



**HAL**  
open science

## Développement d'Edgeworth de deux estimateurs d'une proportion de mesures

Frédéric Bertrand, Myriam Maumy

► **To cite this version:**

Frédéric Bertrand, Myriam Maumy. Développement d'Edgeworth de deux estimateurs d'une proportion de mesures. 39ème Journées de statistique de la SFdS, Jun 2007, Angers, France. pp.83. hal-00160187

**HAL Id: hal-00160187**

**<https://hal.science/hal-00160187>**

Submitted on 5 Jul 2007

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

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

# DÉVELOPPEMENTS D'EDGEWORTH DE DEUX ESTIMATEURS D'UNE PROPORTION DE MESURES

Frédéric Bertrand & Myriam Maumy

*Institut de Recherche Mathématique Avancée, Université Louis Pasteur.*

Email : fbertran@math.u-strasbg.fr, mmaumy@math.u-strasbg.fr

## Résumé

Nous établissons des développements d'Edgeworth de deux estimateurs d'une proportion de mesures  $\pi$ . Le premier est construit à partir de la méthode du maximum de vraisemblance et le second à l'aide des estimateurs sans biais et de variance minimale. Ces résultats sont motivés par leurs applications dans les laboratoires de contrôle ou dans l'industrie chimique, pharmaceutique, agroalimentaire, etc.

*Mots-clefs* : *Développements d'Edgeworth, Développements de Cornish-Fisher, Estimation par maximum de vraisemblance, Estimation sans biais de variance minimale, Contrôle de qualité.*

## Abstract

**Edgeworth expansions of two point estimators of the proportion of measures.**

We establish Edgeworth expansions for two point estimators of a proportion of experimental results  $\pi$ . The first one is based on the maximum likelihood method and the second one on the minimum variance unbiased estimators. These results are motivated by their counterparts in the official quality laboratories and in the industries of various sectors, namely : chemistry, pharmacy, food processing, etc.

*Keywords* : *Edgeworth Expansions, Cornish-Fisher Expansions, Maximum likelihood estimator, Minimum variance unbiased estimator, Quality control.*

## 1 Motivation

Un des problèmes des laboratoires de contrôle ou de l'industrie chimique, pharmaceutique, agroalimentaire, etc, est l'estimation de la proportion d'une population de mesures se situant entre deux limites. La littérature est relativement abondante sur ce sujet : citons par exemple les articles de Boullion, Cascio, Keating (1985) et de Mee (1988) pour de plus amples détails. Supposons qu'une série de mesures suive une loi normale de moyenne  $\mu_M$  et de variance  $\sigma_M^2$  et que la proportion des mesures appartenant à l'intervalle  $[L, U]$  notée  $\pi$ , est le paramètre d'intérêt. Par exemple, dans les laboratoires de contrôle,  $[L, U]$  représente l'étendue des valeurs produites acceptables et  $\pi$  est alors la proportion d'unités conformes. Dans le contexte qui nous intéresse, la moyenne  $\mu_M$  et la variance  $\sigma_M^2$  de l'ensemble des mesures prises sont supposées connues. Nous proposons ici deux estimateurs possibles

pour la proportion  $\pi$  et établissons des développements d'Edgeworth qui permettent de construire des intervalles de confiance. 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.

## 2 Notations

Soit  $\{X_i : 1 \leq i \leq n\}$  une série de mesures d'une procédure analytique qui suit une loi normale de moyenne  $\mu_M$  et de variance  $\sigma_M^2$ . Désignons par  $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  la moyenne empirique et  $s_c^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$  la variance empirique corrigée. Introduisons la proportion  $\pi$  qui se définit par :

$$\pi := \pi(\mu_M, \sigma_M^2) = \mathbb{P}[\mu_T - \lambda \leq X \leq \mu_T + \lambda] = \Phi\left(\frac{\lambda + \mu_T - \mu_M}{\sigma_M}\right) - \Phi\left(\frac{-\lambda + \mu_T - \mu_M}{\sigma_M}\right), \quad (1)$$

où  $\Phi$  désigne la fonction de répartition d'une loi normale centrée réduite,  $-\lambda, \lambda$  sont les « limites d'acceptation », dont la valeur dépend de la norme utilisée, et  $\mu_T$  représente la « vraie valeur » connue de la moyenne de l'échantillon. Posons  $U = \lambda + \mu_T$  et  $L = -\lambda + \mu_T$ .  $\pi$  s'écrit alors :

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

À partir de (2), deux estimateurs de la proportion  $\pi$  sont proposés : le premier, noté  $\hat{\pi}$ , est basé sur la méthode du maximum de vraisemblance (MLE) et le second, noté  $\tilde{\pi}$ , est sans biais et de variance minimale (MVUE).

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

$$\hat{\pi} := \hat{\pi}\left(\bar{X}, \frac{s_c^2}{k^2}\right) = \Phi\left(k \frac{U - \bar{X}}{s_c}\right) - \Phi\left(k \frac{L - \bar{X}}{s_c}\right) = g_{MLE}\left(\bar{X}, \frac{s_c^2}{k^2}\right) \quad \text{où } k^2 = \frac{n}{n-1}. \quad (3)$$

En appliquant la méthode delta exposée, par exemple dans le livre de Iasnogorodski et Lhéritier (2003), nous établissons le théorème suivant :

**Théorème 2.1**  $\hat{\pi} - \pi$  renormalisé par  $\sqrt{n}$  est asymptotiquement gaussien de variance  $\sigma_{MLE}^2$ .

L'expression de l'estimateur  $\tilde{\pi}$ , donnée dans l'article de Kolmogorov (1950), s'écrit :

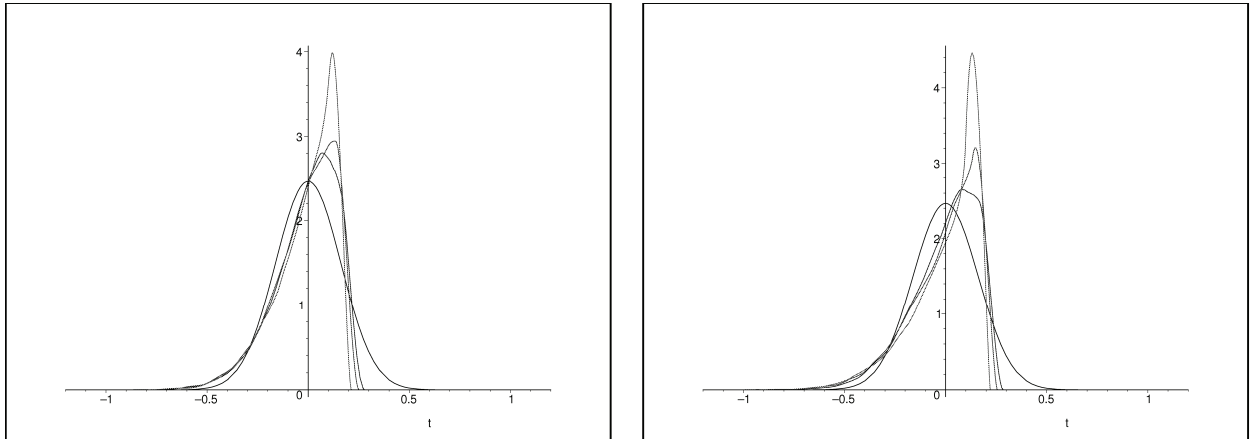
$$\tilde{\pi} := \tilde{\pi}\left(\bar{X}, \frac{s_c^2}{k^2}\right) = (\tilde{\pi}_1 - \tilde{\pi}_2)\left(\bar{X}, \frac{s_c^2}{k^2}\right) \quad \text{où } \tilde{\pi}_i\left(\bar{X}, \frac{s_c^2}{k^2}\right) = \begin{cases} 0 & \text{si } b_i \leq 0 \\ 1 & \text{si } b_i \geq 1 \\ B_{b_i}\left(\frac{n-2}{2}, \frac{n-2}{2}\right) & \text{sinon,} \end{cases} \quad (4)$$

où  $B_{b_i}(\frac{n-2}{2}, \frac{n-2}{2}) = \frac{\Gamma(n-2)}{\Gamma(\frac{n-2}{2})\Gamma(\frac{n-2}{2})} \int_0^{b_i} u^{\frac{n-4}{2}} (1-u)^{\frac{n-4}{2}} du$ ,  $b_1 = \frac{1}{2} \left(1 + k \frac{\bar{X}-L}{s_c \sqrt{n-1}}\right)$  et  $b_2 = \frac{1}{2} \left(1 + k \frac{\bar{X}-U}{s_c \sqrt{n-1}}\right)$ .

**Théorème 2.2**  $\tilde{\pi} - \pi$  renormalisé par  $\sqrt{n}$  est asymptotiquement gaussien de variance  $\sigma_{MVUE}^2 = \sigma_{MLE}^2$ .

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

Nous présentons ici deux graphiques : le premier pour  $\sqrt{n}(\hat{\pi} - \pi)$  et le second pour  $\sqrt{n}(\tilde{\pi} - \pi)$ , où  $\pi = 0.95$ ,  $\lambda = 0.15$ ,  $\delta = \mu_M - \mu_T = 0$  et par ordre décroissant des maxima des courbes, la densité pour  $n = 10$ ,  $n = 15$ ,  $n = 20$  et la loi normale limite lorsque  $n$  tend vers l'infini. 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.



Nous avons également simulé 5 000 réalisations de  $\hat{\pi}$  et  $\tilde{\pi}$  pour  $\pi = 0.95$ ,  $\lambda = 0.15$ ,  $\delta = \mu_M - \mu_T = 0; 0.10$ ,  $n = 5; 10; 20$ . Le tableau ci-dessous résume les résultats obtenus, en rappelant que :

$$\text{Asymétrie} = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^3 / \text{Variance}^{3/2}$$

et

$$\text{Aplatissement} = \frac{1}{n-1} \cdot \sum_{i=1}^n (X_i - \bar{X})^4 / \text{Variance}^2.$$

$\delta$	0			0.10		
$n$	5	10	20	5	10	20
Biais $\hat{\pi}$	-0.0056	-0.0030	0.0017	0.0006	0.0005	0.0003
Variance $\hat{\pi}$	0.0042	0.0021	0.0012	0.0057	0.0027	0.0014
MSE $\hat{\pi}$	0.0042	0.0022	0.0012	0.0057	0.0027	0.0014
Asymétrie $\hat{\pi}$	-1.5591	-1.1598	-0.9218	-1.7346	-1.3491	0.9778
Aplatissement $\hat{\pi}$	5.3362	4.2279	3.7981	5.5692	4.7208	3.8514
Biais $\tilde{\pi}$	0.0012	0.0001	-0.0006	-0.0011	-0.0008	0.0001
Variance $\tilde{\pi}$	0.0057	0.0027	0.0013	0.0061	0.0026	0.0013
MSE $\tilde{\pi}$	0.0057	0.0027	0.0013	0.0061	0.0026	0.0013
Asymétrie $\tilde{\pi}$	-1.7886	-1.3489	-0.9403	-1.7459	-1.3471	-0.9982
Aplatissement $\tilde{\pi}$	5.8684	4.7930	3.8283	5.8648	4.8497	3.9674

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.

## 4 Développements d'Edgeworth pour $\hat{\pi}$ et $\tilde{\pi}$

Afin d'obtenir un développement d'Edgeworth pour  $\hat{\pi}$ , nous utilisons les résultats et les notations du livre de Hall (1992). Soit  $\{X, X_1, \dots, X_n\}$  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 (3) et (4) 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 l'échantillon  $\{(X_i, X_i^2) : 1 \leq i \leq n\}$ .

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

$$F_{0,MLE} : (x, y) \in \mathbb{R}^2 \mapsto \widetilde{g_{MLE}}(x, y) / \sigma_{MLE} \in \mathbb{R} , \quad (5)$$

$$F_{s,MLE} : (x, y, s, t) \in \mathbb{R}^4 \mapsto \widetilde{g_{MLE}}(x, y) / \sqrt{\widehat{\sigma_{MLE}^2}(x, y, z, t)} \in \mathbb{R} \quad (6)$$

où  $\widehat{\sigma_{MLE}^2}(x, y, z, t) = \sum_{i=1}^d \sum_{j=1}^d (\partial \widetilde{g_{MLE}} / \partial x_i \cdot \partial \widetilde{g_{MLE}} / \partial x_j)(x, y) (\phi(i+j) - \phi(i)\phi(j))$ ,  $\phi(k)$  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 4.1** *Supposons que la fonction  $F_{0,MLE}$  (resp.  $F_{s,MLE}$ ) soit  $C^5$  au voisinage du point  $\mathbb{E}[\mathbf{X}]$  (resp.  $\mathbb{E}[\mathbf{X}_s]$ ), que  $F_{0,MLE}(\mathbb{E}[\mathbf{X}]) = 0$  (resp.  $F_{s,MLE}(\mathbb{E}[\mathbf{X}_s]) = 0$ ), que  $\mathbb{E}[\|\mathbf{X}\|^5] < +\infty$  (resp.  $\mathbb{E}[\|\mathbf{X}_s\|^5] < +\infty$ ) et que la fonction caractéristique  $\chi$  de  $\mathbf{X}$  (resp.*

$\chi_s$  de  $\mathbf{X}_s$ ) satisfait la condition  $\lim_{\|\mathbf{t}\| \rightarrow +\infty} \sup |\chi(\mathbf{t})| < 1$  (resp.  $\lim_{\|\mathbf{t}\| \rightarrow +\infty} \sup |\chi_s(\mathbf{t})| < 1$ ). Supposons que  $\sigma_{MLE} > 0$ . Alors

$$\mathbb{P}[\sqrt{n}F_{\bullet,MLE}(\bar{X}, s_c^2/k^2) \leq x] = \Phi(x) + n^{-1/2}p_{1,\bullet,MLE}(x)\phi(x) + n^{-1}p_{2,\bullet,MLE}(x)\phi(x) + O(n^{-3/2}), \quad (7)$$

uniformément en  $x \in \mathbb{R}$  et où  $p_{1,\bullet,MLE}$ ,  $p_{2,\bullet,MLE}$  sont définis par les équations (2.24) et (2.25) page 48 dans le livre de Hall (1992), les coefficients  $k_{1,2,MLE}$ ,  $k_{2,2,MLE}$ ,  $k_{3,1,MLE}$  et  $k_{4,1,MLE}$  sont définis dans l'article de Bertrand et Maumy (2007).

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 en remplaçant  $g_{MLE}$  par  $g_{MVUE}$  dans les expressions (5) et (6). Toutefois un raffinement supplémentaire est nécessaire puisque  $\widetilde{g_{MVUE}}(x, y)$  et de ce fait  $F_{0,MVUE}(x, y)$  et  $F_{s,MVUE}(x, y, s, t)$  dépendent de  $n$ . Nous renvoyons au livre de Hall (1992) pour plus de détails.

**Théorème 4.2** *Supposons que la fonction  $F_{0,MVUE}$  (resp.  $F_{s,MVUE}$ ) soit  $C^5$  au voisinage du point  $\mathbb{E}[\mathbf{X}]$  (resp.  $\mathbb{E}[\mathbf{X}_s]$ ), que  $F_{0,MVUE}(\mathbb{E}[\mathbf{X}]) = 0$  (resp.  $F_{s,MVUE}(\mathbb{E}[\mathbf{X}_s]) = 0$ ), que  $\mathbb{E}[|\mathbf{X}|^5] < +\infty$  (resp.  $\mathbb{E}[|\mathbf{X}_s|^5] < +\infty$ ) et que la fonction caractéristique  $\chi$  de  $\mathbf{X}$  (resp.  $\chi_s$  de  $\mathbf{X}_s$ ) satisfait la condition  $\lim_{\|\mathbf{t}\| \rightarrow +\infty} \sup |\chi(\mathbf{t})| < 1$  (resp.  $\lim_{\|\mathbf{t}\| \rightarrow +\infty} \sup |\chi_s(\mathbf{t})| < 1$ ). Supposons que  $\sigma_{MVUE} > 0$ . Alors*

$$\mathbb{P}[\sqrt{n}F_{\bullet,MVUE}(\bar{X}, s_c^2/k^2) \leq x] = \Phi(x) + n^{-1/2}p_{1,\bullet,MVUE}(x)\phi(x) + n^{-1}p_{2,\bullet,MVUE}(x)\phi(x) + O(n^{-3/2}), \quad (8)$$

uniformément en  $x \in \mathbb{R}$  et où  $p_{1,\bullet,MVUE}$  et  $p_{2,\bullet,MVUE}$  ont été définis au Théorème 4.1 et où  $k_{1,2,MVUE}$ ,  $k_{3,1,MVUE}$ ,  $k_{2,2,MVUE}$  et  $k_{4,1,MVUE}$  sont donnés par :

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

$$k_{3,1,MVUE} = k_{3,1,MLE}, \quad k_{4,1,MVUE} = k_{4,1,MLE},$$

et où les  $b_{0,i}$ ,  $1 \leq i \leq d$ , et les  $b_{2,j}$ ,  $1 \leq j \leq d$ , sont définis par l'équation (2.38) page 58 dans le livre de Hall (1992).

## 5 Développements de Cornish-Fisher

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

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

et

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

Notons  $z_\alpha$  le quantile de niveau  $\alpha$  d'une loi normale centrée réduite. Les équations (7) et (8) permettent d'obtenir un développement de  $w_\alpha$  dans les quatre cas qui nous intéressent en fonction de  $z_\alpha$  :

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

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.

## 6 Conclusion et perspectives

Nous avons donc établi 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 dans la section 1. De tels résultats seront présentés ultérieurement, ainsi que l'utilisation de techniques bootstrap, afin de ne pas avoir à supposer les paramètres  $\mu_M$ ,  $\mu_T$  et  $\sigma_M^2$  connus.

## Bibliographie

- [1] Bertrand, F. et Maumy, M. (2007), Développements d'Edgeworth de deux estimateurs d'une proportion de mesures, *Comptes rendus de l'Académie des sciences, Série I*, Soumis.
- [2] Boullion, T.L., Cascio, G.C., Keating, J.P. (1985), Comparison of estimators of the fraction defective in the normal distribution, *Communications in Statistics : Theory and Methods*, 14, 7, 1511-1529.
- [3] Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, Springer, New-York.
- [4] Iasnogorodski, R. et Lhéritier, H. (2003), *Théorie de l'estimation ponctuelle paramétrique*, EDP Sciences, Les Ulis.
- [5] Kolmogorov, A.N. (1950), Unbiased estimates, *Izvestiya Akademii Nauk SSSR. Seriya Matematicheskaya*, 14, 303-326.
- [6] Mee, R.W. (1988), Estimation of the percentage of a normal distribution lying outside a specified interval, *Communications in Statistics : Theory and Methods*, 17, 1465-1479.