

Développements d'Edgeworth de deux estimateurs d'une proportion de mesures

Frédéric Bertrand¹ et Myriam Maumy¹

¹IRMA, Université Louis Pasteur
Strasbourg, France

SMAI 2007 – Praz sur Arly – 6 juin 2007

Sommaire

- 1 Motivation, notations et résultats préliminaires
- 2 Quelques simulations pour $\hat{\pi}$ et $\tilde{\pi}$
- 3 Développements d'Edgeworth
- 4 Développements de Cornish-Fisher et bootstrap
- 5 Conclusion et perspectives

Motivation

Un des problèmes des laboratoires de contrôle ou de l'industrie chimique, pharmaceutique, agroalimentaire, est l'estimation de la proportion d'une population de mesures se situant entre deux limites.

Le contexte particulier de cet exposé est celui de la validation des procédures analytiques.

Nous supposons qu'une série de mesures suit une loi normale de moyenne μ_M et de variance σ_M^2 .

La proportion des mesures appartenant à l'intervalle $[L, U]$, notée π , est le paramètre d'intérêt.

Par exemple, dans les laboratoires de contrôle de qualité, $[L, U]$ représente l'étendue des valeurs produites acceptables et π est alors la proportion d'unités conformes.

Nous souhaitons déterminer si la méthode de mesure utilisée est conforme aux spécifications des normes internationales, FDA, ISO :

en utilisant cette procédure analytique, a-t-on la garantie que les valeurs observées seront suffisamment proches des vraies valeurs, inconnues, des échantillons ?

Par suffisamment proche, nous entendons que la vraie valeur μ_T appartienne à l'intervalle d'acceptation $[-\lambda, \lambda]$, c'est-à-dire :
 $|X - \mu_T| \leq \lambda.$

La limite d'acceptation λ est fixée a priori par la réglementation en vigueur dans le secteur industriel considéré.

Par exemple, dans le domaine pharmaceutique, λ vaut 15 % pour la règle de la FDA de 2001.

Par garantie nous entendons que $\mathbb{P}[|X - \mu_T| \leq \lambda] \geq \alpha$.

La difficulté provient du fait que nous ne connaissons pas μ_M et σ_M^2 les valeurs des paramètres associés à la méthode de mesures.

L'objectif de la phase de validation d'une procédure analytique est alors de déterminer si, étant donné $\widehat{\mu}_M$ et $\widehat{\sigma}_M^2$ des estimations de μ_M et σ_M^2 , la proportion de mesures qui seront dans l'intervalle d'acceptation est supérieure à une valeur π_0 fixée a priori avec un niveau de confiance de α_0 .

Il n'existe pas de solution exacte à ce problème. Plusieurs auteurs, Resnickoff (1955), Mee (1984) et Mee (1988) ont proposé des procédures approchées pour traiter cette difficulté.

Nous utilisons deux estimateurs possibles pour la proportion π et établissons des développements d'Edgeworth qui permettent de construire **directement** des intervalles de confiance pour π .

Le recours à ces développements se justifie par la **faible taille** des séries de mesures utilisées, souvent comprise **entre 5 et 20**, et par conséquent, par le fait que les **approximations asymptotiques** ne sont plus valides dans ce cas.

Notations

Soit $\{X_i : 1 \leq i \leq n\}$ une série de mesures d'une procédure analytique qui suit une loi normale de moyenne μ_M et de variance σ_M^2 .

Désignons par :

- 1 $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$ la moyenne empirique et
- 2 $s_c^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ la variance empirique corrigée.

Introduisons la proportion π qui se définit par :

$$\pi : = \pi \left(\mu_M, \sigma_M^2 \right) = \mathbb{P}[\mu_T - \lambda \leq X \leq \mu_T + \lambda] \quad (1.1)$$

$$= \Phi \left(\frac{\lambda + \mu_T - \mu_M}{\sigma_M} \right) - \Phi \left(\frac{-\lambda + \mu_T - \mu_M}{\sigma_M} \right), \quad (1.2)$$

où

- 1 Φ désigne la fonction de répartition d'une loi normale centrée réduite,
- 2 $-\lambda, \lambda$ sont les « limites d'acceptation », dont la valeur dépend de la norme utilisée, et
- 3 μ_T représente la « vraie valeur » inconnue de l'échantillon.

Posons $U = \lambda + \mu_T$ et $L = -\lambda + \mu_T$.

π s'écrit alors :

$$\pi(\mu_M, \sigma_M^2) = \Phi\left(\frac{U - \mu_M}{\sigma_M}\right) - \Phi\left(\frac{L - \mu_M}{\sigma_M}\right). \quad (1.3)$$

À partir de (1.3), deux estimateurs de π sont proposés :

- 1 le premier, noté $\hat{\pi}$, est basé sur la méthode du maximum de vraisemblance (MLE)
- 2 le second, noté $\tilde{\pi}$, est sans biais et de variance minimale (MVUE).

Résultats préliminaires

La propriété de l'invariance du maximum de vraisemblance permet d'obtenir l'estimateur $\hat{\pi}$ ainsi :

$$\hat{\pi} := \hat{\pi} \left(\bar{X}, \frac{s_c^2}{k^2} \right) = \Phi \left(k \frac{U - \bar{X}}{s_c} \right) - \Phi \left(k \frac{L - \bar{X}}{s_c} \right) = g_{MLE} \left(\bar{X}, \frac{s_c^2}{k^2} \right) \quad (1.4)$$

où $k^2 = \frac{n}{n-1}$.

Théorème

$\hat{\pi} - \pi$ renormalisé par \sqrt{n} a pour loi limite $\mathcal{N}(0, \sigma_{MLE}^2)$.

Il s'agit d'une application de la « méthode Delta » (cf [6]).

L'expression de l'estimateur $\tilde{\pi}$ (cf [7]) s'écrit :

$$\begin{aligned}\tilde{\pi} &:= \tilde{\pi} \left(\bar{X}, \frac{s_c^2}{k^2} \right) = (\tilde{\pi}_1 - \tilde{\pi}_2) \left(\bar{X}, \frac{s_c^2}{k^2} \right) \\ &= g_{MVUE} \left(\bar{X}, \frac{s_c^2}{k^2} \right)\end{aligned}\quad (1.5)$$

$$\text{où } \tilde{\pi}_i \left(\bar{X}, \frac{s_c^2}{k^2} \right) = \begin{cases} 0 & \text{si } b_i \leq 0 \\ 1 & \text{si } b_i \geq 1 \\ B_{b_i} \left(\frac{n-2}{2}, \frac{n-2}{2} \right) & \text{sinon,} \end{cases}\quad (1.6)$$

$$\text{où } B_{b_i} \left(\frac{n-2}{2}, \frac{n-2}{2} \right) = \frac{\Gamma(n-2)}{\Gamma(\frac{n-2}{2})\Gamma(\frac{n-2}{2})} \int_0^{b_i} u^{\frac{n-4}{2}} (1-u)^{\frac{n-4}{2}} du,$$

$$b_1 = \frac{1}{2} \left(1 + k \frac{\bar{X}-L}{s_c \sqrt{n-1}} \right) \text{ et } b_2 = \frac{1}{2} \left(1 + k \frac{\bar{X}-U}{s_c \sqrt{n-1}} \right).$$

Théorème

$\tilde{\pi} - \pi$ renormalisé par \sqrt{n} a pour loi limite $\mathcal{N}(0, \sigma_{MVUE}^2 = \sigma_{MLE}^2)$.

Sommaire

- 1 Motivation, notations et résultats préliminaires
- 2 Quelques simulations pour $\hat{\pi}$ et $\tilde{\pi}$**
- 3 Développements d'Edgeworth
- 4 Développements de Cornish-Fisher et bootstrap
- 5 Conclusion et perspectives

Quelques simulations pour $\hat{\pi}$ et $\tilde{\pi}$

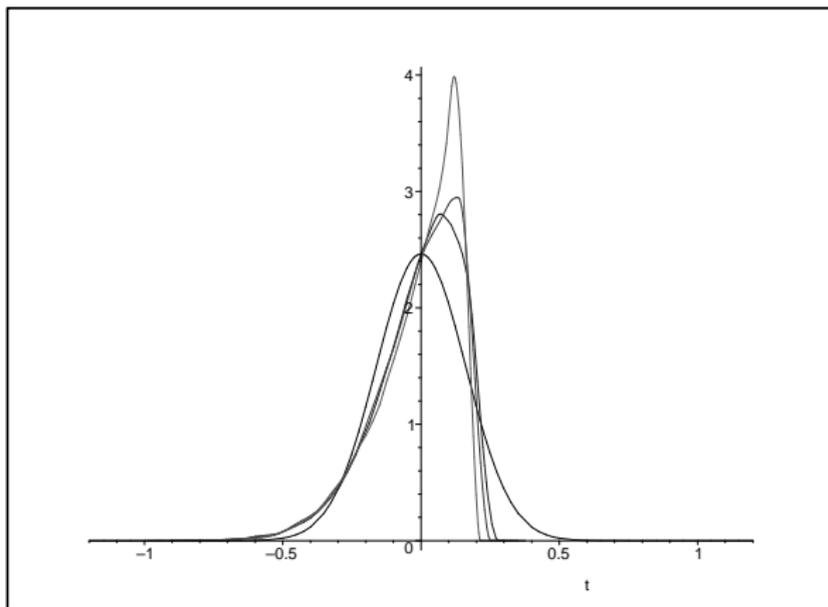
Nous présentons ici deux graphiques :

- 1 le premier pour $\sqrt{n}(\hat{\pi} - \pi)$
- 2 le second pour $\sqrt{n}(\tilde{\pi} - \pi)$

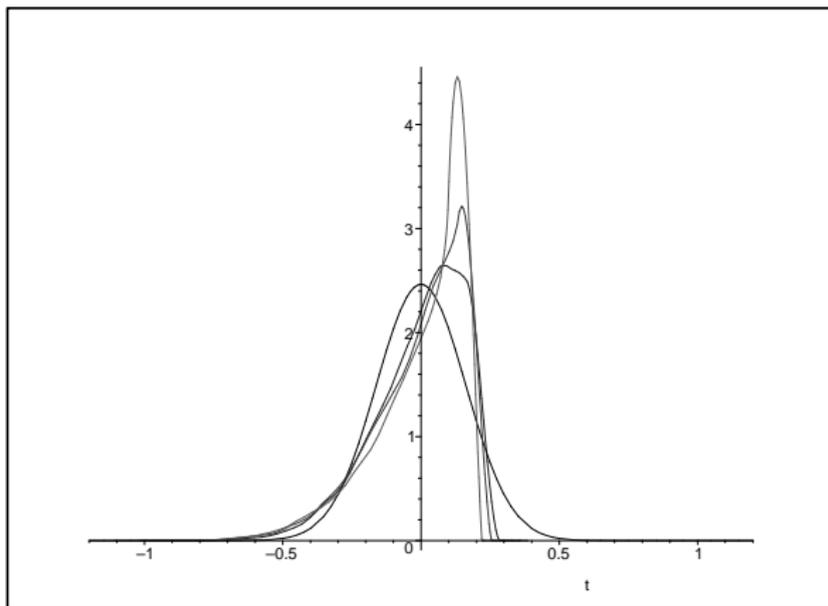
pour $\pi = 0.95$, $\lambda = 0.15$, $\delta = \mu_M - \mu_T = 0$ et par ordre décroissant des maxima des courbes, la densité pour $n = 10$, $n = 15$, $n = 20$ et la loi normale limite lorsque $n \rightarrow +\infty$.

Ces estimations de la densité ont été obtenues à l'aide du logiciel Maple version 10.0 et au moyen d'un noyau d'Epanechnikov et de 10 000 réalisations de chacun des estimateurs.

Densité de $\hat{\pi}$



Densité de $\tilde{\pi}$



Caractéristiques de $\hat{\pi}$ et $\tilde{\pi}$

Nous avons également simulé $r = 10\,000$ réalisations de $\hat{\pi}$ et $\tilde{\pi}$ pour $\pi = 0.95$, $\lambda = 0.15$, $\delta = \mu_M - \mu_T = 0; 0.10$, $n = 5; 10; 15; 20$.

Nous avons calculé certaines caractéristiques numériques de ces estimateurs pour apprécier leur qualité ainsi que leur défaut de normalité :

- 1 Variance = $\frac{1}{r} \cdot \sum_{i=1}^r (x_i - \bar{x})^2$
- 2 MSE = Variance + Biais²,
- 3 Asymétrie = $\frac{1}{r} \cdot \sum_{i=1}^r (x_i - \bar{x})^3 / \text{Variance}^{3/2}$
- 4 Aplatissement = $\frac{1}{r} \cdot \sum_{i=1}^r (x_i - \bar{x})^4 / \text{Variance}^2$.

δ	0			
n	5	10	15	20
Biais $\hat{\pi}$	-0.0056	-0.0030	-0.0033	-0.0017
Variance $\hat{\pi}$	0.0042	0.0022	0.0016	0.0012
MSE $\hat{\pi}$	0.0042	0.0022	0.0016	0.0012
Asymétrie $\hat{\pi}$	-1.5591	-1.1599	-1.0082	-0.9218
Aplatissement $\hat{\pi}$	5.3367	4.2283	3.9795	3.7984
Biais $\tilde{\pi}$	0.0012	0.0001	-0.0011	-0.0006
Variance $\tilde{\pi}$	0.0057	0.0027	0.0018	0.0013
MSE $\tilde{\pi}$	0.0057	0.0027	0.0018	0.0013
Asymétrie $\tilde{\pi}$	-1.7887	-1.3490	-1.0547	-0.9403
Aplatissement $\tilde{\pi}$	5.8690	4.7935	3.8923	3.8286

δ	0.10			
n	5	10	15	20
Biais $\hat{\pi}$	0.0016	0.0012	0.0012	0.0001
Variance $\hat{\pi}$	0.0038	0.0021	0.0014	0.0011
MSE $\hat{\pi}$	0.0038	0.0021	0.0014	0.0011
Asymétrie $\hat{\pi}$	-1.8825	-1.3775	-1.1852	-1.0277
Aplatissement $\hat{\pi}$	7.1694	5.0787	4.7367	4.2284
Biais $\tilde{\pi}$	-0.0011	0.0008	0.0005	0.0001
Variance $\tilde{\pi}$	0.0061	0.0026	0.0018	0.0013
MSE $\tilde{\pi}$	0.0061	0.0026	0.0018	0.0013
Asymétrie $\tilde{\pi}$	-1.7460	-1.3472	-1.1293	-0.9982
Aplatissement $\tilde{\pi}$	5.8653	4.8502	4.2291	3.9678

Ces résultats et, plus particulièrement les valeurs **des coefficients d'asymétrie et d'aplatissement**, nous amènent à des conclusions qui justifient les développements d'Edgeworth proposés ci-après.

Rappelons en effet, que pour une variable aléatoire qui suit une loi normale, le coefficient d'asymétrie vaut 0 et le coefficient d'aplatissement vaut 3.

Sommaire

- 1 Motivation, notations et résultats préliminaires
- 2 Quelques simulations pour $\hat{\pi}$ et $\tilde{\pi}$
- 3 Développements d'Edgeworth**
- 4 Développements de Cornish-Fisher et bootstrap
- 5 Conclusion et perspectives

Soit $\{\mathbf{Z}_n : n \geq 1\}$ une suite de variables aléatoires centrées réduites.

Les quatre premiers cumulants de \mathbf{Z}_n sont :

- 1 $\kappa_{1,n} = \mathbb{E}[\mathbf{Z}_n],$
- 2 $\kappa_{2,n} = \text{Var}[\mathbf{Z}_n],$
- 3 $\kappa_{3,n} = \mathbb{E}[\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]]^3,$
- 4 $\kappa_{4,n} = \mathbb{E}[\mathbf{Z}_n - \mathbb{E}[\mathbf{Z}_n]]^4 - 3(\text{Var}[\mathbf{Z}_n])^2.$

Soit ν un entier compris entre 1 et 4. Le ν -ème cumulant de \mathbf{Z}_n existe ssi $\mathbb{E}[|\mathbf{Z}_n|^\nu] < +\infty.$

Supposons que les quatre premiers cumulants vérifient :

$$\kappa_{1,n} = n^{-1/2}k_{1,2} + O(n^{-3/2}),$$

$$\kappa_{2,n} = 1 + n^{-1}k_{2,2} + O(n^{-3/2}),$$

$$\kappa_{3,n} = n^{-1/2}k_{3,1} + O(n^{-3/2}),$$

$$\kappa_{4,n} = n^{-1}k_{4,1} + O(n^{-3/2}).$$

alors

$$\mathbb{P}[\mathbf{Z}_n \leq x] = \Phi(x) + \left(n^{-1/2}p_1(x) + n^{-1}p_2(x) + \dots \right) \phi(x), \quad (3.1)$$

où $\phi(x) = \Phi'(x)$ est la fonction de densité d'une loi $\mathcal{N}(0, 1)$ et où

$$p_1(x) = - \left(k_{1,2} + \frac{1}{6} k_{3,1} (x^2 - 1) \right), \quad (3.2)$$

$$p_2(x) = -x \left(\frac{1}{2} (k_{2,2} + k_{1,2}^2) + \frac{1}{24} (k_{4,1} + 4k_{1,2}k_{3,1}) (x^2 - 3) + \frac{1}{72} k_{3,1}^2 (x^4 - 10x^2 + 15) \right). \quad (3.3)$$

La détermination explicite des coefficients $k_{i,j}$ peut présenter une difficulté calculatoire.

Développement d'Edgeworth pour $\hat{\pi}$

Nous souhaitons obtenir un développement d'Edgeworth, voir [5], pour $\hat{\pi}$.

Soient $\{X, X_1, \dots, X_n\}$ des variables aléatoires indépendantes et identiquement distribuées de loi $\mathcal{N}(\mu_M, \sigma_M^2)$.

Introduisons $\mathbf{X} = (X, X^2)$ et $\mathbf{X}_s = (X, X^2, X^3, X^4)$.

Les expressions (1.4) et (1.5) permettent de dire que $\hat{\pi}$ et $\tilde{\pi}$ dépendent de \bar{X} et de s_c/k .

Mais nous pouvons aussi dire que $\hat{\pi}$ et $\tilde{\pi}$ sont des fonctions de la moyenne de l'échantillon $\{(X_i, X_i^2) : 1 \leq i \leq n\}$.

Nous allons considérer des développements d'Edgeworth de statistiques qui sont des fonctions de la moyenne d'un échantillon.

Ces développements sont principalement de deux types :

- 1 Le développement de la statistique en supposant sa variance asymptotique connue, en considérant l'échantillon $(\mathbf{X}_i) = (X_i, X_i^2)_{1 \leq i \leq n}$.
- 2 Le développement de la statistique studentisée pour lequel nous estimons également, à distance finie, la variance de la statistique considérée en considérant l'échantillon $(\mathbf{X}_{s,i}) = (X_i, X_i^2, X_i^3, X_i^4)_{1 \leq i \leq n}$.

Maintenant définissons la fonction $F_{0,MLE}$ à valeurs réelles (resp. la fonction studentisée $F_{s,MLE}$ à valeurs réelles) de la façon suivante :

$$F_{0,MLE} : (x, y) \in \mathbb{R}^2 \mapsto \frac{\widetilde{g}_{MLE}(x, y)}{\sigma_{MLE}}, \quad (3.4)$$

$$F_{s,MLE} : (x, y, s, t) \in \mathbb{R}^4 \mapsto \frac{\widetilde{g}_{MLE}(x, y)}{\sqrt{\widehat{\sigma_{MLE}^2}(x, y, z, t)}} \quad (3.5)$$

où $\widehat{\sigma_{MLE}^2}(x, y, z, t) =$

$\sum_{i=1}^d \sum_{j=1}^d (\partial \widetilde{g}_{MLE} / \partial x_i \cdot \partial \widetilde{g}_{MLE} / \partial x_j)(x, y) (\phi(i+j) - \phi(i)\phi(j)),$

avec $\phi(k)$ est la k -ème composante du vecteur (x, y, z, t) et

$\widetilde{g}_{MLE}(x, y) = g_{MLE}(x, y - x^2) - \pi(\mu_M, \sigma_M^2).$

Théorème (3)

Supposons que $\sigma_{MLE} > 0$.

- 1 La fonction $F_{0,MLE}$ (resp. $F_{s,MLE}$) est C^5 au voisinage du point $\mathbb{E}[\mathbf{X}]$ (resp. $\mathbb{E}[\mathbf{X}_s]$),
- 2 $F_{0,MLE}(\mathbb{E}[\mathbf{X}]) = 0$ (resp. $F_{s,MLE}(\mathbb{E}[\mathbf{X}_s]) = 0$),
- 3 $\mathbb{E}[\|\mathbf{X}\|^5] < +\infty$ (resp. $\mathbb{E}[\|\mathbf{X}_s\|^5] < +\infty$)
- 4 la fonction caractéristique χ de \mathbf{X} (resp. χ_s de \mathbf{X}_s) satisfait la condition $\lim_{\|\mathbf{t}\| \rightarrow +\infty} \sup |\chi(\mathbf{t})| < 1$ (resp. $\lim_{\|\mathbf{t}\| \rightarrow +\infty} \sup |\chi_s(\mathbf{t})| < 1$).

Théorème (3 suite)

Alors

$$\begin{aligned} \mathbb{P} \left[\sqrt{n} F_{\bullet, MLE} \left(\bar{X}, s_c^2/k^2 \right) \leq x \right] &= \Phi(x) + n^{-1/2} p_{1, \bullet, MLE}(x) \phi(x) \\ &+ n^{-1} p_{2, \bullet, MLE}(x) \phi(x) \\ &+ O \left(n^{-3/2} \right), \end{aligned} \quad (3.6)$$

uniformément en $x \in \mathbb{R}$ et où $p_{1, \bullet, MLE}$, $p_{2, \bullet, MLE}$ ont été définis précédemment, les coefficients $k_{1,2,MLE}$, $k_{2,2,MLE}$, $k_{3,1,MLE}$ et $k_{4,1,MLE}$ diffèrent d'une statistique à l'autre puisque :

Théorème (3 suite)

$$k_{1,2,MLE} = \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d a_{ij} \mu_{ij},$$

$$k_{3,1,MLE} = \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d a_i a_j a_k \mu_{ijk} + 3 \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d a_i a_j a_{kl} \mu_{ik} \mu_{jl},$$

$$k_{2,2,MLE} = \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d a_i a_{jk} \mu_{ijk} + \frac{1}{2} \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d a_{ij} a_{kl} \mu_{ik} \mu_{jl} +$$

$$\sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d a_i a_{jkl} \mu_{ij} \mu_{kl},$$

Théorème (3 suite)

$$\begin{aligned}
k_{4,1,MLE} = & \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d a_i a_j a_k a_l (\mu_{ijkl} - 3\mu_{ij}\mu_{kl}) + \\
& 12 \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d \sum_{m=1}^d a_i a_j a_k a_l m \mu_{il} \mu_{jkm} \\
& + \frac{3}{2} \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d \sum_{m=1}^d \sum_{o=1}^d a_i a_j a_k a_l a_m a_o (\mu_{ij}\mu_{kl}\mu_{mo} + \\
& 2\mu_{ij}\mu_{km}\mu_{lo} + 4\mu_{ik}\mu_{jl}\mu_{mo} + 16\mu_{ik}\mu_{jm}\mu_{lo}) \\
& + \frac{2}{3} \sum_{i=1}^d \sum_{j=1}^d \sum_{k=1}^d \sum_{l=1}^d \sum_{m=1}^d \sum_{o=1}^d a_i a_j a_k a_l a_m a_o (4\mu_{im}\mu_{kl}\mu_{jo} +
\end{aligned}$$

Théorème (3 suite)

où a est égal à $F_{0,MLE}$ (resp. à $F_{S,MLE}$), en désignant par (i) la i -ème composante d'un vecteur et avec

$$a_{i_1 \dots i_j} = \left(\partial^j / \partial x^{(i_1)} \dots \partial x^{(i_j)} \right) a(\mathbf{X}) |_{\mathbf{X} = \mathbb{E}[\mathbf{X}]}$$

et

$$\mu_{i_1 \dots i_j} = \mathbb{E} \left[(\mathbf{X} - \mu)^{(i_1)} \dots (\mathbf{X} - \mu)^{(i_j)} \right].$$

Développement d'Edgeworth pour $\tilde{\pi}$

De façon similaire, nous obtenons un résultat pour $\tilde{\pi}$.

Nous introduisons la fonction $F_{0,MVUE}$ (ou la version studentisée de cette statistique $F_{s,MVUE}$) définie simplement en remplaçant g_{MLE} par g_{MVUE} dans les expressions (3.4) et (6).

Toutefois un raffinement supplémentaire est nécessaire puisque $\widetilde{g_{MVUE}}(x, y)$ et donc $F_{0,MVUE}(x, y)$ et $F_{s,MVUE}(x, y, s, t)$ dépendent de n (voir [5] pour plus de détails sur la manière de traiter cette difficulté.)

Théorème (4)

Supposons que $\sigma_{MVUE} > 0$.

- 1 La fonction $F_{0,MVUE}$ (resp. $F_{s,MVUE}$) est C^5 au voisinage du point $\mathbb{E}[\mathbf{X}]$ (resp. $\mathbb{E}[\mathbf{X}_s]$),
- 2 $F_{0,MVUE}(\mathbb{E}[\mathbf{X}]) = 0$ (resp. $F_{s,MVUE}(\mathbb{E}[\mathbf{X}_s]) = 0$),
- 3 $\mathbb{E}[\|\mathbf{X}\|^5] < +\infty$ (resp. $\mathbb{E}[\|\mathbf{X}_s\|^5] < +\infty$)
- 4 La fonction caractéristique χ de \mathbf{X} (resp. χ_s de \mathbf{X}_s) satisfait la condition $\lim_{\|\mathbf{t}\| \rightarrow +\infty} \sup |\chi(\mathbf{t})| < 1$ (resp. $\lim_{\|\mathbf{t}\| \rightarrow +\infty} \sup |\chi_s(\mathbf{t})| < 1$).

Théorème (4 suite)

Alors

$$\mathbb{P}\left[\sqrt{n}F_{\bullet, MVUE}\left(\bar{X}, s_c^2/k^2\right) \leq x\right] = \Phi(x) + n^{-1/2}p_{1, \bullet, MVUE}(x)\phi(x) + n^{-1}p_{2, \bullet, MVUE}(x)\phi(x) + O\left(n^{-3/2}\right), \quad (3.7)$$

uniformément en $x \in \mathbb{R}$ et où $p_{1, \bullet, MVUE}$ et $p_{2, \bullet, MVUE}$ ont été définis au Théorème 3.1 et où $k_{1,2, MVUE}$, $k_{3,1, MVUE}$, $k_{2,2, MVUE}$ et $k_{4,1, MVUE}$ sont donnés par :

Théorème (4 suite)

$$k_{1,2,MVUE} = k_{1,2,MLE},$$

$$k_{3,1,MVUE} = k_{3,1,MLE},$$

$$k_{2,2,MVUE} = k_{2,2,MLE} + 2 \sum_{i=1}^d \sum_{j=1}^d b_{0,i} b_{2,j} \mu_{ij},$$

$$k_{4,1,MVUE} = k_{4,1,MLE},$$

et avec les $b_{0,i}$, $1 \leq i \leq d$, et les $b_{2,j}$, $1 \leq j \leq d$, définis dans [5].

Sommaire

- 1 Motivation, notations et résultats préliminaires
- 2 Quelques simulations pour $\hat{\pi}$ et $\tilde{\pi}$
- 3 Développements d'Edgeworth
- 4 Développements de Cornish-Fisher et bootstrap**
- 5 Conclusion et perspectives

Développements de Cornish-Fisher

Nous souhaitons obtenir les deux quantiles de niveau α pour les deux estimateurs dont nous avons obtenu un développement d'Edgeworth qui sont déterminés par :

$$w_{\alpha}^{MLE} = \inf \left\{ x : \mathbb{P} \left[\sqrt{n} F_{\bullet, MLE} \left(\bar{X}, s_c^2/k^2 \right) \leq x \right] \geq \alpha \right\}$$

et

$$w_{\alpha}^{MVUE} = \inf \left\{ x : \mathbb{P} \left[\sqrt{n} F_{\bullet, MVUE} \left(\bar{X}, s_c^2/k^2 \right) \leq x \right] \geq \alpha \right\}.$$

Notons z_α le quantile de niveau α d'une loi normale centrée réduite. Les équations (3.6) et (3.7) permettent d'obtenir un développement de w_α dans les quatre cas qui nous intéressent en fonction de z_α :

$$w_\alpha = z_\alpha + n^{-1/2} p_{11,\bullet,\bullet}(z_\alpha) + n^{-1} p_{21,\bullet,\bullet}(z_\alpha) + \dots + n^{-j/2} p_{j1,\bullet,\bullet}(z_\alpha) + \dots \quad (4.1)$$

où $p_{11,\bullet,\bullet}(x) = -p_{1,\bullet,\bullet}(x)$ et $p_{21,\bullet,\bullet}(x) = p_{1,\bullet,\bullet}(x)p'_{1,\bullet,\bullet}(x) - \frac{1}{2}xp_{1,\bullet,\bullet}^2(x) - p_{2,\bullet,\bullet}(x)$ où $p_{1,\bullet,\bullet}$ et $p_{2,\bullet,\bullet}$ ont été définis ci-dessus.

Utilisation de techniques de bootstrap

La qualité d'intervalles de confiance obtenus par la même méthode n'est pas nécessairement semblable si nous utilisons un intervalle de confiance unilatéral ou un bilatéral.

Les procédures de Mee et de Resnikoff permettent de construire des intervalles de confiance bilatéraux.

L'obtention d'intervalles de confiance unilatéraux pour la proportion de mesures conformes est la préoccupation principale dans le domaine de la validation de procédures analytiques.

Nous proposons des intervalles de confiance pour π de niveau $100(1 - \alpha) \%$:

- 1 **unilatéraux** ou
- 2 **bilatéraux**.

Ces intervalles sont associés aux développements de Cornish-Fisher des estimateurs $\hat{\pi}$ et $\tilde{\pi}$.

Théorème

*Les intervalles de confiance construits à l'aide du développement de Cornish-Fisher obtenu pour $\hat{\pi}$ ci-dessus et d'une technique de bootstrap de type Student ou BC_a sont **exacts à l'ordre deux**.*

Théorème

*Les intervalles de confiance construits à l'aide du développement de Cornish-Fisher obtenu pour $\tilde{\pi}$ ci-dessus et d'une technique de bootstrap de type Student ou BC_a sont **exacts à l'ordre deux**.*

Sommaire

- 1 Motivation, notations et résultats préliminaires
- 2 Quelques simulations pour $\hat{\pi}$ et $\tilde{\pi}$
- 3 Développements d'Edgeworth
- 4 Développements de Cornish-Fisher et bootstrap
- 5 Conclusion et perspectives**

Conclusion et perspectives

Ce travail a permis d'établir les développements d'Edgeworth de deux estimateurs d'une proportion de mesures.

Ces développements doivent maintenant être testés sur des données réelles provenant des secteurs d'applications évoqués au début de cet exposé.

Nous étudions actuellement l'extension de ces résultats à des plans d'échantillonnage plus complexe : modèles emboîtés, mixtes ainsi qu'à l'utilisation d'autres lois d'erreurs que la loi normale.

Remerciements

Les auteurs tiennent à remercier Bruno BOULANGER et Walthère DEWÉ, E. LILLY, Mont-Saint-Guibert (Belgique), pour leur avoir proposé un sujet si intéressant à explorer.

Merci pour votre attention.

-  T.L. BOULLION, G.C. CASCIO ET J.P. KEATING, Comparison of estimators of the fraction defective in the normal distribution, *Communications in Statistics : Theory and Methods*, 14, 7 (1985), 1511-1529.
-  BERTRAND, F. ET MAUMY, M., Développements d'Edgeworth de deux estimateurs d'une proportion de mesures, *Comptes rendus de l'Académie des sciences, Série I*, Soumis.
-  A. GUILLOU, Efficient weighted bootstraps for the mean, *Journal of Statistical Planning and Inference*, 77 (1999), 11-35.
-  A. GUILLOU, Weighted bootstraps for the variance, *Journal of Statistical Planning and Inference*, 81 (1999), 113-120. 

-  P. HALL, *The Bootstrap and Edgeworth Expansion*, Springer, New-York, 1992.
-  R. IASNOGORODSKI ET H. LHÉRITIER, *Théorie de l'estimation ponctuelle paramétrique*, EDP Sciences, Les Ulis, 2003.
-  A.N. KOLMOGOROV, Unbiased estimates, *Izvestiya Akademii Nauk SSSR. Seriya Matematičeskaya*, 14 (1950), 303-326.
-  R.W. MEE, Estimation of the percentage of a normal distribution lying outside a specified interval, *Communications in Statistics : Theory and Methods*, 17 (1988), 1465-1479.

-  G.J. RESNIKOFF, Interval estimation of proportion defective in sampling inspection by variables. *Tech. Report, 21*, Applied Math. and Stat. Lab., Stanford Univ., Stanford, CA (1955).
-  J. SHAO AND D. TU, *The Jackknife and Bootstrap*, Springer, New York, 1996.