PARAMETER UNCERTAINTY IN EXPONENTIAL FAMILY TAIL ESTIMATION

BY

Z. LANDSMAN AND A. TSANAKAS

ABSTRACT

Actuaries are often faced with the task of estimating tails of loss distributions from just a few observations. Thus estimates of tail probabilities (reinsurance prices) and percentiles (solvency capital requirements) are typically subject to substantial parameter uncertainty. We study the bias and MSE of estimators of tail probabilities and percentiles, with focus on 1-parameter exponential families. Using asymptotic arguments it is shown that tail estimates are subject to significant positive bias. Moreover, the use of bootstrap predictive distributions, which has been proposed in the actuarial literature as a way of addressing parameter uncertainty, is seen to double the estimation bias. A bias corrected estimator is thus proposed. It is then shown that the MSE of the MLE, the parametric bootstrap and the bias corrected estimators only differ in terms of order $O(n^{-2})$, which provides decision-makers with some flexibility as to which estimator to use. The accuracy of asymptotic methods, even for small samples, is demonstrated exactly for the exponential and related distributions, while other 1-parameter distributions are considered in a simulation study. We argue that the presence of positive bias may be desirable in solvency capital calculations, though not necessarily in pricing problems.

KEYWORDS

Reinsurance pricing, VaR, parameter uncertainty, bias, bootstrap, exponential families.

1. INTRODUCTION

Actuaries and other insurance risk modellers are often preoccupied by the potential of a portfolio to produce high losses. Hence the tails of loss distributions are of particular interest, for example, in the context of pricing high reinsurance layers or calculating solvency capital requirements.

Severe limitations in the size of available data sets mean that often tails of distributions are estimated from just a few hundreds or even tens of relevant data points. The result is a substantial potential for parameter error in tail estimates.

Astin Bulletin 42(1), 123-152. doi: 10.2143/AST.42.1.2160738

It is thus no surprise that parameter uncertainty has been a recurring theme in the actuarial community, both in academic and practitioner circles; see for example Cairns (2000), Mata (2000), Cummins and Lewis (2003), Powers et al (2003), Verrall and England (2006), Borowicz and Norman (2009), Richards (2009), Saltzmann and Wüthrich (2010), Gerrard and Tsanakas (2010).

The literature is fairly consistent in proposing that parameter uncertainty be reflected in risk calculations by the use of a predictive distribution, that is, a mixture of the loss distribution by a density of estimated parameters. This density of parameters may be obtained by a Bayesian posterior, leading to a Bayesian predictive distribution, or a bootstrap estimate of the sampling distribution, yielding a bootstrap predictive distribution. The rationale behind this approach is that predictive distributions tend to be more volatile than, say, Maximum Likelihood Estimators (MLE), and thus produce more conservative risk estimates. Thus, an implicit risk load for parameter uncertainty is produced.

Nonetheless, the performance of tail estimation procedures based on predictive distributions is usually not considered in relation to standard frequentist criteria such as the bias and Mean-Squared-Error (MSE). This is an issue worth considering; it has been shown by Smith (1998) that the simple MLE estimates of extreme tails often outperform estimates based on Bayesian prediction, when viewed though such a lense.

In Section 2 of the present contribution, we start our discussion with simple analytically tractable examples. We show that the MLE of single parameter exponential/Pareto tail probabilities is subject to significant positive bias, with the bias increasing as one moves further out into the tail. The same holds when considering the MLE of Pareto percentiles. This indicates that the simple MLE of tail functionals, before any predictive distribution is derived, is already in a sense conservative.

In order to generalise these arguments, in Section 3 asymptotic approximations are developed that allow the accurate calculation of the expected value of functions of sample means. Approximations are of "delta-type" and follow from a Taylor expansion around the sample mean and characterisation of the remainder term by combining Edgeworth and Laplace integral asymptotics.

Using these approximations, it is shown in Section 4 that the MLEs of extreme tail probabilities and percentiles in single-parameter exponential families will tend to be positively biased, thus generalising the insights of Section 2.

In Section 5 we turn our attention to the use of bootstrapping in tail estimation. It is shown that the parametric bootstrap estimator of tail probabilities and percentiles is indeed more conservative than the MLE. However, the price one pays for such conservativism is a bias that is double that of the MLE. Consequently, we propose an alternative estimator, which corrects the $O(1/n)$ term of the bias; this correction could be seen as an alternative use of the parametric bootstrap. We then show that the MSE of the three estimators considered differs only in terms of $O(1/n^2)$. Consequently we argue that bias correction is possible without a significant penalty in MSE. In a numerical example involving exponential tail functions, we show that the bias corrected estimator actually has a lower MSE than the others when considering the extreme tail.

In Section 6 we summarise our conclusions and further discuss the results obtained in the paper. In particular, we argue that the desirability of the estimation bias (and hence the choice of estimator) may depend on the application at hand, where positively biased estimators may be quite meaningful in the context of solvency capital calculation, but not necessarily in reinsurance pricing.

Throughout the paper, the performance of asymptotic approximations and estimators is demonstrated with reference to the exponential distribution, for which all quantities considered (e.g. bias and MSE of different tail-function estimators) can be analytically calculated. These calculations are documented in Appendix A. The stated results for the exponential distribution hold identically for distributions of random variables that can be written as increasing transforms of exponential variables. Thus, the cases of distributions such as the one-parameter Pareto (the large loss model most widely used in practice) and Weibull distributions are also implicitly dealt with. To establish further the applicability of our results, a simulation study is presented in Appendix B, considering one-parameter versions of the (log-)Normal, (log-)Gamma and Inverse Gaussian distributions.

2. Bias in tail estimation: two examples

Consider an i.i.d. sample of losses (e.g. insurance claims) $\mathbf{X} = (X_1, \ldots, X_n)$ with density $f(\cdot;\theta)$, where $\theta \in \Theta \subseteq \mathbb{R}$ is an unknown parameter to be estimated from $\mathbf{X}$. Denote by $\hat{\theta}$ the MLE of θ based on $\mathbf{X}$. Henceforth we will assume that $f(\cdot;\theta)$ is positive on $\mathbb{R}_+$ and that the corresponding distribution function $F(\cdot;\theta)$ is invertible.

Usually, rather than the parameter itself, a function of the parameter is of interest. For example, in (re)insurance pricing the tail probability $\bar{F}(x;\theta) = 1 - F(x;\theta)$ is of importance, since its integral over the interval $(d, d+l)$ gives the expected value of reinsurance layer of l in excess of d, $E[\min((X-d)_+, l)] = \int_d^{d+l} \bar{F}(x;0)\,dx$. Alternatively, in a solvency framework one is often interested in estimating the percentile $F^{-1}(p;\theta)$. If a portfolio faces a future loss $Y \sim f(\cdot;\theta)$, then $c = F^{-1}(p;\theta)$ corresponds to the level of capital that needs to be held in order to achieve a portfolio default probability of p (where default is narrowly defined as the event $\{Y > c\}$).

It hence becomes necessary for an insurer to estimate the extreme tail of the loss distribution, in order to be able to price a high layer or limit the default probability to an acceptable level. However in practice data sets, e.g. of insurance claims, can be very small, which leads to possibly large estimation errors. A substantial component of that error may be estimation bias. In the following two examples, we show that the two quantities of interest, the tail probability and the percentile, can be subject to significant positive bias.

Example 1 (Exponential/Pareto tail function). *Consider the case were we are interested in estimating the probability that an exponentially distributed random variable with mean θ exceeds threshold $y > 0$, that is, we seeking to estimate $\bar{F}(y;\theta) = e^{-y/\theta}$. Then the MLE of θ is $\hat{\theta} = \frac{1}{n}\sum_{j=1}^{n} X_j$ and the MLE of $\bar{F}(y;\theta)$ is $\bar{F}(y;\hat{\theta})$.*

$\hat{\theta}$ is unbiased, but $\bar{F}(y;\hat{\theta})$ is not. In fact the bias of $\bar{F}(y;\hat{\theta})$ can be explicitly calculated. Since $\hat{\theta} \sim Gam(n, n/\theta)$, we have

$$E\left[\bar{F}(y;\hat{\theta})\right] = \int_0^\infty e^{-y/t}\frac{t^{n-1}\exp(-tn/\theta)}{(\theta/n)^n\,\Gamma(n)}dt = \frac{2(ny/\theta)^{n/2}}{\Gamma(n)}K_n(2\sqrt{ny/\theta}), \quad (1)$$

where K_n is a modified Bessel function of the second kind (Gradshteyn and Ryzhik, 2007; eq. 3.471/8).

We note that in reinsurance pricing a more common loss model is based on the simple Pareto tail function $(b/y)^\alpha$, $y > b$. If a random variable Y follows an exponential distribution with mean θ, then $\tilde{Y} = be^Y$ follows a Pareto distribution with parameters $\alpha = 1/\theta$, b. Therefore for a Pareto distribution, the equation (1) will also hold, after substituting $\log(y/b)$ for y.

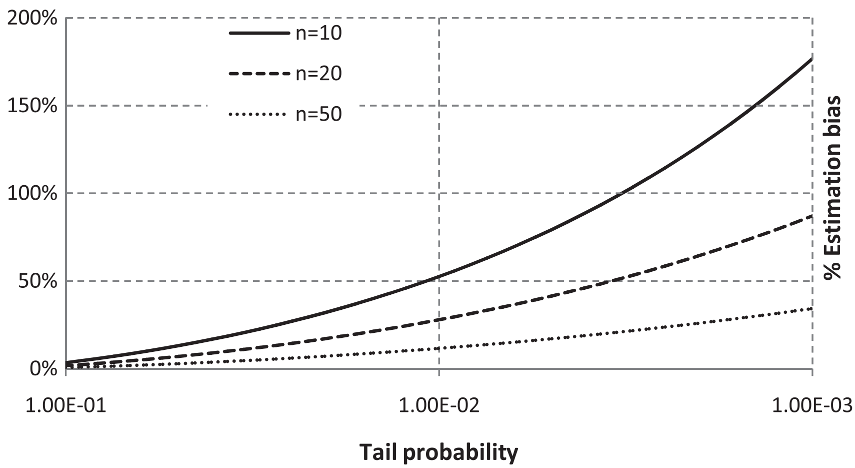

FIGURE 1: Relative bias of tail function estimate $E[\bar{F}(y;\hat{\theta})] / \bar{F}(y;\theta) - 1$ against true tail probability $\bar{F}(y;\theta)$ (inverted scale) for the exponential / Pareto model; sample sizes $n = 10, 20, 50$.

In figure 1, we plot the relative estimation bias $\bar{F}(y;\hat{\theta})/\bar{F}(y;\theta) - 1$ against values of the true tail probability $\bar{F}(y;\theta)$, for sample sizes $n = 10, 20, 50$. It can be seen that there is a substantial positive bias, particularly for low exceedance probabilities (high thresholds) and small sample sizes. Given that it is not uncommon to just have a few tens of samples from which to price a high layer, a relative bias of 28% for $n = 20$, $\bar{F}(y;\theta) = 0.01$ is striking.

Example 2 (Pareto percentiles). *Consider now the case that the distribution is a single-parameter Pareto with tail function* $\bar{F}(x;\theta) = (x/b)^{-1/\theta}$, $x > b$ *(where b is known) and that we are interested in estimating the percentile* $F^{-1}(p;\theta) = b(1-p)^{-\theta}$, *where p is close to* 1, *e.g.* $p = 0.995$, *as required by insurance regulation under the impending Solvency II regime. The MLE of* θ *is* $\hat{\theta} = \frac{1}{n}\sum_j \log(X_j/b)$, *which again follows a* $Gam(n, n/\theta)$ *distribution. Now the expected value of the percentile's MLE is:*

$$E\left[F^{-1}(p;\hat{\theta})\right] = bE\left[e^{-\hat{\theta}\log(1-p)}\right] = b\left[1 + \frac{\theta}{n}\log(1-p)\right]^{-n}. \tag{2}$$

In figures 2 and 3, the relative estimation bias for $F^{-1}(p;\hat{\theta})$ *is plotted, against the confidence level and the sample size respectively, for different values of the*

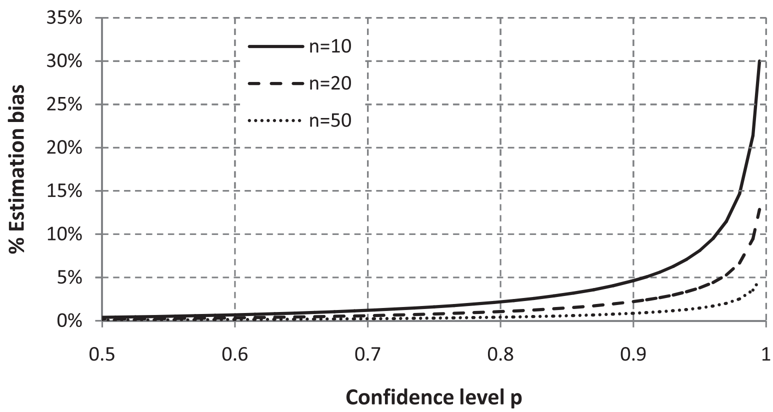

FIGURE 2: Relative bias of Pareto percentile estimate against confidence level p; $\alpha = 2.5$, $p \in [0.5, 0.995]$, $n = 10, 20, 50$.

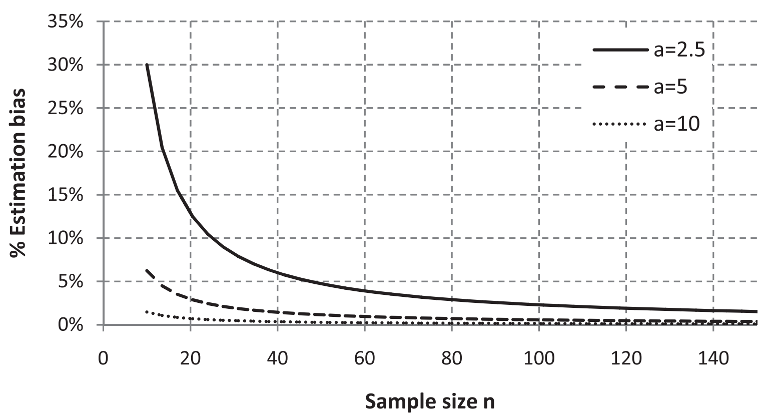

FIGURE 3: Relative bias of Pareto percentile estimate against sample size n; $p = 0.995$, $\alpha = 2.5, 5, 10$.

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press

parameter $\alpha = 1/\theta$. It can be seen that the bias increases dramatically for high confidence levels p, small samples sizes n, and low values of a corresponding to heavier tails.

Besides the issue of bias, for both examples discussed above, small data sizes will imply very substantial estimation errors; this is further discussed in section 5.3.

3. Asymptotic approximations

The analytically tractable examples of section 2 show that tail estimates may be very biased. In order to be able to treat more general cases, we develop in this section asymptotic formulas that can be used to approximate the expected value of non-linear functions of the sample mean. For distributions in a 1-parameter exponential family, such approximations allow calculation of estimation bias and Mean Squared Error to a high degree of accuracy.

First, in Lemma 1 we provide a result characterising the asymptotic behaviour of functions of the form $g(\hat{\mu})(\hat{\mu}-\mu)^k$. Subsequently, in Lemma 2, we derive the approximations that are used in this paper.

In the sequel we use the following standard asymptotic notation. Consider function $\zeta(x,n) : \mathcal{X} \in \mathbb{R} \times \mathbb{Z}_+ \mapsto \mathbb{R}$. We say that $\zeta(x,n) = O(\eta(n))$ as $n \to \infty$, uniformly in x, if for every $x \in \mathcal{X}$ there exist M, $n^* > 0$ such that for all $n > n^*$ it is $|\zeta(x,n)| \leq M|\eta(n)|$. For a family of random variables Z_n, we say that $Z_n = O_P(\eta(n))$ if for any $\varepsilon > 0$ there exists $M_\varepsilon > 0$ such that $P(|Z_n| > M_\varepsilon|\eta(n)|) < \varepsilon$.

Lemma 1. *Consider i.i.d. random variables $X_1, \dots, X_n$ with density $f(\cdot)$ and characteristic function $\varphi(\cdot)$. Denote by $\hat{\mu}$ the sample mean of $X_1, \dots, X_n$ and let $\mu = E(X_1)$, $\mu_i = \int_{-\infty}^{\infty} (x-\mu)^i f(x)dx$. $k \geq 3$ is an odd integer and $g(\cdot)$ a real-valued function. Assume that*

a) $\int_{-\infty}^{\infty} |\varphi(\zeta)|^v d\zeta < \infty$, *for some* $v \geq 1$.

b) $\mu_i < \infty$ *for* $i = 1, \dots, k+4$.

c) The function $g(w)$ has an infinite number of derivatives in some open interval containing $w = \mu$.

d) $\int_{-\infty}^{\infty} \left| g(w)(w-\mu)^k \right| dw < \infty$.

Then

$$E\left[g(\hat{\mu})(\hat{\mu}-\mu)^k\right] = O\left(n^{-\frac{k+1}{2}}\right) \tag{3}$$

as $n \to \infty$, uniformly in μ.

Proof. First write

$$E\left[g(\hat{\mu})(\hat{\mu}-\mu)^k\right] = \int_{-\infty}^{\infty} g(\mu + \sigma n^{-1/2} v)(\sigma n^{-1/2} v)^k f_{S^*}(v)dv,$$

where $f_{S^*}(\cdot)$ is the density of the standardised sample mean $S^* = n^{1/2}\sigma^{-1}(\hat{\mu} - \mu)$. The technical conditions allow an Edgeworth expansion of $f_{S^*}(\cdot)$ and in particular as $n \to \infty$ it is (Feller, 1966; p. 535)

$$f_{S^*}(v) = \phi(v) + \phi(v)\sum_{j=3}^{k+3} n^{-j/2+1}P_j(v) + O\left(n^{-k/2-1}\right)$$

uniformly in v. ϕ is the standard normal density and P_j are polynomials not depending on n or k. The exact form of the polynomials is not of interest as will be explained below. Hence there exists $\lambda > 0$ such that:

$$E\left[g(\hat{\mu})(\hat{\mu} - \mu)^k\right] \leq I_1 + I_2$$

where

$$I_1 = \int_{-\infty}^{\infty} g(\mu + \sigma n^{-1/2}v)(\sigma n^{-1/2}v)^k\left[\phi(v) + \phi(v)\sum_{j=3}^{k+3} n^{-j/2+1}P_j(v)\right]dv$$

$$I_2 = \int_{-\infty}^{\infty}\left|g(\mu + \sigma n^{-1/2}v)(\sigma n^{-1/2}v)^k\right|\lambda n^{-k/2-1}dv.$$

Define integrals of the form:

$$h(k,r) = \int_{-\infty}^{\infty} g(\mu + \sigma n^{-1/2}v)(\sigma n^{-1/2}v)^k v^r\phi(v)\,dv$$

Then we can write I_1 as:

$$I_1 = h(k,0) + \sum_m b_m n^{-c_m} h(k, r_m),$$

where $c_m \geq 1/2$. The above formula derives from observing that the sum of polynomials P_j in the Edgeworth series will produce terms including powers of v and n. In particular by studying integrals of the form $h(k,r)$, we will see that the order of $h(k,r)$ depends on whether $k + r$ is even or odd, but not on the actual value of r. Hence the precise values of the constants b_m, c_m, r_m are not of interest.

We now examine the asymptotics of the integrals $h(k,r)$. By the change of variable $v = (2n)^{1/2}x$ we obtain

$$h(k,r) = \int_{-\infty}^{\infty} g(\mu + \sigma n^{-1/2}v)(\sigma n^{-1/2}v)^k v^r(2\pi)^{-1/2}e^{-v^2/2}dv$$
$$= \sigma^k n^{(1+r)/2}2^{(k+r)/2}\pi^{-1/2}\int_{-\infty}^{\infty} g(\mu + \sigma 2^{1/2}x)x^{k+r}e^{-nx^2}dx.$$

The last integral admits an asymptotic expansion by a modification of Watson's lemma (see e.g. Murray 1974; pp. 24-26):

$$\int_{-\infty}^{\infty} g(\mu+\sigma 2^{1/2}x)x^{k+r}e^{-nx^2}dx \sim \pi^{1/2}n^{-1/2}\left\{a_0+\frac{a_2}{2n}+\frac{1\cdot 3\cdot a_4}{2^2n^2}+\ldots\right\}$$
$$= \pi^{1/2}n^{-1/2}\sum_{j=0}^{\infty} a_{2j}(2j-1)!!(2n)^{-j}.$$

where the a_{2j} are defined by the Taylor expansion

$$g(\mu+\sigma 2^{1/2}x)x^{k+r} = a_0 + a_1 x + a_2 x^2 + \ldots.$$

This yields

$$h(k,r) \sim \sigma^k n^{r/2} 2^{(k+r)/2} \sum_{j=0}^{\infty} a_{2j}(2j-1)!!(2n)^{-j},$$

Expanding the function $x \mapsto g(\mu+\sigma 2^{1/2}x)$ around 0 gives

$$g(\mu+\sigma 2^{1/2}x)x^{k+r} = g(\mu)x^{k+r} + g^{(1)}\sigma 2^{1/2}(\mu)x^{k+r+1}$$
$$+\frac{1}{2!}g^{(2)}(\mu)\sigma^2 2x^{k+r+2}+\ldots$$

We now distinguish between two cases. First consider the case that $k+r$ is even. Then the first non-zero term of even order in the expansion of $g(\mu+\sigma 2^{1/2}x)$ x^{k+r} corresponds to $a_{k+r} = g(\mu)$. Hence

$$h(k,r) \sim \sigma^k n^{r/2} 2^{(k+r)/2} g(\mu)(k+r-1)!!(2n)^{-(k+r)/2} = \sigma^k g(\mu)(k+r-1)!!\,n^{-k/2}$$

On the other hand, if $k+r$ is odd, the first non-zero term of even order in the expansion of $g(\mu+\sigma 2^{1/2}x)x^{k+r}$ corresponds to $a_{k+r+1} = g^{(1)}(\mu)\sigma 2^{1/2}$. Thus

$$h(k,r) \sim \sigma^k n^{r/2} 2^{(k+r)/2} g^{(1)}(\mu)\sigma 2^{1/2}(k+r)!!(2n)^{-(k+r+1)/2}$$
$$= \sigma^{k+1} g^{(1)}(\mu)(k+r)!!\,n^{-(k+1)/2}$$

Consequently, for any k, r, it is $h(k,r) = O(n^{-k/2})$, but for the case that $k+r$ is odd, it is $h(k,r) = O(n^{-(k+1)/2})$. In particular, since k is odd $h(k,0) = O(n^{-(k+1)/2})$. Regarding the other terms in the integral I_1 we have that, since $c_m \geq 1/2$,

$$n^{-c_m}h(k,r_m) = O\left(n^{-k/2-c_m}\right) \Longrightarrow n^{-c_m}h(k,r_m) = O\left(n^{-k/2-1/2}\right).$$

Therefore we conclude that $I_1 = O(n^{-(k+1)/2})$.

We now turn our attention to integral I_2. By change of variable $w = \mu + \sigma n^{-1/2}v$ we have

$$\begin{aligned} I_2 &= \lambda n^{-k/2-1}\int_{-\infty}^{\infty}\left|g(\mu+\sigma n^{-1/2}v)(\sigma n^{-1/2}v)^k\right|dv \\ &= \lambda n^{-k/2-1}\int_{-\infty}^{\infty}\left|g(w)(w-\mu)^k\right|\sigma^{-1}n^{1/2}dw \\ &= O\left(n^{-(k+1)/2}\right) \end{aligned}$$

Since both I_1, I_2 are $O(n^{-(k+1)/2})$, so is $E[g(\hat{\mu})(\hat{\mu}-\mu)^k] \le I_1 + I_2$. □

Lemma 2. *Let $X_1, \ldots, X_n$ be as in Lemma 1, with assumptions a) and b) satisfied. Consider function $\psi(\cdot)$ with a continuous k^{th} derivative, where k is an odd integer. Assume that:*

$$\int_{-\infty}^{\infty}\left|\int_{\mu}^{w}\psi^{(k)}(x)(w-x)^{k-1}dx\right|dw < \infty$$

Then the following approximations hold as $n \to \infty$, uniformly in μ:

i) For $k = 3$,

$$E[\psi(\hat{\mu})] = \psi(\mu) + \frac{1}{2}\psi''(\mu)\frac{\mu_2}{n} + O\left(n^{-2}\right) \tag{4}$$

ii) For $k = 5$,

$$E[\psi(\hat{\mu})] = \psi(\mu) + \frac{1}{2}\psi''(\mu)\frac{\mu_2}{n} + \frac{1}{6}\psi^{(3)}(\mu)\frac{\mu_3}{n^2} + \frac{1}{8}\psi^{(4)}(\mu)\frac{\mu_2^2}{n^2} + O\left(n^{-3}\right) \tag{5}$$

Proof. Taylor expansions of $\psi(\hat{\mu})$ around μ give

$$\psi(\hat{\mu}) = \psi(\mu) + \psi'(\mu)(\hat{\mu}-\mu) + \frac{1}{2}\psi''(\mu)(\hat{\mu}-\mu)^2 + \frac{1}{6}A_3(\hat{\mu}),$$

$$\begin{aligned} \psi(\hat{\mu}) = {} & \psi(\mu) + \psi'(\mu)(\hat{\mu}-\mu) + \frac{1}{2}\psi''(\mu)(\hat{\mu}-\mu)^2 + \frac{1}{6}\psi^{(3)}(\mu)(\hat{\mu}-\mu)^3 \\ & + \frac{1}{24}\psi^{(4)}(\mu)(\hat{\mu}-\mu)^4 + \frac{1}{120}A_5(\hat{\mu}), \end{aligned}$$

where

$$A_k(\hat{\mu}) = \int_{\mu}^{\hat{\mu}} k\psi^{(k)}(x)(\hat{\mu}-x)^{k-1}dx = \psi^{(k)}(\mu^*)(\hat{\mu}-\mu)^k,$$

for $\mu^* = \mu + \alpha_{\hat{\mu}}(\hat{\mu}-\mu)$, $\alpha \in [0,1]$. Hence we can write the expected value of $\psi(\hat{\mu})$ as

$$E[\psi(\hat{\mu})] = \psi(\mu) + \frac{1}{2}\psi''(\mu)\frac{\sigma^2}{n} + \frac{1}{6}E[A_3(\hat{\mu})],$$

$$E[\psi(\hat{\mu})] = \psi(\mu) + \frac{1}{2}\psi''(\mu)\frac{\sigma^2}{n} + \frac{1}{6}\psi^{(3)}(\mu)\frac{\mu_3}{n^2} + \frac{1}{24}\psi^{(4)}(\mu)\frac{\kappa_4 + 3n\sigma^4}{n^3} + \frac{1}{120}E[A_5(\hat{\mu})],$$

where κ_4 is the fourth cumulant of X_1.

Denote $\psi^{(k)}(\mu + \alpha_{\hat{\mu}}(\hat{\mu} - \mu)) = g(\hat{\mu})$. Then the order of $E[A_k(\hat{\mu})]$ follows from Lemma 1 subject to differentiability and absolute integrability of $g(\cdot)$, which we discuss at the end of the proof. For odd k it is:

$$E[A_k(\hat{\mu})] = E\left[g(\hat{\mu})(\hat{\mu} - \mu)^k\right] = O\left(n^{-(k+1)/2}\right)$$

By setting $k = 3, 5$, the approximations (4), (5) follow.

To establish the required technical conditions on $g(\cdot)$, note its alternative form:

$$g(\hat{\mu}) = \frac{A_k(\hat{\mu})}{(\hat{\mu} - \mu)^k} = k(\hat{\mu} - \mu)^{-k} \int_{\mu}^{\hat{\mu}} \psi^{(k)}(x)(\hat{\mu} - x)^{k-1} dx$$

If $\psi^{(k)}(x)$ a continuous function in its domain, from the above expression it follows that $g(\hat{\mu})$ is infinitely differentiable. Absolute integrability of $g(\cdot)$ follows from

$$\begin{aligned} \int_{-\infty}^{\infty} \left| g(w)(w - \mu)^k \right| dw &= \int_{-\infty}^{\infty} \left| \psi^{(k)}(\mu + \alpha_w(w - \mu))(w - \mu)^k \right| dw \\ &= \int_{-\infty}^{\infty} \left| k \int_{\mu}^{w} \psi^{(k)}(x)(w - x)^{k-1} dx \right| dw \\ &< \infty \end{aligned}$$

□

The accuracy of the approximations is demonstrated for the tail function of an exponential distribution, with $\psi(\mu) = e^{-y/\mu}$.

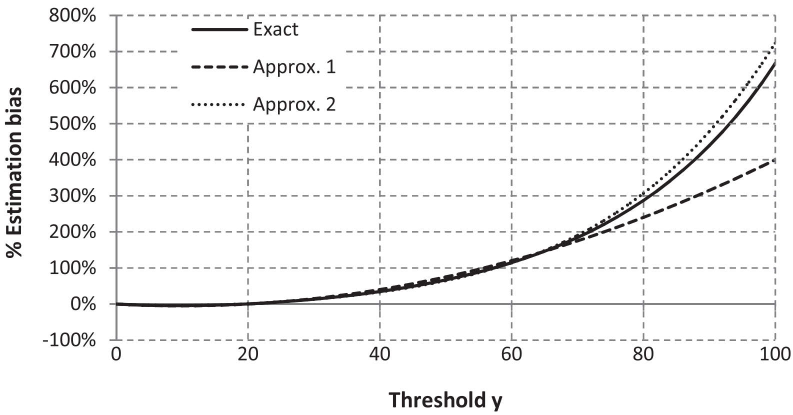

FIGURE 4: Relative bias for exponential tail function estimate against threshold y using exact formulas and approximations with error terms $O(n^{-2})$ (Approx. 1) and $O(n^{-3})$ (Approx. 2); $n = 10$, $\mu = 10$.

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press

In figure 4 the relative bias of $\psi(\hat{\mu})$ is plotted against the threshold y for $n = 10$, $\mu = 10$, using the exact formula (1), and the approximations (4), (5). It is seen that even for such a small sample, the $O(n^{-2})$ approximation performs quite well for values of y until about 70, corresponding to a tail probability $\psi(\mu)$ of approximately 0.001. The $O(n^{-3})$ approximation performs well even for much higher thresholds.

TABLE 1

EXPECTED VALUE OF ESTIMATOR OF EXPONENTIAL TAIL FUNCTION $\psi(\mu) = e^{-y/\mu}$, CALCULATED USING EXACT FORMULAS AND APPROXIMATIONS WITH ERROR TERMS $O(n^{-2})$ (Approx. 1) AND $O(n^{-3})$ (Approx. 2); $\mu = 10$.

a) $y = 30$, $\psi(\mu) \cong 0.05$

n	Exact	Approx. 1	Error (%)	Approx. 2	Error (%)
5	0.0615	0.0647	5.278	0.0610	0.796
6	0.0599	0.0622	3.884	0.0596	0.445
7	0.0587	0.0605	2.978	0.0586	0.267
8	0.0578	0.0591	2.355	0.0577	0.170
9	0.0570	0.0581	1.909	0.0569	0.113
10	0.0564	0.0573	1.578	0.0563	0.078
15	0.0544	0.0548	0.746	0.0544	0.017
20	0.0533	0.0535	0.432	0.0533	0.006
30	0.0522	0.0523	0.198	0.0522	0.001
50	0.0512	0.0513	0.073	0.0512	0.000

b) $y = 46$, $\psi(\mu) \cong 0.01$

n	Exact	Approx. 1	Error (%)	Approx. 2	Error (%)
5	0.0196	0.0221	12.841	0.0183	6.254
6	0.0182	0.0201	10.065	0.0175	4.161
7	0.0172	0.0186	8.134	0.0167	2.923
8	0.0165	0.0176	6.727	0.0161	2.138
9	0.0158	0.0167	5.667	0.0156	1.615
10	0.0153	0.0161	4.844	0.0151	1.251
15	0.0137	0.0141	2.575	0.0136	0.453
20	0.0129	0.0131	1.603	0.0128	0.214
30	0.0120	0.0121	0.796	0.0120	0.072
50	0.0112	0.0113	0.316	0.0112	0.017

In Table 1, for the cases $y = 30$, $\psi(\mu) \cong 0.05$ and $y = 46$, $\psi(\mu) \cong 0.01$, the exact value of $E[\psi(\hat{\mu})]$ is given along with the two approximations and the corresponding approximation errors, for sample sizes from $n = 5$ to $n = 50$. If $\{E[\psi(\hat{\mu})]\}_{appr}$ is an approximation to $E[\psi(\hat{\mu})]$, then the error stated is given by $100\left|\frac{\{E[\psi(\hat{\mu})]\}_{appr}}{E[\psi(\hat{\mu})]}-1\right|\%$. Again the good performance of the approximate formulas can be observed. For example, when considering the high threshold $y = 46$ and with $n = 15$, the approximation error of the approximation with $O(n^{-2})$ error term is 2.575%, while the error of the approximation with $O(n^{-3})$ error is 0.453%.

4. Exponential families

In section 2 it was shown that the estimates of extreme tail probabilities and percentiles may be subject to significant positive bias. Here, using the approximations of section 3, we extend this argument by showing that this is a general property of single-parameter exponential families, of which the exponential distribution is a particular case.

We start with Natural Exponential Families (NEF) with density of the form

$$f(x;\theta) = h(x)e^{\theta x-\kappa(\theta)} \tag{6}$$

Consider a random variable $Y \sim f(\cdot;\theta)$ and denote $\mu = E(Y) = \kappa'(\theta)$, $\sigma^2 = Var(Y) = \kappa''(\theta)$, $\mu_3 = E[(Y-\mu)^3] = \kappa^{(3)}(\theta)$. Throughout this section, we will restrict ourselves to non-negative random variables with an infinite right tail.

The MLE for parameter μ is just the sample mean $\hat{\mu} = \bar{X}$ and hence the MLE for any parameter of the form $\psi(\mu)$ will be $\psi(\hat{\mu})$. Henceforth we will write any parameter of interest in the form $\psi(\mu)$, including $\theta := \theta(\mu) = (\kappa')^{-1}(\mu)$. We denote 2nd and 3rd central moments as functions of μ by $V(\mu) = \kappa''\theta(\mu))$ and $\gamma(\mu) = \kappa^{(3)}(\theta(\mu))$ respectively. It is then easily shown that $\theta'(\mu) = V(\mu)^{-1}$ and $\theta''(\mu) = -\gamma(\mu)V(\mu)^{-3}$.

Equation (4) allows us to characterise the bias of $\psi(\hat{\mu})$; in particular it shows that:

$$Bias(\psi(\hat{\mu})) = E[\psi(\hat{\mu})] - \psi(\mu) \approx \frac{1}{2}\psi''(\mu)\,\frac{\sigma^2}{n} \tag{7}$$

Hence convexity of the function $\psi(\cdot)$ at μ implies that the bias is positive. We now establish increasingness and convexity of tail probabilities and percentiles, as functions of μ.

Lemma 3. *Define the function* $g(m,y) = \bar{F}(y;\theta(m)) = \int_y^\infty f(x;\theta(m))\,dx$. *Let* $g_\mu(m,y) = \frac{\partial g(m,y)}{\partial m}$ *and* $g_{\mu\mu}(m,y) = \frac{\partial^2 g(m,y)}{\partial^2 m}$. *Then:*

i) For $y > \mu$ *it is*

$$g_\mu(\mu,y) > 0. \tag{8}$$

ii) There exists $y^*(\mu) > 0$ *such that for* $y > y^*(\mu)$ *it is*

$$g_{\mu\mu}(\mu,y) > 0. \tag{9}$$

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press

Proof. Part i): Differentiation with respect to μ yields

$$\begin{aligned} g_\mu(\mu, y) &= \int_y^\infty h(x) e^{\theta(\mu)x - \kappa(\theta(\mu))} \theta'(\mu)(x - \kappa'(\theta(\mu)))\,dx \\ &= \theta'(\mu) \int_y^\infty f(x; \theta(\mu))(x - \kappa'(\theta(\mu)))\,dx, \end{aligned}$$

from which it follows that, since $\theta'(\mu) = V(\mu)^{-1} > 0$ it is $g_\mu(\mu, y) > 0$ as long as $y > \kappa'(\theta(\mu)) = \mu$.

Part ii): Differentiating $g_\mu(\mu, y)$ with respect to its first argument we obtain:

$$\begin{aligned} g_{\mu\mu}(\mu, y) &= \theta'(\mu) \left\{ \frac{\theta''(\mu)}{\theta'(\mu)} \int_y^\infty f(x; \theta(\mu))(x - \kappa'(\theta(\mu)))\,dx \right. \\ &\quad + \theta'(\mu) \int_y^\infty f(x; \theta(\mu))(x - \kappa'(\theta(\mu)))^2\,dx \\ &\quad \left. - \int_y^\infty f(x; \theta(\mu))\,dx \right\} \end{aligned}$$

In view of the expressions for the derivatives of κ, θ that were given earlier, we can write the above equations as:

$$\begin{aligned} g_{\mu\mu}(\mu, y) &= \frac{1}{\sigma^2} \left\{ -\frac{\mu_3}{\sigma^4} E[(X-\mu)\mathbf{1}_{X>y}] + \frac{1}{\sigma^2} E[(X-\mu)^2 \mathbf{1}_{X>y}] - P(X>y) \right\} \\ &= \frac{1}{\sigma^4} P(X>y) \left\{ [E(X|X>y)-\mu]^2 - \frac{\mu_3}{\sigma^2}[E(X|X>y)-\mu] - \sigma^2 + Var(X|X>y) \right\}, \end{aligned}$$

since $E[(X-\mu)^2 | X > y] = Var(X|X > y) + [E(X|X > y) - \mu]^2$. As y increases, $[E(X|X > y) - \mu]$ tends to $+\infty$, but $[E(X|X > y) - \mu]^2$ increases to ∞ faster, which makes the expression above positive for large enough y. □

Hence, if we let $\psi(\mu) = g(\mu, y)$, for large enough y it will be $\psi''(\mu) > 0$. Therefore the MLEs of extreme tail probabilities will tend to be positively biased. This property of natural exponential families can be slightly generalised as follows.

Corollary 1. *Let $Y \sim f(\cdot; \theta(\mu))$, $\tilde{Y} = t(Y)$, with t a strictly increasing function. Then the tail probability $P_\theta(\tilde{Y} > \tilde{y}) = g(\mu, t^{-1}(\tilde{y}))$ is convex in μ for $\tilde{y} > t(y^*)$, where y^* is as in Lemma 3.*

This means that our discussion does not only involve distributions such as the exponential, but also distributions obtained by increasing transforms, such as the Pareto, since if $Y \sim Exp(1/\mu)$, $\tilde{Y} = b\exp(Y) \sim Pareto(1/\mu, b)$.

We now turn our attention to the case where percentiles are of interest.

Lemma 4. *Define the function* $q(m,p) = F^{-1}(p;\theta(m))$. *Let* $q_\mu(m,p) = \frac{\partial q(m,p)}{\partial m}$ *and* $q_{\mu\mu}(m,p) = \frac{\partial^2 q(m,p)}{\partial^2 m}$. *Then:*

i) For $p > F(\mu;\theta(\mu))$ *it is*

$$q_\mu(\mu,p) > 0. \tag{10}$$

ii) There exists $p^*(\mu) \in (0,1)$ *such that for* $p > p^*(\mu)$ *it is*

$$q_{\mu\mu}(\mu,p) > 0. \tag{11}$$

Proof. Part i): By the definition of the functions g, q, it is $g(\mu, q(\mu,p)) = 1 - p$. Taking the total derivative wrt μ yields:

$$g_\mu(\mu,q(\mu,p)) + g_y(\mu,q(\mu,p))\, q_\mu(\mu,p) = 0 \Longrightarrow q_\mu(\mu,p) = \frac{g_\mu(\mu,q(\mu,p))}{-\, g_y(\mu,q(\mu,p))}$$

The denominator is just the density $f(q(\mu,p);\theta(\mu))$, therefore positive. For $p > F(\mu;\theta(\mu))$ it is $q(\mu,p) > \mu$ and therefore by Lemma 3i) the numerator is also positive.

Part ii): Taking the second total derivative of $g(\mu, q(\mu,p)) = 1 - p$ wrt to μ yields the equation

$$g_{\mu\mu} + g_{\mu y} q_\mu + \left(g_{y\mu} + g_{yy} q_\mu\right) q_\mu + g_y\, q_{\mu\mu} = 0 \Longrightarrow$$
$$q_{\mu\mu} = (-g_y)^{-1}\left(g_{yy} q_\mu^2 + 2 g_{y\mu} q_\mu + g_{\mu\mu}\right),$$

where $g_{y\mu} = \frac{\partial^2 g}{\partial \mu \partial y}$ and the functions' arguments have been suppressed. Now observe the following

- $-\, g_y(\mu, q(\mu,p)) = f(q(\mu,p);\, \mu) > 0$ for $p > 0$.
- $-\, g_{yy}(\mu, q(\mu,p))$ is the first derivative of the density. For large enough p, by the assumption of an infinite right tail, the density will be decreasing and thus $g_{yy}(\mu, q(\mu,p)) > 0$.
- By differentiating the density with respect to μ it is easily obtained that $g_{y\mu}(\mu, q(\mu,p)) = -V(\mu)^{-1} f(q(\mu,p);\, \mu)\, (q(\mu,p) - \mu)$. By finiteness of the mean, it is $g_{y\mu}(\mu, q(\mu,p)) \to 0$ as $p \to 1$.
- By part i) of the Lemma, for large enough p it is $q_\mu(\mu,p) > 0$.
- By the proof of Lemma 3ii), $g_{\mu\mu}(\mu, q(\mu,p))$ can be made arbitrarily large with increasing p.

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press

From the above observations it follows that that a p large enough can be found such that $q_{\mu\mu}(\mu, p) > 0$. □

Hence MLEs of extreme percentiles will also be convex in μ and hence positively biased. Again we can move slightly beyond natural exponential families.

Corollary 2. *Let $Y \sim f(\cdot; \theta(\mu))$, $\tilde{Y} = t(Y)$, with t a strictly increasing and convex function. Then the p^{th} percentile of $\tilde{Y}$, $t(h(\mu,p))$, is convex in μ for $p > p^*$, where y^* is as in Lemma 4.*

5. Bootstrapping

5.1. Bootstrap-predictive distribution

A method often proposed in order to address the issue of estimation error and associated parameter uncertainty, both in pricing and in solvency applications, is to use a predictive distribution, rather than the 'estimative' one derived from MLE. Predictive distributions arise as mixtures of distributions over distributions of parameters, which may be derived by Bayesian arguments (e.g. Cairns (2000), Verrall and England (2006), Saltzmann and Wüthrich (2010)) or as (bootstrap approximations to) sampling distributions of MLEs (Harris (1989), Mata (2000), Verrall and England (2006)).

Staying within the framework of 1-parameter exponential families, consider again a parameter of interest that can be written as $\psi(\mu)$. Then the parametric bootstrap estimator (PBE) of $\psi(\mu)$ is given by:

$$\psi_{PB}(\hat{\mu}) = \int_0^\infty \psi(m) f_\mu(m; \hat{\mu})\, dm, \tag{12}$$

where $f_{\hat{\mu}}(\cdot; \mu)$ is the density of the sample mean, when the true mean is μ. The link with bootstrapping is established by considering the evaluation of integral (12) via Monte-Carlo simulation.

If the parameter of interest is a tail probability, $\psi(\mu) = \bar{F}(y; \theta(\mu))$, then the function $y \mapsto \psi_{PB}(\hat{\mu})$ is called a bootstrap-predictive tail function. Note that the integral in (12) is formally identical with $E[\psi(\hat{\mu})] = \int_0^\infty \psi(m)\, f_{\hat{\mu}}(m; \mu)\, dm$, with the only difference that in (12) the estimated rather then the true value of the mean is used to evaluate the density of the sample mean. Hence the approximations of Section 3 can be used to evaluate $\psi_{PB}(\hat{\mu})$, yielding

$$\psi_{PB}(\hat{\mu}) = \psi(\hat{\mu}) + \frac{1}{2}\psi''(\hat{\mu})\frac{V(\hat{\mu})}{n} + \frac{1}{6}\psi^{(3)}\frac{\gamma(\hat{\mu})}{n^2} + \frac{1}{8}\psi^{(4)}\frac{V(\hat{\mu})^2}{n^2} + O_P(n^{-3}). \tag{13}$$

In the sequel we will denote the approximation arising from keeping terms up to order n^{-1} by

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press

$$\psi^*(\hat{\mu}) := \psi(\hat{\mu}) + \frac{1}{2}\psi''(\hat{\mu})\frac{V(\hat{\mu})}{n} \tag{14}$$

It is noted that an approximate formula essentially identical to (14) has been obtained by Landsman (2004), in the context of Bayesian estimation.

The above equations imply that bootstrap predictive distributions can be evaluated via simple analytical approximations, without the need to use simulation methods. It furthermore reveals that if the function $\psi(\cdot)$ is convex at μ, as is the case for tail or percentile functions of exponential families, then the PBE tends to be higher than the MLE. Hence, the use of the bootstrap predictive distributions, e.g in pricing applications, is indeed more conservative than just using the MLE. On the other hand, the following Lemma shows also that using the PBE will approximately double the bias in comparison with the MLE.

Lemma 5. *For $\psi^*(\hat{\mu})$ as defined in (14) and assuming the relevant conditions of Lemma 2 fulfilled, it is*

$$E[\psi^*(\hat{\mu})] = \psi(\mu) + \psi''(\mu)\frac{V(\mu)}{n} + O(n^{-2}) \tag{15}$$

Proof. $\psi^*(\hat{\mu})$ can be viewed as a function of $\hat{\mu}$, and thus the approximation (4) can again be used, but now considering the function $\psi^*(\cdot)$ rather than $\psi(\cdot)$. Therefore

$$E[\psi^*(\hat{\mu})] = \psi^*(\mu) + \frac{1}{2}\psi^{*\prime\prime}(\mu)\frac{V(\mu)}{n} + O(n^{-2})$$

It is

$$\psi^*(\mu) = \psi(\mu) + \frac{1}{2}\psi''(\mu)\frac{V(\mu)}{n} \implies \psi^{*\prime\prime}(\mu) = \psi''(\mu) + O(n^{-1})$$

Putting the above expressions together yields:

$$E[\psi^*(\hat{\mu})] = \left[\psi(\mu) + \frac{1}{2}\psi''(\mu)\frac{V(\mu)}{n}\right] + \frac{1}{2}\left[\psi''(\mu) + O(n^{-1})\right]\frac{V(\mu)}{n} + O(n^{-2})$$
$$= \psi(\mu) + \psi''(\mu)\frac{V(\mu)}{n} + O(n^{-2})$$

□

We note that the result of Lemma 5 does not change if an approximation to ψ_{PB} including more terms is used. More accurate approximations for the bias of ψ^*, are given in Appendix A.

5.2. Bias-corrected estimators

Both the MLE $\psi(\hat{\mu})$ and the (approximate) PBE $\psi^*(\hat{\mu})$ are biased. In fact, bootstrapping procedures can be used for correcting such bias (e.g. Hall,

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press

1997). The bootstrap estimate for the bias of the MLE $\psi(\hat{\mu})$ is $\psi_{PB}(\hat{\mu}) - \psi(\hat{\mu})$. Using the approximation $\psi_{PB}(\hat{\mu}) - \psi(\hat{\mu}) \approx \psi^*(\hat{\mu}) - \psi(\hat{\mu}) = \frac{1}{2}\psi''(\hat{\mu})\frac{V(\hat{\mu})}{n}$, we consider the bias-corrected estimator (BCE) $\bar{\psi}(\hat{\mu})$:

$$\bar{\psi}(\hat{\mu}) := \psi(\hat{\mu}) - \frac{1}{2}\psi''(\hat{\mu})\frac{V(\hat{\mu})}{n} \tag{16}$$

This type of bias correction by an estimate of the $O(n^{-1})$ bias term is in essence the one suggested by Cox and Hinkley (1979; Sec 8.4).

The effectiveness of $\bar{\psi}(\hat{\mu})$ in correcting for the bias is shown via the following lemma, whose proof is very similar to that of Lemma 5 and therefore omitted.

Lemma 6. *For $\bar{\psi}(\hat{\mu})$ as defined in (16) and assuming the relevant conditions of Lemma 2 fulfilled, it is*

$$E[\bar{\psi}(\hat{\mu})] = \psi(\mu) + O(n^{-2}) \tag{17}$$

More accurate approximations for the bias of $\bar{\psi}$ are given in Appendix A.

5.3. Mean Squared Errors

So far, three estimators were considered, the MLE $\psi(\hat{\mu})$, the approximate PBE $\psi^*(\hat{\mu})$ and the BCE $\bar{\psi}(\hat{\mu})$. It was shown that the three have different levels of bias, with the PBE approximately doubling the bias of the MLE and the BCE approximately eliminating it. However to effectively compare the three estimators we need to consider their estimation accuracy. We do this by deriving an approximation for the Mean-Squared-Errors (MSE) of the three estimators.

The following lemma shows that three estimators considered have all approximately the same Mean Squared Error.

Lemma 7. *Assuming the relevant conditions of Lemma 2 fulfilled, it is*

$$MSE(\psi(\hat{\mu})) = \left[\psi'(\mu)\right]^2 \frac{V(\mu)}{n} + O(n^{-2}) \tag{18}$$

$$MSE(\psi^*(\hat{\mu})) = \left[\psi'(\mu)\right]^2 \frac{V(\mu)}{n} + O(n^{-2}) \tag{19}$$

$$MSE(\bar{\psi}(\hat{\mu})) = \left[\psi'(\mu)\right]^2 \frac{V(\mu)}{n} + O(n^{-2}) \tag{20}$$

Proof. All three estimators can be written in the form $\psi(\hat{\mu}) + \frac{a}{2}\psi''(\hat{\mu})\frac{V(\hat{\mu})}{n}$. Consider function

$$v(m) = \left(\psi(m) + \frac{a}{2}\psi''(m)\frac{V(m)}{n} - \psi(\mu)\right)^2$$

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press

Then the MSE of an estimator of the form $\psi(\hat{\mu}) + \frac{a}{2}\,\psi''(\hat{\mu})\,\frac{V(\hat{\mu})}{n}$ can be written as $E[v(\hat{\mu})]$. Note that

$$v(\mu) = \left[\frac{a}{2}\psi''(\mu)\,\frac{V(\mu)}{n}\right]^2 = O(n^{-2}).$$

Differentiation of v yields

$$v''(m) = 2\big[\psi'(m)\big]^2 + 2(\psi(m) - \psi(\mu))\,\psi''(m) + O(n^{-1})$$
$$\Longrightarrow v''(\mu) = 2\big[\psi'(\mu)\big]^2 + O(n^{-1}).$$

Substituting the above expressions for $v(\mu), v''(\mu)$ in

$$E[v(\hat{\mu})] = v(\mu) + \frac{1}{2}\,v''(\mu)\,\frac{V(\mu)}{n} + O(n^{-2})$$

yields the required result. □

More accurate approximations for the MSE of the three estimators considered are given in Appendix A.

Therefore, for any of the three estimators considered, the root-Mean Squared Error (rMSE) equals

$$\sqrt{\psi'(\mu)^2\,\frac{\sigma^2}{n} + O(n^{-2})} = n^{-\frac{1}{2}}\sqrt{\psi'(\mu)^2\,\sigma^2 + O(n^{-1})} = |\,\psi'(\mu)\,|\,\sigma n^{-\frac{1}{2}} + O(n^{-\frac{3}{2}}) \tag{21}$$

Therefore the bias of the three estimators differs by terms of the order $O(n^{-1})$, while the rMSE differs by terms of order $O(n^{-3/2})$. Hence, bias correction can be performed without a substantial penalty in terms of estimation accuracy.

Moreover, all three estimators are nearly efficient. For parameter $\psi(\mu)$ it is easily shown that in the natural exponential family of distributions the Fisher information is $I_n(\psi) = n\kappa''(\theta)\,\left(\frac{\partial\theta}{\partial\psi}\right)^2$. By noting that $\kappa''(\theta) = V(\mu)$ and $\frac{\partial\theta}{\partial\psi} = \frac{\partial\theta}{\partial\mu}\frac{\partial\mu}{\partial\psi} = [V(\mu)\,\psi'(\mu)]^{-1}$ it follows that the Cramer-Rao lower bound for an unbiased estimator of $\psi(\mu)$ is $[\psi'(\mu)]^2\,\frac{V(\mu)}{n}$.

It is noted here that all results given above, holding for a distribution in the natural exponential family, also hold for a distribution arising from an increasing transform of a random variable following the original distribution. As in Section 4, consider a random variable $\tilde{Y} = t(Y)$, where $t(\cdot)$ is a strictly increasing function. Fix $P(Y > y) = \psi(\mu)$, such that y corresponds to a fixed percentile of Y. Now set $\tilde{y} = t(y)$ such that again $P(\tilde{Y} > \tilde{y}) = \psi(\mu)$. Let $\tilde{X}_1, \ldots, \tilde{X}_n$ be a sample from the distribution of $\tilde{Y}$. It is straightforward to show that then the MLE of the parameter μ from that sample is given by $\frac{1}{n}\sum_{j=1}^{n} t^{-1}(\tilde{X}_j) =$

$\frac{1}{n}\sum_{j=1}^{n} X_j = \hat{\mu}$. Therefore the statistics $\psi(\hat{\mu})$, $\psi^*(\hat{\mu})$, $\bar{\psi}(\hat{\mu})$ that are of interest in this paper, remain unchanged subject to such increasing transformations.

The following example demonstrates the performance of the estimators introduced in this section, when Y follows an exponential distribution. From the above discussion it follows that the presented results actually hold for a wider range of distributions. For example, if $t(y) = b\exp(y)$, $b > 0$, then $\tilde{Y}$ follows a 1-parameter Pareto distribution, which is the most widely used model in practice for modelling large losses when data are not abundant. Alternatively for $t(y) = y^{1/\gamma}$, $\gamma > 0$, we obtain a Weibull distribution with fixed shape parameter.

Example 3. *Once more, we deal with the example of an exponential tail function, with* $\psi(\mu) = P(Y > y) = e^{-y/\mu}$. *The relative biases of the estimators* $\psi(\hat{\mu})$, $\psi^*(\hat{\mu})$, $\bar{\psi}(\hat{\mu})$ *are plotted against the sample size n in figure 5, for* $\mu = 10$, $y = 30$, *corresponding to* $\psi(\mu) = 0.0498$. *It can be seen how the BPE has a higher bias than the MLE and how with the BCE the bias is nearly eliminated.*

We now compare the three estimators in terms of their rMSE, along with the square root of the CRLB, denoted by rCRLB. In Table 2 we provide values for the rMSEs and the rCRLB for a range of sample sizes. It can be seen that there are some differences, which, as argued in Section 5.3, disappear fairly quickly as the sample size increases, particularly for $y = 30$. *For* $y = 30$, *it can be seen that for very small samples the rMSE of the PBE is lower than that of the MLE, which is lower than that of the BCE. Moreover the rMSE of the MLE and PBE are also lower than the rCRLB. This indicates that for very small samples, there is an element of MSE / bias trade-off, though these effects quickly disappear as the sample size increases. For* $y = 46$, *the picture somewhat changes. The differences between the estimators are more pronounced and the rMSE of the PBE is now the highest.*

TABLE 2

COMPARATIVE rMSE OF ESTIMATORS OF EXPONENTIAL TAIL PROBABILITY $\psi(\mu) = e^{-y/\mu}$ FOR $\mu = 10$.

	a) $y = 30$, $\psi(\mu) \cong 0.05$				b) $y = 46$, $\psi(\mu) \cong 0.01$			
n	MLE	PBE	BCE	rCRLB	MLE	PBE	BCE	rCRLB
5	0.0653	0.0609	0.0715	0.0668	0.0310	0.0360	0.0289	0.0207
6	0.0599	0.0567	0.0645	0.0610	0.0274	0.0320	0.0251	0.0189
7	0.0556	0.0532	0.0592	0.0565	0.0247	0.0288	0.0224	0.0175
8	0.0521	0.0502	0.0550	0.0528	0.0225	0.0263	0.0204	0.0163
9	0.0492	0.0477	0.0516	0.0498	0.0208	0.0243	0.0188	0.0154
10	0.0467	0.0455	0.0488	0.0472	0.0194	0.0226	0.0175	0.0146
15	0.0383	0.0377	0.0393	0.0386	0.0149	0.0170	0.0135	0.0119
20	0.0332	0.0329	0.0339	0.0334	0.0124	0.0139	0.0113	0.0103
30	0.0272	0.0270	0.0275	0.0273	0.0096	0.0106	0.0090	0.0084
50	0.0211	0.0210	0.0212	0.0211	0.0071	0.0076	0.0068	0.0065

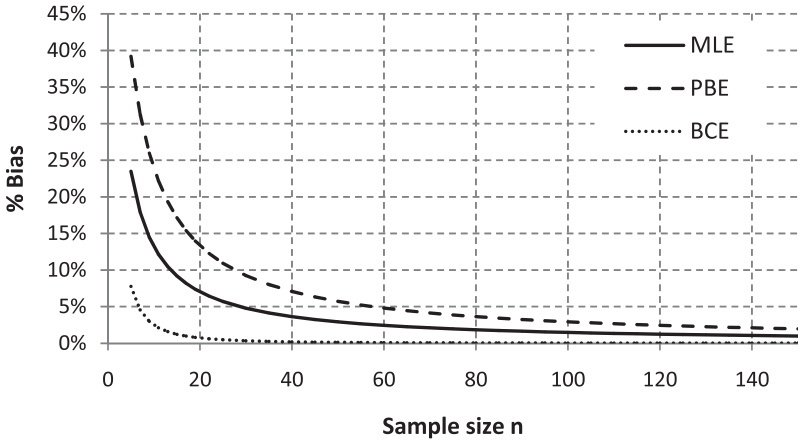

FIGURE 5: Relative biases for Maximum Likelihood, Parametric Bootstrap and Bias Corrected Estimators of exponential tail function against sample size; $\mu = 10$, $y = 30$.

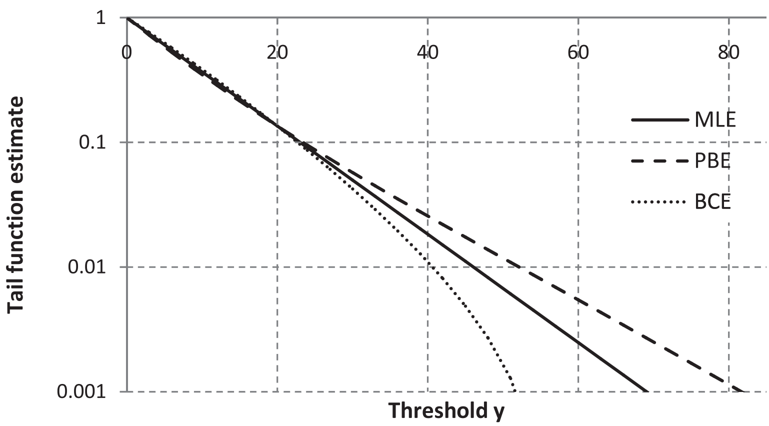

FIGURE 6: Exponential tail function estimates, using Maximum Likelihood, Parametric Bootstrap and Bias Corrected Estimators; $\hat{\mu} = 10$, $n = 10$.

The biases and rMSEs in this example were calculated with exact formulas; these are rather tedious and are given in Appendix A.

Finally we plot the tail functions obtained by using the MLE, the PBE and BCE in figure 6, for $n = 10$ *and* $\hat{\mu} = 10$. *The plot shows that the predictive distribution obtained by PBE is more conservative than the one obtained by MLE. On the other hand, the BCE tail function is not only lower, but for large thresholds also presents a substantial distortion in its shape (for very large thresholds it even becomes negative). Though such high thresholds will typically not be of interest in a pricing problem (especially when starting from a sample as small as 10 data*

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press

points), they would be considered in the rare case of an infinite reinsurance layer or when a tail-based risk measure such as Tail-Value-at-Risk is used.

The performance of the tail function estimators discussed is further studied in Appendix B for distributions that do not arise from increasing transforms of an exponential variable. In particular, 1-parameter versions of the (log-)Normal, (log-)Gamma, and Inverse Gaussian distributions are considered. For those distributions, the bias and MSE cannot in general be calculated analytically and therefore Monte-Carlo simulation is used to evaluate numerically the quantities studied. Appendix B demonstrates that the estimators have properties consistent with the insight obtained by asymptotic theory, for a range of models that is wider than the exponential and associated distributions.

6. Conclusions and discussion

We derived accurate approximations for functionals of sample means, subject to some technical conditions. For single-parameter exponential families, we showed that the tail and percentile functions are convex in the mean for high enough thresholds. These technical considerations allowed us to discuss the following:

- Maximum likelihood estimators of extreme tail probabilities and percentiles are approximately positively biased and accurate approximations of the bias were derived.
- Parametric bootstrap predictive distributions can be evaluated via analytical approximations and are shown to be more conservative than distributions estimated by MLE. However, parametric bootstrap estimators exacerbate the bias of the MLE.
- Analytical bias-corrected estimators can be easily introduced, but may distort the shape of the estimated distribution, especially in the extreme tail.
- The maximum likelihood, parametric bootstrap and bias corrected estimators have approximately the same Mean-Squared-Error, implying that there is only a limited MSE/bias trade-off.

We did not, however, discuss which of the 3 estimators one should use in practice; since their MSEs are approximately the same, this is not a trivial question. Arguably the choice of estimator may depend on the application in mind. For example, in a solvency related context, one may be interested in setting capital such that a given solvency probability is achieved, after allowing for the potentially adverse impact of parameter uncertainty. It was shown in Gerrard and Tsanakas (2010) that such a solvency criterion is best served by the use of a predictive distribution, which would point to the direction of a PBE. In that context, the issue of estimation bias does seem problematic. On the contrary, the presence of some positive bias in capital estimation becomes desirable, as this acts as an implicit risk load against parameter uncertainty.

On the other hand consider a stylised pricing example. A reinsurance company sells r policies, each of which produces a loss, modelled by the random variable Z_j. Losses $Z_1, \ldots, Z_r$ are considered i.i.d. and the premium for each policy is calculated as its expected loss. Parameters for each loss distribution are calculated from a different sample, $\mathbf{X}_j$ and let us also assume that each sample is of the same size n. Denote the premium for the j^{th} policy as $p(\mathbf{X}_j)$. Then a possible criterion for the accuracy of the pricing method is the quadratic deviation between total premium and total loss. Simple manipulations show that this can be written as

$$E\left[\left(\sum_{j=1}^{r} Z_j - \sum_{j=1}^{r} p(\mathbf{X}_j)\right)^2\right] = rVar(Z) + rVar(p(\mathbf{X})) + r^2\left(E(Z) - \pi(\mathbf{X})\right)^2,$$

showing that as the portfolio size r increases, the portfolio pricing error is primarily driven by the bias $(E(Z) - \pi(\mathbf{X}))$, due to the r^2 term, rather than the estimation variance $Var(p(\mathbf{X}))$. In other words, for a large homogenous portfolio the estimation volatility 'diversifies away', while the bias does not. This indicates the potential desirability of bias correction in such a context.

In reality such a homogenous portfolio of independent exposures will not exist, so that the diversification of estimation volatility will never be more than partial. In that case, it may be desirable to allow for some positive bias to act as a safety loading against parameter error. Figure 5 can be viewed in such a way; under the MLE and PBE, a smaller sample size implies a higher positive bias, that would lead to a premium which, on average, would be higher than the expected loss. As the sample size increases, the need for such a loading is eliminated. A difficult question to answer is how much bias one should allow; in other words what should be the value of a in an estimator of the form $\psi(\hat{\mu}) = \psi(\hat{\mu}) + \frac{a}{2}\psi''(\hat{\mu})\frac{V(\hat{\mu})}{n}$? This cannot be answered without a well specified, economically motivated decision criterion. Formulating such criteria is outside the scope of the present investigation, but remains a possible topic for future research.

APPENDIX A

I. HIGHER ORDER APPROXIMATIONS FOR MSE AND BIAS OF ESTIMATORS IN EXPONENTIAL FAMILIES

Notation

We consider a 1-parameter exponential family with mean μ. The second, third and fourth cumulants are denoted as functions of μ by $V(\mu)$, $\gamma(\mu)$, $\delta(\mu)$ respectively. It can be checked that $V'(\mu) = V'(\mu) = \frac{\gamma(\mu)}{V(\mu)}$, $V''(\mu) = \frac{\delta(\mu)}{V(\mu)} - \frac{\gamma(\mu)^2}{V(\mu)^2}$.

We are interested in the performance of a class of estimators for $\psi(\mu)$, that are given by

$$\psi_a(\hat{\mu}) = \psi(\hat{\mu}) + \frac{a}{2}\psi''(\hat{\mu})\frac{V(\hat{\mu})}{n},$$

where $\hat{\mu}$ denotes the sample mean based on a sample of size n.

In what follows, the following approximation will be used:

$$E[v(\hat{\mu})] = v(\mu) + \frac{1}{2}v''(\mu)\frac{\mu_2}{n} + \frac{1}{6}v^{(3)}(\mu)\frac{\mu_3}{n^2} + \frac{1}{8}v^{(4)}(\mu)\frac{\mu_2^2}{n^2} + O(n^{-3}) \qquad (22)$$

where $v(\cdot)$ is a function satisfying the conditions of Lemma 2.

Bias calculation

Let $v(m) = \psi_a(m) - \psi(\mu)$. Differentiating with respect to m, setting $m = \mu$, and collecting terms up to the order of interest, yields:

$$\begin{aligned}
v(\mu) &= \frac{a}{2}\psi''(\mu)\frac{V(\mu)}{n} \\
v'(\mu) &= \psi'(\mu) + \frac{a}{2}\left[\psi^{(3)}(\mu)V(\mu) + \psi''(\mu)V'(\mu)\right]\frac{1}{n} + O(n^{-2}) \\
v''(\mu) &= \psi''(\mu) + \frac{a}{2}\left[\psi^{(4)}(\mu)V(\mu) + \psi^{(3)}(\mu)V'(\mu) + \psi^{(3)}(\mu)V'(\mu) + \psi''(\mu)V''(\mu)\right]\frac{1}{n} \\
&\quad + O(n^{-2}) \\
v^{(3)}(\mu) &= \psi^{(3)}(\mu) + O(n^{-1}) \\
v^{(4)}(\mu) &= \psi^{(4)}(\mu) + O(n^{-1})
\end{aligned}$$

Using (22) and the expressions for the derivatives of $V(\mu)$, an approximation to the bias with error term $O(n^{-3})$ can be obtained.

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press

MSE calculation

Now $v(m) = (\psi_a(m) - \psi(\mu))^2$. Differentiating with respect to m, setting $m = \mu$, and collecting terms up to the order of interest, yields:

$$\begin{aligned} v(\mu) &= \frac{a^2}{4}\psi''(\mu)^2\frac{V(\mu)^2}{n^2} \\ v''(\mu) &= 2\psi'(\mu)^2 + 2a\Big[\psi'(\mu)\big(\psi^{(3)}(\mu)V(\mu) + \psi''(\mu)V'(\mu)\big) + \frac{1}{2}\psi''(\mu)^2 V(\mu)\Big]\frac{1}{n} + O(n^{-2}) \\ v^{(3)}(\mu) &= 6\psi'(\mu)\psi''(\mu) + O(n^{-1}) \\ v^{(4)}(\mu) &= 6\psi''(\mu)^2 + 8\psi'(\mu)\psi^{(3)}(\mu) + O(n^{-1}) \end{aligned}$$

Using (22) and the expressions for the derivatives of $V(\mu)$, an approximation to the MSE with error term $O(n^{-3})$ can be obtained.

II. Exact formulas for MSE of exponential tail function estimators

Let $\psi(\mu) = e^{-y/\mu} \Longrightarrow \psi''(\mu) = e^{-y/\mu}(y^2\mu^{-2} - 2y\mu^{-1})$. We need to calculate

$$MSE\big[\psi_a(\hat{\mu})\big] = E\big[(\psi_a(\hat{\mu}) - \psi(\mu))^2\big],$$

where $\mu \sim Gam(n, n/\mu)$ and $\psi_a(\mu) = \psi(\mu) + \frac{a}{2}\psi''(\mu)\frac{\mu^2}{n}$. The calculation of $MSE[\psi_a(\hat{\mu})]$ will involve a sum of terms including expectations of the form $E[\hat{\mu}^{-k}e^{-ry/\hat{\mu}}]$. To calculate these, define the integrals

$$\eta(c,b,x,k) = \int_0^\infty \frac{t^{c-k-1}e^{-x/t}e^{-t/b}}{\Gamma(c)b^c}dt = \frac{2(x/b)^{(c-k)/2}}{\Gamma(c)b^k}K_{c-k}(2\sqrt{x/b}),$$

where $K_{c-k}(\cdot)$ is a modified Bessel function of the second type. Hence it follows that

$$E[\hat{\mu}^{-k}e^{-ry/\hat{\mu}}] = \eta(n,b,ry,k),$$

where $b = \mu/n$. From developing the expression for the MSE and using the above equation, some rather tedious calculations yield:

$$\begin{aligned} MSE[\psi_a(\hat{\mu})] = {} & \eta(n,b,2y,0) + \frac{a^2y^4}{4n^2}\eta(n,b,2y,4) + \frac{a^2y^2}{n^2}\eta(n,b,2y,2) \\ & + \frac{ay^2}{n}\eta(n,b,2y,2) - \frac{2ay}{n}\eta(n,b,2y,1) - \frac{a^2y^3}{n^2}\eta(n,b,2y,3) \\ & - 2e^{-y/\mu}\Big[\eta(n,b,y,0) + \frac{a}{2n}\big(y^2\eta(n,b,y,2) - 2y\eta(n,b,y,1)\big)\Big] \\ & + e^{-2y/\mu} \end{aligned}$$

Appendix B

Here we consider the performance of the asymptotic approximations to the expected value of $E[\psi(\hat{\mu})]$, for one parameter versions of the Normal, Gamma and Inverse Gamma distributions. Furthermore, the Bias and rMSE of the three tail probability estimators considered (MLE, PBE, BCE) is determined. As for most of those quantities explicit expressions are hard to obtain, evaluation is by Monte-Carlo simulation.

I. Distributions used

Normal distribution

Let Φ, ϕ be respectively the standard normal distribution function and density and note that

$$\phi'(z) = -z\phi(z).$$

As before, set $\psi(\mu)$ equal to the tail probability $P(Y > y)$, calculated under parameter μ. Then, for the Normal distribution with mean μ and fixed variance σ^2, it is:

$$\begin{aligned}
\psi(\mu) &= 1 - \Phi\left(\frac{y-\mu}{\sigma}\right) \\
\psi'(\mu) &= \phi\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma} \\
\psi''(\mu) &= -\phi'\left(\frac{y-\mu}{\sigma}\right)\frac{1}{\sigma^2} \\
\psi^*(\hat{\mu}) &= \psi(\hat{\mu}) + \frac{1}{2}\psi''(\hat{\mu})\frac{\sigma^2}{n} \\
\bar{\psi}(\hat{\mu}) &= \psi(\hat{\mu}) - \frac{1}{2}\psi''(\hat{\mu})\frac{\sigma^2}{n}
\end{aligned}$$

The distribution of the sample mean is $\hat{\mu} \sim N(\mu, \sigma^2/n)$.

It can be easily shown that $E[\psi(\hat{\mu})] = 1 - \Phi\left(\frac{y-\mu}{\sqrt{\sigma^2(1+1/n)}}\right)$, so that this particular quantity can be calculated analytically.

While the normal distribution is not a common model for insurance claims, note that the analysis holds identically for a log-normal distribution, with the μ (relating to a scale parameter in the log-normal case) unknown and the shape parameter σ fixed.

Gamma distribution

Let $G(\cdot;\alpha)$, $g(\cdot;\alpha)$ be respectively the distribution function and density of a $Gam(\alpha,1)$ random variable, such that

$$g(z;\alpha) = \frac{z^{\alpha-1}\exp(-z)}{\Gamma(\alpha)}, \quad g'(z;\alpha) = \left(\frac{1}{z}(\alpha-1)-1\right)g(z;\alpha)$$

For a $Gam(\alpha,\lambda)$ distribution with mean μ, it is $\lambda = \alpha/\mu$, $V(\mu) = \mu^2/\alpha$. Thus:

$$\begin{aligned}
\psi(\mu) &= 1 - G\left(\frac{y\alpha}{\mu};\alpha\right)\\
\psi'(\mu) &= g\left(\frac{y\alpha}{\mu};\alpha\right)\frac{y\alpha}{\mu^2}\\
\psi''(\mu) &= -g'\left(\frac{y\alpha}{\mu};\alpha\right)\frac{(y\alpha)^2}{\mu^4} - 2g\left(\frac{y\alpha}{\mu};\alpha\right)\frac{y\alpha}{\mu^3}\\
\psi^*(\hat{\mu}) &= \psi(\hat{\mu}) + \frac{1}{2}\psi''(\hat{\mu})\frac{\hat{\mu}^2}{\alpha n}\\
\bar{\psi}(\hat{\mu}) &= \psi(\hat{\mu}) - \frac{1}{2}\psi''(\hat{\mu})\frac{\hat{\mu}^2}{\alpha n}
\end{aligned}$$

The distribution of the sample mean is $\hat{\mu} \sim Gam(n\alpha, n\alpha/\mu)$.

Again the discussion of the Gamma distribution also addresses the case of the log-Gamma distribution, which is a distribution with Pareto-type tail and tail index $\lambda = \alpha/\mu$.

Inverse Gaussian distribution

We consider an Inverse Gaussian distribution $IG(\mu,\lambda)$, such that the mean is μ and the variance function is $V(\mu) = \mu^3/\lambda$. More about the Inverse Gaussian distribution can be found in section A.4.1.2 of Klugman et al. (2004). For simplicity, here and in the sequel, we fix $\lambda \equiv 1$.

Let $u = (y+\mu)/\mu$ and $z = (y-\mu)/\mu$. Then, it is:

$$\begin{aligned}
\psi(\mu) &= 1 - \Phi\left(\frac{z}{y^{1/2}}\right) - e^{2/\mu}\Phi\left(-\frac{u}{y^{1/2}}\right)\\
\psi'(\mu) &= \frac{y^{1/2}}{\mu^2}\left[\phi\left(\frac{z}{y^{1/2}}\right) - e^{2/\mu}\phi\left(-\frac{u}{y^{1/2}}\right)\right] + \frac{2}{\mu^2}e^{2/\mu}\Phi\left(-\frac{u}{y^{1/2}}\right)\\
\psi''(\mu) &= -4e^{2/\mu}\Phi\left(-\frac{u}{y^{1/2}}\right)\left[\frac{1}{\mu^3}+\frac{1}{\mu^4}\right]\\
&\quad + e^{2/\mu}\phi\left(-\frac{u}{y^{1/2}}\right)\left[\frac{2y^{1/2}}{\mu^3}+\frac{4y^{1/2}}{\mu^4}-\frac{uy^{1/2}}{\mu^4}\right]\\
&\quad + \phi\left(\frac{z}{y^{1/2}}\right)\left[-\frac{2y^{1/2}}{\mu^3}+\frac{zy^{1/2}}{\mu^4}\right]
\end{aligned}$$

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press

$$\psi^*(\hat{\mu}) = \psi(\hat{\mu}) + \frac{1}{2}\psi''(\hat{\mu})\,\frac{\hat{\mu}^3}{n}$$
$$\bar{\psi}(\hat{\mu}) = \psi(\hat{\mu}) - \frac{1}{2}\psi''(\hat{\mu})\,\frac{\hat{\mu}^3}{n}$$

The distribution of the sample mean is $\hat{\mu} \sim IG(\mu, n)$.

II. Simulation study

Here we present results from a simulation study, assessing the performance of the asymptotic approximations and tail probability estimators introduced in the paper. Consistent parameters are chosen for the three distributions, such that for all of them the mean is equal to 0.16 and the coefficient of variation is 0.4. Hence the parameterisations used are $N(0.16, 0.064^2)$, $Gam(6.25, 39.0625)$, and $IG(0.16, 1)$. For all three distributions, a threshold y corresponding to $\psi(\mu) = 0.05$ is used. For each distribution and value of n, 10^8 pseudo-random samples from the distribution of μ are simulated.

In Table 3, for different values of the sample size n, the biases of the MLE $\psi(\hat{\mu})$, the PBE $\psi^*(\mu)$, and the BCE $\bar{\psi}(\mu)$ are given. The bias of MLE is calculated by simulation (or exactly in the case of the Normal), as well as using an approximation with error of order $O(n^{-2})$. The tabulated results are consistent with the results for the exponential/Pareto distribution presented earlier in the paper. In particular, the approximate value for the bias is quite close to the exact value, the PBE has bias approximately double of that of the MLE, and the BCE reduces the bias to nearly zero. Though the accuracy of asymptotic approximations decreases for very small sample sizes, the bias correction is still quite effective. For example, in the case of the Inverse Gaussian distribution, for $n = 5$, the relative bias of the MLE is $0.01461/0.05 \cong 29\%$, while for the BCE it is $0.00255/0.05 \cong 5\%$. When $n = 10$, the relative bias of the MLE is $0.008/0.05 \cong 16\%$, while for the BCE it becomes $0.00076/0.05 \cong 1.5\%$.

In Table 4, the rMSE of the MLE $\psi(\hat{\mu})$, the PBE $\psi^*(\mu)$, and the BCE $\bar{\psi}(\mu)$ are given. In agreement with the asymptotic arguments of Section 5.3, it is seen that the rMSE of the three estimators are quite close to each other, with the rMSE of the BCE usually lowest. This confirms the previous conclusion that, for distributions in the exponential family, the bias correction proposed does not entail an increase in MSE.

TABLE 3

BIAS OF ESTIMATORS MLE, PBE AND BCE OF TAIL PROBABILITY $\psi(\mu) = 0.05$.

a) Normal distribution, $N(0.16, 0.064^2)$.

n	MLE (exact)	MLE (approx.)	PBE	BCE
5	0.01661	0.01696	0.03278	0.00044
6	0.01390	0.01414	0.02751	0.00029
7	0.01195	0.01212	0.02370	0.00021
8	0.01048	0.01060	0.02081	0.00015
9	0.00933	0.00942	0.01854	0.00012
10	0.00840	0.00848	0.01672	0.00010
15	0.00562	0.00565	0.01121	0.00004
20	0.00422	0.00424	0.00843	0.00002
30	0.00282	0.00283	0.00564	0.00001
50	0.00169	0.00170	0.00339	0.00001

b) Gamma distribution, $Gam(6.25, 39.0625)$.

n	MLE (simul.)	MLE (approx.)	PBE	BCE
5	0.01580	0.01791	0.03003	0.00158
6	0.01343	0.01492	0.02571	0.00115
7	0.01167	0.01279	0.02247	0.00087
8	0.01032	0.01119	0.01997	0.00068
9	0.00925	0.00995	0.01796	0.00055
10	0.00839	0.00895	0.01632	0.00045
15	0.00571	0.00597	0.01121	0.00021
20	0.00433	0.00448	0.00854	0.00012
30	0.00292	0.00298	0.00578	0.00006
50	0.00177	0.00179	0.00352	0.00002

c) Inverse Gaussian distribution, $IG(0.16, 1)$.

n	MLE (simul.)	MLE (approx.)	PBE	BCE
5	0.01461	0.01782	0.02667	0.00255
6	0.01253	0.01485	0.02320	0.00187
7	0.01098	0.01273	0.02053	0.00144
8	0.00976	0.01113	0.01840	0.00113
9	0.00880	0.00990	0.01667	0.00092
10	0.00800	0.00891	0.01524	0.00076
15	0.00552	0.00594	0.01068	0.00036
20	0.00421	0.00445	0.00821	0.00021
30	0.00286	0.00297	0.00563	0.00010
50	0.00174	0.00178	0.00344	0.00003

TABLE 4

rMSE of estimators MLE, PBE and BCE of tail probability $\psi(\mu) = 0.05$.

a) Normal distribution, $N(0.16, 0.064^2)$

n	MLE	PBE	BCE
5	0.06002	0.06634	0.05768
6	0.05304	0.05811	0.05119
7	0.04787	0.05207	0.04636
8	0.04388	0.04742	0.04261
9	0.04069	0.04373	0.03960
10	0.03806	0.04072	0.03712
15	0.02970	0.03125	0.02917
20	0.02509	0.02614	0.02474
30	0.01996	0.02055	0.01976
50	0.01512	0.01540	0.01502

b) Gamma distribution, $Gam(6.25, 39.0625)$.

n	MLE	PBE	BCE
5	0.06468	0.06953	0.06274
6	0.05810	0.06210	0.05654
7	0.05310	0.05646	0.05181
8	0.04913	0.05201	0.04803
9	0.04591	0.04842	0.04497
10	0.04321	0.04542	0.04239
15	0.03438	0.03571	0.03391
20	0.02935	0.03026	0.02903
30	0.02358	0.02410	0.02340
50	0.01801	0.01827	0.01792

c) Inverse Gaussian distribution, $IG(0.16, 1)$.

n	MLE	PBE	BCE
5	0.06486	0.06859	0.06325
6	0.05884	0.06200	0.05752
7	0.05417	0.05688	0.05306
8	0.05042	0.05277	0.04947
9	0.04732	0.04940	0.04651
10	0.04472	0.04657	0.04401
15	0.03603	0.03717	0.03561
20	0.03095	0.03174	0.03066
30	0.02503	0.02550	0.02487
50	0.01922	0.01945	0.01914

References

Borowicz, J.M. and Norman, J.P. (2008) The effects of parameter uncertainty in the extreme event frequency-severity model. *Presented at the 28th International Congress of Actuaries*, Paris, 28 May - 2 June 2008. Available at: www.ica2006.com/Papiers/3020/3020.pdf, Accessed April 12, 2011.

Cairns, A.J.G. (2000) A discussion of parameter and model uncertainty in insurance. *Insurance: Mathematics and Economics* **27**, 313-30.

Cox, D.R. and Hinkley, D.V. (1979) *Theoretical Statistics.* Chapman and Hall.

Cummins, J.D. and Lewis, C.M. (2003) Catastrophic events, parameter uncertainty and the breakdown of implicit long-term contracting: The case of terrorism insurance. *Journal of Risk and Uncertainty* **26(2)**, 153-77.

Feller, W. (1966) *An Introduction to Probability Theory and Its Applications: Vol. 2.* Wiley.

Gerrard, R.J. and Tsanakas, A. (2010) Failure probability under parameter uncertainty. *Risk Analysis*, forthcoming.

Gradshteyn, I.S. and Ryzhik, I.M. (2007) *Table of Integrals, Series, and Products.* 7th ed. Academic Press.

Hall, P. (1997) *The Bootstrap and Edgeworth Expansion.* Springer.

Harris, I.R. (1989) Predictive fit for natural exponential families. *Biometrika*, **76(4)**, 675-84.

Klugman, S.A., Panjer, H.H. and Willmott G.E. (2004) *Loss Models: From Data to Decisions*. Wiley.

Landsman, Z. (2004) Second order Bayes prediction of functionals of exponential dispersion distributions and an application to the prediction of the tails. *ASTIN Bulletin* **34(2)**, 285-298.

Mata, A. (2000) Parameter uncertainty for extreme value distributions. *Presented at 2000 General Insurance Convension*, 25-28 October. Available at: http://www.actuaries.org.uk/sites/all/files/documents/pdf/0151-0173.pdf, Accessed April 12, 2011.

Murray, J.D. (1974) *Asymptotic Analysis.* Clarendon Press.

Powers, M.R., Venezian, E.C. and Jucá, I.B. (2003) Of happy and hapless regulators: the asymptotics of ruin. *Insurance: Mathematics and Economics*, **32**, 317-30.

Richards, S. (2009) Investigating parameter risk for Solvency II and ICAS. *Life and Pensions* Jun: 36-40.

Saltzmann, R. and Wuethrich, M.V. (2010) Cost-of-capital margin for a general insurance liability runoff. *Astin Bulletin*, **40(2)**, 415-451.

Smith, R.L. (1998) Bayesian and frequentist approaches to parametric predictive inference. In Bernardo, J.M., Berger, J.O., Dawid, A.P. and Smith, A.F.M. (eds). *Bayesian Statistics* **6**. Oxford University Press.

Verrall, R.J. and England, P.D. (2006) Predictive distributions of outstanding liabilities in general insurance. *Annals of Actuarial Science* **1(2)**, 221-70.

Z. Landsman
Department of Statistics
University of Haifa

A. Tsanakas (corresponding author)
Cass Business School
City University London
106 Bunhill Row, London EC1Y 8TZ,
United Kingdom
E-Mail: a.tsanakas.1@city.ac.uk

https://doi.org/10.2143/AST.42.1.2160738 Published online by Cambridge University Press