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)

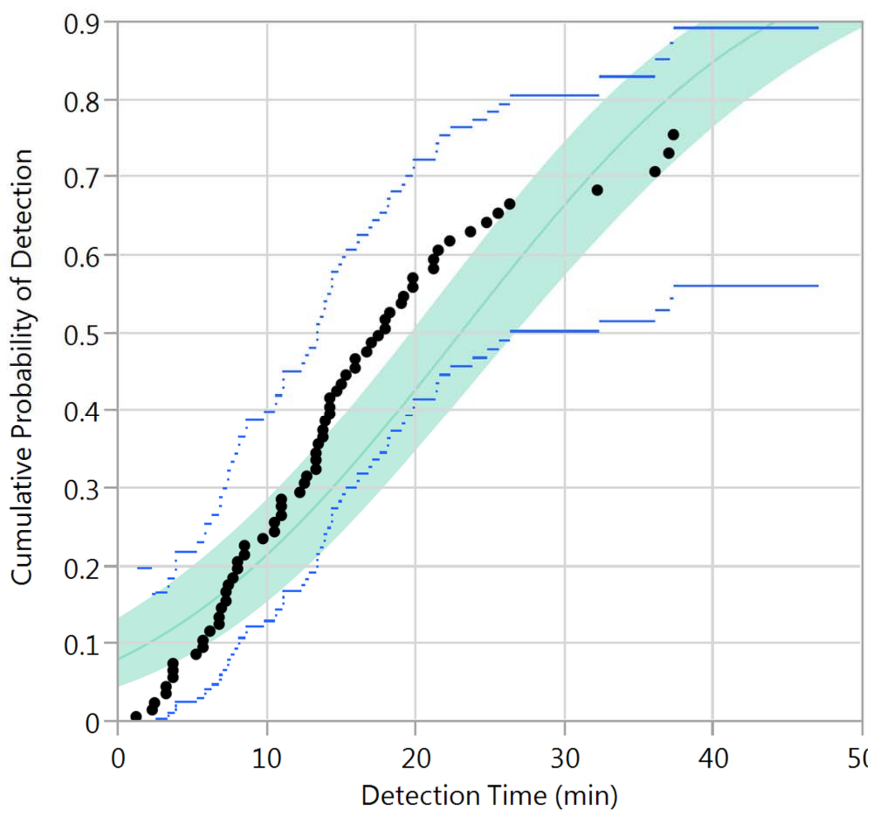
- **Response Variable:** Detection Time
  - Time from first appearance in recordings until operator detection
    - » Failed operator detections resulted in *right censored data*

- **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)

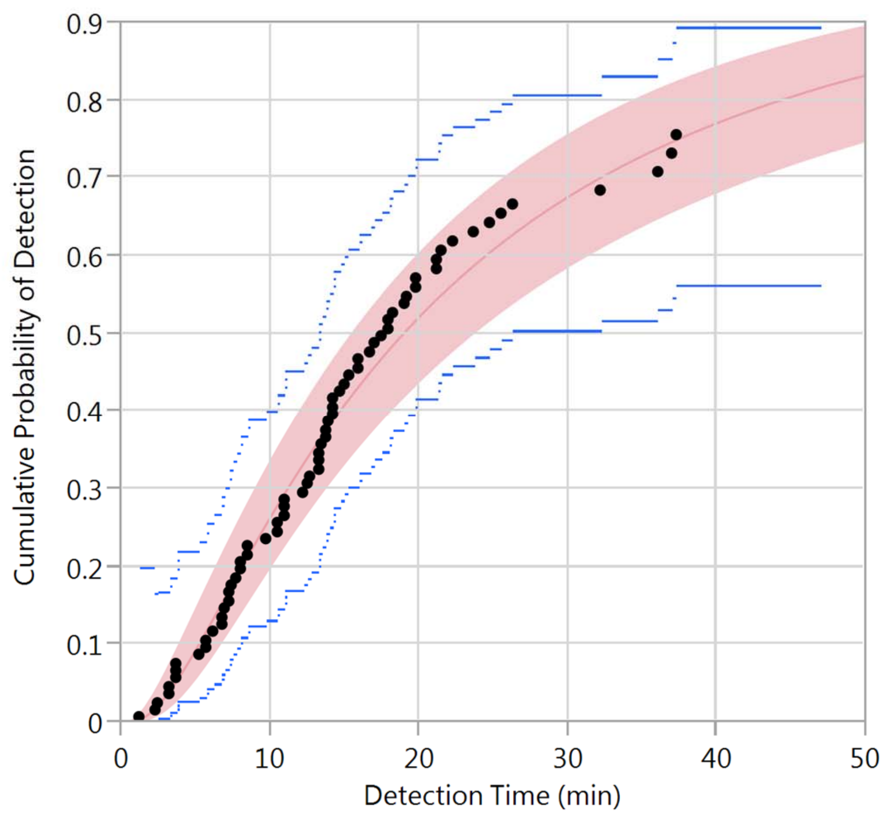
Detection Time Distribution

- **Detection time does not follow a normal distribution**

Normal Distribution

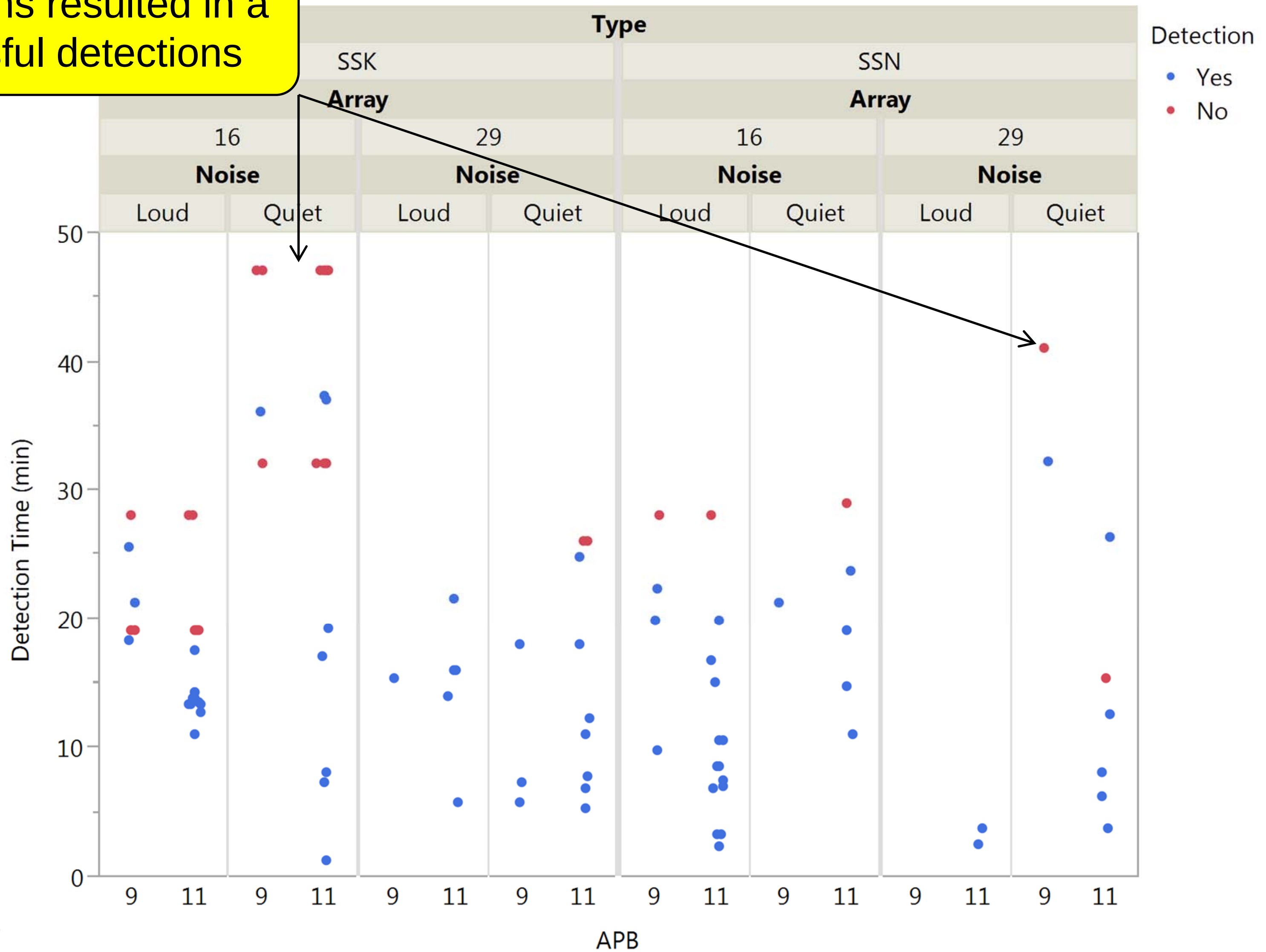


Lognormal Distribution



Failed Detection Opportunities

Not all runs resulted in a successful detections



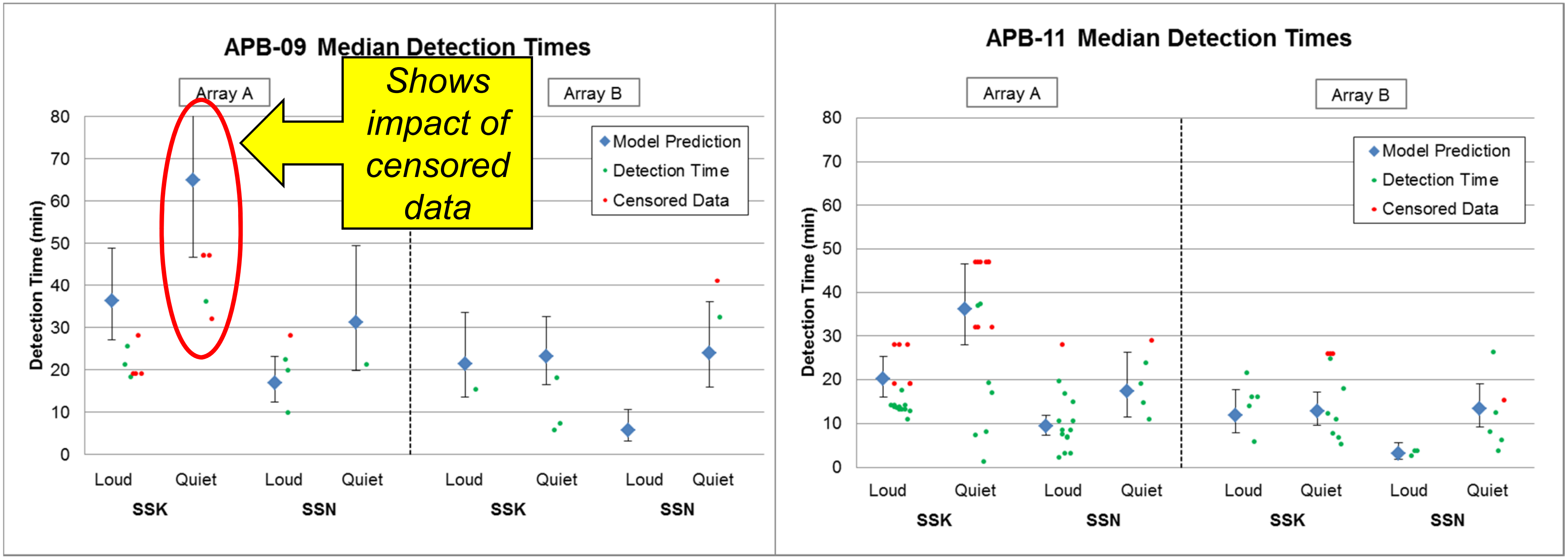
Submarine Detection Time: Analysis

- **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**

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

Submarine Detection Time: Results



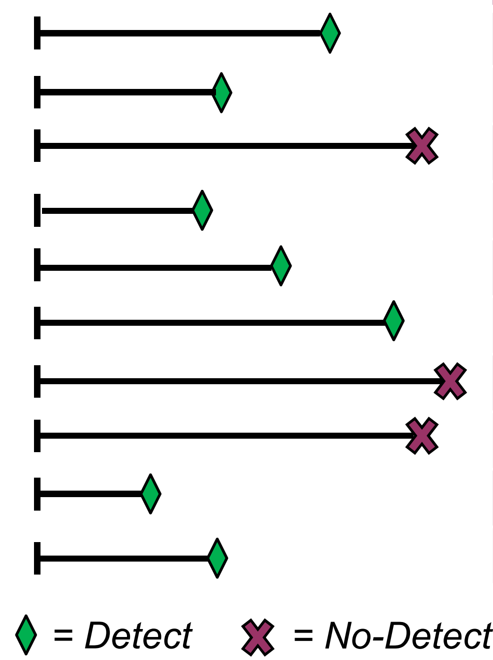
- **Median detection times show a clear advantage of APB-11 over the legacy APB**
- **Confidence interval widths reflect weighting of data towards APB-11**
- **Statistical model provides insights in areas with limited data**
  - Note median detection time in cases with heavy censoring is shifted higher

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)

Run No.	Result	Result Code
1	Detected Target	1
2	Detected Target	1
3	No detect	0
4	Detected Target	1
5	Detected Target	1
6	Detected Target	1
7	No detect	0
8	No detect	0
9	Detected Target	1
10	Detected Target	1

Corresponding Timelines



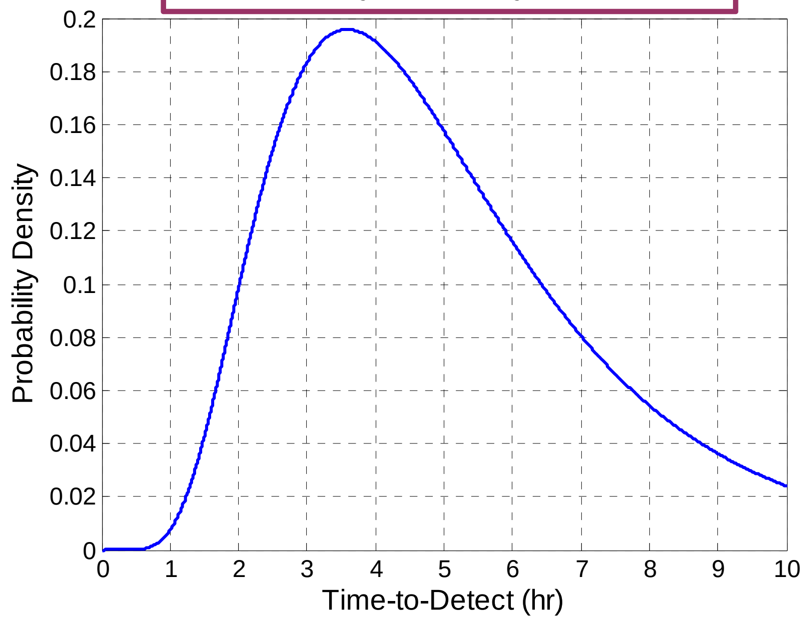
Run No.	Time of Detection (hours after COMEX)
1	**4.4**
2	**2.7**
3	**>6.1**
4	**2.5**
5	**3.5**
6	**5.3**
7	**>6.2**
8	**>5.8**
9	**1.8**
10	**2.7**

Parameterizing Data

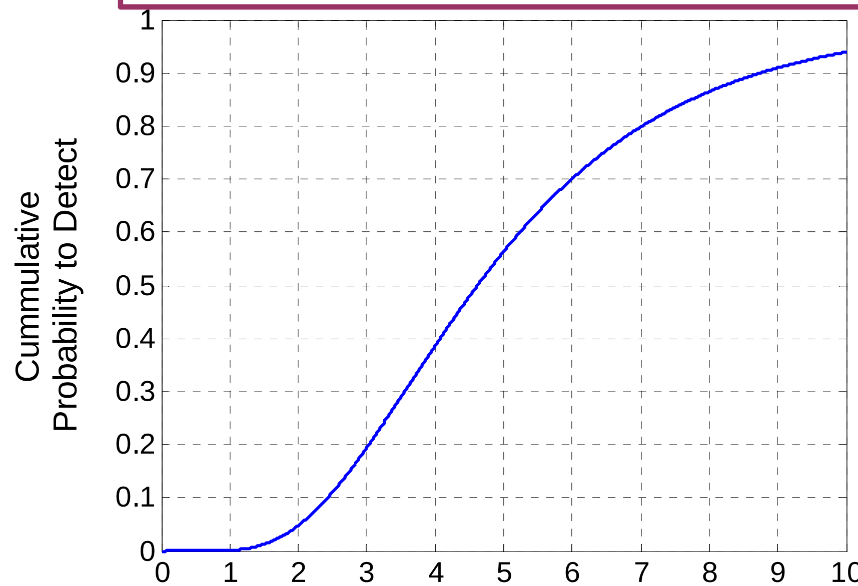
- **Assume that the time data come from an underlying distribution, such as the log-normal distribution**
  - Other distributions may apply – you must consider carefully. See slide 4 where we did it for the submarine detection data

- **That parameterization will enable us to <u>link</u> the time metric to the probability of detection metric.**

Probability Density Function



Cumulative Distribution Function (CDF)

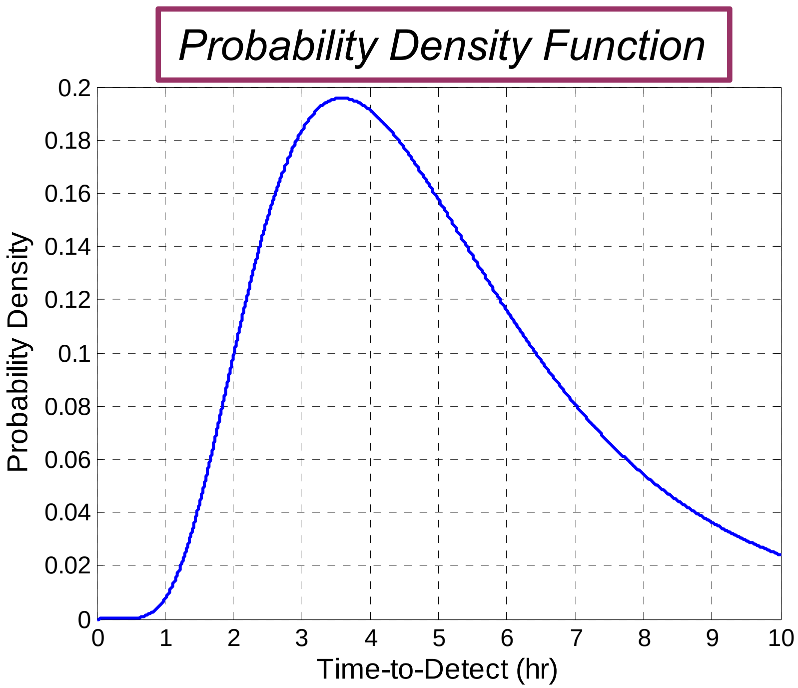


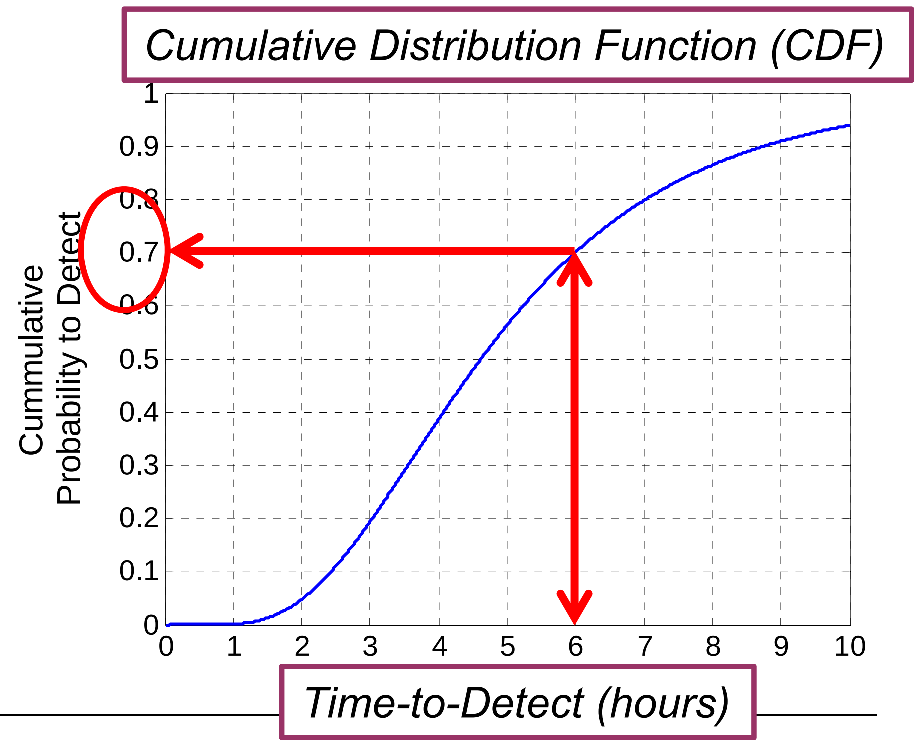
Time-to-Detect (hours)

Parameterizing Data

- **Example: Aircraft must detect the target within it's nominal time on station (6-hours)**
  – Binomial metric was detect/non-detect within time-on-station

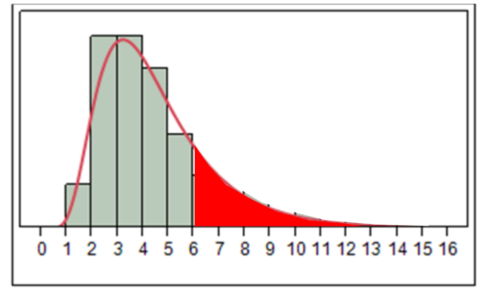
- **If we determine the shape of this curve (i.e., determine the parameters of the PDF/CDF), we can use the time metric to determine the probability to detect!**



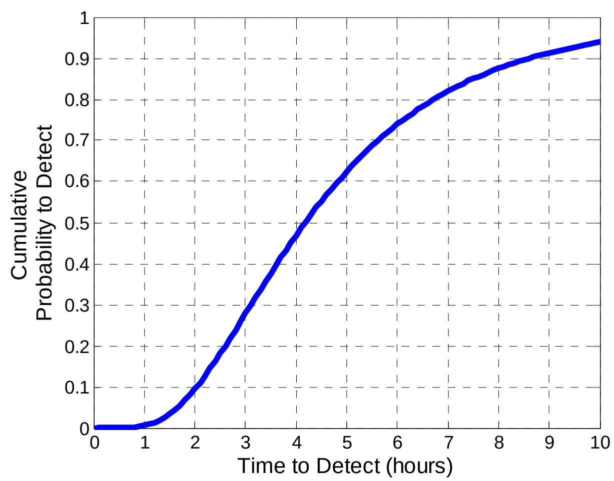


Conceptualizing the Censored-Data Fit

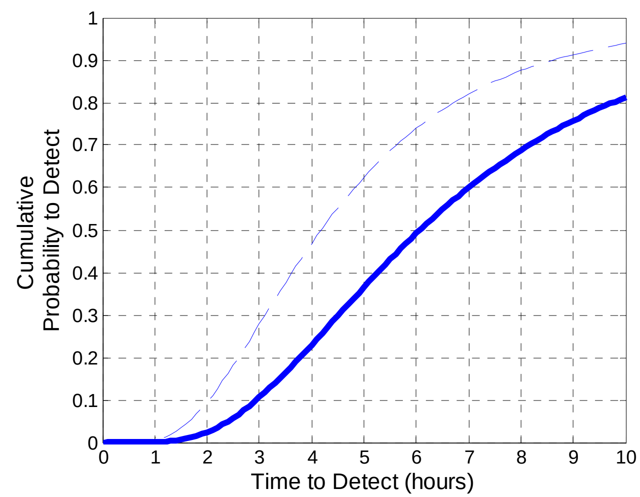
- **For non-censored measurements, the PDF fit is easy to conceptualize**



- **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**
  - **<u>Two</u>** factors examined in the test
  - Run Matrix:

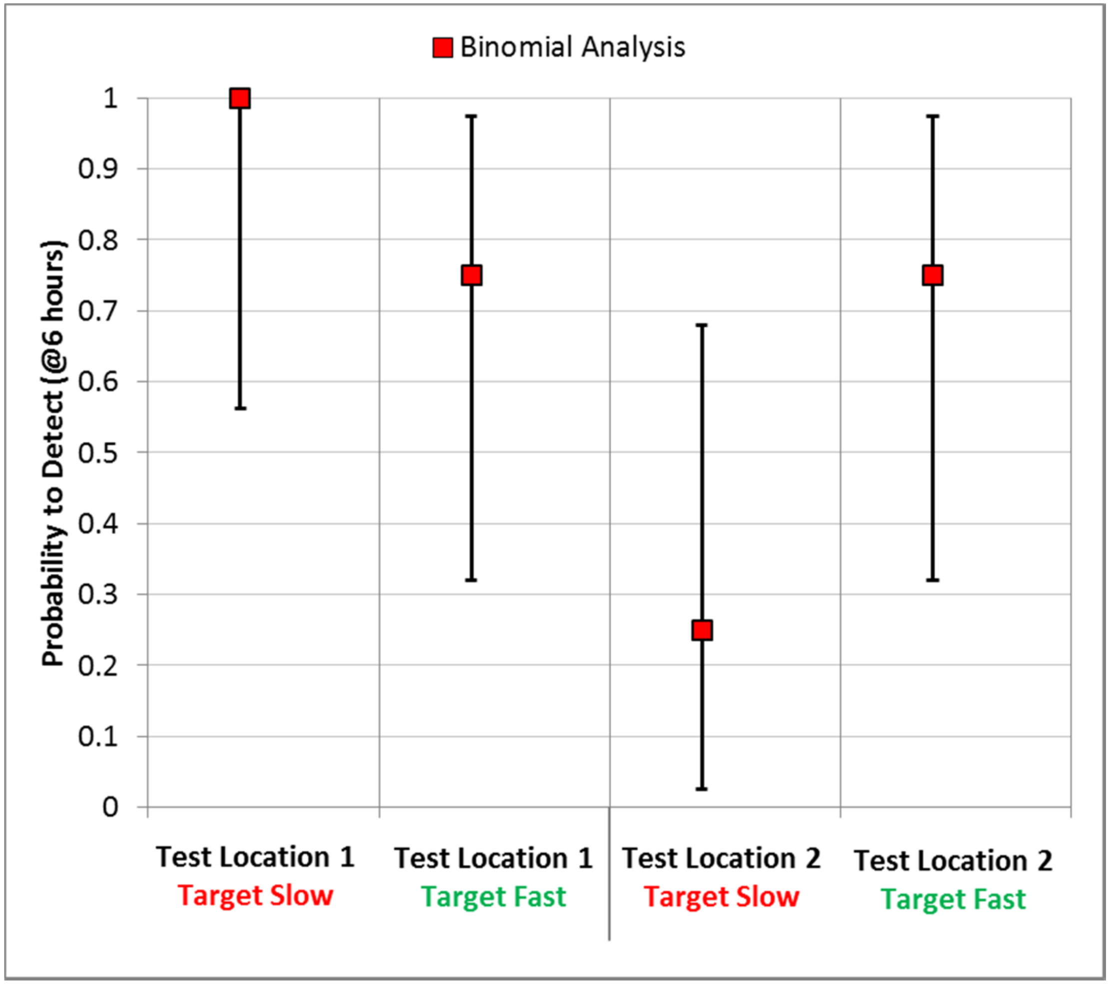
	Target Fast	Target Slow	Totals
Test Location 1	4	4	8
Test Location 2	4	4	8
	8	8	16

  - Detection Results:

	Target Fast	Target Slow	Totals
Test Location 1	3/4	4/4	7/8 (0.875)
Test Location 2	3/4	1/4	4/8 (0.5)
	6/8 (0.75)	5/8 (0.63)	

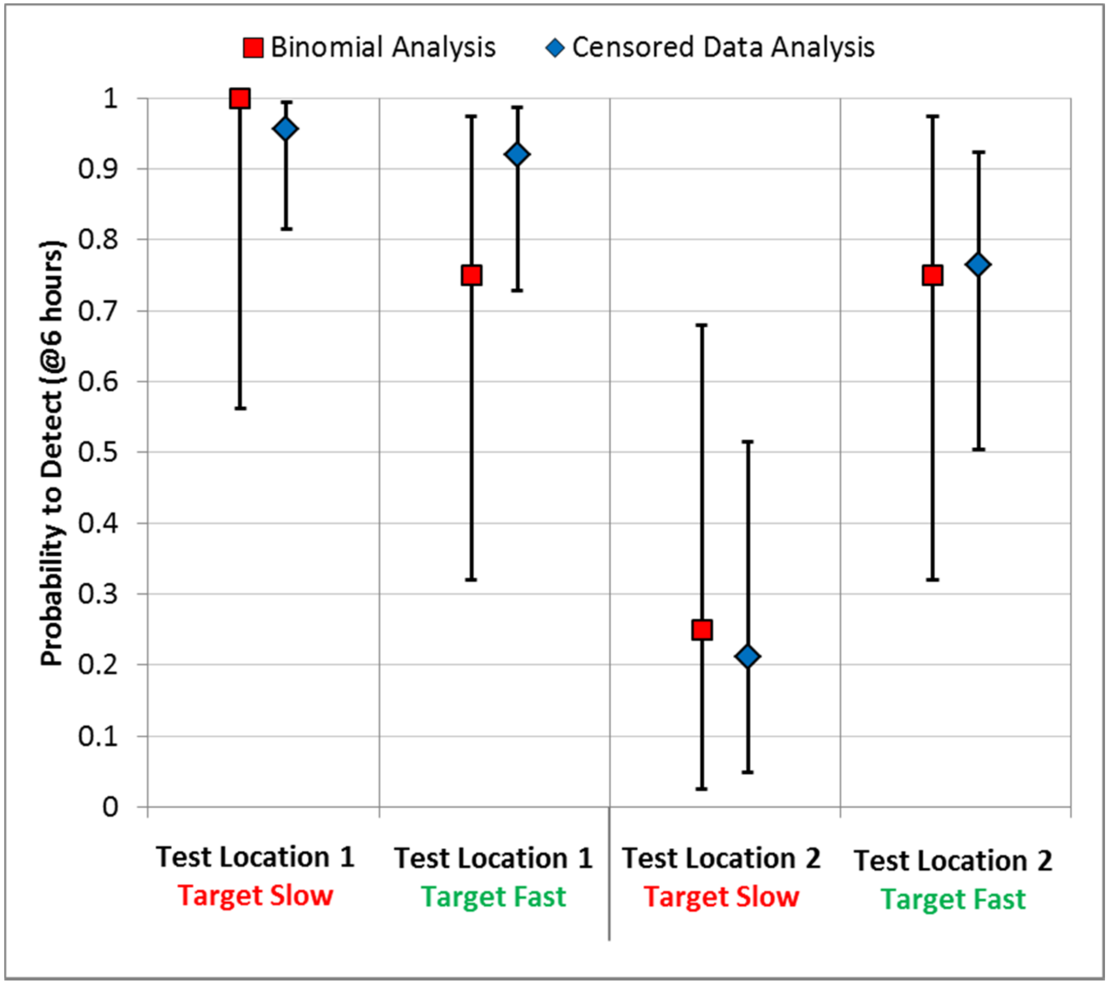
Attempt to Characterize Performance

- **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!



Censored Data Analysis Summary

- **Many binary metrics can be recast using a continuous metrics**
  - 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: $f(y) \sim Normal(\mu, \sigma)$

$$\mu = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_i x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + h.o.t.$$

  - Where, k is the number of factors and h.o.t. are higher order terms.

- **Generali<u>zed</u> Linear Model**
  - Model:

$g^{-}()$ is the inverse "link function" – it literally links the factors to the expected value of the response

$$f(y) \sim ExponentialFamilyDistribution(\alpha, \beta)$$

$$E(Y) = \mu = f(\alpha, \beta)$$

$$\mu = g^{-1}\left(\beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=1}^{k} \beta_i x_i^2 + \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \beta_{ij} x_i x_j + h.o.t \right)$$

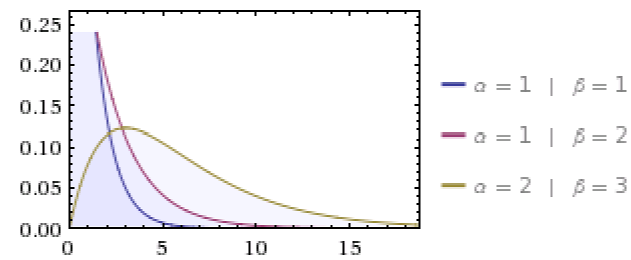
Exponential Family

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

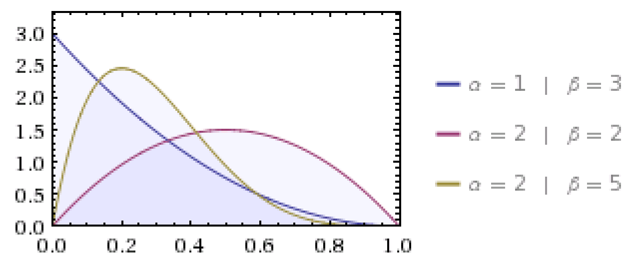
- **Distributions include:**
  - Continuous
    - » Normal
    - » Log-normal
    - » Beta
    - » Gamma
    - » Exponential
  - Discrete:
    - » Binomial/Bernoulli
    - » Poisson
    - » Negative Binomial
  - And several more!

Logistic Regression is a Generalized Linear Model

Gamma Distribution



Beta Distribution



- **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:**

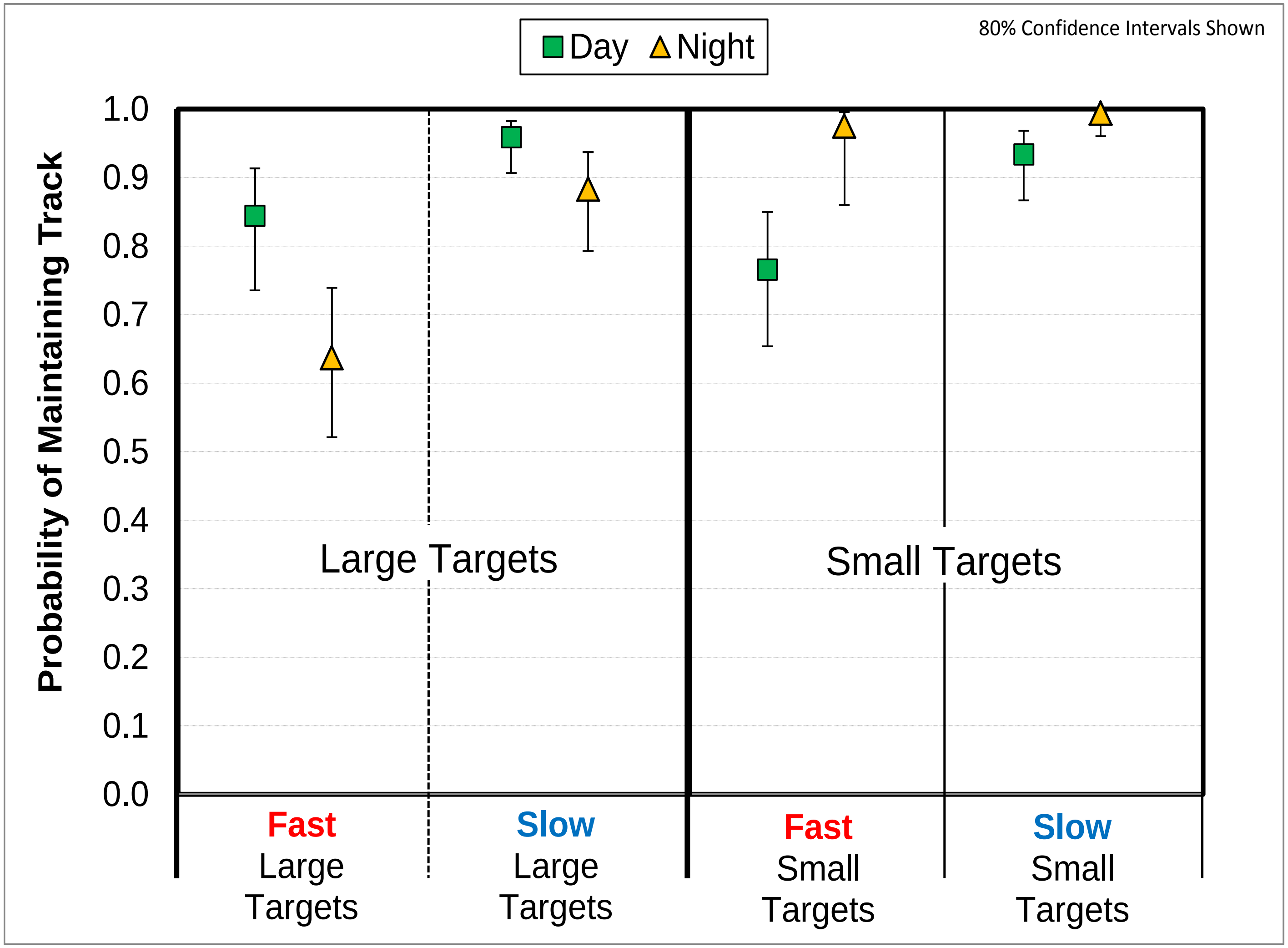
$$f(y) \sim Binomial(n, p)$$

$$\mu = np$$

$$\mu = \frac{\exp\left(\beta_0 + \sum_{i=1}^{k}\beta_i x_i + \sum_{i=1}^{k}\beta_i x_i^2 + \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\beta_{ij} x_i x_j + h.o.t\right)}{1 + \exp\left(\beta_0 + \sum_{i=1}^{k}\beta_i x_i + \sum_{i=1}^{k}\beta_i x_i^2 + \sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\beta_{ij} x_i x_j + h.o.t\right)}$$

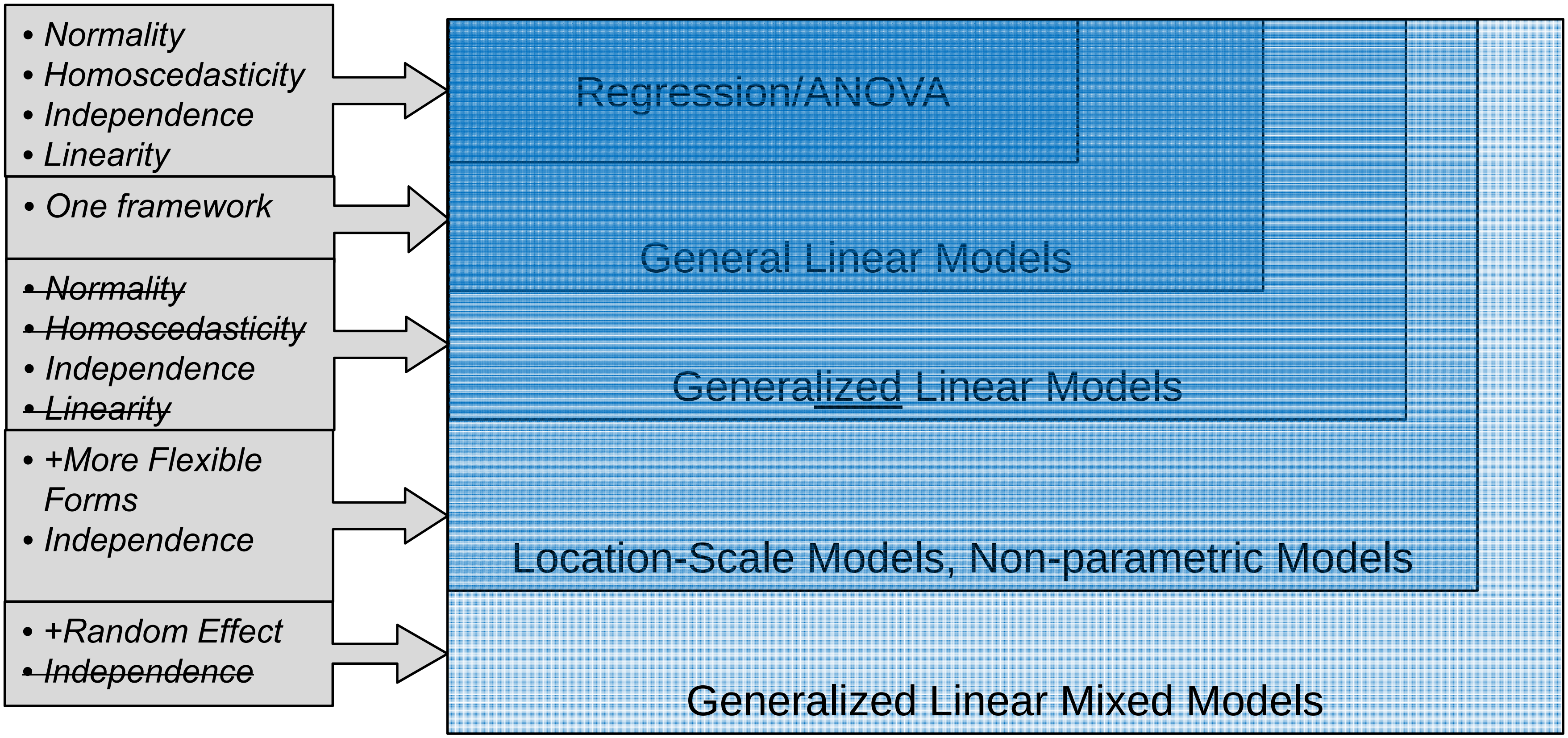
* In JMP: *Fit Model*→ *Generalized Linear Model*→ *Binomial Distribution*→ *Logit Link*

Summarizing Results



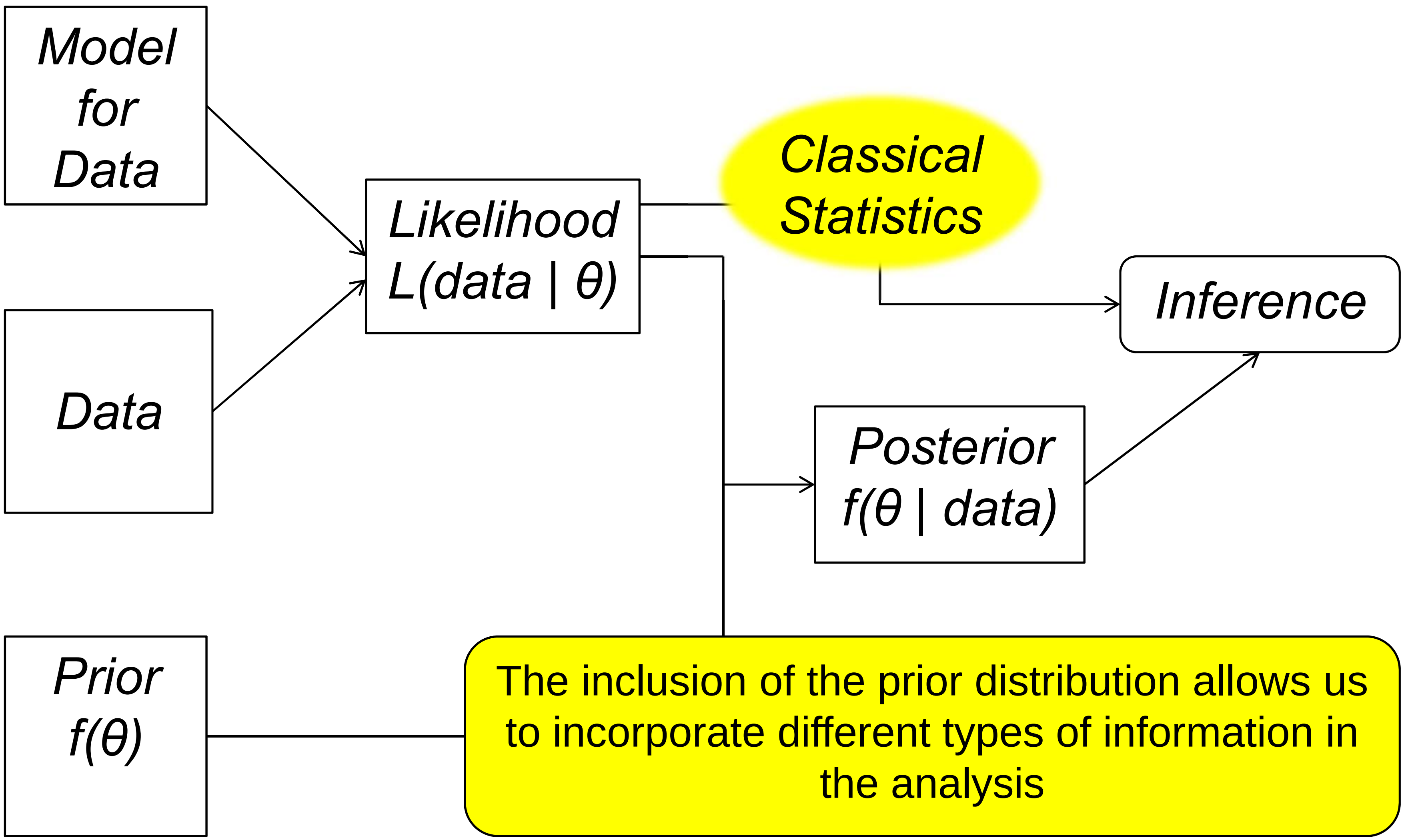
Parametric Statistical Model Hierarchy

- **There is a model for every situation!**



  – 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



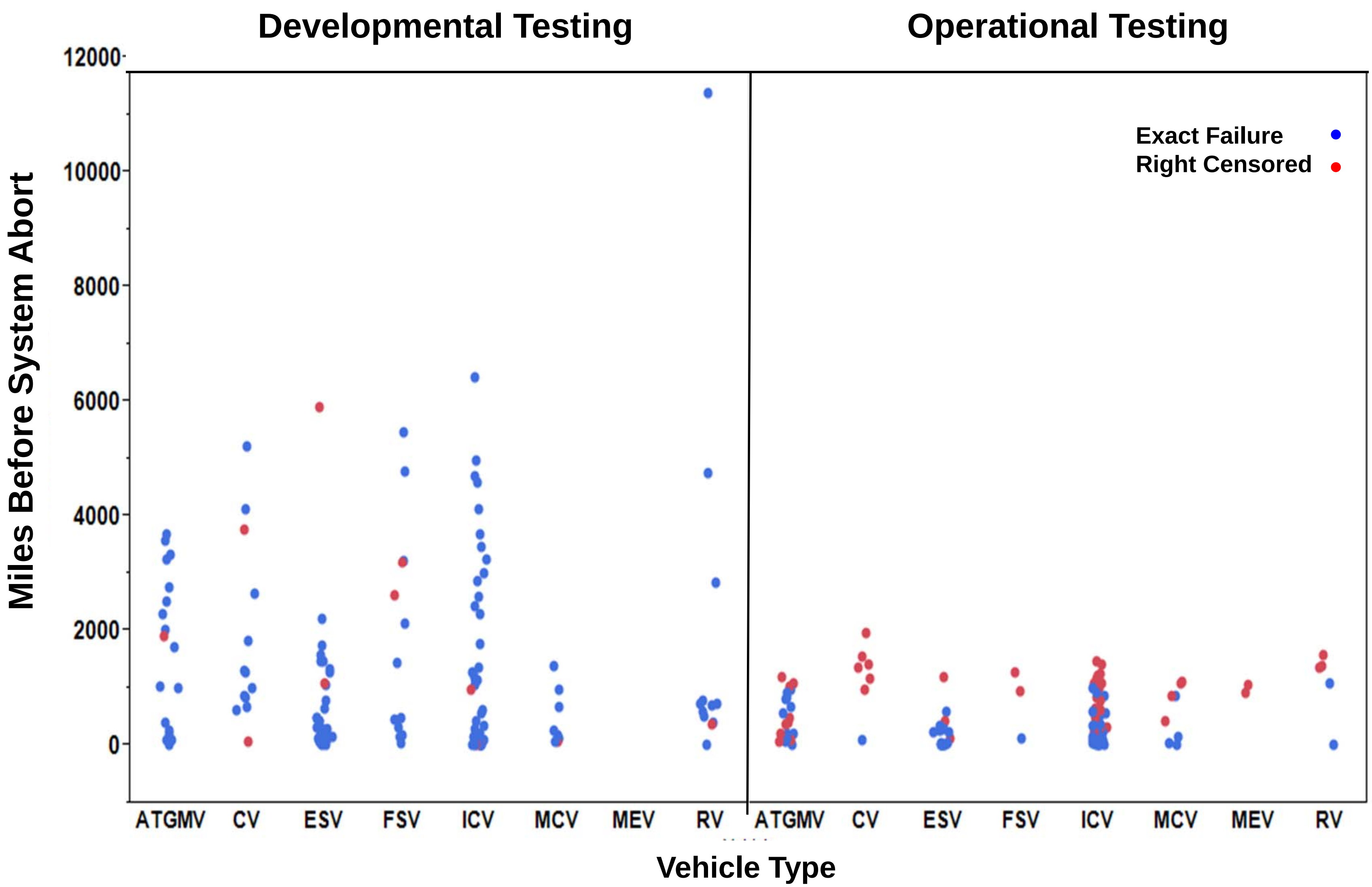
Motivating Example: Stryker Reliability Analysis

- **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:

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

- **Results do not leverage DT data or relationships between vehicles**

The Stryker Reliability Data Set



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

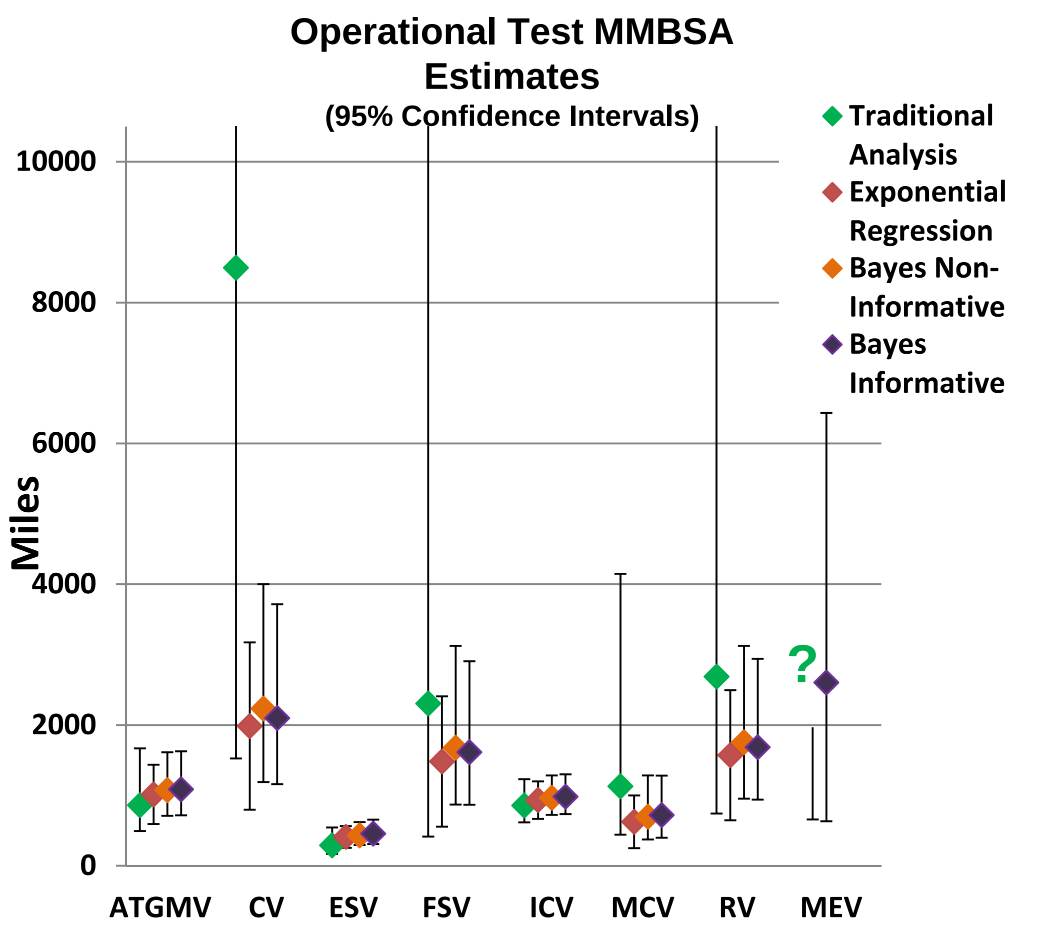
Bayesian Analysis Model:

$$t_{DT} \sim exp(\lambda_i) \quad t_{OT} \sim exp(\lambda_i/\eta)$$

$i = 1,2,\ldots,8$ (vehicle variants including MEV)

$$\lambda_i \sim gamma(a, b)$$
$$\eta \sim beta(1,1)$$
$$a \sim gamma(.001, .001)$$
$$b \sim gamma(.001, .001)$$

Stryker Reliability Results



- **Traditional Approach:**

$$MMBSA = \frac{Miles}{\# Failures}$$

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

- **Exponential Regression Approach & Bayesian Approaches**

$$MMBSA = f(TestPhase \quad Variant)$$

  - 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