Statistics, Vol. 40, No. 4, August 2006, 353–363

Taylor & Francis
Taylor & Francis Group

Bayesian prediction with an asymmetric criterion in a nonparametric model of insurance risk

WOJCIECH NIEMIRO*

Faculty of Mathematics and Computer Science, Nicolaus Copernicus University, Toruń, Poland

(Received 20 March 2003; in final form 19 January 2006)

We consider a nonparametric Bayesian insurance risk model. The claims are seen as a marked point process (T_i, Y_i), where T_i is the time of occurrence of the ith claim and Y_i is its size. We assume that this is a nonhomogeneous Poisson process on $\mathbb{R}_+^2$ with intensity measure $P \times \Theta$. Here P describes the exposure to risk and it is known, whereas Θ is regarded as an unknown risk characteristic. According to the Bayesian paradigm, we assume that the measure Θ is random. Processes with independent increments are used as prior distributions. In particular, Gamma processes are conjugate priors. The problem is to predict the sum of future claims in a given period, given the past of the process. We consider the asymmetric criterion LINEX (linear-exponential) that penalizes underestimation of claims more severely than overestimation. For the conjugate Gamma prior, we construct the best predictor. Under a relaxed assumption on the prior distribution, we construct the best linear predictor.

Keywords: Bayes; Nonparametric; Poisson process; Cox Process; Credibility theory; LINEX

1. Introduction

We consider insurance claims seen as a random collection of pairs (T_i, Y_i), where T_i is the time of occurrence of ith claim and Y_i is its size. Thus we model the risk process as a marked random point process on $\mathbb{R}_+$ or, equivalently, a point process on $\mathbb{R}_+^2$.

Our objective, as in the classical theory of credibility, is to predict future claims in order to calculate an appropriate premium. General references on credibility are Goovaerts *et al.* [1] and Klugman [2], see also a survey in *Encyclopedia of Actuarial Science* [3]. This theory, generally speaking, deals with nonhomogeneous portfolios of risks and adopts the Bayesian or empirical Bayes perspective. It shows how to combine the knowledge about the average risk characteristics in the whole portfolio, with less certain knowledge about an individual contract. The former is usually modeled as a prior distribution, and the latter is given in the form of a random sample.

In this article, we carry over the ideas of credibility to our nonparametric model. We assume that the point process of claims is a Cox or doubly stochastic Poisson process. The unknown risk characteristic is described by a finite measure Θ on $\mathbb{R}_+$. The total mass $\Theta(\mathbb{R}_+)$ is the

*Email: wniem@mat.uni.torun.pl

Statistics
ISSN 0233-1888 print/ISSN 1029-4910 online
http://www.tandf.co.uk/journals
DOI: 10.1080/02331880600618996

expected number of claims per unit of exposure, whereas the normalized measure Θ is the probability distribution of a single claim size, $\mathbb{P}_\Theta(Y_i \le y) = \Theta(0, y]/\Theta(\mathbb{R}_+)$. Conditional on Θ, the point process (T_i, Y_i) is a nonhomogeneous Poisson process with intensity $P \times \Theta$, where the measure P describes the exposure to risk, possibly varying in time but assumed to be known. According to the Bayesian paradigm, Θ is considered as random. In our setting, a prior probability distribution is defined on the set of finite measures on $\mathbb{R}_+$. Natural conjugate priors are given by nonhomogeneous Gamma processes. Our model equipped with the Gamma prior is closely related to nonparametric Bayesian models based on the Dirichlet process [4; 5, chapters 8 and 9]. We also consider more general priors given by some class of processes with independent increments.

For Cox models, Grandell [6] considers the problems of optimal prediction and optimal linear prediction. In the spirit of classical credibility, he derives predictors that minimize the mean square error. Our nonparametric Bayesian model is similar but explicitly includes also the 'size component' of the process. In contrast with the results of Grandell, we choose an asymmetric criterion function known as 'LINEX' (linear-exponential). LINEX seems to be an appealing and well-motivated alternative to the classical quadratic criterion. It is related to the exponential premium principle and penalizes underestimation of claims more severely than overestimation.

For the conjugate Gamma prior, the simplicity of posterior distributions allows us to derive the best LINEX predictor. For more general prior distributions, we only derive the best linear predictor. For comparison, we also provide formulas for the best predictors with respect to the quadratic criterion.

2. Mixed marked process of claims

Consider the following insurance risk model. Claims occur at random points in time and their sizes are also random variables. Denote by T_i and Y_i the time of occurrence and the size, respectively, of the ith claim. The pairs (T_i, Y_i) are seen as a point process on $\mathbb{R}^2_+ = (0, \infty) \times (0, \infty)$. The associated random measure N is generated by

$$N((t, u], (x, y]) = \sum_{i=0}^{\infty} \mathbb{I}(t < T_i \le u, x < Y_i \le y), \tag{1}$$

$0 < t < u < \infty, 0 < x < y < \infty$. Thus, $N((t, u], (x, y])$ is the number of claims which occur during $(t, u]$ and have size in $(x, y]$. The total number of claims in $(t, u]$ is $N((t, u], \mathbb{R}_+) = \sum_{i=0}^{\infty} \mathbb{I}(t < T_i \le u)$. Similarly, the total amount of claims in the same period is

$$S(t, u] = \sum_{i=0}^{\infty} Y_i \mathbb{I}(t < T_i \le u) = \int_t^u \int_0^\infty y N(\mathrm{d}s, \mathrm{d}y). \tag{2}$$

As in the classical Bayesian theory of credibility, we introduce a random quantity Θ interpreted as a parameter that characterizes the risk under consideration. In our model, Θ is a finite random measure on $\mathbb{R}_+$, which describes the intensity of occurrence of claims per unit of exposure. The probability distribution of a single claim size Y_i is $\Theta(\cdot)/\Theta(\mathbb{R}_+)$. Let us also introduce a nonrandom measure P on $\mathbb{R}_+$, which describes the exposure to risk. Thus, $P(t, u]$ is the total amount of risk exposed during $(t, u]$. Assume that P is known and, conditional on Θ, N is a nonhomogeneous Poisson process on $\mathbb{R}^2_+$ with intensity measure $P \times \Theta$. For a definition, construction and properties of nonhomogeneous Poisson processes, we refer the reader to

Kingman [5]. In particular, we have

$$N((t,u],(x,y])|_\Theta \sim \mathrm{Poisson}(P(t,u]\Theta(x,y]), \tag{3}$$

hence $\mathbb{E}_\Theta N((t,u],(x,y]) = P(t,u]\Theta(x,y]$ and consequently,

$$\mathbb{E}_\Theta S(t,u] = P(t,u]\int_0^\infty y\Theta(\mathrm{d}y). \tag{4}$$

Here and in the sequel, $\mathbb{E}_\Theta(\cdot)$ denotes the conditional expectation $\mathbb{E}(\cdot|\Theta)$.

Let us now make suitable assumptions about the prior distribution of the risk 'parameter' Θ. Note that we can identify the random measure Θ with the increasing stochastic process $\Theta(0,t]$.

ASSUMPTION 1 *Suppose that α is a finite measure on a bounded interval $(0, y^*]$ and $\lambda > 0$. Assume that Θ is a Gamma process with shape measure α and inverse scale parameter λ,* i.e. *Θ has independent increments and*

$$\Theta(x,y] \sim \mathrm{Gamma}(\alpha(x,y],\lambda).$$

The Gamma process introduced above is a conjugate prior for the Poisson process. We will also consider another weaker assumption.

ASSUMPTION 2 *Suppose that α is a finite measure on a bounded interval $(0, y^*]$. Assume that Θ is a process with independent increments such that the moment generating function of its increments is of the form*

$$\mathbb{E}\exp\left[r\Theta(x,y]\right] = \exp\left[\psi(r)\alpha(x,y]\right],$$

where ψ is some function. Let $r^* = \sup\{r\colon \psi(r) < \infty\}$. Assume that $r^* > 0$ and $\psi(r) \to \infty$ as $r \nearrow r^*$.

Clearly, if Assumption 1 is fulfilled, then Assumption 2 holds with

$$\psi(r) = \log\frac{\lambda}{\lambda - r} \tag{5}$$

and $r^* = \lambda$.

3. Best prediction with respect to the LINEX criterion

Our objective is to predict the sum of claims in some interval of time $(t,u]$ in the future, given the past history of the process, i.e. pairs (T_i, Y_i) with $T_i \le t$. Thus, a predictor of $S(t,u]$ is a $\mathcal{F}_t^N$-measurable random variable, where $\mathcal{F}_t^N = \sigma(N((v,s],(x,y])\colon 0 < v < s \le t, 0 < x < y < \infty)$. To formalize the problem of prediction, we have to choose a suitable criterion (loss function) L. The best predictor is a solution to the following minimization problem

$$\mathbb{E}L(S(t,u] - H) = \min, \tag{6}$$

over all H that are $\mathcal{F}_t^N$-measurable. The operator $\mathbb{E}$ stands for the unconditional expectation with respect to the joint probability distribution of Θ and N.

The classical square criterion function $L(y) = y^2$ is symmetric. For insurance applications, symmetry may not be desirable: the penalty for underestimation should not be the same as for overestimation. We will use an asymmetric criterion function that seems to be especially suitable for the kind of problems we consider. It is called ‘LINEX’ and is given by

$$L(y) = \mathrm{e}^{\kappa y} - \kappa y - 1,$$

for some fixed $\kappa > 0$ [7]. LINEX is a function with the minimum 0 at $y = 0$. It behaves roughly as an exponential function for $y \to \infty$ and as a linear function for $y \to -\infty$ (figure 1). It penalizes underestimation much more severely than overestimation.

An appealing feature of LINEX is the following fact. The best predictor defined by equation (6) is given by

$$H = \frac{1}{\kappa} \log \mathbb{E} \left(\mathrm{e}^{\kappa S(t,u]} \mid \mathcal{F}_t^N \right). \tag{7}$$

Derivation of this formula is straightforward. It is enough to condition on $\mathcal{F}_t^N$ and to minimize the posterior expectation with respect to H. Let us mention that equation (7) can be interpreted as the ‘posterior premium’ calculated according to the well-known ‘exponential principle’. Derivation of this principle is usually based on utility theory. Discussion of the properties of the exponential principle and its relations with other principles of premium calculation can be found in refs. [1; 8, chapter 5]. Note that the minimizer of the square criterion is $\mathbb{E}\left(S(t,u] \mid \mathcal{F}_t^N\right)$, i.e. the ‘net posterior premium’.

Under the conjugate prior, it is possible to derive an explicit expression for the best predictor (7).

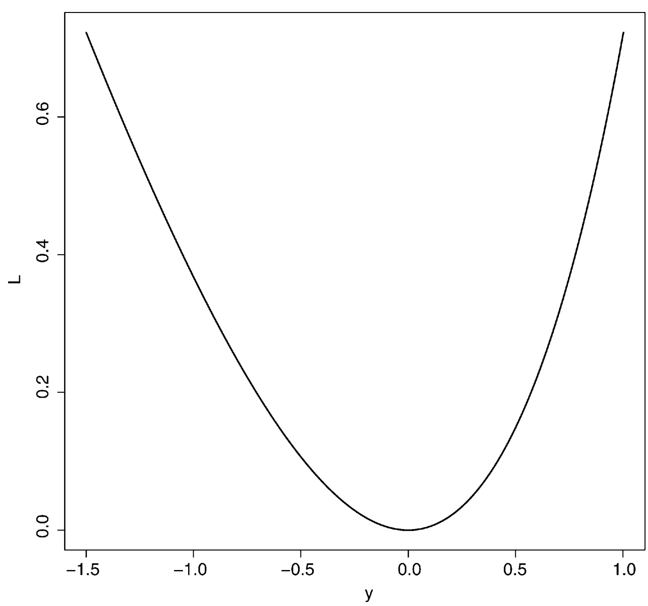

Figure 1. The graph of the LINEX function $L(y)$ for $\kappa = 1$.

THEOREM 1 *If Assumption* 1 *holds, then the best predictor of* $S(t, u]$ *with respect to the LINEX criterion is given by*

$$H = -\frac{1}{\kappa}\int_0^{y^*} \log\left(1 - \frac{P(t,u](e^{\kappa y}-1)}{\lambda + P(0,t]}\right)[\alpha(dy) + N((0,t], dy)],$$

provided that $y^* < (1/\kappa)\log[1 + (\lambda + P(0,t])/P(t,u]]$.

Recall that y^* bounds the support of α from above. Let us rewrite the conclusion of the theorem in a slightly more explicit form

$$H = -\frac{1}{\kappa}\int_0^{y^*} \log\left(1 - \frac{P(t,u](e^{\kappa y}-1)}{\lambda + P(0,t]}\right)\alpha(dy) - \frac{1}{\kappa}\sum_{i:T_i\le t}\log\left(1 - \frac{P(t,u](e^{\kappa Y_i}-1)}{\lambda + P(0,t]}\right).$$

The best predictor is linear in N. We say a predictor H is a linear functional of the counting process N if it is of the form

$$H = b + \int_0^t\int_0^\infty c(s,y)N(ds, dy),$$

for some number b and function $c(s, y)$. The following result describes the best linear predictor under a weaker assumption on the prior distribution.

THEOREM 2 *If Assumption* 2 *holds, then the best linear predictor of* $S(t, u]$ *with respect to the LINEX criterion is given by a function* c *that depends only on* y, i.e. $c(s, y) = c(y)$. *For every* y, $c(y)$ *is the unique solution to the following equation*

$$\psi'(P(t,u](e^{\kappa y}-1) + P(0,t](e^{-\kappa c(y)}-1)) = e^{\kappa c(y)}\psi'(0).$$

The intercept term of the best predictor is

$$b = \frac{1}{\kappa}\int_0^{y^*}\psi(P(t,u](e^{\kappa y}-1) + P(0,t](e^{-\kappa c(y)}-1))\alpha(dy).$$

The above formulas hold under the condition that

$$y^* < \frac{1}{\kappa}\log\left[1 + \frac{(r^* + P(0,t])}{P(t,u]}\right].$$

We should note that the predictors in Theorems 1 and 2 are finite only when the measure α and consequently also the claim size distribution are concentrated on a bounded interval $(0, y^*]$. Moreover, the upper bound for y^* depends on the exposure and tends to 0 as $P(t, u] \to \infty$.

Therefore, the LINEX criterion combined with the Poisson/Gamma model has applicability limited to insurance products for which the claims are bounded random variables. This condition is satisfied by some products with a finite sum insured. The boundedness condition is also satisfied under an excess-loss reinsurance agreement, whereby the ceding company pays the amount bounded above by the retention level.

It is perhaps of some interest to compare the LINEX predictors to the best predictors with respect to the classical square criterion $L(y) = y^2$.

PROPOSITION 1 *If Assumption 1 holds, then the best predictor of $S(t,u]$ with respect to the square criterion is given by*

$$H = \frac{P(t,u] \int_0^\infty y\,[\alpha(\mathrm{d}y) + N((0,t],\mathrm{d}y)]}{\lambda + P(0,t]} = \frac{\lambda}{\lambda + P(0,t]} \mathbb{E}S(t,u] + \frac{P(t,u]}{\lambda + P(0,t]} S(0,t].$$

PROPOSITION 2 *If Assumption 2 holds, then the best linear predictor of $S(t,u]$ with respect to the square criterion is given by the following function c and the intercept term b*

$$c(s,y) = c(y) = \frac{\psi''(0)P(t,u]y}{\psi'(0) + \psi''(0)P(0,t]},$$

$$b = P(t,u]\frac{(\psi'(0))^2}{\psi'(0) + P(0,t]\psi''(0)} \int_0^\infty y\alpha(\mathrm{d}y).$$

Put differently, the best linear predictor is

$$H = \frac{\psi'(0)}{\psi'(0) + \psi''(0)P(0,t]} \mathbb{E}S(t,u] + \frac{\psi''(0)P(t,u]}{\psi'(0) + \psi''(0)P(0,t]} S(0,t].$$

Let us mention that the predictors in Propositions 1 and 2 are not only linear functionals of the counting process N but also linear functions of the random variable $S(0,t]$, in contrast to the LINEX predictors. Let us also note that the LINEX criterion for small κ becomes close to the square criterion. It is indeed not hard to verify that, for $\kappa \to 0$, the formulas in Theorems 1 and 2 reduce to those in Propositions 1 and 2, respectively.

4. Proofs

We will repeatedly use the following lemma.

LEMMA *Suppose that Assumption 2 holds and g is a Borel measurable real function such that $\int_0^\infty |g(y)|\Theta(\mathrm{d}y) < \infty$ a.s. Then,*

$$\mathbb{E}\int_0^\infty g(y)\Theta(\mathrm{d}y) = \psi'(0)\int_0^\infty g(y)\alpha(\mathrm{d}y) \tag{8}$$

in the sense that the LHS exists iff the RHS exists. If (8) is finite, then

$$\mathrm{Var}\int_0^\infty g(y)\Theta(\mathrm{d}y) = \psi''(0)\int_0^\infty g(y)^2\alpha(\mathrm{d}y). \tag{9}$$

Moreover,

$$\mathbb{E}\exp\left[\int_0^\infty g(y)\Theta(\mathrm{d}y)\right] = \exp\left[\int_0^\infty \psi\left(g(y)\right)\alpha(\mathrm{d}y)\right], \tag{10}$$

where both sides of this equation may be finite or both infinite.

This Lemma is well known and rather obvious. Analogous formulas hold also for the process N. Recall that N is, conditional on Θ, a Poisson process. We thus have $\mathbb{E}_\Theta \exp[rN((t,u],(x,y]) = \exp[P(t,u]\Theta(x,y](\mathrm{e}^r - 1)]$. It follows that

$$\mathbb{E}_\Theta \int_0^\infty \int_t^u g(s,y)N(\mathrm{d}s,\mathrm{d}y) = \int_0^\infty \int_t^u g(s,y)P(\mathrm{d}s)\Theta(\mathrm{d}y), \tag{11}$$

$$\mathrm{Var}_\Theta \int_0^\infty \int_t^u g(s,y)N(\mathrm{d}s,\mathrm{d}y) = \int_0^\infty \int_t^u g(s,y)^2 P(\mathrm{d}s)\Theta(\mathrm{d}y), \tag{12}$$

$$\mathbb{E}_\Theta \exp\left[\int_0^\infty \int_t^u g(s,y)N(\mathrm{d}s,\mathrm{d}y)\right] = \exp\left[\int_0^\infty \int_t^u (\mathrm{e}^{g(s,y)} - 1)P(\mathrm{d}s)\Theta(\mathrm{d}y)\right]. \tag{13}$$

Proof of Theorem 1 The prior distribution of Θ is described by Assumption 1. Applying equations (10) and (13) combined with equation (5), we obtain

$$\begin{aligned}\mathbb{E}\mathrm{e}^{\kappa S(t,u]} &= \mathbb{E}\mathbb{E}_\Theta \exp\left[\kappa \int_0^{y^*} \int_t^u yN(\mathrm{d}s,\mathrm{d}y)\right] \\ &= \mathbb{E}\exp\left[\int_0^{y^*} P(t,u](\mathrm{e}^{\kappa y} - 1)\Theta(\mathrm{d}y)\right] \\ &= \exp\left[\int_0^{y^*} \log \frac{\lambda}{\lambda - P(t,u](\mathrm{e}^{\kappa y} - 1)}\alpha(\mathrm{d}y)\right].\end{aligned}$$

Hence,

$$\frac{1}{\kappa}\log \mathbb{E}\mathrm{e}^{\kappa S(t,u]} = -\frac{1}{\kappa}\int_0^{y^*} \log\left(1 - \frac{P(t,u](\mathrm{e}^{\kappa y} - 1)}{\lambda}\right)\alpha(\mathrm{d}y).$$

Let us mention that the above formula can be interpreted as the best prior LINEX predictor.

It is easily seen that the posterior distribution of Θ is also a Gamma process. The prior parameters (α,λ) are replaced with the posterior ones, $(\alpha(\cdot) + N((0,t],\cdot), \lambda + P(0,t])$. Consequently, in view of equation (7), the posterior LINEX predictor H is

$$\frac{1}{\kappa}\log \mathbb{E}(\mathrm{e}^{\kappa S(t,u]}|\mathcal{F}_t^N) = -\frac{1}{\kappa}\int_0^{y^*} \log\left(1 - \frac{P(t,u](\mathrm{e}^{\kappa y} - 1)}{\lambda + P(0,t]}\right)[\alpha(\mathrm{d}y) + N((0,t],\mathrm{d}y)].$$

■

Proof of Theorem 2 We are to minimize the following functional that depends on a real number b and a function $c = c(\cdot,\cdot)$

$$\mathcal{Q}(b,c) = \mathbb{E}L(S - Z(c) - b),$$

where

$$S = \int_0^{y^*} \int_t^u y\,\mathrm{d}N(\mathrm{d}s,\mathrm{d}y), \quad Z(c) = \int_0^{y^*} \int_0^t c(s,y)N(\mathrm{d}s,\mathrm{d}y) \tag{14}$$

and $L(y) = \mathrm{e}^{\kappa y} - \kappa y - 1$. In order to compute $\mathcal{Q}(b,c)$, let us first condition on Θ. We have

$$\begin{aligned}\mathbb{E}_\Theta L(S - Z(c) - b) &= \mathbb{E}_\Theta \exp[\kappa(S - Z(c) - b)] - \mathbb{E}_\Theta[\kappa(S - Z(c) - b)] - 1 \\ &= \mathbb{E}_\Theta \exp[\kappa S]\,\mathbb{E}_\Theta \exp[-\kappa Z(c)]\mathrm{e}^{-\kappa b} - \kappa\mathbb{E}_\Theta S + \kappa\mathbb{E}_\Theta Z(c) + \kappa b - 1,\end{aligned} \tag{15}$$

because S and $Z(c)$ are conditionally independent, given Θ. From equation (13), it follows that

$$\begin{aligned}\mathbb{E}_\Theta \exp[\kappa S] &= \mathbb{E}_\Theta \exp\left[\kappa \int_0^{y^*}\int_0^t yN(\mathrm{d}s,\mathrm{d}y)\right] \\ &= \exp\left[\int_0^{y^*}\int_0^t (\mathrm{e}^{\kappa y}-1)P(\mathrm{d}s)\Theta(\mathrm{d}y)\right]\end{aligned} \tag{16}$$

and

$$\begin{aligned}\mathbb{E}_\Theta \exp[-\kappa Z(c)] &= \mathbb{E}_\Theta \exp\left[-\kappa \int_0^{y^*}\int_0^t c(s,y)N(\mathrm{d}s,\mathrm{d}y)\right] \\ &= \exp\left[\int_0^{y^*}\int_0^t (\mathrm{e}^{-\kappa c(s,y)}-1)P(\mathrm{d}s)\Theta(\mathrm{d}y)\right].\end{aligned} \tag{17}$$

Similarly, from equation (11), it follows that

$$\mathbb{E}_\Theta S = \mathbb{E}_\Theta \int_0^{y^*}\int_0^t yN(\mathrm{d}s,\mathrm{d}y) = \int_0^{y^*}\int_0^t yP(\mathrm{d}s)\Theta(\mathrm{d}y) \tag{18}$$

and

$$\mathbb{E}_\Theta Z(c) = \mathbb{E}_\Theta \int_0^{y^*}\int_0^t c(s,y)N(\mathrm{d}s,\mathrm{d}y) = \int_0^{y^*}\int_0^t c(s,y)P(\mathrm{d}s)\Theta(\mathrm{d}y). \tag{19}$$

We are now in a position to prove that the optimum function c depends only on y and not on s. Indeed, let us define

$$\bar{c}(y) = \frac{\int_0^t c(s,y)P(\mathrm{d}s)}{P(0,t]}. \tag{20}$$

In view of equation (17), by Jensen's inequality, $\mathbb{E}_\Theta \exp[-\kappa Z(c)] \geq \mathbb{E}_\Theta \exp[-\kappa Z(\bar{c})]$. It is clear that $\mathbb{E}_\Theta Z(c) = \mathbb{E}_\Theta Z(\bar{c})$. Consequently, $\mathbb{E}_\Theta L(S - Z(c) - b) \geq \mathbb{E}_\Theta L(S - Z(\bar{c}) - b)$ and therefore $\mathcal{Q}(b,c) \geq \mathcal{Q}(b,\bar{c})$. Hence, it is sufficient to minimize $\mathcal{Q}(b,c)$ under the assumption that $c(s,y) = c(y) = \bar{c}(y)$. This will allow us to write equation (15) in a simpler form. Let

$$\mathcal{E}_\Theta(c) = \exp\left[\int_0^{y^*} (p_1(\mathrm{e}^{\kappa y}-1) + p_0(\mathrm{e}^{-\kappa c(y)}-1))\Theta(\mathrm{d}y)\right] \tag{21}$$

and

$$\mathcal{R}_\Theta(c) = \kappa \int_0^{y^*} (p_1 y - p_0 c(y))\Theta(\mathrm{d}y), \tag{22}$$

where

$$p_0 = P(0,t], \quad p_1 = P(t,u].$$

Combining equations (16)–(19), we can rewrite equation (15) as follows:

$$\mathbb{E}_\Theta L(S - Z(c) - b) = \mathcal{E}_\Theta(c)\mathrm{e}^{-\kappa b} - \mathcal{R}_\Theta(c) + \kappa b - 1.$$

Our next step is to compute $\mathcal{Q}(b,c) = \mathbb{E}\mathbb{E}_\Theta L(S - Z(c) - b)$. Let us write $\mathcal{E}(c) = \mathbb{E}\mathcal{E}_\Theta(c)$ and $\mathcal{R}(c) = \mathbb{E}\mathcal{R}_\Theta(c)$. We proceed in much the same way as before. Combining equation (21) with equation (10), we obtain

$$\mathcal{E}(c) = \exp\left[\int_0^{y^*} \psi\big(p_1(\mathrm{e}^{\kappa y} - 1) + p_0(\mathrm{e}^{-\kappa c(y)} - 1)\big)\alpha(\mathrm{d}y)\right]. \tag{23}$$

Similarly, combining equation (22) with equation (8), we obtain

$$\mathcal{R}(c) = \kappa\psi'(0)\int_0^{y^*} (p_1 y - p_0 c(y))\alpha(\mathrm{d}y). \tag{24}$$

Clearly,

$$\mathcal{Q}(b, c) = \mathbb{E}\mathbb{E}_\Theta L(S - Z(c) - b) = \mathcal{E}(c)\mathrm{e}^{-\kappa b} - \mathcal{R}(c) + \kappa b - 1.$$

We are now going to exploit the fact that $\mathcal{E}(c)$ and $\mathcal{R}(c)$ are expressions that depend only on c and not on b. For a given function c, it is easy to minimize $\mathcal{Q}(b, c)$ with respect to b. Function $\mathcal{Q}(\cdot, c)$ is convex and its unique minimum can be found by equating the derivative to zero

$$\frac{\partial}{\partial b}\mathcal{Q}(b, c) = -\kappa\mathcal{E}(c)\mathrm{e}^{-\kappa b} + \kappa = 0,$$

hence $\mathcal{E}(c) = \mathrm{e}^{\kappa b}$. Therefore, substituting the optimum value $b = (1/\kappa)\log\mathcal{E}(c)$ into $\mathcal{Q}(b, c)$, we obtain

$$\mathcal{Q}(c) = \min_b \mathcal{Q}(b, c) = -\mathcal{R}(c) + \log\mathcal{E}(c). \tag{25}$$

Substituting equations (23) and (24) into equation (25), we obtain

$$\mathcal{Q}(c) = \int \left[\psi(p_1(\mathrm{e}^{\kappa y} - 1) + p_0(\mathrm{e}^{-\kappa c(y)} - 1)) - \kappa\psi'(0)(p_1 y + p_0 c(y))\right]\alpha(\mathrm{d}y).$$

To minimize $Q(c)$, it is sufficient to minimize the expression under the integral with respect to $c(y)$ separately for every y. Slightly modifying the notation, let us fix y and consider the following function of a real variable c

$$\mathcal{K}(c) = \psi(p_1(\mathrm{e}^{\kappa y} - 1) + p_0(\mathrm{e}^{-\kappa c} - 1)) - \kappa\psi'(0)(p_1 y + p_0 c).$$

From the Lèvy–Khinchine representation formula, we know that ψ is a convex function. As

$$\mathcal{K}'(c) = \frac{\mathrm{d}}{\mathrm{d}c}\mathcal{K}(c) = \kappa p_0\left[-\psi'(p_1(\mathrm{e}^{\kappa y} - 1) + p_0(\mathrm{e}^{-\kappa c} - 1))\mathrm{e}^{-\kappa c} + \psi'(0)\right],$$

we infer that the derivative $\mathcal{K}'$ is increasing, thus $\mathcal{K}$ is convex. It can easily be verified that $\mathcal{K}'(c) \to \kappa p_0\psi'(0) > 0$ for $c \to \infty$. If $p_1(\mathrm{e}^{\kappa y} - 1) < r^*$, then $\mathcal{K}'(0) < 0$, otherwise $\mathcal{K}'(c) \to -\infty$ as $c \searrow c_*$ for some $c_* \geq 0$. Consequently, the equation $\mathcal{K}'(c) = 0$ has a unique positive solution.

Thus, we have derived the asserted equation for c, and we have shown that it uniquely defines c. The formula for b follows immediately from equation (23) because $b = (1/\kappa)\log\mathcal{E}(c)$. The proof is complete. ■

Proof of Proposition 1 The best predictor with respect to the square criterion is $\mathbb{E}(S(t, u]|\mathcal{F}_t^N)$. The computation is a straightforward application of equations (8) and (11) to the posterior distribution of Θ. We omit easy details. ■

Proof of Proposition 2 The proof is quite similar to that of Theorem 2 but with equations (9) and (12) taking over the roles played by equations (10) and (13). We are to minimize

$$\mathcal{Q}(b, c) = \mathbb{E}(S - Z(c) - b)^2,$$

where S and $Z(c)$ are defined in equation (14). It is clear that minimum is attained for

$$b = \mathbb{E}(S - Z(c))$$

and

$$\mathcal{Q}(c) = \min_b \mathcal{Q}(b, c) = \operatorname{Var}(S - Z(c)) = \operatorname{Var} \mathbb{E}_\Theta(S - Z(c)) + \mathbb{E} \operatorname{Var}_\Theta(S - Z(c)).$$

By conditional independence and equation (12),

$$\begin{aligned}\operatorname{Var}_\Theta(S - Z(c)) &= \operatorname{Var}_\Theta S + \operatorname{Var}_\Theta Z(c) \\ &= \int_0^\infty \int_t^u y^2 P(\mathrm{d}s)\Theta(\mathrm{d}y) + \int_0^\infty \int_0^t c(s, y)^2 P(\mathrm{d}s)\Theta(\mathrm{d}y).\end{aligned}$$

The conditional expectations $\mathbb{E}_\Theta S$ and $\mathbb{E}_\Theta Z(c)$ are given by equations (18) and (19). Let us define $\bar{c}(y)$ as in the proof of Theorem 2 by equation (20). It is easily seen that $\mathbb{E}_\Theta Z(c) = \mathbb{E}_\Theta Z(\bar{c})$ and $\operatorname{Var}_\Theta Z(c) \geq \operatorname{Var}_\Theta Z(\bar{c})$. Consequently, without loss of generality, we can assume that $c(s, y) = c(y) = \bar{c}(y)$. Now we can write

$$\begin{aligned}\mathbb{E}_\Theta(S - Z(c)) &= \int_0^\infty [p_1 y - p_0 c(y)]\, \Theta(\mathrm{d}y), \\ \operatorname{Var}_\Theta(S - Z(c)) &= \int_0^\infty \left[p_1 y^2 + p_0 c(y)^2\right] \Theta(\mathrm{d}y),\end{aligned} \tag{26}$$

where $p_0 = P(0, t]$ and $p_1 = P(t, u]$. Application of equations (8) and (9) yields

$$\begin{aligned}\mathbb{E} \operatorname{Var}_\Theta(S - Z(c)) &= \psi'(0) \int_0^\infty \left[p_1 y^2 + p_0 c(y)^2\right] \alpha(\mathrm{d}y), \\ \operatorname{Var} \mathbb{E}_\Theta(S - Z(c)) &= \psi''(0) \int_0^\infty [p_1 y - p_0 c(y)]^2\, \alpha(\mathrm{d}y).\end{aligned} \tag{27}$$

Therefore,

$$\mathcal{Q}(c) = \int_0^\infty \left\{\psi''(0)\, [p_1 y - p_0 c(y)]^2 + \psi'(0) \left[p_1 y^2 + p_0 c(y)^2\right]\right\} \alpha(\mathrm{d}y).$$

The minimum of $\mathcal{Q}(c)$ obtains if, for every y, we minimize the function under the integral with respect to $c(y)$. An easy computation gives the formula for c asserted in the theorem. The formula for b follows trivially and the proof is complete. ■

Note that in the proofs of Propositions 1 and 2, we have not used the assumption that $y^* < \infty$ bounds the support of α.

5. Concluding remarks

The choice of the parameters of the prior distribution is certainly important for applications. This problem is not discussed in this article. A rather straightforward part of modeling is to choose the prior intensity measure $\mathbb{E}\Theta$ which is equal to α/λ under Assumption 1. Its natural (empirical Bayes) estimate can be based on the empirical measure that counts past loss occurrences in the whole portfolio of risks. On the other hand, choosing a suitable parameter λ is much more difficult.

Acknowledgement

The author is grateful to the referee who suggested that the time component of the process can be introduced in the model explicitly.

References

[1] Goovaerts, M.J., Kaas, R., van Heervaarden, A.E. and Bauwelinckx, T., 1990, *Effective Actuarial Methods* (North Holland).
[2] Klugman, S.A., 1992, *Bayesian Statistics in Actuarial Science, With Emphasis on Credibility* (Kluwer Academic Publishers).
[3] Teugels, J.L. and Sundt, B. (Eds), 2004, *Encyclopedia of Actuarial Science* (Wiley).
[4] Ferguson, T.S., 1973, A Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209–230.
[5] Kingman, J.F.C., 1993, *Poisson Processes* (Clarendon Press).
[6] Grandell, J., 1975, Doubly stochastic poisson processes. Institute of Actuarial Mathematics and Mathematical Statistics, University of Stockholm.
[7] Zellner, A., 1986, Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, **81**(394), 446–451.
[8] Gerber, H.U., 1979, *An Introduction to Mathematical Risk Theory* (S.S. Huebner Foundation for Insurance Education).