Weak Limit of Poisson-Mixture Sums of Independent but not Identically Distributed Random Variables via Stein's Method

Uwe Schmock

Based on work in progress with Peter Eichelsbacher and Piet Porkert

Financial and Actuarial Mathematics
Vienna University of Technology, Austria
www.fam.tuwien.ac.at

Workshop on New Directions in Stein's Method
National University of Singapore, May 18–29, 2015

Slides: May 25, 2015

The collective model from actuarial mathematics

Let $(X_n)_{n\in\mathbb{N}}$ denote i.i.d. claims, independent of the random number N of claims. Total claim amount is the *random sum*

$$S_N = \sum_{n=1}^{N} X_n \,.$$

If $N, X_1 \in L^1(\mathbb{P})$, then (special case of Wald's equation)

$$\mathbb{E}[S_N] = \mathbb{E}\Big[\underbrace{\mathbb{E}[S_N \,|\, N]}_{\overset{\text{a.s.}}{=} N\,\mathbb{E}[X_1]}\Big] = \mathbb{E}[N]\,\mathbb{E}[X_1]\,.$$

If $N, X_1 \in L^2(\mathbb{P})$, then (Blackwell–Girshick equation)

$$\begin{aligned}\operatorname{Var}(S_N) &= \mathbb{E}\Big[\underbrace{\operatorname{Var}(S_N \,|\, N)}_{\overset{\text{a.s.}}{=} N\operatorname{Var}(X_1)}\Big] + \operatorname{Var}\Big(\underbrace{\mathbb{E}[S_N \,|\, N]}_{\overset{\text{a.s.}}{=} N\,\mathbb{E}[X_1]}\Big)\\ &= \mathbb{E}[N]\operatorname{Var}(X_1) + \operatorname{Var}(N)\,(\mathbb{E}[X_1])^2.\end{aligned}$$

Calculating the distribution of the random sum S_N

If X_1 takes values in $\mathbb{N}_0^d$, then there is a fast algorithm (the extended Panjer recursion) to compute the distribution of the random sum S_N, provided $\mathcal{L}(N) = \text{Panjer}(a, b, k)$.

Definition

A probability distribution $(q_n)_{n\in\mathbb{N}_0}$ is called Panjer(a, b, k) with $a, b \in \mathbb{R}$ and $k \in \mathbb{N}_0$ if $q_0 = q_1 = \cdots = q_{k-1} = 0$ and

$$q_n = \left(a + \frac{b}{n}\right) q_{n-1} \qquad \text{for all } n \in \mathbb{N} \text{ with } n \geq k+1.$$

Determination of all these distributions:

- $k = 0$: Sundt and Jewell (1981)
- $k = 1$: Willmot (1988)
- General $k \in \mathbb{N}_0$: Hess, Liewald and Schmidt (2002)

Basic Panjer class distributions

- Bin(m, p) = Panjer$(\frac{p}{1-p}, -\frac{m+1}{1-p}p, 0)$ with $m \in \mathbb{N}$ and $p \in [0, 1)$
- Poi(λ) = Panjer$(0, \lambda, 0)$ with $\lambda \geq 0$
- NegBin(α, p) = Panjer$(p, (\alpha - 1)p, 0)$ with $\alpha > 0$ and $p \in [0, 1)$
- Log(p) = Panjer$(p, -p, 1)$ with $p \in [0, 1)$ and $q_n = -\frac{p^{n-1}}{c(p)\,n}$ for all $n \in \mathbb{N}$ with $c(p) := -\frac{\log(1-p)}{p}$
- Extended logarithmic distribution: Given $k \in \mathbb{N} \setminus \{1\}$ and $p \in (0, 1]$, define $q_0 = \cdots = q_{k-1} = 0$ and

$$q_n = \frac{\binom{n}{k}^{-1} p^n}{\sum_{l=k}^{\infty} \binom{l}{k}^{-1} p^l} \quad \text{for } n \geq k.$$

  ExtLog(k, p) = Panjer$(p, -kp, k)$, has heavy tails for $p = 1$. Closed-form expression for the series is available.

Basic Panjer class distributions (cont.)

- Extended negative binomial distribution: For $k \in \mathbb{N}$, $\alpha \in (-k, -k+1)$ and $p \in (0,1]$ define $q_0 = \cdots = q_{k-1} = 0$ and

$$q_n = \frac{\binom{\alpha+n-1}{n} p^n}{(1-p)^{-\alpha} - \sum_{j=0}^{k-1} \binom{\alpha+j-1}{j} p^j} \quad \text{for } n \geq k.$$

ExtNegBin$(\alpha, k, p) =$ Panjer$(p, (\alpha-1)p, k)$. It has heavy tails for $p = 1$, which is good for reinsurance companies.

Theorem (Hess, Liewald and Schmidt, 2002)

Let $Q = (q_n)_{n \in \mathbb{N}_0}$ be non-degenerate. Then are equivalent:

- *Q is in* Panjer(a, b, k).
- *Q is the k-truncation of a basic* Panjer(a, b, k') *distribution $Q' = (q'_n)_{n \in \mathbb{N}_0}$ with $k' \leq k$ and $c := \sum_{n=k}^{\infty} q'_n > 0$, i.e., $q_n = 0$ for $n \in \{0, 1, \ldots, k-1\}$ and $q_n = q'_n / c$ for all $n \geq k$.*

Numerical stability of Panjer's recursion

Panjer's recursion is certainly numerically stable when

$$a + \frac{bj}{n} \geq 0 \quad \text{for all } j \in \{1, \ldots, n\}.$$

This is the case when $a \geq 0$ and $b \geq -a$, hence for

- Poisson distribution,
- Negative binomial distribution,
- Logarithmic distribution,
- Truncations of the above.

It is potentially unstable for

- Binomial distribution,
- Extended negative binomial distribution,
- Extended logarithmic distribution.

Application: Poisson–tempered-α-stable mixtures

Definition (τ-tempered α-stable distribution $F_{\alpha,\sigma,\tau}$)

For index $\alpha \in (0,1)$, scale $\sigma > 0$ and tempering $\tau \geq 0$ define

$$F_{\alpha,\sigma,\tau}(y) := \mathbb{E}[e^{-\tau Y} 1_{\{Y \leq y\}}] / \mathbb{E}[e^{-\tau Y}], \quad y \in \mathbb{R}.$$

where Y is α-stable on $[0,\infty)$ with Laplace transform $\mathbb{E}[\exp(-sY)] = \exp(-\gamma_{\alpha,\sigma} s^\alpha)$ for $s \geq 0$, where $\gamma_{\alpha,\sigma} = \frac{\sigma^\alpha}{\cos(\alpha\pi/2)}$.

Theorem (Gerhold, S., Warnung, 2010)

Let $\Lambda \sim F_{\alpha,\sigma,\tau}$ *and* $\mathcal{L}(N|\Lambda) \overset{a.s.}{=} \mathrm{Poi}(\lambda\Lambda)$ *with* $\lambda > 0$. *Then*

$$N \overset{d}{=} N_1 + \cdots + N_M$$

with independent $M \sim \mathrm{Poi}(\gamma_{\alpha,\sigma}((\lambda+\tau)^\alpha - \tau^\alpha))$ *and* $N_m \sim \mathrm{ExtNegBin}\big(-\alpha, 1, \frac{\lambda}{\lambda+\tau}\big)$ *for* $m \in \mathbb{N}$.

Application: Poisson–tempered α-stable mixtures (cont.)

Let $\Lambda \sim F_{\alpha,\sigma,\tau}$ and $\mathcal{L}(N|\Lambda) \stackrel{\text{a.s.}}{=} \mathrm{Poi}(\lambda\Lambda)$ with $\lambda > 0$. Then the stochastic representation $N \stackrel{\text{d}}{=} N_1 + \cdots + N_M$ leads to

$$S = \sum_{j=1}^{N} X_j \stackrel{\text{d}}{=} \sum_{i=1}^{M} \sum_{j=N_1+\cdots+N_{i-1}+1}^{N_1+\cdots+N_i} X_j \stackrel{\text{d}}{=} \sum_{i=1}^{M} \sum_{j=1}^{N_i} X_{i,j}\,,$$

where $(X_{i,j})_{i,j\in\mathbb{N}}$ are i.i.d. with $X_{i,j} \stackrel{\text{d}}{=} X_1$.

Algorithm (numerically stable, $\tau \neq 0$)

- *Panjer recursion for* $\tilde{N} \sim \mathrm{NegBin}\big(1-\alpha, \frac{\lambda}{\lambda+\tau}\big)$
- *Weighted convolution:* $N_1 \sim \mathrm{ExtNegBin}\big(-\alpha, 1, \frac{\lambda}{\lambda+\tau}\big)$
- *Panjer recursion for* $M \sim \mathrm{Poi}(\gamma_{\alpha,\sigma}((\lambda+\tau)^\alpha - \tau^\alpha))$

If $\tau = 0$, use the special algorithm for $N_1 \sim \mathrm{ExtNegBin}(-\alpha, 1, 1)$.

Examples for τ-tempered $\frac{1}{2}$-stable distributions

Definition (Lévy distribution with scale parameter $\sigma > 0$)

A density of $F_{1/2,\sigma,0}$ is

$$f_{\text{Lévy},\sigma}(x) = \Big(\frac{\sigma}{2\pi x^3}\Big)^{1/2} \exp\Big(-\frac{\sigma}{2x}\Big), \quad x > 0.$$

(Distribution of first hitting time of Brownian motion for level σ^2.)

Definition (inverse Gaussian distribution, parameters $\mu, \tilde{\sigma} > 0$)

Define $\sigma = \mu^2/\tilde{\sigma}^2$ and $\tau = 1/(2\tilde{\sigma}^2)$. A density of $F_{1/2,\sigma,\tau}$ is

$$f_{\text{IG},\mu,\tilde{\sigma}}(x) = \frac{\mu}{\sqrt{2\pi\tilde{\sigma}^2 x^3}} \exp\bigg(-\frac{(x-\mu)^2}{2\tilde{\sigma}^2 x}\bigg), \quad x > 0.$$

(Distribution of first hitting time of Brownian motion with drift $1/\tilde{\sigma}$ for level $\mu/\tilde{\sigma}$.)

Additional examples of probability distributions for the Poisson mixture that can be handled

- Generalized τ-tempered α-stable distributions (one additional parameter $m \in \mathbb{N}_0$)
- Inverse gamma distribution (with half-integer shape parameter)
- Generalized inverse Gaussian distribution (with additional half-integer parameter $m + \frac{1}{2}$)

With an additional convolution:

- Reciprocal generalized inverse Gaussian distribution (with additional half-integer parameter $m + \frac{1}{2}$)

References for the explicit calculation of the distribution of the random sum S_N and generalisations

- Uwe Schmock: *Modelling Dependent Credit Risks with Extensions of CreditRisk$^+$ and Application to Operational Risk*, Lecture Notes, Vienna University of Technology, 2004–2015.
- Stefan Gerhold, Uwe Schmock, and Richard Warnung: *A generalization of Panjer's recursion and numerically stable risk aggregation*, Finance & Stochastics, 14, (81–128), 2010.
- Cordelia Rudolph: *A Generalization of Panjer's Recursion for Dependent Claim Numbers and an Approximation of Poisson Mixture Models*, Ph.D. thesis, Vienna University of Technology, Austria, 2014.

Available via my home page.

Lower quantile (value-at-risk) and Kolmogorov metric

Let X, Y be real-valued random variables and $\delta \in (0,1)$ a level.

1. Define the *lower δ-quantile* of X by
$$q_\delta(X) = \min\{x \in \mathbb{R} \mid \mathbb{P}[X \le x] \ge \delta\}$$
2. Define the *Kolmogorov distance* of their distributions by
$$d_{\mathrm{K}}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{z \in \mathbb{R}} \bigl|\mathbb{P}[X \le z] - \mathbb{P}[Y \le z]\bigr|.$$

Lemma (Quantiles and Kolmogorov metric)

Let X and Y be real-valued random variables and denote the Kolmogorov distance of their distributions by d. Then the lower quantiles of X and Y satisfy

1. $q_{\delta-d}(X) \le q_\delta(Y)$ for every level $\delta \in (d,1)$ and
2. $q_\delta(Y) \le q_{\delta+d}(X)$ for every level $\delta \in (0,1-d)$.

Expected shortfall as risk measure

Definition

Let X be a real-valued random variable. Then the *expected shortfall* of the loss variable X at level $\delta \in (0,1)$ is defined as

$$\mathrm{ES}_\delta[X] = \frac{\mathbb{E}\big[X 1_{\{X > q_\delta(X)\}}\big] + q_\delta(X)(\mathbb{P}[X \le q_\delta(X)] - \delta)}{1-\delta}. \qquad (2)$$

Note that $\mathrm{ES}_\delta[X] = \infty$ if $\mathbb{E}[X 1_{\{X > q_\delta(X)\}}] = \infty$.
If $\mathbb{P}[X \le q_\delta(X)] = \delta$, in particular if the distribution function of X is also left-continuous at $x = q_\delta(X)$, then (2) simplifies to

$$\mathrm{ES}_\delta[X] = \mathbb{E}[X \,|\, X > q_\delta(X)]\,.$$

When expected shortfall is taken as a risk measures, then (contrary to VaR) the sizes of large losses exceeding the threshold $q_\delta(X)$ are clearly taken into account. The additional term in (2) is necessary to prove the sub-additivity of expected shortfall (diversification).

Expected shortfall and Wasserstein metric

Let (S, ϱ) be a metric space and let X, Y be S-valued random variables. The Wasserstein distance of their distributions is defined by

$$d_{\mathrm{W}}(\mathcal{L}(X), \mathcal{L}(Y)) = \sup_{b>0} \sup_{h \in \mathcal{H}_b} \mathbb{E}[h(X) - h(Y)]\,,$$

where $\mathcal{H}_b$ denotes the set of all functions $h\colon S \to \mathbb{R}$ with $\|h\|_\infty \le b$ and Lipschitz constant $\mathrm{Lip}(h) \le 1$.

Lemma (Expected shortfall and Wasserstein metric)

Let X and Y be real-valued, integrable random variables. Then the expected shortfall of X and Y satisfies, for every level $\delta \in (0,1)$,

$$\bigl|\mathrm{ES}_\delta[X] - \mathrm{ES}_\delta[Y]\bigr| \le \frac{d_{\mathrm{W}}(\mathcal{L}(X), \mathcal{L}(Y))}{1-\delta}.$$

Definition of the one-factor Bernoulli mixture model

Consider Bernoulli random variables $X_1, \ldots, X_m$, a $[0, \infty)$-valued random variable Λ and $p_1, \ldots, p_m \in [0, 1]$ such that $\max\{p_1, \ldots, p_m\}\Lambda \leq 1$ almost surely. If

$$\mathbb{P}[X_i = 1 \,|\, \Lambda] \overset{\text{a.s.}}{=} p_i \Lambda, \qquad i \in \{1, \ldots, m\},$$

and if $X_1, \ldots, X_m$ are conditionally independent given Λ, i.e.,

$$\mathbb{P}[X_1 = x_1, \ldots, X_m = x_m \,|\, \Lambda] \overset{\text{a.s.}}{=} \prod_{i=1}^{m} \mathbb{P}[X_i = x_i \,|\, \Lambda]$$

for all $x_1, \ldots, x_m \in \{0, 1\}$, then we call $(X_1, \ldots, X_m, \Lambda)$ a *one-factor Bernoulli mixture model*.

Since $\mathbb{P}[X_i = 1] = \mathbb{E}[\mathbb{P}[X_i = 1 \,|\, \Lambda]] = p_i \, \mathbb{E}[\Lambda]$, the case $\mathbb{E}[\Lambda] = 1$ is convenient.

Convergence to Poisson-mixture distributions

In the above one-factor Bernoulli mixture model, define $W = X_1 + \cdots + X_m$ and $\lambda = p_1 + \cdots + p_m$. Consider $\mathbb{N}_0$-valued N such that $\mathcal{L}(N|\Lambda) \stackrel{\text{a.s.}}{=} \text{Poi}(\lambda\Lambda)$. For the total variation distance,

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(N)) = \sup_{f\colon \mathbb{N}_0 \to [0,1]} \mathbb{E}\big[\mathbb{E}[f(W) - f(N)|\Lambda]\big]$$

and almost surely

$$\mathbb{E}[f(W) - f(N)|\Lambda] \le d_{\text{TV}}\big(\mathcal{L}(W|\Lambda), \mathcal{L}(N|\Lambda)\big) \le \frac{1 - e^{-\lambda\Lambda}}{\lambda\Lambda} \sum_{i=1}^{m} (p_i\Lambda)^2,$$

using the classical estimate by Barbour & Hall (1984). Hence

$$d_{\text{TV}}(\mathcal{L}(W), \mathcal{L}(N)) \le \frac{\mathbb{E}[\Lambda(1 - e^{-\lambda\Lambda})]}{\lambda} \sum_{i=1}^{m} p_i^2.$$

Similarly for the Wasserstein distance.
(Implications and improvements are ongoing work with Larry Goldstein.)

Definition of a normal variance mixture distribution

Definition

An $\mathbb{R}^d$-valued random vector Z has a *normal variance mixture distribution* with parameters $\mu \in \mathbb{R}^d$ and $\Sigma \in \mathbb{R}^{d\times d}$ if there exist

1. a dimension $k \in \mathbb{N}$,
2. a matrix $A \in \mathbb{R}^{d\times k}$ with $AA^\top = \Sigma$,
3. a k-dimensional standard normally distributed random vector X, and
4. a variance mixture variable $\Lambda \geq 0$, independent of X,

satisfying the stochastic representation

$$Z \stackrel{d}{=} \mu + \sqrt{\Lambda} A X \,.$$

Examples of normal variance mixture distributions

1. Let Λ be constant. Then $Z \sim \mathcal{N}(\mu, \Lambda\Sigma)$.
2. Let $\Lambda \sim \Gamma(\alpha, \beta)$ with $\alpha, \beta > 0$, then Z has a normal variance-gamma mixture distribution. The special case $\alpha = 1$ corresponds to the d-dimensional Laplace distribution centered at μ.
3. Let $1/\Lambda \sim \Gamma(\alpha, \beta)$ with $\alpha, \beta > 0$. The special case $\alpha = \beta = \nu/2$ corresponds to a d-dimensional t-distribution $t_d(\nu, \mu, \Sigma)$, with location vector $\mu \in \mathbb{R}^d$, dispersion matrix Σ and $\nu > 0$ degrees of freedom. The special case $\nu = 1$ corresponds to the d-dimensional Cauchy distribution.
4. Suppose that Λ has a τ-tempered α-stable distribution with index $\alpha \in (0, 1)$, scale parameter $\sigma > 0$ and tempering parameter $\tau \geq 0$. The special case $\alpha = 1/2$ and $\tau = 0$ corresponds to a Lévy distribution with scale parameter σ.

Setting

- Let $(X_n)_{n\in\mathbb{N}}$ be independent (not necessarily identically distributed) real-valued square-integrable random variables with $\mathbb{E}[X_n]=0$ and $\mathrm{Var}(X_n)=1$ for all $n\in\mathbb{N}$.
- Let Λ be a $(0,\infty)$-valued random variable.
- Given $\lambda>0$, let $\mathcal{L}(N_\lambda|\Lambda)\overset{\text{a.s.}}{=}\mathrm{Poi}(\lambda\Lambda)$ and assume that (N_λ,Λ) is independent of $(X_n)_{n\in\mathbb{N}}$.
- Define the Poisson-mixture sums

$$Z_\lambda=\frac{1}{\sqrt{\lambda}}\sum_{n=1}^{N_\lambda}X_n\,,\qquad \lambda>0.$$

- Define the limiting variable $Z=\sqrt{\Lambda}\,X$, where $X\sim\mathcal{N}(0,1)$ is independent of $(X_n)_{n\in\mathbb{N}}$ and (N_λ,Λ).

Weighted average of third moments

For $\lambda > 0$ we define

$$\varrho(\lambda) := e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^{n-1}}{n!} \sum_{i=1}^{n} \mathbb{E}\big[|X_i|^3\big] \in [1, \infty].$$

Given $P_\lambda \sim \mathrm{Poi}(\lambda)$, then $\mathbb{E}[P_\lambda] = \lambda$ and we can rewrite

$$\varrho(\lambda) = \frac{1}{\mathbb{E}[P_\lambda]} \mathbb{E}\Bigg[\sum_{i=1}^{P_\lambda} \mathbb{E}\big[|X_i|^3\big]\Bigg].$$

We note that

$$\varrho(\lambda) \leq \sup_{i \in \mathbb{N}} \mathbb{E}\big[|X_i|^3\big]$$

for all $\lambda > 0$.

Theorem (Eichelsbacher, Porkert, S.)

Let $Y_\lambda := Z_\lambda/\sqrt{\Lambda}$. On the set $\{\varrho(\lambda\Lambda) < \infty\}$ the conditional expectation $\mathbb{E}\big[|Y_\lambda|\,\big|\Lambda\big]$ is a.s. finite, the Wasserstein distance of $\mathcal{L}(Y_\lambda\,|\,\Lambda)$ to the standard normal distribution $\mathcal{L}(X) = \mathcal{N}(0,1)$ is a.s. well defined, and

$$d_{\mathrm{W}}\big(\mathcal{L}(Y_\lambda|\Lambda), \mathcal{L}(X)\big) \le \frac{4 + 2\varrho(\lambda\Lambda)}{\sqrt{\lambda\Lambda}} \quad \text{a.s.} \tag{3}$$

If $\mathbb{E}[\varrho(\lambda\Lambda)] < \infty$, then

$$d_{\mathrm{W}}\big(\mathcal{L}(Z_\lambda), \mathcal{L}(Z)\big) \le \frac{4 + 2\,\mathbb{E}[\varrho(\lambda\Lambda)]}{\sqrt{\lambda}}. \tag{4}$$

Proof of (3) $\implies$ (4)

Assume that $\mathbb{E}[\varrho(\lambda\Lambda)] < \infty$. Then, for every bounded $h\colon \mathbb{R} \to \mathbb{R}$ with $\mathrm{Lip}(h) \le 1$,

$$\begin{aligned}
\mathbb{E}[h(Z_\lambda) - h(Z)] &= \mathbb{E}\big[\mathbb{E}[h(Z_\lambda) - h(Z)\,|\,\Lambda]\big] \\
&\le \mathbb{E}\big[d_{\mathrm{W}}\big(\mathcal{L}(Z_\lambda|\Lambda), \mathcal{L}(Z|\Lambda)\big)\big] \\
&= \mathbb{E}\big[d_{\mathrm{W}}\big(\mathcal{L}(\sqrt{\Lambda}Y_\lambda|\Lambda), \mathcal{L}(\sqrt{\Lambda}X|\Lambda)\big)\big] \\
&= \mathbb{E}\big[\sqrt{\Lambda}\, d_{\mathrm{W}}\big(\mathcal{L}(Y_\lambda|\Lambda), \mathcal{L}(X)\big)\big],
\end{aligned}$$

where we used the scaling property of the Wasserstein distance in the last step.
Plugging in (3), the upper bound (4) follows.

Remarks

- The scaled random sum Z_λ and the approximation $Z = \sqrt{\Lambda}X$ are coupled through their joint dependence on Λ. This makes the bound (4) possible even when Z_λ and Z are not integrable.
- Due to this coupling we do not need a special Stein equation for the limiting normal variance mixture distribution, it is enough to have the Stein equation for $\mathcal{N}(0,1)$ to prove (3).
- The assumptions on $\varrho(\lambda\Lambda)$ of the theorem and the corollaries below are trivially satisfied if $\sup_{i\in\mathbb{N}} \mathbb{E}[|X_i|^3] < \infty$, because this is an upper bound for the function ϱ.
- The proof of (3) uses the standard techniques combined with size biasing.

Corollaries

1. If $\mathbb{E}[\sqrt{\Lambda}] < \infty$, then Z is integrable. If, in addition, $\mathbb{E}[\varrho(\lambda\Lambda)] < \infty$ for some $\lambda > 0$, then Z_λ is integrable too, the Wasserstein distance of $\mathcal{L}(Z_\lambda)$ and $\mathcal{L}(Z)$ is finite, and
$$d_{\mathrm{W}}(\mathcal{L}(Z_\lambda), \mathcal{L}(Z)) := \sup_{h\in\mathcal{H}} \mathbb{E}[h(Z_\lambda) - h(Z)] \le \frac{4 + 2\,\mathbb{E}[\varrho(\lambda\Lambda)]}{\sqrt{\lambda}},$$
where $\mathcal{H}$ denotes the set of all $h\colon \mathbb{R} \to \mathbb{R}$ with $\mathrm{Lip}(h) \le 1$.
2. If $\mathbb{E}[\varrho(\lambda\Lambda)] = o(\sqrt{\lambda})$ as $\lambda \to \infty$, then Z_λ converges weakly to Z. Using the Cramér–Wold theorem, this transfers to the multi-dimensional setting.
3. Suppose that $\mathbb{E}[\sqrt{\Lambda}] < \infty$. Let $(\lambda_n)_{n\in\mathbb{N}}$ be a sequence in $(0, \infty)$ with $\lambda_n \to \infty$ as $n \to \infty$. If $\mathbb{E}[\varrho(\lambda_n\Lambda)] = o(\sqrt{\lambda_n})$ as $n \to \infty$, then $\{Z_{\lambda_n}\}_{n\in\mathbb{N}}$ is uniformly integrable.

Example illustrating growing third absolute moments (1)

Given $\gamma > 0$, consider independent random variables $(X_n)_{n\in\mathbb{N}}$ with $\mathbb{P}[X_n = e^{\gamma n}] = \mathbb{P}[X_n = -e^{\gamma n}] = e^{-2\gamma n}/2$ and $\mathbb{P}[X_n = 0] = 1 - e^{-2\gamma n}$ for every $n \in \mathbb{N}$.
Then $\mathbb{E}[X_n] = 0$, $\mathbb{E}[X_n^2] = 1$ and $\mathbb{E}[|X_n|^3] = e^{\gamma n}$ for every $n \in \mathbb{N}$.
For $\lambda > 0$ consider $P_\lambda \sim \mathrm{Poi}(\lambda)$. By summing the first P_λ terms of a geometric progression with factor e^γ,

$$\sum_{n=1}^{P_\lambda} \mathbb{E}\big[|X_n|^3\big] = \frac{e^{\gamma P_\lambda} - 1}{1 - e^{-\gamma}},$$

hence by the definition of $\varrho(\lambda)$, using $\frac{e^x - 1}{x} \le e^x$ for $x > 0$,

$$\varrho(\lambda) = \frac{\mathbb{E}[e^{\gamma P_\lambda}] - 1}{\lambda(1 - e^{-\gamma})} = \frac{\exp(\lambda(e^\gamma - 1)) - 1}{\lambda(1 - e^{-\gamma})} \le e^\gamma \exp(\lambda(e^\gamma - 1)).$$

Example illustrating growing third absolute moments (2)

Therefore, $\varrho(\lambda\Lambda) < \infty$ for every $(0,\infty)$-valued random variable Λ and the conditional approximation (3) applies.
If Λ has a gamma distribution $\Gamma(\alpha,\beta)$ with shape parameter $\alpha > 0$ and inverse scale parameter $\beta > 0$, then

$$\mathbb{E}[\varrho(\lambda\Lambda)] \le e^\gamma\,\mathbb{E}[\exp(\lambda\Lambda(e^\gamma - 1))] = e^\gamma \left(\frac{\beta}{\beta - \lambda(e^\gamma - 1)}\right)^\alpha$$

for $\lambda < \beta/(e^\gamma - 1)$. Hence the approximation (4) with the normal variance-gamma mixture applies for these positive λ, and also Corollary 1, because $\mathbb{E}[\sqrt{\Lambda}] = \Gamma(\alpha + 1/2)/(\sqrt{\beta}\,\Gamma(\alpha)) < \infty$.
However, the third absolute moments are increasing too fast to make the convergence of Corollary 2 applicable. No surprise:

$$\mathbb{E}\left[\sum_{n=1}^{N_\lambda} 1_{\{X_n \neq 0\}}\right] \le \sum_{n=1}^{\infty} \mathbb{P}[X_n \neq 0] = \frac{1}{e^{2\gamma} - 1}, \qquad \lambda > 0,$$

for every distribution of the $(0,\infty)$-valued Λ.

Approximation of random sums in the Kolmogorov metric

Theorem (Eichelsbacher, Porkert, S.)

Let $c := \sup_{i\in\mathbb{N}} \mathbb{E}\big[|X_i|^3\big] < \infty$. *For every* $z \in \mathbb{R}$,

$$\Big|\mathbb{P}[Z_\lambda \le z\,|\,\Lambda] - \mathbb{P}[\sqrt{\Lambda}X \le z\,|\,\Lambda]\Big|$$

$$\le \frac{1}{2\sqrt{\lambda\Lambda}}\left(\sqrt{\varrho(\lambda\Lambda)\,\mathbb{E}\left[\frac{1}{\lambda\Lambda}\sum_{n=1}^{N_\lambda+1}\mathbb{E}[|X_n|^3]\,\middle|\,\Lambda\right]} + \sqrt{\pi}\frac{4+\varrho(\lambda\Lambda)}{2\sqrt{2}} + 4\right)$$

$$+\left(1+\frac{\lambda\Lambda}{2}\right)e^{-\lambda\Lambda} + \frac{\sqrt{2}(c/2+2)+(1+\sqrt{2})c}{\sqrt{\lambda\Lambda}}\sqrt{\mathbb{E}[\Lambda]} \quad a.s.$$

Weak limit theory for random sums

- The weak limit behavior of random sums is well studied, see

B. V. Gnedenko and V. Y. Korolev:
Random Summation, Limit Theorems and Applications,
CRC Press, Boca Raton, FL, 1996.

- Normal variance mixture distributions are not surprising as limit distributions of random sums (transfer theorem).
- Normal variance mixture distributions have natural applications, e.g. in mathematical finance (fat tales).

Theorem (Transfer theorem)

Let $(X_i)_{i\in\mathbb{N}}$ be independent random variables and denote $S_n := \sum_{i=1}^n X_i$ for $n \in \mathbb{N}$. Let the sequences of real numbers $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$, $(c_n)_{n\in\mathbb{N}}$, $(d_n)_{n\in\mathbb{N}}$ be such that $b_n, d_n > 0$, $n \in \mathbb{N}$, and $b_n, d_n \to \infty$ as $n \to \infty$ and $Y_n := (S_n - a_n)/b_n \xrightarrow{d} Y$, as $n \to \infty$, for some random variable Y with distribution function F. Let $(N_n)_{n\in\mathbb{N}}$ be $\mathbb{N}$-valued random variables independent of $(X_i)_{i\in\mathbb{N}}$, such that

$$\Big(\frac{b_{N_n}}{d_n}, \frac{a_{N_n} - c_n}{d_n}\Big) \xrightarrow{d} (U, V), \quad n \to \infty,$$

for some random variables U and V. Then

$$\mathbb{P}\Big[\frac{S_{N_n} - c_n}{d_n} < x\Big] \to \mathbb{E}\Big[F\Big(\frac{x - V}{U}\Big)\Big], \quad n \to \infty.$$

Comments

- Often certain model parameters are unkown and subject to statistical variation. One approach to deal with this phenomenon is to model this uncertainty with a random variable and analyze its impact on the final result. In actuarial mathematics this is done e.g. in credibility theory. By considering Poisson-mixture sums we take such an approach.
- According to our best knowledge, considering mixture distributions combined with studying a conditional version of Stein's equation is new.
- The appearance of the size-bias transformation in this context is a new phenomenon, compared to the analysis of

$$\mathbb{E}[f'(W)] \quad \text{and} \quad \mathbb{E}[Wf(W)]\,.$$

Related work

- Let $(X_i)_{i\in\mathbb{N}}$ be a sequence of independent random variables. Under a Lindeberg-type condition, Toda (2012) proves that the weak limit for $\lambda \to \infty$ of the properly normalized random sum $\sum_{i=1}^{N_\lambda} X_i$, where $N_\lambda \sim \mathrm{Geom}(\lambda)$, is the Laplace distribution.
- The Laplace distribution was analyzed via Stein's method by Pike & Ren (2014) and their results were applied to supplement Toda (2012) by a Berry–Esseen type theorem. An upper bound for the bounded Lipschitz distance of the above mentioned normalized random sum and the Laplace distribution is established, but due to a distributional transformation, which is used for the coupling construction, the summands X_i, $i \in \mathbb{N}$, have to satisfy the symmetry condition $\mathbb{P}[X_i > 0] = \mathbb{P}[X_i < 0] = 1/2$.

Related work

- In Döbler (2012), Berry–Esseen type results for the Kolmogorov distance between the above mentioned geometric random sum and the Laplace distribution and the Wasserstein distance of Poisson and binomial random sums and the Laplace distribution were proved by direct calculations, not relying on Stein's method.
- Gaunt (2014) extends Stein's method to the class of variance-gamma distributions, which contains the normal and Laplace distributions as special cases. The Stein equation he derives is a second order differential equation. As an application he derives the limit theorem for an asymptotically variance-gamma distributed statistic.

- Ch. Döbler: On rates of convergence and Berry–Esseen bounds for random sums of centered random variables with finite third moments. *arXiv:1212.5401v1*, 2012.
- R. E. Gaunt: Variance-gamma approximation via Stein's method. *Electron. J. Probab.*, 19:no. 38, 33, 2014. URL http://dx.doi.org/10.1214/EJP.v19-3020.
- John Pike and Haining Ren: Stein's method and the Laplace distribution. *ALEA Lat. Am. J. Probab. Math. Stat.*, 11(1): 571–587, 2014.
- Alexis Akira Toda: Weak limit of the geometric sum of independent but not identically distributed random variables. *arXiv:111.1786v2*, 2012.

Sketch of proof for the Wasserstein distance (1)

For $h \in \mathcal{H}_b$, the function

$$f(x) := \exp\left(\frac{x^2}{2}\right) \int_{-\infty}^{x} (h(y) - \mathbb{E}[h(X)]) \exp\left(-\frac{y^2}{2}\right) dy, \quad x \in \mathbb{R},$$

solves the conditional Stein equation

$$\Psi := \mathbb{E}[h(Y_\lambda)|\Lambda] - \mathbb{E}[h(X)] \overset{\text{a.s.}}{=} \mathbb{E}\left[f'(Y_\lambda) - Y_\lambda f(Y_\lambda)|\Lambda\right]. \qquad (5)$$

For $n \in \mathbb{N}$ we define

$$Y'_\lambda := \frac{1}{\sqrt{\lambda\Lambda}} \sum_{i=1}^{N_\lambda - 1} X_i \text{ and } Y_{\lambda,n} := \frac{1}{\sqrt{\lambda\Lambda}} \sum_{i=1, i\neq n}^{N_\lambda} X_i.$$

Sketch of proof for the Wasserstein distance (2)

Then

$$\mathbb{E}\big[f'(Y_\lambda)\,|\,\Lambda\big] \overset{\text{a.s.}}{=} \mathbb{E}\left[\frac{1}{\lambda\Lambda}\sum_{i=1}^{N_\lambda} f'(Y'_\lambda)\,\middle|\,\Lambda\right] \tag{6}$$

$$\overset{\text{a.s.}}{=} \mathbb{E}\left[\frac{1}{\lambda\Lambda}\sum_{i=1}^{N_\lambda}\big(f'(Y'_\lambda)-f'(Y_{\Lambda,i})\big)\,\middle|\,\Lambda\right] + \mathbb{E}\left[\frac{1}{\lambda\Lambda}\sum_{i=1}^{N_\lambda} f'(Y_{\lambda,i})\,\middle|\,\Lambda\right]$$

$$\overset{\text{a.s.}}{=} \mathbb{E}\left[\frac{1}{\lambda\Lambda}\sum_{i=1}^{N_\lambda}\big(f'(Y'_\lambda)-f'(Y_{\Lambda,i})\big)\,\middle|\,\Lambda\right] + \mathbb{E}\left[\frac{1}{\lambda\Lambda}\sum_{i=1}^{N_\lambda} X_i^2 f'(Y_{\lambda,i})\,\middle|\,\Lambda\right].$$

Sketch of proof for the Wasserstein distance (3)

Furthermore

$$\mathbb{E}[Y_\lambda f(Y_\lambda)\,|\,\Lambda] \overset{\text{a.s.}}{=} \mathbb{E}\left[\frac{1}{\lambda\Lambda}\sum_{i=1}^{N_\lambda} X_i^2 \int_0^1 f'\big(Y_{\lambda,i} + t(Y_\lambda - Y_{\lambda,i})\big)\,dt\,\middle|\,\Lambda\right]. \tag{7}$$

Subtracting (7) from (6) leads to

$$\begin{aligned}\Psi &\overset{\text{a.s.}}{=} \mathbb{E}\left[\frac{1}{\lambda\Lambda}\sum_{i=1}^{N_\lambda}\big(f'(Y'_\lambda) - f'(Y_{\Lambda,i})\big)\,\middle|\,\Lambda\right] \\ &\quad + \mathbb{E}\left[\frac{1}{\lambda\Lambda}\sum_{i=1}^{N_\lambda} X_i^2 \int_0^1 \big(f'(Y_{\lambda,i}) - f'(Y_{\lambda,i} + t(Y_\lambda - Y_{\lambda,i}))\big)\,dt\,\middle|\,\Lambda\right]\end{aligned} \tag{8}$$

To control (8) we use the properties of $\mathcal{H}_b$ ...

Sketch of proof for the Kolmogorov distance

For $n \in \mathbb{N}$ we define

$$Z'_\lambda = \frac{1}{\sqrt{\lambda}} \sum_{i=1}^{N_\lambda - 1} X_i, \qquad Z_{\lambda,n} = \frac{1}{\sqrt{\lambda}} \sum_{\substack{i=1 \\ i \neq n}}^{N_\lambda} X_i, \qquad Z'_{\lambda,n} = \frac{1}{\sqrt{\lambda}} \sum_{\substack{i=1 \\ i \neq n}}^{N_\lambda - 1} X_i,$$

$h_z := 1_{(-\infty,z]}$ and $g_z(\sigma^2) := \mathbb{E}[h_z(\sigma X)]$, and $Y_{\lambda,n} := \frac{X_n}{\sqrt{\lambda}}$ as well as

$$K_{\lambda,n}(t) := \mathbb{E}\Big[Y_{\lambda,n}(\mathbf{1}_{\{0 \le t \le Y_{\lambda,n}} - \mathbf{1}_{Y_{\lambda,n} \le t \le 0})\Big], \quad t \in \mathbb{R}.$$

$$\begin{aligned}
\Phi := \mathbb{P}[Z_\lambda \le z|\Lambda] - \mathbb{P}[\sqrt{\Lambda}X \le z|\Lambda] &\overset{\text{a.s.}}{=} \mathbb{E}\big[h_z(Z_\lambda)\,\big|\,\Lambda\big] - g_z(\Lambda) \\
&\overset{\text{a.s.}}{=} \mathbb{E}\Big[\frac{N_\lambda}{\lambda\Lambda}\big(h_z(Z'_\lambda) - g_z(\Lambda)\big)\,\Big|\,\Lambda\Big] \\
&\overset{\text{a.s.}}{=} \frac{1}{\Lambda}\mathbb{E}\Big[\sum_{n=1}^{N_\lambda}\int_{\mathbb{R}}\big(h_z(Z'_\lambda) - g_z(\Lambda)\big)K_{\lambda,n}(t)\,dt\,\Big|\,\Lambda\Big].
\end{aligned}$$

Sketch of proof for the Kolmogorov distance

Further $\Phi \overset{\text{a.s.}}{=} A + B$ for

$$A := \frac{1}{\Lambda}\,\mathbb{E}\Bigg[\sum_{n=1}^{N_\lambda}\int_{\mathbb{R}}\big(h_z(Z'_\lambda) - h_z(Z_{\lambda,n}+t)\big)K_{\lambda,n}(t)\,dt\,\Bigg|\,\Lambda\Bigg], \tag{9}$$

and

$$B := \frac{1}{\Lambda}\,\mathbb{E}\Bigg[\sum_{n=1}^{N_\lambda}\int_{\mathbb{R}}\big(h_z(Z_{\lambda,n}+t) - g_z(\Lambda)\big)K_{\lambda,n}(t)\,dt\,\Bigg|\,\Lambda\Bigg]. \tag{10}$$

The terms A and B have to be analyzed ...