**Serdica**
Bulgariacae mathematicae
publicationes

**Сердика**
Българско математическо
списание

# ESTIMATION DES PARAMETRES DE LA LOI NORMALE PAR LA METHODE DE LA DISTANCE MINIMALE A L'AIDE DE L'ALGORITHME DE LA REGRESSION MINIMAX

PHILIPPE PILIBOSSIAN, JEAN-PAUL QUELEN

Disposant de réalisations regroupées en classes d'une variable gaussienne, on se propose d'estimer les paramètres de la loi en rendant minimum une distance entre sa fonction de répartion empirique et la famille des fonctions de répartion théoriques. Cette optimisation est réalisée en utilisant l'algorithme de la régression minimax.

**1. Position du problème.** Lorsqu'on ne dispose que de mesures d'une variable alétoire, issues d'une population gaussienne mais regroupées en classe, — soit parce que les valeurs initiales des observations sont perdues, soit parce que l'instrument de mesure est un appareil à seuil, les estimations des paramètres de la loi normale par les méthodes classiques ne sont pas toujours très satisfaisantes. Nous proposons, dans ce cas, de les estimer en rendant minimum le maximum de l'écart absolu entre les fonctions de répartition empirique et théorique. Cette méthode, connue sous le nom d'estimation par la méthode de la distance minimale, a été proposée et utilisée par J. Wolfowitz [8]. Il existe de nombreux travaux théoriques sur cette estimation utilisant différentes distances et sur les problèmes annexes; on trouvera dans [3] une bibliographie détaillée. La distance que nous avons utilisée $\|\cdot\|_\infty$ est connue sous le nom de distance de Kolmogorov — Smirnov. Pour la recherche de la solution numérique nous avons fait appel à une méthode exacte utilisée pour la régression minimax [4, 5].

**2. Méthode**. Soit $X_1, X_2, \dots, X_n$ des variables aléatoires réelles, indépendantes et de même loi. Nous désignerons par $F_\theta$ leur fonction de répartition théorique du paramètre $\theta$ inconnu et par $\widehat{F}_n$ la fonction de répartiton empirique correspondante. Le théorème de Glivenko — Cantelli assure la convergence presque sûre uniforme de $\widehat{F}_n$ vers $F_\theta$ pour presque tout $x$; autrement dit, pour la distance entre $\widehat{F}_n$ et $F_\theta$, on a la réalisation pour presque tout $x$:

$$\|\widehat{F}_n-F_\theta\|_\infty=\sup_{x\in R}|\widehat{F}_n(x)-F_\theta(x)|\xrightarrow{\text{p. s.}}0.$$

On ce propose donc d'estimer le paramètre $\theta$ pour lequel

$$\operatorname*{Inf}_\theta\|\widehat{F}_n-F_\theta\|_\infty$$

est atteint.

Or $\widehat{F}_n$ étant une fonction en escalier non décrossante, il suffit de considérer cet écart absolu uniquement pour les points de discontinuité de $\widehat{F}_n$. Ces points s'obtiennent par classement des $n$ valeurs observées $x_i$ des variables $X_i$ par ordre croissant $x_{1,n}\leq x_{2,n}\leq\dots\leq x_{n,n}$. Si de plus $F_\theta$ est continue partout, les valeurs obtenues $x_{i,n}$ sont presque sûrement distinctes et on a:

$$\|\widehat{F}_n-F_\theta\|_\infty = \sup_{1\leq i\leq n} \operatorname{Max}\{|\frac{i}{n}-F_\theta(x_{i,n})|, |\frac{i-1}{n}-F_\theta(x_{i,n})|\}.$$

**3. Cas de la loi normale** $\mathcal{N}(m, \sigma^2)$. Si $x_1, x_2, \ldots, x_n$ sont les réalisations d'une variable gaussienne d'espérance mathématique $m$ et d'écart — type $\sigma$, on est ramené à déterminer $\widetilde{m}$ et $\widetilde{\sigma}$ de façon à rendre minimum

$$\sup_{1\leq i\leq n} \operatorname{Max}\{|F_0(u_i)-F_0(\frac{x_{i,n}-m}{\sigma})|, |F_0(u_{i-1})-F(\frac{x_{i,n}-m}{\sigma})|\},$$

où $F_0$ désigne la fonction de répratition de la loi normale centrée réduite $\mathcal{N}(0, 1)$ et $u_i=F_0^{-1}(\frac{i}{n})$, le quantile d'ordre $\frac{i}{n}$. Le paramètre $\theta$ est ici pour un $\theta=(\frac{1}{\sigma}, \frac{m}{\sigma})$.

**4. Calcul**. Soit $c_1, c_2, \ldots, c_p$ $(p\leq n)$ les limites supérieures des classes des observations de la variable aléatoire étudiée. Il s'agit de calculer

$$\operatorname*{Min}_{\theta} \sup_{1\leq i\leq p} \operatorname{Max}\{|F_0(u_i)-F_0(\frac{c_i-m}{\sigma})|, |F_0(u_{i-1})-F_0(\frac{c_i-m}{\sigma})|\}.$$

En appliquant la formule de Taylor aux écarts absolus:

$$|F_0(u_i)-F_0(\frac{c_i-m}{\sigma})|=f_0(u_i)\,|u_i-\frac{c_i}{\sigma}+\frac{m}{\sigma}|+o\,(u_i-\frac{c_i}{\sigma}+\frac{m}{\sigma}),$$

$f_0$ désignant ici la densité de probabilité de la loi $\mathcal{N}(0, 1)$ et $o(x)$ un infiniment petit ordre $x^2$, on trouve une expression approchée de ces écarts absolus, linéaires par rapport à $\theta_1=\frac{1}{\sigma}$ et $\theta_2=-\frac{m}{\sigma}$, à savoir:

$$f_0(u_i)u_i-\theta_1 f_0(u_i)c_i-\theta_2 f_0(u_i).$$

On est ainsi ramené à la résolution du problème de la régression minimax [2] pour la variable à expliquer $y(i)=f_0(u_i)u_i$ et deux variables explicatives $x_1(i)=f_0(u_i)c_i$ et $x_2(i)=f_0(u_i)$; les paramètres $\theta_1$ et $\theta_2$ étant définis par

$$\operatorname*{Min}_{\theta} \max_{i} |y(i)-\theta_1 x_1(i)-\theta_2 x_2(i)|.$$

**5. Conclusion.** Pour des fonctions de répartition continues les théorèmes de Smirnov et Kolmogorov nous renseignent sur le comportement asymptotique de la distance normée $\sqrt{n}\|\widehat{F}_n-F_\theta\|_\infty$, dans le cas unilatéral et bilatéral respectivement. On trouvera dans [7] un ensemble de résultats permettant de préciser la vitesse de convergence de la loi asymtotique vers la loi exacte. On peut ainsi établir des tests non paramétriques [1, 5] pour vérifier l'appartenance d'un échantillon à une population déterminée par une fonction de répartition donnée entièrement connue. Or, dans la pratique le cas où les paramètres théoriques d'une famille de lois sont connus ne se présente que très rarement. Diverses estimations sont possibles; dans le présent travail celle pour les paramètres $m$ et $\sigma$ est conduite pour les familles des lois normales $\mathcal{N}(m, \sigma)$ par la méthode de régression minimax. Notons qu'elle a donné de bons résultats en pratique [6].

Dans ce domaine, il y a encore d'intéressants sujets de recherche sur les propriétés limites.

## BIBLIOGRAPHIE


1. M. Fisz. Probability theory and mathematical statistics. New York, 1963, 445—449.
2. J. Hajek. A course in non parametric statistics. Holden-Day, 1969, 62—65.
3. W. C. Parr. Minimum distance estimation: a bibliography. *Commun. Statist. Theor. Meth.*, A 10, 1981, 1205—1225.
4. Ph. Pilibossian. Régression du minimax. *C. R. Acad. Sci. Paris, Ser.* A, **228**, 1979, 921—923.
5. Ph. Pilibossian, G. Der Megreditchian. Recherche d'une fonction de régression par le critère du minimax. *Publ. Inst. Statist. Univ. Paris*, **28**, 1983, fasc. 3, 59—92.
6. Ph. Pilibossian, J.-P. Quelen. Journées de Statistique, 1983.
7. R. J. Serfling. Approximation theorems of mathematical statistics. London, 1980, 56—66.
8. J. Wolfowitz. Minimax estimate of the mean of a normal distribution with known variance. *Ann. Math. Statist.*, **21**, 1950, 218—230.


*Académie de Paris*
*Université de Paris VI Pierre et Marie Curie*
*Mathématiques*
*4, place Jussieu*
*75230 Paris*
*France*

*Received 8. 7. 1986*