*Annals of Actuarial Science, Vol. 13, part 1, pp. 67–79.* © Institute and Faculty of Actuaries 2018
doi:10.1017/S1748499518000039
First published online 7 February 2018

# Real-time Bayesian non-parametric prediction of solvency risk

Liang Hong*
*Department of Mathematics, Robert Morris University, 6001 University Boulevard, Moon Township, PA 15108, USA*

Ryan Martin
*Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC 27695, USA*

## Abstract

Insurance regulation often dictates that insurers monitor their solvency risk in real time and take appropriate actions whenever the risk exceeds their tolerance level. Bayesian methods are appealing for prediction problems thanks to their ability to naturally incorporate both sample variability and parameter uncertainty into a predictive distribution. However, handling data arriving in real time requires a flexible non-parametric model, and the Monte Carlo methods necessary to evaluate the predictive distribution in such cases are not recursive and can be too expensive to rerun each time new data arrives. In this paper, we apply a recently developed alternative perspective on Bayesian prediction based on copulas. This approach facilitates recursive Bayesian prediction without computing a posterior, allowing insurers to perform real-time updating of risk measures to assess solvency risk, and providing them with a tool for carrying out dynamic risk management strategies in today's "big data" era.

## Keywords

Density estimation; Mixture model; Non-parametric Bayes; Risk management; Value-at-risk

## 1. Introduction

Solvency risk is a critical concern for both insurers and insurance regulators. Indeed, insurance regulations often require that insurers monitor their solvency risk continuously and take into consideration of the changes of their risk profiles. For example, the European insurance regulation Solvency II (2009, e.g. Article 63) requires that all insurers must calculate their risk capital at least once a year and monitor it on a continuous basis. Similar regulation can be found in the US regulation Own Risk and Solvency Assessment (2017). From the insurers' point of view, they should go beyond the minimum requirement set by the regulators. That is, they should have clear picture of their financial health at all times, not just at the end of each year. Therefore, it is desirable that they know their solvency risk in real time so that they can respond in case their solvency risk goes above their tolerance level. Since these questions about risk naturally involve future losses, this boils down to a prediction problem, and any sort of risk summary would be best described as a suitable feature of the *predictive distribution* for these future losses.

*Correspondence to: Liang Hong, Department of Mathematics, Robert Morris University, Moon Township, PA 15108-2574, USA. Tel: +412 397 4024. E-mail: hong@rmu.edu

A Bayesian approach is particularly attractive in the context of prediction, since the specification of a joint distribution for all unknowns – including the model parameters and future losses – gives the actuary the opportunity to evaluate a genuine predictive distribution, which is simply the conditional distribution of the future losses, given the observed losses, marginalised over model parameters. A key feature of this Bayesian-style predictive distribution is that the marginalisation actually accounts for parameter uncertainty whereas other non-Bayesian methods predict future losses by simply plugging in estimates of model parameters, effectively ignoring the associated uncertainty, often resulting in an underestimation of solvency risk. Examples of Bayesian methods for evaluating risk measures, such as value-at-risk (VaR) and conditional tail expectation (CTE), to monitoring solvency risk include Gerrard & Tsanakas (2011), Fröhlich & Weng (2015), Hong & Martin (2017*a*), and references therein. The trade-off for accurate uncertainty assessment in prediction is that a sufficiently flexible Bayesian model will typically require posterior computations that cannot be carried out in closed form. There are powerful Markov chain Monte Carlo (MCMC) methods now available (e.g. Scollnik, 2001) to handle these computations, but these methods generally cannot take advantage of the natural Bayesian updating procedure, that is, where the old posterior becomes the new prior. Therefore, formal Bayesian updating the posterior and/or the predictive distribution when a new loss is observed cannot be done recursively or online; instead, it requires reprocessing the entire data set again, which can be computationally prohibitive depending on the rate at which new loss variables are observed and/or how quickly the updated predictions are needed.

So the question is how to maintain the desirable prediction features of a flexible Bayesian model but do so in a recursive way that allows insurers to update the corresponding predictive distribution in real time. Here, in the spirit of Makov (2001), we provide an answer to this question by introducing to actuarial science a new Bayesian-inspired non-parametric estimator, first developed in Hahn *et al.* (2017), that is fast and easy to compute and provides recursive updating of the predictive distribution directly, without requiring any posterior computations via MCMC. Given a recursive estimator of the predictive distribution, it is straightforward to evaluate risk measures such as VaR or CTE, which are just functionals of the predictive distribution, to gauge the insurer's solvency risk. Throughout this paper, we will focus on the risk capital required by Solvency II, that is, VaR at the confidence level 99.5%, but see section 5 for discussion of various extensions.

The remainder of this paper is organised as follows. Section 2 sets up the Bayesian prediction problem and describes a sufficiently flexible non-parametric – or infinite-dimensional parametric – model that is capable of adapting to a variety of distributional forms. The flexibility of a non-parametric model is a necessity in real-time prediction applications since the insurer will not be able to carry out a traditional exploratory analyses to identify a satisfactory parametric model. In section 3, we describe an alternative view of the Bayesian predictive updating through an interesting connection with bivariate copulas, which facilitates direct updating of the predictive distribution without any posterior computations or MCMC. Deriving closed-form expressions for the copula corresponding to the non-parametric models we have in mind is out of reach, but a recursive approximation is available, which is easy to compute and has desirable convergence properties. Two numerical examples are given in section 4 to highlight the quality of fit as well as the online tracking of solvency risk that gives insurers the ability to make adjustments in real time. Some concluding remarks are given in section 5. R code to implement the proposed method and examples is available at the second author's website: `http://www4.stat.ncsu.edu/~rmartin`.



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

## 2. Bayesian Prediction

### 2.1. General setup

While many methods in the frequentist framework are available (e.g. Klugman *et al.*, 2012, chapters 11, 13, 20; Frees *et al.*, 2014, chapters 5, 6, 8), the Bayesian approach is natural in this context, as it provides a genuine predictive distribution for future claims, allowing actuaries to assess prediction uncertainty. This is crucial from the risk management point of view, as described in Gerrard & Tsanakas (2011) and Fröhlich & Weng (2015).

To set the scene, since insurance losses are non-negative and typically long right-tailed, throughout this paper we will work with the natural logarithm of the losses; this removes the non-negativity constraint and provides a more convenient scale on which to visualise the loss distribution. Now, under the Bayesian approach, the actuary assumes a full probability model for all unknowns, that is, a joint probability distribution for the real-valued log-losses $X_1, \ldots, X_n$, the future log-losses $X_{n+1}, X_{n+2}, \ldots$, and any unknown parameters $\theta$. The marginal distribution for $\theta$, denoted by $\Pi$, is called the *prior distribution*, and the conditional distribution of $(X_1, \ldots, X_n)$, given $\theta$, defines a likelihood function, as commonly used in classical statistics. Throughout, we take $(X_1, \ldots, X_n, \ldots)$ to be independent and identically distributed (iid), given $\theta$. This model can be described hierarchically as

$$\theta \sim \pi \quad \text{and} \quad (X_1, \ldots, X_n, \ldots) \mid \theta \overset{\text{iid}}{\sim} P_\theta$$

where $P_\theta$ is a probability measure on (a subset of) $\mathbb{R}$, indexed by $\theta \in \Theta$, assumed to have a density function $p_\theta$ with respect to, say, Lebesgue measure $\mu$.

The full probability model makes inference straightforward, at least conceptually. Indeed, given $(X_1, \ldots, X_n)$, Bayes's formula leads to the *posterior distribution* for $\theta$, that is,

$$\pi_n(A) := \pi(\theta \mid X_1, \ldots, X_n) = \frac{\int_A \prod_{i=1}^n p_\theta(X_i)\, \pi(d\theta)}{\int_\Theta \prod_{i=1}^n p_\theta(X_i)\, \pi(d\theta)} \tag{1}$$

where $A$ is any $\pi$-measurable subset of $\Theta$. Similarly, for predicting a future loss $X_{n+1}$, given $(X_1, \ldots, X_n)$, the *posterior predictive distribution* has a $\mu$-density function given by

$$f_n(x) = \int p_\theta(x) \pi_n(d\theta) \tag{2}$$

which is simply the conditional density of $X_{n+1}$, given $(X_1, \ldots, X_n)$, under the proposed Bayesian model. It will be helpful in what follows to refer to the so-called *prior predictive distribution* which is just like (2) but with no data, that is,

$$f_0(x) = \int p_\theta(x) \pi(d\theta) \tag{3}$$

Although $f_n$ in (2) is the usual Bayesian density estimator based on $(X_1, \ldots, X_n)$, note that it is generally very different from a classical plug-in estimator $p_{\hat{\theta}}$, where $\hat{\theta}$ is some estimator of $\theta$ based on $(X_1, \ldots, X_n)$. In particular, $f_n$ need not be a member of the family $\{P_\theta : \theta \in \Theta\}$ of specified distributions. Moreover, integrating over $\theta$ with respect to the posterior $\pi_n$ implies that uncertainty about $\theta$, given $(X_1, \ldots, X_n)$, which is encoded in $\pi_n$, is accounted for in density estimation. This is what distinguishes $f_n$ from a classical density estimator.



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

From the predictive density $f_n$, actuaries may obtain any feature of the predictive distribution for $X_{n+1}$, such as spread, skewness, quantiles or CTEs. In particular, actuaries can calculate the risk capital level set by Solvency II, that is, the 99.5% VaR, which is simply the predictive distribution quantile

$$\upsilon_n := F_n^{-1}(0.995) \tag{4}$$

where $F_n$ is the cumulative distribution function corresponding to $f_n$ in (2). In what follows, our goal will be to track the sequence $\{\upsilon_n : n \geq 1\}$ as new data arrive to give the insurer an assessment of their capital risk in real time.

The Bayesian framework admits a type of recursive updating in the sense that the new prior is the old posterior. That is, one can readily check that the posterior $\pi_n$ can be re-expressed as

$$\pi_n(A) = \frac{\int_A p_\theta(X_n)\pi_{n-1}(d\theta)}{f_{n-1}(X_n)} \tag{5}$$

In the case where the prior $\pi$ admits a density $g$ on $\Theta$, this update can be written in terms of the posterior density, that is,

$$g_n(\theta) \propto p_\theta(X_n) g_{n-1}(\theta)$$

revealing the recursive nature of the updates. However, in cases where it is not possible to work directly with the posterior density, as we discuss below, this natural Bayesian updating formula is out of reach. For online recursive updates, especially in complex models, a suitable approximation may be needed.

## 2.2. Dirichlet process mixture models

Hong & Martin (2017*b*) argue that the insurer might seek the flexibility of a non-parametric model, largely to avoid potential model misspecification which can erroneously influence capital risk assessments. Here, in our context of estimating the predictive distribution in real time, being able to work with a non-parametric approach is especially important. Indeed, it is not possible to look at the *entire* data set all at once, make a decision about what model is appropriate, and then fit that model. It is necessary to start with a sufficiently flexible non-parametric model that can adapt to the shape of the distribution as they arrive.

In these non-parametric cases, $\theta$ is not a finite-dimensional parameter, it is an infinite-dimensional index of the density function $p_\theta$. While there are a number of ways this can be accomplished (see, e.g. Müller & Quintana, 2004) arguably the most common strategy, in the present context of modelling densities, is the so-called *Dirichlet process mixture model*. In this case, $\theta$ itself corresponds to a distribution over defined on a specified latent variable space, $\mathbb{U}$, possibly different from the sample space, and $p_\theta$ is the mixture

$$p_\theta(x) = \int_{\mathbb{U}} k(x \mid u)\, \theta(du)$$

where $k(x|u)$ is a known kernel – a density function in $x$ for each fixed $u \in \mathbb{U}$. Then the model is completed by introducing a Dirichlet process prior (Ferguson, 1973) for the mixing distribution $\theta$, that is, $\theta \sim \pi := \text{DP}(\alpha, G)$, where the precision parameter $\alpha > 0$ controls the variability, and the base measure $G$ on $\mathbb{U}$ characterises the location. See Ghosal (2010) and Hong & Martin (2017*a*) for more details. The most commonly used kernel is a normal, $k(x \mid u) = \text{N}(x \mid \mu, \sigma^2)$, where $u = (\mu, \sigma^2)$ in $\mathbb{U} = \mathbb{R} \times \mathbb{R}_+$.



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

Whatever the choice of kernel and Dirichlet process parameters in the formulation above, the analysis described in section 2.1 carries through word-for-word. That is, given the observed losses $(X_1, \ldots, X_n)$, the actuary can get a posterior $\pi_n$ for the (mixing distribution) $\theta$, and construct a corresponding posterior predictive density $f_n(x)$ for the next loss $X_{n+1}$ via the formula (2). The only difference is that, since $\theta$ is an infinite-dimensional object, computation of the posterior and corresponding predictive necessarily requires MCMC methods; for the normal kernel, Kalli *et al.* (2011) provide one such algorithm.

While these are indeed powerful computational methods, they do not allow the user to take advantage of the natural Bayesian updating in (5), where the new prior is the old posterior, described above. This means that, given the posterior $\pi_{n-1}$ based on $(X_1, \ldots, X_{n-1})$, when new data $X_n$ arrives, the MCMC algorithm must be rerun on the full data $(X_1, \ldots, X_n)$ to get the posterior $\pi_n$ or the predictive density $f_n$. This can be prohibitively slow, thereby motivating a fast recursive approximation.

## 3. A Recursive Approximation

### 3.1. Formulation

To circumvent the aforementioned computational difficulties in Bayesian updating in non-parametric models, we turn to a new strategy, recently proposed by Hahn *et al.* (2017), for updating the predictive (2), via *copulas* (Nelson, 2006). For observations $(X_1, \ldots, X_{n-1}, X_n)$, (5) and the definition (2) of the predictive $f_n$ imply that

$$\begin{aligned} f_n(x) f_{n-1}(X_n) &= f_{n-1}(X_n) \int p_\theta(x) \pi_n(d\theta) \\ &= f_{n-1}(X_n) \int p_\theta(x) \frac{p_\theta(X_n) \pi_{n-1}(d\theta)}{f_{n-1}(X_n)} \\ &= \int p_\theta(x) p_\theta(X_n)\, \pi_{n-1}(d\theta), \quad n \geq 1 \end{aligned} \tag{6}$$

where $f_0(x)$ is the prior predictive density (3). This defines a symmetric joint density in the arguments $(x, X_n)$ and the marginals are both $f_{n-1}$. Then Sklar (1959) theorem implies that there exists a symmetric copula density $c_n$ such that

$$f_n(x) = c_n(F_{n-1}(x),\, F_{n-1}(X_n)) f_{n-1}(x) \tag{7}$$

where $F_{n-1}$ is the distribution function corresponding the predictive density $f_{n-1}$. That is, for each Bayesian model there exists a unique sequence $\{c_n\}$ of copula densities, depending on the sample only through the size $n$, and can be found, at least in principle, by analysing the ratio $f_n(x)/f_{n-1}(x)$, such that (7) holds. This representation reveals that it is indeed possible to directly and recursively update the predictive distribution, thereby circumventing the need for MCMC methods that tend to be slow. In parametric models, it is possible to derive a closed-form expression for the copula density $c_n$. Below we give one example to illustrate this. For more examples, see Hahn *et al.* (2017).

Example 1. Suppose that the claim amount $X$ exceeding a given threshold $x_0$ follows the single-parameter Pareto distribution with parameter $\theta$, that is

$$p_\theta(x) = \theta x_0^\theta x^{-(1+\theta)}, \quad x > x_0$$



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

This model is taken from Rytgaard (1990) and is widely used in reinsurance industry to model the exceedance of claims over a given limit (Bühlmann & Gisler, 2005). Assume $\theta \sim \text{Gamma}(\alpha, \lambda)$, that is, the prior density is given by

$$\pi(\theta) = \frac{\lambda^\alpha}{\Gamma(\alpha)} \theta^{\alpha-1} \mathrm{e}^{-\lambda\theta}, \quad \theta > 0$$

where $\Gamma(x)$ is the gamma function. Then

$$f_n(x) = \frac{\alpha+n}{x} \frac{\left[\lambda + \log\left(M_n / x_0^n\right)\right]^{\alpha+n}}{\left[\lambda + \log\left(M_n / x_0^n\right) + \log(x/x_0)\right]^{\alpha+n+1}}, \quad x > x_0$$

and

$$F_n(x) = 1 - \left[\frac{\lambda + \log\left(M_n / x_0^n\right)}{\lambda + \log\left(M_n / x_0^n\right)\log(x/x_0)}\right]^{\alpha+n}, \quad x > x_0$$

where $M_n = \prod_{i=1}^n X_i$. It follows that

$$\frac{f_n(x)}{f_{n-1}(x)} = \frac{\alpha+n}{\alpha+n-1} \frac{\{[1-F_{n-1}(x)][1-F_{n-1}(X_n)]\}^{-(\alpha+n)/(\alpha+n-1)}}{\left\{[1-F_{n-1}(x)]^{-1/(\alpha+n-1)} + [1-F_{n-1}(X_n)]^{-1/(\alpha+n-1)} - 1\right\}^{\alpha+n+1}}$$

Therefore, we have the Clayton copula density with parameter $1/(\alpha+n-1)$:

$$c_n(u,v) = \frac{\alpha+n}{\alpha+n-1} \frac{[(1-u)(1-v)]^{-(\alpha+n)/(\alpha+n-1)}}{[(1-u)^{-1/(\alpha+n-1)} + (1-v)^{-1/(\alpha+n-1)} - 1]^{\alpha+n+1}}$$

That is, the genuine Bayesian predictive distribution under this Pareto–gamma model can be updated recursively using formula (7) with the copula density $c_n$ above, avoiding any direct posterior distribution computations.

Whether a formula for $c_n$ is available in closed form or not, the representation (7) states that there is a direct and recursive update of the predictive that does not require computing the posterior via MCMC. This representation even holds for complex non-parametric models like the Dirichlet process mixtures described above, although deriving a closed form for the copula $c_n$ is out of reach. Nevertheless, approximations are available.

For a Dirichlet process mixture model, with kernel $k(x|u) = N(x|u, 1)$, a prior $DP(\alpha, G)$ for the mixing distribution $\theta$ on $\mathbb{U} = \mathbb{R}$, with $G = N(0, \tau^{-1})$, and $(\alpha, \tau)$ fixed, if $n = 1$, then Hahn *et al.* (2017) show that the update from $(f_0, X_1)$ to $f_1$ is given by

$$f_1(x) = (1-\alpha) f_0(x) + \alpha f_0(x) c_\rho(F_0(x), F_0(X_1)) \tag{8}$$

where $\rho = \tau^{-1}$ and $c_\rho$ is the bivariate Gaussian copula density

$$c_\rho(u,v) = \frac{N_2(\Phi^{-1}(u), \Phi^{-1}(v) \mid 0, 1, \rho)}{N(\Phi^{-1}(u) \mid 0, 1) N(\Phi^{-1}(v) \mid 0, 1)}$$

with $N_2(\cdot, \cdot | 0, 1, \rho)$ the standard bivariate normal density with correlation $\rho$, and $\Phi$ the standard normal distribution function. The updates for general $n > 1$ are analytically intractable, but Hahn *et al.* (2017)



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

follow the idea in Newton (2002) to create an algorithm by simply applying the one-step predictive update (8) at every iteration. In particular, we consider the following recursive sequence $(\hat{f}_n)$ of predictive densities

$$\hat{f}_n(x) = (1-\alpha_n)\hat{f}_{n-1}(x) + \alpha_n \hat{f}_{n-1}(x) c_\rho(\hat{F}_{n-1}(x), \hat{F}_{n-1}(X_n)), \quad n \geq 1 \tag{9}$$

where $\hat{f}_0$ is a user-specified prior predictive, not necessarily of the form (3), $\hat{F}_{n-1}$ is the distribution function corresponding to $\hat{f}_{n-1}$, $\rho \in (0,1)$ is a fixed correlation, and the weights $(\alpha_n) \subset (0, 1)$ are vanishing at a suitable rate; see (12). This algorithm is similar to the *predictive recursion* method developed by Newton & Zhang (1999) and Newton (2002), and analysed in Martin & Ghosh (2008), Tokdar *et al.* (2009), and Martin & Tokdar (2009, 2011), except that it has the advantage of directly estimating the predictive density and does not require numerical integration to compute normalising constants.

On the distribution function scale, the algorithm is a bit more transparent, that is,

$$\hat{F}_n(x) = (1-\alpha_n)\hat{F}_{n-1}(x) + \alpha_n C_\rho(\hat{F}_{n-1}(x), \hat{F}_{n-1}(X_n)) \tag{10}$$

where $C_\rho$ is given by

$$C_\rho(u, v) = \Phi\left(\frac{\Phi^{-1}(u) - \rho\Phi^{-1}(v)}{(1-\rho^2)^{1/2}}\right) \tag{11}$$

which is a distribution function in $u$ for fixed $v$. From the expression in (11), it is clear that $(1-\rho^2)^{1/2}$ plays a role similar to that of a bandwidth parameter in kernel-based density estimation, so it makes sense for $\rho$ to be relatively close to 1.

For comparison, the aforementioned kernel density estimators, as described in, for example, Sheather (2004), take the form

$$\tilde{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n} k_h(x - X_i) = \frac{n-1}{n}\tilde{f}_{n-1}(x) + \frac{k_h(x - X_n)}{n}$$

where the kernel $k_h$ is often a normal density function with mean 0 and scale $h > 0$ as a bandwidth parameter. From the latter expression, it appears that $\tilde{f}_n$ is a recursive estimator. However, it is common to let the bandwidth $h = h_n$ depend on $n$ and, for such a choice, the apparent recursive structure breaks down. More importantly, this $\tilde{f}_n$ is effectively just a plug-in estimator that may not properly account for uncertainty in prediction, which is undesirable as argued by Gerrard & Tsanakas (2011) and Fröhlich & Weng (2015).

The take-away message is that there exists a recursive update of the predictive density $f_n$ in the Dirichlet process mixture model formulation, characterised by a copula density, and even though the particular copula density is not available in closed form for all $n$, there is a reasonable approximation $\hat{f}_n$ in (9) or (10). Moreover, the approximation $\hat{f}_n$ retains the desirable flexibility and Bayesian features of the true predictive $f_n$.

## 3.2. Properties

Here, we provide some details about the convergence properties of the sequence $\hat{F}_n$ of predictive distributions as $n \to \infty$ under iid sampling of $X_i$'s from a distribution $F^*$ with density $f^*$. Hahn *et al.* (2017) proved that, for a class of weights $(\alpha_n)$ satisfying

$$\sum_{i=1}^{\infty} \alpha_i = \infty \quad \text{and} \quad \sum_{i=1}^{\infty} \alpha_i^2 < \infty \tag{12}$$



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

if $f^*$, the initial guess $\hat{F}_0$, and the correlation $\rho$ are compatible in the sense that

$$\int \min\left\{\hat{F}_0(x),\, 1-\hat{F}_0(x)\right\}^{-2\rho/(1+\rho)} f^*(x)\,dx < \infty \tag{13}$$

then $\hat{F}_n \to F^*$ in the Kullback–Leibler divergence and, hence, the total variation distance sense, with $F^*$-probability 1, as $n \to \infty$. The condition (12) implies that the weights are vanishing, which is necessary for convergence of the algorithm, but that they are not vanishing too fast that $F_n$ is unable to forget the initial guess $\hat{F}_0$ and adapt to the shape of the true distribution $F^*$. The integrability condition (13), in particular, requires that the support of $\hat{F}_0$ contains that of $F^*$, which is clearly necessary for consistency, since the support of $\hat{F}_n$ is contained in the support of $F_0$ for all $n$. More than that, the integrability condition (13) can fail only if $\hat{F}_0$ has too light of tails compared to $F^*$, which suggests that the insurer's choice of $\hat{F}_0$ be sufficiently heavy-tailed.

We conclude here by saying that, if the insurer is interested in estimating some nice functional $\psi(F^*)$ of the true distribution $F^*$, then consistency of the plug-in estimator $\psi\left(\hat{F}_n\right)$ follows from the theory described above. In particular, the VaR is such a functional so, if the conditions for $\hat{F}_n \to F^*$ are met, then we can expect that $\hat{v}_n = \hat{F}_n^{-1}(0.995)$ as in (4) will converge to $F^{*-1}(0.995)$, the corresponding quantile of the true distribution $F^*$.

## 3.3. Implementation

### 3.3.1. Algorithm

For log-losses $X_1, X_2, \ldots$, the following summarises the recursive algorithm. Software to implement this algorithm is available at `http://www4.stat.ncsu.edu/~rmartin`.

1. Make an initial guess of $\hat{F}_0$ with density $\hat{f}_0$ whose support is the whole real line.
2. Fix a grid of points, $\{\overline{x}_m : m = 1, \ldots, M\}$, in $\mathbb{R}$, covering roughly the entire support of $\hat{f}_0$, where $M$ is a positive integer set by the actuary.
3. For each $m$, compute the sequence $\left(\hat{F}_n(\overline{x}_m)\right)$ using ((10)). Since data $X_i$ is surely not to fall exactly on the specified grid, an interpolation procedure, like approxfun in R, can be used to evaluate $\hat{F}_{i-1}(X_i)$.
4. Given the distribution function $\hat{F}_n(x)$, the corresponding density function $\hat{f}_n(x)$ can be readily evaluated using difference ratios, that is,

$$\hat{f}_n(\overline{x}_m) \approx \frac{\hat{F}_n(\overline{x}_m) - \hat{F}_n(\overline{x}_{m-1})}{\overline{x}_m - \overline{x}_{m-1}}$$

   and, again, interpolation can be used to evaluate $\hat{f}_n(x)$ for points $x$ off the grid. Some additional smoothing, for example, smooth spline in R, can also be used to improve the relatively crude difference-ratio estimate above.

### 3.3.2. Algorithm inputs

Here, we make a few remarks concerning implementation of the recursive algorithm (10). To start, the actuary is required to set the correlation $\rho \in (0, 1)$, the weight sequence $(\alpha_n)$, and the initial guess $F_0$. The choice of $\rho$ is entirely up to the discretion of the actuary, with values closer to 1 corresponding to less smoothing; as a general guideline, we find that $\rho = 0.90$ is a reasonable choice. For the weights, condition (12) suggests a choice like $\alpha_i = (i+1)^{-r}$ for $r \in (0.5, 1]$. In our numerical examples, we take $r = 1$ as a default choice. Finally, in choosing the initial guess $F_0$, the essential point



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

is to adequately capture the support of log-loss distribution. Since this distribution is not known and, by the nature of the recursive estimation problem, there is little or no data to use as a guide, incorporating some prior information is essential. Motivated by the integrability condition (13), we recommend taking $\hat{f}_0$ as a Student-$t$ density with location $\mu$ and scale $\sigma$, with $(\mu, \sigma)$ specified by the insurer based, for example, on data from previous years, etc. Taking the degrees of freedom equal to 2, as we do here, guarantees that the CTE is finite, which is reasonable.

## 4. Examples

### 4.1. Danish fire insurance loss data

The complete Danish data on fire insurance losses, hereby abbreviated as the "Danish data", has been studied by several authors; see, for example, Scollnik & Sun (2012), Cooray & Cheng (2015), Calderín-Ojeda & Kwok (2016), and the references therein. The data are comprised of $n = 2{,}492$ fire insurance loss entries from 1980 to 1990. To account for inflation, the data have been adjusted to reflect 1985 values. All losses are expressed in Danish Krone, and about 94% are between 1 and 7 million Krones. (Note that the data set are traditionally stored in ascending order, which does not resemble an iid sequence, so we work here with a random permutation $X_{i_1}, \ldots, X_{i_n}$ of the sorted data $X_1, \ldots, X_n$.) A histogram of the log-losses is shown in Figure 1(a), along with two final estimators $\hat{f}_n$ from the recursive algorithm presented in section 3.1 based on Student and normal initial guesses and other default settings described in section 3.3. Since this data set is relatively large, as the convergence theory in section 3.2 suggests, the two recursive estimators both are able to adapt to the unusual shape of the data histogram, and provide a satisfactory fit.

The recursive algorithm also provides insurers with online updating of the predictive distribution so that real-time updating of the risk capital is possible. To illustrate this, we treat the permuted data as arriving in real time. We also assume that the insurer has no past data on this line of business. Recall that Solvency II requires the insurer to set its risk capital to VaR at the level 99.5%. Figure 1(b) gives a plot of the evolution of VaR along the data sequence for both $t_2$ and $N(0, 4^2)$ initial guesses with the unit for the vertical axis being log-millions; the choice of $\sigma = 4$ in the normal is to roughly match the VaR of $t_2$, so that the comparison essentially only involves their tail thicknesses, not an overall

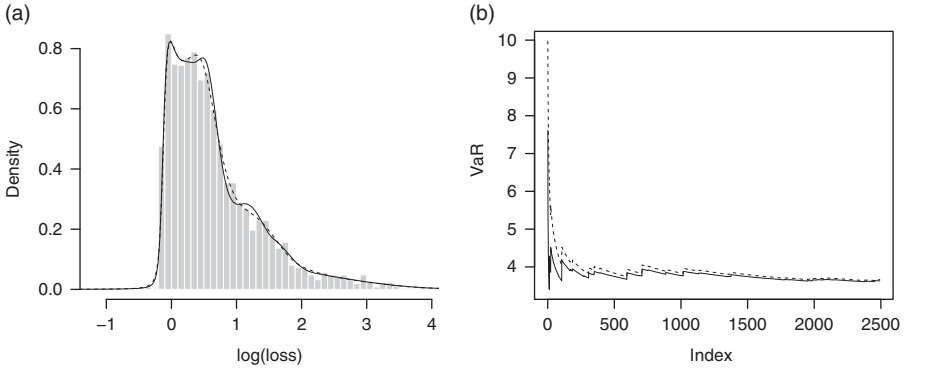


**Figure 1.** Results for the Danish fire insurance data described in section 4.1. (a) Data histogram with the recursive estimators with $t_2$ initial guess (solid) and $N(0, 4^2)$ initial guess (dashed). (b) Evolution of risk capital for the same two initial guesses. VaR, value-at-risk.



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

scale difference. Table 1 also lists the values of VaR and CTE at several chosen observation indices. For both VaR trajectories, the first few iterations have very large VaR as a result of the initial guess, but they stabilise relatively quickly. The insurer may want to adjust its risk capital during the initial, say, 1,000 steps to be prepared for increased solvency risk. Eventually, the asymptotic convergence theory kicks in and the sequence of estimates stabilises, hence the insurer may streamline its capital allocation accordingly. Indeed, the two VaR estimates appear to be merging together towards the end of the data sequence.

## 4.2. 1988 Norwegian fire claims data

Next, we analyse the 1988 Norwegian fire claims data which has been studied by several authors; see, for example, Brazauskas & Kleefeld (2011, 2014, 2016), Nadarajah & Bakar (2015), Scollnik (2014), and references therein. The 1988 Norwegian data consists of $n = 827$ fire loss claims exceeding 500 thousand Norwegian Krones in 1988. Figure 2(a) shows a histogram of the log-loss data, along with the final iteration of the recursive algorithm from section 3, again based on $t_2$ and $N(0, 4^2)$ initial guesses and the default settings, and following a random permutation of the data. In this case, both the two density estimators for different initial guesses are almost indistinguishable.

**Table 1.** Evolution of value-at-risk (VaR) and conditional tail expectation (CTE) values for Danish fire insurance data described in section 4.1.

| | $\hat{F}_0 = t_2$ | | $\hat{F}_0 = N(0, 4^2)$ | |
|---|---|---|---|---|
| Index | VaR | CTE | VaR | CTE |
| 1,000 | 3.769 | 4.530 | 3.837 | 4.647 |
| 1,200 | 3.813 | 4.555 | 3.882 | 4.662 |
| 1,400 | 3.785 | 4.464 | 3.844 | 4.637 |
| 1,600 | 3.709 | 4.493 | 3.758 | 4.592 |
| 1,800 | 3.662 | 4.431 | 3.704 | 4.488 |
| 2,000 | 3.655 | 4.355 | 3.692 | 4.383 |
| 2,200 | 3.649 | 4.279 | 3.684 | 4.439 |
| 2,400 | 3.623 | 4.221 | 3.654 | 4.354 |

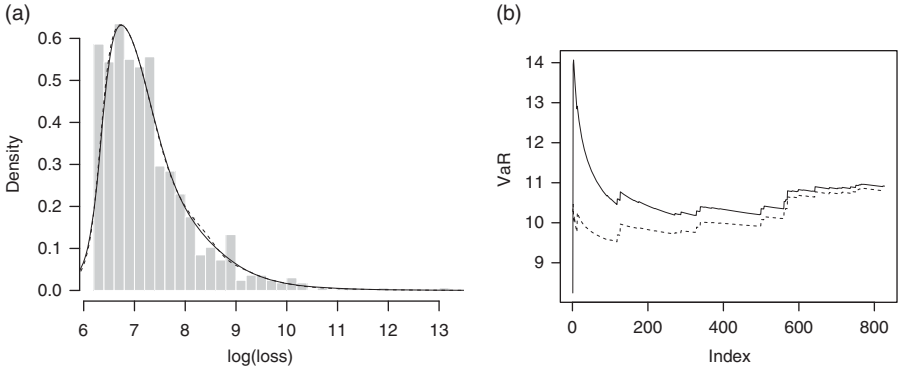


**Figure 2.** Results for the 1988 Norwegian fire claims data described in section 4.2. (a) Data histogram with the recursive estimators with $t_2$ initial guess (solid) and $N(0, 4^2)$ initial guess (dashed). (b) Evolution of risk capital for the same two initial guesses. VaR, value-at-risk.



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

**Table 2.** Evolution of value-at-risk (VaR) and conditional tail expectation (CTE) values for Norwegian fire claims data described in Section 4.2.

| | $\hat{F}_0 = t_2$ | | $\hat{F}_0 = N(0, 4^2)$ | |
|---|---|---|---|---|
| Index | VaR | CTE | VaR | CTE |
| 100 | 10.601 | 11.673 | 9.568 | 9.587 |
| 200 | 10.423 | 11.392 | 9.828 | 10.405 |
| 300 | 10.239 | 11.083 | 9.788 | 10.075 |
| 400 | 10.334 | 11.192 | 9.984 | 10.693 |
| 500 | 10.353 | 11.066 | 10.081 | 10.349 |
| 600 | 10.828 | 11.696 | 10.685 | 11.575 |
| 700 | 10.853 | 11.981 | 10.731 | 11.769 |
| 800 | 10.929 | 11.843 | 10.826 | 11.525 |

But, with smaller $n$ than in the previous example, there is an effect of the truncation at (the log of) 500,000 Krones, something that the recursive estimator is not aware of. If, on the other hand, it were known that "loss < 500" is impossible, then this can be accommodated through the choice of $\hat{f}_0$. In any case, the fit of $\hat{f}_n$ in the right tail is adequate. Figure 2(b) shows the evolution of predictive risk capital as data comes one at a time with vertical axis units log-thousands. Here there are some substantial fluctuations at the early steps, but the two VaR trajectories seem to merge by around 600 steps. Table 2 provides several values of VaR and CTE.

## 5. Concluding Remarks

Sound management of solvency risk requires the insurer to monitor their solvency risk continuously and preferably in real time. While Bayesian analysis has been successful in estimating the risk measures to gauge solvency risk, no existing Bayesian insurance models are able to allow insurers to perform real-time updating of their solvency risk. This is due to the fact implementation of Bayesian models often requires MCMC, which makes real-time updating of predictive distribution infeasible. Motivated by this, our paper introduces to actuarial science a new perspective of Bayesian recursive prediction that allows insurers to recursively update predictive distribution without computing the posterior. This new approach enables insurers to update the predictive distribution recursively in real time. Though we chosen risk capital set by Solvency II as our vehicle for illustration, the same can be done for any other risk measures such as ruin probabilities and CTE. Two real-data examples show how insurers may use this new method to monitor its risk capital and manage its solvency risk.

Throughout we followed Fröhlich & Weng (2015) and considered prediction of a single-future loss; of course, our estimate of the predictive for the log-loss can readily be transformed back to the original scale. But there may be cases where interest is in the predictive distribution of some function $\gamma_N = \gamma(X_{n+1}, \ldots, X_{n+N})$ of $N$ future independent losses. One example of $\gamma_N$ is the sum of the next $N$ future losses and, for this case, there is no conceptual barriers to extending the methodology presented here, since any feature of the distribution of $\gamma_N$ is just a function of the predictive $\hat{f}_n$ for each individual future loss. If $N$ is relatively large, then, by the central limit theorem, the distribution of $\gamma_N$ is approximately normal, and the mean and variance can be derived easily from $\hat{f}_n$. If $N$ is relatively small, then numerical evaluation of the $N$-fold convolution of $\hat{f}_n$ may not be too inconvenient.



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

Though primarily motivated by the need for online prediction of risk measures, the recursive algorithm is also applicable in cases where only batch prediction is needed. In such a situation, the recursive algorithm still has the advantage of being computationally efficient compared to MCMC methods. It is worth mentioning that the estimator $\hat{f}_n$ depends on the order in which the data were processed. In applications involving streaming data, that is, where new data points arrive one at a time, the order-dependence makes sense, but when the data becomes available all at once, the order is arbitrary so order-dependence in $\hat{f}_n$ is difficult to interpret. In the latter case, a (near) order-independent estimator can be obtained by averaging the order-dependent estimators over several randomly chosen permutations of the data sequence. Martin & Tokdar (2012) argue that averaging over roughly 10 random permutations is sufficient to effectively eliminate the order-dependence, especially if $n$ is relatively large. And since each order-dependent $\hat{f}_n$ can be computed very fast, averaging over a few permutations is not a computational burden.

Finally, we believe that the recursive method we introduced in this paper will prove to be more useful for markets with concentration risk such as earthquake insurance, cyber insurance, and terrorism insurance. In such a market, claims may occur on a large scale within just minutes for a company that has written a large number of policies in an affected region, and real-time updating of the predictive distribution becomes critical for managing risks associated with occurred losses even though reported claims are likely to be processed in batches. Therefore, it is crucial that tools for real-time updating of predictive distributions be available for insurers.

### Acknowledgements

The authors thank the editor and two anonymous reviewers for comments and suggestions that led to significant improvements in this article.

## References

Brazauskas, V. & Kleefeld, A. (2011). Folded and log-folded-$t$ distributions as models for insurance loss data. *Scandinavian Actuarial Journal*, **1**, 59–74.

Brazauskas, V. & Kleefeld, A. (2014). Author's reply to "Letter to Editor: Regarding Folded and the Paper by Brazauskas and Kleefeld" by Scollnik. *Scandinavian Actuarial Journal*, **8**, 753–757.

Brazauskas, Y. & Kleefeld, A. (2016). Modeling severity and measuring tail risk of Norwegian fire claims. *North American Actuarial Journal*, **20**(1), 1–16.

Bühlmann, & Gisler (2005). *A Course in Credibility Theory and Its Application*. Springer, New York.

Calderín-Ojeda, E. & Kwok, C.F. (2016). Modeling claims data with composite Stoppa models. *Scandinavian Actuarial Journal*, **9**, 817–836.

Cooray, K. & Cheng, C.I. (2015). Bayesian estimators of the lognormal-Pareto composite distribution. *Scandinavian Actuarial Journal*, **6**, 500–515.

Ferguson, T.S. (1973). Bayesian analysis of some nonparametric problems. *Annals of Statistics*, **1**, 209–230.

Frees, E.W., Derrig, R.A. & Meyers, G. (2014). *Predictive Modeling Applications in Actuarial Science, Vol. I: Predictive Modeling Techniques*. Cambridge University Press, Cambridge.

Fröhlich, A. & Weng, A. (2015). Modeling parameter uncertainty for risk capital calculation. *European Actuarial Journal*, **5**, 79–112.

Gerrard, R. & Tsanakas, A. (2011). Failure probability under parameter uncertainty. *Risk Analysis*, **31**, 727–744.

Ghosal, S. (2010). The Dirichlet process, related priors and posterior asymptotics. In N.L. Hjort, C. Holmes, P. Müller & S.G. Walker (Eds.), *Bayesian Nonparametrics* (pp. 35–79). Cambridge University Press, Cambridge.



https://doi.org/10.1017/S1748499518000039 Published online by Cambridge University Press

Hahn, P.R., Martin, R. & Walker, S.G. (2017). On recursive Bayesian predictive distributions. *Journal of the American Statistical Association*, https://doi.org/10.1080/01621459.2017.1304219.

Hong, L. & Martin, R. (2017a). A flexible Bayesian nonparametric model for predicting future insurance claims. *North American Actuarial Journal*, **21**(2), 228–241.

Hong, L. & Martin, R. (2017b). Dirichlet process mixture models for insurance loss data, *Scandinavian Actuarial Journal*, https://doi.org/10.1080/03461238.2017.1402086.

Kalli, M., Griffin, J.E. & Walker, S.G. (2011). Slice sampling mixture models. *Statistical Computing*, **21**, 93–105.

Klugman, S.A., Panjer, H.H. & Willmot, G.E. (2012). *Loss Models: From Data to Decisions*, 4th edition. Wiley, Hoboken, NJ.

Makov, U.E. (2001). Principal applications of Bayesian methods in actuarial science. *North American Actuarial Journal*, **5**(4), 53–57.

Martin, R. & Ghosh, J.K. (2008). Stochastic approximation and Newton's estimate of a mixing distribution. *Statistical Science*, **23**, 365–382.

Martin, R. & Tokdar, S.T. (2009). Asymptotic properties of predictive recursion: robustness and rate of convergence. *Electronic Journal of Statistics*, **3**, 1455–1472.

Martin, R. & Tokdar, S.T. (2011). Semiparametric inference in mixture models with predictive recursion marginal likelihood. *Biometrika*, **98**, 567–582.

Martin, R. & Tokdar, S.T. (2012). A nonparametric empirical Bayes framework for large-scale multiple testing. *Biostatistics*, **13**, 427–439.

Müller, P. & Quintana, F.A. (2004). Nonparametric Bayesian data analysis. *Statistical Science*, **19**, 95–110.

Nadarajah, S. & Bakar, S.A.A. (2015). New folded models for the log-transformed Norwegian fire claim data. *Communications in Statistics–Theory and Methods*, **44**, 4408–4440.

Own Risk and Solvency Assessment (2017). Available online at the address http://www.naic.org/cipr_topics/topic_own_risk_solvency_assessment.htm [accessed on 8-Aug-2017].

Nelson, R.B. (2006). *An Introduction to Copulas*, 2nd edition. Springer, New York.

Newton, M. (2002). On a nonparametric recursive estimator of the mixing distribution. *Sankhyā: The Indian Journal of Statistics*, **64**, 306–322.

Newton, M. & Zhang, Y. (1999). A recursive algorithm for nonparametric analysis with missing data. *Biometrika*, **86**(1), 15–26.

Rytgaard, M. (1990). Estimation in the Pareto distribution. *ASTIN Bulletin*, **20**, 201–216.

Scollnik, D.P.M. (2001). Actuarial modeling with MCMC and BUGS. *North American Actuarial Journal*, **5**(2), 96–124.

Scollnik, D.P.M. & Sun, C. (2012). Modeling with Weibull–Pareto models. *North American Actuarial Journal*, **16**, 260–272.

Scollnik, D.P.M. (2014). Letter to editor: regarding folded models and the paper by Brazauskas and Kleefeld (2011). *Scandinavian Actuarial Journal*, **2014**(3), 278–281.

Sheather, S.J. (2004). Density estimation. *Statistical Science*, **19**(4), 588–597.

Sklar, M. (1959). Fonctions de répartition á n dimensions et leurs marges. Université Paris, **8**, 229–231.

Solvency II (2009). Available online at the address http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2009:335:0001:0155:en:PDF [accessed 8-Aug-2017].

Tokdar, S.T., Martin, R. & Ghosh, J.K. (2009). Consistency of a recursive estimate of mixing distributions. *Annals of Statistics*, **37**, 2502–2522.