Advanced Analysis Methodologies
Advanced Methodologies

• Censored Data
  – Converting to continuous data often presents analysis challenges
  – For example, if we use detection range, how do we account for non-detects in the analysis
  – Censored data provides a solution

• Generalized Linear Models
  – System performance is often best characterized by non-normal data
    » Time
    » Accuracy
    » Pass/Fail
  – Generalized linear models provide a more flexible analysis framework to handle these non-normal outcomes.

• Bayesian Methodologies
  – Allow for the incorporation of multiple sources of information, when it is appropriate
  – Provide methodologies for finding confidence intervals when there are zero observations
Motivating Example: Submarine Detection Time

- **System Description**
  - Sonar system replica in a laboratory on which hydrophone-level data, recorded during real-world interactions can be played back in real-time.
  - System can process the raw hydrophone-level data with any desired version of the sonar software.
  - Upgrade every two years; test to determine new version is better
  - Advanced Processor Build (APB) 2011 contains a potential advancement over APB 2009 (new detection method capability)

- **Response Variable: Detection Time**
  - Time from first appearance in recordings until operator detection
    - Failed operator detections resulted in right censored data

- **Factors:**
  - Operator proficiency (quantified score based on experience, time since last deployment, etc.)
  - Submarine Type (SSN, SSK)
  - System Software Version (APB 2009, APB 2011)
  - Array Type (A, B)
  - Target Loudness (Quiet, Loud)
Detection Time Distribution

- Detection time does not follow a normal distribution
Failed Detection Opportunities

Not all runs resulted in a successful detections

<table>
<thead>
<tr>
<th>Type</th>
<th>SSK</th>
<th>SSN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Array</td>
<td>16</td>
<td>16</td>
</tr>
<tr>
<td>Noise</td>
<td>29</td>
<td>29</td>
</tr>
</tbody>
</table>

Detection

- Yes
- No

Detection Time (min)

APB

9  11  9  11  9  11  9  11  9  11  9  11
Submarine Detection Time: Analysis

- Advanced statistical modeling techniques incorporated all of the information across the operational space.
  - Generalized linear model with log-normal detection times
  - Censored data analysis accounts for non-detects

- All factors were significant predictors of the detection time

<table>
<thead>
<tr>
<th>Factor/Model Term</th>
<th>Description of Effect</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition Factor</td>
<td>Increased recognition factors resulted in shortened detection times</td>
<td>0.0227</td>
</tr>
<tr>
<td>APB</td>
<td>Detection time is shorter for APB-11</td>
<td>0.0025</td>
</tr>
<tr>
<td>Target Type</td>
<td>Detection time is shorter for SSN targets</td>
<td>0.0004</td>
</tr>
<tr>
<td>Target Noise Level</td>
<td>Detection time is shorter for loud targets</td>
<td>0.0012</td>
</tr>
<tr>
<td>Array Type</td>
<td>Detection time is shorter for Array B</td>
<td>0.0006</td>
</tr>
<tr>
<td>Type* Noise</td>
<td>Additional model terms improve predictions. Third order interaction is marginally significant, therefore all second order terms are retained.</td>
<td>0.0628</td>
</tr>
<tr>
<td>Type* Array</td>
<td></td>
<td>0.9091</td>
</tr>
<tr>
<td>Noise*Array</td>
<td></td>
<td>0.8292</td>
</tr>
<tr>
<td>Type* Noise*Array</td>
<td></td>
<td>0.0675</td>
</tr>
</tbody>
</table>
Submarine Detection Time: Results

- Median detection times show a clear advantage of APB-11 over the legacy APB.
- Confidence interval widths reflect weighting of data towards APB-11.
- Statistical model provides insights in areas with limited data.
  - Note median detection time in cases with heavy censoring is shifted higher.
Introduction to Censored Data Analysis

- Censored data = we didn’t observe the detection directly, but we expect it will occur if the test had continued
  - We cannot make an exact measurement, but there is information we can use. The no detects are on the tail of the distribution!
  - Same concept as a time-terminated reliability trials (failure data)

<table>
<thead>
<tr>
<th>Run No.</th>
<th>Result</th>
<th>Result Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Detected Target</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>Detected Target</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>No detect</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>Detected Target</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>Detected Target</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>Detected Target</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>No detect</td>
<td>0</td>
</tr>
<tr>
<td>8</td>
<td>No detect</td>
<td>0</td>
</tr>
<tr>
<td>9</td>
<td>Detected Target</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>Detected Target</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Run No.</th>
<th>Time of Detection (hours after COMEX)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.4</td>
</tr>
<tr>
<td>2</td>
<td>2.7</td>
</tr>
<tr>
<td>3</td>
<td>&gt;6.1</td>
</tr>
<tr>
<td>4</td>
<td>2.5</td>
</tr>
<tr>
<td>5</td>
<td>3.5</td>
</tr>
<tr>
<td>6</td>
<td>5.3</td>
</tr>
<tr>
<td>7</td>
<td>&gt;6.2</td>
</tr>
<tr>
<td>8</td>
<td>&gt;5.8</td>
</tr>
<tr>
<td>9</td>
<td>1.8</td>
</tr>
<tr>
<td>10</td>
<td>2.7</td>
</tr>
</tbody>
</table>

= Detect  = No-Detect
• Assume that the time data come from an underlying distribution, such as the log-normal distribution
  – Other distributions may apply – you must consider carefully. See slide 4 where we did it for the submarine detection data

• That parameterization will enable us to link the time metric to the probability of detection metric.

**Probability Density Function**

![Probability Density Function](image)

**Cumulative Distribution Function (CDF)**

![Cumulative Distribution Function](image)
Parameterizing Data

- Example: Aircraft must detect the target within its nominal time on station (6-hours)
  - Binomial metric was detect/non-detect within time-on-station

- If we determine the shape of this curve (i.e., determine the parameters of the PDF/CDF), we can use the time metric to determine the probability to detect!

![Probability Density Function](image)

![Cumulative Distribution Function (CDF)](image)
Conceptualizing the Censored-Data Fit

- For non-censored measurements, the PDF fit is easy to conceptualize.
- For censored measurements, the data can’t define the PDF, but we know they contribute to the probability density beyond the censor point.
- Example event from an OT:
  - No Detects (Detect Time > 6 hours) lie somewhere on the tail of the distribution.
  - Detect will eventually occur sometime after 6 hours, pushing the distribution curve to the right.
  - Mathematically, there are ways of calculating the shifted distribution.

Including a bunch of censored (Time > 6 hour) events will push the CDF to the right (see how probability to detect is lower at 6 hours).
Characterizing Performance with Censored Data

- Now let’s employ DOE...

- Consider a test with 16 runs
  - **Two** factors examined in the test
  - Run Matrix:

<table>
<thead>
<tr>
<th></th>
<th>Target Fast</th>
<th>Target Slow</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Location 1</strong></td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td><strong>Test Location 2</strong></td>
<td>4</td>
<td>4</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td><strong>8</strong></td>
<td><strong>8</strong></td>
<td><strong>16</strong></td>
</tr>
</tbody>
</table>

- Detection Results:

<table>
<thead>
<tr>
<th></th>
<th>Target Fast</th>
<th>Target Slow</th>
<th>Totals</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Location 1</strong></td>
<td>3/4</td>
<td>4/4</td>
<td>7/8 (0.875)</td>
</tr>
<tr>
<td><strong>Test Location 2</strong></td>
<td>3/4</td>
<td>1/4</td>
<td>4/8 (0.5)</td>
</tr>
<tr>
<td></td>
<td><strong>6/8 (0.75)</strong></td>
<td><strong>5/8 (0.63)</strong></td>
<td></td>
</tr>
</tbody>
</table>
Attempt to Characterize Performance

• As expected, 4 runs in each condition is insufficient to characterize performance with a binomial metric.

• Cannot tell which factor drives performance or which conditions will cause the system to meet/fail requirements.

• Likely will only report a ‘roll-up’ of 11/16
  – 90% confidence interval: [0.45, 0.87]
Characterizing Performance Better

• Measure *time-to-detect* in lieu of binomial metric, employ censored data analysis…

• Significant reduction in confidence intervals!
  – Now can tell significant differences in performance
    » E.g., system is performing *poorly* in Location 2 against slow targets
  – We can confidently conclude performance is above threshold in three conditions
    » Not possible with a “probability to detect” analysis!

![Graph comparing Binomial Analysis and Censored Data Analysis](image)
Censored Data Analysis Summary

- Many binary metrics can be recast using a continuous metric
  - Care is needed, does not always work, but…
  - Cost saving potential is too great not to consider it!

- With Censored-data analysis methods, we retain the binary information (non-detects), but gain the benefits of using a continuous metric
  - Better information for the warfighter
  - Maintains a link to the “Probability of…” requirements

- Converting to the censored-continuous metric maximizes test efficiency
  - In some cases, as much as 50% reduction in test costs for near identical results in percentile estimates
  - Benefit is greatest when the goal is to identify significant factors (characterize performance)
Generalized Linear Models Overview

• **There are many classes of statistical models:**
  – General linear models (normal distribution)
  – Generalized linear models (Exponential family)
    » Provides a simplified framework for numerically maximizing the likelihood
  – Location-scale regression (location scale, log-location scale)
  – Nonlinear regression (almost everything else)

• **These regression analyses are a logical extension of standard statistical regression analysis**

• **However, methods presented here are more general**
  – Data not necessarily normal
  – Data may not have constant variance
  – Lind between data and response may not be linear

• **Practical T&E problems often cannot be solved with straightforward regression analysis**
Model Specification:
GLM versus Generalized Linear Model

- **General Linear Model (e.g., regression)**
  - Model: \( f(y) \sim \text{Normal}(\mu, \sigma) \)
    \[
    \mu = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_i x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \text{h.o.t.}
    \]
  - Where, \( k \) is the number of factors and h.o.t. are higher order terms.

- **Generalized Linear Model**
  - Model:
    \[
    f(y) \sim \text{ExponentialFamilyDistribution}(\alpha, \beta) \quad E(Y) = \mu = f(\alpha, \beta)
    \]
    \[
    \mu = g^{-1}\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_i x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \text{h.o.t}\right)
    \]

\(g^{-1}\) is the inverse “link function” – it literally links the factors to the expected value of the response.
Exponential Family

- Class of distributions that provides the basis for Generalized Linear Models

- Distributions include:
  - Continuous
    » Normal
    » Log-normal
    » Beta
    » Gamma
    » Exponential
  - Discrete:
    » Binomial/Bernoulli
    » Poisson
    » Negative Binomial
  - And several more!

- Provide flexible shapes that can be used to describe almost any type of data!
Pass/Fail Analysis: A Second Motivating Example

- System’s goal is to maintain a lock on a moving target

- **Response Variable: Maintain track? (Yes/No)**
  - Debatable if a continuous metric could have replaced this binary response. However, no continuous metric was tracked during the test, so we are stuck analyzing pass/fail response.

- **Factors:**
  - Target Size (small/large)
  - Target Speed (slow/fast)
  - Time of Day (day/night)
  - Target Aspect (frontal/quarter)
  - Maneuvering (yes/no)

- **Generalized linear models can be used to fit logistic and probit regression under the same framework!**
Generalized Linear Model: Break Lock?

- Logistic Regression Model:

\[ f(y) \sim \text{Binomial}(n, p) \]
\[ \mu = np \]
\[ \mu = \frac{\exp \left( \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_i x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \text{h.o.t} \right)}{1 + \exp \left( \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_i x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + \text{h.o.t} \right)} \]

* In JMP: Fit Model → Generalized Linear Model → Binomial Distribution → Logit Link
Summarizing Results

- **Day** vs. **Night**
- **Fast** vs. **Slow**
- **Large Targets** vs. **Small Targets**
- 80% Confidence Intervals Shown
There is a model for every situation!

- Normality
- Homoscedasticity
- Independence
- Linearity

Regression/ANOVA

General Linear Models

Generalized Linear Models

Location-Scale Models, Non-parametric Models

Generalized Linear Mixed Models

- x2 for Bayesian versions of these model forms, which can also incorporate prior knowledge
- Note, Bayesian methodologies can make analysis easier by avoiding the complex optimization problem.
Bayesian Methodology – Overview

Model for Data

Data

Likelihood $L(data \mid \theta)$

Prior $f(\theta)$

Classical Statistics

Inference

Posterior $f(\theta \mid data)$

The inclusion of the prior distribution allows us to incorporate different types of information in the analysis
Motivating Example: Stryker Reliability Analysis

- Statistical methods (including DOE) apply to reliability data as well as performance data
- **Stryker Retrospective Case Study**
  - Infantry Carrier Vehicle (ICV) - the infantry/mission-vehicle type
  - Base vehicle for eight separate configurations
  - IOT&E Results:

<table>
<thead>
<tr>
<th>Vehicle Variant</th>
<th>Total Miles Driven</th>
<th>System Aborts</th>
<th>MMBSA</th>
<th>MMBSA 95% LCL</th>
<th>MMBSA 95% UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antitank Guided Missile Vehicle (ATGMV)</td>
<td>10334</td>
<td>12</td>
<td>861</td>
<td>493</td>
<td>1667</td>
</tr>
<tr>
<td>Commander's Vehicle (CV)</td>
<td>8494</td>
<td>1</td>
<td>8494</td>
<td>1525</td>
<td>335495</td>
</tr>
<tr>
<td>Engineer Squad Vehicle (ESV)</td>
<td>3771</td>
<td>13</td>
<td>290</td>
<td>170</td>
<td>545</td>
</tr>
<tr>
<td>Fire Support Vehicle (FSV)</td>
<td>2306</td>
<td>1</td>
<td>2306</td>
<td>414</td>
<td>91082</td>
</tr>
<tr>
<td>Infantry Carrier Vehicle (ICV)</td>
<td>29982</td>
<td>35</td>
<td>857</td>
<td>616</td>
<td>1230</td>
</tr>
<tr>
<td>Mortar Carrier Vehicle (MCV)</td>
<td>4521</td>
<td>4</td>
<td>1130</td>
<td>441</td>
<td>4148</td>
</tr>
<tr>
<td>Medical Evacuation Vehicle (MEV)</td>
<td>1967</td>
<td>0</td>
<td>-</td>
<td>657</td>
<td>-</td>
</tr>
<tr>
<td>Reconnaissance Vehicle (RV)</td>
<td>5374</td>
<td>2</td>
<td>2687</td>
<td>744</td>
<td>22187</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>66749</strong></td>
<td><strong>68</strong></td>
<td><strong>982</strong></td>
<td><strong>774</strong></td>
<td><strong>1264</strong></td>
</tr>
</tbody>
</table>

- Results do not leverage DT data or relationships between vehicles
The Stryker Reliability Data Set

Vehicle Type

Developmental Testing

Operational Testing

Miles Before System Abort

Exact Failure

Right Censored

Vehicle Type
Bayesian Analysis for Incorporating Developmental Test

- **Informative Priors**
  - Based on subject matter expertise (there will be a degradation in OT reliability)
    » Data is already included in model

- **Hierarchical Models**
  - Assumes the parameters are related, the data tells us how closely related
  - Hierarchical models for the Stryker case study allow us to estimate MEV reliability based on other data

**Bayesian Analysis Model:**

\[
t_{DT} \sim \text{exp}(\lambda_i) \quad t_{OT} \sim \text{exp}(\lambda_i/\eta)
\]

\[i = 1,2,\ldots,8\text{ (vehicle variants including MEV)}\]

\[\lambda_i \sim \text{gamma}(a,b)\]

\[\eta \sim \text{beta}(1,1)\]

\[a \sim \text{gamma}(0.001, 0.001)\]

\[b \sim \text{gamma}(0.001, 0.001)\]
Stryker Reliability Results

**Traditional Approach:**

\[ MMBSA = \frac{\text{Miles}}{\# \text{Failures}} \]

- Extremely wide confidence intervals
- Results in unrealistic estimates for the Commander’s Vehicle

**Exponential Regression Approach & Bayesian Approaches**

\[ MMBSA = f(\text{TestPhase, Variant}) \]

- Allows for a degradation in MMBSA from DT to OT (increases could occur as well).
- Leverages all information
  - Better estimates of MMBSA
  - Tighter confidence intervals
Bayesian Methods Summary

• Provide very flexible analysis methods

• Priors allow us to consider other types of data, basing decisions on all available information about a system

• Methods can easily be extended to incorporate other situations:
  – Kill chain analysis
  – Complex system structures reliability analysis
  – Incorporate any relevant prior testing, modeling and simulation, or engineering analysis