Reliability indicators for (hidden) semi-Markov models

Irene Votsi

Processus markoviens, semi-markoviens et leurs applications

Institut du Risque & de l'Assurance
Le Mans Université

LMM
Laboratoire Manceau de Mathématiques
Le Mans Université

Outline

- Introduction
  - Motivation
  - Semi-Markov processes
- Reliability indicators for semi-Markov chains
  - Rate of occurrence of failures

    $\hookrightarrow$ Application : Earthquake occurrences
  - Mean time to failure

    $\hookrightarrow$ Application : Wind energy production
- Reliability indicators for hidden semi-Markov chains
  - Rate of occurrence of failures
  - Particular cases
- Reliability indicators for semi-Markov processes
- Perspectives

Motivation

Context

- How could we estimate, prevent and manage the risk of failures for random systems ?
- Which are the stochastic models to describe such systems ?

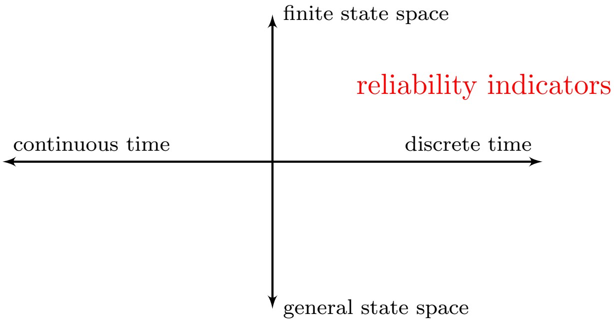
  ↪ Poisson, Markov, Cox, semi-Markov, etc.

Objectives

- Describe random systems by semi-Markov models.
- Estimate empirically reliability indicators

  ↪ ROCOF, reliability, availability, MTTF, MTBF, etc.

Markov processes



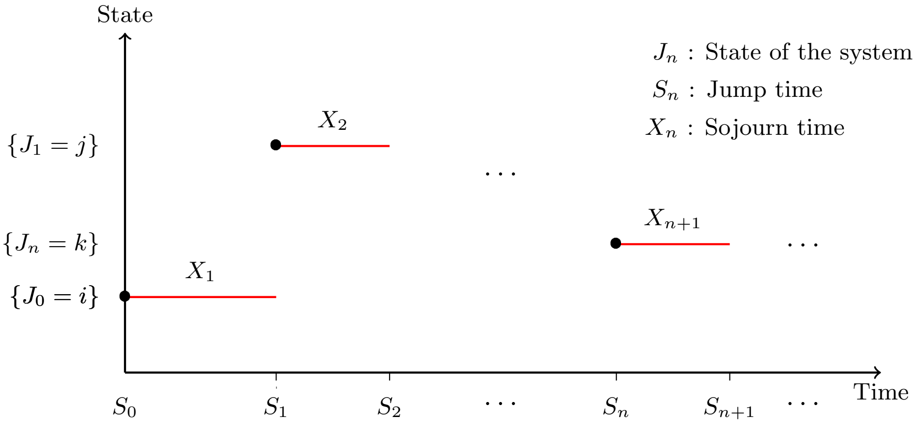
Markov property

$$P(future|present, past)$$

$$=$$

$$P(future|present)$$

Semi-Markov chains



J_n : State of the system

S_n : Jump time

X_n : Sojourn time

Definition

The chain $(\mathbf{J}, \mathbf{S}) = (J_n, S_n)_{n\in\mathbb{N}}$ *is a Markov renewal chain if it satisfies a.s.* $\forall k, n \in \mathbb{N}$, $\forall i, j \in E$

$$P(J_{n+1} = j, X_{n+1} = k | S_0, \ldots, S_n; J_0, \ldots, J_n = i) = P(J_{n+1} = j, X_{n+1} = k | J_n = i).$$

Definition

The semi-Markov chain (SMC) $\mathbf{Z} = (Z_k)_{k\in\mathbb{N}}$ *is defined by* $Z_k = J_{N(k)}$*, where* $N(k) = \max\{n \in \mathbb{N} | S_n \leq k\}$.

Embedded Markov chain

The chain $\mathbf{J} = (J_n)_{n\in\mathbb{N}}$ (*embedded Markov chain*) takes its values in E, describes the state of the system in the $n-$th jump and is characterized by

- initial probabilities
$$\alpha_i = P(J_0 = i), \quad i \in E.$$
- transition probabilities
$$p_{ij} = P(J_{n+1} = j | J_n = i), \quad i, j \in E, \quad n \in \mathbb{N}.$$

Characteristics

- Semi-Markov kernel $\mathbf{q} = \left(q_{ij}(\cdot)\right)_{i,j\in E}$

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

- Conditional sojourn time distribution $\mathbf{f} = \left(f_{ij}(\cdot)\right)_{i,j\in E}$

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

Remark

Note that

$$q_{ij}(k) = p_{ij}\ f_{ij}(k).$$

Particular semi-Markov chains

Case 1:

$$\begin{aligned} f_{i\bullet}(k) &= P(S_{n+1} - S_n = k | J_n = i) \\ q_{ij}(k) &= p_{ij} f_{i\bullet}(k) \end{aligned}$$

Case 2:

$$\begin{aligned} f_{\bullet j}(k) &= P(S_{n+1} - S_n = k | J_{n+1} = j) \\ q_{ij}(k) &= p_{ij} f_{\bullet j}(k) \end{aligned}$$

Case 3:

$$\begin{aligned} f(k) &= P(S_{n+1} - S_n = k) \\ q_{ij}(k) &= p_{ij} f(k) \end{aligned}$$

Markov chains $\hookrightarrow$ a particular case of semi-Markov chains

A Markov chain $(Y_n)_{n\in\mathbb{N}}$ with transition matrix $(\widetilde{p}_{ij})_{i,j\in E}$, $(\widetilde{p}_{ii} \neq 1, \forall i \in E)$ can be seen as a semi-Markov chain with

$$q_{ij}(k) = \begin{cases} \widetilde{p}_{ij}\,(\widetilde{p}_{ii})^{k-1}, & \text{if } i \neq j \text{ and } k \in \mathbb{N}^*, \\ 0, & \text{otherwise}, \end{cases}$$

$$p_{ij} = \begin{cases} \frac{\widetilde{p}_{ij}}{1-\widetilde{p}_{ii}}, & \text{if } i \neq j, \\ 0, & \text{otherwise}, \end{cases}$$

$$f_{ij}(k) = \begin{cases} (1-\widetilde{p}_{ii})\,(\widetilde{p}_{ii})^{k-1}, & \text{if } p_{ij} \neq 0, \\ 0, & \text{otherwise}. \end{cases}$$

Reliability indicators for semi-Markov chains

Rate of occurrence of failures

joint work with N.Limnios (LMAC, UTC)

Double Markov chain

Definition

The sequence of the backward recurrence times is defined by $\mathbf{U} = (U_k)_{k \in \mathbb{N}}$, *where*

$$U_k = k - S_{N(k)}.$$

Theorem (Limnios and Oprişan, 2001)

The chain $(\mathbf{Z}, \mathbf{U}) = (Z_k, U_k)_{k \in \mathbb{N}}$ *is a double Markov chain with initial law* $\widetilde{\alpha}$.

Transition law of (Z, U)

- Transition probabilities

$$\widetilde{P}\big((i,t_1),(j,t_2)\big) = P(Z_{k+1} = j, U_{k+1} = t_2 | Z_k = i, U_k = t_1),$$

$\forall\ (i,t_1),(j,t_2) \in E \times \mathbb{N},\ \forall\ k \in \mathbb{N}.$

- Survival function of sojourn times

$$\overline{H}_i(k) = 1 - \sum_{j\in E} \sum_{n=0}^{k} q_{ij}(n),$$

$\forall i \in E,\ k \in \mathbb{N}.$

Theorem (Chryssaphinou et al., 2008)

$$\widetilde{P}\big((i,t_1),(j,t_2)\big) = \begin{cases} q_{ij}(t_1+1)/\overline{H}_i(t_1), & \text{if} \quad i \neq j,\ t_2 = 0, \\ \overline{H}_i(t_1+1)/\overline{H}_i(t_1), & \text{if} \quad i = j,\ t_2 - t_1 = 1, \\ 0, & \text{otherwise}, \end{cases}$$

$\forall\ (i,t_1),(j,t_2) \in E \times \mathbb{N},\ \forall\ k \in \mathbb{N}.$

Rate of occurrence of failures

Z takes its values in $E = \{1, 2, \ldots, s\}$. We partition $E = U \cup D$ $(U, D \neq \emptyset)$ s.t.

- $U = \{1, 2, \ldots, r\}$ $\hookrightarrow$ up states
- $D = \{r+1, \ldots, s\}$ $\hookrightarrow$ down states
- At time k, the number of transitions of the SMC from U to D is defined by:

$$N_U(k) = \sum_{l=1}^{k} 1_{\{Z_{l-1} \in U, Z_l \in D\}}$$

Definition

The rate of occurrence of failures is the mean transition number of the SMC to D at time k:

$$\widetilde{r}_U(k) = \mathbb{E}[N_U(k) - N_U(k-1)].$$

Literature

- Markov models (continuous time)
  - Discrete state space: Yeh (1997), D'Amico (2015).
- Semi-Markov models (continuous time)
  - Discrete state space: Ouhbi and Limnios (2002).
  - Continuous state space: Limnios (2012).
- (Hidden) semi-Markov models (discrete time)
  - Discrete state space: V. et al. (2014), V. and Limnios (2015), Votsi (2018)
  - Application fields: **Seismology, Energy** etc

Evaluation

Theorem

The rate of occurrence of failures of the SMC at time $k \in \mathbb{N}^*$ is given by

$$\widetilde{r}_U(k) = \sum_{i \in U} \sum_{j \in D} \sum_{m=0}^{k-1} [(\widetilde{a}\widetilde{P}^{k-1})(i,m)] \widetilde{P}\big((i,m),(j,0)\big).$$

Empirical estimation

- Trajectory of the SMC **Z** up to arbitrary time $M \in \mathbb{N}$:

$$H(M) = (J_0, S_1, \ldots, J_{N(M)-1}, S_{N(M)}, J_{N(M)}, U_M).$$

Definition

The estimator of the rate of occurrence of failures is

$$\widehat{\widetilde{r}}_U(k, M) = \sum_{i \in U} \sum_{j \in D} \sum_{m=0}^{k-1} [(\widehat{\widetilde{\alpha}} \widehat{\widetilde{P}}_M^{k-1})(i, m)] \widehat{\widetilde{P}}_M((i, m), (j, 0)),$$

where $(\widehat{\widetilde{\alpha}} \widehat{\widetilde{P}}_M^{k-1})(i, m)$ *is the* (i, m) *element of the vector* $\widehat{\widetilde{\alpha}} \widehat{\widetilde{P}}_M^{k-1}$*, for every* $k \in \mathbb{N}^*$.

Following Barbu and Limnios (2008):

- Semi-Markov kernel:

$$\widehat{q}_{ij}(k, M) = \frac{1}{N_i(M)} \sum_{n=1}^{N(M)} \mathbf{1}_{\{J_{n-1}=i, J_n=j, X_n=k\}}$$

- Survival function of sojourn times:

$$\widehat{\overline{H}}_i(k, M) = 1 - \sum_{j \in E} \sum_{n=0}^{k} \widehat{q}_{ij}(n, M)$$

- Transition probabilities of $(\mathbf{Z}, \mathbf{U})$:

$$\widehat{\widetilde{P}}_M\big((i, t_1), (j, t_2)\big) = \begin{cases} \widehat{q}_{ij}(t_1 + 1, M) / \widehat{\overline{H}}_i(t_1, M), & \text{if } i \neq j,\ t_2 = 0, \\ \widehat{\overline{H}}_i(t_1 + 1, M) / \widehat{\overline{H}}_i(t_1, M), & \text{if } i = j,\ t_2 - t_1 = 1, \\ 0, & \text{otherwise.} \end{cases}$$

Consistency

Proposition

For any fixed, arbitrary $k \in \mathbb{N}^*$, $\widehat{\widetilde{r}}_U(k, M)$ *is strongly consistent, i.e.*

$$\widehat{\widetilde{r}}_U(k, M) \xrightarrow[M\to\infty]{a.s.} \widetilde{r}_U(k).$$

Asymptotic normality

Theorem

Let $\big(Z_k, U_k\big)_{k\in\mathbb{N}}$ be an homogeneous, ergodic Markov chain. For any fixed, arbitrary $k \in \mathbb{N}^*$

$$\sqrt{M}\big(\widehat{\widetilde{r}}_U(k, M) - \widetilde{r}_U(k)\big) \xrightarrow[M\to\infty]{\mathcal{L}} \mathcal{N}(0, \Phi'\Gamma\Phi'^{\top}),$$

where $\Phi : [0,1]^d \to \mathbb{R}^+$ $(d = s^2(M+1)^2)$ is the function

$$\Phi\big(\widetilde{P}\big) = \sum_{i\in U}\sum_{j\in D}\sum_{m=0}^{k-1}\Big(\sum_{s\in E}\widetilde{a}(s,0)\widetilde{P}^{k-1}((s,0),(i,m))\Big)\widetilde{P}((i,m),(j,0))$$

and Γ is the asymptotic covariance matrix of the random vector $F = \big(f_{(i,t_1)(j,t_2)}\big)_{(i,t_1),(j,t_2)\in E\times T_M}$, with

$$f_{(i,t_1)(j,t_2)} = \sqrt{M}\Big(\widehat{\widetilde{P}}_M((i,t_1),(j,t_2)) - \widetilde{P}((i,t_1),(j,t_2))\Big).$$

Real data

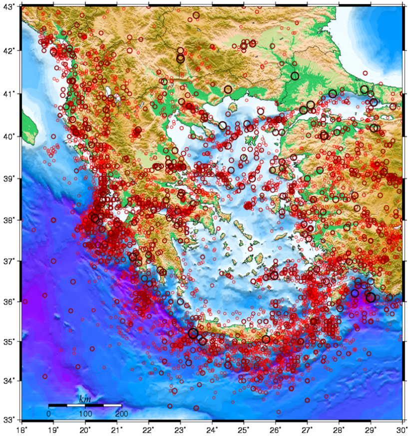


Figure: Epicentral distribution of earthquakes that occurred in the study area from 6^{th} century BC up to May 2011.

Data

- Study area: Greece
- Study period: $[1845, 2016]$
- Magnitudes: $M \geq 6.5$

Semi-Markov model

- U: $M \in [6.5, 7.1]$
- D: $M > 7.1$

Source

http://geophysics.geo.auth.gr//ss

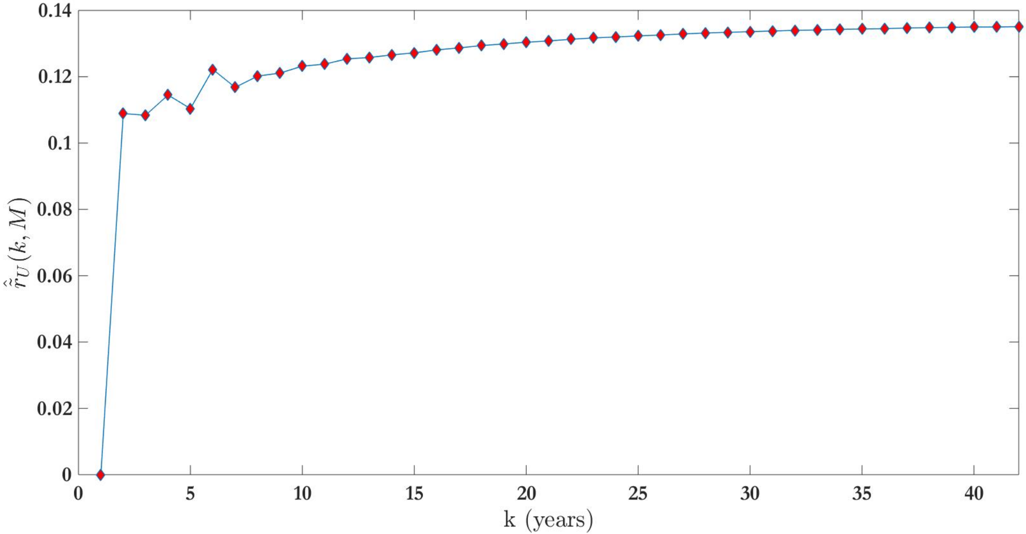


Figure: Occurrence rate of earthquakes with magnitudes $M > 7.1$, $\widehat{\tilde{r}}_U(k, M)$.

Conditional mean time to failure

joint work with A.Brouste (LMM, Le Mans U)

Wind energy production

Motivation

- Estimate failure risks for wind farms via semi-Markov models.
- Provide support for power production management.

Main results

- Asymptotic properties of the estimator of the Conditional Mean Time To Failure (CMTTF).
- Application to wind data: highlight the importance of wind direction on CMTTF.

Conditional mean time to failure

Z takes its values in $E = \{1, 2, \ldots, s\}$. We partition $E = U \cup D$ $(U, D \neq \emptyset)$ s.t.

- $U = \{1, 2, \ldots, r\}$ $\hookrightarrow$ up states
- $D = \{r+1, \ldots, s\}$ $\hookrightarrow$ down states
- The first passage time in D is defined by:

$$T_D = \inf\{k \in \mathbb{N} : Z_k \in D\} \quad \text{and} \quad \inf\{\varnothing\} = \infty.$$

Definition

The conditional mean time to failure is defined by

$$CMTTF_i = \mathbb{E}(T_D | J_0 = i),$$

for any state $i \in U$.

Evaluation

Vector of the conditional mean times to failure:

$$\begin{aligned}\mathbf{CMTTF} &= (CMTTF_1, \ldots, CMTTF_r)^\top \\ &= (\mathbf{I} - \mathbf{P}_{11})^{-1}\mathbf{m_1},\end{aligned}$$

where $\mathbf{P}_{11} = (p_{ij};\ i, j \in U)$ and $\mathbf{m}_1 = (\mathbb{E}(S_1 | J_0 = i);\ i \in U)$.

Empirical estimation

Given a trajectory $H(M)$, we have

$$\widehat{\mathbf{CMTTF}}(M) = (\mathbf{I} - \widehat{\mathbf{P}}_{11}(M))^{-1}\widehat{\mathbf{m}}_1(M),$$

where

- $\widehat{\mathbf{P}}_{11}(M) = (\widehat{p}_{ij}(M);\ i, j \in U)$ and $\widehat{p}_{ij}(M) = \frac{N_{ij}(M)}{N_i(M)}$;
- $\widehat{\mathbf{m}}_1(M) = (\widehat{m}_i(M);\ i \in U)^\top$ and $\widehat{m}_i(M) = \sum_{\ell \geq 0} \widehat{\overline{H}}_i(\ell, M)$.

Consistency

Theorem

For any state $i \in U$, $C\widehat{MTTT}F_i(M)$ *is strongly consistent, i.e.*

$$C\widehat{MTTT}F_i(M) \xrightarrow[M\to\infty]{a.s.} CMTTF_i.$$

Asymptotic normality

Theorem

For any state $i \in U$, the random variable $\widehat{CMTTF}_i(M)$, is asymptotically normal, in the sense that

$$\sqrt{M}(\widehat{CMTTF}_i(M) - CMTTF_i) \xrightarrow[M \to \infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2_{CMTTF_i}),$$

with the asymptotic variance

$$\sigma^2_{CMTTF_i} = \sum_{m \in E} a^2_{im} \mu_{mm} \Big(\sigma^2_m + \sum_{\ell \in E} (\eta_\ell - \widetilde{\eta}_m)^2 p_{m\ell} + 2 \sum_{\ell \in E} \eta_\ell Q_{m\ell} \Big),$$

where $Q_{m\ell} = \sum_{u=1}^{+\infty} (u - m_m) q_{m\ell}(u)$, $a_{ij} = (\mathbf{I} - \mathbf{P}_{11})^{-1}_{ij}$, $\eta_\ell = \sum_{r \in U} m_r a_{r\ell}$, $\widetilde{\eta}_m = \sum_{j \in U} p_{mj} \eta_j$, and σ^2_m is the variance of the sojourn time in state m.

Semi-Markov modeling

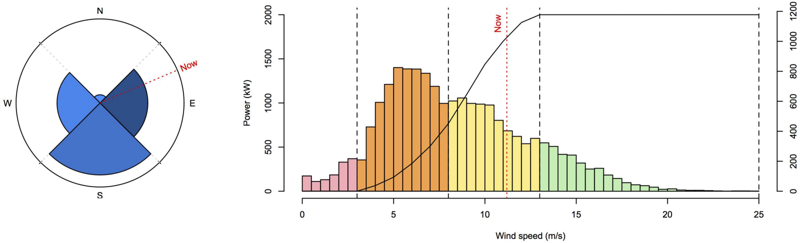


Figure: Wind directions and speeds states. The transfer power function of a 2MW wind turbine is superposed on the wind speeds histogram.

Data Selection

- Real Data $\hookrightarrow$ EREN Group
- Simulated Data $\hookrightarrow$ NRE Lab

States

- $U \hookrightarrow (Direction, Speed)$, $Speed > 3\ m/s$.
- $D \hookrightarrow (Direction, Speed)$, $Speed \leq 3\ m/s$.

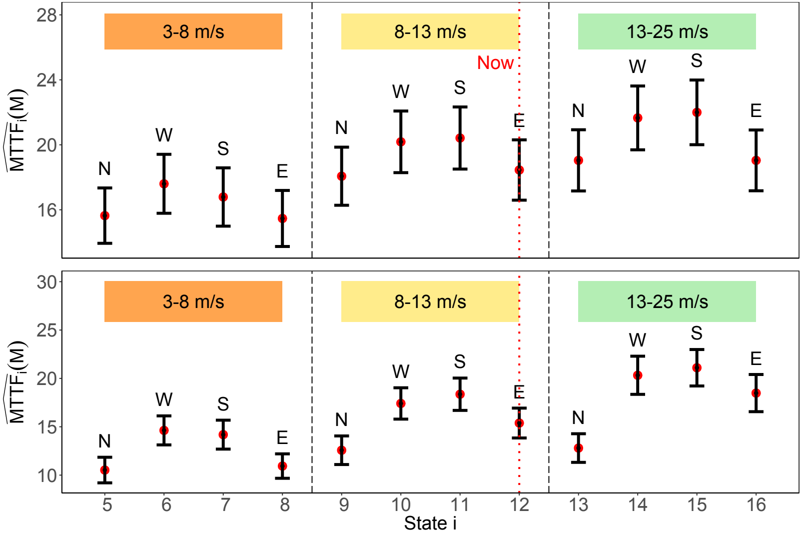


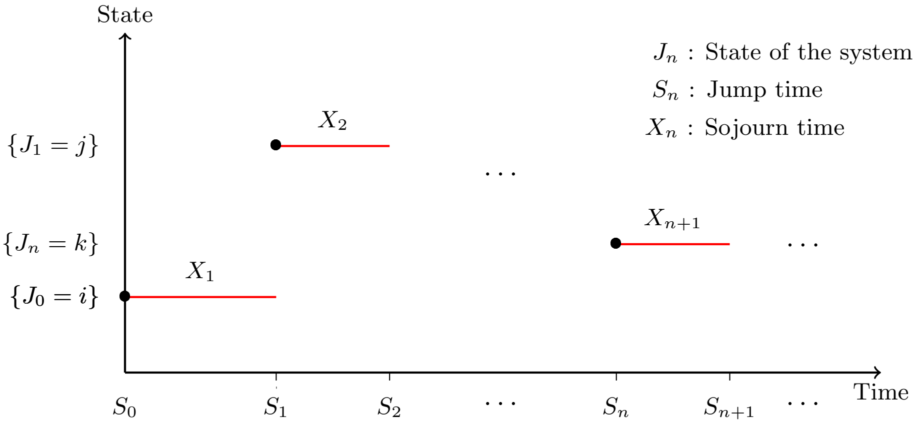
Figure: Empirical estimators of the conditional mean times to failure (in hours), $\widehat{MTTF}_i(M)$, $i \in U$, evaluated for real data (EREN) (upper panel) and simulated data (NREL) (lower panel).

Reliability indicators for hidden semi-Markov chains

Rate of occurrence of failures

joint work with N.Limnios (LMAC, UTC)

Semi-Markov chains



J_n : State of the system

S_n : Jump time

X_n : Sojourn time

Hidden Markov renewal chains

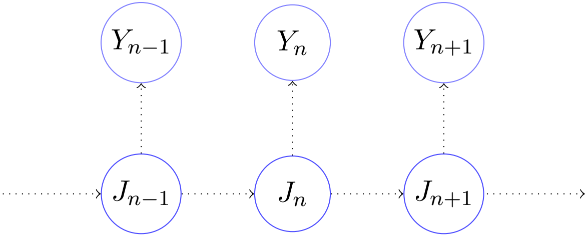


Figure: Trajectory of a hidden Markov renewal chain.

- *Underlying chain* $\hookrightarrow$ semi-Markov chain $\mathbf{Z} = (Z_k)_{k\in\mathbb{N}}$
- *Observation sequence* $\hookrightarrow$ $\mathbf{Y} = (Y_n)_{n\in\mathbb{N}}$
- ⚠ *Assumption:* The process is observed only at jump times.

Definitions

- The process $(\mathbf{J}, \mathbf{S}, \mathbf{Y}) = (J_n, S_n, Y_n)_{n \in \mathbb{N}}$ is a hidden Markov renewal chain.
- State spaces
  - Hidden Markov renewal chain $(\mathbf{J}, \mathbf{S}, \mathbf{Y}) = (J_n, S_n, Y_n)_{n \in \mathbb{N}} \hookrightarrow E^*$
  - Markov renewal chain $(\mathbf{J}, \mathbf{S}) = (J_n, S_n)_{n \in \mathbb{N}} \hookrightarrow L = E \times \mathbb{N}$
  - Observation process $\mathbf{Y} = (Y_n)_{n \in \mathbb{N}} \hookrightarrow A$
- Emission probabilities

$$R_{i;a} = P(Y_n = a | J_n = i), \quad i \in E, a \in A \quad n \in \mathbb{N}.$$

Rate of occurrence of failures

Y takes its values in $A = \{1, 2, \ldots, s\}$. We partition $A = U \cup D$ $(U, D \neq \emptyset)$ s.t.

- $U = \{1, 2, \ldots, r\}$ $\hookrightarrow$ up states
- $D = \{r+1, \ldots, s\}$ $\hookrightarrow$ down states
- At time k, the number of transitions of the observation sequence from U to D is defined by:

$$N_U(k) = \sum_{\ell=1}^{k} 1_{\{Y_{\ell-1} \in U, Y_\ell \in D\}}$$

Definition

The rate of occurrence of failures is the mean transition number of the observation sequence to D at time k:

$$r_U^\sharp(k) = \mathbb{E}[N_U(k) - N_U(k-1)].$$

Evaluation

Theorem

The rate of occurrence of failures at time $k \in \mathbb{N}^*$ is given by

$$r_U^\sharp(k) \;=\; \sum_{(i,y)\in E\times A}\sum_{\ell=1}^{k}\sum_{s_0,s_1\in E}\sum_{k_0=0}^{k} a(i,0)R_{i;y}\mathrm{R}_{s_0}(U)\mathrm{R}_{s_1}(D)q_{is_0}^{(l-1)}(k_0)q_{s_0s_1}(k-k_0).$$

Empirical estimation

- Trajectory of the HMRC $(\mathbf{J}, \mathbf{S}, \mathbf{Y})$ up to arbitrary time $M \in \mathbb{N}$:

$$H(M) = (J_0, S_1, Y_0, \ldots, S_{N(M)}, J_{N(M)}, Y_{N(M)}).$$

Definition

The estimator of the rate of occurrence of failures is

$$\widehat{r}^{\sharp}_U(k, M) = \sum_{(i,y)\in E\times A} \sum_{\ell=1}^{k} \sum_{s_0, s_1 \in E} \sum_{k_0=0}^{k} \widehat{a}(i, M)\widehat{R}_{i;y}(M)\widehat{\mathrm{R}}_{s_0}(U, M)\widehat{\mathrm{R}}_{s_1}(D, M)\widehat{q}^{(l-1)}_{is_0}(k_0, M)\widehat{q}_{s_0 s_1}(k - k_0, M).$$

Consistency

Proposition

For any fixed, arbitrary $k \in \mathbb{N}^*$, $\widehat{r}_U^{\sharp}(k, M)$ *is strongly consistent, i.e.*

$$\widehat{r}_U^{\sharp}(k, M) \xrightarrow[M \to \infty]{a.s.} r_U^{\sharp}(k).$$

Asymptotic normality

Theorem

Under certain conditions, and for any fixed, arbitrary $k \in \mathbb{N}^*$

$$\sqrt{M}\left(\widehat{r}^{\sharp}_U(k,M)\right) \xrightarrow[M\to\infty]{\mathcal{L}} \mathcal{N}(0, \Phi'\Gamma\Phi'^{\top}),$$

where $\Phi : \mathbb{R}^{l_0 d_0} \to \mathbb{R}^+$ *is the function*

$$\Phi\Big((R_{j';m'}, q_{i'j'}(k'-k'_0));\ (i',k'_0) \in L, (i',j',k') \in E^*\Big)$$
$$= \sum_{i\in U}\sum_{j\in D}\sum_{(\ell,y)\in E\times A}\sum_{l=1}^{k}\sum_{s_0,s_1\in E}\sum_{k_0=0}^{k} a(\ell,y)R_{\ell;y}R_{s_0}(i)R_{s_1}(j)q_{is_0}^{(l-1)}(k_0)q_{s_0 s_1}(k-k_0),$$

and Γ *the asymptotic covariance matrix of* $F = \left(f_{(i,t_1)(j,t_2,m)}\right)_{(i,t_1)\in L,(j,t_2,m)\in E^*}$, *where*

$$f_{(i,t_1)(j,t_2,m)} = \sqrt{M}\Big(\widehat{q}_{ij}(t_2-t_1,M)\widehat{R}_j(m,M) - q_{ij}(t_2-t_1)R_j(m)\Big).$$

Hidden semi-Markov chains

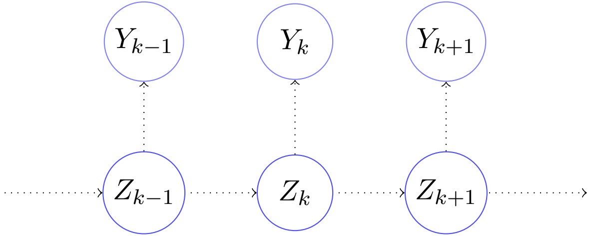


Figure: Trajectory of a hidden Markov renewal chain.

- *Underlying chain* $\hookrightarrow$ semi-Markov chain $\mathbf{Z} = (Z_k)_{k\in\mathbb{N}}$
- *Observation sequence* $\hookrightarrow$ $\mathbf{Y} = (Y_k)_{k\in\mathbb{N}}$
- *Assumption:* The process is observed only at jump times

Evaluation

Theorem

The ROCOF at time $k \in \mathbb{N}^*$ is given by

$$r(k) = \sum_{a_1 \in U} \sum_{a_2 \in D} \sum_{i \in E} \sum_{j \in E} \sum_{u_1 \in T_{k-1}} \sum_{u_2 \in T_k} R_{j;a_2} \widetilde{P}\big((i,u_1),(j,u_2)\big) R_{i;a_1} (aP)^{k-1}(i,u_1).$$

Hidden Markov chains

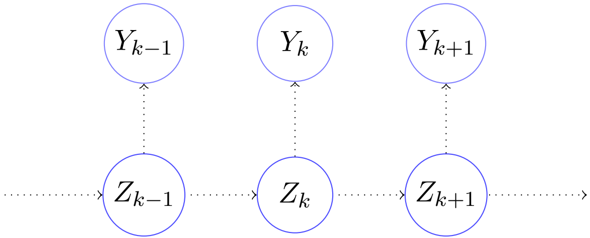


Figure: Trajectory of a hidden Markov chain.

- *Underlying chain* $\hookrightarrow$ Markov chain $\mathbf{Z} = (Z_k)_{k\in\mathbb{N}}$
- *Observation sequence* $\hookrightarrow$ $\mathbf{Y} = (Y_k)_{k\in\mathbb{N}}$

Evaluation

Theorem

The ROCOF at time $k \in \mathbb{N}^*$ is given by

$$r(k) = \sum_{a_1 \in U} \sum_{a_2 \in D} \sum_{i \in E} \sum_{j \in E} \alpha(i_1) p_{i_1 i}^{(k-1)} p_{ij} R_{i;a_1} R_{j;a_2}.$$

Reliability indicators for semi-Markov processes

Context

- $(J_n)_{n\in\mathbb{N}}$ is defined in a discrete state space E;
- $(S_n)_{n\in\mathbb{N}}$ is defined in $\mathbb{R}^+$.

Definition

- *The Markov renewal process* $(J_n, S_n)_{n\in\mathbb{N}}$ *satisfies a.s.*

$$P(J_{n+1}, X_{n+1} \leq x | J_0, \dots, J_n, S_0, \dots, S_n) = P(J_{n+1}, X_{n+1} \leq x | J_n),$$

  $\forall n \in \mathbb{N}$ *and* $\forall x \in \mathbb{R}^+$.

- *The semi-Markov process is defined by* $Z_t = J_{N(t)}$, *where*

$$N(t) = \sup\{n \in \mathbb{N} : S_n \leq t\}, \ t \in \mathbb{R}^+.$$

Semi-Markov kernel

$$Q_{ij}(x) = P(J_{n+1} = j, X_{n+1} \leq x | J_n = i), \quad \forall i, j \in E, \, x \in \mathbb{R}^+$$

Conditional sojourn time distribution

$$F_{ij}(x) = P(X_{n+1} \leq x | J_n = i, J_{n+1} = j), \quad \forall i, j \in E, \, x \in \mathbb{R}^+$$

Survival time distribution

$$H_i(x) = P(X_{n+1} \geq x | J_n = i), \quad \forall i \in E, \, x \in \mathbb{R}^+$$

Empirical estimators

- Observation of a sample path, in the time interval $[0, T]$

$$H_T = (J_0, \dots, J_{N(T)}, S_1, \dots, S_{N(T)}, T - S_{N(T)})$$

- Semi-Markov kernels

$$\widehat{Q}_{ij}(x, T) \quad := \quad \frac{1}{N_i(T)} \sum_{k=1}^{N(T)} 1_{\{J_{k-1}=i, J_k=j\}}, \quad 0 \leq x \leq T, \quad i, j \in E$$

- Conditional sojourn time distribution

$$\widehat{F}_{ij}(x, T) \quad := \quad \frac{1}{N_{i,j}(T)} \sum_{k=1}^{N(T)} 1_{\{J_{k-1}=i, J_k=j\}}, \quad 0 \leq x \leq T, \quad i, j \in E$$

Asymptotic properties

(Limnios and Oprişan, 2001)

For any fixed $x > 0$, we have

$$\sqrt{T}\big(\widehat{Q}_{ij}(x,T) - Q_{ij}(x)\big) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2(ij)),$$

where $\sigma^2(ij) = \mu_{ii} Q_{ij}(x)(1 - Q_{ij}(x))$.

(Limnios and Oprişan, 2001)

For any fixed $x > 0$, we have

$$\sqrt{T}\big(\widehat{F}_{ij}(x,T) - F_{ij}(x)\big) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}(0, \sigma'^2(ij)),$$

where $\sigma'^2(ij) = \frac{\mu_{ii}}{p_{ij}} F_{ij}(x)(1 - F_{ij}(x))$.

Reliability

Z takes its values in $E = \{1, 2, \dots, s\}$. We partition $E = U \cup D$ $(U, D \neq \emptyset)$ s.t.

- $U = \{1, 2, \dots, r\}$ $\hookrightarrow$ up states
- $D = \{r+1, \dots, s\}$ $\hookrightarrow$ down states
- The reliability at time $t \in \mathbb{R}^+$ is defined by:

$$R(t) = P\big(Z(u) \in U, \forall u \in [0, t]\big).$$

- Explicit form

$$R(t) = a(0)(\mathtt{I} - Q_{00}(t))^{(-1)} * (\mathtt{I} - H_0(t))\mathtt{1}_r$$

- Empirical estimator

$$\widehat{R}(t,T) = a(0)(\mathtt{I} - \widehat{Q}_{00}(t,T))^{(-1)} * (\mathtt{I} - \widehat{H}_0(t,T))1_r$$

Asymptotic normality

Theorem (Limnios and Oprisan, 2001)

Under some mild assumptions, for any fixed $t \in \mathbb{R}^+$, *we have*

$$\sqrt{T}\left(\widehat{R}(t,T) - R(t)\right) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}(0, \sigma^2(t)),$$

where

$$\begin{aligned}\sigma^2(t) &= \sum_{i\in U}\sum_{j\in E}\mu_{ii}\Big\{(B^0_{ij}1_{\{j\in U\}} - \sum_{r\in U} a(r)\psi^0_{ri})^2 * Q_{ij}(t) \\ &\qquad -\Big[(B^0_{ij}1_{\{j\in U\}} - \sum_{r\in U} a(r)\psi^0_{ri}) * Q_{ij}(t)\Big]^2\Big\},\end{aligned}$$

$$B_{ij} = \sum_{n\in E}\sum_{k\in U} a(i)B_{nijk} * \left(I - diag(Q(t)1_{kk}\right), \quad B_{irkj}(x) = \sum_{n=1}^{\infty}\sum_{\ell=1}^{n} Q^{(\ell-1)}_{ir} * Q^{(n-\ell)}_{kj}(t).$$

Kernel-type (KT) estimators

Let $K(\cdot)$ be a distribution function on $\mathbb{R}$ and $h_n \in \mathbb{R}^+$ the bandwidth. Following Bouzebda *et al.* (2018), we have:

- Semi-Markov kernel

$$\widetilde{Q}_{ij;h_n}(x,n) := \frac{1}{N_i(n)} \sum_{k=1}^{N(n)} 1_{\{J_{k-1}=i, J_k=j\}} K\left(\frac{x-X_k}{h_n}\right), \quad 0 \le x \le T, \quad i,j \in E.$$

- Conditional sojourn time distribution

$$\widetilde{F}_{ij;h_n}(x,n) := \frac{1}{N_{i,j}(n)} \sum_{k=1}^{N(n)} 1_{\{J_{k-1}=i, J_k=j\}} K\left(\frac{x-X_k}{h_n}\right) \quad 0 \le x \le T, \quad i,j \in E.$$

Bootstrapped estimators

- Let $\mathbf{W} \equiv (W_{nj}, j = 1, \ldots, n, n = 1, 2 \ldots)$ be an array of random variables.

- Semi-Markov kernel

$$\widetilde{Q}_{ij}^{W}(x,T) := \frac{1}{N_i(T)} \sum_{k=1}^{N(T)} W_{N(T)k} 1_{\{J_{k-1}=i, J_k=j\}}, \quad 0 \leq x \leq T, \quad i,j \in E.$$

- Conditional sojourn time distribution

$$\widetilde{F}_{ij}^{W}(x,T) := \frac{1}{N_{i,j}(T)} \sum_{k=1}^{N(T)} W_{N(T)k} 1_{\{J_{k-1}=i, J_k=j\}}, \quad 0 \leq x \leq T, \quad i,j \in E.$$

Bootstrapped KT estimators

- Semi-Markov kernel

$$\widetilde{Q}^{W}_{ij;h_n}(x,n) \quad := \quad \frac{1}{N_i(n)} \sum_{k=1}^{N(n)} W_{N(n)k} 1_{\{J_{k-1}=i, J_k=j\}} K\left(\frac{x - X_k}{h_n}\right), \quad 0 \leq x \leq T, \quad i,j \in E$$

- Conditional sojourn time distribution

$$\widetilde{F}^{W}_{ij;h_n}(x,n) \quad := \quad \frac{1}{N_{i,j}(n)} \sum_{k=1}^{N(n)} W_{N(n)k} 1_{\{J_{k-1}=i, J_k=j\}} K\left(\frac{x - X_k}{h_n}\right), \quad 0 \leq x \leq T, \quad i,j \in E$$

Assumptions - bootstrapped estimators

W.1 *The vector* $W_n = (W_{n1}, \dots, W_{nn})^\top$ *is exchangeable for any* $n = 1, 2, \dots$, *i.e., for any permutation* $\pi = (\pi_1, \dots, \pi_n)$ *of* $(1, \dots, n)$, *the joint distribution of*

$$\pi(W_n) = (W_{n\pi_1}, \dots, W_{n\pi_n})^\top$$

is the same as that of W_n;

W.2 $W_{ni} \geq 0$ *for all* n, i *and* $\sum_{i=1}^n W_{ni} = n$ *for all* n;

W.3 $\limsup_{n\to\infty} \int_0^\infty \sqrt{\mathbb{P}(W_{n1} > u)}du \leq C < \infty$;

W.4

$$\lim_{\lambda\to\infty} \limsup_{n\to\infty} \sup_{t\geq\lambda} t^2 \mathbb{P}(W_{n1} > t) = 0;$$

W.5 $(1/n)\sum_{i=1}^n (W_{ni} - 1)^2 \xrightarrow{\mathbb{P}} c^2 > 0$.

Assumptions - KT estimators

Assumptions

Fs For the derivative of order $s \geq 1$ of $Q_{i,j}(x)$ with respect to x, there exists a constant $0 < C < \infty$ such that for all $i, j \in E$

$$\sup_{x \in \mathbb{R}_+} \left| \frac{d^s}{dx^s} Q_{i,j}(x) \right| < C.$$

C.1. $h_n \to 0$, $nh_n \to \infty$ and $\sqrt{n} h_n^s \to 0$ as $n \to \infty$.

C.2. $k(\cdot)$ is a continuous density function and compactly supported,

C.3. $k(\cdot)$ is of order s.

Asymptotics

Theorem (Bootstrapped estimators)

For any arbitrary but fixed $i, j \in E$, *and fixed* $x \in \mathbb{R}_+$, *we have*

$$T^{1/2}(\widehat{Q}_{ij}^{W}(x,T) - Q_{ij}(x)) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}(0, c^2 b_{ij}^2(x)),$$

where $b_{ij}^2(x) = \mu_{ii} Q_{ij}(x)(1 - Q_{ij}(x))$, *and*

$$T^{1/2}(\widehat{F}_{ij}^{W}(x,T) - F_{ij}(x)) \xrightarrow[T\to\infty]{\mathcal{L}} \mathcal{N}(0, c^2 d_{ij}^2(x)),$$

where $d_{ij}^2(x) = \frac{\mu_{ii}}{p_{ij}} F_{ij}(x)(1 - F_{ij}(x))$.

Reliability

Definition

The bootstrapped estimator is defined by

$$\widetilde{R}^W(t,T) = a(0)(I - \widetilde{Q}^W_{00}(t,T))^{(-1)} * (I - \widetilde{H}^W_0(t,T))1_r,$$

for any $0 < t < T$.

Theorem

For any fixed $t > 0$, *we have*

$$T^{1/2}\big(\widehat{R}^W(t,T) - R(t)\big) \xrightarrow[T\to\infty]{\mathcal{L}} N\big(0, c^2\sigma^2(t)\big),$$

where $N\big(0, c^2\sigma^2(t)\big)$ *is a normal random variable with mean* 0 *and variance* $c^2\sigma^2(t)$.

Perspectives

KT and bootstrapped KT estimators

- Reliability, Availability, Maintainability
- Failure rates
- Asymptotic properties/Assumptions ?

Numerical applications

- Simulated data ?
- Real data ?

Perspectives

KT and bootstrapped KT estimators

- Reliability, Availability, Maintainability
- Failure rates
- Asymptotic properties/Assumptions ?

Numerical applications

- Simulated data ?
- Real data ?

Biblio I

- Barbu, V. & Limnios, N. (2008), Semi-Markov chains and hidden semi-Markov models toward applications, Lecture Notes in Statistics, Springer.
- Barbu, V., D'Amico, G. & Gkelsinis, T. (2021), Sequential interval reliability for discrete-time homogeneous semi-Markov repairable systems, *Mathematics*, 9 (16).
- Barbu, V., Makrides, A. & Karagrigoriou, A. (2019), Estimation and reliability for a special-type of semi-Markov processes, *Math. Stat.*, 15, 259-272.
- Barbu, V., Karagrigoriou, A. & Makrides, A. (2017), Semi-Markov modelling for multistate systems, *Method. Comput. Applied Proba.*, 19, 1011-1028.
- Bouzebda, S., Papamichail, Ch. & Limnios, N. (2018), On a multidimensional general bootstrap for empirical estimator of continuous-time semi-Markov kernels with applications, *J Nonparametr Stat*, 30 (1), 49-86.
- Chryssaphinou, O., Karaliopoulou, M. and Limnios, N. (2008), On discrete time semi-Markov chains and applications in words occurrences. Commun Stat-Theor M 37, 1306-1322.

Biblio II

- D'Amico, G., Petroni, F. & Prattico, F. (2015), Reliability measures for indexed semi-Markov chains applied to wind energy production, *Reliab. Engin. System Safety*, 144, 170-177.
- D'Amico, G., Petroni, F. & Prattico, F. (2015), Performance analysis of second order semi-Markov chains: an application to wind energy production, *Method. Comput. Applied Proba.*, 17(3), 781-794.
- Limnios, N. & Oprişan, G. (2001), Semi-Markov processes and reliability, Birkhauser Boston, 222 p.
- Votsi, I. & Brouste, A. (2019), Confidence intervals for risk indicators in semi-Markov models: an application to wind energy production, *J Appl Stat*, 46(10), 1756-1773.
- Votsi, I. & Limnios, N. (2015), Estimation of the intensity of hitting time for semi-Markov chains and hidden Markov renewal chains, *J Nonparametr Stat*, 27(2), 149-166.
- Votsi, I. (2019), Conditional failure occurrence rates for semi- Markov chains, *J Appl Stat*, https://doi.org/10.1080/02664763.2019.1610164.
- Yeh, L. (1997), The rate of occurrence of failures, *J. Appl. Probab.*, 34, 234–247.

Thank you