Covariate Software Reliability Models and Applications

Lance Fiondella
Department of Electrical and Computer Engineering
Some high profile failures attributed to software

- 1991: MIM-104 Patriot Missile
  - 28 soldiers killed, 99 wounded
- 1996: Ariane 5 Flight 501
  - $370 million
- 1998: Mars Climate Orbiter
  - $193 million
- 1999: Mars Polar Lander
  - $110 million
Prominent reports on reliability

• 2002: NIST
  – The Economic Impacts of Inadequate Infrastructure for Software Testing (22.2,59.5) billion

• 2008: DSB
    • The single most important step ... ensure programs are formulated to execute a viable systems engineering strategy from the beginning, including a robust RAM program.

• 2014: National Academies
  – Reliability Growth: Enhancing Defense System Reliability
    • Recommended use of reliability growth models to direct contractor design and test activities
Percentage of system functions performed by software

F-4 (1960) 8%  F-16 (1982) 45%  F-22 (2000) 80%

Technical challenge: Stability of underlying model fitting algorithms

Initial estimates can behave erratically
Recent Contributions

- Developed open source software reliability tool (SFRAT) to automatically apply reliability models and scripts to further automate
- Expectation conditional maximization (ECM) algorithms
  - Eliminated computationally intensive integration from update rules of ECM
  - Guarantees dimensionality reduction
  - 200-400 factor speed up over past EM and ECM algorithms (conservative estimate)
Summary of frequencies by DTT type and build found

Extracted defects and change requests from major and minor versions exhibiting a large number of events
Cumulative failures

Plot enables comparison of data and model fits
Time between failures

Times between failures should increase (indicates reliability growth)
Failure intensity

Failure intensity should decrease (indicates reliability growth)
Nonhomogeneous Poisson process (NHPP) Software Reliability Model

• Stochastic process counts number of events observed as function of time
  – In context of software reliability (security), NHPP counts number of faults (vulnerabilities) detected by time $t$

• Counting process characterized by mean value function (MVF)
  – Form of MVF of several SRM: $m(t) = \omega \times F(t)$
    • $\omega$ - number of faults (vulnerabilities) that would be detected with indefinite testing
    • $F(t)$ - cumulative distribution function (CDF)
Weibull Software Reliability Model

• Substituting Weibull distribution for $F(t)$
  
  $$m(t) = \omega \left( 1 - e^{-bt^c} \right)$$

  – $b$ - scale parameter
  – $c$ - shape parameters

• $c = 1$ simplifies to exponential distribution
  – Also known as Goel-Okumoto Model

Primary limitation of traditional NHPP software reliability models: Do not explicitly identify underlying software testing activities that lead to fault discovery
Contributions

• A NHPP software reliability model possessing a discrete Cox proportional hazard rate to incorporate covariates
• Expectation conditional maximization (ECM) algorithms to efficiently estimate a model’s numerical parameters
• A generalization of the testing effort allocation problem to covariate models referred to as the optimal testing activity allocation problem to
  – (i) Maximize fault discovery within a budget constraint
  – (ii) Minimize budget required discover a specified number of additional faults
NHPP SRM with Covariates

- Mean value function

$$H_{n;\omega,\theta,\beta} = \omega \sum_{i=1}^{n} p_{i,x_i;\theta,\beta}$$

where

$$p_{i,x_i;\theta,\beta} = \left(1 - (1 - h^0_{i;\theta})g(x_i;\beta)\right) \prod_{k=1}^{i-1} \left(1 - h^0_{k;\theta}g(x_k;\beta)\right)$$

Covariates associated with $y_i$ in the $i^{th}$ testing interval

Vector of Cox model parameters

Vector of model parameters

Number of faults that would be detected with indefinite testing

Discrete Cox proportional hazard model

Baseline hazard function
Baseline Hazard Functions

• Geometric (GM):
  \[ h_{i;b}^0 = b \]
  – where \( b \in (0,1) \)

• Negative binomial of order two (NB):
  \[ h_{i;b}^0 = \frac{ib^2}{1 + b(i - 1)} \]
  – where \( b \in (0,1) \) and exponent 2 indicates order

• Discrete Weibull of order two (DW):
  \[ h_{i;b}^0 = 1 - b^{i^2 - (i-1)^2} \]
  – where \( b \in (0,1) \) and exponent 2 indicates order
Optimal test activity allocation to maximize fault discovery

• Given \( n \) intervals of observed data and budget \( B \) to allocate among \( r \) activities (covariates)
• Maximize total number of faults detected

\[
\text{arg max } \hat{H}_{(n+1); \omega, \theta, \beta}
\]
subject to

\[
\sum_{j=1}^{r} c_j x_{j,(n+1)} \leq B
\]

\[
c_j \left( x_{j,(n+1)} - x_{j,(n)} \right) \geq 0
\]

- \( c_j > 0 \) - cost associated with additional unit of activity \( j \)
Optimal test activity allocation to minimize cost

- Minimize cost to identify a $k$ faults

$$\text{arg min} \sum c_j x_j$$

subject to

$$\hat{H}_{n+1}; \omega, \theta, \beta - \hat{H}_n ; \omega, \theta, \beta \geq k$$

$$\beta_{j+1} - \beta_j \geq 0$$
Parameter Estimation: Maximum Likelihood Estimation (MLE)

- Maximizes likelihood function, also known as joint distribution of failure data
- Likelihood function
  \[ L(\theta, \beta, \omega) = \Pr(Y_1 = y_1, Y_2 = y_2, ..., Y_n = y_n) \]
  \[ = \prod_{i=1}^{n} \exp(-\omega p_{i,x_i;\theta,\beta}) \left( \omega p_{i,x_i;\theta,\beta} \right)^{y_i} \]
  \[ \frac{1}{y_i!} \]
- Log-likelihood function
  \[ LL(\theta, \beta, \omega) = -\omega \sum_{i=1}^{n} p_{i,x_i;\theta,\beta} + \sum_{i=1}^{n} y_i \ln(\omega) + \sum_{i=1}^{n} y_i \ln(p_{i,x_i;\theta,\beta}) - \sum_{i=1}^{n} \ln(y_i!) \]
Expectation Conditional Maximization Algorithm (ECM)

- Preserves monotonicity property of EM algorithm
- Replaces M-step of EM algorithm with $\nu$ conditional maximization (CM)-steps
  - $\nu$ - number of model parameters
  - Divides single $\nu$-dimensional problem into $\nu$ 1-dimensional problems
ECM Algorithm (2)

- (S.1) – Specify log-likelihood function
- (S.2) – Reduce log-likelihood function from $\nu$ to $\nu - 1$ parameter

$$\frac{\partial LL}{\partial \omega} = 0$$

and

$$\hat{\omega} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} p_{i,x_i;\theta,\beta}}$$

- Substitute for $\hat{\omega}$ in $LL$ function to obtain reduced log-likelihood (RLL) function
ECM Algorithm (3)

(S.3) – Derive conditional maximization steps for remaining $\nu - 1$ parameters by computing partial derivatives

$$\frac{\partial RLL}{\partial \theta} = 0$$

and

$$\frac{\partial RLL}{\partial \beta} = 0$$

$\theta$ - all parameters except $\omega$

$\beta$ – Coefficients related to all $p$ covariates
ECM Algorithm (4)

- (S.4) – Cycle through $(\nu - 1)$ CM-steps holding other $(\nu - 2)$ parameters constant until user specified convergence achieved
  - Applies numerical root finding algorithm in each step
  - Determines MLEs of $\hat{\theta}/\omega$
- (S.5) – Computes MLE of $\omega$ by substituting estimates of $\theta$ including $\beta$ into $\hat{\omega}$ equation
- Step (S.1) through (S.5) applied to different hazard functions
Goodness-of-fit measures

- Akaike Information Criterion (AIC)
  - Information theoretic goodness-of-fit measure
  - Quantifies trade-off between model precision and complexity
  - AIC of model $i$ function of maximized log-likelihood and number of model parameters
    \[ AIC_i = 2\nu - 2LL(x_i; \hat{\omega}, \hat{\theta}, \hat{\beta}) \]
  - Model $j$ better than model $i$ if $AIC_{i,j} = AIC_i - AIC_j > 2.0$
Goodness-of-fit measures (2)

- Bayesian Information Criterion (BIC)
  - Function of maximized log-likelihood, number of model parameters, and sample size $n$
    \[ BIC_i = -2LL(x_i; \hat{\omega}, \hat{\theta}, \hat{\beta}) + v \log(n) \]
  - Penalty term proportional to number of parameters multiplied by logarithm of sample size
Goodness-of-fit measures (3)

- Sum of Squares Error (SSE)
  - Also known as residual sum of squares
  - For failure count data
    \[ SSE = \sum_{i=1}^{n} \left( \hat{H}_{i;\omega,\theta,\beta} - Y_i \right)^2 \]
    - \( Y_i = \sum_{j=1}^{i} y_i \) - cumulative number of failures observed in first \( i \) time intervals
Goodness-of-fit measures (4)

• Predictive Sum of Squares Error (PSSE)
  – Compares predictions of model with data not used to perform model fitting
    \[
    PSSE = \sum_{i=n-\ell+1}^{n} (\hat{H}_{i;\omega,\theta,\beta} - Y_i)^2
    \]
  – Maximum likelihood estimates of model parameters determined from first \((n - \ell)\) intervals
  – Common choice for \(\ell\) approximately 10% of data
Model Selection based on Multiple Goodness-of-fit Measures

• Consider $n$ models and $m$ measures
  – Assign normalized score to each measure based on Critic method
    \[ x_{i,j} = \frac{f_{i,j} - f_j^+}{f_j^- - f_j^+} \]
    – $f_j^+$ ($f_j^-$) - best (worst) value of measure $j$ across all models
    – $x_{i,j}$ - indicates how close $j^{th}$ measure of model $i$ to ideal

• Compute median or mean of normalized scores
  – Recommend model with highest scores
Illustrations

- DS1 and DS2 with three covariates considered
  - DS1 with \( n = 17 \) intervals
  - DS2 with \( n = 14 \) intervals
- Covariates considered – E, F, C
  - E – Execution time
  - F – Failure identification work
  - C – Computer time identification work
- Hazard functions considered
  - GM, NB, and DW

<table>
<thead>
<tr>
<th>week</th>
<th>Execution Time (hr)</th>
<th>Failure Identification Work (person hr)</th>
<th>Computer Time-Ident. (hr)</th>
<th>Failure Identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.0531</td>
<td>4</td>
<td>1.0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>.0619</td>
<td>20</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>.1580</td>
<td>1</td>
<td>0.5</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>.0810</td>
<td>1</td>
<td>0.5</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1.0460</td>
<td>32</td>
<td>2.0</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>1.7500</td>
<td>32</td>
<td>5.0</td>
<td>9</td>
</tr>
<tr>
<td>7</td>
<td>2.9600</td>
<td>24</td>
<td>4.5</td>
<td>6</td>
</tr>
<tr>
<td>8</td>
<td>4.9700</td>
<td>24</td>
<td>2.5</td>
<td>7</td>
</tr>
<tr>
<td>9</td>
<td>0.4200</td>
<td>24</td>
<td>4.0</td>
<td>4</td>
</tr>
<tr>
<td>10</td>
<td>4.7000</td>
<td>30</td>
<td>2.0</td>
<td>3</td>
</tr>
<tr>
<td>11</td>
<td>0.9000</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>1.5000</td>
<td>8</td>
<td>4.0</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>2.0000</td>
<td>8</td>
<td>6.0</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>1.2000</td>
<td>12</td>
<td>4.0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>1.2000</td>
<td>20</td>
<td>6.0</td>
<td>2</td>
</tr>
<tr>
<td>16</td>
<td>2.2000</td>
<td>32</td>
<td>10.0</td>
<td>2</td>
</tr>
<tr>
<td>17</td>
<td>7.6000</td>
<td>24</td>
<td>8.0</td>
<td>3</td>
</tr>
</tbody>
</table>

Total: 32.8000 296 60.0 54
Application of ECM to Covariate NHPP SRM: GM (F) on DS2

ECM takes 90° steps in search space
Log-likelihood of covariate model: GM (F) on DS2

ECM algorithm converges monotonically
Model Fit: GM (F) on DS2

Model prediction function of covariates in each interval
**Goodness-of-fit measures for GM on DS1 and DS2**

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\nu$</th>
<th>DS1</th>
<th>DS2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>LLF</td>
<td>AIC</td>
</tr>
<tr>
<td>-</td>
<td>2</td>
<td>-41.47</td>
<td>86.94</td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>-36.13</td>
<td>78.25</td>
</tr>
<tr>
<td>F</td>
<td>3</td>
<td>-31.21</td>
<td>68.43</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>-35.54</td>
<td>77.09</td>
</tr>
<tr>
<td>EF</td>
<td>4</td>
<td>-30.03</td>
<td>68.05</td>
</tr>
<tr>
<td>FC</td>
<td>4</td>
<td>-30.89</td>
<td>69.79</td>
</tr>
<tr>
<td>EC</td>
<td>4</td>
<td>-30.08</td>
<td>68.16</td>
</tr>
<tr>
<td>EFC</td>
<td>5</td>
<td>-28.40</td>
<td>66.81</td>
</tr>
</tbody>
</table>

No combination of covariates preferred by all measures
Model Selection – Normalized scores – GM on DS1

<table>
<thead>
<tr>
<th></th>
<th>–</th>
<th>E</th>
<th>F</th>
<th>C</th>
<th>EF</th>
<th>FC</th>
<th>EC</th>
<th>EFC</th>
</tr>
</thead>
<tbody>
<tr>
<td>LLF</td>
<td>0.00</td>
<td>0.41</td>
<td>0.79</td>
<td>0.45</td>
<td>0.88</td>
<td>0.81</td>
<td>0.87</td>
<td>1.00</td>
</tr>
<tr>
<td>AIC</td>
<td>0.00</td>
<td>0.43</td>
<td>0.92</td>
<td>0.49</td>
<td>0.94</td>
<td>0.85</td>
<td>0.93</td>
<td>1.00</td>
</tr>
<tr>
<td>BIC</td>
<td>0.00</td>
<td>0.44</td>
<td>1.00</td>
<td>0.51</td>
<td>0.97</td>
<td>0.88</td>
<td>0.97</td>
<td>0.99</td>
</tr>
<tr>
<td>SSE</td>
<td>0.00</td>
<td>0.62</td>
<td>0.79</td>
<td>0.54</td>
<td>0.91</td>
<td>0.77</td>
<td>0.97</td>
<td>1.00</td>
</tr>
<tr>
<td>PSSE</td>
<td>1.00</td>
<td>0.94</td>
<td>0.68</td>
<td>0.95</td>
<td>0.79</td>
<td>0.00</td>
<td>0.98</td>
<td>0.62</td>
</tr>
<tr>
<td>Median</td>
<td></td>
<td>0</td>
<td>0.44</td>
<td>0.79</td>
<td>0.51</td>
<td>0.91</td>
<td>0.81</td>
<td>0.97</td>
</tr>
<tr>
<td>Mean</td>
<td>0.2</td>
<td>0.57</td>
<td>0.84</td>
<td>0.59</td>
<td>0.91</td>
<td>0.66</td>
<td>0.94</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Must exercise caution when selecting goodness of fit measures
Model Selection – Overall Recommendation

<table>
<thead>
<tr>
<th>Method</th>
<th>DS1</th>
<th></th>
<th></th>
<th>DS2</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Median</td>
<td>NB (F)</td>
<td>NB (EF)</td>
<td>NB (EFC)</td>
<td>GM (F)</td>
<td>GM (FC)</td>
<td>GM (EF)</td>
</tr>
<tr>
<td>Mean</td>
<td>NB (F)</td>
<td>NB (EF)</td>
<td>NB (EFC)</td>
<td>GM (F)</td>
<td>GM (EF)</td>
<td>GM (FC)</td>
</tr>
</tbody>
</table>

Must consider preferred combination of covariates and practical interpretation
## Optimal Test Activity Allocation

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$\hat{H}^*_{n; \omega, \theta, \beta}$</th>
<th>%$E$</th>
<th>%$F$</th>
<th>%$C$</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td>6.63</td>
<td>100</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>F</td>
<td>2.61</td>
<td>—</td>
<td>100</td>
<td>—</td>
</tr>
<tr>
<td>C</td>
<td>0.41</td>
<td>—</td>
<td>—</td>
<td>100</td>
</tr>
<tr>
<td>EF</td>
<td>5.30</td>
<td>100</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>FC</td>
<td>1.97</td>
<td>—</td>
<td>5.70</td>
<td>94.30</td>
</tr>
<tr>
<td>EC</td>
<td>0.62</td>
<td>40.75</td>
<td>—</td>
<td>59.25</td>
</tr>
<tr>
<td>EFC</td>
<td>1.55</td>
<td>38.03</td>
<td>—</td>
<td>61.97</td>
</tr>
</tbody>
</table>

In practice, allocation may be performed on preferred model.
Sensitivity of Optimal Allocation
NB (EFC) on DS1

Percentage of effort allocated to each activity converges as budget increases
Marginal Utility of NB on DS1

Marginal utility of each activity explains allocation
Optimal test activity allocation to expose $k$ additional faults

Performs better than allocation in final interval of original study, which only exposed two additional faults
Conclusions

• Proposed
  – NHPP software reliability model possessing discrete Cox proportional hazard rate to incorporate covariates
  – ECM algorithms to efficiently fit model
  – *Optimal testing activity allocation problem* to (i) maximize fault discovery within a budget constraint and (ii) minimize budget required to discover a specified number of additional faults

• Explicitly links underlying software testing activities that lead to fault discovery
Future Work & Research

• Tool implementation and automation to promote adoption
• Application to data from past and present programs (DoD, NASA, etc…) to make practical refinements
• Application in context of cybersecurity testing
• Process related studies to identify which combination of testing methods are most effective for a given class of systems
Software reliability tools

• Availability
  – Present (R)
    • https://sasdlc.org/lab/#/research
  – Forthcoming (R and Python)
    • https://lfiondella.sites.umassd.edu/

• Includes
  – Example failure data
  – Link to Github repository
  – Documentation
Acknowledgements

- This work was partially supported by (i) NAVAIR under contract N00421-16-P-0521, (ii) the National Science Foundation under awards #1526128 and #1749635, and NASA under grants #80NSSC18K0154 and 80NSSC20K0276.