Lifetime Dependence Modelling using a Generalized Multivariate Pareto Distribution

Daniel Alai  Zinoviy Landsman

Centre of Excellence in Population Ageing Research (CEPAR)

School of Mathematics, Statistics & Actuarial Science
University of Kent

Department of Statistics
University of Haifa

18 July 2017
Plan

- Introduction

- Multivariate Generalized Pareto Distribution
  - Parameter Estimation
  - Optimal Quantile Selection

- Bulk Annuity Pricing

- Conclusion
Introduction

- Motivation: Provide the means to assess the impact of dependent lifetimes on annuity valuation and risk management.
  - Basis: systematic mortality improvements induce dependence.
    - Could reframe as cohort, or pool of similar-risks, analysis.

- Investigate a multivariate generalized Pareto distribution because:
  - Interesting family with potential for more flexible dependence.
  - More suitable for older-age dependence due to presence of extremes.

- Resolve estimation in the presence of truncation (in a variety of ways).
  - Moment-based estimation (applied to the minimum observation).
  - Quantile-based estimation (with optimal levels).

- Assess the impact of dependence on the risk of a bulk annuity.
  - Dependence increases the risk.
Assume $m$ pools of $n$ lives. Suppose the lives within a pool are dependent.

Let $X_{i,j}$ be the lifetime of individual $i$ in pool $j$.

We apply the following model for lifetimes:

$$X_j \sim h(\theta, \lambda_S), \quad \forall j,$$

where $\lambda_S = \sum_{i=1}^{n} \lambda_i$.

- This means pools are independent.
  - Each pool is one draw from the multivariate distribution.
- The magnitudes of $m$ and $n$ determine the application.
  - $n = 2 \Rightarrow$ joint-life products.

\[ \downarrow \text{Small } m \text{ or } n \text{ might pose difficulties!} \]
Multiply Monotone Generated Distributions

Let $X = (X_1, \ldots, X_n)$ be a multivariate random vector with strictly positive components $X_i > 0$ such that its joint survival function is given by

$$P(X_1 > x_1, \ldots, X_n > x_n) = h\left(\sum_{i=1}^{n} \lambda_i x_i\right), \quad x_i \geq 0,$$

for $\lambda_i > 0, \forall i$, where $h$ is $d$-times monotone, $d \geq n$. That is, for $k \in \{1, \ldots, d\}$,

$$(-1)^k h^{(k)}(x) \geq 0, \quad x > 0.$$

Two well-known examples include the Pareto and Weibull distributions.

{Pareto} \quad h(x) = (1 + x)^{-\frac{1}{\theta}}, \quad x \geq 0, \quad \theta \in \mathbb{R}^+,

{Weibull} \quad h(x) = \exp(-x^{\frac{1}{\theta}}), \quad x \geq 0, \quad \theta \in [1, \infty).

The Pareto generator resembles the Clayton copula generator $(1 + \theta x)^{-1/\theta}$.

The Weibull generator is just the Gumbel copula generator.
The multiply monotone condition on $h$ ensures we have admissible densities for all possible subsets of $X$!

For example, the densities of $X$ and $X_i$ are given by,

\[ f_X(x_1, \ldots, x_n) = (-1)^n \lambda_1 \cdots \lambda_n h^{(n)} \left( \sum_{i=1}^{n} \lambda_i x_i \right) \geq 0, \quad x_i > 0, \]

\[ f_i(x_i) = (-1)\lambda_i h^{(1)}(\lambda_i x_i) \geq 0, \quad x_i > 0. \]

Survival functions are always given by $h$:

\[ P(X_i > x_i, X_j > x_j) = h(\lambda_i x_i + \lambda_j x_j), \quad x_i, x_j \geq 0, i \neq j, \]

\[ P(X_i > x_i) = h(\lambda_i x_i), \quad x_i \geq 0. \]

As such, we require that $h(0) = 1$ and $\lim_{x \to \infty} h(x) = 0$.

There is a clear link to Archimedean survival copulas.
Examples of $h$

Pareto, $\theta \in \left\{ \frac{1}{4}, \frac{1}{2}, 1, 2 \right\}$.

Clayton, $\theta \in \left\{ \frac{1}{4}, \frac{1}{2}, 1, 2 \right\}$.

Gumbel, $\theta \in \left\{ 1, 2, 4, 8 \right\}$.

Frank, $\theta \in \left\{ -4, -1, 1, 4 \right\}$.

AMH, $\theta \in \left\{ -\frac{19}{20}, -\frac{1}{4}, \frac{1}{4}, \frac{3}{4} \right\}$.

Expo-Pareto, $\theta \in \left\{ 1, 2, 4, 8 \right\}$.
Bivariate Marginal Correlations

As well as exhibiting either light or heavy tails, each $h$ produces a different correlation structure between marginals.

\[ \text{Not surprisingly, heavy tailed examples permit only positive correlation, whereas light tailed distributions allow for negative correlation.} \]

For the Pareto and Clayton, \( \text{Corr}(X_i, X_j) = \theta, \) for \( i \neq j \).

For the remaining examples, the bivariate correlation involves either the incomplete gamma, dilogarithm or trilogarithm function.

\[
\Gamma(s, x) = \int_x^\infty t^{s-1} e^{-t} dt,
\]

\[
\text{Li}_2(z) = \int_z^0 \frac{\ln(1 - t)}{t} dt,
\]

\[
\text{Li}_3(z) = -\int_z^0 \frac{\text{Li}_2(t)}{t} dt.
\]

\[ \text{More on correlation later, after we’ve addressed truncation!} \]
Parameter Estimation

We wish to make use of pool statistics to estimate model parameters.

- Mean and Variance;
- Minimum and Maximum;
- Quantiles!

⇒ Within-pool dependence is a clear obstacle, but not the only one!

⇒ We anticipate truncated observations.

We require some theoretical results before we can proceed.
Mixed Truncated Moments

**Theorem (Mixed Moments)**

Consider \( \mathbf{X} = (X_1, \ldots, X_n) \) with distribution generated by \( d \)-times monotone \( h \), \( d \geq n \). Let \( \tau X_i = \{X_i|X > \tau\} \). If finite,

\[
\mathbb{E}\left[ \prod_{i=1}^{n} \tau X_i^{k_i} \right] = h(\lambda S \tau)^{-1} \sum_{j_1=0}^{k_1} \cdots \sum_{j_n=0}^{k_n} h(-\sum_{i=1}^{n} j_i) (\lambda S \tau) \prod_{i=1}^{n} \frac{(-1)^j \tau^{k_i-j_i} k_i!}{(k_i-j_i)! \lambda_i^{k_i}},
\]

where \( \lambda_s = \sum_{i=1}^{n} \lambda_i \), \( k = \sum_{i=1}^{n} k_i \), \( k \in \{1, 2, \ldots, d\} \), and \( k_i \in \{0\} \cup \mathbb{Z}^+ \); furthermore, where \( h^{(-k)}(x) = - \int_{x}^{\infty} h^{(-(k-1))}(y)dy \) and \( h^{(0)}(x) = h(x) \).

- Mean, variance and covariance results are especially relevant.
- This result can be used to find the moments of the minimum (and maximum).

⇒ Let’s take a look at the bivariate correlation plots.
  - They depend on \( \tau \)!
Correlation Plots for $\tau \in \{0, 1, 2, 5\}$

Pareto

Clayton

Gumbel

Frank

Ali-Mikhail-Haq

Exponential-Pareto
Comments on Mean-Variance Matching

Mean, variance and covariance results enable us to determine the expectation of the sample (pool) mean and variance.

- Averaging these, respectively, across pools yields \( \hat{\theta} \) and \( \hat{\lambda}_S \).

Consider the Pareto distribution with \( \lambda_i = \lambda, \forall i \); we have

\[
\mathbb{E}[a_1(\tau X_j)] = \frac{\lambda^{-1} + \tau(n + \theta^{-1} - 1)}{\theta^{-1} - 1},
\]

\[
\mathbb{E}[\tilde{m}_2(\tau X_j)] = \frac{(\lambda^{-1} + \tau n)^2}{(\theta^{-1} - 1)(\theta^{-1} - 2)},
\]

where \( a_1 \) and \( \tilde{m}_2 \) denote the unbiased sample (pool) mean and variance.

Note the relationship with pool size \( n \).

- Inseparable from the truncation point \( \tau \).
- No indication that large \( n \) will produce more accurate estimation.
- Perhaps ideal for a portfolio of many joint-life annuities.
Comments on Minimum-Maximum Matching

Sample moments of minima (or maxima) yield estimates $\hat{\theta}$ and $\hat{\lambda}$.

Focus on minimum, since it looks much more promising.

Consider the Pareto distribution with $\lambda_i = \lambda$, $\forall i$; we have

$$\mathbb{E}[a_1(\tau X_{(1)})] = \frac{\lambda^{-1}/n + \tau \theta^{-1}}{\theta^{-1} - 1},$$

$$\mathbb{E}[\tilde{m}_2(\tau X_{(1)})] = \frac{\theta^{-1}(\lambda^{-1}/n + \tau)^2}{(\theta^{-1} - 1)^2(\theta^{-1} - 2)}.$$ 

Contrast the relationship with pool size $n$ to the mean-variance matching.

This time distinct from $\tau$ and indicative of more accuracy as $n \uparrow$.

Perhaps ideal for a portfolios of employer-based pension schemes.
Quantile Matching

The previous two estimation procedures require sufficiently light tails!
- For the Pareto, \(0 < \theta < \frac{1}{2}\).
- Quantile-based estimation procedures do not impose this restriction!

We apply quantile matching to the sample of pool minima!

\[
q_{\tau X(1)}(p) = \frac{h^{-1}((1 - p)h(\lambda_S \tau))}{\lambda_S}.
\]

- Our estimation procedure requires three \{optimal\} levels \(p_1, p_2, \text{and } p_3\).
Consider a sample of iid $X_1, \ldots, X_n$ with density $f(x, \vartheta)$, $\vartheta \in \Theta \subset \mathbb{R}$, differentiable with respect to $\vartheta$ for almost all $x \in \mathbb{R}$.

The Fisher information about $\vartheta$ contained in statistic $T_n(X_1, \ldots, X_n)$ is

$$I_{T_n}(\vartheta) = \int_{\mathbb{R}} \left( \frac{\partial \ln f_{T_n}(x, \vartheta)}{\partial \vartheta} \right)^2 f_{T_n}(x, \vartheta) dx.$$ 

A higher Fisher information is indicative of more precise estimation.

The Fisher information contained in the sample quantiles, $I_{\hat{q}(p_1), \ldots, \hat{q}(p_k)}(\vartheta)$, $0 = p_0 < p_1 < \ldots < p_{k+1} = 1$, is asymptotically equal to $nI_k(p_1, \ldots, p_k)$;

$$I_k(p_1, \ldots, p_k) = \sum_{i=0}^{k} \frac{(\beta_{i+1} - \beta_i)^2}{p_{i+1} - p_i},$$

where $\beta_i = f(q(p_i), \vartheta) \partial q(p_i) / \partial \vartheta$, $\forall i$ and $\beta_0 = \beta_{k+1} = 0$.

⇒ Find optimal levels $p_1^*, \ldots, p_k^*$, such that $I_k$ is maximized!
The Pareto Distribution

The optimal quantile selection procedure depends heavily on \( h \).

Let us focus on the Pareto distribution.

We want to estimate \( \theta \) (with \( \lambda_S \) unknown) using two quantiles (\( p_1 < p_2 \)).

\[
I_2(p_1, p_2) = \frac{\beta_1^2}{p_1} + \frac{(\beta_2 - \beta_1)^2}{p_2 - p_1} + \frac{\beta_2^2}{1 - p_2}.
\]

For the Pareto distribution, and letting \( \tilde{p}_i = 1 - p_i \), we obtain

\[
\beta_i = \theta \cdot \tilde{p}_i \cdot \ln \tilde{p}_i.
\]

The objective function may be rewritten as follows

\[
I_2(p_1, p_2) = \theta^2 \left( \frac{\tilde{p}_1 \ln^2 \tilde{p}_1}{p_1} + \frac{\tilde{p}_2 \ln \tilde{p}_2 - \tilde{p}_1 \ln \tilde{p}_1}{p_2 - p_1} + \tilde{p}_2 \ln^2 \tilde{p}_2 \right).
\]

Maximizing this does not require knowledge of \( \theta \) and \( \lambda_S \)!

Furthermore, it does not even depend on \( \tau \)!
Finding $p_1^*$ and $p_2^*$ for the Pareto Distribution

$p_2 = 0.25$.  
$p_2 = 0.50$.  
$p_2 = 0.75$.  

$p_2 = 0.90$.  
$p_2 = 0.95$.  
$p_2 = 0.99$.  

The optimal levels are $p_1^* = 0.6385$ and $p_2^* = 0.9265$.  

D. H. Alai (CEPAR, Kent)
Finding $p_3^*$

Armed with $\hat{\theta}$, we consider the optimal quantile level $p_3$ used to estimate $\lambda_5$.

Following the same method, optimal $p_3$ is found by maximizing

$$\tilde{p}_3 \left(1 - \tilde{p}^{\theta}_3\right)^2,$$

This depends on $\theta$, for which we luckily have an estimate!

(a) $\theta = \frac{1}{4}$.

(b) $\theta = \frac{1}{2}$.

(c) $\theta = 1$.

(d) $\theta = 2$.

The lighter the tail, the higher the optimal quantile level.
Optimal Quantiles in General

The Pareto distribution is quite unique!

- The truncation point does not affect the optimal quantile levels.
- \( \theta \) can be estimated optimally without knowledge of \( \lambda_S \).

In general, the truncation point complicates matters significantly.

- But even \( \tau = 0 \) does not imply optimal quantile-levels can always be found.

We can find optimal quantile levels \( p_1^* \) and \( p_2^* \) if we can write

\[
\beta^{(\theta)} = f(\theta, \lambda_S) \times g(p)
\]

for some functions \( f \) and \( g \).

- Achievable for the Pareto, Weibull and exponential-Pareto distributions.

\[
\beta^{(\theta)} \propto \tilde{p} \cdot \ln \tilde{p}, \quad \text{for the Pareto and exponential-Pareto},
\]

\[
\beta^{(\theta)} \propto \tilde{p} \cdot \ln \tilde{p} \cdot \ln(\ln \tilde{p}), \quad \text{for the Weibull}.
\]
Consider a pool of lives \( \tau X = (\tau X_1, \ldots, \tau X_n) \). A bulk annuity pays £1 to each survivor of the pool at the end of each year.

Let \( \tau A \) denote its value at inception \( (t = \tau) \) and let \( \tau S_t \) denote the number of survivors in the pool at time \( t \geq \tau \).

In order to find the mean and variance of \( \tau A \), we need to find the distribution of \( \tau S_t \) and the joint distribution of \( (\tau S_t, \tau S_s) \), \( s > t \).

If the lives are independent, these can readily be found.

What if the lives are dependent?
### The Impact of Dependence ($\delta = 0.02, \mu = 60, \tau = 5$)

<table>
<thead>
<tr>
<th>$n$</th>
<th>Marginal Moments</th>
<th>Independent Pareto</th>
<th>Multivariate Pareto</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$E[\tau X_1]$</td>
<td>$\text{Var}(\tau X_1)^{\frac{1}{2}}$</td>
<td>$E[\tau A]$</td>
</tr>
<tr>
<td>2</td>
<td>75.00</td>
<td>17.32</td>
<td>14.38</td>
</tr>
<tr>
<td>20</td>
<td>75.00</td>
<td>10.95</td>
<td>154.70</td>
</tr>
</tbody>
</table>

Truncation affects the marginal distributions!

Given $n$, we apply appropriate parameters for a fair comparison.
Conclusion

- **Motivation:** Provide the **means** to assess the impact of dependent lifetimes on annuity valuation and risk management.
  - **Basis:** systematic mortality improvements induce dependence.
  - Could reframe as cohort, or pool of similar-risks, analysis.

- Investigate a **multivariate generalized Pareto distribution** because:
  - Interesting family with **potential** for more **flexible** dependence.
  - More suitable for older-age dependence due to presence of **extremes**.

- Resolve estimation in the presence of truncation (in a variety of ways).
  - Moment-based estimation (applied to the **minimum** observation).
  - Quantile-based estimation (with **optimal** levels).

- Assess the impact of dependence on the risk of a **bulk annuity**.
  - Dependence increases the risk.
Thank you!