



HAL
open science

Exemples d'utilisation d'un calculateur en statistique

G. Hisleur

► **To cite this version:**

G. Hisleur. Exemples d'utilisation d'un calculateur en statistique. Modélisation et simulation. Université Joseph-Fourier - Grenoble I, 1969. Français. NNT: . tel-00281656

HAL Id: tel-00281656

<https://theses.hal.science/tel-00281656>

Submitted on 23 May 2008

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

d'ordre :

TV 548

THESES

présentées à
LA FACULTE DES SCIENCES DE GRENOBLE
pour obtenir
LE GRADE DE DOCTEUR INGENIEUR

par

G. Hisleur

Ingénieur I. M. A. G.



Première thèse :

Exemples d'utilisation d'un calculateur en statistique

Deuxième thèse :

SEPARABILITE ET MODELES ELARGIS



Thèses soutenues le 11 Février 1969 devant la commission d'examen :

MM. J. KUNTZMANN

Président

J. R. BARRA

Examineur

Y. MAINGUY

Examineur

L I S T E d e s P R O F E S S E U R S

DOYEN HONORAIRE : M. MORET

DOYEN : M. BONNIER

PROFESSEURS TITULAIRES

MM. NEEL Louis	Chaire de Physique Exérimentale
HEILMANN René	Chaire de Chimie
KRAVTCHENKO Julien	Chaire de Mécanique Rationnelle
CHABAUTY Claude	Chaire de calcul différentiel et Intégral
BENOIT Jean	Chaire de Radioélectricité
CHENE Marcel	Chaire de Chimie Papetière
FELICI Noël	Chaire d'Electrostatique
KUNTZMANN Jean	Chaire de Mathématiques Appliquées
BARBIER Reynold	Chaire de Géologie Appliquée
SANTON Lucien	Chaire de Mécanique des Fluides
OZENDA Paul	Chaire de Botanique
FALLOT Maurice	Chaire de Physique Industrielle
KOSZUL Jean Louis	Chaire de Mathématiques
GALVANI O.	Mathématiques
MOUSSA André	Chaire de Chimie Nucléaire
TRAYNARD Philippe	Chaire de Chimie Générale
SOUTIF Michel	Chaire de Physique Générale
CRAYA Antoine	Chaire d'Hydrodynamique
REULOS R.	Théorie des Champs
BESSON Jean	Chaire de Chimie
AYANT Yves	Physique Approfondie
GALLISSOT	Mathématiques
Mlle LUTZ Elisabeth	Mathématiques
ELAMBERT Maurice	Chaire de Mathématiques
BOUCHEZ Robert	Physique Nucléaire
LLIBOUTRY Louis	Géophysique
MICHEL Robert	Chaire de Minéralogie et Pétrographie
BONNIER Etienne	Chaire d'Electrochimie et d'Electrométallurgie
DESSEAUX Georges	Chaire de Physiologie Animale
PILLET E	Chaire de Physique Industrielle et Electrotechnique
YOCCOZ Jean	Chaire de Physique Nucléaire Théorique
DEBELMAS Jacques	Chaire de Géologie Générale
GERBER R.	Mathématiques
PAUTHENET R.	Electrotechnique
VAUQUOIS B.	Chaire de calcul électronique
BARJON R.	Physique Nucléaire
BARBIER Jean-Claude	Chaire de Physique
SILBER R.	Mécanique des Fluides
BUYLE-BODIN Maurice	Chaire d'Electronique
DREYFUS B.	Thermodynamique

PROFESSEURS TITULAIRES suite

MM. KLEIN J.	Mathématiques
VAILLANT F.	Zoologie et Hydrobiologie
ARNAUD Paul	Chaire de Chimie
SENGEL P.	Chaire de Zoologie
BARNOUD F.	Chaire de Biosynthèse de la Cellulose
BRISSONNEAU P.	Physique
GAGNAIRE	Chaire de Chimie Physique
Mme KOFLER L.	Botanique
DEGRANGE Charles	Zoologie
PEBAY-PEROULA J.C.	Physique
RASSAT A.	Chaire de Chimie Systématique
DUCROS P.	Chaire de Cristallographie Physique
DODU Jacques	Chaire de Mécanique Appliquée I.U.T.
ANGLES D'AURIAC P.	Mécanique des Fluides
LACAZE A.	Thermodynamique

PROFESSEURS SANS CHAIRE

MM. GIDON P.	Géologie et Minéralogie
GIRAUD P.	Géologie
PERRET R.	Servomécanisme
Mme BARBIER M.J.	Electrochimie
Mme SOUTIF J.	Physique
COHEN J.	Electrotechnique
DEPASSEL R.	Mécanique des Fluides
GASTINEL A.	Mathématiques Appliquées
GLENAT R.	Chimie
BARRA J.	Mathématiques Appliquées
COUMES A.	Electronique
PERRIAUX J.	Géologie et Minéralogie
ROBERT A.	Chimie Papetière
BIAREZ J.P.	Mécanique Physique
BONNET G.	Electronique
CAUQUIS G.	Chimie Générale
BONNETAIN L.	Chimie Minérale
DEPOMMIER P.	Etude Nucléaire et Génie Atomique
HACQUES Gérard	Calcul Numérique
POLOUJADOFF M.	Electrotechnique

PROFESSEURS ASSOCIES

MM. NAPP-ZINN	Botanique
RODRIGUES Alexandre	Mathématiques Pures
STANDING Kenneth	Physique Nucléaire

MAITRES DE CONFERENCES

MM. LANCIA Roland	Physique Atomique
Mme KAHANE J.	Physique
DEPORTES C.	Chimie
Mme BOUCHE L.	Mathématiques
SARROT-REYNAUD	Géologie Propédeutique

MAITRES DE CONFERENCES suite

Mme BONNIET M.J.
 KAHANE A.
 DOLIQUE J.M.
 BRIERE G.
 DESRE G.
 LAJZROWICZ J.
 VALENTIN P.
 BERTRANDIAS J.P.
 LAURENT P.
 CAUBET J.P.
 PAYAN J.J.
 Mme BERTRANDIAS F.
 LONGEQUEUE J.P.
 NIVAT M.
 SOHM J.C.
 ZADWORNY F.
 DURAND F.
 CARLER G.
 AUBERT G.
 DELPUECH J.J.
 PFISTER J.C.
 CHIBON P.
 IDELMAN S.
 BOUVARD Maurice
 RICHARD Lucien
 PELMONT Jean
 BLOCH D.
 BOUSSARD J.Claude
 MOREAU René
 BRUGEL L.
 SIBILLE R.
 ARMAND Yves
 BOLLIET Louis
 KUHN Gérard
 GERMAIN Jean Pierre
 CONTE René
 JOLY Jean René
 PIERY Yvette
 BENZAKEN Claude

Chimie
 Physique Générale
 Electronique
 Physique M.P.C.
 Chimie S.P.C.N.
 Physique M.P.C.
 Physique M.P.C.
 Mathématiques Appliquées T.M.P.
 Mathématiques Appliquées T.M.P.
 Mathématiques Pures
 Mathématiques
 Mathématiques Pures M.P.C.
 Physique
 Mathématiques Appliquées
 Electrochimie
 Electronique
 Chimie Physique
 Biologie Végétale
 Physique M.P.C.
 Chimie Organique
 Physique C.P.E.M.
 Biologie Animale
 Physiologie Animale
 Hydrologie
 Botanique
 Physiologie Animale
 Electrotechnique I.P.
 Mathématiques Appliquées I.P.
 Hydraulique I.P.
 Energétique I.U.T.
 Construction Mécanique I.U.T.
 Chimie I.U.T.
 Informatique I.U.T.
 Energétique I.U.T.
 Construction Mécanique I.U.T.
 Thermodynamique
 Mathématiques Pures
 Biologie Animale
 Mathématiques Appliquées

MAITRES DE CONFERENCES ASSOCIES

MM. SAWCZUK A.
 CHEEKE J.
 YAMADA O.
 NATR Lubomir
 NAYLOR Arch
 SILBER Léo
 NAZAKI Akihiro
 RUTLEDGE Joseph
 DONOHO Paul
 EGGER Kurt

Mécanique des Fluides
 Thermodynamique
 Physique du Solide
 B.M.P.V.
 Physique Industrielle
 Radioélectricité
 Mathématiques Appliquées
 Mathématiques Appliquées
 Physique Générale
 B.M.P.V.

Je tiens à exprimer ma très respectueuse gratitude à Monsieur le Professeur J. KUNTZMANN, Directeur de l'Institut de Mathématiques Appliquées de l'Université de Grenoble, qui a bien voulu me faire l'honneur de présider le Jury.

Je remercie particulièrement Monsieur le Professeur J.R. BARRA pour la confiance qu'il m'a accordée en me proposant le sujet de cette étude et qui, par son intérêt et ses conseils, m'a aidé à mener mon travail à bien.

Je remercie Monsieur le Professeur V. MAINGUY d'avoir accepté de participer au Jury et de m'avoir permis, en me proposant le sujet d'une deuxième thèse, d'enrichir mes connaissances.

Je remercie Monsieur E. MORICE, Directeur de la "Revue de Statistique Appliquée" pour les problèmes qu'il m'a proposé de résoudre.

Je remercie enfin tous les membres du Laboratoire de Calcul qui ont contribué à ce travail.

TABLE DES MATIERES

<u>INTRODUCTION</u>	2
<u>CHAPITRE-I : LOIS DE PROBABILITE USUELLES</u>	7
1 - Loi Normale	9
2 - Loi de Poisson	13
3 - Loi Binomiale	17
4 - Loi du χ^2	22
5 - Loi de Student	25
6 - Loi du χ^2 décentrée	28
7 - Loi Gamma décentrée	36
8 - Loi Beta décentrée	42
<u>CHAPITRE -II : EXEMPLE D'UTILISATION DE DONNEES STATISTIQUES</u>	48
<u>CHAPITRE-III : TESTS D'HYPOTHESES</u>	54
1 - Test du χ^2	56
2 - Test de normalité	65
3 - Test de Kolmogorov-Smirnov	73
4 - Taille de l'échantillon dans un test de Student	80
5 - Test d'hypothèses linéaires	98
6 - Test du caractère aléatoire d'une suite	108
<u>CHAPITRE-IV : PROBLEMES D'ESTIMATION SUR ECHANTILLONS GAUSSIENS</u>	115
1 - Estimation de σ , m étant inconnue	116
2 - Estimation de σ , m étant connue	125
3 - Estimation de m et σ simultanément	127
<u>BIBLIOGRAPHIE</u>	147

I N T R O D U C T I O N

L'utilisation des ordinateurs fait entrevoir aux statisticiens des perspectives nouvelles. Des résultats encourageants ont déjà été obtenus dans un certain nombre de domaines par des organismes tels que l'I.N.S.E.E., l'I.F.O.P. et la S.O.F.R.E.S.

Le calculateur peut, en statistique, jouer trois rôles qui sont essentiels pour le chercheur :

- le libérer de tâches souvent fastidieuses, ce qui lui permet de se consacrer à des travaux créatifs, d'un niveau supérieur, qui resteront toujours en dehors des possibilités des machines;
- traiter des problèmes de plus en plus complexes, et lui donner ainsi le moyen d'aborder des questions nouvelles;
- préparer les décisions, par une meilleure exploitation de l'information, ce qui a pour effet d'accroître la sûreté de son jugement.

L'objet de cette thèse est d'établir un ensemble de programmes destinés à montrer que l'on peut systématiquement s'affranchir des calculs numériques. On peut raisonnablement prévoir que, dans l'avenir, l'ordinateur aura la charge de remplacer les tables et d'effectuer toutes les opérations à la fois plus rapidement et plus facilement.

Les programmes que nous donnons sont à usage courant, peut-être même pédagogique. Ils peuvent être comparés à ceux qui sont fournis dans les bibliothèques de calcul des grands ordinateurs. On constatera que nos deux soucis majeurs ont été de rendre leur utilisation aussi pratique que possible et leur enchaînement aisé.

Le langage que nous avons utilisé est ALGOL. Le choix devait être fait entre FORTRAN et ALGOL qui, à notre avis, sont les plus adaptés à nos problèmes. Les difficultés qu'amène FORTRAN avec les expressions mixtes et ses restrictions très nombreuses pour le calcul sur des données vectorielles ou matricielles, qui sont très courantes en statistique, nous ont fait

retenir le second. La lecture assez facile d'ALGOL, son caractère universel nous ont également influencé.

Depuis quelques années on voit apparaître chez certains auteurs (B.E. Cooper [5],[6], J.C. Gower, H.R. Simpson et A.H. Martin [10]) l'utilité d'un langage spécialisé, à usage exclusivement statistique, dont les caractéristiques essentielles semblent être de trois types :

- possibilité d'opérer sur des tableaux, ou des parties de tableaux, à plusieurs dimensions, aussi bien que sur de simples variables;
- grande souplesse lors des analyses de données;
- règles d'entrée beaucoup moins strictes.

Plusieurs ont été proposés, tels ASCOP, METO, SEP et DAPHNE, mais il apparaît que, jusqu'à maintenant, aucun ne soit pleinement satisfaisant.

Le chapitre I est consacré à un certain nombre de procédures de calcul qui ont pour but de remplacer les tables. Elles fournissent les fonctions de répartition des lois de probabilité les plus couramment utilisées en statistique et en analyse de la variance. Elles interviennent dans la majorité des problèmes qui se posent au statisticien.

Notre but a été de constituer un ensemble aussi complet que possible. Ce travail n'a encore jamais été effectué systématiquement, les programmes existants ayant été écrits pour des usages particuliers.

Certaines fonctions sont déjà tabulées : lois Normale, Binomiale, de Poisson, du X^2 et de Student. Les procédures ont donc pu être vérifiées : elles ont confirmé les performances attendues.

Les lois du X^2 , Gamma et Beta décentrées ne sont en général pas tabulées complètement. Ainsi, la première dépend de deux paramètres n et p . Deux seules tables, à notre connaissance, lui sont consacrées. Elles sont dues à Fix [9] et Patnaik [18]. Elles possèdent une large zone commune mais laissent à découvert le domaine en n et p :

$$1 \leq n \leq 30$$

$$12 < p < 40$$

Il nous a semblé intéressant d'effectuer le calcul des valeurs qui faisaient défaut.

Les avantages que présentent de telles procédures sont nombreux. En premier lieu, nous pouvons citer le programme détaillé au paragraphe 8 et qui, ses paramètres étant déclarés réels, permet d'obtenir les fonctions de répartition des lois de Fisher, de Tang ou d'Hotelling. D'autre part, leur intérêt réside dans leur précision qui est souvent bien supérieure à celle que fournissent les tables, le gain de temps très appréciable surtout lorsque le nombre de valeurs désirées est grand, enfin, il n'est plus utile de recourir à des interpolations puisque maintenant le calcul des fonctions est possible en n'importe quel point.

Dans le chapitre II nous montrons à l'aide d'un exemple comment on peut aborder le traitement automatique de données statistiques. Nous avons voulu prouver que des travaux tels que tri, dénombrement et calculs de caractéristiques peuvent être effectués simplement.

Le chapitre III examine les problèmes de tests d'hypothèses.

Les premiers paragraphes traitent respectivement des tests du X^2 , de normalité et de Kolmogorov-Smirnov. Nous pouvons remarquer que, à la différence des procédures du même type que l'on trouve ailleurs, les programmes de traitement ne nécessitent l'entrée que des données et du niveau de signification : ils contiennent les sous-programmes qui calculent les séparatrices.

Ce chapitre se termine par trois problèmes qui ont donné lieu à des résultats nouveaux.

Au paragraphe 4 il s'agit de rechercher la taille de l'échantillon qui est optimale pour réaliser les risques α et β , fixés à l'avance, dans un test de Student. M. Harris, D.G. Horvitz et A.M. Mood [11] ont établi des tables qui ne permettent une résolution que dans quelques cas particuliers. Nous avons généralisé leur technique pour obtenir la taille requise dans tous les cas. Nous avons alors comparé avec la méthode Stein [19].

Le paragraphe 5 traite du test de plusieurs hypothèses linéaires. Le calcul des probabilités d'erreur fait intervenir la fonction de répartition d'un vecteur aléatoire, et non plus d'une variable comme dans un test ordinaire. L'ordinateur a rendu ce calcul possible; nous avons écrit une procédure qui permet de le réaliser.

Le dernier paragraphe est consacré à un test non paramétrique [3] Il donne le moyen de reconnaître si une suite de digits binaires possède l'aspect "aléatoire".

Les essais réalisés sur 50.000 digits de l'écriture en binaire du nombre π confirment [1] que ce nombre ne possède pas l'aspect du hasard, contredisant ainsi les conclusions de Esmenjaud-Bonnardel [7] basées sur l'utilisation des tests usuels.

Le chapitre IV, enfin, examine trois problèmes d'estimation sur échantillons gaussiens.

Les résultats théoriques que nous utilisons, sont désormais classiques en statistique. Il semble pourtant que les deux tables que nous avons obtenues soient nouvelles : des recherches bibliographiques ne nous ont pas permis d'en trouver de semblables.

La première permet de donner un intervalle de confiance de longueur minimum pour la variance d'une population lorsque la moyenne est inconnue. Les deux constantes a et b que nous avons tabulées sont, en pratique, approchées par $Q_{n-1}^{-1}(\frac{\alpha}{2})$ et $Q_{n-1}^{-1}(1-\frac{\alpha}{2})$ $Q_n(z)$ étant la fonction de répartition de la loi du X^2 réduite à n degrés de liberté. Cette approximation est correcte lorsque n est grand mais les écarts sont sensibles pour les petites valeurs de n .

Dans le cas où la moyenne est connue la table précédente est encore valable. Nous montrons qu'il suffit de changer n en $n+1$.

Le dernier paragraphe est consacré à l'estimation simultanée de la moyenne et de la variance. On détermine le domaine d'acceptation de surface minimum pour ces deux paramètres. On compare ensuite avec la méthode usuelle non optimale et les tables de Neyman et Pearson [16]. On constate que le domaine de Neyman et Pearson et celui de surface minimum sont voisins.

Il est certain que les utilisations des possibilités du calculateur ne sont pas limitées aux problèmes particuliers que nous avons étudiés. Notre travail n'est qu'un exemple de ce que pourrait apporter l'usage de moyens de calcul puissants en statistique appliquée.

En conclusion nous pouvons citer les remarques essentielles qui ont été faites au Congrès de Chilton [20] sur la nécessité de posséder un véritable langage statistique et de pouvoir intervenir à tout moment lors du déroulement des programmes, Les multi-traitements et la programmation en temps réel répondent à cette seconde exigence. Pour nous, qui sommes restés sur un plan quasiment pédagogique, le langage que nous avons utilisé nous a semblé commode.

CHAPITRE I

LOIS DE PROBABILITE USUELLES

Ce premier chapitre a pour objet de donner les procédures qui permettent de calculer les fonctions de répartition des lois de probabilité les plus usuelles en statistique.

On peut considérer trois parties.

- Les trois premières distributions que nous étudions (Laplace-Gauss, Poisson, Binomiale) sont considérées comme "élémentaires". Elles régissent dans la pratique de très nombreux phénomènes aléatoires. D'autre part, d'après le théorème central-limite, la plupart des lois de probabilité sont asymptotiquement normalement distribuées. La loi normale est alors d'une grande utilité puisqu'on la prend pour les approcher lorsqu'on est suffisamment près de la limite.
- Les lois du X^2 et de Student, que nous examinons ensuite, servent surtout pour les tests fondamentaux qui dérivent de la loi normale : les tests du X^2 et de Student.
- Les trois derniers paragraphes traitent des lois qui sont les plus importantes en analyse de la variance. Nous verrons, en effet, que les lois de Fisher, d'Hotelling et de Tang, ne sont que des cas particuliers d'une loi Beta décentrée : la procédure BETADEC permet donc de les calculer.

Le but que nous avons cherché à atteindre est la constitution d'un ensemble aussi complet que possible. Il apparaît que ce travail n'a encore jamais été effectué systématiquement : les programmes qui existent ont été écrits pour des usages particuliers.

Dans ce chapitre et dans les chapitres suivants, les programmes de résolution sur machine sont écrits en langage ALGOL pour le compilateur

ALGOL de l'ordinateur IBM-7044 de l'Institut de Mathématiques Appliquées de l'Université de GRENOBLE. Les instructions d'entrée-sortie et les évaluations de précision sont donc liées à cette machine et à ce compilateur.

1 - LOI NORMALE

1.1. - Loi normale centrée réduite $\mathcal{N}(0,1)$

La variable aléatoire X est dite suivre la loi normale centrée réduite $\mathcal{N}(0,1)$ si sa densité de probabilité f(x) est telle que :

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad x \in \mathbb{R}.$$

Le terme "centrée réduite" provient de ce que la moyenne et l'écart-moyen quadratique de cette loi sont respectivement 0 et 1.

1.2. - Loi normale $\mathcal{N}(m, \sigma)$

Soient X une variable aléatoire normale $\mathcal{N}(0,1)$, m et σ deux nombres donnés ($\sigma > 0$); on dira que la variable aléatoire :

$$Y = m + \sigma X$$

suit une loi normale $\mathcal{N}(m, \sigma)$.

La loi de Y admet alors pour densité de probabilité :

$$n(y; m, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-m)^2}{2\sigma^2}}, \quad y \in \mathbb{R}$$

Les deux paramètres m et σ s'interprètent comme la moyenne et l'écart-type de Y.

1.3. - Procédure NORMALE

La procédure NORMALE calcule la valeur au point x de la fonction de répartition d'une loi normale $\mathcal{N}(m, \sigma)$, c'est-à-dire :

$$N(x; m, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{(y-m)^2}{2\sigma^2}} dy$$

et ceci pour tout triplet (x, m, σ) fourni par l'utilisateur. La seule

restriction concerne le paramètre σ qui doit être positif,

Par l'intermédiaire du changement de variable :

$$z = \frac{y - m}{\sigma}$$

on se ramène au calcul de l'intégrale :

$$N(z_0) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z_0} e^{-\frac{z^2}{2}} dz$$

où :

$$z_0 = \frac{x - m}{\sigma}$$

En introduisant la fonction numérique signe (t) telle que :

$$\text{signe (t)} = \begin{cases} 1 & \text{si } t > 0 \\ -1 & \text{si } t \leq 0 \end{cases}$$

nous pouvons écrire :

$$N(z_0) = \frac{1}{2} + \text{signe}(z_0) \cdot \frac{1}{\sqrt{2\pi}} \int_0^{|z_0|} e^{-\frac{z^2}{2}} dz$$

relation qui permet de n'avoir à étudier que des intégrales du type :

$$\frac{1}{\sqrt{2\pi}} \int_0^Z e^{-\frac{z^2}{2}} dz, Z > 0.$$

Deux cas sont à envisager.

a) $|z_0|$ est inférieur à 3.5.

On utilise alors la série convergente :

$$\frac{z^2}{e^2} \int_0^z e^{-\frac{t^2}{2}} dt = z + \frac{z^3}{3} + \frac{z^5}{3.5} + \frac{z^7}{3.5.7} + \dots$$

b) $|z_0|$ est supérieur ou égal à 3.5.

On a recours à la fraction continue :

$$e^{-\frac{z^2}{2}} \int_z^{\infty} e^{-\frac{t^2}{2}} dt = \frac{1}{z} - \frac{1}{z^3} + \frac{2}{z^5} - \frac{3}{z^7} + \frac{4}{z^9} - \dots$$

Le point de changement entre les deux méthodes a été choisi sur des bases de vitesse de convergence.

La précision obtenue est de l'ordre de 10^{-8} , à condition que soit respecté :

$$z_0 \left(= \frac{x-m}{\sigma} \right) \leq 7$$


```
'REEL' 'PROCEDURE' NORMALE(X,MOY,ECART) FF
'VALEUR' X,MOY,ECART FF 'REEL' X,MOY,ECART FF
'DEBUT' 'ENTIER' I FF
      'REEL' SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
      XF=(X-MOY)/ECART FF
      SGNEF=SIGNE(X) FF
      XF=ABS(X) FF
      X2F=X*X FF
      YF=0.3989422804*EXP(-0.5*X2) FF
      'SI' X 'INFER' 3.5 'ALORS'
      'DEBUT' SF=XF=Y*X FF
            'POUR' IF=3,I+2 'TANTQUE' S 'NONEG' T 'FAIRE'
            'DEBUT' TF=S FF
                  XF=X*X2/I FF
                  SF=S+X
            'FIN' FF
            NORMALEF=0.5+SGNE*S
      'FIN'
      'SINON'
      'DEBUT' Q1F=X FF
            P2F=Y*X FF
            P1F=Y FF
            Q2F=X2+1.0 FF
            MF=Y/X FF
            TF=P2/Q2 FF
            'SI' SIGNE=1 'ALORS'
            'DEBUT' MF=1-M FF
                  TF=1-T
            'FIN' FF
            'POUR' IF=2,I+1 'TANTQUE'
            ((M 'NONEG' T) 'ET' (S 'NONEG' T)) 'FAIRE'
            'DEBUT' SF=X*P2+I*P1 FF
                  P1F=P2 FF
                  P2F=S FF
                  SF=X*Q2+I*Q1 FF
                  Q1F=Q2 FF
                  Q2F=S FF
                  SF=M FF
                  MF=T FF
                  TF=P2/Q2 FF
            'SI' SIGNE=1 'ALORS' TF=1-T FF
            'FIN' FF
            NORMALEF=T
      'FIN' FF
FINORMF 'FIN' FF
```



2 - LOI DE POISSON

2.1. - Introduction

On dit qu'une variable aléatoire X suit une loi de Poisson si elle peut prendre toutes les valeurs entières positives avec les probabilités définies par :

$$p_k = \Pr[X=k] = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k = 0, 1, \dots$$

λ étant un paramètre positif.

2.2. - Théorème

"Soit X une variable aléatoire suivant une loi de Poisson de paramètre λ . Alors la variable aléatoire Y telle que :

$$Y = \frac{X - \lambda}{\sqrt{\lambda}}$$

tend en loi vers une variable aléatoire normale centrée-réduite, lorsque $\lambda \rightarrow \infty$ ".

2.3. - Procédure POISSON

La procédure POISSON calcule la probabilité avec laquelle une variable aléatoire X suivant une loi de Poisson de paramètre λ est inférieure strictement à l'entier k :

$$P(\lambda, k) = e^{-\lambda} \sum_{i=0}^{k-1} \frac{\lambda^i}{i!}$$

Elle possède deux paramètres formels :

LAMBDA : paramètre de la distribution,

K : point où l'on désire connaître la valeur de la fonction de répartition.

Lorsque la valeur du paramètre effectif qui remplace LAMBDA est inférieure à 80 on effectue directement la somme ci-dessus.

Dans le cas contraire il faut recourir à une approximation et on commet une erreur négligeable en assimilant alors la loi de Poisson à une loi normale, vers laquelle elle tend, d'après le théorème qui précède.

On utilise donc :

$$P(\lambda, k) \simeq N\left(\frac{k-\lambda-0.5}{\sqrt{\lambda}}; 0, 1\right)$$

quand le paramètre λ est trop grand pour qu'une évaluation directe de la somme impliquée par $P(\lambda, k)$ soit réalisée.

```
'REEL' 'PROCEDURE' POISSON(LAMBDA,K) FF
'VALEUR' LAMBDA,K FF
'REEL' LAMBDA FF 'ENTIER' K FF
'DEBUT'
'REEL' 'PROCEDURE' NORMALE(X,MOY,ECART) FF
'VALEUR' X,MOY,ECART FF 'REEL' X,MOY,ECART FF
'DEBUT' 'ENTIER' I FF
      'REEL' SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
      XF=(X-MOY)/ECART FF
      SGNEF=SIGNE(X) FF
      XF=ABS(X) FF
      X2F=X*X FF
      YF=0.3989422804*EXP(-0.5*X2) FF
      'SI' X 'INFER' 3.5 'ALORS'
      'DEBUT' SF=XF=Y*X FF
            'POUR' IF=3,I+2 'TANTQUE' S 'NONEG' T 'FAIRE'
            'DEBUT' TF=S FF
                  XF=X*X2/I FF
                  SF=S+X
            'FIN' FF
            NORMALEF=0.5+SGNE*S
      'FIN'
      'SINON'
      'DEBUT' Q1F=X FF
            P2F=Y*X FF
            P1F=Y FF
            Q2F=X2+1.0 FF
            MF=Y/X FF
            TF=P2/Q2 FF
            'SI' SIGNE=1 'ALORS'
            'DEBUT' MF=1-M FF
                  TF=1-T
            'FIN' FF
            'POUR' IF=2,I+1 'TANTQUE'
            ((M 'NONEG' T) 'ET' (S 'NONEG' T)) 'FAIRE'
            'DEBUT' SF=X*P2+I*P1 FF
                  P1F=P2 FF
                  P2F=S FF
                  SF=X*Q2+I*Q1 FF
                  Q1F=Q2 FF
                  Q2F=S FF
                  SF=M FF
                  MF=T FF
                  TF=P2/Q2 FF
            'SI' SIGNE=1 'ALORS' TF=1-T FF
            'FIN' FF
            NORMALEF=T
      'FIN' FF
FINORMF 'FIN' FF
      'REEL' AUX,TOT FF
      'ENTIER' I FF
      'SI' LAMBDA 'SUPER' 80.0 'ALORS'
      'DEBUT' POISSONF=NORMALE((K-0.5-LAMBDA)/
            RAC2(LAMBDA),0.0,1.0) FF
            'ALLERA' FINP
      'FIN' FF
      TOTF=AUXF=1.0 FF
      'POUR' IF=1 'PAS' 1 'JUSQUA' K-1 'FAIRE'
      'DEBUT' AUXF=AUX*LAMBDA/I FF
```

TOTF=TOT+AUX
'FIN' FF
POISSONF=TOT*EXP(-LAMBDA) FF
FINPF 'FIN' FF

3 - LOI BINOMIALE

3.1. - Loi de Bernouilli

Soient p et q deux nombres positifs de somme égale à 1. La variable aléatoire X suit une loi de Bernouilli, si elle ne peut prendre que les valeurs 0 et 1 avec les probabilités respectives :

$$\Pr[X=0] = q$$

$$\Pr[X=1] = p.$$

Cette loi de probabilité est celle de la variable aléatoire égale à l'indicatrice d'un événement donné de probabilité p .

3.2 - Loi Binomiale

Soient X_1, X_2, \dots, X_n n variables aléatoires de Bernouilli indépendantes et de même loi, de paramètres p et q . Par définition la variable aléatoire :

$$Y = X_1 + X_2 + \dots + X_n$$

suit une loi Binomiale notée $B(p, n)$.

On voit aisément que cette loi est répartie sur les entiers $0, 1, \dots, n$ avec les probabilités respectives :

$$\Pr[Y=k] = C_n^k p^k (1-p)^{n-k} \quad k = 0, 1, \dots, n$$

En pratique la loi Binomiale définit la correspondance entre la proportion p de défectueux dans un lot et la probabilité d'en trouver k dans un échantillon à n éléments extraits au hasard du lot, lorsqu'au moins une des deux conditions suivantes est remplies :

- a) chaque élément extrait est remis dans le lot avant le tirage suivant.
- b) L'effectif total N est grand par rapport à l'effectif n de l'échantillon.

3.3 - Théorème central limite

"Soient X_1, \dots, X_n n variables aléatoires indépendantes de même loi ayant une espérance mathématique m et un écart moyen quadratique σ . La loi de la variable aléatoire :

$$\sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - m \right)$$

tend vers la loi $\mathcal{N}(0, \sigma)$, lorsque $n \rightarrow \infty$ ".

3.4 - Procédure BINOMIALE

La procédure BINOMIALE calcule la valeur au point k de la fonction de répartition d'une loi B(p,n) :

$$\Pr[Y < k] = \sum_{i=0}^{k-1} C_n^i p^i (1-p)^{n-i}$$

Le nombre réel p doit respecter la condition :

$$0 < p < 1$$

Pour avoir une valeur de la fonction différente de 0 ou 1 il faut que les deux entiers k, n soient tels que :

$$1 \leq k \leq n+1$$

Nous pouvons écrire

$$\sum_{i=0}^{k-1} C_n^i p^i (1-p)^{n-i} = 1 - \sum_{j=0}^{n-k} C_n^j (1-p)^j p^{n-j}$$

Lorsque la valeur de départ $(1-p)^n$, obtenue pour $i = 0$, est inférieure à e^{-88} , plus petite valeur réelle admise par l'ordinateur, la relation précédente permet de se ramener à une valeur initiale p^n , que l'on obtient pour $j = 0$.

Si les deux valeurs $(1-p)^n$ et p^n sont simultanément inférieure à e^{-88} , ce qui peut se produire lorsque n est grand, il faut recourir à une approximation :

- par une loi normale si le produit npq est supérieur à 20 et nous prenons alors :

$$N \left(\frac{k-np-0.5}{\sqrt{npq}} ; 0,1 \right)$$

d'après le théorème qui précède ;

- par une loi de Poisson sinon et nous utilisons :

$$P(np, k)$$

On peut démontrer que, dans le premier cas, l'approximation est effectuée à 1 % près, alors que dans le second, la valeur absolue de l'erreur est majorée par :

$$\exp\left(\frac{np + \frac{1}{4} + n^3 p^3}{2nq}\right) - 1.$$

```
'REEL' 'PROCEDURE' BINOMIALE(K,N,P) FF
'VALEUR' K,N,P FF
'ENTIER' K,N FF
'REEL' P FF
'DEBUT'
'REEL' 'PROCEDURE' NORMALE(X,MOY,ECART) FF
'VALEUR' X,MOY,ECART FF 'REEL' X,MOY,ECART FF
'DEBUT' 'ENTIER' I FF
'REEL' SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
XF=(X-MOY)/ECART FF
SGNEF=SIGNE(X) FF
XF=ABS(X) FF
X2F=X*X FF
YF=0.3989422804*EXP(-0.5*X2) FF
'SI' X 'INFER' 3.5 'ALORS'
'DEBUT' SF=XF=Y*X FF
'POUR' IF=3,I+2 'TANTQUE' S 'NONEG' T 'FAIRE'
'DEBUT' TF=S FF
XF=X*X2/I FF
SF=S+X
'FIN' FF
NORMALEF=0.5+SGNE*S
'FIN'
'SINON'
'DEBUT' Q1F=X FF
P2F=Y*X FF
P1F=Y FF
Q2F=X2+1.0 FF
MF=Y/X FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS'
'DEBUT' MF=1-M FF
TF=1-T
'FIN' FF
'POUR' IF=2,I+1 'TANTQUE'
((M 'NONEG' T) 'ET' (S 'NONEG' T)) 'FAIRE'
'DEBUT' SF=X*P2+I*P1 FF
P1F=P2 FF
P2F=S FF
SF=X*Q2+I*Q1 FF
Q1F=Q2 FF
Q2F=S FF
SF=M FF
MF=T FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS' TF=1-T FF
'FIN' FF
NORMALEF=T
'FIN' FF
FINORME 'FIN' FF
'REEL' A,TOTA,LAMBDA FF
'ENTIER' I FF
'SI' (K=0) 'OU' (K=N+1) 'ALORS'
'DEBUT' BINOMIALEF=K/(N+1) FF
'ALLERA' FIN
'FIN' FF
'SI' N*LN(1-P) 'INFER' -88 'ALORS'
'DEBUT' 'SI' N*LN(P) 'INFER' -88 'ALORS'
'ALLERA' APPROX 'SINON' 'ALLERA' RETOUR
```

```
'FIN' FF
TOTAF=AF=(1-P)**N FF
'SI' K=1 'ALORS'
'DEBUT' BINOMIALEF=TOTA FF
      'ALLERA' FIN FF
'FIN' FF
'POUR' IF=1 'PAS' 1 'JUSQUA' K-1 'FAIRE'
'DEBUT' AF=A*(N-I+1)*P/(I*(1-P)) FF
      TOTAF=TOTA+A FF
'FIN' FF
BINOMIALEF=TOTA FF
'ALLERA' FIN FF
RETOURF TOTAF=AF=P**N FF
'POUR' IF=N 'PAS' -1 'JUSQUA' K+1 'FAIRE'
'DEBUT' AF=A*I*(1-P)/((N-I+1)*P) FF
      TOTAF=TOTA+A FF
'FIN' FF
BINOMIALEF=1-TOTA FF
'ALLERA' FIN FF
APPROXF 'SI' N*P 'INFER' 20.0 'ALORS' 'ALLERA' POISSON FF
AF=(K-0.5-N*P)/RAC2(N*P*(1-P)) FF
BINOMIALEF=NORMALE(A,0.0,1.0) FF
'ALLERA' FIN FF
POISSONFLAMBDAF=N*P FF
'SI' K=1 'ALORS'
'DEBUT' BINOMIALEF=1/EXP(LAMBDA) FF
      'ALLERA' FIN FF
'FIN' FF
TOTAF=AF=1 FF
'POUR' IF=1 'PAS' 1 'JUSQUA' K-1 'FAIRE'
'DEBUT' AF=A*LAMBDA/I FF
      TOTAF=TOTA+A FF
BINOMIALEF=TOTA/EXP(LAMBDA) FF
'FIN' FF
FINE'FIN' FF
```


4 - LOI DU χ^2

4.1 - Introduction

Etant données X_1, \dots, X_n n variables aléatoires indépendantes suivant toutes la loi normale centrée réduite on appelle loi du χ^2 à n degrés de liberté la loi de la variable aléatoire :

$$Z = X_1^2 + \dots + X_n^2$$

La fonction densité de probabilité de Z est alors :

$$q_n(z) = \frac{1}{2^{n/2} \Gamma(n/2)} z^{n/2 - 1} e^{-z/2}, \quad z \geq 0.$$

4.2 - Procédure KI

La procédure KI calcule la valeur au point x de la fonction de répartition d'une loi du χ^2 à n degrés de liberté :

$$Q_n(x) = \frac{1}{2^{n/2} \Gamma(n/2)} \int_0^x z^{n/2 - 1} e^{-z/2} dz$$

Elle utilise la relation de récurrence :

$$Q_n(x) = Q_{n-2}(x) - e^{-x/2} \frac{x^{n/2 - 1}}{2^{n/2 - 1} \Gamma(n/2)}, \quad n > 2$$

avec :

$$Q_2(x) = 1 - e^{-x/2}$$

et :

$$Q_1(x) = \frac{2}{\sqrt{\pi}} \int_0^{\sqrt{x/2}} e^{-z^2} dz = 2.N(\sqrt{x}; 0,1) - 1$$

Remarque

Cette loi aurait pu être obtenue également comme cas particulier de GAMDEC, procédure qui sera introduite au paragraphe 7. Le calcul particulier ci-dessus permet une plus grande précision.

```
'REEL' 'PROCEDURE' KI(N,X) FF
'VALEUR' N,X FF
'ENTIER' N FF 'REEL' X FF
'DEBUT'
'REEL' 'PROCEDURE' NORMALE(X,MOY,ECART) FF
'VALEUR' X,MOY,ECART FF 'REEL' X,MOY,ECART FF
'DEBUT' 'ENTIER' I FF
'REEL' SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
XF=(X-MOY)/ECART FF
SGNEF=SIGNE(X) FF
XF=ABS(X) FF
X2F=X*X FF
YF=0.3989422804*EXP(-0.5*X2) FF
'SI' X 'INFER' 3.5 'ALORS'
'DEBUT' SF=XF=Y*X FF
'POUR' IF=3,I+2 'TANTQUE' S 'NONEG' T 'FAIRE'
'DEBUT' TF=S FF
XF=X*X2/I FF
SF=S+X
'FIN' FF
NORMALEF=0.5+SGNE*S
'FIN'
'SINON'
'DEBUT' Q1F=X FF
P2F=Y*X FF
P1F=Y FF
Q2F=X2+1.0 FF
MF=Y/X FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS'
'DEBUT' MF=1-M FF
TF=1-T
'FIN' FF
'POUR' IF=2,I+1 'TANTQUE'
((M 'NONEG' T) 'ET' (S 'NONEG' T)) 'FAIRE'
'DEBUT' SF=X*P2+I*P1 FF
P1F=P2 FF
P2F=S FF
SF=X*Q2+I*Q1 FF
Q1F=Q2 FF
Q2F=S FF
SF=M FF
MF=T FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS' TF=1-T FF
'FIN' FF
NORMALEF=T
'FIN' FF
FINORMF 'FIN' FF
'ENTIER' I FF
'REEL' AUX,SOM,PI FF
'SI' N-2*ENTIER(N/2)=1 'ALORS'
'ALLERA' NIMP FF
SOMF=AUXF=1.0 FF
'SI' N=2 'ALORS' 'ALLERA' FPAIR FF
'POUR' IF=1 'PAS' 1 'JUSQUA' N/2-1 'FAIRE'
'DEBUT' AUXF=AUX*X/(2*I) FF
SOMF=SOM+AUX
'FIN' FF
```

FPAIRF KIF=1.0-EXP(-X/2)*SOM FF
'ALLERA' FINKI FF

NIMPF PIF=3.141592653590 FF
AUXF=RAC2(2/(PI*X)) FF
SOMF=0.0 FF
'SI' N=1 'ALORS' 'ALLERA' FNIMP FF
'POUR' IF=0 'PAS' 1 'JUSQUA' (N-3)/2 'FAIRE'
'DEBUT' AUXF=AUX*X/(2*I+1) FF
SOMF=SOM+AUX
'FIN' FF

FNIMPF KIF=2*NORMALE(RAC2(X),0.0,1.0)-1.0-EXP(-X/2)*SOM FF
FINKIF'FIN' FF

5 - LOI DE STUDENT

5.1 - Introduction

Soient X une variable aléatoire normale centrée réduite et S une variable aléatoire du χ^2 à n degrés de liberté, X et S étant indépendantes; on appelle loi de Student à n degrés de liberté la loi de la variable aléatoire :

$$T = \sqrt{n} \frac{X}{\sqrt{S}}$$

T admet pour densité de probabilité :

$$s_n(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi} \Gamma(\frac{n}{2})} \frac{1}{(1 + \frac{t^2}{n})^{\frac{n+1}{2}}}, \quad t \in \mathbb{R}.$$

5.2 - Procédure STUDENT

La procédure STUDENT calcule la valeur au point u de la fonction de répartition d'une loi de Student à n degrés de liberté :

$$S_n(u) = \int_{-\infty}^u s_n(t) dt$$

Le changement de variable $x = \frac{t}{\sqrt{n}}$ permet d'écrire :

$$S_n(u) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{1}{2})\Gamma(\frac{n}{2})} \int_{-\infty}^{\frac{u}{\sqrt{n}}} \frac{dx}{(1+x^2)^{\frac{n+1}{2}}}$$

Si l'on introduit maintenant les fonctions $I_m(v)$ de la variable réelle v, indicées par m entier ou demi-entier, et telles que :

$$I_m(v) = \int_{-\infty}^v \frac{dx}{(1+x^2)^m}$$

on peut établir la relation suivante :

$$I_m(v) = \frac{2m-3}{2m-2} I_{m-1}(v) + \frac{1}{2m-2} \frac{v}{(1+v^2)^{m-1}}, \quad m \geq 2.$$

Deux cas peuvent alors se produire :

a) m est entier

La relation précédente permet d'aboutir à $I_1(v)$:

$$I_1(v) = \int_{-\infty}^v \frac{dx}{1+x^2} = \text{Arctg } v + \frac{\pi}{2}$$

b) m est demi-entier

L'évaluation de $I_m(v)$ se ramène alors à celle de $I_{\frac{3}{2}}(v)$:

$$I_{\frac{3}{2}}(v) = \int_{-\infty}^v \frac{dx}{(1+x^2)^{3/2}} = 1 + \frac{v}{\sqrt{1+v^2}}$$

La méthode que nous avons utilisée pour le calcul de $S_n(u)$ consiste à employer de telles fonctions, puisque l'on peut écrire :

$$S_n(u) = \frac{I_{\frac{n+1}{2}}\left(\frac{u}{\sqrt{n}}\right)}{2}$$

On partira alors si n est pair de $I_{\frac{3}{2}}$, sinon de I_1 .

```
'REEL' 'PROCEDURE' STUDENT(N,U) FF
'VALEUR' N,U FF 'ENTIER' N FF 'REEL' U FF
'DEBUT'
  'REEL' 'PROCEDURE' GAMMA(X) FF 'REEL' XFF
  'DEBUT' 'REEL' H,Y FF
  HF=1.0 FF YF=X FF
  AIF 'SI' Y=0 'ALORS' HF=10**5 'SINON'
  'SI' Y=2.0 'ALORS' 'ALLERA' A2 'SINON'
  'SI' Y'INFER'2.0 'ALORS'
  'DEBUT' HF=H/Y FF YF=Y+1.0 FF 'ALLERA' A1 'FIN'
  'SINON' 'SI' Y'SUPEG'3 'ALORS' 'DEBUT'
  YF=Y-1.0 FF HF=H*Y FF 'ALLERA' A1 'FIN'
  'SINON' 'DEBUT' YF=Y-2.0 FF
  HF=((( ((( (( (.0016063118*Y+.0051589951)*Y+.0044511400)*Y+.0721101567)*
  Y+.0821117404)*Y+.4117741955)*Y
  .4227874605)*Y+.9999999758)*H
  'FIN' FF
  A2F GAMMAF=H
  'FIN' GAMMA FF
  'REEL' V,I,A FF
  'ENTIER' M FF
  VF=U/RAC2(N) FF
  'SI' N-2*ENTIER(N/2)=1 'ALORS' 'ALLERA' NIMP FF
  AF=V/RAC2(1+V**2) FF
  IF=1+A FF
  'POUR' MF=1 'PAS' 1 'JUSQUA' N/2-1 'FAIRE'
  'DEBUT' AF=A/(1+V**2) FF
  IF=(2*M*I+A)/(2*M+1) 'FIN' FF
  'ALLERA' FIN FF
NIMPF IF=ARCTAN(V)+1.570796326795 FF
AF=V FF
'POUR' MF=1 'PAS' 1 'JUSQUA' (N-1)/2 'FAIRE'
'DEBUT' AF=A/(1+V**2) FF
IF=((2*M-1)*I+A)/(2*M) 'FIN' FF
STUDENTF='SI' N 'INFER' 60 'ALORS' I*GAMMA((N+1)/2)/
(GAMMA(N/2)*GAMMA(0.5)) 'SINON' I*RAC2(N/6.28318530)*EXP(-0.25/N
1/(24*N**3)) FF
'FIN' FF
```


6 - LOI DU χ^2 DÉCENTRÉE

6.1 - Introduction

Soient X_1, X_2, \dots, X_n des variables aléatoires indépendantes suivant respectivement les lois $(m_i, 1)$. Alors la variable aléatoire Z telle que :

$$Z = X_1^2 + X_2^2 + \dots + X_n^2$$

est dite suivre une loi du χ^2 décentrée à n degrés de liberté.

On appelle

$$\lambda = m_1^2 + m_2^2 + \dots + m_n^2$$

le paramètre de décentrage de la distribution.

Remarque 1 :

La distribution du χ^2 réduite, ou centrée, est un cas particulier de la distribution décentrée. Elle est obtenue pour un paramètre de décentrage λ égal à zéro.

Remarque 2 :

Dans la littérature le paramètre de décentrage est quelquefois défini différemment. Certains auteurs utilisent $\delta = \lambda^{1/2}$, alors que d'autres préfèrent $\delta = \frac{1}{2} \lambda$.

La fonction densité d'une telle variable aléatoire est obtenue à l'aide du théorème suivant [2].

6.2 - Théorème

"La fonction où λ est positif ou nul :

$$(1-2it)^{-\frac{n}{2}} \exp\left(\frac{it\lambda}{1-2it}\right)$$

est fonction caractéristique de la loi de probabilité sur la demi-droite positive
densité :

$$f(z) = 2^{-n/2} z^{n/2-1} e^{-1/2(z+\lambda)} \sum_{k=0}^{\infty} \frac{(\lambda z)^k}{2^{2k} k! \Gamma(\frac{n}{2}+k)}$$

De plus si Z est une variable aléatoire suivant cette loi, elle admet la représentation :

$$Z = X_1^2 + X_2^2 + \dots + X_n^2$$

où les X_i sont des variables aléatoires indépendantes de lois respectives $(m_i, 1)$ sous la seule condition :

$$\lambda = m_1^2 + m_2^2 + \dots + m_n^2."$$

6.3 - Procédure KIDECENT

La procédure KIDECENT calcule la valeur au point x de la fonction de répartition d'une loi du χ^2 décentrée à n degrés de liberté et de paramètre de décentrage p , c'est-à-dire :

$$K_{n,p}(x) = \int_0^x k_{n,p}(z) dz, \quad x > 0.$$

en introduisant la fonction densité de probabilité $k_{n,p}(z)$ telle que :

$$k_{n,p}(z) = 2^{-n/2} z^{n/2-1} e^{-1/2(z+p)} \sum_{i=0}^{\infty} \frac{(pz)^i}{2^{2i} i! \Gamma(\frac{n}{2}+i)}$$

On peut encore écrire

$$K_{n,p}(x) = e^{-p/2} \sum_{i=0}^{\infty} \left[\frac{(p/2)^i}{i! \Gamma(\frac{n}{2}+i)} \int_0^{x/2} t^{n/2+i-1} e^{-t} dt \right]$$

Posons alors :

$$K_{n,p}(x) = e^{-p/2} \sum_{i=0}^{\infty} c_i \cdot d_i$$

avec

$$c_i = \frac{p^i}{2^i i!}$$

$$d_i = \frac{1}{\Gamma(\frac{n}{2}+i)} \int_0^{x/2} t^{n/2+i-1} e^{-t} dt$$

Le problème revient à calculer, pour tout indice i , les trois nombres c_i, a_i, d_i tels que :

$$c_i = c_{i-1} \frac{p}{2i}$$

$$a_i = a_{i-1} \frac{x}{2(\frac{n}{2} + i - 1)}$$

$$d_i = d_{i-1} - a_i$$

sachant que :

$$c_0 = 1$$

$$a_0 = \frac{e^{-x/2} x^{n/2-1}}{2^{n/2-1} \Gamma(\frac{n}{2})}$$

$$d_0 = Q_n(x)$$

Ici $Q_n(t)$ représente la fonction de répartition de la loi du χ^2 réduite à n degrés de liberté.

Avec les notations que nous venons d'introduire on peut écrire :

$$K_{n,p}(x) = e^{-p/2} \sum_{i=0}^N c_i \cdot d_i + R_N$$

où :

$$R_N = e^{-p/2} \sum_{i=N+1}^{\infty} \left[\frac{p^i}{2^i i!} \frac{1}{\Gamma(\frac{n}{2} + i)} \int_0^{x/2} t^{n/2+i-1} e^{-t} dt \right]$$

Il est facile de trouver une majoration de R_N puisque :

$$\frac{1}{\Gamma(\frac{n}{2} + i)} \int_0^{x/2} t^{n/2+i-1} e^{-t} dt$$

représente la valeur au point $\frac{x}{2}$ de la fonction de répartition d'une loi Gamma, et comme telle, est inférieure ou égale à 1, par définition :

On a alors :

$$R_N \leq e^{-p/2} \sum_{i=N+1}^{\infty} \frac{p^i}{2^i i!}$$

on encore

$$R_N \leq 1 - e^{-p/2} \sum_{i=0}^N c_i$$

6.4 - Exemple numérique

Il existe plusieurs applications de la loi du χ^2 décentrée en statistique. Nous allons considérer l'une d'elles et montrer comment une telle distribution intervient dans l'étude des fonctions puissance de certains tests.

Soient n observations x_1, \dots, x_n indépendantes. Si nous faisons l'hypothèse H_0 : les x_i ont été tirées d'une population normale centrée réduite, alors, si H_0 est vraie, la statistique s telle que :

$$s = \sum_{i=1}^n x_i^2$$

sera supérieure à $Q_n^{-1}(1-\alpha)$ avec une probabilité α représentant le niveau de signification du test.

La puissance d'un tel test est donnée par la probabilité pour que s soit supérieure à $Q_n^{-1}(1-\alpha)$ sous d'autres hypothèses.

Nous pouvons par exemple supposer (hypothèse H_1) que les x_i proviennent de populations normales de moyennes respectives a_i différentes et de variance unitaire.

Dans ce cas s obéira à une loi du χ^2 décentrée à n degrés de liberté et de paramètre de décentrage λ tel que :

$$\lambda = \sum_{i=1}^n a_i^2$$

La fonction puissance est alors donnée par :

$$P = 1-\beta = \int_{Q_n^{-1}(1-\alpha)}^{\infty} f_{n,\lambda}(z) dz$$

Elle ne dépend que du seul paramètre λ . Ainsi l'hypothèse simple H_0 est obtenue pour $\lambda = 0$ alors que H_1 est une hypothèse composée qui comprend la famille de toutes les hypothèses simples pour lesquelles :

$$\sum_{i=1}^n a_i^2 = \lambda$$

Les tables relatives à ce test sont les suivantes :

- Fix [9] donne les valeurs de λ correspondant à

$$\beta = 0.1 (0.01) 0.9$$

$$\alpha = 0.01 \text{ et } 0.05$$

$$n = 1(1) 20 (2) 40 (5) 60 (10) 100$$

- Patnaik [18] donne les valeurs de P correspondant à

$$\lambda = 2 (2) 20$$

$$\alpha = 0.05$$

$$n = 2 (1) 10 (2) 20$$

Ces deux tables ont, pour $\alpha = 0.05$, une large zone commune et la seconde complète la première pour λ supérieur à 20. Mais il reste un trou gênant qui, pour $\alpha = 0.01$, correspond à des valeurs de P supérieures à 0.90.

Le calcul d'une table du type de celle de Patnaik paraît donc intéressant. Nous l'avons effectué en utilisant la procédure KIDECENT.

On donnera, à la page suivante, les valeurs de P qui correspondent à :

$$\alpha = 0.01$$

$$\lambda = 12 (2) 20 (5) 40$$

$$n = 2 (2) 10 (5) 30$$

VALEURS DE P EN FONCTION DE n ET λ POUR $\alpha = 0.01$

n \ λ	12	14	16	18	20	25	30	35	40
2	0.721	0.804	0.867	0.911	0.942	0.982	0.995	0.999	0.99
4	0.598	0.695	0.774	0.837	0.886	0.956	0.985	0.995	0.99
6	0.513	0.612	0.700	0.773	0.832	0.927	0.972	0.990	0.99
8	0.450	0.547	0.637	0.715	0.782	0.897	0.956	0.983	0.99
10	0.400	0.494	0.583	0.664	0.735	0.865	0.938	0.974	0.99
15	0.313	0.396	0.480	0.561	0.636	0.789	0.889	0.947	0.97
20	0.257	0.330	0.406	0.483	0.557	0.720	0.838	0.914	0.95
25	0.218	0.282	0.351	0.422	0.494	0.658	0.788	0.879	0.93
30	0.189	0.246	0.309	0.374	0.441	0.603	0.740	0.842	0.91

```
'REEL' 'PROCEDURE' KIDECENT(N,P,X) FF
'VALEUR' N,P,X FF
'ENTIER' N FF 'REEL' P,X FF
'DEBUT'
  'REEL' 'PROCEDURE' GAMMA(X) FF 'REEL' XFF
  'DEBUT' 'REEL' H,Y FF
  HF=1.0 FF YF=X FF
  A1F 'SI' Y=0 'ALORS' HF=10**5 'SINON'
  'SI' Y=2.0 'ALORS' 'ALLERA' A2 'SINON'
    'SI' Y<INFER'2.0 'ALORS'
      'DEBUT' HF=H/Y FF YF=Y+1.0 FF 'ALLERA' A1 'FIN'
      'SINON' 'SI' Y<SUPEG'3 'ALORS' 'DEBUT'
        YF=Y-1.0 FF HF=H*Y FF 'ALLERA' A1 'FIN'
        'SINON' 'DEBUT' YF=Y-2.0 FF
        HF=(((((((0.0016063118*Y+.0051589951)*Y+.0044511400)*Y+.0721101567)*
        Y+.0821117404)*Y+.4117741955)*Y
        .4227874605)*Y+.9999999758)*H
        'FIN' FF
  A2F GAMMAF=H
  'FIN' GAMMA FF
'REEL' 'PROCEDURE' NORMALE(X,MOY,ECART) FF
'VALEUR' X,MOY,ECART FF 'REEL' X,MOY,ECART FF
'DEBUT' 'ENTIER' I FF
  'REEL' SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
  XF=(X-MOY)/ECART FF
  SGNEF=SIGNE(X) FF
  XF=ABS(X) FF
  X2F=X*X FF
  YF=0.3989422804*EXP(-0.5*X2) FF
  'SI' X<INFER'3.5 'ALORS'
    'DEBUT' SF=XF*Y*X FF
      'POUR' IF=3,I+2 'TANTQUE' S<NONEG' T<FAIRE'
        'DEBUT' TF=S FF
          XF=X*X2/I FF
          SF=S+X
        'FIN' FF
      NORMALEF=0.5+SGNE*S
    'FIN'
  'SINON'
    'DEBUT' Q1F=X FF
      P2F=Y*X FF
      P1F=Y FF
      Q2F=X2+1.0 FF
      MF=Y/X FF
      TF=P2/Q2 FF
      'SI' SIGNE=1 'ALORS'
        'DEBUT' MF=1-M FF
          TF=1-T
        'FIN' FF
      'POUR' IF=2,I+1 'TANTQUE'
      ((M<NONEG' T) 'ET' (S<NONEG' T)) 'FAIRE'
        'DEBUT' SF=X*P2+I*P1 FF
          P1F=P2 FF
          P2F=S FF
          SF=X*Q2+I*Q1 FF
          Q1F=Q2 FF
          Q2F=S FF
          SF=M FF
```

MF=T FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS' TF=1-T FF

'FIN' FF
NORMALEF=T

```

'FIN' FF
FINORMF 'FIN' FF
'REEL' 'PROCEDURE' KI(N,X) FF
'VALEUR' N,X FF
'ENTIER' N FF 'REEL' X FF
'DEBUT' 'ENTIER' I FF
'REEL' AUX,SOM,PI FF
'SI' N-2*ENTIER(N/2)=1 'ALORS'
'ALLERA' NIMP FF
SOMF=AUXF=1.0 FF
'SI' N=2 'ALORS' 'ALLERA' FPAIR FF
'POUR' IF=1 'PAS' 1 'JUSQUA' N/2-1 'FAIRE'
'DEBUT' AUXF=AUX*X/(2*I) FF
SOMF=SOM+AUX
'FIN' FF
FPAIRF KIF=1.0-EXP(-X/2)*SOM FF
'ALLERA' FINKI FF
NIMPF PIF=3.141592653590 FF
AUXF=RAC2(2/(PI*X)) FF
SOMF=0.0 FF
'SI' N=1 'ALORS' 'ALLERA' FNIMP FF
'POUR' IF=0 'PAS' 1 'JUSQUA' (N-3)/2 'FAIRE'
'DEBUT' AUXF=AUX*X/(2*I+1) FF
SOMF=SOM+AUX
'FIN' FF
FNIMPF KIF=2*NORMALE(RAC2(X),0.0,1.0)-1.0-EXP(-X/2)*SOM FF
FINKIF 'FIN' FF
'REEL' L,Z,I,A,C,T,R FF
'ENTIER' K,NOR FF
LF=P/2 FF ZF=X/2 FF
NORF=0 FF
AUGF NORF=NOR+1 FF
'SI' (N/2-1)*LN(Z/NOR) 'SUPER' 88.0 'ALORS' 'ALLERA' AUG FF
KF=0 FF
CF=EXP(-L) FF
RF=1-C FF
AF=EXP(-Z)*(Z/NOR)**(N/2-1)*NOR**(N/2-1)/GAMMA(N/2) FF
IF=KI(N,X) FF
TF=I*C FF
BOUCLEF KF=K+1 FF
AF=A*Z/(N/2+K-1) FF
IF=I-A FF
CF=C*L/K FF
TF=T+I*C FF
RF=R-C FF
'SI' R 'SUPER' F*-5 'ALORS' 'ALLERA' BOUCLE FF
FINKIDECFKIDECENTF=T FF
'FIN' FF

```


7 - LOI GAMMA DECENTREE

7.1 - Introduction

La loi du χ^2 décentrée, que nous avons examinée au paragraphe 5, est un cas particulier de la loi Gamma décentrée que nous introduisons maintenant.

Si une variable aléatoire Z a pour fonction densité de probabilité :

$$\gamma(a,p;t) = e^{-p-t} t^{a-1} \sum_{m=0}^{\infty} \frac{(pt)^m}{m! \Gamma(a+m)}$$

où a,p,t sont des nombres réels tels que :

$$a \geq 1$$

$$p,t \geq 0$$

nous dirons alors qu'elle suit une loi Gamma décentrée avec a degrés de liberté et un paramètre de décentrage égal à p.

7.2 - Procédure GAMDEC

La procédure GAMDEC effectue le calcul de la valeur au point z de la fonction de répartition d'une loi Gamma décentrée, soit :

$$G(a,p;z) = \int_0^z \gamma(a,p;t) dt$$

Elle possède trois paramètres formels :

A : degré de liberté

P : paramètre de décentrage

Z : point où l'on désire connaître la valeur de la fonction de répartition.

Remarque 1 :

La relation qui existe entre les deux procédures KIDECENT et GAMDEC est immédiate puisque :

$$K_{n,\lambda}(x) = G\left(\frac{n}{2}, \frac{\lambda}{2}; \frac{x}{2}\right)$$

en utilisant la notation introduite au paragraphe 6.3.

L'intérêt de conserver KIDECENT est cependant certain : le fait que le paramètre n soit entier permet d'obtenir une plus grande précision en l'employant de préférence à GAMDEC pour le calcul de la fonction de répartition d'une variable aléatoire du χ^2 décentrée.

Remarque 2 :

La loi Gamma pour $2a$ entier est également obtenue à l'aide de GAMDEC.

Elle correspond au cas $p = 0$.

Il résulte de la définition de $\gamma(a,p ; t)$ donnée en 7.1 que la fonction G peut s'écrire sous la forme :

$$G(a,p ; z) = e^{-p} \sum_{m=0}^{\infty} \left[\frac{p^m}{m! \Gamma(a+m)} \int_0^z t^{a+m-1} e^{-t} dt \right]$$

ou encore

$$G(a,p ; z) = \sum_{m=0}^{\infty} b_m \cdot i_m$$

en posant :

$$b_m = e^{-p} \frac{p^m}{m!}, \quad m \geq 0$$

$$i_m = \frac{1}{\Gamma(a+m)} \int_0^z t^{a+m-1} e^{-t} dt, \quad m \geq 0.$$

On peut alors introduire les relations de récurrence suivantes :

$$\text{I} \quad \left| \begin{array}{l} b_m = b_{m-1} \frac{p}{m}, \quad m \geq 1 \\ b_0 = e^{-p} \end{array} \right.$$

$$\text{II} \quad \left| \begin{array}{l} i_m = i_{m-1} - \frac{1}{\Gamma(a+m)} e^{-z} z^{a+m-1}, \quad m \geq 1 \\ i_0 = \frac{1}{\Gamma(a)} \int_0^z t^{a-1} e^{-t} dt \end{array} \right.$$

II pouvant être remplacé par II' :

$$\text{II}' \quad \left\{ \begin{array}{l} c_m = c_{m-1} \cdot \frac{z}{a+m-1}, \quad m \geq 1 \\ c_0 = \frac{1}{\Gamma(a)} e^{-z} z^{a-1} \\ i_m = i_{m-1} - c_m, \quad m \geq 1 \\ i_0 = \frac{1}{\Gamma(a)} \int_0^z t^{a-1} e^{-t} dt \end{array} \right.$$

Du point de vue numérique la seule difficulté réside dans l'évaluation de i_0 . Nous pouvons alors établir que :

$$i_0 = z^a e^{-z} \sum_{j=0}^N \frac{z^j}{\Gamma(a+j+1)} + R_N$$

avec

$$R_N = \frac{1}{\Gamma(a+N+1)} \int_0^z e^{-t} t^{a+N} dt$$

Nous avons posé en 7.1 que t ne pouvait prendre que des valeurs positives. Dans de telles conditions la quantité e^{-t} est majorée par 1 et on a :

$$R_N \leq \frac{1}{\Gamma(a+N+1)} \int_0^z t^{a+N} dt$$

ce qui peut encore s'écrire :

$$R_N \leq \frac{z^{a+N+1}}{\Gamma(a+N+2)}$$

Lors de l'évaluation de la valeur de départ i_0 nous négligerons le terme R_N , lorsque l'indice N sera tel que :

$$\frac{z^{a+N+1}}{\Gamma(a+N+2)} \leq 10^{-8}$$

7.3 - Approximation

Le programme général GAMDEC utilise le sous-programme GAMMA qui calcule la fonction Γ . Or les raisonnements qui précèdent ne sont valables que dans le cas où cette fonction est calculable, c'est-à-dire, pour l'ordinateur IBM-7044, lorsque a ne dépasse pas 30. Si cette condition n'est pas vérifiée, il faut recourir à une approximation.

Nous savons que si une variable aléatoire Y suit une loi Gamma de paramètre a et que, par conséquent, sa fonction de répartition est :

$$F(y) = \frac{1}{\Gamma(a)} \int_0^y t^{a-1} e^{-t} dt$$

on peut, lorsque a est grand, considérer que la variable aléatoire :

$$\sqrt{4Y} - \sqrt{4a-1}$$

suit une loi normale centrée réduite.

Lorsque l'indice m est tel que :

$$a+m > 30$$

nous approximos i_m par :

$$N(\sqrt{4z} - \sqrt{4(a+m)-1} ; 0,1)$$

```

'REEL' 'PROCEDURE' GAMDEC(P,A,Z) FF
'VALEUR' P,A,Z FF
'REEL' P,A,Z FF
'DEBUT'
'REEL' 'PROCEDURE' NORMALE(X,MOY,ECART) FF
'VALEUR' X,MOY,ECART FF 'REEL' X,MOY,ECART FF
'DEBUT' 'ENTIER' I FF
'REEL' SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
XF=(X-MOY)/ECART FF
SGNEF=SIGNE(X) FF
XF=ABS(X) FF
X2F=X*X FF
YF=0.3989422804*EXP(-0.5*X2) FF
'SI' X 'INFER' 3.5 'ALORS'
'DEBUT' SF=XF=Y*X FF
'POUR' IF=3,I+2 'TANTQUE' S 'NONEG' T 'FAIRE'
'DEBUT' TF=S FF
XF=X*X2/I FF
SF=S+X
'FIN' FF
NORMALEF=0.5+SGNE*S
'FIN'
'SINON'
'DEBUT' Q1F=X FF
P2F=Y*X FF
P1F=Y FF
Q2F=X2+1.0 FF
MF=Y/X FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS'
'DEBUT' MF=1-M FF
TF=1-T
'FIN' FF
'POUR' IF=2,I+1 'TANTQUE'
((M 'NONEG' T) 'ET' (S 'NONEG' T)) 'FAIRE'
'DEBUT' SF=X*P2+I*P1 FF
P1F=P2 FF
P2F=S FF
SF=X*Q2+I*Q1 FF
Q1F=Q2 FF
Q2F=S FF
SF=M FF
MF=T FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS' TF=1-T FF
'FIN' FF
NORMALEF=T
'FIN' FF
FINORMF 'FIN' FF

```

```

'REEL' 'PROCEDURE' GAMMA(X) FF 'REEL' XFF
'DEBUT' 'REEL' H,Y FF
HF=1.0 FF YF=X FF
A1F 'SI' Y=0 'ALORS' HF=10**5 'SINON'
'SI' Y=2.0 'ALORS' 'ALLERA' A2 'SINON'
'SI' Y'INFER'2.0 'ALORS'
'DEBUT' HF=H/Y FF YF=Y+1.0 FF 'ALLERA' A1 'FIN'
'SINON' 'SI' Y'SUPEG'3 'ALORS' 'DEBUT'
YF=Y-1.0 FF HF=H*Y FF 'ALLERA' A1 'FIN'
'SINON' 'DEBUT' YF=Y-2.0 FF
HF=((( ((( (0.0016063118*Y+.0051589951)*Y+.0044511400)*Y+.0721101567)*

```

```
Y+.0821117404)*Y+.4117741955)*Y
.4227874605)*Y+.9999999758)*H
'FIN' FF
A2F GAMMAF=H
'FIN' GAMMA FF
'ENTIER' M,K,NOR FF
'REEL' I,I0,C,B,D,TOT,TOT1,R FF
NORF=0 FF
AUGF NORF=NOR+1 FF
'SI' (A-1)*LN(Z/NOR) 'SUPER' 88.0 'ALORS' 'ALLERA' AUG FF
'DEBUT' MF=0 FF
TOTF=0.0 FF
BF=EXP(-P) FF
'SI' A 'SUPER' 30.0 'ALORS' 'ALLERA' APPROX FF
DF=GAMMA(A) FF
KF=0 FF
IF=IOF=(Z/NOR)**A*NOR**A*EXP(-Z)/(A*D) FF
RF=(Z/NOR)**(A+1)*NOR**(A+1)/(A*(A+1)*D) FF
RETOURF KF=K+1 FF
IOF=I0*Z/(A+K) FF
RF=R*Z/(A+K+1) FF
IF=I+IO FF
'SI' R 'SUPER' F*-8 'ALORS' 'ALLERA' RETOUR FF
CF=EXP(-Z)*(Z/NOR)**(A-1)*NOR**(A-1)/D FF
TOTF=B*I FF
BOUCLEF MF=M+1 FF
BF=B*P/M FF
'SI' A+M 'SUPER' 30.0 'ALORS' 'ALLERA' APPROX FF
CF=C*Z/(A+M-1) FF
IF=I-C FF
TOT1F=TOT FF
TOTF=TOT+B*I FF
'SI' TOT 'NONEG' TOT1 'ALORS' 'ALLERA' BOUCLE FF
'ALLERA' FINGAM FF
APPROXF IF=NORMALE(RAC2(4*Z)-RAC2(4*(A+M)-1),0.0,1.0) FF
TOT1F=TOT FF
TOTF=TOT+B*I FF
'SI' TOT 'NONEG' TOT1 'ALORS'
'DEBUT' MF=M+1 FF
BF=B*P/M FF
'ALLERA' APPROX FF
'FIN' FF
'FIN' FF
FINGAMF GAMDECF=TOT
'FIN' FF
```

8 - LOI BETA DECENTREE

8.1 - Introduction

La variable aléatoire Z est dite suivre une loi Beta de paramètres a,b et p si elle a pour fonction densité de probabilité :

$$\beta(a,b,p ; t) = e^{-p} \sum_{m=0}^{\infty} \frac{p^m}{m!} \frac{\Gamma(a+b+m)}{\Gamma(a+m)\Gamma(b)} \frac{t^{a+m-1}}{(1+t)^{a+b+m}}$$

où les nombres réels a,b,p,t sont tels que :

$$a,b \geq 1$$

$$p,t \geq 0$$

8.2 - Procédure BETADEC

La procédure BETADEC effectue le calcul de la fonction :

$$BD(a,b,p ; z) = \int_0^z \beta(a,b,p ; t) dt$$

que l'on peut encore écrire :

$$BD(a,b,p ; z) = \sum_{m=0}^{\infty} d_m \cdot i_m$$

en posant

$$d_m = e^{-p} \frac{p^m}{m!}, \quad m > 0$$

$$i_m = \frac{\Gamma(a+b+m)}{\Gamma(a+m)\Gamma(b)} \int_0^z \frac{t^{a+m-1}}{(1+t)^{a+b+m}} dt, \quad m \geq 0$$

La méthode de calcul que nous avons employée nécessite l'introduction des trois relations :

$$\text{I} \quad \left| \begin{array}{l} d_m = d_{m-1} \frac{p}{m}, \\ d_0 = e^{-p} \end{array} \right. \quad m \geq 1$$

$$\text{II} \quad \left| \begin{array}{l} c_m = c_{m-1} \cdot \frac{a+b+m-2}{a+m-1} \cdot \frac{z}{1+z}, \\ c_0 = \frac{\Gamma(a+b-1)}{\Gamma(a)\Gamma(b)} \frac{z^{a-1}}{(1+z)^{a+b-1}} \end{array} \right. \quad m \geq 1$$

$$\text{III} \quad \left| \begin{array}{l} i_m = i_{m-1} - c_m, \\ i_0 = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^z \frac{t^{a-1}}{(1+t)^{a+b}} dt \end{array} \right. \quad m \geq 1$$

Du point de vue numérique il ne reste qu'à calculer i_0 , ce qui pose quelques difficultés par le fait même que les deux paramètres a et b ne sont pas nécessairement entiers.

Nous avons utilisé un sous-programme du à O. Ludwig [14] qui évalue le ratio :

$$\frac{B_x(p,q)}{B_1(p,q)}$$

où :

$$B_x(p,q) = \int_0^x t^{p-1} (1-t)^{q-1} dt$$

dans le cas :

$$0 < x < 1$$

$$p, q > 0 \text{ mais quelconques}$$

ceci avec une précision au moins égale à ϵ , ϵ étant choisie par l'utilisateur.

On peut en effet écrire :

$$\begin{aligned} i_0 &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^{\frac{z}{1+z}} t^{a-1} (1-t)^{b-1} dt \\ &= \frac{B_{\frac{z}{1+z}}(a-1, b-1)}{B_1(a-1, b-1)} \end{aligned}$$

8.3 - Applications

a) Théorème 1

"Soient Y_1 et Y_2 deux variables aléatoires indépendantes suivant respectivement des lois du χ^2 réduites à n_1 et n_2 degrés de liberté. La loi du rapport :

$$Z = \frac{Y_1}{Y_2}$$

est dite loi de Fisher $F(n_1, n_2)$ dont n_1 et n_2 sont les deux paramètres ; sa densité est définie pour $z \geq 0$ par :

$$\frac{\Gamma\left(\frac{n_1+n_2}{2}\right)}{\Gamma\left(\frac{n_1}{2}\right) \Gamma\left(\frac{n_2}{2}\right)} \frac{z^{\frac{n_1}{2}-1}}{(1+z)^{\frac{n_1+n_2}{2}}}$$

b) Théorème 2

"Soient Y_1 et Y_2 deux variables aléatoires indépendantes suivant respectivement des lois du χ^2 décentrée à n_1 degrés de liberté et de paramètre de décentrage λ , et du χ^2 réduite à n_2 degrés de liberté. La loi du rapport :

$$Z = \frac{Y_1}{Y_2}$$

est dite loi de Tang $T(\lambda ; n_1, n_2)$ de paramètres λ, n_1, n_2 ; sa densité de probabilité est définie pour $z \geq 0$ par :

$$e^{-\lambda/2} \sum_{m=0}^{\infty} \frac{\lambda^m}{2^m m!} \frac{\Gamma\left(\frac{n_1+n_2}{2} + m\right)}{\Gamma\left(\frac{n_1}{2} + m\right) \Gamma\left(\frac{n_2}{2}\right)} \frac{z^{\frac{n_1}{2}+m-1}}{(1+z)^{\frac{n_1+n_2}{2} + m}}$$

c) Loi de Fisher

D'après le théorème 1, on voit immédiatement que la valeur de la fonction de répartition au point x d'une telle loi est donnée par :

$$BD\left(\frac{n_1}{2}, \frac{n_2}{2}, 0 ; x\right)$$

d) Loi de Tang

La fonction de répartition en x d'une telle loi est égale à :

$$BD\left(\frac{n_1}{2}, \frac{n_2}{2}, \frac{\lambda}{2}; x\right)$$

en utilisant le théorème 2.

e) Loi de Hotelling

La quantité T^2 de Hotelling a pour fonction densité de probabilité :

$$\frac{e^{-\frac{d^2}{2}}}{(n-1)\Gamma\left(\frac{n-p}{2}\right)} \sum_{i=0}^{\infty} \frac{\left(\frac{d^2}{2}\right)^i \Gamma\left(\frac{n}{2} + i\right) \left(\frac{t^2}{n-1}\right)^{\frac{p}{2} + i - 1}}{i! \Gamma\left(\frac{p}{2} + i\right) \left(1 + \frac{t^2}{n-1}\right)^{n/2 + i}}$$

où n, p et d^2 représentent respectivement la taille de l'échantillon, la dimension de l'espace et le coefficient de décentrage.

On démontre que la variable aléatoire :

$$\frac{T^2}{n-1} \frac{n-p}{p}$$

suit une loi de Tang $T(d^2; p, n-p)$.

La valeur de la fonction de répartition au point x d'une loi de Hotelling est alors :

$$BD\left(\frac{p}{2}, \frac{n-p}{2}, \frac{d^2}{2}, \frac{x(n-p)}{p(n-1)}\right)$$

'REEL' 'PROCEDURE' BETADEC(A,B,P,Z) FF

'VALEUR' A,B,P,Z FF

'REEL' A,B,P,Z FF

'DEBUT'

'REEL' 'PROCEDURE' GAMMA(X) FF 'REEL' XFF

'DEBUT' 'REEL' H,Y FF

HF=1.0 FF YF=X FF

AIF 'SI' Y=0 'ALORS' HF=10**5 'SINON'

'SI' Y=2.0 'ALORS' 'ALLERA' A2 'SINON'

'SI' Y<INFER'2.0 'ALORS'

'DEBUT' HF=H/Y FF YF=Y+1.0 FF 'ALLERA' A1 'FIN'

'SINON' 'SI' Y>SUPEG'3 'ALORS' 'DEBUT'

YF=Y-1.0 FF HF=H*Y FF 'ALLERA' A1 'FIN'

'SINON' 'DEBUT' YF=Y-2.0 FF

HF=(((((1.0016063118*Y+.0051589951)*Y+.0044511400)*Y+.0721101567)*

Y+.0821117404)*Y+.4117741955)*Y

.4227874605)*Y+.9999999758)*H

'FIN' FF

A2F GAMMAF=H

'FIN' GAMMA FF

'REEL' 'PROCEDURE' BETA(X,P,Q,EPS) FF

'VALEUR' X,P,Q FF 'REEL' X,P,Q,EPS FF

'DEBUT' 'REEL' SOM,INFSOM,TOT,TOT1,AUX,AUX1,RECUR,INDEX FF

'BOOLEEN' BOOL FF

'SI' X=0 'OU' X=1 'ALORS'

'DEBUT' BETA F=X FF

'ALLERA' FIN

'FIN' FF

'SI' X<INFER'0.5 'ALORS' BOOLF='FAUX' 'SINON'

'DEBUT' BOOLF='VRAI' FF

TOTF=P FF

PF=Q FF

QF=TOT FF

XF=1-X

'FIN' FF

SOMF=0 FF

AUXF=1 FF

TOTF=1-X FF

RECURF=INDEXF=Q FF

'POUR' INDEXF=INDEX-1 'TANTQUE' INDEX 'SUPER' 0 'FAIRE'

'DEBUT' RECURF=INDEX FF

AUXF=AUX*(RECUR+1)/(TOT*(P+RECUR)) FF

SOMF=SOM+AUX

'FIN' FF

INFSOMF=AUXF=1 FF

INDEXF=0 FF

'POUR' INDEXF=INDEX+1 'TANTQUE'

(AUX/INFSOM) 'SUPER' EPS 'FAIRE'

'DEBUT' AUXF=AUX*X*(INDEX-RECUR)*(P+INDEX-1)/

(INDEX*(P+INDEX)) FF

INFSOMF=INFSOM+AUX

'FIN' FF

TOTF=TOTIF=GAMMA(RECUR) FF

AUXF=AUXIF=GAMMA(RECUR+P) FF

'POUR' INDEXF=RECUR 'PAS' 1 'JUSQUA' (Q-0.5) 'FAIRE'

'DEBUT' TOTIF=TOT1*INDEX FF

AUXIF=AUX1*(INDEX+P)

'FIN' FF

TOTF=X**P*(INFSOM*AUX/(P*TOT)+SOM*AUX1*

(1-X)**Q/(Q*TOT1))/GAMMA(P) FF

```
BETA F= 'SI' BOOL 'ALORS' 1-TOT 'SINON' TOT FF
FINE 'FIN' FF
'REEL' C,D,I,TOT,TOT1,A2,B2 FF
'ENTIER' M,K,KA,KB FF
KAF=ENTIER(A) FF A2F=A-KA FF
KBF=ENTIER(B) FF B2F=B-KB FF
DF=GAMMA(A2+B2)/(GAMMA(A2)*GAMMA(B2)) FF
'POUR' KF=0 'PAS' 1 'JUSQUA' KA-1 'FAIRE'
DF=D*(1+B/(A2+K)) FF
'POUR' KF=0 'PAS' 1 'JUSQUA' KB-1 'FAIRE'
DF=D*(1+A2/(B2+K)) FF
CF=D*EXP((A-1)*LN(Z)-(A+B-1)*LN(1+Z))/(A+B-1) FF
IF=BETA(Z/(1+Z),A,B,F*-8)*D FF
MF=0 FF
DF=1.0 FF
TOTF=D*I FF
BOUCLEF MF=M+1 FF
CF=C*Z*(A+M+B-2)/((1+Z)*(A+M-1)) FF
IF=I-C FF
DF=D*P/M FF
TOT1F=TOT FF
TOTF=TOT+D*I FF
'SI' TOT1 'NONEG' TOT 'ALORS' 'ALLERA' BOUCLE FF
BETADEC F=TOT*EXP(-P)
'FIN' FF
```

CHAPITRE - II

EXEMPLE DE TRAITEMENT DE DONNEES STATISTIQUES

L'étude des phénomènes réels porte sur des résultats d'expériences ou d'observations. La mise en ordre de l'information, sa présentation sous forme utilisable, ressort de la statistique descriptive.

L'objet de ce chapitre est de montrer, par l'étude d'un exemple élémentaire, de quelle manière il est possible d'aborder le traitement automatique des données statistiques, et d'effectuer simplement des travaux tels que tri, dénombrement ou classement. Ces opérations peuvent être relativement simples ou au contraire très complexes. L'ordinateur de par sa puissance de calcul, est parfaitement capable de se plier à n'importe quelle procédure, et permet de prendre en charge des traitements aussi compliqués que l'on veut.

La manipulation de l'information a longtemps été accomplie par la mécanographie classique, mais l'automatisation de la mécanographie reste toujours limitée alors que les capacités du calculateur permettent d'aller beaucoup plus loin. Un programme de dépouillement statistique donne la possibilité de procéder à des études plus élaborées, sans commune mesure avec le simple dénombrement.

1 - Problème proposé

Nous ne considérons ici que le type le plus simple parmi les problèmes usuels, en nous limitant à du matériel statistique qui possède la structure d'échantillon empirique, c'est-à-dire celle d'une suite d'observations indépendantes d'une même variable aléatoire.

Lorsque la taille d'un échantillon empirique, dans le cas où la loi de la variable aléatoire est continue, ou son ordre, si elle est discrète, sont grands, on est conduit à répartir les observations en un certain nombre de classes, afin de permettre une exploitation ultérieure des résultats un peu plus approfondie. Le problème qui se pose alors, et que nous nous proposons de résoudre, est le suivant :

Soient n observations indépendantes x_1, \dots, x_n et s nombres donnés tel que :

$$b_1 < b_2 < \dots < b_s$$

- Répartir les n observations en $s+1$ classes dont les limites sont les nombres b_i ($i = 1, \dots, s$).
- Déterminer l'effectif de chaque classe.
- Calculer les deux principales valeurs caractéristiques de l'échantillon, que sont la moyenne et la variance empiriques.

2 - Procédure APPROX

Le problème étant donné sous la forme précédente, la procédure APPROX effectue les calculs et permet de sortir les résultats. On peut considérer deux catégories de paramètres formels. D'une part les données :

- N : nombre d'observations ;
- S : nombre de séparatrices ;
- X : vecteur des observations ;
- B : vecteur des séparatrices ;

d'autre part les solutions :

COMPT : vecteur (de dimension $S+1$) des effectifs des classes ;

MOY : moyenne empirique ;

ECART : écart moyen quadratique empirique.

Les $S+1$ classes sont définies à l'aide des S séparatrices. Ainsi la classe d'ordre J est déterminée entièrement par $B[J-1]$ et $B[J]$; elle est telle que, si une observation $X[I]$ lui appartient, on a alors :

$$B[J-1] \leq X[I] < B[J]$$

en affectant à $B[0]$ la valeur :- l'infini et à $B[S+1]$ la valeur : + l'infini.

L'organigramme de APPROX est indiqué à la page suivante.

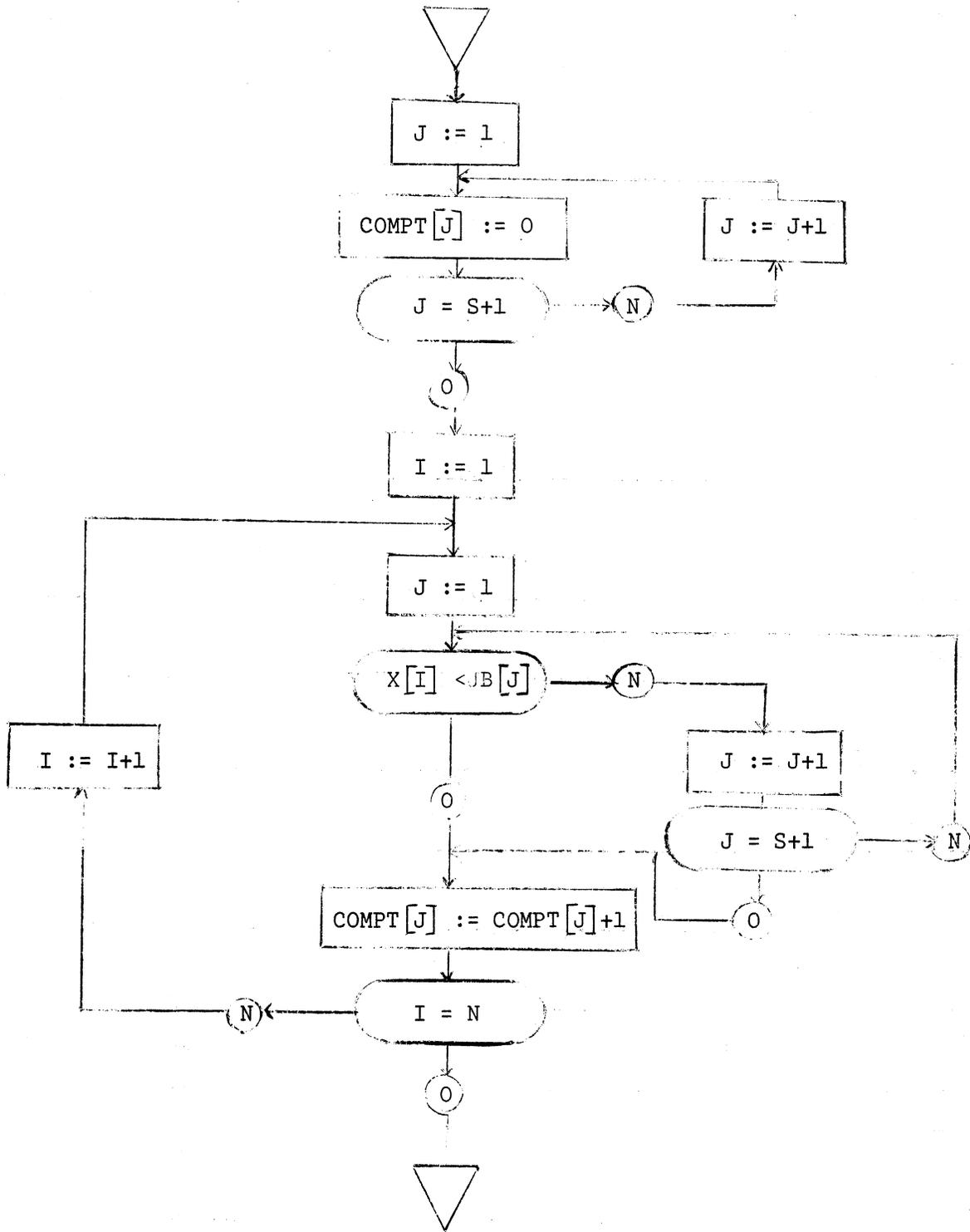
3 - Exemple numérique

Nous avons effectué un essai en considérant 600 observations du niveau de la mer en un point donné de l'Océan Atlantique. 18 classes ont été utilisées, en prenant les séparatrices b_i telles que :

$$b_i = 50 i, \quad i = 1, \dots, 17.$$

A la suite du listage de la procédure APPROX nous donnons les résultats que nous avons obtenus.

ORGANIGRAMME DE APPROX



```
'DEBUT'  
'PROCEDURE' APPROX(N,S,X,B,COMPT,MOY,ECART) FF  
'ENTIER' N,S FF 'TABLEAU' X,B FF  
'ENTIER' 'TABLEAU' COMPT FF 'REEL' MOY,ECART FF  
'DEBUT' 'ENTIER' I,J FF  
      MOYF=ECARTF=0 FF  
'POUR' JF=1 'PAS' 1 'JUSQUA' S+1 'FAIRE'  
      COMPT.(J).F=0 FF  
'POUR' IF=1 'PAS' 1 'JUSQUA' N 'FAIRE'  
'DEBUT' 'POUR' JF=1 'PAS' 1 'JUSQUA' S 'FAIRE'  
      'SI' X.(I). 'INFER' B.(J). 'ALORS'  
      'DEBUT' COMPT.(J).F=COMPT.(J).+1 FF  
      'ALLERA' FINBOUCLE  
      'FIN' FF  
      COMPT.(S+1).F=COMPT.(S+1).+1 FF  
FINBOUCLEF  
      MOYF=MOY+X.(I).  
      'FIN' FF  
      MOYF=MOY/N FF  
'POUR' IF=1 'PAS' 1 'JUSQUA' N 'FAIRE'  
      ECARTF=ECART+(X.(I).-MOY)**2 FF  
      ECARTF=RAC2(ECART/N) FF  
      SORTIE(6,MOD1,1,B.(1),,1,B.(2).) FF  
      MOD1FMODELE(''(30X,31H)PROCEDURE DE PARTAGE EN CLASSES //'  
      10X,26HSUIVANT LES SEPARATRICES F,1F7.2,3H , ,  
      1F7.2,2H , )''') FF  
'POUR' JF=2 'PAS' 1 'JUSQUA' ENTIER(S/2) 'FAIRE'  
'DEBUT' SORTIE(6,MOD2,1,B.(2*J-1),,1,B.(2*J).) FF  
      MOD2FMODELE(''(36X,1F7.2,3H , ,1F7.2,2H , )''')  
'FIN' FF  
'SI' S-ENTIER(S/2)*2 'SUPER' 0 'ALORS'  
'DEBUT' SORTIE(6,MOD2PR,1,B.(S).) FF  
      MOD2PRFMODELE(''(36X,1F7.2,2H .)''')  
'FIN' FF  
      SAUTLIGNE FF  
      SORTIE(6,MOD3,1,N,1,MOY,1,ECART) FF  
      MOD3FMODELE(''(10X,30HNOMBRE TOTAL D OBSERVATIONS NF,115 //'  
      10X,10HMOYENNE F,F10.4 /  
      10X,10HE.M.Q. F,F10.4)''') FF  
      SAUTLIGNE FF  
'POUR' JF=1 'PAS' 1 'JUSQUA' S+1 'FAIRE'  
'DEBUT' SORTIE(6,MOD4,1,J,1,COMPT.(J).) FF  
      MOD4FMODELE(''(10X,22HFREQUENCE DE LA CLASSE,113,  
      3H = ,114)''')  
'FIN' FF  
'FIN' FF  
'ENTIER' N,S,I FF  
'REEL' MOY,ECART FF  
LIRE(N,S) FF  
'DEBUT' 'TABLEAU' X.(1FN),,B.(1FS). FF  
      'ENTIER' 'TABLEAU' COMPT.(1FS+1). FF  
'POUR' IF=1 'PAS' 1 'JUSQUA' N 'FAIRE'  
      X.(I).F=RDONNEE FF  
'POUR' IF=1 'PAS' 1 'JUSQUA' S 'FAIRE'  
      B.(I).F=RDONNEE FF  
      APPROX(N,S,X,B,COMPT,ECART)  
'FIN' FF  
'FIN' FF
```

6400, 0183, 0008, HISLEUR APPROX

PROCEDURE DE PARTAGE EN CLASSES

SUIVANT LES SEPARATRICES : 50.00 , 100.00 ,
150.00 , 200.00 ,
250.00 , 300.00 ,
350.00 , 400.00 ,
450.00 , 500.00 ,
550.00 , 600.00 ,
650.00 , 700.00 ,
750.00 , 800.00 ,
850.00 .

NOMBRE TOTAL D OBSERVATIONS N : 600

MOYENNE : 509.3867

E.M.Q. : 123.7413

FREQUENCE DE LA CLASSE	1 =	0
FREQUENCE DE LA CLASSE	2 =	2
FREQUENCE DE LA CLASSE	3 =	1
FREQUENCE DE LA CLASSE	4 =	7
FREQUENCE DE LA CLASSE	5 =	11
FREQUENCE DE LA CLASSE	6 =	20
FREQUENCE DE LA CLASSE	7 =	22
FREQUENCE DE LA CLASSE	8 =	50
FREQUENCE DE LA CLASSE	9 =	63
FREQUENCE DE LA CLASSE	10 =	96
FREQUENCE DE LA CLASSE	11 =	105
FREQUENCE DE LA CLASSE	12 =	86
FREQUENCE DE LA CLASSE	13 =	46
FREQUENCE DE LA CLASSE	14 =	44
FREQUENCE DE LA CLASSE	15 =	25
FREQUENCE DE LA CLASSE	16 =	10
FREQUENCE DE LA CLASSE	17 =	8
FREQUENCE DE LA CLASSE	18 =	4

15 HEURES 18 MINUTES 13 SECONDES

CHAPITRE - III

TESTS D'HYPOTHESES

Les tests sont de la plus grande importance pour toutes sortes d'applications. Nous nous proposons, dans ce chapitre, d'étudier ceux qui sont généralement considérés comme essentiels.

Les problèmes statistiques que nous rencontrerons entrent tous dans le formalisme suivant :

Soient X une variable aléatoire réelle de loi de probabilité P_X et (X_1, \dots, X_n) un n -échantillon de X . P_X est inconnue mais on suppose qu'elle appartient à une classe \mathcal{P} de lois de probabilité. On considère une partie \mathcal{P}_0 de \mathcal{P} et son complémentaire :

$$\mathcal{P}_1 = \overline{\mathcal{P}_0} \subset \mathcal{P}$$

En utilisant une réalisation (x_1, \dots, x_n) de l'échantillon (X_1, \dots, X_n) on cherche à savoir s'il est raisonnable d'admettre que P_X appartient à \mathcal{P}_0 . A cet effet on introduit les deux hypothèses :

$$H_0 : P_X \in \mathcal{P}_0$$

$$H_1 : P_X \in \mathcal{P}_1$$

Laquelle doit-on choisir ?

On peut considérer quatre parties :

- Les trois premiers paragraphes sont consacrés à des tests d'adéquation.
La partie \mathcal{D}_0 étant réduite à une loi de probabilité unique P_0 , H_0 devient une hypothèse simple. Nous examinons les tests du χ^2 , de normalité et de Kolmogorov-Smirnov, que l'on applique lorsque la loi P_0 est respectivement discrète, normale et continue quelconque.
- Au paragraphe 4 nous traitons numériquement un problème qui se pose lors du test de Student et qui est celui de la taille requise de l'échantillon pour obtenir des risques d'erreur déterminés à l'avance.
- Ensuite nous calculons les probabilités des erreurs que l'on peut commettre en testant deux hypothèses linéaires.
- Nous abordons enfin, au dernier paragraphe, le problème de la reconnaissance de l'aspect aléatoire d'une suite d'éléments d'un même ensemble E , dans le cas particulier où E ne possède que deux éléments distincts.

1 - TEST DU χ^2

1.1 - Introduction

Soit P_0 une loi de probabilité donnée discrète finie, telle que, si la variable aléatoire X suit P_0 , elle ne peut prendre que les valeurs a_1, \dots, a_r avec les probabilités respectives :

$$\Pr\{X = a_j\} = p_j^0, \quad j = 1, \dots, r.$$

et soit un échantillon empirique discret se ramenant aux fréquences empiriques :

$$f_j = \frac{n_j}{n}, \quad j = 1, \dots, r$$

où n_j est le nombre d'observations égales à a_j . On se propose de tester l'hypothèse simple $H_0 : P_X = P_0$ contre l'hypothèse non paramétrique $H_1 : P_X \neq P_0$.

1.2 - Théorème de Karl Pearson

"Soit P_X la loi de probabilité qui aux points a_1, \dots, a_r attribue les probabilités p_1, \dots, p_r et soient n variables aléatoires X_1, \dots, X_n indépendantes et de loi commune P_X .

Introduisons Y_1, \dots, Y_r les variables aléatoires définies par :

$$\underline{Y_j = \text{nombre de variables } X_i \text{ égales à } a_j, \quad j = 1, \dots, r ;}$$

alors la variable aléatoire :

$$Y = \sum_{j=1}^r \frac{(Y_j - np_j)^2}{np_j}$$

tend en loi, lorsque $n \rightarrow \infty$ vers une variable aléatoire du χ^2 à $r-1$ degrés de liber

1.3 - Test du χ^2

Le théorème que nous venons d'énoncer va nous permettre de juger si une hypothèse concernant la loi de probabilité d'une variable aléatoire discrète est compatible avec la réalisation d'un échantillon de cette variable.

Considérons une variable aléatoire X de loi de probabilité P telle que :

$$\Pr[X = a_j] = p_j, \quad j = 1, \dots, r ;$$

$$\sum_{i=1}^r p_i = 1$$

Désignons par Y_j le nombre de répétitions de a_j dans un échantillon (X_1, \dots, X_n) .

Alors on vient de voir que :

$$Y = \sum_{j=1}^r \frac{(Y_j - np_j)^2}{np_j} \rightsquigarrow \chi_{r-1}^2$$

Si l'on considère maintenant l'hypothèse $H_0 : P_X = P_0$, il convient de remplacer dans la variable aléatoire Y les probabilités p_j par les probabilités p_j^0 définies par P_0 . Cette variable aléatoire peut servir de test à H_0 en prenant comme région d'acceptation

$$C = \{ \vec{x} \in \mathbb{R}^n : y(\vec{x}) \leq Q_{r-1}^{-1}(1-\alpha) \}$$

où $Q_r(z)$ est la fonction de répartition de la loi du χ^2 à r degrés de liberté et α le niveau de signification.

Remarque :

L'application du test du χ^2 repose sur le fait que pour tout $j = 1, \dots, r$:

$$\frac{Y_j - np_j}{\sqrt{np_j}} \rightsquigarrow N(0, \sqrt{1-p_j})$$

Pour l'application pratique du test il ne faut pas que np_j soit trop petit. Il a été proposé $np_j \geq 5$. S'il n'en est pas ainsi il est indiqué de réunir des points successifs en une classe unique.

1.4 - Effets de l'estimation de paramètres sur le test du χ^2

Supposons maintenant que la loi de probabilité intervenant dans l'hypothèse H_0 , que l'on veut tester, ne soit pas entièrement spécifiée, mais dépende de paramètres inconnus.

Désignons par $P(\theta)$ la loi de probabilité de la variable aléatoire X :

$$\theta \in \Theta \subset \mathbb{R}^l$$

Soit θ^* un estimateur de θ , obtenu à partir de l'échantillon.

On se propose de rechercher la loi de probabilité de la variable aléatoire :

$$Y(\theta^*) = \sum_{j=1}^r \frac{(Y_j - np_j(\theta^*))^2}{np_j(\theta^*)}$$

où $p_j(\theta^*)$ est la probabilité afférente au point a_j calculée avec $P(\theta^*)$.

Le théorème de Fisher démontre que, si θ^* est un estimateur de maximum de vraisemblance, la variable aléatoire $Y(\theta^*)$ tend en loi vers une variable aléatoire du χ^2 à $r-l-1$ degrés de liberté, lorsque n tend vers l'infini.

On utilisera alors $Y(\theta^*)$ pour tester $H_0 : P_X = P(\theta^*)$ en prenant comme région d'acceptation :

$$C^* = \{ \vec{x} \in \mathbb{R}^n : y^*(\vec{x}) \leq \chi_{r-l-1}^{-1}(1-\alpha) \}$$

1.5 - Procédure ADEQUATION.

La procédure ADEQUATION permet de tester, par la méthode du χ^2 , que nous venons de décrire, l'adéquation de fréquences empiriques réduites à des fréquences théoriques données.

Elle utilise cinq paramètres formels :

N : taille de l'échantillon
R : nombre de valeurs distinctes que peut prendre une observation
FTH : vecteur des fréquences théoriques réduites
FEMP : vecteur des fréquences empiriques réduites
ALPHA : niveau de signification du test.

S'il existe un ou plusieurs indices j tels que :

$$np_j < 5$$

il y a édition d'un message indiquant les classes trop pauvres et arrêt du programme.

1.6 - Exemple numérique

On dispose de 5 pièces de monnaie et on les lance 1000 fois successivement en relevant chaque fois le nombre de "face" obtenus. On obtient ainsi :

nombre de "face"	1	2	3	4	5	6
fréquence observée	0.030	0.150	0.350	0.320	0.125	0.025

Les pièces sont-elles symétriques ?

Le calcul des fréquences théoriques, avec l'hypothèse de symétrie des pièces, donne les résultats :

nombre de "face "	1	2	3	4	5	6
fréquence théorique	0.031	0.156	0.313	0.313	0.156	0.031

On applique alors le test du χ^2 , avec un niveau de signification égal à 5 %.

On est conduit, en examinant les résultats obtenus, à rejeter l'hypothèse de compatibilité des observations avec les fréquences théoriques.

```

'DEBUT'
'PROCEDURE' ADEQUATION(N,R,FTH,FEMP,ALPHA) FF
'ENTIER' N,R FF 'REEL' ALPHA FF
'TABLEAU' FTH,FEMP FF
'DEBUT'
'REEL' 'PROCEDURE' NORMALE(X,MOY,ECART) FF
'VALEUR' X,MOY,ECART FF 'REEL' X,MOY,ECART FF
'DEBUT' 'ENTIER' I FF
'REEL' SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
XF=(X-MOY)/ECART FF
SGNEF=SIGNE(X) FF
XF=ABS(X) FF
X2F=X*X FF
YF=0.3989422804*EXP(-0.5*X2) FF
'SI' X 'INFER' 3.5 'ALORS'
'DEBUT' SF=XF=Y*X FF
'POUR' IF=3,I+2 'TANTQUE' S 'NONEG' T 'FAIRE'
'DEBUT' TF=S FF
XF=X*X2/I FF
SF=S+X
'FIN' FF
NORMALEF=0.5+SGNE*S
'FIN'
'SINON'
'DEBUT' Q1F=X FF
P2F=Y*X FF
P1F=Y FF
Q2F=X2+1.0 FF
MF=Y/X FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS'
'DEBUT' MF=1-M FF
TF=1-T
'FIN' FF
'POUR' IF=2,I+1 'TANTQUE'
((M 'NONEG' T) 'ET' (S 'NONEG' T)) 'FAIRE'
'DEBUT' SF=X*P2+I*P1 FF
P1F=P2 FF
P2F=S FF
SF=X*Q2+I*Q1 FF
Q1F=Q2 FF
Q2F=S FF
SF=M FF
MF=T FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS' TF=1-T FF
'FIN' FF
NORMALEF=T
'FIN' FF
FINORMF 'FIN' FF
'REEL' 'PROCEDURE' KI(N,X) FF
'VALEUR' N,X FF
'ENTIER' N FF 'REEL' X FF
'DEBUT' 'ENTIER' I FF
'REEL' AUX,SOM,PI FF
'SI' N-2*ENTIER(N/2)=1 'ALORS'
'ALLERA' NIMP FF
SOMF=AUXF=1.0 FF
'SI' N=2 'ALORS' 'ALLERA' FPAIR FF

```

```

      'POUR' IF=1 'PAS' 1 'JUSQUA' N/2-1 'FAIRE'
      'DEBUT' AUXF=AUX*X/(2*I) FF
              SOMF=SOM+AUX
      'FIN' FF
FPAIRF KIF=1.0-EXP(-X/2)*SOM FF
      'ALLERA' FINKI FF
NIMPF  PIF=3.141592653590 FF
      AUXF=RAC2(2/(PI*X)) FF
      SOMF=0.0 FF
      'SI' N=1 'ALORS' 'ALLERA' FNIMP FF
      'POUR' IF=0 'PAS' 1 'JUSQUA' (N-3)/2 'FAIRE'
      'DEBUT' AUXF=AUX*X/(2*I+1) FF
              SOMF=SOM+AUX
      'FIN' FF
FNIMPF KIF=2*NORMALE(RAC2(X),0.0,1.0)-1.0-EXP(-X/2)*SOM FF
FINKIF 'FIN' FF
      'REEL' TOTAL FF
      'ENTIER' J FF
      'BOOLEEN' TEST FF
      SAUTPAGE FF
      SORTIE(6,MOD1,1,N,1,R) FF
      MODIFMODELE('((130X,29HPROCEDURE D ADEQUATION DU KI2 ///
10X,23HNOMBRE D OBSERVATIONS F,115,
10X,19HNOMBRE DE CLASSES F,115 / )))) FF
      'POUR' JF=1 'PAS' 1 'JUSQUA' R 'FAIRE'
      'DEBUT' SORTIE(6,MOD2,1,J,1,FTH.(J),1,FEMP.(J),) FF
              MOD2FMODELE('((10X,6HCLASSE,113,10X,21HFREQUENCE THEORIC
F5.3,10X,21HFREQUENCE EMPIRIQUE F,F5.3 ///))')
      'FIN' FF
      TESTF='FAUX' FF
      'POUR' JF=1 'PAS' 1 'JUSQUA' R 'FAIRE'
      'DEBUT' 'SI' N*FTH.(J). 'INFEG' 5.0 'ALORS'
              'DEBUT' TESTF='VRAI' FF
                      SORTIE(6,MOD3,1,J) FF
                      MOD3FMODELE('((10X,17HLA CLASSE NUMERO ,113,1X,
15HEST TROP PAUVRE'))')
              'FIN' FF
      'FIN' FF
      'SI' TEST 'ALORS'
      'DEBUT' SAUTLIGNE FF
              SORTIE(6,MOD6) FF
              MOD6FMODELE('((10X,22HLE TEST EST IMPOSSIBLE))') FF
              'ALLERA' SAUT
      'FIN' FF
      TOTALF=0 FF
      'POUR' JF=1 'PAS' 1 'JUSQUA' R 'FAIRE'
      TOTALF=TOTAL+(FEMP.(J).-FTH.(J).)**2/FTH.(J). FF
      'SI' KI(R-1,N*TOTAL) 'SUPER' (1-ALPHA) 'ALORS'
      'DEBUT'
              SORTIE(6,MOD5,1,ALPHA) FF
              MOD5FMODELE('((10X,23HL ADEQUATION N EST PAS ,
38HACCEPTABLE AU NIVEAU DE SIGNIFICATION ,F5.3))')
      'FIN'
      'SINON'
      'DEBUT'
              SORTIE(6,MOD4,1,ALPHA) FF
              MOD4FMODELE('((10X,31HL ADEQUATION EST ACCEPTABLE AU ,
24HNIVEAU DE SIGNIFICATION ,F5.3))')
      'FIN' FF
SAUTF 'FIN' FF

```

```
'ENTIER' N,R,I FF
'REEL' ALPHA FF
LIRE(N,R,ALPHA) FF
'DEBUT' 'TABLEAU' FTH,FEMP.(1FR). FF
        'POUR' IF=1 'PAS' 1 'JUSQUA' R 'FAIRE'
        FTH.(I).F=RDONNEE FF
        'POUR' IF=1 'PAS' 1 'JUSQUA' R 'FAIRE'
        FEMP.(I).F=RDONNEE FF
ADEQUATION(N,R,FTH,FEMP,ALPHA)
'FIN' FF
'FIN' FF
```

6400, 0183, CC8C, HISLEUR ADEQUAT

PROCEDURE D ADEQUATION DU KI2

NOMBRE D OBSERVATIONS : 1000

NCMBRE DE CLASSES : 6

CLASSE 1 FREQUENCE THECRIQUE : 0.031 FREQUENCE EMPIRIQUE : 0.030

CLASSE 2 FREQUENCE THECRIQUE : 0.156 FREQUENCE EMPIRIQUE : 0.150

CLASSE 3 FREQUENCE THECRIQUE : 0.313 FREQUENCE EMPIRIQUE : 0.350

CLASSE 4 FREQUENCE THECRIQUE : 0.313 FREQUENCE EMPIRIQUE : 0.320

CLASSE 5 FREQUENCE THECRIQUE : 0.156 FREQUENCE EMPIRIQUE : 0.125

CLASSE 6 FREQUENCE THECRIQUE : 0.031 FREQUENCE EMPIRIQUE : 0.025

L ADEQUATION N EST PAS ACCEPTABLE AU NIVEAU DE SIGNIFICATION 0.050

11
11

2 - TEST DE NORMALITE

2.1 - Introduction

Soit un échantillon empirique continu :

$$P_X ; x_1, \dots, x_n$$

et la question :

" P_X est-elle une loi de probabilité gaussienne?"

Il s'agit là d'un test d'adéquation à une hypothèse paramétrique. Une méthode est généralement employée : la droite de Henry.

La question à laquelle nous allons nous efforcer de répondre maintenant possède un caractère moins général et peut se formuler ainsi :

" P_X est-elle une loi de probabilité gaussienne de moyenne m_0 et d'écart quadratique moyen σ_0 donnés?"

2.2 - Test proposé

Soit donc un échantillon empirique continu et une loi de probabilité continue P_0 donnée. Le problème qui consiste à savoir si les observations sont compatibles avec l'hypothèse $H : P_X = P_0$ peut se résoudre au moyen d'un test découlant du théorème de Kolmogorov-Smirnov (3.2).

Pourtant, lorsque le nombre d'observations est grand, l'usage d'un tel test devient rapidement fastidieux et il faut alors rechercher une autre méthode.

Dans le cas particulier que nous traitons, nous allons, comme au paragraphe précédent, appliquer le test du χ^2 , mais après avoir approximé l'échantillon continu considéré par un échantillon empirique discret, en procédant à un partage en classes selon la technique suivante :

Soient r nombres s_1, \dots, s_r ordonnés, définissant donc $r+1$ intervalles C_j ($j=1, \dots, r+1$) tels que :

$$x \in C_1 \Leftrightarrow x < s_1 ;$$

$$x \in C_j \Leftrightarrow s_{j-1} \leq x < s_j, \quad j = 2, \dots, r ;$$

$$x \in C_{r+1} \Leftrightarrow s_r \leq x ;$$

L'échantillon empirique discret est alors défini par les $r+1$ nombres :

$$f_j = \frac{n_j}{n}, \quad j = 1, \dots, r+1$$

où n_j est l'effectif de la classe C_j .

Le problème qu'il faut alors résoudre est celui du choix des r séparatrices s_1, \dots, s_r . Parmi les divers critères existants nous avons choisi celui d'équiprobabilité. Si $N(x ; m_0, \sigma_0)$ représente la fonction de répartition de la loi $\mathcal{N}(m_0, \sigma_0)$ les s_j ($j = 1, \dots, r$) sont tels que :

$$N(s_1 ; m_0, \sigma_0) = \frac{1}{r+1} ;$$

$$N(s_j ; m_0, \sigma_0) - N(s_{j-1} ; m_0, \sigma_0) = \frac{1}{r+1}, \quad j = 2, \dots, r-1 ;$$

$$1 - N(s_r ; m_0, \sigma_0) = \frac{1}{r+1}$$

Puisque nous possédons le vecteur des fréquences empiriques réduites et celui des fréquences théoriques, il suffit alors d'effectuer le test du χ^2 que nous avons décrit au paragraphe 1.

2.3 - Procédure NORMALITE

NORMALITE est une procédure à six paramètres formels.

N : nombre d'observations.
R : nombre de classes équiprobables.
X : vecteur des observations.
MOY : moyenne de la loi de probabilité gaussienne.
SIGMA : écart quadratique moyen de la loi.
ALPHA : niveau de signification du test.

On peut diviser le programme en trois parties :

- détermination des R séparatrices délimitant les classes équiprobables ;
- répartition des N observations dans les classes et calcul des fréquences empiriques réduites ;
- test du χ^2 classique, par une technique analogue à celle de la procédure ADEQUATION.

Remarque :

Nous avons vu en 1.3 qu'une condition était requise pour qu'un test du χ^2 soit valable : le produit np_j doit être supérieur ou égal à 5 pour tout j.

La procédure NORMALITE utilise R+1 classes équiprobables.

On peut donc écrire :

$$p_j = \frac{1}{R+1}, \quad j = 1, \dots, R+1.$$

La condition devient alors :

$$\frac{N}{R+1} > 5$$

Dans le cas où les paramètres effectifs qui remplacent N et R ne la satisfont pas, il y a édition d'un message d'erreur et arrêt du programme.

2.4 - Exemple numérique

Nous donnons ici un exemple de problème qui a été résolu par l'intermédiaire de la procédure NORMALITE.

A partir d'un ensemble qui comportait 300 observations indépendantes, nous avons testé, à un niveau de signification de 5 %, l'adéquation de leur loi commune à une loi normale de moyenne 90 et de variance 45, en utilisant 21 classes équiprobables.

Les résultats édités ont été :

- numéros des classes,
- valeurs des séparatrices,
- fréquences empiriques,
- décision prise.

```

*DEBUT*
*PROCEDURE* NORMALITE(N,R,X,MOY,SIGMA,ALPHA) FF
*ENTIER* N,R FF
*TABLEAU* X FF
*REEL* MOY,SIGMA,ALPHA FF
*DEBUT*
*REEL* *PROCEDURE* KI(N,X) FF
*VALEUR* N,X FF
*ENTIER* N FF *REEL* X FF
*DEBUT*
*REEL* *PROCEDURE* NORMALE(X,MOY,ECART) FF
*VALEUR* X,MOY,ECART FF *REEL* X,MOY,ECART FF
*DEBUT* *ENTIER* I FF
*REEL* SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
XF=(X-MOY)/ECART FF
SGNEF=SIGNE(X) FF
XF=ABS(X) FF
X2F=X*X FF
YF=0.3989422804*EXP(-0.5*X2) FF
*SI* X *INFER* 3.5 *ALORS*
*DEBUT* SF=XF=Y*X FF
*POUR* IF=3,I+2 *TANTQUE* S *NONNEG* T *IRE*
*DEBUT* TF=S FF
XF=X*X2/I FF
SF=S+X
*FIN* FF
NORMALEF=0.5+SGNE*S
*FIN*
*SINON*
*DEBUT* Q1F=X FF
P2F=Y*X FF
P1F=Y FF
Q2F=X2+1.0 FF
MF=Y/X FF
TF=P2/Q2 FF
*SI* SIGNE=1 *ALORS*
*DEBUT* MF=1-M FF
TF=1-T
*FIN* FF
*POUR* IF=2,I+1 *TANTQUE*
((M *NONNEG* T) *ET* (S *NONNEG* T)) *IF* E
*DEBUT* SF=X*P2+I*P1 FF
PIF=P2 FF
P2F=S FF
SF=X*Q2+I*Q1 FF
Q1F=Q2 FF
Q2F=S FF
SF=M FF
MF=T FF
TF=P2/Q2 FF
*SI* SIGNE=1 *ALORS* TF=1-T F
*FIN* FF
NORMALEF=T
*FIN* FF
FINORMF *FIN* FF
*ENTIER* I FF
*REEL* AUX,SOM,PI FF
*SI* N-2*ENTIER(N/2)=1 *ALORS*
*ALLERA* NIMP FF
SOMF=AUXF=1.0 FF

```

```

'SI' N=2 'ALORS' 'ALLERA' FPAIR FF
'POUR' IF=1 'PAS' 1 'JUSQUA' N/2-1 'FAIRE'
'DEBUT' AUXF=AUX*X/(2*I) FF
      SOMF=SOM+AUX
'FIN' FF
FPAIRF KIF=1.0-EXP(-X/2)*SOM FF
      'ALLERA' FINKI FF
NIMPF PIF=3.141592653590 FF
      AUXF=RAC2(2/(PI*X)) FF
      SOMF=0.0 FF
      'SI' N=1 'ALORS' 'ALLERA' FNIMP FF
      'POUR' IF=0 'PAS' 1 'JUSQUA' (N-3)/2 'FAIRE'
      'DEBUT' AUXF=AUX*X/(2*I+1) FF
            SOMF=SOM+AUX
      'FIN' FF
FNIMPF KIF=2*NORMALE(RAC2(X),0.0,1.0)-1.0-EXP(-X/2)*SOM FF
FINKIF 'FIN' FF
'REEL' 'PROCEDURE' NOTLAG(T) FF 'REEL' T FF
'COMMENTAIRE' ETANT DONNE T (0.000003 INFEG T INFEG 0.999997)
      LA PROCEDURE CALCULE NOTLAG(T) TEL QUE
      T = INTEGRALE DE MOINS L INFINI A NOTLAG(T)
      DE EXP(-X*X/2)/RAC2(2*PI) DX
      AVEC UNE EPREUR INFERIEURE A 0.001 .
      L APPROXIMATION EST DUE A C. HASTINGS
      ( APPROXIMATIONS FOR DIGITAL COMPUTERS .
      SHEET 68 PAGE 192 ) FF
'DEBUT' 'REEL' Q,ETA,U,NUM,DEN FF
      'ENTIER' I FF
      'TABLEAU' A,B.(OF3). FF
      A.(0). F= 2.515517 FF B.(0). F= 1.0 FF
      A.(1). F= 0.802853 FF B.(1). F= 1.432788 FF
      A.(2). F= 0.010328 FF B.(2). F= 0.189269 FF
      B.(3). F= 0.001308 FF
      Q F= 'SI' T 'SUPEG' 0.5
            'ALORS' 1.0-T
            'SINON' T FF
      ETA F= -LN(Q*Q) FF
      ETA F= RAC2(ETA) FF
      NUM F= A.(2). FF
      'POUR' I F= 1,0 'FAIRE'
      NUM F= NUM*ETA + A.(I). FF
      DEN F= B.(3). FF
      'POUR' I F= 2,1,0 'FAIRE'
      DEN F= DEN*ETA + B.(I). FF
      U F= ETA - NUM/DEN FF
      NOTLAG F= 'SI' T 'SUPEG' 0.5
                'ALORS' U
                'SINON' -U
      'FIN' PROCEDURE NOTLAG FF
'TABLEAU' SEP.(1FR).,FEMP,FTH.(1FR+1). FF
'ENTIER' I,J FF
'REEL' TOTAL FF
'SI' N 'INFER' 5*(R+1) 'ALORS'
'DEBUT' ECRIRE(' (TEST IMPOSSIBLE) ') FF
      'ALLERA' FINOR
'FIN' FF
'POUR' JF=1 'PAS' 1 'JUSQUA' R 'FAIRE'
SEP.(J).F=MOY+SIGMA*NOTLAG(J/(R+1)) FF
'POUR' JF=1 'PAS' 1 'JUSQUA' R+1 'FAIRE'

```

```
FEMP.(J).F=0.0 FF
'POUR' IF=1 'PAS' 1 'JUSQUA' N 'FAIRE'
'DEBUT' 'POUR' JF=1 'PAS' 1 'JUSQUA' R 'FAIRE'
      'SI' X.(I). 'INFER' SEP.(J). 'ALORS'
      'DEBUT' FEMP.(J).F=FEMP.(J).+1 FF
      'ALLERA' FINBOUCLE
      'FIN' FF
      FEMP.(R+1).F=FEMP.(R+1).+1 FF
FINBOUCLEF
'FIN' FF
ECRIRE('(' CLASSE'))',,(' SEPARA'))',,('TRICES'))',,
      (' FREQUENCE'))' FF
SAUTLIGNE FF
ECRIRE('(' 1))',,(' MOINS L INFINI'))',,SEP.(1).,,
FEMP.(1).) FF
'POUR' IF=2 'PAS' 1 'JUSQUA' R 'FAIRE'
ECRIRE(I,,SEP.(I-1),,SEP.(I),,,FEMP.(I).) FF
ECRIRE(R+1,,SEP.(R),,(' PLUS L INFINI'))',,FEMP.(R+1).) FF
SAUTLIGNE FF
'POUR' JF=1 'PAS' 1 'JUSQUA' R+1 'FAIRE'
'DEBUT' FEMP.(J).F=FEMP.(J)./N FF
      FTH.(J).F=1/(R+1)
'FIN' FF
TOTALF=0.0 FF
'POUR' JF=1 'PAS' 1 'JUSQUA' R+1 'FAIRE'
TOTALF=TOTALF+(FEMP.(J).-FTH.(J).)**2/FTH.(J). FF
'SI' KI(R-1,N*TOTAL) 'SUPER' (1-ALPHA) 'ALORS'
ECRIRE('('NORMALITE ACCEP'))',,('TEE'))'
'SINON' ECRIRE('('NORMALITE REFUS'))',,('EE'))' FF
FINORF
'FIN' FF
'ENTIER' N,R,I FF
'REEL' MOY,SIGMA,ALPHA FF
LIRE(N,R,MOY,SIGMA,ALPHA) FF
'DEBUT' 'TABLEAU' X.(1FN). FF
      'POUR' IF=1 'PAS' 1 'JUSQUA' N 'FAIRE'
      X.(I).F=RDONNEE FF
      NORMALITE(N,R,X,MOY,SIGMA,ALPHA)
'FIN' FF
'FIN' FF
```

6400, C182, C200, HISLEUR, NC

CLASSE	SEPARATRICES	FREQUENCE
1	MCINS L INFINI	.30000000+C2
2	.78805637+02	.33000000+C2
3	.81216502+02	.80000000+C1
4	.82838365+02	.26000000+C2
5	.84123561+02	.24000000+C2
6	.85222605+02	.24000000+C2
7	.86206017+02	.17000000+C2
8	.87113514+02	0
9	.87970467+02	.14000000+C2
10	.88794805+02	.90000000+C1
11	.89600446+02	.14000000+C2
12	.90399551+02	.10000000+C2
13	.91205193+02	.19000000+C2
14	.92029532+02	0
15	.92886484+02	.13000000+C2
16	.93793981+02	.11000000+C2
17	.94777393+02	.70000000+C1
18	.95876436+02	.14000000+C2
19	.97161633+02	.60000000+C1
20	.98783495+02	.13000000+C2
21	.10119436+03 PLUS L INFINI	.80000000+C1

NORMALITE ACCEPTEE 14 HEURES 18 MINUTES 44 SECONDES

3 - TEST DE KOLMOGOROV-SMIRNOV

3.1 - Introduction

Soit un échantillon empirique continu et une loi de probabilité donnée P_0 continue, de fonction de répartition $F_0(x)$. Le problème que nous nous proposons maintenant de résoudre est encore :

"Est-ce que les n observations sont compatibles avec l'hypothèse $P_X = P_0$?".

3.2 - Théorème de Kolmogorov-Smirnov

"Soient $F(x)$ la fonction de répartition supposée continue d'une variable aléatoire X et $F_n(x)$ l'histogramme cumulé d'un échantillon empirique de n observations de X . En posant :

$$D_n = \sup_{x \in \mathbb{R}} |F_n(x) - F(x)|$$

la loi de probabilité de la variable aléatoire $\sqrt{n} D_n$ tend vers une loi de probabilité indépendante de $F(x)$, lorsque $n \rightarrow \infty$."

3.3 - Test de Kolmogorov-Smirnov

Nous nous limitons à étudier le cas où la loi de probabilité P_0 est la loi uniforme sur le segment $(0,1)$.

Supposons que P_0 soit une autre loi de probabilité de fonction de répartition $F(y)$. On peut alors facilement montrer que les deux problèmes Pb1 et Pb2 sont équivalents :

Pb1 : Est-ce que les n observations (y_1, \dots, y_n) sont compatibles avec l'hypothèse la loi de probabilité est P_0 ?

Pb2 : Est-ce que les n observations (x_1, \dots, x_n) sont compatibles avec l'hypothèse la loi de probabilité est uniforme sur le segment $(0,1)$?

à la condition que les x_i soient obtenus à partir des y_i par la transformation :

$$x_i = F(y_i), \quad i = 1, \dots, n.$$

Le problème Pb2 sera donc celui que nous nous attacherons à résoudre, les autres problèmes du type Pbl pouvant tous s'y ramener.

Nous utiliserons un test qui découle immédiatement du théorème de Kolmogorov-Smirnov, énoncé en 3.2.

Nous rejetterons l'hypothèse d'uniformité sur $(0,1)$ de la loi de probabilité P_X , si et seulement si :

$$D_n > D(n, \alpha)$$

où α représente le niveau de signification du test et n la taille de l'échantillon.

La fonction numérique $D(n, \alpha)$ est obtenue de la manière suivante :

a) Pour n inférieur ou égal à 20.

Etant donné la fonction K_n de la variable réelle ε , où ε appartient à l'intervalle $[0, 1]$:

$$K_n(\varepsilon) = \sum_{i=0}^{[n(1-\varepsilon)]} C_n^i \varepsilon \left(\varepsilon + \frac{i}{n}\right)^{i-1} \left(1-\varepsilon - \frac{i}{n}\right)^{n-i}$$

$[n(1-\varepsilon)]$ étant la partie entière de $n(1-\varepsilon)$, $D(n, \alpha)$ est telle que :

$$K_n(D(n, \alpha)) = \frac{\alpha}{2}.$$

Ce résultat est dû à Z.W Birnbaum et F.H. Tingey [4].

'DEB
'REE
'COM
CONG
XIN+
A PA
REMA
DONN
UTIL
MEME

b) Pour n supérieur à 20.

$$D(n, \alpha) = \frac{L^{-1}(1-\alpha)}{\sqrt{n}}$$

la fonction L(z) étant la fonction de répartition limite du théorème 3.2 :

$$\Pr[\sqrt{n} D_n < z] \xrightarrow[n \rightarrow \infty]{} L(z)$$

CETT
L OR
'DEBI

avec

$$L(z) = \begin{cases} 0 & , z \leq 0 \\ \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2 z^2} & , z > 0 \end{cases}$$

3.4 - Procédure KOLMOG

Z
'FCOI

Les paramètres formels de la procédure KOLMOG sont au nombre de trois :

'PROC
'VALE
'ENTI
'TABL
'REEL
'DEBU

- N : nombre d'observations.
- X : vecteur des observations.
- ALPHA : niveau de signification.

On peut diviser le programme en trois parties :

- calcul de D_n ;
- calcul de $D(n, \alpha)$ en employant soit la procédure L lorsque le nombre d'observations est supérieur à 20, soit la procédure KS dans le cas contraire ;
- édition des résultats.

BC

F
'
'
'
'

3.5 - Exemple numérique

La procédure PRM génère des nombres pseudo-aléatoires par congruence multiplicative c'est-à-dire suivant la relation :

$$x_{n+1} \equiv k \cdot x_n \pmod{P}$$

avec :

$$P = 2^{35}$$

$$k = 3125$$

En choisissant le générateur x_0 égal à 77777 nous générons 100 nombres composant le vecteur X.

En prenant 100, X et 0.05 comme paramètres effectifs de KOLMOG, il y a éditon d'un message indiquant que l'adéquation à la loi uniforme sur (0,1) est jugée acceptable à un niveau de signification de 5%.

```

'DEBUT'
'REEL' 'PROCEDURE' PRM(XO) FF
'ENTIER' XO FF
'COMMENTAIRE' CETTE PROCEDURE GENERE DES NOMBRES ALEATOIRES PAR
CONGRUENCE MULTIPLICATIVE (C EST A DIRE SUIVANT LA RELATION
X(N+1)=K*X(N) (MODULE P) .NOUS AVONS PRIS P=2**35,K=3125)
A PARTIR DES GENERATEURS SUIVANTS 77777,14475.
REMARQUE LE NOMBRE DE DEPART (OU GENERATEUR) CHOISI PARI LA LISTE
DONNEE CI-DESSUS DOIT ETRE CHARGE A L EXTERIEUR DES BOUCLE DE CALCUL
UTILISANT CETTE PROCEDURE ,SI L ON NE VEUT PAS OBTENIR TOUJOURS LE
MEME NOMBRE. EXEMPLE D UTILISATION
NO= 77777

```

```

X=PRM(NO)
CETTE PROCEDURE ETANT ECRITE EN CODE DOIT ETRE DECLAREE LA PREMIERE DANS
L ORDRE STATIQUE DES DECLARATIONS DE PROCEDUREFF
'DEBUT' 'ENTIER' A FF 'ENTIER' 'PROCEDURE' C(X) FF
'ENTIER' X FF

```

'CODE'

```

TSX      P2-1,4
SXA      F2+1,2
LDG*     F2+1
MPY      Z
LLS      35
STO*     F2+1
LLS      2
ARS      19
STO      F2
TRA      Z+1
Z        OCT      6065

```

```

'FCODE'FF
AF=262144 FF
PRM=C(XO)/A FF
'FIN' FF
'PROCEDURE' KOLMOG(N,X,ALPHA) FF
'VALEUR' N,ALPHA FF
'ENTIER' N FF
'TABLEAU' X FF
'REEL' ALPHA FF
'DEBUT' 'TABLEAU' XR.(1FN). FF
'REEL' D,FE,X1,X2 FF
'ENTIER' I,J,K FF
'BOOLEEN' TEST FF
'REEL' 'PROCEDURE' L(Z) FF
'VALEUR' Z FF 'REEL' Z FF
'DEBUT' 'ENTIER' K FF 'REEL' A,T FF
TF=-EXP(-2*Z**2) FF
KF=1 FF
BOUCLEFKF=K+1 FF
AF=2*(K*Z)**2 FF

```

```

      'ALLERA' FINKS
'FIN' FF
'SI' EPS=1.0 'ALORS'
'DEBUT' KSF=0.0 FF
      'ALLERA' FINKS
'FIN' FF
ENTF=ENTIER(N*(1-EPS)) FF
TOTF=(1-EPS)**N/EPS FF
CTEF=1.0 FF
'POUR' NUF=1 'PAS' 1 'JUSQUA' ENT 'FAIRE'
'DEBUT' CTEF=CTE*(N-NU+1)/NU FF
      AUXF=(EPS+NU/N)**(NU-1)*(1-EPS-NU/N)**(N-NU) FF
      TOTF=TOT+CTE*AUX
'FIN' FF
KSF=EPS*TOT FF
FINKSF'FIN' FF
XR.(1).F=X.(1). FF
'POUR' IF=2 'PAS' 1 'JUSQUA' N 'FAIRE'
'DEBUT' 'SI' XR.(I-1). 'INFEG' X.(I). 'ALORS'
'DEBUT' XR.(I).F=X.(I). FF 'ALLERA' TERM 'FIN' FF
JF=0 FF
AUGFJF=J+1 FF
'SI' XR.(J). 'INFEG' X.(I). 'ALORS' 'ALLERA' AUG FF
'POUR' KF=1 'PAS' -1 'JUSQUA' J+1 'FAIRE'
XR.(K).F=XR.(K-1). FF
XR.(J).F=X.(I). FF
TERMF'FIN' FF
DF=0.0 FF
KF=0 FF
'POUR' IF=1 'PAS' 1 'JUSQUA' N 'FAIRE'
'DEBUT' 'SI' I 'INFER' K 'ALORS' 'ALLERA' SAUT FF
      FEF=ABS(XR.(I).-(I-1)/N) FF
'SI' FE 'SUPER' D 'ALORS' DF=FE FF
JF=0 FF
REVFJF=J+1 FF
'SI' XR.(I+J). =XR.(I). 'ALORS' 'ALLERA' REV FF
KF=I+J FF
FEF=ABS(XR.(I).-(K-1)/N) FF
'SI' FE 'SUPER' D 'ALORS' DF=FE FF
SAUTF'FIN' FF
'SI' N 'SUPER' 20 'ALORS' 'ALLERA' GRANDA FF
X1F=0.0 FF X2F=1.0 FF
RETFEF=(X1+X2)/2 FF
'SI' KS(FE) -ALPHA/2 'SUPER' 0.0 'ALORS'
X1F=FE 'SINON' X2F=FE FF
'SI' X2-X1 'SUPER' F*-8 'ALORS' 'ALLERA' RET FF
FEF=(X1+X2)/2 FF
TESTF='SI' D 'SUPEG' FE 'ALORS' 'FAUX' 'SINON' 'VRAI' FF
'ALLERA' SORTIE FF
GRANDAFTESTF='SI' L(RAC2(N)*D) 'SUPEG' 1-ALPHA 'ALORS' 'FAUX'
'SINON' 'VRAI' FF
SORTIEF 'SI' 'NON' TEST 'ALORS'
      ECRIRE('('ADEQUATION INAC'),'('CEPTABLE'))
'SINON'
      ECRIRE('('ADEQUATION ACCE'),'('PTABLE')) FF
'FIN' FF
'ENTIER' I,N,NO FF
'REEL' ALPHA FF
LIRE(N,ALPHA) FF

```

NOF=77777 FF

'DEBUT' 'TABLEAU' X.(IFN). FF

'POUR' IF=1 'PAS' 1 'JUSQUA' N 'FAIRE'

X.(I).F=PRM(NO) FF

KOLMOG(N,X,ALPHA)

'FIN' FF

'FIN' FF

4 - DETERMINATION DE LA TAILLE DE L'ECHANTILLON
DANS UN TEST DE STUDENT

4.1 - Introduction

Un problème que l'on rencontre fréquemment en statistique est celui de la taille requise d'un échantillon pour obtenir un résultat déterminé. Nous nous proposons ici de donner huit tables qui permettront la résolution du problème suivant :

disposant d'une estimation préalable s_0^2 avec $n_0 - 1$ degrés de liberté de la variance d'une population normale, quel doit être l'effectif n de l'échantillon à prendre pour que, dans le test de l'hypothèse $H_0 : m = m_0$ contre l'hypothèse $H : m > m_0$ au niveau de signification α , on accepte H_0 avec la probabilité β alors qu'en fait $m = m_0 + a$?

4.2 - Tables

Ce problème a été traité numériquement dans les seuls cas particuliers $\alpha = 0.05 ; \beta = 0.20$ et $\alpha = 0.05 ; \beta = 0.05$ par M. Harris, D.G. Horvitz et A.M. Mood [11]. Nous avons écrit une procédure complète calquée sur celle qu'ils utilisèrent et qui permet de trouver la solution pour tous les couples (α, β) [13]. Nous en rappellerons l'essentiel au paragraphe 4.4.

Nous donnons ici les tables qui correspondent à $\alpha = 0.01 ; \beta = 0.50, 0.20, 0.10, 0.05$ et $\alpha = 0.05 ; \beta = 0.50, 0.20, 0.10, 0.05$.

Dans un article récent [15] E. Morice a établi les graphiques qui fournissent la solution de notre problème pour $\alpha = 0.05 ; \beta = 0.20$ et $\alpha = 0.05 ; \beta = 0.05$.

4.3 - Utilisation

Soit donc à déterminer le nombre $n-1$, nombre de degrés de liberté lors de la seconde évaluation s^2 de la variance.

Calculons d'abord le rapport

$$k = a/s_0$$

puis cherchons dans la table cette valeur qui se trouve à l'intersection de la colonne qui correspond à $n_0 - 1$ et de la ligne qui correspond au $n - 1$ cherché. La taille de l'échantillon à prendre est alors n .

Lorsque k est inférieur à la dernière entrée de la colonne d'entête $m - 1$ la taille requise est alors supérieure à 120 mais on peut la déterminer de la manière suivante :

$$n - 1 = 100(K/k)^2$$

où K est la valeur tabulée pour $n - 1 = 120$ dans la colonne d'en-tête $n_0 - 1$.

Signalons que dans les tables l'interpolation est pratiquement linéaire par rapport à l'inverse de $n_0 - 1$ et également par rapport à l'inverse de la racine carrée de $n - 1$.

Les colonnes qui correspondent à $n - 1$ infini ont été obtenues à partir des tables de la fonction puissance du test de Student de Neyman et Tokarska [16].

4.4 - Problème théorique

Soit une population normale de variance inconnue, on se propose de tester l'hypothèse $H_0 : m = m_0$ contre l'hypothèse $H : m > m_0$, au niveau de signification α .

La méthode usuelle consiste alors à accepter H_0 si la condition suivante est réalisée :

$$\frac{\bar{x} - m_0}{s} \sqrt{n} < t_{1-\alpha}(n-1),$$

\bar{x} et s^2 étant la moyenne et la variance empiriques estimées à l'aide d'un échantillon de taille n et $t_{1-\alpha}(n-1)$ la valeur pour $1-\alpha$ de la variable de Student à $n-1$ degrés de liberté.

Avoir une probabilité β d'accepter H_0 , si en fait $m = m_0 + a$, revient à écrire :

$$\beta = \Pr \left[\frac{\bar{x} - m_0}{s} \sqrt{n} < t_{1-\alpha}(n-1) \right]$$

$$= \Pr \left[\frac{\bar{x} - m}{s} \sqrt{n} + \frac{a}{s} \sqrt{n} < t_{1-\alpha}(n-1) \right]$$

où la probabilité dépend à la fois de m , connu, et σ , inconnu.

Or on dispose d'une information sur σ , de la forme d'une estimation préalable s_0^2 avec n_{0-1} degrés de liberté.

On peut alors écrire :

$$\beta = \Pr [t < t_{1-\alpha}(n-1) - D\sqrt{F}\sqrt{n}]$$

en posant $F = \frac{s_0^2}{s^2}$, variable qui suit une loi de Fisher-Snedecor à n_0-1 et $n-1$ degrés de liberté, et $D = \frac{a}{s_0}$.

C'est la relation utilisée par M. Harris, D.G. Horvitz et A.M. Mood [11], qui ont calculé les valeur de

$D = \frac{m-m_0}{s_0}$ correspondant à des valeurs données de n_0 et n , pour $\alpha = 0.05$; $\beta = 0.05$ et 0.

Quand D est fixé la loi de probabilité qui intervient dans l'équation ci-dessus es la loi des deux variables t et F conditionnée par s_0 et on peut démontrer en utilisant u nouveau changement de variables que considérée comme fonction de D , $\Pr [t < t_{1-\alpha}(n) - D\sqrt{F}\sqrt{n}]$ ne dépend pas de σ et est monotone et strictement croissante.

Il suffit donc de rechercher deux valeurs encadrant la solution, qu'un banal proce sus de bisection permet alors de calculer.

4.5 - Comparaison avec la méthode de Stein

Nous pouvons utiliser une autre méthode bien connue qui est due à C. Stein [19].

Le premier échantillon de taille n_0 permet le calcul de s_0^2 . La taille du second échantillon à prélever est alors $n-n_0$, le nombre entier n étant défini au moyen de l'éga lité :

$$n = \max \left\{ \left[\frac{s_0^2}{z} \right] + 1, n_0 + 1 \right\}.$$

où z est une constante positive qui, ainsi que nous le verrons plus loin, dépend directement de β ; $|q|$ représente la partie entière du nombre q .

Il faut ensuite choisir les nombres réels b_i ($i = 1, \dots, n$) tels que :

$$\sum_{i=1}^n b_i = 1 ; b_1 = b_2 = \dots = b_{n_0} ; \sum_{i=1}^n b_i^2 = \frac{z}{s_0^2}$$

Introduisons maintenant la statistique t définie par :

$$t = \frac{\sum_{i=1}^n b_i x_i - m_0}{\sqrt{z}} = \frac{\sum_{i=1}^n b_i x_i - m}{\sqrt{z}} + \frac{m - m_0}{\sqrt{z}}$$

où :

$$u = \frac{\sum_{i=1}^n b_i x_i - m}{\sqrt{z}}$$

a la distribution de Student à $n_0 - 1$ degrés de liberté.

Pour tester $H_0 : m = m_0$ contre $H : m > m_0$, la région critique est donnée par :

$$\frac{\sum_{i=1}^n b_i x_i - m_0}{\sqrt{z}} > t_{1-\alpha}(n_0 - 1)$$

Accepter H_0 avec une probabilité β , si en fait $m = m_0 + a$, revient à écrire :

$$\beta = \Pr \left[u + \frac{m - m_0}{\sqrt{z}} < t_{1-\alpha}(n_0 - 1) \right]$$

Cette relation est utilisée pour déterminer la valeur de la constante z , qui permet ensuite de trouver la taille requise.

A titre comparatif pour $n_0 = 7$, $s_0^2 = 4$, $a = 3.7$, $\alpha = 0.01$ et $\beta = 0.10$ nous avons obtenu par :

a) la méthode de Stein $n - n_0 = 1$

b) la méthode de Harris, Horvitz et Mood $n = 9$.

Nous pouvons remarquer que la première méthode tient certainement plus compte de l'information apportée par le premier échantillon. Par contre elle introduit une complication supplémentaire puisqu'elle nécessite le calcul des nombres b_i dépendant de n_0 , n et $\frac{z}{s_0}$.

4.6 - Test bilatéral

Les tables sont encore utilisables si l'on désire effectuer un test bilatéral, c'est-à-dire tester l'hypothèse $H_0 : m = m_0$ contre l'hypothèse $H : m \neq m_0$. Dans ce cas le niveau de signification est 2α et les tables correspondent à $\alpha = 0.02 ; \beta = 0.50, 0.20, 0.10, 0.05$ et $\alpha = 0.10 ; \beta = 0.50, 0.20, 0.10, 0.05$.

4.7 - Application : différence entre deux moyennes

Considérons maintenant deux populations normales de même variance mais de moyennes respectives m_1 et m_2 inconnues. Nous nous proposons de résoudre le problème suivant :

disposant d'une estimation préalable s_0^2 avec $n_0 - 1$ degrés de liberté de la variance de l'une quelconque des deux populations, quel doit être l'effectif commun des deux échantillons à prendre pour que, dans le test de l'hypothèse $H_0 : m_1 = m_2$ contre l'hypothèse $H : m_1 > m_2$ au niveau de signification α , on accepte H_0 avec la probabilité β alors qu'en fait $m_1 = m_2 + b$?

L'intervalle d'acceptation de l'hypothèse H_0 est alors :

$$\frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{s_1^2 + s_2^2}{2}}} \sqrt{n} < t_{1-\alpha}(2n-2),$$

\bar{x}_i et s_i^2 étant la moyenne et la variance empiriques estimées à l'aide d'un échantillon de taille n tiré de la population i ($i = 1, 2$).

A partir de là le problème est identique à celui qui a été traité en 4.4. Pour trouver sa solution on procède de la même manière. La seule différence provient de la quantité k à trouver dans la table qui est ici définie de la façon suivante :

$$k = \frac{b}{2s_0}$$

Le nombre $n-1$ que l'on détermine ainsi est le nombre de degrés de liberté lors de l'estimation commune de la variance. La taille de l'échantillon à tirer de chacune des populations est alors $(n-1)/2+1$ si $n-1$ est pair ou $n/2-1$ si $n-1$ est impair.

Dans le cas d'un test bilatéral, c'est-à-dire le test de $H_0 : m_1 = m_2$ contre $H : m_1 \neq m_2$, le problème se transpose facilement en celui que nous avons vu en 4.6.

4.8 - Principe d'une méthode de résolution numérique

En conservant les notations introduites en 4.4, le problème peut se formuler ainsi : les risques α, β et les entiers n_0, n étant fixés déterminer la valeur D^* réalisant :

$$\Pr [t < t_{1-\alpha}(n-1) - D^* \sqrt{F} \sqrt{n}] = \beta$$

où t et F sont des variables aléatoires qui suivent respectivement une loi de Student à $n-1$ degrés de liberté et de Fisher-Snedecor à n_0-1 et $n-1$ degrés de liberté.

t et F ne sont pas indépendantes puisque toutes deux sont des fonctions de s , cependant on peut calculer la densité du couple de la manière qui suit :

La fonction densité de probabilité de l'ensemble (\bar{x}, s_0, s) est le produit des densités individuelles des trois variables, puisqu'elles sont deux à deux indépendantes.

Si maintenant nous transformons en t, F et s et si nous intégrons par rapport à s nous obtenons alors la densité du couple (t, F) :

$$f(t, F) = C \frac{F^{\frac{n_0-3}{2}}}{[t^2 + (n_0-1)F + n-1]^{\frac{n_0+n-1}{2}}}$$

C étant fonction de n_0 et n .

Pour faciliter l'intégration numérique qui est nécessaire pour évaluer la probabilité définie plus haut, faisons encore un changement de variables :

$$x^2 = \frac{(n_0+n-2)t^2}{(n_0-1)F+n-1}$$

$$y = \frac{(n_0-1)F}{(n_0-1)F+n-1}$$

La densité du couple devient alors :

$$g(x,y) = \left[\frac{1}{\sqrt{\pi(n_0+n-2)}} \frac{\Gamma(\frac{n_0+n-1}{2})}{\Gamma(\frac{n_0+n-2}{2})} \frac{1}{(1 + \frac{x^2}{n_0+n-2})^{\frac{n_0+n-1}{2}}} \right] \cdot \left[\frac{\Gamma(\frac{n_0+n-2}{2})}{\Gamma(\frac{n_0-1}{2})\Gamma(\frac{n-1}{2})} y^{\frac{n_0-3}{2}} (1-y)^{\frac{n-3}{2}} \right]$$

de telle sorte que x et y sont distribuées indépendamment, x suivant une loi de Student à n_0+n-2 degrés de liberté et y une loi Beta de paramètres $\frac{n_0-1}{2}$ et $\frac{n-1}{2}$.

4.9 - Procédure DETER

L'intégrale que nous avons à évaluer est la suivante :

$$I(D) = \int_0^1 g_2(y) \left[\int_{-\infty}^{h(D,y)} g_1(x) dx \right] dy$$

où g_1 et g_2 sont les fonctions densité de probabilité d'une loi de Student et d'une loi Beta, $h(D,y)$ étant telle que :

$$h(D,y) = t_{1-\alpha}(n-1) \cdot \sqrt{(1-y) \left(\frac{n_0-1}{n-1} + 1 \right)} - D \sqrt{ny \left(\frac{n-1}{n_0-1} + 1 \right)}$$

On peut encore écrire :

$$I(D) = \int_0^1 g_2(y) \cdot S_{n_0+n-2}(h(D,y)) \cdot dy$$

$S_n(t)$ représentant la fonction de répartition d'une loi de Student à n degrés de liberté.

$I(D)$ considérée comme fonction de la variable D , est décroissante.

Dans une première étape nous recherchons alors deux valeurs D_1 et D_2 telles que :

$$0 < D_1 < D_2$$

$$I(D_1) > \beta > I(D_2)$$

Nous déterminons ensuite la valeur D^* réalisant :

$$I(D^*) = \beta$$

$$D_1 < D^* < D_2$$

par une méthode de bisection.

La procédure DETER effectue le calcul précédent pour N_0, N et BETA donnés. Elle possède le quatrième paramètre formel T qui, au moment de l'appel de DETER, devra être remplacé par un paramètre effectif contenant la valeur $t_{1-\alpha}(n-1)$ introduite en 4.4., cette dernière pouvant soit être prise dans une table, soit être calculée par la procédure STUDENT.

NIM

FIN

```

'REEL' 'PROCEDURE' DETER(NO,N,BETA,T) FF
'VALEUR' NO,N,BETA,T FF
'ENTIER' NO,N FF 'REEL' BETA,T FF
'DEBUT'
'REEL' 'PROCEDURE' GAMMA(X) FF 'REEL' X FF
'DEBUT' 'REEL' H,Y FF
HF=1.0 FF YF=X FF
AIF 'SI' Y=0 'ALORS' HF=10**5 'SIMON'
'SI' Y=2.0 'ALORS' 'ALLERA' A2 'SIMON'
'SI' Y<INFER' 2.0 'ALORS'
'DEBUT' HF=H/Y FF YF=Y+1.0 FF 'ALLERA' A1 'FIN'
'SIMON' 'SI' Y<SUEG' 3 'ALORS' 'DEBUT'
YF=Y-1.0 FF HF=H*Y FF 'ALLERA' A1 'FIN'
'SIMON' 'DEBUT' YF=Y-2.0 FF
HF=(((((0.0016063118*Y+.0051589951)*Y+.0044511400)*Y+.0721101567)*
Y+.0821117404)*Y+.4117741955)*Y
.4227874605)*Y+.9999999758)*H
'FIN' FF
A2F GAMMAF=H
'FIN' GAMMA FF
'REEL' 'PROCEDURE' INSIROF,A,B,ORDMAX,PREC, SORT,RES) FF
'VALEUR' A,B,ORDMAX,PRECFF 'REEL' 'PROCEDURE' FFF 'BOLEEN' SORT FF
'REEL' A,B,PRECFF 'ENTIER' ORDMAX FF 'REEL' 'TABLEAU' RES FF
'DEBUT' 'REEL' L,T,P,MAF 'ENTIER' N,J,I,FACFF
LF=B-AFF SORTF='FAUX' FMAF=RES.(1).F=L*0.5*(F(A)+F(B)) FNF=1 FF
'POUR' JF=1 'PAS' 1 'JUSQUA' ORDMAX 'FAIRE'
'DEBUT' TF=0.0 FFPF=L/NF
'POUR' IF=1 'PAS' 1 'JUSQUA' N 'FAIRE' TF=T+F(A+P*(1-0.5)) FF
RES.(J+1).F=(P*T+RES.(J).)/2.0 FFPF=1 FF
'POUR' IF=J 'PAS' -1 'JUSQUA' 1 'FAIRE'
'DEBUT' FACF=4*FACFF
RES.(I).F=RES.(I+1).+(RES.(I+1).-RES.(I).)/(FAC-1)
'FIN' FF
'SI' ABS((RES.(1).-MA)/RES.(1).)>TNEG 'PREC' 'ALORS'
'ALLER A' TERM FF
MAF=RES.(1). FNF=2*N
'FIN' FF
'ALLERA' AFFECT FF
TERM SORTF='VRAI' FF
AFFECT INSIROF=RES.(1).
'FIN' FF
'REEL' 'PROCEDURE' STUDENT(N,U) FF
'VALEUR' N,U FF 'ENTIER' N FF 'REEL' U FF
'DEBUT' 'REEL' V,I,A FF
'ENTIER' M FF
VF=U/RAC2(N) FF
'SI' N-2*ENTIER(N/2)=1 'ALORS' 'ALLERA' NIMP FF
AF=V/RAC2(1+V**2) FF
IF=1+A FF

```

STUDENTF='SI' N 'INFER' 60 'ALORS' I*GAMMA((N+1)/2)/
 (GAMMA(N/2)*GAMMA(0.5)) 'SINON' I*RAC2(N/6.28318530)*EXP(-0.25/N
 1/(24*N**3)) FF

'FIN' FF

'REEL' CTE,D1,D2,D0 FF

'TABLEAU' RES.(1F9). FF

'BOOLEEN' BOOL FF

CTEF=GAMMA((NO+N)/2-1)/(GAMMA(NO/2-0.5)*GAMMA(N/2-0.5)) FF

'REEL' 'PROCEDURE' I(D) FF

'VALEUR' D FF 'REEL' D FF

'DEBUT' 'REEL' 'PROCEDURE' G(Y) FF

'VALEUR' Y FF 'REEL' Y FF

'DEBUT' 'REEL' H,G1,G2 FF

G1F='SI' NO 'SUPER' 3 'ALORS' Y**((NO-3)/2) 'SINON' 1.0 FF

G2F='SI' N 'SUPER' 3 'ALORS' Y**((N-3)/2) 'SINON' 1.0 FF

Hf=T*RAC2((1-Y)*((NO-1)/(N-1)+1))

D*RAC2(N*Y*((N-1)/(NO-1)+1)) FF

GF=STUDENT(NO+N-2,H)*G1*G2

'FIN' FF

IF=CTE*INSIRO(G,0.0,1.0,8,F*-5,BOOL,RES)

'FIN' FF

D1F=0.5 FF

D2F=1.5 FF

REVFDOF=(D1+D2)/2 FF

'SI' D2-D1 'INFER' F*-2 'ALORS' 'ALLERA' FINDETER FF

'SI' I(D0) 'INFER' BETA 'ALORS' D2F=D0 'SINON' D1F=D0 FF

'ALLERA' REV FF

FINDETERF

DETERF=D0 FF

'FIN' FF

ALPHA = 0.01

BETA = 0.50

N-1 \ NO-1	2	4	6	12	24	INFINI
2	4.022	3.670	3.565	3.464	3.413	3.366
4	1.827	1.677	1.632	1.588	1.569	1.552
6	1.329	1.219	1.187	1.157	1.145	1.130
8	1.094	1.004	0.979	0.955	0.943	0.933
12	0.852	0.782	0.762	0.743	0.736	0.727
16	0.722	0.662	0.646	0.630	0.622	0.616
20	0.640	0.587	0.572	0.559	0.552	0.546
25	0.567	0.520	0.507	0.495	0.489	0.482
30	0.512	0.471	0.458	0.448	0.442	0.438
40	0.441	0.406	0.394	0.386	0.381	0.364
50	0.393	0.361	0.351	0.343	0.339	0.326
60	0.357	0.328	0.319	0.312	0.308	0.298
80	0.308	0.283	0.276	0.269	0.266	0.259
100	0.275	0.253	0.246	0.240	0.237	0.232
120	0.251	0.230	0.224	0.219	0.216	0.212

ALPHA = 0.01

BETA = 0.20

N-1 \ NO-1	2	4	6	12	24	INFINI
2	8.104	6.393	5.940	5.535	5.349	5.173
4	3.528	2.746	2.537	2.350	2.263	2.182
6	2.544	1.976	1.823	1.686	1.623	1.561
8	2.088	1.619	1.494	1.381	1.329	1.280
12	1.625	1.258	1.161	1.072	1.033	0.993
16	1.375	1.065	0.982	0.908	0.873	0.842
20	1.219	0.941	0.870	0.805	0.774	0.742
25	1.079	0.833	0.771	0.712	0.685	0.657
30	0.976	0.755	0.697	0.644	0.619	0.596
40	0.841	0.651	0.600	0.554	0.533	0.495
50	0.748	0.580	0.534	0.493	0.474	0.444
60	0.680	0.528	0.485	0.448	0.431	0.406
80	0.585	0.455	0.419	0.387	0.372	0.352
100	0.521	0.406	0.374	0.345	0.332	0.315
120	0.473	0.370	0.341	0.315	0.303	0.288

ALPHA = 0.01

BETA = 0.10

N-1 \ NO-1	2	4	6	12	24	INFINI
2	12.173	8.470	7.588	6.838	6.506	6.201
4	5.245	3.568	3.166	2.820	2.666	2.522
6	3.776	2.557	2.265	2.011	1.899	1.792
8	3.096	2.092	1.850	1.642	1.550	1.463
12	2.406	1.625	1.437	1.274	1.201	1.134
16	2.038	1.376	1.215	1.078	1.016	0.958
20	1.807	1.219	1.077	0.955	0.900	0.847
25	1.600	1.079	0.953	0.845	0.796	0.749
30	1.447	0.975	0.862	0.764	0.720	0.679
40	1.241	0.840	0.742	0.658	0.620	0.564
50	1.100	0.747	0.660	0.585	0.551	0.506
60	0.996	0.679	0.600	0.532	0.501	0.462
80	0.849	0.587	0.518	0.460	0.433	0.401
100	0.748	0.524	0.462	0.410	0.386	0.359
120	0.674	0.477	0.421	0.374	0.352	0.328

ALPHA = 0.01

BETA = 0.05

N-1 \ NO-1	2	4	6	12	24	INFINI
2	17.698	10.743	9.256	8.063	7.541	7.078
4	7.589	4.474	3.805	3.259	3.023	2.808
6	5.460	3.199	2.711	2.312	2.141	1.980
8	4.475	2.616	2.213	1.885	1.744	1.613
12	3.481	2.030	1.178	1.462	1.349	1.248
16	2.949	1.717	1.452	1.235	1.141	1.055
20	2.601	1.522	1.286	1.094	1.011	0.932
25	2.290	1.347	1.138	0.969	0.895	0.824
30	2.061	1.218	1.029	0.876	0.809	0.747
40	1.749	1.049	0.886	0.754	0.696	0.620
50	1.536	0.933	0.789	0.671	0.620	0.556
60	1.379	0.848	0.717	0.610	0.563	0.508
80	1.196	0.734	0.619	0.527	0.486	0.441
100	1.071	0.656	0.552	0.470	0.434	0.395
120	0.979	0.598	0.503	0.428	0.395	0.361

ALPHA = 0.05

BETA = 0.50

$N-1 \backslash N0-1$	2	4	6	12	24	INFINI
2	1.687	1.552	1.509	1.475	1.455	1.438
4	1.032	0.952	0.931	0.909	0.898	0.890
6	0.815	0.753	0.736	0.718	0.710	0.703
8	0.694	0.642	0.627	0.613	0.605	0.600
12	0.561	0.518	0.506	0.495	0.490	0.483
16	0.483	0.447	0.436	0.426	0.420	0.417
20	0.431	0.399	0.389	0.380	0.375	0.371
25	0.384	0.356	0.347	0.339	0.335	0.331
30	0.350	0.324	0.316	0.309	0.306	0.302
40	0.303	0.280	0.273	0.267	0.264	0.256
50	0.271	0.250	0.244	0.239	0.236	0.230
60	0.247	0.228	0.223	0.218	0.215	0.210
80	0.214	0.197	0.193	0.189	0.186	0.182
100	0.192	0.176	0.172	0.169	0.167	0.163
120	0.175	0.161	0.157	0.154	0.152	0.149

ALPHA = 0.05

BETA = 0.20

N-1 \ NO-1	2	4	6	12	24	INFINI
2	3.510	2.805	2.620	2.448	2.371	2.298
4	2.098	1.669	1.554	1.452	1.405	1.360
6	1.649	1.311	1.220	1.139	1.101	1.066
8	1.405	1.117	1.041	0.970	0.939	0.910
12	1.114	0.902	0.840	0.784	0.758	0.732
16	0.976	0.775	0.722	0.673	0.651	0.631
20	0.873	0.693	0.646	0.602	0.583	0.563
25	0.779	0.619	0.577	0.538	0.520	0.502
30	0.708	0.563	0.525	0.489	0.473	0.456
40	0.613	0.486	0.454	0.423	0.409	0.395
50	0.548	0.435	0.405	0.378	0.365	0.353
60	0.499	0.396	0.369	0.344	0.333	0.322
80	0.432	0.342	0.319	0.298	0.288	0.278
100	0.385	0.306	0.285	0.266	0.257	0.249
120	0.351	0.279	0.260	0.243	0.235	0.227

ALPHA = 0.05

BETA = 0.10

N-1 \ NO-1	2	4	6	12	24	INFINI
2	5.308	3.747	3.372	3.051	2.909	2.778
4	3.152	2.204	1.975	1.779	1.691	1.610
6	2.474	1.727	1.547	1.393	1.324	1.259
8	2.109	1.471	1.317	1.186	1.127	1.070
12	1.701	1.184	1.059	0.953	0.906	0.863
16	1.464	1.020	0.913	0.821	0.780	0.742
20	1.307	0.910	0.815	0.733	0.696	0.661
25	1.167	0.812	0.727	0.653	0.621	0.590
30	1.063	0.739	0.662	0.595	0.565	0.537
40	0.920	0.639	0.572	0.514	0.489	0.458
50	0.822	0.571	0.511	0.460	0.436	0.410
60	0.750	0.521	0.466	0.419	0.398	0.375
80	0.650	0.451	0.403	0.363	0.344	0.326
100	0.581	0.403	0.361	0.324	0.308	0.292
120	0.531	0.368	0.329	0.296	0.281	0.266

ALPHA = 0.05

BETA = 0.05

N-1 \ NO-1	2	4	6	12	24	INFINI
2	7.734	4.771	4.129	3.610	3.385	3.187
4	4.582	2.789	2.397	2.080	1.944	1.820
6	3.595	2.182	1.876	1.623	1.516	1.417
8	3.066	1.858	1.597	1.382	1.290	1.207
12	2.470	1.495	1.285	1.112	1.038	0.971
16	2.129	1.287	1.106	0.956	0.892	0.834
20	1.900	1.149	0.991	0.858	0.798	0.744
25	1.694	1.026	0.884	0.765	0.712	0.663
30	1.543	0.935	0.804	0.695	0.647	0.605
40	1.338	0.809	0.696	0.601	0.560	0.525
50	1.198	0.722	0.622	0.537	0.500	0.469
60	1.095	0.658	0.567	0.490	0.456	0.428
80	0.943	0.569	0.490	0.423	0.395	0.369
100	0.840	0.508	0.438	0.378	0.353	0.329
120	0.767	0.465	0.399	0.345	0.322	0.300

5 - TESTS D'HYPOTHESES LINEAIRES

5.1 - Introduction

Un problème où les calculateurs se révèlent indispensables pour améliorer les résultats connus est celui du test de plusieurs hypothèses linéaires.

Il s'agit en fait d'un problème de décision multiple [8] et le calcul que nous devons effectuer est celui de la fonction de répartition d'un vecteur aléatoire, et non plus d'une simple variable comme c'est le cas avec un test ordinaire.

5.2 - Méthode générale

a) Soient n variables indépendantes X_1, \dots, X_n suivant des lois normales de moyennes respectives m_1, \dots, m_n et de variance commune σ^2 .

On suppose que le vecteur des moyennes \vec{m} appartient à un sous-espace vectoriel V de R^n .

Soient alors W_1, \dots, W_k k sous-espaces vectoriels de V que l'on se bornera à choisir deux à deux orthogonaux.

On se propose, au vu d'une observation \vec{x} de $\vec{X} = (X_1, \dots, X_n)$ de répondre aux k questions : le vecteur \vec{m} admet-il une composante nulle sur W_i ? ($i = 1, \dots, k$).

b) On définit V^* comme étant l'orthogonal de V .

Désignons par $m_{V^*}, m_{W_1}, \dots, m_{W_k}$ les projections orthogonales respectivement sur V^*, W_1, \dots, W_k du vecteur \vec{m} .

Chaque décision multiple, parmi les 2^k que l'on peut prendre, est composée de k décisions individuelles qui sont l'acceptation ou le rejet de l'hypothèse H_i :

$m_{W_i} = 0$ ($i = 1, \dots, k$).

Désignons maintenant par $X_{V^*}, X_{W_1}, \dots, X_{W_k}$ les projections orthogonales du vecteur \vec{X} de R^n sur V^*, W_1, \dots, W_k .

On forme successivement les différents rapports :

$$R_i = \frac{\|X_{W_i}\|^2}{\|X_{V^*}\|^2}, \quad i = 1, \dots, k.$$

Si $m_{V^*} = 0$, hypothèse réalisée puisque \vec{m} est supposée appartenir à V , R_i suit une loi de Tang $T\left(\frac{\|m_{W_i}\|^2}{2}; r, s\right)$, où r représente la dimension de W_i , et s celle de V^* . En effet les deux variables aléatoires :

$$U_i = \frac{\|X_{W_i}\|^2}{\sigma^2},$$

$$U = \frac{\|X_{V^*}\|^2}{\sigma^2}$$

suivent respectivement une loi du χ^2 décentrée à r degrés de liberté avec un paramètre de décentrage $\frac{\|m_{W_i}\|^2}{\sigma^2}$, et une loi du χ^2 réduite à s degrés de liberté, et on peut montrer que leur rapport σ^2 suit la loi de Tang ci-dessus.

La méthode va alors consister à rejeter l'hypothèse H_i lorsque les observations satisfont à :

$$R_i \geq \rho_i$$

où le coefficient ρ_i est déterminé de la manière suivante :

$$\Pr \left[\frac{\|X_{W_i}\|^2}{\|X_{V^*}\|^2} < \rho_i \mid m_{W_i} = 0 \right] = 1 - \alpha$$

α étant un nombre donné qui joue le rôle de niveau de signification.

Puisque l'on suppose m_{W_i} nul, le rapport :

$$\frac{\|X_{W_i}\|^2}{\|X_{V^*}\|^2}$$

qui intervient dans la probabilité ci-dessus, suit une loi de Fisher de paramètres r et s .

5.3 - Application numérique : cas de deux hypothèses

Reprenons le problème posé en 5.2.-a dans le cas simple où k est égal à 2.

Soient alors n_1, n_2, n_3 les dimensions respectives des sous-espaces vectoriels W_1, W_2, V^* . Introduisons les trois variables aléatoires U_1, U_2, U_3 telles que :

$$U_1 = \frac{\|X_{W_1}\|^2}{\sigma^2}$$

$$U_2 = \frac{\|X_{W_2}\|^2}{\sigma^2}$$

$$U_3 = \frac{\|X_{V^*}\|^2}{\sigma^2}$$

Elles suivent respectivement une loi du χ^2 décentrée à n_1 degrés de liberté avec 1 paramètre de décentrage p_1 , une loi identique à n_2 degrés de liberté avec le paramètre de décentrage p_2 , une loi du χ^2 réduite à n_3 degrés de liberté.

Les paramètres p_1 et p_2 sont tels que :

$$p_1 = \frac{\|m_{W_1}\|^2}{\sigma^2}$$

$$p_2 = \frac{\|m_{W_2}\|^2}{\sigma^2}$$

Déterminons les deux coefficients ρ_1 et ρ_2 de la façon suivante :

$$\Pr \left[\frac{U_1}{U_3} < \rho_1 \mid m_{W_1} = 0 \right] = 1 - \alpha$$

$$\Pr \left[\frac{U_2}{U_3} < \rho_2 \mid m_{W_2} = 0 \right] = 1 - \alpha$$

Le test va s'effectuer par l'intermédiaire des deux rapports non indépendants $\frac{u_1}{u_3}$ et $\frac{u_2}{u_3}$, calculés à partir de l'observation \vec{x} : nous acceptons l'hypothèse H_1 : $m_{W_1} = 0$ lorsque $\frac{u_1}{u_3}$ est inférieur à ρ_1 et l'hypothèse H_2 : $m_{W_2} = 0$ lorsque $\frac{u_2}{u_3}$ est inférieur à ρ_2 .

On voit immédiatement que seize situations différentes peuvent apparaître, avec les probabilités $P_{i,j}$ ($i,j = 1,2,3,4$) :

Décision Hypothèse	$m_{W_1} = 0 ;$ $m_{W_2} = 0$	$m_{W_1} = 0 ;$ $\ m_{W_2}\ > 0$	$\ m_{W_1}\ > 0 ;$ $m_{W_2} = 0$	$\ m_{W_1}\ > 0$ $\ m_{W_2}\ > 0$
$m_{W_1} = 0 ; m_{W_2} = 0$	$P_{1,1}$	$P_{1,2}$	$P_{1,3}$	$P_{1,4}$
$m_{W_1} = 0 ; \ m_{W_2}\ > 0$	$P_{2,1}$	$P_{2,2}$	$P_{2,3}$	$P_{2,4}$
$\ m_{W_1}\ > 0 ; m_{W_2} = 0$	$P_{3,1}$	$P_{3,2}$	$P_{3,3}$	$P_{3,4}$
$\ m_{W_1}\ > 0 ; \ m_{W_2}\ > 0$	$P_{4,1}$	$P_{4,2}$	$P_{4,3}$	$P_{4,4}$

Ces différentes probabilités dépendent de n_1, n_2, n_3, α (qui déterminent entièrement ρ_1 et ρ_2) mais aussi des deux paramètres de décentrage p_1 et p_2 . Elles sont fonction des vraies lois de probabilité des deux variables aléatoires U_1 et U_2 .

On peut alors, pour chaque valeur du couple (a_1, a_2) , définir les probabilités des éventualités qui peuvent se présenter en testant simultanément :

$$\left| \begin{array}{l} H_1 : p_1 = 0 \text{ contre } H_1^* : p_1 > 0 \\ H_2 : p_2 = 0 \text{ contre } H_2^* : p_2 > 0 \end{array} \right.$$

La procédure TESTLIN1 réalise le calcul des valeurs des deux séparatrices ρ_1' et ρ_2 , ainsi que des seize probabilités du précédent tableau en fonction des paramètres n_1, n_2, n_3, a_1, a_2 et α .

Nous indiquons ci-dessous les résultats obtenus en prenant :

$$n_1 = n_2 = 10$$

$$n_3 = 5$$

$$a_1 = a_2 = 5$$

$$\alpha = 5 \%$$

Décision Hypothèse	$P_1 = 0 \quad P_2 = 0$	$P_1 = 0 \quad P_2 > 0$	$P_1 > 0 \quad P_2 = 0$	$P_1 > 0 \quad P_2 > 0$
$P_1 = 0 \quad P_2 = 0$	0.927	0.023	0.023	0.027
$P_1 = 0 \quad P_2 = a_2$	0.881	0.069	0.012	0.038
$P_1 = a_1 \quad P_2 = 0$	0.881	0.012	0.069	0.038
$P_1 = a_1 \quad P_2 = a_2$	0.849	0.044	0.044	0.063

Pour la ligne i du tableau, $P_{i,i}$ représente la probabilité de donner une réponse qui soit exacte. Exception faite pour $P_{1,1}$, on remarque que les éléments de la diagonale sont faibles. Cela provient du fait que le test que nous avons décrit ne permet pas de distinguer objectivement des hypothèses voisines et que, dans de telles conditions, les valeurs que nous avons choisies pour a_1 et a_2 sont trop proches de zéro.

Plusieurs essais nous ont d'ailleurs permis de constater que la qualité du test s'améliore en même temps que a_1 et a_2 augmentent.

Considérons alors les deux fonctions $\beta_1(a_1)$ et $\beta_2(a_2)$ telles que :

$$\beta_1(a_1) = \Pr \left[\frac{U_1}{U_3} > \rho_1 \mid p_1 = a_1 \right] ;$$

$$\beta_2(a_2) = \Pr \left[\frac{U_2}{U_3} > \rho_2 \mid p_2 = a_2 \right]$$

Nous pouvons rechercher les deux nombres a_1^* et a_2^* réalisant :

$$\beta_1(a_1^*) = 1-\alpha ;$$

$$\beta_2(a_2^*) = 1-\alpha$$

La probabilité de rejeter avec raison l'hypothèse H_1 (resp. H_2) lorsque $p_1 = a_1^*$ (resp $p_2 = a_2^*$) est égale à $1-\alpha$.

A partir de n_1, n_2, n_3 et α le calcul des valeurs des deux séparatrices ρ_1 et ρ_2 , des deux nombres a_1^* et a_2^* , enfin des seize probabilités est effectué par la procédure TESTLIN dont nous donnons à la fin du paragraphe un listage.

En prenant pour n_1, n_2, n_3 et α les mêmes valeurs que précédemment, nous avons obtenus les résultats ci-dessous. Ils montrent qu'en prenant a_1 et a_2 supérieurs à 104 il devient possible de distinguer valablement les hypothèses.

$$a_1^* = a_2^* = 104.$$

Décision Hypothèse	$p_1 = 0 \quad p_2 = 0$	$p_1 = 0 \quad p_2 > 0$	$p_1 > 0 \quad p_2 = 0$	$p_1 > 0 \quad p_2 > 0$
$p_1 = 0 \quad p_2 = 0$	0.927	0.023	0.023	0.027
$p_1 = 0 \quad p_2 = a_2^*$	0.047	0.901	0.001	0.051
$p_1 = a_1^* \quad p_2 = 0$	0.047	0.001	0.901	0.051
$p_1 = a_1^* \quad p_2 = a_2^*$	0.028	0.021	0.021	0.930

Remarque 1

Le calcul des éléments d'une même ligne s'effectue de la façon suivante.

$$P_{i,1} = \Pr \left[\frac{U_1}{U_3} < \rho_1, \frac{U_2}{U_3} < \rho_2 \mid P_1, P_2 \right]$$

$$P_{i,2} = \Pr \left[\frac{U_1}{U_3} < \rho_1, \frac{U_2}{U_3} > \rho_2 \mid P_1, P_2 \right]$$

$$P_{i,3} = \Pr \left[\frac{U_1}{U_3} > \rho_1, \frac{U_2}{U_3} < \rho_2 \mid P_1, P_2 \right]$$

$$P_{i,4} = \Pr \left[\frac{U_1}{U_3} > \rho_1, \frac{U_2}{U_3} > \rho_2 \mid P_1, P_2 \right]$$

Détaillons à titre d'exemple le calcul de $P_{i,3}$:

$$P_{i,3} = \Pr \left[U_1 > \rho_1 \cdot U_3, U_2 < \rho_2 \cdot U_3 \mid P_1, P_2 \right]$$

En introduisant $f_1(u_1)$, $f_2(u_2)$, $f_3(u_3)$ les fonctions densité de probabilité respectivement de U_1 , U_2 , U_3 on peut écrire :

$$\begin{aligned} P_{i,3} &= \int_0^{\infty} \left[\int_{\rho_1 \cdot u_3}^{\infty} f_1(u_1) du_1 \right] \left[\int_0^{\rho_2 \cdot u_3} f_2(u_2) du_2 \right] f_3(u_3) du_3 \\ &= \int_0^{\infty} [1 - K_{n_1, P_1}(\rho_1 \cdot u_3)] K_{n_2, P_2}(\rho_2 \cdot u_3) \cdot f_3(u_3) du_3 \end{aligned}$$

où $K_{n,p}(x)$ représente la fonction de répartition de la loi du χ^2 décentrée à n degrés de liberté avec un paramètre de décentrage p.

Remarque 2

Ce problème, ainsi que plusieurs autres que nous étudierons plus loin, nécessite pour sa réalisation un calcul d'intégrale. La procédure de quadrature approchée que nous utilisons est INSIRO. Elle calcule la valeur de $\int_A^B f(x) dx$ par une méthode de Romberg (extrapolation de Richardson appliquée à la formule des trapèzes).

```

'PROCEDURE' TESTLIN(N1,N2,N3,ALPHA) FF
'ENTIER' N1,N2,N3 FF
'REEL' ALPHA FF
'DEBUT'

```

```

'REEL' 'PROCEDURE' GAMMA(X) FF 'REEL' XFF
'DEBUT' 'REEL' H,Y FF
HF=1.0 FF YF=X FF
A1F 'SI' Y=0 'ALORS' HF=10**5 'SINON'
'SI' Y=2.0 'ALORS' 'ALLERA' A2 'SINON'
'SI' Y'INFER'2.0 'ALORS'
'DEBUT' HF=H/Y FF YF=Y+1.0 FF 'ALLERA' A1 'FIN'
'SINON' 'SI' Y'SUPEG'3 'ALORS' 'DEBUT'
YF=Y-1.0 FF HF=H*Y FF 'ALLERA' A1 'FIN'
'SINON' 'DEBUT' YF=Y-2.0 FF
HF=(((((((0.0016063118*Y+.0051589951)*Y+.0044511400)*Y+.0721101567)*
Y+.0821117404)*Y+.4117741955)*Y
.4227874605)*Y+.9999999758)*H
'FIN' FF
A2F GAMMAF=H
'FIN' GAMMA FF

```

```

'REEL' 'PROCEDURE' KI(N,X) FF
'VALEUR' N,X FF
'ENTIER' N FF 'REEL' X FF
'DEBUT'

```

```

'REEL' 'PROCEDURE' NORMALE(X,MOY,ECART) FF
'VALEUR' X,MOY,ECART FF 'REEL' X,MOY,ECART FF
'DEBUT' 'ENTIER' I FF
'REEL' SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
XF=(X-MOY)/ECART FF
SGNEF=SIGNE(X) FF
XF=ABS(X) FF
X2F=X*X FF
YF=0.3989422804*EXP(-0.5*X2) FF
'SI' X 'INFER' 3.5 'ALORS'
'DEBUT' SF=XF=Y*X FF
'POUR' IF=3,I+2 'TANTQUE' S 'NONEG' T 'FAIRE'
'DEBUT' TF=S FF
XF=X*X2/I FF
SF=S+X
'FIN' FF
NORMALEF=0.5+SGNE*S

```

```

'FIN'
'SINON'
'DEBUT' Q1F=X FF
P2F=Y*X FF
P1F=Y FF
Q2F=X2+1.0 FF
MF=Y/X FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS'
'DEBUT' MF=1-M FF
TF=1-T
'FIN' FF
'POUR' IF=2,I+1 'TANTQUE'
((M 'NONEG' T) 'ET' (S 'NONEG' T)) 'FAIRE'
'DEBUT' SF=X*P2+I*P1 FF
P1F=P2 FF
P2F=S FF
SF=X*Q2+I*Q1 FF
Q1F=Q2 FF

```

Q2F=S FF
SF=M FF
MF=T FF
TF=P2/Q2 FF

'SI' SIGNE=1 'ALORS' TF=1-T FF

'FIN' FF
NORMALEF=T

FINORMF 'FIN' FF
'FIN' FF
'ENTIER' I FF
'REEL' AUX,SOM,PI FF
'SI' N-2*ENTIER(N/2)=1 'ALORS'
'ALLERA' NIMP FF
SOMF=AUXF=1.0 FF
'SI' N=2 'ALORS' 'ALLERA' FPAIR FF
'POUR' IF=1 'PAS' 1 'JUSQUA' N/2-1 'FAIRE'
'DEBUT' AUXF=AUX*X/(2*I) FF
SOMF=SOM+AUX

FPAIRF 'FIN' FF
KIF=1.0-EXP(-X/2)*SOM FF
'ALLERA' FINKI FF

NIMPF PIF=3.141592653590 FF
AUXF=RAC2(2/(PI*X)) FF
SOMF=0.0 FF
'SI' N=1 'ALORS' 'ALLERA' FNIMP FF
'POUR' IF=0 'PAS' 1 'JUSQUA' (N-3)/2 'FAIRE'
'DEBUT' AUXF=AUX*X/(2*I+1) FF
SOMF=SOM+AUX

FNIMPF 'FIN' FF
KIF=2*NORMALE(RAC2(X),0.0,1.0)-1.0-EXP(-X/2)*SOM FF

FINKIF 'FIN' FF
'REEL' 'PROCEDURE' KIDECENT(N,P,X) FF
'VALEUR' N,P,X FF
'ENTIER' N FF 'REEL' P,X FF
'DEBUT' 'REEL' L,Z,I,A,C,T,R FF
'ENTIER' K,NOR FF
LF=P/2 FF ZF=X/2 FF
NORF=0 FF

AUGF NORF=NOR+1 FF
'SI' (N/2-1)*LN(Z/NOR) 'SUPER' 88.0 'ALORS' 'ALLERA' AUG FF
KF=0 FF
CF=EXP(-L) FF
RF=1-C FF
AF=EXP(-Z)*(Z/NOR)**(N/2-1)*NOR**(N/2-1)/GAMMA(N/2) FF
IF=KI(N,X) FF
TF=I*C FF

BOUCLEF KF=K+1 FF
AF=A*Z/(N/2+K-1) FF
IF=I-A FF
CF=C*L/K FF
TF=T+I*C FF
RF=R-C FF

'SI' R 'SUPER' F*-5 'ALORS' 'ALLERA' BOUCLE FF
FINKIDECFKIDECENTF=T FF

'REEL' A1,A2,V1,V2,P1,P2,TOT,X,Y FF
'BOOLEEN' SORT FF
'ENTIER' I FF
'TABLEAU' RESUL.(1F7). FF
YF=XF=9.4701271 FF

```
'REEL' 'PROCEDURE' FP(A) FF 'REEL' A FF
'DEBUT' 'REEL' TOTAL,B1,B2 FF
  'BOOLEEN' BOOL FF
  'TABLEAU' RES.(1F7) FF
  'REEL' 'PROCEDURE' INTEGR(V) FF 'REEL' V FF
  INTEGRF=EXP(-V/2)*V**(N3/2-1)*KIDECENT(N1,A,V*X) FF
  TOTALF=0.0 FF
  B1F=F*-8 FF B2F=5.0 FF
REV F TOTALF=TOTAL+INSIRO(INTEGR,B1,B2,6,F*-4,BOOL,RES) FF
  'SI' 1-KI(N3,B2) 'INFER' F*-3 'ALORS' 'ALLERA' CONT FF
  B1F=B2 FF
  B2F=B2+5.0 FF
  'ALLERA' REV FF
CONT F FPF=TOTAL/(GAMMA(N3/2)*2**(N3/2)) FF
'FIN' FF
  A1F=0.0 FF
BIS F A1F=A1+2.0 FF
  A2F=FP(A1) FF
  ECRIRE(A1,A2) FF
  'SI' A2 'SUPER' ALPHA 'ALORS' 'ALLERA' BIS FF
  A2F=A1F=A1-1 FF
  'POUR' P1F=0.0,A1 'FAIRE'
  'POUR' P2F=0.0,A2 'FAIRE'
  'POUR' IF=1,2,3 'FAIRE'
'DEBUT' 'REEL' 'PROCEDURE' F(V) FF
  'VALEUR' V FF 'REEL' V FF
  'DEBUT' 'REEL' I1,I2 FF
    I1F=KIDECENT(N1,P1,V*X) FF
    I2F=KIDECENT(N2,P2,V*Y) FF
    'SI' I=3 'OU' I=4 'ALORS' I1F=1-I1 FF
    'SI' I=2 'OU' I=4 'ALORS' I2F=1-I2 FF
    FF=EXP(-V/2)*V**(N3/2-1)*I1*I2
  'FIN' FF
  TOTF=0.0 FF V1F=F*-8 FF V2F=5.0 FF
  RETOURF TOTF=TOT+INSIRO(F,V1,V2,6,F*-4, SORT,RESUL) FF
  'SI' 1-KI(N3,V2) 'INFER' F*-3 'ALORS' 'ALLERA' SUITE FF
  V1F=V2 FF
  V2F=V2+5.0 FF
  'ALLERA' RETOUR FF
  SUITE F ECRIRE(P1,P2,I,TOT/(GAMMA(N3/2)*2**(N3/2))) FF
'FIN' FF
'FIN' FF
```

6 - TEST DU CARACTERE ALEATOIRE D'UNE SUITE

6.1 - Introduction

Un problème très classique de la statistique est la reconnaissance de l'aspect "aléatoire" d'une suite d'éléments d'un même ensemble E.

Il se pose dans un certain nombre de domaines, par exemple en analyse numérique (méthode de Monte-Carlo).

Il est impossible de définir un test optimum aussi il en existe de nombreux dans la littérature et la pratique usuelle. Nous nous proposons d'étudier un nouveau test de fréquence [3].

Soit donc une suite finie $S = (x_1, \dots, x_N)$ d'éléments d'un même ensemble E tel que :

$$\text{card}(E) = 2^k$$

où k est un nombre entier donné.

On cherche à vérifier l'hypothèse représentée par la loi P_0 sur E^N selon laquelle les x_i sont indépendants et de même loi uniforme sur E.

Lorsque k est égal à 1 il s'agit de reconnaître l'aspect aléatoire d'une suite de digits binaires.

Remarque

Tout nombre écrit en mémoire d'un calculateur est exprimé en binaire et dans le cas de l'IBM 7044 occupe 36 bits. Par conséquent, il est facile, en partant d'une suite de digits binaires, de passer à une suite de nombres au hasard. Il suffit de considérer la suite comme formée de blocs successifs de 36 positions, chacun d'eux constituant un nombre au hasard.

6.2 - Test

Définition 1

"Une figure ϕ , d'ordre r , est une partie de E^r ".

Il s'agit en fait d'une condition portant sur r éléments consécutifs du même ensemble.

Définition 2

"La fréquence de la figure ϕ est le nombre de fois où elle apparaît dans S ".

On montre [3] qu'il existe des figures dites fondamentales telles que :

- 1) La fréquence de toute figure s'exprime à l'aide des fréquences des figures fondamentales ;
- 2) Sous l'hypothèse d'uniformité toutes les figures fondamentales sont de probabilité $\frac{1}{2}$ et leurs fréquences sont indépendantes.

Si on considère maintenant p figures fondamentales ϕ_1, \dots, ϕ_p de même ordre r on peut, en utilisant leurs fréquences, définir un vecteur aléatoire \vec{F} :

$$\vec{F} = (F(\phi_1), \dots, F(\phi_p))$$

de moyenne :

$$\vec{m} = (m_1, \dots, m_p)$$

ϕ_1, \dots, ϕ_p étant des figures fondamentales nous avons, d'après la propriété 2 :

$$m_1 = \dots = m_p = \frac{N-r+1}{2}$$

L'idée de base du test est d'exploiter ces fréquences globalement et non plus séparément, ainsi que cela a été fait jusqu'à maintenant.

On considère donc la somme des carrés des différences :

$$F(\phi_i) - m_i, \quad i = 1, \dots, p$$

qui suit une loi de probabilité du χ^2 réduite à p degrés de liberté.

Pour des raisons de commodité, nous sommes obligés de compter d'abord les fréquences de toutes les figures d'ordre r, de passer ensuite aux fréquences des figures fondamentale (propriété 1) et enfin à la somme des carrés.

6.3 - Application

En 1965 une étude statistique des 100.000 premières décimales de π a été publiée [7]. Elle montrait que leur utilisation en tant que nombres au hasard était parfaitement justifiée en s'appuyant sur les résultats obtenus à l'aide des tests des fréquences, du Poker, de série et de suites. Par contre on a montré [1] qu'un certain test n'était pas satisfait par le nombre π .

Nous avons effectué une nouvelle expérience en écrivant π en représentation binaire. Nous avons ainsi pu étudier au moyen du test que nous avons décrit en 6.2 les 50.000 premiers bits.

En prenant un ordre égal à 5 on a observé les fréquences des figures fondamentales suivantes :

'DEB
'PRO
'VAL
'ENT

N° de figure	fréquence	N° de figure	fréquence
16	25 799	24	25 227
17	25 233	25	25 131
18	24 961	26	25 389
19	25 031	27	25 069
20	25 091	28	25 211
21	25 225	29	25 223
22	25 105	30	25 085
23	25 015	31	24 989

TRUC
TMT
FINI
'FCOI
'ENT
'VALI

Avec $r = 5$ nous avons :

$$m_i = 24\,998, \quad i = 16, \dots, 31$$

'FCOI
'PRO
'ENT
'ENT
'DEB
La somme des carrés des différences conduit au résultat : 98. Cette valeur est bien supérieure à la plus grande limite que l'on aurait pu tolérer. Ceci tend à prouver que, sous une forme binaire, le nombre π ne possède pas l'aspect aléatoire.

6.4 - Remarques sur le programme ALGOL

Les 50.000 premiers bits du nombre π sont stockés sur bande magnétique par enregistrements.

- a) Un appel de la procédure REMPLISSAGE provoque la lecture d'un enregistrement sur la bande et son écriture en mémoire dans la zone ($T[I+1], T[I+255]$).
- b) BIT(A,I) est un indicateur de fonction de type entier qui vaut :
 - 0 si en position I de la mémoire A se trouve 0 ;
 - 1 si en cette même position se trouve 1.

c) Le test est effectué par la procédure INDALÉAL qui possède quatre paramètres formels.

K : défini en 6.1 il est tel que $\text{card}(E) = 2^K$.

R : ordre des figures.

NBRE : longueur de la suite S.

T : tableau des fréquences des 2^{KR} figures d'ordre R.

En choisissant un ordre égal à 5, les 2^5 figures que l'on peut rencontrer sont les suivantes :

figure 0 : 0 0 0 0 0

figure 1 : 0 0 0 0 1

figure 2 : 0 0 0 1 0

figure 29 : 1 1 1 0 1

figure 30 : 1 1 1 1 0

figure 31 : 1 1 1 1 1

Nous commençons par remplir le tableau B au moyen de REMPLISSAGE. Il faut ensuite pour chacune des figures précédentes compter le nombre de fois où elle apparaît. La méthode que nous utilisons consiste pour :

$$i = 1, \dots, 49996$$

à calculer le coefficient k tel que :

$$k = (((B[i].2+B[i+1]).2+B[i+2]).2+B[i+3]).2+B[i+4]$$

où $B[i]$ représente la valeur prise par le $i^{\text{ième}}$ bit, et à augmenter le contenu de $FQ[k]$ de 1.

Le test est effectué par la procédure INDALÉAL appelée avec les paramètres effectifs 1,5,50000 et FQ.

DEBUT
PROCEDURE: REMPLISSAGE(T,1) FF
VALEUR: T FF
ENTIER: TABLEAU: T FF ENTIER: T FF

CODE

TSX S*100P,4
IORBS TRUC
PZE S*5J05
LXA R1+1,4
SXA TMT,4
CLA TMT
ADD R1+2
TMT 255
TRA FINI
TRUC PZE *+2,255
BSS 257
PZE ***,TRUC+1
TMT *
FINI EQU

FCODE: FF
ENTIER: PROCEDURE: BIT(A,1) FF
VALEUR: A,1 FF ENTIER: A,1 FF

CODE

CAL F2+2
STA *+2
CAL F2+1
LRS **
ZAC
LLS 1
SLW F2

FCODE: FF
PROCEDURE: INDALCAT(K,R,NBRE,F) FF
ENTIER: K,R,NBRE FF
ENTIER: TABLEAU: F FF
DEBUT: ENTIER: I,IPR,J,JPR,N,P,T,TEST,TEST1,MJ,M FF
ENTIER: TABLEAU: X,Y.(OFR*-1).,SORT.(OFR-1). FF
REEL: T2 FF
PF=R*K FF
TESTIF=0 FF
T2F=0.0 FF
POUR: JF=2**(P-K) PAS: 1 JUSQUA: 2**P-1 FAIRE
DEBUT: TF=0 FF

JPR=J FF

POUR: NF=1 PAS: 1 JUSQUA: P FAIRE
SI: JPP: INFER: 2**(P-N) ALORS: Y.(N-1).F=0 SINDN
DEBUT: Y.(N-1).F=1 FF
JPRF=JPR-2**(P-N)

FIN: FF
POUR: IF=0 PAS: 1 JUSQUA: 2**P-1 FAIRE
DEBUT: IPRF=1 FF
POUR: NF=1 PAS: 1 JUSQUA: P FAIRE

```

'FIN' FF
Ecrire(I,T) FF
T2F=T2+(T-NBRE/2)**2*4/NBRE FF
'SI' (T 'INFEG' NBRE/2-RAC2(NBRE)) 'OU' (T 'SUPEG' NBRE/2+
RAC2(NBRE)) 'ALORS'
'DEBUT' TEST1F=TEST1+1 FF
  'POUR' NF=0 'PAS' 1 'JUSQUA' K-1 'FAIRE'
  'DEBUT' 'POUR' MF=0 'PAS' 1 'JUSQUA' R-1 'FAIRE'
    SORT.(M).F=Y.(M*K+N). FF
    .SORTIE(6,MOD,R,SORT.(0).) FF
    MODFMODELE(''(2012)'')
  'FIN' FF
  Ecrire(J,T) FF
'FIN' FF
'FIN' FF
'SI' TEST1=0 'ALORS'
  Ecrire('('AUCUNE FREQUENC')',('E INCORRECTE')) FF
  Ecrire('('T2')',T2)
'FIN' FF
'ENTIER' I,J,K,L FF
'ENTIER' 'TABLEAU' Y.(IF1535).,B.(1F39).,FQ.(0F31). FF
'POUR' IF=0 'PAS' 1 'JUSQUA' 5 'FAIRE'
REPLISSAGE(Y,255*I) FF
'POUR' KF=0 'PAS' 1 'JUSQUA' 31 'FAIRE' FQ.(K).F=0 FF
IF=2 FF
RETOURF
'POUR' JF=1 'PAS' 1 'JUSQUA' 35 'FAIRE'
B.(J).F=BIT(Y.(I).,J+1) FF
'POUR' JF=1,2,3,4 'FAIRE'
B.(35+J).F=BIT(Y.(I+1).,J+1) FF
'POUR' JF=1 'PAS' 1 'JUSQUA' 35 'FAIRE'
'DEBUT' KF=B.(J). FF
  'POUR' LF=1,2,3,4 'FAIRE'
  KF=K*2+B.(J+L). FF
  FQ.(K).F=FQ.(K).+1
'FIN' FF
IF=I+1 FF
'SI' I 'INFER' 1430 'ALORS' 'ALLERA' RETOUR FF
'POUR' JF=1 'PAS' 1 'JUSQUA' 20 'FAIRE'
B.(J).F=BIT(Y.(I).,J+1) FF
'POUR' JF=1 'PAS' 1 'JUSQUA' 16 'FAIRE'
'DEBUT' KF=B.(J). FF
  'POUR' LF=1,2,3,4 'FAIRE'
  KF=K*2+B.(J+L). FF
  FQ.(K).F=FQ.(K).+1
'FIN' FF
INDALEA1(1,5,50000,FQ) FF
'FIN' FF

```

CHAPITRE - IV

PROBLEMES D'ESTIMATION SUR ECHANTILLONS GAUSSIENS

Considérons un phénomène physique ou économique et une de ses caractéristiques X .

On sait que X est une variable aléatoire dont la loi de probabilité a une forme connue, sans qu'on puisse se prononcer sur la valeur des paramètres qu'elle contient. Si, par exemple, on suppose X de forme absolument continue, sa densité de probabilité sera $f(x, \theta)$, la fonction f étant spécifiée, mais non la valeur du paramètre θ .

On est alors conduit à la réalisation d'un échantillon qui permettra de préciser cette valeur.

Les méthodes classiques proposent deux façons d'estimer les paramètres d'une loi de probabilité : à la réalisation de l'échantillon, l'estimation ponctuelle associe une seule valeur du paramètre alors que l'estimation ensembliste associe un "domaine de confiance" pour le paramètre.

C'est le second procédé que nous étudierons, la loi de probabilité étant supposée normale, dans les trois cas suivants :

- 1 - la moyenne m est inconnue, la variance σ^2 est à estimer.
- 2 - la moyenne m est connue, la variance σ^2 est à estimer.
- 3 - la moyenne m et la variance σ^2 sont toutes deux à estimer.

1 - ESTIMATION DE LA VARIANCE, LA MOYENNE ETANT INCONNUE

1.1 - Introduction

On détermine ici le domaine d'acceptation de longueur minimum pour le paramètre σ^2 d'un échantillon gaussien, de moyenne inconnue, formé de n observations indépendantes de même loi [2].

Ce domaine est entièrement défini par ses deux valeurs extrêmes qui seront donc tabulées pour les trois valeurs les plus usuelles du seuil de l'estimation.

1.2 - Domaine de longueur minimum

a) Théorème

"Soit $q_n(t)$ la fonction du réel t définie pour le paramètre n entier par :

$$q_n(t) = \frac{1}{2^{n/2} \Gamma(\frac{n}{2})} e^{-t/2} t^{n/2 - 1}, \quad t > 0$$

alors si $a(\alpha, n-1)$ et $b(\alpha, n-1)$ sont les solutions en a et b du système d'équations :

$$\begin{cases} a < b ; \\ \int_a^b q_{n-1}(t) dt = 1 - \alpha ; \\ a^2 q_{n-1}(a) = b^2 q_{n-1}(b) \end{cases}$$

le domaine défini par :

$$\frac{ng}{b(\alpha, n-1)} \leq \sigma^2 \leq \frac{ng}{a(\alpha, n-1)}$$

est de longueur minimum par rapport aux domaines de seuil α , g étant une statistique définie par :

$$g(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left[\frac{1}{n} \sum_{i=1}^n x_i \right]^2 ."$$

b) Démonstration

Soient X_1, X_2, \dots, X_n n variables aléatoires indépendantes ayant pour loi commune la loi normale de moyenne m inconnue et de variance σ^2 , et soit G la variable aléatoire associée à la statistique g qui est définie par :

$$g(x_1, x_2, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i^2 - \left(\frac{1}{n} \sum_{i=1}^n x_i \right)^2$$

Alors, d'après le théorème de Fisher, nous savons que la loi de probabilité de $\frac{nG}{\sigma^2}$ est une loi du χ^2 à n-1 degrés de liberté.

Nous cherchons le domaine d'acceptation de longueur minimum pour le paramètre σ^2 . Nous pouvons le faire par l'intermédiaire de la statistique g, soit d(g). Définissons alors E(g) de la façon suivante :

$$\sigma^2 \in d(g) \Leftrightarrow \frac{ng}{\sigma^2} \in E(g)$$

On doit avoir :

$$(1) \int_{E(g)} q_{n-1}(t) dt = 1 - \alpha, \quad \forall g \in \mathbb{R}.$$

De plus la longueur de d(g) est égale à :

$$ng \int_{E(g)} \frac{q_{n-1}(t)}{t^2} dt$$

et donc on minimise, uniformément en g, la longueur de d(g) si l'on choisit E(g) tel que :

$$\int_{E(g)} \frac{q_{n-1}(t)}{t^2} dt$$

soit minimum sous la condition (1) ce qui est obtenu, grâce au lemme de Neyman et Pearson, en prenant :

$$E(g) = [a, b]$$

où a et b sont tels que :

$$a^2 q_{n-1}(a) = b^2 q_{n-1}(b).$$

1.3 - Comparaison avec la méthode usuelle

En pratique on se contente du domaine non optimal suivant :

$$\frac{ng}{b'} \leq \sigma^2 \leq \frac{ng}{a'}$$

où a' et b' sont choisis tels que :

$$a' = Q_{n-1}^{-1} \left(\frac{\alpha}{2} \right) ;$$

$$b' = Q_{n-1}^{-1} \left(1 - \frac{\alpha}{2} \right)$$

$Q_n(z)$ étant la fonction de répartition de la loi du χ^2 réduite à n degrés de liberté.

Ces deux coefficients sont utilisés, en pratique, en raison de leur simplicité de calcul de préférence à $a(\alpha, n-1)$ et $b(\alpha, n-1)$, définis en 1.2-a. Il faut alors remarquer que, lorsque n est grand, cette approximation est correcte mais les écarts sont sensibles pour les petites valeurs.

Ainsi pour $n = 10$ et $\alpha = 0.01$ nous avons obtenu pour

a) l'intervalle de confiance de longueur minimum :

$$(0.310)g \leq \sigma^2 \leq (4.813)g$$

b) l'intervalle de confiance approximatif :

$$(0.424)g \leq \sigma^2 \leq (5.780)g$$

1.4 - Table

Nous avons à résoudre le système suivant :

$$\left\{ \begin{array}{l} a < b ; \\ \int_a^b q_{n-1}(t) dt = 1-\alpha ; \\ a^2 q_{n-1}(a) = b^2 q_{n-1}(b). \end{array} \right.$$

Soit alors la nouvelle variable θ telle que :

$$\theta = \frac{a}{b}$$

On a :

$$0 < \theta < 1$$

On peut exprimer a et b en fonction de θ . On obtient :

$$a = \theta \frac{n+1}{\theta-1} \text{Log } \theta ;$$

$$b = \frac{n+1}{\theta-1} \text{Log } \theta$$

Soit $F(\theta)$ la fonction du réel θ définie par :

$$F(\theta) = Q_{n-1}\left(\frac{n+1}{\theta-1} \text{Log } \theta\right) - Q_{n-1}\left(\theta \frac{n+1}{\theta-1} \text{Log } \theta\right)$$

où $Q_n(z)$ est la fonction de répartition d'une loi du χ^2 réduite à n degrés de liberté.

On peut facilement montrer que la fonction F est monotone et strictement croissante. Un procédé de bisection classique permet alors de trouver la valeur $\theta(\alpha, n-1)$ qui réalise l'égalité avec $1-\alpha$. On en déduit $a(\alpha, n-1)$ et $b(\alpha, n-1)$ puisque nous avons vu que a et b s'expriment en fonction de θ .

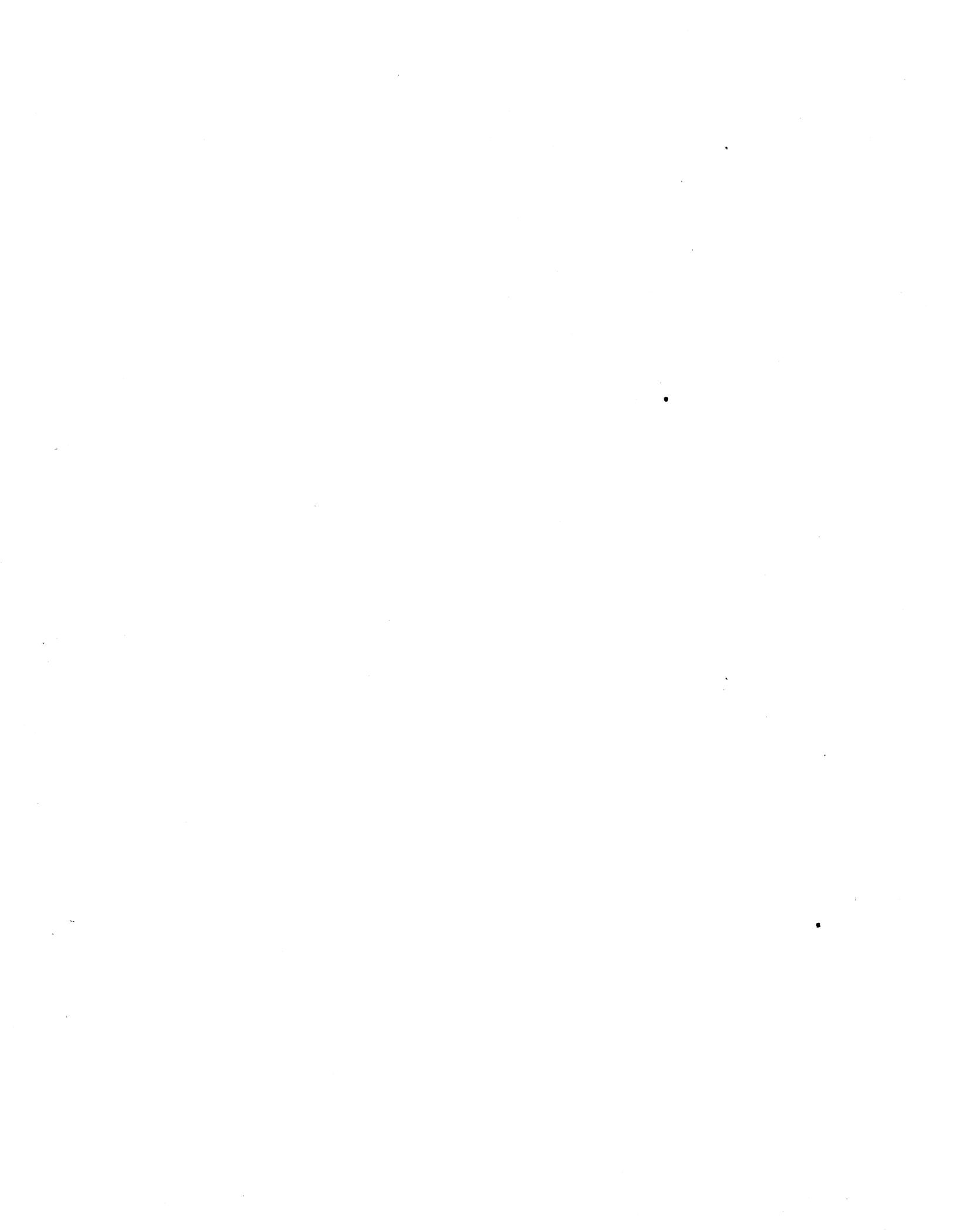
Nous avons tabulé les deux coefficients $a(\alpha, n-1)$ et $b(\alpha, n-1)$ pour $n = 3(1)30$ et $\alpha = 0.001, 0.005, 0.01, 0.05$. En fait le programme que nous avons écrit permet le calcul pour tout couple (α, n) .

```

'PROCEDURE' ESTIMVAR(N,ALPHA,A,B) FF
'VALEUR' N,ALPHA FF
'ENTIER' N FF 'REEL' ALPHA,A,B FF
'DEBUT'
'REEL' 'PROCEDURE' KI(N,X) FF
'VALEUR' N,X FF
'ENTIER' N FF 'REEL' X FF
'DEBUT'
'REEL' 'PROCEDURE' .NORMALE(X,MOY,ECART) FF
'VALEUR' X,MOY,ECART FF 'REEL' X,MOY,ECART FF
'DEBUT' 'ENTIER' I FF
'REEL' SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
XF=(X-MOY)/ECART FF
SGNEF=SIGNE(X) FF
XF=ABS(X) FF
X2F=X*X FF
YF=0.3989422804*EXP(-0.5*X2) FF
'SI' X 'INFER' 3.5 'ALORS'
'DEBUT' SF=XF=Y*X FF
'POUR' IF=3,I+2 'TANTQUE' S 'NONEG' T 'FAIRE'
'DEBUT' TF=S FF
XF=X*X2/I FF
SF=S+X
'FIN' FF
NORMALEF=0.5+SGNE*S
'FIN'
'SINON'
'DEBUT' Q1F=X FF
P2F=Y*X FF
P1F=Y FF
Q2F=X2+1.0 FF
MF=Y/X FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS'
'DEBUT' MF=1-M FF
TF=1-T
'FIN' FF
'POUR' IF=2,I+1 'TANTQUE'
((M 'NONEG' T) 'ET' (S 'NONEG' T)) 'FAIRE'
'DEBUT' SF=X*P2+I*P1 FF
P1F=P2 FF
P2F=S FF
SF=X*Q2+I*Q1 FF
Q1F=Q2 FF
Q2F=S FF
SF=M FF
MF=T FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS' TF=1-T FF
'FIN' FF
NORMALEF=T
'FIN' FF
FINORMF 'FIN' FF
'ENTIER' I FF
'REEL' AUX,SOM,PI FF
'SI' N-2*ENTIER(N/2)=1 'ALORS'
'ALLERA' NIMP FF
SOMF=AUXF=1.0 FF
'SI' N=2 'ALORS' 'ALLERA' FPAIR FF

```

```
'POUR' IF=1 'PAS' 1 'JUSQUA' N/2-1 'FAIRE'  
'DEBUT' AUXF=AUX*X/(2*I) FF  
      SOMF=SOM+AUX  
'FIN' FF  
FPAIRF KIF=1.0-EXP(-X/2)*SOM FF  
'ALLERA' FINKI FF  
NIMPF  PIF=3.141592653590 FF  
      AUXF=RAC2(2/(PI*X)) FF  
      SOMF=0.0 FF  
'SI' N=1 'ALORS' 'ALLERA' FNIMP FF  
'POUR' IF=0 'PAS' 1 'JUSQUA' (N-3)/2 'FAIRE'  
'DEBUT' AUXF=AUX*X/(2*I+1) FF  
      SOMF=SOM+AUX  
'FIN' FF  
FNIMPF KIF=2*NORMALE(RAC2(X),0.0,1.0)-1.0-EXP(-X/2)*SOM FF  
FINKIF 'FIN' FF  
'REEL' TETA,TETA1,TETA2 FF  
'REEL' 'PROCEDURE' F(T) FF  
'VALEUR' T FF 'REEL' T FF  
'DEBUT' 'REEL' AUX FF  
      AUXF=(N+1)*LN(T)/(T-1) FF  
      FF=KI(N-1,AUX)-KI(N-1,T*AUX)  
'FIN' FF  
TETAFF=0.5 FF  
TETA1F=0.0 FF  
TETA2F=1.0 FF  
BOUCLEF 'SI' F(TETA) 'SUPER' 1-ALPHA 'ALORS'  
      TETA2F=TETA 'SINON'  
      TETA1F=TETA FF  
      TETAFF=(TETA1+TETA2)/2 FF  
'SI' TETA2-TETA1 'SUPER' F*-8 'ALORS' 'ALLERA' BOUCLE FF  
      BF=(N+1)*LN(TETA)/(TETA-1) FF  
      AF=TETA*B  
'FIN' FF
```



N	ALPHA = 0.001		ALPHA = 0.005		ALPHA = 0.01		ALPHA = 0.05	
	A	B	A	B	A	B	A	B
3	0.0243	36.6137	0.0717	30.3037	0.1148	27.5107	0.3512	20.7438
4	0.0908	35.9846	0.2069	30.0844	0.2969	27.4603	0.7082	21.0631
5	0.2101	36.2664	0.4113	30.5697	0.5534	28.0269	1.1392	21.8001
6	0.3808	36.9890	0.6747	31.3964	0.8700	28.8927	1.6233	22.7410
7	0.5979	37.9539	0.9871	32.4103	1.2350	29.9228	2.1473	23.7944
8	0.8560	39.0627	1.3406	33.5356	1.6397	31.0506	2.7027	24.9147
9	1.1500	40.2623	1.7288	34.7307	2.0775	32.2396	3.2836	26.0769
10	1.4757	41.5207	2.1469	35.9712	2.5435	33.4684	3.8855	27.2662
11	1.8293	42.8189	2.5908	37.2419	3.0335	34.7235	4.5055	28.4732
12	2.2079	44.1439	3.0573	38.5329	3.5447	35.9962	5.1409	29.6920
13	2.6087	45.4873	3.5439	39.8378	4.0744	37.2808	5.7899	30.9184
14	3.0297	46.8432	4.0483	41.1517	4.6206	38.5733	6.4509	32.1497
15	3.4689	48.2076	4.5689	42.4716	5.1816	39.8706	7.1228	33.3839
16	3.9247	49.5771	5.1039	43.7910	5.7559	41.1709	7.8043	34.6197
17	4.3956	50.9501	5.6523	45.1204	6.3425	42.4727	8.4947	35.8559
18	4.8806	52.3245	6.2128	46.4465	6.9402	43.7748	9.1932	37.0919
19	5.3785	53.6992	6.7845	47.7724	7.5481	45.0765	9.8990	38.3271
20	5.8883	55.0737	7.3666	49.0973	8.1655	46.3771	10.6119	39.5610
21	6.4094	56.4468	7.9582	50.4208	8.7916	47.6763	11.3310	40.7935
22	6.9408	57.8182	8.5588	51.7425	9.4259	48.9736	12.0561	42.0243
23	7.4819	59.1875	9.1677	53.0620	10.0678	50.2689	12.7868	43.2533
24	8.0322	60.5544	9.7845	54.3793	10.7169	51.5619	13.5227	44.4802
25	8.5910	61.9189	10.4086	55.6942	11.3727	52.8525	14.2635	45.7052

N	ALPHA = 0.001		ALPHA = 0.005		ALPHA = 0.01		ALPHA = 0.05	
	A	B	A	B	A	B	A	B
26	9.1530	63.2805	11.0397	57.0064	12.0348	54.1407	15.0089	46.9281
27	9.7327	64.6394	11.6773	58.3160	12.7029	55.4264	15.7588	48.1488
28	10.3146	65.9951	12.3211	59.6229	13.3767	56.7096	16.5128	49.3675
29	10.9034	67.3481	12.9708	60.9272	14.0558	57.9902	17.2707	50.5840
30	11.4989	68.6979	13.6261	62.2286	14.7401	59.2682	18.0324	51.7985

2 - ESTIMATION DE LA VARIANCE, LA MOYENNE ETANT CONNUE

2.1 - Introduction

Le problème d'estimation que nous nous proposons de résoudre est identique à celui qui a été étudié au paragraphe précédent. Le seul changement vient du fait que la moyenne m est maintenant connue.

Nous avons donc à déterminer le domaine d'acceptation de longueur minimum pour le paramètre σ^2 d'un échantillon gaussien, de moyenne m connue, formé de n observations indépendantes.

2.2 - Domaine de longueur minimum

On peut utiliser comme statistique la variance empirique des observations, c'est-à-dire la fonction g telle que :

$$g(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n (x_i - m)^2$$

Si G est la variable aléatoire associée à g , la loi de probabilité de $\frac{nG}{\sigma^2}$ est une loi du χ^2 réduite à n degrés de liberté.

On adopte alors comme domaine de confiance pour σ^2 :

$$\frac{ng}{b(\alpha, n)} \leq \sigma^2 \leq \frac{ng}{a(\alpha, n)}$$

où les deux coefficients $a(\alpha, n)$ et $b(\alpha, n)$ sont les solutions du système d'équations :

$$\left\{ \begin{array}{l} a < b ; \\ \int_a^b q_n(t) dt = 1 - \alpha ; \\ a^2 q_n(a) = b^2 q_n(b) \end{array} \right.$$

Le raisonnement est analogue à celui du paragraphe 1. Il suffit de changer n et $n+1$.

2.3 - Table

La table précédente est encore valable ici.

Si l'échantillon est de taille n les deux valeurs a et b cherchées se trouvent dans la table à l'intersection de la colonne d'en-tête α et de la ligne d'en-tête $n+1$.

3 - ESTIMATION SIMULTANEE DE LA MOYENNE ET DE LA VARIANCE

3.1 - Introduction

On détermine le domaine d'acceptation de surface minimum pour les deux paramètres m et σ d'un échantillon gaussien formé de n observations indépendantes [12].

3.2 - Domaines de surface minimum

a) Théorème

"Soit $f_n(z,t)$ la fonction des deux réels z et t définie pour le paramètre n entier par :

$$f_n(z,t) = \frac{z^{\frac{n}{2}}}{\sqrt{\pi} \cdot 2^{\frac{n}{2}-1} \Gamma(\frac{n-1}{2})} \frac{1}{z^{n+1}} e^{-\frac{n}{2} \frac{1+t^2}{z^2}} ; z > 0$$

et soit $F_n(\lambda)$ la fonction du réel λ :

$$F_n(\lambda) = \iint_{\substack{z \geq 0 \\ f_n(z,t) \geq \lambda}} f_n(z,t) dz dt$$

Alors si $\lambda_{n,\alpha}$ est la solution en λ de l'équation :

$$F_n(\lambda) = 1-\alpha$$

le domaine $D_{n,\alpha}$ de R^2 défini par :

$$(z,t) \in D_{n,\alpha} \iff f_n(z,t) \geq \lambda_{n,\alpha}$$

est le domaine de surface minimum par rapport aux domaines de seuil α ".

b) Démonstration

Soient X_1, X_2, \dots, X_n n variables aléatoires indépendantes, ayant pour loi commune la loi normale de moyenne m et de variance σ^2 , et soit

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$S = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

D'après le théorème de Fisher, nous savons que les deux variables aléatoires :

$$U = \sqrt{n} \frac{\bar{X} - m}{\sigma}$$

$$V = \frac{nS}{\sigma^2}$$

sont indépendantes et suivent respectivement la loi normale centrée réduite et celle du χ^2 à $n-1$ degrés de liberté. Si nous exprimons m et σ en fonction de U et V il vient :

$$m = \bar{X} - \sqrt{S} \frac{U}{\sqrt{V}}$$

$$\sigma = \sqrt{S} \sqrt{\frac{n}{V}}$$

Soient alors les nouvelles variables aléatoires :

$$Z = \sqrt{\frac{n}{V}}$$

$$T = - \frac{U}{\sqrt{V}}$$

Si C est un domaine de R^2 tel que :

$$\text{Prob} \{(Z, T) \in C\} = 1 - \alpha$$

on a :

S. Surface de C = Surface du domaine d'acceptation en m et σ .

D'après le lemme de Neyman et Pearson, le domaine C^* optimum sera défini par :

$$(z,t) \in C^* \Leftrightarrow f_n(z,t) \geq \lambda_{n,\alpha}$$

où $f_n(z,t)$ est la densité de probabilité du couple (Z,T).

Or un calcul élémentaire nous montre que :

$$f_n(z,t) = \frac{\frac{n}{2}}{\sqrt{\pi} \cdot 2^{\frac{n}{2}-1} \cdot \Gamma(\frac{n-1}{2})} \frac{1}{z^{n+1}} e^{-\frac{n}{2} \cdot \frac{1+t^2}{z^2}} ; z \geq 0$$

et ainsi si on a calculé la correspondance :

$$(\alpha, n) \rightarrow \lambda_{n,\alpha} : \iint_{\substack{z \geq 0 \\ f_n(z,t) \geq \lambda_{n,\alpha}}} f_n(z,t) dz dt = 1-\alpha$$

on pourra tracer les domaines C^* comme fonction de n et α sous $D_{n,\alpha}$.

3.3 - Comparaison avec la méthode usuelle

En pratique on se contente du domaine non optimal suivant :

la région critique du test associé à notre problème d'estimation est le complémentaire du rectangle :

$$N^{-1} \left(\frac{\xi}{2} \right) \leq u \leq -N^{-1} \left(\frac{\xi}{2} \right)$$

$$Q_{n-1}^{-1} \left(\frac{\delta}{2} \right) < v < Q_{n-1}^{-1} \left(1 - \frac{\delta}{2} \right)$$

où les deux nombres ξ et δ sont tels que :

$$(1-\xi) (1-\delta) = 1-\alpha,$$

et en ce qui concerne le domaine d'acceptation en m et σ cela donne le trapèze :

$$\sqrt{s} \sqrt{\frac{n}{Q_{n-1}^{-1} \left(1 - \frac{\delta}{2}\right)}} \leq \sigma \leq \sqrt{s} \sqrt{\frac{n}{Q_{n-1}^{-1} \left(\frac{\delta}{2}\right)}}$$

$$\bar{x} + \frac{\sigma}{\sqrt{n}} N^{-1} \left(\frac{\xi}{2}\right) \leq m \leq \bar{x} - \frac{\sigma}{\sqrt{n}} N^{-1} \left(\frac{\xi}{2}\right)$$

s et \bar{x} représentant la moyenne et la variance empiriques des observations.

On retrouvera sur le graphique (a) pour $n = 15$ et $\alpha = 1\%$ le domaine limité par la courbe $f_{15}(z,t) \geq \lambda_{15,0.01}$ et à titre comparatif le trapèze obtenu par la méthode usuelle.

3.4 - Comparaison avec les tables de Neyman et Pearson

Nous pouvons également prendre comme test associé celui qui consiste à vérifier si un échantillon de taille n a été tiré d'une population normale de moyenne m et de variance σ^2 . La méthode utilisée par Neyman et Pearson [16] consiste alors à choisir des valeurs représentatives des deux paramètres, soient \bar{x} et s^2 , et à tester l'hypothèse simple (m, σ^2) contre l'hypothèse simple (\bar{x}, s^2) . On est ainsi ramené à la résolution en μ de l'équation :

$$\iint f_n(z,t) dz dt = 1 - \alpha$$

$$z \geq 0 ; \frac{1}{z^n} e^{-\frac{n}{2} \frac{1+t^2}{z^2}} \geq \mu$$

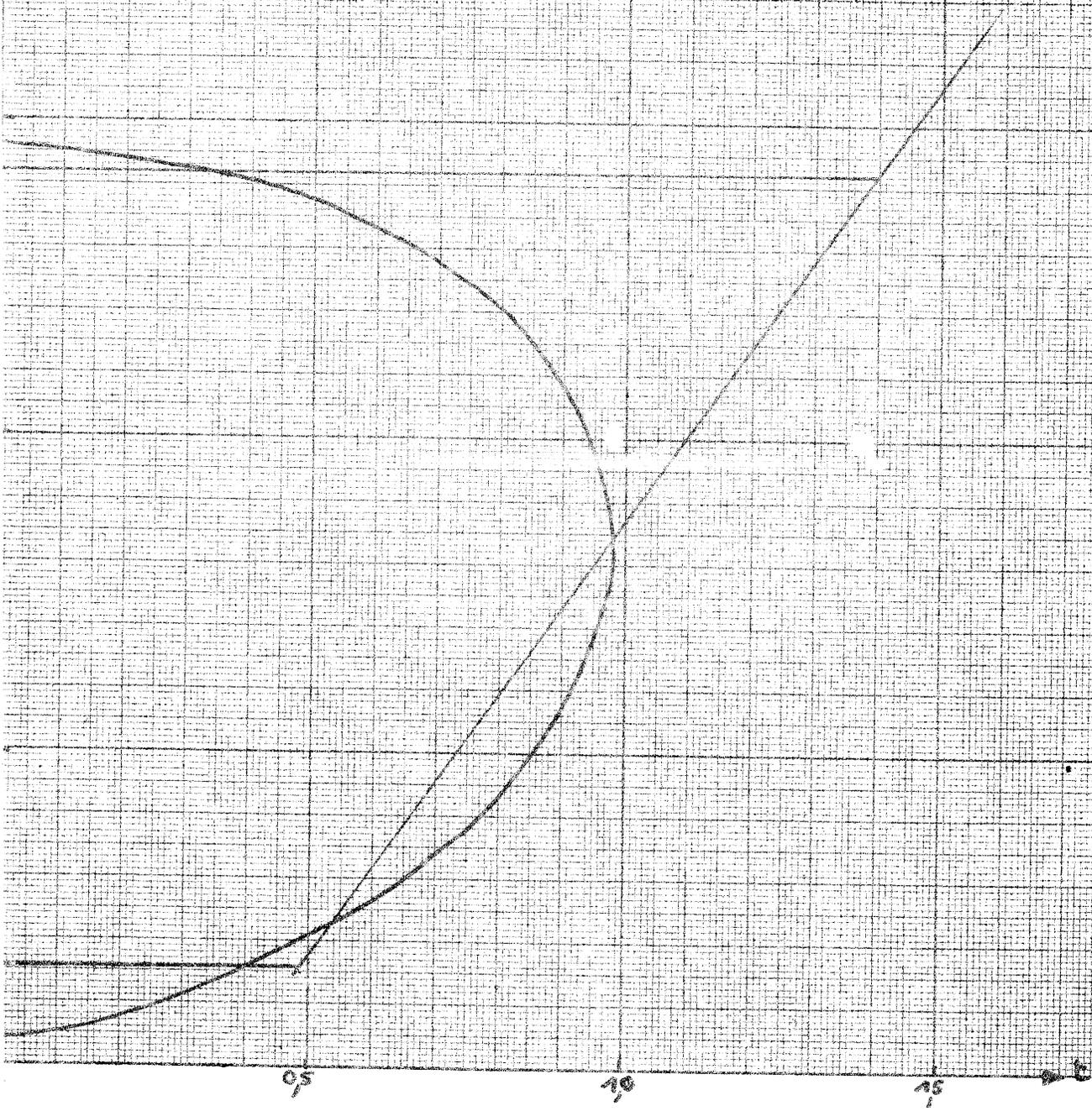
C'est l'intégrale que Neyman et Pearson ont tabulée en fonction du paramètre μ .

Sur le graphique (b) nous avons représenté pour $n = 20$ et $\alpha = 5\%$ le domaine d'acceptation correspondant, ainsi que le domaine de surface minimum.

Graphique (a)

$\alpha = 0.01$

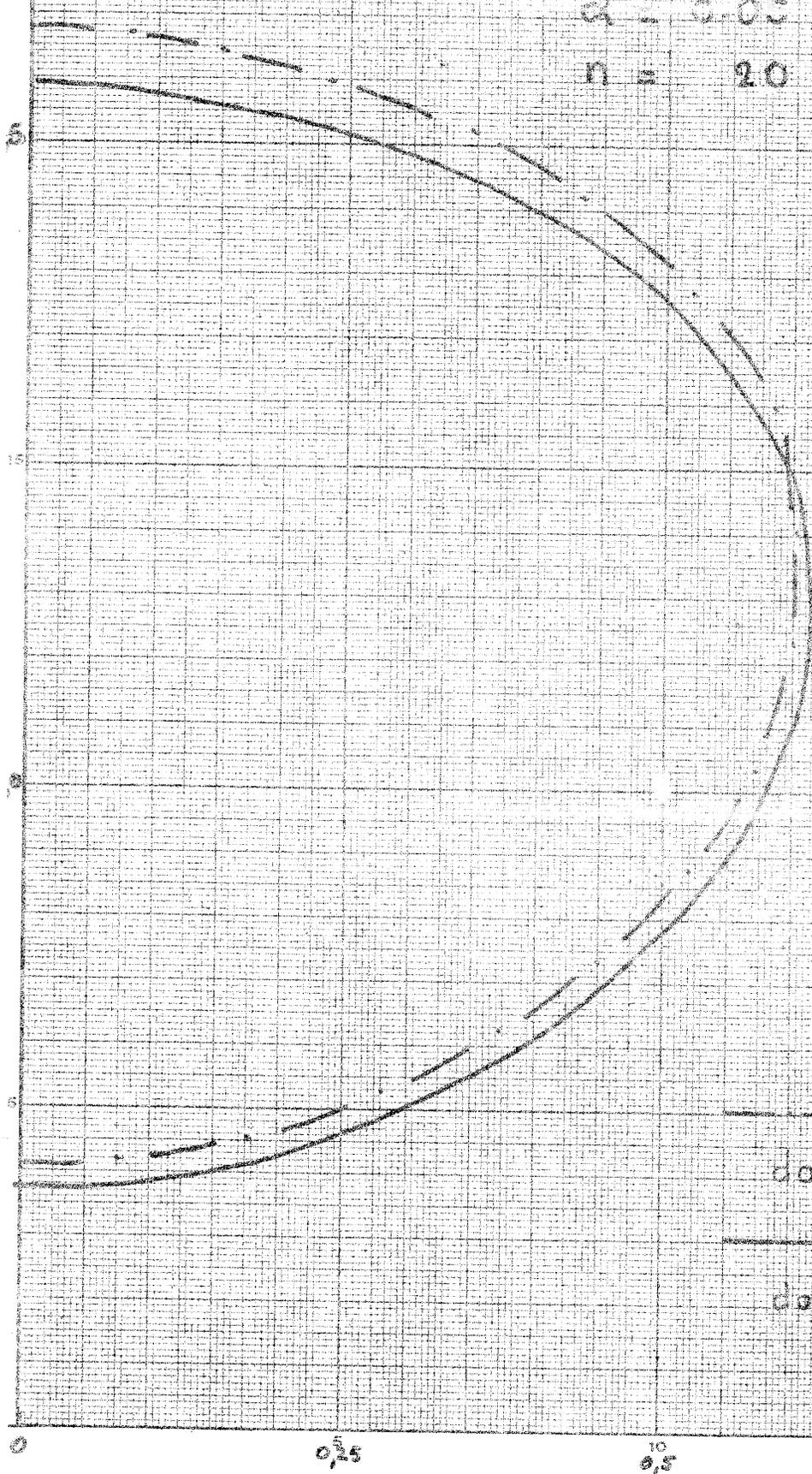
$n = 15$



Graphique (b)

$\alpha = 0.05$

$n = 20$



domaine de Neyman

domaine optimal



3.5 - Utilisation pour l'estimation de σ

Nous pouvons remarquer que les domaines $D_{n,\alpha}$ sont convexes, symétriques par rapport à l'axe des z , et sont de plus limités par la courbe :

$$1+t^2 = 2 \frac{n+1}{n} z^2 [A_{n,\alpha} - \text{Log } z] ; z \geq 0$$

où le coefficient $A_{n,\alpha}$ est simplement déduit de $\lambda_{n,\alpha}$ et n .

z est compris entre les deux valeurs Z_{\min} et Z_{\max} .

On peut penser à utiliser Z_{\min} et Z_{\max} comme intervalle de confiance pour σ^2 .

Il ne sera pas, bien entendu, de longueur minimum mais les résultats seront meilleurs que ceux obtenus en prenant :

$$\frac{n.s}{Q_{n-1}^{-1}(1-\frac{\alpha}{2})} \leq \sigma^2 \leq \frac{n.s}{Q_{n-1}^{-1}(\frac{\alpha}{2})}$$

Ainsi pour $n = 7$ et $\alpha = 0.013$ nous avons obtenu pour :

a) l'intervalle de confiance de longueur minimum

$$(0.27659).s \leq \sigma^2 \leq (5.59552).s$$

b) l'intervalle de confiance établi à l'aide de Z_{\min} et Z_{\max} :

$$(0.27766).s \leq \sigma^2 \leq (5.59559).s$$

c) l'intervalle de confiance approximatif :

$$(0.38212).s \leq \sigma^2 \leq (10.01723).s$$

3.6 - Principe d'une méthode de résolution numérique

En conservant les notations déjà utilisées en 3.2 le problème principal que nous avons à résoudre est le suivant : étant données la fonction $f_n(z,t)$ des deux variables réelles z et t , définie pour z positif par :

$$f_n(z,t) = \frac{\frac{n}{2}}{\sqrt{\pi} \cdot 2^{\frac{n}{2}-1} \Gamma(\frac{n-1}{2})} \frac{1}{z^{n+1}} e^{-\frac{n}{2} \left(\frac{1+t^2}{z^2} \right)} ; z > 0$$

et la fonction $F_n(\lambda)$ du réel λ :

$$F_n(\lambda) = \iint_{z > 0 ; f_n(z,t) > \lambda} f_n(z,t) dz dt$$

pour α et n fixés résoudre l'équation en λ :

$$F_n(\lambda) = 1-\alpha$$

qui admet pour solution $\lambda_{n,\alpha}$.

La fonction $f_n(z,t)$ est paire en t et par conséquent, en se restreignant au demi-espace $t > 0$, notre problème peut se réduire à rechercher la valeur λ réalisant la condition :

$$\iint_{z > 0 ; t > 0 ; f_n(z,t) > \lambda} f_n(z,t) dz dt = \frac{1-\alpha}{2}$$

Afin de simplifier les calculs substituons à $f_n(z,t)$ la fonction $f_n^*(z,t)$ telle que :

$$f_n^*(z,t) = \frac{1}{z^{n+1}} e^{-\frac{n}{2} \left(\frac{1+t^2}{z^2} - 1 \right)} ; z > 0$$

On a alors :

$$f_n^*(z,t) = f_n(z,t) \cdot K_n$$

en posant :

$$K_n = \frac{\sqrt{\pi} \cdot 2^{\frac{n}{2}-1} \cdot \Gamma(\frac{n-1}{2}) \cdot e^{\frac{n}{2}}}{n^{\frac{n}{2}}}$$

Le problème consiste maintenant en la résolution de l'équation en λ :

$$F_n^*(\lambda) = \int_{z>0} \int_{t>0} f_n^*(z,t) dz dt = \frac{1-\alpha}{2} K_n$$

$z>0 ; t>0 ; f_n^*(z,t) > \lambda$

qui admet $\lambda_{n,\alpha}^*$ pour solution.

On voit immédiatement la relation qui existe entre $\lambda_{n,\alpha}$ et $\lambda_{n,\alpha}^*$:

$$\lambda_{n,\alpha}^* = \lambda_{n,\alpha} \cdot K_n$$

Propriété 1

Pour n fixé $F_n^*(\lambda)$ est une fonction décroissante de λ , d'où unicité de la solution.

Propriété 2

Le domaine $D(n,\lambda)$ défini par :

$$D(n,\lambda) = \{(z,t) : z > 0 ; t > 0 ; f_n^*(z,t) > \lambda\}$$

est un domaine qui est non vide si et seulement si :

$$0 < \lambda \leq \text{Max}_{z \geq 0 ; t > 0} f_n^*(z,t)$$

Or il est clair que :

$$\text{Max}_{z > 0 ; t > 0} f_n^*(z,t) = \text{Max}_{z > 0} f_n^*(z,0) = f_n^*(\sqrt{\frac{n}{n+1}}, 0)$$

et la condition devient :

$$0 < \lambda \leq e^{-\frac{1}{2}} \left(1 + \frac{1}{n}\right)^{\frac{n+1}{2}}$$

Propriété 3

$D(n, \lambda)$ contient le point $(t = 0, z = 1)$ et est limité par la courbe $f_n^*(z, t) = \lambda$.

On a alors :

$$1+t^2 = z^2 \left[1 - \frac{2}{n} \text{Log}(\lambda z^{n+1}) \right]$$

ce qui peut encore s'écrire :

$$1+t^2 = -2 \frac{n+1}{n} z^2 [A + \text{Log } z]$$

où le coefficient A est déduit de n et λ :

$$A = \frac{1}{n+1} \left[-\frac{n}{2} + \text{Log } \lambda \right]$$

Propriété 4

On peut encadrer $D(n, \lambda)$ de la façon suivante :

a) la valeur maximum de t est solution de $\frac{\partial f_n^*}{\partial z} = 0$ c'est-à-dire :

$$1+t^2 = \frac{n+1}{n} z^2$$

d'où les coordonnées du point à tangente verticale :

$$z = (\lambda e^{\frac{1}{2} - \frac{1}{n+1}})$$

$$t = \sqrt{\frac{n+1}{n} z^2 - 1}$$

b) z est compris entre les deux valeurs $z_1(n, \lambda)$ et $z_2(n, \lambda)$ solutions de :

$$\frac{1}{z^{n+1}} e^{-\frac{n}{2} \left(\frac{1}{z^2} - 1 \right)} = \lambda$$

Les propriétés 3 et 4 nous permettent, en séparant les variables z et t, d'écrire :

$$F_n^*(\lambda) = e^{\frac{n}{2}} \int_{z_1(n,\lambda)}^{z_2(n,\lambda)} \frac{1}{z^{n+1}} e^{-\frac{n}{2z^2}} \left[\int_0^{t(z)} \frac{nt^2}{e^{2z^2}} dt \right] dz$$

où :

$$t(z) = \sqrt{z^2 \left[1 - \frac{2}{n} \text{Log}(\lambda z^{n+1}) \right] - 1}$$

Nous devons alors rechercher la valeur $\lambda_{n,\alpha}^*$ réalisant :

$$F_n^*(\lambda_{n,\alpha}^*) = \frac{1-\alpha}{2} K_n$$

sachant d'une part (propriété 1) que $F_n^*(\lambda)$ est décroissante, d'autre part (propriété 2) que λ appartient à l'intervalle :

$$\left[0, e^{-\frac{1}{2}} \left(1 + \frac{1}{n} \right)^{\frac{n+1}{2}} \right]$$

Un procédé de bisection permet de résoudre ce problème.

Connaissant $\lambda_{n,\alpha}^*$ il est aisé d'en déduire la valeur du coefficient $A_{n,\alpha}$ introduit en 3.5 puisque (propriété 3) :

$$A_{n,\alpha} = \frac{1}{n+1} \left[-\frac{n}{2} + \text{Log}(\lambda_{n,\alpha}^*) \right]$$

3.7 - Programme ALGOL

La difficulté principale de notre méthode est le calcul de la fonction $F_n^*(\lambda)$ telle que :

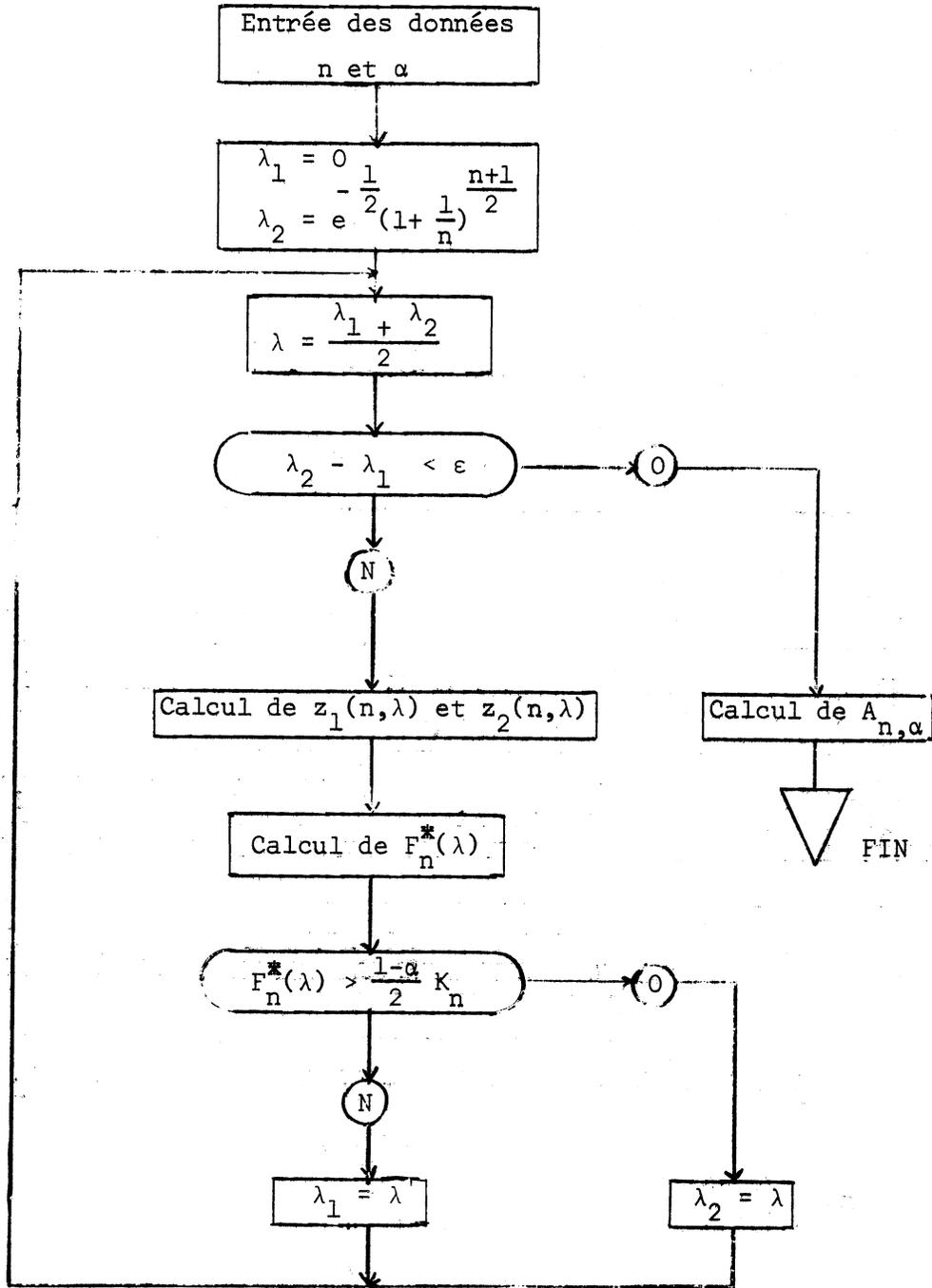
$$F_n^*(\lambda) = e^{\frac{n}{2}} \int_{z_1(n,\lambda)}^{z_2(n,\lambda)} \frac{1}{z^{n+1}} e^{-\frac{n}{2z^2}} \left[\int_0^{t(z)} \frac{nt^2}{e^{2z^2}} dt \right] dz$$

Considérons l'intégrale en t et faisons le changement de variable :

$$u^2 = \frac{nt^2}{z^2}$$

On a alors :

ORGANIGRAMME DE RESOL



$$\int_0^{t(z)} e^{-\frac{nt^2}{2z^2}} dt = \frac{z}{\sqrt{n}} \int_0^{u(z)} e^{-\frac{u^2}{2}} du$$

$$= z \sqrt{\frac{2\pi}{n}} \left[N(u(z); 0,1) - \frac{1}{2} \right]$$

où $N(x; 0,1)$ est la fonction de répartition d'une loi normale centrée, réduite, et $u(z)$ est telle que :

$$u(z) = \frac{n}{z} t(z)$$

$F_n^*(\lambda)$ devient donc :

$$e^{\frac{n}{2}} \sqrt{\frac{2\pi}{n}} \int_{z_1(n,\lambda)}^{z_2(n,\lambda)} \frac{1}{z^n} e^{-\frac{n}{2z^2}} \left[N(u(z); 0,1) - \frac{1}{2} \right] dz$$

on encore :

$$F_n^*(\lambda) = e^{\frac{n}{2}} \sqrt{\frac{2\pi}{n}} \int_{z_1(n,\lambda)}^{z_2(n,\lambda)} H(z) dz$$

intégrale que nous calculons en utilisant une méthode de Romberg (extrapolation de Richardson appliquée à la formule des trapèzes).

Calcul de $z_1(n,\lambda)$ et $z_2(n,\lambda)$.

Les deux valeurs $z_1(n,\lambda)$ et $z_2(n,\lambda)$ doivent réaliser :

$$\frac{1}{z^{n+1}} e^{-\frac{n}{2}\left(\frac{1}{z^2} - 1\right)} = \lambda$$

on encore :

$$\frac{n}{2}\left(1 - \frac{1}{z^2}\right) - \text{Log}(\lambda z^{n+1}) = 0$$

Si nous considérons la fonction $F(z)$ du réel z définie pour z positif ou nul par :

$$F(z) = \frac{n}{2} \left(1 - \frac{1}{z^2}\right) - \text{Log}(\lambda z^{n+1})$$

il suffit de rechercher les deux valeurs pour lesquelles F s'annule.

Or il est facile de voir que $F(z)$ est croissante pour :

$$0 \leq z \leq \sqrt{\frac{n}{n+1}}$$

et décroissante par :

$$z > \sqrt{\frac{n}{n+1}}$$

Par conséquent on a :

$$0 \leq z_1 \leq \sqrt{\frac{n}{n+1}}$$

et une méthode de bisection permet de trouver cette première racine.

En ce qui concerne z_2 le premier problème consiste à encadrer la racine. Nous savons seulement que :

$$z_2 > \sqrt{\frac{n}{n+1}}$$

On peut, par exemple, rechercher une valeur z^* telle que :

$$F(z^*) < 0$$

$$F(z^*-1) \geq 0$$

et par une même méthode que précédemment en déduire z_2 .

3.8 - Tables

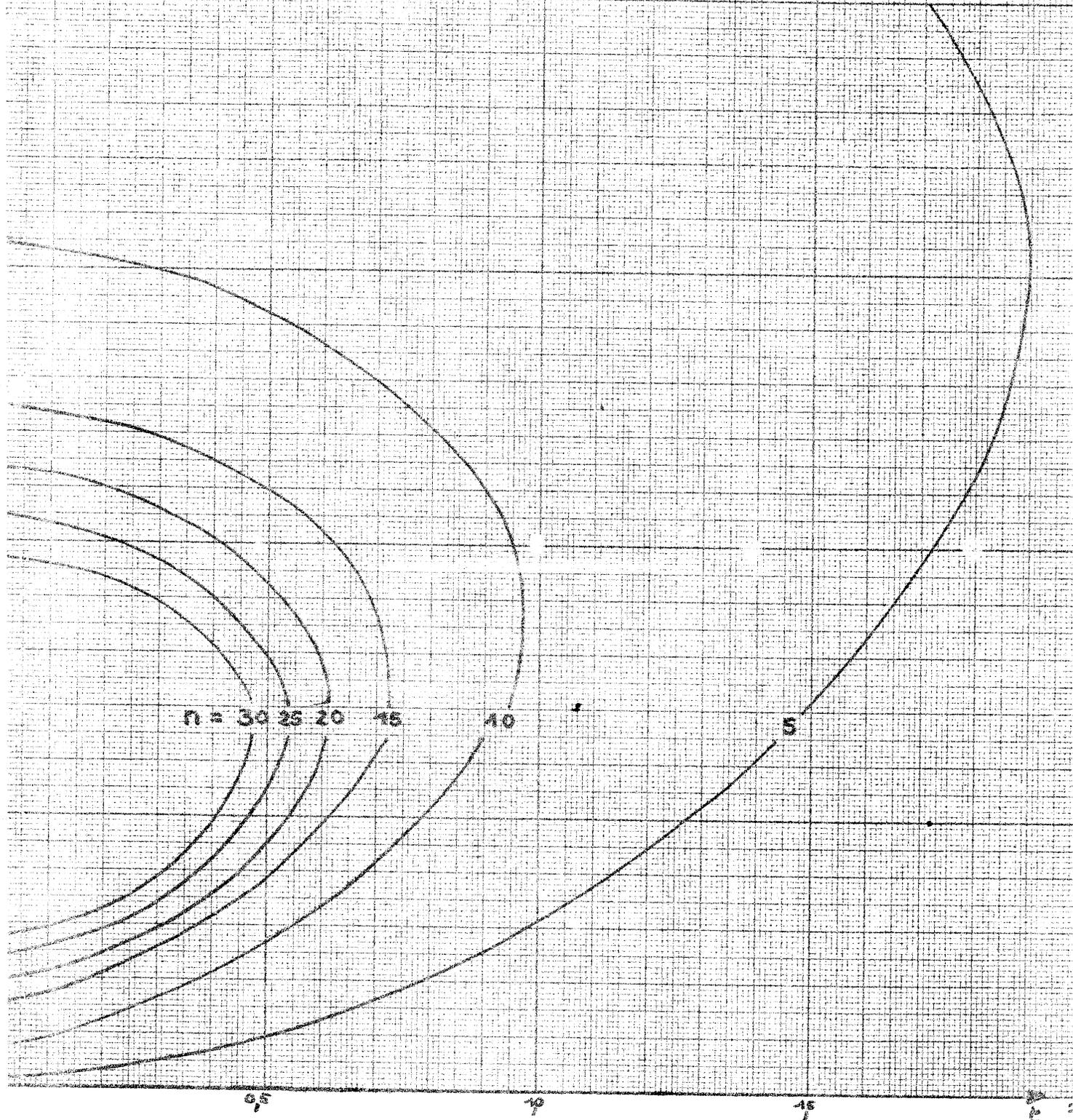
Nous avons tabulé le coefficient $A_{n,\alpha}$ pour $n = 5(1) 30$ et $\alpha = 0.1 \%, 1\%, 5 \%$. De plus pour chaque couple (n,α) nous avons indiqué les deux valeurs Z_{\min} et Z_{\max} .

On trouvera représenté sur le graphique (c) pour $\alpha = 5 \%$ les courbes correspondant aux valeurs $n = 5(5) 30$ et sur le graphe (d) les courbes correspondant aux valeurs $\alpha = 0.1 \%, 1 \%, 5 \%$, n étant égal à 15.

En ce qui concerne la table les calculs ont été réalisés sur calculateur et les valeurs numériques ne sont fournies qu'à titre d'illustration de la procédure établie, qui permet, en fait, de calculer $A_{n,\alpha}$ pour tout couple (n,α) .

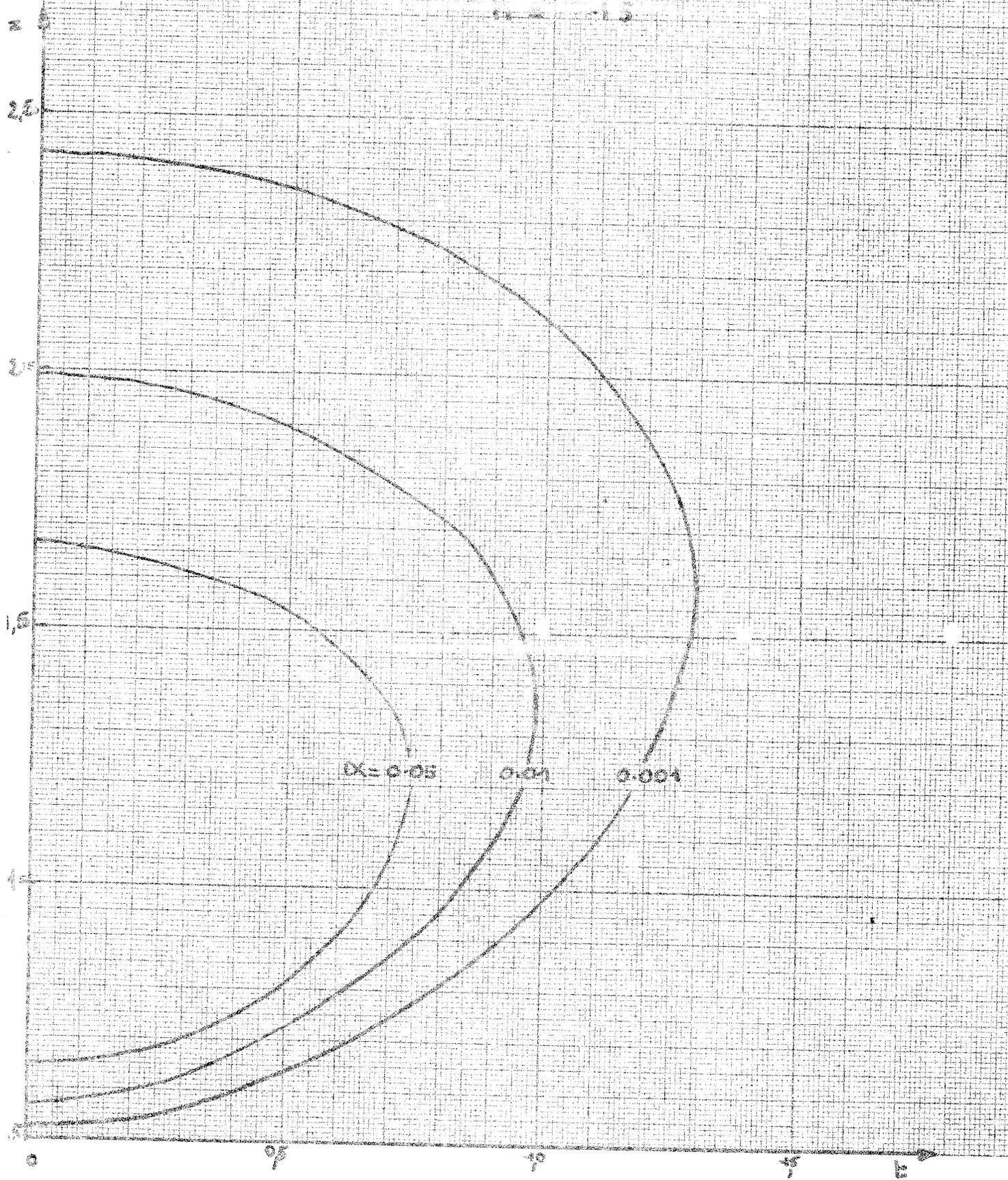
Graphique (c)

$\alpha = 0.05$





Graphique (d)



N	ALPHA=0.001			ALPHA=0.01			ALPHA=0.05		
	A	ZMIN	ZMAX	A	ZMIN	ZMAX	A	ZMIN	ZMAX
4				1.9391	0.3692	6.8942	1.4012	0.4199	3.9576
5				1.5724	0.4116	4.7292	1.1686	0.4639	3.0790
6	1.8100	0.3958	6.0390	1.3541	0.4451	3.7574	1.0305	0.4981	2.6345
7	1.5966	0.4218	4.8450	1.2090	0.4727	3.2109	0.9392	0.5260	2.3655
8	1.4346	0.4451	4.0875	1.1057	0.4960	2.8617	0.8744	0.5492	2.1841
9	1.3176	0.4647	3.6077	1.0289	0.5160	2.6205	0.8262	0.5690	2.0533
10	1.2254	0.4823	3.2631	0.9690	0.5334	2.4418	0.7888	0.5862	1.9536
11	1.1537	0.4975	3.0138	0.9214	0.5489	2.3051	0.7596	0.6011	1.8765
12	1.0921	0.5119	2.8114	0.8824	0.5628	2.1961	0.7534	0.6145	1.8130
13	1.0427	0.5245	2.6561	0.8501	0.5753	2.1075	0.7153	0.6265	1.7601
14	1.0005	0.5361	2.5281	0.8227	0.5867	2.0336	0.6980	0.6376	1.7148
15	0.9642	0.5468	2.4211	0.7992	0.5971	1.9709	0.6835	0.6475	1.6765
16	0.9332	0.5566	2.3318	0.7791	0.6066	1.9177	0.6710	0.6566	1.6433
17	0.9062	0.5657	2.2555	0.7614	0.6155	1.8710	0.6601	0.6649	1.6142
18	0.8816	0.5743	2.1870	0.7458	0.6237	1.8301	0.6505	0.6726	1.5884
19	0.8603	0.5823	2.1286	0.7320	0.6313	1.7938	0.6420	0.6798	1.5654
20	0.8413	0.5897	2.0767	0.7196	0.6385	1.7613	0.6344	0.6865	1.5446
21	0.8241	0.5968	2.0306	0.7085	0.6452	1.7323	0.6275	0.6928	1.5257
22	0.8084	0.6034	1.9886	0.6984	0.6515	1.7058	0.6213	0.6986	1.5085
23	0.7942	0.6098	1.9508	0.6892	0.6575	1.6816	0.6157	0.7042	1.4928
24	0.7817	0.6156	1.9178	0.6810	0.6630	1.6599	0.6106	0.7093	1.4784
25	0.7697	0.6213	1.8862	0.6734	0.6684	1.6398	0.6059	0.7143	1.4651
26	0.7587	0.6267	1.8574	0.6663	0.6735	1.6210	0.6016	0.7189	1.4527
27	0.7490	0.6317	1.8318	0.6599	0.6783	1.6040	0.5977	0.7233	1.4414
28	0.7395	0.6367	1.8069	0.6539	0.6829	1.5879	0.5940	0.7276	1.4306
29	0.7311	0.6413	1.7849	0.6484	0.6872	1.5731	0.5906	0.7316	1.4206
30	0.7229	0.6459	1.7634	0.6432	0.6914	1.5590	0.5874	0.7354	1.4111

```

'PROCEDURE' RESOL(N,ALPHA) FF
'VALEUR' N,ALPHA FF
'ENTIER' N FF 'REEL' ALPHA FF
'DEBUT'
'REEL' 'PROCEDURE' NORMALE(X,MOY,ECART) FF
'VALEUR' X,MOY,ECART FF 'REEL' X,MOY,ECART FF
'DEBUT' 'ENTIER' I FF
'REEL' SGNE,X2,Y,S,T,P1,P2,Q1,Q2,M FF
XF=(X-MOY)/ECART FF

```

```

SGNEF=SIGNE(X) FF
XF=ABS(X) FF
X2F=X*X FF
YF=0.3989422804*EXP(-0.5*X2) FF
'SI' X 'INFER' 3.5 'ALORS'
'DEBUT' SF=XF=Y*X FF
'POUR' IF=3,I+2 'TANTQUE' S 'NONEG' T 'FAIRE'
'DEBUT' TF=S FF
XF=X*X2/I FF
SF=S+X
'FIN' FF
NORMALEF=0.5+SGNE*S

```

```

'FIN'
'SINON'
'DEBUT' Q1F=X FF
P2F=Y*X FF
P1F=Y FF
Q2F=X2+1.0 FF
MF=Y/X FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS'
'DEBUT' MF=1-M FF
TF=1-T
'FIN' FF
'POUR' IF=2,I+1 'TANTQUE'
((M 'NONEG' T) 'ET' (S 'NONEG' T)) 'FAIRE'
'DEBUT' SF=X*P2+I*P1 FF
P1F=P2 FF
P2F=S FF
SF=X*Q2+I*Q1 FF
Q1F=Q2 FF
Q2F=S FF
SF=M FF
MF=T FF
TF=P2/Q2 FF
'SI' SIGNE=1 'ALORS' TF=1-T FF
'FIN' FF
NORMALEF=T

```

```

'FIN' FF
FINORMF 'FIN' FF
'REEL' 'PROCEDURE' INSIRO(F,A,B,ORDMAX,PREC,SORT,RES)FF
'VALEUR' A,B,ORDMAX,PRECFF 'REEL' 'PROCEDURE' FFF 'BOOLEEN' SORTFF
'REEL' A,B,PRECFF 'ENTIER' ORDMAXFF 'REEL' 'TABLEAU' RESFF
'DEBUT' 'REEL' L,T,P,MAFF 'ENTIER' N,J,I,FAFF
LF=B-AFF SORTF='FAUX' FFFMAF=RES.(1).F=L*0.5*(F(A)+F(B))FFNF=IFF
'POUR' JF=1,PAS'1'JUSQUA'ORDMAX'FAIRE'
'DEBUT' TF=0.0FFPF=L/NFF
'POUR' IF=1,PAS'1'JUSQUA'N'FAIRE' TF=T+F(A+P*(I-0.5))FF
RES.(J+1).F=(P*T+RES.(J).)/2.0FFFAFF=IFF

```

```

'POUR' IF=J'PAS' -1 'JUSQUA' 1 'FAIRE'
'DEBUT' FACF=4*FACFF
RES.(1).F=RES.(I+1).+(RES.(I+1).-RES.(I).)/(FAC-1)
'FIN' FF
'SI' ABS((RES.(1).-MA)/RES.(1).)'INFEG'PREC'ALORS'
'ALLER A' TERMFF
MAF=RES.(1).FFNF=0*N
'FIN' FF
'ALLERA' AFECT FF
TERMF SORTF='VRAI' FF
AFECTF INSIROF=RES.(1).
'FIN' FF
'REEL' 'PROCEDURE' GAMMA(X) FF 'REEL' XFF
'DEBUT' 'REEL' H,Y FF
HF=1.0 FF YF=X FF
AIF 'SI' Y=0 'ALORS' HF=10**5 'SINON'
'SI' Y=2.0 'ALORS' 'ALLERA' A2 'SINON'
'SI' Y'INFER'2.0 'ALORS'
'DEBUT' HF=H/Y FF YF=Y+1.0 FF 'ALLERA' A1 'FIN'
'SINON' 'SI' Y'SUPEG'3 'ALORS' 'DEBUT'
YF=Y-1.0 FF HF=H*Y FF 'ALLERA' A1 'FIN'
'SINON' 'DEBUT' YF=Y-2.0 FF
HF=(((0.0016063118*Y+.0051589951)*Y+.0044511400)*Y+.0721101567)*
Y+.0821117404)*Y+.4117741955)*Y
.4227874605)*Y+.9999999758)*H
'FIN' FF
A2F GAMMAF=H
'FIN' GAMMA FF
'REEL' 'PROCEDURE' FN(LAMBDA,Z1,Z2) FF
'VALEUR' LAMBDA,Z1,Z2 FF
'REEL' LAMBDA,Z1,Z2 FF
'DEBUT' 'REEL' 'PROCEDURE' F(Z) FF
'VALEUR' Z FF 'REEL' Z FF
FF=(N/2)*(1-1/Z**2)-LN(LAMBDA*Z**(N+1)) FF
'REEL' 'PROCEDURE' H(Z) FF
'VALEUR' Z FF 'REEL' Z FF
'DEBUT' 'REEL' AUX FF
AUXF=RAC2(N-N/Z**2-2*LN(LAMBDA*Z**(N+1))) FF
HF=EXP(-N/(2*Z**2))*(NORMALE(AUX,0.0,1.0)-0.5)/Z**N
'FIN' FF
'REEL' A1,A2 FF
'TABLEAU' RES.(1F9). FF
'BOOLEEN' SORT FF
A1F=0.0 FF
A2F=RAC2(N/(N+1)) FF
RETOURF Z1F=(A1+A2)/2 FF
'SI' A2-A1 'INFER' F*-8 'ALORS' 'ALLERA' CONT FF
'SI' F(Z1) 'SUPER' 0.0 'ALORS' A2F=Z1 'SINON' A1F=Z1 FF
CONTF A2F=RAC2(N/(N+1)) FF
DEBUTF A1F=A2 FF
A2F=A2+1.0 FF
'SI' F(A2) 'SUPER' 0.0 'ALORS' 'ALLERA' DEBUT FF
RETOURIF Z2F=(A1+A2)/2 FF
'SI' A2-A1 'INFER' F*-8 'ALORS' 'ALLERA' CONT1 FF
'SI' F(Z2) 'INFER' 0.0 'ALORS' A2F=Z2 'SINON' A1F=Z2 FF
'ALLERA' RETOUR1 FF
CONT1F FNF=INSIRO(H,Z1,Z2,8,F*-5,SORT,RES) FF
'SI' 'NON' SORT 'ALORS'
ECRIRE('('PRECISION NON A)'),('TTEINTE'))
'FIN' FF

```

```
'REEL' CTE,MIN,MAX,ZMIN,ZMAX,L FF  
CTEF=GAMMA((N-1)/2)*2**((N/2-2)*(1-ALPHA)/N**((N-1)/2) FF  
MINF=0.0 FF  
MAXF=EXP(-0.5)*(1+1/N)**((N+1)/2) FF  
REVLFF=(MIN+MAX)/2 FF  
'SI' MAX-MIN 'INFER' F*-5 'ALORS' 'ALLERA' STE FF  
'SI' FN(L,ZMIN,ZMAX) 'INFER' 0.0 'ALORS' MAXF=L 'SINON' MINF=L FF  
'ALLERA' REV FF  
STEFECRIRE((LN(L)-N/2)/(N+1),ZMIN,ZMAX)  
'FIN' FF
```

BIBLIOGRAPHIE

- [1] BARRA J.R. (1967) - Contrôle statistique d'une suite de digits aléatoires.
Revue de Statistique Appliquée, Vol. XV, n° 3.
- [2] BARRA J.R. (1964) - Cours de Statistique Mathématique,
Faculté des Sciences de Grenoble.
- [3] BARRA J.R. (1967) - Séminaire de Statistique,
Faculté des Sciences de Grenoble.
- [4] BIRNBAUM Z.W. et TINGEY F.H. (1951) - One-sided confidence contours for distribution
functions,
The Annals of Mathematical Statistics, Vol. 22, 592:596.
- [5] COOPER B.E. (1967) - ASCOP - A statistical computing procedure,
Applied Statistics, Vol. XVI, n° 2.
- [6] COOPER B.E. (1968) - Basic subroutine for the input of numbers, words and special
characters,
The Computer Journal, Vol. 11, N° 2.
- [7] ESMENJAUD BONNARDEL M. (1965) - Etude statistique des décimale de π ,
Revue Française du Traitement de l'Information. Chiffres, Vol 8,
n° 4.
- [8] FERGUSON T.S. (1967) Mathematical Statistics : A decision theoretic approach,
Academic Press New York-London.
- [9] FIX (1949) - Table of the non-central χ^2 ,
University of California, Vol. I, n° 2.
- [10] GOWER J.C., MARTIN A.H. et SIMSON H.R. (1965) - An outline of a programming language
for the analysis of surveys, experiments and multivariate data,
Proc. Int. Symp. for Methods of Field Experimentation, Halle.
- [11] HARRIS M., HORVITZ D.G. et MOOD A.M. (1948) - On the determination of sample sizes in
designing experiments,
Journal fo the American Statistical Association, N° II,
391:402.
- [12] HISLEUR G. (1967) - Une estimation optimale des paramètres d'une loi normale,
Revue de Statistique Appliquée, Vol. XV, N° 3.
- [13] HISLEUR G. (1969) Détermination de la taille de l'échantillon dans un test de
Student
Revue de Statistique Appliquée, Vol. XVII, N° 1.
- [14] LUDWIG O.G. (1963) - Incomplete Beta ratio,
Communication of the A.C.M., Vol. 6, N° 6.
- [15] MORICE E. (1968) - Puissance de quelques tests classiques. Effectif d'échantillon
pour des risques donnés,
Revue de Statistique Appliquée, Vol. XVI, n° 1.

- [16] NEYMAN J. et PEARSON E.S. (1928) - Biometrika 20 A, 175:240.
- [17] NEYMAN J. et TOKARSKA B. (1936) -Errors of the second kind in testing Student's hypothesis,
Journal of the American Statistical Association, Vol. 31, 318:326
- [18] PATNAIK P.B. (1949) - The non-central χ^2 and F distributions,
Biometrika 36, 202 : 232.
- [19] STEIN C. (1945) - A two sample test for a linear hypothesis whose power is independent of the variance,
The Annals of Mathematical Statistics, Vol. 16, 243 : 258.
- [20] MEETING ON "STATISTICAL PROGRAMMING" OF CHILTON (1967) - Discussion,
Applied Statistics, Vol. XVI, N° 2.

VU

Grenoble, le

Le Président de la Thèse

VU

Grenoble, le

Le Doyen de la Faculté des Sciences

Vu, et permis d'imprimer,

Le Recteur de l'Académie de GRENOBLE

