BATHTUB HAZARD MODEL WITH COVARIATES IN THE PRESENCE OF RIGHT- AND INTERVAL-CENSORED DATA

Idari Ismail
INTRODUCTION

LITERATURE REVIEW

METHODOLOGY

RESULTS
   Simulation study
   Real Data Analysis

CONCLUSION & FUTURE WORK
INTRODUCTION
INTRODUCTION

- The specialized fields of mathematical statistics - developed to deal with the special type of time-to-event random variables.
- Reliability analysis - methods related to assessment and prediction of successful operation or performance of products.
Survival analysis is the analysis of time-to-event data.
Deals with modelling and analysis of lifetime data (survival data or failure time).
INTRODUCTION

**CENSORING**

- ♠ Occurs when we do not know the exact time-to-event for an included observation
- ♠ Due to incomplete observations

**TYPES OF CENSORING**

- ✓ Right censoring
- ✓ Left censoring
- ✓ Interval censoring
INTRODUCTION

Failure Rate Function

- Increasing
- Decreasing
- Constant
- Bathtub-shaped
- Upside down bathtub-shaped
INTRODUCTION

Bathtub-shaped failure rates

- Gamma or Weibull - accommodate monotone failure rates.
- Failure rates firstly decrease, then stagnant at a constant level and eventually increase - resembles a bathtub, and thus, known as bathtub-shaped failure rate.
Bathtub-shaped failure rates

Can be observed when studying the lifespan of an industrial product or the lifetime of a biological entity (Dimitrakopoulou et al., 2007).

Example considering a high failure rate in infant mortality which decreases to a certain level, then remains constant for some time, and eventually increases (Gaver & Acar, 1979).

Previous studies

Gaver & Acar (1979), Mudholkar & Srivastava (1993), and Smith & Bain (1975) among others.
Bathtub hazard model

Proposed by Chen (2000)
Bathtub-shaped and increasing-depending on its parameter.

Chen (2000)
no two-parameter distribution that the failure rate exhibit bathtub-shaped.
LITERATURE REVIEW
LITERATURE REVIEW

Wu et al. (2004)

simple method for conducting statistical test with regards to the shape parameter where the method can be applied for a type-II right-censored data.

Wu (2008)

discussed exact confidence interval and exact join confidence region for the parameters under progressive type-II censored sample.
LITERATURE REVIEW

Wang et al. (2014)
Type II censored sample - discussed interval estimations for the parameters in bathtub hazard model.

Sarhan et al. (2012)
parameter estimation of the bathtub hazard model by using maximum likelihood and Bayes method.
LITERATURE REVIEW

Sarhan & Mustafa (2022)
- developed a new lifetime distribution based on bathtub hazard model and generalized exponential distribution.
- parameter estimation using maximum likelihood method and Bayesian procedures.

Chen & Gui (2020); Zhang & Gui (2022)
- discussed parameter estimation of bathtub hazard model and presented confidence intervals for the model’s parameters.
extensive research has been undertaken to study the bathtub hazard model.

the research to date has not focussed on investigating the bathtub hazard model in the presence of interval censored data.

extend the bathtub hazard model by incorporating covariates in the presence of right- and interval-censored data.
Real Data Application – breast cancer data

Midpoint imputation was applied – compared with no imputation
METHODOLOGY
**METHODOLOGY**

Cumulative Hazard Function

\[ F(t) = 1 - e^{-\lambda \left(1 - e^{t\alpha}\right)} , \quad t \geq 0 \]

Probability Density Function

\[ f(t; \lambda, \alpha) = \lambda \alpha t^{\alpha-1} e^{t\alpha} \left(1 - e^{t\alpha}\right) \]

Survival Function

\[ S(t; \lambda, \alpha) = e^{-\lambda \left(1 - e^{t\alpha}\right)} \]

Bathtub hazard model

Hazard Function

\[ h(t; \lambda, \alpha) = \lambda \alpha t^{\alpha-1} e^{t\alpha} \]

The failure rate function becomes bathtub-like as \( \lambda < 1 \) and is increasing when \( \lambda \geq 1 \).
Let the parameter $\lambda$ be a function of the covariates

$$\lambda_i = e^{-\beta_0 - \beta_1 x_i}$$

The failure rate function for a data set with a fixed covariate $x_i$ where $i = 1, 2, ..., n$

$$h(t) = e^{-\beta_0 - \beta_1 x_i} \alpha t_i^{\alpha-1} e^{t_i^\alpha}$$
The likelihood function for the full sample if there are no censored observation is:

\[
L(\theta) = \prod_{i=1}^{n} f(t_i) = \prod_{i=1}^{n} \left\{ e^{(-\beta_0 - \beta_1 x_i)} \alpha t_i^{\alpha-1} e^{t_i \alpha} e^{e^{(-\beta_0 - \beta_1 x_i)}(1-e^{t_i \alpha})} \right\}
\]
Censoring Indicator

- $\delta_{I_i} = 1$ if the subject is interval-censored, 0 otherwise
- $\delta_{R_i} = 1$ if the subject is right censored, 0 otherwise
- $\delta_{E_i} = 1$ if subject is not censored (exact survival time is observed), 0 otherwise
The likelihood function for right-, interval-censored or uncensored data without any imputation is,

\[
L(\theta) = \prod_{i=1}^{n} \left[ f(t_i) \right]^{\delta_{E_i}} \left[ S(t_{R_i}) \right]^{\delta_{R_i}} \left[ S(t_{L_i}) - S(t_{R_i}) \right]^{\delta_{I_i}} \\
= \prod_{i=1}^{n} \left[ e^{(-\beta_0 - \beta_1 x_i)} \alpha t_i^{\alpha-1} e^{\alpha} e^{(-\beta_0 - \beta_1 x_i)(1-e^{t_i^{\alpha}})} \right]^{\delta_{E_i}} \\
\times \left[ e^{(-\beta_0 - \beta_1 x_i)(1-e^{t_{R_i}^{\alpha}})} \right]^{\delta_{R_i}} \times \left[ e^{(-\beta_0 - \beta_1 x_i)(1-e^{t_{L_i}^{\alpha}})} - e^{-e^{e}} \right]^{\delta_{I_i}}
\]
Log-likelihood function

\[ \ell(\theta) = \sum_{i=1}^{n} \delta_{E_i} \left[ (-\beta_0 - \beta_1 x_i) + \ln(\alpha) + (\alpha - 1) \ln(t_i) + t^\alpha + e^{(-\beta_0 - \beta_1 x_i)}(1 - e^{t^\alpha_i}) \right] \\
+ \delta_{R_i} \left[ e^{(-\beta_0 - \beta_1 x_i)}(1 - e^{t_{R_i}^\alpha}) \right] + \delta_{I_i} \ln \left[ e^{(-\beta_0 - \beta_1 x_i)} \left( 1 - e^{t_{I_i}^\alpha} \right) \right] - e^{(-\beta_0 - \beta_1 x_i)} \left( 1 - e^{t_{R_i}^\alpha} \right) \]
The likelihood function for right-, interval- censored or uncensored with imputation method is,

\[
L(\theta) = \prod_{i=1}^{n} \left[ f(t_i) \right]^{\delta_{E_i}} \left[ S(t_{R_i}) \right]^{\delta_{R_i}} f(\tilde{t}_i)^{\delta_{I_i}} \\
= \prod_{i=1}^{n} \left[ e^{-\beta_0 - \beta_1 x_i} \alpha_i t_i^{\alpha-1} e^\alpha e^{-\beta_0 - \beta_1 x_i} \left(1-e^{-\beta_1 x_i} t_i^\alpha\right) \right]^{\delta_{E_i}} \left[ e^{-\beta_0 - \beta_1 x_i} \left(1-e^{-\beta_1 x_i} t_i^\alpha\right) \right]^{\delta_{R_i}} \\
X \left[ e^{-\beta_0 - \beta_1 x_i} \tilde{t}_i^{\alpha-1} e^\alpha e^{-\beta_0 - \beta_1 x_i} \left(1-e^{-\beta_1 x_i} \tilde{t}_i^\alpha\right) \right]^{\delta_{I_i}}
\]
Log-likelihood function

\[ \ell(\theta) = \sum_{i=1}^{n} \delta_{E_i} \left[ (-\beta_0 - \beta_1 x_i) + \ln(\alpha) + (\alpha - 1) \ln(t_i) + t_i^\alpha + e^{-\beta_0 - \beta_1 x_i} (1 - e^{t_i^\alpha}) \right] \\
+ \delta_{R_i} \left[ e^{-\beta_0 - \beta_1 x_i} (1 - e^{t_{R_i}^\alpha}) \right] \\
+ \delta_{I_i} \left[ (-\beta_0 - \beta_1 x_i) + \ln(\alpha) + (\alpha - 1) \ln(\tilde{t}_i) + \tilde{t}_i^\alpha + e^{-\beta_0 - \beta_1 x_i} (1 - e^{\tilde{t}_i^\alpha}) \right] \]
Briefly elaborate on what you want to discuss.

Midpoint imputation - the intervals are replaced by the midpoint.

Imputation technique - has an advantage over other methods due to its simplicity and ease of implementation (Chen & Sun, 2010).
RESULTS
RESULTS

Simulation Study

\[ N = 1000 \]

\[ n = 50, 100, 150, 250 \]

The values of the parameters \( \alpha, \beta_0 \) and \( \beta_1 \) were particularly set at 0.4, 3.3 and 0.9 respectively.

R programming language
Simulation study

Step 1
Generate covariate values $x_i$ from a standard normal distribution

Step 2
Generate a sequence of random numbers $u_i$ from a standard uniform distribution on the unit interval $(0,1)$ to obtain the event time, $t_i$ for $i = 1, 2, ..., n$

Step 3
Generate censoring times, $c_i$ from an exponential distribution with the value of $\mu$ would be modified to obtain the desired censoring proportion (cp).

Step 4
Generate survival time $t_i$
Simulation Study

Generate survival time $t_i$

$$t_i = \left( \ln \left( 1 - \frac{\ln (1 - U_i)}{e^{(-\beta_0 - \beta_1 x_i)}} \right) \right)^{1/\alpha}$$
### RESULTS

#### Summary of standard error (SE) values for parameter estimates

<table>
<thead>
<tr>
<th>Estimates</th>
<th>n</th>
<th>No Imputation</th>
<th>Midpoint Imputation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>cp = 0</td>
<td>cp = 10%</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>50</td>
<td>0.02366</td>
<td>0.02622</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.01598</td>
<td>0.01726</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>0.01313</td>
<td>0.01431</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.01000</td>
<td>0.01070</td>
</tr>
<tr>
<td>$\hat{\beta}_0$</td>
<td>50</td>
<td>0.34859</td>
<td>0.35192</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.23250</td>
<td>0.24050</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>0.19378</td>
<td>0.19813</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.14554</td>
<td>0.15099</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>50</td>
<td>0.19702</td>
<td>0.20701</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.13490</td>
<td>0.13417</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>0.10798</td>
<td>0.10984</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.08325</td>
<td>0.08127</td>
</tr>
</tbody>
</table>
### Results

#### Summary of root mean square error (RMSE) values for parameter estimates

<table>
<thead>
<tr>
<th>Estimates</th>
<th>n</th>
<th>No Imputation</th>
<th></th>
<th></th>
<th>Midpoint Imputation</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>cp = 0</td>
<td>cp = 10%</td>
<td>cp = 20%</td>
<td>cp = 30%</td>
<td>cp = 0</td>
<td>cp = 10%</td>
</tr>
<tr>
<td>$\alpha$</td>
<td>50</td>
<td>0.02469</td>
<td>0.02692</td>
<td>0.02752</td>
<td>0.04108</td>
<td>0.02481</td>
<td>0.02448</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.01651</td>
<td>0.01790</td>
<td>0.03225</td>
<td>0.03607</td>
<td>0.01649</td>
<td>0.01632</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>0.01329</td>
<td>0.01634</td>
<td>0.02426</td>
<td>0.03574</td>
<td>0.01329</td>
<td>0.01366</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.01010</td>
<td>0.01391</td>
<td>0.02391</td>
<td>0.04283</td>
<td>0.01010</td>
<td>0.01092</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>50</td>
<td>0.36208</td>
<td>0.35199</td>
<td>0.36146</td>
<td>0.38923</td>
<td>0.36461</td>
<td>0.35911</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.23819</td>
<td>0.24060</td>
<td>0.28328</td>
<td>0.28088</td>
<td>0.23791</td>
<td>0.23926</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>0.19541</td>
<td>0.20160</td>
<td>0.22416</td>
<td>0.25231</td>
<td>0.19546</td>
<td>0.19498</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.14630</td>
<td>0.15929</td>
<td>0.19029</td>
<td>0.24768</td>
<td>0.14626</td>
<td>0.15030</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>50</td>
<td>0.20190</td>
<td>0.21165</td>
<td>0.21389</td>
<td>0.22505</td>
<td>0.20183</td>
<td>0.20307</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.13617</td>
<td>0.13527</td>
<td>0.15815</td>
<td>0.15488</td>
<td>0.13623</td>
<td>0.14236</td>
</tr>
<tr>
<td></td>
<td>150</td>
<td>0.10897</td>
<td>0.11018</td>
<td>0.11420</td>
<td>0.11811</td>
<td>0.10905</td>
<td>0.11000</td>
</tr>
<tr>
<td></td>
<td>250</td>
<td>0.08368</td>
<td>0.08127</td>
<td>0.08766</td>
<td>0.09195</td>
<td>0.08370</td>
<td>0.08428</td>
</tr>
</tbody>
</table>
RESULTS

Line plot of SE for each parameter estimate at cp=0

Line plot of SE for each parameter estimate at cp=30
RESULTS

Line plot of SE for each parameter estimate at n=50

Line plot of SE for each parameter estimate at n=250
RESULTS

Line plot of RMSE for each parameter estimate at cp=0

Line plot of RMSE for each parameter estimate at cp=0
RESULTS

Line plot of RMSE for each parameter estimate at n=50

Line plot of RMSE for each parameter estimate at n=250
Study design

Objective

Sample

Event of interest

retrospective study by Finkelstein & Wolfe (1985).

To investigate the comparison between radiation therapy (RT) alone and radiation therapy combined with adjuvant chemotherapy (RCT).


time to cosmetic deterioration
 REAL DATA ANALYSIS

59.6%
Interval-censored

40.4%
Right-censored
## REAL DATA ANALYSIS

### Summary of maximum likelihood estimates for bathtub hazard model with and without covariate

<table>
<thead>
<tr>
<th>Parameters</th>
<th>Estimate (Standard error)</th>
<th>t value</th>
<th>Pr(&gt;t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Without covariate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.4049(0.0216)</td>
<td>18.921</td>
<td>0.000</td>
</tr>
<tr>
<td>$\lambda$</td>
<td>0.0123(0.0043)</td>
<td>2.857</td>
<td>0.00427</td>
</tr>
<tr>
<td>With covariate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\alpha$</td>
<td>0.4255(0.0218)</td>
<td>19.565</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_0$</td>
<td>4.1756(0.3548)</td>
<td>11.769</td>
<td>0.000</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.9585 (0.2821)</td>
<td>3.397</td>
<td>0.00068</td>
</tr>
</tbody>
</table>

### Descriptive Statistics of time to cosmetic deterioration by treatment

<table>
<thead>
<tr>
<th>Treatment</th>
<th>n</th>
<th>Mean</th>
<th>Standard error</th>
<th>Standard deviation</th>
<th>Median</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT</td>
<td>46</td>
<td>28.11</td>
<td>2.10</td>
<td>14.23</td>
<td>31.50</td>
</tr>
<tr>
<td>RCT</td>
<td>48</td>
<td>21.66</td>
<td>1.53</td>
<td>10.58</td>
<td>20.50</td>
</tr>
</tbody>
</table>

$H_0: \beta_1 = 0$

$H_1: \beta_1 \neq 0$
CONCLUSION
midpoint imputation outperforms since it obtains smaller SE and RMSE values for most cp levels and sample sizes.

- bathtub hazard model shows a good fit to the data
- the results indicate that the treatment received by breast cancer patients has a significant effect on time to development of cosmetic deterioration.
- patients receiving RCT have a higher risk of developing deterioration
- Beadle et al. (1984) - adjuvant chemotherapy tends to increase development of breast retraction
CONCLUSION

Future Work

• Include other imputation approach

Conclusion

• This study provides support with the notion that the lifetime model with bathtub-shaped failure rates has an important contribution in medical study.
• Demonstrates how data modelling could be undertaken considering the presence of interval-censored data.
THANK YOU

idarismail@gmail.com