Proportional hazards model estimation under dependent censoring using copulas and penalized likelihood - an application to a dementia dataset

Jun Ma

Department of Statistics, Macquarie University
Australia

(Joint work with Dr Kenny Xu, Dr Michael Connors and Prof Henry Brodaty)

ViCBiostat, Melbourne

Contents

Introduction

- In practical survival analysis, the event time and censoring time are sometimes correlated
- In this talk, we discuss semi-parametric Cox model fitting when dependent censoring presents
- Particularly, we discuss how to perform likelihood or penalized likelihood estimation of the model parameters, including
  (i) regression coefficients; and
  (ii) the baseline hazard
- We adopt copulas to describe the dependence

Introduction

- In practical survival analysis, the event time and censoring time are sometimes correlated
- In this talk, we discuss semi-parametric Cox model fitting when dependent censoring presents
- Particularly, we discuss how to perform likelihood or penalized likelihood estimation of the model parameters, including
  (i) regression coefficients; and
  (ii) the baseline hazard
- We adopt copulas to describe the dependence

Introduction

- In practical survival analysis, the event time and censoring time are sometimes correlated
- In this talk, we discuss semi-parametric Cox model fitting when dependent censoring presents
- Particularly, we discuss how to perform likelihood or penalized likelihood estimation of the model parameters, including
  (i) regression coefficients; and
  (ii) the baseline hazard
- We adopt copulas to describe the dependence

Introduction

- In practical survival analysis, the event time and censoring time are sometimes correlated
- In this talk, we discuss semi-parametric Cox model fitting when dependent censoring presents
- Particularly, we discuss how to perform likelihood or penalized likelihood estimation of the model parameters, including
  (i) regression coefficients; and
  (ii) the baseline hazard
- We adopt copulas to describe the dependence

Introduction

- Dependence can be modeled using copula or frailty
- Two existing copula-based methods:
- Huang and Zhang (2008, Biometrics) adopted a self-consistent estimator together with partial likelihood
- They used a computational expensive bootstrap method to estimate the standard errors
- Chen (2010, JRSSB) used the maximum likelihood (ML) method and developed consistency and asymptotic properties of the estimators
- Chen used a built-in MATLAB optimizer for constrained optimization which may be inefficient when the number of constraints is large
- Both methods do not provide smooth baseline hazard estimate, and thus, the baseline hazard results can be difficult to interpret

Introduction

- Dependence can be modeled using copula or frailty
- Two existing copula-based methods:
- Huang and Zhang (2008, Biometrics) adopted a self-consistent estimator together with partial likelihood
- They used a computational expensive bootstrap method to estimate the standard errors
- Chen (2010, JRSSB) used the maximum likelihood (ML) method and developed consistency and asymptotic properties of the estimators
- Chen used a built-in MATLAB optimizer for constrained optimization which may be inefficient when the number of constraints is large
- Both methods do not provide smooth baseline hazard estimate, and thus, the baseline hazard results can be difficult to interpret

Introduction

- Dependence can be modeled using copula or frailty
- Two existing copula-based methods:
- Huang and Zhang (2008, Biometrics) adopted a self-consistent estimator together with partial likelihood
- They used a computational expensive bootstrap method to estimate the standard errors
- Chen (2010, JRSSB) used the maximum likelihood (ML) method and developed consistency and asymptotic properties of the estimators
- Chen used a built-in MATLAB optimizer for constrained optimization which may be inefficient when the number of constraints is large
- Both methods do not provide smooth baseline hazard estimate, and thus, the baseline hazard results can be difficult to interpret

Introduction

- Dependence can be modeled using copula or frailty
- Two existing copula-based methods:
- Huang and Zhang (2008, Biometrics) adopted a self-consistent estimator together with partial likelihood
- They used a computational expensive bootstrap method to estimate the standard errors
- Chen (2010, JRSSB) used the maximum likelihood (ML) method and developed consistency and asymptotic properties of the estimators
- Chen used a built-in MATLAB optimizer for constrained optimization which may be inefficient when the number of constraints is large
- Both methods do not provide smooth baseline hazard estimate, and thus, the baseline hazard results can be difficult to interpret

Introduction

- Dependence can be modeled using copula or frailty
- Two existing copula-based methods:
- Huang and Zhang (2008, Biometrics) adopted a self-consistent estimator together with partial likelihood
- They used a computational expensive bootstrap method to estimate the standard errors
- Chen (2010, JRSSB) used the maximum likelihood (ML) method and developed consistency and asymptotic properties of the estimators
- Chen used a built-in MATLAB optimizer for constrained optimization which may be inefficient when the number of constraints is large
- Both methods do not provide smooth baseline hazard estimate, and thus, the baseline hazard results can be difficult to interpret

Introduction

- Dependence can be modeled using copula or frailty
- Two existing copula-based methods:
- Huang and Zhang (2008, Biometrics) adopted a self-consistent estimator together with partial likelihood
- They used a computational expensive bootstrap method to estimate the standard errors
- Chen (2010, JRSSB) used the maximum likelihood (ML) method and developed consistency and asymptotic properties of the estimators
- Chen used a built-in MATLAB optimizer for constrained optimization which may be inefficient when the number of constraints is large
- Both methods do not provide smooth baseline hazard estimate, and thus, the baseline hazard results can be difficult to interpret

Introduction

- Dependence can be modeled using copula or frailty
- Two existing copula-based methods:
- Huang and Zhang (2008, Biometrics) adopted a self-consistent estimator together with partial likelihood
- They used a computational expensive bootstrap method to estimate the standard errors
- Chen (2010, JRSSB) used the maximum likelihood (ML) method and developed consistency and asymptotic properties of the estimators
- Chen used a built-in MATLAB optimizer for constrained optimization which may be inefficient when the number of constraints is large
- Both methods do not provide smooth baseline hazard estimate, and thus, the baseline hazard results can be difficult to interpret

Dementia example

- Data from the PRIME study
- A longitudinal study followed 970 patients, with either dementia or mild cognitive impairment in Australia
- Interested in identifying predictors for time to institutionalization
- Patients were assessed annually with additional visits at 3 and 6 months
- Predictors: age, sex, education level, living alone, dementia type, baseline cognition ability (MMSE), baseline function ability (SMAF), baseline neuropsychiatric symptoms (total NPI), baseline dementia severity (CDR), baseline care-giver burden (ZBI), medication types, changes in cognition ability, function ability, neuropsychiatric symptoms, at 3 months,

Dementia example

- Data from the PRIME study
- A longitudinal study followed 970 patients, with either dementia or mild cognitive impairment in Australia
- Interested in identifying predictors for time to institutionalization
- Patients were assessed annually with additional visits at 3 and 6 months
- Predictors: age, sex, education level, living alone, dementia type, baseline cognition ability (MMSE), baseline function ability (SMAF), baseline neuropsychiatric symptoms (total NPI), baseline dementia severity (CDR), baseline care-giver burden (ZBI), medication types, changes in cognition ability, function ability, neuropsychiatric symptoms, at 3 months,

Dementia example

- Data from the PRIME study
- A longitudinal study followed 970 patients, with either dementia or mild cognitive impairment in Australia
- Interested in identifying predictors for time to institutionalization
- Patients were assessed annually with additional visits at 3 and 6 months
- Predictors: age, sex, education level, living alone, dementia type, baseline cognition ability (MMSE), baseline function ability (SMAF), baseline neuropsychiatric symptoms (total NPI), baseline dementia severity (CDR), baseline care-giver burden (ZBI), medication types, changes in cognition ability, function ability, neuropsychiatric symptoms, at 3 months,

Dementia example

- Data from the PRIME study
- A longitudinal study followed 970 patients, with either dementia or mild cognitive impairment in Australia
- Interested in identifying predictors for time to institutionalization
- Patients were assessed annually with additional visits at 3 and 6 months
- Predictors: age, sex, education level, living alone, dementia type, baseline cognition ability (MMSE), baseline function ability (SMAF), baseline neuropsychiatric symptoms (total NPI), baseline dementia severity (CDR), baseline care-giver burden (ZBI), medication types, changes in cognition ability, function ability, neuropsychiatric symptoms, at 3 months,

Dementia example

- Data from the PRIME study
- A longitudinal study followed 970 patients, with either dementia or mild cognitive impairment in Australia
- Interested in identifying predictors for time to institutionalization
- Patients were assessed annually with additional visits at 3 and 6 months
- Predictors: age, sex, education level, living alone, dementia type, baseline cognition ability (MMSE), baseline function ability (SMAF), baseline neuropsychiatric symptoms (total NPI), baseline dementia severity (CDR), baseline care-giver burden (ZBI), medication types, changes in cognition ability, function ability, neuropsychiatric symptoms, at 3 months,

Dementia example

- 583 patients with complete data
- Among them
  - 156 patients (26.8%) were institutionalized during the study;
  - 146 (25.0%) withdrew before the 3-year period; and
  - 281 (48.2%) not institutionalized after full 3-year
- In Brodaty et al (2014) initial analyses assumed independent censoring
- Patients withdrew from the studies appeared more likely to be institutionalized than the patients remained in the studies (e.g. older, more severe dementia, etc)
- Wish to develop Cox model fitting methods allowing both dependent and independent censorings

Dementia example

- 583 patients with complete data
- Among them
  - 156 patients (26.8%) were institutionalized during the study;
  - 146 (25.0%) withdrew before the 3-year period; and
  - 281 (48.2%) not institutionalized after full 3-year
- In Brodaty et al (2014) initial analyses assumed independent censoring
- Patients withdrew from the studies appeared more likely to be institutionalized than the patients remained in the studies (e.g. older, more severe dementia, etc)
- Wish to develop Cox model fitting methods allowing both dependent and independent censorings

Dementia example

- 583 patients with complete data
- Among them
  - 156 patients (26.8%) were institutionalized during the study;
  - 146 (25.0%) withdrew before the 3-year period; and
  - 281 (48.2%) not institutionalized after full 3-year
- In Brodaty et al (2014) initial analyses assumed independent censoring
- Patients withdrew from the studies appeared more likely to be institutionalized than the patients remained in the studies (e.g. older, more severe dementia, etc)
- Wish to develop Cox model fitting methods allowing both dependent and independent censorings

Dementia example

- 583 patients with complete data
- Among them
  - 156 patients (26.8%) were institutionalized during the study;
  - 146 (25.0%) withdrew before the 3-year period; and
  - 281 (48.2%) not institutionalized after full 3-year
- In Brodaty et al (2014) initial analyses assumed independent censoring
- Patients withdrew from the studies appeared more likely to be institutionalized than the patients remained in the studies (e.g. older, more severe dementia, etc)
- Wish to develop Cox model fitting methods allowing both dependent and independent censorings

Dementia example

- 583 patients with complete data
- Among them
  - 156 patients (26.8%) were institutionalized during the study;
  - 146 (25.0%) withdrew before the 3-year period; and
  - 281 (48.2%) not institutionalized after full 3-year
- In Brodaty et al (2014) initial analyses assumed independent censoring
- Patients withdrew from the studies appeared more likely to be institutionalized than the patients remained in the studies (e.g. older, more severe dementia, etc)
- Wish to develop Cox model fitting methods allowing both dependent and independent censorings

Copulas

- Let T = event time; C = dependent censoring time
- Dependence between T and C is captured via a copula function
- Let $K(a, b; \alpha)$ be a copula function
  - where α is the association parameter; and
  - it can be converted to τ = Kendall's rank corr coef
- At time t, the joint survival function of T and C is modeled by

$$S_{T,C}(t,t) = Pr(T > t, C > t) = K(S_T(t), S_C(t); \alpha)$$

where $S_T(t)$ and $S_C(t)$ are respectively marginal survival functions of T and C

Copulas

- Let T = event time; C = dependent censoring time
- Dependence between T and C is captured via a copula function
- Let $K(a, b; \alpha)$ be a copula function
  - where α is the association parameter; and
  - it can be converted to τ = Kendall's rank corr coef
- At time t, the joint survival function of T and C is modeled by

$$S_{T,C}(t,t) = Pr(T > t, C > t) = K(S_T(t), S_C(t); \alpha)$$

where $S_T(t)$ and $S_C(t)$ are respectively marginal survival functions of T and C

Copulas

- Let T = event time; C = dependent censoring time
- Dependence between T and C is captured via a copula function
- Let $K(a, b; \alpha)$ be a copula function
  - where α is the association parameter; and
  - it can be converted to τ = Kendall's rank corr coef
- At time t, the joint survival function of T and C is modeled by

$$S_{T,C}(t,t) = Pr(T > t, C > t) = K(S_T(t), S_C(t); \alpha)$$

where $S_T(t)$ and $S_C(t)$ are respectively marginal survival functions of T and C

Copulas

- Let T = event time; C = dependent censoring time
- Dependence between T and C is captured via a copula function
- Let $K(a, b; \alpha)$ be a copula function
  - where α is the association parameter; and
  - it can be converted to τ = Kendall's rank corr coef
- At time t, the joint survival function of T and C is modeled by

$$S_{T,C}(t,t) = Pr(T > t, C > t) = K(S_T(t), S_C(t); \alpha)$$

where $S_T(t)$ and $S_C(t)$ are respectively marginal survival functions of T and C

Copulas

- we consider the Archimedean copulas
- For $a, b \in [0, 1]$, copula function K adopts

$$K(a, b; \alpha) = \phi^{-1}(\phi(a; \alpha) + \phi(b; \alpha))$$

  where $\phi \geq 0$ is called the generator of K
- $\phi(w)$ must satisfy
  (i) $\phi(1) = 0$; and
  (ii) $\phi(w)$: convex and decreasing

Copulas

- we consider the Archimedean copulas
- For $a, b \in [0, 1]$, copula function K adopts

$$K(a, b; \alpha) = \phi^{-1}(\phi(a; \alpha) + \phi(b; \alpha))$$

where $\phi \geq 0$ is called the generator of K

- $\phi(w)$ must satisfy
(i) $\phi(1) = 0$; and
(ii) $\phi(w)$: convex and decreasing

Copulas

- we consider the Archimedean copulas
- For $a, b \in [0, 1]$, copula function K adopts

$$K(a, b; \alpha) = \phi^{-1}(\phi(a; \alpha) + \phi(b; \alpha))$$

  where $\phi \geq 0$ is called the generator of K
- $\phi(w)$ must satisfy
  (i) $\phi(1) = 0$; and
  (ii) $\phi(w)$: convex and decreasing

Copulas

- Some examples of Archimedean copulas

  (i) Independent copula $K(a, b) = ab$

  (ii) Clayton copula: $\phi(a) = a^{-1} - 1$ for $\alpha > 1$

  (iii) Gumbel-Hougaard copula: $\phi(a) = (-\log a)^{\alpha}$ for $\alpha > 1$

  (iv) Frank copula: $\phi(a) = \log \frac{(e^{\alpha a}-1)}{(e^{\alpha}-1)}$ where $-\infty < \alpha < \infty$
- How to generate dependent survival and censoring times in R?
- Use the Frank copula as an example
- Firstly, (U, V) is obtained by R function "`frankCopula`"
- Then, (t, c) are generated by $t = S_T^{-1}(U)$, $c = S_C^{-1}(V)$ (inversion method)

Copulas

- Some examples of Archimedean copulas

  (i) Independent copula $K(a, b) = ab$

  (ii) Clayton copula: $\phi(a) = a^{-1} - 1$ for $\alpha > 1$

  (iii) Gumbel-Hougaard copula: $\phi(a) = (-\log a)^{\alpha}$ for $\alpha > 1$

  (iv) Frank copula: $\phi(a) = \log \frac{(e^{\alpha a} - 1)}{(e^{\alpha} - 1)}$ where $-\infty < \alpha < \infty$
- How to generate dependent survival and censoring times in R?
- Use the Frank copula as an example
- Firstly, (U, V) is obtained by R function "`frankCopula`"
- Then, (t, c) are generated by $t = S_T^{-1}(U)$, $c = S_C^{-1}(V)$ (inversion method)

Copulas

- Some examples of Archimedean copulas

  (i) Independent copula $K(a, b) = ab$

  (ii) Clayton copula: $\phi(a) = a^{-1} - 1$ for $\alpha > 1$

  (iii) Gumbel-Hougaard copula: $\phi(a) = (-\log a)^{\alpha}$ for $\alpha > 1$

  (iv) Frank copula: $\phi(a) = \log \frac{(e^{\alpha a} - 1)}{(e^{\alpha} - 1)}$ where $-\infty < \alpha < \infty$
- How to generate dependent survival and censoring times in R?
- Use the Frank copula as an example
- Firstly, (U, V) is obtained by R function "`frankCopula`"
- Then, (t, c) are generated by $t = S_T^{-1}(U)$, $c = S_C^{-1}(V)$ (inversion method)

Copulas

- Some examples of Archimedean copulas

  (i) Independent copula $K(a, b) = ab$

  (ii) Clayton copula: $\phi(a) = a^{-1} - 1$ for $\alpha > 1$

  (iii) Gumbel-Hougaard copula: $\phi(a) = (-\log a)^{\alpha}$ for $\alpha > 1$

  (iv) Frank copula: $\phi(a) = \log \frac{(e^{\alpha a}-1)}{(e^{\alpha}-1)}$ where $-\infty < \alpha < \infty$
- How to generate dependent survival and censoring times in R?
- Use the Frank copula as an example
- Firstly, (U, V) is obtained by R function "`frankCopula`"
- Then, (t, c) are generated by $t = S_T^{-1}(U)$, $c = S_C^{-1}(V)$ (inversion method)

Copulas

- Some examples of Archimedean copulas

  (i) Independent copula $K(a, b) = ab$

  (ii) Clayton copula: $\phi(a) = a^{-1} - 1$ for $\alpha > 1$

  (iii) Gumbel-Hougaard copula: $\phi(a) = (-\log a)^{\alpha}$ for $\alpha > 1$

  (iv) Frank copula: $\phi(a) = \log \frac{(e^{\alpha a} - 1)}{(e^{\alpha} - 1)}$ where $-\infty < \alpha < \infty$
- How to generate dependent survival and censoring times in R?
- Use the Frank copula as an example
- Firstly, (U, V) is obtained by R function "`frankCopula`"
- Then, (t, c) are generated by $t = S_T^{-1}(U)$, $c = S_C^{-1}(V)$ (inversion method)

Copula based likelihood

- Let T_i = event time for individual i, where $i = 1, \ldots, n$
- We consider the case where some right censoring times are dependent on their associated event times
- Observed survival time: $X_i = \min(T_i, C_i, A_i)$, where Ci = dependent right censoring time and A_i = independent right censoring time
- Individual i has a record $(X_i, \Delta_{iT}, \Delta_{iC}, \mathbf{Z}_i^T)$, where Δ_{iT} = event indicator Δ_{iC} = dependent censoring indicator $\mathbf{Z}_i$ = a vector of p covariate values
- Assume $\mathbf{Z}_i$ is not a function of time t
- Let $(x_i, \delta_{iT}, \delta_{iC})$ be observed value of $(X_i, \Delta_{iT}, \Delta_{iC})$

Copula based likelihood

- Let $T_i =$ event time for individual i, where $i = 1, \ldots, n$
- We consider the case where some right censoring times are dependent on their associated event times
- Observed survival time: $X_i = \min(T_i, C_i, A_i)$, where $Ci =$ dependent right censoring time and $A_i =$ independent right censoring time
- Individual i has a record $(X_i, \Delta_{iT}, \Delta_{iC}, \mathbf{Z}_i^T)$, where $\Delta_{iT} =$ event indicator $\Delta_{iC} =$ dependent censoring indicator $\mathbf{Z}_i =$ a vector of p covariate values
- Assume $\mathbf{Z}_i$ is not a function of time t
- Let $(x_i, \delta_{iT}, \delta_{iC})$ be observed value of $(X_i, \Delta_{iT}, \Delta_{iC})$

Copula based likelihood

- Let T_i = event time for individual i, where $i = 1, \ldots, n$
- We consider the case where some right censoring times are dependent on their associated event times
- Observed survival time: $X_i = \min(T_i, C_i, A_i)$, where
  Ci = dependent right censoring time and
  A_i = independent right censoring time
- Individual i has a record $(X_i, \Delta_{iT}, \Delta_{iC}, \mathbf{Z}_i^T)$, where
  Δ_{iT} = event indicator
  Δ_{iC} = dependent censoring indicator
  $\mathbf{Z}_i$ = a vector of p covariate values
- Assume $\mathbf{Z}_i$ is not a function of time t
- Let $(x_i, \delta_{iT}, \delta_{iC})$ be observed value of $(X_i, \Delta_{iT}, \Delta_{iC})$

Copula based likelihood

- Let T_i = event time for individual i, where $i = 1, \ldots, n$
- We consider the case where some right censoring times are dependent on their associated event times
- Observed survival time: $X_i = \min(T_i, C_i, A_i)$, where Ci = dependent right censoring time and A_i = independent right censoring time
- Individual i has a record $(X_i, \Delta_{iT}, \Delta_{iC}, \mathbf{Z}_i^T)$, where Δ_{iT} = event indicator Δ_{iC} = dependent censoring indicator $\mathbf{Z}_i$ = a vector of p covariate values
- Assume $\mathbf{Z}_i$ is not a function of time t
- Let $(x_i, \delta_{iT}, \delta_{iC})$ be observed value of $(X_i, \Delta_{iT}, \Delta_{iC})$

Copula based likelihood

- Let $T_i =$ event time for individual i, where $i = 1, \ldots, n$
- We consider the case where some right censoring times are dependent on their associated event times
- Observed survival time: $X_i = \min(T_i, C_i, A_i)$, where
  $Ci =$ dependent right censoring time and
  $A_i =$ independent right censoring time
- Individual i has a record $(X_i, \Delta_{iT}, \Delta_{iC}, \mathbf{Z}_i^T)$, where
  $\Delta_{iT} =$ event indicator
  $\Delta_{iC} =$ dependent censoring indicator
  $\mathbf{Z}_i =$ a vector of p covariate values
- Assume $\mathbf{Z}_i$ is not a function of time t
- Let $(x_i, \delta_{iT}, \delta_{iC})$ be observed value of $(X_i, \Delta_{iT}, \Delta_{iC})$

Copula based likelihood

- Let T_i = event time for individual i, where $i = 1, \ldots, n$
- We consider the case where some right censoring times are dependent on their associated event times
- Observed survival time: $X_i = \min(T_i, C_i, A_i)$, where
  Ci = dependent right censoring time and
  A_i = independent right censoring time
- Individual i has a record $(X_i, \Delta_{iT}, \Delta_{iC}, \mathbf{Z}_i^T)$, where
  Δ_{iT} = event indicator
  Δ_{iC} = dependent censoring indicator
  $\mathbf{Z}_i$ = a vector of p covariate values
- Assume $\mathbf{Z}_i$ is not a function of time t
- Let $(x_i, \delta_{iT}, \delta_{iC})$ be observed value of $(X_i, \Delta_{iT}, \Delta_{iC})$

Copula based likelihood

- Proportional hazards models for T and C:

$$\lambda_T(x|\mathbf{Z}_i) = \lambda_{0T}(x)e^{\mathbf{Z}_i\boldsymbol{\beta}}$$
$$\lambda_C(x|\mathbf{Z}_i) = \lambda_{0C}(x)e^{\mathbf{Z}_i\boldsymbol{\phi}}$$

  where
  $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ are regression coefficient vectors, and
  $\lambda_{0T}(x)$, $\lambda_{0C}(x)$ are baseline hazard functions (nonparametric)
- These are marginal hazard functions
- Require $\lambda_{0T}(x) \geq 0$ and $\lambda_{0C}(x) \geq 0$
- Main interest is in estimating $\boldsymbol{\beta}$ and $\lambda_{0T}(x)$
- But estimation of $\boldsymbol{\phi}$ and $\lambda_{0C}(x)$ are unavoidable

Copula based likelihood

- Proportional hazards models for T and C:

$$\lambda_T(x|\mathbf{Z}_i) = \lambda_{0T}(x)e^{\mathbf{Z}_i\boldsymbol{\beta}}$$
$$\lambda_C(x|\mathbf{Z}_i) = \lambda_{0C}(x)e^{\mathbf{Z}_i\boldsymbol{\phi}}$$

  where
  $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ are regression coefficient vectors, and
  $\lambda_{0T}(x)$, $\lambda_{0C}(x)$ are baseline hazard functions (nonparametric)
- These are marginal hazard functions
- Require $\lambda_{0T}(x) \geq 0$ and $\lambda_{0C}(x) \geq 0$
- Main interest is in estimating $\boldsymbol{\beta}$ and $\lambda_{0T}(x)$
- But estimation of $\boldsymbol{\phi}$ and $\lambda_{0C}(x)$ are unavoidable

Copula based likelihood

- Proportional hazards models for T and C:

$$\lambda_T(x|\mathbf{Z}_i) = \lambda_{0T}(x)e^{\mathbf{Z}_i\boldsymbol{\beta}}$$
$$\lambda_C(x|\mathbf{Z}_i) = \lambda_{0C}(x)e^{\mathbf{Z}_i\boldsymbol{\phi}}$$

  where
  $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ are regression coefficient vectors, and
  $\lambda_{0T}(x)$, $\lambda_{0C}(x)$ are baseline hazard functions (nonparametric)
- These are marginal hazard functions
- Require $\lambda_{0T}(x) \geq 0$ and $\lambda_{0C}(x) \geq 0$
- Main interest is in estimating $\boldsymbol{\beta}$ and $\lambda_{0T}(x)$
- But estimation of $\boldsymbol{\phi}$ and $\lambda_{0C}(x)$ are unavoidable

Copula based likelihood

- Proportional hazards models for T and C:

$$\lambda_T(x|\mathbf{Z}_i) = \lambda_{0T}(x)e^{\mathbf{Z}_i\boldsymbol{\beta}}$$
$$\lambda_C(x|\mathbf{Z}_i) = \lambda_{0C}(x)e^{\mathbf{Z}_i\boldsymbol{\phi}}$$

  where
  $\boldsymbol{\beta}$, $\boldsymbol{\phi}$ are regression coefficient vectors, and
  $\lambda_{0T}(x)$, $\lambda_{0C}(x)$ are baseline hazard functions (nonparametric)
- These are marginal hazard functions
- Require $\lambda_{0T}(x) \geq 0$ and $\lambda_{0C}(x) \geq 0$
- Main interest is in estimating $\boldsymbol{\beta}$ and $\lambda_{0T}(x)$
- But estimation of $\boldsymbol{\phi}$ and $\lambda_{0C}(x)$ are unavoidable

Copula based likelihood

- Let $K_1(a,b) = \frac{\partial K(a,b)}{\partial a}$, $K_2(a,b) = \frac{\partial K(a,b)}{\partial b}$
- The likelihood based on the observations is

$$\boxed{L = \prod_{i=1}^{n} L_{iT}^{\delta_{iT}} L_{iC}^{\delta_{iC}} L_{iA}^{1-\delta_{iT}-\delta_{iC}}}$$

  where

  $L_{iT} = f_T(x_i) K_1(S_T(x_i), S_C(x_i)) S_A(x_i)$
  $L_{iC} = f_C(x_i) K_2(S_T(x_i), S_C(x_i)) S_A(x_i)$
  $L_{iA} = f_A(x_i) K(S_T(x_i), S_C(x_i))$

  Here S_T, S_C and S_A are survival functions of T, C and A respectively
- Note that f_A and S_A are not related to parameters of interest

Copula based likelihood

- Let $K_1(a,b) = \frac{\partial K(a,b)}{\partial a}$, $K_2(a,b) = \frac{\partial K(a,b)}{\partial b}$
- The likelihood based on the observations is

$$\boxed{L = \prod_{i=1}^{n} L_{iT}^{\delta_{iT}} L_{iC}^{\delta_{iC}} L_{iA}^{1-\delta_{iT}-\delta_{iC}}}$$

where

$$L_{iT} = f_T(x_i)K_1(S_T(x_i), S_C(x_i))S_A(x_i)$$
$$L_{iC} = f_C(x_i)K_2(S_T(x_i), S_C(x_i))S_A(x_i)$$
$$L_{iA} = f_A(x_i)K(S_T(x_i), S_C(x_i))$$

Here S_T, S_C and S_A are survival functions of T, C and A respectively

- Note that f_A and S_A are not related to parameters of interest

Copula based likelihood

- Let $K_1(a,b) = \frac{\partial K(a,b)}{\partial a}$, $K_2(a,b) = \frac{\partial K(a,b)}{\partial b}$
- The likelihood based on the observations is

$$\boxed{L = \prod_{i=1}^{n} L_{iT}^{\delta_{iT}} L_{iC}^{\delta_{iC}} L_{iA}^{1-\delta_{iT}-\delta_{iC}}}$$

  where

$$L_{iT} = f_T(x_i) K_1(S_T(x_i), S_C(x_i)) S_A(x_i)$$
$$L_{iC} = f_C(x_i) K_2(S_T(x_i), S_C(x_i)) S_A(x_i)$$
$$L_{iA} = f_A(x_i) K(S_T(x_i), S_C(x_i))$$

  Here S_T, S_C and S_A are survival functions of T, C and A respectively
- Note that f_A and S_A are not related to parameters of interest

Copula based likelihood

- The log likelihood is

$$\boxed{l = \sum_{i=1}^{n} \{\delta_{iT} l_{iT} + \delta_{iC} l_{iC} + (1 - \delta_{iT} - \delta_{iC}) l_{iA}\}}$$

where

$$\begin{aligned} l_{iT} &= \log \lambda_{0T}(x_i) + \boldsymbol{Z}_i\beta - \Lambda_T(x_i) + \log K_1(e^{-\Lambda_T(x_i)}, e^{-\Lambda_C(x_i)}) \\ l_{iC} &= \log \lambda_{0C}(x_i) + \boldsymbol{Z}_i\phi - \Lambda_C(x_i) + \log K_2(e^{-\Lambda_T(x_i)}, e^{-\Lambda_C(x_i)}) \\ l_{iA} &= \log K(e^{-\Lambda_T(x_i)}, e^{-\Lambda_C(x_i)}) \\ \Lambda_T(x_i) &= \Lambda_{0T}(x_i) e^{\boldsymbol{Z}_i\beta} \quad \text{and} \quad \Lambda_C(x_i) = \Lambda_{0C}(x_i) e^{\boldsymbol{Z}_i\phi} \end{aligned}$$

Copula based likelihood

- Estimating $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$ are ill-conditioned as they are infinite dimensional parameters
- Approximations:

$$\lambda_{0T}(t) = \sum_{u=1}^{m} \theta_u \psi_u(t)$$

$$\lambda_{0C}(t) = \sum_{u=1}^{m} \gamma_u \psi_u(t)$$

$\psi_u(t) \geq 0$ are basis functions; they can be
(a) indicator functions $\Rightarrow$ piecewise const $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$);
(b) M-splines;
(c) normal density functions; etc

Copula based likelihood

- Estimating $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$ are ill-conditioned as they are infinite dimensional parameters
- Approximations:

$$\lambda_{0T}(t) = \sum_{u=1}^{m} \theta_u \psi_u(t)$$

$$\lambda_{0C}(t) = \sum_{u=1}^{m} \gamma_u \psi_u(t)$$

$\psi_u(t) \geq 0$ are basis functions; they can be
(a) indicator functions $\Rightarrow$ piecewise const $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$);
(b) M-splines;
(c) normal density functions; etc

Penalized likelihood estimation

- Let $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ be respectively vectors for θ_u and γ_u
- Constrained MPL estimation for $\boldsymbol{\eta} = (\beta^T, \phi^T, \boldsymbol{\theta}^T, \gamma^T)^T$:

$$\boxed{\widehat{\boldsymbol{\eta}} = \underset{\boldsymbol{\theta} \geq 0, \gamma \geq 0}{\arg\max} \{\Phi(\boldsymbol{\eta}) = l(\boldsymbol{\eta}) - h_1 J(\boldsymbol{\theta}) - h_2 J(\gamma)\}}$$

  where $h_1, h_2 \geq 0$ are smoothing parameters
- $J(\boldsymbol{\theta})$ and $J(\gamma)$ are penalty functions
- We adopt roughness penalties:

$$J(\boldsymbol{\theta}) = \int_t \lambda_{0T}''(x)^2 dx = \boldsymbol{\theta}^T \mathbf{R} \boldsymbol{\theta}$$

$$J(\gamma) = \int_t \lambda_{0C}''(x)^2 dx = \gamma^T \mathbf{R} \gamma$$

Penalized likelihood estimation

- Let $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ be respectively vectors for θ_u and γ_u
- Constrained MPL estimation for $\boldsymbol{\eta} = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T, \boldsymbol{\theta}^T, \boldsymbol{\gamma}^T)^T$:

$$\boxed{\widehat{\boldsymbol{\eta}} = \underset{\boldsymbol{\theta}\geq 0, \boldsymbol{\gamma}\geq 0}{\arg\max} \{\Phi(\boldsymbol{\eta}) = l(\boldsymbol{\eta}) - h_1 J(\boldsymbol{\theta}) - h_2 J(\boldsymbol{\gamma})\}}$$

where $h_1, h_2 \geq 0$ are smoothing parameters

- $J(\theta)$ and $J(\gamma)$ are penalty functions
- We adopt roughness penalties:

$$J(\theta) = \int_t \lambda_{0T}''(x)^2 dx = \boldsymbol{\theta}^T \mathbf{R} \boldsymbol{\theta}$$

$$J(\gamma) = \int_t \lambda_{0C}''(x)^2 dx = \boldsymbol{\gamma}^T \mathbf{R} \boldsymbol{\gamma}$$

Penalized likelihood estimation

- Let $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ be respectively vectors for θ_u and γ_u
- Constrained MPL estimation for $\boldsymbol{\eta} = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T, \boldsymbol{\theta}^T, \boldsymbol{\gamma}^T)^T$:

$$\boxed{\widehat{\boldsymbol{\eta}} = \underset{\boldsymbol{\theta}\geq 0, \boldsymbol{\gamma}\geq 0}{\arg\max} \{\Phi(\boldsymbol{\eta}) = l(\boldsymbol{\eta}) - h_1 J(\boldsymbol{\theta}) - h_2 J(\boldsymbol{\gamma})\}}$$

where $h_1, h_2 \geq 0$ are smoothing parameters
- $J(\boldsymbol{\theta})$ and $J(\boldsymbol{\gamma})$ are penalty functions
- We adopt roughness penalties:

$$J(\theta) = \int_t \lambda''_{0T}(x)^2 dx = \boldsymbol{\theta}^T \mathbf{R} \boldsymbol{\theta}$$

$$J(\gamma) = \int_t \lambda''_{0C}(x)^2 dx = \boldsymbol{\gamma}^T \mathbf{R} \boldsymbol{\gamma}$$

Penalized likelihood estimation

- Let $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ be respectively vectors for θ_u and γ_u
- Constrained MPL estimation for $\boldsymbol{\eta} = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T, \boldsymbol{\theta}^T, \boldsymbol{\gamma}^T)^T$:

$$\boxed{\widehat{\boldsymbol{\eta}} = \underset{\boldsymbol{\theta}\geq 0, \boldsymbol{\gamma}\geq 0}{\arg\max} \left\{ \Phi(\boldsymbol{\eta}) = l(\boldsymbol{\eta}) - h_1 J(\boldsymbol{\theta}) - h_2 J(\boldsymbol{\gamma}) \right\}}$$

  where $h_1, h_2 \geq 0$ are smoothing parameters
- $J(\boldsymbol{\theta})$ and $J(\boldsymbol{\gamma})$ are penalty functions
- We adopt roughness penalties:

$$J(\boldsymbol{\theta}) = \int_t \lambda''_{0T}(x)^2 dx = \boldsymbol{\theta}^T \mathbf{R} \boldsymbol{\theta}$$

$$J(\boldsymbol{\gamma}) = \int_t \lambda''_{0C}(x)^2 dx = \boldsymbol{\gamma}^T \mathbf{R} \boldsymbol{\gamma}$$

Penalized likelihood estimation

- A constrained optimization problem
- Why penalty?
  (i) to smooth $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$;
  (ii) to make the number and location of knots for basis functions less important
- Can we use transformation for $\theta_u, \gamma_u \geq 0$? e.g. $\theta_u = \alpha_u^2$?
- Be careful! Transformations can change a concave function into a function with multiple maxima!

Penalized likelihood estimation

- A constrained optimization problem
- Why penalty?
  (i) to smooth $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$;
  (ii) to make the number and location of knots for basis functions less important
- Can we use transformation for $\theta_u, \gamma_u \geq 0$? e.g. $\theta_u = \alpha_u^2$?
- Be careful! Transformations can change a concave function into a function with multiple maxima!

Penalized likelihood estimation

- A constrained optimization problem
- Why penalty?
  (i) to smooth $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$;
  (ii) to make the number and location of knots for basis functions less important
- Can we use transformation for $\theta_u, \gamma_u \geq 0$? e.g. $\theta_u = \alpha_u^2$?
- Be careful! Transformations can change a concave function into a function with multiple maxima!

Penalized likelihood estimation

- A constrained optimization problem
- Why penalty?
  (i) to smooth $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$;
  (ii) to make the number and location of knots for basis functions less important
- Can we use transformation for $\theta_u, \gamma_u \geq 0$? e.g. $\theta_u = \alpha_u^2$?
- Be careful! Transformations can change a concave function into a function with multiple maxima!

Penalized likelihood estimation

- We developed an alternating algorithm; for each iteration:
  (i) update $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, then
  (ii) update $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$
- Let $\boldsymbol{\eta}_1 = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)^T$ and $\boldsymbol{\eta}_2 = (\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T)^T$
- Hence, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \boldsymbol{\eta}_2^T)^T$
- KKT conditions:

$$\frac{\partial \Phi}{\partial \beta_j} = 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \phi_j} = 0,$$

$$\frac{\partial \Phi}{\partial \theta_u} = 0 \quad \text{if} \quad \theta_u > 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \theta_u} < 0 \quad \text{if} \quad \theta_u = 0$$
$$\frac{\partial \Phi}{\partial \gamma_u} = 0 \quad \text{if} \quad \gamma_u > 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \gamma_u} < 0 \quad \text{if} \quad \gamma_u = 0$$

for $j = 1, \dots, p$ and $u = 1, \dots, m$

Penalized likelihood estimation

- We developed an alternating algorithm; for each iteration:
  (i) update $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, then
  (ii) update $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$
- Let $\boldsymbol{\eta}_1 = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)^T$ and $\boldsymbol{\eta}_2 = (\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T)^T$
- Hence, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \boldsymbol{\eta}_2^T)^T$
- KKT conditions:

$$\frac{\partial \Phi}{\partial \beta_j} = 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \phi_j} = 0,$$

$$\frac{\partial \Phi}{\partial \theta_u} = 0 \quad \text{if} \quad \theta_u > 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \theta_u} < 0 \quad \text{if} \quad \theta_u = 0$$
$$\frac{\partial \Phi}{\partial \gamma_u} = 0 \quad \text{if} \quad \gamma_u > 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \gamma_u} < 0 \quad \text{if} \quad \gamma_u = 0$$

for $j = 1, \ldots, p$ and $u = 1, \ldots, m$

Penalized likelihood estimation

- We developed an alternating algorithm; for each iteration:
  (i) update $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, then
  (ii) update $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$
- Let $\boldsymbol{\eta}_1 = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)^T$ and $\boldsymbol{\eta}_2 = (\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T)^T$
- Hence, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \boldsymbol{\eta}_2^T)^T$
- KKT conditions:

$$\frac{\partial \Phi}{\partial \beta_j} = 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \phi_j} = 0,$$

$$\frac{\partial \Phi}{\partial \theta_u} = 0 \quad \text{if} \quad \theta_u > 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \theta_u} < 0 \quad \text{if} \quad \theta_u = 0$$

$$\frac{\partial \Phi}{\partial \gamma_u} = 0 \quad \text{if} \quad \gamma_u > 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \gamma_u} < 0 \quad \text{if} \quad \gamma_u = 0$$

for $j = 1, \ldots, p$ and $u = 1, \ldots, m$

Penalized likelihood estimation

- We developed an alternating algorithm; for each iteration:
  (i) update $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$, then
  (ii) update $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$
- Let $\boldsymbol{\eta}_1 = (\boldsymbol{\beta}^T, \boldsymbol{\phi}^T)^T$ and $\boldsymbol{\eta}_2 = (\boldsymbol{\theta}^T, \boldsymbol{\gamma}^T)^T$
- Hence, $\boldsymbol{\eta} = (\boldsymbol{\eta}_1^T, \boldsymbol{\eta}_2^T)^T$
- KKT conditions:

$$\frac{\partial \Phi}{\partial \beta_j} = 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \phi_j} = 0,$$

$$\frac{\partial \Phi}{\partial \theta_u} = 0 \quad \text{if} \quad \theta_u > 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \theta_u} < 0 \quad \text{if} \quad \theta_u = 0$$

$$\frac{\partial \Phi}{\partial \gamma_u} = 0 \quad \text{if} \quad \gamma_u > 0 \quad \text{and} \quad \frac{\partial \Phi}{\partial \gamma_u} < 0 \quad \text{if} \quad \gamma_u = 0$$

for $j = 1, \ldots, p$ and $u = 1, \ldots, m$

Penalized likelihood estimation

- Let $\boldsymbol{\eta}^{(k)}$ be the estimate of $\boldsymbol{\eta}$ at iteration k
- At iter. $k+1$, first update η_1 by Newton (or Gauss-Newton):

$$\eta_1^{(k+1)} = \eta_1^{(k)} + \omega_1^{(k)} \mathbf{S}_1^{(k)} \frac{\partial \Phi(\eta_1^{(k)}, \eta_2^{(k)})}{\partial \eta_1}$$

- Then update η_2 by the MI algorithm:

$$\eta_2^{(k+1)} = \eta_2^{(k)} + \omega_2^{(k)} \mathbf{S}_2^{(k)} \frac{\partial \Phi(\eta_1^{(k+1)}, \eta_2^{(k)})}{\partial \eta_2}$$

where
$\mathbf{S}_2^{(k)} = \mathrm{diag}(\mathbf{S}_{21}^{(k)}, \mathbf{S}_{22}^{(k)})$
$\omega_1^{(k)}$ and $\omega_2^{(k)}$ are line search step sizes assuring
$\Phi(\eta_1^{(k+1)}, \eta_2^{(k+1)}) \geq \Phi(\eta_1^{(k+1)}, \eta_2^{(k)}) \geq \Phi(\eta_1^{(k)}, \eta_2^{(k)})$

Penalized likelihood estimation

- Let $\boldsymbol{\eta}^{(k)}$ be the estimate of $\boldsymbol{\eta}$ at iteration k
- At iter. $k+1$, first update $\boldsymbol{\eta}_1$ by Newton (or Gauss-Newton):

$$\boldsymbol{\eta}_1^{(k+1)} = \boldsymbol{\eta}_1^{(k)} + \omega_1^{(k)} \mathbf{S}_1^{(k)} \frac{\partial \Phi(\boldsymbol{\eta}_1^{(k)}, \boldsymbol{\eta}_2^{(k)})}{\partial \boldsymbol{\eta}_1}$$

- Then update η_2 by the MI algorithm:

$$\eta_2^{(k+1)} = \eta_2^{(k)} + \omega_2^{(k)} \mathbf{S}_2^{(k)} \frac{\partial \Phi(\eta_1^{(k+1)}, \eta_2^{(k)})}{\partial \eta_2}$$

where
$\mathbf{S}_2^{(k)} = \mathrm{diag}(\mathbf{S}_{21}^{(k)}, \mathbf{S}_{22}^{(k)})$
$\omega_1^{(k)}$ and $\omega_2^{(k)}$ are line search step sizes assuring
$\Phi(\eta_1^{(k+1)}, \eta_2^{(k+1)}) \geq \Phi(\eta_1^{(k+1)}, \eta_2^{(k)}) \geq \Phi(\eta_1^{(k)}, \eta_2^{(k)})$

Penalized likelihood estimation

- Let $\boldsymbol{\eta}^{(k)}$ be the estimate of $\boldsymbol{\eta}$ at iteration k
- At iter. $k+1$, first update $\boldsymbol{\eta}_1$ by Newton (or Gauss-Newton):

$$\boldsymbol{\eta}_1^{(k+1)} = \boldsymbol{\eta}_1^{(k)} + \omega_1^{(k)} \mathbf{S}_1^{(k)} \frac{\partial \Phi(\boldsymbol{\eta}_1^{(k)}, \boldsymbol{\eta}_2^{(k)})}{\partial \boldsymbol{\eta}_1}$$

- Then update $\boldsymbol{\eta}_2$ by the MI algorithm:

$$\boldsymbol{\eta}_2^{(k+1)} = \boldsymbol{\eta}_2^{(k)} + \omega_2^{(k)} \mathbf{S}_2^{(k)} \frac{\partial \Phi(\boldsymbol{\eta}_1^{(k+1)}, \boldsymbol{\eta}_2^{(k)})}{\partial \boldsymbol{\eta}_2}$$

  where
  $\mathbf{S}_2^{(k)} = \mathrm{diag}(\mathbf{S}_{21}^{(k)}, \mathbf{S}_{22}^{(k)})$
  $\omega_1^{(k)}$ and $\omega_2^{(k)}$ are line search step sizes assuring
  $\Phi(\boldsymbol{\eta}_1^{(k+1)}, \boldsymbol{\eta}_2^{(k+1)}) \geq \Phi(\boldsymbol{\eta}_1^{(k+1)}, \boldsymbol{\eta}_2^{(k)}) \geq \Phi(\boldsymbol{\eta}_1^{(k)}, \boldsymbol{\eta}_2^{(k)})$

Penalized likelihood estimation

- The MI algorithm is very efficient when m is large
- It is easy to implement; only demands the first derivative
- No need to solve a difficult linear system in each iteration
- Our experience shows it performs well and generally has a good convergence speed
- The rational of this algorithm is that it solves the KKT conditions efficiently
- Its convergence properties have been investigated in Chan and Ma (2012) (IEEE Trans. Image Processing)

Penalized likelihood estimation

- The MI algorithm is very efficient when m is large
- It is easy to implement; only demands the first derivative
- No need to solve a difficult linear system in each iteration
- Our experience shows it performs well and generally has a good convergence speed
- The rational of this algorithm is that it solves the KKT conditions efficiently
- Its convergence properties have been investigated in Chan and Ma (2012) (IEEE Trans. Image Processing)

Penalized likelihood estimation

- The MI algorithm is very efficient when m is large
- It is easy to implement; only demands the first derivative
- No need to solve a difficult linear system in each iteration
- Our experience shows it performs well and generally has a good convergence speed
- The rational of this algorithm is that it solves the KKT conditions efficiently
- Its convergence properties have been investigated in Chan and Ma (2012) (IEEE Trans. Image Processing)

Penalized likelihood estimation

- The MI algorithm is very efficient when m is large
- It is easy to implement; only demands the first derivative
- No need to solve a difficult linear system in each iteration
- Our experience shows it performs well and generally has a good convergence speed
- The rational of this algorithm is that it solves the KKT conditions efficiently
- Its convergence properties have been investigated in Chan and Ma (2012) (IEEE Trans. Image Processing)

Penalized likelihood estimation

- The MI algorithm is very efficient when m is large
- It is easy to implement; only demands the first derivative
- No need to solve a difficult linear system in each iteration
- Our experience shows it performs well and generally has a good convergence speed
- The rational of this algorithm is that it solves the KKT conditions efficiently
- Its convergence properties have been investigated in Chan and Ma (2012) (IEEE Trans. Image Processing)

Penalized likelihood estimation

- The MI algorithm is very efficient when m is large
- It is easy to implement; only demands the first derivative
- No need to solve a difficult linear system in each iteration
- Our experience shows it performs well and generally has a good convergence speed
- The rational of this algorithm is that it solves the KKT conditions efficiently
- Its convergence properties have been investigated in Chan and Ma (2012) (IEEE Trans. Image Processing)

Asymptotic results

- Let $\boldsymbol{\xi}_0 = (\boldsymbol{\beta}_0, \phi_0, \lambda^0_{0T}(t), \lambda^0_{0C}(t))$ be true parameters, and $\widehat{\boldsymbol{\xi}} = (\widehat{\boldsymbol{\beta}}, \widehat{\phi}, \widehat{\lambda}_{0T}(t), \widehat{\lambda}_{0C}(t))$ be MPL estimates.
- Let $\mu_{1n} = h_{1n}/n, \mu_{2n} = h_{2n}/n$

Theorem

Assume $m = n^v$ where $0 < v < 1$, and μ_{1n} and $\mu_{2n} \to 0$ as $n \to \infty$ Assume $\lambda_{0T}(x)$ and $\lambda_{0C}(x)$ have up to $r \geq 1$ derivatives. Then, under certain conditions and when $n \to \infty$

1. $\|\widehat{\beta} - \beta_0\|^2 \to 0$ *(a.s.) and* $\|\widehat{\phi} - \phi_0\|^2 \to 0$*(a.s.).*
2. $\sup_{t\in[x(1),x(n)]} |\widehat{\lambda}_{0T}(t) - \lambda^0_{0T}(t)| \to 0$ *(a.s.) and* $\sup_{t\in[x(1),x(n)]} |\widehat{\phi}_{0T}(t) - \phi^0_{0T}(t)| \to 0$ *(a.s).*

- The results on $\lambda^0_{0T}(t)$ and $\lambda^0_{0C}(t)$ are nonparametric

Asymptotic results

- Let $\xi_0 = (\beta_0, \phi_0, \lambda^0_{0T}(t), \lambda^0_{0C}(t))$ be true parameters, and $\widehat{\xi} = (\widehat{\beta}, \widehat{\phi}, \widehat{\lambda}_{0T}(t), \widehat{\lambda}_{0C}(t))$ be MPL estimates.
- Let $\mu_{1n} = h_{1n}/n, \mu_{2n} = h_{2n}/n$

Theorem

Assume $m = n^v$ where $0 < v < 1$, and μ_{1n} and $\mu_{2n} \to 0$ as $n \to \infty$ Assume $\lambda_{0T}(x)$ and $\lambda_{0C}(x)$ have up to $r \geq 1$ derivatives. Then, under certain conditions and when $n \to \infty$

1. $\|\widehat{\beta} - \beta_0\|^2 \to 0$ *(a.s.) and* $\|\widehat{\phi} - \phi_0\|^2 \to 0$*(a.s.).*
2. $\sup_{t\in[x(1),x(n)]} |\widehat{\lambda}_{0T}(t) - \lambda^0_{0T}(t)| \to 0$ *(a.s.) and* $\sup_{t\in[x(1),x(n)]} |\widehat{\phi}_{0T}(t) - \phi^0_{0T}(t)| \to 0$ *(a.s).*

- The results on $\lambda^0_{0T}(t)$ and $\lambda^0_{0C}(t)$ are nonparametric

Asymptotic results

- Let $\boldsymbol{\xi}_0 = (\boldsymbol{\beta}_0, \phi_0, \lambda^0_{0T}(t), \lambda^0_{0C}(t))$ be true parameters, and $\widehat{\boldsymbol{\xi}} = (\widehat{\boldsymbol{\beta}}, \widehat{\phi}, \widehat{\lambda}_{0T}(t), \widehat{\lambda}_{0C}(t))$ be MPL estimates.
- Let $\mu_{1n} = h_{1n}/n, \mu_{2n} = h_{2n}/n$

Theorem

Assume $m = n^v$ where $0 < v < 1$, and μ_{1n} and $\mu_{2n} \to 0$ as $n \to \infty$ Assume $\lambda_{0T}(x)$ and $\lambda_{0C}(x)$ have up to $r \geq 1$ derivatives. Then, under certain conditions and when $n \to \infty$

1. $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 \to 0$ *(a.s.) and* $\|\widehat{\phi} - \phi_0\|^2 \to 0$*(a.s.).*
2. $\sup_{t\in[x(1),x(n)]} |\widehat{\lambda}_{0T}(t) - \lambda^0_{0T}(t)| \to 0$ *(a.s.) and* $\sup_{t\in[x(1),x(n)]} |\widehat{\phi}_{0T}(t) - \phi^0_{0T}(t)| \to 0$ *(a.s).*

- The results on $\lambda^0_{0T}(t)$ and $\lambda^0_{0C}(t)$ are nonparametric

Asymptotic results

- Let $\boldsymbol{\xi}_0 = (\boldsymbol{\beta}_0, \phi_0, \lambda^0_{0T}(t), \lambda^0_{0C}(t))$ be true parameters, and $\widehat{\boldsymbol{\xi}} = (\widehat{\boldsymbol{\beta}}, \widehat{\phi}, \widehat{\lambda}_{0T}(t), \widehat{\lambda}_{0C}(t))$ be MPL estimates.
- Let $\mu_{1n} = h_{1n}/n, \mu_{2n} = h_{2n}/n$

Theorem

Assume $m = n^v$ where $0 < v < 1$, and μ_{1n} and $\mu_{2n} \to 0$ as $n \to \infty$ Assume $\lambda_{0T}(x)$ and $\lambda_{0C}(x)$ have up to $r \geq 1$ derivatives. Then, under certain conditions and when $n \to \infty$

1. $\|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0\|^2 \to 0$ *(a.s.) and* $\|\widehat{\phi} - \phi_0\|^2 \to 0$*(a.s.).*
2. $\sup_{t\in[x(1),x(n)]} |\widehat{\lambda}_{0T}(t) - \lambda^0_{0T}(t)| \to 0$ *(a.s.) and* $\sup_{t\in[x(1),x(n)]} |\widehat{\phi}_{0T}(t) - \phi^0_{0T}(t)| \to 0$ *(a.s).*

- The results on $\lambda^0_{0T}(t)$ and $\lambda^0_{0C}(t)$ are nonparametric

Asymptotic results

- Asymptotics useful for inferences are obtained with fixed m
- Need to consider $\theta_u = 0$ and $\gamma_u = 0$; boundary values will invalid inv(-Hessian) as covariance!
- Assume: q and r active constraints for $\boldsymbol{\theta}$ and γ respectively
- Let $\mathbf{U}_{[2(m+p)-q-r]\times[2(m+p)]}$ be matrix to indicate active constraints
- Let $\mathbf{G}_0(\boldsymbol{\eta}) = -n^{-1}E_{\eta_0}\partial^2 l(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T$
- Let $\widetilde{\mathbf{G}}_0(\eta)^{-1} = \mathbf{U}(\mathbf{U}^T\mathbf{G}_0(\eta)\mathbf{U})^{-1}\mathbf{U}^T$

Asymptotic results

- Asymptotics useful for inferences are obtained with fixed m
- Need to consider $\theta_u = 0$ and $\gamma_u = 0$; boundary values will invalid inv(-Hessian) as covariance!
- Assume: q and r active constraints for θ and γ respectively
- Let $\mathbf{U}_{[2(m+p)-q-r]\times[2(m+p)]}$ be matrix to indicate active constraints
- Let $\mathbf{G}_0(\eta) = -n^{-1}E_{\eta_0}\partial^2 l(\eta)/\partial\eta\partial\eta^T$
- Let $\widetilde{\mathbf{G}}_0(\eta)^{-1} = \mathbf{U}(\mathbf{U}^T\mathbf{G}_0(\eta)\mathbf{U})^{-1}\mathbf{U}^T$

Asymptotic results

- Asymptotics useful for inferences are obtained with fixed m
- Need to consider $\theta_u = 0$ and $\gamma_u = 0$; boundary values will invalid inv(-Hessian) as covariance!
- Assume: q and r active constraints for $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ respectively
- Let $\mathbf{U}_{[2(m+p)-q-r]\times[2(m+p)]}$ be matrix to indicate active constraints
- Let $\mathbf{G}_0(\boldsymbol{\eta}) = -n^{-1}E_{\boldsymbol{\eta}_0}\partial^2 l(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T$
- Let $\widetilde{\mathbf{G}}_0(\boldsymbol{\eta})^{-1} = \mathbf{U}(\mathbf{U}^T\mathbf{G}_0(\boldsymbol{\eta})\mathbf{U})^{-1}\mathbf{U}^T$

Asymptotic results

- Asymptotics useful for inferences are obtained with fixed m
- Need to consider $\theta_u = 0$ and $\gamma_u = 0$; boundary values will invalid inv(-Hessian) as covariance!
- Assume: q and r active constraints for $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ respectively
- Let $\mathbf{U}_{[2(m+p)-q-r]\times[2(m+p)]}$ be matrix to indicate active constraints
- Let $\mathbf{G}_0(\boldsymbol{\eta}) = -n^{-1}E_{\eta_0}\partial^2 l(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T$
- Let $\widetilde{\mathbf{G}}_0(\boldsymbol{\eta})^{-1} = \mathbf{U}(\mathbf{U}^T\mathbf{G}_0(\boldsymbol{\eta})\mathbf{U})^{-1}\mathbf{U}^T$

Asymptotic results

- Asymptotics useful for inferences are obtained with fixed m
- Need to consider $\theta_u = 0$ and $\gamma_u = 0$; boundary values will invalid inv(-Hessian) as covariance!
- Assume: q and r active constraints for $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ respectively
- Let $\mathbf{U}_{[2(m+p)-q-r]\times[2(m+p)]}$ be matrix to indicate active constraints
- Let $\mathbf{G}_0(\boldsymbol{\eta}) = -n^{-1}E_{\boldsymbol{\eta}_0}\partial^2 l(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T$
- Let $\widetilde{\mathbf{G}}_0(\eta)^{-1} = \mathbf{U}(\mathbf{U}^T\mathbf{G}_0(\eta)\mathbf{U})^{-1}\mathbf{U}^T$

Asymptotic results

- Asymptotics useful for inferences are obtained with fixed m
- Need to consider $\theta_u = 0$ and $\gamma_u = 0$; boundary values will invalid inv(-Hessian) as covariance!
- Assume: q and r active constraints for $\boldsymbol{\theta}$ and $\boldsymbol{\gamma}$ respectively
- Let $\mathbf{U}_{[2(m+p)-q-r]\times[2(m+p)]}$ be matrix to indicate active constraints
- Let $\mathbf{G}_0(\boldsymbol{\eta}) = -n^{-1}E_{\boldsymbol{\eta}_0}\partial^2 l(\boldsymbol{\eta})/\partial\boldsymbol{\eta}\partial\boldsymbol{\eta}^T$
- Let $\widetilde{\mathbf{G}}_0(\boldsymbol{\eta})^{-1} = \mathbf{U}(\mathbf{U}^T\mathbf{G}_0(\boldsymbol{\eta})\mathbf{U})^{-1}\mathbf{U}^T$

Asymptotic results

- Let η_0 be the true η corresponding to the fixed m

Theorem

Assume m is fixed. Assume μ_{1n} and μ_{2n} are $o(n^{-1/2})$. Then

1. *The MPL estimator $\widehat{\eta}$ is consistent for η_0 and,*
2. *the distribution of $\sqrt{n}(\widehat{\eta} - \eta_0)$ converges, when $n \to \infty$, to $N(\mathbf{0}_{2(p+m)\times 1}, \widetilde{\mathbf{G}}_0(\eta_0)^{-1})$.*

Asymptotic results

- Let $\boldsymbol{\eta}_0$ be the true $\boldsymbol{\eta}$ corresponding to the fixed m

Theorem

Assume m is fixed. Assume μ_{1n} and μ_{2n} are $o(n^{-1/2})$. Then

1. *The MPL estimator $\widehat{\boldsymbol{\eta}}$ is consistent for $\boldsymbol{\eta}_0$ and,*
2. *the distribution of $\sqrt{n}(\widehat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ converges, when $n \to \infty$, to $N(\mathbf{0}_{2(p+m)\times 1}, \widetilde{\mathbf{G}}_0(\boldsymbol{\eta}_0)^{-1})$.*

Simulations

- Used 2 covariates: $Z_1 \sim$ Bernoulli(0.5) and $Z_2 \sim$ U(-10, 10)
- Marginal hazards for T and C:

$$\lambda_T(t) = \lambda_{0T}(t) exp(-0.5Z_1 + 0.1Z_2)$$
$$\lambda_C(t) = \lambda_{0C}(t) exp(0.3Z_1 + 0.2Z_2)$$

  where $\lambda_{0C}(t) = 1/5$ and $\lambda_{0T}(t) = 2t/\lambda_t^2$
- The simulation involved both dependent and independent censoring
- Distribution for the independent censoring time was U(0, 10)
- Random samples generated using the Frank copula
- We used Kendall's $\tau = 0.8$ and $\tau = 0.3$

Simulations

- Used 2 covariates: $Z_1 \sim$ Bernoulli(0.5) and $Z_2 \sim$ U(-10, 10)
- Marginal hazards for T and C:

$$\lambda_T(t) = \lambda_{0T}(t) exp(-0.5Z_1 + 0.1Z_2)$$
$$\lambda_C(t) = \lambda_{0C}(t) exp(0.3Z_1 + 0.2Z_2)$$

where $\lambda_{0C}(t) = 1/5$ and $\lambda_{0T}(t) = 2t/\lambda_t^2$

- The simulation involved both dependent and independent censoring
- Distribution for the independent censoring time was U(0, 10)
- Random samples generated using the Frank copula
- We used Kendall's $\tau = 0.8$ and $\tau = 0.3$

Simulations

- Used 2 covariates: $Z_1 \sim$ Bernoulli(0.5) and $Z_2 \sim$ U(-10, 10)
- Marginal hazards for T and C:

$$\lambda_T(t) = \lambda_{0T}(t) exp(-0.5Z_1 + 0.1Z_2)$$
$$\lambda_C(t) = \lambda_{0C}(t) exp(0.3Z_1 + 0.2Z_2)$$

  where $\lambda_{0C}(t) = 1/5$ and $\lambda_{0T}(t) = 2t/\lambda_t^2$
- The simulation involved both dependent and independent censoring
- Distribution for the independent censoring time was U(0, 10)
- Random samples generated using the Frank copula
- We used Kendall's $\tau = 0.8$ and $\tau = 0.3$

Simulations

- Used 2 covariates: $Z_1 \sim$ Bernoulli(0.5) and $Z_2 \sim$ U(-10, 10)
- Marginal hazards for T and C:

$$\lambda_T(t) = \lambda_{0T}(t) exp(-0.5Z_1 + 0.1Z_2)$$
$$\lambda_C(t) = \lambda_{0C}(t) exp(0.3Z_1 + 0.2Z_2)$$

where $\lambda_{0C}(t) = 1/5$ and $\lambda_{0T}(t) = 2t/\lambda_t^2$

- The simulation involved both dependent and independent censoring
- Distribution for the independent censoring time was U(0, 10)
- Random samples generated using the Frank copula
- We used Kendall's $\tau = 0.8$ and $\tau = 0.3$

Simulations

- Used 2 covariates: $Z_1 \sim$ Bernoulli(0.5) and $Z_2 \sim$ U(-10, 10)
- Marginal hazards for T and C:

$$\lambda_T(t) = \lambda_{0T}(t) exp(-0.5Z_1 + 0.1Z_2)$$
$$\lambda_C(t) = \lambda_{0C}(t) exp(0.3Z_1 + 0.2Z_2)$$

  where $\lambda_{0C}(t) = 1/5$ and $\lambda_{0T}(t) = 2t/\lambda_t^2$
- The simulation involved both dependent and independent censoring
- Distribution for the independent censoring time was U(0, 10)
- Random samples generated using the Frank copula
- We used Kendall's $\tau = 0.8$ and $\tau = 0.3$

Simulations

- Used 2 covariates: $Z_1 \sim$ Bernoulli(0.5) and $Z_2 \sim$ U(-10, 10)
- Marginal hazards for T and C:

$$\lambda_T(t) = \lambda_{0T}(t) exp(-0.5Z_1 + 0.1Z_2)$$
$$\lambda_C(t) = \lambda_{0C}(t) exp(0.3Z_1 + 0.2Z_2)$$

  where $\lambda_{0C}(t) = 1/5$ and $\lambda_{0T}(t) = 2t/\lambda_t^2$
- The simulation involved both dependent and independent censoring
- Distribution for the independent censoring time was U(0, 10)
- Random samples generated using the Frank copula
- We used Kendall's $\tau = 0.8$ and $\tau = 0.3$

Simulations

- Let
  π_t = proportion of events
  π_c = proportion of dependent censoring
- For $\tau = 0.8, \lambda_t = 2 \Rightarrow \{\pi_t = 46\%, \pi_c = 35\%\}$;
  for $\tau = 0.8, \lambda_t = 3.5 \Rightarrow \{\pi_t = 22\%, \pi_c = 48\%\}$
- For $\tau = 0.3, \lambda_t = 2 \Rightarrow \{\pi_t = 47\%, \pi_c = 35\%\}$
- For each simulated data set, we computed the MPL or ML estimates with an assumed τ (denoted by $\widetilde{\tau}$)
- $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$ were approximated by piecewise constant functions (indicator basis)
- We applied a single smoothing parameter to all replication samples

Simulations

- Let
  π_t = proportion of events
  π_c = proportion of dependent censoring
- For $\tau = 0.8, \lambda_t = 2 \Rightarrow \{\pi_t = 46\%, \pi_c = 35\%\}$;
  for $\tau = 0.8, \lambda_t = 3.5 \Rightarrow \{\pi_t = 22\%, \pi_c = 48\%\}$
- For $\tau = 0.3, \lambda_t = 2 \Rightarrow \{\pi_t = 47\%, \pi_c = 35\%\}$
- For each simulated data set, we computed the MPL or ML estimates with an assumed τ (denoted by $\tilde{\tau}$)
- $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$ were approximated by piecewise constant functions (indicator basis)
- We applied a single smoothing parameter to all replication samples

Simulations

- Let
  π_t = proportion of events
  π_c = proportion of dependent censoring
- For $\tau = 0.8, \lambda_t = 2 \Rightarrow \{\pi_t = 46\%, \pi_c = 35\%\}$;
  for $\tau = 0.8, \lambda_t = 3.5 \Rightarrow \{\pi_t = 22\%, \pi_c = 48\%\}$
- For $\tau = 0.3, \lambda_t = 2 \Rightarrow \{\pi_t = 47\%, \pi_c = 35\%\}$
- For each simulated data set, we computed the MPL or ML estimates with an assumed τ (denoted by $\tilde{\tau}$)
- $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$ were approximated by piecewise constant functions (indicator basis)
- We applied a single smoothing parameter to all replication samples

Simulations

- Let
  π_t = proportion of events
  π_c = proportion of dependent censoring
- For $\tau = 0.8, \lambda_t = 2 \Rightarrow \{\pi_t = 46\%, \pi_c = 35\%\}$;
  for $\tau = 0.8, \lambda_t = 3.5 \Rightarrow \{\pi_t = 22\%, \pi_c = 48\%\}$
- For $\tau = 0.3, \lambda_t = 2 \Rightarrow \{\pi_t = 47\%, \pi_c = 35\%\}$
- For each simulated data set, we computed the MPL or ML estimates with an assumed τ (denoted by $\widetilde{\tau}$)
- $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$ were approximated by piecewise constant functions (indicator basis)
- We applied a single smoothing parameter to all replication samples

Simulations

- Let
  π_t = proportion of events
  π_c = proportion of dependent censoring
- For $\tau = 0.8, \lambda_t = 2 \Rightarrow \{\pi_t = 46\%, \pi_c = 35\%\}$;
  for $\tau = 0.8, \lambda_t = 3.5 \Rightarrow \{\pi_t = 22\%, \pi_c = 48\%\}$
- For $\tau = 0.3, \lambda_t = 2 \Rightarrow \{\pi_t = 47\%, \pi_c = 35\%\}$
- For each simulated data set, we computed the MPL or ML estimates with an assumed τ (denoted by $\widetilde{\tau}$)
- $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$ were approximated by piecewise constant functions (indicator basis)
- We applied a single smoothing parameter to all replication samples

Simulations

- Let
  $\pi_t =$ proportion of events
  $\pi_c =$ proportion of dependent censoring
- For $\tau = 0.8, \lambda_t = 2 \Rightarrow \{\pi_t = 46\%, \pi_c = 35\%\}$;
  for $\tau = 0.8, \lambda_t = 3.5 \Rightarrow \{\pi_t = 22\%, \pi_c = 48\%\}$
- For $\tau = 0.3, \lambda_t = 2 \Rightarrow \{\pi_t = 47\%, \pi_c = 35\%\}$
- For each simulated data set, we computed the MPL or ML estimates with an assumed τ (denoted by $\widetilde{\tau}$)
- $\lambda_{0T}(t)$ and $\lambda_{0C}(t)$ were approximated by piecewise constant functions (indicator basis)
- We applied a single smoothing parameter to all replication samples

Simulations

- Based on a random selected sample, m chosen by minimizing AIC; each bin contained approximately an equal number of observations
- Given an m, fixed $h_2 = 0$ and selected h_1 by maximizing the approximate CV criterion
- We used $N = 500$ replication samples in the simulation

Simulations

- Based on a random selected sample, m chosen by minimizing AIC; each bin contained approximately an equal number of observations
- Given an m, fixed $h_2 = 0$ and selected h_1 by maximizing the approximate CV criterion
- We used $N = 500$ replication samples in the simulation

Simulations

- Based on a random selected sample, m chosen by minimizing AIC; each bin contained approximately an equal number of observations
- Given an m, fixed $h_2 = 0$ and selected h_1 by maximizing the approximate CV criterion
- We used $N = 500$ replication samples in the simulation

Simulations

TABLE 1 Summary of simulation results for the maximum penalized likelihood estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\phi}$

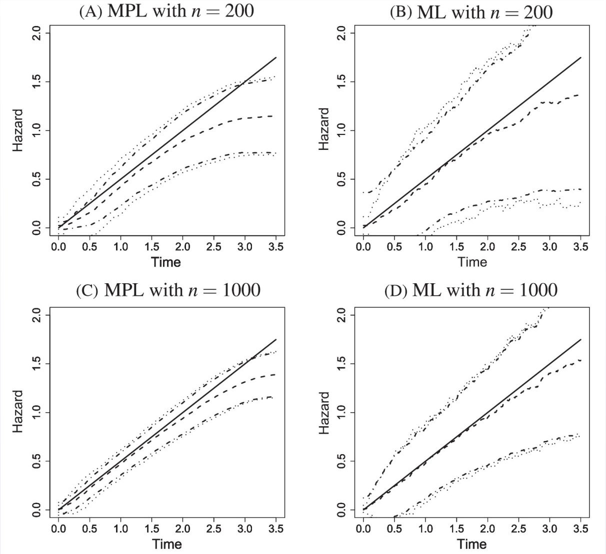
	BIAS	SD	AASD	MSE	CP(95%)	BIAS	SD	AASD	MSE	CP(95%)
$\tau = 0.8, \lambda_t = 2(\pi_t = 46\%, \pi_c = 35\%)$ $n = 200$ [Frank copula and $\tilde{\tau} = 0.8$]						$\tau = 0.8, \lambda_t = 2(\pi_t = 46\%, \pi_c = 35\%)$ $n = 1000$ [Frank copula and $\tilde{\tau} = 0.8$]				
β_1	0.009	0.166	0.171	0.028	0.952	0.012	0.073	0.075	0.005	0.954
β_2	−0.010	0.016	0.016	0.000	0.918	−0.005	0.007	0.007	0.000	0.878
ϕ_1	0.053	0.191	0.197	0.039	0.944	0.008	0.087	0.087	0.007	0.952
ϕ_2	−0.002	0.024	0.024	0.001	0.938	−0.005	0.011	0.011	0.000	0.910
$\tau = 0.8, \lambda_t = 2(\pi_t = 46\%, \pi_c = 35\%), n = 200$ [Frank copula and $\tilde{\tau} = 0.4$]						[Independent censoring]				
β_1	−0.149	0.177	0.196	0.054	0.914	−0.266	0.199	0.221	0.110	0.826
β_2	−0.035	0.018	0.019	0.002	0.568	−0.065	0.020	0.022	0.005	0.154
ϕ_1	0.225	0.212	0.233	0.096	0.898	0.247	0.236	0.251	0.117	0.886
ϕ_2	0.029	0.027	0.028	0.002	0.872	0.037	0.029	0.030	0.002	0.814
$\tau = 0.8, \lambda_t = 2(\pi_t = 46\%, \pi_c = 35\%), n = 200$ [Clayton copula and $\tilde{\tau} = 0.8$]						[Gumbel copula and $\tilde{\tau} = 0.8$]				
β_1	0.048	0.174	0.176	0.033	0.940	0.064	0.163	0.171	0.031	0.932
β_2	−0.007	0.017	0.017	0.000	0.930	−0.008	0.016	0.017	0.000	0.938
ϕ_1	0.024	0.210	0.199	0.045	0.924	−0.033	0.188	0.196	0.036	0.964
ϕ_2	−0.006	0.027	0.024	0.001	0.906	−0.015	0.022	0.021	0.001	0.842
$\tau = 0.8, \lambda_t = 3.5(\pi_t = 22\%, \pi_c = 48\%), n = 200$ [Frank copula and $\tilde{\tau} = 0.8$]						[Independent censoring]				
β_1	−0.007	0.216	0.217	0.047	0.962	−0.491	0.316	0.337	0.341	0.730
β_2	−0.013	0.023	0.023	0.001	0.890	−0.11	0.035	0.037	0.013	0.148
ϕ_1	0.017	0.182	0.187	0.033	0.956	0.131	0.204	0.210	0.059	0.916
ϕ_2	−0.008	0.021	0.022	0.000	0.924	0.018	0.024	0.024	0.001	0.910
$\tau = 0.3, \lambda_t = 2(\pi_t = 47\%, \pi_c = 35\%), n = 200$ [Frank copula and $\tilde{\tau} = 0.3$]						[Independent censoring]				
β_1	0.046	0.192	0.195	0.039	0.940	−0.053	0.215	0.216	0.049	0.946
β_2	−0.010	0.020	0.020	0.000	0.918	−0.031	0.021	0.022	0.001	0.706
ϕ_1	0.024	0.211	0.229	0.045	0.970	0.093	0.229	0.247	0.061	0.954
ϕ_2	−0.006	0.027	0.026	0.001	0.928	0.003	0.029	0.027	0.001	0.944
$\tau = 0.3, \lambda_t = 2(\pi_t = 47\%, \pi_c = 35\%)$ $n = 200$ [Frank copula and $\tilde{\tau} = 0.8$]						$\tau = 0.8, \lambda_t = 2(\pi_t = 46\%, \pi_c = 35\%)$ $n = 200$ [Frank copula and $\tilde{\tau} = 0.8$ Chen's ML]				
β_1	0.271	0.152	0.158	0.096	0.586	−0.017	0.175	0.174	0.031	0.958
β_2	0.013	0.016	0.016	0.000	0.872	−0.001	0.017	0.017	0.000	0.936
ϕ_1	−0.276	0.152	0.164	0.099	0.614	0.012	0.197	0.213	0.039	0.915
ϕ_2	−0.057	0.019	0.018	0.004	0.158	0.004	0.025	0.027	0.001	0.879

Simulations

TABLE 2 Results for the maximum penalized likelihood (MPL) and maximum likelihood (ML) estimates of $\lambda_{0T}(t)$ assuming Frank copula with $\tilde{\tau} = 0.8$

t	0.5	1.0	1.5	2.0	2.5	3.0	3.5
MPL							
BIAS	−0.092	−0.074	−0.076	−0.110	−0.218	−0.374	−0.605
SD	0.120	0.147	0.145	0.165	0.183	0.196	0.207
AASD	0.065	0.100	0.120	0.145	0.165	0.180	0.192
MSE	0.023	0.027	0.027	0.039	0.081	0.178	0.409
ML							
BIAS	−0.060	−0.048	−0.040	−0.042	−0.161	−0.210	−0.384
SD	0.212	0.302	0.324	0.393	0.414	0.531	0.562
AASD	0.210	0.245	0.281	0.350	0.396	0.463	0.495
MSE	0.048	0.094	0.107	0.157	0.197	0.326	0.464
Chen's ML							
BIAS	1.932	9.858	4.967	5.418	6.602	41.14	17.97
SD	19.82	190.4	40.12	31.08	41.33	622.7	137.3
AASD	8.355	6.538	6.771	7.184	7.924	10.17	11.37
MSE	396.7	36341	1634	995.5	1752	389463	19165

Simulations



Dementia data analysis

- We analyzed the data by assuming a Frank copula model, and
- performed a sensitivity analysis, i.e. we fit the model repeatedly using Kendall's $\tau = 0, 0.2, 0.5, 0.8$
- Results based on Clayton or Gumbel copulas are similar
- Adopted equal number of observations in each bin
- Used piecewise constant function approximating $\lambda_{0T}(t)$
- AIC gives $m = 26$ regardless of τ
- For $\tau = 0, 0.2, 0.5, 0.8$ optimal $h_1 = 1.2 \times 10^5, 10^5, 8 \times 10^4, 3 \times 10^4$

Dementia data analysis

- We analyzed the data by assuming a Frank copula model, and
- performed a sensitivity analysis, i.e. we fit the model repeatedly using Kendall's $\tau = 0, 0.2, 0.5, 0.8$
- Results based on Clayton or Gumbel copulas are similar
- Adopted equal number of observations in each bin
- Used piecewise constant function approximating $\lambda_{0T}(t)$
- AIC gives $m = 26$ regardless of τ
- For $\tau = 0, 0.2, 0.5, 0.8$
  optimal $h_1 = 1.2 \times 10^5, 10^5, 8 \times 10^4, 3 \times 10^4$

Dementia data analysis

- We analyzed the data by assuming a Frank copula model, and
- performed a sensitivity analysis, i.e. we fit the model repeatedly using Kendall's $\tau = 0, 0.2, 0.5, 0.8$
- Results based on Clayton or Gumbel copulas are similar
- Adopted equal number of observations in each bin
- Used piecewise constant function approximating $\lambda_{0T}(t)$
- AIC gives $m = 26$ regardless of τ
- For $\tau = 0, 0.2, 0.5, 0.8$
  optimal $h_1 = 1.2 \times 10^5, 10^5, 8 \times 10^4, 3 \times 10^4$

Dementia data analysis

- We analyzed the data by assuming a Frank copula model, and
- performed a sensitivity analysis, i.e. we fit the model repeatedly using Kendall's $\tau = 0, 0.2, 0.5, 0.8$
- Results based on Clayton or Gumbel copulas are similar
- Adopted equal number of observations in each bin
- Used piecewise constant function approximating $\lambda_{0T}(t)$
- AIC gives $m = 26$ regardless of τ
- For $\tau = 0, 0.2, 0.5, 0.8$ optimal $h_1 = 1.2 \times 10^5, 10^5, 8 \times 10^4, 3 \times 10^4$

Dementia data analysis

- We analyzed the data by assuming a Frank copula model, and
- performed a sensitivity analysis, i.e. we fit the model repeatedly using Kendall's $\tau = 0, 0.2, 0.5, 0.8$
- Results based on Clayton or Gumbel copulas are similar
- Adopted equal number of observations in each bin
- Used piecewise constant function approximating $\lambda_{0T}(t)$
- AIC gives $m = 26$ regardless of τ
- For $\tau = 0, 0.2, 0.5, 0.8$ optimal $h_1 = 1.2 \times 10^5, 10^5, 8 \times 10^4, 3 \times 10^4$

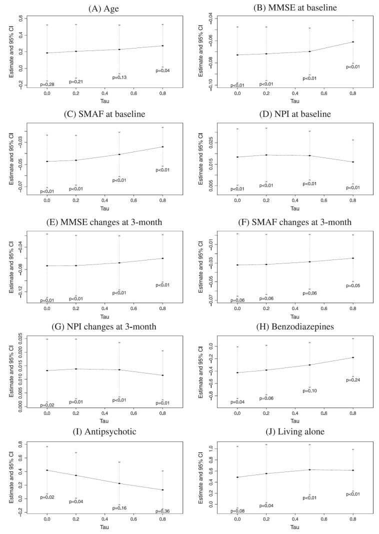
Dementia data analysis

- We analyzed the data by assuming a Frank copula model, and
- performed a sensitivity analysis, i.e. we fit the model repeatedly using Kendall's $\tau = 0, 0.2, 0.5, 0.8$
- Results based on Clayton or Gumbel copulas are similar
- Adopted equal number of observations in each bin
- Used piecewise constant function approximating $\lambda_{0T}(t)$
- AIC gives $m = 26$ regardless of τ
- For $\tau = 0, 0.2, 0.5, 0.8$ optimal $h_1 = 1.2 \times 10^5, 10^5, 8 \times 10^4, 3 \times 10^4$

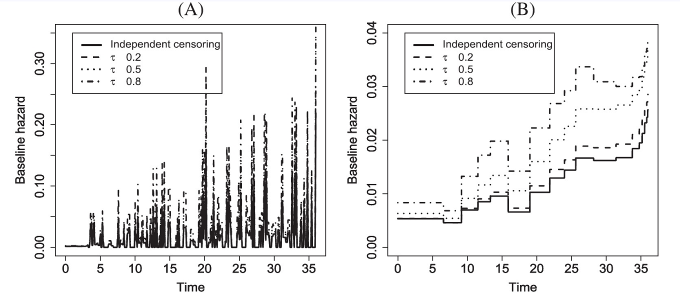
Dementia data analysis

- We analyzed the data by assuming a Frank copula model, and
- performed a sensitivity analysis, i.e. we fit the model repeatedly using Kendall's $\tau = 0, 0.2, 0.5, 0.8$
- Results based on Clayton or Gumbel copulas are similar
- Adopted equal number of observations in each bin
- Used piecewise constant function approximating $\lambda_{0T}(t)$
- AIC gives $m = 26$ regardless of τ
- For $\tau = 0, 0.2, 0.5, 0.8$
  optimal $h_1 = 1.2 \times 10^5, 10^5, 8 \times 10^4, 3 \times 10^4$

Dementia data analysis



Dementia data analysis



Conclusions

- We developed a novel estimation method for Cox proportional hazard models under dependent censoring using copulas and penalized likelihood
- It can provide quality MPL estimates if the selected copula can reasonably correctly model dependence between the failure and censoring times
- An incorrect copula has less effect than incorrect τ
- The MPL estimate of baseline hazard usually offers smaller MSE, as well as much improved interpretations, over its ML counterpart
- As true τ is unknown, recommend to conduct a sensitivity analysis for a range of τ values

Conclusions

- We developed a novel estimation method for Cox proportional hazard models under dependent censoring using copulas and penalized likelihood
- It can provide quality MPL estimates if the selected copula can reasonably correctly model dependence between the failure and censoring times
- An incorrect copula has less effect than incorrect τ
- The MPL estimate of baseline hazard usually offers smaller MSE, as well as much improved interpretations, over its ML counterpart
- As true τ is unknown, recommend to conduct a sensitivity analysis for a range of τ values

Conclusions

- We developed a novel estimation method for Cox proportional hazard models under dependent censoring using copulas and penalized likelihood
- It can provide quality MPL estimates if the selected copula can reasonably correctly model dependence between the failure and censoring times
- An incorrect copula has less effect than incorrect τ
- The MPL estimate of baseline hazard usually offers smaller MSE, as well as much improved interpretations, over its ML counterpart
- As true τ is unknown, recommend to conduct a sensitivity analysis for a range of τ values

Conclusions

- We developed a novel estimation method for Cox proportional hazard models under dependent censoring using copulas and penalized likelihood
- It can provide quality MPL estimates if the selected copula can reasonably correctly model dependence between the failure and censoring times
- An incorrect copula has less effect than incorrect τ
- The MPL estimate of baseline hazard usually offers smaller MSE, as well as much improved interpretations, over its ML counterpart
- As true τ is unknown, recommend to conduct a sensitivity analysis for a range of τ values

Conclusions

- We developed a novel estimation method for Cox proportional hazard models under dependent censoring using copulas and penalized likelihood
- It can provide quality MPL estimates if the selected copula can reasonably correctly model dependence between the failure and censoring times
- An incorrect copula has less effect than incorrect τ
- The MPL estimate of baseline hazard usually offers smaller MSE, as well as much improved interpretations, over its ML counterpart
- As true τ is unknown, recommend to conduct a sensitivity analysis for a range of τ values