Generalized Control Limits of George Obreja Type

Florin-Cristian Gheorghe, George Obreja+, Marin-Nicolae Popescu

Abstract We apply the estimation theory to *dependent* random variables when these variables form a simple and constant Markov chain. This type of study was initiated by F. Chartier. We are able to obtain maximum likelihood estimations for the chain's parameters. Finally we obtain the control limits for the average value in the case of dependent variables.

I. Introduction

Let $x_0, x_1, \ldots, x_n$ be a simple and constant Markov chain (the random variables are real). We shall assume that these variables form an *ordered selection* and follow a theoretical probabilistic law depending upon the real parameters $m, \sigma > 0$ and $-1 < \rho < 1$.

More precisely, the transition probabilities

$$P\left(x_i \in A \,\middle|\, x_{i-1} = \xi\right)$$

have the real densities of the form:

$$f(x_{i-1}, x_i, m, \sigma, \rho) = \frac{1}{\sqrt{2\pi}\,\sigma\sqrt{1-\rho^2}} \exp\left(-\frac{\left[(x_i - m) - \rho(x_{i-1} - m)\right]^2}{2\sigma^2(1-\rho^2)}\right)$$

(1)

We shall estimate the parameters m, σ, ρ.

According to Billingsley, we shall take log-likelihood of the selection $(x_0, x_1, \ldots, x_n)$:

$$L_n(m, \sigma, \rho) = \sum_{i=1}^{n} l_n f(x_{i-1}, x_i, m, \sigma, \rho) \qquad (2)$$

for sufficiently large n.

We shall see in the discussion that follows that in some particular cases one can obtain estimations with remarkable properties.

The maximum likelihood system is obtained (for fixed n) via differentiation of $L_n(m, \sigma, \rho)$ with respect to m, σ, ρ.

Namely, this system will be:

$$\frac{\partial L_n(m, \sigma, \rho)}{\partial m} = \frac{\partial L_n(m, \sigma, \rho)}{\partial \sigma} = \frac{\partial L_n(m, \sigma, \rho)}{\partial \rho} = 0$$

(3)

According to (1) and (2), the solution of (3) will be given by the following formulae:

$$\begin{cases} m = \dfrac{\sum_{i=1}^{n} x_i - \rho \sum_{i=1}^{n} x_{i-1}}{n(1-\rho)} \\ \sigma^2 = \dfrac{1}{n(1-\rho^2)} \sum_{i=1}^{n} \left[(x_i - m) - \rho(x_{i-1} - m)\right]^2 \\ \rho = \dfrac{\sum_{i=1}^{n} (x_i - m)(x_{i-1} - m)}{\sum_{i=1}^{n} (x_{i-1} - m)} \end{cases} \qquad (4)$$

We shall take into account three particular cases:

a) Assuming that σ^2 and ρ are known, we obtain the following estimation for m:

$$\overline{m} = \frac{\sum_{i=1}^{n} x_i - \rho \sum_{i=1}^{n} x_{i-1}}{n(1-\rho)},$$

Which is an absolutely correct and efficient estimation of parameter m.

b) Now assuming m and ρ being known, we obtain the following estimation for σ^2 :

$$\overline{\sigma}^2 = \frac{1}{n(1-\rho^2)} \sum_{i=1}^{n} \left[(x_i - m) - \rho(x_{i-1} - m)\right]^2$$

which is an absolutely correct estimation of σ^2.

c) Finally, assuming m and σ being known, we obtain the following estimation for ρ (using R.A. Fischer's results about the estimation of the covariance):

$$\overline{\rho} = \frac{n \sum_{i=1}^{n} x_{i-1} x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_{i-1}}{n \sum_{i=1}^{n} x_{i-1}^2 - \left(\sum_{i=1}^{n} x_{i-1}\right)^2}$$

which is also absolutely correct.

ISBN:978-988-17012-3-7

Now we return to *the general case*. Resolving the system (3) according to m, σ and ρ we obtain the following estimation of maximum likelihood:

$$\overline{\rho} = \frac{n\sum_{i=1}^{n} x_{i-1}x_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} x_{i-1}}{n\sum_{i=1}^{n} x_{i-1}^2 - \left(\sum_{i=1}^{n} x_{i-1}\right)^2} \qquad (5)$$

$$\overline{m} = \frac{\sum_{i=1}^{n} x_i - \overline{\rho}\sum_{i=1}^{n} x_{i-1}}{n(1-\overline{\rho})} \qquad (6)$$

(with $\overline{\rho}$ as above)

$$\overline{\sigma}^2 = \frac{1}{n\left(1-\overline{\rho}^2\right)}\sum_{i=1}^{n}\left[(x_i - \overline{m}) - \overline{\rho}(x_{i-1} - \overline{m})\right]^2 \qquad (7)$$

(with $\overline{\rho}$ and $\overline{m}$ as above).

So, everything can be expressed in terms at x_0, $x_1, \dots, x_n$.

We demonstrated that the estimations (5), (6) *and* (7) *are consistent (they converge in the probability to the parameters m,* σ^2 *and ρ). We observe that these estimations are different from those obtained if we assume the* x_0, $x_1, \dots, x_n$ *are independent in the probabilities.*

II. Generalized limits of control type George Obreja

We assume that σ *and* ρ *are known from the other studies regarding the fabrication process (calculated using the equations* (5), (6) *and* (7)*, using many trials). Then we show that the estimation* $\overline{m}$ *of medium value m is an absolute correct and efficient estimation, which tends at the limit to a normal repartition.*

We have

$$M(\overline{m}) = m\,;\; D(\overline{m}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{1+\rho}{1-\rho}}$$

and the statistic

$$Z_n = \frac{(\overline{m}-m)\sqrt{n}}{\sigma\sqrt{\frac{1+\rho}{1-\rho}}} \to N(0,1)$$

The control limits for the verification of the hypothesis $H_0 : m = T_0$ *(*T_0 *being the midpoint of the tolerance interval) with the alternative* $H_1 : m \neq T_0$ *result from the equation*

$$P\left(|Z| < Z_{1-\frac{\alpha}{2}}\right) = 1-\alpha \qquad (8)$$

where α *is the risk of the first order.*

Equation (8) *leads to the following conclusions:*

a) we accept the hypothesis H_0 *with the probability* $1-\alpha$*, if*

$$T_0 - Z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\sqrt{\frac{1+\rho}{1-\rho}}\right) \le \overline{m} \le T_0 + Z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\sqrt{\frac{1+\rho}{1-\rho}}\right) \qquad (9)$$

b) we reject the hypothesis H_0 *and we accept the hypothesis* H_1 *with the probability equal to* α *in the opposite case.*

In conclusion we obtain the limits of generalized control called George Obreja, in the memory of our well known colleague. These limits are:

$$Lcs = T_0 + Z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\sqrt{\frac{1+\rho}{1-\rho}}\right)$$

$$Lci = T_0 - Z_{1-\frac{\alpha}{2}}\left(\frac{\sigma}{\sqrt{n}}\sqrt{\frac{1+\rho}{1-\rho}}\right)$$

(10)

We observe that for $\rho = 0$*, the equations become:*

$$Lcs = T_0 + Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

$$Lci = T_0 - Z_{1-\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$$

which represent classical limits of control, calculated with the hypothesis that the observations are independent.

Observations. The limits of generalized control depend also on the correlation coefficient ρ*, which takes into account the strength of the liaison between the variables of the chain.*

III. Example

Checking the length of a set of industrial products, the standard (accept length) was 42.5 cm with tolerances: +0.2 cm and -0.3 cm. So, the acceptable limits are *l*=42.2 cm, L=42.7 cm.

Observing 100 items (in the order of their manufacture), the observed values $x_0, x_1, \dots, x_n$ (e.g. $x_0 = 42{,}40$, $x_1=42{,}45 \dots$, $x_{99}=42{,}79$) were tabulated.

Using the formulae (5), (6) (7) for $\overline{\rho}, \overline{m}, \overline{\sigma}^2$ successively, the following values were obtained:

$$\overline{\rho} \cong 0{,}966,$$
$$\overline{m} \cong 42{,}42,$$
$$\overline{\sigma}^2 \cong 0{,}37747 \Rightarrow \overline{\sigma} \cong 0{,}193$$

Taking $\alpha = 0{,}05$.

The centre of the confidence interval is $T_0 = \frac{42{,}7 + 42{,}2}{2} = 42{,}45$

Therefore

$$-Z_{1-\frac{\alpha}{2}} \le Z = \frac{\overline{m} - T_0}{\overline{\sigma}\sqrt{\frac{1+\rho}{1-\rho}}} \le Z_{1-\frac{\alpha}{2}}$$

and $Z_{1-\frac{\alpha}{2}} = Z_{0,975}$ must be such that

$$\int_{-Z_{1-\frac{\alpha}{2}}}^{Z_{1-\frac{\alpha}{2}}} \Phi(x)dx = 1 - \alpha = 0,95$$

i.e. $\frac{2}{\sqrt{2\pi}} \int_0^{Z_{1-\frac{\alpha}{2}}} \exp\left(-\frac{x^2}{2}\right)dx = 0,95$

So $Z_{0,975} = 1,64$.

So, the control limits are:

$$l_1 = 42,45 - 1.64 \cdot 0,136 \cong 42,230$$

$$L_1 = 42,45 + 1,64 \cdot 0,136 \cong 42,670$$

(accepted from standard point of view), with significance level 0, 05.

IV. Notes

A. One can previously estimate $\rho = \overline{\rho}$ and $\sigma = \overline{\sigma}$ and get $m = \overline{m}$. Then one can compute the statistic Z.

B. For $\rho = 0$ we obtain the case of the classical independent (more precisely uncorrelated) observations and the classical limits are $l_1^{/} = 42,42$ and $L_1^{/} = 42,68$.

References

1. P., Billinsley, Statistical Inference for Markov Chains, University of Chicago Press, Chicago (1961).
2. M., Craiu, Gh. Mihoc, Statistical Inference for dependent variables, Ed. Academiei RSR,1972 (in Romanian).
3. F., Chartier, Estimation statistique dans le cas d'observation non independantes (These) Paris, 1947.
4. G Obreja, FC Gheorghe Determination of the statistical control limits for dependent variables, Revista de Statistica, 1974, no 11, p 28-32, (article in Romanian)
5. F.C.Gheorghe Markov chains in demography, doctoral thesis, 1974 Centrul de Statistica al Academiei Romane (in Romanian)