«ΣΠΟΥΔΑΙ», Τόμος 47, Τεύχος Ιο-2ο, Πανεπιστήμιο Πειραιώς / «SPOUDAI», Vol. 47, No 1-2, University of Piraeus

ESTIMATION OF A POPULATION DISTRIBUTION BASED ON CLASSIFICATION PROBABILITY MODEL: APPLICATION IN THE CASE OF HUNTING GUN CRIMES IN GREECE

By

Vassilios Hombas

University of Athens

Abstract

This work has a goal to suggest two iteration processes and the sufficient conditions under which these processes converge and have useful properties. In other words it investigates the probability and the sufficient conditions that have to be satisfied. The use of maximum likelihood equation as well as of the iteration processes that are usually applied is justified only in large samples. An original sample of hunting gun crimes in Greece led the author, based on a discrete probability model, to estimate the distribution of five categories of crimes. Some comparisons are made between the suggested methods concerning the speed of their convergence and accuracy of their results. (JEL C13)

1. Introduction

There has been much discussion in the old and recent statistical literature concerning the optimal way of formulating a model, in an attempt to estimate some parameters in it, fitting it and then assessing the goodness of fit of the data to the model.

Gani (1989) has commented on various aspects of statistics, considering prediction as the one having special importance. Field (1984) favors the view that one should finish the course being able to set up models, test hypotheses or estimate parameters and interpret the results in the classical statistical setting.

In this paper we will estimate the distribution of a population with the help of a sample and a probability model, especially a multinomial model. A standard approach to parameter estimation for such a model is the method of maximum likelihood.

The exact determination of maximum likelihood estimate (MLE) is often quite difficult. In such cases (MLE) is approximated by using some iteration process, the usual being the method of scoring for parameters as given by Rao (1973, pp. 165-167). If $\hat{\theta}_0$ is the trial solution of the likelihood equation, θr the r^{th} iterate and $\hat{\theta}$ the MLE, then the use of an iteration process is justified if the error $|\theta_r - \hat{\theta}|$ decreases with increased iterations and tends to zero as r tends to infinity. We consider two iteration processes in large samples at least.

2. Estimation Methods

Under the regularity conditions the following well-known results have already been proved by Cramér (1974).

Result 1. With probability approaching certainty as n, the size of the sample, tends to infinity the likelihood equation admits a consistent solution.

Result 2. $(1/n)\ (\partial^2 \log L/\partial\ \theta^2)\ \theta = \hat{\theta}$ converges in probability, to $-I(\theta_0)$ as $n \to +\infty$, where L is the maximum likelihood function and $I(\theta)$ is ths well known Fisher's information function.

Result 3. The consistent solution of the likelihood equation is unique and $P\ [\partial^2 \log L/\partial\theta^2)\ \theta = \hat{\theta}<0] \to 1$ as $n \to +\infty$.

Let $\psi(\theta)$ be a differential function of θ, which has no zero in a neighbourhood of $\hat{\theta}$, the root of likelihood equation which we assume to exist.

Define

$$\varphi(\theta) = \theta - \psi(\theta)\ \frac{\partial \log L}{\partial\theta} \tag{1}$$

Consider the iteration process $\theta_{r+1} = [\varphi(\theta)]_{\theta=\theta_r}$ or $\theta_{r+1} = \theta_r - \psi(\theta_r)\ (\partial \log L/\ \partial\theta)_{\theta=\theta_r}$. Let $e_r = |\ \theta_r - \hat{\theta}|$ be the error at the rth iteration, then the choice of $\psi((\theta)$ is to be made in such a way that $e_{r+1} < e_r$ and $e_r \to 0$. We will consider two different iteration processes corresponding to the different choices of $\psi(\theta)$.

(A) First consider the Newton-Raphson process, where $\psi(\theta) = -[\partial^2 \log L/\partial\theta^2]^{-1}$, then by (1) we obtain

$$\psi(\theta) = \theta + \frac{\partial \log L}{\partial\theta} \Big/ \frac{\partial^2 \log L}{\partial\theta^2} \tag{2}$$

Results 1, 2, 3 and the Householder's conditions guarantee that if $\hat{\theta}_0$, the initial estimate, is consistent then the process converges with probability approaching unity as n tends to infinity.

(B) Now consider the method of scoring for parameters (SOP) Here

$$\psi(\theta) = \frac{1}{nI(\theta)}, \text{ where } I(\theta) = E\left(- \frac{\partial^2 \log L}{\partial^2 \theta} \right) \text{ and } 0 < I(\theta) < \infty$$

Now $[I(\theta)]^{-1}$ has no zero in the neighbourhood of $\hat{\theta}$ and $I(\theta)$ is differentiable.
We have

$$\varphi(\theta) = \theta + \frac{\partial \log L}{\partial \theta} \Big/ nI(\theta) \tag{3}$$

In 1950 Frechet shows (see John F. (1991) that the above method converges with probability approaching unity as $n \to +\infty$.

By way of comparison of the two above iteration processes, only the N-R is of second order, while the (SOP) is of first order. The N-R process is accordingly rapidly convergent. However this process usually involves more cumbersome calculations than any other. The method of (SOP) can be applied if $I(\theta)$ is differentiable for θ which is usually the case.

3. Application

In order to achieve much of the above and to compare the methods I thought of carrying out this program on the basis of a real random sample which two Greek members of an ecological organization Tsirimokou and Gouras (1991) "fished" daily from a greek newspaper named "Eleftherotypia". The sample presents the crimes commited in Greece with hunting guns during the period 1985-1987. In Table 1 five categories of such accidents are presented.

TABLE 1
Crimes with hunting guns in Greece 1985-1987

Inguries	Suicides	Unwilfull Deaths	Murders	Threats
640	30	29	67	514

n = 1280.

Of interest is thequestion: What is the estimate of the probability that one of the above types of accidents happens.

The data set has led us to choose the multinomial random vector, where x_i i = 1, 2, 3, 4 are the frequencies of the five classes. The five multinomial probabilities are specified with the help of one parameter $\left[\frac{7-9\theta}{11}, \frac{\theta}{11}, \frac{\theta}{11}, \frac{3\theta}{11}, \frac{4+4\theta}{11}\right]$ with admissible parameter $0 \leq \theta \leq 7/9$.

A standard approach to parameter estimation for such a model is the method of maximum likelihood. For families of distributions satisfying appropriate reqularity conditions, standard large sample results guarantee the existence of solutions to the likelihood equations that are consistent, efficient and asympotically normal (see Lehman 1980 and Cox 1984).

The distribution of $\underline{X}$ is given by

$$f(\underline{X}) = n! \prod_{i=1}^{5} P_i^{x_i} / x_i ! \tag{4}$$

for nonnegative integers x_i, i = 1, 2, 3, 4, 5 satisfying $\Sigma x_i = n$ and $\Sigma p_i = 1$. (see Davis and Jones (1992) or Agresti (1990), chapter 3).

Equation (4) is exactly the likelihood function $L(x, \theta)$. Since $L(x, \theta)$ and $\ln L(x, \theta)$ are maximized for the same value of the parameter θ, we have $\ln L(\theta) = \ln H + x_1 \ln(7-9\theta) + x_2 \ln\theta + x_3 \ln\theta + x_4 \ln 3\theta + x_5 \ln(4+4\theta)$ where H is a quantity independent of θ.

Therefore the maximum likelihood equation (MLE) is

$$\frac{-9x_1}{7-9\theta} + \frac{x_2}{\theta} + \frac{x_3}{\theta} + \frac{x_4}{\theta} + \frac{4x_5}{4+4\theta} = 0 \tag{5}$$

Introducing the observed frequencies x_i, i = 1, 2, 3, 4, 5 from the Table 1 into (5), we obtain the quadratic equation

$$5760\theta^2 + 1207\theta - 441 = 0 \tag{6}$$

Such cases can be handled analytically using formulae of a quatratic equation. Another question has been raised here: What is going on if the (MLE) is of third or more? The answer in this case is that the roots can be handled with the help of an iteration process like N-R (see J. Stoer and R. Bulirsch (1980)), scoring for parameters cases we saw in (A) and (B) of this paper. The equation

(6) has two roots, one positive and one negative. The positive root is the required of $\hat{\theta}$, namely 0.19109755. As the exact value of $\hat{\theta}$ is available, the comparison of the methods was easier.

The N-R iteration procedure is

$$\theta_{r+1} = \theta_r + \left(\frac{\partial \mathrm{InL}}{\partial \theta}\right)_{\theta=\theta_r} \Bigg/ \left(-\frac{\partial \mathrm{I}^2\mathrm{nL}}{\partial \theta^2}\right)_{\theta=\theta_r} \tag{7}$$

We need a starting point. Knowing that $E(x_i) = np_i$ and $v(x_i) = np_i(1 - p_i)$, $i = 1, 2, 3, 4, 5$, we propose as a starting point.

$$\hat{\theta}_0 = \frac{1}{n}(-4x_1 + 300x_2 - 278x_3 - 25x_4 + 7x_5) = 0.23515625$$

Note that this is consistent since, as easily can be shown,

$E(\hat{\theta}_0) = \theta$ and $V(\theta_0) = 0\left\langle\frac{1}{n}\right\rangle$.

We now need the quantities $\frac{\partial \ln L}{\partial \theta}$, $\left(\frac{-\partial^2 \ln L}{\partial \theta^2}\right)$ and $I(\theta)$. They are calculated for the data concerned.

The first is easy to compute

$$\left(\frac{\partial \ln L}{\partial \theta}\right)_{\hat{\theta}} = \frac{-9x_1}{7 - 9\hat{\theta}_0} + \frac{x_2 + x_3 + x_4}{\hat{\theta}_0} + \frac{4x_5}{4 + 4\hat{\theta}_0} = -228.3743 \tag{8}$$

The second comes from the first by differentiating and changing the sign.

$$\left(-\frac{\partial \mathrm{I}^2\mathrm{nL}}{\partial \theta^2}\right)_{\hat{\theta}_0} = \frac{81x_1}{(7-9\hat{\theta}_0)^2} + \frac{x_2 + x_3 + x_4}{\hat{\theta}_0^2} + \frac{16x_5}{(4+4\hat{\theta}_0)^2} = 5542.6 \tag{9}$$

The third comes from (9) taking its expected value

$$nI(\theta_0) = nE\left(-\frac{\partial \mathrm{I}^2\mathrm{nL}}{\partial \theta^2}\right)_{\hat{\theta}_0} = \frac{n^2}{11}\left[\frac{81}{7-9\hat{\theta}_0} + \frac{5}{\hat{\theta}_0} + \frac{16}{4+4\hat{\theta}_0}\right] = 6129105.453$$

Substituting $\hat{\theta}_0$ in (2) we find the (N - R) θ_1 iterate

$\theta_1 = 0.23515625 - 228.3743/\ 5542.6 = 0.193952794$ while the (SOP) θ_1 iterate is $\theta_1 = 0.23515625 - 228.3743/6129105.453 = 0.23511898$.

Using computer we can obtain the following Table 2 which shows the successive iterates accordindy. The entry Adjustment gives the value of correction to θ_4. The "Error" gives the value of $|\theta_5 - \hat{\theta}_0|$.

TABLE 2.
Successive iterations by two different methods listed above

Iterates	I (N - R)	II (SOP)
θ_0	0.23515625	0.23515625
θ_1	0.193952794	0.23511899
θ_2	0.191111196	0.20235096
$*\theta_3$	0.191097551	0.19403251
θ_4	0.191097550	0.19203567
θ_5	0.19097550	0.191097550*

Adjustment = 0.000001122
Error = 0.040587

In Table 3 we finally illustrate the probabilities of belonging to the five classes taking into account that $\hat{\theta} = 0.19109755$.

TABLE 3
Probabilities of crimes with hunting guns in Greece

Injuries	Suicides	Unwilfull Deaths	Murders	Threats
48&	1.74&	1.74&	5.21&	43.31&

A chi-square goodness of fit test at the level 2,5& gives $x^2 = 8.72 < x^2_{0025,3} = 9.348$. Hence, the null hypothesis cannot be rejected. In other words, the model provides a "good fit". (see Freund and Simon (1992) p. 380).

4. Conclusion

By means of the above somewhat peculiar and interesting sample, one manages to touch upon topics like setting a parametric model, estimating via (MLE) method the parameter, testing of hypothesis and interpretation the results.

Furthermore we saw that in large samples application of iterative processes is made on the MLE. The convergence of these processes is assured by the work of Lehmann (1980). Also we see that the (N-R) method is rapidly convergent and is asymptotically better than the method (SOP). The true value of root springs up at the 3rd iterate of the (N-R) method, while in the (SOP) method at the 5th iterate. Note that for purposes of simplicity we constructed the above model so that no third order or more equation arises.

Finally Table 3 shows that out of 100 greek persons hit by hunting guns almost 48 are injured 2 commit suiside, 2 have unwilfull deaths, 5 are murdered and 43 receive threats.

References

Agresti, A. (1990). *Categorical Data Analysis,* John Wiley, New York.

Cox, C. (1984). Elementary Introduction to Maximum Likelihood Estimation for Multinomial Models: Birch's Theorem and the Delta Method. *The American Statistician.* November, Vol. 38, No 4, pp. 283-287.

Cramer, H. (1974) *Mathematical Methods of Statistics* (13th printing). Princenton University Press.

Davis, C. and Jones, M. (1992) "Maximum Likelihood Estimation for the Multinomial Distribution. *Teaching Statistics,* Volume 14, No. 3. Autumn.

Field, C.A. (1984). A "Reverse Order" Elementery Statistics Course. *The American Statistician,* Vol. 38. No 2, pp. 117-119.

Freud, J. and Simon, G. (1992) *Modern Elementary Statistics* 8th ed., Englewood Cliffs, New Jersey.

Gani, J. (1989). The Many Faces of Statistics. *Teaching Statistics,* Vol. 11, No 3. Autumn, pp. 71-74.

John, F. (1991). *Lectures on Numerical Analysis* Gordon and Breach, New York.

Lehmann, E. L. (1980). Efficient Likelihood Estimators. *The American Statistician,* 34, pp. 233-235.

Rao, C. R. (1973), *Linear Statistical Inference and its Applications* 2nd ed John Wiley and Sons Inc. New York.

Stoer, J. and Bulirsch, R. (1980) *Introduction to Numerical Analysis,* Springer and Verlag, New York, p. 246.

Tsirimokou, G. and Gouras, G. (1991). Statistics Newspaper *"Eleftherotypia",* October 31, p. 9.