HAL
open science

Estimation of foreseeable and unforeseeable risks in motor insurance

Weihong Ni, Corina Constantinescu, Alfredo D Egídio dos Reis, Véronique Maume-Deschamps

To cite this version:

Weihong Ni, Corina Constantinescu, Alfredo D Egídio dos Reis, Véronique Maume-Deschamps. Estimation of foreseeable and unforeseeable risks in motor insurance. 2019. hal-02126453

HAL Id: hal-02126453

https://hal.science/hal-02126453v1

Preprint submitted on 16 May 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimation of foreseeable and unforeseeable risks in motor insurance

Weihong Ni [1], Corina Constantinescu [2], Alfredo D. Egídio dos Reis [3]
& Véronique Maume-Deschamps [4, *† ‡].

[1] *CEMAPRE, ISEG, Universidade de Lisboa; wni.grace@gmail.com*
[2] *Institute for Financial and Actuarial Mathematics, Department of Mathematical Sciences, University of Liverpool, UK; c.constantinescu@liverpool.ac.uk*
[3] *ISEG and CEMAPRE, Universidade de Lisboa, Pt; alfredo@iseg.ulisboa.pt*
[4] *Institut Camille Jordan, Institut de Science Financière et dAssurances (ISFA); Université Claude Bernard Lyon 1, Lyon, Fr, veronique.maume@univ-lyon1.fr*

Abstract

This project works with the risk model developed by [6] and quests modelling, estimating and pricing insurance for risks brought in by innovative technologies, or other emerging or latent risks. The model considers two different risk streams that arise together, however not clearly separated or observed. Specifically, we consider a risk surplus process where premia are adjusted according to past claim frequencies, like in a Bonus-Malus (BM) system, when we consider a *classical* or *historical risk* stream and an *unforeseeable risk* one. These are unknown risks which can be of high uncertainty that, when pricing insurance (ratemaking and experience rating), suggest a sensitive premium adjustment strategy. It is not clear for the actuary to observe which claim comes from one or the other stream. When modelling such risks it is crucial to estimate the behaviour of such claims, occurrence and their severity. Premium calculation must fairly reflect the nature of these two kinds of risk streams.

We start proposing a model, separating claim counts and severities, then propose a premium calculation method, and finally a parameter estimation procedure. In the modelling we assume a Bayesian approach as used in credibility theory, a credibility approach for premium calculation and the use of the Expectation-Maximization (EM) algorithm in the estimation procedure.

Keywords: Mixed Poisson processes; Foreseeable risks; Unforeseeable risks; Bayesian estimation; Ratemaking; Experience rating; Credibility; Bonus-malus; EM algorithm.

Authors thank Fundación MAPFRE, through 2016 Research Grants Ignacio H. de Larramendi, for the financial support given to the project ["TAPAS (Technology Advancement on Pricing Auto inSurance)"*].

†Authors ([1]),([3]) gratefully acknowledge support from FCT/MEC (Fundação para a Ciência e a Tecnologia/Portuguese Foundation for Science and Technology) through national funds and when applicable co-financed financed by FEDER, under the Partnership Agreement PT2020

‡Authors ([1]),([4]) gratefully acknowledge support by the LABEX MILYON (ANR-10-LABX-0070) of Université de Lyon, within the program "Investissements d'Avenir" (ANR-11-IDEX- 0007) operated by the French National Research Agency (ANR)

1 Introduction

This project aims at developping and also applying statistical tools for parameter estimation into the model introduced by [6], where the so-called unforeseeable risks were taken into account. These risks refer to those that are not really predictable due to presence of very little knowledge about them. With the ever smarter way of living, the evaluation of risks could probably be different from classical criteria. For instance, the introduction of autonomous cars would possibly change the magnitude of risks on roads, yet we do not know in which direction precisely, although they were designed to reduce *human error*. To model such unforeseeable feature, a defective random variable was incorporated in the risk model by [6]. There, unforeseeable risks were reflected in claim frequencies only. Letting premium adjust according to claim numbers imitates the operation of a Bonus-malus system (briefly, BM system or BMS). This project takes the above cited idea, develops their work on the claim frequencies topic, incorporates the two risk stream proposal into the claim severity, add appropriate premium calculation and estimation, and works an estimation procedure, for the distribution parameters of the claim counts and severities, as well as for the premium estimation. For the premium calculation/estimation credibility theory is used, parameter estimation procedure uses the Expectation-Maximization algorithm (briefly EM algorithm).

EM algortithm is developped in [3], a *guided tour* can be found in [2]. We will be using mixture distributions, particular algorithm application for these can be found in [8]. For premium calculation we will be using Bayesian credibility methods. Introductory notions on credibility can be found in [5], Chapters 17-19, and more advanced in [1], completely devoted to credibility theory and applications.

The output of the project should demonstrate a more reasonable pricing model, especially applied for BMS, taking into account those uncertainties (e.g. emerging risk estimations). By embedding the premium adjustment rules in a classical collective insurance risk model, some new risk measurements will be introduced, so that they can be used in risk management. The work provides a theoretically structured method.

The insurance risk process we base our developments is that of defined by Equation (3) by [6], retrieved from [4], where the Poisson parameter λ is a realization of a random variable Λ. From there, the premium income along time uses a Bayesian approach and is estimated by the posterior mean $\hat{\lambda}(t) = \mathbb{E}[\Lambda|N(t)]$ (it is one of [4]'s proposed estimates). We mean, the risk process is driven by equation

$$U(t) = u + c\int_0^t \hat{\lambda}(s)ds - S(t)\,, \quad t \geq 0\,, \tag{1}$$

where, $S(t) = \sum_{j=0}^{N(t)} Y_j$ is the aggregate claims up to time t, $u = U(0)$ is the initial surplus, $\{Y_j\}_{j=1}^{\infty}$ is a sequence of independent and idenciatlly distributed (briefly, i.i.d.) random variables with common distribution $H_Y(.)$, with existing mean $\mu = \mathbb{E}[Y_1]$ and $Y_0 \equiv 0$. The premium rates are dynamically adjusted, the underlying counting process $\{N(t), t \geq 0\}$ is a mixed Poisson process with random itensity Λ. The other premium comonent, c, denotes the time constant part of the premium income, and often defining a loading as a proportion of the pure premium, here $c = (1+\theta)\mathbb{E}[Y_1]$, where θ (> 0) is the loading coefficient.

Another [4]'s neat idea that taken by [6] was to consider a positive probability $\mathbb{P}[\Lambda = 0] = p$ (> 0), so that the counting process $\{N(t)\}$ is a mixed Poisson process conditional on $\Lambda > 0$. The latter authors generalised to consider that the process $\{N(t) = N^{(1)}(t) + N^{(2)}(t)\}$ is a sum of two mixed counting processes, where $\{N^{(1)}(t)\}$ represents the *historical risk stream* of claim counts and $\{N^{(2)}(t)\}$ represents the corresponding *unforeseeable risk stream*. This is the process that we recapture for further developments. Concerning the first process the randomized intensity parameter has a classical behaviour, it is positive, or $\mathbb{P}\{\Lambda^{(1)} = 0\} = 0$, whereas in the second process and we set $\mathbb{P}\{\Lambda^{(2)} = 0\} =$

$p > 0$. The first one is assigned to a *historical* risk behaviour, and the second to an *unforeseeable* risk one. Putting this idea into (1), [6] set $\hat{\lambda}(t) = \mathbb{E}[\Lambda^{(1)} + \Lambda^{(2)}|N(t)]$.

Modelling in this way is assuming that the second stream may be either not as risky as the classical one, or much riskier. If we think of the autonomous cars, experiments to date have suggested that they are potentially safer than human driving (Waymo Team, 2016). On the other hand, if we think of the asbestos problem it may be the other way around. Not much interpretation is given to the situation where $p > 0$. About the possible outbreak of unforeseeable clims, we quote [6] at the start of their Section 2 where they say that

> *...once they broke out in a negative way, it would possibly be too late for an insurance company to control the losses... the expectation of* Λ *conditioning on* $\{\Lambda > 0\}$ *is very large and practically it could be assumed that it could be much more than the average number of claims in the "historical" stream.*

Modelling two different streams of risks for the same portfolio can be done exclusively on the claim count process like [4] and [6] did, or also in the claim severity if we consider dependence between $\{N(t), t \geq 0\}$ and sequence $\{Y_j\}_{j=1}^{\infty}$. This latter means that the different streams may (also) *bring* different severity behaviour. This is not dealt by the previous authors. We start dealing with the model as the above authors did, we mean work first the claim frequency component separately. [6] were more concerned in ruin probability calculation, embodying the effect of the unforesseable stream in the claim counts only.

The manuscript is organised as follows. In the next section we work the claim frequency component of the model, in Section 3 we complete the starting model with the introduction of claim severity, since we admit that the severity may bring some information on the stream origin. In Section 4 we deal with the premium calculation/estimation, where a Bayesian credibility approch is used. Section 5 is devoted to parameter estimation where the EM algortithm is used, dealing separately the claim count and severity estimation. We finish the section by adding some discussion on the estimation procedure and results. Our manuscript is finished inserting some notes on how to do future estimation by setting a *global* likelihood in order to estimate all parameters together, from both claim frequency and severity distributions, regarding the application of estimation algorithms like the EM algorithm.

2 Claim frequency component

The total claim numbers over a time period $(0, t]$ within the portfolio is given by

$$N(t) = N^{(1)}(t) + N^{(2)}(t)\,. \tag{2}$$

Due to the randomness in their intensity parameters, we know that $\{N^{(1)}(t)\}$ and $\{N^{(2)}(t)\}$ are both mixed counting processes. More specifically, we consider now that $\{N^{(1)}(t)\}$ is a mixed Poisson process, whereas $\{N^{(2)}(t)\}$ is mixed Poisson conditional on $\{\Lambda^{(2)} > 0\}$, since the intensity must be positive. We consider that the event $\{\Lambda^{(2)} = 0\}$ implies that $\{N^{(2)}(t) = 0\}$, with probability one, like behaving as a limiting situation of a Poisson intensity where the probability

$$\lim_{\lambda \downarrow 0} \mathbb{P}\{N^{(2)}(t) = 0 | \Lambda = \lambda\} = \lim_{\lambda \downarrow 0} e^{-\lambda\, t} = 1\,.$$

It was proved by [6] [see their Lemma 1] that $\{N(t)\}$ is also a mixed Poisson process with intensity $\Lambda^{(1)} + \Lambda^{(2)}$ given that $N^{(1)}(t)$ and $N^{(2)}(t)$ are independent. Under this global process, $\{N(t) = N^{(1)}(t) + N^{(2)}\}$, we have that when a claim arrives it comes either from process $\{N(t)^{(1)}\}$, or from

process $\{N(t)^{(2)}\}$. We remark that in practice actuaries will mostly observe the realization of the global claim process, which means they'll have to *guess* which process the arriving claim belongs to.

Let's define and denote Ξ as the *split rate*, in favour of process $\{N(t)^{(1)}\}$ in $\{N(t)\}$, between the two previous processes. It is well known that for independent Poisson processes, in our case for given and positive rates $\Lambda^{(1)} = \lambda^{(1)}$ and $\Lambda^{(2)} = \lambda^{(2)} > 0$, we have that the *split* between processes "(1)" and "(2)" in the combined process is going to be given by the probabilities, respectively as

$$\xi \;=\; \frac{\lambda^{(1)}}{\lambda^{(1)}+\lambda^{(2)}} \quad \text{and} \quad (1-\xi) = \frac{\lambda^{(2)}}{\lambda^{(1)}+\lambda^{(2)}}\,. \tag{3}$$

Unconditionally, considering that ξ is particular outcome of the random variable Ξ, then we can write, *extending* to the unconditional mixed process as defined in (2):

$$\Xi \;=\; \frac{\Lambda^{(1)}}{\Lambda^{(1)}+\Lambda^{(2)}}\,. \tag{4}$$

Note that

$$\{\Lambda^{(2)} = 0\} \iff \{\Xi = 1\} \Rightarrow \mathbb{P}[\Lambda = 0] = \mathbb{P}[\Xi = 1] = p\,. \tag{5}$$

We will return to this issue later in the text. We denote the distribution function of Ξ as $A_{\Xi}(.)$.

We will now set further assumptions. From now onwards, we assume assume that the Poisson intensities follow gamma distributions: $\Lambda^{(1)} \sim Gamma(\alpha_1, \beta_1)$ and, conditional on $\Lambda^{(2)} > 0$, $\Lambda^{(2)}|\Lambda^{(2)} > 0 \sim Gamma(\alpha_2, \beta_2)$. For simplicity, in the sequel we denote $\Lambda^{(2)}_{+} = \Lambda^{(2)}|\Lambda^{(2)} > 0$. Also, without loss of generality, we will be working with the random variable $N(1) = N$, denoting N for simplicity. Likewise with related quantities, such as $N^{(1)}(1) = N^{(1)}$ and $N^{(2)}(1) = N^{(2)}$.

Under a more restrict assumption for the gamma distributions above, we can arrive at the following result:

Lemma 1 *If $\beta_1 = \beta_2 = \beta$, then $N(1) = N$ is a mixture of two Negative Binomial random variables, such that*

$$\mathbb{P}(N=n) = p\binom{n+\alpha_1-1}{n}\left(\frac{\beta}{\beta+1}\right)^{\alpha_1}\left(\frac{1}{\beta+1}\right)^{n} + (1-p)\binom{n+\alpha_1+\alpha_2-1}{n}\left(\frac{\beta}{\beta+1}\right)^{\alpha_1+\alpha_2}\left(\frac{1}{\beta+1}\right)^{n}.$$

Proof. Let $N^{(1)}(1) = N^{(1)}$ and $N^{(2)}(1) = N^{(2)}$, for simplicity. Taking the moment generating function (briefly, mgf) of N, as function of ρ, yields

$$\begin{aligned} M_N(\rho) &= \mathbb{E}\left[e^{\rho\,(N^{(1)}+N^{(2)})}\right] = M_{N^{(1)}}(\rho)M_{N^{(2)}}(\rho) \\ &= \left(\frac{1-\frac{1}{1+\beta}}{1-e^{\rho}\frac{1}{1+\beta}}\right)^{\alpha_1}\left[p+(1-p)\left(\frac{1-\frac{1}{1+\beta}}{1-e^{\rho}\frac{1}{1+\beta}}\right)^{\alpha_2}\right] \\ &= p\left(\frac{1-\frac{1}{1+\beta}}{1-e^{\rho}\frac{1}{1+\beta}}\right)^{\alpha_1} + (1-p)\left(\frac{1-\frac{1}{1+\beta}}{1-e^{\rho}\frac{1}{1+\beta}}\right)^{\alpha_1+\alpha_2}. \end{aligned}$$

We know that this corresponds to a weighted average of two Negative Binomial distributions, with weights p and $1-p$ respectively. ■

Remark 1 *Note that $N^{(2)}$ follows a 'Zero Modified' Negative Binomial distribution, briefly, $N^{(2)} \frown ZM\ Negative\ Binomial(\alpha_2, \beta)$. It belongs to the $(a, b, 1)$ recursion class of distributions, see Section 6.6 of [5]. Since*

$$\begin{aligned} M_{N^{(2)}}(\rho) &= \mathbb{E}\left[e^{N^{(2)}}|\Lambda^{(2)}=0\right] \times \mathbb{P}\left[\Lambda^{(2)}=0\right] + \mathbb{E}\left[e^{N^{(2)}}|\Lambda^{(2)}>0\right] \times \mathbb{P}\left[\Lambda^{(2)}>0\right] \\ &= p + (1-p)\left(\frac{1-\frac{1}{1+\beta}}{1-e^{\rho}\frac{1}{1+\beta}}\right)^{\alpha_2}, \end{aligned}$$

which is a weighted average of the mgf's of a degenerate distribution at $\{0\}$ and a Negative Binomial, member of the $(a, b, 0)$ class. □

Alternatively, instead of focusing on the claim arrival process $\{N(t)\}$ which consists of combining two counting processes, $\{N^{(1)}(t)\}$ and $\{N^{(2)}(t)\}$, as explained above, we could model the underlying claim counts via mixing random variables, mixing two Gamma random variables with an independent Bernoulli random variable.

Lemma 2 *Assume that the randomized Poisson parameter Λ follows a prior distribution as a mixture of two Gamma random variables, such that*

$$\Lambda = I\, Z_1 + (1-I)\, Z_2, \tag{6}$$

where $Z_1 \sim Gamma(\alpha_1, \beta)$ and $Z_2 \sim Gamma(\alpha_1+\alpha_2, \beta)$ and I is a Bernoulli(p) random variable independent of Z_1 and Z_2. Then the marginal distribution of $N(t) = N$ for a fixed t is a mixture of two Negative Binomial random variables as shown in Lemma 1 above.

Proof. Recall that given $\Lambda = \lambda$, $\mathbb{P}\{N = n|\Lambda = \lambda\} = e^{-\lambda}\lambda^n/n!$. Now we integrate over Λ whose density, denoted as $\pi(.)$, can be written as

$$\pi(\lambda) = p \cdot \frac{\lambda^{\alpha_1-1}\beta^{\alpha_1}e^{-\beta\lambda}}{\Gamma(\alpha_1)} + (1-p) \cdot \frac{\lambda^{(\alpha_1+\alpha_2)-1}\beta^{\alpha_1+\alpha_2}e^{-\beta\lambda}}{\Gamma(\alpha_1+\alpha_2)}. \tag{7}$$

Hence,

$$\begin{aligned} \mathbb{P}\{N=n\} &= \int_\Lambda \frac{e^{-\lambda}\lambda^n}{n!}\left(p \cdot \frac{\lambda^{\alpha_1-1}\beta^{\alpha_1}e^{-\beta\lambda}}{\Gamma(\alpha_1)} + (1-p) \cdot \frac{\lambda^{(\alpha_1+\alpha_2)-1}\beta^{\alpha_1+\alpha_2}e^{-\beta\lambda}}{\Gamma(\alpha_1+\alpha_2)}\right) \mathrm{d}\lambda \\ &= p \cdot \int_\Lambda \frac{e^{-\lambda}\lambda^n}{n!}\frac{\lambda^{\alpha_1-1}\beta^{\alpha_1}e^{-\beta\lambda}}{\Gamma(\alpha_1)}\,\mathrm{d}\lambda + (1-\mathrm{p}) \cdot \int_\Lambda \frac{\mathrm{e}^{-\lambda}\lambda^{\mathrm{n}}}{\mathrm{n}!}\frac{\lambda^{(\alpha_1+\alpha_2)-1}\beta^{\alpha_1+\alpha_2}\mathrm{e}^{-\beta\lambda}}{\Gamma(\alpha_1+\alpha_2)}\,\mathrm{d}\lambda \\ &= p\binom{n+\alpha_1-1}{n}\left(\frac{\beta}{\beta+1}\right)^{\alpha_1}\left(\frac{1}{\beta+1}\right)^n \\ &\quad +(1-p)\binom{n+\alpha_1+\alpha_2-1}{n}\left(\frac{\beta}{\beta+1}\right)^{\alpha_1+\alpha_2}\left(\frac{1}{\beta+1}\right)^n. \end{aligned}$$

This distribution coincides with the one shown in Lemma 1. ■

The above enhances the fact that the distribution for claim counts $N(t)$ over a fixed period t is equivalent to a mixing over a Poisson with Gamma mixtures. In this way, under a Bayesian set-up, considering a prior distribution of Gamma mixtures, we will be able to compute the posterior distribution. This is important for computing estimates aiming experience rating. We show this in Section 4. In fact, the equivalence of the previous two constructions can be explained further. Consider the following remark.

Remark 2 *Let* $Z_1 = \Lambda^{(1)}$, $Z_2 = \Lambda^{(1)} + \Lambda_+^{(2)}$, *where* $\Lambda_+^{(2)} = \Lambda^{(2)}|\Lambda^{(2)} > 0$. *Also, define* $\Lambda^{(2)} = (1-I)\,\Lambda_+^{(2)} + I\Lambda_0^{(2)} = (1-I)\,\Lambda_+^{(2)}$, *with* $\Lambda_0^{(2)} = 0$ *and* $I \sim$ *Bernoulli*(p). *We have,*

$$\begin{aligned}\Lambda &= I\,Z_1 + (1-I)\,Z_2 = I\,\Lambda^{(1)} + (1-I)\,(\Lambda^{(1)} + \Lambda_+^{(2)}) \\ &= \Lambda^{(1)} + (1-I)\,\Lambda_+^{(2)} = \Lambda^{(1)} + \Lambda^{(2)}\,.\end{aligned}$$

Since $N|\Lambda \sim$ *Poisson* $(\Lambda^{(1)} + \Lambda^{(2)})$, *it is true that* $N = N^{(1)} + N^{(2)}$ *where* $N^{(1)}|\Lambda^{(1)} \sim$ *Poisson*$(\Lambda^{(1)})$ *and* $N^{(2)}|\Lambda_+^{(2)} \sim$ *Poisson*$(\Lambda_+^{(2)})$. □

In the model presented so far, we distinguish the *unforeseeable* stream of risks from the *historical* stream by considering two different, and independent, claim counting processes where the randomized intensities are of different nature. For now, these parameters only cause influence in the claim number process, not the claim size. However, we should also consider the possibility that the amounts are affected by the risk type. The introduction of a randomness itself in the intensities does not necessarily bring a dependence between the number of claims and sizes. However, as we put together the two risk streams in a global process, this is the one observed, and we consider that the claim size behaviour may depend on the risk stream, we need to admit dependence. This sets us away from the classical risk consideration that claim arrival process is independent of the claim amount one. We may assume the twofold stage:

1. On a starting stage, we can consider that for a given $\Lambda^{(i)}$, $i = 1, 2$, the independence is only (conditional) inside each (i). Or,

2. On a further stage, a model with *extended* dependence.

The first situation seems quite feasible for our model, if we consider no independence. Our next developments will clear the situation. We remark that parameters cannot be observed, only claims can. Further, we, as actuaries, only observe the realization of the total claim process $\{N(t) = N^{(1)}(t) + N^{(2)}(t)\}$, from there we'll have to *guess*/estimate which process each claim arrival belongs to. Then, using data and the model as behaving like is defined in Lemma 1 we have four parameters to estimate, p, α_1, α_2, and β.

As defined in (4), Ξ is viewed as the (random) split rate for process between processes $\{N^{(1)}(t)\}$ and $\{N^{(2)}(t)\}$ in $\{N(t)\}$. Let's consider the event $\{\Lambda^{(1)} + \Lambda^{(2)} = \lambda,\, \Xi = \xi,\, 0 < \xi < 1\}$ as given, so that $\{N(t)\}$ conditionally is a Poisson process. Consider the following lemma:

Lemma 3 *Under* $\{\Lambda^{(1)} + \Lambda^{(2)} = \lambda, \Xi = \xi, 0 < \xi < 1\}$, *the split rate, of process* $\{N(t) = N^{(1)}(t) + N^{(2)}(t)\}$ *in favour of* $\{N^{(1)}(t)\}$, *is given by* $\xi = \frac{\Lambda^{(1)}}{\Lambda^{(1)}+\Lambda^{(2)}}$.

Proof. Let us assume there are $n = n_1 + n_1$ number of arrivals within a fixed time $(0, t]$ where n_1 came from the historical stream and n_2 arrived from the unforeseeable stream. We could thus write, conditioning on $\{\Lambda^{(1)} + \Lambda^{(2)} = \lambda, \Xi = \xi\}$

$$\begin{aligned}&\mathbb{P}\left\{N^{(1)}(t) = n_1, N^{(2)}(t) = n_2 \middle| \Lambda^{(1)} + \Lambda^{(2)} = \lambda, \Xi = \epsilon\right\} \\ &\quad = \sum_{n=0}^{\infty} \mathbb{P}\left\{N^{(1)}(t) = n_1, N^{(2)}(t) = n_2 \middle| N^{(1)}(t) + N^{(2)}(t) = n, \Lambda^{(1)} + \Lambda^{(2)} = \lambda, \Xi = \xi\right\} \\ &\qquad \times \mathbb{P}\left\{N^{(1)}(t) + N^{(2)}(t) = n \middle| \Lambda^{(1)} + \Lambda^{(2)} = \lambda, \Xi = \xi\right\}.\end{aligned}$$

But $\mathbb{P}\left\{N^{(1)}(t)=n_1, N^{(2)}(t)=n_2 \middle| N^{(1)}(t)+N^{(2)}(t)=n, \Lambda^{(1)}+\Lambda^{(2)}=\lambda, \Xi=\xi\right\}=0$ when $n \neq n_1+n_2$, then equation above continues as

$$\begin{aligned}
&\mathbb{P}\left\{N^{(1)}(t)=n_1, N^{(2)}(t)=n_2 \middle| \Lambda^{(1)}+\Lambda^{(2)}=\lambda, \Xi=\epsilon\right\} \\
&\quad = \mathbb{P}\left\{N^{(1)}(t)=n_1, N^{(2)}(t)=n_2 \middle| N^{(1)}(t)+N^{(2)}(t)=n_1+n_2, \Lambda^{(1)}+\Lambda^{(2)}=\lambda, \Xi=\xi\right\} \\
&\qquad \times \mathbb{P}\left\{N^{(1)}(t)+N^{(2)}(t)=n_1+n_2 \middle| \Lambda^{(1)}+\Lambda^{(2)}=\lambda, \Xi=\xi\right\}.
\end{aligned}$$

The second probability on the right-hand side of the expression above does not depend on ξ, thus

$$\mathbb{P}\left\{N^{(1)}(t)+N^{(2)}(t)=n_1+n_2 \middle| \Lambda^{(1)}+\Lambda^{(2)}=\lambda, \Xi=\xi\right\} = \frac{(\lambda t)^{n_1+n_2} e^{-\lambda t}}{(n_1+n_2)!}.$$

Recall that under $\{\Lambda^{(1)}+\Lambda^{(2)}=\lambda, \Xi=\xi\}$, we have ξ denoting the *split rate* for the underlying Poisson process with parameter λ. Precisely, ξ means the probability of drawing an event from the first stream. Hence, given that n_1+n_2 events occurred, each event has probability ξ of being a Stream 1 event and probability $1-\xi$ of being a Stream 2 event, then we have a binomial distributed event coming into play. Hence, the first probability in the same (above) expression is given by

$$\begin{aligned}
&\mathbb{P}\left\{N^{(1)}(t)=n_1, N^{(2)}(t)=n_2 \middle| N^{(1)}(t)+N^{(2)}(t)=n_1+n_2, \Lambda^{(1)}+\Lambda^{(2)}=\lambda, \Xi=\xi\right\} \\
&\quad = \binom{n_1+n_2}{n_1} \xi^{n_1}(1-\xi)^{n_2}.
\end{aligned}$$

Hence,

$$\mathbb{P}\left\{N^{(1)}(t)=n_1, N^{(2)}(t)=n_2 \middle| \Lambda^{(1)}+\Lambda^{(2)}=\lambda, \Xi=\xi\right\} = \binom{n_1+n_2}{n_1} \xi^{n_1}(1-\xi)^{n_2} \frac{(\lambda t)^{n_1+n_2} e^{-\lambda t}}{(n_1+n_2)!}$$

With simple calculation, noting that we can write $\lambda=\lambda\,\xi+\lambda\,(1-\xi)$, we have

$$\mathbb{P}\left\{N^{(1)}(t)=n_1, N^{(2)}(t)=n_2 \middle| \Lambda^{(1)}+\Lambda^{(2)}=\lambda, \Xi=\xi\right\} = \left(\frac{e^{-\lambda\xi t}(\lambda\xi t)^{n_1}}{n_1!}\right)\left(\frac{e^{-\lambda(1-\xi)t}(\lambda(1-\xi)t)^{n_2}}{n_2!}\right).$$

We have independence between two Poisson random variables with means $\lambda^{(1)}t=\lambda\xi t$ and $\lambda^{(2)}t=\lambda\,(1-\xi)\,t$, and ξ is such that

$$\xi=\frac{\lambda^{(1)}}{\lambda}=\frac{\lambda^{(1)}}{\lambda^{(1)}+\lambda^{(2)}}.$$

■

Remark 3 *Remark that $\Xi=\xi$ is the probability that an event arriving from process $\{N(t)=N^{(1)}+N^{(2)}\}$ is generated by $\{N^{(1)}\}$. If $\{\Lambda^{(1)}+\Lambda^{(2)}=\lambda, \Xi=\xi\}$ is given then $\Lambda^{(1)}$ is also given since $\Lambda^{(1)}=\lambda\xi$. Then $\Lambda^{(2)}=\lambda-\Lambda^{(1)}=\lambda(1-\xi)$. If $\xi=1 \Rightarrow \Lambda^{(2)}=0 \Leftrightarrow \Lambda=\Lambda^{(1)}=\lambda$ and $N^{(2)}(t)=0\ \forall\, t\geq 0$ almost surely. In addition, since the values of $\lambda, \xi, \lambda^{(1)}$ are arbitrary, It should be also true that*

$$(\Lambda^{(1)}+\Lambda^{(2)})\Xi=\Lambda^{(1)} \Leftrightarrow \Xi=\frac{\Lambda^{(1)}}{\Lambda^{(1)}+\Lambda^{(2)}}.$$

The variable Ξ represents the random split rate of process $\{N(t)\}$ in favour of process $\{N^{(1)}(t)\}$ previously defined in the beginning of this section. □

In the next section we are going to model the influence of each risk stream in each claim severity.

3 Claim severity component

For now, we are going to admit that the distribution of the individual claim size depend on the stream type, either *historical* or *unforeseeable*. We keep presuming that the actuary may not be able to disclose which stream the claim comes from. At least, he cannot be certain.

Let F and G denote the distributions for claim severities in the *historical* and *unforeseeable* streams, respectively. Consider that the individual claim size, taken at random, say Y, follows a distribution function denoted as $H(y)$.

Lemma 4 *For a given claim Y, its distribution function, conditional on $\Xi = \xi$, can be represented by*

$$\mathbb{P}\{Y \leq y | \Xi = \xi\} := H_\xi(y) = \xi F(y) + (1-\xi)G(y), \tag{8}$$

where $\Xi = \frac{\Lambda^{(1)}}{\Lambda^{(1)}+\Lambda^{(2)}} \in (0,1]$, and F, G correspond to the distributions for claim severities in the historical and unforeseeable streams, respectively. The set $\{\Xi = 1\} = \{\Lambda^{(2)} = 0\}$ has a probability measure p.

Proof. The proof is straightforward using the law of total probability.

$$\begin{aligned}
\mathbb{P}\{Y \leq y|\xi\} &= \mathbb{P}\{Y \leq y|Y = Y^{(1)}, \xi\}\mathbb{P}\{Y = Y^{(1)}|\xi\} + \mathbb{P}\{Y \leq y|Y = Y^{(2)}, \xi\}\mathbb{P}\{Y = Y^{(2)}|\xi\} \\
&= F(y)\mathbb{P}\{Y = Y^{(1)}|\xi\} + G(y)\mathbb{P}\{Y = Y^{(2)}|\xi\} \\
&= F(y) \cdot \xi + G(y) \cdot (1-\xi).
\end{aligned}$$

where $Y^{(1)}$ and $Y^{(2)}$ denote random variables of the size of claims stemming from Stream 1 (the *historical stream*) and Stream 2 (the *unforeseeable stream*) respectively. Note that when $\Xi = 1$, i.e., $\Lambda^{(2)} = 0$, $N(t) = N^{(1)}(t)$ and the probability of having a claim from Stream 1, i.e., $\mathbb{P}\{Y = Y^{(1)}|\xi = 1\} = 1$, where as $\mathbb{P}\{Y = Y^{(2)}|\Xi = 1\} = 0$. This special situation does not affect Equality (8). ■

It has been illustrated in [6], see their Lemmas 2 and 3, that the independence between $\frac{\Lambda^{(1)}}{\Lambda^{(1)}+\Lambda_+^{(2)}}$ and $\Lambda^{(1)} + \Lambda_+^{(2)}$, conditional on $\Xi \neq 1$, result in $\Xi|\Xi \neq 1$ being distributed as a $Beta(\alpha_1, \alpha_2)$ law. Then, it is easy to derive the unconditional distribution of Y under these assumptions.

Lemma 5 *Assume that $\frac{\Lambda^{(1)}}{\Lambda^{(1)}+\Lambda_+^{(2)}}$ and $\Lambda^{(1)} + \Lambda_+^{(2)}$ are independent, and that $\Lambda^{(1)} \sim Gamma(\alpha_1, \beta)$ and $\Lambda_+^{(2)} \sim Gamma(\alpha_2, \beta)$. A given claim size is distributed according to a mixture law of $F(\cdot)$ and $G(\cdot)$, i.e., with density,*

$$h_Y(y) = \nu \cdot f(y) + (1-\nu) \cdot g(y), \tag{9}$$

where $\nu = p + (1-p)\frac{B(\alpha_1+1,\alpha_2)}{B(\alpha_1,\alpha_2)}$ and $B(\cdot,\cdot)$ being the Beta function.

Proof. Under the assumptions, we know that $\Xi|\Xi \neq 1 \frown Beta(\alpha_1, \alpha_2)$ according to [7]'s proportion-sum independence theorem. The subsequent proof is then straightforward. $A_\Xi(\xi)$ denotes the distribution function of Ξ (a mixture), we have

$$\begin{aligned}
\mathbb{P}\{Y \leq y\} &= \int_{(0,1]} H_\xi(y) dA(\xi) = pF(y) + (1-p)\int_{(0,1)} H_\xi(y) \cdot \frac{\xi^{\alpha_1-1}(1-\xi)^{\alpha_2-1}}{B(\alpha_1,\alpha_2)} d\xi \\
&= pF(y) + (1-p)\left[\frac{B(\alpha_1+1,\alpha_2)}{B(\alpha_1,\alpha_2)} \cdot F(y) + \frac{B(\alpha_1,\alpha_2+1)}{B(\alpha_1,\alpha_2)} \cdot G(y)\right] \\
&= \left[p + (1-p)\frac{B(\alpha_1+1,\alpha_2)}{B(\alpha_1,\alpha_2)}\right] \cdot F(y) + (1-p)\frac{B(\alpha_1,\alpha_2+1)}{B(\alpha_1,\alpha_2)} \cdot G(y) \\
&= \nu \cdot F(y) + (1-\nu) \cdot G(y)\,.
\end{aligned}$$

■

Remark 4 *If we look carefully at the end of the proof of [6]' Lemma 2 (see the Appendix, it is not clear from the Lemma) we can conclude that, under the conditions of Lemma 5, Y is independent from N. This will allow us to consider the parameter estimation using separately the claim frequency and the severity component.*

In the previous section we specified a distribution for the random variable $N(t)$, as well as its parametrization. The unconditional distribution for $N(t)$ was found starting from the Poisson distribution, this is commonly accepted as a starting assumption for claim count data in actuarial science, particularly in motor insurance. We ended up developing a particular mixture of two Negative Binomial distributions for $N(t)$, having started from a different Poisson distribution for each stream of risks. We'll see in Section 5 that the assumption is reasonable and may fit actual data. For the claim severity we will consider a similar reasoning, basing in a particular case from the Gamma distribution family. The Gamma distribution family is also a common fit for insurance severity data. First, we assume a particular form for the distribution of the individual claim severity as a mixture of two distributions, denoted as $F(\cdot)$ and $G(\cdot)$. They represent the behaviours of the claim severities from the two streams, separately the historical and unforeseeable streams, respectively $F(\cdot)$ and $G(\cdot)$. Then, if we specify $F(\cdot)$ and $G(\cdot)$, we can move for the parameter estimation.

Under many different possibilities we took a simple case but an illustrative one. Consider that both the densities of the two streams come from a common ground, and that in general the claim sizes are distributed as the exponential distribution. $Y \sim Exp(\Theta)$, where Θ^{-1} is the mean, and there is a dependence structure embedded in a random parameter Θ, such that:

1. $\Theta = \mu$ with probability one if we consider the historical stream; and,
2. $\Theta \sim Gamma(\delta, \sigma)$ if we consider the unforeseeable one.

This implies that claims in the historical stream conform to the Exponential distribution with mean μ^{-1}, whereas those of the unforeseeable stream conform to a Pareto distribution, $Pareto(\delta, \sigma)$. Their respective unconditional densities are

$$f(y) = \mu e^{-\mu y} \quad \text{and} \quad g(y) = \frac{\delta \sigma^{\delta}}{(\sigma + y)^{\delta+1}}, \quad \mu, \delta, \sigma > 0\,. \tag{10}$$

Our model will be complete, and ready for parameter estimation/distribution fit, once the premium calculation is defined. It is of course essential for the risk modelling as well as for risk pricing. So, in the next section we're going to address the premium calculation. Previous developments lead us naturally to deal premium calculation under a Bayesian framework. As explained in the above we will work separately the claim counts and severity premium calculation.

4 Bayesian premium

A Bayesian approach and credibility in premium calculation is naturally opened. We'll work with risk random variables that are observable and these variables follow distributions that are mixtures of other distributions. In turn, these latter distributions are assumed to be members of the linear exponential family. Associated to these we consider conjugate priors as distributions for the parameters in a Bayesian setting. Under this assumption credibility is shown to be exact, i.e., the credibility premium

coincides with the Bayesian premium. In another words, the Bayesian premium is a linear function with the respect to the observable risk data, making premium calculation straightforward.

Retrieving some essentials in the theory of credibility premium, particularly *credibility exact*, consider a generic a random variable, say X, following a distribution belonging to the linear exponential family, depending of some parameter, say θ. Its probability or density function is of the form [see [5], Chapters 5 and 15]:

$$f_{X|\Theta}(x|\theta) = \frac{p(x)e^{r(\theta)x}}{q(\theta)} \,. \tag{11}$$

If parameter θ is considered to be an outcome of a random variable Θ, whose density function is of the following form, denoted as $\pi(.)$:

$$\pi(\theta) = \frac{[q(\theta)]^{-k} e^{\mu k r(\theta)} r'(\theta)}{c(\mu, k)} \,, \tag{12}$$

then we are in the presence of a conjugate prior. In the above formulae, (11) and (12), $p(.)$ is some function not depending of parameter θ, $r(.)$, $q(.)$ are functions of the parameter, and $c(.)$ is a normalizing function of given parameters.

In this manuscript we are working with finite mixtures of prior distributions of form (12) and distributions of the linear exponential family (11). Recall that in the case of the claim counts we considered a Poisson random variable mixed with the mixture given by (7), and that in the case of the claim severity we considered an exponential mixed with a prior that can be given as a mixture of a degenerate distribution and a Gamma (see end of Section 3).

Consider from now on that the prior distribution is of the following form:

$$\pi(\theta) = \sum_{i=1}^{\eta} \omega_i \frac{[q(\theta)]^{-k_i} e^{\mu_i k_i r(\theta)} r'(\theta)}{c_i(\mu_i, k_i)} \,, \tag{13}$$

where ω_i, $i = 1, \cdots, \eta$, are given weights.

Theorem 1 *Suppose that given $\Theta = \theta$, the observable random variables $X_1, \ldots, X_n =: \mathbf{X}$ are i.i.d. with common probability function given by* (11)*, and that the prior distribution of Θ, $\pi(\theta)$, is of the form given by* (13)*. Then, the posterior distribution, denoted as $\pi(\theta|\mathbf{x})$ with $\mathbf{x} = \{x_1, \ldots, x_n\}$, is of a mixture form as* (13)*:*

$$\pi(\theta|\mathbf{x}) \;=\; \sum_i \omega_i^* \frac{[q(\theta)]^{-k_i^*} e^{\mu_i^* k_i^* r(\theta)} r'(\theta)}{c_i(\mu_i^*, k_i^*)}, \tag{14}$$

where

$$\mu_i^* \;=\; \frac{\mu_i k_i + \sum_j x_j}{k_i + n}, \tag{15}$$

$$k_i^* \;=\; k_i + n, \tag{16}$$

$$w_i^* \;=\; \frac{\omega_i \frac{c_i(\mu_i^*, k_i^*)}{c_i(\mu_i, k_i)}}{\sum_i \omega_i \frac{c_i(\mu_i^*, k_i^*)}{c_i(\mu_i, k_i)}} \,. \tag{17}$$

Proof. With observations $\mathbf{X} = \mathbf{x}$, the posterior distribution is

$$\begin{aligned}
\pi(\theta|\mathbf{x}) &= \frac{\frac{\prod_j p(x_j) e^{r(\theta)\sum_j x_j}}{[q(\theta)]^n} \cdot \sum_{i=1}^m \omega_i \frac{[q(\theta)]^{-k_i} e^{\mu_i k_i r(\theta)} r'(\theta)}{c_i(\mu_i, k_i)}}{\int_\Theta \left(\frac{\prod_j p(x_j) e^{r(\theta)\sum_j x_j}}{[q(\theta)]^n} \cdot \sum_{i=1}^m \omega_i \frac{[q(\theta)]^{-k_i} e^{\mu_i k_i r(\theta)} r'(\theta)}{c_i(\mu_i, k_i)} \right) \mathrm{d}\theta} \\
&= \frac{\sum_i \omega_i \frac{[q(\theta)]^{-k_i-n} e^{(\mu_i k_i + \sum_j x_j) r(\theta)} r'(\theta)}{c_i(\mu_i, k_i)}}{\sum_i \omega_i \int_\Theta \frac{[q(\theta)]^{-k_i-n} e^{(\mu_i k_i + \sum_j x_j) r(\theta)} r'(\theta)}{c_i(\mu_i, k_i)} \mathrm{d}\theta} \\
&= \frac{\sum_i \omega_i \frac{[q(\theta)]^{-k_i-n} e^{(\mu_i k_i + \sum_j x_j) r(\theta)} r'(\theta)}{c_i(\mu_i, k_i)}}{\sum_i \omega_i \frac{c_i(\mu_i^*, k_i^*)}{c_i(\mu_i, k_i)} \int_\Theta \frac{[q(\theta)]^{-k_i^*} e^{k_i^* \mu_i^* r(\theta)} r'(\theta)}{c_i^*(\mu_i^*, k_i^*)} \mathrm{d}\theta} = \sum_i \omega_i^* \frac{[q(\theta)]^{-k_i^*} e^{\mu_i^* k_i^* r(\theta)} r'(\theta)}{c_i(\mu_i^*, k_i^*)}.
\end{aligned}$$

■

From the posterior distribution we can calculate the Bayesian premium, as an estimate for the risk premium [see [5], Chapters 17 and 18] defined as $\mathbb{E}[X|\theta] = \mu(\theta)$. The Bayesian premium is given by estimate for next rating period, $n+1$, $\mathbb{E}[X_{n+1}|\mathbf{X} = \mathbf{x}]$ (with n, $\mathbf{X}$ and $\mathbf{x}$ as defined in Theorem 1) and can be calculated via the posterior distribution $\pi(\theta|\mathbf{x})$, as

$$\mathbb{E}[X_{n+1}|\mathbf{x}] = \mathbb{E}[\mu(\Theta)|\mathbf{x}]\,.$$

If this expectation is a linear function on the observations then we have exact credibility. This is the case when we work with conjugate distributions.

Moving to our particular concern, the prior distributions under study in our work, both for the claim counts and severity are of the above (mixture) form. As said in Remark 4, we can calculate the premium components, for the claim counts and severity, separately. We are going to consider first the claim frequency and then the severity component.

Before going to that we must write some considerations on premium calculation and estimation. Premia are set to be calculated for rating periods, most commonly the year period. For estimating, risk observations must be organized by periods, suppose that we observed the risk for m years, or rating periods. Then, our observation vector for the claim counts is $\mathbf{n} = \{n_1, \ldots, n_m\}$. For each year, say year j, we'll have n_j individual claims observed, say vector $\mathbf{y}_j = \{y_{j1}, \ldots, y_{jn_j}\}$. This means that observations in year j are set in the two fold vector $(n_j, \mathbf{y}_j)$. Of course that if we needed to calculate just the annual premium we would just need annual amounts on aggregate, since a premium is set for an aggregate quantity. However, we may consider that claim severities bring some information on an unforseeable stream. Then, it would be interesting to consider what is the premium contribution for that. Besides, we may also consider that information in both claim counts and severity is the same, then there is no need to double that in premium calculation.

Besides calculating Bayesian premia, we will still need to estimate parameters to feed in the Bayesian estimates (*Empirical Bayes*). This is done in Section 5.

4.1 Frequency component

If we consider a prior to be a Gamma mixture of the form given by (7), we reach to a posterior with a similar mixture, with updated paramenters. This is given in Lemma 6 that follows.

Lemma 6 *Let the prior distribution be* (7)*. With a Poisson model distribution, the posterior distribution is still a mixture of Gamma distributions, with updated parameters:*

$$\begin{aligned}\pi(\lambda|\mathbf{n}) &= w\cdot\frac{(\beta+m)^{\sum_j n_j+\alpha_1}\lambda^{\sum_j n_j+\alpha_1-1}e^{-(m+\beta)\lambda}}{\Gamma(\sum_j n_j+\alpha_1)}\\ &\quad+(1-w)\cdot\frac{(\beta+m)^{\sum_j n_j+\alpha_1+\alpha_2}\lambda^{\sum_j n_j+\alpha_1+\alpha_2-1}e^{-(m+\beta)\lambda}}{\Gamma(\sum_j n_j+\alpha_1+\alpha_2)},\end{aligned}\tag{18}$$

where

$$\begin{aligned}w &= \frac{1}{1+G(p,\alpha_1,\alpha_2,\beta,\mathbf{n})}\\ G(p,\alpha_1,\alpha_2,\beta,\mathbf{n}) &= \frac{1-p}{p}\cdot\frac{B(\alpha_1,\alpha_2)}{B(\sum_j n_j+\alpha_1,\alpha_2)}\cdot\left(\frac{\beta}{\beta+m}\right)^{\alpha_2},\end{aligned}\tag{19}$$

$\mathbf{n}=\{n_1,\ldots,n_m\}$ *is the vector of claim count observations and* m *is the sample size* ($\sum_j\cdot=\sum_{j=1}^m\cdot$)*.*

Proof. Under the observations $\mathbf{n}=\{n_1,\ldots,n_m\}$,

$$\begin{aligned}\pi(\lambda|\mathbf{n}) &= \frac{\prod_j\frac{e^{-\lambda}\lambda^{n_j}}{n_j!}\left\{p\cdot\frac{\lambda^{\alpha_1-1}\beta^{\alpha_1}e^{-\beta\lambda}}{\Gamma(\alpha_1)}+(1-p)\cdot\frac{\lambda^{(\alpha_1+\alpha_2)-1}\beta^{\alpha_1+\alpha_2}e^{-\beta\lambda}}{\Gamma(\alpha_1+\alpha_2)}\right\}}{\int_\Lambda\prod_j\frac{e^{-\lambda}\lambda^{n_j}}{n_j!}\left\{p\cdot\frac{\lambda^{\alpha_1-1}\beta^{\alpha_1}e^{-\beta\lambda}}{\Gamma(\alpha_1)}+(1-p)\cdot\frac{\lambda^{(\alpha_1+\alpha_2)-1}\beta^{\alpha_1+\alpha_2}e^{-\beta\lambda}}{\Gamma(\alpha_1+\alpha_2)}\right\}\mathrm{d}\lambda}\\ &= \frac{p\cdot\frac{\lambda^{\sum_j n_j+\alpha_1-1}\beta^{\alpha_1}e^{-(m+\beta)\lambda}}{\Gamma(\alpha_1)}+(1-p)\cdot\frac{\lambda^{\sum_j n_j+(\alpha_1+\alpha_2)-1}\beta^{\alpha_1+\alpha_2}e^{-(m+\beta)\lambda}}{\Gamma(\alpha_1+\alpha_2)}}{p\cdot\frac{\Gamma(\sum_j n_j+\alpha_1)}{\Gamma(\alpha_1)}\left(\frac{\beta}{\beta+m}\right)^{\alpha_1}\left(\frac{1}{\beta+m}\right)^{\sum_j n_j}+(1-p)\cdot\frac{\Gamma(\sum_j n_j+\alpha_1+\alpha_2)}{\Gamma(\alpha_1+\alpha_2)}\left(\frac{\beta}{\beta+m}\right)^{\alpha_1+\alpha_2}\left(\frac{1}{\beta+m}\right)^{\sum_j n_j}}\\ &= \frac{1}{1+G(p,\alpha_1,\alpha_2,\beta,\mathbf{n})}\cdot\frac{(\beta+m)^{\sum_j n_j+\alpha_1}\lambda^{\sum_j n_j+\alpha_1-1}e^{-(m+\beta)\lambda}}{\Gamma(\sum_j n_j+\alpha_1)}\\ &\quad+\frac{G(p,\alpha_1,\alpha_2,\beta,\mathbf{n})}{1+G(p,\alpha_1,\alpha_2,\beta,\mathbf{n})}\cdot\frac{(\beta+m)^{\sum_j n_j+\alpha_1+\alpha_2}\lambda^{\sum_j n_j+\alpha_1+\alpha_2-1}e^{-(m+\beta)\lambda}}{\Gamma(\sum_j n_j+\alpha_1+\alpha_2)}.\end{aligned}$$

The result is as described. ■

For a fixed individual claim size, The pure premium estimator will be considered as the posterior mean of the claim frequency component, i.e.,

$$\mathbb{E}[N_{m+1}|\mathbf{n}]=\mathbb{E}[\Lambda|\mathbf{n}]=w\cdot\frac{\sum_j n_j+\alpha_1}{\beta+m}+(1-w)\cdot\frac{\sum_j n_j+\alpha_1+\alpha_2}{m+\beta},\tag{20}$$

where z is given by (19).

We note that the Bayesian premium does not have the credibility form as it may look at first [see [5], Chapters 17 and 18]. Our Bayesian premium estimator is not linear on the observations, the $G(\cdot)$ function depend on vector $\mathbf{n}$, through a Beta function, and it is not a linear function on the observations. Furthermore, due to Stirling's formula we have that the quocient of the Gamma functions $\Gamma(n+a)/\Gamma(n+b)\sim n^{a-b}$ as $n\to\infty$, allowing us to write that the function

$$G(p,\alpha_1,\alpha_2,\beta,\mathbf{n})\sim\frac{1-p}{p}\frac{\Gamma(\alpha_1)}{\Gamma(\alpha_1+\alpha_2)}\left(\frac{\beta\sum_j n_j}{\beta+m}\right)^{\alpha_1}\text{ as }\sum_j n_j\to\infty\,.$$

In fact, the Bayesian premium is of credibility form when we work with conjugate priors. What we have here is a mix of conjugate priors. We have, though, a mixture of two credibility premia. For instance, we have that the formula

$$\frac{\sum_j n_j + \alpha_1}{\beta + m} = \frac{m}{\beta + m} \cdot \left(\frac{\sum_j n_j}{m} \right) + \frac{\beta}{\beta + m} \cdot \left(\frac{\alpha_1}{\beta} \right) \tag{21}$$

can be viewed as the credibility premium as if there was only the historical stream claim frequency. Similarly, $(\sum_j n_j + \alpha_1 + \alpha_2)/(m + \beta)$ gives a corresponding formula when considering the existence of an unforeseeable stream. Then, Formula (20) is a weighted average of these two premia, where the weights depend on the observations.

In order to calculate the Bayesian premium, we still need the claim severity component. A similar methodology can be implemented for the claim severity component.

4.2 Severity component

For the severity component, following what we wrote on the mean of the exponetial distribution at the end of Section 3, we can set the prior as

$$\pi(\theta) = \nu \Delta_\theta(\{\mu\}) + (1 - \nu) \frac{\theta^{\delta - 1} \sigma^\delta e^{-\sigma\theta}}{\Gamma(\delta)}, \tag{22}$$

where $\Delta_\theta(\{\mu\}) = \mathbf{1}_{\{\mu\}}(\theta)$ is the Dirac measure, and ν is as given in (9). The posterior is still a mixture, as follows. As said, given $\Theta = \theta$, the conditional distribution of a single severity is exponential with mean $1/\theta$.

We are calculating a premium estimate for the severity component only, based on the observed single quantities. Let's consider that the sample is generically of size m^* and the observation vector is $\mathbf{y} = \{y_1, \ldots, y_{m^*}\}$. If sample size of claim counts $\mathbf{N} = \mathbf{n}$ is m then $m^* = \sum_{j=1}^m n_j$, $n_j \geq 1$. The posterior distribution is set in the following Lemma 7 and the premium follows after in Expression (26).

Lemma 7 *Let the prior distribution be* (22) *and the conditional distribution of a single severity, given* $\Theta = \theta$*, be exponential with mean* $1/\theta$*. The posterior distribution is still a mixture distribution in terms of the prior form with updated parameters, such that*

$$\pi(\theta|\mathbf{y}) = \omega \cdot \Delta_\theta(\{\mu\}) + (1 - \omega) \cdot \frac{(\sigma + \sum_i y_i)^{m^* + \delta} \, \theta^{m^* + \delta - 1} \, e^{-(\sigma + \sum_i y_i)\theta}}{\Gamma(m^* + \delta)}, \tag{23}$$

where

$$\begin{aligned} \omega &= \frac{1}{1 + \varphi(\nu, \mu, \delta, \sigma, \mathbf{y})} \\ \varphi(\nu, \mu, \delta, \sigma, \mathbf{y}) &= \frac{1 - \nu}{\nu} \cdot \frac{\Gamma(m^* + \delta)\sigma^\delta}{\Gamma(\delta)(\sigma + \sum_i y_i)^{m^* + \delta}} \cdot \mu^{-m^*} e^{\mu \sum_i y_i}, \end{aligned} \tag{24}$$

with ν *as given in* (9)*,* $\mathbf{y} = \{y_1, \ldots, y_{m^*}\}$ *and* $\sum_i \cdot = \sum_{i=1}^{m^*} \cdot$.

Proof. Under the observations $\mathbf{y} = \{y_1, \ldots, y_{m^*}\}$, and let $\Pi_\Theta(\cdot)$ be the distribution function of Θ

$$\begin{aligned}
\pi(\theta|\mathbf{n}) &= \frac{\left(\prod_{i=1}^{m^*} \mu e^{-\mu y_i}\right)\{\nu \Delta_\theta(\{\mu\})\} + \left(\prod_{i=1}^{m^*} \theta e^{-\theta y_i}\right)\left\{(1-\nu)\frac{\theta^{\delta-1}\sigma^\delta e^{-\sigma\theta}}{\Gamma(\delta)}\right\}}{\int_\Theta \left(\prod_{i=1}^{m^*} \theta e^{-\theta y_i}\right) \mathrm{d}\Pi_\Theta(\theta)} \\
&= \frac{\nu\left(\mu^{m^*} e^{-\mu\sum_i y_i}\right)\Delta_\theta(\{\mu\}) + (1-\nu)\frac{\theta^{m^*+\delta-1}\sigma^\delta e^{-(\sigma+\sum_i y_i)\theta}}{\Gamma(m^*+\delta)}}{\nu\left(\mu^{m^*} e^{-\mu\sum_i y_i}\right) + (1-\nu)\frac{\Gamma(m^*+\delta)\sigma^{m^*+\delta}}{\Gamma(\delta)(\sigma+\sum_i y_i)^{m^*+\delta}}} \\
&= \frac{1}{1+\varphi(\nu,\mu,\delta,\sigma,\mathbf{y})}\Delta_\theta(\{\mu\}) + \frac{\varphi(\nu,\mu,\delta,\sigma,\mathbf{y})}{1+\varphi(\nu,\mu,\delta,\sigma,\mathbf{y})}\frac{(\sigma+\sum_i y_i)^{m^*+\delta}\,\theta^{m^*+\delta-1}\,e^{-(\sigma+\sum_i y_i)\theta}}{\Gamma(m^*+\delta)}.
\end{aligned} \tag{25}$$

The result is as described. ■

Unlike (20), the Bayesian premium for the severity component is not given directly by the posterior mean. That is given by the posterior expectation $\mathbb{E}[\mu(\Theta)|\mathbf{x}]$, and is the weighted average between the mean of the exponential $(1/\mu)$ and the mean of the (posterior) $Gamma(m^*+\delta, \sigma+\sum_i y_i)$:

$$\mathbb{E}[Y_{n+1}|\mathbf{y}] = \mathbb{E}[\mu(\Theta)|\mathbf{x}] = \omega\,\frac{1}{\mu} + (1-\omega)\,\frac{m^*+\delta}{\sum_{i=1}^{m^*} y_i + \sigma}\,, \tag{26}$$

with ω calculated through (24).

As a remark, we can see that as $m^* \to 0$, $\sum_i y_i \to 0$, and

$$\varphi(\nu,\mu,\delta,\sigma,\mathbf{y}) \sim \frac{1-\nu}{\nu}.$$

Hence, the Bayesian premium, estimate for the next rating period $m+1$ and denoted as P_{m+1} can be given by multiplying (20) with (26), feeding (19) and (24),

$$\begin{aligned}
\mathbb{E}[P_{m+1}|\mathbf{n},\mathbf{y}] &= \left(w\cdot\frac{\sum_{j=1}^m n_j + \alpha_1}{\beta+m} + (1-w)\cdot\frac{\sum_{j=1}^m n_j + \alpha_1 + \alpha_2}{m+\beta}\right) \\
&\qquad \times\left(\omega\cdot\frac{1}{\mu} + (1-\omega)\cdot\frac{m^*+\delta}{\sum_{j=1}^{m^*} y_i + \sigma}\right).
\end{aligned} \tag{27}$$

In this formula we have a set of prior parameters that are most often unkown, and they need to be estimated from observed data, leading to the *Empirical Bayesian Premium*. This is not an easy task, commonly when we have to deal with mixtures of distributions, we work with an unusual quantity of estimating parameters.

In the coming subsection we present a numerical example to address the asymptotics of functions G and φ, from (19) and (24), in the derived formulae, and to understand the non-linearity of these functions with the observed claims (counts and costs).

4.3 Numerical example

For illustration, we work with a numerical example setting a scenario where the claim counts follow a $(0, 2, 1)$ pattern in every three consecutive years over a 12-year period. That is to say, we investigate

Period	$\sum n_j$	$\sum_{i=1}^{m^*} y_i$	Premium
0	0	0	0.8840
1	0	0	0.6049
2	2	2.52	0.8135
3	3	3.98	0.8445
4	3	3.98	0.7473
5	5	6.60	0.8361
6	6	8.07	0.8047
7	6	8.07	0.7360
8	8	10.59	0.7968
9	9	11.82	0.7965
10	9	11.82	0.7423
11	11	14.50	0.7852
12	12	15.83	0.7813

Table 1: Annual Bayesian premia with $(0, 2, 1)$ pattern

one policyholder's claim histories and his/her claim counts for each consecutive three years are 0, 2, 1, respectively.

Here, each individual claim size is simulated to be the prior mean

$$\mathbb{E}[Y] = \nu\frac{1}{\mu} + (1-\nu)\frac{\delta}{\sigma},$$

plus a Normal distributed random error with mean 0 and standard deviation 0.1. Further, we assume that $\alpha_1 = 3$, $\alpha_2 = 1$, $\beta = 0.5$, $p = 0.6$, $\mu = 1$, $\delta = 2$, $\sigma = 0.5$. As a consequence, applying Formula (27), we compute annual Bayesian premia, figures and graph are shown in Table 1 and Figure 1, respectively.

The pattern of the premia reflects the periodical pattern we assigned for the claim counts over years. Note that the no claim situation in a particular year leads to a significant decrease in the premium for the upcoming year, yet the amount of reduction diminishes over time. In addition, the rate of increase in the annual premium is generally higher with two claims when compared to that with one claim only, as expected. Looking at the graph on the right hand side, in the long run, the premium converges to a fixed value with the growth of cumulative claim counts and costs.

The premium function depends on two quantities which are functions of past claim records. They are function $G(\cdot)$ and $\varphi(\cdot)$ from (19) and (24), respectively. Now, we turn to look at some numerical results for them and the corresponding weight functions they generate, respectively. We first show how $G(\cdot)$ varies according to years elapsed. It can be seen that the values of $G(\cdot)$ also behaves in a periodical manner due to the scenario setting, and eventually the speed of growth slows down. We refer to Figure 2 and on the left hand side graph. Graph on the right hand side is the weight function governed by $G(\cdot)$. That shows the weights in the mixture posterior distribution for the claim frequency component. A large $G(\cdot)$ value implies a smaller weight assigned for the historical risk stream. Both graphs demonstrate that $G(\cdot)$ as well as the historical weight will converge to a fixed value ultimately with extensive fluctuations in earlier years. One should be aware that over time cumulative claim counts and costs are non-decreasing. The graphs also verify the asymptotics of $G(\cdot)$ and the weight function when accumulated claim counts $m^* = \sum_j n_j \to \infty$, as mentioned earlier.

Next we address the φ and its associated weight function. Similarly, we can justify the asymptotic result of $\varphi(\cdot)$ when $\sum_i y_i \to 0$ is given by a fixed value dependent on ν. Additionally, $\varphi(\cdot)$ has a

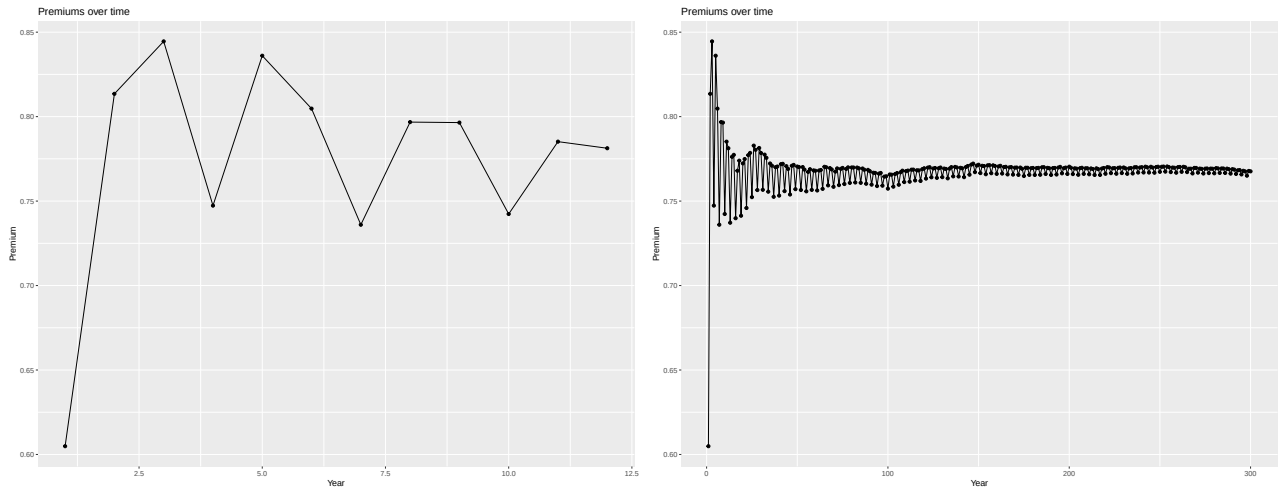

Figure 1: Premium evolution with $(0, 2, 1)$ pattern

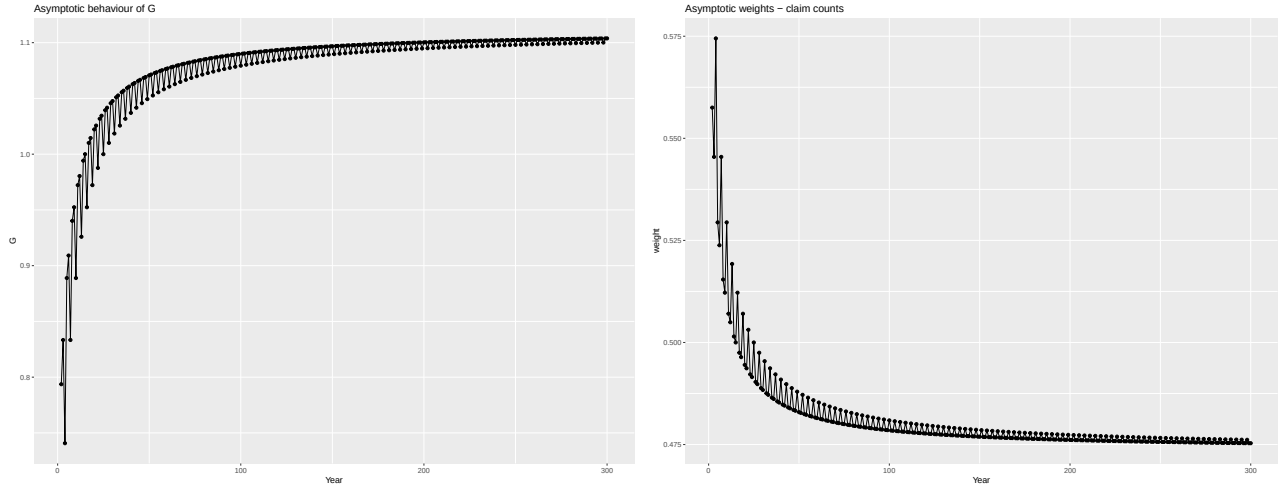

Figure 2: Asymptotic behaviour of function G and the posterior weight generated by G

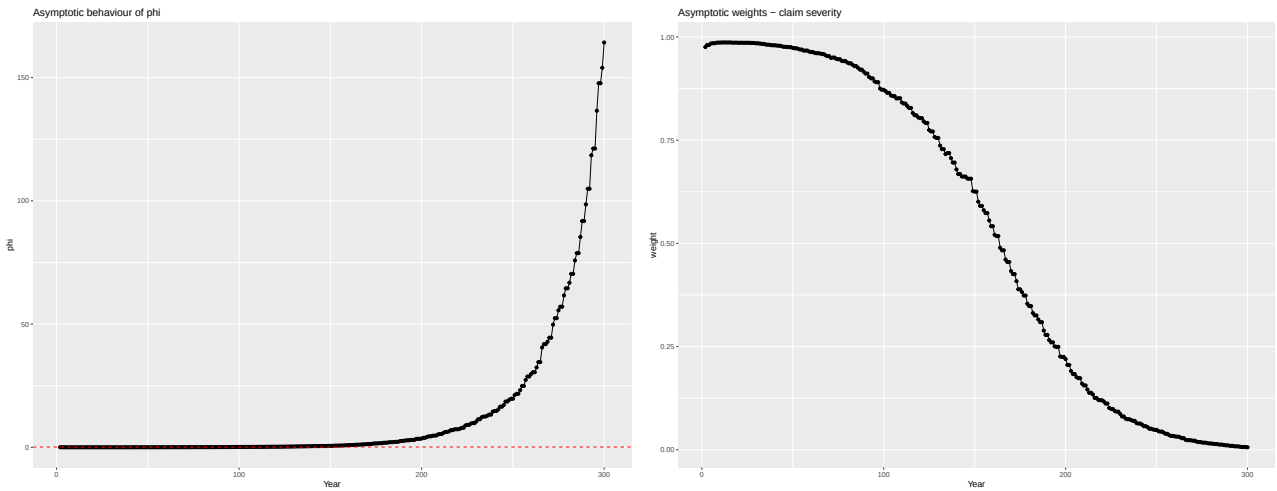

Figure 3: Asymptotic behaviour of function φ and the posterior weight generated by φ

different behaviour compared to $G(\cdot)$ overtime. See Figure 3. Firstly, $G(\cdot)$ and $\varphi(\cdot)$ differ in convexity, which leads to opposite asymptomatic directions. Secondly, when $\varphi(\cdot)$ gets larger, the variations tend to be slightly larger as well. Graph on the right shows the weight function for posterior distribution governed by $\varphi(\cdot)$. When the weight is larger, claim sizes are dominated by the exponential part in the mixture. And a smaller weight implies dominance by the gamma part. We could see that the weight starts at values closer to 1 and gradually lean towards values closer to 0. That means as time goes by, more weights are being assigned to the unforseeable risk stream when calculating premiums.

5 Parameter estimation and EM algorithm

We have seven prior parameters to estimate, the direct use of the Maximum Likelihood Estimation (briefly MLE) method is not workable. Looking a the literature we found that the use the Expectation-Maximization (EM) algorithm suited our purpose, it is said to find also, say, *best fit* estimates. It is an iterative way to approximate Maximum Likelyhood Estimates (MLE), it uses a multi-step process. It is said to be a method to find MLE for model parameters in the presence of incomplete data [see [3]], has missing data points, or when existing unobserved (hidden) latent variables. See also [2].

It works by choosing random values, guesses, for the missing data points, then uses those guesses to estimate a second set of data. The new values are used to create a better guess for the first set, and the process continues until the algorithm converges to a fixed point.

In this section we explain briefly how we employed the EM algorithm to estimate the prior parameters of our model based on a data set from a Portuguese insurer. Unfortunately we could not find a data set having both the claim counts and the corresponding severities. Often data is recorded on aggregate. The implementation of the algorithm depend on the distribution choice. We assumed the use of a simple distribution as a choice for the severity, other realistic choices may bring other problems to the estimation procedure, like numerical ones. However, some other choices are doable likewise.

5.1 Data description

We were provided with quarterly claim count data from a motor insurance portfolio. Namely, it records the total number of claims arising from the portfolio every quarter and there were a total of

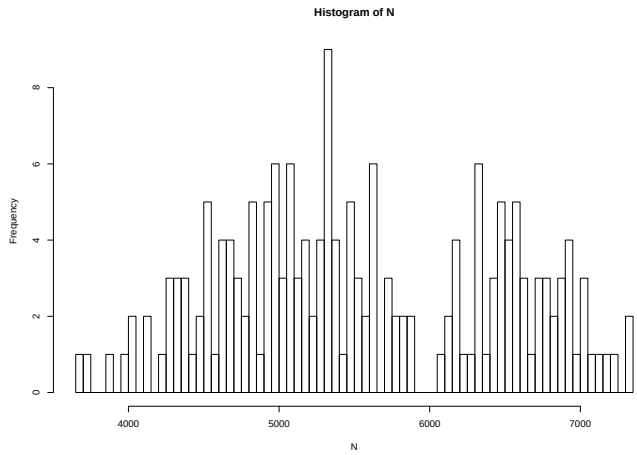

Figure 4: Histogram of claim counts

approximately 180 quarters recorded. Therefore, we could treat each entry as an observation for $N(t)$ for a fixed t which stands for a quarter here. We start to write N for short in the sequel since $N(t)$ is a random variable for fixed t. We took care of the Third Party Liability claims only as those are the obligatory components for a car insurance policy. It was assumed that the portfolio is closed over the underlying period and that claim counts for each quarter are independent observations for N.

Table 2 shows a brief summary of the data under consideration. The last two columns show the standard deviation and the coefficient of variation for this dataset. Its distribution can be also visualised in the histogram shown by Figure 4. Clearly, there is a separation around 6000, which serves as a clue for the adoption of a mixture model as we theoretically derived above. We will explain next in more details about the estimation procedures.

5.2 EM algorithm

5.2.1 Claim frequency component

Recall that we work with a mixture model and that we only observe a global $N_i \in \mathbb{N}$, we mean, we do not know which negative binomial distribution it comes from. Here, we introduce a latent variable, denoted as Z, representing the missing information of which distribution it takes for each observation. Since we only have two mixed distributions, $Z \frown Bernoulli(p)$ is a Bernoulli random variable with $\mathbb{P}\{Z = 1\} = q = 1 - \mathbb{P}\{Z = 0\}$. Now, the complete information is given by vector $\{N, Z\}$. We can

Min.	**1st Qu.**	**Median**	**Mean**	**3rd Qu.**	**Max.**	**S.D.**	**C.V.**
3685	4879	5373	5552	6432	7316	902.844	0.1626

Table 2: Summary of statistics

write the complete likelihood function as follows.

$$
\begin{aligned}
L(\theta|N,Z) &= \prod_{i=1}^{m}\left[p\binom{n_i+\alpha_1-1}{n_i}\left(\frac{\beta}{\beta+1}\right)^{\alpha_1}\left(\frac{1}{\beta+1}\right)^{n_i}\right]^{z_i} \\
&\quad \times\left[(1-p)\binom{n_i+\alpha_1+\alpha_2-1}{n_i}\left(\frac{\beta}{\beta+1}\right)^{\alpha_1+\alpha_2}\left(\frac{1}{\beta+1}\right)^{n_i}\right]^{1-z_i},
\end{aligned}
$$

and according to the law of total probability we have,

$$
\mathbb{P}\{N\} = \mathbb{P}\{N|Z=1\}\mathbb{P}\{Z=1\} + \mathbb{P}\{N|Z=0\}\mathbb{P}\{Z=0\}.
$$

Here, $\theta = \{\alpha_1, \alpha_2, \beta, p\}$ are the parameters we are interested in estimating for the claim count distribution. Correspondingly, the log-likelihood function is given by, from above,

$$
\begin{aligned}
\mathcal{L}(\theta|N,Z) &= \sum_i z_i \log\left[p\binom{n_i+\alpha_1-1}{n_i}\left(\frac{\beta}{\beta+1}\right)^{\alpha_1}\left(\frac{1}{\beta+1}\right)^{n_i}\right] \\
&\quad + \sum_i (1-z_i)\log\left[(1-p)\binom{n_i+\alpha_1+\alpha_2-1}{n_i}\left(\frac{\beta}{\beta+1}\right)^{\alpha_1+\alpha_2}\left(\frac{1}{\beta+1}\right)^{n_i}\right]. \quad (28)
\end{aligned}
$$

We summarize how the algorithm works. It is an iterative process where each iteration consists of two steps, the E-step and the M-step, standing for Expectation and Maximisation, respectively.

1. We begin with an initially determined parameter values $\theta^{(0)} = \{\alpha_1^{(0)}, \alpha_2^{(0)}, \beta^{(0)}, p^{(0)}\}$.

2. E-Step

   For the $(l+1)$-th iteration, $l = 0, 1, \ldots$, we first seek for the expected value of Z_i conditional on the observations together with the current parameter estimates $\theta^{(l)}$, i.e., estimates from the previous l-th iteration. This is denoted as τ_i,

$$
\begin{aligned}
\tau_i &= \mathbb{E}[Z_i|N,\theta^{(l)}] = 1\times\mathbb{P}\{Z_i=1|N_i,\theta^{(l)}\} + 0\times\mathbb{P}\{Z_i=0|N_i,\theta^{(l)}\} = \frac{\mathbb{P}\{Z_i=1, N_i=n_i|\theta^{(l)}\}}{\mathbb{P}\{N_i|\theta^{(l)}\}} \\
&= \frac{p^{(l)}\binom{n_i+\alpha_1^{(l)}-1}{n_i}\left(\frac{\beta^{(l)}}{\beta^{(l)}+1}\right)^{\alpha_1^{(l)}}\left(\frac{1}{\beta^{(l)}+1}\right)^{n_i}}{p^{(l)}\binom{n_i+\alpha_1^{(l)}-1}{n_i}\left(\frac{\beta^{(l)}}{\beta^{(l)}+1}\right)^{\alpha_1^{(l)}}\left(\frac{1}{\beta^{(l)}+1}\right)^{n_i} + (1-p^{(l)})\binom{n_i+\alpha_1^{(l)}+\alpha_2^{(l)}-1}{n_i}\left(\frac{\beta^{(l)}}{\beta^{(l)}+1}\right)^{\alpha_1^{(l)}+\alpha_2^{(l)}}\left(\frac{1}{\beta^{(l)}+1}\right)^{n_i}}
\end{aligned}
$$

   Subsequently, based on Expression (28), we compute the expectation the log-likelihood function with respect to the conditional distribution of Z, given N under the current estimates $\theta^{(l)}$,

denoted as $Q(\theta|\theta^{(l)})$. Considering we have we have m independent observations, we have

$$
\begin{aligned}
& Q(\theta|\theta^{(l)}) = \mathbb{E}_{Z_i|N_i,\theta^{(l)}}[\mathcal{L}(\theta|N_i, Z_i)] \qquad (29)\\
= & \sum_{i=1}^{m} \tau_i \left[\log p^{(l)} + \log \binom{n_i + \alpha_1^{(l)} - 1}{n_i} + \alpha_1^{(l)} \log \frac{\beta^{(l)}}{\beta^{(l)}+1} + n_i \log \frac{1}{\beta^{(l)}+1} \right] \\
& + \sum_{i=1}^{m} (1-\tau_i) \left[\log(1-p^{(l)}) + \log \binom{n_i + \alpha_1^{(l)} + \alpha_2^{(l)} - 1}{n_i} + (\alpha_1^{(l)} + \alpha_2^{(l)}) \log \frac{\beta^{(l)}}{\beta^{(l)}+1} + n_i \log \frac{1}{\beta^{(l)}+1} \right] \\
= & \log p^{(l)} \sum_i \tau_i + \sum_i \tau_i \log \binom{n_i + \alpha_1^{(l)} - 1}{n_i} + \alpha_1^{(l)} m \log \frac{\beta^{(l)}}{\beta^{(l)}+1} + \log \frac{1}{\beta^{(l)}+1} \sum_i n_i \\
& + \log(1-p^{(l)}) \sum_i (1-\tau_i) + \sum_i (1-\tau_i) \log \binom{n_i + \alpha_1^{(l)} + \alpha_2^{(l)} - 1}{n_i} + \alpha_2^{(l)} \log \frac{\beta^{(l)}}{\beta^{(l)}+1} \sum_i (1-\tau_i).
\end{aligned}
$$

3. M-Step
   The maximisation step is set to find the parameter values that maximises function (29) and they become the estimates to be used in the next iteration, i.e.,

$$\theta^{(l+1)} = \underset{\theta}{\operatorname{argmax}}\, Q(\theta|\theta^{(l)}) \qquad (30)$$

   For this, we take the gradient of $Q(\theta|\theta^{(l)})$, equate to zero and solve for $\{\alpha_1, \alpha_2, \beta, p\}$ simultaneously.

$$
\begin{aligned}
\frac{\partial Q}{\partial \alpha_1} &= \sum_i \tau_i [\digamma(n_i + \alpha_1) - \digamma(\alpha_1)] + \sum_i (1-\tau_i)[\digamma(n_i + \alpha_1 + \alpha_2) - \digamma(\alpha_1 + \alpha_2)] + m \log \frac{\beta}{1+\beta} = 0; \\
\frac{\partial Q}{\partial \alpha_2} &= \sum_i (1-\tau_i) \left[\digamma(n_i + \alpha_1 + \alpha_2) - \digamma(\alpha_1 + \alpha_2)] + \log \frac{\beta}{1+\beta} \right] = 0; \\
\frac{\partial Q}{\partial \beta} &= \frac{\alpha_1 m + \alpha_2 \sum_i (1-\tau_i) - \beta \sum_i n_i}{\beta(1+\beta)} = 0; \\
\frac{\partial Q}{\partial p} &= \frac{\sum_i \tau_i}{p} - \frac{\sum_i (1-\tau_i)}{1-p} = 0, \qquad (31)
\end{aligned}
$$

   where $\digamma(\cdot)$ is the digamma function denoting the logarithmic derivative of a gamma function, i.e.,

$$\digamma(x) = \frac{d}{dx} \log(\Gamma(x)) = \frac{\Gamma'(x)}{\Gamma(x)}.$$

   The estimates for the subsequent iteration $(l+1)$-th will be the solutions of the above equations. Equation (31) does not depend on others than parameter p, it can be solved directly with explicit representation

$$p^{(l+1)} = \frac{\sum_i \tau_i}{m}.$$

   For the other three parameters, we can only solve numerically. The *nleqslv* package in R was employed here, it solves systems of non-linear equations. As a consequence, we can obtain estimated parameters at this iteration $\theta^{(l+1)} = \{\alpha_1^{(l+1)}, \alpha_2^{(l+1)}, \beta^{(l+1)}, p^{(l+1)}\}$.

4. Plug $\theta^{(l+1)}$ into the $(l+2)$-th iteration and repeat the above steps until convergence.

Initiating values using the method of moments yields $\alpha_1^{(0)} = 96.14042$; $\alpha_2^{(0)} = 31.54888$; $\beta^{(0)} = 0.01927362$; $p^{(0)} = 0.6555556$. Then we implemented EM algorithm on the chosen dataset mentioned

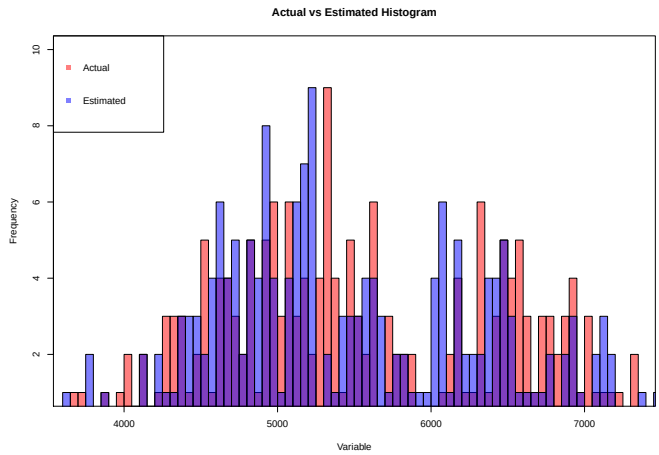

Figure 5: Histogram Actual *vs.* Estimated

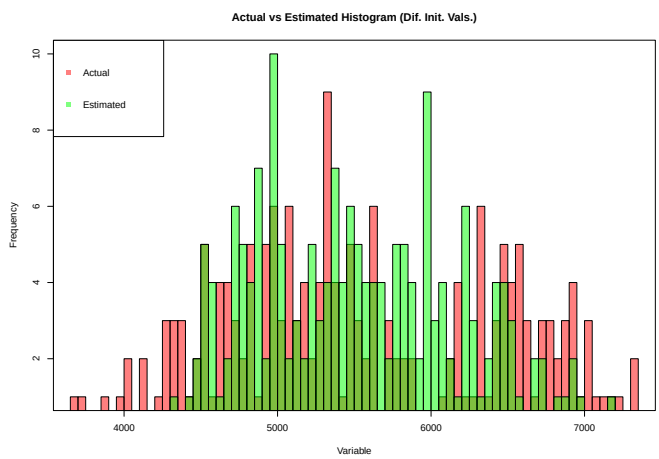

Figure 6: Histogram Actual *vs.* Estimated using a different set of init. vals

earlier. At the tolerance level of 0.001, we reached convergence with 75 iterations and the resulting estimates for $\alpha_1, \alpha_2, \beta, p$ are

$$\alpha_1 = 97.55820446;\ \alpha_2 = 30.14706672;\ \beta = 0.01978072;\ p = 0.5929959.$$

Based on these parameters, we randomly generated a sequence of numbers and compared them with the observed data. We put together both histograms, in different colors, in Figure 5. Visually it appears to be a good fit. Note that, however, this result stemmed from initial values derived from moments. In addition, we found that estimates vary much with different chosen initial values. For instance, if we begin with $\alpha_1^{(0)} = 150; \alpha_2^{(0)} = 20; \beta^{(0)} = 0.02$. Convergence happened at the 308th iteration at the same toleration level above. Nevertheless, corresponding estimates lie closer to the initial values.

$$\alpha_1 = 141.98069243; \alpha_2 = 28.02122426; \beta = 0.02773337, p = 0.5723845.$$

As can be seen in Figure 6, we still have a quite good fit visually. But we could already tell that it possibly does not provide a fitting as good as, if not worse than the previous one.

Reference Distribution	Kolmogorov-Smirnov test	Chi-square test
NB1(97.55820446, 0.01978072) and NB2(127.70527118, 0.01978072)	0.3983	0.2769
NB1(141.98069243, 0.02773337) and NB2(170.00191669, 0.02773337)	0.006767	0.0004998

Table 3: Summary of goodness-of-fit tests

On the other hand, the Kolmogorov-Smirnov and Chi-square test results are summarised in Table 3. In the former case, both p-values are greater than 25%. That is to say, we cannot reject the null hypothesis that the observed data agree with the proposed model at 25% significance level. On the contrary, neither tests present p-values greater than 0.01 for the second example. Therefore, we could conclude the second model does not fit the data well even at 1% significance level.

5.2.2 Claim severity component

We are now ready to implement the steps of EM algorithm for estimating parameters in the claim severity model. Before proceeding to the EM steps, let us look into the parameter ν. In fact, ν is connected to the parameters in the claim frequency component. Therefore, once the sample of claim counts is given, we would get an estimated value for ν. Recall from previous sections that,

$$\nu = p + (1-p) \cdot \frac{B(\alpha_1 + 1, \alpha_2)}{B(\alpha_1, \alpha_2)} \,. \tag{32}$$

If we substitute in (32) the estimated values of p, α_1, α_2, we could obtain that the point estimation for ν would be $\nu = 0.9039196$. ν could then be treated as a constant onwards, and we follow an iteration procedure similar to that presented in the previous subsection. As follows,

1. We begin with an initially determined parameter values $\vartheta^{(0)} = \{\mu^{(0)}, \delta^{(0)}, \sigma^{(0)}\}$.

2. E-Step

   For the $(l+1)$-th iteration, $l = 0, 1, \ldots$, we first seek for the expected value of the random value U_i, representing the latent variable, conditional on the observations together with the current parameter estimates $\vartheta^{(l)}$, i.e., estimates from the l-th iteration. Let τ_i denote this expectation

$$\begin{aligned} \tau_i &= \mathbb{E}[U_i|Y, \vartheta^{(l)}] = 1 \times \mathbb{P}\{U_i = 1|Y_i, \vartheta^{(l)}\} + 0 \times \mathbb{P}\{U_i = 0|Y_i, \vartheta^{(l)}\} \\ &= \frac{\mathbb{P}\{U_i = 1, Y_i|\vartheta^{(l)}\}}{\mathbb{P}\{Y_i, \vartheta^{(l)}\}} = \frac{\nu\mu^{(l)}e^{-\mu^{(l)}y_i}}{\nu\mu^{(l)}e^{-\mu^{(l)}y_i} + (1-\nu)\frac{\delta^{(l)}\sigma^{(l)\delta^{(l)}}}{(\sigma^{(l)}+y_i)^{\delta^{(l)}+1}}} \end{aligned}$$

   Subsequently, we compute the expectation of its log-likelihood function with respect to the conditional distribution of U given Y under the current estimates $\vartheta^{(l)}$. Considering that we have m^*

independent observations, we have,

$$
\begin{aligned}
Q(\vartheta|\vartheta^{(l)}) &= \mathbb{E}_{U_i|Y_i,\vartheta^{(l)}}[\mathcal{L}(\vartheta|Y_i,U_i)] \qquad (33)\\
&= \sum_i^{m^*}\tau_i\left[\log\nu+\log\mu^{(l)}-\mu^{(l)}y_i\right]\\
&\quad+\sum_i^{m^*}(1-\tau_i)\left[\log(1-\nu)+\log\delta^{(l)}+\delta^{(l)}\log\sigma^{(l)}-(\delta^{(l)}+1)\log(\sigma^{(l)}+y_i)\right]\\
&= \log\nu\sum_i\tau_i+\log\mu^{(l)}\sum_i\tau_i-\mu^{(l)}\sum_i\tau_iy_i+\log(1-\nu)\sum_i(1-\tau_i)\\
&\quad+\log\delta^{(l)}\sum_i(1-\tau_i)+\delta^{(l)}\log\sigma^{(l)}\sum_i(1-\tau_i)-(\delta^{(l)}+1)\sum_i(1-\tau_i)\log(\sigma^{(l)}+y_i).
\end{aligned}
$$

3. M-Step

   The maximisation step is set to find the parameter values that maximises (33) and they are the estimates to be used in the next iteration, i.e.,

$$
\vartheta^{(l+1)}=\underset{\vartheta}{\operatorname{argmax}}\,Q(\vartheta|\vartheta^{(l)}) \qquad (34)
$$

   In order to do this, we take the gradient of $Q(\vartheta|\vartheta^{(l)})$, equate to zero and solve the system for $\{\mu,\delta,\sigma\}$ simultaneously.

$$
\begin{aligned}
\frac{\partial Q}{\partial\mu} &= \frac{\sum_i\tau_i}{\mu}-\sum_i\tau_iy_i=0;\\
\frac{\partial Q}{\partial\delta} &= \frac{n-\sum_i\tau_i}{\delta}+n\log\sigma-\log\sigma\sum_i\tau_i-\sum_i(1-\tau_i)\log(\sigma+y_i)=0;\\
\frac{\partial Q}{\partial\sigma} &= \frac{(n-\sum_i\tau_i)\delta}{\sigma}-(\delta+1)\sum_i\frac{1-\tau_i}{\sigma+y_i}=0\,.
\end{aligned}
$$

   The estimates for the subsequent iteration $(l+1)$-th are the solutions to these equations. It is obvious that μ is independent from other parameters and could be solved directly with an explicit representation

$$
\mu^{(l+1)}=\frac{\sum_i\tau_i}{\sum_i\tau_iy_i}.
$$

   For the other two parameters, we could only solve numerically. Again we apply the *nleqslv* package in R. As a consequence, we can obtain estimated parameters at this iteration $\vartheta^{(l+1)}=\{\mu^{(l+1)},\delta^{(l+1)},\sigma^{(l+1)}\}$.

4. Plug $\vartheta^{(l+1)}$ into the $(l+2)$-th iteration and repeat the above steps until convergence.

As we said, we do not have the claim severity data corresponding to the claim count one, we simulated those based on the claim counts data we used before. Then we have for claim counts in each period followed by a sequence of claim sizes randomly generated according to (9) when $f(\cdot)$ and $g(\cdot)$ take $Exponential(\mu)$ and $Pareto(\delta,\sigma)$ forms, respectively.

Parameters	μ	δ	σ
Predefined	1	2	0.5
Initial value	1.5	2.5	0.2
After 1 iteration	1.0018065	1.572045	0.1476012
After 10 iterations	0.9918480	2.011409	0.4177998
After 110 iterations (at convergence)	0.9999930	1.957350	0.4856047

Table 4: Performance of estimation using EM algorithm on simulated claim severities without ν

Then, we implemented the algorithm and were able to achieve the desired parameters within reasonable amount of iterations. In Table 4 we show our estimates against the predefined parameters. This verifies the effectiveness of our algorithm.

As an alternative, we we have the option to estimate ν directly from claim size data. Implementing the EM algorithm to find this estimate $\hat{\nu}$ is very similar to that we applied to p in Subsection 5.2.1. Recall that the partial derivative equation with respect to p in that subsection is independent from all other variables. A similar situation exists here for ν. We simply need to introduce one additional equation respecting ν estimation, note that it does not affect the remainder equations. As so, we find

$$\hat{\nu} = \frac{\sum_i \tau_i}{m^*}. \tag{35}$$

After running the EM algorithm for this construction, we obtain results in Table 5 and we can compare them with the "true" values.

Table 5: Performance of estimation using EM algorithm on simulated claim severities with ν

Parameters	μ	δ	σ	ν
Predefined	1	2	0.5	0.9039196
Initial value	1.5	2.5	0.2	0.9
After 1 iteration	1.000203	1.596950	0.1515525	0.9311160
After 100 iterations	1.005289	1.819941	0.3791028	0.9196332
After 1552 iterations (at convergence)	0.9991517	2.012381	0.4928861	0.904398

Notice that we have input the predefined value estimate for ν as the estimate from the claim count data. In this way, we could further consider the value estimated from this method, i.e., 0.904398 from Table 5, to be a secondary estimate for ν. Iteration reached convergence though at a much lower speed than before (intuitively, we could expect this). It still led to good estimation of all the parameters including ν, converging closely to its "true" value, eventually.

5.3 A short discussion

To conclude, on the model and estimation procedure, this manuscript has carried out a develoment and and subsequent parameter estimation for the so-called unforeseeable risks discussed in [6]. Precisely, for these risks, a probability p has been assigned at mass point $\{0\}$ so that their corresponding counting process is distinguished from the classical one. Since we could only observe the entire claim counts from an insurance portfolio, this missing information could be estimated using the EM algorithm. Under certain assumptions for the distribution of heterogeneities within the portfolio (which was denoted as

Λ in this work), we could derive the randomness of the total claim counts given a fixed period to be a Negative Binomial mixture distribution. Thus, the likelihood function could be presented explicitly with which EM algorithm was implemented properly, based on a set of claim counts data for the third party liability insurance portfolio. However, the resulting estimates seem to be very sensitive to chosen initial values. Hence, we employed starting values which were computed via method of moments. Both the Kolmogorov-Smirnov and Chi-square tests suggest a good fit on the observed data.

We added to the initial model a study on missing information on the claim severity, assuming that it could bring some information, may be duplicate, on the unforeseeable stream. We could also apply a similar estimation procedure to that of the claim count one.

According to the dataset considered here, the estimated probability of non-occurrence of claims per period is approximately 60%. This is surprisingly larger than what we would expect, and we can argue about it. It may be due to the quality of the data set of course, but also note that it corresponds to the probability for a fixed time period which is a quarter here. Nevertheless, it may show that many policies have not filed their claims to the system, yet. Also, Similar to the concerns usually put into the right tail extreme, i.e, extreme value theory, it could also be interesting to devote some particular atention to the left extreme. The latter one is what an insurer deals with on a daily basis, it could be related with the "Bonus Hunge" problem, common in motor insurance. It could be quite dangerous when someone does not show any claims before a large claim. This could be many small scratches unreported, leaving no history, that lead to a big accident. To detect the potential of policyholders moving from left towards right end tail would probably be more secure for an insurer. It also applies to introduction of new policies covering autonomous cars for instance where initial information is missing and accurate estimation of risks is needed. This study, however, serves as a starting tool to help identifying some of these risks.

6 Combined parameter estimation and global likelihood

In our model we could consider the calculation of Bayesian premia separately for claim counts and respective severities since we built a model where there is stochastic independence between these two quantities. However, it would be nice to compute a premium where we estimate parameters $\vartheta := (\alpha_1, \alpha_2, \beta, \mu, \delta, \sigma, p)$ altogether, using a likelihood where N and Y are jointly distributed. We will refer to it as the global likelihood function and it will serve as a direct extension of the current work.

In general, a random sample will be composed by a sequence of m independent pairs of dependent observervations $(N_i, \overrightarrow{Y_i})$, where N_i represents the number of claims in the i-th period and $\overrightarrow{Y_i} = (Y_{i1}, \ldots, Y_{in_i})$ is the corresponding sequence of claim severities. It is clear that if we consider in general that both claim counts and severities may bring information about each stream, foreseable or unforeseable, we should consider that stochastic dependence between N_i and corresponding $\overrightarrow{Y_i}$ is present. However, they are conditionaly independent, given $\Lambda = \Lambda_i$ $(i = 1, 2)$ that is, for each individual stream we can consider the classical asumption of independence between claim counts and severities.

Observed values are represented by corresponding lower case letters $(n_i, \overrightarrow{y_i})$. We also note that in each pair the dimension of vector $\overrightarrow{y_i}$ depend on the observed n_i. Now, we could write a likelihood function considering the joint random vector $(N, \overrightarrow{Y})$ assuming we have m groups of observations and conditional independence of each Y_{ij}, $j = 1, \ldots, n_i$, for a given i:

$$L(\vartheta|\mathbf{n}, \overrightarrow{y}) = \prod_{i=1}^{m} f_{N,\overrightarrow{Y}}(n_i, y_{i1}, \ldots, y_{in_i}) \quad = \quad \prod_{i=1}^{m} f_{\overrightarrow{Y}|N}(\overrightarrow{y_i}|n_i)\mathbb{P}(n_i) = \prod_{i=1}^{m} \left\{ \prod_{j=1}^{n_i} f_{\overrightarrow{Y}|N}(y_{ij}|n_i) \right\} \mathbb{P}(n_i)\,,$$

where $\mathbb{P}(n_i) = \mathbb{P}(N = n_i)$, for simplification.

For our model, we can write the corresponding global log-likelihood function, denoted as $\mathcal{L}(\vartheta|\mathbf{n}, \overrightarrow{y})$. If we let claim counts conform to Lemma 1 and claim severities follow (9), i.e. $f(\cdot)$ and $g(\cdot)$ are respectively Exponential and Pareto densities, we have

$$\begin{aligned}
\mathcal{L}(\vartheta|\mathbf{n}, \overrightarrow{y}) &= \sum_{i=1}^{m} \log \mathbb{P}(n_i) + \sum_{i=1}^{m}\sum_{j=1}^{n_i} \log f_{\overrightarrow{Y}|N}(\overrightarrow{y_i}|n_i) \\
&= \sum_{i=1}^{m} \log \left\{ p\binom{n+\alpha_1-1}{n}\left(\frac{\beta}{\beta+1}\right)^{\alpha_1}\left(\frac{1}{\beta+1}\right)^{n} \right. \\
&\qquad \left. +(1-p)\binom{n+\alpha_1+\alpha_2-1}{n}\left(\frac{\beta}{\beta+1}\right)^{\alpha_1+\alpha_2}\left(\frac{1}{\beta+1}\right)^{n} \right\} \\
&\qquad + \sum_{i=1}^{m}\sum_{j=1}^{n_i} \log \left\{ \nu\mu e^{-\mu y_{ij}} + (1-\nu)\frac{\delta\sigma^{\delta}}{(\sigma+y_{ij})^{\delta+1}} \right\}.
\end{aligned}$$

Instead of building a likelihood function based on a sample of observations of the pair $(N, \overrightarrow{Y})$, we can build it using interarrival time and corresponding severity, where each observation is a bivariate pair, denoted as (T_j, Y_j), of dependent random varibles, in general. Let's denote the density and the distribution function of T_j by $\phi(\cdot)$ and $\Phi(\cdot)$, respectively.

For finding the distribution of (T_j) recall Lemma 2 and that, conditional on $\Lambda = \lambda$, $\{N(t)\}$ is a Poisson process with intensity λ. Then, given $\Lambda = \lambda$, conditional interarrival time $T_j|\Lambda = \lambda \frown Exponential(\lambda)$, with mean λ^{-1}. If Poisson parameter is an outcome of a random variable Λ and is distributed as (7), we have that the unconditional distribution of T_j is a mixture of two Pareto distributions whose density is given by

$$\phi(t_j) = p\,\frac{\alpha_1\beta^{\alpha_1}}{(\beta+t_j)^{\alpha_1+1}} + (1-p)\,\frac{(\alpha_1+\alpha_2)\beta^{\alpha_1+\alpha_2}}{(\beta+t_j)^{\alpha_1\alpha_2+1}} \tag{36}$$

On the other hand, the density function of Y_j, conditional on a given $\Xi = \xi$, is given by the mixture, see (8),

$$h_\xi(y_j) = \xi f(y_j) + (1-\xi)g(y_j)\,, \tag{37}$$

and the unconditional density is given by (9). In particular, we assumed in our model that $f \frown exp(\mu)$, $g \frown Pareto(\delta, \sigma)$.

Recall Remark 4, we know that with $\Xi = \xi$ fixed, T_j and Y_j are independent. A random sample of observations are made of independent pairs (Y_j, T_j), each j of dependent Y_i and T_i. Each pair, j, has density function

$$\begin{aligned}
\phi(y_i, t_i) &= \int_{\xi} h_\xi(y_i)\phi(t_i)dA(\xi) \qquad (38)\\
&= (1-p)\int_{\xi\neq 1} h_\xi(y_i)\phi(t_i)dA(\xi) + p\,h_1(y_i)\frac{\alpha_1\beta^{\alpha_1}}{(\beta+t_i)^{\alpha_1+1}}
\end{aligned}$$

where $A(\xi)$ stands for the distribution function of Ξ. Note that that we separated the situations $\xi \neq 1$ and $\xi = 1$. Also, the events are equivalent: $\{\Xi = 1\} \Leftrightarrow \{\Lambda^{(2)} = 0\}$, so that $Pr\{\Xi = 1\} = Pr\{\Lambda^{(2)} = 0\} = p$.

A short discussion on the distributions above, (36) and (37), sounds apropriate. First, we consider the unconditional distribution of the arrival time (unconditional of Λ) however, we considered the conditional of the individual claim size Y, given $\Xi = \xi$. This ratio is given but that only means that

$\Lambda^{(1)}$ is a given proportion of Λ, it remains random, it only means that the distribution of $\Lambda^{(1)}$ is a scaled distribution of that of Λ, or *vice versa*. Looking at the density $h_\xi(y_i)$ we see that the split rate ξ gives the weight for the foreseeable stream claim amount, and looking at the density $\phi(y_j)$ we could think that the probability p gives a similar meaning regarding the claim count stream. It does not seem the case, it is not clear that the second part in the mixture represents the unforeseeable only.

Now, let's back to joint density (38). We build the likelihood function over m^* pairs of observations (m^* may be different from sample size m from random vector $(N, \overrightarrow{Y})$ above):

$$\begin{aligned} L(\vartheta) &= \prod_{i=1}^{m^*} \Phi(y_i, t_i) = \prod_{i=1}^{m^*} \int_\xi h_\xi(y_i)\phi(t_i)dA(\xi) \\ &= \prod_{i=1}^{m^*} \left[(1-p) \int_{\xi \neq 1} \left(\xi f(y_i) + (1-\xi) g(y_i)\right) \phi(t_i) Beta(\xi; \alpha_1, \alpha_2) d\xi + p\, f(y_i)\, \frac{\alpha_1 \beta^{\alpha_1}}{(\beta + t_i)^{\alpha_1+1}} \right], \end{aligned}$$

where $Beta(\xi; \alpha_1, \alpha_2)$ denotes the Beta density function of Ξ, given $\Xi \neq 1$, and $\frac{\alpha_1 \beta \alpha_1}{(\beta+t_i)^{\alpha_1+1}}$ is the density $\phi(t_j)$ when $\xi = 1$. Remark that we cannot interchange the product $(\prod_{i=1}^{m^*})$ and the integral.

References

[1] Hans Bühlmann and Alois Gisler. *A course in credibility theory and its applications*. Springer Science & Business Media, 2006.

[2] Christophe Couvreur. The EM algorithm: A guided tour. In *Computer Intensive Methods in Control and Signal Processing*, pages 209–222. Springer, 1997.

[3] Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (methodological)*, pages 1–38, 1977.

[4] André Dubey. Probabilité de ruine lorsque le paramètre de poisson est ajusté a posteriori. *Mitteilungen der Vereinigung Schweiz Versicherungsmathematiker*, 2:130–141, 1977.

[5] Stuart A Klugman, Harry H Panjer, and Gordon E Willmot. *Loss models: from data to decisions*, volume 715. John Wiley & Sons, 2012.

[6] Bo Li, Weihong Ni, and Corina Constantinescu. Risk models with premiums adjusted to claims number. *Insurance: Mathematics and Economics*, 65:94–102, 2015.

[7] Eugene Lukacs. A characterization of the gamma distribution. *The Annals of Mathematical Statistics*, 26(2):319–324, 1955.

[8] Richard A Redner and Homer F Walker. Mixture densities, maximum likelihood and the em algorithm. *SIAM review*, 26(2):195–239, 1984.

[9] Waimo team, 2016.