Reliability/survival analysis of semi-Markov systems: modeling and estimation

Vlad Stefan Barbu

Université de Rouen, France
barbu@univ-rouen.fr

IRMA, Université de Strasbourg

15 mai 2012

Plan

Infinite matrices

Semi-Markov kernel and convolution

Markov renewal equation

Applications in reliability/survival analysis

Nonparametric estimation

This talk :

- semi-Markov processes in discrete time
- SM processes - important generalization with respect to Markov processes
- the state space - infinitely countable
- need of considering infinite matrices
- some domains of application: reliability, survival analysis

Plan

Infinite matrices:

- rather few references and works on the topic
- Cooke (1955), Kemeny, Snell, and Knapp (1976)

Consider : a random system with countable state space, $E = \{1, \ldots, s\}$, $s < \infty$, or $E = \mathbb{N} - \{0\}$

Let us denote by :

- $\mathcal{M}_E$ - set of real matrices on $E \times E$
- $\mathcal{M}_E^b = \{\mathbf{A} = (A_{ij})_{i,j \in E} \in \mathcal{M}_E \mid \exists\, M, 0 < M < \infty, \text{ such that } \mid A_{ij} \mid \leq M, i, j \in E\}$
- $\mathcal{M}_E^{sub} = \{\sum_{i=1}^{n} \lambda_i \mathbf{A}_i \mid \mathbf{A}_1, \ldots, \mathbf{A}_n \text{ substochastic matrices}, \lambda_1, \ldots, \lambda_n \in \mathbb{R}, n \in \mathbb{N}\}$.

Infinite matrices:

- rather few references and works on the topic
- Cooke (1955), Kemeny, Snell, and Knapp (1976)

Consider : a random system with countable state space, $E = \{1, \ldots, s\}$, $s < \infty$, or $E = \mathbb{N} - \{0\}$

Let us denote by :

- $\mathcal{M}_E$ - set of real matrices on $E \times E$
- $\mathcal{M}_E^b = \{\mathbf{A} = (A_{ij})_{i,j \in E} \in \mathcal{M}_E \mid \exists\, M, 0 < M < \infty, \text{ such that } \mid A_{ij} \mid \leq M, i, j \in E\}$
- $\mathcal{M}_E^{sub} = \{\sum_{i=1}^{n} \lambda_i \mathbf{A}_i \mid \mathbf{A}_1, \ldots, \mathbf{A}_n \text{ substochastic matrices}, \lambda_1, \ldots, \lambda_n \in \mathbb{R}, n \in \mathbb{N}\}.$

Operations with infinite matrices :

1. addition - well defined ; denote by $\mathbf{0}_E \in \mathcal{M}_E$ the zero matrix;
2. multiplication of a matrix by a real number - well defined;
3. for $\mathbf{A}$ and $\mathbf{B} \in \mathcal{M}_E$, the product matrix $\mathbf{AB}$ is defined as usual if $\sum_{k \in E} A_{ik} B_{kj}$ is well defined and finite for all $i, j \in E$.

Remarks

- Obviously, if $\mathbf{A}, \mathbf{B} \in \mathcal{M}_E$ have arbitrary real entries, their product is not always well defined.
- The product of two bounded matrices is not always well defined.
- For $\mathbf{A}$ and $\mathbf{B} \in \mathcal{M}_E^{sub}$, the product $\mathbf{AB}$ is in $\mathcal{M}_E^{sub}$, so it is well defined, with finite entries.
- If $\mathbf{A} \in \mathcal{M}_E^{sub}$ and $\mathbf{B} \in \mathcal{M}_E^{b}$, then the product $\mathbf{AB}$ is well defined, with finite entries.
- The problem : the associativity of matrix multiplication does not always hold for infinite matrices.
- Note also that the uniqueness of the inverse rests upon associativity.

Remarks

- Obviously, if $\mathbf{A}, \mathbf{B} \in \mathcal{M}_E$ have arbitrary real entries, their product is not always well defined.
- The product of two bounded matrices is not always well defined.
- For $\mathbf{A}$ and $\mathbf{B} \in \mathcal{M}_E^{sub}$, the product $\mathbf{AB}$ is in $\mathcal{M}_E^{sub}$, so it is well defined, with finite entries.
- If $\mathbf{A} \in \mathcal{M}_E^{sub}$ and $\mathbf{B} \in \mathcal{M}_E^{b}$, then the product $\mathbf{AB}$ is well defined, with finite entries.
- The problem : the associativity of matrix multiplication does not always hold for infinite matrices.
- Note also that the uniqueness of the inverse rests upon associativity.

Lemma (Kemeny et al., 1976)

1. *Nonnegative matrices associate under multiplication.*
2. *Matrices associate if the product of their absolute values has only finite entries.*

Plan

Infinite matrices

Semi-Markov kernel and convolution

Markov renewal equation

Applications in reliability/survival analysis

Nonparametric estimation

- $Z = (Z_k)_{k \in \mathbb{N}}$, chain with state space $E = \{1, 2, \ldots, s\}$
- $S = (S_n)_{n \in \mathbb{N}}$, jump times
  $J = (J_n)_{n \in \mathbb{N}}$, visited states
  $X = (X_n)_{n \in \mathbb{N}}$, sojourn times of Z

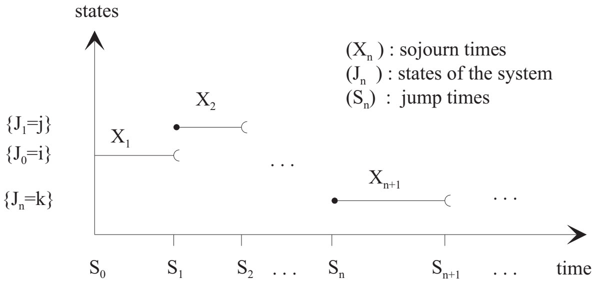

Figure: A sample path of a semi-Markov chain

If (J,S) verifies:

$$
\begin{aligned}
& \mathbb{P}(J_{n+1}=j, S_{n+1}-S_n=k \mid J_0,\cdots,J_n;S_1,\cdots,S_n) \\
= \ & \mathbb{P}(J_{n+1}=j, S_{n+1}-S_n=k \mid J_n), j\in E, k\in\mathbb{N}
\end{aligned}
$$

- (J,S) Markov renewal chain (MRC)
- $Z=(Z_k)_{k\in\mathbb{N}}-$ semi-Markov chain (SMC) associated to the MRC (J,S)

If (J, S) verifies:

$$\begin{aligned} & \mathbb{P}(J_{n+1} = j, S_{n+1} - S_n = k | J_0, \cdots, J_n; S_1, \cdots, S_n) \\ = \ & \mathbb{P}(J_{n+1} = j, S_{n+1} - S_n = k | J_n), j \in E, k \in \mathbb{N} \end{aligned}$$

- (J, S) Markov renewal chain (MRC)
- $Z = (Z_k)_{k \in \mathbb{N}} -$ semi-Markov chain (SMC) associated to the MRC (J, S)

$$Z_k := J_{N(k)} \quad \Leftrightarrow \quad J_n = Z_{S_n}$$

with $N(k) := \max\{n \in \mathbb{N} \mid S_n \leq k\}, \ k, n \in \mathbb{N}$

Remark : $J = (J_n)_{n \in \mathbb{N}}$ is a Markov chain, called the embedded Markov chain (EMC).

Notation/Definitions

the initial distribution $\alpha(i) := \mathbb{P}(J_0 = i)$
the homogeneous SM kernel $\mathbf{q} = \big(q_{ij}(\cdot)\big)_{i,j \in E}$

$$q_{ij}(k) := \begin{cases} \mathbb{P}(J_{n+1} = j, X_{n+1} = k \mid J_n = i), & k \in \mathbb{N}^* \\ 0, & k = 0 \end{cases}$$

Notation/Definitions

the initial distribution $\alpha(i) := \mathbb{P}(J_0 = i)$
the homogeneous SM kernel $\mathbf{q} = \left(q_{ij}(\cdot)\right)_{i,j\in E}$

$$q_{ij}(k) := \begin{cases} \mathbb{P}(J_{n+1} = j, X_{n+1} = k \mid J_n = i), & k \in \mathbb{N}^* \\ 0, & k = 0 \end{cases}$$

the conditional sojourn time distributions $\mathbf{f} = \left(f_{ij}(\cdot)\right)_{i,j\in E}$

$$f_{ij}(k) := \mathbb{P}(X_{n+1} = k \mid J_n = i, J_{n+1} = j),\ f_{ij}(0) := 0$$

the transition matrix of the MC $(J_n)_{n\in\mathbb{N}}$, $\mathbf{p} = (p_{ij})_{i,j\in E}$

$$p_{ij} := \mathbb{P}(J_{n+1} = j \mid J_n = i),\ p_{ii} := 0$$

Note that $q_{ij}(k) = p_{ij}\, f_{ij}(k)$

- $\mathcal{M}_E(\mathbb{N}) := \{f : \mathbb{N} \to \mathcal{M}_E\}$
- $\mathcal{M}_E^b(\mathbb{N}) := \{f : \mathbb{N} \to \mathcal{M}_E^b\}$
- $\mathcal{M}_E^{sub}(\mathbb{N}) := \{f : \mathbb{N} \to \mathcal{M}_E^{sub}\}$

Operations with $\mathbf{A}, \mathbf{B} \in \mathcal{M}_E(\mathbb{N})$:

- $\mathbf{A} * \mathbf{B}$ = the discrete-time matrix convolution product,

$$\mathbf{AB}(k) := \sum_{l=0}^{k} \mathbf{A}(k-l)\,\mathbf{B}(l), \quad k \in \mathbb{N},$$

  provided that all the matrix products $\mathbf{A}(k-l)\,\mathbf{B}(l)$, $k \in \mathbb{N}$, $l = 0, \ldots, k$, are well defined and all their entries are finite.
- $\delta\mathbf{I}$ = the identity element, defined by $\delta\mathbf{I}(0) := \mathbf{I}$ and $\mathbf{I}(k) := \mathbf{0}$ if $k \neq 0$
- $\mathbf{A}^{(n)}$ = the n-fold convolution of $\mathbf{A}$, provided that $\mathbf{A}$ is self-associative
- $\mathbf{A}^{(-1)}$ = the left convolution inverse of $\mathbf{A}$ (if it exists)

Plan

Markov renewal equation

Problem : solve equations of the type

$$\mathbf{L}(k) = \mathbf{G}(k) + \mathbf{q} * \mathbf{L}(k), \; k \in \mathbb{N}, \qquad (1)$$

with

- $G \in \mathcal{M}_E(\mathbb{N})$ known
- $L \in \mathcal{M}_E(\mathbb{N})$ unknown.

Theorem

Suppose that $\mathbf{L}$ and $\mathbf{G} \in \mathcal{M}_2(\mathbb{N})$, where $\mathcal{M}_2 = \{\mathbf{A} = (A_{ij})_{i,j \in E} \in \mathcal{M}_E \mid (A_{ij})_{i \in E} \in l^2 \text{ for all } j \in E, (A_{ij})_{j \in E} \in l^2 \text{ for all } i \in E\}$, $l^2 = \{(x_n)_{n \in \mathbb{N}} \mid x_n \in \mathbb{R}, \sum_{n \geq 0} x_n^2 < \infty\}$. Then, the Markov renewal equation (1) has the unique solution

$$\mathbf{L}(k) = (\delta I - q)^{(-1)} * G(k) = \left(\sum_{n=0}^{k} \mathbf{q}^{(n)} \right) * \mathbf{G}(k).$$

Theorem (asymptotic behavior)

Let $\mathbf{L}$ *and* $\mathbf{G} \in \mathcal{M}_2(\mathbb{N})$ *such that* $\sum_{i \in E} \sum_{n \in \mathbb{N}} \mid G_{ij}(n) \mid < \infty$.
Then:

$$L_{ij}(k) = (\psi * G)_{ij}(k) \xrightarrow[k \to \infty]{} \sum_{i \in E} \sum_{n \in \mathbb{N}} \frac{1}{\mu_{ii}} G_{ij}(n),$$

where μ_{ii} *is the mean recurrence time of state* i *for the semi-Markov chain.*

Theorem (asymptotic behavior)

Let $\mathbf{L}$ *and* $\mathbf{G} \in \mathcal{M}_2(\mathbb{N})$ *such that* $\sum_{i \in E} \sum_{n \in \mathbb{N}} \mid G_{ij}(n) \mid < \infty$.
Then:

$$L_{ij}(k) = (\psi * G)_{ij}(k) \xrightarrow[k \to \infty]{} \sum_{i \in E} \sum_{n \in \mathbb{N}} \frac{1}{\mu_{ii}} G_{ij}(n),$$

where μ_{ii} *is the mean recurrence time of state* i *for the semi-Markov chain.*

Let $\mathbf{P} = \left(P_{ij}(\cdot)\right)_{i,j \in E}$ be the semi-Markov transition function,

$$P_{ij}(k) := \mathbb{P}(Z_k = j \mid Z_0 = i), \; i,j \in E, \; k \in \mathbb{N}$$

Theorem

The semi-Markov transition function $\mathbf{P}$ *verifies*

$$\mathbf{P} = \mathbf{I} - \mathbf{H} + \mathbf{q} * \mathbf{P},$$

the unique solution is given by

$$\mathbf{P}(k) = (\delta\mathbf{I} - \mathbf{q})^{(-1)} * (\mathbf{I} - \mathbf{H})(k), \ k \in \mathbb{N},$$

and

$$\lim_{k\to\infty} P_{ij}(k) = \frac{1}{\mu_{jj}} m_j,$$

where

$H_i(n) := \mathbb{P}(X_1 \leq n \mid J_0 = i) = \sum_{l=1}^{n} \sum_{j\in E} q_{ij}(l)$

$\mathbf{H}(n) := diag(H_i(n); \ i \in E), i \in E, n \in \mathbb{N}$

m_i is the mean sojourn time in state i.

Plan

Framework

- Discrete-time semi-Markov system with state space E
- Partition $E = U \cup U^c$
- Suppose that the initial state of the chain belongs to U.
- Problems: compute/estimate the mean time needed to hit $U^c = D$, the mean up/down times, the reliability/survival function, the availability function, the failure rate, etc.

Framework

- Discrete-time semi-Markov system with state space E
- Partition $E = U \cup U^c$
- Suppose that the initial state of the chain belongs to U.
- Problems: compute/estimate the mean time needed to hit $U^c = D$, the mean up/down times, the reliability/survival function, the availability function, the failure rate, etc.
- Reorder E such that all the elements of U precede the elements of U^c.
- Partition every matrix or matrix-valued function according to the partition $\{U, U^c\}$.

$$\mathbf{q}(k) = \begin{array}{c} \\ \\ \end{array} \begin{array}{cc} U & U^c \\ \left(\begin{array}{c} \mathbf{q}_{11}(k) \\ \mathbf{q}_{21}(k) \end{array} \right. & \left. \begin{array}{c} \mathbf{q}_{12}(k) \\ \mathbf{q}_{22}(k) \end{array} \right) \end{array} \begin{array}{c} U \\ U^c \end{array}$$

Consider the equation

$$V_i = \begin{cases} m_i + (\mathbf{p}\mathbf{V})_i & \text{if } i \in U, \\ 0 & \text{if } i \in U^c, \end{cases} \tag{2}$$

where $\mathbf{V} = (V_i;\ i \in E) \in \mathbb{R}^E$ is an unknown column vector. In matrix form, this equation can be written

$$\mathbf{V}_1 = \mathbf{m}_1 + \mathbf{p}_{11}\mathbf{V}_1.$$

Theorem (Kemeny et al., 1976)

If the matrix $(\mathbf{I} - \mathbf{p}_{11})$ is invertible, then $\mathbf{N}\mathbf{m_1}$ is the minimal nonnegative solution of Equation (2), where $\mathbf{N} = \sum_{k \geq 0} \mathbf{p}_{11}^k$.

Mean Time To Failure

Let T_D denote the lifetime of the system. We want to compute $MTTF = \mathbb{E}[T_D]$.

For any state $i \in U$, we introduce:

- $MTTF_i := \mathbb{E}_i[T_D]$ - the MTTF of the system, given that it starts in state $i \in U$;
- $\mathbf{MTTF} := (MTTF_i; i \in U)^\top$

We can show that **MTTF** satisfies equation

$$\mathbf{MTTF} = \mathbf{m}_1 + \mathbf{p}_{11}\mathbf{MTTF}.$$

Theorem

If the matrix $(\mathbf{I} - \mathbf{p}_{11})$ is invertible, then the MTTF of the system is given by

$$MTTF = \alpha_1(\mathbf{I} - \mathbf{p}_{11})^{-1}\mathbf{m}_1.$$

Reliability/Survival function

$$
\begin{aligned}
T_D &:= \inf\{n \in \mathbb{N} \mid Z_n \in D\} \\
R(k) &:= \mathbb{P}(T_D > k) = \mathbb{P}(Z_n \in U, n \in \{0, \ldots, k\})
\end{aligned}
$$

Reliability/Survival function

$$T_D := \inf\{n \in \mathbb{N} \mid Z_n \in D\}$$
$$R(k) := \mathbb{P}(T_D > k) = \mathbb{P}(Z_n \in U, n \in \{0, \dots, k\})$$

Theorem

The reliability of a discrete-time SM system is given by

$$R(k) = \alpha_1 \, \mathbf{P}_{11}(k) \, \mathbf{1}_U = \alpha_1 \, \psi_{11} * (\mathbf{I} - \mathbf{H}_{11})(k) \, \mathbf{1}_U.$$

Reliability estimator

$$\begin{aligned} \widehat{R}(k, M) &:= \widehat{\alpha}_1 \, \widehat{\mathbf{P}}_{11}(k, M) \, \mathbf{1}_U \\ &= \widehat{\alpha}_1 \left[\widehat{\psi}_{11}(\cdot, M) * \left(\mathbf{I} - \widehat{\mathbf{H}}(\cdot, M)_{11} \right) \right](k) \mathbf{1}_U \end{aligned}$$

Proof 1 : direct

- $Y = (Y_k)_{k \in \mathbb{N}}$ a new semi-Markov chain, of state space $E_Y = U \cup \{\Delta\}$, with Δ an absorbing state:

$$Y_k = \begin{cases} Z_k & \text{if } k < T_D, \\ \Delta & \text{if } k \geq T_D, \end{cases} \quad k \in \mathbb{N}.$$

- The SM chain Y : kernel $\mathbf{q}_Y$, transition matrix $\mathbf{P}_Y$

$$\mathbf{q}_Y(k) = \begin{bmatrix} \mathbf{q}_{11}(k) & \mathbf{q}_{12}(k)\,\mathbf{1}_D \\ \mathbf{0}_{1,s_1} & 0 \end{bmatrix}, \; k \in \mathbb{N}.$$

The reliability of the semi-Markov system :

$$\begin{aligned} R(k) &= \mathbb{P}(Z_l \in U, \forall l \in \{0, \ldots, k\}) \\ &= \mathbb{P}(Y_k \in U) \\ &= \sum_{j \in U} \sum_{i \in U} \mathbb{P}(Y_k = j \mid Y_0 = i)\, \mathbb{P}(Y_0 = i) \\ &= \sum_{j \in U} \sum_{i \in U} (P_Y)_{ij}(k)\, \mathbb{P}(Y_0 = i) \end{aligned}$$

For $i, j \in U$

$$(P_Y)_{ij}(k) = \mathbb{P}(Z_k = j, Z_l \in U, l = 1, \ldots, k-1 \mid Z_0 = i) = (\mathbf{P}_{11})_{ij}(k)$$

$\square$

Proof 2: Markov renewal equation

Define:

- $R_i(k) := \mathbb{P}(T_D > k \mid Z_0 = i), \; i \in U$
- $\mathbf{R}(k) := (R_1(k), \ldots, R_{s_1}(k))^\top \;\Rightarrow\; R(k) = \alpha_1 \mathbf{R}(k)$

Proof 2: Markov renewal equation

Define:

- $R_i(k) := \mathbb{P}(T_D > k \mid Z_0 = i),\ i \in U$
- $\mathbf{R}(k) := (R_1(k), \ldots, R_{s_1}(k))^\top \ \Rightarrow\ R(k) = \alpha_1 \mathbf{R}(k)$

For all $i \in U$

$$
\begin{aligned}
R_i(k) &= \mathbb{P}_i(T_D > k, S_1 > k) + \mathbb{P}_i(T_D > k, S_1 \leq k) \\
&\vdots \\
&= 1 - H_i(k) + \sum_{j \in U} \sum_{m=1}^{k} q_{ij}(m) R_j(k-m)
\end{aligned}
$$

$$
\mathbf{R}(k) = (\mathbf{I} - \mathbf{H}_{11})(k) + \mathbf{q}_{11}(k) * \mathbf{R}(k);\ \mathbf{R}(k) = (\delta \mathbf{I} - \mathbf{q}_{11})^{(-1)} * (\mathbf{I} - \mathbf{H}_{11})(k)
$$

$\square$

Plan

Two estimation procedures

1. adapted to reliability analysis

   consider one SM sample path of length M
   estimate the quantities of interest
   look for the asymptotic properties as $M \to \infty$

2. adapted to survival analysis

   consider K SM sample paths
   estimate the quantities of interest
   look for the asymptotic properties as $K \to \infty$

Consider

$$\mathcal{H}(M) := (J_0, X_1, \dots, J_{N(M)-1}, X_{N(M)}, J_{N(M)}, u_M),$$

where $u_M := M - S_{N(M)}$, $M \in \mathbb{N}$.

The associated likelihood function is

$$L(M) = \alpha(J_0) \prod_{k=1}^{N(M)} p_{J_{k-1}J_k} f_{J_{k-1}J_k}(X_k) \overline{H}_{J_{N(M)}}(u_M), \tag{3}$$

with $\overline{H}_{J_{N(M)}}(u_M) = \mathbb{P}(X_{N(M)+1} > u_M \mid J_{N(M)})$.

Consider

$$\mathcal{H}(M) := (J_0, X_1, \dots, J_{N(M)-1}, X_{N(M)}, J_{N(M)}, u_M),$$

where $u_M := M - S_{N(M)}$, $M \in \mathbb{N}$.

The associated likelihood function is

$$L(M) = \alpha(J_0) \prod_{k=1}^{N(M)} p_{J_{k-1}J_k} f_{J_{k-1}J_k}(X_k) \overline{H}_{J_{N(M)}}(u_M), \qquad (3)$$

with $\overline{H}_{J_{N(M)}}(u_M) = \mathbb{P}(X_{N(M)+1} > u_M \mid J_{N(M)})$.

We neglect $\overline{H}_{J_{N(M)}}(u_M)$ in (3) $\Rightarrow$ maximize :

$$L_1(M) = \alpha(J_0) \prod_{k=1}^{N(M)} p_{J_{k-1}J_k} f_{J_{k-1}J_k}(X_k).$$

Theorem

Let $(Z_n)_{n\in\mathbb{N}}$ be an irreducible and aperiodic SMC, with finite mean sojourn times, $m_i := \mathbb{E}_i(X_1) < \infty$. Let $\mathcal{H}(M)$ be a censored sample path of the chain. Then, the approached maximum likelihood estimators of $p_{ij}, f_{ij}(k)$ and $q_{ij}(k)$ are:

- $\widehat{p}_{ij}(M) = N_{ij}(M)/N_i(M)$
- $\widehat{f}_{ij}(k,M) = N_{ij}(k,M)/N_{ij}(M)$
- $\widehat{q}_{ij}(k,M) = N_{ij}(k,M)/N_i(M)$

where:

$$N_i(M) := \sum_{n=0}^{N(M)-1} \mathbf{1}_{\{J_n=i\}}$$

$$N_{ij}(M) := \sum_{n=1}^{N(M)} \mathbf{1}_{\{J_{n-1}=i,J_n=j\}}$$

$$N_{ij}(k,M) := \sum_{n=1}^{N(M)} \mathbf{1}_{\{J_{n-1}=i,J_n=j,X_n=k\}}.$$

$\Rightarrow$ obtain plug-in estimators for the quantities of interest

Theorem (Asymptotic normality)

For fixed $i, j \in E$ and $k \in \mathbb{N}$, the estimator of $q_{ij}(k)$ is asymptotically normal:

$$\sqrt{M}[\widehat{q}_{ij}(k, M) - q_{ij}(k)] \xrightarrow[M\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2_{q_{ij}}(k)),$$

$$\sigma^2_{q_{ij}}(k) \quad := \quad \mu_{ii} q_{ij}(k)[1 - q_{ij}(k)].$$

Proof 1 : main steps

- Use the central limit theorem for Markov renewal chains (Pyke and Schaufele, 1964 ; Moore and Pyke, 1968)
- There exists a measurable function $f : E \times E \times \mathbb{R} \to \mathbb{R}$ such that

$$\sqrt{M}[\widehat{q}_{ij}(k,M) - q_{ij}(k)] = \frac{M}{N_i(M)} \frac{1}{\sqrt{M}} \sum_{n=1}^{N(M)} f(J_{n-1}, J_n, X_n)$$

$$f(m,l,u) := \mathbf{1}_{\{m=i,l=j,u=k\}} - q_{ij}(k)\mathbf{1}_{\{m=i\}}$$

- For a MRC:

$$N_i(M)/M \xrightarrow[M\to\infty]{a.s.} 1/\mu_{ii}$$

□

Proof 2 : main steps

- Use the Lindeberg-Lévy CLT for martingales (Billingsley, 1961 ; 1995)
-

$$\begin{aligned}
&\sqrt{M}[\widehat{q}_{ij}(k,M) - q_{ij}(k)] \\
= \quad & \frac{M}{N_i(M)} \frac{1}{\sqrt{M}} \sum_{n=1}^{N(M)} [\mathbf{1}_{\{J_n=j, X_n=k\}} - q_{ij}(k)] \mathbf{1}_{\{J_{n-1}=i\}}
\end{aligned}$$

- $\mathcal{F}_n := \sigma(J_l, X_l; l \leq n)$ $Y_n := \mathbf{1}_{\{J_{n-1}=i, J_n=j, X_n=k\}} - q_{ij}(k)\mathbf{1}_{\{J_{n-1}=i\}}$
- $(Y_n)_{n\in\mathbb{N}}$ is a $\mathcal{F}_n-$difference martingale

- $\frac{1}{n}\sum_{l=1}^{n}\mathbb{E}(Y_l^2\mathbf{1}_{\{|Y_l|>\epsilon\}}) \xrightarrow[n\to\infty]{} 0$
- By the Lindeberg-Lévy theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{l=1}^{n} Y_l \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,\sigma^2); \quad \frac{1}{\sqrt{N(M)}}\sum_{l=1}^{N(M)} Y_l \xrightarrow[M\to\infty]{\mathcal{D}} \mathcal{N}(0,\sigma^2)$$

- σ^2 is given by

$$\sigma^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n}\mathbb{E}(Y_l^2 \mid \mathcal{F}_{l-1})$$

□

- $\frac{1}{n}\sum_{l=1}^{n}\mathbb{E}(Y_l^2 \mathbf{1}_{\{|Y_l|>\epsilon\}}) \xrightarrow[n\to\infty]{} 0$
- By the Lindeberg-Lévy theorem, we have

$$\frac{1}{\sqrt{n}}\sum_{l=1}^{n} Y_l \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0,\sigma^2); \quad \frac{1}{\sqrt{N(M)}}\sum_{l=1}^{N(M)} Y_l \xrightarrow[M\to\infty]{\mathcal{D}} \mathcal{N}(0,\sigma^2)$$

- σ^2 is given by

$$\sigma^2 = \lim_{n\to\infty}\frac{1}{n}\sum_{l=1}^{n}\mathbb{E}(Y_l^2 \mid \mathcal{F}_{l-1})$$

□

$\Rightarrow$ asymptotic normality of $\widehat{\psi}_{ij}(k,M)$ and $\widehat{P}_{ij}(k,M)$

Reliability/survival function estimation

Theorem

For a discrete-time SM system, for fixed $k \in \mathbb{N}$:

- $\widehat{R}(k,M) := \widehat{\alpha}_1 \left[\widehat{\psi}_{11}(\cdot,M) * \left(\mathbf{I} - \widehat{\mathbf{H}}(\cdot,M)_{11} \right) \right](k) \mathbf{1}_U$ *is strongly consistent, as* $M \to \infty$
- $\sqrt{M}[\widehat{R}(k,M) - R(k)] \xrightarrow[M\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma_R^2(k))$

$$\sigma_R^2(k) = \sum_{i=1}^{s} \mu_{ii} \left\{ \sum_{j=1}^{s} \left[D_{ij}^U - \mathbf{1}_{\{i\in U\}} \left(\sum_{t\in U} \alpha(t) \Psi_{ti} \right) \right]^2 * q_{ij}(k) - \left[\sum_{j=1}^{s} \left(D_{ij}^U * q_{ij} - \mathbf{1}_{\{i\in U\}} \left(\sum_{t\in U} \alpha(t) \psi_{ti} \right) * Q_{ij} \right) \right]^2 (k) \right\}$$

$$D_{ij}^U := \sum_{n\in U} \sum_{r\in U} \alpha(n) \psi_{ni} * \psi_{jr} * \left(I - diag(Q \cdot \mathbf{1}) \right)_{rr}$$

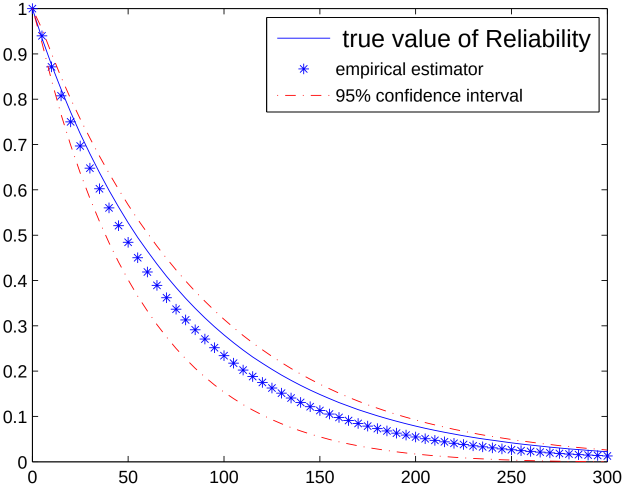

Figure: Confidence interval of reliability

References

- Barbu, V. and Limnios, N. (2008), *Semi-Markov Chains and Hidden Semi-Markov Models toward Applications–Their use in Reliability and DNA Analysis*, Lecture Notes in Statistics, vol. 191, Springer, New York.
- Barbu, V. and Limnios, N. (2010), Some algebraic methods in semi-Markov processes, In *Algebraic Methods in Statistics and Probability*, volume 2, series Contemporary Mathematics edited by AMS, Urbana, 19–35.
- Cooke, R.G. (1955), *Infinite Matrices and Sequence Spaces*, Dover Publications, New York.
- Kemeny, J.G., Snell, J.L. and Knapp, A.W. (1976), *Denumerable Markov chains*, Springer, New York.
- Trevezas, S. and Limnios, N. (2011), Exact MLE and asymptotic properties for nonparametric semi-Markov models, *Journal of Nonparametric Statistics*, 23 (3), 719–739.

CLT for Markov renewal chains

Let (J, S) be a MRC and $f : E \times E \times \mathbb{R} \to \mathbb{R}$ measurable.

$$W_f(M) := \sum_{i,j=1}^{s} \sum_{n=1}^{N_{ij}(M)} f(i, j, X_{ijn}) = \sum_{n=1}^{N(M)} f(J_{n-1}, J_n, X_n), M \in \mathbb{N}.$$

Put:

$$A_{ij} := \sum_{x=1}^{\infty} f(i, j, x) q_{ij}(x), \quad A_i := \sum_{j=1}^{s} A_{ij},$$

$$B_{ij} := \sum_{x=1}^{\infty} f^2(i, j, x) q_{ij}(x), \quad B_i := \sum_{j=1}^{s} B_{ij},$$

$$n_i := \sum_{j=1}^{s} A_j \frac{\mu_{ii}^*}{\mu_{jj}^*}, \quad m_f := \frac{n_i}{\mu_{ii}}, B_f := \frac{\sigma_i^2}{\mu_{ii}},$$

$$\sigma_i^2 := -n_i^2 + \sum_{j=1}^{s} B_j \frac{\mu_{ii}^*}{\mu_{jj}^*} + 2 \sum_{r=1}^{s} \sum_{l \neq i} \sum_{k \neq i} A_{rl} A_k \mu_{ii}^* \frac{\mu_{li}^* + \mu_{ik}^* - \mu_{lk}^*}{\mu_{rr}^* \mu_{kk}^*},$$

where μ_{ii}^* is the mean return time in state i for the chain J.

Theorem (Pyke and Schaufele, 1964; Moore and Pyke, 1968)

For an irreducible Markov renewal chain, with finite mean sojourn times, such that all the above sums are finite,we have:

$$\frac{1}{\sqrt{M}} [W_f(M) - M\ m_f] \xrightarrow[M \to \infty]{\mathcal{D}} \mathcal{N}(0, B_f).$$

return

The Lindeberg-Lévy CLT for martingales

Theorem (Billingsley, 1961 ; 1995)

Let $X_n, n \in \mathbb{N}^*$, *be a martingale with respect to the filtration* $\mathcal{F} = (\mathcal{F}_n)_{n\in\mathbb{N}}$ *and let* $Y_n := X_n - X_{n-1}, n \in \mathbb{N}^*$, *be the difference martingale (with* $Y_1 := X_1$*). If:*

1. $\frac{1}{n}\sum_{k=1}^n \mathbb{E}[Y_k^2 \mid \mathcal{F}_{k-1}] \xrightarrow[n\to\infty]{P} \sigma^2$,
2. $\frac{1}{n}\sum_{k=1}^n \mathbb{E}[Y_k^2 \mathbf{1}_{\{|Y_k|>\epsilon\}}] \xrightarrow[n\to\infty]{} 0$, *for all* $\epsilon > 0$,

then

$$\frac{1}{\sqrt{n}}X_n = \frac{1}{\sqrt{n}}\sum_{k=1}^n Y_k \xrightarrow[n\to\infty]{\mathcal{D}} \mathcal{N}(0, \sigma^2).$$

return