

Monte Carlo Methods for Insurance Risk Computation

Shaul K. Bar-Lev¹ & Ad Ridder²

¹ Department of Statistics, University of Haifa, Haifa, Israel

² School of Business and Economics, Vrije Universiteit, Amsterdam, Netherlands

Correspondence: Shaul Bar-Lev, Department of Statistics, University of Haifa, Haifa 31905, Israel. Tel: 972-4-8240178. E-mail: barlev@stat.haifa.ac.il

Received: March 24, 2019 Accepted: April 17, 2019 Online Published: April 24, 2019

doi:10.5539/ijsp.v8n3p54 URL: https://doi.org/10.5539/ijsp.v8n3p54

Abstract

In this paper we consider the problem of computing tail probabilities of the distribution of a random sum of positive random variables. We assume that the individual claim variables follow a reproducible natural exponential family (NEF) distribution, and that the random number has a NEF counting distribution with a cubic variance function. This specific modeling is supported by data of the aggregated claim distribution of an insurance company. Large tail probabilities are important as they reflect the risk of large losses, however, analytic or numerical expressions are not available. We propose several simulation algorithms which are based on an asymptotic analysis of the distribution of the counting variable and on the reproducibility property of the claim distribution. The aggregated sum is simulated efficiently by importance sampling using an exponential change of measure. We conclude by numerical experiments of these algorithms, based on real car insurance claim data.

Keywords: aggregated claim distributions, natural exponential families, reproducibility, Tweedie scale, variance functions, discrete counting variables, Monte Carlo simulations, importance sampling

1. Introduction

Let Y_1, Y_2, \dots be i.i.d. positive random variables representing the individual claims at an insurance company, and let $N \in \mathbb{N}_0 = \{0, 1, \dots\}$ designate the total number of claims occurring during a certain time period, where N and the Y_i 's are independent. For modeling convenience, we define $Y_0 \equiv 0$. The total claimed amount is called the aggregated claim variable, denoted by

$$S_N = \sum_{k=0}^N Y_k.$$

A major issue for insurance companies is the uncertainty of the occurrence of a large aggregated claim, because, if this happens, the company faces large losses that may ultimately lead to a ruin. Thus, an important quantity to compute is the insurance risk factor

$$\ell(x) = \mathbb{P}\left(\sum_{k=0}^N Y_k > x\right), \quad (1)$$

for large levels x . (Other risk/loss factors or models can be found in the fundamental monograph by Klugman et al.(2008)). Because of its importance for insurance companies, many actuarial studies deal with this problem, see the monograph of Kaas et al. (2008). However, there are many other practical situations in which the object of interest is a random sum of i.i.d. random variables (Bahnemann, 2015). For instance, S_N might represent the total loss of a financial institute due to defaults of N obligors with credit sizes Y_1, Y_2, \dots

Two rather moderate remarks should be noted at this stage. The first relates to another approach of dealing with the random aggregated sum. Such an approach considers $N = N(t)$ to form a renewal counting process up to time t , in which case the aggregated sum variable becomes a renewal risk model (Dickson, 1998; Asmussen and Albrecher, 2010). The second remark relates to the notion of reproducibility or reproductive distributions. Indeed, denote by $f(n)$ the probability mass function (p.m.f.) of N and by G the common cumulative distribution function (c.d.f.) of the Y_k 's. Then the c.d.f. of S_N is

$$\sum_{n=0}^{\infty} f(n)G_n^*(s), \quad s \geq 0,$$

where G_n^* stands for the n -fold convolution of G with itself. In the case where G is the gamma distribution then the convolution of i.i.d. gamma variates is also gamma, a fact which led to an easy-to-calculate exact formula for the aggregate distribution function (Bahnemann, 2015, p. 113). However, as opposed to Bahnemann's claim that "this desirable

reproductive property—the distribution of a sum of identical independent random variables having the same distribution type as the components—is shared by just a few families of distributions (notably the normal distributions, which are not generally useful as claim-size distributions)”, is not entirely adequate as the class of reproducible families is huge and contains all natural exponential families (NEF’s) having a power variance function that belong to the Tweedie’s scale (Tweedie, 1984; Bar-Lev and Enis, 1986; Bar-Lev and Cassalis, 2003). In particular all NEF’s generated by stable distributions are reproductive, though stable distributions by themselves have complicated analytical forms (see in the sequel). As a consequence, ”actuaries have since the mid-1900s sought to develop various procedures for calculating values of an aggregate distribution. Among these are several approximations using easily calculable parametric distributions, algorithms featuring recursive formulas, Fourier-transform-based methods, and Monte Carlo simulation” (Bahnmann, 2015, p. 115). Also, see Vinogradov et al. (2012, 2013) regarding the case of positive stable distributions.

Now, for doing the actual computations of the risk factor in (1), one needs to fit a model for the counting distribution of N and the claim size distribution of Y . Nowadays we see that the Poisson and the gamma distributions, respectively, are often being used (Bowers et al., 1997). Other proposals include negative binomial for the counting number and inverse Gaussian for the claim size and positive stable distributions.

However, due to large uncertainties, many realistic data show large overdispersion. In fact, our study is motivated by available data of a car insurance company for which the traditional distributions clearly do not fit properly. The (empirical) variance of the counting number data shows a power law with respect to the (empirical) mean, with a power close to three. This observation was the reason that we decided to consider counting distributions with tails that go beyond (are heavier than) the Poisson and negative binomial. A natural modeling technique to introduce families of distributions is by considering the concept of NEF’s (Letac and Mora, 1990; Smyth and Jorgensen, 2002; Dunn and Smyth, 2005, 2008). In our case we are interested in NEF’s with cubic variance functions (Letac and Mora, 1990). Concerning the counting variable N , we shall investigate

- The NEF generated by the Abel distribution (hereafter the Abel distribution);
- the NEF generated by the strict arcsine distribution (hereafter the strict arcsine distribution);
- the NEF generated by the Takács distribution (hereafter the Takács distribution).

These are new distributions for insurance modeling, and have to our knowledge not been considered before in computation and simulation studies. As said above, our objective is to execute numerical computations of the insurance risk factor, for which we consider using Monte Carlo simulations. This was motivated because there are no easily workable analytic expressions available for the probability functions of the Abel, strict arcsine arcsine and Takács distributions. Thus, a main part of our paper deals with developing simulation algorithms for generating samples from these distributions.

Also concerning the claim size distributions, we propose modeling by NEF’s. Specifically, we consider

- gamma distribution;
- positive stable distributions;
- inverse Gaussian distribution.

These are well-known distributions in insurance modeling, for which simulation algorithms for generating samples have been established (Shuster, 1968; Chambers et al., 1976; Michael et al., 1976; Devroye, 1986). However, please note that when we write positive stable distributions (for which no integer moments exist) we mean NEF’s generated by positive stable distributions (for which all moments exist).

In this way, our aggregate models become Tweedie models in the sense that both the distributions of the counting number and the distributions of the claim size belong to NEF’s (Smyth and Jorgensen, 2002; Dunn and Smyth, 2005, 2008). Hence, we shall investigate whether the statistical procedures for estimating the parameters in these models can be applied to our data, or whether we need to develop other procedures. Commonly, one models the mean and dispersion in terms of risk factors, for instance by regression models or by generalized linear models (Smyth and Jorgensen, 2002). However, we propose to directly compute the risk as a tail probability of the aggregated claim distribution by executing Monte Carlo simulations.

The simulation algorithm exploits two efficiency improvements with respect to standard Monte Carlo. Firstly, the claim size distributions show the reproducibility property (Bar-Lev and Enis, 1986), which says that convolutions can be considered being affine transformations of univariates. Thus, for example, a single sample of the inverse Gaussian distribution suffices for generating a sum of i.i.d. inverse Gaussian. Secondly, we apply importance sampling by implementing a change of measure which is based on the exponentially tilting the probability distributions (Asmussen and Glynn, 2007(@, Chapter VI). The optimal tilting factor is determined by a saddle-point equation, and results in a logarithmically efficient estimator.

The paper is organized as follows. Section 2 summarizes the concepts of Tweedie NEF distributions and reproducibility. The main contribution of the paper is contained in Section 3 where we analyze the three counting distributions which leads to the construction of the simulation algorithms for generating samples. Section 4 summarizes a few aspects of the claim distributions. The aggregated claim risks are computed in Section 5 by Monte Carlo simulation using the algorithms that we have developed. We show how these risks for large levels can be computed efficiently by an appropriate change of measure for importance sampling. Finally, Section 6 gives details of the data that motivated this work. Briefly, Section 6 demonstrates that the best fit for the data of a Swedish claims at a car insurance company is obtained for the pair (arcsine, positive stable) with p -value equals .7460. All fit ranking after are, respectively, (arcsine, inverse Gaussian, p -value 0.4224), (Takács, gamma, p -value 0.4159), (Abel, positive stable, p -value 0.3089), (Takács, inverse Gaussian, p -value 0.2800), (Takács, positive stable, p -value 0.2701), (Abel, inverse Gaussian, p -value 0.2459) and (Abel, gamma, p -value 0.2101). As opposed to these, the worst fit has been obtained for pairs of the Poisson along with the gamma, inverse Gaussian and positive stable distributions with p -value less than .00001, see Table 3 of Section 6 for a reference.

2. Natural Exponential Family and Reproducibility

We summarize some relevant concepts and properties of distributions NEF's, see Letac and Mora (1990); Smyth and Jorgensen (2002); Dunn and Smyth (2005, 2008).

Definition 1. Let ν be a non-Dirac positive Radon measure on \mathbb{R} , and $L(\theta) = \int e^{\theta x} \nu(dx)$ its Laplace transform. Assuming that $\Theta = \text{int}\{\theta : L(\theta) < \infty\} \neq \emptyset$, then the NEF generated by ν is defined by the probability distributions

$$\mathcal{F} = \left\{ F_\theta : F_\theta(dx) = e^{\theta x - \kappa(\theta)} \nu(dx), \theta \in \Theta \right\}, \tag{2}$$

where $\kappa(\theta) = \log L(\theta)$, the cumulant transform of ν , is strictly convex and real analytic on Θ .

The generating measure ν is called the kernel of the NEF. We may associate a random variable X_θ with the NEF distribution F_θ . Then

$$\mathbb{E}[X_\theta] = \kappa'(\theta); \quad \text{Var}[X_\theta] = \kappa''(\theta).$$

Since κ' is invertible, we obtain the NEF parameter θ by

$$\theta = \theta(m) = (\kappa')^{-1}(m).$$

This means that if we let $\kappa(m) = \kappa(\theta(m))$, we can represent the NEF equivalently by

$$\mathcal{F} = \left\{ F_m : F_m(dx) = e^{\theta(m)x - \kappa(m)} \nu(dx), m \in \mathcal{M} \right\},$$

where \mathcal{M} is called the mean domain of \mathcal{F} and is given by $\mathcal{M} = \kappa'(\Theta)$. Such a parametrization is called the mean value parametrization of the NEF \mathcal{F} . Finally, if also the variance $V(m) = \kappa''(\theta(m))$ of an NEF distribution is given as function of the mean m , the pair (V, \mathcal{M}) uniquely determines an NEF within the class of NEF's. Simple algebra shows that θ and $\kappa(\theta)$ can be represented in terms of m by

$$\begin{aligned} \theta_A(m) &= \int \theta'(m) dm = \int \frac{1}{V(m)} dm + A, \\ \kappa_B(m) &= \int \kappa'(m) dm = \int \frac{m}{V(m)} dm + B, \end{aligned} \tag{3}$$

where A and B are constants. The constants A and B need to be chosen appropriately such that the corresponding F_θ is a probability distribution. A thorough survey on this issue as well on the mean value parameterization of NEF's can be found in Bar-Lev and Kokonendji (2017).

We call an NEF a Tweedie NEF when V has the form $V(m) = \alpha m^\gamma$, $\alpha > 0$, $\gamma \in \mathbb{R} \setminus \{(0, 1)\}$; i.e., when V is a power function of the mean (Tweedie, 1984; Bar-Lev and Enis, 1986; Jorgensen, 1987). Furthermore, the following reproducibility concept will be a key element in our analysis. It has been developed in (Bar-Lev and Enis, 1986).

Definition 2. Let \mathcal{F} be a NEF as in (2), and suppose that $X_1, X_2, \dots \stackrel{iid}{\sim} F_\theta$. Denote $S_n = \sum_{k=1}^n X_k$. The NEF is said to be reproducible if there exist a sequence of real numbers $(c_n)_{n \geq 1}$, and a sequence of mappings $\{g_n : \Theta \rightarrow \Theta, n \geq 1\}$, such that for all $n \in \mathbb{N}$ and for all $\theta \in \Theta$,

$$c_n S_n \stackrel{\mathcal{D}}{\sim} F_{g_n(\theta)} \in \mathcal{F}.$$

Bar-Lev and Enis (1986) and Bar-Lev and Cassalis (2003) showed that an NEF \mathcal{F} is reproducible iff it belongs to the Tweedie scale. They also provided the corresponding explicit forms of c_n and $g_n(\theta)$ for all NEF's belonging to the Tweedie scale (see in the sequel).

3. Counting Distributions

In this section we analyze discrete counting NEF's that are given by a cubic variance function (VF), see Letac and Mora (1990). As said in the introduction, we are motivated by data in a case study having a variance showing indeed such a power law. Our distributions will be used for computing the insurance risk factor by simulations, and, thus, the issue is how to generate samples from these distributions. Our analysis will lead to the construction of sampling algorithms that are based on the accept-reject method. As dominating proposal distribution, we can use the same distribution that is used to sample from the Zipf distribution (Devroye, 1986).

Consequently, we analyze the Abel, the arcsine, and the Takács NEF's whose kernels and VF's are given in Letac and Mora (1990). For each NEF we introduce the VF, develop relevant asymptotic, and then propose our simulation procedure. It should be noted at this point, however, that the Abel distribution is also commonly known as the generalized Poisson distribution - see Consul (1989); Consul and Shoukri (1988), whereas the Takács distribution is known as the generalized negative binomial distribution - c.f., Devroye (1992).

3.1 The Abel NEF

The VF is given by

$$V(m) = m \left(1 + \frac{m}{p}\right)^2, \quad m > 0, p > 0. \tag{4}$$

By (3) we deduce that the NEF-parameter function $\theta_A(m)$ and the cumulant function $\kappa_B(m)$ as functions of mean m are derived by

$$\theta_A(m) = \int \frac{1}{V(m)} dm = \int \frac{1}{m \left(1 + \frac{m}{p}\right)^2} dm = \log \frac{m}{m+p} + \frac{p}{m+p} + A \text{ (a constant)} \tag{5}$$

$$\kappa_B(m) = \int \frac{m}{V(m)} dm = \int \frac{1}{\left(1 + \frac{m}{p}\right)^2} dm = -\frac{p^2}{m+p} + B \text{ (a constant)}. \tag{6}$$

The kernel is given in (Letac and Mora, 1990):

$$v(n) = \frac{1}{n!} p(p+n)^{n-1}, \quad n \in \mathbb{N}_0. \tag{7}$$

Proposition 1. $A = -1$ and $B = p$.

Proof. The constant B follows from (Letac and Mora, 1990, Proposition 4.4):

$$v(0) = e^{\kappa_B(0)} \Leftrightarrow 1 = e^{-p+B} \Leftrightarrow B = p.$$

Hence,

$$\kappa_B(m) = -\frac{p^2}{m+p} + p = \frac{mp}{m+p} \Rightarrow \frac{\kappa_B(m)}{p} = \frac{m}{m+p}. \tag{8}$$

The constant A follows from (Letac and Mora, 1990, Example C). The Abel distribution has generating function

$$g(z) = e^z = \sum_{n=0}^{\infty} \frac{1}{n!} z^n,$$

and the kernel is computed by

$$v(n) = \frac{p}{p+n} \frac{1}{n!} \left(\frac{d}{dz}\right)^n g^{n+p}(z) \Big|_{z=0}.$$

In this we get (7). Furthermore, display (4.27) in Letac and Mora (1990) says

$$e^{\theta_A(m)+\kappa_B(m)/p} = g^{-1}(e^{\kappa_B(m)/p}).$$

Substituting $g^{-1} = \log$ and the expression for $\kappa_B(m)/p$ in display (8), we get:

$$e^{\theta_A(m)+m/(m+p)} = m/(m+p) \Leftrightarrow \theta_A(m) = \log \frac{m}{m+p} - \frac{m}{m+p} = \log \frac{m}{m+p} + \frac{p}{m+p} - 1.$$

Conclusion $A = -1$. □

Define

$$\begin{aligned} \nu_0(n) &= \nu(n)e^{-n-p} = \frac{pe^{-p}}{n!}(p+n)^{n-1}e^{-n} \\ \theta(m) &= \theta_A(m) - A = \log \frac{m}{m+p} + \frac{p}{m+p} \\ \kappa(m) &= \kappa_B(m) - B = -\frac{p^2}{m+p} \end{aligned}$$

Then, the NEF Abel counting probability mass function (p.m.f.) of the associated counting variable N_θ is:

$$\mathbb{P}(N_\theta = n) = f_\theta(n) = \nu_0(n) e^{n\theta(m)-k(m)}. \tag{9}$$

Conveniently, we omit the NEF parameter θ in our notations when there is no confusion.

3.1.1 Analysis

First we consider an asymptotic of the modified kernel $\nu_0(n)$, using the Stirling approximation:

$$n! \sim \sqrt{2\pi n} (ne^{-1})^n, \quad n \rightarrow \infty,$$

where \sim means that the ratio converges to 1 (for $n \rightarrow \infty$). This gives

$$\nu_0(n) \sim \frac{pe^{-p}(p+n)^{n-1}e^{-n}}{\sqrt{2\pi n}(ne^{-1})^n} = \frac{pe^{-p}}{\sqrt{2\pi}} \frac{1}{n\sqrt{n}} \left(1 + \frac{p}{n}\right)^{n-1} \sim \frac{p}{\sqrt{2\pi}} \frac{1}{n\sqrt{n}}.$$

The right-hand side shows correspondence with a Zipf distribution (Devroye, 1986):

$$z(n) = \frac{1}{\zeta(3/2)} \frac{1}{n\sqrt{n}}, \quad n = 1, 2, \dots,$$

where $\zeta(\cdot)$ is the Riemann zeta function. Sampling from this Zipf distribution is done by an accept-reject algorithm using the dominating p.m.f. $b(n)$ of the random variable $\lfloor U^{-2} \rfloor$, where U is the random variable uniformly distributed on $(0, 1)$.

$$b(n) = \frac{1}{\sqrt{n+1}} \left(\sqrt{1 + \frac{1}{n}} - 1 \right), \quad n \in \mathbb{N}. \tag{10}$$

The multiplication factor for $z(n) \leq cb(n)$ is (Devroye, 1986)

$$c = \frac{\sqrt{2}}{\zeta(3/2)(\sqrt{2}-1)}.$$

We show that we can use $b(n)$ also as proposal dominating p.m.f. for our NEF p.m.f. $f(n)$. However, because the domains of $b(n)$ and $f(n)$ differ (\mathbb{N} versus \mathbb{N}_0), we need a minor tweak. Denote the conditional p.m.f. by $f(n|n \geq 1) = \mathbb{P}(N = n|N \geq 1)$.

Lemma 1. *There is a constant C (not dependent on n) such that*

$$f(n|n \geq 1) \leq Cb(n), \quad n = 1, 2, \dots$$

Proof. Note that $f(n|n \geq 1) = f(n)/(1 - f(0))$, with $f(0) = \nu_0(0)e^{-\kappa(m)} = e^{-p-\kappa(m)}$. Furthermore, for $n \geq 1$ using the lower bound $n! \geq \sqrt{2\pi n} n^n e^{-n}$, we get:

$$\begin{aligned} \frac{(p+n)^{n-1} e^{-n}}{n!} &\leq \frac{(p+n)^{n-1}}{\sqrt{2\pi n} \sqrt{n} n^{n-1}} \\ &= \frac{1}{\sqrt{2\pi n} \sqrt{n}} \left(1 + \frac{p}{n}\right)^{n-1} \leq \frac{1}{\sqrt{2\pi n} \sqrt{n}} e^p. \end{aligned}$$

Moreover, clearly

$$\theta(m) = \log \frac{m}{m+p} + \frac{p}{m+p} \leq 0.$$

Taking all together we conclude,

$$\begin{aligned} f(n|n \geq 1) &= \frac{pe^{-p}}{(1-f(0))n!} (p+n)^{n-1} e^{-n} e^{n\theta(m)-\kappa(m)} \leq \frac{pe^{-\kappa(m)}}{(1-f(0))\sqrt{2\pi n} \sqrt{n}} \\ &= \frac{\zeta(3/2)pe^{-\kappa(m)}}{(1-f(0))\sqrt{2\pi}} z(n) \leq \frac{\zeta(3/2)pe^{-\kappa(m)}}{(1-f(0))\sqrt{2\pi}} \frac{\sqrt{2}}{\zeta(3/2)(\sqrt{2}-1)} b(n) = \underbrace{\frac{pe^{-\kappa(m)}}{(1-f(0))\sqrt{\pi}(\sqrt{2}-1)}}_{=C} b(n), \end{aligned}$$

where $e^{-\kappa(m)} = e^{p^2/(p+m)}$. □

Remark 1. Note that when $m \gg p$, the constant C is of order p which is reflected in the acceptance probability $1/C$ in the accept-reject sampling algorithm. However, for large dispersion parameters p the larger constant C deteriorates this algorithm. In that case one might improve bounding the kernel and the probabilities. Our case study gave $m \gg p$, so we decided to implement the bounding as given above. That gave acceptance probability 0.25.

Summarizing, the Monte Carlo algorithm for simulating from the Abel distribution (9) becomes:

```

SAMPLING ALGORITHM FOR NEF ABEL DISTRIBUTION
1: Generate  $U \stackrel{\mathcal{D}}{\sim} \mathbf{U}(0, 1)$ .
2: if  $U < f(0)$  then
3:    $N \leftarrow 0$ .
4: else
5:   repeat
6:     Generate  $U \stackrel{\mathcal{D}}{\sim} \mathbf{U}(0, 1)$ .
7:      $N \leftarrow \lfloor U^{-2} \rfloor$ .
8:     Compute  $P = \frac{f(N|N \geq 1)}{C b(N)}$ .
9:     Generate  $U \stackrel{\mathcal{D}}{\sim} \mathbf{U}(0, 1)$ .
10:  until  $U < P$ 
11: end if
12: return  $N$ .
    
```

3.2 The Arcsine NEF

The VF is given by

$$V(m) = m\left(1 + \frac{m^2}{p^2}\right) = \frac{m}{p^2}(m^2 + p^2), \quad m > 0, p > 0. \tag{11}$$

By (3) we deduce that the NEF-parameter function $\theta_A(m)$ and the log-moment generating function $\kappa_B(m)$ are derived by

$$\theta_A(m) = \int \frac{1}{V(m)} dm = \int \frac{p^2}{m(m^2 + p^2)} dm = \ln m - \frac{1}{2} \log(m^2 + p^2) + A = -\frac{1}{2} \log(1 + (p^2/m^2)) + A; \tag{12}$$

$$\kappa_B(m) = \int \frac{m}{V(m)} dm = \int \frac{1}{(1 + \frac{m^2}{p^2})} dm = p \arctan(m/p) + B. \tag{13}$$

The kernel is given in (Letac and Mora, 1990):

$$\nu(2n) = \frac{1}{(2n)!} \prod_{i=0}^{n-1} ((2i)^2 + p^2)$$

$$\nu(2n + 1) = \frac{p}{(2n + 1)!} \prod_{i=0}^{n-1} ((2i + 1)^2 + p^2).$$

Proposition 2. $A = 0$ and $B = 0$.

Proof. The constant B follows from (Letac and Mora, 1990, Proposition 4.4):

$$\nu(0) = e^{\kappa_B(0)} \Leftrightarrow 1 = e^B \Leftrightarrow B = 0.$$

Hence,

$$\kappa_B(m) = p \arctan(m/p) \Rightarrow \frac{\kappa_B(m)}{p} = \arctan(m/p). \tag{14}$$

The generating function of the arcsine kernel is (Letac and Mora, 1990, Example C),

$$f(z) = \sum_{n=0}^{\infty} \nu(n)z^n = e^{p \arcsin z}.$$

Because, $\kappa(\theta) = \log f(e^\theta)$, and $\kappa_B(m) = \kappa(\theta_A(m))$, we get

$$\begin{aligned} \kappa_B(m) &= \log e^{p \arcsin e^{\theta_A(m)}} = p \arcsin e^{\theta_A(m)} \\ &= p \arcsin e^{-\log \sqrt{1+(p^2/m^2)+A}} = p \arcsin \frac{e^A}{\sqrt{1+(p^2/m^2)}} \\ \Rightarrow \sin \frac{\kappa_B(m)}{p} &= \frac{e^A}{\sqrt{1+(p^2/m^2)}}. \end{aligned}$$

According to display (14):

$$\sin \frac{\kappa_B(m)}{p} = \sin \arctan(m/p) = \frac{m/p}{\sqrt{1+(m^2/p^2)}},$$

the last equation a well-known identity of trigonometric functions. Equating:

$$\frac{e^A}{\sqrt{1+(p^2/m^2)}} = \frac{m/p}{\sqrt{1+(m^2/p^2)}} \Leftrightarrow e^A = 1.$$

Conclusion $A = 0$. □

Denote $\theta(m) = \theta_A(m)$, and $\kappa(\theta) = \kappa_B(m)$, with our choices $A = B = 0$. Hence, we get the NEF arcsine counting p.m.f. of the counting variable N , omitting NEF parameter θ in the index notation:

$$\mathbb{P}(N = n) = f(n) = \nu(n) e^{n\theta(m)-k(m)}, \quad n \in \mathbb{N}_0. \tag{15}$$

3.2.1 Analysis

In Appendix A we show that there is a constant K such that for $n = 1, 2, \dots$

$$\nu(2n) \leq K \frac{1}{n \sqrt{n}}.$$

Thus, for these even terms we recognize again the Zipf distribution. This will be helpful to find a dominating proposal distribution.

Lemma 2. Define the double Zipf dominating distribution $b_2(n), n = 2, 3, \dots$ by

$$b_2(2n) = b_2(2n + 1) = \frac{1}{2}b(\lfloor n/2 \rfloor),$$

where $b(n), n = 1, 2, \dots$ is the p.m.f. that dominates the Zipf p.m.f., defined in (10). Then there is a constant C such that

$$f(n|n \geq 2) = \mathbb{P}(N = n|N \geq 2) \leq Cb_2(n), \quad n = 2, 3, \dots$$

Proof. The NEF parameter satisfies

$$\theta(m) = -\frac{1}{2} \log(1 + (p^2/m^2)) \leq 0.$$

Let $A = f(0) + f(1) = \mathbb{P}(N = 0) + \mathbb{P}(N = 1)$. Because $f(n) = \nu(n)e^{\theta(m)n - \kappa(m)}$, we can bound the probabilities $f(2n|n \geq 1) = \mathbb{P}(N = 2n|N \geq 2)$ by

$$f(2n|n \geq 1) = \frac{f(2n)}{1 - A} \leq \frac{Ke^{-\kappa(m)}}{1 - A} \frac{1}{n\sqrt{n}} \leq \tilde{C}b(n),$$

where

$$\tilde{C} = \frac{Ke^{-\kappa(m)}}{1 - A} \frac{\sqrt{2}}{\sqrt{2} - 1}.$$

Then $f(2n|n \geq 1) \leq Cb_2(2n)$ for $C = 2\tilde{C}$.

The constant K depends on a threshold i^* such that for $n > i^*$, $f(2n + 1) < f(2n)$, and for $n \leq i^*$, $\nu(2n + 1) \leq Kn^{-3/2}$ (see Appendix A). Thus also all odd terms satisfy $f(2n + 1|n \geq 1) \leq Cb_2(2n + 1)$. \square

Remark 2. Similarly to our algorithm for sampling from the Abel distribution, also the constant C becomes larger for larger p , deteriorating the accept-reject sampling method. In our implementation we included one more term in the bounding procedure that is described in Appendix A. This gave an acceptance ratio of 0.34.

Summarizing, the Monte Carlo algorithm for simulating from the arcsine distribution (15) becomes:

```

SAMPLING ALGORITHM FOR NEF ARCSINE DISTRIBUTION
1: Generate  $U \stackrel{\mathcal{D}}{\sim} U(0, 1)$ .
2: if  $U < f(0)$  then
3:    $N \leftarrow 0$ .
4: else
5:   if  $U < f(0) + f(1)$  then
6:      $N \leftarrow 1$ .
7:   else
8:     repeat
9:       Generate  $U \stackrel{\mathcal{D}}{\sim} U(0, 1)$ .
10:       $Y \leftarrow \lfloor U^{-2} \rfloor$ .
11:      Generate  $U \stackrel{\mathcal{D}}{\sim} U(0, 1)$ .
12:      if  $U < 0.5$  then
13:         $N \leftarrow 2Y$ .
14:      else
15:         $N \leftarrow 2Y + 1$ .
16:      end if
17:      Compute  $P = \frac{f(N|N \geq 2)}{Cb_2(N)}$ .
18:      Generate  $U \stackrel{\mathcal{D}}{\sim} U(0, 1)$ .
19:    until  $U < P$ 
20:   end if
21: end if
22: return  $N$ .
    
```


3.3 The Takács NEF

The VF is given by

$$V(m) = m\left(1 + \frac{m}{p}\right)\left(1 + \frac{2m}{p}\right), \quad m, p > 0. \tag{16}$$

By (3), we deduce that the NEF-parameter function $\theta_A(m)$ and the log-moment generating function $\kappa_B(m)$ are derived using partial-fraction decomposition:

$$\begin{aligned} \theta_A(m) &= \int \frac{1}{V(m)} dm = \int \left(\frac{1}{m} + \frac{1/p}{1+m/p} - \frac{4/p}{1+2m/p}\right) dm \\ &= \int \left(\frac{1}{m} + \frac{1}{p+m} - \frac{4}{p+2m}\right) dm = \log m + \log(p+m) - 2 \log(p+2m) + A \\ &= \log \frac{m(p+m)}{(p+2m)^2} + A. \\ \kappa_B(m) &= \int \frac{m}{V(m)} dm = \int \left(\frac{-1}{1+m/p} + \frac{2}{1+2m/p}\right) dm \\ &= \int \left(\frac{-p}{p+m} + \frac{2p}{p+2m}\right) dm = -p \log(p+m) + p \log(p+2m) + B \\ &= p \log \frac{p+2m}{p+m} + B. \end{aligned}$$

The kernel is given in (Letac and Mora, 1990):

$$\nu(n) = \frac{p}{n+p} \frac{1}{n!} (n+p)(n+p+1) \cdots (n+p+n-1), \quad n \in \mathbb{N}_0. \tag{17}$$

Proposition 3. $A = 0$ and $B = 0$.

Proof. The constant B follows from (Letac and Mora, 1990, Proposition 4.4):

$$\nu(0) = e^{\kappa_B(0)} \Leftrightarrow 1 = e^B \Leftrightarrow B = 0.$$

Hence,

$$\kappa_B(m) = p \log \frac{p+2m}{p+m} \Rightarrow \frac{\kappa_B(m)}{p} = \log \frac{p+2m}{p+m}. \tag{18}$$

The constant A follows from (Letac and Mora, 1990, Theorem 4.5). The Takács distribution has generating function

$$g(z) = (1-z)^{-1} = \sum_{n=0}^{\infty} z^n,$$

and kernel

$$\nu(n) = \frac{p}{p+n} \frac{1}{n!} \left(\frac{d}{dz}\right)^n g^{n+p}(z) \Big|_{z=0}.$$

In this way we obtain (17). Furthermore, display (4.27) in Letac and Mora (1990) says

$$e^{\theta_A(m)+\kappa_B(m)/p} = g^{-1}\left(e^{\kappa_B(m)/p}\right).$$

Substituting $g^{-1}(y) = (y-1)/y = 1 - (1/y)$ and the expression for $\kappa_B(m)/p$ in display (18), we get:

$$y = e^{\kappa_B(m)/p} = \frac{p+2m}{p+m} \Rightarrow g^{-1}\left(e^{\kappa_B(m)/p}\right) = 1 - \frac{p+m}{p+2m} = \frac{m}{p+2m},$$

thus,

$$\begin{aligned} e^{\theta_A(m)+\kappa_B(m)/p} &= g^{-1}\left(e^{\kappa_B(m)/p}\right) \\ \Leftrightarrow \theta_A(m) + \log \frac{p+2m}{p+m} &= \log \frac{m}{p+2m} \\ \Leftrightarrow \theta_A(m) &= \log \frac{m}{p+2m} - \log \frac{p+2m}{p+m} = \log \frac{m(p+m)}{(p+2m)^2}. \end{aligned}$$

Conclusion $A = 0$. □

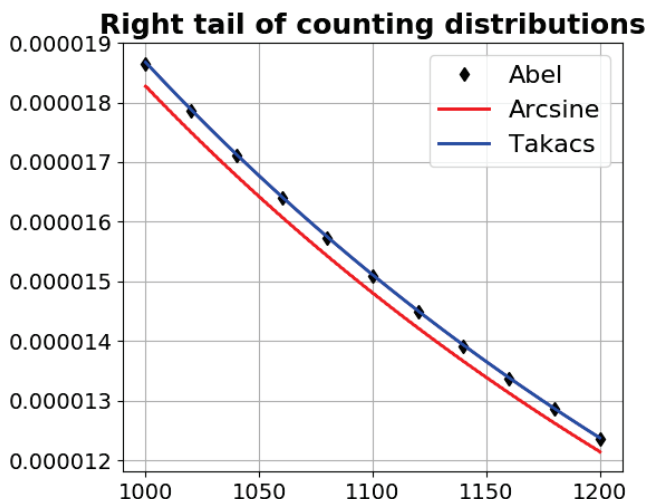


Figure 1. Part of the probability mass functions

Define (with $A = B = 0$)

$$\nu_0(n) = \nu(n)e^{\theta_A(m)n}; \quad \theta(m) = 0; \quad \kappa(m) = \kappa_B(m)$$

The NEF Takács counting p.m.f. of the random variable N becomes

$$\mathbb{P}(N = n) = f(n) = \nu_0(n) e^{-\kappa(m)}. \tag{19}$$

In Appendix B we show that there is a constant K such that

$$\nu_0(n) \leq K \frac{1}{n \sqrt{n}}, \quad n = 1, 2, \dots$$

Lemma 3. *There is a constant C (not dependent on n) such that*

$$f(n|n \geq 1) \leq Cb(n), \quad n = 1, 2, \dots$$

Proof. It follows immediately,

$$\begin{aligned} f(n|n \geq 1) &= \frac{f(n)}{1 - f(0)} = \nu_0(n) \frac{e^{-\kappa(m)}}{1 - f(0)} \leq \frac{K e^{-\kappa(m)}}{1 - f(0)} \frac{1}{n \sqrt{n}} \\ &= \frac{K e^{-\kappa(m)} \zeta(3/2)}{1 - f(0)} z(n) \leq \frac{K e^{-\kappa(m)} \sqrt{2}}{(1 - f(0))(\sqrt{2} - 1)} b(n). \end{aligned}$$

□

The associated Monte Carlo algorithm for generating Takács samples is similar as the Abel algorithm. The acceptance ratio in our case study is 0.23.

3.4 Concluding Remarks

These three counting distributions have tails that are much fatter than the more often used Poisson and negative binomial distributions. As an example, we consider in Section 6 a case study where the data show a mean $m \approx 70$, and the variance $V \approx 52000$, which indicates a power function $V(m) \approx m^r$ with $r \approx 2.5$. This was one of the reasons to consider our specific counting distributions.

An important feature of these distributions is their large tails. Figure 1 shows the probability mass functions for $1000 \leq n \leq 1200$. The Poisson probabilities in this region are virtually zero. The Abel and Takács distributions behave in the tails equivalently, while the arcsine shows slightly lighter tails.

4. NEF Claim Distributions

For modeling the individual claim Y , we consider positive reproducible NEF densities represented by

$$f(y; \theta, p) = f(y) e^{\theta y - \kappa(\theta)}, \quad y > 0.$$

- The gamma NEF given by kernel

$$v(y) = \frac{y^{p-1} e^{-y}}{\Gamma(p)}, \quad y > 0,$$

with dispersion parameter $p > 0$. The VF $V(m) = \frac{m^2}{p}$ yields by Section 2

$$\theta(m) = 1 - \frac{p}{m}; \quad \kappa(m) = p \log \frac{m}{p}.$$

By inversion we get

$$m(\theta) = \frac{p}{1 - \theta}; \quad \kappa(\theta) = p \log \frac{1}{1 - \theta},$$

for $\theta < 1$. Hence,

$$f(y; \theta, p) = \frac{(1 - \theta)^p y^{p-1} e^{-(1-\theta)y}}{\Gamma(p)}.$$

We observe that we actually deal with a gamma distribution with shape parameter p and scale parameter $1 - \theta$, and thus generating samples can be easily done (Devroye, 1986).

Finally, let $Y_1, \dots, Y_n \stackrel{\mathcal{D}}{\sim} f(y; \theta, p)$ i.i.d., and $S_n = \sum_{i=1}^n Y_i$. Then $S_n \stackrel{\mathcal{D}}{\sim} f(y; \theta, np)$. Thus, the gamma NEF is reproducible when considered as a two-parameter NEF with parameters θ and p (Bar-Lev and Enis, 1986).

- The inverse Gaussian NEF given by kernel

$$v(y) = \frac{1}{\sqrt{2\pi p y^3}} e^{-\frac{1}{2py}}, \quad y > 0,$$

with dispersion parameter $p > 0$. The VF $V(m) = pm^3$ yields by Section 2

$$\theta(m) = -\frac{1}{2pm^2}; \quad \kappa(m) = -\frac{1}{pm}.$$

By inversion we get

$$m(\theta) = \frac{1}{\sqrt{-2p\theta}}; \quad \kappa(\theta) = -\sqrt{\frac{-2\theta}{p}},$$

for $\theta < 0$. Hence,

$$f(y; \theta, p) = \sqrt{\frac{1}{2\pi p y^3}} e^{-\frac{1}{2py} + \theta y + \sqrt{\frac{-2\theta}{p}}} = \frac{1}{\sqrt{2\pi p y^3}} e^{\frac{\theta}{y} (y + \sqrt{\frac{1}{-2p\theta}})^2}.$$

Setting

$$\delta = \frac{1}{\sqrt{p}}; \quad \gamma = \sqrt{-2\theta},$$

we recognize the more traditional form of the inverse Gaussian pdf for which a simulation algorithm has been developed (Shuster, 1968; Michael et al., 1976).

Finally, let $Y_1, \dots, Y_n \stackrel{\mathcal{D}}{\sim} f(y; \theta, p)$ i.i.d., and $S_n = \sum_{i=1}^n Y_i$. Then

$$S_n \stackrel{\mathcal{D}}{\sim} c_n f(c_n y; g_n(\theta), p),$$

where

$$c_n = \frac{1}{n^2}; \quad g_n(\theta) = n^2 \theta.$$

See (Bar-Lev and Enis, 1986) for details. Substituting these, we get after algebra,

$$S_n \stackrel{\mathcal{D}}{\sim} f(y; \theta, p/n^2). \tag{20}$$

- The positive stable NEF. Recall that a random variable Y has a stable distribution with index α , denoted $Y \stackrel{\mathcal{D}}{\sim} S_\alpha(\sigma, \beta, \mu)$, if its characteristic function satisfies (for convenience $\alpha \neq 1$):

$$\log \phi(t) = -\sigma^\alpha |t|^\alpha \left(1 - i\beta \text{sign}(t) \tan \frac{\pi\alpha}{2} \right) + i\mu t,$$

for $t \in \mathbb{R}$, where the parameters satisfy

$$\alpha \in (0, 2]; \beta \in [-1, 1]; \mu \in \mathbb{R}; \sigma > 0,$$

see e.g. Samorodnitsky and Taqqu (1994); Nolan (2010). Since we consider positive variables Y , we get the so-called positive α -stable distribution by setting $\alpha \in (0, 1), \beta = 1, \mu \geq 0$. Furthermore we set location parameter $\mu = 0$ in which case

$$\sigma = \left(\cos \frac{\pi\alpha}{2} \right)^{1/\alpha},$$

and the cumulant generating function becomes (Feller, 1971)

$$\kappa(\theta) = -(-\theta)^\alpha, \quad \theta < 0.$$

Both moments of the NEF-distributions $F(Y; \theta, p)$ are finite, whereas these are infinite for the kernel distribution $F(y)$ which is positive α -stable.

Note that with this modeling the p.d.f. $f(y), y > 0$ is only parameterized by index α , but it is not given in explicit form. However, it generates a NEF with a power VF (Tweedie, 1984; Bar-Lev and Enis, 1986; Jorgensen, 1987)

$$V(m) = am^p,$$

where

$$p = \frac{2 - \alpha}{1 - \alpha} > 2; a = (1 - \alpha)\alpha^{1/(\alpha-1)} > 0.$$

Also we obtain the θ and κ function of mean m and index α :

$$\theta(m) = -\left(\frac{m}{\alpha}\right)^{1/(\alpha-1)}; \kappa(m) = -\left(\frac{m}{\alpha}\right)^{\alpha/(\alpha-1)}$$

Thus, given mean m and variance $V(m)$ we compute the parameters θ and α for the NEF distribution with p.d.f.

$$f(y; \theta, \alpha) = f(y) e^{\theta y - \kappa(\theta)}, \quad y > 0.$$

Generating samples from the NEF distribution is done by an accept-reject algorithm, using $f(y)$ as proposal p.d.f. and $C = e^{-\kappa(\theta)} = e^{(-\theta)^\alpha} \geq 1$ as dominating factor. This follows directly from $\theta y \leq 0$. Furthermore, generating from the proposal p.d.f. $f(y)$ is based on (i) generating from $S_\alpha(1, 1, 0)$ distribution by the Chambers algorithm (Chambers et al., 1976), and (ii) the property $S_\alpha(\sigma, 1, 0) \stackrel{\mathcal{D}}{=} \sigma S_\alpha(1, 1, 0)$.

SAMPLING FROM θ -NEF POSITIVE α -STABLE DISTRIBUTION

- 1: **repeat**
- 2: Generate $X \stackrel{\mathcal{D}}{\sim} S_\alpha(1, 1, 0)$.
- 3: $Y \leftarrow \sigma X$.
- 4: Compute acceptance probability $P = e^{\theta Y}$.
- 5: Generate $U \stackrel{\mathcal{D}}{\sim} U(0, 1)$.
- 6: **until** $U < P$
- 7: **return** Y .

Finally, positive α -distributions satisfy the reproducibility property (Bar-Lev and Enis, 1986): let $Y_1, \dots, Y_n \stackrel{\mathcal{D}}{\sim} f(y; \theta, \alpha)$ i.i.d., and $S_n = \sum_{i=1}^n Y_i$. Then

$$S_n \stackrel{\mathcal{D}}{\sim} c_n f(c_n y; g_n(\theta), \alpha),$$

where

$$c_n = n^{-1/\alpha}; g_n(\theta) = \frac{\theta}{c_n} = \theta n^{1/\alpha}.$$

Table 1. Estimates of $\mathbb{P}(S_N > x)$ by Monte Carlo simulation for Abel counting and inverse Gaussian claim distributions

x	M	$\widehat{\ell}$	std. error
5000	9000	1.08e-02	1.09e-03
10000	37000	2.59e-03	2.64e-04
15000	150000	6.47e-04	6.56e-05
20000	410000	2.37e-04	2.40e-05
25000	1020000	9.51e-05	9.66e-06

5. Computing Insurance Risk

The goal of our study is to compute efficiently the tail probability $\ell = \mathbb{P}(S_N > x)$ for large thresholds x , where $S_N = \sum_{j=0}^N Y_j$ is the random sum. We consider Monte Carlo simulation while applying two ideas: (i) reproducibility, and (ii) importance sampling. The reproducibility ensures that given $N = n$ has been generated or observed, we generate S as a single random variable instead of a sum (convolution).

The standard Monte Carlo algorithm is trivial. Let M be the sample size, then the Monte Carlo estimator is

$$\widehat{\ell} = \frac{1}{M} \sum_{i=1}^M \mathbb{I}\{S_{N^{(i)}}^{(i)} > x\},$$

where in the i -th replication, the counting number $N^{(i)}$ is generated from the counting distribution of interest (Abel, arcsine, or Takács), according to the algorithms of Section 3. Given $N^{(i)} = n$, the aggregated claim size $S_n^{(i)}$ is generated from the claim distribution of interest (gamma, inverse Gaussian, or positive α -stable) using the reproducibility property of Section 2. From the observations $\mathbb{I}\{S_{N^{(i)}}^{(i)} > x\}$, $i = 1, \dots, M$, we compute the usual estimator and standard error (or confidence interval) statistics.

However, if the threshold $x \gg \mathbb{E}[S_N] = \mathbb{E}[N]\mathbb{E}[Y]$, we have difficulties in observing the event $\{S_N > x\}$ when we apply the standard Monte Carlo algorithm. As an illustration, let N be Abel and Y be inverse Gaussian, where the parameters are fitted by data in our case study of Section 6. The mean aggregate claim size $\mathbb{E}[S_N] \approx 330$. Because our distributions have large tails, we consider large levels x . As sample size we choose M so large that the standard error is about 10% of the estimate. We see in Table 1 that the required sample sizes grow exponentially with level x which means that very small probabilities are practically impossible to compute.

5.1 Importance Sampling Algorithm

The idea of importance sampling is to change the underlying probability measure of the stochastic system in such a way that more samples are generated from the target event. An unbiased estimator is obtained by multiplying the observations with the likelihood ratio. Denote the random variables that are generated in importance sampling by \widetilde{N} and \widetilde{S} , respectively. Suppose that $\widetilde{N} = n$, and $\widetilde{S} = s$ are simulated, then the associated likelihood ratio is

$$W(n, s) = \frac{\mathbb{P}(N = n)}{\mathbb{P}(\widetilde{N} = n)} \times \frac{f_S(s)}{f_{\widetilde{S}}(s)}.$$

The importance sampling estimator becomes

$$\widehat{\ell} = \frac{1}{M} \sum_{i=1}^M \mathbb{I}\{\widetilde{S}_{\widetilde{N}^{(i)}}^{(i)} > x\} W(\widetilde{N}^{(i)}, \widetilde{S}^{(i)}).$$

We have implemented the following importance sampling algorithm. Let the parameters of the counting distribution be (θ_N, p_N, m_N) (see Section 3), and of the claim distribution (θ_Y, p_Y, m_Y) (see Section 4). These parameters are fitted to the data, but note that the NEF-parameter θ follows from the mean-parameter m , and vice-versa, thus one of these suffices. For the change of measure we propose changing the NEF-parameter (and consequently the mean-parameter), but not the dispersion parameter p . In fact, we apply an exponential change of measure using a common tilting parameter, say θ^* , for both the counting and the claim-size distribution. This parameter is obtained as follows. Let $\kappa(\theta) = \log \mathbb{E}[\exp(\theta S_N)]$ be the cumulant generating function of the aggregated sum. Then θ^* solves the saddle point equation $\kappa'(\theta) = x$; thus

$$m_N(\theta_N + \theta^*) \times m_Y(\theta_Y + \theta^*) = x.$$

Table 2. Estimates of $\mathbb{P}(S_N > x)$ by importance sampling simulation for Abel counting and inverse Gaussian claim distributions

x	M	$\widehat{\ell}$	std. error
5000	4000	1.01e-02	9.09e-04
10000	6000	2.46e-03	2.43e-04
15000	10000	7.18e-04	6.88e-05
20000	14000	2.22e-04	2.23e-05
25000	16000	8.48e-05	8.40e-06
30000	20000	3.59e-05	3.65e-06
35000	26000	1.29e-05	1.25e-06
40000	34000	4.42e-06	4.41e-07
45000	34000	2.18e-06	2.16e-07
50000	40000	7.68e-07	7.78e-08

The interpretation is that under the change of measure the most likely samples of S_N are generated around our target level x . It is well-known in the rare-event theory that such a change of measure yields a logarithmically efficient (or, asymptotically optimal) estimator in case of a fixed number of light-tailed claims, i.e. $\mathbb{P}(S_n > x)$, see Asmussen and Glynn (2007(@, Chapter VI Section 2) or Bucklew (2004, Chapter V Section 2). However, by a conditioning argument one can show that the same holds true for a random sum. This means that the required sample sizes grow polynomially in level x , which we can clearly see in Table 2. Our algorithm contains a minor tweak in that after \widetilde{N} has been generated, say $\widetilde{N} = n$, we check whether $\mathbb{E}[S_n] > x$. In that case, we generate S_n from the original claim distribution, and otherwise we apply the change of measure also for the claims.

6. Case Study

Data are available of claims at a car insurance company in Sweden in a specific year (Hallin and Ingenbleek, 1983; Smyth, 2011). The data consist of 2182 categories of 7 variables specifying per category: kilometers, zone, bonus, make, insured, claims, payment. Let I be the set of categories, with $|I|=2182$. For any $i \in I$, we model the random variables

- N_i : the number of claims in category i ;
- Y_i : the claim size of a claimer in category i ;
- S_i : the total amount of claims in category i .

The data give the numbers n_i of claimers and s_i of total claim amount, but not the individual claim sizes.

For subcategories $J \subset I$ we propose that the $N_j, j \in J$ are i.i.d. as N , and that the $Y_j, j \in J$ are i.i.d. as Y . Also we propose that N and Y are independent. Data available are $n_j, j \in J$ observations from N , and $s_j = \sum_{k=1}^{n_j} y_{jk}$ observations from $S = \sum_{k=1}^N Y_k$ given N . Let θ_N be the vector of parameters of the probability distribution of the counting variable N , and θ_Y of the claim size distribution of Y . Due to the reproducibility property of Y , the distribution of sum $S|(N = n)$ has the same parameter vector θ_Y (and the given number n). For estimating these parameters we considered the two-moment fit method because all our distributions are derived from the mean and variance.

That is, let \widehat{m}_N and \widehat{v}_N be the sample average and variance of the counting data $(n_j)_{j \in J}$. Then we fit a distribution for N such that

$$\mathbb{E}[N] = \widehat{m}_N \text{ and } \text{Var}(N) = \widehat{v}_N.$$

For the counting distributions of Section 3 we get

$$\begin{aligned} \text{Abel : } p &= \frac{\widehat{m}_N \sqrt{\widehat{m}_N}}{\sqrt{\widehat{v}_N} - \sqrt{\widehat{m}_N}} \\ \text{arcsine : } p &= \frac{\widehat{m}_N \sqrt{\widehat{m}_N}}{\sqrt{\widehat{v}_N} - \widehat{m}_N} \\ \text{Takacs : } p &= \frac{4\widehat{m}_N \sqrt{\widehat{m}_N}}{\sqrt{8\widehat{v}_N + \widehat{m}_N} - 3\sqrt{\widehat{m}_N}} \end{aligned}$$

Similarly, let \widehat{m}_Y and \widehat{v}_Y be the sample average and variance of the claim data $(y_{jk})_{j \in J, k=1, \dots, n_j}$. Then we fit a distribution for Y such that

$$\mathbb{E}[Y] = \widehat{m}_Y \text{ and } \text{Var}(Y) = \widehat{v}_Y.$$

Note that the individual claim data (y_{jk}) are not observed, but that their sample average can be computed:

$$\widehat{m}_Y = \frac{\sum_j s_j}{\sum_j n_j}.$$

And for the sample variance of the individual claims we use the well-known identity for the variance of the aggregated sum $S = \sum_{k=1}^N Y_k$:

$$\text{Var}(S) = (\mathbb{E}[N])(\text{Var}(Y)) + (\text{Var}(N))(\mathbb{E}[Y])^2.$$

6.1 Subcategories Larger Cities

630 data have insured customers from major cities.

data	average	variance
claim number	70.60	52181.52
aggregate claim size	329.22	1153532.32
individual claim size	4.66	265.34

These yield values of dispersion parameter p of the counting distributions:

	p
Abel	2.695844
arcsine	2.598444
Takács	3.821015

The parameters of the claim distributions were obtained as explained above:

	θ	p	α
gamma	0.982425	0.081960	
IG	-0.008788	2.616360	
stable	-0.015496	2.134192	0.118315

With these parameters we have fitted the counting distribution and the claim distribution. Then we ran simulations of aggregated claim sizes in these models and executed the chi-square test for goodness-of-fit (hypothesizing that the samples came from the same distribution). As an example, below we show the histograms of the data S_N and the simulated S_N in case of the Poisson-gamma, Abel-IG and arcsine-stable combinations.

Table 3 summarizes the test results in terms of p -values.

We may conclude that the arcsine counting variable with positive stable claim size gives the best fit. The computations of the risk probabilities in this model are easily implemented by the algorithms that we exposed in Section 3 for the arcsine samples, in Section 4 for the positive stable samples, and in Section 5 for the Monte Carlo and importance sampling simulations. Table 4 shows results for both standard Monte Carlo and importance sampling simulations. Again we see the exponential versus polynomial increase of the required sample sizes. For levels $x \leq 25000$ the estimates fall in their corresponding confidence intervals (in most cases), while for large levels, $x > 25000$, we have no Monte Carlo results.

7. Conclusion

We analyzed insurance claim data and modeled the accumulated claim during a certain a period as a random sum of positive random variables representing the individual claims. The data showed that both the random sum and the random claim size have variances as large as cubic powers of their means. For fitting distributions with cubic VF's to the insurance data we used the NEF modeling. In this way we considered three discrete counting variables for fitting the random sum, and three positive continuous distributions for fitting the claim size, all coming from NEF's. We gave a thorough analysis of the nontrivial discrete counting variables for the purpose of developing sampling algorithms. These sampling algorithms are all accept-reject based, where the dominating proposal distribution is a Zipf distribution. Our claim size distributions are commonly known and sampling algorithms can be found in the literature.

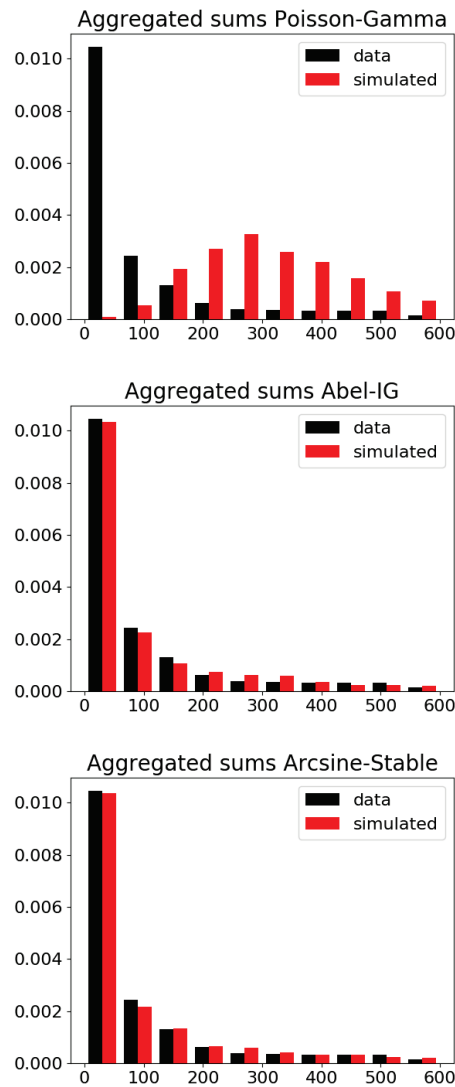


Figure 2. Histograms of the data and 2000 simulated samples (normed to form pdf's)

Table 3. p -values of the fitted models

count	claim	p -value
Poisson	gamma	≈ 0
Poisson	IG	≈ 0
Poisson	stable	≈ 0
Abel	gamma	0.2101
Abel	IG	0.2459
Abel	stable	0.3089
arcsine	gamma	0.1306
arcsine	IG	0.4224
arcsine	stable	0.7460
Takács	gamma	0.4159
Takács	IG	0.2800
Takács	stable	0.2701

Table 4. Estimates of $\mathbb{P}(S_N > x)$ by Monte Carlo and importance sampling simulation for arcsine counting and positive α -stable claim distributions

x	Monte Carlo			importance sampling		
	M	$\widehat{\ell}$	std. error	M	$\widehat{\ell}$	std. error
5000	10000	1.02e-02	1.00e-03	4000	9.89e-03	9.39e-04
10000	46000	2.11e-03	2.14e-04	7000	2.23e-03	2.23e-04
15000	128000	7.64e-04	7.73e-05	9000	8.12e-04	8.20e-05
20000	394000	2.49e-04	2.51e-05	14000	2.21e-04	2.25e-05
25000	1360000	7.13e-05	7.24e-06	16000	8.37e-05	8.52e-06
30000				20000	4.18e-05	4.13e-06
35000				24000	1.42e-05	1.42e-06
40000				30000	5.10e-06	5.15e-07
45000				36000	2.31e-06	2.35e-07
50000				38000	1.08e-06	1.08e-07

Being able to sample from the aggregate claim distribution, we executed Monte Carlo simulations for computing tail probabilities, specifically for large losses. The efficiency of these simulations was improved by two techniques. The first being that the claim size distributions satisfy the reproducibility property implying that convolutions come from the same family as the individual distribution. The second improvement is the application of importance sampling. Our numerical experiments showed that the exponential complexity of standard Monte Carlo is reduced to polynomial complexity.

Acknowledgements

The part of work of Shaul Bar-Lev is partially supported by the Netherlands Organization for Scientific Research (NWO) project number 040.11.608. The part of work of Ad Ridder is partially supported by the Zimmerman foundation while he was visiting the University of Haifa in January 2017.

References

Asmussen, S., & Albrecher, H. (2010). *Ruin Probabilities* (2nd ed.). World Scientific, Singapore. <https://doi.org/10.1142/7431>

Asmussen, S., & Glynn, P. W. (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer-Verlag, New York. <https://doi.org/10.1007/978-0-387-69033-9>

Bahnemann, D. (2015). *Distributions for Actuaries*. Casualty Actuarial Society, Arlington, Virginia.

Bar-Lev, S. K., & Cassalis, M. M. (2003). A classification of reproducible natural exponential families in the broad sense. *Journal of Theoretical Probability*, 16, 175-196. <https://doi.org/10.1023/A:1022286606451>

Bar-Lev, S. K., & Enis, P. (1986). Reproducibility and natural exponential families with power variance functions. *Annals of Statistics*, 14(4), 1507-1522. <https://doi.org/10.1214/aos/1176350173>

Bar-Lev, S. K., & Kokonendji, C. C. (2017). On the mean value parameterization of natural exponential families – a Revisited Review. *Mathematical Methods of Statistics*, 26(3), 159-175. <https://doi.org/10.3103/S1066530717030012>

Bowers, N. L., Gerber, H. U., Hickman, J. C., Jones, D. A., & Nesbitt, C. J. (1997). *Actuarial Mathematics* (2nd ed.). Society of Actuaries, Itasca, Illinois. <https://doi.org/10.1017/S0071368600009812>

Bucklew, J. A. (2004). *Introduction to Rare Event Simulation*. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4757-4078-3>

Chambers, J. M., Mallows, C. L., & Stuck, B. W. (1976). A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354), 340-344. <https://doi.org/10.1080/01621459.1976.10480344>

Consul, P. C. (1989). *Generalized Poisson Distributions: Properties and Applications*. Marcel Dekker, New York.

Consul, P. C., & Shoukri, M. M. (1988). Some chance mechanisms related to a generalized Poisson probability model. *American Journal of Mathematical and Management Sciences*, 8, 1-2, 181-202. <https://doi.org/10.1080/01966324.1988.10737237>

Devroye, L. (1986). *Non-Uniform Random Variate Generation*. Springer-Verlag, New York. <https://doi.org/10.1007/978-1-4613-8643-8>

- Devroye, L. (1992). The branching process method in Lagrange random variate generation. *Communications in Statistics - Simulation and Computation*, 21(1), 1–14. <https://doi.org/10.1080/03610919208813005>
- Dickson, D. C. M. (1998). On a class of renewal risk processes. *North American Actuarial Journal*, 2(3), 60-68, 72-73. <https://doi.org/10.1080/10920277.1998.10595723>
- Dunn, P. K., & Smyth, G. K. (2005). Series evaluation of Tweedie exponential dispersion densities. *Statistics and Computing*, 15(4), 267–280. <https://doi.org/10.1007/s11222-005-4070-y>
- Dunn, P. K., & Smyth, G. K. (2008). Evaluation of Tweedie exponential dispersion model densities by Fourier inversion. *Statistics and Computing*, 18(1), 73-86. <https://doi.org/10.1007/s11222-007-9039-6>
- Feller, W. (1971). *An Introduction to Probability Theory and its Applications, Volume II* (2nd ed.). Wiley, New York.
- Hallin, M., & Ingenbleek, J.-F. (1983). The Swedish automobile portfolio in 1977. *Scandinavian Actuarial Journal*, 1983(1), 49–64. <https://doi.org/10.1080/03461238.1983.10408691>
- Jorgensen, B. (1987). Exponential dispersion models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 49(2), 127-162. <https://doi.org/10.1111/j.2517-6161.1987.tb01685.x>
- Kaas, R., Goovaerts, M., Dhaene, J., & Denuit, M. (2008). *Modern Actuarial Risk Theory* (2nd ed.). Springer, Heidelberg. <https://doi.org/10.1007/978-3-540-70998-5>
- Klugman, S. A., Panjer, H. H., & Willmot, G. E. (2008). *Loss Models: From Data to Decisions* (3rd ed.). John Wiley. <https://doi.org/10.1002/9780470391341>
- Natural real exponential families with cubic variance functions. *The Annals of Statistics*, 18(1), 1-37. <https://doi.org/10.1214/aos/1176347491>
- Michael, J. R., Schucany, W. R., & Haas, R. W. (1976). Generating random variates using transformations with multiple roots. *The American Statistician*, 30(2), 88–90. <https://doi.org/10.1080/00031305.1976.10479147>
- Nolan, J. P. (2010). *Stable Distributions: Models for Heavy Tailed Data*. Birkhäuser, Boston.
- Samorodnitsky, G., & Taqqu, M. S. (1994). *Stable Non-Gaussian Random Processes*. Chapman and Hall, New York. <https://doi.org/10.1201/9780203738818>
- Shuster, J. (1968). On the inverse Gaussian distribution function. *Journal of the American Statistical Association*, 63(324), 1514-1516. <https://doi.org/10.2307/2285899>
- Smyth, G. K. (2011). Third party motor insurance in Sweden. Australasian Data and Story Library (OzDASL). <http://www.statsci.org/data/>
- Smyth, G. K., & Jorgensen, B. (2002). Fitting Tweedie's compound Poisson model to insurance claims data: dispersion modelling. *Astin Bulletin* 32, 143–157. <https://doi.org/10.2143/AST.32.1.1020>
- Tweedie, M. C. K. (1984). An index which distinguishes between some important exponential families. In J.K. Ghosh and J. Roy (Eds.), *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Golden Jubilee International Conference*. Indian Statistical Institute, Calcutta, pp. 579–604.
- Vinogradov, V., Paris, R. B., & Yanushkevichiene, O. O. (2012). New properties and representations for members of the power-variance family. I. *Lithuanian Mathematical Journal*, 52(4), 444-461. <https://doi.org/10.1007/s10986-012-9186-0>
- Vinogradov, V., Paris, R. B., & Yanushkevichiene, O. O. (2013). New properties and representations for members of the power-variance family. II. *Lithuanian Mathematical Journal*, 53(1), 103–120. <https://doi.org/10.1007/s10986-013-9197-5>

Appendix A

Arcsine Bound

Recall the kernel on the even outcomes

$$\nu(2n) = \frac{1}{(2n)!} \prod_{i=0}^{n-1} ((2i)^2 + p^2) = \prod_{i=0}^{n-1} \frac{(2i)^2 + p^2}{(2i+1)(2i+2)} = \prod_{i=0}^{n-1} \underbrace{\frac{4i^2 + p^2}{4i^2 + 6i + 2}}_{=\rho_i}$$

We shall bound the ρ_i 's for sufficiently large i . The threshold lies at i^* that is such that both $i^* + 1 \geq 6$, and

$$\frac{3}{2i} - \left(\frac{9+p^2}{4i^2} - \frac{3p^2}{8i^3} + \frac{9p^2}{4i^4} \right) > 0 \quad \text{for all } i \geq i^* + 1.$$

First, dividing by $4i^2$ we easily get (for all $i \geq 1$),

$$\begin{aligned} \rho_i &= \frac{1 + p^2/(4i^2)}{1 + 6/(4i) + 2/(4i^2)} \leq \frac{1 + p^2/(4i^2)}{1 + 6/(4i)} \\ &\leq \left(1 + \frac{p^2}{4i^2}\right) \left(1 - \frac{6}{4i} + \frac{36}{(16i^2)}\right) \\ &= 1 - \frac{3}{2i} + \underbrace{\left(\frac{9+p^2}{4i^2} - \frac{3p^2}{8i^3} + \frac{9p^2}{4i^4}\right)}_{=\epsilon_i}. \end{aligned}$$

Thus for $i \geq i^* + 1$ we have that $\frac{3}{2i} - \epsilon_i > 0$, and therefore we can bound

$$\log \rho_i \leq \log \left(1 - \left(\frac{3}{2i} - \epsilon_i\right)\right) \leq -\left(\frac{3}{2i} - \epsilon_i\right).$$

Define $G = \prod_{i=0}^{i^*} \rho_i$. So we get for $n \geq i^* + 1$,

$$\begin{aligned} \log \nu(2n) &= \log G + \sum_{i=i^*+1}^{n-1} \log \rho_i \\ &\leq \log G - \sum_{i=i^*+1}^{n-1} \left(\frac{3}{2i} - \epsilon_i\right) = \log G - \frac{3}{2} \sum_{i=i^*+1}^{n-1} \frac{1}{i} + \sum_{i=i^*+1}^{n-1} \epsilon_i. \end{aligned}$$

The second term is easy:

$$\sum_{i=i^*+1}^{n-1} \frac{1}{i} = \sum_{i=1}^{n-1} \frac{1}{i} - \sum_{i=1}^{i^*} \frac{1}{i} \geq \log n - \sum_{i=1}^{i^*} \frac{1}{i}.$$

The ϵ_i 's are clearly positive, and for $i \geq 6$,

$$\epsilon_i = \frac{9+p^2}{4i^2} - \frac{3p^2}{8i^3} + \frac{9p^2}{4i^4} = \frac{9}{4i^2} + \frac{p^2}{8i^2} \underbrace{\left(2 - \frac{3}{i} + \frac{18}{i^2}\right)}_{\leq 2} \leq \frac{9+p^2}{4i^2}.$$

Hence, for $n \geq i^* + 1$,

$$\begin{aligned} \sum_{i=i^*+1}^{n-1} \epsilon_i &\leq \frac{9+p^2}{4} \sum_{i=i^*+1}^{n-1} \frac{1}{i^2} = \frac{9+p^2}{4} \left(\sum_{i=1}^{n-1} \frac{1}{i^2} - \sum_{i=1}^{i^*} \frac{1}{i^2}\right) \\ &\leq \frac{9+p^2}{4} \left(\sum_{i=1}^{\infty} \frac{1}{i^2} - \sum_{i=1}^{i^*} \frac{1}{i^2}\right) = \frac{9+p^2}{4} \left(\zeta(2) - \sum_{i=1}^{i^*} \frac{1}{i^2}\right), \end{aligned}$$

where $\zeta(\cdot)$ is the Riemann-zeta function. Wrapping up we get by exponentiating, $\nu(2n) \leq K_1 n^{-3/2}$ for $n \geq i^* + 1$, where

$$K_1 = G \exp\left(\frac{3}{2} \sum_{i=1}^{i^*} \frac{1}{i} + \frac{9+p^2}{4} \left(\zeta(2) - \sum_{i=1}^{i^*} \frac{1}{i^2}\right)\right).$$

Find K_0 such that both $\nu(2n) \leq K_0 n^{-3/2}$ for $n = 1, \dots, i^*$. We demand this inequality also for the odd terms; i.e., $\nu(2n + 1) \leq K_0 n^{-3/2}$, $n = 1, \dots, i^*$. Then by setting $K = \max\{K_0, K_1\}$,

$$\nu(2n) \leq K \frac{1}{n \sqrt{n}}, \quad n \geq 1.$$

Again we recognize the Zipf distribution, which will be useful for an accept-reject sampling algorithm.

Appendix B

Takács Bound

Recall the adapted kernel

$$\begin{aligned} \nu_0(n) &= \nu(n)e^{\theta_0(m)n} = \frac{p}{n+p} \frac{1}{n!} (n+p)(n+p+1) \cdots (n+p+n-1) e^{\theta_0(m)n} \\ &= \frac{p}{n+p} \frac{(2n+p-1)!}{n!(n+p-1)!} e^{\theta_0(m)n}, \quad n = 0, 1, \dots \end{aligned}$$

The Stirling bounds of $n!$ are

$$\sqrt{2\pi n} (ne^{-1})^n < n! < \sqrt{2\pi n} (ne^{-1})^n e^{1/12}, \quad n = 1, 2, \dots$$

Applying these bounds, we get

$$\nu_0(n) \leq \frac{p}{n+p} \frac{e^{1/12}}{\sqrt{2\pi}} \sqrt{\frac{2n+p-1}{n(n+p-1)}} \left(\frac{2n+p-1}{n+p-1}\right)^{p-1} \left(\frac{(2n+p-1)^2}{n(n+p-1)} e^{\theta_0(m)}\right)^n$$

The factors of this expression are worked out below:

$$\sqrt{\frac{2n+p-1}{n(n+p-1)}} = \frac{1}{\sqrt{n}} \sqrt{\frac{2+\frac{p-1}{n}}{1+\frac{p-1}{n}}} \leq \frac{1}{\sqrt{n}} \sqrt{2},$$

because in our models $p > 1$. Thus, also,

$$\left(\frac{2n+p-1}{n+p-1}\right)^{p-1} = \left(\frac{2+\frac{p-1}{n}}{1+\frac{p-1}{n}}\right)^{p-1} \leq 2^{p-1}.$$

Finally,

$$\begin{aligned} \frac{(2n+p-1)^2}{n(n+p-1)} e^{\theta_0(m)} &= \frac{(2n+p-1)^2}{n(n+p-1)} \frac{m(p+m)}{(p+2m)^2} \\ &= \frac{4n^2+4np-4n+(p-1)^2}{n^2+np-n} \frac{mp+m^2}{p^2+4mp+4m^2} \\ &= \frac{n^2+np-n+(p-1)^2/4}{n^2+np-n} \frac{mp+m^2}{p^2/4+mp+m^2} \\ &= \left(1 + \frac{(p-1)^2/4}{n^2+n(p-1)}\right) \left(1 + \frac{p^2/4}{m^2+mp}\right)^{-1}. \end{aligned}$$

This expression is less than 1 for all $n \geq m$. Putting it all together, we get

$$\nu_0(n) \leq \frac{p}{n+p} \frac{e^{1/12}}{\sqrt{2\pi}} \frac{\sqrt{2}}{\sqrt{n}} 2^{p-1}, \quad n \geq m.$$

Let

$$K_1 = p \frac{e^{1/12}}{\sqrt{2\pi}} \sqrt{2} 2^{p-1},$$

then $\nu_0(n) \leq K_1 n^{-3/2}$ for all $n \geq m$. Find K_0 such that $\nu_0(n) \leq K_0 n^{-3/2}$ for $n = 1, \dots, m$. Then by setting $K = \max\{K_0, K_1\}$, we conclude

$$\nu_0(n) \leq K \frac{1}{n \sqrt{n}}, \quad n \geq 1.$$

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).