Long-Term Care with Multi-State Models

Quentin Guibert

ISFA, Université de Lyon, Université Claude-Bernard Lyon-I
Autorité de Contrôle Prudentiel et de Résolution
Email: quentin.guibert@acpr.banque-france.fr

7-9 September 2015
Longevity 11 - Lyon

Joint work with F. Planchet (ISFA and Prim’Act)

The views expressed in this presentation are those of the authors and do not necessarily reflect those of the Autorité de Contrôle Prudentiel et de Résolution (ACPR), neither those of the Banque de France.
Outline

1. Context and Motivations
2. Literature overview
3. Acyclic multi-state model
4. Application
Demographic and insurance context

- Significant increase of health costs for elderly people in recent decades
- This trend will continue in future with a lot of uncertainty...
- Long-term care (LTC) insurance products in addition to social benefits
- Payment of benefits depends on the level of dependency (functional disability)
- No uniform definition and grid to measure the severity. Generally, contractual grids use criteria depending on the number of ADLs (wash, eat, dress, move, ...)
- A wide range of insurance products (short and long terms). LTC insurance may also be individual or collective.
- In France, contracts contain lots of policy clauses (whole life annuity vs. policy term, deferred period, maximum age, deductible)
Regulatory context

- Solvency II offers a great role for actuaries
- Need for realistic (best estimate) assumptions. Actuaries are responsible for the data quality (accuracy, completeness) and the adequacy between data and models for reserving
- Pay close attention to bias (selection bias, information bias,...) and to the type of available data (e.g. continuous, discrete time, censorship) to select the best inference methods
- Need to regularly check biometric assumptions
- External data and expert opinion should be justified
- For LTC insurance, take account for the appropriately granular level and risk dynamics are great challenges
Current practices and available data

- Multi-state models are the most natural tools for pricing and reserving LTC guarantees, e.g. the illness-death model for only one heavy dependency state

\[
\begin{align*}
0: \text{Health} & \xrightarrow{\mu_{01}(t)} 1: \text{Disability} \\
& \xrightarrow{\mu_{02}(t)} 2: \text{Death} \\
& \xrightarrow{\mu_{12}(t)}
\end{align*}
\]

- In the literature, large aggregated national dataset are usually used
- Introduce a Markov process \( X \) that describes the current state of a policyholder
- Quantities of interest:

\[
p_{hj}(s, t) = \mathbb{P}(X_t = j \mid X_s = h) \quad \text{and} \quad \mu_{hj}(t) = \lim_{\Delta t \to 0} \frac{p_{hj}(t, t + \Delta t)}{\Delta t}
\]
Current practices and available data

- Researchers assume that the Markov assumption is satisfied and are interested in fitting the quantities of interest (e.g. Haberman and Pitacco, 1998; Pritchard, 2006; Levantesi and Menzietti, 2012; Fong et al., 2015)

- Inference methodology $\rightarrow$ GLM Poisson models that depend on age $x$ (CMIR12, 1991)

$$
\eta \left( \mathbb{E} \left[ \frac{d_{hj}(x)}{e_h(x)} \right] \right) = a_k x^k + a_{k-1} x^{k-1} + \ldots + a_0
$$

- Lack of (detailed) national data. No covariate. Determining trends is quite complex (Gouriéroux and Lu, 2014)
Motivations

- Insurers should use their own data = \textit{longitudinal data} in continuous time with censorship and truncation $\implies$ we do not discuss the other cases

- It is time to develop statistical methods for multi-state models taking into account the data features. Non-parametric techniques $\implies$ goodness-of-fit checks

- Practitioners often use methods developed for survival analysis (Guibert and Planchet, 2014)

- The Markov assumption is clearly not satisfied

\textbf{Figure:} Fitted forces of mortality for LTC claimants (Tomas and Planchet, 2013)
Markov case: non-parametric inference for censored data

- Inference technique application to all Markov multi-state models
- C is an independent, non-informative, right censoring variable. We observe the censored process
- Based on counting process theory (Andersen et al., 1993)

\[ N_{hj}(t) = \# \{0 \leq \tau \leq t : X_\tau = j, X_{\tau^-} = h, 0 \leq \tau \leq C \} \]

\[ L_h(t) = \mathbb{1}_{\{X_{\tau^-}=h, 0 \leq t \leq C\}} \text{ and } N(t) = \sum_{h,j} N_{hj}(t) \]

- Under \( (\mathcal{F}_t) \) the canonical filtration generated by \( N \) and \( X_0 \) for all \( h \to j \)

\[ N_{hj}(t) - \int_0^t L_h(\tau) dA_{hj}(\tau) = N_{hj} - \int_0^t L_h(\tau) \mu_{hj}(\tau) d\tau \]

are martingale.
Markov case: non-parametric inference for censored data

- Transition intensities are estimated by the Nelson-Aalen estimators

\[ \hat{A}_{ij}(t) = \int_0^t \frac{dN_{ij}(\tau)}{L_h(\tau)} = \sum_{\{k : t_k \leq t\}} \frac{d_{ij}(t_k)}{L_h(t_k)} \]

- Heterogeneous population can be modeled with semi-parametric approaches, e.g. the Cox proportional hazard model

\[ \mu_{ij}(t | Z_{ij}, \theta) = \mu_{0ij}(t) \exp(\theta^\top Z_{ij}) \]

- Transition probabilities matrices \( p \) are obtained with the Aalen-Johansen estimators

\[ \hat{p}(s, t) = \mathcal{P} \left( \text{Id} + d\hat{A}(\tau) \right)_{\tau \in [s,t]} \]

with \( \mathcal{P} \) the integral-product operator

- Generalization of the Kaplan-Meier (KM) estimator for survival data (Kaplan and Meier, 1958)
Example: LTC insurance data

- Database from a large French LTC insurer (Guibert and Planchet, 2014)
- Entry in dependency is distinguished by pathology (different waiting periods)
- \( \simeq 210,000 \) contracts observed during the period 1998-2010 after cleaning the database and almost 70% are censored

4 types of pathology and 2 direct exit causes.

<table>
<thead>
<tr>
<th>Exit causes</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>( e_1 ) Neurologic pathologies</td>
<td>2.5%</td>
</tr>
<tr>
<td>( e_2 ) Various pathologies</td>
<td>2.7%</td>
</tr>
<tr>
<td>( e_3 ) Terminal cancers</td>
<td>2.4%</td>
</tr>
<tr>
<td>( e_4 ) Dementia</td>
<td>5.4%</td>
</tr>
<tr>
<td>( e_5 ) Death</td>
<td>52.2%</td>
</tr>
<tr>
<td>( e_6 ) Cancel</td>
<td>34.8%</td>
</tr>
</tbody>
</table>
Example: LTC insurance data

Actuaries are interested in the inception rates

$$q_j(t) = p_{0j}(t, t + 1)$$

$$\hat{q}_j(t) = \sum_{\{k : t < t_k \leq t + 1\}} \frac{\hat{S}(t_k)}{\hat{S}(t)} \frac{d_{0j}(t_k)}{L_0(t_k)}$$

Figure: Inception rates estimates with approximate pointwise 95% confidence intervals.
Semi-Markov model

- Let $S_1 < S_2 < \ldots < S_k < \ldots$ be the ordered jump times for the process $X$
- Let $J_k$ be the discrete time process which gives the state occupied by $X$ between times $S_k$ and $S_{k+1}$
- Duration time in the current state $U_t = t - S_{N(t)}$

Definition

If the discrete time process $(S_k, J_k)$ is a Markov process. It is called a Markov renewal process and is built up by an initial distribution and a semi-Markov kernel

$$Q_{hj}(s, t) = P(\Delta S_{k+1} \leq t, J_{k+1} = j \mid S_k = s, J_k = h)$$

Then, $(X_t, U_t)$ is Markov and the process $(X_t)$ is called a semi-Markov process with its canonical filtration $(\mathcal{F}_t)$.

- Transition probabilities: $p_{hj}(s, t, u, v) = P(X_t = j, U_t \leq v \mid X_s = h, U_s = u)$
- Transition intensities: $\mu_{hj}(t, u) = \lim_{\Delta t \to 0} \frac{p_{hj}(t, t + \Delta t, u, \infty)}{\Delta t}$
Inference for homogeneous semi-Markov model

- Homogeneous semi-Markov process: $Q_{hj}(s, t) = Q_{hj}(t)$
- No problem to infer parametric model. We regard non-parametric model
- Model without loop: can be estimated similarly to a Markov model
  $\Rightarrow$ Many situations in actuarial science
- Model with loops: the semi-Markov kernel is estimated non-parametrically
  (Gill, 1980) by
  \[ \hat{Q}_{hj}(t) = \int_0^t \left(1 - \hat{H}_h(\tau)\right) \frac{dN_{hj}(\tau)}{L_h(\tau)}, \]
  where $H_h(u) = P(\Delta S_{k+1} \leq u \mid J_k = h)$ and this function is estimated with
  Kaplan-Meier.
- The processes $N_{hj}(u)$ and $L_h(u)$ depend on the time $u$ spends in state $h$
- But transition probabilities $\Psi_{hj}(t) = P(X_t = j \mid X_0 = h)$ are tricky to compute (Spitoni et al., 2012)
Non-homogeneous semi-Markov and non-Markov models

- **Non-homogeneous semi-Markov without loop:**
  - Most of the time, one of the time variable is considered as a covariate
  - The splitting of state approach (Haberman and Pitacco, 1998)
  - Actuarial approach for the disability model: estimating the survival function in the disability state, for e.g. by Kaplan-Meier, splitting the sample by age (integer) \( \sim \) survival data with staggered entry
  - Cox semi-Markov model (Andersen and Perme, 2008)

\[
\mu_{hj} (t \mid Z_{hj,i}, U_t, \theta) = \mu_{0hj} (t) \exp \left( \theta_0 f (U_t) + \theta^\top Z_{hj,i} \right)
\]

- **Non-homogeneous semi-Markov with loops:**
  - General framework not available for non-parametric inference
  - Cox specification is also applicable (Dabrowska, 1995)
  - Parametric approaches where intensities or kernels are written as a product of two uni-dimensional functions (Monteiro et al., 2006; Mathieu et al., 2007)
Approaches based on direct probabilities

- Meira-Machado et al. (2006): transition probabilities for an acyclic illness-death model without the Markov assumption
- Let $S$, the lifetime in healthy state and $T$ the overall lifetime
- Let $C$ a independent right-censored variable. We observe

  \[
  \begin{align*}
  Y &= \min (S, C) \quad \text{and} \quad \gamma = \mathbb{1}_{\{s \leq C\}} \\
  Z &= \min (T, C) \quad \text{and} \quad \delta = \mathbb{1}_{\{T \leq C\}}
  \end{align*}
  \]

- Transition probabilities are viewed as functional under the joint distribution of $(S, T)$ and estimated using Kaplan-Meier integral

\[
\begin{align*}
p_{00}(s, t) &= \frac{\mathbb{P}(S > t)}{\mathbb{P}(S > s)} \\
p_{01}(s, t) &= \frac{\mathbb{P}(s < S \leq t < T)}{\mathbb{P}(S > s)} = \frac{\mathbb{E} \left[ \varphi_{st}^{(1)} (S, T) \right]}{\mathbb{P}(S > s)} \\
p_{11}(s, t) &= \frac{\mathbb{P}(S \leq s, t < T)}{\mathbb{P}(S \leq s < T)} = \frac{\mathbb{E} \left[ \varphi_{st}^{(2)} (S, T) \right]}{\mathbb{E} \left[ \varphi_{ss}^{(2)} (S, T) \right]}
\end{align*}
\]
Acyclic multi-state model

Let an acyclic multi-state model which refers to a situation where both terminal and non-terminal events.

Formally, two lifetimes are identified:

- \( S \), the lifetime in healthy state
  \[
  S = \inf \{ t : X_t \neq a_0 \},
  \]
- \( T \), the overall lifetime
  \[
  T = \inf \{ t : X_t \in \{d_1, \ldots, d_{m_2}\} \},
  \]
where \((X_t)_{t \geq 0}\) is the current state of the individual.
Main goals

- With independent right-censoring (non informative) variable
- No Markov assumption

Main goals

- Non-parametric estimation of transition probabilities for a such a right censoring acyclic multi-state model
- Define association measure between the failure time in healthy state and the overall lifetime
Motivations

Literature overview

Acyclic multi-state model

Application

Summary

Existing Estimators for Competing Risks Data

- Let $V$ be the indicator of the type of failure. The Aalen-Johansen (AJ) estimator for the cumulative incidence function (CIF) which is the joint distribution of $(T, V)$ is

$$F^{(v)}(t) = \mathbb{P}(T \leq t, V = v)$$

Non-parametric estimator for CIF

- i.i.d. observations are composed of $(Z_i, \delta_i, \delta_i V_i, )_{1 \leq i \leq n}$

- Estimator can be expressed as a sum considering the ordered $Z$-values

$$\hat{F}_n^{(v)}(z) = \sum_{i=1}^{n} \tilde{W}_{in} J_{[i:n]}^{(v)} 1\{Z_{i:n} \leq z\}, \quad \tilde{W}_{in} = \frac{\delta_{[i:n]}}{n-i+1} \prod_{j=1}^{i-1} \left( \frac{n-j}{n-j+1} \right)^{\delta_{[j:n]}}$$

- $\tilde{W}_{in}$ is the Kaplan-Meier (KM) weights and $J_{i}^{(v)} = 1\{V_i=v\}$

- $\hat{F}_n^{(v)}(\cdot)$ converges w.p.1 to $F^{(v)}(\cdot)$ and is asymptotically normal
Bivariate Competing Risks Data

- **Idea:** our model = a bivariate competing risks models with a unique right-censoring variable. Non-parametric inference is studied by Cheng et al. (2007) for a more general case.

- Let \((S, V_1)\) and \((T, V)\) be 2 competing risks processes where:
  - \(V_1\): indicator taking its values in the set of arrival states by direct transition from \(a_0\)
  - \(V = (V_1, V_2)\) with is \(V_2\) indicator taken its values in the set of arrival states from non-terminal events

**Bivariate CIF estimator**

\[
\hat{F}_{0n}^{(v)}(y, z) = \sum_{i=1}^{n} \tilde{W}_{in} J_{[i:n]}^{(v)} \mathbb{1}_{Y_{[i:n]} \leq y, Z_{i:n} \leq z}
\]

- Simple form for the weights as \((S, V_1)\) is observed whether \(T\) is observed
- \(\hat{F}_{0n}\) is weakly convergent under independent censoring
Aalen-Johansen Integrals Estimators

- Consider an integral of the form \( S_{(v)}(\varphi) = \int \varphi \, dF_{0_{(v)}} \) with \( \varphi \) a generic function

- \( S \) can be considered as a covariate

**AJ integrals**

\[
\hat{S}_{n_{(v)}}(\varphi) = \int \varphi(s, t) \hat{F}_{0_{(v)}}(ds, dt) = \sum_{i=1}^{n} \hat{W}_{in_{(v)}}(Y_{i:n}, Z_{i:n}) \varphi, \quad 0 \leq s \leq t \leq \tau Z.
\]

- \( W_{in_{(v)}} = W_{in_{J_{[i:n]}}} \), AJ weights (Suzukawa, 2002) for competing risks data

- **Possibility** to take into account the left-truncation \( L \) considering

\[
\hat{W}_{in_{(v)}} = \frac{\delta_{[i:n]}J_{[i:n]}}{nC_n(Z_{i:n})} \prod_{j=1}^{i-1} \left( 1 - \frac{1}{nC_n(Z_{i:n})} \right)^{\delta_{[j:n]}}.
\]

where \( C_n(x) = n^{-1} \sum_{i=1}^{n} \mathbb{1}_{L_i \leq x \leq Z_i} \)
Transition Probabilities Estimators

Application for estimating key probabilities in actuarial science i.e.

\[
p_{0e}(s, t, \eta) = \frac{\mathbb{P}(s < S \leq \min(t, t-\eta), T > t, V_1 = e)}{\mathbb{P}(S > s)},
\]

\[
p_{ee}(s, t) = \frac{\mathbb{P}(S \leq s, T > t, V_1 = e)}{\mathbb{P}(S \leq s, T > s, V_1 = e)},
\]

\[
p_{ed}(s, t, \eta, \zeta) = \frac{\mathbb{P}(\eta < T - S \leq \zeta, s < S \leq t, V = (e, d))}{\mathbb{P}(T - S > \eta, s < S \leq t, V_1 = e)}.
\]

Remarking that \(\{V_1 = e\} = \{V_1 = e, V_2 \in C_e\}\) where \(C_e\) is the set of children (i.e. transition states from \(e\)) related to the state \(e\), we can refer to our AJ integrals estimators.
Transition Probabilities Estimators

\[ \hat{p}_{0e}(s, t, \eta) = \frac{\hat{S}_{n}^{(e, C_{e})}(\varphi_{s, t, \eta})}{1 - \hat{H}_{n}(s)}, \text{ with } \varphi_{s, t, \eta}(x, y) = \mathbb{1}\{s < x \leq \min(t, t - \eta), y > t\}, \]

\[ \hat{p}_{ee}(s, t) = \frac{\hat{S}_{n}^{(e, C_{e})}(\varphi_{s, t})}{\hat{S}_{n}^{(e, C_{e})}(\varphi_{s, s})}, \text{ with } \varphi_{s, t}(x, y) = \mathbb{1}\{x \leq s, y > t\}, \]

\[ \hat{p}_{ed}(s, \eta, \zeta) = \frac{\hat{S}_{n}^{(e, d)}(\varphi_{s, \zeta})}{\hat{S}_{n}^{(e, C_{e})}(\varphi_{s, \eta})}, \text{ with } \varphi_{s, \zeta}(x, y) = \mathbb{1}\{s < x \leq t, \eta < y - x \leq \zeta\}, \]

\[ \varphi_{s, \eta}(x, y) = \mathbb{1}\{s < x \leq t, \eta < y - x\} \text{ and } \hat{H}_{n} \text{ is the KM estimator of the distribution function of } S. \]

⇒ Our estimators generalize those of Meira-Machado et al. (2006)
Association measures

- Multivariate competing risks model (Scheike and Sun, 2012) → we introduce local association measures based on cross-odds ratio

\[ \pi_0^{(e,d)}(s,t) = \frac{\text{odds}(T \leq t, V_2 = d \mid S \leq s, V_1 = e)}{\text{odds}(T \leq t, V_2 = d \mid V_1 = e)}, \]

where \( \text{odds}(A) = \frac{\text{P}(A)}{1 - \text{P}(A)} \).

- Measure dependence between the lifetime in healthy state and the overall lifetime per cause

- Non-parametric estimator

\[ \hat{\pi}_{0n}^{(e,d)}(s,t) = \frac{\hat{F}_{0n}^{(e,d)}(s,t)}{\frac{\hat{H}_{0n}^{(e)}(s) - \hat{F}_{0n}^{(e,d)}(s,t)}{\hat{F}_{n}^{(e,d)}(t)}} = \frac{\frac{\hat{F}_{0n}^{(e,d)}(s,t)}{\hat{H}_{0n}^{(e)}(\infty) - \hat{F}_{n}^{(e,d)}(t)}}{\hat{F}_{n}^{(e,d)}(t)}, \]

where \( \hat{H}_{0n}^{(e)} \) is the estimator of the CIF of \( S \) for cause \( V_1 = e \) and \( \hat{F}_{n}^{(e,d)} \) is that of \( T \) for cause \( V = (e,d) \).
AJ integrals estimators

Theorem (Consistency)

Assume that

- $\varphi$ is an $F_0$-integrable function,
- $F_0$ and censoring distribution function $G$ are continuous,
- $C$ is independent from the vector $(S, T, V)$.

Then, we have

$$\hat{S}_n^{(v)} (\varphi) \rightarrow S_\infty^{(v)} (\varphi) = \int \mathbb{1}_{\{t<\tau\}} \varphi(s, t) F_0^{(v)} (ds, dt), \ v \in \mathcal{V} \ \text{w.p.1.}$$

- **proof:** Apply a similar strategy than Stute (1993)
AJ integrals estimators

Theorem (Weak convergence)

Assume that:

1. \[ \int \frac{\varphi(S, T)^2 \delta}{(1 - G(T))^2} dP < \infty, \]
2. \[ \int |\varphi(S, T)| \sqrt{C_0(T)} \mathbb{1}_{\{T < \tau_Z\}} dP < \infty, \]

where \( C_0(x) = \int_0^x \frac{G(dy)}{(1 - M(y))(1 - G(y))} \) and \( M(z) = \mathbb{P}(Z \leq z) \).

With the previous assumptions and assuming \( \text{supp}(Z) \subseteq \text{supp}(C) \), we have

\[ \sqrt{n} \left\{ \hat{S}_n(\varphi) - S(\varphi) \right\} \xrightarrow{d} \mathcal{N}(0, \Sigma(\varphi)). \]

- Extendable considering additional (discrete) covariates \( U = (U_1, \ldots, U_p) \) and assuming

\[ \mathbb{P}(T \leq C | S, T, U, V) = \mathbb{P}(T \leq C | T, U, V). \]

- **proof:** Follows ideas used by Stute (1995)
Transition probabilities and association measures

Proposition (Asymptotic results for transition probabilities)

\( \hat{p}_{0e}(s, t, \eta), \hat{p}_{ee}(s, t) \) and \( \hat{p}_{ed}(s, t, \eta, \zeta) \) are consistent w.p.1 if the support of \( Z \) is included in that of \( C \). These estimators admit a weak convergence result.

- Provide estimators when the Markov assumption is released.
- Application to goodness-of-fit testing. Practitioners often use simple multi-state Markov model or Cox semi-Markov model. Misspecification may lead to important errors.

Proposition (Asymptotic results for association measures)

\( \hat{\pi}_{0n}^{(e,d)}(s, t) \) is consistent w.p.1 if the support of \( Z \) is included in that of \( C \) and admits a weak convergence result.

Possible applications to goodness-of-fit testing for models based on cross-odds ratios specification (see Scheike and Sun, 2012).
LTC insurance data

- Same dataset

<table>
<thead>
<tr>
<th>Exit causes</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_1$ Neurologic pathologies</td>
<td>2.5%</td>
</tr>
<tr>
<td>$e_2$ Various pathologies</td>
<td>2.7%</td>
</tr>
<tr>
<td>$e_3$ Terminal cancers</td>
<td>2.4%</td>
</tr>
<tr>
<td>$e_4$ Dementia</td>
<td>5.4%</td>
</tr>
<tr>
<td>$d_1$ Death</td>
<td>52.2%</td>
</tr>
<tr>
<td>$d_2$ Cancel</td>
<td>34.8%</td>
</tr>
</tbody>
</table>

4 types of pathology and 2 direct exit causes.
Transition probabilities

- Estimate annual transition probabilities to become dependent and stay at least one month in a disability state
- Compute pointwise 95% confidence interval from 500 bootstrap resamples
Transition probabilities

- Estimated surface of monthly death rates from each dependent state but quality is low due to missing data

\[ e_1 - \text{Neurologic pathologies.} \]

\[ e_2 - \text{Various pathologies.} \]
Transition probabilities

- Estimated surface of monthly death rates from each dependent state but quality is low due to missing data.

$e_3$-Terminal cancers.

$e_4$-Dementia.
Summary

- Non-parametric estimation for AJ-integrals are applied to estimate this type of acyclic multi-state model under right-censoring.
- These estimators and their properties stay valid if we consider covariates.
- We provide new non-parametric estimators for transition probabilities.
- We exhibit a non-parametric estimator for local association measures.
- We apply them to LTC insurance data to estimate key probabilities.

- Many outlooks
  - Consider framework for regression models.
  - Develop more relevant bootstrap approach for AJ-integrals estimation.
  - Develop semi-parametric approaches based on our local association measure.
  - Consider general estimators for non-homogeneous semi-Markov models.
Thank you for your kind attention.
Some References I


Some References II


Some References III


