a_M = 0$ for some $M \in \mathbb{N}$ and $a_t > 0$ for each $t \neq M$, the sequence is not extendible for the uniform model of superpopulation.

3 Conclusion

The examples show that Bayesian superpopulation models do not necessarily exist for every prior $f_X(x)$.

We indicate a way for making statistical inference on finite populations under a strict subjectivistic approach (de Finetti (1975, v.1)). As sampling plans are unnecessary in this approach, it can be used when information is based on intentional samples, providing a formal inference for the situation. One might choose samples y having more "interesting" expected likelihood functions $L_y(t)$, for instance.

A more general setting would be given by consideration of loss functions $U(y, X)$ (see Kadane and Sedransk, p.335, for a similar approach).

References

Basu, D. (1969). Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyã A*, **31**: 441–454.

Cochran, W.E. (1977). *Sampling techniques.* 3.ed. John Wiley, N. York.

de Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henry Poincaré*, v.7, p.1–68. Translated in Kyburg Jr., H.E.; Smokler, H.E. (1964). Studies in subjective probability. John Wiley & Sons.

de Finetti, B. (1975). *Theory of probability.* J. Wiley, N. York.

Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Ann. Statist.*, **7**:269–281.

Ericson, W.A. (1969a). Subjective Bayesian models in sampling finite population (with discussion). *J. Roy. Statist. Soc.*, **31**: 195–224.

Ericson, W.A. (1969b). A note on the posterior mean of a population mean. *J. Roy. Statist. Soc.*, **31**:332–324. ???

Feller, W. (1968). *An introduction to probability theory and its applications.* v.1. J. Wiley, N. York.

Iglesias-Zuazola, P. (1993). *Finite forms of the de Finetti's theorem: a predictivistic approach to statistical inference in finite populations.* (In portuguese) Doctoral Thesis. Instituto de Matemática e Estatística, Universidade de São Paulo.

Kadane, J.B. and Sedransk, N. (1982). Toward a more ethical clinical trial. In *Bayesian Statistics*, **1**:339–346.

Smith, T.M.F. (1976). The foundations of survey sampling: a review (with discussion). *J. Roy. Statist. Soc.*, *A*, **39**:183–204.

Shohat, J.A and Tamarkin. J.D. (1943). The problem of moments. *Amer. Math. Soc.*, N. York.

Wechsler, S. (1993). Exchangeability and predictivism. *Erkenntnis – International Journal of Analytic Philosophy*, **38**: 343–350.

Zabell, S.L. (1989). The rule of sucession. *Erkenntnis – International Journal of Analytic Philosophy*, **31**:283–321.

ULTIMOS RELATORIOS TECNICOS PUBLICADOS

9401 - PAULA, G. A., SEN P.K., Tests of ordered Hypotheses in Linkage in Heredity, 12p.

9402 - PAULA, G. A., SEN P. K. One-Sided Tests in Generalized Linear Models with Parallel Regression Lines, 18p.

9403 - PAULA, G. A., .Influence Diagnostics in Proper Dispersion Models,22 p.

9404 - SALINAS-TORRES, V.H. & PEREIRA, C.A.B. and TIWARI, R.C. - Convergence of Dirichlet Measures Arising in Context of Bayesian Analysis of Competing Risk Models, 17p.

9405 - CRIBARI-NETO, F. & FERRARI, S.L.P. - Second Order Asymptotics for Score Tests in Generalized Linear Models, 25p.

The complete list of Relatórios do Departamento de Estatística, IME-USP, will be sent upon request.

- Departamento de Estatística
IME-USP
Caixa Postal 20.570
01452-990 - São Paulo, Brasil