Séminaire de probabilités (Strasbourg)

Iain M. Johnstone
Brenda MacGibbon

# Une mesure d'information caractérisant la loi de Poisson

*Séminaire de probabilités (Strasbourg)*, tome 21 (1987), p. 563-573
<http://www.numdam.org/item?id=SPS_1987__21__563_0>

# UNE MESURE D'INFORMATION CARACTERISANT LA LOI DE POISSON

Iain M. Johnstone
Department of Statistics
Stanford University
Palo Alto, U.S.A.

et

Brenda MacGibbon
Département de mathématiques
et d'informatique
Université de Sherbrooke
Sherbrooke (Québec)
Canada J1K 2R1

## Résumé

On définit une mesure d'information analogue à l'information de Fisher pour les mesures de probabilité dont le support est l'ensemble des entiers non-négatifs. Cette information possède des propriétés similaires à l'information de Fisher et donne deux caractérisations différentes de la loi de Poisson. Ceci conduit à une caractérisation des suites de mesures de probabilité ayant comme points d'accumulation des lois de Poisson.

## 0. Introduction

L'idée d'utiliser des mesures d'information pour démontrer des théorèmes limites en probabilité semble due à Linnik. Dans [9] il donne une démonstration du théorème de la limite centrale dans le cas où la condition de Lindeberg est satisfaite, en utilisant la fonction d'information suivante:

$$I(X) = -\left(\int_{-\infty}^{+\infty} p(x)\log p(x)\,dx + \frac{1}{2}\log \int_{-\infty}^{+\infty} x^2 p(x)\,dx\right)$$

où le premier terme est l'entropie de Shannon d'une loi continue ([16], ch. III, section 19). Cette fonction a aussi été utilisée par McKean dans [12]. En utilisant des mesures d'information, Renyi [14] a démontré l'ergodicité des chaînes de Markov homogènes dans le cas fini et D.G. Kendall [7] a donné la démonstration dans le cas dénombrable. Indépendamment, L.D. Brown [3] a donné une démonstration du théorème de la limite centrale classique basée sur l'information de Fisher des suites de sommes normalisées. Cette démonstration utilise une simple propriété des fonctions propres des polynômes d'Hermite. Barron [1] a généralisé les résultats de Brown pour démontrer un théorème de la limite centrale pour les densités en établissant la convergence monotone au sens de l'entropie relative. Ici nous définissons une notion analogue à l'information de Fisher pour les mesures de probabilité dont le support est l'ensemble des entiers non-négatifs. Nous utilisons une propriété des fonctions propres des polynômes de Poisson-Charlier pour fournir une démonstration unifiée des "lois des petits nombres".

---

Une partie de ce travail a été complété lorsque les auteurs étaient membres du MSRI à Berkeley. Les auteurs sont reconnaissants aux NSF et CRSNG.

L'information de Fisher d'une variable aléatoire Z de fonction de densité f absolument continue, est définie par $\mathbb{E}[(f'(z)/f(z))^2]$. On considère ici la classe $\mathcal{P}_0$ des lois dont le support est $\mathbb{N}$ (l'ensemble des entiers non-négatifs). On définit une mesure d'information discrète I analogue à l'information de Fisher pour l'ensemble $\mathcal{X}_0$ des variables aléatoire X dont la loi de X est élément de $\mathcal{P}_0$. On montre que I a des propriétés similaires à celles de l'information de Fisher et que I peut être utilisée pour donner deux caractérisations différentes de la loi de Poisson:

1) $\forall\, X \in \mathcal{X}_0$, $I(X) \geq (\mathrm{var}(X))^{-1}$ avec égalité si et seulement si X a une loi de Poisson.

2) $\forall\, X, Y \in \mathcal{X}_0$ (X et Y indépendantes), $\frac{I(X) + I(Y)}{4} - I(X*Y) \geq 0$, où $X*Y$ est la convolution de X et Y. De plus, $\frac{I(X)}{2} = I(X*X)$ si et seulement si X a une loi de Poisson.

Cette deuxième propriété nous permet de donner une condition nécessaire et suffisante pour qu'une suite de lois ait une loi de Poisson comme limite.

## 1. Mesure d'information

*Définition 1.1: Pour chaque $P \in \mathcal{P}_0$ soit:*

$$I(P) = \sum_{x=0}^{\infty} \frac{(p_x - p_{x-1})^2}{p_x} \quad \text{où} \quad p_x = P(x) \quad \text{et} \quad p_{-1} = 0\,.$$

Par analogie, si $X \in \mathcal{X}_0$ et a comme loi P, alors I(X) est définit comme étant égal à I(P). Notons que I(P) peut être infinie; par exemple pour $p_x = c\, \exp(-2^x)$ où c est la constante de normalisation.

*Lemme 1.2: Si $X \in \mathcal{X}_0$, alors $I(X) \geq (\mathrm{var}(X))^{-1}$ avec égalité si et seulement si X a une loi de Poisson.*

Démonstration: Si $\mathbb{E}(X) = \infty$, l'inégalité est vraie. Soit $\lambda = \sum x p_x$, $0 < \lambda < \infty$. Il vient en vertu de l'inégalité de Cauchy-Schwarz

$$1 = [\sum_x (p_x - p_{x-1})(x-\lambda)]^2 \leq I(X) \sum_x (x-\lambda)^2\, p_x\,, \tag{1}$$

avec égalité si et seulement si:

$$p_x - p_{x-1} = c(1-\frac{x}{\lambda})\, p_x\,, \quad x \geq 0\,.$$

Puisque $p_0 > 0$ il s'en suit que $c = 1$, et on a égalité (dans $\mathcal{X}_0$) si et seulement si P est une loi de Poisson de paramètre $\lambda$.

Les propriétés suivantes de $I$ sont entièrement semblables à celles de l'information de Fisher (*cf.* Huber [6], §4.4).

*Lemme 1.3*: *Soit* $\mathcal{B} = \{\phi : \mathbb{N} \to \mathbb{R},\ \phi \text{ bornée}\}$. *Si* $P \in \mathcal{P}_0$ *alors*

$$I(P) = \sup_{\phi\in\mathcal{B}} \left[\frac{[\sum_x(\phi_{x+1} - \phi_x)p_x]^2}{\sum_x \phi_x^2 p_x}\right].$$

Démonstration: Introduisons l'opérateur $T_P(\phi)$ suivant:

$$T_P(\phi) = \sum_x \phi_x(p_x - p_{x-1}) = \sum_x (\phi_x - \phi_{x+1})p_x ,$$

lequel est bien défini sur $\mathcal{B}$. En écrivant $\|\phi\|_P^2 = \sum_x \phi_x^2 p_x$, il est clair que $T_P^2(\phi) \le I(P)\|\phi\|_P^2$ pour tout $\phi \in \mathcal{B}$.

Cependant, pour la suite $\phi_n(x) = \frac{p_x - p_{x-1}}{p_x} I\{0 \le x \le n\}$, où $I$ représente la fonction indicatrice, nous avons $\|\phi_n\|_P^2 = \sum_{x=0}^{n} \frac{(p_x - p_{x-1})^2}{p_x}$,

$$T_P^2(\phi_n) = \left[\sum_{x=0}^{n} \frac{(p_x - p_{x-1})^2}{p_x}\right]^2 ,$$

donc $T_P^2(\phi_n)\|\phi_n\|_P^{-2} \to I(P)$ quand $n\to\infty$. Cette preuve montre aussi que lorsque $I(P) < \infty$, cette quantité est alors égale au carré de la norme de $T_P$ dans l'espace dual de $L^2(\mathbb{N},P)$.

*Corollaire 1.4*: *$I(P)$ est semi-continue inférieurement sur $\mathcal{P}_0$; c'est-à-dire que $P_n \xrightarrow{\mathcal{D}} P \in \mathcal{P}_0$ implique que $I(P) \le \liminf I(P_n)$.*

*Lemme 1.5*: *$I(P)$ est convexe sur $\mathcal{P}_0$.*

Démonstration: Puisque $I(P)$ est un supremum, il suffit de vérifier que la fonction $t \to \frac{T_{P_t}^2(\phi)}{\|\phi\|_{P_t}^2}$ est convexe, où $P_t = (1-t)P_0 + t\,P_1$ pour $P_0, P_1 \in \mathcal{P}_0$, $t\in[0,1]$. Puisque les deux fonction $t \to T_{P_t}(\phi)$ et $t \to \|\phi\|_{P_t}^2$ sont linéaires en $t$ alors la convexité suit (*cf.* Huber [6], lemme 4.4).

*Remarque 1.6:* $I(P)$ n'a pas la même interprétation asymptotique que celle de l'information de Fisher dans les problèmes qui concernent les paramètres de position, mais dans le cas des lois exponentielles discrètes, elle joue le même rôle que la borne inférieure de Cramer-Rao. En effet, soit $P_\theta(x) = \theta^x t_x \phi(\theta)$, la fonction de densité d'une famille exponentielle discrète sur $N$. Grâce au théorème de Rao-Blackwell, $d_0(x) = \frac{t_{x-1}}{t_x}$ est l'estimateur non-biaisé de $\theta$ à variance minimale.

Or,

$$\mathrm{var}(d_0) = \sum_x \left(\frac{\theta t_x - t_{x-1}}{\theta t_x}\right)^2 \theta^2 P_\theta(x) = \theta^2\, \mathbb{E}_\theta \left(1 - \frac{P_\theta(x-1)}{P_\theta(x)}\right)^2 = \theta^2 I(P_\theta)$$

Donc, si $T(x)$ est un estimateur non-biaisé de $\theta$, on a $\mathrm{var}_\theta(T) \geq \theta^2 I(P_\theta)$.

*Remarque 1.7:* Un principe d'incertitude.

L'équation (1) du lemme 1.2 peut s'écrire: $I(X)\sum(x-\lambda)^2 p_x \geq 1$. Nous savons (voir Dynkin et Jushkewitsch [5]) que si nous définissons la matrice tridiagonale $Q = \{q_{xy}\}_{x,y \in N}$ par:

$$q_{x,x+1} = 1,\ q_{x,x-1} = \frac{p_{x-1}}{p_x},\ q_{xx} = -1 - \frac{p_{x-1}}{p_x} \quad \text{et} \quad q_{xy} = 0 \ \text{sinon pour } x > 0$$

et par:

$$q_{xx} = -1,\ q_{x,x+1} = +1 \quad \text{et} \quad q_{xy} = 0 \quad \text{sinon pour } x = 0$$

alors: 1) $Q$ est le générateur infinitésimal d'un processus irréductible de naissance et de mort défini sur $\mathbb{N}$, noté $X_t$, et

2) $P(x) = p_x$ $x \in \mathbb{N}$ représente la distribution stationnaire associée à ce processus.

Avec les modifications adéquates permettant d'envisager d'éventuelles "explosions" comme il est indiqué dans Dynkin et Jushkewitsch [5] ou Johnstone [8] ces taux correspondent à un processus markovien continu sur $\mathbb{N}\cup\{\infty\}$.

Si nous désignons par $V(x) = \lim\limits_{o} E \frac{(X_t | X_o = x) - x}{t}$, c'est-à-dire, la vitesse moyenne de ce processus en $x$, un calcul élémentaire montre que $V(x) = \frac{p_x - p_{x-1}}{p_x}$, qui est la fonction "score". L'équation (1) du lemme 1.2 peut maintenant s'écrire:

$$E([X-E(X)]^2) \cdot E([V(x) - E(V(x))]^2) \geq 1$$

où l'espérance est calculée par rapport à la mesure de probabilité $P(x)$. Ce résultat établit que le produit de l'incertitude associée à la position par l'incertitude associée à la vitesse (telle que mesurée par les divergences ci-dessus) est toujours plus grande ou égale à un. Ce produit n'est égal à un que si et seulement si $P(x)$ est une loi de Poisson.

## 2. Deuxième caractérisation de la loi de Poisson

*Lemme 2.1:* *Soit $X$, $Y$ deux variables aléatoires indépendantes dans $X_0$ telles que $I(X) < \infty$ et $I(Y) < \infty$. Alors $I(X*Y) \leq \min\{I(X), I(Y)\}$.*

**Démonstration:** Soit $Z = X*Y$ et supposons que $I(X) \leq I(Y)$. Posons:

$$\tau_x^X = \frac{p_x^X - p_{x-1}^X}{p_x^X} \quad \text{où} \quad p_x^X = P(X = x) .$$

Nous avons:

$$\tau_s^Z = \frac{p_s^Z - p_{s-1}^Z}{p_s^Z} \quad \text{où} \quad p_s^Z = \sum_{t=0}^{s} p_t^Y p_{s-t}^X . \quad \text{Ainsi,}$$

$$\tau_s^Z = \sum_{t=0}^{s} \frac{p_t^Y p_{s-t}^X \tau_{s-t}^X}{p_s^Z} = \mathbb{E}[\tau^X \mid X+Y = s] \tag{2}$$

Donc $I(Z) = \mathbb{E}[(\tau^Z)^2] = \mathbb{E}[\mathbb{E}^2[\tau^X \mid X+Y]] \leq \mathbb{E}[(\tau^X)^2] = I(X)$ .

*Proposition 2.2:* *Soit $X$ et $Y$ deux variables aléatoires indépendantes dans $X_0$ tel que $I(X) < \infty$ et $I(Y) < \infty$. On a alors:*

1) $\dfrac{I(X) + I(Y)}{4} - I(X*Y) \geq 0$, *et*

2) $\dfrac{I(X)}{2} - I(X*X) = 0$ *si et seulement si $X$ suit une loi de Poisson.*

**Démonstration:** Soit $Z = X*Y$. Une version symétrique de (2) est donnée par $\tau_s^Z = \frac{1}{2} E[\tau^Y + \tau^X \mid X+Y = s]$. Nous avons:

$$\begin{aligned}
\frac{I(X) + I(Y)}{4} - I(X*Y) &= \frac{1}{4} \{\mathbb{E}(\tau^X)^2 + \mathbb{E}(\tau^Y)^2 - \mathbb{E}(\mathbb{E}^2(\tau^X + \tau^Y \mid X+Y)\} \\
&= \frac{1}{4} \{\mathbb{E}(\tau^X + \tau^Y)^2 - \mathbb{E}(\mathbb{E}^2(\tau^X + \tau^Y \mid X+Y)\} \\
&= \frac{1}{4} \mathbb{E}(\mathrm{Var}(\tau^X + \tau^Y \mid X+Y)) \geq 0
\end{aligned} \tag{3}$$

Supposons que les lois de $X$ et $Y$ soient les mêmes; ($L(X) = L(Y)$). Alors on a égalité en (3) si et seulement si $\tau_x^X + \tau_{t-s}^Y = c_t$, $\forall\, t \geq 0$, $0 \leq x \leq t$, c'est-à-dire si et seulement si:

$$\frac{p_{x-1}}{p_x} + \frac{p_{t-x-1}}{p_{t-x}} = c_t .$$

Pour $x = 0$, $c_t = \frac{p_{t-1}}{p_t}$ et alors $c_x + c_y = c_{x+y}$ $\forall\, x, y \in \mathbb{N}$. Puisque $c_0 = 0$, ceci implique que $c_t = t/\lambda$ pour $\lambda > 0$ et ainsi $X$ suit une loi de Poisson de paramètre $\lambda$. On écrit $X \sim P_\lambda$.

*Corollaire 2.3:*

$$\frac{I(X) + I(Y)}{4} - I(X*Y) = \frac{1}{4} E\, E\{\{\tau^X + \tau^Y - E(\tau^X + \tau^Y \mid X+Y)\}^2 \mid X+Y\}$$

$$= \frac{1}{4} E(\tau^X + \tau^Y - 2\tau^{X+Y})^2 .$$

## 3. Lemme de projection et polynômes de Poisson-Charlier

En vue de la démonstration du théorème 4.4 on donne ici une borne inférieure à $\mathbb{E}[(v(X) + v(Y) - w(X+Y))^2]$; où $X$ et $Y$ sont des variables aléatoires indépendantes de loi $P_a$ (Poisson de paramètre a).

*Lemme 3.1: Soient $X$ et $Y$ des variables aléatoires indépendantes de loi de Poisson $P_a$ et $v, w : \mathbb{N} \to \mathbb{R}$ des fonctions mesurables arbitraires telles que $\mathbb{E}[v^2(X)] < \infty$. Alors*

$$\mathbb{E}[(v(X) + v(Y) - w(X+Y))^2] \geq \inf_{\alpha,\beta} \mathbb{E}[(v(X) - \alpha X - \beta)^2] \tag{4}$$

Démonstration: En choisissant $w(X+Y) = \mathbb{E}[v(X) + v(Y) \mid X+Y]$ on minimise le terme de gauche de (4). Le terme de droite est égal à

$$\mathbb{E}[(v(X) - \mathbb{E}[v(X)] - \alpha_v(X-\mathbb{E}[X]))^2], \text{ où } \alpha_v = \frac{\operatorname{cov}(X,v(X))}{\operatorname{var}(X)} .$$

Posant: $\overline{v}(X) = v(X) - \mathbb{E}[v(X)] - \alpha_v(X - \mathbb{E}[X])$, (4) devient

$$\mathbb{E}[\{\overline{v}(X) + \overline{v}(Y) - \mathbb{E}[\overline{v}(X) + \overline{v}(Y) \mid X + Y]\}^2]$$

ou, de façon équivalente,

$$\mathbb{E}[(\mathbb{E}[\overline{v}(X) \mid X+Y])^2] \leq \frac{1}{4} \mathbb{E}[\overline{v}^2(X)] . \tag{5}$$

Pour prouver (5) il est commode (mais probablement pas nécessaire) d'employer les polynômes de Poisson-Charlier (*cf.* [15]). Soit $\{p_{n,a}\}$ une suite orthonormale complète pour $L^2(P_a)$, où $P_a$ représente une mesure de Poisson de paramètre $a$ sur $\mathbb{N}$.

On définit les polynôme $P_{n,a}$ par la fonction génératrice suivante

$$\rho_a(t,X) = e^{-t}(1+t/a)^X = \sum_0^\infty \frac{t^n P_{n,a}(X)}{n!} \tag{6}$$

Alors Parthasarathy [13, p. 221] montre que: $p_{n,a} = \frac{a^{n/2} P_{n,a}}{n!}$ forme une suite orthonormale.

On peut montrer que l'application $T: L^2(P_a) \to L^2(P_{2a})$ définie par $v(X) \to \mathbb{E}[v(X) \mid X+Y]$ envoie $p_{n,a}$ sur $2^{-n/2} p_{n,2a}$. Utilisant la fonction génératrice de (6) nous avons:

$$\mathbb{E}[\rho_a(t,X) \mid X+Y=z] = e^{-t} \sum_{k=0}^{z} \binom{z}{k} (1 + \frac{t}{a})^k \, 2^{-z}$$

$$= e^{-t}(1+1+\frac{t}{a})^z \, 2^{-z} = \rho_{2a}(t,z) \; .$$

Et de cette même relation on tire:

$$\mathbb{E}[P_{n,a}(X) \mid X+Y] = P_{n,2a}(X+Y).$$

Pour compléter la preuve, notons que dans (5) $\bar{v}$ appartient au sous-espace fermé $M$ engendré par $\{p_{n,a},\ n \geq 2\}$ puisque $p_{0,a} = 1$, $p_{1,a} = \frac{X}{a} - 1$. De plus (3) est vérifié dans $M$ puisque:

$$\mathbb{E}[\mathbb{E}^2[p_{n,a}(X) \mid X+Y]] = \mathbb{E}[(2^{-n/2} p_{n,2a}(X+Y))^2] = 2^{-n} \, \mathbb{E}[p_{n,a}^2(X)] \; .$$

## 4. Caractérisation des suites ayant comme points d'accumulation des lois de Poisson.

La section 3 nous permet de donner une condition nécessaire et suffisante en terme de $I(Q_n)$ et $I(Q_n * Q_n)$ pour qu'une suite $\{Q_n\}$ de mesures de probabilité à support dans $\mathbb{N}$ (mais pas nécessairement égal à $\mathbb{N}$) ait comme limite une loi de Poisson de paramètre $\lambda$, où $\lambda$ est tel que $\mathrm{var}(Q_n) \to \lambda$ .

Il est clair que $Q_n \xrightarrow{\mathcal{D}} P_\lambda$ n'implique pas nécessairement que $\mathrm{var}(Q_n) \to \lambda$. En effet, il suffit de considérer la suite $Q_n = P_\lambda + a_n \delta_{\{n\}} - a_{n-1} \delta_{\{n-1\}}$ avec $a_n$ convenablement choisi. Pour éviter de telles situations on va se restreindre à la classe suivante de suites de mesures de probabilité:

<u>*Définition 4.1:*</u> *$\{Q_n\}_{n=1}^{\infty}$ est une suite de mesures de probabilité du type $B_2$ si le support de $Q_n$ est inclu dans $\mathbb{N}$ pour tout $n$ et s'il existe une fonction $f: \mathbb{N} \to \mathbb{R}$ telle que $Q_n(x) \leq f(x)$ pour tout $n$ et $\sum_{x=0}^{\infty} x^2 f(x) < \infty$ .*

<u>*Définition 4.2:*</u> *Si $\mathrm{var}(P) < \infty$, on dit que $Q_n \xrightarrow{\mathcal{D}_2} P$ si $Q_n \xrightarrow{\mathcal{D}} P$ et $\mathrm{var}(Q_n) \to \mathrm{var}(P)$.*

*Remarque 4.3:* Ce type de $\mathcal{D}_2$-convergence a été à l'origine définie par Bickel et Freedman [2]. Ce type de convergence est mètrisable, par exemple, en utilisant la métrique $d_2$ de Mallows. La $d_2$-distance entre $P$ et $Q$ (telles que $\sum x^2 p(x)$ et $\sum x^2 q(x)$ soient finies) est définie comme la borne inférieure de $E[(X-Y)^2]$ sur toutes les distributions conjointes de paires de variables aléatoires $X$ et $Y$ dont les distributions marginales sont données par $P$ et $Q$ fixées à l'avance.

Cette métrique fut introduite par Mallows [11] et Tanaka [17] et reliée à la métrique de Vassershtein introudite par Dobrushin [4]. Pour des détails supplémentaires sur la métrique $d_2$ voir Bickel et Freedman [2].

*Théorème 4.4:* *Soit $\{Q_n\}$ une suite de mesures de probabilité du type $\mathcal{B}_2$ à support dans $\mathbb{N}$ telle que $\lim\inf Q_n(\{0\}) = q_0 > 0$. Alors l'ensemble des points d'accumulation (pour la $\mathcal{D}_2$-convergence) de $\{Q_n\}$ est contenu dans $\{P_\lambda,\ \lambda \in [0,\infty)\}$ si et seulement si pour tout $\delta > 0$ suffisamment petit on a*

$$I(Q_{n*\delta}) - 2I(Q_{n*\delta} * Q_{n*\delta}) \to 0 \tag{7}$$

*où* $Q_{n*\delta} = Q_n * P_\delta$.

*Remarque 4.5:* La condition concernant $q_0$ a été introduite pour exclure les situations dans lesquelles la condition (6) est vérifiée mais $Q_n$ converge vers une loi de Poisson tronquée, par exemple $Q(x) = P_\lambda\{X=x \mid X \geq 1\}$. (De tels exemples peuvent facilement être construits).

Démonstration du théorème: Supposons d'abord que l'ensemble (non vide) des points limites (pour la $\mathcal{D}_2$-convergnece) de $\{Q_n\}$ est composé de lois de Poisson. On va établir (7) en montrant que pour toute sous-suite $\{n'\}$, il existe une sous-suite $\{n''\}$ pour laquelle (7) est vérifiée. Ainsi, en utilisant le fait que $\{Q_n\}$ est tendue (en vertu de la définition 4.1) on peut se restreindre au cas où $Q_n \xrightarrow{\mathcal{D}_2} P_\lambda$. De simples calculs montrent que:

$$\tau_{n*\delta}(x) = \frac{Q_{n*\delta}(x) - Q_{n*\delta}(x-1)}{Q_{n*\delta}(x)} \dashrightarrow 1 - \frac{x}{\lambda+\delta}$$

point par point en $x$ lorsque $n \to \infty$. Pour transformer ceci en l'affirmation $I(Q_{n*\delta}) \to \frac{1}{\lambda+\delta}$, il est suffisant de constater, à partir de (2) dans la section 2, que $\tau^2_{n*\delta}(x) \leq 2(1+x^2/\delta^2)$, $Q_{n*\delta} \xrightarrow{\mathcal{D}_2} P_{\lambda+\delta}$ et d'appliquer l'extension de Young du théorème de convergence dominée (Loève [10], pp. 164-165). Un argument similaire montre que $I(Q_{n*\delta} * Q_{n*\delta}) \to \frac{1}{2(\lambda+\delta)}$, et ceci implique (7).

Pour l'implication inverse notons d'abord que pour $n$ grand nous avons:

$Q_{n*\delta}(x) \geq \frac{q_0}{2} P_\delta(x)$. Du corollaire 2.3 et du lemme 3.1 il suit que:

$$I(Q_{n*\delta}) - 2\, I(Q_{n*\delta} * Q_{n*\delta}) \geq \frac{q_0^2}{8} \inf\{\sum_x [\tau_{n*\delta}(x) + \alpha x + \beta]^2 P_\delta(x) : (\alpha,\beta) \in \mathbb{R}^2\}$$

La condition (7) implique qu'il existe $\alpha_n, \beta_n$ tels que:

$$\tau_{n*\delta}(x) + \alpha_n x + \beta_n \longrightarrow 0, \ \forall\, x \in \mathbb{N}, \text{ quand } n \to \infty. \tag{8}$$

Maintenant, supposons qu'il existe une sous-suite $\{n'\}$ telle que $Q_{n'} \xrightarrow{\mathcal{D}_2} P$ ; ainsi:

$\tau_{n'*\delta}(x) \longrightarrow \tau_{*\delta}(x) = 1 - \frac{P_{*\delta}(x-1)}{P_{*\delta}(x)}$, où $P_{*\delta}(\cdot)$ est définie ci-dessous. En prenant $x = 0$ dans l'équation (8) on voit que $\lim \beta_{n'}$ existe et est égale à 1; et en prenant $x = 1$ dans la même équation que $\lim \alpha_{n'}$ existe et est égale à $\alpha$.

Ainsi $\frac{P_{*\delta}(x-1)}{P_{*\delta}(x)} = \alpha x$, et alors $P_{*\delta}(x) = \frac{e^{-1/\alpha}\, \alpha^{-x}}{x!}$. Donc $P$ est une loi de Poisson de paramètre $(\alpha^{-1} - \delta)$.

Bien que chacun des résultats suivants peut être démontré de façon élémentaire on peut à l'aide du théorème 4.3 fournir une démonstration unifiée des "lois des petits nombres".

*Corollaire 4.6:*

a) *Soit $Q_n = Bin(n,p_n)$ une suite de lois binomiales telle que: $np_n \to \lambda$ ($p_n$ étant la probabilité d'un succès).*

b) *Soit $Q_n = BN(n,p_n)$ une suite de lois binomiales-négatives telle que: $np_n \to \lambda$. ($p_n$ étant la probabilité d'un échec).*

c) *Soit $Q_n = PB(n,p_1,\ldots,p_n)$ une suite de lois Poisson-binomiales ($Q_n = Bin(1,p_1) * Bin(1,p_2) * \ldots * Bin(1,p_n)$) telle que:*

*$\alpha = max\{p_i : 1 \leq i \leq n\} \to 0$ et $\sum_{i=1}^{n} p_i \to \lambda$. Dans ces trois cas on a $Q_n \xrightarrow{\mathcal{D}_2} P_\lambda$.*

Démonstration: Dans chacun de ces trois cas, $\{Q_n\}$ est une suite de mesures de probabilité du type $\mathcal{B}_2$ puisque $Q_n(X)$ est dominé par $\frac{(\lambda+\varepsilon)^x}{x!}$ pour $n$ suffisamment grand.

Soit $X \sim Bin(n,p_n)$ dans le cas a); $X \sim BN(n,p_n)$ dans le cas b); $X \sim PB(n,p_1,\ldots,p_n)$ dans le cas c) et $Y \sim P_\delta$. En vertu de (2) de la section 2 nous avons:

$$\tau_{n*\delta} = \mathbb{E}[\tau^Y \mid X+Y = s] = \mathbb{E}[1 - \frac{s-X}{\delta} \mid X+Y = s]. \tag{9}$$

Dans le cas (a) soit:
$Y_i (i=1,\ldots,n)$ une suite de variables aléatoires indépendantes et équidistribuées suivant une loi de Poisson de paramètre $\delta/n$ et soit $X_i$ une suite de variables aléatoires indépendantes et équidistribuées suivant une loi de Bernouilli de paramètre $p_n$=probabilité de succès. Soit A l'événement $\bigcap_{i=1}^{n} \{X_i+Y_i \leq 1\}$ .

Nous avons de façon évidente:

$$
\begin{aligned}
E[X|X+Y=s] &= E[X|X+Y=s \text{ et } A] + O(\tfrac{1}{n}) \\
&= s\ E[X_i|X_i+Y_i=1] \\
&= \frac{s\ p_n}{p_n+(1-p_n)\delta/n} + o_n(1) .
\end{aligned}
$$

Le cas (b) se traite de manière similaire en considérant une suite $X_i$ de variables aléatoires indépendantes et équidistribuées suivant une loi géométrique de paramètre $p_n$=probabilité d'échec.

Dans le cas (c) soit $X_i \sim \text{Bin}(1,p_i)$ $i=1,2,\ldots,n$, les $X_i$ étant indépendantes et équidistribuées, et soit $Y_i$ une suite de variables aléatoires indépendantes suivant des lois de Poisson de paramètre $\delta_i^* = p_i/p_+$ , $i = 1\ldots n$ où $p_+ = \sum_i p_i$, et soit $A = \bigcap_{i=1}^{n} \{X_i+Y_i \leq 1\}$ .

$$
\begin{aligned}
\text{Alors,}\quad E(X \mid X+Y=s) &= \sum_{i=1}^{n} P(X_i=1 \mid \textstyle\sum(X_i+Y_i)=s \text{ et } A) + o_n(1) \\
&= \sum_{i=1}^{n} P(X_i=1 \mid X_i+Y_i=1)P(X_i+Y_i=1 \mid \textstyle\sum(X_i+Y_i)=s \text{ et } A) + o_n(1) \\
&= \sum_{i=1}^{n} \frac{p_i}{p_i+(1-p_i)\delta_i}\ s\ q_i, \quad \text{où} \quad q_i = \frac{p_i+\delta p_i/p_+}{p_+ + \delta} \\
&= \frac{s\ p_+}{p_+ + \delta} + o_n(1).
\end{aligned}
$$

Par conséquent, dans les trois cas nous avons

$$\tau^2_{n*\delta}(x) = \left[1 - \frac{x}{\lambda+\delta}\right]^2 + o_n(1). \tag{10}$$

Des arguments similaires montrent que:

$$\tau^2_{n*\delta*n*\delta}(x) = \left[1 - \frac{x}{2(\lambda+\delta)}\right]^2 + o_n(1). \tag{11}$$

Puisque que $Q_n$ est une suite de probabilités du type $\mathcal{B}_2$ alors en vertu des équations (10) et (11) nous avons $I(Q_{n*\delta}) - 2I(Q_{n*\delta}*Q_{n*\delta}) \longrightarrow 0$, $\forall\ \delta < \delta'$ et ainsi, en vertu du théorème 4.4, nous avons dans les trois cas, $Q_n \xrightarrow{\mathcal{D}_2} P_\lambda$ .

## Références


[1] Barron, A.R., Entropy and the central limit theorem, Annals of Probability 14, (1986), 336-342.

[2] Bickel, P.J. et Freedman, D.A., Some asymptotic theory for the bootstrap, Ann. Statist. 9 (1981), 1196-1217.

[3] Brown, L.D., A proof of the central limit theorem motivated by the Cramer-Rao Inequality, Statistics and Probability: Essays in Honor of C.R. Rao, Kallianpur, G. Krishnaian, P.R., Ghosh, J.K., eds. North-Holland (1982), 141-148.

[4] Dobrushin, R.L., Describing a system of random variables by conditional distributions, Theory Probab. Appl. 15 (1970) 458-486.

[5] Dynkin E.B. et Yushekvich, A.A., Markov Processes; Theorems and Problems, Plenum Press, New York (1969).

[6] Huber, P.J., Robust Statistics, John Wiley and Sons (1981).

[7] Kendall, D.G., Information theory and the limit theorem for Markov chains and processes with a countable infinity of states, Ann. Inst. Statist. Math. 15, (1964), 137-143.

[8] Johnstone, I., Admissibility, difference equations and recurrence in estimating a Poisson mean, Ann. Statist. 12, (1984), 1173-1198.

[9] Linnik, Y.V., An information-theoretic proof of the central limit theorem with the Lindeberg condition, Theory Probab. Applic. 4 (1959), 288-299.

[10] Loève, M.M., Probability Theory, 3rd ed., Van Nostrand, Princeton (1963).

[11] Mallows, C.L., A note on asymptotic joint normality, Ann. Math. Statist. 43, (1972), 508-515.

[12] McKean, H.P. Jr., Speed of approach to equilibrium for Kac's caricature of a Maxwellian gas, Arch. Rational Mech. Anal. 21 (1967), 343-367.

[13] Parthasarathy, K.R., Introduction to Probability and Measure, Springer, New York, (1977).

[14] Rényi, A., On measures of entropy and information, Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability, University of California Press Berkeley, 1 (1961), 541-561.

[15] Schmidt, E., Über die Charlier-Jordansche Entwicklung einer willkürlichen funktion nach der Poissonschen funktion und ihren Ableitungen, Ztschr. f. angew. Math. und Mech. 13 (1933), 139-142.

[16] Shannon C.E. et Weaver, W., The Mathematical Theory of Communications, Univ. of Illinois Press, Urbana (1949).

[17] Tanaka H., An inequality for a functional of probability distribution and its application to Kac's one-dimensional model of a Maxwellian gas, Z. Warsch. verw. Gebiete 27 (1973), 47-52.