A Model for Competing Risks Data with Cure Rate and its' Application

Ayon Ganguly

July 18, 2026

Department of Mathematics
Indian Institute of Technology Guwahati

Presented at: Advanced in Statistics and Data Sciences, IIT Kanpur

Publication

This talk is based on the following publication:

Ganguly, A., Sultana, F., Kundu, D. and Pal, A (2026).
A Model based on Mixture of Weibull Distributions for Depending Competing Risks Data in the Presence of Long-term Survivors, and it's Application to Malignant Melanoma Cancer Data.
Statistics in Medicine, 45:e70466

Main Sections

Cure Rate Model

Competing Risks

Motivating Data

Proposed Model

Likelihood Inference

Analysis of Melanoma Data

Cure Rate Model

Description

The Concept: The population consists of two groups

- Susceptible – consists of items that are subject to the event of interest
- Cure/Immune – consists of items that are not subject to the event of interest

Examples:

- A hospital tracks the time to relapse for patients who have undergone surgery for early-stage cancer
- A bank tracks thousands of newly issued personal loans to see how long it takes for a borrower to default on their payments
- Tracking a cohort of individuals released from a correctional facility to measure the time until they

Mixture Cure Rate Model

- T : Time-to-event
- I : Indicator (0 : cured, 1 : susceptible)
- The SF of T is

$$S_{pop}(t) = p_0 + (1 - p_0)S(t)$$

- $p_0 = P(I = 0)$: Cure probability/Incidence
- $S(t) = P(T > t \mid I = 1)$: Latency distribution

Competing Risks

Description

The Concept: A competing risk is an event whose occurrence prevents the other events of interest from happening

Examples:

- In a study of time to leukemia relapse, *death in remission* is a competing risk.
- When analyzing time to heart transplant for waitlisted patients, *death before transplant* or *removal from the list due to recovery* act as competing events.
- For a multi-component system, failure of one component (e.g., a power supply) halts the system, precluding the observation of failure in another component (e.g., a motor).

Mixture Model for Competing Risks

- The SF of the time-to-event is

$$S_{mix}(t) = P(T > t) = \sum_{k=1}^{K} \pi_k S_k(t)$$

- K : Number of competing risks
- $\widetilde{N}$: Failure mode of a subject
- $\pi_k = P(\widetilde{N} = k)$: Probability of failure due to k-th cause
- $S_k(t) = P(T > t | \widetilde{N} = k)$: Mode-specific conditional SF

Motivating Data

Malignant Melanoma Cancer Data Set (1/4)

- Melanoma : Skin cancer
- Collected at Odense University Hospital during 1962 to 1977
- Two hundred and five patients
- Number of days after the operation
  - until they died
  - until they left the study
  - until the termination of the study in the year 1977

Malignant Melanoma Cancer Data Set (2/4)

no	status	days	ulc	thick	sex
789	3	10	1	676	1
13	3	30	0	65	1
97	2	35	0	134	1
16	3	99	0	290	0
21	1	185	1	1208	1
469	1	204	1	484	1

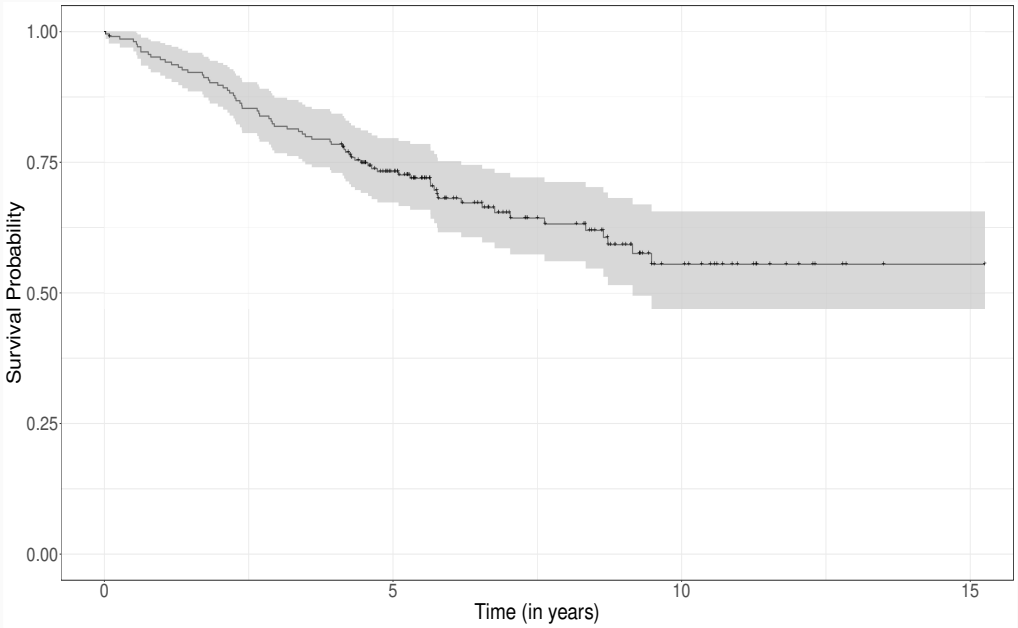
- no : Patient code
- status : Survival status (1: dead from melanoma, 2: alive, 3: dead from other cause)
- days : Survival time (in days)
- ulc : Ulceration (1: present, 0: absent)
- thick : Tumour thickness (in 1/100 mm)
- sex : Gender (0: female, 1: male)

Malignant Melanoma Cancer Data Set (3/4)

- Melanoma cancer: 28%
- Other causes: 7%
- Right censored: 65%
- Sex (Male 39%, Female 61%)
- Status of ulcer (Present 44%, Absent 56%)
- Thickness of tumor (Mean 2.92 mm, SD 2.92 mm)
- Available in `timereg` package

Malignant Melanoma Cancer Data Set (4/4)

- Patients died due to other causes: Right censored
- The Kaplan-Meier estimate



- Levels off at a non-zero proportion
- Possibility of non-zero cure proportion

Malignant Melanoma Cancer Data Set (5/5)

- Presence of cure rate
- Presence of competing risks
- Presence of censoring
- Presence of covariates

Proposed Model

Proposed Model (1/4)

- The population consists of two groups
  - Susceptible
  - Cure
- There are K mutually exclusive and exhaustive causes

Proposed Model (2/4)

- $\widetilde{T}$: Time-to-event
- $\widetilde{I}$: Indicator (0 : cured, 1 : susceptible)

Competing Risks:

$$S_{mix}(t) = P(\widetilde{T} > t|\widetilde{I} = 1) = \sum_{k=1}^{K} \pi_k S_k(t)$$

Cure Rate:

$$S_{pop}(t) = P(\widetilde{T} > t) = p_0 + (1 - p_0) \sum_{k=1}^{K} \pi_k S_k(t)$$

- K : Number of competing risks
- $\widetilde{N}$: Failure mode of a susceptible subject
- $\pi_k = P(\widetilde{N} = k|\widetilde{I} = 1)$: Failure probability due to k-th cause
- $S_k(t) = P(\widetilde{T} > t|\widetilde{N} = k, \widetilde{I} = 1)$: Mode-specific SF
- $p_0 = P(\widetilde{I} = 0)$: Cure probability

Proposed Model (3/4)

- $k-$th mode-specific time-to-event follows Weibull distribution
- $S_k(t) = \mathbf{exp}\left\{-\lambda_k t^{\alpha_k}\right\}$ for $t > 0$
- $\lambda_k > 0$: Scale parameter
- $\alpha_k > 0$: Shape parameter

Proposed Model (4/4)

- Probability of a subject being cured depends on covariates
- Binary regression models
- Logistic link function

$$p_0(\boldsymbol{x};\, \boldsymbol{\beta}) = \frac{1}{1 + \exp(\boldsymbol{x}^T \boldsymbol{\beta})}$$

- $\boldsymbol{x} = (1,\, x_1,\, \ldots,\, x_L)$: Vector of covariates
- $\boldsymbol{\beta} = (\beta_0,\, \beta_1,\, \ldots,\, \beta_L)$: Parameter vector

Likelihood Inference

Likelihood Inference (Form of data)

- Form of available data :

$$\{(t_\ell, \delta_\ell, \boldsymbol{x}_\ell) : \ell = 1, 2, \ldots, n\}$$

- t_ℓ : Observed value of $T_\ell = \min(\widetilde{T}_\ell, C_\ell)$
- δ_ℓ : Observed value of $\Delta_\ell = \widetilde{N} \times I(\widetilde{T}_\ell \leq C_\ell)$
- $\widetilde{T}$: Time-to-event
- C_ℓ : Censoring time
- $\boldsymbol{x}_\ell$: Covariate vector

Likelihood Inference (Likelihood function)

- Likelihood function (non-informative censoring)

$$L(\boldsymbol{\gamma}) \propto \prod_{\ell=1}^{n} \left[\left\{ \prod_{k=1}^{K} \{(1 - p_0(\boldsymbol{x}_\ell; \boldsymbol{\beta}))\pi_k f_k(t_\ell; \boldsymbol{\theta}_k)\}^{I(\delta_\ell = k)} \right\} \{S_{pop}(t_\ell; \boldsymbol{\gamma})\}^{I(\delta_\ell = 0)} \right]$$

- $\boldsymbol{\gamma} = (\alpha_1, \ldots, \alpha_K, \lambda_1, \ldots, \lambda_K, \pi_1, \ldots, \pi_K, \beta_0, \beta_1, \ldots, \beta_L)$
- $\boldsymbol{\theta}_k = (\alpha_k, \beta_k)$
- MLEs can be found by maximizing likelihood function over

$$\boldsymbol{\Gamma} = \Bigg\{ \boldsymbol{\gamma} : \alpha_k > 0,\ \lambda_k > 0,\ k = 1, 2, \ldots, K,$$

$$0 < \pi_k < 1,\ k = 1, 2, \ldots, K,\ \sum_{k=1}^{K} \pi_k = 1,$$

$$\beta_\ell \in \mathbb{R},\ \ell = 0, 1, 2, \ldots, L \Bigg\}$$

- $(3K + L + 1)$ dimensional constrained optimization problem

Likelihood Inference (EM based method)

- $\widetilde{I}$ is latent if $\delta = 0$
- $\widetilde{I} = 1$ if $\delta \neq 0$
- EM algorithm is a natural choice to obtain the MLEs
- The complete data :

$$\{(t_\ell, \delta_\ell, \boldsymbol{x}_\ell, i_\ell) : \ell = 1, 2, \ldots, n\}$$

- i_ℓ : Realized value of $\widetilde{I}$ for ℓ-th subject

Likelihood Inference (Complete data log-likelihood)

$$\begin{aligned}\ln L_c(\gamma) = & \sum_{\ell \in J_1} \beta' x_\ell + \sum_{\ell \in J_1} \sum_{k=1}^{K} I(\delta_\ell = k) \ln \pi_k \\ & + \sum_{\ell \in J_1} \sum_{k=1}^{K} I(\delta_\ell = k) \ln f_k(t_\ell; \theta_k) \\ & + \sum_{\ell \in J_0} i_\ell \beta' x_\ell + \sum_{\ell \in J_0} i_\ell \ln S_{mix}(\tau_\ell; \psi) \\ & - \sum_{\ell=1}^{n} \ln(1 + \exp\{\beta' x_\ell\})\end{aligned}$$

- $J_1 = \{l : \delta_\ell \neq 0\}$ and $J_0 = \{l : \delta_\ell = 0\}$
- $\psi = (\alpha_1, \ldots, \alpha_K, \lambda_1, \ldots, \lambda_K, \pi_1, \ldots, \pi_K)$

Likelihood Inference (E and M step)

- At the $(m+1)-$st step, the expectation of complete data log-likelihood function is

$$\widetilde{Q}^{(m+1)}(\boldsymbol{\gamma}) = \widetilde{Q}_1^{(m+1)}(\boldsymbol{\beta}) + \widetilde{Q}_2^{(m+1)}(\boldsymbol{\psi})$$

- $\widetilde{Q}_1^{(m+1)}(\cdot)$ is a (non-linear) function of $\boldsymbol{\beta}$ only
- $\widetilde{Q}_2^{(m+1)}(\cdot)$ is a (non-linear) function of $\boldsymbol{\psi}$ only
- Maximize $\widetilde{Q}_1^{(m+1)}(\cdot)$ and $\widetilde{Q}_2^{(m+1)}(\cdot)$ separately
- Max $\widetilde{Q}_1^{(m+1)}(\cdot)$: $(L+1)$ dimensional non-linear optimization
- Max $\widetilde{Q}_2^{(m+1)}(\cdot)$: K one-dimensional non-linear optimization

MLE of Cure Rate

- Once the MLE of $\boldsymbol{\gamma}$ is obtained, the MLE of the cure fraction is found using the invariance property of the MLE
- MLE of the cure fraction for a subject with covariate $\boldsymbol{x}$

$$\hat{p}_0 = \frac{1}{1 + \exp(\boldsymbol{x}^T \hat{\boldsymbol{\beta}}_{\text{MLE}})}$$

Likelihood Inference (Confidence interval)

- Model parameters – Luis missing value principle
- Cure rate – δ-method

Analysis of Melanoma Data

Analysis of Melanoma Data (Recall)

no	status	days	ulc	thick	sex
789	3	10	1	676	1
13	3	30	0	65	1
97	2	35	0	134	1
16	3	99	0	290	0
21	1	185	1	1208	1
469	1	204	1	484	1

- no : Patient code
- status : Survival status (1: dead from melanoma, 2: alive, 3: dead from other cause)
- days : Survival time (in days)
- ulc : Ulceration (1: present, 0: absent)
- thick : Tumour thickness (in 1/100 mm)
- sex : Gender (0: female, 1: male)

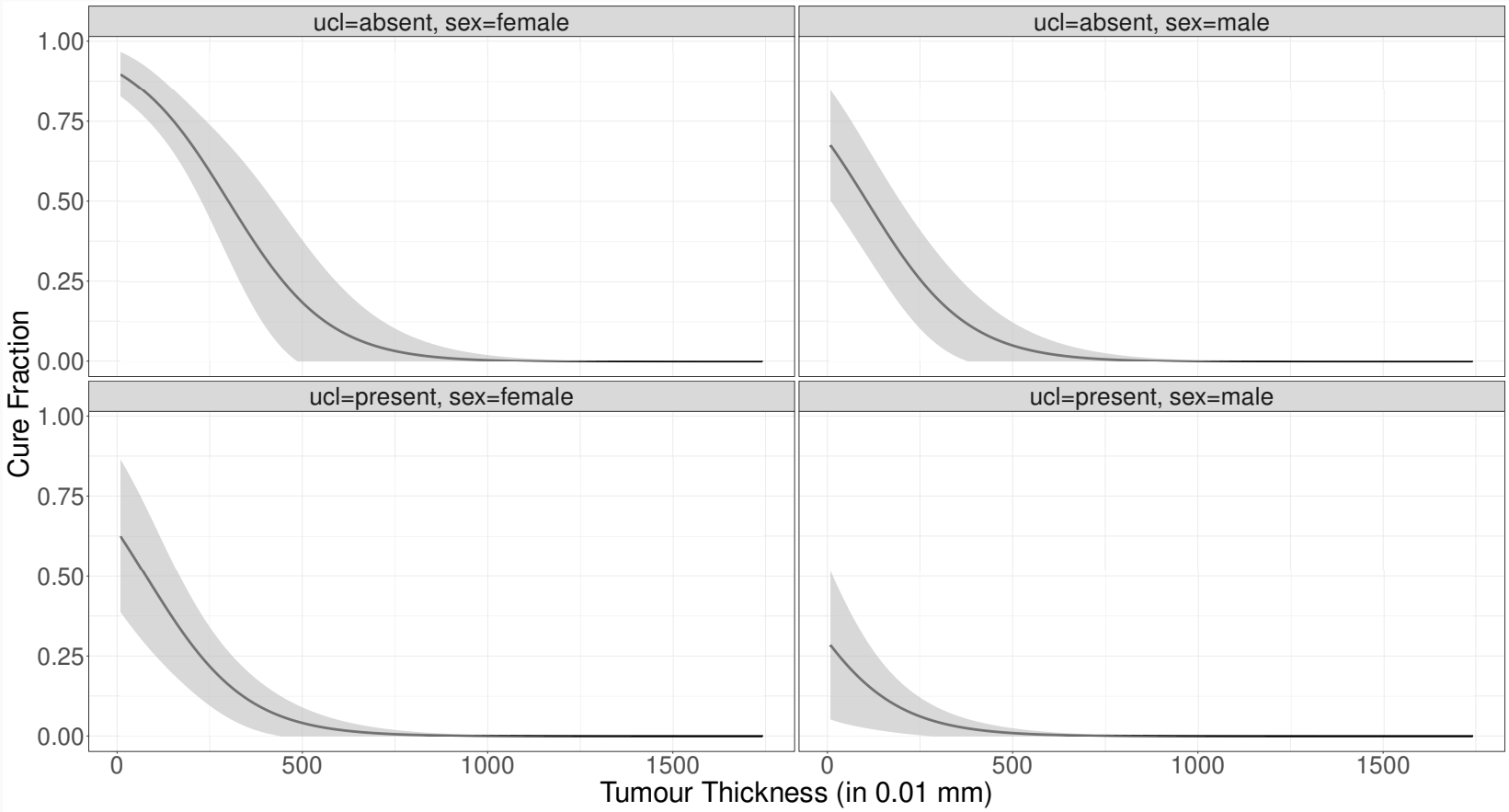
Analysis of Melanoma Data (Modelling)

- Mode 1 – death due to malignant melanoma
- Mode 2 – death due to any other causes
- 3 covariates (ulcer status, tumour thickness and sex)
- Scale survival time dividing by 365
- Weibull model for cause specific distribution
- Logistic link function

Analysis of Melanoma Data (Estimates)

Parameters	MLE	95% CI
α_1 (shape, mode 1)	1.603	1.348 – 1.858
α_2 (shape, mode 2)	0.655	0.321 – 0.988
λ_1 (scale, mode 1)	0.085	0.051 – 0.119
λ_2 (scale, mode 2)	0.081	0.039 – 0.123
π_1 (mixture coefficient, mode 1)	0.564	0.418 – 0.710
π_2 (mixture coefficient, mode 2)	0.436	0.206 – 0.666
β_0 (intercept)	−2.238	−3.014 – −1.461
β_1 (ulcer status)	1.654	0.796 – 2.512
β_2 (tumour thickness)	0.007	0.004 – 0.011
β_3 (sex)	1.432	0.608 – 2.256

Analysis of Melanoma Data (Plots of cure fraction)



Further Extension

- In many studies, analysts are interested in one specific mode of failure (e.g., mode 1)
- The scale parameter of the Weibull distribution corresponding to the mode of interest may be modeled using a log-linear relationship to the covariates

$$\lambda_{1l}(\boldsymbol{x};\boldsymbol{\eta}) = \exp(\boldsymbol{\eta}^T \boldsymbol{x}_l)$$

- At the $(m+1)$st step, the expectation of the complete data log-likelihood:

$$\tilde{Q}^{(m+1)}(\boldsymbol{\gamma}) = \tilde{Q}_1^{(m+1)}(\boldsymbol{\beta}) + \tilde{Q}_2^{(m+1)}(\boldsymbol{\zeta})$$

- Max $\tilde{Q}_1^{(m+1)}(\boldsymbol{\beta})$: Solving a $(L+1)$ dim non-linear equation
- Max $\tilde{Q}_2^{(m+1)}(\boldsymbol{\zeta})$: Solving K systems of non-linear equations
  - One system involving $(L+2)$ variables
  - The remaining systems each involving one variable

Bibliography

Bibliography

- Jiang, Siyuan and Dimitri Kececioglu (1992). “Maximum likelihood estimates, from censored data, for mixed-Weibull distributions”. In: *IEEE Transactions on Reliability* 41.2, pp. 248–255.
- Pal, Suvra and N. Balakrishnan (2016). “Destructive negative binomial cure rate model and EM-based likelihood inference under Weibull lifetime”. In: *Statistics & Probability Letters* 116, pp. 9–20.
- Rodrigues, J. et al. (2011). “Destructive weighted Poisson cure rate models.”. In: *Lifetime Data Analysis* 17, pp. 333–346.

Thank You

---

aganguly@iitg.ac.in
https://ayonganguly.github.io/