SECOND ORDER BAYES PREDICTION OF FUNCTIONALS OF EXPONENTIAL DISPERSION DISTRIBUTIONS AND AN APPLICATION TO THE PREDICTION OF THE TAILS

BY

ZINOVIY LANDSMAN

ABSTRACT

Second order Bayes estimators, being the main tool in second order optimal statistical theory, provide a natural basis for a new approach to the problem of the prediction of functions of expectation functional for members of an exponential dispersion family. A general formula, providing such prediction up to the term of the order $1/n$, is suggested and the application to the problem of the prediction of the tail of distributions is demonstrated. The results are illustrated with normal and gamma claim sizes. The numerical experiment demonstrates the high effectiveness of the approach even for small sample sizes.

KEYWORDS

Second order Bayes estimator, predictive tail, predictive mean, credibility formula, lower truncated mean, variance functions.

1. INTRODUCTION

Let θ be a risk parameter characterizing a member of a risk collective, and given θ, let P_θ, the distribution of the claim X, be a member of a family of distributions $\{P_\theta, \theta \in \Theta \subset R^1\}$. Furthermore, let $\pi(\theta)$ be the prior distribution of θ, the so-called *structure distribution*. Then $m = \int_\Theta E(X|\theta) d\pi(\theta)$ is called the collective fair premium, which is also known as an "industrial" mean. Let

$$x_1, x_2, ..., x_n,$$

be n observations of individual experience, which conditionally on θ are independently distributed in accordance with P_θ. Then $\bar{x}_n = 1/n \sum_{i=1}^n x_i$ is a consistent estimator of the fair premium $\mu(\theta) = E(X|\theta)$.

At the center of the Bayesian credibility theory is the fact, stated by Jewell (1974), that the classical credibility formula (Bühlmann (1967)

$$(1 - z_n) m + z_n \bar{x}_n, \tag{1}$$

coincides with the predicted mean, i.e.

$$(1 - z_n)\, m + z_n \bar{x}_n = E(x_{n+1}|x_1, x_2, ..., x_n),$$

if the distribution of the claim size is a member of the *Natural Exponential Family* (NEF)

$$dP_\theta = e^{x\theta - k(\theta)}dQ(x), \; x \in S \subset R^1, \; \theta \in \Theta \subset R^1, \tag{2}$$

$\Theta = \{\theta \,|\, k(\theta) = \ln \int e^{\theta x}dQ(x) < \infty\}$, and the conjugate prior density is given by

$$d\pi(\theta) \propto e^{n_0(x_0\theta - k(\theta))}d\theta. \tag{3}$$

More details about the relevance of conjugate prior densities were well investigated in Diaconis and Ylvisaker (1979). Kaas, Dannenburg and Goovaerts (1997), Nelder and Verrall (1997) and Landsman and Makov (1998) showed that this phenomenon holds for the more general *Exponential Dispersion Family* (EDF),

$$dP_{\theta,\lambda} = e^{\lambda(x\theta - k(\theta))}dQ_\lambda(x), \; \theta \in \Theta \subset R^1, \; \lambda \in \Lambda \subset R^+, \tag{4}$$

considered in Tweedie (1984), Nelder and Wedderburn (1972), and Jorgensen (1986, 1987, 1992, 1997); here λ is called dispersion parameter. For this family the Bayes estimator of the mean μ under the conjugate prior (3), given λ, continues to be linear with respect to the sample mean $\bar{x}_n$, as follows:

$$E(x_{n+1}|x_1,...,x_n) = E(\mu(\theta)|x_1,...,x_n) = \frac{n_0}{n_0 + n\lambda}\, m + \frac{n\lambda}{n_0 + n\lambda}\, \bar{x}_n. \tag{5}$$

Although predicting the mean of a claim size is of immense importance, the prediction of other functionals of loss distributions, namely variance, moments, skewness and kurtosis, plays a significant role not only in credibility theory, but also, for example, in predictive premium evaluation. However, exact expressions for predicting such functionals are intractable, even for Bayesian prediction with conjugate priors. One of the most important and difficult characteristic to predict is the tail of a loss distribution, i.e.,

$$P(x_{n+1} > T \,|\, x_1, ..., x_n), \tag{6}$$

the expected propensity of large claims. This is especially significant in a reinsurance context, where dramatically deviated claims are the main concern. However, the exact expression for (6) is quite complicated to evaluate even for claim distributions belonging to EDF and the use of conjugate priors.

    In this paper we evaluate the formula for the prediction of general type functionals of distributions up to the term of order $O_p(\frac{1}{n})$ and as an application we consider the problem of predicting the tail area. This general result can certainly be applied immediately to other functionals, for example, the well-known variance premium. Such applications are beyond the goals of this paper and

will be discussed elsewhere. We use the approach based on the second order (s.o.) Bayes estimators of some function of the mean, i.e., the estimators that coincide with the Bayes estimators up to the term of order $O_P(\frac{1}{n})$. The natural impact of the s.o. Bayes estimators on credibility theory was shown in Landsman (2002). Second order Bayes estimators represent a main tool in the s.o. optimal statistics theory. Levit (1980) and Bickel (1981) used the s.o. Bayes estimators for constructing the s.o. minimax estimator of the mean parameter of normal distribution. These results were generalized by Levit (1982) for the case of an s-dimensional bounded parametric space and for more general distributions. Landsman and Levit (1990, 1991) studied the case where nuisance parameters are present. Landsman (2001) explored the s.o. Bayes estimators of the mean parameter for members of the EDF in the context of s.o. minimax improvement of the sample mean $\bar{x}_n$ (which is the maximum likelihood estimator (MLE) for the EDF) for the unbounded space of all possible values of μ. We extend the approach suggested in Landsman (2001) for the evaluation of the s.o. Bayes estimators of the function of the mean in the case of the EDF.

In section 2 we derive the s.o. Bayes estimators of some general function of μ, and in section 3 we apply this result for predicting the probability that claim x_{n+1} exceeds some large threshold T given previous claims $x_1, ..., x_n$. In section 4 we illustrate the results with Normal and Gamma distributed claims and report the results with a numerical experiment.

Finally, let us recall that the statement for random variable $Y_n = O_P(1/n)$ about random variable Y_n is taken to mean that $n|Y_n|$ is bounded in probability, i.e., for any $\varepsilon > 0$ there exists $M_\varepsilon > 0$ that $P(n|Y_n| > M_\varepsilon) < \varepsilon$; and $Y_n = o_P(1/n)$ is taken to mean that $nY_n \xrightarrow{P} 0$, $n \to \infty$ (see, for example, Fuller (1996), Ch. 5).

2. Second order Bayes estimator of a function of the mean

Let $\psi(\mu)$ be some function defined on the space $M = k'(\Theta)$ of all possible values of the mean μ. It is well known that

$$\psi_n = E\left(\psi(\mu(\theta))\,\middle|\,x_1, x_2, ... x_n\right) = \int_\Theta \psi(\mu(\theta)) f(\theta\,|\,x_1, x_2, ... x_n)\, d\theta \tag{7}$$

is the Bayes estimator of function $\psi(\mu)$, given $x_1, x_2, ..., x_n$. We say estimator ψ_n^* is a s.o. Bayes estimator of $\psi(\mu)$ if it coincides with ψ_n up to the term $O_P(\frac{1}{n})$. The following theorem represents the s.o. Bayes estimator of $\psi(\mu)$.

We assume that the NEF which generates the EDF (see Jorgensen (1997), Sect. 3.1) is regular, i.e., the parameter set Θ is open or at least steep, i.e. (see Brown (1986), Proposition 3.3),

$$\int_S |x|\, dP_{\theta,\lambda} = \infty, \ \text{ for } \theta \in \Theta \backslash int(\Theta)$$

Here $int(\Theta)$ is the set of all intern points of Θ. Then mean function (Brown (1986), Theorem 3.6)

$$\mu = \mu(\theta) = E(X|\theta) = \int_S x\, dP_{\theta,\lambda} = k'(\theta)$$

is one-to-one (X is a random variable distributed according to $P_{\theta,\lambda}$. The variance is

$$Var(X|\theta) = \frac{1}{\lambda} k''(\theta) = V(\mu)\sigma^2, \tag{8}$$

where $V(\mu)$ is called the *variance function*. The predictive mean (5) is denoted by

$$\mu_n = \frac{n_0}{n_0 + n\lambda} m + \frac{n\lambda}{n_0 + n\lambda} \bar{x}_n. \tag{9}$$

Suppose that $\Theta = (a,b)$ is an open finite or infinite interval, i.e., $-\infty \leq a < b \leq \infty$. (In the case of a steep family, instead of θ one should write $int(\Theta)$.)

Theorem 1. *For $x_0 > 0$, $n_0 > 0$ let*

$$\int_{\Theta} k'(\theta)^2 e^{n_0(x_0\theta - k(\theta))} d\theta < \infty,$$

and $\psi(x)$ be a twice continuously differentiable function. Then under the conjugate prior (3)

$$E\left(\psi\left(\mu(\theta)\right)|x_1, x_2, \ldots x_n\right) = \psi(\mu_n) + \frac{1}{2} \frac{V(\mu_n)}{(n_0 + n\lambda)} \psi''(\mu_n) + o_P\left(\frac{1}{n}\right). \tag{10}$$

Proof. As $\psi(x)$ is twice differentiable, using the Taylor expansion around μ_n one can write

$$\psi(\mu) = \psi(\mu_n) + (\mu - \mu_n)\psi'(\mu_n) + \frac{1}{2}(\mu - \mu_n)^2 \psi''(\mu_n + \alpha(\mu - \mu_n)), \tag{11}$$

where $|\alpha| \leq 1$. Under the conjugate prior (3) one can show, straightforwardly, that a posterior density has the form

$$f(\theta|x_1, x_2, \ldots x_n) = \exp\left(n_0'\left(x_0'\theta - k(\theta)\right)\right)\left(\int_{\Theta} \exp\left(n_0'\left(x_0'\theta - k(\theta)\right)d\theta\right)^{-1}, \tag{12}$$

where

$$x_0' = \mu_n, \tag{13}$$
$$n_0' = n_0 + n\lambda. \tag{14}$$

Then for ψ_n we have from (7), (11), (12),

$$\psi_n = \psi(\mu_n) + \frac{1}{\int_{\Theta} \exp\left(n_0'\left(x_0'\theta - k(\theta)\right)d\theta\right)}$$
$$\times \left\{\psi'(\mu_n)\int_{\Theta} (\mu(\theta) - \mu_n)\exp\left(n_0'\left(x_0'\theta - k(\theta)\right)d\theta\right.\tag{15}$$

$$+ \frac{1}{2} \int_{\Theta} (\mu(\theta) - \mu_n)^2 \psi''(\mu_n + \alpha(\mu(\theta) - \mu_n)) \exp\left(n_0'\left(x_0'\theta - k(\theta)\right)d\theta\right\}.$$

Let us note first that for density (12)

$$\int_{\Theta} \mu(\theta) f(\theta | x_1, x_2, \dots, x_n) \, d\theta = x_0' = \mu_n,$$

hence the second integral in the right hand side of (15) equals 0.

Using the Laplace method we evaluate the asymptotics of the right hand side of (15), when $n_0' = (n_0 + n\lambda) \to \infty$. In reference to the Laplace method see, for example, Lindley (1980) and Tierney and Kadane (1986). For a general form of this method, see Barndorff-Nielsen (1989, Sect. 6.2), where in our case $r_{n_0'}(y) = -n_0'(x_0'y - k(y))$. Recall that in our case $r_n(y)$ has only one minimum equal to $\bar{y}_n = \theta(x_0')$ in $\Theta = (a,b)$, where the function $\theta = \theta(\mu)$ is the inverse function of function $\mu = \mu(\theta) = k'(\theta)$. The Laplace method gives

$$\int_{\Theta} \exp(n_0'(x_0'\theta - k(\theta))d\theta = \sqrt{2\pi} \exp(n_0'(x_0'\theta(x_0') - k(\theta(x_0'))))$$
$$\times (n_0' k''(\theta(x_0')))^{-1/2} (1 + o_P(1)), \quad n_0' \to \infty. \tag{16}$$

To derive the asymptotics of the third integral in the right hand side of (15) one should pass to a new variable of integration, μ, and write

$$A_n = \int_{\Theta} (\mu(\theta) - \mu_n)^2 \psi''(\mu_n + \alpha(\mu(\theta) - \mu_n)) \exp(n_0'(x_0'\theta - k(\theta))d\theta$$
$$= \int_M (\mu - \mu_n)^2 \psi''(\mu_n + \alpha(\mu - \mu_n)) \exp(n_0'(x_0'\theta(\mu) - k(\theta(\mu))))\theta'(\mu)d\mu,$$

where $M = k'(\Theta)$ is the set of the all possible values of expectations of EDF. As function $k'(\theta)$ is monotone, continuous and a one-to-one function, Θ (or $int(\Theta)$ for a steep family) is an open interval (finite or infinite), M is also an interval (finite or infinite). Introduce

$$h(\mu) = x_0'\theta(\mu) - k(\theta(\mu)).$$

The function $h(\mu)$ has only one maximum $\bar{\mu} = x_0' = \mu_n$, and

$$h''(x_0') = [\theta''(\mu)(x_0' - \mu) - \theta'(\mu)]_{|\mu = x_0'} = -\theta'(x_0') = -V(x_0')^{-1}.$$

Then a slight generalization of the asymptotics of the Laplace-type integral (see Landsman (1996), Lemma 2) gives, putting in the cited Lemma $k = 2$ (taking into account that $\theta'(\mu) = V(\mu)^{-1}$ and using (13))

$$A_n = 2(\pi/2)^{1/2} \psi''(\mu_n)\theta'(x_0')$$
$$\times \exp(n_0'(x_0'\theta(x_0') - k(\theta(x_0'))))(n_0'V(x_0')^{-1})^{-3/2}(1 + o_P(1))$$
$$= 2(\pi/2)^{1/2} \psi''(\mu_n)$$
$$\times \exp(n_0'(x_0'\theta(x_0') - k(\theta(x_0'))))(n_0')^{-3/2}V(\mu_n)^{1/2}(1 + o_P(1)), \tag{17}$$
$$n_0' \to \infty.$$

Substituting (16) and (17) into (15) and taking into account that $k''(\theta(\mu)) = V(\mu)$ and using (13), (14), we obtain (10).   ∎

Remark 1. *Although the approach considered in Tierney and Kadane (1986) is also based on the asymptotics of the Laplace-type integrals, the suggested approach, based on Lemma 2 (Landsman (1996)) seems preferable because it results in terms of the maximal point $\hat{\theta}_n$ of function L (in the notation of Tierney and Kadane (1986)) or μ_n (in the notation of the paper), which does not depend on g (in the notation of Tierney and Kadane (1986)) or ψ (in the notation of the paper). At the same time the results of Tierney and Kadane (1986) require additional knowledge of the maximal point $\hat{\theta}_n^*$ of the second function L^*, which depends on function g (in the notation of Tierney and Kadane (1986)) or ψ (in the notation of the paper).*

  If $\psi(x)$ has $k > 2$ derivatives, Lemma 2 (Landsman (1996)) allows to clearing the reminder term up to the term of order $O_P((n_0 + n\lambda)^{-k/2})$. In particular if $\psi(x)$ is three times continuously differentiable, by the same approach one can write expression (10) more precisely as

$$E(\psi(\mu(\theta))|x_1, x_2, ..., x_n) = \psi(\mu_n) + \frac{1}{2}\frac{V(\mu_n)}{(n_0 + n\lambda)}\psi''(\mu_n) + O_P\left(\frac{1}{(n_0 + n\lambda)^{3/2}}\right).$$

3. SECOND ORDER PREDICTION OF THE TAILS

In this section we apply the results of the previous section for predicting the probability that a claim exceeds a threshold T. Recall that $\theta = \theta(\mu)$ is the inverse function of $\mu = \mu(\theta) = k'(\theta)$. Let

$$\bar{F}(T|\theta) = P(X > T|\theta) = \int_T^\infty \exp(\lambda(x\theta - k(\theta))\,dQ_\lambda(x) \qquad (18)$$

be the tail of the distribution of claim size,

$$\mu_T(\mu) = \int_T^\infty x \exp(\lambda(x\theta(\mu) - k(\theta(\mu))))\,dQ_\lambda(x) \qquad (19)$$

be a lower truncated expectation of the claim size, and

$$V_T(\mu) = \lambda \int_T^\infty (x - \mu)^2 \exp(\lambda(x\theta(\mu) - k(\theta(\mu))))\,dQ_\lambda(x) \qquad (20)$$

be a lower truncated variance function. It is clear that $\bar{F}(T|\theta)|_{T=0} = 1$, $\mu_T(\mu)|_{T=0} = \mu$, $V_T(\mu)|_{T=0} = V(\mu)$. Denoting $K = n_0/\lambda$, we observe that K plays the role of Bühlmann's coefficient in the case of EDF, because $K = E_\pi V(X|\theta)/V_\pi(E(X|\theta)$ (see, for example, Landsman and Makov (1999), Sect. 3). Recall that $E_\pi V(X|\theta)$ and $V_\pi(E(X|\theta)$ are called expected process variance and variance of the hypothetical means (Young (1998)), respectively.

Theorem 2. *Under the conditions of Theorem 1*

$$P(x_{n+1} > T \mid x_1,...,x_n) = \bar{F}(T \mid \theta(\mu_n))$$
$$+ \frac{1}{2} \frac{1}{K+n} \Delta_T(\mu_n) + o_P\left(\frac{1}{n}\right),$$

where

$$\Delta_T(\mu) = \frac{V_T(\mu)}{V(\mu)} - \bar{F}(T \mid \theta(\mu)) - (\mu_T(\mu) - \mu\bar{F}(T \mid \theta(\mu)))\frac{V'(\mu)}{V(\mu)} \qquad (21)$$

Proof. Setting

$$\psi(\mu) = \bar{F}(T \mid \theta(\mu)) \qquad (22)$$

and taking into account that $\theta'(\mu) = 1/\mu'(\theta) = 1/V(\mu)$, one can write straight-forwardly from (18), after differentiation under the integral sign,

$$\psi'(\mu) = \lambda[\mu_T(\mu) - \mu\bar{F}(T \mid \theta(\mu))]/V(\mu),$$

and from (19), (20)

$$\begin{aligned}
\psi''(\mu) &= \frac{\lambda}{V(\mu)^2} \{[\theta'(\mu)\lambda\int_T (x-\mu)^2 \exp(\lambda(x\theta(\mu) - k(\theta(\mu)))dQ_\lambda(x) \\
&\quad - \bar{F}(T \mid \theta(\mu))]V(\mu) \\
&\quad - V'(\mu)\int_T (x-\mu)\exp(\lambda(x\theta(\mu) - k(\theta(\mu)))dQ_\lambda(x)\} \\
&= \frac{\lambda}{V(\mu)}\left[\frac{V_T(\mu)}{V(\mu)} - \bar{F}(T \mid \theta(\mu)) - (\mu_T(\mu) - \mu\bar{F}(T \mid \theta(\mu)))\frac{V'(\mu)}{V(\mu)}\right].
\end{aligned} \qquad (23)$$

The proof of Theorem 2 follows from Theorem 1 after substituting (22) and (23) into (10), from the well-known fact that the predictive density is given by

$$f(x_{n+1} \mid x_1,...,x_n) = \int_\Theta f(x_{n+1} \mid \theta) f(\theta \mid x_1, x_2,...,x_n) d\theta \qquad (24)$$

(see Aitchinson and Dunsmore (1975)) and from Fubini's Theorem, providing the equality

$$P(x_{n+1} > T \mid x_1,...,x_n) = E(\bar{F}(T \mid \theta) \mid x_1,...x_n). \qquad \blacksquare$$

4. APPLICATIONS

In this section we apply the results of the previous section for the prediction of the probability that a future claim, whose distribution is Normal or Gamma, will exceed a large threshold.

4.1. Predictive tails for Normal claim sizes

Conditionally on μ, σ^2, the claim sizes $x_1, x_2, ..., x_n, x_{n+1}$ are assumed to be normally distributed, $N(\mu, \sigma^2)$. It is well-known that $N(\mu, \sigma^2)$ is a member of EDF with $\theta = \mu$, $\lambda = 1/\sigma^2$, $k(\theta) = 1/2\theta^2$ and $V(\mu) = 1$. Integrating by parts we get from (20)

$$
\begin{aligned}
V_T(\mu) &= \frac{1}{\sqrt{2\pi}\,\sigma} \int_T^\infty (x - \mu)^2 \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) \\
&= \sigma^{-1}(T - \mu)\varphi(\sigma^{-1}(T - \mu)) + \overline{\Phi}(\sigma^{-1}(T - \mu)),
\end{aligned}
\tag{25}
$$

where $\varphi(x)$ and $\Phi(x)$ are standard Normal density and distribution functions, $\overline{\Phi}(x) = 1 - \Phi(x)$. Noticing that in this case $V'(\mu) = 0$ we obtain, substituting (25) into (21),

$$
\Delta_T(\mu) = \sigma^{-1}(T - \mu)\varphi(\sigma^{-1}(T - \mu)).
$$

Then, from Theorem 2, we get the prediction of the tail of the coming claim up to the term $O_P\left(\frac{1}{n}\right)$,

$$
\begin{aligned}
P(x_{n+1} > T \,|\, x_1, ..., x_n) &= \overline{\Phi}(\sigma^{-1}(T - \mu_n)) \\
&+ \frac{1}{2}\frac{\lambda}{n_0 + n\lambda}\frac{(T - \mu_n)}{\sigma}\varphi(\sigma^{-1}(T - \mu_n)) + o_P\left(\frac{1}{n}\right).
\end{aligned}
$$

Recall that the conjugate prior distribution for normal claims is also normal, $N(m, \sigma_0^2)$, where m is the collective fair premium and variance $\sigma_0^2 = \frac{1}{n_0}$. Denote $\kappa = \sigma^2/\sigma_0^2$, Bühlmann's coefficient, then

$$
\mu_n = \frac{n}{\kappa + n}\bar{x} + \frac{\kappa}{\kappa + n}m
$$

is the credible mean, and

$$
\begin{aligned}
P(x_{n+1} > T \,|\, x_1, ..., x_n) &= \overline{\Phi}(\sigma^{-1}(T - \mu_n)) \\
&+ \frac{1}{2}\frac{1}{n + \kappa}\sigma^{-1}(T - \mu_n)\varphi(\sigma^{-1}(T - \mu_n)) + o_P\left(\frac{1}{n}\right).
\end{aligned}
\tag{26}
$$

Remark 2. *For Normal claims under normal prior one can calculate the exact expression for the predictive tail. In fact, formulas (24) and (12) yield*

$$
f(x_{n+1} | x_1, ..., x_n) = n(0, \sigma^2) * n(x_0', 1/n_0'),
$$

where the symbol $*$ means the convolution between two normal densities with parameters $(0, \sigma^2)$ and $(x_0', 1/n_0')$ respectively. Here $x_0' = \mu_n$, $n_0' = n_0 + n\lambda = \sigma_0^{-2} + n\sigma^{-2}$ (cf. (14)). Then $f(x_{n+1} | x_1, ..., x_n)$ is the density of $N(\mu_n, \sigma^2 + 1/(\sigma_0^{-2} + n\sigma^{-2}))$ and

$$P(x_{n+1} > T | x_1, ..., x_n) = \overline{\Phi}\left(\frac{T - \mu_n}{\sigma\left(\sqrt{1 + 1/(n + \kappa)}\right)}\right), \tag{27}$$

where $\kappa = \sigma^2/\sigma_0^2$. As

$$\frac{1}{\sqrt{1 + 1/(n + \kappa)}} = 1 - \frac{1}{2}\frac{1}{n + k} + o\left(\frac{1}{n}\right),$$

Taylor expansion of $\overline{\Phi}$ shows that (26) well conforms with (27).

4.2. Predictive tails for Gamma claim sizes

Conditionally on λ, α, the claim sizes $x_1, x_2, ..., x_n, x_{n+1}$ are assumed to be gamma distributed with shape and scale parameters $\lambda, \alpha \in R_+$ respectively. It is well known (see, for example, Jorgensen (1997), Ch.3), that the gamma family is a member of the EDF with canonical parameter $\theta = -\alpha/\lambda$, dispersion parameter λ, cumulant function

$$k(\theta) = -\log(-\theta), \tag{28}$$

mean $\mu = k'(\theta) = \lambda/\alpha$, and variance function

$$V(\mu) = \mu^2. \tag{29}$$

Let us calculate the all ingredients of Theorem 2. Denote by $G(x, \lambda, \alpha)$ the gamma distribution function and let $g(x, \lambda, \alpha)$ be the gamma density function. Integrating by parts we have

$$\begin{aligned}
\mu_T(\mu) &= \int_T^\infty g(x, \lambda, \alpha(\mu))\, dx \\
&= \frac{\alpha(\mu)^{\lambda-1}}{\Gamma(\lambda)} T^\lambda \exp(-\alpha(\mu)T) + \mu\bar{G}(T, \lambda, \alpha(\mu)),
\end{aligned} \tag{30}$$

where $\bar{G}(x, \lambda, \alpha) = 1 - G(x, \lambda, \alpha)$, $\alpha(\mu) = \lambda/\mu$, and

$$\begin{aligned}
V_T(\mu) &= \lambda \int_T^\infty (x - \mu)^2 g(x, \lambda, \alpha(\mu))\, dx = \lambda\frac{\alpha(\mu)^{\lambda-1}}{\Gamma(\lambda)} T^{\lambda+1} \exp(-a(\mu)T) \\
&+ (1 - \lambda)\mu\mu_T(\mu) + \lambda\mu^2\bar{G}(T, \lambda, \alpha(\mu)).
\end{aligned} \tag{31}$$

Substituting (30) into (31) we obtain

$$\begin{aligned}
V_T(\mu) &= \frac{\alpha(\mu)^{\lambda-1}}{\Gamma(\lambda)} T^\lambda \exp(-a(\mu)T)(\lambda T + (1 - \lambda)\mu) \\
&+ \mu^2 \bar{G}(T, \lambda, \alpha(\mu)).
\end{aligned} \tag{32}$$

Substituting (30) and (32) into (21) and taking into account (29) we can derive $\Delta_T(\mu)$ from Theorem 2:

$$
\begin{aligned}
\Delta_T(\mu) &= \frac{V_T(\mu)}{V(\mu)} - \bar{G}(T,\lambda,\alpha(\mu)) - \frac{2}{\mu}(\mu_T(\mu) - \mu\bar{G}(T|\lambda,\alpha(\mu))) \\
&= \frac{\alpha(\mu)^{\lambda-1}}{\Gamma(\lambda)} T^\lambda \exp(-\alpha(\mu)T)\left(\frac{\lambda}{\mu^2}T + \frac{(1-\lambda)}{\mu} - \frac{2}{\mu}\right) \\
&= \frac{1}{\mu}\frac{\alpha(\mu)^{\lambda-1}}{\Gamma(\lambda)} T^\lambda \exp(-\alpha(\mu)T)(\alpha(\mu)T - (\lambda+1)).
\end{aligned}
\tag{33}
$$

Let $G(x,\lambda)$ and $g(x,\lambda)$ be standard gamma distribution and density functions, i.e., having a scale parameter equal to 1. Then (33) reduces to

$$
\Delta_T(\mu) = \lambda\frac{T}{\mu}\left(\frac{T}{\mu} - \left(1 + \frac{1}{\lambda}\right)\right)g\left(\lambda\frac{T}{\mu},\lambda\right)
\tag{34}
$$

Recall that μ_n is the predicted mean of the future claim, calculated by the credibility formula (9) and set

$$
R_n = \frac{T}{\mu_n}.
$$

Then for a predictive tail we yield the following result from Theorem 2 and (34)

$$
\begin{aligned}
P(x_{n+1} > T | x_1,...,x_n) &= \bar{G}(\lambda R_n,\lambda) \\
&+ \frac{1}{2}\frac{\lambda}{K+n}R_n\left[R_n - \left(1 + \frac{1}{\lambda}\right)\right]g(\lambda R_n,\lambda) + o_P\left(\frac{1}{n}\right).
\end{aligned}
\tag{35}
$$

Putting in (35) $\lambda = 1$ we obtain the predictive tail for exponentially distributed, $EXP(\theta)$, claim sizes

$$
P(x_{n+1} > T | x_1,...,x_n) = \exp(-R_n)\left(1 + \frac{1}{2}\frac{1}{K+n}R_n(R_n - 2)\right) + o_P\left(\frac{1}{n}\right). \tag{36}
$$

Remark 3. *For exponential claims under conjugate (gamma) prior we can calculate the exact expression for the predictive tails. In fact, formulas (24) and (12) yield, taking into account (28),*

$$
\begin{aligned}
f(x_{n+1}|x_1,...,x_n) &= \frac{\left(n_0'x_0'\right)^{n_0'+1}}{\Gamma\left(n_0'+1\right)}\int_0^\infty \theta\exp(-x_{n+1}\theta)\exp\left(-n_0'x_0'\theta\right)\theta^{n_0'}d\theta \\
&= \frac{n_0'+1}{n_0'x_0'}\left(1 + \frac{x_{n+1}}{n_0'x_0'}\right)^{-\left(n_0'+2\right)}.
\end{aligned}
$$

Then, taking into account (14) and (13), we have

$$P\left(x_{n+1} > T \mid x_1,...,x_n\right) = \frac{n_0' + 1}{n_0' x_0'} \int_T^\infty \left(1 + \frac{x_{n+1}}{n_0' x_0'}\right)^{-\left(n_0' + 2\right)} dx_{n+1}$$

$$= \left(1 + \frac{T}{\mu_n(n_0 + n)}\right)^{-(n_0 + n + 1)} \tag{37}$$

Using the Taylor expansion of the function $\ln(1 + \alpha)$ up to the term $O(\alpha^3)$, $\alpha \to 0$ we get

$$-(n_0 + n)\ln\left(1 + \frac{T}{\mu_n(n_0 + n)}\right) = -\frac{T}{\mu_n} + \frac{1}{2}\frac{T^2}{\mu_n^2(n_0 + n)} + o_P\left(\frac{1}{n}\right). \tag{38}$$

Then, taking into account that

$$\left(1 + \frac{T}{\mu_n(n_0 + n)}\right)^{-1} = 1 - \frac{T}{\mu_n(n_0 + n)} + o_P\left(\frac{1}{n}\right),$$

we obtain from (37) and (38)

$$P\left(x_{n+1} > T \mid x_1,...,x_n\right) = \exp\left(-\frac{T}{\mu_n} + \frac{1}{2}\frac{T^2}{\mu_n^2(n_0 + n)} + o_P\left(\frac{1}{n}\right)\right)\left(1 + \frac{T}{\mu_n(n_0 + n)}\right)^{-1}$$

$$= \exp\left(-\frac{T}{\mu_n}\right)\left(1 + \frac{1}{2}\frac{T}{\mu_n(n_0 + n)}\left(\frac{T}{\mu_n} - 2\right)\right) + o_P\left(\frac{1}{n}\right),$$

which well conforms with (36).

4.3. Numerical experiment

In this subsection we illustrate how sharp the s.o. predictive formula performs. We generated 1000 samples of sizes $n = 8$ in Normal-Normal and Exponential-Gamma experiments respectively with Bühlmann coefficient $K = 4$. For each experiment we calculated predictive value-at-risks (VaR) using exact formulas for predictive tails ((27) and (37)), their first order (f.o.) approximations (obtained by simply substituting the predictive mean into tail's functional (18) before prediction) and s.o. predictive tails' formulas. Figures 1, 2, show that the s.o. tails' prediction piercingly estimates the exact predictive tails and is significantly superior to the f.o. prediction.
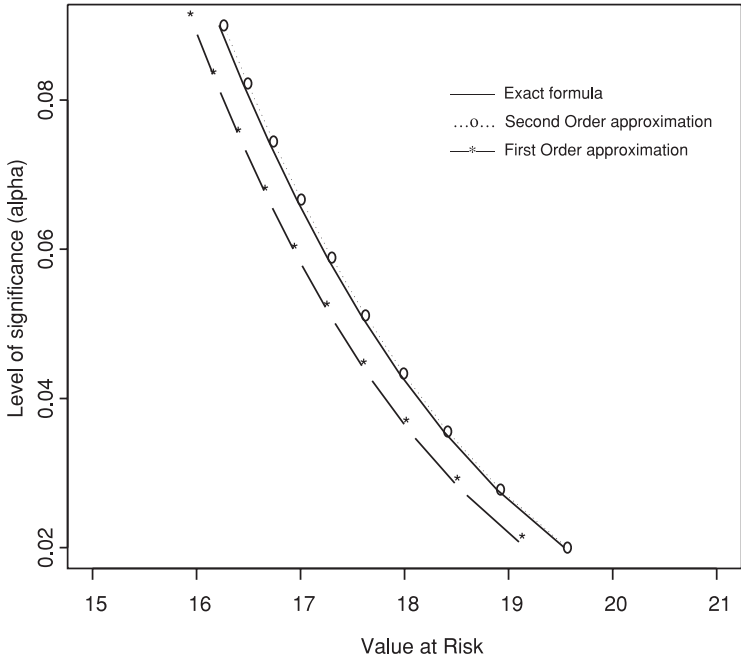
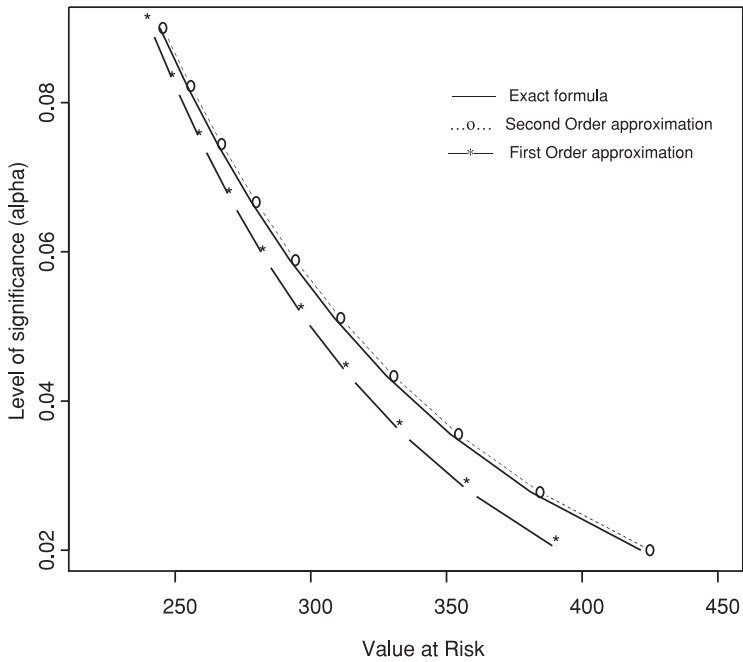FIGURE 1: Predictive VaRs for normal-normal experiment, $m = 10$.

FIGURE 2: Predictive VaRs for exponential losses, $m = 100$.

References

AITCHINSON, J., DUNSMORE, I.R. (1975) *Statistical Prediction Analysis*. Cambridge University Press, Cambridge.

BARNDORFF-NIELSEN, O.E. (1989) *Asymptotic Techniques for Use in Statistics*. Chapman and Hall, London, New York

BICKEL, P.J. (1981) Minimax estimation of the mean of a normal distribution when the parameter space is restricted. *Ann. Statist.*, **9(6)**, 1301-1309.

BROWN, L.D. (1986) *Fundamentals of Statistical Exponential Families*. Lecture Notes – Monograph series, Hayward, CA.

BÜHLMANN, H. (1967) Experience rating and probability. *The Astin Bull.* **4**, 199-207.

DIACONIS, P. and YLVISAKER, D. (1979) Conjugate priors for exponential families. *The Annals of Statist.* **7**, 269-281.

FULLER, W.A. (1996). *Introduction to Statistical Time Series*, Wiley. New York.

JEWELL, W.S. (1974) Credible means are exact Bayesian for exponential families. *Astin Bull.* **8**, 77-90.

JORGENSEN, B. (1986) Some properties of exponential dispersion models. *Scan. J. Statist.* **13**, 187-198.

JORGENSEN, B. (1987) Exponential dispersion models (with discussion). *J. Roy. Statist. Soc. Ser. B* **49**, 127-162.

JORGENSEN, B. (1992) Exponential dispersion models and extensions: A review. *Internat. Statist. Rev.* **60**, 5-20.

JORGENSEN, B. (1997) *The Theory of Dispersion Models*. Chapman and Hall, London.

KAAS, R., DANNENBURG, D. and GOOVAERTS, M. (1997) Exact credibility for weighted observations. *Astin Bull.* **27(2)**, 287-295.

LANDSMAN, Z. (1996) Sample quantiles and additive statistics: information, sufficiency, estimation. *Journal of Statistical Planning and Inference*, **52**, 93-108.

LANDSMAN, Z. (2001) Second order minimax estimation of the mean value for Exponential Dispersion models. *Journal of Statistical Planning and Inference*, **98**, 57-71.

LANDSMAN, Z. (2002) Credibility theory: A new view from the theory of second order optimal statistics. *Insurance: Mathematics & Economics*, **30**, 351-362.

LANDSMAN, Z. and LEVIT, B. (1990) The second order minimax estimation: nuisance parameters and hypoelliptic operators. *Probability Theory and Mathematical Statistics*, Proceedings of the Fifth International Vilnius Conference, **2**, 47-58, VSP, Utrecht.

LANDSMAN, Z. and LEVIT, B. (1991) Second order asymptotic minimax estimation in the presence of a nuisance parameter. *Problems of Information Transmissions*, **26(3)**, 226-244.

LANDSMAN, Z. and MAKOV, U. (1998) Exponential dispersion models and credibility. *Scand. Actuarial J.*, **1**, 89-96.

LANDSMAN, Z. and MAKOV, U. (1999) On stochastic approximation and credibility. *Scand. Actuaial J.*, **1**, 15-31.

LEVIT, B. (1980) Second order minimaxity. *Theory of Prob. Appl.* **25**, 3, 561-576.

LEVIT, B. (1982) Second order minimax estimation and positive solutions of elliptic equations. *Theory Prob. Appl.* **27(3)**, 525-546.

LINDLEY, D.V. (1980) Approximate Bayesian Methods. *Bayesian Statistics*. Proceedings of the 1st International Meetings held in Valencia, 223-237.

NELDER, J.A. and VERRALL, R.J. (1997) Credibility theory and generalized linear models. *Astin Bull.*, **27(3)**, 71-82.

NELDER, J.A. and WEDDERBURN, R.W.M. (1972) Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135**, 370-384.

TIERNEY, L. and KADANE, J.B. (1986) Accurate approximations for posterior moments and marginal densities *JASA*, **81**, 393, 82-86.

TWEEDIE, M.C.K. (1984) An index which distinguishes between some important exponential families. In *Statistics: Applications and New Directions. Proceedings of the Indian Statistical Golden Jubilee International Conference* (Eds. J.K. Ghosh and J. Roy), 579-604. Indian Statistical Institute.

YOUNG, V.R. (1998) Credibility using a loss function from spline theory: Parametric models with a one-dimensional sufficient statistic. *North American Actuarial Journal*, **2**, 1, 101-117.

ZINOVIY LANDSMAN
Department of Statistics
University of Haifa
Mount Carmel, Haifa 31905, Israel.
E-mail: landsman@stat.haifa.ac.il