

Nice 7 au 12 mai 1973

ESTIMATION DE PROBABILITES FAIBLES

PAR SIMULATIONS.

O. MACCHI et J.P. CALISTE

Laboratoire d'Etude des Phénomènes Aléatoires

Bât. 210 - Université de Paris - Sud, Centre d'Orsay

91405 ORSAY

RESUME

La probabilité de fausse alarme (F.A.) d'un système de détection constitue une de ses plus importantes caractéristiques. Lorsqu'elle est impossible à évaluer analytiquement, on a souvent recours à des simulations sur ordinateur. Dans cet exposé on rappelle la précision atteinte par les simulations en fonction du nombre d'essais. Dans la pratique pour des F.A. petites et une précision correcte, le temps de calcul est souvent prohibitif.

On étudie une méthode de simulation rapide qui, grâce à une réduction de variance, diminue notablement le nombre d'essais. La méthode est illustrée par le calcul de la fonction d'erreur, et celui de la F.A. d'un détecteur de type non paramétrique.

SUMMARY

The false alarm probability (F.A.) of any detection system is one of its most important characteristics. If its evaluation is intractable by analytical means, one commonly uses computer simulations. In this paper we recall the accuracy of this technique in terms of the number of simulations. In practice for small F.A. and a correct accuracy, the computation time is often prohibitive.

We study a fast simulation method which decreases the variance and therefore the number of trials. We illustrate it by computing the standardized error function, and the F.A. of a non parametric detector.

ESTIMATION DE PROBABILITES FAIBLES PAR SIMULATIONS

O. MACCHI et J.P. CALISTE

I - INTRODUCTION ET EXEMPLES

Soit X un processus aléatoire (p.a.) c'est-à-dire une application mesurable d'un espace de probabilité $(\Omega, \mathcal{A}, \Pi)$ dans un espace $(\mathcal{X}, \mathcal{B})$. Soit \mathcal{C} une transformation mesurable de $(\mathcal{X}, \mathcal{B})$ dans un espace $(\mathcal{Y}, \mathcal{C})$. La loi de probabilité P de X étant supposée connue, celle du processus aléatoire $Y = \mathcal{C}(X)$ est entièrement déterminée par la fonction \mathcal{C} . Cependant le calcul analytique à partir de P et de \mathcal{C} est souvent inextricable.

Tout en évitant ce calcul analytique, on cherche à évaluer la probabilité α d'un évènement C de \mathcal{C} :

$$(1) \quad \alpha = \Pr[I_C(Y) = 1] = E(I_C(Y)),$$

où $I_C(Y)$ est la variable aléatoire (v.a.) indicatrice de l'évènement C.

Le p.a. X peut être une tranche temporelle $X(t)$, $t \in [0, T]$, d'une fonction aléatoire (f.a.), ou d'un processus ponctuel, une v.a. $X = (X^1, \dots, X^k)$ k-dimensionnelle etc ... Le traitement X est souvent une opération non linéaire avec mémoire par exemple, dans le cas d'une f.a. séparable $X(t)$

$$(2) \quad Y = \sup_{t \in [0, T]} X(t)$$

ou bien

$$(3) \quad Y = (Y^1, \dots, Y^\ell),$$

$$(4) \quad Y^i = \int_0^T f_i(X(t)) dt \quad i = 1, \dots, \ell,$$

$f_i(\cdot)$ étant une opération instantanée.

Exemples : Ce problème d'évaluation de probabilité peut surgir dans de nombreux contextes de la physique ou des statistiques (fiabilité, files d'attente, etc ...). Il intervient aussi dans tous les problèmes de communication où l'information transmise est discrète. Le récepteur doit associer à toute épreuve observée du p.a. X, un et un seul message \hat{a} parmi les ℓ messages possibles a_1, \dots, a_ℓ . Pour cela, il opère sur X un traitement déterministe, dont le résultat Y est une v.a. (Y^1, \dots, Y^ℓ) ℓ -dimensionnelle ; le message \hat{a} choisi est alors a_j si

$$(5) \quad H(Y^j - Y^k) = 1 \quad k = 1, \dots, \ell,$$

où H(y) est la fonction de Heaviside

$$(6) \quad H(y) = \begin{cases} 1 & \text{si } y \geq 0, \\ 0 & \text{si } y < 0. \end{cases}$$



(Le plus souvent (5) définit $\hat{a} = a_j$ de manière univoque sauf sur un sous-ensemble de \mathcal{X} de mesure nulle pour P). L'évènement C_j de la tribu \mathcal{C} associé à la décision j a donc la forme

$$(7) \quad \prod_{k=1}^l (Y^j \geq Y^k) :$$

$$I_{C_j}(Y) = \prod_{k=1}^l H(Y^j - Y^k),$$

et il est essentiel de pouvoir calculer les probabilités α_j de chaque décision car elles permettent de juger la qualité du récepteur.

Le fonctionnement d'un détecteur par exemple radar ou sonar illustre le cas où $l = 2$ (2 décisions possibles) et peut être schématisé par la figure 1,

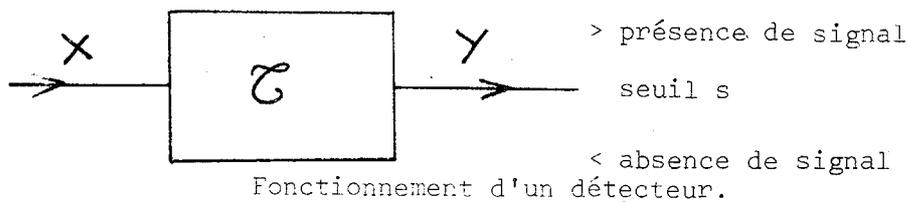


Figure 1.

où la sortie Y de l'opération \mathcal{C} est une v.a. unidimensionnelle ($\mathcal{Y} = \mathbb{R}$). En l'absence de signal (hypothèse H_0), le p.a. X est constitué de bruit, de loi P_0 sur $(\mathcal{X}, \mathcal{B})$. Lorsqu'un signal utile se mélange au bruit (hypothèse H_1), la loi de X devient P_1 .

Le traitement \mathcal{C} est choisi de manière à "faire ressortir" le signal, lorsqu'il est présent, dans le bruit de fond, de sorte que les grandes valeurs de Y ($Y \geq s$) sont associées à la décision "Signal présent" et déclenchent une alarme. L'évènement C dont on veut évaluer la probabilité est constitué par cette alarme :

$$(8) \quad I_C(Y) = H(Y - s),$$

et sa probabilité dépend du seuil s . En l'absence de signal ("fausse alarme") elle vaut

$$(9) \quad \alpha(s) = \int_{\mathcal{X}} H(\mathcal{C}(X) - s) dP_0(X),$$

et reste toujours très faible, de l'ordre de 10^{-3} ou 10^{-6} .

En présence de signal ("probabilité de détection")



elle vaut

$$(10) \quad P_d(s) = \int_{\mathcal{X}} H(\mathcal{C}(X) - s) dP_1(X),$$

et n'est jamais très petite, variant, par exemple entre 0,2 et 1.

Les courbes de coordonnées paramétriques $\alpha(s)$, $P_d(s)$ caractérisent les performances du détecteur \mathcal{C} , et même pour des lois P_0 et P_1 connues, leur détermination analytique n'est généralement pas possible. Leur tracé implique donc l'estimation des deux probabilités $\alpha(s)$ et $P_d(s)$ pour chaque seuil, la première étant la plus délicate à cause de l'ordre de grandeur.

II - LA SIMULATION DIRECTE ET SES LIMITES.

A. Principe

Lorsque le calcul de α s'avère inaccessible, on a maintenant recours, de plus en plus fréquemment, à une simulation sur ordinateur dont le principe est tout simplement d'évaluer la moyenne d'ensemble (1) de la v.a. $I_C(Y)$ en répétant des expériences indépendantes Y_i de Y . L'estimateur choisi est donc

$$(11) \quad \hat{\alpha} = \frac{1}{N} \sum_{i=1}^N I_C(Y_i),$$

qui est évidemment sans biais.

Par une technique de Monte-Carlo on génère une suite pseudo-aléatoire de N points de l'espace \mathcal{X} , qui peuvent être considérés comme N tirages indépendants X_i , $i = 1, \dots, N$ du p.a. X , si l'algorithme de Monte-Carlo a été choisi convenablement en fonction de la loi de probabilité P que l'on veut simuler [1] - [4]. La transformation - déterministe - étant en général aisée à programmer, on obtient N tirages Y_i indépendants du p.a. Y pour lesquels le test de la réalisation de l'évènement C fournit la valeur de la v.a. $I_C(Y_i)$.

On sait [5] que (11) est un estimateur efficace de α : parmi les estimateurs sans biais de α basés sur N tirages indépendants du processus X de loi P , il minimise la variance.

Dans la suite de ce paragraphe, nous admettons que les simulations de Monte-Carlo des p.a. désirés sont statistiquement parfaites, et en nous fixant une limite pour l'erreur sur α , nous évaluons le nombre N de simulations nécessaires et le temps de calcul correspondant.



B. Nombre de simulations

L'erreur relative δ d'un estimateur sans biais de moyenne α et de variance u peut être définie comme le rapport

$$(12) \quad \delta = \frac{\sqrt{u}}{\alpha} .$$

Pour qu'avec certitude $(1-\eta)$, le résultat $\hat{\alpha}$ de la simulation soit vrai avec une précision relative au moins égale à k

$$(13) \quad \Pr \left[\frac{|\hat{\alpha} - \alpha|}{\alpha} > k \right] \leq \eta$$

il suffit, d'après l'inégalité de Tchébycheff

$$(14) \quad \Pr \left[\frac{|\hat{\alpha} - \alpha|}{\alpha} > k \right] \leq \frac{E(|\hat{\alpha} - \alpha|^2)}{k^2 \alpha^2} = \frac{\delta^2}{k^2} ,$$

que l'erreur relative de l'estimateur $\hat{\alpha}$ vérifie

$$(15) \quad \delta^2 \leq k^2 \eta .$$

Si comme cela se produira au paragraphe III, l'estimateur est un "moyenneur" sans être nécessairement de la forme (11)

$$(16) \quad \hat{\alpha} = \frac{1}{N} \sum_{i=1}^N Z_i ,$$

les v.a. Z_i étant indépendantes, équidistribuées, de moyenne α (grandeur cherchée) et de variance σ_Z^2 , l'erreur relative est liée à N par

$$(17) \quad \delta^2 = \frac{1}{N} \cdot \frac{\sigma_Z^2}{\alpha^2} .$$

D'après (15) le nombre N doit satisfaire à

$$(18) \quad N > \frac{1}{\eta} \cdot \frac{1}{k^2} \cdot \frac{\sigma_Z^2}{\alpha^2} .$$

C'est le cas de l'estimateur (11) associé à une simulation directe, avec $\sigma_Z^2 = \alpha - \alpha^2$. D'où un minorant du nombre de simulations nécessaires

$$(19) \quad N \geq \frac{1}{\eta} \cdot \frac{1}{k^2} \cdot \left(\frac{1}{\alpha} - 1 \right)$$

Lorsque le nombre N est assez grand pour qu'on puisse considérer le moyenneur (16) comme une v.a. gaussienne, on peut améliorer la borne sur N en utilisant l'équation

$$\Pr \left[\frac{|\hat{\alpha} - \alpha|}{\alpha} > k \right] = 2 \cdot Q \left(\frac{k}{\delta} \right) ,$$

cù $Q(x)$ est la fonction d'erreur définie par

$$Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^{\infty} e^{-\frac{u^2}{2}} du.$$

De la sorte, pour obtenir l'intervalle de confiance (13), il faut que δ vérifie

$$(20) \quad \frac{1}{\delta^2} \geq [Q^{-1}(\frac{\eta}{2})]^2 \cdot \frac{1}{k^2}.$$

D'après (17) cela implique pour N

$$(21) \quad N \geq [Q^{-1}(\frac{\eta}{2})]^2 \cdot \frac{1}{k^2} \cdot \frac{\sigma^2 Z}{\alpha^2}.$$

Dans le cas présent où $\hat{\alpha}$ est donné par (11), on trouve

$$(22) \quad N \geq [Q^{-1}(\frac{\eta}{2})]^2 \cdot \frac{1}{k^2} \cdot (\frac{1}{\alpha} - 1).$$

Les équations (18) et (21) d'une part, (19) et (22) d'autre part donnent, à un facteur près, le même minorant pour N . Ce facteur ne dépend que de la confiance $(1-\eta)$ exigée pour la validité du résultat. Pour les exemples simulés nous avons fixé cette confiance à la valeur standard de 0,95. Il en résulte que

$$(21\text{bis}) \quad \frac{1}{\eta} = 20, \quad [Q^{-1}(\frac{\eta}{2})]^2 = 3,84,$$

et qu'on gagne un facteur 5,2 sur N , lorsqu'on peut appliquer le théorème de la limite centrale au moyennneur. Dans la pratique des simulations, c'est toujours le cas et nous retiendrons seulement les bornes (21) et (22).

Donnons deux exemples numériques de la borne (22), permettant de préciser le nombre de simulations requises.

a) La probabilité à estimer est $\alpha = \frac{1}{2}$; on désire la précision au 1/100è, avec certitude 0,99; on trouve comme minimum pour le nombre de simulations $N_0 = 66\ 200$, ce qui est un chiffre étonnamment élevé.

b) La probabilité à estimer est $\alpha = 10^{-3}$; comme elle est très faible, seul compte son ordre de grandeur et on se contente de la précision relative 2/10; on veut une confiance 0,95 dans le résultat. Le calcul montre qu'alors le nombre minimum de simulations est $N_0 = 95\ 900$.

On voit qu'il est hors de question par une simulation directe, d'évaluer de manière satisfaisante une probabilité de



l'ordre de 10^{-3} à l'aide de 10^4 essais. Pourtant, dans la littérature, on voit fréquemment employer cette méthode dans des conditions comparables. Il y a alors lieu de se demander si la précision avec laquelle les résultats sont donnés par les auteurs n'est pas illusoire.

C. Temps de calcul

Le nombre important de simulations requises constitue le facteur qui limite l'emploi de la méthode par l'intermédiaire du temps de calcul. Ainsi pour $N_0 = 10^5$, celui-ci devient prohibitif dès que le temps nécessaire pour générer sur ordinateur une épreuve X_i , la traiter par \mathcal{C} , et la trier, est plus grand que $10 \mu s$ ou $100 \mu s$, ce qui est pratiquement toujours le cas.

Nous en donnons maintenant un exemple où le temps de calcul est évalué pour une simulation sur UNIVAC 1108.

Exemple : Le p.a. X est constitué par K échantillons successifs X_1, \dots, X_K d'un bruit gaussien centré, à spectre plat dans une bande de fréquence de largeur B . Le pas d'échantillonnage est pris égal à $\frac{1}{2B}$, pour que les échantillons successifs soient indépendants.

La transformation \mathcal{C} est illustrée par la figure 2. Elle comporte un filtrage numérique \mathcal{R} de la forme :

$$(23) \quad X'_i = \sum_{j=1}^C X_{i-j} R_j \quad i = C + 1, \dots, k,$$

puis l'opération qui recherche l'échantillon le plus grand :

$$(24) \quad Y = \sup_{i=C+1, \dots, k} X'_i.$$

L'évènement C (dit alarme) correspond à $Y \geq s$ (cf. (8)).

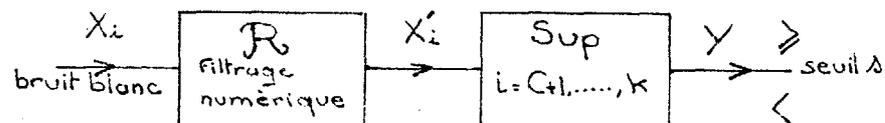


FIGURE 2. Détecteur basé sur le maximum de sortie du filtre adapté

On peut évaluer à 10 μ s le temps nécessaire pour effectuer le produit $X_{i-j} R_j$, l'additionner au résultat précédent, et à 15 μ s le temps nécessaire au test de fin de boucle $j \leq C$. Ainsi le calcul de (23) nécessite

$$(25) \quad t_R = (K-C) \cdot C \cdot 25 \mu s.$$

Chaque variable pseudo-gaussienne X_i est la somme de 12 variables pseudo-aléatoires E_j^i , indépendantes et uniformément réparties sur l'intervalle $[0, 1]$:

$$(26) \quad X_i = \sum_{j=1}^{12} E_j^i.$$

La suite pseudo-aléatoire E_j^i est engendrée par la relation de récurrence

$$(27) \quad E_j^i \equiv a E_{j-1}^i \pmod{m},$$

de sorte que, par un calcul analogue au précédent, la génération de chaque X_i nécessite $12 \times 25 \mu s = 300 \mu s$. Enfin les $(K-C)$ tests successifs de maximum effectués par l'ordinateur pour mettre en oeuvre (24) nécessitent $(K-C) 15 \mu s$. D'où le temps $t = [300 K + 25 C (K-C) + 15 (K-C)] \mu s$ d'ordinateur pour une seule simulation, et pour N simulations du système

$$(28) \quad T = N(300 K + 25 C (K-C) + 15 (K-C)) \mu s.$$

Cet exemple vise à simuler sur ordinateur le fonctionnement d'un détecteur sonar ; c'est pourquoi la durée des signaux traités et leurs largeurs de bande imposent un nombre assez grand d'échantillons à traiter simultanément : $K = 700$. La longueur du filtre est de $C = 70$ échantillons. Cherchant à estimer une probabilité de fausse alarme de $\alpha = 10^{-3}$, nous devons choisir $N = 10^5$, d'après le paragraphe II-B.b). Ces chiffres donnent à (28) la valeur énorme

$$(29) \quad T = 1,32 \cdot 10^5 s = 36 \text{ heures } 40'.$$



Ainsi, malgré la simplicité du détecteur, il est exclu d'envisager de mesurer ses performances à l'aide d'une simulation directe, car le temps de calcul est rédhibitoire.

III - TECHNIQUE DE SIMULATION RAPIDE

A. Principe

Tout en conservant le principe de la simulation qui est d'obtenir l'espérance (1) en moyennant le résultat d'expériences indépendantes et identiques (cf. formule (11)), on songe à changer de v.a. observée. Si Z est une v.a. d'espérance α , de variance plus faible que $I_C(Y)$, le moyeneur (16) constitue un estimateur sans biais de α , meilleur que (11), que l'on calculera sur ordinateur pourvu que l'on puisse simuler la variable Z . Parce que (11) est un estimateur efficace de α pour l'observation de N tirages indépendants du processus X , cette réduction de variance n'est possible qu'en changeant la loi de probabilité du p.a. traité.

D'après les formules (18) et (21) la réduction du nombre d'essais assurant le même intervalle de confiance est proportionnelle à la réduction de variance définie par

$$(30) \quad \lambda = \sigma_Z^2 / \text{var.} (I_C(Y)).$$

Ainsi, dans l'exemple ci-dessus, si l'on accepte 10' de temps d'ordinateur, il faut pouvoir simuler une v.a. Z de moyenne α , de variance 220 fois plus petite que celle de $I_C(Y)$.

B. Réduction de variance [6] - [8].

Pour réduire la variance, nous adoptons une méthode qui, à notre connaissance, a été introduite par Hansen [7], [8] pour l'appliquer à des problèmes de communications : on concentre la distribution des échantillons X_i vers les valeurs pour lesquelles se produit effectivement l'évènement C après la transformation \mathcal{C} . Autrement dit, introduisant le sous-ensemble

$$(31) \quad B = \mathcal{C}^{-1}(C)$$

de la tribu \mathcal{B} , on cherche un autre p.a., soit U , sur l'espace probabilisable $(\mathcal{X}, \mathcal{B})$ dont la loi de probabilité Q soit plus concentrée sur B que la loi initiale P . Supposons que l'on sache simuler U sur ordinateur, comme on sait le faire pour X . Pour ne pas biaiser le résultat de la simulation, on ne choisira pas comme estimateur

$$\hat{\alpha} = \frac{1}{N} \sum_{i=1}^N I_C(\mathcal{C}(U_i)),$$

mais, supposant la loi Q équivalente à la loi P , on posera

$$(32) \quad \hat{\alpha} = \frac{1}{N} \sum_{i=1}^N \frac{dP}{dQ}(U_i) I_C(\mathcal{C}(U_i)),$$

où $\frac{dP}{dQ}$ est la dérivée de Radon-Nicodým de P relativement à Q . Cela revient à pondérer chaque apparition de l'évènement C dans la proportion inverse de celle $(\frac{dQ}{dP}(u))$ où l'on a forcé la fréquence d'apparition du point u correspondant de \mathcal{X} , en remplaçant la loi P par la loi Q [9].

La v.a. Z dont on cherche la moyenne par simulation est donc

$$(33) \quad Z = \frac{dP}{dQ}(U) \cdot I_C(\mathcal{C}(U)),$$

où U est le p.a. $(\mathcal{X}, \mathcal{B}, Q)$. D'après le théorème de Radon-Nicodým $E(Z) = \alpha$, et l'estimateur (16) ou (32) est sans biais. La réduction de variance vaut

$$(34) \quad \lambda = (\alpha - \alpha^2)^{-1} \left[\int_B \left[\frac{dP}{dQ}(u) \right]^2 dQ(u) - \alpha^2 \right].$$

Dans le cas qui nous intéresse où $\alpha \ll 1$, on peut écrire

$$(35) \quad \lambda < \alpha^{-1} \int_B \left[\frac{dP}{dQ}(u) \right]^2 dQ(u) \leq \lambda_0,$$

où λ_0 est la borne inférieure essentielle de $\frac{dP}{dQ}(u)$ sur B , relativement à la mesure Q

$$(36) \quad \lambda_0 = \sup_{u \in B} \text{ess.} \frac{dP}{dQ}(u) = \inf_{\mu \in \mathbb{R}} \mu \text{ t.q. } \frac{dP}{dQ}(u) \leq \mu \text{ } Q_{pp} \text{ sur } B.$$



L'inégalité (32) traduit quantitativement cette idée intuitive : la réduction de variance est d'autant meilleure que la loi Q est plus concentrée sur B que la loi P .

Dans la suite, convenons d'appeler "simulation rapide" la méthode qui consiste à simuler sur ordinateur l'estimateur (32).

C. Choix de la loi "distorbue" Q

Le choix de Q sera bon si la valeur $\lambda_0(32)$, correspondante est très faible. Or celle-ci dépend pour une grande part de l'ensemble B . Pour bien choisir Q il faut donc avoir une idée correcte de la géométrie de B . Cette géométrie est unidimensionnelle liée au traitement \mathcal{E} et d'autant plus complexe que le p.a. X a un plus grand nombre de degrés de liberté. C'est ainsi qu'un bon choix de Q doit permettre d'évaluer la probabilité α pour une grande variété de lois P à l'entrée du traitement. Pour optimiser le choix de la loi distorbue Q , il faut minimiser la réduction de variance λ . La loi optimum - si elle existe - dépend évidemment de P . C'est ainsi que pour

$$(37) \quad \left\{ \begin{array}{l} Q = \frac{1}{\alpha} P \quad \text{sur } B \\ Q = 0 \quad \text{sur } B^c \end{array} \right.$$

il apparaît sur (34) que l'estimateur (32) est parfait puisque de variance nulle. Evidemment dans la pratique il est impossible de choisir Q suivant (37), puisque cela implique la connaissance de α , qui est l'inconnue du problème. Cependant on peut remarquer qu'on aura intérêt à choisir Q "à peu près" proportionnelle à P sur l'ensemble B pour réduire la variance.

Dans la suite nous ne nous intéressons pas à l'optimisation du choix de Q , pour une loi P donnée.

Il nous importe seulement de trouver une loi Q qui assure une précision relative k avec la confiance $1-\eta$ réduite, pour l'intervalle des valeurs de α que l'on considère, à l'aide d'un nombre de simulations N acceptable (défini auparavant, et géné-

ralement limité par le temps de calcul que l'on peut se permettre).

IV - EXEMPLE SIMPLE

Pour tester l'efficacité de la technique de simulation rapide, nous l'avons utilisée pour évaluer la fonction d'erreur

$$(38) \quad \alpha(s) = \int_s^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

pour les grandes valeurs de s . Pour cela, on choisit pour X une v.a. unidimensionnelle de loi normale $N(0, 1)$, de densité de probabilité

$$(39) \quad p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Le traitement \mathcal{C} est l'identité, et l'évènement C a pour indicateur $I_C = H(X - s)$ (cf. (8)). Evidemment l'ensemble B correspondant est l'intervalle (s, ∞) .

Comme loi distordue, nous avons choisi successivement une loi normale décalée au seuil s , $N(s, 1)$, et une loi répartie sur les plus petites valeurs de B , par exemple la loi $\text{Rect}[s, 2s]$ équirépartie sur l'intervalle $[s, 2s]$.

Ainsi qu'on pouvait s'y attendre, le résultat est obtenu avec une moins bonne précision dans le second cas que dans le premier. En effet, le rapport $\frac{\text{Rect}(s, 2s)}{N(0, 1)}$ varie beaucoup plus vite sur B que le rapport $\frac{N(s, 1)}{N(0, 1)}$. Disons seulement que pour la loi distordue $\text{Rect}(s, 2s)$, en faisant $N = 5000$ simulations, et en utilisant l'estimateur (32) nous avons trouvé les valeurs de $\alpha(s)$ avec une erreur relative toujours inférieure à $(1/20)$ et pratiquement constante, pour $s = 2.0, 2.1, \dots, 5.9, 6$, ce qui fait varier $\alpha(s)$ de $2,27 \cdot 10^{-2}$ à $1,01 \cdot 10^{-9}$.

Lorsque $Q = N(s, 1)$, la densité de probabilité de la v.a. U utilisée pour la simulation rapide est

$$(40) \quad q(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(u-s)^2}{2}}$$



de sorte que, d'après (33),

$$(41) \quad Z = e^{-Us} + \frac{s^2}{2} H(U - s), \quad E(Z) = \alpha(s),$$

et que

$$(42) \quad \sigma_Z^2 = e^{s^2} \alpha(2s) - \alpha^2(s).$$

On a choisi de limiter à $N = 1000$ le nombre de tirages de U . Cette valeur permet de considérer que $\hat{\alpha}$ est gaussienne. Ainsi pour une confiance 0,95, (21) montre que le résultat est vrai à la précision relative k si

$$(43) \quad N \geq \frac{3,84}{k^2} \left(e^{s^2} \frac{\alpha(2s)}{\alpha(s)} - 1 \right).$$

La fonction $f(s) = e^{s^2} \frac{\alpha(2s)}{\alpha^2(s)} - 1$ est représentée graphiquement

sur la figure 3. Elle admet la droite $y(s) = \sqrt{\frac{\pi}{2}} s - 1$ comme asymptote. L'équation (43) montre que pour $N = 1000$, si l'on exige une précision $k = 20\%$, les résultats de la simulation ne seront valables que pour les valeurs $s \leq s_0$ avec

$$f(s_0) = \frac{1000 \times 0,04}{3,84} = 10,4. \text{ D'après la figure 3, } s_0 = 9.$$

C'est pourquoi nous n'avons simulé $\alpha(s)$ que pour $s \leq 9$. Le calcul a été fait pour les 8 valeurs $s = 2, 3, \dots, 9$, et le résultat était correct avec une erreur relative toujours inférieure à 12,5% (l'erreur augmente avec s). Le temps d'ordinateur théorique, évalué comme au paragraphe II-C, est de 2,6 s d'Univac 1108 pour la détermination de ces 8 valeurs. Le temps réel fut de 2 s, ce qui donne un temps de 7 s déclaré sur le listing puisqu'il s'ajoute toujours 5 s de temps d'entrée-sortie au temps de calcul.

C'est ainsi qu'avec 0,3 s d'Univac 1108, on évalue par simulation à l'aide de 1000 épreuves une probabilité $\alpha(9) = 1,14 \cdot 10^{-19}$, avec une erreur relative de 12,4%.

Etant donné le coût très faible de ce calcul, une deuxième série de simulations a été faite en prenant $N = 5000$, et s variable de 2 à 9 par pas de 0,1. Les résultats sont donnés sur la figure 4 ; l'erreur relative a toujours été inférieure à 9/100. Cette erreur effective est à comparer avec la précision k que l'on peut espérer atteindre avec une certitude 0,95 et qui découle de (43) : $k = 9\%$ pour $N = 5000$ et $s \leq 9,1$.

Autrement dit le résultat de la simulation a toujours atteint la précision requise par l'intervalle de confiance.

V - SIMULATION DU TEST DE SPEARMAN-RHO.

Ayant testé l'efficacité et la précision d'une simulation rapide dans un cas où le résultat est déjà connu, et avant d'appliquer la méthode à des exemples entièrement originaux, nous avons voulu essayer une simulation rapide sur le même exemple que Hansen [7], le test de Spearman-Rho, afin de confronter nos résultats aux siens, en précisant les intervalles de confiance associés, ce qu'il ne faisait pas.

Le test de Spearman-Rho est utilisé en statistiques pour détecter la présence d'une tendance à l'augmentation dans les K éléments consécutifs x^1, \dots, x^K d'une permutation x des entiers $1, \dots, K$.

Ce test peut être utilisé en radar. On traite séparément l'onde reçue au cours de chaque période $[(i-1)T, iT]$ de récurrence d'émission, par exemple à l'aide du filtre adapté, de manière à déterminer un nombre y^i caractéristique de la i ème tranche d'observation. On veut prendre une décision globale de présence ou d'absence d'un écho durant toute la période KT , à l'aide des K observations consécutives traitées $y^i, i = 1, \dots, K$. En l'absence de cet écho les v.a. Y^1, \dots, Y^K correspondant aux épreuves y^1, \dots, y^K sont indépendantes (ceci parce que la durée de corrélation des bruits filtrés est bien plus faible que T) et de même loi. De la sorte, si X^i est le rang de Y^i dans l'ensemble



des valeurs Y^1, \dots, Y^K classées par ordre croissant, la permutation $X = (X^1, \dots, X^K)$ est une v.a. de loi équadistribuée sur l'ensemble \mathcal{X} des $K!$ permutations possibles. Remarquons au passage que la loi P ne dépend pas de la forme de la loi des échantillons de bruit Y^1, \dots, Y^K . Il n'en est pas de même en présence d'un écho, car alors les v.a. Y^1, \dots, Y^K ne sont plus de même loi. En particulier, leurs moyennes S_1, \dots, S_K sont distinctes et peuvent par un réarrangement convenable être supposées croissantes avec K . Dans ces conditions, le détecteur pourra raisonnablement fonder sa décision de présence du signal, sur le degré de tendance à l'augmentation présenté par la permutation X .

C'est ce degré que mesure le test de Spearman-Rho

$$(44) \quad Y = \mathcal{C}(X) = \sum_{i=1}^K i X^i,$$

et la décision d'alarme est choisie si

$$(45) \quad I_c(Y) = H(Y - s) = 1.$$

D'après une remarque antérieure, la loi de probabilité du test $\mathcal{C}(X)$ ne dépend pas de la loi du bruit (test dit "non paramétrique") de sorte que la probabilité de fausse alarme $\alpha(s)$ pour le seuil s n'en dépend pas non plus. Naturellement il est essentiel d'évaluer la fonction $\alpha(s)$.

Puisque P est équirépartie, le calcul direct de $\alpha(s)$ exige d'engendrer les $K!$ permutations, évaluer (44) pour chacune et dénombrer celles qui vérifient $Y \geq s$. Cela n'est envisageable que pour les petites valeurs de K , par exemple $K < 8$ ($8! = 40.320$ rend déjà cette opération impraticable). Dans la pratique K est plus grand et on doit adopter d'autres techniques par exemple approcher la loi de Y à l'aide d'une série de Gram-Charlier [10]. Cette méthode ne s'applique guère aux très petites valeurs des probabilités, c'est pourquoi la simulation rapide est bien indiquée pour ce problème.

L'espace \mathcal{X} a $K!$ éléments $x_1, \dots, x_{K!}$. La tribu \mathcal{B} est

celle de toutes les parties de \mathcal{X} . Pour faciliter la génération sur ordinateur de permutations pseudo-aléatoires, on représente une permutation $x = (x^1, \dots, x^K)$ à l'aide du vecteur p des rangs relatifs $r^1, \dots, r^j, \dots, r^K$: r^j est le rang qu'occupe l'entier $(K - j + 1)$ dans les $(K - j + 1)$ éléments qui restent à la permutation x après extraction des $(j - 1)$ plus grandes valeurs. Ainsi r^1 est la place de 16, r^2 la place de 15 dans les 15 éléments restants etc ..., et il est très aisé à partir du vecteur r de reconstruire la permutation x .

On voit facilement que la permutation aléatoire X est équidistribuée dans \mathcal{X} si et seulement si le vecteur aléatoire R des rangs relatifs a ses composantes R^j indépendantes, chacune d'elles étant équidistribuée sur l'ensemble $\{1, 2, \dots, K - j + 1\}$. Pour générer une loi distordue Q plus concentrée dans le domaine B des permutations croissantes, on utilise cette dernière représentation : on gardera l'indépendance des rangs relatifs, mais on choisira la loi du rang R^j plus concentrée vers les grandes valeurs. On adopte

$$(46) \quad q_{ij} = \Pr[R^j = i] = a \cdot \left(\frac{1}{q}\right)^i$$

$$j = 1, \dots, K, \quad i = 1, \dots, K - j + 1,$$

$$(47) \quad q < 1.$$

Dans ces conditions le nombre K aura tendance à être placé à droite de la permutation, le nombre $K - 1$ sera fréquemment à droite des valeurs restantes ... ce qui produit la tendance à l'augmentation recherchée.

La normalisation de (46) exige que

$$(48) \quad q_{ij} = \frac{1 - q}{1 - q^{K - j + 1}} q^{K - j + 1 - i},$$

et, à cause de l'indépendance, la loi distordue des rangs relatifs vaut



$$(49) \quad Q(x^1, \dots, x^K) = \frac{(1-q)^K}{\prod_{j=1}^K (1-q^j)} \cdot q^{\left[\frac{K(K+1)}{2} - \sum_{j=1}^K r^j \right]}$$

Pour $q=1$ on retrouve la loi équirépartie initiale. Appelons q le coefficient de distorsion.

Il est clair que les lois P et Q ainsi définies sur $(\mathcal{X}, \mathcal{B})$ sont équivalentes. On peut donc appliquer l'estimateur (32) de simulation rapide à la fausse alarme $\alpha(s)$ du test de Spearman-Rho.

D'après (33) et (49), la variable aléatoire Z dont on cherche la moyenne $\alpha(s)$ à l'aide de simulations, vaut ici

$$(50) \quad Z = \frac{1}{K!} \prod_{j=1}^K \frac{1-q^j}{1-q} \cdot q^{-\frac{K(K+1)}{2} + \sum_{j=1}^K R^j} \cdot H\left(\sum_{j=1}^K j X^j - s\right),$$

où $R = (R^1, \dots, R^K)$ est le vecteur des rangs correspondant à la permutation $X = (X^1, \dots, X^K)$, et tiré à l'aide de la loi distordue (49).

L'estimateur de simulation rapide basé sur (50) vaut évidemment

$$(51) \quad \hat{\alpha}(s) = \frac{1}{N} \sum_{\ell=1}^N Z_{\ell},$$

en fonction des N tirages indépendants Z_{ℓ} de la v.a. Z .

Naturellement le coefficient de distorsion q doit être adapté au domaine des valeurs α que l'on estime (donc au seuil s) pour assurer à l'estimateur (51) l'intervalle de confiance requis. La détermination de q en fonction de s implique de calculer la variance de Z , (50), ou tout au moins un de ses majorants. D'après (49) et (50)

$$(52) \quad E(Z^2) = \frac{1}{(K!)^2} \sum_{m=1}^{K!} \frac{1}{Q(r_m)} \cdot H\left(\sum_{j=1}^K j x_m^j - s\right),$$

où r_m décrit l'ensemble à $K!$ éléments de tous les vecteurs rangs



possibles et où $x_m = (x_m^1, \dots, x_m^K)$ est la permutation correspondante. Soit B_s la partie de \mathcal{X} pour laquelle le test de Spearman-Rho est supérieur à s . Puisque la loi initiale des permutations est équirépartie sur \mathcal{X} , B_s comporte exactement $K! \alpha(s)$ permutations, de sorte que

$$(53) \quad E(Z^2) = \frac{1}{(K!)^2} \sum_{x \in B_s} \frac{1}{Q(r)} \leq \frac{\alpha(s)}{K!} \sup_{x \in B_s} \frac{1}{Q(r)}.$$

D'après la forme (49) de $Q(r)$, on doit trouver une borne inférieure $M(s)$ de $\sum_{j=1}^K r^j$ pour tout r dans B_s :

$$(54) \quad \sum_{j=1}^K j x^j \geq s \quad \Rightarrow \quad \sum_{j=1}^K r^j \geq M(s).$$

Pour cela on considère la partition de \mathcal{X}

$$\mathcal{X} = \sum_{M=K}^a \mathcal{X}_M, \quad a = \frac{K(K+1)}{2},$$

$$(55) \quad \mathcal{X}_M = \{x; \sum_{j=1}^K r^j = M\},$$

et on génère itérativement une permutation r_M et une seule dans chaque \mathcal{X}_M , qui attribue une grande valeur y_M au test de Spearman-Rho. L'algorithme adopté est le suivant : on choisit $r_M^1 = K$ si $M \geq 2K-1$, puis $r_M^2 = K-1$ si $M \geq 3K-3$ etc... ; par exemple, pour $M = 2K+1$, $r_M^1 = K$, $r_M^2 = 3$, $r_M^3 = 1$, ..., $r_M^K = 1$. Cela correspond à donner à K le rang le plus élevé possible tout en respectant $\sum_{j=1}^K r^j = M$, puis faire de même pour $K-1$ etc... Il est

aisé de montrer que les valeurs y_M forment une suite strictement croissante. Par ailleurs pour les grandes valeurs du test qui seules nous intéressent, donc pour les grandes valeurs de M , y_M forme pratiquement une progression arithmétique de raison 1 et la valeur y du test est constante sur \mathcal{X}_M (à quelques termes près).

Ainsi en posant

$$(56) \quad M(s) = \{M; y_{M-1} < s \leq y_M\},$$



on peut considérer que pour les seuils élevés

$$B_s \subset \sum_{M=M(s)}^{\infty} \mathcal{L}_M,$$

ce qui implique, pour x dans B_s

$$\sum_{j=1}^K r^j \geq M(s).$$

(cf. équation (54)). A l'aide de ce minorant $M(s)$, (53) devient, d'après (49)

$$(57) \quad E(Z^2) \leq f(q, s) \cdot \alpha(s),$$

où la fonction $f(q, s)$ est définie par

$$(58) \quad f(q, s) = q^{-\frac{K(K+1)}{2} + M(s)} \prod_{j=1}^K \frac{1-q^j}{j(1-q)}.$$

Pour la simulation on a choisi $K = 16$. Sur la figure 5 les fonctions (58) ont été tracées en échelles semi-logarithmiques pour les valeurs de q comprises entre 0 et 1 et pour les valeurs de $M(s)$ correspondant à des seuils variables de 1300 à 1496 qui est la plus grande valeur possible. Elles présentent toutes un minimum unique, qui fournit donc pour Z^2 un minimum (pour la borne) de la variance. C'est la valeur correspondante q_s de q qui est prise dans chacune des simulations.

D'après l'étude générale du II - formules (21) et (21 bis) - pour assurer avec la confiance $1-\eta = 0,95$ une précision relative $k = 0,2$, il faut que

$$(59) \quad E(Z^2) \leq \left(\frac{N}{96} + 1\right) \alpha^2(s).$$

Il en est ainsi lorsque

$$(60) \quad \alpha^2(s) \cdot \left(\frac{N}{96} + 1\right) \geq f(q_s, s) \alpha(s).$$



Nous avons vérifié sur le tracé des courbes $f(q, s)$ (cf. figure 5) que (60) est effectivement vérifié pour chaque valeur du seuil, c'est-à-dire que l'estimateur assure la précision désirée. Nous avons aussi vérifié expérimentalement la validité de chaque simulation à l'aide des valeurs expérimentales de $\alpha(s)$ et de $E(Z^2)$ qui ont toujours satisfait à la condition (59). Le temps d'UNIVAC 1108 nécessaire est faible puisque 4 minutes ont permis le calcul de 9 valeurs de la fausse alarme. Le nombre N de simulations était 1000.

Les résultats pour la fausse alarme $\alpha(s)$ sont portés sur la figure 6.

Quelques vérifications analytiques sont possibles. D'abord pour le dernier seuil $s = 1496$, il n'existe qu'une seule permutation qui déclenche l'alarme donc

$$\alpha(1496) = \frac{1}{16!} = 0,48 \cdot 10^{-13}.$$

En outre, par un calcul analytique approché basé sur un développement de la loi du test en série de Gram-Charlier, Glasser et Winter [10] ont déterminé le seuil s correspondant à certaines valeurs de $\alpha(s)$ supérieures à 10^{-3} . Les points correspondants ($\alpha = 5 \cdot 10^{-2}, 2,5 \cdot 10^{-2}, 10^{-2}, \dots, 5 \cdot 10^{-3}, 10^{-3}$) sont portés sur la figure. L'accord est excellent.

VI - CONCLUSION

La méthode de simulation rapide que nous venons d'exposer a pour but de mesurer la probabilité d'un évènement "intéressant". Cet évènement concerne la sortie d'un système de traitement qui reçoit une entrée aléatoire. La méthode est adaptée à la mesure de probabilités faibles. Sur les deux exemples que nous venons d'exposer, on constate sa grande efficacité du point de vue temps de calcul. La simulation rapide peut même donner des résultats spectaculaires, puisque, dans certains cas elle permet d'évaluer avec précision des probabilités de l'ordre de 10^{-13} . Cependant son

[9] Il suffit que P soit absolument continue par rapport à Q ;
mais l'équivalence fait mieux apparaître le sens physique de la pondération qui intervient dans (32).

[10] Glasser G.J., WINTER R.F. "Critical values of rank correlation for testing the hypothesis of independence"
Biometrika, vol. 48, p. 444-448, 1961.



VARIANCE DE LA VARIABLE NORMALE SIMULEE
(Exemple 1)

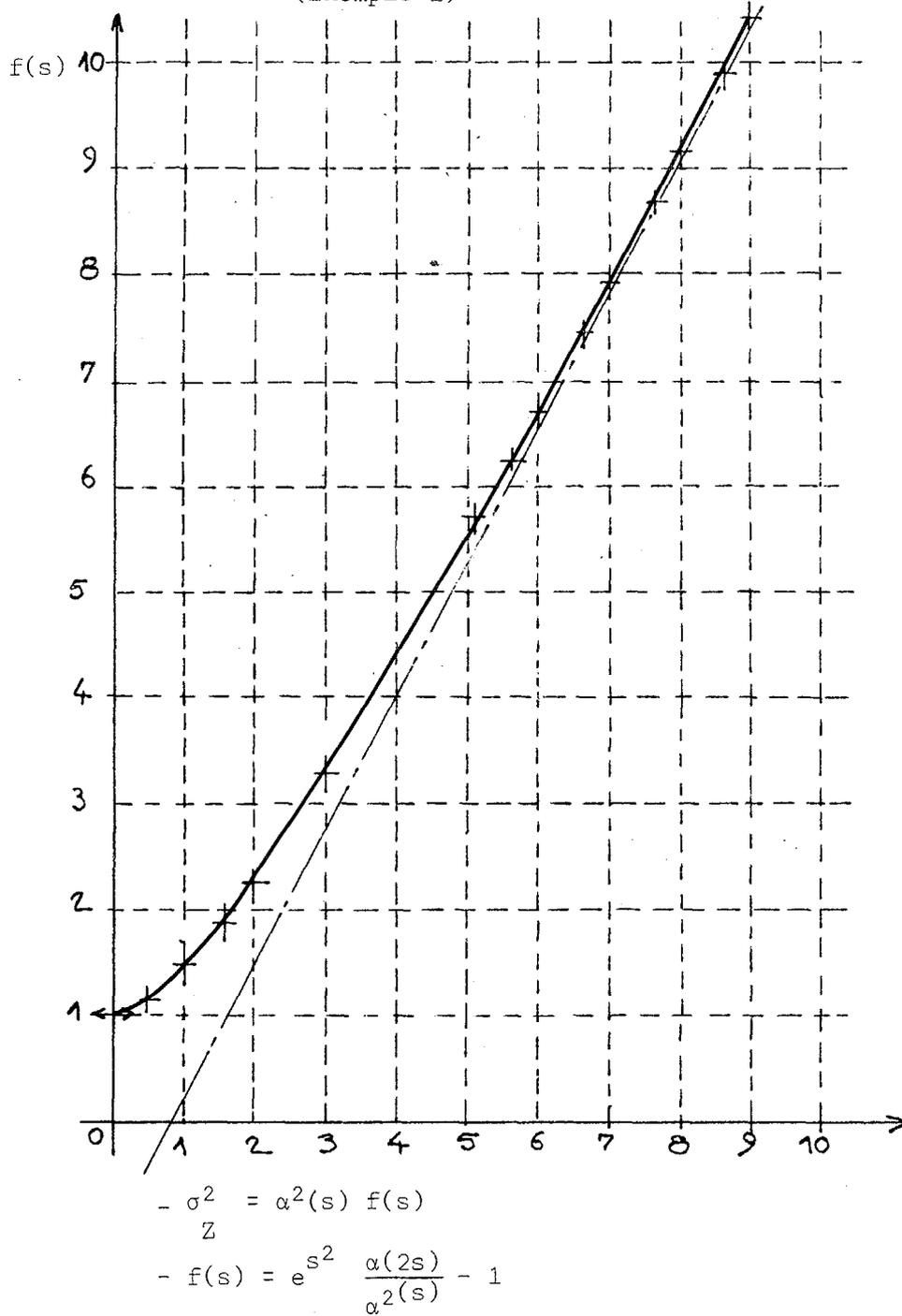
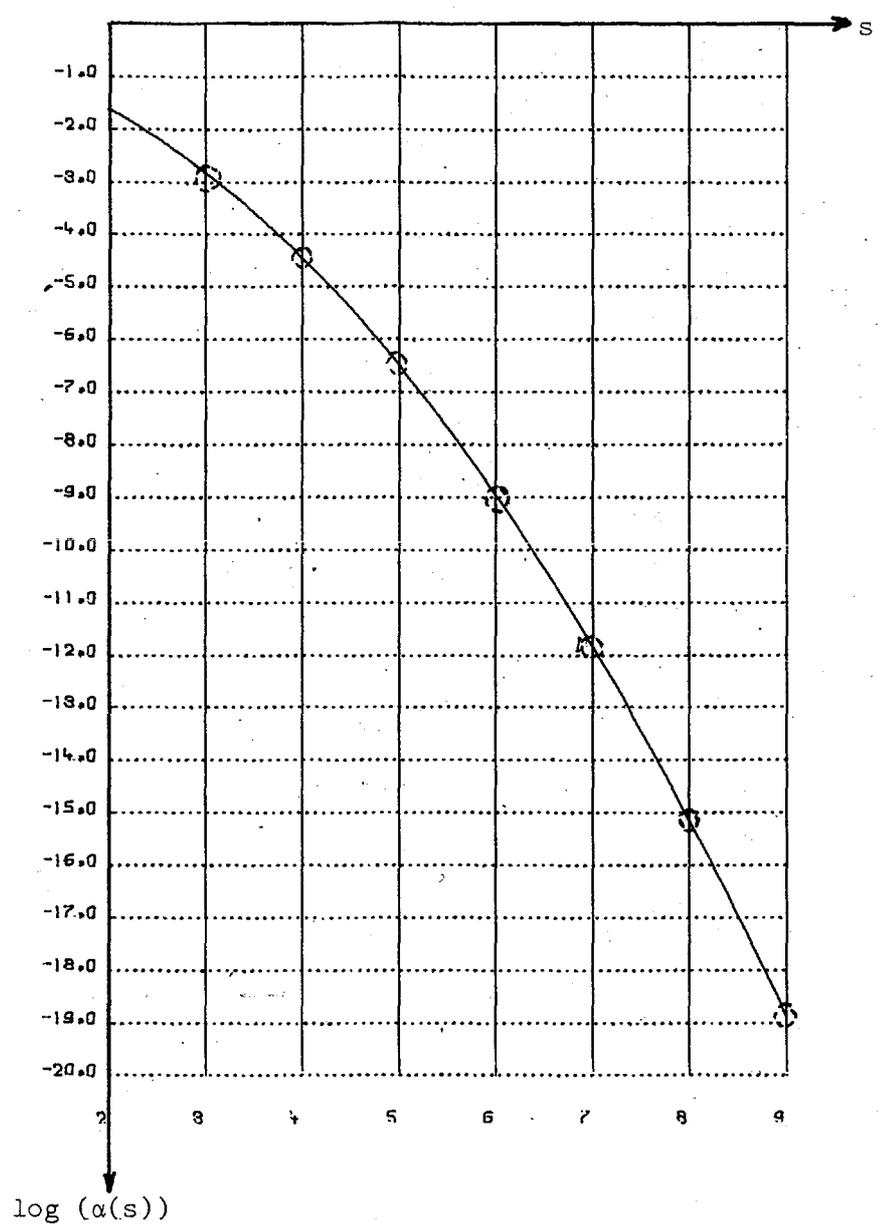


FIGURE 3

LA FONCTION D'ERREUR
 évaluée par simulation rapide

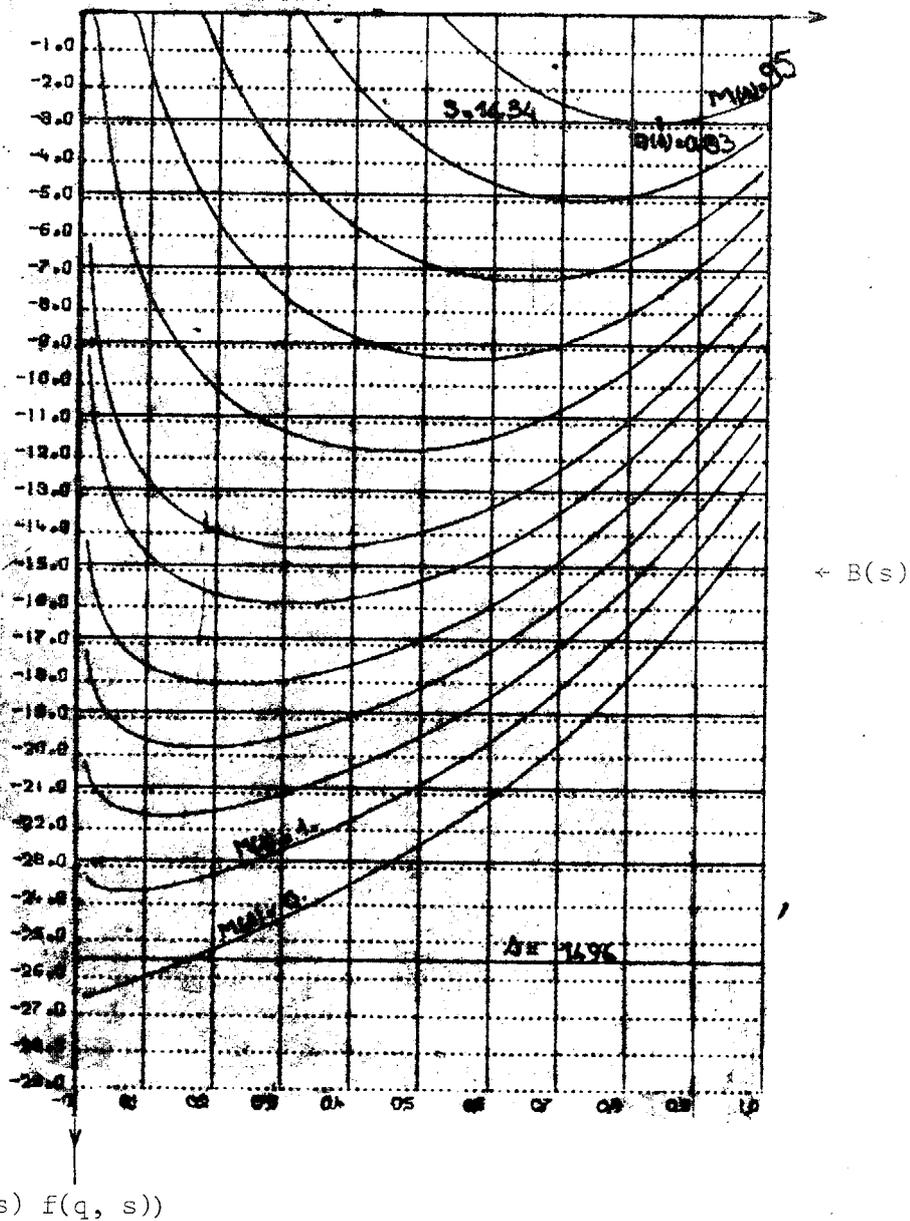


Les points marqués d'un cercle ont été
 calculés analytiquement.

FIGURE 4



VARIANCE DE LA VARIABLE SIMULEE. TEST de SPEARMAN-RHO (exemple 2)

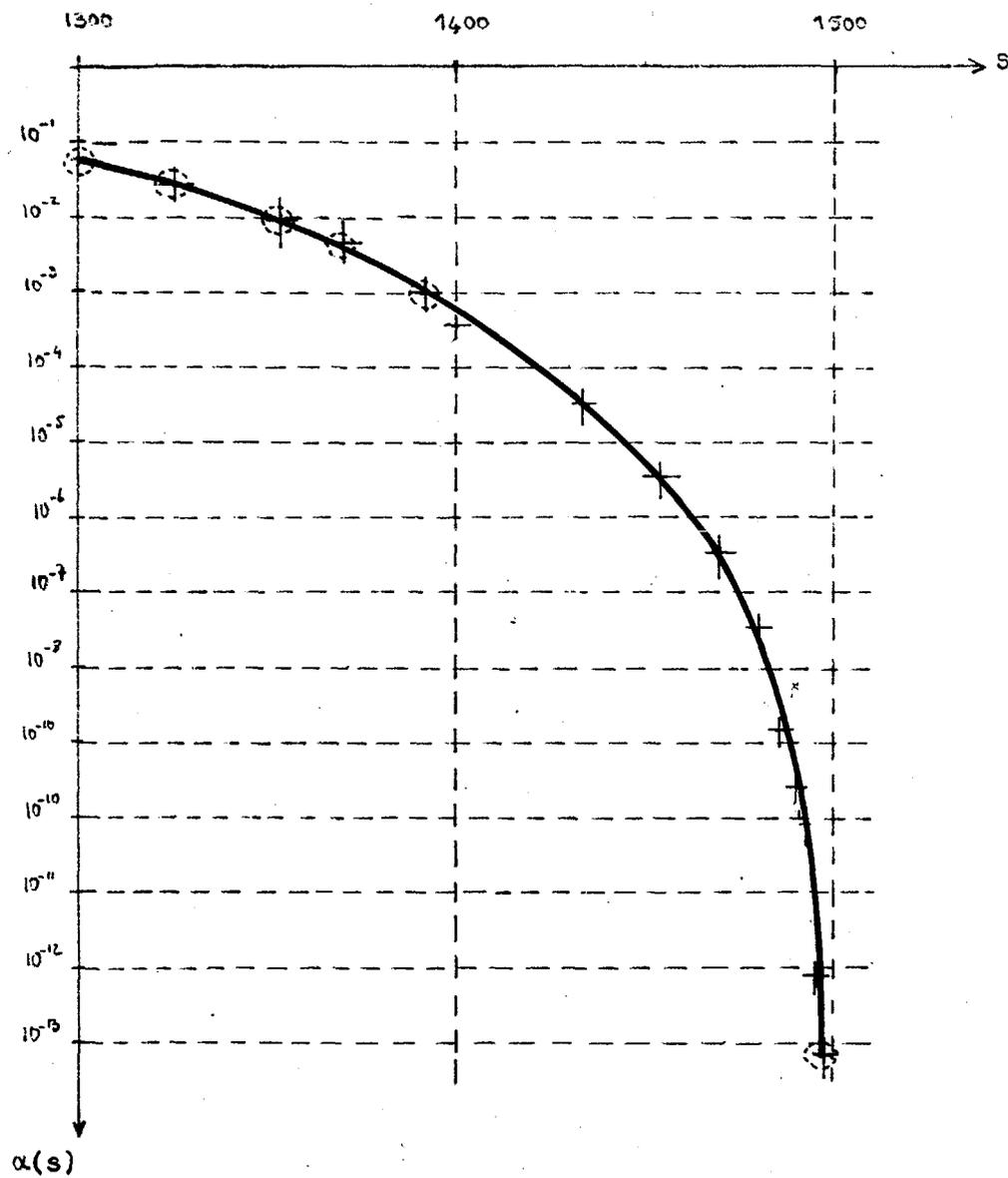


- Sur chacune des courbes nous vérifions l'inégalité (60) $\alpha^2(s) \left(\frac{N}{96} + 1\right) \geq f(q_s, s) \alpha(s)$
- $B(s) = \alpha^2(s) \left(\frac{N}{96} + 1\right)$

FIGURE 5



PROBABILITE DE FAUSSE ALARME DU
TEST DE SPEARMAN-RHO
évaluée par simulation rapide



Les points marqués d'un cercle ont été calculés analytiquement.

FIGURE 6