A Bayesian Approach to Army Reliability Evaluations

Randy Griffiths and Andrew Thompson
Army Evaluation Command (AEC)
23 October 2014

U.S. Army Test and Evaluation Command

Introduction

- AEC, the Army's independent evaluator, is traditionally frequentist
- Independent Evaluator does not trust "expert opinion" in the Army acquisition process
- Many systems are upgrades or similar versions of legacy systems that ATEC has tested
- With recent budget constraints have come greater interest in utilizing all sources of data to a greater extent than previously done

Purpose

- We provide a method for developing prior distributions in the context of Army reliability evaluation
- The Bayesian paradigm provides a probabilistic approach to incorporating all sources of data into an evaluation while quantifying the uncertainty (value) of a given data source.

Bayesian vs Frequentist

Bradley Efron in *Bayesians, Frequentists, and Scientists*, ASA 2004 presidential address

- Bayesian
  - Trying to use all the information at its disposal to make the quickest possible progress
  - Tend to be aggressive and optimistic with their modeling assumptions
- Frequentist
  - Aims for universally acceptable conclusions that will stand up to adversarial scrutiny
  - More cautious and defensive

“The FDA for example doesn’t care about Pfizer’s prior opinion of how well its new drug will work, it wants objective proof. Pfizer, on the other hand may care very much about its own opinions in planning future drug development.”

Issues for Army Reliability Evaluations

- As independent evaluator we do not trust "expert opinion" (developer)
- Increasingly asked to leverage non-traditional reliability test data into the evaluation without compromising accuracy of inference
- DA PAM 73-1 states that reliability requirements must be demonstrated with "statistical confidence." This is currently interpreted in the context of frequentist demonstration testing.

General Approach to Prior Development

We want **conservative** and **defendable** priors

- Priors developed from past test data of similar systems
- Use MLE from previous test as a starting point for the mode of the prior distribution
- Base quantiles of prior distributions on confidence bounds of parameters from previous test data
- Account for uncertainty in system changes by increasing the variance of the prior

Missile System

- A new increment has been developed. The Program Manager needs to demonstrate to the Army "with confidence" that the reliability of the new missile is above the probability of success requirement.
- Only change to system was an upgraded payload
- ATEC tested and evaluated the previous increments for reliability
- ATEC developmental testing (DT) is essentially the same as operational testing (OT) for this system
- No changes were made to the system between tests

Missile System: Increment A data

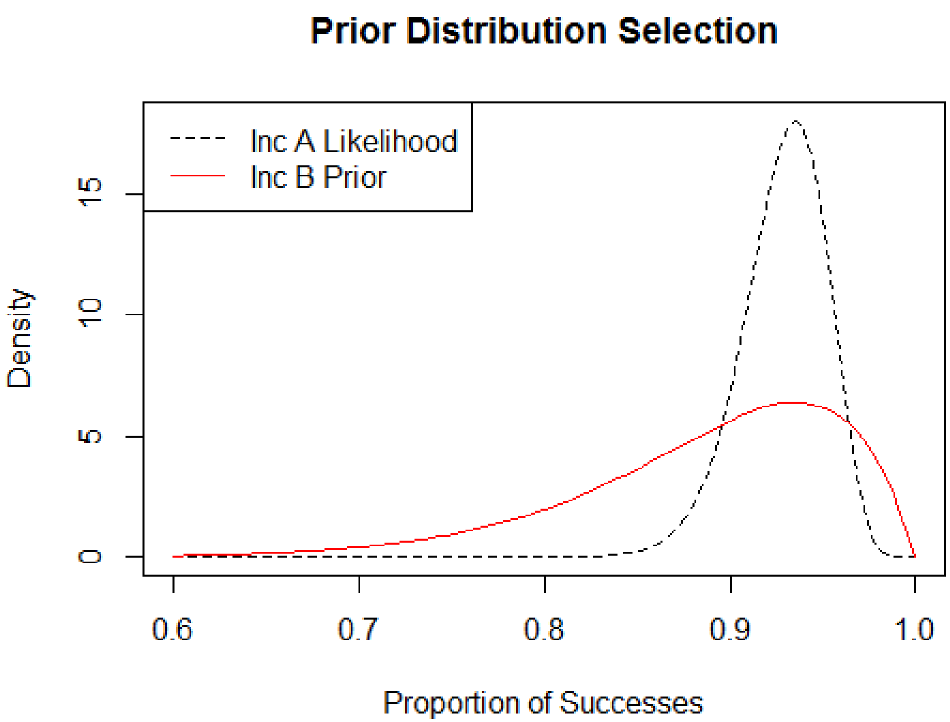
Recent Increment A OT demonstration test

- 124 trials; 116 successes, 8 failures
- MLE for the probability of success is 0.935
- ATEC used Clopper-Pearson confidence interval
  - Beta(α/2; failures, successes + 1) < p < Beta(1 - α/2; failures + 1, successes)

Missile System: Increment B Prior

- We believe the reliability of Increment B is about the same as the reliability of Increment A but we do have some uncertainty in that claim because ATEC has not yet tested the new payload
- Our prior for Increment B based on the MLE from Increment A and a Beta(116+1, 8+1) (confidence interval)
  - Note the variance of Beta(117, 9) is 0.00052
- Set the mode of the prior equal to the MLE from Increment A but increase the variance from the "likelihood" of Increment A (uncertainty in effect of new payload)
- Beta(16, 2.03) has mode 0.935 and variance 0.00525
  - Proportion of success is between 0.74 and 0.98 with 95% probability

Missile System: Increment B Prior

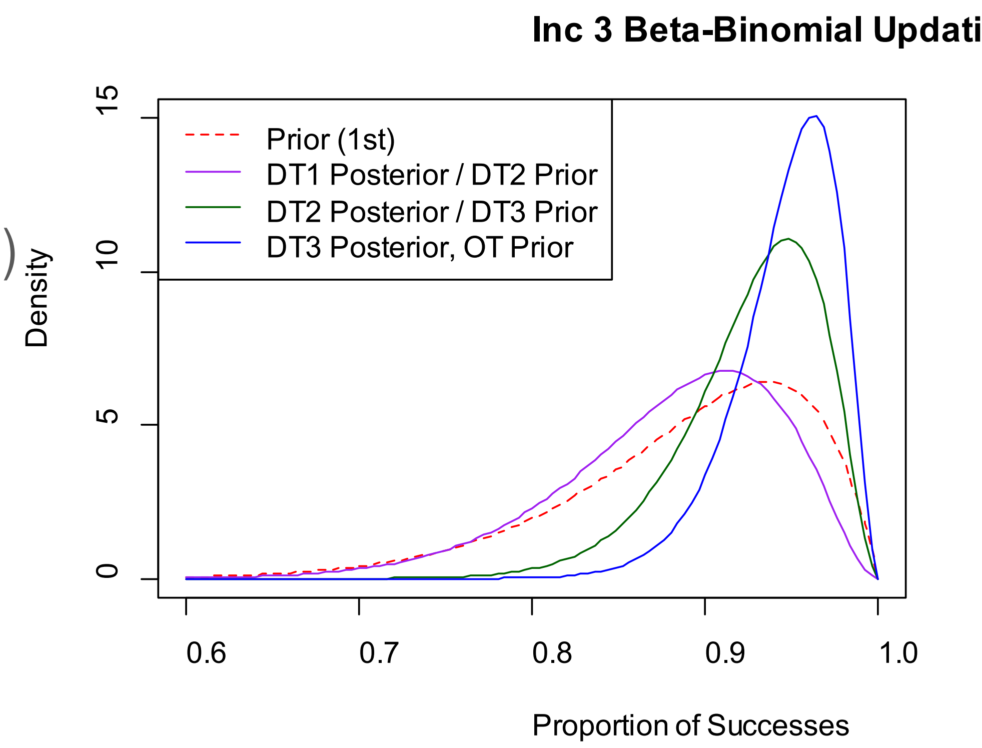


Useful R code for finding a Beta distribution based on a given mode and first parameter

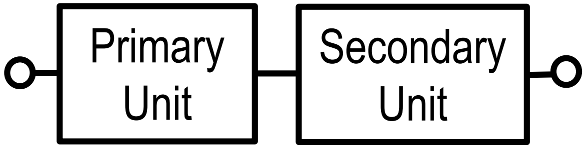
```
estBetaMode <- function(mode, alpha){
  beta <- (alpha*(1-mode) + 2*mode - 1)/mode
  mean <- alpha/(alpha+beta)
  var <- alpha*beta/((alpha + beta)^2*(alpha + beta + 1))
  return(params = list(beta = beta, mean = mean, variance = var))
}
```

Missile System: Increment B Test Data

- DT1 had 6 successes, 1 failure
  - Posterior is Beta(16+6, 2.03+1)
- DT2 had 16 successes, 0 failures
  - Posterior is Beta(16+6+16, 2.03+1)
- DT3 had 15 successes, 0 failures
- 0 failure frequentist demonstration test is 20 rounds (1 failure Bayesian assurance test with our OT prior)



Counter IED System



Let X be the failure count in t hours

$X \sim Poisson(t*\lambda_{sys})$ $\quad \lambda_{PU} \sim Gamma(a,b)$

$\lambda_{sys} = \lambda_{PU} + \lambda_{SU}$ $\quad \lambda_{SU} \sim Gamma(c,d)$

- The Primary Unit has been upgraded from a legacy system. Army is interested in buying small number of upgraded systems. Limited (0 allowable failures) final operational test planned with system mounted on vehicle and used by soldiers in mission scenarios.
- No significant reliability testing done on legacy system
- Each subsystem was tested independently in chamber testing to stress components for typical failure modes
- There is minimal interaction between operators and system, but system will be mounted on a vehicle

Counter IED: Data

Chamber test data available on upgraded Primary Unit (PU) and Secondary Unit (SU) subsystems

- PU test yielded 3 failures in 6849 hours in chamber
  - λ_{PU} MLE = 0.00043 with 80% CI = (0.00016, 0.00078)
- SU test yielded 1 failure in 7055 hours in chamber
  - λ_{SU} MLE = 0.00014 with 80% CI = (0.000014, 0.000551)

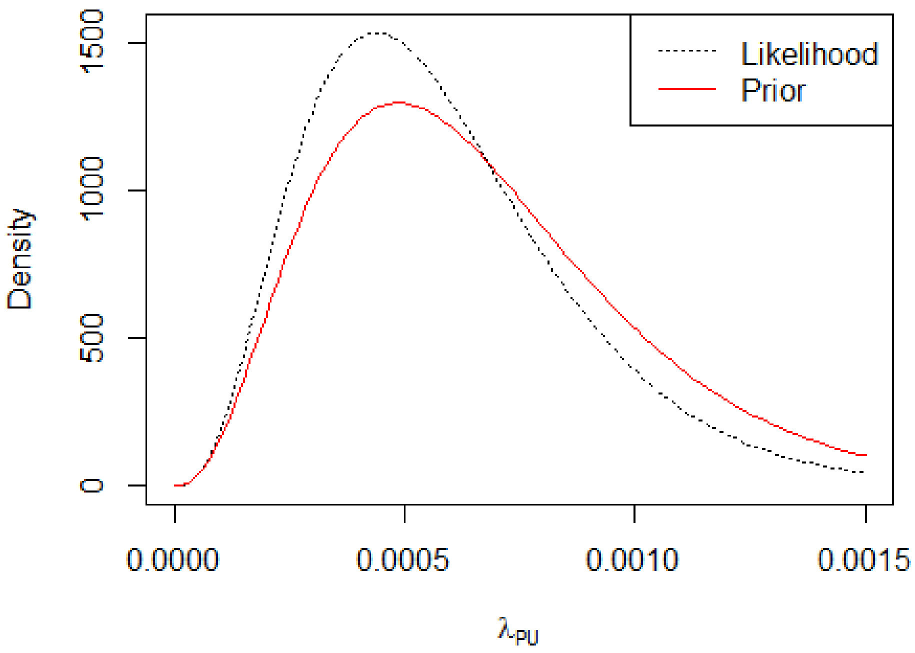
Counter IED: Priors

Approach

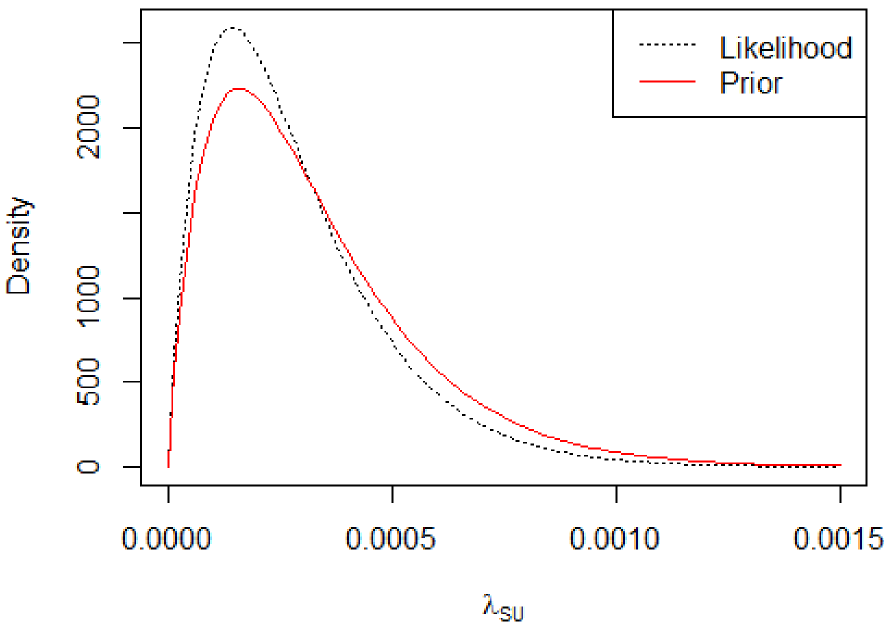
- We believe the chamber test was fairly robust for providing representative estimates for operational reliability, but we (conservatively) expect a small drop in reliability and are less certain of that estimate due to change in test conditions
- AEC typically plans on a 10% degradation in system MTBF going from DT to OT. Set mode equal to 10% degradation in MTBF from chamber test and set 0.90 quantile equal to upper 80% confidence limit from previous test (plus 5% to account for uncertainty in the effect of changing test conditions)
- Useful R function: gamma.parms.from.quantiles (http://www.medicine.mcgill.ca/epidemiology/joseph/pbelisle/GammaParmsFromQuantiles.html)

Counter IED: Priors

Primary Unit Prior Selection



Secondary Unit Prior Selection



Counter IED: Posteriors

- Posterior distributions for λs will be of the form Gamma(a + failures, (t*b + 1)/b)
- Since PU and SU are independent, Monte Carlo simulation can provide a posterior distribution for λ_{sys}
- These priors yield a 1 failure test if failure is on SU, or a 2 failure test if both failures on the PU.
- Based on chamber testing estimates, the probability of having a 1280 hour test with 0 failures is 47%; the probability of having 1 or fewer failures is 83%

Conclusions

- Objective data often available for prior development
- A conservative prior can be established based on MLE and sampling distribution from past test data
- With high reliability prior data, can provide a similar quality evaluation with a decrease in either test scope or Producer's Risk
- Certain commodity areas work better than others for this approach
- Still must have a test long enough to verify requirement in OT with "statistical confidence"

Room For Improvement

- Accounting for fixes to system between test events
- Accounting for changes in reliability due to environmental changes
- Developing guidelines to protect against inappropriate prior data use (i.e. systems too dissimilar, data unverifiable)