

# Bayesian prediction in doubly stochastic Poisson process

Alicja Jokiel-Rokita · Daniel Lazar ·  
Ryszard Magiera

Received: 3 July 2013 / Published online: 15 February 2014  
© The Author(s) 2014. This article is published with open access at Springerlink.com

**Abstract** A stochastic marked point process model based on doubly stochastic Poisson process is considered in the problem of prediction for the total size of future marks in a given period, given the history of the process. The underlying marked point process  $(T_i, Y_i)_{i \geq 1}$ , where  $T_i$  is the time of occurrence of the  $i$ th event and the mark  $Y_i$  is its characteristic (size), is supposed to be a non-homogeneous Poisson process on  $\mathbb{R}_+^2$  with intensity measure  $P \times \Theta$ , where  $P$  is known, whereas  $\Theta$  is treated as an unknown measure of the total size of future marks in a given period. In the problem of prediction considered, a Bayesian approach is used assuming that  $\Theta$  is random with prior distribution presented by a gamma process. The best predictor with respect to this prior distribution is constructed under a precautionary loss function. A simulation study for comparing the behavior of the predictors under various criteria is provided.

**Keywords** Bayes prediction · Doubly stochastic Poisson process · Random measure · Precautionary loss

## 1 Introduction

Marked point process models are used in studying the reliability of technical systems, in failure-repair process analysis, in survival analysis for biological units, in insurance problems, in meteorology, in seismology and in other sciences. There is a vast literature on the theory of point processes and their applications; see, for example, Andersen et al. (1993), Cox (2006), Grandell (1975), Kallenberg (1986), Kingman (1964), Kingman (1993), Kingman (2006), Last and Brandt (1995), Mikosch (2009) and Møller and Waagepetersen (2002).

---

A. Jokiel-Rokita · D. Lazar · R. Magiera (✉)  
Institute of Mathematics and Computer Science, Wrocław University of Technology, Wrocław, Poland  
e-mail: ryszard.magiera@pwr.wroc.pl

Let  $(T_i)_{i \geq 1}$  be a sequence of time points in which events occur ( $T_0 = 0$ ), and let  $(Y_i)_{i \geq 1}$  be a sequence of the corresponding marks (additional characteristics of the events, for instance censoring indicators, damage sizes, cost values, repair types, repair times, stress values, atmospheric precipitation levels). Considering the double sequence  $(T_i, Y_i)_{i \geq 1}$  (the marked point process) one usually associates to it the counting processes  $\{N(t, A), t \geq 0\}$ , where

$$N(t, A) = \sum_{i=1}^{\infty} \mathbf{1}(T_i \leq t) \mathbf{1}(Y_i \in A), \quad A \subset \mathbb{R},$$

with  $N(t) := N(t, \mathbb{R})$ . It is then assumed that there exists a finite expectation measure (intensity measure)  $\Lambda(t, A) = E(N(t, A)) = \int_0^t \lambda(s, A) ds = \int_0^t \int_A \lambda(s, dy) ds$ . The measure  $\Lambda(t, A)$  is the compensator of the process  $\{N(t, A)\}$  and  $\{N(t, A) - \Lambda(t, A), t \geq 0\}$ ,  $A \subset \mathbb{R}$ , is a martingale.

In the purpose of the statistical problem considered, we describe the counting process  $N$  in the form

$$N((t, u], (x, y]) = \sum_{i=1}^{\infty} \mathbf{1}(t < T_i \leq u, x < Y_i \leq y). \quad (1.1)$$

In this paper we deal with the problem of predicting the total value (size) of the marks (e.g. the total damage value, total precipitation level or the total claim amount) appeared in the future time interval  $(t, u]$  in a special case of a marked counting process, namely in a doubly stochastic Poisson process model. Doubly stochastic Poisson processes have been frequently used in many areas of science.

The underlying marked point process  $(T_i, Y_i)_{i \geq 1}$  is supposed to be a non-homogeneous Poisson process on  $\mathbb{R}_+^2$  with intensity measure  $P \times \Theta$ , where  $P$  is known, whereas  $\Theta$  is treated as an unknown measure of the total size of future marks in a given period. In the problem of prediction considered, a Bayesian approach is used assuming that  $\Theta$  is random with prior distribution presented by a gamma process. The model was used by [Niemi \(2006\)](#) in the problem of prediction under the LINEX loss function. The Bayesian prediction under quadratic loss function was considered by [Grandell \(1975\)](#). In this paper the best predictor is derived under a precautionary loss function. From a practical point of view, in many situations the precautionary loss function considered can be more appropriate than the LINEX or a quadratic loss because of its right boundedness property (see next section).

The paper is organized as follows. In Sect. 2 the process model considered is presented. Section 3 contains the comparison of the precautionary loss function with the LINEX loss function and a quadratic loss in the context of their use in the problem of prediction considered. The best predictor under a precautionary loss function is derived in Sect. 4. In Sect. 5 a simulation study for comparing the behavior of the predictors under various criteria is provided.

## 2 The model

We will consider the counting process  $N$  of (1.1) based on a doubly stochastic Poisson process. The observed random vectors  $(T_i, Y_i)$  define the random measure  $N$  which counts the events of specified size occurred in a time period  $(t, u]$ . The value  $N((t, u], (x, y])$  is then the number of points in the rectangle  $\{(t, u], (x, y]\}$ , where  $0 \leq t < u < \infty; 0 < x < y < \infty$ .

Let us note that

$$N((t, u], \mathbb{R}) = \sum_{i=1}^{\infty} \mathbf{1}(t < T_i \leq u)$$

denotes the number of events appeared in the time interval  $(t, u]$ .

Assume that  $P$  is a known Radon measure on  $\mathbb{R}_+$ . It describes the diversity of risk over time and can model for example seasonality. On the other hand,  $\Theta$  is a finite random measure. It characterizes the size of marks and also the general intensity of occurrence of events. It is supposed that, conditional on  $\Theta$ ,  $N$  is a non-homogeneous Poisson process with mean measure  $P \times \Theta$ . We can also think of  $N$  as a marked Poisson process  $(N_1(\cdot), Y_i)_{i \geq 1}$ , where  $N_1(\cdot) = \sum_{i=1}^{\infty} \mathbf{1}_{(\cdot)}(T_i)$  is a non-homogeneous Poisson process with mean measure  $P(\cdot)\Theta(\mathbb{R}_+)$  and  $(Y_i)$  is the mark sequence with distribution  $\Theta(\cdot)/\Theta(\mathbb{R}_+)$ .

In the sequel we shall assume that the process  $N$  is the two-dimensional non-homogeneous Poisson process with the expectation measure

$$\Lambda((t, u], (x, y]) = P(t, u]\Theta(x, y],$$

i.e.

$$N((t, u], (x, y]) \Big|_{\Theta} = \mathcal{P}\left(P(t, u]\Theta(x, y)\right),$$

where  $\mathcal{P}(\lambda)$  denotes the Poisson distribution with parameter  $\lambda$ .

Let us also remark that the conditional expected value of the process  $N$ , given the realization of the process  $\Theta(x, y]$  is given by the formula

$$E_{\Theta} N((t, u], (x, y]) = P(t, u]\Theta(x, y].$$

Assuming the points  $(T_i, Y_i)$  are sorted with respect to the first axis, the points  $T_i$  form one-dimensional point process with rescaled expectation measure  $P(\cdot)$ . On the other hand, assuming the points  $(T_i, Y_i)$  are sorted with respect to the second axis, the points  $Y_i$  form one-dimensional point process with rescaled expectation measure  $\Theta(\cdot)$ . Denoting these processes by  $N_1(t, u]$  and  $N_2(x, y]$ , correspondingly, we observe that  $N_1(t, u]$  and  $N_2(x, y]$  are the Poisson processes with the corresponding expectation measures  $P(t, u]\Theta(\mathbb{R}_+)$  and  $P(\mathbb{R}_+)\Theta(x, y]$ .

Our goal is to predict the total value of the marks (e.g. the total damage size, total precipitation level or the total claim amount) appeared in the time interval  $(t, u]$ , i.e. we would like to predict the value

$$S(t, u] = \sum_{i=1}^{\infty} Y_i \mathbf{1}(t < T_i \leq u) = \int_t^u \int_0^{\infty} yN(ds, dy). \quad (2.1)$$

Note that the notation above follows from the Riemann-Stieltjes integral construction, and that  $S(t, u]$  is a stochastic process.

### 3 Precautionary loss versus LINEX and quadratic losses

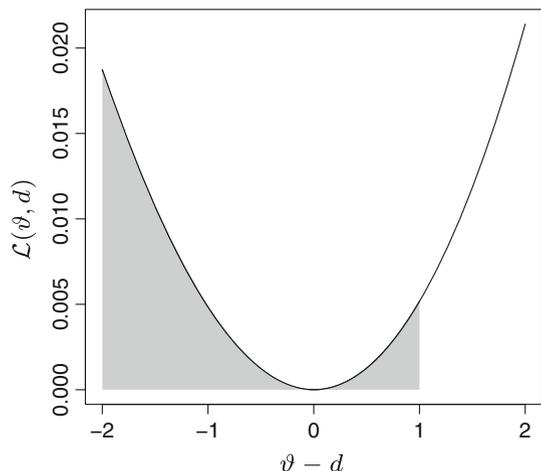
Our goal is to predict  $S(t, u]$ , i.e. the total value (size) of marks in the time period  $(t, u]$ . This predictor could be used for calculating, for example, the total size of damages in a reliability system or capital reserves needed to cover claims that will occur in the future. Underestimating of these values could lead to endangering the functionality of a reliability system or to insolvency of an insurance company. That is why underestimating can be more dangerous than overestimating. Using asymmetric criteria is a natural way to deal with this kind of situations. A very well known asymmetric loss function is the LINEX function, i.e. the loss function of the form

$$\mathcal{L}(\vartheta, d) = K \{\exp[\kappa(\vartheta - d)] - \kappa(\vartheta - d) - 1\}, \quad K > 0, \kappa > 0. \quad (3.1)$$

It is a function with many applications and worth being considered although there are also some disadvantages. It may be difficult to choose the right parameter  $\kappa$  especially in the situation where we desire to estimate a nonnegative parameter, whereas the LINEX loss function is defined for any real value of the parameter and its estimate. For a wrong choice of  $\kappa$  the most of the exponential part of the LINEX curve (which is penalizing underestimating) might fall into a region where the estimate should be negative (see Fig. 1).

In many practical problems it is desirable to take into account a loss function which tends to  $\infty$  as  $d \rightarrow 0^+$  independently of the function parameters. A family of functions fulfilling this assumption was introduced by [Norström \(1996\)](#):

**Fig. 1** The LINEX loss function for  $K = 1$ ,  $\kappa = 0.1$  and  $\vartheta = 1$ . Only the *shadowed part* is being used when estimating a nonnegative parameter



$$\mathcal{L}(\vartheta, d) = w(\vartheta) \frac{(\vartheta - d)^2}{d^k}, \quad 0 < k \leq 2, \tag{3.2}$$

where  $w(\vartheta) > 0$  is a weight function and  $d > 0$ . Note that for  $k = 0$  the above formula is also correct and it is a weighted quadratic loss function. In the case  $0 < k \leq 2$  it is called *precautionary loss function*. This loss was considered by [Norström \(1996\)](#) in parametric problems for risk analysis. A precautionary index  $k$  regulates how downside damaging the loss function is. The loss function becomes more precautionary as  $k$  increases. A way to determine  $k$  (see [Norström 1996](#)) is to: (1) select two values of  $d$ :  $\vartheta - \varepsilon$  and  $\vartheta + b\varepsilon$ , where  $b > 1$ , which have equal loss for the decision maker, (2) solve  $\mathcal{L}(\vartheta, \vartheta - \varepsilon) = \mathcal{L}(\vartheta, \vartheta + b\varepsilon)$  with respect to  $k$ . For example  $b = 2$  gives  $k = 1$ , and  $b = 3$  gives  $k = 2$ .

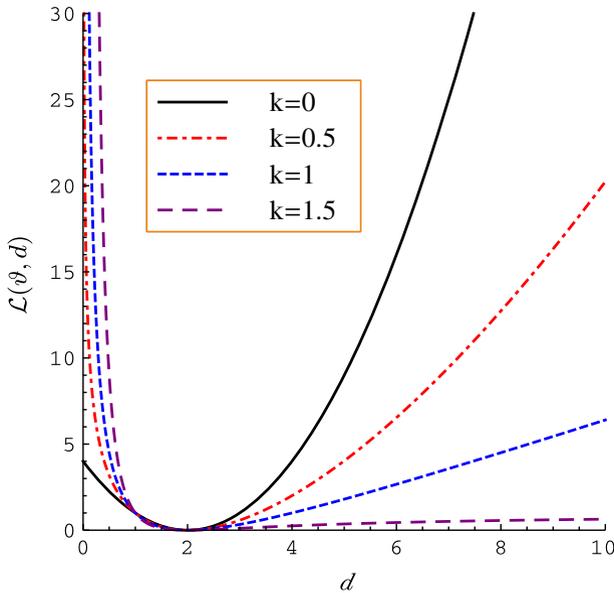
The condition  $k \leq 2$  means that the loss increases as the difference  $\vartheta - d$  increases. Example curves for different values of  $k$  are shown in [Fig. 2](#).

For  $k = 2$  and a given value  $\vartheta$ ,  $\mathcal{L}(\vartheta, d) \rightarrow w(\vartheta)$  as  $d \rightarrow \infty$ , which means that the precautionary loss function is right bounded (see [Figs. 3, 4](#)).

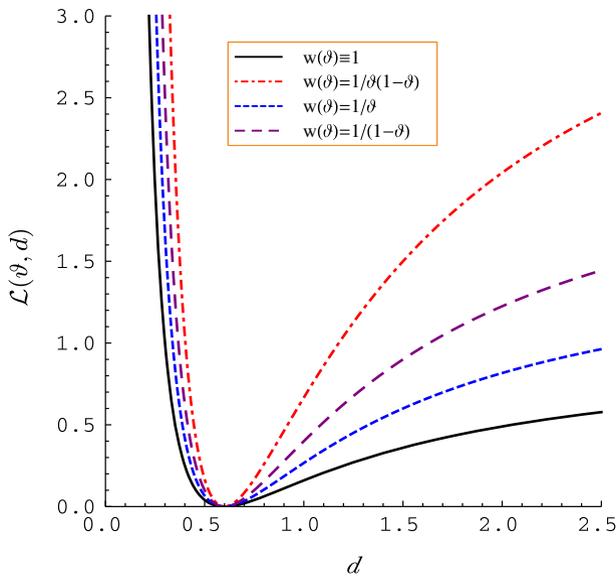
Since in the model considered  $\vartheta = S(t, u)$  is equal to 0 with positive probability, we omit the weight function  $w(\vartheta)$ . Thus we use the loss function

$$\mathcal{L}_k(\vartheta, d) = \frac{(\vartheta - d)^2}{d^k}, \quad 0 < k \leq 2. \tag{3.3}$$

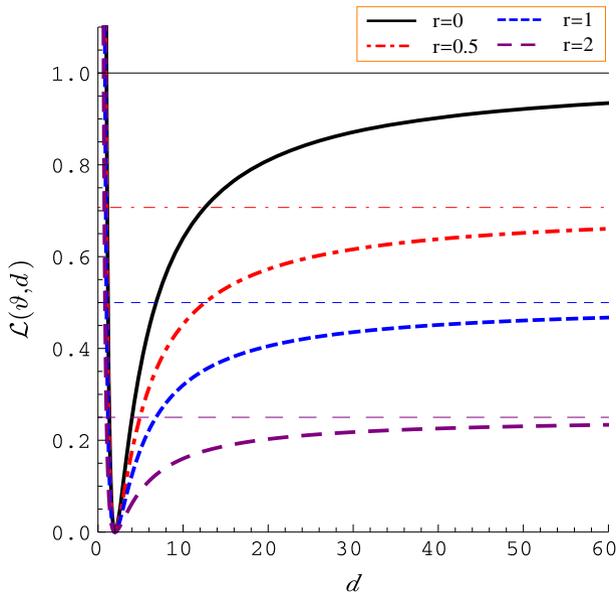
Let us note that the precautionary loss function converges to the square error loss function



**Fig. 2** The plots of the precautionary loss function for chosen values of  $k$  and  $w(\vartheta) \equiv 1$  ( $\vartheta = 2$ )



**Fig. 3** The plots of the precautionary loss function for  $k = 2$  and various weights  $w(\vartheta)$  ( $\vartheta = 0.6$ )



**Fig. 4** The plots of the precautionary loss function for  $k = 2$  and the weight function  $w(\vartheta) = 1/\vartheta^r$  for various  $r$  ( $\vartheta = 2$ )

$$\mathcal{L}(\vartheta, d) = (\vartheta - d)^2 \tag{3.4}$$

as  $k \rightarrow 0$ , and that for  $\kappa$  close to 0 the LINEX loss is close to the quadratic loss function.

### 4 The Bayes prediction under the precautionary loss function

#### 4.1 The Bayes approach

We assume that the finite random measure  $\Theta$  is a gamma process:

$$\Theta(x, y] \sim \text{Gamma}(\alpha(x, y], \lambda), \tag{4.1}$$

such that  $\int_0^\infty y^2 \Theta(dy) < \infty$  almost sure, where  $\alpha$  is a finite measure on  $(0, \infty)$  describing the shape of the gamma distribution and  $\lambda$  is the scale parameter. In other words we impose on  $\Theta(x, y]$  the prior distribution  $\pi(\theta) = \pi(\theta; \alpha(x, y], \lambda)$  to be the gamma distribution with the parameters  $\alpha(x, y]$  and  $\lambda$ . The use of the gamma process is justified by the fact that it is conjugate prior to the Poisson process and thereby the posterior process is also a gamma process.

The posterior distribution of  $\Theta(x, y]$  given  $N((0, t], (x, y])$  is proportional to

$$\begin{aligned} & P\left(N((0, t], (x, y]) = k \mid \Theta(x, y] = \theta\right) \pi(\theta) \\ &= \exp(-P(0, t]\theta) \frac{(P(0, t]\theta)^k}{k!} \frac{\lambda^{\alpha(x, y]}}{\Gamma(\alpha(x, y])} \theta^{\alpha(x, y]-1} \exp(-\lambda\theta) \\ &\propto \theta^{k+\alpha(x, y]-1} \exp[-(P(0, t] + \lambda)\theta]. \end{aligned}$$

Thus the posterior process  $\Theta(x, y] \mid N((0, t], (x, y])$  is the gamma process  $\text{Gamma}(N((0, t], (x, y]) + \alpha(x, y], P(0, t] + \lambda)$ .

Remark that the moment generating function of the increments of the process  $\Theta$  is of the form:

$$E \exp[r\Theta(x, y)] = \exp[\alpha(x, y]\psi(r)], \tag{4.2}$$

where  $\psi(r) = \ln\left(\frac{\lambda}{\lambda-r}\right)$ . This fact will be needed later.

#### 4.2 The best predictor

Let

$$\mathcal{F}_t = \sigma\{N((v, s], (x, y]) : 0 \leq v < s \leq t, 0 < x < y < \infty\}$$

be the  $\sigma$ -algebra of events generated by the process  $N$  up to time  $t$ . We will determine the form of the  $\mathcal{F}_t$ -measurable Bayes predictor of the random variable  $S(t, u]$  under the precautionary loss function. We find the optimal predictor by minimizing the expected risk, i.e. the expectation of the loss with respect to the joint measure  $P \times \Theta$ , over all  $\mathcal{F}_t$ -measurable decisions  $d$  (estimators). By the well known fact from Bayesian decision theory we find the best predictor by minimizing the corresponding posterior risk, i.e. we find a decision  $d^*$  that satisfies the condition

$$E(\mathcal{L}(S, d^*) \mid \mathcal{F}_t) = \inf_{d \in \mathcal{D}} E(\mathcal{L}(S, d) \mid \mathcal{F}_t),$$

where  $S = S(t, u)$ ,  $\mathcal{D}$  is the set of all  $\mathcal{F}_t$ -measurable functions and  $E(\mathcal{L}(S, d)|\mathcal{F}_t)$  denotes the expectation with respect to the posterior distribution of  $S$  (the decision  $d$  is fixed).

**Lemma 1** *The Bayes predictor  $d_k$  of  $S = S(t, u)$  under the loss function  $\mathcal{L}_k$  given by (3.3) is determined by the following formula*

$$d_k = \begin{cases} \frac{(1-k)E(S|\mathcal{F}_t) + \sqrt{(1-k)^2[E(S|\mathcal{F}_t)]^2 + k(2-k)E(S^2|\mathcal{F}_t)}}{2-k}, & 0 < k < 2, \\ \frac{E(S^2|\mathcal{F}_t)}{E(S|\mathcal{F}_t)}, & k = 2. \end{cases} \tag{4.3}$$

*Proof* The proof is analogous to that one of [Norström \(1996\)](#). We can treat the posterior risk as the function of a decision  $d$ . We have

$$E(\mathcal{L}(S, d)|\mathcal{F}_t) = \frac{1}{d^k} E(S^2|\mathcal{F}_t) - \frac{2}{d^{k-1}} E(S|\mathcal{F}_t) + \frac{1}{d^{k-2}} =: f(d). \tag{4.4}$$

The derivative of  $f$  is

$$f'(d) = -\frac{k}{d^{k+1}} E(S^2|\mathcal{F}_t) + \frac{2(k-1)}{d^k} E(S|\mathcal{F}_t) - \frac{k-2}{d^{k-1}}. \tag{4.5}$$

The sign of  $f'(d)$  is the same as the sign of

$$g(d) = -kE(S^2|\mathcal{F}_t) + 2(k-1)dE(S|\mathcal{F}_t) - (k-2)d^2. \tag{4.6}$$

Considering the two situations  $k = 2$  and  $0 < k < 2$  it is easy to find the roots of the above function, which minimize the posterior risk. □

The explicit form of the Bayes predictor in the point process model considered is given in [Theorem 1](#). To prove the theorem we need the following lemmas.

**Lemma 2** (See [Mikosch 2009](#)) *Let  $N$  be a non-homogeneous Poisson process on  $A \subset \mathbb{R}^d$  with mean measure  $\mu$  and  $f$  be a real-valued measurable function on  $A$ .*

1. *Assume that  $\int_A |f(x)|\mu(dx) < \infty$ . Then*

$$E\left(\int_A f dN\right) = \int_A f(x)\mu(dx). \tag{4.7}$$

2. *Assume that  $\int_A \max([f(x)]^2, |f(x)|)\mu(dx) < \infty$ . Then*

$$Var\left(\int_A f dN\right) = \int_A [f(x)]^2\mu(dx). \tag{4.8}$$

Using Lemma 2 we obtain the following form of the conditional expected value of the process  $S(t, u]$

$$\begin{aligned}
 E_{\Theta} S(t, u] &= E_{\Theta} \left( \int_t^u \int_0^{\infty} y N(ds, dy) \right) = \int_t^u \int_0^{\infty} y P(ds) \Theta(dy) \\
 &= P(t, u] \int_0^{\infty} y \Theta(dy),
 \end{aligned}
 \tag{4.9}$$

and for the conditional variance

$$\begin{aligned}
 Var_{\Theta} S(t, u] &= Var_{\Theta} \left( \int_t^u \int_0^{\infty} y N(ds, dy) \right) = \int_t^u \int_0^{\infty} y^2 P(ds) \Theta(dy) \\
 &= P(t, u] \int_0^{\infty} y^2 \Theta(dy).
 \end{aligned}
 \tag{4.10}$$

**Lemma 3** (See Niemiro 2006) *Let  $\alpha$  be a finite measure on the interval  $(0, \infty)$ . Assume that  $\Theta$  is the process with independent increments, such that the moment generating function of the increments has the following form*

$$E \exp[r\Theta(x, y)] = \exp[\psi(r)\alpha(x, y)],
 \tag{4.11}$$

where  $\psi$  is a known function. Let  $r^* = \sup\{r : \psi(r) < \infty\}$ . Suppose that  $r^* > 0$  and  $\psi(r) \rightarrow \infty$  as  $r \nearrow r^*$ . Moreover, let  $g$  be a real measurable function such that  $\int_0^{\infty} |g(y)|\Theta(dy) < \infty$  a.s. Then

$$\begin{aligned}
 E \int_0^{\infty} g(y)\Theta(dy) &= \psi'(0) \int_0^{\infty} g(y)\alpha(dy), \\
 Var \int_0^{\infty} g(y)\Theta(dy) &= \psi''(0) \int_0^{\infty} g^2(y)\alpha(dy), \\
 E \exp \left[ \int_0^{\infty} g(y)\Theta(dy) \right] &= \exp \left[ \int_0^{\infty} \psi(g(y))\alpha(dy) \right].
 \end{aligned}$$

**Theorem 1** *If assumption (4.1) is satisfied then the Bayes predictor  $d_k$  of  $S(t, u]$ , under the loss function  $\mathcal{L}_k$  given by (3.3), is determined by the following formula:*

$$d_k = \begin{cases} (1-k)\frac{aI_1}{2-k} + \sqrt{\left(\frac{aI_1}{2-k}\right)^2 + k(1+a)\frac{aI_2}{2-k}}, & 0 < k < 2, \\ aI_1 + (1+a)\frac{I_2}{I_1}, & k = 2, \end{cases} \quad (4.12)$$

where

$$a = \frac{P(t, u]}{\lambda + P(0, t]}, \quad (4.13)$$

$$I_1 = \int_0^\infty y(\alpha(dy) + N((0, t], dy)), \quad (4.14)$$

$$I_2 = \int_0^\infty y^2(\alpha(dy) + N((0, t], dy)). \quad (4.15)$$

*Proof* To determine the Bayesian predictors with respect to the precautionary loss function we need to evaluate  $E(S(t, u]|\mathcal{F}_t)$  and  $E(S^2(t, u]|\mathcal{F}_t)$ . To do this we will first figure out  $ES(t, u]$  and  $ES^2(t, u]$  and then switch to the parameters of the posterior process. We can do so because we have chosen a conjugate prior.

Let us begin with  $ES(t, u]$ . We have already shown that

$$E_{\Theta}S(t, u] = P(t, u] \int_0^\infty y\Theta(dy). \quad (4.16)$$

Applying Lemma 3 to integral (4.16) we get

$$\begin{aligned} ES(t, u] &= EE_{\Theta}S(t, u] = P(t, u]E \int_0^\infty y\Theta(dy) \\ &= P(t, u]\psi'(0) \int_0^\infty y\alpha(dy) = \frac{P(t, u]}{\lambda} \int_0^\infty y\alpha(dy). \end{aligned} \quad (4.17)$$

Now we substitute the parameters of the prior process with the ones of the posterior process. We then have

$$E(S(t, u]|\mathcal{F}_t) = \frac{P(t, u]}{\lambda + P(0, t]} \int_0^\infty y(\alpha(dy) + N((0, t], dy)). \quad (4.18)$$

To evaluate  $ES^2(t, u]$  we will use the formula

$$ES^2(t, u] = VarS(t, u] + (ES(t, u])^2. \quad (4.19)$$

The last term was already calculated. To figure out  $VarS(t, u]$  we will use the law of total variance

$$VarS(t, u] = EVar_{\Theta}S(t, u] + VarE_{\Theta}S(t, u]. \tag{4.20}$$

From Lemma 3 we have

$$\begin{aligned} VarE_{\Theta}S(t, u] &= P^2(t, u)Var \int_0^{\infty} y\Theta(dy) = P^2(t, u)\psi''(0) \int_0^{\infty} y^2\alpha(dy) \\ &= \frac{P^2(t, u]}{\lambda^2} \int_0^{\infty} y^2\alpha(dy). \end{aligned} \tag{4.21}$$

We already know that

$$Var_{\Theta}S(t, u] = P(t, u] \int_0^{\infty} y^2\Theta(dy) \tag{4.22}$$

Applying Lemma 3 to the above integral we get

$$\begin{aligned} EVar_{\Theta}S(t, u] &= P(t, u]E \int_0^{\infty} y^2\Theta(dy) = P(t, u]\psi'(0) \int_0^{\infty} y^2\alpha(dy) \\ &= \frac{P(t, u]}{\lambda} \int_0^{\infty} y^2\alpha(dy). \end{aligned} \tag{4.23}$$

Thus,

$$ES^2(t, u] = \left(1 + \frac{P(t, u]}{\lambda}\right) \frac{P(t, u]}{\lambda} \int_0^{\infty} y^2\alpha(dy) + \left[ \frac{P(t, u]}{\lambda} \int_0^{\infty} y\alpha(dy) \right]^2. \tag{4.24}$$

Considering the posterior process gives

$$\begin{aligned} E(S^2(t, u]|\mathcal{F}_t) &= \left(1 + \frac{P(t, u]}{\lambda + P(0, t]}\right) \frac{P(t, u]}{\lambda + P(0, t]} \int_0^{\infty} y^2(\alpha(dy) + N((0, t], dy)) \\ &\quad + \left[ \frac{P(t, u]}{\lambda + P(0, t]} \int_0^{\infty} y(\alpha(dy) + N((0, t], dy)) \right]^2. \end{aligned} \tag{4.25}$$

Combining equations (4.18), (4.25) and (4.3) we obtain the conclusion. □

### 4.3 Comparison to the square error loss function

It might be of interest to compare the derived predictors with the Bayes predictor under the square error loss function defined by (3.4). Recall that the Bayes predictor under loss function (3.4) is  $E(S|\mathcal{F}_t)$ . In the model considered in this paper we have

$$E(S(t, u)|\mathcal{F}_t) = \frac{P(t, u)}{\lambda + P(0, t]} \int_0^\infty y(\alpha(dy) + N((0, t], dy)) = aI_1 =: d_0. \quad (4.26)$$

It is easy to see that  $d_k$  given in Theorem 1 tends to  $d_0$  when  $k \rightarrow 0$ , and also the best predictor under LINEX loss tends to  $d_0$  when the parameter  $\kappa \rightarrow 0$ .

## 5 Simulation study

A simulation study is provided to illustrate the properties of the Bayes predictors, which cannot be deduced by analyzing its explicit formulas. The Bayes predictors under various loss criteria are compared and their behavior is studied according to length of the future time interval.

### 5.1 Assumptions

Recall that in the model considered, the process  $N$  is assumed to be a doubly stochastic Poisson process with mean measure  $P \times \Theta$ . The measure  $P$  is supposed to be known.

Let us assume, for example, that

$$P(0, t] = ct + \frac{b}{\pi} \left[ 1 - \cos\left(\frac{\pi}{b}t\right) \right] \quad (5.1)$$

where  $c > 1, b > 0$ . The Radon-Nikodym derivative of the above measure with respect to the Lebesgue measure is of the form

$$\frac{dP}{d\mu}(t) = c + \sin\left(\frac{\pi}{b}t\right), \quad (5.2)$$

where  $\mu$  denotes the Lebesgue measure. The function  $dP/d\mu$  is the intensity function of a Poisson process with mean measure  $P$ . The intensity (5.2) is a periodic function with period equal  $2b$ .

If it is convenient to take a month as the basic time unit, then we can put  $b = 6$ . In this case it is easy to see that the function reaches its maximum at the third month of each year. This means that the most events will appear around the third month. Analogously, there will be the fewest events around the ninth month of each year.

Let us remind that conditional on  $\Theta$ ,  $N_1$  is a Poisson process with mean measure  $P(\cdot)\Theta(\mathbb{R}_+)$ . The measure  $\Theta$  is a gamma process with shape measure  $\alpha$  and scale

parameter  $\lambda$ . In our simulations we will assume that the random variables  $Y_i$  are bounded by a value, say,  $y^*$ . This assumption is justified in most practical situations because it is unreasonable to suppose, for example, that damage values, precipitation levels or claim sizes are infinite. This assumption ensures also finiteness of the needed integrals. For the measure  $\alpha$  we will take the standard Lebesgue measure on the interval  $(0, y^*]$ .

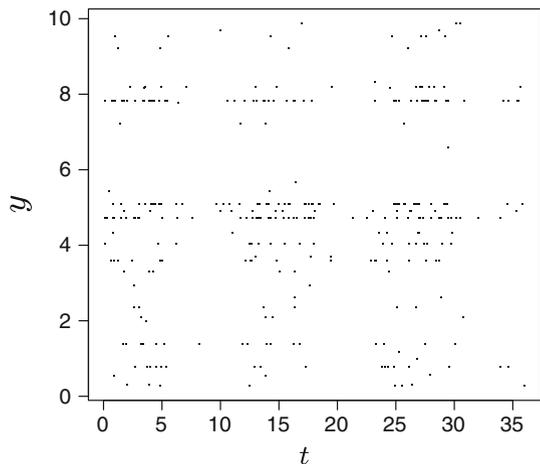
### 5.2 Algorithms

To proceed our simulations sample paths of the process  $N$  on  $[0, T] \times [0, y^*]$  will be needed. They can be obtained by generating independently the event moments  $(T_i)_{i=1}^n$  and their mark sizes  $(Y_i)_{i=1}^n$ . Note that in general the random distribution  $\Theta(\cdot)/\Theta(\mathbb{R}_+)$  of the marks is a Dirichlet process with base distribution  $\alpha(\cdot)/\alpha(\mathbb{R}_+)$  and concentration parameter  $\alpha(\mathbb{R}_+)$ . This means  $Y_1, \dots, Y_n \sim iid \Theta(\cdot)/\Theta(\mathbb{R}_+)$  can be generated via the Blackwell-MacQueen urn scheme (see [Blackwell and MacQueen 1973](#)). The algorithm for generating  $N$  is the following:

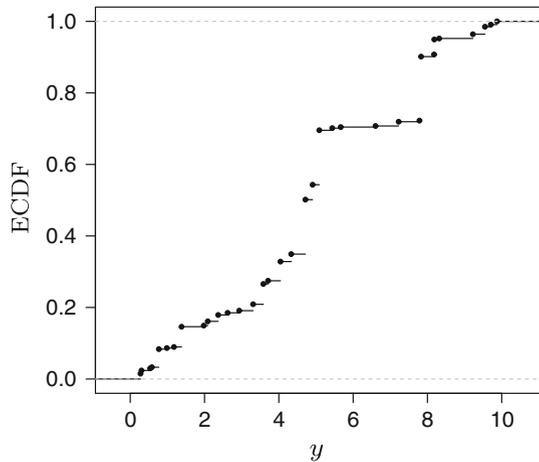
1. Generate  $\Theta(0, y^*] \sim Gamma(\alpha(0, y^*], \lambda)$ ;
2. generate the number of events  $n \sim \mathcal{P}(P(0, T] \Theta(0, y^*])$ ;
3. generate  $X_1, X_2, \dots, X_n$  according to distribution  $P(\cdot)/P(0, T]$ ;
4. set  $T_i = X_{i:n}$ , for  $i = 1, 2, \dots, n$ ;
5. generate  $Y_1, Y_2, \dots, Y_n$  according to distribution  $\Theta(\cdot)/\Theta(0, y^*]$  using the Blackwell-MacQueen urn scheme.

A sample realization of the process  $N$  is shown on Fig. 5. We observe that the points of the process form clusters in time intervals with high intensity  $dP/d\mu$ . High concentration of the marks around a few values is visible too. This is due to the fact that the probability measure  $\Theta/\Theta(0, y^*]$  is discreet. Figure 6 shows the ECDF of  $(Y_i)_{i=1}^n$ .

**Fig. 5** Sample realization of  $N$  ( $P$  given by (5.1) for  $c = 1.1, b = 6$ ;  $(Y_i)_{i=1}^n$  like in Fig. 6)



**Fig. 6** ECDF of  $(Y_i)_{i=1}^n$  ( $y^* = 10, \lambda = 1, \alpha$ —Lebesgue measure)



### 5.3 Bayes predictors examination

The main goal of the simulation study is to compare the predictors of  $S(t, u]$  under the following expected loss criteria

$$K_j(d_k) = E\mathcal{L}_j(S, d_k) = E \frac{(S - d_k)^2}{d_k^j}, \quad j = 0, 1, 2; k = 0, 0.5, 1, 1.5, 2,$$

$$K_3(d_k) = P(d_k < S), \quad k = 0, 0.5, 1, 1.5, 2,$$

where  $d_k$  is the Bayes predictor under the loss  $\mathcal{L}_k$ . The indexes  $k = 0.5, 1, 1.5, 2$  correspond to the precautionary loss function and  $k = 0$  corresponds to the quadratic loss.

The process  $N$  is generated on  $(0, u] \times (0, y^*]$ , where  $u = 30$  and  $y^* = 2$ . The measures  $P$  (with  $b = 6, c = 1.1$ ) and  $\Theta$  are as described in Sect. 5.1. The scale parameter  $\lambda$  of the process  $\Theta$  is set to 1, and the upper bound  $y^*$  of its support is set to 2. With  $\alpha$  being the Lebesgue measure, the value  $y^*$  has an essential impact on the number of events (note that  $E\Theta(0, y^*] = y^*/\lambda$ ). Assuming the history up to time  $t = 24$  is known, predictors of  $S(t, t + h]$  will be constructed, for  $h = 1, 2, \dots, 6$ .

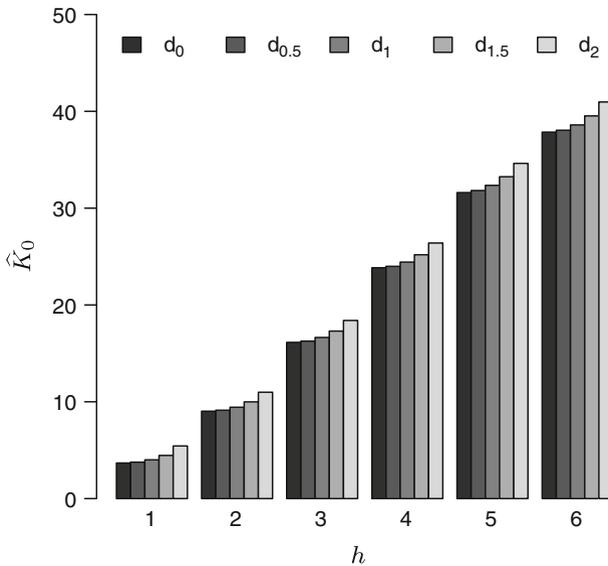
Average values

$$\widehat{K}_j(d_k) := E\widehat{\mathcal{L}}_j(S, d_k) = \frac{1}{m} \sum_{i=1}^m \mathcal{L}_j(S_i, d_{k,i}),$$

$j = 0, 1, 2$ , and

$$\widehat{K}_3(d_k) := \frac{1}{m} \sum_{i=1}^m \mathbf{1}(d_{k,i} < S_i)$$

(the occurrence rate of events  $d < S(t, t + h]$ ) of the four loss criteria are evaluated via Monte-Carlo method for  $m = 10,000$  repetitions.



**Fig. 7** Estimates of the criterion  $K_0$  for the best predictors of  $S(t, t+h]$  under the *square* error loss function and the precautionary loss function for parameters  $k = 0.5, 1, 1.5, 2$

It is obvious that for  $k = 0$  the average loss is minimized by the best predictor under the quadratic loss function. However, we would like to see how are the other predictors doing compared to the best one. Analogously we will look at the loss criteria  $K_1$  and  $K_2$  which are minimized by the predictors under the precautionary loss function for  $k = 1$  and  $k = 2$ , respectively. The loss criterion  $K_3$  is the probability of underestimating. One of the main reasons for using the precautionary loss function is to prevent underestimation. The simulation study shows that

$$\widehat{K}_j(d_j) = \min_k \{\widehat{K}_j(d_k)\}, \quad j = 0, 1, 2,$$

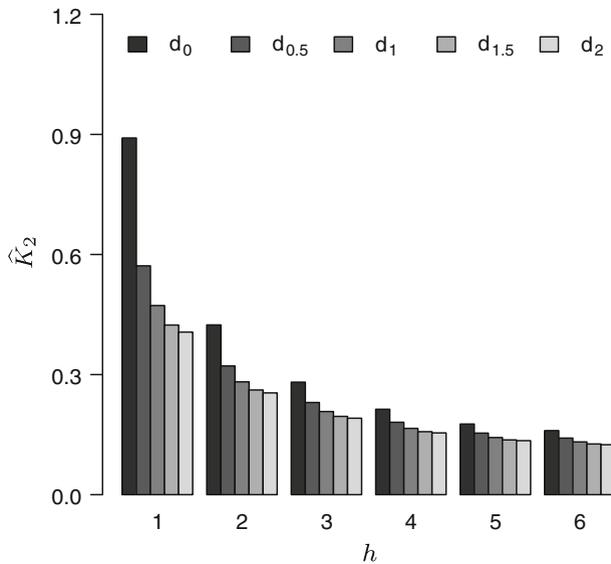
what confirms the theoretical properties of the predictors considered.

Let us take into account the best predictor of  $S(t, u]$  under the LINEX loss function (for  $K = 1$ ), which was presented by Niemiro (2006) The predictor is of the form:

$$d_\kappa^L = -\frac{1}{\kappa} \int_0^{y^*} \log \left( 1 - \frac{P(t, u](e^{\kappa y} - 1)}{\lambda + P(0, t]} \right) (\alpha(dy) + N((0, t], dy)), \quad (5.3)$$

provided that

$$y^* < \frac{1}{\kappa} \log \left( 1 + \frac{\lambda + P(0, t]}{P(t, u]} \right). \quad (5.4)$$



**Fig. 8** Estimates of the criterion  $K_2$  for the best predictors of  $S(t, t + h]$  under the *square* error loss function and the precautionary loss function for parameters  $k = 0.5, 1, 1.5, 2$

**Table 1** Estimates of the criterion  $K_0$  for the best predictors of  $S(t, t + h]$  under the square error loss function, the precautionary loss function for parameter  $k = 2$  and the LINEX loss function for the parameters  $\kappa = 0.01, 0.1, 0.2$

$h$	$d_0$	$d_2$	$d_{0.01}^L$	$d_{0.1}^L$	$d_{0.2}^L$
1	3.68	5.44	3.68	3.74	3.98
2	9.03	10.98	9.03	9.42	10.90
3	16.14	18.40	16.15	17.44	22.30
4	23.84	26.40	23.87	26.77	37.93
5	31.61	34.62	31.67	36.99	57.33
6	37.85	40.97	37.92	45.32	74.83

We compare the predictors  $d_\kappa^L$  for different parameter values with the predictors  $d_0$  and  $d_2$  in terms of the criteria  $K_0$  and  $K_3$ , because of the closeness of  $d_\kappa^L$  to  $d_0$  for small values of  $\kappa$  and of course because of its underestimation preventing properties.

### 5.4 Concluding remarks from the simulation study

The predictors  $d_0, d_1$  and  $d_2$  minimize the average losses  $\widehat{K}_0, \widehat{K}_1, \widehat{K}_2$  respectively, which can be seen in Figs. 7 and 8 (we omit the graph for  $\widehat{K}_1$ ). The underestimation rate of predictors  $d_k$  is the lower, the higher the parameter  $k$  is (Table 2). Of course, lower underestimation rates causes higher mean square errors. That is especially good to see for the predictors due to the LINEX loss function (Table 1). It is interesting that for high values of  $\kappa$  the underestimation rate decreases as the length  $h$  of the prediction horizon increases (Table 2). The predictors due to the precautionary loss function behave different. Their underestimation rates are bigger for bigger  $h$ . Note

**Table 2** Estimates of the criterion  $K_3$  for the best predictors of  $S(t, t + h]$  under the square error loss function, the precautionary loss function for parameters  $k = 1, 2$  and the LINEX loss function for the parameters  $\kappa = 0.01, 0.2$

$h$	$d_0$	$d_1$	$d_2$	$d_{0.01}^L$	$d_{0.2}^L$
1	0.43	0.32	0.20	0.42	0.36
2	0.45	0.36	0.27	0.45	0.35
3	0.46	0.38	0.30	0.45	0.31
4	0.47	0.39	0.32	0.46	0.28
5	0.46	0.39	0.33	0.45	0.26
6	0.47	0.40	0.33	0.45	0.24

that the predictor  $d_2$  achieves very low underestimation rates for  $h = 1$ , much lower than  $d_\kappa^L$  for the parameter values considered. In some situations it might be difficult to decrease the underestimation rates for  $d_\kappa^L$  for small  $h$ , because of the relation (5.4) between  $y^*$  and  $\kappa$ .

**Acknowledgments** The authors wish to thank the referee for valuable comments.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

## References

- Andersen PK, Borgan Ø, Gill RD, Keiding N (1993) Statistical models based on counting processes. Springer series in statistics. Springer, New York
- Blackwell D, MacQueen JB (1973) Ferguson distributions via Pólya urn schemes. *Ann. Stat.* 1:353–355
- Cox DR (2006) Principles of statistical inference. Cambridge University Press, Cambridge. doi:[10.1017/CBO9780511813559](https://doi.org/10.1017/CBO9780511813559)
- Grandell J (1975) Doubly stochastic poisson processes. University of Stockholm, Institute of Actuarial Mathematics and Mathematical Statistics
- Kallenberg O (1986) Random measures, 4th edn. Akademie-Verlag, Berlin
- Kingman JFC (1964) On doubly stochastic Poisson processes. *Math. Proc. Camb. Philos. Soc.* 60:923–930. doi:[10.1017/S030500410003838X](https://doi.org/10.1017/S030500410003838X)
- Kingman JFC (1993) Poisson processes. Oxford studies in probability, vol. 3. The Clarendon Press Oxford University Press, New York. Oxford Science Publications
- Kingman JFC (2006) Poisson processes revisited. *Probab. Math. Stat.* 26(1):77–95
- Last G, Brandt A (1995) Marked point processes on the real line. Probability and its applications (New York) Springer, New York. The dynamic approach
- Mikosch T (2009) Non-life insurance mathematics, 2nd edn. Universitext, Springer, Berlin. doi:[10.1007/978-3-540-88233-6](https://doi.org/10.1007/978-3-540-88233-6) An introduction with the Poisson process
- Møller J, Waagepetersen RP (2002) Statistical inference for Cox processes. Spatial cluster modelling. Chapman & Hall/CRC, Boca Raton, pp 37–60
- Niemiro W (2006) Bayesian prediction with an asymmetric criterion in a nonparametric model of insurance risk. *Statistics* 40(4):353–363
- Norstrøm JG (1996) The use of precautionary loss function in risk analysis. *IEEE Trans Reliab* 45(3):400–403