

ANDRÉ G. LAURENT

Distribution d'échantillon et de caractéristiques d'échantillon quand la population de référence est Laplace-gaussienne de paramètres inconnus

Journal de la société statistique de Paris, tome 96 (1955), p. 262-296

http://www.numdam.org/item?id=JSFS_1955__96__262_0

© Société de statistique de Paris, 1955, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

VI

VARIÉTÉ

Distribution d'échantillon et de caractéristiques d'échantillon quand la population de référence est Laplace-Gaussienne de paramètres inconnus.

INTRODUCTION

Étant donné un échantillon de n observations indépendantes, extraites au hasard d'une population de référence Laplace-Gaussienne $N(m, \sigma)$, de paramètres m et σ , la distribution des caractéristiques d'échantillon les plus élémentaires : moyenne, variance, moment d'ordre deux, etc... s'obtient facilement quand m et σ sont connus. Ces résultats classiques figurent dans tous les manuels. Ils donnent le moyen de construire, réciproquement, des « intervalles de confiance » pour « l'estimation » de différentes caractéristiques de la population sur la base de l'information — incomplète — fournie par un échantillon et permettent le « test d'hypothèses » relatives aux valeurs de tels paramètres. (En contrôle de fabrication, par exemple, la connaissance de la distribution de la moyenne et de la variance d'un échantillon permet de construire des « cartes de contrôle » en vue de « tester » la stabilité du processus de production, lorsque la moyenne et l'écart-type initiaux du processus sont connus.)

Dans de nombreuses situations, par contre, l'un des paramètres de la population de référence ou même leur ensemble n'est pas connu ; les paramètres inconnus doivent être remplacés par des « estimations » et la distribution des caractéristiques d'échantillon en fonction des paramètres estimés pose un problème nouveau. En 1929, J. O. Irwin [1] a donné la loi de distribution d'un échantillon lorsque la moyenne de la population de référence est remplacée par la moyenne de l'échantillon. En 1935, W. R. Thompson [2] a obtenu la loi de distribution d'une observation choisie au hasard dans un échantillon lorsque la moyenne de la population et son écart-type sont remplacés par les caractéristiques correspondante de l'échantillon (1). En 1946, A. H. Bowker et plusieurs autres statisticiens étudient la probabilité intégrale de la loi de Laplace-Gauss quand les paramètres sont remplacés par la moyenne et l'écart-type d'un échantillon (des tables figurent in « *Techniques of Statistical Analysis* », Mac Graw Hill, 1947).

Dans la présente note, deux situations sont essentiellement considérées : I) L'échantillon est extrait d'une population $N(m, \sigma)$, m ou (et) σ inconnu(s), l'information sur le(s) paramètre(s) étant fournie par un échantillon indépendant. II) L'échantillon est extrait d'un échantillon plus important de moyenne et (ou) d'écart-type donnée(s) extrait lui-même d'une population $N(m, \sigma)$, m

(1) Quoique mentionné par Cramer (*Mathematical methods of Statistics*, p. 390), ce résultat semble peu connu : G. J. Lieberman le démontre à nouveau quand il en a besoin in « *Sampling Plans for Inspection by Variables* » (*J. A. S. A.*, June 1955, pp. 457-517).

ou (et) σ inconnu(s). Cette situation couvre notamment le cas où un échantillon est dit être extrait d'une population finie Laplace-Gaussienne, une telle population n'étant en fait elle-même qu'un échantillon plus important prélevé dans une population de référence Laplace-Gaussienne.

I. — ÉCHANTILLON D'EFFECTIF i EXTRAIT DE $N(m, \sigma)$, σ ET (OU) m INCONNU(S)

Soient $\xi = (\xi_1 \ \xi_2 \ \dots \ \xi_i)$ l'échantillon considéré, $\bar{\xi}$ sa moyenne et S_{ξ}^2 sa variance.

Soient $X = (X_1 \ X_2 \ \dots \ X_n)$ un échantillon indépendant, \bar{X}_n sa moyenne et S_n^2 sa variance.

Dans le cas où n est important, m ou (et) σ , est (sont) remplacé(s) par leur(s) estimation(s) : \bar{X}_n ou (et) S_n et la distribution de ξ est approximativement normale. Il n'en est plus de même si n est faible.

I.1. m INCONNU, σ CONNU

m étant estimée par \bar{X}_n , on se propose d'étudier l'échantillon $\xi' = \xi - \bar{X}_n = (\xi_1 - \bar{X}_n, \dots, \xi_i - \bar{X}_n)$.

I.1.1. Distribution de ξ'

On écrit la distribution simultanée de ξ et \bar{X}_n , on effectue le changement de variables $\xi' = \xi - \bar{X}_n$ et l'on intègre par rapport à \bar{X}_n , on obtient

$$\frac{\sqrt{n/n+i}}{(2\pi)^{i/2} \sigma^i} e^{-\frac{Q'}{2\sigma^2}} d(\xi - \bar{X})$$

où Q' est la forme quadratique $Q' = (\xi - \bar{X}_n) B (\xi - \bar{X}_n)'$
et

$$B = \left(\delta_{jk} - \frac{1}{n+i} \right)_{i \times i}$$

L'inverse de la matrice des covariantes est $\frac{B}{\sigma^2} = -\frac{1}{(n+i)\sigma^2} (-(n+i)\delta_{jk} + 1)_{i \times i}$

Les termes diagonaux de la matrice des covariances sont alors :

$$\sigma_{\xi'_1}^2 = \dots = \sigma_{\xi'_i}^2 = \frac{\left| -\frac{1}{(n+i)\sigma^2} (-(n+i)\delta_{jk} + 1) \right|_{(i-1) \times (i-1)}}{\left| -\frac{1}{(n+i)\sigma^2} (-(n+i)\delta_{jk} + 1) \right|_{i \times i}}$$

En tenant compte du fait que les valeurs caractéristiques d'une matrice $[(d-1)\delta_{jk} + 1]_{l \times l}$ sont $\alpha - 1$ d'ordre de multiplicité $l - 1$ et $\alpha + l - 1$, on trouve

$$\boxed{\sigma_{\xi'_j}^2 = \sigma^2 \frac{n+1}{n}} \quad j = 1 \text{ à } i$$

La matrice des corrélations est alors $P = \frac{n}{n-1} B^{-1}$; mais si ρ est le coefficient de corrélation entre deux variables ξ'_j et ξ'_k , $P = \rho \left[\left(\frac{1}{\rho} - 1 \right) \delta_{jk} + 1 \right]$; en égalant les déterminants de ces deux expressions, on trouve

$$\boxed{\rho = \frac{1}{n+1}}$$

ξ'_j est distribué de façon Laplace-Gaussienne et la matrice des covariances est $\sigma^2 \left(\delta_{jk} + \frac{1}{n} \right)$.

I.1.2. Distribution de la moyenne $\bar{\xi}' = \bar{\xi} - \bar{X}_n$.

$\bar{\xi}'$ a une distribution Laplace-Gaussienne de moyenne $\boxed{m_{\bar{\xi}'} = 0}$

Sa variance est égale à la somme des éléments de la matrice des covariances de I.1.1. divisée par i^2 , soit :

$$\boxed{\sigma_{\bar{\xi}'}^2 = \frac{n+i}{n} \frac{i}{i} \sigma^2}$$

I.1.3. Distribution du moment d'ordre deux : $m_2(\xi')$

La fonction caractéristique $\varphi_{m_2}(t)$ de la variable m_2 est donnée par

$$\varphi_{m_2}(t) = \int \int \int \frac{|P^{-1}|^{1/2}}{(2\pi)^{i/2} \sigma_{\xi'}^2} e^{-\frac{[(\xi - \bar{x}_n) P^{-1} (\xi - \bar{x}_n)]' - 2 \vec{t} m_2 \sigma_{\xi'}^2}{2 \sigma_{\xi'}^2}} d\xi'$$

$= \frac{|P^{-1}|^{1/2}}{\Delta^{1/2}}$ où Δ est le déterminant de la forme quadratique entre crochets. α_j ($j = 1$ à i) étant les valeurs caractéristiques de la matrice de corrélation P , (à savoir $\frac{n}{n+1}$, $i-1$ fois et $\frac{n+i}{n+1}$) on obtient

$$\varphi_{m_2}(t) = \frac{1}{\prod \left(1 - 2 \vec{t} i \sigma_{\xi'}^2 \alpha_j \frac{t}{i} \right)^{1/2}} = \left[1 - 2 \vec{t} i \frac{\sigma^2}{i} t \right]^{-\frac{i-1}{2}} \left[1 - 2 \vec{t} i \frac{\sigma^2}{i} \frac{n+i}{n} t \right]^{-\frac{1}{2}}$$

d'où, en consultant une table d'intégrales de Fourier.

$$\boxed{p(m_2) dm_2 = \sqrt{\frac{n}{n+i}} \cdot \left(\frac{im_2}{2\sigma^2} \right)^{\frac{i-2}{2}} \frac{e^{-\frac{im_2}{2\sigma^2}}}{\Gamma\left(\frac{i}{2}\right)} \cdot {}_1F_1\left(\frac{1}{2}, \frac{i}{2}; \frac{i}{n+i} \frac{im_2}{2\sigma^2}\right) d\left(\frac{im_2}{2\sigma^2}\right)}$$

${}_1F_1$ désignant la fonction hypergéométrique confluyente

$${}_1F_1(\alpha, \beta, Z) = 1 + \frac{\alpha}{\beta} \frac{Z}{1!} + \frac{\alpha(\alpha+1)}{\beta(\beta+1)} \frac{Z^2}{2!} + \dots$$

I.2. m CONNU, σ INCONNU

σ étant estimé par S_n , on se propose d'étudier l'échantillon ζ' .

$$\zeta' = \frac{\xi - m}{S_n} = \left(\frac{\xi_1 - m}{S_n}, \dots, \frac{\xi_i - m}{S_n} \right)$$

I.2.1. Distribution de ζ' .

On écrit la distribution simultanée de ξ et S_n , on effectue le changement de variable $\zeta' = \frac{\xi - m}{S_n}$, on intègre par rapport à S_n , on obtient :

$$\frac{\Gamma\left(\frac{n+i-1}{2}\right)}{\pi^{i/2} n^{i/2} \Gamma\left(\frac{n-1}{2}\right)} \left[1 + \frac{1}{n} \left(\frac{\xi - m}{S_n}\right)^2 \right]^{-\frac{n+i-1}{2}} d\left(\frac{\xi - m}{S_n}\right)$$

C'est une généralisation à plusieurs variables de la distribution de Student. Lorsque $i = 1$, on obtient évidemment l'usuelle distribution de Student avec $n - 1$ degrés de liberté.

I.2.2. Distribution de la moyenne $\bar{\zeta}' = \frac{\bar{\xi} - m}{S_n}$.

Il s'agit d'un cas particulier de I.2.1., celui où $i = 1$ et où la variable étudiée est $\bar{\xi} \sqrt{i}$ au lieu de ξ .

on obtient

$$\frac{1}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)} \left[1 + \left(\frac{\bar{\xi} - m}{S_n}\right)^2 \frac{i}{n} \right]^{-\frac{n}{2}} d\left(\sqrt{\frac{i}{n}} \frac{\bar{\xi} - m}{S_n}\right)$$

distribution de Student à $n - 1$ degrés de liberté, résultat bien connu; l'usuelle variable t à utiliser pour les tables est $t = \frac{\bar{\xi} - m}{S_n/\sqrt{i}} \sqrt{\frac{n-1}{n}}$

I.2.3. Distribution de la variance $S_{\zeta'}^2$.

$S_{\zeta'}^2 = \frac{S_{\xi}^2}{S_n^2}$, il est bien connu que $F = \frac{S_{\xi}^2 i/i - 1}{S_n^2 n/n - 1}$ suit une distribution de Snedecor avec $i - 1$ et $n - 1$ degrés de liberté.

$$F(i - 1, n - 1) = \frac{i(n-1)}{n(i-1)} S_{\zeta'}^2$$

I.2.4. Distribution du moment d'ordre deux $m_2(\zeta')$.

Pour les mêmes raisons

$$F(i, n - 1) = \frac{n-1}{n} m_2$$

I.3. m ET σ INCONNUS

m et σ étant estimés par \bar{X}_n et S_n , On se propose d'étudier l'échantillon

$$\Theta = \frac{\xi - \bar{X}_n}{S_n} = \frac{\xi'}{S_n} = \left(\frac{\xi_1 - \bar{X}_n}{S_n}, \dots, \frac{\xi_i - \bar{X}_n}{S_n} \right)$$

I.3.1. Distribution de Θ .

La distribution de ξ' a été obtenue en § I.1.1. On écrit la distribution simultanée de ξ' et de S_n , on effectue le changement de variable $\Theta = \frac{\xi'}{S_n}$, on intègre par rapport à S_n , on obtient :

$$\frac{\Gamma\left(\frac{n+i-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)} \cdot \frac{1}{\pi^{i/2} n^{\frac{i-1}{2}} (n+i)^{1/2}} \left[1 + \frac{1}{n} \Theta B \Theta'\right]^{-\frac{n+i-1}{2}} d\Theta$$

où
$$B = \left(\partial_{jk} - \frac{1}{n+i} \right)$$

C'est une généralisation de la distribution de Student qui dans le cas de $i = 1$ conduit à une distribution de Student à $n - 1$ degrés de liberté (la variable t à utiliser pour les tables étant

$$t = \frac{\xi - \bar{X}_n}{S_n} \cdot \sqrt{\frac{n-1}{n+1}}$$

I.3.2. Distribution de la forme quadratique $\frac{1}{n} \Theta B \Theta' = \Theta F \Theta'$.

Les valeurs caractéristiques de la matrice F sont : $\frac{1}{n}$ (ordre de multiplicité $i - 1$) et $\frac{1}{n+i}$. Il existe une transformation orthogonale des variables Θ en variables Y transformant F en sa forme diagonale, $\Theta F \Theta'$ en la forme $\frac{Y_1^2}{n+i} + \sum_{j=2}^i \frac{Y_j^2}{n} d\Theta$ en dY . Les surfaces de densité de probabilité constante pour la loi I.3.1. sont des ellipsoïdes de révolution de grand axe proportionnel à $\sqrt{n+i}$ et de petits axes proportionnels à \sqrt{n} . Le changement de variables

$$Z_1 = \frac{Y_1}{\sqrt{n+i}} \quad Z_j = \frac{Y_j}{\sqrt{n}} \quad j = 2 \text{ à } i$$

transforme la forme quadratique en ZZ' et la distribution en

$$\frac{\Gamma\left(\frac{n+i-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right) \pi^{i/2}} [1 + Z Z']^{-\frac{n+i-1}{2}} dZ,$$

les ellipsoïdes étant transformés en sphères; la distribution de Z est une distribution sphérique, il en résulte que la distribution de ZZ' est obtenue en multipliant la densité de probabilité par

$$(ZZ')^{\frac{i-2}{2}} \frac{\pi^{i/2}}{\Gamma\left(\frac{i}{2}\right)}$$

L'élément de volume $dZ = \frac{1}{2} (ZZ')^{\frac{i-2}{2}} dZZ' d\Sigma_i$; intégrant $d\Sigma_i$, on obtient

$$\Sigma_i = \frac{2\pi^{i/2}}{\Gamma\left(\frac{i}{2}\right)}$$

d'où la loi

$$\frac{1}{B\left(\frac{i}{2}, \frac{n-1}{2}\right)} (1 + ZZ')^{-\frac{n+i-1}{2}} (ZZ')^{\frac{i-2}{2}} dZZ'$$

Il en résulte que

$$\frac{n-1}{i} ZZ' = \frac{n-1}{i n} \sum \left(\nu_{jk} - \frac{1}{n+i} \right) \Theta_j \Theta_k = F(i, n-1)$$

suit une distribution de Snedecor à i et $n-1$ degrés de liberté.

I.3.3. Fonction caractéristique de Θ .

a) Du fait que la distribution de Z est sphérique, il est facile d'obtenir sa fonction caractéristique.

C'est un résultat bien connu de la transformation de Fourier que si la densité de probabilité de Z est sphérique, c'est-à-dire de la forme $f(ZZ')$, la fonction caractéristique $\Phi_z(u)$ de Z avec $u = (u_1 \dots u_i)'$ est également sphérique et

$$\Phi_z(u'u) = (2\pi)^{i/2} (u'u)^{\frac{1-i}{4}} \mathcal{H}_{\frac{i}{2}-1} \left\{ (ZZ')^{\frac{i-1}{4}} f(ZZ'); (u'u)^{1/2} \right\}$$

où \mathcal{H} désigne la transformation de Hankel.

Il en résulte que

$$\Phi_z(u'u) = (2\pi)^{1/2} (u'u)^{\frac{i-1}{4}} \mathcal{H}_{\frac{i}{2}-1} \left\{ (ZZ')^{\frac{i-1}{4}} \frac{\Gamma\left(\frac{n+i-1}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right) \pi^{i/2}} (1 + ZZ')^{-\frac{n+i-1}{2}}; (u'u)^{1/2} \right\}$$

on trouvera ce type de transformées dans une table de transformées de Hankel, (Erdelyi, *Tables of Integral Transforms*, tome II). Après quelques simplifications, il vient :

$$\Phi_z(u) = \frac{1}{2^{\frac{n-3}{2}} \Gamma\left(\frac{n-1}{2}\right)} \frac{(u'u)^{(n-1)/4}}{K_{\frac{n-1}{2}} [(u'u)^{1/2}]}$$

ou K est une fonction de Bessel de 3^e espèce.

b) Il est alors aisé d'obtenir la fonction caractéristique de Θ .

En effet, on a $\Theta = Z M C$, C désignant la matrice de la transformation de Θ en Y et M la matrice de la transformation de Y en Z .

La fonction caractéristique de Θ est :

$$\Phi_{\Theta}(V) = \int \int e^{i\vec{v} \cdot \Theta} f(\Theta) d\Theta = \int \int e^{i\vec{Z} M C V} f(Z) |M| dZ$$

ou

$$\Phi_Z(u) = \int \int e^{i\vec{Z} \cdot u} f(Z) |M| dZ$$

on obtient donc la fonction caractéristique de Θ en remplaçant u par $M C V$ dans la fonction caractéristique de Z , c'est-à-dire $u'u$ par $(M C V)' (M C V) = V' C' M' M C V$

Mais du fait que $ZZ' = \Theta F \Theta' = Z M C F C' M' Z'$ il résulte que $C' M' M C = F^{-1}$.

On doit donc remplacer $u'u$ par $V' F^{-1} V$ d'où :

$$\Phi_{\Theta}(V) = \frac{1}{2^{\frac{n-3}{2}}} \cdot \frac{(V' F^{-1} V)^{\frac{n-1}{4}}}{\Gamma\left(\frac{n-1}{2}\right)} K_{\frac{n-1}{2}} [(V' F^{-1} V)^{1/2}]$$

Pour expliciter $V' F^{-1} V$, il faut inverser la matrice F . On calcule d'abord les éléments diagonaux de F^{-1} , comme quotients des mineurs de $|F|$ à $|F|$ et l'on trouve : $n + 1$; puis tenant compte de ce que les autres éléments sont tous égaux de valeur λ , on doit avoir

$$|F^{-1}| = (n + 1 - \lambda)^{i-1} [n + 1 + (i - 1) \lambda]$$

mais, connaissant les valeurs caractéristiques de F

$$|F^{-1}| = n^{i-1} (n + i) \quad \text{d'où } \lambda = 1$$

et

$$V' F^{-1} V = \sum (n \vartheta_{jk} + 1) V_j V_k$$

1.3.4. Distribution de $S_{\Theta}^2, \bar{\Theta}, M_2(\Theta)$.

La forme quadratique $\Theta B \Theta'$ peut encore s'écrire

$$\Theta B \Theta' = i m_2 - \frac{i^2}{n + i} \bar{\Theta}^2 = i S^2 \Theta + \frac{i n}{n + i} \bar{\Theta}^2$$

Il résulte des considérations géométriques du § 1.3.2 que

$$Y_1 = \sqrt{i} \cdot \bar{\Theta} \quad \text{et} \quad \sum_{j=2}^i Y_j^2 = i S^2 \Theta$$

Le grand axe des ellipsoïdes de révolution porte la moyenne $\bar{\Theta}$. A $\bar{\Theta}$ constant, la distribution de l'échantillon est donc sphérique et ne dépend que de S_{Θ}^2 .

$$\text{Le volume élémentaire } d\Theta = \frac{1}{2} i^{1/2} (S_{\Theta}^2)^{\frac{i-3}{2}} dS_{\Theta}^2 d\bar{\Theta} d\Sigma_{i-1}$$

Intégrant par rapport à $d\Sigma_{i-1}$, on obtient :

$$\frac{\Gamma\left(\frac{n+i-1}{2}\right) i^{1/2}}{\pi^{1/2} \Gamma\left(\frac{n-1}{2}\right) \Gamma\left(\frac{i-1}{2}\right) n^{\frac{i-1}{2}} (n+i)^{1/2}} \left[1 + \frac{iS_{\Theta}^2}{n} + \frac{i\bar{\Theta}^2}{n+i}\right]^{-\frac{n+i-1}{2}} (S_{\Theta}^2)^{\frac{i-s}{2}} dS_{\Theta}^2 d\bar{\Theta}$$

ou

$$\frac{\Gamma\left(\frac{n+i-1}{2}\right) i^{1/2}}{\pi^{1/2} \Gamma\left(\frac{n-1}{2}\right) \Gamma\left(\frac{i-1}{2}\right) n^{\frac{i-1}{2}} (n+i)^{1/2}} \left[1 + \frac{im_2}{n} - \frac{i^2\bar{\Theta}^2}{n(n+i)}\right]^{-\frac{n+i-1}{2}} (m_2 - \Theta^2)^{\frac{i-s}{2}} dm_2 d\bar{\Theta}$$

I.3.4.1. Distribution marginale de $\bar{\Theta}$.

Pour $\bar{\Theta}$ donnée, S_{Θ}^2 varie de 0 à $+\infty$; on pose :

$$u = \frac{i}{n} S_{\Theta}^2 / 1 + \frac{i\bar{\Theta}^2}{n+i} + \frac{iS_{\Theta}^2}{n}$$

Intégrant u de 0 à 1, on obtient après simplifications :

$$\frac{\left(1 + \frac{i\bar{\Theta}^2}{n+i}\right)^{-\frac{n}{2}}}{B\left(\frac{1}{2}, \frac{n-1}{2}\right)} d\left(\sqrt{\frac{i}{n+i}} \bar{\Theta}\right)$$

Distribution de Student à $n-1$ degrés de liberté, résultat évident *a priori*.

(L'usuelle variable à utiliser pour les tables est $t = \sqrt{\frac{i(n-1)}{n+i}} \bar{\Theta}$)

I.3.4.2. Distribution de S_{Θ}^2 .

Pour S_{Θ}^2 donné, $\bar{\Theta}$ varie de $-\infty$ à $+\infty$.

Posant

$$u^{-1} = 1 + \frac{i}{n+i} \frac{\bar{\Theta}^2}{a^2} \quad \text{et} \quad a^2 = 1 + \frac{iS_{\Theta}^2}{n}$$

et intégrant par rapport à u de 0 à 1, on obtient après groupement de termes :

$$\frac{1}{B\left(\frac{n-1}{2}, \frac{i-1}{2}\right)} \left(\frac{iS_{\Theta}^2}{n}\right)^{\frac{i-s}{2}} \left[1 + \frac{iS_{\Theta}^2}{n}\right]^{-\frac{n+i-2}{2}} d\left(\frac{iS_{\Theta}^2}{n}\right)$$

Il en résulte que

$$\frac{i(n-1)}{n(i-1)} S_{\Theta}^2 = F(i-1, n-1)$$

a une distribution de Snedecor avec $i - 1$ et $n - 1$ degrés de liberté, résultat évident *a priori*.

I.3.4.3. Distributions conditionnelles de $\bar{\Theta}$ et S_{Θ}^2 .

Il résulte de I.3.4 et I.3.4.2 et I.3.4.1. que

pour S_{Θ} donnée, $\bar{\Theta}$ a une distribution de Student à $n + i - 2$ degrés de liberté

$$\left(\text{variable à utiliser pour les tables } t = \sqrt{\frac{i(n+i-2)}{n+i}} \frac{\bar{\Theta}}{\sqrt{1 + \frac{i S_{\Theta}^2}{n}}} \right)$$

pour $\bar{\Theta}$ donnée, S_{Θ}^2 a une distribution de Snedecor :

$$F(i-1, n) = \frac{i}{i-1} S_{\Theta}^2 / 1 + \frac{i \bar{\Theta}^2}{n+i}$$

I.3.4.4. Distribution du moment d'ordre deux : m_2 (Θ).

On doit intégrer la distribution I.3.4 par rapport à $\bar{\Theta}$ dans l'intervalle :

$$-\sqrt{m_2}, +\sqrt{m_2}$$

Posant $u = \frac{i^2 m_2}{n(n+i)} / 1 + \frac{i m_2}{n}$ et $t = \frac{\bar{\Theta}^2}{m_2}$, on a

$$K(m_2)^{\frac{i-2}{2}} \left(1 + \frac{i m_2}{n}\right)^{-\frac{n+i-1}{2}} \int_0^1 [1-tu]^{-\frac{n+i-1}{2}} (1-t)^{\frac{i-2}{2}} t^{-\frac{1}{2}} dt$$

L'intégrale est bien connue, c'est la fonction hypergéométrique ${}_2F_1\left(\frac{n+i-1}{2}, \frac{1}{2}; \frac{i}{2}; u\right)$, valable si $u < 1$, donc toujours valable ici puisque u varie de 0 à $\frac{i}{n+i}$.

La distribution est donc :

$$\sqrt{\frac{n}{n+i}} \frac{\left(\frac{i m_2}{n}\right)^{\frac{i-2}{2}} \left[1 + \frac{i m_2}{n}\right]^{-\frac{n+i-1}{2}}}{B\left(\frac{i}{2}, \frac{n-1}{2}\right)} {}_2F_1\left(\frac{n+i-1}{2}, \frac{1}{2}; \frac{i}{2}; \frac{i}{n+i} \frac{i m_2/n}{1 + i m_2/n}\right) d\left(\frac{i m_2}{n}\right)$$

II. — ÉCHANTILLON D'EFFECTIF i

EXTRAIT D'UNE POPULATION FINIE D'EFFECTIF N

Soit $\xi = (\xi_1, \dots, \xi_i)$ l'échantillon considéré; la partie restante de la population constitue un second échantillon $X' = (X_1, \dots, X_n)$ et $n = N - i$.

On se propose de généraliser au cas d'une population finie les résultats classiques obtenus pour une population infinie.

Soient

$\bar{\xi}$ et S_{ξ}^2 la moyenne et la variance de ξ
 \bar{X}_n et S_n^2 celles de X'
 \bar{X} et S^2 celles de la population finie.

La plupart des résultats obtenus peuvent l'être à partir des résultats de la partie I au moyen de changements appropriés de variables, on peut aussi les obtenir directement.

Les changements de variables intéressants sont les suivants (quoique fort élémentaires, ils se révèlent précieux dans un grand nombre de problèmes) :

$$\Theta = \frac{\xi - \bar{X}_n}{S_n} \quad t = \frac{\xi - \bar{X}}{S} \quad \Theta' = \frac{X' - \bar{X}_n}{S_n} \quad t' = \frac{X' - \bar{X}}{S}$$

On a, entre autres relations :

$$\bar{X} - \bar{X}_n = \frac{i \bar{\Theta}}{n+i} S_n = \frac{i \bar{t} S}{n}$$

$$t'_n = \Theta' \frac{S_n}{S} + \frac{\bar{X}_n - \bar{X}}{S}$$

$$X' - \bar{X}_n = X' - \bar{X} + \frac{i}{n} \bar{t} S$$

$$\xi - \bar{X}_n = \Theta S_n = \left(t + \frac{i}{n} \bar{t} \right) S$$

$$X' - \bar{X} = X' - \bar{X}_n - \frac{i \bar{\Theta}}{n+i} S_n$$

$$\xi - \bar{X} = \left(\Theta - \frac{i \bar{\Theta}}{n+i} \right) S_n = t S$$

$$(n+i) S^2 = n S_n^2 - \frac{i^2 \bar{\Theta}^2}{n+i} S_n^2 + \sum \Theta_j^2 S_n^2$$

$$n S_n^2 = (n+i) S^2 - \frac{i^2 \bar{t}^2}{n} S^2 - \sum t_j^2 S^2$$

$$\Theta' = \frac{S}{S_n} \left(t' + \frac{i \bar{t}}{n} \right)$$

$$\Theta = \frac{S}{S_n} \left(t + \frac{i \bar{t}}{n} \right) = \left(t + \frac{i \bar{t}}{n} \right) / \sqrt{\frac{n+i}{n} - \frac{i^2}{n^2} \bar{t}^2 - \frac{\sum t_j^2}{n}}$$

$$t' = \frac{S_n}{S} \left(\Theta' - \frac{i \bar{\Theta}}{n+i} \right)$$

$$= \frac{S_n}{S} \left(\Theta - \frac{i \bar{\Theta}}{n+i} \right) = \left(\Theta - \frac{i \bar{\Theta}}{n+i} \right) / \sqrt{\frac{n}{n+i} - \frac{i^2}{(n+i)^2} \bar{\Theta}^2 + \frac{\sum \Theta_j^2}{n+i}}$$

II.1. *m* INCONNU, \bar{X} et σ CONNUS

On se propose d'étudier $\xi'' = \xi - \bar{X} = (\xi_1 - \bar{X}, \dots, \xi_i - \bar{X})$

II.1.1. *Distribution de ξ''*

Irwin a montré que ξ'' était distribué de façon Laplace-Gaussienne.

$$\frac{1}{(2\pi)^{i/2} \sigma^i} \sqrt{\frac{N}{N-i}} e^{-\frac{Q}{2\sigma^2}} d(\xi - \bar{X})$$

ou

$$Q = (\xi - \bar{X}) A (\xi - \bar{X})' \text{ et } A = \left(\lambda_{jk} + \frac{1}{N-i} \right)$$

La matrice des covariances est alors $\sigma^2 A^{-1}$ et l'on en déduit

$$\sigma_{ij}^2 = \frac{N-1}{N} \sigma^2 \quad j = 1 \text{ à } i$$

et les coefficients de corrélation des variables deux à deux égalent

$$\rho = -\frac{1}{N-1}$$

la matrice des covariances est donc $\sigma^2 \left(\lambda_{jk} - \frac{1}{N} \right)$.

Ces résultats peuvent être obtenus de façon élémentaire, on peut aussi utiliser les résultats de I.1.1. et effectuer le changement de variables

$$\xi' = \xi'' + \frac{i}{n} \bar{\xi}''$$

en tenant compte que

$$\frac{D(\xi')}{D(\xi'')} = \left| \lambda_{jk} + \frac{1}{n} \right| = \frac{N}{N-i}$$

II.1.2. *Distribution de ξ'' .*

Elle s'obtient de la même façon qu'en I.1.2.

$\bar{\xi}''$ est Laplace-Gaussienne de variance $\sigma_{\bar{\xi}''}^2 = \frac{\sigma^2}{i} \frac{N-i}{N}$

II.1.3. *Distribution du moment d'ordre deux : $m_2(\xi'')$.*

La fonction caractéristique est obtenue comme en I.1.3. Les valeurs caractéristiques de la matrice de corrélation étant

$$\frac{N}{N-1}, i-1 \text{ fois et } \frac{N-i}{N-1}, \text{ on a}$$

$$\Phi_{m_2}(t) = \left[1 - 2 \frac{\sigma^2}{i} t \right]^{-\frac{i-1}{2}} \left[1 - 2 \frac{\sigma^2}{i} \frac{N-i}{N} t \right]^{-\frac{1}{2}}$$

Il en résulte que la distribution est

$$\left(\frac{N}{N-i} \right)^{1/2} \left(\frac{i m_2}{2 \sigma^2} \right)^{\frac{i-2}{2}} \frac{e^{-\frac{i m_2}{2 \sigma^2}}}{\Gamma\left(\frac{i}{2}\right)} {}_1F_1 \left(\frac{1}{2}, \frac{i}{2}, \frac{i}{N-i}, \frac{i m_2}{2 \sigma^2} \right) d \left(\frac{i m_2}{2 \sigma^2} \right)$$

II.2. σ INCONNU, m et S^2 CONNUS

On se propose d'étudier l'échantillon $\bar{\zeta}'' = \frac{\xi - m}{S} = \left(\frac{\xi_1 - m}{S}, \dots, \frac{\xi_i - m}{S} \right)$

II.2.1. Distribution de ζ'' .

La distribution est obtenue à partir de I.2.1 en effectuant le changement de variables :

$$\zeta' = \zeta'' / \sqrt{\frac{N}{N-i} - \frac{\sum \xi_j^2}{N-i}}$$

Le Jacobien de la transformation est

$$\frac{D(\zeta')}{D(\zeta'')} = \frac{N}{N-i} \left(\frac{N}{N-i} - \frac{\sum \xi_j^2}{N-i} \right)^{-\frac{1}{2} - 1}$$

la distribution est donc

$$\frac{\Gamma\left(\frac{N-1}{2}\right)}{\pi^{1/2} \Gamma\left(\frac{N-i-1}{2}\right) N^{i/2}} \left[1 - \frac{\sum \zeta_j''^2}{N} \right]^{\frac{N-i-3}{2}} d \zeta''$$

II.2.2. Distribution de la moyenne $\bar{\zeta}'' = \frac{\bar{\xi} - m}{S}$.

Ce cas se ramène à celui d'une seule observation $\sqrt{i} \bar{\zeta}''$ au lieu de ζ'' d'où la distribution

$$\frac{1}{B\left(\frac{1}{2}, \frac{N-2}{2}\right)} \left[1 - \frac{i \bar{\zeta}''^2}{N} \right]^{\frac{N-4}{2}} d \sqrt{\frac{i}{N}} \bar{\zeta}''$$

c'est-à-dire que $\frac{i \bar{\zeta}''^2}{N}$ suit une *distribution Bêta incomplète* de paramètres $\frac{1}{2}$ et $\frac{N-2}{2}$.

II.2.3. Distribution du moment d'ordre deux : $m_2(\zeta'')$.

La distribution de ζ'' est sphérique, il en résulte que la distribution de

$i m_2 = (\zeta'') (\zeta'')'$ est obtenue en multipliant la densité de probabilité de ζ'' par $(i m_2)^{\frac{i-2}{2}} \frac{\pi^{i/2}}{\Gamma(\frac{i}{2})}$ soit :

$$\frac{(i m_2/N)^{\frac{i-2}{2}}}{B\left(\frac{i}{2}, \frac{N-i-1}{2}\right)} \left[1 - \frac{i m_2}{N}\right]^{\frac{N-i-3}{2}} d\left(\frac{i m_2}{N}\right)$$

c'est-à-dire une distribution Bêta incomplète de paramètres $\frac{N+i-1}{2}$ et $\frac{i}{2}$.

II.2.4. Fonction caractéristique de ζ'' .

La distribution de $\frac{\zeta''}{\sqrt{N}}$ étant sphérique, la fonction caractéristique est donnée par :

$$\frac{\Phi_{\zeta''}}{\sqrt{N}}(u) = (2\pi)^{i/2} (u'u)^{-\frac{i-1}{4}} \mathcal{H}_{\frac{i-1}{2}-1} \left[\frac{\Gamma\left(\frac{N-1}{2}\right)}{\pi^{i/2} \Gamma\left(\frac{N-i-1}{2}\right)} \left[1 - \frac{(\xi'')(\xi'')'}{N}\right]^{\frac{N-i-3}{2}} \left[\frac{(\xi'')(\xi'')'}{N}\right]^{\frac{i-1}{4}}; (u'$$

on trouve cette transformée dans une table de transformées de Hankel (par exemple : Erdelyi, *op. cit.*, p. 26).

$$\frac{\Phi_{\zeta''}}{\sqrt{N}}(u) = \Gamma\left(\frac{N-1}{2}\right) \left(\frac{u'u}{4}\right)^{-\frac{N-3}{4}} J_{\frac{N-3}{2}}[(u'u)^{1/2}]$$

II.3. m ET σ INCONNUS SONT ESTIMÉS PAR X ET S .

On se propose d'étudier l'échantillon

$$t = \frac{-\bar{X}}{S} = \left(\frac{\xi_1 - \bar{X}}{S}, \dots, \frac{\xi_i - \bar{X}}{S}\right)'$$

II.3.1. Distribution de t

On obtient la distribution à partir de I.3.1. au moyen des changements de variables appropriés.

On a vu que :

$$\Theta = \left(t + \frac{i}{N-i} \bar{t}\right) / \sqrt{\frac{N}{N-i} - \frac{i^2 \bar{t}^2}{(N-i)^2} - \frac{t t'}{N-i}}$$

a) Le likelihood de Θ se transforme en

$$K \left[1 - \frac{N-i}{N} \left(t + \frac{i \bar{t}}{N-i}\right) \left(\bar{t} - \frac{1}{N}\right) \left(1 + \frac{i \bar{t}}{N-i}\right)'\right]^{\frac{N-1}{2}}$$

Soit

$$K \left[1 - \frac{1}{N} \left(\frac{i^2 \bar{t}^2}{N-i} + t t' \right) \right]^{\frac{N-1}{2}}$$

ou

$$K \left[1 - \frac{1}{N} t \left(\alpha_{,k} + \frac{1}{N-i} \right) t' \right]^{\frac{N-1}{2}}$$

avec

$$K = \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-i-1}{2}\right) \pi^{i/2}} \frac{1}{(N-i)^{\frac{i-1}{2}} N^{1/2}}$$

b) Le Jacobien $\frac{D(\Theta)}{D(t)}$ de la transformation s'obtient de la façon suivante :

Posons

$$l = \frac{1}{N-i} \left(N - \frac{i^2 \bar{t}^2}{N-i} - t t' \right)$$

$$\varphi = t + \frac{i \bar{t}}{N-i}$$

on trouve

$$\frac{D(\Theta)}{D(t)} = \frac{l^{-\frac{3i}{2}}}{(N-i)^i} \left| [(N-i) \alpha_{,k} + 1] l + \varphi_j \varphi_k \right|_{i \times i}$$

le déterminant à calculer est le déterminant de la matrice :

$$(N-i) l \left(I + \frac{1}{N-i} u' u + \frac{1}{(N-i) l} \varphi' \varphi \right)$$

où I est la matrice unité et $u = (1 \ 1 \ \dots \ 1 \ 1)$.

Il résulte de l'identité :

$$\left| I + X_1 Y_1 + X_2 Y_2 \right| = \begin{vmatrix} 1 + Y_1 X_1' & Y_2 X_1' \\ Y_1 X_2' & 1 + Y_2 X_2' \end{vmatrix}$$

que ce déterminant a pour développement :

$$(N-i)^i l \left[\left(1 + \frac{\varphi' \varphi}{(N-i) l} \right) \left(1 + \frac{i}{N-i} \right) - \frac{i^2 \bar{t}^2}{(N-i)^2 l} \right] = (N-i)^{i-2} l^{i-1} N^2$$

d'où, après simplifications :

$$\frac{D(\Theta)}{D(t)} = \left(\frac{N-i}{N} \right)^{\frac{i-2}{2}} \left[1 - \frac{1}{N} \left(\frac{i^2 \bar{t}^2}{N-i} + t t' \right) \right]^{-\frac{i+2}{2}}$$

c) La distribution cherchée est donc :

$$\frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-i-1}{2}\right) \pi^{i/2}} \frac{1}{N^{\frac{i-1}{2}} (N-i)^{1/2}} \left[1 - \frac{1}{N} \left(\frac{i^2 \bar{t}^2}{N-i} + t t' \right) \right]^{\frac{N-i-3}{2}} dt$$

ou

$$\frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-i-1}{2}\right) \pi^{i/2} N^{\frac{i-1}{2}} (N-i)^{1/2}} [1 - t D t']^{\frac{N-i-3}{2}} dt$$

avec $D = \left(\frac{\partial^2 t}{\partial N} + \frac{1}{N(N-i)} \right)$

Les valeurs caractéristiques de la matrice D sont $\frac{1}{N}$, $i - 1$ fois et $\frac{1}{N-i}$.

Le champ de variation de t est celui qui rend le crochet positif, la distribution est intérieure à l'ellipsoïde $t D t' = 1$ et les surfaces de densité constantes sont les ellipsoïdes $t D t' = \frac{1}{C^2}$.

La distribution généralise à plusieurs variables la distribution Bêta incomplète.

d) Un cas particulier est celui où $i = 1$ et où l'on obtient :

$$\frac{1}{B\left(\frac{1}{2}, \frac{N}{2} - 1\right)} \left[1 - \left(\frac{\xi - \bar{X}}{S} \right)^2 / (N-1) \right]^{\frac{N-4}{2}} d \left(\frac{\xi - \bar{X}}{S} / \sqrt{N-1} \right) \quad |t| < \sqrt{N-1}$$

c'est-à-dire que $\left(\frac{\xi - \bar{X}}{S} \right)^2 / (N-1)$ suit une distribution Bêta incomplète de paramètre $\frac{1}{2}$ et $\frac{N}{2} - 1$. C'est le résultat trouvé par Thompson. Dans le cas plus particulier où $N = 4$, on obtient une distribution rectangulaire.

e) Dans le cas particulier où $i = N - 3$, la densité est constante dans l'ellipsoïde.

II.3.2. Distribution de la forme quadratique $t D t'$

Suivant une marche analogue à celle utilisée en I.3.2. on effectue une transformation orthogonale de t en Y diagonalisant D , puis le changement de variables

$$Z_1 = \frac{Y_1}{\sqrt{N-i}} \quad Z_j = \frac{Y_j}{\sqrt{N}} \quad j = 2 \text{ à } i$$

on obtient pour Z la distribution sphérique :

$$\frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-i-1}{2}\right) \pi^{i/2}} (1 - ZZ')^{\frac{N-i-3}{2}} dZ \quad \text{avec } ZZ' = t D t'$$

après intégration de l'hyperangle $d\Sigma_i$ on obtient pour ZZ' la distribution :

$$\frac{1}{B\left(\frac{i}{2}, \frac{N-i-1}{2}\right)} (1-ZZ')^{\frac{N-i-3}{2}} (ZZ')^{\frac{i-2}{2}} dZZ'$$

qui est une *distribution Bêta incomplète* de paramètres $\frac{i}{2}, \frac{N-i-1}{2}$.

II.3.3. Fonction caractéristique de t .

La distribution de Z étant sphérique, sa fonction caractéristique est donnée par

$$\Phi_Z(u) = (2\pi)^{1/2} (u'u)^{-\frac{i-1}{4}} \mathcal{J}C_{\frac{i}{2}-1} \left\{ \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-i-1}{2}\right) \pi^{i/2}} (1-ZZ')^{\frac{N-i-3}{2}} (ZZ')^{\frac{i-1}{4}} \right\}$$

on trouve la transformée de Hankel dans une table de transformée (Erdelyi, *op. cit.*, p. 26).

$$\Phi_Z(u) = {}_2F_1\left(\frac{i}{2}, \frac{i}{2}; \frac{N-1}{2}; -\frac{u^2}{4}\right) = \Gamma\left(\frac{N-1}{2}\right) \left(\frac{u'u}{4}\right)^{-\frac{N-3}{4}} \mathcal{J}_{\frac{N-3}{2}} [(u'u)^{1/2}]$$

où $\mathcal{J}_{\frac{N-3}{2}}$ est une fonction de Bessel de première espèce.

D'après les remarques formulées au § I.3.3, la fonction caractéristique de t est donnée par

$$\Phi_t(V) = \Gamma\left(\frac{N-1}{2}\right) \left[\frac{V' D^{-1} V}{4}\right]^{-\frac{N-3}{4}} \mathcal{J}_{\frac{N-3}{2}} [(V' D^{-1} V)^{1/2}]$$

on explicite $V' D^{-1} V$ en inversant la matrice D , on trouve :

$$D^{-1} = (N \lambda_{jk} - 1) \quad \text{d'où} \quad V' D^{-1} V = \sum (N \lambda_{jk} - 1) V_j V_k$$

II.3.4. Distribution de \bar{t} , S_i^2 , $m_2(t)$.

On voit aisément (mêmes remarques qu'en I.3.4) que t est porté par l'axe de révolution des ellipsoïdes de densité de probabilité constante et que pour \bar{t} donnée, la distribution est sphérique; en intégrant $d\Sigma_{i-1}$ dans l'élément de volume $dt = \frac{1}{2} i^{1/2} (S_i^2)^{\frac{i-3}{2}} dS_i^2 dt d\Sigma_{i-1}$, on obtient à partir de la formule I.3.1.C.

$$\frac{i^{1/2} \Gamma\left(\frac{N-1}{2}\right)}{N^{\frac{i-1}{2}} (N-i)^{1/2} \pi^{1/2} \Gamma\left(\frac{N-i-1}{2}\right) \Gamma\left(\frac{i-1}{2}\right)} \left[1 - \frac{i}{N} S_i^2 - \frac{i}{N-i} \bar{t}^2\right]^{\frac{N-i-3}{2}} (S_i^2)^{\frac{i-3}{2}} dS_i^2 d\bar{t}$$

et

$$\frac{i^{1/2} \Gamma\left(\frac{N-1}{2}\right)}{N^{\frac{t-1}{2}} (N-i)^{1/2} \pi^{1/2} \Gamma\left(\frac{N-i-1}{2}\right) \Gamma\left(\frac{i-1}{2}\right)} \left[1 - \frac{i^2}{N} \frac{t^2}{N-i} - \frac{i m_2}{N}\right]^{\frac{N-t-3}{2}} (m_2 - \bar{t}^2)^{\frac{t-3}{2}} d m_2 d \bar{t}$$

II.3.4.1 Distribution de t .

Pour une valeur donnée de t , S_t^2 varie de 0 à

$$\frac{N}{i} \left(1 - \frac{i}{N-i} \bar{t}^2\right)$$

Effectuant le changement de variable

$$u = S / \sqrt{\frac{N}{i} \left(1 - \frac{i}{N-i} \bar{t}^2\right)},$$

et intégrant de 0 à 1, on obtient :

$$B\left(\frac{1}{2}, \frac{N-2}{2}\right) \left(1 - \frac{i}{N-i} \bar{t}^2\right)^{\frac{N-4}{2}} d \sqrt{\frac{i}{N-i} \bar{t}^2}$$

c'est-à-dire que $\frac{i}{N-i} \bar{t}^2$ est distribuée suivant une *Bêta incomplète de paramètre* $\frac{1}{2}$, et $\frac{N-2}{2}$.

La fonction caractéristique de :

$$\frac{i}{N-i} \bar{t}^2 \text{ est } {}_1F_1\left(\frac{1}{2}; \frac{N-1}{2}; i u\right).$$

$$E[\bar{t}] = 0 \quad E[\bar{t}^2] = \sigma_t^2 = \frac{N-i}{i} \cdot \frac{1}{N-1}$$

II.3.4.2. Distribution de S_t^2 .

Pour S_t^2 donné, \bar{t}^2 varie de 0 à $\frac{N-i}{i} \left(1 - \frac{i}{N} S_t^2\right)$

Posant :

$$\nu = t / \sqrt{\frac{N-i}{i} \left(1 - \frac{i}{N} S_t^2\right)}$$

et intégrant ν de -1 à $+1$ on obtient :

$$B\left(\frac{i-1}{2}, \frac{N-i}{2}\right) \left(1 - \frac{i}{N} S_t^2\right)^{\frac{N-t-2}{2}} \left(\frac{i}{N} S_t^2\right)^{\frac{t-3}{2}} d\left(\frac{i}{N} S_t^2\right)$$

$\frac{i}{N} S_t^2$ a une distribution *Bêta incomplète de paramètres* $\frac{i-1}{2}$ et $\frac{N-i}{2}$.

La fonction caractéristique de :

$$\frac{i}{N} S_i^2 \text{ est } {}_1F_1 \left(\frac{i-1}{2}; \frac{N-1}{2}; \vec{t} u \right) \quad E [S_i^2] = \frac{N (i-1)}{i (N-1)}$$

II.3.4.3. Distributions conditionnelles de \bar{t} et S_i^2 .

A partir de II.3.4, II.3.4.1, II.3.4.2, on trouve aisément que :

1) pour S_i donné, $\bar{t}^2 / \frac{N-i}{i} \left(1 - \frac{i}{N} S_i^2 \right)$ a une distribution de B incomplète de paramètres $\frac{1}{2}$ et $\frac{N-i-1}{2}$.

2) pour \bar{t} donnée, $S_i^2 / \frac{N}{i} \left(1 - \frac{i \bar{t}^2}{N-i} \right)$ a une distribution Bêta incomplète de paramètres $\frac{i-1}{2}, \frac{N-i-1}{2}$.

II.3.4.4. Distribution du moment d'ordre deux : $m_2(t)$.

a) L'intérêt de la variable $i m_2(t)$ est qu'elle représente le carré de la distance des observations ξ à la moyenne \bar{X} de la population mesurée en écart-types S de la population, elle est au problème actuel ce que le χ^2 est au problème classique d'un échantillon extrait d'une population infinie.

b) De nombreuses techniques peuvent être employées pour obtenir la distribution de $i m_2 = R^2$.

II.3.4.4.1. On peut employer la méthode du *facteur discontinu*, en écrivant que la probabilité intégrale de R^2 , soit $P \{ R^2 \leq r^2 \}$ est l'espérance mathématique de la fonction caractéristique de l'ensemble des points de l'espace échantillon pour lesquels $R \leq r$.

$$P \{ R^2 \leq r^2 \} = E [U (r - R)]$$

ou
$$U (r - R) = \begin{cases} 1 & \text{si } R \leq r \\ 0 & \text{si } R > r \end{cases}$$

soit
$$U (r - R) = \left(\frac{r}{2\pi} \right)^{i/2} \int \int e^{\vec{t} \cdot \mathbf{v}} \frac{J^{i/2} \{ r (V' V)^{1/2} \}}{(V' V)^{i/4}} dV$$

d'où
$$P = \left(\frac{r}{2\pi} \right)^{i/2} \int \int \frac{J^{i/2} \{ r (V' V)^{1/2} \}}{(V' V)^{i/4}} \Phi_t (V) dV$$

où $\Phi_t (V)$ est la fonction caractéristique de t , c'est-à-dire :

$$J_{\frac{N-3}{2}} [(V' D^{-1} V)^{1/2}] \left[\frac{V' D^{-1} V}{4} \right]^{-\frac{N-3}{4}}$$

c'est une méthode assez pénible. On développe en série la fonction de Bessel et on a à calculer les moments de $V' V$ par rapport à la « loi » $\Phi_t (V) dV$.

II.3.4.4.2. La méthode la plus *directe* consiste à intégrer la distribution II.3.4. par rapport à \bar{t} .

Posant

$$\tau^2 = \frac{i^2 \bar{t}^2}{N(N-i)} \Big/ 1 - \frac{i m_2}{N} \quad \text{et } W^2 = \frac{N(N-i)}{i^2 m_2} \left(1 - \frac{i m_2}{N}\right),$$

la distribution de m_2 et t devient :

$$K \cdot \frac{(N-i)^{1/2} N^{i/2}}{i^{\frac{i+1}{2}}} \left(\frac{i m_2}{N}\right)^{\frac{i-3}{2}} \left(1 - \frac{i m_2}{N}\right)^{\frac{N-i-2}{2}} (1 - \tau^2)^{\frac{N-i-3}{2}} (1 - W^2 \tau^2)^{\frac{i-3}{2}} d\tau \frac{d i m_2}{N}$$

Si $i m_2 < N - i$ $i \bar{t}^2$ varie de 0 à $i m_2$, τ^2 de 0 à $\frac{1}{W^2}$ et $W^2 > 1$

Si $i m_2 > N - i$ $i \bar{t}^2$ varie de 0 à $\left(1 - \frac{i m_2}{N}\right)N(N-i)$, τ^2 de 0 à 1 et $W^2 < 1$.

Après arrangements de termes on obtient :

$$\sqrt{\frac{N}{i}} \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-i-1}{2}\right)\Gamma\left(\frac{i-1}{2}\right)} \pi^{1/2} \left(\frac{i m_2}{N}\right)^{\frac{i-3}{2}} \left(1 - \frac{i m_2}{N}\right)^{\frac{N-i-2}{2}} h(\tau) d\tau d \frac{i m_2}{N}$$

avec

$$h(\tau) = (1 - \tau^2)^{\frac{N-i-3}{2}} (1 - W^2 \tau^2)^{\frac{i-3}{2}}$$

Pour obtenir la distribution de m_2 ,

a) Si $i m_2 < N - i$, on intègre $h(\tau)$, τ variant de $-\frac{1}{W}$ à $\frac{1}{W}$.

Posant : $u = \frac{1}{\tau^2}$ il vient

$$\int_{-\frac{1}{W}}^{\frac{1}{W}} h(\tau) d\tau = \int_{W^2}^{\infty} u^{\frac{N-3}{2}} (u-1)^{\frac{N-i-3}{2}} (u-W^2)^{\frac{i-3}{2}} du$$

On reconnaît une intégrale de Weyl, dont on trouve le type dans Erdelyi (op. cit.). Après simplification, on obtient :

$$B\left(\frac{i-1}{2}, \frac{1}{2}\right) \frac{1}{W} {}_2F_1\left(-\frac{N-i-3}{2}, \frac{1}{2}; \frac{1}{2}; \frac{1}{W^2}\right)$$

d'où la distribution de m_2 :

$$B\left(\frac{N-i-1}{2}, \frac{i}{2}\right) \frac{\sqrt{\frac{N}{N-i}}}{\left(\frac{i m_2}{N}\right)^{\frac{i-2}{2}} \left(1 - \frac{i m_2}{N}\right)^{\frac{N-i-3}{2}}} {}_2F_1\left(-\frac{N-i-3}{2}, \frac{1}{2}; \frac{i}{2}; \frac{i m_2}{N} / 1 - \frac{i m_2}{N}\right) d\left(\frac{i m_2}{N}\right)$$

b) Si $i m_2 > N - i$

On intègre $h(\tau)$, τ variant de -1 à $+1$,

Posant $V = \frac{1}{\tau^2 W^2}$

$$\int_{-1}^{+1} h(\tau) d\tau = \frac{1}{W} \int_{\frac{1}{W^2}}^{W^2} V^{-\frac{N-2}{2}} \left(V - \frac{1}{W^2}\right)^{\frac{N-i-3}{2}} (V-1)^{\frac{i-3}{2}} dV$$

On reconnaît une intégrale de Weyl dont on trouvera le type dans Erdelyi (op. cit.). Après simplification, on trouve :

$$B \left(\frac{N-i-1}{2}, \frac{1}{2} \right) {}_2F_1 \left(-\frac{i-3}{2}, \frac{1}{2}; \frac{N-i}{2}; W^2 \right)$$

d'où la distribution de m_2 :

$$\frac{\sqrt{\frac{N}{i}}}{\frac{i-1}{2}, \frac{N-i}{2}} \left(\frac{im_2}{N} \right)^{\frac{i-3}{2}} \left(1 - \frac{im_2}{N} \right)^{\frac{N-i-3}{2}} {}_2F_1 \left[-\frac{i-3}{2}, \frac{1}{2}; \frac{N-i}{2}; \frac{N-i}{i} \left(1 - \frac{im_2}{N} \right) / \frac{im_2}{N} \right] d \left(\frac{im_2}{N} \right)$$

e) Si l'on pose $im_2 = R^2$, on voit que de a) en b) la densité de probabilité demeure invariante si l'on change simultanément i en $N-i$ et $\frac{R^2}{N}$ en $1 - \frac{R^2}{N}$.

Le résultat est évident a priori, car si m_2 est le moment de t et m'_2 le moment de la partie restante t' de la population, et si $im_2 = R^2$ et $(N-i)m'_2 = R'^2$ on a

$$\frac{R^2}{N} = 1 - \frac{R'^2}{N}$$

et l'effectif i' du reste de la population est $i' = N - i$.

Si $R^2 < N - i$:

$$P \left\{ \frac{R^2}{N} > \lambda \right\} = f_1(i, N-i, \lambda) \text{ (formule a)}$$

$$\text{mais } P \left\{ \frac{R^2}{N} > \lambda \right\} = 1 - P \left\{ \frac{R'^2}{N} > 1 - \lambda \right\}$$

comme

$$R^2 > N - i \quad P \left\{ \frac{R'^2}{N} > 1 - \lambda \right\} = f_2(i', N - i', 1 - \lambda) \text{ (formule b)}$$

d'où :

$$f_1(i, N-i, \lambda) = 1 - f_2(N-i, i, 1-\lambda)$$

a) L'obtention de la probabilité intégrale de m_2 , n'offre pas de difficultés essentielles, du moins dans certains cas.

Si $im_2 > N - i$ et si i est impair,

le développement en série de ${}_2F_1$ est limité à :

$$\sum_{K=1}^{(i-3)/2} \frac{(-1)^K}{K!} \frac{\left[\frac{i-1}{2} - 1 \right] \left[\frac{i-1}{2} - 2 \right] \dots \left[\frac{i-1}{2} - K \right] \left(\frac{1}{2} \right) \left(\frac{3}{2} \right) \dots \left(\frac{1}{2} + K - 1 \right)}{\left(\frac{N-i}{2} \right) \left(\frac{N-i}{2} + 1 \right) \dots \left(\frac{N-i}{2} + K - 1 \right)} \left(\frac{N-i}{i} \right)^K \left(1 - \frac{im_2}{N} \right)^K \left(\frac{im_2}{N} \right)^{-K}$$

d'où après réarrangement et groupement de termes, la loi de m_2 :

$$\sqrt{\frac{N}{i}} \sum_{k=0}^{(t-3)/2} (-1)^k \left(\frac{im_2}{N}\right)^{\frac{t-1}{2}-k-1} \left(1 - \frac{im_2}{N}\right)^{\frac{N-i}{2}+k-1} \frac{B\left(K + \frac{1}{2}, \frac{1}{2}\right) \left(\frac{N-i}{i}\right)^K}{\pi B\left(\frac{i-1}{2} - K, \frac{N-i}{2} + K\right)} d\left(\frac{im_2}{N}\right)$$

d'où

$$P\left\{\frac{im_2}{N} > x\right\} = \frac{1}{\pi} \sqrt{\frac{N}{i}} \sum_{k=0}^{(t-3)/2} (-1)^k \left(\frac{N-i}{i}\right)^K B\left(\frac{1}{2}, K + \frac{1}{2}\right) \left[1 - I_x\left(\frac{i-1}{2} - K, \frac{N-i}{2} + K\right)\right]$$

valable pour $x > (N-i)/N$, en intégrant la formule précédente entre x et 1.

$I_x\left(\frac{i-1}{2} - K, \frac{N-i}{2} + K\right)$ désigne l'intégrale de la loi de distribution

Bêta incomplète, de paramètres $\frac{i-1}{2} - K$ et $\frac{N-i}{2} + K$.

Si $im_2 < N - i$ et Si $N - i$ est impair, les mêmes méthodes conduisent à :

$$P\left\{\frac{im_2}{N} \leq x\right\} = \frac{1}{\pi} \sqrt{\frac{N}{N-i}} \sum_{k=0}^{(N-i-3)/2} (-1)^k \left(\frac{i}{N-i}\right)^K B\left(\frac{1}{2}, K + \frac{1}{2}\right) I_x\left(\frac{i}{2} + K, \frac{N-i-1}{2} - K\right)$$

valable pour $x < (N-i)/N$

Si i est impair et N pair, la tabulation de la probabilité intégrale de m_2 n'implique que des développements en séries limitées; elle est aisée quand on dispose d'une table de la fonction Bêta incomplète...

Si ces deux conditions ne sont pas remplies ensemble, on doit intégrer une ou deux séries illimitées; on peut montrer que les séries intégrées sont convergentes et que l'intégration terme à terme est valable, mais les séries se prêtent moins aisément à la tabulation.

Si, par exemple, i est pair, et $x > \frac{N-i}{N}$.

$$P\left\{\frac{im_2}{N} > x\right\} = P_1 + R_1$$

avec

$$P_1 = \frac{1}{\pi} \sqrt{\frac{N}{i}} \sum_{k=0}^{(t-3)/2} (-1)^k \left(\frac{N-i}{i}\right)^K B\left(\frac{1}{2}, K + \frac{1}{2}\right) \left[1 - I_x\left(\frac{i-1}{2} - K, \frac{N-i}{2} + K\right)\right]$$

tandis que R_1 implique le calcul d'intégrales du type $\int_x^1 \frac{(1-X)^\alpha}{X^\beta} dX$ avec $\beta > 1$ $\alpha > 1$. Cette question fera l'objet d'une prochaine note.

II.3.4.4.3. On peut encore tenter d'obtenir directement la probabilité intégrale de $im_2 = tt'$, en intégrant la loi de probabilité de t dans le domaine : $t' \leq r^2$ et $t D t' \leq 1$

$$P\{im_2 \leq r^2\} = \int \int_{t' \leq r^2} K (1 - t D t')^{\frac{N-i-3}{2}} dt$$

En développant en série le binôme, on est ramené à une intégration terme à terme qui est au moins valable si le développement est limité, c'est-à-dire si $N - i - 3$ est positif et pair.

Dans ce cas

$$P \{ i m_2 \leq r^2 \} = \sum \binom{N-i-3}{2} (-1)^j K J^j \quad \text{avec } J_j = \int \int_{t' \leq r^2} (t D t')^j dt$$

il est possible de montrer que si $r^2 \leq N - i$

$$J_j = \pi^{i/2} \frac{(r^2)^{\frac{i}{2} + j}}{\Gamma \left(\frac{i}{2} + j + 1 \right)} M_j, \quad \text{où } M_j \text{ est le moment d'ordre } j \text{ de } \frac{u u'}{2}$$

quand $u = (u_1 \dots u_i)$ a une distribution de Laplace-Gauss de matrice de covariance D (1).

Les valeurs caractéristiques de la matrice D étant $\frac{1}{N}$, $i - 1$ fois et $\frac{1}{N - i}$ la fonction caractéristique de $\frac{u u'}{2}$ est

$$\Phi_{\frac{u u'}{2}}(\lambda) = \left(1 - \frac{\lambda \vec{i}}{N - i} \right)^{-\frac{i}{2}} \left(1 - \frac{\lambda \vec{i}}{N} \right)^{-\frac{i-1}{2}}$$

m peut développer en série $\Phi(\lambda)$ quand $\lambda < N - i$.

$$\Phi_{\frac{u u'}{2}}(\lambda) = \sum_0^{\infty} \binom{i}{2}_j \left(\frac{1}{N - i} \right)^j {}_2F_1 \left(-j, \frac{i-1}{2}; \frac{i}{2}; \frac{\lambda \vec{i}}{N} \right) \left(\frac{\lambda \vec{i}}{j!} \right)^j$$

(Voir par exemple, Erdelyi : *Higher Transcendental Functions*, I, p. 82).

On a donc :

$$J_j = \frac{\pi^{i/2} (r^2)^{\frac{i}{2} + j}}{\Gamma \left(\frac{i}{2} + j + 1 \right)} \binom{i}{2}_j \frac{1}{(N - i)^j} {}_2F_1 \left(-j, \frac{i-1}{2}; \frac{i}{2}; \frac{\lambda \vec{i}}{N} \right) \text{ si } r^2 < N - i$$

Toutes réductions faites, on obtient :

$$P \{ i m_2 \leq r^2 \} = \frac{\sum (-1)^k \binom{N-i-3}{2} \binom{N}{N-i}^{k+\frac{1}{2}}}{\binom{i}{2} + k} B \left(\frac{N-i-1}{2}, \frac{i}{2} \right) {}_2F_1 \left(-K, \frac{i-1}{2}; \frac{i}{2}; \frac{\lambda \vec{i}}{N} \right) \left(\frac{r^2}{N} \right)^{\frac{i}{2} + K} \text{ si } r^2 < N - i$$

on obtiendrait une formule analogue pour $r^2 > N - i$.

(1) Dans un ordre d'idées voisin, voir « *Notes on the Distribution of a Definite Quadratic Form* », par J. Pachares, in *A. M. S.* 26, 1, pp. 128-132.

**III. ÉCHANTILLON EXTRAIT D'UNE POPULATION FINIE
DE PARAMÈTRES \bar{X} ET S , $\bar{X} = x$ ET $S = s$. RÉSULTATS CONDITIONNELS**

On se propose d'étudier la distribution d'un échantillon ξ extrait d'une population de paramètres $\bar{X} = x$ et $S = s$, \bar{X} et S n'étant plus des variables aléatoires mais des valeurs numériques.

III.1. La distribution conditionnelle de ξ pour \bar{X} et S donnés s'obtient en divisant la distribution simultanée de ξ , \bar{X} , S^2 , par la distribution simultanée de \bar{X} et S^2 .

On écrit la distribution simultanée de ξ , \bar{X}_n et S_n^2 , on effectue le changement de variables :

$$\bar{X}_n = \frac{N}{N-i} \bar{X} - \frac{i \bar{\xi}}{N-i}$$

$$S_n^2 = \frac{N}{N-i} S^2 - \frac{i^2 (\bar{\xi} - \bar{X})^2}{(N-i)^2} - \frac{\sum (\xi_i - \bar{X})^2}{N-i},$$

on obtient

$$\frac{D(\xi \bar{X}_n S_n^2)}{D(\xi \bar{X} S^2)} = \left(\frac{N}{N-i} \right)^2 \text{ pour le Jacobien de la transformation.}$$

Divisant par la distribution simultanée de \bar{X} et S^2 , on obtient après simplification.

$$\frac{1}{\Pi^{1/2}} \frac{\Gamma\left(\frac{N-1}{2}\right)}{\Gamma\left(\frac{N-i-1}{2}\right)} \frac{1}{(N-i)^{1/2} N^{\frac{i-1}{2}}} \left[1 - \sum \frac{(\xi_i - \bar{X})^2}{N S^2} - \frac{i^2 (\bar{\xi} - \bar{X})^2}{(N-i) N S^2} \right]^{\frac{N-i-3}{2}} \frac{d\xi}{S_i}$$

le crochet peut s'écrire encore

$$\left[1 - \frac{i}{N-i} \frac{(\bar{X} - \bar{\xi})^2}{S^2} - \frac{i S_n^2}{N S^2} \right]$$

ou

$$\left[1 - \left(\frac{\xi - \bar{X}}{S} \right) D \left(\frac{\xi - \bar{X}}{S} \right)' \right]$$

En rapprochant ces résultats de ceux qui ont été obtenus en II.3.1, on voit que la distribution conditionnelle de $\frac{\xi - \bar{X}}{S}$ est la même que la distribution de t , elle est indépendante de \bar{X} et S . Il en résulte que tous les résultats obtenus en III.3 pour la distribution non conditionnelle des caractéristiques de $t = \frac{\xi - \bar{X}}{S}$ sont valables pour les caractéristiques de $\frac{\xi - \bar{X}}{S}$ à \bar{X} et S donnés, et, par la transformation $\xi = tS + \bar{X}$, pour les caractéristiques de ξ pour \bar{X} et S données, c'est-à-dire que le problème de la généralisation au cas d'une population finie (\bar{X} , S) des résultats valables pour une population infinie (m , σ) est résolu.

Le rapprochement de la loi de distribution de ξ , à \bar{X} et S donnés, et de la loi de distribution de $(\bar{\xi}$ et $S^2)$ à \bar{X} et S donnés et le fait que le champ de variation de ξ est défini par celui du couple $(\bar{\xi}, S^2)$, montrent que la loi de distribution de ξ à \bar{X} et S donnés admet $\bar{\xi}, S^2$ pour résumé exhaustif.

Ces résultats permettent de tester des hypothèses concernant la population finie et d'estimer, sur la base de l'information apportée par un échantillon, celles des caractéristiques de la population finie qui sont fonction de \bar{X} et de S .

Dans le cas où $i = N - 2$, on obtient la distribution des observations d'un échantillon de moyenne et de variance données; on sait que le point représentatif d'un tel échantillon se déplace sur une hypersphère à $n - 2$ dimensions et que le paramètre de position de ce point est un hyperangle ψ de loi de distribution

$$\frac{d\psi}{2\pi^{\frac{N-1}{2}}/\Gamma\left(\frac{N-1}{2}\right)}$$

En particulier, la loi de distribution de Thompson est une loi conditionnelle donnant

$$P\left\{\xi_i \leq x \mid \bar{X}, S\right\} = P\left\{\frac{\xi - \bar{X}}{S} \leq \frac{x - \bar{X}}{S} \mid \bar{X}, S\right\} = P\left\{t_i \leq \frac{x - \bar{X}}{S}\right\},$$

c'est-à-dire la probabilité qu'une observation donnée tirée d'une population finie (\bar{X}, S) soit au plus égale à une valeur x , elle généralise, au cas d'une population finie, la probabilité intégrale Laplace-Gaussienne $P\left\{\xi_i \leq x \mid m, \sigma\right\}$; elle peut permettre la construction d'un graphique analogue à celui qui est utilisé dans la méthode de la droite de Henri : étant donnée la fréquence cumulative $\frac{k}{i}$ des observations d'un échantillon dont la valeur est inférieure ou au plus égale à x , on peut porter en ordonnée de chaque abscisse x d'un graphique, non pas $\frac{k}{i}$, mais t_k défini par la relation $P(t \leq t_k) = \frac{k}{i}$, le graphique obtenu devra être sensiblement linéaire si la population dont l'échantillon est extrait est Laplace-Gaussienne et l'abscisse à l'origine de la droite estimée fournira une estimation de \bar{X} , tandis que la pente fournira une estimation de l'inverse de S .

Étant donnée une valeur x et un échantillon de taille i extrait d'une population finie (\bar{X}, S) , la fréquence $\frac{k}{i}$ des observations de valeur au plus égale à x dans l'échantillon est une variable aléatoire susceptible de varier de 0 à 1 (si, par exemple la population finie est un lot de produits manufacturés et x une spécification minimum pour X , $\frac{k}{i}$ est la proportion d'articles défectueux dans l'échantillon); dans le cas où l'échantillon est extrait d'une population infinie (m, σ) , la distribution de $\frac{k}{i}$ est une distribution de Bernoulli, de moyenne $E\left[\frac{k}{i}\right] = \int_{-\infty}^{\infty} N_x(m, \sigma) dx = P\left[Z \leq \frac{x - m}{\sigma}\right]$; dans le cas d'une population finie la distribution peut être obtenue en notant que $\frac{k}{i} = \frac{1}{i} \sum_1^k W_i$, où $W_i = \mathbb{1}$

· si $\xi_j \leq x$ et $W_j = 0$ Si $\xi_j > x$ on peut en effet obtenir tous les moments de $\frac{k}{i}$;
 (du moins théoriquement) :

$$m_1 = E \left[\left(\frac{k}{i} \right) | \bar{X} S \right] = \frac{1}{i} E [\Sigma W_j | \bar{X} S] = E [W_j | \bar{X} S] = P [\xi_j \leq x | \bar{X} S] = P \left\{ t_j \leq \frac{x - \bar{X}}{S} \right\}$$

.....

$$m_l = E \left[\left(\frac{k}{i} \right)^l | \bar{X} S \right] = \frac{1}{i^l} E [(\Sigma W_j)^k | \bar{X} S] = \frac{\Sigma}{i^l} E \left[W_1^{\alpha_1} W_2^{\alpha_2} \dots W_l^{\alpha_l} \frac{l!}{\alpha_1! \alpha_2! \dots \alpha_l!} | \bar{X} S \right]$$

$$= \frac{\Sigma}{i^l} \frac{l!}{\alpha_1! \alpha_2! \dots \alpha_l!} P \left\{ t_1 \text{ et } t_2 \text{ et } \dots \text{ et } t_l \leq \frac{x - \bar{X}}{S} \right\}$$

(avec $\alpha_1 + \alpha_2 + \dots + \alpha_l = l$ et une variable t_i ne figurant dans le crochet que pour autant que $\alpha_i \neq 0$).

Il suffit ensuite de résoudre le système :

$$p(0) + p\left(\frac{1}{i}\right) + \dots + p(1) = 1$$

$$p\left(\frac{1}{i}\right) + \frac{2}{i} p\left(\frac{2}{i}\right) + \dots + p(1) = m_1$$

.....

$$p\left(\frac{1}{i}\right) + \left(\frac{2}{i}\right)^l p\left(\frac{2}{i}\right) + \dots + p(1) = m_l$$

.....

$$p\left(\frac{1}{i}\right) + \left(\frac{2}{i}\right)^t p\left(\frac{2}{i}\right) + \dots + p(1) = m_t$$

pour obtenir la distribution cherchée.

Malheureusement, l'obtention des moments exige le calcul d'intégrales multiples qui n'a pas encore pu être effectué. Il faut se contenter à l'heure actuelle du premier moment $m_1 = E \left[\left(\frac{k}{i} \right) | \bar{X} S \right] = P \left\{ t_j \leq \frac{x - \bar{X}}{S} \right\}$ -.

Ce premier moment est une estimation « non biaisée » de la proportion p d'observations de valeur au plus égale à x dans la population de référence, soit $p = P \left\{ Z \leq \frac{x - m}{\sigma} \right\}$; la proportion $\frac{k}{i}$ d'observations de valeur au plus égale à x dans l'échantillon en constitue une autre estimation. Il résulte du théorème de Blackwell et d'un théorème de Lehmann et Scheffé que $m_1 = p^1 \left(\frac{x - \bar{X}}{S} \right)$ constitue l'estimation ayant uniformément une variance minimum dans la classe des estimations non biaisées de p , - l'estimation p^1 fondée sur la moyenne et l'écart-type du lot doit être préférée à celle qui est donnée par la fréquence cumulative dans n'importe quel échantillon extrait du lot.

Il est aisé de donner, grâce à une représentation paramétrique, la distribution de la fonction $p^1 \left(\frac{x - \bar{X}}{S} \right)$ — estimation de p :

Étant donné n'importe quelle valeur ϖ

$$P \{ p^1 \leq \varpi \} = P \left\{ \frac{x - \bar{X}}{S} \leq t_{\varpi} \right\} \text{ si } t_{\varpi} \text{ désigne la valeur limite telle que}$$

$$P \{ t_{\varpi} \leq t_{\varpi} \} = \varpi \text{ (} t_{\varpi} \text{ est fournie par une table de la distribution de Thompson).}$$

d'où

$$p^1 \leq \varpi \} = P \left\{ \frac{\frac{(\bar{X} - m) \sqrt{N} - \frac{(x - m)}{\sigma} \sqrt{N}}{\frac{\sqrt{N} S}{\sqrt{N-1} \sigma}} > -\sqrt{N-1} t_{\varpi}}{\chi / \sqrt{N-1}} \leq -\sqrt{N-1} t_{\varpi}} \right\} = 1 - P \left\{ \frac{Z - Z_p \sqrt{N}}{\chi / \sqrt{N-1}} \leq -\sqrt{N-1} t_{\varpi} \right\}$$

Cette probabilité fonction de ϖ et p et N est fournie par une table de la distribution de Student non centrée « non central Student's distribution » (1). D'où la possibilité de construire des « cartes de contrôle » pour la fonction d'échantillon p^1 ou de tester, au moyen de p^1 , des hypothèses concernant la proportion p . Cette possibilité a récemment été exploitée par Lieberman (article cité) en « Sampling Inspection ».

Dans les applications de la distribution de ξ à \bar{X} et S donnés à l'étude des séries temporelles ou au contrôle statistique des fabrications, il est souvent utile d'avoir la distribution de la j ^{ième} observation connaissant les $j - 1$ premières observations.

On écrit la distribution de j observations et on divise par la distribution de $j - 1$ observations en tenant compte des relations :

$$j \bar{\xi} = (j - 1) \bar{\xi}^* + \xi,$$

$$S_{\xi}^2 = \frac{j - 1}{j} S_{\xi^*}^2 + \frac{(j - 1)}{j^2} (\xi_j - \bar{\xi}^*)^2$$

Si $\bar{\xi}$ et S_{ξ}^2 désigne la moyenne et la variance de j observations et $\bar{\xi}^*$ et $S_{\xi^*}^2$ la moyenne et la variance de $j - 1$ observations.

On obtient après groupements de termes :

$$\frac{1 - \frac{N - j + 1}{N(N - j)} S^2 \left[\xi_j - \bar{X} - \frac{j - 1}{N - j + 1} (\bar{X} - \bar{\xi}^*) \right]^2 - \frac{j - 1}{N - j + 1} \left(\frac{\bar{X} - \bar{\xi}^*}{S} \right)^2 - \frac{j - 1}{N} \frac{S_{\xi^*}^2}{S^2} \left\{ \frac{N - j - 2}{2} \right\}}{K' \left\{ 1 - \frac{j - 1}{N - j + 1} \left(\frac{\bar{X} - \bar{\xi}^*}{S} \right)^2 - \frac{j - 1}{N} \frac{S_{\xi^*}^2}{S^2} \left\{ \frac{N - j - 2}{2} \right\} \right\}} \frac{d \xi_j}{S}$$

(1) Dans les Tables de Johnson et Welch, *Biometrika*, 51, 1939, on devra chercher $\varpi = -Z_p \sqrt{N}$ et $t = -\sqrt{N-1} t_{\varpi}$.

Il en résulte que la variable.

$$\tau = \frac{N-j+1}{N(N-j)} \frac{\left[\frac{\xi_j - \bar{X}}{S} + \frac{j-1}{N-j+1} \left(\frac{\bar{\xi}^* - \bar{X}}{S} \right) \right]^2}{1 - \frac{j-1}{N-j+1} \left(\frac{\bar{\xi}^* - \bar{X}}{S} \right)^2 - \frac{j-1}{N} \frac{S_i^2}{S^2}}$$

à \bar{X} , S , ξ_1, \dots, ξ_{j-1} ,
donnés,
valable pour $j < N - 2$

suit une loi de Bêta incomplète de paramètre $\frac{1}{2}$ et $\frac{N-j-1}{2}$

On a de nouveau une loi de Thompson.

On peut démontrer que la distribution précédente

$$P \{ \xi_j \mid \xi_1 \xi^2 \dots \xi_{j-1} \bar{X} S \mid d \xi_j \text{, est également } P \{ \xi_j \mid \bar{\xi}^* S_i^2 \bar{X} S \mid d \xi_j \text{.}$$

IV. APPLICATIONS

Les applications des formules obtenues précédemment sont nombreuses. Un grand nombre toutefois constituent des résultats classiques qu'on peut obtenir directement.

Estimation et test

1. Une des applications les plus importantes des formules obtenues en III résulte du fait qu'en vertu du théorème de Blackwell-Lehmann-Scheffé, \bar{X} et S constituant un « résumé exhaustif » « complet » des paramètres (m, σ) de la population de référence dont la population finie $(\bar{X} S)$ est extraite, si $\varphi(\xi)$ est une estimation « non biaisée » d'une caractéristique « estimable » de la population de référence, l'espérance mathématique $E[\varphi(\xi) \mid \bar{X} S]$ de φ , à \bar{X} et S donnés, est une estimation « non biaisée » de cette caractéristique, ayant uniformément une variance minimum. Cette propriété rend la distribution des ξ à \bar{X} et S donnés, c'est-à-dire la distribution de t particulièrement intéressante.

Tantôt elle permet d'obtenir de telles estimations, et l'estimation de la probabilité intégrale de la loi de Laplace-Gauss $P[X \leq x]$ donnée au § III illustre cette méthode; tantôt elle permet d'obtenir l'espérance mathématique de certaines caractéristiques d'un échantillon tiré d'une population finie. Un exemple « trivial » est : $\frac{N}{N-1} S^2$ est l'estimation « non biaisée » de variance uniformément minimum de σ^2 et $\frac{i S_i^2}{i-1}$ est une estimation « non biaisée » de σ^2 , donc

$$E \left[\frac{i S_i^2}{i-1} \mid \bar{X} S \right] = \frac{N S}{N-1}$$

d'où

$$E[S_i^2] = \frac{i-1}{i} \frac{N}{N-1} S^2, \text{ à } \bar{X} \text{ et } S \text{ donnés}$$

2. La technique, citée au § III, permet d'établir un plan de « contrôle d'acceptation par variable » de lots de produits industriels, lorsque ces lots sont « infinis », et qu'il existe une spécification minimum L (ou maximum).

Étant donné un échantillon ξ d'effectif i , de moyenne $\bar{\xi}$ et d'écart-type S_{ξ} , extrait d'un lot « infini » de moyenne m , d'écart-type σ , contenant la proportion p d'éléments de valeur au plus égale à L : $p = \int_{-\infty}^L N_{\sigma}(m, \sigma) dx$, la meilleure estimation de p est

$$p^1 = P \left[\xi_i \leq L \left[\bar{\xi}, S_{\xi} \right] = P \left\{ \frac{\xi - \bar{\xi}}{S_{\xi}} \leq \frac{L - \bar{\xi}}{S_{\xi}} \right\} = f \left(\frac{L - \bar{\xi}}{S_{\xi}} \right)$$

obtenue au moyen de la distribution de Thompson, et la loi de distribution de l'estimation p^1 est

$$P \{ p^1 \leq \varpi \} = 1 - P \left\{ \frac{Z - Z_p \sqrt{i}}{\chi / \sqrt{i-1}} \leq -\sqrt{i-1} t_{-} \right\} = g(-Z_p \sqrt{i}, -\sqrt{i-1} t_{-})$$

obtenue au moyen de la loi de Student non centrée, où t_{-} est fournie par $f(t_{-}) = \varpi$ (loi de Thompson) et Z_p par $P \{ Z \leq Z_p \} = p$ (loi de Laplace-Gauss); il en résulte que le « test », ou règle d'acceptation :

$$\begin{cases} \text{Accepter le lot si } p^1 \leq \varpi \\ \text{Rejeter le lot si } p^1 > \varpi \end{cases}$$

est caractérisé par une *courbe caractéristique efficace* (« Operating Characteristic Curve ») définie par $g(-Z_p \sqrt{i}, -\sqrt{i-1} t_{-})$, probabilité d'accepter un lot de proportion défectueuse p quand on suit la règle d'acceptation définie par ϖ .

Une règle équivalente caractérisée par une valeur limite de $\frac{L - \bar{\xi}}{S_{\xi}}$ est la suivante :

$$\begin{cases} \text{Accepter le lot si } \frac{L - \bar{\xi}}{S_{\xi}} \leq K \\ \text{Rejeter le lot si } \frac{L - \bar{\xi}}{S_{\xi}} > K \end{cases}$$

à condition que $K = t_{-}$.

De tels plans d'échantillonnage ont fait l'objet d'une littérature importante et sont couramment utilisés.

3. Dans le cas où les lots soumis au contrôle d'acceptation sont finis, les techniques décrites plus haut ne sont plus valables, elles sont cependant couramment utilisées. Notre propos est d'étudier les conséquences de l'application des plans d'acceptation du type précédent quand les lots soumis à l'inspection sont des lots finis de moyenne \bar{X} et d'écart-type S , au lieu de lots infinis de moyenne m et d'écart-type σ — situation qui se rencontre chaque fois que l'effectif N des lots n'est pas très important.

Remarquons tout d'abord que

$$P \left[\xi_i \leq L / \bar{\xi} S_{\xi} \mid \bar{X} S \right] = P \left[\xi_i \leq L / \bar{\xi} S_{\xi} \right];$$

c'est-à-dire que la probabilité pour qu'une observation quelconque dans un échantillon ξ soit au plus égale à \bar{L} , à $\bar{\xi}$ et S_{ξ} donnés, est la même que $\bar{X} S$ soient donnés ou non, c'est-à-dire que le lot soit infini ou non; on peut le vérifier directement par le calcul; on peut aussi remarquer que $\bar{\xi} S_{\xi}$ étant un résumé

exhaustif pour $\bar{X} S$, la distribution de ξ_i à $\bar{\xi}$ et S_i donnés est nécessairement indépendante de \bar{X} et S et dépend seulement de $\bar{\xi}$ et S_i . Il en résulte que $p^1 = E \left[\frac{K}{i} \mid \bar{\xi} S_i \right] = E \left[\frac{K}{i} \mid \bar{\xi} S_i \bar{X} S \right] = f \left(\frac{L - \bar{\xi}}{S_i} \right)$. L'estimation p^1 obtenue dans le cas d'un lot infini, n'est pas modifiée par le fait que le lot est fini de moyenne \bar{X} et S , elle est indépendante de \bar{X} et S . Mais alors que la proportion défectueuse d'un lot infini est une fonction de m et σ , $p = \int_{-\infty}^L N_x(m, \sigma) dX$, il n'en est pas de même d'un lot fini :

Étant donnés \bar{X} et S , p^1 la proportion défectueuse d'un lot fini est une variable aléatoire susceptible de varier de 0 à 1 et dont la loi de probabilité peut être, en théorie, déterminée en fonction de \bar{X} et S_1 (en calculant les moments de p^1 — voir § III).

Un lot fini est donc insuffisamment décrit par \bar{X} et S , pour autant qu'on s'intéresse aux éléments défectueux. Il en résulte que $\bar{\xi}$ et S_i ne constituent plus un résumé exhaustif des paramètres $\bar{X} S p^1$ d'un tel lot; partant, p^1 n'est pas une estimation « non biaisée » et de variance uniformément minimum de p^1 .

Examinons quelle est la distribution de p^1 dans ces conditions :

$P [p^1 \leq \varpi] = P \left[\frac{L - \bar{\xi}}{S_i} \leq t \right]$ si t est défini par $P \left[\frac{\xi - \bar{\xi}}{S_i} \leq t \right] = \varpi$ obtenu au moyen de la distribution de Thompson.

$$a) \quad P \left\{ \frac{L - \bar{\xi}}{S_i} \leq t \right\} = g(-Z_p \sqrt{i}, \sqrt{i-1} t)$$

Si \bar{X} et S ne sont pas fixés, p étant la « fraction défectueuse » de la population d'où provient le lot.

b) Si \bar{X} et S sont donnés, c'est-à-dire si l'on définit un lot par sa moyenne \bar{X} et sa standard déviation S .

$$P \left\{ \frac{L - \bar{\xi}}{S_i} \leq t \mid \bar{X} S \right\} = 1 - P \left\{ \frac{\bar{\xi} - L}{S_i} \leq -t \mid \bar{X} S \right\} = 1 - P \left\{ \frac{\frac{\bar{\xi} - \bar{X}}{S} - \frac{L - \bar{X}}{S}}{\frac{S_i}{S}} \leq -t \mid \bar{X} S \right\}$$

$$= 1 - P \left\{ \frac{t - (L - \bar{X}) / S}{S_i} \leq -t \mid \bar{X} S \right\}$$

et la distribution de

$$U = \frac{t - (L - \bar{X}) / S}{S_i}$$

n'est plus une distribution de « Student » non centrée.

La distribution de t et S_i^2 est donnée par

$$\frac{i^{i/2}}{N^{\frac{i-1}{2}} (N-i)^{i/2}} \frac{\Gamma \left(\frac{N-1}{2} \right)}{\pi^{1/2} \Gamma \left(\frac{N-i-1}{2} \right) \Gamma \left(\frac{i-1}{2} \right)} \left[1 - \frac{i}{N} S_i^2 - \frac{i}{N-i} t^2 \right]^{\frac{N-i-3}{2}} \frac{t^{-3}}{(S_i^2)^2} dS_i^2 dt$$

à l'intérieur d'une demi-ellipse de centre ($\bar{t} = o$, $S_i = o$) et de demi-axes $\sqrt{\frac{N-i}{i}}$ et $\sqrt{\frac{N}{i}}$; si I est le point $(S_i = o, \bar{t} = \frac{L-\bar{X}}{S})$ et M le point de coordonnées courantes (\bar{t}, S_i), une valeur constante pour u correspond à un angle constant φ de IM avec l'axe des \bar{t} : $U = \text{Co } t \varphi$.

Pour obtenir la loi de distribution de U, il convient donc de calculer la loi de distribution de φ , c'est-à-dire la masse de probabilité contenue dans un secteur de sommet I et d'angle $d\varphi$. Si I est intérieur à l'ellipse, ce calcul est simplifié si on opère une affinité sur l'ellipse, définie par $Y = \lambda \bar{t}$ de façon à ce que l'ellipse transformée ait un foyer en I. Il suffit pour cela de prendre

$$\lambda = \sqrt{\frac{N}{i}} / \sqrt{\frac{N-i}{i} - \left(\frac{L-\bar{X}}{S}\right)^2}.$$

U est transformé en $U' = \lambda U$

la distribution de \bar{t} , S_i^2 devient

$$\frac{i^{i/2}}{\lambda N^{\frac{i-1}{2}} (N-i)^{1/2}} \frac{\Gamma\left(\frac{N-1}{2}\right)}{\pi^{1/2} \Gamma\left(\frac{N-i-1}{2}\right) \Gamma\left(\frac{i-1}{2}\right)} \left[1 - \frac{Y^2}{a^2} - \frac{S_i^2}{b^2}\right]^{\frac{N-i-3}{2}} (S_i^2)^{\frac{i-3}{2}} dS_i^2 dY.$$

avec

$$a = \sqrt{\frac{N-i}{i}} \lambda \quad b = \sqrt{\frac{N}{i}} \quad e = \frac{\sqrt{a^2 - b^2}}{a} = \sqrt{\frac{i}{N-i} \frac{L-\bar{X}}{S}}$$

en supposant $L > \bar{X}$.

φ est transformé en φ' .

En passant en coordonnées polaires, l'équation de la demi-ellipse devient

$$\rho^* = \frac{b^2}{a(1 + e \cos \varphi')};$$

en tenant compte de

$$\frac{\sin^2 \varphi'}{b^2} + \frac{\cos^2 \varphi'}{a^2} = \frac{1}{b^2} \left[1 - \frac{\sqrt{a^2 - b^2}}{a} \cos \varphi'\right] \left[1 + \frac{\sqrt{a^2 - b^2}}{a} \cos \varphi'\right]$$

la distribution devient :

$$K (\sin^2 \varphi')^{\frac{i-2}{2}} d\varphi' \left\{ \left(\frac{1 - e^2 \cos^2 \varphi'}{b^2} \right) \left[-\rho + \frac{b^2}{a(1 + e \cos \varphi')} \right] \left[\rho + \frac{b^2}{a(1 - e \cos \varphi')} \right] \right\}^{\frac{N-i-3}{2}} \rho^{\frac{i-1}{2}} d\rho$$

En intégrant par rapport à ρ de 0 à ρ^* , on obtient une intégrale de Riemann Louville dont on trouvera le type dans une table, par exemple dans Erdelyi (*op. cit.*, p. 186). Après regroupement des termes et simplifications, on trouve pour la distribution de φ' , la densité de probabilité :

$$\frac{\Gamma(i) \Gamma\left(\frac{N-1}{2}\right)}{\pi^{1/2} \Gamma\left(\frac{i-1}{2}\right) \Gamma\left(\frac{N+i-1}{2}\right)} \left[1 - \frac{i}{N-i} \left(\frac{L-\bar{X}}{S}\right)^2\right]^{\frac{N-2}{2}} \frac{(\sin^2 \varphi')^{\frac{i-2}{2}}}{(1 + e \cos \varphi')^2} {}_2F_1 \left[-\frac{N-i-3}{2}, i; \frac{N+i-1}{2}; -\frac{1-e \cos \varphi'}{1+e \cos \varphi'} \right]$$

L'intégration par rapport à $\cos\varphi'$ est plus ou moins aisée suivant la parité de i et de N , on la réalise par l'intégration terme à terme du développement en série obtenu après expansion de ${}_2F_1$.

Pour obtenir la probabilité intégrale

$\Phi_\lambda(u) = P\{U \leq u \mid \bar{X}, S\}$ il faut évidemment effectuer cette intégration de

$$\varphi' = -\pi \text{ à } \varphi' = \text{Arc cot } \lambda u$$

Quand $\bar{X} < L < \bar{X} + \sqrt{\frac{N-i}{i}}$, la distribution de p^1 à \bar{X} et S donnés est donc :

$$P[p^1 \leq \pi \mid \bar{X}, S] = 1 - \Phi_\lambda(-t_-);$$

cette distribution dépend de \bar{X} et S par l'intermédiaire du paramètre λ , c'est-à-dire de $\frac{L - \bar{X}}{S}$.

Il en résulte que lorsqu'on applique à l'inspection sur échantillon d'un lot fini, la règle d'acceptation :

$$\left\{ \begin{array}{l} \text{Accepter le lot si } p^1 \leq \pi \text{ ou } \frac{L - \bar{X}}{S} \leq t_+ \\ \text{Rejeter le lot si } p^1 > \pi \text{ ou } \frac{L - \bar{X}}{S} > t_- \end{array} \right.$$

la courbe caractéristique efficace du plan est fonction de $\frac{L - \bar{X}}{S}$, non de p^1 , c'est-à-dire que dans un tel plan, on contrôle $\frac{L - \bar{X}}{S}$, non p^1 .

On notera toutefois que :

a) $f\left(\frac{L - \bar{X}}{S}\right)$ donnée par la loi de Thompson est $E[p^1 \mid \bar{X}, S]$, l'espérance mathématique de la fraction défectueuse d'un lot fini de moyenne \bar{X} et d'écart-type S ; $f\left(\frac{L - \bar{X}}{S}\right)$ étant une fonction monotone de $\frac{L - \bar{X}}{S}$, on peut exprimer la courbe caractéristique efficace du plan, en fonction de $f\left(\frac{L - \bar{X}}{S}\right)$, c'est-à-dire que si les lots de mêmes \bar{X} et S étaient groupés, la courbe donnerait la proportion de ces lots qui est acceptée en fonction de la « fraction défectueuse » obtenue en les groupant en un lot unique.

b) Si l'on peut admettre valablement que les lots examinés proviennent d'une même population caractérisée par une « proportion défectueuse » p d'ailleurs inconnue, l'application du procédé d'acceptation ci-dessus donne lieu à une courbe caractéristique efficace, fonction de p , soit $g(-Z_p \sqrt{i}, \sqrt{i-1} t_-)$; par ailleurs, p^1 proportion défectueuse dans le lot est une variable aléatoire suivant une loi binomiale $\binom{N}{N p^1} p^{N p^1} (1-p)^{N(1-p^1)}$; pour chaque valeur de p^1 ne peut construire un intervalle de confiance pour p , couvrant la valeur p avec une probabilité donnée; pour chaque valeur de p^1 on peut donc construire un intervalle de confiance pour la probabilité $g(-Z_p \sqrt{i}, -\sqrt{i-1} t_-)$

d'accepter un lot et substituer ainsi la courbe caractéristique du plan, en fonction de p' , une bande de confiance pour cette courbe.

4. Pour estimer \bar{X} et \bar{S} , paramètres d'une population finie, on ne peut attendre de la méthode « maximum likelihood » qu'elle fournisse des estimations optima puisque le champ de variations des variables est fonction des paramètres. Toutefois une telle méthode est intuitivement satisfaisante.

Les estimations obtenues sont :

$$\widehat{\bar{X}} = \bar{\xi} \quad \text{et} \quad \left\{ \begin{array}{l} \widehat{S^2} = S_i^2 \frac{N-3}{N} \text{ si } i < N-3 \\ \widehat{S^2} = S_i^2 \frac{i}{N} \text{ si } i > N-3 \end{array} \right.$$

Comme $\widehat{S^2} = S^2 \frac{N-3}{N} S_i^2$ et S_i^2 a une Bêta distribution, l'estimation « non biaisée » de S^2 est $S^{2/2} = \frac{i}{i-1} \cdot \frac{N-1}{N} S_i^2$

$\bar{\xi}$ est une estimation « non biaisée » de \bar{X} de variance

$$S^2 \frac{N-i}{i} \frac{1}{N-1} = \sigma_{\bar{\xi}}^2,$$

un résultat bien connu.

5. Quant à « tester » une « hypothèse » $\bar{X} = \bar{X}_0$ contre $\bar{X} \neq \bar{X}_0$, on peut appliquer la règle du « maximum likelihood ratio ».

Le « maximum de vraisemblance » pour S variant seul est proportionnel à

$$\left[\frac{(\bar{X} - \bar{\xi})^2}{N-i} + \frac{S_i^2}{N} \right]^{-\frac{i}{2}},$$

tandis qu'il est proportionnel à $(S_i^2)^{-\frac{i}{2}}$ au cas où S et \bar{X} varient tous les deux; le critère d'acceptation ou de rejet de l'« hypothèse nulle » est donc fondé sur le comportement de la quantité $\left[\frac{(\bar{X} - \bar{\xi})^2}{N-i S_i^2} + 1 \right]^{-\frac{i}{2}}$, c'est-à-dire sur celui

$$\left(\frac{\bar{X} - \bar{\xi}}{S_i} \right)^2 = \frac{\bar{t}^2}{S_i^2}.$$

D'après la distribution obtenue au § II.3.4, en effectuant le changement de variable

$$t = u S_i \\ V^2 = S_i^2 \left(\frac{i}{N} + \frac{i}{N-i} u^2 \right)$$

on obtient, après intégration de V^2 de 0 à 1 et regroupement de termes :

$$\frac{1}{B\left(\frac{1}{2}, \frac{i-1}{2}\right)} \frac{1}{\left(1 + \frac{i}{N-i} \frac{\bar{t}^2}{S_i^2}\right)^{i/2}} d\left(\frac{\bar{t}}{S_i} \sqrt{\frac{N}{N-i}}\right)$$

c'est-à-dire $\frac{\bar{t}}{S_i} \sqrt{\frac{N}{N-i}}$ suit une loi de Student à $i-1$ degrés de liberté.

(La variable à utiliser pour les tables est $\sqrt{\frac{N(i-1)}{N-i} \frac{\bar{t}}{S_i}}$).

Si τ est la variable utilisée dans les tables, le test à appliquer est donc :
Rejeter l'hypothèse nulle si

$$\bar{\xi} < \bar{X}_o + \sqrt{\frac{N-i}{N}} \tau_{\frac{\alpha}{2}} S_i / \sqrt{i-1}$$

ou si

$$\bar{\xi} > \bar{X}_o + \sqrt{\frac{N-i}{N}} \tau_{1-\frac{\alpha}{2}} S_i / \sqrt{i-1}$$

dans le cas où le niveau de signification est α .

La distribution de $\frac{\bar{X} - \bar{\xi}}{S_i}$ sert également à construire des intervalles de confiance pour l'estimation de \bar{X} , du type $\bar{\xi} \pm \sqrt{\frac{N-i}{N}} \tau_{\frac{\alpha}{2}} S_i / \sqrt{i-1}$ pour un niveau de confiance de $1 - \alpha$.

6. Pour tester une « hypothèse » $S^2 = S_o^2$ contre $S^2 \neq S_o^2$, la même méthode conduit à un « likelihood ratio » proportionnel à $\left(1 - \frac{i}{N} \frac{S_i^2}{S^2}\right)^{\frac{N-i-3}{2}} \left(\frac{S_i^2}{S^2}\right)^{i/2}$ c'est-à-dire $\left(1 - \frac{i}{N} S_i^2\right)^{\frac{N-i-3}{2}} (S_i^2)^{i/2}$, quantité qui n'est pas une fonction monotone de S_i^2 , dans l'intervalle de variation de S_i^2 , du moins si $i < N - 3$, ou si i est compris dans l'intervalle ouvert $\frac{N}{2} \pm \frac{\sqrt{N^2 - 12}}{2}$. Des calculs très pénibles en résultent, si l'on veut obtenir la distribution du « likelihood ratio ».

Si l'on veut tester l'hypothèse $S^2 = S_o^2$ contre $S^2 > S_o^2$, il sera en général suffisant de rejeter l'« hypothèse nulle » si S_i^2 est trop grand; c'est-à-dire si $S_i^2 > (S_i^2)_{1-\alpha}$ si α est le niveau de signification du test.

D'après la distribution du § II.3.4.2., $\tau'_{1-\alpha}$ étant le quantile $1 - \alpha$ de la distribution Bêta incomplète $I_{\nu} \left(\frac{i-1}{2}, \frac{N-i}{2}\right)$, on rejettera l'hypothèse nulle si $S_i^2 > \frac{N}{i} S_o^2 \tau'_{1-\alpha}$.

7. De tels tests peuvent être utilement employés en *contrôle statistique des fabrications* par « cartes de contrôle », quand on commence un contrôle, au moins comme procédé de dégrossissage.

Les traités classiques supposent en effet que, lors de la construction de telles cartes pour le contrôle de la stabilité d'un processus $N_i(m, \sigma)$, les paramètres m et σ sont connus ou du moins peuvent être très exactement estimés à partir des données passées; il suffit alors de construire des cartes de limites $\bar{\xi}_{\frac{\alpha}{2}} = m + Z_{\frac{\alpha}{2}} \sigma / \sqrt{i}$, $\bar{\xi}_{1-\frac{\alpha}{2}} = m + Z_{1-\frac{\alpha}{2}} \sigma / \sqrt{i}$ pour le contrôle de la moyenne, et de limite $S_{1-\alpha} = \chi_{1-\alpha} \cdot \frac{\sigma}{\sqrt{i}}$ pour le contrôle de l'écart-type, les échantillons prélevés étant d'effectif i .

En pratique, il arrive très souvent que m et σ soient inconnus et il est alors indiqué d'effectuer un contrôle *a posteriori*.

Ayant prélevé K échantillons d'effectif i , on calcule la moyenne \bar{X} et la variance S^2 de la population finie d'effectif iK ainsi obtenue, on construit alors, pour la moyenne, une carte de contrôle de limites

$$\xi_{\frac{\alpha}{2}} = \bar{X} - \sqrt{(K-1)} U_{1-\frac{\alpha}{2}} S, \xi_{1-\frac{\alpha}{2}} = \bar{X} + \sqrt{(K-1)} U_{1-\frac{\alpha}{2}} S,$$

$U_{1-\frac{\alpha}{2}}$ étant le quantile $1 - \alpha$ de la distribution Bêta incomplète $I_u \left(\frac{1}{2}, \frac{Ki-2}{2} \right)$ (voir II.3.4.1), et pour l'écart-type une carte de limite

$$S_{1-\frac{\alpha}{2}} = \sqrt{K} S \sqrt{\tau'_{1-\frac{\alpha}{2}}},$$

$\tau'_{1-\frac{\alpha}{2}}$ étant le quantile $1 - \alpha$ de la Bêta incomplète $I_{\tau'} \left(\frac{i-1}{2}, \frac{i(K-1)}{2} \right)$

La sortie d'un point des limites de contrôle indique absence de contrôle. Le défaut d'une telle méthode est que les points représentatifs ne sont pas indépendants. Une méthode séquentielle plus raffinée sera exposée ultérieurement.

On peut également utiliser une carte où l'on porte les valeurs successives

de $Q = \frac{S_i^2}{K} + \frac{(\xi_i - \bar{X})^2}{K-1}$, avec limite, de contrôle $Q_{\frac{\alpha}{2}}, Q_{1-\frac{\alpha}{2}}$, données par

$$Q_{\frac{\alpha}{2}} = S^2 \cdot \lambda_{\frac{\alpha}{2}}, Q_{1-\frac{\alpha}{2}} = S^2 \lambda_{1-\frac{\alpha}{2}}$$

où $\lambda_{\frac{\alpha}{2}}$ et $\lambda_{1-\frac{\alpha}{2}}$ sont les quantiles $\frac{\alpha}{2}$ et $1 - \frac{\alpha}{2}$, de la distribution Bêta incomplète

$I_{\lambda} \left(\frac{i}{2}, \frac{i(K-1)-1}{2} \right)$ (Voir II.3.2), pour le contrôle de S ,

et une carte où l'on porte les valeurs de

$$Y = \frac{\xi_i - \bar{X}}{S} / \sqrt{(K-1) \left(1 - \frac{S_i^2}{S^2 K} \right)}$$

avec limites de contrôle

$$Y_{\frac{\alpha}{2}} = -\sqrt{\lambda_{\alpha}}, Y_{1-\frac{\alpha}{2}} = +\sqrt{\lambda_{\alpha}}$$

où λ_{α} est le quantile α de la Bêta incomplète $I_{\lambda} \left(\frac{i}{2}, \frac{i(K-1)-1}{2} \right)$, (voir II.3.4.3), pour le contrôle de X .

8. La distribution de $im_2(t) = \Sigma \left(\frac{\xi_i - \bar{X}}{S} \right)^2$ est celle de la « distance » d'un échantillon au centre de la population finie d'où l'échantillon est extrait; elle est l'homologue de la distribution de $\chi^2 = \Sigma \frac{(\xi_i - m)^2}{\sigma^2}$ dans le cas d'une

population infinie. D'un point de vue purement intuitif, elle peut servir de point de départ pour la construction d'un *test visant à déterminer si un échantillon ξ est extrait ou non de la population (\bar{X}, S)* , à savoir : étant donnée la « distance » $\Sigma \left(\frac{\xi_i - \bar{X}}{S} \right)^2$ d'un échantillon observé ξ à la population (\bar{X}, S) , conclure que l'échantillon a été extrait d'une autre population, si $P \left\{ im_2 > \Sigma \left(\frac{\xi_i - \bar{X}}{S} \right)^2 \right\}$ est trop faible. Il conviendrait d'étudier le « pouvoir » d'un tel test contre des « alternatives » spécifiques.

Un tel test serait précieux pour les problèmes de *discrimination*. Étant donnés deux échantillons (X_n, S_n) et (X'_n, S'_n) extraits de deux populations inconnues (m, σ) et (m', σ') , chacun d'effectif n , on se demande si un troisième échantillon $(\bar{\xi}, S_\xi)$ d'effectif i , est extrait de (m, σ) ou de (m', σ') . L'ensemble $(\bar{\xi}, S_\xi)$ et (X_n, S_n) constitue une population finie (\bar{X}, S) et l'ensemble $(\bar{\xi}, S_\xi)$ et (X'_n, S'_n) une population finie (\bar{X}', S') ; le problème se ramène à savoir si $(\bar{\xi}, S_\xi)$ est extrait d'une population finie (\bar{X}, S) provenant de (m, σ) ou si $(\bar{\xi}, S_\xi)$ est extrait d'une population finie (\bar{X}', S') provenant de (m', σ') .

9. On notera que les formules des § II et III sont intéressantes pour *l'étude des séries temporelles*, quand la moyenne et la variance de la série sont inconnues et que la série observée est limitée à un nombre fini de termes. Elles peuvent servir de point de départ à la « dérivation » de nombreuses distributions obtenues dans l'hypothèse où la population de référence est infinie mais qui n'ont pas encore fait l'objet d'investigations lorsque cette population est finie. Elles peuvent enfin être généralisées au cas de population finie à plusieurs variables.

RÉFÉRENCES

- [1] IRWIN, J. O. — « On the frequency distribution of any number of deviates from the mean of a sample from a normal distribution ». *J. R. S. S.*, **92**, 580.
- [2] THOMPSON W. R. — « On a criterion for the rejection of observations ». *A. M. S.*, **6**, 214.
- [3] G. J. LIEBERMAN and G. J. REMIKOFF. — « Sampling plans for inspection by variables ». *J. A. S. A.*, **50**, 457.

ANDRÉ G. LAURENT.
*Institut National de la Statistique
 et Michigan State University.*
