Approximating the Integrated Tail Distribution

Ants Kaasik & Kalev Pärna
University of Tartu

VIII Tartu Conference on Multivariate Statistics
26th of June, 2007

Motivating example

TARTU ÜLIKOOL

Let W be a steady-state waiting time of a $M/G/1$ queue process with service time distribution B

μ :=mean service time, ν :=mean interarrival time, $\mu < \nu < \infty$

$B^I(x) := \int_0^x \bar{B}(y)dy/\mu$, where $\bar{B} = 1 - B$, is the integrated tail distribution

Motivating example

TARTU ÜLIKOOL

Let W be a steady-state waiting time of a $M/G/1$ queue process with service time distribution B

μ :=mean service time, ν :=mean interarrival time, $\mu < \nu < \infty$

$B^I(x) := \int_0^x \bar{B}(y)dy/\mu$, where $\bar{B} = 1 - B$, is the integrated tail distribution

Then

$$W \stackrel{D}{=} \sum_{i=1}^{N} Y_i,$$

Y_i are IID with distribution B^I and independent of N which has a geometric distribution (the parameter of which can be expressed in terms of μ and ν).

The simulation approach

When B is known, $\mathbb{P}(W > u)$ can be estimated with excellent precision regardless how big u is (this is true regardless of the tail of B).

The simulation approach

When B is known, $\mathbb{P}(W > u)$ can be estimated with excellent precision regardless how big u is (this is true regardless of the tail of B).

Thus we have the following (simulation) approach:

data $\xrightarrow{(1)}$ estimate of B $\xrightarrow{(2)}$ estimate of $\mathbb{P}(W > u)$

The simulation approach

When B is known, $\mathbb{P}(W > u)$ can be estimated with excellent precision regardless how big u is (this is true regardless of the tail of B).

Thus we have the following (simulation) approach:

data $\xrightarrow{(1)}$ estimate of B $\xrightarrow{(2)}$ estimate of $\mathbb{P}(W > u)$

The estimation process (1) can be the cause of a major error! (we do not know even the right class of distributions)

The simulation approach

When B is known, $\mathbb{P}(W > u)$ can be estimated with excellent precision regardless how big u is (this is true regardless of the tail of B).

Thus we have the following (simulation) approach:

data $\xrightarrow{(1)}$ estimate of B $\xrightarrow{(2)}$ estimate of $\mathbb{P}(W > u)$

The estimation process (1) can be the cause of a major error! (we do not know even the right class of distributions)

It would be better to find $\mathbb{P}(W > u)$ directly as B is of no interest

Subexponential distributions

Random variable X has a subexponential distribution iff $X : \Omega \to (0, \infty)$ and

$$\frac{\overline{F^{*2}}(x)}{\overline{F}(x)} \to 2, \quad x \to \infty,$$

where F is the cdf of X.

Important members of the subexponential family of distributions are distributions with a regularly varying tail, lognormal distribution and Weibull distribution.

The proposed approach

Let B^I be subexponential. Then

$$\mathbb{P}(W > u) \sim \mathbb{E}N \cdot \bar{B}^I(u),$$

where $a(x) \sim b(x)$ iff $\lim_{x \to \infty} a(x)/b(x) = 1$.

The proposed approach

Let B^I be subexponential. Then

$$\mathbb{P}(W > u) \sim \mathbb{E}N \cdot \bar{B}^I(u),$$

where $a(x) \sim b(x)$ iff $\lim\limits_{x \to \infty} a(x)/b(x) = 1$.

Idea:

$\mu \approx \mu_n$ (LLN)

$B(y) \approx B_n(y)$ (Glivenko-Cantelli)

thus

$B^I(x) = \int_0^x \bar{B}(y)dy/\mu \overset{?}{\approx} \int_0^x \bar{B}_n(y)dy/\mu_n$

Main result

Let X_n be a sequence of IID positive random variables with a finite mean μ and cumulative distribution function B with B_n its empirical counterpart. Also denote the sample mean with $\mu_n = (X_1 + \ldots + X_n)/n$. Then the following result holds

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

Simulation study (1)

The theorem does not say anything about the rate of convergence. We studied it for the Pareto ($\bar{B}(x) = (1+x)^{-\alpha}$, $\alpha > 1$) and Weibull ($\bar{B}(x) = e^{-x^{\beta}}$, $0 < \beta < 1$) case with the help of simulations.

Let ϵ_n be "half-width of the confidence interval of B" i.e. $\mathbb{P}(\sup |B_n^I(x) - B^I(x)| > \epsilon_n) = 0.05$

Simulation study (1)

The theorem does not say anything about the rate of convergence. We studied it for the Pareto ($\bar{B}(x) = (1+x)^{-\alpha}$, $\alpha > 1$) and Weibull ($\bar{B}(x) = e^{-x^{\beta}}$, $0 < \beta < 1$) case with the help of simulations.

Let ϵ_n be "half-width of the confidence interval of B" i.e. $\mathbb{P}(\sup|B_n^I(x) - B^I(x)| > \epsilon_n) = 0.05$

How big is the quantile ϵ_n? (initial width)
How does the ratio ϵ_{2n}/ϵ_n behave? (rate of convergence)
How big is $\mathbb{E}(\sup|B_n^I(x) - B^I(x)|)$? (average supremum error of the estimate)

Simulation study (2)

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

Table: Half-width of the 95%-confidence interval for the Pareto 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: Half-width of the 95%-confidence interval for the Weibull case

Simulation study (3)

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)

Table: Quantile ratio with 95%-confidence intervals for the Pareto 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: Quantile ratio with 95%-confidence intervals for the Weibull case

Simulation study (4)

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

Table: Mean supremum absolute error for the Pareto 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

Table: Mean supremum absolute error for the Weibull case

Onwards...

TARTU ÜLIKOOL

The asymptotic equivalence $\mathbb{P}(W > u) \sim \mathbb{E}N \cdot \bar{B}^I(u)$ remains valid for the case of a $GI/G/1$ queue (we need an additional assumption that also B is subexponential). There is also a dual problem in insurance risk context (the ultimate ruin probability for a company dealing with subexponential claims).

Onwards...

The asymptotic equivalence $\mathbb{P}(W > u) \sim \mathbb{E}N \cdot \bar{B}^I(u)$ remains valid for the case of a $GI/G/1$ queue (we need an additional assumption that also B is subexponential). There is also a dual problem in insurance risk context (the ultimate ruin probability for a company dealing with subexponential claims).

Our proposed approach has an obvious shortfall in practice: support of the approximating distribution is a finite interval.

Onwards...

The asymptotic equivalence $\mathbb{P}(W > u) \sim \mathbb{E}N \cdot \bar{B}^I(u)$ remains valid for the case of a $GI/G/1$ queue (we need an additional assumption that also B is subexponential). There is also a dual problem in insurance risk context (the ultimate ruin probability for a company dealing with subexponential claims).

Our proposed approach has an obvious shortfall in practice: support of the approximating distribution is a finite interval.

The way onwards is clear: sample data must be used to fit a generalized Pareto distribution to the tail, so that we have an approximating distribution support of which is $(0, \infty)$.

Thank you for listening!