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.

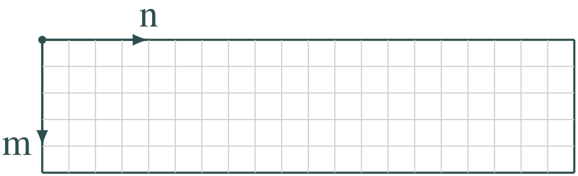
Modelling Dependent Lifetimes

Assume m pools of n lives. $\rightsquigarrow$ Suppose the lives within a pool are dependent.
$\rightarrow$ Let $X_{i,j}$ be the lifetime of individual i in pool j.
We apply the following model for lifetimes:

$$\mathbf{X}_j \sim h(\theta, \lambda_S), \quad \forall j,$$

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.



↳ Small m or n might pose difficulties!

Multiply Monotone Generated Distributions

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

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

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

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

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

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

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

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

↳ The Weibull generator is just the Gumbel copula generator.

Joint Densities of Subsets of $\mathbf{X}$

The multiply monotone condition on h ensures we have admissible densities for all possible subsets of $\mathbf{X}$!

For example, the densities of $\mathbf{X}$ and X_i are given by,

$$f_{\mathbf{X}}(x_1, \ldots, x_n) = (-1)^n \lambda_1 \cdots \lambda_n h^{(n)}\left(\sum_{i=1}^{n} \lambda_i x_i\right) \geq 0, \qquad x_i > 0,$$

$$f_i(x_i) = (-1)\lambda_i h^{(1)}(\lambda_i x_i) \geq 0, \qquad x_i > 0.$$

Survival functions are always given by h:

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

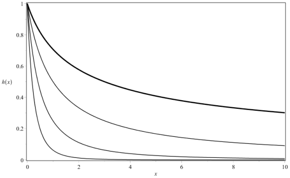
$$P(X_i > x_i) = h(\lambda_i x_i), \qquad x_i \geq 0.$$

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

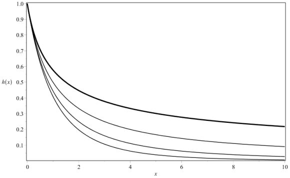
↳ There is a clear link to Archimedean survival copulas.

cepar ARC CENTRE OF EXCELLENCE IN POPULATION AGEING RESEARCH University of Kent

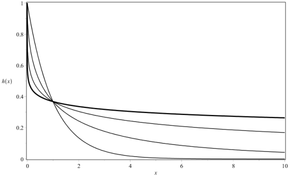
Examples of h



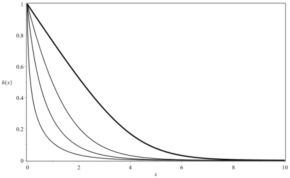
Pareto, $\theta \in \left\{\frac{1}{4}, \frac{1}{2}, 1, \mathbf{2}\right\}$.



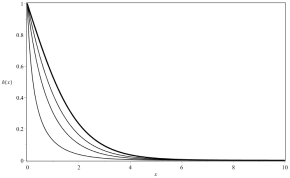
Clayton, $\theta \in \left\{\frac{1}{4}, \frac{1}{2}, 1, \mathbf{2}\right\}$.



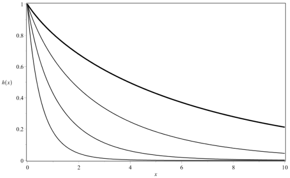
Gumbel, $\theta \in \{1, 2, 4, \mathbf{8}\}$.



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



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



Expo-Pareto, $\theta \in \{1, 2, 4, \mathbf{8}\}$.

Bivariate Marginal Correlations

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

↳ Not surprisingly, heavy tailed examples permit only positive correlation, whereas light tailed distributions allow for negative correlation.

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

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

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

$$\mathrm{Li}_2(z) = \int_z^0 \frac{\ln(1-t)}{t} dt,$$

$$\mathrm{Li}_3(z) = -\int_z^0 \frac{\mathrm{Li}_2(t)}{t} dt.$$

↳ 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!

$\Rightarrow$ 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* ${}_{\boldsymbol{\tau}}X_i = \{X_i | \mathbf{X} > \boldsymbol{\tau}\}$. *If finite,*

$$\mathbb{E}\left[\prod_{i=1}^{n} {}_{\boldsymbol{\tau}}X_i^{k_i}\right] = h(\lambda_S \tau)^{-1} \sum_{j_1=0}^{k_1} \cdots \sum_{j_n=0}^{k_n} h^{(-\sum_{l=1}^{n} j_l)}(\lambda_S \tau) \prod_{i=1}^{n} \frac{(-1)^{j_i} \tau^{k_i - j_i}}{(k_i - j_i)!} \frac{k_i!}{\lambda_i^{j_i}},$$

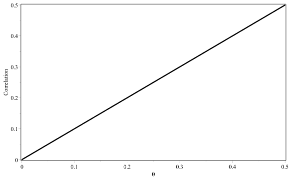
where $\lambda_S = \sum_{i=1}^{n} \lambda_i$, $k = \sum_{i=1}^{n} k_i, k \in \{1, 2, \ldots, d\}$, *and* $k_i \in \{0\} \cup \mathbb{Z}^+$; *furthermore, where* $h^{(-k)}(x) = -\int_x^{\infty} h^{(-(k-1))}(y) dy$ *and* $h^{(0)}(x) = h(x)$.

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

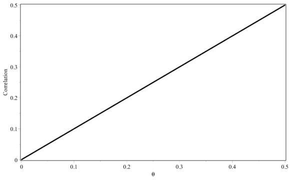
⇒ Let's take a look at the bivariate correlation plots.

- ↳ They depend on τ!

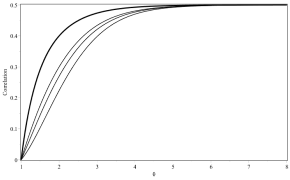
Correlation Plots for $\tau \in \{\mathbf{0}, 1, 2, 5\}$



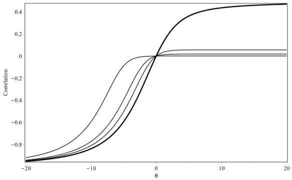
Pareto



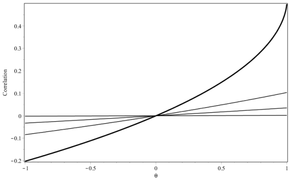
Clayton



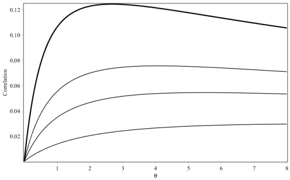
Gumbel



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 $\widehat{\theta}$ and $\widehat{\lambda}_S$.

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

$$\mathbb{E}[a_1({}_{\boldsymbol{\tau}}\mathbf{X}_j)] = \frac{\lambda^{-1} + \tau(n + \theta^{-1} - 1)}{\theta^{-1} - 1},$$

$$\mathbb{E}[\widetilde{m}_2({}_{\boldsymbol{\tau}}\mathbf{X}_j)] = \frac{(\lambda^{-1} + \tau n)^2}{(\theta^{-1} - 1)(\theta^{-1} - 2)},$$

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

Note the relationship with pool size n.

↳ Inseparable from the truncation point τ.

↳ 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 $\widehat{\theta}$ and $\widehat{\lambda}$.

↳ Focus on minimum, since it looks much more promising.

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

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

$$\mathbb{E}[\widetilde{m}_2({}_{\boldsymbol{\tau}}\mathbf{X}_{(1)})] = \frac{\theta^{-1}(\lambda^{-1}/n + \tau)^2}{(\theta^{-1} - 1)^2(\theta^{-1} - 2)}.$$

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

$\Rightarrow$ This time distinct from τ and indicative of more accuracy as $n \nearrow$.

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

Quantile Matching

The previous two estimation procedures require sufficiently light tails!

↳ For the Pareto, $0 < \theta < 1/2$.

↳ 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, and p_3.

Fisher Information: Establishing the Objective Function

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 ϑ for almost all $x \in \mathbb{R}$.
The Fisher information about ϑ 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, $I_{\widehat{q}(p_1), \ldots, \widehat{q}(p_k)}(\vartheta)$, $0 = p_0 < p_1 < \ldots < p_{k+1} = 1$, 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 $\beta_0 = \beta_{k+1} = 0$.
$\Rightarrow$ Find optimal levels $p_1^\star, \ldots, p_k^\star$, 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 θ (with λ_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 $\check{p}_i = 1 - p_i$, we obtain

$$\beta_i = \theta \cdot \check{p}_i \cdot \ln \check{p}_i.$$

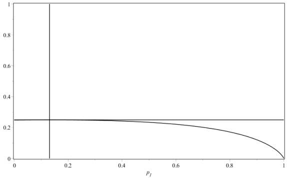
The objective function may be rewritten as follows

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

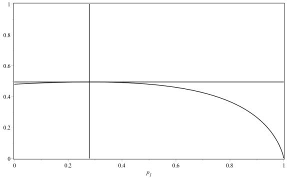
↳ Maximizing this does not require knowledge of θ and λ_S!

↳ Furthermore, it does not even depend on τ!

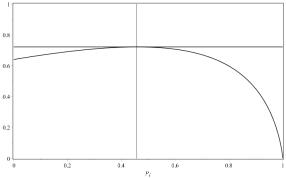
Finding $p_1^\star$ and $p_2^\star$ for the Pareto Distribution



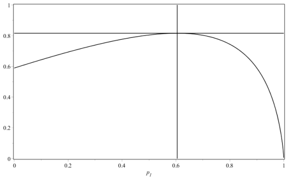
$p_2 = 0.25.$



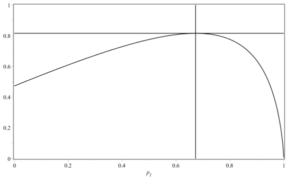
$p_2 = 0.50.$



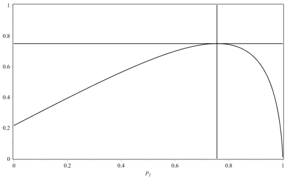
$p_2 = 0.75.$



$p_2 = 0.90.$



$p_2 = 0.95.$



$p_2 = 0.99.$

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

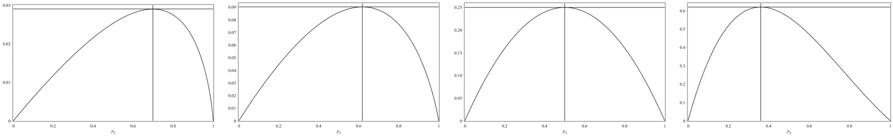
Finding $p_3^\star$

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

Following the same method, optimal p_3 is found by maximizing

$$\frac{\check{p}_3\left(1-\check{p}_3^{\theta}\right)^2}{p_3}.$$

↳ This depends on θ, 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.

↳ θ can be estimated optimally without knowledge of λ_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 $p_1^\star$ and $p_2^\star$ 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.

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

$$\beta^{(\theta)} \propto \check{p} \cdot \ln \check{p} \cdot \ln(-\ln \check{p}), \quad \text{for the Weibull.}$$

The Bulk Annuity

Consider a pool of lives ${}_{\tau}\mathbf{X} = ({}_{\tau}X_1, \ldots, {}_{\tau}X_n)$. 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 ($\delta = 0.02$, $\mu = 60$, $\tau = 5$)

	Marginal Moments		Independent Pareto		Multivariate Pareto	
n	$\mathbb{E}[_{\boldsymbol{\tau}}X_1]$	$\mathrm{Var}(_{\boldsymbol{\tau}}X_1)^{\frac{1}{2}}$	$\mathbb{E}[_{\tau}A]$	$\mathrm{Var}(_{\tau}A)^{\frac{1}{2}}$	$\mathbb{E}[_{\tau}A]$	$\mathrm{Var}(_{\tau}A)^{\frac{1}{2}}$
2	75.00	17.32	14.38	11.50	14.38	13.11
20	75.00	10.95	154.70	32.79	154.70	52.07

↳ 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!