MATHÉMATIQUES ET SCIENCES HUMAINES

LOUIS MAILHOT

Quelques aspects statistiques des lois de probabilité réelles tronquées à droite

Mathématiques et sciences humaines, tome 90 (1985), p. 45-80

<http://www.numdam.org/item?id=MSH_1985__90__45_0>

Math. Sci. hum. (23^{e} année, n°90, *1985*, p.45-81).

QUELQUES ASPECTS STATISTIQUES DES LOIS DE PROBABILITE REELLES TRONQUEES A DROITE

Louis MAILHOT*

INTRODUCTION

L'origine de ce travail est dans le traitement statistique de données recueillies par des psychologues auprès d'enfants handicapés mentaux : il s'agissait essentiellement de comparer les performances de deux populations de sujets à une même épreuve.

Nous avons constaté que les résultats obtenus par les deux groupes étaient bien ajustés par deux lois gaussiennes de même moyenne, de même écart-type, tronquées à droite (c'est-à-dire ne conservant de la courbe "en cloche" que la partie située à gauche d'un nombre a , éventuellement différent d'une population à l'autre).

Cette situation nous a paru assez fréquente : en particulier les épreuves déterminant un quotient intellectuel sont étalonnées de sorte que le résultat d'un individu pris au hasard dans la population de référence obéisse à une loi gaussienne (de moyenne 100, d'écart-type 15 pour le WISC par exemple). Dans ces conditions une loi gaussienne tronquée à droite paraît être un modèle mathématique privilégié pour le score d'une personne ayant un certain handicap mental.

Pour répondre à la question des psychologues un test de Student nous a paru inadapté (malgré sa robustesse) les conditions de normalité théoriquement nécessaires étant loin d'être vérifiées.... Nous avons eu recours à une procédure non paramétrique, en l'occurence le test de Mann-Whitney.

*Département de Mathématiques Appliquées - Université de Clermont II

Nous nous sommes néanmoins posé la question de l'existence d'un test statistique plus puissant que le test de Mann-Whitney, adapté à cette situation : comparaison des points de troncature de deux lois gaussiennes réelles de mêmes paramètres, tronquées à droite. Il nous a paru également intéressant d'étudier l'estimation du point de troncature qui peut être considéré comme un indice mesurant le degré de réussite d'une population à une épreuve.

Nous énonçons d'abord des résultats théoriques dont la démonstration peut être trouvée dans notre thèse [7]. Nous en déduisons des conséquences sur l'estimation ponctuelle et les tests d'hypothèses portant sur un point de troncature. Nous abordons ensuite la comparaison de deux ou plusieurs points de troncature d'une même loi : après l'énoncé de quelques résultats "connus" nous étudions (par des méthodes de simulation) la puissance de trois tests. Notre travail ne se limite pas à la troncature de lois gaussiennes mais les applications numériques données concernent essentiellement ce cas, si important dans la pratique.

1 - Définitions et résultats concernant la variance d'une loi réelle tronquée

1.1 - Nous verrons que la log-concavité de certaines fonctions joue un grand rôle dans ce travail. Il nous paraît nécessaire d'en rappeler la définition.

DEFINITION 1 : Une fonction $g : \mathbb{R}^n \to \mathbb{R}^+ \cup \{+\infty\}$ est dite log-concave sur une partie convexe D de $\mathbb{R}^n$ si :

$$\forall \alpha \in]0,1[\ , \ \forall (x,y) \in D \times D$$

$$(1) \quad g(\alpha x + (1-\alpha)y) \geqq (g(x))^{\alpha} \ (g(y))^{1-\alpha}$$

Remarques :

a) Cette définition est préférable à : "Log g concave" car elle s'applique aussi aux points en lesquels g s'annule ; en particulier g étant à valeurs dans $\mathbb{R}^+ \cup \{+\infty\}$ nous conviendrons que $0 \times \infty = 0$.

b) La log-convexité d'une fonction g se définit de manière analogue, en renversant le sens de l'inégalité dans (1).

1.2 - Soit F la fonction de répartition (f.r.) d'une loi de probabilité réelle d'espérance et de variance finies et soit $A = \{a \in \mathbb{R} / 0 < F(a) < 1\}$.

DEFINITION 2 : Pour un élément a de A nous définissons la f.r. de la loi tronquée à droite en a par :

$$F_a(x) = \frac{F(x)}{F(a)} \quad \forall x \leqq a$$

$$= 1 \qquad \forall x > a$$

Si Y est une variable aléatoire (v.a.) de f.r. F, $F_a(x)$ est la probabilité de l'événement $\{Y < x\}$ conditionné par $\{Y < a\}$.

Soit :

$$M(a) = \int_{\mathbb{R}} x \, dF_a(x) \quad \text{et} \quad V(a) = \int_{\mathbb{R}} (x-M(a))^2 \, dF_a(x)$$

respectivement l'espérance et la variance de la loi tronquée à droite en a.

La propriété de croissance stochastique de la famille $\{F_a, a \in A\}$ implique la croissance de la fonction M. Nous avons établi (voir [7]) une condition générale sur F pour qu'il en soit de même de la variance V.
Posons :

$$\forall x \in \mathbb{R} \quad , \; F_{[1]}(x) = \int_{-\infty}^{x} F(t) \, dt$$

$$\text{et} \qquad F_{[2]}(x) = \int_{-\infty}^{x} F_{[1]}(t) \, dt \; .$$

Dans le cas où F estcontinue nous avons la proposition suivante :

PROPOSITION 3 : V est croissante si et seulement si $F_{[2]}$ est log-concave.

COROLLAIRE a) Si F ou $F_{[1]}$ est log-concave, V est croissante.

b) Si F a une densité f log-concave, V est croissante.

Un cas particulier important de ce corollaire est celui d'une loi normale $\mathcal{N}(m \; ; \sigma)$.
Nous avons un résultat analogue pour les troncatures à gauche : dans le cas d'une loi de densité log-concave la variance est une fonction décroissante du point de troncature. Nous en déduisons une conséquence intéressante sur la variance W d'une loi de densité log-concave, tronquée bilatéralement. W est en effet une fonction croissante de l'intervalle de troncature au sens suivant :

$$]a_1,b_1[\supset]a_2,b_2[\quad \Longrightarrow \quad W(a_1,b_1) \geqq W(a_2,b_2)$$

Dans la suite de cet article, nous omettrons le plus souvent le qualificatif "à droite" (toujours sous entendu) pour désigner les lois tronquées.

Un résultat semblable pour certaines lois discrètes peut être démontré (mais plus difficilement ...). Considérons une loi de probabilité ayant pour support S un intervalle de $\mathbb{Z}$; sa distribution est $\{(k, p_k), k \in \mathbb{Z}\}$.

PROPOSITION 4 : Sous l'hypothèse

$$\forall k \in S \ , \ F^2(k) \geqq F(k-1)\ F(k+1)$$

V est croissante.

COROLLAIRE : Sous l'hypothèse

$$\forall k \in S \ , \ p_k^2 \geqq p_{k-1}\ p_{k+1}$$

V est croissante.

Les lois de Bernoulli, binomiales, de Poisson vérifient cette hypothèse.

2 - Estimation ponctuelle du point de troncature

Différents travaux ont été faits à ce sujet. Nous citerons ceux qui nous ont paru les plus intéressants et verrons des conséquences de nos résultats du paragraphe précédent ; pour avoir une bibliographie plus complète le lecteur peut consulter l'important ouvrage de M.G. KENDALL et A. STUART [6]. Les estimations fournies sont basées sur le maximum d'un échantillon de v.a.r. ; nous verrons dans un prochain paragraphe qu'il en est de même pour les tests d'hypothèses. Nous donnons d'abord quelques propriétés de ce maximum.

2.1 - Propriétés de la loi du maximum d'un échantillon de v.a.r.

Soit F une v.a.r. (pour le moment quelconque),

F_a la f.r. obtenue à partir de F par troncature à droite en a,

$\underset{\sim}{X} = (X_1,\ldots,X_n)$ un n-échantillon d'une v.a.r. de f.r. F_a,

$\underset{\sim}{Y} = (Y_1,\ldots,Y_n)$ un n-échantillon d'une v.a.r. de f.r. F,

$\underset{\sim}{X}^* = (X_1^*,\ldots,X_n^*)$ et $\underset{\sim}{Y}^* = (Y_1^*,\ldots, Y_n^*)$ sont les statistiques d'ordre correspondant à $\underset{\sim}{X}$ et $\underset{\sim}{Y}$ respectivement et Z_k une v.a.r. dont la loi est obtenue à partir de celle de Y_k^* par troncature à droite en a.

LEMME 5 : X_n^* et Z_n ont la même répartition. Il n'en est pas ainsi, en général pour X_k^* et Z_k, $k < n$.

Démonstration : Pour k=n, $\forall x \leq a$, $P(X_n^* < x) = P(X_1 < x, \ldots, X_n < x) = \prod_{i=1}^{n} F_a(x)$

$$= \frac{F^n(x)}{F^n(a)} = \frac{P(Y_n^* < x)}{P(Y_n^* < a)} = P(\underset{\sim}{z}_n < x)$$

$$\forall x > a, \; P(X_n^* < x) = P(\underset{\sim}{z}_n < x) = 1.$$

Remarquons que $(F_a)^n = (F^n)_a$. Pour $k < n$, la f.r. de X_k^* est donnée par :

$$F_k^*(x) = \sum_{r=k}^{n} C_n^r \, (F_a(x))^r \, (1-F_a(x))^{n-r}$$

$$= \frac{n!}{(k-1)!(n-k)!} \int_0^{F_a(x)} t^{k-1}(1-t)^{n-k} dt.$$

La f.r. de $\underset{\sim}{z}_k$ est définie par :

$$\forall x \leq a, \; G_k(x) = \frac{\int_0^{F(x)} t^{k-1}(1-t)^{n-k} dt}{\int_0^{F(a)} t^{k-1}(1-t)^{n-k} dt}$$

$$\forall x > a, \; G_k(x) = 1.$$

F_k^* et G_k ne sont pas égales en général. Par exemple pour k=1,

$$\forall x \leq a, \; F_1^*(x) = 1 - (1 - \frac{F(x)}{F(a)})^n$$

tandis que :

$$G_1(x) = \frac{1-(1-F(x))^n}{1-(1-F(a))^n} \; .$$

La proposition 6 permet l'introduction de lois tronquées dans certaines distributions conditionnelles d'un échantillon.

PROPOSITON 6 : Soit $\underset{\sim}{Y} = (Y_1, \ldots, Y_n)$ un n-échantillon d'une v.a.r. Y de densité f, de f.r. F ; $\underset{\sim}{Y}^* = (Y_1^*, \ldots, Y_n^*)$ la statistique d'ordre associée ; alors la densité de $(Y_1^*, \ldots, Y_{n-1}^*)$ conditionnée par Y_n^* est celle d'une statistique d'ordre $\underset{\sim}{X}^* = (X_1^*, \ldots, X_{n-1}^*)$ d'une v.a.r. X dont la f.r. est donnée par la troncature de F en Y_n^* .

Démonstration : Soit $(y_1,\ldots,y_n)\in\mathbb{R}^n$, $y_1 \leq \ldots \leq y_n$. La densité de $(Y_1^*,\ldots, Y_{n-1}^*)$ conditionnée par Y_n^* est donnée par :

$$f_{y_n}(y_1,\ldots,y_{n-1}) = \frac{\frac{\partial^n}{\partial y_1\ldots\partial y_n} P(Y_1^* < y_1,\ldots,Y_{n-1}^* < y_{n-1}, Y_n^* < y_n)}{\frac{d}{dy_n} P(Y_n^* < y_n)}$$

Nous venons de voir que :

$$P(Y_n^* < y_n) = F^n(y_n)$$

$$\frac{d}{dy_n} P(Y_n^* < y_n) = n\, f(y_n)\, F^{n-1}(y_n)$$

$$P(Y_1^* < y_1,\ldots, Y_{n-1}^* < y_{n-1}, Y_n^* < y_n) =$$

$$\sum_{p=0}^{n-1} \left\{ \frac{n!}{(n-p)!} \left[F(y_{p+1}) - F(y_p)\right]^{n-p} \prod_{i=1}^{p} \left[F(y_i) - F(y_{i-1})\right] \right\}$$

avec, par convention, $F(y_0) = 0$.

Dans le calcul de la dérivée $n^{\text{ième}}$, seul le terme correspondant à $p = n-1$ intervient (c'est le seul qui contient $y_1,\ldots, y_n$).

Donc :

$$\frac{\partial^n}{\partial y_1\ldots\partial y_n} P(Y_1^* < y_1,\ldots,Y_n^* < y_n) = n! \frac{\partial^n}{\partial y_1\ldots\partial y_n} \left\{ \prod_{i=1}^{n} (F(y_i)-F(y_{i-1})) \right\}$$

$$= n! \prod_{i=1}^{n} f(y_i).$$

Par suite :

$$f_{y_n}(y_1,\ldots, y_{n-1}) = \frac{n! \prod_{i=1}^{n} f(y_i)}{n f(y_n) F^{n-1}(y_n)}$$

$$f_{y_n}(y_1,\ldots, y_{n-1}) = (n-1)\,! \prod_{i=1}^{n-1} \left[\frac{f(y_i)}{F(y_n)}\right].$$

Par ailleurs soit $\tilde{X} = (X_1,\ldots, X_{n-1})$ un $(n-1)$-échantillon d'une v.a.r. X de f.r. F_{y_n} (F tronquée à droite en y_n) ; la densité de $\tilde{X}$ est donnée par :

$$g(y_1,\ldots, y_{n-1}) = \prod_{i=1}^{n-1} \left[\frac{f(y_i)}{F(y_n)}\right]$$

et celle de $\tilde{X}^*$, la statistique d'ordre associée, est :

$$h(y_1,\ldots,y_{n-1}) = \frac{\partial^{n-1}}{\partial y_1 \ldots \partial y_{n-1}} P(X_1^* < y_1,\ldots, X_{n-1}^* < y_{n-1})$$

$$= (n-1)\,! \prod_{i=1}^{n-1} \left[\frac{f(y_i)}{F(y_n)} \right]$$

(calculs analogues à ceux vus plus haut) pour :

$$y_1 \leqq \cdots \leqq y_{n-1} .$$

Nous avons donc bien $f_{y_n}(y_1,\ldots, y_{n-1}) = h(y_1,\ldots, y_{n-1})$ pour :

$$y_1 \leqq \cdots \leqq y_{n-1} .$$

Bien sûr, s'il existe $i < j$ tel que $y_i > y_j$

$$f_{y_n}(y_1,\ldots,y_{n-1}) = h(y_1,\ldots, y_{n-1}) = 0 . \blacksquare$$

Remarques :

a) Nous n'avons pas de proposition analogue dans le cas discret. La distribution de $(Y_1^*,\ldots, Y_{n-1}^*)$ conditionnée par Y_n^* est alors donnée par :

$$P_{y_n}(y_1,\ldots,y_{n-1}) = \frac{P(Y_1^* = y_1,\ldots, Y_n^* = y_n)}{P(Y_n^* = y_n)} .$$

Posons $p_i = P(Y=y_i)$.

Lorsque $y_1 < \ldots < y_{n-1} < y_n$, $P_{y_n}(y_1,\ldots,y_{n-1}) = \dfrac{n! \prod_{i=1}^{n} p_i}{np_n F^{n-1}(y_n)}$

soit

$$P_{y_n}(y_1,\ldots, y_{n-1}) = (n-1)\,! \prod_{i=1}^{n-1} \frac{p_i}{F(y_n)} .$$

Par contre, si certains y_i sont égaux, nous avons une expression différente ; soit par exemple $y_1 = y_2$.

$$\frac{P(Y_1^* = y_1 \,,\, Y_2^* = y_1)}{P(Y_2^* = y_1)} = \frac{p_1^2}{2p_1 F(y_1)} = \frac{p_1}{2F(y_1)}$$

qui n'a pas la forme précédente.

b) La proposition 6 serait fausse si nous remplacions le (n-1)-échantillon par un k-échantillon, $k < n-1$. Calculons par exemple :

$$f_{y_3}(y_1) = \frac{\frac{\partial^2}{\partial y_1 \partial y_3} P(Y_1^* < y_1, Y_3^* < y_3)}{\frac{d}{dy_3} P(Y_3^* < y_3)} \quad , \; y_1 \leq y_3$$

$$P(Y_1^* < y_1, Y_3^* < y_3) = F^3(y_1) + 3F^2(y_1)\left[F(y_3) - F(y_1)\right]$$

$$+ 3F(y_1)\left[F(y_3) - F(y_1)\right]^2$$

$$= F^3(y_1) - 3F^2(y_1)\,F(y_3) + 3F(y_1)\,F^2(y_3)$$

$$\frac{\partial^2}{\partial y_1 \partial y_3} P(Y_1^* < y_1, Y_3^* < y_3) = 6f(y_1)\,f(y_3)\left[F(y_3) - F(y_1)\right]$$

et :

$$\frac{d}{dy_3} P(Y_3^* < y_3) = 3f(y_3)\,F^2(y_3)$$

d'où :

$$f_{y_3}(y_1) = \frac{2f(y_1)\left[F(y_3) - F(y_1)\right]}{F^2(y_3)} \neq \frac{f(y_1)}{F(y_3)}$$

2.2 - Estimation du point de troncature par la méthode du maximum de vraisemblance.

Rappelons qu'étant donné un n-échantillon $\underset{\sim}{X} = (X_1,\ldots,X_n)$ d'une v.a.r. X de distribution $f_\theta(x)$, dépendant d'un paramètre θ à estimer, l'ensemble des valeurs possibles de θ étant Θ , la vraisemblance est la fonction de Θ dans $\mathbb{R}^+$ qui, pour une valeur observée $\underset{\sim}{x}$ de $\underset{\sim}{X}$ associe à θ le nombre $L_{\underset{\sim}{x}}(\theta) = \prod_{i=1}^{n} f_\theta(x_i)$. Une application T de l'espace $\mathcal{X}$ des observations dans Θ qui réalise un maximum de $L_{\underset{\sim}{x}}$ pour tout $\underset{\sim}{x}$ est un estimateur du maximum de vraisemblance pour θ :

$$L_{\underset{\sim}{x}}(T(\underset{\sim}{x})) \geq L_{\underset{\sim}{x}}(\theta) \quad , \quad \forall \underset{\sim}{x} \in \mathcal{X} \quad , \quad \forall \theta \in \Theta \quad .$$

Nous considérons ici le cas d'une f.r. F de densité f, le paramètre est le point de troncature a, Θ est un intervalle de $\mathbb{R}$. Pour une valeur observée $\underset{\sim}{x} = (x_1,\ldots, x_n)$ de $\underset{\sim}{X}$,

$$L_{\underset{\sim}{x}}(a) = F^{-n}(a) \prod_{i=1}^{n} f(x_i)\, 1_{\{\max x_i \leq a\}} \; ;$$

F^{-n} étant une fonction décroissante de a, $T(\underset{\sim}{x})$ associe à $\underset{\sim}{x}$ la plus petite des valeurs possibles de a. Comme $a \geq \text{Max}(x_i)$, nous prenons $T(\underset{\sim}{x}) = \text{Max } x_i$, soit avec les notations ci-dessus :

$$T(X_1,\ldots,X_n) = X_n^* .$$

De plus T est un estimateur <u>exhaustif</u> pour a. En effet rappelons qu'une condition nécessaire et suffisante pour qu'il en soit ainsi est que pour toute valeur observée $\underset{\sim}{x}$ de $\underset{\sim}{X}$ la vraisemblance puisse s'écrire comme le produit de deux fonctions : l'une, g, dépendant seulement de a et de $T(\underset{\sim}{x})$, l'autre, h, seulement de $\underset{\sim}{x}$.

Prenant :

$$g(a,T(\underset{\sim}{x})) = F^{-n}(a)\ 1_{\{X_n^*(\underset{\sim}{x}) \leq a\}} \quad \text{et } h(\underset{\sim}{x}) = \prod_{i=1}^{n} f(x_i)$$

nous voyons que

$$L_{\underset{\sim}{X}}(a) = g(a,T(x)).h(\underset{\sim}{x})$$

et en déduisons que X_n^* est exhaustive pour a.

<u>PROPOSITION 7</u> : Soit F une f.r. telle que $\int_{\mathbb{R}} x^2\, dF(x) < \infty$, a appartenant à A et $a' = \text{Inf}\{\, x/F(x) = F(a)\,\}$. X_n^* est alors un estimateur de a' asymptotiquement sans biais et convergent.

<u>Démonstration</u> :

a) Une intégration par parties montre que l'espérance M(a) d'une v.a.r. de f.r. F_a s'écrit :

$$M(a) = \int_{\mathbb{R}} x\, dF_a(x) = a - \int_{-\infty}^{a} F_a(x)dx$$

$$= a - \int_{-\infty}^{a} \frac{F(x)}{F(a)}\, dx.$$

Pour X_n^* nous obtenons donc :

$$E(X_n^*) = a - \int_{-\infty}^{a} \left(\frac{F(x)}{F(a)}\right)^n dx$$

et aussi :

$$E(X_n^*) = a - \int_{-\infty}^{a'} \left(\frac{F(x)}{F(a)}\right)^n dx - \int_{a'}^{a} dx$$

$$= a' - \int_{-\infty}^{a'} \left(\frac{F(x)}{F(a)}\right)^n dx$$

X_n^* est donc un estimateur <u>biaisé</u> de a et de a'. Par exemple, dans le cas de la troncature d'une loi uniforme sur $[0,1]$, $E(X_n^*) = \frac{n}{n+1}\, a$, $\frac{n+1}{n}\ X_n^*$ étant alors non biaisé.

Toutefois $\lim_{n\to\infty} E(X_n^*) = a'$.

Pour s'en convaincre, il suffit d'appliquer le théorème de convergence dominée de Lebesgue.

La condition : $\forall x < a'$, $F(x) < F(a)$ implique $(\frac{F(x)}{F(a)})^n \xrightarrow[n\to\infty]{} 0$.

Or,

$$\forall n \geq 1 \ , \ (\frac{F(x)}{F(a)})^n \leq \frac{F(x)}{F(a)} \quad \text{avec} \quad \int_{-\infty}^{a} \frac{F(x)}{F(a)}\, dx = a - E(X) < \infty .$$

b) Montrons que X_n^* est convergent et pour ceci que var $(X_n^*) \longrightarrow 0$. Des intégrations par parties conduisent à (voir [7]) :

$$\text{var}(X_n^*) = 2a \int_{-\infty}^{a} (\frac{F(x)}{F(a)})^n\, dx - (\int_{-\infty}^{a} (\frac{F(x)}{F(a)})^n\, dx)^2$$

$$-2 \int_{-\infty}^{a} x\, (\frac{F(x)}{F(a)})^n\, dx .$$

Tenant compte du fait que $\forall x \in]a',a]$, $F(x)=F(a)$, nous obtenons :

$$\begin{aligned}
\text{var}(X_n^*) &= 2a\left[\int_{-\infty}^{a'} (\frac{F(x)}{F(a)})^n\, dx + a-a'\right] - \left[\int_{-\infty}^{a'} (\frac{F(x)}{F(a)})^n\, dx + a-a'\right]^2 \\
&\quad - 2\left[\int_{-\infty}^{a'} x(\frac{F(x)}{F(a)})^n\, dx + \frac{a^2 - a'^2}{2}\right] \\
&= 2a \int_{-\infty}^{a'} (\frac{F(x)}{F(a)})^n\, dx - \left[\int_{-\infty}^{a'} (\frac{F(x)}{F(a)})^n dx\right]^2 - 2(a-a') \int_{-\infty}^{a'} (\frac{F(x)}{F(a)})^n\, dx \\
&\quad - 2 \int_{-\infty}^{a'} x\, (\frac{F(x)}{F(a)})^n\, dx .
\end{aligned}$$

Appliquant de nouveau le théorème de Lebesgue pour chacune des intégrales nous constatons que $\lim \text{var}(X_n^*) = 0$. En particulier pour la dernière intégrale nous avons :

$$|x(\frac{F(x)}{F(a)})^n| \leq |x \frac{F(x)}{F(a)}| ,$$

fonction d'intégrale finie (car $\int_{\mathbb{R}} x^2\, dF(x) < +\infty$ par hypothèse) et

$$\forall x < a' \ , \ x(\frac{F(x)}{F(a)})^n \xrightarrow[n\to\infty]{} 0 .$$

COROLLAIRE :

a) Soit F strictement croissante sur A ; alors pour tout a de A, X_n^* est un estimateur asymptotiquement sans biais et convergent de a.

b) Soit F la f.r. d'une v.a.r. discrète dont les valeurs a_i, $i \in \mathbb{Z}$, sont numérotées par ordre croissant ; si les v.a.r. X_j, $j=1,\ldots,n$ ont pour f.r. F_{a_i} alors X_n^* est un estimateur asymptotiquement sans biais et convergent de a_{i-1}.

La proposition 8 établit une propriété de la quantité d'information de Fisher apportée par le n-échantillon $\underset{\sim}{X}$ sur a lorsque varie n ou a, dans le cas où F a une densité f.

PROPOSITION 8 : La quantité d'information de Fisher $I_{\underset{\sim}{X}}(a)$ apportée par le n-échantillon $\underset{\sim}{X}$ est proportionnelle à n^2. Pour n fixé, $I_{\underset{\sim}{X}}(a)$ est une fonction décroissante (respectivement croissante) de a dans le domaine de log-concavité (respectivement de log-convexité) de F.

Démonstration : Rappelons que $I_{\underset{\sim}{X}}(a) = E_a\left(\frac{\partial}{\partial a} \text{Log } L_{\underset{\sim}{X}}(a)\right)^2$

$$\text{Log } L_{\underset{\sim}{x}}(a) = -n \text{ Log } F(a) + \sum_{i=1}^{n} \text{Log } f(x_i) \text{ si } \max x_i \leqq a$$

$$\frac{\partial}{\partial a} \text{Log } L_{\underset{\sim}{X}}(a) = -n \frac{f(a)}{F(a)} \text{ si } \max x_i \leqq a$$

et :

$$I_{\underset{\sim}{X}}(a) = \int_{-\infty}^{a} \cdots \int_{-\infty}^{a} \left(-n \frac{f(a)}{F(a)}\right)^2 \frac{1}{F^n(a)} \prod_{i=1}^{n} f(x_i)\, d\underset{\sim}{x}$$

$$\boxed{I_{\underset{\sim}{X}}(a) = n^2\left(\frac{f(a)}{F(a)}\right)^2}$$

soit :

$$I_{\underset{\sim}{X}}(a) = n^2\left(\frac{d}{da} \text{Log } F(a)\right)^2.$$

Nous en déduisons la deuxième partie de la proposition. ■

Remarques :

a) X_n^* étant exhaustive $I_{\underset{\sim}{X}}(a) = I_{X_n^*}(a)$.

b) La propriété d'additivité de la quantité d'information lorsque le support ne dépend pas du paramètre étudié, n'est pas vraie ici :

$$I_{\underset{\sim}{X}}(a) = n^2 I_X(a) \neq n I_X(a) .$$

PROPOSITION 9 : Lorsque F est log-concave, la variance de X_n^* est une fonction croissante de a.

Démonstration : Supposant F log-concave, F^n l'est aussi. Or la f.r. de X_n^* est $(F^n)_a$ (d'après le lemme 5). La proposition 4 permet alors d'affirmer que la variance de X_n^* est une fonction croissante de a. ■

Conséquence : En tenant compte du fait que X_n^* est asymptotiquement sans biais, l'inverse de la variance mesure pour n grand la précision de cet estimateur ; la proposition 9 nous dit que lorsque F est log-concave, cette précision est d'autant plus faible que la valeur à estimer est grande.

Pour n quelconque définissant la précision comme étant l'inverse de l'erreur quadratique moyenne :

$$e(T_n) = \frac{1}{E(T_n-a)^2}$$

nous avons la proposition suivante :

PROPOSITION 10 : Lorsque F est log-concave, la précision de X_n^* est une fonction décroissante de a.

Démonstration : Pour n fixé soit :

$$g(a) = E(X_n^* - a)^2 \quad \left(= \frac{1}{e(X_n^*)}\right)$$

$$g(a) = E\left[(X_n^* - E(X_n^*)) + (E(X_n^*) - a)\right]^2$$

$$= E(X_n^* - E(X_n^*))^2 + (E(X_n^*) - a)^2 .$$

La f.r. de X_n^* étant $(F^n)_a$, et F^n étant log-concave d'après l'hypothèse sur F, le premier terme est une fonction croissante de a. D'autre part, $h(a)=a-E(X_n^*)$ est positif, et, $h(a) = \frac{(F^n)_{[1]}(a)}{F^n(a)}$ où $(F^n)_{[1]}(a) = \int_{-\infty}^{a} F^n(x)\,dx$.

Or, si une fonction est log-concave, sa primitive qui s'annule en $-\infty$ est elle-même log-concave : voir par exemple A. PREKOPA [9]. $(F^n)_{[1]}$ est donc log-concave. Par suite h est croissante ainsi que h^2. Il en est de même pour g d'où la proposition.

2.3 - Amélioration de l'estimation de a

L'application de résultats importants établis par Blackwell, Rao, Lehmann et Scheffé (voir par exemple M.G. KENDALL et A. STUART [6]) a permis à R.F.TATE [11] d'obtenir un estimateur sans biais, de variance minimum de a, basé sur X_n^*, lorsque F a une densité f. D'une façon plus générale si ξ est une fonction de A dans $\mathbb{R}$, dérivable, le seul estimateur sans biais de $\xi(a)$ est donné par :

$$T_n = \xi(X_n^*) + \frac{\xi'(X_n^*)}{n} \frac{F(X_n^*)}{f(X_n^*)} .$$

Si $\xi(a) = a$, nous avons :

$$T_n = \Psi_n(X_n^*) = X_n^* + \frac{1}{n} \frac{F(X_n^*)}{f(X_n^*)}$$

Remarque : Dans le cas de la troncature d'une loi uniforme, nous retrouvons $T_n = \frac{n+1}{n} X_n^*$.

Pour une loi gaussienne $\mathcal{N}(0;1)$ tronquée en a, notons la $\mathcal{N}_a(0;1)$,

$T_n = X_n^* - \frac{1}{nM(X_n^*)}$ où $M(a) = E(X_a)$ est l'espérance d'une v.a. de loi $\mathcal{N}_a(0;1)$.

Il est en effet facile de vérifier que pour une telle loi :

$$M(a) = -\frac{\Phi'(a)}{\Phi(a)} \quad , \text{ où } \quad \Phi(a) = \frac{1}{\sqrt{2\Pi}} \int_{-\infty}^{a} \exp(-\frac{x^2}{2})dx$$

Nous avons construit une table donnant $\Psi_n(t)$ pour n=10, 20, 30 et $-3 \leqq t \leqq 3$ par pas de 0.2 dans le cas d'une loi $\mathcal{N}_a(0;1)$. Pour une loi $\mathcal{N}_a(m;\sigma)$, nous avons :

$$F(X_n^*) = \Phi(\frac{X_n^* - m}{\sigma}) \text{ et } f(X_n^*) = \frac{1}{\sigma} \Phi'(\frac{X_n^* - m}{\sigma}).$$

Dans ce cas, nous calculons d'abord $y_n = \frac{x_n^* - m}{\sigma}$, puis $t_n = x_n^* - \frac{\sigma}{nM(y_n)}$ donnera une estimation ponctuelle de a lorsque m et σ sont connus.

Tableau 1 - Valeurs numériques de $\Psi_n(t)$ pour n=1Ø, 2Ø, 3Ø (cas d'une loi normale $\mathcal{N}(0;1)$).

T	n=1Ø	n=2Ø	n=3Ø
-3	-2.9695	-2.9848	-2.9898
-2.8	-2.7677	-2.7839	-2.7892
-2.6	-2.5657	-2.5818	-2.5886
-2.4	-2.3634	-2.3817	-2.3878
-2.2	-2.16Ø8	-2.18Ø4	-2.1869
-2	-1.9579	-1.9789	-1.986Ø
-1.8	-1.7545	-1.7772	-1.7848
-1.6	-1.55Ø6	-1.5753	-1.5835
-1.4	-1.3461	-1.373Ø	-1.382Ø
-1.2	-1.14Ø7	-1.17Ø4	-1.18Ø2
-1	- .9344	- .9672	- .9781
- .8	- .7269	- .7634	- .7756
- .6	- .5177	- .5588	- .5726
- .4	- .3Ø64	- .3532	- .3688
- .2	- .Ø924	- .1462	- .1641
Ø	.1253	.Ø627	.Ø418
.2	.3481	.2741	.2494
.4	.578Ø	.489Ø	.4593
.6	.8178	.7Ø89	.6726
.8	1.Ø721	.936Ø	.89Ø7
1	1.3477	1.1739	1.1159
1.2	1.6557	1.4279	1.3519
1.4	2.Ø139	1.7Ø7Ø	1.6Ø46
1.6	2.4521	2.Ø261	1.884Ø
1.8	3.Ø211	2.41Ø6	2.2Ø7Ø
2	3.81ØØ	2.9Ø5Ø	2.6Ø33
2.2	4.9797	3.5899	3.1266
2.4	6.8288	4.6144	3.8763
2.6	9.9278	6.2639	5.Ø426
2.8	15.4Ø12	9.1ØØ6	7.ØØØ4
3	25.5335	14.2667	1Ø.5112

La table de $\Psi_n(t)$ dans le cas d'une loi gaussienne nous montre que $\lambda_n(t) = \Psi_n(t)-t$ est alors croissante. Ce résultat est plus général.

PROPOSITION 11 : Pour n fixé, λ_n est croissante (respectivement décroissante) si F est log-concave (respectivement log-convexe sur A).
Cette proposition est une conséquence immédiate de l'expression de λ_n en fonction de F et f.

Comme cas particulier, citons celui où F est la f.r. d'une loi exponentielle négative. λ_n est alors constante :

$$F(a) = \exp(\theta a),\ \theta > 0,\quad a \leqq 0$$

$$\lambda_n(a) = \frac{1}{n\theta}\ .$$

Days le cas d'une loi $\mathcal{N}_a(0;1)$ nous donnons ci-après une valeur approchée de $\mathrm{var}(T_n)$, obtenue par simulation, pour différentes valeurs de a et de n.

Tableau 2 - Estimations non biaisées de l'espérance et de la variance de deux estimateurs du point de troncature a d'une loi gaussienne $\mathcal{N}(0;1)$, à partir de 50 échantillons de taille n.

n=1Ø	a	$\bar{t}$	s'2	$\bar{x}^*$	s^{*2}
	-1	-Ø.9818	Ø.ØØ2Ø	-1.Ø458	Ø.ØØ19
	-Ø.5	-Ø.5Ø79	Ø.ØØ98	-Ø.59Ø8	Ø.ØØ89
	Ø	Ø.ØØØ7	Ø.Ø112	-Ø.1145	Ø.ØØ96
	Ø.5	Ø.4985	Ø.Ø291	Ø.3298	Ø.Ø225
	1	Ø.9653	Ø.Ø8Ø3	Ø.7Ø97	Ø.Ø513
	1.5	1.4279	Ø.1964	1.Ø298	Ø.Ø91Ø
	2	2.Ø934	Ø.57Ø7	1.3738	Ø.1417
	2.5	2.8758	2.954Ø	1.5651	Ø.2799
	3	2.8512	8.8731	1.4781	Ø.2641
n=3Ø	-1	-1.ØØ17	Ø.ØØØ5	-1.Ø233	Ø.ØØØ5
	-Ø.5	-Ø.5Ø72	Ø.ØØ15	-Ø.5357	Ø.ØØ14
	Ø	-Ø.ØØØ3	Ø.ØØ18	-Ø.Ø4Ø8	Ø.ØØ17
	Ø.5	Ø.4941	Ø.ØØ45	Ø.4327	Ø.ØØ4Ø
	1	1.ØØ49	Ø.ØØ68	Ø.9Ø19	Ø.ØØ54
	1.5	1.5152	Ø.Ø328	1.3271	Ø.Ø211
	2	2.Ø887	Ø.13ØØ	1.71Ø5	Ø.Ø535
	2.5	2.5745	Ø.5747	1.9144	Ø.129Ø
	3	3.Ø288	2.4548	1.9964	Ø.2Ø37

La première colonne donne la vraie valeur de a, dans les deuxième et troisième figurent les estimations $\bar{t}$ et s'^2 de $a = E(T_n)$ ($T_n = X_n^* + \frac{1}{n}\frac{\Phi(X_n^*)}{\Phi'(X_n^*)}$) et de var (T_n) ; l'estimation $\bar{x}^*$ de $E(X_n^*)$ se trouve dans la quatrième et celle, s^{*2}, de $\text{var}(X_n^*)$ dans la dernière.

Notons que $\bar{t}$ donne une meilleure approximation de a que ne le fait $\bar{x}^*$. Remarquons aussi la croissance de s'^2 et s^{*2} en fonction de a : l'estimation de a par T_n est d'autant plus précise que la valeur à estimer est plus faible.

Pour a assez grand (≥ 1) la variance de l'estimateur sans biais est très importante ce qui constitue un sérieux handicap pour l'utilisation de T_n.

Suivant des travaux déjà anciens de M. QUENOUILLE, D.S. ROBSON et J.H.WHITLOCK [10] ont construit d'autres estimateurs de a, certes biaisés , mais ne nécessitant pas le calcul de $F(X_n^*)$ et $f(X_n^*)$ donc très intéressants lorsque certains autres paramètres (m ou σ pour une loi normale par exemple) sont inconnus. La condition de cette estimation est que F admette un développement de Taylor au voisinage de a et $F'(a) \neq 0$ pour tout a de A.

Partant d'un n-échantillon $\tilde{X}$, pour obtenir un estimateur dont le biais est d'ordre $n^{-(k+1)}$ il suffit de prendre :

$$T_n^{(k)} = \sum_{i=0}^{k} (-1)^i C_{k+1}^{i+1} X_{n-i}^*$$

pour k=0 nous retrouvons $T_n^{(0)} = X_n^*$

et : $$T_n^{(1)} = 2 X_n^* - X_{n-1}^*$$

$$T_n^{(2)} = 3 X_n^* - 3 X_{n-1}^* + X_{n-2}^*$$

(rappelons que $\tilde{X}^* = (X_1^*, \ldots, X_n^*)$ est la statistique d'ordre de $\tilde{X}=(X_1,\ldots,X_n)$).

Cependant, le fait qu'interviennent plusieurs X_{n-i} a en général pour conséquence une diminution de la précision. Robson et Whitlock ont en effet montré que l'erreur quadratique est :

$$E(T_n^{(k)} - a)^2 = C_{2k}^k \, n^{(-2)} \left(\frac{F(a)}{F'(a)}\right)^2 + 0(n^{-3}).$$

Toutefois, pour k=0 et k=1, $T_n^{(0)}$ et $T_n^{(1)}$ ont la même erreur quadratique sauf éventuellement pour les termes d'ordre n^{-3}. Pour a assez grand, nous limitant au terme principal (d'ordre n^{-2}), nous remarquons encore que :

- si F est log-concave, l'erreur quadratique croît avec a donc la précision de $T_n^{(k)}$ est une fonction décroissante de a
- si F est log-convexe sur A, la précision de $T_n^{(k)}$ croît avec a.

Dans le cas simple d'une loi uniforme $T_n^{(1)}$ est sans biais ; comparant sa précision à celle de $T_n' = \frac{n+1}{n} X_n^*$ (aussi sans biais) nous constatons que :

$$\lim_{n\to\infty} \frac{\mathrm{Var}(T_n^{(1)})}{\mathrm{var}(T_n')} = 2$$

c'est à dire que T'_n est asymptotiquement deux fois plus précis que $T_n^{(1)}$.

Nous terminons ce paragraphe en signalant d'autres travaux traitant de l'estimation de certains paramètres de lois tronquées lorsque a est connu ou non. R.L. PLACKETT [8] s'est intéressé à l'estimation de λ pour une loi de Poisson tronquée en a=1. A.C. COHEN et J. WOODWARD [3] établissent des estimations de m et σ pour une loi gaussienne $\mathcal{N}(m;\sigma)$ tronquée à gauche en un point a connu. A.HALD reprend dans [5] ces résultats et indique qu'il peut être intéressant d'utiliser du papier gausso-arithmétique pour avoir une valeur approchée de m, σ et a lorsqu'est connu le nombre $\tau = \Phi(\frac{a-m}{\sigma})$ (qui est en quelque sorte le "degré" de troncature). Si H_n est la fonction de répartition empirique associée à un n-échantillon $\tilde{X}$ d'une v.a r X de loi $\mathcal{N}_a(m;\sigma)$ nous avons :

$$\forall x \leqq a \quad H_n(x) \simeq \frac{1}{\tau} \Phi(\frac{x-m}{\sigma})$$

$$\forall x > a \quad H_n(x) \simeq 1$$

donc :

$$\forall x \leqq a \quad x \simeq m+\sigma \Phi^{-1}(u)$$

avec :

$$u = \tau H_n(x).$$

Les points $(x, \Phi^{-1}(\tau H_n(x)))$ sont donc approximativement alignés pour $x \leqq a$. Sur le papier gausso-arithmétique portant x en abscisse (échelle arithmétique) et en ordonnée (échelle gaussienne) $100\ \tau\ H_n(x)$ nous obtenons une valeur approchée de m en prenant l'abscisse du point d'ordonnée 50 ; l'écart-type σ s'obtient en prenant la différence d'abscisse entre le point d'ordonnée 84.13 et m (ou le point d'ordonnée 15.87 et m). Si τ est très faible, il peut être nécessaire de prolonger au delà du point d'abscisse a cette demi-droite pour faire ces estimations. Nous donnons ci-après deux exemples :

a) m=10 , σ = 2 ; a=11 ; n=200 (donc $\tau \simeq 0.6914$)

b) m=20 , σ = 3 ; a=17 ; n=500 (donc $\tau \simeq 0.1587$)

Le cas où F est la f.r. d'une loi discrète a été étudié en particulier par C. CHARALAMBIDES [2] .

Tableau 3 - Valeurs numériques utilisées pour la représentation sur papier gausso-arithmétique

a) échantillon de taille 200 d'une v.a.r. $\mathcal{N}(10;2)$ tronquée en 11

b) échantillon de taille 500 d'une v.a.r. $\mathcal{N}(20;3)$ tronquée en 17

a) TAU = .691462461

Abscisse	Ordonnée	F.R. Empirique	F.R. Théorique
4	.34	.ØØ49	.ØØ19
4.5	1.Ø3	.Ø149	.ØØ43
5	1.38	.Ø199	.ØØ89
5.5	2.76	.Ø399	.Ø176
6	5.18	.Ø75Ø	.Ø329
6.5	5.87	.Ø849	.Ø579
7	8.29	.1199	.Ø966
7.5	18.82	.2ØØØ	.1527
8	17.63	.255Ø	.2294
8.5	27.65	.4ØØØ	.3277
9	34.91	.5Ø49	.4462
9.5	44.59	.645Ø	.58Ø3
1Ø	52.89	.7649	.7231
1Ø.5	62.23	.8999	.8658
11	69.14	1	1

b) TAU = .158655254

Abscisse	Ordonnée	F.R. Empirique	F.R. Théorique
1Ø	.Ø3	.ØØ2Ø	.ØØ27
1Ø.5	.Ø3	.ØØ2Ø	.ØØ48
11	.Ø3	.ØØ2Ø	.ØØ85
11.5	.19	.Ø119	.Ø145
12	.28	.Ø179	.Ø241
12.5	.44	.Ø279	.Ø391
13	.98	.Ø62Ø	.Ø618
13.5	1.42	.Ø899	.Ø953
14	2.18	.1379	.1433
14.5	3.2Ø	.2Ø19	.21Ø3
15	4.82	.3Ø4Ø	.3Ø12
15.5	6.79	.4279	.421Ø
16	9.36	.59ØØ	.5749
16.5	12.59	.7339	.7668
17	15.86	1	1

Exemple a) :

$100\ \tau\ H_{200}(x_j)$

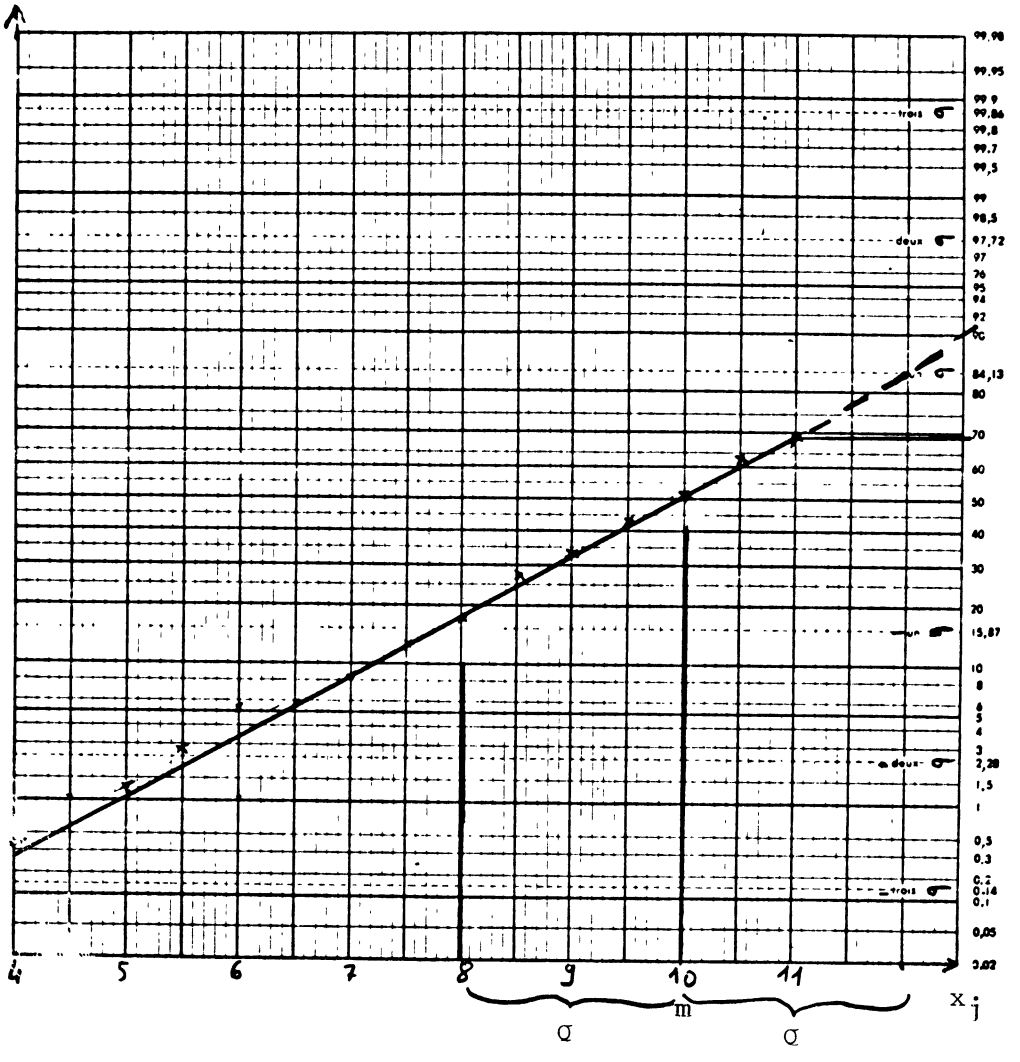

Figure 1 - Echantillon de taille 200 d'une loi $\mathcal{N}(10;2)$ tronquée en 11

Exemple b) :

$100\ \tau\ H_{500}(x_j)$

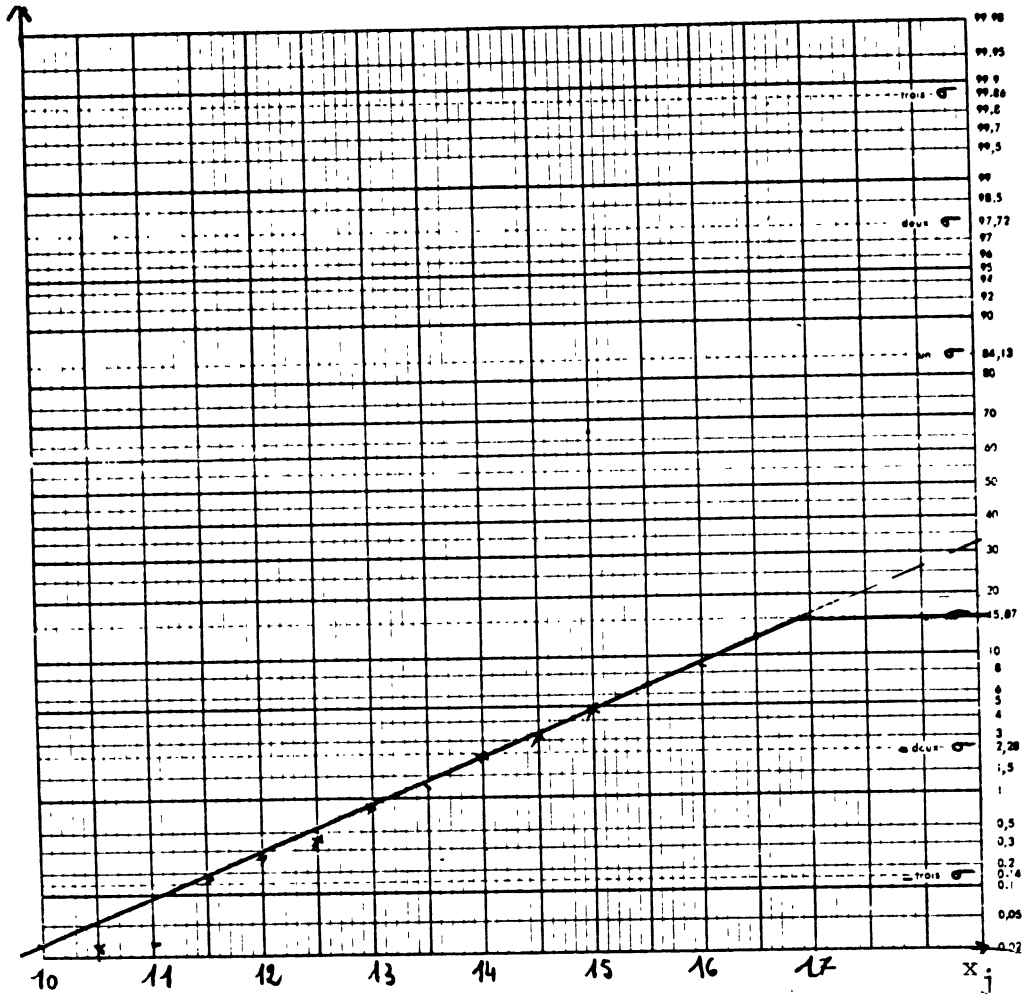

Figure 2 - Echantillon de taille 500 d'une loi $\mathcal{N}(20;3)$ tronquée en 17

L'estimation de m et σ est d'autant plus délicate que τ est faible

3 - Tests d'hypothèses portant sur un point de troncature dans le cas d'une seule population

Soit $\underset{\sim}{X} = (X_1,\ldots,X_n)$ un n-échantillon de la v.a.r. X de f.r. F_a. Pour la commodité de l'exposé et compte tenu de notre objectif (application à une loi gaussienne tronquée) nous supposons que F est absolument continue par rapport à la mesure de Lebesgue sur $\mathbb{R}$. La plupart de nos résultats sont encore valables sans cette hypothèse ; la différence essentielle est que dans certains cas de lois présentant des masses ponctuelles (par exemple lois discrètes), il est parfois nécessaire de faire un "tirage au sort" avant de prendre une décision. Dans ce paragraphe, nous étudions les tests :

1° $H_0 : a \leq a_o$ contre $H_1 : a > a_o$ au seuil α

2° $H_0 : a \geq a_o$ contre $H_1 : a < a_o$ au seuil α

3° $H_0 : a = a_o$ contre $H_1 : a \neq a_o$ au seuil α,

α étant un réel donné, $0 < \alpha < 1$.

Nous avons vu au paragraphe précédent que $T = X_n^*$ est une statistique exhaustive pour a. Nos tests sont construits à partir de T par l'application du lemme de Neyman et Pearson (voir par exemple l'ouvrage de T.S. FERGUSON [4], page 201).

3.1 - Test $H_0 : a \leq a_o$ contre $H_1 : a > a_o$; seuil α

PROPOSITION 12 : Le test défini par le domaine de rejet $\mathcal{R}$ de H_0 :

$$\mathcal{R} =]C(a_o), +\infty[\text{, } C(a_o) \text{ tel que } F(C(a_o)) = (1-\alpha)^{1/n} F(a_o)$$

est uniformément le plus puissant (U.P.P.) pour tester H_0 contre H_1 au seuil α. De plus il est convergent.

Démonstration : Appliquant le lemme de Neyman et Pearson au test entre hypothèses simples : $H_0' : a = a_o$ contre $H_1' : a = a_1 > a_o$ au seuil α, nous obtenons pour domaine de rejet de H_0' :

$$\mathcal{R} =]C(a_o), +\infty[$$

où $C(a_o)$ est tel que :

$$P_{a_o}(T > C(a_o)) = \alpha$$

c'est à dire :

$$1 - \frac{F^n(C(a_o))}{F^n(a_o)} = \alpha$$

donc :

$$F(C(a_o)) = (1-\alpha)^{1/n} F(a_o).$$

puisque $C(a_o)$ est indépendant de a_1 nous en déduisons que ce test est uniformément le plus puissant pour tester H_0' contre H_1 : $a > a_o$ au seuil α .

Par ailleurs, la fonction

$$\beta_n : \mathbb{R} \longrightarrow [0,1]$$

$$a \rightsquigarrow \beta_n(a) = P_a(T \in \mathscr{R}) = 1 - \frac{F^n(C(a_o))}{F^n(a)}$$

est croissante ; donc :

$$\forall a \leq a_o \ , \ P_a(T \in \mathscr{R}) \leq \alpha \ .$$

Le test proposé est donc U.P.P. pour tester H_0 contre H_1.

Remarque : Une autre façon d'obtenir ce résultat est de vérifier que la famille $\{F_a, a \in A\}$ a un rapport de vraisemblance monotone et d'appliquer ensuite un théorème de Karlin et Rubin (1956) : voir [4] , p. 210. Ce test est convergent c'est à dire que la probabilité d'une erreur de deuxième espèce tend vers zéro lorsque n augmente indéfiniment. En effet cette probabilité est donnée, pour $a > a_o$, par :

$$P_a(T \leq C(a_o)) = \left(\frac{F(C(a_o))}{F(a)}\right)^n = (1-\alpha)\left(\frac{F(a_o)}{F(a)}\right)^n \underset{n\to+\infty}{\longrightarrow} 0 \quad \blacksquare$$

Pour effectuer le test, il n'est pas nécessaire de calculer $C(a_o)$ mais seulement $F(t)$ et $F(a_o)$ où t est la valeur observée de $T = X_n^*$. H_0 est rejetée au seuil α si et seulement si $\boxed{F^n(t) > (1-\alpha) F^n(a_o)}$. F étant croissante la borne inférieure $C(a_o)$ du domaine de rejet de H_0 est inférieure à a_o ; elle est une fonction croissante de n et décroissante de α . Dans le cas particulier de la troncature d'une loi $\mathscr{N}(0;1)$, nous avons obtenu, par inversion de Φ , des valeurs approchées de $C(a_o)$ pour $\alpha = 0.05$, n=10 et n=50. Nous désignons par $\lambda(a_o)$ l'expression $a_o - C(a_o)$.

Tableau 4 - Borne inférieure du domaine de rejet de H_0 : "$a \leq a_o$" au seuil 0.05

a_o	n=1Ø		n=5Ø	
	$C(a_o)$	$\lambda(a_o)$	$C(a_o)$	$\lambda(a_o)$
-3	-3.ØØ16	Ø.ØØ16	-3.ØØØ3	Ø.ØØØ3
-2	-2.ØØ22	Ø.ØØ22	-2.ØØØ4	Ø.ØØØ4
-1	-1.ØØ34	Ø.ØØ34	-1.ØØØ7	Ø.ØØØ7
Ø	-Ø.ØØ64	Ø.ØØ64	-Ø.ØØ13	Ø.ØØ13
1	Ø.9824	Ø.Ø176	Ø.9964	Ø.ØØ36
2	1.9149	Ø.Ø851	1.9818	Ø.Ø182
3	2.486Ø	Ø.514Ø	2.8237	Ø.1763

Nous remarquons que λ est croissante.

Ce résultat est plus général ainsi que l'indique la proposition suivante :

PROPOSITION 13 : Si F est log-concave la fonction λ est croissante. Si F est log-convexe sur A, λ est décroissante.

Démonstration : D'après la définition de C(a), $\text{Log } F(C(a)) = \frac{1}{n}\text{Log}(1-\alpha) + \text{Log } F(a)$

En dérivant, nous obtenons : $C'(a) = \dfrac{\frac{f(a)}{F(a)}}{\frac{f(C(a))}{F(C(a))}}$.

Nous savons que $C(a) \leq a$. Supposant F log-concave, c'est à dire $\frac{f}{F}$ décroissante, nous voyons que $C'(a) \leq 1$. Par suite, $\lambda'(a) = 1-C'(a) \geq 0$ prouvant que λ est croissante. De façon analogue, si F est log-convexe sur A, $\lambda'(a) \leq 0$ ■

Remarque : Dans le cas de la troncature d'une loi exponentielle négative (F log-concave et log-convexe), nous vérifions que λ est constante :

$$F(a) = \exp(\theta a), \quad \theta > 0 \quad , \quad a \leq 0$$

$$\lambda(a) = a - C(a) = -\frac{1}{n}\,\frac{\text{Log}(1-\alpha)}{\theta} \quad .$$

Fonction puissance de ce test

Rappelons qu'il s'agit de la fonction $\beta_n : \mathbb{R} \longrightarrow [0,1]$ définie par :

$$\beta_n(a) = P_a(T \in \mathcal{R}).$$

Ici :

$$\beta_n(a) = 1 - (1-\alpha)\left(\frac{F(a_o)}{F(a)}\right)^n \quad \text{si} \quad a \geq C(a_o)$$

$$= 0 \quad \text{si } a < C(a_o).$$

Nous donnons ci-dessous, pour un exemple, les courbes représentatives de β_{10} et β_{50}.

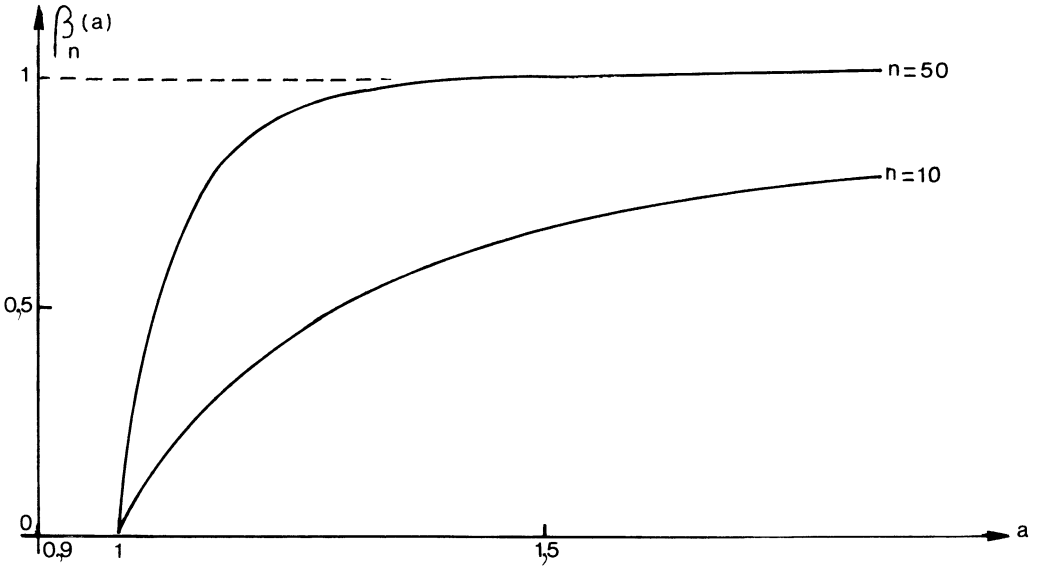

Figure 3 - Fonction puissance du test $H_0 : a \leq 1$ contre $H_1 : a > 1$ au seuil $\alpha = 0.05$ pour des échantillons de taille n=1Ø et n=5Ø d'une v.a.r. de loi $\mathcal{N}_a(0;1)$

PROPOSITION 14 : Au voisinage de a_o la puissance du test proposé est une fonction décroissante (respt. croissante) de a_o lorsque F est log-concave (respt. log-convexe sur A).

Démonstration :

$$\beta_n(a) = 1 - (1-\alpha)\left(\frac{F(a_o)}{F(a)}\right)^n$$

a pour dérivée

$$\beta_n'(a) = n(1-\alpha)\left(\frac{F(a_o)}{F(a)}\right)^n \frac{f(a)}{F(a)} \quad , \quad C(a_o) \leq a < +\infty$$

donc

$$\beta_n'(a_o) = n(1-\alpha)\frac{f(a_o)}{F(a_o)}$$

$\beta_n'(a_o)$ est une fonction décroissante de a_o si F est log-concave ($\frac{f(a_o)}{F(a_o)}$ est décroissante), croissante si F est log-convexe sur A, d'où la proposition puisque $\beta_n'(a_o) > 0$ ■

3.2 - Test $H_0 : a \geq a_o$ contre $H_1 : a < a_o$ au seuil α .

PROPOSITION 15 : Le test défini par le domaine de rejet $\mathcal{R}$ de H_0 : $\mathcal{R} =]-\infty, C(a_o)[$, $C(a_o)$ tel que $F(C(a_o)) = \alpha^{1/n} F(a_o)$ est uniformément le plus puissant pour tester H_0 contre H_1 au seuil α . C'est un test convergent.

Démonstration : Appliquant le lemme de Neyman et Pearson au test entre hypothèses simples :

$$H_0' : a = a_o \text{ contre } H_1' : a = a_1 < a_o \text{ au seuil } \alpha$$

nous obtenons pour domaine de rejet de H_0' : $\mathcal{R} =]-\infty, C(a_o)[$ où $C(a_o)$ est tel que $P_{a_o}(T < C(a_o)) = \alpha$ c'est-à-dire :

$$\frac{F^n(C(a_o))}{F^n(a_o)} = \alpha$$

donc $F(C(a_o)) = \alpha^{1/n} F(a_o)$.

Le domaine de rejet de H_0' étant indépendant de a_1 et la fonction $\beta_n : a \longrightarrow P_a(T \in \mathcal{R})$ étant décroissante (donc $\beta_n(a) \leq \alpha \ \forall a \geq a_o$) nous en déduisons le caractère U.P.P. du test de H_0 contre H_1 au seuil α . La fonction puissance de ce test est définie par :

$$\beta_n(a) = P_a(T < C(a_o)) = \left(\frac{F(C(a_o))}{F(a)}\right)^n = \alpha\left(\frac{F(a_o)}{F(a)}\right)^n \text{ si } a \geq C(a_o)$$

$$= 1 \text{ si } a < C(a_o).$$

Pour montrer la convergence du test, remarquons que quand n croît indéfiniment $C(a_o)$ tend vers a_o ; par suite : $\forall a < a_o$, $\beta_n(a) \underset{n\to\infty}{\longrightarrow} 1$, le test est bien convergent. ■

La borne supérieure du domaine de rejet de H_0 est inférieure à a_o ; c'est une fonction croissante de n et décroissante de α .

La proposition 13 est encore vraie ici pour $\lambda(a) = a - C(a)$. La valeur maximum de C est C_o telle que $F(C_o) = \alpha^{1/n}$. De même la proposition 14 reste vraie : $|\beta_n'(a_o)|$ est une fonction décroissante de a_o lorsque F est log-concave, croissante de a_o lorsque F est log-convexe sur A.

Les valeurs approchées de $C(a_o)$ sont les suivantes dans le cas de la troncature d'une loi gaussienne normalisée, pour $\alpha = 0.05$.

Tableau 5 - Borne supérieure du domaine de rejet de H_0 : "$a \geqq a_o$" au seuil 0.05

a_o	n=1Ø		n=5Ø	
	$C(a_o)$	$\lambda(a_o)$	$C(a_o)$	$\lambda(a_o)$
-3	-3.Ø9Ø1	Ø.Ø9Ø1	-3.Ø182	Ø.Ø182
-2	-2.1234	Ø.1234	-2.Ø251	Ø.Ø251
-1	-1.1871	Ø.1871	-1.Ø389	Ø.Ø389
Ø	-Ø.33Ø4	Ø.33Ø4	-Ø.Ø73Ø	Ø.Ø73Ø
1	Ø.3148	Ø.6852	-Ø.8148	Ø.1852
2	Ø.5954	1.4Ø46	1.4Ø79	Ø.5921
3	Ø.6436	2.3564	1.5598	1.44Ø2

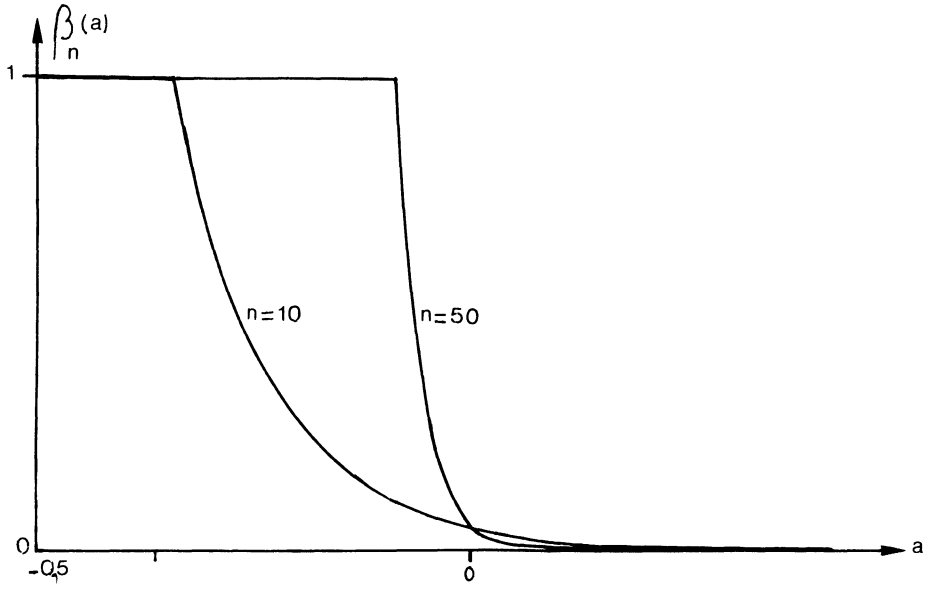

Figure 4 - Fonction puissance du test H_0 : $a \geqq 0$ contre H_1 : $a < 0$ au seuil $\alpha = 0.05$ pour des échantillons de taille n=1Ø et n=5Ø d'une v.a.r. de loi $\mathcal{N}_a(0;1)$

3.3 - Test $H_0 : a = a_o$ contre $H_1 : a \neq a_o$; seuil α .

PROPOSITION 16 : Le test défini par le domaine de rejet $\mathscr{R}$ de H_0 :

$$\mathscr{R} =]-\infty, C(a_o)[\cup]a_o, +\infty[\text{ où } C(a_o) \text{ est tel que } F(C(a_o)) = \alpha^{1/n} F(a_o)$$

est U.P.P. pour tester H_0 contre H_1. De plus il est convergent.

Démonstration : Les expressions obtenues pour les fonctions puissances des deux tests précédents -qui sont U.P.P.- exigent de vérifier que :

a) $\forall a < a_o$, $P_a(T \in \mathscr{R}) = \alpha \dfrac{F^n(a_o)}{F^n(a)}$ si $a \geq C(a_o)$

$= 1$ si $a < C(a_o)$

b) $\forall a > a_o$, $P_a(T \in \mathscr{R}) = 1 - (1-\alpha) \dfrac{F^n(a_o)}{F^n(a)}$.

Pour $a < a_o$ la propriété est triviale car $C(a_o)$ a la même expression que dans 2°.

Pour $a > a_o$ $P_a(T \in \mathscr{R}) = \dfrac{F^n(C(a_o))}{F^n(a)} + 1 - \dfrac{F^n(a_o)}{F^n(a)}$ mais $F^n(C(a_o)) = \alpha F^n(a_o)$

par hypothèse donc :

$$P_a(T \in \mathscr{R}) = 1 - (1-\alpha) \frac{F^n(a_o)}{F^n(a)} .$$

La convergence de ce test est une conséquence immédiate de l'expression de sa fonction puissance ■

Notons que la proposition 14 reste valable dans ce cas aussi : la démonstration est facile en considérant les dérivées à gauche et à droite de β_n en a_o.

Nous donnons ci-après les représentations graphiques de la fonction puissance de ce test dans le cas de lois gaussiennes (deux valeurs de n).

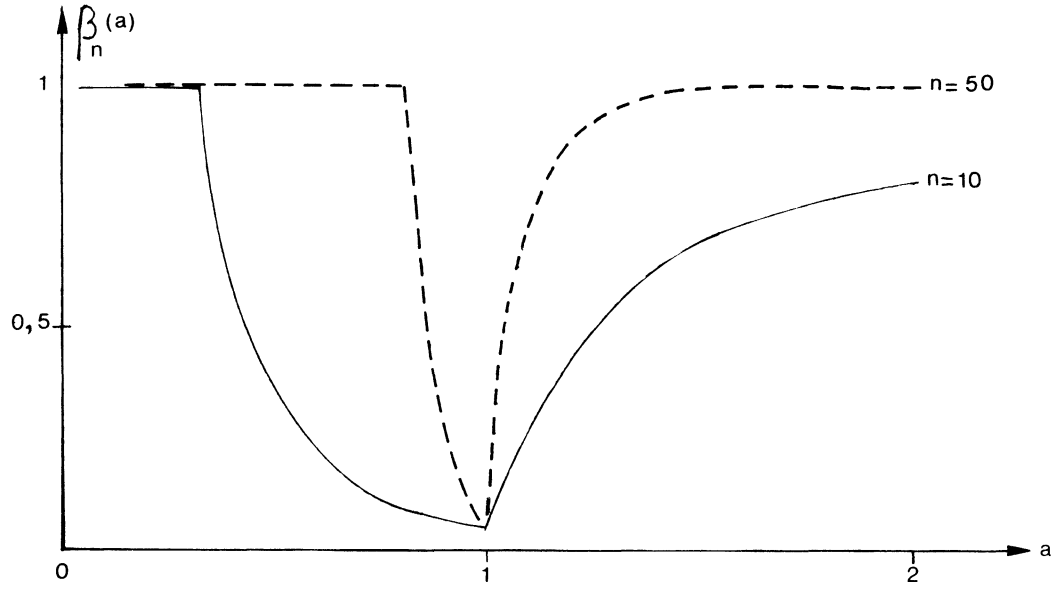

Figure 5 - Fonction puissance du test H_0 : a=1 contre H_1 : a≠1 au seuil $\alpha = 0.05$ pour des échantillons de taille n=10 et n=50 d'une v.a.r. de loi $\mathcal{N}_a(0;1)$

Conséquence : Recherche d'un intervalle de confiance pour a : $I_\alpha(a)$

Partant de la statistique exhaustive $T = X_n^*$ dont la valeur observée est t, nous pouvons construire un intervalle de confiance de a basé qur le test U.P.P. ci-dessus. Au seuil α nous prenons pour I_α l'ensemble des valeurs a_o telles que l'hypothèse H_0 : $a = a_o$ ne soit pas rejetée lorsqu'on la teste contre l'alternative H_1 : $a \neq a_o$, au seuil α , la valeur observée étant t. Donc :

$$I_\alpha = \{a/C(a) \leqq t \leqq a\}$$

où C(a) est tel que $F(C(a)) = \alpha^{1/n} F(a)$.

L'application C : a ⟶ C(a) étant croissante (α et n fixés) la borne supérieure de I_α est un élément a_1 de $\mathbb{R} \cup \{+\infty\}$ tel que $C(a_1) = t$ c'est à dire :

$$F(a_1) = \alpha^{-1/n} F(t) . \qquad (2)$$

En conséquence, ou bien $F(t) < \alpha^{1/n}$ et a_1 est un nombre réel (unique si F est strictement croissante ; sinon nous choisissons $a_1 = \inf\{a/F(a) = \alpha^{-1/n} F(t)\}$), ou bien $F(t) > \alpha^{1/n}$: (2) n'a pas de solution ; nous prenons $a_1 = +\infty$ (le seuil réel de l'intervalle de confiance est alors inférieur à α), ou bien $F(t) = \alpha^{1/n}$ nous prenons $a_1 = \inf\{a/F(a) = 1\}$, a_1 peut être fini ou infini. En résumé, pour intervalle de confiance de a, nous prenons $I_\alpha(a) = [T, a_1(T)]$ (intervalle aléatoire) où $a_1(T)$ vérifie $F(a_1(T)) = \alpha^{-1/n} F(t)$ si cette équation a une solution, sinon $a_1(T) = +\infty$.

Exemples :

a) Cas d'une loi uniforme sur $[0,1]$ tronquée en a :

$$I_\alpha(a) = [t,\ \alpha^{-1/n}\ t]$$

b) Cas d'une loi normale $\mathcal{N}(0;1)$ tronquée en a.

Supposons $\alpha = 0.05$; n=1Ø et 5Ø ; pour différentes valeurs de t nous donnons la borne supérieure approchée de $I_{0.05}(a)$ (tableau 6).

Tableau 6 - Borne supérieure de $I_{0.05}(a)$ dans le cas d'une loi $\mathcal{N}_a(0;1)$

t	n=1Ø		n=5Ø	
	$\alpha^{-1/n}\ \Phi(t)$	$a_1(t)$	$\alpha^{-1/n}\Phi(t)$	$a_1(t)$
-3	Ø.ØØ18	-2.92Ø6	Ø.ØØ14	-2.9983
-2	Ø.Ø3Ø8	-1.8692	Ø.Ø242	-1.9739
-1	Ø.2141	-Ø.7933	Ø.1685	-Ø.96Ø1
Ø	Ø.6746	Ø.4527	Ø.53Ø9	Ø.Ø775
1	1.1352	$+\infty$	Ø.8932	1.2437
2	1.3185	$+\infty$	1.Ø375	$+\infty$

Dans ce cas particulier nous constatons que la fonction :

$$\eta : \mathbb{R} \longrightarrow \bar{\mathbb{R}}$$

$$t \rightsquigarrow \eta(t) = a_1(t) - t \quad (=\text{longueur de } I_\alpha \text{ si } a_1(t) < \infty)$$

est croissante. Nous avons plus généralement la proposition suivante :

PROPOSITION 17 : Si F est log-concave η est croissante ; si F est log-convexe sur A, η est décroissante.

Démonstration : $\text{Log } F(a_1(t)) = -\frac{1}{n} \text{Log } \alpha + \text{Log } F(t)$ d'où :

$$a_1'(t)\ \frac{f(a_1(t)}{F(a_1(t))} = \frac{f(t)}{F(t)}$$

$$\eta'(t) = a_1'(t) - 1$$

$$= \frac{\frac{f(t)}{F(t)}}{\frac{f(a_1(t))}{F(a_1(t))}} - 1 .$$

3 - Test d'hypothèses portant sur deux ou plusieurs populations

Considérons k populations $\mathcal{P}_1,\ldots,\mathcal{P}_k (k \geq 2)$ et une v.a.r. de f.r. F_{a_i} sur $\mathcal{P}_i$ obtenue par troncature en a_i d'une f.r. F. Nous testons les hypothèses :

(3) $H_0 : a_1 = a_2 = \ldots = a_k = a_o$ fixé contre $H_1 : a_i \neq a_o$ pour au moins un i, i=1,..., k ; seuil α ,

ou (4) : $H_0 : a_1 = a_2 = \ldots = a_k$ (valeur commune non spécifiée) contre $H_1 : a_i \neq a_j$ pour au moins un (i,j), i≠j, i, j=1,..., k ; seuil α .

Les résultats que nous rappelons ci-dessous obtenus notamment par D.R.BARR [1], sont basés sur une propriété remarquable des lois uniformes. Soit U_i, i=1,... k,k v.a.r. de loi uniforme sur $[0,1]$, indépendantes, alors $U=-2 \text{ Log}(\prod_{i=1}^{k} U_i)$ obéit à une loi de χ^2 à 2k degrés de liberté (voir par exemple M.G. KENDALL et A. STUART [6]). Dans le but de tester (3) ou (4), prenons un n_i-échantillon $X_{i,1},\ldots, X_{i,n_i}$ de v.a.r. de loi F_{a_i} i=1,...,k ; $X^*_{i,1},\ldots, X^*_{i,n_i}$ est la statistique d'ordre correspondante. Il est alors possible de montrer que sous l'hypothèse H_0 de (3), la statistique :

$$T = -2 \text{ Log } \frac{\prod_{i=1}^{k} F^{n_i}(X^*_{i,n_i})}{F^N(a_o)} \quad , \text{ où } N = \sum_{i=1}^{k} n_i \ ,$$

obéit à une loi de χ^2 à 2k degré de liberté.

Au seuil α le domaine de rejet de H_0 associé à la statistique de test T est de la forme $\mathcal{R} =]C,+\infty[$ (C lu sur une table de χ^2 à 2k degrés de liberté). D.R.BARR a montré qu'il n'existe pas de test U.P.P. pour (3) (k ≧ 2). Le test construit ci-dessus est sans biais mais il n'est pas U.P.P. parmi les tests sans biais.

Dans le cas du test (4), en posant $X^*_N = \underset{\{i \in 1,\ldots,k\}}{\text{Max}} X^*_{i,n_i}$, la statistique :

$$T' = -2 \text{ Log } \frac{\prod_{i=1}^{k} F^{n_i}(X^*_{i,n_i})}{F^N(X^*_N)}$$

obéit à une loi de χ^2 à 2(k-1) degrés de liberté, sous l'hypothèse H_0.

Le domaine de rejet de H_0 est encore de la forme $\mathcal{R} =]C,+\infty[$, C étant obtenu par une table de χ^2 à 2(k-1) degrés de liberté.

Remarquons que la loi de T', comme celle de T, n'est pas une approximation. Par ailleurs D.R. BARR a montré que le test construit pour (4) est U.P.P. parmi les tests sans biais pour k=2 ; pour k > 2 il n'existe pas de test U.P.P. même parmi les tests sans biais.

Dans le cas k=2 l'expression de T' se simplifie un peu :

$$T' = 2\, n_1 \operatorname{Log}\left(\frac{F(X^*_{2,n_2})}{F(X^*_{1,n_1})}\right) \qquad \text{si } X^*_{1,n_1} \leqq X^*_{2,n_2}$$

$$T' = 2\, n_2 \operatorname{Log}\left(\frac{F(X^*_{1,n_1})}{F(X^*_{2,n_2})}\right) \qquad \text{si } X^*_{1,n_1} > X^*_{2,n_2}$$

Pour avoir une idée de la puissance du test de Barr par rapport à deux tests non paramétriques (de Mann-Whitney et de Kolmogorov-Smirnov) couramment utilisés pour la comparaison des f.r. de deux v.a.r. nous avons effectué des simulations.

Considérant l'hypothèse nulle H_0 : $a_1 = a_2$ (non spécifié) et l'hypothèse alternative H_1 : $a_1 \neq a_2$ (seuil 0.05), nous avons calculé la fréquence de rejet de H_0 lorsque cette hypothèse est fausse, sur 200 échantillons ; chaque échantillon est composé de 20 v.a.r. : n_1=10 de loi $\mathcal{N}_{a_1}(0;1)$ et n_2=10 de loi $\mathcal{N}_{a_2}(0;1)$; et ceci pour différents couples (a_1,a_2).

Nous avons recommencé l'expérience avec 100 échantillons de taille 100 (avec $n_1 = n_2 = 50$). Les fréquences de rejet obtenues figurent dans le tableau ci-après.

Tableau 7 - Fréquence de rejet de l'hypothèse H_0 :"$a_1 = a_2$",au seuil 0.05 dans le cas de deux lois gaussiennes $\mathcal{N}(0;1)$ tronquées en a_1 et a_2

$n_1 = n_2 = 1Ø$		Barr	Mann-Whitney	Kolmogorov-Smirnov
a_1	a_2			
Ø	-1	1	Ø.78	Ø.7Ø
	-Ø.8	1	Ø.64	Ø.43
	-Ø.6	Ø.99	Ø.34	Ø.22
	-Ø.4	Ø.73	Ø.2Ø	Ø.Ø3
	-Ø.2	Ø.17	Ø.11	Ø.Ø3
	Ø.2	Ø.11	Ø.Ø3	Ø
	Ø.4	Ø.295	Ø.11	Ø.Ø55
	Ø.6	Ø.7Ø	Ø.17	Ø.Ø8
	Ø.8	Ø.93	Ø.3Ø	Ø.12
	1	Ø.9Ø	Ø.32	Ø.12

$n_1 = n_2 = 1Ø$		Barr	Mann-Whitney	Kolmogorov-Smirnov
a_1	a_2			
1	Ø.2	Ø.77	Ø.18	Ø.Ø55
	Ø.4	Ø.26	Ø.12	Ø.Ø35
	Ø.6	Ø.14	Ø.Ø85	Ø.Ø3
	Ø.8	Ø.Ø5	Ø.Ø5	Ø.Ø1
	1.2	Ø.Ø4	Ø.Ø4	Ø.Ø2
	1.4	Ø.Ø45	Ø.Ø6	Ø.Ø15
	1.6	Ø.Ø4	Ø.Ø4	Ø
	1.8	Ø.Ø9	Ø.Ø5	Ø.Ø3
	2	Ø.1Ø	Ø.Ø9	Ø.Ø3
	3	Ø.13	Ø.Ø5	Ø.Ø25

$n_1 = n_2 = 5Ø$		Barr	Mann-Whitney	Kolmogorov-Smirnov
a_1	a_2			
Ø	-Ø.2	1	Ø.28	Ø.2Ø
	-Ø.1	Ø.87	Ø.11	Ø.Ø8
	-Ø.Ø5	Ø.2Ø	Ø.Ø9	Ø.Ø7
	Ø.Ø5	Ø.16	Ø.Ø8	Ø.Ø5
	Ø.1	Ø.78	Ø.Ø7	Ø.Ø3
	Ø.2	Ø.99	Ø.16	Ø.Ø8

Commentaire : La puissance du test de Barr est nettement supérieure à celle du test de Mann-Whitney et surtout de Kolmogorov-Smirnov lorsque la troncature de l'une au moins des deux lois est "assez forte" (par exemple en 0), et ceci pour les deux tailles d'échantillons considérées. Pour des troncatures plus faibles (a assez grand, par exemple $a_1 \geqq 1$, $a_2 \geqq 0.8$) les trois tests ont des puissances voisines, mais faibles.

La supériorité du test de Barr n'est pas surprenante puisqu'il est spécifique au problème posé.

Le test de Mann-Whitney est surtout intéressant pour des alternatives de position mais son utilisation est en fait plus générale.

Le test de Kolmogorov-Smirnov s'applique à des alternatives très générales mais a une puissance faible pour des alternatives particulières.

Ci-dessous (Figure 6) nous représentons graphiquement le résultat de nos simulations dans le cas $n_1 = n_2 = 10$; $a_1 = 0$.

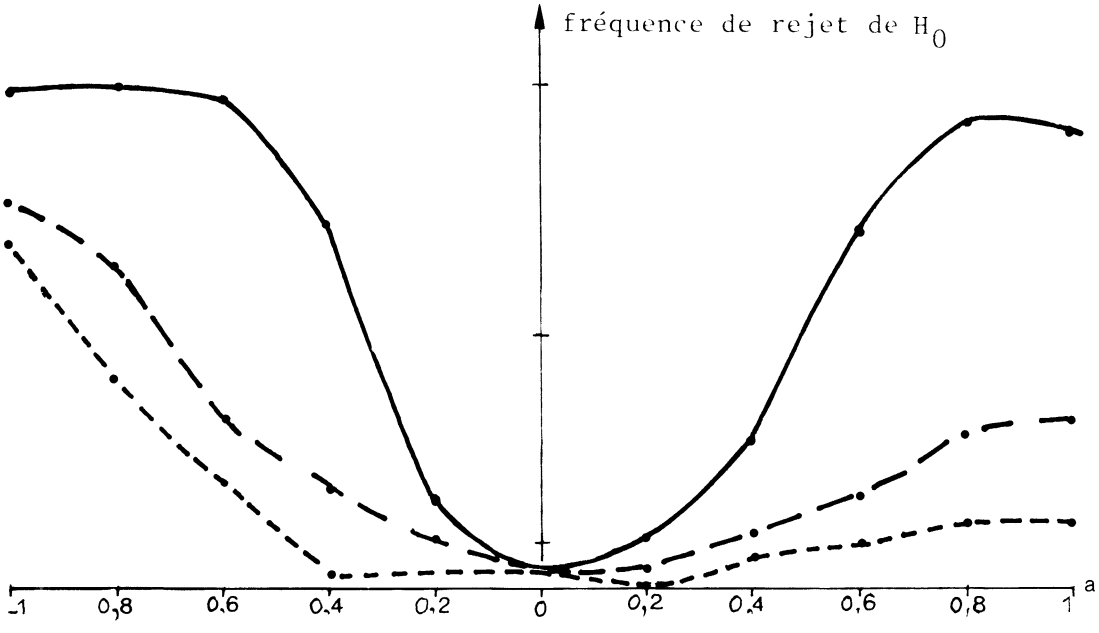

Figure 6 - Fonctions puissances des tests, obtenues par simulations

$H_0 : a_1 = a_2$; $\alpha = 0.05$

——— Test de Barr ; — — — Test de Mann-Whitney

- - - - Test de Kolmogorov-Smirnov

Or $a_1(t) \geqq t$ donc :

$$\frac{f(t)}{F(t)} \geqq \frac{f(a_1(t))}{F(a_1(t))} \quad \text{si F est log-concave et alors } \eta'(t) \geqq 0$$

$$\frac{f(t)}{F(t)} \leqq \frac{f(a_1(t))}{F(a_1(t))} \quad \text{si F est log-convexe et alors } \eta'(t) \leqq 0 \ \blacksquare$$

Pour terminer ce paragraphe, nous revenons sur les travaux de D.S. ROBSON et J.H. WHITLOCK [10] qui ont donné une valeur approchée de la limite supérieure de confiance de a au seuil α : $\bar{z}_n(\alpha) = X_n^* + \frac{(1-\alpha)}{\alpha}(X_n^* - X_{n-1}^*)$

c'est à dire :

$$P_a(X_n^* + \frac{1-\alpha}{\alpha}(X_n^* - X_{n-1}^*) > a) \simeq 1-\alpha \ .$$

Ceci permet d'obtenir des intervalles de confiance "approximatifs" de a : $[\bar{z}_n(1-\alpha_1), \bar{z}_n(\alpha-\alpha_1)]$ où α_1 est un nombre quelconque compris entre 0 et α .

La longueur d'un tel intervalle est :

$$(X_n^* - X_{n-1}^*) \ \frac{(1-\alpha)}{(1-\alpha_1)(\alpha-\alpha_1)}$$

Comme $P_a(X_n^* > a) = 0$, nous avons :

$$P_a(a \in [X_n^*, \bar{z}_n(\alpha)]) \simeq 1-\alpha \ .$$

La longueur de cet intervalle est : $\frac{1-\alpha}{\alpha}(X_n^* - X_{n-1}^*)$ qui est inférieure à

$\frac{1-\alpha}{(1-\alpha_1)(\alpha-\alpha_1)}(X_n^* - X_{n-1}^*)$ pour tout α_1, $0 < \alpha_1 < \alpha$.

Le "meilleur" (i.e. de longueur minimale) intervalle de confiance approché de a obtenu par cette méthode est donc $[X_n^*, \bar{z}_n(\alpha)]$. Ces auteurs montrent que $\bar{z}_n(\alpha)$ est une limite supérieure de confiance exacte au seuil α dans le cas de la troncature d'une loi uniforme.

Bien qu'il n'entre pas dans le cadre strict de ce travail, nous signalons l'article de B.J. WILLIAMS [12] dans lequel est étudiée la perte de puissance d'un test portant sur la moyenne d'une loi normale lorsqu'en fait il s'agit d'une loi $\mathcal{N}(m;\sigma)$ tronquée en a, pour diverses valeurs de a.

B I B L I O G R A P H I E

[1] BARR D.R., "On testing the equality of uniform and related distribution", J.Amer. Statist. Ass. 61 (1966), 856-864.

[2] CHARALAMBIDES C., "Minimum variance unbiased estimation for a class of left-truncated discrete distributions", Sankhya A, 36(1974), 397-418.

[3] COHEN A.C., WOODWARD J., "Tables of Pearson - Lee - Fisher functions of singly truncated normal distributions", Biometrics, 9 (1953), 489-497.

[4] FERGUSON T.S., Mathematical Statistics. A decision theoretic approach, Academic Press, 1967.

[5] HALD A., Statistical theory with engineering applications, New York, Wiley, 1967.

[6] KENDALL M.G., STUART A., The advanced theory of statistics, Tomes 1 et 2, Londres, Griffin, 1967.

[7] MAILHOT L., Etude des lois de probabilités réelles tronquées à droite et applications statistiques, Thèse de 3ème cycle, Clermont-Fd, Université Clermont II.

[8] PLACKETT R.L., "The truncated Poisson distribution", Biometrics, 9 (1953), 485-488.

[9] PREKOPA A., "On logarithmic concave measures and functions", Acta Sci. Math., 34 (1973), 335-343.

[10] ROBSON D.S., WHITLOCK J.H., "Estimation of a truncation point", Biometrika, 51 (1964), 33-39.

[11] TATE R.F., "Unbiased estimation : functions of location and scale parameters", Ann. Math. Statist., 30 (1959), 341-366.

[12] WILLIAMS B., "The effect of truncation on tests of hypotheses for normal populations", Ann. Math. Statist., 36 (1965), 1504-1510.