

Record Number: 16620
Author, Monographic: Yagouti, A.//Ouarda, T. B. M. J.//Bobée. B.
Author Role:
Title, Monographic: Étude des tests de séries d'événements pour l'analyse du risque
Translated Title:
Reprint Status:
Edition:
Author, Subsidiary:
Author Role:
Place of Publication: Québec
Publisher Name: INRS-Eau
Date of Publication: 1999
Original Publication Date: 30 avril 1999
Volume Identification:
Extent of Work: 101
Packaging Method: pages incluant un annexe
Series Editor:
Series Editor Role:
Series Title: INRS-Eau, rapport de recherche
Series Volume ID: 546
Location/URL:
ISBN: 2-89146-322-6
Notes: Rapport annuel 1998-1999
Abstract: Réalisé pour le centre de recherches pour la Défense Valcartier
BNC et BNQ aucun prix
Call Number: R000546
Keywords: rapport/ ok/ dl

**ÉTUDE DES TESTS DE SÉRIES
D'ÉVÉNEMENTS POUR L'ANALYSE
DU RISQUE**

**ÉTUDE DES TESTS DE SÉRIES D'ÉVÉNEMENTS
POUR L'ANALYSE DU RISQUE**

Rapport final présenté à

**Irène Abi-Zeid
Centre de Recherches pour la Défense Valcartier**

Contrat XSK-8-00191 (303)

Rapport rédigé par

**Abderrahmane Yagouti
Taha B.M.J. Ouarda
Bernard Bobée**

**Chaire industrielle Hydro-Québec / CRSNG en Hydrologie statistique
Institut national de la recherche scientifique, INRS-Eau
2800 rue Einstein, case postale 7500, SAINTE-FOY (Québec) G1V 4C7**

Rapport de recherche No R-546

30 avril 1999

Table de matières

1. INTRODUCTION	1
2. PROCESSUS PONCTUELS ET RISQUE.....	3
2.1 Processus ponctuels	3
2.1.1 Définitions	3
2.1.2 Processus de Poisson	5
2.1.2.1 Le processus de Poisson homogène (PPH)	5
2.1.2.2 Le processus de Poisson non-homogène (PPNH).....	6
2.1.2.3 Processus de Poisson composé ou multiple	6
2.1.2.4 Processus de Poisson groupé (PPG) "clusters process"	6
2.1.2.5 Processus de Poisson doublement stochastique	7
2.1.2.6 Processus linéaire auto-excité	7
2.1.3 Le Processus de renouvellement	8
2.1.3.1 Processus de renouvellement homogène (PRH)	8
2.1.3.2 Processus semi-markovien	9
2.1.3.3 Processus de Moran à intervalles dépendants deux à deux.....	9
2.1.3.4 Processus à intervalles dépendants et Markoviens.....	10
2.1.3.5 Processus autoregressif et à moyenne mobile.....	10
2.1.4 Application des processus ponctuels	13
2.2 Analyse du risque	13
3. TESTS D'UNE SÉRIE D'ÉVÉNEMENTS	15
3.1 Notions générales sur les tests statistiques	15
3.2 Exemple d'illustration	17
3.3 Analyse graphique	19
3.3.1 Nombre d'événements en fonction du temps d'occurrence (détection de tendance).....	19
3.3.2 La moyenne du taux d'occurrence (homogénéité)	19
3.3.3 Fonction d'intensité (homogénéité).....	20
3.3.4 Fonction de la variance temporelle (stationnarité)	20
3.3.5 Le spectre du nombre d'événements (analyse de périodicité).....	21
3.4 Analyse de tendance	21
3.4.1 Analyse de régression des intervalles.....	21
3.4.2 Analyse de régression des nombres d'événements	24
3.4.3 Autres méthodes de détection de la tendance.....	25
3.4.3.1 A partir du taux d'occurrence	25
3.4.3.2 A partir de l'analyse des intervalles	27
3.5 Tests des processus de Poisson et de renouvellement	29
3.5.1 Tests d'ajustement indépendants de la distribution "free-distribution"	29
3.5.1.1 Généralités	29
3.5.1.2 Tests bilatéral et unilatéral de Kolmogorov-Smirnov.....	30
3.5.1.3 Test de Cramer-von Mises	31
3.5.1.4 Test de Stépens (Gomes et al, 1975).....	32

3.5.1.5	Test de Sherman (Gomes et al, 1975)	32
3.5.1.6	Autres tests	32
3.5.1.7	Modification de Durbin (Durbin, 1961)	33
3.5.2	Tests relatifs au processus de Poisson	34
3.5.2.1	Test d'uniformité conditionnelle (intervalles).....	34
3.5.2.2	Application de la transformation de Durbin des U_i	35
3.5.2.3	Test de dispersion (nombre d'événements).....	36
3.5.2.4	Tests avec l'hypothèse alternative : processus de renouvellement.....	37
3.5.2.5	Détection de rupture pour des observations poissonniennes.....	38
3.5.2.6	Tests entre le processus de Poisson et le processus de Poisson doublement stochastique	43
3.5.3	Tests relatifs aux processus de renouvellement.....	48
3.5.3.1	Tests du coefficient d'autocorrélation entre les intervalles	48
3.5.3.2	Tests du coefficient d'autocorrélation des rangs	49
3.5.3.3	Tests basés sur l'analyse spectrale	50
3.6	Tests entre processus ponctuels et séries chronologiques.....	51
3.6.1	Test circulaire d'association (Wolfe, 1980).....	52
3.6.2	Processus de Poisson et série chronologique (Wolfe, 1980)	53
3.6.3	Détection de tendance en utilisant une série chronologique (Brillinger, 1994).....	54
3.7	Test de sélection d'un processus ponctuel	56
3.8	Tests d'une série de deux ou de plusieurs types d'événements	59
3.8.1	Superposition de p processus ponctuels	59
3.8.2	Propriétés et analyse des processus superposés.....	62
4.	COMPARAISON ENTRE DEUX OU PLUSIEURS SÉRIES D'ÉVÉNEMENTS.....	67
4.1	Comparaison de deux processus ponctuels de Poisson.....	67
4.1.1	Méthode basée sur la distribution de Poisson.....	67
4.1.2	Méthode basée sur la distribution exponentielle.....	69
4.2	Comparaison de $k > 2$ processus ponctuels de Poisson	70
4.2.1	Test basé sur la distribution de Poisson.....	70
4.2.2	Test du rapport de maximum de vraisemblance.	71
4.2.3	Test basé sur la distribution exponentielle.....	71
4.3	Comparaison des taux d'occurrence des processus non-poissonien	72
4.3.1	Test des rangs de Wilcoxon-Mann-Whitney (Kendall et Stuart, 1979, Kanji, 1997).....	73
4.3.2	Test des sommes de rangs de Kruskal-wallis (k-population, Kanji, 1997)	73
4.4	Test de stationnarité pour plusieurs séries d'événements	75
5.	TESTS DE SERIES D'ÉVENEMENT DANS LES OUTILS INFORMATIQUES	77
5.1.1	Méthodologie de la recherche.....	77
5.1.2	Résultats de la recherche.....	78
6.	CONCLUSION ET RECOMMANDATIONS	81
7.	REFERENCES BIBLIOGRAPHIQUES	85
8.	ANNEXE A.....	91

Liste des figures

Figure-1: Schématisation d'un processus ponctuel.....	4
Figure 2 : Illustration des liens entre les différents processus ponctuels	12
Figure 3: Visualisation des erreurs de type I et II.	17
Figure 4 : Nombre cumulatif des événements en fonction du temps d'occurrence	18
Figure 5 : Récapitulatif des différentes analyses pour une série d'un seul type d'événements.....	58
Figure 6 : Superposition de deux processus ponctuels.	60
Figure 7 : Schématisation du processus alterné	61
Figure 8 : Organigramme de la modélisation d'une série d'événements par un processus ponctuel valable pour un seul type.	83

Liste des tableaux

Tableau 1 : Identification des tests statistiques appliqués au processus alterné (Abi-Zeid, 1997).....	62
Tableau 2 : Principales procédures des logiciels statistiques	79
Tableau 3 : Identification des tests statistiques de six logiciels	80
Tableau A 1 : Tests de stationnarité basés sur la moyenne (Faucher et al, 1997).....	91
Tableau A 2 : Tests de stationnarité basés sur la variance (Faucher et al, 1997)	92
Tableau A 3 : Test d'homogénéité appliqués à une tendance en escalier (Ondo et al, 1997).....	92
Tableau A 4 : Test d'homogénéité appliqués à une tendance en saut (Ondo et al, 1997).....	93
Tableau A 5 : Tests d'indépendance (Ondo et al, 1997)	93

1. INTRODUCTION

L'analyse de séries d'événements a été souvent traitée par la théorie des processus stochastiques ou ponctuels. Un processus ponctuel est défini comme étant un processus stochastique pour lequel chaque réalisation constitue une collection de points. Ce modèle est de plus en plus utilisé dans plusieurs domaines de recherche, là où l'occurrence des événements n'est pas régulière (tremblements de terre, crues extrêmes, écrasement d'un avion, ...). Comme tout modèle, un processus ponctuel est basé sur une série d'hypothèses à vérifier. En effet, la phase de validation du modèle est nécessaire pour tester l'adéquation des résultats théoriques avec les données empiriques. Cette étape de modélisation peut être effectuée en utilisant une nouvelle série de données ou par l'application des tests statistiques. Ces tests peuvent être subdivisés en deux catégories : 1/ La première catégorie consiste à tester si une série d'événements peut être représentée par un processus ponctuel bien défini. À titre d'exemple, nos hypothèses H_0 et H_1 peuvent être formulées de la façon suivante :

H_0 : la série d'événements provient d'un processus de Poisson homogène ;

H_1 : la série d'événements ne provient pas d'un processus de Poisson homogène .

2/ La deuxième catégorie de tests suit le choix du modèle représentant les données, et consiste à tester les différentes hypothèses du processus ponctuel, telles que l'ajustement de la loi statistique du nombre d'occurrence ou la stationnarité de la fonction d'intensité.

L'objectif de cette recherche bibliographique est de mettre en évidence les différents tests statistiques appliqués aux séries d'événements. Cette étude portera également sur l'identification des outils informatiques qui existent dans le marché des logiciels statistiques contenant des tests statistiques en général et ce genre de tests en particulier. Ce travail est réalisé avec l'idée d'une application ultérieure dans le cadre de l'analyse du risque. En effet, l'étude statistique de séries d'événements par les processus ponctuels peut être complétée

par une analyse du risque, qui comprend deux étapes : l'évaluation du risque et la gestion du risque. L'analyse du risque permet l'élaboration d'un outil d'aide à la prise de décision en avenir incertain. Ainsi, la prise d'une décision est souvent basée sur l'optimisation de certains critères ou indices. En effet, il est plus commode de définir le risque par des indices de risque, qui servent à juger et à comparer plusieurs politiques de gestion dans la prise de décision. Dans une étude d'analyse du risque appliquée aux étiages des cours d'eau, Abi-Zeid (1997) a considéré cinq indices du risque basés sur les durées des étiages, et qui ont servis à l'identification des politiques de gestion de la ressource en eau.

La section suivante de ce rapport comporte une présentation succincte de différents processus ponctuels existant dans la littérature et une brève introduction à l'analyse du risque. Au troisième chapitre, on s'intéresse à la description des tests statistiques d'une série d'événements. Une procédure de classification des tests ainsi, qu'une identification des tests applicables à une série de deux ou de plusieurs types d'événements seront abordées. Le quatrième chapitre est consacré à la comparaison entre deux ou plusieurs séries d'événements d'un seul type. Le dernier chapitre sera consacré à l'identification d'outils informatiques traitant de la modélisation des processus ponctuels. Finalement, on termine par une conclusion générale de notre recherche bibliographique et une présentation des recommandations concernant la possibilité et l'utilité du développement d'un outil informatique.

2. PROCESSUS PONCTUELS ET RISQUE

Dans cette section, nous allons présenter une description générale des processus de Poisson et de renouvellement. Différentes généralisations de ces processus ponctuels seront également présentées. Enfin, on termine par l'illustration de quelques définitions du risque, et une brève introduction à l'analyse du risque.

2.1 Processus ponctuels

2.1.1 Définitions

On désigne sous le nom de modèle aléatoire tout schéma abstrait de nature probabiliste, susceptible de représenter certains phénomènes réels. Ces modèles jouent un grand rôle dans l'évolution de la théorie de décision, d'autant plus que leur champ d'application est immense car ils sont d'une très grande généralité. Les éléments aléatoires sont des fonctions aléatoires avec un argument qui est le temps. On dit alors que ce sont des processus aléatoires (Girault, 1965).

On suppose qu'un événement aléatoire peut se produire, soit à chaque instant, soit à certains moments déterminés. Une réalisation d'un tel processus est décrite par une suite de points $\{P_1, P_2, \dots, P_n, \dots\}$ sur l'axe des temps, ou d'une manière équivalente, par une suite croissante de dates $\{t_1, t_2, \dots, t_n, \dots\}$, d'où le nom de processus ponctuel donné à un tel schéma.

Une définition en terme de probabilité peut être postulée de la façon suivante :

Soit un ensemble d'intervalles disjoints $(t_1, t_1'') ; (t_2, t_2'') ; \dots ; (t_k, t_k'')$ et N_1, N_2, \dots, N_k les nombres aléatoires qui se produisent dans ces intervalles. Pour définir en probabilité le processus ponctuel, il faut, par exemple, se donner pour toute valeur de k et en fonction de

$[t_1, t_1''; \dots; t_k, t_k'']$, la fonction de répartition de l'ensemble aléatoire (N_1, N_2, \dots, N_k) , nombres d'événements qui se produisent dans des intervalles considérés. (Girault, 1965).

$$\left\{ F_{(n_1, n_2, \dots, n_k)}^{(t_1, t_1'', \dots, t_k, t_k'')} \right\} \quad k = 1, 2, \dots, \infty$$

$$t_1' \leq t_1'' \leq \dots \leq t_k' \leq t_k''$$

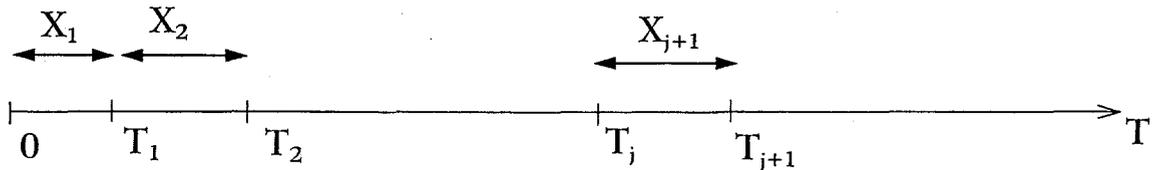


Figure-1: Schématisation d'un processus ponctuel

T_j est une variable aléatoire de réalisation t_j .

Un processus ponctuel, schématisé par la figure-1 et ne permettant pas l'occurrence d'événements multiples simultanés, peut être spécifié par trois manières :

- en fonction de la distribution conjointe du nombre des points $N(A)$ sur un intervalle A , où $N(A)$ est le nombre d'événements dans l'intervalle A ;
- en fonction de la distribution conjointe des intervalles $\{X_j\}$ entre les points successifs, où $X_j = T_j - T_{j-1}, j > 1$. ; ou
- en utilisant la fonction d'intensité complète $\lambda(t, H_t)$ qui représente le taux moyen instantané d'occurrence d'un événement

$$\lambda(t, H_t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(N_{t+\Delta t} = 1 / H_t)}{\Delta t} ; H_t = t_1, t_2, \dots, t_n \quad (2-1)$$

H_t : représente l'historique du processus au temps t , c'est-à-dire tous les points dans l'intervalle $(0, t]$.

Dans ce cas la dimension de l'espace de réalisation des événements est égale à l'unité. Il s'agit de l'espace temps. Cependant, on peut définir un processus ponctuel spatial (Diggle, 1983) ou bien spatio-temporel (Cox et Isham, 1980).

2.1.2 Processus de Poisson

2.1.2.1 Le processus de Poisson homogène (PPH)

A cause de sa simplicité et de sa généralité, ce processus a été le plus couramment utilisé. Il est caractérisé par les hypothèses de :

- a) stationnarité : la probabilité d'occurrence d'un événement est proportionnelle à la longueur d'un petit intervalle et indépendante du temps d'occurrence de l'événement lui-même ;
- b) non-multiplicité : la probabilité de deux ou plusieurs événements dans un petit intervalle de temps $(t, t+\Delta t]$ est négligeable ; et
- c) indépendance : le nombre d'événements dans un intervalle donné est indépendant du nombre d'événements dans d'autres intervalles disjoints.

Un processus ponctuel (PP) est un processus de Poisson homogène (PPH) de taux λ constant si les conditions suivantes sont vérifiées $\forall t$ et $\Delta t \rightarrow 0^+$:

$$P\{N_{t,t+\Delta t} = 1 / H_t\} = \lambda \Delta t + o(\Delta t) \quad (2-2)$$

$$P\{N_{t,t+\Delta t} > 1 / H_t\} = o(\Delta t) \quad (2-3)$$

Le nombre de points dans un intervalle de longueur t est distribué selon une loi de Poisson de moyenne λt . Les intervalles X_1, X_2, \dots, X_j entre les points successifs sont indépendants et distribués selon une loi exponentielle de paramètre λ (Girault, 1965) :

$$f_x(x) = \lambda \exp(-\lambda x) \quad (2-4)$$

La variable aléatoire T_j suit une loi d'Erlang, dont la densité est définie par (Cox et Lewis, 1966) :

$$g(t) = \frac{\lambda(\lambda t)^{j-1} e^{-\lambda t}}{(j-1)!}, \quad t \geq 0 \quad (2-5)$$

2.1.2.2 Le processus de Poisson non-homogène (PPNH)

La fonction d'intensité d'un processus de Poisson est une fonction du temps, $\lambda = \lambda(t)$. Ce processus est dit processus de Poisson non-homogène. Le nombre d'événements N_t est alors distribué selon une loi de Poisson de paramètre (Cox et Isham, 1980) :

$$\tau(t) = \int_0^t \lambda(u) du \quad (2-6)$$

τ : représente la valeur moyenne du nombre d'observations dans l'intervalle $(0, t]$. Ainsi, l'indépendance des intervalles entre deux événements successifs est conditionnelle à la variable temps.

2.1.2.3 Processus de Poisson composé ou multiple

Comme son appellation l'indique, ce processus ne tient pas compte de l'hypothèse b) de non-multiplicité définie dans le processus de Poisson homogène (paragraphe, 2.1.2.1). Chaque point du processus de Poisson est remplacé par un nombre aléatoire M d'occurrences simultanées. Les hypothèses a) et c) d'indépendance et de stationnarité sont conservées. Ce processus peut être défini par la fonction génératrice de probabilité (probability generating functional) suivante (Cox et Isham, 1980) :

$$G[\xi] = \exp \left\{ -\lambda \int_{-\infty}^{+\infty} [1 - G_M \{ \xi(t) \}] dt \right\} \quad (2-7)$$

où :

$G_M(z) = E(z^M) = E[\exp(M \log(z))]$: est la fonction génératrice de probabilité pour le nombre M .

2.1.2.4 Processus de Poisson groupé (PPG) "clusters process"

Il s'agit d'associer à chaque point du processus de Poisson un groupe de points. Ainsi, un processus de Poisson groupé est un processus ponctuel des centres de ces groupes, on associe à chaque centre un nombre aléatoire de points formant un processus subsidiaire.

Afin de spécifier un PPG, il est nécessaire de déterminer le processus qui génère les centres des groupes et les mécanismes qui gouvernent chaque groupe. Ainsi, il est important de tenir en compte dans ce processus les propriétés suivantes :

- a. les tailles des groupes sont indépendantes et identiquement distribuées ;
- b. les centres des groupes constituent un processus de Poisson.

Deux cas spécifiques du PPG sont le processus de Neyman-Scott et celui de Bartlett-Lewis. Ils ont été utilisés dans plusieurs études (Cox et Isham, (1980), Kavvas, (1975), Waymire et Gupta, (1981)).

2.1.2.5 Processus de Poisson doublement stochastique

Le processus de Poisson non-homogène est caractérisé par un taux d'occurrence qui est une fonction du temps. Cependant, ce dernier peut être généralisé en utilisant un taux d'occurrence $\Lambda(t)$ représenté par un processus stochastique non observé (Cox et Lewis, 1966). Ce cas correspond au processus de Poisson doublement stochastique. Ainsi, on peut écrire le taux d'occurrence du processus comme suit (Cox et Isham, 1980):

$$\begin{aligned} \rho(t, H_t, H_t^\Lambda) &= \lim_{\delta \rightarrow 0^+} \delta^{-1} \text{prob}\{N(t, t + \delta) > 0 / H_t, \Lambda(s) = \lambda(s) (-\infty \leq s \leq t)\} \\ &= \lambda(t) \end{aligned} \quad (2-8)$$

où :

H_t^Λ : représente l'historique du processus aléatoire de la fonction Λ au temps t .

2.1.2.6 Processus linéaire auto-excité

Le processus linéaire auto-excité est le processus le plus général défini par la fonction d'intensité (Cox et Isham, 1980) :

$$\Lambda(t, H_t) = \gamma + \int_{-\infty}^t w(t-z) dN(z) \quad (2-9)$$

où : $\gamma > 0$ est une constante et w est une fonction non-négative telle que :

$$w^*(0) = \int_0^{+\infty} w(x) dx < 1.$$

Si on suppose que le processus est stationnaire, on peut vérifier que (Cox et Isham, 1980) :

$$\lambda(1 - w^*(0)) = \gamma \quad (2-10)$$

C'est un processus plus général comparativement aux autres modèles. Il est obtenu à partir d'un processus de Poisson en permettant à l'intensité de dépendre du processus ponctuel lui-même. Ainsi, le nombre d'événements N_t passés et leurs temps d'occurrence peuvent influencer le nombre d'événements futurs et leurs temps d'occurrence.

2.1.3 Le Processus de renouvellement

2.1.3.1 Processus de renouvellement homogène (PRH)

Les intervalles séparant les points consécutifs dans un processus de Poisson sont indépendants et distribués selon une loi exponentielle. Ce résultat peut être généralisé, afin d'obtenir un processus de renouvellement homogène, lorsque les intervalles $\{X_1, X_2, \dots, X_j\}$ sont indépendants et distribués selon une loi commune arbitraire. L'intensité $\lambda(t, H_t)$ d'un PRH dépend uniquement de l'instant d'occurrence du dernier événement, donc $\lambda(t, H_t) = \lambda(t, t_{N_t}) = \lambda(t_{N_t})$. On définit la fonction d'intensité conditionnelle $h(t)$ comme la probabilité instantanée d'un événement dans l'intervalle $(t, t+\Delta t]$, étant donné qu'un événement s'est réalisé à l'origine (Cox et Isham, 1980) :

$$h(t) = \lim_{\Delta t \rightarrow 0^+} \frac{P(N_{t, t+\Delta t} > 0)}{\Delta t} \quad (2-11)$$

Une quantité importante pour le PRH est le taux d'avarie (Hazard rate) $\varphi(x)$. Elle est définie par:

$$\varphi(x) = \lim_{\Delta x \rightarrow 0^+} \frac{P(x \leq X \leq x + \Delta x / x < X)}{\Delta x} \quad (2-12)$$

$\varphi(x)$ peut être également définie à partir de la fonction densité de probabilité des intervalles $f(x)$:

$$\varphi(x) = \frac{f(x)}{1 - F(x)} \quad (2-13)$$

où : $F(x) = \int_0^x f(u)du$: représente la fonction de répartition.

$\varphi(x)\Delta t$: représente la probabilité d'occurrence d'un événement étant donné que le temps écoulé depuis le dernier événement est x .

2.1.3.2 Processus semi-markovien

La distribution des intervalles d'un processus stochastique est représentée par une chaîne de Markov. A titre d'exemple, soient k fonctions de répartition F_1, \dots, F_k et supposons que le processus est dans l'état j à un instant t avec une fonction F_j . Alors, on suppose que la séquence des états du processus est déterminée par une chaîne de Markov avec une matrice des probabilités de transition $P=(P_{ij})$, telle que si le processus est dans l'état i , la probabilité que le prochain état sera dans l'état j est P_{ij} (Cox et Isham, 1980). Le processus ponctuel déterminé par ce type de séquences est appelé processus semi-markovien. Dans le cas d'un processus de renouvellement non homogène, la distribution de la séquence des intervalles X_i est supposée connue. Cependant, le processus semi-markovien représente la séquence des intervalles par une chaîne de Markov.

2.1.3.3 Processus de Moran à intervalles dépendants deux à deux

Une autre façon pour généraliser la séquence des intervalles indépendamment et identiquement distribués d'un processus de renouvellement est d'assouplir le critère d'indépendance en considérant les paires d'intervalles indépendants, au lieu des intervalles pris séparément. Ainsi, la séquence d'intervalles pour ce processus est $\dots, U_{i-1}, V_{i-1}, U_i, V_i, U_{i+1}, V_{i+1}, \dots$, où (U_i, V_i) est une paire d'intervalles indépendants ayant la même distribution marginale, (Cox et Isham, 1980).

2.1.3.4 Processus à intervalles dépendants et Markoviens

Dans ce type de processus, les intervalles séparant les points d'occurrence sont dépendants et forment une séquence de Markov, telle que la distribution de $X_{j+1} / [X_i, i \leq j]$ ne dépend pas de $[X_i, i < j]$. Plus généralement, les intervalles peuvent être pris pour constituer une séquence de Markov m-dépendante telle que la distribution de $X_{j+m} / [X_i, i \leq j+m-1]$ ne dépend pas de $[X_i, i < j]$, (Cox et Isham, 1980).

2.1.3.5 Processus autoregressif et à moyenne mobile

Une importante classe de processus ponctuels spécifiés par les séquences d'intervalles dépendants est reliée à la fonction autorégressive ou/et à la fonction moyenne mobile. Dans chaque cas, la séquence des intervalles est stationnaire et suit une distribution marginale exponentielle (Cox et Isham, 1980). De tels processus peuvent constituer une bonne alternative au processus de Poisson. Soient :

- $\{\varepsilon_i, i = 0, 1, \dots\}$ une séquence d'intervalles indépendants et distribués selon une loi exponentielle de paramètre λ ;
- $\{U_i, i = 1, 2, \dots\}$ et $\{V_i, i = 1, 2, \dots\}$ sont indépendants de $\{\varepsilon_i\}$;
- $\{U_i, i = 1, 2, \dots\}$ et $\{V_i, i = 1, 2, \dots\}$ sont deux séquences indépendamment et identiquement distribuées selon une loi de Bernoulli de paramètre $\alpha = \text{prob}(U_i = 0)$ et $\beta = \text{prob}(V_i = 0)$

Ainsi, on peut définir les séquences d'intervalles suivantes :

$$\begin{aligned}
 X_0 &= \varepsilon_0 \\
 X_n &= \alpha X_{n-1} + U_n \varepsilon_n \quad (n = 1, 2, \dots) \\
 Y_n &= \beta \varepsilon_n + V_n \varepsilon_{n-1} \quad (n = 1, 2, \dots) \\
 Z_n &= \beta \varepsilon_n + V_n X_{n-1} \quad (n = 1, 2, \dots)
 \end{aligned}
 \tag{2-14}$$

Le processus ponctuel défini par la séquence d'intervalles $\{X_n\}$ est appelé un processus exponentiel autoregressif d'ordre 1, noté EAR1, $\{Y_n\}$ est un processus exponentiel à

moyenne mobile, noté EMM1, et $\{Z_n\}$ est un processus exponentiel autorégressif-moyenne mobile, noté EARMM(1,1).

La figure-2 résume les hypothèses de la majorité des processus ponctuels décrits dans ce chapitre, et illustre les étapes de passage entre le processus le plus général aux processus les moins compliqués.

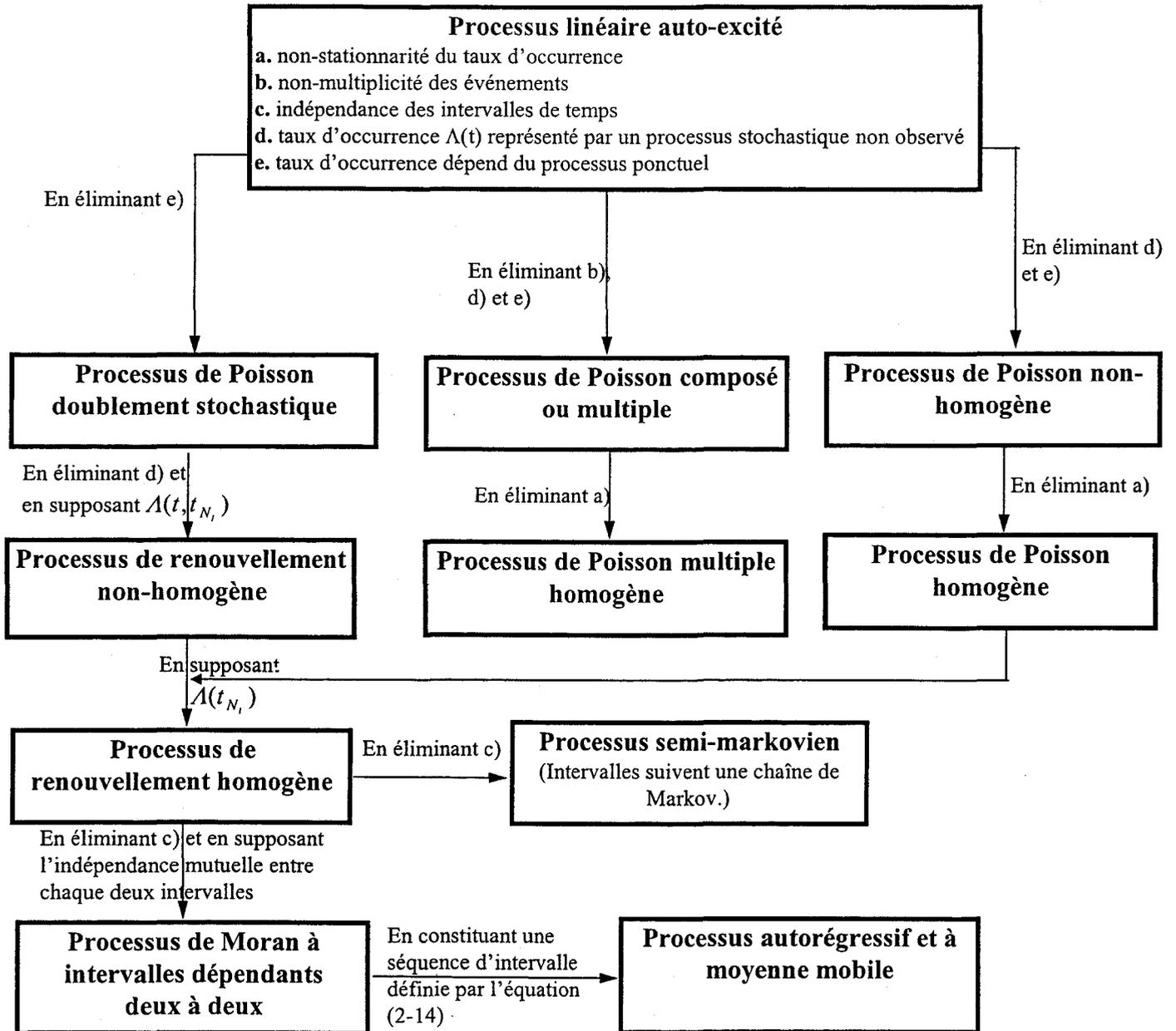


Figure 2 : Illustration des liens entre les différents processus ponctuels

2.1.4 Application des processus ponctuels

Les processus ponctuels ont attiré l'attention de plusieurs chercheurs, d'où la diversité de leurs champs d'applications. Ainsi, on trouve dans la littérature plusieurs ouvrages qui traitent particulièrement ce type de processus, (Cox et Lewis, 1966, Snyder, 1975, Cox et Isham, 1980). Dans le cadre de la modélisation des phénomènes naturels par ces processus, Kavvas (1975) et Smith (1980) ont modélisé les précipitations en utilisant différents types de processus ponctuels, tout en vérifiant les hypothèses de chaque processus. Diggle (1983) a employé la théorie des processus ponctuels spatiaux pour la modélisation de plusieurs problématiques en biologie, telle que la modélisation du nombre des arbres dans une forêt afin de mieux connaître la répartition des oiseaux dans cette zone d'étude. Karr (1986) a étudié un certain nombre de processus ponctuels et il s'est notamment intéressé aux inférences statistiques à partir de ces modèles. Thompson (1988) a également utilisé les processus ponctuels pour modéliser les déplacements des points, ce qu'il a appelé les points d'arrivée, dans le temps selon des mécanismes donnés. Cette application dans le domaine de la physique nucléaire concernait les arrivées des particules radioactives dans un compteur Geiger, les collisions successives d'une particule avec d'autres particules dans la théorie cinétique d'un gaz parfait, etc. Dans une autre étape, il a exploité ses résultats dans le domaine de la sécurité et de la fiabilité des systèmes. Brillinger (1994) a appliqué les processus ponctuels dans le domaine de l'hydrologie et de la sismologie. Abi-Zeid (1997) a effectué une évaluation des processus ponctuels pour modéliser les occurrences et les durées des étiages et des surplus afin de caractériser et quantifier les risques qui en résultent, tout en tenant compte de la non-stationnarité du phénomène.

2.2 Analyse du risque

Nous avons présenté dans les sections précédentes différents processus ponctuels pour modéliser les séries d'événements. Cette modélisation doit être complétée par la

Risque et probabilité d'erreur

α et β sont les probabilités d'erreur de première et deuxième espèce :

- α représente la probabilité de choisir H_1 alors que H_0 est vraie ;
- β représente la probabilité de choisir H_0 alors que H_1 est vraie.

Il faut noter que la quantité $(1-\beta)$ représente la puissance du test statistique, qui est la probabilité de rejeter H_0 alors que H_1 est vraie.

Dans la pratique des tests statistiques, il est de règle de fixer α (les valeurs courantes sont par exemple 0.05 ; 0.01 ou 0.1) de préférence en fonction du risque de première espèce couru, ce qui fait jouer à H_0 un rôle prééminent. Cependant, il ne faut pas négliger l'importance de la probabilité β , qui est liée directement à la puissance du test.

En effet, on peut déterminer la valeur de l'erreur de deuxième espèce β en fixant α et la taille d'échantillon n . Afin de mieux visualiser la relation entre les deux probabilités α et β , nous allons traiter le test de la moyenne d'un échantillon normal de moyenne \bar{X} et de variance σ^2 . Les hypothèses du test sont :

$$H_0 : \mu = \mu_0 = 500$$

$$H_1 : \mu = \mu_1 = 650$$

Les valeurs de μ_0 et de μ_1 sont choisies arbitrairement. On peut alors exprimer, respectivement, l'erreur du premier type et du second type par :

$$\alpha = \text{Pr ob}(AH_0 / H_0) = \text{Pr ob}(\bar{X} > k / H_0) \quad (3-1)$$

$$\beta = \text{Pr ob}(AH_0 / H_1) = \text{Pr ob}(\bar{X} \leq k / H_1) \quad (3-2)$$

La valeur du seuil k est donnée par :

$$k = \mu_0 + Z(1-\alpha) \frac{\sigma}{\sqrt{n}} \quad (3-3)$$

où $Z(1-\alpha)$ est le quantile $(1-\alpha)$ de la loi normale centrée réduite $N(0,1)$.

Ces deux probabilités sont schématisées dans la figure-3. On constate qu'en diminuant la valeur de α , la valeur de β augmente, d'où l'importance de considérer ces deux probabilités dans un test statistique.

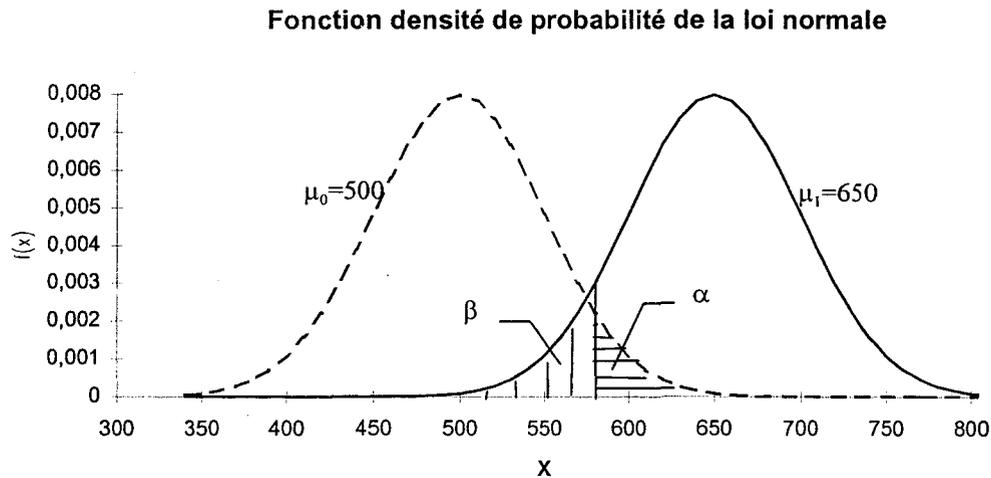


Figure 3: Visualisation des erreurs de type I et II.

En développant l'équation (3-2), on peut alors écrire :

$$\begin{aligned} \beta &= \text{Prob}(\bar{X} \leq k \mid \mu = \mu_1) = \text{Prob}\left(\frac{\bar{X} - \mu_1}{\sigma / \sqrt{n}} \leq \frac{k - \mu_1}{\sigma / \sqrt{n}}\right) \\ &= \text{Prob}\left(Z \leq \frac{\mu_0 - \mu_1}{\sigma / \sqrt{n}} + Z(1 - \alpha)\right) \end{aligned} \quad (3-4)$$

En examinant l'équation (3-4), on constate que la probabilité β dépend de α , du rapport

$d = \frac{\mu_0 - \mu_1}{\sigma / \sqrt{n}}$ et de n . Il existe dans la littérature des diagrammes de variations de β en

fonction de ces trois paramètres. Ces courbes sont appelées les courbes OC.

3.2 Exemple d'illustration

Avant de décrire les différents tests visant l'analyse de séries d'événements par les processus ponctuels, nous présentons un exemple d'une série d'événements tiré de Abi-Zeid (1997) afin de bien mettre en évidence l'importance de l'application des tests statistiques pour la vérification des hypothèses d'un processus ponctuel. Il s'agit des données de

précipitations enregistrées à la station pluviométrique Bar-sur-Seine située sur la rivière la Seine en France.

On note les intervalles séparants les événements successifs enregistrés par x_1, x_2, \dots, x_n . Une série équivalente à la série des intervalles est représentée par les temps d'occurrence des événements à partir du début de la période d'observation : $t_1=x_1, t_2=t_1+x_2, \dots, t_r=t_{r-1}+x_r$, où t_r est le temps d'occurrence des événements. La période totale des observations est notée par t_0 . Il sera plus commode de représenter les données graphiquement afin de mieux les analyser.

On peut représenter la variation du taux moyen d'occurrence des événements par deux méthodes de représentation graphique : l'une basée sur le nombre cumulé et l'autre basée sur le nombre d'occurrence individuel (Cox et Lewis, 1966). Le graphique le plus simple est celui du nombre total des événements enregistrés à ou avant l'instant t en fonction du temps t (figure-4).

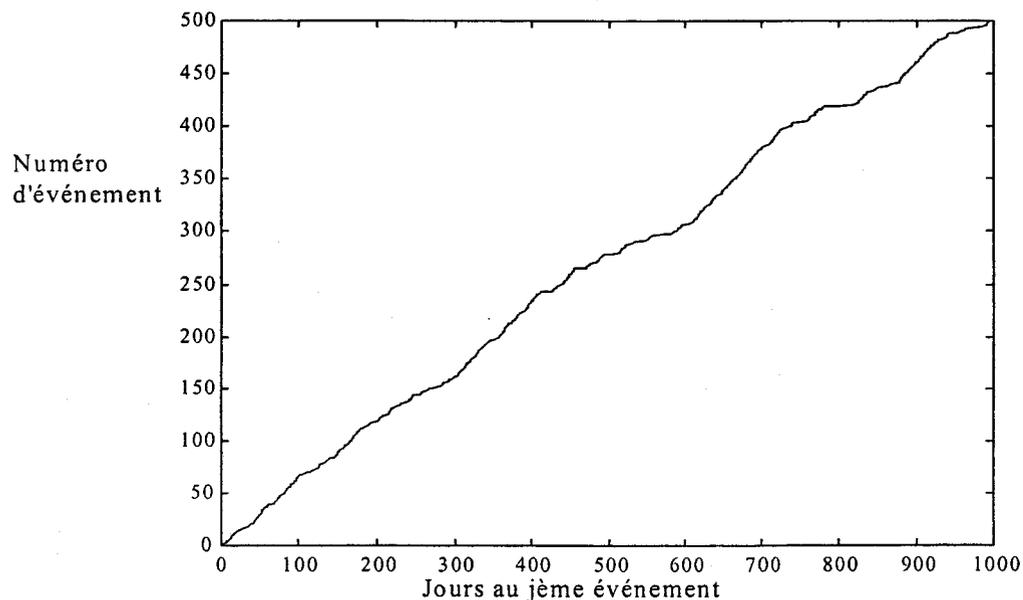


Figure 4 : Nombre cumulé des événements en fonction du temps d'occurrence

Une analyse préliminaire de ce graphique nous révèle une tendance dans le taux moyen d'occurrence des événements avec la variable temps. Ainsi, l'hypothèse de stationnarité du

processus de Poisson homogène est violée, d'où la nécessité de l'application des tests statistiques aux séries d'événements avant de choisir le modèle approprié à cette série.

3.3 Analyse graphique

Comme nous avons vu dans l'exemple de la section 3-2, l'analyse graphique des séries de données est une étape nécessaire pour détecter toute forme de tendance ou de dépendance. En effet, la stationnarité, l'homogénéité et l'indépendance des données empiriques constituent des hypothèses principales dans les processus ponctuels. Kavvas (1975) a proposé cinq représentations graphiques pour analyser les données d'une série d'un seul type d'événements.

3.3.1 Nombre d'événements en fonction du temps d'occurrence (détection de tendance)

C'est le graphique qui lie le nombre total d'occurrence et le temps cumulatif. La pente de la courbe obtenue représente le taux moyen d'occurrence par unité de temps. Dans le cas d'un processus stationnaire, cette quantité doit être constante (figure-4).

3.3.2 La moyenne du taux d'occurrence (homogénéité)

Le taux moyen d'occurrence au temps t est donné par :

$$\lambda_{\tau} = \frac{n(t, t + \tau)}{\tau} \quad (3-5)$$

Pour cette analyse, on doit prendre une période de temps τ constante. La variable $n(t, t + \tau)$ représente le nombre d'occurrence durant l'intervalle $(t, t + \tau)$, $\forall t$.

3.3.3 Fonction d'intensité (homogénéité)

Sous l'hypothèse de stationnarité, la fonction intensité est donnée par :

$$m_f(\tau) = \frac{dM_f(\tau)}{d\tau} = \lim_{\Delta t \rightarrow 0} \frac{\text{Prob}(\text{événement} \in (t + \tau, t + \tau + \Delta t) / \text{événement a } t)}{\Delta t} \quad (3-6)$$

Dans un processus de renouvellement, $M_f(\tau)$ est appelée la fonction densité de renouvellement. Dans le cas du processus de Poisson homogène, $m_f(\tau) = \lambda$; la fonction d'intensité est alors représentée par une droite horizontale.

3.3.4 Fonction de la variance temporelle (stationnarité)

La fonction de variance $V(t)$ du nombre N_t est définie par :

$$V(t) = \text{var}(N_t) = E(N_t^2) - E^2(N_t) \quad (3-7)$$

Sous l'hypothèse d'un processus stationnaire, la fonction $V(t)$ peut être exprimée en fonction du taux moyen d'occurrence et de la fonction intensité (Cox et Lewis, 1966) :

$$V(t) = mt + 2 \int_0^t \int_0^v m(m_f(u) - m) du dv \quad (3-8)$$

où :

$$m = \frac{dM(t)}{dt} = \frac{1}{E(X)}$$

L'analyse de la stationnarité par cette fonction n'est pas toujours recommandée, car les estimations de $V(t)$ pour différentes valeurs de t sont très corrélées entre elles (Cox et Lewis, 1966).

3.3.5 Le spectre du nombre d'événements (analyse de périodicité)

Le spectre des dénombrements, $g(w)$, pour un processus stochastique stationnaire est donné par (Cox et Lewis, 1966) :

$$g(w) = \frac{m}{2\pi} + \frac{m}{2\pi} \int_{-\infty}^{+\infty} \{m_f(\tau) - m\} e^{-i w \tau} d\tau, \quad -\infty < w < +\infty \quad (3-9)$$

Si le spectre est défini pour des fréquences positives, alors la fonction de ce spectre est :

$$g_+(w) = 2g(w)$$

la valeur du spectre à l'origine est donnée par :

$$g_+(0) = \frac{m}{\pi} + \frac{m}{\pi} \int_{-\infty}^{+\infty} \{m_f(\tau) - m\} d\tau$$

Pour un processus de Poisson homogène, on peut montrer que (Cox et Lewis, 1966) :

$$g_+(w) = \frac{\lambda}{\pi}$$

Kavvas (1975) a indiqué que l'application de l'analyse spectrale est nécessaire pour la détection de la périodicité dans la série de données. Cependant, cette analyse dépend fortement de l'hypothèse de stationnarité.

3.4 Analyse de tendance

3.4.1 Analyse de régression des intervalles

Dans cette section, l'hypothèse nulle à vérifier est exprimée de la façon suivante :

H_0 : Il n'existe aucune tendance dans la série d'événements.

Supposons que l'on dispose d'une série d'observations assez longue et que l'on veuille examiner l'existence d'une tendance dans le taux d'occurrence. On choisit l un entier de

préférence inférieure à quatre pour ne pas avoir de changement dans le taux d'occurrence durant la réalisation des événements. Soient y_1 le temps observé du début jusqu'au $l^{\text{ième}}$ événement, y_2 le temps du $l^{\text{ième}}$ jusqu'au $2l^{\text{ième}}$ événement ...etc.

Il sera nécessaire de ne pas prendre en considération quelques événements, si n n'est pas un multiple de l . Dans ce cas, il est préférable d'éliminer des événements au milieu de la série. La raison pour cette omission est due au fait que pour la plupart des processus, à l'exception du processus de Poisson, la longueur du temps mesurée à partir d'un point arbitraire jusqu'au premier événement ne possède pas la même distribution que les intervalles entre les événements successifs (Cox et Lewis, 1966).

Ainsi, on a une série d'intervalles y_1, \dots, y_r . Si le processus est poissonnien, on a :

$$E(\text{Log}Y_i) = -\text{Log}\lambda + e_i \quad (3-10)$$

$$\text{var}(\text{Log}Y_i) = v_i \quad (3-11)$$

e_i et v_i sont des constantes indépendantes de λ et en particulier, on a :

$$v_i \approx \frac{1}{l-1/2} \quad (3-12)$$

La distribution des $\text{Log} Y_i$ est une loi Log Chi-deux à $2n$ degrés de liberté (Cox et Lewis, 1966).

On suppose que les observations sont générées par un processus de Poisson défini par les conditions b) et c) de la section 2.1.2.1 et que le taux d'occurrence du processus varie légèrement avec la variable temps. Alors, on construit le modèle suivant :

$$\text{Log}\lambda_o = \alpha + \beta z_i \quad (3-13)$$

$$\Rightarrow \lambda_i = e^{\alpha + \beta z_i} ; i=1, \dots, r$$

z_i est une variable indépendante et peut être représentée par :

- la valeur centrale de l'intervalle y_i ;
- le numéro d'événement au centre de l'intervalle y_i . Ce cas se présente si λ est exprimé en fonction du nombre d'événements ;
- la moyenne des valeurs comprises dans l'intervalle y_i .

En remplaçant la valeur de $\text{Log} \lambda_i$ (équation, 3-13) dans l'équation (3-10), on obtient :

$$E(\text{Log} Y_i) = -(\alpha' + \beta z_i) \quad (3-14)$$

où :

$$\alpha' = \alpha - e_1$$

$$\text{var}(\text{Log} Y_i) = v_1$$

Ainsi, on peut:

- estimer les valeurs de α' et β par la méthode des moindres carrés ;
- comparer les résidus quadratiques de la régression avec la valeur théorique v_1 pour obtenir un appui de la validité du modèle ;
- tester l'hypothèse nulle que $\beta=0$ et obtenir l'intervalle de confiance de β par un modèle de régression.

$$H_0 : \beta = 0$$

$$H_1 : \beta \neq 0$$

La statistique du test est donnée par :

$$t = \frac{bS_z}{\frac{S_{y.z}}{(n-1)^{1/2}}} \quad (3-15)$$

$$b = E(\beta), \quad S_z = \frac{\sum_{i=1}^n (z_i - \bar{z})^2}{n-1} \quad \text{et} \quad S_{y.z} = \frac{1}{n-2} \sum_{i=1}^n \left\{ \text{Log} y_i - \frac{1}{n} \sum_{i=1}^n \text{Log} y_i - b(z_i - \bar{z}) \right\}^2$$

\bar{z} représente la valeur moyenne des z_i .

La statistique t suit une loi de Student à $(n-2)$ degrés de liberté. Ce test exige la normalité de la variable dépendante dans le modèle de régression.

La transformation logarithmique assure une variance constante et permet de considérer une relation exponentielle, qui est la relation la plus simple du taux d'occurrence λ . On peut également utiliser un modèle de régression multiple :

$$\text{Log} \lambda_i = \alpha + \beta z_i + \gamma z_i^2 \quad (3-16)$$

Le modèle exprimé par l'équation (3-13) où $\text{var}(\text{Log} Y_i) = \sigma^2$ et σ^2 est inconnu, permet d'étudier la tendance temporelle dans le processus. Il faut noter que dans la plupart des processus où une tendance existe, il y a une corrélation entre les différentes valeurs de

$\text{Log}Y_i$. Cependant, les intervalles successifs entre les événements doivent être indépendamment distribués avec un coefficient de variation constant.

Sous l'hypothèse de ce modèle, il est également possible d'obtenir des tests asymptotiquement efficaces et estimer α' et β par la méthode du maximum de vraisemblance. Ainsi, on peut mesurer l'efficacité asymptotique de l'estimateur des moindres carrés de β par le rapport entre l'estimateur du maximum de vraisemblance et celui des moindres carrés. Cox (1955) a montré qu'au voisinage de $\beta=0$, l'efficacité est la même pour le cas du processus de Poisson homogène.

Finalement, il est possible de détecter l'existence d'une tendance, en supposant que les événements forment un processus de Poisson et en utilisant des tests d'ajustement pour :

- la distribution des intervalles entre les événements successifs ;
- la distribution du nombre d'événements dans un intervalle de longueur constante.

Cox et Lewis, (1966) n'ont pas recommandé l'utilisation de ces tests, car ils possèdent une faible puissance pour discriminer entre l'hypothèse nulle et l'hypothèse alternative.

3.4.2 Analyse de régression des nombres d'événements

Par analogie à la section 3.4.1, on peut établir un modèle de régression en fonction du nombre d'événements n dans un intervalle de temps donné ou en fonction de ses valeurs transformées. La transformation des données est effectuée pour obtenir un modèle linéaire.

Cox et Lewis (1966) ont adopté les variables transformées suivantes :

$$\text{Log}(N+1/2) \text{ et } \sqrt{N+1/4}.$$

Les valeurs 1/2 et 1/4 sont incorporées pour raffiner les résultats. Ainsi, N suit une loi de Poisson de moyenne μ , la valeur estimée de la transformation de N ($\mu = \lambda t \rightarrow \infty$):

$$E\left\{\text{Log}\left(N + \frac{1}{2}\right)\right\} = \text{Log}\mu + o\left(\frac{1}{\mu}\right) \quad (3-17)$$

et :

$$E\left\{\left(\sqrt{n + \frac{1}{4}}\right)\right\} = \sqrt{\mu} + o\left(\frac{1}{\mu^{3/2}}\right) \quad (3-18)$$

Tandis que :

$$E\{Log(N)\} = Log\mu + o\left(\frac{1}{\mu}\right) \text{ et } E\{(\sqrt{n})\} = \sqrt{\mu} + o\left(\frac{1}{\mu^{1/2}}\right)$$

On démontre numériquement que la transformation exprimée par l'équation (3-17) est la plus proche estimation pour des valeurs de $\mu > 1$, (Cox et Lewis, 1966).

3.4.3 Autres méthodes de détection de la tendance

3.4.3.1 A partir du taux d'occurrence

Supposons l'existence d'une certaine tendance dans le taux d'occurrence, représentée par un changement dans le temps :

$$\lambda(t) = e^{\alpha + \beta t} \quad (3-19)$$

Au voisinage de $\beta = 0$, la tendance est linéaire.

Dans le processus de Poisson non homogène, on fait l'hypothèse que les intervalles de temps sont indépendants, et que la probabilité d'occurrence d'un événement dans l'intervalle $(t_{i+1}, t_{i+1} + \Delta t]$ est donnée par :

$$\lambda(t_{i+1}) \exp\left\{-\int_{t_i}^{t_{i+1}} \lambda(u) du\right\} \Delta t + o(\Delta t) \quad (3-20)$$

Si la série est observée dans l'intervalle $(0, t_0]$ et si différents événements résultent aux temps t_1, \dots, t_n , la fonction de vraisemblance est :

$$\begin{aligned} & \lambda(t_1) \exp\left\{-\int_0^{t_1} \lambda(u) du\right\} \lambda(t_2) \exp\left\{-\int_{t_1}^{t_2} \lambda(u) du\right\} \dots \lambda(t_n) \exp\left\{-\int_{t_{n-1}}^{t_n} \lambda(u) du\right\} \exp\left\{-\int_{t_n}^{t_0} \lambda(u) du\right\} \\ & = \left\{\prod_{i=1}^n \lambda(t_i)\right\} \exp\left\{-\int_0^{t_0} \lambda(u) du\right\} \end{aligned} \quad (3-21)$$

En remplaçant $\lambda(u)$ par sa valeur (équation, 3-19), on obtient :

$$\exp\left\{n\alpha + \beta \sum_{i=1}^n t_i - \frac{e^\alpha (e^{\beta t_0} - 1)}{\beta}\right\} \quad (3-22)$$

La distribution conditionnelle est obtenue en divisant (3-22) par la probabilité marginale des n événements. Ainsi, la fonction densité de probabilité conditionnelle des observations est donnée par :

$$\frac{n! \beta^n}{(e^{\beta t_0} - 1)^n} \exp\left(\beta \sum_{i=1}^n t_i\right) \quad (3-23)$$

où $0 \leq t_1 \leq \dots \leq t_n \leq t_0$

Le logarithme de la fonction de vraisemblance conditionnelle est exprimé par :

$$L(\beta) = n \log \beta - n \log(e^{\beta t_0} - 1) + \beta \sum_{i=1}^n t_i + \log n! \quad (3-24)$$

La dérivée première de la fonction $L(\beta)$:

$$L'(\beta) = \begin{cases} \frac{n}{\beta} - \frac{nt_0}{1 - e^{-\beta t_0}} + \sum_{i=1}^n t_i & \beta \neq 0 \\ -\frac{1}{2}nt_0 + \sum_{i=1}^n t_i & \beta = 0 \end{cases} \quad (3-25)$$

et la fonction d'information :

$$I(\beta) = E\left\{-\frac{\partial^2 L}{\partial \beta^2}(\beta)\right\} = \begin{cases} n \left\{ \frac{1}{\beta^2} - \frac{t_0^2 e^{-\beta t_0}}{(1 - e^{-\beta t_0})^2} \right\} & \beta \neq 0 \\ \frac{nt_0^2}{12} & \beta = 0 \end{cases} \quad (3-26)$$

L'estimateur $\hat{\beta}$ du maximum de vraisemblance doit satisfaire $L'(\hat{\beta}) = 0$. Pour tester l'hypothèse :

$$H_0 : \beta = \beta_0$$

On peut utiliser les deux équations (3-25) et (3-26) pour construire la statistique suivante :

$$S_L = \frac{L'(\beta_0)}{\sqrt{I(\beta_0)}} \quad (3-27)$$

Sous l'hypothèse nulle, cette statistique suit asymptotiquement une loi normale centrée réduite $N(0,1)$. Pour des valeurs vérifiant $e^{|\beta_0 t_0|} \gg 1$, l'approximation alternative de la distribution deviendra préférable.

Si on veut tester l'hypothèse :

$$H_0 : \beta = 0$$

On considère la statistique $S = \sum_{i=1}^n T_i$ qui suit une distribution rectangulaire de la somme de n variables aléatoires indépendantes. Ainsi, Cramer (1946) propose la statistique U basée sur l'équation (3-27) pour $\beta = 0$:

$$U = \frac{S - \frac{1}{2}nt_0}{t_0 \sqrt{n/12}} \quad (3-28)$$

qui tend rapidement à la forme normale centrée réduite. Ce test est optimum quand l'hypothèse alternative est basée sur $\lambda(t) = e^{\alpha + \beta t}$ où $\beta \neq 0$.

3.4.3.2 A partir de l'analyse des intervalles

Dans le cas précédent, nous avons supposé que le taux d'occurrence est une fonction exponentielle du temps. Soit la relation suivante :

$$\lambda_i = \theta + \varphi z_i \quad (3-29)$$

Les paramètres θ et φ sont estimés par la méthode du maximum de vraisemblance.

La stationnarité dans ce cas est vérifiée en testant l'hypothèse suivante :

$$H_0 : \varphi = 0$$

Cox et Lewis (1966) ont considéré la statistique $\sum x_i z_i$ conditionnellement à la statistique $\sum x_i$.

Sous l'hypothèse nulle, nous avons un processus de Poisson homogène de durée $\sum x_i$. Soit la statistique d'ordre $U_{(1)}, \dots, U_{(n-1)}$ constituée de $(n-1)$ variables aléatoires indépendantes et distribuées selon une loi uniforme sur l'intervalle $(0, \sum x_i]$, où $U(0) = 0$ et

$U(n) = \sum x_i$. Conditionnellement à $\sum x_i$, la variable aléatoire X peut être écrite sous forme de :

$$x_i = U_{(i)} - U_{(i-1)}$$

Ainsi, on peut formuler la statistique du test :

$$R = \sum_{i=1}^n z_i \{U_{(i)} - U_{(i-1)}\} = \sum_{i=1}^{n-1} (z_i - z_{i+1}) U_{(i)} + z_n \sum_{i=1}^n x_i \quad (3-30)$$

La moyenne et la variance de la statistique R sont données par Cox et Lewis (1966) :

$$E(R) = a \sum_i z_i \quad (3-31)$$

$$\text{var } R = b \sum_i (z_i - \bar{z})^2 \quad (3-32)$$

où a et b sont deux constantes indépendantes de z_i . A titre d'exemple, si $z_i=1, \forall i$ ou si $z_1=1, z_2=...=z_n=0$, alors nous avons :

$$E(R) = \bar{z} \sum_i x_i \text{ et } \text{var } R = \frac{\sum_i (z_i - \bar{z})^2 \left(\sum_i x_i \right)^2}{n(n+1)}$$

Quand la valeur de n est assez grande, la statistique tend vers une loi normale de moyenne exprimée par l'équation (3-31) et de variance donnée par l'équation (3-32).

Cox et Lewis, (1966) ont montré que la statistique :

$$R' = \frac{\sum_i (z_i - \bar{z})(x_i - \bar{x})}{\left(\sum_i (x_i) \right) \left\{ \frac{\sum_i (z_i - \bar{z})^2}{n(n+1)} \right\}^{1/2}} \quad (3-33)$$

suit une loi normale de moyenne zéro et de variance unitaire.

Finalement, on mentionne qu'il est possible d'effectuer une analyse par la méthode du maximum de vraisemblance des nombres d'événements résultant dans différents intervalles.

3.5 Tests des processus de Poisson et de renouvellement

Dans ce paragraphe, nous allons considérer les tests permettant de vérifier la validité du modèle ou du processus ponctuel choisi pour représenter la série d'événements observés. Parmi les modèles choisis, le processus de Poisson et celui de renouvellement, car d'une part, ce sont des processus moins compliqués comparativement à ceux présentés dans la section 2-1 et d'autre part, ils constituent la base de chaque processus plus général. Parmi les hypothèses du processus de renouvellement figure celle de l'indépendance des intervalles et leurs distributions selon une loi donnée. La vérification de cette hypothèse fera l'objet des tests présentés dans cette section. On suppose que l'hypothèse de stationnarité est vérifiée.

3.5.1 Tests d'ajustement indépendants de la distribution "free-distribution"

Ces tests sont appliqués dans le but de trouver la distribution adéquate pour les intervalles de temps du processus (modèle de renouvellement). Cependant, ils peuvent être utilisés dans d'autres approches concernant l'analyse de séries d'événements.

3.5.1.1 Généralités

On suppose que l'on dispose d'une série de variables aléatoires (v.a) représentée par une fonction de répartition inconnue $F_X(x)$. Soit $F_0(x)$ une distribution spécifique telle que les hypothèses suivantes sont respectées :

$$H_0 : F_X = F_0;$$

$$H_1^{(a)} : F_X \neq F_0, \text{ (test bilatéral) ; ou}$$

$$H_1^{(b)} : F_X \geq F_0, \text{ (test unilatéral) ; ou}$$

$$H_1^{(c)} : F_X \leq F_0, \text{ (test unilatéral).}$$

Les statistiques utilisées sont basées sur la distance entre $F_0(x)$ et la distribution empirique

$F_n(x)$, où : $F_n(x) = \frac{\text{nombre de } x_i \leq x}{n}$. Ces distributions possèdent l'appellation "free-

distribution" dans le sens que sous l'hypothèse nulle, la distribution statistique ne dépend pas de l'égalité $F_0(x)=F_x(x)$.

Si $x_{(i)}$ ($i=1, \dots, n$) sont les observations d'ordre où $x_{(0)}=0$ et $x_{(n+1)}=\infty$, alors le pas de $F_n(x)$ est égal à $1/n$ pour chaque point $x_{(i)}$ et $F_n(x) = i/n$ pour chaque x tel que : $x_{(i)} \leq x \leq x_{(i+1)}$. Une forme plus simple est définie par :

$$I(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x - x_i) = \text{moyenne de } I(x - x_i)$$

Dans ce qui suit, nous présentons un théorème utilisé par Cox et Lewis, (1966) dans leur analyse de la série d'événements :

- **Théorème de la transformation intégrale de probabilité**

Si une variable aléatoire X suit une distribution $F_x(x)$, alors $U = F_x(x)$ est une v.a uniformément distribuée sur un intervalle $(0,1]$.

Ainsi, si on a n v.a X_1, X_2, \dots, X_n distribuées selon $F_x(x)$ et si l'on considère les statistiques d'ordre $X_{(1)} \leq X_{(2)} \leq \dots \leq X_{(n)}$ alors $U_1 = F_x(X_1), U_2 = F_x(X_2), \dots, U_n = F_x(X_n)$ suivent une loi uniforme sur un intervalle $(0,1]$, où les statistiques d'ordre sont : $U_{(1)} = F_x(X_{(1)}), U_{(2)} = F_x(X_{(2)}), \dots, U_{(n)} = F_x(X_{(n)})$. Ainsi, le problème d'ajustement de la loi des variables X_i revient à tester si les points observés $u_1 = F_0(x_1), u_2 = F_0(x_2), \dots, u_n = F_0(x_n)$ divisent l'intervalle $(0,1]$ uniformément.

3.5.1.2 Tests bilatéral et unilatéral de Kolmogorov-Smirnov

On définit les statistiques du test de Kolmogorov-Smirnov pour les cas suivants :

- test unilatéral à droite :

$$D_n^+ = \sup_{-\infty < x < +\infty} \{F_n(x) - F_0(x)\} = \max_{1 \leq i \leq n} \left\{ \frac{i}{n} - F_0(x_{(i)}) \right\} \quad (3-34)$$

- test unilatéral à gauche :

$$D_n^- = \sup_{-\infty < x < +\infty} \{F_0(x) - F_n(x)\} = \max_{1 \leq i \leq n} \left\{ F_0(x_{(i)}) - \frac{i-1}{n} \right\} \quad (3-35)$$

- test bilatéral :

$$D_n = \sup_{-\infty < x < +\infty} |F_n(x) - F_0(x)| = \max(D_n^+, D_n^-) \quad (3-36)$$

Il faut noter que les distributions asymptotiques ne suivent pas une loi normale. Le test bilatéral rejette l'hypothèse H_0 à un niveau de signification α ($\text{Prob}(D_n > d_{n,\alpha}) = \alpha$) si la valeur de D_n observée est supérieure à $d_{n,\alpha}$. Ce test est consistant, mais il est biaisé (Massey, 1950), cependant les tests unilatéraux sont consistants et non biaisés (Cox et Lewis, 1966).

Pour une distribution empirique donnée, le test basé sur la statistique D_n acceptera H_0 à un niveau de signification α , $\forall x$, si :

$$F_n(x) - d_{n,\alpha} \leq F_0(x) \leq F_n(x) + d_{n,\alpha} \quad (3-37)$$

3.5.1.3 Test de Cramer-von Mises

La statistique du test est définie par la quantité suivante :

$$W_n^2 = \int_{-\infty}^{+\infty} \{F_n(x) - F_0(x)\}^2 dF_0(x) \quad (3-38)$$

Cette quantité mesure l'erreur quadratique moyenne ou la distance entre $F_n(x)$ et $F_0(x)$. La distribution asymptotique est donnée par ;

$$W_n^2 = \frac{1}{12n^2} + \frac{1}{n} \sum_{i=1}^n \left\{ F_0(x_{(i)}) - \frac{2i-1}{2n} \right\}^2 \quad (3-39)$$

Cette distribution, qui ne suit pas une loi normale, a été obtenue par Smirnov (1949).

Pour les échantillons de petite taille, W_n^2 ne converge pas toujours rapidement. Cependant, pour $n > 20$, la distribution asymptotique peut être utilisée.

3.5.1.4 Test de Stépens (Gomes et al, 1975)

La statistique du test est donnée par :

$$U_n^2 = n \int \int_{-\infty-\infty}^{+\infty+\infty} \left\{ F_n(x) - F(x) - \int_{-\infty}^{+\infty} [F_n(y) - F(y)] dF(y) \right\}^2 dF(x) \quad (3-40)$$

dont la distribution asymptotique est :

$$P(U^2 \leq z) = \lim_{n \rightarrow \infty} \text{Pr ob}(U_n^2 \leq z) = 1 - 2 \sum_{k=1}^{+\infty} (-1)^{k-1} \exp(-2\pi^2 k^2 z) = Q(\pi\sqrt{z}) \quad (3-41)$$

où $Q(x)$ est la fonction de distribution limite de la statistique de Kolmogorov-Smirnov.

3.5.1.5 Test de Sherman (Gomes et al, 1975)

La statistique du test est donnée par l'équation suivante :

$$S_n = \frac{1}{2} \sum_{i=1}^{n+1} \left| F(x_{(i)}) - F(x_{(i-1)}) - \frac{1}{n+1} \right| \quad (3-42)$$

où :

$$F(x_{(0)}) = 0 \quad \text{et} \quad F(x_{(n+1)}) = 1$$

La distribution asymptotique de la variable aléatoire :

$$\left(\frac{ne^2}{2e-5} \right)^{1/2} \left(S_n - \frac{1}{e} \right) \text{ est une distribution normale centrée réduite, } N(0,1).$$

3.5.1.6 Autres tests

Gomes et al (1975) ont présenté des tests d'ajustement qui ne sont applicables que si la distribution de la population testée est absolument continue, c'est-à-dire, s'il existe $f(x)=F'(x)$. Ces tests sont basés sur les statistiques suivantes:

$$D_{1n} = \frac{1}{n} \sum_{j=1}^n \text{Log}f(x_j) \quad (3-43)$$

$$D_{2n} = \frac{1}{n} \sum_{j=1}^n f^2(x_j) \quad (3-44)$$

Ces statistiques sont asymptotiquement distribuées selon une loi normale. La moyenne et la variance sont calculées à partir de la fonction densité de probabilité f considérée.

Les tests basés sur les statistiques D_n et W_n^2 sont des tests indépendants de la distribution statistique, consistants et n'exigent pas une subdivision des données. Cependant, le test du Chi-deux est consistant, mais ne possède pas les deux autres propriétés, d'où l'avantage de l'utilisation des deux premières statistiques (Cox et Lewis, 1966).

3.5.1.7 Modification de Durbin (Durbin, 1961)

On considère maintenant une importante modification des tests d'ajustement indépendants de la distribution due à Durbin (1961) et qui est basée sur la transformation des observations au lieu de la modification de la statistique du test, telle que D_n ou W_n^2 . L'objectif de cette transformation est d'accroître la puissance des tests.

Sous l'hypothèse nulle, les v.a $U_{(i)}$ sont distribuées selon une loi uniforme d'intervalle (0,1] et possède une fonction densité de probabilité $f(u_{(1)}, u_{(2)}, \dots, u_{(n)}) = n!$ où $0 \leq u_{(1)} \leq \dots \leq u_{(n)} \leq 1$. Soit le changement de variable suivant : $C_1 = U_{(1)}$, $C_2 = U_{(2)} - U_{(1)}$, ..., $C_i = U_{(i)} - U_{(i-1)}$, $i=2, \dots, n$ et $C_{n+1} = 1 - U_{(n)}$.

Le Jacobien de cette transformation est égal à l'unité, ainsi les valeurs de C_i ont une fonction densité de probabilité exprimée par :

$$f(C_1, \dots, C_n) = n! \quad \text{où : } C_i \geq 0, \dots, C_{n+1} \geq 0 \text{ et } \sum_{i=1}^{n+1} C_i = 1$$

$$f(C_{(1)}, \dots, C_{(n)}) = n!(n+1)! \quad \text{où : } 0 \leq C_{(1)} \leq \dots \leq C_{(n+1)}, \text{ et } \sum_{i=1}^{n+1} C_{(i)} = 1$$

Finalement, les valeurs de $C_{(i)}$ sont transformées de la façon suivante:

$$C'_i = (n+2-i)(C_{(i)} - C_{(i-1)}) \quad (3-45)$$

où : $C_0 = 0$ et $i=1, \dots, n+1$.

Le Jacobien de cette transformation est égal à $1/(n+1)!$. Ainsi, la fonction densité de probabilité des C'_i est :

$$f(C'_1, \dots, C'_n) = n! \text{ où } C'_1 \geq 0, \dots, C'_{n+1} \geq 0 \text{ et } \sum_{i=1}^{n+1} C'_i = 1$$

Cette fonction de densité est identique à celle des C_i , on peut former les sommes suivantes :

$$\begin{aligned} U'_{(i)} &= \sum_{r=1}^i C'_r \\ &= C'_{(1)} + \dots + C'_{(i-1)} + (n+2-i)C'_{(i)} \quad (i = 1, \dots, n) \end{aligned} \quad (3-46)$$

Les variables d'ordre $U'_{(i)}$ sont des variables aléatoires distribuées selon une loi uniforme d'intervalle (0,1]. Finalement, on peut appliquer les tests d'ajustement cités précédemment à la nouvelle variable $U'_{(i)}$.

3.5.2 Tests relatifs au processus de Poisson

3.5.2.1 Test d'uniformité conditionnelle (intervalles)

Soit une série d'événements de taille n_0+1 dont les temps d'occurrence sont t_1, \dots, t_{n_0+1} . X_i représentent les temps séparant ces événements ($i=1, \dots, n_0+1$). Ainsi, l'hypothèse nulle est formulée de la façon suivante :

$$\mathbf{H}_0 : F(x_1, \dots, x_{n_0+1}) = \prod_{i=1}^{n_0+1} \left\{ 1 - \exp\left(-\frac{x_i}{E(X)}\right) \right\} : \text{ la fonction de répartition de } (n_0+1)$$

intervalles (Processus de Poisson homogène).

$$E(X) = 1/\lambda > 0.$$

Barnard (1953) a proposé un changement de variable $U_{(i)} = \frac{T_i}{T_{n_0+1}}$ conditionnellement à la

valeur observée t_{n_0+1} de T_{n_0+1} . $U_{(i)}$ est une v.a indépendamment distribuée selon une loi uniforme sur l'intervalle (0,1]. Ainsi, $E(X) = 1/\lambda$, "paramètre de nuisance", est éliminé. Le test est appelé test d'uniformité conditionnelle. Finalement, on peut tester cette hypothèse par un test d'ajustement, (section 3-5-1).

Ce test est sensible à l'hypothèse de stationnarité. Ainsi, on peut l'utiliser comme un moyen pour détecter l'existence d'une tendance dans la série observée d'un processus de Poisson.

La statistique $S=U_1 + \dots + U_n$ fournit le plus puissant test de l'hypothèse nulle (Cox et Lewis, 1966) sachant que l'hypothèse alternative est :

H_1 : La série d'événements provient d'un processus de Poisson non homogène.

Cette hypothèse peut être représentée par l'équation (3-19). La statistique S , sous l'hypothèse nulle, tend rapidement vers une distribution normale $N\left(\frac{1}{2}n_0, \frac{1}{12}n_0\right)$.

Bartholmew (1956) a comparé la puissance du test basé sur la statistique S avec la puissance du test d'uniformité conditionnelle en utilisant les statistiques $D_{n_0}^+, D_{n_0}^-, D_{n_0}$ en considérant l'hypothèse alternative de tendance exprimée par l'équation (3-19). Il a trouvé que dans le cas où une tendance de la fonction d'occurrence est présente, le test basé sur la statistique S possède une plus grande puissance comparativement aux autres tests.

Il faut remarquer que les résultats qu'on vient d'établir sont valables seulement sous l'hypothèse que les variables aléatoires U_i soient indépendamment et identiquement distribuées.

3.5.2.2 Application de la transformation de Durbin des U_i

Soient les variables $(X_{(i)} - X_{(i-1)})$, $i=1, \dots, n_0+1$ où $X_0=0$ indépendamment distribuées selon une loi exponentielle avec une moyenne $\frac{1}{\lambda(n_0 + 2 - i)}$, (Cox et Lewis, 1966). Ainsi, les v.a $(n_0+2-i)(X_{(i)} - X_{(i-1)})$ représentent les temps successifs entre les événements du processus de Poisson de n_0+1 événements et $T_{n_0+1} = t_{n_0+1}$. Les variables :

$U'_{(i)} = \frac{1}{t_{n_0+1}} \{X_{(1)} + X_{(2)} + \dots + (n_0 + 2 - i)X_{(i)}\}$ ($i=1, \dots, n_0$) sont les statistiques d'ordre d'un

échantillon aléatoire de taille n_0 tiré d'une loi uniforme d'intervalle $(0,1]$.

Une statistique, qui est simple à calculer, est la statistique du test moyen modifié (Lewis, 1965) basée sur :

$$S' = \sum_{i=1}^{n_0} U'_{(i)} = \frac{2}{t_{n_0+1}} \sum_{i=1}^{n_0} (n_0 - i + 1) X_{(i)} \quad (3-47)$$

Identiquement à la statistique S, la statistique S' tend rapidement sous l'hypothèse nulle vers une distribution normale $N\left(\frac{1}{2}n_0, \frac{1}{12}n_0\right)$.

3.5.2.3 Test de dispersion (nombre d'événements)

Soient n_1, n_2, \dots, n_k k observations d'une v.a discrète N et soit $\tilde{n} = \frac{n_1 + \dots + n_k}{k}$ la moyenne des k observations. Un test permettant de vérifier l'hypothèse que n_i proviennent d'une loi de Poisson de paramètre μ est le test de dispersion d'homogénéité basé sur la statistique :

$$d = \sum_{i=1}^k \frac{(n_i - \tilde{n})^2}{\tilde{n}} \quad (3-48)$$

Sous l'hypothèse nulle, cette statistique suit asymptotiquement une distribution Chi-deux à (k-1) degrés de liberté (Rao et Chakravatti, 1956). Sukhatme (1938) a montré que l'approximation de d est adéquate pour des valeurs de $k > 5$ si $\tilde{n} > 1$ ou bien pour des valeurs $k > 15$ si $\tilde{n} < 1$. Cunnane (1979) a indiqué que l'approximation de la loi de la statistique d est valable pour des valeurs de $\tilde{n} > 5$.

La statistique $d/(k-1)$ représente le rapport de la variance estimée de N à sa moyenne estimée et le test est basé sur la comparaison de cette quantité et du rapport observé. Les hypothèses alternatives peuvent être formulées selon deux formes : n_i suit une loi de Poisson avec un paramètre $\mu_1 \neq \mu$ ou n_i ne suit pas une loi de Poisson.

On peut montrer que ce test appliqué au processus de Poisson est équivalent au test d'uniformité conditionnelle quand le test d'ajustement Chi-deux est choisi pour vérifier la loi uniforme des U_i .

Le test de dispersion est fiable quand l'hypothèse alternative représente l'hétérogénéité de l'échantillon (Les intervalles choisis ont des taux d'occurrence différents). Pour d'autres types d'hypothèses alternatives, le choix de t ou de k représente un problème majeur dans l'application de ce test (Cox et Lewis, 1966).

3.5.2.4 Tests avec l'hypothèse alternative : processus de renouvellement

Dans cette section, on considère les tests vérifiant l'hypothèse que le processus est poissonien et en fixant comme hypothèse alternative le processus de renouvellement :

H_0 : la série d'événements provient d'un processus de Poisson ;

H_1 : la série d'événements provient d'un processus de renouvellement.

Ces deux hypothèses peuvent être formulées d'une façon équivalente :

H_0 : les intervalles de temps entre les événements sont indépendants et suivent une distribution exponentielle ;

H_1 : les intervalles suivent une distribution Gamma à deux paramètres.

On rappelle que la fonction densité de la loi Gamma à deux paramètres, (μ, κ) est donnée par :

$$f(x) = \frac{\kappa x^{\kappa-1} e^{-\kappa x / \mu}}{\mu \Gamma(\kappa)} \quad (3-49)$$

Selon la notation utilisée dans l'équation (3-49), on veut tester

H_0 : $\kappa=1$

H_1 : $\kappa \neq 1$

avec le paramètre μ non spécifié. L'estimation de κ est issu de la méthode du maximum de vraisemblance (Cox et Lewis, 1966).

La statistique du test est tirée de Moran (1951) :

$$l_n = \frac{-\text{Log} \bar{x} + \frac{1}{n} \sum_{i=1}^n \text{Log} x_i}{1 + \frac{n+1}{6n}} \quad (3-50)$$

Sous l'hypothèse nulle, l_n est approximativement distribué selon une loi du Chi-deux à $(n-1)$ degrés de liberté. Le dénominateur a été introduit par Bartlett (1937) pour améliorer cette approximation. Ce test demeure non efficace sur des échantillons de faible taille.

Bartholomew, (1957) a comparé la puissance de ce test avec d'autres statistiques. Parmi ces tests, figure la statistique proposée par Sherman (1950). Cette statistique, dans sa forme originale, est une statistique de test d'ajustement indépendant de la distribution.

L'estimation de la statistique d'ordre $E(U_{(i)}) = \frac{i}{n+1}, i=1, \dots, n$ ainsi que

$\{E(U_{(i)} - U_{(i-1)})\} = \frac{1}{n+1}$. La statistique de Sherman mesure l'écart entre cette estimation

et la valeur de la différence $(U_{(i)} - U_{(i-1)})$:

$$\bar{w}_n = \frac{1}{2} \sum_{i=1}^{n+1} \left| U_{(i)} - U_{(i-1)} - \frac{1}{n+1} \right| = \frac{\sum_{i=1}^{n+1} |X_i - \bar{X}|}{2(n+1)\bar{X}} \quad (3-51)$$

$$\text{où : } \bar{X} = \frac{X_1 + \dots + X_{n+1}}{n+1} = \frac{t_{n+1}}{n+1}$$

Sous l'hypothèse nulle, on a :

$$E(\bar{w}_n) \approx \frac{1}{e} = 0,368 \text{ et } \text{var}(\bar{w}_n) = \frac{0,05908}{n} - \frac{0,07145}{n^2} + o\left(\frac{1}{n^3}\right)$$

et la statistique centrée réduite suit asymptotiquement une loi normale $N(0,1)$.

Bartolomew, (1957) a montré que les tests basés sur cette statistique sont consistants en considérant l'hypothèse alternative H_1 .

3.5.2.5 Détection de rupture pour des observations poissonniennes

Ghorbanzadeh et Lounes (1996) ont présenté deux méthodes pour la détection de rupture des observations poissonniennes. Soit X_1, \dots, X_n une suite de variables aléatoires. Les hypothèses du test sont :

$$H_0: P(X_i = x_i) = \exp\left(-n_i \lambda_0 \frac{(n_i \lambda_0)^{x_i}}{x_i!}\right) \quad i=1, \dots, n$$

$$H_1^{(N_n)} := \bigcup_{k=1}^{n-1} H_1\left(\frac{N_k}{N_n}\right)$$

où :

$$H_1\left(\frac{N_k}{N_n}\right) : P(X_i = x_i) = \begin{cases} \exp\left(-n_i \lambda_1 \frac{(n_i \lambda_1)^{x_i}}{x_i!}\right) & \text{si } i = 1, \dots, k \\ \exp\left(-n_i \lambda_2 \frac{(n_i \lambda_2)^{x_i}}{x_i!}\right) & \text{si } i = k+1, \dots, n \end{cases}$$

où

$$\sum_{i=1}^n n_i = N_n, \quad \sum_{i=1}^k n_i = N_k, \quad \sum_{i=k+1}^n n_i = N_k^*$$

λ_0, λ_1 et λ_2 sont les paramètres de la loi de Poisson.

Il faut noter que chaque variable X_i peut être considérée comme une somme de n_i variables aléatoires indépendantes et distribuées suivant la loi de Poisson.

Le test proposé dans cette étude est basé sur la statistique de la Log-vraisemblance maximale pénalisée. La loi asymptotique de cette statistique se comporte comme celle d'une fonctionnelle d'un pont brownien (Ghorbanzadeh et Lounes, 1996).

On définit un Pont Brownien par :

Pour tout $t \in (0,1]$, $V(t)$ est un pont brownien, c'est-à-dire un processus gaussien sur $(0,1]$ conditionné par $V(0)=V(1)=0$, on sait alors que pour $t_1 \in (0,1]$ et $t_2 \in (0,1]$:

$$\text{Cov}(V(t_1), V(t_2)) = \min(t_1, t_2) - t_1 t_2$$

3.5.2.5.1 Test du rapport de vraisemblance

La statistique Log-vraisemblance est définie par :

$$A_{N_n} = \max_{1 < k < n} A_{N_n}^* \left(\frac{N_k}{N_n} \right) \quad (3-52)$$

où :

$$A_{N_n}^* \left(\frac{N_k}{N_n} \right) = N_k \bar{X}_{N_k} \log \left(\frac{\bar{X}_{N_k}}{\bar{X}_{N_n}} \right) + N_k^* \bar{X}_{N_k}^* \log \left(\frac{\bar{X}_{N_k}^*}{\bar{X}_{N_n}} \right)$$

où

$$\frac{1}{N_n} \sum_{i=1}^n X_i = \bar{X}_{N_n}, \quad \frac{1}{N_k} \sum_{i=1}^k X_i = \bar{X}_{N_k} \text{ et } \frac{1}{N_k^*} \sum_{i=k+1}^n X_i = \bar{X}_{N_k}^*$$

La valeur \hat{k} de k réalisant le maximum de $A_{N_n}^* \left(\frac{N_k}{N_n} \right)$ est l'estimateur du maximum de vraisemblance de k sous l'hypothèse alternative.

Soit l'application suivante :

$$g_n: [0,1] \rightarrow [0,1]$$

$$t \rightarrow g_n(t) = \frac{1}{N_n} \sum_{i=1}^{[nt]} n_i$$

[A] désigne la partie entière du réel A.

on remarque que si $t = \frac{k}{n}$ alors $g_n \left(\frac{k}{n} \right) = \frac{N_k}{N_n}$

On peut démontrer que sous l'hypothèse H_0 , $A_{N_n}^*$ converge au sens de la convergence en loi des marginales de dimension finie vers le processus Λ_0 défini par :

$$\lambda_0(t) = \frac{V^2(g_n(t))}{2g_n(t)(1-g_n(t))} \quad (3-53)$$

où V est un pont brownien.

Afin d'éliminer les problèmes de bord où la fonction Λ_0 n'est pas continue pour des valeurs de t au voisinage de 0 ou 1. Ghorbanzadeh et Lounes (1996) ont adopté la méthode suivante :

soit la suite de processus $\Gamma_{N_n}(t) = \Psi(t)\Psi(1-t)\Lambda_{N_n}^*(t)$ où Ψ est une fonction continue

par morceaux sur $(0,1]$ vérifiant $\int_0^1 \left(\frac{\Psi(t)}{t}\right)^2 dt < \infty$.

Posons $\Gamma(t) = \Psi(t)\Psi(1-t)\Lambda_0(t)$ le processus continu pour $t \in (0,1]$ en considérant les moments d'ordre 4 de $\{\Gamma_{N_n}(t), t \in (0,1]\}$, on montre que $\{\Gamma_{N_n}(t), t \in (0,1]\}$ converge en loi vers $\{\Gamma(t), t \in (0,1]\}$.

Alors sous l'hypothèse H_0 , pour tout $\eta \geq 0$, on a :

$$\lim_{N_n \rightarrow \infty} P\left(\sup_{t \in [0,1]} \Gamma_{N_n}(t) \geq \eta\right) = P\left(\sup_{t \in [0,1]} \Gamma(t) \geq \eta\right) \quad (3-54)$$

Ainsi, la zone de rejet est :

$$\left\{ \sup_{t \in [0,1]} \Gamma_{N_n}(t) \geq \eta \right\} \quad (3-55)$$

où pour un α donné dans $(0,1]$, le seuil η_α est déterminé par :

$$P\left(\sup_{t \in [0,1]} \Gamma(t) \geq \eta\right) \leq \alpha$$

3.5.2.5.2 Approche bayésienne

Dans cette partie, Ghorbanzadeh et Lounes (1996) ont proposé le modèle suivant :

$$X_i \sim \text{Poisson}(n_i \lambda_1) \text{ si } i = 1, \dots, k$$

$$X_i \sim \text{Poisson}(n_i \lambda_2) \text{ si } i = k + 1, \dots, n$$

Le test étudié dans ce cas est le suivant :

$$H_0 : k=n \text{ (pas de rupture } \lambda_1 = \lambda_2)$$

$$H_1 : 1 \leq k \leq n-1 \text{ (rupture à } k)$$

On suppose que les trois paramètres k , λ_1 et λ_2 sont indépendants.

Le test est basé sur la loi a posteriori du paramètre k de l'instant de rupture. L'avantage de cette méthode est qu'elle n'exige pas de loi asymptotique.

Cependant, il faut choisir pour les paramètres λ_1 , λ_2 et k les lois à priori. Ghorbanzadeh et Lounes (1996) ont choisi la forme suivante pour λ_1 et λ_2 :

$$f(\lambda_j) = \frac{1}{\lambda_j}, j = 1, 2.$$

Ils ont justifié ce choix par des raisons techniques et par la conformité avec ce qui est utilisé dans la littérature.

Pour le paramètre k , plusieurs lois a priori ont été proposées par Ghorbanzadeh et Lounes (1996):

$$v_0(k) = pI_{\{k=n\}} + \frac{1-p}{n-1} I_{\{1 \leq k \leq n-1\}} \quad (3-56)$$

$$v_1(k) = pI_{\{k=n\}} + \frac{p(1-p)^k}{1-(1-p)^{n-1}} I_{\{1 \leq k \leq n-1\}} \text{ (tirée d'une loi géométrique)} \quad (3-57)$$

$$v_2(k) = pI_{\{k=n\}} + \frac{C_{n-1}^k p^k (1-p)^{n-k}}{1-(1-p)^{n-1}} I_{\{1 \leq k \leq n-1\}} \text{ (tirée d'une loi binomiale)} \quad (3-58)$$

Les lois a posteriori pour le paramètre k sont (pour $j=0,1,2$) :

$$\pi_j(k / x_1, \dots, x_n) \propto \begin{cases} v_j(k) \frac{\gamma(s_n)}{N_n^{s_n}} & \text{si } k = n \\ v_j(k) \frac{\gamma(s_k) \gamma(s_k^*)}{N_n^{s_n} (N_k^*)^{s_k^*}} & \text{si } k \neq n \end{cases} \quad (3-59)$$

où $\gamma(\cdot)$ désigne la fonction Gamma, $s_k = \sum_{i=1}^k x_i$, $s_k^* = \sum_{i=k+1}^n x_i$.

Ainsi, une probabilité a posteriori maximale nous indique le point de la rupture dans les données.

Chacune des procédures possède ses avantages et ses inconvénients. La première procédure semble plus appropriée dans le cas de l'application des Ghorbanzadeh et Lounes (1996), car les données sont nombreuses ce qui justifie l'utilisation de l'approximation asymptotique. Cependant, cette méthode de détection nécessite un grand temps de calcul, des seuils asymptotiques et un choix des fonctions de pénalisation. La méthode bayésienne est plus

simple à mettre en oeuvre, cependant, ses résultats sont fortement liés aux choix de la loi a priori du paramètre de rupture.

3.5.2.6 Tests entre le processus de Poisson et le processus de Poisson doublement stochastique

Driancourt et Streit (1983) ont étudié des tests statistiques qui permettent de choisir entre un processus ponctuel de Poisson et un processus ponctuel doublement stochastique de Poisson.

3.5.2.6.1 Modèles de tendance exponentielle

Les événements sont réalisés dans l'intervalle $(0, s]$ et $N(s)$ est le nombre des événements ponctuels dans cet intervalle et soient $T(1) < T(2) < \dots < T(s)$ leurs temps d'occurrence. Les événements sont représentés par un processus ponctuel P_{θ^*} .

Les deux hypothèses à tester sont :

H₀ : Etant donné $\Lambda_0 = \lambda_0$, le processus ponctuel conditionnel $P_{\theta^*} | \Lambda_0 = \lambda_0$ est un processus ponctuel homogène de Poisson à intensité globale $\lambda(t) = \lambda_0(t) = \lambda_0$. Λ_0 suit une fonction de répartition connue F_Λ où $F_\Lambda(0) = 0$.

H₁ : Etant donné $\Lambda_1 = \lambda_1$ et $B = \beta$, le processus ponctuel conditionnel $P_{\theta^*} | \Lambda_1 = \lambda_1, B = \beta$ est un processus non-homogène de Poisson à fonction d'intensité $\lambda(t) = \lambda_1(t) = \lambda_1 e^{-\beta t}$.

Les distributions de Λ_1 et B sont spécifiées par les hypothèses suivantes concernant leurs fonctions de répartition.

$\Lambda_1 \sim F_\Lambda$ où $F_\Lambda(0) = 0$.

$B \sim F_B$ où $F_B(0) = 0$.

Driancourt et Streit (1983) ont traité plusieurs cas selon les paramètres du modèle :

Premier cas :

Considérons le cas où λ_0 est connu et β_0 est inconnu, c'est-à-dire que on désire discriminer entre les hypothèses suivantes :

H_0 : Les données sont représentées par un processus de Poisson homogène

H_1 : Les données sont représentées par un processus de Poisson non homogène.

Un test uniformément le plus puissant par rapport à $\beta_0 > 0$ pour comparer H_0 et H_1 sous les conditions indiquées utilise la statistique :

$$U(s) = \sum_{i=1}^{N(s)} T(i) \quad (3-60)$$

La région critique est définie par :

$$u(s) < U(s, \alpha)$$

où $U(s, \alpha)$ est la valeur réelle la plus grande satisfaisant :

$$P(U(s) < U(s, \alpha); \lambda(t) = \lambda_0) \leq \alpha; \alpha \text{ est le niveau de signification choisi.}$$

Asymptotiquement, pour $s \rightarrow \infty$, la région critique est définie par :

$$\sqrt{\frac{3}{\lambda_0 s^3}} \left(u(s) - \frac{\lambda_0 s^2}{2} \right) < z(\alpha) \quad (3-61)$$

où $z(\alpha)$ est déterminé par $\Phi(z(\alpha))$, la fonction de répartition de la distribution normale centrée réduite.

Ce test est également le test du rapport généralisé des vraisemblances pour comparer H_0 et H_1 où $\beta_0 > 0$ non-spécifié. Le test du rapport des vraisemblances compare le maximum de la fonction de vraisemblance par rapport à $\beta_0 \geq 0$ et à sa valeur $\beta_0 = 0$. On utilise la statistique :

$$A_{\beta_0 \geq 0}^* = \exp \left[2 \lambda_0 s (\hat{B}_0 s)^{-1} (\exp(-\hat{B}_0 s) - 1) + \lambda_0 s (\exp(-\hat{B}_0 s) + 1) \right] \quad (3-62)$$

\hat{B}_0 est l'estimateur selon la méthode du maximum de vraisemblance de β_0 . $A_{\beta_0 \geq 0}^*$ est une fonction croissante en $\hat{B}_0 = \hat{\beta}_0$. Ainsi, \hat{B}_0 est une statistique équivalente à $A_{\beta_0 \geq 0}^*$.

La région critique prend la forme :

$$\hat{\beta}_0 > \hat{B}_0 (1 - \alpha)$$

où :

$$P(\hat{B}_0 > \hat{B}_0 (1 - \alpha); \lambda(t) \equiv \lambda_0) \leq \alpha$$

\hat{B}_0 satisfait à la relation :

$$U(s) = \lambda_0 \hat{B}_0^{-2} \left[1 - e^{-\hat{B}_0 s} - \hat{B}_0 s e^{-\hat{B}_0 s} \right]$$

pourvu que $\hat{B}_0 \neq 0$ ($U(s) < \frac{1}{2} \lambda_0 s^2$)

On trouve également pour $\hat{B}_0 \neq 0$ que

$$(\hat{B}_0 s)^{-2} \gamma(2, \hat{B}_0 s) = U(s) (\lambda_0 s^2)^{-1}$$

où :

$\gamma(2, x)$ est la fonction gamma incomplète, ainsi, \hat{B}_0 peut être déterminé numériquement.

Comme $(\hat{B}_0 s)^{-2} \gamma(2, \hat{B}_0 s)$ est une fonction décroissante de $\hat{B}_0 = \hat{\beta}_0$, la région critique indiquée est équivalente à $u(s) < U(s, 1-\alpha)$.

Pour déterminer la valeur critique $U(s, \alpha)$, il faut connaître la distribution de $U(s)$ sous l'hypothèse H_0 .

Pour s fini, la loi de probabilité de $U(s)$ est spécifiée par :

$$P(U(s) = 0; \lambda_0, \beta_0 = 0) = e^{-\lambda_0 s}$$

$$f_{U(s)}(u; \lambda_0, \beta_0 = 0) = \lambda_0 e^{-\lambda_0 s} \sum_{\nu=0}^{\infty} \left[(\nu!)^{-1} \lambda_0^{\frac{\nu-1}{2}} (u - \nu s)_+^{\frac{\nu-1}{2}} 2I_{|\nu-1|} \sqrt{\lambda_0 (u - \nu s)_+} \right] \quad (3-63)$$

I_ν désigne la fonction modifiée de Bessel d'ordre ν et $(x)_+ = \frac{x + |x|}{2}$.

Deuxième cas

Dans le cas où $B = \beta_0$ est fixé et Λ suit une distribution non-uniponctuelle. Soit la fonction de répartition F_Λ de Λ et soit β_0 connu. Dans ce cas, on considère l'aspect stochastique de la fonction d'intensité tout en fixant le paramètre β_0 . Driancourt et Streit (1983) ont défini la statistique du test en se basant sur le résultat du théorème suivant :

Soit :

$$P_{n, F_\Lambda}(t) = \int_0^\infty e^{-\lambda t} (\lambda t)^n (n!)^{-1} dF_\Lambda(\lambda); \quad [n \in \{0, 1, 2, \dots\}, t \geq 0]$$

Sous les conditions mentionnées, le test le plus puissant pour comparer H_0 et H_1 est obtenu en utilisant comme statistique du test :

$$\Lambda^* = (\beta_0 s)^{N(s)} (1 - e^{-\beta_0 s})^{-N(s)} \exp[-\beta_0 U(s)] P_{N(s), F_A}(\beta_0^{-1} (1 - e^{-\beta_0 s})) [P_{N(s), F_A}(s)]^{-1} \quad (3-64)$$

Λ^* ne dépend que de l'observation de P_{θ^*} qu'en terme de la statistique bivariée $(N(s), U(s))$.

Troisième cas

Dans ce cas tous les paramètres du modèle sont des variables aléatoires. On analyse le cas où Λ et B suivent indépendamment les uns des autres des lois de probabilité non-uniponctuelles à fonction de répartition F_A et F_B , dont on suppose que toutes les sommes existent. Sous ces conditions, le test le plus puissant pour comparer H_0 et H_1 au niveau de signification α est exprimé par la statistique :

$$\Lambda^* = L(F_A, F_B; N(s); T(I), \dots, T(N(s))) [L(F_A, F_{B=0}; N(s); T(I), \dots, T(N(s)))]^{-1} \quad (3-65)$$

où :

L est la fonction du maximum de vraisemblance :

$$L(F_A, F_B; N(s); T(I), \dots, T(N(s) = t_n)) = \sum_{m=0}^{\infty} \sum_{v=0}^{\infty} \sum_{w=0}^{\infty} \frac{(-I)^{m+v+w} s^{m+w}}{v!(m+w)!} S_{m+w}^{(m)} \left(\sum_{l=1}^n t_l \right)^v E_{F_A} [A^{n+m}] E_{F_B} [B^{n+m}]$$

où

$S_{m+w}^{(m)}$ désigne un nombre de seconde espèce de Stirling. Ce nombre est déterminé à partir de la formule suivante (Abramowitz et Stegun, 1964) :

$$S_n^{(m)} = \frac{I}{m!} \sum_{k=0}^m (-I)^{m-k} C_m^k k^n$$

La région critique est définie de la façon suivante :

$\Lambda^* > \Lambda^*(1-\alpha)$ où $\Lambda^*(1-\alpha)$ est la valeur de $\tilde{\lambda}^*$ la plus petite satisfaisant l'inégalité

$$P_{\theta^*}(\Lambda^* > \tilde{\lambda}^*) \leq \alpha$$

3.5.2.6.2 Modèles alternatifs à déviation relativement lente du niveau original d'intensité

Des techniques similaires aux cas précédents peuvent être appliquées pour construire des tests comparant H_0 à l'hypothèse alternative, que le processus est un **processus de Poisson non-homogène**. Deux cas particuliers ont été étudiés par Driancourt et Streit (1983):

cas 1 :

$$\lambda_i(t) = \lambda_0 \left(1 + \frac{\tau_0}{\lambda_0} (1 - e^{-\beta_0 t}) \right) \quad (3-66)$$

où λ_0, β_0 et τ_0 sont connus. $\lambda_0, \beta_0 > 0$ et $-\lambda_0 \leq \tau_0 < \infty$; $\tau_0 \neq 0$

La statistique du test optimal fournie par le théorème fondamental de Neyman-Pearson est :

$$A^* = \prod_{i=1}^{N(s)} \left\{ 1 + \frac{\tau_0}{\lambda_0} (1 - e^{-\beta_0 T(i)}) \right\} \exp(\lambda_0 s - f(s)) \quad (3-67)$$

où :

$$f(s) = (\lambda_0 + \tau_0)s + \beta_0^{-1} \tau_0 e^{-\beta_0 s} - \beta_0^{-1} \tau_0$$

Cette statistique est équivalente à :

$$S^* = \sum_{i=1}^{N(s)} \text{Log} \left[1 + \frac{\tau_0}{\lambda_0} (1 - e^{-\beta_0 T(i)}) \right] \quad (3-68)$$

Une analyse conditionnelle est basée sur la loi de S^* , étant donné la réalisation de $N(s)$, on trouve :

$$\begin{aligned} E[S^* | N(s) = n; \lambda(t) \equiv \lambda_0] &= \frac{n}{s} \int_0^s \text{Log} \left(1 + \frac{\tau_0}{\lambda_0} (1 - e^{-\beta_0 t}) \right) dt \\ &= -n \sum_{m=1}^{\infty} \left[\left(\sum_{k=1}^m \left(\frac{\tau_0}{\lambda_0} \right)^k (k-1)! S_m^{(k)} \right) \frac{(-\beta_0 s)^m}{(m+1)!} \right] \end{aligned} \quad (3-69)$$

Cas 2 :

$$\lambda_i(t) = \lambda_0 \text{Log}(t + e)$$

On remarque que pour l'analyse globale $\sum_{i=1}^{N(s)} [\text{Log}(\text{Log}(T(i)) + e)]$ est une statistique

exhaustive pour le problème. En effet, on a :

$$A^* = \prod_{i=1}^{N(s)} [\text{Log}(T(i) + e)] \exp(\lambda_0 s - \lambda_0 \int_0^s \text{Log}(t + e) dt) \quad (3-70)$$

3.5.3 Tests relatifs aux processus de renouvellement

Parmi les hypothèses du processus ponctuel de renouvellement figure celle de l'indépendance des intervalles entre les événements. Cette hypothèse peut être vérifiée par la fonction d'autocorrélation ou par d'autres tests basés sur la théorie spectrale.

3.5.3.1 Tests du coefficient d'autocorrélation entre les intervalles

Le coefficient d'autocorrélation d'ordre k , ρ_k , est défini par la formule suivante :

$$\rho_k = \frac{n}{n-k} \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3-71)$$

Les hypothèses du test sont :

H_0 : $\rho_k = 0$ (indépendance des intervalles)

H_1 : $\rho_k \neq 0$ (dépendance des intervalles)

La statistique du test est donnée par :

$$\hat{\rho}_k \sqrt{n-1} \quad (3-72)$$

où $\hat{\rho}_k$ est l'estimation de ρ_k .

Cette statistique suit une loi normale $N(0,1)$ à condition que le nombre d'observations soit assez grand.

La zone de rejet de l'hypothèse d'indépendance est définie par :

$$|\hat{\rho}_k \sqrt{n-1}| > Z_{\alpha/2}$$

$Z_{\alpha/2}$ est le quantile $\alpha/2$ de la loi normale centrée réduite $N(0,1)$.

L'application de ce test exige d'une part une grande taille d'échantillon, en d'autres termes, il faut que la série des intervalles suive une loi normale. Cependant, la distribution des

intervalles n'est pas connue et elle dépend des temps qui séparent les événements. D'autre part, on suppose que la corrélation entre les différents coefficients d'autocorrélation est nulle :

$Corr(\hat{\rho}_j, \hat{\rho}_k) = 0$ pour $j \neq k$, ce qui n'est pas toujours le cas.

Cox et Lewis,(1966) ont présenté quelques statistiques qui sont supposées plus robustes à ces deux contraintes. Il s'agit en particulier de la statistique suivante :

$$r_j = \sum_{i=1}^{n-j} x_{i+j} x_i \quad (3-73)$$

Une forme standardisée de cette statistique est donnée par:

$\frac{r_j - E(r_j)}{\sqrt{\text{var}(r_j)}}$ et suit une loi normale $N(0,1)$ pour des tailles $n > 10$.

Cunnane (1979) a considéré que le calcul du coefficient de corrélation représente un outil non significatif pour la mesure de l'indépendance des intervalles de temps dans la série d'événements. Ceci est dû aux durées non constantes séparant chaque deux événements successifs. Cependant, il a examiné cette détection par une analyse graphique des données.

3.5.3.2 Tests du coefficient d'autocorrélation des rangs

La procédure de ce test reste identique à celle vue en section (3-5-3-1). Il suffit de remplacer les valeurs de x_i par leur rang.

Le rang de $x_k = 1 +$ (nombres vérifiant $x_i < x_k, i \neq k$).

A partir de la statistique définie par l'équation (3-73), on obtient :

$$E(R_j) = \frac{1}{12} (n-1)(n+1)(3n+2) \quad (3-74)$$

$$\text{var}(R_j) = \frac{5n^6 + 16n^5 - 14n^4 - 80n^3 - 35n^2 + 64n + 44}{720(n-1)} \quad (3-75)$$

La loi asymptotique de la statistique R_j est une loi normale de moyenne et de variance définies respectivement par les équations (3-74) et (3-75), (Wald et Wolfowitz, 1943).

3.5.3.3 Tests basés sur l'analyse spectrale

On définit l'estimateur du spectre des intervalles par la fonction $I_n(w_p)$:

$$I_n(w_p) = \frac{1}{2\Pi n} \sum_{s=1}^n \sum_{u=1}^n X_s X_u e^{i(s-u)w_p} \quad (3-76)$$

Dans le cas d'un processus de renouvellement, l'espérance de la quantité définie par l'équation (3-76) est égale à (Cox et Lewis, 1966):

$$E(I_n(w_p)) = \frac{\sigma^2}{2\Pi} = \sigma^2 f(w)$$

Cette quantité représente la puissance du spectre du processus de renouvellement.

La forme générale de la fonction $f(w)$ est donnée par l'équation suivante :

$$f(w) = \frac{1}{2\Pi} \left\{ 1 + 2 \sum_{k=1}^{\infty} \rho_k \cos(kw) \right\}, \quad 0 \leq w \leq \Pi \quad (3-77)$$

Les hypothèses du test sont formulées de la façon suivante :

H₀ : L'estimateur du spectre $I_n(w_p)$ est indépendamment et asymptotiquement distribué selon une loi exponentielle de moyenne $\sigma^2 f(w)$

H₁ : Existence d'une tendance (le spectre n'est pas constant).

Le rapport $\frac{I_n(w_p)}{f(w_p)}$ peut être considéré comme une estimation de la variance à partir d'une

population normale. Alors, on peut appliquer le test d'homogénéité des variances de Bartlett (Bartlett, 1937). On subdivise les l valeurs de $I(w_p)$ en k classes de taille v ($l=kv$) et on considère la somme suivante :

$$s_i^2 = \sum_{p=(i-1)v+1}^{iv} \frac{I_n(w_p)}{f(w_p)}, \quad (i = 1, \dots, k)$$

On rappelle que dans notre cas $f(w_p)=1/2\Pi, \forall p$. Alors la statistique du test est donnée par :

$$H = \frac{2l \operatorname{Log} \left(\frac{\sum_{i=1}^k s_i^2}{2l} \right) - \sum_{i=1}^k 2\nu \operatorname{Log} \left(\frac{s_i^2}{2\nu} \right)}{(6\nu - 2) / (6\nu - 3)} \quad (3-78)$$

Si $2\nu > 5$ la statistique H suit approximativement une loi du Chi-deux à $(k-1)$ degrés de liberté et l'hypothèse nulle est rejetée si H est grand (Cox et Lewis, 1966).

Le deuxième test d'indépendance des intervalles est basé sur l'estimation du spectre et sur les tests d'ajustement indépendants de la distribution. Soit la quantité:

$$U_{(i)} = \frac{\sum_{p=1}^i \frac{I(w_p)}{f(w_p)}}{\sum_{p=1}^l \frac{I(w_p)}{f(w_p)}} \quad (3-79)$$

Si $U_{(i)}$ suit une loi uniforme alors l'hypothèse H_0 est vérifiée (indépendance des intervalles). La condition d'uniformité des variables $U_{(i)}$ est réalisée par le biais des tests d'ajustement vus en section 3-5-1. Cox et Lewis, (1966) ont montré que la statistique de Cramer-von Mises (équation, 3-38) est plus appropriée pour la vérification de cette condition que celle de Kolmogorov-Smirnov.

3.6 Tests entre processus ponctuels et séries chronologiques

L'analyse d'un processus ponctuel par le biais de la fonction d'intensité des séries d'événements, $\lambda(t)$, peut être liée aux valeurs d'une fonction continue, $Z(t)$, représentant une série chronologique temporelle. A titre d'exemple, on peut tenir compte des données de précipitations journalières ou mensuelles (modélisées par une série chronologique) pour l'explication ou l'analyse de la variabilité de la fonction d'intensité de la série des débits dépassant un seuil (modélisés par un processus ponctuel). Ainsi, quelques tests ont été

élaborés dans le sens de la mise en évidence de cette liaison. D'après Wolfe (1980), Brown a développé une théorie pour les processus de Poisson non-homogènes et les séries chronologiques continues multivariées.

Wolfe (1980) a présenté, dans le cadre d'une étude motivée par des besoins médicaux, un test de l'association d'un processus ponctuel et les séries chronologiques continues. Le test est exact pour les processus de Poisson et pour une classe de processus ponctuels vérifiant l'hypothèse d'homogénéité.

3.6.1 Test circulaire d'association (Wolfe, 1980)

La procédure du test circulaire entraîne le calcul de la covariance entre le processus ponctuel, $\lambda(t)$, et la série chronologique, $Z(t)$, pour tous les délais (lag) considérés et compare ces covariances avec celle au délai zéro.

Si la covariance au délai zéro est plus grande que la proportion $(1-\alpha)$ des covariances, l'hypothèse nulle de l'association entre le processus ponctuel et les séries chronologiques continues est rejetée à un niveau de signification α . Les hypothèses du test sont :

H_0 : Il existe une association entre un processus ponctuel et une série chronologique.

H_1 : Il n'existe aucune association entre un processus ponctuel et une série chronologique.

Soient $T_i, i=1,2, \dots, n$, les observations des temps d'occurrence de n événements durant un intervalle $(0,T]$ et $Z(t)$ une fonction mesurable du temps sur cet intervalle qui exprime les valeurs d'un facteur suspect d'être relié aux occurrences des événements. On définit la fonction cyclique de $Z(t)$ sur $(0,2T]$ par :

$$Z^*(t) = \begin{cases} Z(t) & \text{si } 0 < t \leq T \\ Z(t-T) & \text{si } T < t \leq 2T \end{cases} \quad (3-80)$$

Pour des valeurs de temps discrètes, où $Z(t)$ est constante pour $s \in (t-1, t]$, le test circulaire est particulièrement simple et la valeur critique du test est facile à calculer. En effet, si on considère l'extension circulaire de $Z(t)$ définie par l'équation (3-80), la statistique du test est :

$$S(s) = \sum_{t=1}^T N(t)Z^*(t+s) \quad s = 0, 1, \dots, T-1 \quad (3-81)$$

$N(t)$ est le nombre d'événements observés durant la période $(t-1, t]$. L'hypothèse nulle, qui favorise l'existence d'une relation au délai 0, est rejetée si la quantité $\sum_{t=1}^T Z(t)N(t)$ est plus grande que $(1-\alpha)T\%$ des T valeurs de $S(s)$, $s=0, 1, \dots, T-1$.

Le test circulaire d'association est valide pour tout les processus ponctuels possédant les propriétés suivantes :

soient T_1, \dots, T_N les temps d'occurrence de N événements durant une période $(0, T]$. Conditionnellement à N , la distribution jointe de T_1, \dots, T_N est identique à la distribution de $T_1+s(\text{mod } T), \dots, T_N+s(\text{mod } T)$ pour tout s . La notation $c(\text{mod } T)$ signifie le nombre $c+kT$ où k est un entier satisfaisant $0 < c+kT \leq T$. Cette condition résume tout simplement l'homogénéité des données.

3.6.2 Processus de Poisson et série chronologique (Wolfe, 1980)

La relation entre un processus ponctuel de Poisson et une série chronologique peut être illustrée à travers la fonction d'intensité ou bien par la fonction densité des événements :

$$\lambda(t) = \lim_{\Delta t \rightarrow 0^+} \frac{\text{Pr ob}\{\text{événement dans } (t, t + \Delta t]\}}{\Delta t} = \lambda \exp(bZ(t)) \quad (3-82)$$

D'après Wolfe (1980), Brown a montré qu'un test optimal pour l'hypothèse suivante :

$$H_0: b \leq 0$$

est conditionnel au nombre d'événements n et basé sur la statistique suivante :

$$S_0 = \sum_{i=1}^n Z(t_i) \text{ où } n \text{ et } t_i \text{ sont, respectivement, le nombre d'observations et les temps}$$

d'occurrence des événements dans l'intervalle $(0, T]$. Pour des grandes valeurs de n , on rejette l'hypothèse H_0 à un niveau de signification α quand :

$$(S_0 - n\bar{Z}) / (n \text{ var}(Z))^{1/2} > Z_{1-\alpha} \quad (3-83)$$

\bar{Z} est la moyenne des $Z(s)$ sur $(0, T]$ et $Z_{1-\alpha}$ est le quantile $(1-\alpha)$ de la loi normale de moyenne zéro et de variance unité.

Wolfe (1980) a examiné la validité de l'approximation, exprimée par (3-83), en fonction du nombre d'événements par une série de simulations. Il a proposé deux formes de la fonction chronologique $Z(t)$ pour une période $T=50$:

$$\begin{aligned} Z(t) &= t \\ Z(t) &= \sin(2\pi t / T) \end{aligned} \quad t=1, 2, \dots, 50$$

Les résultats de la simulation ont montré une bonne performance du test en utilisant la forme linéaire de $Z(t)$, ce qui n'était pas le cas pour la forme sinusoïdale.

Finalement, Wolfe (1980) a appliqué ce test pour détecter l'existence d'une association entre le nombre d'échecs d'une transplantation du coeur (processus ponctuel) et le niveau de régularité des cellules lymphocyte*-T (série chronologique). Le résultat du test a confirmé l'existence d'une relation significative entre les deux variables.

* : Les lymphocytes T sont les globules blancs portant chacun des récepteurs protéiques capables de reconnaître un antigène, c'est-à-dire une substance étrangère à l'organisme (Housset et al, 1997)

3.6.3 Détection de tendance en utilisant une série chronologique (Brillinger, 1994)

Supposons que $N(t)$ est le nombre de points dans l'intervalle $(0, t]$ d'un processus ponctuel, et soit $dN(t)$ égale à l'unité si un point est enregistré dans l'intervalle $(t, t+dt]$ et zéro dans le cas contraire. Sous l'hypothèse de stationnarité, on définit la fonction d'auto-intensité $h_{NN}(u)$ au délai u par :

$$h_{NN}(u)du = \text{Pr ob}\{dN(t+u) = 1 / \text{événement } t\} \quad u \neq 0$$

Ce paramètre est analogue à la fonction d'auto-covariance. Si les points $\tau_1 < \tau_2 \dots < \tau_{N(t)} < T$ du processus N sont disponibles, alors $h_{NN}(u)$ peut être estimée par :

$$h_{NN}^T(u) = \# \left\{ \left| \tau_k - \tau_j - u \right| < b \right\} / 2bN(T) \quad \text{où } b \text{ est une "fenêtre" assez faible.}$$

Cette estimation représente essentiellement l'histogramme des valeurs $(\tau_k - \tau_j)$.

Brillinger (1994) a utilisé l'approche du maximum de vraisemblance, qui est développée à partir de la forme conditionnelle de la fonction d'intensité :

$$\text{Prob}\{dN(t)=1/H_t\}.$$

La théorie des séries chronologiques peut être utilisée pour l'analyse des processus ponctuels. On définit la série temporelle indicatrice de N_t :

$$\begin{cases} 1 & \text{un événement} \in (t, t+b] \\ 0 & \text{aucun événement dans } (t, t+b] \end{cases} ; t = 0, \pm b, \pm 2b, \dots \quad (3-84)$$

Ainsi, on construit la forme de probabilité suivante :

$$\Pi_t = \text{Prob}\{N_t = 1 / H_{t-1}\} \quad (3-85)$$

Brillinger (1994) a supposé la forme suivante de la cette probabilité :

$$\Pi_t = \phi\left(S(t) + \sum_{u=1}^U a(u)N_{t-u}\right) \quad (3-86)$$

où :

ϕ est la fonction de distribution de la loi normale. Le logarithme de la fonction de vraisemblance est donné par :

$$\sum [N_t \log \Pi_t + (1 - N_t) \log(1 - \Pi_t)] \quad (3-87)$$

Brillinger (1994) a considéré deux formes de la fonction de tendance $S(t)$. La forme paramétrique $S(t) = \alpha + \beta t$ qui a été traitée dans la section (3-4-1). La deuxième forme de tendance est non-paramétrique, c'est une fonction de lissage. Dans ce cas l'estimation de la fonction $S(t)$ peut être tirée de la fonction du maximum de vraisemblance pénalisée, donnée par (Brillinger 1994) :

$$\sum_t [N_t \log \Pi_t + (1 - N_t) \log(1 - \Pi_t)] + \lambda \sum_t [S(t) - 2S(t-1) + S(t+2)]^2 \quad (3-88)$$

où :

$$\Pi_t = \text{Prob}\{N_t = 1 / H_{t-1}\}$$

3.7 Test de sélection d'un processus ponctuel

Smith et Karr (1985) ont développé une procédure de sélection entre plusieurs processus ponctuels représentant des données de précipitations. L'approche est basée sur les estimateurs du maximum de vraisemblance des paramètres de chaque modèle. Smith et Karr (1985) ont considéré trois processus ponctuels pour modéliser les précipitations. Il s'agit des processus de Cox, de Neyman-Scott et du processus de renouvellement. L'inférence statistique est basée sur la fonction d'intensité stochastique :

$$\lambda(t) = \lim_{s \rightarrow 0^+} \frac{E[N(t+s) - N(t) | N(u), u < t]}{s}$$

La procédure de sélection n'est pas une procédure standard, cependant, elle peut être appropriée pour d'autres modèles qui ne visent pas la modélisation des précipitations. Elle permet de déterminer le modèle qui représente au mieux la série de données. Toutefois, cette méthode de sélection n'est pas toujours optimale.

La procédure est basée sur le rapport de vraisemblance :

$$A(n, s_1, \dots, s_n) = - \int_0^t (\lambda_0(u) - \lambda_1(u)) du + \int_0^t \log \left(\frac{\lambda_0(u)}{\lambda_1(u)} \right) dN(u) \quad (3-89)$$

où :

λ_0, λ_1 sont les intensités stochastiques, respectivement, sous les hypothèses H_0 et H_1 .

La procédure du choix entre le modèle exprimé par l'hypothèse nulle et celui exprimé par l'hypothèse alternative H_1 est constituée de trois étapes :

- étape 1 : calculer les estimateurs de maximum de vraisemblance pour les modèles relatifs aux hypothèses H_0 et H_1 ;
- étape 2 : construire le rapport de vraisemblance en utilisant les estimateur de l'étape 1 ;
- étape 3 : le modèle de H_0 est adopté si le logarithme du rapport de vraisemblance est supérieur à zéro, sinon le modèle de H_1 est choisi.

Smith et Karr (1985) ont indiqué que cette procédure est applicable que dans le cas où il est possible de calculer la fonction d'intensité stochastique.

Enfin, nous avons représenté sur la figure 5 un récapitulatif des différents tests et analyses pour une série d'un seul type d'événements.

STATIONNARITE

Analyse de tendance

- Analyse graphique :
 - Nombre d'événements en fonction du temps d'occurrence
 - Fonction de la variance temporelle
- Analyse de régression des intervalles (Processus de Poisson)
- Analyse de régression des nombres d'événements (Processus de Poisson)
- Méthodes basées sur la fonction du maximum de vraisemblance (Processus de Poisson)
 - A partir du taux d'occurrence
 - A partir de l'analyse des intervalles
 - Détection de rupture pour des observations poissonniennes
 - Test du rapport de vraisemblance
 - Approche bayésienne
- Tests d'association entre processus ponctuels et séries chronologiques
 - Test circulaire d'association
 - Processus de Poisson et série chronologique
 - Détection de tendance en utilisant une série chronologique

Homogénéité : Analyse graphique

- La moyenne du taux d'occurrence
- Fonction d'intensité

Analyse de périodicité

- Le spectre du nombre d'événements

TESTS SUR LES INTERVALLES

Indépendance des intervalles

- tests relatifs au processus de renouvellement
 - Tests du coefficient d'autocorrélation entre les intervalles
 - Tests du coefficient d'autocorrélation des rangs
 - Tests basés sur l'analyse spectrale

Adéquation de la distribution

- Adéquation des distributions des intervalles :
 - Tests d'ajustements indépendants de la distribution
 - Test d'uniformité conditionnelle (intervalles)
- Nombre d'événements (Poisson)
 - Test de dispersion

TEST DE SÉLECTION D'UN PROCESSUS PONCTUEL PAR ÉLIMINATION

- Procédure de sélection entre les processus de Cox, de Neyman-Scott et de renouvellement
- Tests entre le processus de Poisson et le processus de Poisson doublement stochastique
- Tests entre processus de Poisson et processus de renouvellement

Figure 5 : Récapitulatif des différentes analyses pour une série d'un seul type d'événements

3.8 Tests d'une série de deux ou de plusieurs types d'événements

Dans cette section, nous allons nous intéresser à un cas qui est fréquemment rencontré dans la pratique. Il s'agit de l'analyse d'un processus ponctuel global (output pooled) obtenu en superposant p processus ponctuels. Dans la littérature, il n'existe pas de tests statistiques ou de procédures d'analyse de ce type de données. Toutefois, nous allons examiner la possibilité d'appliquer les tests présentés dans ce chapitre pour un processus ponctuel global. Une étude des propriétés de ce type de modèle s'avère nécessaire pour tester la validité de ses hypothèses de base.

3.8.1 Superposition de p processus ponctuels

Afin de mieux visualiser ce type de processus, nous allons représenter sur la figure 6 un schéma de superposition de deux processus ponctuels où :

t_k^i : représente le temps d'occurrence du $k^{\text{ième}}$ événement du processus ponctuel i , $i=1, 2$.

X_k^i : représente la durée séparant deux événements successifs enregistrés aux temps t_k^i et t_{k+1}^i pour le processus ponctuel i , $i=1, 2$.

Z_k^i : représente la durée séparant deux événements du même type i , $i=1, 2$ dans le processus ponctuel global.

On rappelle que l'hypothèse de non-multiplicité b) définie dans la section (2-1-2-1) est respectée dans le schéma de la figure-6 et que le principe de superposition reste le même pour $p > 2$ processus ponctuels.

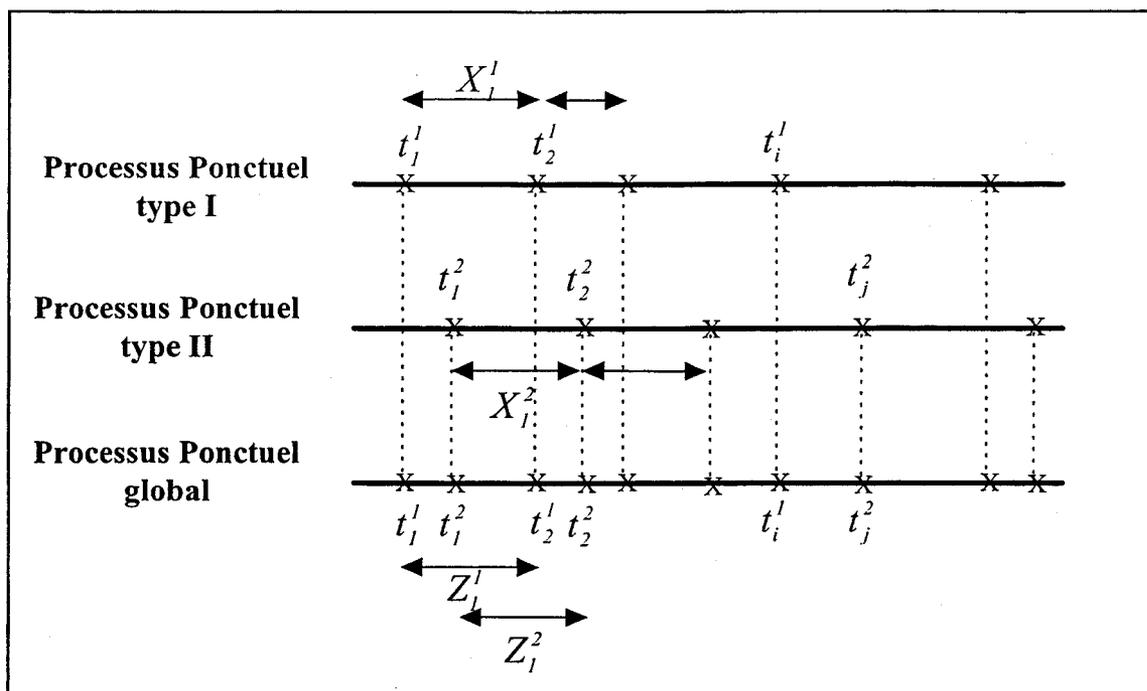


Figure 6 : Superposition de deux processus ponctuels.

Une première constatation consiste à voir que la variable Z_k^i est la somme de deux durées de temps provenant de deux processus différents. D'où la difficulté de la modélisation de la série des intervalles par la théorie des processus ponctuels.

Un cas particulier de la superposition des processus ponctuels est le processus alterné. Dans ce cas le processus global est constitué de deux types d'événements, type I et type II, qui sont alternés dans le temps. Abi-Zeid (1997) a modélisé les séries des débits journaliers par ce type de processus (figure-7). En fixant un seuil s_0 , elle a obtenu deux séries d'événements. Une série des surplus construite à partir des débits supérieurs ou égaux au seuil s_0 et une série des étiages obtenue en considérant les débits inférieurs au seuil s_0 . Ainsi, le processus alterné est composé, d'une part, d'un processus de rupture des surplus, défini par les temps d'occurrence $t_1^1, t_2^1, t_3^1, \dots$ etc, et par les intervalles de temps entre ces événements $Z_1^1, Z_2^1, Z_3^1, \dots$ etc, et d'autre part, d'un processus de rupture des étiages défini

par les temps d'occurrence $t_1^2, t_2^2, t_3^2, \dots$, etc, et par les intervalles de temps entre ces événements $Z_1^2, Z_2^2, Z_3^2, \dots$ etc. L'analyse du processus ponctuel global (alterné dans ce cas) a montré que l'intervalle de temps séparant deux événements successifs de surplus est la somme de deux variables aléatoires correspondantes aux durées d'étiage et de surplus. Pour la validation des hypothèses du processus ponctuel alterné, Abi-Zeid (1997) a appliqué quelques tests statistiques valables aux durées seulement (tableau-1). Chebchoub et al (1999) les ont appliqués à des données du réseau hydrographique tunisien, cependant, leurs résultats ne sont pas encore publiés. Les tests identifiés dans la figure-5 et qui sont valables pour les intervalles d'un processus ponctuel à un seul type d'événements pourront être appliqués aux durées des deux processus constituant le processus alterné.

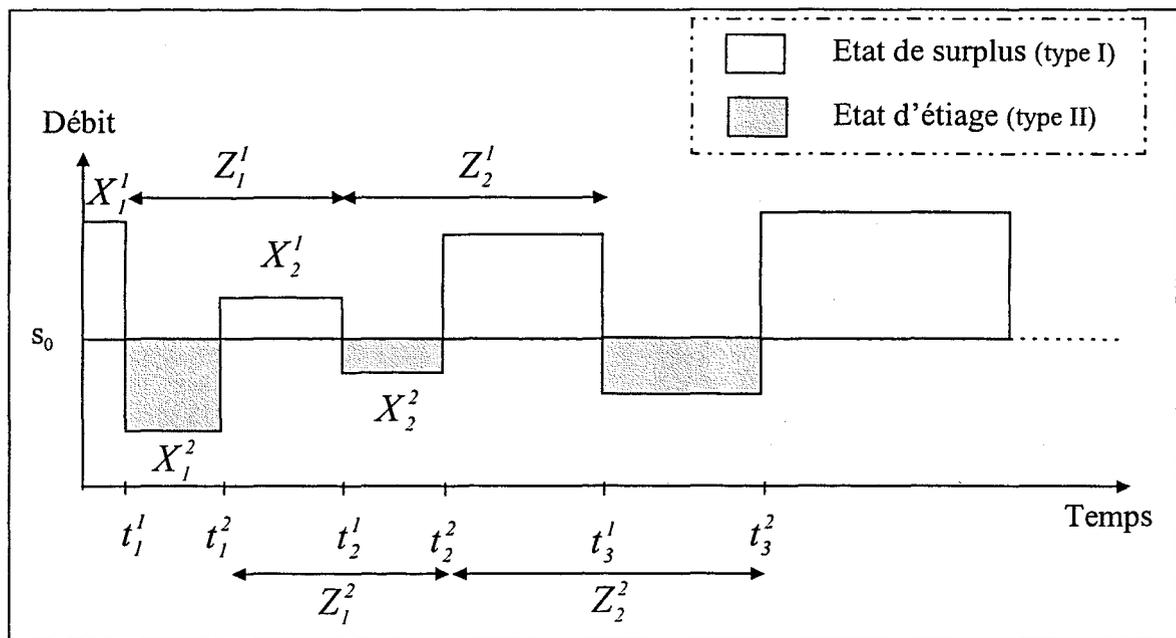


Figure 7 : Schématisation du processus alterné

Ainsi, nous avons jugé utile de présenter dans cette section quelques propriétés des processus ponctuels superposés afin d'examiner la possibilité de définir les distributions des intervalles et du nombre d'événements dans un processus ponctuel global.

Tableau 1 : Identification des tests statistiques appliqués au processus alterné (Abi-Zeid, 1997)

Tests statistiques	Processus Ponctuel
Fonction de survie logarithmique (analyse graphique)	PPNH
Le test du périodogramme appliqué aux durées homogénéisées de chaque processus : indépendance des durées	PPNH, PRNH*, PNHC**
Adéquation de la loi exponentielle aux durées homogénéisées : statistique de Kolmogorov-Smirnov et Anderson	PPNH, PNHC
Adéquation de la loi Gamma aux durées homogénéisées : statistiques de Kolmogorov-Smirnov et Anderson	PRNH
Adéquation de la loi des durées non homogénéisées (North, 1981)	PPNH

* : Processus de Renouvellement non-homogène.

** : Processus de Poisson non-homogène conditionnel, (Abi-Zeid, 1997).

3.8.2 Propriétés et analyse des processus superposés

Cox et Smith (1954) ont réalisé le premier travail de recherche sur la superposition de p processus de renouvellement homogènes. Ils ont montré que la fonction densité de probabilité des intervalles du processus global est donnée par :

$$f_p(x) = -\frac{d}{dx} \left[R(x) \left\{ m \int_x^\infty R(u) du \right\}^{p-1} \right] \quad (3-90)$$

$$\text{où : } m = m(t) = \frac{dM(t)}{dt} = \frac{dE(N_t)}{dt}$$

$R(x)$ représente la fonction de survie marginale des p processus ponctuels. Quand $p \rightarrow \infty$ la fonction définie par l'équation (3-90) tend rapidement à une loi exponentielle de moyenne

$\frac{1}{pm}$, et le nombre d'événements dans un intervalle de taille $\frac{\nu}{pm}$ suit une distribution de

Poisson de moyenne ν , c'est-à-dire que le processus global est un processus de Poisson homogène.

Cox et Lewis (1966) ont présenté quelques résultats probabilistiques du processus ponctuel global. Les hypothèses de stationnarité et d'indépendance sont nécessaires pour le développement des résultats.

Soient p processus ponctuels ayant chacun un taux d'occurrence $m_i(\tau)$. La densité des covariances est :

$$\gamma_+(\tau) = \lim_{\Delta t \rightarrow 0} \frac{\text{cov}(\Delta N_t, \Delta N_{t+\tau})}{(\Delta t)^2} = m(m_f(\tau) - m) \quad (3-91)$$

Sous l'hypothèse d'indépendance des p processus ponctuels, le taux d'occurrence du processus global est égal à :

$$m^{(p)} = pm \quad (3-92)$$

La notation (p) en exposant indique le processus global.

On a également les égalités suivantes :

$$V^{(p)} = pV(t) \quad (3-93)$$

$$\gamma_+^{(p)}(\tau) = p\gamma_+(\tau) \quad (3-94)$$

$$g^{(p)}(w) = pg(w) \quad (3-95)$$

Cox et Lewis (1966) ont considéré l'approximation suivante pour la fonction de variance $V(t)$:

$$V(t) \sim tm \Rightarrow V^{(p)}(t) \sim tpm = tm^{(p)} \quad (3-96)$$

Cette approximation est valable pour des faibles valeurs de t comparées aux valeurs de

$$\frac{1}{m} = \frac{p}{m^{(p)}}.$$

Dans le cas où les p processus ponctuels sont des processus de Poisson, le processus global est également un processus de Poisson et la relation (3-96) est valable pour toute valeur de t (Cox et Lewis, 1966). L'équation (3-96) signifie que pour des grandes valeurs du paramètre p , le processus global tend vers un processus de Poisson. Dans le cas où les p processus ne sont pas représentés par un processus de Poisson, la validité de l'approximation donnée par l'équation (3-96) doit être vérifiée.

Thompson (1988) a présenté un résultat fondamental sur le comportement asymptotique du processus ponctuel global. Il s'agit du théorème de Grigelionis :

On suppose que :

$$\lim_{p \rightarrow \infty} \min_{i=1, \dots, p} \text{Pr ob}(N_{pi}(t) = 0) = 1 \quad (3-97)$$

$$\lim_{p \rightarrow \infty} \sum_{i=1}^p \text{Pr ob}(N_{pi}(t) \geq 2) = 0 \quad (3-98)$$

où N_{pi} est le nombre d'événements du processus ponctuel i enregistré jusqu'au temps t .

Pour un système vérifiant les conditions décrites par les équations (3-97) et (3-98), une condition nécessaire et suffisante pour que le nombre d'événements du processus global suive une distribution de Poisson est :

$$\lim_{p \rightarrow \infty} \sum_{i=1}^p \text{Pr ob}(N_{pi}(s, t) = 1) = M(t) - M(s) \quad (3-99)$$

où :

s et t sont deux valeurs du temps telle que $s < t$ et $M(t) = E(N(t))$.

Sous les conditions de ce théorème, le processus global est approximativement représenté par un processus de Poisson, même si les p processus ponctuels ne le sont pas. Ainsi, le nombre d'événements du processus global est défini par :

$$\text{Pr ob}(N(t) \geq 1) \cong 1 - \exp\left[-\sum_{i=1}^p \text{Pr ob}(N_{pi}(t) = 1)\right] \quad (3-100)$$

Les intervalles du processus global suivent approximativement une distribution exponentielle :

$$R(w) \cong \exp[-(M(t+w) - M(t))] \quad (3-101)$$

$R(w)$ est la fonction de survie des temps d'arrivés w . C'est une autre conséquence du théorème de Grigelionis.

Chandrakohan et Liang (1985) ont précisé que la superposition de deux processus de renouvellement indépendants l'un de l'autre est un processus de renouvellement si et

seulement si les trois processus ont les propriétés d'un processus de Poisson. Ils ont examiné le cas où les deux processus ponctuels sont dépendants et ils ont défini les conditions nécessaires pour obtenir un processus global de renouvellement. Soit le théorème suivant :

On considère deux processus ponctuels de nombre d'événements N^0 et N^1 et soit N le nombre d'événements total. Le processus global est un processus de renouvellement si et seulement si les distributions jointes de N^0 et N^1 vérifient la condition suivante :

$$\text{Prob}(N_t^0 = k_0, N_t^1 = k_1) = f(k_0 + k_1, k_0) [F^{*(k_0+k_1)}(t) - F^{*(k_0+k_1+1)}(t)] \quad (3-102)$$

où :

$$f(n, k) = \text{Prob}\left(\sum_{j=1}^n I_{\{0\}}(X_j) = k\right) = \text{Probabilité que la variable } X \text{ reçoit la valeur zéro } k$$

fois.

$$X_n = \begin{cases} 0 & \text{n}^{\text{ième}} \text{ événement du processus } N^0 \\ 1 & \text{sinon} \end{cases} \quad (3-103)$$

$$\text{Prob}(N_t = n) = F^{*n}(t) - F^{*(n+1)}(t)$$

Cinlar et Agnew (1968) ont présenté une série de théorèmes concernant la superposition de deux processus ponctuels. Ils ont défini un processus ponctuel par deux variables (T, X_n) où X_n est la variable définie par l'équation (3-103). Nous allons citer quelques résultats démontrés par Cinlar et Agnew (1968) :

Théorème 1

Soit (N_1, N_2) un système de processus ponctuels muni du processus indicateur X et soit le processus global N . On suppose que X est un processus de Bernoulli indépendant de N . Alors, le système est indépendant si et seulement si :

$$P\{N(t) = k\} = \exp\{-\lambda(t)\} \frac{[\lambda(t)]^k}{k!} \quad (t \geq 0) ; (k=0, 1, \dots)$$

$\lambda(t)$ est une fonction croissante sur l'intervalle $[0, \infty)$ et $\lambda(0)=0$.

On rappelle qu'un processus de Bernoulli est défini par :

$$\text{Prob}(X_1=b) = 1 - \text{Prob}(X_1=a)$$

Théorème 2

Soit (N_1, N_2) un système de deux processus ponctuels de Poisson. On suppose que X est représenté par un processus de Bernoulli indépendant de N . Alors, le processus global est un processus de Poisson si et seulement si le système est indépendant.

Théorème 3

Soit (N_1, N_2) un système de deux processus ponctuels de Poisson. On suppose que X est une chaîne de Markov stationnaire indépendante du nombre total d'événements du processus global et que les taux d'occurrence de N_1 et N_2 ne sont pas égaux. Alors, le processus global est un processus de renouvellement si et seulement si le système est indépendant.

Théorème 4

Soient (N_1, N_2) un système de deux processus ponctuels de Poisson indépendants et X le processus indicateur. Si N_1 est un processus de Poisson et X est indépendant de N , alors N est un processus de Poisson si et seulement si N_2 est un processus de Poisson.

D'autres travaux de recherche ont été réalisés sur la superposition des processus ponctuels. On cite ceux de Ambartzumian, (1965), Cinlar, (1968), Lawrance, (1973), Ripley (1976), Ito, (1977), et Lam et al (1991). Cependant, on ne trouve pas dans ces travaux des tests statistiques spécifiques à une série de deux ou de plusieurs types d'événements. Les auteurs supposent généralement les deux hypothèses de stationnarité et d'indépendance des p processus ponctuels, qui sont représentés par un processus de Poisson ou un processus de renouvellement.

L'application des tests statistiques présentés dans ce chapitre à une série de plusieurs types d'événements est possible dans le cas asymptotique, où $p \rightarrow \infty$. Ils sont également applicables sous les conditions des hypothèses présentées dans cette section. Dans le cas contraire, on ne retient que les tests ou les procédures sur les durées ou le nombre d'événements. En ce qui concerne les intervalles de temps entre deux événements successifs de même type, il faut définir la forme de la distribution de ces intervalles.

4. COMPARAISON ENTRE DEUX OU PLUSIEURS SÉRIES D'ÉVÉNEMENTS

4.1 Comparaison de deux processus ponctuels de Poisson

4.1.1 Méthode basée sur la distribution de Poisson

On suppose que l'on dispose de deux séries d'observations indépendantes tirées de deux processus de Poisson, de taux d'occurrence λ_1 et λ_2 . On désire comparer ces deux quantités. Soient les deux périodes d'observations des événements t'_0 et t''_0 , et soient n_1 et n_2 les nombres d'événements enregistrés respectivement durant ces deux périodes. N_1 et N_2 sont deux variables aléatoires indépendantes et suivent chacune une distribution de Poisson de moyenne $\mu_1 = \lambda_1 t'_0$ et $\mu_2 = \lambda_2 t''_0$. Ainsi, on peut écrire :

$$\text{Prob}(N_1 = n_1, N_2 = n_2) = \frac{\exp(-\mu_1) \mu_1^{n_1}}{n_1!} \frac{\exp(-\mu_2) \mu_2^{n_2}}{n_2!} \quad (4-1)$$

La probabilité conditionnelle de n_2 est exprimé par ;

$$\begin{aligned} \text{Prob}(N_2 = n_2 | N_1 + N_2 = n_1 + n_2) &= \frac{\text{Prob}(N_1 = n_1, N_2 = n_2)}{\text{Prob}(N_1 + N_2 = n_1 + n_2)} \\ &= \frac{\frac{\mu_1^{n_1} \mu_2^{n_2}}{n_1! n_2!} \exp[-(\mu_1 + \mu_2)]}{\frac{(\mu_1 + \mu_2)^{n_1 + n_2}}{(n_1 + n_2)!} \exp[-(\mu_1 + \mu_2)]} = C_{n_1 + n_2}^{n_1} \theta^{n_2} (1 - \theta)^{n_1} \end{aligned} \quad (4-2)$$

où :

$$\theta = \frac{\rho}{1 + \rho} ; \rho = \frac{\mu_2}{\mu_1}$$

L'hypothèse nulle à tester est :

$$\mathbf{H}_0 : \lambda_1 = \lambda_2$$

Sous l'hypothèse nulle, la distribution définie par l'équation (4-2) suit une loi binomiale de paramètre $\frac{t'_0}{t'_0 + t''_0}$ et en particulier de paramètre 1/2 si les périodes d'observations sont égales.

Le rapport $\rho = \frac{\mu_2}{\mu_1}$ est un outil de comparaison entre les deux processus de Poisson. Alors,

l'hypothèse nulle peut être écrite sous la forme suivante :

$$\mathbf{H}_0 : \rho = \rho_0 \text{ i.e. } \theta = \theta_0$$

En utilisant l'approximation normale, on construit la statistique suivante

$$Z = \frac{|n_2 - n\theta_0| - 1/2}{\sqrt{n\theta_0(1-\theta_0)}} \sim N(0,1) \quad (4-3)$$

On peut comparer les taux d'occurrence par le biais de la différence $d_t = \lambda_1 - \lambda_2$. Cette méthode exige une grande taille des échantillons, afin d'utiliser l'approximation normale.

L'estimateur non biaisé de d_t est donné par :

$$\hat{d}_t = \hat{\lambda}_1 - \hat{\lambda}_2 = \frac{n_1}{t'_0} - \frac{n_2}{t''_0}$$

et sa variance non biaisée est donnée par :

$$\text{var } \hat{d}_t = \frac{n_1}{t'^2_0} + \frac{n_2}{t''^2_0} \quad (4-4)$$

La statistique du test peut alors s'écrire :

$$Z_d = \frac{d_t - \hat{d}_t}{\sqrt{\text{var}(\hat{d}_t)}} \sim N(0,1) \quad (4-5)$$

Cox et Lewis (1966) ont indiqué que cette analyse est robuste à la taille d'échantillon, i.e. le résultat du test ne change pas d'une façon significative si la taille d'échantillon est petite.

Cependant, la première méthode, basée sur le rapport des taux d'occurrence, reste préférable car la variable ρ est adimensionnelle

4.1.2 Méthode basée sur la distribution exponentielle

Cette méthode est basée sur les temps enregistrés (intervalles) pour un certain nombre d'événements. Dans le cas précédent, on avait fixé les périodes d'observations et on a construit les distributions statistiques du nombre d'événements de chaque série. Nous allons exploiter le résultat suivant :

Si T est une variable aléatoire représentant le temps d'occurrence du $n_0^{\text{ième}}$ événement, alors la statistique $2\lambda T$ suit une distribution du Chi-deux à $2n_0$ degrés de liberté.

Ainsi, on peut construire le rapport R qui suit une distribution de Fisher :

$$R = \frac{(2\lambda_2 T_2)/(2n_0'')}{(2\lambda_1 T_1)/(2n_0')} = \frac{n_0' T_2 \lambda_2}{n_0'' T_1 \lambda_1} \sim F(2n_0', 2n_0'') \quad (4-6)$$

Si α est le niveau de signification du test :

$$\text{Prob}(f_{1-\alpha} < R < f_\alpha) = 1 - 2\alpha$$

On peut construire alors l'intervalle de confiance du rapport des deux taux d'occurrence :

$$f_{1-\alpha} \frac{t_1 n_0''}{t_2 n_0'} < \frac{\lambda_2}{\lambda_1} < f_\alpha \frac{t_1 n_0''}{t_2 n_0'} \quad (4-7)$$

Quand n_0' et n_0'' sont assez grand, une transformation logarithmique peut être appliquée :

$$Z_R = \text{Log} R = \text{Log} \left(\frac{\lambda_2}{\lambda_1} \right) + \text{Log} \left(\frac{n_0' T_2}{n_0'' T_1} \right) \quad (4-8)$$

Z_R est approximativement distribuée selon une loi normale de moyenne $\frac{1}{2} \left(\frac{1}{n_0'} - \frac{1}{n_0''} \right)$ et

d'écart-type $\left(\frac{1}{n_0'} + \frac{1}{n_0''} \right)^{1/2}$

4.2 Comparaison de $k > 2$ processus ponctuels de Poisson

On considère k processus de Poisson, $k > 2$, et on désire tester l'égalité de leur taux d'occurrence $\lambda_1, \dots, \lambda_k$. L'hypothèse alternative reste non spécifiée.

4.2.1 Test basé sur la distribution de Poisson

Supposons la réalisation des nombres d'événements, n_1, \dots, n_k pendant les périodes de temps $t_0^{(1)}, \dots, t_0^{(k)}$. Chaque variable aléatoire N_i suit une distribution marginale de Poisson, de moyenne $\mu_i = \lambda_i t_0^{(i)}$, $i = 1, \dots, k$. On définit la probabilité suivante :

$$\text{Prob}(N_i = n_i, i = 1, \dots, k) = \prod_{i=1}^k \frac{\{\lambda_i t_0^{(i)}\}^{n_i}}{n_i!} = \exp\left(-\sum_{i=1}^k \lambda_i t_0^{(i)}\right) \prod_{i=1}^k \frac{(\lambda_i t_0^{(i)})^{n_i}}{n_i!} \quad (4-9)$$

L'hypothèse nulle s'écrit de la façon suivante :

$$\mathbf{H}_0: \lambda_i = \lambda; i = 1, \dots, k$$

On construit un test basé sur l'indice de dispersion d :

$$d = \sum_{i=1}^k \frac{(n_i - t_0^{(i)} \hat{\lambda})^2}{t_0^{(i)} \hat{\lambda}} = \sum_{i=1}^k \frac{n_i^2}{t_0^{(i)} \hat{\lambda}} - kn \quad (4-10)$$

où

$$\hat{\lambda} = \frac{n}{t_0}$$

Dans le cas où les temps $t_0^{(1)}, \dots, t_0^{(k)}$ sont égaux, la statistique d devient ;

$$d = \frac{\sum_{i=1}^k (n_i - \bar{n})^2}{\bar{n}} \quad (4-11)$$

où : $\bar{n} = \frac{n}{k}$ est le nombre moyen par série d'événements.

Quand $\lambda \rightarrow \infty$, la distribution de d est une loi du Chi-deux à $(k-1)$ degrés de liberté (Cramèr, 1946)

4.2.2 Test du rapport de maximum de vraisemblance.

Afin de simplifier les calculs, on suppose que les périodes d'observations des événements sont égales. L'hypothèse à tester est :

$$\mathbf{H}_0 : \mu_1 = \dots = \mu_k$$

Le test est basé sur la statistique $H = -2 \text{Log } w$ où w représente le rapport entre le maximum de la fonction du maximum de vraisemblance sous l'hypothèse nulle et celui sous l'hypothèse alternative (Cox et Lewis, 1966). H s'écrit alors de la façon suivante :

$$H = -2 \text{Log } w = 2 \left(\sum_{i=1}^k n_i \text{Log } n_i - n \text{Log } \bar{n} \right) \quad (4-12)$$

Pour des périodes de temps non égales, H est exprimé par :

$$H = 2 \left\{ \sum_{i=1}^k n_i \text{Log} \left(\frac{n_i}{t_0^{(i)}} \right) - n \text{Log} \left(\frac{n}{t_0} \right) \right\} \quad (4-13)$$

Sous l'hypothèse nulle et pour des grandes valeurs de μ , H suit une distribution du Chi-deux à $(k-1)$ degrés de liberté.

Les statistiques H et d sont asymptotiquement équivalentes et ceci en considérant que

$$n_i = \bar{n}(1 - \varepsilon_i) \text{ où } \sum_{i=1}^k \varepsilon_i = 0$$

4.2.3 Test basé sur la distribution exponentielle

On observe, t_1, \dots, t_k , les temps d'occurrence de k processus de Poisson pour $n_0^{(1)}, \dots, n_0^{(k)}$ événements enregistrés. On rappelle que si t_i est la réalisation de la variable aléatoire T_i , alors la statistique $2\lambda_i T_i$ est distribuée selon une loi du Chi-deux à $2n_0^{(i)}$ degrés

de liberté, λ_i est le taux d'occurrence du $i^{\text{ème}}$ processus. Un test peut être basé sur la somme des carrés pondérés suivante :

$$\sum_{i=1}^k n_0^{(i)} \left\{ t_i - \frac{n_0^{(i)}}{\hat{\lambda}} \right\}^2 \quad (4-14)$$

où

$$\hat{\lambda} = \frac{\sum_{i=1}^k n_0^{(i)}}{\sum_{i=1}^k t_i}$$

On peut également remplacer t_i par sa transformation logarithmique. Cependant, Cox et Lewis (1966) ont favorisé le test du rapport du maximum de vraisemblance :

$$H = 2 \left\{ \sum_{i=1}^k n_0^{(i)} \text{Log} \left(\frac{n_0^{(i)}}{t_i} \right) - n_0 \text{Log} \left(\frac{n_0}{t} \right) \right\} \quad (4-15)$$

$$t = \sum_{i=1}^k t_i$$

Sous l'hypothèse H_0 , la statistique H suit une distribution du Chi-deux à $(k-1)$ degrés de liberté.

4.3 Comparaison des taux d'occurrence des processus non-poissonien

Cox et Lewis (1966) ont suggéré l'utilisation des tests non paramétriques, tel que le test de Wilcoxon. Ce choix est justifié par le fait que ces tests n'exigent pas une connaissance de la distribution des nombres d'événements ou des intervalles de temps, cependant, ces tests exigent l'indépendance des intervalles (Processus de renouvellement). Dans ce cadre, on présente deux tests non paramétriques pour la comparaison de la moyenne de deux échantillons ou plus.

4.3.1 Test des rangs de Wilcoxon-Mann-Whitney (Kendall et Stuart, 1979, Kanji, 1997)

Ce test permet de vérifier l'hypothèse que deux échantillons indépendants proviennent de deux populations de moyennes égales.

La procédure du test consiste à combiner deux échantillons et à les classer par ordre croissant en affectant à chaque observation son rang dans la série d'ordre. Soit R la somme des rangs de l'échantillon de petite taille et soit N le nombre d'observations pour l'échantillon combiné. La somme des rangs du deuxième échantillon est donnée par :

$$R' = n(N+1) - R \quad (4-16)$$

où n est la taille du plus petit échantillon.

Ainsi, les valeurs de R et R' sont comparés, respectivement, avec les valeurs critiques. Si R et R' sont inférieures à la valeur critique, alors l'hypothèse nulle de l'égalité des deux moyennes des échantillons est rejetée.

Dans le cas où les deux échantillons sont de taille identique, la somme des rangs R est :

$$R = \min(R, R')$$

La limitation du test vient de la condition que les deux échantillons doivent avoir deux distributions fréquentielles continues.

4.3.2 Test des sommes de rangs de Kruskal-wallis (k-population, Kanji, 1997)

L'objectif du test est de vérifier l'hypothèse que k échantillons aléatoires proviennent de k populations ayant la même moyenne. Ce test est également basé sur l'analyse des rangs. En effet, on combine les k échantillons et on les classe en ordre croissant en attribuant à chaque observation son rang correspondant. Soit R_j la somme des rangs du $j^{\text{ième}}$ échantillon et n_j la taille de cet échantillon. La statistique du test est donnée par :

$$H_{K-w} = \left\{ \frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_i^2}{n_j} \right\} - 3(N+1) \quad (4-17)$$

Cette statistique suit une loi du Chi-deux à (k-1) degrés de liberté. L'hypothèse nulle est rejetée si la valeur critique est supérieure à la valeur de la statistique.

L'utilisation du test nécessite une taille d'échantillon minimale, $n > 4$ et la distribution fréquentielle de chaque échantillon doit être continue.

Une procédure a été présentée par Cox et Lewis, (1966). Elle est basée sur les propriétés du second ordre des processus ponctuels.

On définit la quantité suivante :

$$I(t) = \frac{V(t)}{M(t)} \sim I = \frac{V'(\infty)}{m} \quad (4-18)$$

où :

$$M(t) = mt$$

$$V(t) = \text{var}(N_t)$$

L'approximation (4-18) nous permet d'estimer I(t) à partir de la pente asymptotique de la courbe de variance temporelle.

Dans cette procédure, on suppose que l'estimation de I(t) est disponible et on estime m par :

$$\hat{m} = \frac{n}{t}$$

sa variance est donnée par :

$$\text{var}(\hat{m}) = I \frac{\hat{m}}{t}$$

En utilisant ces deux estimations, on peut appliquer une approximation normale pour comparer deux taux d'occurrence. On constitue un rapport des deux taux d'occurrence et on adapte l'équation (4-8) à ce cas. On prend $\text{Log} \hat{m} = \text{Log} \left(\frac{n}{t} \right)$ qui suit une loi normale de

moyenne $\text{Log}(m) + \frac{I}{2n}$ et de variance $\frac{I}{n}$.

4.4 Test de stationnarité pour plusieurs séries d'événements

Dans ce test, on considère que les séries d'événements sont indépendantes et que les périodes d'observations pour chaque série sont exprimées par t'_0, t''_0, \dots

Identiquement au test basé sur l'équation (3-28), section 3.4.3.1, on peut considérer un modèle de tendance exponentielle pour chaque série d'événements. Ainsi pour la $j^{\text{ième}}$ série, on suppose que le taux d'occurrence est représenté par le modèle suivant :

$$e^{\alpha(j)+\beta t}$$

Pour tester l'hypothèse nulle $\beta = 0$, Cox et Lewis (1966) ont proposé la statistique suivante :

$$U_K = \frac{\left(\sum_i t'_i + \sum_i t''_i + \dots \right) - \frac{1}{2} (n' t'_0 + n'' t''_0 + \dots)}{\left\{ \frac{1}{12} (n' t'^2_0 + n'' t''^2_0 + \dots) \right\}^{\frac{1}{2}}} \quad (4-19)$$

où :

(n', n'', \dots) et (t'_i, t''_i, \dots) représentent, respectivement, les nombres d'événements et les temps d'occurrence enregistrés dans chaque série.

Sous l'hypothèse nulle, la statistique, exprimée par l'équation (4-19), suit asymptotiquement une distribution normale centrée réduite $N(0,1)$.



5. TESTS DE SERIES D'EVENEMENT DANS LES OUTILS INFORMATIQUES

Parmi les objectifs de cette étude, figure l'identification d'un logiciel statistique qui traite la modélisation et l'analyse des séries d'événements. Ainsi, nous nous proposons, dans ce chapitre, d'identifier les outils informatiques mis à la disposition des utilisateurs et traitant les processus ponctuels en général et les tests statistiques de séries d'événements, en particulier.

5.1.1 Méthodologie de la recherche

Il existe dans le marché de l'informatique une large gamme de logiciels statistiques. Ces outils informatiques sont élaborés et orientés selon la demande et le besoin des utilisateurs. Ainsi, on peut rencontrer des logiciels non spécialisés conçus pour une clientèle générale et d'autres qui visent une catégorie particulière de chercheurs ou de scientifiques. Notre recherche vise principalement les logiciels de statistique distribués à une grande échelle tels que le logiciel Splus, SAS, STATISTICA, SYSTAT ou SPSS. Ce choix est justifié par les raisons suivantes : la réputation et la diversité des techniques de ces logiciels et la disponibilité de l'information sur ces outils.

La recherche des outils informatiques traitant les processus ponctuels a été effectuée en deux étapes. Nous avons réalisé, dans un premier temps, une recherche exhaustive dans les revues de statistiques et d'informatique. Dans un deuxième temps, nous avons exploité l'information via le réseau INTERNET et plus particulièrement, dans les sites "web" des logiciels statistiques et des départements de statistique des universités.

5.1.2 Résultats de la recherche

Le premier résultat obtenu lors de cette recherche est un logiciel élaboré par Moore et Masse (1989). Ce logiciel fonctionne sous l'environnement DOS et il a été conçu dans le but de vérifier si un ensemble observé de points correspond à une réalisation d'un processus ponctuel donné. Les méthodes statistiques sont tirées de Diggle (1983). Ainsi, ce logiciel analyse les processus ponctuels spatiaux, qui n'ont pas été l'objet de notre étude.

Becue et al (1994) ont réalisé une enquête auprès des statisticiens et des utilisateurs de logiciels statistiques en France. L'objectif de leur étude était de connaître les types de logiciels utilisés et l'opinion des utilisateurs sur ces outils informatiques. Becue et al (1994) ont fourni aux répondants les différentes procédures qui peuvent exister dans un logiciel statistique et ils ont demandé aux répondants d'identifier les techniques qu'ils utilisent dans leurs applications. En analysant les résultats de cette partie, nous avons constaté l'absence d'un traitement statistique spécifique aux processus ponctuels.

Borgard et Guégan, (1996) ont comparé huit logiciels statistiques permettant l'étude et l'analyse des séries chronologiques linéaires à temps discret. Cependant, ils ont indiqué dans leur travail de recherche les différentes applications que chaque logiciel est capable d'effectuer. Encore une fois, la modélisation des processus ponctuels ne figurait pas parmi ces procédures.

Dans une récente étude, Morgan, (1998) a évalué et a comparé la performance de huit logiciels statistiques utilisés par une clientèle de non statisticiens. Cependant, l'utilisation de ces outils informatiques par cette classe d'utilisateurs reste un moyen important pour l'analyse des données et pour la prise de la décision. Les logiciels examinés sont : GraphPad Prism, InStat, ISP, NCSS, SigmaStat, Statistix, Statmost et Winks. Morgan (1998) a appliqué un certain nombre de tests sur ces outils et il a comparé les résultats avec ceux obtenus par le logiciel SAS. L'information que nous désirons tirer de cette étude est

l'identification d'un module qui traite la modélisation des processus ponctuels ou leur analyse. Morgan (1998) a énuméré les différentes techniques statistiques de chaque logiciel. Cependant, nous avons constaté que le terme processus ponctuel ne figure pas parmi les noms des procédures présentées dans cette étude.

Il existe dans la littérature peu de travaux effectués sur l'identification ou la comparaison des logiciels statistiques. Cependant, on peut rencontrer un nombre important de publications commerciales qui s'intéressent plus à l'aspect convivial et informatique du produit qu'à son contenu statistique. Toutes ces raisons nous ont obligé à effectuer plusieurs recherches via le réseau INTERNET. Ces recherches ont confirmé l'absence d'un logiciel ou de module d'un logiciel qui modélise les séries d'événements par les processus ponctuels temporels. Cependant, nous avons identifié le module S+SpatialStats du logiciel S+, considéré comme le premier programme orienté objet qui traite l'analyse des données spatiales. Les logiciels rencontrés possèdent généralement une ou plusieurs techniques illustrées dans le tableau 2.

Le tableau 3 représente les différents tests statistiques disponibles dans les six logiciels suivants : S+, STATISTICA, SPSS, SAS, MINITAB et SYSTAT.

Tableau 2 : Principales procédures des logiciels statistiques

PROCEDURES STATISTIQUES	
Statistiques descriptives	Modèle de régression
Séries chronologiques	Analyse de variance (ANOVA)
Analyse factorielles et classifications	Tests non-paramétriques
Analyse discriminante	Analyse de survie
Modélisation économique	Plan d'expérience

Tableau 3 : Identification des tests statistiques de six logiciels

Tests statistiques	S+	SPSS	SAS	MINITAB	STATISTICA	SYSTAT
Test du Student à un et deux échantillons et pour des échantillons appariés	X		X	X	X	X
Test de Spearman	X	X	X	X	X	X
Test de Kendall	X					
Test de Wilcoxon	X	X	X	X		X
Test de Pearson	X	X	X	X	X	X
Test de McNemar pour la symétrie	X	X				
Test de Mantel Haenszel	X					X
Test de Fisher	X	X	X		X	
Test de Chi-deux	X	X	X		X	
Test des proportions exactes	X					
Test de Friedman	X	X	X	X		X
Test de Kruskal-Wallis	X	X	X	X		X
Test du rapport de vraisemblance		X				
Test de Wald Wolfowitz		X	X			X
Test de Kolmogorov-Smirnov (1 et 2 échantillons)		X				X
Test de la médiane		X		X		
Test Jonckheere-Terpstra		X				
Test de Cochran		X				X
Test de Mann-Whitney		X	X	X		X
Test du signe		X	X	X		X
Test d'association linéaire		X				
Test d'homogénéité des variances	X		X	X		

6. CONCLUSION ET RECOMMANDATIONS

Cette étude a été menée dans le but d'identifier les tests statistiques d'une série d'un seul ou de plusieurs types d'événements pour l'analyse du risque. Il a été également proposé d'examiner la possibilité de l'élaboration d'un outil informatique qui modélise et analyse les séries d'événements. Les modèles adoptés pour la modélisation de ce type de données sont les processus ponctuels temporels. Ainsi, nous avons réalisé une revue exhaustive de la littérature sur les tests statistiques appliqués aux séries d'événements et une recherche d'un ou de plusieurs logiciels qui traitent ces données. Les principales conclusions tirées de cette étude sont :

- la grande majorité des travaux de recherche utilisent des processus ponctuels basés sur les hypothèses de stationnarité et d'indépendance pour modéliser les séries d'événements (processus de Poisson homogène, processus de renouvellement homogène). Cependant, les données disponibles dans la réalité ne sont pas toujours conformes à ces deux hypothèses ;
- nous avons identifié plusieurs méthodes pour l'analyse des séries d'un seul type d'événements, en partant de l'étape préliminaire qui est l'analyse graphique des données jusqu'aux tests statistiques basés sur des théories assez complexes. Les tests identifiés visent essentiellement l'analyse de tendance dans la série de données, l'indépendance et l'adéquation de la distribution statistique des intervalles, la discrimination entre deux ou plusieurs processus ponctuels et finalement la détection d'une association entre les processus ponctuels et les séries chronologiques temporelles.
- nous avons examiné la possibilité d'appliquer ces tests statistiques à une série de deux ou de plusieurs types d'événements. Cette opération est possible pour les durées de chaque type d'événement. Cependant, les tests appliqués aux nombres et aux temps d'occurrence ne sont pas toujours valables pour ce type de données. Une raison qui nous a poussée à étudier les propriétés des processus superposés. La superposition des

processus ponctuels utilisés dans la littérature est basée principalement sur le processus de Poisson homogène et le processus de renouvellement homogène. Ainsi, on a pu démontrer à titre d'exemple que la superposition de p processus ponctuels de Poisson est un processus de Poisson quand p tend vers l'infini. D'autres résultats de la superposition des processus ponctuels sont présentés au chapitre 4.

- nous avons également identifié quelques tests statistiques permettant la comparaison de deux ou de plusieurs séries d'un seul type d'événements.
- finalement, nous avons effectué des recherches dans la littérature et via le réseau INTERNET pour identifier les logiciels qui permettent la modélisation et l'analyse des séries d'événements. Les résultats de la recherche ont montré l'existence de deux logiciels qui traitent les processus ponctuels spatiaux et qu'il n'existe aucun outil informatique qui permette une analyse complète des processus ponctuels temporels.

Sur la base de ces résultats, on peut suggérer le développement d'un outil informatique permettant la modélisation des processus ponctuels temporels en partant des résultats du travail de recherche de Abi-Zeid (1997) et en incorporant les autres procédures d'analyses identifiées dans ce rapport. Nous avons construit un organigramme contenant les différentes étapes de la modélisation d'une série d'événements par un processus ponctuel (figure-7). Si la série de données ne représente pas une forme de tendance ou dépendance, il est raisonnable de modéliser cette série par un processus de Poisson homogène. Les tests d'hypothèses de ce modèle sont décrits dans le chapitre-3. Dans le cas où le processus de Poisson homogène n'est pas le modèle adéquat, on peut utiliser la procédure de sélection entre deux ou plusieurs processus ponctuels proposée par Smith et Karr (1985), section 3-7. Les tests d'hypothèses du modèle sélectionné sont résumés dans la figure-5, cependant, nous avons présenté les procédures d'analyse disponibles dans la littérature des processus ponctuels suivants : Poisson non-homogène, renouvellement homogène, renouvellement non-homogène, Poisson doublement stochastiques.

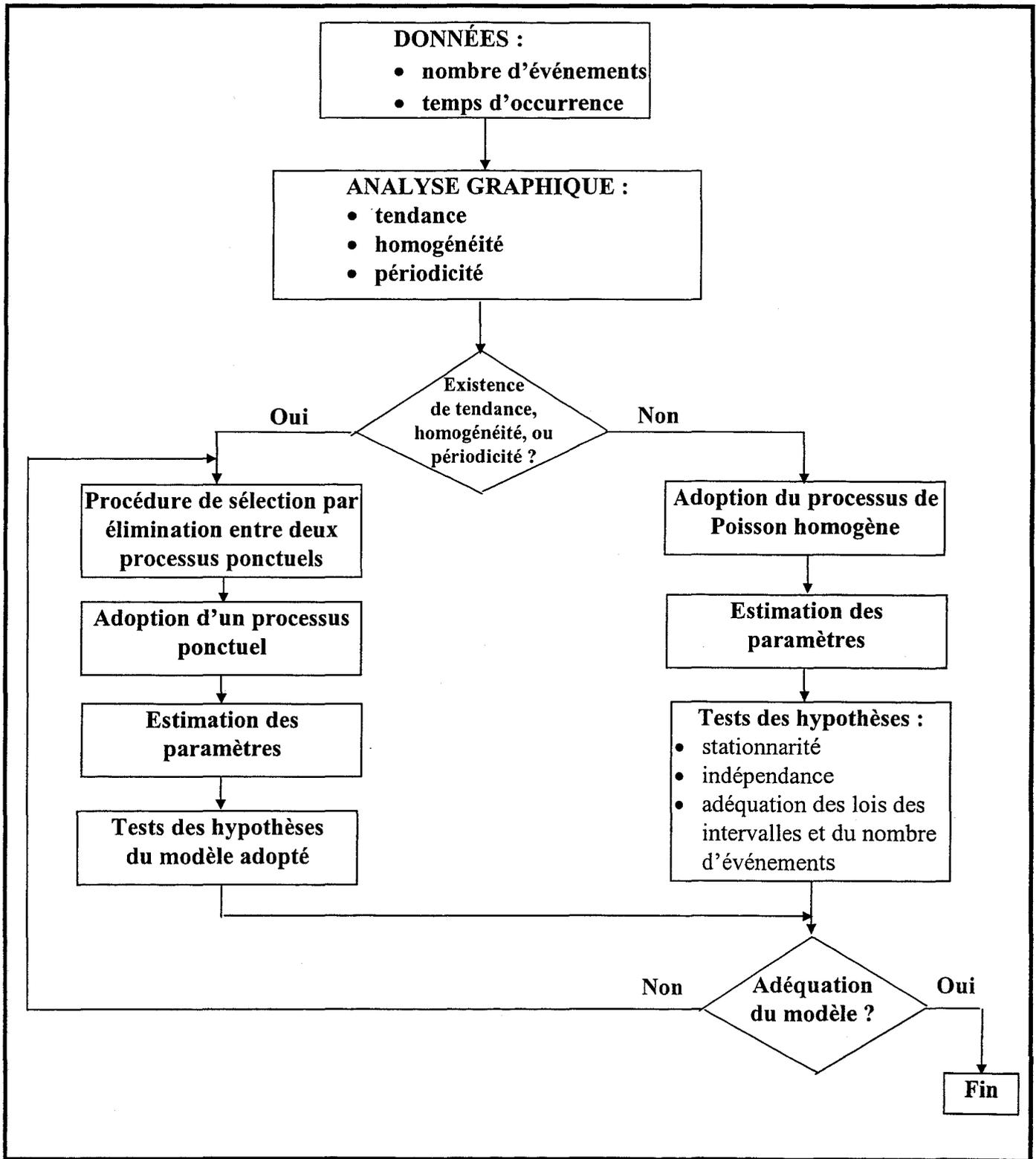


Figure 8 : Organigramme de la modélisation d'une série d'événements par un processus ponctuel valable pour un seul type.



7. REFERENCES BIBLIOGRAPHIQUES

- Abi-Zeid I. (1997). La modélisation stochastique des étiages et de leurs durées en vue de l'analyse du risque. Thèse de Doctorat. INRS-Eau. 278 pages.
- Abramowitz, M. et Stegun, I. A. (1964). Handbook of mathematical functions with formulas, graphs and mathematical tables. National bureau of standard applied. mathematics series. 55. Quatrième édition.
- Ambartzumian, R. V. (1965). Two inverse problems concerning the superposition of recurrent point processes. Journal of the Applied. Probability. No. 2 : 449- 454.
- Barnard, G. A. (1953). Time intervals between accidents. Biometrika. Vol. 40 : 212-213.
- Bartholmew, D. J. (1956). Tests for randomness in a series of events when the alternative is a trend. Journal of R. statist. Soc. B, 18 : 234-239.
- Bartholmew, D. J. (1957). Testing for départure from the exponential distribution. Biometrika. Vol. 44 : 253-257.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. Proc. Roy. Soc. A. 160 : 268-282.
- Becue, M, Dials, M. K, Grangé, D, Haeuster, L, Lechevallier, Y, Majjid, Y, Ringenbach, M, Peres, V et Scrmier, P. (1994). Enquête sur l'utilisation des logiciels de statistiques - ASU 1992. Revue statistique appliquée. Vol. XLII (3) : 5-12.
- Bernier, J. (1998). Risque et décisions en gestion de l'eau. Essai d'analyse de la rationalité du dialogue entre hydrologue et gestionnaire. [Tribune libre] Revue des Sciences de l'eau. 11 (1) : 3-24.
- Borgard, F et Guégan, D. (1996). Etude des séries chronologiques linéaire à temps discret. Comparaison des logiciels. Revue statistique appliquée. Vol. XLIV (4) : 59-80.
- Brillinger, D. R. (1994). Trend analysis : Time series and point process problems. Environmetrics, vol. 5 : 1-19.

- Chandramohan, J. et Liang, L. K. (1985). Bernoulli, multinomial and Markov chain thinning of some point processes and some results about the superposition of dependent renewal processes. *J. of Applied Probability*. 22 : 828- 835.
- Chebchoub, A, Bargaoui, Z, Abi-Zeid, I et Bobée, B. (1999). Modélisation des étiages à partir des débits moyens journaliers à l'extrême nord de la Tunisie. Programme de l'atelier : Estimation des crues et des étiages. École Nationale d'Ingénieurs de Tunis.
- Cinlar, E. (1968). On the superposition of m-dimensional point processes. *Journal of the Applied Probability*. 5 : 169-176.
- Cinlar, E. et Agnew, R. A. (1968). On the superposition of point processes. *Journal of the Royal. Statistical Society. Serie. B.30* : 576- 581
- Colson, G. (1985). Theories of risk and MCDM, Multiple criteria decision methods and applications. Eds. G. Fandel and J. Spronk, Springer-verlag : 171-196.
- Cox, D. R et Isham, V. (1980). *Point process*. Chapman et Hall. 188 pages.
- Cox, D. R. et Lewis, P. A. W. (1966) *The statistical analysis of series of events*. Methuen and Co LTD. 285 pages.
- Cox, D. R. et Smith, W, L. (1954). On the superposition of renewal processes. *Biometrika*, Vol. 41 : 91-99.
- Cox, D.R. (1955). Some statistical methods connected with series of events. *J. R. Statist. Soc. B*, 17 : 129-164.
- Cramer, H. (1946). *Mathematical methods of statistics*. Princeton University press.
- Cunnane, C. (1979). A Note on the Poisson assumption partial duration series models. *Water resources research*. Vol. 15, No. 2 : 489-494.
- Diggle, P. J. (1983). *Statistical analysis of spatial point patterns*. Academic press. 148 pages.
- Driancourt, C. et Streit, F. (1983). Mise en évidence de tendances pour des processus ponctuels de Poisson à l'aide de tests statistiques. *Pub. Inst. Stat. Univ. XXVIII*, fascicule .3 : 1-20
- Durbin, J. (1961). Some methods of constructing exact tests. *Biometrika*. Vol. 48 : 41-55.

-
- Faucher, D, Ouarda, T. B. M. J, et Bobée, B. (1997). Revue bibliographique des tests de stationnarité. Rapport de recherche R-499. Chaire en hydrologie statistique, INRS-eau. 65 pages.
 - Ghorbanzadeh, D. Et Lounes, R. (1996). Deux procédures de rupture pour des observations poissonniennes groupés. Revue de la statistique appliquée. XLIV(3) : 47-61.
 - Girault M. (1965). Processus aléatoires. Edition Dunod. 150 pages.
 - Gomes, M.I, Baroso, H. M et Amaral, M. A. (1975). Etude expérimentale de tests d'ajustements. Revue de statistique appliquée. Vol. XXIII (2) : 5-18.
 - Housset, D. Mazza, G. Gregoire, C. Piras, C. Malissen, B., et Fontecilla-Camps, JC.(1997). The three-dimensional structure of a T-cell antigen receptor. EMBO Journal. Vol. 14 : 4205-4216.
 - Ito, Y. (1977). Superposition of distinguishable point processes. Journal of the Applied Probability. 14 : 200-204.
 - Kanji, G. K. (1997). 100 statistical tests. SAGE Publication. 213 pages.
 - Karr, A. F. (1986). Point processes and their statistical inference. Marcel Dekker inc. 490 pages.
 - Kavvas, M. L.(1975). Point stochastic analysis of daily rainfall with application to Indiana. Thèse de doctorat à l'université de Purdue. 185 pages.
 - Kendall S.M., et Stuart A. (1979), The advanced theory of statistics. Vol. 2 : Inference and relationship. Edition Charles Griffin et Co Ltd (4ème édition).
 - Lam, C. T. et Lehoczky, J. P. (1991). Superposition of renewal processes. Advanced in Applied Probability. 23 : 64-85.
 - Lawrance, A. J. (1973). Dependency of intervals between events in superposition processes. Journal of the Royal. Statistical Society. Serie. B. 35 : 306-315
 - Lewis, P. A. W. (1965). Some results on tests for Poisson processes. Biometrika, Vol. 52 : 67-78.
 - Massey, F. J. (1950). A note on the power of a non-parametric test. Ann. Math. Statist. 20 : 440-442.

- Moore, M et Masse, J. Y. (1989). Un logiciel pour l'analyse statistique des processus ponctuels. Département de mathématiques appliqués. Ecole polytechnique de Montréal.
- Moran, P. A. P. (1951). The random division of a an interval. Part II. J. R. statist. Soc. B. 13 : 147-150.
- Morgan, W. T. (1998). A review of eight statistics software parckages for general use. The American Statistician. Vol. 52, N°. 1 : 70-82.
- North, M. (1981). Processus hydrologique intermittents. Etude de modélisations stochastiques et applications. Thèse de Doctorat, no. 419, Ecole Polytechnique Fédérale de Lausanne. 194 pages.
- Ondo, J. C, Ouarda, T. B. M. J, et Bobée, B. (1997). Revue bibliographique des tests d'homogénéité et d'indépendance. Rapport de recherche N° R-500. Chaire en hydrologie statistique, INRS-Eau. 78 pages.
- Rao, C. R. et Chakravati, I. M. (1956). Some small sample tests of signficance for a Poisson distribution. Biometrika. 12 : 264-282.
- Ripley, B. D. (1976). On stationarity and superposition of point processes. The Annals of Probabability. Vol. 4, No. 6 : 999-1005
- Sherman, B. (1950). Arandom variable related to the spacing of sample values. Ann. Math. Statist. 21 : 339-361.
- Smirnov, N. V. (1949). On the Cramer-von Mises criterion. Uspehi, Matem, Nauk. (NS), 4 : 196-197.
- Smith, J. A. (1980) Point process models of rainfall. Thèse de doctorat à l'université de Johns Hopkins. 181 pages.
- Smith, J. A. et Karr, A. F. (1985). Statistical inference for point process models of rainfall. Water resources research. Vol. 21. No. 1 : 73-79
- Snyder, D. L. (1975). Random point processes. John Wiley & sons, Inc. 485 pages.
- Sukhatme, P. V. (1938). On the distribution of χ^2 in samples of the Poisson series. Journal of the royal statistical society. 5. Suppl. 1 : 75-79.
- Thompson, W. A. (1988). Point process models with application to safety and reliability. Chapman et Hall. 146 pages.

-
- Wald, A. et Walfowitz, J. (1943). An exact test of randomness in the nonparametric case based on serial correlation. *Ann. Math. Statist.* 14 : 378-388.
 - Waymire, E. et Gupta, V. K. (1981). The mathematical structure of rainfall representations : 2. A review of the theory of point processes. *Water Resources Research*, vol. 17, No. 5 : 1273-1285.
 - Wolfe, A. (1980). A circular test of the association between a point process and a continuous time series. *Communication in statistics. Theory and method.* A9(12) : 1299-1307.

8. ANNEXE A

Tableau A 1 : Tests de stationnarité basés sur la moyenne (Faucher et al, 1997)

Test statistique	Saut	Tendance monotone	Tendance en escalier
<i>TESTS PARAMETRIQUES</i>			
Student	x		
Déviations cumulées	x		
Worsley	x		
Hinkley	x		
Cramer	x		x
Régression			
Kite		x	
Bayes	x	x	
Hubert			x
<i>TESTS NON PARAMETRIQUES</i>			
Wilcoxon	x		
Mann-Whitney	x		
Lepage	x		
Terry-Hoeffding	x		
Van der Waerden	x		
Bell-Doksum	x		
Terpstra-Jonckheere	x		x
Kruskall-Wallis	x		x
Mann		x	
Kendall turning point		x	

Suite du tableau A1

Gold		x	
Pettitt	x		

x : le test est applicable

Tableau A 2 : Tests de stationnarité basés sur la variance (Faucher et al, 1997)

Tests	Comparaison de 2 variances	Comparaison de k variances
TESTS PARAMETRIQUES		
Fisher	x	
Bartlett		x
Hartley		x
TESTS NON PARAMETRIQUES		
Mood	x	
Ansari-Bradely	x	
Moses	x	
Kruskall-Wallis		x

Tableau A 3 : Test d'homogénéité appliqués à une tendance en escalier (Ondo et al, 1997)

Test statistique	Fonction de répartition empirique	Moyenne empirique
TESTS NON PARAMETRIQUES		
Birnbaum-Hall (k=3)	x	
Smirnov à k >2 sous-échantillons (unilatéral)	x	
Smirnov à k >2 sous-échantillons (bilatéral)	x	
Kruskal-Wallis		x
Van der Warden		x

Tableau A 4 : Test d'homogénéité appliqués à une tendance en saut (Ondo et al, 1997)

Test statistique	Fonction de répartition empirique	Moyenne empirique	Autres
TESTS PARAMETRIQUES			
Chi-deux	x		
Normality test		x	
Student		x	
Test de la médiane			x
Déviations cumulées			x
Rapport de vraisemblance de worsley			
Procédures bayésienne			
TESTS NON PARAMETRIQUES			
Kolmogorov-Smirnov	x		
Cramer-von Mises	x		
Mann-Whitney		x	
Wald-Wolfowitz			x

x : le test est applicable

Tableau A 5 : Tests d'indépendance (Ondo et al, 1997)

Test statistique	H ₁ : Persistance	H ₁ : Tendance	H ₁ : Effet cyclique
TESTS PARAMETRIQUES			
Ljung-box	x		
Box-Pierce	x		
Bartlet	x		
Anderson	x		
Brock, Dechert et Scheinkman	x		

Suite du tableau A5

TESTS NON PARAMETRIQUE			
Test des groupes	x		
Wallis et Moore	x		
Von Neumann	x		
Wald-Wolfowitz	x		
Foster et Stuart		x	
Cox-stuart		x	
Difference-sign		x	
Rangs de Spearman		x	
Rangs de Kendall		x	
Mann-Whitney saisonnier			x

x : le test est applicable