Bayesian Data Analysis in R/STAN

Dr. Kassie Fronczyk
Institute for Defense Analyses

Dr. James Brownlow
US Air Force
Outline

• Fundamentals of Bayesian Analysis
• Case study: Littoral Combat Ship (LCS)
• Case Study: Bio-chemical Detection System (BDS)
• STAN Examples
Acceptance of Bayesian Techniques

Bayesian methods are commonly used and becoming more widely accepted

• Applications
  – FAA/USAF in estimating probability of success of launch vehicles
  – Delphi Automotive for new fuel injection systems
  – Science-based Stockpile Stewardship program at LANL for nuclear warheads
  – Army for estimating reliability of new aircraft systems
  – FDA for approving new medical devices

• Recent High Profile Successes:
  – During the search for Air France 447 (2009-2011), black box location
  – The Coast Guard in 2013 found the missing fisherman, John Aldridge
Classical versus Bayesian Statistics

• Experiment 1:
  – A fine musician, specializing in classical works, tells us that he is able to distinguish if Haydn or Mozart composed some classical song. Small excerpts of the compositions of both authors are selected at random and the experiment consists of playing them for identification by the musician. The musician makes 10 correct guesses in 10 trials.

• Experiment 2:
  – The guy next to you at the bar says he can correctly guess in a coin toss what face of the coin will fall down. Again, after 10 trials the man correctly guesses the outcomes of the 10 throws.
Classical versus Bayesian Statistics

- **Classical Statistics Analysis**
  - You have the same confidence in the musician’s ability to identify composers as in the bar guy’s ability to predict coin tosses. In both cases, there were 10 successes in 10 trials.

- **Bayesian Statistics Analysis**
  - Presumably, you are inclined to have more confidence in the musician’s claim than the guy at the bar’s claim. Post analysis, the credibility of both claims will have increased, though the musician will continue to have more credibility than bar guy.
Bayesian Statistics 101

Simple Example

• We have a system comprised of 2 components: Component 1 and Component 2.

• For each of the two components, 10 pass/fail tests are administered and results are recorded. Component 1 fails twice and Component 2 fails zero times.
  – We can calculate the reliability of each component, $R_1$ and $R_2$.
  – We also want an assessment of the system reliability, assuming the components work in series.

$$R_{system} = R_1 \times R_2 = \left(1 - \frac{2}{10}\right) \times \left(1 - \frac{0}{10}\right) = 0.8 \times 1 = 0.8$$

** For the purposes of the next few slides, focus on Component 1.**
Bayesian Statistics Basics

1. Construct prior from prior information, \( f(\theta) \)

2. Construct likelihood from test data,

\[
L(x \mid \theta) = \prod_{i=1}^{n} f(x_i \mid \theta)
\]

3. Estimate posterior distribution using Bayes Theorem

\[
f(\theta \mid x) = \frac{L(x \mid \theta)f(\theta)}{\int L(x \mid \theta)f(\theta)} \propto L(x \mid \theta)f(\theta)
\]

Each of these steps requires careful consideration of both the system and the statistics!
• The prior distribution of the reliability, $f_{\text{prior}}(R)$, is constructed from previous data or expert knowledge. This is your first assessment of the system.

• Say Component 1 was previously tested and failed 3 out of 40 tests: use Beta distribution.

$$f_{\text{prior}}(R_1) \propto R_1^{n_p} (1 - R_1)^{n_p(1-p)}$$

• $p$ is the reliability estimate and $n_p \geq 0$ weights the relevance of the prior test data.

Careful thought should always be put into the prior distribution!
Likelihood

- Tests are performed and the resulting test data is used in the likelihood function, $L(x|R)$.

- The binary test data of Component 1 follows a Binomial distribution with probability of a pass of $R_1$

\[
L(x|R_1) \propto R_1^{s_1}(1 - R_1)^{f_1}
\]

$s_1$ is the number of successes and $f_1$ is the number of failures from Component 1.
Posterior Distribution

- Bayes’ theorem is used to find the posterior reliability distribution, \( f_{\text{posterior}}(R|\text{data}) \).

- The posterior distribution is proportional to the product of the prior distribution and the likelihood function.

- For our example, choosing the **Beta distribution** as a prior is ideal for a few reasons: it ensures that \( R \) is between (0, 1) and it is the “conjugate” prior for the Binomial distribution.

\[
\begin{align*}
    f_{\text{posterior}}(R_1) &\propto R_1^{s_1}(1 - R_1)^{f_1} R_1^{n p} (1 - R_1)^{n p (1 - p)} \\
    &\propto R_1^{s_1 + n p} (1 - R_1)^{f_1 + n p (1 - p)}
\end{align*}
\]

\[\rightarrow\] *Also a Beta distribution with updated parameters: \( s_1 + n p p + 1 \) and \( f_1 + n p (1 - p) + 1 \).*
Visualizing Bayesian Statistics

Classical Estimate: 0.8 (0.55, 0.95)
When Should We Think About Using Bayesian Techniques

- To obtain interval estimates (credible intervals) when there are zero failures
  - Mean time between failure for short tests or for highly reliable systems
  - Interval estimates in kill-chain analysis where zero failures occur at any point along the kill-chain

- If you are assessing a complex system mission reliability
  - LCS Example - Confidence intervals are not straightforward to obtain using frequentist methods, impossible with zero failures in any sub-system

- If there is relevant prior information to be incorporated in your analysis – this may include previous developmental (or operational) test data, engineering analyses, or information from modeling and simulation.
  - BDS Example
Case Study: LCS

- The Capability Development Document for LCS provides a reliability threshold for Core Mission functional area.

<table>
<thead>
<tr>
<th>Critical Subsystem</th>
<th>Total System Operating Time</th>
<th>Operational Mission Failures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Ship Computing Environment (full-time)</td>
<td>4500 hours</td>
<td>1</td>
</tr>
<tr>
<td>Sea Sensors and Controls (underway)</td>
<td>2000 hours</td>
<td>3</td>
</tr>
<tr>
<td>Communications (full-time)</td>
<td>4500 hours</td>
<td>0</td>
</tr>
<tr>
<td>Sea Engagement Weapons (on-demand)</td>
<td>11 missions</td>
<td>2</td>
</tr>
</tbody>
</table>

- The target reliability for Core Mission is 0.80 in 720 hours.

- Assume the functional area is a series system: system is up if all subsystems are up.

Data are notional, based on preliminary results.
Prior Assumptions: LCS

- **On-demand system**
  - Assume no belief in the relevance of prior knowledge, $n_p = 0$

- **Continuous systems**
  - The Gamma prior parameter $a$ is set to 1, giving large variance. To ensure the 50th percentile is set at $\lambda_{50} = 1/\text{MTBF}_{\text{guess}}$, choose $b = \log(2) \times \text{MTBF}_{\text{guess}}$
  - $\text{MTBF}_{\text{guess}}$ chosen by solving the reliability function at the requirement

**Guiding Principles in Prior Selection:**

- **Start with the properties of the parameter of interest**
- **Decide on what prior information to use**
- **Allow for the analysis to change freely based on the data observed**
- **Priors specified at the system level, as opposed to mission level – check impact on system prior**
Prior Specification: On Demand System
Visualizing the Prior Specification: LCS
Visualizing the Prior Specification: LCS
Note the core mission prior is somewhat informative – We will want to check the impact of this in the analysis.
```r
> for(i in 1:B){
+   lambdaTC[i]=rgamma(1, TSCE+aT, TSCET+bT)
+   lambdaSSC[i]=rgamma(1, SensCont+aS, SensContT+bS)
+   lambdaCOMM[i]=rgamma(1, Comm+aC, CommT+bC)
+   pSEW[i]=rbeta(1, SEWs + pguess*pweight, SEWf+(1-pguess)*pweight)
+   Rsys[i]=exp(-720*lambdaTC[i])*exp(-720*lambdaSSC[i])*exp(-720*lambdaCOMM[i])*pSEW[i]
+ }
```
### Comparison of Results

<table>
<thead>
<tr>
<th></th>
<th>Classical MTBOMF</th>
<th>Classical Reliability at 720hrs</th>
<th>Bayesian MTBOMF</th>
<th>Bayesian Reliability at 720hrs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TSCE</strong></td>
<td>4500 hrs (1156 hrs, 42710 hrs)</td>
<td>0.85 (0.54,0.98)</td>
<td>3630 hrs (1179 hrs, 6753 hrs)</td>
<td>0.73 (0.54,0.90)</td>
</tr>
<tr>
<td><strong>SSC</strong></td>
<td>667 hrs (299 hrs, 1814 hrs)</td>
<td>0.33 (0.09,0.67)</td>
<td>697 hrs (332 hrs, 1172 hrs)</td>
<td>0.31 (0.11,0.54)</td>
</tr>
<tr>
<td><strong>Comm</strong></td>
<td>&gt; 2796 hrs</td>
<td>&gt; 0.77*</td>
<td>10320 hrs (1721 hrs, 18210 hrs)</td>
<td>0.83 (0.66,0.96)</td>
</tr>
<tr>
<td><strong>SEW</strong></td>
<td></td>
<td>0.82 (0.58,0.95)</td>
<td></td>
<td>0.77 (0.62,0.91)</td>
</tr>
<tr>
<td><strong>Core Mission</strong></td>
<td></td>
<td>??????</td>
<td></td>
<td>0.15 (0.05, 0.27)</td>
</tr>
</tbody>
</table>

*Zero failures occurred in the notional on-demand system data*

**Note the impact of the prior is greater in the one failure system**

**Many ways to think about calculating this, none of which are particularly satisfactory**

**Full Mission mean is comparable with the simple point estimate**

---

TSCE: Total Ship Computing Environment  
SSC: Sea Sensors and Controls  
Comm: Communications  
SEW: Sea Engagement Weapons
Core Mission Reliability Over Time

Posterior reliability as a function of time for TSCE (red), SSC (blue), and Communications (green)
Reporting of Results

Posterior mean and 80% intervals for each subsystem and the total system reliability over 15 days (light blue) and 30 days (dark blue) for the notional example.
Value of Bayesian Statistics for LCS

- Avoids unrealistic reliability estimates when there are no observed failures.

- In our notional example (zero failures for the Communications system), the Bayesian approach helped us solve an otherwise intractable problem.

- Obtaining interval estimates is straightforward for system reliability
  - Frequentist methods would have to employ the Delta method, Normal approximations, or bootstrapping.

- Flexibility in developing system models
  - We used a series system for the core mission reliability
  - Many other system models are possible and we can still get full system reliability estimates with intervals.
Case Study: Biochem Detection System

- Bio-chemical Detection System analyzes environmental samples and identifies chemical, biological, radiological agents. Each subsystem is comprised of a collection of components of various sensitivity.

- KPP performance requirement for each subsystem: detect 85-90% of samples that come into the lab.

- Multiple Tiers of testing
  - Tier 2: component level testing with agent in pristine matrix to each device (vendor testing)
    » 5 components: total of almost 2000 runs
  - Tier 3: component level testing with agent in various matrices, such as soil, food, or swab (vendor testing)
    » 8 components: total of about 3600 runs
  - DT/OT: subsystem level test with agent in matrices, triage procedures (government testing)
    » 80-90 samples tested on multiple components, final call made by operator based on component output
Case Study: Biochem Detection System

- DT/OT: set concentration levels, comparatively small sample size

- Standard logistic regression on the Tier 3 data could be problematic
  - All detections or non-detections

- Bayesian approach with a dispersed prior:
  \[
  \text{logit}(P_D) = \beta_1 \cdot \text{conc} + \beta_2^{\text{matrix}} + \beta_3^{\text{agent}}
  \]
  \[
  (\beta_1, \beta_2, \beta_3) \sim \text{Multivariate Normal}(0, W)
  \]
  - Explicitly forcing a dependence on concentration.
  - Leverage all device runs to learn about each agent/matrix combination performance curve.
```r
post=function(beta1, beta2, beta3) {
  betavec=c(beta1, beta2, beta3)

  val=dtnorm(betavec[1], 0, 10^-3, lower=0, log=TRUE) +
  dmvnorm(betavec[-1], rep(0, length(betavec[-1])), diag(rep(10^-3, length(betavec)-1)), log=TRUE) +
  sum(dbinom(y, 1, mylogit(beta1*log(X[, 1]) + beta2[X[, 2]] + beta3*X[, 3]), log=TRUE))

  return(val)
}
```
> for(i in 1:size){
+    #update beta1
+    cand.beta1=rnorm(1,beta1,0.1)
+    if(cand.beta1 > 0){
+      r = post(cand.beta1,beta2,beta3) - post(beta1,beta2,beta3)
+      u = runif(1) <= exp(r)
+      arate1 = arate1 + u
+      beta1 = cand.beta1*(u==1) + beta1*(u==0)
+    }
+
+    #update beta2
+    for(j in 1:length(unique(X[,2]))){
+      cand.beta2=beta2
+      cand.beta2[j]=rnorm(1,beta2[j],0.7)
+      r = post(beta1,cand.beta2,beta3) - post(beta1,beta2,beta3)
+      u = runif(1) <= exp(r)
+      arate2[j] = arate2[j] + u
+      beta2[j] = cand.beta2[j]*(u==1) + beta2[j]*(u==0)
+    }
+
+    #update beta3
+    for(j in 2:length(unique(X[,3]))){
+      cand.beta3=beta3
+      cand.beta3[j]=rnorm(1,beta3[j],0.7)
+      r = post(beta1,beta2,cand.beta3) - post(beta1,beta2,beta3)
+      u = runif(1) <= exp(r)
+      arate3[j] = arate3[j] + u
+      beta3[j] = cand.beta3[j]*(u==1) + beta3[j]*(u==0)
+    }
+
+    par[i,] = c(beta1,beta2,beta3) #save current parameter values
+  }

4/27/2016-27
R code implementation

```r
# update beta2
for(j in 1:length(unique(X[,2]))) {
    cand.beta2[j] = rnorm(1, beta2[j], 0.7)
    r = post(beta1, cand.beta2, beta3) - post(beta1, beta2, beta3)
    u = runif(1) <= exp(r)
    arate2[j] = arate2[j] + u
    beta2[j] = cand.beta2[j]*(u==1) + beta2[j]*(u==0)
}
```
Component: Chemical Detector 1
Agent: Chemical A
Matrix: Soil, Swab
Case Study: Biochem Detection System

**Chemical Detector 2**

**Biological Detector 1**
Case Study: Biochem Detection System

Chemical Detector 3

Chemical Detector 4
Value of Bayesian Statistics for Biochem Detection System

• Tier 2 and Tier 3 produced a lot of data which we can leverage to make informed decisions.

• By knowing the concentrations of agents within various matrices that each component can detect, we can determine concentrations that the system of devices might be easy or difficult for the operators to identify in DT/OT.

• This analysis can serve as the basis for the analysis of the DT/OT data.
Discussion: When Is it Worth the Effort?

• Inclusion of prior information from prior testing, modeling and simulation, or engineering analyses only when it is relevant to the current test. We do not want to bias the OT results.

• Even when including prior information, the prior must have enough variability to allow the estimates to move away from what was previously seen if the data support such values.

• We can use very flexible models for many types of test data (e.g. kill chains, complex system structures, linking EFFs to SA) and obtain estimates more readily than with the frequentist paradigm. The model and assumptions have to make sense for the test at hand.
Discussion: Other Resources

• For other R packages that provide easy to implement tools and short but informative how to guides with examples, see
  https://cran.r-project.org/web/views/Bayesian.html

  ➢ arm, bayesm, and bayesSurv are good places to start

As with any new statistical method, it is important to have an expert review your work the first few times you apply these techniques. There are many ways to accidentally do bad statistics!
Backup Slides
## Common Conjugate Models

<table>
<thead>
<tr>
<th>Likelihood</th>
<th>Parameter</th>
<th>Prior</th>
<th>Posterior</th>
</tr>
</thead>
<tbody>
<tr>
<td>Binomial((s+f, R))</td>
<td>(0 \leq R \leq 1)</td>
<td>Beta((a,b)) (a &gt; 0, b &gt; 0)</td>
<td>Beta((a',b')) (a' = a + s) (b' = b + f)</td>
</tr>
<tr>
<td>Poisson((\lambda))</td>
<td>(\lambda &gt; 0)</td>
<td>Gamma((a,b)) (a &gt; 0, b &gt; 0)</td>
<td>Gamma((a',b')) (a' = a + n) (b' = b + \Sigma t)</td>
</tr>
<tr>
<td>Exponential((\lambda))</td>
<td>(\lambda &gt; 0)</td>
<td>Gamma((a,b)) (a &gt; 0, b &gt; 0)</td>
<td>Gamma((a',b')) (a' = a + n) (b' = b + \Sigma t)</td>
</tr>
</tbody>
</table>

**For more examples, see Wikipedia page or “Bayesian Reliability” pg 48**