Modeling Joint Mortality
and its Impact on Annuity Contracts

AP - TU Delft

Pasquale Cirillo

Longevity 12

Multivariate mortality

- As discussed in Frees et al. (1996), most of the models of multivariate mortality assume the independence of individuals, so that the joint survival function is nothing more than the product of the marginals.
- Independence is an extremely strong (and unrealistic) assumption.
- In the literature, there are some interesting models, which do not assume independence:
    1. copulas.
    2. joint shock models.
    3. bivariate Gompertz.
    4. doubly stochastic mortality.

My Idea

Using some results I recently proposed in urn-based shock models, I introduce a constructive approach to dependent mortality.
The building blocks are:

- Generalized Polya Sequences.
- an intuitive dependence structure (based on common shocks).

The model shows good performances, learns from data and it is easy to simulate.

Generalized Polya sequence

Consider a sequence of random variables $\{T_n\}_{n\geq 1}$ with values in $\mathbb{N}_0^+$. Let $\{\alpha_j, \beta_j, j \in \mathbb{N}_0^+\}$ be such that

1. $\alpha_j, \beta_j \geq 0$, for all j,
2. $\alpha_j + \beta_j > 0$, for all j,
3. $\lim_{n\to\infty} \prod_{j=0}^{n} \beta_j/(\alpha_j + \beta_j) = 0$.

We say that $\{T_n\}_{n\geq 1}$ is a GPS if

$$P[T_1 = t] = \frac{\alpha_t}{\alpha_t + \beta_t} \prod_{j=0}^{t-1} \frac{\beta_j}{\alpha_j + \beta_j}, \quad (1)$$

$$P[T_{n+1} = t | \mathbf{T}_n = \mathbf{t}_n] = \frac{\alpha_t + m_t(\mathbf{t}_n)}{\alpha_t + \beta_t + s_t(\mathbf{t}_n)} \prod_{j=0}^{t-1} \frac{\beta_j + r_j(\mathbf{t}_n)}{\alpha_j + \beta_j + s_j(\mathbf{t}_n)},$$

where $\mathbf{t}_n = (t_1, ..., t_n)$, $m_j(\mathbf{t}_n) = \sum_{k=1}^{n} \mathbf{1}_{[t_k=j]}$, $r_j(\mathbf{t}_n) = \sum_{k=1}^{n} \mathbf{1}_{[t_k>j]}$ and $s_j(\mathbf{t}_n) = m_j(\mathbf{t}_n) + r_j(\mathbf{t}_n)$.

Generalized Polya sequence

The name *Generalized Polya sequence* comes from the fact that $\{T_n\}$ can be generated in the following way:

- Consider $j = 0, 1, 2, 3, \ldots$ Polya urns U_j. Every urn contains α_j white balls and β_j black balls.
- Sample urn U_0. If the sampled ball is white, set $T_1 = 0$ and reinforce the urn with 1 (or $\gamma \in \mathbb{N}_0^+$ in general) extra white ball. If the ball is black, add another black ball and move to urn U_1.
- Sample urn U_1, if white $T_1 = 1$, otherwise move to U_2. And so on until a white ball is extracted in t, so that $T_1 = t$.
- To generate T_2 start again from U_0 and notice that, up to step t, all the urns U_j, $j = 0, \ldots, t$, have been Polya reinforced.
- Etc.

GPS generalize the Bayesian paradigm embedded in simple Polya urns.

Generalized Polya sequence

Walker and Muliere (1997) have shown that

- the sequence $\{T_n\}$ is exchangeable;
- the de Finetti measure of $\{T_n\}$, i.e. the random distribution function F, such that, given F, the random variables T_n are i.i.d. with distribution F, is a beta-Stacy process.

This is important, since

- the beta-Stacy process is a special case of neutral to the right process and it is conjugate to right-censored observations;
- the beta-Stacy process is widely used in Bayesian nonparametrics.

Censoring

Let $T_1, ..., T_n$ be independent and identically distributed random variables, subject to right-censoring. What we observe is $(T_1^*, \delta_1), ..., (T_n^*, \delta_n)$, with

$$T_i^* = t, \delta_i = 0 \text{ if a censoring took place, i.e. } T_i > t,$$
$$T_i^* = t, \delta_i = 1 \text{ if no censoring took place, i.e. } T_i = t.$$

With a quadratic loss function, the predictive distribution of T_{n+1} given $(\mathbf{T}_n, \delta_n)$ is the Bayes estimator for the random distribution function.

Censoring

Under a beta-Stacy prior, we have the following result for the survival function (see Bulla, 2005):

$$\hat{S}(t) = P[T_{n+1} > t | \mathbf{T}_n^* = \mathbf{t}_n, \delta_n = \mathbf{d}_n] = \prod_{j=0}^{t} \left[1 - \frac{\alpha_j + m_j^*(\mathbf{t}_n, \mathbf{d}_n)}{\alpha_j + \beta_j + s_j(\mathbf{t}_n)} \right] \quad (2)$$

where $m_j^*(\mathbf{t}_n, \mathbf{d}_n) = \sum_{k=1}^{n} \mathbf{1}_{[t_k = j, d_k = 1]}$ and $\mathbf{d}_n \in \{0, 1\}^n$.
Notice that:

- in case of no censoring, eq. 2 is given by $\hat{S}(t) = \prod_{j=0}^{t} (\beta_j + r_j(\mathbf{t}_n) / (\alpha_j + \beta_j + s_j(\mathbf{t}_n))$.
- for $\alpha_j, \beta_j \to 0$ for all j, eq. 2 reduces to the standard Kaplan-Meier estimator.

TUDelft

Constructing a bivariate reinforced process

Following Cirillo (2008, 2011, 2015) we want to build a bivariate random process $\{(X_n, Y_n), n \geq 1\}$ that provides a model for coupled lifetimes. At the same time, we want this process to be reinforced in a way similar to that of GPS, so that we can try to perform some Bayesian nonparametric analysis.

Here the ingredients:

- let $\{V_n\}_{n\geq 1}$, $\{W_n\}_{n\geq 1}$ and $\{Z_n\}_{n\geq 1}$ be three independent sequences from GPS with respectively parameters (α_j^V, β_j^V), (α_j^W, β_j^W) and (α_j^Z, β_j^Z);
- define the random process $\{(X_n, Y_n), n \geq 1\}$ with

$$X_n = Z_n + V_n, \tag{3}$$

$$Y_n = Z_n + W_n. \tag{4}$$

Constructing a bivariate reinforced process

By construction:

- for every couple (X_j, Y_j), each individual has a common element Z_j and a specific one, V_j or W_j;
- in this way, we build a dependence without creating a parametric model;
- conditionally on Z_j, X_j and Y_j are independent;
- $\sigma(\mathbf{Z}_n, \mathbf{V}_n, \mathbf{W}_n) = \sigma(\mathbf{Z}_n, \mathbf{X}_n, \mathbf{Y}_n)$;
- the dependence structure is given by

$$Cov(X_1, Y_1) = Var(Z_1) \geq 0, \quad (5)$$

$$Cov(X_{n+1}, Y_{n+1} | \mathbf{Z}_n, \mathbf{V}_n, \mathbf{W}_n) = Var(Z_{n+1} | \mathbf{Z}_n), \; n \geq 1. \quad (6)$$

TU Delft

Main Properties

It can be shown that:

- The couples $\{(X_n, Y_n), n \geq 1\}$ are exchangeable. (Walker and Muliere, 1997)
- Let F_X be the marginal distribution of $\{X_n\}$, we have

$$F_X = F_Z * F_V, \quad (7)$$

$$F_Y = F_Z * F_W. \quad (8)$$

  Hence the marginal distributions of $\{X_n\}$ and $\{Y_n\}$ are convolutions of beta-Stacy processes.
- If P is the probability function associated with F,

$$P_{XY}(x, y) = \sum_{z=0}^{x \wedge y} P_Z(z) P_V(x-z) P_W(y-z), \ \forall (x, y) \in \mathbb{N}_0^2. \quad (9)$$

TU Delft

Main Properties

- If $\sigma_Z^2 = Var_{F_Z}(Z)$, then

$$Cov_{F_{XY}}(X, Y) = \sigma_Z^2. \quad (10)$$

- From a Bayesian point of view, the use of a bivariate reinforced process to study coupled lifetimes is equivalent to the definition of a probability measure $\mathcal{P}_2$ on the space of the bivariate functions on $\mathbb{N}_0^2$.

TU Delft

Estimating the bivariate survival function

We consider, on $\mathbb{N}_0^2$, the bivariate survival function $S(x, y) = P[X > x, Y > y]$.
Let $(\mathbf{X}_n, \mathbf{Y}_n)$ be an independent and identically distributed sample from S. Our interest is related to the predictive

$$\hat{S}(x, y) = P[X_{n+1} > x, Y_{n+1} > y | \mathbf{X}_n = \mathbf{x}_n, \mathbf{Y}_n = \mathbf{y}_n]. \quad (11)$$

- If the elements in $(\mathbf{X}_n, \mathbf{Y}_n)$ are not subject to right-censoring, it is possible to compute eq. 11 as (see Cirillo, 2008)

$$\hat{S}(x, y) = \frac{P[X_{n+1} > x, Y_{n+1} > y, \mathbf{X}_n = \mathbf{x}_n, \mathbf{Y}_n = \mathbf{y}_n]}{P[\mathbf{X}_n = \mathbf{x}_n, \mathbf{Y}_n = \mathbf{y}_n]}. \quad (12)$$

TUDelft

Estimating the bivariate survival function

If some of the elements in $(\mathbf{X}_n, \mathbf{Y}_n)$ are right-censored, it is not possible to obtain a closed-form expression for

$$\hat{S}(x,y) = P[X_{n+1} > x, Y_{n+1} > y | \mathbf{X}_n^* = \mathbf{x}_n, \delta_n = \mathbf{d}_n, \mathbf{Y}_n^* = \mathbf{y}_n, \epsilon_n = \mathbf{e}_n]. \quad (13)$$

Anyway it is always possible to use a MCMC estimation.

MCMC algorithm

Set the parameters of the model, then compute $\hat{S}(x, y)$ using the following algorithm:

- Given $(\mathbf{X}_n^*, \delta_n, \mathbf{Y}_n^*, \epsilon_n)$, the full conditional of Z_n is

$$\begin{aligned} & P_{Z_n|Z_{n-1},\mathbf{X}_n^*,\delta_n,\mathbf{Y}_n^*,\epsilon_n} \\ \propto\ & P[V_n^* = x_n - z_n, \delta_n = d_n | \mathbf{V}_{n-1}^* = \mathbf{x}_{n-1} - \mathbf{z}_{n-1}, \delta_{n-1} = \mathbf{d}_{n-1}] \\ \times\ & P[W_n^* = y_n - z_n, \epsilon_n = e_n | \mathbf{W}_{n-1}^* = \mathbf{y}_{n-1} - \mathbf{z}_{n-1}, \epsilon_{n-1} = \mathbf{e}_{n-1}] \\ \times\ & P[Z_n = z_n | \mathbf{Z}_{n-1} = \mathbf{z}_{n-1}] \end{aligned}$$

  where, for example,

$$\begin{aligned} & P[W_n^* = w, \epsilon_n = e | \mathbf{W}_{n-1}^* = \mathbf{w}_{n-1}, \epsilon_{n-1} = \mathbf{e}_{n-1}] \\ = & \begin{cases} P[W_n^* \geq w | \mathbf{W}_{n-1}^* = \mathbf{w}_{n-1}, \epsilon_{n-1} = \mathbf{e}_{n-1}] & \text{if } e = 0, \\ P[W_n^* = w | \mathbf{W}_{n-1}^* = \mathbf{w}_{n-1}, \epsilon_{n-1} = \mathbf{e}_{n-1}] & \text{if } e = 1. \end{cases} \end{aligned}$$

- Since $\{Z_n\}$ is exchangeable, all the conditionals $P_{Z_n|\mathbf{Z}_{-j},\mathbf{X}_n^*,\delta_n,\mathbf{Y}_n^*,\epsilon_n}$, where $\mathbf{Z}_{-j} = \{Z_i\}_{i=1}^n \setminus \{Z_j\}$, have the same form.
- Compute $\mathbf{V}_n^* = \mathbf{X}_n^* - \mathbf{Z}_n$ and $\mathbf{W}_n^* = \mathbf{Y}_n^* - \mathbf{Z}_n$.
- Z_{n+1}, V_{n+1} and W_{n+1} are sampled w.r.t $P_{Z_{n+1}|\mathbf{Z}_n}$, $P_{V_{n+1}|\mathbf{V}_n^*,\delta_n}$ and $P_{W_{n+1}|\mathbf{W}_n^*,\epsilon_n}$.

Specifying the prior knowledge

- In Bayesian analysis, it is fundamental to elicit some prior distribution that reflects the prior knowledge we have about the phenomenon under study.
- In the univariate case, if F is a beta-Stacy process (which is conjugate), it is possible to center it on a given discrete distribution G, so that $E[F(\{j\})] = G(\{j\})$, by considering $c_j > 0 \ \forall j$, and setting

$$\alpha_j = c_j(G(\{j\}) \qquad \beta_j = c_j\left(1 - \sum_{i=0}^{j}(G(\{i\})\right). \tag{14}$$

Hence G represents the initial guess, while c_j is the so-called strength of belief.

Elicitation

- Given the construction we have used, it is interesting to notice that, in our bivariate model, it is not necessary to have a deep knowledge of the bivariate survival distribution of lifetimes, but it is sufficient to have some a priori guesses about $Cov(X, Y)$ and the marginals of X and Y.

Elicitation

- Using eqs. 7 and 10, we can formulate the following procedure:
  1. Express an initial distribution F_Z^0 for Z. If a prior knowledge is available, use it! Otherwise, the choice is free, the only constraint being $\sigma_Z^2 = Cov(X, Y)$.
  2. Determine α_j^Z and β_j^Z.
  3. Given F_Z^0 and the prior guesses F_X^0 and F_Y^0 (these can be extrapolated from data), solve eq. 7 and get F_V^0 and F_W^0.
  4. Compute $\alpha_j^V, \beta_j^V, \alpha_j^W, \beta_j^W$.

Elicitation

Some interesting special cases:

- If F_Z^0 is assumed to be a Dirichlet process degenerate at 0, i.e. $\alpha_0^Z > 0$, $\alpha_j^Z = 0$ for all $j \geq 1$, and $\beta_j^Z = 0$ for all $j \geq 0$, there is no dependence between X and Y, and the joint distribution is the product of the two marginals.
- If $\alpha_j^V, \beta_j^V, \alpha_j^W, \beta_j^W \to 0$ for all j, $\hat{S}(x, y)$ is simply the product of the two Kaplan-Meier estimators of X and Y.

TU Delft

Annuity data

- The data set [1] used in this application is the same one of Frees et al. (1996) and Luciano et al. (2008).
- It contains information about 14947 contracts of a large Canadian insurance company from 29.12.1988 to 31.12.1993.
- The contracts are joint and last-survivor annuities in payout status over the observation period.
- For every contract we know:
  1. dates of birth and, if applicable, of death,
  2. date of contract initiation,
  3. income class of the individuals (not used here),
  4. sex of each annuitant.

[1] I am grateful to E. Valdez for providing the data.

Statistics

Number of contracts by sex, age and mortality:

Age	Alive	Dead	Total
MALES			
Less than 60	1170	42	1212
60-70	7620	534	8154
70-80	4355	806	5161
More than 80	229	177	406
Total	13374	1559	14933
FEMALES			
Less than 60	2962	30	2992
60-70	8222	239	8461
70-80	3014	245	3259
More than 80	186	63	249
Total	14384	577	14961

Statistics

- Male deaths are roughly 3 times as many as female deaths.
- This is due to:
  - the average entry age for males is 68, against 65 for females,
  - males have a higher mortality rate (as known).
- Other facts:
  - youngest couple: (32,30); oldest: (109,99) [both censored!]
  - youngest dead couple: (57,52); oldest (98,99)
  - youngest male death: 39; female: 38.
  - oldest male death: 98; female: 99.
  - percentage of dead couples: 1.5%
  - percentage of couples with survivors: 12.75%
  - sex primary annuitant: 80.52% males.
  - sex secondary annuitant: 80.61% females.

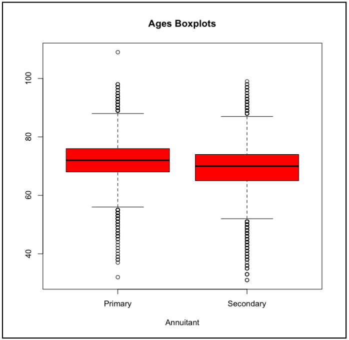
TU Delft

Statistics

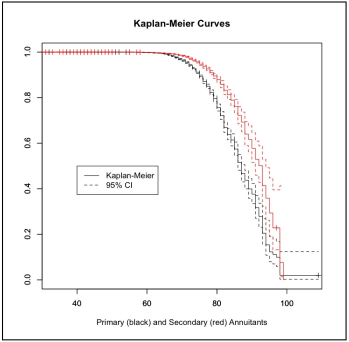
- If we focus on the 231 pairs of deaths, we have the following about time separation:
  - 29 occur in 1 day (simultaneous deaths): 12.55%,
  - 63 within 5 days: 27.27%,
  - 70 within 10 days: 30.30%,
  - 86 within 30 days: 37.23%,
  - 128 within 1 year: 55.41%,
  - 175 within 2 years: 75.32%,
  - 100% within 4 years.

The marginals

- Individuals' ages are (from now on) discretized.
- To consider all couples, even same sex ones, we no more distinguish between males and females, but between the roles in the annuity contract.



Boxplots Comparison



Kaplan-Meier Estimators

TU Delft

The marginals

Frees et al. (1996) have shown that the ecdf of the marginals (for sex, but also for the role in the contract) are well approximated by Gumbel distributions of the form

$$F(x) = 1 - \exp\left(e^{-\frac{\mu}{\sigma}}\left(1 - e^{\frac{x}{\sigma}}\right)\right). \tag{15}$$

- $\mu = 68.4$ and $\sigma = 9.8$ for males.
- $\mu = 64.7$ and $\sigma = 11.1$ for females.

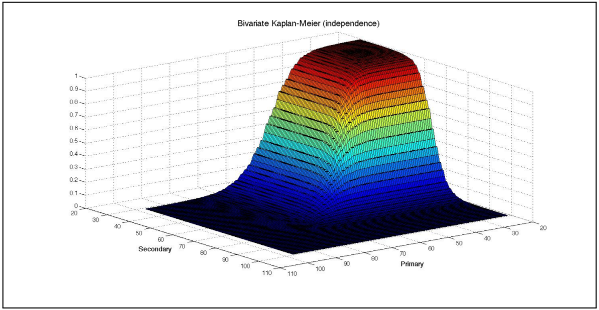
We have also computed:

- $\mu = 68.5$ and $\sigma = 9.8$ for primary annuitants.
- $\mu = 65.5$ and $\sigma = 11.1$ for secondary annuitants.

It is interesting to notice how the estimates for males and primary annuitants, and for females and secondary annuitants are similar.

The Joint

To have an idea of the joint distribution, we can use the Kaplan-Meier estimator by assuming independence of X and Y.
The result is the following for annuitants:



Similar results do hold for sex.

The dependence between X and Y

In reality X and Y are not independent:

- Using Spearman's rank correlation, as suggested in Frees et al. (1996), we found out that the correlation is $\approx 0.43 \pm 0.11$.
- Even with all the caveats, this simple measure suggests dependent lives.
- Moreover we can rely on all the analyses performed by Frees et al. (1996) and Luciano et al. (2008).
- Since we will need it later: $Cov(X, Y) = 36.01$.

Initialization of the model

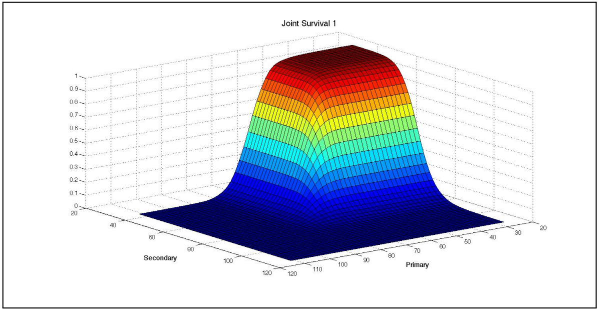
To initialize our model we need to choose the priors for Z (with variance $Cov(X, Y)$), X and Y.
We do not have any strict requirement, apart from the one on the variance of F_Z. We can choose simple distributions to deal with!
In particular, for its closure under convolution and its broad use in mortality studies, we may choose a Poisson distribution.

- We center F_Z on $Poi(36)$, F_X on $Poi(71)$ and F_Y on $Poi(69)$, where 71 and 69 are the rounded values for the average ages of primary and secondary annuitants.
- By solving eq. 7 we get that F_V is $Poisson(35)$ and F_W is $Poisson(33)$.
- For the strengths of belief, that we use to center our GPS, we set
  - $c_j^Z, c_j^V, c_j^W = 1\ \forall j \rightarrow$ CASE 1
  - $c_j^Z, c_j^V, c_j^W = 5\ \forall j \rightarrow$ CASE 2

TU Delft

Results

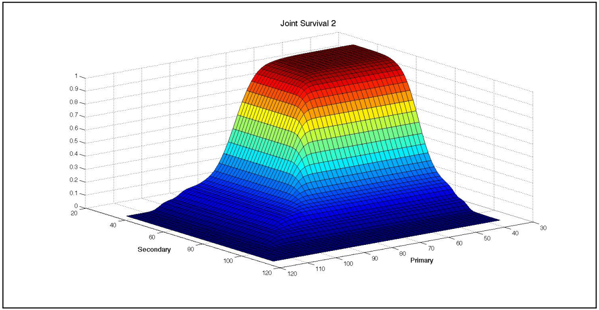
This is the joint survival function under CASE 1.
It is possible to see how the graph is a smoothed version of the Kaplan-Meier estimator we have seen before.
In particular, remember that we can get KM, by setting F_Z degenerate at 0, $\alpha_0^Z = 1000$, $\alpha_j^Z = 0$ for all $j \geq 1$, $\beta_j^Z = 0$ for all $j \geq 0$, and all the other α and β close to 0.



TU Delft

Results

This is the joint survival function under CASE 2.
In this case we have given a higher weight to our initial guesses, and the result is a distribution that gives more mass to middle-high values in the data.
In other words, the posterior is more dependent on our prior knowledge.



Results

Since we are in a Bayesian framework, we can use the algorithm we have developed to compute our bivariate posterior distribution, to perform some prediction.

In particular, we can always ask: *We have n couples, what is the probability that couple* $n + 1$ *is (70*-75*,65-70)?*

The answer obviously depends on the model we are using:

- CASE 1: 0.15
- CASE 2: 0.19

TU Delft

Thanks!