Nontrivial Generalization of the Markov Chains and Phase-Type Distributions: A Markov Mixture Approach

Budhi Arta Surya

School of Mathematics and Statistics
Victoria University of Wellington
New Zealand

budhi.surya@vuw.ac.nz

Wrocław, February 14, 2018

Outline

This talk is based on the following work:

Reference: B.A. Surya. (2018). Distributional Properties of the Mixture of Continuous-Time Absorbing Markov Chains Moving at Different Speeds. *INFORMS Journal Stochastic Systems*.

Outline

Markov chain with an absorbing state

The Markov chain and its PHD have been among of the most important probabilistic tools in the analysis of complex stochastic systems evolution.

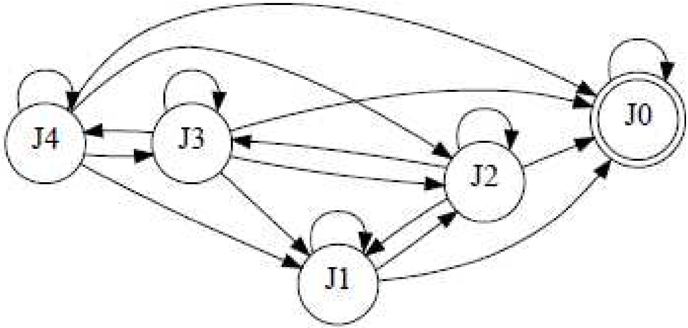


Figure: State diagram of a Markov chain $(X, \mathbb{P})$.

Applications of the Markov model

Markov chain and PHD have been used in various fields, e.g.,

- **Finance/Credit Risk**: Jarrow and Turnbull [22] , Jarrow et.al. [21].
- **Actuarial science**: (Albrecher and Asmussen [4], Lee and Lin [24], Lin and Liu [25], Rolski et al. [29]),
- **Option pricing**: (Asmussen et al. [5], Rolski et. al [29]),
- **Queueing theory**: (Badila et al. [8], Chakravarthy and Neuts [14], Buchholz et al. [13], Breuer and Baum [12], Asmussen [6]),
- **Reliability theory**: (Assaf and Levikson [7], Okamura and Dohi [28]),
- **Survival analysis, Biostatistics**: (Aalen [3], Aalen and Gjessing [2]).
- **Ecological modelling**: Balzter [9], etc.
- **Marketing**: Berger and Nasr (1998), Pfeifer and Carraway (2000).

Remarks

Markov model allows some analytically tractable results in applications.

Outline

Distributional properties of the Markov chain

The time propagation of state changes is represented by a Markov jump process $X = \{X_t : t \geq 0\}$ with finite state space $\mathbb{S} = E \cup \{\Delta\}$, where for some integer $m \geq 1$, $E = \{i : i = 1, ..., m\}$ is transient states and $\Delta = \{m+1\}$ is the absorbing state. The rates at which the process X moves on the transient states E is described by intensity matrix $\mathbf{Q}$:

$$\mathbf{Q} = \begin{pmatrix} \mathbf{T} & \boldsymbol{\delta} \\ \mathbf{0} & 0 \end{pmatrix}, \tag{1}$$

where $\mathbf{T}$ is $m \times m-$ nonsingular matrix with exit vector $\boldsymbol{\delta}$ satisfying

$$\mathbf{T1} + \boldsymbol{\delta} = \mathbf{0}, \tag{2}$$

with $\mathbf{1} = (1, ..., 1)^\top$, as the rows of the intensity matrix $\mathbf{Q}$ sums to zero:

$$q_{ii} \leq 0, \quad q_{ij} \geq 0, \quad \sum_{j \neq i} q_{ij} = -q_{ii} = q_i, \quad (i,j) \in \mathbb{S}. \tag{3}$$

As $\mathbf{1}^\top \boldsymbol{\delta} > 0$, it implies following (2) that $\mathbf{T}$ is negative definite.

Transition probability matrix

By spacial homogeneity of X, for all $t \geq 0$ the transition probability matrix $\mathbf{P}(t) := [p_{ij}(t) = \mathbb{P}\{X_t = j | X_0 = i\}, i, j \in \mathbb{S}]$ solves Kolmogorov equation:

$$\frac{d}{dt}\mathbf{P}(t) = \mathbf{QP}(t), \quad \text{with } \mathbf{P}(0) = \mathbf{I}.$$

Following Theorem 3.4 and Corollary 3.5 in Asmussen (2003),

$$\mathbf{P}(t) = \exp(\mathbf{Q}t), \quad t \geq 0, \tag{4}$$

where $\exp(\mathbf{Q}t)$ is the $(m+1) \times (m+1)$ matrix exponential:

$$\exp(\mathbf{Q}t) = \sum_{n=0}^{\infty} \frac{1}{n!}(\mathbf{Q}t)^n. \tag{5}$$

Partition of the transition matrix

Following the infinite series expansion (5), the transition probability matrix $\mathbf{P}(t)$ (4) of an absorbing Markov chain has the following matrix partition:

Proposition

Let the phase-generator matrix $\mathbf{T}$ *of* $\mathbf{Q}$ *be non-singular. Then,*

$$\mathbf{P}(t) = \begin{pmatrix} e^{\mathbf{T}t} & \mathbf{T}^{-1}(e^{\mathbf{T}t} - \mathbf{I})\delta \\ \mathbf{0} & 1 \end{pmatrix}. \qquad (6)$$

Remark

By the Markov property, the conditional transition probability matrix $\mathbf{P}(t,s) := [p_{ij}(t,s) = \mathbb{P}\{X_s = j | X_t = i\} : i,j \in \mathbb{S}, s \geq t]$ *is given by*

$$\mathbf{P}(t,s) = \mathbf{P}(s-t).$$

Joint distribution of waiting time and the jump

For this purpose, define $\tau_0 = 0$ and the consecutive stopping times $\{\tau_k : k = 1, \ldots\}$ at which the Markov chain makes a new jump, i.e.,

$$\tau_k = \inf\{t > \tau_{k-1} : X_t \neq X_{\tau_{k-1}}, X_{t-} = X_{\tau_{k-1}}\}. \tag{7}$$

Proposition

Let $\{\tau_k\}$ be the stopping times (7). Then, for $x \in E$, $y \in \mathbb{S}$ with $y \neq x$,

$$\mathbb{P}\{\tau_k - \tau_{k-1} \leq t, X_{\tau_k} = y \big| X_{\tau_{k-1}} = x, X_s, s \leq \tau_{k-1} - \} = \big(1 - e^{-q_x t}\big)\frac{q_{xy}}{q_x}$$

Corollary (Likelihood function)

By the Markov property, the likelihood of observing the sample paths of X:

$$L_Q = \prod_x e^{-q_x T_x} \prod_{y \neq x} (q_{xy})^{N_{xy}}$$

Independence of waiting time and the jump

Corollary

$$\mathbb{P}\{\tau_k - \tau_{k-1} \leq t \big| X_{\tau_{k-1}} = x, X_s, s \leq \tau_{k-1}- \} = (1 - e^{-q_x t})$$
$$\mathbb{P}\{X_{\tau_k} = y \big| X_{\tau_{k-1}} = x, X_s, s \leq \tau_{k-1}- \} = \frac{q_{xy}}{q_x}$$

Therefore, conditional on the event $\{X_{\tau_{k-1}} = x\}$, *we observe that*

$$(\tau_k - \tau_{k-1}) \perp\!\!\!\perp X_{\tau_k}.$$

Remarks

The entry q_{xy} *has probabilistic interpretation:*

- $1/q_x$ *is the expected length of time that* X *remains in state* $x \in E$
- q_{xy}/q_x *is the probability that when* X *makes a transition out of state* $x \in E$, *it is to state* $y \in \mathbb{S}$, $y \neq x$.

Lifetime distribution of the Markov chain

The lifetime τ of X and its distribution $\overline{F}(t)$ are defined by

$$\tau = \inf\{t \geq 0 : X_t = \Delta\} \quad \text{and} \quad \overline{F}(t) = \mathbb{P}\{\tau > t\}. \tag{8}$$

Let $\boldsymbol{\pi}$ be $E - \textit{vector}$ representing the initial distribution of X on E.

By Theorem 8.2.3 and Theorem 8.2.5 in Rolski et al. (1998),

$$\overline{F}(t) = \boldsymbol{\pi}^\top e^{\mathbf{T}t}\mathbf{1} \quad \text{and} \quad f(t) = \boldsymbol{\pi}^\top e^{\mathbf{T}t}\boldsymbol{\delta}. \tag{9}$$

The **Phase-Type distribution** $\overline{F}$ is uniquely specified by $(\boldsymbol{\pi}, \mathbf{T})$. We refer among others to Neuts (1975, 1981) and Rolski et al. (1998).

Remarks

- *PHD is closed under finite convex mixtures and finite convolutions*
- *dense property: approximate distribution on* $\mathbb{R}_+$ *arbitrarily well.*

Lifetime intensity / hazard rate

The hazard rate $\alpha(t)dt = \mathbb{P}\{\tau \leq t + dt | \tau > t\}$ is given by

$$\alpha(t) = \frac{f(t)}{\overline{F}(t)} = \frac{\boldsymbol{\pi}^\top e^{\mathbf{T}t}\boldsymbol{\delta}}{\boldsymbol{\pi}^\top e^{\mathbf{T}t}\mathbf{1}}. \tag{10}$$

It has been used in greater applications in survival analysis, Aalen (1995).

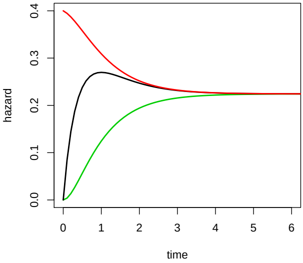


Figure: Birth-death process with different inital state.

Outline

Drawbacks of the Markov chain

Remarks

Some apparent limitations of Markov chain in its practical use:

(i) *each state represents pool of homogeneous objects*

(ii) *unable to model heterogeneity (e.g., age, speeds, track record, etc.) of each individuals within the same pool/state*

(iii) *unable to incorporate its past information in the conditional distribution; the process is always Markovian regardless of the age*

(iv) *regardless of the age and past information, the process never changes the speeds when moving to other states*

(v) *waiting time to move out certain state is exponentially distributed*

(vi) *residual lifetime, default intensity are the same for everyone, etc*

Empirical evidences against the Markov model

Empirical evidences found in Frydman [19], Frydman and Schuerman [18]:

(i) there is strong evidence that firms of the same credit rating may move at different speeds to other ratings, a feature that is lacking in the Markov model.

(ii) the incorporation of past credit information helps improve the out-of-sample prediction of the Nelson-Aalen estimate of the cumulative default intensity for corporate bonds.

Empirical evidences against Markov model: cont'd

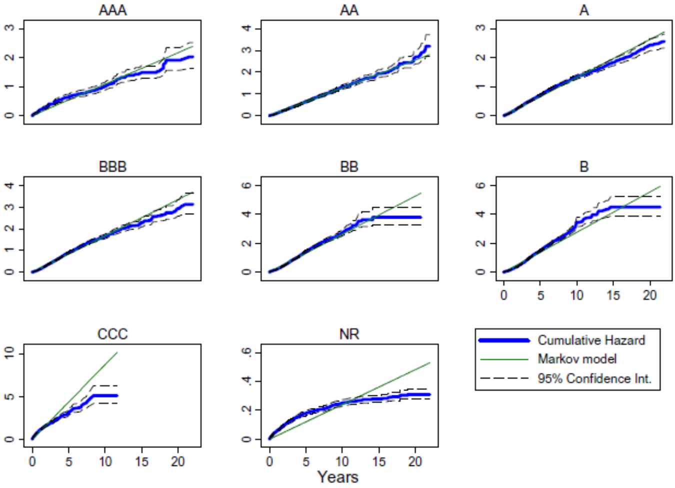


Figure: Source: Frydman and Schuermann, *J. Bank & Financ* (2008)

Outline

Attempts to generalize the Markov chains

There have been some attempts to generalize the Markov chains such as

- Semi-Markov process in which the inter-arrival (waiting) time follows more general distribution other than exponential distribution.

  However, **this approach lacks in explicit identities** regarding distributional properties of the process, and furthermore, it **preserves the Markov property**. See for e.g. **S.I. Resnick (2002)**

- Hidden Markov model in which **the state space is assumed to be unobservable**. Again, the model **preserves the Markov property**. See e.g. **Elliott et al. (2008)** for more details.

- Markov regime switching model. It is a **complicated model**, **no explicit identities available**, and **NO Bayesian updates on the switching probability**

The Markov mixture process: A new approach

Let $\{X_t^{(0)}\}_{t\geq 0}$ and $\{X_t^{(1)}\}_{t\geq 0}$ be finite-state continuous-time Markov chains with intensity matrices $\mathbf{Q}$ and $\mathbf{G}$ resp. **defined on state space** $\mathbb{S}$.

Conditionally on the initial state $X_0 = i_0 \in E$, the mixture is defined by

$$X = \begin{cases} X^{(1)}, & \phi = 1 \\ X^{(0)}, & \phi = 0. \end{cases} \tag{11}$$

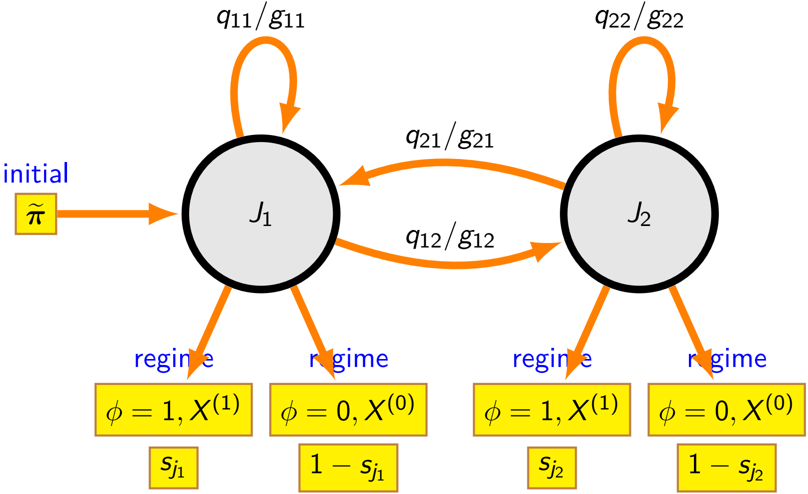
NOTE: the variable ϕ represents the **unobservable speed regime**.

For initial state $X_0 = i_0 \in E$, there is a separate mixing distribution:

$$s_{i_0} = \mathbb{P}\{\phi = 1 | X_0 = i_0\} \quad \text{and} \quad 1 - s_{i_0} = \mathbb{P}\{\phi = 0 | X_0 = i_0\}, \tag{12}$$

Denote by $\widetilde{\boldsymbol{\pi}} = (\boldsymbol{\pi}^\top, \pi_{m+1})^\top$ initial distribution of X, with $\boldsymbol{\pi}$ an $E-$vector.

State Diagram of the Markov mixture process



Throughout the remaining of this talk, we define the intensity matrix **G** as

$$\mathbf{G} = \begin{pmatrix} \mathbf{\Psi T} & \mathbf{\Psi}\delta \\ \mathbf{0} & 0 \end{pmatrix}.$$

where $\mathbf{\Psi}$ is an $(m \times m)$-matrix as such that **G** is an intensity matrix.

Remarks

- *The model was first introduced by Blumen et al. [11] in 1955 as a stochastic model for jobs mobility dynamics (**mover-stayer model**).*
- *Frydman [19], extended the model to a mixture of finite-state Markov chains and provided EM algorithm for the model estimation.*
- *Frydman and Schuermann [18] applied the result [19] for the estimation of credit rating dynamics with relative speed*

$$\mathbf{\Psi} = diag(\psi_1, \psi_2, ..., \psi_m). \tag{13}$$

The Markov mixture process: cont'd

Remarks

- *In general the Markov chains $X^{(0)}$ and $X^{(1)}$ have different rates of leaving the states, i.e., $q_i \neq g_i$,* ***but under [18]*** *both of them may have the same probability of leaving state $i \in E$ to state $j \in \mathbb{S}$, $j \neq i$,*

$$\frac{q_{ij}}{q_i} = \frac{g_{ij}}{g_i}.$$

- *Depending on the value of ψ_i, the Markov chain $X^{(1)}$*
  - *never moves out of state i when $\psi_i = 0$ (**reduced to the mover-stayer model of Blumen et al. (1955)**),*
  - *moves out of state i at lower rate when $0 < \psi_i < 1$*
  - *moves at higher rate when $\psi_i > 1$ (or the same for $\psi_i = 1$) than $X^{(0)}$.*
  - *If $\psi_i = 1$ $\forall i = 1, ..., m$, X* ***simplifies to simple Markov process****.*

The Markov mixture process: cont'd

To incorporate past information, denote by $\mathcal{F}_{t-} = \{X_s, 0 \leq s \leq t-\}$ the available past realizations of the mixtures process X (11) prior to t, and by

$$\mathcal{F}_{i,t} = \mathcal{F}_{t-} \cup \{X_t = i\}$$

all previous and current information of X arriving in state $i \in \mathbb{S}$ at time t.

Remarks

The set $\mathcal{F}_{t-}$ may contain complete, partial or maybe no information of X.

The **Bayesian update of switching probability** is defined by the quantity

$$s_i(t) = \mathbb{P}\{\phi = 1 | \mathcal{F}_{i,t}\}, \quad i \in \mathbb{S}. \tag{14}$$

Outline

Distributional properties of the mixture process

Depending on the availability of the past information $\mathcal{F}_{i,t}$ of X, the elements $s_i(t)$, of the information matrix $\mathbf{S}(t)$ are given below.

Lemma (Bayesian Update of Switching Probability)

For a given $t \geq 0$ and any state $i, i_0 \in \mathbb{S}$, we have

(i) *in case of limited information with $\mathcal{F}_{i,t} = \{X_0 = i_0\} \cup \{X_t = i\}$,*

$$s_i(t) = \frac{\mathbf{e}_{i_0}^\top \mathbf{S}_0 e^{\mathbf{G}t} \mathbf{e}_i}{\mathbf{e}_{i_0}^\top \big(\mathbf{S}_0 e^{\mathbf{G}t} + (\mathbf{I} - \mathbf{S}_0) e^{\mathbf{Q}t}\big) \mathbf{e}_i}. \tag{15}$$

(ii) *in case of limited information with $\mathcal{F}_{i,t} = \{X_t = i\}$,*

$$s_i(t) = \frac{\widetilde{\boldsymbol{\pi}}^\top \mathbf{S}_0 e^{\mathbf{G}t} \mathbf{e}_i}{\widetilde{\boldsymbol{\pi}}^\top \big(\mathbf{S}_0 e^{\mathbf{G}t} + (\mathbf{I} - \mathbf{S}_0) e^{\mathbf{Q}t}\big) \mathbf{e}_i}. \tag{16}$$

Distributional properties of the mixture process: cont'd

Lemma (Bayesian Update of Switching Probability: Cont'd)

(iii) *in case of complete information* $\mathcal{F}_{it} = \{X_s : s \leq t-\} \cup \{X_t = i\}$ *of* X

$$s_i(t) = \frac{s_{i_0} L_G}{s_{i_0} L_G + (1 - s_{i_0}) L_Q}, \tag{17}$$

where L_Q *and* L_G *are likelihood of observing* X *in* $\mathcal{F}_{it}$ *under* $X^{(0)}$ *and* $X^{(1)}$,

$$L_Q = \prod_k e^{-q_k T_k} \prod_{j \neq k} (q_{kj})^{N_{kj}} \quad \text{and} \quad L_G = \prod_k e^{-g_k T_k} \prod_{j \neq k} (g_{kj})^{N_{kj}}.$$

Note that we denoted by τ_k and N_{kj} respectively represent the length of time X stays in state k and the number of transitions from state k to j,

$$T_k = \int_0^t \mathbf{1}_{\{X_s = k\}} ds \quad \text{and} \quad N_{kj} = \sum_{0 < s \leq t} \mathbf{1}_{\{X_{s-} = k, X_s = j\}}.$$

Distributional properties of the mixture process: cont'd

Proposition (Long-term composition)

Let $\mathbf{T}$ *(*$\mathbf{\Psi T}$*) have distinct eigenvalues* $\varphi_j^{(1)}$ $(\varphi_j^{(2)})$, $j = 1, ..., m$, *with* $\varphi_{p_k}^{(k)} = \max\{\varphi_j^{(k)}, j = 1, ..., m\}$ *and* $p_k = argmax_j\{\varphi_j^{(k)}\}$, $k = 1, 2$. *Then,*

$$\lim_{t\to\infty} s_i(t) = \begin{cases} \begin{cases} 1, & \text{if } \varphi_{p_2}^{(2)} > \varphi_{p_1}^{(1)} \\ 0, & \text{if } \varphi_{p_2}^{(2)} < \varphi_{p_1}^{(1)} \\ \dfrac{\boldsymbol{\pi}^\top \mathbf{S}_m L_{p_2}(\mathbf{\Psi T})\mathbf{e}_i}{\boldsymbol{\pi}^\top \Big(\mathbf{S}_m L_{p_2}(\mathbf{\Psi T}) + (\mathbf{I}_m - \mathbf{S}_m) L_{p_1}(\mathbf{T})\Big)\mathbf{e}_i}, & \text{if } \varphi_{p_2}^{(2)} = \varphi_{p_1}^{(1)} \end{cases} & \text{for } i \in E \\ \begin{cases} \dfrac{\boldsymbol{\pi}^\top \mathbf{S}_m \mathbf{T}^{-1}\boldsymbol{\delta}}{\boldsymbol{\pi}^\top \mathbf{T}^{-1}\boldsymbol{\delta}}, \end{cases} & \text{for } i = m+1. \end{cases} \tag{18}$$

where $L_{p_1}(\mathbf{T})$ *and* $L_{p_2}(\mathbf{\Psi T})$ *are the Lagrange interpolation coefficients defined by*

$$L_{p_1}(\mathbf{T}) = \prod_{j=1, j\neq p_1}^{m} \left(\frac{\mathbf{T} - \varphi_j^{(1)}\mathbf{I}_m}{\varphi_{p_1}^{(1)} - \varphi_j^{(1)}}\right) \quad and \quad L_{p_2}(\mathbf{\Psi T}) = \prod_{j=1, j\neq p_2}^{m} \left(\frac{\mathbf{\Psi T} - \varphi_j^{(2)}\mathbf{I}_m}{\varphi_{p_2}^{(2)} - \varphi_j^{(2)}}\right) \tag{19}$$

Remark (15)

Recall that under the diagonal assumption (13) on $\mathbf{\Psi}$*, we have following the previous Proposition that when the mixture process* X *ageing*

$$\lim_{t\to\infty} \mathbf{S}_m(t) = \begin{cases} \mathbf{I}_{m\times m}, & when\ \psi_i < 1,\ for\ i = 1, ..., m \\ \mathbf{0}_{m\times m}, & when\ \psi_i > 1,\ for\ i = 1, ..., m \end{cases} \quad (20)$$

From this observation we can see that ***in the long-run (prior moving to absorbing state)*** X ***tends to move at slow speed, and is Markovian****.*

We will see this more following the next theorem and proposition.

Distributional properties of the mixture process: cont'd

Below we present $\mathcal{F}_t-$conditional transition matrix of X. Define

$$\mathbf{S}(t) = \mathrm{diag}(s_1(t), ..., s_{m+1}(t))$$

whose $i-$th element is given by the quantity $s_i(t) = \mathbb{P}\{\phi = 1|\mathcal{F}_{i,t}\}$ (14).

Theorem (Transition Probability Matrix)

For a given $t \geq 0$ and any state $i, i_0 \in \mathbb{S}$, the $\mathcal{F}_t-$conditional $(t, s)-$transition probability matrix of X is given by

$$\mathbf{P}(t, s) = \mathbf{S}(t)e^{\mathbf{G}(s-t)} + (\mathbf{I} - \mathbf{S}(t))e^{\mathbf{Q}(s-t)}. \quad (21)$$

Corollary

*Set $\mathbf{\Psi} = \mathbf{I}$ in (21), $\mathbf{P}(t, s) = e^{\mathbf{Q}(s-t)}$, i.e., X **is a simple Markov chain**.*

Joint distribution of waiting time and the jump

The distribution can be used **to simulate sample paths of the mixture process**. For this purpose, define $\tau_0 = 0$ and the consecutive stopping times $\{\tau_k : k = 1, ...\}$ at which the mixture process makes a new jump,

$$\tau_k = \inf\{t > \tau_{k-1} : X_t \neq X_{\tau_{k-1}}, X_{t-} = X_{\tau_{k-1}}\}. \tag{22}$$

Proposition

Let $\{\tau_k\}$ be the stopping times (22). Then, for $x \in E$, $y \in \mathbb{S}$ with $y \neq x$,

$$\begin{aligned}\mathbb{P}\{\tau_k - \tau_{k-1} \leq t, X_{\tau_k} = y \big| X_{\tau_{k-1}} = x, \mathcal{F}_{\tau_{k-1}-}\} \\ = (1 - e^{-g_x t})\frac{g_{xy}}{g_x} s_x(\tau_{k-1}) + (1 - e^{-q_x t})\frac{q_{xy}}{q_x}(1 - s_x(\tau_{k-1})),\end{aligned} \tag{23}$$

where the switching probability $s_x(t)$ is described by eqns. (15)-(17).

Dependence of waiting time and the jump

Corollary

$$\mathbb{P}\{\tau_k - \tau_{k-1} \leq t \big| X_{\tau_{k-1}} = x, \mathcal{F}_{\tau_{k-1}-}\}$$
$$= \big(1 - e^{-g_x t}\big) s_x(\tau_{k-1}) + \big(1 - e^{-q_x t}\big)\big(1 - s_x(\tau_{k-1})\big)$$
$$\mathbb{P}\{X_{\tau_k} = y \big| X_{\tau_{k-1}} = x, \mathcal{F}_{\tau_{k-1}-}\}$$
$$= \frac{g_{xy}}{g_x} s_x(\tau_{k-1}) + \frac{q_{xy}}{q_x}\big(1 - s_x(\tau_{k-1})\big)$$

*Therefore, **unless** $\mathbf{\Psi} = \mathbf{I}$, conditional on the information $\mathcal{F}_{x,\tau_{k-1}}$,*

$$(\tau_k - \tau_{k-1}) \not\perp\!\!\!\perp X_{\tau_k}.$$

Remark

*For **finite mixtures**, the waiting time has hyperexponential distribution.*

Simulation of sample paths

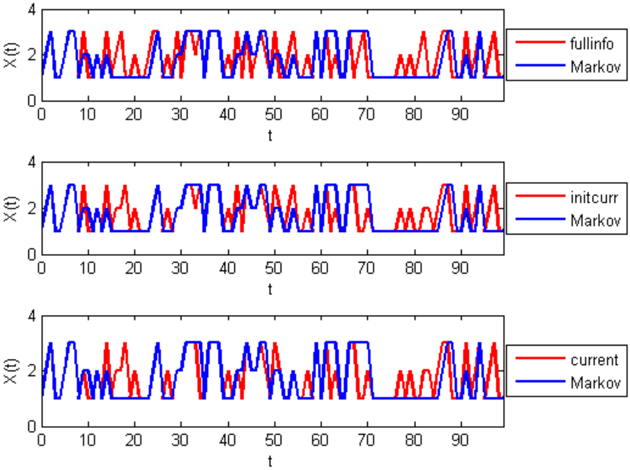


Figure: Sample paths of Markov chains along with Markov mixture process.

Partition of transition probability matrix

Proposition

The conditional transition probability matrix (21) has the partition

$$\mathbf{P}(t,s) = \begin{pmatrix} \mathbf{P}_{11}(t,s) & \mathbf{P}_{12}(t,s) \\ \mathbf{0} & 1 \end{pmatrix}, \tag{24}$$

where the block matrices $\mathbf{P}_{11}(t,s)$ *and* $\mathbf{P}_{12}(t,s)$ *are defined by*

$$\begin{aligned} \mathbf{P}_{11}(t,s) &= \mathbf{S}_m(t)e^{\mathbf{\Psi}\mathbf{T}(s-t)} + (\mathbf{I}_m - \mathbf{S}_m(t))e^{\mathbf{T}(s-t)}, \\ \mathbf{P}_{12}(t,s) &= \mathbf{S}_m(t)(\mathbf{\Psi}\mathbf{T})^{-1}(e^{\mathbf{\Psi}\mathbf{T}(s-t)} - \mathbf{I}_m)\mathbf{\Psi}\boldsymbol{\delta} \\ &\quad + (\mathbf{I}_m - \mathbf{S}_m(t))\mathbf{T}^{-1}(e^{\mathbf{T}(s-t)} - \mathbf{I}_m)\boldsymbol{\delta}. \end{aligned}$$

Long-term behavior under partial information

Corollary (Long-term behavior)

By Remark (16) and eqn. (24), we have under (13) that as $t \uparrow \infty$

$$\mathbf{P}(t,s) = \begin{cases} \begin{pmatrix} e^{\boldsymbol{\Psi}\mathbf{T}(s-t)} & (\boldsymbol{\Psi}\mathbf{T})^{-1}\big(e^{\boldsymbol{\Psi}\mathbf{T}(s-t)} - \mathbf{I}\big)\boldsymbol{\Psi}\boldsymbol{\delta} \\ \mathbf{0} & 1 \end{pmatrix}, & \psi_i < 1, i \in E \\ \\ \begin{pmatrix} e^{\mathbf{T}(s-t)} & \mathbf{T}^{-1}\big(e^{\mathbf{T}(s-t)} - \mathbf{I}\big)\boldsymbol{\delta} \\ \mathbf{0} & 1 \end{pmatrix}, & \psi_i > 1, i \in E \end{cases}$$

Conclusion: Under partial information, as the mixture process aging,

- it tends to move at lower speed
- behave Markovianly, see eqn. (6).

Lifetime distributions of Markov mixture process

Default time τ is defined as time until absorption of X to Δ, i.e.,

$$\tau = \inf\{t > 0 : X_t = \Delta\}.$$

Theorem below states the $\mathcal{F}_{it}-$conditional distribution and density of τ.

Theorem (Conditional Lifetime Distribution)

Given $\mathcal{F}_{it}$ and $i \in E$, we have for all $s, t \geq 0$ that

$$\mathbb{P}\{\tau > t + s | \mathcal{F}_{it}\} = \mathbf{e}_i^\top \big(\mathbf{S}_m(t) e^{\mathbf{\Psi}\mathbf{T}s} + (\mathbf{I} - \mathbf{S}_m(t)) e^{\mathbf{T}s}\big)\mathbf{1}$$

$$(ds)^{-1}\mathbb{P}\{t + s < \tau \leq t + s + ds | \mathcal{F}_{it}\} = \mathbf{e}_i^\top \big(\mathbf{S}_m(t) e^{\mathbf{\Psi}\mathbf{T}s}\mathbf{\Psi} + (\mathbf{I} - \mathbf{S}_m(t)) e^{\mathbf{T}s}\big)\delta$$

Unconditional lifetime distribution

Theorem (Generalized Mixture of Phase-Type Distributions)

The unconditional distribution is given for any $t \geq 0$ by

$$\begin{aligned} \overline{F}(t) &= \boldsymbol{\pi}^\top \Big(\mathbf{S}_m e^{\boldsymbol{\Psi}\mathbf{T}t} + (\mathbf{I}_m - \mathbf{S}_m) e^{\mathbf{T}t} \Big) \mathbf{1}_m, \\ f(t) &= \boldsymbol{\pi}^\top \Big(\mathbf{S}_m e^{\boldsymbol{\Psi}\mathbf{T}t} \boldsymbol{\Psi} + (\mathbf{I}_m - \mathbf{S}_m) e^{\mathbf{T}t} \Big) \boldsymbol{\delta}. \end{aligned} \tag{25}$$

Corollary (Phase-Type Distribution, Neuts [27])

Set $\boldsymbol{\Psi} = \mathbf{I}$ in (25), then $\overline{F}(t) = \boldsymbol{\pi}^\top e^{\mathbf{T}t} \mathbf{1}_m$ and $f(t) = \boldsymbol{\pi}^\top e^{\mathbf{T}t} \boldsymbol{\delta}$.

Example (Convex mixture of PHD)

Set $\mathbf{S}_m = \alpha \mathbf{I}_m$, with $0 \leq \alpha \leq 1$, in (25). Then,

$$\overline{F}(t) = \alpha \boldsymbol{\pi}^\top e^{\boldsymbol{\Psi}\mathbf{T}t} \mathbf{1}_m + (1-\alpha) \boldsymbol{\pi}^\top e^{\mathbf{T}t} \mathbf{1}_m.$$

Unconditional lifetime distribution

Below are the Laplace transform of (25) and its nth moment.

Theorem

Let $\overline{F}$ be the phase-type distribution $\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m)$. Then,

(i) *the Laplace transform $\widehat{F}[\theta] = \int_0^\infty e^{-\theta u} f(u) du$ is given by*

$$\widehat{F}[\theta] = \boldsymbol{\pi}^\top \Big(\mathbf{S}_m (\theta \mathbf{I}_m - \boldsymbol{\Psi}\mathbf{T})^{-1} \boldsymbol{\Psi} + (\mathbf{I}_m - \mathbf{S}_m)(\theta \mathbf{I}_m - \mathbf{T})^{-1} \Big) \boldsymbol{\delta}. \quad (26)$$

(ii) *the nth moment, for $n = 0, 1, ...,$ of τ (8) is given by*

$$\mathbb{E}\{\tau^n\} = (-1)^n n! \boldsymbol{\pi}^\top \Big(\mathbf{S}_m (\boldsymbol{\Psi}\mathbf{T})^{-n} + (\mathbf{I}_m - \mathbf{S}_m) \mathbf{T}^{-n} \Big) \mathbf{1}_m. \quad (27)$$

Erlang distribution

Example (Erlang distribution)

Let F be $\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m)$ *distribution with* $\boldsymbol{\Psi} = \mathbf{I} = \mathbf{S}_m$, *and*

$$\boldsymbol{\pi} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{T} = \begin{pmatrix} -\beta & \beta & 0 \\ 0 & -\beta & \beta \\ 0 & 0 & -\beta \end{pmatrix},$$

with $\beta > 0$. *Then,* $F := F(t; m, \beta)$, $m = 3$, *has the following form:*

$$F(t; m, \beta) = 1 - \sum_{k=0}^{m-1} \frac{(\beta t)^k}{k!} e^{-\beta t}. \tag{28}$$

Mixture of Erlang distributions

Example

Let F be $\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m)$ *distribution with* $\boldsymbol{\pi} = (1,0,0)^\top$,

$$\mathbf{T} = \begin{pmatrix} -\beta_2 & \beta_2 & 0 \\ 0 & -\beta_2 & \beta_2 \\ 0 & 0 & -\beta_2 \end{pmatrix}, \; \boldsymbol{\Psi} = \begin{pmatrix} \beta_1/\beta_2 & 0 & 0 \\ 0 & \beta_1/\beta_2 & 0 \\ 0 & 0 & \beta_1/\beta_2 \end{pmatrix}, \; \mathbf{S}_m = \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \alpha & 0 \\ 0 & 0 & \alpha \end{pmatrix}.$$

with $0 \leq \alpha \leq 1$ *and* $\beta_1, \beta_2 > 0$. *Then, following (25)* F *is given by*

$$F(t) = \alpha F(t; m, \beta_1) + (1-\alpha) F(t; m, \beta_2).$$

where $F(t; m, \beta_i)$, *for* $i = 1, 2$ *and* $m = 3$, *is given by (28).*

Equivalence with PH Distribution

Then, $\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m)$ can be represented by $\mathrm{PH}(\widetilde{\boldsymbol{\pi}}, \widetilde{\mathbf{T}})$

$$\widetilde{\boldsymbol{\pi}} = \begin{pmatrix} \alpha \\ 0 \\ 0 \\ 1-\alpha \\ 0 \\ 0 \end{pmatrix}, \quad \widetilde{\mathbf{T}} = \begin{pmatrix} -\beta_1 & \beta_1 & 0 & 0 & 0 & 0 \\ 0 & -\beta_1 & \beta_1 & 0 & 0 & 0 \\ 0 & 0 & -\beta_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -\beta_2 & \beta_2 & 0 \\ 0 & 0 & 0 & 0 & -\beta_2 & \beta_2 \\ 0 & 0 & 0 & 0 & 0 & -\beta_2 \end{pmatrix},$$

Lemma (PH Matrix Representation)

The distribution $\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m)$ *has* $\mathrm{PH}(\widetilde{\boldsymbol{\pi}}, \widetilde{\mathbf{T}})$ *matrix representation,*

$$\widetilde{\boldsymbol{\pi}}^\top = \left((\mathbf{S}_m\boldsymbol{\pi})^\top, ((\mathbf{I}_m - \mathbf{S}_m)\boldsymbol{\pi})^\top\right) \quad \text{and} \quad \widetilde{\mathbf{T}} = \begin{pmatrix} \boldsymbol{\Psi}\mathbf{T} & \mathbf{0} \\ \mathbf{0} & \mathbf{T} \end{pmatrix}.$$

Dense and closure property of GPH

Closure property of lifetime distribution

Theorem (Closure property)

The lifetime distribution $\mathrm{GPH}(\boldsymbol{\pi}, \mathbf{T}, \boldsymbol{\Psi}, \mathbf{S}_m)$ *possesses closure property under finite convex mixtures and finite convolutions.*

Dense property of lifetime distribution

Theorem (Dense property)

The lifetime distribution forms a dense class of distributions on $\mathbb{R}_+$.

Forward intensity representation of the distribution

Motivated by Duan et al. [15], consider so-called **forward intensity**

$$\lambda_{it}(s)ds = \mathbb{P}\{t+s<\tau\leq t+s+ds \big| \tau > t+s, \mathcal{F}_{it}\}, \quad s,t\geq 0. \tag{29}$$

Theorem (Forward Intensity)

For a given $i\in E$ and fixed $t\geq 0$, we have for all $s\geq 0$,

$$\lambda_{it}(s) = \frac{\mathbf{e}_i^\top\big(\mathbf{S}_m(t)e^{\mathbf{\Psi}\mathbf{T}s}\mathbf{\Psi} + (\mathbf{I}_m - \mathbf{S}_m(t))e^{\mathbf{T}s}\big)\boldsymbol{\delta}}{\mathbf{e}_i^\top\big(\mathbf{S}_m(t)e^{\mathbf{\Psi}\mathbf{T}s} + (\mathbf{I}_m - \mathbf{S}_m(t))e^{\mathbf{T}s}\big)\mathbf{1}_m}. \tag{30}$$

Forward intensity representation: cont'd

Theorem (Cont'd: Forward Intensity Representation)

The $\mathcal{F}_{it}$−conditional probability of survival to time $t+s$ is given by

$$\mathbb{P}\{\tau > t+s \big| \mathcal{F}_{it}\} = \exp\Big(-\int_0^s \lambda_{it}(u)du\Big), \quad i \in E, s, t \geq 0 \tag{31}$$

whereas the $\mathcal{F}_{it}$−conditional probability of default in $[t, t+s]$ is given by

$$\mathbb{P}\{t < \tau \leq t+s \big| \mathcal{F}_{it}\} = \int_0^s \exp\Big(-\int_0^u \lambda_{it}(\nu)d\nu\Big)\lambda_{it}(u)du. \tag{32}$$

Long-term intensity

Proposition

Let $\mathbf{T}$ $(\mathbf{\Psi T})$ *have distinct eigenvalues* $\varphi_j^{(1)}$ $(\varphi_j^{(2)})$, $j = 1, ..., m$ *with* $\varphi_{p_k}^{(k)} = \max\{\varphi_1^{(k)}, ..., \varphi_m^{(k)}\}$ *and* $p_k = argmax_j\{\varphi_j^{(k)}\}$, $k = 1, 2$. *Then, for any* $t \geq 0$ *and* $i \in E$,

$$\lim_{s\to\infty} \lambda_i(t,s) = \begin{cases} \dfrac{\mathbf{e}_i^\top \mathbf{S}_m(t) L_{p_2}(\mathbf{\Psi T})\mathbf{\Psi}\delta}{\mathbf{e}_i^\top \mathbf{S}_m(t) L_{p_2}(\mathbf{\Psi T})\mathbf{1}_m}, & \varphi_{p_2}^{(2)} > \varphi_{p_1}^{(1)} \\[2ex] \dfrac{\mathbf{e}_i^\top (\mathbf{I}_m - \mathbf{S}_m(t)) L_{p_1}(\mathbf{T})\delta}{\mathbf{e}_i^\top (\mathbf{I}_m - \mathbf{S}_m(t)) L_{p_1}(\mathbf{T})\mathbf{1}_m}, & \varphi_{p_2}^{(2)} < \varphi_{p_1}^{(1)} \\[2ex] \dfrac{\mathbf{e}_i^\top \Big(\mathbf{S}_m(t) L_{p_2}(\mathbf{\Psi T})\mathbf{\Psi} + (\mathbf{I}_m - \mathbf{S}_m(t)) L_{p_1}(\mathbf{T})\Big)\delta}{\mathbf{e}_i^\top \Big(\mathbf{S}_m(t) L_{p_2}(\mathbf{\Psi T}) + (\mathbf{I}_m - \mathbf{S}_m(t)) L_{p_1}(\mathbf{T})\Big)\mathbf{1}_m}, & \varphi_{p_2}^{(2)} = \varphi_{p_1}^{(1)} \end{cases} \quad (33)$$

where $L_{p_1}(\mathbf{T})$ *and* $L_{p_2}(\mathbf{\Psi T})$ *are Lagrange interpolation coefficients (19).*

Instantaneous default intensity

However, when we set $s = 0$ in eqn. (29), we obtain default intensity $\lambda_i(t)dt = \mathbb{P}\{t < \tau \leq t + dt | \tau > t, \mathcal{F}_{it}\}$ for event occurrence in $(t, t + dt]$.

Corollary

Given $i \in E$ and past information $\mathcal{F}_{i,t}$, we have for any $t \geq 0$,

$$\lambda_i(t) = \mathbf{e}_i^\top (\mathbf{I}_m + \mathbf{S}_m(t)(\mathbf{\Psi} - \mathbf{I}_m))\delta. \tag{34}$$

Remark

By setting $\mathbf{\Psi} = \mathbf{I}_m$ in (34), we have for each $i \in E$ $\lambda_i(t) = \delta_i$ $\forall t \geq 0$.

Residual occupation time

Define state occupation time of X as $T_j(t) = \int_0^t \mathbf{1}_{\{X_s=j\}} ds$. Define

$$\overline{T}_j(t,s) = \mathbb{E}\{ T_j(s) - T_j(t) \big| \mathcal{F}_{i,t} \}. \tag{35}$$

Theorem

For a given $s \geq t \geq 0$ and past information $\mathcal{F}_{i,t}$, we have

$$\overline{T}_j(t,s) = \begin{cases} \mathbf{e}_i^\top \overline{\mathbf{F}}_{11}(t,s)\mathbf{e}_j, & \forall i,j \in E \\ \mathbf{e}_i^\top \overline{\mathbf{F}}_{12}(t,s)\mathbf{e}_{m+1}, & \forall i \in E \\ 0, & \forall j \in E, i = m+1 \\ (s-t), & \textit{for } i,j = m+1, \end{cases} \tag{36}$$

Residual occupation time: cont'd

Theorem (Cont'd)

where the matrices $\overline{\mathbf{F}}_{11}(t,s)$ *and* $\overline{\mathbf{F}}_{12}(t,s)$ *are defined by*

$$\begin{aligned}
\overline{\mathbf{F}}_{11}(t,s) &= \mathbf{S}_m(t)\big[\mathbf{\Psi}\mathbf{T}\big]^{-1}\big(e^{\mathbf{\Psi}\mathbf{T}(s-t)} - \mathbf{I}_m\big) \\
&\quad + (\mathbf{I}_m - \mathbf{S}_m(t))\big[\mathbf{T}\big]^{-1}\big(e^{\mathbf{T}(s-t)} - \mathbf{I}_m\big), \\
\overline{\mathbf{F}}_{12}(t,s) &= \mathbf{S}_m(t)\Big[(s-t)\mathbf{I}_m - [\mathbf{\Psi}\mathbf{T}]^{-1}\big(e^{\mathbf{\Psi}\mathbf{T}(s-t)} - \mathbf{I}_m\big)\Big]\mathbf{1}_m \\
&\quad + (\mathbf{I}_m - \mathbf{S}_m(t))\Big[(s-t)\mathbf{I}_m - [\mathbf{T}]^{-1}\big(e^{\mathbf{T}(s-t)} - \mathbf{I}_m\big)\Big]\mathbf{1}_m.
\end{aligned}$$

Residual lifetime of the mixture process

Corollary

For a given $t \geq 0$ and past information $\mathcal{F}_{i,t}$, we have

$$R_i(t) = -\mathbf{e}_i^\top \Big(\mathbf{S}_m(t)(\boldsymbol{\Psi}\mathbf{T})^{-1} + (\mathbf{I}_m - \mathbf{S}_m(t))\mathbf{T}^{-1} \Big) \mathbf{1}_m, \qquad (37)$$

for any initial state $i \in E$ of X at time t and is zero for $i = m + 1$.

Remark

Set $\boldsymbol{\Psi} = \mathbf{I}_m$ in (37), for each $i \in E$, $R_i(t) = -\mathbf{e}_i^\top \mathbf{T}^{-1} \mathbf{1}_m \;\; \forall t \geq 0$.

Outline

Example in survival analysis

Example (Married & Divorced Model)

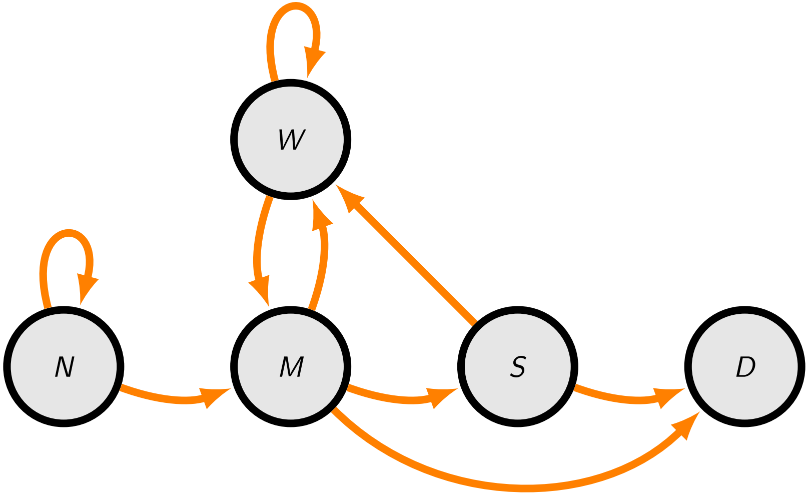
Let the intensity matrix $\mathbf{Q}$ *be given by the following*

$$\mathbf{Q} = \begin{pmatrix} -0.95 & 0.95 & 0 & 0 & 0 \\ 0 & -0.37 & 0.25 & 0.05 & 0.07 \\ 0 & 0 & -0.6 & 0.1 & 0.5 \\ 0 & 0.85 & 0 & -0.85 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \qquad (38)$$

The values of parameters $\boldsymbol{\pi}$, $\boldsymbol{\Psi}$ *and* $\mathbf{S}_m$ *are given respectively by*

$$\begin{aligned} \boldsymbol{\pi} &= (0.5, 0.3, 0.1)^\top \\ \boldsymbol{\Psi} &= diag(0.25, 0.25, 0.25) \\ \mathbf{S}_m &= diag(0.5, 0.5, 0.5, 0.5, 0.5) \end{aligned}$$

State Diagram of the Married-Divorced Model



Numerical result based on limited information

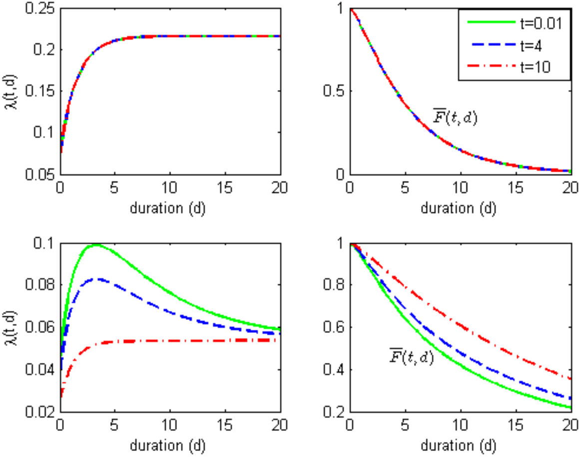


Figure: Divorced intensity $\lambda_{it}(s)$ and distribution $\overline{F}_{it}(s)$ for different cohort t.

Empirical facts: Aalen et al. (2008)

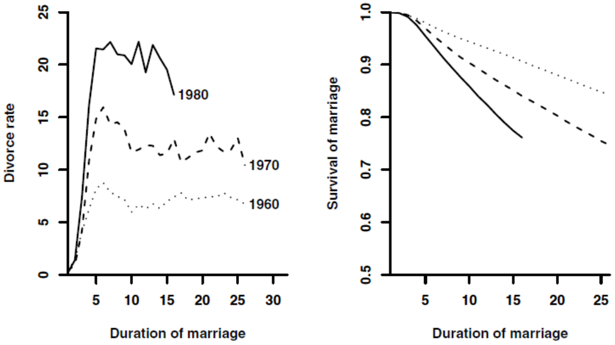


Fig. 5.3 *Rates of divorce per 1000 marriages per year (left panel) and empirical survival curves (right panel) for marriages contracted in 1960, 1970, and 1980.*

Figure: Rates of divorce per 1000 marriages per year (left panel) and empirical survival curves (right panel) for marriages contracted in 1960, 1970, and 1980. (Source: Figure 5.3 on page 222 of Aalen, Borgan, Gjessing (2008))

Numerical result: cont'd

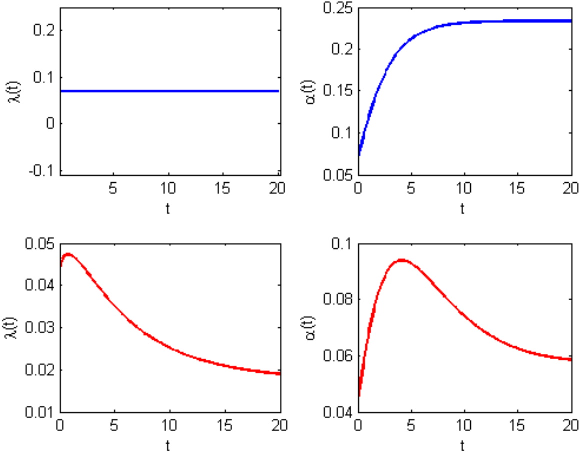


Figure: Intensity $\lambda_i(t)$ (34) and baseline $\alpha(t)dt = \mathbb{P}\{t < \tau \leq t + dt | \tau > t\}$.

Residual marriage lifespan

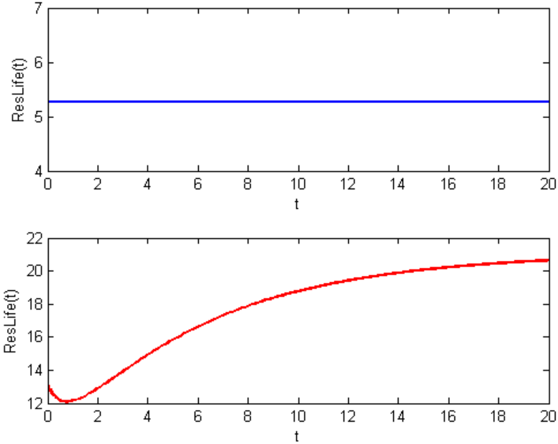


Figure: Residual marriage lifetime against (marriage) age t

U-bend shape of lifetime

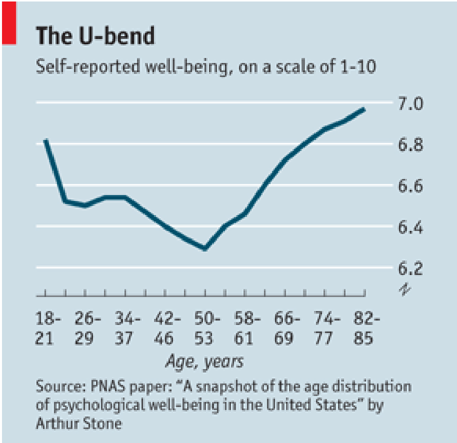


Figure: Age distribution of psychological well-being in the US. Source: *The Economist*, December 16th, 2010. http://www.economist.com/node/17722567

Thank You!