Sums of lognormal random variables

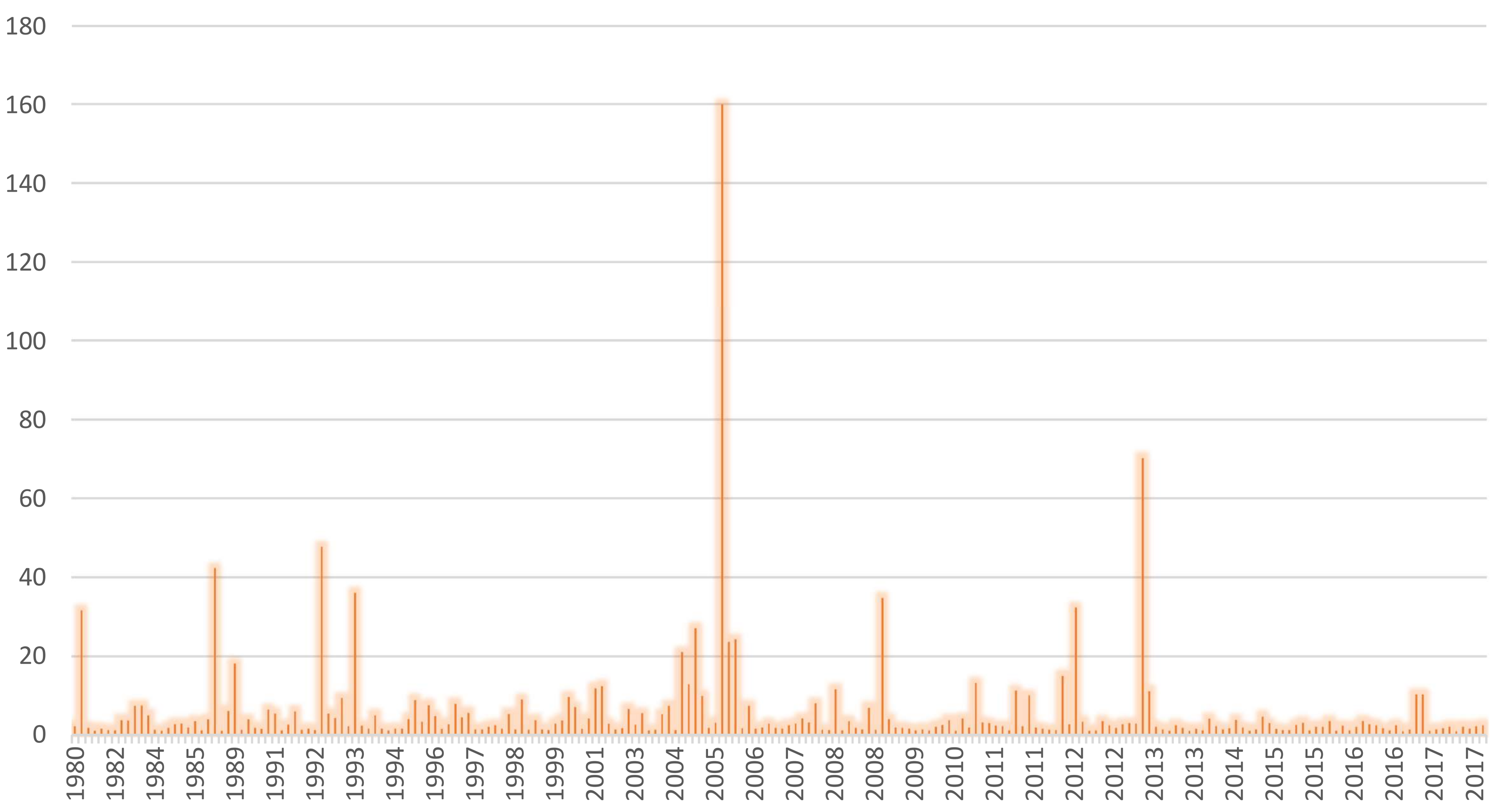
University of Queensland and Aarhus University Denmark

Patrick J. Laub

p.laub@[uq.edu.au|math.au.dk]

A Most Gentle Introduction

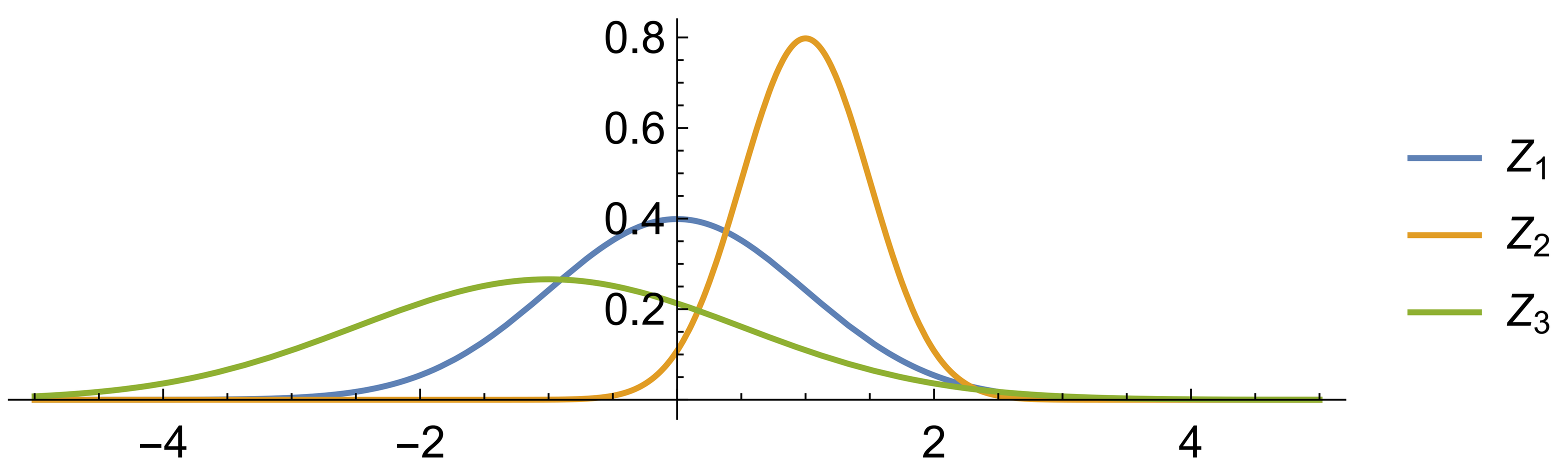
The following plot shows the estimated financial cost of some large natural disasters in the USA. What do you think is the most striking feature about the plot?



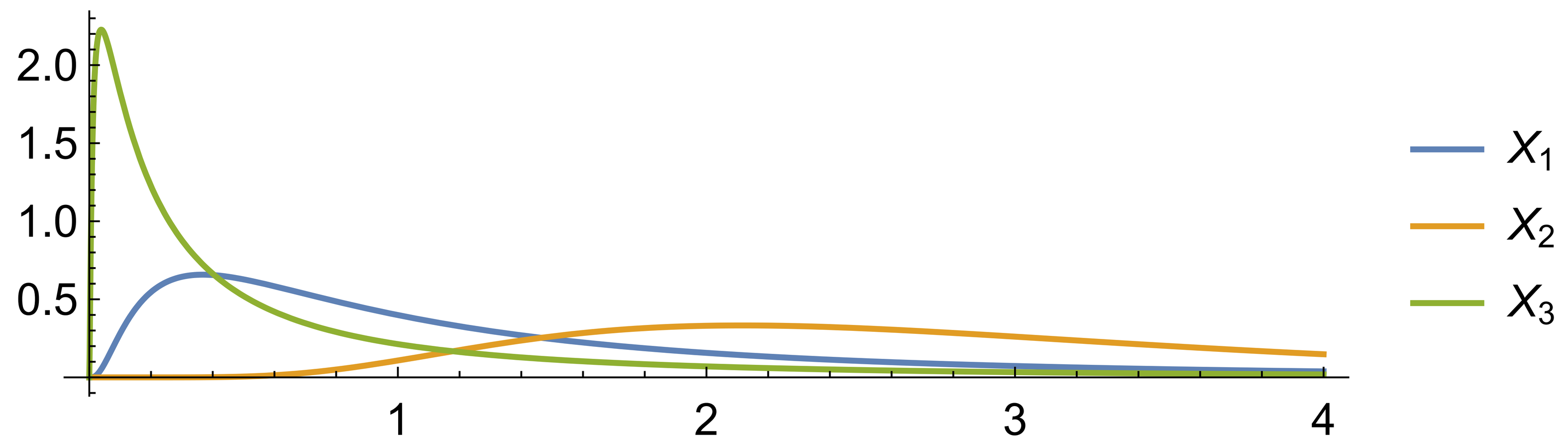
I hope you noticed that it is remarkably spiky. Most of these disasters were near the \$1bn threshold, but a few (such as the biggest spike for Hurricane Katrina) were much more massive. This behaviour makes life difficult for an insurer, as it must set premiums high enough that an event like Hurricane Katrina would not send them bankrupt, but not too high as to lose all their customers.

Thus the insurer would like to know the distribution of the total cost of a disaster, which is simply the aggregation of many individual insurance claims. Say the first claim is of size X_1 dollars, the second X_2 dollars, and the last is X_d dollars, so the total payout for the insurer is $S = X_1 + \cdots + X_d$ dollars.

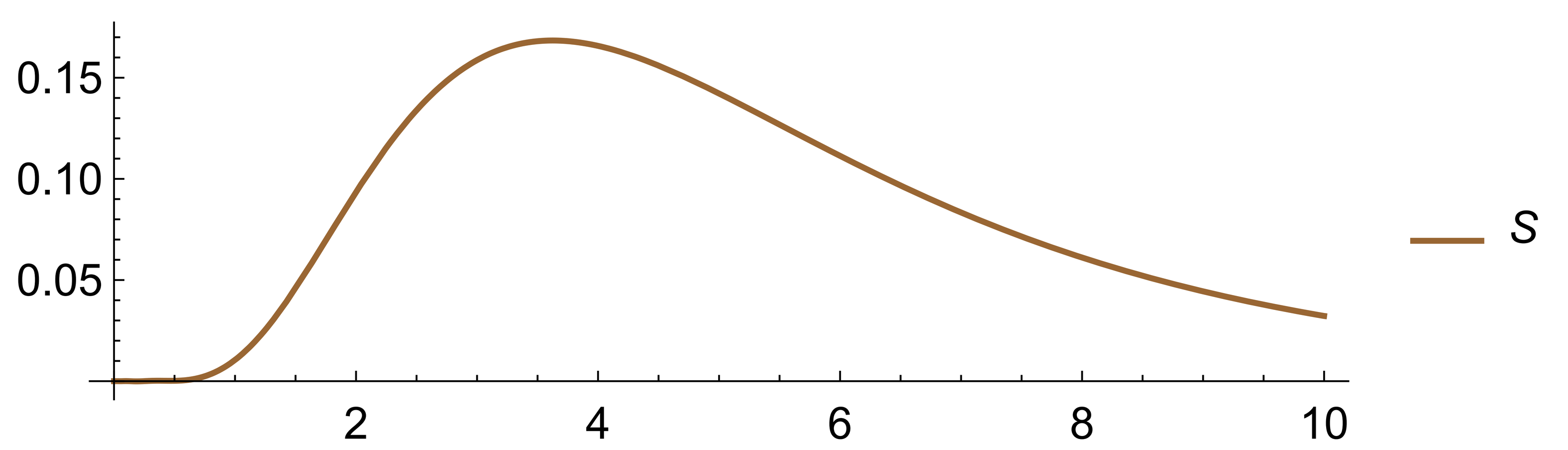
How do we model the distribution of the X_i's? You may have been taught that everything is a bell curve$^{\text{TM}}$(a.k.a. the normal distribution). Remember these densities look like:



However the normal distribution doesn't work here. It won't generate the spikiness we observed, and it would allow for X_i's which are negative. Instead, we do a common trick which is to say $X_i = \exp\{Z_i\}$ where the Z_i's are normally distributed. Then the densities of the X_i's are:



This makes the X_i's always positive and 'heavy-tailed' (that is our jargon for spikey); we say these X_i's are lognormally distributed. The density of the sum $S = X_1 + X_2 + X_3$ is:



Our approximation was used to make this last curve since there is no simple density for the sum of lognormals, and we desperately want one for insurance/financial/engineering applications.

The approximation

Say $f(x)$ is the lognormal sum density that we wish to estimate. We approximate the expansion

$$f(x) \;=\; f_\nu(x)\sum_{k=0}^{\infty} a_k Q_k(x) \;\approx\; f_\nu(x)\sum_{k=0}^{K} \widehat{a}_k Q_k(x)\,,$$

where $f_\nu(x)$ is a reference density, the a_k's and $\widehat{a}_k$'s are coefficients in $\mathbb{R}$, the Q_k's are special polynomials of order k, and $K \in \mathbb{N}$.

We first select an f_ν which then determines the appropriate Q_k to use, then we must choose a method to approximate a_k with $\widehat{a}_k$, and a truncation order K. We considered f_ν to be normally, gamma, and lognormally distributed.

Choosing an f_ν and the resulting Q_k

The reference density must satisfy $f/f_\nu \in L^2(\nu)$ which just means that $\int_{\mathbb{R}} f(x)^2/f_\nu(x)\,\mathrm{d}x < \infty$. The Q_k polynomials must then be orthonormal w.r.t. the reference density, i.e. for all $i, j \in \mathbb{N}$

$$\int_{\mathbb{R}} Q_i(x)Q_j(x)f_\nu(x)\,\mathrm{d}x \;=\; \langle Q_i, Q_j\rangle_\nu \;=\; \mathbb{I}_{\{i=j\}}\,.$$

1) Normal reference distribution

If we choose $f_\nu(x)$ to be the density of a $\mathsf{Normal}(\mu,\sigma^2)$ distribution, i.e.

$$f_\nu(x) \;\propto\; \exp\{-(x-\mu)^2/(2\sigma^2)\} \quad \text{then} \quad Q_k(x) \;\propto\; H_k\big((x-\mu)/(\sigma\sqrt{2})\big),$$

where $\{H_k\}_{k\in\mathbb{N}}$ are the (physicists') Hermite polynomials. This particular expansion is also called a Gram–Charlier expansion, or in some applications the Edgeworth expansion.

2) Gamma reference distribution

If we choose $f_\nu(x)$ to be the density of a $\mathsf{Gamma}(r,m)$ distribution, i.e.

$$f_\nu(x) \;\propto\; x^{r-1}\mathrm{e}^{-x/m} \quad \text{then} \quad Q_k(x) \;\propto\; L_k^{r-1}(x/m),$$

where $\{L_k^{r-1}\}_{k\in\mathbb{N}}$ denote the generalised Laguerre polynomials. Unfortunately, we can never satisfy the $f/f_\nu \in L^2(\nu)$ requirement, but we can get $f_\theta/f_\nu \in L^2(\nu)$ where

$$f_\theta(x) \;=\; \mathrm{e}^{-\theta x}f(x)/\mathbb{E}[\mathrm{e}^{-\theta S}] \quad \text{so} \quad f(x) \;\approx\; \mathrm{e}^{\theta x}\,\mathbb{E}[\mathrm{e}^{-\theta S}]\,f_\nu(x)\sum_{k=0}^{K}\widehat{a}_k Q_k(x)\,.$$

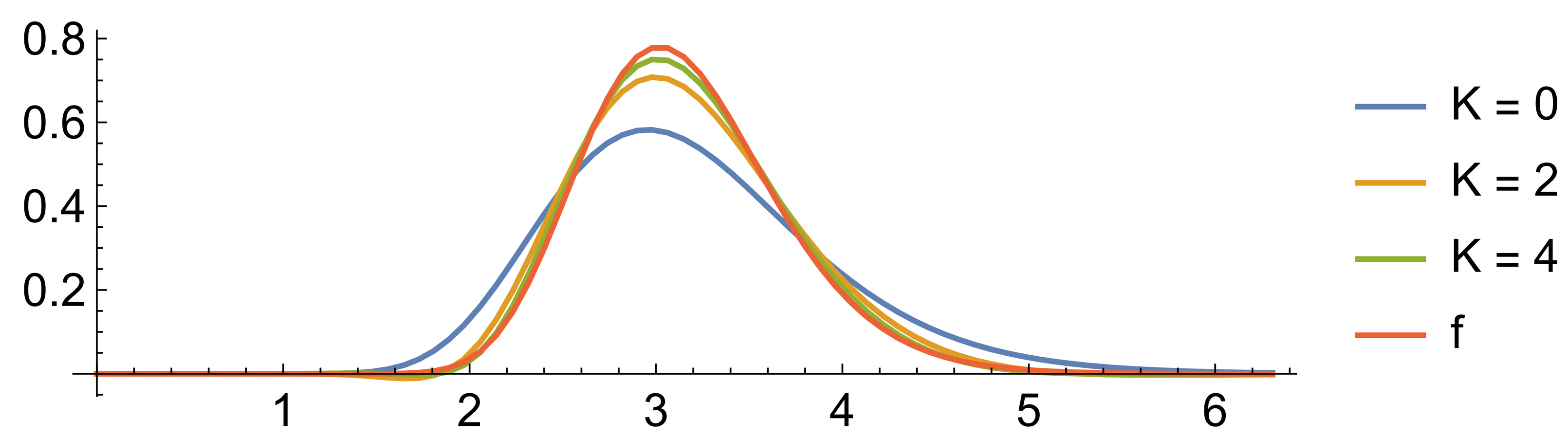
The f_θ is called the exponentially-tilted distribution. We estimate $\mathbb{E}[\mathrm{e}^{-\theta S}]$ using Laub et al. (2016).

3) Lognormal reference distribution

We tried to let $f_\nu(x)$ to be the density of a $\mathsf{LogNormal}(\mu,\sigma^2)$ distribution, and we were the first to describe the associated orthonormal polynomial system. However, we have proved that the resulting approximation does not converge as K increases (this relates to the moment problem for lognormals).

Estimating the a_k's and choosing a K

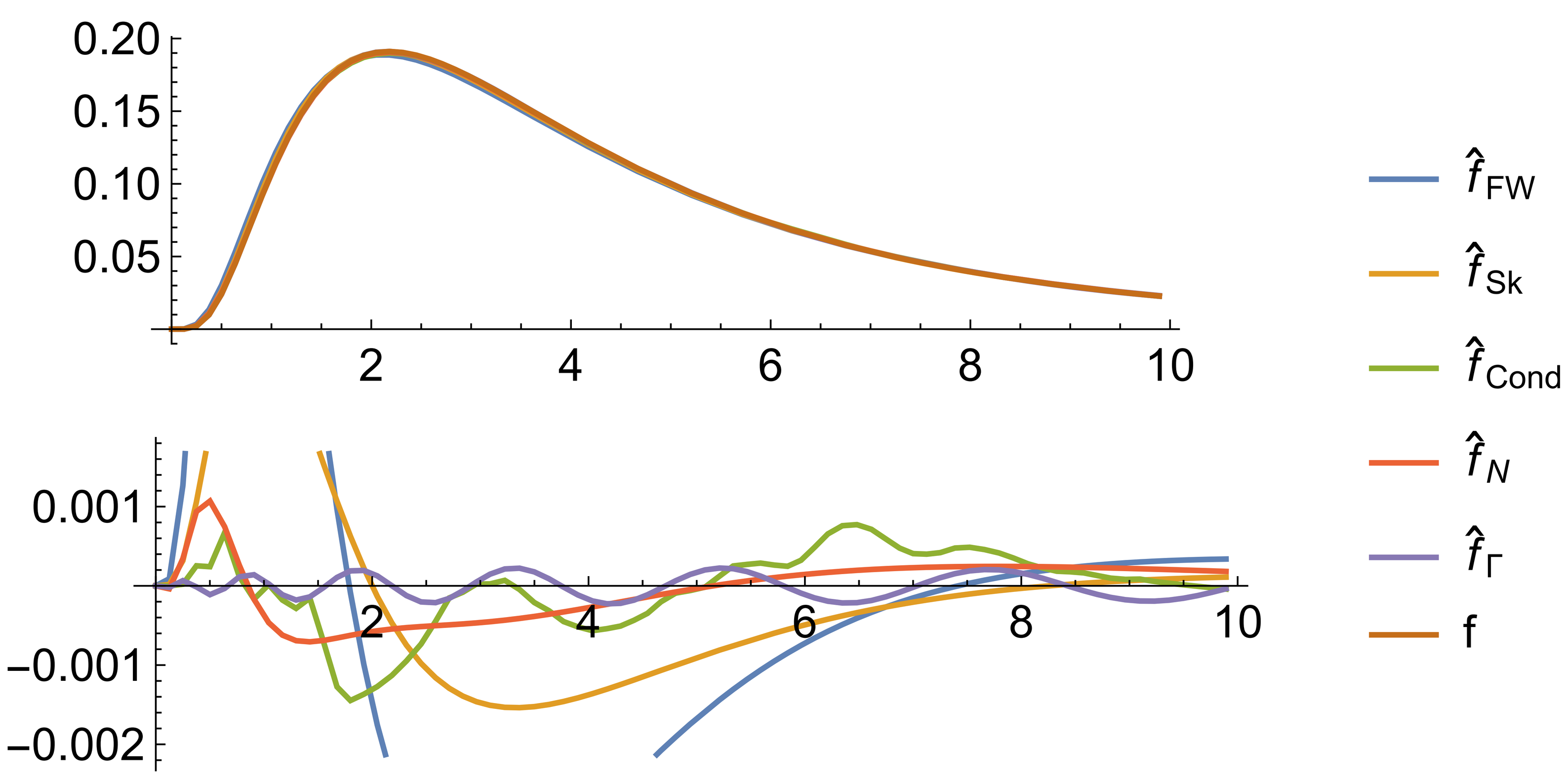
The exact coefficients are $a_k = \langle f/f_\nu, Q_k\rangle_\nu = \mathbb{E}[Q_k(S)]$. If we use these a_k's and let $K = \infty$ then the approximation is exact. Yet even for small K the truncation error can disappear quickly:



The coefficients of the polynomial expansion, $\{a_k\}_{k\in\mathbb{N}}$, tend toward 0 as $k \to \infty$, thus the size of the truncation error depends upon how swiftly the coefficients decay. We use approximate coefficients $\widehat{a}_k$ found by (quasi-) Monte Carlo integration on $\mathbb{E}[Q_k(S)]$.

Results

We compare against 3 other methods denoted $\widehat{f}_{\mathrm{FW}}$, $\widehat{f}_{\mathrm{Sk}}$, and $\widehat{f}_{\mathrm{Cond}}$. An example test is:



	$\widehat{f}_{\mathrm{FW}}$	$\widehat{f}_{\mathrm{Sk}}$	$\widehat{f}_{\mathrm{Cond}}$	$\widehat{f}_N$	$\widehat{f}_\Gamma$
Mean Squared Error	9.48×10^{-3}	3.71×10^{-3}	1.60×10^{-3}	1.18×10^{-3}	3.53×10^{-4}

Applications

Since the density approximation is very simple — it is a combination of polynomials and the reference density — closed-form solutions quantities of interest should be simple. For example, Dufresne and Li (2014) use this to price Asian options.

References

- Asmussen, S., Goffard, P.O., Laub, P.J. (2017) 'Orthonormal polynomial expansions and lognormal sum densities'. To appear in the Festschrift for Ragnar Norberg.
- Dufresne, D., & Li, H. (2014). 'Pricing Asian options: Convergence of Gram-Charlier series.'
- Botev, Z., Salomone, R., & MacKinlay, D. (2017). 'Fast and Accurate Computation of the Distribution of Sums of Dependent Log-Normals'. arXiv preprint arXiv:1705.03196.
- Laub, P.J., Asmussen, S., Jensen, J.L., Rojas-Nandayapa, L. (2016) 'Approximating the Laplace transform of the sum of dependent lognormals'. Advances in Applied Probability, 48(A), pp. 203–215.