Statistique et analyse des données

V. BRAILOVSKY

Remark concerning a paradoxical situation in behaviour of error rate in discriminant analysis

Statistique et analyse des données, tome 6, nº 1 (1981), p. 28-38

<http://www.numdam.org/item?id=SAD_1981__6_1_28_0>

Statistique et Analyse des données
1981 - 1 - pp. 28- 38

REMARK CONCERNING A PARADOXICAL SITUATION IN BEHAVIOUR OF ERROR RATE IN DISCRIMINANT ANALYSIS

V. BRAILOVSKY

RESUME : *On considère quelques problèmes d'analyse discriminante sur des échantillons ayant des données manquantes, soit sur des observations multivariées et enregistrées à l'aide d'un dispositif à un seul canal. En particulier on examine, en liaison avec les résultats de [3] et [5], la situation paradoxale où ajouter de nouveaux échantillons conduit à la détérioration de la qualité de la règle de décision.*

ABSTRACT : *Some problems of discriminant analysis, when samples have missing values or some multivariate observations are to be registered with help of a single-channel device, are considered. Especially the paradoxal situation when adding new samples leads to deterioration of decision rule quality is discussed in connection with results obtained in [3] and [5].*

Mots-clés : *Analyse discriminante.*

Note de l'éditeur : *V. BRAILOVSKY est un scientifique soviétique considéré comme "dissident", et il a été impossible de réaliser avec lui des aller et retour habituels entre l'auteur d'un manuscrit et l'éditeur de la revue. Vu son intérêt et à titre de témoignage, nous avons décidé de publier tel quel son article.*
V. BRAILOVSKY a été fait Docteur Honoris Causa de l'Université de Lausanne.
Nous avons appris tout récemment que V. BRAILOVSKY a été arrêté.

1. INTRODUCTION

As it is well known Anderson's linear discriminant function [1] gives the minimum-error-rate solution for two category classification problem for the case when the categories are discribed by multivariate normal densities with different means μ_1 and μ_2 and identical nonsingular covariance matrices $\Sigma = ||p_{ij}||$. This minimum error rate takes the form

$$\Delta = p_1 \, \Phi(-\frac{D}{2} + C) + p_2 \, \Phi(-\frac{D}{2} - C) \tag{1.1}$$

Here D stands for the Mahalanobis distance between categories

$$D^2 = (\mu_1 - \mu_2)^t \Sigma^{-1} (\mu_1 - \mu_2) \tag{1.2}$$

$$C = \frac{1}{D} \log p_2 / p_1$$

p_1, p_2 stand for a priori probabilities of categories 1 and 2, respectively, Φ (x) stands for c.d.f. of univariate normal density N(0,1).

M. Okamoto [2] considered the case when the values μ_1 , μ_2 , $\Sigma = ||p_{ij}||$ are unknown and they are estimated with help of usual unbiased estimators by sample set consisting of N_1 and N_2 complete samples from the first and second category, respectively. These estimates replaced the corresponding values in the Anderson's linear discriminant function. As a result one obtained the decision rule with additional (in comparison with (1-1)) error rate, the formula for this one being obtained in [2].

In [3] the author for the simple case $p_1 = p_2$ and $N_1 = N_2 = N$ and, assuming the covariance matrix Σ being known and only mean vectors μ_1 and μ_2 must be estimated with help of a sample set, considered the situation when the samples were not complete, i.e. there where missing values in them or multivariate observations were to be registered with help of a single-channel device and so on. The investigation proved that in this case a paradoxical phenomenon took place. Namely on some conditions if new samples are added to the sample set and as a result the quality of parameter estimates used in the decision rule becomes better, in the same time the quality of the decision rule itself becomes worse.

J.R. Barra discussed this situation in [4]. In [5], using the fact that in the model, considered in [3], the covariance matrix Σ being known, J.R. Barra suggested unbiased estimators for μ_1 and μ_2 of a new form and such ones that their quality is better than one of estimators used in [3] and no paradoxical phenomenon takes place. So one may think that the phenomenon is a result of improper choice of estimators.

The autor think that the considered phenomenon is of more general nature and in many cases the very existence of "proper estimators" with above mentioned properties is questionable or their obtaining is a difficult problem.

In this article a model of discriminant analysis where the mean values μ_1 and μ_2 as well as covariance matrix Σ are unknown is considered an one proves that on some conditions the above mentioned paradoxical phenomenon exists for the model. In the same time the estimators, similar to ones suggested by J.R. Barra in [5], cannot be used in this model because all parameters used in them are unknown.

2. DESCRIPTION OF THE MODEL

Let there be two categories each of them be described by a bivariate normal density, the covari nce matrices being identical, nonsingular : $N(x_1,x_2,\mu_1,\Sigma)$, $N(x_1,x_2,\mu_2,\Sigma)^*$. μ_1 and μ_2 stand for mean values; x_1, x_2-arguments. For simplification we shall suppose that a priori probabilities $p_1 = p_2 = 1/2$.

Anderson's linear discriminant function for the case has the form [1]

$$x_1 = ax_2+b \qquad (2.1.)$$

Coefficients a and b depends on components of means vectors $\mu_1(\mu_{11},\mu_{12})$ and $\mu_2(\mu_{21},\mu_{22})$ and covariance matrix Σ (p_{ij}) $(i,j = 1,2)$. But these values are unknown. If one replace these values by their statistical estimates we obtain another discriminant function of the form

$$x_1 = \hat{a}x_2+\hat{b} \qquad (2.2.)$$

------

* For our purposes (see Section 1) it is enough to considered bivariate case.

For the considered case

$$\hat{a} = \frac{\hat{p}_{11}(\hat{\mu}_{12}-\hat{\mu}_{22})+\hat{p}_{12}(\hat{\mu}_{21}-\hat{\mu}_{11})}{\hat{p}_{22}(\hat{\mu}_{21}-\hat{\mu}_{11})+\hat{p}_{12}(\hat{\mu}_{12}-\hat{\mu}_{22})} ;$$

$$\hat{b} = \frac{\hat{p}_{22}(\hat{\mu}_{21}^2-\hat{\mu}_{11}^2)+2\hat{p}_{12}(\hat{\mu}_{11}\hat{\mu}_{12}-\hat{\mu}_{21}\hat{\mu}_{22})+\hat{p}_{11}(\hat{\mu}_{22}^2-\hat{\mu}_{12}^2)}{2\,[\hat{p}_{22}(\hat{\mu}_{21}-\hat{\mu}_{11})+\hat{p}_{12}(\hat{\mu}_{12}-\hat{\mu}_{22})]} \tag{2.3}$$

Let optimal discriminant function (2.1) divide space of arguments (x_1,x_2) into two regions R_1 and R_2 for the first and second category, respectively. Then for any sample set used to obtain (2.2) the additional error rate, connected with use (2.2) instead of (2.1), may be written as follows

$$f(\hat{\mu}_1,\hat{\mu}_2,\hat{\Sigma}) = \frac{1}{2}\int_{R_{21}}[N(x,\mu_2,\Sigma)-N(x,\mu_1,\Sigma)]\,dx + \frac{1}{2}\int_{R_{12}}[N(x,\mu_1,\Sigma)-N(x,\mu_2,\Sigma)]dx \tag{2.4}$$

R_{12} (R_{21}) stands for the part of R_1 (R_2) , where discriminant function (2.2) gets wrong classification R_2 (R_1).

Because of the fact that $p_{12} = p_{21}$ and $\hat{p}_{12} = \hat{p}_{21}$ for any sample set, the function (2.4) depends on 7 variables : $\hat{\mu}_{11}$, $\hat{\mu}_{12}$, $\hat{\mu}_{21}$, $\hat{\mu}_{22}$, $\hat{p}_{11}$, $\hat{p}_{12}$, $\hat{p}_{22}$. Following the approach described in [3] let us consider the Taylor's formula for the function (2.4) with center in the point (μ_1,μ_2,Σ), i.e. in the point corresponding the genuine values of respective parameters and let us average the formula over all possible sample sets with size N. One means that from both of categories one obtains sample sets with the same size N. Let us take into account that :

(1) We use the Taylor's formula with the terms of the first and second orders (and remainder). It follows the partial derivatives of the first and second orders are calculated in the center point and does not depend on sample variables.

(2) We shall use only unbiased statistical estimates for μ_{ij} and p_{ij} and $\mathbb{E}(\hat{\mu}_{ij}-\mu_{ij}) = \mathbb{E}(\hat{p}_{ij}-p_{ij}) = 0 \quad i,j = 1,2.$

(3) Samples from the first and second categories are taken independently and (see also the form of statistical estimators for $\hat{\mu}_{ij}$) $cov(\hat{\mu}_{1i},\hat{\mu}_{2j}) = 0$ $(i,j = 1,2)$.

(4) We shall use the estimator for $\hat{\mu}_{ij}$ and $\hat{p}_{ij}$ of usual form and as a result $cov(\hat{\mu}_{ij},\hat{p}_{k\ell}) = 0$ [6]. $i,j,k,\ell = 1,2$.

(5) It may be proved that averaged value of remainder has the order higher, than 1/N.

As a result one obtains

$$\mathbb{E}\, f(\hat{\mu}_1,\hat{\mu}_2,\hat{\Sigma}) = \frac{1}{2}\Big[\sum_{i,j=1,2} \frac{\partial^2 f}{\partial \mu_{ij}^2} \text{Var}\, \hat{\mu}_{ij} + 2 \sum_{i,j,k=1,2;\, j>k} \frac{\partial^2 f}{\partial \mu_{ij}\partial \mu_{ik}} cov(\hat{\mu}_{ij},\hat{\mu}_{ik}) +$$

$$+ \sum_{\substack{i,j=1,2\\ j\geq i}} \frac{\partial^2 f}{\partial p_{ij}^2} \text{Var}\, \hat{p}_{ij} + 2 \sum_{\substack{i,j,k,\ell=1,2\\ j\geq i,\, \ell\geq k,\, [(i>k)\cup(i=k)(j>\ell)]}} \frac{\partial^2 f}{\partial p_{ij}\partial p_{k\ell}} cov(\hat{p}_{ij},\hat{p}_{k\ell})\Big] + \bar{\bar{O}}(1/N) \qquad (2.5)$$

All indices i,j,k,ℓ are equal to 1 or 2 with some restrictions pointed under each symbol Σ. For the last one $j\geq i$ and $\ell\geq k$ and in the same time either $i > k$ or $i = k$ and $j > \ell$ simultaneously.
As it is easy to see the expression in the square brackets in (2.5) has the order 1/N. Thus to estimate $\mathbb{E}\, f(\hat{\mu}_1,\hat{\mu}_2,\hat{\Sigma})$ with the accuracy of magnitude 1/N one must calculate Eq. (2.5)

Let us assume

$$\mu_{1i} = 0 \quad (i = 1,2)\ ; \quad \mu_{21} > 0\ ; \ \mu_{22} = 0 \qquad (2.6)$$

From (2.1)-(2.4) and (2.6) the values of partial derivatives used in (2.5) may be calculated. Their values are given in the Appendix 1.

As for values of variances and covariances from Eq. (2.5) they depend on sample set structure an we shall consider it in the next Section.

3. RESULTS OF CALCULATIONS

1. Let one has N randomly drawn complete samples from the first and those from the second category. Let us use the estimators for unknown parameters as follows.

$$\hat{\mu}_{1i} = \frac{1}{N}\sum_{r=1}^{N} x_i^r \; ; \qquad \hat{\mu}_{2i} = \frac{1}{N}\sum_{r=1}^{N} \tilde{x}_i^r \tag{3.1}$$

$$\hat{p}_{ij} = \frac{1}{2(N-1)}\left[\sum_{r=1}^{N} (x_i^r-\hat{\mu}_{1i})(x_j^r-\hat{\mu}_{1j})+\sum_{r=1}^{N} (\tilde{x}_i^r-\hat{\mu}_{2i})(\tilde{x}_j^r-\hat{\mu}_{2j})\right] \tag{3.2}$$

$$i,j = 1,2 \; ; \; j \geq i.$$

x_i^r ($\tilde{x}_i^r$) stands for i^{th} component of rth sample in sample set from the first (second) category.

Estimators (3.1) and (3.2) are unbiased and in the Appendix 2 one gives the values of variances and covariances of estimates (3.1), (3.2) to calculate Eq. (2.5) [6].

Using the values given in Appendices 1, 2 after calculation of expression (2.5) for $N \gg 1$ one obtains

$$\Delta_1 = \mathbb{E}\, f(\hat{\mu}_1,\hat{\mu}_2,\hat{\Sigma}) = \frac{\tilde{K}}{2N}\left[\frac{2|\Sigma|^2}{\mu_{21}p_{22}^{1/2}} + p_{11}p_{22}^{3/2}\mu_{21}-p_{12}^2 p_{22}^{1/2}\mu_{21}\right]; \tag{3.3}$$

Taking into account that for considered case Mahalanobis distance between categories (1.2) $D = \frac{\mu_{21}\sqrt{p_{22}}}{|\Sigma|^{1/2}}$ one can see the formula (3.3) coincides with the formula obtained for the case by Okamoto [2].

2. Let in addition to the sample sets described in the previous part one obtain KN independently drawn samples from the first variable in first category and KN independtly drawn samples from the second variable and first category. The additional samples of the same kind are obtained from the second category. Let us use the unbiased estimators as follows

$$\hat{\mu}_{1i} = \frac{1}{N(1+K)}\sum_{r=1}^{N(1+K)} x_i^r \; ; \quad \hat{\mu}_{2i} = \frac{1}{N(1+K)}\sum_{r=1}^{N(1+K)} \tilde{x}_i^r \tag{3.4}$$

$$\hat{p}_{ii} = \frac{1}{2[N(1+K)-1]}\left[\sum_{r=1}^{N(1+K)} (x_i^r-\hat{\mu}_{1i})^2+\sum_{r=1}^{N(1+K)} (x_i^r-\hat{\mu}_{2i})^2\right]; \quad i=1,2 \tag{3.5}$$

In (3.4) and (3.5) summations are performed over all possible samples

$$\hat{p}_{12} = \frac{1}{2[N - \frac{2K+1}{(1+K)^2}]} [\sum_{r=1}^{N} (x_1^r - \hat{\mu}_{11})(x_2^r - \hat{\mu}_{12}) + \sum_{r=1}^{N} (\tilde{x}_1^r - \hat{\mu}_{21})(\tilde{x}_2^r - \hat{\mu}_{22})] \tag{3.6}$$

In (3.6) summation are performed over sample sets with complete samples only. The form of coefficient in (3.6) is determined by the condition $\hat{p}_{12}$ (3.6) to be unbiased.

In Appendix 3 the result of calculation of variances and covariances of estimates (3.4) (3.5), (3.6) to calculate the Eq. (2.5) is given. Using the values given in Appendices 1, 3 after calculation of expression (2.5) one obtains

$$\Delta_2 = \mathbb{E}\, \mathcal{E}(\hat{\mu}_1, \hat{\mu}_2, \hat{\Sigma}) = \frac{\tilde{K}}{2N} \left[\frac{2|\Sigma|^2}{\mu_{21} P_{22}^{1/2} (1+K)} + \frac{K+2}{K+1} \frac{\mu_{21} P_{11} P_{22}^{3/2}}{2} + \frac{K^2-K-2}{(K+1)^2} \frac{\mu_{21} P_{12}^2 P_{22}^{1/2}}{2} \right] \tag{3.7}$$

It is easy to obtain the condition when the paradoxical phenomenon, discussed in the Section 1, takes place, i.e. $\Delta_2 > \Delta_1$ (see (3.3), (3.7)). After some algebra one obtains : $\Delta_2 > \Delta_1$ if and only if

$$P_{12}^2 > \frac{|\Sigma|}{D} \left(\frac{2}{D} + \frac{D}{2}\right) \tag{3.8}$$

It is interesting to note that if the condition (3.8) is fulfilled, $\Delta_2 > \Delta_1$ for any $k > 0$. It means that adding of any quantity of samples of one-variate form (see the beginning of this part) leads to increasing of error rate. Comparing this example with similar one for the case when the covariance matrix Σ is known and only mean vectors μ_1 and μ_2 must be estimated (the latter example was considered in the first version [3] and reproduced in [4] and [5]), one sees for the latter case the condition similar to $\Delta_2 > \Delta_1$ may take place only for $0 < k < k^* < \infty$. There always exist a quantity k^* such that for $k > k^*$ no paradoxical phenomenon take place. In the considered for the latter case numerical example for $D = 3$, coefficient of correlation $\rho = 0.8$ the paradoxical phenomenon takes place for $0 < k < 1.47$. In the considered here case for $D = 3$ and $\rho = 0.8$ this phenomenon takes place for $0 < k < \infty$ (see (3.8)).

3. Let us now consider the situation when samples have randomly missing values. So let one have independently drawn complete samples from the first and second category (see part 1 of this section), and let for each component of each sample the value is missing with the probability λ_1 and independently on the fact whether some other values are missing or not. Let from each category one obtains $\frac{N'}{1-\lambda_1}$ samples with missing values of such a structure. Because of the fact that our analysis is valid for $N' \gg 1$ we consider sample set from each category have the structure as follows. There are $N'(1-\lambda_1)$ complete samples, $N'.\lambda_1$ samples where only the first component is registered and $N'\lambda_1$ ones where only the second component is registered. So the structure of sample set coincides with one considered in the part 2 of this section and all the results are the same if $N = N'(1-\lambda_1)$ and $K = \lambda_1/(1-\lambda_1)$; let us denote the result of (3.7) after such a replacing Δ'.

Let one have two sample sets from each category : the first one with frequency ratio of missing values λ_1 and size $\frac{N'}{1-\lambda_1}$ and the second one with respective parameters λ_2 and $\frac{cN'}{1-\lambda_1}$; if one unites both of them one obtains for each of category $N'[(1+c)-(\lambda_1+c\lambda_2)]$ complete samples and $N'(\lambda_1+c\lambda_2)$ samples, where only the first component is registered and those for the second component. This situation coincide with that for part 2 of this Section if $N = N'[(1+c)-(\lambda_1+c\lambda_2)]$ and $K = \frac{\lambda_1+c\lambda_2}{[(1+c)-(\lambda_1+c\lambda_2)]}$; let us denote the result of (3.7) after such a replacing Δ''.

The paradoxical effect takes place if $\Delta'' > \Delta'$. We shall not discuss any general formula but let us consider the numerical example discussed in [3] for similar situation. Let $\lambda_1 = 0$, $D = 2$, $p_{11} = p_{22} = 0.526$, $p_{12} = 0.474$, $|\Sigma| = 0.052$, $\mu_{21} = 0.629$. For our situation the paradoxical effect takes place if $c < \frac{0.11436\lambda_2-0.03557}{0.03557(1-\lambda_2)}$; Taking into account $c > 0$, the effect takes place if $\lambda_2 > \frac{0.03557}{0.11436} = 0.311$.

Appendix 1

On condition (2.6) one can obtain

$$\frac{\partial^2 f}{\partial \mu_{11}^2} = \frac{\partial^2 f}{\partial \mu_{21}^2} = \tilde{K} \frac{\mu_{21} p_{22}^{3/2}}{4}$$

$$\frac{\partial^2 f}{\partial \mu_{12}^2} = \frac{\partial^2 f}{\partial \mu_{22}^2} = \tilde{K}\, \mu_{21}\, p_{22}^{3/2} \left(\frac{|\Sigma|^2}{p_{22}^3 \mu_{21}^2} + \frac{p_{12}^2}{4 p_{22}^2} \right)$$

$$\frac{\partial^2 f}{\partial \mu_{11} \partial \mu_{12}} = \frac{\partial^2 f}{\partial \mu_{21} \partial \mu_{22}} = -\tilde{K} \frac{\mu_{21} p_{12} p_{22}^{1/2}}{4}$$

$$\frac{\partial^2 f}{\partial p_{11}^2} = 0$$

$$\frac{\partial^2 f}{\partial p_{22}^2} = \tilde{K} \frac{\mu_{21} p_{12}}{p_{22}^{3/2}}$$

$$\frac{\partial^2 f}{\partial p_{12}^2} = \tilde{K}\, \mu_{21} p_{22}^{1/2}$$

$$\frac{\partial^2 f}{\partial p_{11} \partial p_{12}} = 0$$

$$\frac{\partial^2 f}{\partial p_{22} \partial p_{12}} = -\tilde{K} \frac{\mu_{21} p_{12}}{p_{22}^{1/2}}$$

$$\frac{\partial^2 f}{\partial p_{11} \partial p_{22}} = 0$$

Here $\tilde{K} = \frac{1}{2\sqrt{2\pi}\,|\Sigma|^{3/2}} \exp\left(-\frac{p_{22}\mu_{21}^2}{8|\Sigma|}\right)$; $|\Sigma| = \det \|\Sigma\|$

All values of partial derivatives are calculated in the center point of expansion.

Appendix 2

For estimators (3.1), (3.2) one has

$$\operatorname{var}\hat{\mu}_{11} = \operatorname{var}\hat{\mu}_{21} = \frac{p_{11}}{N} \; ; \; \operatorname{var}\hat{\mu}_{12} = \operatorname{var}\hat{\mu}_{22} = \frac{p_{22}}{N} \; ;$$

$$\operatorname{cov}(\hat{\mu}_{11},\hat{\mu}_{12}) = \operatorname{cov}(\hat{\mu}_{21},\hat{\mu}_{22}) = \frac{p_{12}}{N}$$

$$\operatorname{var}\hat{p}_{11} = \frac{p_{11}^2}{N-1} \; ; \; \operatorname{var}\hat{p}_{22} = \frac{p_{22}^2}{N-1} \; ; \; \operatorname{var}\hat{p}_{12} = \frac{p_{11}p_{22}+p_{12}^2}{2(N-1)}$$

$$\operatorname{cov}(\hat{p}_{11},\hat{p}_{12}) = \frac{p_{11}p_{12}}{N-1} \; ; \; \operatorname{cov}(\hat{p}_{22},\hat{p}_{12}) = \frac{p_{22}\cdot p_{12}}{N-1} \; ;$$

$$\operatorname{cov}(\hat{p}_{11},\hat{p}_{22}) = \frac{p_{12}^2}{N-1} \; ;$$

Appendix 3

For estimators (3.4), (3.5), (3.6) for N >> 1 one has

$$\operatorname{var}\hat{\mu}_{11} = \operatorname{var}\hat{\mu}_{21} = \frac{p_{11}}{N(1+K)} \; ; \; \operatorname{var}\hat{\mu}_{12} = \operatorname{var}\hat{\mu}_{22} = \frac{p_{22}}{N(1+K)} \; ;$$

$$\operatorname{cov}(\hat{\mu}_{11},\hat{\mu}_{12}) = \operatorname{cov}(\hat{\mu}_{21},\hat{\mu}_{22}) = \frac{p_{12}}{N(1+K)^2} \; ;$$

$$\operatorname{var}\hat{p}_{22} = \frac{p_{22}^2}{N(1+K)-1} \approx \frac{p_{22}^2}{N(1+K)} \; ;$$

$$\operatorname{var}\hat{p}_{12} = \frac{p_{12}^2\left[2N+\frac{4K(N-1)}{(1+K)^2} - 2\frac{(2K+1)^2}{(1+K)^4}\right] + 2p_{11}p_{22}\left[(N-1)+\frac{K^2}{(1+K)^2}\right]}{4\left[N-\frac{2K+1}{(1+K)^2}\right]^2} \approx$$

$$\approx \frac{1}{2N}\left[p_{11}p_{22}+p_{12}^2\left(1+\frac{2K}{(1+K)^2}\right)\right] \; ;$$

$$\operatorname{cov}(\hat{p}_{22},\hat{p}_{12}) = \frac{\frac{2K+1}{(1+K)^2}\, p_{12}p_{22}}{[N - \frac{2K+1}{(1+K)^2}]} \approx \frac{2K+1}{N(1+K)^2}\, p_{12}p_{22} \quad ;$$

While calculating one used above mentioned approximations obtained on condition $N >> 1$ and taking into account $\frac{2k+1}{(1+K)^2} \leq 1$ for $0 \leq k < \infty$.

REFERENCES

1. Anderson T.W. - An introduction to Multivariate Statistical Analysis J. Wiley, New York, 1958.
2. Okamoto M. - An Asymptotic Expansion for the Distribution of the Linear Discriminant Function. Ann. Math. Stat., 1963, 34, 4, pp. 1286-1301 (with correction 39, 1358-1359).
3. Brailovsky V. - On Influence of Sample Set Structure on Decision Rule Quality for the Case of Linear Discriminant Function. To appear in Trans. IEEE PAMI.
4. Barra J.R. - Influence de l'estimation des paramètres sur une probabilité d'erreur en analyse discriminante (selon V. Brailovsky) Séminaire de l'Ecole Polytechnique. Novembre 1979.
5. Barra J.R. - A propos d'un résultat de Brailovsky concernant une probabilité d'erreur en analyse discriminante. Annales de l'Institut Henri Poincaré B. Vol. 17 n° 1.
6. Cramer H. - Mathematical Methods of Statistics, 1945. Princeton University Press.