ACTA ET COMMENTATIONES UNIVERSITATIS TARTUENSIS DE MATHEMATICA
Volume 12, 2008

Approximating the integrated tail distribution

Ants Kaasik and Kalev Pärna

Abstract. We propose a natural approximation for the estimation of the distribution function of an integrated tail distribution of a subexponential distribution and prove that the approximation is almost surely uniformly convergent. The behaviour of the approximating error is studied using simulation. Knowing the distribution function of an integrated tail distribution is useful in the context of GI/G/1 queue with heavy-tailed service times and in the context of risk process with heavy-tailed claims.

1. Introduction

We start with two examples. Consider a standard M/G/1 queue process with service time distribution B. The traffic intensity, denoted by ρ, is defined as the ratio of mean service time μ and mean interarrival time, both of which are assumed to be finite. It is natural to assume that $\rho < 1$. The integrated tail distribution of B is defined as $B^I(x) = \int_0^x \bar{B}(y)dy/\mu$, where $\bar{B} = 1 - B$. It is well known that W, the steady-state waiting time of the queue, has the same distribution as $S_N = Y_1 + \ldots + Y_N$, where $Y_i > 0$ are independent and have distribution B^I and N is an independent geometric random variable with $\mathbb{P}(N = n) = (1 - \rho)\rho^n$ (see [1, p. 237]). But it holds even for GI/G/1 queue process that

$$\mathbb{P}(W > u) \sim \frac{\rho}{1-\rho}\bar{B}^I(u), \tag{1}$$

whenever B^I and B are subexponential (subexponential distributions are a sub-class of heavy-tailed distributions, see Section 2 for precise definition) (see [1, p. 296]); we have used the notation $a(x) \sim b(x)$ for expressing that $\lim_{x\to\infty} a(x)/b(x) = 1$. In insurance risk context, the probability of ultimate ruin of a company with initial reserve u can also be calculated as $\mathbb{P}(S_N > u)$

Received October 12, 2007.
2000 *Mathematics Subject Classification.* 62G30, 60K20.
Key words and phrases. Cramér-Lundberg model, GI/G/1 queue, subexponential distribution, integrated tail distribution.

if we assume the Cramér-Lundberg model. The above approximation can be used whenever the integrated tail distribution of claims is subexponential (see [1, p. 399]).

When the distribution B is known, B^I can be directly calculated and either the approximation (1) or simulation can be used for finding the approximate probability of a waiting time exceeding a large value u or a company with initial reserve u going bankrupt. In [2] two effective algorithms for simulating $\mathbb{P}(S_N > u)$ in case B^I has a regularly varying or a Weibull-like tail are presented. However normally B is not known beforehand and thus one firstly needs to choose a suitable class of distributions for the service times or claims and then estimate the parameter(s). There is room for subjective decisions when it comes to fixing a class of distributions, however. Thus it would seem more natural to avoid fixing a specific class and try to approximate the distribution function of B^I directly. This paper is concerned with providing and examining such an approximation.

The paper is organized as follows. In Section 2 we first present an approximation and then prove that our approximation is uniformly convergent. In Section 3 we study the convergence rate of the approximation numerically (and find out that the rate is dependent on the tail of B). Finally, in Section 4 we present a brief discussion.

2. Main result

We start this section by making precise the definition of subexponentiality.

Definition 2.1. A positive random variable X with distribution function B is called subexponential if for all $n \in \mathbb{N}$ it holds that

$$\lim_{x\to\infty} \frac{\overline{B^{*n}}(x)}{\bar{B}(x)} = n, \tag{2}$$

where B^{*n} denotes the n-fold convolution of B.

Remark 2.1. For most practical cases B is subexponential when B^I is and vice versa, thus it is usally enough to assume that the service time distribution is subexponential to be able to use the approximation (1); good reference for subexponential distributions is [4, pp. 36–57].

Consider the integrated tail distribution $B^I(x) = \int_0^x \bar{B}(y)dy/\mu$ as before. The idea is to replace μ with the sample mean and $\bar{B}(y)$ with the empirical distribution function. The following theorem shows that such an approximation has good theoretical properties.

Theorem 2.1. *Let X_n be a sequence of independent identically distributed (IID) positive random variables with a finite mean μ and cumulative distribution function B with B_n its empirical counterpart. Denote the sample mean by $\mu_n = (X_1 + \ldots + X_n)/n$. Then*

$$\mathbb{P}\left(\sup_x \left|\frac{\int_0^x \bar{B}_n(y)dy}{\mu_n} - \frac{\int_0^x \bar{B}(y)dy}{\mu}\right| \xrightarrow{n} 0\right) = 1. \tag{3}$$

Proof. Strong law of large numbers (SLLN) and Glivenko-Cantelli theorem hold simultaneously on a set which has probability one. Fix an ω from that set and also fix an $\epsilon > 0$. Because μ is finite there exists a $K > 0$ such that

$$\int_K^\infty \bar{B}(y)dy < \frac{\epsilon\mu}{6}. \tag{4}$$

Due to SLLN we have n_1, n_2 and n_3 such that when $n > n_1$

$$|\mu_n - \mu| < \min\left\{\frac{\epsilon\mu}{6}, \mu\frac{-1 + \sqrt{1 + 2\epsilon/3}}{2}\right\}, \tag{5}$$

when $n > n_2$

$$|\bar{B}_n(K) - \bar{B}(K)| < \frac{\epsilon\mu}{12K}, \tag{6}$$

and when $n > n_3$

$$\left|\int_K^\infty y d\bar{B}_n(y) - \int_K^\infty y d\bar{B}(y)\right| < \frac{\epsilon\mu}{12}. \tag{7}$$

Due to Glivenko-Cantelli theorem, there exists n_4 such that when $n > n_4$

$$\sup_y |\bar{B}_n(y) - \bar{B}(y)| < \frac{\epsilon\mu}{6K}. \tag{8}$$

Thus when $n > \max\{n_1, n_2, n_3, n_4\}$ we have

$$
\begin{aligned}
&\frac{1}{\mu}\sup_x\left|\frac{\mu}{\mu_n}\int_0^x \bar{B}_n(y)dy - \int_0^x \bar{B}(y)dy\right| \\
&\qquad \leqslant \frac{1}{\mu}\sup_x\left|\frac{\mu}{\mu_n}\int_0^x \bar{B}_n(y)dy - \int_0^x \bar{B}_n(y)dy\right| \\
&\qquad\quad + \frac{1}{\mu}\sup_x\left|\int_0^x \bar{B}_n(y)dy - \int_0^x \bar{B}(y)dy\right| \\
&\qquad \overset{(5)}{<} \frac{\epsilon}{6} + \frac{1}{\mu}\sup_x\left|\int_0^x \bar{B}_n(y)dy - \int_0^x \bar{B}(y)dy\right| \\
&\qquad \leqslant \frac{\epsilon}{6} + \frac{1}{\mu}\sup_{x\leqslant K}\int_0^x \left|\bar{B}_n(y)dy - \bar{B}(y)\right| dy \\
&\qquad\quad + \frac{1}{\mu}\sup_{x>K}\left|\int_0^x \bar{B}_n(y)dy - \int_0^x \bar{B}(y)dy\right| \\
&\qquad \overset{(8)}{<} \frac{\epsilon}{3} + \frac{1}{\mu}\sup_{x>K}\left|\int_0^x \bar{B}_n(y)dy - \int_0^x \bar{B}(y)dy\right| \\
&\qquad \leqslant \frac{\epsilon}{3} + \frac{1}{\mu}\sup_{x>K}|\mu_n - \mu| + \frac{1}{\mu}\int_K^\infty \bar{B}_n(y)dy + \frac{1}{\mu}\int_K^\infty \bar{B}(y)dy \\
&\qquad \overset{(4),(5)}{<} \frac{2\epsilon}{3} + \frac{1}{\mu}\int_K^\infty \bar{B}_n(y)dy \\
&\qquad = \frac{2\epsilon}{3} - \frac{K}{\mu}\bar{B}_n(K) + \frac{1}{\mu}\int_K^\infty y\ d\bar{B}_n(y) \\
&\qquad \overset{(6),(7)}{<} \frac{5\epsilon}{6} + \frac{1}{\mu}\int_K^\infty \bar{B}(y)dy \\
&\qquad \overset{(4)}{<} \epsilon,
\end{aligned}
$$

as required. □

Remark 2.2. Theorem 2.1 is general and does not require the data to come from any specific class of distributions, only the finiteness of mean is necessary. However for the approximation (1) to work, the subexponentiality requirements, as mentioned in Section 1, must be satisfied.

The empirical approximation $B_n^I(x) = \int_0^x \bar{B}_n(y)dy/\mu_n$ is not an empirical distribution function in the traditional sense: while being a proper distribution function and piecewise linear it does not possess any discontinuity points like empirical distribution functions do, however the distribution B_n^I does have finite support (the largest member of the sample is the right endpoint). This is the main reason why the supremum absolute error rather than the supremum relative error is considered in the next section.

3. Behaviour of the approximation error

First consider a cumulative distribution function $F(x)$ and its empirical counterpart $F_n(x)$ based on IID random sample of size n, and denote the supremum absolute error $A_n = \sup_x |F(x) - F_n(x)|$. The Dvoretzky-Kiefer-Wolfowitz inequality (given in [3]) states that we can estimate $\mathbb{P}(A_n > \epsilon) \leqslant Ce^{-2n\epsilon^2}$. The inequality can be used to construct a "confidence interval" for $F(x)$. Roughly speaking it states that the half-width of the confidence interval converges to zero with rate $1/\sqrt{n}$. Also due to the same inequality (strengthened by [5]),

$$\mathbb{E}A_n \leqslant \int_0^1 2e^{-2nx^2}\,dx = \sqrt{\frac{\pi}{2n}}(2\Phi(2\sqrt{n}) - 1) < \sqrt{\frac{\pi}{2n}},$$

meaning that the bound for the mean supremum absolute error of an empirical cumulative distribution function converges to zero with rate $1/\sqrt{n}$.

In our case the Dvoretzky-Kiefer-Wolfowitz inequality obviously cannot be used and we will use simulation to study the convergence rates of the half-width of the confidence interval and mean of the supremum absolute error. In the role of B we will use the finite-mean Pareto distribution where $\bar{B}(x) = (1+x)^{-\alpha}$ and $\alpha > 1$, the heavy-tailed Weibull distribution where $\bar{B}(x) = e^{-x^\beta}$ and $0 < \beta < 1$, and the log-normal distribution where $\bar{B}(x) = \bar{\Phi}((\ln x)/\sigma)$ where $\sigma > 0$, three the most prominent members of the subexponential distributions.

For the Pareto distribution with parameter α, the integrated tail distribution is also Pareto but with parameter $\alpha - 1$. First we simulated differently-sized samples from the Pareto distribution with parameter 2 and 3 and estimated the ϵ_n for which $\mathbb{P}(\sup |B_n^I(x) - B^I(x)| > \epsilon_n) = 0.05$ (this is basically an estimation of the 0.95-quantile). The point estimates (produced by 20000 replicates) given in Table 1 seem to confirm Theorem 2.1 but show that the approximation needs a huge sample size to give meaningful results.

Table 1. Half-width of the 95%-confidence interval for the Pareto case

n	α=2	α=3
100	0.2564	0.1769
1000	0.1192	0.0636
10000	0.0470	0.0212
100000	0.0181	0.0068

To gain insight into the rate of convergence of the half-width of the confidence interval we simulate the ratio ϵ_{2n}/ϵ_n. This ratio is $1/\sqrt{2} = 0.7071$ if the rate of convergence is $1/\sqrt{n}$, bigger than that if the convergence is slower, and smaller otherwise. The results are complemented with confidence intervals and presented in Table 2. One point estimate is based on 100000 total simulations. The conclusion from Table 2 is that the convergence accelerates with the increase of the sample size but the rate does not exceed $1/\sqrt{n}$. Also the heavier the tail, the slower the rate of convergence is. Further simulations (not presented here) show that for $\alpha \approx 1$ the rate of convergence can be arbitrarily slow at first, and for large parameter values, on the contrary, the rate of convergence approaches $1/\sqrt{n}$ extremely fast.

Table 2. Quantile ratio with 95%-confidence intervals for the Pareto case

n	α=2	α=3
100	0.7970 (0.7835;0.8094)	0.7459 (0.7342;0.7570)
1000	0.7593 (0.7424;0.7752)	0.7265 (0.7158;0.7373)
10000	0.7541 (0.7382;0.7706)	0.7092 (0.6995;0.7183)
100000	0.7350 (0.7201;0.7476)	0.7131 (0.7039;0.7222)

Finally we take a look at the average supremum error, which can be regarded as a typical supremum error when using the aproximation. The results in Table 3 show that it is roughly half of the 0.95-quantile. The heaviness of the tail, once again, plays an important part in the size of the error.

Table 3. Mean supremum absolute error for the Pareto case

n	α=2	α=3
100	0.1319	0.0855
1000	0.0572	0.0306
10000	0.0230	0.0103
100000	0.0087	0.0033

We carried out similar simulations (in terms of replications and principles) for the Weibull distribution with parameter 1/3 and 1/2. In those cases $B^I(x) = 1 - e^{-\sqrt[3]{x}}(\sqrt[3]{x} + \sqrt[3]{x^2}/2 + 1)$ and $B^I(x) = 1 - e^{-\sqrt{x}}(\sqrt{x} + 1)$, respectively. The results are similar to the ones obtained in the Pareto case. Table 4 shows that for small β the approximation can be rather inaccurate when the sample size is not extremely large. The rate of convergence is dependent on the tail of the distribution and with the increase of the sample size it approaches but does not exceed $1/\sqrt{n}$. The average supremum error depends on the tail of B and is roughly the 0.95-quantile divided by 2.

Table 4. Half-width of the 95%-confidence interval for the Weibull case

n	β=1/3	β=1/2
100	0.3675	0.2135
1000	0.1463	0.0704
10000	0.0481	0.0224
100000	0.0154	0.0071

Table 5. Quantile ratio with 95%-confidence intervals for the Weibull case

n	β=1/3	β=1/2
100	0.7688 (0.7612;0.7774)	0.7259 (0.7168;0.7351)
1000	0.7305 (0.7212;0.7399)	0.7046 (0.6952;0.7132)
10000	0.7100 (0.7017;0.7183)	0.7115 (0.7033;0.7195)
100000	0.7128 (0.7039;0.7215)	0.7009 (0.6922;0.7101)

Table 6. Mean supremum absolute error for the Weibull case

n	β=1/3	β=1/2
100	0.1885	0.1027
1000	0.0700	0.0342
10000	0.0236	0.0110
100000	0.0076	0.0035

Finally, we simulated the log-normal case for the log-normal distribution $B^I(x) = xe^{-\sigma^2/2}\bar{\Phi}[(\log x)/\sigma] + \Phi[(\log x)/\sigma - \sigma]$ (the number of simulations was left unchanged). The results are similar to the previous cases. Table 7 shows that when σ gets bigger (the tail of the original and the integrated tail distribution gets heavier) the accuracy of the approximation decreases.

Table 7. Half-width of the 95%-confidence interval for the log-normal case

n	σ=1	σ=2
100	0.1431	0.4172
1000	0.0479	0.2118
10000	0.0152	0.0900
100000	0.0048	0.0326

The rate of convergence is also dependent on the tail of the distribution and with the increase of the sample size it approaches but does not exceed $1/\sqrt{n}$.

The average supremum error is again roughly the 0.95-quantile divided by 2 as can be seen from Table 9.

TABLE 8. Quantile ratio with 95%-confidence intervals for the log-normal case

n	σ=1	σ=2
100	0.7309 (0.7210;0.7414)	0.8201 (0.8121;0.8274)
1000	0.7086 (0.7001;0.7179)	0.7888 (0.7759;0.8025)
10000	0.7082 (0.6995;0.7169)	0.7407 (0.7290;0.7542)
100000	0.7064 (0.6980;0.7147)	0.7220 (0.7082;0.7331)

TABLE 9. Mean supremum absolute error for the log-normal case

n	σ=1	σ=2
100	0.0692	0.2367
1000	0.0231	0.1069
10000	0.0074	0.0429
100000	0.0023	0.0155

4. Conclusion

The approximation for the integrated tail distribution proposed in this paper can be made to work in practice, but when the tail of the original distribution is too heavy, the approximation falls apart in a sense that the required sample size for an acceptable approximating error becomes very large. Thus if size of the sample is under our control, the decision of how large a sample to use still remains subjective. However, the main result of the paper at least guarantees that the approximation is uniformly convergent whenever the sample is made up of IID random variables with finite mean. Simulations show that the convergence rate does not exceed $1/\sqrt{n}$, but it is reached with relatively small sample size when the random variables are not too heavy-tailed. The rate of convergence can be arbitrarily slow, at first, when the mean of the random variables "approaches infinity".

Acknowledgement. The research was partially supported by Estonian Science Foundation Grant 7044.

References

[1] S. Asmussen (2003), *Applied Probability and Queues*, Second Edition, Springer-Verlag, New York.

[2] S. Asmussen and D. P. Kroese (2006), *Improved algorithms for rare event simulation with heavy tails*, Adv. in Appl. Probab. **38**(2), 545–558.

[3] A. Dvoretzky, J. Kiefer, and J. Wolfowitz (1956), *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*, Ann. Math. Statist. **27**(3), 642–669.

[4] P. Embrechts, T. Mikosch, and C. Klüppelberg (1997), *Modelling Extremal Events. For Insurance and Finance*, Springer-Verlag, Berlin.

[5] P. Massart (1990), *The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality*, Ann. Probab. **18**(3), 1269–1283.

Institute of Mathematical Statistics, University of Tartu, Liivi 2, 50409 Tartu, Estonia

E-mail address: `ants.kaasik@ut.ee`

E-mail address: `kalev.parna@ut.ee`