Bayesian Reliability: Combining Information

Alyson Wilson, Ph.D., PStat®

Department of Statistics
North Carolina State University
agwilso2@ncsu.edu

March 7, 2016
From J. de Mast, T. Akkerhuis, T. Erdmann in their Quality Engineering (2014) paper from the First Stu Hunter Research Conference (The Statistical Evaluation of Categorical Measurements: “Simple Scales, but Treacherous Complexity Underneath”): In line with the philosophy of the Stu Hunter Research Conference, where this article was presented, the article focuses less on novel technical contributions but instead aims to explore where the field is now and make the case for a certain direction that, in the view of the authors, is a fruitful way to make progress.
Outline

1. Background and Context
2. Bayesian Basics
   - Illustrative Examples
   - Computation
   - Prior Distributions
3. System Reliability
4. Moving Forward
   - Planning Data Collection
   - Reliability Growth
   - Mission Reliability
5. References
At the First Stu Hunter Research Conference, Bill Meeker looked ahead to the changes coming to reliability due to “Big Data.”

- Sensors that will collect system operating/environmental (SOE) data
- System health management, condition-based maintenance, early warning of emerging reliability issues, prediction of remaining life for individual systems
When be Bayesian?

From Meeker and Hong (2014), “Many of the applications described in this article and particularly in this concluding section will require combining information from different sources (e.g., data, inexact physics-based knowledge, and certain kinds of expert opinion). Additionally, statistical models being used will often contain multiple sources of variability. Bayesian statistical methods provide a natural approach for combining such information.”
Use the slide to introduce some of the areas for further discussion.

Introduce stockpile surveillance/reliability assessment for the US nuclear stockpile as an example. This is a classic example of multiple sources of data (and variability as stated in Meeker and Hong).

Meeker and Hong discuss extrapolation, a common desire in reliability analyses: “Extrapolation will be more reliable if predictions are based on a combination of science-based models of reliability (e.g., knowledge of the physics of well-understood failure modes) and data are used to develop predictive models for a failure time distribution.” (Come back to this in the discussion of reliability growth models.)

Introduce the DoD acquisition test and evaluation problem, where there are data collected from a series of test events and the system changes/is improved. Introduce test planning as a goal.
Why be Bayesian?

- Allows incorporation of prior information
  - May supplement limited data
  - May provide improvements in cost or precision
  - Provides a formal framework to think about how to combine information
- Computational simplifications
  - Censored data
  - When framed as a Bayesian problem, complex models can often be fit relatively easily using Markov chain Monte Carlo or other computational algorithms.
- Straightforward to produce estimates and credible intervals for complicated functions of model parameters (e.g., predictions, probability of failure, quantiles of lifetime distribution)
- Philosophical
Examples

Basic Statistics Problem: Unknown population parameter ($\theta$) must be estimated.

1. **Example 1:**
   $\theta = $ Probability of a successful launch of a new vehicle by an inexperienced agent. During the period 1980–2002, eleven launches of new vehicles were performed by companies or agencies with little launch vehicle design experience. Of these eleven, three were successful and eight were failures.

2. **Example 2:**
   $\theta = $ Mean and standard deviation of (log) viscosity breakdown times for a particular lubricating fluid. Data are collected (in 1000s of hours) for 50 samples.
Discuss the traditional types of reliability data.

- Bernoulli Success/Failure Data
- Failure Count Data
- Lifetime/Failure Time Data
- Degradation Data
- Covariates to allow modeling of all of the above

Mention the “less traditional” sources of information, like information from previous tests/historical data, related systems, computer models, engineering judgment
## Launch Vehicle Outcome Data

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pegasus</td>
<td>Success</td>
</tr>
<tr>
<td>Percheron</td>
<td>Failure</td>
</tr>
<tr>
<td>AMROC</td>
<td>Failure</td>
</tr>
<tr>
<td>Conestoga</td>
<td>Failure</td>
</tr>
<tr>
<td>Ariane 1</td>
<td>Success</td>
</tr>
<tr>
<td>India SLV-3</td>
<td>Failure</td>
</tr>
<tr>
<td>India ASLV</td>
<td>Failure</td>
</tr>
<tr>
<td>India PSLV</td>
<td>Failure</td>
</tr>
<tr>
<td>Shavit</td>
<td>Success</td>
</tr>
<tr>
<td>Taepodong</td>
<td>Failure</td>
</tr>
<tr>
<td>Brazil VLS</td>
<td>Failure</td>
</tr>
</tbody>
</table>
Viscosity Breakdown Times

Viscosity breakdown times (in 1000s of hours) for 50 samples of a lubricating fluid.

<p>| | | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5.45</td>
<td>16.46</td>
<td>15.70</td>
<td>10.39</td>
<td>6.71</td>
<td>3.77</td>
<td>7.42</td>
<td>6.89</td>
<td>9.45</td>
<td>5.89</td>
<td></td>
</tr>
<tr>
<td>7.39</td>
<td>5.61</td>
<td>16.55</td>
<td>12.63</td>
<td>8.18</td>
<td>10.44</td>
<td>6.03</td>
<td>13.96</td>
<td>5.19</td>
<td>10.96</td>
<td></td>
</tr>
<tr>
<td>14.73</td>
<td>6.21</td>
<td>5.69</td>
<td>8.18</td>
<td>4.49</td>
<td>3.71</td>
<td>5.84</td>
<td>10.97</td>
<td>6.81</td>
<td>10.16</td>
<td></td>
</tr>
<tr>
<td>4.34</td>
<td>9.81</td>
<td>4.30</td>
<td>8.91</td>
<td>10.07</td>
<td>5.85</td>
<td>4.95</td>
<td>7.30</td>
<td>4.81</td>
<td>8.44</td>
<td></td>
</tr>
<tr>
<td>6.56</td>
<td>9.40</td>
<td>11.29</td>
<td>12.04</td>
<td>1.24</td>
<td>3.45</td>
<td>11.28</td>
<td>6.64</td>
<td>5.74</td>
<td>6.79</td>
<td></td>
</tr>
</tbody>
</table>
Examples

Step 1 of both the Bayesian and non-Bayesian formulations is to choose a statistical model (sampling distribution) for the data.

- **Example 1:**
  One choice for $f(y \mid \theta)$ is that the number of successful launches ($Y$) follows a *binomial* distribution with $n$ launches and the probability of any one test being a “success” denoted as $\theta$.

- **Example 2:**
  One choice for $f(y \mid \theta)$ is that natural logarithm of the viscosity breakdown times ($\log(Y)$) has an *normal* distribution with parameters $\theta = (\mu, \sigma^2)$. 
Non-Bayesian Analysis

1. All pertinent information enters the problem through the likelihood function in the form of data \((Y_1, \ldots, Y_n)\). For example,

\[
L(\theta) = \prod_{i=1}^{n} f(y_i | \theta)
\]

2. Many software packages have this capability

3. Maximum likelihood, unbiased estimation, etc.

4. Confidence intervals, tests of hypotheses
Non-Bayesian Analysis

For the launch vehicle data, a point estimate of the success probability of a new launch system developed by an inexperienced manufacturer is provided by the maximum likelihood estimate:

$$\hat{\theta} = \frac{y}{n} = \frac{3}{11} = 0.272$$

An interval estimate for the population proportion of success can be obtained using the asymptotic normal sampling distribution of the MLE $\hat{\theta}$. For $\theta$, we have a 90% confidence interval of

$$(0.272 - 1.645 \times 0.134, 0.272 + 1.645 \times 0.134)) = (0.052, 0.492).$$

In repeated sampling, one expects the confidence interval to include the unknown parameter $\theta$ with probability close to 0.90.
Bayesian Analysis

1. Data enters through the likelihood function
2. However, other information can be incorporated through the *prior distribution*
3. Prior distribution: Before any data collection, the view of/information about the parameter
   - Expressed as a probability distribution on $\theta$
   - Can come from expert opinion, historical studies, previous research, or general knowledge of a situation
   - There exist “noninformative” ("flat,” “diffuse”) priors that represent states of ignorance.
Briefly discuss the idea of noninformative priors. Make the distinction between something like a Jeffreys prior, which is developed to satisfy mathematical concepts of invariance, versus a prior that is relatively diffuse with respect to the likelihood. Mention that a uniform prior for the launch vehicle case is not particularly uninformative, although it is diffuse.
Bayesian Analysis

4 Bayes’ Theorem:

\[ p(\theta \mid y_1, \ldots, y_n) = \frac{\prod_{i=1}^{n} f(y_i \mid \theta) \times \pi(\theta)}{\int \prod_{i=1}^{n} f(y_i \mid \theta) \times \pi(\theta) d\theta}. \]

The \textit{posterior distribution}, \( p(\theta \mid y_1, \ldots, y_n) \), is a constant multiplied by the likelihood, \( \prod_{i=1}^{n} f(y_i \mid \theta) \), multiplied by the \textit{prior distribution}, \( \pi(\theta) \). (The posterior distribution is proportional to the prior times the likelihood.)

5 Posterior distribution: In light of the data, the updated view of/information about the parameter.

6 All inference is based on the posterior distribution.
Censored Data

Censored data requires no new methodology in a Bayesian analysis. It is incorporated into the likelihood.

<table>
<thead>
<tr>
<th>Type of Observation</th>
<th>Failure Time</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Uncensored</td>
<td>$T = t$</td>
<td>$f(t)$</td>
</tr>
<tr>
<td>Left censored</td>
<td>$T \leq t_L$</td>
<td>$F(t_L)$</td>
</tr>
<tr>
<td>Interval censored</td>
<td>$t_L &lt; T \leq t_R$</td>
<td>$F(t_R) - F(t_L)$</td>
</tr>
<tr>
<td>Right censored</td>
<td>$T &gt; t_R$</td>
<td>$1 - F(t_R)$</td>
</tr>
</tbody>
</table>
Bayesian Analysis

1. “Subjective” (what/whose information is contained in the prior distribution?)
2. Fewer software packages have this capability (SAS PROC MCMC, OpenBUGS, JAGS, STAN, NIMBLE, R packages)
3. Result is a probability distribution
4. *Credible intervals* use the language that everyone wants to use. (Probability that $\theta$ is in the interval is 0.90.)
Mention the “subjectivity” inherent in the specification of the likelihood and discuss that well-done Bayesian analyses are up front about how the data is used and the sensitivity of the posterior distribution to any assumptions.
Example 1
Bayesian Analysis

A convenient choice to represent prior information about the probability of a successful launch is the *Beta* distribution. One interpretation of the parameters defining this distribution are the number of *a priori* successes and failures.

For example, if an expert hypothesizes that her opinion about the probability of successful launch is worth 8 vehicle launches and further expects successful launches 6 times, we would reflect this with a *Beta*(6, 2) distribution.
Example 1

- Non-Bayesian analysis: If our data are $Binomial(n, \theta)$ then we would calculate $Y/n$ as our estimate and use a confidence interval formula for a proportion.

- Bayesian analysis: If our data are $Binomial(n, \theta)$ and our prior distribution is $Beta(a, b)$, then our posterior distribution is $Beta(a + y, b + n - y)$.
Example 1
Bayesian Analysis

• In this case, the posterior distribution is $Beta(6 + 3, 2 + 11 − 3) = Beta(9, 10)$.

• This means that we can say that the probability that $\theta$ is in the interval $(0.291, 0.659)$ is 0.90.

• Notice that we don’t have to address the problem of “in repeated sampling”; this is a direct probability statement that relies on the prior distribution.
Example 1
Beta(6,2) Prior

\[ n = 11, y = 3, a = 6, b = 2 \]
Example 1

Diffuse (Uniform(0,1)) Prior Distribution

\[ n = 11, \ y = 3, \ a = 1, \ b = 1 \]
Example 1

Large $n$, Beta(6,2) Prior Distribution

$n = 110, y = 30, a = 6, b = 2$
Example 1

Large $n$, Diffuse Prior Distribution

$n = 110, y = 30, a = 1, b = 1$
Example 1
All Four Posterior Distributions

![Graph showing posterior distributions with different parameters and sample sizes.](image-url)
90% confidence interval: (0.052, 0.492) 90% credible intervals:

- Beta(6,2), n = 11: (0.291,0.659)
- Uniform(0,1), n = 11: (0.123,0.527)
- Beta(6,2), n = 110: (0.238,0.376)
- Uniform(0,1), n = 110: (0.210,0.348)

Spend some time with this slide interpreting the differences among the confidence interval and credible intervals. Discuss the traditional subjective Bayesian interpretation of the Beta(6,2) interval and give examples where expert judgment may appropriately influence inference and how that can be presented to a decision-maker.
Non-Conjugate Prior

Suppose that instead of capturing our prior information using a $Beta(6, 2)$ distribution, we capture it using a $NegativeLogGamma(2, 0.5)$ distribution, with

$$
\pi(\theta) = 0.25[-\log(\theta)]\theta^{-0.5}
$$
What is plotted is the Beta(6,2) in solid and the NLG(2,0.25) in dotted
Negative Log Gamma Prior

For our example, the posterior distribution is

\[
p(\theta \mid y_1, \ldots, y_n) = \frac{\prod_{i=1}^{n} f(y_i \mid \theta) \times \pi(\theta)}{c(y)}
\]

\[
\propto \prod_{i=1}^{n} f(y_i \mid \theta) \times \pi(\theta)
\]

\[
\propto \theta^{2.5} (1 - \theta)^{8} [- \log(\theta)]
\]
Computation

- What do we do with the joint posterior distribution of $\theta$?
- Note that the dimension of $\theta$ is often large in real problems.
- Only for very simple cases is $c(y)$ known.
- Direct numerical integration (for example, to determine means or marginal distributions) is problematic.
- We could construct a normal approximation ...
- The answer is: We sample. In particular, we figure out how to draw a random sample from $p(\theta \mid y_1, \ldots, y_n)$. Then we can compute quantities of interest using Monte Carlo techniques.
There are a variety of algorithms that can be used to get our random samples:

- Rejection sampling
- Importance sampling
- Sampling importance resampling
- Gibbs sampling
- Metropolis-Hastings algorithm
Example 2

We now turn to an example involving continuous-valued random variables. The particular data set we consider represents viscosity breakdown times for 50 samples of a lubricant.
Example 2

- Recall: $\theta = (\mu, \sigma^2)$ are the mean and variance of the natural logarithm of the lubricant measurements.

- A convenient choice for $\pi(\theta)$ is a normal distribution for $\mu$ and an inverse gamma distribution for $\sigma^2$. We'll assume mutual independence.

- If $X_1, \ldots, X_n$ are normal with mean $\mu$ and known variance $\sigma^2$ \([x_1 \ldots x_n \sim N(\mu, \sigma^2)]\) AND the prior distribution of $\mu$ is also normal with mean $m$ and variance $v^2$ \([\mu \sim N(m, v^2)]\) then the posterior distribution of $\theta$ is also normal with mean $\frac{m\sigma^2 + nv^2 \bar{x}}{\sigma^2 + nv^2}$ and variance $\frac{v^2\sigma^2}{\sigma^2 + nv^2}$.
Example 2

- System engineers who have studied the viscous properties of these lubricants believe that $\log(Y)$ should be centered at 2, but that value is known only with standard deviation 1.
- They have very little knowledge about the variability they expect to see.
- The process of getting a prior distribution from statements made by experts is called *elicitation*.
- Role of prior predictive distribution in assessing multivariate prior distributions:

$$p(y) = \int f(y | \theta) \pi(\theta) d\theta,$$

This distribution reflects what we would expect for a randomly selected fluid breakdown time in the presence of all *a priori* uncertainty.
Prior Predictive Distribution for Viscosity Breakdown Times

\[ \mu \sim \text{Normal}(2, 1), \sigma^2 \sim \text{InverseGamma}(1, 5) \]
Launch Vehicle Revisited
Data from Johnson et al. (2005)

<table>
<thead>
<tr>
<th>Vehicle</th>
<th>Outcome</th>
<th>Quality Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pegasus</td>
<td>9/10</td>
<td>64</td>
</tr>
<tr>
<td>Percheron</td>
<td>0/1</td>
<td>34</td>
</tr>
<tr>
<td>AMROC</td>
<td>0/1</td>
<td>51</td>
</tr>
<tr>
<td>Conestoga</td>
<td>0/1</td>
<td>55</td>
</tr>
<tr>
<td>Ariane 1</td>
<td>9/11</td>
<td>76</td>
</tr>
<tr>
<td>India SLV-3</td>
<td>3/4</td>
<td>62</td>
</tr>
<tr>
<td>India ASLV</td>
<td>2/4</td>
<td>72</td>
</tr>
<tr>
<td>India PSLV</td>
<td>6/7</td>
<td>75</td>
</tr>
<tr>
<td>Shavit</td>
<td>2/4</td>
<td>68</td>
</tr>
<tr>
<td>Taepodong</td>
<td>0/1</td>
<td>56</td>
</tr>
<tr>
<td>Brazil VLS</td>
<td>0/2</td>
<td>47</td>
</tr>
</tbody>
</table>
Hierarchical Models

Launch Vehicles Revisited

To more accurately model these data, it makes sense to introduce parameters $\pi_i$ that denote the long-term probability that the launch of the $i$th vehicle is successful.

The model proposed in Johnson et al. (2005) for the probability of a successful launch of vehicle $i$ on the $j^{th}$ launch:

$$X_{ij} \sim Bernoulli(\pi_{ij})$$

$$-\log(-\log(\pi_{ij})) = -\log(-\log(\pi_i)) + \alpha_0 W_{j,1} + \alpha_1 W_{j,2}$$

$$\pi_i \sim Beta(K\gamma_i, K(1-\gamma_i))$$

$$-\log(-\log(\gamma_i)) = x_i^T \beta$$

- $\gamma_i$ is the mean of the Beta distribution, and $K$ controls its precision.
- $\alpha_0, \alpha_1, \beta_0, \beta_1$ have (improper) uniform priors
- $K \sim Exponential(1)$
• Discuss hierarchical specification. Give a few examples where a hierarchical prior makes sense.

• There is strength borrowed from the data of other manufacturers, but there is an offset based on experience \((j \leq 2 \rightarrow W_{j,1} = 1, j > 2 \rightarrow W_{j,2} = 1)\).

• There is also a covariate that adjusts the probability of launch success based on assessed quality indices.
Reflections

- Science is subjective (what about the choice of a likelihood?)
- Bayesian analyses use all available information
- Bayesian analyses often make interpretation easier
- **Bad News**: So far, we have looked at straightforward cases
- **Good News**: Bayesian analyses are possible (and practical) with advanced computational procedures.
Mention that there are prior specifications that weight the expert opinion by how well it is calibrated with observed data, e.g., Reese et al. (2004) and subsequent; Ibrahim and Chen power priors (e.g., Statistical Science (2000) 15(1): 46-60
DoD systems experience system design, contractor testing, developmental testing, and operational testing.

Later in the lifecycle, we see new variants of systems, life extension programs, . . . .

Within the Department of Energy, the goal of science-based stockpile stewardship is the assessment of safety and reliability in aging warheads in the absence of nuclear testing.
Considering systems is one area that Bayesian methods have particular utility. The basic ideas that have just been introduced for modeling individual components can then be put together into more complex structures.
Statistical Areas of Interest

- **Data**
  - Multilevel
  - Multiple types: binary, lifetime, degradation, expert judgement, computer model

- **Systems**
  - Representation
  - Assessment: model checking and diagnostics, model fit

- **Planning data collection**
Statistical Areas of Interest

- **Data**
  - Multilevel
  - Multiple types: binary, lifetime, degradation, expert judgement, computer model

- **Systems**
  - Representation
  - Assessment: model checking and diagnostics, model fit

- Planning data collection
Statistical Areas of Interest

- **Data**
  - Multilevel
  - Multiple types: binary, lifetime, degradation, expert judgement, computer model

- **Systems**
  - Representation
  - Assessment: model checking and diagnostics, model fit

- Planning data collection
Multilevel Data
Multilevel Pass/Fail Data, Series System

- Information collected at $C_0$, $C_1$, $C_2$, and $C_3$
- Information at $C_0$ provides partial information about $C_1$, $C_2$, and $C_3$
- Goal: simultaneous inference about system and component reliabilities

<table>
<thead>
<tr>
<th>Component</th>
<th>Successes</th>
<th>Failures</th>
<th>Trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Component 1</td>
<td>8</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Component 2</td>
<td>7</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Component 3</td>
<td>3</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>System</td>
<td>10</td>
<td>2</td>
<td>12</td>
</tr>
</tbody>
</table>
System Representations

Reliability Block Diagram

Fault Tree

Bayesian Network
Mention these as three ways that are commonly used to represent systems. Reliability block diagrams and fault trees are the same, but depend on whether you are looking at system success or failure. A Bayesian network is a generalization that allows for a probability of system failure depending on the states of the components.
System Representations
Developing the model for a complex system
Mention the thread of work in the elicitation of complex systems (e.g., Wilson et al. (2007) and the subsequent work to develop tools, e.g. Graves et al. (2008).

Also leads to model checking, e.g., Anderson-Cook (2008), Guo and Wilson (2013), Zhang and Wilson (2016).

Inference does depend on getting the system structure correct. The reliability problem is different from the machine learning problem where we are trying to use the data to learn the conditional independence structure (also represented using a Bayesian network).
Example
Multilevel Pass/Fail Data, Bayesian Network, Reliability Changing with Time

<table>
<thead>
<tr>
<th>Age</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>19/19</td>
<td>35/35</td>
<td>15/16</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>47/48</td>
<td>14/14</td>
</tr>
<tr>
<td>3</td>
<td>16/19</td>
<td>37/38</td>
<td>12/14</td>
</tr>
<tr>
<td>4</td>
<td>12/12</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>-</td>
<td>44/45</td>
<td>13/14</td>
</tr>
<tr>
<td>6</td>
<td>-</td>
<td>35/37</td>
<td>11/12</td>
</tr>
<tr>
<td>7</td>
<td>9/13</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>-</td>
<td>33/42</td>
<td>5/16</td>
</tr>
<tr>
<td>9</td>
<td>-</td>
<td>-</td>
<td>12/19</td>
</tr>
<tr>
<td>10</td>
<td>3/10</td>
<td>30/39</td>
<td>8/14</td>
</tr>
</tbody>
</table>
The intent is not to cover the example in detail, but to give the flavor of the data, likelihood, and inference.
Model

- “Component” probabilities, logistic regression,

\[
p_A(t) = \frac{\exp(\alpha_A + \beta_A t)}{1 + \exp(\alpha_A + \beta_A t)}
\]

\[
p_B(t) = \frac{\exp(\alpha_B + \beta_B t)}{1 + \exp(\alpha_B + \beta_B t)}
\]

- Conditional probabilities

\[
\tau_{11} = P(C = 1 | A = 1, B = 1)
\]

\[
\tau_{10} = P(C = 1 | A = 1, B = 0)
\]

\[
\tau_{01} = P(C = 1 | A = 0, B = 1)
\]

\[
\tau_{00} = P(C = 1 | A = 0, B = 0)
\]
Model

- “Component” probabilities, logistic regression,

\[ p_A(t) = \frac{\exp(\alpha_A + \beta_A t)}{1 + \exp(\alpha_A + \beta_A t)} \]
\[ p_B(t) = \frac{\exp(\alpha_B + \beta_B t)}{1 + \exp(\alpha_B + \beta_B t)} \]

- Conditional probabilities

\[ \tau_{11} = \mathbb{P}(C = 1 \mid A = 1, B = 1) \]
\[ \tau_{10} = \mathbb{P}(C = 1 \mid A = 1, B = 0) \]
\[ \tau_{01} = \mathbb{P}(C = 1 \mid A = 0, B = 1) \]
\[ \tau_{00} = \mathbb{P}(C = 1 \mid A = 0, B = 0) \]
Model

- “System” probability

\[
p_C(t) = \tau_{11} \frac{\exp(\alpha_A + \alpha_B + (\beta_A + \beta_B)t)}{(1 + \exp(\alpha_A + \beta_A t))(1 + \exp(\alpha_B + \beta_B t))} \\
+ \tau_{10} \frac{\exp(\alpha_A + \beta_A t)}{(1 + \exp(\alpha_A + \beta_A t))(1 + \exp(\alpha_B + \beta_B t))} \\
+ \tau_{01} \frac{\exp(\alpha_B + \beta_B t)}{(1 + \exp(\alpha_A + \beta_A t))(1 + \exp(\alpha_B + \beta_B t))} \\
+ \tau_{00} \frac{1}{(1 + \exp(\alpha_A + \beta_A t))(1 + \exp(\alpha_B + \beta_B t))}
\]
Likelihood and Prior

Data at $t = 3$

<table>
<thead>
<tr>
<th>Age</th>
<th>A</th>
<th>B</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>16/19</td>
<td>37/38</td>
<td>12/14</td>
</tr>
</tbody>
</table>

Likelihood at $t = 3$

$$L(\alpha_A, \alpha_B, \beta_A, \beta_B, \tau_{11}, \tau_{10}, \tau_{01}, \tau_{00}) = p_A(3)^{16}(1 - p_A(3))^3 p_B(3)^{37}(1 - p_B(3)) p_C(3)^{12}(1 - p_C(3))^2$$
Posterior Distributions
Extensions

This basic approach has been extended in many directions.

- Lifetime and degradation data at the components (Wilson et al. (2006))
- Multiple diagnostics measured at the components (Anderson-Cook et al. (2008))
- Lifetime data at the components and system (Reese et al. (2011))
- Binary, lifetime, or degradation data at components and system (Guo and Wilson (2013))
- Parallel line of research focused on developing system models: elicitation, software, representations (e.g., Wilson et al. (2007), Anderson-Cook (2008))
- Prior distributions (to capture knowledge about parameters “before” this experiment)
  - “Naive” specifications can lead to surprisingly bad results
  - Variable selection priors
• There has been substantial progress in the last 20 years with models, methods, and tools for Bayesian reliability.
• How do we design tests and optimize data collection when we acknowledge that we have many kinds of information and considerable relevant historical information?
• How do we embed reliability growth models in a Bayesian framework?
• How do we think about the idea of mission reliability?
Note the link between physics-based and reliability growth models. The reliability growth model is more probabilistic than mechanistic. Gets back to the question raised in the introduction about extrapolation and prediction.
Resource Allocation

One approach that we will not cover in detail is often called resource allocation. For given constraints, typically of cost or time, what set of tests do we perform that optimizes some metric?

- Often the metric is the length of the credible interval for some quantity of interest.
- The optimization is done over not just one possible source of data, but over a range of possibilities, from more field tests to additional computer modeling.
- Often there is often an explicit tradeoff between data that can be collected about a system and data that can be collected about its components.
Reliability Demonstration and Assurance

- Example: Using minimal assumptions, to demonstrate that reliability at time $t_0$ hours is 0.99, with 90% confidence, requires testing at least 230 units for $t_0$ hours with zero failures. To have a 80% chance of passing the test, requires that the true reliability be approximately 0.999.

- For complicated, expensive systems, traditional reliability demonstration is usually not practical.

- Reliability assurance is an alternative: Use whatever relevant knowledge you have in a principled Bayesian approach to plan the test.
Bayesian Binomial Test Plan

- Suppose we want to develop a Bayesian binomial test plan.
- We want to determine \((n, c)\) where \(n\) is the test sample size and \(c\) is the number of systems allowed to fail before the “test is failed.”
- What criteria do we use to choose \(n\) and \(c\)?
Test Criteria

There are two errors we could make:

- We could decide the “test is failed” when the system reliability $\pi$ is higher than a specified $\pi_P$

Posterior Producer’s Risk: Choose a test plan so that if the test is failed, there is a small probability that the reliability at $t_I$ (the time of interest) is high.
Test Criteria

There are two errors we could make:

- We could decide the “test is failed” when the system reliability $\pi$ is higher than a specified $\pi_P$

Posterior Producer’s Risk: Choose a test plan so that if the test is failed, there is a small probability that the reliability at $t_I$ (the time of interest) is high
Test Criteria

- We could decide the “test is passed” when the system reliability is lower than a specified $\pi_C$.

Posterior Consumer’s Risk: Choose a test plan so that if the test is passed, there is a small probability that the reliability at $t_I$ is low.

- Reliable Life Criterion: Choose a test plan so that if the test is passed, there is a high probability that the $1 - \alpha$ quantile of the distribution is greater than $t_I$. 
Test Criteria

- We could decide the “test is passed” when the system reliability is lower than a specified $\pi_C$.

Posterior Consumer’s Risk: Choose a test plan so that if the test is passed, there is a small probability that the reliability at $t_I$ is low.

- Reliable Life Criterion: Choose a test plan so that if the test is passed, there is a high probability that the $1 - \alpha$ quantile of the distribution is greater than $t_I$. 
Posterior Producer’s Risk

Posterior Producer’s Risk

\[ = \mathbf{P}(\pi \geq \pi_P \mid \text{Test Is Failed}, x) \]

\[ = \int_{\pi_P}^{1} p(\pi \mid y > c, x) d\pi \]

\[ = \int_{\pi_P}^{1} \frac{f(y > c \mid \pi)p(\pi \mid x)}{\int_{\pi_P}^{1} f(y > c \mid \pi)p(\pi \mid x) d\pi} d\pi \]

\[ = \int_{\pi_P}^{1} \left[ \sum_{y=c+1}^{n} \binom{n}{y}(1-\pi)^y \pi^{n-y} \right] p(\pi \mid x) d\pi \]

\[ = \int_{0}^{1} \left[ \sum_{y=c+1}^{n} \binom{n}{y}(1-\pi)^y \pi^{n-y} \right] p(\pi \mid x) d\pi \]

\[ = \int_{\pi_P}^{1} \left[ 1 - \sum_{y=0}^{c} \binom{n}{y}(1-\pi)^y \pi^{n-y} \right] p(\pi \mid x) d\pi \]

\[ = 1 - \int_{0}^{1} \left[ \sum_{y=0}^{c} \binom{n}{y}(1-\pi)^y \pi^{n-y} \right] p(\pi \mid x) d\pi \]
Posterior Consumer’s Risk

\[
\begin{align*}
\text{Posterior Consumer’s Risk} & = P(\pi \leq \pi_C \mid \text{Test Is Passed, } x) \\
& = \int_0^{\pi_C} p(\pi \mid y \leq c, x) d\pi \\
& = \int_0^{\pi_C} \frac{f(y \leq c \mid \pi)p(\pi \mid x)}{\int_0^1 f(y \leq c \mid \pi)p(\pi \mid x) d\pi} d\pi \\
& = \int_0^{\pi_C} \left[ \sum_{y=0}^{c} \binom{n}{y}(1-\pi)^y \pi^{n-y} \right] \frac{p(\pi \mid x)}{\int_0^1 \left[ \sum_{y=0}^{c} \binom{n}{y}(1-\pi)^y \pi^{n-y} \right] p(\pi \mid x) d\pi} d\pi.
\end{align*}
\]

We evaluate these integrals using Monte Carlo integration and draws from the posterior distribution of \( p(\pi \mid x) \).
Assurance Testing Example

- Missile system consisting of 10 components.
- The 10 components are labeled A – K, and each component has two or three versions as denoted by the numbers following the identifying letters. For example, component A1 is component A, version 1.
- The missile is a series system.
- Over time, seven variants of the missile have been tested.
- For some systems we do not have information about which variant of the component was tested.
Example drawn from Hamada et al. (2014)
## Components Used in Variants of System

<table>
<thead>
<tr>
<th>System Variants</th>
<th>I</th>
<th>II</th>
<th>III</th>
<th>IV</th>
<th>V</th>
<th>VI</th>
<th>VII</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A1</td>
<td>A2</td>
<td>A1</td>
<td>A2</td>
</tr>
<tr>
<td>B1</td>
<td>B1</td>
<td>B1</td>
<td>B1</td>
<td>B1</td>
<td>B1</td>
<td>B2</td>
<td>B2</td>
</tr>
<tr>
<td>C1</td>
<td>C1</td>
<td>C2</td>
<td>C3</td>
<td>C3</td>
<td>–</td>
<td>C3</td>
<td>–</td>
</tr>
<tr>
<td>D1</td>
<td>D2</td>
<td>D2</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>E1</td>
<td>E2</td>
<td>E2</td>
<td>E1</td>
<td>E2</td>
<td>E2</td>
<td>E2</td>
<td>E2</td>
</tr>
<tr>
<td>F1</td>
<td>F2</td>
<td>F3</td>
<td>F3</td>
<td>F3</td>
<td>F3</td>
<td>F3</td>
<td>F2</td>
</tr>
<tr>
<td>G1</td>
<td>G2</td>
<td>G3</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>H1</td>
<td>H2</td>
<td>H2</td>
<td>H2</td>
<td>H2</td>
<td>H2</td>
<td>H2</td>
<td>H2</td>
</tr>
<tr>
<td>J1</td>
<td>J2</td>
<td>J3</td>
<td>J3</td>
<td>J3</td>
<td>J3</td>
<td>J3</td>
<td>J3</td>
</tr>
<tr>
<td>K1</td>
<td>K1</td>
<td>K1</td>
<td>K1</td>
<td>K1</td>
<td>K1</td>
<td>K1</td>
<td>K2</td>
</tr>
</tbody>
</table>
# Previous Tests

<table>
<thead>
<tr>
<th>Component</th>
<th>Tests</th>
<th>Successes</th>
<th>Component</th>
<th>Tests</th>
<th>Successes</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>638</td>
<td>633</td>
<td>F2</td>
<td>338</td>
<td>336</td>
</tr>
<tr>
<td>A2</td>
<td>70</td>
<td>65</td>
<td>F3</td>
<td>196</td>
<td>195</td>
</tr>
<tr>
<td>B1</td>
<td>662</td>
<td>651</td>
<td>G1</td>
<td>174</td>
<td>174</td>
</tr>
<tr>
<td>B2</td>
<td>46</td>
<td>43</td>
<td>G2</td>
<td>296</td>
<td>294</td>
</tr>
<tr>
<td>C1</td>
<td>470</td>
<td>468</td>
<td>G3</td>
<td>144</td>
<td>141</td>
</tr>
<tr>
<td>C2</td>
<td>144</td>
<td>141</td>
<td>H1</td>
<td>174</td>
<td>172</td>
</tr>
<tr>
<td>C3</td>
<td>90</td>
<td>89</td>
<td>H2</td>
<td>534</td>
<td>529</td>
</tr>
<tr>
<td>D1</td>
<td>174</td>
<td>174</td>
<td>J1</td>
<td>174</td>
<td>172</td>
</tr>
<tr>
<td>D2</td>
<td>440</td>
<td>435</td>
<td>J2</td>
<td>296</td>
<td>296</td>
</tr>
<tr>
<td>E1</td>
<td>194</td>
<td>192</td>
<td>J3</td>
<td>238</td>
<td>237</td>
</tr>
<tr>
<td>E2</td>
<td>514</td>
<td>512</td>
<td>K1</td>
<td>664</td>
<td>650</td>
</tr>
<tr>
<td>F1</td>
<td>174</td>
<td>174</td>
<td>K2</td>
<td>42</td>
<td>42</td>
</tr>
</tbody>
</table>
We assume that the versions of a component are similar so that we can model the reliabilities hierarchically. In particular, we model the successes of the $j$th version of the $i$th component as:

$$X_{ij} | \pi_{ij} \sim \text{Binomial}(n_{ij}, \pi_{ij}),$$
$$\pi_{ij} | \delta_i, \gamma_i \sim \text{Beta}(\delta_i, \gamma_i),$$
$$\delta_i \sim \text{Uniform}(0, 5000),$$
$$\gamma_i \sim \text{Uniform}(0, 5000)$$
Designing a Test

Consider the situation where a new version of component A is under consideration for the system, while the other components remain at the most recent version. Draws for $\pi$ for this new system that we can use in our Monte Carlo evaluation of consumer and posterior risk are

$$\pi^{(k)} = \pi^{(k)}_A \pi^{(k)}_{B2} \pi^{(k)}_{C3} \pi^{(k)}_{D2} \pi^{(k)}_{E2} \pi^{(k)}_{F3} \pi^{(k)}_{G3} \pi^{(k)}_{H2} \pi^{(k)}_{J3} \pi^{(k)}_{K2}$$

where we use the predictive distribution for $\pi_1$ (component A) and posterior distributions for $\pi_{ij}$ (specific versions of the rest of the components).

We set values for $\pi_P, \pi_C$, posterior producer’s risk, and posterior consumer’s risk and then solve for $n$ and $c$ that satisfy the conditions.
Additional Problems to Investigate

- Reliability Growth
- Mission Reliability
Reliability Growth

- When predicting/extrapolating reliability, combining information from the data and from a physics-based model can improve the quality of the prediction.
- Can we achieve similar gains using a more probabilistic description like those in reliability growth models?
Reliability Growth

Models change and improvement in the reliability of a system as it goes through testing and corrective action periods.
Reliability Growth

- Popular reliability growth models used in the DoD: Duane Model, Crow-AMSAA, AMSAA PM2
- Parameters include initial system MTBF, average fix effectiveness factor, and proportion of uncovered defects repaired
- As currently used in the DoD, reliability growth models are not based on nor are they updated using data!
Reliability

The ability of an item to perform a required function, under given environmental and operating conditions and for a stated period of time (ISO 8402, International Standard: Quality Vocabulary, 1986)

Reliability assessments have traditionally focused on understanding the reliability of components or systems. However,

- Most complex defense systems serve more than one required function (e.g., ships may provide transportation, defense, self-protection, etc.)
- Most are deployed to multiple operating environments: desert, littoral (close to shore), mountain, etc.
- Operating conditions vary depending on mission
- The total mission time is often stated.

How do we formulate a reasonable definition of *mission reliability* and a strategy for assessment and test design?
Favorite Books

Bayesian Reliability


Introductory Bayesian Methods

Favorite Books

Bayesian Computing

Bayesian Reliability Articles I


Bayesian Reliability Articles II


Bayesian Reliability Articles III


Bayesian Reliability Articles IV


Bayesian Reliability Articles V
