Schéma d'occupation d'une infinité de boîtes

Marie Devaine, Jehanne Richet
Sujet proposé et encadré par Jean Bertoin

24 Juin 2008

Table des matières

Introduction

On s'intéresse à un modèle d'occupation qui a de nombreuses applications notamment en écologie, en statistiques ou même en littérature. Le problème est le suivant : on cherche à estimer le nombre d'espèces et leur proportions dans un environnement donné alors que l'on ne dispose que d'échantillons. On peut aussi s'intéresser au nombre et à la fréquence des mots utilisés par un certain auteur à partir de l'étude d'une seule de ses oeuvres. Cela peut s'avérer assez compliqué quand les espèces considérées sont très rares ou quand les mots sont très peu souvent employés. On modélise alors les espèces par des boîtes et les échantillons par des lancers de boules.

On se donne donc une suite de boîtes et on y lance aléatoirement des boules de manière indépendante. La probabilité qu'une boule tombe dans la $i^{\text{ème}}$ boîte est p_i. On se place dans le cas d'une infinité dénombrable de boîtes, c'est-à dire que

$$\forall i \in \mathbb{N}, \quad p_i > 0.$$

On peut voir l'ensemble des boîtes comme une partition dénombrable de $[0,1]$, *les* p_i *sont alors les longueurs des boîtes.*

1 Poissonisation

1.1 Position du problème

On note $X_{n,i}$ le nombre de boules après le $n^{\text{ème}}$ lancer dans la $i^{\text{ème}}$ boîte. On va étudier $K_n = |\{j : X_{n,j} > 0\}|$ qui est le nombre de boîtes contenant au moins une boule au bout du $n^{\text{ème}}$ lancer, et $K_{n,r} = |\{j : X_{n,j} = r\}|$ qui est le nombre de boîtes contenant exactement r boules au bout du $n^{\text{ème}}$ lancer. On a $K_n = \sum_r K_{n,r}$ et $\sum_r rK_{n,r} = n$. On peut également exprimer la loi de $(X_{n,j})_{j\in\mathbb{N}^*}$ pour tout $n \in \mathbb{N}^*$:

$$\mathbb{P}((X_{n,j})_{j\in\mathbb{N}^*} = (n_j)_{j\in\mathbb{N}^*}) = \frac{n!}{\prod_j n_j!} \prod_j p_j^{n_j} \quad \text{avec} \quad \sum_j n_j = n.$$

Pour calculer les moments de K_n, on utilise :

$$K_{n,k} = \sum_j 1_{X_{n,j}=k}, \quad K_n = \sum_j 1_{X_{n,j}>0}.$$

On note $\Phi_n = \mathbb{E}(K_n)$ et $\Phi_{n,r} = \mathbb{E}(K_{n,r})$ et on obtient :

$$\Phi_n = \sum_i (1 - (1-p_i)^n) \tag{1}$$

et

$$\Phi_{n,r} = \binom{n}{r} \sum_i (1-p_i)^{n-r} p_i^r \tag{2}$$

Malheureusement, comme les événements $(\{X_{n,j} = k\})_{j\geq 1}$ ne sont pas indépendants, les calculs et les expressions des moments sont très vite compliqués. Cependant l'expression des variances va nous être utile. On pose $V_n = \mathrm{Var}(K_n)$ et $V_{n,r} = \mathrm{Var}(K_{n,r})$. Un calcul, donné en appendice, nous donne l'expression suivante :

$$V_n = \Phi_{2n} - \Phi_n + \sum_{i\neq j} \left[(1-p_j-p_i)^n - (1-p_j)^n(1-p_j)^n\right] \tag{3}$$

On a également :

$$\begin{aligned} V_{n,r} &= \binom{n}{r} \sum_{i\neq j} \left[\binom{n-r}{r} (1-p_j-p_i)^{n-2r} - \binom{n}{r} (1-p_j)^{n-r}(1-p_i)^{n-r} \right] \\ &\quad + \Phi_{n,r} - \frac{\binom{n}{r}\binom{n}{r}}{\binom{2n}{2r}} \Phi_{2n,2r} \end{aligned} \tag{4}$$

Pour surmonter la difficulté liée à la dépendance des variables $X_{n,j}$, on va "poissonniser" les temps de lancer des boules, c'est-à-dire que plutôt que de jeter les boules une à une, on va les jeter "aléatoirement" au cours du temps, de manière à rendre le nombre de boules dans une boîte au bout d'un temps t indépendant du nombre de boules dans les autres boîtes. Plus précisément on va les jeter selon un processus de Poisson.

1.2 Processus de Poisson

On va faire intervenir le nombre de balles tombées dans les boîtes pendant un certain intervalle de temps, on est donc amené à introduire des processus de comptage.

1.2.1 Définitions

Definition Un processus stochastique $(N_t)_{t\in\mathbb{R}_+}$ est un processus de comptage si pour tout $t \in \mathbb{R}_+$, N_t représente le nombre de "tops" (pour nous le nombre de lancers de boule) s'étant produit dans l'intervalle de temps $[0, t]$ et si $N_0 = 0$.

Ainsi, pour tous $0 < a < b$, $N_b - N_a$ représente le nombre de tops ayant eu lieu dans l'intervalle de temps $[a, b]$.

Definition Un processus de comptage est dit à *accroissements indépendants* si les nombres de tops se produisant dans des intervalles de temps disjoints sont indépendants.

Definition Un processus de comptage est dit à *accroissements stationnaires* si pour tous $t_1, t_2, a \in \mathbb{R}_+$, $N_{t_1+a} - N_{t_2+a}$ et $N_{t_1} - N_{t_2}$ ont la même loi.

Definition Un processus de comptage $(N_t)_{t\in\mathbb{R}_+}$ est un *Processus de Poisson de densité* $\lambda > 0$ s'il est à accroissements indépendants et si

$$\forall t, s \in \mathbb{R}_+, \quad \forall n \in \mathbb{N}, \quad \mathbb{P}(N_{s+t} - N_t = n) = e^{-\lambda s}\frac{(\lambda s)^n}{n!}.$$

On peut remarquer qu'un processus de Poisson est à accroissements stationnaires.

Ainsi définis, les processus de Poisson ont de bonnes propriétés d'additivité que l'on verra plus loin et qui seront utiles à notre problème. Cependant, pour construire effectivement ces processus il est intéressant d'envisager une autre caractérisation.

1.2.2 Caractérisation à l'aide de temps exponentiels

Soit $(N_t)_{t\in\mathbb{R}_+}$ un processus de Poisson de paramètre λ. On pose pour tout $k \in \mathbb{N}^*$

$$S_k = \inf\{t|N_t = k\} \text{ et } T_k = S_k - S_{k-1} \text{ avec } S_0 = 0.$$

Les T_k peuvent être vus comme les temps d'attente entre les lancers, et S_k comme le temps de lancer de la $k^{\text{ème}}$ boule.

Proposition 1. *$(T_k)_{k\in\mathbb{N}^*}$ est une famille de variables aléatoires indépendantes et identiquement distribuées (i.i.d.) de loi exponentielle de paramètre λ.*

Démonstration. Montrer que T_k suit la même loi qu'une famille Y_k de v.a.i.i.d. de loi exponentielle de paramètre λ revient au même que montrer que S_n suit la même loi que celle des $X_n = \sum_{k=1}^{n} Y_k$.

Il est de plus évident que $S_1 \leq S_2 \cdots \leq S_n \cdots$ mais on a également les inégalités strictes p.s. En effet, si on avait pour $i < j$, $S_i = S_j = t_0$, alors pour $t > t_0$ on aurait $N_t - N_{t_0} > 0$. Or

$$\mathbb{P}\left(\exists i < j,\ S_i = S_j\right) = \lim_{N \to \infty} \mathbb{P}\left(\exists i < j,\ S_i = S_j \leq N\right).$$

Pour tout N et pour tout n, on note

$$[a_k, b_k] = \left[\frac{k}{2n}, \frac{k+2}{2n}\right]$$

pour $k \in \{0, 2, \cdots, 2nN - 2\}$.

On a alors

$$\begin{aligned}\mathbb{P}(\exists i < j,\ S_i = S_j \leq N) &\leq \mathbb{P}(\exists k \in \{0, 2, \cdots, 2(nN-1)\},\ N_{b_k} - N_{a_k} \geq 2) \\ &\leq nNe^{-\frac{\lambda}{n}} \sum_{p \geq 2} \frac{\lambda^p}{n^p p!}\end{aligned}$$

qui à N fixé tend vers 0 quand n tend vers l'infini. Donc pour tout N,

$$\mathbb{P}\left(\exists i < j, S_i = S_j \leq N\right) = 0 \quad \text{et} \quad (S_n)_{n \in \mathbb{N}} \in \Lambda = \{0 < s_1 < s_2 \cdots\} \text{ p.s.}$$

Mais $(X_n)_{n \in \mathbb{N}}$ est lui aussi p.s. dans Λ, car pour tout n, $\mathbb{P}(Y_n = 0) = 0$, et donc pour tout n, $Y_n = 0$ p.s.

Il s'agit donc de montrer que

$$\mathbb{P}\left((S_n)_{n \in \mathbb{N}} \in P\right) = \mathbb{P}((X_n)_{n \in \mathbb{N}} \in P)$$

pour tout P appartenant à un Π-système de la tribu induite sur Λ.

Or l'ensemble

$$\pi = \{]a_1, b_1] \times \cdots \times]a_n, b_n] \times]b_n, \infty[\text{ tq } i \in \mathbb{N},\ 0 \leq a_i < b_i \leq a_{i+1},\ n \in \mathbb{N}\}$$

est stable par intersection finie et engendre Λ. On se donne n un entier et $a_1 < b_1 \leq a_2 \cdots < b_n$ des réels positifs. Calculons

$$\begin{aligned}&\mathbb{P}(S_1 \in]a_1, b_1],\ \cdots,\ S_n \in]a_n, b_n],\ S_{n+1} > b_n) \\ &\quad = \mathbb{P}(N_{a_1} = 0,\ N_{b_1} - N_{a1} = 1,\ N_{a_2} - N_{b_1} = 0,\ \cdots,\ N_{b_n} - N_{a_n} = 1) \\ &\quad = \mathbb{P}(N_{a_1} = 0)\mathbb{P}(N_{b_1} - N_{a1} = 1)\mathbb{P}(N_{a_2} - N_{b_1} = 0) \cdots (N_{b_n} - N_{a_n} = 1) \\ &\quad = e^{-\lambda a_1} e^{-\lambda(b_1 - a_1)} \lambda(b_1 - a_1) e^{-\lambda(b_1 - a_2)} \cdots e^{-\lambda(b_n - a_n)} \lambda(b_n - a_n) \\ &\quad = \lambda^n (b_1 - a_1) \cdots (b_n - a_n) e^{-\lambda b_n}.\end{aligned}$$

D'autre part, en utilisant le changement de variable $(y_1, \cdots, y_n) \mapsto (y_1,\ y_+ y_2,\ \cdots,\ y_1 + \cdots + y_n)$, on obtient :

$$\begin{aligned}&\mathbb{P}(X_1 \in]a_1, b_1],\ \cdots,\ X_n \in]a_n, b_n],\ X_{n+1} > b_n) \\ &\quad = \int \cdots \int 1_{y_1 \in]a_1, b_1],\ \cdots,\ y_1 + \cdots + y_n \in]a_n, b_n]} \lambda^n e^{-\lambda(y_1 + \cdots + y_n)} dy_1 \cdots dy_n \\ &\quad = \int \cdots \int 1_{a_1 < x_1 \leq b_1 \leq a_2 \cdots \leq a_n < x_n \leq b_n > x_{n+1}} \lambda^{n+1} e^{-\lambda x_{n+1}} dx_1 \cdots dx_n \\ &\quad = \lambda^n (b_1 - a_1) \cdots (b_n - a_n) e^{-\lambda b_n}.\end{aligned}$$

Ce qui achève la preuve. □

On va maintenant montrer une sorte de réciproque qui sera très utile pour construire concrètement des processus de Poisson.

Soit $(T_n)_{n\in\mathbb{N}^*}$ une famille de variables aléatoires i.i.d. de loi exponentielle de paramètre $\lambda > 0$. On lui associe le processus de comptage $(N_t)_{t\geq 0}$ tel que le $n^{\text{ème}}$ top ait lieu à l'instant $S_n = T_1 + \cdots + T_n$. En posant $S_0 = 0$, on a donc :

$$N_t = \sum_{n\geq 1} 1_{\{S_n\leq t\}}.$$

Proposition 2. *Le processus $(N_t)_{t\in\mathbb{R}_+}$ ainsi défini est un processus de Poisson.*

Démonstration. Comme $(S_n)_{n\in\mathbb{N}^*}$ est croissante, $\{S_n \leq t < S_{n+1}\} = \{N_t = n\}$.

Soit $n,\ a_1,\ \cdots,\ a_n$ des entiers et $t_1 \leq \cdots \leq t_n$ des réels positifs. En utilisant à la quatrième ligne le même changement de variable que dans la preuve précédente, on obtient

$$\begin{aligned}
&\mathbb{P}(N_{t_1} = a_1,\ N_{t_2} - N_{t_1} = a_2,\ \cdots,\ N_{t_n} - N_{t_{n-1}} = a_n)\\
&= \mathbb{P}(S_1,\ \cdots,\ S_{a_1} \leq t_1 \text{ et } S_{a_1+1},\ \cdots S_{b_2} \in]t_1, t_2] \text{ et } \cdots \text{ et} S_{b_n+1} > t_n)\\
&= \lambda^{b_n+1} \int \dots \int e^{-\lambda(s_1+\cdots+s_{b_n+1})} 1_{s_1\leq\cdots\leq s_{a_1}\leq t_1 < s_{a_1+1}\leq\cdots< s_{b_n+1}} ds_1 \cdots ds_{b_n+1}\\
&= \lambda^{b_n} \left(\int \dots \int 1_{s_1\leq\cdots\leq s_{a_1}\leq t_1} ds_1 \cdots ds_{a_1}\right)\\
&\quad \left(\int \dots \int 1_{t_1< s_{a_1+1}\leq\cdots\leq s_{b_2}\leq t_2} ds_{a_1+1} \cdots ds_{b_2}\right) \dots \left(\int_{t_n}^{\infty} \lambda e^{-\lambda s_{b_n+1}} ds_{b_n+1}\right)\\
&= \lambda^{b_n+1} \frac{t_1^{a_1}}{a_1!} \frac{(t_2-t_1)^{a_2}}{a_2!} \dots \frac{(t_n - t_{n-1})^{a_n}}{a_n!}
\end{aligned}$$

où $b_1 = a_1$ et pour $i \geq 1 \quad b_{i+1} = b_i + a_{i+1}$. Ce qui montre que $(N_t)_{t\geq 0}$ est un processus de Poisson. □

Ici, on va jeter la première boule au bout d'un temps exponentiel T_1 de paramètre 1, puis la deuxième boule au bout d'un temps exponentiel T_2 de paramètre 1, et ainsi de suite. D'après ce qu'on vient de voir le processus de comptage qui compte le nombre de balles lancées au temps t sera un processus de Poisson. Cependant il n'est pas évident par cette définition que l'on ait l'indépendance voulue. On va donc construire des processus de Poisson indépendants sur chaque boîte et montrer que la superposition de ces processus est bien un processus de Poisson.

1.2.3 Théorème de superposition

Proposition 3 (Additivité de la loi de Poisson)**.** *Soit $(X_n)_{n\in\mathbb{N}}$ une famille de variables aléatoires indépendantes telle que pour tout entier naturel n, X_n suit la loi de probabilité $\mathcal{P}(\mu_n)$ (loi de poisson de paramètre μ_n).*

Si la série de terme général μ_i converge, alors la série de terme général X_i converge p.s. et

$$S = \sum_{i=0}^{\infty} X_i$$

suit la loi de Poisson de paramètre μ, où

$$\mu = \sum_{i=0}^{\infty} \mu_i.$$

Démonstration. Remarquons que si X et Y suivent respectivement les lois $\mathcal{P}(\mu)$ et $\mathcal{P}(\lambda)$ et sont indépendantes, alors $X + Y$ suit la loi $\mathcal{P}(\mu + \lambda)$. En effet,

$$\begin{aligned}
\mathbb{P}(X + Y = n) &= \sum_{k=0}^{n} \mathbb{P}(X = k, Y = n - k) \\
&= \sum_{k=0}^{n} \mathbb{P}(X = k)\mathbb{P}(Y = n - k) \\
&= \sum_{k=0}^{n} \frac{\mu^k e^{-\mu}}{k!} \frac{\lambda^{n-k} e^{-\lambda}}{(n-k)!} \\
&= \frac{e^{-(\mu+\lambda)}}{n!} \sum_{k=0}^{n} \binom{n}{k} \mu^k \lambda^{n-k} \\
&= \frac{(\mu + \lambda)^n e^{-(\mu+\lambda)}}{n!}
\end{aligned}$$

On a donc par récurrence sur n, que

$$S_n = \sum_{i=0}^{n} X_i$$

suit la loi $\mathcal{P}(m_n)$, où $m_n = \sum_{i=0}^{n} \mu_i$. Ainsi, pour tout k,

$$\mathbb{P}\left(S_n \leq k\right) = \sum_{i=0}^{k} e^{-m_n} \frac{m_n^i}{i!}$$

.

De plus les événements $\{S_n \leq k\}$ sont décroissants et leur intersection est $\{S \leq k\}$, donc

$$\mathbb{P}\left(S \leq k\right) = \lim_{n \to \infty} \mathbb{P}(S_n \leq k),$$

et par continuité de $x \mapsto e^x \frac{x^i}{i!}$ on a

$$\mathbb{P}\left(S \leq k\right) = \sum_{i=0}^{k} e^{-\mu} \frac{\mu^i}{i!}.$$

Ainsi S est finie p.s. et suit la loi $\mathcal{P}(\mu)$. □

Montrons maintenant le théorème de superposition dans notre cas particulier.

Théorème 4. *Soit $(N_1, \cdots, N_n \cdots)$ des processus de Poissons indépendants sur $\mathbb{R}_+$ de paramètres respectifs l_n, avec l_n la longueur associée à la $n^{\text{ème}}$ boîte (et $\sum_n l_n = 1$), et tels que, pour tout i, $N_{n,t}$ est le nombre de balles tombées dans la boîte n pendant l'intervalle de temps $[0, t]$.*

Alors $(N_t)_{t \in \mathbb{R}_+}$, où $N_t = \sum_n N_{n,t}$, est un processus de poisson de paramètre 1.

Démonstration. Il est immédiat que $(N_t)_{t\in\mathbb{R}_+}$ est un processus de comptage qui compte le nombre de boules lancées dans l'ensemble des boîtes pendant l'intervalle de temps $[0,t]$. De plus, d'après la proposition précédente :

$$\forall t,s\in\mathbb{R}_+,\quad \mathbb{P}(N_{t+s}-N_t=i)=e^{-s}\frac{s^i}{i!}.$$

Il suffit donc de montrer que $(N_t)_{t\in\mathbb{R}}$ est à accroissements indépendants. Soit

$$t_1<t_2<t_3<t_4 \text{ et } i,j\in\mathbb{N}.$$

$$\begin{aligned}
&\mathbb{P}\left(N_{t_2}-N_{t_1}=i,\ N_{t_4}-N_{t_3}=j\right)\\
&= \mathbb{P}\left(\sum_{n=1}^{\infty}(N_{n,t_2}-N_{nt_1})=i,\ \sum_{n=1}^{\infty}(N_{n,t_4}-N_{n,t_3})=j\right)\\
&= \sum_{\substack{\sum_n k_n=i\\ \sum_n m_n=j\\ (k_n)_{n\in\mathbb{N}},(m_n)_{n\in\mathbb{N}}}} \mathbb{P}\left(\forall n\in\mathbb{N},\ N_{n,t_2}-N_{n,t_1}=k_n,\ N_{n,t_4}-N_{n,t_3}=m_n\right)\\
&= \sum_{(k_n),(m_n)}\prod_{n=1}^{\infty}\mathbb{P}\left(N_{n,t_2}-N_{n,t_1}=k_n,\ N_{n,t_4}-N_{n,t_3}=m_n\right)\\
&= \sum_{(k_n),(m_n)}\prod_{n=1}^{\infty}\mathbb{P}\left(N_{n,t_2}-N_{n,t_1}=k_n\right)\mathbb{P}\left(N_{n,t_4}-N_{n,t_3}=m_n\right)\\
&= \mathbb{P}\left(\sum_{n=1}^{\infty}(N_{n,t_2}-N_{n,t_1})=i\right)\mathbb{P}\left(\sum_{n=1}^{\infty}(N_{n,t_4}-N_{n,t_3})=j\right).
\end{aligned}$$

Ce qui achève la preuve. □

Cette propriété de superposition des processus de Poisson reste vraie dans un cadre plus général ; on pourra se référer à [5].

1.3 Notations et calcul des moments

On choisit de noter les quantités associées au problème continu (poissonisé) comme des fonctions du temps, réservant l'indice au problème originel (où la quantité de balles lancées est fixée). Ainsi on notera $X_j(t):=X_{N_t,j}$ le nombre de balles à l'instant t dans la $j^{ème}$ boîte. On note également

$$K(t):=K_{N_t}=\sum_j 1_{X_j(t)>0},\ \text{ et } K_r(t):=K_{N_t,r}=\sum_j 1_{X_j(t)=r}.$$

On note de même $\Phi(t):=\mathbb{E}[K(t)]$, et $\Phi_r(t):=\mathbb{E}[K_r(t)]$, et on a, par définition d'un processus de Poisson :

$$\Phi(t)=\sum_j\mathbb{E}\left[1_{X_j(t)>0}\right]=\sum_j(1-e^{-tp_j}),$$

et

$$\Phi_r(t)=\sum_j\mathbb{E}\left[1_{X_j(t)=r}\right]=\frac{t^r}{r!}\sum_j p_j^r e^{-tp_j}.$$

On peut aussi, grâce à l'indépendance des variables aléatoires $X_j(t)$, avoir des formules simples pour la variance :

$$\begin{aligned}
V(t) := \mathrm{Var}[K(t)] &= \sum_j \mathrm{Var}\left(1_{X_j(t)>0}\right) \\
&= \sum_j \left(\mathbb{E}\left[1^2_{X_j(t)>0}\right] - \mathbb{E}\left[1_{X_j(t)>0}\right]^2\right) \\
&= \sum_j \mathbb{E}\left[1_{X_j(t)>0}\right] - \sum_j \mathbb{E}\left[1_{X_j(t)>0}\right]^2 \\
&= \sum_j \left(1 - e^{-tp_j}\right) - \sum_j \left(1 - e^{-tp_j}\right)^2
\end{aligned}$$

D'où

$$V(t) = \Phi(2t) - \Phi(t) \tag{5}$$

et

$$\begin{aligned}
V_r(t) := \mathrm{Var}\left[K_r(t)\right] &= \sum_j \mathrm{Var}\left(1_{X_j(t)=r}\right) \\
&= \sum_j \left(\mathbb{E}\left(1^2_{X_j(t)=r}\right) - \mathbb{E}\left(1_{X_j(t)=r}\right)^2\right) \\
&= \sum_j \left(\mathbb{E}\left(1_{X_j(t)=r}\right) - \mathbb{E}\left(1_{X_j(t)=r}\right)^2\right) \\
&= \Phi_r(t) - \left(\frac{t^r}{r!}\right)^2 \sum_j p_j^{2r} e^{-2tp_j}
\end{aligned}$$

D'où,

$$V_r(t) = \Phi_r(t) - 2^{-2r}\binom{2r}{r}\Phi_{2r}(2t). \tag{6}$$

Enfin, on introduit la mesure suivante, permettant une plus grande clarté des calculs :

$$\nu(\mathrm{d}x) := \sum_j \delta_{p_j}(\mathrm{d}x),$$

mesure sur $[0,1]$, où δ_x représente la masse de Dirac en x. Ainsi, pour toute fonction positive f, on a :

$$\sum_j f(p_j) = \int_0^1 f(x)\nu(\mathrm{d}x)$$

On a alors :

$$\begin{aligned}
\Phi_n &= \int_0^1 \left(1 - (1-x)^n\right)\nu(\mathrm{d}x), && (7) \\
\Phi_{n,r} &= \binom{n}{r}\int_0^1 x^r(1-x)^{n-r}\nu(\mathrm{d}x), && (8) \\
\Phi(t) &= \int_0^1 \left(1 - e^{-tx}\right)\nu(\mathrm{d}x), && (9) \\
\Phi_r(t) &= \frac{t^r}{r!}\int_0^1 x^r e^{-tx}\nu(\mathrm{d}x). && (10)
\end{aligned}$$

2 Loi des grands nombres

2.1 Quelques résultats préliminaires

K_n et $K(t)$ sont croissantes et tendent vers l'infini quand n (respectivement t) tend vers l'infini : en effet chaque boîte est éventuellement découverte par une balle. Par convergence monotone on a de même $\Phi_n \uparrow \infty$ et $\Phi(t) \uparrow \infty$.

Cependant, on a toujours $\Phi_n \ll n$ $(n \to \infty)$ et $\Phi(t) \ll t$ $(t \to \infty)$. En effet, si on ignore les J premières boîtes et en notant D_J le nombre de boîtes découvertes parmi les restantes, on a :

$$D_J = \sum_{j>J} 1_{X_j>0} \quad \text{et} \quad \mathbb{E}\left[D_J\right] = \sum_{j>J} \mathbb{E}\left(1_{X_j>0}\right) = \sum_{j>J} \left(1-(1-p_j)^n\right).$$

L'inégalité classique $1-(1-x)^n \leq nx$ pour $0 \leq x \leq 1$ donne $\mathbb{E}[D_J] \leq n\sum_{j>J} p_j$, puis $\Phi_n \leq J + n\sum_{j>J} p_j$, et en choisissant $J(\varepsilon)$ assez grand on a $\Phi_n/n \leq \varepsilon$. En utilisant l'inégalité classique $1-e^{-nx} \leq nx$, on obtient de même $\Phi(t) \ll t$ $(t \to \infty)$.

2.2 Retour au problème originel ou dépoissonisation

La poissonisation facilite les calculs tout en restant une bonne approximation du problème initial. Ainsi le lemme suivant permet d'estimer la proximité des moments du modèle de Poisson de ceux du modèle originel.

Lemme 5. *Les inégalités suivantes sont vérifiées quand n tend vers l'infini.*

$$\begin{aligned}
|\Phi(n)-\Phi_n| &\leq \frac{2}{n}\Phi_2(n) \to 0,\\
|\Phi_r(n)-\Phi_{n,r}| &\leq \frac{c}{n}\max\left\{\Phi_r(n),\ \Phi_{r+1}(n),\ \Phi_{r+2}(n)\right\} \to 0,\\
|V(n)-V_n| &\leq \frac{c}{n}\max\left\{\Phi_{n,1}^2,\ \Phi_2(n),\ \Phi_2(2n)\right\}\\
&\leq \frac{d}{n}\max\left\{\Phi_2(n),\ \Phi_2(2n),\ \Phi_1^2(n),\ \Phi_3^2(n),\ \Phi_2^2(n)\right\},\\
|V_r(n)-V_{n,r}| &\leq \frac{c}{n}\max\left\{\Phi_r^2(n),\ \Phi_{r+1}^2(n),\ \Phi_{r+1}(n),\ \Phi_{r+2}(n),\ \Phi_{2r}(2n)\right\},
\end{aligned}$$

où c et d désignent des constantes.

Démonstration.

1. Pour démontrer les deux premières inégalités on utilise l'inégalité classique suivante :

$$0 \leq e^{-nx}-(1-x)^n \leq nx^2e^{-nx} \quad \text{pour} \quad 0 \leq x \leq 1. \tag{11}$$

En utilisant (7), (9), (10) et (11) on obtient :

$$\begin{aligned}
|\Phi(n)-\Phi_n| &= \int_0^1 (e^{-nx}-(1-x)^n)\nu(\mathrm{d}x)\\
&\leq n\int_0^1 x^2e^{-nx}\nu(\mathrm{d}x)\\
&= \frac{2}{n}\Phi_2(n) \leq \frac{2}{n}\Phi(n) \to 0.
\end{aligned}$$

2. De même, d'après (8) et (10), on a :

$$\begin{aligned}|\Phi_r(n) - \Phi_{n,r}| &= \left|\frac{n^r}{r!}\int_0^1 x^r e^{-nx}\nu(\mathrm{d}x) - \binom{n}{r}\int_0^1 x^r(1-x)^{n-r}\nu(\mathrm{d}x)\right| \\ &\leq \underbrace{\left|\frac{n^r}{r!} - \binom{n}{r}\right|\int_0^1 x^r e^{-nx}\nu(\mathrm{d}x)}_{a_1(n)} \\ &\quad + \underbrace{\binom{n}{r}\int_0^1 x^r\left(e^{-(n-r)x} - e^{-nx}\right)\nu(\mathrm{d}x)}_{a_2(n)} \\ &\quad + \underbrace{\binom{n}{r}\int_0^1 x^r\left(e^{-(n-r)x} - (1-x)^{n-r}\right)\nu(\mathrm{d}x)}_{a_3(n)}\end{aligned}$$

Le premier terme est évalué grâce à la formule de Stirling :

$$a_1(n) \sim \frac{r^2 r!}{2n}\Phi_r(n).$$

On majore le second terme grâce à l'inégalité classique $0 \leq 1 - e^{-rx} \leq rx$ pour x positif. Ainsi :

$$0 \leq a_2(n) \leq c\frac{\Phi_{r+1}(n)}{n}.$$

Enfin on majore le dernier terme grâce à l'inégalité (11),

$$0 \leq \ a_3(n) \leq c\frac{\Phi_{r+2}(n)}{n}.$$

3. Pour démontrer les deux autres inégalités on utilise le résultat suivant :

$$(a-b)^n \leq a^n - na^{n-1}b + cn^2a^{n-2}b^2, \quad (a > b) \tag{12}$$

où c est une constante indépendante de n, de a et de b. D'après (3) et (5) on a :

$$\begin{aligned}|V_n - V(n)| &\leq |\Phi_{2n} - \Phi(2n)| + |\Phi_n - \Phi(n)| \\ &\quad + \sum_{i\neq j}\left[(1-p_i)^n(1-p_j)^n - (1-p_i-p_j)^n\right].\end{aligned}$$

La majoration des deux premiers termes utilise le premier point de ce lemme. Pour majorer le dernier terme on utilise (12) avec

$$a = (1-p_i)(1-p_j) \text{ et } b = p_ip_j.$$

On obtient alors :

$$\begin{aligned}
&\sum_{i\neq j}\left[(1-p_i)^n(1-p_j)^n-(1-p_i-p_j)^n\right]\\
&\quad\leq\ n\sum_{i\neq j}\left(((1-p_i)(1-p_j))^{n-1}\,p_ip_j\right)\\
&\qquad+n^2c\sum_{i\neq j}\left(((1-p_i)(1-p_j))^{n-2}\,p_i^2p_j^2\right)\\
&\quad\leq\ n\int_0^1\int_0^1(1-x)^{n-1}(1-y)^{n-1}xy\nu(\mathrm{d}x)\nu(\mathrm{d}y)\\
&\qquad+cn^2\int_0^1\int_0^1(1-x)^{n-2}(1-y)^{n-2}x^2y^2\nu(\mathrm{d}x)\nu(\mathrm{d}y)\\
&\quad\leq\ \frac{1}{n}\Phi_{n,1}^2+c\frac{1}{n^2}\Phi_{n,2}^2.
\end{aligned}$$

et le deuxième point de ce lemme permet de conclure.

4. On démontre de même la dernière inégalité en utilisant (4), (6) et (12).

□

Comme $\Phi(n)\uparrow\infty$ la première inégalité implique $\Phi_n \underset{n\to\infty}{\sim} \Phi(n)$.

2.3 Loi des grands nombres

Le nombre moyen de boîtes occupées satisfait l'inégalité suivante :

$$\Phi(t+s)-\Phi(t)\leq\Phi(s)\quad\forall s,t>0.$$

En effet, il y a en moyenne $\Phi(t)$ boîtes distinctes qui ont été touchées par au moins une balle pendant l'intervalle de temps $[s,s+t]$ (un processus de Poisson est à accroissements constants). Cependant certaines de ces boîtes n'étaient pas vides et avaient pu être remplies avant l'instant s et ne contribuent donc pas à $K(t+s)-K(s)$.

On peut aussi remarquer que Φ est concave d'après (9). Le même raisonnement garantit que Φ_n vérifie la même inégalité. On utilise cette inégalité pour majorer les variances comme suit :

$$V_n\leq\Phi_{2n}-\Phi_n\leq\Phi_n=\mathbb{E}\left[K_n\right] \tag{13}$$

et

$$V(t)=\Phi(2t)-\Phi(t)\leq\Phi(t)=\mathbb{E}\left[K(t)\right]. \tag{14}$$

Lemme 6. *Les quantités $K(t)/\Phi(t)$ et K_n/Φ_n convergent en probabilité vers 1 et pour tout $s\geq1$, les quantités suivantes*

$$\frac{\sum_{r\geq s}K_{n,r}}{\sum_{r\geq s}\Phi_{n,r}}\quad et\quad\frac{\sum_{r\geq s}K_r(t)}{\sum_{r\geq s}\Phi_r(t)}$$

convergent en probabilité vers 1. De plus $K_r(t)/\Phi_r(t)$ et $K_{n,r}/\Phi_{n,r}$ convergent en probabilité vers 1 si $\Phi_r(t)\to\infty$.

Démonstration.

1. Soit ε positif, on a d'après l'inégalité de Bienaymé-Tchebichev et (14) :

$$\mathbb{P}\left(\left|\frac{K(t)}{\Phi(t)}-1\right|>\varepsilon\right)\leq\frac{V(t)}{\varepsilon^2\Phi(t)^2}\leq\frac{1}{\varepsilon^2\Phi(t)}.$$

   Donc, comme $\Phi(t)\to\infty$,

$$\mathbb{P}\left(\left|\frac{K(t)}{\Phi(t)}-1\right|>\varepsilon\right)\to 0,$$

   et $K(t)/\Phi(t)$ converge en probabilité vers 1. De même K_n/Φ_n converge en probabilité vers 1.

2. Soit $s\geq 1$. Comme on peut écrire

$$\sum_{r\geq s}K_r(t)=\sum_j 1_{X_j(t)\geq s},$$

   c'est donc une somme de variables de Bernoulli indépendantes et de paramètre $\sum_{r\geq s}e^{-tp_j}\frac{(tp_j)^r}{r!}$. On a donc :

$$\begin{aligned}\mathrm{Var}\left(\sum_{r\geq s}K_r(t)\right) &= \sum_j\sum_{r\geq s}\left(e^{-tp_j}\frac{(tp_j)^r}{r!}\right)-\sum_j\left(\sum_{r\geq s}e^{-tp_j}\frac{(tp_j)^r}{r!}\right)^2\\ &\leq \mathbb{E}\left(\sum_{r\geq s}K_r(t)\right),\end{aligned}$$

   puis, en utilisant l'inégalité de Bienaymé-Tchebichev :

$$\mathbb{P}\left(\left|\frac{\sum_{r\geq s}K_r(t)}{\sum_{r\geq s}\Phi_r(t)}-1\right|>\varepsilon\right)\leq\frac{1}{\varepsilon^2\sum_{r\geq s}\Phi_r(t)}\underset{t\to\infty}{\to}0$$

   Pour le problème originel on va "dépoissoniser" le problème. On pose

$$A_{n,s}:=\sum_{r\geq s}K_{n,r}\quad\text{et}\quad A_s(t):=\sum_{r\geq s}K_r(t),$$

$$c_n:=n+cn^{1/2}\quad\text{et}\quad d_n:=n-cn^{1/2}.$$

   Comme on peut écrire

$$N_n=\sum_{k=0}^{n-1}N_{k+1}-N_k,$$

   et qu'un processus de Poisson est à accroissements indépendants et stationnaires, le théorème central limite nous garantit que pour $c(\varepsilon)$ assez grand

$$\mathbb{P}\left(d_n\leq N_n\leq c_n\right)\underset{n\to\infty}{\to}\int_{-c}^{c}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}\mathrm{d}x\geq 1-\varepsilon.$$

Comme $(A_{n,s})_n$ est croissante, on a :

$$\begin{aligned}
&\mathbb{P}\left(A_{c_n,s} > (1+\eta)\sum_j \sum_{r\geq s}(np_j)^r \frac{e^{-np_j}}{r!}\right) \\
&\quad \leq \quad \mathbb{P}\left(c_n < N_n\right) + \mathbb{P}\left(A_s(n) > (1+\eta)\sum_j \sum_{r\geq s}(np_j)^r \frac{e^{-np_j}}{r!}\right),
\end{aligned}$$

puis

$$\mathbb{P}\left(A_{c_n,s} > (1+\eta)\sum_j \sum_{r\geq s}(np_j)^r \frac{e^{-np_j}}{r!}\right) \leq 3\varepsilon.$$

On a de même

$$\mathbb{P}\left(A_{d_n,s} < (1-\eta)\sum_j \sum_{r\geq s}(np_j)^r \frac{e^{-np_j}}{r!}\right) \leq 3\varepsilon,$$

et ainsi,

$$\mathbb{P}\left(\left|A_{n,s} - \sum_j \sum_{r\geq s}(np_j)^r \frac{e^{-np_j}}{r!}\right| < \eta \sum_j \sum_{r\geq s}(np_j)^r \frac{e^{-np_j}}{r!}\right) \to 1$$

On conclut en remarquant que :

$$\mathbb{E}\left(A_{n,s}\right) \underset{n\to\infty}{\sim} \mathbb{E}\left(A_s(n)\right).$$

3. D'après (6), $V_r(t) \leq \Phi_r(t)$. On obtient donc, en utilisant l'inégalité de Bienaymé-Tchebichev, que $K_r(t)/\Phi_r(t)$ converge en probabilité vers 1 si $\Phi_r(t) \to \infty$. D'après le lemme 5, si $\Phi_r(t) \to \infty$ alors $\Phi_{n,r} \sim \Phi_r(n)$ quand $n \to \infty$. D'après l'inégalité de Bienaymé-Tchebichev :

$$\mathbb{P}\left(\left|\frac{K_{n,r}}{\Phi_{n,r}} - 1\right| > \varepsilon\right) \leq \frac{V_{n,r}}{\varepsilon^2 \Phi_{n,r}^2},$$

et du lemme 5 on déduit :

$$\begin{aligned}
V_{n,r} \quad &\leq \quad V_r(n) \\
&\qquad + \frac{c}{n}\max\left\{\Phi_r(n)^2,\ \Phi_{r+1}(n)^2,\ \Phi_{r+1}(n),\ \Phi_{r+2}(n),\ \Phi_{2r}(2n)\right\} \\
&\leq \quad V_r(n) + \frac{c'\max\left\{\Phi_r(n),\ \Phi_r(n)^2\right\}}{n},
\end{aligned}$$

où c et c' désignent deux constantes. La dernière inégalité découle de (10) dont on déduit que $r \mapsto r!\Phi_r(n)$ est croissante et que

$$\Phi_{2r}(2n) \leq \frac{r!}{(2r)!}2^r \Phi_r(n).$$

Ainsi si $\Phi_r(t) \to \infty$, on a $\Phi_{n,r} \sim \Phi_r(n)$ et

$$\mathbb{P}\left(\left|\frac{K_{n,r}}{\Phi_{n,r}} - 1\right| > \varepsilon\right) \leq \frac{V_r(n)}{\varepsilon^2 \Phi_{n,r}^2} + \frac{c'\max\left\{\Phi_r(n),\ \Phi_r(n)^2\right\}}{n\varepsilon^2\Phi_{n,r}^2},$$

ainsi $K_{n,r}/\Phi_{n,r}$ converge en probabilité vers 1.

$\square$

On peut maintenant énoncer le théorème fondamental de cette section, une loi des grands nombres :

Théorème 7 (Loi des grands nombres)**.** *On a toujours* $K_n \sim_{p.s.} \Phi_n \ (n \to \infty)$ *et* $K(t) \sim_{p.s.} \Phi(t) \ (t \to \infty)$*. On a également pour tout* $s \geq 1$

$$\sum_{r \geq s} K_{n,r} \underset{n\to\infty}{\sim} \sum_{r \geq s} \Phi_{n,r} \quad et \quad \sum_{r \geq s} K_r(t) \underset{t\to\infty}{\sim} \sum_{r \geq s} \Phi_r(t) \quad p.s.$$

Cependant, on n'a pas nécessairement $K_{n,r} \sim_{p.s.} \Phi_{n,r}$ *et* $K_r(t) \sim_{p.s.} \Phi_r(t)$*.*

Démonstration. La fonction Φ est continue, strictement croissante et tend vers ∞ quand t tend vers ∞, il est donc possible de construire une suite croissante $(t_m, m \in \mathbb{N}^*)$ telle que pour tout m on ait $m^2 < \Phi(t_m) < m^2 + 1$. D'après (14) et l'inégalité de Bienaymé-Tchebichev, on a pour tout m :

$$\forall \varepsilon > 0 \quad \mathbb{P}\left(\left| \frac{K(t_m)}{\Phi(t_m)} - 1 \right| > \varepsilon \right) < \varepsilon^{-2} m^{-2},$$

puis d'après le lemme de Borel-Cantelli

$$\frac{K(t_m)}{\Phi(t_m)} \underset{m\to\infty}{\to} 1 \quad \text{p.s.}$$

Pour $t_m < t < t_{m+1}$, comme K et Φ sont croissantes, on a

$$K(t_m) \leq K(t) \leq K(t_{m+1}) \text{ et } \Phi(t_m) \leq \Phi(t) \leq \Phi(t_{m+1}).$$

Ainsi :

$$\frac{K(t_m)}{\Phi(t_{m+1})} \leq \frac{K(t)}{\Phi(t)} \leq \frac{K(t_{m+1})}{\Phi(t_m)},$$

où les deux côtés convergent presque sûrement vers 1. Ainsi

$$K(t) \underset{t\to\infty}{\sim} \Phi(t) \quad \text{p.s.}$$

Un raisonnement analogue permet de conclure que

$$K_n \underset{n\to\infty}{\sim} \Phi_n \quad \text{p.s.}$$

De même on a

$$\sum_{r \geq s} K_{n,r} \underset{n\to\infty}{\sim} \sum_{r \geq s} \Phi_{n,r} \quad \text{p.s.} \quad \text{et} \quad \sum_{r \geq s} K_r(t) \underset{t\to\infty}{\sim} \sum_{r \geq s} \Phi_r(t) \quad \text{p.s.}$$

$\square$

3 Théorème central limite

3.1 Théorème central limite pour le problème "poissonisé"

$K(n)$ vérifie un théorème central limite, que l'on démontre à partir de la condition de Lindeberg-Feller.

Proposition 8 (Théorème central limite pour $K(n)$)**.** *La suite de variables aléatoires*

$$\left((K(n)-\Phi(n))/V(n)^{\frac{1}{2}},\ n\in\mathbb{N}^*\right)$$

converge en loi vers $\mathfrak{N}$ si $V(n)\to\infty$, où $\mathfrak{N}$ est une variable aléatoire suivant la loi normale centrée et réduite.

Démonstration. On commence par démontrer que $K(n)$ vérifie la condition de Lindeberg-Feller puis on démontrera que cette condition entraîne bien le théorème.

On note pour tout $n,k\in\mathbb{N}^*$

$$x_{n,k} := \frac{1_{X_k(n)}-\mathbb{E}(1_{X_k(n)})}{\sqrt{\sum_{k=1}^{\infty}\mathrm{Var}(1_{X_k(n)})}} = \frac{1_{X_k(n)}-(1-e^{-np_k})}{\sqrt{V(n)}}, \tag{15}$$

$F_{n,k}$ la loi de $x_{n,k}$ et

$$u_n := \sum_{k=1}^{\infty} x_{n,k} = \frac{K(n)-\Phi(n)}{\sqrt{V(n)}}.$$

Les variables $x_{n,k}$ sont indépendantes et centrées et donc u_n est centrée et normalisée. On veut montrer que u_n converge en loi vers $\mathfrak{N}$.

- On montre d'abord que pour tout $t>0$,

$$\lim_{n\to\infty}\sum_{k=1}^{\infty}\int_{|x|\geq t} x^2 F_{n,k}(\mathrm{d}x)=0. \tag{16}$$

  Lorsque les $F_{n,k}$ vérifient (16) on dit qu'ils vérifient la condition de Lindeberg.
  Soit $t>0$ et $\varepsilon>0$. D'après (15), pour tout $n,k\in\mathbb{N}^*$

$$x_{n,k} = \begin{cases} \frac{e^{-np_k}}{\sqrt{V(n)}} & \text{avec probabilité} \quad 1-e^{-np_k} \\ \frac{-(1-e^{-np_k})}{\sqrt{V(n)}} & \text{avec probabilité} \quad e^{-np_k} \end{cases}$$

  ainsi

$$\forall n,k\in\mathbb{N}^*,\quad |x_{n,k}|\leq\frac{2}{\sqrt{V(n)}},$$

  et comme $V(n)\to\infty$ la condition de Lindeberg est bien vérifiée.
- On peut alors montrer le théorème. Soit $\psi_{n,k}$ la fonction caractéristique de $x_{n,k}$ et ψ_n celle de u_n. Comme

$$|x_{n,k}|\leq 2/\sqrt{V(n)} \quad \text{alors} \quad \mathrm{Var}(x_{n,k}) \underset{n\to\infty}{\to} 0 \quad \text{uniformément en } k.$$

  On fixe ξ. Pour tout $k,n\in\mathbb{N}^*$, $x_{n,k}$ est centrée et on a donc :

$$\psi_{n,k}(\xi)-1=\int_{-\infty}^{\infty}\left(e^{i\xi x}-1-i\xi x\right)F_{n,k}(\mathrm{d}x),$$

  puis l'inégalité de Taylor à l'ordre 1 donne, pour n assez grand (avec uniformité en k) :

$$|\psi_{n,k}(\xi)-1|\leq\frac{1}{2}\xi^2\mathrm{Var}(x_{n,k})<\frac{1}{2}\varepsilon.$$

Comme

$$\sum_{k=1}^{\infty} |\log(1+z_{n,k})-z_{n,k}| \leq \sum_{k=1}^{\infty} |z_{n,k}|^2 \leq \varepsilon \sum_{k=1}^{\infty} |z_{n,k}| \quad \text{pour} \quad |z_{n,k}| < \varepsilon < \frac{1}{2},$$

où on a posé $z_{n,k} = \psi_{n,k}(\xi) - 1$, on a :

$$\sum_{k=1}^{\infty} \log(\psi_{n,k}(\xi)) \underset{n\to\infty}{\sim} \sum_{k=1}^{\infty} (1 - \psi_{n,k}) \, .$$

Comme $\log(\psi_n(\xi)) = \sum_{k=1}^{\infty} \log(\psi_{n,k}(\xi))$, il suffit de montrer que

$$\sum_{k=1}^{\infty} (1 - \psi_{n,k}) \to -\frac{1}{2}\xi^2.$$

Pour ce faire on considère :

$$\begin{aligned}
\left|\psi_{n,k} - 1 + \frac{1}{2}\mathrm{Var}(x_{n,k})\xi^2\right| &= \left|\int_{-\infty}^{\infty} \left(e^{i\xi x} - 1 - i\xi x + \frac{\xi^2 x^2}{2}\right) F_{n,k}(\mathrm{d}x)\right| \\
&\leq \int_{|x|<\varepsilon} \left|e^{i\xi x} - 1 - i\xi x + \frac{\xi^2 x^2}{2}\right| F_{n,k}(\mathrm{d}x) \\
&\quad + \int_{|x|\geq\varepsilon} \left|e^{i\xi x} - 1 - i\xi x + \frac{\xi^2 x^2}{2}\right| F_{n,k}(\mathrm{d}x) \\
&\leq \frac{\varepsilon|\xi|^3 \mathrm{Var}(x_{n,k})}{6} + |\xi|^2 \int_{|x|\geq\varepsilon} x^2 F_{n,k}(\mathrm{d}x),
\end{aligned}$$

où on a utilisé pour la première intégrale l'inégalité de Taylor à l'ordre 2 et à l'ordre 1 pour la deuxième intégrale. D'après la condition de Lindeberg (16)

$$\lim_{n\to\infty} \sum_{k=1}^{\infty} (1 - \psi_{n,k}(\xi)) = -\frac{1}{2} \sum_{k=1}^{\infty} \mathrm{Var}(x_{n,k})\xi^2,$$

puis, comme

$$\sum_{k=1}^{\infty} \mathrm{Var}(x_{n,k}) = 1$$

on a bien

$$\lim_{n\to\infty} \log(\psi_n(\xi)) = -\frac{1}{2}\xi^2.$$

□

La condition de Lindeberg utilisée et démontrée dans la démonstration du théorème 8 n'est pas la condition de Lindeberg "classique" que l'on énonce ci-après ; sa démonstration reprend les mêmes idées. On pourra se réferer à [3].

Théorème 9 (Condition de Lindeberg "classique")**.** *Soit $X_1, X_2, \ldots$ des variables aléatoires indépendantes et centrées. On note leur fonctions de répartition $F_1, F_2, \ldots$. On pose $\sigma_k^2 := Var(X_k)$ et $s_n^2 := \sigma_1^2 + \cdots + \sigma_n^2$. On suppose de plus que pour tout $\varepsilon > 0$,*

$$\lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{n} \int_{|x|>\varepsilon s_n} x^2 F_k(dx) = 0. \tag{17}$$

Si cette hypothèse (condition de Lindeberg-Feller) est vérifiée, alors la somme normalisée
$$S_n^* := \frac{X_1 + \cdots + X_n}{s_n}$$
converge en loi vers $\mathfrak{N}$.

Feller a énoncé une réciproque à cette condition à savoir : si $S_n^* \to \mathfrak{N}$ et que $\sigma_n/s_n \to 0$, $s_n \to \infty$, alors la condition de Lindeberg (17) est vérifiée.

Le théorème central limite pour le problème poissinisé va nous permettre de déduire le théorème central limite pour le problème originel. Il faut pour cela "dépoissoniser".

3.2 Dépoissonisation

Le théorème central limite pour le problème poissonisé fait apparaître la condition $V(n) \to \infty$. On regarde donc ce que signifie cette condition pour le problème originel ; c'est l'objet du lemme suivant.

Lemme 10. *Les conditions* $V_n \to \infty$ *et* $V(n) \to \infty$ *sont équivalentes et si elles sont vérifiées on a* $V_n \sim V(n)$.

Démonstration. On pose $\vec{\nu}(x) := \nu([x, \infty[)$.

On montre d'abord que $\Phi_1(t)^2 \ll tV(t)$. On note $\psi(t) := t^{-1}\Phi_1(t) = \Phi'(t)$, et d'après (9), on a $\psi'(t) < 0$. De plus comme $\vec{\nu}(x) < \infty$ pour $x > 0$, on a
$$\psi(t) \underset{t\to\infty}{\sim} \int_0^\varepsilon x e^{-tx} \nu(\mathrm{d}x)$$
pour tout $\varepsilon > 0$. En appliquant l'inégalité de Cauchy-Schwarz à la mesure $x\nu(\mathrm{d}x)$ on obtient :
$$\psi(t)^2 \sim \left[\int_0^\varepsilon x e^{-tx}\nu(\mathrm{d}x)\right]^2 \leq \int_0^\varepsilon x\nu(\mathrm{d}x) \int_0^\varepsilon x e^{-2tx}\nu(\mathrm{d}x) \sim \psi(2t)\int_0^\varepsilon x\nu(\mathrm{d}x),$$
et en faisant tendre ε vers 0 on obtient
$$\psi(t)^2 \ll \psi(2t). \tag{18}$$
D'après (5) et comme ψ est décroissante on a bien ce qu'on voulait :
$$V(t) = \Phi(2t) - \Phi(t) = \int_t^{2t} \psi(u)\mathrm{d}x > t\psi(2t) \gg t\psi(t)^2 = t^{-1}\Phi_1(t)^2.$$
Donc si $V(n) \to \infty$ on déduit de cette estimation et du lemme 5 que $V_n \sim V(n)$ et en particulier que $V_n \underset{n\to\infty}{\to} \infty$.

Réciproquement, si $V_n \to \infty$, alors $\Phi_{2n} - \Phi_n \to \infty$ car le terme mixte dans (3) est négatif. Ainsi d'après le lemme 5 on a également
$$V(n) = \Phi(2n) - \Phi(n) \to \infty.$$
□

On peut alors énoncer le théorème central limite pour le problème originel :

Théorème 11 (Théorème central limite pour K_n). *Les conditions $V(n) \to \infty$ et $V_n \to \infty$ sont équivalentes et si elles sont vérifiées, $(K_n - a_n)/b_n^2$ converge en loi vers $\mathfrak{N}$, où on peut choisir pour a_n soit $\Phi(n)$ soit Φ_n et pour b_n soit $V(n)$ soit V_n.*

Démonstration. On reprend les notations de la preuve du lemme 10. On travaille avec $c > 0$ quelconque et d'après (18) on a, pour tout $t > 0$:

$$\begin{aligned} \frac{\Phi(n + cn^{1/2}) - \Phi(n)}{(V(n))^{1/2}} &= \frac{\Phi(n + cn^{1/2}) - \Phi(n)}{(\Phi(2n) - \Phi(n))^{1/2}} \\ &= \frac{\int_n^{n+cn^{1/2}} \psi(u)\mathrm{d}u}{\left(\int_n^{2n} \psi(u)\mathrm{d}u\right)^{1/2}} \leq \frac{cn^{1/2}\psi(n)}{(n\psi(2n))^{1/2}} \to 0 \end{aligned} \tag{19}$$

On a donc

$$\frac{V(n + cn^{1/2})}{V(n)} \to 1 \tag{20}$$

si $V(n) \to \infty$. En effet,

$$\begin{aligned} \frac{V(n + cn^{1/2})}{V(n)} &= \frac{\Phi(2n + 2cn^{1/2}) - \Phi(n + cn^{1/2})}{\Phi(2n) - \Phi(n)} \\ &= \left(\frac{\Phi(2n + 2cn^{1/2}) - \Phi(2n)}{\Phi(4n) - \Phi(2n)}\right)\left(\frac{\Phi(4n) - \Phi(2n)}{\Phi(2n) - \Phi(n)}\right) \\ &\quad - \frac{\Phi(n + cn^{1/2}) - \Phi(n)}{\Phi(2n) - \Phi(n)} + 1. \end{aligned} \tag{21}$$

De plus, comme Φ est concave on a $(\Phi(4t) - \Phi(2t))/(\Phi(2t) - \Phi(t)) < 2$. Ce dernier fait, combiné avec (21) et (19) donne (20).

Du lemme 5 et de l'inégalité (19), on déduit :

$$\begin{aligned} \mathbb{P}\left(K_{n+cn^{1/2}} - K_n > \varepsilon V(n)^{1/2}\right) &\leq \frac{\Phi_{n+cn^{1/2}} - \Phi_n}{\varepsilon V(n)^{1/2}} \\ &= \frac{\Phi(n + cn^{1/2}) - \Phi(n)}{\varepsilon V(n)^{1/2}} + o(1) \to 0. \end{aligned}$$

De même que pour (19), on majore

$$\mathbb{P}\left(K_n - K_{n-cn^{1/2}} > \varepsilon V(n)^{1/2}\right)$$

en utilisant l'inégalité de Markov, le lemme 5 et (18) :

$$\begin{aligned}
\mathbb{P}\left(K_n - K_{n-cn^{1/2}} > \varepsilon V(n)^{1/2}\right) &\leq \frac{\Phi_n - \Phi_{n-cn^{1/2}}}{\varepsilon V(n)^{1/2}} \\
&\leq \frac{(\Phi(n) - \Phi_n) + \left(\Phi_{n-cn^{1/2}} - \Phi(n-cn^{1/2})\right)}{\varepsilon V(n)^{1/2}} \\
&\quad + \frac{\Phi(n) - \Phi(n-cn^{1/2})}{\varepsilon V(n)^{1/2}} \\
&\leq \frac{\int_{n-cn^{1/2}}^{n} \psi(u)\mathrm{d}u}{\varepsilon \left(\int_n^{2n} \psi(u)\mathrm{d}u\right)^2} + o(1) \\
&\leq \frac{cn^{1/2}\psi(n-cn^{1/2})}{\varepsilon \left(n\psi(2n)\right)^{1/2}} + o(1) \\
&\ll \frac{c\psi(2n-2cn^{1/2})^{1/2}}{\varepsilon\psi(2n)^{1/2}} \leq 1,
\end{aligned}$$

où la dernière inégalité découle de la décroissance de ψ.

Les calculs précedents permettent d'affirmer que

$$\left|K_{n\pm cn^{1/2}} - K_n\right| / V(n)^{1/2}$$

converge en probabilité vers 0. De plus, d'après le théorème central limite classique, on peut choisir un c suffisamment grand pour que

$$\mathbb{P}\left(n - cn^{1/2} < N_n < n + cn^{1/2}\right) \geq 1 - \varepsilon,$$

puis

$$\mathbb{P}(K_{n-cn^{1/2}} \leq K(n) \leq K_{n+cn^{1/2}}) \geq 1 - \varepsilon.$$

Ainsi $(K_n - K(n)) / V(n)^{1/2}$ converge en probabilité vers 0. Le théorème central limite pour $K(n)$ (théorème 8) permet de conclure, le lemme 5 garantissant le choix pour a_n alors que le lemme 10 garantit celui de b_n.

□

Cependant, la condition $V(n) \to \infty$ n'est pas toujours satisfaite, comme on le voit dans l'exemple développé ci-dessous.

Exemple 12 (La condition du lemme 10 n'est pas toujours vérifiée). *Si on a $p_i = 2^{-i}$ pour tout entier naturel non nul i, alors*

$$\forall n \quad V(n) = \Phi(2n) - \Phi(n) = n \int_0^\infty e^{-nx} \nabla\nu(x)\, dx = 1$$

avec

$$\nabla\nu(x) := \nu\left([x/2, \infty[\right) - \nu\left([x, \infty[\right) = \mathit{Card}\left(\{j : x/2 \leq p_j < x\}\right).$$

Ainsi, pour tout $x \in]0, 1[$, $\nabla\nu(x) = 1$.

4 Fonctions variant régulièrement et applications

La théorie sur les fonctions à variation régulière va nous permettre d'avoir de plus amples connaissances sur le comportement asymptotique du schéma d'occupation d'une infinité de boîtes dans certains cas.

4.1 Définitions et propriétés

Definition Une fonction l positive, mesurable, définie sur un voisinage de $+\infty$ du type $[X, \infty[$ varie lentement si et seulement si

$$\frac{l(\lambda x)}{l(x)} \underset{x\to+\infty}{\to} 1 \quad \forall \lambda > 0.$$

Par exemple, log et $1/\log$ sont deux fonctions variant lentement, l'une tendant vers $+\infty$ en ∞ et l'autre vers 0. Les fonctions qui convergent vers une limite non nulle en $+\infty$ varient lentement.

On suppose souvent que l est définie sur $\mathbb{R}_+$ car modifier l sur $[0, X]$ n'altère pas son comportement asymptotique.

Théorème 13 (Convergence Uniforme)**.** *Si l varie lentement, alors*

$$\frac{l(\lambda x)}{l(x)} \underset{x\to\infty}{\longrightarrow} 1$$

uniformément sur tout compact de $[0, \infty[$ que λ parcourt.

Démonstration. On pose $h(x) := log(l(e^x))$ et on a par hypothèse sur l,

$$h(x+u) - h(x) \underset{x\to\infty}{\longrightarrow} 0 \quad \forall u \in \mathbb{R}.$$

Il s'agit de montrer l'uniforme convergence sur tous les compacts de $\mathbb{R}$ que u parcourt. On va montrer l'uniforme convergence sur $[0, A]$ avec $A > 0$ et le résultat s'en déduira par translation.

On prend $\varepsilon \in]0, A[$. Pour $x > 0$, on pose

$$\begin{aligned} I_x &:= [x, x+2A] \\ E_x &:= \left\{ t \in I_x : |h(t) - h(x)| \geq \frac{1}{2}\varepsilon \right\} \\ E_x^* &:= \left\{ t \in [0, 2A] : |h(x+t) - h(x)| \geq \frac{1}{2}\varepsilon \right\} \end{aligned}$$

Comme l est mesurable, E_x et E_x^* le sont aussi et sont de mesures (de Lebesgue) égales.

Soit $t \in \mathbb{R}$,

$$1_{E_x^*}(t) \underset{x\to\infty}{\to} 0 \quad \text{et} \quad 1_{E_x^*} \leq 1_{[0,2A]},$$

par convergence dominée on obtient que la mesure de E_x^*, notée $|E_x^*|$, tend vers 0 quand x tend vers l'infini. On a donc pour un certain x_0,

$$\forall x \geq x_0, \quad |E_x| < \frac{1}{2}\varepsilon.$$

Soit $c \in [0, A]$, $|I_x \cap I_{x+c}| = 2A - c \geq A$, tandis que

$$|E_x \cup E_{x+c}| \leq |E_x| + |E_{x+c}| \leq \varepsilon < A,$$

donc $(I_x \cap I_{x+c}) \backslash (E_x \cup E_{x+c})$ est de mesure non nulle donc est non vide. Soit $u \in (I_x \cap I_{x+c}) \backslash (E_x \cup E_{x+c})$, on a :

$$|h(u) - h(x)| < \frac{1}{2}\varepsilon \quad \text{et} \quad |h(u) - h(x+c)| < \frac{1}{2}\varepsilon,$$

et donc

$$|h(x) - h(x+c)| < \varepsilon,$$

ce qui achève la preuve.

□

En fait, toutes les fonctions à variation lente possèdent la représentation suivante :

$$l(x) = c(x) \exp\left(\int_a^x \frac{\varepsilon(t)}{t} dt\right)$$

où $x \geq a > 0$, $c(.)$ est mesurable et converge vers une limite strictement positive en l'infini, et $\varepsilon(.)$ converge vers 0 en l'infini. On pourra se reporter à [1].

Definition Une fonction $f > 0$ mesurable, définie sur un voisinage de $+\infty$ du type $[X, \infty[$ est à variation régulière d'indice $\rho \in \mathbb{R}$ si

$$\forall \lambda > 0, \quad \frac{f(\lambda x)}{f(x)} \underset{x \to +\infty}{\to} \lambda^\rho$$

On a en fait une extension du théorème d'uniforme continuité que l'on admet. On pourra se réferer à [1].

Théorème 14. *Si f est à variation régulière d'indice ρ, alors*

$$\frac{f(\lambda x)}{f(x)} \underset{x \to \infty}{\to} \lambda^\rho \quad \textit{uniformément en } \lambda$$

$$\textit{sur chaque intervalle de la forme} \begin{cases} [a, b] & (a \leq b) & \textit{si} & \rho = 0 \\ ]0, b] & (0 < b) & \textit{si} & \rho > 0 \\ [a, \infty[& (0 < a) & \textit{si} & \rho < 0 \end{cases}$$

On peut maintenant énoncer un théorème qui va nous être très utile.

Théorème 15 (Théorème de Karamata)**.** *Soit $l > 0$ localement intégrable sur $[X, \infty[$. Si $\alpha + 1 < 0$, alors l varie lentement si et seulement si*

$$\frac{x^{\alpha+1} l(x)}{\int_x^\infty t^\alpha l(t) dt} \underset{x \to \infty}{\longrightarrow} |\alpha + 1|.$$

De manière similaire, si $\alpha + 1 > 0$, alors l varie régulièrement si et seulement si

$$\frac{x^{\alpha+1} l(x)}{\int_0^x t^\alpha l(t) dt} \underset{x \to \infty}{\longrightarrow} \alpha + 1.$$

Démonstration.

- sens direct : Supposons que l varie lentement. Soit

$$\rho = \frac{\alpha+1}{2} < 0,$$

alors $f(x) := x^{\rho} l(x)$ est à variation régulière d'indice ρ. On a :

$$\frac{\int_x^\infty t^\alpha l(t)dt}{x^{\alpha+1}l(x)} + \frac{1}{\alpha+1} = \int_1^\infty \left(\frac{f(ux)}{f(x)} - u^\rho\right) u^{\rho-1} du.$$

D'après le Théorème 14,

$$\frac{f(ux)}{f(x)} - u^\rho \underset{x\to\infty}{\longrightarrow} 0 \quad \text{uniformément en } u.$$

De plus, $u^{\rho-1}$ étant intégrable sur $[1,\infty[$, l'intégrale du terme de droite tend vers 0.
- sens indirect : Soit

$$g(x) := \frac{x^{\alpha+1}l(x)}{\left(\int_x^\infty t^\alpha l(t)dt\right)}.$$

On a :

$$\begin{aligned}\int_X^x \frac{g(t)}{t}dt &= \int_X^x \frac{t^\alpha l(t)dt}{\int_t^\infty u^\alpha l(u)du} \\ &= \log\left(\frac{\int_X^\infty t^\alpha l(t)dt}{\int_x^\infty t^\alpha l(t)dt}\right).\end{aligned}$$

Alors,

$$\begin{aligned} l(x) &= \left(\int_x^\infty t^\alpha l(t)\, dt\right) x^{-(\alpha+1)} g(x) \\ &= \left(\int_X^\infty t^\alpha l(t) dt\right) x^{-(\alpha+1)} g(x) \exp\left(-\int_X^x \frac{g(t)}{t} dt\right) \\ &= \left(\int_X^\infty t^\alpha l(t) dt\right) X^{-(\alpha+1)} g(x) \exp\left(-\int_X^x \frac{g(t)+\alpha+1}{t} dt\right). \end{aligned}$$

On a donc, pour $\lambda > 0$

$$\frac{l(\lambda x)}{l(x)} = \frac{g(\lambda x)}{g(x)} \exp\left(-\int_x^{\lambda x} \frac{g(t)+\alpha+1}{t}dt\right).$$

Or,

$$\frac{g(\lambda x)}{g(x)} \underset{x\to\infty}{\longrightarrow} 1$$

car

$$g(x) \underset{x\to\infty}{\longrightarrow} |\alpha+1| > 0$$

et

$$\exp\left(-\int_x^{\lambda x} \frac{g(t)+\alpha+1}{t}dt\right) = \exp\left(-\int_1^{\lambda} \frac{g(xt)+\alpha+1}{t}dt\right) \underset{x\to\infty}{\longrightarrow} 1.$$

Ce qui permet de conclure que l varie lentement.

$\square$

On considère à présent U, une mesure concentrée sur $[0,\infty]$ et telle que sa transformée de Laplace

$$\omega(\lambda) = \int_0^\infty e^{-\lambda x} U(dx)$$

existe pour $\lambda > 0$.

On pose $\overrightarrow{U}(x) = U([0,x])$. On a alors des relations entre le comportement de ω en 0 et celui de $\overrightarrow{U}$ en $+\infty$; de telles relations sont données par les théorèmes Taubériens. En voici deux qui vont nous être particulièrement utiles mais que l'on va admettre. On pourra se reporter à [3].

Théorème 16 (Théorème Taubérien)**.** *Si l varie lentement et $0 \leq \rho < \infty$, alors*

$$\omega(\tau) \underset{\tau \to 0}{\sim} \tau^{-\rho} l\left(\frac{1}{\tau}\right) \tag{22}$$

si et seulement si

$$\overrightarrow{U}(t) \underset{t\to\infty}{\sim} \frac{1}{\Gamma(\rho+1)} t^\rho l(t) \tag{23}$$

Théorème 17 (Théorème de la densité monotone)**.** *Si U possède une densité u qui est monotone sur un certain intervalle $]a,\infty[$, alors U varie régulièrement avec indice $\rho > 0$ si et seulement si u varie régulièrement avec indice $\rho - 1$ et dans ce cas*

$$u(x) \underset{x\to\infty}{\sim} \rho \frac{U(x)}{x}.$$

4.2 Applications au schéma d'occupation d'une infinité de boîtes

On rappelle les définitions suivantes :

$$\nu(\mathrm{d}x) := \sum_j \delta_{p_j}(\mathrm{d}x)$$

et

$$\vec{\nu}(x) := \nu\left([x,\infty[\right) = \nu\left([x,1]\right).$$

On remarque d'abord deux propriétés de $\vec{\nu}(x)$:

- $\vec{\nu}(1/m) < m$. En effet, il y a au plus m fréquences supérieures à $1/m$.
- $\vec{\nu}(x) \ll x^{-1}$ quand x tend vers 0. En effet, une intégration par partie donne pour $x > 0$ (la mesure ν est finie sur $[x,1]$) :

$$\int_x^1 u\nu(\mathrm{d}u) = x\vec{\nu}(x) + \int_x^1 \vec{\nu}(u)\mathrm{d}u. \tag{24}$$

  De plus, quand x tend vers 0 l'intégrale de gauche croît vers 1 et l'intégrale de droite converge (par monotonie), ainsi $x\vec{\nu}(x)$ converge aussi. Si $\lim x\vec{\nu}(x)$ appartenait à $]0,\infty]$, alors l'intégrale de droite divergerait, forçant l'intégrale de gauche à diverger aussi, ce qui est absurde. On a donc $x\vec{\nu}(x) \to 0$.

– La démonstration précédente donne également

$$\int_0^1 \vec{\nu}(u)\mathrm{d}u = 1. \tag{25}$$

Dans le cas du schéma d'occupation d'une infinité de boîtes, on dit que la suite des fréquences (p_j) varie régulièrement si elle vérifie :

$$\vec{\nu}(x) \underset{x\to 0}{\sim} l(1/x)x^{-\alpha} \quad , \tag{26}$$

où l est une fonction variant lentement à l'infini et $\alpha \in [0,1]$. Si $\alpha = 0$ on parle de variation lente et si $\alpha = 1$ on parle de variation rapide. On définit les mesures suivantes pour $r \in \mathbb{N}^*$:

$$\nu_r(\mathrm{d}x) := x^r\nu(\mathrm{d}x) = \sum_i p_i^r \delta_{p_i}(\mathrm{d}x).$$

La mesure ν_1 correspond donc à la loi de la fréquence de la première boîte découverte. De plus, comme $\sum_j p_j = 1$, on a $\nu_1[0,1] = \int_0^1 u\nu(\mathrm{d}u)$.

Proposition 18. *Soit $\alpha \in]0,1[$. La suite des fréquences (p_j) vérifie la relation (26) (c'est à dire est à variation régulière d'exposant α) si et seulement si la condition suivante est satisfaite :*

$$\nu_1[0,x] \sim \frac{\alpha}{1-\alpha}x^{1-\alpha}l(1/x) \quad (x \downarrow 0). \tag{27}$$

De plus, si ces relations sont vérifiées, elles impliquent alors pour $r \geq 1$:

$$\nu_r[0,x] \sim \frac{\alpha}{1-\alpha}x^{r-\alpha}l(1/x) \quad (x \downarrow 0). \tag{28}$$

Démonstration. D'après (24) et (25) on a :

$$\nu_1[0,x] = 1 - \int_x^1 u\nu(\mathrm{d}u) = -x\vec{\nu}(x) + \int_0^x \vec{\nu}(u)\mathrm{d}u. \tag{29}$$

Si (26) est vérifiée, le théorème de Karamata (théorème 15) donne :

$$\frac{x^{-2+1}\vec{\nu}(1/x)}{\int_0^{1/x}\vec{\nu}(u)\mathrm{d}u} \underset{x\to\infty}{\to} |-2+1+\alpha| = 1-\alpha$$

dont on déduit (27) par le changement de variable $t \mapsto 1/t$. Pour montrer la réciproque, on utilise l'égalité suivante, obtenue par une intégration par parties :

$$\vec{\nu}(x) = \int_x^\infty u^{-1}\nu_1(\mathrm{d}u) = \nu_1[0,1] - x^{-1}\nu_1[0,x] + \int_x^\infty t^{-2}\nu_1[0,t]\mathrm{d}t. \tag{30}$$

Si on suppose (27), le théorème de Karamata (théorème 15) donne :

$$\frac{x^{1+0}\nu_1[0,1/x]}{\int_0^x \nu_1[0,1/t]\mathrm{d}t} \underset{x\to\infty}{\to} 1+0+\alpha-1 = \alpha$$

et on a bien (26). De plus, une intégration par parties donne :

$$\nu_r[0,x] = \int_0^x u^r\nu(\mathrm{d}u) = -x^r\vec{\nu}(x) + r\int_0^x u^{r-1}\vec{\nu}(u)\mathrm{d}u. \tag{31}$$

Ainsi si on suppose (26), le théorème de Karamata (théorème 15) nous donne (28). □

Pour les cas des fonctions variant rapidement ou lentement, les théorèmes sont moins forts.

Proposition 19. *Soit l variant lentement. Si la suite des fréquences varie rapidement, c'est à dire si la relation suivante est satisfaite :*

$$\vec{\nu}(x) \sim x^{-1} l(1/x) \quad (x \downarrow 0), \tag{32}$$

alors on a :

$$\nu_1[0, x] \sim l_1(1/x) \quad (x \downarrow 0), \tag{33}$$

avec

$$l_1(y) := \int_y^\infty u^{-1} l(u) du. \tag{34}$$

La fonction l_1 est également une fonction variant lentement et on a $l_1 \gg l$. De plus si $r > 1$ la relation (32) implique :

$$\nu_r[0, x] \sim \frac{1}{r-1} x^{r-1} l(1/x) \quad (x \downarrow 0). \tag{35}$$

En général, la relation (33) où l_1 est une fonction à variation lente implique seulement :

$$\vec{\nu}(x) \ll x^{-1} l_1(1/x) \quad (x \downarrow 0), \tag{36}$$

mais n'implique pas nécessairement que $\vec{\nu}(x)$ est à variation régulière. Cependant si on sait en plus que $\vec{\nu}(x)$ est à variation régulière, alors $\vec{\nu}(x)$ vérifie (32) et l et l_1 sont reliées par (34).

Démonstration. On suppose que la relation (32) est vérifiée. La fonction l_1 est bien définie. En effet

$$u^{-1} l(u) \underset{u \to \infty}{\sim} \vec{\nu}(1/u)$$

et (24) garantit que $\vec{\nu}$ est intégrable au voisinage de 0. De plus, comme l_1 est définie, on peut utiliser le dernier point du théorème de Karamata (théorème 15) et on obtient

$$\frac{x^{-1}\vec{\nu}(1/x)}{\int_x^\infty t^{-2}\vec{\nu}(1/t)\mathrm{d}t} \underset{x \to \infty}{\to} 0$$

et ainsi

$$l(x) \sim x^{-1}\vec{\nu}(1/x) \ll l_1(x) \quad (x \to \infty).$$

Comme $l \ll l_1$, la relation (29) nous donne :

$$\nu_1[0, x] \underset{x \to \infty}{\sim} \int_0^x \vec{\nu}\mathrm{d}u \underset{x \to \infty}{\sim} \int_0^x u^{-1} l(1/u)\mathrm{d}u = l_1(1/x).$$

Le théorème de Karamata (théorème 15) (valable car $r > 1$) et les relations (31) et (32) donnent (35).
Réciproquement, si on suppose la relation (33) vérifiée, alors d'après le théorème de Karamata (théorème 15)

$$\int_x^\infty u^{-2}\nu_1[0, u]\mathrm{d}u \sim_{x \to 0} x^{-1}\nu_1[0, x],$$

et (30) entraîne (36).
On suppose en plus que $\vec{\nu}$ est à variation régulière d'exposant α. Supposons par l'absurde que $\alpha \in]0,1[$. La proposition 18 donnerait alors

$$\vec{\nu}(x) \sim \frac{1-\alpha}{\alpha} x^{-1} l_1(1/x),$$

ce qui contredit (36). Si l'on avait $\alpha = 0$, alors le théorème de Karamata (théorème 15) nous donnerait $x^{-1}l_1(1/x) \ll \overrightarrow{\nu}(x)$ ce qui contredit également (36). En conclusion, si on suppose (33) et que $\vec{\nu}$ est à variation régulière, alors (32) et (34) sont vérifiées. □

Proposition 20. *Soit l_0 à variation lente. Alors la relation*

$$\nu_1[0,x] \sim x l_0(1/x) \quad (x \downarrow 0) \tag{37}$$

implique

$$\vec{\nu}(x) \sim l(1/x) \quad (x \downarrow 0), \tag{38}$$

avec

$$l(y) := \int_1^y u^{-1} l_0(u)\,du \quad (y > 1). \tag{39}$$

La fonction l est alors à variation lente et $l \gg l_0$.
De plus si $r > 1$, la relation (37) implique

$$\nu_r[0,x] \sim \frac{1}{r} x^r l_0(1/x) \quad (x \downarrow 0). \tag{40}$$

En général, la relation (38), où l est à variation régulière, entraîne seulement

$$\nu_1[0,x] \ll x l(1/x) \quad (x \downarrow 0) \tag{41}$$

et n'implique pas nécessairement (37). Cependant si on sait en plus que $\nu_1[0,x]$ est à variation régulière, alors (37) est satisfaite et l_0 et l sont reliées par (39).

Démonstration. On remarque que pour tout $u \geq 1$, $\nu_1[0,u] = 1$, ainsi

$$\int_x^\infty u^{-2} \nu_1[0,u] \mathrm{d}u = \int_0^x \nu_1[0,1/u] \mathrm{d}u$$

est bien définie. On peut donc appliquer le dernier point du théorème de Karamata (théorème 15) et on obtient :

$$x\nu_1[0,1/x] \underset{x\to\infty}{\ll} \int_0^x \nu_1[0,1/u]\mathrm{d}u.$$

On a donc

$$\vec{\nu}(x) \underset{x\to 0}{\sim} \int_0^1 u^{-2}\nu_1[0,u]\mathrm{d}u \sim_{x\to 0} l(1/x).$$

De plus, une intégration par parties donne

$$\nu_r[0,x] = \int_0^x u^{r-1}\nu_1(\mathrm{d}u) = \nu_1[0,x]x^{r-1} - (r-1)\int_0^x \nu_1[0,u]u^{r-2}\mathrm{d}u,$$

et en utilisant une nouvelle fois le théorème de Karamata (théorème 15) on obtient (40).
Réciproquement, si on suppose seulement (38), le théorème de Karamata (théorème 15) donne

$$\int_0^x \vec{\nu}(u)\mathrm{d}u \underset{x\to 0}{\sim} x\vec{\nu}(x)$$

et on conclut grâce à (30).
On suppose de plus que $\nu_1[0,x]$ est à variation régulière d'exposant α. Si on avait $\alpha > -1$, le théorème de Karamata donnerait

$$\vec{\nu}(x) \sim \frac{-\alpha}{1+\alpha} x^{-1}\nu_1[0,x],$$

ce qui contredirait (41). Si on avait $\alpha < -1$, la proposition 18 donnerait que $\vec{\nu}$ est à variation régulière d'exposant $1+\alpha \neq 0$, ce qui est absurde. Ainsi $\alpha = -1$ et on a bien (37) et (39). □

Les théorèmes Taubériens énoncés dans le paragraphe précédent, associés aux propositions 18, 19 et 20 nous permettent de traduire les relations (26), (32) et (38) en relations faisant intervenir les moyennes et donc plus facilement exploitables. C'est l'objet des propositions suivantes.

Proposition 21. *Soit $\alpha \in]0,1[$. La suite des fréquences (p_j) vérifie la relation (26) (c'est à dire est à variation régulière d'exposant α) si et seulement si l'une des deux conditions équivalentes suivantes est satisfaite :*

$$\Phi(t) \sim \Gamma(1-\alpha)t^\alpha l(t) \quad (t \to \infty) \tag{42}$$

ou

$$\Phi_1(t) \sim \alpha\Gamma(1-\alpha)t^\alpha l(t) \quad (t\to\infty). \tag{43}$$

Si c'est la cas, alors la relation suivante est satisfaite :

$$\Phi_r(t) \sim \frac{\alpha\Gamma(1-\alpha)}{r!} t^\alpha l(t) \quad (t\to\infty). \tag{44}$$

Démonstration. En faisant une intégration par parties dans (9) on obtient

$$\Phi(t) = t\int_0^\infty e^{-tx}\vec{\nu}(x)\mathrm{d}x, \tag{45}$$

et le théorème Taubérien 17 (avec $\rho = 1-\alpha$) donne l'équivalence entre (26) et (42). De plus, d'après (10)

$$\Phi_1(t) = t\int_0^\infty e^{-tx}\nu_1(\mathrm{d}x)$$

et en utilisant que $\Gamma(1+x) = x\Gamma(x)$. Le théorème Taubérien 16 avec $\rho = 1-\alpha$ donne l'équivalence entre (27) et (43). Ce théorème associé à la proposition 18 permet aussi de déduire (44) de (43). Ce même théorème avec $\rho = r-\alpha$ montre l'équivalence entre (28) et (44). □

Dans le cas où $(p_i)_{i\in\mathbb{N}}$ est à variation rapide, la proposition 21 devient :

Proposition 22. *La relation (32) implique*

$$\Phi(t) \sim \Phi_1(t) \sim tl_1(t) \quad (t \to \infty), \tag{46}$$

où l_1 est définie par (34). De plus, pour $r > 1$ cela implique également

$$\Phi_r(t) \sim \frac{1}{r(r-1)} tl(t) \quad (t \to \infty). \tag{47}$$

De plus, la relation (33) est équivalente à $\Phi_1(t) \sim tl_1(t) \quad (t \to \infty)$.

Démonstration. Comme précédemment, on démontre les relations faisant intervenir les $\Phi_r(t), r \geq 1$ grâce au théorème Taubérien 16 avec $\rho = r - 1$ ($\Gamma(r-1) = (r-2)!$). Cependant on ne peut plus utiliser le théorème Taubérien 17, car il n'est pas valable pour $\rho = 0$. On doit revenir à la démonstration de la proposition 19. On avait montré que si la relation (32) était vérifiée alors d'après (29),

$$\nu_1[0, x] \underset{x \to 0}{\sim} \int_0^x \vec{\nu}(u) \mathrm{d}u.$$

On peut alors utiliser le théorème Taubérien 16, valable pour $\rho = 0$ et on obtient l'équivalence entre (32) et

$$\Phi(t) \underset{t \to \infty}{\sim} tl_1(t).$$

□

Dans le cas des variations lentes, la proposition 21 devient :

Proposition 23. *La relation (37) implique*

$$\Phi(t) \sim l(t) \quad (t \to \infty), \tag{48}$$

où l est définie par (39). De plus cela implique également pour $r \geq 1$

$$\Phi_r(t) \sim \frac{1}{r} l_0(t) \quad (t \to \infty). \tag{49}$$

De plus, la relation (37) est équivalente à $\Phi_1(t) \sim l_0(t) \quad (t \to \infty)$, et la relation (38) est équivalente à (48) et entraîne $\Phi_r(t) \ll \Phi(t) \quad (t \to \infty)$ pour tout $r \geq 1$.

Démonstration. Comme précédemment, on démontre les relations faisant intervenir les $\Phi_r(t), r \geq 1$ grâce au théorème Taubérien 16 avec $\rho = r$ ($\Gamma(r) = (r-1)!$). Si on suppose (37), alors d'après la proposition 20 et le théorème Taubérien 17 avec $\rho = 1$ on a bien ($\Gamma(1) = 1$)

$$\Phi(t) \underset{t \to \infty}{\sim} l(t).$$

□

Les deux corollaires suivants permettent de résumer les résultats des propositions précédentes.

Corollaire 24. *On suppose que $\vec{\nu}$ est à variation régulière d'exposant α.*
- *Si $\alpha \in]0, 1[$, alors pour tout $r \geq 1$, $\Phi_r(t)$ est du même ordre de croissance que $\Phi(t)$, et le ratio $\Phi_r(t)/\Phi(t)$ converge vers $(-1)^r \binom{\alpha}{r}$.*

- *Si $\alpha = 1$, alors $\vec{\nu}$ varie rapidement et les $\Phi_r(t)$ sont du même ordre de croissance pour $r > 1$, mais $\Phi(t) \sim \Phi_1(t) \gg \Phi_r(t)$ pour $r > 1$. Ainsi, la plupart des boîtes occupées le sont par une seule boule.*
- *Si $\alpha = 0$ et si l'on suppose en plus que la condition (37) vérifiée, alors $r^{-1}\Phi_r(t) \sim \Phi_1(t) \ll \Phi(t)$.*

Corollaire 25. *Soit $\alpha \in]0,1[$, alors $\vec{\nu}$ vérifie la condition (26) si et seulement si l'une des conditions suivantes est vérifiée*

$$K_n \sim_{p.s.} \Gamma(1-\alpha)n^\alpha l(n) \quad (n \to \infty), \text{ ou } K(t) \sim_{p.s.} \Gamma(1-\alpha)t^\alpha l(t) \quad (t \to \infty),$$

et cela implique pour $r \in \mathbb{N}^*$

$$K_{n,r} \sim \frac{\alpha\Gamma(r-\alpha)}{r!} n^\alpha l(n) \quad (n \to \infty), \text{ et } K_r(t) \sim \frac{\alpha\Gamma(r-\alpha)}{r!} t^\alpha l(t) \quad (t \to \infty).$$

Démonstration. Le corollaire découle de la loi des grands nombres (théorème 7) et de la proposition 21. La loi des grands nombres pour $K_r(t)$, qui n'est pas vraie dans le cas général, est vérifiée ici et découle de la loi des grands nombres pour $\sum_{s>r} K_s(t)$ et du fait que $\sum_{s>r} \Phi_s(t) \asymp \Phi(t) \asymp t^\alpha l(t)$. □

La section suivante présente le conjugué de Bruijn d'une fonction variant lentement qui va permettre de formuler les propriétés des variations régulières directement en termes des fréquences (p_j).

4.3 Inversion

On va admettre quelques résultats issus de la théorie des fonctions à variation régulière. On pourra se reporter à [1].

Théorème 26 (Inverse asymptotique)**.** *Soit f fonction à variation régulière d'indice $\rho > 0$. Il existe g à variation régulière d'indice $1/\rho$, telle que*

$$f(g(x)) \sim g(f(x)) \sim x \quad (x \to \infty).$$

On dit alors que g est l'inverse asymptotique de f. L'inverse asymptotique est unique à équivalence asymptotique près et un des représentant est l'inverse généralisée $\overleftarrow{f}(x) := inf\{y : f(y) > x\}$.

Théorème 27 (conjugué de De Bruijn)**.** *Soit l variant lentement, il existe une fonction variant lentement $l^\sharp$, unique à équivalence asymptotique près, avec*

$$l(x)l^\sharp(xl(x)) \to 1, \quad l^\sharp(x)l(xl^\sharp(x)) \to 1 \quad (x \to \infty).$$

On a de plus $l^{\sharp\sharp} = l$.

Lemme 28. *Soit h et g inverses asymptotiques l'une de l'autre, alors pour $\alpha > 0$ et l variant lentement :*

$$h(x) \underset{x\to\infty}{\sim} x^\alpha l(x) \Leftrightarrow g(x) \underset{x\to\infty}{\sim} x^{1/\alpha}(l^{1/\alpha}(x^{1/\alpha}))^\sharp$$

Démonstration. Par l'unicité de l'inverse asymptotique il suffit de montrer que G, où $G(x) = x^\alpha l(x)$, et H, avec $H(x) = x^{1/\alpha}(l^{1/\alpha}(x^{1/\alpha}))^\sharp$, sont inverses l'une de l'autre. On remarque que si u et v sont inverses l'une de l'autre alors u^a et $v(.^{1/a})$ sont aussi inverses l'une de l'autre. Il suffit alors de montrer que $\overline{G} = G^{1/\alpha}$ et $\overline{H} = H(.^\alpha)$ sont inverses l'une de l'autre. L'indice de $\overline{G}$ et de $\overline{H}$ est 1, et on a bien

$$\overline{H}(\overline{G}(x)) = xl^{1/\alpha}(x)(l^{1/\alpha}(xl^{1/\alpha}(x)))^\sharp \underset{x\to\infty}{\sim} x$$

car $l^{1/\alpha}$ et $(l^{1/\alpha})^\sharp$ sont conjugués. De même,

$$\overline{G}(\overline{H}(x)) = x(l^{1/\alpha}(x))^\sharp l^{1/\alpha}(x(l^{1/\alpha}(x))^\sharp) \underset{x\to\infty}{\sim} x.$$

□

On pose alors

$$l^*(x) = \frac{1}{(l^{1/\alpha}(x^{1/\alpha}))^\sharp},$$

qui dépend de α.

On classe les p_i dans l'ordre décroissant et on étudie le comportement asymptotique de la suite $(p_i)_{i\in\mathbb{N}}$.

Proposition 29. *La condition* (26) *avec* $0 < \alpha < 1$ *est équivalente à*

$$p_i \underset{i\to\infty}{\sim} l^*(i)i^{-1/\alpha} \tag{50}$$

et implique

$$\sum_{i>k} p_i \underset{k\to\infty}{\sim} \frac{\alpha}{1-\alpha} l^*(k)k^{1-1/\alpha}$$

Démonstration. On considère

$$h(x) = \vec{\nu}(1/x) \quad \text{et} \quad g(x) = \frac{1}{p_{\lfloor x \rfloor}}$$

où $\lfloor x \rfloor$ est la partie entière de x.

On remarque que

$$\vec{\nu}(1/x) = \max\left\{i : p_i \geq \frac{1}{x}\right\} = \inf\left\{y : \frac{1}{p_{\lfloor y \rfloor}} > x\right\},$$

donc h est l'inverse généralisé de g et c'est donc son inverse asymptotique.

D'après le Lemme 28 la relation $h(x) \underset{x\to\infty}{\sim} x^\alpha l(x)$ est équivalente à

$$g(x) \underset{x\to\infty}{\sim} x^{1/\alpha}(l^{1/\alpha}(x^{1/\alpha}))^\sharp,$$

ce qui donne facilement le résultat souhaité. □

On définit les $S_k := min\{n : K_n = k\}$, $k \in \mathbb{N}^*$, les "temps" auxquels on découvre de nouvelles boîtes (pour le problème originel pour lequel n n'est pas aléatoire). On définit de même pour le problème poissonisé $T_k := min\{t : K(t) = k\}$. On a donc par définition : $K_{S_k} = K(T_k) = k$, $k \in \mathbb{N}^*$. La proposition suivante nous fournit la relation inverse sous l'hypothèse de variation régulière d'exposant $\alpha \in]0, 1[$.

Proposition 30. *La condition* (26) *implique pour* $0 < \alpha < 1$ *:*

$$S_k \sim_{p.s.} T_k \sim_{p.s.} \left(\frac{k}{\Gamma(1-\alpha)}\right)^{1/\alpha} \frac{1}{l^*(k)} \qquad (k \to \infty)$$

On a aussi,

$$s_k \sim \left(\frac{k}{\Gamma(1-\alpha)}\right)^{1/\alpha} \frac{1}{l^*(k)} \iff K_{s_k} \sim_{p.s.} K(s_k) \sim_{p.s.} k \quad (k \to \infty)$$
$$k_n \sim \Gamma(1-\alpha) n^\alpha l(n) \iff S_{k_n} \sim_{p.s.} T_{k_n} \sim_{p.s.} n \qquad (n \to \infty)$$

Démonstration. On pose $f(x) = \inf\{z : K_{\lfloor z \rfloor} > x\}$ l'inverse généralisé de $K_{\lfloor x \rfloor}$, on a $f(k) \underset{k\to\infty}{\sim} S_k$. On déduit le résultat du Lemme 28, de l'unicité de l'inverse asymptotique et du corollaire 25. □

5 Appendice

On donne le calcul de la variance dans le cas fixe.

$$\begin{aligned}
V_n &= \mathrm{Var}\,(K_n) \\
&= \mathbb{E}\left(K_n^2\right) - \mathbb{E}\,(K_n)^2 \\
&= \mathbb{E}\left(\sum_i \sum_j 1_{X_{n,j}>0} 1_{X_{n,i}>0}\right) - \Phi_n^2 \\
&= \sum_{i\neq j} \mathbb{P}\,(X_{n,j} > 0 \text{ et } X_{n,i} > 0) + \Phi_n - \Phi_n^2 \\
&= \sum_{i\neq j} \sum_{r=1}^{n-1} \sum_{l=1}^{n-r} \mathbb{P}\,(X_{n,j} = r \text{ et } X_{n,i} = l) + \Phi_n - \Phi_n^2 \\
&= \sum_{i\neq j} \sum_{r=1}^{n-1} \sum_{l=1}^{n-r} \left(\frac{n!}{l!r!(n-r-l)!} p_j^r p_i^l (1-p_i-p_j)^{n-r-l}\right) + \Phi_n - \Phi_n^2 \\
&= \sum_{i\neq j} \left(\sum_{r=1}^{n-1} \frac{n!}{r!(n-r)!} p_j^r (1-p_j)^{n-r} - \sum_{r=1}^{n-1} \frac{n!}{(n-r)!r!} p_j^r (1-p_i-p_j)^{n-r}\right) \\
&\quad +\Phi_n - \Phi_n^2 \\
&= \sum_{i\neq j} \left(1-(1-p_j)^n - p_j^n - ((1-p_i)^n - (1-p_i-p_j)^n - p_j^n)\right) + \Phi_n \\
&\quad -\sum_i \sum_j (1-(1-p_i)^n)(1-(1-p_j)^n) \\
&= \sum_{i\neq j} ((1-p_j-p_i)^n - (1-p_j)^n(1-p_i)^n) \\
&\quad +\sum_j \left(1-(1-p_j)^n - \left(1+(1-p_j)^{2n} - 2(1-p_j)^n\right)\right)
\end{aligned}$$

D'où

$$V_n = \Phi_{2n} - \Phi_n + \sum_{i\neq j} [(1-p_j-p_i)^n - (1-p_j)^n(1-p_j)^n]$$

Nous remercions Jean Bertoin et Marie Théret de leur aide qui nous a été très précieuse.

Références

[1] N.H. Bingham, C.M. Goldie et J.L. Teugels, *Regular variation*, Cambridge University Press, 2006.

[2] M.Dutko, Central limit theorems for infinite urn models, *Ann. Probab.* 17 :1255-1263, 1989.

[3] W. Feller, *An introduction to probability theory and its applications*, vol. II, Wiley, NY, 1971.

[4] A. Gnedin, B. Hansen et J. Pitman, Notes on the occupancy problem with infinitely many boxes : general asymptotics and power laws, *Probability Survey* 4 :146-171, 2007.

[5] J.F.C. Kingman, *Poisson processes*, Oxford studies in probability, 1993.