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.
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. \( \lim_{n \to \infty} \prod_{j=0}^{n} \frac{\beta_j}{\alpha_j + \beta_j} = 0 \).

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

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

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

where \( t_n = (t_1, ..., t_n) \), \( m_j(t_n) = \sum_{k=1}^{n} 1_{[t_k = j]} \), \( r_j(t_n) = \sum_{k=1}^{n} 1_{[t_k > j]} \)
and \( s_j(t_n) = m_j(t_n) + r_j(t_n) \).
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 \( \alpha_j \) white balls and \( \beta_j \) 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.
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 $T_1, \ldots, T_n$ be independent and identically distributed random variables, subject to right-censoring. What we observe is $(T^*_1, \delta_1), \ldots, (T^*_n, \delta_n)$, with

- $T^*_i = t, \delta_i = 0$ if a censoring took place, i.e. $T_i > t$,
- $T^*_i = t, \delta_i = 1$ if no censoring took place, i.e. $T_i = t$.

With a quadratic loss function, the predictive distribution of $T_{n+1}$ given $(T_n, \delta_n)$ 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):

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

(2)

where \( m_j^*(t_n, d_n) = \sum_{k=1}^{n} 1[t_k = j, d_k = 1] \) and \( d_n \in \{0, 1\}^n \).

Notice that:

- in case of no censoring, eq. 2 is given by
  \[
  \hat{S}(t) = \prod_{j=0}^{t} (\beta_j + r_j(t_n))/(\alpha_j + \beta_j + s_j(t_n)).
  \]

- for \( \alpha_j, \beta_j \to 0 \) for all \( j \), eq. 2 reduces to the standard Kaplan-Meier estimator.
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 \( (\alpha^V_j, \beta^V_j) \), \( (\alpha^W_j, \beta^W_j) \) and \( (\alpha^Z_j, \beta^Z_j) \);
- define the random process \( \{(X_n, Y_n), n \geq 1\} \) with
  \[
  X_n = Z_n + V_n, \\
  Y_n = Z_n + W_n.
  \]
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;
- \(\sigma(Z_n, V_n, W_n) = \sigma(Z_n, X_n, Y_n)\);
- the dependence structure is given by

\[
\begin{align*}
\text{Cov}(X_1, Y_1) &= \text{Var}(Z_1) \geq 0, \\
\text{Cov}(X_{n+1}, Y_{n+1} \mid Z_n, V_n, W_n) &= \text{Var}(Z_{n+1} \mid Z_n), \ n \geq 1.
\end{align*}
\]
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
  \[
  F_X = F_Z \ast F_V, \\
  F_Y = F_Z \ast F_W. 
  \]
  (7)
  (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 \( F \),
  \[
  P_{XY}(x, y) = \sum_{z=0}^{x \wedge y} P_Z(z)P_V(x - z)P_W(y - z), \quad \forall (x, y) \in \mathbb{N}_0^2. 
  \]
  (9)
Main Properties

• If $\sigma_Z^2 = \text{Var}_{F_Z}(Z)$, then

$$\text{Cov}_{F_{XY}}(X, Y) = \sigma_Z^2.$$  \hspace{1cm} (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 $\mathbb{N}_0^2$. 
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 \((X_n, Y_n)\) be an independent and identically distributed sample from \( S \). Our interest is related to the predictive

\[ \hat{S}(x, y) = P[X_{n+1} > x, Y_{n+1} > y | X_n = x_n, Y_n = y_n]. \quad (11) \]

- If the elements in \((X_n, Y_n)\) are not subject to right-censoring, it is possible to compute eq. 11 as (see Cirillo, 2008)

\[ \hat{S}(x, y) = \frac{P[X_{n+1} > x, Y_{n+1} > y, X_n = x_n, Y_n = y_n]}{P[X_n = x_n, Y_n = y_n]}. \quad (12) \]
Estimating the bivariate survival function

If some of the elements in \((X_n, Y_n)\) are right-censored, it is not possible to obtain a closed-form expression for

\[
\hat{S}(x, y) = P[X_{n+1} > x, Y_{n+1} > y | X_n^* = x_n, \delta_n = d_n, Y_n^* = y_n, \epsilon_n = e_n].
\]

(13)

Anyway it is always possible to use a MCMC estimation.
Set the parameters of the model, then compute $\hat{S}(x, y)$ using the following algorithm:

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

$$P_{Z_n \mid Z_{n-1}, X^*_n, \delta_n, Y^*_n, \epsilon_n} \propto P[V^*_n = x_n - z_n, \delta_n = d_n \mid V^*_{n-1} = x_{n-1} - z_{n-1}, \delta_{n-1} = d_{n-1}]$$

$$\times P[W^*_n = y_n - z_n, \epsilon_n = e_n \mid W^*_{n-1} = y_{n-1} - z_{n-1}, \epsilon_{n-1} = e_{n-1}]$$

$$\times P[Z_n = z_n \mid Z_{n-1} = z_{n-1}]$$

where, for example,

$$P[W^*_n = w, \epsilon_n = e \mid W^*_{n-1} = w_{n-1}, \epsilon_{n-1} = e_{n-1}]$$

$$= \begin{cases} P[W^*_n \geq w \mid W^*_{n-1} = w_{n-1}, \epsilon_{n-1} = e_{n-1}] & \text{if } e = 0, \\ P[W^*_n = w \mid W^*_{n-1} = w_{n-1}, \epsilon_{n-1} = e_{n-1}] & \text{if } e = 1. \end{cases}$$

- Since $\{Z_n\}$ is exchangeable, all the conditionals $P_{Z_n \mid Z_{-j}, X^*_n, \delta_n, Y^*_n, \epsilon_n}$, where $Z_{-j} = \{Z_i\}_{i=1}^n \setminus \{Z_j\}$, have the same form.

- Compute $V^*_n = X^*_n - Z_n$ and $W^*_n = Y^*_n - Z_n$.

- $Z_{n+1}, V_{n+1}$ and $W_{n+1}$ are sampled w.r.t $P_{Z_{n+1} \mid Z_n}$, $P_{V_{n+1} \mid V^*_n, \delta_n}$ and $P_{W_{n+1} \mid W^*_n, \epsilon_n}$. 

---

**TU Delft**

Pasquale Cirillo (AP - TU Delft)  
Modeling Joint Mortality  
Longevity 12  
15 / 32
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

\[
\alpha_j = c_j(G(\{j\})) \quad \quad \beta_j = c_j \left(1 - \sum_{i=0}^{j} (G(\{i\}))\right).
\]

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 $\text{Cov}(X, Y)$ and the marginals of $X$ and $Y$. 
Elicitation

• Using eqs. 7 and 10, we can formulate the following procedure:

1. Express an initial distribution $F_0^Z$ for $Z$. If a prior knowledge is available, use it! Otherwise, the choice is free, the only constraint being $\sigma_Z^2 = \text{Cov}(X, Y)$.
2. Determine $\alpha_j^Z$ and $\beta_j^Z$.
3. Given $F_0^Z$ and the prior guesses $F_0^X$ and $F_0^Y$ (these can be extrapolated from data), solve eq. 7 and get $F_0^V$ and $F_0^W$.
4. Compute $\alpha_j^V$, $\beta_j^V$, $\alpha_j^W$, $\beta_j^W$. 
Elicitation

Some interesting special cases:

- If $F^0_Z$ is assumed to be a Dirichlet process degenerate at 0, i.e. $\alpha^Z_0 > 0$, $\alpha^Z_j = 0$ for all $j \geq 1$, and $\beta^Z_j = 0$ 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 $\alpha^V_j, \beta^V_j, \alpha^W_j, \beta^W_j \to 0$ for all $j$, $\hat{S}(x, y)$ is simply the product of the two Kaplan-Meier estimators of $X$ and $Y$. 
Annuity data

• The data set used in this application is the same one of Frees et al. (1996) and Luciano et al. (2008).


• 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.

---

I am grateful to E. Valdez for providing the data.
Statistics

Number of contracts by sex, age and mortality:

<table>
<thead>
<tr>
<th>Age</th>
<th>Alive</th>
<th>Dead</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MALES</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 60</td>
<td>1170</td>
<td>42</td>
<td>1212</td>
</tr>
<tr>
<td>60-70</td>
<td>7620</td>
<td>534</td>
<td>8154</td>
</tr>
<tr>
<td>70-80</td>
<td>4355</td>
<td>806</td>
<td>5161</td>
</tr>
<tr>
<td>More than 80</td>
<td>229</td>
<td>177</td>
<td>406</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>13374</td>
<td>1559</td>
<td>14933</td>
</tr>
<tr>
<td><strong>FEMALES</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than 60</td>
<td>2962</td>
<td>30</td>
<td>2992</td>
</tr>
<tr>
<td>60-70</td>
<td>8222</td>
<td>239</td>
<td>8461</td>
</tr>
<tr>
<td>70-80</td>
<td>3014</td>
<td>245</td>
<td>3259</td>
</tr>
<tr>
<td>More than 80</td>
<td>186</td>
<td>63</td>
<td>249</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>14384</td>
<td>577</td>
<td>14961</td>
</tr>
</tbody>
</table>
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.
• 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
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

\[ F(x) = 1 - \exp \left( e^{-\frac{\mu}{\sigma}} \left( 1 - e^{\frac{x}{\sigma}} \right) \right). \]  

(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:

![Bivariate Kaplan-Meier (independence)](image)

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: $\text{Cov}(X, Y) = 36.01$. 
Initialization of the model

To initialize our model we need to choose the priors for $Z$ (with variance $\text{Cov}(X, Y)$), $X$ 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 $\text{Poi}(36)$, $F_X$ on $\text{Poi}(71)$ and $F_Y$ on $\text{Poi}(69)$, 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 $\text{Poisson}(35)$ and $F_W$ is $\text{Poisson}(33)$.
- For the strengths of belief, that we use to center our GPS, we set
  - $c_j^Z, c_j^V, c_j^W = 1 \forall j \rightarrow \text{CASE 1}$
  - $c_j^Z, c_j^V, c_j^W = 5 \forall j \rightarrow \text{CASE 2}$
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 $\alpha$ and $\beta$ close to 0.
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.
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
Thanks!