Sequential quasi credibility for scale dispersion models

Zinoviy Landsman and Udi Makov
Department of Statistics, University of Haifa

Abstract

The sequential approach to credibility, developed in Landsman and Makov (1999a,b) is extended to the *scale dispersion family,* which contains distributions often used in actuarial science: Log-normal, Weibull, half Normal, Stable, to mention only a few. For members of this family a *sequential quasi credibility formula* is devised , which can also be used for heavy tailed claims.

1 Introduction

In this paper we extended the sequential approach to a problem of credibility evaluation, recently developed in (Landsman and Makov (1999a,b)) and applied there for a *location dispersion family,* to a *scale-dispersion family* (SDF). We consider SDF in it's most general form (Jorgensen (1997), Sect. 1.4.2) as follows

$$dP_{\theta,\lambda} = a(\lambda)x^{-1}\exp(\lambda u(\theta x))dx, \quad x \in R_{+} \tag{1}$$

Here $\theta \in R_{+}$ is the risk parameter of $P_{\theta,\lambda}$, the distribution of the claim size X, and $\lambda \in R_{+}$ is a dispersion parameter. SDF contains many popular distributions. For example the lognormal family, frequently used successfully to represent loss data, is a member of SDF, with $u(x) = -\frac{1}{2}\ln^2 x$. A considerably rich 3-parameter class of SDF was introduced by Fergusson (1962),

$$dP_{\theta,\lambda,\gamma} = \frac{\lambda^{\lambda}|\gamma|}{\Gamma(\lambda)}\frac{1}{x}\exp(\lambda(\gamma\ln(x\theta) - (x\theta)^{\gamma}), \tag{2}$$

Distribution	γ	λ
Gamma	1	$\lambda \in R_+$
Weibull	$\gamma > 0$	1
Stable	-1	1/2
Uniform($U([0,\theta])$)	$\to \infty$	$1/\gamma$
Half-normal	2	1/2

Table 1: Fergusson's Family

including many popular models. Here $\gamma \in R\backslash\{0\}$ is a shape parameter. Varying with γ and sometimes with λ one may obtain different models, some of which are given in the Table 1.

Notice that a *scale family*, where only 1 parameter, scale, is specified, can be considered as SDF with $\lambda = 1$.

Remark 1 *We either assume that λ is known a priori, or, if not, several alternatives are available. In Landsman and Makov (1998, 1999c, 2000) the following cases were considered:*

1. *The prior density of $\lambda, g(\lambda)$, is known.*

2. *$g(\lambda)$ is not specified, however, the first two moments of λ is known.*

3. *$g(\lambda)$ is not specified, but the probability that the claim X exceeds a given threshold T is known.*

4. *A non-Bayesian approach: λ is established by the method of moments or MLE (see Landsman and Makov (1998)). For example, for lognormal claims MLE of λ can be obtained straightforwardly*

$$\hat{\lambda}_n = (\frac{1}{n}\sum_{i=1}^{n} \ln^2(x_i\hat{\theta}_n))^{-1},$$

where

$$\hat{\theta}_n = (\prod_{i=1}^{n} x_i)^{-1/n}$$

Let

$$\bar{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i$$

be the average of n years of individual experience $x_1, ..., x_n$, and m is the collective fair premium. The traditional linear credibility formula

$$\hat{\mu}_n = (1-\alpha_n)m + \alpha_n \bar{x}_n \tag{3}$$

provides, in the Bayesian set up, the exact predicted mean.

$$(1-\alpha_n)m + \alpha_n \bar{x}_n = E(x_{n+1} \mid x_1, x_2, ...x_n) \tag{4}$$

This equality holds only for the class of distributions $P_{\theta,\lambda}$ belonging to the *Exponential Dispersion Family* (EDF), i.e., having the form

$$dP_{\theta,\lambda} = f(x \mid \theta, \lambda)dx \ = e^{\lambda(x\theta - k(\theta))} q_\lambda(x)dx, \theta \in \Theta \subset R, \lambda \in R^+, \tag{5}$$

$\Theta = \{\theta \mid \lambda k(\theta) = \ln \int e^{\lambda\theta x} q_\lambda(x)dx < \infty\}$ (Herzog (1994), Kaas, Dannenburg, Goovaerts (1997), Nelder and Verrall (1997), Landsman and Makov (1998)). The prior distribution $\pi(\theta)$, which produces (4), is called a conjugate prior distribution and has the form

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

In this context $m = \int_\Theta \mu(\theta)\pi(\theta)d\theta$, where $\mu = \mu(\theta) = E(X|\theta)$ is the mean of the claim size X. For references on the EDF see Tweedie (1984), Nelder and Wedderburn (1972) and Jorgensen (1986, 1987, 1992, 1997).

For claim distributions which are not members of the EDF the predicted mean is no longer linear with respect to the data (see Diaconis and Ylvisaker (1979)) and the identity between the predicted mean claim and the linear credibility formula is no longer true. Landsman and Makov (1999a,b) exploited a relationship between the standard linear credibility formula and stochastic approximation and suggested a sequential approach to credibility. This approach led to generalized sequential credibility formulas for the *Location Dispersion and Symmetric Location Dispersion Families* (LDF and SLDF). In the current paper we will extend this approach to the SDF.

Let us recall the main futures of stochastic approximation, originally proposed by Robbins and Monroe (1951), a methodology aimed at finding the root of a function which is neither known nor directly observable.

Let $g(t)$ be a trend of some stochastic process $Z(t)$ with a unique root τ i.e.

$$g(t) = EZ(t) = \int xf\left(x|t\right)\right) dz,$$

$$g(\tau) = 0,$$

where $f(x \mid t)$ is the probability density function of $Z(t)$.

Stochastic approximation theory examines the random sequence $t_0, t_1, \ldots,$ in which t_0 is an arbitrary number and $t_1, t_2, \ldots,$ are evaluated in terms of the observed $Z_1(t_1), Z_2(t_2), \ldots,$ by

$$t_{n+1} = t_n - a_n Z_n(t_n), \tag{6}$$

where a_n is a gain sequence of positive numbers. The convergence of (6) was established under various conditions. See Gladyshev (1965) for a possible set of conditions.

Over the years, stochastic approximation has been extensively studied and proved to be a powerful and useful tool. For a recent reference see Kushner and Yin (1997).

In Landsman and Makov (1999a) the generalized sequential credibility formula was introduced as a variant of stochastic approximation that corresponds to the process

$$Z(t) = \mu'(\mu^{-1}(t))\tilde{J}(X|t, \lambda). \tag{7}$$

Here $\mu^{-1}(\cdot)$ is the inverse function of $\mu(\cdot)$ and

$$\tilde{J}(x|\mu, \lambda) = \frac{\partial}{\partial\mu} \log \tilde{f}(x|\mu, \lambda)$$

is the score of X with respect to parameter μ ($\tilde{f}$ is the density of X after reparametrization from θ to μ). Starting from $\hat{\mu}_0 = m$ we write the generalized sequential credibility formula as follows

$$\hat{\mu}_n = \hat{\mu}_{n-1} - a_n\mu'(\mu^{-1}(\hat{\mu}_{n-1}))\tilde{J}(x_n|\hat{\mu}_{n-1}, \lambda). \tag{8}$$

For the EDF we have, from (5) and (7), that

$$Z(t) = \lambda(t - X)$$

and then (8) has the form $\hat{\mu}_0 = m$,

$$\hat{\mu}_n = \hat{\mu}_{n-1} - a_n(\hat{\mu}_{n-1} - x_n). \tag{9}$$

Taking into account that

$$\bar{x}_n = \bar{x}_{n-1} - \frac{1}{n}(\bar{x}_{n-1} - x_n)$$

it is straight forward to establish that (9) is equivalent to (3) with $\alpha_n = na_n$.

2 Scale Dispersion Family

Let $P_{\theta,\lambda}$ be a member of SDF, i.e. be of the form (1). Then one straightforwardly obtains

$$\mu = \mu(\theta) = \frac{\mu_0(\lambda)}{\theta},$$

where $\mu_0(\lambda)$ is the expected value of (1) for $\theta = 1$, i.e.,

$$\mu_0(\lambda) = E_{\theta=1,\lambda}X = a(\lambda)\int \exp(\lambda u(x))dx. \tag{10}$$

For SDF we write down the process $Z(t)$, described in (7),

$$Z(t) = \lambda x u'(x\frac{\mu_0(\lambda)}{t}), \tag{11}$$

and the sequential formula (8)

$$\hat{\mu}_n = \hat{\mu}_{n-1} - a_n\lambda x_n u'(x_n\frac{\mu_0(\lambda)}{\hat{\mu}_{n-1}}). \tag{12}$$

For some distributions the process $Z(t)$ is equal to -$\lambda x u'(x\frac{\mu_0(\lambda)}{t})$ and consequently the sequential formula takes the form

$$\hat{\mu}_n = \hat{\mu}_{n-1} + a_n\lambda x_n u'(x_n\frac{\mu_0(\lambda)}{\hat{\mu}_{n-1}}) \tag{13}$$

3 Convergence of the sequential procedure

A full list of conditions guaranteeing the convergence of (6) to the true value τ can find in Landsman and Makov(1999a). In this section we consider only three conditions: Two on the gain sequences

$$C_1 : \sum_{n=1}^{\infty} a_n = \infty,$$
$$C_2 : \sum_{n=1}^{\infty} a_n^2 < \infty,$$

i.e., that the gain sequence should decrease to zero as an harmonic series, and the most difficult condition to verify

$$C_3 : \inf_{\{t:\varepsilon<|t-\tau|<\varepsilon^{-1}\}} (t-\tau)g(t) > 0, \text{ for each } \varepsilon > 0 ,$$

which implies, in fact, that $g(t)$ has only one root τ. Based on these conditions we can now state the following theorem.

Theorem 1 *Let $u(x)$ in (1) be concave (convex), twice differentiable and for any $k > 0$, $xu'(kx)^2$, and $xu''(kx)$ be integrable with respect to the measure $\exp(\lambda u(x))dx$. Then for the gain sequence, satisfying conditions C_1, C_2, sequential procedure (12) (or (13)) converges a.s. to the fair premium μ.*

Instead of the concavity (convexity) of $u(x)$ one can assume that $\tilde{u}(x) = u(\exp(x))$ is concave (convex).

Proof. The only condition needed to establish convergence is condition C_3. From (11) we have

$$g(t) = \lambda a(\lambda) \int_0^\infty u'(x\frac{\mu_0}{t}) \exp(\lambda u(x\frac{\mu_0}{\mu}))dx. \tag{14}$$

Changing variables in the integral (14) and denoting $k = \mu/t$ we obtain

$$g(t) = \tilde{g}(k) = \lambda a(\lambda)\frac{\mu}{\mu_0} \int_0^\infty u'(kx) \exp(\lambda u(x))dx. \tag{15}$$

The conditions of the Theorem allow us to differentiate the integral in (15) with respect to k under the sign of the integral and for concave $u(x)$ (convex) to obtain

$$\tilde{g}'(k) = \lambda a(\lambda)\frac{\mu}{\mu_0} \int_0^\infty xu''(kx) \exp(\lambda u(x))dx < 0 \ (> 0)$$

for any k. Then $\tilde{g}(k)$ has only one root $k = 1$, and consequently $g(t)$ has only one root $t = \mu$. C_3 is proved.

If $u(x)$ happens not to be concave (convex), but $\tilde{u}(x) = u(\exp(x))$ is concave (convex), then denoting $p = \ln k$ and replacing the variable x in the integral in (15) to a new variable of integration $z = \ln x$ we obtain

$$g(t) = \lambda a(\lambda)\frac{t}{\mu_0}\int_{-\infty}^{\infty} \tilde{u}'(z+p)\exp(\lambda u(z))dz.$$

As $\tilde{u}(x)$ is concave (convex), the function

$$g_1(p) = \int_{-\infty}^{\infty} \tilde{u}'(z+p)\exp(\lambda u(z))dz,$$

whose derivative for any $p \in R$

$$g_1'(p) = \int_{-\infty}^{\infty} \tilde{u}''(z+p)\exp(\lambda u(z))dz < 0 \;(> 0),$$

has only one root $p = 0$, and consequently the function $g(t)$ has only one root $\tau = \mu \exp(-p)_{|p=0} = \mu$. Condition C_3 is established and thus the Theorem is proved. ■

3.0.1 Examples

- Lognormal distribution is SDF with $u(x) = -\frac{1}{2}\ln^2 x$. Then since

  $$\tilde{u}(x) = -\frac{1}{2}x^2$$

  is concave, Theorem 1 holds.

- Fergusson's Family (see (2)) is SDF with $u(x) = \gamma \ln x - x^\gamma$. Then

  $$\tilde{u}(x) = \gamma x - \exp(\gamma x),$$

  and for any $\gamma \neq 0$ and $x \in R$, $\tilde{u}''(x) = -\gamma^2 \exp(\gamma x) < 0$ and thus Theorem 1 holds.

4 Stepwhise optimal gain sequence

In Landsman and Makov (1999a) it was established that for members of the EDF, the linear sequential credibility formula (9) coincides with the classical credibility formula if the gain sequence a_n is optimal with respect to it's risk for any step n. Such a property was called a stepwise optimal gain property. For the LDF, whose members, except the normal distribution, are non members of the EDF, such principle was modified to the so-called first order stepwise optimal principle. Now we adopt it for the SDF.

Under conditions of Theorem 1 and Taylor expansion we obtain from (13) for $n \to \infty$,

$$\begin{aligned}\mu_n - \mu &= (\mu_{n-1} - \mu)(1 + a_n \lambda \frac{x_n^2}{\mu^2} \mu_0(\lambda) u''(x_n \frac{\mu_0(\lambda)}{\mu})) - a_n \lambda x_n u'(x_n \frac{\mu_0(\lambda)}{\mu}), \\ &\quad + o_p(a_n(\mu_{n-1} - \mu)), \end{aligned} \tag{16}$$

taking into account that for $\mu \neq 0$,

$$\frac{\mu}{\mu_{n-1}} - 1 = O_p(\frac{\mu - \mu_{n-1}}{\mu}).$$

Ignoring the last term in the right hand side of (16) we obtain the recursion

$$\mu_n^* - \mu = (\mu_{n-1}^* - \mu)(1 + a_n \lambda \frac{x_n^2}{\mu^2} \mu_0(\lambda) u''(x_n \frac{\mu_0(\lambda)}{\mu})) - a_n \lambda x_n u'(x_n \frac{\mu_0(\lambda)}{\mu}) \tag{17}$$

which we call the first order approximation of (16).

Let $E_{\mu,\lambda}(\cdot)$ be expectation for given parameters μ, λ; $E_\lambda(\cdot)$ be the expectation with respect to measure $\pi(\mu) dP_{\mu,\lambda} d\mu$ with prior density $\pi(\mu)$ of parameter μ, given λ, and

$$R_n = E_\lambda(\mu_n^* - \mu)^2 = \int E_{\mu,\lambda}(\mu_n^* - \mu)^2 \pi(\mu) d\mu$$

the Bayes risk μ_n^*.

Definition 1 *A gain sequence $\{a_n\}$ is called the first order optimal stepwise gain sequence if it minimizes (Bayes) risk R_n of μ_n^* at every step n .*

Denote by

$$I_\lambda = \lambda^2 a(\lambda) \int_0^\infty x u'(x)^2 \exp(\lambda u(x)) dx = -\lambda a(\lambda) \int_0^\infty x u''(x) \exp(\lambda u(x)) dx, \tag{18}$$

$$B_\lambda = \lambda^2 a(\lambda) \int_0^\infty x^3 u''(x)^2 \exp(\lambda u(x)) dx, \tag{19}$$

and

$$D_\lambda = \lambda^2 a(\lambda) \int_0^\infty x^2 u'(x) u''(x) \exp(\lambda u(x)) dx. \tag{20}$$

I_λ is the Fisher information about parameter θ at point $\theta = 1$.

We now establish the first order optimal stepwise gain sequence

Theorem 2 *Suppose that the random variable μ has a finite variance $V(\mu) = \int (\mu(\theta) - m)^2 \pi(\theta) d\theta$ and denote by $E(\mu^2) = \int \mu(\theta)^2 \pi(\theta) d\theta$ and $CV(\mu) = \sqrt{V(\mu)}/m$ the coefficient of variation. In addition to the conditions of Theorem 1 let the function $x^3 u''(x)^2$ be integrable with respect to the measure $\exp(\lambda u(x))dx$. Then starting from $\mu_0^* = m(1 + CV(\mu)^2)$, the first order optimal stepwise gain sequence takes the form*

$$a_n = \frac{\mu_0(\lambda) R_{n-1}}{E\mu^2 + \kappa R_{n-1}} \tag{21}$$

where

$$\kappa = B_\lambda / I_\lambda \tag{22}$$

and the corresponding Bayes risk

$$R_n = \frac{E\mu^2 R_{n-1}}{\kappa R_{n-1} + E\mu^2} + o(R_{n-1}). \tag{23}$$

If $D_\lambda = 0$ the first order optimal stepwise gain and risk sequences take the form (21), (23) starting from any point μ_0^*.

Proof. For the proof of Theorem we prove the following lemma.

Lemma 1 *Under conditions of Theorem 2*

$$E_\lambda \mu(\mu_n^* - \mu) = 0, \tag{24}$$

if starting point $\mu_0^* = m(1 + CV(\mu)^2)$. If $D_\lambda = 0$, (24) holds for any starting point μ_0^*.

Proof. *As* μ_{n-1}^* and x_n are independent we can write from (16)

$$E_{\mu,\lambda}\mu(\mu_n^* - \mu) = \mu E_{\mu,\lambda}(\mu_{n-1}^* - \mu)(1 - a_n \frac{\lambda}{\mu_0(\lambda)} I_\lambda), \tag{25}$$

taking into account that

$$E_{\mu,\lambda}(x_n u'(x_n \frac{\mu_0(\lambda)}{\mu})) = 0. \tag{26}$$

Because I_λ doesn't depend on μ, we obtain after integrating both sides of (25) with respect to prior measure $\pi(\mu)d\mu$,

$$E_\lambda \mu(\mu_n^* - \mu) = (1 - a_n \frac{\lambda}{\mu_0(\lambda)} I_\lambda) E_\lambda \mu(\mu_{n-1}^* - \mu). \tag{27}$$

The sequential calculations gives

$$\begin{aligned} E_\lambda \mu(\mu_n^* - \mu) &= \prod_{i=1}^n ((1 - a_i \frac{\lambda}{\mu_0(\lambda)} I_\lambda) E_\lambda \mu(\mu_0^* - \mu)) = \\ & \prod_{i=1}^n ((1 - a_i \frac{\lambda}{\mu_0(\lambda)} I_\lambda)(\mu_0^* m - E\mu^2) \end{aligned}$$

Starting from $\mu_0^* = m(1 + V(\mu)/m^2) = E\mu^2/m$ we obtain (24). ∎

Return to the proof of Theorem. Squaring the both sides of (17), using (26) and again taking into account that μ_{n-1}^* and x_n are independent, we obtain

$$\begin{aligned} E_{\mu,\lambda}(\mu_n^* - \mu)^2 &= (1 - 2a_n \frac{I_\lambda}{\mu_0(\lambda)} + a_n^2 \frac{B_\lambda}{\mu_0(\lambda)^2}) E_{\mu,\lambda}(\mu_{n-1}^* - \mu)^2 + a_n^2 \frac{B_\lambda}{\mu_0(\lambda)^2} \mu^2 I_\lambda \\ & -2a_n^2 \frac{D_\lambda}{\mu_0(\lambda)^2} \mu E_{\mu,\lambda}(\mu_{n-1}^* - \mu). \end{aligned} \tag{28}$$

As $I_\lambda, B_\lambda, D_\lambda$ do not depend on μ, after integrating the both sides of (28) with respect to measure $\pi(\mu)d\mu$ we obtain

$$\begin{aligned} R_n &= (1 - 2a_n \frac{I_\lambda}{\mu_0(\lambda)} + a_n^2 \frac{B_\lambda}{\mu_0(\lambda)^2}) R_{n-1} + a_n^2 \frac{I_\lambda}{\mu_0(\lambda)^2} E\mu^2 \\ & -2a_n^2 \frac{D_\lambda}{\mu_0(\lambda)^2} E_\lambda(\mu(\mu_{n-1}^* - \mu)) \end{aligned}$$

If $D_\lambda = 0$

$$R_n = (1 - 2a_n \frac{I_\lambda}{\mu_0(\lambda)} + a_n^2 \frac{B_\lambda}{\mu_0(\lambda)^2}) R_{n-1} + a_n^2 \frac{I_\lambda}{\mu_0(\lambda)^2} E\mu^2, \tag{29}$$

if $D_\lambda \neq 0$, and a starting point $\mu_0^* = m(1+CV(\mu)^2)$, from Lemma 1 we again have (29).

Setting the first derivative of the right part of (29) with respect to a_n to 0 we obtain (21). Substituting this optimal a_n into (29) we straightforwardly obtain (23). ■

Following Landsman and Makov (1999a) we modify the risk and gain sequences, dropping the second order term on the right hand side of (23). Then the modified stepwise optimal gain and risk sequences are, respectively,

$$\begin{cases} R_n^* = \frac{E\mu^2}{\kappa R_{n-1} + E\mu^2} R_{n-1}^*, \quad n = 1, 2, ..., \\ \quad \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \quad R_0^* = R_0 = E_\lambda(\mu - \mu_0^*)^2, \end{cases} \tag{30}$$

and

$$a_n^* = \frac{\mu_0(\lambda)}{E\mu^2 + \kappa R_{n-1}^*} R_{n-1}^*, \quad n = 1, 2, ... \tag{31}$$

We now establish the modified first order stepwise optimal gain and risk sequences.

Theorem 3 *Under conditions of Theorem 2 suppose that $V_{\mu,\lambda}(X) < \infty$. Then the modified first order stepwise optimal gain and risk sequences are*

$$R_n^* = \frac{R_0}{n\kappa R_0/(c_\lambda V_\lambda(X)) + 1}, \tag{32}$$

$$a_n^* = \frac{\mu_0(\lambda)}{n\kappa + c_\lambda V_\lambda(X)/R_0}, \quad n = 1, 2, ...,$$

where

$$c_\lambda = \frac{\mu_0(\lambda)^2}{V_{\mu_0(\lambda),\lambda}(X)} \tag{33}$$

Proof. From (30) and (31) we obtain

$$\frac{1}{R_n^*} = \frac{\kappa}{E\mu^2} + \frac{1}{R_{n-1}^*}$$

and

$$a_n^* = \frac{\mu_0(\lambda)}{\kappa + E\mu^2/R_{n-1}^*} = \frac{\mu_0(\lambda)}{2\kappa + E\mu^2/R_{n-2}^*} = \cdots = \frac{\mu_0(\lambda)}{n\kappa + E\mu^2/R_0}.$$

The proof of Theorem follows since

$$V_{\mu,\lambda}(X) = \frac{\mu^2}{\mu_0(\lambda)^2} V_{\mu_0(\lambda),\lambda}(X),$$

and then $E\mu^2 = c_\lambda V_\lambda(X)$. ■

Remark 2 *If under conditions of Theorem 2 the distribution of the claim size X is heavy tailed in the sense that $V_{\mu,\lambda}(X) = \infty$, but the prior distribution of mean μ has a finite variance, i.e., $V_\lambda(\mu) < \infty$, then the modified first order stepwise optimal gain and risk sequences can still be calculated by the formulae*

$$R_n^* = \frac{R_0}{n\kappa R_0/E\mu^2 + 1},$$

$$a_n^* = \frac{\mu_0(\lambda)}{n\kappa + E\mu^2/R_0}, \quad n = 1, 2, ...$$

The following theorem clears the relation between κ, I_λ and V and is important for comparing the risks of the sequential quasi-linear and the linear procedure.

Theorem 4 *Under the conditions of Theorem 2*

$$\kappa \geq I_\lambda \geq V_{\mu_0(\lambda),\lambda}(X)^{-1} \tag{34}$$

and equality $\kappa = I_\lambda = V_{\mu_0(\lambda),\lambda}(X)^{-1}$ holds iff X is a Gamma distributed random variable.

Proof. The left inequality in (34) follows from Cauchy-Schwartz inequality, (18), (19) and (22) as follows

$$\begin{aligned} I_\lambda^2 &= (\lambda a(\lambda)\int_0^\infty x^2 u''(x)\frac{1}{x}\exp(\lambda u(x))dx)^2 \\ &\leq \lambda^2 a(\lambda)\int_0^\infty x^4 u''(x)^2\frac{1}{x}\exp(\lambda u(x))dx \cdot a(\lambda)\int_0^\infty \frac{1}{x}\exp(\lambda u(x))dx = B_\lambda \end{aligned}$$

Moreover the equality in (34) holds iff

$$x^2 u''(x) = c \;\; a.s.,$$

where c is some constant. As $u(x)$ is twice differentiable it should have (after a double integration of c/x^2) the form

$$u(x) = \alpha \ln x + \beta x + \gamma,$$

where α, β, γ are some constants. That means that $a(\lambda)\frac{1}{x}\exp(\lambda u(x))$ is the density of the Gamma distribution.

The right inequality in (34) follows from the well-known Rao-Cramer inequality (see, for example, Zacks (1971), Sect. 1.4). ■

As a result of the section we obtained the recursion

$$\hat{\mu}_n = \hat{\mu}_{n-1} - \frac{\lambda\mu_0(\lambda)}{n\kappa + c_\lambda V_\lambda(X)/R_0} x_n u'(x_n\frac{\mu_0(\lambda)}{\hat{\mu}_{n-1}}), n = 1, 2, .. \tag{35}$$

$$\hat{\mu}_0 = \begin{cases} m, \text{ if } D_\lambda = 0 \\ m(1 + CV(\mu)^2), \text{ if } D_\lambda \neq 0 \end{cases}, \tag{36}$$

$$R_0 = \begin{cases} V(\mu), \text{ if } D_\lambda = 0, \\ V(\mu)(1 + CV(\mu)), \text{ if } D_\lambda \neq 0, \end{cases}$$

for a finite variance claims distributions, and for heavy tailed claims distributions in the sense of infinite claim variance, the recursion (35) is modified to be

$$\hat{\mu}_n = \hat{\mu}_{n-1} - \frac{\lambda\mu_0(\lambda)}{n\kappa + E\mu^2/R_0} x_n u'(x_n\frac{\mu_0(\lambda)}{\hat{\mu}_{n-1}}), n = 1, 2, .. \tag{37}$$

We call both of these recursions *the sequential quasi credibility formula.*

As a comparison we give the linear sequential credibility formula, which follows immediately from the traditional credibility formula (of the Bühlmann type, Bühlmann, 1967) for the EDF (see Landsman and Makov (1999a), (8), (31)).

$$\begin{aligned} \hat{\mu}_n &= \hat{\mu}_{n-1} - \frac{\lambda}{n\lambda + \lambda V_\lambda(X)/V(\mu)}(\hat{\mu}_{n-1} - x_n), n = 1, 2, \ldots . \\ \hat{\mu}_0 &= m, \end{aligned} \tag{38}$$

We can see that (35) is the generalization of (38), where in the denominator the coefficient multiplying n is κ instead of λ. It tunes the score of the current claim for the sake of predicting the next one. The coefficient multiplying $V_\lambda(X)/R_0$ is c_λ instead of λ. Let us recall that $V_\lambda(X)$ and $V(\mu)$ are called respectively (Young (1996)) the *expected process variance* and the *variance of the hypothetical means.*

Remark 3 *Performing the Bühlmann credibility formula (cf. (38)) one doesn't need the full information about the prior density. It is enough to know only 2 moments:*

$V_\lambda(X)$ *- the expectation of process variance and*

$R_0 = V(\mu)$ *- the variance, corresponding to the prior of* μ.

The sequential quasi credibility formula (35) preserves this property.

As a matter of fact $V(\mu)$, *the collective (or "industry") variance may be known a priory. If not, it can be estimated from portfolio data in a way similar to the Bühlmann empirical Bayes approach.*

Remark 4 *From Theorem 4 we can see that for non gamma claims the modified Bayes risk* R_n^* *(see (32)) of the sequential quasi linear procedure (37) is less than the Bayes risk* R_n *of the linear procedure (38) , which is equal to*

$$R_n = \left(\frac{n}{V_\lambda(X)} + \frac{1}{R_0}\right)^{-1} \tag{39}$$

(see Landsman and Makov (1999a), Remark 3). In fact, for given λ, *SDF can be easy reparametrized such that* $\mu_0(\lambda)$ *becomes equaled to* 1, *i.e., one can rewrite (1) as follows*

$$dP_{\theta,\lambda} = a(\lambda)x^{-1}\exp(\lambda\breve{u}(\theta x))dx, \quad x \in R_+,$$

where $\breve{u}(x) = u(\mu_0(\lambda)x)$, and θ is a new scale parameter. Since $c_\lambda \le \kappa$ (cf. (33), (34))

$$R_n \ge \left(\frac{\kappa n}{c_\lambda V_\lambda(X)} + \frac{1}{R_0}\right)^{-1} = R_n^*. \quad (40)$$

The equality in the left hand side of (40) holds only for gamma claims.

Formula (37) can be used for the cases for which the exact linear credibility (38) does not exists.

5 Applications

5.1 Gamma claims

Recall that Gamma distribution is a member of the SDF with $u(x) = \ln x - x$. Then from (10) we have $\mu_0(\lambda) = 1$, $V_{\mu_0(\lambda),\lambda}(X) = 1/\lambda$. From (18)-(20) it follows that $I_\lambda = \lambda$, $B_\lambda = \lambda^2$, $D_\lambda = 0$, consequently

$$\kappa = B_\lambda / I_\lambda = I_\lambda = V_{\mu_0(\lambda),\lambda}(X)^{-1} = c_\lambda = \lambda$$

(cf. Theorem 4), and the sequential quasi credibility formula (35), (36) takes the form of the linear sequential credibility formula (cf. Landsman and Makov (1999a), (8), (31)).

5.2 Fergusson's family

Let us apply our quasi credibility formulae for the Fergusson's class of distributions given in (2), which is SDF with

$$u(x) = \gamma \ln x - x^\gamma.$$

Then from (2) one straightforwardly gets

$$\mu_0(\lambda) = \frac{\Gamma(\lambda + 1/\gamma)}{\lambda^{1/\gamma}\Gamma(\lambda)}, \quad V_{\mu_0(\lambda),\lambda}(X) = \frac{\Gamma(\lambda)\Gamma(\lambda + 2/\gamma) - \Gamma(\lambda + 1/\gamma)^2}{\lambda^{2/\gamma}\Gamma(\lambda)^2}$$

and consequently from (33)

$$c_\lambda = c_{\lambda,\gamma} = \left(\frac{\Gamma(\lambda)\Gamma(\lambda + 2/\gamma)}{\Gamma(\lambda + 1/\gamma)^2} - 1\right)^{-1}. \quad (41)$$

From (18)-(20) it follows that

$$I_\lambda = \lambda\gamma^2, \quad B_\lambda = \lambda\gamma^2(\lambda\gamma^2 + (\gamma-1)^2),$$

and consequently

$$\kappa = \lambda\gamma^2 + (\gamma-1)^2.$$

Let us write the sequential quasi credibility (35) for this family

$$\hat{\mu}_n = \hat{\mu}_{n-1} - \frac{\lambda}{n(\lambda\gamma^2 + (\gamma-1)^2) + c_{\lambda,\gamma}V_\lambda(X)/R_0}\gamma(\hat{\mu}_{n-1} - \mu_0(\lambda)^\gamma x_n^\gamma/\hat{\mu}_{n-1}^{\gamma-1})), \tag{42}$$

with c_λ given in (41). Notice that for gamma distributed claims, $\gamma = 1$, and (42) coincides with (38).

Example 1 *The Weibull family has $\gamma > 0$ and $\lambda = 1$ (see Table 1). Then*

$$\begin{aligned}
\kappa &= \gamma^2 + (\gamma-1)^2, \\
\mu_0(1) &= \Gamma(\frac{1}{\gamma}+1), \\
V_{\mu_0(1),1}(X) &= \Gamma(\frac{2}{\gamma}+1) - \Gamma(\frac{1}{\gamma}+1)^2 \\
c_{1,\gamma} &= c_\gamma = (\Gamma(\frac{2}{\gamma}+1)/\Gamma(\frac{1}{\gamma}+1)^2 - 1)^{-1}
\end{aligned}$$

and the sequential quasi credibility (42) takes the form

$$\hat{\mu}_n = \hat{\mu}_{n-1} - \frac{\gamma}{n(\gamma^2 + (\gamma-1)^2) + c_\gamma V_\gamma(X)/R_0}(\hat{\mu}_{n-1} - \Gamma(\frac{1}{\gamma}+1)^\gamma x_n^\gamma/\hat{\mu}_{n-1}^{\gamma-1})), \tag{43}$$

where $V_\gamma(X)$ is the expected process variance for given γ.

5.3 Lognormal claims.

Lognormal claims are SDF with $u(x) = -\frac{1}{2}\ln^2 x$. Then we straightforwardly obtain

$$\mu_0(\lambda) = \exp(1/(2\lambda)), \quad V_{\mu_0(\lambda),\lambda}(X) = \exp(1/\lambda)(\exp(1/\lambda) - 1)$$

and then from (33) we have

$$c_\lambda = (\exp(1/\lambda) - 1)^{-1}.$$

From (18)-(20) it follows that $I_\lambda = \lambda, B_\lambda = \lambda(1+\lambda)$, and then

$$\kappa = B_\lambda / I_\lambda = (1+\lambda).$$

To the end we can write from (35) the sequential quasi credibility formula

$$\begin{aligned}\hat{\mu}_n &= \hat{\mu}_{n-1} - \\ &\frac{\lambda}{n(\lambda+1) + V_\lambda(X)/((\exp(1/\lambda)-1)R_0)}(\ln \hat{\mu}_{n-1} - \ln x_n - 1/(2\lambda)).\end{aligned}$$

6 Numerical Example

Given 30 claims from Weibull-Gamma experiment: 39.757, 35.846, 48.797, 1420.9, 2836.6, 66.781, 949.8, 5156.9, 1492.6, 454.65, 309.63, 2458.8, 5762.4, 2547.7, 2313.4, 2027.2, 535.74, 70.119, 811.6, 4360.1, 1208.3, 3782.1, 3152.1, 2559.7, 127.93, 21997.0, 357.61, 1442.7, 140.05, 5283.3 on some scale, say \$s. In particular these data are the realization of Weibull distribution with $\gamma = 0.8$, where μ is being the realization of Gamma distribution with mean $m = 3000$, and variance $V(\mu) = 500^2 = R_0$. Then $\Gamma(1/0.8+1) = 1.133, c_\gamma V_\gamma(X)/R_0 = 37$ and the sequential quasi credibility formula (43) takes the form

$$\hat{\mu}_n = \hat{\mu}_{n-1} - \frac{0.8}{0.68n+37}(\hat{\mu}_{n-1} - 1.1051 x_n^{0.8}/\hat{\mu}_{n-1}^{-0.2})). \tag{44}$$

As $c_\gamma^{-1} = (\Gamma(\frac{2}{\gamma}+1)/\Gamma(\frac{1}{\gamma}+1)^2 - 1) = 1.5889, V_\gamma(X)/R_0 = 58.789$, and the linear sequential credibility formula is of the form (see (9))

$$\hat{\mu}_n = \hat{\mu}_{n-1} - \frac{1}{n+58.789}(\hat{\mu}_{n-1} - x_n) \tag{45}$$

For a comparison of the two credibility formulae we calculate the sample mean of the predicted errors subtracted by sample variance $S^2 = 1.65 \times 10^7$, the estimator of the fixed component of the mean squared predicted errors (see Landsman and Makov (1999a), Remark 2),

$$\frac{1}{29}\sum_{i=2}^{30}(x_i - \hat{\mu}_{i-1})^2 - S^2 = \begin{cases} 60225, \text{ for formula (44)} \\ 107838, \text{ for formula (45).}\end{cases}$$

and sample variances of the predicted values

$$S^2(\hat{\mu}_n) = \frac{1}{30}\sum_{i=1}^{30}(\hat{\mu}_i - \overline{\hat{\mu}})^2 = \begin{cases} 5228.22, \text{ for formula (44)} \\ 7301.76, \text{ for formula (45)} \end{cases},$$

where $\overline{\hat{\mu}} = 1/30\sum_{i=1}^{30}\hat{\mu}_i$. We can see that with respect to both measures the sequential quasi credibility formula is superior to the linear sequential credibility formula.

Acknowledgment

We are grateful to the Actuarial Education and Research Fund for their financial support and to the Project Oversight Group for their useful comments.

References

[1] Bühlmann, H. (1967). Experience rating and probability. *The Astin Bul.* 4, 199-207.

[2] Diaconis, P. and Ylvisaker, D. (1979). Conjugate priors for exponential families. *Annals of Statist.* **7** , 269-281

[3] Fergusson, T.S. (1962). Location and scale parameters in exponential families of distributions, *Annals of Math. Statist*, **33**, 986-1001.

[4] Gladyshev, E.G. (1965). On stochastic approximation. Theor. Prob. Appl., 10, 275-278.

[5] Herzog, T.N. (1994). *Introduction to Credibility Theory*, ACTEX Publications, Winsted.

[6] Jorgensen, B. (1983). Maximal likelihood estimation and large - sample inference for generalized linear and nonlinear regression models. *Biometrika,* **70**, 1, 19-28.

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

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

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

[10] Jorgensen, B. (1997). *The Theory of Dispersion Models.* Chapman and Hall, London

[11] Kaas, R., Dannenburg, D., Goovaerts, M. (1997). Exact Credibility for weighted observations. *Astin Bulletin*, **27**, 2, 287-295.

[12] Kushner, H. J. and Yin, G. G.(1997). Stochastic approximation algorithms and applications. Applications of Mathematics, **35**. Springer-Verlag, New York.

[13] Landsman, Z. and Makov, U. (1998). Exponential dispersion models and Credibility. *Scand. Actuarial J.*, 1, 89-96.

[14] Landsman, Z. and Makov, U. (1999a). On Stochastic Approximation and Credibility. *Scand. Actuarial J.,1, 15-31.*

[15] Landsman, Z. and Makov, U. (1999b). Sequential credibility evaluation for symmetric location claim distributions. *Insurance: Mathematics & Economics*, **24**, 291-300.

[16] Landsman, Z. and Makov, U. (1999c). Credibility evaluations for exponential dispersion families. *Insurance: Mathematics & Economics*, 1999, **24**, 33-39.

[17] Landsman, Z. and Makov, U. (2000). On Credibility Evaluation and the tail area of the Exponential Dispersion Family. To appear in: *Insurance: Mathematics & Economics.*

[18] Nelder J. A. and Wedderburn, R.W.M. (1972). Generalized linear models. *J. Roy. Statist. Soc. Ser. A* **135**, 370-384.

[19] Robbins, H. and Monro, S. (1951). A Stochastic approximation method. *Ann. Math. Stat.* **22**, 400-407.

[20] 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

[21] Young, V.,R. Credibility and persistency (1996). *Astin Bulletin*, **26**,1, 53-70.

[22] Zacks, S. (1971). *The Theory of Statistical Inference.* Wiley, New York.