Imprecise inference for warranty contract analysis

Lev V. Utkin[1], Frank P.A. Coolen[2a], Sergey V. Gurov[1]

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

[2] *Department of Mathematical Sciences, Durham University, UK*

Abstract

This paper presents an investigation into generalized Bayesian analysis of warranty contracts, using sets of prior distributions within the theory of imprecise probability. Explicit expressions are derived for optimal lower and upper bounds for the expected profit for the manufacturer of a product, corresponding to an imprecise negative binomial model for which two sets of prior distributions are studied. The results can be used to set a maximum value of compensation such that the manufacturer's expected profit remains positive, under vague prior knowledge.

Keywords: Generalized Bayesian analysis, imprecise probability, lower and upper expected profit, warranties

1. INTRODUCTION

Warranties are important aspects of many contracts between consumers and manufacturers. Typically, decisions about such contracts must be made at an early stage, when the available knowledge about the product reliability might be vague. While the Bayesian approach is attractive to investigate warranties, meaningfully assigning a single prior distribution might be difficult and it might not fully reflect available information. In particular, if one attempts to model lack of prior information, the generalized Bayesian approach using theory of imprecise probability, in which sets of prior distributions are used instead of a single prior distribution, provides an attractive framework for inference that can be used to analyse warranty contracts.

[a] Corresponding author: frank.coolen@durham.ac.uk

An introduction to general theory of imprecise probability has been presented by Augustin et al [1], while an earlier detailed mathematical introduction to such theory was presented by Walley [2]. Introductions and overviews of imprecise probability with specific attention to topics in reliability and risk have been presented by the current authors [3, 4, 5]. The problem studied in this paper concerns a basic model for warranties, proposed by Singpurwalla [6] and also mentioned by Aven [7]. It does not include detailed analysis of real-world warranty data, which is an important and challenging topic which could benefit from analysis with the use of statistical methods based on imprecise probabilities. Recent contributions to statistical methods for analysis of real-world warranty data, including many further references, have been presented by Wu [8] and Gupta et al [9]. Standard Bayesian analysis of warranty claim data has been proposed by many authors, for example, Stephens and Crowder [10], Chen and Popova [11], Wu and Huang [12], Akbarov and Wu [13].

Section 2 introduces the basic setting for the analysis of warranty contracts considered in this paper. Section 3 presents a standard Bayesian approach for such an analysis, which is generalized through the use of an imprecise probability model in Section 4. While this model is closely related to popular imprecise probability models, it has a quite obvious disadvantage which is addressed in Section 5, effectively by using a restricted set of prior distributions. The main results presented in this paper are explicit expressions for the lower and upper expected profits for the manufacturer with a specified warranty contract. These are optimal lower and upper bounds and they enable valuation of compensation under this contract in order for the expected profit to remain positive. The presented imprecise probability models only assume vague prior knowledge and explicitly reflect this through these lower and upper bounds. The results are illustrated by examples in the respective sections. The paper ends with some concluding remarks in Section 6. Detailed proofs of the propositions in this paper are presented in an appendix.

2. WARRANTY CONTRACT ANALYSIS

Consider a scheme of typical warranty contracts as proposed by Singpurwalla [6] and also considered by Aven [7]. The scheme models the exchange of items from a large collection of similar items between a manufacturer (seller A) and a consumer (buyer B).

Let n be the number of items that the buyer would like to purchase. These items are supposed to be exchangeable with regard to their intended functioning and to have independent and identically distributed (iid) failure behaviour. Each item is required to be used for τ units of time, so the iid assumption for their lifetimes implies that each item meets this requirement with the same probability and the random success of any item in doing so is independent of that of other items, conditional upon the value of this probability. Throughout this paper, and in line with common practice, the probability of an item functioning successfully over the period considered is assumed to be high, so failures are relatively rare. It is assumed that an item can only fail once.

Suppose that the buyer B is willing to pay x monetary units, say dollars, per item, and is prepared to tolerate at most a total of z failures for all the n units in the time interval $[0, \tau]$. For each failure in excess of z, the buyer B needs to be compensated at the rate of y dollars per item. In effect, the quantity τ can be viewed as the duration of a warranty. One of the questions of interest is determination of the maximum compensation y per item in order for the seller to keep a non-negative expected profit.

Suppose that it costs c dollars to produce a single unit of the item, then the sale of n units at price x leads to income $n(x - c)$ dollars for seller A. If the buyer B experiences z or fewer failures in $[0, \tau]$, then this income is equal to A's profit. However, if B experiences $i > z$ failures in $[0, \tau]$ then A's liability is $(i - z)y$ leading to total profit of $n(x - c) - (i - z)y$ dollars.

Formally, the number of failing items in the given time period of length τ should be modelled by a Binomial distribution. However, due to the reasonable assumption that failures during this period are relatively rare, it is common practice [6] to use the Poisson distribution as approximation, this simplifies computation and is assumed henceforth in this paper. In this model, the parameter reflecting the quality of the items is the failure rate λ, which represents the average number of failing items among n items during a unit time interval. Let $p(i|\lambda)$ denote the probability for the event that exactly i items will fail during the time interval $[0, \tau]$. For known value of the parameter λ, this probability is

$$p(i|\lambda) = \frac{(\lambda\tau)^i \exp(-\lambda\tau)}{i!}. \tag{1}$$

Note that, while these probabilities are positive for all integers $i \geq 0$, the assumption that items will only fail quite rarely implies that $p(i|\lambda)$ for $i > n$

will be neglectably small, hence the approximation mentioned above remains reasonable. The corresponding *expected profit* for seller A, denoted by G (for 'gain'), for known value of λ, is

$$\mathbb{E}_\lambda G = n(x-c) - y \sum_{i=z+1}^{n} (i-z)p(i|\lambda). \tag{2}$$

In this paper, the scenario considered is that seller A will aim at non-negative expected profit, so $\mathbb{E}_\lambda G \geq 0$. Of course, this could be replaced with a different target for the expected profit, the mathematical analysis would be easily adapted and is not discussed further. If the seller has strong background information concerning the failure rate λ, it may be possible to consider it to be known. However, in many applications such information is not available. The Bayesian approach, reviewed in the following section, is the standard method for dealing with a not fully known failure rate.

3. STANDARD BAYESIAN APPROACH

If the parameter λ is unknown, it can be considered as a random quantity for which a probability density function $\pi(\lambda|\theta)$ can be assumed. In this case, the Bayesian approach can be applied for computing the expected profit, which is determined as the unconditional expected value

$$\mathbb{E}G = \int_\Omega \mathbb{E}_\lambda G \cdot \pi(\lambda|\theta)\mathrm{d}\lambda = n(x-c) - y \sum_{i=z+1}^{n} (i-z) \int_\Omega p(i|\lambda) \cdot \pi(\lambda|\theta)\mathrm{d}\lambda.$$

Here θ is the vector of parameters of π and $\Omega = \mathbb{R}_+$ is the set of possible values of λ. The corresponding probability of exactly k failures occurring during a period of length τ is

$$P(k) = \int_\Omega p(k|\lambda) \cdot \pi(\lambda|\theta)\mathrm{d}\lambda.$$

The Bayesian approach enables prior information, mainly based on expert judgement, to be combined with data. Suppose that the prior distribution $\pi(\lambda|\theta)$ reflects the expert's opinion about the possible values for λ prior to collecting any information. Suppose that data become available of the following form: n items have been tested for m periods, which can be of variable length. Suppose that the number of failing items, out of n, during period $j \in \{1, \ldots, m\}$ is k_j, and that the length of this period is τ_j. It should be noted that a more general scenario, with numbers of items being tested during the different periods not being equal to n, is quite straightforward to analyse following a similar setting but with the parameter λ explicitly related to a single item; this is left as an exercise for the reader, the current restriction simplifies the presentation and does not really limit the model with regard to the main new results as presented in the following sections.

It is convenient in Bayesian analysis to choose a prior distribution such that resulting computations in order to derive the posterior distribution are easy, which particularly occurs when a conjugate prior distribution is used. This leads to a posterior distribution belonging to the same parametric family of distributions as the prior distribution [14]. The Gamma distribution is a conjugate prior for the parameter λ of the Poisson distribution. Its parameters are $\theta = (a, b)$, with $a > 0, b > 0$, and it has the probability density function

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

where $\Gamma(a)$ is the gamma function.

Suppose that data become available for n items during m time periods, as described above, and let $K = k_1 + \ldots + k_m$ and $T = \tau_1 + \ldots + \tau_m$. The corresponding posterior predictive probability for the event that, out of n further items, k will fail during a time period length τ can be derived by standard Bayesian methods [14]. These probabilities, for $k \geq 0$, are given by the Negative Binomial distribution and are equal to

$$P(k) = \frac{\Gamma(a+K+k)}{\Gamma(a+K)k!} \left(\frac{b+T}{b+T+\tau}\right)^{a+K} \left(\frac{\tau}{b+T+\tau}\right)^k. \tag{3}$$

Note again that these $P(k)$ are positive for all $k \geq 0$, but with relatively few items failing these probabilities for $k > n$ will be extremely small, ensuring that the approximate model does not lead to complications.

Returning to the warranty model analyzed in this paper, as introduced in Section 2, the expected profit for the manufacturer is equal to

$$\mathbb{E}G = n(x-c) - y \sum_{k=z+1}^{n} \frac{(k-z)\Gamma(a+K+k)}{\Gamma(a+K)k!} \left(\frac{b+T}{b+T+\tau}\right)^{a+K} \left(\frac{\tau}{b+T+\tau}\right)^k. \tag{4}$$

This standard Bayesian scenario is illustrated by Example 1.

Example 1. *Suppose a buyer is considering to purchase* $n = 100$ *items at a cost of* $x = 20$ *dollars per item. Suppose that it costs* $c = 16$ *to produce each single item. Let the time period considered be of length* $\tau = 1$ *and suppose that the buyer would be willing to accept only* $z = 1$ *failure for all 100 items during this period. Suppose that also 100 items have been tested, over three time periods which were all also of length 1, and assume that the numbers of failing items per time period were* $k_1 = 0$, $k_2 = 1$, $k_3 = 1$. *To derive the posterior probabilities for the model described above, the sufficient statistics for these data are the total length of the periods for the tests,* $T = 3$, *and the total number of failing items during these tests,* $K = k_1 + k_2 + k_3 = 2$. *Assume that the prior distribution was the Gamma distribution with parameters* $a = 1$ *and* $b = 1$, *then the corresponding expected profit is equal to*

$$\mathbb{E}G = 100 \cdot 4 - y \sum_{k=2}^{100} \frac{(k-1)\Gamma(3+k)}{\Gamma(3)k!} \left(\frac{4}{5}\right)^3 \left(\frac{1}{5}\right)^k = 100 \cdot 4 - y \cdot 0.262.$$

This implies that the expected profit for the seller A is nonnegative if and only if

$$100 \cdot 4 - y \cdot 0.262 \geq 0$$

and hence that A would be willing to pay up to $1,527$ *dollars in compensation per item, for any number of items that would fail during the warranty period of length 1 apart from one item, which was the number deemed to be acceptable to fail by the buyer B. Of course, there are likely to be further aspects which the seller A may need to take into account to set a realistic level of compensation, for example additional costs and risk of large losses. Including such aspects is conceptually straightforward in combination with the posterior distribution presented here.*

It should be noted that the negative binomial distribution is widely applied in many areas, including marketing research, insurance and risk management, where events of interest are relatively rare. When applying the Bayesian approach, as outlined in this section, the choice of prior distribution in general, or more specifically the choice of the parameters a and b if one assumes the conjugate Gamma prior distribution, is nontrivial if one has little or no meaningful prior information about the frequency of failures. It has become a standard procedure, in such situations, to use a so-called 'noninformative' prior distribution.

Many methods for determining noninformative prior distributions in the Bayesian framework have been proposed in the literature. Many methods apply the Bayes-Laplace postulate, which is also known as the principle of insufficient reason [15]. According to this principle, the prior distribution should be uniform. However, this choice meets some difficulties or problems. The first problem is that the uniform distribution is not invariant under reparametrization. If one has no information, for instance, about a parameter ϕ, then one also has no information about $1/\phi$, but a uniform distribution for ϕ does not correspond to a uniform distribution for $1/\phi$. Another possible 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. The corresponding posterior distribution may well be proper, but for example in case of data containing zero failures, in the setting considered in this paper, the posterior distribution may remain improper, hence not enabling conclusions in terms of expected values. Walley [16] gives a number of examples illustrating possible problems and shortcomings of the principle of insufficient reason. A detailed review of other methods for constructing a noninformative prior has been presented by Syversveen's [17].

Another interesting approach to modelling absence of prior information within a Bayesian framework of statistics is based on using a class $\mathcal{M}$ of prior distributions instead of a single prior distribution [1, 16]. This overcomes many of the problems related to selecting a single noninformative prior. Typically, when interest is in an event U, the class of prior distributions is reflected by corresponding lower and upper probabilities for the event U, denoted by $\underline{P}(U)$ and $\overline{P}(U)$, respectively, which are defined by as

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

As pointed out by Walley [16] and Syversveen [17], the class $\mathcal{M}$ is "not a class of reasonable priors, but a reasonable class of priors". This means that each single member of the class is not a reasonable model for prior ignorance, because no single distribution can model ignorance satisfactorily. However, the whole class can be considered to provide a reasonable model for prior ignorance. When one has 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. One can interpret this, in terms of precise probabilities, as reflecting that the prior probability for the event of interest can be anywhere within the range from 0 to 1. It is particularly attractive to define such classes of prior distributions in a manner such that the conjugacy property is maintained, hence leading to quite straightforward updating of the prior class to a class of posterior distributions when data become available. Examples of such models are Walley's imprecise Dirichlet model [16], which has been applied to a variety of scenarios in reliability [18, 19, 20], and, more generally, imprecise probability models for inference in exponential families [21]. In the following sections, two related models using classes of prior distributions in the generalized Bayesian framework are presented for the basic warranty problem discussed in this paper.

4. IMPRECISE NEGATIVE BINOMIAL MODEL I

To simplify presentation, it is convenient to let us replace the parameters a and b in the expression for the Negative Binomial distribution $P(k)$ as by α and s, such that $a = s\alpha$ and $b = s$. This replacement is proposed by Quaeghebeur and de Cooman [21] in their paper devoted to imprecise models for inference in exponential families, and follows a similar parametrization used by Walley [16] for the imprecise Dirichlet model. Then

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

A convenient way to construct an imprecise probability model is by using the set of all Negative Binomial distributions with fixed hyperparameter s and with arbitrary $\alpha \geq 0$. Note that this corresponds to applying the same reparametrization for the underlying Gamma prior distribution for the parameter λ of the Poisson model, and taking the corresponding class of prior distributions. By dealing with the set of distributions instead of a single distribution, one derives lower and upper bounds for $\mathbb{E}G$ instead of a precise value as corresponds to a single distribution, which is in line with the lower and upper probabilities as discussed above. These lower and upper bounds can be obtained by minimizing and maximizing $\mathbb{E}G$ over all values of α in $[0, \infty)$.

The expected profit for the seller A, with the replaced parameters, is of the form

$$\mathbb{E}G = n(x-c) - y \sum_{k=z+1}^{n} (k-z) \frac{\Gamma(s\alpha + K + k)}{\Gamma(s\alpha + K) \cdot k!} \left(\frac{s+T}{s+T+\tau} \right)^{s\alpha + K} \left(\frac{\tau}{s+T+\tau} \right)^{k}.$$

The hyperparameter $s > 0$ determines the influence of the prior distribution on posterior probabilities and the expected profit. In particular, if $s = 0$, then the posterior distribution

$$P(k) = \frac{\Gamma(K+k)}{\Gamma(K)k!} \left(\frac{T}{T+\tau}\right)^K \left(\frac{\tau}{T+\tau}\right)^k.$$

is totally determined only by information in the form of K and T.

Proposition 1. *If $y \geq 0$, then the expected profit $\mathbb{E}G$ as a function of the parameter α has a single minimum.*

The expected number of failures X under conditions $K = 0$ and $T = 0$ is computed as

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

So the parameter α for the Negative Binomial distribution can be interpreted as the prior expected number of failing items out of n items, which are all used for a time period of unit length. Assuming that the prior expected number of failures may be arbitrary from 0 to ∞ and using Proposition 1, the following optimal lower and upper bounds for the expected profit are derived corresponding to the model presented in this section. Note that the upper expected profit is not proven as it is straightforward.

Proposition 2. *The upper expected profit, denoted by $\overline{\mathbb{E}}G$, is achieved for $\alpha \to \infty$ and is equal to*

$$\overline{\mathbb{E}}G = n(x - c).$$

It follows from Proposition 2 that the upper bound for the expected profit is noninformative. It assumes an ideal case when we get the maximally possible expected profit. The value of compensation y can be accepted arbitrarily in this case.

Proposition 3. *The lower bound for the expected profit, denoted by $\underline{\mathbb{E}}G$, is achieved at the point α_0 which is the root of the equation*

$$\sum_{k=z+1}^{n} Z_k \cdot \frac{\Gamma(s\alpha + K + k)}{\Gamma(s\alpha + K)} v^{s\alpha + K} \left(\sum_{i=0}^{k-1} \frac{1}{s\alpha + K + i} + \ln v\right) = 0, \qquad (5)$$

where

$$Z_k = \frac{(k-z)(1-v)^k}{k!}, \quad v = \frac{s+T}{s+T+\tau}.$$

The point α_0 belongs to interval $[\alpha_1, \alpha_{n-z}]$. Here α_1 and α_{n-z} are roots of equations

$$\sum_{i=0}^{z} \frac{1}{s\alpha + K + i} + \ln\left(\frac{s+T}{s+T+\tau}\right) = 0,$$

$$\sum_{i=0}^{n-1} \frac{1}{s\alpha + K + i} + \ln\left(\frac{s+T}{s+T+\tau}\right) = 0,$$

respectively.

Proposition 3 provides a simple way for numerical computation of the lower bound for the expected profit $\underline{\mathbb{E}}G$. According to the proof of this proposition, equation (5) has a unique root. Moreover, bounds for possible values of the root can be simply computed. This implies that (5) can be solved by means of one of the well-known numerical methods, for example, gradient methods or the bisection method.

Note that the limit value of the expected profit for $\alpha \to 0$, before any observations, is $n(x-c)$, i.e., values of the expected profit for $\alpha \to \infty$ and for $\alpha \to 0$ coincide and are equal to the upper bound $\overline{\mathbb{E}}G$. This interesting fact is explained in the following way by setting $K = T = 0$. The case $\alpha \to 0$ means that the number of failing items tends to zero and seller A does not need to compensate the failed items. Non-zero probabilities of failures $P(k)$ by $\alpha \to \infty$ are concentrated at values $k \gg n$. The restricted value of n is the main reason of the unexpected behaviour of the lower bound for the expected profit $\underline{\mathbb{E}}G$ as a function of the parameter α. This fact gives us the idea to study the expected profit under condition of large values of n.

Proposition 4. *If $y \geq 0$ and $n \to \infty$, then the expected profit $\mathbb{E}G$ is a decreasing function of the parameter α.*

Proposition 4 implies that we can apply the property of monotonicity of the expected profit for very large values of n.

Example 2. *Consider again the scenario of Example 1, which is now used to illustrate the imprecise model presented in this section, in order to compute an interval for the values of y corresponding to nonnegative expected lower and upper bounds for the profit for seller A. Assume that the hyperparameter s is set equal to 1, then the upper bounds for the expected profit is $\overline{\mathbb{E}}G = 400$.*

The lower bound for the expected profit is the root of the equation

$$\sum_{k=2}^{100} \frac{(k-1)\Gamma(\alpha+2+k)}{\Gamma(\alpha+2)\,k!} 0.8^{\alpha+2} \cdot 0.2^k \left(\ln 0.8 + \sum_{i=0}^{k-1} \frac{1}{\alpha+2+i} \right) = 0,$$

which is $\alpha_0 = 330$*. At that, the bounds for* α_0 *can also be computed using Proposition 3. They are* $\alpha_1 = 2.48$ *and* $\alpha_{n-z} = 398.5$*. The function*

$$Q(\alpha) = \sum_{k=z+1}^{n} (k-z) P(k,\alpha)$$

is shown in Fig. 1. Moreover, the function $Q(\alpha)$ *at point* $\alpha_0 = 330$ *has value* $Q(330) = 77.112$*. The pessimistic value of* y *can be derived from the equation* $100 \cdot 4 - y \cdot 77.112 = 0$*. The solution is* $y = 5.19$*.*

Hence the upper bound for y *is undetermined. The lower bound for* y *is determined from the equation* $\underline{\mathbb{E}}G = 0$ *and is* $\underline{y} = 5.19$*. This means that, in the worst case scenario with all items failing, compensation of* 5.19 *dollars per item would be the maximum in order to avoid a loss, which follows from the difference between the selling price and production costs. This situation is addressed further in the following section.*

In order to compare these values it is interesting just to apply the Poisson distribution with the parameter $\lambda = K/T = 2/3$ *in the standard warranty model, so without learning in the Bayesian framework. The corresponding expected profit is*

$$\begin{aligned} \mathbb{E}G &= 100 \cdot 4 - y \sum_{i=2}^{100} (i-1) \frac{(2/3)^i \exp(-2/3)}{i!} \\ &= 100 \cdot 4 - y \cdot 0.18. \end{aligned}$$

This would lead to $y = 2,222$ *as maximum possible compensation in order for* A *to keep nonnegative expected profit. Clearly, the above calculated lower value for* y *forms an interval which contains this value corresponding to replacing* λ *by the empirical value* K/T*.*

5. IMPRECISE NEGATIVE BINOMIAL MODEL II

The imprecise Negative Binomial model I, as presented in the previous section, has one major problem when applied to the warranty model considered in this paper. Namely, the upper bound for the expected profit, as

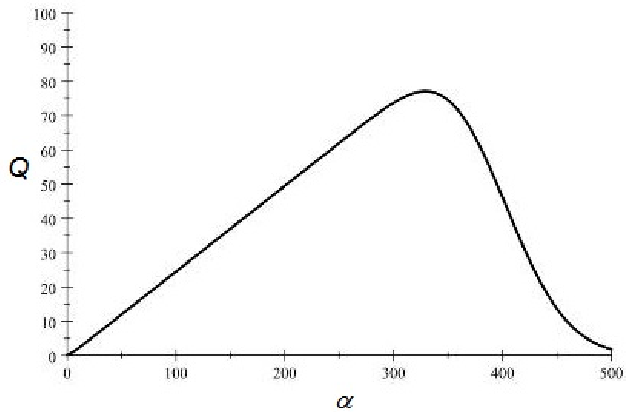

Figure 1: The function $Q(\alpha)$

derived in Proposition 2, does not depend on data observations due to it being attained for the limit situation $\alpha \to \infty$. Therefore, another model for constructing a set of Negative Binomial distributions for the warranty model is now proposed. Similar to model I, this is based on the generalized Bayesian approach with a class of Gamma prior distributions for the parameter λ of the Poisson distribution, but now the possible range of values for α is limited. The proposal followed here is to take as the set of parameters (a, b) all values within the triangle given by end-points $(0, 0)$, $(s_a, 0)$, $(0, s_b)$, with hyperparameters $s_a > 0$ and $s_b > 0$. Note that this implies that, a priori, all 'rates of occurrence of failures' $a/b \in [0, \infty)$ are represented by pairs (a, b) within this set. This can again be considered to represent lack of prior information. Next the posterior lower and upper expected profits in the warranty model, using this set of parameters, are derived.

However, before getting the bounds for the expected profit, we consider some properties of $\mathbb{E}G$ as a function of the second parameter b.

Proposition 5. *If $y \geq 0$, then the expected profit $\mathbb{E}G$ of equation (4), as a function of the parameter b, has a single minimum in the interval $[0, \infty)$ or it is decreasing over this interval.*

Let us represent the function $\mathbb{E}G$ in equation (4) as $\mathbb{E}G = n(x - c) - yQ(a, b)$. Then the minimum of $\mathbb{E}G$ corresponds to the maximum of $Q(a, b)$. It follows from the proof of Proposition 5 that the condition for the function

$Q(a,b)$ to have its maximum at point $b \geq 0$ for fixed a is

$$\sum_{k=z+1}^{n} (k-z)P(k)\left((a+K)\tau - (b+T)k\right).$$

Here $P(k)$ is defined by equation (3). If the optimal value of b is negative, then the function $Q(a,b)$ is decreasing.

If we take s_a such that it is less than some $a_{\text{opt}}(b)$ which provides the maximum of $Q(a,b)$ for a fixed b, then $Q(a,b)$ increases for $a \leq s_a$, for every b. This implies that the smallest value of $Q(a,b)$ is achieved at $a = 0$. Then the minimum of $Q(0,b)$ is achieved at $b = s_b$ if s_b is larger than the optimal value of b providing the minimum of $Q(0,b)$. Together with Proposition 5, this enables us to formulate the following proposition which determines the upper bound for the expected profit.

Proposition 6. *Suppose that $Q(0,b)$ achieves its maximum at $b = b_{opt}$. Let us take $s_b \geq b_{opt}$. Then the upper expected profit $\overline{\mathbb{E}}G$ is achieved at $(a,b) = (0, s_b)$ and is equal to*

$$\overline{\mathbb{E}}G = n(x-c) - y\sum_{k=z+1}^{n} \frac{(k-z)\Gamma(K+k)}{\Gamma(K)k!}\left(\frac{s_b+T}{s_b+T+\tau}\right)^{K}\left(\frac{\tau}{s_b+T+\tau}\right)^{k}.$$

Here b_{opt} is defined as the solution of the equation:

$$\sum_{k=z+1}^{n} \frac{(k-z)\Gamma(K+k)}{\Gamma(K)k!}\left(\frac{b+T}{b+T+\tau}\right)^{K}\left(\frac{\tau}{b+T+\tau}\right)^{k}(K\tau - (b+T)k) = 0. \tag{6}$$

If $b_{opt} < 0$, then the function $\overline{\mathbb{E}}G$ is increasing as a function of b and we take $b_{opt} = 0$.

In the same way, we formulate another proposition which determines the lower bound for the expected profit.

Proposition 7. *Let us consider a set of values of $b \in [0, s_b]$. The lower expected profit $\underline{\mathbb{E}}G$ is determined by means of the optimization problem:*

$$\underline{\mathbb{E}}G = n(x-c) - y \cdot \max_{b\in[0,s_b]} Q(a_0(b), b),$$

where $a_0(b)$ is the root of equation (5) in Proposition 3 under the condition that $s\alpha$ is replaced by a and v is computed as

$$v = \frac{b+T}{b+T+\tau}.$$

If the obtained value of $a_0(b)$ satisfies condition $a_0(b) > s_a$, then $a_0(b)$ takes the value s_a.

It follows from Proposition 7 that $\underline{\mathbb{E}}G$ can only be computed numerically by considering all possible values $b \in [0, s_b]$ in a predefined grid.

Corollary 1. *Before taking any observations into account, hence solely based on the set of prior distributions as described for model II in this section, and assuming $s_a > 0$ and $s_b > 0$, the optimal upper bound for the expected profit for seller A is $\overline{\mathbb{E}}G = n(x - c)$.*

It is interesting to point out that the lower bound for the expected profit does not have a similar trivial form due to restricted values of n.

The Negative Binomial model II for the basic warranty scenario considered in this paper, is illustrated in the following example.

Example 3. *Consider again the scenario of Example 1, where now model II is applied in order to determine the interval of values of y, following the same arguments as in Example 2, and assuming that $s_a = 1$ and $s_b = 1$.*

First, we solve equation (6), i.e.,

$$\sum_{k=2}^{100} \frac{(k-1)\Gamma(2+k)}{\Gamma(2)k!} \left(\frac{b+3}{b+3+1}\right)^2 \left(\frac{1}{b+3+1}\right)^k (2 \cdot 1 - (b+3)k) = 0.$$

This equation has a negative solution. Therefore, the function $Q(0, b)$ is decreasing as function of b and the value of the hyperparameter s_b can be taken arbitrarily in the interval $(0, \infty)$.

$$\begin{aligned}\overline{\mathbb{E}}G &= 100 \cdot 4 - y \sum_{k=2}^{100} \frac{(k-1)\Gamma(2+k)}{\Gamma(2)k!} \left(\frac{1+3}{1+3+1}\right)^2 \left(\frac{1}{1+3+1}\right)^k \\ &= 100 \cdot 4 - y \cdot 0.14.\end{aligned}$$

Let us solve the problem $\max_{b\in[0,1]} Q(a_0(b), b)$*. It turns out that* $Q(a_0(b), b)$ *achieves its maximum at* $b = 1$*, where* $a_0(1) = 332.8$*. Hence, the optimal lower bound for the expected profit is attained for* $a = s_a = 1$*, leading to*

$$\underline{\mathbb{E}}G = 100 \cdot 4 - y \sum_{k=2}^{100} \frac{(k-1)\Gamma(1+2+k)}{\Gamma(1+2)k!} \left(\frac{1+3}{1+3+1}\right)^{1+2} \left(\frac{1}{1+3+1}\right)^{k}$$
$$= 100 \cdot 4 - y \cdot 0.262.$$

Hence, the upper bound for y*, determined by setting* $\overline{\mathbb{E}}G = 0$*, is* $\overline{y} = 2,857$*. The lower bound for* y*, determined by setting* $\underline{\mathbb{E}}G = 0$*, is* $\underline{y} = 1,527$*, which is identical to the value for the "precise" model as derived in Example 1. The bounds differ substantially from the value under model I as derived in Example 2, which is a direct consequence of the restriction of the prior range of values of* α*.*

To consider some further aspects of interest in the proposed model II, Figure 2 shows the lower and upper bounds for y *for various values of the hyperparameters* $s_a = s_b = s$*. This clearly illustrates the increased imprecision for larger values of* s*, the specific value to use must be based on judgement of the topic experts.*

It is further of interest to illustrate the dependence of the lower and upper bounds for $\ln(y)$ *on* z*. For the model with* $s = 1$*, Figure 3 shows these bounds for some values of* z*, where it is assumed that not too many of the* $n = 100$ *items are likely to fail during the time period considered, as failures are assumed to be quite rare. Both these lower and upper bounds for* y *are, of course, increasing as functions of* z*, with particularly the upper bound increasing rapidly due to the small number of failing items in the observed data.*

Finally, it is interesting to consider how the total time T *of the data observations influences the lower and upper bounds for* $\ln(y)$*. For fair comparison, the corresponding values of the total number* K *of failures over this time period are defined such that the empirical failure rate, given by* K/T*, is constant and is kept at the value* $2/3$ *(see Example 2). The corresponding values are shown in Figure 4. Clearly, the imprecision, that is the width of the interval* $[\underline{y}, \overline{y}]$*, decreases as* T *increases, which is in line with intuition as the amount of imprecision logically decreases as function of the number of available data observations, and* y *tends to a limit value which can be determined from Equation (2) with* $p(i|\lambda = 2/3)$*.*

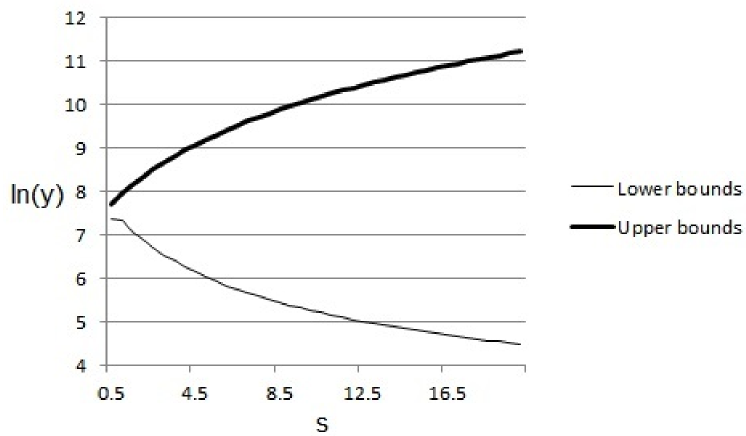

Figure 2: The lower and upper bounds for $\ln(y)$ for various values of the hyperparameter s.

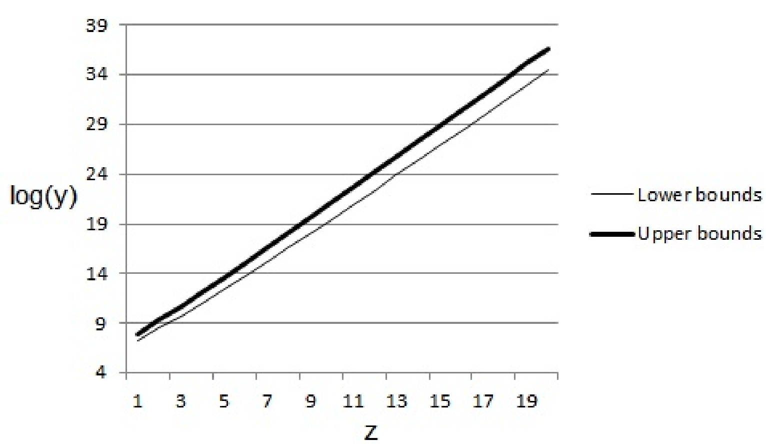

Figure 3: The lower and upper bounds for $\ln(y)$ for various values of z.

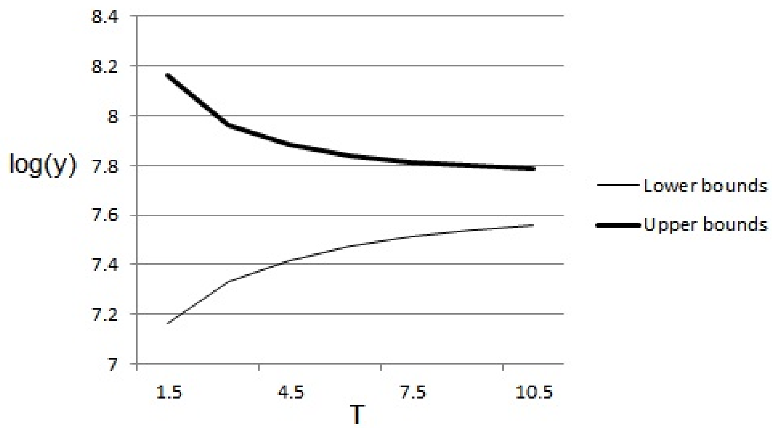

Figure 4: The lower and upper bounds for $\ln(y)$ for various values of T.

6. CONCLUDING REMARKS

This paper has introduced two imprecise probability models for a basic warranty scenario, which can be used to guide the compensation offered by a seller of units in case too many items fail during the warranty period. The proposed models are closely related, with model I being arguably the intuitively more logical one, but it has a disadvantage that is overcome by model II. The explicit derivations of formulae for the lower and upper expected profits for these two models are powerful results that make application and analysis of the models straightforward. Of course, there are many aspects related to practical warranty decisions that require more detailed study in order to develop imprecise probability models and inferential approaches for them, these provide interesting challenges for future research.

Acknowledgement

We would like to express our appreciation to the anonymous referees whose very valuable comments have resulted in an improved paper.

References

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

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

[3] Utkin L.V., Coolen F.P.A. (2007). Imprecise reliability: an introductory overview. In: *Computational Intelligence in Reliability Engineering, Volume 2: New Metaheuristics, Neural and Fuzzy Techniques in Reliability*, Levitin G. (Ed.), Springer, Berlin, pp. 261–306.

[4] Coolen F.P.A., Utkin L.V. (2011). Imprecise reliability. In: *International Encyclopedia of Statistical Science*, Lovric M. (Ed.), Springer, Berlin, pp. 649–650.

[5] Coolen F.P.A., Utkin L.V. (2014). Reliability and risk. In: *Introduction to Imprecise Probabilities*, Augustin et al (Eds.), Wiley, Chichester, pp. 305–317.

[6] Singpurwalla N.D. (2000). Warranty contracts and equilibrium probabilities. In: *Statistical Science in the Courtroom*, Gastwirth J.L. (Ed.), Springer, New York, pp. 363–377.

[7] Aven T. (2003). *Foundations of Risk Analysis: A Knowledge and Decision-Oriented Perspective*. Wiley, Chichester.

[8] Wu S. (2013). A review on coarse warranty data and analysis. *Reliability Engineering and System Safety* **114** 1–11.

[9] Gupta S.K., De S., Chatterjee A. (2014). Warranty forecasting from incomplete two-dimensional warranty data. *Reliability Engineering and System Safety* **126** 1–13.

[10] Stephens D., Crowder M. (2004). Bayesian analysis of discrete time warranty data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **53** 195–217.

[11] Chen T.M., Popova E. (2000). Bayesian maintenance policies during a warranty period. *Stochastic Models* **16** 121–142.

[12] Wu S.J., Huang S.R. (2010). Optimal warranty length for a Rayleigh distributed product with progressive censoring. *IEEE Transactions on Reliability* **59** 661–666.

[13] Akbarov A., Wu S. (2012). Forecasting warranty claims considering dynamic over-dispersion. *International Journal of Production Economics* **139** 615–622.

[14] Bernardo J.M., Smith A.F.M. (1994). *Bayesian Theory*. Wiley, Chichester.

[15] Novick M.R., Hall W.J. (1965). A Bayesian indifference procedure. *Journal of the American Statistical Association* **60** 1104–1117.

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

[17] Syversveen A.R. (1998). Noninformative Bayesian priors. Interpretation and problems with construction and applications. Preprint Statistics 3, Department of Mathematical Sciences, NTNU, Trondheim.

[18] Coolen F.P.A. (1997). An imprecise Dirichlet model for Bayesian analysis of failure data including right-censored observations. *Reliability Engineering and System Safety* **56** 61–68.

[19] Troffaes M.C.M., Coolen F.P.A. (2009). Applying the Imprecise Dirichlet Model in cases with partial observations and dependencies in failure data. *International Journal of Approximate Reasoning* **50** 257–268.

[20] Kelly D., Atwood C. (2011). Finding a minimally informative Dirichlet prior distribution using least squares. *Reliability Engineering and System Safety* **96** 398–402.

[21] Quaeghebeur E., de Cooman G. (2005). Imprecise probability models for inference in exponential families. In: *Proceedings of the 4rd International Symposium on Imprecise Probabilities and Their Applications, ISIPTA'05*, Bernard J.M. et al (Eds.), Carnegie Mellon University, Pittsburgh, pp. 287–296.

Appendix

A. Proof of Proposition 1

Let us introduce notation $w = s\alpha + K$, $v = (s+T)/(s+T+\tau)$ and consider the function

$$Q(w) = \sum_{k=z+1}^{n} (k-z) \frac{\Gamma(w+k)}{\Gamma(w)\cdot k!} (1-v)^k v^w.$$

We aim to prove that the function $Q(w)$ has a single maximum in interval $[0,\infty)$ for w (the expected profit has a single minimum in interval $[0,\infty)$ for α). Without loss of generality, we take $z = 0$ and integer values of w for simplicity. Let us transform the function $Q(w)$ as follows:

$$Q\left(w\right) = \sum_{k=1}^{n} \frac{1}{(k-1)!} \frac{\Gamma\left(w+k\right)}{\Gamma\left(w\right)} (1-v)^k v^w. \tag{7}$$

Hence, we can represent the function $Q(w)$ through the following derivatives:

$$\begin{aligned} Q\left(w\right) &= \frac{v^w\left(1-v\right)}{\Gamma\left(w\right)} \sum_{k=1}^{n} \left(w+k-1\right)...k \cdot \left(1-v\right)^{k-1} \\ &= \left(-1\right)^w \frac{v^w\left(1-v\right)}{\Gamma\left(w\right)} \sum_{k=1}^{n} \left(\left(1-v\right)^{w+k-1}\right)_v^{(w)} \end{aligned}$$

where $(\cdot)_v^{(w)}$ denotes the w-th derivative over v. The above implies that

$$Q\left(w\right) = \left(-1\right)^w \frac{v^w\left(1-v\right)}{\Gamma\left(w\right)} \left(\frac{\left(1-v\right)^w - \left(1-v\right)^{w+n}}{v}\right)_v^{(w)}.$$

It is easy to show that

$$\left(\frac{\left(1-v\right)^w}{v}\right)_v^{(w)} = \left(\frac{1}{v}\right)^{(w)} = \left(-1\right)^w \frac{w!}{v^{w+1}}.$$

Then we get

$$Q\left(w\right) = \left(-1\right)^w \frac{v^w\left(1-v\right)}{\Gamma\left(w\right)} \left(\left(-1\right)^w \frac{w!}{v^{w+1}} - \left(\frac{\left(1-v\right)^{w+n}}{v}\right)_v^{(w)}\right)$$

or

$$Q\left(w\right)=\frac{1-v}{v}w-\left(-1\right)^{w}\frac{v^{w}\left(1-v\right)}{\Gamma\left(w\right)}\left(\frac{\left(1-v\right)^{w+n}}{v}\right)_{v}^{(w)}.$$

Let us use the expression for differentiation of the product of two functions

$$(f\cdot g)^{(n)}=\sum_{k=0}^{n}\binom{n}{k}f^{(n-k)}g^{(k)}.$$

Here $\binom{n}{k}=n!/(k!(n-k)!)$. Then we get

$$\begin{aligned}&\left(\left(1-v\right)^{w+n}\cdot\frac{1}{v}\right)_{v}^{(w)}\\&=\sum_{k=0}^{w}\binom{w}{k}\left(\left(1-v\right)^{w+n}\right)^{(w-k)}\left(\frac{1}{v}\right)^{(k)}\\&=\left(-1\right)^{w}\sum_{k=0}^{w}\binom{w}{k}\left(w+n\right)...\left(n+k+1\right)\cdot\left(1-v\right)^{n+k}\frac{k!}{v^{k+1}}.\end{aligned}$$

As a result, we obtain

$$\begin{aligned}Q\left(w\right)&=\frac{1-v}{v}w\\&-w\left(\sum_{k=0}^{w}\frac{1}{\left(w-k\right)!}\left(w+n\right)...\left(n+k+1\right)\cdot\left(1-v\right)^{n+k+1}v^{w-k-1}\right)\end{aligned}$$

or

$$\begin{aligned}Q\left(w\right)&=\frac{1-v}{v}w\\&-w\left(\sum_{k=0}^{w}\frac{1}{k!}\left(w+n\right)...\left(w+n-k+1\right)\cdot\left(1-v\right)^{w+n-k+1}v^{k-1}\right).\end{aligned}$$

Hence, there holds

$$\begin{aligned}Q\left(w\right)&=\frac{1-v}{v}w\\&-\frac{1-v}{v}w\left(\sum_{k=0}^{w}\binom{w+n}{k}\left(1-v\right)^{w+n-k}v^{k}\right),\end{aligned}$$

or

$$Q\left(w\right)=\frac{1-v}{v}w\sum_{k=w+1}^{w+n}\binom{w+n}{k}\left(1-v\right)^{w+n-k}v^{k}.$$

After replacing the summation index, we finally get

$$Q\left(w\right)=\left(1-v\right)v^{w-1}w\sum_{k=1}^{n}\binom{w+n}{n-k}\left(1-v\right)^{n-k}v^{k}$$

or

$$Q\left(w\right)=\left(1-v\right)v^{w-1}w\sum_{k=0}^{n-1}\binom{w+n}{k}\left(1-v\right)^{k}v^{n-k}.$$

For arbitrary values of w, we have

$$\begin{aligned}Q\left(w\right)&=\left(1-v\right)v^{w-1}w\\&\times\left(\sum_{k=1}^{n-1}\frac{\left(1-v\right)^{k}v^{n-k}}{k!}\left(w+n\right)...\left(w+n-k+1\right)+v^{n}\right).\end{aligned}$$

Another form of the same function is

$$\begin{aligned}Q\left(w\right)&=\frac{1-v}{v}v^{w+n}w\\&\times\left(\sum_{k=1}^{n-1}\frac{1}{k!}\left(\frac{1-v}{v}\right)^{k}\left(w+n\right)...\left(w+n-k+1\right)+1\right).\end{aligned}\tag{8}$$

Both polynomials before v^w in (7) and (8) have the n-th power. The derivative of function (8) is

$$\begin{aligned}Q'\left(w\right)&=\frac{1-v}{v}v^{w+n}\sum_{k=1}^{n-1}\frac{1}{k!}\left(\frac{1-v}{v}\right)^{k}\left(w+n\right)...\left(w+n-k+1\right)\\&\times\left(1+\sum_{i=0}^{k-1}\frac{w}{w+n-i}+w\ln v\right)+\frac{1-v}{v}v^{w+n}\left(1+w\ln v\right).\end{aligned}$$

For clarity, we write the above sum starting from index zero

$$\begin{aligned} Q'(w) &= \frac{1-v}{v} v^{w+n} \sum_{k=0}^{n-1} \frac{1}{k!} \left(\frac{1-v}{v} \right)^k \\ &\times (w+n) \ldots (w+n-k+1) \\ &\times \left(1 + \sum_{i=0}^{k-1} \frac{w}{w+n-i} + w \ln v \right). \end{aligned}$$

One can see that the root of the equation $Q'(w) = 0$ is between $w_1(k_1)$ and $w_1(k_1+1)$ for some k, where $w_1(k_1)$ is the root of an equation produced from the expression in parentheses, which is of the form:

$$\frac{1}{w} + \sum_{i=0}^{k_1-1} \frac{1}{w+n-i} = -\ln v.$$

Then we have to find k, k_1 and w satisfying the system of equations

$$\begin{cases} \sum_{i=0}^{k-1} \frac{1}{w+i} = -\ln v, \\ \frac{1}{w} + \sum_{i=0}^{k_1-1} \frac{1}{w+n-i} = -\ln v. \end{cases}$$

If a solution of the above system is unique, then we have proven the proposition. Suppose that the solution is not unique. Let $w_0(k)$ be a root of the equation

$$\sum_{i=0}^{k-1} \frac{1}{w+i} = -\ln v,$$

and $w_1(k)$ be a root of the equation

$$\frac{1}{w} + \sum_{i=0}^{k-1} \frac{1}{w+n-i} = -\ln v, \;\; k = 1, \ldots, n.$$

It is obvious that the sequences $\{w_0(k)\}$ and $\{w_1(k)\}$ are increasing. It is also obvious that there exist k and k_1 such that the root w^* of the equation $Q'(w) = 0$ satisfies the following inequalities:

$$\sum_{i=0}^{k-1} \frac{1}{w^*+i} \geq -\ln v \geq \sum_{i=0}^{k} \frac{1}{w^*+i},$$

$$\frac{1}{w^*}+\sum_{i=0}^{k_1-1}\frac{1}{w^*+n-i}\geq-\ln v\geq\frac{1}{w^*}+\sum_{i=0}^{k_1}\frac{1}{w^*+n-i}.$$

So, we get a new expression for $Q(w)$ following from (8)

$$Q\left(w\right)=\frac{1-v}{v}w\sum_{k=0}^{n-1}\frac{1}{k!}\left(w+n\right)...\left(w+n-k+1\right)\left(1-v\right)^{k}v^{w+n-k},$$

which can be rewritten for integer w as

$$Q\left(w\right)=\frac{1-v}{v}w\sum_{k=0}^{n-1}\binom{w+n}{k}\left(1-v\right)^{k}v^{w+n-k}.$$

We introduce the function

$$P\left(w\right)=\frac{Q\left(w\right)}{\frac{1-v}{v}w}=\sum_{k=0}^{n-1}\binom{w+n}{k}\left(1-v\right)^{k}v^{w+n-k}.$$

First, we will prove that this function is monotone and decreasing. Second, we will prove that the function $-P'\left(w\right)/P\left(w\right)$ is increasing. Then the equation $Q'\left(w\right)=0$ has a unique root which defines the maximum of the function $Q\left(w\right)$. Indeed, it follows from the equality $Q'\left(w\right)=0$ that

$$\frac{1-v}{v}P\left(w\right)+\frac{1-v}{v}wP'\left(w\right)=0.$$

Hence, there holds

$$-\frac{P'\left(w\right)}{P\left(w\right)}=\frac{1}{w}. \tag{9}$$

Consequently, if the function $-P'\left(w\right)/P\left(w\right)$ is increasing, then equality (9) has a unique root.

Now we have to prove that the function $P\left(w\right)$ is decreasing, this is done in two steps.

First, we show that the inequality $P\left(w+1\right)\leq P\left(w\right)$ holds. It follows from the equality

$$\binom{w+n+1}{k}=\binom{w+n}{k}+\binom{w+n}{k-1}$$

that the function $P(w+1)$ is represented as

$$P(w+1) = \sum_{k=1}^{n-1} \binom{w+n+1}{k} (1-v)^k v^{w+n-k+1} + v^{w+n+1}$$
$$= \sum_{k=1}^{n-1} \left(\binom{w+n}{k} + \binom{w+n}{k-1} \right) (1-v)^k v^{w+n-k+1} + v^{w+n+1}.$$

Hence

$$P(w+1) = \sum_{k=1}^{n-1} \binom{w+n}{k} (1-v)^k v^{w+n-k+1}$$
$$+ \sum_{k=0}^{n-2} \binom{w+n}{k} (1-v)^{k+1} v^{w+n-k} + v^{w+n+1}.$$

The above implies

$$P(w+1) = \sum_{k=1}^{n-1} \binom{w+n}{k} (1-v)^k v^{w+n-k} (v+1-v)$$
$$- \binom{w+n}{n-1} (1-v)^n v^{w+1} + (1-v) v^{w+n} + v^{w+n+1}.$$

Consequently, we can write

$$P(w+1) = P(w) - \binom{w+n}{n-1} (1-v)^n v^{w+1} \leq P(w). \tag{10}$$

As the second step, we prove that the function $-P'(w)/P(w)$ is increasing, i.e., we prove that

$$-\frac{P'(w+1)}{P(w+1)} \geq -\frac{P'(w)}{P(w)}.$$

Equality (10) can be rewritten as follows:

$$P(w+1) = P(w) - \frac{1}{(n-1)!} (w+n) \ldots (w+2) (1-v)^n v^{w+1}.$$

By differentiating we get

$$P'(w+1) = P'(w) - \binom{w+n}{n-1}(1-v)^n v^{w+1} \times \left(\sum_{i=0}^{n-2} \frac{1}{w+n-i} + \ln(v)\right) \tag{11}$$

Now we consider the difference

$$P'(w+1)\,P(w) - P'(w)\,P(w+1)$$

and prove that it is not larger than zero. By using (10) and (11), we obtain

$$\begin{aligned} & P'(w+1)\,P(w) - P'(w)\,P(w+1) \\ = & \left(P'(w) - \binom{w+n}{n-1}(1-v)^n v^{w+1}\left(\sum_{i=0}^{n-2}\frac{1}{w+n-i} + \ln(v)\right)\right)P(w) \\ & -P'(w)\left(P(w) - \binom{w+n}{n-1}(1-v)^n v^{w+1}\right) \\ = & -\binom{w+n}{n-1}(1-v)^n v^{w+1}\left(\sum_{i=0}^{n-2}\frac{1}{w+n-i} + \ln(v)\right)P(w) \\ & +\binom{w+n}{n-1}(1-v)^n v^{w+1}P'(w). \end{aligned}$$

Now we have to prove that

$$P'(w) - \left(\sum_{i=0}^{n-2}\frac{1}{w+n-i} + \ln(v)\right)P(w) \le 0.$$

This is obvious due to

$$\begin{aligned} P'(w) &= \sum_{k=0}^{n-1} C_{w+n}^{k}(1-v)^k v^{w+n-k}\left(\sum_{i=0}^{k-1}\frac{1}{w+n-i} + \ln(v)\right) \\ &\le P(w)\left(\sum_{i=0}^{n-2}\frac{1}{w+n-i} + \ln(v)\right), \end{aligned}$$

which completes the proof.

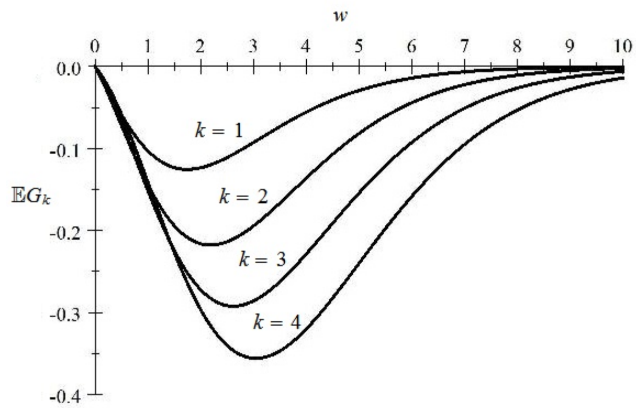

Figure 5: Typical curves of the function $\mathbb{E}G_k$ for $k = 1, ..., 4$

B. Proof of Proposition 3

We consider the k-th term $\mathbb{E}G_k$ of the sum in the expression for $\mathbb{E}G$, without parameters which do not depend on α and k,

$$\mathbb{E}G_k = -(k-z)\frac{\Gamma(w+k)}{\Gamma(w)\cdot k!}v^w (1-v)^k .$$

Here $w = s\alpha + K$ and $v = (s+T)/(s+T+\tau)$. Denote

$$Z = \frac{(k-z)(1-v)^k}{k!} \geq 0,$$

and rewrite the k-th term as follows:

$$\mathbb{E}G_k = -Z\frac{\Gamma(w+k)}{\Gamma(w)}v^w.$$

Differentiating the above expression gives

$$\frac{\mathrm{d}\mathbb{E}G_k}{\mathrm{d}w} = Z \cdot \frac{\Gamma(w+k)}{\Gamma(w)}v^w (\psi(w) - \psi(w+k) - \ln v) .$$

Here $\psi(x) = \frac{\mathrm{d}}{\mathrm{d}x}\ln\Gamma(x)$ is the digamma function. Hence, we can write the following condition for the extremum:

$$\ln v + \psi(w+k) - \psi(w) = 0.$$

Note that $0 < v \leq 1$, which implies that $\ln v \leq 0$. On the other hand, $\psi(w+k) - \psi(w)$ is a decreasing function of $w \geq 0$ because

$$\psi(w+k) - \psi(w) = \sum_{i=0}^{k-1} \frac{1}{w+i}.$$

Moreover, $\psi(w+k) - \psi(w) \geq 0$. Therefore, there exists a single point w_0 for which $\psi(w_0+k) - \psi(w_0) = -\ln v$. If $w \leq w_0$ then

$$\psi(w+k) - \psi(w) \geq \psi(w_0+k) - \psi(w_0)$$

due to fact that function $\psi(w+k) - \psi(w)$ is decreasing. This implies that $\ln v + \psi(w+k) - \psi(w) \geq 0$ and

$$\frac{\mathrm{d}\mathbb{E}G_k}{\mathrm{d}w} \leq 0,$$

because $Z \geq 0$ and $v^w \Gamma(w+k) / \Gamma(w) \geq 0$. If $w > w_0$ then

$$\psi(w+k) - \psi(w) \leq \psi(w_0+k) - \psi(w_0)$$

and

$$\frac{\mathrm{d}\mathbb{E}G_k}{\mathrm{d}w} \geq 0.$$

The above implies that w_0 is a global minimum of $\mathbb{E}G_k$.

According to Proposition 1, there is a unique minimum point α_0. Typical curves of the function $\mathbb{E}G_k$ for different k are depicted in Fig. 5. One can see from the figure that every term $\mathbb{E}G_k$ has a single minimum whose existence has been proved above. Moreover, the value of α_k corresponding to the minimum of $\mathbb{E}G_k$ is less than the value of α_{k+1} corresponding to the minimum of $\mathbb{E}G_{k+1}$. This follows from the inequality

$$\sum_{i=0}^{k-1} \frac{1}{w+i} \leq \sum_{i=0}^{k} \frac{1}{w+i}$$

and condition $\ln v + \psi(w+k) - \psi(w) = 0$. This implies that the minimum of the sum of all $\mathbb{E}G_k$ should be located between α_1 and α_{n-z}.

C. Proof of Proposition 4

Without loss of generality, we again take $z = 0$ for simplicity. Then we can write

$$\begin{aligned}\mathbb{E}G &= n(x-c) - y\sum_{k=1}^{\infty} kP(k) \\ &= n(x-c) - y \cdot \mathbb{E}X.\end{aligned}$$

Here the expectation $\mathbb{E}X$ for the Negative Binomial distribution is defined as $\mathbb{E}X = \tau(a+K)/(b+T)$. Hence, there holds

$$\mathbb{E}G = n(x-c) - y\tau(s\alpha + K)/(s+T),$$

as was to be proved.

D. Proof of Proposition 5

Let us represent the function $\mathbb{E}G$ as $\mathbb{E}G = n(x-c) - yQ(b)$. Then we have to prove that the function $Q(b)$ has a single maximum. Without loss of generality we again take $z = 0$ for simplicity. By differentiating $Q(b)$ we obtain the following condition for the maximum

$$\sum_{k=1}^{n} kP(k,b)\left((a+K)\tau - (b+T)k\right) = 0.$$

Here $P(k,b)$ is used to denote the probability given in equation (3).

Let us consider the case $n = 1$, which gives

$$(a+K)\tau = b+T.$$

It is obvious that we have a single non-negative root of the above equation if $(a+K)\tau - T \geq 0$. Suppose that the proposition is valid for some n, i.e., there is a value of b denoted b_n such that $Q(b)$ achieves the maximum at point b_n. By induction, we write the following condition of the maximum for the case $n+1$:

$$b_{n+1} + T = \frac{(a+K)\tau\sum_{k=1}^{n} kP(k,b) + (a+K)\tau(n+1)P(n+1,b)}{\sum_{k=1}^{n} k^2P(k,b) + (n+1)^2P(n+1,b)}.$$

We will prove that there is a single value of b_{n+1} satisfying the above condition. We write the above equality in the following short form:

$$\beta\left(\alpha + P(n+1,b)\right) = \gamma + \lambda P(n+1,b),$$

where $\beta = b_{n+1} + T$, $\alpha = (a+K)\tau/(b_n + T)/(n+1)^2$, $\gamma = (a+K)\tau/(n+1)^2$, $\lambda = (a+K)\tau/(n+1)$. First, note that $P(n+1,b)$ has a maximum (see the similar case $n = 1$). Now we prove that $\beta\,(\alpha + P(n+1,b))$ has a maximum if $P(n+1,b)$ has a maximum. Indeed, in this case $P'(n+1,b) = 0$ holds. Then for the function $\beta P(n+1,b)$ we can write

$$\begin{aligned}(\beta P(n+1,b))' &= \beta' P(n+1,b) + \beta P'(n+1,b)\\ &= P(n+1,b) + \beta P'(n+1,b) = 0.\end{aligned}$$

The above condition is valid if $P'(n+1,b) \leq 0$ at point b_{n+1}. Then $\beta P'(n+1,b)$ is increasing function by $b \geq b_{n+1}$. At the same time $P(n+1,b)$ is decreasing as function of $b \geq b_{n+1}$. This implies that they intersect in a single point, as was to be proved.

The case when the function $Q(b)$ is decreasing is obvious if $b_{n+1} \leq 0$.