Journal of Uncertain Systems
Vol.9, No.3, pp.198-214, 2015
Online at: www.jus.org.uk

Imprecise Individual Risk Models of Insurance

Lev V. Utkin[1], Frank P.A. Coolen[2,*]

[1] *Department of Control, Automation and System Analysis, St.Petersburg State Forest Technical University Institutski per. 5, 194021, St.Petersburg, Russia*

[2] *Department of Mathematical Sciences, Durham University, Durham, DH1 3LE, England*

Received 17 January 2015; Revised 4 Jun 2015

Abstract

Two imprecise individual risk models of insurance are presented and studied in this paper. The models use the beta-binomial and the negative binomial imprecise statistical models, which can be regarded as sets of the corresponding distributions of claims. The sets of distributions strongly depend on prior statistical information about claims and their use can give an advantage when we have little prior information. The results of these models are interval-valued probabilities for the event that aggregate claims will exceed the premium. Moreover, it is shown how to construct the corresponding fuzzy probabilities. Numerical examples illustrate the proposed imprecise models.

Keywords: beta-binomial distribution, claims, fuzzy sets, imprecise probabilities, insurance, negative binomial distribution

1 Introduction

Insurance is a protection in order to provide an individual or business compensation in case of a possible loss or damage. Insurance is a well-known risk sharing strategy and there is a huge literature devoted to various aspects of insurance. The random nature of claims does not allow unique determination of the parameters of insurance and there is some risk of ruin for the insurer. Actuarial methods aim at insurance parameter evaluation in order to minimize risks. Obviously, correctness and usefulness of actuarial methods is mainly defined by adequacy of the corresponding actuarial models depending on completeness of statistical information about claims.

The model we consider in the paper is known as the individual risk model of insurance and we study the accumulated claim of an insurance portfolio. Claim means the amount that the insurance company has to pay out to its policy holders. Accumulated means that all claims from different policies and possibly different branches are looked at in aggregation over some time period. Portfolio means that we consider an aggregate of insurance contracts. We suppose that the claim size is constant, but the number of claims in a given period of time is uncertain and can be modelled by a discrete probability distribution.

The established models of insurance assume that there are precise probability distributions of random variables characterizing the insurance process, or similarly precise probability distributions for parameters of such probability distributions if Bayesian methods are used. However, many real applications meet some difficulties with getting high quality statistical information in order to justify such a level of precision, or alternatively high quality expert judgement in order to do so. For many insurance areas, the prior statistical information about claims might be partial and imprecise. Moreover, this information might be entirely absent, especially when a new insurance policy is introduced and analyzed. Therefore, the well-known statistical models might not provide suitable methods for dealing with the varying kinds of uncertainties involved, and models which explicitly take indeterminacy into account are attractive.

One of the interesting ways for dealing with unreliable and partial information in insurance is the Bayesian approach [3]. Many papers study the well-known Bayesian models which can be directly or indirectly applied to

*Corresponding author.
Emails: lev.utkin@gmail.com (L.V. Utkin), frank.coolen@durham.ac.uk (F.P.A. Coolen).

insurance problems [4, 6, 8, 24]. They typically consider the well-known Beta-binomial and negative binomial distributions which can be successfully applied to the individual risk model of insurance. However, in spite of the efficiency of Bayesian models, they have an important disadvantage. For these models, a noninformative prior has to be constructed if there is no suitable prior information available. There is a variety of methods for determining noninformative priors in the literature, but we will show that such methods which use a single prior distribution in an attempt to be 'non-informative' may not be suitable as they may lead to unsatisfactory decisions.

Many methods and models of insurance have been proposed which use fuzzy set theory for representing initial information, e.g. [9, 15, 16, 20, 21]. Some elements of imprecise probability theory in insurance problems have been also presented by Jeleva [12] and by Jeleva and Villeneuve [13] in order to relax some strong assumptions concerning the probability distributions of insurance parameters. Robust insurance models based on the ε-contaminated (robust) models [10] have been presented by Carlier et al. [5].

In this paper, we combine the Bayesian approach and imprecise probability theory [2] to present new models for cautious decision making in insurance, taking into account incompleteness and imprecision of data about claims. We consider the imprecise beta-binomial and imprecise negative binomial models, which can be regarded as generalizations of the beta-binomial and negative binomial distributions. They are used for cautious probabilistic inference for the event that aggregate claims will be less than the premium collected. We also show that the proposed imprecise models can be represented in terms of fuzzy insurance models. Numerical examples illustrate these models. It should be emphasized that the imprecise probability models used in this paper are closely related to the imprecise Dirichlet model (IDM) presented by Walley [26]. The imprecise beta-binomial model is a special case of the IDM and was already presented by Walley [25]. The main novelty of this paper is the application of these models to probabilistic individual risk models of insurance, and to develop and investigate the corresponding imprecise risk models.

The paper is organized as follows. Some elements of the individual risk model of insurance are introduced in Section 2. Applications of the Bayesian approach to modelling the individual risk are given in Section 3, where the beta-binomial and negative binomial distributions are reviewed as they form the bases of the new imprecise insurance models presented in Section 4. A novel way to derive fuzzy models from the proposed imprecise models is explained in Section 5. The paper ends with some concluding remarks in Section 6.

2 Individual Risk Model

We briefly review the well-known individual risk model of insurance, which is widely used in applications, especially in life and health insurance. Assume that a portfolio consists of N identical insurance policies for a given period of time t, and the claim made in respect of the policy i is denoted by X_i. This random variable X_i is usually presented as $X_i = I_i \cdot y_i$, where y_i is the claim amount (size) and the random binary variable I_i indicates if the i-th policy involved a claim: if this is the case then $I_i = 1$, while if there has not been any claim under policy i then $I_i = 0$. We denote the corresponding probabilities as follows, $\Pr\{I_i = 1\} = q$ and $\Pr\{I_i = 0\} = 1 - q$, so I_i is Bernoulli distributed with parameter q. We assume that each insurance premium is c. Then the total premium for the time t is $\Pi(t) = cN$ and the total or aggregate amount of claims is defined as $R(t) = X_1 + \cdots + X_N$. We further assume that all claim amounts are equal, so $y_i = y$ for all i. It should be remarked that the simplifying assumptions made here are mainly included in order to present the new imprecise methods without further complications. However, generalizations along the same lines are possible in order to relax these assumptions, detailed research into this is an important topic for the future.

The probability that aggregate claims will be less than the total premium collected is determined by

$$P = \Pr\{\Pi(t) \geq R(t)\} = \Pr\{cN \geq X_1 + \cdots + X_N\}.$$

Of course, this implies that the probability that the aggregate claims will exceed the premium is given by

$$Q = \Pr\{\Pi(t) \leq R(t)\} = 1 - P.$$

If the random variables $X_1, ..., X_N$ are mutually independent and the parameters c, N, y are fixed for all $i = 1, ..., N$, then the probability P can be computed by introducing the random number K of claims for the time t, which then has a discrete probability distribution $p(k, w)$, $k = 0, 1, 2, ...$, with a parameter (or a set of parameters) w as

$$P = \sum_{k=0}^{M} p(k, w).$$

Here $M = \lceil \Pi(t)/y \rceil = \lceil cN/y \rceil$ is the largest number of claims which can be paid by the insurer.

If the probability q is known, then $p(k, w)$ is the binomial distribution with parameter $w = q$, so in this case we have

$$P = \sum_{k=0}^{M} \binom{N}{k} \cdot q^k (1-q)^{N-k}.$$

Under the assumption that only a relatively small proportion of policies will involve a claim, the Poisson distribution is commonly used as a suitable approximation to the binomial distribution. For this model, initial information may be represented by the claim rate λ, i.e., roughly speaking, by the mean number of claims occurring in a unit time interval (e.g. a year). In this case, the probability distribution with parameter $w = \lambda$ is given by

$$P = \sum_{k=0}^{M} \frac{(\lambda t)^k \exp(-\lambda t)}{k!}.$$

It is obvious that the considered model is simple enough and may particularly be suitable for small time intervals t. Nevertheless, even for such simple models, we need to know a precise value for the parameter w, so the probability q or the claim rate λ.

3 Bayesian Approach

If the parameter w is unknown, then it can be regarded as a random variable with some probability density function $\pi(w, \theta)$. In this case, the Bayesian approach can be applied for computing the probability P for the event that the aggregate claims will be less than the total premium collected, which is determined as the following unconditional expected value:

$$\begin{aligned} P &= \int_{\Omega} \sum_{k=0}^{M} p(k, w) \cdot \pi(w, \theta) \mathrm{d}w \\ &= \sum_{k=0}^{M} \int_{\Omega} p(k, w) \cdot \pi(w, \theta) \mathrm{d}w = \sum_{k=0}^{M} P(k). \end{aligned}$$

Here θ is the vector of parameters of π; Ω is the set of values of w; $P(k)$ is the probability of exactly k claims. For the binomial model, with $w = q$, we have $\Omega = [0, 1]$, while for the Poisson model, with $w = \lambda$, we have $\Omega = \mathbb{R}_+$.

The well-known Bayesian approach is concerned with generating the posterior distribution of the unknown parameters given both the data and some prior distribution for these parameters. We suppose that the prior distribution $\pi(w, \theta)$ represents our beliefs about the parameter w, considered as a random variable, prior to collecting any information in the form of the set $\mathbf{k} = (k_1, \cdots, k_n)$ of observed numbers of claims for each of the policies.

The likelihood function $L(\theta|\mathbf{k})$ is the density function for the observed data given θ, which is computed as

$$L(\theta|\mathbf{k}) = p(k_1, w) \cdots p(k_n, w).$$

Then the posterior distribution $\pi(w, \theta|\mathbf{k})$, which is the conditional distribution of θ given the observed data $\mathbf{k}$, is computed as

$$\pi(w, \theta|\mathbf{k}) \propto L(\theta|\mathbf{k}) \cdot \pi(w, \theta).$$

This posterior distribution $\pi(w, \theta|\mathbf{k})$ represents our updated opinion about the possible values θ can take on, now that we have some information $\mathbf{k}$ about the specific individual.

The prior distribution is often chosen to facilitate calculation of the posterior, especially through the use of conjugate priors. When the posterior distribution $\pi(w, \theta|\mathbf{k})$ and the prior distribution $\pi(w, \theta)$ both belong to the same distribution family, π and p are called conjugate distributions and π is the conjugate prior for p. For example, the beta distribution is a conjugate prior for the binomial distribution, the gamma distribution is a conjugate prior for the Poisson distribution. By applying the beta-binomial probability distribution and the negative binomial distribution to the individual risk models of insurance, we follow the papers [4, 8, 14] as explained next.

3.1 Beta-binomial Distribution

It follows from the above that if we assume that the claims are binomially distributed numbers with parameter θ, then it is convenient to choose as the distribution π of the parameters θ the beta distribution. The prior beta distribution for the random variable θ, denoted by $Beta(a, b)$, has the following probability density function:

$$\pi(\theta) = Beta(a, b) = \frac{1}{\mathrm{B}(a, b)} \theta^{a-1}(1-\theta)^{b-1}, \ 0 \le \theta \le 1.$$

Here $a > 0$, $b > 0$ are parameters of the beta distribution and $\mathrm{B}(a, b)$ is the Beta function.

Suppose that we have observations consisting of $K = \sum_{i=1}^{n} k_i$ claims from the total number of premiums N^*. Here k_i is the number of claims from the i-th source of data, this is the information that will be used to update our prior distribution. Then the corresponding posterior beta distribution $\pi(\theta|\mathbf{k})$ is of the form:

$$\pi(\theta|\mathbf{k}) = Beta(a + K, b + N^* - K).$$

Hence, the probability $P(k)$ for the event that exactly k claims will occur in future for N premiums and with $\theta = q$ is

$$\begin{aligned} P(k) &= \int_0^1 \binom{N}{k} \cdot q^k (1-q)^{N-k} \cdot Beta(a + K, b + N^* - K)\mathrm{d}q \\ &= \binom{N}{k} \cdot \frac{\mathrm{B}(a + k + K, b + N + N^* - k - K)}{\mathrm{B}(a + K, b + N^* - K)}. \end{aligned}$$

This is the beta-binomial distribution, so the probability that the aggregate claims will be less than the total premium collected is equal to

$$P = \sum_{k=0}^{M} \binom{N}{k} \cdot \frac{\mathrm{B}(a + k + K, b + N + N^* - k - K)}{\mathrm{B}(a + K, b + N^* - K)}.$$

Example 1. *Consider a short-term insurance. Suppose that an insurance company has $N = 100$ premiums. The claim size is $y = 10$. The premium is $c = 1.2$. Then the largest number of claims which can be paid by the insurer is $M = \lceil cN/y \rceil = 1.2 \cdot 100/10 = 12$. Suppose also that the probability of the claim occurrence is known to be $q = 0.1$. This can be interpreted as the ratio K/N^* of former observations being approximately equal to 0.1. Note that $P = 0.802$ when we apply the binomial distribution with $q = 0.1$. Suppose that we had a possibility to get data on only $N^* = 50$ premiums. Since we do not have any prior information about the parameters a and b, so-called noninformative priors would seem to be appropriate. One of the established noninformative priors for this model is Jeffreys' prior [11], which is the beta distribution with parameters $a = 1/2$ and $b = 1/2$. Let us generate 10 random values of K in accordance with the binomial distribution having parameters $q = 0.1$ and $N^* = 50$. It is obvious that the random numbers are close to 5. Let us take three numbers of total claims $4, 5, 6$ and compute the probability P by using the binomial distribution with the parameter $\hat{q} = K/N^*$ and the Beta-binomial distributions with parameters $a = 1/2$ and $b = 1/2$. The computation results are shown in Table 1.*

Now we suppose that the probability of the claim occurrence is $q = 0.01$ and other conditions and parameters are the same as in the previous case. Note that the probability that aggregate claims will exceed the total premium is $1 - P = 3.164 \times 10^{-11}$ when we apply the binomial distribution with $q = 0.01$. The most probable values of K in this case are $0, 1, 2$. The corresponding probabilities that aggregate claims will exceed the premium are shown in Table 2.

It can be seen from the results that deviations of the computed probabilities from the 'ideal' ones are significant. These computational results are caused by the fact that when we have a small number of observations

Table 1: Probabilities that aggregate claims will be less than the premium collected

	K		
distribution	4	5	6
binomial	0.944	0.802	0.576
beta-binomial	0.794	0.666	0.526

Table 2: Probabilities that aggregate claims will exceed the premium

	K		
distribution	0	1	2
binomial	0	1.149×10^{-7}	1.831×10^{-4}
beta-binomial	9.4×10^{-4}	1.041×10^{-2}	4.133×10^{-2}

these do not allow us to get reasonable statistical estimates. A further reason for these results is the use of Jeffreys' in case of such small numbers of observations. Therefore, more cautious approaches should be developed for computing these probabilities.

3.2 Negative Binomial Distribution

The negative binomial distribution and its extensions in insurance have been studied, for example, by Shi and Valdez [22]. If the number of claims has the Poisson distribution, which might be assumed if the underlying model is known to be binomial yet if claims are relatively rare, hence the Poisson distribution is a suitable approximation, then choosing the prior distribution π to be the gamma distribution is convenient as this is a conjugate prior. The prior gamma distribution of the random variable θ, denoted by $Gamma(a, b)$, has the following probability density function:

$$\pi(\theta) = Gamma(a, b) = \frac{1}{\Gamma(a)} b^a \theta^{a-1} \exp(-b\theta),\ \theta > 0.$$

Here $a > 0$, $b > 0$ are parameters of the gamma distribution and $\Gamma(a)$ is the gamma function.

Suppose that we had $K = \sum_{i=1}^{n} k_i$ claims observed during n periods of time. Here k_i is the number of claims during the i-th period of observations. Then the posterior distribution $\pi(\theta|\mathbf{k})$ is again a gamma distribution, namely:

$$\pi(\theta|\mathbf{k}) = Gamma(a + K, b + n).$$

The corresponding probability $P(k)$ for the event that exactly k claims will occur in future for N premiums, with $\theta = \lambda$ and $t = 1$, is

$$\begin{aligned} P(k) &= \int_0^\infty \frac{\lambda^k \exp(-\lambda)}{k!} \cdot Gamma(a, b) \mathrm{d}\lambda \\ &= \frac{\Gamma(a+k)}{\Gamma(a)k!} \cdot \left(\frac{b}{b+1}\right)^a \left(\frac{1}{b+1}\right)^k. \end{aligned}$$

This is the negative binomial distribution. Hence the probability for the event that aggregate claims will be less than the total premium collected is

$$P = \sum_{k=0}^{M} \frac{\Gamma(a+k)}{\Gamma(a)k!} \cdot \left(\frac{b}{b+1}\right)^a \left(\frac{1}{b+1}\right)^k.$$

So, based on known values of the parameters a and b of the prior distribution and with available data, we can compute the posterior probabilities for specific events of interest. The probability of exactly k claims can be computed using the following forward recursion formula:

$$P(k) = \begin{cases} \left(\frac{b}{b+1}\right)^a, & k = 0 \\ \frac{a+k-1}{k(b+1)} \cdot P(k-1) & k \geq 1. \end{cases}$$

Now we consider the more general case that periods of observations, which led to the available data, may have been different, say of length $t_1, ..., t_n$, respectively. Let the numbers of claims during these periods be $k_1, ..., k_n$, respectively. Let $T = t_1 + \cdots + t_n$ be the total time of observations, and $K = k_1 + \cdots + k_n$ the total number of claims over this total time T. Let $n = 1$. Applying Bayes theorem, we get the posterior density

$$\frac{\Pr\{X = k_1|\lambda\} Gamma(\lambda|a, b)}{\Pr\{X = k_1\}} = Gamma(a + k_1, b + t_1).$$

Hence, the probability of exactly k claims during time t, under the condition that k_1 claims were observed during the observation time t_1, is

$$\begin{aligned} P(k) &= \int_0^\infty \frac{(\lambda t)^k \exp(-\lambda t)}{k!} \cdot Gamma(a + k_1, b + t_1) \mathrm{d}\lambda \\ &= \frac{\Gamma(a + k_1 + k)}{\Gamma(a + k_1) k!} \cdot \left(\frac{b + t_1}{b + t_1 + t}\right)^{a+k_1} \left(\frac{t}{b + t_1 + t}\right)^k. \end{aligned}$$

A similar expression for probability of exactly k claims can be obtained for arbitrary $n \geq 1$. It can be seen from the obtained probability that it does not depend on the actual values $k_1, ..., k_n$ and $t_1, ..., t_n$ but only on the total claim number K and the total observation time T, i.e.,

$$P(k) = \frac{\Gamma(a + K + k)}{\Gamma(a + K) k!} \cdot \left(\frac{b + T}{b + T + t}\right)^{a+K} \left(\frac{t}{b + T + t}\right)^k.$$

In conclusion, the probability for the event that the aggregate claims will be less than the total premium is

$$P = \sum_{k=0}^{M} \frac{\Gamma(a + K + k)}{\Gamma(a + K) k!} \cdot \left(\frac{b + T}{b + T + t}\right)^{a+K} \left(\frac{t}{b + T + t}\right)^k.$$

Example 2. *Consider a short-term insurance. Suppose that an insurance company has $N = 100$ premiums. The claim size is $y = 10$. The premium is $c = 1.2$. Then the largest number of claims which can be paid by the insurer is $M = \lceil cN/y \rceil = 1.2 \cdot 100/10 = 12$. Suppose that $k_1 = 2$, $k_2 = 0$ numbers of claims occurred after $T = 2$ periods of observations, i.e., $K = k_1 + k_2 = 2$. By using the parameters $a = 1$ and $b = 1$, we can compute the probability that the aggregate claims will be less than the total premium collected for $t = 3$*

$$P = \sum_{k=0}^{12} \frac{\Gamma(1 + 2 + k)}{\Gamma(1 + 2) \cdot k!} \cdot \left(\frac{1 + 2}{1 + 2 + 3}\right)^{1+2} \left(\frac{3}{1 + 2 + 3}\right)^k = 0.996.$$

Similarly, if the total number of claims in the data would be $K = 3$, then $P = 0.989$, while if this number would be $K = 4$, then $P = 0.975$.

It should be noted that the beta-binomial and negative binomial distributions are widely applied in marketing research, actuarial and risk analyses due to interesting and useful properties. However, a critical feature of any Bayesian analysis is the choice of a prior, so for the models considered here the choice of the parameters a and b, in particular when we have no prior information about the parameters a and b. In this case, a noninformative prior has to be constructed. Many methods for determining noninformative priors have been presented in the literature. Many of these are based on the Bayes-Laplace postulate, which is also called the principle of insufficient reason. According to this principle, the prior distribution should be uniform. For instance, by applying this principle to the beta distribution, the parameters should be chosen as $a = 1$, $b = 1$. However, this choice meets some problems. The first problem is that the uniform distribution is not invariant under reparametrization. If we have no information, for instance, about a parameter θ, then we also have no information about $1/\theta$, but assigning a uniform distribution for both θ and $1/\theta$ does not lead to the same results. Another problem with the uniform prior is that if the parameter space is infinite, the uniform prior is improper because it does not integrate to one. Walley [26] gives further discussion and a number of examples illustrating possible problems and shortcomings of the principle of insufficient reason. Some interesting approaches for determining noninformative priors can be found in works of Perks [18] and Jeffreys [11]. A detailed review of methods for constructing a noninformative prior was presented by Syversveen [23].

The difficulties of presenting lack of prior data or information through a single prior distribution are now well understood. A fundamentally different approach to statistical inference in such situations is through the use of a class $\mathcal{M}$ of prior distributions π, instead of a single prior distribution. Such a class can be considered through the corresponding lower $\underline{P}$ and upper $\overline{P}$ probabilities of an event A, which are such that as

$$\underline{P}(A) = \sup\{P_\pi(A) : \pi \in \mathcal{M}\}, \quad \overline{P}(A) = \inf\{P_\pi(A) : \pi \in \mathcal{M}\}.$$

As pointed out by Syversveen [23] and Walley [26], the class $\mathcal{M}$ is aimed, based on some assumptions, to be 'not a class of reasonable priors, but a reasonable class of priors'. This means that each single member of the class is not in itself considered to be a reasonable model for prior ignorance, because no single distribution can model ignorance satisfactory. But the whole class is a reasonable model to reflect prior ignorance. When we have little prior information, the upper probability of a non-trivial event should be close to one and the lower probability should be close to zero. This means that the prior probability of the event may be arbitrary from 0 to 1.

Examples of the considered models are Walley's imprecise Dirichlet model [26], a bounded derivative model [27] and imprecise probability models for inference in exponential families [19], which can be regarded as extensions of Walley's imprecise Dirichlet model [17]. Such models have not yet been used in insurance analysis. Therefore, our aim is to develop new imprecise individual risk models of insurance which take into account the lack of prior knowledge about parameters of insurance processes. This is presented next, in relation to the two insurance models discussed in this Section.

4 Imprecise Models

This section presents the new contribution of this paper, consisting of imprecise probabilistic models applied to insurance scenarios. These models generalize the beta-binomial model and the negative binomial model reviewed in the previous section.

4.1 Imprecise Beta-binomial Model

In order to simplify the generalization to imprecise probabilities, which occurs by using a class of prior distributions instead of a single prior distribution, we change notation from that used in Section 3, where the more common notation was used in order to introduce the basic models. We replace the parameters a and b in the expression for the beta-binomial distribution $P(k)$ by parameters s and α, such that $a = s\alpha$ and $b = s - s\alpha$, so

$$P(k) = \binom{N}{k} \cdot \frac{\mathrm{B}(s\alpha + k + K, s - s\alpha + N + N^* - k - K)}{\mathrm{B}(s\alpha + K, s - s\alpha + N^* - K)}.$$

Here the hyperparameter $s > 0$ determines the influence of the prior distribution on posterior probabilities. In particular, if $s = 0$, then the posterior distribution is totally determined by the information in the form of N^* and K. The parameter α belongs to the interval from 0 to 1. Walley [26] proposed the imprecise model which can be defined as the set of all beta-binomial distributions with the fixed parameter s and the set of parameters $0 \leq \alpha \leq 1$. By dealing with the set of distributions instead of a single distribution, it is natural to use for inferences the optimal lower and upper bounds for $P(k)$ or P, corresponding to all distributions in this class, instead of precise values corresponding to one distribution. Of course, these lower and upper bounds, which are lower and upper probabilities, can be obtained by minimizing and maximizing $P(k)$ or P over all values α in $[0, 1]$, for fixed parameter s. The hyperparameter s also determines how quickly upper and lower probabilities of events converge as statistical data accumulate. Walley [26] defined s as a number of observations needed to reduce the imprecision (difference between upper and lower probabilities) to half its initial value. Smaller values of s produce faster convergence and stronger conclusions, whereas large values of s produce more cautious inferences.

It should be noted that the expectation of the number of claims X under conditions $K = 0$ and $N^* = 0$ is computed in terms of the introduced parameters as

$$\mathbb{E}X = N \cdot \frac{a}{a+b} = N\alpha.$$

The variance of X is of the form:

$$\mathrm{var}(X) = \frac{Nab(a+b+N)}{(a+b)^2(a+b+1)} = \frac{N\alpha(1-\alpha)(s+N)}{(s+1)} = N\alpha(1-\alpha)\frac{s+N}{s+1}.$$

It follows from the above that the parameter α for the beta-binomial distribution determines the relative prior expected number of claims.

Walley [26] states that the probability P, which in this paper is the probability that aggregate claims will be less than the total premium collected, decreases as the parameter α increases. This implies that the smallest value of P corresponding to the class of prior distributions, so the lower probability $\underline{P}$, is achieved at $\alpha = 1$,

$$\underline{P} = \sum_{k=0}^{M} \binom{N}{k} \cdot \frac{\mathrm{B}(s+k+K, N+N^*-k-K)}{\mathrm{B}(s+K, N^*-K)}.$$

Similarly, the largest value of P, so the upper probability $\overline{P}$, is achieved at $\alpha = 0$,

$$\overline{P} = \sum_{k=0}^{M} \binom{N}{k} \cdot \frac{\mathrm{B}(k+K, s+N+N^*-k-K)}{\mathrm{B}(K, s+N^*-K)}.$$

Note that, if $K = N$ then $\underline{P} = 0$ and if $K = 0$ then $\overline{P} = 1$. Before any data are taken into account, we have $K = N^* = 0$ so $\underline{P} = 0$ and $\overline{P} = 1$. In other words, if we do not have any data about claims, then the probability that aggregate claims will be less than the total premium collected can be arbitrary from 0 to 1. This is an important property as it reflects lack of prior knowledge about this specific event, which may e.g. be suitable in case of a new insurance product. Note further that the precise model is attained by setting $s = 0$, but as discussed this leads to only a single prior distribution to be used, hence it reduces the flexibility to model indeterminacy, which is the main advantage of the imprecise probability approach, e.g. because it enables cautious inference through focus on the lower probability for the event that the aggregate claims will be less than the total premium collected.

Example 3. *Let us return to the scenario of Example 1. In contrast to Example 1, we now use the imprecise model for computing lower and upper probabilities. First, we study the case $q = 0.1$ and $N^* = 50$. Using $s = 1$, we get the lower and upper probabilities, presented in Table 3, for the event that the aggregate claims will be less than the total premium collected. Next, suppose that the probability of the claim occurrence is $q = 0.01$. Then the corresponding probabilities that aggregate claims will exceed the total premium are shown in Table 4. Similar probabilities (Table 5 and Table 6) are obtained under the same conditions but using $s = 2$.*

Tables 3-6 show that the intervals of probabilities for different s are embedded in each other, e.g. for $K = 5$, $[0.596, 0.732] \subset [0.474, 0.745]$. This is another important feature of the imprecise statistical models which allows us to construct fuzzy probabilities and to analyze the insurance problem in terms of the fuzzy set theory, as will be explained in the next section.

Fig.1 illustrates the relationship between the upper probability $\overline{P}$ for the event that the aggregate claims will be less than the total premium collected (the thin curve), the lower probability $\underline{P}$ (the thick curve) and the precise probability P (the dashed curve) computed by applying Jeffreys prior [11] with parameters $a = 1/2$ and $b = 1/2$ (see Example 1) as functions of K, with $s = 1$. The same functions are depicted in Fig. 2 but with $s = 4$. It can be seen from the figures that the largest difference between lower and upper probabilities is observed for $K = 5$.

4.2 Imprecise Negative Binomial Model

Similar to the change of parameters in the previous model, in order to simplify the introduction and discussion of the imprecise probabilistic generalization of the standard model, we replace the parameters a and b in the expression for the negative binomial distribution $P(k)$ by s and α such that $a = s\alpha$ and $b = s$, so

$$P(k) = \frac{\Gamma(s\alpha+K+k)}{\Gamma(s\alpha+K)k!} \cdot \left(\frac{s+T}{s+T+t}\right)^{s\alpha+K} \left(\frac{t}{s+T+t}\right)^{k}.$$

Table 3: Lower and upper probabilities that aggregate claims will be less than the total premium ($s = 1$)

	K		
probabilities	4	5	6
lower	0.732	0.596	0.456
upper	0.848	0.732	0.596

Table 4: Lower and upper probabilities that aggregate claims will exceed the total premium ($s = 1$)

	K		
probabilities	0	1	2
lower	0	3.885×10^{-3}	2.232×10^{-2}
upper	3.885×10^{-3}	2.232×10^{-2}	6.880×10^{-2}

Table 5: Lower and upper probabilities that aggregate claims will be less than the total premium ($s = 2$)

	K		
probabilities	4	5	6
lower	0.612	0.474	0.346
upper	0.857	0.745	0.612

Table 6: Lower and upper probabilities that aggregate claims will exceed the total premium ($s = 2$)

	K		
probabilities	0	1	2
lower	0	3.550×10^{-3}	2.061×10^{-2}
upper	2.061×10^{-2}	6.419×10^{-2}	0.143

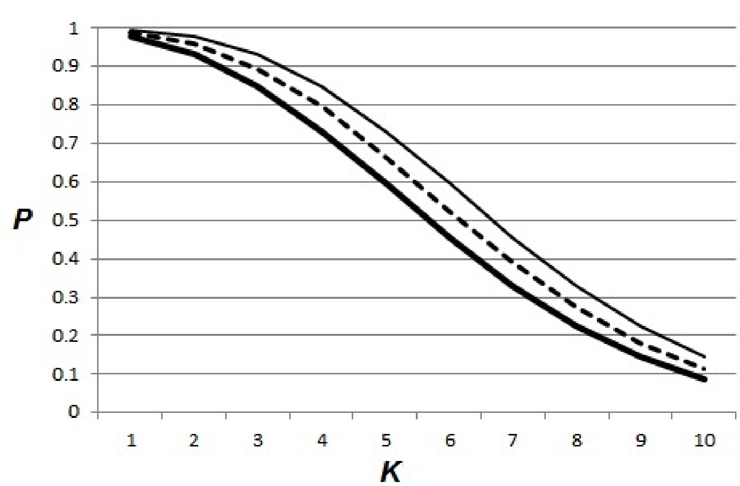

Figure 1: Probability P as function of K^* ($s = 1$)

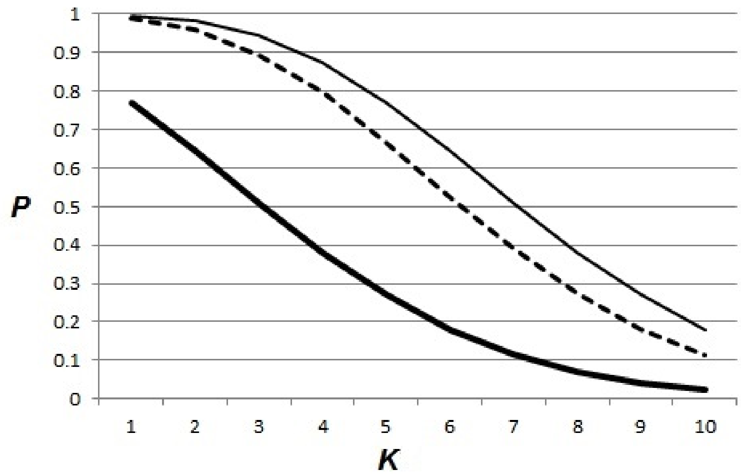

Figure 2: Probability P as function of K^* $(s = 4)$

It should be noted that the expectation of the number of claims X under conditions $K = 0$ and $N^* = 0$ is computed in terms of the introduced parameters as

$$\mathbb{E}X = \frac{a}{b} = \alpha.$$

The variance of X is of the form:

$$var(X) = \frac{a}{b} + \frac{a}{b^2} = \alpha + \frac{\alpha}{s}.$$

Therefore, the parameter α for the negative binomial distribution can be interpreted as the prior expected number of claims.

We consider the imprecise model defined as the set of all negative binomial distributions with fixed hyperparameter s and with arbitrary $\alpha \geq 0$. As for the previous model, because we are dealing with this set of distributions instead of a single distribution, we derive optimal lower and upper bounds for $P(k)$ or P, and these are lower and upper probabilities for the event of interest. These lower and upper probabilities can be obtained by minimizing and maximizing $P(k)$ or P over all values α in $[0, \infty)$.

According to this model, with the parameters s and α, the probability for the event that the aggregate claims will be less than the total premium collected over a time period of length t, is

$$P = \sum_{k=0}^{M} \frac{\Gamma(s\alpha + K + k)}{\Gamma(s\alpha + K)k!} \cdot \left(\frac{s+T}{s+T+t}\right)^{s\alpha+K} \left(\frac{t}{s+T+t}\right)^{k}.$$

Using Beta functions, this can also be written as

$$P = 1 - \frac{\mathrm{B}_q(M+1, r)}{\mathrm{B}(M+1, r)}.$$

where

$$r = s\alpha + K, \; q = \frac{t}{s+T+t},$$

and $\mathrm{B}_q(M+1, r)$ is the incomplete Beta-function,

$$\mathrm{B}_q(M+1, r) = \int_0^q x^M (1-x)^{r-1} \mathrm{d}x.$$

If $q = 1$, then the incomplete Beta-function is equal to the Beta-function, i.e., $\mathrm{B}_1(M+1, r) = \mathrm{B}(M+1, r)$. The following result simplifies the derivation of the lower and upper probabilities in the generalization of this model as presented in this paper.

Proposition 4. *The probability P for the event that the aggregate claims will be less than the total premium collected is decreasing as function of α.*

Proof. Denote $r+\Delta r = s(\alpha+\Delta\alpha)+K$, $\Delta\alpha > 0$. In order to prove that the probability P decreases, we have to prove the following inequality

$$1-\frac{\mathrm{B}_q(M+1,r)}{\mathrm{B}(M+1,r)} \geq 1-\frac{\mathrm{B}_q(M+1,r+\Delta r)}{\mathrm{B}(M+1,r+\Delta r)}$$

or

$$\frac{\int_0^q x^M(1-x)^{r-1}\mathrm{d}x}{\int_0^1 x^M(1-x)^{r-1}\mathrm{d}x} \leq \frac{\int_0^q x^M(1-x)^{r-1}(1-x)^{\Delta r}\mathrm{d}x}{\int_0^1 x^M(1-x)^{r-1}(1-x)^{\Delta r}\mathrm{d}x}. \tag{1}$$

Denote

$$\int_0^1 x^M(1-x)^{r-1}\mathrm{d}x = A,\ \int_0^q x^M(1-x)^{r-1}\mathrm{d}x = B,\ \int_q^1 x^M(1-x)^{r-1}\mathrm{d}x = C.$$

According to the second mean value theorem, we can write

$$\int_0^q x^M(1-x)^{r-1}(1-x)^{\Delta r}\mathrm{d}x = z_1 B,$$

$$\int_q^1 x^M(1-x)^{r-1}(1-x)^{\Delta r}\mathrm{d}x = z_2 C.$$

The right side of inequality (1) can be rewritten as

$$\frac{z_1 B}{z_1 B + z_2 C}.$$

Here $z_1, z_2 \geq 0$ are the values of $(1-x)^{\Delta r}$ at some points from 0 to 1 or mean values of integrals B and C in accordance with the second mean value theorem. Then the condition of monotonicity can be rewritten as

$$\frac{B}{A} \leq \frac{z_1 B}{z_1 B + z_2 C}.$$

Since the function $(1-x)^{\Delta r}$ is decreasing, then $z_1 \geq z_2$ or $z_1 = z_2 + \varepsilon$. By using the conditions $A = B + C$ and $B \leq A$, we obtain

$$\frac{z_1 B}{z_1 B + z_2 C} = \frac{z_1 B}{z_2 B + z_2 C + \varepsilon B} = \frac{z_1 B}{z_2 A + \varepsilon B} = \frac{z_2 B + \varepsilon B}{z_2 A + \varepsilon B} \geq \frac{z_2 B}{z_2 A} = \frac{B}{A},$$

as was to be proved. □

This proposition implies that the largest value of P over the class of distributions considered, so the upper probability $\overline{P}$, is achieved for $\alpha \to 0$ and is equal to

$$\overline{P} = 1 - \frac{\mathrm{B}_q(M+1,K)}{\mathrm{B}(M+1,K)}$$

which can also be written as

$$P = \sum_{k=0}^{M} \frac{\Gamma(K+k)}{\Gamma(K)k!} \cdot \left(\frac{s+T}{s+T+t}\right)^K \left(\frac{t}{s+T+t}\right)^k.$$

The proposition similarly implies that the smallest value of P over the class of distributions considered, so the lower probability $\underline{P}$, is achieved for $\alpha \to \infty$ and is equal to $\underline{P} = 0$. This can be simply proved by using the second mean value theorem as follows. Denote

$$\int_0^q x^M(1-x)^{K-1}\mathrm{d}x = A,\ \int_q^1 x^M(1-x)^{K-1}\mathrm{d}x = B.$$

Then

$$\underline{P} = 1 - \lim_{\alpha\to\infty} \frac{\int_0^q x^M(1-x)^{s\alpha+K-1}\mathrm{d}x}{\int_0^1 x^M(1-x)^{s\alpha+K-1}\mathrm{d}x} = 1 - \lim_{\alpha\to\infty} \frac{z_1^{s\alpha}A}{z_1^{s\alpha}A + z_2^{s\alpha}B}.$$

Here $z_1, z_2 \geq 0$ are the values of $(1-x)$ at some points from 0 to 1 or mean values of integrals B and C in accordance with the second mean value theorem. It follows from the condition $z_1 \geq z_2$ that

$$\lim_{\alpha\to\infty} \frac{z_1^{s\alpha}A}{z_1^{s\alpha}A + z_2^{s\alpha}B} = \lim_{\alpha\to\infty} \frac{1}{1+\left(\frac{z_2}{z_1}\right)^{s\alpha}\frac{B}{A}} = 1.$$

Hence $\underline{P} = 0$.

As in the previous model, if we consider this model before any data are included, so with $K = T = 0$, then $\underline{P} = 0$ and $\overline{P} = 1$. This again implies that the model provides vacuous lower and upper probabilities for this event of interest, before data are taken into account, and hence that it reflects complete lack of prior knowledge for this specific event.

However, the fact that the lower probability above is always zero, for any data set, is perhaps against intuition and rather inconvenient, it is of course due to the fact that the set of possible values for α is unlimited. To resolve this, we can change the model by restricting this set of possible values for α by choosing a suitable upper bound. We suggest to do this by considering a condition of successful working of the insurance company in terms of the expected values, i.e.,

$$\mathbb{E}\Pi(t) \geq \mathbb{E}R(t)$$

This immediately leads to the condition

$$cN \geq y \cdot \mathbb{E}X = y\alpha.$$

which provides a convenient constraint for α,

$$\alpha \leq \frac{cN}{y}.$$

This is nothing else but the largest number of claims M which can be paid by the insurer, i.e., $\alpha \leq M$. This ensures that the lower probability considered above is no longer always equal to 0, in fact it is easily derived to be equal to

$$\underline{P} = 1 - \frac{\mathrm{B}_q(M+1, sM+K)}{\mathrm{B}(M+1, sM+K)}.$$

Of course, with this limited range of values for α, we again would get the precise model in the case $s = 0$, yet this keeps having the restriction of being unable to reflect lack of prior information adequately.

Example 5. *We return to Example 2 and apply the imprecise model to derive the lower and upper probabilities for the event that the aggregate claims will be less than the total premium collected, over a time period of length* $t = 3$. *We again take* $T = 2$, *and* $K = 2$. *If* $K = 2$ *and* $s = 0.5$, *then* $\underline{P} = 0.761$, $\overline{P} = 0.997$. *If* $K = 2$ *and* $s = 1$, *then* $\underline{P} = 0.423$, $\overline{P} = 0.999$. *The precise probability when taking* $s = 0$ *is* 0.973. *We can see that the interval of probabilities when taking* $s = 1$ *is rather large. The same can be seen from curves shown in Fig. 3, which depict the upper* $\overline{P}$, *lower* $\underline{P}$ *and precise* P *probabilities for the event that aggregate claims will be less than the total premium collected, as functions of* K *and with* $s = 0.5$. *By increasing* s *to* 1, *we get wider intervals of probabilities (see the corresponding Fig. 4).*

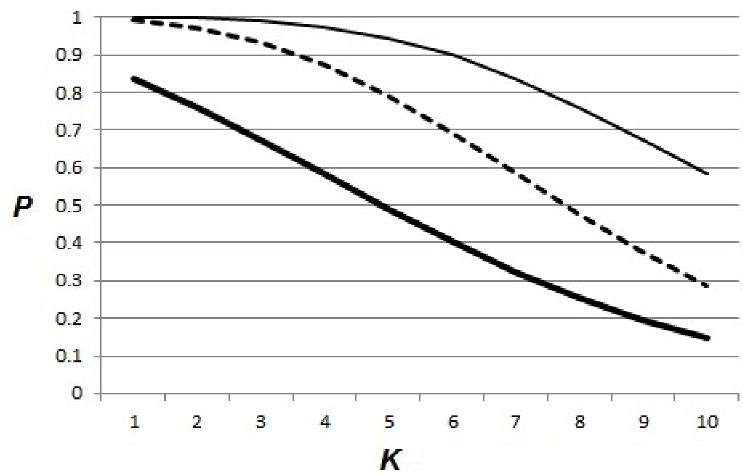

Figure 3: Probability P as function of K^* ($s = 0.5$)

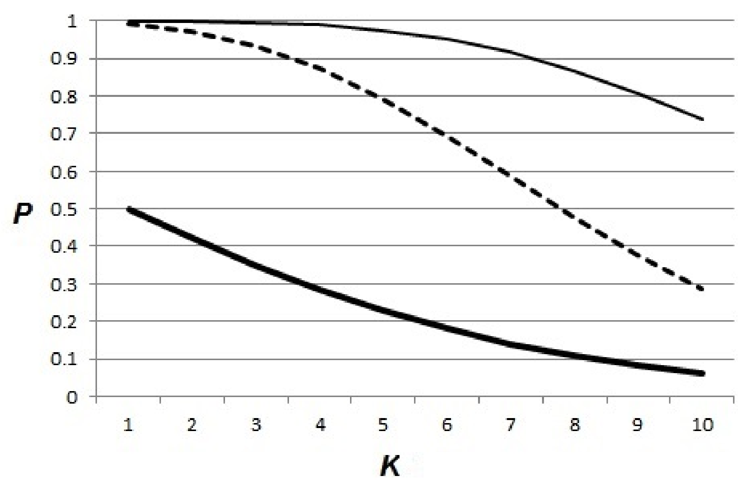

Figure 4: Probability P as function of K^* ($s = 1$)

5 The Hyperparameter and Fuzzy Probabilities

One of the important questions for the imprecise probability insurance models presented in this paper involves the choice of the hyperparameter s. There is an argument in favour of setting s as small as possible in order to avoid large imprecision. On the other hand, small values of s may lead to risky insurance decisions and fail to reflect lack of information adequately. A unique answer to the question how to select the hyperparameter s, that would be suitable in all cases, cannot be provided. Moreover, the meaning of s may not be clear to practitioners and different values of s typically lead to different optimal decisions.

A natural way to overcome this difficulty is to find a way to avoid the procedure of selecting the hyperparameter. In other words, we could consider all possible intervals $[\underline{P}(s), \overline{P}(s)]$ for all possible s, where obvious new notation is used to emphasize the dependence of the lower and upper probabilities for the event of interest on the value of s. A nice property of both imprecise probability insurance models introduced in this paper is that the intervals $[\underline{P}(s), \overline{P}(s)]$ are nested. This property allows us to construct a fuzzy probability [7] for the event that aggregate claims will be less than the total premium collected.

Note that $\underline{P}(s) = \overline{P}(s) = P$ in case $s = 0$. Moreover, $\underline{P}(s)$ decreases and $\overline{P}(s)$ increases as s increases. Let us introduce a function $\mu(s)$ with the following properties: $\mu(0) = 1$, $\mu(\infty) = 0$, and $\mu(s)$ is non-increasing with s. Examples of such functions are $\mu(s) = \exp(-s)$, $\mu(s) = (1+s)^{-1}$. So, we have a set of nested intervals characterized by real numbers $\mu \in [0,1]$. This set can be viewed as a fuzzy probability $\widetilde{P}$ for the event of interest, with the membership function μ. Every interval $[\underline{P}(s), \overline{P}(s)]$ can now be regarded as an α-cut set [7] of the fuzzy set $\widetilde{P}$, where $\alpha = \mu(s)$.

Example 6. *We return to Example 2 and find the fuzzy probabilities $\widetilde{P}$ for the event that aggregate claims will exceed the total premium under the same initial parameters of insurance. Fig. 5 shows the membership functions μ of the fuzzy probabilities for $K = 4, 5, 6$. Here we take $\mu(s) = (1+s)^{-1}$. Fig. 6 illustrates the same fuzzy probabilities, but under condition that $\mu(s) = \exp(-s)$. One can see from the figures that the form of the function $\mu(s)$ significantly impacts on the fuzzy probability when applying the imprecise Beta-binomial model.*

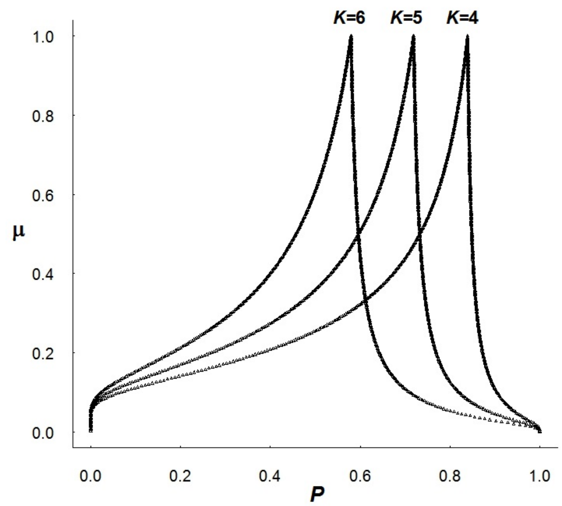

Figure 5: The fuzzy probabilities P with $\mu(s) = (1+s)^{-1}$

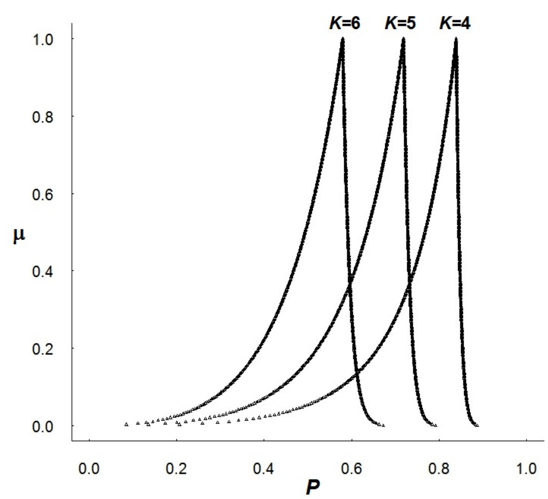

Figure 6: The fuzzy probabilities P with $\mu(s) = \exp(-s)$

Example 7. *We return to Example 5 and find the fuzzy probabilities $\widetilde{P}$ for the event that the aggregate claims will exceed the total premium under the same initial parameters of insurance. Figs. 7-8 show the membership functions μ of the fuzzy probabilities with $\mu(s) = (1+s)^{-1}$ and $\mu(s) = \exp(-s)$, respectively. It can be seen from the figures that the form of the function $\mu(s)$ does not significantly impact on the fuzzy probability when applying the imprecise negative binomial model.*

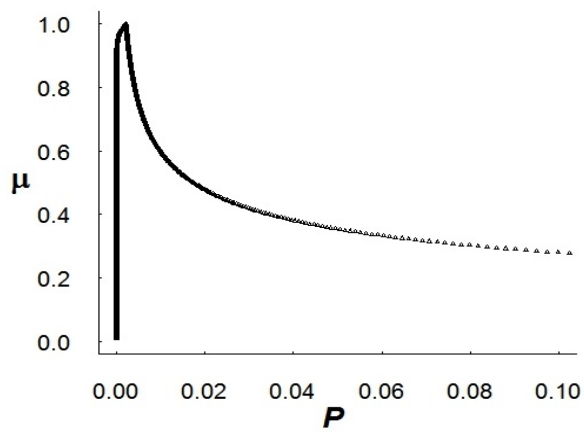

Figure 7: The fuzzy probability P with $\mu(s) = (1+s)^{-1}$

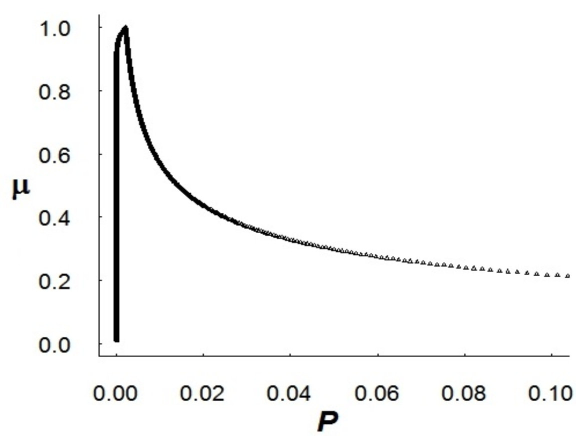

Figure 8: The fuzzy probability P with $\mu(s) = \exp(-s)$

This suggestion to link the imprecise probability models to theory of fuzzy probabilities, in order to select the hyperparameter s, provides one avenue of research that may be attractive in certain applications. Of course, the analyst will need to explain fuzzy probabilities and numbers, if using this approach to elicit information from topic specialists, in order to assess a suitable value for s. This is important in order to strike the right balance between caution and the risk of being too cautious, the latter case could lead to inability to use the model for decision support.

6 Conclusion

Two imprecise probabilistic individual risk models of insurance have been proposed in this paper. The models use a combination of ideas underlying the Bayesian approach and imprecise statistical models. The results of the insurance analysis are intervals of probabilities for the event that aggregate claims will be less than the total premium collected, as functions of the hyperparameter s of the imprecise statistical models. In order to overcome the difficulty of selecting the hyperparameter, a link to simple fuzzy models has been proposed. The fuzzy representation of the probabilities may actually be useful in extending these models, and the general approach, to more compex scenarios, this provides interesting topics for future research.

It is important to mention that real-valued probabilities, which are of course a special case of interval-valued probabilities, may be more suitable in some applications. In particular if there is much information available and no need to be cautious, the standard approach probably suffices to solve real problems and it is probably easier understood by practitioners. It is important, however, to realize that lower and upper probabilities reflect the amount of information available on which the conclusions of the model application are based.

Typically, a lower probability reflects the amount of information available in favour of the event of interest while the upper probability reflects the amount of information available against the event of interest. When information is limited, this is reflected through the imprecision. Therefore, if one has limited information and wishes to be cautious, then the interval-valued probabilities are attractive for use in decision support. Mostly, we would recomment considering the real-valued probabilities as well as the interval-valued probabilities, and the latter for a range of values of the hyperparameter s in these models, in order to get a good picture of the scenario and, ideally, some insight into the robustness of a final decision.

It should be noted that only the simplest individual risk models of insurance have been considered in this paper. This was due to our explicit aim to present an idea of the possibilities to generalize the well-known basic models by the use of imprecise probabilities. For the presented models, the computational algorithms for calculating insurance probabilities were presented. For more complex models and applications, for example, multivariate insurance models [1], both suitable ways to generalize to imprecise probabilities and the corresponding computational algorithms will be more demanding, these provide interesting challenges for future research.

Acknowledgements

The authors are grateful to a reviewer whose careful reading of our paper enabled us to improve the presentation.

References

[1] Anastasiadis, S., and S. Chukova, Multivariate insurance models: an overview, *Insurance: Mathematics and Economics*, vol.51, no.1, pp.222–227, 2012.

[2] Augustin, T., Coolen, F.P.A., de Cooman, G., and M.C.M. Troffaes, (Eds.) *Introduction to Imprecise Probabilities*, Wiley, Chichester, 2014.

[3] Bernardo, J.M., and A.F.M. Smith, *Bayesian Theory*, Wiley, Chichester, 1994.

[4] Bradlow, E.T., Hardie, B.G.S., and P.S. Fader, Bayesian inference for the negative binomial distribution via polynomial expansions, *Journal of Computational and Graphical Statisitics*, vol.11, no.1, pp.189–201, 2002.

[5] Carlier, G., Dana, R., and N. Shahdi, Efficient insurance contracts under epsilon-contaminated utilities, *The Geneva Papers on Risk and Insurance Theory*, vol.28, pp.59–71, 2003.

[6] Dionne, G., and C. Vanasse, A generalization of automobile insurance rating model: the negative binomial distribution with a regression component, *ASTIN Bulletin*, vol.19, no.2, pp.199–212, 1989.

[7] Dubois, D., and H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York, 1988.

[8] Fader, P.S., Hardie, B.G.S., and K.L. Lee, "Counting your customers" the easy way: an alternative to the Pareto/NBD model, *Marketing Science*, vol.24, no.2, pp.275–284, 2005.

[9] Huang, T., Zhao, R., and W. Tang, Risk model with fuzzy random individual claim amount, *European Journal of Operational Research*, vol.192, no.3, pp.879–890, 2009.

[10] Huber, P.J., *Robust Statistics*, Wiley, New York, 1981.

[11] Jeffreys, H., *Theory of Probability*, Oxford University Press, Oxford, 1961.

[12] Jeleva, M., Background risk, demand for insurance, and choquet expected utility preferences, *The GENEVA Papers on Risk and Insurance Theory*, vol.25, no.1, pp.7–28, 2000.

[13] Jeleva, M., and B. Villeneuve, Insurance contracts with imprecise probabilities and adverse selection, *Economic Theory*, vol.23, no.4, pp.777–794, 2004.

[14] Jerath, K., Fader, P.S., and B.G.S. Hardie, New perspectives on customer "death" using a generalization of the Pareto/NBD model, *Marketing Science*, vol.30, no.5, pp.866–880, 2011.

[15] Lemaire, J., Fuzzy insurance, *ASTIN Bulletin*, vol.20, no.1, pp.33–55, 1990.

[16] Liu, Y.J., and W.G. Zhang, Fuzzy portfolio optimization model under real constraints, *Insurance: Mathematics and Economics*, vol.53, no.3, pp.704–711, 2013.

[17] Moral, S., Imprecise probabilities for representing ignorance about a parameter, *International Journal of Approximate Reasoning*, vol.53, no.3, pp.347–362, 2012.

[18] Perks, W., Some observations on inverse probability including a new indifference rule (with discussion), *Journal of the Institute of Actuaries*, vol.73, pp.285–334, 1947.

[19] Quaeghebeur, E., and G. de Cooman, Imprecise probability models for inference in exponential families, *Proceeding of the 4th International Symposium on Imprecise Probabilities and Their Applications, ISIPTA'05*, 2005.

[20] Shapiro, A.F., Fuzzy logic in insurance, *Insurance: Mathematics and Economics*, vol.35, no.2, pp.399–424, 2004.

[21] Shapiro, A.F., Modeling future lifetime as a fuzzy random variable, *Insurance: Mathematics and Economics*, vol.53, no.3, pp.864–870, 2013.

[22] Shi, P., and E.A. Valdez, Multivariate negative binomial models for insurance claim counts, *Insurance: Mathematics and Economics*, vol.55, pp.18–29, 2014.

[23] Syversveen, A.R., Noninformative Bayesian priors. interpretation and problems with construction and applications, *Preprint Statistics*, no.3, 1998.

[24] van Oest, R., and G. Knox, Extending the BG/NBD: a simple model of purchases and complaints, *International Journal of Research in Marketing*, vol.28, pp.30–37, 2011.

[25] Walley, P., *Statistical Reasoning with Imprecise Probabilities*, Chapman and Hall, London, 1991.

[26] Walley, P., Inferences from multinomial data: learning about a bag of marbles, *Journal of the Royal Statistical Society, Series B*, vol.58, pp.3–57, 1996 (with discussion).

[27] Walley, P., A bounded derivative model for prior ignorance about a real-valued parameter, *Scandinavian Journal of Statistics*, vol.24, no.4, pp.463–483, 1997.