Abstract

We use a Bayesian hierarchical model to assess the reliability of the Joint Light Tactical Vehicle (JLTV), which is a family of vehicles. The proposed model effectively combines information across three phases of testing and across common vehicle components. The analysis yields estimates of failure rates for specific failure modes and vehicles as well as an overall estimate of the failure rate for the family of vehicles. We are also able to obtain estimates of how well vehicle modifications between test phases improve failure rates. In addition to using all data to improve on current assessments of reliability and reliability growth, we illustrate how to leverage the information learned from the three phases to determine appropriate specifications for subsequent testing that will demonstrate if the reliability meets a given reliability threshold.

1 Introduction

2 Data

3 Methodology

3.1 Modeling Reliability

A standard reliability analysis employed by the Department of Defense (DoD) test community considers each test phase independently and uses the exponential distribution to model the miles between failure [?]. The traditional analysis is overly simplistic, relies on correct modeling assumptions and ignores valuable information learned about the individual vehicles and their failure modes.

In considering the following alternative approach we will begin by introducing a hierarchical model structure that can use data across all test phases and incorporate known similarities between vehicle failure modes. Next, we will look at a number different modeling and distributional assumptions. We discuss different model diagnostics to consider when choosing a final model. Then we illustrate how this decision can effect our assessments of system reliability. This will lead us into the next section on how this modeling can be used for assurance test planning.

3.1.1 Exponential Model

In Test Phase 1, we assume the vehicle miles at failure, y, follow an exponential distribution with a failure rate parameter λ_{ij}. Introducing notation,

$$y_{ijk} \mid \lambda_{ij} \sim Exp(\lambda_{ij}), \quad i = 1, 2, .., v \quad j = 1, 2, .., s \quad k = 1, 2, .., n_{ij} \tag{1}$$

where y_{ijk} are the miles between failure for vehicle i failure mode j, v is the number of vehicles, s is the number of failure modes, and n_{ij} are the number of failures of vehicle i failure mode j. The number of failure modes is assumed fixed and known *a priori*.

The prior distribution on the exponential failure rate parameter, λ_{ij}, depends on whether failure mode j is considered to be common across vehicles or related but not identical. For the related failure modes, we place a gamma prior distribution on the collection of λ_{ij}; in other words, we assume each vehicle has a distinct failure rate in failure mode j but they arise from a common gamma distribution.

If failure mode j is considered common across vehicles, the collection of failure rates is collapsed to a single parameter, $\lambda_{ij} = \lambda_j$. As with the related failure modes, a gamma prior distribution is placed on the single failure rate. The prior distributions are independent across failure mode, and can potentially have different hyperparameters.

The Phase 1 analysis yields an estimate of the failure rate λ_{ij} for each of the vehicles for failure modes that are related. We are assuming the vehicles are conditionally independent, therefore the failure rate estimate for the family of vehicles for such failure modes can be found by $\sum_i \lambda_{ij}$. For failure modes that are common across vehicles the Phase 1 analysis yields a λ_j, which is the failure rate for the family of vehicles. Under the exponential modeling assumption the overall failure rate across all failure modes can be found by $\sum_j \lambda_j$.

3.1.2 Fix Effectiveness

After the first CAP, Test Phase 2 begins with the repaired vehicles. To capture these revisions, the PM2 reliability growth model [**?**] is often used. This model explicitly captures testing phases, choices about which failure modes to correct, and the potential of not completely eliminating a failure upon repair. One of the downsides of PM2 is that many parameters of potential interest, such as the Fix Effectiveness Factor (FEF), which measures how much repairs improve failure rates, are typically fixed. A common value for FEF is 0.70. We follow the premise of this type of model, but allow a more flexible and data-driven result that is less dependent on hard-coded assumptions.

One normal assumption used in reliability growth modeling is non-decreasing failure rates; that is either the fixes were effective or had no effect, but did not degrade the family of vehicles. This should generally be the case, but because we are dealing with complex systems we will sometimes see decreases in failure rates after adjustments are made. Therefore for the Phase 2 data we write the rate parameters as a function of the rate parameters found in Phase 1. In particular, we define $\lambda_{ij}^{P2} = (\rho_j)\lambda_{ij}^{P1}$ where ρ_j represents the between phase change in failure mode j. Given this definition of λ_{ij}^{P2}, we again model the miles to failure for a given vehicle and failure mode using the exponential distribution. We assume the prior distribution for the ρ_j is a gamma distribution. If ρ_j is less than one, this represents an improvement in reliability. After Phase 2 we can look again at failure rates across failure modes and vehicles and obtain an overall estimate of the rate for the family of vehicles. The analysis of Test Phase 3 follows the same pattern as that shown in Phase 2. At the end of Phase 3, we can look at failure rates across failure m odes and vehicles and obtain an overall estimate of the rate for the family of vehicles. Future tests will be planned based on the inferences of Phase 3.

3.1.3 Weibull Model

The exponential model is by far the most common parametric distribution used in reliability modeling because of its desirable math mathematical properties and simple interpretations. Despite its common uses the assumption of a constant failure rate over time is rarely justifiable. It has been well documented (Statistics, Testing, and Defense Acquisition: Background Papers chapter - 2 http://www.nap.edu/catalog/9655.html) the issues that can arise when this assumption is violated. We will now consider the same hierarchical model structure while using the Weibull distribution for each miles between failure observation.

$$y_{ijk} \mid \lambda_{ij}\kappa_i \sim Weibull(\lambda_{ij}), \quad i = 1, 2, .., v \quad j = 1, 2, .., s \quad k = 1, 2, .., n_{ij} \tag{2}$$

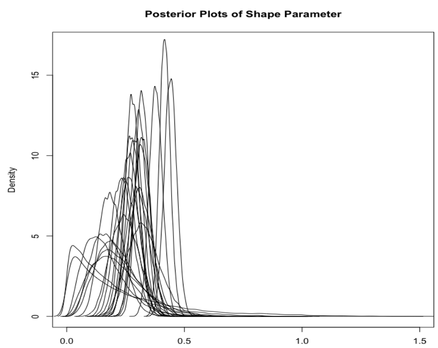
The Weibull distribution is a more flexible model with both a rate parameter λ_{ij} and a shape parameter κ_j. The exponential is a special case of the Weibull, when $\kappa = 1$.
It should be noted that we are currently indexing both ρ and κ, only by j and not i, in other words we are assuming a single shape and rate change parameter for each failure mode across vehicles. All of these parameters could also be indexed by i and modeled hierarchical or with some combination of both. This is something we will omit for this paper but hope to explore in the future. One special case we will discuss is when a single shape parameter can be assumed across all failure modes. When this is justifiable it can greatly simplify test planning calculations, as we will show in the assurance testing section.

3.1.4 JLTV Model Selection

So far we have discussed a number of different models and the assumptions that accompany them. We will now discuss the process of model selection. The decision of which model to use can have a drastic impact on reliability assessment and future test planning. We will use the JLTV dataset to demonstrate the process.

The first model selection question we will explore is the parametric form, exponential versus Weibull. In statistical modeling we often face the trade-off between fit and interpretability, and this situation is no different. Because the Weibull is a more flexible model, it will always, in a sense, fit the data better, but this comes at a price. The exponential's convenient form makes both computation and interpretation straightforward. When the Weibull's shape parameter is introduced this advantage is lost. Thus, when the overall fit is close to the same between the exponential and the Weibull we will default to using the exponential.

The first check we used to decide between the exponential and Weibull models is to fit the Weibull model and look at the posterior distributions of the shape parameters and determine if one is a reasonable value. Looking at the plot below from the JLTV data, we see that for the 26 components it appears that the value of one falls in the extreme tails of the distributions. This is our first clue that the exponential model will not be a good fit to this data.



This is not a surprising result, because the exponential model assumes constant failure rate. For most vehicle components we would expect the failure rate to increase as the distance driven grows larger, and this corresponds with a shape parameter between zero and one. On the other hand, when the shape parameter is larger than 1, the failure rate will decrease with time/distance.

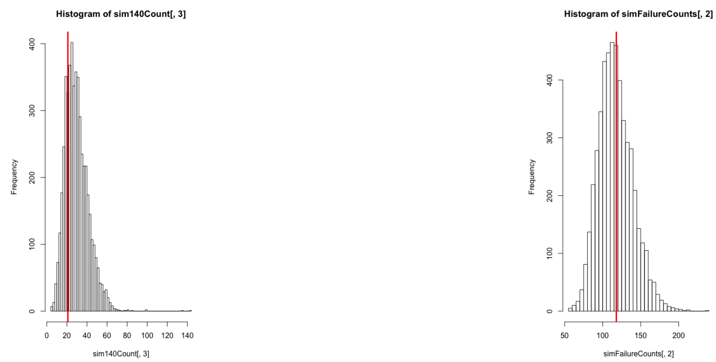
Now a more formal diagnostic tool used for model selection is the Deviance information criterion (DIC). This is a popular method for comparing the goodness of fit of multiple models. The DIC method gives a slight penalty for larger numbers of parameters. The DIC value is a unitless measure with lower values indicating a better fit to the data. In the table below we show the DIC results for the different distribution and structure combinations we considered with the JLTV dataset.

Goodness of Fit

Distribution	Structure	DIC
Exponential	Single Rate	23258
	Hierarchical Rate	23022
Weibull	Single Rate	18750
	Hierarchical Rate	**18556**
	Hierarchical Rate (One shape)	18677

The single rate structure represents the non-hierarchical model, using one rate parameter all 8 vehicles for a given failure mode. The hierarchical rate structure uses the common gamma distribution for the failure modes that are common across vehicles. The final entry in the table is the model that uses one shape parameter across all 26 failure modes.

While we have shown that the hierarchical Weibull model fits much better than the exponential, these test still do not tell us that this model is a good fit for the data. The last method we will present is called posterior predictive



(a) Plot 1 (b) Plot 2

Figure 1: Posterior Predictive Plots

checking ***reference***. Here we are given important features of the data. In the JLTV case we were interested in the total failure counts for each phase and the number of time the miles between failures was less than 140. We then used the final model to simulate 5,000 new datasets. We then plot histograms for each of the 8 vehicles for all 3 phases. In Figure 1 are examples of the histograms produced, with the line showing where the value of the true dataset fell. For this method we dont expect all of the true values to fall in the center of the distribution. Values in the tales are to be expected in a random processes like this one. We will only be concerned if we find true values falling in the extreme tails or if we see a common bias to the high or low side of the distributions.

3.1.5 JLTV Reliability Results

Results plots
Exponential vs. Weibull interpretation

3.2 Assurance Test Planning

For all statisticical test planning we are interested in minimizing error rates while using the resoruces avalible as efficently as possible. In the DoD acquisition process the error rates will be refrenced in terms of risk. Consumer risk being the probability of a unacceptable product passing a given test.

Producer risk being the probability of an acceptable product failing a given test.

3.2.1 Traditional Approach

Traditional methods do not use any past data for determining the testing proceedure. The only rely on distributional assumptions and asymptoic results.

3.2.2 Bayesian Approach

With Bayesian statistical approach we can incorperate the supplementary data from previous testing phases with the hopes of minimizing the resorces needed for testing.

R(t) = reliability at time t (miles in our case)

t_{*c} : time of interest to consumer
t_{*p} : time of interest to producer

- Consumer Risk : Prob($R(t_{*c}) \leq \pi_c$| Test is passed)
- Producer Risk : Prob($R(t_{*p}) \geq \pi_p$| Test is failed)

π_c : minimum reliability acceptable to the consumer at t_{*c}
π_p : minimum reliability goal of the producer at t_{*p}

We would like to get reliability into a inequality in terms of something we have a distribution for so we can evaluate the conditional probability statements.

$R(t_{*c}) \leq \pi_c \Rightarrow$ average number of system failures in 80 miles is greater than 2
$R(t_{*p}) \geq \pi_p \Rightarrow$ average number of system failures in 140 miles is less than 2

3.2.3 Poisson Process

Reliability: $R(t) = 1 - F(t)$
Hazard or failure rate function: $\lambda(t) = \frac{f(t)}{R(t)}$
Series system reliability: $R_S(t) = \prod_{i=1}^{N} R_i(t)$

Result:

$$
\begin{aligned}
R_S(t) &= \prod_{i=1}^{N} R_i(t) \\
\frac{dR_S(t)}{dt} &= \frac{d}{dt} \prod_{i=1}^{N} R_i(t) \ \text{ (derivative of both sides)} \\
\frac{dR_S(t)}{dt} &= \sum_{i=1}^{N} \left[\frac{dR_i(t)}{dt} \frac{dR_S(t)}{dR_i(t)} \right] \ \text{ (product rule)} \\
\frac{-\frac{dR_S(t)}{dt}}{R_S(t)} &= \sum_{i=1}^{N} \left[\frac{-\frac{d}{dt} R_i(t)}{dR_i(t)} \right] \ \text{ (divide both sides by } -R_S(t)\text{)} \\
\lambda_S(t) &= \sum_{i=1}^{N} \lambda_i(t) \ \text{ (because } \lambda(t) = \frac{f(t)}{R(t)} = \frac{-\frac{dR(t)}{dt}}{R(t)} \text{)}
\end{aligned}
$$

3.2.4 Non-Homogeneous Poisson Process

Definitions:
Number of failures before time t: $N(t) \sim \text{Poisson}(m(t))$
Mean function: $m(t) = \int_0^t \lambda(s) ds$ (represents the expected number of failures before time t)
Weibull(γ, β) failure rate: $\lambda(t) = \gamma \beta t^{\beta - 1}$

Result:
If we assume a constant shape parameter β then,

$$
\begin{aligned}
\lambda_S(t) &= \sum_{i=1}^{N} \lambda_i(t) \\
&= \sum_{i=1}^{N} \gamma_i \beta t^{\beta-1} \\
&= \beta t^{\beta-1} \sum_{i=1}^{N} \gamma_i
\end{aligned}
$$

Then Solving for the mean function of the system,

$$
\begin{aligned}
m_S(t) &= \int_0^t \lambda_S(s)ds \\
&= \int_0^t \beta t^{\beta-1} \sum_{i=1}^{N} \gamma_i \\
&= \sum_{i=1}^{N} \gamma_i \int_0^t \beta t^{\beta-1} \\
&= t^{\beta} \sum_{i=1}^{N} \gamma_i
\end{aligned}
$$

This gives us: $N(t) \sim \text{Poisson}(t^{\beta} \sum_{i=1}^{N} \gamma_i)$

3.2.5 Exponential Case

If we model the reliability of all 26 components in the system as $Y_{ij} \sim$ exponential (γ_i) random variables, this leads to the system reliability being the minimum or $Y_{system} \sim$ exponential $(\sum_{n=1}^{26} \gamma_i)$

- Looking at consumer risk first:
want to find Prob($R(t_{*c}) \leq \pi_c |$ Test is passed)
Let $\sum_{n=1}^{26} \gamma_i = \lambda_S$
The expected number of failures for the system per mile is $\mathbf{E}(Y_{system}) = \lambda_S$

This leads to our consumer risk probability constraint as follows. Given the test is passed, the consumer would like the probability of the expected number of failures in 80 miles being greater than 2 to be smaller than α.

Prob($\lambda_S \cdot (80) \geq 2$ | Test is passed) $\leq \alpha$

Let W be the number of failures during the test and $W \leq c \Rightarrow$ Test is passed and $W > c \Rightarrow$ Test is failed.

Because failures are exponential $W \sim Poisson(\lambda_S T)$ where T is the number of miles run during the test.

$$
\begin{aligned}
P(\lambda_S \geq 2/80 \mid W \leq c) &= \int_{1/40}^{\infty} P(\lambda_S \,|W < c) d\lambda_S \\
&= \int_{1/40}^{\infty} \frac{f(W < c\,|\lambda_S) p(\lambda_S)}{f(W < c)} d\lambda_S \\
&= \int_{1/40}^{\infty} \frac{f(W < c\,|\lambda_S) p(\lambda_S)}{\int_0^{\infty} f(W < c\,|\lambda_S) p(\lambda_S) d\lambda_S} d\lambda_S \\
&= \frac{\int_{1/40}^{\infty} [\sum_{W=0}^{c} \frac{(\lambda_S T)^W exp(-\lambda T)}{W!}] p(\lambda_S) d\lambda_S}{\int_0^{\infty} [\sum_{W=0}^{c} \frac{(\lambda T)^W exp(-\lambda_S T)}{W!}] p(\lambda_S) d\lambda_S}
\end{aligned}
$$

For simplicity we fix c to be zero (The number of failures needed to pass the test) and take N posterior draws $\lambda_S^{(j)}$

$$
\begin{aligned}
P(\lambda_S \geq 1/40 \mid W = 0) &= \frac{\int_{1/40}^{\infty} exp(-\lambda_S T) p(\lambda_S) d\lambda_S}{\int_0^{\infty} exp(-\lambda_S T) p(\lambda_S) d\lambda_S} \\
&\approx \frac{\sum_{j=1}^{N} exp(-\lambda_S^{(j)} T) I(\lambda_S^{(j)} \geq \frac{1}{40})}{\sum_{j=1}^{N} exp(-\lambda_S^{(j)} T)}
\end{aligned}
$$

Using the same technique for producer risk we get the following. The producer would like, given the test is failed, the probability of the expected number of failures in 140 miles being less than 2 to be smaller than β.

$$
\begin{aligned}
P(\lambda_S \leq 1/70 \mid W > 0) &= \frac{\int_0^{1/70} [1 - exp(-\lambda_S T)] p(\lambda_S) d\lambda_S}{\int_0^{\infty} [1 - exp(-\lambda_S T)] p(\lambda_S) d\lambda_S} \\
&\approx \frac{\sum_{j=1}^{N} [1 - exp(-\lambda_S^{(j)} T)] I(\lambda_S^{(j)} \leq \frac{1}{70})}{\sum_{j=1}^{N} [1 - exp(-\lambda_S^{(j)} T)]}
\end{aligned}
$$

By constraining these two probabilities to acceptable risk levels we can solve for the smallest T that satisfies both.

Consumer Risk : $P(\lambda_S \geq 2/80 \,|W = 0) \leq \alpha$
Producer Risk : $P(\lambda_S \leq 2/140 \,|W > 0) \leq \beta$

3.2.6 Weibull Case

Now if we model the reliability of all 26 components in the system as $Y_{ij} \sim$ Weibull (γ_i, β) random variables we are able to use the the non-homogeneous result from part two to build our assurance test. This follows the same process as the exponential test plan with all mile variables adjusted by the β exponent and the N posterior draws will use both $\lambda_S^{(j)}$ and $\beta^{(j)}$.

Consumer risk:

$$P(80^{\beta}(\lambda_S) \geq 2 \mid W = 0) \approx \frac{\sum_{j=1}^{N} exp(-\lambda_S^{(j)} T^{\beta^{(j)}}) I(80^{\beta^{(j)}}(\lambda_S^{(j)}) \geq 2)}{\sum_{j=1}^{N} exp(-\lambda_S^{(j)} T^{\beta^{(j)}})}$$

Producer risk:

$$P(140^{\beta}(\lambda_S) \leq 2 \mid W > 0) \approx \frac{\sum_{j=1}^{N} [1 - exp(-\lambda_S^{(j)} T^{\beta^{(j)}})] I(140^{\beta^{(j)}}(\lambda_S^{(j)}) \leq 2)}{\sum_{j=1}^{N} [1 - exp(-\lambda_S^{(j)} T^{\beta^{(j)}})]}$$

Important note: For the exponential test plan, the producer and consumer risks were in terms of expected number of failures in a certain number of miles, say t. Because the exponential has a constant hazard rate this can be considered the expected number of failures for a given number of miles regardless of how many miles have been driven prior. On the other hand the Weibull does not have a constant hazard rate. For the Weibull test plan presented here the producer and consumer risk statements are now in terms of expected number of failures in the first t miles driven.

3.2.7 Results