Semi-parametric inference for effective age models

Eric Beutner, Maastricht University

Joint work with Laurent Bordes and Laurent Doyen

Troyes, January 29

Overview

Introduction /Recap

Inference hazard and link function

Inference hazard and effective age function

Introduction /Recap

Recurrent events

- Some examples of recurrent events:
  - Re-occurrence of a tumour after surgical removal in cancer studies;
  - Migraines;
  - Outbreak of a disease;
  - Failure of a mechanical or electronic system;
  - Discovery of a bug in an operating system software or of an error in a scientific article.

Recurrent events

- Some examples of recurrent events:
  - Re-occurrence of a tumour after surgical removal in cancer studies;
  - Migraines;
  - Outbreak of a disease;
  - Failure of a mechanical or electronic system;
  - Discovery of a bug in an operating system software or of an error in a scientific article.
- Monitoring an observational unit (e.g. patient, mechanical system) during the time interval $[0, \tau]$, the data consist of:
  - The times $T_1, T_2, \ldots,$ between successive event occurrences;
  - The number of event occurrences $K := \max\{k \in \mathbb{N}_0 | S_k \leq \tau\}$, where $S_k = \sum_{\ell=1}^{k} T_\ell$;
  - Additional covariates $\mathbf{Z}$.

Counting process formulation

- Alternatively, the information at time t can be represented by

$$\{\mathbf{Z}, \{N(s), 0 \leq s \leq t\}, \{Y(s), 0 \leq s \leq t\}\},$$

where

  - $N(s) = \sum_{\ell=1}^{\infty} \mathbb{1}_{\{S_\ell \leq s\}}$;
  - $Y(s) = \mathbb{1}\{s \leq \tau\}$;
  - Z covariates as above.

A probabilistic model

Hollander & Peña (2004) introduced the following model:

A probabilistic model

Hollander & Peña (2004) introduced the following model:

- $N = \{N(s) | s \in [0, \tau]\}$ point process;

A probabilistic model

Hollander & Peña (2004) introduced the following model:

- $N = \{N(s) | s \in [0, \tau]\}$ point process;
- It is assumed that compensator A of N given the covariates $\mathbf{Z}$ equals

$$A(t|\mathbf{Z}) = \int_0^t Y_i(s) \lambda_0 \left(\varepsilon(s, \omega) \right) \psi(\boldsymbol{\beta}_0' \mathbf{Z}) \, ds,$$

where

A probabilistic model

Hollander & Peña (2004) introduced the following model:

- $N = \{N(s)|s \in [0, \tau]\}$ point process;
- It is assumed that compensator A of N given the covariates $\mathbf{Z}$ equals

$$A(t|\mathbf{Z}) = \int_0^t Y_i(s)\lambda_0\left(\varepsilon(s,\omega)\right)\psi(\boldsymbol{\beta}_0'\mathbf{Z})\, ds,$$

  where

  - ψ is a known link function;

A probabilistic model

Hollander & Peña (2004) introduced the following model:

- $N = \{N(s)|s \in [0, \tau]\}$ point process;
- It is assumed that compensator A of N given the covariates $\mathbf{Z}$ equals

$$A(t|\mathbf{Z}) = \int_0^t Y_i(s)\lambda_0\left(\varepsilon(s,\omega)\right)\psi(\boldsymbol{\beta}_0'\mathbf{Z})\,ds,$$

  where

  - ψ is a known link function;
  - $\boldsymbol{\beta}_0$ is a (unknown) parameter vector;

A probabilistic model

Hollander & Peña (2004) introduced the following model:

- $N = \{N(s) | s \in [0, \tau]\}$ point process;
- It is assumed that compensator A of N given the covariates $\mathbf{Z}$ equals

$$A(t|\mathbf{Z}) = \int_0^t Y_i(s)\lambda_0\left(\varepsilon(s,\omega)\right)\psi(\boldsymbol{\beta}_0'\mathbf{Z})\,ds,$$

  where

  - ψ is a known link function;
  - $\boldsymbol{\beta}_0$ is a (unknown) parameter vector;
  - $\varepsilon(s, \omega)$ is a possibly random function that describes the effective age at time s;

A probabilistic model

Hollander & Peña (2004) introduced the following model:

- $N = \{N(s) | s \in [0, \tau]\}$ point process;
- It is assumed that compensator A of N given the covariates $\mathbf{Z}$ equals

$$A(t|\mathbf{Z}) = \int_0^t Y_i(s) \lambda_0 \left(\varepsilon(s, \omega) \right) \psi(\boldsymbol{\beta}_0' \mathbf{Z}) \, ds,$$

  where

  - ψ is a known link function;
  - $\boldsymbol{\beta}_0$ is a (unknown) parameter vector;
  - $\varepsilon(s, \omega)$ is a possibly random function that describes the effective age at time s;
  - λ_0 is an unknown hazard rate function.

Examples

- Taking $\psi(u) = \exp(u)$ and $\varepsilon(s, \omega) = s$ leads to the Cox model.

Examples

- Taking $\psi(u) = \exp(u)$ and $\varepsilon(s, \omega) = s$ leads to the Cox model.
- Taking $\psi(u) = 1$ and $\varepsilon(s, \omega) = s - S_{N(s-,\omega)}$ leads i.i.d. inter-occurrence times, i.e., renewal process.

Examples

- Taking $\psi(u) = \exp(u)$ and $\varepsilon(s, \omega) = s$ leads to the Cox model.
- Taking $\psi(u) = 1$ and $\varepsilon(s, \omega) = s - S_{N(s-,\omega)}$ leads i.i.d. inter-occurrence times, i.e., renewal process.
- Taking $\psi(u) = 1$ and $\varepsilon(s, \omega) = s - \theta S_{k-1}(\omega)$ on $(S_{k-1}(\omega), S_k(\omega)]$ leads to an Arithmetic Reduction Age model of Type 1 (ARA_1).
  - For $\theta = 0$ we have $\varepsilon(s, \omega) = s \Rightarrow$ Poisson process for which effective age = calendar time;
  - For $\theta = 1$ we have $\varepsilon(s, \omega) = s - S_{k-1}(\omega) \Rightarrow$ renewal process;
  - For $\theta \in (0, 1) \Rightarrow$ imperfect repair;
  - Gonzales et al. (2005) applied the model with $\theta \in \{0, 0.5, 1\}$ to the response of patients suffering from a non-curable cancer to a therapy.

Semi-parametric inference on hazard rate function and link function

Inference

- Let (N_i, Y_i, Z_i), $1 \leq i \leq m$, be m copies of (N, Y, Z).

Inference

- Let (N_i, Y_i, Z_i), $1 \leq i \leq m$, be m copies of (N, Y, Z).
- For inference for the above model recall that an event at calendar time s is in general not "caused by" $\lambda(s)$ $\Rightarrow$ Keep track of both time scales.

Inference

- Let (N_i, Y_i, Z_i), $1 \leq i \leq m$, be m copies of (N, Y, Z).
- For inference for the above model recall that an event at calendar time s is in general not "caused by" $\lambda(s)$ $\Rightarrow$ Keep track of both time scales.
- To do so Peña et al. (2007) introduced double indexed stochastic processes:

$$N_i^d(s,t) = \int_0^s H_i(v,t)\, dN_i(v), \text{ and}$$

$$A_i^d(s,t) = \int_0^s H_i(v,t)\, dA_i(v),$$

where $H_i(s,t) = \mathbb{1}_{\{\epsilon_i(s) \leq t\}}$.

Inference

- Let (N_i, Y_i, Z_i), $1 \leq i \leq m$, be m copies of (N, Y, Z).
- For inference for the above model recall that an event at calendar time s is in general not "caused by" $\lambda(s) \Rightarrow$ Keep track of both time scales.
- To do so Peña et al. (2007) introduced double indexed stochastic processes:

$$N_i^d(s,t) = \int_0^s H_i(v,t)\, dN_i(v), \text{ and}$$

$$A_i^d(s,t) = \int_0^s H_i(v,t)\, dA_i(v),$$

where $H_i(s,t) = \mathbb{1}_{\{\epsilon_i(s) \leq t\}}$.

- $H_i(s,t)$ indicates whether at calendar time s the age is at most t.

Inference

- Let (N_i, Y_i, Z_i), $1 \le i \le m$, be m copies of (N, Y, Z).
- For inference for the above model recall that an event at calendar time s is in general not "caused by" $\lambda(s) \Rightarrow$ Keep track of both time scales.
- To do so Peña et al. (2007) introduced double indexed stochastic processes:

$$N_i^d(s,t) = \int_0^s H_i(v,t)\, dN_i(v), \text{ and}$$

$$A_i^d(s,t) = \int_0^s H_i(v,t)\, dA_i(v),$$

where $H_i(s,t) = \mathbb{1}_{\{\epsilon_i(s) \le t\}}$.

- $H_i(s,t)$ indicates whether at calendar time s the age is at most t.
- $N_i^d(s,t)$ gives the number of events during 0 and s with effective age at most t.

Inference (cont'd)

- The difference between N_i^d and A_i^d equals

$$M_i^d(s,t) = \int_0^s H_i(v,t)\, dM_i(v) \text{ with } M_i = N_i - A_i.$$

Notice that M_i^d is not directly amenable to inference on λ_0, as it involves the time-transformed λ_0, i.e. $\lambda_0 \circ \epsilon$.

Inference (cont'd)

- The difference between N_i^d and A_i^d equals

$$M_i^d(s,t) = \int_0^s H_i(v,t)\, dM_i(v) \text{ with } M_i = N_i - A_i.$$

  Notice that M_i^d is not directly amenable to inference on λ_0, as it involves the time-transformed λ_0, i.e. $\lambda_0 \circ \epsilon$.
- Solution: De-couple λ_0 and ϵ.

Inference (cont'd)

- The difference between N_i^d and A_i^d equals

$$M_i^d(s,t) = \int_0^s H_i(v,t)\, dM_i(v) \text{ with } M_i = N_i - A_i.$$

Notice that M_i^d is not directly amenable to inference on λ_0, as it involves the time-transformed λ_0, i.e. $\lambda_0 \circ \epsilon$.

- Solution: De-couple λ_0 and ϵ.
- Applying a change of variable leads to

$$M_i^*(s,t) = N_i^*(s,t) - \int_0^t Y_i^d(s,u,\boldsymbol{\beta})\, d\Lambda(u),$$

where

$$Y_i^d(s,t,\boldsymbol{\beta}) = \text{size of risk set at calendar time } s \text{ with age } t$$

is a 'time-transformed' at risk process.

Inference (cont'd.)

- For a given $\boldsymbol{\beta}$ the above representation suggests the following method of moment estimator for $\Lambda_0(t)$ at calendar time s:

$$\widehat{\Lambda}(t|s,\boldsymbol{\beta}) = \int_0^t \left(J(s,t) \Big/ \sum_{i=1}^m Y_i^d(s,u,\boldsymbol{\beta}) \right) \left[\sum_{i=1}^m N_i^d(s,du) \right],$$

where $J(s,t) = 1$ if $\sum_{i=1}^m Y_i^d(s,u,\boldsymbol{\beta}) > 0$ and zero otherwise.

Inference (cont'd.)

- For a given $\boldsymbol{\beta}$ the above representation suggests the following method of moment estimator for $\Lambda_0(t)$ at calendar time s:

$$\widehat{\Lambda}(t|s,\boldsymbol{\beta}) = \int_0^t \left(J(s,t) / \sum_{i=1}^m Y_i^d(s,u,\boldsymbol{\beta}) \right) \left[\sum_{i=1}^m N_i^d(s,du) \right],$$

  where $J(s,t) = 1$ if $\sum_{i=1}^m Y_i^d(s,u,\boldsymbol{\beta}) > 0$ and zero otherwise.
- Inserting $\widehat{\Lambda}(t|s,\boldsymbol{\beta})$ in the full likelihood Peña et al. (2007) obtain a profile likelihood function for estimating $\boldsymbol{\beta}$.

Inference (cont'd.)

- For a given $\boldsymbol{\beta}$ the above representation suggests the following method of moment estimator for $\Lambda_0(t)$ at calendar time s:

$$\widehat{\Lambda}(t|s,\boldsymbol{\beta}) = \int_0^t \left(J(s,t) / \sum_{i=1}^{m} Y_i^d(s,u,\boldsymbol{\beta}) \right) \left[\sum_{i=1}^{m} N_i^d(s,du) \right],$$

  where $J(s,t) = 1$ if $\sum_{i=1}^{m} Y_i^d(s,u,\boldsymbol{\beta}) > 0$ and zero otherwise.
- Inserting $\widehat{\Lambda}(t|s,\boldsymbol{\beta})$ in the full likelihood Peña et al. (2007) obtain a profile likelihood function for estimating $\boldsymbol{\beta}$.
- Recall that, for instance, in the Cox model this leads to consistent and asymptotically normally distributed estimates.

Results

- Dorado et al. (1997) weak convergence results for $\Lambda_0 := \int \lambda_0(u)\,du$ for a model slightly more general than ARA_1.
- Gärtner (2003) also weak convergence results for the same model but different data collection process.
- Adekpedjou and Stocker (2015) weak convergence results for $\Lambda_0 := \int \lambda_0(u)\,du$ and β for an ARA_1-type model.
- Very recently Peña (2014) obtained weak convergence results for $\Lambda_0 := \int \lambda_0(u)\,du$ and β without restricting the effective age function.

Results

- Dorado et al. (1997) weak convergence results for $\Lambda_0 := \int \lambda_0(u)\,du$ for a model slightly more general than ARA_1.
- Gärtner (2003) also weak convergence results for the same model but different data collection process.
- Adekpedjou and Stocker (2015) weak convergence results for $\Lambda_0 := \int \lambda_0(u)\,du$ and β for an ARA_1-type model.
- Very recently Peña (2014) obtained weak convergence results for $\Lambda_0 := \int \lambda_0(u)\,du$ and β without restricting the effective age function.
- In these articles it is assumed that the effective age function is entirely known $\Rightarrow$ the way the interventions influence the effective age must be known by the statistician.

Semi-parametric inference for **hazard rate function** and **effective age function**

Introduction

- Already seen: Models where the age function ϵ depends on a parameter ($\epsilon = s - \theta S_{k-1}$).

Introduction

- Already seen: Models where the age function ϵ depends on a parameter ($\epsilon = s - \theta S_{k-1}$).
- Assuming, for instance, θ to be unknown it is tempting to use the same inferential procedure.

Introduction

- Already seen: Models where the age function ϵ depends on a parameter ($\epsilon = s - \theta S_{k-1}$).
- Assuming, for instance, θ to be unknown it is tempting to use the same inferential procedure.
- However, notice this model does not fit directly into the above inferential procedure, because unknown are (θ_0, λ_0) where θ_0 unknown parameter of age function ϵ in contrast to $\boldsymbol{\beta}$ which is an unknown parameter of link function ψ.

Introduction

- Already seen: Models where the age function ϵ depends on a parameter ($\epsilon = s - \theta S_{k-1}$).
- Assuming, for instance, θ to be unknown it is tempting to use the same inferential procedure.
- However, notice this model does not fit directly into the above inferential procedure, because unknown are (θ_0, λ_0) where θ_0 unknown parameter of age function ϵ in contrast to $\boldsymbol{\beta}$ which is an unknown parameter of link function ψ.

Model

- Let $N = \{N(s) | s \in [0, \tau]\}$ be a counting process.
- It is assumed that the compensator A of N is given by

$$A(t) = \int_0^t Y(s)\lambda(\varepsilon^{\boldsymbol{\theta}}(s))\, ds.$$

Model

- Let $N = \{N(s) | s \in [0, \tau]\}$ be a counting process.
- It is assumed that the compensator A of N is given by

$$A(t) = \int_0^t Y(s)\lambda(\varepsilon^{\boldsymbol{\theta}}(s))\, ds.$$

- Moreover, for every $\boldsymbol{\theta} \in \Theta$, $\Theta \subset \mathbb{R}^d$, we have that the process $\varepsilon^{\boldsymbol{\theta}} = \{\epsilon^{\boldsymbol{\theta}}(s), 0 \leq s \leq \tau\}$ fulfils
  - $\varepsilon^{\boldsymbol{\theta}}(0, \omega) = c_0$ a.s. for some $c_0 \in \mathbb{R}_+$;
  - $s \to \varepsilon^{\boldsymbol{\theta}}(s, \omega)$ is a.s. non-negative;
  - $s \to \varepsilon^{\boldsymbol{\theta}}(s, \omega)$ is a.s. continuous and increasing on $(S_{k-1}, S_k]$, $k \in \mathbb{N}$.

Model

- Let $N = \{N(s) | s \in [0, \tau]\}$ be a counting process.
- It is assumed that the compensator A of N is given by

$$A(t) = \int_0^t Y(s)\lambda(\varepsilon^{\boldsymbol{\theta}}(s))\, ds.$$

- Moreover, for every $\boldsymbol{\theta} \in \Theta$, $\Theta \subset \mathbb{R}^d$, we have that the process $\varepsilon^{\boldsymbol{\theta}} = \{\epsilon^{\boldsymbol{\theta}}(s), 0 \leq s \leq \tau\}$ fulfils
  - $\varepsilon^{\boldsymbol{\theta}}(0, \omega) = c_0$ a.s. for some $c_0 \in \mathbb{R}_+$;
  - $s \to \varepsilon^{\boldsymbol{\theta}}(s, \omega)$ is a.s. non-negative;
  - $s \to \varepsilon^{\boldsymbol{\theta}}(s, \omega)$ is a.s. continuous and increasing on $(S_{k-1}, S_k]$, $k \in \mathbb{N}$.
- Example ARA$_1$: Then $\Theta = [0, 1]$ and $\varepsilon^{\theta}(s, \omega) = s - \theta S_{k-1}(\omega)$ on $(S_{k-1}(\omega), S_k(\omega)]$.

Profile likelihood

- Let $N_1, \ldots, N_m$ be m independent copies of N. Then full likelihood equals $L_{m,F}(s|\lambda, \varepsilon^{\boldsymbol{\theta}}, \mathbf{D}_m(s))$

$$\prod_{i=1}^{m} \prod_{u=0}^{s} \left[Y_i(u)\lambda(\varepsilon_i^{\boldsymbol{\theta}}(u))\right]^{N_i(\Delta u)} \exp\left[-\sum_{i=1}^{m} \int_0^s Y_i(u)\lambda(\varepsilon_i^{\boldsymbol{\theta}}(u))\, du\right]$$

$$= \prod_{i=1}^{m} \prod_{u=0}^{s} \left[Y_i(u)\lambda(\varepsilon_i^{\boldsymbol{\theta}}(u))\right]^{N_i(\Delta u)} \exp\left[-\int_0^\infty S_m^{\boldsymbol{\theta}}(s,u)\, d\Lambda(u)\right],$$

where $\mathbf{D}_m(s)$ denotes the data at time s and

$$\mathcal{S}_m^{\boldsymbol{\theta}}(s,t) := \sum_{i=1}^{m} \sum_{j=1}^{N_i(s-)} \gamma_{i,j-1}^{\boldsymbol{\theta}}(t) \cdot \mathbb{1}_{(\varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j-1}+), \varepsilon_{i,j-1}^{\boldsymbol{\theta}}(S_{i,j})]}(t)$$
$$+ \sum_{i=1}^{m} \gamma_{i,N_i(s-)}^{\boldsymbol{\theta}}(t) \cdot \mathbb{1}_{(\varepsilon_{i,N_i(s-)}^{\boldsymbol{\theta}}(S_{i,N_i(s-)}+), \varepsilon_{i,N_i(s-)}^{\boldsymbol{\theta}}(s\wedge\tau_i)]}(t).$$

Profile likelihood (cont'd.)

- To profile out the infinite-dimensional parameter we use the method-of-moment estimator proposed by Peña et al. that equals here for fixed $\boldsymbol{\theta}$

$$\widehat{\Lambda}_m(s,t|\boldsymbol{\theta}) := \int_0^t \frac{J_m^{\boldsymbol{\theta}}(s,u)}{S_m^{\boldsymbol{\theta}}(s,u)} \left[\sum_{i=1}^m N_i^{d,\boldsymbol{\theta}}(s,du) \right].$$

Profile likelihood (cont'd.)

- To profile out the infinite-dimensional parameter we use the method-of-moment estimator proposed by Peña et al. that equals here for fixed $\boldsymbol{\theta}$

$$\widehat{\Lambda}_m(s,t|\boldsymbol{\theta}) := \int_0^t \frac{J_m^{\boldsymbol{\theta}}(s,u)}{S_m^{\boldsymbol{\theta}}(s,u)} \left[\sum_{i=1}^m N_i^{d,\boldsymbol{\theta}}(s,du)\right].$$

- Worth mentioning that $\widehat{\Lambda}_m$ can be justified as NPMLE.
- Hence full likelihood after plugging in $\widehat{\Lambda}_m$ can be considered profile likelihood function

$$\ell_{m,P}(s|\boldsymbol{\theta},\widehat{\Lambda}_m,\mathbf{D}_m(s)) = -\int_0^s \sum_{i=1}^m \log\left(\mathcal{S}_m(s,\varepsilon_i^{\boldsymbol{\theta}}(w))\right)\, dN_i(w),$$

Preparation for main theorem

- To motivate the main result consider log profile likelihood

$$-\int_0^s \sum_{i=1}^{m} \log\left(\mathcal{S}_m^{\boldsymbol{\theta}}(s, \varepsilon_i^{\boldsymbol{\theta}}(w))\right)\, dN_i(w),$$

for an ARA_1 model and $\theta = 0$ and $\theta = 1$, respectively.

Preparation for main theorem

- To motivate the main result consider log profile likelihood

$$-\int_0^s \sum_{i=1}^{m} \log\left(\mathcal{S}_m^{\boldsymbol{\theta}}(s, \varepsilon_i^{\boldsymbol{\theta}}(w))\right)\, dN_i(w),$$

  for an ARA_1 model and $\theta = 0$ and $\theta = 1$, respectively.
- Take $m = 2$, let both samples be Type-II censored and consider arbitrary event times $s_{1,1}, \ldots, s_{1,n_1}$ and $s_{2,1}, \ldots, s_{2,n_2}$.

Preparation for main theorem

- To motivate the main result consider log profile likelihood

$$-\int_0^s \sum_{i=1}^m \log\left(\mathcal{S}_m^{\boldsymbol{\theta}}(s, \varepsilon_i^{\boldsymbol{\theta}}(w))\right)\, dN_i(w),$$

  for an ARA_1 model and $\theta = 0$ and $\theta = 1$, respectively.
- Take $m = 2$, let both samples be Type-II censored and consider arbitrary event times $s_{1,1}, \ldots, s_{1,n_1}$ and $s_{2,1}, \ldots, s_{2,n_2}$.
- Then with $s = \max\{s_{1,n_1}, s_{2,n_2}\}$ the function $\sum_{i=1}^2 \mathcal{S}_2^0(s, \cdot)$ equals:

$$\mathbb{1}_{(0,s_{1,1}]}(\cdot) + \mathbb{1}_{(s_{1,1},s_{1,2}]}(\cdot) + \ldots + \mathbb{1}_{(s_{1,n_1-1},s_{1,n_1}]}(\cdot) + \ldots + \mathbb{1}_{(s_{2,n_2-1},s_{2,n_2}]}(\cdot)$$

Preparation for main theorem

- To motivate the main result consider log profile likelihood

$$-\int_0^s \sum_{i=1}^{m} \log\left(\mathcal{S}_m^{\boldsymbol{\theta}}(s, \varepsilon_i^{\boldsymbol{\theta}}(w))\right)\, dN_i(w),$$

  for an ARA_1 model and $\theta = 0$ and $\theta = 1$, respectively.
- Take $m = 2$, let both samples be Type-II censored and consider arbitrary event times $s_{1,1}, \ldots, s_{1,n_1}$ and $s_{2,1}, \ldots, s_{2,n_2}$.
- Then with $s = \max\{s_{1,n_1}, s_{2,n_2}\}$ the function $\sum_{i=1}^{2} \mathcal{S}_2^{0}(s, \cdot)$ equals:

$$\mathbb{1}_{(0,s_{1,1}]}(\cdot) + \mathbb{1}_{(s_{1,1},s_{1,2}]}(\cdot) + \ldots + \mathbb{1}_{(s_{1,n_1-1},s_{1,n_1}]}(\cdot) + \ldots + \mathbb{1}_{(s_{2,n_2-1},s_{2,n_2}]}(\cdot)$$

  whereas the function $\sum_{i=1}^{2} \mathcal{S}_2^{1}(s, \cdot)$ equals

$$\mathbb{1}_{(0,s_{1,1}]}(\cdot) + \mathbb{1}_{(0,s_{1,2}]}(\cdot) + \ldots + \mathbb{1}_{(0,s_{1,n_1}]}(\cdot) + \ldots + \mathbb{1}_{(0,s_{2,n_2}]}(\cdot).$$

Main result part (a)

Theorem: Let $(\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m$ the m-fold product measure of $\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0}$ and the samples Type-II censored.
(a) Denote by $A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$ the set of all ω's such that for all pairs (i,j), $1 \le i \le m$, $1 \le j \le J_i(s^*)$, and all pairs (k,ℓ) $1 \le k \le m$, $1 \le \ell \le J_k(s^*)$, we have that

$$\varepsilon^{\boldsymbol{\theta}}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\boldsymbol{\theta}}_{k,\ell-1}(S_{k,\ell}(\omega))$$

implies that

$$\varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}(\omega)) < \varepsilon^{\tilde{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega)).$$

Then we have

$$\begin{aligned} & (\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m \left(\ell_{P,m}\left(s^*|\boldsymbol{\theta},\widehat{\Lambda}_m,\mathbf{D}_m(s)\right) \ge \ell_{P,m}\left(s^*|\tilde{\boldsymbol{\theta}},\widehat{\Lambda}_m,\mathbf{D}_m(s)\right)\right) \\ \ge \quad & (\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m \left(A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}\right). \end{aligned}$$

Remarks on part (a)

- The theorem provides a tool for pairwise comparison of the profile likelihood function.

Remarks on part (a)

- The theorem provides a tool for pairwise comparison of the profile likelihood function.
- Intuitively, the condition leads to the result, because
  - $\mathcal{S}_m^{\boldsymbol{\theta}}$ consists (roughly) of indicator functions of the form

$$\mathbb{1}_{(\varepsilon^{\boldsymbol{\theta}}_{i,j-1}(S_{i,j-1}+),\infty)}(\cdot)$$

    and $\mathcal{S}_m^{\tilde{\boldsymbol{\theta}}}$ is of the form

$$\mathbb{1}_{(\varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}+),\infty)}(\cdot).$$

Remarks on part (a)

- The theorem provides a tool for pairwise comparison of the profile likelihood function.
- Intuitively, the condition leads to the result, because
  - $\mathcal{S}_m^{\boldsymbol{\theta}}$ consists (roughly) of indicator functions of the form

$$\mathbb{1}_{(\varepsilon^{\boldsymbol{\theta}}_{i,j-1}(S_{i,j-1}+),\infty)}(\cdot)$$

    and $\mathcal{S}_m^{\tilde{\boldsymbol{\theta}}}$ is of the form

$$\mathbb{1}_{(\varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}+),\infty)}(\cdot).$$

  - We evaluate these indicators at

$$\varepsilon^{\boldsymbol{\theta}}_{k,\ell-1}(S_{k,\ell}(\omega)) \text{ and } \varepsilon^{\tilde{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega)), \text{ respectively}.$$

Main result part (b)

Theorem (cont'd) (b) Denote by $B_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$ the set of all $\omega \in A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$ for which we additionally have that there are at least two pairs $(\underline{i}, \underline{j})$, $1 \le \underline{i} \le m$, $1 \le \underline{j} \le J_{\underline{i}}(s^*)$, and $(\underline{k}, \underline{\ell})$, $1 \le \underline{k} \le m$, $1 \le \underline{\ell} \le J_{\underline{k}}(s^*)$, such that

$$\varepsilon^{\tilde{\boldsymbol{\theta}}}_{\underline{i},\underline{j}-1}(S_{\underline{i},\underline{j}-1}(\omega)) < \varepsilon^{\tilde{\boldsymbol{\theta}}}_{\underline{k},\underline{\ell}-1}(S_{\underline{k},\underline{\ell}}(\omega))$$

but

$$\varepsilon^{\boldsymbol{\theta}}_{\underline{i},\underline{j}-1}(S_{\underline{i},\underline{j}-1}(\omega)) \ge \varepsilon^{\boldsymbol{\theta}}_{\underline{k},\underline{\ell}-1}(S_{\underline{k},\underline{\ell}}(\omega)).$$

Then we have

$$(\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m \left(\ell_{P,m}\left(s^*|\boldsymbol{\theta}, \widehat{\Lambda}_m, \mathbf{D}_m(s)\right) > \ell_{P,m}\left(s^*|\tilde{\boldsymbol{\theta}}, \widehat{\Lambda}_m, \mathbf{D}_m(s)\right)\right)$$
$$\ge (\mathbb{P}^{\lambda_0,\boldsymbol{\theta}_0})^m \left(B_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}\right).$$

Remarks on part (b)

- As in part (a) we get a tool for pairwise comparison of the profile likelihood function.

Remarks on part (b)

- As in part (a) we get a tool for pairwise comparison of the profile likelihood function.
- We only consider $\omega \in A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$. For those ω we already know that $\ell_{P,m}$ at $\boldsymbol{\theta}$ is at least as large as at $\tilde{\boldsymbol{\theta}}$

Remarks on part (b)

- As in part (a) we get a tool for pairwise comparison of the profile likelihood function.
- We only consider $\omega \in A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$. For those ω we already know that $\ell_{P,m}$ at $\boldsymbol{\theta}$ is at least as large as at $\tilde{\boldsymbol{\theta}}$
- Intuitively, the condition leads to the result, because
  - $\mathcal{S}_m^{\boldsymbol{\theta}}$ and $\mathcal{S}_m^{\tilde{\boldsymbol{\theta}}}$ consists (roughly) of indicator functions of the form

$$\mathbb{1}_{(\varepsilon^{\boldsymbol{\theta}}_{i,j-1}(S_{i,j-1}+),\infty)}(\cdot) \text{ and } \mathbb{1}_{(\varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}+),\infty)}(\cdot), \text{ respectively.}$$

Remarks on part (b)

- As in part (a) we get a tool for pairwise comparison of the profile likelihood function.
- We only consider $\omega \in A_{m,\boldsymbol{\theta},\tilde{\boldsymbol{\theta}}}$. For those ω we already know that $\ell_{P,m}$ at $\boldsymbol{\theta}$ is at least as large as at $\tilde{\boldsymbol{\theta}}$
- Intuitively, the condition leads to the result, because
  - $\mathcal{S}_m^{\boldsymbol{\theta}}$ and $\mathcal{S}_m^{\tilde{\boldsymbol{\theta}}}$ consists (roughly) of indicator functions of the form

$$\mathbb{1}_{(\varepsilon^{\boldsymbol{\theta}}_{i,j-1}(S_{i,j-1}+),\infty)}(\cdot) \text{ and } \mathbb{1}_{(\varepsilon^{\tilde{\boldsymbol{\theta}}}_{i,j-1}(S_{i,j-1}+),\infty)}(\cdot), \text{ respectively}.$$

  - We evaluate these indicators at

$$\varepsilon^{\boldsymbol{\theta}}_{k,\ell-1}(S_{k,\ell}(\omega)) \text{ and } \varepsilon^{\tilde{\boldsymbol{\theta}}}_{k,\ell-1}(S_{k,\ell}(\omega)), \text{ respectively}.$$

Remarks on proof

- Only difficulty of the proof: To make the "roughly" precise.

Remarks on proof

- Only difficulty of the proof: To make the "roughly" precise.
- Anyhow, it can be entirely based on the following simple facts:
- **Fact 1:** Let $I_1 \subset J$ and $I_2 \subset J$ with J finite, $I_1 \neq J$, $|I_1| = |I_2|$ and $\exists i_1 \in I_1$ such that $i_1 \notin I_2$. Then $\exists i_2 \in J$ such that $i_2 \in I_2$, but $i_2 \notin I_1$.

Remarks on proof

- Only difficulty of the proof: To make the "roughly" precise.
- Anyhow, it can be entirely based on the following simple facts:
- **Fact 1:** Let $I_1 \subset J$ and $I_2 \subset J$ with J finite, $I_1 \neq J$, $|I_1| = |I_2|$ and $\exists i_1 \in I_1$ such that $i_1 \notin I_2$. Then $\exists i_2 \in J$ such that $i_2 \in I_2$, but $i_2 \notin I_1$.

- **Fact 2:** Let $x_i \in \mathbb{R}_+$, $y_i \in \mathbb{R}_+$, $\tilde{x}_i \in \mathbb{R}_+$, $\tilde{y}_i \in \mathbb{R}_+$, $1 \leq i \leq I$. Let $G, \tilde{G} : \{1, \ldots, I\} \to \mathbb{N}$ be defined by

$$G(j) := \sum_{i=1}^{I} \mathbb{1}_{(x_i,\infty)}(y_j) \text{ and } \tilde{G}(j) := \sum_{i=1}^{I} \mathbb{1}_{(\tilde{x}_i,\infty)}(\tilde{y}_j).$$

Then

(i) If $\forall i \in \{1, \ldots, I\}$: $x_i < y_j \Rightarrow \tilde{x}_i < \tilde{y}_j$, then $G(j) \leq \tilde{G}(j)$.

(ii) If additionally $\exists \underline{i} \in \{1, \ldots, I\}$ such that $\tilde{x}_{\underline{i}} < \tilde{y}_j$ but $x_{\underline{i}} \geq y_j$ then $G(j) < \tilde{G}(j)$.

Consistency

- Main result not immediately a tool to show inconsistency of profile likelihood method. Need some kind of uniformity.

Consistency

- Main result not immediately a tool to show inconsistency of profile likelihood method. Need some kind of uniformity.
- **Corollary:** Denote by $B(\boldsymbol{\theta}_0, \epsilon)$ an ϵ-ball around $\boldsymbol{\theta}_0$ and assume that $\boldsymbol{\theta}$ is such that for some $m' \in \mathbb{N}$ we have for all $m \geq m'$ that $(\mathbb{P}^{\lambda_0, \boldsymbol{\theta}_0})^m \left(B_{m, \boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}}\right) \geq c$, $c > 0$, $\forall \tilde{\boldsymbol{\theta}} \in B(\boldsymbol{\theta}_0, \epsilon)$. Then

$$\hat{\boldsymbol{\theta}}_m \overset{\mathbb{P}}{\nrightarrow} \boldsymbol{\theta}_0, \text{ as } m \to \infty.$$

Example ARA_1

- No guarantee that the main results can be applied to well-known effective age models.

Example ARA_1

- No guarantee that the main results can be applied to well-known effective age models.
- For an ARA_1 condition (a) reads as

$$\begin{aligned} S_{i,j-1}(\omega) - \theta S_{i,j-1}(\omega) &< S_{k,\ell}(\omega) - \theta S_{k,\ell-1}(\omega) \\ \Rightarrow S_{i,j-1}(\omega) - \tilde{\theta} S_{i,j-1}(\omega) &< S_{k,\ell}(\omega) - \tilde{\theta} S_{k,\ell-1}(\omega). \end{aligned}$$

Example ARA_1

- No guarantee that the main results can be applied to well-known effective age models.
- For an ARA_1 condition (a) reads as

$$\begin{aligned} S_{i,j-1}(\omega) - \theta S_{i,j-1}(\omega) &< S_{k,\ell}(\omega) - \theta S_{k,\ell-1}(\omega) \\ \Rightarrow S_{i,j-1}(\omega) - \tilde{\theta} S_{i,j-1}(\omega) &< S_{k,\ell}(\omega) - \tilde{\theta} S_{k,\ell-1}(\omega). \end{aligned}$$

For $0 \leq \theta < \tilde{\theta} \leq 1$, $x \in \mathbb{R}_+$, $y \in \mathbb{R}_+$ and $z \in \mathbb{R}_+$ with $y < z$ we have by linearity

$$x - \theta x < z - \theta y \Rightarrow x - \tilde{\theta} x < z - \tilde{\theta} y.$$

Example ARA$_1$

- No guarantee that the main results can be applied to well-known effective age models.
- For an ARA$_1$ condition (a) reads as

$$\begin{aligned} S_{i,j-1}(\omega) - \theta S_{i,j-1}(\omega) &< S_{k,\ell}(\omega) - \theta S_{k,\ell-1}(\omega) \\ \Rightarrow S_{i,j-1}(\omega) - \tilde{\theta} S_{i,j-1}(\omega) &< S_{k,\ell}(\omega) - \tilde{\theta} S_{k,\ell-1}(\omega). \end{aligned}$$

For $0 \leq \theta < \tilde{\theta} \leq 1$, $x \in \mathbb{R}_+$, $y \in \mathbb{R}_+$ and $z \in \mathbb{R}_+$ with $y < z$ we have by linearity

$$x - \theta x < z - \theta y \Rightarrow x - \tilde{\theta} x < z - \tilde{\theta} y.$$

- Hence, for $0 \leq \theta < \tilde{\theta} \leq 1$, any $m \in \mathbb{N}$ and any (λ_0, θ_0)

$$\left(\mathbb{P}^{\lambda_0,\theta_0}\right)^m \left(\ell_{P,m}\left(s^*|\theta, \widehat{\Lambda}_m, \mathbf{D}_m(s)\right) \geq \ell_{P,m}\left(s^*|\tilde{\theta}, \widehat{\Lambda}_m, \mathbf{D}_m(s)\right)\right) = 1.$$

Examples ARA$_1$ cont'd.

- Last result implies that $\ell_{m,P}$ is decreasing as function of θ. Could still be flat.

Examples ARA$_1$ cont'd.

- Last result implies that $\ell_{m,P}$ is decreasing as function of θ. Could still be flat.
- To check for strictly decreasing with positive probabilities we can use part (b).

Examples ARA_1 cont'd.

- Last result implies that $\ell_{m,P}$ is decreasing as function of θ. Could still be flat.
- To check for strictly decreasing with positive probabilities we can use part (b). For ARA_1 the condition reads as

$$S_{i,j-1} - \tilde{\theta} S_{i,j-1} < S_{k,\ell} - \tilde{\theta} S_{k,\ell-1}, \text{ but } S_{i,j-1} - \theta S_{i,j-1} \geq S_{k,\ell} - \theta S_{k,\ell-1}.$$

Examples ARA$_1$ cont'd.

- Last result implies that $\ell_{m,P}$ is decreasing as function of θ. Could still be flat.
- To check for strictly decreasing with positive probabilities we can use part (b). For ARA$_1$ the condition reads as

$$S_{i,j-1} - \tilde{\theta} S_{i,j-1} < S_{k,\ell} - \tilde{\theta} S_{k,\ell-1}, \text{ but } S_{i,j-1} - \theta S_{i,j-1} \geq S_{k,\ell} - \theta S_{k,\ell-1}.$$

- With $i = 1$, $j = 1$, $k = 2$ and $\ell = 2$ the above event has probability

$$\int_{\mathbb{R}^2} \left[F_{\lambda_0,\theta_0}^{S_{1,1}} \left(\frac{s_{2,2} - s_{2,1}}{1 - \tilde{\theta}} + s_{2,1} \right) - F_{\lambda_0,\theta_0}^{S_{1,1}} \left(\frac{s_{2,2} - s_{2,1}}{1 - \theta} + s_{2,1} \right) \right] dF_{\lambda_0,\theta_0}^{S_{2,2},S_{2,1}}(s_{2,2}, s_{2,1}).$$

- Take $[\theta_0 - \epsilon, \theta_0 + \epsilon]$ and $\theta < \theta_0 - \epsilon$. Then lower bound for $(\mathbb{P}^{\lambda_0,\theta_0})^m (B_{m,\theta,\theta_0-\epsilon})$ independent of m.

Example ARA$_\infty$

- For an ARA$_\infty$ condition (a) reads as

$$s_{i-1} - \theta \sum_{\ell=1}^{i-1} (1-\theta)^{i-1-\ell} s_\ell < \bar{s}_k - \theta \sum_{\ell=1}^{k-1} (1-\theta)^{k-1-\ell} \bar{s}_\ell$$

$$\Rightarrow s_{i-1} - \tilde{\theta} \sum_{\ell=1}^{i-1} (1-\tilde{\theta})^{i-1-\ell} s_\ell < \bar{s}_k - \tilde{\theta} \sum_{\ell=1}^{k-1} (1-\tilde{\theta})^{k-1-\ell} \bar{s}_\ell.$$

Example ARA$_\infty$

- For an ARA$_\infty$ condition (a) reads as

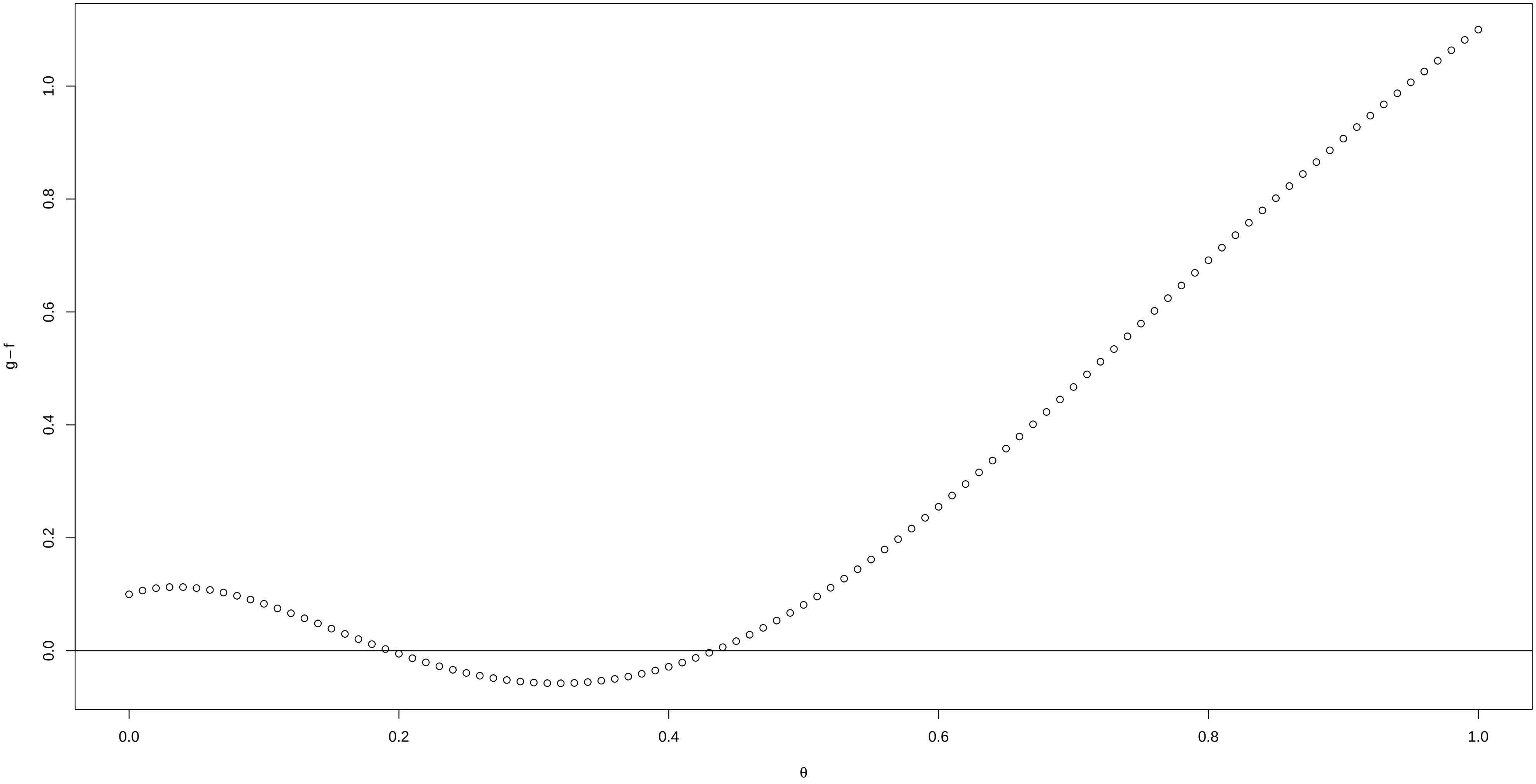
$$s_{i-1} - \theta \sum_{\ell=1}^{i-1} (1-\theta)^{i-1-\ell} s_\ell < \bar{s}_k - \theta \sum_{\ell=1}^{k-1} (1-\theta)^{k-1-\ell} \bar{s}_\ell$$

$$\Rightarrow s_{i-1} - \tilde{\theta} \sum_{\ell=1}^{i-1} (1-\tilde{\theta})^{i-1-\ell} s_\ell < \bar{s}_k - \tilde{\theta} \sum_{\ell=1}^{k-1} (1-\tilde{\theta})^{k-1-\ell} \bar{s}_\ell.$$

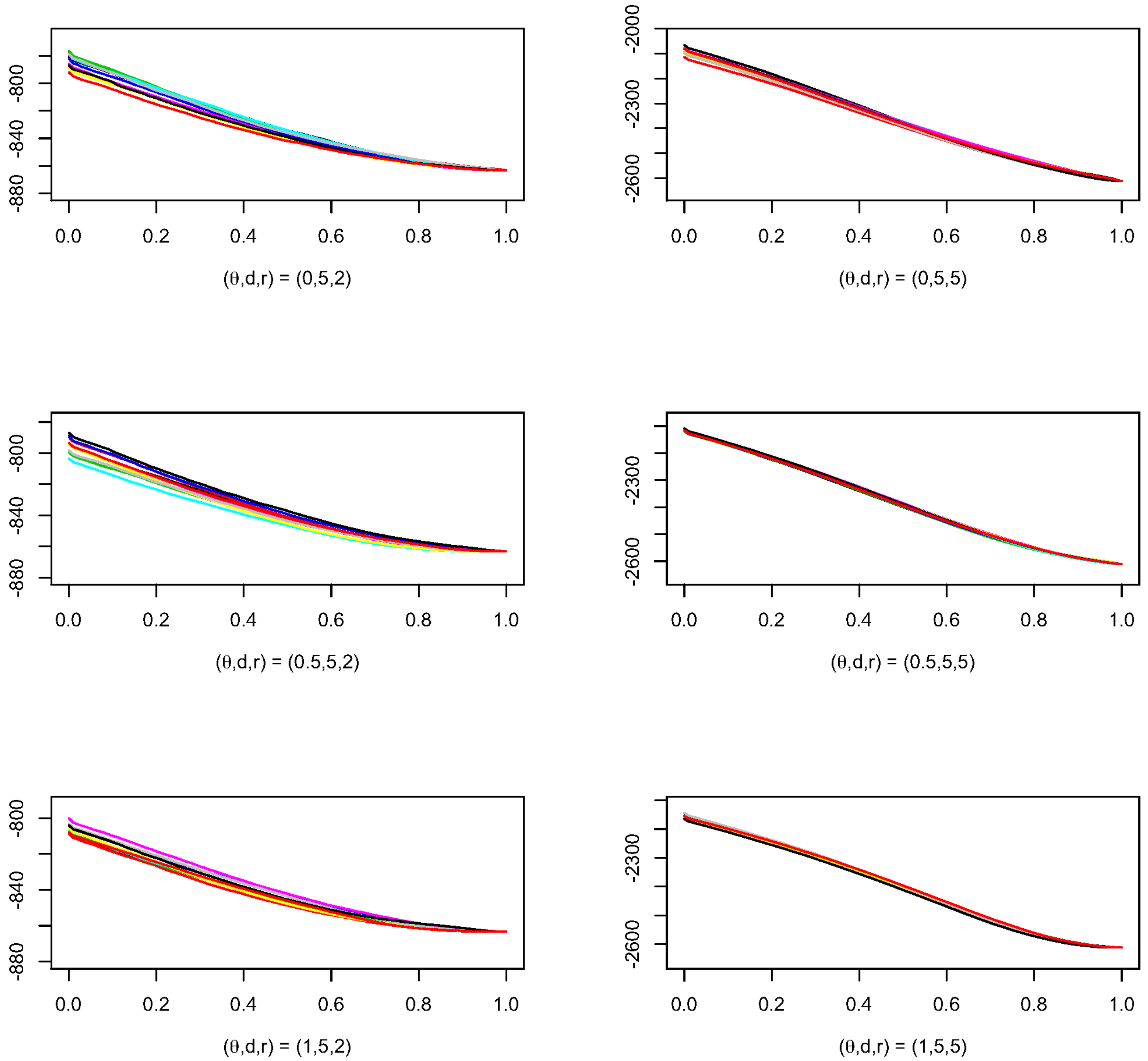
- May not hold for every pair $(\theta, \tilde{\theta})$ with $\theta < \tilde{\theta}$ regardless of $s_1 < \ldots < s_{i-1}$ and $\bar{s}_1 < \ldots < \bar{s}_k$.
- However, it holds for $0 \leq \theta < 1$ and $\tilde{\theta} = 1$ so that

$$(\mathbb{P}^{\lambda_0, \theta_0})^m \left(l_{P,m}(s^* | \theta) \geq l_{P,m}(s^* | 1) \right) = 1, \quad 0 \leq \theta < 1.$$

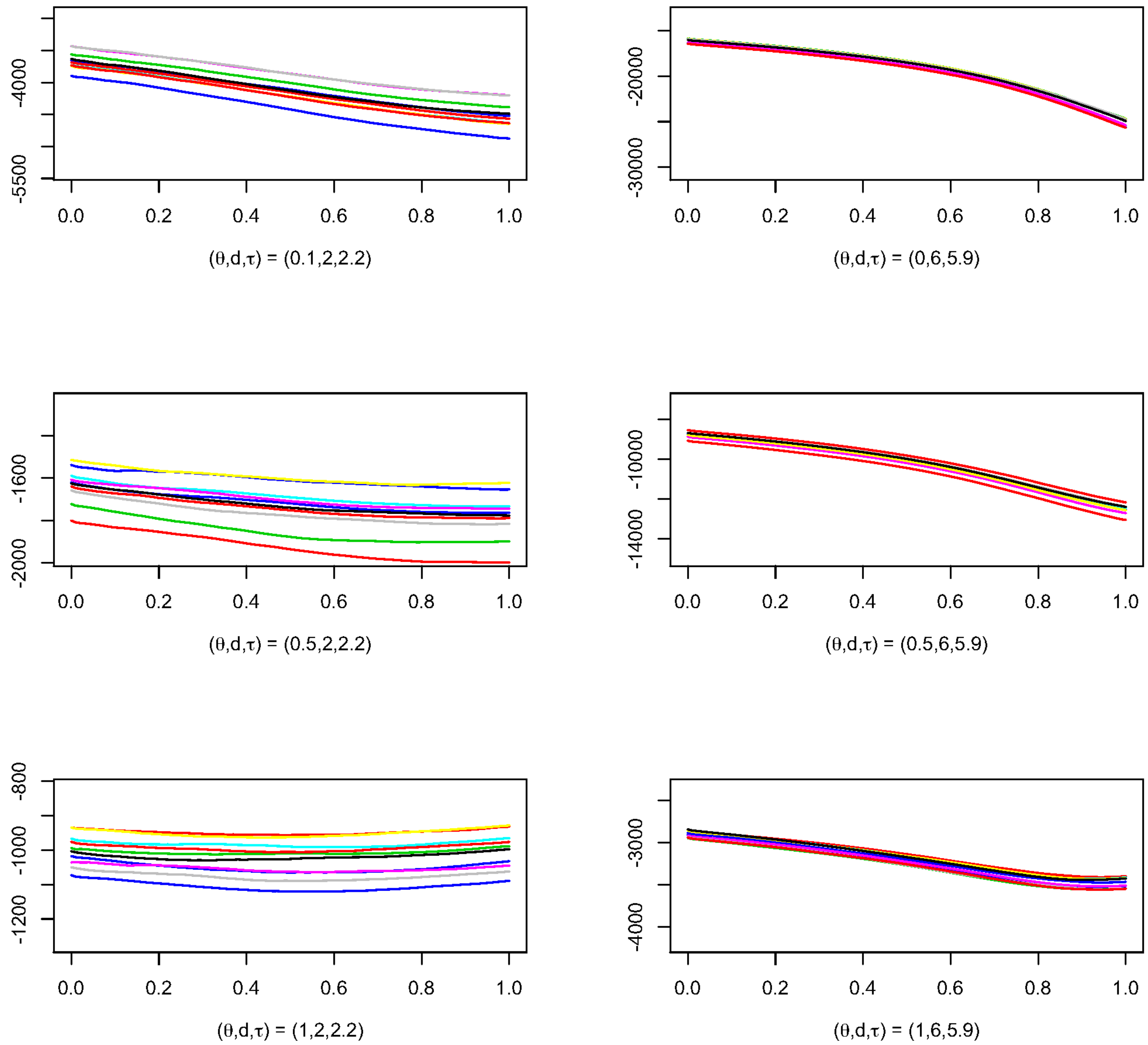
ARA$_\infty$ not monotone



ARA_1 Type 2

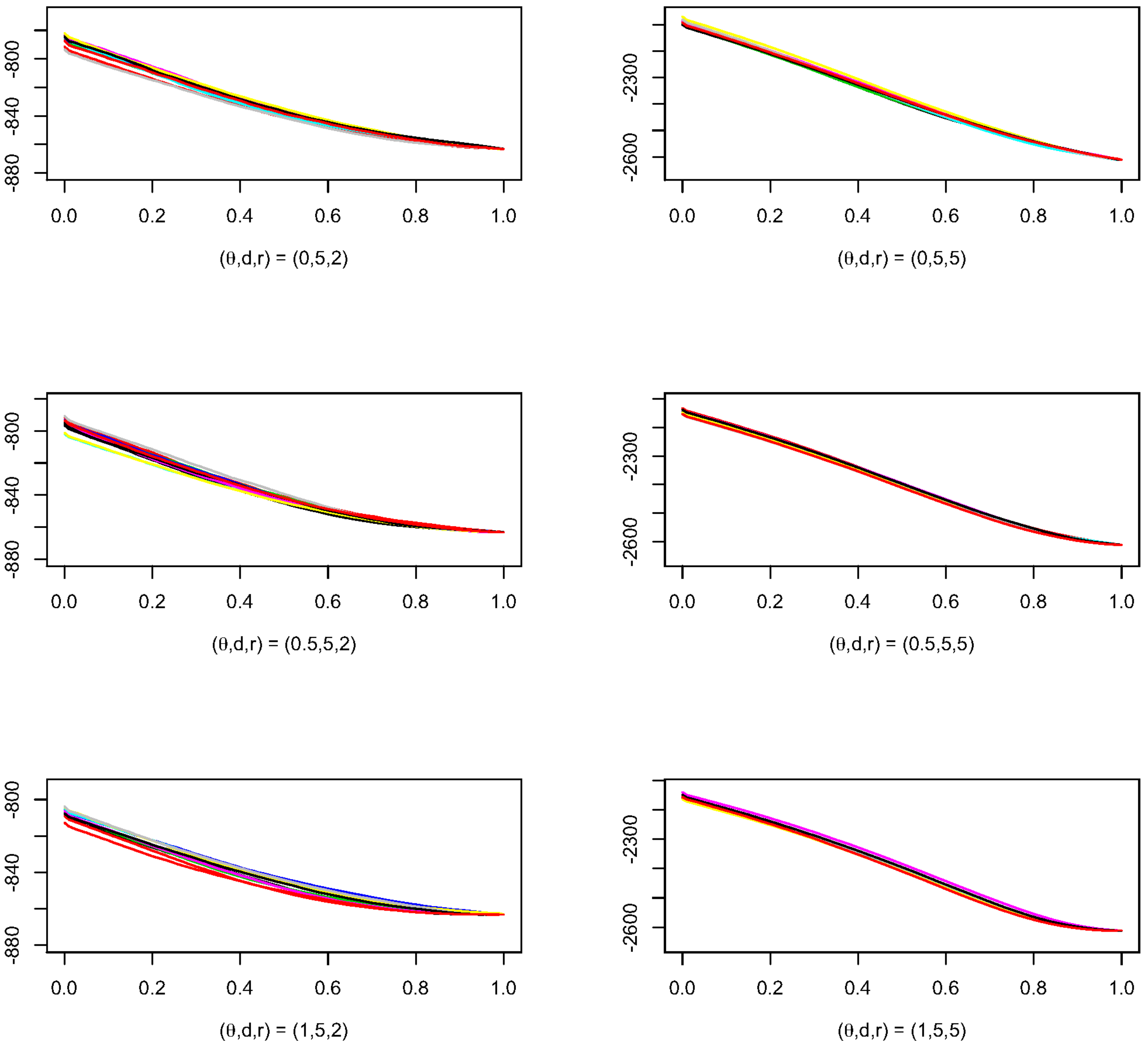

-800
-840
-880
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,r) = (0,5,2)
-2000
-2300
-2600
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,r) = (0,5,5)
-800
-840
-880
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,r) = (0.5,5,2)
-2300
-2600
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,r) = (0.5,5,5)
-800
-840
-880
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,r) = (1,5,2)
-2300
-2600
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,r) = (1,5,5)

ARA_1 Type 1



-4000
-5500
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,τ) = (0.1,2,2.2)
-20000
-30000
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,τ) = (0,6,5.9)
-1600
-2000
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,τ) = (0.5,2,2.2)
-10000
-14000
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,τ) = (0.5,6,5.9)
-800
-1000
-1200
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,τ) = (1,2,2.2)
-3000
-4000
0.0 0.2 0.4 0.6 0.8 1.0
(θ,d,τ) = (1,6,5.9)

ARA$_\infty$ Type 2



ARA$_1$ Type 2 discrete