Approximation of Heavy-tailed distributions via infinite dimensional phase–type distributions

Leonardo Rojas-Nandayapa
The University of Queensland

ANZAPW March, 2015.
Barossa Valley, SA, Australia.
Summary of Mogens’ talk...
Phase–type
Distributions of first passage times associated to a Markov Jump Process with a finite number of states.

Facts

1. Dense class in the nonnegative distributions.
3. However, it is a subclass of light–tailed distributions.
Key problem
For most applications, a phase–type approximation will suffice. However, for certain important problems this fail.

A partial solution
Consider Markov Jump processes with infinite number of states. This class contains heavy–tailed distribution but also many other degenerate cases.
A more refined approach: Infinite mixtures of PH

Consider $X$ a random variable with phase–type distribution, and let $S$ be a nonnegative discrete random variable. Then the distribution of the random variable

$$Y := S \cdot X$$

is an *infinite mixture of phase–type distributions*. Such a class is mathematically tractable and obviously dense in the nonnegative distributions.
Let $F$, $G$ and $H$ be the cdf of $Y$, $S$ and $X$ respectively. Then

$$F(y) = \int_0^\infty G(y/s) dH(s), \quad y \geq 0.$$ 

The right hand side is called the Mellin–Stieltjes transform.

If $G$ is phase–type, the $F$ is absolutely continuous with density

$$f(y) = \int_0^\infty g(y/s) dH(s), \quad y \geq 0.$$
Some interesting questions related to this class of distributions?
1. What are the possible tail behaviors in this class?
2. How to approximate a “theoretic” heavy-tailed distribution?
3. Given a sample $y_1, \ldots, y_n$. How to estimate the parameters of a distribution in this class?
Heavy Tails
Recall that...

Heavy–tailed distributions

A nonnegative random variable $X$ has a heavy–tailed distribution iff

$$\mathbb{E}[e^{\theta X}] = \infty, \quad \forall \theta > 0.$$  

Equivalently,

$$\limsup_{x \to \infty} \frac{\mathbb{P}(X > x)}{e^{-\theta x}} = \infty, \quad \forall \theta > 0.$$  

Otherwise we say $X$ is light–tailed.
Characterization via Convolutions

Let $X_1, X_2$ be nonnegative iid rv with unbounded support,

1. It holds that

$$\liminf_{x \to \infty} \frac{\mathbb{P}(X_1 + X_2 > x)}{\mathbb{P}(X_1 > x)} \geq 2.$$

2. Moreover, $X_1$ is heavy–tailed iff

$$\liminf_{x \to \infty} \frac{\mathbb{P}(X_1 + X_2 > x)}{\mathbb{P}(X_1 > x)} = 2.$$

3. Furthermore, $X$ is subexponential iff

$$\lim_{x \to \infty} \frac{\mathbb{P}(X_1 + X_2 > x)}{\mathbb{P}(X_1 > x)} = 2.$$
Heavy-tailed random variables arise as

- Exponential transformations of some light-tailed random variables (normal—lognormal, exponential—Pareto).
- As weak limits of normalized maxima (also normalized sums).
- As reciprocals of certain random variables.
- Random sums of light-tailed distributions.
- Products of light-tailed distributions.
Problem 1:
What are the possible tail behaviours in this class?
Theorem
Let $X$ be a phase–type distribution and $S$ be a nonnegative random variable. Define

$$Y := S \cdot X$$

$Y$ is heavy–tailed iff $S$ has unbounded support.
Example

Let $X \sim \exp(\lambda)$ and $S$ any nonnegative discrete distribution supported over a countable set \{\(s_1, s_2, \ldots\)\}.

\[
\lim_{y \to \infty} \frac{\mathbb{P}(S \cdot X > y)}{e^{-\theta y}} = \lim_{y \to \infty} \sum_{k=1}^{\infty} \frac{e^{-\lambda y/s_k} \mathbb{P}(S = s_k)}{e^{-\theta y}} = \infty.
\]

(choose $s_k$ large enough and such that $\lambda/s_k < \theta$).
Theorem (generalization)

Let $S \sim H_1$ and $X \sim H_2$, and let $Y := S \cdot X$.

1. If there exist $\theta > 0$ and $\xi(x)$ a nonnegative function such that
   \[
   \limsup_{x \to \infty} e^{\theta x} \left( \frac{H_1(x/\xi(x))}{\xi(x)} + H_2(\xi(x)) \right) = 0,
   \]
   then $Y$ is light–tailed.

2. If there exists $\xi(x)$ a nonnegative function such that for all $\theta > 0$ it holds that
   \[
   \limsup_{x \to \infty} e^{\theta x} \frac{H_1(x/\xi(x))}{\xi(x)} \cdot H_2(\xi(x)) = \infty,
   \]
   then $Y$ is heavy–tailed.
Example

Let $S \sim \text{Weibull}(\lambda, p)$ and $Y \sim \text{Weibull}(\beta, q)$. Then $Y := S \cdot X$ is light–tailed iff

$$\frac{1}{p} + \frac{1}{q} \leq 1. $$

Otherwise it is heavy–tailed.
What are the possible maximum domains of attraction?
Breiman’s Lemma
If $S$ is a $\alpha$–regularly varying random variable and there exists $\delta > 0$ such that $\mathbb{E}[X^{\alpha+\delta}] < \infty$, then

$$Y := S \cdot X$$

has an $\alpha$–regularly varying distribution. Moreover

$$\mathbb{P}(Y > x) \sim \mathbb{E}[X]\mathbb{P}(S > x).$$
Breiman’s Lemma
If $L_{1/S}$ is a $\alpha$–regularly varying function then

$$Y := S \cdot X$$

has an $\alpha$–regularly varying distribution. Moreover

$$\mathbb{P}(Y > x) \sim \mathbb{E}[X] \mathbb{P}(S > x).$$
Let $V(x) = (-1)^{\eta-1} \mathcal{L}_{1/S}^{(n-1)}(x)$.\(^1\)

**Theorem**

If $V(\cdot)$ is a von Mises function, then $F \in \text{MDA}(\Lambda)$. Moreover, if

$$\liminf_{x \to \infty} \frac{V(tx) V'(x)}{V'(tx) V(x)} > 1, \quad \forall t > 1,$$

then $F$ is subexponential.

---

\(^1\) $\eta$ is the largest dimension among the Jordan blocks associated to the largest eigenvalue of the sub–intensity matrix.
Problem 2:
How to approximate theoretical heavy–tailed distributions?
An approach for approximating theoretic distributions

- Take a sequence of Erlang rvs

\[ X_n \sim \text{Erlang}(n, n). \]

- Consider increasing sequences

\[ S_n := \{s_{n,0}, s_{n,1}, s_{n,2}, \ldots\}, \quad n \in \mathbb{N} \]

such that \( s_{n,0} = 0, \) \( s_{n,i} \to \infty \) as \( i \to \infty \) and

\[ \Delta_n := \sup\{s_{n,i+1} - s_{n,i} : i \in \mathbb{N}\} \to 0, \]

as \( n \to 0 \) For each \( n, \) define the cdf’s

\[ F_n(x) = F(s_{n,i+1}), \quad x \in [s_{n,i}, s_{n,i+1}). \]
Construct a sequences of random variables

\[ S_n \sim F_n, \quad X_n \sim \text{Erlang}(n, n). \]

Since \( X_n \rightarrow 1 \), then Slutsky’s theorem implies that

\[ S_n \cdot X_n \rightarrow X \sim F. \]

Note: It can be modified and will also work for light-tailed distributions.
Ruin problem

Definition
Let $F$ and $G$ be two distribution functions with non-negative support. The relative error from $F$ to $G$ from 0 to $s$ is defined by

$$\sup_{x \in [0,s]} \frac{|F(x) - G(x)|}{|F(x)|}.$$
Theorem

Fix $u > 0$ and let $F$ and $G$ be such that

$$
\sup_{x \in [0,u]} \frac{|F_I(x) - G_I(x)|}{|F_I(x)|} \leq \epsilon(u)
$$

where $\epsilon : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ is some non-decreasing function. Let $F_I$ be the integrated tail of $F$ and let $G$ and an approximation of $F$. Further assume that $F$ dominates $G$ stochastically. Then

$$
|\psi_F(u) - \hat{\psi}_G(u)| \leq \epsilon(u) \mathbb{E} \left( \text{Number of down-crossings}; A \right),
$$

where $A := \{\text{Ruin happens in the last down-crossing}\}$. 


Problem 3: Statistical inference?
Key Idea

See the first passages times as incomplete data set from the evolution of the Markov Jump Process. Employ the EM algorithm for estimating the parameters.
Density of an infinite mixture of PH

\[ f_A(y) = \sum_{i=1}^{\infty} \pi_i(\theta) \alpha e^{T y/s_i} t/s_i, \quad x \geq 0. \]

Complete Data Likelihood

\[ \ell_c(\theta, \alpha, T) = \sum_{i=1}^{\infty} L^i \log \pi_i(\theta) + \sum_{i=1}^{\infty} \sum_{k=1}^{p} B^i_k \log \alpha_k + \sum_{i=1}^{\infty} \sum_{k \neq \ell}^{\infty} N^i_{k\ell} \log \left( \frac{T_{k\ell}}{s_i} \right) \]

\[ - \sum_{i=1}^{\infty} \sum_{k \neq \ell}^{\infty} \frac{T_{k\ell}}{s_i} Z^i_k + \sum_{i=1}^{\infty} \sum_{k=1}^{p} A^i_k \log \left( \frac{t_k}{s_i} \right) - \sum_{i=1}^{\infty} \sum_{k=1}^{p} \frac{t_k}{s_i} Z^i_k. \]
EM estimators
Assume \((\theta, \alpha, T)\) is the current set of parameters in the EM algorithm. The one step EM-estimators are thus given by

\[
\hat{\theta} = \arg\max_\Theta \sum_{i=1}^{\infty} (\alpha U_i t) \cdot \log \pi_i(\Theta),
\]

\[
\hat{\alpha} = \frac{\text{diag}(\alpha) U t}{N},
\]

\[
\hat{T}_{kl} = (\text{diag}(R)^{-1} T \circ R)_{k\ell}, \quad k \neq \ell,
\]

\[
\hat{t} = \alpha U \text{ diag}(t) \text{ diag}(R)^{-1},
\]

where

\[
R = R(\theta, \alpha, T) = \sum_{i=1}^{\infty} \sum_{j=1}^{N} \frac{\pi_i(\theta) J(y_j/s_i; \alpha, T, t)}{s_i} f_A(y_j; \theta, \alpha, T),
\]

\[
U_i = U_i(\theta, \alpha, T) = \sum_{j=1}^{N} \frac{\pi_j(\theta) \exp\left(\frac{T y_j}{s_i}\right)}{s_i} f_A(y_j; \theta, \alpha, T),
\]
In the above
\[
J_i(y) := \int_0^y e^{T_i(y-u)} t_i \alpha e^{T_iu} du.
\]

Proposition
Let \( c = \max\{ T_{kk} : k = 1, \ldots, p \} \) and set \( K := c^{-1} T + I \). Define
\[
D(s) := D(s; \alpha, T) = \frac{1}{c} \sum_{r=0}^s K^r \alpha K^{s-r}
\]
Then
\[
J(y; \theta, \alpha, T) = \sum_{s=0}^{\infty} \frac{(cy)^{s+1} e^{-cy}}{(s + 1)!} D(s).
\]
An alternative model
Let $r \in (0, 1)$.

$$f_B(y) = r \cdot \alpha e^{T_y t} + (1 - r) \sum_{i=1}^{\infty} \pi_i(\theta) \frac{\lambda_i^q y^{q-1} e^{-\lambda_i y}}{\Gamma(q - 1)},$$
Proposition

The EM estimators take the form

\[ \hat{r} = \frac{r\alpha Ut}{N} \]
\[ \hat{\alpha} = \frac{\text{diag} (\alpha) Ut}{\alpha Ut} \]
\[ \hat{T}_{k\ell} = \left( \text{diag} (R)^{-1} T \circ R' \right)_{k\ell} \]
\[ \hat{t} = \alpha U \text{diag} (t) \cdot \text{diag} (R)^{-1} \]
\[ \hat{\theta} = \arg\max_{\Theta} \sum_{i=1}^{\infty} w_i \log \pi_i (\Theta) \]
\[ \hat{\lambda} = \frac{\sum_{i=1}^{\infty} w_i}{\sum_{i=1}^{\infty} \frac{v_i}{s_i}}. \]
where

\[
R = R(\alpha, T, \lambda, \theta) = \sum_{j=1}^{N} \frac{J(y_j; \alpha, T, t)}{f_B(y_j; r, \alpha, T, \theta, \lambda)}.
\]

\[
U = U(\alpha, T, \lambda, \theta) = \sum_{j=1}^{N} \frac{\exp(Ty_j)}{f_B(y_j; r, \alpha, T, \theta, \lambda)}.
\]

\[
w_i = w_i(\alpha, T, \lambda, \theta) = \pi_i(\theta) \sum_{j=1}^{N} \frac{g(y_j; q, \lambda/s_i)}{f_B(y_j; r, \alpha, T, \theta, \lambda)}
\]

\[
v_i = v_i(\alpha, T, \lambda, \theta) = \pi_i(\theta) \sum_{j=1}^{N} \frac{y_j \cdot g(y_j; q, \lambda/s_i)}{f_B(y_j; r, \alpha, T, \theta, \lambda)}
\]
Advantages

▶ Fast.
▶ Recovers the parameters of simulated data.
▶ Works reasonable well with real data.
▶ Converges to the MLE estimators most of the times.
▶ Provides a better approximation for the tails.

Disadvantages

▶ There are some (minor) numerical issues yet to be resolved.
▶ MLE estimators are not appropriate for estimating “tail parameters”. EVT methods could be implemented.
▶ Sensitive with respect to the tail parameters.
▶ Kullback-Leibler does not work properly. Though we now have the first simpler approach.