Performance of Empirical Bayes Estimation Techniques Used in Probabilistic Risk Assessment on Failure Data collected in U.S NRC Reactor Operating Experience Database

Andrei Gribok, Vivek Agarwal, Vaibhav Yadav

2018 Probabilistic Safety Assessment & Management Conference (PSAM 14)
Los Angeles, CA, USA September 16-21, 2018
Two Parts of PRA Analysis

Level 1: Population variability

Level 2: Plant-specific data

- Plant 1, $\lambda_1$, $t_1$ $\rightarrow$ $X_1 \sim \text{Poisson}(\lambda_1 t_1)$
- Plant 2, $\lambda_2$, $t_2$ $\rightarrow$ $X_2 \sim \text{Poisson}(\lambda_2 t_2)$
- ...$\rightarrow$ $X_n \sim \text{Poisson}(\lambda_m t_m)$

Population-variability distribution, $g$
Bayesian analysis depends on prior elicitation/selection

Prior elicitation/selection is a fundamental problem of Bayesian inference

To produce a prior we need to convert information about observable values (number of failures, time) into information about unobservable variables (probability of failure, failure rate)

Is there the best prior? Is there a best method to elicit prior? Is there the best method for parameter estimation?

The historical data is our best, most objective, and reliable source of information about true probability of failure or failure rate

The credibility of PRA analysis depends on how closely it reflects industry-wide historical data

*ALI MOSLEH, “PRA: A PERSPECTIVE ON STRENGTHS, CURRENT LIMITATIONS, AND POSSIBLE IMPROVEMENTS”, NUCLEAR ENGINEERING AND TECHNOLOGY, VOL.46 NO.1 FEBRUARY 2014
Bayesian Parameter Estimation

Bayesian inference

\[
\pi(\lambda/x, \alpha) = \frac{L(x/\lambda) \cdot \pi(\lambda/\alpha)}{\int L(x/\lambda) \cdot \pi(\lambda/\alpha) d\lambda}
\]

Prior predictive distribution

\[
\pi(x/\alpha) = \int_0^\infty L(x/\lambda) \cdot \pi(\lambda/\alpha) d\lambda
\]

Empirical Bayes

\[
\pi(\lambda/x, \alpha) = \frac{L(x/\lambda) \cdot \pi(\lambda/\hat{\alpha})}{\int L(x/\lambda) \cdot \pi(\lambda/\hat{\alpha}) d\lambda}
\]
Gamma distribution is a continuous distribution.

Poisson distribution is a discrete distribution.

Negative binomial (NB) distribution is a discrete distribution.
Prior Predictive Distribution for Gamma-Poisson Compound Distribution Model

\[ P(X = x / \alpha, \beta, T) = \text{NB} \left( x; \alpha, \frac{T}{\beta + T} \right) \]

\[ = \frac{\Gamma(x + \alpha)}{x! \Gamma(\alpha)} \cdot \left[ \frac{T}{\beta + T} \right]^x \cdot \left[ \frac{\beta}{\beta + T} \right]^\alpha \]

\[ = \int_0^\infty \frac{(\lambda \cdot T)^x}{x!} \cdot e^{-\lambda \cdot T} \cdot \frac{\beta^\alpha \cdot \lambda^{\alpha-1} \cdot e^{-\lambda \cdot \beta}}{\Gamma(\alpha)} d\lambda \]
Why Do We Need Nonparametric Bayes?

\[
\text{Gamma}(\lambda/\alpha, \beta) = \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda \beta}}{\Gamma(\alpha)}, \quad \mu(\lambda) = \frac{\alpha}{\beta}, \quad \sigma(\lambda) = \sqrt{\frac{\alpha}{\beta}}
\]
Performance of Different Estimation Techniques on NROD MDP FR Data

- RMSE = $\sqrt{\frac{1}{95} \sum_{i=1}^{95} (\lambda_i^{est} - \lambda_i^{true})^2}$
- Posterior mean is posterior summary statistics
- Plant-specific MLE
- Industry average MLE
- The James–Stein (JS) Estimator
- Jeffreys noninformative prior, Gamma(0.5,0)
- Constrained noninformative prior (CNI), Gamma (0.5,0.5·Industry average)
- Method of moments
- Maximum likelihood fitting of Gamma ($\alpha,\beta$)
- Maximum Likelihood II (MLE2)
- Nonparametric empirical Bayses (NPFR)
• NROD contains failure data for thousands of components
• The data typically span the period from 1998 to 2015 (2016, 2017 is now available)
• Binomial data are presented as number of failures and number of demands
• Poisson data are presented as number of failures and time period
### NROD Data for Motor-Driven Pump (MDP)
#### Failure to Run (FR)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Plant 1</td>
<td>3</td>
<td>2</td>
<td>54938.82</td>
<td>2</td>
<td>268467.12</td>
</tr>
<tr>
<td>Plant 2</td>
<td>2</td>
<td>0</td>
<td>26280.00</td>
<td>1</td>
<td>131400.00</td>
</tr>
<tr>
<td>Plant 3</td>
<td>6</td>
<td>0</td>
<td>88100.52</td>
<td>0</td>
<td>346328.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Plant 94</td>
<td>9</td>
<td>5</td>
<td>169077.24</td>
<td>4</td>
<td>918448.20</td>
</tr>
<tr>
<td>Plant 95</td>
<td>12</td>
<td>3</td>
<td>139221.72</td>
<td>4</td>
<td>816829.16</td>
</tr>
<tr>
<td>Total</td>
<td>624</td>
<td>55</td>
<td>9312985.75</td>
<td>168</td>
<td>50313464.88</td>
</tr>
</tbody>
</table>

$$\lambda_i^{est} = \frac{(Failure Count)_i}{(Run hours)_i}$$

$$\lambda_i^{true} = \frac{(Failure Count)_i}{(Run hours)_i}$$
The chi-square test of homogeneity (data pullability)

- Test if population variability distribution is not a degenerate one
- $H_0: \lambda_1 = \lambda_2 = \ldots = \lambda_k$, all failure rates are the same
- $H_1: \lambda_1 \neq \lambda_2 = \ldots = \lambda_k$, at least one $\lambda_i$ is different from others

\[
\chi^2 = \sum_{i=1}^{k} \left( \frac{\lambda_i^{obs} - \lambda_i^{exp}}{\lambda_i^{exp}} \right)^2,
\text{ if } \chi^2 > \chi^2_{0.95}(df=k-1) \text{ then } H_0 \text{ is rejected and } H_1 \text{ is accepted}
\]
- For MDP FR data, $\chi^2 = 143.5877$, $\chi^2_{0.95}(94) = 117.6317$
Performance of Different Estimation Techniques on NROD MDP FR Data

RMSE and 95% CI between "true" failure rate and its estimates

- MLE
- Industry average
- JS
- Jeffreys
- CNI
- Moments Gamma
- MLE
- MLE2
- NPEB

RMSE

$\times 10^{-10}$
Performance of Different Parameter Estimation Methods for Different Components on NROD Data Set.

- ACX_FR: Accumulator Fail to Run (ACX)
- AHU_NR_FTR: Air Handling Unit Normally Running Fails to Run (AHU)
- CHL_FR: Chiller Unit Fails to Run (CHL)
- MDC_FR: Motor Driven Compressor Fail to Run (MDC)
- MDP_FR: Motor Driven Pump Fail to Run (MDP)
Empirical Bayes is a shrinkage estimator. It is biased.
Why do not we always use EB in lieu of MLE?

MLE is better for 38 plants, including 13 nonzero $\lambda$s
Number of plants for which Plant-Specific MLE Outperforms NPEB

<table>
<thead>
<tr>
<th>Component</th>
<th># of plants in NROD database</th>
<th># plants for which NPEB has larger RMSE than plant-specific MLE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACX</td>
<td>34</td>
<td>13</td>
</tr>
<tr>
<td>AHU</td>
<td>34</td>
<td>14</td>
</tr>
<tr>
<td>CHL</td>
<td>22</td>
<td>3</td>
</tr>
<tr>
<td>MDC</td>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>MDP</td>
<td>99</td>
<td>33</td>
</tr>
</tbody>
</table>
Conclusions

- Nonparametric EB method is more flexible, allowing to change prior’s mean and variance independently
- Nonparametric EB method requires selection of two parameters: kernel width for kernel density estimation and regularization parameter for Bayesian deconvolution
- EB and hierarchical Bayes improve industry-wide performance with respect to plant-specific MLE, however, they do not guarantee an improvement for a specific plant
- Single-stage noninformative prior is likely to degrade the accuracy of the estimate in comparison to MLE
- Bayesian methods are shrinkage (biased) estimators. The variance of the estimate is shrunk toward zero and the value of the estimator is shrunk towards the mean value of prior distribution
- If we need industry-wide improvement for the accuracy of parameter estimates, EB is the way to go, however, if we need to improve estimate for a specific plant, all options should be considered
Fredholm Integral Equation of the First Kind
Bayesian Deconvolution

\[ g(x) = \int_a^b K(\lambda, x) f(\lambda) d\lambda, \quad c \leq x \leq d \]

\[ \hat{g}(x)_{\text{predictive}} = \int_0^\infty \frac{(\lambda \cdot T)^x}{x!} e^{-\lambda \cdot T} \cdot f(\lambda) \ d\lambda, \ 0 \leq \lambda \leq \infty, T \text{ specified}, x = 0, 1, 2 ... \]

\[
\begin{align*}
\arg\min_{f(\lambda)} & \left\{ \left\| \int_0^\infty \frac{(\lambda \cdot T)^x}{x!} e^{-\lambda \cdot T} \cdot f(\lambda) \ d\lambda - \hat{g}(x)_{\text{predictive}} \right\|_2^2 
+ \varepsilon^2 \int_0^\infty \left( \frac{d^2 f(\lambda)}{d\lambda^2} \right)^2 d\lambda \right\}
\end{align*}
\]
Discretization of the Integral Equation

\[
\begin{pmatrix}
\hat{g}(x_1) \\
\hat{g}(x_2) \\
\vdots \\
\hat{g}(x_m)
\end{pmatrix}_{\text{predictive}} = 
\begin{pmatrix}
\frac{(\lambda_1 T)^{x_1}}{x_1!} e^{-\lambda_1 T} \\
\frac{(\lambda_1 T)^{x_2}}{x_2!} e^{-\lambda_1 T} \\
\vdots \\
\frac{(\lambda_1 T)^{x_m}}{x_m!} e^{-\lambda_1 T}
\end{pmatrix} \cdot 
\begin{pmatrix}
\frac{(\lambda_2 T)^{x_1}}{x_1!} e^{-\lambda_2 T} \\
\frac{(\lambda_2 T)^{x_2}}{x_2!} e^{-\lambda_2 T} \\
\vdots \\
\frac{(\lambda_2 T)^{x_m}}{x_m!} e^{-\lambda_2 T}
\end{pmatrix} \cdot 
\begin{pmatrix}
\frac{(\lambda_3 T)^{x_1}}{x_1!} e^{-\lambda_3 T} \\
\frac{(\lambda_3 T)^{x_2}}{x_2!} e^{-\lambda_3 T} \\
\vdots \\
\frac{(\lambda_3 T)^{x_m}}{x_m!} e^{-\lambda_3 T}
\end{pmatrix} 
\begin{pmatrix}
\tilde{f}_1 \\
\tilde{f}_2 \\
\vdots \\
\tilde{f}_n
\end{pmatrix}_{\text{prior}}
\]

\[
\hat{g} = P \cdot \tilde{f}
\]

\[
\text{argmin} \left\{ \| P \cdot \tilde{f} - \hat{g} \|^2_2 + \lambda^2 \cdot \| L \cdot \tilde{f} \|^2_2 \right\}, 
L = \begin{pmatrix}
1 & -2 & 1 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 1 & -2 & 1
\end{pmatrix} \in \mathbb{R}^{(n-2) \times n}
\]

\[
\text{argmin} \left\| \left( \begin{pmatrix} P \\ \lambda L \end{pmatrix} \right) \cdot \tilde{f} - \left( \begin{pmatrix} \hat{g} \\ 0 \end{pmatrix} \right) \right\|^2_2,
\text{subject to } \tilde{f} \geq 0
\]
Prior Selection Methods

- Method of moments
- Maximum likelihood fitting of Gamma (\(\alpha, \beta\))
- Maximum Likelihood II (Evidence maximization)
- Nonparametric empirical Bayes (NPEB)

\[
\text{Gamma}(30, 3)
\]

\[
\lambda_1 \quad \lambda_2 \quad \lambda_3
\]

\[
x_1 = \text{Poisson}(\lambda_1 t) \quad x_2 = \text{Poisson}(\lambda_2 t) \quad x_3 = \text{Poisson}(\lambda_3 t)
\]

\[
x_1, x_2, x_3, t - \text{specified}
\]
Inference for a New Plant

\[ x_{new} = \text{Poisson}(\lambda_{true}^{new} \cdot t) \quad \leftarrow \lambda_{true}^{new} \leftarrow \text{Gamma}(30,3) \]

\[ \lambda_{est}^{new} = \int_{0}^{\infty} \lambda \cdot p(\lambda) d\lambda \]
Performance Measures

Root Mean Squared Error (RMSE) between true prior and estimated one

\[ RMSE = \frac{\|f_{true} - f_{est}\|_2}{\sqrt{n}} \]

Kullback–Leibler (KL) divergence

\[ KL = \sum_{i=1}^{n} f_{true}^i \cdot \log \left( \frac{f_{true}^i}{f_{est}^i} \right) \]

MSE between \( \lambda_{true}^{new} \) and \( \lambda_{est}^{new} \)

\[ MSE = [\lambda_{true}^{new} - \lambda_{est}^{new}]^2 \]
Comparison of Different Prior Estimation Techniques

RMSE and 95% CI between true prior and estimated one

KL divergence and 95% CI between true prior and estimated one

MSE and 95% CI between true failure rate for the new plant and its estimates
Applying NPEB to MDP FR Data