PROBABILITY
AND
MATHEMATICAL STATISTICS

Vol. 8 (1987), p. 111–116

THE JACKKNIFE METHOD AND THE GAUSS-MARKOV ESTIMATION

BY

CHRISTIAN LAVERGNE AND JEAN-RENÉ MATHIEU (TOULOUSE)

Summary. It is shown that in many cases the Jackknife estimator is the least squares estimator in a convenient linear model; moreover, under some light assumptions about the distribution of the observations, this least squares estimator is also the Gauss-Markov estimator.

1. Introduction. Let $x = (x_1, \ldots, x_n)$ be a sample so that x_i is an observation of a random variable X_i whose distribution depends on a real unknown parameter θ to be estimated.

$T = T(x_1, \ldots, x_n)$, $n > 1$, denotes an estimator of θ; for each $i = 1, \ldots, n$, T_{-i} denotes the corresponding estimator of θ based on the sample where the i-th observation is deleted; T_J denotes the Jackknife estimator defined by

$$T_J = nT - \frac{n-1}{n}\sum_i T_{-i}.$$

If, for each θ, $E_\theta(T) = \theta + \delta/n$, where δ is unknown and is constant with respect to n, then it is well known that T_J is a unbiased estimator of θ (e.g. see [1], [5]).

In the following, a slightly more general situation is considered; so the following definition is introduced:

Definition. T is an *α-biased estimator of θ* if α exists, α known and strictly positive, so that $E_\theta(T) = \theta + \delta/n^\alpha$, where δ does not depend on n and is unknown.

2. The Jackknife method and the linear model. T being an α-biased estimator of θ, let us consider the linear model

(1) $\quad \mathcal{T} = A\beta + \varepsilon, \qquad \mathcal{T} = (T, T_{-1}, \ldots, T_{-n})' \in R^N,\ N = n+1,\ \beta = (\theta, \delta)',$

A is the $(N\times 2)$-matrix whose columns are $\mathbf{1}=(1,\ldots,1)'$ and $a_\alpha=(1/n^\alpha, 1/(n-1)^\alpha,\ldots,1/(n-1)^\alpha)'$.

Moreover, $E[\varepsilon]=0$.

In the following, $E[\varepsilon\varepsilon']$ is denoted by Γ and the subspace of $\mathbf{R}^N$ spanned by the vectors $\mathbf{1}$ and a_α is also denoted by A.

THEOREM 1. *If T is a α-biased estimator of θ, the least square estimator of θ in the linear model* (1), *defined as*

$$\hat{\theta}=\frac{1}{n^\alpha-(n-1)^\alpha}\Big(n^\alpha T-\frac{(n-1)^\alpha}{n}\sum_{i=1}^{n}T_{-i},$$

is an unbiased estimator of θ.

In the particular case $\alpha=1$, $\hat{\theta}=T_J$.

Proof. The model (1) is a simple linear regression model; so $\hat{\theta}=(A'A)^{-1}A'\mathscr{T}$. But a more simple way to calculate $\hat{\theta}$ is to remark that A is spanned by the vectors $e_1=(1,0,\ldots,0)'$ and $e_2=(0,1,\ldots,1)'$. Hence the orthogonal projection of $\mathscr{T}$ on A, say $\mathscr{T}_A$, with respect to the inner product, defined by the identity matrix, satisfies:

$$\mathscr{T}_A=\hat{\theta}\mathbf{1}+\hat{\delta}a_\alpha,\quad (\mathscr{T}-\mathscr{T}_A)'\cdot e_1=0,\quad (\mathscr{T}-\mathscr{T}_A)'\cdot e_2=0.$$

Then

$$T=\hat{\theta}+\frac{\hat{\delta}}{n^\alpha}\quad\text{and}\quad \frac{1}{n}\sum_i T_i=\hat{\theta}+\frac{\hat{\delta}}{(n-1)^\alpha},$$

from which the expression of $\hat{\theta}$ follows.

Now it is interesting to examine whether $\hat{\theta}$ is the best linear unbiased estimator of θ in the model (1); to do that it is necessary to make some additional assumptions: $\mathscr{M}_N$ denotes the set of the non-singular $(N\times N)$-matrices so that

$$\Gamma\in\mathscr{M}_N\Leftrightarrow\Gamma=\begin{bmatrix} v_1 & c_1 & \cdots & \cdots & c_1\\ & v_2 & c_2 & \cdots & c_2\\ & & \ddots & & \vdots\\ & & & \ddots & c_2\\ & & & & v_2\end{bmatrix}.$$

THEOREM 2. *If T is an α-biased estimator of θ and if Γ belongs to $\mathscr{M}_N$, then $\hat{\theta}$ is the Gauss-Markov estimator of θ in the linear model* (1).

Proof. According to [2], the least squares estimator and the Gauss-Markov estimator are identical if and only if $\Gamma A\subseteq A$ (this results holds even if Γ is singular). Now this assumption is fulfilled if Γ belongs to $\mathscr{M}_N$, for, in this case,

$$\Gamma e_1 = (v_1, c_1, \ldots, c_1)' = v_1 e_1 + c_1 e_2 \in A,$$

$$\Gamma e_2 = (nc_1, v_2+(n-1)c_2, \ldots, v_2+(n-1)c_2)' = nc_1 e_1 + [v_2+(n-1)c_2] c_2 \in A.$$

In the most frequent case $X_1, \ldots, X_n$ are independent and identically distributed and T is a symmetric function of $X_1, \ldots, X_n$; hence the assumption $M \in \mathscr{M}_n$ is fulfilled.

Remark 1. As a consequence of the above propositions, if T_1 and T_2 are some α-biased estimators of respectively θ_1 and θ_2, then the Jackknife estimator of $\theta_1+\theta_2$ is $\hat{\theta}_1+\hat{\theta}_2$.

Remark 2. Let us assume that Γ belongs to $\mathscr{M}_n$. It is easy to see that, for every vector u belonging to $A^{\perp}$, the orthogonal complement of A, $u = cu$ with $c = v_2 - c_2$ (c is strictly positive if the correlation coefficient of T_{-i} and $T_{-i'}$, $i \neq i'$, is different from 1). Then, following theorem 3 of Kruskall [2], $c\|\mathscr{T} - A\hat{\beta}\|^2_{\Gamma^{-1}} = S^2$ with $S^2 = \sum_i (T_{-i} - \bar{T})^2$, $\bar{T} = n^{-1} \sum_i T_i$. Hence, if $\mathscr{T}$ is multinormally distributed, S^2/c is independent from $\hat{\beta}$ and is χ^2_{n-1}; then

$$\frac{\hat{\theta}-\theta}{\sqrt{\dfrac{S^2 \operatorname{Var}(\hat{\theta})}{c(n-1)}}} \sim t_{n-1}$$

$$\operatorname{Var}(\hat{\theta}) = \frac{1}{[n^\alpha - (n-1)^\alpha]^2} \left[n^{2\alpha} v_1 + \frac{(n-1)^{2\alpha} v_2}{n} - 2[n(n-1)]^\alpha c_1 + \frac{(n-1)^{2\alpha+1} c_2}{n} \right].$$

This result is a generalisation of theorem 3 of Miller [4] concerning the preservation of normality.

3. A generalisation of the Jackknife. In the following it is assumed that n is greater than 2. Let T be a biased estimator of θ so that

$$(2) \qquad E_\theta[T] = \theta + \frac{\delta}{n^\alpha} + \frac{\delta'}{n^{\alpha'}},$$

where α and α' are known strictly positive real numbers, $\alpha \neq \alpha'$.

To obtain an unbiased estimator of θ, let us introduce the random vector

$$U = (T, T_{-1}, \ldots, T_{-n}, T_{-(1,2)}, \ldots, T_{-(n,n-1)})',$$

where $T_{-(i,j)}$ is the estimator of θ based on the sample where the i-th and j-th observations, $i \neq j$, are deleted; then the following linear model can be considered:

$$(3) \qquad U = B\gamma + \varepsilon, \quad U \in \boldsymbol{R}^N, \; N = n^2+1, \quad \gamma = (\theta, \delta, \delta').$$

B is the $(N\times 3)$-matrix whose column vectors are $\mathbb{1}$, b_α and $b_{\alpha'}$ with

$$b_\alpha = \Big(\frac{1}{n^\alpha}, \underbrace{\frac{1}{(n-1)^\alpha}, \ldots, \frac{1}{(n-1)^\alpha}}_{n \text{ terms}}, \underbrace{\frac{1}{(n-2)^\alpha}, \ldots, \frac{1}{(n-2)^\alpha}}_{n(n-1) \text{ terms}}\Big)',$$

ε is a centred random variable and Σ denotes $E[\varepsilon\cdot\varepsilon']$.

THEOREM 3. *If T fulfills* (2), *an unbiased estimator of θ exists, say $\hat{\theta}$, which is a linear function of U; moreover if $X_1, \ldots, X_n$ are independent and identically distributed and if T is a symmetric function of $X_1, \ldots, X_n$, then $\hat{\theta}$ is the Gauss-Markov estimator of θ in the linear model* (3).

Proof. The subspace $\mathcal{B}$ of $\mathbf{R}^n$ spanned by the columns of B is also spanned by the vectors:

$$f_1 = (1, 0, \ldots, 0)', \quad f_2 = (0, \underbrace{1, \ldots, 1}_{n \text{ terms}}, 0, \ldots, 0)', \quad f_3 = (0, 0, \ldots, 0, \underbrace{1, \ldots, 1}_{n(n-1) \text{ terms}})'.$$

The orthogonal projection of U on $\mathcal{B}$, say U_B, is the solution of:

$$(U-U_B)'\cdot f_1 = 0, \quad (U-U_B)'\cdot f_2 = 0, \quad (U-U_B)'\cdot f_3 = 0,$$
$$U_B = \hat{\theta}\mathbb{1} + \hat{\delta} b_\alpha + \hat{\delta}' b_{\alpha'}.$$

Hence the linear system is to be solved:

$$T = \hat{\theta} + \frac{\hat{\delta}}{n^\alpha} + \frac{\hat{\delta}'}{n^{\alpha'}},$$

$$T_2 = \frac{1}{n}\sum_i T_{(-i)} = \hat{\theta} + \frac{\hat{\delta}}{(n-1)^\alpha} + \frac{\hat{\delta}'}{(n-1)^{\alpha'}},$$

$$T_3 = \frac{1}{n(n-1)} \sum_{\substack{i,j\\ i\neq j}} T_{-(i,j)} = \hat{\theta} + \frac{\hat{\delta}}{(n-2)^\alpha} + \frac{\hat{\delta}'}{(n-2)^{\alpha'}}.$$

Clearly there is only one solution.

Moreover, in order to apply Kruskal's result, let us prove that $\Sigma\mathcal{B} \subseteq \mathcal{B}$: $\Sigma = E[(U-m)(U-m)']$ with $E[U] = m$.

For each vector f_l, $l = 1, 2, 3$, $\Sigma f_l = E[(U-m)(U-m)' f_l]$ with

$$(U-m)' f_1 = T - E(T),$$

$$(U-m)' f_2 = \sum_i [T_{-i} - E(T_{-i})], \quad (U-m)' f_3 = \sum_i \sum_j [T_{-(i,j)} - E(T_{-(i,j)})].$$

Each of these three real random variables is centred and is a symmetric function of $X_1, \ldots, X_n$; hence the last $n(n-1)$ components of the random variable $(U-m)(U-m)'f_l$ are identically distributed and the $(n+1)$-first components, except the first, are identically distributed; hence $(U-m)(U-m)'f_l$ is a linear combination of f_1, f_2 and f_3.

In the particular case, $\alpha = 1$ and $\alpha' = 2$, the Gauss-Markov estimator of θ is:

$$\hat{\theta} = \frac{1}{2}[n^2 T - 2(n-1)^2 T_2 + (n-2)^2 T_3];$$

$\hat{\theta}$ is the estimator given in [7].

The estimator $\hat{\theta}$ has a smaller variance than the unbiased estimator given by Mantel [3]:

$$\hat{\theta}' = n^2 T - (2n-1)(n-1) T_2 + (n-1)(n-2) T_3.$$

Remark 3. If (2) holds, it is impossible to get a linear unbiased estimator of θ depending on $\mathcal{T}$. Indeed, let us consider the linear model in R^N, $N = n+1$, $\mathcal{T} = B^* \gamma + \varepsilon$, where B^* is the $(N \times 3)$-matrix whose three column vectors are 1, a_α and $a_{\alpha'}$: each is a linear combination of e_1 and e_2; hence the range of B^* is 2 and it is only possible to estimate the linear combinations of θ, δ and δ' belonging to the subspace of R^3 spanned by the column vectors of $B^{*\prime} B^*$. This subspace is spanned by $(1, 1/n^\alpha, 1/n^{\alpha'})'$ and $(1, 1/(n-1)^\alpha, 1/(n-1)^{\alpha'})'$ and the vector $(1, 0, 0)'$ does not belong to this subspace.

Remark 4. In paper [7] the following situation is considered: $T_1, T_2, \ldots, T_{k+1}$ are $k+1$ estimations of θ such that

$$E[T_j] = \theta + \sum_{i=1}^{k} f_{ij}(n) b_i(\theta) \qquad (j = 1, \ldots, k+1)$$

and the proposed unbiased estimator of θ is

$$\hat{\theta} = \frac{\begin{vmatrix} t_1 & t_2 & \ldots & t_{k+1} \\ f_{11} & f_{12} & \ldots & f_{1,k+1} \\ \ldots & \ldots & \ldots & \ldots \\ f_{k1} & f_{k2} & \ldots & f_{k,k+1} \end{vmatrix}}{\begin{vmatrix} 1 & 1 & \ldots & 1 \\ f_{11} & f_{12} & \ldots & f_{1,k+1} \\ \ldots & \ldots & \ldots & \ldots \\ f_{k1} & f_{k2} & \ldots & f_{k,k+1} \end{vmatrix}}.$$

If this estimator exists, then it is the Gauss-Markov estimator of θ in the following linear model:

$$\mathscr{T} = \begin{pmatrix} T_1 \\ \dots \\ T_{k+1} \end{pmatrix} = A \begin{pmatrix} \theta \\ b_1(\theta) \\ \dots \\ b_k(\theta) \end{pmatrix} + \varepsilon$$

with $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon') = \Gamma$.

In effect, if the rank of A is $k+1$, the subspace A is R^{k+1} and then $\Gamma A \subseteq A$.

REFERENCES

[1] B. Efron, *Nonparametric estimates of standard error: The Jackknife, the bootstrap, and other methods*, Biometrika 68 (1981), p. 589–599.

[2] W. Kruskal, *When are Gauss-Markov and least squares estimators identical? A coordinate free approach*, Ann. Math. Statistics 39 (1968), p. 70–75.

[3] N. Mantel, *Assumption-free estimators using U-statistics and a relationship to the Jackknife method*, Biometrics 23 (1967), p. 567–571.

[4] R. G. Miller, *Trustworthy Jackknife*, Ann. Math. Statist. 35 (1964), p. 1594–1605.

[5] – *The Jackknife: A review*, Biometrika 61 (1974), p. 1–15.

[6] W. C. Paar, *A note on the Jackknife, the bootstrap and the delta method estimators of bias and variance*, ibidem 70 (1983), p. 719–722.

[7] W. R. Schucany, H. L. Gray and D. B. Owen, *On bias reduction in estimation*, J. Amer. Statist. Ass. 66 (1971), p. 524–533.

Laboratoire de Statistique et Probabilités
Université Paul Sabatier
U.A.-C.N.R.S. 745
31062 Toulouse Cedex, France

Received on 20. 8. 1985