UNIVERSITÉ CATHOLIQUE DE LOUVAIN
INSTITUT DE STATISTIQUE

Recursions for Actuaries and Applications in the Field of Reinsurance and Bonus-Malus Systems

Membres du jury :
Prof. José Paris (Promoteur)
Prof. Léopold Simar (Président du jury)
Prof. Jean-Marie Rolin
Prof. Michel Denuit
Prof. Christian Jaumain
Prof. Paul Embrechts (ETH Zürich)
Prof. Harry Panjer (University of Waterloo)

Thèse présentée en vue de l'obtention du grade de Docteur en sciences (orientation statistique) par

Jean-François Walhin

Louvain-la-Neuve
Septembre 2000

Contents

Acknowledgments

First of all, I would like to thank my parents for the education I was given and for the opportunity they offered me to become a civil engineer and an actuary at the Université Catholique de Louvain.

This thesis would not have been possible if I had not encountered Professor José Paris. I had already written three theses under his supervision before this PhD thesis. He not only taught me probability theory but gave me the taste for actuarial sciences. He always told me that a good theoretical background is necessary in order to find relevant practical applications. I hope the reader will be convinced of the truth of this saying after reading this PhD thesis. Most important has been the relationship between Professor Paris and myself during the last six years. It were rather relations between two colleagues than between a Professor and his assistant (student). This was already the case for our research work but also for the teaching part of our job at the University. Let me thank you, Professor, for having accepted to be the supervisor of my PhD thesis after having become emeritus.

There has been another important encounter that has contributed to the completion of this PhD thesis. Didier Boussemart was my first CEO at the former company Le Mans Assurances. Not being an actuary, he realized that the actuary had an important part to play in an insurance company. On my first work day, he asked me to be in charge of reinsurance. This brought me into contact with the practical world of reinsurance and clearly influenced my research topics (as well as my knowledge of some of the best restaurants in Brussels). Practical reinsurance has been a motivation for writing this PhD thesis.

I would also like to thank all the members of my jury. Their helpful remarks at the private defence stage allowed me to enhance the quality of the text. Special thanks to the non UCL members, namely Professor Paul Embrechts and Professor Harry Panjer, for having accepted to sit in my jury.

I have also derived much benefit from of a lot of useful conversations with people at the Institut de Statistique or in Insurance and Reinsurance companies. I would like to thank, for practical or theoretical discussions Pierre Ars, Valentin Patilea, Michel Denuit, Ivo De Bondt, Philippe De Longueville, Mohamed Snoussi,...
Some parts of this thesis have been used in my teaching to my students. That was obviously an excellent experimental laboratory and I thank all my students for their patience and attention.

At the Université Catholique de Louvain, I benefitted of great Professors when I was myself studying to become a civil engineer, a statistician or an actuary and I thank them all. As actuarial science has become my favourite subject matter, it is a pleasure to thank those

Professors who have taught me quantitative actuarial sciences, namely Professor Paris for statistics in non life insurance (from a theoretical point of view with the practical case in mind), Professor Jaumain for life insurance and financial mathematics (from the practical point of view with the theoretical aspects in mind) and Professor De Vylder for life and non life insurance (from a mathematical point of view with some mathematical intuition).

You will have noticed that English is not my mother tongue. That is why some parts of this thesis have been reread by Inge Staelens and Ernest Van de Winkel. A lot of thanks to both of you, for making the text more English-like.

Some parts of this thesis have already appeared elsewhere in literature. This is the opportunity for me to thank those anonymous referees who did a great job by helping me to write more pedagogical papers. I have really benefited from enlightened advice regarding the scientific communication.

Informatics is important when writing a PhD thesis. I have benefitted of excellent advice regarding the typesetting of this work from Jean-Luc Marrion.

Mainly thanks to my employers in the private sector, I was offered the opportunity to present some parts of this thesis at various congresses. My thanks also go, for the same reasons, to the Fonds National de la Recherche Scientifique, the Communauté Française de Belgique and the Institut de Statistique.
Writing a PhD thesis means having at one's disposal some time to do it. I therefore thank the Institut de Statistique where I have been, during five years, a half-time teaching assistant. I also thank my present employer, Secura Belgian Re, who offered me facilities in order to complete this PhD thesis in a reasonable time.

Finally I thank my wife for her encouragements and her support during those years when I remained an eternal student. I thank her in particular for having accepted to share our week-ends with my (our ?) computer. However, Laurence, keep in mind that some papers are not ready yet and that I have not given up writing on other subjects. He who starts research work is, as it were, driven by an endeavouring mind that he cannot resist.

Notations

The following notations are common in this thesis :

- iid stands for independently and identically distributed
- $Coll$ is an abbreviation for Collective risk model
- Ind is an abbreviation for Individual risk model
- $p(n)$ denotes the probability function of a counting variable N :

$$p(n) = \mathbb{P}[N = n]$$

- $p(n, m)$ denotes the joint probability function of a bivariate counting vector (N, M) :

$$p(n, m) = \mathbb{P}[N = n, M = m]$$

- $f_X(x)$ denotes the probability function of the claim amount X :

$$f_X(x) = \mathbb{P}[X = x]$$

- $F_X(x)$ denotes the cumulative density function of the claim amount X :

$$F_X(x) = \mathbb{P}[X \leq x]$$

- $\psi_X(u)$ denotes the probability generating function of X
- $g_i(x)$ also denotes a probability function (in the context of the individual risk model)
- $G_i(x)$ also denotes a cumulative density function (in the context of the individual risk model)
- $\psi_i(u)$ denotes the probability generating function of risk i in the context of the individual risk model
- I is an indicator function
- S denotes the aggregate claims distribution
- $f_X^{*n}(x)$ denotes the probability function of $X_1 + \cdots + X_n$
- $f_S(x)$ denotes the probability function of the aggregate claims distribution

- $F_S(x)$ denotes the cumulative density function of the aggregate claims distribution
- $\psi_S(u)$ denotes the probability generating function of S
- $f_{S,T}(x,y)$ denotes the joint probability function of a bivariate vector

$$f_{S,T}(x,y) = \mathbb{P}[S = x, T = y]$$

- a has different meanings :
  (a) The number of types of claim amounts in the individual risk model
  (b) One of the parameters for the (a, b, m) class
  (c) One of the parameters of the Hofmann Distribution
  (d) One of the parameters of Weighted Poisson Distribution
- b has different meanings :
  (a) The number of types of claim probability in the individual risk model
  (b) One of the parameters for the (a, b, m) class
- p has different meanings :
  (a) The proportion of the driving population that reports all accidents
  (b) One of the parameters of the Hofmann Distribution
  (c) One of the parameters of the ZIP Distribution
- r has different meanings :
  (a) The order of the De Pril's approximation
  (b) One of the parameters for the (r, s, m) class which is a reparametrization of the (a, b, m) class
  (c) One of the parameters of the Negative Binomial Distribution
  (d) One of the parameters of Weighted Poisson Distribution
  (e) The number of mass points of the nonparametric mixing distribution in the Mixed Poisson Process
  (f) The maximum number of claims in the pro rata temporis reinstatements frame
- n has different meanings :
  (a) The number of risks in the individual risk model
  (b) The number of correlated random variables in the multivariate extensions
  (c) The observed value of the random variable N
  (d) The size of an univariate sample
- m has different meanings :
  (a) One of the parameters of the (a, b, m) class

(b) The observed value of the random variable M
(c) The number of excess of loss reinsurance layers
(d) The number of claims already reported in Lemaire's algorithm

- q denotes the number of layers in an excess of loss protection
- β has different meanings :
  (a) one of the parameters of the Negative Binomial Distribution
  (b) one of the parameters of the Mixed Bivariate Independent Hofmann Distribution
  (c) the actualisation rate in Lemaire's algorithm
  (d) the expenses percentage of the Reinsurer in a profit participation
- t denotes the time
- s has different meanings :
  (a) The number of classes of a bonus-malus system
  (b) The observed value of S
  (c) One of the parameters for the (r, s, m) class which is a reparametrization of the (a, b, m) class
- c has different meanings :
  (a) One of the parameters of the Hofmann Distribution
  (b) The price of reinstatements
  (c) The average retention of a portfolio within the bonus hunger frame
- α has different meanings :
  (a) The profit percentage in a profit participation
  (b) The size of a bivariate sample
  (c) A parameter of the $P - G(\alpha, \delta, \theta)$ Distribution
  (d) The fraction between average cost of bodily injury claims and average cost of material damage claims
  (e) The safety loading with the expected value principle
- δ has two meanings :
  (a) A parameter of the Ho+Po Distribution
  (b) A parameter of the $P - G(\alpha, \delta, \theta)$ Distribution
- λ has two meanings :
  (a) The parameter of the Poisson Distribution
  (b) One of the parameters of the Weighted Poisson Distribution
  (c) One of the parameters of the Generalized Poisson Distribution

- γ has different meanings :
  (a) The skewness
  (b) The scale parameter of an exponential utility function
  (c) The solidarity index within the bonus-malus system with an exponential loss function
  (d) The coefficient loading with the standard deviation principle
- κ denotes the kurtosis
- σ denotes the standard deviation
- $\mathbf{S} = (S_1, \ldots, S_n)$
- D denotes the deductible of an excess of loss cover
- L denotes the limit of an excess of loss cover
- u and v are arguments of bivariate probability generating functions
- $u_1, \ldots, u_n$ are arguments of multivariate probability generating functions
- k denotes the number of reinstatements
- h denotes the discretization span

Introduction

Nowadays, it is commonly considered that there are three types of actuaries : life actuaries, non-life actuaries and actuaries of the third kind : we could call them financial actuaries.
More and more actuaries have to face problems related with assets and liabilities. This is the reason why finance is taking a larger place in the education of actuaries. I am certainly aware of this fact and think it is good practice.

The present thesis, however, will not follow such a modern approach. It will only concentrate on non-life actuarial mathematics because I think that some recent techniques deserve particular attention. A few of these modern techniques worth mentionning are : stochastic ordering, relaxation of the independence hypotheses, multivariate recursive techniques.

Life insurance is not forgotten and some applications will be given for life portfolios, based on traditional non-life techniques.

The basic motivation of this thesis is obviously the Panjer's recursion. This recursion is a tool in order to find the aggregate claims distribution under specific circumstances without using the classical brute force convolution formula.
It is not the aim of this thesis to give an exhaustive overview of the work that was published in this area after Panjer's paper in 1981.
I will present some multivariate extensions of classical recursions instead with practical applications in mind. The individual risk model and the collective risk model will be extended in a multivariate setting.
This will give us the necessary tools for the evaluation of excess of loss reinsurance covers with specific clauses. In particular we will be able to find the ruin probability of the Cedent when it buys such a protection.
It is remarkable in literature that excess of loss reinsurance has received such great attention from an academical point of view. However, in practice, these treaties are often accompanied by clauses and then the classical derivations need multivariate extensions of the classical techniques.

After having introduced correlation in the aggregate claims distribution due to the correlation in the claim amount distributions, we will concentrate on the counting distribution.
Very recent models will be briefly reviewed and we will essentially insist on the Hofmann Distribution because this distribution has a great potential, being Mixed Poisson and Compound Poisson distributed, encompassing a lot of classical counting distributions, and presenting some more theoretical and practical properties.
Recently, the dependence between the number of claims has been discovered in actuarial sciences. A typical application is the study of the number of material damage claims and bodily

injury claims in motor insurance. Some results, in fact, come from probability theory. Once again I will not give an exhaustive overview of the literature. I will rather concentrate on bivariate extensions of the Hofmann Distribution. Special attention will be paid to the recursive evaluation of the compound bivariate distributions. This will give us the opportunity to use the results derived for the multivariate extension of the collective risk model.

Finally, considering the importance of the subject nowadays in Belgium, I will comment on the practical replacement of a bonus-malus system. This will be an application of two different types of counting distributions : the Parametric Mixed Poisson Distribution and the Nonparametric Mixed Poisson Distribution.

The text has been written to enable the reader to distinguish what is known literature from what is new. In the latter case, the section or chapter begins by mentioning where the results have been published. Proofs are only given for new results.

Practical numerical applications are provided for all concepts described in this PhD thesis.

Chapter 1

Some well-known univariate recursions in actuarial sciences

In this chapter we will introduce two types of popular models in actuarial sciences : the individual risk model and the collective risk model. Both models have received great attention in the literature related to the problem of evaluating the probability function of aggregate claims distribution. This problem is tackled by using recursive schemes. We will review these recursions in the present chapter. Some amendments will be given as well as some historical observations. A model extending the recursion of the collective risk model is the (a, b, m) class. The recursions coming from the (a, b, m) class will be reviewed in this chapter as well. In order to run the recursions, we will be obliged to work with arithmetic claim amount distributions. A random variable is said to be arithmetic if it is distributed on $0, 1, 2, \ldots$.
If the claim amount distribution is not arithmetic, then chapter 2 will provide a solution.

1.1 The individual risk model

1.1.1 Definition

Let us assume a portfolio of n independent policies. Each policy may generate at most one claim during the period considered (typically one year). This is obviously the case in life insurance and it is also quite realistic for fire insurance.
The probability that policy i leads to a claim is q_i. If a claim has occurred, the claim amount distribution is modelled by a random variable X_i the cumulative density function of which is denoted by $G_i(x)$. Let us define the indicator I_i : $I_i = 1$ if risk i leads to a claim, $I_i = 0$ else. We assume that the random variables I_i, X_i , $i = 1, \ldots, n$ are mutually independent.
The cumulative density function of a single risk writes

$$F_i(x) = p_i \mathbb{I}(x) + q_i G_i(x)$$

where $\mathbb{I}(x) = 0$ if $x < 0$ and $\mathbb{I}(x) = 1$ if $x \geq 0$ denotes the classical scale function.
A particular case of the individual risk model is the individual life model where the claim amounts are non random, being simply deterministic risk sums.
$F_S(x)$ is the cumulative density function of

$$S = I_1 X_1 + \cdots + I_n X_n$$

and is the aggregate claims distribution of the portfolio. We have

$$F_S(x) = F_1 * \cdots * F_n(x)$$

It is clear that evaluating the last convolution product is very time consuming. This is the reason why different algorithms have been given in literature in order to find the distribution of S, or an approximation of it, as fast as possible.
Literature has been developped since the 80's. Kornya (1983) gave an approximate formula to solve the individual life model whereas De Pril (1986) gave an exact formula as well as an approximate formula based on the exact one (De Pril (1988)). The formulae have been extended to the individual risk model by De Pril (1989). Note that the exact formula of De Pril is difficult to use and very time consuming.
Recently Dhaene and Vandebroek (1995) gave a new exact recursion for the individual risk model improving the formula of De Pril (1989).
The algorithms will be given in the next section.

1.1.2 The algorithms

We often have policies with the same characteristics (q_i and $G_i(x)$) in a portfolio. It is then possible to rewrite the model in an homogeneous form. This form leads to faster algorithms and we will therefore adopt the notation of the homogeneous individual risk model.
The policies are classified into $a \times b$ homogeneous classes in the following way : the policies of row j have a probability q_j $(j = 1, \ldots, b)$ of producing a claim and the conditional distribution of the claim for a policy in column i given that a claim has occurred is $g_i(x)\,(i = 1, \ldots, a)$.
Let n_{ij} be the number of policies having the characteristics i and j and n the total number of policies.

		claim amount distribution					
		$g_1(x)$	$g_2(x)$	$\cdots$	$g_i(x)$	$\cdots$	$g_a(x)$
	q_1						
	q_2						
claim	$\vdots$						
probability	q_j				n_{ij}		
	$\vdots$						
	q_b						

We have

$$\sum_{j=1}^{b} \sum_{i=1}^{a} n_{ij} = n$$

It is assumed that $0 < q_j < 1$ and that the claim amounts (X_i) are distributed on $x = 1, 2, \ldots, m_i$, i.e. with a maximal claim amount m_i.
$M = \sum\limits_{i=1}^{a} \sum\limits_{j=1}^{b} n_{ij} m_i$ is the maximal amount for the aggregate claims.
Let $f_S(s), s = 0, \ldots, M$ be the probability function of the aggregate claims.
In order to find this distribution, we will use the probability generating function :

$$\psi_S(u) = \sum_{s=0}^{M} f_S(s) u^s = \prod_{i=1}^{a} \prod_{j=1}^{b} (p_j + q_j \psi_i(u))^{n_{ij}}$$

where $\psi_i(u) = \sum\limits_{x=1}^{m_i} g_i(x) u^x$.

Different techniques have been described in order to evalute $f_S(s)$. Dhaene and Vandebroek (1995) gave an exact algorithm for the evaluation of the $f_S(s)$ with the following steps:

$$\begin{aligned}
f_S(0) &= \prod_{i=1}^{a} \prod_{j=1}^{b} (1-q_j)^{n_{ij}} \\
s f_S(s) &= \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} v_{ij}(s) \quad , \quad s = 1, \ldots, M \\
v_{ij}(s) &= \frac{q_j}{p_j} \sum_{x=1}^{m_i} g_i(x) \left(x f_S(s-x) - v_{ij}(s-x) \right) \quad , \quad s \geq 1 \\
& 0 \quad \text{otherwise}
\end{aligned}$$

De Pril (1989) proposed an approximate algorithm valid if $q_j < \frac{1}{2} \quad \forall j$: you choose the degree of approximation r and you evaluate

$$\begin{aligned}
f_S^{(r)}(0) &= \prod_{i=1}^{a} \prod_{j=1}^{b} (1-q_j)^{n_{ij}} \\
s f_S^{(r)}(s) &= \sum_{i=1}^{a} \sum_{k=1}^{\min(r,s)} A(i,k) \sum_{x=k}^{\min(s,km_i)} x g_i^{*k} f_S^{(r)}(s-x) \quad , \quad s = 1, \ldots, M \\
A(i,k) &= \frac{(-1)^{k+1}}{k} \sum_{j=1}^{b} n_{ij} \left(\frac{q_j}{p_j} \right)^k \quad , \quad 1 \leq i \leq a \quad , \quad 1 \leq k \leq r
\end{aligned}$$

The De Pril's approximation can also be described as a two stage algorithm :

$$\begin{aligned}
t^{(r)}(0) &= \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \ln p_j \\
t^{(r)}(x) &= \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \sum_{k=1}^{r} \frac{(-1)^{k+1}}{k} \left[\frac{q_j}{p_j} \right]^k g_i^{*k}(x) \\
f_S^{(r)}(0) &= e^{t^{(r)}(0)} \\
s f_S^{(r)}(s) &= \sum_{x=1}^{s} x t^{(r)}(x) f_S^{(r)}(s-x) \quad , \quad s = 1, \ldots, M
\end{aligned} \tag{1.1}$$

Dhaene and De Pril (1994) gave a bound on the error commited by the approximate formula of De Pril in term of the total variation distance :

$$2 d_{TV}(S, S^{(r)}) = \sum_{s=0}^{\infty} |f_S(s) - f_S^{(r)}(s)| \leq e^{\epsilon} - 1$$

where

$$\epsilon = \sum_{x=0}^{\infty} |t(x) - t^{(r)}(x)| \leq \frac{1}{r+1} \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} \frac{p_j}{p_j - q_j} \left(\frac{q_j}{p_j} \right)^{r+1} \tag{1.2}$$

1.1.3 The number of arithmetic operations

Dhaene and Vandebroek (1995) relied on the one stage formula in order to draw conclusions about the number of multiplications needed to perform the De Pril's approximate algorithm:

$$(3ar + r(r+1)\sum_{i=1}^{a}(m_i - 1))s$$

They also found that the number of multiplications for their exact algorithm is

$$2b(a + \sum_{i=1}^{a} m_i)s$$

s is the value up to which the probability function $f_S(s)$ is evaluated.
We can see, however, that the De Pril's approximation is always better when it is considered as a two stage algorithm : $t^{(r)}(x)$ has to be evaluated from $x = 0$ to $x = r \max_i m_i$ which needs $2abr^2 \max_i m_i$ multiplications. This number does not depend on s and may be neglected.
The recursion (1.1) needs $(2r \max_i m_i + 1)s$ multiplications.
Now the conclusion of this section is that the efficiency if by far more important if one works with the two stage algorithm for the De Pril's approximation. It will always be better than the one stage formula :

$$(2r \max_i m_i + 1)s < (3ar + r(r+1)\sum_{i=1}^{a}(m_i - 1))s$$

It will be better than the exact Dhaene and Vandebroek (1995) formula if

$$r \max_i m_i < b \sum_{i=1}^{a} m_i$$

This will certainly be the case when $r < b$ and may be the case when $r > b$, precisely when

$$r < b \frac{\sum_{i=1}^{a} m_i}{\max_i m_i}$$

1.2 The collective risk model

In the collective risk model, we do not look at the policies individually. The policies are anonymous in the whole portfolio.
We assume that the number of claims, N, is given by a Poisson random variable whereas the claim amount random variable, X, is described by its cumulative density function $F_X(x)$.
We are interested in the distribution of the aggregate claims distribution of

$$S = X_1 + \cdots + X_N$$

We say that S has a compound Poisson distribution.
The cumulative density function of S is given by

$$F_S(x) = \sum_{n=0}^{\infty} e^{-\lambda} \frac{\lambda^n}{n!} F_X^{*n}(x) \quad , \quad x \geq 0 \tag{1.3}$$

This formula is of course very time consuming. Furthermore it needs an infinite sum. Fortunately, the Poisson distribution belongs to the Panjer's class of counting distribution for which there exists a recursive algorithm (see section 1.3 for details) :

$$\begin{aligned} f_S(0) &= e^{-\lambda(1-f_X(0))} \\ f_S(x) &= \lambda \sum_{i=1}^{x} \frac{i}{x} f_X(i) f_S(x-i) \quad , \quad x \geq 1 \end{aligned}$$

Note that when the q_i's are small, the collective risk model may be viewed as a good approximation of the individual risk model. This is just the extension of the well known result saying that the binomial distribution can be approximated by a Poisson distribution in certain circumstances.
Let us define S^{Ind} the aggregate claims distribution with the individual risk model and S^{coll} the aggregate claims distribution under the collective risk model.
It is well known that the collective risk model is a good approximation of the individual risk model when

$$\begin{aligned} \lambda &= \sum_{j=1}^{b} \sum_{i=1}^{a} n_{ij} q_j \\ f_X(x) &= \frac{\sum_{j=1}^{b} \sum_{i=1}^{a} n_{ij} q_i g_i(x)}{\lambda} \end{aligned}$$

De Pril and Dhaene (1992) gave upper bounds on the error commited when using the collective risk model instead of the individual risk model :

$$d_K(S^{Ind}, S^{Coll}) \leq \frac{1}{2} \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} q_j^2$$

$$d_{TV}(S^{Ind}, S^{Coll}) \leq \sum_{i=1}^{a} \sum_{j=1}^{b} n_{ij} q_j^2$$

where d_K denotes the Kolmogorov distance whereas d_{TV} denotes the total variation distance :

$$\begin{aligned} d_K(X,Y) &= \max_x |F_X(x) - F_Y(x)| \\ d_{TV}(X,Y) &= \sup_{A \subseteq \mathbb{Z}^+} |\mathbb{P}(X \in A) - \mathbb{P}(Y \in A)| \\ &= \frac{1}{2} \sum_{m=0}^{\infty} |\mathbb{P}(X = m) - \mathbb{P}(Y = m)| \end{aligned}$$

1.3 Recursions for the (a,b,m) class

In this section we use a more general model than the collective model. We relax the hypothesis of a Poisson random variable.
Let N be a counting random variable. The probability function of N is given by $p(n)\,, n \geq 0$.
Let X be an arithmetic random variable representing claim amounts. The probability function of X is given by $f_X(x)\,, x \geq 0$. We assume that N and X are mutually independent.
Like in section 1.2 we are interested in the distribution of the aggregate claims :

$$S = X_1 + \cdots + X_N$$

the probability function of which is given by

$$f_S(x) = \sum_{n=0}^{\infty} p(n) f_X^{*n}(x) \quad , \quad x \geq 0$$

It is necessary to find algorithms giving the probability function of S faster than with the brute convolution formula.
In this section we concentrate on particular classes of counting variables.
A very well known class of counting distributions in actuarial sciences is the Panjer's class. This class of counting distributions received Panjer's name in the actuarial literature because Panjer (1981) discovered a recursive algorithm allowing to compute easily the aggregate claims distribution if the counting distribution belongs to the Panjer's class.
The Panjer's class contains the counting random variables satisfying the relation

$$p(n) = (a + \frac{b}{n}) p(n-1) \quad , \quad n > 1$$

In fact this class of random variables was discovered by Katz (1965) who studied the family of distributions such that

$$\frac{p(n+1)}{p(n)} = \frac{\alpha + \beta n}{1+n} \quad , \quad n \geq 0$$

This leads to the Panjer's class with the reparametrization :

$$\begin{aligned} \alpha &= a + b \\ \beta &= a \end{aligned}$$

Katz (1965) showed that only three distributions belong to the Panjer's class :

- Poisson (Po) : a=0 , $b = \lambda$
- Negative Binomial (NB) : $a = \frac{\beta}{1+\beta}$, $b = (r-1)\frac{\beta}{1+\beta}$
- Binomial (Bi) : $a = -\frac{p}{1-p}$, $b = (N+1)\frac{p}{1-p}$

where the random variables are parametrized as :

$$\begin{aligned} f_{Po}(n) &= e^{-\lambda}\frac{\lambda^n}{n!} \quad , \quad n \geq 0 \\ f_{BN}(n) &= \frac{\Gamma(r+n)}{\Gamma(r)n!}\left(\frac{\beta}{1+\beta}\right)^n \left(\frac{1}{1+\beta}\right)^r \quad , \quad n \geq 0 \\ f_{Bi}(n) &= \frac{N!}{(N-n)!n!} p^n (1-p)^{N-n} \quad , \quad 0 \leq n \leq N \end{aligned}$$

This result is often attributed to Sundt and Jewell (1981) in the actuarial literature while it was shown by Katz (1965).
We will continue to use the term Panjer's class even if the class has been discovered by Katz because its importance lies in the algorithm derived by Panjer (1981) for evaluating the probability function of the compound distribution with counting variable being in the Panjer's class.

The Poisson distribution is important because it it used in the collective risk model (see section 1.2). The Negative Binomial has received great attention in the actuarial literature as it often gives better fits than the Poisson distribution. The Negative Binomial distribution presents the good property of overdispersion, which is common in insurance portfolios. The Binomial distribution is probably the least interesting distribution. It is a distribution with the property of underdispersion.

If N belongs to the Panjer's class of counting distributions, we have the following algorithm due to Panjer (1981) :

$$\begin{aligned} f_S(0) &= \psi_N(f_X(0)) \\ f_S(x) &= \frac{1}{1-af_X(0)}\sum_{i=1}^{x}(a+b\frac{i}{x})f_X(i)f_S(x-i) \quad , \quad x \geq 1 \end{aligned}$$

where $\psi_N(u) = \mathbb{E}[u^N]$ is the probability generating function of N.
If N belongs to the Panjer's class, then its probability generating function is given by

$$\psi_N(u) = \left(\frac{1-au}{1-a}\right)^{-\frac{a+b}{a}}$$

with the limiting case $a \to 0$:

$$\psi_N(u) = e^{-b(1-u)}$$

Let us note that the result in the particular case $a = 0$, i.e. the Poisson case, was discovered by Adelson (1966) and Kemp (1967). In Kemp's paper, the probability generating function of the compound distribution is given by

$$\begin{aligned} \psi_S(u) &= exp(-\lambda(1-\psi_X(u))) \\ &= exp(\sum_{i=1}^{\infty} a_i(u^i-1)) \end{aligned}$$

Kemp (1967) derives the following algorithm :

$$\begin{aligned} f_S(0) &= exp(-\sum_{i=1}^{\infty} a_i) \\ (x+1)f_S(x+1) &= \sum_{i=0}^{x}(i+1)a_{i+1}f_S(x-i) \quad , \quad x \geq 0 \end{aligned}$$

which is clearly the same as the Panjer's algorithm in the case $a = 0$. However Kemp's algorithm does not clearly reveal the probability function of X.
Note that an interesting advantage of Kemp's notations is that they lead to a recursive

algorithm for the probability function of the Hermite distribution of which the probability generating function is

$$\psi(u) = e^{a_1(u-1)+a_2(u^2-1)}$$

The typical way of proving the Panjer's algorithm is to write a relation between the probability generating function of S and its first derivative. This is the classical method of using ordinary generating functions that will be used in the sequel. There also exists another proof, more intuitive, based on conditional expectations (see Panjer (1981)).

The Panjer's class is often called the $(a, b, 0)$ class.
Sundt and Jewell (1981) have extended the Panjer's class to a more general class of counting distributions, the (a, b, m) class.
The (a, b, m) class is obtained by the random variables satisfying the relation

$$p(n) = (a + \frac{b}{n})p(n-1) \quad , \quad n > m \quad , \quad m = 0, 1, 2, \ldots \tag{1.4}$$

If N belongs to the (a, b, m) class of counting distributions, the following algorithm was proved by Sundt and Jewell (1981) :

$$\begin{aligned} f_S(0) &= \psi_N(f_X(0)) \\ f_S(x) &= \frac{1}{1 - af_X(0)} \left[\sum_{i=1}^{x} (a + b\frac{i}{x}) f_X(i) f_S(x-i) \right. \\ &\qquad \left. + \sum_{n=1}^{m} (p(n) - (a + \frac{b}{n})p(n-1)) f_X^{*n}(x) \right] \quad , \quad x \geq 1 \end{aligned}$$

We immediately see that the Panjer's class is the particular case with $m = 0$. This is the reason why the Panjers' class is aften called the $(a, b, 0)$ class.
The members of the $(a, b, 1)$ class have received a name in the literature. The reason is that the $(a, b, 1)$ class is a class of counting variables which is zero modified. For $m \geq 2$, the members are more abstract. Note however that we will use the case $m = 2$ in section 5.1.9.
With $m = 1$ the members of the zero truncated $(a, b, 1)$ class are

- Zero-Truncated Poisson : $a = 0$, $b = \lambda$
- Zero-Truncated Binomial : $a = -\frac{q}{1-q}$, $b = (N + 1)\frac{q}{1-q}$
- Zero-Truncated Negative Binomial : $a = \frac{\beta}{1+\beta}$, $b = (r - 1)\beta \quad , \quad r > 0$
- Extended Zero-Truncated Negative Binomial : $a = \frac{\beta}{1+\beta}$, $b = (r - 1)\beta \quad , \quad -1 < r < 0$
- No name : $a = 1$, $b = -(\alpha + 1)$, $0 < \alpha < 1$

The last distribution did not receive a name because, as mentionned in Panjer and Willmot (1992), this distribution has an infinite mean and is of no use in insurance problems. We will not refer anymore to this distribution.
We can obviously extend the class of distributions to those which are Zero-Modified.
The probability generating function writes

$$\psi(u) = \left(1 - \frac{1 - p^{[1]}(0)}{1 - p(0)}\right) + \frac{1 - p^{[1]}(0)}{1 - p(0)} \left(\frac{1 - au}{1 - a}\right)^{-\frac{a+b}{a}}$$

with $p(0)$ being the probability of 0 event under the $(a,b,0)$ class and $p^{[1]}(0)$ being the probability of 0 event under the $(a,b,1)$ class. $p(0)$ is a function of a and b while $p^{[1]}(0)$ is a free parameter.
The initialization of (1.4) with $m=1$ is easily given by

$$p^{[1]}(1)=\frac{1-p^{[1]}(0)}{1-p(0)}p(1)$$

with $p^{[1]}(1)$ and $p(1)$ being the respective probabilities of one event in the $(a,b,1)$ class and $(a,b,0)$ class.
For example with the Zero-Modified Extended Truncated Negative Binomial with $p^{[1]}(0)=0.25$, $r=-0.4$ and $\beta=5$ we find

$$\begin{aligned} p^{[1]}(1) &= \frac{1-0.25}{1-p(0)}p(1) \\ &= \frac{1-0.25}{1-(1+5)^{0.4}}(-0.4)5\frac{1}{(1+5)^{1-0.4}} \\ &= 0.48862 \end{aligned}$$

1.4 Other recursions

It is not the aim of this chapter to make an exhaustive list of the recursions that have been derived in actuarial sciences. There exists a lot of other types of recursions for more complicated classes of random variables. See for example Hesselager (1994) who studied the class of random variables satisfying the recursion

$$\frac{p(n)}{p(n-1)}=\frac{\sum\limits_{i=0}^{k}a_i n^i}{\sum\limits_{i=0}^{k}b_i n^i} \quad , \quad n\geq 1$$

As mentionned by Hesselager (1994), particular members of this class are the Hypergeometric distribution, the Polya-Eggenberg distribution, the Waring distribution and the Generalized Waring distribution.
Sundt (1992) studied the class of random variables satisfying the relation

$$p(n)=\sum_{i=1}^{k}\left(a_i+\frac{b_i}{n}\right)p(n-i) \quad , \quad n\geq 1$$

This class of distributions does not lead to practical particular cases apart from those already known from the $(a,b,0)$ class.
Ambagaspitya and Balakrishnan (1994) derived a recursion for the compound Generalized Poisson Distribution. The Generalized Poisson Distribution will be briefly discussed in chapter 5.
In the following chapters we will use the properties of the (a,b,m) class to construct more elaborate recursions for more general counting random variables as well as their compound distributions. We will also extend some results in bivariate or multivariate settings.

1.5 Recursions, transforms, numerical integration and simulation

Regarding the problem of evaluating the probability function of the aggregate claims distribution, I will concentrate in this PhD thesis on the possible recursions, as introduced in this chapter. Moreover I will assume that the severity distribution is of lattice type. Some comments regarding the possible discretization of a continuous distribution will be given in chapter 2.

However, there exist other ways to solve the problem : the Fast Fourier Transform, the numerical integration of Volterra equations and the use of simulation techniques. I just mention these techniques while it has not been my aim to make comparisons within these techniques. Obviously this is a topic for future research.

1.5.1 The Fast Fourier Transform

The Fast Fourier Transform (FFT) is a classical tool when one has to deal with convolutions. The methodology is the following :

- Find the Fast Fourier Transform of a sequence $a_j, j \in \mathbb{Z}$
- Make useful calculations in the Fourier space
- Apply the Inverse Fast Fourier Transform (IFFT) in order to come back in the space of the sequences

See Klugman, Panjer and Willmot (1998) for a short description and Robertson (1992) for a detailed description with applications to aggregate loss calculation.
In the particular case of aggregate loss calculation the procedure is (see Klugman, Panjer and Willmot (1998)) :

- Discretize the claims distribution (see chapter 2 for methods) and obtain the sequence

$$f_X(0), f_X(1), \ldots, f_X(n-1)$$

  with $n = 2^r$ (this is the fact why we call the method FFT and not FT ; it allows for a reduced computing time which is $O(n \log_2 n)$) for some integer r and n the maximal value we need in $f_S(x)$.
- Obtain the FFT of the discretized claims distribution. The FFT transform is a sequence $\hat{f}_X$
- Transform the obtained vector into the FT of the aggregate claims distribution by using $\hat{f}_S = \psi_N(\hat{f}_X)$
- Apply the IFFT

Bühlmann (1984) compared the computing time of the Panjer's recursion for a Compound Poisson Distribution and the FFT. Defining n the maximal value for S we want to obtain and m the maximal value of X, Bühlmann (1984) obtained the following criterion (based on the number of complex multiplications) :

If

$$\log_2 n + \log_2(m+1) + \frac{m(m+1)}{8n} > \frac{m}{4}$$

Then recursion beats FFT

Bühlmann (1984) concluded that if $m \geq 255$ the FFT practically always beats the recursion whereas if $m \leq 63$ the contrary holds.
It should be noted that, if the claims distribution has been truncated at n, the recursion method will provide exact results for $f_S(0), \ldots, f_S(n-1)$ whereas this is not the case for the transform method as pointed out by Embrechts et al. (1993) where they use the terminology "wraparound effect". This is especially important if X is heavy-tailed. Grübel and Hermesmeier (1999) use the term aliasing error and show how to bound the aliasing error as well as how to eliminate it in practical purposes.
Let us note that the FFT method allows to work with any kind of counting distribution. It also allows to work with random variables having possible negative values.

1.5.2 Numerical integration

Panjer (1981) showed that in the case where the claims severity is a real random variable, the Panjer's algorithm write :

$$f_S(s) = p_1 f_X(s) + \int_0^s (a + b\frac{x}{s}) f_X(x) f_S(s-x) dx \quad , \quad s > 0$$

This is a Volterra integral equation of the second kind.
Appendix D in Panjer and Willmot (1992) discussed some numerical methods giving a solution to this kind of equation.
As far as I know, very few papers discuss this method of finding the density function of the aggregate claims distribution S. Ströter (1985) introduced the single-step and multi-step methods of Baker (1977) in the actuarial context. An error bound is provided but we have no idea of the computing time associated with these numerical methods.
Based on cubic splines, den Isegeer et al. (1997) proposed a method the accuracy of which is better than that obtained when the claims distribution is discretized. Moreover they give an idea of the computing time of their method.
It should be recalled that very few papers discuss numerical techniques solving the Volterra equations and comparing computing time and accuracy with other methods.

1.5.3 Simulation

Obviously it is always possible to resort to simulation in order to find the cumulative density function of the aggregate claims distribution. We just have to simulate values for N and conditionaly values for X. It then remains to compute the simulated values of S. Using the empirical cumulative distribution function based on the simulated values of S we are able to approximate $F_S(s)$. Moreover this methodology authorizes to work with a more complicated model than the classical Compound model, e.g. by introducing dependencies in the model. Nevertheless one should be extremely cautious with the behaviour of the right tail of the simulated distribution. In some cases it is important to have a good approximation of the tail and this may probably require a huge amount of generated values.

Chapter 2

A global way to discretize a distribution

The present chapter is based on Walhin and Paris (1998).
For the recursions described in chapter 1, the computing time will be significantly reduced if the claim amount distribution is equispaced because in this case working on $0, h, 2h, \ldots$ is the same as working on $0, 1, 2, \ldots$.
This is the reason why we look in this chapter for some methods used to find equispaced discrete distributions from any distribution (continuous or discrete).
The following hypothetical claim severity distribution will be used for illustration :

X	0	7	12	17	21	23	28	39	46	53	67
f_X	0.05	0.1	0.1	0.15	0.05	0.05	0.05	0.1	0.1	0.15	0.1

Table 2.1: Claim severity distribution

We first give a review of the existing methods.

2.1 The rounding method

Gerber and Jones (1976) proposed a very intuitive rounding method. Suppose we want to transform the claim amount distribution X into an equispaced distribution on $x = 0, h, 2h, \ldots$.
We will denote the equispaced distribution X_{app} because it is an approximation of the original distribution.
The probabilities are simply given by

$$\begin{aligned} f_{X_{app}}(0) &= F_X(\frac{h}{2} - 0) \\ f_{X_{app}}(xh) &= F_X(xh + \frac{h}{2} - 0) - F_X(xh - \frac{h}{2} - 0) \quad , \quad x = 1, 2, \ldots \end{aligned}$$

Assume we choose $h = 20$. We find the approximate equispaced distribution (X_{app}) :

X_{app}	0	20	40	60
$f_{X_{app}}$	0.15	0.40	0.20	0.25

Table 2.2: Approximate distribution; Rounding, $h = 20$

The rounding method has the disadvantage that the approximate distribution does not preserve any moment of the exact distribution.
Panjer and Willmot (1992) propose two other types of rounding :

1. Rounding to lower unit :

$$\begin{aligned} f_{X_{low}}(0) &= F_X(h-0) \\ f_{X_{low}}(xh) &= F_X(xh+h-0) - F_X(xh-0) \quad , \quad x = 1, 2, \ldots \end{aligned}$$

2. Rounding to upper unit

$$\begin{aligned} f_{X_{up}}(0) &= 0 \\ f_{X_{up}}(xh) &= F_X(xh+0) - F_X(xh-h+0) \quad , \quad x = 1, 2, \ldots \end{aligned}$$

These types of rounding are interesting because they provide upper and lower bounds in the sense of the stochastic order (see section 4.3.4).

2.2 The local moment matching method

Gerber (1982) proposed a method that matches some moments : the local moment matching method (LMM).
Suppose for example we want to match two moments. Let the interval $(x_k, x_k + 2h]$. The masses m_0^k, m_1^k, m_2^k are associated to the points x_k, x_{k+h}, x_{k+2h} in order to preserve the first two moments :

$$\sum_{j=0}^{2} (x_k + jh)^r m_j^k = \int_{x_k}^{x_k+2h} x^r dF_X(x) \quad , \quad r = 0, 1, 2 \tag{2.1}$$

Using the Lagrange formula, Gerber (1982) gives the solution to the system (2.1) :

$$m_j^k = \int_{x_k}^{x_k+2h} \prod_{i \neq j} \frac{x - x_k - ih}{(j-i)h} dF_X(x) \quad , \quad j = 0, 1, 2$$

In our numerical example the integral is a sum.
The method might be applied in order to match more than two moments.

By taking $h = 20$, we find

m_0^0	m_1^0	m_2^0	m_0^1	m_1^1	m_2^1
0.1318	0.4389	0.0793	0.0836	0.2704	−0.0040

Table 2.3: Intermediate masses; LMM

and thus the following approximate distribution :

X	0	20	40	60	80
$f_{X_{app}}$	0.1318	0.4389	0.1629	0.2704	−0.0040

Table 2.4: Approximate distribution; LMM, $h = 20$

This example shows an important disadvantage of the local moment matching method : the probabilities may take negative values.
With $h = 17$, we find

X	0	17	34	51	68
$f_{X_{app}}$	0.0998	0.4268	0.0921	0.3009	0.0804

Table 2.5: Approximate distribution; LMM, $h = 17$

2.3 Minimizing the Kolmogorov distance

The methods presented in the sections 2.1 and 2.2 are essentially local and they present some important disadvantages. In the sequel we will propose a global method consisting in the minimization of a distance between the exact distribution and the approximate distribution. Having chosen the equispaced mass points of the approximate distribution, we minimize the Kolmogorov distance in order to find the probabilities associated with the equispaced points. The Kolmogorov distance between two random variables is given by

$$d_K(X, Y) = \max_x |F_X(x) - F_Y(x)| \tag{2.2}$$

This distance is used for the Kolmogorov-Smirnov adjustment test between two distributions. A small d_K is in favour of the null hypothesis that the two distributions are the same whereas a large value of d_K is in favour of the alternative hypothesis that the two distributions are different.
Let us assume the same claim amount distribution as in the previous section. We want to approximate it by a distribution with masses at $0, 20, 40, 60, 80$ i.e. such that $h = 20$. The problem is how to find $f_{X_{app}}(0), \ldots, f_{X_{app}}(80)$ in order to minimize $d_K(X, X_{app})$. This is easily done numerically.

We find

X_{app}	0	20	40	60	80
$f_{X_{app}}$	0.2250	0.2976	0.2976	0.0899	0.0899

$$d_K(X, X_{app}) = 0.175$$

Table 2.6: Approximate distribution ; Kolmogorov, $h = 20$

Other approximate distributions might be obtained with the same Kolmogorov distance. This suggests that we might minimize d_K under the constraint that the first moment of X is preserved. We then find

X_{app}	0	20	40	60	80
$f_{X_{app}}$	0.225	0.2960	0.2604	0.1312	0.0874

$$d_K(X, X_{app}) = 0.175$$

Table 2.7: Approximate distribution ; Kolmogorov, $h = 20$, first moment constrained

If the minimization is proceeded with the constraint of the first two moments, we find

X_{app}	0	20	40	60	80
$f_{X_{app}}$	0.1833	0.25	0.4206	0.1156	0.0305

$$d_K(X, X_{app}) = 0.2167$$

Table 2.8: Approximate distribution ; Kolmogorov, $h = 20$, first two moments constrained

We see that in this case the Kolmogorov distance is higher with two constrained moments than with only one. The cost of matching several moments is such that the d_K is higher.
The preceding considerations suggest the following algorithm when using the Kolmogorov distance for the construction of an equispaced distribution :

1. Choose a set of equispaced points.
2. m=0
3. Minimize $d_K(X, X_{app})$ with m moments constrained. Let $d_K(m) = d_K(X, X_{app})$.
4. Minimize $d_K(X, X_{app})$ with $m+1$ moments constrained. Let $d_K(m+1) = d_K(X, X_{app})$.
5. If $d_K(m) = d_K(m+1)$ then $m = m+1$ and go to 3.
   Else keep the approximate distribution given by the minimization of $d_K(m)$.
6. Note that the necessity of keeping n moments may overrule the rule 4.
7. If the distance cannot be accepted then go back to 1 and change the equispaced points.

2.4 The effect of the approximation on the compound distribution

Now let us have a look at the Kolmogorov distance between the compound distribution ($f_S(x)$) and the approximate compound distribution ($f_{S_{app}}(x)$).
Suppose that N is Poisson distributed with mean $\lambda = 0.1$.
The following distances have been calculated

	$d_K(X, X_{app})$	$d_K(S, S_{app})$
$h = 10; 4$ moments	0.125	0.0115
$h = 10; 5$ moments	0.125	0.0115
$h = 10; 6$ moments	0.1273	0.0117
$h = 17; LMM$	0.1696	0.0157
$h = 17; 2$ moments	0.1395	0.0131
$h = 20; Rounding$	0.25	0.0228
$h = 20; 1$ moment	0.175	0.0161
$h = 20; 2$ moments	0.2167	0.0202
$h = 20; 3$ moments	0.2311	0.0212
$h = 25; 1$ moments	0.225	0.0207
$h = 25; 2$ moments	0.2646	0.0242

Table 2.9: Kolmogorov distances with $\lambda = 0.1$

We see that the $d_K(X, X_{app})$ increases with h. This is not a surprise because, at the limit $h = 0$ we would trivially have $d_K(X, X_{app}) = 0$.
Concerning $d_K(S, S_{app})$, the results are excellent. In each case, it is less than 10% of $d_K(X, X_{app})$.
Obviously the parameters of the claim frequency distribution must play a role (see Theorem 2.1). In our example, there is only one parameter : λ. Intuitively, we expect the $d_K(S, S_{app})$ to grow with λ. The following calculations show it is not necessarily the case.

We choose $h = 10$ and 5 moments constrained ($d_K(X, X_{app}) = 0.125$). We find

λ	$d_K(S, S_{app})$
0.05	0.005992
0.10	0.011488
0.20	0.021113
1	0.053471
2	0.048496
3	0.036778
4	0.027632
5	0.023072
20	0.010845

Table 2.10: Kolmogorov distances; $h = 10$, first five moments constrained

As this example shows, the $d_K(S, S_{app})$ grows until around $\lambda = 1$ and then decreases. In this example, the exact distribution X and the approximate distribution X_{app} are so closed that high order convolutions of the distributions remain very close. This explains why the Kolmogorov distance between the compound distributions decreases when λ becomes very large.
Obviously this is not the general rule. The following example shows that the Kolmogorov distance between the compound distributions increases with λ when the approximation is bad :

X	0	2	4
f_X	0.4	0.2	0.4

X_{app}	0	3
$f_{X_{app}}$	0.3	0.7

$$d_K(X, X_{app}) = 0.4$$

λ	$d_K(S, S_{app})$
0.1	0.037062
1	0.185621
10	0.126143
50	0.147345
100	0.180262
500	0.344425
1000	0.464542

Table 2.11: Kolmogorov distances; pathological example

Note that the discretization error was studied by Embrechts et al. (1993) and Grübel and Hermesmeier (2000). In these papers it is shown that this discretization error can be reduced by using the so-called Richardson's deferred approach to the limit which is a numerical technique borrowed from numerical analysis (see Johnson and Riess (1977)).

2.5 Upper and lower bounds for the Kolmogorov distance

With formula (1.3), we easily find an upper bound for the Kolmogorov distance between the exact distribution and the approximate distribution.

Theorem 2.1

$$d_K(S, S_{app}) \leq \mathbb{E}N d_K(X, X_{app})$$

Proof

$$\begin{aligned}
d_K(S, S_{app}) &= \max_x |F_S(x) - F_{S_{app}}(x)| \\
&= \max_x \left| \sum_{n=0}^{\infty} p(n)(F_X^{*^n}(x) - F_{X_{app}}^{*^n}(x)) \right| \\
&\leq \sum_{n=0}^{\infty} p(n) \max_x |F_X^{*^n}(x) - F_{X_{app}}^{*^n}(x)| \\
&= \sum_{n=0}^{\infty} p(n) d_K(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_{app_i}) \\
&\leq \sum_{n=0}^{\infty} p(n) \sum_{i=1}^{n} d_K(X_i, X_{app_i}) \\
&= \sum_{n=0}^{\infty} n p(n) d_K(X, X_{app}) \\
&= \mathbb{E}N d_K(X, X_{app}) \qquad (2.3)
\end{aligned}$$

■

A trivial lower bound is given by

$$d_K(S, S_{app}) \geq |\psi_N(f_X(0)) - \psi_N(f_{X_{app}}(0))|$$

Obviously $d_K(X^{*^n}, X_{app}^{*^n})$ is always ≤ 1. Therefore the upper bound can be sharpened :

$$\begin{aligned}
d_K(S, S_{app}) \leq\ & \mathbb{E}N d_K(X, X_{app}) \\
& - \sum_{n=2}^{\infty} p(n) \max(0, 1 - n d_K(X, X_{app}))
\end{aligned}$$

Let us take the example of table 2.10 :

λ	$d_K(S, S_{app})$	lower bound	upper bound	sharpened upper bound
0.1	0.011488	0.000712	0.0125	0.0125
0.2	0.021113	0.001297	0.0250	0.025
1	0.053471	0.003042	0.125	0.124999

Table 2.12: Comparison between the Kolmogorov distance and the bounds; $h = 10$

Even when λ (or $\mathbb{E}N$ in a more general setting) is high, a precise bound can be obtained. There will be a cost for this operation : the calculation of some convolutions of X and X_{app}.
Define $\Gamma_{\mathbb{E}N,a,b} = \{\lfloor \mathbb{E}N \rfloor - a; \cdots ; \lfloor \mathbb{E}N \rfloor + b\}$
We have

$$\begin{aligned} d_K(S, S_{app}) \quad &\leq \quad \max_x \left| \sum_{j \in \Gamma_{\mathbb{E},a,b}} p(j)(F_X^{*j}(x) - F_{X_{app}}^{*j}(x)) \right| \\ &\quad + \sum_{j \in (\mathbb{Z}^+ \setminus \Gamma_{\mathbb{E},a,b})} p(j) \min(1, j d_K(X, X_{app})) \end{aligned}$$

$d_K(S, S_{app})$	0.053471
twice sharpened upper bound $\Gamma = \{1, 2\}$	0.086197
twice sharpened upper bound $\Gamma = \{1, 3\}$	0.063503
twice sharpened upper bound $\Gamma = \{1, 4\}$	0.055846

Table 2.13: Comparison between the Kolmogorov distance and the bounds; $h = 10$, $\lambda = 1$

2.6 Berry-Esseen type results

In the previous sections, we noticed that whenever $\mathbb{E}N$ is high, the upper bound (2.3) is useless, even if, in practice, the approximation based on the Kolmogorov distance remains a relevant one. This fact, not really surprising, can be confirmed by a Berry-Esseen result.
Michel (1993) gave a Berry-Esseen bound for the normal approximation of a compound Poisson distribution.

Theorem 2.2 *Let S be a compound Poisson distribution. Assume the the skewness of S is finite ($\gamma_S < \infty$). We have*

$$\sup_{t \in \mathbb{R}} \left| \mathbb{P}\left(\frac{S - \mathbb{E}S}{\sigma_S} \leq t \right) - \Phi(t) \right| \leq 0,8\gamma_S$$

where $\Phi(t)$ *is the cumulative distribution of a standard normal distribution*

$\gamma_S = \frac{\mathbb{E}S^3 - 3\mathbb{E}S\mathbb{E}S^2 + 2(\mathbb{E}S)^3}{\sigma_S^3}$ *is the skewness of S*

For our numerical example, whenever N is a Poisson random variable with mean λ, we obtain the following upper bound for the difference between the distribution of S and the normal approximation :

λ	bound
$0,1$	$3,4965$
1	$1,1057$
10	$0,3496$
100	$0,1105$
1000	$0,0349$

Table 2.14: Berry-Esseen bound in function of λ

The result is not surprising : whenever the mean of the counting distribution is low the approximation by the normal distribution is bad but in that case the method described in the previous section can be applied. Whenever this mean is high, the central limit approximation is correct. Nevertheless, the techniques described in the previous sections remain interesting if you want an approximation of S by a discrete random variable S_{app}.
Assume the respective skewness are γ and γ_{app}. If X is a $N(\mathbb{E}S, \mathbb{V}arS)$ random variable, then by the Berry-Esseen result of Michel and the triangular inequality, we obtain

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(S < t) - \mathbb{P}(S_{app} < t)| \leq 0,8\gamma + 0,8\gamma_{app} + \sup_{t\in\mathbb{R}} |\mathbb{P}[X < t] - \mathbb{P}[Y < t]| \tag{2.4}$$

where $X \sim N(\mathbb{E}S; \mathbb{V}arS)$
$Y \sim N(\mathbb{E}S_{app}; \mathbb{V}arS_{app})$
In particular, when S_{app} is constructed in such a way that the first two moments of X and Y are the same, (2.4) becomes

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(S < t) - \mathbb{P}(S_{app} < t)| \leq 0,8\gamma + 0,8\gamma_{app}$$

Notes :

1) Whenever the mean of N is very large, a technical problem arises : the initialization of the recursion gives an underflow. This problem was treated by Wang and Panjer (1994) but the technique they propose can not be controlled if one wants an arithmetic distribution. However if a normal approximation is sufficient, then it is very well controlled by Michel's (1993) result.

2) Michel (1993) also gives a non-uniform result under the same hypothesis :

$$\left|\mathbb{P}\left(\frac{S - \mathbb{E}S}{\sigma_S} \leq t\right) - \Phi(t)\right| \leq \frac{30,6}{1 + |t|^3}\gamma_S$$

Unfortunately this result is only of more interest than the uniform bound if $|t| \geq 3,34$, i.e. in the very far tails of the distribution. This is the reason why we concentrate only on the uniform result. Obviously the non-uniform result of Michel is interesting to use when approximating quantities like stop-loss premiums for example.

Let us point out that the proof of Michel's (1993) result is given for the case when N, the number of components of the random sum, is a Poisson random variable but as the proof shows, the crucial point for the extension of the Berry-Esseen result to the random case is the fact that the distribution of S is infinitely divisible. Consequently, N has to be infinitely divisible.
In fact, in Feller (1971) (see also section 5.1.1), there is a logical equivalence between infinite divisibility on integers and Compound Poisson distributions.
Then the result of Michel (1993) is immediately extended to every infinite divisible distribution on integers.

Chapter 3

Multivariate extension of the individual and collective risk models

The motivation for looking at multivariate extensions of the individual and collective risk models results from the fact that practical excess of loss treaties have specific clauses on the market. These clauses, that will be described in section 4.1, imply that the aggregate claims distribution of the ceding company depends on its own retention as well as on the aggregate claims distribution of the reinsurers. This accounts for the necessity of multivariate extensions of the known results in the univariate case.
We will see in chapter 6 that the bivariate Panjer's algorithm has another application when looking for the compound distribution where the counting distribution is a bivariate distribution.
The present chapter is mainly based on Walhin and Paris (2000i) and Walhin and Paris (2000c). Note that the multivariate Panjer's algorithm was presented by the author at the 1997 meeting of the Belgian Statistical Society. It was however published independently by two other authors, namely Sundt (1999b) and Ambagaspitya (1999).
We first introduce the concept of ordinary generating functions that will be used throughout this chapter as well as in chapter 6.
Note that we will write, in this chapter, the formulae in a bivariate setting for a pedagogical purpose. The transition to the multivariate setting is always immediate.

3.1 Ordinary generating functions

In the following sections we will use the concept of bivariate ordinary generating functions. This is a generalization of the ordinary generating function (see Panjer and Willmot (1992) for an application in actuarial sciences). Multivariate ordinary generating functions are trivially extended.
Let us assume a sequence $\{a_{n,m}\,,\, n \in \mathbb{N}, m \in \mathbb{N}\}$ of real numbers.
The ordinary generating function of this sequence is defined as

$$T_{a_{n,m}}(u_1, u_2) = \sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{n,m} u_1^n u_2^m$$

Obviously u_1 and u_2 must be chosen in such a way that the sum exists.
Ordinary generating functions have the following nice properties :

- There is a one-to-one correspondence between $\{a_{n,m}\ ,\ n \in \mathbb{N}, m \in \mathbb{N}\}$ and $T_{a_{n,m}}(u_1,u_2)$
- $a_{n,m} = \frac{1}{n!m!}\frac{\partial^n\partial^m T_{a_{n,m}}(u_1,u_2)}{\partial u_1^n \partial u_2^m}\Big|_{u_1=0,u_2=0}$
- $c_{n,m} = \alpha a_{n,m} + \beta b_{n,m} \Leftrightarrow T_{c_{n,m}}(u_1,u_2) = \alpha T_{a_{n,m}}(u_1,u_2) + \beta T_{b_{n,m}}(u_1,u_2)$
- $c_{n,m} = \sum\limits_{k=0}^{n}\sum\limits_{l=0}^{m} a_{k,l} b_{n-k,m-l} \Leftrightarrow T_{c_{n,m}}(u_1,u_2) = T_{a_{n,m}}(u_1,u_2) T_{b_{n,m}}(u_1,u_2)$
- $T_{na_{n,m}}(u_1,u_2) = u_1 \frac{\partial}{\partial u_1} T_{a_n}(u_1,u_2)$

The philosophy behind using ordinary generating functions is the following :

- look for a relation between some sequences $a_{n,m}, b_{n,m}, c_{n,m}, \ldots$
- go into the (u_1,u_2) map where the calculations become easier (think of the convolution that becomes a product)
- go back into the initial map by inverting the expression in (u_1,u_2) thanks to the properties

In this thesis, the sequences $a_{n,m}$ will be probability functions. As a consequence we will not have any problems of convergence for the ordinary generating functions : $|u_1| < \infty$, $|u_2| < \infty$. In the present case, ordinary generating functions are just probability generating functions. ¿From now on we will only refer to probability generating functions and we will use them extensively in the sense of ordinary generating functions.

3.2 The multivariate individual risk model

Recently, Dickson and Waters (1999) and Sundt (1999a) extended De Pril's approximation to the multivariate setting. The main contribution of this section is to extend the result of Dhaene and Vandebroek (1995) to a multivariate setting and to compare the different results.

3.2.1 Notations

¿From now on we will assume that the random variable giving the claim amount is bivariate. We will use the following definitions :

(X_{1i}, X_{2i})	:	random vector of claim amount for risk of type i $\forall j$
$g_i(x_1,x_2)$	:	probability function of (X_{1i}, X_{2i})
(m_{1i}, m_{2i})	:	maximum claim amounts for risk of type i $\forall j$
$\psi_i(u_1,u_2)$	:	probability generating function of (X_{1i}, X_{2i})
(S_1,S_2)	:	random vector of the aggregate claim amounts
$f_{S_1,S_2}(s_1,s_2)$	:	probability function of (S_1,S_2)
$\psi_{S_1,S_2}(u_1,u_2)$	:	probability generating function of (S_1,S_2)
q_j	:	probability of no claim for the risk of type j $\forall i$
$w(x) = O(x)$	:	$w(x)$ is of the same order as x

3.2.2 The multivariate approximate De Pril formula

In this section we will extend the methodology developed in Dhaene and De Pril (1994). We have

$$\begin{aligned}\psi_{S_1,S_2}(u_1,u_2) &= \sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty} f_{S_1,S_2}(s_1,s_2)u_1^{s_1}u_2^{s_2}\\ \ln\psi_{S_1,S_2}(u_1,u_2) &= \sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty} t(x_1,x_2)u_1^{x_1}u_2^{x_2} \qquad (3.1)\end{aligned}$$

Note that as $f_{S_1,S_2}(0,0) > 0$, (3.1) is well defined in the neighbourhood of $(0,0)$.
As we immediately have

$$\begin{aligned}u_1\frac{\partial}{\partial u_1}\psi_{S_1,S_2}(u_1,u_2) &= \psi_{S_1,S_2}(u_1,u_2)u_1\frac{\partial}{\partial u_1}\ln\psi_{S_1,S_2}(u_1,u_2)\\ u_2\frac{\partial}{\partial u_2}\psi_{S_1,S_2}(u_1,u_2) &= \psi_{S_1,S_2}(u_1,u_2)u_2\frac{\partial}{\partial u_2}\ln\psi_{S_1,S_2}(u_1,u_2)\end{aligned}$$

we get, inverting the preceding expressions

$$\begin{aligned}f_{S_1,S_2}(0,0) &= e^{t(0,0)}\\ s_1 f_{S_1,S_2}(s_1,s_2) &= \sum_{x_1=1}^{s_1}\sum_{x_2=0}^{s_2} x_1 t(x_1,x_2) f_{S_1,S_2}(s_1-x_1,s_2-x_2) \quad , \quad s_1\geq 1 \qquad (3.2)\\ s_2 f_{S_1,S_2}(s_1,s_2) &= \sum_{x_1=0}^{s_1}\sum_{x_2=1}^{s_2} x_2 t(x_1,x_2) f_{S_1,S_2}(s_1-x_1,s_2-x_2) \quad , \quad s_2\geq 1 \qquad (3.3)\end{aligned}$$

Note that when $s_1 \geq 1$ and $s_2 \geq 1$ both recursions (3.2) and (3.3) are valid.
In the particular case of the individual risk model, we have

$$\psi_{S_1,S_2}(u_1,u_2) = \prod_{i=1}^{a}\prod_{j=1}^{b}[p_j + q_j\psi_i(u_1,u_2)]^{n_{ij}}$$

A Taylor expansion of $\ln\psi_{S_1,S_2}(u_1,u_2)$ gives

$$\ln\psi_{S_1,S_2}(u_1,u_2) = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}\ln p_j + \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\left(\frac{q_i}{p_i}\right)^k \psi_i^k(u_1,u_2)$$

Inverting this expression gives

$$\begin{aligned}t(0,0) &= \sum_{i=1}^{n}\ln p_i\\ t(x_1,x_2) &= \sum_{i=1}^{a}\sum_{k=1}^{\max(x_1,x_2)}\sum_{j=1}^{b} n_{ij}\frac{(-1)^{k+1}}{k}\left(\frac{q_j}{p_j}\right)^k g_i^{*k}(x_1,x_2)\end{aligned}$$

Calculating the $t(x_1,x_2)$ in a first stage and inserting these quantities in the bivariate recursion (3.2),(3.3) in a second stage will give us an exact algorithm.

As explained in section 1.1.3, we will always find the $t(x_1, x_2)$ in a first stage before inserting these values in the bivariate recursion. We certainly do not try to find a one stage bivariate recursion.
However this two stage bivariate recursion is $O(s_1^2 s_2^2)$ and would be very time consuming. Nevertheless there is a trivial approximation for it which will be $O(s_1 s_2)$ consisting in truncating the sum in k to a certain integer r which will be very small most of the time.
Let $f^{(r)}_{S_1,S_2}(s_1, s_2)$ be the r^{th} order approximation for $f_{S_1,S_2}(s_1, s_2)$.
$t^{(r)}(x_1, x_2)$ be the r^{th} order approximation for $t(x_1, x_2)$.
We have

$$
\begin{aligned}
f^{(r)}_{S_1,S_2}(0,0) &= e^{t^{(r)}(0,0)} \\
s_1 f^{(r)}_{S_1,S_2}(s_1, s_2) &= \sum_{x_1=1}^{s_1} \sum_{x_2=0}^{s_2} x_1 t^{(r)}(x_1, x_2) f^{(r)}_{S_1,S_2}(s_1 - x_1, s_2 - x_2) \quad , \quad s_1 \geq 1 \\
s_2 f^{(r)}_{S_1,S_2}(s_1, s_2) &= \sum_{x_1=0}^{s_1} \sum_{x_2=1}^{s_2} x_2 t^{(r)}(x_1, x_2) f^{(r)}_{S_1,S_2}(s_1 - x_1, s_2 - x_2) \quad , \quad s_2 \geq 1 \\
t^{(r)}(0,0) &= \sum_{i=1}^{n} \ln p_i \qquad (3.4) \\
t^{(r)}(x_1, x_2) &= \sum_{i=1}^{a} \sum_{k=1}^{\min(r,\max(x_1,x_2))} \sum_{j=1}^{b} n_{ij} \frac{(-1)^{k+1}}{k} \left(\frac{q_j}{p_j}\right)^k g_i^{*^k}(x_1, x_2) \qquad (3.5)
\end{aligned}
$$

It is clear that our bivariate derivation trivially extends to a multivariate setting. As we will see in the next sections our algorithm written in the map of probability functions authorizes to have a look at the calculation time. This is not the case for the derivation of Dickson and Waters (1999) because (3.4)-(3.5) are related to the ordinary generating functions in their paper.
Note that the algorithm can also be discovered by using the multivariate De Pril transform as mentioned in Sundt (1999a).

Using the same arguments as Dhaene and De Pril (1994), it is easy to extend their theorem 1 (see for example Sundt (1998)):

Theorem 3.1 *If there exists a real number ϵ such that*

$$\sum_{x_1=0}^{\infty} \sum_{x_2=0}^{\infty} |t(x_1, x_2) - t^{(r)}(x_1, x_2)| \leq \epsilon \qquad (3.6)$$

then the following error bound holds

$$\sum_{s_1=0}^{\infty} \sum_{s_2=0}^{\infty} |f_{S_1,S_2}(s_1, s_2) - f^{(r)}_{S_1,S_2}(s_1, s_2)| \leq e^{\epsilon} - 1$$

For our bivariate De Pril's approximation, it is also easy to show that the bound for (3.6) remains the same as in the univariate case :

$$\sum_{x_1=0}^{\infty}\sum_{x_2=0}^{\infty}|t(x_1,x_2)-t^{(r)}(x_1,x_2)| \leq \frac{1}{r+1}\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}\frac{p_j}{p_j-q_j}\left(\frac{q_j}{p_j}\right)^{r+1}$$

So the error bound for the $f_{S_1,S_2}(s_1,s_2)$ is controlled in the same way as in the univariate case.

3.2.3 The multivariate exact Dhaene and Vandebroek formula

In this section we extend the exact formula given by Dhaene and Vandebroek (1995) to the bivariate setting.

Theorem 3.2 *With the notations introduced in section 3.2.1 , an exact solution for the individual risk model is given by*

$$\begin{aligned}
f_{S_1,S_2}(0,0) &= \prod_{i=1}^{a}\prod_{j=1}^{b} p_j^{n_{ij}} && (3.7)\\
s_1 f_{S_1,S_2}(s_1,s_2) &= \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij} v_{ij}(s_1,s_2) \quad , \quad s_1 \geq 1\,,\ s_2 \geq 0 && (3.8)\\
v_{ij}(s_1,s_2) &= \frac{q_j}{p_j}\sum_{x_1=0}^{m_{1i}}\sum_{x_2=0}^{m_{2i}} g_i(x_1,x_2)(x_1 f_{S_1,S_2}(s_1-x_1,s_2-x_2) && (3.9)\\
&\quad -v_{ij}(s_1-x_1,s_2-x_2)) \quad , \quad s_1 \geq 1\,,\ s_2 \geq 0\\
&= 0 \quad \textit{otherwise}\\
s_2 f_{S_1,S_2}(s_1,s_2) &= \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij} w_{ij}(s_1,s_2) \quad , \quad s_2 \geq 1\,,\ s_1 \geq 0 && (3.10)\\
w_{ij}(s_1,s_2) &= \frac{q_j}{p_j}\sum_{x_1=0}^{m_{1i}}\sum_{x_2=0}^{m_{2i}} g_i(x_1,x_2)(x_2 f_{S_1,S_2}(s_1-x_1,s_2-x_2) && (3.11)\\
&\quad -w_{ij}(s_1-x_1,s_2-x_2)) \quad , \quad s_2 \geq 1\,,\ s_1 \geq 0\\
&= 0 \quad \textit{otherwise}
\end{aligned}$$

Proof :

$$\begin{aligned}
\psi_{S_1,S_2}(u_1,u_2) &= \sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty} f_{S_1,S_2}(s_1,s_2)u_1^{s_1}u_2^{s_2}\\
&= \prod_{i=1}^{a}\prod_{j=1}^{b}[p_j+q_j\psi_i(u_1,u_2)]^{n_{ij}}
\end{aligned}$$

$$\frac{\partial}{\partial u_1}\psi_{S_1,S_2}(u_1,u_2) = \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}V_{ij}(u_1,u_2) \tag{3.12}$$

$$\begin{aligned} V_{ij}(u_1,u_2) &= \sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty} v_{ij}(s_1+1,s_2)u_1^{s_1}u_2^{s_2} \\ &= \frac{q_j\frac{\partial}{\partial u_1}\psi_i(u_1,u_2)\psi_{S_1,S_2}(u_1,u_2)}{p_j+q_j\psi_i(u_1,u_2)} \end{aligned} \tag{3.13}$$

$$\begin{aligned} \frac{\partial}{\partial u_2}\psi_{S_1,S_2}(u_1,u_2) &= \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}W_{ij}(u_1,u_2) \\ W_{ij}(u_1,u_2) &= \sum_{s_1=0}^{\infty}\sum_{s_2=0}^{\infty} w_{ij}(s_1,s_2+1)u_1^{s_1}u_2^{s_2} \\ &= \frac{q_j\frac{\partial}{\partial u_2}\psi_i(u_1,u_2)\psi_{S_1,S_2}(u_1,u_2)}{p_j+q_j\psi_i(u_1,u_2)} \end{aligned}$$

Multiplying both sides of (3.12) by u_1 and inverting gives (3.8).
Multiplying both sides of (3.13) by u_1 and rearranging gives

$$u_1V_{ij}(u_1,u_2) = \frac{q_j}{p_j}\left(u_1\frac{\partial}{\partial u_1}\psi_i(u_1,u_2)\psi_{S_1,S_2}(u_1,u_2) - \psi_i(u_1,u_2)u_1V_{ij}(u_1,u_2)\right)$$

which gives (3.9) after inverting.
(3.10) and (3.11) are obtained similarly. ■

The number of multiplications for the exact bivariate algorithm is

$$2b(\sum_{i=1}^{a} m_{1i}m_{2i} + a)s_1s_2$$

whereas the number of multiplications for the approximate bivariate algorithm is

$$(2r^2 \max_i m_{1i} \max_i m_{2i} + 1)s_1s_2$$

plus some initial work not depending on s_1, s_2 in order to construct the $t^{(r)}(x_1,x_2)$.
We can immediately see that the approximate formula reveal more efficient than the exact one if

$$r^2 < b\frac{\sum_{i=1}^{a} m_{1i}m_{2i}}{\max_i m_{1i}\max_i m_{2i}}$$

Note that in this bivariate case the fraction may be $>$ or $<$ than 1.

3.3 The multivariate collective risk model

We will use the following notation :

$$\sum_{x_1,\ldots,x_k}^{s_1,\ldots,s_k} w(x_1,\ldots,x_k) = \sum_{x_1=0}^{s_1}\ldots\sum_{x_k=0}^{s_k} w(x_1,\ldots,x_k) - w(0,\ldots 0)$$

3.3.1 The multivariate Panjer's algorithm

Let us point out that the multivariate extension of Panjer's recursion was discovered independently by Sundt (1999b), Ambagaspitya (1999) and the author of the present work. We give the proof of the theorem because it is based on probability generating functions whereas Sundt's proof is based on conditional expectations. Ambagaspitya's proof is also based on probability generating functions.

Theorem 3.3 *Let*

$$(S_1, S_2) = (\sum_{i=1}^{N} X_i, \sum_{i=1}^{N} Y_i)$$

with (X_i, Y_i) *independent and identically distributed, arithmetic and independent of* N
X_i *and* Y_i *are not necessarily independent*
N *belonging to Panjer's class, i.e. such that*

$$\frac{p(n)}{p(n-1)} = a + \frac{b}{n} \quad , \quad n > 0$$

Then, if $\psi_N(u)$ *denotes the probability generating function of* N*, we have*

$$f_{S_1,S_2}(0,0) = \psi_N(f_{X_1,X_2}(0,0)) \tag{3.14}$$

$$f_{S_1,S_2}(s_1,s_2) = \frac{1}{1 - af_{X_1,X_2}(0,0)} \sum_{x_1}^{s_1}\sum_{x_2}^{s_2}[a + b\frac{x_1}{s_1}] \times f_{S_1,S_2}(s_1-x_1, s_2-x_2) f_{X_1,X_2}(x_1,x_2) \quad , \quad s_1 \geq 1 \tag{3.15}$$

$$f_{S_1,S_2}(s_1,s_2) = \frac{1}{1 - af_{X_1,X_2}(0,0)} \sum_{x_1}^{s_1}\sum_{x_2}^{s_2}[a + b\frac{x_2}{s_2}] \times f_{S_1,S_2}(s_1-x_1, s_2-x_2) f_{X_1,X_2}(x_1,x_2) \quad , \quad s_2 \geq 1 \tag{3.16}$$

Proof
Let $\psi_{X_1,X_2}(u_1,u_2)$ and $\psi_{S_1,S_2}(u_1,u_2)$ be the probability generating functions of (X_1, X_2) and (S_1, S_2) respectively.
From

$$f_{S_1,S_2}(s_1,s_2) = \sum_{n=0}^{\infty} p(n) f^{*n}_{X_1,X_2}(s_1,s_2)$$

we get

$$\psi_{S_1,S_2}(u_1,u_2) = \sum_{n=0}^{\infty} p(n) \psi^{n}_{X_1,X_2}(u_1,u_2)$$

from which (3.14) follows immediately.
By hypothesis, one has

$$np(n) = a(n-1)p(n-1) + (a+b)p(n-1)$$

Let us multiply on both sides by $\psi^{n-1}_{X_1,X_2}(u_1,u_2)u_1\frac{\partial}{\partial u_1}\psi_{X_1,X_2}(u_1,u_2)$ and sum over $n = 1 \to \infty$, we get

$$u_1\frac{\partial}{\partial u_1}\psi_{S_1,S_2}(u_1,u_2) = a\psi_{X_1,X_2}(u_1,u_2)u_1\frac{\partial}{\partial u_1}\psi_{S_1,S_2}(u_1,u_2) + (a+b)u_1\frac{\partial}{\partial u_1}\psi_{X_1,X_2}(u_1,u_2)\psi_{S_1,S_2}(u_1,u_2)$$

which gives, inverting in the initial space

$$\begin{aligned} s_1 f_{S_1,S_2}(s_1 s_2) &= a \sum_{x_1=0}^{s_1} \sum_{x_2=0}^{s_2} f_{X_1,X_2}(x_1,x_2)(s_1-x_1) f_{S_1,S_2}(s_1-x_1, s_2-x_2) \\ &\quad +(a+b) \sum_{x_1=0}^{s_1} \sum_{x_2=0}^{s_2} f_{X_1,X_2}(x_1,x_2) x_1 f_{S_1,S_2}(s_1-x_1, s_2-x_2) \end{aligned}$$

(3.15) is easily derived after an algebraic manipulation.
(3.16) follows similarly which completes the proof. ■

It should be noted that both recursions (3.15) and (3.16) are valid when $s_1 \geq 1$ and $s_2 \geq 1$ simultaneously. Sundt(1999b) suggested to use the following recursion

$$\begin{aligned} f_{S_1,S_2}(s_1,s_2) &= \frac{1}{1 - a f_{X_1,X_2}(0,0)} \sum_{x_1}^{s_1} \sum_{x_2}^{s_2} [a + b\frac{x_1+x_2}{s_1+s_2}] \times \\ &\quad f_{S_1,S_2}(s_1-x_1, s_2-x_2) f_{X_1,X_2}(x_1,x_2) \end{aligned}$$

valid unless $(s_1,s_2) = (0,0)$.
The bivariate recursion is easily extended to the following multivariate recursion.

Theorem 3.4 *Let*

$$\mathbf{S} = (S_1,\ldots,S_n) = (\sum_{i=1}^{N} X_{1i},\ldots,\sum_{i=1}^{N} X_{ni})$$

with $(X_{1i},\ldots,X_{ni})$ *iid, arithmetic and independent of* N
The components of the vector $\mathbf{X}$ *are not necessarily independent*
N *belonging to Panjer's class, i.e. such that*

$$\frac{p(n)}{p(n-1)} = a + \frac{b}{n} \quad , \quad n \geq 1$$

Then

$$\begin{aligned} f_{\mathbf{S}}(0) &= \Psi_N(f_{\mathbf{X}}(0)) \\ f_{\mathbf{S}}(s_1,\ldots s_n) &= \frac{1}{(1 - a f_{\mathbf{X}}(0))} \sum_{x_1,\ldots,x_n}^{s_1,\ldots,s_n} [a + b\frac{x_1}{s_1}] f_{\mathbf{S}}(s_1-x_1,\ldots,s_n-x_n) f_{\mathbf{X}}(x_1,\ldots,x_n) \\ &\quad s_1 \geq 1 \\ &\ldots \\ f_{\mathbf{S}}(s_1,\ldots s_n) &= \frac{1}{(1 - a f_{\mathbf{X}}(0))} \sum_{x_1,\ldots,x_n}^{s_1,\ldots,s_n} [a + b\frac{x_n}{s_n}] f_{\mathbf{S}}(s_1-x_1,\ldots,s_n-x_n) f_{\mathbf{X}}(x_1,\ldots,x_n) \\ &\quad s_n \geq 1 \end{aligned}$$

3.3.2 The multivariate collective risk model as an approximation of the multivariate individual risk model

As shown in De Pril and Dhaene (1992), the individual risk model can be approximated by a collective risk model. They derive error bounds when working with the collective instead of the individual risk model. It is easy to see that their results remain in the bivariate setting. Namely, we have :

$$f^{Coll}_{S_1,S_2}(s_1,s_2)=\sum_{k=0}^{\infty} e^{-\lambda}\frac{\lambda^k}{k!} f^{*k}_{X_1,X_2}(s_1,s_2)$$

with

$$\begin{aligned}\lambda &= \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}q_j\\ f_{X_1,X_2}(x_1,x_2) &= \frac{1}{\lambda}\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}q_j g_i(x_1,x_2)\end{aligned}$$

With this bivariate Compound Poisson approximation, we have the following results :

$$d_K((S_1^{Ind},S_2^{Ind}),(S_1^{Coll},S_2^{Coll}))\le \frac{1}{2}\sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}q_j^2$$

$$d_{TV}((S_1^{Ind},S_2^{Ind}),(S_1^{Coll},S_2^{Coll}))\le \sum_{i=1}^{a}\sum_{j=1}^{b} n_{ij}q_j^2$$

where d_K denotes the Kolmogorov distance whereas d_{TV} denotes the total variation distance :

$$\begin{aligned}d_K((X_1,X_2),(Y_1,Y_2)) &= \sup_{(x_1,x_2)\in\mathbb{N}^2} |\mathbb{P}(X_1\le x_1, X_2\le x_2)-\mathbb{P}(Y_1\le x_1,Y_2\le x_2)|\\ d_{TV}((X_1,X_2),(Y_1,Y_2)) &= \sup_{A\subset\mathbb{N}^2} |\mathbb{P}((X_1,X_2)\in A)-\mathbb{P}((Y_1,Y_2)\in A)|\end{aligned}$$

Naturally, the multivariate extension is immediate.

3.3.3 Multivariate recursions with the (a,b,m) class

Obviously more complicated bivariate recursions can be derived in the same way, for example with the Sundt and Jewell (1981) family of counting distributions :

$$\frac{p(n)}{p(n-1)}=a+\frac{b}{n} \quad , \quad n\ge m \quad , \quad m\ge 1$$

Theorem 3.5 *Let*

$$(S_1,S_2)=(\sum_{i=1}^{N} X_i,\sum_{i=1}^{N} Y_i)$$

with (X_i, Y_i) *iid, arithmetic and independent of* N
X_i *and* Y_i *are not necessarily independent*
N *belonging to Sundt and Jewell family of counting distributions,*
Then, if $\psi_N(z)$ *denotes the probability generating function of* N*, we have*

$$\begin{aligned}
f_{S_1,S_2}(0,0) &= \psi_N(f_{X_1,X_2}(0,0)) \\
f_{S_1,S_2}(s_1,s_2) &= \frac{1}{1-af_{X_1,X_2}(0,0)}\left(\sum_{x_1}^{s_1}\sum_{x_2}^{s_2}[a+b\frac{x_1}{s_1}]f_{S_1,S_2}(s_1-x_1,s_2-x_2)f_{X_1,X_2}(x_1,x_2)\right. \\
&\quad \left.+\sum_{i=1}^{m}p(i)f^{*i}_{X_1,X_2}(x_1,x_2)\right) \quad , \quad s_1 \geq 1 \\
f_{S_1,S_2}(s_1,s_2) &= \frac{1}{1-af_{X_1,X_2}(0,0)}\left(\sum_{x_1}^{s_1}\sum_{x_2}^{s_2}[a+b\frac{x_2}{s_2}]f_{S_1,S_2}(s_1-x_1,s_2-x_2)f_{X_1,X_2}(x_1,x_2)\right. \\
&\quad \left.+\sum_{i=1}^{m}p(i)f^{*i}_{X_1,X_2}(x_1,x_2)\right) \quad , \quad s_2 \geq 1
\end{aligned}$$

The multivariate extension is immediate.

3.4 Recursion for multivariate convolutions

In this section we consider a portfolio with n identical policies. ($p = p_j \quad \forall j$ and $g = g_i \quad \forall i$).
¿From theorem 3.2 we get

$$\begin{aligned}
f_{S_1,S_2}(0,0) &= p^n \\
s_1 f_{S_1,S_2}(s_1,s_2) &= nv(s_1,s_2) \\
v(s_1,s_2) &= \frac{q}{p}\sum_{x_1=0}^{\min(s_1,m_1)}\sum_{x_2=0}^{\min(s_2,m_2)} f_{X_1,X_2}(x_1,x_2)(x_1 f_{S_1,S_2}(s_1-x_1,s_2-x_2) - \\
&\quad v(s_1-x_1,s_2-x_2) \quad s_1 \geq 1 \\
s_2 f_{S_1,S_2}(s_1,s_2) &= nw(s_1,s_2) \\
w(s_1,s_2) &= \frac{q}{p}\sum_{x_1=0}^{\min(s_1,m_1)}\sum_{x_2=0}^{\min(s_2,m_2)} f_{X_1,X_2}(x_1,x_2)(x_2 f_{S_1,S_2}(s_1-x_1,s_2-x_2) - \\
&\quad w(s_1-x_1,s_2-x_2) \quad s_2 \geq 1
\end{aligned}$$

which immediately gives

$$\begin{aligned}
f_{S_1,S_2}(0,0) &= p^n \\
s_1 f_{S_1,S_2}(s_1,s_2) &= \frac{q}{p}\sum_{x_1=0}^{\min(s_1,m_1)}\sum_{x_2=0}^{\min(s_2,m_2)}\left(\frac{(n+1)x_1-s_1}{n}\right)\times \\
&\quad f_{X_1,X_2}(x_1,x_2)f_{S_1,S_2}(s_1-x_1,s_2-x_2) \quad , \quad s_1 \geq 1 \\
s_2 f_{S_1,S_2}(s_1,s_2) &= \frac{q}{p}\sum_{x_1=0}^{\min(s_1,m_1)}\sum_{x_2=0}^{\min(s_2,m_2)}\left(\frac{(n+1)x_2-s_2}{n}\right)\times
\end{aligned}$$

$$f_{X_1,X_2}(x_1,x_2)f_{S_1,S_2}(s_1-x_1,s_2-x_2) \quad , \quad s_2 \geq 1$$

which is the bivariate extension of De Pril's (1985) formula for the convolutions of arithmetic distributions. The multivariate extension is trivial.
This formula can also be deduced from the multivariate Panjer's algorithm :

Theorem 3.6 *Let*

$$(X_1,X_2)^{*^n} = (X_{11}+\cdots+X_{1n}, X_{21}+\cdots+X_{2n})$$

for which one has

$$f^{*^{n+1}}_{X_1,X_2}(s_1,s_2) = \sum_{x_1=0}^{s_1}\sum_{x_2=0}^{s_2} f^{*^n}_{X_1,X_2}(x_1,x_2)f_{X_1,X_2}(s_1-x_1,s_2-x_2)$$

Let (X_{1i},X_{2i}) *be defined on the non-negative integers and* $f_{X_1,X_2}(0,0)>0$.
Then the following recursion holds :

$$\begin{aligned}
f^{*^n}_{X_1,X_2}(0,0) &= f^n_{X_1,X_2}(0,0) \\
f^{*^n}_{X_1,X_2}(s_1,s_2) &= \frac{1}{f_{X_1,X_2}(0,0)}(\sum_{x_1}^{s_1}\sum_{x_2}^{s_2}\left[\frac{n+1}{s_1}x_1-1\right] f^{*^n}_{X_1,X_2}(s_1-x_1,s_2-x_2)f_{X_1,X_2}(s_1,s_2) \\
&\quad , \quad s_1 \geq 1 \\
f^{*^n}_{X_1,X_2}(s_1,s_2) &= \frac{1}{f_{X_1,X_2}(0,0)}(\sum_{x_1}^{s_1}\sum_{x_2}^{s_2}\left[\frac{n+1}{s_2}x_2-1\right] f^{*^n}_{X_1,X_2}(s_1-x_1,s_2-x_2)f_{X_1,X_2}(x_1,x_2) \\
&\quad , \quad s_2 \geq 1
\end{aligned}$$

Proof :

¿From Theorem 3.3 and by assuming

$$(R_1,R_2) = (W_{1_1}+\cdots+W_{1_K}, W_{2_1}+\cdots+W_{2_K})$$

with $K \sim Bi(n, 1-f_{X_1,X_2}(0,0))$
(W_1,W_2) such that $f_{W_1,W_2}(x,y) = \frac{f_{X_1,X_2}(x_1,x_2)}{1-f_{X_1,X_2}(0,0)}$
we immediately get the desired result because $(R_1,R_2)=(S_1,S_2)$. ■

Note that this recursion needs an atom on $(0,0)$ or may be generalized if there is an atom on (c,c) with no mass on the other points (x,y) such that $x \leq c$ and $y \leq c$.

This result was established in Sundt (1999b) and the multivariate extension is immediate.

Chapter 4

Pricing reinstatements and ruin of the Cedent when there is excess of loss reinsurance with clauses

In this chapter we first review the concept of excess of loss reinsurance. In particular we study some typical clauses that may be found on the market. These clauses are generally neglected in the actuarial literature. We then show how to price these clauses and specifically the case of reinstatements.
Next, we adopt the point of view of the Ceding Company and we study the retained risk, in particular when there are reinstatements.

4.1 Excess of loss reinsurance

4.1.1 Definition

Excess of loss reinsurance is widely used when the concern of the ceding company is to cut the high claims individually. If D is the retention of the Insurance Company, then the Reinsurer and the Cedent are liable for

$$
\begin{aligned}
R_i &= \max(0, X_i - D) \qquad (4.1)\\
C_i &= X_i - R_i \qquad (4.2)
\end{aligned}
$$

Often the Reinsurer limits its liability to a level L. $[D, D + L]$ is called the layer. (4.1) and (4.2) become

$$
\begin{aligned}
R_i &= \min(L, \max(0, X_i - D)) \qquad (4.3)\\
C_i &= X_i - R_i \qquad (4.4)
\end{aligned}
$$

where the subscript i denotes the ith claim in the case of the collective risk model or the ith risk in the case of the individual risk model.
If there are N claims during the year, the aggregate claims distributions of the Reinsurer and the Cedent are

$$
\begin{aligned}
S_R &= R_1 + \cdots + R_N\\
S_C &= C_1 + \cdots + C_N
\end{aligned}
$$

in the collective risk model.
If there are n risks in the portfolio, the aggregate claims distribution of the Reinsurer and the Cedent are

$$\begin{aligned} S_R &= I_1 R_1 + \cdots + I_n R_n \\ S_C &= I_1 C_1 + \cdots + I_n C_n \end{aligned}$$

in the individual risk model.
Excess of loss reinsurance is an individual reinsurance in the sense that the payments of the Reinsurer depend on the individual claim amounts. There is a theoretical result (see Denuit and Vermandele (1998)) that says that among all the individual reinsurances with the same expected reinsurance loss and the same reinsurance premium, excess of loss reinsurance is optimal in the sense that it is not possible to find another type of reinsurance with a retained risk less risky in the sense of the stop-loss order. This theoretical result explains why excess of loss reinsurance is so common on the reinsurance market. However it is generally assorted with some specific clauses. Before studying these clauses, let us mention that there may be more than one layer protecting the retention of the Cedent .
Assume that the excess of loss reinsurance is split in m layers.
Let X be a random variable. Let C and $R_1, \ldots, R_m$ be the respective parts of the Ceding Company and the different reinsurance layers :

$$\begin{aligned} R_{1i} &= \min(L_1, \max(X_i - D_1, 0)) \quad \text{cover } L_1\, xs\, D_1 \\ R_{2i} &= \min(L_2, \max(X_i - D_2, 0)) \quad \text{cover } L_2\, xs\, D_2 \\ &\cdots \\ R_{mi} &= \min(L_m, \max(X_i - D_m, 0)) \quad \text{cover } L_m\, xs\, D_m \\ C_i &= X_i - (R_{1i} + \cdots + R_{mi}) \end{aligned}$$

We define S_C and $S_{R_1}, \ldots, S_{R_m}$ the aggregate parts of the Cedent and Reinsurers.
In general we will have $D_i = D_{i-1} + L_{i-1} \quad , \quad i = 2, \ldots, m$. In this case, we might question why it is necessary to cut out in different layers. The reason will be clear when we will study the notion of reinstatements (see section 4.1.4).
Note that in general, L_m will be equal to ∞ if the insured risk is a liability risk for which the insurer has to assume infinite claim amounts although it will be equal to the maximum loss in property insurance.
In the next subsections we mention some typical clauses accompanying excess of loss reinsurance.
Let us define

- P the insurance premium
- P_{Re} the reinsurance premium

For the rest of this chapter we suggest the following conventions. P is a risk premium, i.e. the expected loss plus some loading for fluctuation. P_{Re} is a commercial premium, i.e. the risk premium plus administrative and profit costs, brokerage and possibly tax.
Finally we define P_{Ced} the "net of reinsurance" insurance risk premium, i.e.

$$P_{Ced} = P - P_{Re}$$

¿From now on we will assume in order to make things easier that there is only one layer. Obviously all the results remain true in the case of more than one layer.

The aggregate claims distribution of the Reinsurer is a function of R and N or n.

$$S_{Re} = f(R, N \text{ or } n)$$

The aggregate claims distribution of the Cedent is a function of R, C, N or n.

$$S_{Ced} = f(R, C, N \text{ or } n)$$

In the rest of the chapter we will show how to calculate the reinsurance premium. Sometimes this reinsurance premium will be a random variable :

$$P_{Re} = P_{Re}^{det} + P_{Re}^{rand}$$

where P_{Re}^{det} is the part of the reinsurance premium that is deterministic whereas P_{Re}^{rand} is the random part of P_{Re}. P_{Re}^{det} is the part of the reinsurance premium that will be paid for sure to the Reinsurer. P_{Re}^{rand} is the part of the reinsurance premium that will be paid in function of the observed claims hitting the Reinsurer. It is therefore a random variable.
Furthermore we will study the ruin probability of the ceding company. In order to apply the algorithms related to ruin probability, it is necessary to have a Cedent premium P_{Ced} that is not random.

$$P_{Ced} = P - P_{Re}^{det}$$

This is the reason why we will refer to a Cedent claims distribution S_{Ced} that is constructed such that P_{Ced} is deterministic. This means that the randomness of the P_{Ced}, due to P_{Re} will be transferred to S_{Ced} :

$$S_{Ced} = f(R, C, N \text{ or } n) + P_{Re}^{rand}$$

The reinsurance premium, P_{Re} (or more precisely P_{Re}^{det}) will be calculated according to a premium principle H (see section 4.2.2, 4.2.3 and 4.2.4 for a discussion about some premium principles applied to excess of loss reinsurance with reinstatements). Note that the definition of P_{Re} as a commercial premium may be overruled if there is a deterministic and a random reinsurance premium. Indeed, the deterministic premium automatically includes brokerage whereas the random part of the premium may not.
We now review different kinds of clauses.

4.1.2 Aggregate

Unfortunately for the Ceding Company, the Reinsurer often limits his aggregate liability. In practice we speak of an aggregate. Let us define Agg the maximum amount paid by the reinsurer during one year. Then the aggregate claims distributions and the premiums become

$$\begin{aligned}
S_{Re} &= \min(S_R, Agg) \\
S_{Ced} &= S_C + \max(0, S_R - Agg) \\
P_{Re} &= H(S_{Re}) \\
P_{Ced} &= P - P_{Re}
\end{aligned}$$

The main reason for the Reinsurer's limitation is that he is not able to propose unlimited capacity to the market. So he limits the offered capacity on the market with an aggregate. Most of the time, the impact on the reinsurance premium will be low because it will concern only the very far right tail of S_R.

4.1.3 Aggregate deductible

If the Ceding Company wants to pay a lower reinsurance premium, a good way is to ask for an aggregate deductible AD. This means that the aggregate liability of the Reinsurer has a deductible equal to AD. The aggregate claims distributions and premiums become

$$\begin{aligned} S_{Re} &= \max(0, S_R - AD) \\ S_{Ced} &= S_C + \min(S_R, AD) \\ P_{Re} &= H(S_{Re}) \\ P_{Ced} &= P - P_{Re} \end{aligned}$$

4.1.4 Reinstatements

Often the aggregate clause takes another form. The aggregate (Agg) is a multiple of the length of the layer L. Assume that this multiple is $k+1$. It means that the layer may be consumed $k+1$ times. We say that there are k reinstatements. These reinstatements may be free or paid. The case of free reinstatements is just a pure aggregate. In case of paid reinstatements, the reinstatements premiums are a fraction c_i , $1 \le i \le k$ of the initial reinsurance premium P_{Re}^{det} in proportion of the part of the layer hit by the claims. We say that the reinstatement premiums are payable pro rata capita of the amount consumed.
The payment of the reinstatement premium is compulsory.
Sundt (1991) gave the methodology to price excess of loss treaties with reinstatements.
The retained risk of the Cedent is composed of three parts :

- The claims that are not covered by the reinsurance : S_C
- The aggregate claims that exceed the aggregate liability of the Reinsurer : $\max(0, S_R - (k+1)L)$
- The reinstatement premiums : $P_{Re}^{rand} = \frac{P_{Re}^{det}}{L} \sum_{i=1}^{k} c_i \min(L, \max(0, S_R - (i-1)L))$

The aggregate claims distributions and premiums are :

$$\begin{aligned} S_{Re} &= \min(S_R, (k+1)L) - P_{Re}^{rand} \\ S_{Ced} &= S_C + \max(0, S_R - (k+1)L) + P_{Re}^{rand} \\ P_{Re}^{det} &= H(S_{Re}) \\ P_{Ced} &= P - P_{Re}^{det} \end{aligned}$$

in the case there is only one layer.
If there are multiple layers (m) we have

$$S_{Ced} = S_C + \sum_{j=1}^{m} \max(0, S_{R_j} - (k_j+1)L_j) + \sum_{j=1}^{m} \frac{P_{Re\,j}^{det}}{L_j} \sum_{i=1}^{k_j} c_{j_i} \min(L, \max(0, S_{R_j} - (i-1)L_j))$$

The number of reinstatements (k_j) may be different for different layers. Each layer has its own premium $P_{Re\,j}^{det}$. This represents what the Reinsurers do in practice. However we might also think of a single premium P_{Re}^{det} for the whole layers. The reinstatement premiums would be function of this single initial premium P_{Re}^{det} :

$$S_{Ced} = S_C + \sum_{j=1}^{m} \max(0, S_{R_j} - (k_j+1)L_j) + \sum_{j=1}^{m} \frac{P_{Re}^{det}}{L_j} \sum_{i=1}^{k_j} c_{j\,i} \min(L, \max(0, S_{R_j} - (i-1)L_j))$$

This is however not used in practice.

4.1.5 Proportional coinsurance

In the case where the Reinsurer wants to leave some liability to the Ceding Company on each claim hitting the reinsurance, a good way is to impose a proportion of the excess claim to be paid by the Ceding Company. Let us define α the fraction of each excess claim paid by the Cedent . We have

$$\begin{aligned} R_i &= (1-\alpha)\min(L, \max(0, X_i - D)) \\ C_i &= \min(X_i, D) + \alpha \min(L, \max(0, X_i - D)) \end{aligned}$$

This type of reinsurance implies a reduction of the reinsurance premium with a factor $1-\alpha$.

4.1.6 Sliding scale

For any type of reinsurance, the Ceding Company may ask for a slide premium if it believes that the aggregate claims distribution will not be important. In this case, the initial premium will be lower than with a classical premium. The reinsurance premium writes :

$$P_{Re} = \begin{cases} P_1 & \text{if } \; S_R < \frac{P_1}{f} \\ fS_R & \text{if } \; \frac{P_1}{f} < S_R < \frac{P_2}{f} \\ P_2 & \text{if } \; S_R > \frac{P_2}{f} \end{cases}$$

where f is a loading coefficient. Generally $f = \frac{100}{70}$ or $\frac{100}{80}$. The reinsurance premium consists in a fixed (minimum) premium $P_1 = P_{Re}^{det}$ and a random premium : P_{Re}^{rand}.

$$P_{Re} = P_{Re}^{det} + P_{Re}^{rand}$$

$$P_{Re}^{rand} = \begin{cases} 0 & \text{if } \; S_R < \frac{P_1}{f} \\ fS_R - P_1 & \text{if } \; \frac{P_1}{f} < S_R < \frac{P_2}{f} \\ P_2 - P_1 & \text{if } \; S_R > \frac{P_2}{f} \end{cases}$$

Note that this kind of clause may be applied to any type of reinsurance, not only to the excess of loss reinsurance. The aggregate claims distributions and the premiums are

$$\begin{aligned} S_{Re} &= S_R - P_{Re}^{rand} \\ S_{Ced} &= S_C + P_{Re}^{rand} \\ P_{Re}^{det} &= H(S_{Re}) \\ P_{Ced} &= P - P_{Re}^{det} \end{aligned}$$

4.1.7 Profit participation

Sometimes the Reinsurer offers a profit participation. This means that after the aggregate claims have been observed, in case of a good statistic, the Reinsurer gives back a part of the reinsurance premium to the Ceding Company. Once again this kind of clause is not typical of excess of loss reinsurance.
A common type of profit participation is to give back

$$PP = \alpha \max(0, \beta P_{Re}^{det} - S_R)$$

where $1-\beta$ represents the administrative costs of the Reinsurer and α represents the fraction of the profit of the Reinsurer that is given back to the ceding company. The aggregate parts of the Cedent and Reinsurer become :

$$\begin{aligned} S_{Re} &= S_R + PP \\ S_{Ced} &= S_C + \alpha \min(\beta P_{Re}^{det}, S_R) \\ P_{Re}^{det} &= H(S_{Re}) \\ P_{Ced} &= P - P_{Re}^{det} + \alpha\beta P_{Re}^{det} \end{aligned}$$

Another type of profit participation is

$$PP = \alpha(P_{Re}^{det} - S_R) \quad \text{if } \frac{S_R}{P_{Re}^{det}} < \beta$$

where β is a critical loss ratio under which the profit participation is accorded and α is the fraction of the profit of the Reinsurer that is given back to the Cedent . The parts of the Cedent and Reinsurer become

$$\begin{aligned} S_{Re} &= S_R + PP \\ S_{Ced} &= S_C + \alpha S_R \mathbb{I}_{[\frac{S_R}{P_{Re}^{det}} < \beta]} + \alpha P_{Re}^{det} \mathbb{I}_{[\frac{S_R}{P_{Re}^{det}} > \beta]} \\ P_{Re}^{det} &= H(S_{Re}) \\ P_{Ced} &= P - P_{Re}^{det} + \alpha P_{Re}^{det} \end{aligned}$$

As a special case of this type of profit participation, we find the no claim bonus which gives a profit participation only if S_R=0 :

$$PP = \alpha P_{Re}^{det} \quad \text{if } S_R = 0$$

The parts of the Cedent and Reinsurer become

$$\begin{aligned} S_{Re} &= S_R + PP \\ S_{Ced} &= S_C + \alpha P_{Re}^{det} \mathbb{I}_{[S_R > 0]} \\ P_{Re}^{det} &= H(S_{Re}) \\ P_{Ced} &= P - P_{Re}^{det} + \alpha P_{Re}^{det} \end{aligned}$$

A third type of profit participation is the slide commission. In this case, the Cedent receives a fraction of the reinsurance premium. The fraction varies according to the level of the loss ratio :

$$\begin{aligned} PP &= \alpha_1 P_{Re}^{det} \quad \text{if } \beta_1 < \frac{S_R}{P_{Re}^{det}} < \beta_2 \\ &\dots \\ &= \alpha_i P_{Re}^{det} \quad \text{if } \beta_i < \frac{S_R}{P_{Re}^{det}} < \beta_{i+1} \\ &\dots \\ &= \alpha_r P_{Re}^{det} \quad \text{if } \beta_r < \frac{S_R}{P_{Re}^{det}} < \infty \end{aligned}$$

We make the logical assumption that $\alpha_1 > \ldots > \alpha_r$. The aggregate claims distributions and premiums become

$$\begin{aligned} S_{Re} &= S_R + PP \\ S_{Ced} &= S_C + \sum_{i=1}^{r} (\alpha_1 - \alpha_i) P_{Re}^{det} \mathbb{I}_{[\beta_i < \frac{S_R}{P_{Re}^{det}} < \beta_{i+1}]} \\ P_{Re}^{det} &= H(S_{Re}) \\ P_{Ced} &= P - P_{Re}^{det} + \alpha_1 P_{Re}^{det} \end{aligned}$$

4.2 The pricing of reinstatements

Sections 4.2.1, 4.2.2, 4.2.3 and 4.2.4 are mainly based on Walhin and Paris (1999a).

4.2.1 Notations

We will use the following notations. Some of them are inspired by Sundt (1991).

- $p = \frac{P_{Re}^{det}}{L}$ is the rate on line
- $r_i = \min(\max(0, S_R - iL), L)$ is the coverage of the ith reinstatement
- $c_i P_{Re}^{det} \frac{r_{i-1}}{L}$ is the ith reinstatement premium (pro rata capita)
- $R_k = \sum_{i=0}^{k} r_i = \min(S_R, (k+1)L)$ is the aggregate claim payments of the Reinsurer
- $P_{Re} = P_{Re}^{det}(1 + \frac{1}{L} \sum_{i=1}^{k} c_i r_{i-1})$ is the total reinsurance premium. It is a random variable.
- $d_i = \mathbb{E} r_i$
- $D_k = \mathbb{E} R_k$
- $v_{ij} = \mathbb{C}ov(r_i, r_j)$

- $v_i = \mathbb{V}ar\, r_i$
- $V_k = \mathbb{V}ar\, R_k$
- γ : security loading with the standard deviation principle
- ρ : risk aversion index for PH transform principle
- α : safety loading with the expected value principle

We will work with the following numerical example :
The claim amount distribution is given by

X	1	2	3	4	5	6	8	10	12	14
$f_X(x)$	0.2	0.15	0.15	0.2	0.06	0.06	0.06	0.05	0.04	0.03

Table 4.1: Claim amount distribution

The number of claims is Poisson distributed with mean $\lambda = 3$.
There are two reinsurance layers : $4\,xs\,6$ and $4\,xs\,10$. We will also consider the layer $8\,xs\,6$.

4.2.2 Expected value principle

The methodology was worked out by Sundt (1991).
An initial premium P_{Re}^{det} is paid. Reinstatement premiums, functions of P_{Re}^{det}, might be paid, following the occurence of an event. This random part of the premium income is

$$P_{Re}^{det} \sum_{j=1}^{k} \frac{c_j}{L} \min(L, \max(0, S_R - (j-1)L))$$

The reinsurer covers $(k+1)$ times the layer :

$$\min(S_R, (k+1)L)$$

The premium P_{Re}^{det} is the solution to the problem of equating the mean premium and the mean aggregate claims. Indeed, in order to calculate the fair premium amount, we have to set the expected premium collections equal to the expected loss :

$$\mathbb{E}[P_{Re}^{det}(1 + \sum_{j=1}^{k} c_j \min(1, \max(0, \frac{S_R - (j-1)L}{L}))] = \mathbb{E}\min(S_R, (k+1)L)$$

This immediately gives

$$P_{Re}^{det} = \frac{\mathbb{E}\min(S_R, (k+1)L)}{1 + \sum_{j=1}^{k} \frac{c_j}{L} \mathbb{E}\min(L, \max(0, S_R - (j-1)L))} \tag{4.5}$$

This is the premium given by the expected value premium principle developed in Sundt (1991).

c_i	$k=0$	1	2	3
0%	1.4592	1.7550	1.7955	1.7996
50%		1.4843	1.4724	1.4697
100%		1.2859	1.2479	1.2420
150%		1.1343	1.0828	1.0754
$1st : 100\%\ 2nd : 0\%$			1.3155	
$1st : 0\%\ 2nd : 100\%$			1.6718	

Table 4.2: Pure Reinsurance premiums

The following tables give the reinsurance premium in function of the number of reinstatements (k) as well as the percentage (c_i) of the reinstatements premiums :

Note the interesting behaviour of the reinsurance premium when the reinstatements are payable : it diminishes with the number of reinstatements for this particular example.
Now we give the premiums calculated according to the expected value principle with safety loading $\alpha = 18.27\%$:

c_i	$k=0$	1	2	3
0%	1.7258	2.0757	2.1236	2.1284
50%		1.7555	1.7415	1.7383
100%		1.5209	1.4760	1.4690
150%		1.3416	1.2607	1.2720

Table 4.3: Loaded reinsurance premiums

The safety loading has been chosen so that the loaded premium with 3 reinstatements at 150% should remain the same within the three types of loadings (expected value principle, standard deviation premium principle and PH transform premium principle).

4.2.3 Standard deviation principle

This case was studied in Sundt (1991). The reinsurance premium P_{Re}^{det} is obtained from the equation

$$\mathbb{E}(P_{Re}) = \mathbb{E}R_k + \gamma\sqrt{Var(R_k - P_{Re})} \tag{4.6}$$

Introducing some further notations :

$$\begin{aligned} A &= L + \sum_{i=1}^{k} c_i d_{i-1} \\ B &= \mathbb{V}ar(\sum_{i=1}^{k} c_i r_{i-1}) \end{aligned}$$

$$C \;=\; \mathbb{C}ov(\sum_{i=1}^{k} c_i r_{i-1}, R_k)$$

The rate on line p is deduced from the equation

$$pA - D_k = \gamma\sqrt{V_k + p^2 B - 2pC} \tag{4.7}$$

Sundt (1991) claims that an acceptable solution will be provided at least if $\gamma < \frac{A}{\sqrt{B}}$. We now show the exact condition on γ.
(4.7) leads to

$$p^2(A^2 - \gamma^2 B) - 2(AD_k - C\gamma^2)p + D_k^2 - \gamma^2 V_k = 0 \tag{4.8}$$

The determinant will be positive if

$$\gamma < \sqrt{\frac{A^2 V_k + BD_k^2 - 2CAD_k}{BV_k - C^2}}$$

Obviously this bound on γ is better than Sundt's bound :

$$\begin{aligned} \frac{A^2}{B} &< \frac{A^2 V_k + BD_k^2 - 2CAD_k}{BV_k - C^2} \\ &\Leftrightarrow \\ A^2 BV_k - A^2 C^2 &< A^2 BV_k + B^2 D_k^2 - 2ABCD_k \\ &\Leftrightarrow \\ 0 &< A^2 (C + B\frac{D_k}{A})^2 \end{aligned}$$

which is always true.
The rate on line p is then given by the largest solution of (4.8).
Note that (4.6) can be written as

$$P_{Re}^{det} = \mathbb{E}S_{Re} + \gamma\sqrt{Var(S_{Re})} \tag{4.9}$$

with S_{Re} a function of P_{Re}^{det}. So P_{Re}^{det} can also be found by an iterative process through (4.9).
Note that if one uses the algorithm

$$(P_{Re}^{det})^{(i)} = f((P_{Re}^{det})^{(i-1)})$$

with $f = \mathbb{E}S_{Re} + \gamma\sqrt{Var(S_{Re})}$, f should satisfy the Lipschitz condition in order to get a fixed point. In practice this is not always the case and divergence may occur. Then another iterative algorithm has to be used like the bisection method for example.
Nevertheless the solution, if any, is always explicitely available for the standard deviation principle.

With a coefficient loading $\gamma = 0.25$ we find the following reinsurance premiums :

c_i	$k=0$	1	2	3
0%	1.9125	2.3537	2.4265	2.4355
50%		1.9249	1.8820	1.8698
100%		1.6334	1.5391	1.5177
150%		1.4217	1.3032	1.2774

Table 4.4: Reinsurance premiums

4.2.4 PH transform principle

Wang (1996) introduced several risk adjusted premium calculation principles.
These principles were studied by Silva and Centeno (1998) and they concluded that only the proportional hazard (PH) premium principle has a different behaviour from the expected value principle.
In this section we will concentrate on the evaluation of the premium of an excess of loss treaty with reinstatements with the PH premium principle.
For a risk S, the premium calculated according to the PH premium priciple is given by

$$P = \int_0^\infty (1 - F_S(t))^{1/\rho} dt \quad , \quad \rho \geq 1$$

where ρ is called the risk aversion index.
In our case, we are interested in

$$P_{Re}^{det} = \int_0^\infty (1 - F_{S_{Re}}(t))^{1/\rho} dt$$

where S_{Re} is itself a function of P_{Re}^{det}. The premium P_{Re}^{det} will be given by iterations.
The observation of the preceding section about divergence is still valid.
Note that, if $c_i p > 1$, a case that would not happen in practice, the distribution of S_{Re} is not on the positive numbers and so the PH transform has to be extended as

$$P_{Re}^{det} = \int_{-\infty}^0 (1 - F_{S_{Re}}(t))^{1/\rho} - 1 \; dt + \int_0^\infty (1 - F_{S_{Re}}(t))^{1/\rho} dt$$

With a loading coefficient $\rho = 1.2675$ we find

c_i	$k=0$	1	2	3
0%	1.8022	2.3118	2.4174	2.4347
50%		1.8868	1.8754	1.8695
100%		1.5938	1.5320	1.5176
150%		1.3795	1.2948	1.2771

Table 4.5: Reinsurance premiums

4.2.5 ROL method

This section is based on Walhin (2000d).
Throughout the section we will use a numerical example, based on the following assumptions:
The X_i, N and T are independent random variables.
N : Poisson with mean λ
X : truncated Pareto with parameters A, B and α

$$\begin{aligned} F_X(x) &= 0 \quad , \quad x \leq A \\ F_X(x) &= \frac{A^{-\alpha} - x^{-\alpha}}{A^{-\alpha} - B^{-\alpha}} \quad , \quad A < x < B \\ F_X(x) &= 1 \quad , \quad x \geq B \end{aligned}$$

T : Beta distributed with parameters a and b :

$$f_T(t) = \frac{1}{B(a,b)} t^{a-1}(1-t)^{b-1} \quad , \quad 0 \leq t \leq 1 \quad , \quad a > 0\,, b > 0$$

where $B(a,b)$ is the Beta function :

$$B(a,b) = \int_0^1 t^{a-1}(1-t)^{b-1} dt$$

Note that in the particular case where $a = b = 1$ the Beta distribution degenerates into the Uniform Distribution.
We will take $\lambda = 1$ or 2, $A = 20$, $B = 50$, $\alpha = 1.5$, $a = 0.5, 1, 5$ and $b = 0.5, 1, 5$.

We will study 4 excess of loss covers :

- 30 xs 20
- 10 xs 20
- 10 xs 30
- 10 xs 40

As we will use convolutions or the algorithm of Panjer (1981), we need a discretization of the random variable X. The discrete distribution X_{dis} has been obtained by the minimization of the Kolmogorov distance between the exact distribution and the approximate one (see chapter 2). The minimization has been handled with the following constraints

- X_{dis} has the following support : $20, 21, \ldots, 49, 50$
- $F_{X_{dis}}(B) = 1$
- $f_{X_{dis}}(x) \geq 0 \quad , \quad \forall x = 20, \ldots, 50$
- $\mathbb{E}X_{dis} = \mathbb{E}X$
- $\mathbb{V}arX_{dis} = \mathbb{V}arX$
- $\mathbb{E}\min(10, \max(0, X_{dis} - 20)) = \mathbb{E}\min(10, \max(0, X - 20))$
- $\mathbb{E}\min(10, \max(0, X_{dis} - 30)) = \mathbb{E}\min(10, \max(0, X - 30))$

in order to keep the first two moments fixed as well as the expectations on each layer we will study. The expectation on the layer 10 xs 40 is automatically kept by linear combination of the defined constraints.
The Kolmogorov distance between the exact and the discretized distribution is 0.0472.

The rate on line for an unlimited free reinstatements treaty is defined as

$$ROL = \frac{\mathbb{E}N\mathbb{E}R}{L}$$

It is the premium for a treaty with unlimited free reinstatements covering the layer $[D, D+L]$ divided by the length of the layer L.
Let us assume that we do not know the distribution S_R. Instead, we know the rate on line. The rate on line method assumes that there are only total losses, i.e. losses hitting the layer completely . For such losses, the claims frequency is the ROL.
Let us define two new random variables :

$R' =$ L with probability 1

N' has the same distribution as N but with mean ROL

It is then easy to show that the formula (4.5) for P_{Re}^{det} becomes

$$P_{Re}^{det} = L \frac{\sum_{i=0}^{k} \mathbb{P}(N' > i)}{1 + \sum_{i=1}^{k} c_i \mathbb{P}(N' > i - 1)}$$

This formula is obviously easier to evaluate than the exact one.
The following table gives the premium for the 30 xs 20 cover with uniform price $c_j \equiv c$ of the reinstatement.

c/k	0	1	2	3	4
0%	8.75	9.48	9.51	9.52	9.52
50%		8.28	8.21	8.21	8.21
100%		7.34	7.23	7.22	7.22
150%		6.59	6.45	6.45	6.45

Table 4.6: 30 xs 20 : pure premium

Our numerical example gives the following approximate premiums :

c/k	0	1	2	3	4
0%	8.15	9.38	9.50	9.51	9.52
50%		8.26	8.22	8.21	8.21
100%		7.37	7.24	7.22	7.22
150%		6.66	6.47	6.45	6.45

Table 4.7: 30 xs 20 : approximate pure premium

We note that, in case of free reinstatements, the exact premium is always higher than the approximate one. This is obviously not a safe approximation and it can be shown on numerical examples that the error may exceed 10%.
We will now prove that the conjecture mentionned above is always true.

Theorem 4.1 *Let N be such that its probability generating function $\psi_N(t) = W(\theta(t-1))$ where W is a function independent of the parameter θ.*
Then the approximate premium is always lower than the exact premium if the reinstatements are free.

Proof :
Let us define the random variable R'' :

x	$\mathbb{P}[R'' = x]$
0	$1 - \frac{\mathbb{E}R}{L}$
L	$\frac{\mathbb{E}R}{L}$

Table 4.8: Probability function of R''

By a theorem of Panjer and Willmot (1984) we have

$$\begin{aligned}
\psi_{S_{R''}} &= \psi_N(\psi_{R''};\theta) \\
&= \psi_N(1 - \frac{\mathbb{E}R}{L} + \frac{\mathbb{E}R}{L}\psi_{R'};\theta) \\
&= \psi_N(\psi_{R'};\frac{\mathbb{E}R}{L}\theta) \\
&= \psi_{S_{R'}}
\end{aligned}$$

from which we conclude that $S_{R''} = S_{R'}$.
¿From the crossing condition (see Goovaerts et al. (1990) for a reference), it is clear that $R \preccurlyeq_{sl} R''$:

$$SL(R'', t) \geq SL(R, t) \quad \forall t$$

where $SL(R, t)$ denotes the stop loss premium for a risk R with retention t :

$$SL(R, t) = \int_t^{\infty} (x - t) dF_R(x)$$

By Bühlmann et al. (1977), the stop loss order is preserved under compounding :

$$SL(S_{R''}, t) \geq SL(S_R, t) \quad \forall t$$

With free reinstatements we have

$$\begin{aligned} P &= \mathbb{E}S_R - SL(S_R, (t+1)L) \\ P' &= \mathbb{E}S'_R - SL(S'_R, (t+1)L) \end{aligned}$$

Combined with the fact that $\mathbb{E}S_R = \mathbb{E}S_{R'}$ we find

$$P' \leq P$$

■

The proposition is particularly interesting when N is Poisson ($W(x) = e^x$ with $\theta = \lambda$) and Negative Binomial ($W(x) = (1 - x)^{-r}$ with $\theta = \pi$).

The following tables give the fraction $\frac{\text{approximate premium}}{\text{exact premium}}$ for the layers 30 xs 20, 10 xs 20, 10 xs 30 and 10 xs 40 with a uniform price (c) of the reinstatement.

c/k	0	1	2	3	4
0%	0.93174	0.98889	0.99898	0.99993	0.99999
50%		0.99757	1.00049	1.00007	1.00000
100%		1.00438	1.00165	1.00017	1.00001
150%		1.00988	1.00256	1.00026	1.00001

Table 4.9: 30 xs 20 : approximate / exact premium

c/k	0	1	2	3	4
0%	0.98554	0.99379	0.99837	0.99970	0.99996
50%		0.99659	0.99983	1.00010	1.00003
100%		0.99849	1.00073	1.00034	1.00008
150%		0.99986	1.00135	1.00050	1.00011

Table 4.10: 10 xs 20 : approximate / exact premium

c/k	0	1	2	3	4
0%	0.99100	0.99834	0.99983	0.99999	1.00000
50%		0.99924	1.00001	1.00001	1.00000
100%		0.99997	1.00016	1.00002	1.00000
150%		1.00058	1.00028	1.00003	1.00000

Table 4.11: 10 xs 30 : approximate / exact premium

c/k	0	1	2	3	4
0%	0.98841	0.99957	0.99999	1.00000	1.00000
50%		0.99990	1.00000	1.00000	1.00000
100%		1.00022	1.00002	1.00000	1.00000
150%		1.00052	1.00003	1.00000	1.00000

Table 4.12: 10 xs 40 : approximate / exact premium

We notice that the fraction increases with the price (c) of the reinstatements. This result can be shown analytically.
We also see that the fraction tends to unity with the number of reinstatements (k). This is logical as when $k \to \infty$, the claims distribution (R) is not important. The formula for unlimited uniformly paid reinstatements is

$$P = \frac{\mathbb{E}N\mathbb{E}R}{1 + \frac{c}{L}\mathbb{E}N\mathbb{E}R}$$

where we only use $\mathbb{E}R$.

Now let us assume that there is an aggregate deductible (AD), i.e. the Reinsurer is liable for the aggregate claims in excess of AD. It is not difficult to extend the formulae and the following example shows that the approximation is not controlled and may give very bad results.
Let us assume an aggregate deductible (AD=15) for the cover 30 xs 20. We find

c/k	0	1	2	3	4
0%	1.23881	1.27606	1.28599	1.28773	1.28796
50%		1.26636	1.27298	1.27403	1.27415
100%		1.25740	1.26113	1.26157	1.26160
150%		1.24909	1.25029	1.25021	1.25015

Table 4.13: 30 xs 20 with $AD = 15$: approximated / exact premium

The results are bad and this is not surprising as we know from Bühlmann et al. (1977) that

$$SL(S'_R, AD) > SL(S_R, AD)$$

i.e. the unlimited free reinstatement premium is higher for the approximate case.

4.2.6 Pro rata temporis reinstatements

This section is based on Walhin (2000d).
The pro rata temporis clause is essentially used for catastrophe reinsurance where typically only one reinstatement is offered by the reinsurer. In this section we will mainly consider this particular case.
In some cases it might be necessary to account for seasonality. Indeed hurricanes or tornadoes are seasonal. On the other side, earthquakes occur uniformly throughout the year.
We now have to take into account the time remaining until the renewal after each claim because the reinstatement premium is proportional to the remaining time in the reinsurance contract after an occurrence.
Thus we introduce a new random variable T, the elapsed time before the claim. We also have to study the random variable Y_i, part of the ith claim leading to the reinstatement premium. The Y_i are defined as

$$\begin{aligned} Y_1 &= R_1 \\ Y_i &= \min[\max(0, L - \sum_{l=1}^{i-1} Y_l), R_i] \quad , \quad i \geq 2 \end{aligned}$$

Let us recall that the R_i are assumed to be independent. So we have

$$f_{R_1,\ldots,R_r}(x_1, \ldots, x_r) = f_{R_1}(x_1) \times \cdots \times f_{R_r}(x_r)$$

The Y_i are associated with the order statistic of T : $T_{(1:r)} \leq \ldots \leq T_{(r:r)}$.
¿From now on $i \in \{1, \cdots, r\}$ will also denote the index of the ith order statistic : $T_{(i:r)}$ when we consider r claims.

We only consider the case of pure premiums.
The exact premium for an excess of loss reinsurance with pro rata capita and pro rata temporis reinstatements is :

$$P_{Re}^{det} = \frac{\mathbb{E}\min(S_R, 2L)}{1 + \sum_{n=1}^{r} \mathbb{P}(N = n) \sum_{i=1}^{n} \frac{c}{L} \mathbb{E}(1 - T_{(i:n)}) \mathbb{E} Y_i} \tag{4.10}$$

Unfortunately, evaluating $\mathbb{E}Y_i$ even for i small is very time consuming. Indeed, the random variables $Y_1, \ldots, Y_r$ are highly correlated. So, evaluating $\mathbb{E}Y_i$ requires an i-multiple integral (or sum in the discrete case).
The expectations $\mathbb{E}T_{(i:r)}$ are not too complicated. Indeed, it is not difficult to show that the distribution of the i-th order statistic of T is given by

$$F_{T_{(i:r)}}(u) = \sum_{l=i}^{r} \frac{r!}{l!(r-l)!} F_T^l(u)(1 - F_T(u))^{r-l}$$

Then we immediately have

$$\mathbb{E}(T_{(i:r)}) = \int_0^{\infty} \left(1 - \sum_{l=i}^{r} \frac{r!}{l!(r-l)!} F_T^l(u)(1 - F_T(u))^{t-l}\right) du$$

Note that a recursive formula was given by David (1970) in order to compute these expectations :

$$i\mathbb{E}(T_{(i+1:r)}) + (r-i)\mathbb{E}(T_{(i:r)}) = r\mathbb{E}(T_{(i:r-1)})$$

For the practical evaluation of a premium, we have to assume a maximum number of claims r. This is not a problem as these treaties are essentially used for natural perils for which we do not expect a high frequency.
For our numerical example, we will use a truncated Poisson random variable. Let us choose to truncate at $r = 4$. The truncation is such that $\mathbb{P}(N \geq 4)$ accumulates at 4.
In order to make different modelizations of the time pattern, we will work with a Beta distribution and let its parameters vary.
We find

i	1	2	3	4
$\mathbb{E}Y_i$	9.52	8.30	5.84	3.42

Table 4.14: $\mathbb{E}Y_i$ for the 30 xs 20 cover

r/i	1	2	3	4
1	0.5	–	–	–
2	0.58593	0.41407	–	–
3	0.62889	0.5	0.37110	–
4	0.65630	0.54666	0.45333	0.34369

Table 4.15: $\mathbb{E}(1 - T_{(i:r)})$ for $a = b = 5$

	P $a = b = 5$	P $a = 5, b = 0.5$	P no pro rata temporis
$c = 50\%$	8.80	9.32	8.25
$c = 100\%$	8.23	9.20	7.32
$c = 150\%$	7.73	9.07	6.57

Table 4.16: Premiums for the 30 xs 20 cover

Obviously the premiums with no pro rata temporis clause are smaller as there will be no discount for remaining time for the potential reinstatement premiums.

A natural approximation is now proposed.
It is clear that a trivial simplification of the formula (4.10) is to consider that each claim occurs at time $\mathbb{E}T$. Then we assume a mean time remaining equal to $\mathbb{E}(1-T)$ for each claim. This is a very natural approximation.

$$
\begin{aligned}
P_{Re}^{det} &= \frac{\mathbb{E}\min(S_R, 2L)}{1 + \mathbb{E}(1-T)\sum_{n=1}^{r} \mathbb{P}(N=n) \sum_{i=1}^{n} \frac{c}{L}\mathbb{E}Y_i} \\
&= \frac{\mathbb{E}\min(S_R, 2L)}{1 + \mathbb{E}(1-T)\frac{c}{L}\mathbb{E}\min(L, \max(0, S_R - L))}
\end{aligned}
$$

Clearly this formula is far easier to use than the exact formula.
Most of the time, the reinstatements are payable at 100%. We will henceforth work with $c = 100\%$. We have made some comparisons between the exact and the approximate formulae. The tables give the fraction $\frac{\text{exact}}{\text{approximated}}$ for λ and a and b varying as well as the different layers. We find

	$a = b = 5$	$a = 0.5, b = 5$	$a = 5, b = 0.5$	$a = 0.5, b = 0.5$	$a = b = 1$
$\lambda = 0.5$	0.9995	0.9996	0.9996	0.9988	0.9990
$\lambda = 1$	0.9980	0.9988	0.9986	0.9953	0.9961
$\lambda = 2$	0.9929	0.9959	0.9948	0.9836	0.9865

Table 4.17: 30 xs 20 : exact / approximate

	$a=b=5$	$a=0.5, b=5$	$a=5, b=0.5$	$a=0.5, b=0.5$	$a=b=1$
$\lambda=0.5$	0.9968	0.9981	0.9977	0.9925	0.9938
$\lambda=1$	0.9899	0.9946	0.9920	0.9766	0.9806
$\lambda=2$	0.9734	0.9867	0.9767	0.9395	0.9497

Table 4.18: 10 xs 20 : exact / approximate

	$a=b=5$	$a=0.5, b=5$	$a=5, b=0.5$	$a=0.5, b=0.5$	$a=b=1$
$\lambda=0.5$	0.9994	0.9996	0.9996	0.9987	0.9989
$\lambda=1$	0.9980	0.9988	0.9986	0.9955	0.9963
$\lambda=2$	0.9941	0.9967	0.9956	0.9863	0.9887

Table 4.19: 10 xs 30 : exact / approximate

	$a=b=5$	$a=0.5, b=5$	$a=5, b=0.5$	$a=0.5, b=0.5$	$a=b=1$
$\lambda=0.5$	0.9999	0.9999	0.9999	0.9999	0.9999
$\lambda=1$	0.9999	0.9999	0.9999	0.9997	0.9998
$\lambda=2$	0.9996	0.9998	0.9997	0.9992	0.9993

Table 4.20: 10 xs 40 : exact / approximate

It seems clear that the approximate premium is always higher than the exact premium, which is a conservative approximation.
Moreover the error does not seem to be important. This is logical.
In the case of 0 claim, there is no error.
In the case of 1 claim, $\mathbb{E}T = \mathbb{E}T_{(1:1)}$ and so there is no error either.
Thus, the error occurs in case $[N \geq 2]$, that happens with low probability in catastrophe treaties.

We note that case $10\,xs\,20$ is the worst, which is logical. Indeed, for that layer, EY_1 will be the largest because it is the layer for which the probability of having a total loss is the highest. If, moreover, $\mathbb{E}(1-T_{(1:r)})$ is quite different from $\mathbb{E}(1-T)$ (case $a=b=0.5$), the approximation is bad. This is almost the worst conceivable case for our approximation.

We will now prove that the above mentionned conjecture is always true.

Lemma 4.1 *For the order statistic $T_{(1:r)}, \ldots, T_{(r:r)}$, we have*

$$\sum_{i=1}^{r} \mathbb{E}T_{(i:r)} = r\mathbb{E}T$$

Proof :
This result immediately comes from the almost sure equality :

$$T_{(1:r)} + \cdots + T_{(r,r)} = T_1 + \cdots + T_r \quad \text{a.s.}$$

■

Theorem 4.2 *With pro rata temporis reinstatement premiums, the approximate premium is always larger than the exact premium.*

Proof :
We have to prove that

$$\sum_{n=1}^{r} \mathbb{P}(N=n) \sum_{i=1}^{n} \frac{c}{L} \mathbb{E}(1 - T_{(i:n)}) \mathbb{E}Y_i > \mathbb{E}(1-T) \sum_{n=1}^{r} \mathbb{P}(N=n) \sum_{i=1}^{n} \frac{c}{L} \mathbb{E}Y_i$$

Let us show that the inequality is true for each term of the sum in n :

$$\sum_{i=1}^{n} \mathbb{E}(1 - T_{(i:n)}) \mathbb{E}Y_i > \mathbb{E}(1-T) \sum_{i=1}^{n} \mathbb{E}Y_i$$

Let ϵ be such that

$$\mathbb{E}T_{(\epsilon:n)} \leq \mathbb{E}T \leq \mathbb{E}T_{(\epsilon+1:n)}$$

We have

$$\sum_{i=1}^{\epsilon} \mathbb{E}T_{(i:n)} \mathbb{E}Y_i + \sum_{i=\epsilon+1}^{n} \mathbb{E}T_{(i:n)} \mathbb{E}Y_i < \sum_{i=1}^{\epsilon} \mathbb{E}T\mathbb{E}Y_i + \sum_{i=\epsilon+1}^{n} \mathbb{E}T\mathbb{E}Y_i$$

which is equivalent to

$$\sum_{i=\epsilon+1}^{n} (\mathbb{E}T_{(i:n)} - \mathbb{E}T)\mathbb{E}Y_i < \sum_{i=1}^{\epsilon} (\mathbb{E}T - \mathbb{E}T_{(i:n)})\mathbb{E}Y_i$$

and the last inequality is always true because
i) $\mathbb{E}Y_1 \geq \mathbb{E}Y_2 \geq \cdots \geq \mathbb{E}Y_n$
ii) $\sum\limits_{i=\epsilon+1}^{n} (\mathbb{E}T_{(i:n)} - \mathbb{E}T) = \sum\limits_{i=1}^{\epsilon} (\mathbb{E}T - \mathbb{E}T_{(i:n)})$ because of lemma 4.1. ■

We now propose some extensions :

In the case of an aggregate deductible AD, the random variables R_i are transformed into the random variables Z_i :

$$\begin{aligned}
W_1 &= \max(0, R_1 - AD) \\
W_i &= \max(0, R_i - \max(0, (AD - \sum_{l=1}^{i-1} R_l))) \quad , \quad i \geq 2 \\
Z_1 &= W_1 \\
Z_i &= \min[\max(0, L - \sum_{l=1}^{(i-1)} Z_l), W_i] \quad , \quad i \geq 2 \\
P_{Re}^{det} &= \frac{\mathbb{E}\min(S_R, 2L)}{1 + \sum_{n=1}^{r} \mathbb{P}(N=n) \sum_{i=1}^{n} \frac{c}{L}\mathbb{E}(1 - T_{(i:n)})\mathbb{E}Z_i}
\end{aligned}$$

For $\lambda = 1$ and $c = 100\%$, we have evaluated the ratio $\frac{\text{exact}}{\text{approximate}}$ for two aggregate deductibles. We find

	$a=b=5$	$a=0.5, b=5$	$a=5, b=0.5$	$a=b=5$	$a=b=1$
$AD=15$	1.0036	1.0021	1.0025	1.0085	1.0070
$AD=45$	1.0005	1.0003	1.0003	1.0013	1.0010

Table 4.21: 30 xs 20 : exact / approximate

We can see that the approximation is not conservative. This is due to the fact that the $\mathbb{E}Z_i$ are no longer ordered. Note that for some AD, the approximation is conservative.
Note also that even if the approximation is not conservative, it is better than without an aggregate deductible. This is due to the fact that $\mathbb{E}Z_i < \mathbb{E}Y_i$ for the small values of i, i.e. the ones with the larger probability because the frequency is low. So the error decreases. For our numerical example we find

i	1	2	3	4
$\mathbb{E}Y_i$	9.52	8.30	5.84	3.42
$\mathbb{E}Z_i$	0	1.21	3.57	5.40

Table 4.22: $\mathbb{E}Y_i$ and $\mathbb{E}Z_i$ for the 30 xs 20 cover

Now let us extend the formulae to the case of two reinstatements payable at prices c_1 and c_2. The case of multiple reinstatements is similar but has no practical interest. This is the reason why we limit ourselves to the case of two reinstatements.
The random variables Y_i are transformed into the random variables Z_i. Note that the random variables Z_i take into account the price of the reinstatement.

$$\begin{aligned}
Z_1 &= R_1 \\
Z_2 &= c_1(L - R_1) + c_2(R_1 + R_2 - L) \quad \text{if } R_1 + R_2 \geq L
\end{aligned}$$

$$
\begin{array}{rcl}
 & & c_1 R_2 \quad \text{else} \\
Z_k & = & c_1 R_k \quad \text{if } \sum_{i=1}^{k} R_i \leq L \\
 & & c_2 \max(0, 2L - \sum_{i=1}^{k-1} R_i) \quad \text{if } \sum_{i=1}^{k} R_i \geq 2L \\
 & & c_1 \max(0, L - \sum_{i=1}^{k-1} R_i) + c_2[\min(L, \sum_{i=1}^{k-1} R_i) + R_k - L] \quad \text{else} \\
P_{Re}^{det} & = & \frac{\mathbb{E}\min(S_R, 2L)}{1 + \sum_{n=1}^{r} \mathbb{P}(N = n) \sum_{i=1}^{n} \frac{1}{L}\mathbb{E}(1 - T_{(i:n)})\mathbb{E}Z_i}
\end{array}
$$

As a numerical example, we take $a = b = 1$ and $\lambda = 1$. We find

c_1	0	0	1	2	1	1	2	2
c_2	1	2	0	0	1	2	1	2
ratio	0.97	0.97	0.99	1.02	0.99	1.00	1.02	1.02

Table 4.23: 10 xs 20 : exact / approximate

We observe that the approximation cannot be controlled.

4.3 Ruin of the Cedent in the presence of reinstatements

4.3.1 Ruin probability and algorithms

Let us define the surplus of an Insurance Company in discrete time :

$$U_t = u + ct - (S_1 + \cdots + S_t)$$

where u is the initial surplus, c the earned premium, t the year and S_i the aggregate claims distribution of year i. We assume that the S_i are iid.
The ruin probability of the company is given by

$$\psi(u, t) = \mathbb{P}(\exists i | U_i < 0, i = 1, 2, \ldots, t)$$

De Vylder and Goovaerts (1988) gave the following recursive algorithm for the computation of $\psi(u, t)$:

$$
\begin{array}{rcl}
\psi(u, 1) & = & 1 - F_S(u + c) \\
\psi(u, t) & = & 1 - F_S(u + c) + \int_0^{u+c} \psi(u + c - y, t - 1) dF_S(y)
\end{array}
$$

In order to use this algorithm it is necessary that the distribution of S should be arithmetic. We will assume it is the case. Moreover it needs that c and u should also be arithmetic. In practice this will not be the case. A solution is then to give the ruin probabilities by linear interpolation :

$$\psi(u+c-j,n-1) \sim (u+c-j-(\underline{u+c-j}))\psi(\overline{u+c-j},n-1)-(u+c-j-(\overline{u+c-j}))\psi(\underline{u+c-j},n-1)$$

where $\underline{x}$ denotes the largest integer contained in x whereas $\overline{x}$ denotes the smallest integer larger than x.
These problems of rounding can be solved with the algorithm of Klugman, Panjer and Willmot (1998) which is based on the brute force convolution formula.
Let us define a generalized surplus at time t :

$$\begin{aligned} U(u,t) &= u + ct + \sum_{i=1}^{t} C_i - \sum_{i=1}^{t} S_i \\ &= U(u,t-1) + c + C_t - S_t \\ &= U(u,t-1) + W(t) \end{aligned}$$

where C_i denotes any cash flow other than the collection of premiums and losses. We will use it as an interest on the positive reserve. It might also be used as a payment of dividend.
Klugman, Panjer and Willmot (1998) defined a second process

$$\begin{aligned} W^*(t) &= 0 \quad \text{if} \quad U^*(t-1) < 0 \\ &= W(t) \quad \text{if} \quad U^*(t-1) \geq 0 \\ U^*(t) &= U^*(t-1) + W^*(t) \end{aligned}$$

Then we have

$$\psi(u,t-1) = \mathbb{P}(U^*(t-1) < 0)$$

and the distribution of the non-negative surplus is given by

$$f(j) = \mathbb{P}(U^*(t-1) = u_j) \quad , \quad j = 1, \ldots, n$$

where u_n is the maximal value for $U^*(t-1)$.
Let us define

$$g(j,k) = \mathbb{P}(W(t) = w_{j,k} | U^*(t-1) = u_j)$$

Then Klugman, Panjer and Willmot (1998) show that

$$\begin{aligned} \psi(u,t) &= \psi(u,t-1) + \sum_{j=1}^{n} \sum_{w_{j,k} < -u_j} g(j,k) f(j) \\ \mathbb{P}(U^*(t) = x) &= \sum_{j=1}^{n} \sum_{w_{j,k} + u_j = x} g(j,k) f(j) \end{aligned}$$

As mentionned in Klugman, Panjer and Willmot (1998), these formulae, although intimidating are easy to implement.
A problem arises when the number of u_j values becomes large. In that case, a good way to reduce the calculation-time is to round the u values to some multiple of a span h.
We will use both algorithms in the sequel.

4.3.2 Risk quantities of interest and comparisons of treaties

There are other interesting quantities if we want to compare two treaties.
For the case of infinite time ruin probability we have the Lundberg's bound

$$\lim_{t\to\infty} \psi(u,t) \le e^{-Ru} \tag{4.11}$$

where R, the adjustment coefficient, is given by the only strictly positive solution of

$$e^{-cr}\mathbb{E}[e^{rS}] = 1 \tag{4.12}$$

The gain of the Insurance Company is

$$G = c - S$$

The expected gain is

$$\mathbb{E}G = c - \mathbb{E}S$$

The variance of the gain of the Insurance Company is

$$\mathbb{V}arG = \mathbb{V}arS$$

Now let us compare the different treaties of the previous section.
We continue with the numerical example of section 4.2.1.
We assume that the Cedent loading is 50% and the Reinsurer loading is 100%. The premiums are calculated according to the expected value principle.
We find

c	$k = 0$	1	2	3
0%	0.1019	0.1142	0.1223	0.1252
50%		0.1064	0.1070	0.1065
100%		0.1008	0.0972	0.0953
150%		0.0965	0.0906	0.0880
$1st$: 100% $2nd$: 0%			0.1064	
$1st$: 0% $2nd$: 100%			0.1068	
$\mathbb{E}G$	4.9758	4.6799	4.6395	4.6353

Table 4.24: Adjustment coefficient and expected gain

We can make the following comments :

1 the expected gain diminishes with the number of reinstatements because the loading of the Reinsurer is higher than the loading of the Cedent.

2 if we have the choice only between paid reinstatements at 100%, then we will obviously choose 1 reinstatement because this situation gives the higher adjustment coefficient with the higher expected gain.

3 if we need 2 or 3 reinstatements, that are payable, we will always choose 2 for the same reason.

4 for a given number of reinstatements, that are constant, the adjustement coefficient decreases when the percentage increases

Comments 2 and 3 are not general and show the interest for the Cedent to make the calculations in order to choose the best treaty according to the two criterions : best expected gain; best adjustment coefficient.
Comment 4 can be shown to be always true :

Theorem 4.3 *Let*

$$\begin{aligned} S_{\text{Ced},0} &= S_C + \max(0, S_R - (k+1)L) + c_0 p_0 \min(k, \frac{S_R}{L}) + p_0 \\ S_{\text{Ced},1} &= S_C + \max(0, S_R - (k+1)L) + c_1 p_1 \min(k, \frac{S_R}{L}) + p_1 \end{aligned}$$

p_i and R_i are associated with c_i according to $S_{Ced,i}$.
The p_i are calculated with the expected value principle.
If

$$c_0 < c_1$$

Then

$$R_0 > R_1$$

Proof

Let

$$\begin{aligned} g_0(S_R)|S_C &= S_C + \max(0, S_R - (k+1)L) + c_0 p_0 \min(k, \frac{S_R}{L}) + p_0 \\ g_1(S_R)|S_C &= S_C + \max(0, S_R - (k+1)L) + c_1 p_1 \min(k, \frac{S_R}{L}) + p_1 \\ f(S_R) &= \min(k, \frac{S_R}{L}) \end{aligned}$$

We use Ohlin's (1969) lemma with the random variable $S_R|S_C$. The hypotheses of Ohlin's lemma are verified :

$$\begin{aligned} g_0(x)|S_C & \quad \text{is increasing with } x \\ g_1(x)|S_C & \quad \text{is increasing with } x \\ \mathbb{E}[g_0(S_R)|S_C] &= \mathbb{E}[g_1(S_R)|S_C] \end{aligned}$$

We have

$$\begin{aligned} g_0(S_R)|S_C &\leq g_1(S_R)|S_C \quad S_R \geq x_0 \\ g_0(S_R)|S_C &\geq g_1(S_R)|S_C \quad S_R \leq x_0 \end{aligned}$$

with

$$x_0 = f^{-1}\left(\frac{p_0 - p_1}{c_1 p_1 - c_0 p_0}\right)$$

Then by Ohlin's lemma we have

$$\mathbb{E}[e^{rg_0(S_R)}|S_C] \leq \mathbb{E}[e^{rg_1(S_R)}|S_C]$$

Multiplying both sides by $\mathbb{P}[S_C = s_C]$ and summing on s_c gives the desired result. ■

Now let us assume that there are two reinsurance layers. The second one is $4\,xs\,10$. Let us assume one free reinstatement for each layer. We find :

$p_1 = 3.5101$	$r = 0.1242$
$p_2 = 1.1971$	$\mathbb{E}G = 4.0813$

Table 4.25: Reinsurance premiums, expected gain and adjustment coefficient 1

Now let us assume a paid reinstatement at 100% for each layer. We have three cotations. Reinsurer 1 gives the cotation by application of the expected value premium principle. You might observe cotations 2 and 3 on the market.

	Cotation 1	Cotation 2	Cotation 3
p_1	2.5719	2.8	2.4
p_2	1.0494	0.8	1.24
$p_1 + p_2$	3.6213	3.60	3.64
r	0.1050	0.1040	0.1057
$\mathbb{E}G$	4.0813	4.0545	4.0985

Table 4.26: Reinsurance premiums, expected gain and adjustment coefficient 2

We can make the following observations :

1 the adjustement coefficient is better when the reinstatements are free ; this is connected with theorem 4.3.

2 with the paid reinstatements, if we look only at $p_1 + p_2$ we would choose cotation 2. However this is a nonsense because cotation 2 has the worst expected gain as well as the worst adjustment coefficient. On the contrary, the best treaty for the cedent is given by cotation 3 even if for this cotation $p_1 + p_2$ is the highest. This shows that it is a nonsense to aggregate the reinsurance premiums of different layers when we have to compare different cotations. The structure of the agreement is such that the expected gain and the adjustement coefficient should be kept in mind.

An interesting exercise is to compare the influence of the structure of the layers. Hereunder we give the figures for the double layer structure (T1 and T2) described above as well as the figures for a simple layer structure (T : $8\,xs\,6$). The Reinsurers apply the expected value principle. We find

	2 rein @ 100% on T1 1 rein @ 100% on T2	2 free rein on T1 1 free rein on T2	3 rein @ 100% on T1 1 rein @ 100% on T2	3 free rein on T1 1 free rein on T2
p_1	2.49591	3.59103	2.48419	3.59928
p_2	1.04941	1.19715	1.04373	1.19992
$p_1 + p_2$	3.54532	4.78818	3.52792	4.79920
r	0.10093	0.13261	0.09846	0.13568
$\mathbb{E}G$	4.04103	4.04103	4.03551	4.03550
	1 rein @ 100% on T	1 free rein on T	2 rein @ 100% on T	2 free rein on T
p	3.75916	4.76885	3.69682	4.79867
r	0.10783	0.13030	0.10301	0.13586
$\mathbb{E}G$	4.05149	4.05143	4.03651	4.03640

Table 4.27: Reinsurance premiums, expected gain and adjustment coefficient 3

This table shows that if the reinstatements are payable @ 100%, it is better for the Cedent to work with one layer whereas with free reinstatements, a choice has to be made.
The combination of two reinstatements (paid @ 100% or free) on the low layer and one on the high layer are less interesting than two reinstatements on one large layer.
Once again we observe that the reinsurance premiums are not a good criterion in order to see which is the best cover for the Cedent.

The same kind of calculations can be performed with the multivariate individual risk model. Indeed let us assume a fire portfolio. With the notations of the individual risk model we have:

		g_1	g_2	g_3	g_4
	14				0.1
	12				0.1
	10			0.125	0.1
	8			0.125	0.1
	6			0.125	0.1
	5			0.125	0.1
	4		0.25	0.125	0.1
	3		0.25	0.125	0.1
	2	0.5	0.25	0.125	0.1
	1	0.5	0.25	0.125	0.1
q_1	0.001	200	140	120	100
q_2	0.002	170	140	120	100
q_3	0.003	100	140	120	100

Table 4.28: Fire portfolio

The loading of the cedent is 50%.
The loading of the reinsurer is 100%.

The premiums are calculated according to the expected value premium principle.
The reinsurer sells excess of loss cover $8\,xs\,6$.
The pure reinsurance premium is given by :

0 reinst. 1.61962	1 free reinst. 1.73527	2 free reinst. 1.73987	3 free reinst. 1.74
	1 paid reinst. @ 100% 1.44311	2 paid reinst. @ 100% 1.42975	3 paid reinst. @100% 1.42917

Table 4.29: Reinsurance premiums

The fact that the premium for paid reinstatements diminishes with the number of reinstatements is not illogical because it would be exceptional that the layer might be completely hit more than two times. So with two or more reinstatements, the ceding company has more capacity but it will pay more reinstatements premiums for a capacity that will most probably not be necessary.

Let us now consider the expected gain and the adjustment coefficient of the Cedent for different types of excess of loss covers. We find :

	0 reinst.	1 free reinst.	2 free reinst.	3 free reinst.
$\mathbb{E}G$ R	3.76538 0.127647	3.64973 0.14688	3.64513 0.150515	3.645 0.150802
		1 paid reinst. at 100%	2 paid reinst. at 100%	3 paid reinst. at 100%
$\mathbb{E}G$ R		3.64973 0.12643	3.64513 0.12322	3.645 0.12266

Table 4.30: Expected gain and adjustment coefficient

We can make the following observations :

1 The covers with 2 or 3 paid reinstatements are clearly not interesting

2 The free reinstatements are more interesting than the paid ones. This result has been established in general in Theorem 4.3.

3 Clearly a choice has to be made between 0 reinstatement that gives higher expected gain but a poor adjustement coefficient and 1 free reinstatement that has the opposite properties.

The same results are found using the Poisson approximation with the bivariate Panjer's algorithm :

	0 rein.	1 free rein.	2 free rein.	3 free rein.
$\mathbb{E}G$		3.64989	3.64524	3.64509
R		0.14694	0.150308	0.150478
		1 paid rein. @ 100%	2 paid rein. @ 100%	3 paid rein. @ 100%
$\mathbb{E}G$		3.6499	3.64525	3.64511
R		0.126349	0.123041	0.122429

Table 4.31: Expected gain and adjustment coefficient with the approximation of the collective risk model

Note that in practice the calculation will be done with the reinsurance premiums given by the Reinsurer and not with the ones calculated by ourselves with the expected value premium principle. As most of the time several Reinsurers make cotations, it can give arbitrage opportunities to the user of such models.

Obviously multiple layers may be treated with the trivial multivariate extension of the individual risk model.

In order to make the calculations so as to find the adjustment coefficient, it was only necessary to compute the bivariate aggregate distribution of the portfolio for the risks g_3 and g_4 because the other risks never reach the layer. For them the moment generating function is sufficient in order to make the calculations.
In order to give an application of the result of section 3.3.2, we compare the bounds and the true errors for the Compound Poisson approximation of

		f_3					f_4								
		0	1	2	3	4	0	1	2	3	4	5	6	7	8
	14														
	12														
	10														
	8														
	6	0.125		0.125		0.125	0.1		0.1		0.1		0.1		0.1
	5	0.125					0.1								
	4	0.125					0.1								
	3	0.125					0.1								
	2	0.125					0.1								
	1	0.125					0.1								
q_1	0.001	120					100								
q_2	0.002	120					100								
q_3	0.003	120					100								

Table 4.32: Part of the portfolio to be reinsured

We find

$$\begin{aligned} d_{TV}(S^{Ind}, S^{CP}) &= 0.0005942 \quad \leq \quad \text{bound} = 0.00308 \\ d_K(S^{Ind}, S^{CP}) &= 0.0004117 \quad \leq \quad \text{bound} = 0.00154 \end{aligned}$$

We see that the approximation is very satisfactory. This is not a surprise as the q_j are very small.

4.3.3 Organization of the calculations

In this section we first study the number of multiplications required to find the multivariate distribution $(S_1, \ldots, S_n)$ when using the multivariate Panjer's algorithm:

$$\sum_{x_1=0}^{s_1} \cdots \sum_{x_n=0}^{s_n} (3(\min(x_1, m_1) + 1) \cdots (\min(x_n, m_n) + 1) + 1) \tag{4.13}$$

where $(m_1, \ldots, m_n)$ are the maximal values taken by $(X_1, \ldots, X_n)$
$(s_1, \ldots, s_n)$ are the maximal values for which $(S_1, \ldots, S_n)$ are evaluated
An upper bound for (4.13) is

$$3(s_1 + 1) \cdots (s_n + 1)(m_1 + 1) \cdots (m_n + 1) \tag{4.14}$$

For our trivariate example, we had

X_1	:	part of the Cedent	$m_1 = 6$
X_2	:	part of the Reinsurer, low layer	$m_2 = 4$
X_3	:	part of the Reinsurer, high layer	$m_3 = 4$
S_C	:	aggregate part for which the Cedent is liable	$s_1 = 64$
S_{R1}	:	aggregate part for which the first layer Reinsurer is liable	$s_2 = 44$
S_{R2}	:	aggregate part for which the second layer Reinsurer is liable	$s_3 = 34$

The number of multiplications needed was :

Exact	46 290 821
Upper bound	53 849 250

for a precision given by

$$\sum_{i=0}^{64} \sum_{j=0}^{44} \sum_{k=0}^{34} \mathbb{P}(S_1 = i, S_{R1} = j, S_{R2} = k) = 0.999999905971$$

Obviously it is possible to take advantage of the particular structure of the distribution of the random vector (X_1, X_2, X_3).
For our particular case, the number of multiplications is given by

$$\sum_{x_1=1}^{s_1} [3(1 + \min(x_1, m_1) + 1] + \sum_{x_1=m_1}^{s_1} \sum_{x_2=0}^{s_2} [3((m_1 + (1 + \min(x_2, m_2))) + 1]$$

$$+\sum_{x_1=m_1}^{s_1}\sum_{x_2=m_2}^{s_2}\sum_{x_3=1}^{s_3}[3((m_1+(1+m_2+(1+\min(x_3,m_3))+1]$$

An upper bound is given by

$$s_1(3m_1+4)+(s_1-m_1+1)(s_2+1)(3(m_1+m_2)+4)$$
$$+(s_1-m_1+1)(s_2-m_2+1)s_3 3(m_1+m_2+m_3)+7)$$

With this organization, the number of multiplications needed was :

Exact	4 074 885
Upper bound	4 121 732

Obviously, in the particular case Poisson ($a=0$), the recursion is simplified. The number of multiplications is given by :

$$\sum_{x_1=1}^{s_1}[3\min(x_1,m_1)+1]+\sum_{x_1=m_1}^{s_1}\sum_{x_2=1}^{s_2}[3\min(x_2,m_2)+1]+\sum_{x_1=m_1}^{s_1}\sum_{x_2=m_2}^{s_2}\sum_{x_3=1}^{s_3}[3\min(x_3,m_3)+1]$$

An upper bound is given by

$$s_1(3m_1+1)+(s_1-m_1+1)s_2(3m_2+1)+(s_1-m_1+1)(s_2-m_2+1)s_3(3m_3+1)$$

With this organization, the number of multiplications needed was :

Exact	1 059 513
Upper bound	1 104 162

Now for our purpose that consists in evaluating the adjustment coefficient of a Ceding Company when it buys an excess of loss cover with reinstatements, let us remind that we only need the moment generating function of S_{Ced}.
If we restrict ourselves to a counting distribution that is Poisson distributed, we have the following property :
Let X be the severity of claims
Let N be Poisson distributed
Let D_1 be the lower deductible.
We immediately have

$$\begin{aligned} S &= S_1+S_2 \\ S &= Y_1+\cdots+Y_{N_1}+Z_1+\cdots+Z_{N_2} \end{aligned}$$

with $Y_1=_d X|X\leq D_1$
$N_1=\mathbb{I}_{X_1\leq D_1}+\cdots+\mathbb{I}_{X_N\leq D_1}$
$Y_2=_d X|X>D_1$
$N_2=\mathbb{I}_{X_1>D_1}+\cdots+\mathbb{I}_{X_N>D_1}$
S_1 independent of S_2

The moment generating function (mgf) of S_{Ced} is then the product of the mgf of S_1 and of the mgf a function of $(Z_1,\ldots,Z_{N_2})$.
Then we save time of calculation at two levels :

- The distribution of S_1 need not be calculated
- The multivariate distribution based on $(Z_1, \ldots, Z_{N_2})$ will be evaluated more rapidly because the mean of N_2 is lower that the one of N. Note that most of the time the difference will be very important.

If, as it was the case in our trivariate example, the high claims liable to the cedent are completely cut by the reinsurance, then the multivariate distribution based on $(Z_1, \ldots Z_{N_2})$ is simplified because the C which is part of the Cedent is degenerated at D_1, which reduces the dimensionality of the problem.
With the following hypotheses

C	:	part of the Cedent ; degenerated at 6
R_1	:	part of the Reinsurer, low layer ; $m_2 = 4$
R_2	:	part of the Reinsurer, high layer ; $m_3 = 4$
S_C^Z	:	aggregate claims distribution for which the Cedent is liable ; $s_1 = 48 = 8 \times 6$
$S_{R_1}^Z$	:	aggregate claims distribution for which the first layer Reinsurer is liable ; $s_2 = 30$
$S_{R_2}^Z$	:	aggregate claims distribution for which the second layer Reinsurer is liable ; $s_3 = 25$

The frequency of the claims hitting the reinsurance layers is $\lambda = 0.54$.
Using this technique, the number of multiplications needed was

$$\sum_{x_1=1}^{8} \sum_{x_2=0}^{s_2} \sum_{x_3=0}^{s_3} (1 + \min(x_2, m_2))(1 + \min(x_3, m_3)) + 1$$

An upper bound is given by

$$8(s_2 + 1)(s_3 + 1)((m_2 + 1)(m_3 + 1) + 1)$$

With this organization, the number of multiplications needed was :

Exact	145 648
Upper bound	167 648

for a precision given by

$$\sum_{i=0}^{48} \sum_{j=0}^{30} \sum_{k=0}^{25} \mathbb{P}(S_C^Z = i, S_{R_1}^Z = j, S_{R_2}^Z = k) = 0.999999989229$$

4.3.4 Some results on stochasting ordering

This section is mainly based on Walhin (2000c).
In this section we study some results about stochasting ordering. Our aim is to find upper and lower bounds on S_{Ced} in some sense. These bounds will be used in order to find bounds on the quantities to be calculated : expected gain, variance of the gain, probability of ruin and adjustment coefficient.
We first introduce the notion of stochastic order, that is well known in actuarial sciences.

Definition 4.1 *Let us assume two random variables X and Y. We say that X is smaller than Y in the sense of the stochastic order when*

$$X \preccurlyeq_{st} Y \Leftrightarrow \mathbb{P}[X \leq x] \geq \mathbb{P}[Y \leq x] \quad , \quad \forall x \in \mathbb{R}$$

Note that an equivalent definition is

$$X \preccurlyeq_{st} Y \Leftrightarrow \mathbb{P}[X < x] \geq \mathbb{P}[Y < x] \quad , \quad \forall x \in \mathbb{R}$$

Definition 4.2 *Let us assume a random variable X. Let $h > 0$ be a given span. We define two approximations of X, namely X^{low} and X^{up} :*

$$\begin{aligned} f_{X^{low}}(0) &= F_X(h-0) \\ f_{X^{low}}(xh) &= F_X(xh+h-0) - F_X(xh-0) \quad x = 1, 2, \ldots \\ f_{X^{up}}(0) &= 0 \\ f_{X^{up}}(xh) &= F_X(xh+0) - F_X(xh-h+0) \quad x = 1, 2, \ldots \end{aligned}$$

where $f_X(x)$ denotes the probability function of X.

Definition 4.3 *Let I_i (resp. I_i^{low} and I_i^{up}) be Bernoulli random variables with mean q_i (resp. q_i^{low} and q_i^{up}).*
We will assume

$$q_i^{low} \leq q_i \leq q_i^{up}$$

¿From definitions 4.2 and 4.3, and with the notations of section 1.1.1, it is clear that we have

$$I_i^{low} X_i \preccurlyeq_{st} I_i X_i \preccurlyeq_{st} I_i^{up} X_i \tag{4.15}$$

Moreover, the stochastic order has some interesting properties. We will use some of them (see e.g. Goovaerts et al. (1990)):

Lemma 4.2 *Let the independent random variables $U_1, U_2, \ldots, U_n$ (resp. $V_1, V_2, \ldots, V_n$) be such that $U_i \preccurlyeq_{st} V_i \quad \forall i = 1, \ldots, n$. We have*

$$\sum_{i=1}^{n} U_i \preccurlyeq_{st} \sum_{i=1}^{n} V_i$$

Lemma 4.3 *Let the independent random variables $U_1, U_2, \ldots$ and N (resp. $V_1, V_2, \ldots$ and N) be such that $U_i \preccurlyeq_{st} V_i \quad \forall i = 1, 2, \ldots$. We have*

$$\sum_{i=1}^{N} U_i \preccurlyeq_{st} \sum_{i=1}^{N} V_i$$

Lemma 4.4 *If $U \preccurlyeq_{st} V$ then $f(U) \preccurlyeq_{st} f(V)$ for every non decreasing function f.*

By using equation (4.15) and lemma 4.2 (in the case of the individual risk model) or lemma 4.3 (in the case of the collective risk model) we immediately get

$$S^{low} \preccurlyeq_{st} S \preccurlyeq_{st} S^{up} \tag{4.16}$$

where we do not specify the upperscript *ind* or *coll*, the result being valid in both cases.

By using equation 4.16 and lemma 4.4, the following results are shown immediately :

Proposition 4.1 *With the notations and hypotheses of this section we have*

$$\begin{array}{rcccr} \mathbb{E}G^{low} & \geq & \mathbb{E}G & \geq & \mathbb{E}G^{up} \\ \mathbb{V}arG^{low} & \leq & \mathbb{V}arG & \leq & \mathbb{V}arG^{up} \\ R^{low} & \geq & R & \geq & R^{up} \\ \psi^{low}(u,t) & \leq & \psi(u,t) & \leq & \psi^{up}(u,t) \end{array}$$

Our aim of is to extend these formulae in a multivariate setting.

¿From now on we will adopt the following definition for increasing multivariate functions :

Definition 4.4 *Let* $\mathbf{s} = (s_1, \ldots, s_m)$ *be a multivariate vector. Let* $f(\mathbf{s})$ *be a real-valued function of* $\mathbf{s}$. *We will say that* f *is an increasing function if and only if* $f(s_1, \ldots, s_i, \ldots, s_m)$ *is increasing in* $s_i \quad \forall i = 1, \ldots, m$ *with all the other variables being fixed.*

We have borrowed the following definition of the multivariate stochastic order from Barlow and Proschan (1975).

Definition 4.5 **S** *is stochastically smaller than* **S'** *if*

$$f(\mathbf{S}) \preccurlyeq_{st} f(\mathbf{S}')$$

for all real-valued increasing functions $f(s)$. *We write* $\mathbf{S} \preccurlyeq_{st} \mathbf{S}'$

Some related concepts, also taken from Barlow and Proschan (1975) will be needed :

Definition 4.6 *A random variable* T *is stochastically increasing in random variables* $S_1, \ldots, S_m$ *if* $\mathbb{P}[T > t | S_1 = s_1, \ldots, S_m = s_m]$ *is increasing in* $s_1, \ldots, s_m$.

Definition 4.7 *Random variables* $T_1, \ldots, T_m$ *are conditionally increasing in sequence if* T_i *is stochastically increasing in* $T_1, \ldots, T_{i-1}$ *for* $i = 2, \ldots, m$.

The following theorem (Barlow and Proschan (1975)) gives us the tool to derive a multivariate stochastic order when using multivariate vectors.

Theorem 4.4 *Let* $\mathbf{U} = (U_1, \ldots, U_m)$ *be conditionally increasing in sequence.*
Let us assume that

$$U_1 \preccurlyeq_{st} V_1$$

and

$$U_j | U_1 = x_1, \ldots, U_{j-1} = x_{j-1} \preccurlyeq_{st} V_j | V_1 = x_1, \ldots, V_{j-1} = x_{j-1}\,,\ \forall x_1, \ldots, x_{j-1}\,,\ j = 2, \ldots, m$$

Then

$$\mathbf{U} \preccurlyeq_{st} \mathbf{V}$$

Theorem 4.4 is a key result because it gives sufficient conditions in order to get to the multivariate stochastic order. In the next section we will verify its hypotheses.

4.3.5 Application of the multivariate stochastic order

In this section we will show the results in a bivariate setting. Obviously they immediately extend in a multivariate setting.
For X^{low} and X^{up} we use the same definition as above with a restriction as regards the deductible and limit :

Definition 4.8 *Let h be a span with $j \in \mathbb{N}$ and $k \in \mathbb{N}$ such that $D = jh$ and $L = kh$. Let X^{up} and X^{low} be the random variables constructed as*

$$\begin{aligned}
\mathbb{P}(X^{low} = 0) &= F_X(h - 0) \\
\mathbb{P}(X^{low} = xh) &= F_X(xh + h - 0) - F_X(xh - 0) \\
\mathbb{P}(X^{up} = 0) &= 0 \\
\mathbb{P}(X^{up} = xh) &= F_X(xh + 0) - F_X(xh - h + 0)
\end{aligned}$$

We want to prove the following property :

$$S_{Ced}^{low} \preccurlyeq_{st} S_{Ced} \preccurlyeq_{st} S_{Ced}^{up}$$

Let us show each hypothesis of theorem 4.4.

Proposition 4.2

$$\begin{array}{ccccc}
C^{low} & \preccurlyeq_{st} & C & \preccurlyeq_{st} & C^{up} \\
R^{low} & \preccurlyeq_{st} & R & \preccurlyeq_{st} & R^{up}
\end{array}$$

Proof
Let us prove that

$$C^{low} \preccurlyeq_{st} C$$

If $x < D$ we have

$$\mathbb{P}[C \leq x] = \mathbb{P}[X \leq x] \geq \mathbb{P}[X^{low} \leq x] = \mathbb{P}[C^{low} \leq x]$$

If $x \geq D$ we have

$$\mathbb{P}[C \leq x] = \mathbb{P}[X \leq x + L] \geq \mathbb{P}[X^{low} \leq x + L] = \mathbb{P}[C^{low} \leq x]$$

which gives the announced result.
The other cases are similar. ■

Proposition 4.3

$$\begin{array}{ccccc}
S_C^{low} & \preccurlyeq_{st} & S_C & \preccurlyeq_{st} & S_C^{up} \\
S_R^{low} & \preccurlyeq_{st} & S_R & \preccurlyeq_{st} & S_R^{up}
\end{array}$$

Proof
By combining proposition 4.2 with lemma 4.2 (in the case of the individual risk model) or lemma 4.3 (in the case of the collective risk model) , we immediately get the proof. ■

Proposition 4.4 *(S_C, S_R) is conditionally increasing in sequence.*

Proof
We have to show that

$$\mathbb{P}[R_1 + \cdots + R_n > t | C_1 + \cdots + C_n = x]$$

is increasing with x $\forall n \in \mathbb{N}$.
This gives the result for the individual risk model and also for the collective risk model by conditioning on $N = n$.
We will use induction.
The initial step is $n = 1$: $\mathbb{P}[R > t | C = x]$ is increasing in x.
This is obviously true if we recall the definition of R and C :
For $t > L$, we have $\mathbb{P}[R > t | C = x] = 0 \quad \forall x$.
For $t \leq L$ we find
a) $x < D$: $\mathbb{P}[R > t | C = x] = 0$
b) $x = D$: $\mathbb{P}[R > t | C = x] = \mathbb{P}[X > D + t]$
c) $x > D$: $\mathbb{P}[R > t | C = x] = 1$.

Let us assume that the hypothesis is true for $n = k$. Let us prove it for $n = k + 1$:

$$\begin{aligned}
& \mathbb{P}[R_1 + \cdots + R_{k+1} > t | C_1 + \cdots + C_{k+1} = x] \\
= \; & \mathbb{P}[f(X_1) + \cdots + f(X_{k+1}) > t | X_1 - f(X_1) + \cdots + X_k - f(X_{k+1}) = x] \\
= \; & \sum_y \mathbb{P}[X_{k+1} = y] \times \\
& \mathbb{P}[f(X_1) + \cdots + f(X_k) > t - f(y) | X_1 - f(X_1) + \cdots + X_k - f(X_k) = x + f(y) - y]
\end{aligned}$$

where $f(X_i) = R_i \quad \forall i$.
However, by the inductive hypothesis, we have that
$\mathbb{P}[f(X_1) + \cdots + f(X_k) > t - f(y) | X_1 - f(X_1) + \cdots + X_k - f(X_k) = x + f(y) - y]$ is increasing in x $\forall y$.
This closes the proof. ■

Proposition 4.5

$$R^{low} | C^{low} \preccurlyeq_{st} R | C \preccurlyeq_{st} R^{up} | C^{up}$$

Proof
We have

$$\begin{aligned}
\mathbb{P}[R \leq r | C = c] &= 1 && \text{if } c < D \\
\mathbb{P}[R \leq r | C = D] &= \tfrac{\mathbb{P}[X \leq r + D]}{\mathbb{P}[D \leq X \leq D + L]} && \text{if } 0 \leq r \leq L \\
\mathbb{P}[R \leq r | C = c] &= 0 && \text{if } c > D, r < L \\
\mathbb{P}[R \leq r | C = c] &= 1 && \text{if } c > D, r \geq L
\end{aligned}$$

The same relations are true for $(R^{up} | C^{up})$ and $(R^{low} | C^{low})$.
Let us prove

$$R | C \preccurlyeq_{st} R^{up} | C^{up}$$

We now have only to demonstrate that

$$\frac{\mathbb{P}[X \leq r + D]}{\mathbb{P}[D \leq X \leq D + L]} \geq \frac{\mathbb{P}[X^{up} \leq r + D]}{\mathbb{P}[D \leq X^{up} \leq D + L]}$$

For the numerator we have

$$\mathbb{P}[X \leq r + D] \geq \mathbb{P}[X^{up} \leq r + D]$$

because

$$X \preccurlyeq_{st} X^{up}$$

For the denominator we have

$$\mathbb{P}[D \leq X \leq D + L] \leq \mathbb{P}[D \leq X^{up} \leq D + L]$$

because

$$\mathbb{P}[X \leq D + L] = \mathbb{P}[X^{up} \leq D + L]$$

by Definition 4.8 and

$$\mathbb{P}[X < D] \geq \mathbb{P}[X^{up} < D]$$

because

$$X \preccurlyeq_{st} X^{up}$$

∎

Proposition 4.6

$$S_R^{low} | S_C^{low} \preccurlyeq_{st} S_R | S_C \preccurlyeq_{st} S_R^{up} | S_C^{up}$$

Using Proposition 4.5 and the independence hypothesis between the X_i we have the following relations :

$$\begin{aligned}
R_1 | C_1 = x &\preccurlyeq_{st} R_1^{up} | C_1^{up} = x \\
R_2 | C_2 = y &\preccurlyeq_{st} R_2^{up} | C_2^{up} = y \\
R_1 | C_1 = x, C_2 = y &=_{st} R_1 | C_1 = x && (4.17) \\
R_1^{up} | C_1^{up} = x, C_2^{up} = y &=_{st} R_1^{up} | C_1^{up} = x && (4.18) \\
R_2 | C_1 = x, C_2 = y &=_{st} R_2 | C_2 = x && (4.19) \\
R_2^{up} | C_1^{up} = x, C_2^{up} = y &=_{st} R_2^{up} | C_2^{up} = x && (4.20)
\end{aligned}$$

By the closure property (lemma 4.2) of the convolution we have

$$R_1 | C_1 = x + R_2 | C_2 = y \preccurlyeq_{st} R_1^{up} | C_1^{up} = x + R_2^{up} | C_2^{up} = y \tag{4.21}$$

Combining equations (4.17), (4.18), (4.19) , (4.20) and (4.21) we find

$$R_1 | C_1 = x, C_2 = y + R_2 | C_1 = x, C_2 = y \preccurlyeq_{st} R_1^{up} | C_1^{up} = x, C_2^{up} = y + R_2^{up} | C_1^{up} = x, C_2^{up} = y$$

As a particular case we find

$$R_1 + R_2 | C_1 + C_2 = x + y \preccurlyeq_{st} R_1^{up} + R_2^{up} | C_1^{up} + C_2^{up} = x + y$$

By extension we find

$$\sum_{i=1}^{n} R_i | \sum_{i=1}^{n} C_i \preccurlyeq_{st} \sum_{i=1}^{n} R_i^{up} | \sum_{i=1}^{n} C_i^{up} \tag{4.22}$$

Now we have the extension of the closure property for random convolution :

$$\begin{aligned} \mathbb{P}[S_R \leq x | S_C = y] &= \int_{n\in\mathbb{N}} \mathbb{P}[\sum_{i=1}^{N} R_i \leq x | \sum_{i=1}^{N} C_i = y, N = n] dF_N(n) \\ &\geq \int_{n\in\mathbb{N}} \mathbb{P}[\sum_{i=1}^{N} R_i^{up} \leq x | \sum_{i=1}^{N} C_i^{up} = y, N = n] dF_N(n) \\ &= \mathbb{P}[S_R^{up} \leq x | S_C^{up} = y] \end{aligned}$$

which gives the desired result. ■

Now we can prove the main result of this section.

Theorem 4.5 *If f is an increasing function and*

$$S_{Ced} = f(S_C, S_R)$$

then

$$S_{Ced}^{low} \preccurlyeq_{st} S_{Ced} \preccurlyeq_{st} S_{Ced}^{up}$$

By combining theorem 4.4 and propositions 4.3 , 4.4 and 4.6 we immediately have

$$(S_C^{low}, S_R^{low}) \preccurlyeq_{st} (S_C, S_R) \preccurlyeq_{st} (S_C^{up}, S_R^{up})$$

As $S_{Ced} = f(S_C, S_R)$ with f an increasing function, we immediately get the proof by using Definition 4.5. ■

4.3.6 Numerical application in the collective risk model

In this section we will use the algorithm of De Vijlder and Goovaerts (1988) in order to find the finite time ruin probability.
The numerical example we use is based on the following hypothesis :

- The company has initial surplus u
- The company faces small claims of which the annual aggregate claims distribution is normally distributed with mean $\mu = 30$ and standard deviation $\sigma = 9$
- The company faces annually large claims with the following characteristics : the number of large claims is Poisson distributed with mean $\lambda = 0.75$. The claim amount distribution of the large claims is limited Pareto distributed :

$$F_X(x) = \frac{A^{-\alpha} - x^{-\alpha}}{A^{-\alpha} - B^{-\alpha}} \quad , \quad A \leq x \leq B$$

  with $A = 5$, $B = 150$ and $\alpha = 1.5$. The number of claims is assumed to be independent of the claim amounts.

- The small claims are assumed to be independent of the large claims.
- The premium income of the Company is calculated according to the expected value principle with loading $\xi = 25\%$:

$$P = (1 + 25\%)(30 + \frac{1}{2}(8.8961 + 9.6461)) = 49.09$$

where 8.8961 is the expected value of X^{low} while 9.6461 is the expected value of X^{up}.

- A Reinsurer sells excess of loss protection with different deductibles ($D = 5$, $D = 50$, $D = 100$) with one reinstatement free or at 100% additional premium. The Reinsurer calculates the premium according to the PH transform principle with loading $\rho = 1.25$.

The reinsurance premiums are given in the next table. They are also calculated as the mean of the reinsurance premiums with X^{low} and X^{up}.

	145 xs 5	100 xs 50	50 xs 100
1 free reinst.	8.8524	1.4584	0.2779
1 paid reinst.@100%	8.3440	1.4375	0.2764

Table 4.33: Loaded reinsurance premiums

As a comparison the pure reinsurance premiums are given in the following table

	145 xs 5	100 xs 50	50 xs 100
1 free reinst.	5.5212	0.5494	0.0800
1 paid reinst.@100%	5.3180	0.5464	0.0799

Table 4.34: Pure reinsurance premiums

The following tables give some characteristics if one works with X^{low} and X^{up}.

	$145\,xs\,5$		$100\,xs\,50$		$50\,xs\,100$	
	inf	sup	inf	sup	inf	sup
$\mathbb{E}G$	6.4871	6.4871	8.5433	9.2740	9.2469	9.9930
$\mathbb{V}arG$	9.9875	9.9875	15.6274	16.1162	17.537	18.032
R	0.1243	0.1243	0.0435	0.0476	0.0311	0.0333
$\psi(1,20)$	0.0043	0.0073	0.0507	0.0639	0.0489	0.0608
$\psi(2,20)$	0.0106	0.0199	0.0799	0.1061	0.0798	0.1025
$\psi(3,20)$	0.0151	0.0303	0.0970	0.1335	0.1001	0.1313
$\psi(4,20)$	0.0179	0.0378	0.1077	0.1524	0.1140	0.1519
$\psi(5,20)$	0.0197	0.0431	0.1149	0.1658	0.1236	0.1672
$\psi(10,20)$	0.0224	0.0544	0.1287	0.1974	0.1441	0.2050
$\psi(1,100)$	$8\,10^{-8}$	$1\,10^{-7}$	$7\,10^{-5}$	0.0001	0.0011	0.0015
$\psi(2,100)$	$1\,10^{-7}$	$2\,10^{-7}$	0.0002	0.0004	0.0025	0.0036
$\psi(3,100)$	$2\,10^{-7}$	$3\,10^{-7}$	0.0005	0.0010	0.0039	0.0057
$\psi(4,100)$	$3\,10^{-7}$	$4\,10^{-7}$	0.0008	0.0017	0.0051	0.0078
$\psi(5,100)$	$4\,10^{-7}$	$5\,10^{-7}$	0.0011	0.0024	0.0062	0.0098
$\psi(10,100)$	$9\,10^{-7}$	$2\,10^{-6}$	0.0020	0.0054	0.0093	0.0169

Table 4.35: 1 free reinstatement

We can make following observations :

1 Whether the reinstatement is paid or free has practically no effect. We will therefore concentrate on the case with one paid reinstatement.

2 The $145\,xs\,5$ option is by far the best from the point of view of protection. However, the cost is very high. We will therefore only comment on the other two options.

3 We note that the ruin probability after one year and two years is smaller with the $50\,xs\,100$ option whereas the contrary is true for time horizon ≥ 3 year for the initial reserve $u = 20$.

4 With an initial reserve $u = 100$ the difference of the ruin probabilities between the two options is by far larger.

4.3.7 Numerical application in the individual risk model

Let us assume a portfolio with payments in case of death X_i. For some risks, the payments are doubled in case of an accidental death. This is denoted by d. The risks are classified by age layer of 5 years.

Let us assume that the probability of an accidental death is not well known. One hesitates between 0.1 (model A) and 0.2 (model B).
Let us assume that a reinsurance cover protecting all claims over 3 is bought.
The technical direct insurance premium is calculated as the mean of the pure premium with an accidental death probability of 0.1 and 0.2 plus a fluctuation loading of 20%. The reinsurance

	$145\,xs\,5$		$100\,xs\,50$		$50\,xs\,100$	
	inf	sup	inf	sup	inf	sup
$\mathbb{E}G$	6.6560	6.6998	8.5562	9.2872	9.2480	9.9941
$\mathbb{V}arG$	10.1854	10.1623	15.6523	16.1413	17.5394	18.0345
R	0.1209	0.1223	0.0433	0.0477	0.0311	0.0333
$\psi(1, 20)$	0.0052	0.0108	0.0507	0.0640	0.0489	0.0608
$\psi(2, 20)$	0.0124	0.0292	0.0799	0.1066	0.0798	0.1025
$\psi(3, 20)$	0.0176	0.0445	0.0970	0.1343	0.1001	0.1313
$\psi(4, 20)$	0.0209	0.0559	0.1078	0.1533	0.1140	0.1519
$\psi(5, 20)$	0.0229	0.0643	0.1150	0.1670	0.1236	0.1672
$\psi(10, 20)$	0.0263	0.0841	0.1288	0.1991	0.1441	0.2052
$\psi(1, 100)$	$2\,10^{-6}$	$2\,10^{-6}$	$7\,10^{-5}$	0.0001	0.0011	0.0015
$\psi(2, 100)$	$4\,10^{-6}$	$5\,10^{-6}$	0.0002	0.0005	0.0025	0.0037
$\psi(3, 100)$	$6\,10^{-6}$	$7\,10^{-6}$	0.0005	0.0011	0.0039	0.0059
$\psi(4, 100)$	$8\,10^{-6}$	$1\,10^{-5}$	0.0008	0.0018	0.0051	0.0080
$\psi(5, 100)$	$1\,10^{-5}$	$1\,10^{-5}$	0.0011	0.0025	0.0062	0.0099
$\psi(10, 100)$	$2\,10^{-5}$	$3\,10^{-5}$	0.0020	0.0057	0.0093	0.0171

Table 4.36: 1 reinstatement paid @ 100%

premium is calculated according to the same rule. It gives

$$\begin{aligned} P &= 7.67 \\ P^{Re} &= 3.11 \end{aligned}$$

We will also assume that for commercial reasons the Reinsurer gives a no claim bonus of 30% of the reinsurance premium.
Let us define the claim amount distributions under model A (resp. B) : X_A (resp. X_B). We immediately have

$$X_A \preccurlyeq_{st} X_B$$

This implies that

$$\begin{array}{ccccc} \mathbb{E}G_A & \geq & \mathbb{E}G & \geq & \mathbb{E}G_B \\ \mathbb{V}arG_A & \leq & \mathbb{V}arG & \leq & \mathbb{V}arG_B \\ R_A & \geq & R & \geq & R_B \end{array}$$

We find

$$\begin{aligned} \mathbb{E}G_A &= 1.33 \\ \mathbb{E}G_B &= 1.29 \\ \mathbb{V}arG_A &= 11.70 \\ \mathbb{V}arG_B &= 11.86 \\ R_A &= 0.1927 \\ R_B &= 0.1854 \end{aligned}$$

Evaluating the ruin probability is more complicated or time-consuming. We have chosen to apply the algorithm of Klugman, Panjer and Willmot (1998) because it allows an interest

	q_x	1	1d	2	2d	3	3d	4	4d	5	5d	10	10d	15	15d	20	20d
$20-24$	0.0009	10	8	5	1	1	1	1	0	0	0	1	0	0	0	0	0
$25-29$	0.0011	15	17	17	13	10	9	13	8	5	4	3	1	1	0	0	0
$30-34$	0.0014	12	15	13	3	17	16	4	8	1	0	0	2	0	0	0	0
$35-39$	0.0019	11	3	12	17	0	2	7	4	0	2	1	1	0	0	1	0
$40-44$	0.0027	7	11	12	10	3	9	7	4	1	3	1	2	2	0	1	2
$45-49$	0.0041	7	4	7	2	2	3	1	0	0	1	2	1	1	0	1	1
$50-54$	0.0063	12	7	13	1	12	10	8	3	4	7	4	2	0	1	2	2
$55-59$	0.0098	3	1	7	3	9	4	2	0	1	0	2	1	2	0	0	0

Table 4.37: Life portfolio

rate on the reserve and it does not need to round the premium.
On the other side a rounding has to be processed in order to limit the number of u values in the algorithm. The rounding has been done from $t = 3$ and it is an upper and lower rounding giving rise to bounds on the ruin probability. This is the reason why I give bounds on the ruin probability for both models A and B.

We have run the algorithm with an interest on the reserve equal to 4% and an initial reserve $u = 10$.
The results for model A are

t	1	2	3	4	5	6	7
$\psi_A^{low}(u,t)$	0.0021	0.0074	0.0117	0.0147	0.0167	0.0180	0.0188
$\psi_A^{up}(u,t)$	0.0021	0.0074	0.0143	0.0208	0.0263	0.0308	0.0344

Table 4.38: Ruin for model A and $i = 4\%$

The results for model B are

t	1	2	3	4	5	6	7
$\psi_B^{low}(u,t)$	0.0022	0.0079	0.0125	0.0157	0.0179	0.0193	0.0203
$\psi_B^{up}(u,t)$	0.0022	0.0079	0.0153	0.0222	0.0282	0.0331	0.0371

Table 4.39: Ruin for model B and $i = 4\%$

As we know from section 4.3.5 that

$$\psi_A(u,t) \leq \psi_B(u,t)$$

we have that the true ruin probabilities are between $\psi_A^{low}(u,t)$ and $\psi_B^{up}(u,t)$.

The same exercise has been processed with no interest rate on the reserve. This gives the classical ruin probability :

t	1	2	3	4	5	6	7
$\psi^{low}(u,t)$	0.0039	0.0118	0.0234	0.0257	0.0307	0.0347	0.0376
$\psi^{up}(u,t)$	0.0039	0.0118	0.0194	0.0359	0.0480	0.0593	0.0694

Table 4.40: Ruin for model A and $i = 0\%$

We immediately note the impact of using an interest rate of 4% on the reserve.
The results for model B are

t	1	2	3	4	5	6	7
$\psi_B^{low}(u,t)$	0.0041	0.0124	0.0206	0.0274	0.0328	0.0371	0.0405
$\psi_B^{up}(u,t)$	0.0041	0.0124	0.0248	0.0382	0.0512	0.0632	0.0742

Table 4.41: Ruin for model B and $i = 0\%$

Now let us assume that the Reinsurer does not give a profit commission ($\alpha = 0\%$). With no interest on the reserve the results become :

$$\begin{aligned}
\mathbb{E}G_A &= 0.77 \\
\mathbb{E}G_B &= 0.75 \\
\mathbb{V}arG_A &= 9.80 \\
\mathbb{V}arG_B &= 9.90 \\
R_A &= 0.1386 \\
R_B &= 0.1330
\end{aligned}$$

t	1	2	3	4	5	6	7
$\psi_A^{low}(u,t)$	0.0040	0.0129	0.0225	0.0313	0.0390	0.0455	0.0509
$\psi_B^{up}(u,t)$	0.0042	0.0135	0.0289	0.0474	0.0668	0.0861	0.1048

Table 4.42: Ruin with $i = 0\%$ and $\alpha = 0\%$

We notice that the profit commission has an enormous effect on the expected gain as well as on the risk measures.

Chapter 5

A review of some counting distributions useful in insurance

In this chapter, the main emphasis will be laid on the Hofmann Distribution that will be completely described because it is the author's opinion that it is really a good candidate for overdispersed data sets. Other types of counting variables will be included either because we will use them (e.g. the Nonparametric Mixed Poisson Distribution) or for the sake of completeness. Indeed very recently some new types of counting variables were introduced in literature and it is interesting to find some points of comparison.
Note that there exists specific literature about the discrete distributions, namely Johnson, Kotz and Kemp (1992) in which a lot of discrete distributions have been listed. The present chapter is different in the sense that it makes an effort of globalization with the introduction of the Hofmann Distribution. For the rest the other distributions are only sketched. Note that the Weighted Poisson Distribution and the Poisson Gontcharov Distribution date from later than the book of Johnson, Kotz and Kemp (1992).

We note that a common way of constructing new distributions is to use the notion of mixing or compounding. There is a lot of confusion about these notions in the literature. We will not enter in details. The notions we adopt are typically accepted in actuarial sciences (see for example the book of Panjer and Willmot (1992)). We will say that

- If the distribution of a random variable X depends on a parameter θ and that this parameter θ is a realization of the random variable Θ, then the random variable Y is a mixed random variable with mixing variable Θ. The cumulative density function of Y is given by
$$F_Y(x) = \int_{\theta=0}^{\theta=\infty} \mathbb{P}(X = x|\Theta = \theta) dF_\Theta(\theta)$$
- Let us assume two independent random variables X and N. Let us assume a sample $X_1, X_2, \ldots$ of X. Then the random variable S is a compound random variable with compounding distribution N and primary distribution X :
$$S = X_1 + \cdots + X_N$$

It often happens in statistical literature that these definitions interfer or that other terms are used (see Kemp (1967) or Johnson, Kotz and Kemp (1992) for details).

A typical motor portfolio will be fitted by maximum likelihood for each type of distribution we will encounter. This portfolio came from Switzerland in 1961 and was published by Bühlmann (1970) and used by, among others, Lemaire (1985), Tremblay (1992), Denuit (1997) and Walhin and Paris (1999b). The data set gives, as usual, the number of policyholders for each number of claims reported to the company. This portfolio will be our reference portfolio in this chapter as well as in chapter 7 for the numerical examples.

Number of accidents	Number of policyholders
0	103704
1	14075
2	1766
3	255
4	45
5	6
6	2

Table 5.1: Reference portfolio

For the grouping rule in order to find the χ^2 statistic, we try to obey the rule B in Lemaire (1995), i.e. each theoretical frequency is at least 1 and 80% of the theoretical frequencies are at least 5. Note also that for the point of comparison we will take 6 for the number of classes during all our numerical example with this data set because if we group the classes differently according to the fitted distribution, we can find results that are not comparable.
Note that the types of models we will study are also useful in other areas like e.g. ecology, biology, genetics, physics, operations research, ...

5.1 The Hofmann Distribution

This section is mainly based on Walhin and Paris (2000e).
In Grandell (1997), the author argued that the Hofmann Distribution has no practical relevance. The aim of this section is to show the contrary.

Count data occur in many practical problems. When only randomness looks present, they seem to be described by a Poisson distribution which is the first choice for nonnegative integer valued random variables. In many cases, this choice is rejected by the usual χ^2 test for goodness of fit. It is immediately observed that the Poisson model underestimates the variance because overdispersion occurs (the variance is greater than the mean), indicating that the population heterogeneity has not been taken into account by this model and its single parameter.
This indicates that more parameters are needed to describe the distribution of the data.
As the sequence of signs of the difference between observed and expected frequencies under the Poisson distribution is + , - , + , the result of Shaked (1980) indicates that it is natural to try a Mixed Poisson Distribution.

Moreover infinite divisibility of the data is observed and especially whenever the frequency is low, excess of zeroes often arise.
Many models have been built in order to try to solve each part of the problem separately. Here we suggest a methodology and a model to solve it in the aggregate. We prove that this model is the most general one of this type.

In many cases, data are collected in a period of time which is eliminated when taken as unity. Nevertheless the model has to be extended to stochastic processes. Our model is also adapted to this situation, to data reported on several periods and to the problem met in industry where the defaults are reported on periods of different lengths.

Let us analyze the fit of our reference portfolio by a Poisson Distribution :

	Obs	Po	Obs-Po
0	103704	102629.55	+
1	14075	15921.95	−
2	1766	1235.07	+
3	255	63.87	+
4	45	2.48	+
5	6	0.07	+
6	2	0.00	+
7		0.00	
8		0.00	

Table 5.2: Poisson fit and the signs rule

If we try to fit a Poisson Distribution with parameter λ, estimated by maximum likelihood estimation : $\hat{\lambda} = \bar{N} = 0.15514$, the usual χ^2 statistic of goodness of fit, with 3 degrees of freedom is 2550.93 and leads to the rejection of the Poisson Distribution. Moreover the sequence of signs is the Shaked one. This suggests a Poisson mixture.
Some other tests are useful.
The usual asymptotic Poisson overdispersion test (see Gart (1975) and Böhning (1994)) is based on

$$\sqrt{\frac{n-1}{2}}\left(\frac{S^2}{\bar{N}}-1\right) \sim N(0,1)$$

In our case the value of the statistic is 38.14, indicating a strong presence of overdispersion.
The rejection of the Poisson Distribution within the infinitely divisibile distributions :

$$\begin{aligned} H0 &: N \sim Po \\ H1 &: N \text{ is infinitely divisible} \end{aligned}$$

can be tested with the corrected version of the statistic (see Gupta, Móri, Székely (1994))

$$k_2k_4 - k_3^2 \sim N(0, \frac{8\lambda^4}{n}(1+12\lambda+3\lambda^2))$$

where k_j is the k-statistic of order j of Fisher and λ is the maximum likelihood estimator of the Poisson parameter.

Let us define $M_r = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^r$. The k-statistic write

$$\begin{aligned}
k_1 &= \bar{X} \\
k_2 &= \frac{n}{n-1}M_2 \\
k_3 &= \frac{n^2}{(n-1)(n-2)}M_3 \\
k_4 &= \frac{n^2}{(n-1)(n-2)(n-3)}[(n+1)M_4 - 3(n-1)M_2^2]
\end{aligned}$$

With our numerical example we find an observed statistic $k_2k_4 - k_3^2 = 0.0149$ and the variance of the normal distribution $\frac{8\lambda^4}{n}(1+12\lambda+3\lambda^2) = 1.1344\,10^{-7}$. The reduced observed statistic is 44.32 indicating a strong presence of infinite divisibility. This suggests a Compound Poisson Model by a result of Feller (1971).
The test for deviation in a single cell (see Rao (1973) p395) can also be applied. In the particular case of the Poisson distribution, the test statistic for the deviation in the zero cell is based on

$$\frac{N_0 - ne^{-\hat{\lambda}}}{\sqrt{ne^{-\hat{\lambda}}(1 - e^{-\hat{\lambda}} - \hat{\lambda}e^{-\hat{\lambda}})}} \sim N(0,1)$$

The observed value, 32.18, indicates a strong deviation in the zero cell. Due to Jensen's inequality, the Mixed Poisson Distribution has the following related property :

$$\int_0^\infty e^{-\lambda}dU(\lambda) \geq e^{-\int_0^\infty \lambda dU(\lambda)}$$

5.1.1 The model

Let $N(t)$ be a counting (pure birth) process on the interval $(0,t]$ $(N(0) = 0)$. The main problem in risk theory is the determination of the probability function of the random sum

$$S(t) = X_1 + \cdots + X_{N(t)}$$

where the X_i are iid and represent the cost of claims, $N(t)$ is the number of claims in $(0,t]$ and is supposed to be independent of the X_i.
The two main objectives are :

- the evaluation of the probability function of $S_{N(t)}$ by a procedure which avoids the use of convolutions

- the evalutation of the intensity of the process

  $$\mathbb{E}[N(t+1) - N(t)|N(t)]$$

  on which the premium for the period $[t, t+1]$ is based.

As the data are counts of accident insurance policies reporting exactly a number of claims during a particular year, it is natural from the preceding considerations to choose for $N(t)$ a Mixed Poisson Process characterized by

$$\Pi(n,t) = \mathbb{P}[N(t) = n] = \int_0^\infty e^{-\lambda t}\frac{(\lambda t)^n}{n!}dU(\lambda) \tag{5.1}$$

In fact the population heterogeneity is unobserved. We assume that the population consists of several subpopulations of Poisson type. It is however not possible to observe the individual memberships. Therefore we assume that the heterogeneity is adequately described by a density function $u(\lambda)$.
¿From a probabilistic point of view, in order to construct a Mixed Poisson Process, it is sufficient to specify the distribution function U of the random variable Λ. This leads to analytical expressions for the probabilities $\Pi(n,t)$ from equation 5.1, which can also be rewritten in the form

$$\Pi(n,t) = (-1)^n \frac{t^n}{n!} \frac{d^n}{dt^n} \Pi(0,t) \tag{5.2}$$

This relation indicates that it is sufficient to know $\Pi(0,t)$ to be able to deduce the probability law of $N(t)$.

¿From a statistical point of view, it is $N(t)$ which is observed and, usually, U is unknown. Therefore we can take another approach. We know that $\Pi(0,t)$ is the Laplace transform of U and determines not only U but also the probability function of $N(t)$. Thus it looks interesting to obtain $\Pi(0,t)$ directly. Due to the Bernstein-Widder theorem (see Feller (1971)) $\Pi(0,t)$ is a completely monotonic function and estimating such a function with infinitely numerous restrictions is a difficult task.
As we have observed infinite divisibility, let us choose

$$\Pi(0,t) = e^{-\theta(t)}$$

with $\theta(t)$ a Bernstein function :

$$\begin{aligned} \theta(t) &\geq 0 \\ \theta(0) &= 0 \\ \frac{d}{dt}\theta(t) & \quad \text{completely monotonic} \end{aligned}$$

Feller (1971) shows that this corresponds to the case where the probability distribution (characterized by U) of the random variable Λ is infinitely divisible.
The probability generating function of $N(t)$ is easily obtained by

$$\begin{aligned} \psi_{N(t)}(u) &= \sum_{n=0}^{\infty} \Pi(n,t) u^n \\ &= \int_0^{\infty} e^{-\lambda t(1-u)} dU(\lambda) \\ &= e^{-\theta(t-tu)} \end{aligned} \tag{5.3}$$

Taking the derivative of the logarithm of $\psi_{N(t)}$ we obtain

$$\frac{d}{du} \ln \psi_{N(t)}(u) = t \frac{d}{du} \theta(t-tu)$$

where the function $\frac{d}{du}\theta(t-tu)$, taken as a function of u, is absolutely monotonic. Thus it can be expanded in series as

$$\frac{d}{du}\theta(t-tu) = \sum_{n=0}^{\infty} r_n(t) u^n$$

By equating identical powers of u in the two parts of

$$\begin{aligned}
\frac{d}{du}\psi_{N(t)}(u) &= \psi_{N(t)}(u)t\frac{d}{du}\theta(t-tu) \\
\sum_{n=1}^{\infty} n\Pi(n,t)u^{n-1} &= \sum_{n=0}^{\infty}\Pi(n,t)u^n t\sum_{m=0}^{\infty} r_m(t)u^m
\end{aligned}$$

we obtain a recursive relationship

$$\begin{aligned}
\Pi(0,t) &= e^{-\theta(t)} && (5.4)\\
(n+1)\Pi(n+1,t) &= t\sum_{j=0}^{\infty} r_j(t)\Pi(n-j,t) && (5.5)
\end{aligned}$$

where $r_j \geq 0 \quad \forall j$ because $\theta(t-tu)$ is an absolutely monotonic function (using Bernstein's theorem).
These relationships are identical to those of Steutel (1973) (see also Katti (1967)) which characterize infinitely divisible probability distributions for discrete random variables.
Note that, as Λ is infinitely divisible, Maceda (1948) shows that $N(t)$ is also infinitely divisible. Then, by Feller (1971), $N(t)$ is a Compound Poisson Distribution :

$$N(t) = \Xi_1 + \cdots + \Xi_{L(t)}$$

where $L(t)$ is Poisson distributed and the Ξ_i are iid and independent of $L(t)$.

In this section we have used the following theorems :

Theorem 5.1 (Lévy's Theorem (Feller(1971))**)** *A discrete probability distribution on the nonnegative integers is infinitely divisible if and only if it is Compound Poisson distributed.*

Theorem 5.2 (Maceda's Theorem (Maceda(1948))**)** *Consider a Mixed Poisson Distribution where the mixing distribution has nonnegative support. The Mixed Poisson Distribution is infinitely divisible if and only if the mixing distribution is infinitely divisible*

Theorem 5.3 (Bernstein-Widder's Theorem (Feller(1971))**)** *The function f defined on the support $[0,\infty)$ is the Laplace transform of a probability distribution if and only if f is completely monotonic with $f(0)=1$.*

Theorem 5.4 (Construction theorem (Feller (1971))**)** *The function f is the Laplace transform of an infinitely divisible probability distribution if and only if $f = exp(-g)$ where the derivative of g is completely monotonic with $g(0)=0$.*

5.1.2 A remarkable family of probability distributions

A probability distribution for $N(t)$, constructed as in the previous section, can be interpreted as a Mixed Poisson Process or as a Compound Poisson Process. To use it, it is sufficient to know $\frac{d}{dt}\theta(t)$. Let us take the family characterized by

$$\frac{d}{dt}\theta(t) = \frac{p}{(1+ct)^a} \qquad (5.6)$$

which corresponds to the class of infinitely divisible Mixed Poisson Process introduced by Hofmann (1955) and used by Thyrion (1961), Kestemont and Paris (1985) and Walhin and Paris (1999b). By integration, we immediately find

$$\begin{aligned}\theta(t) &= pt \quad \text{if } a=0 \\ &= \frac{p}{c}\ln(1+ct) \quad \text{if } a=1 \\ &= \frac{p}{c(1-a)}[((1+ct))^{1-a}-1]\end{aligned}$$

Thus, the parameter a distinguishes among the different distributions. For $a=0$ we find the ordinary Poisson Distribution, for $a=\frac{1}{2}$ the Poisson Inverse Gaussian Distribution , for $a=1$ the Negative Binomial Distribution and for $a=2$ the Polya-Aeppli Distribution. Likewise, if $a\to\infty$ and $c\to 0$ such that $ac\to b$ we find the Neyman Type A Distribution for which

$$\theta(t)=\frac{p}{b}(1-e^{-bt})$$

For notation purpose we will write $N(1)\sim Ho(p,c,a)$.

5.1.3 Separation of the family

In our family of probability distributions, we can distinguish two parts :

1 the one for which

$$\lim_{t\to\infty}\theta(t)=\infty$$

corresponding to $0\le a\le 1$.
This leads to

$$\lim_{t\to\infty}\Pi(0,t)=0=\lim_{\lambda\to 0}U(\lambda)$$

from Tauberian results (see Feller (1971)).
This result explains why the Poisson Distribution ($a=0$) and the Negative Binomial Distribution ($a=1$) have received such a great attention in counting distributions.

2 the one for which

$$\lim_{t\to\infty}\theta(t)=d>0$$

corresponding to $a>1$ and yielding

$$\lim_{t\to\infty}\Pi(0,t)=\lim_{\lambda\to 0}U(\lambda)=e^{-d}$$

In this case the random variable Λ has a point mass at the origin.
Letting

$$\begin{aligned}\theta(t) &= d\left[1-(1-\frac{\theta(t)}{d})\right] \\ Q(0,t) &= 1-\frac{\theta(t)}{d}\end{aligned}$$

we see that

$$\begin{aligned} Q(0,0) &= 0 \\ Q(0,t) &\geq 0 \\ Q(0,t) &\text{ is } \quad \text{completely monotonic} \end{aligned}$$

and

$$\Pi(0,t) = e^{-d[1-Q(0,t)]}$$

is also a convenient choice.
In the particular case $a > 1$ in the Hofmann Process

$$Q(0,t) = (1+ct)^{1-a}$$

is the Laplace transform of a Gamma random variable. This offers another way to deduce the Hofmann Distribution in the case $a > 1$ only.

5.1.4 Properties of the family

¿From (5.6), we find

$$\frac{d^n}{dt^n}\theta(t) = (-1)^{n-1}pc^{n-1}\frac{\Gamma(a+n-1)}{\Gamma(a)}(1+ct)^{1-a-n}$$

and from (5.4)-(5.5) we have

$$\begin{aligned} \Pi(0,t) &= e^{-\theta(t)} \\ (n+1)\Pi(n+1,t) &= \frac{pt}{(1+ct)^a}\sum_{j=0}^{n}\frac{\Gamma(a+j)}{j!\Gamma(a)}\left(\frac{ct}{1+ct}\right)^j \Pi(n-j,t) \quad , \quad n \geq 0 \end{aligned} \tag{5.7}$$

Unfortunately, this formula requires the knowledge of all the probabilities. If $a = 0, \frac{1}{2}, 1$ or 2, we have simplified formulae for calculation. The cases $a = 0$ and $a = 1$ are in the $(a, b, 0)$ class.
The case $a = \frac{1}{2}$ can be evaluated by the formula

$$\begin{aligned} \Pi(0,t) &= e^{-\frac{2p}{c}(\sqrt{1+ct}-1)} \\ \Pi(1,t) &= \frac{p}{\sqrt{1+ct}}\Pi(0,t) \\ \Pi(n,t) &= \frac{ct}{1+ct}(1-\frac{3}{2n})\Pi(n-1,t) + \frac{p^2}{n(n-1)(1+ct)}\Pi(n-2,t) \quad , \quad n \geq 2 \end{aligned}$$

The case $a = 2$ can be evaluated by the formula

$$\begin{aligned} \Pi(0,t) &= e^{-\frac{p}{1+ct}} \\ (n+1)\Pi(n+1,t) &= (\frac{p}{(1+ct)^2} + 2\frac{ct}{1+ct}n)\Pi(n,t) - (\frac{ct}{1+ct})^2(n-1)\Pi(n-1,t) \quad , \quad n \geq 0 \end{aligned}$$

given by Evans (1953). However this formula is unstable due to close numbers that have to be subtracted.

Let us note that the recursion for the particular case Neyman Type A was obtained by Beall (1940).

The factorial cumulant generating function for $N(t)$

$$\begin{aligned} \ln \mathbb{E}[(1+u)^{N(t)}] &= \ln \Pi(0,-tu) \\ &= -\theta(-tu) \end{aligned}$$

yields the factorial cumulants

$$\begin{aligned} \kappa_{[1]} &= pt \\ \kappa_{[2]} &= pcat^2 \\ \kappa_{[3]} &= pc^2a(a+1)t^3 \\ \kappa_{[4]} &= pc^3a(a+1)(a+2)t^4 \end{aligned}$$

from which we derive

$$\begin{aligned} \mathbb{E}N(t) &= pt \\ \mathbb{V}arN(t) &= pt + pcat^2 \end{aligned}$$

Except for the case $a = 0$, we always have

$$\mathbb{V}arN(t) > \mathbb{E}N(t)$$

and for fixed c, the difference increases with a.
The skewness and kurtosis, expressed in the manner proposed by Anscombe (1950), using factorial cumulants are

$$\begin{aligned} \frac{\kappa_{[3]}}{p(ac)^2} &= 1 + \frac{1}{a} \\ \frac{\kappa_{[4]}}{p(ac)^3} &= \left(1+\frac{1}{a}\right)\left(1+\frac{2}{a}\right) \end{aligned}$$

These are two decreasing functions of a.
The classical skewness and kurtosis are given by

$$\begin{aligned} \gamma &= \frac{pt(1+3act+ac^2t^2+a^2c^2t^2)}{(pt(1+act))^{3/2}} \\ \kappa &= \frac{pt(1+7act+6ac^2t^2+6a^2c^2t^2+2ac^3t^3+3a^2c^3t^3+a^3c^3t^3)}{(pt+acpt^2)^2} \end{aligned}$$

As discussed in Panjer and Willmot (1992), it is interesting to write the skewness in function of $\mu = \mathbb{E}N(t)$ and $\sigma^2 = \mathbb{V}arN(t)$:

$$\gamma = \frac{1}{\sigma^3}(3\sigma^2 - 2\mu + C\frac{(\sigma^2-\mu)^2}{\mu})$$

In the case of the Hofmann Distribution, we have

$$C = \frac{1+a}{a}$$

which is a decreasing function of a.

The equation (5.3) shows that, if $N_1(t)$ has a Hofmann Distribution with parameters p_1, c, a, and if $N_2(t)$ has a Hofmann Distribution with parameters p_2, c, a, when they are independent, their sum $N_1(t)+N_2(t)$ also follows the same type of distribution with parameters p_1+p_2, c, a.

The ordinary Poisson Distribution is a limiting case. If we consider those with the same expected value, pt, as introduced above, that is

$$p(n,t) = e^{-pt}\frac{(pt)^n}{n!} \quad , \quad n = 0, 1, 2, \dots$$

we have, because of Jensen's inequality,

$$p(0,t) \leq \Pi(0,t)$$

which explains the excess of zeroes.
We also have

$$\frac{p(1,t)}{p(0,t)} = pt \geq \frac{\Pi(1,t)}{\Pi(0,t)} = t\frac{d}{dt}\theta(t)$$

The preceding result is known since Feller (1943) :

Theorem 5.5 *Consider X a random variable with a Mixed Poisson Distribution with a non degenerate mixing distribution and Y a random variable that is Poisson distributed with the same mean as X. We have*

$$\begin{array}{ccc} \mathbb{P}(X=0) & > & \mathbb{P}(Y=0) \\ \dfrac{\mathbb{P}(X=1)}{\mathbb{P}(X=0)} & < & \dfrac{\mathbb{P}(Y=1)}{\mathbb{P}(Y=0)} \end{array}$$

For fixed t, $\frac{d}{dt}$ is a decreasing function of a and thus, $\Pi(0,t)$ is increasing in a. Then we have

$$\lim_{a\to 0} \Pi(0,t) = p(0,t)$$

However it is not possible to say if $\Pi(1,t)$ is systematically larger or smaller than $p(1,t)$.

¿From the Chebyshev inequality for integrals of decreasing functions, we have

$$\Pi(0,t+s) \geq \Pi(0,t)\Pi(0,s)$$

which is an evident result in insurance.

The Cauchy-Schwartz inequality gives

$$\frac{n\Pi(n,t)}{\Pi(n-1,t)} \leq \frac{(n+1)\Pi(n+1,t)}{\Pi(n,t)}$$

whereas these relationships are constant for the ordinary Poisson Distribution.

Let us assume that $N \sim Ho(p,c,a)$. Let us introduce the random variables :

$$\begin{aligned} Y_i &= 1 \text{ if } X_i > D \\ &= 0 \text{ if } X_i \leq D \end{aligned}$$

Let us analyse the distribution of

$$N^{'} = Y_1 + \cdots + Y_N$$

i.e. we count the number of events with characteristic $X > D$. We immediately have

$$\begin{aligned}
\psi_{N'}(u) &= \psi_N(\psi_Y(u)) \\
&= \psi_N(u\mathbb{P}(X > D) + \mathbb{P}(X \leq D)) \\
&= e^{-\theta(t - tu(1-F_X(D)) - tF_X(D))} \\
&= e^{-\theta(t(1-F_X(D)(1-u))}
\end{aligned}$$

which shows that $N^{'} \sim Ho(p(1 - F_X(D)), c(1 - F_X(D)), a)$.

5.1.5 The probability distribution of the components of $N(t)$

The representation of $N(t)$ as a Compound Poisson Process of the type

$$N(t) = \Xi_1 + \cdots + \Xi_{L(t)}$$

is easily obtained from the probability generating function of $N(t)$:

$$\psi_{N(t)}(u) = e^{-\theta(t-tu)}$$

If we write

$$\theta(t - tu) = \theta(t)\left[1 - (1 - \frac{\theta(t-tu)}{\theta(t)})\right]$$

and define

$$\psi_{\Xi(t)}(u) = 1 - \frac{\theta(t-tu)}{\theta(t)}$$

we immediately see that

$$\begin{aligned}
\psi_{\Xi(t)}(0) &= 0 \\
\psi_{\Xi(t)}(u) &= \sum_{n=1}^{\infty} p_{\Xi(t)}(n)u^n
\end{aligned}$$

where

$$\begin{aligned}
p_{\Xi(t)}(n) &= \mathbb{P}[\Xi(t) = n] \\
&= (-1)^{n-1}\frac{t^n}{n!}\frac{\frac{d^n}{dt^n}\theta(t)}{\theta(t)} \qquad (5.8)
\end{aligned}$$

This shows that the probability distribution of the Ξ_i can be immediately deduced from the knowledge of the function $\theta(t)$.
Thus $N(t)$ has probability generating function

$$\psi_{N(t)} = e^{-\theta(t)[1-\psi_{\Xi(t)}(u)]} \qquad (5.9)$$

which shows that $N(t)$ is the sum of a random number $L(t)$ of iid random variables $\Xi_i(t)$ characterized by equation (5.8) and $L(t)$ follows a non-homogeneous Poisson Process with mean $\theta(t)$.
The distribution of Ξ is summarized in the following table :

Parameter a	Distribution of $\Xi(t)$
$a = 0$	Degenerate $\mathbb{P}[\Xi(t) = 1] = 1$
$0 < a < 1$	Extended Truncated Negative Binomial
$a = 1$	Logarithmic Distribution
$a > 1$	Truncated Negative Binomial
$a \to \infty$	Truncated Poisson

Table 5.3: Candidates for the distribution of $\Xi(t)$

5.1.6 The probability function of the aggregate claims distribution

In insurance problems one is naturally interested in the compound distribution

$$S(t) = X_1 + \cdots + X_{N(t)}$$

where the X_i are iid random variables representing the claim amounts. They are assumed to be independent of the number of claims $N(t)$. They are also assumed to be arithmetic.
We have already seen that if $N(t)$ belongs to the $(a, b, 0)$ class, or more generally to the (a, b, m) class, it is possible to find the probability function of $S(t)$ recursively, without using the brute force convolution formula.

¿From now on, we will reparametrize the (a, b, m) class by the (r, s, m) class in order to avoid confusion with the a of the Hofmann Distribution.

Obviously, the Hofmann Distribution does not belong to the (r, s, m) class. However the $\Xi_i(t)$ belong to the $(r, s, 1)$ class. Indeed we have

$$\frac{p_{\Xi(t)}(n)}{p_{\Xi(t)}(n-1)} = \frac{pt}{1+ct}(1 + \frac{a-2}{n}) \quad , \quad n > 1$$

Therefore it is interesting to evaluate the probability function of $S(t)$ in a two stage procedure. We first evaluate the probability function of an intermediary random variable

$$U(t) = X_1 + \cdots + X_{\Xi(t)}$$

which is easy because $\Xi(t)$ belongs to the $(r, s, 1)$ class.
Then we find the probability function of $S(t)$:

$$S(t) = U_1 + \cdots + U_{L(t)}$$

which is also easy because $L(t)$ is Poisson distributed, which belongs to the $(r, s, 0)$ class.
In fact we have just applied the fact that

$$\begin{aligned} \psi_{S(t)}(u) &= \psi_{N(t)}(\psi_X(u)) \\ &= \psi_{L(t)}(\psi_{\Xi(t)}(\psi_X(u))) \end{aligned}$$

When applying the recursive schemes described in chapter 1, we immediately find :

$$\begin{aligned}
f_{U(t)}(0) &= 1 - \frac{\theta(t - tf_X(0))}{\theta(t)} \\
f_{U(t)}(x) &= \frac{1}{1 - \frac{ct}{1+ct} f_X(0)} \Bigg[\left(\frac{ct}{1+ct} \right) \sum_{i=1}^{x} (1 + (a-2)\frac{i}{x}) f_X(i) f_{U(t)}(x-i) \\
&\quad + p_{\Xi(t)}(1) f_X(x) \Big] \quad , \quad x \geq 1 \\
f_{S(t)}(0) &= e^{-\theta(t)(1 - f_{U(t)}(0))} \\
f_{S(t)}(x) &= \frac{\theta(t)}{x} \sum_{i=1}^{x} i f_{U(t)}(i) f_{S(t)}(x-i) \quad , \quad x \geq 1
\end{aligned}$$

In fact the probability function of $N(t)$ is also easily available via the Panjer's algorithm due to the fact that $L(t)$ is Poisson distributed :

$$\begin{aligned}
f_{N(t)}(0) &= e^{-\theta(t)} \\
f_{N(t)}(x) &= \frac{\theta(t)}{x} \sum_{i=1}^{x} i(-1)^{i-1} \frac{t^i}{i!} \frac{\theta^{(i)}(t)}{\theta(t)} f_{N(t)}(x-i) \\
&= \frac{1}{x} \sum_{i=1}^{x} i(-1)^{i-1} \frac{t^i}{i!} (-1)^{i-1} p c^{i-1} \frac{\Gamma(a+i-1)}{\Gamma(a)} (1+ct)^{1-a-i} f_{N(t)}(x-i) \\
&= \frac{pt}{(1+ct)^a} \frac{1}{x} \sum_{i=1}^{x} \frac{1}{(i-1)!} \frac{\Gamma(a+i-1)}{\Gamma(a)} \left(\frac{ct}{1+ct} \right)^{i-1} f_{N(t)}(x-i) \quad , \quad x \geq 1
\end{aligned}$$

which is equivalent to

$$(x+1) f_{N(t)}(x+1) = \frac{pt}{(1+ct)^a} \sum_{i=0}^{x} \frac{1}{i!} \frac{\Gamma(a+i)}{\Gamma(a)} \left(\frac{ct}{1+ct} \right)^i f_{N(t)}(x-i) \quad , \quad x \geq 0 \tag{5.10}$$

5.1.7 Prior probability distribution of Λ

$\Pi(0,t)$ determines completely U. The cumulant generating function for Λ is obtained by

$$\ln \mathbb{E}[e^{u\Lambda}] = \ln \Pi(0, -u) = -\theta(-u) \tag{5.11}$$

and the cumulants of Λ in the Hofmann Process are

$$\begin{aligned}
\kappa_{[1]} &= p \\
\kappa_{[j]} &= p(ac)^{j-1}(1 + \frac{1}{a})(1 + \frac{2}{a}) \ldots (1 + \frac{j-2}{a}) \quad , \quad j \geq 2
\end{aligned}$$

In particular we find

$$\begin{aligned}
\mathbb{E}\Lambda &= p \\
\mathbb{V}ar\Lambda &= pac
\end{aligned}$$

and the coefficient of variation of Λ, $\sqrt{\frac{ac}{p}}$, is an a priori measure of heterogeneity of the grouped elements.

It is possible to find the distribution of Λ. For some particular cases, Λ has a easy form. For $a=0$, Λ is degenerated. For $a=1$, Λ is Gamma distributed. For $a=\frac{1}{2}$, Λ is Inverse Gaussian distributed. For the general case, Panjer and Willmot (1992) show that the probability function of Λ is given by

$$\begin{aligned} u(\lambda) &= e^{\frac{p}{c(1-a)}}\left(\frac{c(1-a)}{p}\right)^{1-a}\frac{1}{c}e^{-\frac{x}{c}}\frac{1}{\pi}\times \\ &\sum_{k=1}^{\infty}\frac{\Gamma(k(1-a)-1}{k!}(-1)^{k-1}\left(\frac{x}{c}(\frac{c(1-a)}{p})^{1-a}\right)^{-(k(1-a)-1)}\sin((1-a)k\pi) \end{aligned}$$

This expression is cumbersome. However, and fortunately, it is not necessary for all our developments. Let us mention that the Insurer prefers to control the distribution of $N(t)$ instead of the distribution of Λ because whatever the liable driver, the Insurer pays for the claim.

For the particular case $a=\frac{1}{2}$, Λ is Inverse Gaussian distributed :

$$u(\lambda)=p(\pi c\lambda^3)^{-1/2}exp\{-\frac{(\lambda-p)^2}{c\lambda}\} \quad , \quad \lambda\geq 0$$

For the particular case $a=1$, Λ is Gamma distributed :

$$u(\lambda)=\left(\frac{1}{c}\right)^{-p/c}e^{-\lambda/c}\frac{\lambda^{p/c-1}}{\Gamma(p/c)}$$

For notation purpose, we will say that Λ is $Homix(p,c,a)$ distributed.

5.1.8 Bayesian analysis of the model

If we observe the stochastic process $N(t)$ covering the period t, can we predict the period τ which follows ? Let H_t be the observation made during $(0,t]$. Working conditionally on Λ and using the properties of the Poisson Process on disjoint intervals, we have

$$\mathbb{P}[N(t+\tau)-N(t)=n|H_t]=\frac{1}{\mathbb{P}(H_t)}\int_0^{\infty}e^{-\lambda t}\frac{(\lambda t)^n}{n!}\mathbb{P}[H_t|\Lambda]dU(\lambda)$$

¿From the Bayes formula we obtain the posterior distribution of Λ

$$dU(\lambda|H_t)=\frac{\mathbb{P}[H_t|\lambda]dU(\lambda)}{\mathbb{P}[H_t]}$$

In particular

$$\mathbb{E}[N(t+\tau)-N(t)|H_t]=\tau\mathbb{E}[\Lambda|H_t]$$

One special case occurs when $H_t=\{N(t)=k\}$. Then we find

$$\begin{aligned} \mathbb{E}[N(t+1)-N(t)|N(t)=k] &= \frac{k+1}{t}\frac{\Pi(k+1,t)}{\Pi(k,t)} \\ &= \mathbb{E}[\Lambda|N(t)=k] \end{aligned}$$

a formula we will use in chapter 7 in order to construct bonus-malus systems.

A surprising result is that

$$\mathbb{E}[\Lambda|N(t)=0]=\frac{d}{dt}\theta(t)$$

and so this regression function characterizes the distribution of the two random variables $N(t)$ and Λ.

5.1.9 Introduction of a pure random effect

As $\frac{d}{dt}\theta(t)$ is a decreasing function of t, we have

$$\lim_{t\to\infty} \frac{d}{dt}\theta(t) = 0$$

In this case, at the limit, the drivers who reported zero claim during the long period t will pay a very low premium for the period $[t, t+1]$. This is unacceptable. Indeed any driver, even the best, might by chance cause an accident. So we must apply a basic minimum premium and replace $\frac{d}{dt}\theta(t)$ by

$$\frac{d}{dt}\theta_1(t) = \delta + \frac{d}{dt}\theta(t)$$

This function $\frac{d}{dt}\theta_1(t)$ is also completely monotonic and from a general result on completely monotonic functions (see Berg and Forst (1975)) this is the most general situation of this type. We can say that the model studied in the previous sections is a particular case of the present model, which is the most general model.
The corresponding counting process $N_1(t)$ is the sum of two independent components : the first $N^*(t)$ is a simple Poisson Process with mean δt which describes the purely random part and the second $N(t)$ is a mixed infinitely divisible Poisson Process related to the behaviour of the driver. $N(t)$ and $N^*(t)$ are independent.
Note that for the particular case when $N(t)$ is Negative Binomial distributed, one speaks of the Lüders Distribution (Lüders (1934)) or Delaporte Distribution (Delaporte (1959)). In the case $N(t)$ is Neyman Type A distributed, the resulting distribution is called the Short Distribution (introduced by Cresswell and Frogatt (1963)).

We know that $N_1(t)$ is a Compound Poisson Process :

$$\Xi_1^* + \cdots + \Xi_{L(t)}^*$$

Using similar arguments than in section 5.1.6, it is not difficult to show that the probability function of the $\Xi^*(t)$ is given by

$$\begin{aligned} p_{\Xi^*(t)}(n) &= 0 \quad \text{if } n = 0 \\ &= \frac{t\frac{d}{dt}\theta(t) + \delta t}{\theta(t) + \delta t} \quad , \quad n = 1 \\ &= (-1)^{n-1}\frac{t^n}{n!}\frac{\frac{d^n}{dt^n}\theta(t)}{\theta(t) + \delta t} \quad , \quad n \geq 2 \end{aligned}$$

whereas $L(t)$ is Poisson distributed : $L(t) \sim Po(\theta(t) + \delta t)$.
Applying the Panjer's algorithm we immediately find the probability function of $N_1(t)$:

$$\begin{aligned} f_{N_1(t)}(0) &= e^{-\theta(t)-\delta t} \\ f_{N_1(t)}(1) &= \delta t + \frac{pt}{(1+ct)^a} \\ f_{N_1(t)}(x) &= \frac{f_{N_1(t)}(x-1)}{x}\left(\delta t + \frac{pt}{(1+ct)^a}\right) + \\ & \frac{pt}{x(1+ct)^a}\sum_{i=2}^{x}\frac{1}{(i-1)!}\left(\frac{ct}{1+ct}\right)^{i-1}\frac{\Gamma(a+i-1)}{\Gamma(a)}f_{N_1(t)}(x-i) \quad , \quad x \geq 2 \end{aligned}$$

whereas a two stage algorithm is also available for the evalutation of the probability function of the aggregate claims distribution :

$$S(t) = X_1 + \cdots + X_{N_1(t)}$$

Indeed we have that the Ξ^* belong to the (r,s,2) class :

$$\begin{aligned}
p_{\Xi^*(t)}(0) &= 0 \\
p_{\Xi^*(t)}(1) &= \frac{t\frac{d}{dt}\theta(t) + \delta t}{\theta(t) + \delta t} \\
\frac{p_{\Xi^*(t)}(n)}{p_{\Xi^*(t)}(n-1)} &= \frac{ct}{1+ct}\left(1 + \frac{a-2}{n}\right) \quad , \quad n = 3, 4, 5, \ldots
\end{aligned}$$

Then we easily find

$$\begin{aligned}
f_{U(t)}(0) &= 1 - \frac{\theta(t - tf_X(0)) + \delta(t - tf_X(0))}{\theta(t) + \delta t} \\
f_{U(t)}(x) &= \frac{1}{1 - \frac{ct}{1+ct}f_X(0)}\left[\frac{ct}{1+ct}\sum_{i=1}^{x}(1 + (a-2)\frac{i}{x})f_X(i)f_{U(t)}(x-i)\right. \\
&\qquad \left. + p_{\Xi^*(t)}(1)f_X(x) + (p_{\Xi^*(t)}(2) - \frac{ct}{1+ct}\frac{a}{2}p_{\Xi^*(t)}(1))f_X^{*2}(x)\right] \quad , \quad x \geq 1 \\
f_{S(t)}(0) &= e^{-\theta(t)(1 - f_{U(t)}(0))} \\
f_{S(t)}(x) &= \frac{\theta(t) + \delta t}{x}\sum_{i=1}^{x} i f_{U(t)}(i) f_{S(t)}(x-i) \quad , \quad x \geq 1
\end{aligned}$$

Some characteristics are :

$$\begin{aligned}
\mu &= \delta t + pt \\
\sigma^2 &= \delta t + pt + pcat^2 \\
\gamma &= \frac{\delta t + pt + 3acpt^2 + a(1+a)c^2pt^3}{(t(\delta + p + acpt))^{3/2}} \\
\kappa &= \frac{\delta + p + 7acpt + 6ac^2pt^2 + 6a^2c^2pt^2 + 2ac^3pt^3 + 3a^2c^3pt^3 + a^3c^3pt^3}{(t(\delta + p + acpt))^2}
\end{aligned}$$

In function of μ and σ^2, the skewness writes

$$\gamma = \frac{1}{\sigma^3}(3\sigma^2 - 2\mu + C\frac{(\sigma^2 - \mu)^2}{\mu})$$

with

$$C = \frac{1+a}{a}\frac{p+\delta}{p}$$

5.1.10 Estimation of the parameters

The Hofmann Process has three parameters, one of them distinguishes among the different probability laws. Several methods can be proposed for the estimation of the parameters. The

moment method is inadequate to nonnegative integer valued random variables. So the first method we apply is the maximum likelihood. From Hürliman (1990) we know that the sample mean is the maximum likelihood estimator of the parameter p but it is not possible to obtain explicit formulae for the estimators of a and c. They can be obtained numerically and their efficiency can be estimated.
We fix the time to $t = 1$ as the observation is on a one year basis.

With our reference portfolio we find

	Obs	Ho				
0	103704	103704.60	l	-54609.59	$as\sigma_p$	0.0012
1	14075	14072.52	χ^2	0.434	$as\sigma_c$	0.0687
2	1766	1769.26	df	2	$as\sigma_a$	0.0824
3	255	255.23	$p - value$	0.574	$as\rho(p, c)$	0.0463
4	45	41.98	p	0.15514	$as\rho(p, a)$	0.0002
5	6	7.58	c	0.3480	$as\rho(c, a)$	-0.9718
6	2	1.46	a	0.4483		
7		0.29				
8		0.06				

Table 5.4: Hofmann fit

We note that the asymptotic correlation between the estimates of p and a is almost 0. This property has been observed on every used data set. It is a good property saying that the estimates of p and a are almost independent, which strenghtens our opinion that a chooses the distribution whereas p is just the average claims frequency. The asymptotic correlation between the estimates of c and a is almost -1, which is not surprising as we know that the the product ac is part of the variance. So we expect the estimates of these two parameters to move in opposite directions.
Note that the asymptotic variance-covariance matrix has been obtained as the inverse of the observed Information matrix, that is the negative of the second partial derivative of the log-likelihood. A better way to estimate this variance-covariance matrix is to compute the Fisher Information matrix, which is based only on the first partial derivatives of the log of the $\Pi(n, 1)$. The elements of the Fisher Information matrix write

$$a_{ij}(\theta) = n \sum_{k=0}^{\infty} \Pi(k, 1)\left(\frac{\partial}{\partial \theta_i} \ln \Pi(k, 1)\right)\left(\frac{\partial}{\partial \theta_j} \ln \Pi(k, 1)\right)$$

where θ is the vector of parameters to be estimated.
Special attention should be laid here as the infinite sum will have to be truncated. With this method we find

$as\sigma_p$	$as\sigma_c$	$as\sigma_a$	$as\rho(p, c)$	$as\rho(p, a)$	$as\rho(c, a)$
0.001223	0.068904	0.082642	0.046447	0	-0.971996

Table 5.5: Asymptotic variance-covariance by the Fisher Information

Note that it is necessary to truncate the sum in k far enough in order to get convergence. In particular this is true for the $as\rho(p,a)$. Moreover derivating the $\Pi(k,1)$ with respect to the parameters p, c and a is a tedious task. Fortunately formulae giving these derivatives are given in Panjer and Willmot (1992) for the parametrization under the form of the Generalized Poisson Pascal Distribution (see section 5.1.11). Having the variance-covariance matrix within this parametrization , it is not difficult to find the variance-covariance matrix under the Hofmann parametrization (see also Panjer and Willmot (1992) page 313).
It turns out that $\hat{p}$ and $\hat{a}$ are asymptotically uncorrelated. This fact seems to be true even for very small samples and for any values of the parameters. Nevertheless this is only a conjecture. It remains to be proved.
An alternative method proposed by Kestemont and Paris (1985) with equivalent properties as the maximum likelihood is the following : estimate p by the sample mean and the parameters a and c with the relations

$$\begin{aligned} \frac{n_0}{n} &= e^{-\theta(1)} \\ \frac{n_1}{n} &= \left.\frac{d}{dt}\theta(t)\right|_{t=1} e^{-\theta(1)} \end{aligned}$$

where n_i is the number of observation in the class i.
It is possible to analyze the quality of this procedure with the two statistics

$$\begin{aligned} T &= k_2 - p(1+ac) \\ V &= k_3 - (pc^2a(a+1) + 3pca + p) \end{aligned}$$

where k_2 and k_3 are the Fisher k statistics.
With our numerical example we have

	Obs	Ho		
0	103704	103704.00	l	-54609.60
1	14075	14075.00	χ^2	0.438
2	1766	1766.78	df	2
3	255	255.39	$p-value$	0.803
4	45	42.26	T	-0.00006
5	6	7.69	V	-0.00076
6	2	1.50	p	0.15514
7		0.30	c	0.3546
8		0.06	a	0.4406

Table 5.6: Hofmann fit with the proportion estimation method

Another method of estimation is the minimum χ^2 method presented by Berkson (1980) as an alternative to maximum likelihood. We will not use this method because of the thickness of the right tail of the distribution and the difficulties related to the grouping of the classes with low frequency.

We now study the case of the Hofmann + Poisson Distribution :

	Obs	Ho + Po				
0	103704	103703.67	l	-54609.53	$as\sigma_\delta$	0.1244
1	14075	14076.17	χ^2	1.16	$as\sigma_p$	0.1244
2	1766	1763.51	df	1	$as\sigma_c$	0.3319
3	255	258.60	$p-value$	0.280	$as\sigma_a$	0.2600
4	45	42.18	δ	0.0524	$as\rho(\delta,p)$	-0.9999
5	6	7.28	p	0.1027	$as\rho(\delta,c)$	-0.9805
6	2	1.30	c	0.2581	$as\rho(\delta,a)$	0.9952
7		0.24	a	0.9119	$as\rho(p,c)$	0.9805
8		0.04			$as\rho(p,a)$	-0.9952
					$as\rho(c,a)$	-0.9947

Table 5.7: Hofmann+Poisson fit

We note that the results are very bad. The asymptotic standard deviations are high and the asymptotic correlations are near 1 in absolute value. This is due to the fact that the model finds it difficult to separate the effect of δ from p. At the limit, in model $Po + Po$, we would have a problem of identification between the two parameters. Another reason is perhaps that introducing parameter δ makes the model overparametrized and leads to instability of the estimates.
We are not surprised that the asymptotic correlation between δ and a is near 1. Indeed, as δ represents the pure Poisson part of the process, it is clear that if we give a great importance to the Poisson part, the model will tend to a great a in order to compensate the effect of δ.
As a conclusion, we can say that the extended model has to be interpreted with care due to the high sampling errors. Nevertheless, in absence of data on a longer period, maximum likelihood estimation is the best we can do.
If data are available on a longer period, parameter δ might be estimated with the asymptotic relations :

$$\begin{aligned}\lim_{t\to\infty}\frac{\Pi_1(t+1,0)}{\Pi_1(t,0)} &= e^{-\delta}\\ \lim_{t\to\infty}\frac{1}{t}\frac{\Pi_1(1,t)}{\Pi_1(0,t)} &= \delta\end{aligned}$$

5.1.11 Other parametrizations for the Hofmann Distribution in literature

In literature, the Hofmann Distribution exists with other parametrizations.
The first one is known under the name Generalized Poisson Pascal Distribution and is introduced in Panjer and Willmot (1992). The probability generating function of the Generalized Poisson Pascal Distribution is

$$\psi(u) = exp\{\mu((1-\beta(u-1))^{-r}-1)\}$$

With the following change of variables, we show that the Generalized Poisson Pascal Distribution is nothing else but the Hofmann Distribution :

$$r = a-1$$

$$\begin{aligned} \beta &= c \\ \mu &= \frac{p}{c(a-1)} \end{aligned}$$

Clearly the parametrization of the Hofmann Distribution is better. It gives immediately $\hat{p} = \bar{N}$. Moreover the asymptotic correlations between the estimates of the parameters of the Generalized Poisson Pascal Distribution is always high whereas we have seen that it is not the case for the Hofmann Distribution.
For our numerical example we find

$\hat{\mu}$	=	-0.8081
$\hat{\beta}$	=	0.3480
$\hat{r}$	=	-0.5517
$as\sigma_\mu$	=	0.0534
$as\sigma_\beta$	=	0.0687
$as\sigma_r$	=	0.0824
$as\rho(\mu,\beta)$	=	0.9947
$as\rho(\mu,r)$	=	-0.9909
$as\rho(\beta,r)$	=	-0.9719

Table 5.8: Characteristics of the estimates with the GPPD

A second parametrization is due to Hougaard et al. (1997). In this paper, the authors define a $P-G(\alpha,\delta,\theta)$ distribution as a mixture of Poisson distribution with the mixing distribution being distributed $G(\alpha,\delta,\theta)$. The family G is defined from its Laplace transform :

$$\mathbb{E}[e^{-u\Lambda}] = e^{-\frac{\delta}{\alpha}[(\theta+u)^\alpha - \theta^\alpha]}$$

The link between the parameters of the $P-G$ distribution and those of the Hofmann Distribution is :

$$\begin{aligned} a &= 1-\alpha \\ c &= \frac{1}{\theta} \\ p &= \delta\theta^{\alpha-1} \end{aligned}$$

It is possible to show that the probabilities $p(n)$ write

$$p(n) = p(0)\sum_{i=1}^{n} c_{n,i}(a)p^i c^{n-i}(1+c)^{i-ai-n} \tag{5.12}$$

with

$$\begin{aligned} c_{n,1}(a) &= \frac{\Gamma(n-1+a)}{\Gamma(a)} \\ c_{n,i}(a) &= c_{n-1,i-1}(a) + (n-1-i+ai)c_{n-1,i}(a) \\ c_{n,n}(a) &= 1 \\ p(0) &= e^{-\frac{p}{c(1-a)}((1+c)^{1-a}-1)} \end{aligned}$$

Now let us use another change of variables :

$$\begin{aligned} a &= a \\ p &= \rho \frac{e^w}{(1-e^w)^a} \\ c &= \frac{e^w}{1-e^w} \end{aligned}$$

then (5.12) becomes :

$$p(n) = p(0)e^{nw} \sum_{i=1}^{n} c_{n,i}(a)\rho^i \tag{5.13}$$

which shows that the Hofmann Distribution belongs to the exponential family with canonical parameter w when the parameters a et ρ are fixed.
The parametrization of Hougaard et al. (1997) is not natural. Its only advantage is that it shows the fact that the Hofmann Distribution belongs to an exponential family.
Moreover Hougaard (1986) has shown that for $\alpha \geq 0$, i.e. $a \leq 1$, the distribution is unimodal. It is known in the literature that for case $a > 1$, the distribution may not be unimodal.

5.2 The Nonparametric Mixed Poisson Distribution

This section is mainly based on Walhin and Paris (1999b).
We still assume a Mixed Poisson Process in order to describe the number of claims in the period $(0, t]$.
In the previous section, we assumed that the mixing distribution was parametric . We assumed that a U function had been chosen and that all that remained was to estimate the parameters.
In the present section we assume a Mixed Poisson Distribution for which we do not specify a parametric distribution $U(\lambda)$ for Λ.
In that case, Simar (1976) shows that the maximum likelihood estimate of U will be attained for a discrete distribution function $U(\lambda)$ with a maximum number m of growing points.
The probabilities $\Pi(n, t)$ are then given by

$$\Pi(n,t) = \sum_{j=1}^{r} p_j e^{-\lambda_j t} \frac{(\lambda_j t)^n}{n!} \quad , \quad n = 0, 1, 2, \ldots \tag{5.14}$$

with $\sum_{j=1}^{r} p_j = 1$, $p_j \geq 0 \quad \forall j$ and r, the number of support points, i.e. the number of homogeneous classes of risks.
We will say that $N(t) \sim NPMPD(p_1, \ldots, p_{r-1}, \lambda_1, \ldots, \lambda_r)$.
We will assume in the sequel that $0 \leq \lambda_1 < \lambda_2 \ldots < \lambda_r$.
Simar (1976) gives an algorithm to find the nonparametric maximum likelihood estimators for an automobile portfolio. Unfortunately the log-likelihood is not concave everywhere and the algorithm does not converge with certainty to the global maximum. In particular, Simar (1976) did not verify the fact that

$$\hat{\mathbb{E}N} = \bar{N} \tag{5.15}$$

in his numerical example (this property is valid for all nonparametric mixtures of the exponential family (Lindsay (1995))).

The maximum likelihood can also be found using a classical Newton-Raphson technique. The property (5.15) can be used to simplify the procedure by reducing the number of parameters to be estimated.
About the number of mass points r, Simar (1976) shows that the maximum likelihood estimator will be unique in the following conditions

$$
\begin{aligned}
r &\leq \min\left(q, \left\lfloor \frac{N+2}{2} \right\rfloor\right) \quad \text{if } \lambda_1 = 0 \\
r &\leq \min\left(q, \left\lfloor \frac{N+1}{2} \right\rfloor\right) \quad \text{if } \lambda_1 > 0
\end{aligned}
$$

where q is the number of classes for which the observation is different from 0
N is the maximum number of claims per risk

For our reference portfolio, we find the following maximum likelihood estimation :

	Obs	NPMP		
0	103704	103704.53	l	-54609.46
1	14075	14075.07	χ^2	0.125
2	1766	1765.21	df	0
3	255	255.73	$p-value$	???
4	45	43.63	p_1	0.56189
5	6	7.50	p_2	0.41463
6	2	1.16	p_3	0.02348
7		0.16	λ_1	0.05461
8		0.02	λ_2	0.24599
			λ_3	0.95618

Table 5.9: Nonparametric Mixed Poisson (NPMP) fit

Obviously, the log-likelihood is higher than in any parametric case.
The fit is excellent because we have 5 free parameters for 7 classes. We are in a case where it is not possible to compute the p-value of the χ^2 test due to the high number of parameters. The model seems to be overfitting. However when tested against the simpler model with two components (see section 5.7) the rejection of the overfitted model is not strong.
Note that the distribution is not infinitely divisible in this case.
The nonparametric case gives a physical interpretation of the heterogeneity of the portfolio : 56% of the risks follow a Poisson Distribution with parameter $\lambda = 0.05461$, 41% of the risks follow a Poisson Distribution with parameter $\lambda = 0.24599$ and 2% of the risks follow a Poisson Distribution with parameter $\lambda = 0.95618$.
Note that the maximum likelihood method described in this section gives more information than the simple good guy bad guy model of Lemaire (1985). Indeed the procedure gives the number of mass points needed to have the highest likelihood whereas the good guy bad guy model imposes two mass points.

5.3 The Zero Inflated Poisson Distribution

As mentioned by Böhning (1998), practical data sets often show a large proportions of zeroes. In fact one should say a larger proportion of zeroes than the one predicted by the Poisson distribution. See also the introductory comments of section 5.1. This is the reason why the Zero Inflated Poisson model has been introduced.
The Zero Inflated Poisson model consists in inflating the number of zeroes in the Poisson model :

$$\Pi(0,t) = p + (1-p)e^{-\lambda t}$$

The rest of the distribution writes

$$\Pi(n,t) = (1-p)e^{-\lambda t}\frac{(\lambda t)^n}{n!} \quad , \quad n \geq 1$$

This model is in fact a particular case of the Nonparametric Mixed Poisson Distribution in the sense that it is a mixture of two Poisson Distributions with the parameter of the first Poisson Distribution being equal to 0.
We can write

$$\Pi(n,t) = \sum_{j=1}^{2} p_j e^{-\lambda_j t}\frac{(\lambda_j t)^n}{n!} \quad , \quad n \geq 0 \tag{5.16}$$

with $p_1 = p \geq 0$, $p_2 = 1-p$ and $\lambda_1 = 0$.
We have

$$\Pi(n,t) = pI_{[n=0]} + (1-p)e^{-\lambda t}\frac{(\lambda t)^n}{n!} \quad , \quad n \geq 0 \tag{5.17}$$

As this model is a mixture model, we have an easy maximum likelihood equation (Lindsay (1995)) :

$$(1-\hat{p})\hat{\lambda} = \bar{N} \tag{5.18}$$

Writing the log-likelihood of the model gives

$$l = n_0 \ln(p + (1-p)e^{-\lambda}) + (n - n_0)(\ln(1-p) - \lambda) + n\bar{N}\ln(\lambda)$$

Derivating with respect to p gives a score equation :

$$\frac{n_0}{n} = \hat{p} + (1-\hat{p})e^{-\hat{\lambda}} \tag{5.19}$$

Solving equations (5.18) and (5.19) gives the maximum likelihood estimates.
The fit on our reference portfolio gives :

	Obs	ZIP		
0	103704	103704	l	-54668.41
1	14075	13928.35	χ^2	149.05
2	1766	2012.05	df	2
3	255	193.76	$p-value$	0.000
4	45	14.00	p	0.46302
5	6	0.81	λ	0.28891
6	2	0.04	$as\sigma_p$	0.0099
7		0.00	$as\sigma_\lambda$	0.0057
8		0.00	$as\rho(p,\lambda)$	0.9189

Table 5.10: Zero Inflated Poisson (ZIP) fit

We note that the fit is not adequate especially in the tail. The model concentrates too much on the Zero Inflation. The tail is so badly fitted that it has been necessary to work with 5 classes in this example in order to assure all theoretical frequencies larger than 1 for the χ^2 statistic.
We note that the estimates are highly correlated. This is due to our parametrization. Another parametrization, the Zero-Modified Poisson Distribution (see p21 and see Klugmann, Panjer and Willmot (1998) for details), does not present this drawback. Within the Zero-Modified Poisson Distribution, we have

$$\begin{aligned}
\Pi(0,t) &= p_0^M \\
\Pi(n,t) &= \frac{1-p_0^M}{1-e^{-\lambda t}} e^{-\lambda t} \frac{(\lambda t)^n}{n!} \quad , \quad n \geq 1
\end{aligned}$$

Within this reparametrization ($p_0^M = p + (1-p)e^{-\lambda t}$), it is easy to show that $\hat{p}_0^M$ and $\hat{\lambda}$ are uncorrelated. We find

$\hat{p}_0^M$	0.8652
$\hat{\lambda}$	0.2889
$as\sigma_{p_0^M}$	0.0009
$as\sigma_\lambda$	0.0057
$as\rho(p_0^M,\lambda)$	0

Table 5.11: Maximum likelihood estimates within the ZM Poisson Distribution

5.4 The Generalized Poisson Distribution

The Generalized Poisson Distribution was introduced by Consul and Jain (1973). It was also fully described in Consul's book (1989).
A discrete random variable N is Generalized Poisson distributed if its probability function writes :

- If $\theta > 0$ and $0 \leq \lambda < 1$:

$$p(n) = \theta(\theta + n\lambda)^{n-1} \frac{e^{-\theta - n\lambda}}{n!} \quad , \quad n \geq 0$$

- If $\theta > 0$ and $\lambda < 0$:

$$\begin{aligned} p(n) &= \theta(\theta + n\lambda)^{n-1} \frac{e^{-\theta - n\lambda}}{n!} \quad , \quad n = 0, \dots, m \\ &= 0 \quad , \quad n > m \end{aligned}$$

where $\max(-1, -\frac{\theta}{m}) \leq \lambda < 1$ and m is usually the largest possible integer such that $\theta + m\lambda > 0$.

We immediately note that the definition is not very intuitive and that there are a lot of unnatural restrictions on the parameters.
Moreover when $\lambda < 0$ the Generalized Poisson Distribution is truncated. A normalization should then be used in order that the probabilities sum to 1. Consul (1989) argues that the truncation error is always less than 0.07%.

The mean and the variance are given, after some tedious calculations by

$$\begin{aligned} \mu &= \frac{\theta}{1 - \lambda} \\ \sigma^2 &= \frac{\theta}{(1 - \lambda)^3} \end{aligned}$$

Note that these formulae, as well as other more complicated theoretical results, are only valid for the case $0 \leq \lambda < 1$. In some papers it is tacitely assumed that these formulae are valid also for negative values of λ (see e.g. Famoye and Consul (1995)).

The Generalized Poisson Distribution has the property that it can be a model for underdispersed ($\lambda < 0$) or overdispersed ($\lambda > 0$) data sets. The case $\lambda = 0$ reduces to the Poisson case.
As the case $\lambda < 0$ causes problems in the sense that the probabilities do not sum to 1, we will only consider the case $\lambda \geq 0$.

Consul (1989) shows that the maximum likelihood estimate for the mean of the Generalized Poisson Distribution is given by the empirical mean :

$$\frac{\hat{\theta}}{1 - \hat{\lambda}} = \bar{N}$$

This means that only one parameter has to be estimated by numerical technique when fitting a data set by maximum likelihood.

¿From a physical point of view, Consul (1989) shows that the Generalized Poisson Distribution represents the cumulative effect of two stochastic processes. In motor insurance, θ represents the cumulative effect of the conditions of the roads, the amount of traffic, the sense of road discipline of the drivers, ... The parameter λ represents the average effect of the number of passengers in the cars.

The Generalized Poisson Distribution is neither a Mixed Poisson Distribution nor a Compound Poisson Distribution. It does not present the interesting physical properties of the Hofmann Distribution.

The problem of evaluating the probability function of a compound distribution when the counting distribution is Generalized Poisson distributed was solved by Ambagaspitya and Balakrishnan (1994) where recursive algorithms are derived.
We now fit our reference portfolio :

	Obs	GPD		
0	103704	103722.22	l	-54612.96
1	14075	14003.68	χ^2	7.32
2	1766	1838.19	df	3
3	255	248.48	$p-value$	0.062
4	45	34.62	θ	-0.1445
5	6	4.95	λ	-0.0683
6	2	0.72	$as\sigma(\theta)$	0.0011
7		0.11	$as\sigma(\lambda)$	0.0027
8		0.02	$as\rho(\theta,\lambda)$	-0.1804

Table 5.12: Generalized Poisson Distribution (GPD) fit

This fit would not be rejected at the significance level 5%.

5.5 The Poisson Gontcharov Distribution

The Poisson Gontcharov Distribution is a three-parameter distribution that was proposed by Denuit (1997). Its probability function is

$$p(n) = G(n)e^{-\theta_1-\theta_2 n-\theta_3 n^2} \quad , \quad n \geq 0$$

with

$$\begin{aligned}
\theta_1 &< 0 \\
\theta_2 &\leq 0 \\
\theta_3 &\leq 0 \\
G(0) &= 1 \\
G(n) &= \frac{1}{n!} - \sum_{i=0}^{n-1} \frac{(-\theta_1-\theta_2 i-\theta_3 i^2)^{n-i}}{(n-i)!} G(i) \quad , \quad n \geq 1
\end{aligned}$$

To the knowledge of the author, there is no simple form for the first moments of the Poisson Gontcharov Distribution.

A physical interpretation of the Poisson Gontcharov Distribution was given by Denuit (1997) :

$$N = N_{Poisson} + N_{extra}^{(1)} + N_{extra}^{(2)} + \cdots$$

where $N_{Poisson}$ follows a Poisson distribution with parameter $-\theta_1$ and $N_{extra}^{(j)}$ given $N_{Poisson}$, $N_{extra}^{(1,2,\cdots,j-1)}$ is a Poisson Distribution with a complicated parameter to write out. This gives a physical interpretation of the Poisson Gontcharov model : some accident processes add and each type of accident gives a number of accidents conditionnally distributed along a Poisson Distribution.

A particular case of the Poisson Gontcharov Distribution is the Generalized Poisson Distribution introduced in the previous section. Indeed if one takes $\theta_3 = 0$, we find the Generalized Poisson Distribution. Note that the Poisson Gontcharov Distribution authorizes only negative values for the parameters. Thus we do not have the same problems as with the underdispersed Generalized Poisson Distribution. The overdispersed Generalized Poisson Distribution is a particular case of the Poisson Gontcharov Distribution with

$$\begin{aligned}\theta_1 &= -\theta \\ \theta_2 &= -\lambda \\ \theta_3 &= 0\end{aligned}$$

The numerical determination of the estimates of the parameters by maximum likelihood does not pose any problems. Let us note that each data set the author had fitted presented the following property

$$\bar{N} = \hat{\mathbb{E}N}$$

Whether this property is always true or not remains to be proved. Nevertheless if an analytical form of $\mathbb{E}N$ is not known, the property will not reduce the dimensionality of the optimization problem.

We note that the Poisson Gontcharov is neither Mixed Poisson nor Compound Poisson distributed.

The problem of compounding with a Poisson Gontcharov has not yet been treated in literature. Up to now there does not exist a recursive formula giving the probability function of the compound distribution when the counting distribution is Poisson Gontcharov distributed. We now fit our reference portfolio :

	Obs	Po Gont		
0	103704	103705.63	l	-54609.89
1	14075	14067.00	χ^2	0.88
2	1766	1777.93	df	2
3	255	251.47	$p-value$	0.643
4	45	41.06	θ_1	-0.1447
5	6	7.72	θ_2	-0.0565
6	2	1.65	θ_3	-0.0082
7		0.40		
8		0.11		

Table 5.13: Poisson Gontcharov fit

The fit is excellent.

5.6 The Weighted Poisson Distribution

Del Castillo and Perez-Casany (1998) introduced the Weighted Poisson Distribution. The model is given by

$$p(n) = \frac{(n+a)^r \lambda^n e^{-\lambda}}{n! C(\lambda, r, a)} \quad , \quad n \geq 0$$

where

$$C(\lambda, r, a) = \sum_{n=0}^{\infty} \frac{\lambda^n (n+a)^r}{n!}$$

This distribution, that has no closed form, is just a weighted Poisson distribution with weight $(k+a)^r$. The parameter r discriminates between overdispersion ($r < 0$) and underdispersion ($r > 0$). The case $r = 0$ reduces to the Poisson case.
The model of Del Castillo and Perez-Casany (1998) is purely technical and does not present any physical interpretation.

The numerical determination of the estimates of the parameters by maximum likelihood does not pose any problems whereas Del Castillo and Perez-Casany (1998) processed maximum likelihood estimation with likelihood profiles in function of a. This is clearly not necessary. It does not seem that a moment equation is a likelihood equation.

Based on two examples, Del Castillo and Perez-Casany (1998) claim that their weighted Poisson Distribution give better fits than the traditional counting distributions. Based on the fits obtained for the data sets given in Denuit (1997), among the counting distributions with 3 parameters, this fact is obviously not true.

We now fit our reference portfolio :

	Obs	WPD		
0	103704	103709.89	l	-54609.95
1	14075	14053.67	χ^2	1.11
2	1766	1788.85	df	2
3	255	251.45	$p-value$	0.574
4	45	40.07	λ	40.211
5	6	7.21	r	-45.757
6	2	1.45	a	7.54797
7		0.32		
8		0.08		

Table 5.14: Weighted Poisson Distribution (WPD) fit

The fit is also excellent.

5.7 Nested and non-nested models

In this section we will try to compare the different models based on the quality of their fits. We have different statistics at our disposal :

- the log-likelihood : l
- the χ^2 statistic
- the p-value of the χ^2 test
- the Akaike Information Criterion (see Akaike (1973)) : $AIC = l - dim$
- the Bayesian Information Criterion (see Schwartz(1978)) : $BIC = l - \frac{1}{2}\ln(n)dim$

where dim denotes the number of parameters that have been estimated and n is the size of our observed sample. The AIC and BIC penalize the log-likelihood against overfitting. The BIC takes into account the size of the sample whereas it is not the case for the AIC. The latter tends to accept overparametrized models when the size of the sample is high. The AIC and BIC, as well as the l are just statistics. They can be used for the sake of comparison.
Within nested models we have at our disposal a statistical test in order to analyze the overfitting in order to respect the principle of parcimony : the Likelihood Ratio Test (LRT).
Let a general model and its parameters estimated by maximum likelihood. After having imposed k restrictions on the parameters, it is possible to process restricted maximum likelihood estimation on the remaining parameters. We test
H0 : restricted model (with log-lokelihood l_0)
H1 : general model (with log-likelihood l)
The test statistic is $LRT = 2\ln(\frac{l}{l_0})$. Under classical regularity conditions, this statistic is asymptotically distributed as a χ^2 with k degrees of freedom.
Unfortunately, in some cases of interest, tested parameters lie on the boundary which makes the regularity conditions fail. This subject was treated in Self and Liang (1987) from where we can deduce the following alternative tests :

- if the only restricted parameter is on the boundary, the asymptotic distribution of the LRT is distributed as $\frac{1}{2}\chi^2$ with 1 df.
- if two parameters are restricted with one on the boundary, the asymptotic distribution of the LRT is distributed as $\frac{1}{2}\chi^2$ with 3 df.

Let us study some nested models.

H0 : $N \sim Ho(p, c, a)$
H1 : $N \sim Ho(p, c, a) + Po(\delta)$
The LRT is 0.00000256 and the null hypothesis cannot be rejected. Note that the p-value of the test should be evaluated with the $\frac{1}{2}\chi^2$ distribution with 1 df. Working with the four-parameter model induces overparametrization. This partly explains the bad behaviour of the asymptotic standard deviations and correlations between the estimates of the parameters of the model under H1. Nevertheless the general model is physically justified. The parameters are identifiable. It is clear that with such a low frequency, the risks causing claims by chance imply that the Hofmann part of the process is ejected, which causes the bad behaviour of the efficiency of the estimates of the model. We should recall that the most general model is Ho+Po. In the case when it is physically justified, the most general model should not be rejected. Moreover it is clear that with typical motor portfolio, the number of classes is too low to distinguish statistically which is the better model.

H0 : $N \sim NPMPD(p, \lambda_1, \lambda_2)$
H1 : $N \sim NPMPD(p_1, \lambda_1, p_2, \lambda_2, \lambda_3)$
The LRT is 3.24. The p-value is $\mathbb{P}(\frac{1}{2}\chi^2(3) > 3.24) = 0.09$ which is against rejection of $H0$ at the 5% level.

H0 : $N \sim ZIP$
H1 : $N \sim NPMPD(p, \lambda_1, \lambda_2)$
The LRT is 115 which implies rejection of H0. This is not surprizing as we noted a bad fit of the ZIP Distribution.

H0 : $N \sim GPP$
HI : $N \sim$ *PoGont*
The LRT is 6.14. The p-value is $\mathbb{P}(\frac{1}{2}\chi^2(1) > 6.14) = 0.00046$ which indicates a strong rejection of H0.

We now move to the comparison between all models including the non-nested models. The following table gives the l, AIC,BIC, χ^2 and p-value of the χ^2 test.

	l	AIC	BIC	χ^2	p-value
Poisson	-55108.45	-55109.45	-55114.30	2550	0
NB	-54615.31	-54617.31	-54627.01	12.36	0.006
PIG	-54609.75	-54611.75	-54621.45	0.77	0.85
Hofmann	-54609.59	-54612.59	-54627.13	0.43	0.80
Ho + Po	-54609.53	-54613.53	-54632.92	1.16	0.28
NPMPD(3)	-54609.46	-54614.46	-54638.09	0.12	??
NPMPD(2)	-54611.08	-54614.08	-54628.62	2.79	0.24
ZIP	-54668.40	-54670.40	-54680.01	149	0
GPD	-54612.95	-54614.95	-54624.65	7.31	0.06
Poisson Gontcharov	-54609.88	-54612.88	-54627.42	0.88	0.64
WPD	-54609.95	-54612.95	-54627.49	1.10	0.57

Table 5.15: Comparison of the models

We will not base our conclusions on the χ^2 statistic nor on the p-value associated with the χ^2 test because of the pitfalls of that statistic.
Based on the AIC the best fit is given by the PIG Distribution. The same conclusion is drawn with the BIC. A likelihood ratio test would also be in favour of the PIG Distribution (against the Hofmann Distribution). This is due to the fact that the estimated parameter a in the three parameters model is 0.44, close to the chosen a of the PIG Distribution ($a = 0.5$). Clearly one will always find "a best fit based on penalized log-likelihood" by choosing different values for the parameter a (e.g. $a = 0.1$, $a = 0.2$, ...) and possibly giving a name to these distributions. The NB and PIG Distributions are recognized distributions in literature. Should the PIG Distribution be the winner on account of that fact ? For motor portfolio's I am tempted to answer yes. Indeed all data sets given in Denuit (1997) show that the Hofmann Distribution is not able to beat the PIG Distribution based on a penalizing criterion. This is coherent with

the study made by Willmot (1987) who compared the PIG and NB Distributions. Note that in other areas than motor portfolios, other members of the Hofmann Distributions might fit well. This fact was advocated by Lemaire (1991) who is in favour of the NB Distribution. As the general model with three parameters is physically justified, I would not reject it. Nevertheless, for motor portfolios the PIG Distribution is an excellent candidate.

Note that based on the BIC, the NB Distribution should be chosen in comparison with the Hofmann Distribution. This is due to the large sample we use and is clearly not acceptable in view of the poor fit of the NB Distribution.

The Poisson Gontcharov, the Hofmann and the WPD fits are highly comparable, which is not illogical due to the type of data we have at hand. We have very few classes and three parameters for the fitting. Clearly the preference should be given to the model with physical intuition.

The Nonparametric Mixed Poisson Distribution with three components is clearly overparametrized in comparison with parametric models with three parameters.

As already mentionned above, the Ho + Po Distribution is overfitting in comparison with the Ho Distribution. Once again this is also partly due to the type of data we use. If the physical intuition tells us that a pure random part of the frequency is present in the model, it should not be rejected based on the bad statistics.

5.8 Fits with varying size

In this section (taken from Walhin and Paris (2000b)), we study the same numerical example as in Bissel (1972) with a more general distribution in order to get a better fit. Moreover we will be able to distinguish between the pure random part of the frequency and the non pure random part of it.
The data are from Bissel (1972) and give the number of faults in rolls of textile fabric with varying sizes for the rolls :

Roll no.	Roll length	No. of faults		Roll no.	Roll length	No. of faults
1	551	6		17	543	8
2	651	4		18	842	9
3	832	17		19	905	23
4	375	9		20	542	9
5	715	14		21	522	6
6	868	8		22	122	1
7	271	5		23	657	9
8	630	7		24	170	4
9	491	7		25	738	9
10	372	7		26	371	14
11	345	6		27	735	17
12	441	8		28	749	10
13	895	28		29	495	7
14	458	4		30	716	3
15	642	10		31	952	9
16	492	4		32	417	2

Table 5.16: Data set

We will assume that $N_1(t)$ is the sum of two independent components

$$N_1(t) = N^*(t) + N(t)$$

The former, $N^*(t)$, which is a pure Poisson Process with mean δt represents the purely random defaults. The latter, $N(t)$ which is a Mixed and Compound Poisson process represents the defaults related to the quality of the production. We assume that $N(t)$ is Hofmann distributed. We apply maximum likelihood in order to find estimates for p, c, a and δ. Standard numerical maximization techniques allow these estimates. However it is necessary to be careful because the log-likelihood is not concave everywhere.

Type	δ	p	c	a	ac	l	AIC	BIC
Poisson	0	0.0151		0		−93.91	−94.91	−95.64
NB	0	0.0153	0.0024	1		−88.01	−90.01	−91.47
PIG	0	0.0150	0.0037	0.5		−87.54	−89.54	−91.00
Hofmann	0	0.0150	0.0086	0.2402		−87.53	−90.53	−92.72
Neyman Type A	0	0.0151			0.0015	−88.02	−90.02	−91.48
NB + Poisson	0.0097	0.0055	0.0077	1		−87.35	−90.35	−92.54
PIG + Poisson	0.0066	0.0083	0.0076	0.5		−87.35	−90.35	−92.54
NTA + Poisson	0.0117	0.0032			0.0099	−86.93	−89.93	−92.12

Table 5.17: Characteristics of the fit

Our model has the following good properties (see section 5.1) :

$$\mathbb{E}[N_1(t+1) - N_1(t) | N_1(t) = 0] = \frac{d}{dt}\theta(t) + \delta \tag{5.20}$$

$$\lim_{t\to\infty} \mathbb{P}(N(t) = 0) > 0 \quad \text{if } a > 1 \tag{5.21}$$

Clearly we expect that a roll without default implies the smallest frequency of defaults in a longer roll. So we would like to have the smallest value for (5.20). It is interesting to note, that with our model constrained to fixed ac, $\frac{d}{dt}\theta(t)$ is minimized for the Neyman Type A choice (this is easily seen by a Taylor expansion).
Furthermore we expect that the quality of the production implies that it is possible to have infinite rolls without defaults. So (5.21) implies that $N(t)$ should have the property $a > 1$. Clearly we can not have infinite rolls without defaults because of the purely random effects. This is the reason why a Poisson component is introduced in our model. Finally, it is clear that from an industrial point of view, we expect that most of the defaults should occur randomly. The best fit (based on the log-likelihood), given by the Neyman Type A + Poisson shows very much this desirable property : 78.5% of the frequency comes from the Poisson component. However the AIC and BIC remain in favour of the Negative Binomial Distribution due to its two parameters against three in the Neyman Type A + Poisson model.

Therefore a choice has to be made between Negative Binomial Distribution and Neymann Type A + Poisson. Based on the physical interpretation of the model and the fact that NTA+Poisson is not overfitting too much, I would choose the latter.

Chapter 6

Bivariate counting distributions

Bivariate counting distributions are extensively described in Kocherlakota and Kocherlakota (1992) and become more and more popular in actuarial sciences (see Hesselager (1996), Vernic (1997), Partrat (1994), Walhin and Paris (2000h)). The aim of this chapter is to unify and extend some previous results by using the interesting characteristics of the Hofmann Distribution.
We will use the following notation :

$$p(n,m) = \mathbb{P}[N = n, M = m]$$

A special attention is drawn to the physical interpretation of the proposed models. Moreover I will closely consider the possible recursions for the probability function of the bivariate model as well as the possible recursions for the compound bivariate model.
Two techniques will be used :

- Mixing the Bivariate Poisson model :
  The Bivariate Poisson model has been known for a long time. Its probability generating function writes

  $$\psi(u,v) = e^{\lambda_1(u-1)+\lambda_2(v-1)+\lambda_0(uv-1)}$$

  It is known from Teicher (1954) that the following recursion is useful in order to evaluate the probability function

  $$\begin{aligned} p(0,0) &= e^{-\lambda_0-\lambda_1-\lambda_2} \\ np(n,m) &= \lambda_1 p(n-1,m) + \lambda_0 p(n-1,m-1) \quad , \quad n \geq 1 \\ mp(n,m) &= \lambda_2 p(n,m-1) + \lambda_0 p(n-1,m-1) \quad , \quad m \geq 1 \end{aligned}$$

  The Bivariate Poisson Distribution can be derived along different methods. One of these methods is the trivariate reduction method that will be described hereunder.
  A simple way of giving other bivariate distribution is to mix the Bivariate Poisson Distribution. Kocherlakota (1988) proposes the following model for the probability generating function :

  $$\psi(u,v) = \int_0^\infty e^{\lambda(\lambda_1(u-1)+\lambda_2(v-1)+\lambda_0(uv-1))} dU(\lambda)$$

  We will study this model as well as its particular case ($\lambda_0 = 0$) in the sections 6.1.1 and 6.1.2 with the mixing distribution leading to the Hofmann Distribution.

- Using the Trivariate Reduction Method (TRM) introduced by Mardia (1970) :
  Let us assume that N_i are independent random variables for $i = 0, 1, 2$. Then a bivariate model can be constructed by

$$\begin{aligned} N &= N_0 + N_1 \\ M &= N_0 + N_2 \end{aligned}$$

  This model will be studied by assuming that the random variables N_i are Hofmann distributed. See section 6.2.

Section 6.3 is concerned by the two kinds of models but with the philosophy of Zero Inflation. A data set will be fitted by maximum likelihood for each model we will derive in the sequel. The data set comes from Arbous and Kerrich (1951) and is concerned by accidents sustained by members of a sample of 122 shunters in two consecutive 2-years periods. It was used in Famoye and Consul (1995). Note that Famoye and Consul (1995) fitted a Bivariate Generalized Poisson Distribution by using the Trivariate Reduction Method. The same model was also presented by Vernic (1997). Unfortunately in both papers, the parameters are estimated by the method of moments. Our model, based on the Hofmann Distribution, is more interesting because it allows two procedures of construction of a bivariate distribution.

N/M	0	1	2	3	4	5	6	7
0	21	18	8	2	1	0	0	0
1	13	14	10	1	4	1	0	0
2	4	5	4	2	1	0	1	0
3	2	1	3	2	0	1	0	0
4	0	0	1	1	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	1	0	0	0	0	0	0

Table 6.1: Reference observed bivariate distribution

Note that this kind of bivariate counting distributions are also useful in social science, economics, physics, ... (see e.g. Westman (1971), Hamdan and Tsokos (1971), Rao et al. (1973), Galliher et al. (1959), Patil and Bildikar (1967) or Upton and Lampitt (1981)).

6.1 The Mixed Bivariate Hofmann Distribution

We first study the case of mixing a Bivariate Independent Poisson model because it has interesting physical interpretations and it leads to a generalized type of bonus-malus systems as we will see in sections 7.2 and 7.3. Then we will study the mixing of the Bivariate Poisson Distribution that has less good properties apart from the fact that it is more general.

6.1.1 Mixing the Bivariate Independent Poisson model

This section is mainly based on Walhin and Paris (2000d).
We are going to study the random vector (N, M) of counting variables. We will obtain

the distribution of (N, M) by mixing the conditional distribution of (N, M) with a random variable Λ with cumulative density function $U(\lambda)$:

$$\mathbb{P}(N = n, M = m) = \int_0^\infty \mathbb{P}(N = n, M = m|\Lambda = \lambda)dU(\lambda) \tag{6.1}$$

Furthermore we assume that

- Conditionally on Λ the random variables N and M are independent
- The conditional distributions of N and M given that $\Lambda = \lambda$ are univariate Poisson with parameter respectively λ and $\beta\lambda$

The probability generating function of (N, M) writes

$$\psi_{N,M}(u, v) = \int_0^\infty e^{\lambda(u-1)+\beta\lambda(v-1)}dU(\lambda)$$

Kemp (1981) introduced the notion of Homogeneous Bivariate Distribution :

Definition 6.1 *A bivariate probability generating function $\psi(u, v)$ is said to be of the homogeneous type if*

$$\psi(u, v) = H(\sigma_1 u + \sigma_2 v)$$

with

$$H(\sigma_1 + \sigma_2) = 1$$

If we choose H such that

$$H(x) = \int_0^\infty e^{-\lambda(1+\beta)}e^{\lambda x}dU(\lambda)$$

we immediately get that (N, M) is a Bivariate Homogeneous Distribution with $\sigma_1 = 1$ and $\sigma_2 = \beta$.
Kocherlakota and Kocherlakota (1992) gave the following characterization theorem.

Theorem 6.1 *The probability generating function $\psi(u, v)$ is of the homogeneous type if and only if the conditional distribution of N given $N+M = z$ is Binomial distributed : $Bi(z, \frac{\sigma_1}{\sigma_1+\sigma_2})$.*

In our case we have

$$N|N + M = z \sim Bi(z, \frac{1}{1 + \beta})$$

a result that was also obtained by Partrat (1994) and Hesselager (1996).
¿From Hesselager (1996) it is possible to extend the result of Kocherlakota and Kocherlakota (1992) by

$$\begin{aligned}
\rho_1 &= \frac{\sigma_1}{\sigma_1 + \sigma_2} \\
\rho_2 &= \frac{\sigma_2}{\sigma_1 + \sigma_2} \\
\psi_{N,M}(u, v) &= \psi_{N+M}(\rho_1 u + \rho_2 v)
\end{aligned}$$

Let us use the Hofmann Distribution in our bivariate case. Let us assume that $U(\lambda)$ is the mixing distribution leading to the Hofmann Distribution ($\Lambda \sim Homix(p, c, a)$). The

probability function of the mixing distribution can be obtained but it is not necessary for our developments.
¿From (6.1) we immediately get

$$\mathbb{P}(N=n, M=m) = \frac{(n+m)!}{n!m!} \frac{\beta^m}{(1+\beta)^{n+m}} \Pi(n+m, 1+\beta) \tag{6.2}$$

where it is easy to see that

$$\Pi_{p,c,a}(n+m, 1+\beta) = \Pi_{(1+\beta)p,(1+\beta)c,a}(n+m, 1)$$

In fact our model introduces dependency such that :

$$\begin{aligned} N &\sim Ho(p, c, a) \\ M &\sim Ho(p\beta, c\beta, a) \\ N+M &\sim Ho(p(1+\beta), c(1+\beta), a) \end{aligned}$$

This clearly generalizes the reasoning of Partrat (1994) where only Λ Gamma or Inverse Gaussian distributed are considered.

Let us assume that we have observed a sample $(n_i, m_i) \quad 1 \le i \le q$ of (N, M).
The log-likelihood writes

$$\begin{aligned} l(\beta, p, c, a) &= \ln \prod_{i=1}^{q} \mathbb{P}(N=n_i, M=m_i) \\ &= C + \ln\beta \sum_{i=1}^{q} m_i - \ln(1+\beta) \sum_{i=1}^{q} (n_i+m_i) + l_{n+m}(\beta, p, c, a) \end{aligned}$$

where $l_{n+m}(\beta, p, c, a)$ is the log-likelihood for the univariate Hofmann Distribution $Ho((1+\beta)p, (1+\beta)c, a)$ with the sample $(n_i+m_i), 1 \le i \le q$ and C is a constant not depending on the unknown parameters.
As shown in Besson and Partrat (1992), $\frac{\partial}{\partial\beta} l_{n+m}(\beta, p, c, a) = 0$ at the maximum likelihood estimate. So we immediately get

$$\hat{\beta} = \frac{\bar{m}}{\bar{n}}$$

where $\bar{n}$ (resp. $\bar{m}$) is the experimental mean of N (resp. M).
By standard results on the univariate Hofmann Distribution (see section 5.1.10), we know that the maximum likelihood estimate of the mean is the observed frequency. Therefore maximizing $l_{n+m}(\beta, p, c, a)$ implies that

$$\hat{p}(1+\hat{\beta}) = \bar{n} + \bar{m}$$

So the estimates $\hat{p}$ and $\hat{\beta}$ are derived analytically. The other two estimates $\hat{c}$ and $\hat{a}$ are to be found by standard numerical maximization techniques. Note that one has to be careful because the likelihood may be very flat and local extrema are not excluded.

Let us fit our reference data set :

N/M		0	1	2	3	4	5	6	7		
Obs Fitted	0	21 21.72	18 16.63	8 8.24	2 3.34	1 1.20	0 0.40	0 0.13	0 0.04	l χ^2	-341.775 2.199
Obs Fitted	1	13 12.77	14 12.66	10 7.70	1 3.70	4 1.54	1 0.58	0 0.20	0 0.07	df $p-value$	8 0.974
Obs Fitted	2	4 4.86	5 5.91	4 4.26	2 2.36	1 1.11	0 0.47	1 0.18	0 0.07	p β	1.2705 0.7677
Obs Fitted	3	2 1.51	1 2.18	3 1.81	2 1.14	0 0.60	1 0.28	0 0.12	0 0.05	c a	0.2851 1.0238
Obs Fitted	4	0 0.42	0 0.70	1 0.66	1 0.46	0 0.27	0 0.14	0 0.06	0 0.00	AIC BIC	-345.775 -351.383
Obs Fitted	5	0 0.11	0 0.20	0 0.21	0 0.16	0 0.11	0 0.06	0 0.00	0 0.00		
Obs Fitted	6	0 0.03	0 0.05	0 0.06	0 0.05	0 0.04	0 0.00	0 0.00	0 0.00		
Obs Fitted	7	0 0.01	1 0.01	0 0.02	0 0.02	0 0.00	0 0.00	0 0.00	0 0.00		

Table 6.2: Mixed Independent Bivariate Hofmann fit

As we see the model is quite significant. In order to compute the χ^2 statistic, some cells have been grouped in order to have the majority (about 80% like in rule B in Lemaire (1985)) of the theoretical frequencies larger than 5 and each theoretical frequency larger than 1.
The grouped cells are : $(0,3+),(1,3+),(2,3+),(3+,0+)$ which makes a total of 13 classes. We will keep the same grouping rule for the fits with the other models in order that the χ^2 statistics remain comparable.

In the conclusion of his paper, Partrat (1994) addressed the problem of finding recursions like the Panjer's recursion (Panjer (1981)) in order to give the distribution of the aggregate claims.
We will first derive a trivial recursion that unfortunately will prove to be unstable.
Let X_i be the random variable representing the ith claim amount of type N and Y_i the random variable representing the ith claim amount of type M. We will assume, as usual, that the X_i and Y_i are mutually independent random variables. They are also arithmetic. The X_i are identically distributed. The Y_i are also identically distributed. We further assume that the X_i and Y_i are independent of N and M. We are interested in the distribution of

$$(S,T)=(\sum_{i=1}^{N}X_i,\sum_{i=1}^{M}Y_i)$$

Theorem 6.2 *A simple recursion for $p(n,m)$ is given by*

$$p(0,0) \quad = \quad \Pi_{((1+\beta)p,(1+\beta)c,a)}(0,1) \tag{6.3}$$

$$np(n,m) = \frac{1}{\beta}(m+1)p(n-1,m+1) \quad \text{if} \quad n>0 \tag{6.4}$$
$$mp(n,m) = \beta(n+1)p(n+1,m-1) \quad \text{if} \quad m>0 \tag{6.5}$$

Proof

(6.3) follows immediately from (6.1).
(6.4) is given by

$$\begin{aligned}
\Pi_{p,c,a}(n+m,1+v) &= p(n,m)\frac{n!m!}{(n+m)!}\frac{(1+v)^{n+m}}{v^m} \\
&= p(n-1,m+1)\frac{(n-1)!(m+1)!}{(n+m)!}\frac{(1+v)^{n+m}}{v^{m+1}} \\
&\Leftrightarrow \\
np(n,m) &= \frac{1}{v}(m+1)p(n-1,m+1)
\end{aligned}$$

and (6.5) holds similarly. ■

Note that in order to run this recursion, it is necessary to have a full initialization that is $p(n,0)\,, n>0$ or $p(0,m)\,, m>0$:

$$\begin{aligned}
p(n,0) &= \frac{1}{(1+\beta)^n}\Pi_{p,c,a}(n,1+\beta) \\
p(0,m) &= \frac{\beta^m}{(1+\beta)^m}\Pi_{p,c,a}(m,1+\beta)
\end{aligned}$$

Theorem 6.3 *A simple recursion for $f_{S,T}(s,t)\,,\ s>0\,,\ t>0$ is given by*

$$\sum_{j=1}^{t} jf_Y(j)sf_{S,T}(s,t-j) = \frac{1}{\beta}\sum_{i=1}^{s} if_X(i)tf_{S,T}(s-i,t) \tag{6.6}$$

Proof

Multiplying (6.4) by $\psi_X^{n-1}(u)\frac{d}{du}\psi_X(u)\psi_Y^m(v)\frac{d}{dv}\psi_Y(v)$ on both sides, rearranging and inverting immediately gives (6.6). ■

The recursion (6.6) will clearly not be stable due to numbers close to zero that have to be substracted. Moreover the initialization ($f_{S,T}(s,0)$ and $f_{S,T}(0,t)$) is not given because the structure of $\Pi_{p,c,a}(n+m,1+\beta)$ is not used.
So we should look for another type of recursions.
In the case of the Mixed Bivariate Negative Binomial Distribution the answer was given by Hesselager (1996). In his paper, Hesselager (1996) gave a stable algorithm for the evaluation of the joint probability function of (S,T) for the particular case of the Mixed Bivariate Negative Binomial Distribution, i.e. when Λ is Gamma distributed.

We now use the same methodology as in Hesselager (1996) in order to derive stable algorithms for the distribution of (S,T). As we know that U is infinitely divisible, it follows from Maceda (1948) that the distribution of (N,M) is also infinitely divisible. Then, according to Sundt (2000) we deduce that (N,M) can be interpreted as a Bivariate Poisson Compound Distribution :

$$(N,M) = (\sum_{i=1}^{L} \Xi_i, \sum_{i=1}^{L} \Omega_i)$$

where the Ξ_i and Ω_i are not independent and L is Poisson distributed independently of the (Ξ_i, Ω_i).
We use the same methodology as in section 5.1.6 :

$$\begin{aligned}(U,V) &= (\sum_{i=1}^{\Xi} X_i, \sum_{i=1}^{\Omega} Y_i)\\ (S,T) &= (\sum_{i=1}^{L} U_i, \sum_{i=1}^{L} V_i)\end{aligned}$$

The bivariate Panjer's algorithm described in section 3.3.1 will be used in order to find the distribution of (S,T) knowing the distribution of (U,V).
In a first time we are interested to derive the distribution of (U,V). Therefore we first need to derive the distribution of (Ξ, Ω).
Remember that we have

$$\psi_{N,M}(u,v) = \psi_{N+M}(\rho_1 u + \rho_2 v)$$

We find

$$\begin{aligned}\psi_{N,M}(u,v) &= e^{-\theta(1+\beta-(1+\beta)(\rho_1 u+\rho_2 v))}\\ &= e^{-\theta(1+\beta)(1-\psi_{\xi,\Omega}(u,v))}\end{aligned}$$

where

$$\begin{aligned}\psi_{\Xi,\Omega}(0,0) &= 0\\ \psi_{\Xi,\Omega}(u,v) &= 1 - \frac{\theta((1+\beta)(1-(\rho_1 u+\rho_2 v)))}{\theta(1+\beta)}\end{aligned} \qquad (6.7)$$

(Ξ, Ω) has also a bivariate homogeneous distribution. Indeed

$$\begin{aligned}\psi_{\Xi,\Omega}(u,v) &= G(\rho_1 u + \rho_2 v)\\ G(x) &= 1 - \frac{\theta((1+\beta)(1-x)}{\theta(1+\beta)}\\ G(\rho_1+\rho_2) &= 1\end{aligned}$$

A Taylor expansion around $(1+\beta)$ of (6.7) gives after a few calculations

$$f_{\Xi,\Omega}(n,m) = \frac{\theta^{(m+n)}(1+\beta)}{\theta(1+\beta)} \frac{(-1)^{m+n-1}}{(m+n)!} \beta^m \frac{(n+m)!}{n!m!}$$

As we have

$$\begin{aligned}\theta(1+\beta) &= \frac{p}{c(1-a)}\left[(1+c(1+\beta))^{1-a}-1\right]\\ \theta^{(n+m)}(1+\beta) &= (-1)^{n+m-1}pc^{n+m-1}\frac{\Gamma(a+m+n-1)}{\Gamma(a)}(1+c(1+\beta))^{1-a-n-m}\end{aligned}$$

we immediately get

$$\begin{aligned}f_{\Xi,\Omega}(n,m) &= \frac{(n+m)!}{n!m!}\frac{c^{n+m}}{(n+m)!}\frac{1-a}{\Gamma(a)}\beta^m\Gamma(a+m+n-1)\frac{[1+c(1+\beta)]^{1-a-n-m}}{[1+c(1+\beta)]^{1-a}-1}\\ &= \frac{(n+m)!}{n!m!}\rho_1^n\rho_2^m\mathbb{P}(W=n+m)\end{aligned}$$

with

$$W = \Xi + \Omega$$

Indeed

$$\begin{aligned}\psi_W(u) &= \psi_{\Xi,\Omega}(u,u)\\ &= 1-\frac{\theta((1+\beta)(1-u))}{\theta(1+\beta)}\end{aligned} \tag{6.8}$$

A Taylor expansion around $(1+\beta)$ of (6.8) immediately shows that

$$\mathbb{P}(W=w) = \frac{c^w}{w!}\frac{1-a}{\Gamma(a)}\Gamma(a+w-1)(1+\beta)^w\frac{[1+c(1+\beta)]^{1-a-w}}{[1+c(1+\beta)]^{1-a}-1}$$

With the particular form of the distribution of W, we immediately find, with $f_W(w)=\mathbb{P}(W=w)$

$$\frac{f_W(w)}{f_W(w-1)} = \frac{c(1+\beta)}{1+c(1+\beta)}+\frac{c(1+\beta)(a-2)}{1+c(1+\beta)}\frac{1}{w} \quad , \quad w>1$$

which is well in the form of the $(r,s,1)$ class with

$$\begin{aligned}r &= \frac{c(1+\beta)}{1+c(1+\beta)}\\ s &= \frac{c(1+\beta)(a-2)}{1+c(1+\beta)}\\ f_W(0) &= 0\\ f_W(1) &= c(1+\beta)(1-a)\frac{[1+c(1+\beta)]^{-a}}{[1+c(1+\beta)]^{1-a}-1}\end{aligned}$$

We conclude that the distribution of W is a member of the $(r,s,1)$ class of counting distributions.

Now let us study the distribution of

$$(U,V) = (X_1+\cdots+X_\Xi, Y_1+\cdots+Y_\Omega)$$

If W is a member of the $(r,s,1)$ class we have :

$$\frac{d}{du}\psi_W(u)[1-ru] = w(1)+(r+s)\psi_W(u) \tag{6.9}$$

We are now able to extend Hesselager's methodology to find the aggregate claims distribution.

Theorem 6.4 *We have*

$$\begin{aligned}
f_{\Xi,\Omega}(1,0) &= \rho_1 f_W(1) && (6.10)\\
f_{\Xi,\Omega}(n,m) &= \rho_1\left(r+\frac{s}{n}\right)f_{\Xi,\Omega}(n-1,m)\\
&\quad +r\rho_2 f_{\Xi,\Omega}(n,m-1) \quad , \quad n\geq 1 \quad \textit{unless if } (n,m)=(1,0) && (6.11)\\
f_{\Xi,\Omega}(0,1) &= \rho_2 f_W(1) && (6.12)\\
f_{\Xi,\Omega}(n,m) &= \rho_2\left(r+\frac{s}{n}\right)f_{\Xi,\Omega}(n,m-1)\\
&\quad +r\rho_1 f_{\Xi,\Omega}(n-1,m) \quad , \quad m\geq 1 \quad \textit{unless if } (n,m)=(0,1) && (6.13)
\end{aligned}$$

Proof

We have already noticed that

$$\psi_{\Xi,\Omega}(u,v)=\psi_W(\rho_1 u+\rho_2 v) \tag{6.14}$$

By differentiating (6.14) with respect to u, and using (6.9), we get

$$(1-r\rho_1 u-r\rho_2 v)\frac{\partial}{\partial u}\psi_{\Xi,\Omega}(u,v)=\rho_1(f_W(1)+(r+s)\psi_{\Xi,\Omega}(u,v))$$

Inverting this expression we immediately get (6.10) and (6.11).
(6.12) and (6.13) are similarly derived. ∎

Theorem 6.5 *We have*

$$\begin{aligned}
f_{U,V}(0,0) &= 1-\frac{\theta((1+\beta)(1-(\rho_1 f_X(0)+\rho_2 f_Y(0))))}{\theta(1+\beta)} && (6.15)\\
f_{U,V}(x,0) &= \frac{1}{1-r\rho_1 f_X(0)}\,\Bigg(\rho_1 f_W(1)f_X(x)\\
&\quad +\rho_1\sum_{i=1}^{x}(r+\frac{si}{x})f_X(i)f_{U,V}(x-i,0)\Bigg) \quad , \quad x>0 && (6.16)\\
f_{U,V}(x,y) &= \frac{1}{1-r\rho_1 f_X(0)-r\rho_2 f_Y(0)}\left(\rho_1\sum_{i=1}^{x}(r+\frac{si}{x})f_X(i)f_{U,V}(x-i,y)\right.\\
&\quad \left. +\rho_2 r\sum_{j=1}^{y}f_Y(j)f_{U,V}(x,y-j)\right) \quad , \quad x>0\,,y>0 && (6.17)\\
f_{U,V}(0,y) &= \frac{1}{1-r\rho_2 f_Y(0)}\,\Bigg(\rho_2 f_W(1)f_Y(y)\\
&\quad +\rho_2\sum_{i=1}^{y}(r+\frac{si}{y})f_Y(i)f_{U,V}(0,y-i)\Bigg) \quad , \quad y>0 && (6.18)
\end{aligned}$$

$$\begin{aligned} f_{U,V}(x,y) &= \frac{1}{1 - r\rho_2 f_Y(0) - r\rho_1 f_X(0)} \left(\rho_2 \sum_{i=1}^{y} (r + \frac{si}{y}) f_Y(i) f_{U,V}(x, y-i) \right. \\ & \left. +\rho_1 r \sum_{j=1}^{x} f_X(j) f_{U,V}(x-j, y) \right) \quad , \quad y > 0\,, x > 0 \end{aligned} \tag{6.19}$$

Proof

As $f_{U,V}(0,0) = \psi_{\Xi,\Omega}(f_X(0), f_Y(0))$ we immediately find (6.15) by using equation (6.7).
¿From (6.11) we get

$$\begin{aligned} f_{\Xi,\Omega}(n,m) &= \rho_1(r + \frac{s}{n}) f_{\Xi,\Omega}(n-1,m) + r\rho_2 f_{\Xi,\Omega}(n,m-1) & (6.20) \\ &\Leftrightarrow \\ n f_{\Xi,\Omega}(n,m) &= r\rho_1(n-1) f_{\Xi,\Omega}(n-1,m) \\ & +\rho_1(r+s) f_{\Xi,\Omega}(n-1,m) + r\rho_2 n f_{\Xi,\Omega}(n,m-1) & (6.21) \end{aligned}$$

Multiplying both sides of (6.21) by $u\psi_X^{n-1}(u)\frac{d}{du}\psi_X(u)\psi_Y^m(v)$ and summing on $n = 1 \to \infty, m = 1 \to \infty$ gives

$$\begin{aligned} u\frac{\partial}{\partial u}\psi_{U,V}(u,v) &= r\rho_1 u \frac{\partial}{\partial u}\psi_{U,V}(u,v)\psi_X(u) \\ & +(r+s)\rho_1 u \frac{d}{du}\psi_X(u)\psi_{U,V}(u,v) + r\rho_2 \psi_{U,V}(u,v)\psi_Y(v) \end{aligned}$$

Inverting and rearranging this expression gives (6.17).
Multiplying both sides of (6.21) by $u\psi_X^{n-1}(u)\frac{d}{du}\psi_X(u)\psi_Y^m(0)$, summing on $n = 2 \to \infty, m = 0 \to \infty$ and adding $f_{\Xi,\Omega}(1,0)u\frac{d}{du}\psi_X(u)$ on both sides gives

$$\begin{aligned} u\frac{\partial}{\partial u}\psi_{U,V}(u,v) &= \rho_1 f_W(1)\psi_X(u) + r\rho_1 u \frac{\partial}{\partial u}\psi_{U,V}(u,v)\psi_X(u) \\ & +(r+s)\rho_1 u \frac{d}{du}\psi_X(u)\psi_{U,V}(u,v) \end{aligned}$$

Inverting and rearranging this expression gives (6.16).
(6.18) and (6.19) are derived similarly. ■

Knowing the distribution of (U, V) we must finally evaluate the distribution of

$$(S,T) = (\sum_{i=1}^{L} U_i, \sum_{i=1}^{L} V_i)$$

This is easily done with the bivariate Panjer's algorithm described section 3.3.1 for the particular case $L \sim Po(\theta(1+v))$.

We have

$$\begin{aligned}
f_{S,T}(0,0) &= e^{-\theta(1+\beta)(1-f_{U,V}(0,0))} & (6.22)\\
f_{S,T}(s,t) &= \sum_x^s \sum_y^t [\theta(1+\beta)\frac{x}{s}] f_{S,T}(s-x,t-y) f_{U,V}(x,y) \quad , \quad s \geq 1 & (6.23)\\
f_{S,T}(s,t) &= \sum_x^s \sum_y^t [\theta(1+\beta)\frac{y}{t}] f_{S,T}(s-x,t-y) f_{U,V}(x,y) \quad , \quad t \geq 1 & (6.24)
\end{aligned}$$

Note :

In the case of the Mixed Bivariate Negative Binomial Distribution, a one stage algorithm is given in Hesselager (1996).

6.1.2 Mixing the Bivariate Poisson model

This section is mainly based on Walhin (2000a).
Let the Bivariate Poisson Distribution, defined by its probability generating function :

$$\psi_{N,M}(u,v) = e^{\lambda_1(u-1)+\lambda_2(v-1)+\lambda_0(uv-1)}$$

Let Λ be a random variable characterizing an individual in the population. We define a new bivariate model by mixing with the random variable Λ. The probability generating function of the new model is given by

$$\begin{aligned}
\psi_{N,M}(u,v|\Lambda) &= e^{\Lambda(\lambda_1(u-1)+\lambda_2(v-1)+\lambda_0(uv-1))}\\
\psi_{N,M}(u,v) &= \int_0^\infty e^{\lambda(\lambda_1(u-1)+\lambda_2(v-1)+\lambda_0(uv-1))} dU(\lambda)\\
\psi_{N,M}(u,v) &= \phi_\Lambda(\lambda_1(u-1)+\lambda_2(v-1)+\lambda_0(uv-1))
\end{aligned}$$

where $\phi_\Lambda(u)$ is the moment generating function of Λ. In the present section we will choose Λ being $Homix(1,c,a)$. The reason why we take $p=1$ is to make the parameters identifiable. As was shown in Kocherlakota (1988), there exists a recursion giving the probability function :

$$\begin{aligned}
p(n,0) &= \frac{\lambda_1^n}{n!} \left.\frac{d^n}{dz^n}\phi_\Lambda(z)\right|_{z=\gamma}\\
p(0,m) &= \frac{\lambda_1^m}{m!} \left.\frac{d^m}{dz^m}\phi_\Lambda(z)\right|_{z=\gamma}\\
p(n,m) &= \frac{\lambda_2}{\lambda_1}\frac{n+1}{m}p(n+1,m-1) + \frac{\lambda_0}{\lambda_1}\frac{n-m+1}{m}p(n,m-1) \quad , \quad n \geq m\\
p(n,m) &= \frac{\lambda_1}{\lambda_2}\frac{m+1}{n}p(n-1,m+1) + \frac{\lambda_0}{\lambda_2}\frac{m-n+1}{n}p(n-1,m) \quad , \quad n \leq m
\end{aligned}$$

where γ stands for $-(\lambda_1+\lambda_2+\lambda_0)$.
Obviously, in our case, derivating $\phi_\Lambda(z)$ is tedious. However it is possible to find easier

formulae for evaluating $p(n,0)$ and $p(0,m)$. Indeed

$$\begin{aligned} p(n,0) &= \frac{\lambda_1^n}{n!}\int_0^\infty \lambda^n e^{\lambda\gamma} dU(\lambda) \\ &= \left(\frac{\lambda_1}{-\gamma}\right)^n \int_0^\infty \frac{(\lambda(-\gamma))^n}{n!} e^{-(-\gamma)\lambda} dU(\lambda) \\ &= \left(\frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_0}\right)^n \Pi(n,\lambda_1+\lambda_2+\lambda_0) \end{aligned}$$

and $\Pi(n,\lambda_1+\lambda_2+\lambda_0)$ is easily given by the recursion (5.7).
Similarly we have

$$p(0,m) = \left(\frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_0}\right)^m \Pi(m,\lambda_1+\lambda_2+\lambda_0)$$

Therefore it is easy to find the probability function of the bivariate vector (N,M).
Note that numerical examples show that the following maximum likelihood equations are true:

$$\begin{aligned} \bar{N} &= \hat{\lambda}_1 + \hat{\lambda}_0 \\ \bar{M} &= \hat{\lambda}_2 + \hat{\lambda}_0 \end{aligned}$$

Whether this is true or not remains to be shown. Note that if $\lambda_0 = 0$, this result has been shown in section 6.1.1 with the reparametrization :

$$\begin{aligned} \lambda_0 &= 0 \\ \lambda_1 &= p \\ \lambda_2 &= p\beta \\ c &= \frac{c'}{p} \\ a &= a' \end{aligned}$$

Let us fit our reference data set :

This fit is also quite acceptable.

The marginal distributions of N and M are

$$\begin{aligned} N &\sim Ho(\lambda_0+\lambda_1, c(\lambda_0+\lambda_1), a) \\ M &\sim Ho(\lambda_0+\lambda_2, c(\lambda_0+\lambda_2), a) \end{aligned}$$

whereas the distribution of the sum $N+M$ is a Mixed Hermite Distribution with probability generating function :

$$\psi_{N+M}(u) = \int_0^\infty e^{\lambda((\lambda_1+\lambda_2)(u-1)-\lambda_0(u^2-1))} dU(\lambda)$$

Unfortunately it does not seem possible to find a recursive scheme in this model in order to find the probability function of the bivariate compound distribution when the bivariate counting distribution is Mixed Bivariate Hofmann.

N/M		0	1	2	3	4	5	6	7		
Obs	0	21	18	8	2	1	0	0	0	l	-341.607
Fitted		21.82	16.74	8.00	3.06	1.03	0.32	0.09	0.03	χ^2	1.546
Obs	1	13	14	10	1	4	1	0	0	df	7
Fitted		12.57	13.23	8.06	3.76	1.49	0.53	0.18	0.05	$p-value$	0.981
Obs	2	4	5	4	2	1	0	1	0	a	0.8799
Fitted		4.51	6.05	4.52	2.51	1.16	0.47	0.18	0.06	c	0.2953
Obs	3	2	1	3	2	0	1	0	0	λ_1	1.1847
Fitted		1.30	2.12	1.89	1.22	0.65	0.30	0.12	0.05	λ_2	0.8896
Obs	4	0	0	1	1	0	0	0	0	λ_0	0.0858
Fitted		0.33	0.63	0.65	0.49	0.29	0.15	0.07	0.02	AIC	-346.607
Obs	5	0	0	0	0	0	0	0	0	BIC	-353.61
Fitted		0.08	0.17	0.20	0.17	0.11	0.06	0.02	0.00		
Obs	6	0	0	0	0	0	0	0	0		
Fitted		0.02	0.04	0.06	0.05	0.04	0.01	0.00	0.00		
Obs	7	0	1	0	0	0	0	0	0		
Fitted		0.00	0.01	0.01	0.01	0.01	0.00	0.00	0.00		

Table 6.3: Mixed Bivariate Hofmann fit

6.2 The TRM Bivariate Hofmann Distribution

This section is mainly based on Walhin and Paris (2000h).
Ahmed (1961) and Papageorgiou and David (1995) discuss some bivariate counting distributions, namely, the joint distribution (N, M) where $N = N_0 + N_1$ and $M = N_0 + N_2$ with N_0, N_1 and N_2 independent random variables such that

$$\begin{aligned}
\mathbb{P}(N_0 = n) &= \int_0^\infty e^{-\lambda} \frac{\lambda^n}{n!} dU(\lambda) \quad , \quad n \geq 0 \\
\mathbb{P}(N_1 = n) &= e^{-\lambda_1} \frac{\lambda_1^n}{n!} \quad , \quad n \geq 0 \\
\mathbb{P}(N_2 = n) &= e^{-\lambda_2} \frac{\lambda_2^n}{n!} \quad , \quad n \geq 0
\end{aligned}$$

i.e. N_1 and N_2 are Poisson distributed while N_0 is Mixed Poisson distributed with mixing distribution Λ.
The joint probability function of (N, M) is given by

$$\mathbb{P}(N = n, M = m) = \sum_{k=0}^{\min(n,m)} \mathbb{P}(N_0 = k)\mathbb{P}(N_1 = n - k)\mathbb{P}(N_2 = m - k)$$

For some choices of the mixing distribution Λ, Papageorgiou and David (1995) give the probability function of (N, M) by using Stirling numbers of the second kind, C-numbers and modified Bessel functions of the third kind.
Using a the Hofmann Distribution, it is possible to give simple expressions for the joint distribution of (N, M) which avoids these numbers.

We will assume two models. In the former it is assumed that N_1 and N_2 are Poisson distributed whereas an extension of model 1, namely model 2, will consider all N_i being Hofmann distributed.

The reason why we call this model the TRM Bivariate Hofmann Distribution is that this Bivariate Hofmann Distribution is obtained by the Trivariate Reduction Method (TRM).

6.2.1 Model 1

In this section we work with the model

$$(N, M) = (N_0 + N_1, N_0 + N_2)$$

where N_0 is Ho(p,c,a) and N_1 and N_2 are respectively Po(λ_1) and Po(λ_2). The three random variables are assumed to be mutually independent.

Theorem 6.6

$$\begin{aligned}
p(0,0) &= e^{-\theta(1)-\lambda_1-\lambda_2} \\
np(n,m) &= \theta(1) \sum_{k=1}^{\min(n,m)} kf_{\Xi_0}(k)p(n-k,m-k) + \lambda_1 p(n-1,m) \quad , \quad n > 0 \\
mp(n,m) &= \theta(1) \sum_{k=1}^{\min(n,m)} kf_{\Xi_0}(k)p(n-k,m-k) + \lambda_2 p(n,m-1) \quad , \quad m > 0
\end{aligned}$$

Proof
We have

$$\begin{aligned}
\psi_{N,M}(u,v) &= \mathbb{E}[u^N v^M] = \mathbb{E}[(uv)^{N_0} u^{N_1} v^{N_2}] \\
&= \psi_{N_0}(uv)\psi_{N_1}(u)\psi_{N_2}(v) \\
&= e^{-\theta(1)[1-\psi_{\Xi_0}(uv)]} e^{-\lambda_1(1-u)} e^{-\lambda_2(1-v)}
\end{aligned}$$

Differentiating with respect to u and multiplying by u gives

$$u\frac{\partial \psi_{N,M}(u,v)}{\partial u} = \theta(1)u\frac{\partial \psi_{\Xi_0}(uv)}{\partial u}\psi_{N,M}(u,v) + u\lambda_1\psi_{N,M}(u,v)$$

Inverting this expression gives

$$np(n,m) = \theta(1) \sum_{k=1}^{\min(n,m)} kf_{\Xi_0}(k)p(n-k,m-k) + \lambda_1 p(n-1,m) \quad , \quad n > 0$$

Differentiating with respect to v gives the symmetric recursion. ■

Let us note that for the particular cases where N_0 is Negative Binomial or Poisson Inverse Gaussian, we have easier recursions.
The case Negative Binomial ($a = 1$) is given in Hesselager (1996) where the fact that the

Negative Binomial belongs to the $(r, s, 0)$ class is used.
In this case the recursion becomes (Hesselager (1996)) :

$$\begin{aligned}
p(0,0) &= (1+c)^{-\frac{p}{c}} e^{-\lambda_1-\lambda_2} \\
p(n,m) &= (r+\frac{s}{n})p(n-1,m-1) + \frac{\lambda_1}{n}p(n-1,m) \\
&\quad -\frac{\lambda_1 r}{n}p(n-2,m-1) \quad , \quad n>0 \\
p(n,m) &= (r+\frac{s}{m})p(n-1,m-1) + \frac{\lambda_2}{m}p(n,m-1) \\
&\quad -\frac{\lambda_2 r}{m}p(n-1,m-2) \quad , \quad m>0
\end{aligned}$$

The case Poisson Inverse Gaussian ($a = \frac{1}{2}$) is derived without using the particular structure of the probability generating function of N (equation (5.9)) :

$$\begin{aligned}
\psi_{N,M}(u,v) &= e^{-\frac{2p}{c}((1+c(1-uv))^{\frac{1}{2}}-1)} e^{-\lambda_1(1-u)} e^{-\lambda_2(1-v)} \qquad (6.25) \\
\frac{\partial}{\partial u}\psi_{N,M}(u,v) &= (\lambda_1 + \frac{pv}{1+c(1-uv)^{\frac{1}{2}}})\psi_{N,M}(u,v) \\
\frac{\partial^2}{\partial u^2}\psi_{N,M}(u,v) &= \lambda_1^2\psi_{N,M}(u,v) + \frac{1}{2}\frac{cpv^2}{(1+c(1-uv))^{\frac{3}{2}}}\psi_{N,M}(u,v) \\
&\quad + \frac{p^2v^2}{1+c(1-uv)}\psi_{N,M}(u,v) + \frac{2\lambda_1 pv}{(1+c(1-uv))^{\frac{1}{2}}}\psi(u,v) \\
(1+c(1-uv))\frac{\partial^2}{\partial u^2}\psi_{N,M}(u,v) &= \psi_{N,M}(u,v)(-\lambda_1^2(1+c) + \lambda_1^2 cuv - \frac{1}{2}\lambda_1 cv + p^2v^2) \\
&\quad + \frac{\partial}{\partial u}\psi_{N,M}(u,v)(2\lambda_1(1+c) - 2\lambda_1 cuv + \frac{1}{2}cv) \qquad (6.26)
\end{aligned}$$

The initializing terms are easily found from (6.25) :

$$\begin{aligned}
p(0,0) &= e^{-\lambda_1-\lambda_2-2\frac{p}{c}((1+c)^{0.5}-1)} \\
p(1,0) &= \lambda_1 p(0,0) \\
p(0,1) &= \lambda_2 p(0,0) \\
p(1,1) &= (\lambda_1\lambda_2 + \frac{p}{(1+c)^{\frac{1}{2}}})p(0,0)
\end{aligned}$$

Inverting (6.26) and using its similar expression in v gives

$$\begin{aligned}
(1+c)n(n-1)p(n,m) &= c(n-1)(n-\frac{3}{2})p(n-1,m-1) - \lambda_1 c(2n-\frac{7}{2})p(n-2,m-1) \\
&\quad +2\lambda_1(1+c)(n-1)p(n-1,m) - \lambda_1^2(1+c)p(n-2,m) \\
&\quad +\lambda_1^2 cp(n-3,m-1) + p^2p(n-2,m-2) \quad , \quad n\geq 2 \\
(1+c)m(m-1)p(n,m) &= c(m-1)(m-\frac{3}{2})p(n-1,m-1) - \lambda_2 c(2m-\frac{7}{2})p(n-1,m-2) \\
&\quad +2\lambda_2(1+c)(m-1)p(n,m-1) - \lambda_2^2(1+c)p(n,m-2) \\
&\quad +\lambda_2^2 cp(n-1,m-3) + p^2p(n-2,m-2) \quad , \quad m\geq 2
\end{aligned}$$

Note that for both simplified recursions, there is a risk of numerical instability due to numbers close to zero that have to be substracted.

Let us fit our reference data set :

N/M		0	1	2	3	4	5	6	7		
Obs Fitted	0	21 17.13	18 17.46	8 8.90	2 3.02	1 0.77	0 0.16	0 0.03	0 0.00	l χ^2	-345.25 3.006
Obs Fitted	1	13 12.40	14 15.92	10 9.78	1 3.89	4 1.14	1 0.26	0 0.05	0 0.01	df $p-value$	7 0.884
Obs Fitted	2	4 4.49	5 6.95	4 5.51	2 2.80	1 1.02	0 0.28	1 0.06	0 0.01	p_0 a_0	0.2514 140.866
Obs Fitted	3	2 1.08	1 1.96	3 1.99	2 1.34	0 0.64	1 0.22	0 0.06	0 0.01	c_0 λ_1	0.00194 0.7240
Obs Fitted	4	0 0.20	0 0.41	1 0.51	1 0.45	0 0.28	0 0.13	0 0.04	0 0.01	λ_2 AIC	1.0191 -350.25
Obs Fitted	5	0 0.03	0 0.07	0 0.10	0 0.11	0 0.09	0 0.05	0 0.02	0 0.01	BIC	-357.26
Obs Fitted	6	0 0.00	0 0.01	0 0.02	0 0.02	0 0.02	0 0.02	0 0.01	0 0.00		
Obs Fitted	7	0 0.00	1 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00		

Table 6.4: TRM Bivariate Hofmann fit ; model 1

Once again the fit is quite acceptable. Note that one might think that the component N_0 tends to a Neyman Type A distribution. The particular case $N_0 \sim$ Neyman Type A has been fitted but it does not show a higher log-likelihood than the model proposed hereabove.

We now study the probability function of the compound bivariate distribution when the bivariate counting distribution comes from model 1:

$$(S,T) = (X_1 + \cdots + X_{N_0+N_1}, Y_1 + \cdots + Y_{N_0+N_2})$$

the probability function of which is given by

$$\mathbb{P}[S = x, T = y] = f_{S,T}(x,y) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} p(n,m) f_X^{*n}(x) f_Y^{*m}(y)$$

Our aim is to give a recursive scheme in order to derive the probability function $f_{S,T}(x,y)$. We have

$$\begin{aligned} \psi_{S,T}(u,v) &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} p(n,m)\psi_X^n(u)\psi_Y^m(v) \\ &= \psi_{N,M}(\psi_X(u), \psi_Y(v)) \\ &= e^{-\theta(1)[1-\psi_{\Xi_0}(\psi_X(u)\psi_Y(v))]} e^{-\lambda_1(1-\psi_X(u))} e^{-\lambda_2(1-\psi_Y(v))} \end{aligned}$$

Differentiating with respect to u and multiplying by u gives

$$u\frac{\partial \psi_{S,T}(u,v)}{\partial u} = \theta(1)u\frac{\partial \psi_{\Xi_0}(\psi_X(u)\psi_Y(v))}{\partial u}\psi_{S,T}(u,v) + \lambda_1 u \frac{\partial \psi_X(u)}{\partial u}\psi_{S,T}(u,v) \tag{6.27}$$

where $\psi_{U,V}(u,v) = \psi_{\Xi_0}(\psi_X(u)\psi_Y(v))$ is the probability generating function of the pair

$$(U = X_1 + \cdots + X_{\Xi_0}, V = Y_1 + \cdots + Y_{\Xi_0})$$

the probability function of which will be denoted by

$$f_{U,V}(i,j) = \mathbb{P}(X_1 + \cdots + X_{\Xi} = i, Y_1 + \cdots + Y_{\Xi} = j) \quad , \quad i \geq 0 \quad , \quad j \geq 0$$

The probability function of (U,V) is easily given by theorem 3.5.
Inverting (6.27) gives

$$x f_{S,T}(x,y) = \theta(1) \sum_{i=0}^{x} \sum_{j=0}^{y} i f_{U,V}(i,j) f_{S,T}(x-i,y-j) + \lambda_1 \sum_{i=0}^{x} i f_X(i) f_{S,T}(x-i,y) \quad , \quad x > 0 \tag{6.28}$$

With the symmetric expression of (6.28) we have the following result

Theorem 6.7

$$\begin{aligned}
f_{S,T}(0,0) &= e^{-\theta(1-f_X(0)f_Y(0))} \\
f_{S,T}(x,y) &= \theta(1) \sum_{i=0}^{x} \sum_{j=0}^{y} \frac{i}{x} f_{U,V}(i,j) f_{S,T}(x-i,y-j) + \lambda_1 \sum_{i=0}^{x} \frac{i}{x} f_X(i) f_{S,T}(x-i,y) \quad , \quad x > 0 \\
f_{S,T}(x,y) &= \theta(1) \sum_{i=0}^{x} \sum_{j=0}^{y} \frac{j}{y} f_{U,V}(i,j) f_{S,T}(x-i,y-j) + \lambda_2 \sum_{j=0}^{y} \frac{j}{y} f_Y(j) f_{S,T}(x,y-j) \quad , \quad y > 0
\end{aligned}$$

Proof

Using (6.28), its similar expression valid for $y > 0$ gives the proof. ■

6.2.2 Model 2

In this section we consider the following model

$$(N, M) = (N_0 + N_1, N_0 + N_2)$$

where N_i are independent $Ho(p_i, c_i, a_i)$. The corresponding probability laws of the Ξ_i are noted f_{Ξ_i}.
Then we have the following results. The proofs are similar to those given in section 6.2.1. So we neglect them.

Theorem 6.8

$$\begin{aligned}
p(0,0) &= e^{-\theta_0(1)-\theta_1(1)-\theta_2(1)} \\
p(n,m) &= \sum_{i=1}^{\min(n,m)} \frac{i}{n} \theta_0(1) f_{\Xi_0}(i) p(n-i,m-i) + \sum_{i=1}^{n} \frac{i}{n} \theta_1(1) f_{\Xi_1}(i) p(n-i,m) \quad , \quad n > 0 \\
p(n,m) &= \sum_{i=1}^{\min(n,m)} \frac{i}{m} \theta_0(1) f_{\Xi_0}(i) p(n-i,m-i) + \sum_{i=1}^{n} \frac{i}{m} \theta_2(1) f_{\Xi_2}(i) p(n,m-i) \quad , \quad n > 0
\end{aligned}$$

Theorem 6.9

$$
\begin{aligned}
f_{S,T}(0,0) &= e^{-\theta_0(1-f_X(0)f_Y(0))}e^{-\theta_1(1-f_X(0))}e^{-\theta_2(1-f_Y(0))} \\
f_{S,T}(x,y) &= \theta_0(1)\sum_{i=1}^{x}\sum_{j=0}^{y}\frac{i}{x}f_{U,V}(i,j)f_{S,T}(x-i,y-j)+\theta_1(1)\sum_{i=1}^{x}\frac{i}{x}f_P(i)f_{S,T}(x-i,y) \quad , \quad x>0 \\
f_{S,T}(x,y) &= \theta_0(1)\sum_{i=0}^{x}\sum_{j=1}^{y}\frac{j}{y}f_{U,V}(i,j)f_{S,T}(x-i,y-j)+\theta_2(1)\sum_{j=1}^{y}\frac{j}{y}f_Q(j)f_{S,T}(x,y-j) \quad , \quad y>0
\end{aligned}
$$

where

$$
\begin{aligned}
(U,V) &= (X_1+\cdots+X_{\Xi_0}, Y_1+\cdots+Y_{\Xi_0}) \\
P &= X_1+\cdots+X_{\Xi_1} \\
Q &= Y_1+\cdots+Y_{\Xi_2}
\end{aligned}
$$

Note : for the case where the N_i belong to the $(r,s,0)$ class, Hesselager (1996) gives an easier algorithm. However, for the case Negative Binomial which is a member of the $(r,s,0)$ class, numerical examples show that this algorithm is not stable whereas the combination of theorems 3.5 and 6.9 give stable recursions. See Panjer and Wang (1993) for comments on the stability of recursions.
Let us fit our reference data set :

N/M		0	1	2	3	4	5	6	7		
Obs	0	21	18	8	2	1	0	0	0	l	-340.79
Fitted		21.56	16.20	7.89	3.15	1.12	0.37	0.11	0.03	χ^2	1.832
Obs	1	13	14	10	1	4	1	0	0	df	3
Fitted		12.34	14.75	8.63	3.80	1.44	0.49	0.16	0.05	$p-value$	0.608
Obs	2	4	5	4	2	1	0	1	0	p_0	0.1383
Fitted		3.95	6.10	4.99	2.62	1.10	0.40	0.14	0.04	a_0	$\rightarrow \infty$
Obs	3	2	1	3	2	0	1	0	0	c_0	$\rightarrow 0$
Fitted		1.05	1.79	1.82	1.25	0.61	0.25	0.09	0.03	a_0c_0	0.0822
Obs	4	0	0	1	1	0	0	0	0	p_1	0.6712
Fitted		0.30	0.49	0.53	0.43	0.26	0.12	0.05	0.02	c_1	6.6765
Obs	5	0	0	0	0	0	0	0	0	a_1	0.0781
Fitted		0.12	0.16	0.16	0.13	0.09	0.05	0.02	0.01	p_2	0.9663
Obs	6	0	0	0	0	0	0	0	0	c_2	0.2824
Fitted		0.06	0.07	0.06	0.04	0.03	0.02	0.01	0.00	a_2	1.0114
Obs	7	0	1	0	0	0	0	0	0	AIC	-349.79
Fitted		0.03	0.04	0.03	0.02	0.01	0.0100	0.00	0.00	BIC	-362.40

Table 6.5: TRM Bivariate Hofmann fit ; model 2

We are not surprized to note that this model is overparametrized.
It seems that, as an extension of the result in the univariate case, we have the following maximum likelihood equations :

$$\hat{p}_0+\hat{p}_1 = \bar{N}$$

$$\hat{p}_0 + \hat{p}_2 = \bar{M}$$

This is a conjecture that remains to be shown.

6.3 The Bivariate Zero Inflated Poisson model

This section is mainly based on Walhin (2000b). We will first use the mixing method with the particular nonparametric case leading to the mixture of a Poisson Distribution with a degenerate at 0 distribution. Then we will study the Trivariate Reduction Method with Zero Inflated Poisson Distributions.

6.3.1 Model 1

The classical Bivariate Poisson Distribution is modelled by its probability generating function (see Kocherlakota and Kocherlakota (1992)) :

$$\psi_{N,M}(u,v) = e^{\lambda_1(u-1)+\lambda_2(v-1)+\lambda_0(uv-1)}$$

It is not difficult to find a more general bivariate distribution by mixing :

$$\psi_{N,M}(u,v|\Lambda) = e^{\Lambda\lambda_1(u-1)+\Lambda\lambda_2(v-1)+\Lambda\lambda_0(uv-1)}$$

By choosing the nonparametric structure for the random variable Λ with two mass points : 0 and 1 with probabilities p and $1-p$ we have

$$\psi_{N,M}(u,v) = p + (1-p)e^{\lambda_1(u-1)+\lambda_2(v-1)+\lambda_0(uv-1)}$$

This is clearly a Zero Inflated model in the bivariate setting.
The probability function is easily given by

$$\begin{aligned}
f(0,0) &= p + (1-p)e^{-\lambda_1-\lambda_2-\lambda_0} \\
f(1,0) &= \lambda_1 f(0,0) - p\lambda_1 \\
f(1,1) &= \lambda_1 f(0,1) + \lambda_0 f(0,0) - p\lambda_0 \\
f(n,m) &= \frac{1}{n}\left(\lambda_1 f(n-1,m) + \lambda_0 f(n-1,m-1)\right), n > 0, (n,m) \neq (1,0),(1,1) \quad (6.29) \\
f(0,1) &= \lambda_2 f(0,0) - p\lambda_2 \\
f(1,1) &= \lambda_2 f(1,0) + \lambda_0 f(0,0) - p\lambda_0 \\
f(n,m) &= \frac{1}{m}\left(\lambda_2 f(n,m-1) + \lambda_0 f(n-1,m-1)\right), m > 0, (n,m) \neq (0,1),(1,1) \quad (6.30)
\end{aligned}$$

These recursions are found by equating the terms in $u^n v^m$ in the following equations :

$$\begin{aligned}
\frac{\partial\psi_{N,M}(u,v)}{\partial u} &= \lambda_1\psi_{N,M}(u,v) + \lambda_0 v\psi_{N,M}(u,v) - p\lambda_1 - p\lambda_0 v \\
\frac{\partial\psi_{N,M}(u,v)}{\partial v} &= \lambda_2\psi_{N,M}(u,v) + \lambda_0 u\psi_{N,M}(u,v) - p\lambda_2 - p\lambda_0 u
\end{aligned}$$

Let us define $\alpha = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty}\alpha_{ij}$ where α_{ij} is the number of observations corresponding to $(N=i, M=j)$.
We have the following maximum likelihood result :

Theorem 6.10

$$\begin{aligned}
\hat{p} + (1-\hat{p})e^{-(\hat{\lambda}_0+\hat{\lambda}_1+\hat{\lambda}_2)} &= \frac{\alpha_{0,0}}{\alpha} \\
(1-\hat{p})(\hat{\lambda}_0+\hat{\lambda}_1) &= \bar{n} \\
(1-\hat{p})(\hat{\lambda}_0+\hat{\lambda}_2) &= \bar{m}
\end{aligned}$$

Proof :
As we are interested in the maximum likelihood estimation, we write the loglikelihood :

$$l_{n,m}(p,\lambda_1,\lambda_2,\lambda_0) = \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \alpha_{n,m} \ln(f(n,m))$$

The score equations will require the knowledge of $\frac{\partial \ln(f(n,m))}{\partial \theta}$ where θ is any one of the four parameters of model 2.
Therefore we study the derivative of the probability generating function with respect to the four parameters :

$$\begin{aligned}
\frac{\partial P(u,v)}{\partial \lambda_1} &= uP(u,v) - P(u,v) - pu + p \\
\frac{\partial P(u,v)}{\partial \lambda_2} &= vP(u,v) - P(u,v) - pv + p \\
\frac{\partial P(u,v)}{\partial \lambda_0} &= uvP(u,v) - P(u,v) - puv + p \\
\frac{\partial P(u,v)}{\partial p} &= \frac{1-P(u,v)}{1-p}
\end{aligned}$$

Equating the coefficients of the terms in $u^n v^m$ gives

$$\begin{aligned}
\frac{\partial f(0,0)}{\partial \lambda_1} &= p - f(0,0) \\
\frac{\partial f(1,0)}{\partial \lambda_1} &= f(0,0) - f(1,0) - p \\
\frac{\partial f(n,m)}{\partial \lambda_1} &= f(n-1,m) - f(n,m) \quad , \quad (n,m) \neq (0,0),(1,0) \\
\frac{\partial f(0,0)}{\partial \lambda_2} &= p - f(0,0) \\
\frac{\partial f(0,1)}{\partial \lambda_2} &= f(0,0) - f(0,1) - p \\
\frac{\partial f(n,m)}{\partial \lambda_2} &= f(n-1,m) - f(n,m) \quad , \quad (n,m) \neq (0,0),(0,1) \\
\frac{\partial f(0,0)}{\partial \lambda_0} &= p - f(0,0) \\
\frac{\partial f(1,0)}{\partial \lambda_0} &= -f(1,0)
\end{aligned}$$

$$\begin{aligned}
\frac{\partial f(0,1)}{\partial \lambda_0} &= -f(0,1) \\
\frac{\partial f(1,1)}{\partial \lambda_0} &= f(0,0) - f(1,1) - p \\
\frac{\partial f(n,m)}{\partial \lambda_0} &= f(n-1,m) - f(n,m) \quad , \quad (n,m) \neq (0,0),(0,1),(1,0),(1,1) \\
\frac{\partial f(0,0)}{\partial p} &= \frac{1-f(0,0)}{1-p} \\
\frac{\partial f(1,0)}{\partial p} &= \frac{-f(n,m)}{1-p}
\end{aligned}$$

The score equation for p writes :

$$\begin{aligned}
\frac{\partial l_{n,m}(\lambda_1,\lambda_2,\lambda_0,p)}{\partial p} &= \sum_{n=0}^{\infty}\sum_{m=0}^{\infty} \alpha_{n,m} \frac{\frac{\partial f(n,m)}{\partial p}}{f(n,m)} = 0 \\
&= \frac{\alpha_{0,0}}{(1-p)f(0,0)} - \frac{\alpha}{1-p} = 0
\end{aligned}$$

implying

$$\hat{p} + (1-\hat{p})e^{-\hat{\lambda}_1-\hat{\lambda}_2-\hat{\lambda}_0} = \frac{\alpha_{0,0}}{\alpha}$$

The score equation for λ_1, λ_2 and λ_0 write, after some calculations :

$$\frac{\partial l_{n,m}}{\partial \lambda_1} = 0 = \frac{\alpha_{0,0}p}{f(0,0)} + \frac{\alpha_{1,1}f(0,1)}{f(1,1)} + \frac{\alpha_{1,0}}{\lambda_1} - \alpha + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \alpha_{n,m} \frac{f(n-1,m)}{f(n,m)} \tag{6.31}$$

$$\frac{\partial l_{n,m}}{\partial \lambda_2} = 0 = \frac{\alpha_{0,0}p}{f(0,0)} + \frac{\alpha_{1,1}f(1,0)}{f(1,1)} + \frac{\alpha_{0,1}}{\lambda_2} - \alpha + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \alpha_{n,m} \frac{f(n,m-1)}{f(n,m)} \tag{6.32}$$

$$\frac{\partial l_{n,m}}{\partial \lambda_0} = 0 = -\alpha + \frac{\alpha_{0,0}p}{f(0,0)} - \frac{\alpha_{1,1}p}{f(1,1)} + \frac{\alpha_{1,1}f(0,0)}{f(1,1)} + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \alpha_{n,m} \frac{f(n-1,m-1)}{f(n,m)}$$

According to equations (6.29) and (6.30) we are able to write

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \alpha_{n,m} \frac{f(n-1,m)}{f(n,m)} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \alpha_{n,m} \left[\frac{n}{\lambda_1} - \frac{\lambda_0}{\lambda_1}\frac{f(n-1,m-1)}{f(n,m)}\right] \tag{6.33}$$

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \alpha_{n,m} \frac{f(n,m-1)}{f(n,m)} = \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \alpha_{n,m} \left[\frac{m}{\lambda_2} - \frac{\lambda_0}{\lambda_2}\frac{f(n-1,m-1)}{f(n,m)}\right] \tag{6.34}$$

The score equation in λ_0 can be rewritten as

$$\sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \alpha_{n,m} \frac{f(n-1,m-1)}{f(n,m)} = \alpha - \frac{\alpha_{0,0}p}{f(0,0)} + \frac{\alpha_{1,1}p}{f(1,1)} - \frac{\alpha_{1,1}f(0,0)}{f(1,1)} \tag{6.35}$$

Inserting (6.35) and (6.33) in (6.31) we find

$$\frac{\alpha_{0,0}p}{f(0,0)} + \frac{\alpha_{1,1}f(0,1)}{f(1,1)} - \frac{\alpha_{1,1}}{\lambda_1} - \alpha + \frac{\alpha\bar{n}}{\lambda_1} - \frac{\lambda_0}{\lambda_1}\left[\alpha - \frac{\alpha_{0,0}p}{f(0,0)} - \frac{\alpha_{1,1}f(0,0)}{f(1,1)} + \frac{\alpha_{1,1}p}{f(1,1)}\right] = 0$$

Some algebra leads immediately to the following equation :

$$\bar{n} = (\hat{\lambda}_1 + \hat{\lambda}_0)(1 - \hat{p})$$

Similarly we have the other moment equation.

$$\bar{m} = (\hat{\lambda}_2 + \hat{\lambda}_0)(1 - \hat{p})$$

In conclusion the maximum likelihood estimates of model 2 are given as the solution of the following system :

$$\begin{aligned}
\frac{\alpha_{0,0}}{\alpha} &= \hat{p} + (1-\hat{p})e^{-\hat{\lambda}_1 - \hat{\lambda}_2 - \hat{\lambda}_3} \\
\bar{n} &= (\hat{\lambda}_1 + \hat{\lambda}_0)(1-\hat{p}) \\
\bar{m} &= (\hat{\lambda}_2 + \hat{\lambda}_0)(1-\hat{p}) \\
0 &= -\alpha + \frac{\alpha_{0,0}p}{f(0,0)} - \frac{\alpha_{1,1}p}{f(1,1)} + \frac{\alpha_{1,1}f(0,0)}{f(1,1)} + \sum_{n=1}^{\infty}\sum_{m=1}^{\infty} \alpha_{n,m}\frac{f(n-1,m-1)}{f(n,m)}
\end{aligned}$$

■

Let us fit our reference data set :

N/M		0	1	2	3	4	5	6	7		
Obs Fitted	0	21 21.00	18 14.56	8 8.32	2 3.17	1 0.91	0 0.21	0 0.04	0 0.01	l χ^2	−344.55 3.573
Obs Fitted	1	13 10.53	14 14.83	10 10.07	1 4.45	4 1.45	1 0.37	0 0.08	0 0.01	df $p-value$	8 0.893
Obs Fitted	2	4 4.35	5 7.29	4 5.79	2 2.95	1 1.09	0 0.31	1 0.07	0 0.01	p λ_0	0.0677 0.2199
Obs Fitted	3	2 1.20	1 2.33	3 2.13	2 1.24	0 0.51	1 0.17	0 0.04	0 0.01	λ_1 λ_2	0.8264 1.1429
Obs Fitted	4	0 0.25	0 0.55	1 0.57	1 0.37	0 0.17	0 0.06	0 0.02	0 0.00	AIC BIC	−348.55 −354.15
Obs Fitted	5	0 0.04	0 0.10	0 0.12	0 0.09	0 0.05	0 0.02	0 0.01	0 0.00		
Obs Fitted	6	0 0.01	0 0.02	0 0.02	0 0.02	0 0.01	0 0.00	0 0.00	0 0.00		
Obs Fitted	7	0 0.00	1 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00	0 0.00		

Table 6.6: Mixed Bivariate ZIP fit

This model gives an excellent fit.

6.3.2 Model 2

A classical way of obtaining a bivariate distribution is to use the trivariate reduction method:

$$N = N_0 + N_1$$
$$M = N_0 + N_2$$

where the N_i , $i = 0, 1, 2$ are independent random variables. If the N_i are Poisson distributed, we find the classical bivariate Poisson Distribution. We will assume now that the N_i are ZIP distributed. This model is certainly less intuitive than the previous one in the sense that the construction of the bivariate dependence is more technical. However as we will see it gives good fits.
Let us assume that

$$\begin{aligned}
\mathbb{P}[N_i = 0] &= (1 - p_i) + p_i e^{-\lambda_i} \\
\mathbb{P}[N_i = y] &= (1 - p_i) \frac{\lambda_i^y}{y!} e^{-\lambda_i} \quad , \quad y > 0
\end{aligned}$$

A particular case would be to assume that $p_1 = p_2 = 0$. This is the situation where the zero inflation would be assumed to come only from the common N_0.
The probability function of (N, M) is given by

$$p(n, m) = \sum_{k=0}^{\min(n,m)} \mathbb{P}[N_0 = k]\mathbb{P}[N_1 = n - k]\mathbb{P}[N_2 = m - k]$$

Due to the form of the probability generating function of this bivariate model :

$$\psi_{N,M}(u, v) = (p_0 + (1 - p_0)e^{-\lambda_0(1-uv)})(p_1 + (1 - p_1)e^{-\lambda_1(1-u)})(p_2 + (1 - p_2)e^{-\lambda_2(1-v)})$$

it does not seem to be possible to find a recursion in order to find the probability function easily.
Once again, as an extension of the result in the univariate case, it seems that we have the following maximum likelihood equations :

$$\begin{aligned}
(1 - \hat{p}_0)\hat{\lambda}_0 + (1 - \hat{p}_1)\hat{\lambda}_1 &= \bar{n} \\
(1 - \hat{p}_0)\hat{\lambda}_0 + (1 - \hat{p}_2)\hat{\lambda}_2 &= \bar{m}
\end{aligned}$$

Let us fit our reference data set :

N/M		0	1	2	3	4	5	6	7		
Obs	0	21	18	8	2	1	0	0	0	l	-343.69
Fitted		20.17	15.62	9.37	3.75	1.12	0.27	0.05	0.01	χ^2	3.269
Obs	1	13	14	10	1	4	1	0	0	df	7
Fitted		10.84	12.82	8.46	4.07	1.43	0.39	0.09	0.02	$p-value$	0.859
Obs	2	4	5	4	2	1	0	1	0	p_0	0.3729
Fitted		5.22	6.42	5.19	2.79	1.16	0.37	0.10	0.02	λ_0	0.4173
Obs	3	2	1	3	2	0	1	0	0	p_1	0.2321
Fitted		1.68	2.44	2.16	1.36	0.64	0.24	0.07	0.02	λ_1	0.9632
Obs	4	0	0	1	1	0	0	0	0	p_2	0.1419
Fitted		0.40	0.68	0.71	0.50	0.27	0.11	0.04	0.01	λ_2	1.1994
Obs	5	0	0	0	0	0	0	0	0	AIC	-349.69
Fitted		0.08	0.15	0.18	0.15	0.09	0.04	0.02	0.01	BIC	-358.10
Obs	6	0	0	0	0	0	0	0	0		
Fitted		0.01	0.03	0.04	0.04	0.02	0.01	0.01	0.00		
Obs	7	0	1	0	0	0	0	0	0		
Fitted		0.00	0.00	0.01	0.01	0.01	0.00	0.00	0.00		

Table 6.7: Bivariate ZIP fit

This model gives an excellent fit.

Chapter 7

Bonus-malus systems

This chapter deals with the construction of optimal bonus-malus systems, derived from the Mixed Poisson models we studied in the previous chapters. Moreover we will study how to construct practical bonus-malus systems and we will take into account the hunger for bonus that is induced on the market because of the bonus-malus system.

7.1 Univariate optimal bonus-malus systems

Sections 7.1.1 and 7.1.2 are based on Walhin and Paris (1999b).

7.1.1 The expected value principle

The bonus-malus system depends only on the number of accidents caused by the insured in the past. In our Mixed Poisson model, it is easy to see that it is sufficient to consider the total number of accidents without reference to the history of the accidents :

$$\begin{aligned}
& dU(\lambda|N(t)-N(t-1)=k_t,\ldots,N(1)-N(0)=k_1) \\
= \quad & \frac{\mathbb{P}[N(t)-N(t-1)=k_t,\ldots,N(1)-N(0)=k_1|\lambda]dU(\lambda)}{\mathbb{P}[N(t)-N(t-1)=k_t,\ldots,N(1)-N(0)=k_1]} \\
= \quad & \frac{\frac{e^{-\lambda t}\lambda^k}{\prod_{j=1}^t k_j!}dU(\lambda)}{\int_0^\infty \frac{e^{-\lambda t}\lambda^k}{\prod_{j=1}^t k_j!}dU(\lambda)}
\end{aligned}$$

where $k=\sum_{j=1}^t k_j$.

The premium for the first year is an a priori premium because there is no information concerning the risk :

$$\mathbb{E}N(1)=\mathbb{E}\Lambda$$

For the t^{th} year, as the history of the accidents is unimportant, we take into account the information about the number of accidents during the first t years and the premium is

$$\begin{aligned}\mathbb{E}[N(t+1)-N(t)|N(t)=k] &= \mathbb{E}(\Lambda|N(t)=k)\\ &= \frac{k+1}{t}\frac{\Pi(k+1,t)}{\Pi(k,t)}\end{aligned}$$

This expression is general. It reduces to

$$\frac{p+kc}{1+ct} = p\frac{1}{1+ct} + \frac{k}{t}\frac{ct}{1+ct}$$

in the Negative Binomial case and is by far easier to use than the formulae derived by Tremblay (1992) for the particular case of the Poisson Inverse Gaussian Distribution.
Assuming that the first premium paid is 100, we can construct a bonus-malus table depending on k and t with the formula :

$$P_{k,t} = \frac{100}{\mathbb{E}\Lambda}\frac{k+1}{t}\frac{\Pi(k+1,t)}{\Pi(k,t)} \tag{7.1}$$

The following good properties justify the "optimal bonus-malus" denomination :

1. The system is financially balanced each year :

$$\sum_{k=0}^{\infty}\Pi(k,t)\mathbb{E}(\Lambda|N(t)=k) = \mathbb{E}\Lambda \quad \forall t$$

2. The more accidents you cause , the higher the premium :

$$\mathbb{E}[\Lambda|N(t)=k+1] > \mathbb{E}[\Lambda|N(t)=k] \quad \forall t\,,k$$

3. The premium always decreases when no more accidents are caused :

$$\frac{\partial}{\partial t}\mathbb{E}[\Lambda|N(t)=k] \leq 0 \quad \forall t\,,k$$

Property 1 is just an application of the iterative property of the expectation.
Properties 2 and 3 are easily shown by using the Cauchy-Schwartz inequality.
With the Hofmann fit of the reference portfolio, we find :

t/k	0	1	2	3	4
1	87	162	279	424	582
2	79	138	229	342	465
3	73	122	195	287	389
4	68	110	172	249	334
5	64	100	154	220	294
6	60	93	139	197	262
7	58	87	128	180	237
8	55	82	119	165	217
9	53	77	111	153	200
10	51	73	104	142	186
20	39	52	68	88	111
50	27	33	39	47	56
100	20	23	26	30	34

Table 7.1: Bonus-malus table with the Hofmann fit

The asymptotic behaviour of the bonus-malus premium is

$$\lim_{t\to\infty} P_{k,t} = 0 \quad \forall k$$

¿From a theoretical point of view it is not acceptable that a driver pays a premium equal to zero because there is always a positive probability that he (she) will cause a claim. Therefore we introduce the model $Ho+Po$ (see section 5.1.9) where a homogeneous part of the premium implies that the premium will never be zero. Obviously this discussion is essentially theoretical because for classical frequencies, the premiums are acceptable within the range of the driving ages.
The asymptotic behaviour of the bonus-malus premium for the $Ho+Po$ fit is given by

$$\lim_{t\to\infty} P_{k,t} = 100\frac{\delta}{\delta+p} \quad \forall k$$

The next table gives the bonus-malus premiums for the $Ho+Po$ fit :

t/k	0	1	2	3	4
1	87	162	284	421	556
2	79	136	234	347	460
3	73	119	199	295	392
4	68	106	173	256	342
5	65	97	153	226	302
6	62	89	138	202	271
7	60	83	125	183	245
8	58	78	115	167	224
9	56	74	107	153	206
10	54	71	100	142	190
20	47	53	65	83	107
50	40	41	44	47	53
100	37	38	38	39	40

Table 7.2: Bonus-malus table with the Ho + Po fit

The next table is constructed with the nonparametric fit :

$$P_{k,t} = \frac{100}{\mathbb{E}\Lambda} \frac{\sum_{j=1}^{m} p_j e^{-\lambda_j t} \lambda_j^{k+1}}{\sum_{j=1}^{m} p_j e^{-\lambda_j t} \lambda_j^{k}}$$

t/k	0	1	2	3	4
1	87	162	280	440	554
2	79	138	222	359	505
3	72	124	185	283	435
4	67	114	163	228	354
5	63	106	150	193	281
6	59	99	142	173	227
7	55	93	136	162	194
8	52	87	131	156	176
9	50	81	126	152	166
10	47	76	120	148	161
20	37	44	66	109	142
50	35	35	35	36	38
100	35	35	35	35	35

Table 7.3: Bonus-malus table with the NP fit

This table differs greatly from the parametric case table.
In fact the form of the bonus-malus table reflects the discontinuity of Λ. The nonparametric fit shows that there are three classes of risks : those with $\lambda = 0.05461; \lambda = 0.24599$ and $\lambda = 0.95618$. We find those three classes in the bonus-malus table ; locally, the table has

the same behaviour as a table constructed with a simple Poisson process : the premiums are almost indistinguishable because of the lack of heterogeneity. We have

$$\begin{aligned} \frac{100}{\mathbb{E}\Lambda}\lambda_3 &= 616.33 \\ \frac{100}{\mathbb{E}\Lambda}\lambda_2 &= 158.56 \\ \frac{100}{\mathbb{E}\Lambda}\lambda_1 &= 35.20 \end{aligned}$$

The somewhat curious behaviour of our table and the three classes are very detectable for very bad risks. The following graph shows the evolution of the premium for $k = 15$.

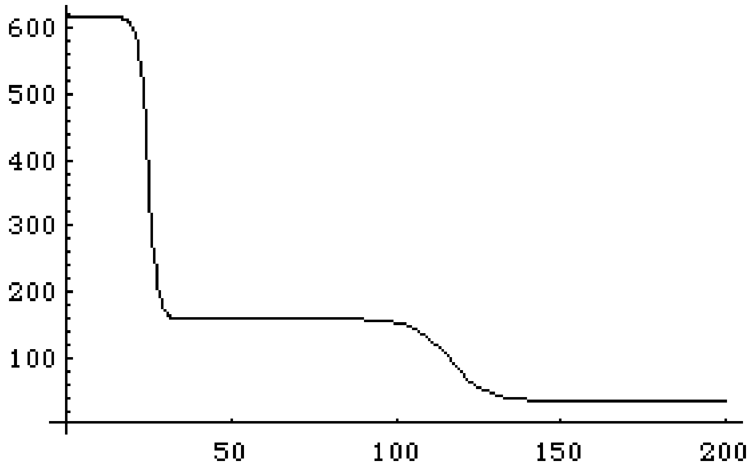

Figure 7.1: Optimal bonus-malus premium with nonparametric fit ; $k = 15$

Obviously, because of that, the concavity of the premiums changes.
Such a curious behaviour of the bonus-malus table seems very difficult to apply and so the nonparametric fit should not be used for the construction of bonus-malus tables.
The surprising results of the nonparametric method are due to the fact that the estimation of the distribution function U of the random variable Λ is only based on the observation of $N(1)$. Even if the period of observation is longer, the trouble will remain because as the frequency is low, the number of points of increase of U is always low and the number of classes of risks is low.
The asymptotic behaviour of the bonus-malus table is given by :
If $\lambda_1 > 0$:

$$\lim_{t\to\infty} P_{k,t} = \min_{\lambda_j} \lambda_j \frac{100}{\mathbb{E}\Lambda}$$

If $\lambda_1 = 0$:

$$\begin{aligned} \lim_{t\to\infty} P_{k,t} &= \min_{\lambda_j>0} \lambda_j \frac{100}{\mathbb{E}\Lambda} \quad \text{if } k > 0 \\ &= 0 \quad \text{if } k = 0 \end{aligned}$$

7.1.2 The zero utility principle with exponential utility function

This section is mainly based on Walhin and Paris (1999b).
Following Lemaire (1985) and Tremblay (1992), we can construct a bonus-malus system for charged premiums using an exponential utility function

$$\mu(x) = \frac{1}{\gamma}(1 - e^{-\gamma x}) \quad , \quad \gamma > 0$$

with the principle of zero utility.
Using the formulae (5.3) and (5.11) the a priori premium becomes (Gerber (1979)) :

$$\begin{aligned} P &= \frac{1}{\gamma} \ln \mathbb{E}[e^{\gamma N(1)}] \\ &= \frac{1}{\gamma} \ln \mathbb{E}[e^{w\Lambda}] \quad \text{where } w = e^{\gamma} - 1 \end{aligned}$$

The a posteriori premium is given in the same way as in the previous section :

$$\begin{aligned} P &= \frac{1}{\gamma} \ln \mathbb{E}[e^{w\Lambda} | N(t) = k] \\ &= \frac{1}{\gamma} \ln \left\{ \frac{1}{\Pi(k,t)} \int_0^\infty e^{w\lambda} e^{-\lambda t} \frac{(\lambda t)^k}{k!} dU(\lambda) \right\} \\ &= \frac{1}{\gamma} \ln \left\{ \left(\frac{t}{t-w} \right)^k \frac{\Pi(k, t-w)}{\Pi(k,t)} \right\} \end{aligned}$$

By normalizing such that the first premium is 100, the bonus-malus table is constructed with the formula

$$100 \, \frac{\ln \left\{ \left(\frac{t}{t-w} \right)^k \frac{\Pi(k,t-w)}{\Pi(k,w)} \right\}}{\ln\{\Pi(0,-w)\}} \qquad \text{with } w = e^{\gamma} - 1 \tag{7.2}$$

Once again, this formula is more general and more simple than in Tremblay (1992).
We find the following premium tables

t/k	0	1	2	3	4
1	87	163	282	430	590
2	78	138	229	343	468
3	72	121	195	287	389
4	67	109	171	248	333
5	63	99	152	218	292
6	59	92	138	196	260
7	57	86	127	178	235
8	54	80	117	163	215
9	52	76	110	151	197
10	50	72	103	141	183
20	39	51	67	87	109
50	27	32	38	46	54
100	20	22	26	29	33

Table 7.4: Bonus-malus table for loaded premium with $\gamma = 0.25$

This table is comparable with table 7.1. We see in the next table that, even for unreasonable values of γ, the difference with table 7.1 is small.

t/k	0	1	2	3	4
1	82	165	298	462	638
2	72	133	229	347	476
3	65	114	188	280	382
4	60	100	161	236	319
5	56	91	141	205	275
6	53	83	127	181	242
7	50	77	115	163	216
8	48	72	106	148	196
9	46	68	98	136	179
10	44	64	92	126	165
20	34	44	59	76	96
50	23	27	33	40	47
100	17	19	22	25	28

Table 7.5: Bonus-malus table for loaded premium with $\gamma = 1$

In the nonparametric case the asymptotic behaviour is :
If $\lambda_1 > 0$:

$$\lim_{t\to\infty} P_{(k,t)} = \min_{\lambda_j} \lambda_j \frac{w}{\ln \Pi(0,-w)}$$

If $\lambda_1 = 0$:

$$\begin{aligned} \lim_{t\to\infty} P_{(k,t)} &= \min_{\lambda_j>0} \lambda_j \frac{w}{\ln \Pi(0,-w)} \qquad \text{if } k > 0 \\ &= 0 \quad \text{if } k = 0 \end{aligned}$$

t/k	0	1	2	3	4
1	87	162	284	440	546
2	78	138	224	362	501
3	72	123	185	287	436
4	66	113	162	229	357
5	62	105	148	192	284
6	58	98	140	171	229
7	54	92	133	159	194
8	51	86	128	153	174
9	49	80	123	149	163
10	47	75	118	145	158
20	36	43	65	107	140
50	34	34	35	35	37
100	34	34	34	34	34

Table 7.6: Bonus-malus table for loaded premium with $\gamma = 0.25$ and the NP fit

7.1.3 Exponential loss functions

This section is mainly based on Walhin and Paris (2000g).
The construction of a bonus-malus system with an exponential loss function was described in Lemaire (1979) and used by Morillo et Bermudez (1999).
The motivation for introducing such a bonus-malus system is the following. The classical construction is based on the expected value principle :

$$\lambda_{t+1}(k) = \mathbb{E}(\Lambda|N(t) = k)$$

which is in fact the conditional intensity of the process.
The latter comes from the minimization of the quadratic loss function

$$\int_0^\infty (\lambda - \lambda_{t+1}(k))^2 dU(\lambda|N(t) = k)$$

As noted in Lemaire (1979), this method treats the maluses and bonuses symmetrically for the policyholders having reported k claims in t years. If the Insurance Company wishes to maintain a high level of solidarity among its policyholders, it needs to use a method weighting the maluses and bonuses differently.
Let us choose an exponential loss function :

$$\int_0^\infty \frac{1}{\gamma}(e^{-\gamma(\lambda - \lambda_{t+1}(k))} - 1) dU(\lambda|N(t) = k)$$

Such a loss function gives us a tool in order to index our wishes about the bonuses and maluses to use. For example if $\gamma = 0.25$, two policyholders with an under-tarification of 0.02 compensate one policyholder with an over-tarification of 0.04 whereas with $\gamma = 25$, 4 policyholders with an under-tarification of 0.02 are needed in order to compensate one policyholder with an over-tarification of 0.04. The latter situation thus represents a larger solidarity among the policyholders.

We can give a solution to the minimization program

$$\min_{\lambda_{t+1}(k)} \int_0^\infty \frac{1}{\gamma}(e^{-\gamma(\lambda-\lambda_{t+1}(k))} - 1)dU(\lambda|N(t) = k)$$

only if $\lambda_{t+1}(k)$ is constrained.
The natural constraint gives the financial equilibrium of the system :

$$\mathbb{E}\lambda_{t+1}(N(t)) = \mathbb{E}\Lambda$$

The solution to this minimization program with constraint is given by Lemaire (1979) thanks to the Lagrangian and with a more intuitive proof by Morillo et Bermudez (1999) thanks to the Jensen's inequality.
The solution writes :

$$\lambda_{t+1}(k) = \mathbb{E}\Lambda + \frac{1}{\gamma}\left[\mathbb{E}\ln\mathbb{E}(e^{-\gamma\Lambda}|N(t)) - \ln\mathbb{E}(e^{-\gamma\Lambda}|N(t) = k)\right]$$

Some algebra gives

$$\lambda_{t+1}(k) = \mathbb{E}\Lambda + \frac{1}{\gamma}\left[\mathbb{E}\ln\left\{(\frac{t}{t+\gamma})^{N(t)}\frac{\Pi(N(t), t+\gamma)}{\Pi(N(t), t)}\right\} - \ln\left\{(\frac{t}{t+\gamma})^{k}\frac{\Pi(k, t+\gamma)}{\Pi(k, t)}\right\}\right]$$

The premium will be a percentage, in reference to the a priori premium, $\mathbb{E}\Lambda$. The percentage is :

$$P_{t+1}(k) = 100 + \frac{100}{\mathbb{E}\Lambda}\frac{1}{\gamma}\left[\mathbb{E}\ln\left\{(\frac{t}{t+\gamma})^{N(t)}\frac{\Pi(N(t), t+\gamma)}{\Pi(N(t), t)}\right\} - \ln\left\{(\frac{t}{t+\gamma})^{k}\frac{\Pi(k, t+\gamma)}{\Pi(k, t)}\right\}\right]$$

	Ho fit					Ho+Po fit				
t/k	0	1	2	3	4	0	1	2	3	4
1	92	140	212	301	400	92	137	214	307	401
2	86	126	186	261	344	86	123	187	267	350
3	80	116	168	232	303	82	112	166	237	311
4	76	108	153	209	272	78	104	150	213	279
5	72	101	141	191	246	74	97	137	193	253
6	69	95	131	176	226	71	91	127	176	232
7	66	90	123	163	209	69	86	118	163	214
8	63	85	115	152	194	67	82	110	151	198
9	61	81	109	143	182	65	79	104	141	184
10	59	78	104	135	171	63	76	98	132	172
20	46	57	72	90	110	53	59	68	84	105
50	31	36	42	49	58	43	45	47	50	55
100	22	25	28	32	36	39	40	40	41	42

Table 7.7: Bonus-malus table with low solidarity $\gamma = 5$

	Ho fit					Ho + Po fit				
t/k	0	1	2	3	4	0	1	2	3	4
1	95	125	169	224	285	96	121	165	223	285
2	91	118	156	204	257	92	113	151	203	259
3	87	111	145	188	235	89	107	140	186	237
4	83	106	137	175	218	86	102	131	172	219
5	80	101	129	164	203	83	98	123	160	204
6	77	97	122	154	190	81	94	117	150	190
7	75	93	117	146	180	79	90	111	142	179
8	73	89	112	139	170	77	87	106	134	169
9	70	86	107	133	162	75	85	101	127	159
10	68	83	103	127	154	73	82	98	121	151
20	54	64	76	91	108	62	67	74	85	102
50	36	41	47	54	61	49	51	53	55	59
100	26	28	31	35	39	43	43	44	45	46

Table 7.8: Bonus-malus table with high solidarity $\gamma = 15$

7.2 Bivariate optimal bonus-malus systems

This section is based on Walhin and Paris (2000d).
We first treat the case of the Mixed Independent Bivariate Poisson model.
In this section we construct a bonus-malus system depending on two types of claims. For example we will use the data set with bodily injury and material damage claims.

Let us remind that the pure premium the drivers are asked is

$$\mathbb{E}N\mathbb{E}C \tag{7.3}$$

where N is the number of claims and C is the cost of claims.
In an a posteriori rating we have

$$\mathbb{E}(N(t+1) - N(t)|N(t))\mathbb{E}C \tag{7.4}$$

Our aim is to introduce the aspect of bodily injury and material damage claims in order to give more liability to drivers with bodily injury claims.
(7.3) becomes

$$\mathbb{E}N\mathbb{E}MD + \mathbb{E}M\mathbb{E}BI$$

where N and M are the number of material damage and bodily injury claims respectively and MD and BI are the respective costs.
(7.4) becomes

$$\mathbb{E}(N(t+1) - N(t)|N(t), M(t))\mathbb{E}MD + \mathbb{E}(M(t+1) - M(t)|N(t), M(t))\mathbb{E}BI$$

Unfortunately we come to

$$\mathbb{E}(N(t+1) - N(t)|N(t) = n, M(t) = m) \quad = \quad \frac{\int_0^\infty \lambda e^{-\lambda t(1+\beta)}(\lambda t)^{n+m} dU(\lambda)}{\int_0^\infty e^{-\lambda t(1+\beta)}(\lambda t)^{n+m} dU(\lambda)}$$

$$\mathbb{E}(M(t+1)-M(t)|N(t)=n, M(t)=m) \quad = \quad \beta\frac{\int_0^\infty \lambda e^{-\lambda t(1+\beta)}(\lambda t)^{n+m}dU(\lambda)}{\int_0^\infty e^{-\lambda t(1+\beta)}(\lambda t)^{n+m}dU(\lambda)}$$

i.e. these quantities depend only of $n+m$. So the model is not able to catch the difference between two kinds of claims in an a posteriori rating.
We will use univariate formulae :

$$\mathbb{E}(N(t+1)-N(t)|N(t))\mathbb{E}MD + \mathbb{E}(M(t+1)-M(t)|M(t))\mathbb{E}BI \tag{7.5}$$

In order to work with percentage premiums, we need to simplify the expression (7.5). Let us assume that the rapport between $\mathbb{E}BI$ and $\mathbb{E}MD$ is α :

$$\alpha = \frac{\mathbb{E}BI}{\mathbb{E}MD}$$

Then we have

$$[\mathbb{E}(N(t+1)-N(t)|N(t)=n) + \alpha\mathbb{E}(M(t+1)-M(t)|M(t)=m)]\,\mathbb{E}MD \tag{7.6}$$

Percentage premiums in function of $\mathbb{E}MD$ will be given by

$$100\frac{\mathbb{E}(N(t+1)-N(t)|N(t)=n) + \alpha\mathbb{E}(M(t+1)-M(t)|M(t)=m)}{\mathbb{E}N(1) + \alpha\mathbb{E}N(1)\beta}$$

We have applied this formula with the fit of our reference data set and with an $\alpha = 15$. Obviously our data set is not concerned with two types of claims but we consider they are in order to have a numerical application.
We find

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$n=0$	77	99	120	141	163
$n=1$	79	100	122	143	164
$n=2$	80	102	123	144	165
$n=3$	82	103	124	146	167
$n=4$	83	104	126	147	168

Table 7.9: BM table $t=1$

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$n=0$	46	59	72	84	97
$n=1$	47	60	72	85	98
$n=2$	48	60	73	86	99
$n=3$	48	61	74	87	100
$n=4$	49	62	75	88	100

Table 7.10: BM table $t=9$

Note that the bonus malus tables hereabove have the good property to be financially balanced.

We now treat the case of the Mixed Bivariate Poisson model.
Unfortunately numerical examples show that the premium always depends only on $n+m$ in this case as well. However it does not seem easy to prove. We have

$$\begin{aligned}
\mathbb{E}[N(t+1)-N(t)|N(t)=n, M(t)=m] &= \\
\frac{\int_0^\infty \sum_{k=0}^{\min(n,m)} \lambda(\lambda_0+\lambda_1) e^{-\lambda\lambda_0 t}\frac{(\lambda\lambda_0 t)^k}{k!} e^{-\lambda\lambda_1 t}\frac{(\lambda\lambda_1 t)^{n-k}}{(n-k)!} e^{-\lambda\lambda_2 t}\frac{(\lambda\lambda_2 t)^{m-k}}{(m-k)!}}{\mathbb{P}[N(t)=n, M(t)=m]} &= \\
(\lambda_0+\lambda_1)\frac{\sum_{k=0}^{\min(n,m)} \frac{(n+m-k+1)!}{k!(n-k)!(m-k)!}\frac{1}{t}\frac{\lambda_0^k \lambda_1^{n-k}\lambda_2^{m-k}}{(\lambda_0+\lambda_1+\lambda_2)^{n+m-k+1}} \Pi(n+m-k+1, t(\lambda_0+\lambda_1+\lambda_2))}{\mathbb{P}[N(t)=n, M(t)=m]}
\end{aligned}$$

We find the following bonus-malus tables :

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$n=0$	63	79	95	111	127
$n=1$	79	93	108	124	140
$n=2$	95	108	123	139	154
$n=3$	111	124	139	154	169
$n=4$	127	140	154	169	184

Table 7.11: BM table $t=1$

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$n=0$	32	39	46	54	61
$n=1$	39	46	53	61	68
$n=2$	46	53	60	68	75
$n=3$	54	61	68	75	82
$n=4$	61	68	75	82	90

Table 7.12: BM table $t=1$

These tables show that the model is not able to make the distinction between the two types of claims.
We apply once again the following formula with $\alpha=15$:

$$100\frac{\mathbb{E}(N(t+1)-N(t)|N(t)=n)+\alpha\mathbb{E}(M(t+1)-M(t)|M(t)=m)}{\mathbb{E}N(1)+\alpha\mathbb{E}N(1)\beta}$$

We find

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$n=0$	80	99	118	138	158
$n=1$	81	100	119	139	159
$n=2$	82	101	121	140	160
$n=3$	84	102	122	142	162
$n=4$	85	104	123	143	163

Table 7.13: BM table $t=1$

	$m=0$	$m=1$	$m=2$	$m=3$	$m=4$
$n=0$	50	62	73	85	97
$n=1$	51	62	74	86	98
$n=2$	52	63	75	86	98
$n=3$	53	64	75	87	99
$n=4$	53	65	76	88	100

Table 7.14: BM table $t=4$

7.3 Univariate optimal bonus-malus system with changing frequency

This section is mainly based on Walhin and Paris (2000d).
In this section we extend the hypothesis of section 6.1.1 in order to construct bonus-malus systems allowing a mean frequency to vary geometrically from a year to another.
Using the same notations as for introducing the Hofmann Distribution, we introduce a new process :

$$\mathbb{P}(N(t)=n)=\Pi(n,\frac{1-\beta^t}{1-\beta})$$

This distribution obeys the hypotheses given in Besson and Partrat (1992) :

- $N_1-N_0|\Lambda,\cdots,N_{t+1}-N_t|\Lambda$ are independent
- $N_i-N_{i-1}\sim Po(\lambda\beta^{i-1})$

Then by section 7.1.2 we get the percentage premium for the expected value premium principle (non loaded premium) :

$$\frac{100}{p}\frac{k+1}{t}\frac{\Pi(k+1,\frac{1-\beta^t}{1-\beta})}{\Pi(k,\frac{1-\beta^t}{1-\beta})}$$

and for the exponential utility premium principle (loaded premium) :

$$100\frac{\ln\left(\frac{t}{t-w}\right)^k\frac{\Pi(k,\frac{1-\beta^{t-w}}{1-\beta})}{\Pi(k,\frac{1-\beta^t}{1-\beta})}}{\ln\Pi(0,\frac{1-\beta^{-w}}{1-\beta})}$$

with security loading w.
With our reference portfolio, a bonus-malus table under the hypothesis that the frequency diminishes of 5% per year is given by

t/k	0	1	2	3	4
1	87	162	279	424	582
2	79	138	227	337	459
3	73	121	193	282	380
4	69	110	169	243	325
5	65	101	151	214	285
6	63	94	138	192	254
7	60	88	127	175	229
8	58	84	118	161	210
9	57	80	111	150	193
10	55	76	105	140	180
20	47	59	73	91	111
50	41	46	51	58	65
100	40	42	45	48	51

Table 7.15: BM table with yearly decreasing frequency at rate 5%

We note that the bonus for absence of claims is slower than in the case of a stable frequency. This is coherent with our intuition. Obviously a more rapidly changing frequency will have a more dramatic effect.
Let us analyze a frequency increasing of 5% each year :

t/k	0	1	2	3	4
1	87	162	279	424	582
2	79	139	231	346	472
3	72	122	198	293	397
4	66	110	174	254	343
5	62	100	156	225	302
6	58	92	141	202	271
7	55	85	129	184	245
8	52	80	119	169	224
9	49	75	111	156	206
10	47	71	104	145	191
20	32	46	64	86	111
50	15	20	28	37	48
100	5	8	12	17	23

Table 7.16: BM table with yearly increasing frequency at rate 5%

7.4 Bonus-malus system with finite number of classes

This section is mainly based on Walhin and Paris (2000f).
The optimal bonus-malus tables described in section 7.1 seem difficult to apply because of the infinite number of classes in both directions k and t. Such a system will certainly be

too complicated for the policyholders. Therefore most European countries use bonus-malus systems with a finite number of classes.

Basically, the bonus-malus systems with finite number of classes (s) have the Markov property (if not, a redefinition of the system may be needed to show the Markov property).

s	0	1	2	3	$\dots$	$s-1$
C_s	C_0	C_1	C_2	C_3	$\dots$	C_{s-1}

Table 7.17: Premium levels

Transition rules are given indicating how the drivers move inside the bonus-malus system according to their annual number of reported claims.
If we assume that the number of classes as well as the transition rules were chosen, the premium levels need to be calculated. This is explained in Coene and Doray (1996) where they minimize a certain distance between the C_i and the $P_{(k,t)}$ in order to determine the C_i.
We will use the same methodology as Coene and Doray (1996) with some amendments.
Firstly, it is necessary to construct a table of $C_{(k,t)}$ parallel to the table of $P_{(k,t)}$. $C_{(k,t)}$ gives the different possible values for C_i in function of k and t.

Let us assume a bonus-malus system with the following characteristics :

- $s = 9$: 9 classes numbered $0, 1, \dots, 8$. 0 is the minimum class. 8 is the maximum class.
- Entry of the system is in class 4.
- In the case of a claims free year, the policyholder comes down one class
- In the case of claim(s) being reported, the policyholder goes to level 8 whatever the number of claims.

We have

t/k	0	1	2	$\dots$
0	C_4	$-$	$-$	$\dots$
1	C_3	C_8	C_8	$\dots$
2	C_2	C_8/C_7	C_8/C_7	$\dots$
3	C_1	$C_8/C_7/C_6$	$C_8/C_7/C_6$	$\dots$
4	C_0	$C_8/C_7/C_6/C_5$	$C_8/C_7/C_6/C_5$	$\dots$
5	C_0	$C_8/C_7/C_6/C_5/C_4$	$C_8/C_7/C_6/C_5/C_4$	$\dots$
$\vdots$				

Table 7.18: $C_{(k,t)}$

Coene and Doray (1996) suggest to choose the maximum class for $C_{(k,t)}$ because it is often the most probable class.
However it is not difficult to find the probabilities of the following events : $[N(t) = k, C_{(k,t)} = C_i]$. Let us define

$$w(k,t,i) = \mathbb{P}[N(t) = k, C_{(k,t)} = C_i]$$

The following general formula is valid for a Mixed Poisson Process :

$$\mathbb{P}[N(n) - N(n-1) = k_n, \dots, N(1) - N(0) = k_1] = \frac{\Pi(\sum_{i=1}^n k_i, n)}{n^{\sum_{i=1}^n k_i}} \frac{(\sum_{i=1}^n k_i)!}{k_1! \dots k_n!}$$

Therefore it is always possible to write the probabilities $w(k,t,i)$ as $\alpha \Pi(k,t)$ where α depends on the different ways to reach C_i at time t with k claims.
For example

$$\begin{aligned}
\mathbb{P}[C_{(1,1)} = C_8 | N(1) = 1] &= 1 \\
\mathbb{P}[C_{(1,1)} = C_i | N(1) = 1] &= 0 \quad \forall i < 8 \\
\mathbb{P}[C_{(1,2)} = C_8 | N(2) = 1] &= \frac{\mathbb{P}[N(1) - N(0) = 0, N(2) - N(1) = 1]}{\mathbb{P}[N(2) = 1]} = \frac{1}{2} \\
\mathbb{P}[C_{(1,2)} = C_7 | N(2) = 1] &= \frac{\mathbb{P}[N(1) - N(0) = 1, N(2) - N(1) = 0]}{\mathbb{P}[N(2) = 1]} = \frac{1}{2} \\
\mathbb{P}[C_{(1,2)} = C_i | N(2) = 1] &= 0 \quad \forall i < 7
\end{aligned}$$

The most obvious way of calculating the C_i is to minimize the following quadratic error with natural weights :

$$\sum_{(k,t,i)} w(k,t,i) [P_{(k,t)} - C_i]^2$$

This minimization procedure is quite similar to the one derived in Coene and Doray (1996) but it is more natural because we use the exact values of $C_{(k,t)}$ and we do not use the stationary distribution of the drivers in the bonus-malus system.
As mentionned in Coene and Doray (1996) we now have a problem of optimization with s variables. Obviously some constraints have to be taken into account :

- The entry class in the bonus-malus system is generally at level 100. So in our example we have to constrain $C_4 = 100$
- Corresponding to the natural properties

$$\begin{array}{rcl} \frac{\partial}{\partial t}P_{(k,t)} & \leq & 0 \quad \forall k \\ P_{(k+1,t)} & \geq & P_{(k,t)} \quad \forall t \end{array}$$

  we have the natural constraint : $C_i \leq C_{i+1} \quad \forall i$

- As we are interested in having integer percentages C_i, it may be interesting to constrain the C_i to being integer, i.e. to make optimization with integers.
- As a consequence of the financial equilibrium of the optimal bonus-malus system

$$\sum_{k=0}^{\infty} \Pi(k,t)P_{(k,t)} = 100 \quad \forall t$$

  the following constraints may be imposed

$$\begin{array}{rcll} \sum_{i=0}^{s-1} e_\infty(i)C_i & \geq & 100 & \\ \sum_{i=0}^{s-1} e_t(i,\mathbf{e_0})C_i & \geq & 100 & \text{for some } t \text{ with } \mathbf{e_0} \text{ given} \end{array}$$

  where $e_\infty(i)$ denotes the i^{th} component of the stationary distribution of the drivers in the bonus-malus system ($\mathbf{e}_\infty$) and $e_t(i,\mathbf{e_0})$ denotes the i^{th} component of the transient distribution at time t of the drivers in the bonus-malus system ($\mathbf{e_t}(\mathbf{e_0})$) with initial distribution $\mathbf{e_0}$.

Numerical examples will be given in section 7.5.3.

Note that other techniques are described in Norberg(1976), Gilde and Sundt (1989), Borgan et al. (1981).

7.5 The actual claim amount and frequency distributions within a bonus malus system

This section is taken from Walhin and Paris (2000a).

7.5.1 Introduction

When an insurer uses a bonus-malus system independent of the claim amounts, he notes a propensity of the insured not to declare the small claims. Indeed it is in some cases more interesting for the insured to defray himself the Third Party than to declare the claim and to see the malus degree increase. Lemaire (1977) called this fact the hunger for bonus. See also Lemaire (1995).

The hunger for bonus induces that the introduction of a (new) bonus-malus system creates a censored view of the claim amount and frequency distributions. Indeed some of the lowest claim amounts will not be declared to the Companies. Obviously, for the policyholder, the natural question is : "under what level of claim amount is it interesting for me to defray the cost myself ?"

Lemaire (1977) answered this question using an algorithm related to dynamic programming. In this section, we use Lemaire's (1977) algorithm and the Nonparametric Mixed Poisson fit of an automobile portfolio in order to refind the true claim amount and frequency distributions. This problem was already implicitely posed by Lemaire (1977) in his paper where he states that he has to use old claim amount data because the recent ones are influenced by the introduction of the bonus-malus system.

Throughout this section we will mention the following example :
The bonus-malus system has 9 classes (from 0 to 8). A new driver enters the system in class 4. If there is no claim in a year, he (or she) falls one class (with minimum class 0). He (or she) climbs up 3 classes (with maximum class 8) per claim during the year.

s	0	1	2	3	4	5	6	7	8
C_s	75	80	90	95	100	150	170	185	250

Table 7.19: Percentage premiums

The reference (observed) portfolio is the one of chapter 5 (table 5.1).
For the claim amounts we use the following hypothetical data set :

6	6	10	11	17	18	20	26	27	34
42	44	47	54	59	60	61	61	61	61
64	64	65	66	67	68	71	71	73	75
76	81	85	87	93	94	101	103	105	109
110	110	113	116	116	129	134	134	141	141
151	154	156	159	167	171	172	173	174	179
181	183	185	187	195	195	203	226	235	240
251	255	273	340						

Table 7.20: Observed (and ordered) claim amounts

7.5.2 Lemaire's algorithm and extensions

Lemaire's (1977) algorithm needs the following hypotheses:

Let a bonus-malus system be with s classes : $i = 0, \ldots, s-1$
- the claims frequency of a policyholder be Poisson distributed
- the claim amount distribution be X, with cumulative density function (cdf) $F_X(x)$
- β be the actualisation rate forecast for the future
- P be the total premium, i.e. the base premium at level 100%, including security loading, administration expenses, profit and taxes.
- $1-t$ with $0 \leq t < 1$ be the time remaining until the next premium payment
- m be the number of claims reported to the Company in $[0, t)$

With these hypotheses an iterative algorithm can be performed in order to find the optimal policy of the driver as a function of his bonus-malus level. The optimal policy is simply the optimal retention of the driver as a function of his bonus-malus level. It is the level of claim amount up to which it is interesting for the policyholder to bear the cost himself and not to report the claim to the company. Obviously, the optimal policy is also a function of t, the time at which the claim occurs and m, the number of claims reported before t unless one assumes $t = 0$. Optimal frequencies of the driver are also given by the algorithm.

The solution of the algorithm is shown to be unique if $\beta < 1$, which is always the case if the interest rate is positive.

In short, the algorithm of Lemaire gives the optimal frequency and the optimal retention of a driver based on the actual claim amount and frequency distributions of the driver.

7.5.3 The stationary distribution of the policyholders within a bonus-malus system

In order to perform our calculations we will assume that the bonus-malus has existed for a long time and that it has reached its stationary distribution.

Let $p(x)$ be the probability that a driver with average claims frequency λ causes x claims during a given year. The transition probability matrix (Q) of this driver within the bonus-malus system described in the introduction is thus

s	0	1	2	3	4	5	6	7	8
0	$p(0)$	0	0	$p(1)$	0	0	$p(2)$	0	$1-p(0)-p(1)-p(2)$
1	$p(0)$	0	0	0	$p(1)$	0	0	$p(2)$	$1-p(0)-p(1)-p(2)$
2	0	$p(0)$	0	0	0	$p(1)$	0	0	$1-p(0)-p(1)$
3	0	0	$p(0)$	0	0	0	$p(1)$	0	$1-p(0)-p(1)$
4	0	0	0	$p(0)$	0	0	0	$p(1)$	$1-p(0)-p(1)$
5	0	0	0	0	$p(0)$	0	0	0	$1-p(0)$
6	0	0	0	0	0	$p(0)$	0	0	$1-p(0)$
7	0	0	0	0	0	0	$p(0)$	0	$1-p(0)$
8	0	0	0	0	0	0	0	$p(0)$	$1-p(0)$

Table 7.21: Transition probability matrix

As is the case for each bonus-malus system, we have an irreducible (there are no cycles) Markov chain the states of which are all ergodic (each state can be attained from another state). Under those conditions, there is a stationary probability distribution that is given by :

$$\mathbf{e}_\infty(\lambda) = \lim_{n\to\infty} Q^n \mathbf{e_0}(\lambda)$$

where $\mathbf{e_0}(\lambda)$ denotes any initial distribution of the drivers in the bonus-malus system. The stationary probability distribution is independent of $\mathbf{e_0}(\lambda)$.
The stationary probability distribution is also given by solving

$$\mathbf{e}_\infty(\lambda) = \mathbf{e}_\infty(\lambda) Q$$

with the normalizing condition

$$\sum_{i=0}^{s-1} e_\infty(i;\lambda) = 1$$

where $e_\infty(i;\lambda)$ is the i^{th} component of the vector $\mathbf{e}_\infty(\lambda)$.
If we are interested in the stationary distribution of the portfolio ($\mathbf{e}_\infty$), we only have to take the weighted average of the stationary distributions for the different types of policyholders (see Walhin and Paris (1999b) for details). With our Nonparametric Mixed Poisson fit, we have

$$\mathbf{e}_\infty = \sum_{i=1}^{r} p_i \mathbf{e}_\infty(\lambda_i)$$

For our numerical example we find

	$\lambda = 0.05461$	$\lambda = 0.24600$	$\lambda = 0.95619$	Portfolio
0	0.8278	0.2598	0.0005	0.5728
1	0.0464	0.0724	0.0008	0.0561
2	0.0490	0.0926	0.0022	0.0660
3	0.0518	0.1185	0.0057	0.0783
4	0.0095	0.0876	0.0145	0.0420
5	0.0075	0.0942	0.0369	0.0441
6	0.0052	0.0977	0.0939	0.0457
7	0.0014	0.0880	0.2386	0.0429
8	0.0009	0.0888	0.6066	0.0516

Table 7.22: Stationary distribution of the drivers

Note that if the stationary distribution has not yet been reached, it is not a problem to work with the transient probabilities. The distribution of the drivers within the bonus-malus system after T years is given by

$$e_T(\lambda) = Q^T e_0(\lambda)$$

7.5.4 Formulation of the problem

When collecting data on a market where a bonus-malus system is in use, we do not observe the actual claim amount and frequency distributions. Indeed they are influenced by the hunger for bonus.

The actual claim amount distribution should have a lower mean whereas the actual claims frequency distribution should have a higher mean.

Let us assume that there is a proportion p of the driving population that reports all accidents whereas $(1-p)\%$ only reports the claims exceeding the optimal retention given by Lemaire's algorithm.
Let us assume that a nonparametric fit for the claims frequency distribution has been performed on the observed portfolio (i.e. the reported claims). It reveals r types of risks λ_j with probability p_j $(j = 1, \dots, r)$. This distribution (N) is not the distribution of the number of accidents but the distribution of the number of accidents reported to the Company.
The probability function (pf) of N writes

$$\Pi(k,1) = \sum_{j=1}^{r} p_j e^{-\lambda_j} \frac{\lambda_j^k}{k!} \quad , \quad k = 0, 1, 2, \dots$$

Let $N^{'}$ be the actual distribution of the number of claims. Its pf writes :

$$\Pi'(k,1) = \sum_{j=1}^{r} p_j^{'} e^{-\lambda_j^{'}} \frac{(\lambda_j^{'})^k}{k!} \quad , \quad k = 0, 1, 2, \dots$$

We will assume that $p_j^{'} = p_j \quad \forall j$ i.e. that the proportions of different risks for both distributions are the same.

Let X be the random variable representing the reported claim amounts.
Let Z be the random variable representing the actual claim amounts. This random variable is unknown whereas X is the observed one.
The pf of X is a function of the pf of Z and writes

$$f_X(x) = pf_Z(x) + (1-p)\frac{f_Z(x)}{1-F_Z(c)}\mathbb{I}_{\{x \geq c\}} \quad , \quad x \geq 0$$

where c is the average retention limit of the portfolio.

Our aim is to find the distribution of Z and $N^{'}$.

7.5.5 An algorithm to solve the problem

A solution to the problem described in section 7.5.4 will be given by means of an iterative algorithm using the nonparametric fit of the portfolio and an inversion of the algorithm of Lemaire (1977).

Step 0 : Initializing step

Have an initial guess for the parameter c. Choose a parametric distribution to fit the random variable Z.

Step 1 : Correction of the claim amount distribution

We use the average optimal retention (c) as a censor in order to find a new estimate for the vector of parameters of the distribution of Z. Therefore we maximize the likelihood :

$$L(\theta, p|c) = \prod_{i=1}^{n} \left[pf_Z(x_i;\theta) + (1-p)\frac{f_Z(x_i;\theta)}{1-F_Z(c;\theta)}\mathbb{I}_{\{x_i \geq c\}} \right] \tag{7.7}$$

With the new estimate $(\hat{\theta}, \hat{p}) = argmaxL(\theta, p|c)$ we move to step 2.

Step 2 : Correction of the frequency distribution

In view of the nonparametric fit, there are r types of policyholders. For each of them we repeat the following :
Let $\lambda_j \quad , \quad j = 1, \ldots, r$ be the observed frequency (found by the nonparametric fit of the reference portfolio).
Let $\lambda_j^{'} \quad , \quad j = 1, \ldots, r$ be the actual frequency.
Let $\lambda_j^{''} \quad , \quad j = 1, \ldots, r$ be the optimal frequency given by Lemaire's algorithm.
We then have :

$$\lambda_j = p\lambda_j^{'} + (1-p)\lambda_j^{''} \tag{7.8}$$

λ_j is our observation. $\lambda_j^{'}$ is in fact the entry of the algorithm of Lemaire (1977). $\lambda_j^{''}$ is a by-product of this algorithm.
We apply a trial-error scheme on the entry $\lambda_j^{'}$ in order to match the observation λ_j in connection with equation (7.8).
For this $\lambda_j^{'}$, the algorithm of Lemaire (1977) gives the optimal policy c_j.

The average optimal retention is given by

$$c = \sum_{j=1}^{r} c_j p_j$$

where p_j is the weight associated to λ_j.
With this new average optimal retention c we go back to step 1.

Stopping rule :

We stop the process when convergence occurs. We cannot prove this convergence but in practice it is bound to happen.

Let us note that the following intuitive result is easily shown by maximum likelihood :

$$\hat{p} = \frac{F_n(c)}{F_Z(c)}$$

Proof :
Let $x_{(1)}, x_{(2)}, \ldots, x_{(k)}, c, x_{(k+1)}, \ldots x_{(n)}$ be the order statistic of our observation.
The log-likelihood of (7.7) writes

$$\prod_{i=1}^{k} p f_Z(x_{(i)}) \prod_{i=k+1}^{n} [p f_Z(x_{(i)}) + (1-p) \frac{f_Z(x_{(i)})}{1 - F_Z(c)}]$$

The normal equation for p gives

$$\hat{p} = \frac{\frac{k}{n}}{F_Z(c)}$$

■

Thus clearly the estimate of p, the proportion of insured reporting all accidents will depend on the parametric distribution chosen for Z.

In short the algorithm writes

Step 0 : initialization

Do

Step 1 : correction of the claim amount distribution

Maximize (7.7)

Step 2 : correction of the frequency distribution

For $j = 1$ to r
Do
Try a value for λ_j' and apply Lemaire's algorithm
Until (7.8) is verified
Next j
Find the average optimal retention

Until convergence

7.5.6 Numerical example

For the numerical example, we use the observed data set as well as the bonus-malus system described in section 7.5.1.

In order to use Lemaire's algorithm, we set up the following hypothesis :

1. $\beta = \frac{1}{1+0.06}$
2. $t = \frac{1}{2}$
3. $P = 35$

The total premium may seem very high but in fact it is not. Let us look at the composition of a Third Party Liability premium in Belgium :

- If we work with the observed data sets, the pure premium is

$$\mathbb{E}N \times \mathbb{E}X = 0.155 \times 113.40 = 17.58$$

- Let us assume a 10% fluctuation loading. The premium becomes

$$17.58 \times 1.1 = 19.34$$

- Let us assume that administration expenses are 10%. The premium becomes

$$19.34 \times 1.1 = 21.27$$

- Brokerage is usually 17%. The commercial premium is

$$\frac{21.27}{1 - 17\%} = 25.63$$

- Taxes are 27.10%. The total premium thus amounts to

$$25.63 \times 1.2710 = 32.57$$

This shows that an average total premium of 35 is certainly not exaggerated.

The initializing step of the algorithm is chosen as

- $c = 30$
- Z is exponentially distributed with mean μ :

$$f_Z(x;\mu) = \frac{1}{\mu} e^{-\frac{x}{\mu}} \quad , \quad x \geq 0$$

We will now describe in detail the first iteration of the algorithm.
Step 1 : by maximizing (7.7) we get

$$\begin{aligned} \hat{\mu} &= 97.137 \\ p &= 0.4577 \end{aligned}$$

Step 2 : for $j = 1, 2, 3$ we have to match equation (7.8). We describe in detail the trial-error scheme for $j = 1$.
Let us try an actual claim frequency $\lambda_1^{'} = 0.075$.
The application of Lemaire's algorithm gives the following optimal retentions and frequencies:

s	0	1	2	3	4	5	6	7	8
$m = 0$	18	28	52	76	102	145	119	87	53
$m = 1$	83	108	128	91	54	0	0	0	0
$m = 2$	89	53	0	0	0	0	0	0	0

Table 7.23: Optimal retentions

s	0	1	2	3	4	5	6	7	8
$m = 0$	0.062	0.056	0.043	0.034	0.026	0.016	0.022	0.030	0.043
$m = 1$	0.031	0.024	0.020	0.029	0.043	0.075	0.075	0.075	0.075
$m = 2$	0.029	0.043	0.075	0.075	0.075	0.075	0.075	0.075	0.075

Table 7.24: Optimal frequencies

We assume that the stationary distribution is attained within our bonus-malus system. This stationary distribution was obtained in column 1 of table 7.22. The parameter of the Poisson distribution in the transition probability matrix is of course 0.05461 because the drivers move in the bonus-malus system according to the frequency of reported claims.
Average values for the optimal frequencies and retentions are then easily given by the scalar product between the stationary probability vector and the optimal frequency or retention vector. We find

	Retention limit	Frequency
$m = 0$	25	0.0587
$m = 1$	85	0.0310
$m = 2$	76	0.0362
$m \geq 3$	0	0.075

Table 7.25: Average retention limits and frequencies in function of m

We now want figures that are independent of m. Therefore we look for an average value of the optimal retention and frequency by applying the formulae :

$$\begin{aligned} c_j &= \sum_{m=0}^{\infty} e^{-\lambda_j t} \frac{(\lambda_j t)^m}{m!} c_j(m) \quad j = 1, \ldots, r \\ \lambda_j^{''} &= \sum_{m=0}^{\infty} e^{-\lambda_j t} \frac{(\lambda_j t)^m}{m!} \lambda_j^{''}(m) \quad j = 1, \ldots, r \end{aligned}$$

Retention limit : c_1	Frequency : λ_1''
27	0.0587

Table 7.26: Average retention limit and frequency

Equation (7.8) writes :

$$0.05461 \neq 0.4577\,0.075 + (1 - 0.4577)\,0.0587 = 0.0661$$

We then proceed by trial-error until equation (7.8) is matched. This happens with $\lambda_1' = 0.062$. For $j = 2$ (resp. $j = 3$) we find $\lambda_2' = 0.3392$ (resp. $\lambda_3' = 1.0745$).

The second and subsequent iterations may now be completed. We find

Iteration	c	μ	p	λ_1'	λ_2'	λ_3'
1	30	97.137	0.4577	0.062	0.3392	1.0745
2	47.7386	85.2208	0.4096	0.0637	0.3628	1.1112
3	47.6437	85.3233	0.4105	0.0637	0.3624	1.1107
4	47.6434	85.3239	0.4105	0.0637	0.3624	1.1107

Table 7.27: Iterations until convergence

As we see, convergence occurs. The actual claim amount distribution is then exponentially distributed with mean $\mu = 85.32$. The model shows that 41% of the policyholders report all the claims while 59% use the optimal retention. The actual claims frequency distribution is Nonparametric Mixed Poisson distributed with

$\lambda_1' = 0.0637$	$p_1 = 0.56189$
$\lambda_2' = 0.3628$	$p_2 = 0.41463$
$\lambda_3' = 1.1107$	$p_3 = 0.02348$

Table 7.28: Parameters of the actual claims frequency distribution

With these actual distributions, the pure premium should have been

$$\mathbb{E}N \times \mathbb{E}X = 0.2122 \times 85.32 = 18.11$$

although it was

$$\mathbb{E}N \times \mathbb{E}X = 0.155 \times 113.40 = 17.58$$

with the observed distributions. As expected, the pure premium is higher with the actual distributions because in the case of the observed distribution, some claims are withheld by the policyholders which makes the aggregate claim amount distribution less important.

The frequencies now compare as

	actual frequency	frequency with bm	increase
λ_1	0.0637	0.0546	17%
λ_2	0.3622	0.2459	47%
λ_2	1.1107	0.9561	16%

Table 7.29: Comparison of the frequencies

This is not surprising as the bad drivers remain in the higher classes of the bonus-malus system and are less interested by the hunger for bonus because of the maximal penalty.

In order to analyze the impact of some hypothesis I have also conducted the calculations with an actualisation rate $\beta = \frac{1}{1+1\%}$ and $\beta = \frac{1}{1+11\%}$ on the one hand and with $t = 0.15$ and $t = 0.85$ on the other hand. The results are summarized in the following table :

	c	μ	p	$\lambda_1^{'}$	$\lambda_2^{'}$	$\lambda_3^{'}$	$\mathbb{E}N \times \mathbb{E}X$
$\beta = 1\%$	54.88	73.31	0.3788	0.0656	0.3948	1.1848	18.11
$\beta = 11\%$	42.47	87.32	0.3859	0.0633	0.3550	1.0887	18.19
$t = 0.15$	48.79	84.07	0.3990	0.0642	0.3771	1.0835	18.32
$t = 0.85$	46.58	84.67	0.3832	0.0639	0.3586	1.1056	17.82

Table 7.30: Characteristics of the distributions with changes hypothesis

We are not surprized to observe that the optimal retention is lower with higher interest rates. In this case, the policyholders will keep their money due to the high interest rate they can obtain. Similarly the optimal retention is lower when the renewal is closer. Further investigations should be necessary in order to draw general conclusions.

7.6 The practical replacement of a bonus-malus system

This section is taken from Walhin and Paris (2000f).

7.6.1 Notations

We will use the following notations :

- N is the random variable representing the number of observed claims within the current bonus-malus system
- $N^{'}$ is the random variable representing the number of actual claims
- $N^{''}$ is the random variable representing the number of claims under the optimal policy for the drivers using Lemaire's algorithm within the current bonus-malus system

- N''' is the random variable representing the number of future observed claims within the new bonus-malus system
- N'''' is the random variable representing the number of claims under the optimal policy for the drivers using Lemaire's algorithm within the new bonus-malus system

7.6.2 An iterative algorithm to find the new bonus-malus system

Let us assume that the number of classes of the new bonus-malus system has been chosen. The transition rules are also known.
On the one hand we have to determine the premium levels $(C_i\,,\, i = 0, \ldots, s-1)$ associated with each class of the bonus-malus system.
On the other hand the conception of the bonus-malus system is based on the future observed claims frequency distribution $((\lambda_j''' , p_j)\,, j = 1, \ldots, r)$ within the bonus-malus system. The problem is that the future observed claims frequency distribution is not known and depends on the premium levels C_i.
The solution is thus to find the bonus-malus levels as well as the future observed claims frequency distribution iteratively. The following algorithm is proposed.

Algorithm :

Initializing step :

Choose the number of classes of the practical bonus-malus system as well as the transition rules.
Use the actual claims frequency distribution as an initial guess for the future observed claims frequency distribution.

Iterations :

Do
  Use the future observed claims frequency distribution obtained as a nonparametric mixture of Poisson Distributions in order to find the observed claims frequency distribution under the form of a Hofmann Distribution by matching the first three moments.
  With the Hofmann Distribution obtain an optimal bonus-malus table.
  From the optimal bonus-malus table obtain the premium levels of the practical bonus-malus system.
    Make an initial choice for the future observed claims frequency distribution
    Run Lemaire's algorithm to obtain the optimal claims frequency distribution of the drivers and thus the future observed frequency distribution.
    Do
      Use the observed claims frequency distribution obtained by Lemaire's algorithm in order to rerun Lemaire's algorithm
    Until convergence
Until convergence.

7.6.3 Numerical example

We will now set up the practical bonus-malus system based on the hypothesis made in the introduction.
The optimization program used in order to match as best as possible the $P_{(k,t)}$ and the C_i is the following :

$$Min \sum_{k=0}^{4} \sum_{t=0}^{10} \sum_{i=0}^{8} w_{k,t,i} [P_{(k,t)} - C_i]^2 \tag{7.9}$$

under the following constraints

$$\begin{aligned} C_i & \quad \text{are integers} \\ C_4 &= 100 \\ C_{i+1} - C_i &\geq 0 \\ \sum_{i=0}^{8} e_{\infty}(i) C_i &\geq 100 \\ \sum_{i=0}^{8} e_0(i) C_i &\geq 100 \\ \sum_{i=0}^{8} e_2(i, \mathbf{e_0}) C_i &\geq 100 \\ \sum_{i=0}^{8} e_4(i, \mathbf{e_0}) C_i &\geq 100 \end{aligned}$$

where $\mathbf{e_0} = \{0.5729, 0.0562, 0.0661, 0.0784, 0.0421, 0.0441, 0.0457, 0.0429, 0.0516\}$ i.e. the stationary distribution of the drivers in the current bonus-malus system.
We will now fully describe the first iteration of our algorithm.
The hypotheses for the algorithm of Lemaire are the following

- $\beta = \frac{1}{1+6\%}$
- $P = 35$
- $t = \frac{1}{2}$
- X is exponentially distributed with mean 85.32. This is the actual claim amount distribution derived in section 7.5.6

As an initializing step we use the actual claims frequency distribution :

$\lambda_1^{'} = 0.0637$	$p_1 = 0.56189$
$\lambda_2^{'} = 0.3628$	$p_2 = 0.41463$
$\lambda_3^{'} = 1.1107$	$p_3 = 0.02348$

Table 7.31: Parameters of the actual claims frequency distribution (nonparametric fit)

The corresponding Hofmann Distribution is given by matching the first three moments :

$p = 0.2123$
$c = 0.2153$
$a = 0.8915$

Table 7.32: Parameters of the actual claims frequency distribution (Hofmann parametric form)

An optimal bonus-malus table is immediately obtained with the expected value premium principle :

k/t	0	1	2	3	4
1	84	158	237	317	399
2	73	136	203	271	340
3	64	119	177	236	296
4	57	106	157	210	263
5	52	96	142	189	236
6	48	87	129	171	214
7	44	80	118	157	196
8	41	74	109	145	181
9	38	69	101	135	168
10	36	65	95	126	157

Table 7.33: Optimal bonus-malus table

The practical bonus-malus table is derived by solving the minimization procedure (7.9):

s	0	1	2	3	4	5	6	7	8
C_s	89	89	93	99	100	130	134	141	155

Table 7.34: Practical bonus-malus levels associated with table 10

By iterating Lemaire's algorithm, we find the future observed claims frequency distribution :
The future observed claims frequency distribution is then

Iteration	$\lambda_1^{'''}$	$\lambda_2^{'''}$	$\lambda_3^{'''}$
1	0.0637	0.3628	1.1107
2	0.0455	0.3134	1.0839
3	0.0449	0.3095	1.0832
4	0.0449	0.3091	1.0831
5	0.0449	0.3091	1.0831

Table 7.35: Iterations in order to find the $\lambda_j^{'''}$

$\lambda_1^{'''} = 0.0449$	$p_1 = 0.56189$
$\lambda_2^{'''} = 0.3091$	$p_2 = 0.41463$
$\lambda_3^{'''} = 1.0831$	$p_3 = 0.02348$

Table 7.36: Future observed claims frequency distribution

This is the end of the first iteration of our algorithm.

The following iterations are summarized in the next table :

Iteration	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	λ_1'''	λ_2'''	λ_3'''
1	89	89	93	99	100	130	134	141	155	0.0449	0.3091	1.0831
2	82	85	95	100	100	133	143	163	204	0.0380	0.2708	1.0415
3	81	81	93	100	100	137	149	171	215	0.0368	0.2620	1.0332
4	79	84	93	100	100	140	152	174	218	0.0359	0.2581	1.0318
5	79	82	93	100	100	139	152	175	220	0.0359	0.2570	1.0295
6	79	82	93	100	100	139	152	175	220	0.0359	0.2570	1.0295

Table 7.37: Iterations of the algorithm

Let us note that with this bonus-malus system the future premium income is

t	Income
0	100
2	113.51
4	117.57
∞	118.38

Table 7.38: Future premium income

This means that in such a situation the Insurer will be able to offer a reduction of the premium because the financial desequilibrium is in his favour. Note that if the Insurer decides to do so, it may have an influence on the behaviour of the drivers using Lemaire's algorithm. Then a new observed frequency distribution should be calculated as well as the new transient and stationary distributions.

Let us have a look at the bonus-malus system obtained if we ease the constrain

$$\sum_{i=0}^{8} e_0(i)C_i \geq 100$$

We find

Iteration	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7	C_8	λ_1'''	λ_2'''	λ_3'''
1	58	75	79	89	100	112	123	151	185	0.0359	0.2658	1.0463
2	57	74	80	92	100	129	145	173	218	0.0332	0.2430	1.0228
3	58	73	80	92	100	132	149	176	225	0.0329	0.2399	1.0170
4	58	73	80	92	100	133	149	177	226	0.0328	0.2393	1.0162
5	58	73	80	92	100	133	150	177	226	0.0328	0.2392	1.0164
6	58	72	81	92	100	134	150	177	226	0.0328	0.2391	1.0167
7	58	73	81	92	100	133	150	177	226	0.0328	0.2392	1.0165
8	58	72	81	92	100	134	150	177	226	0.0328	0.2391	1.0167

Table 7.39: Iterations of the algorithm

As we can see, the solution is given by iteration 7 or iteration 8. The reason why we do not converge is that the C_i are restricted to be integers.
In this case the future premium income is

t	Income
0	86.07
2	100
4	104.17
∞	105.51

Table 7.40: Future premium income

If the company adopts this bonus-malus system, it accepts to lose money during the first two years. The profit made in subsequent years will rapidly absorb the losses of the first two years.

Conclusion

In this PhD thesis, we have extended two kinds of dependencies.
In the former case, with the multivariate versions of the individual risk model or collective risk model, we can discern dependencies within the claim amounts. In this case we are able to find recursions to evaluate the probability function of the aggregate claims distribution :

$$\mathbf{S} = \mathbf{X_1} + \cdots + \mathbf{X}_N$$

In the latter case, we have extended some results where dependence is introduced at the level of the counting distribution. In that case we are able to treat recursively the evaluation of the probability function of

$$(S, T) = (X_1 + \cdots + X_N, Y_1 + \cdots + Y_M)$$

with X independent of Y.
Whether it is possible to find recursions for the case where both types of dependencies are present, i.e.

$$(S, T) = (X_1 + \cdots + X_N, Y_1 + \cdots + Y_M)$$

with X not independent of Y is an open problem for future research.

The multivariate extensions of the collective risk model or individual risk model are easily treated by using multivariate ordinary generating functions. The classical univariate results of Panjer (1981), De Pril (1989) or Dhaene and Vandebroek (1995) are extended in a multivariate setting.

The natural application of these multivariate extensions is the study of the ruin probability of an Insurance Company buying excess of loss covers with clauses such as reinstatements, deductible aggregate, sliding scale, ...
In these cases, ignored in literature, the aggregate claims distribution of the Cedent is a function of a multivariate aggregate claims distribution which makes it compulsory to know the multivariate joint distribution. Henceforth the importance of the extensions of the collective risk model and individual risk model.
We have also pointed out that the bivariate extension of the collective risk model is useful for the evaluation of the aggregate claims distribution of a bivariate model where the bivariate counting distribution is Mixed Bivariate Hofmann distributed.
Obviously there will be other applications of the multivariate extensions in literature. This is also an open problem for future research. At Secura Belgian Re, we currently investigate the importance of the bivariate collective risk model for the tarification of excess of loss treaties.

As mentioned in chapter 2, it is necessary to have a discretized version of the claim amount

distributions in order to run the recursions. Moreover, it is advisable to have arithmetic distributions with a large span. It is interesting to note that a multivariate extension of the classical stochastic order is available and authorizes to find bounds on the different quantities of interest when studying the risk characteristics of an Insurance Company buying excess of loss cover.
One should note that recursions are not the only techniques existing for the calculation of aggregate claims distributions. Obviously the Fast Fourier Transform, simulation and the numerical solution of the corresponding Volterra equations deserve attention. In particular a research topic might be to try to make a systematic comparison of these techniques from the point of view computing time and accuracy in the scope of not only the univariate case but also the multivariate case.

Chapter 5 has been devoted to the problem of finding adequate distributions for the fitting of univariate counting distributions. A three-parameter model seems to be the best from a physical point of view : the Hofmann Distribution. This model, which can also be viewed as a Mixed Process, is both infinitely divisible and Mixed Poisson distributed. Moreover, it has fine features like a moment equation for the maximum likelihood estimation, recursive schemes for the evaluation of the probabilities as well as the distribution of the aggregate claims, easy construction of optimal bonus-malus systems with natural extensions. It also encompasses classical counting distributions : the Poisson, the Poisson Inverse Gaussian, The Negative Binomial, The Polya-Aeppli and the Neyman Type A Distributions.

Other models exist in literature but are only interesting from the point of view of fitting, namely the Generalized Poisson Distribution or the Weighted Poisson Distribution.
The Nonparametric Mixed Poisson Distribution proved its importance when necessary to take an average value of complicated functions of the parameter λ. The classical example is the stationary distribution of the drivers within a bonus-malus system with finite number of classes.

Bivariate counting distributions are becoming more and more popular nowadays. In this thesis, I have only shown how to construct Bivariate Hofmann Distributions. For this purpose I have used the property of mixing the Bivariate Poisson Distribution and also the trivariate reduction method. In both cases it was possible to find recursions for the probabilities as well as for the bivariate aggregate claims distributions. The same kind of methodology was used with the Zero Inflation model.

Finally bonus-malus systems are investigated into both from the theoretical and practical points of view. The construction of bonus-malus systems with an infinite number of classes is reviewed with the expected value principle, with the zero utility principle with an exponential utlity function, and with an exponential loss function. Bonus-malus tables for bivariate counting distributions are also constructed. However, the result is not satisfactory as the premiums always depend on the observed sum of both types of claims, which does not allow for a clear distinction between the frequencies. There is room for future research.
The practical construction of bonus-malus systems is also investigated into and we have extended the methodology proposed by Coene and Doray (1996). Moreover, we have analysed the problem of the influence of the bonus hunger and have proposed an algorithm to find the actual claim amount and frequency distributions within a bonus-malus system. With this tool in mind we are also able to construct bonus-malus systems that anticipate the effect of

the bonus hunger. Obviously, this topic should be further investigated into in the future as the problem of the a priori tarification (covariates) has not been taken into account in the present work.

In conclusion, the problem of studying recursions useful for actuaries has led to two important applications on the Belgian market : the control of reinsurance programmes of Insurance Companies and the construction or replacement of our bonus-malus system.

Bibliography

Bibliography

Adelson, R.M. (1966). Compound Poisson Distributions. *Operations Research Quarterly*, 17:73–75.

Ahmed, M.S. (1961). On a Locally Most Powerful Boundary Randomized Similar Test for the Independence of two Poisson Variables. *Ann. Math. Statist.*, 32:809–827.

Akaike, H. (1973). *Information Theory and an Extension of the Maximum Likelihood Principle*, pages 267–281. In (Petrov, B.N. and Csaki, F., 1973).

Ambagaspitya, R.S. (1999). On the Distributions of Two Classes of Correlated Aggregate Claims. *Insurance : Mathematics and Economics*, 24:301–308.

Ambagaspitya, R.S. and Balakrishnan, N. (1994). On the Compound Generalized Poisson Distributions. *Astin Bulletin*, 24:255–263.

Anscombe, F.J. (1950). Sampling Theory of the Negative Binomial and Logarithmic Series Distributions. *Biometrika*, 37:358–382.

Arbous, A.G. and Kerrich, J.E. (1951). Accidents Statistics and the Concept of Accident Proneness. *Biometrics*, 7:340–432.

Baker, C. (1977). *The Numerical Treatment of Integral Equations*. Clarendon Press, Oxford.

Barlow, R.E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing*. Holt, Rinehart and Winston.

Beall, G. (1940). The Fit and Significance of Contagious Distributions when Applied to Observations on Larval Insects. *Ecology*, 21:460–474.

Berg, C. and Forst, G. (1975). *Potential Theory on Locally Compact Abelian Groups*. Springer-Verlag, Berlin.

Berkson, J. (1980). Minimum Chi-Square, not Maximum Likelihood. *The Annals of Statistics*, 8:457–487.

Besson, J.L. and Partrat, C. (1992). Trend et Systèmes de Bonus-Malus. *Astin Bulletin*, 22:11–31.

Bissel, A. (1972). A Negative Binomial Model with Varying Element Sizes. *Biometrika*, 59:435–441.

Böhning, D. (1994). A Note on a Test for Poisson Overdispersion. *Biometrika*, 81:418–419.

Böhning, D. (1998). Zero-Inflated Poisson Models and C.A.MAN : A Tutorial Collection of Evidence. *Biometrical Journal*, 7:833–843.

Borgan, O., Hoem, J.M., and Norberg, R. (1981). A Nonasymptotic Criterion for the Evaluation of Automobile Bonus Systems. *Scandinavian Actuarial Journal*, pages 165–178.

Bühlmann, H. (1970). *Mathematical Models in Risk Theory*. Springer-Verlag, Berlin.

Bühlmann, H. (1984). Numerical Evaluation of the Compound Poisson Distribution : Recursion or Fast Fourier Transform ? *Scandinavian Actuarial Journal*, pages 116–126.

Bühlmann, H., Gagliardi, B., Gerber, H.U., and Straub, E. (1977). Some Inequalities for Stop-Loss Premiums. *Astin Bulletin*, 9:75–83.

Coene, G. and Doray, L.G. (1996). A Financially Balanced Bonus-Malus System. *Astin Bulletin*, 26:107–115.

Consul, P.C. (1989). *Generalized Poisson Distributions : Properties and Applications*. Marcel Dekker Inc., New York.

Consul, P.C. and Jain, G.C. (1973). A Generalization of the Poisson Distribution. *Technometrics*, 15:791–799.

Cresswell, W.L. and Frogatt, P. (1963). *The Causation of Bus Driver Accidents : An Epidemiological Study*. Oxford University Press, London.

David, H.A. (1970). *Order Statistics*. John Wiley and Sons, Inc, New York.

De Pril, N. (1985). Recursions for Convolutions of Arithmetic Distributions. *Astin Bulletin*, 15:135–139.

De Pril, N. (1986). On the Exact Computation of the Aggregate Claims distribution in the Individual Life Model. *Astin Bulletin*, 16:21–45.

De Pril, N. (1988). Improved Approximations for the Aggregate Claims Distributions of a Life Insurance Portfolio. *Scandinavian Actuarial Journal*, pages 61–68.

De Pril, N. (1989). The Aggregate Claims Distribution in the Individual Model with Arbitrary Positive Claims. *Astin Bulletin*, 19:9–24.

De Pril, N. and Dhaene, J. (1992). Error Bounds for Compound Poisson Approximations of the Individual Risk Model. *Astin Bulletin*, 22:137–148.

De Vylder, F. and Goovaerts, M.J. (1988). Recursive Calculation of Finite-Time Ruin Probabilities. *Insurance : Mathematics and Economics*, 7:1–7.

Del Castillo, J. and Perez-Casany, M. (1998). Weighted Poisson Distributions for Overdispersion and Underdispersion Situations. *Ann. Inst. Statist. Math.*, 50:567–585.

Delaporte, P. (1959). Quelques Problèmes de Statistique Mathématique Posés par l'Assurance et le Bonus pour Non Sinistre. *Bulletin Trimestriel de l'Institut des Actuaires Français*, 227:87–102.

den Isegeer, P.W., Smith, M.A.J., and Dekker, R. (1997). Computing Compound Distributions Faster ! *Insurance : Mathematics and Economics*, 20:23–34.

Denuit, M. (1997). A New Distribution of Poisson-Type for the Number of Claims. *Astin Bulletin*, 27:229–242.

Denuit, M. and Vermandele, C. (1998). Optimal Reinsurance and Stop-Loss Order. *Insurance : Mathematics and Economics*, 22:229–233.

Dhaene, J. and De Pril, N. (1994). On a Class of Approximative Computation Methods in the Individual Risk Model. *Insurance : Mathematics and Economics*, 14:181–196.

Dhaene, J. and Vandebroek, M. (1995). Recursions for the Individual Model. *Insurance : Mathematics and Economics*, 16:31–38.

Dickson, D.C.M. and Waters, H.R. (1999). Multi-Period Aggregate Loss Distributions for a Life Portfolio. *Astin Bulletin*, 29:295–309.

Embrechts, P., Grübel, R., and Pitts, S.M. (1993). Some Applications of the Fast Fourier Transform Algorithm in Insurance Mathematics. *Statistica Neerlandica*, 47:59–75.

Evans, D. (1953). Experimental Evidence Concerning Contagious Distributions in Ecology. *Biometrika*, 40:186–211.

Famoye, F. and Consul, P.C. (1995). Bivariate Generalized Poisson Distribution with some Applications. *Metrika*, 42:127–138.

Feller, W. (1971). *An Introduction to Probability Theory and its Applications Vol II (3ed).* Wiley, New-York.

Feller, W. (1943). On a General Class of Contagious Distributions. *Annals of Mathematical Statistics*, 14:389–400.

Galliher, H.P., Morse, R.M., and Simond, M. (1959). Dynamics of Two Classes of Continuous-Review Inventory Systems. *Operations Research.*, 7:362–383.

Gart, J.J. (1975). *The Poisson Distribution : The Theory and Applications of some Conditional Tests*, pages 125–140. Volume 2 of (Patil, G.P. et al., 1975).

Gerber, H. and Jones, D. (1976). Some Practical Considerations in Connection with the Calculation of Stop-Loss Premiums. *Transactions of the Society of Actuaries*, 28:215–231.

Gerber, H.U. (1979). *An Introduction to Mathematical Risk Theory.* S.S. Huebner Foundation Monograph 8, University of Pennsylvania, Philadelphia.

Gerber, H.U. (1982). On the Numerical Evaluation of the Distribution of Aggregate Claims and its Stop-Loss Premiums. *Insurance : Mathematics and Economics*, 1:13–18.

Gilde, V. and Sundt, B. (1989). On Bonus Systems with Credibility Scales. *Scandinavian Actuarial Journal*, pages 13–22.

Goovaerts, M.J., Kaas, R., van Heerwaarden, A.E, and Bauwelinckx, T. (1990). *Effective Actuarial Methods*. North-Holland, Amsterdam.

Grandell, J. (1997). *Mixed Poisson Processes*. Chapman and Hall.

Grübel, R. and Hermesmeier, R. (1999). Computation of Compound Distributions I : Aliasing Errors and Exponential Tilting. *Astin Bulletin*, pages 197–214.

Grübel, R. and Hermesmeier, R. (2000). Computation of Compound Distributions II : Discretization Errors and Richardson Extrapolation. *to be published in Astin Bulletin*.

Gupta, A.K., Móri, T.F., and Székely, G.J. (1994). Testing for Poissonity-Normality versus Other Infinite Divisibility. *Statistics* & *Probability Letters*, 19:245–248.

Hamdan, M.A. and Tsokos, C.P. (1971). A Model for Physical and Biological Problems : The Bivariate-Compounded Poisson Distribution. *Int. Statist. Rev*, 39:59–63.

Hesselager, O. (1994). A Recursive Procedure for Calculation of Some Compound Distributions. *Astin Bulletin*, 24:19–32.

Hesselager, O. (1996). Recursions for Certain Bivariate Counting Distributions and their Compound Distributions. *Astin Bulletin*, 26:35–52.

Hofmann, M. (1955). Über zusammengesetzte Poisson-Prozesse und ihre Anwendungen in der Unfallversicherung. *Bulletin of the Swiss Actuaries*, 55:499–575.

Hougaard, P. (1986). Survival Models for Heterogeneous Populations Derived from Stable Distributions. *Biometrika*, 73:387–396.

Hougaard, P., Ting Lee, M.L., and Whitmore, G.A. (1997). Analysis of Overdispersed Count Data by Mixtures of Poisson Variables and Poisson Processes. *Biometrics*, 53:1225–1238.

Hürlimann, W. (1990). On Maximum Likelihood Estimation for Count Data Models. *Insurance : Mathematics and Economics*, 9:39–49.

Johnson, L.W. and Riess, R.D. (1977). *Numerical Analysis*. Addison-Wesley, Reading.

Johnson, N.L., Kotz, S., and Balakrishnan, N. (1997). *Discrete Multivariate Distributions*. John Wiley and Sons, Inc, New York.

Johnson, N.L., Kotz, S., and Kemp, A. (1992). *Univariate Discrete Distributions*. John Wiley and Sons, Inc, New York.

Katti, S.K. (1967). Infinite Divisibility of Integer Valued Random Variables. *Ann. Math. Statist.*, 38:1306–1308.

Katz, L. (1965). *Unified Treatment of a Broad Class of Discrete Probability Distributions*, pages 175–182. In (Patil, 1965).

Kemp, A.W. (1981). Computer Sampling from Homogeneous Bivariate Discrete Distributions. *ASA Proceedings of the Statistical Computing Section*, pages 173–175.

Kemp, C.D. (1967). Stuttering-Poisson Distributions. *Journal of the Statistical and Social Inquiry Society of Ireland*, 21:151–157.

Kestemont, R.M. and Paris, J. (1985). Sur l'Ajustement du Nombre de Sinistres. *Bulletin of the Swiss Actuaries*, 85:157–163.

Klugman, S.A., Panjer, H.H., and Willmot, G.E. (1998). *Loss Models. From Data to Decisions.* Wiley.

Kocherlakota, S. (1988). On the Compounded Bivariate Poisson Distribution : A Unified Treatment. *Annals of the Institute of Statistical Mathematics*, 40:61–76.

Kocherlakota, S. and Kocherlakota, K. (1992). *Bivariate Discrete Distributions.* Marcel Dekker, New-York.

Kornya, P. (1983). Distribution of Aggregate Claims in the Individual Model. *Transactions of the Society of Actuaries*, 35:823–836.

Lemaire, J. (1977). La Soif du Bonus. *Astin Bulletin*, 9:181–190.

Lemaire, J. (1985). *Automobile Insurance : Actuarial Models.* Kluwer-Nijhoff, Netherlands.

Lemaire, J. (1979). How to Define a Bonus-Malus System with an Exponential Utility Function. *Astin Bulletin*, 19:274–282.

Lemaire, J. (1991). Negative Binomial or Poisson-Inverse Gaussian ? *Astin Bulletin*, 21:187–197.

Lemaire, J. (1995). *Bonus-Malus Systems in Automobile Insurance.* Boston : Kluwer.

Lindsay, B. (1995). *Mixture Models : Theory, Geometry and Applications.* NSI-CBMS Regional Conference Series in Probability and Statistics, Vol.5, Institute of Mathematical Statistics, Hayward, California.

Lüders, R. (1934). Die Statistik der seltenen Ereignisse. *Biometrika*, 26:108–128.

Maceda, E. (1948). On the Compound and Generalized Poisson Distributions. *Annals of Mathematical Statistics*, 19:414–416.

Mardia, K.V. (1970). *Families of Bivariate Distributions.* Charles Griffin and Sons, London.

Michel, R. (1993). On Berry-Esseen Results for the Compound Poisson Distribution. *Insurance : Mathematics and Economics*, 13:35–37.

Morillo, I. and Bermúdez, L. (1999). An Optimal Bonus-Malus System. *Papers presented at the third congress on Insurance : Mathematics and Economics*, 6.

Norberg, R. (1976). A Credibility Theory for Automobile Bonus Systems. *Scandinavian Actuarial Journal*, pages 92–107.

Ohlin, J. (1969). On a Class of Measures of Dispersion with Application to Optimal Reinsurance. *Astin Bulletin*, 5:248–266.

Panjer, H.H. (1981). Recursive Evaluation of a Family of Compound Distributions. *Astin Bulletin*, 12:22–26.

Panjer, H.H. and Wang, S. (1993). On the Stability of Recursive Formulas. *Astin Bulletin*, 23:227–258.

Panjer, H.H. and Willmot, G. (1984). Computational Techniques in Reinsurance Models. *Transactions of the 22nd International Congress of Actuaries, Sydney*, 4:111–120.

Panjer, H.H. and Willmot, G.E. (1992). *Insurance Risk Models*. Society of Actuaries.

Papageorgiou, H. and David, K.M. (1995). On a Class of Bivariate Compound Poisson Distributions. *Statistics & Probability Letters*, 23:93–104.

Partrat, C. (1994). Compound Model for Two Dependent Kinds of Claims. *Insurance : Mathematics and Economics*, 15:219–231.

Patil, G., editor (1965). *Classical and Contagious Discrete Distributions*. Statistical Publishing Society, Calcutta.

Patil, G.P. and Bildikar, S. (1967). Multivariate Logarithmic Series Distribution as a Probability Model in population and community Ecology and some of its Statistical Properties. *Journal of the American Statistical Association*, 62:655–674.

Patil, G.P., Kotz, S., and Ord, J.K., editors (1975). *Statistical Distributions in Scientific Work*. D. Reidel Publishing Company.

Patil, G.P., Pielou, E.C., and Waters, W.E., editors (1971). *Statistical Ecology, 1 : Spatial Patterns and Statistical Distribution*. University Park, Pennsylvania : Penn State University Press.

Petrov, B.N. and Csaki, F., editors (1973). *Second International Symposium on Information Theory*. Akademiai Kiado, Budapest.

Rao, B.R., Mazumdar, S., Waller, J.H., and Li, C.C. (1973). Correlation Between the Numbers of Two Types of Children in a Family. *Biometrics*, 29:271–279.

Rao, C. (1973). *Linear Statistical Inference and Its Applications*. Wiley.

Robertson, J. (1992). The Computation of Aggregate Loss Distributions. *Proceedings of the Casualty Actuarial Society*, LXXIX:57–133.

Schwartz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6:461–464.

Self, S. and Liang, K. (1987). Asymptotic Properties of Maximum Likelihood Estimators and Likelihood Ratio Tests Under Nonstandard Conditions. *Journal of the American Statistical Association*, 82:605–610.

Shaked, M. (1980). On Mixtures from Exponential Families. *Journal of the Royal Statistical Society B*, 42:192–198.

Silva, J.M. and Centeno, M. (1998). Comparing Risk Adjusted Premiums from the Reinsurance Point of View. *Astin Bulletin*, 28:221–239.

Simar, L. (1976). Maximum Likelihood Estimation of a Compound Poisson Process. *The Annals of Statistics*, 4:1200–1209.

Steutel, F.W. (1973). Some Recent Results in Infinite Divisibility. *Stochastic Processes and their Applications*, 1:123–143.

Ströter, B. (1985). The Numerical Evaluation of the Aggregate Claims Density Function via Integral Equations. *Blätter der Deutsche Gesellschaft für Versicherungsmathematik*, 17:1–14.

Sundt, B. (1991). On Excess of Loss Reinsurance with Reinstatement. *Bulletin of the Swiss Actuaries*, 1991:1–15.

Sundt, B. (1992). On Some Extensions of Panjer's Class of Counting Distributions. *Astin Bulletin*, 22:61–80.

Sundt, B. (1998). On Error Bounds for Multivariate Distributions.

Sundt, B. (1999a). Discussion on "D.C.M Dickson and H.R. Waters Multi-Period Aggregate Loss Distribution for a Life Portfolio. *Astin Bulletin*, 29:311–314.

Sundt, B. (1999b). On Multivariate Panjer's Recursions. *Astin Bulletin*, 29:29–46.

Sundt, B. (2000). Multivariate Compound Poisson Distributions and Infinite Divisibility.

Sundt, B. and Jewell, W.S. (1981). Further Results on Recursive Evaluation of Compound Distributions. *Astin Bulletin*, 12:27–39.

Teicher, (1954). On the Multivariate Poisson Distribution. *Skandinavisk Aktuarietidskrift*, 37:1–9.

Thyrion, P. (1961). Contribution à l'Etude du Bonus pour non Sinistre en Assurance Automobile. *Astin Bulletin*, 1:142–162.

Tremblay, L. (1992). Using the Poisson Inverse Gaussian Distribution in Bonus-Malus Systems. *Astin Bulletin*, 22:97–106.

Upton, G.J.G. and Lampitt, G.A. (1981). A Model for Interyear Change in the Size of Bird Populations. *Biometrics*, 37:113–127.

Vernic, R. (1997). On the Bivariate Generalized Poisson Distribution. *Astin Bulletin*, 27:23–31.

Walhin, J.F. (2000a). The Bivariate Hofmann Distribution and Applications. *submitted to Metrika.*

Walhin, J.F. (2000b). A Bivariate ZIP Model. *in revision.*

Walhin, J.F. (2000c). On the Use of the Multivariate Stochastic Order in Risk Theory. *submitted to Insurance : Mathematics and Economics.*

Walhin, J.F. (2000d). Some Comments on Two Approximations Used for the Pricing of Reinstatements. *Proceedings of the XXXIst International Astin Colloquium 2000, Porto Cervo*, pages 223–239.

Walhin, J.F. and Paris, J. (1998). On the Use of Equispaced Discrete Distributions. *Astin Bulletin*, 28:241–255.

Walhin, J.F. and Paris, J. (1999a). Excess of Loss Reinsurance with Reinstatements : Premium Calculation and Ruin Probability of the Cedant. *Proceedings of the third IME Congress, London.*

Walhin, J.F. and Paris, J. (1999b). Using Mixed Poisson Distributions in Connection with Bonus-Malus Systems. *Astin Bulletin*, 29:81–99.

Walhin, J.F. and Paris, J. (2000a). The Actual Claim Amount and Frequency Distributions Within a Bonus-Malus System.

Walhin, J.F. and Paris, J. (2000b). A Compound Poisson Model with Varying Element Sizes.

Walhin, J.F. and Paris, J. (2000c). The Effect of Excess of Loss Reinsurance with Reinstatements on the Cedent's Portfolio. *to be pusblished in Blätter der Deutsche Gesellschaft für Versicherungsmathematik.*

Walhin, J.F. and Paris, J. (2000d). A General Family of Bivariate Mixed Poisson Distributions.

Walhin, J.F. and Paris, J. (2000e). A Large Family of Discrete and Overdispersed Probability Laws. *forthcoming.*

Walhin, J.F. and Paris, J. (2000f). The Practical Replacement of a Bonus-Malus System. *submitted to Astin Bulletin.*

Walhin, J.F. and Paris, J. (2000g). Processus de Poisson Mélange et Formules Unifiées pour Systèmes Bonus-Malus. *to be published in Bulletin Français d'Actuariat.*

Walhin, J.F. and Paris, J. (2000h). Recursive Formulae for Some Bivariate Counting Distributions Obtained by the Trivariate Reduction Method. *Astin Bulletin*, 30:141–155.

Walhin, J.F. and Paris, J. (2000i). Some Comments on the Individual Risk Model and Multivariate Extension.

Wang, S. (1996). Premium Calculation by Transforming the Layer Premium Density. *Astin Bulletin*, 26:71–92.

Wang, S. and Panjer, H.H. (1994). Proportional Convergence and Tail-Cutting Techniques in Evaluating Aggregate Claim Distributions. *Insurance : Mathematics and Economics*, 14:129–138.

Westman, W.E. (1971). *Mathematical Models of Contagion and their Relation to Density and Basal Area Sampling Techniques*, pages 515–536. In (Patil, G.P. et al., 1971).

Willmot, G. (1987). The Poisson-Inverse Gaussian as an Alternative to the Negative Binomial. *Scandinavian Actuarial Journal*, pages 113–127.