

Improved Survival Modeling Using A Piecewise Exponential Approach

Gang Han, Michael J. Schell, and Jongphil Kim

Biostatistics Core
H.Lee Moffitt Cancer Center & Research Institute

Joint Statistical Meetings
Washington, DC
August 03, 2009

- 1 Introduction
- 2 A motivating example
- 3 The method
- 4 Summary and future research

Models for time-to-event data

- Non-parametric models, e.g., Product-Limits estimates, Life-Table analysis.
 - Strength:** Easy to understand and apply, distribution-free.
 - Weakness:** Generally the uncertainty is high, less efficient than proper parametric models.
- Parametric models, e.g., Exponential, Weibull, Lognormal, Gamma.
 - Strength:** Parsimonious, makes use of all the data.
 - Weakness:** Use of the parametric models is restricted by the distribution assumption; more effort is needed to ensure parametric modeling.

Models for time-to-event data

Middle ground: Piecewise parametric modeling.

- **Gains most of strengths of both while minimizing the weaknesses.**

- Piecewise exponential distribution is a simple and flexible distribution for modeling time to event data. (Kim and Proschan 1991, Qiu et al., 1999, Demarqui et al. 2008)
- **Challenge** determination of the pieces.

A non-small cell lung cancer (NSCLC) example

- Lung cancer has the highest mortality among all types of cancer.
- Non-small cell lung cancer (NSCLC) is the most common and most severe lung cancer. (NSCLC median overall survival is about 8 to 10 months.)
- We model a survival data set in a clinical trial at Moffitt Cancer Center for NSCLC patients from recent phase II trials. Sample size $N = 178$.

Results and comparison

Using Weibull survival model can result in large bias

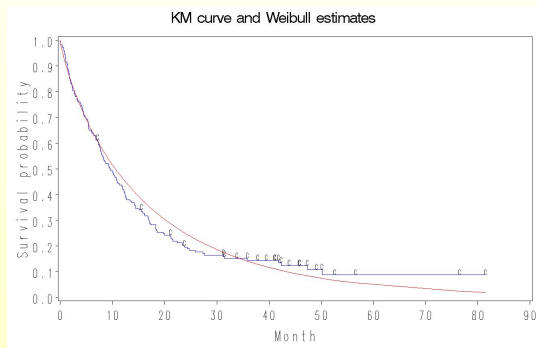


Figure 1. Plot of KME (blue step function) and Weibull estimates (the continuous curve). The label “C” indicates censored cases. The estimated survival function from Weibull distribution has the form $S(t) = e^{-(t/16.27)^{0.85}}$.

Results and comparison

The proposed piecewise exponential model fits the data much better.

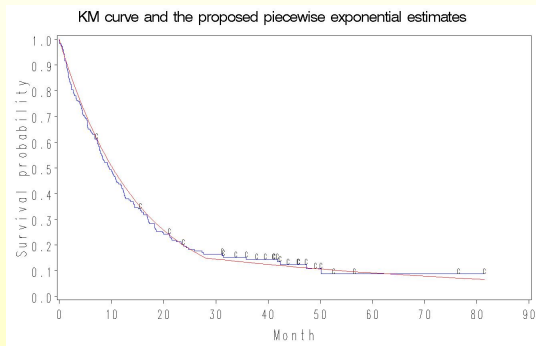


Figure 2. Plot of KME (the step function) and the proposed piecewise estimate (the continuous curve).

The method

The idea is to integrate three major components:

- 1 A likelihood ratio (UMPU) test;
- 2 A backward elimination procedure;
- 3 An optional pool-adjacent-violators algorithm.

The 1st component of the proposed method

A likelihood ratio test (LRT)

- For exponentially distributed survival times, the total time on test (TTOT) between two events follows $\text{Gamma}(\alpha, \lambda)$;
 α : the number of events (known);
 λ : unknown but its MLE is TTOT/α .

- We attempt to test the equivalence of the scale parameters (λ) of two Gamma distributions.

A likelihood ratio test (LRT)

- Two independent random variables $X_1 \sim \Gamma(\alpha_1, \lambda_1)$, $X_2 \sim \Gamma(\alpha_2, \lambda_2)$; α_1, α_2 are known.

Test: $H_0 : \lambda_1 = \lambda_2$ vs. $H_1 : \lambda_1 \neq \lambda_2$.

- The LRT is to reject H_0 if $\Phi(X_1, X_2) = \left(\frac{X_1}{X_1+X_2}\right)^{\alpha_1} \left(\frac{X_2}{X_1+X_2}\right)^{\alpha_2}$ is small.
- **Theorem** The above test is the uniformly most powerful unbiased (UMPU) test under certain conditions.

Quantifying the P-value and power

Let $Y = \frac{X_1}{X_1 + X_2}$, then $\Phi(X_1, X_2) = Y^{\alpha_1} (1 - Y)^{\alpha_2} = h(Y)$.

Facts

- 1 Y is distributed as Beta(α_1, α_2) under H_0 .
- 2 function $h(Y)$ is increasing if $Y \in \left[0, \frac{\alpha_1}{\alpha_1 + \alpha_2}\right]$ and is decreasing if $Y \in \left[\frac{\alpha_1}{\alpha_1 + \alpha_2}, 1\right]$. So $h(Y) < C_\alpha$ iff $Y \in (0, a_1) \cup (a_2, 1)$ where $a_1 \leq a_2$ and $h(a_1) = h(a_2) = C_\alpha$.
- 3 p-value: $\text{Prob}(Y \in (0, a_1) \cup (a_2, 1))$ under H_0 .
- 4 power: $\text{Prob}(Y \in (0, a_1) \cup (a_2, 1))$ under H_1 .

An example

If $\alpha_1 = \alpha_2 = 1$, then $Y \sim \text{uniform}(0, 1)$ under H_0 . Size 5% UMPU test rejects H_0 if $x_1/(x_1 + x_2) \in (0, 0.025) \cup (0.975, 1)$, where (x_1, x_2) is a realization of (X_1, X_2) .

- $0.025 = \frac{1}{40}$ and $0.975 = \frac{39}{40}$. So we reject $\lambda_1 = \lambda_2$ if one survival is at least 39 times the other one.

The 2nd component of the proposed method

Backward elimination (BE)

- 1 Start with hazard rate changes for each distinct failure time.
- 2 Combine the two adjacent hazards with the highest p-value from the LRT.
- 3 Repeat until the p-value is below some critical “naive” p-value α^* .
- 4 α^* can be found using a Monte Carlo approach (Schell and Singh, 1997) in order to achieve an α level test.

The 3rd component of the proposed method

An optional preprocessing: The pool-adjacent-violators algorithm (PAVA) for decreasing failure rate functions

- Robertson et al. (1988), “Order restricted statistical inference.”
- If the hazard function is “*known*” to be non-increasing, one can pool the adjacent cases where the non-increasing assumption is violated (Robertson et al.). Then the BE approach can be used.
- PAVA might lead to a different solution.

Estimation and confidence interval(CI)

- Suppose t has a piecewise exponential distribution with pieces divided by times $0 = t_0 < t_1 < t_2 < \dots < t_p$ and model parameters are $\lambda = (\lambda_1, \lambda_2, \dots, \lambda_{p+1})$.
Define $E_0 = 1$ and $E_j = \prod_{k=1}^j e^{-(t_k - t_{k-1})/\lambda_k}$.
 - $S(t|\lambda) = E_{i-1} e^{-(t-t_{i-1})/\lambda_i}$ for all $t \in [t_{i-1}, t_i)$;
 - $E(t|\lambda) = \lambda_1 + \sum_{i=1}^p E_i \times (\lambda_{i+1} - \lambda_i)$;
 - $\text{Median}(t|\lambda) = t_m + \lambda_{m+1} \log(2E_m)$ for $E_m \geq 0.5$ and $E_{m+1} \leq 0.5$.
- Assuming that the t_i s are known, we let
 - $\hat{\lambda} \equiv \text{MLE of } \lambda$;
CIs of λ have been derived in (Epstein and Sobel, 1953);
 - $\hat{S}(t|\lambda) \equiv S(t|\hat{\lambda})$; we estimate $E(t|\lambda)$ and $\text{Median}(t|\lambda)$ in the same way.
 - Similarly, CIs for $S(t|\lambda)$, $E(t|\lambda)$, and $\text{Median}(t|\lambda)$ build on the CIs of λ .

Demo of our method in the NSCLC example

- We use PAVA with non-increasing hazard rate assumption because the risk of NSCLC cancer is believed to decrease over time.
- There were 141 distinct failure times in total; 6 times left after PAVA: (12.7, 18.6, 23.5, 24.7, 28.0, 47.9).

Demo of our method in the NSCLC example

- BE 1st iteration ($p = 6$):

t_0, \dots, t_p	0	12.7	18.6	23.5	24.7	28.0	47.9
$\hat{\lambda}_1, \dots, \hat{\lambda}_{p+1}$	14.8	15.7	19.9	20.6	24.9	61.7	82.5
n_1, \dots, n_{p+1}	112	22	10	2	4	6	1
p-value for t_1, \dots, t_p	–	0.809	0.529	0.966	0.840	0.191	0.798

- BE 2nd iteration ($p = 5$):

t_0, \dots, t_p	0	12.7	18.6	24.7	28.0	47.9
$\hat{\lambda}_1, \dots, \hat{\lambda}_{p+1}$	14.8	15.7	20.0	24.9	61.7	82.7
p-value for t_1, \dots, t_p	–	0.809	0.493	0.707	0.191	0.798

- BE 3rd iteration ($p = 4$):

t_0, \dots, t_p	0	18.6	24.7	28.0	47.9
$\hat{\lambda}_1, \dots, \hat{\lambda}_{p+1}$	15.0	20.0	24.9	61.7	82.5
p-value for t_1, \dots, t_p	–	0.317	0.707	0.191	0.798

Demo of our method in the NSCLC example

- BE 4th iteration ($p = 3$):

t_0, \dots, t_p	0	18.6	24.7	28.0
$\widehat{\lambda}_1, \dots, \widehat{\lambda}_{p+1}$	15.0	20.0	24.9	64.6
p-value for t_1, \dots, t_p	–	0.317	0.707	0.162

- BE 5th iteration ($p = 2$):

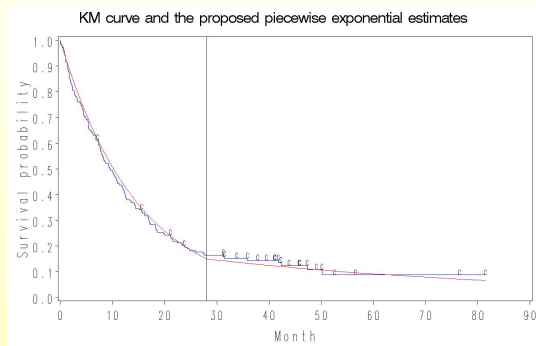
t_0, \dots, t_p	0	18.6	28.0
$\widehat{\lambda}_1, \dots, \widehat{\lambda}_{p+1}$	15.0	21.3	64.6
p-value for t_1, \dots, t_p	–	0.167	0.011

- BE 6th iteration ($p = 1$):

t_0, \dots, t_p	0	28.0
$\widehat{\lambda}_1, \dots, \widehat{\lambda}_{p+1}$	15.6	64.6
n_1, \dots, n_{p+1}	150	7
p-value for t_1, \dots, t_p	–	0.000

Demo of our method in the NSCLC example

- $(\hat{\lambda}_1, \hat{\lambda}_2) = (15.6, 64.6)$; $t_1 = 28.0$.



- $S(t_1 | \lambda_1) \approx 0.17 < 1/2$.
Thus the estimated median survival: $10.8 = \log 2 \times \hat{\lambda}_1$;
95% CI of the median survival: (9.6, 12.4).

Summary and future research

- **By integrating the three components, the proposed method can greatly improve survival modeling.**
- Four of the future research topics are
 - More generalized piecewise models (e.g., piecewise Gompertz model);
 - Instead of PAVA, using umbrella alternatives for certain data sets (e.g., breast cancer data);
 - Obtain α^* formulae for different censoring rates;
 - Goodness-of-fit quantitative assessments of models.