SUR UN THÉORÈME DE LINNIK EN THÉORIE DES ERREURS

ABDELMAJID BEN HADJ SALEM, DIPL.-ENG.

Résumé. Dans cette note, nous donnons une démonstration d'un théorème de Linnik concernant la théorie des erreurs, énnoncé dans son livre "*Méthode des moindres carrés et les bases mathématiques de la théorie statistique du traitement des observations*", sans démonstration.

Abstract. In this note, we give a proof of a theorem of Linnik concerning the theory of errors, stated in his book "*Least squares method and the mathematical bases of the statistical theory of the treatment of observations*", without proof.

1. Introduction

Dans son fascicule sur la théorie des erreurs [1],[2], l'auteur P. Hottier mentionne un théorème cité par Linnik [1] concernant la comparaison de l'ellipsoïde indicateur de la loi relative à l'estimateur des moindres carrés par rapport à celui de tout autre estimateur non biaisé et normal, sans avoir connaissance une démonstration de théorème en question.

C'est vrai que le théorème en question est cité sans démonstration dans le livre de Linnik [4].

2. Cadre du Théorème de Linnik

Le théorème de Linnik a été cité dans ([1],[2]) comme conséquence du théorème suivant :

On suppose à présent :

$$L \in \mathcal{N}(\dot{L}, \sqrt{P^{-1}}\sigma_0)$$

ce qui entraîne évidemment $E(L - \dot{L}) = 0$, d'où :

1. Yuri Vladimirovich Linnik (1915 - 1972) : mathématicien soviétique était actif en théorie des nombres, probabilité et statistiques mathématiques.

Théorème 2.1. *Si les observations sont normales, les estimateurs des moindres carrés sont totalement efficaces.*

Autrement dit, parmi tous les estimateurs sans biais et normaux [a], *ce sont les plus précis (variance minimale).*

a. On remarquera qu'en pratique toute méthode de calcul fournit des estimateurs normaux (si les observations sont normales) ; en effet si $\tilde{X}'$ désigne un estimateur de $\dot{X}$, c'est une fonction des observations L : $\tilde{X}' = f(L) = f(\dot{L} + (L - \dot{L})) \approx f(\dot{L})+$ fonction linéaire d'erreurs normales ; de plus si la méthode de calcul est correcte, c'est-à-dire que par définition elle fournit des estimateurs non biaisés : $\dot{X} = f(\dot{L})$.

Ce théorème capital par ses conséquences théoriques et pratiques découle de la théorie de l'estimation des paramètres d'une loi de probabilité donnée à partir d'un échantillon de variables aléatoires relevant de cette loi.

Il est facile à établir :

- l'échantillonage se compose ici des n composantes du vecteur aléatoire normal :

$$ {}_nL_1 \in \mathcal{N}(\dot{L}, \sqrt{P^{-1}}\sigma_0) $$

La densité de probabilité de la loi est :

$$ f(L) = \frac{1}{(2\pi)^{n/2}} \cdot \frac{\sqrt{P}}{(\sigma_0^2)^{n/2}} e^{-\dfrac{1}{2\sigma_0^2}(L - \dot{L})^T P (L - \dot{L})} $$

soit encore :

$$ f(L) = \frac{1}{(2\pi)^{n/2}} \cdot \frac{\sqrt{P}}{(\sigma_0^2)^{n/2}} e^{-\dfrac{1}{2\sigma_0^2}(A\dot{X} - K)^T P (A\dot{X} - K)} $$

Or d'après la théorie de l'estimation des paramètres inconnus d'une loi donnée (les paramètres inconnus sont ici les r composantes de $\dot{X}$) à partir d'un échantillon (ici L), on sait que la matrice variance de $\tilde{X}$, estimateur de $\dot{X}$, soit $\Gamma_{\tilde{X}}$ est telle que :

$$ Z^T \Gamma_{\tilde{X}} Z \geq Z^T F^{-1} Z, \ \ \forall Z $$

F^{-1} étant l'inverse de F matrice d'information de Fisher donnée par :

$$ F = -E(F_{ij}) = -E\left(\frac{\partial^2 Log f}{\partial X_i \partial X_j}\right) \tag{2.1} $$

Or on a ici : $Log f = Cte - \dfrac{1}{2\sigma_0^2}(A\dot{X} - K)^T P (A\dot{X} - K)$. D'où on déduit :

$$ \frac{\partial Log f}{\partial X} = \left(\frac{\partial Log f}{\partial X_i}\right) = -\frac{1}{2\sigma_0^2}(A^T P A\dot{X} - A^T P K) \quad \text{vecteur} \ \ r \times 1 $$

Puis :

$$\left(\frac{\partial^2 Logf}{\partial X_i \partial X_j}\right) = -\frac{N}{\sigma_0^2} \Longrightarrow F = \frac{N}{\sigma_0^2}$$

On a donc :

$$Z^T \Gamma_{\tilde{X}} Z \geq Z^T N^{-1} \sigma_0^2 Z, \ \forall \, {}_r Z_1 \tag{2.2}$$

Mais il se trouve que pour l'estimateur des moindres carrés, $\Gamma_{\tilde{X}}$ est justement égal à la matrice variance limite $N^{-1}\sigma_0^2$; on dit alors que l'estimateur des moindres carrés est un estimateur totalement efficace.

Ce qui signifie que si $\tilde{X}'$ est un autre estimateur non biaisé de $\dot{X}$, on a :

$$\boxed{Z^T \Gamma_{\tilde{X}'} Z > Z^T \Gamma_{\tilde{X}} Z = Z^T N^{-1} \sigma_0^2 Z, \ \forall \ Z}$$

3. Conséquences

i) Si on considère d'abord une inconnue scalaire $\dot{Y}$, alors parmi tous les estimateurs non biaisés et normaux, celui des moindres carrés est de **variance minimale**.

Autrement dit, si $\dot{Y}'$ est un autre estimateur non biaisé et normal de $\dot{Y}$, on a :

$$\sigma_{\tilde{Y}'} > \sigma_{\tilde{Y}}$$

ou encore, pour tout $\epsilon > 0, \exists N$ tel que :

$$n > N \Longrightarrow Pr(||\tilde{Y} - \dot{Y}|| < \epsilon) > Pr(||\tilde{Y}' - \dot{Y}|| < \epsilon)$$

ii) Plus généralement si on considère un ensemble de r grandeurs géométriquement indépendantes ${}_r\dot{Y}_1$ (les inconnues ${}_r\dot{X}_1$ sont un cas particulier) dont les estimateurs au sens des moindres carrés sont les composantes du vecteur ${}_r\dot{Y}_1$ de matrice variance $\Gamma_{\tilde{Y}}$, alors si $\tilde{Y}'$ est un autre estimateur non biaisé et normal de $\dot{Y}$, on a :

$$Z^T \Gamma_{\tilde{Y}'} Z > Z^T \Gamma_{\tilde{Y}} Z, \ \forall \ Z \tag{3.1}$$

Si Z est le vecteur unitaire d'une certaine direction, cette inégalité implique que la variance marginale relative à cette direction est minimale si on choisit l'estimateur des moindres carrés.

Cette inégalité implique que l'ellipsoïde indicateur de la loi relative à l'estimateur des moindres carrés $\tilde{Y}$ (d'équation $Z^T \Gamma_{\tilde{Y}}{}^{-1} Z = 1$) est intérieur à celui relatif à $\tilde{Y}'$, d'où le théorème :

Théorème 3.1. ***(Théorème de Linnik** [4])* *Si deux vecteurs estimateurs normaux et non biaisés $\tilde{X}$, et $\tilde{X}'$ ($E(\tilde{X}) = E(\tilde{X}')$), de même direction et $\tilde{X}$ est un estimateur des moindres carrés, alors l'ellipsoïde de variance de $\tilde{X}$ est intérieur à celui relatif à $\tilde{X}'$.*

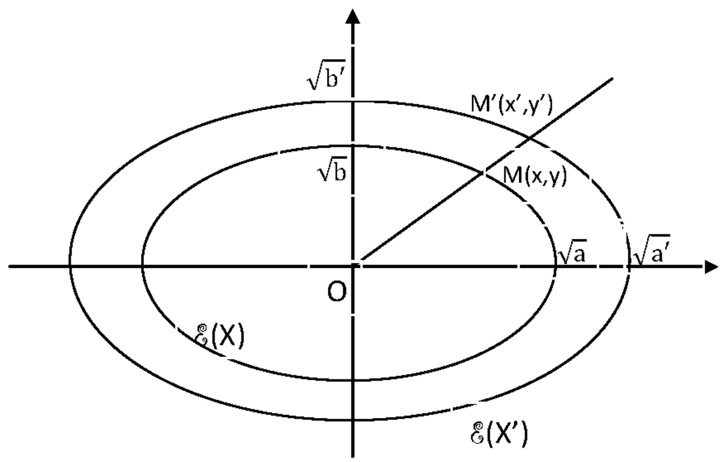

FIGURE 1. Ellipses de variance de $\tilde{X}$ et $\tilde{X}'$

Nous donnons ci-après une preuve de ce théorème :

Démonstration. On considère le cas où le vecteur $\dot{X}$ des inconnues est de dimension 2 et que les inconnues sont indépendantes non corrélées. Soient $\tilde{X}$ et $\tilde{X}'$ deux estimateurs de $\dot{X}$ non biaisés ($E(\tilde{X}) = E(\tilde{X}') = \dot{X}$) avec $\tilde{X}$ est un estimateur des moindres carrés. De 3.1, on écrit :

$$Z^T\Gamma_{\tilde{X}'}Z > Z^T\Gamma_{\tilde{X}}Z, \ \forall\ Z \tag{3.2}$$

Pour l'estimateur des moindres carrés $\tilde{X}$, on sait que : $\Gamma_{\tilde{X}} = \sigma_0^2N^{-1}$. La matrice $\sigma_0^2N^{-1}$ est de la forme :

$$\sigma_0^2N^{-1} = \begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}, \quad a > 0, b > 0$$

car N est symétrique, définie positive et que les 2 composantes de X sont des variables aléatoires indépendantes. La matrice $\Gamma_{\tilde{X}'}$ s'écrit :

$$\Gamma_{\tilde{X}'} = \begin{pmatrix} a' & c' \\ c' & b' \end{pmatrix}, \quad a'b' - c'^2 > 0 \Longrightarrow a'b' > 0$$

car $\Gamma_{\tilde{X}'}$ est aussi symétrique défine positive. Soit $Z = (x, y)$, l'équation (3.2) s'écrit :

$$(x, y).\begin{pmatrix} a' & c' \\ c' & b' \end{pmatrix}.\begin{pmatrix} x \\ y \end{pmatrix} > (x, y).\begin{pmatrix} a & 0 \\ 0 & b \end{pmatrix}.\begin{pmatrix} x \\ y \end{pmatrix} \Longrightarrow (a'-a)x^2+2c'xy+(b'-b)y^2 > 0 \tag{3.3}$$

Soit
$$P(x,y) = (a'-a)x^2 + 2c'xy + (b'-b)y^2 > 0$$
C'est une forme quadratique définie positive, dans ce cas on a nécessairement $\Delta' = c'^2 - (a'-a)(b'-b) < 0$. Posons $\alpha = a'-a, \beta = b'-b$. D'où $0 \leq c'^2 < (a'-a)(b'-b) \Longrightarrow (a'-a)(b'-b) = \alpha.\beta > 0 \Longrightarrow \alpha > 0$ et $\beta > 0$ ou $\alpha < 0$ et $\beta < 0$. Or si $x = 0, y \neq 0 \Longrightarrow b'-b > 0 \Longrightarrow b' > b$ ou $y = 0, x \neq 0 \Longrightarrow a'-a > 0 \Longrightarrow a' > 0$. Maintenant on pose $t = x/y$ avec $y \neq 0$, on obtient $P(x,y) = \dfrac{1}{y^2}(\alpha t^2 + 2c't + \beta) > 0 \Longrightarrow \alpha t^2 + 2c't + \beta > 0$. Comme $\Delta = c'^2 - \alpha.\beta < 0$, alors le coefficient α du polynôme $\alpha t^2 + 2c't + \beta$ est positif. Il s'ensuit que seul le cas $a'-a > 0$ et $b'-b > 0$ est acceptable. Donc $a' > a$ et $b' > b$.

Soit $\mathcal{E}(\tilde{X})$ l'ellipse de variance de l'estimateur $\tilde{X}$. Son équation est donnée par :
$$Z^T \Gamma_{\tilde{X}}^{-1} Z = 1 \quad Z \in \mathbb{R}^2 \tag{3.4}$$
Prenons $Z = \boldsymbol{OM} = (x,y)^T, x, y$ étant les composantes du vecteur $\boldsymbol{OM}$ (Fig. :1), alors (3.4) s'écrit :
$$\frac{x^2}{a} + \frac{y^2}{b} = 1 \tag{3.5}$$
Comme a, b étant >0 et supposons que $a > b$, on a obtenu l'équation d'une ellipse de demi-grand axe $\sqrt{a}$ et demi-petit axe $\sqrt{b}$.

De même, soit $\mathcal{E}(\tilde{X}')$ l'ellipse de variance de l'estimateur $\tilde{X}'$. Son équation est donnée par :
$$Z'^T \Gamma_{\tilde{X}'}^{-1} Z' = 1 \quad Z' \in \mathbb{R}^2 \tag{3.6}$$
Prenons $Z' = \boldsymbol{OM'} = (x', y')^T, x', y'$ étant les composantes du vecteur $\boldsymbol{OM'}$ (Fig. :1), alors (3.6) s'écrit :
$$\frac{x'^2}{a'} + \frac{y'^2}{b'} = 1 \tag{3.7}$$
Comme a', b' étant >0 et supposons que $a' > b'$, on a obtenu l'équation d'une ellipse de demi-grand axe $\sqrt{a'}$ et demi-petit axe $\sqrt{b'}$.

Comme les deux vecteurs estimateurs $\tilde{X}, \tilde{X}'$ sont colinéaires, alors $\boldsymbol{OM'} = \lambda \boldsymbol{OM}$ avec $\lambda > 0$. Comme $a' > a, b' > b \Longrightarrow \sqrt{a'} > \sqrt{a}$ et $\sqrt{b'} > \sqrt{b}$, alors l'ellipse $\mathcal{E}(\tilde{X})$ est à l'intérieur de l'ellipse $\mathcal{E}(\tilde{X}')$.

Dans le cas où $\tilde{X}, \tilde{X}'$ sont de dimension supérieure à 2, on obtient que l'ellipsoïde $\mathcal{E}(\tilde{X})$ est à l'intérieur de l'ellipsoïde $\mathcal{E}(\tilde{X}')$.

Fin de la preuve

$\square$

Références

[1] **P. Hottier**. 1980. Théorie des erreurs. Ecole Nationale des Sciences Géographiques (ENSG). IGN France.

[2] **P. Hottier, A. Ben Hadj Salem**. 2020. Théorie des erreurs. Nouvelle édition numérique. 123 pages. www.vixra.org/abs/2008.0065

[3] **P. Hottier**. 1980. Cours de statistique. ENSG. IGN France.

[4] **Yuri V. Linnik**. 1958. Méthode des Moindres Carrés et les Bases Mathématiques de la théorie Statistique et le traitement des Observations (en russe). Publications de Littérature, Physique et Mathématique de l'Etat. Moscou, 336 pages.

$(Z-M)^T B_X (Z-M)$. Совокупность точек $(z_1, \ldots, z_n)$, где эта дисперсия равна единице, будет задаваться уравнением

$$(Z-M)^T B_X (Z-M) = 1. \tag{2.4.10}$$

Для каждого вектора в направлении $Z-M$ эллипсоид (2.4.10) задает длину, при которой дисперсия соответствующей проекции будет единичной.

Мы видим, что в левых частях уравнений (2.4.5) и (2.4.10) стоят взаимно обратные квадратичные формы.

Теорема 2.4.2 *). *Если для двух n-мерных нормальных векторов X и Y с общим вектором средних $M = E(X) = E(Y)$ корреляционный эллипсоид Y лежит внутри корреляционного эллипсоида X, то эллипсоид постоянной дисперсии X лежит внутри эллипсоида постоянной дисперсии Y, и обратно.*

Мы не будем приводить здесь доказательства этой теоремы.

§ 5. Сопоставление различных нормальных распределений

Допустим, что мы производим измерение некоторой физической величины m, причем наблюдения сопровождаются случайными погрешностями ξ без систематической погрешности, так что результат наблюдений может быть выражен случайной величиной $X = m + \xi$. Мы будем считать, что X нормальна с параметрами m, σ:

$$X \in N(m, \sigma).$$

Если другой способ измерения той же величины (или другой наблюдатель) дает результаты, выражаемые случайной величиной $X_1 = m + \xi_1 \in N(m, \sigma_1)$, причем $\sigma_1 < \sigma$, то естественно считать эти результаты точнее прежних, так как величина ξ_1 „более сосредоточена" вблизи нуля, чем ξ. Точный смысл этого высказывания следующий: если задать какое-либо число $\varepsilon > 0$, то уклонения наблюдений от m более чем на ε будут более вероятны в первом случае, чем во втором, т. е.

$$P\{|X_1 - m| > \varepsilon\} < P\{|X - m| > \varepsilon\}. \tag{2.5.1}$$

В самом деле, имеем

$$P\{|X_1 - m| > \varepsilon\} = \frac{2}{\sqrt{2\pi}\,\sigma_1}\int_{\varepsilon}^{\infty} e^{-\frac{t^2}{2\sigma_1^2}}\,dt = \frac{2}{\sqrt{2\pi}}\int_{\varepsilon/\sigma_1}^{\infty} e^{-u^2/2}\,du$$

и аналогично

$$P\{|X - m| > \varepsilon\} = \frac{2}{\sqrt{2\pi}}\int_{\varepsilon/\sigma}^{\infty} e^{-u^2/2}\,du.$$

*) Автору неизвестно, нова ли эта теорема.

FIGURE 2. Citation du théorème 2.4.2. de Linnik dans [4]

Abdelmajid Ben Hadj Salem, Résidence Bousten 8, Mosquée Raoudha, Bloc B, 1181 Soukra Raoudha, Tunisia

E-mail address: abenhadjsalem@gmail.com