

UNIVERSITÉ DE LIÈGE
Faculté des Sciences
Institut de Mathématique



PROCESSUS DE POISSON

Année académique 2006-2007

Mémoire présenté par
Christel Ruwet en vue de
l'obtention
du grade de licenciée en
Sciences Mathématiques

Introduction

Il existe plusieurs types de processus de Poisson. Dans ce mémoire, ce sont les processus "temporels" qui seront abordés, mais il existe également des processus de Poisson dans le plan (comme la projection sur un plan de la répartition des étoiles dans le ciel) ou dans d'autres espaces. En langage non mathématique, un processus de Poisson dans le temps est le processus qui est souvent le mieux adapté pour expliquer un processus "d'arrivées", ce dernier mot étant pris au sens large. En effet, une arrivée peut être une panne se produisant sur une machine, un coup de téléphone arrivant à un standard, un client accédant à un guichet, Comme nous le verrons dans la suite, les processus de Poisson temporels se subdivisent en plusieurs types.

La première partie de ce travail reprend les aspects théoriques des différents processus de Poisson temporels. Après un premier chapitre ayant pour but d'asseoir les principes probabilistes et statistiques utiles tout au long de ce travail, le deuxième chapitre consiste à définir de façon rigoureuse le plus connu et le plus simple d'entre eux, appelé ici processus de Poisson de base, ainsi que d'en chercher les caractéristiques principales. Cette partie du travail était sans doute la plus simple car beaucoup d'ouvrages traitent ce sujet en profondeur. Cependant, pour adhérer encore plus à la réalité, il ne faut pas s'arrêter là et il faut étudier les généralisations possibles de ce processus. On parle alors de processus de Poisson non-homogène ou de processus de Poisson composé. Le troisième chapitre est donc consacré au processus de Poisson non-homogène permettant de modéliser des situations où l'intensité des occurrences n'est pas constante au cours du temps. Par un changement d'échelle, ce processus se ramène à un processus de Poisson de base. Le quatrième chapitre permet lui de lever l'hypothèse selon laquelle deux arrivées ne peuvent se produire en même temps en considérant le processus de Poisson dit composé. Il cache en fait un processus de Poisson de base. Leur lien avec le processus de base explique pourquoi ces deux processus ne sont souvent que cités dans la littérature. Comme le but de ce mémoire était de comparer ces trois processus entre eux, il ne suffisait pas de donner les relations les liant. La première difficulté a donc consisté à trouver des moyens pour adapter les propriétés d'un processus de Poisson de base à ces généralisations. Dans certains cas, cela n'a malheureusement pas été possible. Le cinquième chapitre aborde quant à lui le cas de la superposition de plusieurs processus de Poisson et de l'amincissement d'un tel processus.

Commence alors la seconde et dernière partie. Elle est consacrée à l'application sur des exemples concrets de la théorie développée dans la première partie. Cette seconde partie a été le point de départ de nouvelles embûches. La première est venue du fait que les processus stochastiques en général, et donc les processus de Poisson, sont étudiés principalement par des probabilistes qui se soucient parfois peu de la mise en pratique. Cela étant, dans les références les plus populaires dans ce domaine, les méthodes de simulations et de tests sont

peu développées. Pour les besoins de cette partie, il a donc été nécessaire de trouver des techniques statistiques utilisables dans le contexte et de les comparer afin de voir leur efficacité respective. La seconde difficulté n'était pas inhérente au sujet des processus de Poisson mais il était indispensable de l'aborder quand même. Il s'agit de la problématique liée à l'estimation des paramètres lors d'un test d'ajustement de Kolmogorov-Smirnov. Par souci de clarté, ce point est abordé dans le premier chapitre avec les autres notions liées indirectement aux processus de Poisson. Enfin, le dernier problème fut la recherche des données à analyser. Il est habituellement considéré acquis que les arrivées à la poste sont poissonniennes. C'est un exemple parmi tant d'autres qu'il serait intéressant d'analyser afin de contredire ou de confirmer ces a priori bien connus. Les données utilisées sont des données réelles (non simulées) qu'il a fallu récolter "à la main".

Tout au long de ce travail, l'utilisation d'un logiciel statistique a été nécessaire. C'est le logiciel R qui a été choisi pour ses facilités de programmation.

Je tiens tout d'abord à remercier toutes celles et ceux qui m'ont aidé dans la réalisation de ce mémoire, que ce soit en me permettant d'utiliser leurs données pour les analyses, en me procurant un endroit calme où travailler ou encore en me donnant de nombreux conseils.

Des conseils, j'en ai reçu beaucoup de ma promotrice, Madame Haesbroeck, que je remercie vivement. Ses nombreuses suggestions et questions m'ont permis de mener à bien ce travail. Je la remercie également pour la découverte de cette théorie ainsi que pour sa disponibilité tout au long de l'année.

Enfin, je remercie toute ma famille pour leur soutien au long de mes études, mes camarades pour les bons moments passés ensemble et Olivier pour sa patience et ses encouragements.

Table des matières

1	Notions de base	1
1.1	Fonctions génératrices et caractéristique	1
1.2	Quelques lois de distributions	6
1.3	Tests d'ajustement	14
1.4	Introduction aux processus stochastiques	16
2	Processus de Poisson de base	19
2.1	Distribution des marges finies du processus	20
2.2	Définition alternative	24
2.3	Temps d'inter-arrivées et temps d'occurrence	26
2.4	Distributions tronquées	34
3	Processus de Poisson non homogène	35
3.1	Distribution des marges finies du processus	36
3.2	Temps d'inter-arrivées et temps d'occurrence	43
3.3	Lien avec le processus de Poisson de base	48
4	Processus de Poisson composé	51
4.1	Distribution des marges finies du processus	51
4.2	Temps d'inter-arrivées et temps d'occurrence	59
5	Autres processus de Poisson	61
5.1	Amincissement d'un processus de Poisson	61
5.2	Superposition de processus de Poisson	67
6	Processus de Poisson en pratique	73
6.1	Processus de base	73
6.1.1	Simulation	73
6.1.2	Tests	74
6.1.3	Intervalle de confiance pour le paramètre d'intensité ν	75
6.1.4	Comparaison des intensités de deux processus de Poisson	76
6.2	Processus de Poisson composé	76
6.2.1	Simulation	76
6.2.2	Tests	77
6.3	Processus de Poisson non homogène	77
6.3.1	Simulation	77
6.3.2	Tests	78

6.3.3 Estimation de la fonction moyenne	79
7 Applications	84
7.1 Données provenant de la littérature	84
7.2 Données observées	87
A Logiciel R	96
Bibliographie	105

Chapitre 1

Notions de base

Ce premier chapitre a pour but de présenter brièvement les différents concepts de base relatifs aux probabilités ainsi qu'aux processus stochastiques qui seront nécessaires dans la suite. Malgré que cela ne soit pas en lien direct avec le sujet de ce travail, une section est également consacrée aux tests d'ajustement. Pour les applications abordées dans le chapitre 7, il a été très utile d'étudier les tests qui sont le plus souvent employés mais aussi leurs inconvénients afin d'essayer de les palier. Pour plus de précision ou d'information, il est conseillé de consulter les ouvrages référencés. En outre, il sera supposé que les fonctions de densité abordées sont toujours bien définies et continues.

1.1 Fonctions génératrices et caractéristique

Ces fonctions sont très importantes car elles ont la particularité de caractériser complètement la variable aléatoire à laquelle elles se rapportent. En d'autres termes, connaître la fonction génératrice ou la fonction caractéristique d'une variable aléatoire est équivalent à connaître sa distribution. Commençons alors par la plus simple d'entre elles.

Définition 1.1. *Soit X une variable aléatoire discrète ne prenant que des valeurs entières positives ou nulles. La fonction génératrice des probabilités de la variable X est la fonction*

$$G_X(t) = E[t^X] = \sum_{n=0}^{+\infty} t^n \mathbb{P}[X = n] \quad (1.1)$$

définie pour tout $t \in \mathbb{R}$ tel que l'espérance existe (c-à-d tel que la série soit absolument convergente).

Pour $t \in [-1, 1]$, cette série converge absolument ; la fonction génératrice des probabilités est donc au moins définie dans l'intervalle $[-1, 1]$.

Comme son nom l'indique, cette fonction permet de générer les probabilités de la variable X .

Proposition 1.1. *Soit X une variable aléatoire discrète ne prenant que des valeurs entières positives ou nulles. Pour tout $k \in \mathbb{N}$, on a*

$$\mathbb{P}[X = k] = \frac{D_t^k G_X(t)|_{t=0}}{k!}. \quad (1.2)$$

La fonction génératrice des probabilités d'une variable aléatoire discrète à valeurs entières positives ou nulles X caractérise donc complètement cette dernière.

Démonstration. Cela a un sens de calculer les dérivées en zéro car cette valeur appartient à l'ensemble des points où la fonction est définie. Comme la série converge absolument sur $[-1, 1]$ et que le terme général de cette série est dérivable sur cet intervalle, on peut¹ faire porter la dérivée sur chacun des termes de la série. En évaluant en zéro les dérivées successives de la fonction génératrice des probabilités, on obtient alors la loi de probabilité de la variable :

$$\begin{aligned} G_X(0) &= \mathbb{P}[X = 0] \\ D_t G_X(t)|_{t=0} &= \sum_{n=1}^{+\infty} n t^{n-1} \mathbb{P}[X = n]|_{t=0} = \mathbb{P}[X = 1] \\ &\vdots \\ D_t^k G_X(t)|_{t=0} &= \sum_{n=k}^{+\infty} n(n-1) \dots (n-k+1) t^{n-k} \mathbb{P}[X = n]|_{t=0} = k! \mathbb{P}[X = k] \end{aligned}$$

pour tout $k \in \mathbb{N}$.

Or une variable aléatoire discrète est complètement définie par sa loi de probabilité. ■

Cette proposition montre l'importance de considérer des variables à valeurs entières positives ou nulles car les termes pour $n < 0$ ne sont pas dérivables en zéro.

Si les moments non-centrés d'ordre inférieur ou égal à p de la variable X existent, il est possible de les retrouver en évaluant en 1 les dérivées successives de la fonction génératrice des probabilités de X . En effet, en passant la dérivée à l'intérieur de la série, il vient

$$\begin{aligned} D_t G_X(t)|_{t=1} &= \sum_{n=0}^{+\infty} n \mathbb{P}[X = n] = E[X] \tag{1.3} \\ D_t^2 G_X(t)|_{t=1} &= \sum_{n=1}^{+\infty} n(n-1) \mathbb{P}[X = n] = \sum_{n=0}^{+\infty} n^2 \mathbb{P}[X = n] - \sum_{n=0}^{+\infty} n \mathbb{P}[X = n] = E[X^2] - E[X] \end{aligned}$$

ce qui donne

$$E[X^2] = D_t^2 G_X(t)|_{t=1} + D_t G_X(t)|_{t=1} \tag{1.4}$$

⋮

$$D_t^k G_X(t)|_{t=1} = \sum_{n=k}^{+\infty} n(n-1) \dots (n-k+1) \mathbb{P}[X = n] = \sum_{i=1}^k C_i E[X^i]$$

en isolant le k -ième moment non-centré, cela donne $E[X^k] = \sum_{i=1}^k C_i' D_t^i G_X(t)|_{t=1}$, pour tout $k \leq p$.

Proposition 1.2. *La fonction génératrice des probabilités d'une somme de variables indépendantes à valeurs entières non négatives, définie en tout point t en lequel la fonction génératrice des probabilités de chacune des variables est définie, est le produit des fonctions génératrices des probabilités de ces variables.*

¹cfr F.BASTIN, *ANALYSE NON LINEAIRE, Partie 1*, Université de Liège, Département de Mathématique, Année académique 2004-2005.

Démonstration. Soient X_1, X_2, \dots, X_n des variables indépendantes à valeurs entières non négatives et $Y = X_1 + X_2 + \dots + X_n$. Pour t réel tel que $E[t^{X_i}]$ existe, $\forall i = 1, 2, \dots, n$,

$$G_Y(t) = E[t^Y] = E[t^{X_1} t^{X_2} \dots t^{X_n}] = E[t^{X_1}] E[t^{X_2}] \dots E[t^{X_n}] = G_{X_1}(t) G_{X_2}(t) \dots G_{X_n}(t),$$

où la troisième égalité est due à l'indépendance des variables. ■

Cette fonction, bien que très utile et facile d'utilisation, a deux inconvénients, à savoir qu'elle n'est définie que pour des variables discrètes à valeurs positives ou nulles et que, même pour ces variables, elle n'est pas définie partout. Pour résoudre le premier, on introduit la fonction génératrice des moments.

Définition 1.2. *Pour toute variable aléatoire X , discrète ou continue, la fonction génératrice des moments est définie par*

$$M_X(t) = E[e^{tX}] = \begin{cases} \sum_x e^{tx} \mathbb{P}[X = x] & \text{si } X \text{ est discrète} \\ \int_{-\infty}^{+\infty} e^{tx} f(x) dx & \text{si } X \text{ est continue et de densité } f \end{cases} \quad (1.5)$$

pour tout $t \in \mathbb{R}$ tel que l'espérance soit définie.

Cette fonction possède des propriétés similaires à la fonction génératrice des probabilités.

Proposition 1.3. *Si les moments non-centrés d'ordre inférieur ou égal à p de la variable X existent, la fonction génératrice des moments de X permet de les retrouver grâce à la formule suivante valable $\forall k \leq p$:*

$$E[X^k] = D_t^k M_X(t) |_{t=0}. \quad (1.6)$$

Démonstration. On va être intéressé par cette fonction au point 0, ce qui est légitime puisque en $t = 0$, la série est absolument convergente et l'intégrale existe.

Cela étant, l'opérateur de dérivation passe à l'intérieur de la série ou de l'intégrale², ce qui donne

$$\begin{aligned} D_t M_X(t) |_{t=0} &= E [D_t(e^{tX})] |_{t=0} = E [X e^{tX}] |_{t=0} = E[X] \\ D_t^2 M_X(t) |_{t=0} &= E [D_t^2(e^{tX})] |_{t=0} = E [X^2 e^{tX}] |_{t=0} = E[X^2] \\ &\vdots \\ D_t^k M_X(t) |_{t=0} &= E [D_t^k(e^{tX})] |_{t=0} = E [X^k e^{tX}] |_{t=0} = E[X^k] \end{aligned}$$

pour tout $k \leq p$. ■

Proposition 1.4. *Soient deux variables aléatoires X et Y telles que $M_X(t) = M_Y(t)$ pour tout t en lequel ces expressions existent. Alors ces deux variables ont la même distribution. Cela étant, la fonction génératrice des moments caractérise complètement la distribution.*

Démonstration. Traitons les cas discret et continu séparément.

Cas discret : Supposons que X prenne les valeurs $\{x_1, \dots, x_n\}$ et que les valeurs possibles pour Y soient $\{y_1, \dots, y_m\}$. Notons $p_i^X = \mathbb{P}[X = x_i]$ et $p_j^Y = \mathbb{P}[Y = y_j]$ pour tout $i = 1, \dots, n$

²cfr J.SCHMETS, *Analyse Mathématique, Introduction au Calcul Intégral*, Liège, Éditions Derouaux, 1994.

et $j = 1, \dots, m$. Alors l'hypothèse d'égalité des fonctions génératrices des moments se traduit comme suit :

$$\sum_{i=1}^n p_i^X e^{tx_i} = \sum_{j=1}^m p_j^Y e^{ty_j}, \forall t.$$

En posant

$$b_k = \begin{cases} x_i & \text{si } k = i \\ y_j & \text{si } k = j \end{cases}$$

et

$$a_k = \begin{cases} p_i^X & \text{si } k = i \text{ et } x_i \neq y_j \forall j \\ p_j^Y & \text{si } k = j \text{ et } y_j \neq x_i \forall i \\ p_i^X - p_j^Y & \text{si } k \in \{i, j\} \text{ et } x_i = y_j \end{cases},$$

il est possible de réécrire cela sous la forme

$$\sum_k a_k e^{tb_k} = 0, \forall t \tag{1.7}$$

où les b_k sont tous distincts. Comme les fonctions e^{tb_k} sont alors indépendantes, l'équation (1.7) est équivalente à la nullité de tous les a_k . Cela étant, $n = m$ et $p_i^X = p_j^Y$ pour tout $i, j = 1, \dots, n$. Cela montre que les deux variables discrètes ont même distribution.

Cas continu : Notons f_X et f_Y les fonctions de densité de X et Y respectivement. Il vient

$$M_X(t) = \int_{\mathbb{R}} e^{tx} f_X(x) dx = \mathcal{L}_{-t} f_X,$$

où $\mathcal{L}_t f_X$ est la transformée de Laplace de f_X en t . L'hypothèse d'égalité des fonctions génératrices des moments se traduit alors par

$$\begin{aligned} \mathcal{L}_{-t} f_X &= \mathcal{L}_{-t} f_Y, \forall t \\ \Leftrightarrow \mathcal{L}_{-t} (f_X - f_Y) &= 0, \forall t \\ \Leftrightarrow f_X - f_Y &= 0 \text{ presque partout} \\ \Leftrightarrow f_X &= f_Y \text{ car les densités sont supposées continues.} \end{aligned}$$

Les variables ont donc la même distribution. ■

Proposition 1.5. *La fonction génératrice des moments d'une somme de variables aléatoires indépendantes, définie en tout point t en lequel la fonction génératrice des moments de chacune des variables est définie, est le produit des fonctions génératrices des moments des variables considérées.*

La preuve est similaire à celle de la proposition 1.2.

On constate que si cette fonction s'applique à tout type de variable aléatoire, elle n'est pas nécessairement définie pour tout $t \in \mathbb{R}$. Pour pallier à ce second inconvénient, il faut introduire la fonction caractéristique.

Définition 1.3. *Soit X une variable aléatoire quelconque. Sa fonction caractéristique est définie pour tout $t \in \mathbb{R}$ par*

$$\phi_X(t) = E[e^{itX}] = \begin{cases} \sum_x e^{itx} IP[X = x] & \text{si } X \text{ est discrète} \\ \int_{-\infty}^{+\infty} e^{itx} f(x) dx & \text{si } X \text{ est continue et de densité } f. \end{cases} \tag{1.8}$$

Dans le cas où la variable X est continue, la fonction caractéristique correspond à la transformée de Fourier de sa fonction de densité.

Comme précédemment, la fonction caractéristique caractérise complètement une variable. Plusieurs démonstrations de ce fait sont possibles ; celle introduite ici suit le même schéma que pour les fonctions génératrices et donne un moyen simple de retrouver les moments non-centrés lorsqu'ils existent.

Proposition 1.6. *Si les moments d'ordre inférieur ou égal à p existent, alors*

$$D_t^k \phi_X(t)|_{t=0} = i^k E[X^k], \quad \forall k \leq p. \quad (1.9)$$

Démonstration. Comme précédemment, la dérivée peut passer sous le signe d'intégration ou à l'intérieur de la série. Pour tout $k \leq p$, il vient

$$D_t^k \phi_X(t)|_{t=0} = E[D_t^k e^{itX}]|_{t=0} = E[(iX)^k e^{itX}]|_{t=0} = i^k E[X^k].$$

■

Proposition 1.7. *Soient deux variables aléatoires X et Y telles que $\phi_X(t) = \phi_Y(t)$ en tout point t . Alors ces deux variables ont la même distribution. Cela étant, la fonction caractéristique caractérise complètement la distribution.*

Démonstration. Traitons les cas discret et continu séparément.

Cas discret : En utilisant les mêmes notations que dans la preuve de la proposition 1.4, l'égalité des fonctions caractéristiques est équivalente à

$$\begin{aligned} \sum_{j=1}^n e^{itx_j} p_j^X &= \sum_{k=1}^m e^{ity_k} p_k^Y, \quad \forall t \\ \Leftrightarrow \sum_{j=1}^n e^{\frac{2i\pi t x_j}{2\pi}} p_j^X &= \sum_{k=1}^m e^{\frac{2i\pi t y_k}{2\pi}} p_k^Y, \quad \forall t \end{aligned}$$

et comme la fonction caractéristique est 2π -périodique, un résultat concernant les séries trigonométriques de Fourier donne

$$\Leftrightarrow n = m \text{ et } p_j^X = p_k^Y \quad \forall j, k = 1, \dots, n.$$

Pour le cas discret, la conclusion est donc immédiate.

Cas continu : Notons f_X et f_Y les fonctions de densité de X et Y . Il vient

$$\begin{aligned} \phi_X(t) &= \phi_Y(t), \quad \forall t \\ \Leftrightarrow \int_{\mathbb{R}} e^{itx} f_X(x) dx &= \int_{\mathbb{R}} e^{itx} f_Y(x) dx, \quad \forall t \\ \Leftrightarrow \int_{\mathbb{R}} e^{itx} (f_X(x) - f_Y(x)) dx &= 0, \quad \forall t \end{aligned}$$

et, en introduisant la notation $\mathcal{F}_t^+ f_X$ pour la transformée de Fourier, cela donne

$$\Leftrightarrow \mathcal{F}_t^+(f_X - f_Y) = 0 = \mathcal{F}_t^+ 0, \quad \forall t.$$

L'injectivité de la transformée de Fourier implique alors

$$\Leftrightarrow f_X = f_Y,$$

ce qui suffit pour conclure. ■

Proposition 1.8. *La fonction caractéristique d'une somme de variables aléatoires indépendantes est le produit des fonctions caractéristiques de ces variables.*

La preuve suit le même schéma que celle donnée dans le cas des fonctions génératrices des probabilités.

Voici quelques propriétés importantes de cette fonction, conséquences directes du fait que $\phi_X(t) = E[e^{itX}]$.

Proposition 1.9. *La fonction caractéristique de toute variable aléatoire X vérifie les propriétés suivantes :*

- (a) $\phi_X(0) = 1$;
- (b) $|\phi_X(t)| \leq 1$;
- (c) $\phi_{aX+b}(t) = e^{ibt} \phi_X(at)$, $\forall a, b \in \mathbb{R}$.

Lorsque l'on considère des variables aléatoires non négatives, on peut remplacer avec profit la fonction caractéristique par la transformée de Laplace de la fonction de densité dont voici la définition.

Définition 1.4. *Soit X une variable aléatoire non négative, sa transformée de Laplace est définie par*

$$\psi_X(s) = E[e^{-sX}] = \begin{cases} \sum_x e^{-sx} \mathbb{P}[X = x] & \text{si } X \text{ est discrète} \\ \int_{-\infty}^{+\infty} e^{-sx} f(x) dx & \text{si } X \text{ est continue et de densité } f \end{cases} \quad (1.10)$$

pour une variable imaginaire $s = r + it$ avec $r \in \mathbb{R}^+$ et $t \in \mathbb{R}$.

Quand s est purement imaginaire, $s = it$, on a

$$\psi_X(s) = \phi_X(-t). \quad (1.11)$$

De même, lorsque s est purement réel, $s = r$, on a

$$\psi_X(s) = M_X(-r). \quad (1.12)$$

1.2 Quelques lois de distributions

Pour comprendre la suite de ce travail, il sera utile de connaître les définitions et quelques propriétés des lois les plus rencontrées. En effet, celles-ci seront alors appliquées sans autre explication que la référence à cette section. Pour les autres, un simple rappel sera effectué en cas de besoin.

Définition 1.5. Une variable aléatoire discrète X suit une loi de Poisson de paramètre $\nu \in \mathbb{R}^+$, ce que l'on note $X \sim \mathcal{P}(\nu)$, si $X(\Omega) = \mathbb{N}$ et si, $\forall x \in \mathbb{N}$,

$$\mathbb{P}[X = x] = e^{-\nu} \frac{\nu^x}{x!}.$$

Les graphiques représentés aux figures 1.1, 1.2, 1.3, 1.4 permettent de voir l'impact du paramètre ν sur la loi de la variable. Ils ont été obtenus dans le logiciel R par la commande `ppois()`.

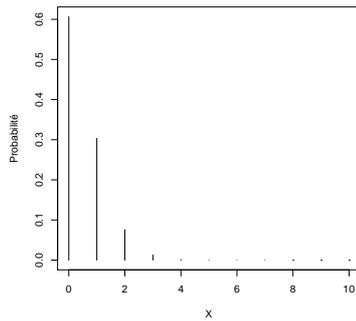


FIG. 1.1 – Poisson : $\nu = 0,5$

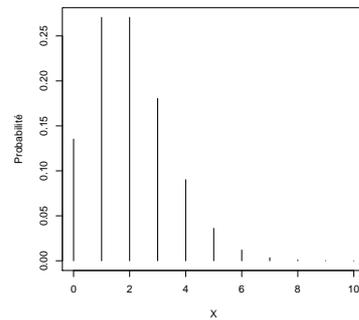


FIG. 1.2 – Poisson : $\nu = 2$

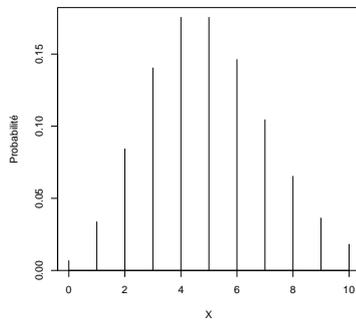


FIG. 1.3 – Poisson : $\nu = 5$

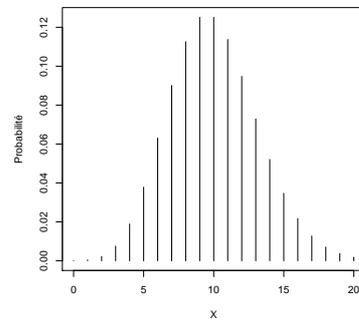


FIG. 1.4 – Poisson : $\nu = 10$

Proposition 1.10. La fonction génératrice des probabilités d'une variable aléatoire discrète distribuée selon une loi de Poisson de paramètre ν est donnée par

$$G_X(t) = e^{\nu(t-1)}, \quad \forall t \in \mathbb{R}.$$

Démonstration. En utilisant successivement les définitions 1.1 et 1.5 puis le développement en série de l'exponentielle, on obtient

$$G_X(t) = E[t^X] = \sum_{n=0}^{+\infty} t^n \mathbb{P}[X = n] = \sum_{n=0}^{+\infty} \frac{t^n e^{-\nu} \nu^n}{n!} = e^{-\nu} \sum_{n=0}^{+\infty} \frac{(t\nu)^n}{n!} = e^{-\nu} e^{t\nu} = e^{\nu(t-1)},$$

pour tout $t \in \mathbb{R}$ car le développement en série de la fonction exponentielle est valable sur \mathbb{R} . ■

Au départ de cette expression et des formules (1.3) et (1.4), il est possible de trouver l'espérance et la variance de X , toutes deux égales à ν .

Revenons sur l'interprétation des graphiques précédents. Lorsque ν est petit, l'espérance et la variance de la variable aléatoire le sont aussi. Donc la valeur la plus probable pour la variable est proche de 0 et la dispersion est faible. La masse de probabilité est par conséquent concentrée sur les petites valeurs, comme le montrent les graphiques 1.1 et 1.2. Ensuite, lorsque ν augmente, la masse de probabilité se déplace vers la droite tout en s'étalant de plus en plus. Cela est illustré par les graphiques 1.3 et 1.4.

Proposition 1.11. *Les fonctions génératrice des moments et caractéristique d'une variable aléatoire discrète distribuée selon une loi de Poisson de paramètre ν sont données respectivement par*

$$M_X(t) = e^{\nu(e^t-1)}$$

et

$$\phi_X(t) = e^{\nu(e^{it}-1)}$$

pour tout $t \in \mathbb{R}$.

Démonstration. Établissons le résultat de la fonction caractéristique uniquement, le développement étant similaire pour la fonction génératrice des moments. Comme dans la preuve de la proposition 1.10, il vient

$$\phi_X(t) = \sum_{n=0}^{+\infty} e^{itn} \mathbb{P}[X = n] = \sum_{n=0}^{+\infty} e^{itn} e^{-\nu} \frac{\nu^n}{n!} = e^{-\nu} \sum_{n=0}^{+\infty} \frac{(\nu e^{it})^n}{n!} = e^{-\nu} e^{\nu e^{it}} = e^{\nu(e^{it}-1)}.$$

■

Définition 1.6. *Une variable aléatoire continue X suit une loi exponentielle négative de paramètre $\nu > 0$, ce que l'on note $X \sim \text{Exp}(\nu)$, si la fonction de densité correspondante est donnée par $f(x) = \begin{cases} \nu e^{-\nu x} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$*

La fonction de répartition est alors donnée par $F(x) = \begin{cases} 1 - e^{-\nu x} & \text{si } x \geq 0 \\ 0 & \text{sinon.} \end{cases}$

Regardons à nouveau l'effet du paramètre ν sur la distribution de la variable avec les quelques graphiques de la page 9. La commande à taper dans R est `dexp()`.

Proposition 1.12. *La fonction génératrice des moments d'une variable aléatoire X distribuée suivant une loi exponentielle de paramètre ν est donnée par*

$$M_X(t) = \frac{\nu}{\nu - t}, \quad \forall t < \nu.$$

Démonstration. En utilisant successivement les définitions 1.2 et 1.6, nous obtenons

$$M_X(t) = E[e^{tX}] = \int_{-\infty}^{+\infty} e^{tx} f(x) dx = \int_0^{+\infty} e^{tx} \nu e^{-\nu x} dx = \nu \int_0^{+\infty} e^{-(\nu-t)x} dx.$$

Pour $t \geq \nu$, la fonction n'est pas intégrable en l'infini, tandis que pour $t < \nu$, on a $M_X(t) = \frac{\nu}{\nu-t}$.

■

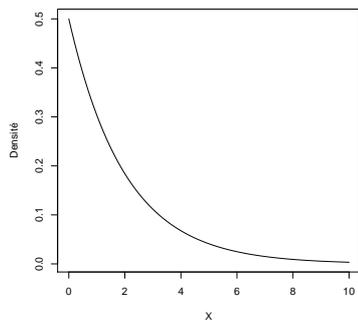


FIG. 1.5 – Exponentielle : $\nu = 0,5$

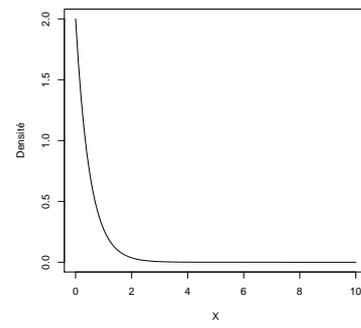


FIG. 1.6 – Exponentielle : $\nu = 2$

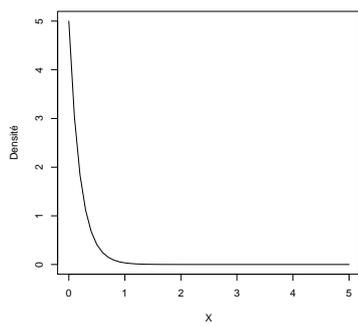


FIG. 1.7 – Exponentielle : $\nu = 5$

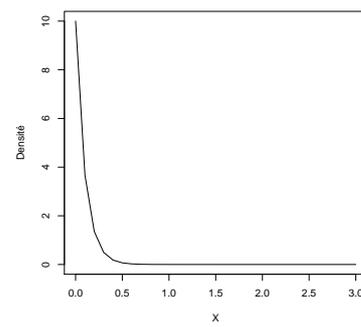


FIG. 1.8 – Exponentielle : $\nu = 10$

On peut dès lors trouver les moments non centrés.

Proposition 1.13. *Les moments non centrés d'une variable aléatoire X distribuée suivant une loi exponentielle de paramètre ν sont donnés par*

$$E[X^n] = \frac{n!}{\nu^n}.$$

Démonstration. La formule (1.6) donne

$$E[X^n] = \frac{\nu n!}{(\nu - t)^{n+1}} \Big|_{t=0} = \frac{n!}{\nu^n}.$$

■

En utilisant la formule précédente, on obtient que l'espérance d'une telle variable vaut $E[X] = \frac{1}{\nu}$ et sa variance $V[X] = \frac{1}{\nu^2}$. Plus ν augmente, plus l'espérance est donc proche de zéro et plus la dispersion est faible, ce que l'on peut voir sur les graphiques de la page 9. La loi exponentielle a une propriété très particulière, à savoir qu'elle est sans mémoire.

Proposition 1.14 (Propriété d'oubli de l'exponentielle). *Soit X une variable aléatoire exponentielle de paramètre ν . Pour tous $s, t \geq 0$,*

$$\mathbb{P}[X > s + t | X > s] = \mathbb{P}[X > t].$$

Démonstration. Soient $t, s \geq 0$.

$$\mathbb{P}[X > s+t | X > s] = \frac{\mathbb{P}[X > s+t \text{ et } X > s]}{\mathbb{P}[X > s]} = \frac{\mathbb{P}[X > s+t]}{\mathbb{P}[X > s]} = \frac{e^{-\nu(s+t)}}{e^{-\nu s}} = e^{-\nu t} = \mathbb{P}[X > t].$$

■

Cette propriété est aussi une condition suffisante pour que la variable X soit exponentielle mais cela ne nous sera pas nécessaire dans la suite.

Définition 1.7. *Fixons $\alpha > 0$ et $\nu > 0$. Une variable aléatoire continue X suit une **loi gamma de paramètres α et ν** , ce que l'on note $X \sim \text{gamma}(\alpha, \nu)$, si elle a pour densité*

$$\text{la fonction } f_X(x) = \begin{cases} \frac{\nu^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\nu x} & \text{si } x > 0 \\ 0 & \text{si } x \leq 0 \end{cases}$$

où $\Gamma(\alpha)$ dénote la fonction gamma de Euler définie pour tout $\alpha > 0$ par

$$\Gamma(\alpha) = \int_0^{+\infty} u^{\alpha-1} e^{-u} du.$$

Si $\alpha = 1$, il s'agit tout simplement une exponentielle négative de paramètre ν . Nous allons regarder comment la fonction de densité varie selon que α ou ν varient. Commençons par regarder l'influence du paramètre α en laissant ν fixe, nous regarderons ensuite ce qu'il se passe lorsque α est fixe et ν augmente. Les graphiques 1.9, 1.10, 1.11, 1.12 montrent que l'augmentation de α implique le déplacement de la masse de probabilité vers la droite. Les graphiques 1.13, 1.14, 1.15, 1.16 montrent quant à eux que lorsque ν augmente, la masse de probabilité se concentre de plus en plus vers l'origine. Ces graphiques ont été obtenus dans \mathbb{R} avec la commande `dgamma(x, alpha, nu)`.

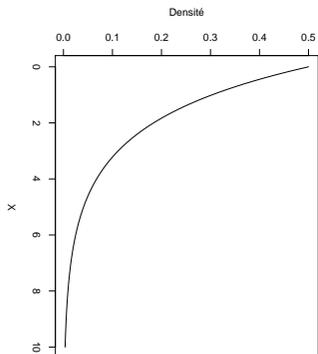


FIG. 1.9 – Gamma : $\alpha = 1, \nu = 0,5$

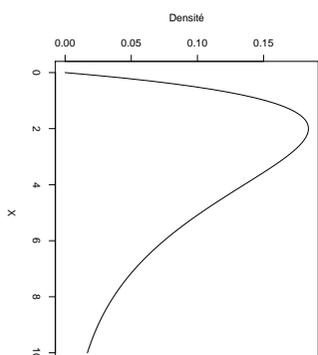


FIG. 1.10 – Gamma : $\alpha = 2, \nu = 0,5$

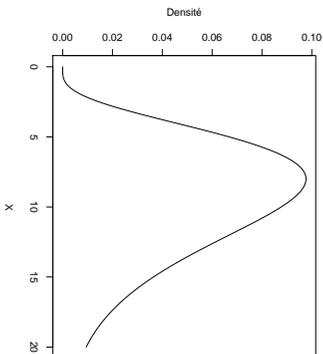


FIG. 1.11 – Gamma : $\alpha = 5, \nu = 0,5$

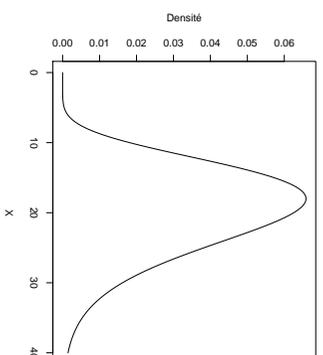


FIG. 1.12 – Gamma : $\alpha = 10, \nu = 0,5$

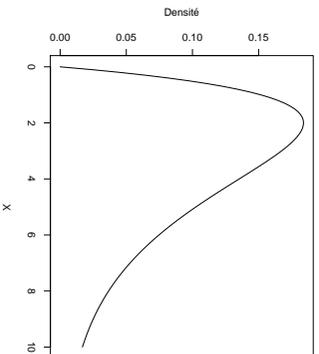


FIG. 1.13 – Gamma : $\alpha = 2, \nu = 0,5$

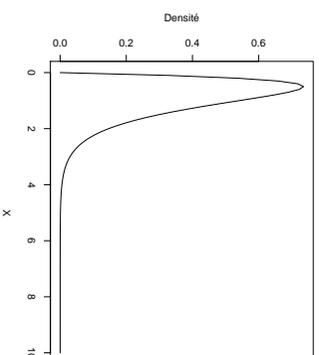


FIG. 1.14 – Gamma : $\alpha = 2, \nu = 2$

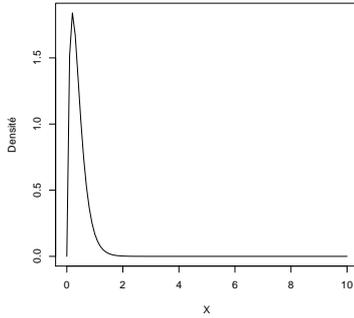


FIG. 1.15 – Gamma : $\alpha = 2, \nu = 5$

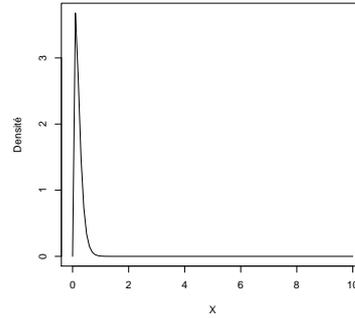


FIG. 1.16 – Gamma : $\alpha = 2, \nu = 10$

Proposition 1.15. *La fonction génératrice des moments d’une variable aléatoire X distribuée selon une loi gamma de paramètres α et ν est donnée par*

$$M_X(t) = \left(\frac{\nu}{\nu - t} \right)^\alpha, \quad \forall t < \nu.$$

Démonstration. Par la définition de l’espérance d’une fonction de variable aléatoire continue et la définition 1.7, nous obtenons

$$M_X(t) = E[e^{tX}] = \int_0^{+\infty} e^{tx} \frac{\nu^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\nu x} dx = \frac{\nu^\alpha}{\Gamma(\alpha)} \int_0^{+\infty} x^{\alpha-1} e^{-(\nu-t)x} dx.$$

Si $t \geq \nu$, l’intégrale n’existe pas. Sinon, nous avons

$$\begin{aligned} M_X(t) &= \frac{\nu^\alpha}{(\nu - t)^\alpha \Gamma(\alpha)} \int_0^{+\infty} ((\nu - t)x)^{\alpha-1} e^{-(\nu-t)x} (\nu - t) dx \\ &= \frac{\nu^\alpha}{(\nu - t)^\alpha \Gamma(\alpha)} \int_0^{+\infty} v^{\alpha-1} e^{-v} dv \quad \text{en posant } v = (\nu - t)x \\ &= \frac{\nu^\alpha}{(\nu - t)^\alpha \Gamma(\alpha)} \Gamma(\alpha) = \left(\frac{\nu}{\nu - t} \right)^\alpha. \end{aligned}$$

■

Cette proposition permet de calculer les moments non centrés.

Proposition 1.16. *Si $X \sim \text{gamma}(\alpha, \nu)$, alors ses moments sont donnés par*

$$E[X^n] = \frac{\Gamma(n + \alpha)}{\Gamma(\alpha) \nu^n}.$$

Démonstration. Par la formule (1.6), on obtient

$$E[X^n] = \alpha(\alpha + 1) \dots (\alpha + n - 1) \frac{\nu^\alpha}{(\nu - t)^{\alpha+n}} \Big|_{t=0} = \frac{\Gamma(\alpha + n)}{\Gamma(\alpha) \nu^n}$$

car³ $\Gamma(n) = (n - 1)! \quad \forall n \in \mathbb{N}$.

■

³cf. J.SCHMETS, ANALYSE MATHÉMATIQUE, Introduction au Calcul Intégral, Liège, Éditions De-rouaux, 1994.

On en déduit que

$$E[X] = \frac{\Gamma(\alpha + 1)}{\Gamma(\alpha)\nu} = \frac{\alpha\Gamma(\alpha)}{\Gamma(\alpha)\nu} = \frac{\alpha}{\nu}$$

vu les propriétés de la fonction gamma de Euler².

Pour la variance, on obtient

$$V[X] = E[X^2] - (E[X])^2 = \frac{\alpha(\alpha + 1)\Gamma(\alpha)}{\Gamma(\alpha)\nu^2} - \left(\frac{\alpha}{\nu}\right)^2 = \frac{\alpha}{\nu^2}.$$

La loi gamma possède une propriété d'additivité assez intéressante :

Proposition 1.17. *Soient X_1, X_2, \dots, X_k des variables aléatoires indépendantes telles que, pour $j = 1, 2, \dots, k$, $X_j \sim \text{gamma}(\alpha_j, \nu)$. La distribution de la variable aléatoire $Y = \sum_{j=1}^k X_j$ est la loi $\text{gamma}(\alpha, \nu)$ avec $\alpha = \sum_{j=1}^k \alpha_j$.*

Démonstration. En tout temps $t \in \mathbb{R}$ tel que l'espérance soit définie, il vient

$$\begin{aligned} M_Y(t) &= E[e^{tY}] = E[e^{t(X_1+X_2+\dots+X_k)}] \\ &= E[e^{tX_1} e^{tX_2} \dots e^{tX_k}] \\ &= E[e^{tX_1}] E[e^{tX_2}] \dots E[e^{tX_k}] \text{ par indépendance des } X_i \\ &= M_{X_1}(t) M_{X_2}(t) \dots M_{X_k}(t) \\ &= \left(\frac{\nu}{\nu-t}\right)^{\alpha_1} \left(\frac{\nu}{\nu-t}\right)^{\alpha_2} \dots \left(\frac{\nu}{\nu-t}\right)^{\alpha_k}, \forall t < \nu \text{ par la propriété 1.15} \\ &= \left(\frac{\nu}{\nu-t}\right)^{\alpha_1+\alpha_2+\dots+\alpha_k}, \forall t < \nu. \end{aligned}$$

On reconnaît alors la fonction génératrice des moments de la loi $\text{gamma}(\alpha, \nu)$ avec $\alpha = \sum_{j=1}^k \alpha_j$. Vu que les fonctions génératrices caractérisent complètement les variables aléatoires, on conclut que $Y \sim \text{gamma}(\alpha, \nu)$ avec $\alpha = \sum_{j=1}^k \alpha_j$. ■

Puisque la loi $\text{gamma}(1, \nu)$ est la loi $\text{Exp}(\nu)$, le corollaire suivant est immédiat :

Corollaire 1.18. *Si X_1, X_2, \dots, X_k sont des variables aléatoires indépendantes et identiquement distribuées selon la loi $\text{Exp}(\nu)$, ce que l'on notera dans la suite*

$$X_i \stackrel{iid}{\sim} \text{Exp}(\nu),$$

et si $Y = \sum_{j=1}^k X_j$, alors $Y \sim \text{gamma}(k, \nu)$.

1.3 Tests d'ajustement

Lorsque l'on désire vérifier une hypothèse statistique concernant la distribution d'un échantillon de données, plusieurs tests statistiques sont à notre disposition, comme par exemple celui de Kolmogorov-Smirnov, celui du χ^2 , celui de Cramer-von Mises, . . . , les plus utilisés étant les deux premiers (car ils sont faciles à réaliser avec n'importe quel logiciel statistique). Cependant, ils ne sont pas sans failles. La distribution exacte de la statistique D_n du test de Kolmogorov-Smirnov n'est connue que si la distribution sous H_0 est complètement spécifiée, c'est-à-dire qu'elle ne peut pas dépendre de paramètres inconnus que l'on estimerait sur l'échantillon à tester. Ce qui n'est pas le cas du second test ; lorsque l'on estime des paramètres sur l'échantillon, il suffit de diminuer le degré de liberté de la statistique du nombre de paramètres estimés. Mais ce test s'applique à des variables discrètes, si bien qu'en présence d'une variable continue, il faut la catégoriser. Le malheur, c'est que la statistique, et donc le résultat, de ce test dépendent de la répartition en classes choisies. Les données représentées par les histogrammes 1.17 et 1.18 illustrent ce phénomène. Une répartition en 18 classes donne une p-value de 0,04 alors qu'une

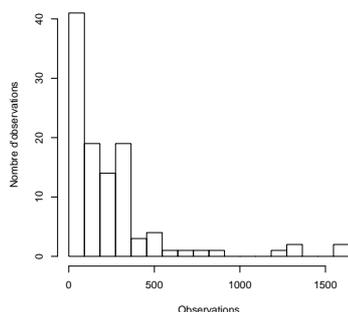


FIG. 1.17 – Histogramme des observations avec 18 classes

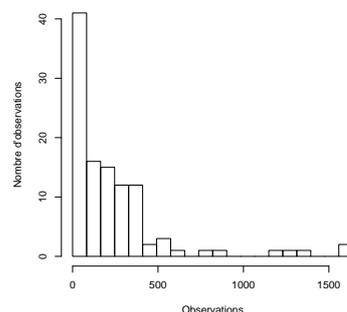


FIG. 1.18 – Histogramme des observations avec 20 classes

répartition en 20 classes fournit une p-value acceptable de 0,07.

Le problème rencontré dans certains chapitres sera de tester si un vecteur d'observations $w = (w_1, \dots, w_n)$, qui seront en fait des temps, suit une distribution exponentielle de paramètre ν inconnu. Il faudra l'estimer sur le vecteur w ; ce qui peut se faire en utilisant, par exemple, l'estimateur du maximum de vraisemblance donné par

$$\hat{\nu} = \frac{n}{\sum_{i=1}^n w_i} = \frac{1}{\bar{w}}.$$

Il s'agit donc d'un test d'ajustement sur une distribution continue de paramètre inconnu, situation pour laquelle le test de Kolmogorov-Smirnov n'est pas adéquat et le test du χ^2 pourrait ne pas être fiable. J. DURBIN s'est intéressé au test de Kolmogorov-Smirnov en cas de paramètres inconnus dans un article intitulé "Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings" publié dans *Biometrika* 62 en 1975 (pp.5-22). Cet article présente une technique pour obtenir la distribution exacte de la statistique D_n de Kolmogorov-Smirnov lorsque les paramètres sont estimés sur l'échantillon à tester. Cette technique est alors appliquée au cas d'une distribution exponentielle pour lequel une table des seuils critiques c_n tels que $\mathbb{P}[\sqrt{n}D_n > c_n] = \alpha$ est donnée pour plusieurs valeurs de α . Cette table ne reprend pas toutes les valeurs de n possibles et s'arrête à $n = 100$. Pour les valeurs de n plus petites que 100 mais ne se trouvant pas dans la table, une interpolation linéaire a été réalisée. Le vecteur des seuils critiques d_n tel que $d_n = \frac{c_n}{\sqrt{n}}$ avec $n \leq 100$ est donnée en annexe pour $\alpha = 0,05$. Pour les valeurs plus grandes que 100, la méthode du bootstrap paramétrique⁴ a été utilisée de manière à obtenir une estimation de la valeur du seuil critique. Le graphique 1.19 montre que l'on peut considérer que pour $n > 100$, $\sqrt{n}d_n = 1,08$ peut être pris comme approximation de la valeur réelle.

⁴Une introduction à cette méthode est donnée dans le livre de EFRON B. & TIBSHIRANI R.J., *An introduction to the bootstrap*, New-York, Chapman & Hall, 1993.

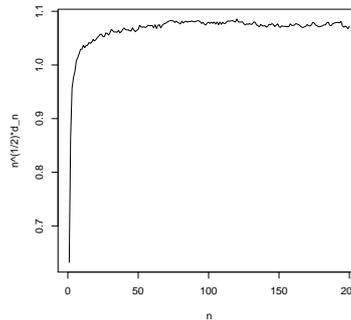


FIG. 1.19 – Seuils critiques obtenus par la méthode du Bootstrap paramétrique

1.4 Introduction aux processus stochastiques

Définition 1.8. *Un processus stochastique est un ensemble infini $\{X_t : t \in T\}$ de variables aléatoires X_t , ou encore une application de T vers l'espace des variables aléatoires définies sur un même espace Ω et à valeurs dans \mathcal{S} .*

A titre d'exemple, si $T = \{1, 2, \dots\}$, la variable X_n pourrait être le résultat obtenu lors du n -ième lancer d'un dé, $n \in T$. Une réalisation du processus serait 5, 1, 3, 2, 2, 4, 1, 6, 3, 6, \dots . L'ensemble T est l'espace des paramètres, aussi appelé espace du temps car le paramètre $t \in T$ est souvent un paramètre temporel.

Si T est dénombrable, on parle de processus stochastique à temps discret et si T est un intervalle de \mathbb{R} , on parle alors de processus stochastique à temps continu. Lorsque le processus est à temps continu, on remplace X_t par $X(t)$.

La même distinction est faite sur l'espace \mathcal{S} des valeurs des variables aléatoires X_t , appelé espace d'états; on parle de processus stochastique discret ou continu.

Pour un événement ω fixé dans Ω , la fonction $t \mapsto X_t(\omega)$ est appelée trajectoire du processus associée au point ω . Pour visualiser la notion de trajectoire, voici un processus stochastique très simple.

Exemple 1.1. Soient des variables aléatoires X_1, \dots, X_n qui suivent une distribution de Bernoulli de paramètre p , c-à-d que $\forall i = 1, \dots, n$, X_i peut prendre la valeur 1 avec une probabilité p et la valeur 0 avec une probabilité $1 - p$. On définit alors le processus discret à temps discret $\{Y_n, n \in \mathbb{N}\}$ par $Y_n = \sum_{i=1}^n X_i$. Prenons par exemple $n = 10$ et $p = 0,6$. Trois échantillons de 10 variables de Bernoulli de paramètre 0,6 ont été générés grâce à la commande `rbinom(10,1,0.6)`. A partir de ces échantillons, les valeurs des Y_n pour $n = 1, \dots, 10$ ont alors été calculées. La figure 1.20 représente trois trajectoires possibles du processus.

Définition 1.9. *On appelle marge du processus $\{X_t, t \in T\}$ tout vecteur aléatoire*

$$(X_{t_1}, X_{t_2}, \dots, X_{t_k}), k \in \mathbb{N}_0, t_1 < t_2 < \dots < t_k \in T.$$

Les distributions de ces vecteurs aléatoires sont appelées les distributions de dimension finie du processus $\{X_t, t \in T\}$.

Lorsque l'espace d'états est réel, les distributions de dimension finie caractérisent complètement le processus stochastique. Plus généralement même, les marges finies de tout vecteur

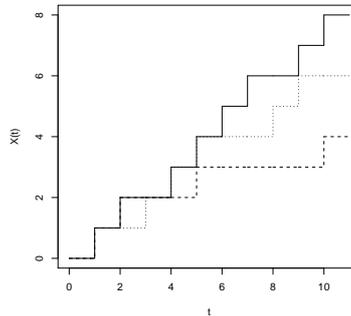


FIG. 1.20 – Différentes trajectoires pour $\{Y_n, n \in \mathbb{N}\}$.

infini de variables aléatoires réelles caractérisent celui-ci⁵.

Tous les processus que nous allons rencontrer dans la suite seront à espace d'états réel. Pour déterminer univoquement les processus qui seront étudiés, il suffira de donner les distributions de dimension finie de ceux-ci.

Dans la suite nous nous intéresserons souvent aux deux propriétés suivantes des processus stochastiques :

Définition 1.10. *Le processus stochastique $\{X_t, t \in T\}$ a des **accroissements indépendants** si $\forall k \in \mathbb{N}_0$ et $\forall t_0 < t_1 < t_2 < \dots < t_k$, les variables aléatoires $X_{t_1} - X_{t_0}$, $X_{t_2} - X_{t_1}, \dots, X_{t_k} - X_{t_{k-1}}$ sont indépendantes. Le processus stochastique $\{X_t, t \in T\}$ a des **accroissements stationnaires** si $\forall h > 0$, les variables aléatoires $X_{t+h} - X_t$ ont la même distribution $\forall t \in \mathbb{R}^+$.*

Intuitivement, la première propriété signifie que les occurrences du processus dans un intervalle sont indépendantes des occurrences du processus sur tout autre intervalle disjoint et la seconde veut dire que la distribution des occurrences du processus dans tout intervalle dépend juste de la longueur de cet intervalle.

Définition 1.11. *Deux processus stochastiques $\{X_t, t \in T\}$ et $\{Y_t, t \in T\}$ sont dits **indépendants** si tout événement défini à partir du premier est indépendant de tout événement défini à partir du second, c'est-à-dire si pour tout $n \in \mathbb{N}_0$ et tout ensemble de n éléments t_1, \dots, t_n de T , les vecteurs aléatoires $(X_{t_1}, \dots, X_{t_n})$ et $(Y_{t_1}, \dots, Y_{t_n})$ sont indépendants.*

Voici encore deux nouveaux concepts utiles en théorie des processus stochastiques, ce sont les fonctions moyenne et covariance.

Définition 1.12. *Soit $\{X_t, t \in T\}$ un processus stochastique quelconque. Sous réserve d'existence, on appelle **fonction moyenne** de ce processus la fonction qui, à tout $t \in T$, associe $E[X_t]$ et **fonction covariance** ou **auto-covariance** la fonction qui, à tout couple $(t, s) \in T \times T$, associe $Cov[X_s, X_t]$.*

⁵Une démonstration de ce fait, qui sera essentiel pour la suite de ce travail, peut être trouvée dans le livre A.N.KOLMOGOROV, *Foundations of the theory of probability*, Second English Edition by N.MORRISON, Chelsea Publishing Compagny, New-York, 1956, pp.27-30.

Le type de processus stochastique étudié dans la suite porte le nom de processus de comptage. Supposons que des événements se produisent au cours du temps et notons T_n la variable aléatoire "Temps d'occurrence du n -ième événement" pour tout $n \in \mathbb{N}$. Le processus $\{T_n, n \in \mathbb{N}\}$, appelé *processus des temps d'occurrence*, est continu à temps discret. Par convention, on pose $T_0 = 0$. A partir de ce processus des temps d'occurrence, il est possible de définir le processus de comptage.

Définition 1.13. *Un processus stochastique $\{N(t), t \in \mathbb{R}^+\}$ est un processus de comptage si $N(t)$ représente le nombre total d'événements qui se sont produits entre 0 et t ; il doit donc satisfaire*

- i) $N(t) \geq 0$;
- ii) $N(t)$ a des valeurs entières uniquement;
- iii) Si $s < t$, alors $N(s) \leq N(t)$;
- iv) Pour $s < t$, $N(t) - N(s)$ est le nombre d'événements qui ont eu lieu entre s et t .

Un processus de comptage est un processus discret à temps continu. Un second processus peut être associé au processus des temps d'occurrence; le *processus des temps d'inter-arrivées* $\{W_n, n \in \mathbb{N}_0\}$ où $\forall n \in \mathbb{N}_0$, la variable aléatoire W_n est le temps d'attente entre les $(n - 1)$ -ième et n -ième occurrences, c-à-d

$$W_n = T_n - T_{n-1}.$$

Les relations suivantes sont triviales à vérifier :

- $W_n = T_1 + T_2 + \dots + T_n \quad \forall n \geq 1$;
- $N(t) = \sup\{n \in \mathbb{N} | T_n \leq t\}$;
- $\mathbb{P}[N(t) = n] = \mathbb{P}[T_n \leq t < T_{n+1}]$;
- $\mathbb{P}[N(t) \geq n] = \mathbb{P}[T_n \leq t]$;
- $\mathbb{P}[s < T_n < t] = \mathbb{P}[N(s) < n \leq N(t)]$.

Pour terminer ce chapitre, ces différentes notions peuvent être visualisées sur le schéma suivant qui représente une trajectoire possible pour chacun d'entre eux.

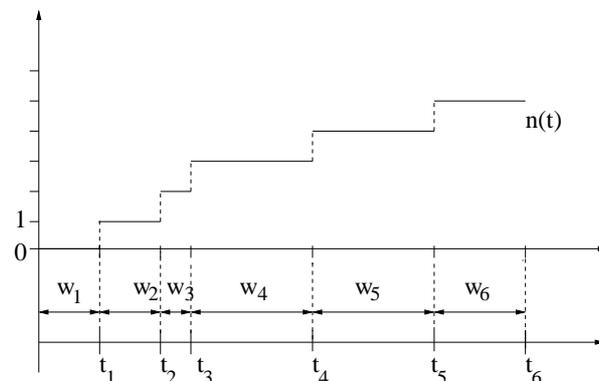


FIG. 1.21 – Processus de comptage et processus associés

Chapitre 2

Processus de Poisson de base

C'est ici que débute réellement l'étude des processus de Poisson. Le premier type abordé est le plus simple et aussi le plus répandu dans la littérature. Ce processus de base requiert notamment les hypothèses de stationnarité des accroissements et de "disjonction" des événements (hypothèse qui empêche que plusieurs événements se produisent simultanément) qui seront levées dans les chapitres suivants.

Définition 2.1. *Un processus de comptage $\{N(t), t \geq 0\}$ est un processus de Poisson d'intensité $\nu > 0$ si*

- $N(0) = 0$;
- *Le processus est à accroissements stationnaires;*
- *Le processus est à accroissements indépendants;*
- $\forall 0 \leq s < t$, *la variable aléatoire $N(t) - N(s)$ suit une loi de Poisson de paramètre $\nu(t - s)$.*

Le paramètre d'intensité ν d'un processus de Poisson est tel que, $\forall t \geq 0$, $E[N(t)] = \nu t$ car $N(t) \sim \mathcal{P}(\nu t)$. Cela étant, pour $t = 1$, il vient $E[N(1)] = \nu$. Or $N(1)$ représente le nombre d'événements qui ont eu lieu dans l'intervalle $[0, 1]$. Par stationnarité des accroissements, le paramètre d'intensité ν est le nombre attendu d'événements par unité de temps.

L'impact de ce paramètre sur la fonction $N(t)$ peut être visualisé sur les figures 2.1 et 2.2 qui représentent deux trajectoires du processus $\{N(t), t \geq 0\}$ pour deux valeurs différentes de ν . Le nombre d'événements atteint plus vite 10 lorsque l'intensité est plus grande, ce qui est logique vu l'interprétation de ν donnée avant. Il sera expliqué à la section 2.3 comment générer de telles trajectoires.

Rien qu'avec la définition 2.1, il est possible de démontrer une propriété très importante des processus de Poisson, à savoir que *les processus de Poisson sont sans mémoire*, c-à-d que le fait de connaître le comportement du processus à un instant donné ne modifie pas les prévisions pour le futur. En effet, la probabilité d'avoir k événements dans l'intervalle $[t, t + s]$

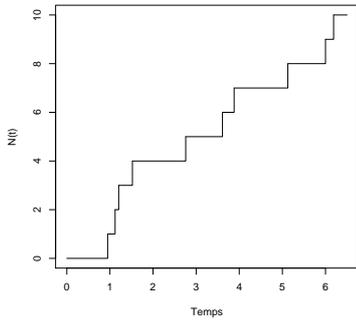


FIG. 2.1 – Trajectoire lorsque $\nu = 2$

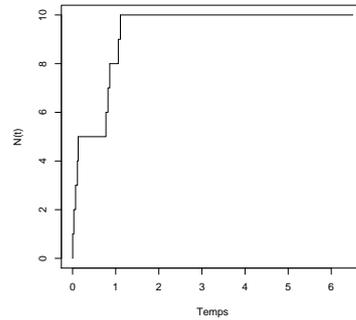


FIG. 2.2 – Trajectoire lorsque $\nu = 10$

est indépendante du nombre d'événements qui ont eu lieu avant t :

$$\begin{aligned}
 \mathbb{P}[N(t+s) - N(t) = k | N(t) = n] &= \frac{\mathbb{P}[N(t+s) - N(t) = k \text{ et } N(t) - N(0) = n]}{\mathbb{P}[N(t) = n]} \\
 &= \frac{\mathbb{P}[N(t+s) - N(t) = k] \mathbb{P}[N(t) = n]}{\mathbb{P}[N(t) = n]} && \text{par indépendance} \\
 &= \mathbb{P}[N(s) = k] && \text{par stationnarité} \\
 &= \frac{e^{-\nu s} (\nu s)^k}{k!}
 \end{aligned}$$

pour tous $k, n \in \mathbb{N}$ et pour tous $t, s \in \mathbb{R}$ tels que $t \geq 0$ et $s > 0$.

Pour illustrer cette propriété, prenons l'exemple de la pêche. Soit $N(t)$ le nombre de poissons attrapés sur l'intervalle de temps $[0, t]$. Sous les hypothèses que le nombre de poissons disponible est grand, qu'en tout instant, ils sont susceptibles de mordre à l'hameçon et que tous les pêcheurs ont la même chance d'en attraper, le processus $\{N(t), t \geq 0\}$ peut être considéré comme un processus de Poisson. La propriété d'être sans mémoire s'exprime alors comme ceci : un pêcheur n'ayant encore rien pêché à l'instant t a la même chance de pêcher un poisson d'ici l'instant $t + s$ qu'un autre pêcheur qui lui en aurait déjà pêché 5.

2.1 Distribution des marges finies du processus

Proposition 2.1. *Un processus de comptage $\{N(t), t \geq 0\}$ est un processus de Poisson si et seulement si il vérifie les conditions suivantes :*

- (C1) $N(0) = 0$;
- (C2) Les accroissements sont indépendants ;
- (C3) $\forall t > 0, 0 < \mathbb{P}[N(t) > 0] < 1$;
- (C4) $\forall t \geq 0, \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) \geq 2]}{\mathbb{P}[N(t+h) - N(t) = 1]} = 0$;
- (C5) Les accroissements sont stationnaires.

Démonstration. Vérifions d'abord que la condition est nécessaire. Les conditions (C1), (C2) et (C5) sont vérifiées vu la définition 2.1.

Pour tout $t > 0$, $\mathbb{P}[N(t) > 0] = 1 - \mathbb{P}[N(t) = 0] = 1 - e^{-\nu t}$ car $N(t) \sim \mathcal{P}(\nu t)$ pour tout $t > 0$. On a donc de suite que $\mathbb{P}[N(t) > 0] < 1$. L'autre inégalité vient du fait que $e^{-\nu t} = 1$ ssi $t = 0$ car $\nu > 0$. Ce qui donne (C3).

Enfin, par stationnarité, la condition (C4) équivaut à $\lim_{h \rightarrow 0} \frac{\mathbb{P}[N(h) \geq 2]}{\mathbb{P}[N(h) = 1]} = 0$. Montrons que cette limite est bien nulle.

$$\begin{aligned} \lim_{h \rightarrow 0} \frac{\mathbb{P}[N(h) \geq 2]}{\mathbb{P}[N(h) = 1]} &= \lim_{h \rightarrow 0} \frac{1 - \mathbb{P}[N(h) = 0] - \mathbb{P}[N(h) = 1]}{\mathbb{P}[N(h) = 1]} \\ &= \lim_{h \rightarrow 0} \frac{1 - e^{-\nu h} - e^{-\nu h} \nu h}{e^{-\nu h} \nu h} \end{aligned}$$

qui est un cas d'indétermination que le théorème de l'Hospital peut lever

$$\begin{aligned} &= \lim_{h \rightarrow 0} \frac{\nu e^{-\nu h} + \nu e^{-\nu h} \nu h - \nu e^{-\nu h}}{\nu e^{-\nu h} - \nu e^{-\nu h} \nu h} \\ &= \lim_{h \rightarrow 0} \frac{\nu h}{1 - \nu h} = 0. \end{aligned}$$

Passons alors au caractère suffisant de la preuve.

Vu les hypothèses (C1), (C2) et (C5), les trois premiers points de la définition 2.1 sont vérifiés. Comme les accroissements sont stationnaires, il suffit de trouver une constante $\nu > 0$ telle que $\forall t > 0$, $N(t) \sim \mathcal{P}(\nu t)$. Pour cela, montrons qu'il existe une constante $\nu > 0$ telle que la fonction génératrice des probabilités de $N(t)$, que l'on notera $G(z, t)$ pour rappeler l'indice t , soit

$$e^{\nu t(z-1)}, \quad \forall |z| \leq 1. \quad (2.1)$$

Vu la proposition 1.10, cette expression correspond à la fonction génératrice des probabilités d'une variable aléatoire distribuée selon une loi de Poisson de paramètre νt même si elle n'est valable que sur $[-1, 1]$ car en évaluant en 1 les dérivées, on obtient (comme cela est expliqué à la page 2) les moments qui caractérisent complètement la variable. Montrons d'abord que l'on obtient effectivement l'expression (2.1) pour la fonction génératrice des probabilités lorsqu'il existe $\nu > 0$ vérifiant les trois propriétés suivantes :

$$\lim_{h \rightarrow 0^+} \frac{1 - \mathbb{P}[N(h) = 0]}{h} = \nu; \quad (2.2)$$

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(h) = 1]}{h} = \nu; \quad (2.3)$$

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(h) \geq 2]}{h} = 0. \quad (2.4)$$

Pour cela, on considère

$$E[z^{N(t+h)}] = E[z^{N(t+h)-N(t)+N(t)}] = E[z^{N(t+h)-N(t)} z^{N(t)-N(0)}] = E[z^{N(h)}] E[z^{N(t)}]$$

vu les hypothèses (C2) et (C5).

Vu cela, et par définition des fonctions génératrices des probabilités, il vient $G(z, t+h) = G(z, h)G(z, t)$ ou encore,

$$\frac{G(z, t+h) - G(z, t)}{h} = \frac{G(z, h)G(z, t) - G(z, t)}{h} = \frac{G(z, t)(G(z, h) - 1)}{h}. \quad (2.5)$$

Or,

$$\lim_{h \rightarrow 0^+} \frac{G(z, h) - 1}{h} = \lim_{h \rightarrow 0^+} \left[\frac{z^0 \mathbb{P}[N(h) = 0] - 1}{h} + \frac{z^1 \mathbb{P}[N(h) = 1]}{h} + \frac{\sum_{n=2}^{+\infty} z^n \mathbb{P}[N(h) = n]}{h} \right]$$

et, pour $|z| \leq 1$, la série $\sum_{n=2}^{+\infty} z^n \mathbb{P}[N(h) = n]$ converge absolument, ce qui permettra d'échanger la limite et la série car le terme de cette série est fini lorsque h tend vers zéro¹. De plus, comme nous avons $\forall n \geq 2$,

$$0 \leq \frac{\mathbb{P}[N(h) = n]}{h} \leq \frac{\mathbb{P}[N(h) \geq 2]}{h},$$

la relation (2.4) et le théorème de l'étau² impliquent alors que

$$\lim_{h \rightarrow 0} \frac{\mathbb{P}[N(h) = n]}{h} = 0, \quad \forall n \geq 2.$$

En mettant tout cela avec les relations (2.2) et (2.3), on obtient

$$\begin{aligned} \lim_{h \rightarrow 0^+} \frac{G(z, h) - 1}{h} &= \lim_{h \rightarrow 0^+} \frac{z^0 \mathbb{P}[N(h) = 0] - 1}{h} + z \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(h) = 1]}{h} + \sum_{n=2}^{+\infty} z^n \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(h) = n]}{h} \\ &= -\nu + \nu z = \nu(z - 1). \end{aligned}$$

De là, en repartant de (2.5), on a

$$\frac{\partial}{\partial t} G(z, t) = \lim_{h \rightarrow 0^+} \left[\frac{G(z, t+h) - G(z, t)}{h} \right] = G(z, t) \nu(z - 1), \quad \forall |z| \leq 1.$$

Pour tout z fixé dans $[-1, 1]$, c'est une équation différentielle à second membre séparé de condition initiale $G(z, 0) = E[z^{N(0)}] = 1$ dont la solution est donnée par³ $G(z, t) = e^{\nu t(z-1)}$; ce qui donne bien (2.1).

Il reste à montrer qu'il existe une constante $\nu > 0$ satisfaisant les équations (2.2), (2.3) et (2.4). Posons $P_k(t) = \mathbb{P}[N(t) = k]$, pour $k \in \mathbb{N}$. Pour $t_1, t_2 \in \mathbb{R}^+$,

$$\begin{aligned} \mathbb{P}[N(t_1 + t_2) = 0] &= \mathbb{P}[N(t_1 + t_2) - N(t_1) = 0 \text{ et } N(t_1) = 0] \\ &= \mathbb{P}[N(t_1 + t_2) - N(t_1) = 0] \mathbb{P}[N(t_1) = 0]. \end{aligned}$$

Vu l'hypothèse (C5), $N(t_1 + t_2) - N(t_1) \sim N(t_2)$, d'où $P_0(t_1 + t_2) = P_0(t_1)P_0(t_2)$.

Comme $P_0(t)$ est une fonction bornée, on a ⁴ $P_0 \equiv 0$ ou $\exists \nu > 0$ telle que $P_0(t) = e^{-\nu t} \forall t \geq 0$.

Vu l'hypothèse (C3), $P_0(t)$ n'est pas identiquement nul et donc il existe ν tel que $\mathbb{P}[N(t) = 0] = e^{-\nu t}, \forall t \geq 0$. En effet, $\forall t > 0, 0 < \mathbb{P}[N(t) > 0] = 1 - P_0(t) < 1$ implique que, $\forall t > 0, 0 < P_0(t) < 1$.

¹cfr J.SCHMETS, *ANALYSE MATHÉMATIQUE, Introduction aux Espaces Fonctionnels*, Université de Liège, année académique 2000-2001

²cfr J.SCHMETS, *ANALYSE MATHÉMATIQUE*, Liège, Éditions Derouaux, 1996.

³cfr J.SCHMETS, *ANALYSE MATHÉMATIQUE*, Liège, Éditions Derouaux, 1996.

⁴Si f est une fonction réelle telle que $f(t+s) = f(t)f(s) \quad \forall s, t > 0$ et qui est bornée dans tout intervalle fini, alors soit $f \equiv 0$ soit $\exists \nu > 0$ telle que $f(t) = e^{-\nu t}$.

(voir PARZEN, E., *Stochastic Processes*, Stanford University, Holden-Day, p.121).

Ainsi, $\mathbb{P}[N(h) = 0] = e^{-\nu h} = \sum_{n=0}^{+\infty} \frac{(-\nu h)^n}{n!} = 1 - \nu h + o(h)$ où la fonction $o(h)$ est telle que

$\lim_{h \rightarrow 0} \frac{o(h)}{h} = 0$, c-à-d que la fonction $o(h)$ tend vers zéro plus vite que h .

Dès lors, $\lim_{h \rightarrow 0^+} \frac{1 - \mathbb{P}[N(h) = 0]}{h} = \lim_{h \rightarrow 0^+} \frac{1 - 1 + \nu h - o(h)}{h} = \nu - \lim_{h \rightarrow 0^+} \frac{o(h)}{h} = \nu$, ce qui permet d'obtenir (2.2).

Posons $Q(t) = \mathbb{P}[N(t) \geq 2] = \sum_{k=2}^{+\infty} P_k(t)$ et constatons que

$$\frac{P_1(h)}{h} \left(1 + \frac{Q(h)}{P_1(h)} \right) = \frac{P_1(h)}{h} + \frac{Q(h)}{h} = \frac{P_1(h) + Q(h)}{h} = \frac{1 - P_0(h)}{h}.$$

Vu les hypothèses (C4) et (C5), on a $\lim_{h \rightarrow 0^+} \frac{Q(h)}{P_1(h)} = 0$.

Dès lors, d'une part, on a

$$\lim_{h \rightarrow 0^+} \left[\frac{P_1(h)}{h} \left(1 + \frac{Q(h)}{P_1(h)} \right) \right] = \lim_{h \rightarrow 0^+} \frac{P_1(h)}{h} = \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(h) = 1]}{h},$$

et d'autre part,

$$\lim_{h \rightarrow 0^+} \left[\frac{P_1(h)}{h} \left(1 + \frac{Q(h)}{P_1(h)} \right) \right] = \lim_{h \rightarrow 0^+} \frac{1 - P_0(h)}{h} = \nu,$$

ce qui permet de démontrer (2.3).

Enfin, vu que $\nu = \lim_{h \rightarrow 0^+} \frac{P_1(h)}{h}$, il vient

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(h) \geq 2]}{h} = \lim_{h \rightarrow 0^+} \frac{Q(h)}{h} = \lim_{h \rightarrow 0^+} \frac{Q(h)}{h} \nu \frac{1}{\nu} = \nu \lim_{h \rightarrow 0^+} \frac{Q(h)}{h} \frac{h}{P_1(h)} = \nu \lim_{h \rightarrow 0^+} \frac{Q(h)}{P_1(h)} = 0,$$

et (2.4) est ainsi démontré. ■

Il est maintenant possible de donner la loi de toute marge finie du processus de Poisson $\{N(t), t \geq 0\}$.

$\forall 0 < t_1 < t_2 < \dots < t_k, \forall n_1 \leq n_2 \leq \dots \leq n_k \in \mathbb{N}, k \in \mathbb{N}_0$,

$$\mathbb{P}[N(t_1) = n_1, N(t_2) = n_2, \dots, N(t_k) = n_k]$$

$$= \mathbb{P}[N(t_1) - N(0) = n_1, N(t_2) - N(t_1) = n_2 - n_1, \dots, N(t_k) - N(t_{k-1}) = n_k - n_{k-1}]$$

$$= \mathbb{P}[N(t_1) = n_1] \mathbb{P}[N(t_2) - N(t_1) = n_2 - n_1] \dots \mathbb{P}[N(t_k) - N(t_{k-1}) = n_k - n_{k-1}] \text{ par indépendance}$$

$$= \mathbb{P}[N(t_1) = n_1] \mathbb{P}[N(t_2 - t_1) = n_2 - n_1] \dots \mathbb{P}[N(t_k - t_{k-1}) = n_k - n_{k-1}] \text{ par stationnarité}$$

$$= e^{-\nu t_1} \frac{(\nu t_1)^{n_1}}{n_1!} e^{-\nu(t_2 - t_1)} \frac{(\nu(t_2 - t_1))^{n_2 - n_1}}{(n_2 - n_1)!} \dots e^{-\nu(t_k - t_{k-1})} \frac{(\nu(t_k - t_{k-1}))^{n_k - n_{k-1}}}{(n_k - n_{k-1})!}$$

$$= e^{-\nu t_k} \prod_{j=1}^k \frac{(\nu(t_j - t_{j-1}))^{n_j - n_{j-1}}}{(n_j - n_{j-1})!} \quad \text{où } t_0 = 0 \text{ et } n_0 = 0.$$

Voici l'expression des fonctions moyenne et covariance du processus.

Proposition 2.2. *Soit un processus de Poisson $\{N(t), t \geq 0\}$ d'intensité ν . Pour tous $t, s \geq 0$,*

$$E[N(t)] = \nu t \tag{2.6}$$

$$Cov[N(t), N(s)] = \nu \min(s, t). \tag{2.7}$$

Démonstration. Le résultat de l'espérance est immédiat car $N(t) \sim \mathcal{P}(\nu t)$. Notons de plus que $Cov[N(t), N(t)] = V[N(t)] = \nu t = \nu \min(t, t)$.

Pour le reste du calcul, la relation suivante est nécessaire

$$N(t)N(s) = \frac{(N(t))^2 + (N(s))^2 - (N(t) - N(s))^2}{2}, \quad \forall s < t. \tag{2.8}$$

Prenons alors $s < t$,

$$\begin{aligned} Cov[N(t), N(s)] &= E[(N(t) - E[N(t)])(N(s) - E[N(s)])] \\ &= E[(N(t) - \nu t)(N(s) - \nu s)] \quad \text{vu (2.6)} \\ &= E[N(t)N(s)] - \nu t E[N(s)] - \nu s E[N(t)] + \nu^2 ts \quad \text{vu la linéarité de l'espérance} \\ &= \frac{E[(N(t))^2] + E[(N(s))^2] - E[(N(t) - N(s))^2]}{2} - \nu^2 ts \quad \text{en utilisant (2.8) et (2.6)} \\ &= \frac{V[N(t)] + (E[N(t)])^2 + V[N(s)] + (E[N(s)])^2 - V[N(t) - N(s)] - (E[N(t) - N(s)])^2}{2} - \nu^2 ts \\ &= \frac{\nu t + (\nu t)^2 + \nu s + (\nu s)^2 - \nu(t - s) - \nu^2(t - s)^2}{2} - \nu^2 ts \quad \text{par stationnarité} \\ &= \nu s = \nu \min(s, t) \end{aligned}$$

car $s < t$. ■

2.2 Définition alternative

La définition 2.1 n'est pas la définition la plus répandue puisque, lorsque l'on parle de processus de Poisson, on rencontre plus souvent cette définition-ci :

Définition 2.2. *Un processus de comptage $\{N(t), t \geq 0\}$ est un processus de Poisson d'intensité $\nu > 0$ si*

- $N(0) = 0$;
- *Le processus est à accroissements stationnaires;*
- *Le processus est à accroissements indépendants;*
- $IP[N(t+h) - N(t) = k] = IP[N(h) = k] = \begin{cases} 1 - \nu h + o(h) & \text{si } k = 0 \\ \nu h + o(h) & \text{si } k = 1 \\ o(h) & \text{si } k \geq 2. \end{cases}$

La proposition suivante justifie l'emploi de la définition 2.1.

Proposition 2.3. *Les définitions 2.1 et 2.2 sont équivalentes.*

Démonstration. Montrons tout d'abord que la définition 2.2 implique l'autre. Pour tout $t > 0$, posons $\psi_t = \psi_{N(t)}(u) = E[e^{-uN(t)}]$, pour $u \in \mathbb{R}^+$ fixé, la transformée de Laplace de la variable $N(t)$. Par stationnarité et indépendance des accroissements, on obtient

$$\psi_{t+h} = E[e^{-uN(t+h)}] = E[e^{-uN(t)}] E[e^{-u(N(t+h)-N(t))}] = \psi_t E[e^{-uN(h)}]. \quad (2.9)$$

Cherchons alors l'expression de $E[e^{-uN(h)}] = \sum_{n=0}^{+\infty} e^{-un} \mathbb{P}[N(h) = n]$. Puisque

$$\sum_{n=2}^{+\infty} e^{-un} \mathbb{P}[N(h) = n] \leq \sum_{n=2}^{+\infty} \mathbb{P}[N(h) = n] = \mathbb{P}[N(h) \geq 2] = 1 - \mathbb{P}[N(h) = 0] - \mathbb{P}[N(h) = 1],$$

les hypothèses et le théorème de l'étau impliquent que $\sum_{n=2}^{+\infty} e^{-un} \mathbb{P}[N(h) = n]$ est de l'ordre de $o(h)$. De là,

$$\begin{aligned} E[e^{-uN(h)}] &= \mathbb{P}[N(h) = 0] + e^{-u} \mathbb{P}[N(h) = 1] + o(h) \\ &= 1 - \nu h + o(h) + e^{-u}(\nu h + o(h)) + o(h) \quad \text{par hypothèse} \\ &= 1 - \nu h + e^{-u} \nu h + o(h). \end{aligned}$$

A partir de (2.9), il vient donc que $\psi_{t+h} = \psi_t(1 - \nu h + e^{-u} \nu h) + o(h)$.

Par conséquent, on a $\frac{\psi_{t+h} - \psi_t}{h} = \psi_t(e^{-u} - 1)\nu + \frac{o(h)}{h}$, et en passant à la limite sur h tendant vers zéro, on obtient

$$D_t \psi_t = (e^{-u} - 1)\nu \psi_t.$$

C'est une équation différentielle à second membre séparé dont une solution est donnée par $C \exp(\nu t(e^{-u} - 1))$. Or en $t = 0$, on a $\psi_0 = 1$, ce qui donne

$$\psi_t = \exp(\nu t(e^{-u} - 1)).$$

Le lien entre la transformée de Laplace et la fonction génératrice des moments est tel que⁵, $\forall u \in \mathbb{R}^+$, $M_{N(t)}(u) = \psi_{N(t)}(-u) = \exp(\nu t(e^u - 1))$. Comme cela a été vu dans le chapitre 1, cette dernière expression correspond à la fonction génératrice des moments d'une variable aléatoire dont la distribution est une loi de Poisson de paramètre νt . Ainsi, on a donc $N(t) \sim \mathcal{P}(\nu t)$, $\forall t > 0$.

Comme par hypothèses, les accroissements sont stationnaires et indépendants et $N(0) = 0$, on retrouve bien la définition 2.1.

Considérons alors l'autre implication. La définition 2.1 implique que $\mathbb{P}[N(h) = k] = e^{-\nu h} \frac{(\nu h)^k}{k!}$.

$$\text{Pour } k = 0, \mathbb{P}[N(h) = 0] = e^{-\nu h} = \sum_{i=0}^{+\infty} \frac{(-\nu h)^i}{i!} = 1 - \nu h + o(h).$$

$$\text{Pour } k = 1, \mathbb{P}[N(h) = 1] = e^{-\nu h} \nu h = \nu h(1 - \nu h + o(h)) = \nu h + o(h).$$

$$\text{Pour } k \geq 2, \mathbb{P}[N(h) = k] = e^{-\nu h} \frac{(\nu h)^k}{k!} = (1 - \nu h + o(h)) \frac{(\nu h)^k}{k!} = o(h).$$

Cela suffit pour conclure. ■

⁵voir équation (1.12)

Cette définition est plus simple à interpréter car elle exprime clairement que la probabilité d'avoir simultanément deux événements ou plus est nulle. Cependant, le choix de la définition 2.1 a été fait pour être cohérent avec les définitions qui seront introduites dans les chapitres ultérieurs.

2.3 Temps d'inter-arrivées et temps d'occurrence

La proposition suivante donne un moyen simple de générer des processus de Poisson ; celui-ci sera expliqué juste après.

Proposition 2.4. *Soient $\{N(t), t \geq 0\}$ un processus de comptage et $\{W_n, n \in \mathbb{N}_0\}$ le processus des temps d'inter-arrivées associé à ce processus. Le processus de comptage est un processus de Poisson d'intensité $\nu > 0$ si et seulement si les variables aléatoires W_n sont indépendantes et identiquement distribuées selon une loi exponentielle de paramètre ν .*

Démonstration. Commençons par le sens nécessaire de cet énoncé.

Vu le lien direct entre les variables T_n et W_n et puisqu'il existe une relation simple entre les variables T_n et le processus de comptage $\{N(t), t \geq 0\}$ qui est mieux connu, cherchons alors à déterminer la fonction de densité jointe des temps d'occurrence. Pour cela, considérons $\forall n \in \mathbb{N}_0$ des instants $s_1 < t_1 < s_2 < t_2 < \dots < s_n < t_n$.

$$\begin{aligned} & \mathbb{P}[s_1 < T_1 < t_1, s_2 < T_2 < t_2, \dots, s_n < T_n < t_n] \\ &= \mathbb{P}[N(s_1) = 0, N(t_1) - N(s_1) = 1, N(s_2) - N(t_1) = 0, N(t_2) - N(s_2) = 1, \dots, \\ & \quad N(t_{n-1}) - N(s_{n-1}) = 1, N(s_n) - N(t_{n-1}) = 0, N(t_n) - N(s_n) \geq 1] \\ &= \mathbb{P}[N(s_1) = 0] \mathbb{P}[N(t_1) - N(s_1) = 1] \mathbb{P}[N(s_2) - N(t_1) = 0] \mathbb{P}[N(t_2) - N(s_2) = 1] \dots \\ & \quad \mathbb{P}[N(t_{n-1}) - N(s_{n-1}) = 1] \mathbb{P}[N(s_n) - N(t_{n-1}) = 0] \mathbb{P}[N(t_n) - N(s_n) \geq 1] \end{aligned}$$

par indépendance des accroissements. De là, sachant que $N(t) \sim \mathcal{P}(\nu t)$ et que les accroissements sont stationnaires, il vient

$$\begin{aligned} & \mathbb{P}[s_1 < T_1 < t_1, s_2 < T_2 < t_2, \dots, s_n < T_n < t_n] \\ &= e^{-\nu s_1} \nu (t_1 - s_1) e^{-\nu(t_1 - s_1)} e^{-\nu(s_2 - t_1)} \nu (t_2 - s_2) e^{-\nu(t_2 - s_2)} \dots \nu (t_{n-1} - s_{n-1}) \\ & \quad e^{-\nu(t_{n-1} - s_{n-1})} e^{-\nu(s_n - t_{n-1})} (1 - e^{-\nu(t_n - s_n)}) \\ &= \nu^{n-1} e^{-\nu s_n} (1 - e^{-\nu(t_n - s_n)}) (t_1 - s_1) \dots (t_{n-1} - s_{n-1}) \end{aligned}$$

car toutes les exponentielles se simplifient entre elles sauf $e^{-\nu s_n}$. Ainsi,

$$\mathbb{P}[s_1 < T_1 < t_1, s_2 < T_2 < t_2, \dots, s_n < T_n < t_n] = \nu^{n-1} (e^{-\nu s_n} - e^{-\nu t_n}) (t_1 - s_1) \dots (t_{n-1} - s_{n-1}). \quad (2.10)$$

Il est aisé de réécrire (2.10) sous la forme d'une intégrale multiple de la densité sur le domaine fixé :

$$\int_{s_1}^{t_1} \int_{s_2}^{t_2} \dots \int_{s_{n-1}}^{t_{n-1}} \int_{s_n}^{t_n} e^{-\nu y_n} \nu^n dy_1 \dots dy_{n-1} dy_n.$$

D'où,

$$f_{(T_1, \dots, T_n)}(y_1, \dots, y_n) = \nu^n e^{-\nu y_n} \chi_{\{(y_1, y_2, \dots, y_n) \in \mathbb{R}^n : y_1 < y_2 < \dots < y_n\}}.$$

Pour revenir aux temps d'inter-arrivées, le changement de variables

$$\begin{cases} T_1 = W_1 \\ T_2 = W_1 + W_2 \\ \vdots \\ T_n = W_1 + W_2 + \dots + W_n \end{cases}$$

est utilisé. On vérifie aisément que le jacobien vaut 1, et par la formule de changement de variables, on obtient

$$f_{(W_1, W_2, \dots, W_n)}(w_1, w_2, \dots, w_n) = \nu^n e^{-\nu(w_1 + w_2 + \dots + w_n)} \chi_{\{(w_1, w_1 + w_2, \dots, w_1 + \dots + w_n) \in \mathbb{R}^n : w_1 < w_1 + w_2 < \dots < w_1 + \dots + w_n\}}.$$

Comme la variable W_n est à valeurs strictement positives $\forall n > 0$, on a donc

$$f_{(W_1, W_2, \dots, W_n)}(w_1, w_2, \dots, w_n) = \nu^n e^{-\nu w_1} e^{-\nu w_2} \dots e^{-\nu w_n}. \quad (2.11)$$

En prenant $n = 1$, on voit que $W_1 \sim \text{Exp}(\nu)$. Ensuite pour $n = 2$, calculons la distribution de W_2 conditionnelle à $W_1 = w_1$,

$$f_{W_2|W_1=w_1}(w_2) = \frac{f_{W_1, W_2}(w_1, w_2)}{f_{W_1}(w_1)} = \frac{\nu^2 e^{-\nu w_1} e^{-\nu w_2}}{\nu e^{-\nu w_1}} = \nu e^{-\nu w_2}.$$

Cela montre que W_2 est indépendante de W_1 et que $W_2 \sim \text{Exp}(\nu)$.

Pour tout $k \geq 2$, on montre que

$$f_{W_k|W_{k-1}=w_{k-1}, \dots, W_1=w_1}(w_k) = \nu e^{-\nu w_k}.$$

En effet,

$$\begin{aligned} f_{W_k|W_{k-1}=w_{k-1}, \dots, W_1=w_1}(w_k) &= \frac{f_{(W_1, W_2, \dots, W_k)}(w_1, w_2, \dots, w_k)}{f_{(W_1, \dots, W_{k-1})}(w_1, \dots, w_{k-1})} \\ &= \frac{\nu^k e^{-\nu w_1} e^{-\nu w_2} \dots e^{-\nu w_k}}{\nu^{k-1} e^{-\nu w_1} e^{-\nu w_2} \dots e^{-\nu w_{k-1}}} \quad \text{par (2.11)} \\ &= \nu e^{-\nu w_k}. \end{aligned}$$

Comme le résultat est indépendant de w_1, w_2, \dots, w_{k-1} , on a obtenu la densité marginale de W_k , $\forall k \geq 2$. Elle correspond bien à la densité d'une exponentielle de paramètre ν .

Pour conclure, on remarque que

$$f_{(W_1, W_2, \dots, W_n)}(w_1, w_2, \dots, w_n) = f_{W_1}(w_1) f_{W_2}(w_2) \dots f_{W_n}(w_n)$$

ce qui prouve l'indépendance des variables W_1, \dots, W_n .

Pour ce qui est du sens suffisant, commençons par montrer que $N(t) \sim \mathcal{P}(\nu t)$ pour tout $t > 0$. Soit $n \in \mathbb{N}_0$. Le lien entre les variables $N(t)$ et T_n donne que

$$\begin{aligned} \mathbb{P}[N(t) = n] &= \mathbb{P}[N(t) \geq n] - \mathbb{P}[N(t) \geq n + 1] \\ &= \mathbb{P}[T_n \leq t] - \mathbb{P}[T_{n+1} \leq t] \end{aligned}$$

où $T_n = W_1 + \dots + W_n \sim \text{gamma}(n, \nu)$ vu le corollaire 1.18. Cela implique que

$$\begin{aligned}
 \mathbb{P}[N(t) = n] &= \int_0^t \frac{\nu^n}{(n-1)!} s^{n-1} e^{-\nu s} ds - \int_0^t \frac{\nu^{n+1}}{n!} s^n e^{-\nu s} ds \\
 &= \int_0^t \left(\frac{\nu^n}{(n-1)!} s^{n-1} e^{-\nu s} - \frac{\nu^{n+1}}{n!} s^n e^{-\nu s} \right) ds \\
 &= \int_0^t D_s \left(\frac{\nu^n}{n!} s^n e^{-\nu s} \right) ds \\
 &= e^{-\nu t} \frac{(\nu t)^n}{n!}.
 \end{aligned} \tag{2.12}$$

En se servant de cela,

$$\begin{aligned}
 \mathbb{P}[N(t) = 0] &= 1 - \sum_{n=1}^{+\infty} e^{-\nu t} \frac{(\nu t)^n}{n!} \\
 &= 1 - \left(e^{-\nu t} \sum_{n=0}^{+\infty} \frac{(\nu t)^n}{n!} - e^{-\nu t} \right) \\
 &= e^{-\nu t}.
 \end{aligned}$$

Pour la suite, la notation suivante sera nécessaire : W_i^s est le temps d'attente du i -ième événement se produisant après s . Par la propriété d'oubli de l'exponentielle, W_1^s est encore une variable aléatoire exponentielle de paramètre ν , les autres aussi car ce sont simplement des inter-arrivées.

Montrons que les accroissements sont stationnaires. Soient $0 \leq s < t$ et $n \in \mathbb{N}$.

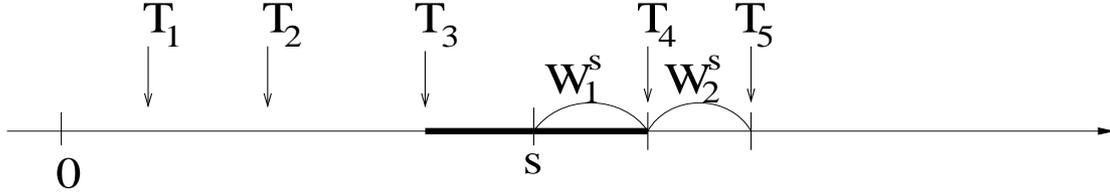


FIG. 2.3 – Nouvelle notation

$$\begin{aligned}
 \mathbb{P}[N(t) - N(s) = n] &= \mathbb{P}[N(t) - N(s) \geq n] - \mathbb{P}[N(t) - N(s) \geq n + 1] \\
 &= \mathbb{P} \left[\sum_{i=1}^n W_i^s \leq t - s \right] - \mathbb{P} \left[\sum_{i=1}^{n+1} W_i^s \leq t - s \right].
 \end{aligned}$$

Comme les sommes sont deux variables aléatoires de loi $\text{gamma}(n, \nu)$ et $\text{gamma}(n + 1, \nu)$ respectivement, le même développement que précédemment, permet d'obtenir

$$\mathbb{P}[N(t) - N(s) = n] = e^{-\nu(t-s)} \frac{(\nu(t-s))^n}{n!}. \tag{2.13}$$

Les équations (2.12) et (2.13) impliquent que $\mathbb{P}[N(t) - N(s) = n] = \mathbb{P}[N(t-s) = n]$. Ainsi, $N(t) - N(s) \sim N(t-s)$, $\forall 0 \leq s < t$.

Passons alors à l'indépendance des accroissements. Soient $0 \leq t_1 < t_2 < t_3$ et $n, m \in \mathbb{N}$.

$$\begin{aligned} & \mathbb{P}[N(t_3) - N(t_2) = n \text{ et } N(t_2) - N(t_1) = m] \\ &= \mathbb{P}[\{\{N(t_3) - N(t_2) \geq n\} \setminus \{N(t_3) - N(t_2) \geq n + 1\}\} \text{ et} \\ & \quad (\{N(t_2) - N(t_1) \geq m\} \setminus \{N(t_2) - N(t_1) \geq m + 1\})\}] \\ &= \mathbb{P} \left[\left(\left\{ \sum_{i=1}^n W_i^{t_2} \leq t_3 - t_2 \right\} \setminus \left\{ \sum_{i=1}^{n+1} W_i^{t_2} \leq t_3 - t_2 \right\} \right) \text{ et} \right. \\ & \quad \left. \left(\left\{ \sum_{i=1}^m W_i^{t_1} \leq t_2 - t_1 \right\} \setminus \left\{ \sum_{i=1}^{m+1} W_i^{t_1} \leq t_2 - t_1 \right\} \right) \right]. \end{aligned}$$

Comme les sommes de temps d'inter-arrivées portent sur des inter-arrivées différentes, elles sont indépendantes et la factorisation des probabilités est possible, ce qui donne

$$\begin{aligned} & \mathbb{P}[N(t_3) - N(t_2) = n \text{ et } N(t_2) - N(t_1) = m] \\ &= \mathbb{P} \left[\left\{ \sum_{i=1}^n W_i^{t_2} \leq t_3 - t_2 \right\} \setminus \left\{ \sum_{i=1}^{n+1} W_i^{t_2} \leq t_3 - t_2 \right\} \right] \\ & \quad \mathbb{P} \left[\left\{ \sum_{i=1}^m W_i^{t_1} \leq t_2 - t_1 \right\} \setminus \left\{ \sum_{i=1}^{m+1} W_i^{t_1} \leq t_2 - t_1 \right\} \right] \\ &= \mathbb{P}[\{N(t_3) - N(t_2) \geq n\} \setminus \{N(t_3) - N(t_2) \geq n + 1\}] \\ & \quad \mathbb{P}[\{N(t_2) - N(t_1) \geq m\} \setminus \{N(t_2) - N(t_1) \geq m + 1\}] \\ &= \mathbb{P}[N(t_3) - N(t_2) = n] \mathbb{P}[N(t_2) - N(t_1) = m]. \end{aligned}$$

Ce qui prouve bien l'indépendance des accroissements. ■

Par les propriétés de l'exponentielle, $E[W_n] = \frac{1}{\nu} \forall n \in \mathbb{N}_0$. Cela signifie que le temps moyen entre deux occurrences du processus vaut $\frac{1}{\nu}$. Autrement dit, il se produit ν événements dans un intervalle de temps unitaire. On retrouve bien la même interprétation du paramètre ν que celle donnée au début du chapitre.

Étant donné la proposition 2.4, pour générer les instants d'occurrence d'un processus de Poisson d'intensité ν , on génère des variables indépendantes et identiquement distribuées selon une loi exponentielle de paramètre ν . Pour $n \in \mathbb{N}_0$, notons w_n la réalisation de W_n qui représente le temps d'attente entre les $(n - 1)^e$ et n^e événements. Les temps d'occurrence t_n s'obtiennent alors par la formule

$$\begin{cases} t_0 = 0 \\ t_n = \sum_{i=1}^n w_i. \end{cases}$$

C'est cette méthode qui a permis de générer les trajectoires représentées aux figures 2.1 et 2.2 de la page 20. On a en effet généré deux groupes de variables indépendantes et exponentiellement distribuées, un avec une moyenne $\frac{1}{2}$ et l'autre avec $\frac{1}{10}$ pour moyenne. A partir de ces échantillons, on a calculé les différents temps d'occurrence dans les deux cas, ce qui permet d'obtenir la valeur de $N(t)$ en tout temps $t \in \mathbb{R}^+$.

A partir du lien entre les variables aléatoires W_n et T_n et de la proposition précédente, on peut déduire la distribution de T_n pour tout $n \geq 1$. En effet, on sait que $\forall n \geq 1$, $T_n = W_1 + W_2 + \dots + W_n$ et $W_n \stackrel{\text{iid}}{\sim} \text{Exp}(\nu)$. Par la propriété d'additivité des exponentielles (cfr. corollaire 1.18), il vient que

$$T_n \sim \text{gamma}(n, \nu) \quad \forall n \geq 1. \quad (2.14)$$

C'est-à-dire que $\forall n \in \mathbb{N}$, la fonction de densité de T_n est donnée par

$$f_{T_n}(t) = \frac{\nu^n}{\Gamma(n)} t^{n-1} e^{-\nu t} \chi_{[0, +\infty[}(t). \quad (2.15)$$

Pour terminer cette section, voici deux résultats sur les temps d'occurrence d'un processus de Poisson de base. Ce sont des résultats conditionnels au nombre d'occurrences du processus. Le premier répondra aux deux questions suivantes :

1. Sachant qu'au temps fixé T , n événements se sont déjà produits, quelle est la distribution du temps d'occurrence du n -ième ?
2. Sachant qu'au temps fixé T , n événements se sont déjà produits, quelle est la distribution du temps d'occurrence du $(n + 1)$ -ième ?

Le second sera utile pour tester si un ensemble de données peut être considéré comme une trajectoire d'un tel processus.

Proposition 2.5. *Soit $\{N(t), t \geq 0\}$ un processus de Poisson d'intensité $\nu > 0$. Sachant que n événements se sont produits dans l'intervalle de temps $[0, T]$, T fixé, la fonction de densité du temps d'occurrence du n -ième est donnée par*

$$f_{T_n|N(T)=n}(t) = \frac{nt^{n-1}}{T^n} \chi_{[0, T]}(t), \quad (2.16)$$

cela indépendamment de l'intensité du processus. Dans les mêmes conditions, la densité du temps d'occurrence du $(n + 1)$ -ième événement est donnée par

$$f_{T_{n+1}|N(t)=n}(t) = e^{-\nu(t-T)} \nu \chi_{]T, +\infty[}(t). \quad (2.17)$$

Démonstration. Pour trouver l'expression de la densité, cherchons celle de la fonction de répartition $F_{T_n|N(T)=n}(t) = \mathbb{P}[T_n \leq t | N(T) = n]$. Tout d'abord, il faut remarquer que si $t \geq T$, alors la fonction de répartition vaut 1 puisqu'on a un événement certain. Ensuite, pour $t < T$, il vient

$$\begin{aligned} F_{T_n|N(T)=n}(t) &= \mathbb{P}[N(t) \geq n | N(T) = n] \\ &= \mathbb{P}[N(t) > n | N(T) = n] + \mathbb{P}[N(t) = n | N(T) = n] \end{aligned}$$

mais comme $t < T$, la première probabilité est nulle. Dès lors, il vient

$$F_{T_n|N(T)=n}(t) = \frac{\mathbb{P}[N(t) = n \text{ et } N(T) - N(t) = 0]}{\mathbb{P}[N(T) = n]}$$

l'indépendance des accroissements, leur stationnarité et la distribution de $N(t)$ impliquent

$$\begin{aligned} F_{T_n|N(T)=n}(t) &= \frac{e^{-\nu t}(\nu t)^n}{n!} e^{-\nu(T-t)} \frac{n!}{e^{-\nu T}(\nu T)^n} \\ &= \left(\frac{t}{T}\right)^n. \end{aligned}$$

En dérivant, on obtient alors la fonction de densité

$$f_{T_n|N(T)=n}(t) = \begin{cases} \frac{n}{T^n} t^{n-1} & \text{si } t < T \\ 0 & \text{sinon.} \end{cases}$$

Passons à la seconde partie de l'énoncé et procédons de la même façon. La fonction de répartition $F_{T_{n+1}|N(T)=n}(t) = \mathbb{P}[T_{n+1} \leq t | N(T) = n]$ vaut 0 lorsque $t \leq T$. Dans l'autre cas, cela vaut

$$\begin{aligned} F_{T_{n+1}|N(T)=n}(t) &= \mathbb{P}[N(t) \geq n+1 | N(T) = n] \\ &= \frac{\mathbb{P}[N(T) = n \text{ et } N(t) - N(T) \geq 1]}{\mathbb{P}[N(T) = n]} \end{aligned}$$

ce qui donne, après utilisation de l'indépendance des accroissements, de leur stationnarité et de la distribution de $N(t)$

$$\begin{aligned} F_{T_{n+1}|N(T)=n}(t) &= \frac{e^{-\nu T}(\nu T)^n}{n!} \left(1 - e^{-\nu(t-T)}\right) \frac{n!}{e^{-\nu T}(\nu T)^n} \\ &= 1 - e^{-\nu(t-T)}. \end{aligned}$$

L'expression de la fonction de densité est alors

$$f_{T_{n+1}|N(T)=n}(t) = e^{-\nu(t-T)} \nu, \quad \text{si } t > T.$$

■

Le premier résultat est particulièrement intéressant à interpréter. Considérons le processus d'arrivée des trains sur une voie d'une gare d'une petite ville (qui n'est reliée qu'à une unique grande ville) et supposons qu'il soit un processus de Poisson. La proposition précédente implique que connaissant le nombre de trains qui sont passés depuis le matin, le temps d'arrivée du dernier train à la même distribution quel que soit le taux d'arrivée des trains. Le second résultat donne la distribution du temps d'arrivée du prochain train en fonction de l'intensité des arrivées. Elle est valable quel que soit le nombre de trains déjà passés. Si $T = 0$, cette expression correspond bien à la distribution du premier temps d'inter-arrivée qui est aussi le premier temps d'occurrence.

Avant la dernière proposition, le lemme suivant est nécessaire.

Lemme 2.6. *La densité conjointe des statistiques d'ordre de n variables aléatoires U_1, \dots, U_n indépendantes et distribuées uniformément sur $[a, b]$ est donnée par*

$$f_{(U_{(1)}, \dots, U_{(n)})}(u_1, \dots, u_n) = \begin{cases} \frac{n!}{(b-a)^n} & \text{si } a \leq u_1 \leq u_2 \leq \dots \leq u_n \leq b \\ 0 & \text{sinon} \end{cases}$$

Démonstration. Par définition des variables uniformes, la densité des variables U_i est la fonction $f(x) = \frac{1}{b-a}\chi_{[a,b]}(x)$ et leur fonction de répartition est la fonction

$$F(x) = \frac{x-a}{b-a}\chi_{[a,b]}(x) + \chi_{]b,+\infty[}(x) \text{ pour tout } i = 1, 2, \dots, n.$$

Soient des réels u_i et h_i tels que $a \leq u_1 < u_1 + h_1 < u_2 < \dots < u_i < u_i + h_i < u_{i+1} < \dots < u_n < u_n + h_n < u_{n+1} = b$.

$$\mathbb{P}[u_i < U_{(i)} \leq u_i + h_i, i = 1, 2, \dots, n] = \mathbb{P} \left[\bigcup_{\mu \text{ permutation de } \{1, \dots, n\}} (u_i < U_{\mu_i} \leq u_i + h_i, i = 1, 2, \dots, n) \right].$$

Vu le caractère disjoint des intervalles $[u_i, u_i + h_i]$, ces événements sont incompatibles, ce qui donne

$$\mathbb{P}[u_i < U_{(i)} \leq u_i + h_i, i = 1, 2, \dots, n] = \sum_{\mu \text{ permutation de } \{1, \dots, n\}} \mathbb{P}[u_i < U_{\mu_i} \leq u_i + h_i, i = 1, 2, \dots, n].$$

L'indépendance des variables U_i implique alors

$$\begin{aligned} \mathbb{P}[u_i < U_{(i)} \leq u_i + h_i, i = 1, 2, \dots, n] &= \sum_{\mu \text{ permutation de } \{1, \dots, n\}} \prod_{i=1}^n \mathbb{P}[u_i < U_{\mu_i} \leq u_i + h_i] \\ &= \sum_{\mu \text{ permutation de } \{1, \dots, n\}} \prod_{i=1}^n \frac{h_i}{b-a}. \end{aligned}$$

Comme il y a $n!$ permutations de $1, 2, \dots, n$, cela devient

$$\mathbb{P}[u_i < U_{(i)} \leq u_i + h_i, i = 1, 2, \dots, n] = n! \frac{h_1 h_2 \dots h_n}{(b-a)^n}. \quad (2.18)$$

Par définition de la densité, on a

$$f(x) = D_x \mathbb{P}[X \leq x] = \lim_{h \rightarrow 0} \frac{\mathbb{P}[X \leq x+h] - \mathbb{P}[X \leq x]}{h} = \lim_{h \rightarrow 0} \frac{\mathbb{P}[x < X \leq x+h]}{h}.$$

En utilisant (2.18), il vient

$$f_{(U_{(1)}, \dots, U_{(n)})}(u_1, \dots, u_n) = \lim_{h_i \rightarrow 0 \forall i} \frac{\mathbb{P}[u_i < U_{(i)} \leq u_i + h_i, i = 1, 2, \dots, n]}{h_1 h_2 \dots h_n} = \frac{n!}{(b-a)^n},$$

pour tous $a \leq u_1 \leq \dots \leq u_n \leq b$. Et si les u_i ne sont pas dans l'intervalle $[a, b]$ ou ne sont pas croissants, cette expression est nulle puisqu'il s'agit alors d'un événement de probabilité nulle. ■

Proposition 2.7. Soit $\{N(t), t \geq 0\}$ un processus de Poisson d'intensité $\nu > 0$. La densité conditionnelle de (T_1, T_2, \dots, T_n) sachant que $N(t)=n$ est donnée par

$$f_{T_1, T_2, \dots, T_n | N(t)=n}(t_1, t_2, \dots, t_n) = \begin{cases} \frac{n!}{t^n} & \text{si } 0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq t \\ 0 & \text{sinon} \end{cases}$$

qui est la densité des statistiques d'ordre d'un ensemble de n variables indépendantes et identiquement distribuées selon une loi uniforme sur $[0, t]$.

Cette proposition stipule que, sachant que n événements se sont produits dans l'intervalle $[0, t]$, les temps d'occurrence de ces n événements, considérés comme des statistiques d'ordre car $T_1 < T_2 < \dots < T_n$, sont indépendamment et uniformément distribués dans l'intervalle $[0, t]$. Pour $n = 1$, cela signifie que chaque intervalle de $[0, t]$ a la même probabilité de contenir l'unique événement qui a eu lieu entre 0 et t .

Démonstration. Soient des réels t_i et h_i tels que $0 \leq t_1 < t_1 + h_1 < t_2 < \dots < t_i < t_i + h_i < t_{i+1} < \dots < t_{n+1} = t$. Considérons alors la probabilité suivante

$$\mathbb{P}[t_i < T_i \leq t_i + h_i, i = 1, 2, \dots, n | N(t) = n].$$

Vu le choix des h_i , elle est égale à

$$\frac{\mathbb{P}[\text{"un et un seul événement dans l'intervalle } [t_i, t_i + h_i], i = 1, 2, \dots, n \text{ et aucun ailleurs"}]}{\mathbb{P}[N(t) = n]}.$$

En effet, on a $0 \leq t_1 < t_1 + h_1 < t_2 < t_2 + h_2 < t_3 < \dots < t_{n-1} + h_{n-1} < t_n < t_n + h_n < t$. Cette probabilité se calcule alors comme suit :

$$\begin{aligned} & \mathbb{P}[t_i < T_i \leq t_i + h_i, i = 1, 2, \dots, n | N(t) = n] \\ &= \frac{1}{\mathbb{P}[N(t) = n]} \mathbb{P}[N(t_1) = 0, N(t_1 + h_1) - N(t_1) = 1, N(t_2) - N(t_1 + h_1) = 0, \\ & \quad N(t_2 + h_2) - N(t_2) = 1, \dots, N(t_n + h_n) - N(t_n) = 1, N(t_{n+1}) - N(t_n + h_n) = 0]. \end{aligned}$$

On utilise alors l'indépendance et la stationnarité des accroissements ainsi que la distribution de $N(t)$ pour obtenir que

$$\begin{aligned} & \mathbb{P}[t_i < T_i \leq t_i + h_i, i = 1, 2, \dots, n | N(t) = n] \\ &= \frac{n!}{e^{-\nu t} (\nu t)^n} [e^{-\nu t_1} e^{-\nu h_1} (\nu h_1) e^{-\nu(t_2 - t_1 - h_1)} e^{-\nu h_2} (\nu h_2) e^{-\nu(t_3 - t_2 - h_2)} \dots \\ & \quad e^{-\nu(t_n - t_{n-1} - h_{n-1})} e^{-\nu h_n} (\nu h_n) e^{-\nu(t - t_n - h_n)}] \\ &= \frac{n!}{\nu^n t^n} \nu^n h_1 h_2 \dots h_n \end{aligned}$$

car les exponentielles se simplifient. Comme

$$\frac{\mathbb{P}[t_i < T_i \leq t_i + h_i, i = 1, 2, \dots, n | N(t) = n]}{h_1 h_2 \dots h_n} = \frac{n!}{t^n}$$

où le second membre est indépendant des h_i , on obtient

$$f_{(T_1, \dots, T_n | N(t) = n)}(t_1, \dots, t_n) = \frac{n!}{t^n} \quad \text{si } 0 \leq t_1 \leq \dots \leq t_n \leq t.$$

Sinon, $\exists i < j$ tels que $t_i > t_j$ et on ne saurait pas avoir $T_i \leq t_i$ et $T_j \leq t_j$ en même temps. Le lemme précédent permet alors de conclure. ■

Si $n = 1$ dans la relation (2.16), la densité obtenue est uniforme sur $[0, T]$, ce qui est annoncé par le résultat précédent.

2.4 Distributions tronquées

En pratique, cela semble parfois difficile de considérer un processus dans son ensemble. C'est pourquoi les distributions tronquées peuvent être utiles. La troncature d'un processus de comptage peut porter sur le temps d'observation ou sur le nombre d'événements observés.

Dans le premier cas, chaque temps d'occurrence doit se situer dans un intervalle fini $[0, T]$ où T est le temps limite d'observation. Si l'astérisque représente la variable aléatoire tronquée, voici ce que nous obtenons pour la fonction de répartition :

$$F_{T_n^*}(t) = \mathbb{P}[T_n \leq t | 0 \leq T_n \leq T] = \frac{\mathbb{P}[0 \leq T_n \leq \min(t, T)]}{\mathbb{P}[0 \leq T_n \leq T]} = \begin{cases} 0 & \text{si } t < 0 \\ \frac{\mathbb{P}[0 \leq T_n \leq t]}{\mathbb{P}[0 \leq T_n \leq T]} & \text{si } 0 \leq t \leq T \\ 1 & \text{sinon.} \end{cases}$$

Cela implique que la densité de T_n tronquée sur l'intervalle $[0, T]$ est donnée par

$$f_{T_n^*}(t) = \frac{f_{T_n}(t)}{F_{T_n}(T)} \chi_{[0, T]}(t) \quad (2.19)$$

où le dénominateur est une constante par rapport à t , appelée *constante de normalisation*.

Si maintenant, le nombre d'événements observés est fixé à N , la distribution de la variable $N(t)$ n'est plus la même puisque les valeurs plus grandes que N ne sont plus permises. Cette distribution devient alors

$$\mathbb{P}[N^*(t) = n] = \mathbb{P}[N(t) = n | 0 \leq N(t) \leq N] = \begin{cases} \frac{\mathbb{P}[N(t)=n]}{\mathbb{P}[0 \leq N(t) \leq N]} & \text{si } n \in [0, N] \\ 0 & \text{sinon.} \end{cases} \quad (2.20)$$

Comme dans le cas précédent, le dénominateur ne dépend pas de la valeur de n , c'est à nouveau une *constante de normalisation*.

L'expression (2.20) peut être réécrite de façon plus explicite : supposons que l'intensité du processus soit $\nu \in \mathbb{R}^+$, $\forall n \in [0, N]$,

$$\begin{aligned} \mathbb{P}[N^*(t) = n] &= \frac{e^{-\nu t} (\nu t)^n}{n!} \frac{1}{e^{-\nu t} \sum_{i=0}^N \frac{(\nu t)^i}{i!}} \\ &= \frac{(\nu t)^n}{n! \sum_{i=0}^N \frac{(\nu t)^i}{i!}}. \end{aligned} \quad (2.21)$$

Chapitre 3

Processus de Poisson non homogène

La différence entre un processus de Poisson de base et un processus de Poisson non homogène réside dans le fait que les accroissements ne sont plus stationnaires. Cela est plus réaliste car, en général, le taux d'occurrence d'événements dépend du temps. Par exemple, pour un commerce, le taux d'entrée des clients varie au cours du temps. On imagine bien que pour un restaurant, ce taux sera élevé entre 11h et 13h puis faible dans l'après-midi (voir nul si le restaurant ferme durant cette période) puis à nouveau élevé entre 18h et 20h, après quoi il diminuera graduellement pour retomber à zéro.

Il est courant de parler d'un processus de Poisson non homogène en terme de processus de Poisson non-stationnaire par opposition au processus de Poisson stationnaire qui est le processus de Poisson de base étudié dans la section précédente pour lequel les accroissements étaient stationnaires.

Définition 3.1. *Un processus de comptage $\{N(t), t \geq 0\}$ est un processus de Poisson non homogène s'il satisfait les conditions suivantes :*

(NH1) $N(0)=0$;

(NH2) *Le processus est à accroissements indépendants ;*

(NH3) $\forall t \geq 0, \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) \geq 2]}{\mathbb{P}[N(t+h) - N(t) = 1]} = 0$;

(NH4) *Il existe une fonction à valeurs positives $\nu(t)$ telle que*

$$\lim_{h \rightarrow 0^+} \frac{1 - \mathbb{P}[N(t+h) - N(t) = 0]}{h} = \nu(t), \forall t \geq 0.$$

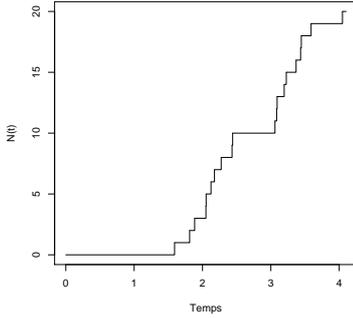
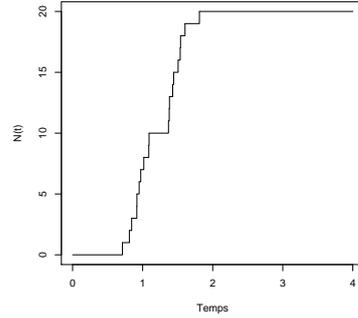
La fonction $\nu(t)$ dont il est question dans la définition est la *fonction d'intensité* du processus. Lorsqu'elle est constante, il s'agit simplement d'un processus de base. Dans ce cas, l'espérance satisfait l'équation $E[N(t)] = \nu t = \int_0^t \nu dt'$. Dans le cas d'un processus de Poisson non homogène, cette équation n'est plus valable. On introduit alors la *fonction moyenne* $m(t)$ définie par

$$m(t) = \int_0^t \nu(t') dt'.$$

Très souvent, et ce sera le cas ici, la fonction $m(t)$ est supposée continue et différentiable. Étant donné que la fonction d'intensité $\nu(t)$ est positive, la fonction moyenne est positive et non-décroissante.

La proposition 3.1 mettra en évidence l'analogie avec le cas de base, à savoir que cette fonction moyenne correspond au nombre attendu d'occurrences dans l'intervalle $[0, t]$.

Les figures 3.1 et 3.2 donnent la représentation d'une trajectoire de deux processus de Poisson non homogènes dont les fonctions d'intensité sont respectivement les fonctions $\nu(t) = 3t$ et $\nu(t) = 15t$. Dans les deux cas, les événements se succèdent plus vite lorsque le temps augmente.


 FIG. 3.1 – Trajectoire pour $\nu(t) = 3t$

 FIG. 3.2 – Trajectoire pour $\nu(t) = 15t$

3.1 Distribution des marges finies du processus

Proposition 3.1. Soit $\{N(t), t \geq 0\}$ un processus de Poisson non homogène de fonction d'intensité $\nu(t)$. Alors la fonction génératrice des probabilités de $N(t)$ est donnée, $\forall |z| \leq 1$, par $G(z, t) = e^{m(t)(z-1)}$ avec $m(t) = \int_0^t \nu(t') dt'$.

Démonstration.

$$\begin{aligned}
 G(z, t+h) &= E \left[z^{N(t+h)} \right] = E \left[z^{N(t+h)-N(t)+N(t)} \right] = E \left[z^{N(t+h)-N(t)} z^{N(t)} \right] \\
 &= E \left[z^{N(t+h)-N(t)} \right] E \left[z^{N(t)} \right] \quad \text{vu l'hypothèse (NH2)} \\
 &= G(z, t) E \left[z^{N(t+h)-N(t)} \right]. \tag{3.1}
 \end{aligned}$$

En utilisant la définition de l'espérance, on a $\frac{E[z^{N(t+h)-N(t)}]-1}{h} =$

$$\frac{\mathbb{P}[N(t+h) - N(t) = 0] - 1}{h} + z \frac{\mathbb{P}[N(t+h) - N(t) = 1]}{h} + \frac{\sum_{n=2}^{+\infty} z^n \mathbb{P}[N(t+h) - N(t) = n]}{h},$$

où la série est absolument convergente si $z \in [-1, 1]$. Par la même justification que dans la preuve de la proposition 2.1, on pourra échanger la limite sur $h \rightarrow 0^+$ et la série. En passant à la limite pour $h \rightarrow 0^+$ et en exploitant la condition (NH4), on obtient

$$\begin{aligned}
 -\nu(t) + z \lim_{h \rightarrow 0^+} \frac{1 - \mathbb{P}[N(t+h) - N(t) = 0] - \mathbb{P}[N(t+h) - N(t) \geq 2]}{h} + \\
 \sum_{n=2}^{+\infty} z^n \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) = n]}{h}. \tag{3.2}
 \end{aligned}$$

Cherchons l'expression des deux limites $\lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) = n]}{h}$ et

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) \geq 2]}{h}.$$

$$\frac{\mathbb{P}[N(t+h) - N(t) \geq 2]}{h} = \frac{\mathbb{P}[N(t+h) - N(t) \geq 2] \mathbb{P}[N(t+h) - N(t) \geq 1]}{\mathbb{P}[N(t+h) - N(t) \geq 1] h}.$$

Ainsi, on a

$$0 \leq \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) \geq 2]}{h} \leq \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) \geq 2] \mathbb{P}[N(t+h) - N(t) \geq 1]}{\mathbb{P}[N(t+h) - N(t) \geq 1] h}.$$

Par la condition (NH3), la première fraction du majorant tend vers 0, et par la condition (NH4), $\frac{\mathbb{P}[N(t+h) - N(t) \geq 1]}{h} = \frac{1 - \mathbb{P}[N(t+h) - N(t) = 0]}{h}$ tend vers $\nu(t)$. Le théorème de l'étau permet de conclure que

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) \geq 2]}{h} = 0.$$

De plus, $\mathbb{P}[N(t+h) - N(t) \geq 2] = \sum_{l=2}^{+\infty} \mathbb{P}[N(t+h) - N(t) = l]$. Cela entraîne que, $\forall n \geq 2$,

$$0 \leq \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) = n]}{h} \leq \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) \geq 2]}{h}.$$

Vu le calcul précédent, la dernière limite est nulle. On conclut de nouveau par le théorème de l'étau que

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) = n]}{h} = 0, \quad \forall n \geq 2.$$

En revenant à l'expression (3.2), et en utilisant (NH4), on obtient

$$\lim_{h \rightarrow 0^+} \frac{E[z^{N(t+h) - N(t)}] - 1}{h} = -\nu(t) + z\nu(t) = \nu(t)(z - 1).$$

Par conséquent, nous avons

$$\begin{aligned} \lim_{h \rightarrow 0^+} \frac{G(z, t+h) - G(z, t)}{h} &= \frac{\partial}{\partial t} G(z, t) \\ &= \lim_{h \rightarrow 0^+} G(z, t) \frac{E[z^{N(t+h) - N(t)}] - 1}{h} \quad \text{vu (3.1)} \\ &= G(z, t) \nu(t) (z - 1). \end{aligned} \tag{3.3}$$

Pour tout z fixé dans l'intervalle $[-1, 1]$, l'expression (3.3) est une équation différentielle à second membre séparé qui, avec pour condition initiale $G(z, 0) = 1$, admet¹ $G(z, t) = e^{m(t)(z-1)}$ pour solution où $m(t) = \int_0^t \nu(t') dt'$. ■

Comme on l'a vu dans les notions de base, $G(z, t) = e^{m(t)(z-1)}$ est la fonction génératrice d'une variable aléatoire suivant une loi de Poisson de paramètre $m(t)$. Cela veut donc dire que $\forall t > 0$, $N(t) \sim \mathcal{P}(m(t))$ et $E[N(t)] = m(t)$.

¹ $m(t) = \int_0^t \nu(t') dt' = \int \nu(t) dt$ car les primitives sont définies à une constante additive près.

Lorsque la fonction d'intensité $\nu(t)$ n'est pas simplement positive mais strictement positive, la fonction moyenne $m(t)$ est alors strictement croissante. Dans ce cas, on retrouve la condition (C3) de la définition 2.1. En effet,

$$\mathbb{P}[N(t) > 0] = 1 - \mathbb{P}[N(t) = 0] = 1 - e^{-m(t)}$$

qui n'est nulle que quand $m(t)$ est nulle, c-à-d en $t = 0$ car $\nu(t) > 0$ pour tout $t > 0$. Tant que c'est possible, la condition $\nu(t) > 0$ ne sera pas imposée. Cependant, pour les applications pratiques, il viendra un moment où cette condition rendra les choses plus simples.

Avant de continuer dans l'étude de ces processus, regardons quelques exemples de fonctions d'intensité et moyennes.

Fonction d'intensité linéairement croissante

Il faut considérer une fonction d'intensité dépendante du temps (sinon le processus est homogène); prenons le cas très simple d'une fonction linéaire pour commencer. Posons donc

$$\nu(t) = ct, \text{ avec } c > 0, \quad \forall t \geq 0.$$

Dans ce cas,

$$m(t) = \int_0^t \nu(y) dy = c \frac{t^2}{2} \quad \forall t \geq 0.$$

Ces deux fonctions sont représentées aux figures 3.3 et 3.4.

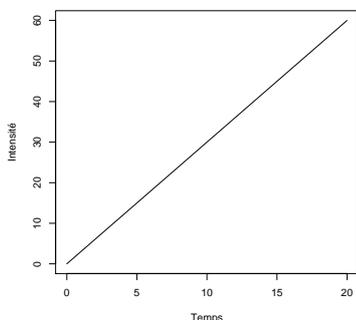


FIG. 3.3 – Intensité linéaire

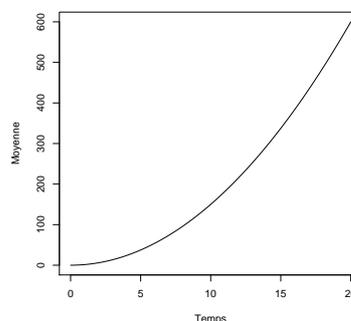


FIG. 3.4 – Moyenne linéaire

Fonction d'intensité parabolique

C'est le cas de la cafétéria du B37, ou de tout autre service de repas, qui connaît un pic d'intensité à midi.

$$\nu(t) = (at^2 + bt + c)\chi_{\{t \geq 0: at^2 + bt + c \geq 0\}}(t), \quad \text{avec } a < 0, \quad \forall t \geq 0.$$

La figure 3.5 montre la forme de cette fonction d'intensité.

Ici, la fonction $m(t)$ n'augmente pas indéfiniment; elle est croissante jusqu'en un temps t' tel que $\nu(t) = 0 \forall t \geq t'$ où elle devient stationnaire comme cela est illustré à la figure 3.6.

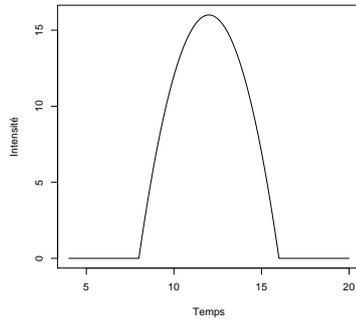


FIG. 3.5 – Intensité parabolique

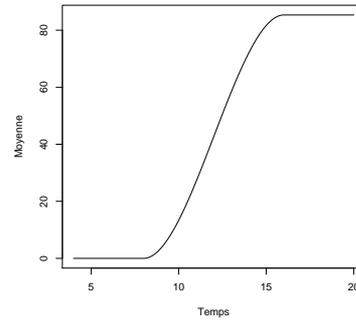


FIG. 3.6 – Moyenne parabolique

Fonction d'intensité exponentiellement décroissante

Un exemple d'utilité de cette fonction est la modélisation de l'intensité des décès causés par les accidents de la route.

$$\nu(t) = \alpha e^{-\beta t}, \quad \text{avec } \alpha, \beta > 0, \quad \forall t \geq 0.$$

La figure 3.7 montre la forme de cette fonction d'intensité.

$$m(t) = \frac{\alpha}{\beta} \left(-e^{-\beta t} + 1 \right) \quad \forall t \geq 0.$$

On a $\lim_{t \rightarrow 0} m(t) = 0$ et $\lim_{t \rightarrow +\infty} m(t) = \frac{\alpha}{\beta}$. La figure 3.8 nous donne la forme de cette fonction.

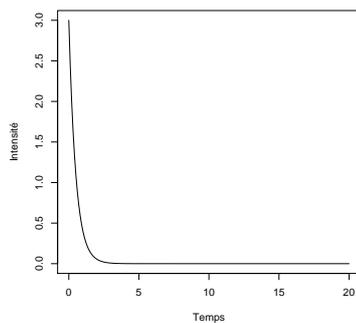


FIG. 3.7 – Intensité exponentiellement décroissante

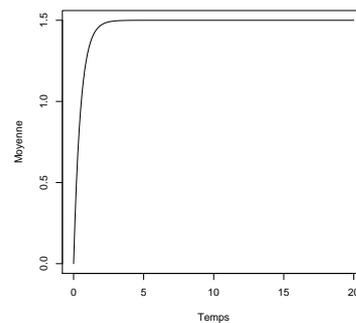


FIG. 3.8 – Moyenne exponentiellement décroissante

Comme nous l'avons déjà signalé, les accroissements ne sont plus stationnaires. De ce fait, la proposition 3.1 ne nous apprend rien sur la distribution de $N(t+s) - N(t)$. C'est pourquoi nous avons besoin de la proposition suivante.

Proposition 3.2. Si $\{N(t), t \geq 0\}$ est un processus de Poisson non homogène de fonction d'intensité $\nu(t)$, alors

$$\mathbb{P}[N(t+s) - N(t) = n] = e^{-(m(t+s)-m(t))} \frac{(m(t+s) - m(t))^n}{n!} \quad \forall t, s > 0, n \in \mathbb{N}$$

où $m(t) = \int_0^t \nu(t') dt'$.

Autrement dit, $N(t+s) - N(t) \sim \mathcal{P}(m(t+s) - m(t))$.

Démonstration. Soit t fixé. Posons $P_n(s) = \mathbb{P}[N(t+s) - N(t) = n]$ et procédons par récurrence sur n .

Cas de base $n = 0$:

$$\begin{aligned} P_0(s+h) &= \mathbb{P}[N(t+s+h) - N(t) = 0] \\ &= \mathbb{P}["0 \text{ événement dans l'intervalle } [t, t+s+h]"'] \\ &= \mathbb{P}["0 \text{ événement dans } [t, t+s]"'] \text{ et } ["0 \text{ événement dans } [t+s, t+s+h]"'] \\ &= \mathbb{P}[N(t+s) - N(t) = 0] \mathbb{P}[N(t+s+h) - N(t+s) = 0] \end{aligned} \quad (3.4)$$

où la dernière ligne est obtenue grâce à l'indépendance des accroissements.

Or, la condition (NH4) implique que

$$\lim_{h \rightarrow 0^+} \left(\nu(t+s) + \frac{o(h)}{h} \right) = \nu(t+s) = \lim_{h \rightarrow 0^+} \frac{1 - \mathbb{P}[N(t+s+h) - N(t+s) = 0]}{h},$$

ce qui donne

$$\mathbb{P}[N(t+s+h) - N(t+s) = 0] = 1 - h\nu(t+s) + o(h).$$

A partir de (3.4), on obtient $P_0(s+h) = P_0(s)(1 - h\nu(t+s) + o(h))$.

Ce qui mène à
$$\frac{P_0(s+h) - P_0(s)}{h} = -\nu(t+s)P_0(s) + \frac{o(h)}{h}.$$

En passant à la limite sur $h \rightarrow 0^+$, on a $D_s P_0(s) = -\nu(t+s)P_0(s)$, ce qui correspond à une équation différentielle à second membre séparé. La condition initiale étant $P_0(0) = 1$, elle admet la solution $P_0(s) = e^{-(m(t+s)-m(t))}$.

Récurrence :

Supposons alors la proposition vraie $\forall k < n$ et démontrons-la pour $k = n$.

$$\begin{aligned} P_n(s+h) &= \mathbb{P}[N(t+s+h) - N(t) = n] \\ &= \mathbb{P}["n \text{ événements dans l'intervalle } [t, t+s+h]"'] \\ &= \sum_{k=0}^n \mathbb{P}["n-k \text{ événements dans } [t, t+s]"'] \text{ et } ["k \text{ événements dans } [t+s, t+s+h]"'] \\ &= \sum_{k=0}^n \mathbb{P}[N(t+s) - N(t) = n-k] \mathbb{P}[N(t+s+h) - N(t+s) = k] \quad \text{par indépendance} \\ &= P_n(s)(1 - h\nu(t+s) + o_0(h)) + P_{n-1}(s)(h\nu(t+s) + o_1(h)) + \sum_{k=2}^n P_{n-k}(s)o_k(h) \end{aligned}$$

où les fonctions $o_k(h)$ sont de l'ordre de h .

La dernière égalité est due au fait que² $\mathbb{P}[N(t+s+h) - N(t+s) \geq 2] = o(h)$ et que $\mathbb{P}[N(t+s+h) - N(t+s) = 1] = 1 - \mathbb{P}[N(t+s+h) - N(t+s) = 0] - \mathbb{P}[N(t+s+h) - N(t+s) \geq 2]$.

D'où

$$\frac{P_n(s+h) - P_n(s)}{h} = -\nu(t+s)P_n(s) + P_{n-1}(s)\nu(t+s) + P_n(s)\frac{o_0(h)}{h} + P_{n-1}(s)\frac{o_1(h)}{h} + \sum_{k=2}^n P_{n-k}(s)\frac{o_k(h)}{h}.$$

En passant à la limite sur $h \rightarrow 0^+$, on a $D_s P_n(s) = -\nu(t+s)P_n(s) + \nu(t+s)P_{n-1}(s)$.

Comme c'est une équation différentielle à second membre linéaire, la solution générale est

$$\begin{aligned} P_n(s) &= e^{-\int \nu(t+s)ds} \left(\int e^{\int \nu(t+s)ds} \nu(t+s) P_{n-1}(s) ds + C \right) \\ &= e^{-(m(t+s)-m(t))} \left(\int e^{m(t+s)-m(t)} \nu(t+s) P_{n-1}(s) ds + C \right). \end{aligned}$$

On utilise alors l'hypothèse de récurrence selon laquelle $P_{n-1}(s) = e^{-(m(t+s)-m(t))} \frac{(m(t+s)-m(t))^{n-1}}{(n-1)!}$ pour obtenir

$$P_n(s) = e^{-(m(t+s)-m(t))} \left(\int \nu(t+s) \frac{(m(t+s)-m(t))^{n-1}}{(n-1)!} ds + C \right).$$

En posant $u = m(t+s) - m(t)$, cela donne

$$\begin{aligned} P_n(s) &= e^{-(m(t+s)-m(t))} \left(\frac{(m(t+s)-m(t))^n}{n} \frac{1}{(n-1)!} + C \right) \\ &= e^{-(m(t+s)-m(t))} \frac{(m(t+s)-m(t))^n}{n!} + e^{-(m(t+s)-m(t))} C. \end{aligned}$$

La condition initiale étant $P_n(0) = 0$, il vient alors $C = 0$ et $P_n(s)$ a bien la forme attendue. ■

Voici maintenant la distribution des marges finies du processus.

$\forall 0 < t_1 < t_2 < \dots < t_k, \forall n_1 \leq n_2 \leq \dots \leq n_k \in \mathbb{N}, k \in \mathbb{N}_0$,

$$\begin{aligned} &\mathbb{P}[N(t_1) = n_1, N(t_2) = n_2, \dots, N(t_k) = n_k] \\ &= \mathbb{P}[N(t_1) - N(0) = n_1, N(t_2) - N(t_1) = n_2 - n_1, \dots, N(t_k) - N(t_{k-1}) = n_k - n_{k-1}] \\ &= \mathbb{P}[N(t_1) = n_1] \mathbb{P}[N(t_2) - N(t_1) = n_2 - n_1] \dots \mathbb{P}[N(t_k) - N(t_{k-1}) = n_k - n_{k-1}] \text{ par indépendance} \\ &= e^{-m(t_1)} \frac{m(t_1)^{n_1}}{n_1!} e^{-(m(t_2)-m(t_1))} \frac{(m(t_2)-m(t_1))^{n_2-n_1}}{(n_2-n_1)!} \dots e^{-(m(t_k)-m(t_{k-1}))} \frac{(m(t_k)-m(t_{k-1}))^{n_k-n_{k-1}}}{(n_k-n_{k-1})!} \\ &= e^{-m(t_k)} \prod_{j=1}^k \frac{(m(t_j) - m(t_{j-1}))^{n_j - n_{j-1}}}{(n_j - n_{j-1})!} \quad \text{où } t_0 = 0 \text{ et } n_0 = 0. \end{aligned}$$

²voir dans la preuve de la proposition précédente.

Réciproquement à la proposition 3.2, nous avons le résultat suivant.

Proposition 3.3. *Soit $\{N(t), t \geq 0\}$ un processus de comptage tel que*

1. $N(0) = 0$;
2. *Les accroissements sont indépendants ;*
3. $\forall 0 \leq t$ et $0 < s$, $N(t+s) - N(t) \sim \mathcal{P}(m(t+s) - m(t))$ où la fonction m est continue, différentiable et croissante.

Alors ce processus est un processus de Poisson non homogène de fonction moyenne $m(t)$ et de fonction d'intensité donnée par $\nu(t) = D_t m(t)$.

Démonstration. Les conditions (NH1) et (NH2) sont satisfaites par hypothèses. Pour la condition (NH3), soit $t \geq 0$.

$$\begin{aligned} \lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) \geq 2]}{\mathbb{P}[N(t+h) - N(t) = 1]} &= \lim_{h \rightarrow 0^+} \frac{1 - \mathbb{P}[N(t+h) - N(t) = 0] - \mathbb{P}[N(t+h) - N(t) = 1]}{\mathbb{P}[N(t+h) - N(t) = 1]} \\ &= \lim_{h \rightarrow 0^+} \frac{1 - e^{-(m(t+h)-m(t))}}{e^{-(m(t+h)-m(t))}(m(t+h) - m(t))} - 1. \end{aligned}$$

Le théorème de l'Hospital permet ensuite d'obtenir

$$\lim_{h \rightarrow 0^+} \frac{\mathbb{P}[N(t+h) - N(t) \geq 2]}{\mathbb{P}[N(t+h) - N(t) = 1]} = \lim_{h \rightarrow 0^+} \frac{1}{1 - (m(t+h) - m(t))} - 1 = 0$$

par continuité de m .

Passons alors à la condition (NH4). Soit encore $t \geq 0$.

$$\lim_{h \rightarrow 0^+} \frac{1 - \mathbb{P}[N(t+h) - N(t) = 0]}{h} = \lim_{h \rightarrow 0^+} \frac{1 - e^{-(m(t+h)-m(t))}}{h}$$

et, à nouveau par le théorème de l'Hospital, cela donne

$$\lim_{h \rightarrow 0^+} \frac{1 - \mathbb{P}[N(t+h) - N(t) = 0]}{h} = \lim_{h \rightarrow 0^+} \frac{e^{-(m(t+h)-m(t))} \nu(t+h)}{1} = \nu(t)$$

par continuité de m et ν en t . ■

Voici l'expression des fonctions moyenne et covariance du processus de Poisson non homogène.

Proposition 3.4. *Soit un processus de Poisson non homogène $\{N(t), t \geq 0\}$ de fonction moyenne $m(t)$. Pour tous $t, s \geq 0$,*

$$E[N(t)] = m(t) \text{ et } Cov[N(t), N(s)] = m(\min(s, t)).$$

Démonstration. Le résultat de l'espérance est immédiat car $N(t) \sim \mathcal{P}(m(t))$. De plus, on a $Cov[N(t), N(t)] = V[N(t)] = m(t)$.

Le reste de la preuve est semblable à celui de la proposition 2.2 avec comme seule différence que $E[N(t)] = V[N(t)] = m(t)$. Pour $s < t$, il vient

$$\begin{aligned} Cov[N(t), N(s)] &= E[(N(t) - E[N(t)])(N(s) - E[N(s)])] \\ &= \frac{m(t) + m(t)^2 + m(s) + m(s)^2 - (m(t) - m(s)) - (m(t) - m(s))^2}{2} - m(t)m(s) \\ &= m(\min(s, t)). \end{aligned}$$
■

3.2 Temps d'inter-arrivées et temps d'occurrence

Sans l'hypothèse de stationnarité des accroissements, la distribution des temps d'inter-arrivées n'est plus forcément exponentielle. Lorsque l'on connaît l'instant t_{n-1} auquel la dernière occurrence a eu lieu, la forme de la fonction de répartition de W_n est assez directe. Pour $t \geq 0$, il vient

$$\begin{aligned}
 F_{W_n|T_{n-1}=t_{n-1}}(t) &= \mathbb{P}[W_n \leq t | T_{n-1} = t_{n-1}] \\
 &= 1 - \mathbb{P}[W_n > t | T_{n-1} = t_{n-1}] \\
 &= 1 - \mathbb{P}[N(t_{n-1} + t) - N(t_{n-1}) = 0] \\
 &= 1 - \exp[-(m(t_{n-1} + t) - m(t_{n-1}))] \quad \forall n \in \mathbb{N}_0.
 \end{aligned} \tag{3.5}$$

Si $m(t) = a + \nu t$, c-à-d lorsque le processus est homogène car $\nu(t)$ vaut alors simplement ν , cette fonction est indépendante de t_{n-1} et correspond à la fonction de répartition d'une exponentielle négative de paramètre ν .

Afin d'obtenir la distribution marginale de W_n , il est possible d'exploiter la relation (3.5) en intégrant sur toutes les valeurs possibles pour la variable T_{n-1} .

Traitons d'abord le cas plus simple $n = 1$.

$$F_{W_1}(t) = \mathbb{P}[W_1 \leq t] = 1 - \mathbb{P}[W_1 > t] = 1 - \mathbb{P}[N(t) = 0] = 1 - e^{-m(t)}, \quad \forall t \in \mathbb{R}^+.$$

Delà, on tire $f_{W_1}(t) = e^{-m(t)}\nu(t)\chi_{[0,+\infty[}(t)$.

Passons alors au cas $n \geq 2$:

$$\begin{aligned}
 F_{W_n}(t) &= 1 - \mathbb{P}[W_n > t] = 1 - \int_0^{+\infty} \mathbb{P}[W_n > t | T_{n-1} = s] f_{T_{n-1}}(s) ds \quad ^3 \\
 &= 1 - \int_0^{+\infty} \mathbb{P}[N(s+t) - N(s) = 0] f_{T_{n-1}}(s) ds \\
 &= 1 - \int_0^{+\infty} e^{-(m(t+s)-m(s))} f_{T_{n-1}}(s) ds, \quad \forall t \in \mathbb{R}^+.
 \end{aligned}$$

Il faut alors rechercher la fonction de densité $f_{T_{n-1}}$ de la variable T_{n-1} . Pour cela, recherchons la fonction de répartition de T_{n-1} .

$$\begin{aligned}
 F_{T_{n-1}}(s) &= \mathbb{P}[T_{n-1} \leq s] = \mathbb{P}[N(s) \geq n-1] \\
 &= 1 - \mathbb{P}[N(s) < n-1] = 1 - \sum_{k=0}^{n-2} \mathbb{P}[N(s) = k] \\
 &= 1 - \sum_{k=0}^{n-2} e^{-m(s)} \frac{m(s)^k}{k!} \\
 &= 1 - e^{-m(s)} - \sum_{k=1}^{n-2} e^{-m(s)} \frac{m(s)^k}{k!}, \quad \forall s \in \mathbb{R}^+
 \end{aligned}$$

où la somme n'a pas lieu d'être si $n = 2$.

³cfr. S.M. ROSS, *Introduction to probability models*, Second edition, New-York, Academic Press, 1980.

On obtient la fonction de densité en dérivant la fonction de répartition

$$\begin{aligned}
 f_{T_{n-1}}(s) &= \nu(s)e^{-m(s)} - \sum_{k=1}^{n-2} \left(-e^{-m(s)}\nu(s)\frac{m(s)^k}{k!} + e^{-m(s)}\frac{m(s)^{k-1}\nu(s)}{(k-1)!} \right) \\
 &= \sum_{k=0}^{n-2} e^{-m(s)}\nu(s)\frac{m(s)^k}{k!} - \sum_{k=1}^{n-2} e^{-m(s)}\nu(s)\frac{m(s)^{k-1}}{(k-1)!} \\
 &= e^{-m(s)}\nu(s) \left(\sum_{k=0}^{n-2} \frac{m(s)^k}{k!} - \sum_{k=0}^{n-3} \frac{m(s)^k}{k!} \right) \\
 &= e^{-m(s)}\nu(s)\frac{m(s)^{n-2}}{(n-2)!}, \quad \forall s \in \mathbb{R}^+.
 \end{aligned} \tag{3.6}$$

L'expression (3.6) est valable $\forall n \geq 2$. Ainsi, pour tout $t \geq 0$, on trouve

$$\begin{aligned}
 F_{W_n}(t) &= 1 - \int_0^{+\infty} e^{-(m(t+s)-m(s))} e^{-m(s)}\nu(s)\frac{m(s)^{n-2}}{(n-2)!} ds \\
 &= 1 - \int_0^{+\infty} e^{-m(t+s)}\nu(s)\frac{m(s)^{n-2}}{(n-2)!} ds.
 \end{aligned} \tag{3.7}$$

En utilisant le théorème des intégrales paramétriques⁴, on obtient, $\forall t \in \mathbb{R}^+, \forall n \geq 2$,

$$f_{W_n}(t) = \int_0^{+\infty} e^{-m(t+s)}\nu(t+s)\nu(s)\frac{m(s)^{n-2}}{(n-2)!} ds \chi_{[0,+\infty[}(t). \tag{3.8}$$

Comme cette distribution dépend de n , on a perdu le caractère "identiquement distribué" des temps d'inter-arrivées; cependant, ils restent indépendants comme le montre le résultat suivant.

Proposition 3.5. *Soit un processus de Poisson non homogène $\{N(t), t \geq 0\}$ et soit $\{W_n, n \in \mathbb{N}_0\}$ le processus des temps d'inter-arrivées associé. Pour tous $n' < n$,*

$$\mathbb{P}[W_n > t | W_{n'} = s] = \mathbb{P}[W_n > t].$$

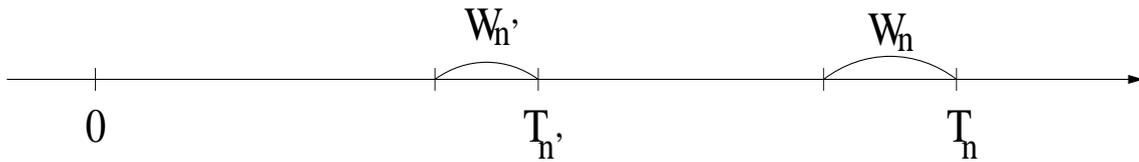
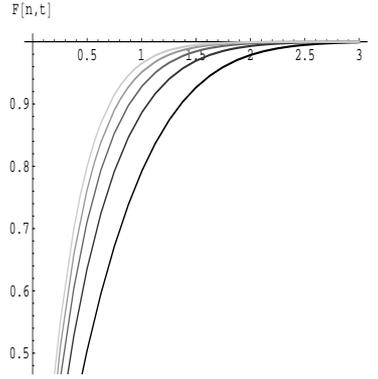


FIG. 3.9 – Indépendance des inter-arrivées

Démonstration. $W_{n'}$ ne dépend que des valeurs de $N(t)$ entre 0 et l'instant d'occurrence du n' -ième événement $T_{n'}$. W_n ne dépend lui que des valeurs de $N(t)$ entre $T_{n'}$ et T_n . Comme $n' < n$ implique que $T_{n'} < T_n$, et que des accroissements d'intersection vide sont indépendants, on peut conclure à l'indépendance de $W_{n'}$ et W_n . ■

⁴cfr J.SCHMETS, *ANALYSE MATHÉMATIQUE Introduction au Calcul Intégral*, Liège, Éditions De-rouaux, 1994.


 FIG. 3.10 – Fonction de répartition de W_n dans le cas où $\nu(t) = t$

Dans la littérature, ce sujet n'est pas plus traité. La distribution d'un temps d'inter-arrivées quelconque a fait l'objet de quelques recherches supplémentaires qui n'ont rien apporté de bien concret dans un délai raisonnable.

Passons alors aux variables temps d'occurrence du processus. Comme cela a été fait dans le développement de la fonction de répartition des temps d'inter-arrivées W_n , on peut retrouver l'équation (3.6) selon laquelle la fonction de densité du temps d'arrivée T_n du n^e événement vaut

$$f_{T_n}(s) = e^{-m(s)} \nu(s) \frac{m(s)^{n-1}}{(n-1)!} \chi_{[0, +\infty[}(t). \quad (3.9)$$

Revenons aux exemples de la page 38, et plus particulièrement sur celui de l'intensité linéaire,

$$\nu(t) = ct, \quad c > 0.$$

Les expressions des fonctions de répartition et de densité des temps d'inter-arrivées ne sont pas intégrables facilement sans ordinateur, et cela malgré que cet exemple soit le plus simple possible. C'est pourquoi les autres exemples introduits précédemment ne seront pas abordés ici. Le logiciel mathematica a donc été utilisé pour obtenir la représentation graphique de l'évolution de la fonction de répartition $F[n, t]$ suite à des variations des paramètres n et t reprise à la figure 3.10. Les différentes courbes correspondent à des valeurs de n différentes, de 1 à 6 en partant de la plus foncée.

Cette figure montre que la probabilité qu'un temps inter-arrivées W_n soit plus petit qu'un temps t augmente avec t et, si t est fixé, cette probabilité augmente aussi avec n . La première caractéristique est simplement due à la croissance de la fonction de répartition. La seconde est propre aux cas où la fonction d'intensité est strictement croissante. En effet, à mesure que le temps s'écoule, l'intensité augmente et il s'en suit que les temps d'inter-arrivées sont plus petits. Ainsi, pour un temps t fixé, si $n < m$, alors $\mathbb{P}[W_n \leq t] \leq \mathbb{P}[W_m \leq t]$ puisque l'événement n se produit avant l'événement m et donc $W_n \geq W_m$.

L'équation (3.9) nous donne, pour $t \geq 0$,

$$f_{T_n}(t) = e^{-\frac{c}{2}t^2} ct \left(\frac{c}{2}t^2\right)^{n-1} \frac{1}{(n-1)!} = \frac{c^n}{2^{n-1}(n-1)!} t^{2n-1} e^{-\frac{c}{2}t^2}.$$

La formule de changement de variable pour $Y = T_n^2$ donne

$$\begin{aligned} f_Y(y) &= f_{T_n}(t(y)) \frac{\partial t(y)}{\partial y} \quad \text{où } t(y) = \sqrt{y} \\ &= \frac{c^n}{2^{n-1}(n-1)!} y^{n-1} \sqrt{y} e^{-\frac{c}{2}y} \frac{1}{2\sqrt{y}} \\ &= \left(\frac{c}{2}\right)^n \frac{1}{\Gamma(n)} y^{n-1} e^{-\frac{c}{2}y} \end{aligned}$$

qui est la fonction de densité d'une variable aléatoire suivant une loi gamma de paramètre n et $\frac{c}{2}$ comme le montre la définition 1.7.

Ainsi, on a montré que, dans le cas d'une intensité linéaire, $T_n^2 \sim \text{gamma}(n, \frac{c}{2})$. Ce résultat pourrait être plus utile que la loi de T_n car la loi gamma est bien connue.

De manière analogue au processus de Poisson de base, on peut obtenir une adaptation des propriétés 2.5 et 2.7 :

Proposition 3.6. *Soit $\{N(t), t \geq 0\}$ un processus de Poisson non homogène de fonction d'intensité $\nu(t)$. Sachant qu'en un instant fixé T , n événements ont déjà eu lieu, les densités des temps d'occurrence des n -ième et $(n+1)$ -ième événements sont données par*

$$f_{T_n|N(T)=n}(t) = \frac{n(m(t))^{n-1}\nu(t)}{(m(T))^n} \chi_{[0,T]}(t), \tag{3.10}$$

$$f_{T_{n+1}|N(T)=n}(t) = e^{-(m(t)-m(T))} \nu(t) \chi_{]T,+\infty]}(t). \tag{3.11}$$

où $m(t) = \int_0^t \nu(y)dy$ est la fonction moyenne.

La démonstration est similaire à celle donnée dans le cas d'un processus de base, la seule différence étant les distributions de $N(t)$ et de $N(t) - N(T)$.

Les conclusions à tirer sont les mêmes; sachant que n événements ont eu lieu entre 0 et T , la distribution du temps d'occurrence du n -ième est indépendante de l'intensité et la distribution du $(n+1)$ -ième est indépendante du paramètre n et, si $T = 0$, elle correspond à la distribution du premier temps d'inter-arrivées.

Proposition 3.7. *Soit $\{N(t), t \geq 0\}$ un processus de Poisson non homogène de fonction moyenne $m(t) = \int_0^t \nu(u)du$. Sous la condition que $N(t) = n$, les temps $T_1 < T_2 < \dots < T_n$ dans l'intervalle $[0, t]$ auxquels se sont produits les n événements ont la même distribution que les statistiques d'ordre de n variables aléatoires U_1, U_2, \dots, U_n indépendantes et de même distribution*

$$F_{U_j}(u) = \frac{m(u)}{m(t)} \quad \text{si } 0 \leq u \leq t.$$

C'est-à-dire que

$$f_{(T_1, T_2, \dots, T_n | N(t)=n)}(t_1, t_2, \dots, t_n) = \begin{cases} \frac{n!}{m(t)^n} \nu(t_1) \dots \nu(t_n) & \text{si } 0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq t \\ 0 & \text{sinon.} \end{cases}$$

Démonstration. La preuve suit le même schéma que celle de la proposition 2.7. Soient des réels t_i et h_i tels que

$0 \leq t_1 < t_1 + h_1 < t_2 < \dots < t_i < t_i + h_i < t_{i+1} < \dots < t_{n+1} = t$. Il vient alors

$$\begin{aligned} & \mathbb{P}[t_i < T_i \leq t_i + h_i, i = 1, 2, \dots, n | N(t) = n] \\ &= \frac{1}{\mathbb{P}[N(t)=n]} \mathbb{P}[N(t_1) = 0, N(t_1 + h_1) - N(t_1) = 1, N(t_2) - N(t_1 + h_1) = 0, \dots, \\ & \qquad \qquad \qquad N(t_n + h_n) - N(t_n) = 1, N(t_{n+1}) - N(t_n + h_n) = 0] \\ &= \frac{n!}{e^{-m(t)} m(t)^n} e^{-m(t_1)} e^{-(m(t_1+h_1)-m(t_1))} (m(t_1 + h_1) - m(t_1)) e^{-(m(t_2)-m(t_1+h_1))} \dots \\ & \qquad \qquad \qquad e^{-(m(t_n+h_n)-m(t_n))} (m(t_n + h_n) - m(t_n)) e^{-(m(t)-m(t_n+h_n))} \\ &= \frac{n!}{m(t)^n} (m(t_1 + h_1) - m(t_1)) (m(t_2 + h_2) - m(t_2)) \dots (m(t_n + h_n) - m(t_n)) \end{aligned}$$

car toutes les exponentielles se simplifient.

Par définition de la fonction de densité, on a

$$\begin{aligned} f_{(T_1, T_2, \dots, T_n | N(t)=n)}(t_1, t_2, \dots, t_n) &= \frac{n!}{m(t)^n} \lim_{h_1 \rightarrow 0} \frac{m(t_1 + h_1) - m(t_1)}{h_1} \dots \lim_{h_n \rightarrow 0} \frac{m(t_n + h_n) - m(t_n)}{h_n} \\ &= \frac{n!}{m(t)^n} \nu(t_1) \dots \nu(t_n) \end{aligned}$$

si $0 \leq t_1 \leq \dots \leq t_n \leq t$ et $f_{(T_1, T_2, \dots, T_n | N(t)=n)}(t_1, t_2, \dots, t_n) = 0$ sinon.

Il reste à montrer que cette densité correspond bien à celle des statistiques d'ordre de n variables U_j indépendantes et de même distribution

$$F_{U_j}(u) = \frac{m(u)}{m(t)} \quad \text{si } 0 \leq u \leq t.$$

Pour $0 \leq u \leq t$, on a donc $f_{U_j}(u) = \frac{\nu(u)}{m(t)}$ et $f_{U_j}(u) = 0$ si $u \notin [0, t]$.

Par indépendance des U_j ,

$$f_{(U_1, \dots, U_n)}(u_1, \dots, u_n) = \begin{cases} \frac{\nu(u_1) \dots \nu(u_n)}{m(t)^n} & \text{si } 0 \leq u_i \leq t \quad \forall i = 1, 2, \dots, n \\ 0 & \text{sinon.} \end{cases}$$

Par le même raisonnement que dans la preuve de le lemme 2.6, on obtient finalement

$$f_{(U_{(1)}, \dots, U_{(n)})}(u_1, \dots, u_n) = \begin{cases} \frac{n!}{m(t)^n} \nu(u_1) \dots \nu(u_n) & \text{si } 0 \leq u_i \leq t \quad \forall i = 1, 2, \dots, n \\ 0 & \text{sinon} \end{cases}$$

■

Comme dans le chapitre précédent, il est possible de définir les distributions tronquées. Les équations (2.19) et (2.20) restent valables, seule l'équation (2.21) doit être adaptée. Si la fonction moyenne du processus $\{N(t), t \geq 0\}$ est $m(t)$, $\forall t \in [0, N]$, (2.21) devient

$$\mathbb{P}[N^*(t) = n] = \frac{m(t)^n}{n! \sum_{i=0}^N \frac{m(t)^i}{i!}}. \quad (3.12)$$

De plus, comme le processus n'est plus stationnaire, il faudrait chercher l'expression de la distribution de $N(t) - N(s)$ tronquée dans $[0, N]$ pour tout $s < t$. Celle-ci prend la forme finale

$$\mathbb{P}[N^*(t) - N^*(s) = n] = \frac{(m(t) - m(s))^n}{n! \sum_{i=0}^N \frac{(m(t) - m(s))^i}{i!}}$$

pour tout $n \in [0, N]$.

3.3 Lien avec le processus de Poisson de base

Cette section a pour but de montrer qu'un processus de Poisson non homogène peut toujours être transformé en un processus de Poisson d'intensité unitaire pour autant que la fonction moyenne soit continue et non-décroissante (hypothèses faites au départ). Si l'hypothèse que la fonction moyenne est strictement croissante est également émise, la réciproque sera vraie.

Considérons d'abord le cas d'une fonction moyenne croissante. Pour cela, il faut introduire une nouvelle fonction définie à partir de la fonction moyenne $m(t)$, notons-la m^{-1} . Elle est définie par

$$m^{-1}(u) = \inf\{t \geq 0 : m(t) \geq u\}, \quad \forall u \geq 0.$$

Lorsque la fonction m est bornée, il existe des valeurs de u pour lesquelles $m(t) < u$ pour tout $t \geq 0$. Notons alors u_c le plus petit $u \geq 0$ tel que cela. Comme m est croissante, cela implique que $\lim_{t \rightarrow +\infty} m(t) = u_c$. Pour tout $u \geq u_c$, l'ensemble $\{t \geq 0 : m(t) \geq u\}$ est vide et $m^{-1}(u)$ n'est donc pas défini dans \mathbb{R} . Dans ce cas, nous posons $m^{-1}(u) = +\infty$.

Soit $\{N(t), t \geq 0\}$ un processus de Poisson non homogène de fonction moyenne m . Si m est bornée, l'intensité retombe à zéro à un certain moment et $N("+\infty") = \lim_{t \rightarrow +\infty} N(t)$ est quand même fini. Ainsi, $N(m^{-1}(u))$ est constant pour $u \geq u_c$. L'étude du processus n'est donc intéressante que pour des $u \in [0, u_c]$.

Lemme 3.8. *Soit une fonction m continue et différentiable, qui est non-décroissante. Soient encore la fonction m^{-1} et $u_c \in \mathbb{R}^+$ définis comme précédemment. La fonction m^{-1} est strictement croissante sur $[0, u_c]$ et pour tout $u \in [0, u_c]$,*

$$m(m^{-1}(u)) = u.$$

Si, en outre, m est strictement croissante, alors $m^{-1}(m(t)) = t$, pour tout $t \geq 0$. Dans ce cas, la fonction m^{-1} est l'inverse de la fonction m .

Démonstration. Commençons par la croissance de la fonction m^{-1} . Soient $0 \leq u < v \leq u_c$. Comme m est croissante, il vient que l'équation $m(t) \geq u$ est satisfaite par des valeurs de la fonction m plus petites que celles satisfaisant $m(t) \geq v$ et donc par des t plus petits. De là, on tire que $m^{-1}(u) < m^{-1}(v)$ et donc que m^{-1} est strictement croissante.

Comme m est continue, le théorème des valeurs intermédiaires⁵ implique l'existence de $t' \geq 0$ tel que $m(t') = u$ pour tout $u \in]0, u_c[$. Si il en existe plusieurs, on considère le plus petit d'entre eux. De plus, comme m est non-décroissante, il n'existe pas de $t'' < t'$ tel que $m(t'') > m(t')$. De là, pour $u \in]0, u_c[$, il vient que

$$m(m^{-1}(u)) = m(\inf\{t \geq 0 : m(t) \geq u\}) = m(t') = u.$$

⁵cfr J.SHMETS, *ANALYSE MATHÉMATIQUE*, Liège, Éditions Derouaux, 1996.

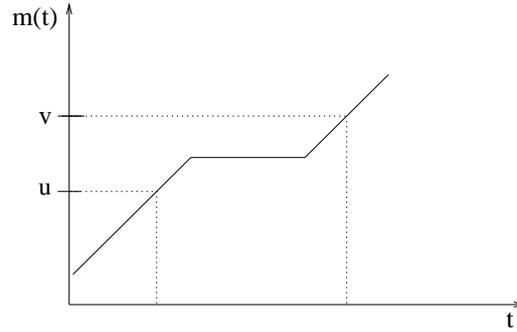


FIG. 3.11 – Illustration d’une fonction croissante

Pour $u = 0$ et $u = u_c$, cela vient du fait que $m(0) = 0$ et $m(+\infty) = \lim_{t \rightarrow +\infty} m(t) = u_c$. Si m est supposée strictement croissante, le théorème de la fonction inverse³, entraîne que m soit inversible. On sait déjà que $m(m^{-1}(u)) = u$ pour $u \in [0, u_c]$. La fonction m^{-1} sera l’inverse de la fonction m si en plus, $m^{-1}(m(t)) = t$ pour tout $t \geq 0$. Or,

$$m^{-1}(m(t)) = \inf\{t' \geq 0 : m(t') \geq m(t)\} = t$$

car m est strictement croissant donc il n’existe pas de $t'' < t$ tel que $m(t'') \geq m(t)$. ■

Proposition 3.9. Soit $\{N(t), t \geq 0\}$ un processus de Poisson non homogène de fonction moyenne m . Considérons la fonction m^{-1} et u_c définis comme précédemment. Si on pose $M(u) = N(m^{-1}(u))$, alors le processus $\{M(u), u \geq 0\}$ est un processus de Poisson d’intensité $\nu = 1$.

Démonstration. Comme cela a été expliqué avant, on ne considère le processus que pour des valeurs plus petites que la valeur critique u_c .

Soient $0 \leq u_1 < u_2 < u_3 \leq u_c$. Vu le lemme 3.8, on obtient

$$M(u_3) - M(u_2) = N(m^{-1}(u_3)) - N(m^{-1}(u_2))$$

est indépendant de

$$M(u_2) - M(u_1) = N(m^{-1}(u_2)) - N(m^{-1}(u_1))$$

car $m^{-1}(u_1) < m^{-1}(u_2) < m^{-1}(u_3)$ et car les accroissements de $\{N(t), t \geq 0\}$ sont indépendants.

Pour conclure, nous allons voir que, pour tous $0 \leq v < u \leq u_c$, $M(u) - M(v) \sim \mathcal{P}(u - v)$. En effet,

$$M(u) - M(v) = N(m^{-1}(u)) - N(m^{-1}(v)) \sim \mathcal{P}(m(m^{-1}(u)) - m(m^{-1}(v)))$$

par hypothèse sur $\{N(t), t \geq 0\}$. Vu le lemme 3.8, on a bien $M(u) - M(v) \sim \mathcal{P}(u - v)$. Comme $M(u - v) = N(m^{-1}(u - v))$ suit aussi une loi de Poisson de paramètre $u - v$, cela démontre la stationnarité des accroissements. De plus, cela permet de conclure en vertu de la définition 2.1. ■

Si l'on fait l'hypothèse supplémentaire que la fonction moyenne m est strictement croissante, la réciproque est également vraie :

Proposition 3.10. *Soient $\{M(u), u \geq 0\}$ un processus de Poisson de base d'intensité $\nu = 1$ et m une fonction continue, différentiable et strictement croissante. Si l'on pose $N(t) = M(m(t))$, alors le processus $\{N(t), t \geq 0\}$ est un processus de Poisson non homogène de fonction moyenne m .*

Démonstration. Il suffit de vérifier les conditions de la proposition 3.3.

1. $N(0) = M(0) = 0$.
2. Pour $0 \leq t_1 < t_2 < t_3$,

$$N(t_3) - N(t_2) = M(m(t_3)) - M(m(t_2)) \text{ et } N(t_2) - N(t_1) = M(m(t_2)) - M(m(t_1))$$

qui sont indépendants car m est strictement croissante et $\{M(u), u \geq 0\}$ est un processus de Poisson.

3. Pour $0 \leq s < t$,

$$N(t) - N(s) = M(m(t)) - M(m(s)) \sim M(m(t) - m(s)) \sim \mathcal{P}(m(t) - m(s))$$

par hypothèse sur $\{M(u), u \geq 0\}$. ■

Pour régler ce problème de croissance de la fonction moyenne, il faudra poser $t = 0$ pour le temps t tel que $t = \min\{t' \geq 0 : \nu(t') > 0\}$ et arrêter le processus en un temps T tel que $T = \min\{t' > 0 : \nu(t') = 0\}$, c-à-d qu'il faudra considérer un processus tronqué sur $[0, T]$.

Une autre manière d'exprimer la proposition précédente serait celle-ci :

Proposition 3.11. *Soient $\{N(t), t \geq 0\}$ un processus de Poisson d'intensité $\nu = 1$, $\{T_n, n \in \mathbb{N}\}$ le processus des temps d'occurrence associé et une fonction m continue, différentiable et strictement croissante. Posons $T'_n = m^{-1}(T_n)$. Le processus $\{T'_n, n \in \mathbb{N}\}$ est alors le processus des temps d'occurrence associé à un processus de Poisson non homogène de fonction moyenne m .*

Démonstration. Rappelons que, $\forall n \in \mathbb{N}$ fixé, $T_n \sim \text{gamma}(n, \nu)$ avec ici $\nu = 1$:

$$f_{T_n}(t) = \frac{1}{\Gamma(n)} t^{n-1} e^{-t} \chi_{[0, +\infty[}(t).$$

La formule du changement de variable dans ce cas est donnée par

$$f_{T'_n}(t') = f_{T_n}(t(t')) \frac{\partial(t(t'))}{\partial t'} \text{ où } t(t') = m(t'),$$

car m est inversible.

Ainsi, $f_{T'_n}(t') = \frac{1}{(n-1)!} m(t')^{n-1} e^{-m(t')} \nu(t')$, qui est l'expression de la densité du n -ième temps d'occurrence d'un processus de Poisson non homogène de fonction moyenne $m(t)$ comme le montre l'équation (3.9). ■

On voit ainsi qu'un processus non-stationnaire peut être transformé en processus stationnaire simplement en étirant ou en comprimant l'échelle du temps.

Chapitre 4

Processus de Poisson composé

Jusqu'à présent les processus étudiés ne permettaient pas que plusieurs événements se produisent en même temps, cela à cause de l'hypothèse

$$\lim_{h \rightarrow 0} \frac{\mathbb{P}[N(t+h) - N(t) \geq 2]}{h} = 0$$

conséquence de (C4) et de (NH3). Le processus de Poisson composé lève cette hypothèse et permet ainsi des arrivées en "grappes".

Définition 4.1. *Un processus de comptage $\{X(t), t \geq 0\}$ est un **processus de Poisson composé** si $\forall t \geq 0$, $X(t) = \sum_{n=0}^{N(t)} Y_n$ où $\{N(t), t \geq 0\}$ est un processus de Poisson, $\{Y_n, n = 1, 2, \dots\}$ est une famille de variables aléatoires indépendantes et identiquement distribuées selon une loi de probabilité quelconque F , appelée distribution de gain, et $Y_0 = 0$. Le processus $\{N(t), t \geq 0\}$ et la suite $\{Y_n, n = 1, 2, \dots\}$ sont supposés indépendants.*

Un exemple d'un processus de Poisson composé pourrait être le nombre de litres d'essence vendus à une pompe d'une station service. Le processus $\{N(t), t \geq 0\}$ est le processus d'arrivée des clients à cette pompe et pour le n -ième client, Y_n est le nombre de litres qu'il a achetés.

4.1 Distribution des marges finies du processus

Proposition 4.1. *Un processus de Poisson composé $\{X(t), t \geq 0\}$ a des accroissements stationnaires et indépendants. Sa fonction caractéristique est donnée par*

$$\phi_{X(t)}(u) = e^{\nu t(\phi_Y(u)-1)}$$

où

- ϕ_Y est la fonction caractéristique des variables Y_n ,
- ν est l'intensité du processus $\{N(t), t \geq 0\}$ sous-jacent au processus $\{X(t), t \geq 0\}$.

De plus, si $E[Y^2] < +\infty$, alors

$$\begin{aligned} E[X(t)] &= \nu t E[Y], \\ V[X(t)] &= \nu t E[Y^2], \\ \text{cov}[X(t), X(s)] &= \nu E[Y^2] \min(s, t). \end{aligned}$$

Démonstration. Soient $0 \leq t_1 < t_2 < t_3$ et considérons les deux accroissements $X(t_3) - X(t_2)$ et $X(t_2) - X(t_1)$.

$$X(t_3) - X(t_2) = \sum_{n=0}^{N(t_3)} Y_n - \sum_{n=0}^{N(t_2)} Y_n = \sum_{n=N(t_2)+1}^{N(t_3)} Y_n$$

et, de la même manière,

$$X(t_2) - X(t_1) = \sum_{n=N(t_1)+1}^{N(t_2)} Y_n.$$

Ainsi, vu le caractère indépendant des variables Y_i , ces accroissements sont indépendants. Pour montrer que les accroissements sont stationnaires et donner la forme de la fonction caractéristique, il suffit de prouver que $\forall 0 \leq s < t$, on a

$$\phi_{X(t)-X(s)}(u) = e^{\nu(t-s)(\phi_Y(u)-1)}$$

car alors, on aura $\phi_{X(t-s)}(u) = \phi_{X(t-s)-X(0)}(u) = e^{\nu(t-s-0)(\phi_Y(u)-1)} = \phi_{X(t)-X(s)}(u)$.

Pour ce faire, commençons par chercher l'expression de l'espérance de $e^{iu(X(t)-X(s))}$ conditionnelle à $N(t) - N(s) = n$.

$\forall n \geq 1$,

$$\begin{aligned} E[e^{iu(X(t)-X(s))} | N(t) - N(s) = n] &= \sum_{k=0}^{+\infty} e^{iuk} \mathbb{P}[X(t) - X(s) = k | N(t) - N(s) = n] \\ &= \sum_{k=0}^{+\infty} e^{iuk} \mathbb{P}\left[\sum_{l=N(s)+1}^{N(t)} Y_l = k | N(t) - N(s) = n\right] \\ &= E\left[e^{iu \sum_{l=N(s)+1}^{N(t)} Y_l} | N(t) - N(s) = n\right] \\ &= E\left[\prod_{l=N(s)+1}^{N(t)} e^{iu Y_l} | N(t) - N(s) = n\right] \\ &= E\left[\prod_{l=N(s)+1}^{N(s)+n} e^{iu Y_l}\right] \\ &= \prod_{l=N(s)+1}^{N(s)+n} E[e^{iu Y_l}] \quad \text{par indépendance des } Y_n \\ &= \prod_{l=N(s)+1}^{N(s)+n} \phi_Y(u) \\ &= (\phi_Y(u))^n \quad \forall n = 1, 2, \dots \end{aligned} \tag{4.1}$$

Pour $n = 0$, $E[e^{iu(X(t)-X(s))} | N(t) - N(s) = 0] = E[e^{iu0}] = 1$, car si $N(t) = N(s)$, alors $X(t) = X(s)$ par définition. L'équation (4.1) reste donc valable pour $n = 0$.

Ainsi, on a

$$\begin{aligned}\phi_{X(t)-X(s)}(u) &= E \left[e^{iu(X(t)-X(s))} \right] = \sum_{k=0}^{+\infty} e^{iuk} \mathbb{P}[X(t) - X(s) = k] \\ &= \sum_{k=0}^{+\infty} e^{iuk} \sum_{n=0}^{+\infty} \mathbb{P}[X(t) - X(s) = k | N(t) - N(s) = n] \mathbb{P}[N(t) - N(s) = n]\end{aligned}$$

où l'on peut permuter les séries en se référant aux notes de cours d'ANALYSE MATHÉMATIQUE, Introduction aux espaces fonctionnels de J.SCHMETS.

$$\begin{aligned}\phi_{X(t)-X(s)}(u) &= \sum_{n=0}^{+\infty} \sum_{k=0}^{+\infty} e^{iuk} \mathbb{P}[X(t) - X(s) = k | N(t) - N(s) = n] \mathbb{P}[N(t) - N(s) = n] \\ &= \sum_{n=0}^{+\infty} E \left[e^{iu(X(t)-X(s))} | N(t) - N(s) = n \right] \mathbb{P}[N(t) - N(s) = n]\end{aligned}$$

ce qui donne en utilisant (4.1)

$$\phi_{X(t)-X(s)}(u) = \sum_{n=0}^{+\infty} (\phi_Y(u))^n \mathbb{P}[N(t) - N(s) = n].$$

Comme $N(t) - N(s) \sim N(t-s) \sim \mathcal{P}(\nu(t-s))$, on obtient

$$\begin{aligned}\phi_{X(t)-X(s)}(u) &= \sum_{n=0}^{+\infty} (\phi_Y(u))^n e^{-\nu(t-s)} \frac{(\nu(t-s))^n}{n!} \\ &= e^{-\nu(t-s)} e^{\phi_Y(u)\nu(t-s)} \\ &= e^{\nu(t-s)(\phi_Y(u)-1)}.\end{aligned}$$

Ce qui correspond bien à la forme voulue. De plus, en prenant $s = 0$, on obtient

$$\phi_{X(t)}(u) = e^{\nu t(\phi_Y(u)-1)}.$$

Passons maintenant aux calculs de l'espérance et de la variance. La relation (1.9) donne un moyen facile pour les calculer à partir de la fonction caractéristique.

$$E[X(t)] = -i \left[D_u \phi_{X(t)}(u) \right]_{u=0} = -i \left[e^{\nu t(\phi_Y(u)-1)} \nu t D_u \phi_Y(u) \right]_{u=0} = -i \nu t i E[Y] = \nu t E[Y],$$

car $\phi_Y(0) = 1$ comme le dit la proposition 1.9.

$$\begin{aligned}\text{De même, } E[(X(t))^2] &= - \left[D_u^2 \phi_{X(t)}(u) \right]_{u=0} \\ &= - \left[e^{\nu t(\phi_Y(u)-1)} (\nu t)^2 (D_u \phi_Y(u))^2 + e^{\nu t(\phi_Y(u)-1)} \nu t D_u^2 \phi_Y(u) \right]_{u=0} = (\nu t)^2 (E[Y])^2 + \nu t E[Y^2].\end{aligned}$$

De là, on tire que $V[X(t)] = (\nu t)^2 (E[Y])^2 + \nu t E[Y^2] - (\nu t)^2 (E[Y])^2 = \nu t E[Y^2]$.

Pour finir, la preuve du résultat sur la covariance suit le même schéma que la preuve de la proposition 2.2. Soient $s < t$,

$$\begin{aligned}\text{cov}[X(t), X(s)] &= E[(X(t) - E[X(t)])(X(s) - E[X(s)])] \\ &= E[X(s)X(t)] - \nu t E[Y] E[X(s)] - \nu s E[Y] E[X(t)] + \nu^2 t s (E[Y])^2.\end{aligned}$$

On utilise alors (2.8) pour obtenir (après quelques simplifications)

$$\begin{aligned} \text{cov}[X(t), X(s)] &= \frac{\nu s E[Y^2]}{2} + \frac{\nu s E[Y^2]}{2} + \nu^2 st (E[Y])^2 - \nu^2 st (E[Y])^2 \\ &= \nu E[Y^2] \min(s, t). \end{aligned}$$

■

Exemple 4.1. Si $\{N(t), t \geq 0\}$ est un processus de Poisson d'intensité ν et si les variables aléatoires Y_n sont toutes des variables de Bernoulli de paramètre p indépendantes, on a alors $X(t) \sim \mathcal{P}(\nu tp)$.

En effet, pour une variable $Y \sim \text{Bern}(p)$, on a $\phi_Y(u) = pe^{iu} + (1-p)$, $E[Y] = p$ et $E[Y^2] = p$. Vu la proposition 4.1, on a

$$\begin{aligned} E[X(t)] &= \nu tp, \\ V[X(t)] &= \nu tp, \\ \phi_{X(t)}(u) &= e^{\nu t(pe^{iu} - p)} = e^{\nu tp(e^{iu} - 1)}. \end{aligned}$$

On retrouve ainsi la fonction caractéristique d'une variable de Poisson de paramètre νtp . Pour visualiser ces processus, nous avons généré avec le logiciel R deux suites de 10 nombres aléatoires. La première, dont la distribution est exponentielle avec $\nu = 2$ pour paramètre, correspond aux temps d'inter-arrivées du processus. La seconde, ayant une distribution de Bernoulli de paramètre $p = 0,4$, est donnée par

$$(0, 0, 1, 0, 1, 0, 0, 0, 0, 0).$$

Elle donne une réalisation de la famille des Y_n .

Comme cela a été expliqué à la page 29, on obtient les temps d'occurrence. A partir de ces temps d'occurrence, il est possible de tracer une trajectoire de $N(t)$, comme cela est fait à la figure 4.1. Et avec la formule $X(t) = \sum_{n=0}^{N(t)} Y_n$, on obtient le graphe 4.2 d'une trajectoire du processus $\{X(t), t \geq 0\}$.

On voit bien que le nombre d'événements par unité de temps diminue car l'espérance est

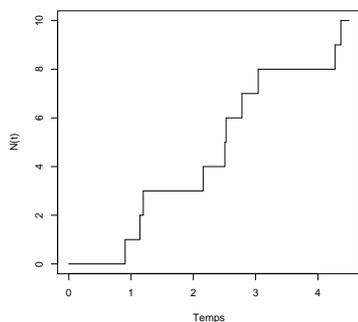


FIG. 4.1 – Trajectoire de $N(t)$

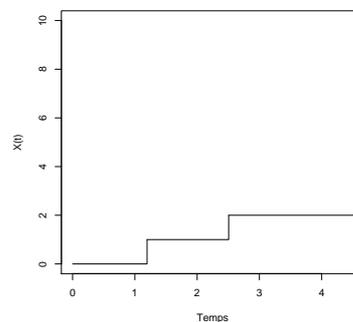


FIG. 4.2 – Trajectoire de $X(t)$

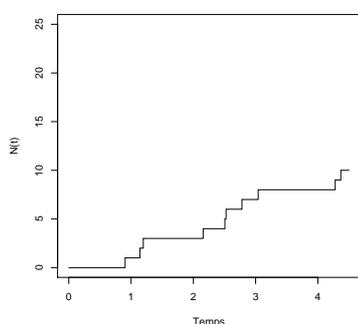
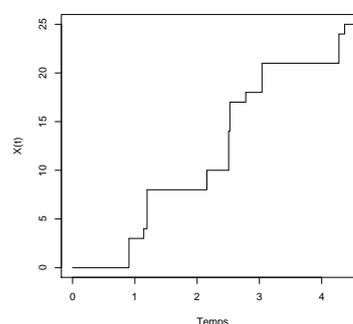
multipliée par 0,4 et passe donc de 2 à 0,8.

Exemple 4.2. L'exemple précédent est intéressant car il présente le cas d'un processus de Poisson composé qui est aussi un processus de Poisson de base. Mais il n'illustre pas le nouveau phénomène d'arrivées simultanées. Pour visualiser ce cas, considérons une distribution binomiale à la place d'une bernoulli. Ainsi, la réalisation suivante suit une loi binomiale de paramètre $n = 5$ et $p = 0,4$:

$$(3, 1, 4, 2, 4, 3, 1, 3, 3, 1).$$

Le processus composé $X(t)$ obtenu avec le même processus sous-jacent que dans l'exemple 4.1 est représenté à la figure 4.4. La graphique 4.3 est une représentation du processus sous-jacent avec la même échelle que celle utilisée pour le processus composé afin de pouvoir les comparer facilement.

Lorsque deux ou plusieurs événements ont lieu en même temps, le saut dans le graphique de

FIG. 4.3 – Trajectoire de $N(t)$ FIG. 4.4 – Trajectoire de $X(t)$

$X(t)$ est plus grand. La représentation graphique d'un processus est donc un moyen simple de voir s'il est composé ou non. Ces figures illustrent aussi le fait que les temps d'occurrence des processus sous-jacent et composé sont les mêmes.

Pour l'instant aucune hypothèse n'a été émise concernant la distribution F des Y_n , mais une première intuition pousserait à la supposer discrète puisque la variable Y_n représente le nombre d'événements se produisant lors de la n -ième occurrence du processus de Poisson sous-jacent. A ce propos, dans le cas où la famille de variables aléatoires $\{Y_n, n = 1, 2, \dots\}$ ne prend que des valeurs entières et strictement positives, on peut écrire explicitement la loi de probabilité de $X(t)$ pour tout $t \geq 0$ fixé. Pour cela, nous avons besoin d'un lemme. Il est démontré ici dans le cas d'une variable aléatoire discrète et à valeurs entières strictement positives, mais il se généralise facilement. Il est de plus supposé que les espérances rencontrées existent toutes.

Lemme 4.2. Soient $N \sim \mathcal{P}(\lambda)$, $\{Y_n, n = 1, 2, \dots\}$ une famille de variables aléatoires à valeurs entières et strictement positives, indépendantes entre elles et indépendantes de N et $Y_0 = 0$. Si $W = \sum_{i=0}^N Y_i$, alors, pour toute fonction $g(x)$, on a

$$E[Wg(W)] = \lambda E[Yg(W + Y)]$$

où Y a la même distribution que les Y_n .

Démonstration. En conditionnant par rapport aux valeurs de N , il vient

$$\begin{aligned} E[Wg(W)] &= \sum_{n=0}^{+\infty} E[Wg(W)|N = n]\mathbb{P}[N = n] \\ &= \sum_{n=0}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} E \left[\sum_{i=0}^n Y_i g \left(\sum_{j=0}^n Y_j \right) \right]. \end{aligned}$$

Par linéarité de l'espérance et par distribution identique des Y_i , on obtient donc

$$E[Wg(W)] = \sum_{n=0}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} \sum_{i=0}^n E \left[Y_n g \left(\sum_{j=0}^n Y_j \right) \right].$$

Comme la somme sur i est inutile, un facteur n apparaît et implique que le terme pour $n = 0$ soit nul,

$$\begin{aligned} E[Wg(W)] &= \sum_{n=1}^{+\infty} e^{-\lambda} \frac{\lambda^n}{n!} n E \left[Y_n g \left(\sum_{j=0}^n Y_j \right) \right] \\ &= \lambda \sum_{n=1}^{+\infty} e^{-\lambda} \frac{\lambda^{n-1}}{(n-1)!} \sum_{y=0}^{+\infty} E \left[Y_n g \left(\sum_{j=0}^{n-1} Y_j + Y_n \right) | Y_n = y \right] \mathbb{P}[Y_n = y]. \end{aligned}$$

Utilisons alors le fait que Y à la même distribution que les Y_n ,

$$E[Wg(W)] = \lambda \sum_{m=0}^{+\infty} e^{-\lambda} \frac{\lambda^m}{m!} \sum_{y=0}^{+\infty} y E \left[g \left(\sum_{j=0}^m Y_j + y \right) \right] \mathbb{P}[Y = y],$$

et en se référant aux notes de cours d'ANALYSE MATHÉMATIQUE, *Introduction aux espaces fonctionnels* de J.SCHMETS, il est possible de permuter les séries,

$$\begin{aligned} E[Wg(W)] &= \lambda \sum_{y=0}^{+\infty} y \sum_{m=0}^{+\infty} e^{-\lambda} \frac{\lambda^m}{m!} E \left[g \left(\sum_{j=0}^m Y_j + y \right) \right] \mathbb{P}[Y = y] \\ &= \lambda \sum_{y=0}^{+\infty} y \sum_{m=0}^{+\infty} E [g(W + y) | N = m] \mathbb{P}[N = m] \mathbb{P}[Y = y] \\ &= \lambda \sum_{y=0}^{+\infty} y E[g(W + y)] \mathbb{P}[Y = y] \\ &= \lambda \sum_{y=0}^{+\infty} E[Yg(W + Y) | Y = y] \mathbb{P}[Y = y] \\ &= \lambda E[Yg(W + Y)]. \end{aligned}$$

■

Il est alors possible de donner les moments de la variable $X(t)$ lorsque la famille $\{Y_n, n = 1, 2, \dots\}$ ne prend que des valeurs entières et strictement positives.

Proposition 4.3. *Soit $\{X(t), t \geq 0\}$ un processus de Poisson composé d'intensité ν tel que la famille de variables aléatoires $\{Y_n, n = 1, 2, \dots\}$ ne prenne que des valeurs entières et strictement positives. Soit encore Y une variable aléatoire de même distribution que les Y_n . Posons $\alpha_j = \mathbb{P}[Y_n = j]$, $\forall j \geq 1$ et $P_j = \mathbb{P}[X(t) = j]$, $\forall j \geq 0$ et $\forall t \geq 0$ fixés. On a alors*

$$E[(X(t))^n] = \nu t \sum_{j=0}^{n-1} C_{n-1}^j E[(X(t))^j] E[Y^{n-j}] \quad \forall n \geq 1,$$

$$P_0 = e^{-\nu t} \text{ et } P_j = \frac{\nu t}{j} \sum_{i=1}^j i \alpha_i P_{j-i} \quad \forall j \geq 1.$$

Démonstration. Commençons par la première égalité. Pour cela, fixons $n \geq 1$ et considérons la fonction $g(x) = x^{n-1}$ afin d'appliquer le lemme précédent. Nous avons bien les hypothèses nécessaires pour l'appliquer :

- $N(t) \sim \mathcal{P}(\nu t)$,
 - $\{Y_n, n = 1, 2, \dots\}$ est une famille de variables aléatoires à valeurs entières et strictement positives indépendantes entre elles et indépendantes de $N(t) \forall t \geq 0$,
 - $X(t) = \sum_{i=0}^{N(t)} Y_i$.
- Cela étant, il vient

$$\begin{aligned} E[(X(t))^n] &= E[X(t)g(X(t))] \\ &= \nu t E[Yg(X(t) + Y)] \quad \text{vu le lemme 4.2} \\ &= \nu t E \left[Y(X(t) + Y)^{n-1} \right] \\ &= \nu t E \left[Y \sum_{j=0}^{n-1} C_{n-1}^j (X(t))^j Y^{n-1-j} \right] \\ &= \nu t \sum_{j=0}^{n-1} C_{n-1}^j E \left[(X(t))^j \right] E[Y^{n-j}] \quad \text{par indépendance.} \end{aligned}$$

Ceci finit la première partie de la démonstration. Passons à la seconde.

Pour $j = 0$, on a

$$P_0 = \mathbb{P}[X(t) = 0] = \mathbb{P} \left[\sum_{n=0}^{N(t)} Y_n = 0 \right] = \mathbb{P}[N(t) = 0] = e^{-\nu t}$$

car les Y_n étant tous strictement positifs, le seul moyen que leur somme soit nulle est que celle-ci ne comporte aucun terme.

$$\forall j > 0, \text{ posons } g_j(x) = \begin{cases} 0 & \text{si } x \neq j \\ \frac{1}{j} & \text{si } x = j. \end{cases}$$

On a donc $\forall k \in \mathbb{R}, kg_j(k) = \begin{cases} 0 & \text{si } k \neq j \\ 1 & \text{si } k = j \end{cases}$, ce que l'on note $kg_j(k) = \delta_{j,k}$.

Dans ce cas, la définition de l'espérance donne

$$E[X(t)g_j(X(t))] = \sum_{k=0}^{+\infty} kg_j(k)\mathbb{P}[X(t) = k] = \mathbb{P}[X(t) = j].$$

Vu le lemme 4.2 et la relation ci-dessus, il vient

$$\begin{aligned} P_j &= \mathbb{P}[X(t) = j] = \nu t E[Y g_j(X(t) + Y)] \\ &= \nu t \sum_{i=1}^{+\infty} E[Y g_j(X(t) + Y) | Y = i] \mathbb{P}[Y = i]. \end{aligned}$$

Mais comme $Y > 0$, $X(t) \geq 0$ et $X(t) + Y = j$ impliquent $0 < Y \leq j$, cela équivaut à

$$\begin{aligned} P_j &= \mathbb{P}[X(t) = j] = \nu t \sum_{i=1}^j E[Y g_j(X(t) + Y) | Y = i] \mathbb{P}[Y = i] \\ &= \nu t \sum_{i=1}^j i E[g_j(X(t) + i)] \alpha_i \\ &= \nu t \sum_{i=1}^j i \sum_{k=0}^{+\infty} g_j(k+i) \mathbb{P}[X(t) = k] \alpha_i \\ &= \nu t \sum_{i=1}^j i \frac{1}{j} \mathbb{P}[X(t) = j-i] \alpha_i \quad \text{car } g_j(k+i) = \frac{1}{j} \delta_{j, k+i} \\ &= \frac{\nu t}{j} \sum_{i=1}^j i \alpha_i P_{j-i}. \end{aligned}$$

■

Voici alors l'expression de la distribution de toute marge finie du processus de Poisson composé dans le cas où les variables Y_n sont à valeurs entières et strictement positives : $\forall 0 < t_1 < t_2 < \dots < t_k, \forall n_1 \leq n_2 \leq \dots \leq n_k \in \mathbb{N}, k \in \mathbb{N}_0$, on a

$$\begin{aligned} &\mathbb{P}[X(t_1) = n_1, X(t_2) = n_2, \dots, X(t_k) = n_k] \\ &= \mathbb{P}[X(t_1) - X(0) = n_1, X(t_2) - X(t_1) = n_2 - n_1, \dots, X(t_k) - X(t_{k-1}) = n_k - n_{k-1}] \\ &= \mathbb{P}[X(t_1) = n_1] \mathbb{P}[X(t_2) - X(t_1) = n_2 - n_1] \dots \mathbb{P}[X(t_k) - X(t_{k-1}) = n_k - n_{k-1}] \text{ par indépendance} \\ &= \mathbb{P}[X(t_1) = n_1] \mathbb{P}[X(t_2 - t_1) = n_2 - n_1] \dots \mathbb{P}[X(t_k - t_{k-1}) = n_k - n_{k-1}] \text{ par stationnarité} \\ &= \frac{\nu t_1}{n_1} \left(\sum_{j=1}^{n_1} j \mathbb{P}[Y = j] \mathbb{P}[X(t_1) = n_1 - j] \right) \dots \\ &\quad \frac{\nu(t_k - t_{k-1})}{n_k - n_{k-1}} \left(\sum_{j=1}^{n_k - n_{k-1}} j \mathbb{P}[Y = j] \mathbb{P}[X(t_k - t_{k-1}) = n_k - n_{k-1} - j] \right) \end{aligned}$$

$$= \prod_{i=1}^k \left(\frac{\nu(t_i - t_{i-1})}{n_i - n_{i-1}} \sum_{j=1}^{n_i - n_{i-1}} j \mathbb{P}[Y = j] \mathbb{P}[X(t_i - t_{i-1}) = n_i - n_{i-1} - j] \right) \text{ où } t_0 = 0 \text{ et } n_0 = 0.$$

On n'obtient plus une formule directe comme dans les cas précédents mais plutôt une formule de récurrence. Cela est dû à la forme des probabilités d'états.

4.2 Temps d'inter-arrivées et temps d'occurrence

Comme dans les cas précédents, intéressons nous à la distribution des temps d'inter-arrivées W_n d'un processus de Poisson composé.

Proposition 4.4. *Soit $\{X(t), t \geq 0\}$ un processus de Poisson composé d'intensité ν défini au moyen du processus de Poisson $\{N(t), t \geq 0\}$ et de la famille de variables aléatoires $\{Y_n, n = 0, 1, 2, \dots\}$. Les temps d'inter-arrivées sont indépendants et identiquement distribués selon une loi exponentielle de paramètre ν .*

Démonstration. Par définition d'un processus de Poisson composé, ses temps d'inter-arrivées sont ceux du processus de Poisson sous-jacent $\{N(t), t \geq 0\}$. Vu le résultat (2.4), on a immédiatement que

$$W_n \stackrel{\text{iid}}{\sim} \text{Exp}(\nu).$$

■

Voyons maintenant ce qu'il en est de la distribution des temps d'occurrence T_n d'un processus de Poisson composé.

Le lien entre les variables aléatoires T_n et W_n est le même que dans le processus de base, c-à-d que l'on a

$$T_n = W_1 + W_2 + \dots + W_n \quad \forall n \geq 1.$$

Par la proposition précédente et la propriété d'additivité des exponentielles (proposition 1.18), on obtient que

$$T_n \sim \text{gamma}(n, \nu)$$

où ν est l'intensité du processus de Poisson $\{N(t), t \geq 0\}$ sous-jacent au processus composé. On a également les mêmes résultats que dans le cas d'un processus de base à savoir :

Proposition 4.5. *Soit $\{X(t), t \geq 0\}$ un processus de Poisson composé dont le processus sous-jacent $\{N(t), t \geq 0\}$ est d'intensité ν . Sachant que $N(T) = n$, T étant fixé, la fonction de densité du temps d'occurrence de la n -ième "grappe" est donnée par*

$$f_{T_n|N(T)=n}(t) = \frac{nt^{n-1}}{T^n} \chi_{[0, T]}(t), \quad (4.2)$$

cela indépendamment de l'intensité du processus. Dans les mêmes conditions, la densité du temps d'occurrence de la $(n+1)$ -ième "grappe" d'événements est donnée par

$$f_{T_{n+1}|N(t)=n}(t) = e^{-\nu(t-T)} \nu \chi_{[T, +\infty]}(t). \quad (4.3)$$

Proposition 4.6. Soit $\{X(t), t \geq 0\}$ un processus de Poisson composé dont le processus de Poisson sous-jacent est $\{N(t), t \geq 0\}$ d'intensité ν . La densité conditionnelle de (T_1, T_2, \dots, T_n) sachant que $N(t) = n$ est donnée par

$$f_{(T_1, T_2, \dots, T_n | N(t)=n)}(t_1, t_2, \dots, t_n) = \begin{cases} \frac{n!}{t^n} & \text{si } 0 \leq t_1 \leq t_2 \leq \dots \leq t_n \leq t \\ 0 & \text{sinon} \end{cases}$$

qui est la densité des statistiques d'ordre d'un ensemble de n variables indépendantes et identiquement distribuées selon une loi uniforme sur $[0, t]$.

Les preuves sont les mêmes que celles données au chapitre 2.

En ce qui concerne les distributions tronquées, les relations (2.19) et (2.20) sont toujours valables pour le processus sous-jacent $\{N(t), t \geq 0\}$ et en procédant de la même manière que lors du développement aboutissant à (2.20), il est aisé de voir que cette expression reste la même pour la variable $X(t)$. Cependant, il n'est pas aisé de la transformer en une forme plus explicite comme cela avait été le cas au chapitre 2.

Chapitre 5

Autres processus de Poisson

Dans ce chapitre sont présentés quelques processus de Poisson qui sont définis à partir d'un ou plusieurs autres processus de Poisson. Comme, pour chacun des cas abordés, il s'agira de processus de Poisson, il ne sera pas nécessaire de regarder les propriétés de ceux-ci car elles ont déjà été étudiées dans les chapitres précédents.

5.1 Amincissement d'un processus de Poisson

Définition 5.1. *Considérons un processus de Poisson $\{N(t), t \geq 0\}$ d'intensité constante ν . Soient un naturel r et (p_1, p_2, \dots, p_r) un vecteur de probabilités tel que chaque événement du processus a une probabilité p_j d'être de type j , pour tout $j = 1, 2, \dots, r$. Supposons en outre que les attributions des types sont indépendantes les unes des autres. Pour chaque $j \in \{1, 2, \dots, r\}$, le processus de comptage des événements de type j , $\{N_j(t), t \geq 0\}$, est appelé **amincissement du processus** $\{N(t), t \geq 0\}$.*

On peut démontrer qu'un amincissement d'un processus de Poisson est encore un processus de Poisson. Cependant, un lemme est nécessaire.

Lemme 5.1. *Soit G une variable aléatoire discrète de loi géométrique¹ de paramètre $p \in]0, 1[$, c-à-d que $IP[G = g] = (1 - p)^{g-1}p, \forall g \in \mathbb{N}_0$. Soient encore des variables aléatoires W_i indépendantes et identiquement distribuées selon une loi exponentielle de paramètre $\nu > 0$. La variable aléatoire Y définie par*

$$Y = \sum_{k=1}^G W_k$$

suit une loi exponentielle de paramètre νp .

Démonstration. Cherchons l'expression de la fonction génératrice des moments de la va-

¹Pour rappel, une variable aléatoire est dite géométrique si elle indique le rang du premier succès dans une succession d'essais indépendants et identiques d'une expérience aléatoire n'ayant que deux résultats possibles : l'échec avec une probabilité $1 - p$ et le succès avec une probabilité p . On note alors $G \sim G(p)$.

riable Y . Pour tout $t \in \mathbb{R}$ tel que l'espérance existe,

$$\begin{aligned} M_Y(t) &= E[e^{tY}] = \sum_{g=1}^{+\infty} E[e^{tY} | G = g] \mathbb{P}[G = g] \\ &= \sum_{g=1}^{+\infty} E\left[e^{t \sum_{k=1}^g W_k}\right] (1-p)^{g-1} p. \end{aligned}$$

En utilisant l'indépendance des variables W_k , il vient

$$M_Y(t) = \sum_{g=1}^{+\infty} \prod_{k=1}^g E[e^{tW_k}] (1-p)^{g-1} p.$$

La proposition 1.12 donne alors

$$\begin{aligned} M_Y(t) &= \sum_{g=1}^{+\infty} \left(\frac{\nu}{\nu-t}\right)^g (1-p)^{g-1} p, \forall t < \nu \\ &= \frac{\nu p}{\nu-t} \sum_{g=0}^{+\infty} \left(\frac{\nu}{\nu-t}(1-p)\right)^g. \end{aligned}$$

Si $t < \nu p$, alors la série correspond à la série géométrique² de raison $\frac{\nu}{\nu-t}(1-p)$, d'où l'on tire que

$$\begin{aligned} M_Y(t) &= \frac{\nu p}{\nu-t} \frac{1}{1 - \frac{\nu}{\nu-t}(1-p)} \\ &= \frac{\nu p}{\nu p - t}. \end{aligned}$$

Cela correspond bien à la fonction génératrice des moments d'une variable exponentielle de paramètre νp . ■

Proposition 5.2. *Dans les conditions de la définition 5.1, les processus $\{N_j(t), t \geq 0\}$ sont des processus de Poisson d'intensité νp_j pour tout $j = 1, 2, \dots, r$. De plus, ils sont indépendants les uns des autres.*

Démonstration. Soit $j \in \{1, 2, \dots, r\}$, montrons que les temps d'inter-arrivées du processus $\{N_j(t), t \geq 0\}$ sont indépendants et identiquement distribués selon une loi exponentielle de paramètre νp_j . Pour cela, notons W_1^j le temps d'attente pour le premier événement de type j . Il peut s'écrire sous la forme suivante :

$$W_1^j = \sum_{k=1}^G W_k$$

où W_1, W_2, \dots sont les temps d'inter-arrivées du processus $\{N(t), t \geq 0\}$ et G est une variable aléatoire qui dénote le nombre d'événements à observer avant d'obtenir le premier événement

²cfr J.SCHMETS, *ANALYSE MATHÉMATIQUE*, Liège, Éditions Derouaux, 1996.

de type j . La variable aléatoire G est donc une variable géométrique de paramètre p_j ; vu le lemme 5.1, $W_1^j \sim \text{Exp}(\nu p_j)$.

Soit alors $n > 1$. De la même manière, il est possible d'écrire le n -ième temps d'inter-arrivées du processus $\{N_j(t), t \geq 0\}$ en fonction des temps d'inter-arrivées du processus $\{N(t), t \geq 0\}$,

$$W_n^j = \sum_{k=1}^G W_{n,k}$$

où $W_{n,1}$ est le temps d'attente entre le $(n-1)$ -ième événement de type j et l'événement suivant du processus $\{N(t), t \geq 0\}$ (c'est donc un temps d'inter-arrivées de ce processus), les autres $W_{n,k}$ sont les temps d'inter-arrivées suivants de $\{N(t), t \geq 0\}$ et G est une variable aléatoire qui dénote le nombre d'événements à observer avant d'obtenir un nouvel événement de type j . Le lemme précédent donne à nouveau que $W_n^j \sim \text{Exp}(\nu p_j)$.

Comme les sommes qui définissent les W_n^j portent sur des temps d'inter-arrivées W_n différents, elles sont indépendantes les unes des autres. Vu la proposition 2.4, le processus $\{N_j(t), t \geq 0\}$ est un processus de Poisson d'intensité νp_j , pour tout $j \in \{1, 2, \dots, r\}$.

Pour ce qui est de l'indépendance de ces processus, procédons par récurrence sur r .

Cas de base $r = 2$:

Pour tout $t \geq 0$, calculons la distribution conjointe de $N_1(t)$ et $N_2(t)$.

$$\mathbb{P}[N_1(t) = n_1 \text{ et } N_2(t) = n_2] = \sum_{k=0}^{+\infty} \mathbb{P}[N_1(t) = n_1 \text{ et } N_2(t) = n_2 | N(t) = k] \mathbb{P}[N(t) = k].$$

Comme $N(t) = N_1(t) + N_2(t)$, il ne reste que le terme $k = n_1 + n_2$ et il vient

$$\mathbb{P}[N_1(t) = n_1 \text{ et } N_2(t) = n_2] = \mathbb{P}[N_1(t) = n_1 \text{ et } N_2(t) = n_2 | N(t) = n_1 + n_2] \mathbb{P}[N(t) = n_1 + n_2].$$

Le premier facteur correspond à la probabilité d'avoir n_1 succès (événement de type 1) parmi $n_1 + n_2$ épreuves indépendantes pour lesquelles le succès s'obtient avec une probabilité p_1 . On obtient donc³

$$\begin{aligned} \mathbb{P}[N_1(t) = n_1 \text{ et } N_2(t) = n_2] &= C_{n_1+n_2}^{n_1} p_1^{n_1} (1-p_1)^{n_2} e^{-\nu t} \frac{(\nu t)^{n_1+n_2}}{(n_1+n_2)!} \\ &= e^{-\nu p_1 t} \frac{(\nu p_1 t)^{n_1}}{n_1!} e^{-\nu(1-p_1)t} \frac{(\nu(1-p_1)t)^{n_2}}{n_2!} \end{aligned}$$

et, comme $p_1 + p_2 = 1$, on a finalement

$$\mathbb{P}[N_1(t) = n_1 \text{ et } N_2(t) = n_2] = \mathbb{P}[N_1(t) = n_1] \mathbb{P}[N_2(t) = n_2].$$

Ainsi, les processus $\{N_1(t), t \geq 0\}$ et $\{N_2(t), t \geq 0\}$ sont indépendants.

³La variable aléatoire X qui compte le nombre de succès obtenus lors de n essais indépendants d'une expérience aléatoire n'ayant que deux issues possibles, le succès avec une probabilité p et l'échec avec une probabilité $1-p$, est appelée variable binomiale et est notée $X \sim B(n, p)$. Pour tout $k \in \mathbb{N}$, $\mathbb{P}[X = k] = C_n^k p^k (1-p)^{n-k}$.

Récurrance :

Considérons les deux types d'événements A et B où A rassemble les événements de type j et B rassemble tous les autres types. Vu le cas de base, les processus $\{N_A(t), t \geq 0\}$ et $\{N_B(t), t \geq 0\}$ sont indépendants. On recommence à amincir le processus $\{N_B(t), t \geq 0\}$ en considérant deux types d'événements C et D où C rassemble les événements de type i et D tous les autres. Par hypothèse de récurrence, $\{N_C(t), t \geq 0\}$ et $\{N_D(t), t \geq 0\}$ sont indépendants entre eux et, par le cas précédent, ils sont indépendants de $\{N_A(t), t \geq 0\}$. En continuant ainsi, on obtient l'indépendance des r processus. ■

Pour illustrer ce phénomène, revenons sur l'exemple de processus de Poisson composé 4.1. Le processus composé $\{X(t), t \geq 0\}$ est un amincissement du processus de Poisson sous-jacent $\{N(t), t \geq 0\}$. En effet, le n -ième événement peut être de type 0 ou 1 selon que la variable Y_n égale 0 ou 1 et on ne considère que les événements de type 1.

Les figures 5.1, 5.2 et 5.3 représentent une trajectoire du processus de Poisson $\{N(t), t \geq 0\}$ d'intensité $\nu = 2$ et les deux amincissements correspondants de celui-ci avec le vecteur (p_0, p_1) égal à $(0.6, 0.4)$.

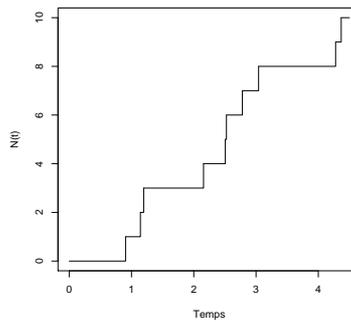


FIG. 5.1 – Processus $\{N(t), t \geq 0\}$.

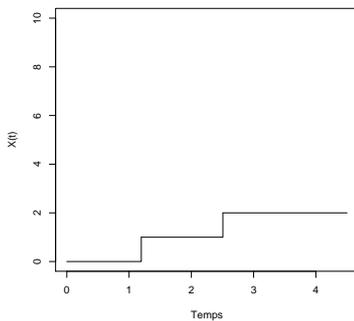


FIG. 5.2 – Processus $\{N_1(t), t \geq 0\}$

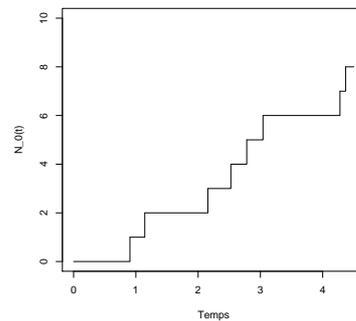


FIG. 5.3 – Processus $\{N_0(t), t \geq 0\}$

Dans les ouvrages consultés, rien n'est dit concernant l'amincissement d'un processus de Poisson non homogène ou composé. La question de savoir ce qu'il se passe alors était donc

ouverte. Pour les processus de Poisson non homogènes, il est possible d'adapter la définition 5.1 et la proposition 5.2.

Définition 5.2. *Considérons un processus de Poisson $\{N(t), t \geq 0\}$ non homogène de fonction moyenne $m(t)$. Soient un naturel r et (p_1, p_2, \dots, p_r) un vecteur de probabilités tel que chaque événement du processus a une probabilité p_j d'être de type j , pour tout $j = 1, 2, \dots, r$. Supposons en outre que les attributions des types sont indépendantes les unes des autres. Pour chaque $j \in \{1, 2, \dots, r\}$, le processus de comptage des événements de type j , $\{N_j(t), t \geq 0\}$, est appelé **amincissement du processus** $\{N(t), t \geq 0\}$.*

Proposition 5.3. *Dans les conditions de la définition 5.2, les processus $\{N_j(t), t \geq 0\}$ sont des processus de Poisson non homogènes de fonction moyenne $m(t)p_j$ pour tout $j = 1, 2, \dots, r$. De plus, ils sont indépendants les uns des autres.*

Démonstration. Pour avoir un processus de Poisson non homogène, il suffit de vérifier les conditions de la proposition 3.3.

Tout d'abord, $N_j(0) = 0$ pour tout $j = 1, 2, \dots, r$ car $N(0) = 0$.

Avec les mêmes notations que dans la proposition 5.2, il est possible d'écrire les temps d'inter-arrivées du processus $\{N_j(t), t \geq 0\}$ sous la forme

$$\begin{cases} W_1^j = \sum_{k=1}^G W_k \\ W_n^j = \sum_{k=1}^G W_{n,k} \quad \forall n > 1. \end{cases}$$

Comme les temps d'inter-arrivées du processus $\{N(t), t \geq 0\}$ sont indépendants entre eux et que les sommes portent sur des temps d'inter-arrivées différents, les variables W_n^j , $n \geq 1$, sont indépendantes. Pour démontrer l'indépendance des accroissements, la notation suivante, illustrée par la figure 5.4, sera nécessaire⁴ : $W_i^{j,s}$ est le temps d'attente du i -ième événement de type j se produisant après s . Si le i -ième événement de type j après s est le n -ième de type j , il vient

$$\begin{cases} W_1^{j,s} = W_n^j - \left(s - \sum_{l=1}^{n-1} W_l^j \right) \\ W_i^{j,s} = W_n^j \quad \forall i > 1. \end{cases}$$

Cette réécriture montre que, pour tout j et s fixés, les variables $W_i^{j,s}$ sont indépendantes. Soient alors $0 \leq t_1 < t_2 < t_3$ et $k, k' \in \mathbb{N}$.

$$\begin{aligned} & \mathbb{P}[N_j(t_3) - N_j(t_2) = k \text{ et } N_j(t_2) - N_j(t_1) = k'] \\ &= \mathbb{P}[\{N_j(t_3) - N_j(t_2) \geq k\} \setminus \{N_j(t_3) - N_j(t_2) \geq k+1\}] \text{ et} \\ & \quad (\{N_j(t_2) - N_j(t_1) \geq k'\} \setminus \{N_j(t_2) - N_j(t_1) \geq k'+1\})] \\ &= \mathbb{P} \left[\left(\left\{ \sum_{i=1}^k W_i^{j,t_2} \leq t_3 - t_2 \right\} \setminus \left\{ \sum_{i=1}^{k+1} W_i^{j,t_2} \leq t_3 - t_2 \right\} \right) \text{ et} \right. \\ & \quad \left. \left(\left\{ \sum_{i=1}^{k'} W_i^{j,t_1} \leq t_2 - t_1 \right\} \setminus \left\{ \sum_{i=1}^{k'+1} W_i^{j,t_1} \leq t_2 - t_1 \right\} \right) \right]. \end{aligned}$$

⁴Une notation semblable avait déjà été introduite lors de la preuve de la proposition 2.4.

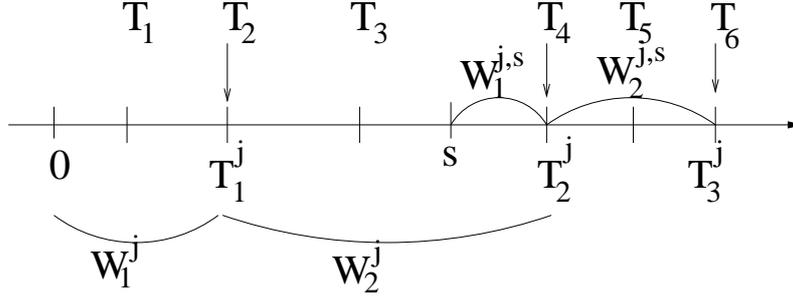


FIG. 5.4 – Nouvelle notation

Comme les sommes de temps d'inter-arrivées portent sur des inter-arrivées différentes, elles sont indépendantes et la factorisation des probabilités est possible, ce qui donne

$$\begin{aligned}
 & \mathbb{P}[N_j(t_3) - N_j(t_2) = n \text{ et } N_j(t_2) - N_j(t_1) = m] \\
 &= \mathbb{P} \left[\left\{ \sum_{i=1}^n W_i^{j,t_2} \leq t_3 - t_2 \right\} \setminus \left\{ \sum_{i=1}^{n+1} W_i^{j,t_2} \leq t_3 - t_2 \right\} \right] \\
 & \quad \mathbb{P} \left[\left\{ \sum_{i=1}^m W_i^{j,t_1} \leq t_2 - t_1 \right\} \setminus \left\{ \sum_{i=1}^{m+1} W_i^{j,t_1} \leq t_2 - t_1 \right\} \right] \\
 &= \mathbb{P}[\{N_j(t_3) - N_j(t_2) \geq n\} \setminus \{N_j(t_3) - N_j(t_2) \geq n + 1\}] \\
 & \quad \mathbb{P}[\{N_j(t_2) - N_j(t_1) \geq m\} \setminus \{N_j(t_2) - N_j(t_1) \geq m + 1\}] \\
 &= \mathbb{P}[N_j(t_3) - N_j(t_2) = n] \mathbb{P}[N_j(t_2) - N_j(t_1) = m].
 \end{aligned}$$

Le même raisonnement peut être tenu si l'on considère plus de deux accroissements. Il reste donc à montrer que pour tout $0 \leq s < t$, et pour tout $k \in \mathbb{N}$,

$$\mathbb{P}[N_j(t) - N_j(s) = k] = e^{-(m(t)-m(s))p_j} \frac{((m(t) - m(s))p_j)^k}{k!}.$$

Sachant que n événements ont eu lieu dans l'intervalle $[s, t]$, la probabilité d'en avoir eu k de type j est donnée par $C_n^k p_j^k (1 - p_j)^{n-k}$. Cela étant,

$$\begin{aligned}
 \mathbb{P}[N_j(t) - N_j(s) = k] &= \sum_{n=0}^{+\infty} \mathbb{P}[N_j(t) - N_j(s) = k | N(t) - N(s) = n] \mathbb{P}[N(t) - N(s) = n] \\
 &= \sum_{n=k}^{+\infty} C_n^k p_j^k (1 - p_j)^{n-k} e^{-(m(t)-m(s))} \frac{(m(t) - m(s))^n}{n!} \\
 &= e^{-(m(t)-m(s))} \frac{p_j^k (m(t) - m(s))^k}{k!} \sum_{n=k}^{+\infty} \frac{(1 - p_j)^{n-k} (m(t) - m(s))^{n-k}}{(n - k)!} \\
 &= e^{-(m(t)-m(s))} \frac{((m(t) - m(s))p_j)^k}{k!} e^{(1-p_j)(m(t)-m(s))} \\
 &= e^{-(m(t)-m(s))p_j} \frac{((m(t) - m(s))p_j)^k}{k!}.
 \end{aligned}$$

Pour ce qui est de l'indépendance entre les processus amincis, on procède par récurrence sur r exactement comme cela a été fait dans la preuve de la proposition 5.2 mis à part que νt doit être remplacé par $m(t)$ dans le développement. ■

Si l'on souhaite amincir un processus de Poisson composé, deux façons de faire sont envisageables. La première consiste à considérer les grappes entières d'événements comme étant d'un certain type. Cela nous ramène dans le cas d'un amincissement d'un processus de base et il est donc clair que le processus aminci est encore un processus de Poisson. Si maintenant, nous voulons amincir plus en profondeur, c-à-d que dans une même grappe, les événements peuvent être de types différents, alors le problème est plus difficile et le temps disponible n'a pas été suffisant pour le résoudre. Dans ce cas, il est déjà possible de dire que le processus sous-jacent au processus aminci sera toujours un processus de Poisson mais dont l'intensité dépendra peut-être de la distribution de la famille de variables aléatoires $\{Y_n, n \in \mathbb{N}\}$, ce qui devient gênant pour l'indépendance du processus sous-jacent avec la nouvelle famille $\{Y_n^j, n \in \mathbb{N}\}$.

Avant de passer à la notion de superposition de processus de Poisson, voici un exemple de processus de Poisson aminci. Si l'hypothèse que les arrivées de clients dans un magasin suivent le modèle d'un processus de Poisson de base ou non homogène est valable, alors le processus d'arrivée des hommes dans ce magasin est un processus de Poisson aminci du processus des arrivées des clients. Il en est évidemment de même pour les arrivées des dames.

5.2 Superposition de processus de Poisson

Définition 5.3. Soient, pour tout $j = 1, 2, \dots, r$, $\{N_j(t), t \geq 0\}$ des processus de Poisson d'intensité ν_j qui sont indépendants les uns des autres. Pour tout $t \geq 0$, posons

$$N(t) = N_1(t) + N_2(t) + \dots + N_r(t).$$

Le processus $\{N(t), t \geq 0\}$ est appelé **superposition des processus** $\{N_j(t), t \geq 0\}$, $j = 1, 2, \dots, r$.

Proposition 5.4. Dans les conditions de la définition 5.3, le processus $\{N(t), t \geq 0\}$ est un processus de Poisson d'intensité $\nu = \sum_{i=1}^r \nu_i$.

Démonstration. La première chose dont il faut s'assurer est que le processus superposé n'est pas composé, c-à-d qu'en un temps t quelconque il n'est pas possible d'avoir une occurrence pour plusieurs des processus $\{N_j(t), t \geq 0\}$. Cette propriété s'appelle "disjonction" des processus. Or,

$$\mathbb{P}[\exists i \neq j \text{ et } n, n' : T_n^i = T_{n'}^j] = \mathbb{P}[\exists i \neq j \text{ et } n, n' : T_n^i - T_{n'}^j = 0] = 0 \quad (5.1)$$

vu le caractère continu des variables aléatoires caractérisant des temps d'occurrence. Passons maintenant à la vérification des différentes propriétés des processus de Poisson. Pour cela, procédons par récurrence sur r .

Cas de base $r = 2$:

Commençons par montrer que les accroissements sont bien indépendants. Pour cela, prenons $0 \leq t_1 < t_2 < t_3$. Par définition de $N(t)$, on a

$$N(t_3) - N(t_2) = N_1(t_3) - N_1(t_2) + N_2(t_3) - N_2(t_2)$$

et

$$N(t_2) - N(t_1) = N_1(t_2) - N_1(t_1) + N_2(t_2) - N_2(t_1).$$

Par indépendance des accroissements de $\{N_1(t), t \geq 0\}$ et $\{N_2(t), t \geq 0\}$ et par indépendance de ces deux processus entre eux, les deux accroissements considérés sont indépendants.

Ensuite, il faut montrer que les accroissements sont stationnaires. Soient $s, t > 0$. De nouveau par définition de $N(t)$, il vient

$$N(t+s) - N(s) = N_1(t+s) - N_1(s) + N_2(t+s) - N_2(s) \sim N_1(t) + N_2(t) = N(t)$$

par stationnarité des accroissements des processus $\{N_1(t), t \geq 0\}$ et $\{N_2(t), t \geq 0\}$.

Il ne reste plus qu'à montrer que $N(t) \sim \mathcal{P}(\nu t), \forall t > 0$. Cela se fait en utilisant les fonctions génératrices des probabilités. La définition donne

$$G_{N(t)}(u) = E[u^{N(t)}] = E[u^{N_1(t)} u^{N_2(t)}].$$

Par l'indépendance des processus, il vient

$$G_{N(t)}(u) = E[u^{N_1(t)}] E[u^{N_2(t)}].$$

En vertu du caractère poissonien des deux processus et de la proposition 1.10, on obtient

$$G_{N(t)}(u) = e^{\nu_1 t(u-1)} e^{\nu_2 t(u-1)} = e^{\nu t(u-1)}$$

qui est la fonction génératrice des probabilités d'une variable aléatoire discrète suivant une loi de Poisson de paramètre $\nu t = (\nu_1 + \nu_2)t$.

Ainsi, $\{N(t), t \geq 0\}$ est un processus de Poisson d'intensité ν qui, par hypothèse d'indépendance des processus $\{N_j(t), t \geq 0\}, j = 1, 2, \dots, r$, est indépendant des processus $\{N_j(t), t \geq 0\}$ pour $j = 3, \dots, r$.

Réurrence :

Posons

$$N_A(t) = N_1(t) + N_2(t) + \dots + N_{r-1}(t), \forall t \geq 0,$$

et supposons, par hypothèse de récurrence, que le processus $\{N_A(t), t \geq 0\}$ soit un processus de Poisson d'intensité $\nu_A = \nu_1 + \nu_2 + \dots + \nu_{r-1}$ et qu'il soit indépendant du processus $\{N_r(t); t \geq 0\}$ d'intensité ν_r . En appliquant le cas de base à ces deux processus, nous obtenons que $\{N(t), t \geq 0\}$ est un processus de Poisson d'intensité $\nu = \nu_1 + \nu_2 + \dots + \nu_r$. ■

Il existe des preuves plus formelles (et en dehors du sujet de ce travail) de la disjonction des processus que celle donnée ici ; mais la relation 5.1 permet déjà de bien faire sentir ce fait. En pratique, la condition de disjonction des processus peut ne pas être réalisée puisque le manque de précision des observations implique souvent que les variables continues apparaissent discrètes.

Avec les mêmes notations, la probabilité qu'un événement du processus superposé $\{N(t), t \geq 0\}$ soit un événement du processus $\{N_i(t), t \geq 0\}$ est donnée par $\frac{\nu_i}{\nu}$ indépendamment du temps auquel a lieu cet événement. En effet, un événement a lieu en t si $N(t)$ est plus grand d'une

unité que $N(t - h)$ pour h suffisamment petit ; cela étant, pour $t \geq 0$, il est possible d'écrire la probabilité recherchée sous la forme

$$\begin{aligned} & \lim_{h \rightarrow 0} \mathbb{P}[N_1(t) - N_1(t - h) = 0, \dots, N_{i-1}(t) - N_{i-1}(t - h) = 0, N_i(t) - N_i(t - h) = 1, \\ & \quad N_{i+1}(t) - N_{i+1}(t - h) = 0, \dots, N_r(t) - N_r(t - h) = 0 | N(t) - N(t - h) = 1] \\ &= \lim_{h \rightarrow 0} \left(\frac{1}{\mathbb{P}[N(t) - N(t - h) = 1]} \mathbb{P}[N_1(t) - N_1(t - h) = 0, \dots, N_{i-1}(t) - N_{i-1}(t - h) = 0, \right. \\ & \quad \left. N_i(t) - N_i(t - h) = 1, N_{i+1}(t) - N_{i+1}(t - h) = 0, \dots, \right. \\ & \quad \left. N_r(t) - N_r(t - h) = 0, N(t) - N(t - h) = 1] \right) \\ &= \lim_{h \rightarrow 0} \left(\frac{1}{\mathbb{P}[N(t) - N(t - h) = 1]} \mathbb{P}[N_1(t) - N_1(t - h) = 0, \dots, N_{i-1}(t) - N_{i-1}(t - h) = 0, \right. \\ & \quad \left. N_i(t) - N_i(t - h) = 1, N_{i+1}(t) - N_{i+1}(t - h) = 0, \dots, N_r(t) - N_r(t - h) = 0] \right) \end{aligned}$$

car les événements sont équivalents. On utilise alors l'indépendance des processus, leur stationnarité et leur caractère poissonien

$$\begin{aligned} &= \lim_{h \rightarrow 0} \frac{e^{-\nu_1 h} e^{-\nu_2 h} \dots e^{-\nu_{i-1} h} e^{-\nu_i h} \nu_i h e^{-\nu_{i+1} h} \dots e^{-\nu_r h}}{e^{-\nu h} \nu h} \\ &= \frac{\nu_i}{\nu} \end{aligned}$$

car $\nu = \nu_1 + \nu_2 + \dots + \nu_r$.

Illustrons ce phénomène avec quelques graphiques. Les figures 5.5 et 5.6 représentent deux trajectoires des processus $\{N_A(t), t \geq 0\}$ et $\{N_B(t), t \geq 0\}$ d'intensité respective $\nu_1 = 2$ et $\nu_2 = 10$. Le graphique 5.7 est celui de la trajectoire correspondante du processus superposé $\{N(t), t \geq 0\}$ d'intensité $\nu = 12$.

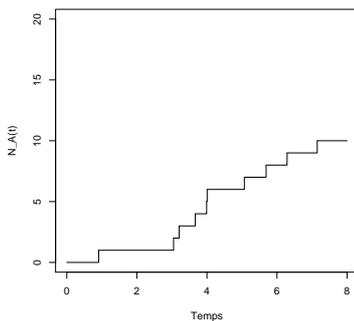


FIG. 5.5 – Processus $\{N_A(t), t \geq 0\}$ d'intensité $\nu = 2$

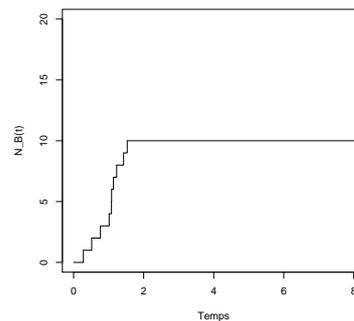


FIG. 5.6 – Processus $\{N_B(t), t \geq 0\}$ d'intensité $\nu = 10$

Comme pour l'amincissement, malgré que cela ne figure pas dans les livres consultés, la notion de superposition peut être étendue aux processus de Poisson non homogènes.

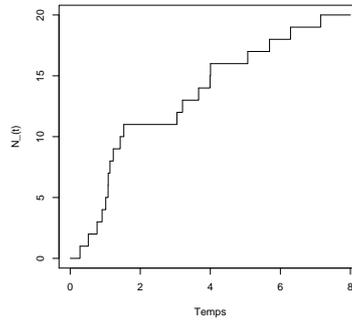


FIG. 5.7 – Processus $\{N(t), t \geq 0\}$

Définition 5.4. Soient, pour tout $j = 1, 2, \dots, r$, $\{N_j(t), t \geq 0\}$ des processus de Poisson de fonction moyenne $m_j(t)$ qui sont indépendants les uns des autres. Pour tout $t \geq 0$, posons $N(t) = N_1(t) + N_2(t) + \dots + N_r(t)$.

Le processus $\{N(t), t \geq 0\}$ est appelé **superposition des processus** $\{N_j(t), t \geq 0\}$, $j = 1, 2, \dots, r$.

Proposition 5.5. Dans les conditions de la définition 5.4, le processus $\{N(t), t \geq 0\}$ est un processus de Poisson de fonction moyenne $m(t) = m_1(t) + m_2(t) + \dots + m_r(t)$.

Démonstration. Tout d’abord, la propriété de disjonction reste valable dans ce cas. Vérifions alors les conditions de la proposition 3.3. La preuve se fait à nouveau par récurrence sur r et seul le cas de base sera refait ici car l’argument de récurrence est le même que dans la proposition 5.4.

Premièrement, l’indépendance des accroissements se fait encore de la même façon que dans la proposition 5.4.

Deuxièmement, on a bien $N(0) = 0$ car $N_j(0) = 0 \forall j \in \{1, 2\}$. Il ne reste donc plus qu’à vérifier que $\forall 0 \leq s < t$ et $\forall k \in \mathbb{N}$,

$$\mathbb{P}[N(t) - N(s) = k] = e^{-(m(t)-m(s))} \frac{(m(t) - m(s))^k}{k!}.$$

Pour démontrer cela, prenons donc $0 \leq s < t$ et $k \in \mathbb{N}$.

$$\begin{aligned} \mathbb{P}[N(t) - N(s) = k] &= \mathbb{P}[N_1(t) - N_1(s) + N_2(t) - N_2(s) = k] \\ &= \sum_{n=0}^{+\infty} \mathbb{P}[N_1(t) - N_1(s) + N_2(t) - N_2(s) = k \text{ et } N_1(t) - N_1(s) = n] \end{aligned}$$

par la formule des probabilités totales. Comme ces probabilités sont nulles pour $n > k$, on a

$$\begin{aligned} \mathbb{P}[N(t) - N(s) = k] &= \sum_{n=0}^k \mathbb{P}[N_2(t) - N_2(s) = k - n \text{ et } N_1(t) - N_1(s) = n] \\ &= \sum_{n=0}^k \mathbb{P}[N_2(t) - N_2(s) = k - n] \mathbb{P}[N_1(t) - N_1(s) = n] \end{aligned}$$

par indépendance des processus entre eux.

$$\begin{aligned} \mathbb{P}[N(t) - N(s) = k] &= \sum_{n=0}^k e^{-(m_2(t)-m_2(s))} \frac{(m_2(t) - m_2(s))^{k-n}}{(k-n)!} e^{-(m_1(t)-m_1(s))} \frac{(m_1(t) - m_1(s))^n}{n!} \\ &= e^{-(m(t)-m(s))} \frac{1}{k!} \sum_{n=0}^k C_k^n (m_2(t) - m_2(s))^{k-n} (m_1(t) - m_1(s))^n \end{aligned}$$

ce qui donne en utilisant la formule du binôme de Newton⁵

$$= e^{-(m(t)-m(s))} \frac{(m(t) - m(s))^k}{k!}.$$

Par la proposition 3.3, $\{N(t), t \geq 0\}$ est un processus de Poisson de fonction moyenne $m(t) = m_1(t) + m_2(t)$. Comme il n'est défini qu'à partir des processus $\{N_1(t), t \geq 0\}$ et $\{N_2(t), t \geq 0\}$, il est indépendant des processus $\{N_j(t), t \geq 0\}$ pour $j = 3, \dots, l$. ■

Comme dans le cas de processus de base, la probabilité qu'un événement du processus superposé soit de type i peut être calculée. Le raisonnement est le même sauf que l'argument de stationnarité ne s'applique plus, ce qui fait que l'on obtient

$$\begin{aligned} &\lim_{h \rightarrow 0} \mathbb{P}[N_1(t) - N_1(t-h) = 0, \dots, N_{i-1}(t) - N_{i-1}(t-h) = 0, N_i(t) - N_i(t-h) = 1, \\ &\quad N_{i+1}(t) - N_{i+1}(t-h) = 0, \dots, N_r(t) - N_r(t-h) = 0 | N(t) - N(t-h) = 1] \\ &= \lim_{h \rightarrow 0} \left(\frac{1}{e^{-(m(t)-m(t-h))} (m(t) - m(t-h))} e^{-(m_1(t)-m_1(t-h))} \dots e^{-(m_{i-1}(t)-m_{i-1}(t-h))} \right. \\ &\quad \left. e^{-(m_i(t)-m_i(t-h))} (m_i(t) - m_i(t-h)) e^{-(m_{i+1}(t)-m_{i+1}(t-h))} \dots e^{-(m_r(t)-m_r(t-h))} \right) \\ &= \lim_{h \rightarrow 0} \frac{m_i(t) - m_i(t-h)}{m(t) - m(t-h)}. \end{aligned}$$

Vu la continuité des fonctions moyennes, c'est un cas d'application du théorème de l'Hospital qui donne

$$= \lim_{h \rightarrow 0} \frac{\nu_i(t-h)}{\nu(t-h)} = \frac{\nu_i(t)}{\nu(t)}$$

par continuité des fonctions d'intensité.

Voici ce qu'il se passe pour des processus de Poisson composés.

Définition 5.5. Soient $\{X_j(t), t \geq 0\}$ des processus de Poisson composés pour $j \in \{1, 2, \dots, r\}$ indépendants. Notons $\{N_j(t), t \geq 0\}$ les processus de Poisson sous-jacents (d'intensité ν_j) et $\{Y_n^j, n \in \mathbb{N}\}$ les familles de variables aléatoires associées définies comme dans la définition 4.1. Le processus $\{X(t), t \geq 0\}$ défini par $X(t) = X_1(t) + X_2(t) + \dots + X_r(t)$, pour $t \geq 0$, est appelé **superposition des processus** $\{X_j(t), t \geq 0\}$, $j = 1, 2, \dots, r$.

Proposition 5.6. La superposition de processus de Poisson composés indépendants est encore un processus de Poisson composé.

⁵cfr J.SCHMETS, ANALYSE MATHÉMATIQUE, Liège, Éditions Derouaux, 1996.

Démonstration. Notons d'abord que la propriété de disjonction est encore valable pour le processus sous-jacent $\{N(t), t \geq 0\}$, avec $N(t) = N_1(t) + \dots + N_r(t)$, qui est donc encore un processus de Poisson d'intensité $\nu = \sum_{j=1}^r \nu_j$. De plus, $\forall t \geq 0$,

$$\begin{aligned} X(t) &= X_1(t) + X_2(t) + \dots + X_r(t) \\ &= \sum_{n_1=0}^{N_1(t)} Y_{n_1}^1 + \sum_{n_2=0}^{N_2(t)} Y_{n_2}^2 + \dots + \sum_{n_r=0}^{N_r(t)} Y_{n_r}^r \\ &= \sum_{n=0}^{N(t)} Y_n \end{aligned}$$

où $Y_n = Y_{n_j}^j$ si la n -ième occurrence du processus superposé $\{N(t), t \geq 0\}$ est la n_j -ième du processus $\{N_j(t), t \geq 0\}$.

La famille $\{Y_n, n \in \mathbb{N}\}$ associée au processus superposé $\{X(t), t \geq 0\}$ est donc un mélange des familles $\{Y_n^j, n \in \mathbb{N}\}$. Comme, pour tout j fixé, les variables Y_n^j sont indépendantes pour tout $n \in \mathbb{N}$ et que les processus sont indépendants entre eux, les variables Y_n sont bien indépendantes entre elles. De plus, pour tout j , chaque élément de la famille $\{Y_n, n \in \mathbb{N}\}$ est indépendant des variables $N_j(t)$ et cela pour tout $t \geq 0$, que ce soit par indépendance des processus entre eux si $Y_n = Y_{n_i}^i$ avec $i \neq j$ ou par hypothèse sur $\{X_j(t), t \geq 0\}$ si $Y_n = Y_{n_j}^j$. Ainsi, chaque élément de cette famille est indépendant de $N(t) = N_1(t) + N_2(t) + \dots + N_r(t)$ pour tout $t \geq 0$. ■

Pour illustrer la superposition de processus de Poisson, voici un exemple concret. Si l'on suppose que les processus d'ouverture des tulipes d'une certaine couleur dans un jardin sont des processus de Poisson de même type (tous stationnaires ou tous non-stationnaires ou encore tous composés) indépendants les uns des autres, alors le processus d'ouverture de toutes les tulipes est un processus de Poisson superposé.

Chapitre 6

Processus de Poisson en pratique

Ce chapitre est consacré à une vision plus appliquée des processus de Poisson étudiés dans les chapitres précédents. Pour chacun d'entre eux, la partie "simulation" sera consacrée à la génération de tels processus, tandis que la partie "test" expliquera comment admettre qu'un ensemble de données possède le caractère poissonien. Les ensembles de données à notre disposition seront soit les temps d'inter-arrivées, soit les temps d'occurrence d'un processus ; il faudra alors voir s'ils correspondent à des temps d'inter-arrivées ou d'occurrence relatifs à un processus de Poisson.

6.1 Processus de base

6.1.1 Simulation

Dans ce cas, la simulation a déjà été abordée. En effet, comme cela a été expliqué au chapitre 2, il suffit de générer des valeurs $\{w_1, \dots, w_n\}$ de variables exponentiellement distribuées. Ce sont les inter-arrivées du processus et au moyen de la formule (2.3), on trouve les temps d'occurrence $\{t_1, t_2, \dots, t_n\}$ correspondants. Le paramètre utilisé pour la distribution exponentielle est le taux d'occurrence du processus que l'on désire obtenir. Par exemple, pour un processus dont le nombre moyen d'arrivées par unité de temps est 5, il faut utiliser une exponentielle de paramètre 5.

Plusieurs méthodes de simulation de variables exponentielles sont à notre disposition. Le logiciel R peut bien sûr le faire via la commande `rexp(x, nu)`, tout comme tout autre logiciel statistique. Excel possède aussi un moyen de simuler de telles variables. Dans l'outil "Utilitaire d'analyse", il suffit de générer des valeurs $\{u_1, \dots, u_n\}$ de variables uniformes sur $[0, 1]$ et de leur appliquer la transformation

$$w_n = \frac{-1}{\nu} \ln(1 - u_n)$$

pour obtenir un échantillon $\{w_1, \dots, w_n\}$ de variables exponentielles de paramètre ν . Cette méthode se base sur la proposition suivante :

Proposition 6.1. *Soient F une fonction de répartition continue et U une variable aléatoire uniforme sur l'intervalle $[0, 1]$. Si l'on pose $F^{-1}(y) = \inf\{x \in \mathbb{R}^+ : F(x) \geq y\}$ pour tout $y \in [0, 1]$, alors $X = F^{-1}(U)$ est une variable aléatoire de fonction de répartition F .*

Démonstration. Par le même raisonnement que dans la preuve de la proposition 3.8, on montre que $F(F^{-1}(y)) = y, \forall 0 \leq y \leq 1$. Ensuite, il vient

$$\mathbb{P}[X \leq x] = \mathbb{P}[F^{-1}(U) \leq x] = \mathbb{P}[F(F^{-1}(U)) \leq F(x)] = \mathbb{P}[U \leq F(x)] = F(x)$$

par définition d'une variable uniforme sur $[0, 1]$. ■

6.1.2 Tests

Comme la distribution des temps d'inter-arrivées est bien connue dans le cadre d'un processus de Poisson de base, il est possible de s'en servir pour vérifier si un ensemble de temps constitue les temps d'inter-arrivées d'un processus de Poisson. En effet, il a déjà été démontré (cfr. proposition 2.4) que le processus est un processus de Poisson de base d'intensité ν si et seulement si les temps d'inter-arrivées sont indépendants et identiquement distribués selon une loi exponentielle de paramètre ν , c-à-d $W_i \stackrel{\text{iid}}{\sim} \text{Exp}(\nu)$. A partir des temps d'inter-arrivées observés $\{w_1, w_2, \dots, w_n\}$, il suffit donc de tester s'ils constituent un échantillon aléatoire et simple provenant d'une exponentielle de paramètre ν .

Pour cela, le test d'ajustement de Kolmogorov-Smirnov peut être exploité.

$$H_0 : F = \text{Exp}(\nu) \quad \text{vs} \quad H_1 : F \neq \text{Exp}(\nu).$$

Dans le logiciel R, la commande pour réaliser un test de Kolmogorov-Smirnov est `ks.test()`. Attention, il ne faut pas tenir compte de la p-value donnée par R lorsque le paramètre ν a été estimé sur l'échantillon à tester. Dans ce cas, il faut utiliser les seuils critiques présentés à la section 1.3 et repris à la page 96.

Une analyse graphique permet d'obtenir de l'information sur le caractère exponentiel ou non des temps d'inter-arrivées. Il s'agit simplement d'exploiter l'intuition sur laquelle le test de Kolmogorov-Smirnov est basée. Si F_n représente la distribution empirique et si la distribution F est supposée valide, on doit avoir $F_n(x) \approx F(x)$ pour tout $x \in \mathbb{R}$ et en particulier en toutes valeurs w_1, \dots, w_n observées. C'est-à-dire que

$$\begin{aligned} \frac{i}{n} &\approx 1 - e^{-\nu w_{(i)}} \\ \Leftrightarrow w_{(i)} &\approx \frac{-1}{\nu} \ln \left(1 - \frac{i}{n} \right) \\ \Leftrightarrow \left\{ \left(w_{(i)}, -\ln \left(1 - \frac{i - \frac{1}{2}}{n} \right) \right), 1 \leq i \leq n \right\} &\text{ forment un nuage de points qui s'ajuste plus ou} \\ &\text{moins à une droite de pente } \nu. \end{aligned}$$

Le terme $\frac{1}{2}$ est ajouté pour empêcher d'avoir "ln(0)" lorsque $i = n$. Si les points du graphe ainsi construit sont plus ou moins alignés, on peut supposer que le processus est bien de type poissonien et on peut estimer le paramètre ν grâce à la pente de la droite. Ce type de graphique porte le nom de "quantile-quantile plot", ou plus simplement "qq-plot", car il compare les quantiles d'un échantillon de données à des quantiles théoriques.

Il est également possible de tester qu'un ensemble de données forme un processus de Poisson en utilisant la proposition 2.7. Soit le vecteur (t_1, t_2, \dots, t_n) des temps d'occurrence d'un processus observé sur $[0, T]$, il faut alors lui appliquer le test de Kolmogorov-Smirnov d'hypothèses

$$H_0 : F = U[0, T] \quad \text{vs} \quad H_1 : F \neq U[0, T].$$

Cependant, une étude comparative des puissances de ces deux tests montre que celui basé sur les temps d'inter-arrivées est plus puissant quelque soit la vraie distribution des observations et quelque soit la taille de l'échantillon. C'est pourquoi il sera toujours privilégié dans la suite.

6.1.3 Intervalle de confiance pour le paramètre d'intensité ν

Lorsqu'on sait que l'on est en présence d'un processus de Poisson (en utilisant le test précédent, par exemple), on peut trouver un intervalle de confiance pour ν . Pour cela, la proposition suivante est nécessaire :

Proposition 6.2. *Soit $\{N(t), t \geq 0\}$ un processus de Poisson d'intensité ν . On dénote par T_n le temps d'occurrence du n -ième événement. On a alors*

$$2\nu T_n \sim \chi_{2n}^2.$$

Pour rappel, une variable aléatoire X suit une loi χ^2 à ν degrés de liberté ($\nu \in \mathbb{N}_0$), ce qui se note $X \sim \chi_\nu^2$, si sa fonction de densité est donnée par

$$f_X(x) = \frac{1}{2^{\frac{\nu}{2}} \Gamma(\frac{\nu}{2})} e^{-\frac{x}{2}} x^{\frac{\nu}{2}-1} \chi_{[0,+\infty[}(x).$$

Démonstration. La preuve est directe en appliquant la formule du changement de variable à partir de la densité de T_n donnée par l'équation 2.15. ■

Par conséquent, on a

$$\mathbb{P} \left[Q_{\chi_{2n}^2} \left(\frac{\alpha}{2} \right) \leq 2\nu T_n \leq Q_{\chi_{2n}^2} \left(1 - \frac{\alpha}{2} \right) \right] = 1 - \alpha$$

où $Q_{\chi_{2n}^2}(\alpha)$ est tel que $\mathbb{P}[Y \leq Q_{\chi_{2n}^2}(\alpha)] = \alpha$, si $Y \sim \chi_{2n}^2$. Ce qui permet d'obtenir, par la méthode du pivotage, un intervalle de confiance au niveau $1 - \alpha$ pour ν :

$$IC_{\alpha,n}(\nu) = \left[\frac{Q_{\chi_{2n}^2} \left(\frac{\alpha}{2} \right)}{2T_n}, \frac{Q_{\chi_{2n}^2} \left(1 - \frac{\alpha}{2} \right)}{2T_n} \right].$$

En pratique, on dispose d'un échantillon des temps d'occurrence pour construire l'intervalle. Pour analyser l'évolution de cet intervalle en fonction du paramètre n sans devoir générer les temps d'occurrence, il serait profitable de les remplacer par des valeurs possibles pour ceux-ci, comme par exemple leur espérance. Comme cette dernière dépend de ν , il faut remplacer ce paramètre par une valeur quelconque, prenons $\nu = 2$. L'intervalle devient alors

$$IC_{\alpha,n}(\nu = 2) = \left[\frac{Q_{\chi_{2n}^2} \left(\frac{\alpha}{2} \right)}{n}, \frac{Q_{\chi_{2n}^2} \left(1 - \frac{\alpha}{2} \right)}{n} \right].$$

Comme n se trouve au dénominateur, la largeur de l'intervalle diminue lorsque n augmente. Voyons cela sur le graphique 6.1.

Tous les intervalles contiennent bien la valeur $\nu = 2$ mais il vaut mieux prendre une valeur de n élevée pour que l'intervalle soit plus précis.

Si maintenant la valeur de ν est beaucoup plus grande, par exemple $\nu = 50$, le graphique 6.2 montre que la précision de l'intervalle apparaît meilleure pour des petites valeurs de ν .

Dans le logiciel R, il est possible d'obtenir les quantiles de la distribution χ^2 via la commande `qchisq(x,ddl)`.

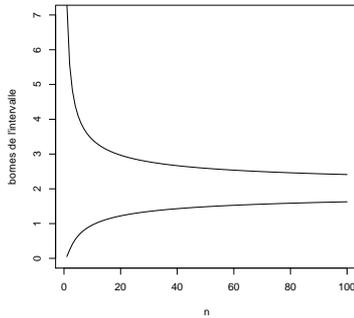


FIG. 6.1 – Limites des intervalles de confiance pour le paramètre $\nu=2$

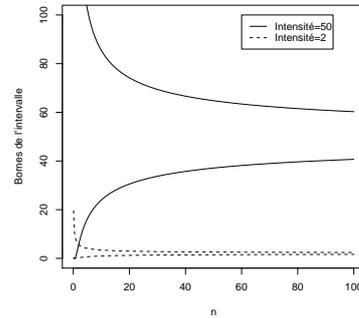


FIG. 6.2 – Comparaison des intervalles de confiance pour le paramètre ν

6.1.4 Comparaison des intensités de deux processus de Poisson

Soient $\{N(t), t \geq 0\}$ et $\{N'(t), t \geq 0\}$ deux processus de Poisson indépendants d'intensité respective ν et ν' . Il serait intéressant de tester si les deux intensités peuvent être supposées égales. Soient deux entiers n et n' , T_n le temps d'occurrence du n -ième événement du processus $\{N(t), t \geq 0\}$ et $T_{n'}$ celui du n' -ième événement du processus $\{N'(t), t \geq 0\}$. De la proposition 6.2, on peut déduire que $2\nu T_n \sim \chi_{2n}^2$ et $2\nu' T_{n'} \sim \chi_{2n'}^2$. De plus, comme les processus sont indépendants, ces deux variables le sont aussi. Sous l'hypothèse d'égalité des intensités, $\frac{n'T_n}{nT_{n'}}$ suit une loi F à n et n' degrés de liberté.¹ Ainsi, l'hypothèse d'égalité sera rejetée au niveau de confiance $1 - \alpha$ si $\frac{n'T_n}{nT_{n'}} < Q_{F_{n,n'}}(\frac{\alpha}{2})$ ou si $\frac{n'T_n}{nT_{n'}} > Q_{F_{n,n'}}(1 - \frac{\alpha}{2})$, où $Q_{F_{n,n'}}(\alpha)$ est tel que $\mathbb{P}[Y \leq Q_{F_{n,n'}}(\alpha)] = \alpha$ si $Y \sim F_{n,n'}$. Si l'hypothèse d'égalité est rejetée, il est possible d'utiliser des tests unilatéraux pour déterminer quel paramètre est le plus grand.

6.2 Processus de Poisson composé

Dans les différentes propriétés déjà étudiées, on a constaté que celles relatives aux processus de Poisson composés se déduisaient souvent assez facilement de leur équivalent dans le cas de base ou avaient le même schéma de démonstration. C'est encore le cas ici, et c'est pourquoi nous abordons d'abord ce cas.

6.2.1 Simulation

Si l'intensité ν du processus de Poisson sous-jacent $\{N(t), t \geq 0\}$ et la distribution des variables aléatoires Y_n , $n \in \mathbb{N}$ sont données, alors il est possible de générer le processus $\{X(t), t \geq 0\}$ avec $X(t) = \sum_{n=0}^{N(t)} Y_n$. Pour ce faire, on génère les inter-arrivées $\{w_1, w_2, \dots, w_n\}$ de $\{N(t), t \geq 0\}$, ce qui permet d'obtenir les temps d'occurrence du processus sous-jacent mais

¹Soient $X_1 \sim \chi_{\nu_1}^2$ et $X_2 \sim \chi_{\nu_2}^2$ indépendantes. La variable aléatoire Y définie par

$$Y = \frac{\frac{X_1}{\nu_1}}{\frac{X_2}{\nu_2}}$$

est distribuée selon une loi F à ν_1 et ν_2 degrés de liberté, ce que l'on note $Y \sim F_{\nu_1, \nu_2}$.

également ceux de $\{X(t), t \geq 0\}$ qui sont les mêmes par définition. On génère également un échantillon aléatoire et simple $\{y_n, n \in \mathbb{N}\}$ de même distribution que les Y_n . Connaissant une réalisation de $N(t)$ pour tout $t \geq 0$ et de Y_n pour tout $n \in \mathbb{N}$, on peut trouver la réalisation de $X(t)$ correspondante pour tout $t \geq 0$.

6.2.2 Tests

Si le processus observé présente plusieurs temps d'occurrence égaux, il faut de suite rejeter l'hypothèse que le processus pourrait être un processus de Poisson de base. Par contre, il est possible de tester s'il ne serait pas un processus de Poisson composé.

Si le nombre d'événements observés est très grand, il sera utile d'exploiter un algorithme pour détecter ce fait. Ensuite, on ne considère qu'une seule fois chaque temps d'occurrence, ce qui correspond au processus sous-jacent. On teste alors son caractère poissonien comme dans la section précédente. Il faut cependant faire l'hypothèse supplémentaire que lorsque la variable Y_n était nulle, il a quand même été noté qu'une occurrence du processus sous-jacent a eu lieu. Par exemple, pour étudier le processus des personnes descendant du bus à un certain arrêt, il faut noter les temps d'occurrence auxquels un bus passe à l'arrêt même si personne n'en descend. Dans ce cas, la variable Y_n est simplement nulle. Cela permet de connaître exactement le processus sous-jacent.

Ensuite, l'indépendance entre la famille de variables $\{Y_n, n \in \mathbb{N}\}$ et le processus de Poisson sous-jacent doit être testée. Cela peut se faire à l'aide du test d'indépendance du χ^2 entre un échantillon aléatoire issu de la distribution de gain et un échantillon aléatoire issu de la distribution de Poisson de paramètre νt pour t quelconque qui s'effectue simplement dans R via la commande `chisq.test()`.

6.3 Processus de Poisson non homogène

Comme précédemment, c'est cette partie qui est la plus délicate.

6.3.1 Simulation

A partir d'une fonction $m(t)$ croissante, continue et différentiable, on désire générer un processus de Poisson non homogène de fonction moyenne $m(t)$. Rappelons que préciser la fonction moyenne est équivalent à donner la fonction d'intensité vu leur lien : $m(t) = \int_0^t \nu(y) dy$.

Plusieurs méthodes existent mais elles ne sont pas toujours toutes applicables. Le cas le plus simple est celui où la fonction moyenne est strictement croissante et donc inversible. Nous considérerons ce cas en premier, pour passer ensuite au cas plus général d'une fonction moyenne non-décroissante.

Fonction moyenne strictement croissante

Lorsque la fonction moyenne $m(t)$ est strictement croissante, c'est la proposition 3.11 qui donne un moyen simple de simuler un processus de Poisson non homogène. Il suffit de générer un processus de Poisson d'intensité constante $\nu = 1$, comme expliqué à la section 6.1.1, et de transformer le processus des temps d'occurrence $\{T_n, n \in \mathbb{N}\}$ en un processus des temps

d'occurrence $\{T'_n, n \in \mathbb{N}\}$ d'un processus de Poisson de fonction moyenne $m(t)$ par la transformation $T'_n = m^{-1}(T_n)$.

Fonction moyenne quelconque

Cette méthode est valable pour toute fonction moyenne $m(t)$ mais est plus difficile à implémenter. Elle se base sur la proposition 3.7. Si l'on souhaite obtenir des événements répartis dans l'intervalle de temps $[0, T]$, voici ce qu'il faut faire :

- Générer une valeur N d'une variable aléatoire de Poisson d'intensité $m(T)$.
- Générer N valeurs des variables aléatoires $V_i, i = 1, 2, \dots, N$ ayant la fonction $F(u) = \frac{m(u)}{m(T)}$ pour fonction de répartition.
- Les temps d'occurrence du processus non homogène sont les valeurs ordonnées des variables V_i .

En effet, la proposition 3.7 stipule que lorsque l'on connaît le nombre d'événements ayant eu lieu entre 0 et T , les temps d'occurrence de ces événements ont la même distribution que les statistiques d'ordre de variables aléatoires indépendantes et de même distribution F . Vu la distribution choisie, la valeur N de la méthode ci-dessus est ce nombre d'événements entre 0 et T .

La performance de cet algorithme dépend de la capacité à générer des valeurs de variables aléatoires de distribution $F(u)$. Cela peut se faire en utilisant les variables uniformes comme cela a été montré à la proposition 6.1.

6.3.2 Tests

Comme la perte de la stationnarité des accroissements implique que les temps d'inter-arrivées ne sont plus identiquement distribués, ceux-ci ne peuvent plus être utilisés si facilement pour les tests. Plusieurs méthodes existent pour tester si un ensemble de données peut être considéré comme étant issu d'un processus de Poisson non homogène. Cependant cela semble moins facile que pour les autres cas car il faut au préalable trouver une bonne forme pour la fonction $m(t)$. Les deux tests présentés ici supposent donc que la fonction moyenne est connue ou a été estimée. En fonction de ses caractéristiques, il faut utiliser l'un ou l'autre test. Une section sera consacrée à l'estimation de la fonction $m(t)$.

Test basé sur les temps d'occurrence

Ce test se base sur la proposition 3.7. On observe le processus que l'on veut tester durant un certain temps T pendant lequel il se produit n événements aux temps t_1, t_2, \dots, t_n . Si le processus est un processus de Poisson non homogène, l'échantillon obtenu doit être une réalisation du vecteur aléatoire (U_1, \dots, U_n) tel que $U_i \stackrel{\text{iid}}{\sim} F_U$ où $F_U(u) = \frac{m(u)}{m(T)}$. Il faut donc connaître la fonction moyenne du processus ou l'estimer. On effectue ensuite un test statistique d'hypothèses

$$H_0 : F = F_U \quad \text{vs} \quad H_1 : F \neq F_U.$$

Test basé sur le lien entre un processus de Poisson non homogène et un processus de Poisson d'intensité 1

La section 3.3 nous apprend que si la fonction $m(t)$ est strictement croissante, $\{N(t), t \geq 0\}$ est un processus de Poisson non homogène de fonction moyenne $m(t)$ si et seulement si le processus $\{M(u), u \geq 0\}$ où $M(u) = N(m^{-1}(u))$ est un processus de Poisson d'intensité $\nu = 1$. Donc si la fonction moyenne est strictement croissante, on peut tester que le processus $\{N(t), t \geq 0\}$ est un processus de Poisson non homogène de fonction moyenne $m(t)$ en testant que le processus $\{M(u), u \geq 0\}$ est un processus de Poisson d'intensité $\nu = 1$.

6.3.3 Estimation de la fonction moyenne

Pour cette estimation, il faut être en présence d'un processus de Poisson non homogène. Soient S un instant fixé et k réalisations indépendantes du processus sur $[0, S]$ avec $k > 1$ (cette condition sur le nombre de réalisations sera discutée plus tard).

Posons, $\forall t \in [0, S]$, $N_j(t)$ le nombre d'événements observés dans l'intervalle de temps $[0, t]$ lors de la j^e réalisation du processus et

$$N_k^*(t) = \sum_{j=1}^k N_j(t)$$

le nombre d'événements observés dans le même intervalle lors des k réalisations. Le graphique 6.3 représente $k = 4$ réalisations d'un processus non homogène. En tout temps t , la réalisation de la variable $N_4^*(t)$ est la somme des quatre valeurs de $N(t)$ pour chacune des réalisations. Par exemple, $N_4^*(3) = 4 + 6 + 10 + 10 = 30$.

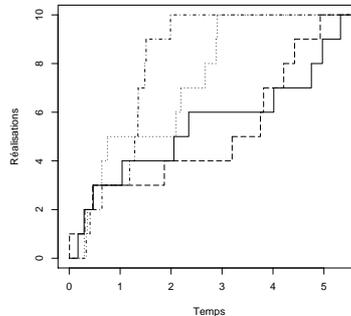


FIG. 6.3 – Réalisations d'un même processus non homogène

Notons encore $n = N_k^*(S)$, c-à-d le nombre total d'événements lors des k réalisations indépendantes du processus sur $[0, S]$ et $t_{(1)}, \dots, t_{(n)}$ les temps d'occurrence des événements des k réalisations ordonnés par ordre croissants. On pose $t_{(0)} = 0$ et $t_{(n+1)} = S + \varepsilon$ où $\varepsilon > 0$ est suffisamment petit.

Proposition 6.3. Avec les notations introduites ci-dessus, la variable aléatoire $\tilde{m}_k(t) = \frac{N_k^*(t)}{k}$ converge presque sûrement vers $m(t)$ pour tout t fixé, c-à-d

$$\lim_{k \rightarrow +\infty} \tilde{m}_k(t) = m(t) \text{ avec une probabilité égale à 1.}$$

Démonstration. Vu la définition des $N_j(t)$, on a $N_j(t) \stackrel{\text{iid}}{\sim} \mathcal{P}(m(t))$, $\forall t \in [0, S]$. Cela étant, $E[N_j(t)] = V[N_j(t)] = m(t)$, $\forall t \in [0, S]$.

En appliquant la loi forte des grands nombres, on obtient que $\tilde{m}_k(t)$ converge presque sûrement vers $m(t)$. ■

En conséquence, pour tout t fixé et pour un entier k quelconque, $\tilde{m}_k(t)$ peut être exploité comme estimateur de $m(t)$. La proposition 6.3 montre que cet estimateur est consistant. Le théorème de la limite centrée permet de dire que

$$\frac{\tilde{m}_k(t) - m(t)}{\frac{\sqrt{m(t)}}{\sqrt{k}}} \xrightarrow{\mathcal{L}} N(0, 1)$$

tandis que le théorème de Slutsky² appliqué à $m(t)$ et $\tilde{m}_k(t)$ avec la fonction $g(x) = \sqrt{\frac{m(t)}{x}}$ implique que

$$\sqrt{\frac{m(t)}{\tilde{m}_k(t)}} \xrightarrow{\mathcal{L}} 1.$$

Le problème de discontinuité de la fonction g doit être évité en prenant pour $t = 0$ le temps en lequel la fonction d'intensité commence à ne plus être nulle. Comme cela, $m(t) > 0$ pour tout $t > 0$ et

$$\begin{aligned} \mathbb{P}[\tilde{m}_k(t) = 0] &= \mathbb{P}\left[\sum_{j=1}^k N_j(t) = 0\right] = \mathbb{P}[N_j(t) = 0, \forall j = 1, \dots, k] \\ &= \prod_{j=1}^k \mathbb{P}[N_j(t) = 0] \quad \text{par hypothèse d'indépendance des réalisations} \\ &= \prod_{j=1}^k e^{-m(t)} = e^{-m(t)k} \rightarrow 0 \text{ si } k \rightarrow +\infty. \end{aligned}$$

Ainsi, $\tilde{m}_k(t) = 0$ est un événement stochastiquement impossible.

Vu ces deux limites, on a donc

$$\frac{\tilde{m}_k(t) - m(t)}{\sqrt{\frac{\tilde{m}_k(t)}{k}}} = \sqrt{\frac{m(t)}{\tilde{m}_k(t)}} \frac{\tilde{m}_k(t) - m(t)}{\sqrt{\frac{m(t)}{k}}} \xrightarrow{\mathcal{L}} N(0, 1).$$

De là, il est possible de construire un intervalle de confiance au niveau $1 - \alpha$ pour $m(t)$:

$$IC_\alpha(m(t)) = \left[\tilde{m}_k(t) - Q_z\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\tilde{m}_k(t)}{k}}, \tilde{m}_k(t) + Q_z\left(1 - \frac{\alpha}{2}\right) \sqrt{\frac{\tilde{m}_k(t)}{k}} \right]$$

où l'on a $\mathbb{P}[Z \leq Q_z(1 - \frac{\alpha}{2})] = 1 - \frac{\alpha}{2}$ avec $Z \sim N(0, 1)$.

Le problème de cet estimateur est qu'il n'est pas continu. Or la théorie a été développée sous l'hypothèse de continuité de $m(t)$. En partant de cette première estimation, il est possible d'en

²Si X_n converge en loi vers X (ce qui est le cas notamment lorsque X_n converge presque sûrement vers X), alors $g(X_n)$ converge en loi vers $g(X)$ pour toute fonction continue g .

construire une autre qui soit continue cette fois.

Posons

$$\hat{m}_k(t) = U_k \tilde{m}_k(t) + R_k(t) \quad (6.1)$$

avec

$$U_k = \frac{n}{n+1} = \frac{N_k^*(S)}{N_k^*(S) + 1} \quad (6.2)$$

$$R_k(t) = \frac{U_k}{k} \left[\frac{t - t_{(N_k^*(t))}}{t_{(N_k^*(t)+1)} - t_{(N_k^*(t))}} \right] \quad (6.3)$$

Si $N_k^*(t) = i$, c-à-d si $t_{(i)} \leq t < t_{(i+1)}$, cette expression peut se réécrire sous la forme

$$\hat{m}_k(t) = \frac{n}{n+1} \frac{1}{k} \left(i + \frac{t - t_{(i)}}{t_{(i+1)} - t_{(i)}} \right) \quad (6.4)$$

pour tout $i = 0, 1, \dots, n$.

Proposition 6.4. *Avec les nouvelles notations introduites ci-dessus, la variable aléatoire $\hat{m}_k(t)$ converge presque sûrement vers $m(t)$ pour tout t fixé, c-à-d*

$$\lim_{k \rightarrow +\infty} \hat{m}_k(t) = m(t) \text{ avec une probabilité égale à } 1.$$

De plus, cette estimation est continue.

Démonstration. Pour la convergence, il est nécessaire de calculer plusieurs limites.

Premièrement, $\lim_{k \rightarrow \infty} N_k^*(S) = \infty$ puisque le nombre de processus observés tend vers l'infini.

Ensuite, vu la relation (6.2), on a

$$\lim_{k \rightarrow \infty} U_k = \lim_{k \rightarrow \infty} \frac{N_k^*(S)}{N_k^*(S) + 1} = 1.$$

De plus, comme $t_{(i)} \leq t < t_{(i+1)}$ implique

$$0 \leq \frac{t - t_{(i)}}{t_{(i+1)} - t_{(i)}} < 1,$$

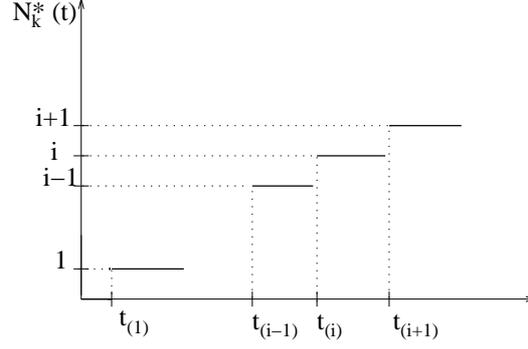
l'équation (6.3) donne $0 \leq R_k(t) < \frac{U_k}{k}$. En passant à la limite sur k tendant vers l'infini, il vient

$$\lim_{k \rightarrow \infty} R_k(t) = 0.$$

Tout cela mis ensemble dans la relation (6.1) montre que

$$\lim_{k \rightarrow \infty} \hat{m}_k(t) = \lim_{k \rightarrow \infty} \tilde{m}_k(t) = m(t)$$

avec une probabilité 1 vu la proposition 6.3.


 FIG. 6.4 – Illustration de $N_k^*(t)$

Passons alors à la continuité. Vu l'expression (6.4), il suffit de montrer la continuité en $t_{(i)}$ pour tout $i = 0, 1, \dots, n$. En partant alors des expressions (6.1), (6.2) et (6.3), on obtient

$$\lim_{t \rightarrow t_{(i)}^-} \hat{m}_k(t) = \frac{n}{n+1} \lim_{t \rightarrow t_{(i)}^-} \left(\frac{N_k^*(t)}{k} \right) + \frac{n}{(n+1)k} \lim_{t \rightarrow t_{(i)}^-} \left(\frac{t - t_{(N_k^*(t))}}{t_{(N_k^*(t)+1)} - t_{(N_k^*(t))}} \right).$$

La limite à considérer est la limite à gauche d'une fonction en escaliers qui vaut $i-1$ avant $t_{(i)}$ et i en $t_{(i)}$. La limite vaut donc $i-1$, comme cela est illustré par la figure 6.4. Ainsi,

$$\begin{aligned} \lim_{t \rightarrow t_{(i)}^-} \hat{m}_k(t) &= \frac{n}{(n+1)k} (i-1) + \frac{n}{(n+1)k} \frac{t_{(i)} - t_{(i-1)}}{t_{(i)} - t_{(i-1)}} \\ &= \frac{ni}{(n+1)k}. \end{aligned}$$

Grâce aux mêmes équations, il vient

$$\begin{aligned} \lim_{t \rightarrow t_{(i)}^+} \hat{m}_k(t) &= \frac{n}{n+1} \lim_{t \rightarrow t_{(i)}^+} \left(\frac{N_k^*(t)}{k} \right) + \frac{n}{(n+1)k} \lim_{t \rightarrow t_{(i)}^+} \left(\frac{t - t_{(N_k^*(t))}}{t_{(N_k^*(t)+1)} - t_{(N_k^*(t))}} \right) \\ &= \frac{n}{(n+1)k} i + \frac{n}{(n+1)k} \frac{t_{(i)} - t_{(i)}}{t_{(i+1)} - t_{(i)}} \\ &= \frac{ni}{(n+1)k}. \end{aligned}$$

Ce qui permet de conclure à la continuité. ■

Comme cela a déjà été fait, le théorème de Slutsky permet de prouver que $\sqrt{\frac{m(t)}{\hat{m}_k(t)}}$ converge en loi vers 1. De là,

$$\begin{aligned} \frac{\hat{m}_k(t) - m(t)}{\sqrt{\frac{\hat{m}_k(t)}{k}}} &= \frac{U_k \tilde{m}_k(t) + R_k(t) - m(t)}{\sqrt{\frac{m(t)}{k}}} \sqrt{\frac{m(t)}{\hat{m}_k(t)}} \\ &= \sqrt{\frac{m(t)}{\hat{m}_k(t)}} \left(\frac{R_k(t)}{\sqrt{\frac{m(t)}{k}}} + \frac{U_k \tilde{m}_k(t) - m(t)}{\sqrt{\frac{m(t)}{k}}} \right). \end{aligned}$$

Vu les relations précédentes, le premier terme converge vers zéro quand k tend vers l'infini car

$$0 \leq \frac{R_k(t)}{\sqrt{\frac{m(t)}{k}}} < \frac{U_k}{k} \sqrt{\frac{k}{m(t)}} = \frac{U_k}{\sqrt{k}\sqrt{m(t)}}$$

et le second converge vers $\lim_{k \rightarrow \infty} \frac{\tilde{m}_k(t) - m(t)}{\sqrt{\frac{m(t)}{k}}}$ qui suit une loi gaussienne par le théorème de la limite centrée. Au total,

$$\frac{\hat{m}_k(t) - m(t)}{\sqrt{\frac{\hat{m}_k(t)}{k}}} \xrightarrow{\mathcal{L}} N(0, 1).$$

Un intervalle de confiance pour $m(t)$ basé sur $\hat{m}_k(t)$ est donc donné par

$$IC_\alpha(m(t)) = \left[\hat{m}_k(t) - Q_z(1 - \frac{\alpha}{2}) \sqrt{\frac{\hat{m}_k(t)}{k}}, \hat{m}_k(t) + Q_z(1 - \frac{\alpha}{2}) \sqrt{\frac{\hat{m}_k(t)}{k}} \right].$$

Pour tester un processus observé afin de voir s'il est un processus de Poisson non homogène, il y a donc deux étapes. La première est d'estimer $m(t)$ en supposant que c'est un processus de Poisson non homogène. La seconde est de voir si, avec cette fonction $m(t)$ estimée, le caractère poissonien du processus observé est acceptable. Pour cela, il faut donc disposer d'au moins deux réalisations du processus ; une pour estimer la fonction moyenne et l'autre pour tester le caractère poissonien. En effet, lorsqu'une seule réalisation est disponible, l'estimateur $\tilde{m}_1(t)$ est simplement $N(t) = N_1^*(t)$ et donc

$$\hat{m}_1(t) = \frac{n}{n+1} \left(N(t) + \frac{t - t_{(N(t))}}{t_{(N(t)+1)} - t_{(N(t))}} \right).$$

Ainsi, en chaque temps d'occurrence, cette estimation vaut $\frac{n}{n+1}N(t)$, ce qui est proche de $N(t)$, et en d'autres temps, elle diffère juste un peu de $N(t)$ car le second terme est toujours compris entre 0 et 1. La différence majeure entre ces deux fonctions est que l'estimation \hat{m}_1 est continue.

Lors du test du χ^2 , il faut utiliser une répartition en classes, dont les bornes de la k -ième peuvent être notées e_{k-1} et e_k , et chercher les effectifs observés et théoriques de ces différentes classes. L'effectif observé de la k -ième classe vaut $N(e_k) - N(e_{k-1})$ tandis que l'effectif théorique est donné par

$$\hat{m}_1(e_k) - \hat{m}_1(e_{k-1}) \approx N(e_k) - N(e_{k-1}).$$

Pour chaque classe, l'effectif théorique est donc proche de l'effectif observé et la statistique du test

$$\chi_{\text{obs}}^2 = \sum_{k=1}^{\text{nbr de classes}} \frac{(\text{eff}_{\text{obs}}[k] - \text{eff}_{\text{th}}[k])^2}{\text{eff}_{\text{th}}[k]}$$

est donc proche de 0, ce qui conduit à une p-value proche de 1 puisqu'une variable aléatoire suivant une loi χ^2 est toujours à valeurs positives. Ce résultat qui semble très beau en pratique ne l'est pas car il n'est dû qu'au manque d'information sur le processus observé. Avec plusieurs réalisations du processus, l'estimation de la fonction moyenne se baserait sur plus d'informations. Ainsi, malgré qu'elle ne serait plus aussi proche des fonctions $N_j(t)$, elle n'en serait que meilleure.

Chapitre 7

Applications

Ce chapitre a pour but d'utiliser les méthodes présentées dans les chapitres théoriques. Pour les illustrer, il faut des données à analyser ; celles-ci ont été collectées dans la littérature ou observées. Par souci de clarté, aucune explication déjà donnée dans ce qui précède ne sera retransmise. C'est pourquoi la première partie de ce travail constitue un réel pré-requis pour celle entamée ici. Toutes les données se trouvent en annexe. Même lorsque cela n'est pas dit explicitement, le niveau d'incertitude choisi vaut 5%.

7.1 Données provenant de la littérature

Il n'est pas aisé d'obtenir des données à analyser lorsque celles-ci doivent se présenter sous la forme des temps d'inter-arrivées ou d'occurrence d'un phénomène. La plupart des bases de données reprennent le nombre d'événements qui se sont produits sur un certain laps de temps, ce qui n'est d'aucune utilité dans l'analyse à réaliser ici. Par ailleurs, l'observation de phénomènes intéressants dans cette théorie demande beaucoup de temps. C'est pourquoi la littérature pouvait être d'une grande aide. Les articles choisis ne sont pas récents. Cela permet de comparer les techniques utilisées avant la grande dispersion des logiciels statistiques avec les techniques développées dans ce travail.

Les explosions dans les mines de Grande-Bretagne impliquant le décès de dix personnes ou plus.

Ce premier ensemble de données provient de l'article de B.A. MAGUIRE, E.S. PEARSON et A.H.A WYNN intitulé "The time intervals between industrial accidents" et publié dans la revue *Biometrika* 39 en 1952 (p.169). Les données correspondent aux nombres de jours entre deux explosions causant au moins dix décès dans les mines de Grande-Bretagne. La période de temps couverte va du 6 décembre 1875 jusqu'au 29 mai 1951. La figure 7.1 représente la fonction de répartition empirique de ces données.

Comme il y a 109 observations et que le test de Kolmogorov-Smirnov donne une statistique vérifiant $D_{109} = 0,0906 < d_{109} = \frac{1,08}{\sqrt{109}} = 0,103$, il est donc possible d'affirmer à un niveau de confiance de 95% que les données correspondent aux temps d'inter-arrivées d'un processus de Poisson dont l'intensité se trouve, avec 5% de risque d'erreur, dans l'intervalle $[0,003; 0,005]$. Si l'on regarde le graphique associé à ce test, représenté à la figure 7.2, on remarque que les premiers points suivent bien une même droite. Cependant ce n'est pas celle donnée par l'ajus-

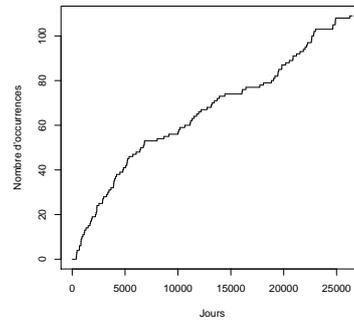


FIG. 7.1 – Évolution du nombre d'explosions en fonction du temps

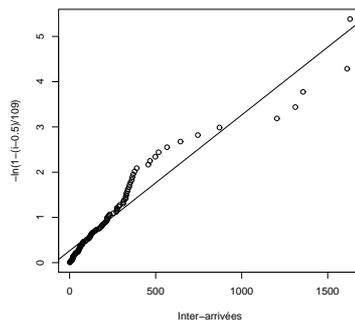


FIG. 7.2 – Qq-plot des inter-arrivées des explosions

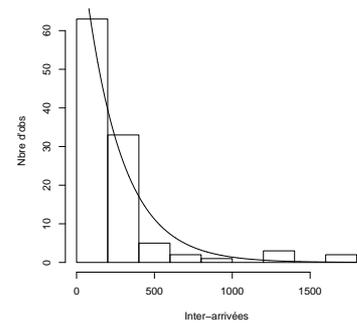


FIG. 7.3 – Histogramme des inter-arrivées des explosions

tement linéaire car elle est déviée par les derniers points. Cela signifie donc qu'il y a quelques temps d'inter-arrivées qui sont fort grands pour être considérés comme issus d'une distribution exponentielle. Pour confirmer cette idée, regardons l'histogramme 7.3 correspondant aux données. Il montre bien que les observations entre 1200 et 1400 puis entre 1600 et 1800 sont trop nombreuses. Malgré cela, le test accepte quand même la distribution exponentielle avec un niveau de risque de 5%. L'article allait aussi dans ce sens, à savoir que l'ajustement exponentiel n'est pas parfait mais tout de même acceptable.

Les accidents successifs dans une partie d'une mine de Grande-Bretagne.

Dans le même article mais à la page 170, se trouve aussi des données correspondant aux durées, exprimées en jours, séparant deux accidents dans une partie d'une mine.

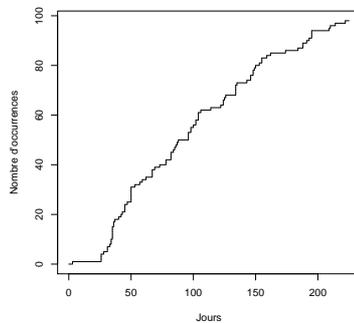


FIG. 7.4 – Évolution du nombre d'accidents dans une partie d'une même mine au cours du temps

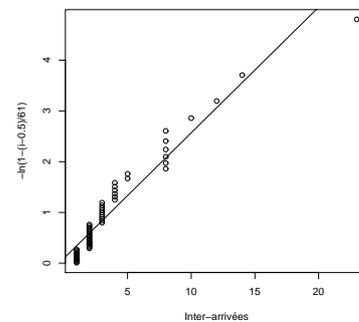


FIG. 7.5 – Qq-plot des temps d'inter-arrivées des accidents dans une partie d'une même mine

Il y a 99 données mais il est possible de voir sur le graphique 7.4 que certains accidents ont eu lieu le même jour puisque certaines marches sont plus hautes. Il faut alors travailler avec le processus sous-jacent, c-à-d qu'on ne doit pas considérer les valeurs nulles des inter-arrivées. Nous allons tester si le processus sous-jacent correspond à un processus de Poisson.

Il n'y a plus que 61 données et le test de Kolmogorov-Smirnov sur les inter-arrivées donne une valeur de la statistique D_{61} égale à 0.2403. On doit refuser l'hypothèse selon laquelle les inter-arrivées sont distribuées exponentiellement car $D_{61} > d_{61} = 0.1369$. La méthode graphique permet de se convaincre de cela car la figure 7.5 montre bien qu'un ajustement linéaire n'est pas recommandé.

Ce graphique présente une particularité supplémentaire par rapport au graphique 7.2, à savoir qu'il montre plusieurs ensembles de points. Les points d'un même ensemble sont caractérisés par une même valeur du temps d'inter-arrivées. Cela traduit un manque de précision dans le relevé des données car, en théorie, il n'est pas possible d'avoir deux valeurs égales pour une même variable continue. Dans cet exemple, il ne se passe pas exactement un nombre entier de jours entre deux accidents, mais les données ont été arrondies à l'entier le plus proche.

L'article suppose que le rejet du modèle exponentiel est dû au fait que le risque d'accident n'est pas constant. Si l'on disposait d'au moins deux réalisations de ce processus, il serait possible de tester le modèle d'un processus de Poisson non homogène. Cela n'étant pas le cas,

la seule conclusion possible sur ces données est que le processus sous-jacent ne forme pas un processus de Poisson de base et donc que les temps entre deux accidents ne constituent pas un ensemble de temps d'inter-arrivées d'un processus de Poisson composé.

7.2 Données observées

Passons ensuite aux données qui ont été récoltées dans le cadre de ce travail. Ces données n'ont pas été choisies n'importe comment ; certaines font partie d'un ensemble d'exemples qui sont réputés pour être des processus de Poisson, d'autres ont été sélectionnés pour vérifier une intuition personnelle.

Sinistres subis par des personnes assurées par un particulier

Le premier exemple traité est celui des sinistres subis par des assurés auprès d'un assureur indépendant qui travaille pour deux compagnies d'assurances différentes. Ce dernier nous a fourni les données concernant les accidents de voiture et les incendies. Il serait utile pour lui de savoir si ses clients des deux compagnies se comportent de la même façon ou s'il y a des différences entre ces deux groupes de clients. Les données ont été récoltées au cours des années 2004 et 2005. Commençons par analyser celles relatives aux accidents de voitures.

Pour la première compagnie, nous disposons de 90 données ; et pour la seconde, il y en a 56. Elles sont représentées aux figures 7.6 et 7.7. Comme certaines "marches" de ces deux

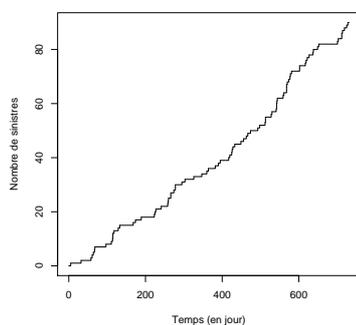


FIG. 7.6 – Évolution du nombre de sinistres "voitures" de la première compagnie

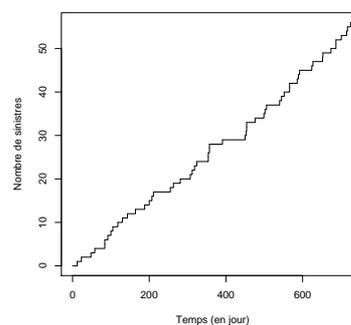


FIG. 7.7 – Évolution du nombre de sinistres "voitures" de la seconde compagnie

graphiques sont plus hautes que les autres, le processus n'est pas un processus de Poisson de base mais est peut-être un processus de Poisson composé. Pour voir cela, il faut considérer les processus sous-jacents à ces deux processus. Les deux ensembles de données correspondent bien à un processus de Poisson car les statistiques vérifient $D_{81} = 0,1053 < d_{81} = 0,1192$ et $D_{50} = 0,1269 < d_{50} = 0,1509$. Pour illustrer ces résultats, les figures 7.8 et 7.9 sont les histogrammes des données relatives à la première et à la seconde compagnie respectivement. Le choix de représenter l'histogramme plutôt que le qq-plot est dû au manque de précision dans l'observation des données ; celles-ci sont exprimées en jours, ce qui fait que certaines valeurs sont observées un grand nombre de fois comme c'était le cas pour l'exemple des accidents dans les mines traité précédemment.

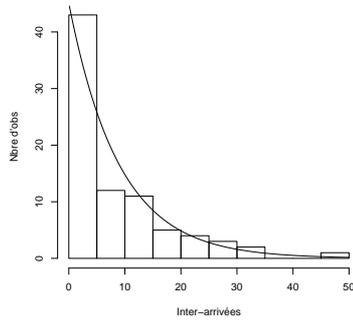


FIG. 7.8 – Histogramme des sinistres "voitures" de la première compagnie

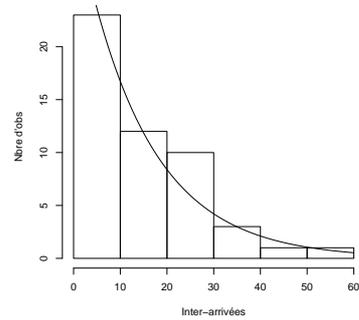


FIG. 7.9 – Histogramme des sinistres "voitures" de la seconde compagnie

Les intensités de ces deux processus sont estimées par la méthode du maximum de vraisemblance qui donne $\nu = 0,11$ pour la première compagnie et $\nu = 0,07$ pour la seconde. Essayons de voir si l'on peut supposer que ces intensités sont comparables grâce à la méthode développée à la page 76. Pour cela, il faut l'indépendance entre les processus, ce qui est acquis car le test du χ^2 donne une p-value de 0,2724. Considérons au choix deux entiers plus petits que les nombres d'occurrence des deux processus, par exemple 45 et 12. Notons T_i et T'_j les temps d'occurrence du i -ième accident pour la première compagnie et du j -ième pour la seconde. Pour pouvoir pencher en faveur de l'égalité des intensités, il faut que le quotient $\frac{12 * T_{45}}{45 * T'_{12}}$ soit compris entre les quantiles 0,025 et 0,975 de la distribution de Fisher-Snedecor de paramètres 45 et 12, ce qui est bien le cas. On peut donc conclure que ces processus sont comparables en intensité.

Passons alors à l'analyse des données concernant les incendies. Les données sont représentées dans les graphiques 7.10 et 7.11. Il y en a 20 pour la première compagnie et 16 pour la seconde. Certains temps d'inter-arrivées étant nuls, on travaillera avec les processus sous-jacents. Pour le premier ensemble de données, le test de Kolmogorov-Smirnov sur les temps d'inter-

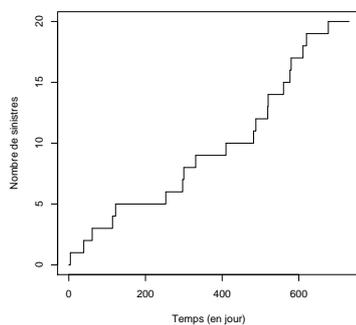


FIG. 7.10 – Évolution du nombre de sinistres "incendies" de la première compagnie

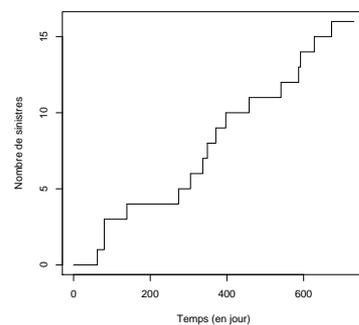


FIG. 7.11 – Évolution du nombre de sinistres "incendies" de la deuxième compagnie

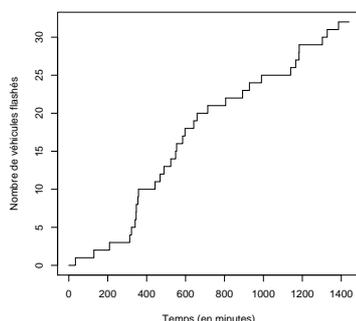


FIG. 7.12 – Évolution du nombre de véhicules flashés le 1 janvier 2006

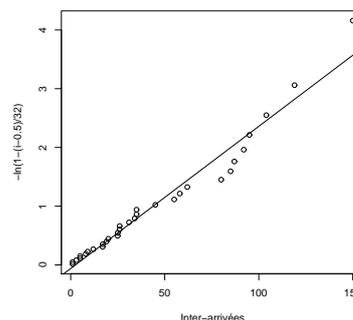


FIG. 7.13 – Qq-plot des temps d'inter-arrivées des flash

arrivées donne une statistique vérifiant $D_{15} = 0,1971 < d_{15} = 0,2686$. Pour le second, on obtient $D_{20} = 0,1498 < d_{20} = 0,2345$. Il en résulte que les deux processus sous-jacents sont des processus de Poisson dont les intensités respectives sont estimées à 0,03 et 0,02. La p-value du test d'indépendance des deux processus valant 0,98, on peut essayer de les comparer. Comme le quotient $\frac{9 * T_{14}}{14 * T_9}$ est compris entre les quantiles 0,025 et 0,975 de la distribution de Fisher-Snedecor de paramètres 14 et 9, les intensités sont bien comparables.

Comme l'intensité des accidents est la même pour les deux compagnies, cette analyse permet donc de mettre en avant le fait que les clients de l'assureur concerné affiliés chez l'une ou l'autre des deux compagnies se comportent de la même façon au niveau des assurances voitures et incendies.

Radar automatique du tunnel de Cointe

Les données qui sont à notre disposition correspondent aux temps (exprimés en minutes) auxquels un véhicule trop rapide s'est fait flasher sous le tunnel de Cointe le 1^{er} janvier 2006 ; elles sont au nombre de 32.

Le test de Kolmogorov-Smirnov sur les temps d'inter-arrivées donne une statistique $D_{32} = 0,0923$ qui est inférieure au seuil $d_{32} = 0,1872$. Ceci implique que l'on peut supposer le caractère poissonien pour ces données avec une intensité comprise dans l'intervalle $[0,016; 0,032]$ et dont l'estimation du maximum de vraisemblance vaut 0,023. La méthode graphique donne la figure 7.13.

L'espérance du nombre de véhicules flashés sur une heure de temps vaut donc $0,023 * 60 = 1,38$. Ainsi, si un policier veut verbaliser (pour excès de vitesse) 10 conducteurs sur sa journée de travail, il lui suffit de se poster près du radar fixe du tunnel de Cointe pendant environ 7h15 et d'arrêter tout véhicule se faisant flasher par le radar.

Appels et visites arrivant à l'accueil d'une école

Les appels téléphoniques arrivant à une centrale sont réputés former un processus de Poisson ; c'est un exemple souvent cité. C'est pour cette raison qu'il est étudié dans cette section. Les données ont été récoltées à l'accueil du Collège Saint-François d'Assise à Ans

entre 9h et 13h un jour quelconque de la semaine.

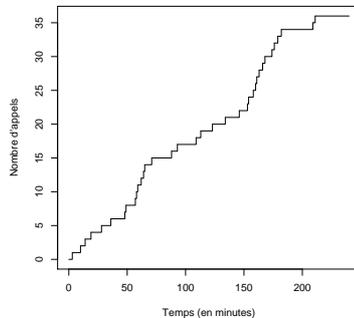


FIG. 7.14 – Évolution du nombre d'appels en fonction du temps

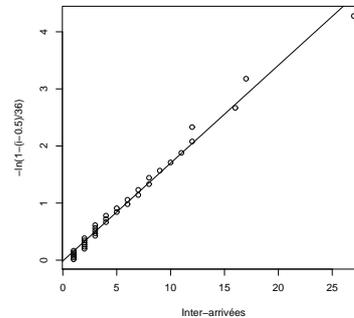


FIG. 7.15 – Qq-plot des temps d'inter-arrivées des appels

Il y a 36 données et le test de Kolmogorov-Smirnov sur les temps d'inter-arrivées permet de conclure qu'elles sont issues d'un processus de Poisson car la statistique D_{36} de ce test est de 0,1569 et le seuil critique d_{36} vaut 0,1769. L'ajustement linéaire représenté à la figure 7.15 permet de voir que le modèle exponentiel est relativement bon et qu'une estimation de l'intensité peut-être donnée par 0,17 qui est la pente de la droite d'ajustement.

La connaissance de ce fait permet notamment de calculer les probabilités d'avoir un certain nombre d'appels durant n'importe quel intervalle de temps. Nous allons ainsi rechercher combien de temps l'employée de l'accueil peut quitter son bureau sans risquer de rater trop d'appels. Pour cela, un programme a été développé dans le logiciel R. Pour une intensité donnée, un nombre a compris entre 0 et 1 et un temps t , il rend la plus longue durée h inférieure ou égale à t durant laquelle la probabilité de ne pas recevoir d'appel est plus grande que ce nombre a .

Dans notre exemple, l'intensité vaut 0,17, et pour $a = 0,5$ et $t = 10$, le programme rend $h = 4$. Cela signifie que la personne qui travaille à l'accueil peut s'absenter pendant 4 minutes durant lesquels la probabilité de rater un ou plusieurs appels est plus petite que 0,5. Plus on est exigeant quant à la probabilité de ne pas rater d'appel, c'est-à-dire plus a est grand, plus cette durée h est petite. Pour $a = 0,8$, h vaut 1, et pour $a \geq 0,85$, h est nul.

Si l'on regarde maintenant les données concernant les visites reçues au bureau d'accueil durant la même période, qui sont représentées par le graphique 7.16, on voit que deux personnes sont arrivées en même temps après la 200^e minute. En supprimant cette donnée, il en reste 15 qui correspondent au processus sous-jacent. Le résultat du test est l'acceptation de l'hypothèse nulle car $D_{15} = 0,2644 < 0,2686 = d_{15}$. On peut donc supposer que ces données constituent un processus de Poisson dont l'estimation par maximum de vraisemblance de l'intensité vaut 0,06. Le test d'indépendance du χ^2 entre le processus sous-jacent et la famille de variables donnant le nombre de personnes arrivées lors de chaque occurrence nous permet de conclure à l'indépendance (car la p-value vaut 0,3563) et donc au fait que le processus des arrivées de personnes au bureau d'accueil est bien un exemple de processus de Poisson composé.

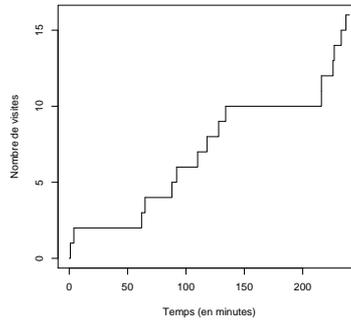


FIG. 7.16 – Évolution du nombre de visites au cours du temps

Accidents sur les autoroutes de la province de Liège

Intéressons-nous maintenant aux données relatives aux accidents qui se sont produits sur les autoroutes de la province de Liège entre le 1^{er} janvier et le 1^{er} juin 2006, c'est-à-dire sur 4344 heures. Ces données sont exprimées en heure et sont au nombre de 147. Le test de Kolmogorov-Smirnov sur les temps d'inter-arrivées du processus donne une statistique D_{147} de 0,094 qui est légèrement plus grande que le seuil $d_{147} = \frac{1,08}{\sqrt{147}} = 0,089$. Ceci signifie donc que le processus ne peut pas être considéré comme un processus de Poisson.

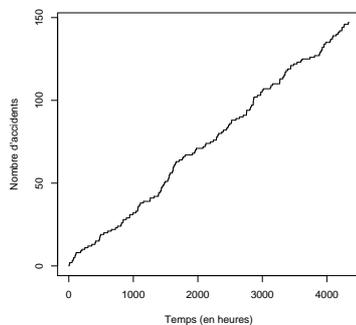


FIG. 7.17 – Évolution du nombre d'accidents au cours du temps

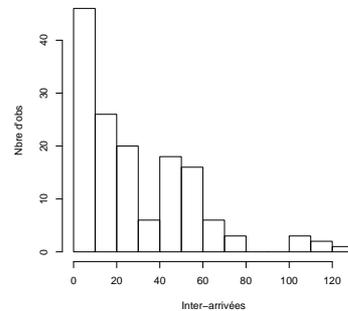


FIG. 7.18 – Histogramme des temps d'inter-arrivées des accidents sur les autoroutes de la provinces de Liège

Il est aisé de croire que l'intensité des accidents se produisant sur les autoroutes est différente selon le trafic (heures d'affluence, nuit, ...). Peut-être qu'un modèle non homogène conviendrait mieux. Comme une seule réalisation est à notre disposition, il n'est pas possible de tester cette hypothèse. En supposant que ce modèle conviendrait, on peut donner l'estimation \hat{m}_1 de la fonction moyenne qui est représentée à la figure 7.19. Comme cela a été expliqué à la page 83, cette estimation est proche de l'évolution du nombre d'accidents donnée par le graphique 7.17.

La figure 7.20 montre quant à elle que le modèle stationnaire qui a été rejeté est fort semblable

au modèle non stationnaire que l'on pourrait tester sur les données. Cela est dû au fait que le rejet du modèle stationnaire était assez limite.

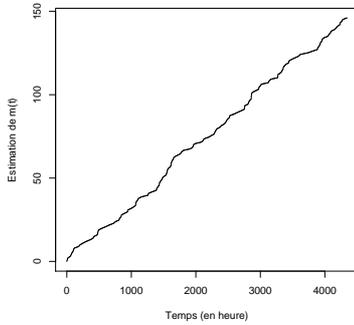


FIG. 7.19 – Estimation de la fonction moyenne des accidents sur les autoroutes de la province de Liège

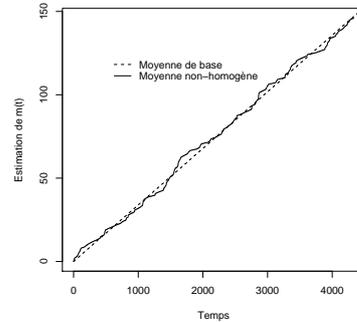


FIG. 7.20 – Comparaison des modèles homogène et non homogène

Païement par bancontact à la caisse d'un supermarché

Les données concernent les paiements effectués par bancontact à l'AD delhaize de Pepinster sur les quatre heures du dimanche matin. Quatre échantillons, de taille respective 54, 62, 63 et 64, sont disponibles ; ils correspondent à quatre dimanches différents. Cependant aucun des quatre échantillons ne permet de conclure que ce processus serait un processus de Poisson car les valeurs de la statistique sont trop élevées ($D_{54} = 0,2159$, $D_{62} = 0,2302$, $D_{63} = 0,2458$ et $D_{64} = 0,2438$). Les histogrammes 7.21, 7.22, 7.23 et 7.24 illustrent cela.

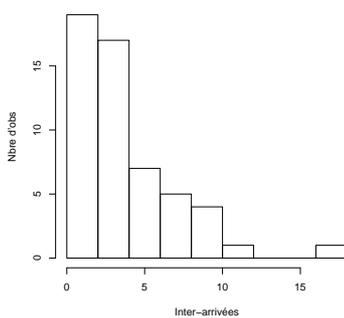


FIG. 7.21 – Histogramme des temps d'inter-arrivées des paiements bancontact le 3/12/06

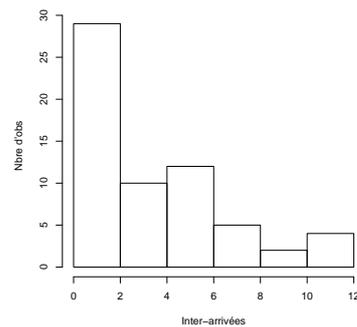


FIG. 7.22 – Histogramme des temps d'inter-arrivées des paiements bancontact le 10/12/06

Nous allons essayer de voir si le modèle d'un processus de Poisson non homogène conviendrait mieux. Cette fois, trois échantillons sont à notre disposition pour estimer la fonction

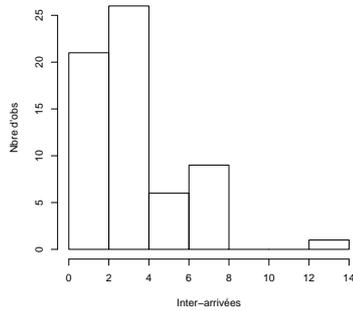


FIG. 7.23 – Histogramme des temps d’inter-arrivées des paiements bancontact le 14/01/07

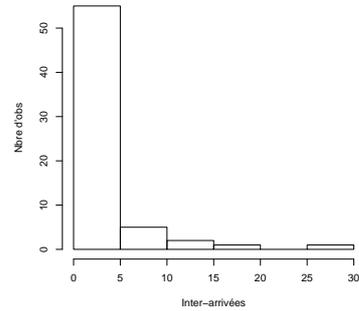


FIG. 7.24 – Histogramme des temps d’inter-arrivées des paiements bancontact le 18/02/07

moyenne, ce qui donne l’estimation représentée par la figure 7.25. Il faut vérifier que le vecteur des temps d’occurrence du quatrième jour a pour distribution une loi dont la fonction de répartition est $F(t) = \frac{\hat{m}(t)}{\hat{m}(240)}$. Comme la p-value du test du χ^2 vaut 0.13, cette distribution peut être acceptée et l’on peut alors conclure que les paiements par bancontact le dimanche matin à l’AD delhaize de Pepinster se font suivant un processus de Poisson dont l’estimation de la fonction moyenne est donnée par le graphique 7.25. La figure 7.26 permet de voir que le modèle d’un processus de Poisson homogène est différent du modèle accepté. La droite discontinue représente la fonction moyenne d’un processus de Poisson de base d’intensité estimée sur le quatrième dimanche.

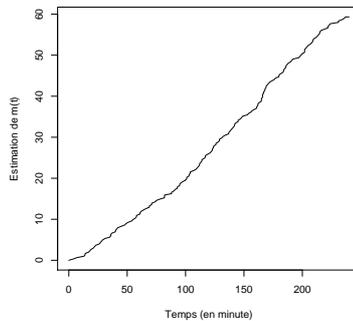


FIG. 7.25 – Estimation de la fonction moyenne des paiements bancontact au delhaize de Pepinster

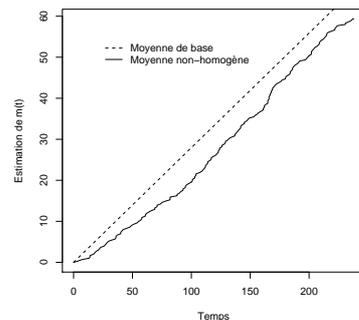


FIG. 7.26 – Comparaison des modèles homogène et non homogène

Tous ces exemples montrent bien que la théorie développée sur les processus de Poisson n’est pas simplement une belle théorie probabiliste mais qu’elle s’applique à la vie de tous les jours. Cependant, mis à part le cas d’un processus de Poisson de base, les étapes sont nombreuses pour arriver à déterminer si un ensemble de temps d’inter-arrivées est un processus de Poisson. C’est pourquoi il était intéressant de développer un programme dans R, appelé

`Poisson(w,k,e)`, qui fait toutes les étapes intermédiaires et rend la conclusion qui s'impose. Celui-ci est repris en annexe. Ses paramètres sont l'échantillon `w` des temps d'inter-arrivées que l'on souhaite tester, le nombre `k` de réalisations à utiliser pour estimer $m(t)$ et le vecteur `e` des temps d'occurrence des `k` réalisations intervenant dans l'estimation que l'on a ordonnés chronologiquement sans tenir compte de la réalisation dont ils étaient issus. Il est possible de ne spécifier que le vecteur `w` et de laisser `k` et `e` avec leur valeur par défaut `NULL`; dans ce cas, le programme affichera soit qu'il s'agit d'un processus de Poisson de base ou composé, soit un avertissement comme de quoi le programme a besoin des arguments `k` et `e` pour vérifier l'hypothèse d'un processus de Poisson non homogène. Avec l'avertissement, une estimation de la fonction moyenne sur l'unique réalisation disponible s'affiche. Attention que cette estimation suppose que le modèle d'un processus de Poisson non homogène soit correct, ce que l'on ne sait pas vérifier sur une réalisation. La fonction `ks.test2` réalise le test d'ajustement habituel de Kolmogorov-Smirnov sans afficher l'avertissement concernant le calcul des p-values exactes puisqu'elles n'interviennent pas dans les décisions.

Comme il est très facile maintenant de voir si l'on est en présence d'un processus de Poisson, considérons encore les quelques exemples suivants.

Arrivées de clients à la cafétéria du B37

```
> Poisson(Inter_cafet)
Warning message:
Il s'agirait peut-être d'un processus de Poisson non homogène, veuillez
relancer la fonction avec les paramètres k et e pour confirmer cette hypothèse.
in: Poisson(Inter_cafet)
```

Il ne s'agit ni d'un processus de Poisson de base ni composé. Comme une seule réalisation est disponible, il n'est pas possible de voir si c'est un processus de Poisson non homogène. Mais il est cependant possible de voir l'estimation de la fonction moyenne de ce processus s'il était non homogène. Cette estimation est donnée à la figure 7.27. Elle est assez différente de l'estimation obtenue par un modèle stationnaire car, cette fois-ci, le rejet est plus marqué (la statistique du test de Kolmogorov-Smirnov est $D_{170} = 0,1675$ et le seuil critique vaut $0,083$). L'exemple des arrivées de clients dans un service de repas avait déjà été introduit au début du chapitre 3. Le graphique 7.27 a bien la même forme que celle prédite pour ce genre de système et représentée à la figure 3.6, c'est-à-dire avec une pente beaucoup plus faible aux extrémités qu'au centre.

Déchargement d'une benne de terrassement

Le déchargement se fait à intervalle relativement régulier, ce qui laisse imaginer que le modèle d'un processus de Poisson de base serait bon. Cependant, si l'on regarde les déchargements effectués sur une journée entière, cela ne fonctionne pas. Ce problème est justement dû au fait que le déchargement est très régulier (environ toutes les 10 minutes) mais qu'un des temps d'inter-arrivées vaut 42 minutes, correspondant à la pause de midi du conducteur. Ce temps est beaucoup trop grand par rapport aux autres et fait ainsi rejeter l'hypothèse d'un processus de Poisson de base. Par contre, un processus de Poisson non homogène pourrait être imaginé avec l'estimation de la fonction moyenne donnée par le graphique 7.28. Cette

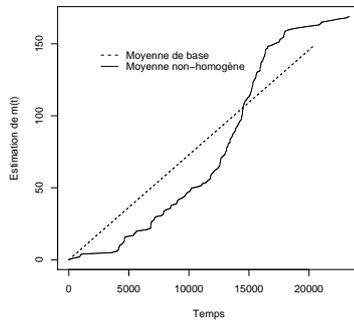


FIG. 7.27 – Estimations des fonctions moyenne pour les arrivées à la cafétéria

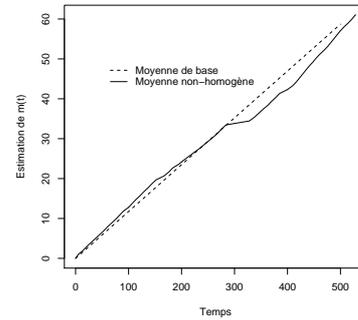


FIG. 7.28 – Estimations des fonctions moyenne pour les déchargements d’une benne

estimation est linéaire (comme pour un processus de Poisson de base) sauf sur le temps de midi où elle est constante.

Arrivées des enfants à la crèche

```
> Poisson(Inter_crèche01)
```

Ce vecteur correspond aux temps d’inter-arrivées d’un processus de Poisson composé d’intensité 0.05652174 et de famille de variables indépendantes 1 1 2 4 3 2 2 1 5 2 2 1 1 .

Sur une matinée, les arrivées des enfants à la crèche du Tilleul (Ciplet) constituent un processus de Poisson composé d’intensité $\nu = 0.057$ et de famille de variables Y_n données par le vecteur 1, 1, 2, 4, 3, 2, 2, 1, 5, 2, 2, 1, 1.

Beaucoup d’autres exemples pourraient encore être traités comme, par exemple, les fautes d’orthographe ou de dactylographie dans un texte, les assurances-vie, les rayons émis par une substance radio-active, les tremblements de terre, . . . Ceux abordés dans le cadre de ce travail montrent déjà que le processus de Poisson et ses généralisations s’adaptent bien à la réalité.

Annexe A

Logiciel R

Valeurs du seuil critique du test de Kolmogorov-Smirnov pour une distribution exponentielle de moyenne inconnue

n	d	n	d	n	d
2	0.6132737	35	0.1793248	68	0.1298778
3	0.5507922	36	0.1768967	69	0.1289573
4	0.4843500	37	0.1745687	70	0.1280568
5	0.4420259	38	0.1723343	71	0.1271696
6	0.4085341	39	0.1701874	72	0.1263011
7	0.3811394	40	0.1681225	73	0.1254506
8	0.3589628	41	0.1661189	74	0.1246175
9	0.3404000	42	0.1641880	75	0.1238012
10	0.3243864	43	0.1623256	76	0.1230012
11	0.3103306	44	0.1605277	77	0.1222170
12	0.2981148	45	0.1587906	78	0.1214480
13	0.2871683	46	0.1571024	79	0.1206938
14	0.2774439	47	0.1554687	80	0.1199539
15	0.2685914	48	0.1538869	81	0.1192256
16	0.2606000	49	0.1523543	82	0.1185107
17	0.2532314	50	0.1508683	83	0.1178089
18	0.2464974	51	0.1494183	84	0.1171197
19	0.2402442	52	0.1480107	85	0.1164428
20	0.2344741	53	0.1466434	86	0.1157779
21	0.2290961	54	0.1453146	87	0.1151245
22	0.2240953	55	0.1440226	88	0.1144824
23	0.2193780	56	0.1427656	89	0.1138512
24	0.2149631	57	0.1415422	90	0.1132306
25	0.2108000	58	0.1403508	91	0.1126183
26	0.2068829	59	0.1391902	92	0.1120160
27	0.2031696	60	0.1380589	93	0.1114236
28	0.1996597	61	0.1369482	94	0.1108407
29	0.1963264	62	0.1358647	95	0.1102670
30	0.1931635	63	0.1348073	96	0.1097024
31	0.1901266	64	0.1337750	97	0.1091467
32	0.1872348	65	0.1327668	98	0.1085995
33	0.1844771	66	0.1317818	99	0.1080607
34	0.1818434	67	0.1308190	100	0.1075300

Programmation

Programme utilisé pour obtenir la figure 1.19

```
bootstrap3<-function(i,I){
  taille<-1:i
  quant<-1:i
  par<-runif(1,0.1,10)
  M_sim<-NULL
  for (n in 1:i){
    N_res<-NULL
    x_sim<-1:I
    x_sim<-rexp(I,par)
    M_sim<-rbind(M_sim,x_sim)
    for ( j in 1:I){
      stat<-sqrt(n)*ks.test(M_sim[,j],"pexp",1/mean(M_sim[,j]))$statistic
      N_res<-c(N_res,stat)
    }
    quant[n]<-quantile(N_res,0.95)
  }
  plot(taille,quant,xlab="n",ylab="n^(1/2)*d_n",type="l")
}
```

Programme concernant l'exemple sur les appels téléphoniques :

```
proba<-function(lambda,a,t){
  d<-0
  for (h in 0:t){
    if (dpois(0,lambda*h)>a){
      d<-h
    }
  }
  return(d)
}
```

Programmes intervenant dans le programme final :

```
estimation<-function(t,k,vecteur){
  n<-length(vecteur)-1
  if (t>=vecteur[n]){
    estimation<-(((n-1)*(n-1))/(n*k))
  }
  else{
    for (i in 1:n){
      if ( (vecteur[i]<=t) & (t<vecteur[i+1]) ){
        estimation<-(((i-1)*(n-1))/(n*k))+(((n-1)*(t-vecteur[i]))/(n*k*(vecteur[i+1]-vecteur[i])))
      }
    }
  }
}
```

```
return(estimation)
}
```

```
est_vect<-function(t,k,vecteur){
est_vect<-rep(0,times=length(t))
for(i in 1:length(t)){
est_vect[i]<-estimation(t[i],k,vecteur)
}
return(est_vect)
}
```

```
repartition<-function(t,k,vecteur){
n<-length(vecteur)-1
return(estimation(t,k,vecteur)/estimation(vecteur[n],k,vecteur))
}
```

```
rep_vect<-function(t,k,vecteur){
rep_vect<-rep(0,times=length(t))
for(i in 1:length(t)){
rep_vect[i]<-repartition(t[i],k,vecteur)
}
return(rep_vect)
}
```

```
class<-function(x){
Sturges<-1+3.322*log(length(x),10)
class<-rep(0,times=Sturges+2)
long<-max(x)/Sturges
for (i in 2:(Sturges+2)){
class[i]<-class[i-1]+long
}
return(class)
}
```

```
NB<-function(x){
class<-class(x)
NB<-rep(0,times=(length(class)-1))
for (i in 1:length(NB)){
for (j in 1:length(x)){
if ((class[i]<=x[j])&(x[j]<class[i+1])){
NB[i]<-NB[i]+1
}
}
}
return(NB)
}
```

```

vect_p<-function(x,k,y){
Sturges<-1+3.322*log(length(x),10)
p<-rep(0,times=Sturges+1)
w<-c(0,y,max(y)+1)
class_x<-class(x)
p[1]<-repartition(class_x[2],k,w)-repartition(0,k,w)
for (i in 2:length(p)){
p[i]<-repartition(class_x[i+1],k,w)-repartition(class_x[i],k,w)
}
return(p)
}

```

```

test_chisq<-function(x,k,y){
vect_class<-class(x)
obs<-NB(x)
vect_proba<-vect_p(x,k,y)
th<-vect_proba*sum(obs)
l<-length(th)-1
h<-length(th)
while (min(th)<5){
for (i in 1:l){
if (i<h){
if (th[i]<5){
th[i+1]<-th[i+1]+th[i]
obs[i+1]<-obs[i+1]+obs[i]
vect_proba[i+1]<-vect_proba[i+1]+vect_proba[i]
if (i==1){
vect_class<-c(vect_class[1],vect_class[3:h])
th<-th[2:h]
obs<-obs[2:h]
vect_proba<-vect_proba[2:h]
}
else{
th<-c(th[1:i-1],th[i+1:h])
obs<-c(obs[1:i-1],obs[i+1:h])
vect_class<-c(vect_class[1:i],vect_class[i+2:h+1])
vect_proba<-c(vect_proba[1:i-1],vect_proba[i+1:h])
}
}
h<-h-1
th<-th[1:h]
obs<-obs[1:h]
vect_proba<-vect_proba[1:h]
}
}
}
if(th[h]<5){
th[h-1]<-th[h-1]+th[h]
}
}
}
}

```

```

obs[h-1]<-obs[h-1]+obs[h]
vect_proba[h-1]<-vect_proba[h-1]+vect_proba[h]
th<-th[1:h-1]
obs<-obs[1:h-1]
vect_proba<-vect_proba[1:h-1]
vect_class<-c(vect_class[1:h-1],vect_class[h+1])
}
}
M<-rbind(obs,vect_proba)
return(M)
}

m<-function(t,k){
T<-c(0,t,(max(t)+1))
s<-seq(0,max(T),by=0.1)
est<-est_vect(s,k,T)
plot(s,est,xlab="Temps",ylab="Estimation de m(t)",type="l")
}

```

Programme final :

```

Poisson<-function(w,k=NULL,e=NULL){
alpha=0.05
n<-length(w)
if(min(w)>0){
t<-1:n
t[1]<-w[1]
for (i in 2:n){t[i]<-t[i-1]+w[i]}
if(n<=100){
if(ks.test2(w,"pexp",1/mean(w))$statistic<d[n]){
cat("Ce vecteur correspond bien aux inter-arrivées d'un processus de Poisson
d'intensité ",1/mean(w),".","\n")}
else {
if ((is.null(k)==FALSE)&(is.null(e)==FALSE)){
M<-test_chisq(t,k,e)
if(chisq.test(M[1,],p=M[2,])$p.value>alpha){
cat("Ce vecteur correspond aux inter-arrivées d'un processus de Poisson
non homogène dont voici la représentation de la fonction moyenne.", "\n")
m(e,k)
}
}
else{cat("Ce vecteur ne correspond à aucun processus de Poisson","\n")}
else{warning("Il s'agirait peut-être d'un processus de Poisson non homogène
(dont voici une estimation de la fonction moyenne), veuillez relancer la
fonction avec les paramètres k et e pour confirmer cette hypothèse.", "\n")
m(t,1)
}
}
}
}
}
}

```

```

else{
if(ks.test2(w,"pexp",1/mean(w))$statistic<1.08/sqrt(n)){
cat("Ce vecteur correspond bien aux inter-arrivées d'un processus de Poisson
d'intensité ",1/mean(w),".","\n")}
else{
if((is.null(k)==FALSE)&(is.null(e)==FALSE)){
M<-test_chisq(t,k,e)
if(chisq.test(M[1,],p=M[2,])$p.value>alpha){
cat("Ce vecteur correspond aux inter-arrivées d'un processus de Poisson
non homogène dont voici la représentation de la fonction moyenne.", "\n")
m(e,k)
}
else{cat("Ce vecteur ne correspond à aucun processus de Poisson.", "\n")}}
else{warning("Il s'agirait peut-être d'un processus de Poisson non homogène
(dont voici une estimation de la fonction moyenne), veuillez relancer la
fonction avec les paramètres k et e pour confirmer cette hypothèse.", "\n")
m(t,1)
}
}
}
}
}
else{
W<-w[w>0]
N<-length(W)
y<-rep(1,times=N)
i<-1:n
I<-i[w>0]
for (j in 1:N-1){y[j]<-I[j+1]-I[j]}
if (max(I)!=n){y[N]<-y[N]+n-max(I)}
T<-1:N
T[1]<-W[1]
for (i in 2:N){T[i]<-T[i-1]+W[i]}
if(N<=100){
if(ks.test2(W,"pexp",1/mean(W))$statistic<d[N]){
if(chisq.test(W,y,simulate.p.value=TRUE)$p.value>alpha) {
cat("Ce vecteur correspond aux temps d'inter-arrivées d'un processus de
Poisson composé d'intensité ",1/mean(W)," et de famille de
variables indépendantes ", y, ".","\n")}
else{cat("Le processus sous-jacent est un processus de Poisson d'intensité
estimée à ",1/mean(W)," mais le processus n'est pas un processus de Poisson
composé.", "\n")}}
else{ cat("Le processus sous-jacent n'est pas un processus de Poisson.", "\n")}}
else{
if(ks.test2(W,"pexp",1/mean(W))$statistic<1.08/sqrt(N)){
if(chisq.test(W,y,simulate.p.value=TRUE)$p.value>alpha){
cat("Le processus est un processus de Poisson composé d'intensité ",1/mean(W),"
et de famille", y, ".","\n")}

```

```

else{
cat("Le processus sous-jacent est un processus de Poisson d'intensité ",1/mean(W),"
mais le processus n'est pas un processus de Poisson composé.", "\n")}}
else {cat("le processus sous-jacent n'est même pas un processus de Poisson.", "\n")}}
}

```

Analyse de données

Données issues de la littérature

Les explosions dans les mines de Grande-Bretagne impliquant le décès de dix personnes ou plus

```

> Inter_mine
378  36  15  31 215  11 137   4  15  72  96 124  50 120 203
176  55  93 59 315  59  61   1  13 189 345  20  81 286 114
108 188 233  28  22  61  78  99 326 275  54 217 113  32  23
151 361 312 354  58 275  78  17 1205 644 467 871  48 123 457
498  49 131 182 255 195 224 566 390  72 228 271 208 517 1613
54  326 1312 348 745 217 120 275  20  66 291  4 369 338 336
19  329 330 312 171 145  75 364  37  19 156  47 129 1630  29
217   7  18 1357

```

Les accidents dans une région d'une mine en Grande-Bretagne

```

> Inter_mine2
3 23  0  0  2  3  0  2  1  0  1  0  0  0  1  0  1  3  2  1  2  0  0  2  3
0  0  0  0  0  3  4  2  3  5  0  0  2  4  5  0  4  0  0  2  1  1  1  1  8
0  0  2  0  2  2  0  2  0  0  2  8  8  2  0  1  1  8  0  0  0  1  8  3  0  2
0  1  1  3  2  0  4  3 12 10  4  0  3  2  2  0  0 14  1  4  8

```

Données observées

Sinistres subis par des personnes assurées auprès d'un particulier

Les données relatives aux accidents de voitures de la première compagnie sont :

```

> Inter_assauto1
5 27 26  3  3  3  1 29 14  3  1  0  3 11  4 35  6 15 34  2  2 14 17  1  1  6
0  8  4  0 17  8 23 21 13  4 18  8  5 22  3  5  1  2  5 16  7  7  4  7 19  5
14  1  0 15  2 11  1  0  1  1 15  2  7  0  0  2  4  3  1  3 21  0 15  2  3  5
10  1 11  3 49  2  9  0  2  5  6  3

```

Les données relatives aux accidents de voitures de la seconde compagnie sont :

```

> Inter_assauto2
12 11 25 10 26  0  8  8  5 13 12 13 21 24 12  7  4 44  8 18 26  5  6  6 30
0  3  0 34 59  3  1  0 22 23  3  4 34  5  7 14  0 20  3  3 32  3 25  1 21
13  0 14 14  2  8

```

Voici alors les données concernant les incendies pour la première compagnie d'assurance :

```
> Inter_assinc1
4 35 22 53 8 131 44 3 31 79 72 6 31 1 40 17 3 31 9 57
```

Et pour la deuxième compagnie :

```
> Inter_assinc2
62 18 0 59 135 31 32 12 22 26 61 83 46 5 36 45
```

Radars automatiques du tunnel de Cointe

```
> Inter_flash01
34 95 80 104 9 19 5 1 8 3 85 26 20 35 25 5 31 12
45 17 55 92 87 35 62 150 26 17 1 119 25 58
```

Appels et visites arrivant à l'accueil d'une école

Voici les données concernant les appels téléphoniques :

```
> Inter_tel
3 7 4 5 9 8 12 1 8 1 1 3 2 1 6 17 5 16 4 10 11 12
7 1 4 2 1 2 3 2 6 2 3 3 27 2
```

Et pour les visites :

```
> Inter_visite<-c(1,3,58,3,23,4,18,8,10,6,82,0,10,1,6,4)
```

Accidents sur les autoroutes de la province de Liège

```
> Inter_acc
7 8 37 15 10 19 8 11 68 22 44 49 60 50 9 54 8 5 10
50 61 55 64 31 55 7 25 3 48 50 1 53 43 28 3 5 21 13
48 102 3 58 68 3 20 19 4 8 23 13 14 41 12 8 5 9 33
14 2 6 14 15 10 48 50 16 27 112 36 10 14 106 32 12 72 43
47 6 16 14 50 23 45 23 11 20 26 3 68 54 58 51 1 2 51
9 12 28 1 2 3 8 51 52 2 22 16 111 10 25 109 1 4 45
14 17 4 15 26 44 1 44 55 56 22 121 71 75 6 18 17 3 26
3 24 68 10 28 6 54 28 21 36 2 29 9 60
```

Paiement par bancontact à la caisse d'un supermarché

```
> Inter_delhaize1
4 10 3 5 4 1 9 1 3 17 4 5 5 11 6 2 3 8 3 3 4 1 7 4
1 5 5 3 1 2 1 1 3 2 2 5 7 1 1 2 1 1 1 2 4 3 9 4
10 2 7 7 3 3
> Inter_delhaize2
7 6 5 5 5 12 1 8 5 1 3 3 3 5 5 1 7 11 2 1 2 3 2 1
5 3 1 1 2 1 5 2 2 2 1 2 12 10 8 4 1 3 11 1 2 1 1 1
6 5 5 2 4 1 4 1 1 7 9 3 2 1
```

```

> Inter_delhaize3
 14 5 2 8 2 4 1 6 3 3 3 7 4 7 2 7 4 6 3 4 5 3 7 4
 3 4 2 3 6 5 2 3 3 3 8 1 2 3 3 1 1 1 1 2 1 4 4 4
 3 1 3 7 1 3 1 2 2 3 2 3 8 1 7
> Inter_delhaize4
 4 30 1 1 14 2 3 1 2 3 1 4 5 5 1 3 1 1 4 7 2 5 1 3
 3 18 1 3 6 1 1 1 3 1 1 1 3 8 3 1 2 1 3 2 1 1 2 9
 1 1 1 1 5 1 8 3 2 2 1 3 1 14 4 1

```

Arrivées des clients à la cafétéria du B37

```

> Inter_cafet
300 600 60 80 2534 421 142 2 75 30 180 112 90 4 3
80 675 50 147 32 850 328 34 1 10 36 57 183 15 101
524 116 21 44 203 333 12 35 445 30 47 85 277 191 101
77 122 210 40 70 552 312 4 385 47 207 67 10 20 88
114 94 152 140 102 66 9 16 25 24 20 120 148 84 11
87 14 6 85 25 27 148 1 1 68 103 22 58 94 95
3 47 92 71 58 58 64 118 1 11 57 3 14 6 25
19 1 47 31 68 23 149 57 131 66 38 29 60 13 38
9 46 10 8 73 53 96 27 10 43 31 254 8 1 25
35 7 25 113 47 24 80 28 35 14 2 36 155 1 376
226 341 40 69 180 119 2 49 8 117 378 639 815 900 166
83 690 600 779 197

```

Déchargement d'une benne de terrassement

```

> Inter_benne
 5 9 7 8 8 8 8 7 8 8 7 7 10 7 7 8 7 8 7 8 15 9 7 12 9
 10 10 10 9 9 9 8 7 8 44 10 8 8 9 7 8 7 15 10 7 6 6 6 6 7
 6 7 8 7 6 6 6 6 7 8 7 6

```

Arrivées des enfants à la crèche

```

> Inter_crèche
20 10 30 0 10 0 0 0 10 0 0 5 0 10 0 5 10 0 0 0 0 25 0 35 0 20 40

```

Bibliographie

- [1] BASTIN F., *Analyse non-linéaire*, Université de Liège, 2004-2005, Notes de cours.
- [2] BELISLE C., *Processus de Poisson*, Université de Laval, Notes de cours.
- [3] BRAOUEZEC Y., *Compléments d'ingénierie financière : introduction au processus stochastiques à sauts*, ESILV, 2004, Notes de cours.
- [4] CIZEK P., HÄRDLE W. and WERON R., *Statistical Tools for Finance and Insurance*, 2005, Consulté sur Internet le 27/11/2006 via <http://www.quantlet.com/mdstat/scripts/stf/html/stfhtmlframe172.html>.
- [5] DURBIN J., Kolmogorov-Smirnov tests when parameters are estimated with applications to tests of exponentiality and tests on spacings, *Biometrika*, 1975, Volume **62**, n°1, pp5-22.
- [6] EFRON B. & TIBSHIRANI R.J., *An introduction to the bootstrap*, New-York, Chapman & Hall, 1993.
- [7] HAESBROECK G., *Éléments de statistique mathématique, Probabilités et statistique mathématique*, Université de Liège, 2004-2005, Notes de cours.
- [8] HAESBROECK G., *Éléments de statistique mathématique, Statistique inférentielle*, Université de Liège 2004-2005, Notes de cours.
- [9] HAESBROECK G., *Introduction aux processus stochastiques*, Université de Liège, 2005-2006, Notes de cours.
- [10] HOEL P.G., *Introduction to mathematical statistics*, Second edition, New-York, John Wiley & Sons, 1958.
- [11] KINGMAN J.F.C., *Poisson Processes*, Oxford, Clarendon Press, 1993.
- [12] KARLIN S., *Initiation aux processus aléatoires*, Paris, Dunod, 1969, Traduit par F. ROSTAND.
- [13] KARLIN S. & TAYLOR H.M., *A first course in stochastic processes*, Second edition, New-York, Academic Press, 1975.
- [14] KARLIN S. & TAYLOR H.M., *A second course in stochastic processes*, New-York, Academic Press, 1981.
- [15] KOLMOGOROV A.N., *Foundations of the theory of probability*, Second English edition, New-York, Chelsea publishing company, 1956, Traduit par N. MORRISON.
- [16] KOTS S. & JOHNSON N.L., *Breakthroughs in Statistics, Methodology and distribution*, New-York, Springer-Verlag, 1993.
- [17] LAMBERT R., *Le L^AT_EX*, Cinquième édition, Liège, Les Éditions de l'Université de Liège, 2005.

- [18] LEEMIS L.M., *Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process*, 1990, Consulté sur Internet via <http://www.math.wm.edu/leemis/1991mgmtsci.pdf> le 27/11/2006.
- [19] MAGIS D., *Introduction au logiciel R*, Université de Liège, 2003-2004, Notes de cours.
- [20] MAGUIRE B.A., PEARSON E.S. and WYNN A.H.A., The time intervals between industrial accidents, *Biometrika*, 1952, volume **39**, pp168-180.
- [21] MAWHIN J., *Analyse : Fondements, techniques, évolution*, Bruxelles, De Boeck-Wesmael, 1992.
- [22] PARZEN E., *Stochastic processes*, San Francisco, Holden-Day, 1967.
- [23] ROSS S.M., *Introduction to probability models*, Second edition, New-York, Academic Press, 1980.
- [24] ROSS S.M., *Stochastic processes*, Second edition, New-York, John Wiley & Sons, 1996.
- [25] SCHMETS J., *Analyse mathématique*, Liège, Derouaux, 1996-1997.
- [26] SCHMETS J., *Analyse mathématique, Introduction au Calcul Intégral*, Liège, Derouaux, 1994.
- [27] SCHMETS J., *Analyse mathématique, Introduction aux espaces fonctionnels*, Université de Liège, 2000-2001, Notes de cours.