Comparing Methods for Assessing Reliability Uncertainty Based on Pass/Fail Data Collected Over Time

Abes, Jeff I.
Hamada, Michael Scott
Hills, Charles Robert

To be published in: Quality Engineering

DOI to publisher's version: 10.1080/08982112.2017.1417600

Quality Engineering

ISSN: 0898-2112 (Print) 1532-4222 (Online) Journal homepage: http://www.tandfonline.com/loi/lqen20

Comparing Methods for Assessing Reliability Uncertainty Based on Pass/Fail Data Collected Over Time

Jeff I. Abes, Michael S. Hamada & Charles R. Hills

Comparing Methods for Assessing Reliability Uncertainty Based on Pass/Fail Data Collected Over Time

Jeff I. Abes, Michael S. Hamada, and Charles R. Hills

Los Alamos National Laboratory

Los Alamos, NM 87545

110117 1800

Abstract

In this article, we compare statistical methods for analyzing pass/fail data collected over time; some methods are traditional and one (the RADAR or Rationale for Assessing Degradation Arriving at Random) was recently developed. These methods are used to provide uncertainty bounds on reliability. We make observations about the methods' assumptions and properties. We illustrate the differences between two traditional methods, logistic regression and Weibull failure time analysis, and the RADAR method using a numerical example.

Key Words: Bayesian inference, left-censored, logistic regression, lognormal distribution, normal distribution, pass/fail data, probit regression, RADAR method, right-censored, Weibull distribution, zero-inflated failure time distribution.

Introduction

Various weapons organizations collect pass/fail data over time. For example, missiles, one-shot devices, are tested and they either perform successfully or not. Such data are used to assess reliability and there is a need to quantify the uncertainty on reliability when limited amounts of data are collected. The RADAR (Rationale for Assessing Degradation Arriving at Random) method (Vander Wiel et al., 2011) is one method that provides uncertainty bounds on reliability.

Recently, we were asked to conduct a survey of alternative methods to the RADAR method for analyzing pass/fail data over time. That is, this article compares various methods for analyzing pass/fail data collected over time including the RADAR method.

The survey of industry and academia identified two traditional methods for analyzing such data. One method is logistic regression with time as the predictor or independent variable that NASA uses (Dezfuli et al., 2009); SRFYDO, a Los Alamos National Laboratory (LANL) software product used by the DoD, uses probit (related to logistic) regression in reliability assessments (Anderson-Cook et al., 2012). The second method assumes a failure time distribution such as the Weibull distribution (Olwell and Sorell, 2001) and treats the pass/fail data as right-/left-censored data. In this article, we will compare these methods in terms of their assumptions and properties. We also compare the Weibull failure time analysis, logistic regression, and the RADAR method using a numerical example.

Methods

Pass/fail data are collected over time. That is, one-shot devices are tested and they either work, i.e., pass or they fail. At time t, suppose that n_t pass/fail data are collected consisting of x_t fails. Here we consider time in years, so that $t = 0, 1, \ldots, U$ for time 0, i.e., at birth, and U subsequent years. Note that in some years data collection can be skipped but we do not consider that scenario in the article. Next, we present the three methods in turn.

Logistic Regression and Other Related Models

Hosmer and Lemeshow (2013) provides an entire book on logistic regression. Dezfuli et al. (2009) presents the use of logistic regression for a reliability analysis. In logistic regression, x_t fails in n_t trials (i.e., tests) is assumed to follow a binomial distribution with probability of a fail at time $t \geq 0$, $Prob(Fail\ at\ time\ t)$, denoted by p_t and expressed as:

$$\text{logit}(p_t) = \log \frac{p_t}{1 - p_t} = \beta_0 + \beta_1 t. \tag{1}$$

Note we will use the shorthand $Prob_t(Fail)$ to indicate indexing at time t as well as $Prob_t(Pass)$. Using Equation (1), $Prob_t(Fail)$ can be expressed as

$$Prob_t(Fail) = p_t = \frac{\exp(\beta_0 + \beta_1 t)}{1 + \exp(\beta_0 + \beta_1 t)}, \tag{2}$$

where Equation (2) is the cumulative distribution function (cdf) of a random variable that has a logistic distribution for $t \geq 0$. Now $Prob_t(Pass) = 1 - p_t$ is the reliability at time t denoted by $R(t)$, where $R(t)$ can be expressed as

$$R(t) = 1 - \frac{\exp(\beta_0 + \beta_1 t)}{1 + \exp(\beta_0 + \beta_1 t)}. \tag{3}$$

Anderson-Cook et al. (2012) uses probit regression to analyze pass/fail data collected over time. Instead of the logistic distribution given in Equation (3), probit regression uses the normal distribution so that

$$Prob_t(Fail) = p_t = \Phi\left(\frac{t - \mu}{\sigma}\right),$$

where $\Phi(\cdot)$ is the standard normal cdf. That is, probit regression depends on a normal cdf with mean μ and standard deviation $\sigma > 0$. The related expression to Equation (1) for probit regression is

$$\text{probit}(p_t) = \Phi^{-1}(p_t) = \beta_0 + \beta_1 t = -\frac{\mu}{\sigma} + \frac{1}{\sigma} t,$$

where $\Phi^{-1}(\cdot)$ is the inverse of the normal cdf. The corresponding reliability function is

$$R(t) = Prob_t(Pass) = 1 - Prob_t(Fail) = 1 - p_t = 1 - \Phi\left(\frac{t - \mu}{\sigma}\right). \tag{4}$$

Weibull Failure Time Analysis

Olwell and Sorell (2001) assumes a Weibull failure time distribution and treats the pass and fail data as right-censored and left-censored data, respectively. If failure time T follows a Weibull distribution with scale ψ and shape β parameters, the reliability $R(t) = Prob(T > t)$ for a Weibull distribution can be expressed as

$$R(t) = \exp(-(t/\psi)^{\beta}). \tag{5}$$

A fail is treated as a left-censored observation, where $T \leq t$; that is, the failure occurred before t. Also a pass is treated as a right-censored observation, where $T > t$; that is, the failure has not occurred yet and will occur sometime after t. Consequently,

$$Prob_t(Pass) = Prob(T > t) = \exp(-(t/\psi)^{\beta})$$

and

$$Prob_t(Fail) = Prob(T \leq t) = 1 - \exp(-(t/\psi)^{\beta}), \tag{6}$$

where Equation (6) is the Weibull cdf. In summary, this method is a traditional failure time analysis assuming the Weibull distribution with left- and right-censored data.

Bayesian Inference

Before describing the RADAR method which relies on Bayesian inference, here we briefly describe what Bayesian inference is. Hamada et al. (2008) applies Bayesian inference to reliability assessment for various data types and models. Bayesian inference, i.e., analysis, combines prior information with observed data to produce a posterior distribution for the parameters $\boldsymbol{\theta}$ using Bayes' theorem that takes the form:

$$\pi(\boldsymbol{\theta} \,|\, \mathbf{y}) \;=\; \frac{L(\mathbf{y} \,|\, \boldsymbol{\theta})\; \pi(\boldsymbol{\theta})}{\int_{\Theta} \, L(\mathbf{y} \,|\, \boldsymbol{\theta})\, \pi(\boldsymbol{\theta})\, d\boldsymbol{\theta}}, \tag{7}$$

where $\boldsymbol{\Theta}$ denotes the range of values for the parameters $\boldsymbol{\theta}$. (In logistic and probit regression, $\boldsymbol{\theta} = (\beta_0, \beta_1)$ and for the Weibull failure time analysis, $\boldsymbol{\theta} = (\psi, \beta)$.) The likelihood function denoted by $L(\mathbf{y} \,|\, \boldsymbol{\theta})$ describes the information about the model parameters $\boldsymbol{\theta}$ provided by $\mathbf{y}$, the vector of observed data. For logistic and probit regression, the likelihood is the product of $U+1$ binomial probabilities; for example, the probability of x_t fails in n_t trials is equal to

$$\begin{pmatrix} x_t \\ n_t \end{pmatrix} p_t^{x_t}(1-p_t)^{n_t - x_t}. \tag{8}$$

For the Weibull failure time analysis, there are x_t fails (i.e., left-censored observations) and $n_t - x_t$ passes (i.e., right-censored observations) so that the likelihood has the form:

$$p_t^{x_t}(1-p_t)^{n_t - x_t}.$$

ACCEPTED MANUSCRIPT

We see that the logistic (as well as probit) regression and Weibull failure time likelihoods are the equivalent except for the specific expression of p_t because the $\begin{pmatrix} x_t \\ n_t \end{pmatrix}$ term in Equation (8) does not depend on the parameters.

The available information about $\boldsymbol{\theta}$ is summarized by the prior distribution (or prior) $\pi(\boldsymbol{\theta})$ before observing the data $\mathbf{y}$. The prior for $\boldsymbol{\theta}$ should reflect what information is available about $\boldsymbol{\theta}$. Bayes' theorem in Equation (7) shows how the data and prior information are combined to obtain the posterior distribution (or posterior) of $\boldsymbol{\theta}$ denoted by $\pi(\boldsymbol{\theta} \mid \mathbf{y})$.

In many applications, an analytical expression for the integral in Equation (7) does not exist. Instead, a Markov chain Monte Carlo (MCMC) algorithm is used to simulate samples $\{\boldsymbol{\theta}^{(m)}, m = 1, \ldots, M\}$ from the posterior distribution $\pi(\boldsymbol{\theta} \mid \mathbf{y})$. See Casella and George (1992), Chib and Greenberg (1995) and Gelman et al. (2013) for discussions of popular MCMC algorithms.

Estimates of the parameters $\boldsymbol{\theta}$ and their uncertainties are then based on the posterior samples. For each model parameter, empirical quantiles of its posterior samples, e.g., the 0.025 and 0.975 quantiles, are calculated to obtain a central 95% credible interval (i.e., the Bayesian analog of a Frequentist confidence interval). The median or 0.5 quantile of its posterior samples can be used as an estimate. Here, what is of most interest is reliability $R(t)$, which is a function of $\boldsymbol{\theta}$ as given above for logistic and probit regression and Weibull failure time analysis in Equations (3), (4) and (5). A posterior sample of R is easily obtained by evaluating $R(t)$ with a posterior sample of $\boldsymbol{\theta}$.

For logistic and probit regression, prior distributions need to be specified for β_0 and β_1. For the Weibull failure time analysis, prior distributions need to be specified for ψ and β.

The RADAR Method

Vander Wiel et al. (2011) proposes the RADAR method with the following prior model for reliability over time. RADAR is motivated by "A stockpile will likely maintain high reliability from one year to the next under a small, but ever-present, possibility that reliability could begin to decline at any time. (Vander Wiel et al., 2011)" One can then envision downward reliability trajectories over time, where degradations to reliability occur randomly and compound yearly

ACCEPTED MANUSCRIPT

once they happen. For example, if a 0.9 degradation occurs in year 6, then the year 6 reliability is 0.9 times the year 5 reliability, the year 7 reliability is $0.9^2 = 0.81$ of the year 5 reliability, and so on. Additional degradations can occur at random that also compound yearly once the happen. For example, suppose that in year 10, there is an additional 0.8 degradation, then the year 10 reliability is $0.9^5 \times 0.8$ of the year 5 reliability and the year 11 reliability is $0.9^6 \times 0.8^2$ of the year 5 reliability, and so on.

Using the Vander Wiel et al. (2011) notation, let π_t be the reliability at year t (which we previously denoted by $R(t)$ in the logistic and probit regression and Weibull failure time analysis methods) that is defined as follows:

$$\pi_0 = \rho_0,$$
$$\pi_t = \rho_t \pi_{t-1}, t = 1, 2, \ldots,$$

where the annual degradation rate ρ_t is defined as

$$\begin{aligned} \rho_t &= r_t, t = 0, 1 \\ \rho_t &= r_t \rho_{t-1}, t = 2, 3, \ldots. \end{aligned} \tag{9}$$

Moreover, independent shocks r_t are distributed as $r_t \sim OneBeta(a_t, b_t, p_t)$, where the OneBeta distribution is defined as

$$\begin{cases} 1 \text{ with probability } p_t, \\ Beta(a_t, b_t) \text{ with probability } 1 - p_t. \end{cases}$$

That is, with probability $1 - p_t$, the shock r_t is less than 1 and drawn from a $Beta(a_t, b_t)$ distribution. Equation (9) implies that this less-than-one shock r_t is applied in year t and every year thereafter. Note that this p_t is different from that used previously in logistic and probit regression and Weibull failure time analysis. Moreover, $p_t = \frac{\exp(c+dt)}{1+\exp(c+dt)}$ so that p_t decreases over time ($d < 0$), where the prior distribution for (c, d) used by the Vander Wiel et al. (2011) in its examples is a multivariate normal distribution (MVN):

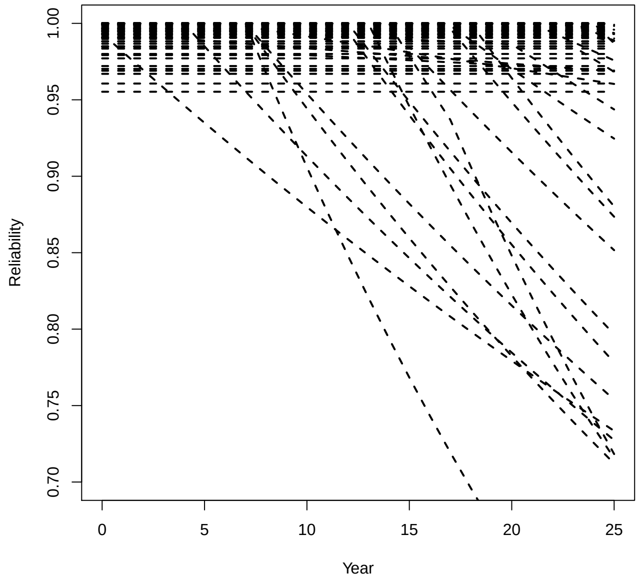
$$\begin{pmatrix} c \\ d \end{pmatrix} \sim MVN\left(\begin{pmatrix} 5.7 \\ -0.076 \end{pmatrix}, 10^{-3} \begin{pmatrix} 78 & -2.2 \\ -2.2 & 0.066 \end{pmatrix} \right). \tag{10}$$

We also use $(a_0, b_0, p_0) = (99, 1, 0.5)$ and $(a_t, b_t) = (99, 1)$ that Vander Wiel et al. (2011) used in its examples. From Vander Wiel et al. (2001), the (c, d) prior was chosen so that the 0.15

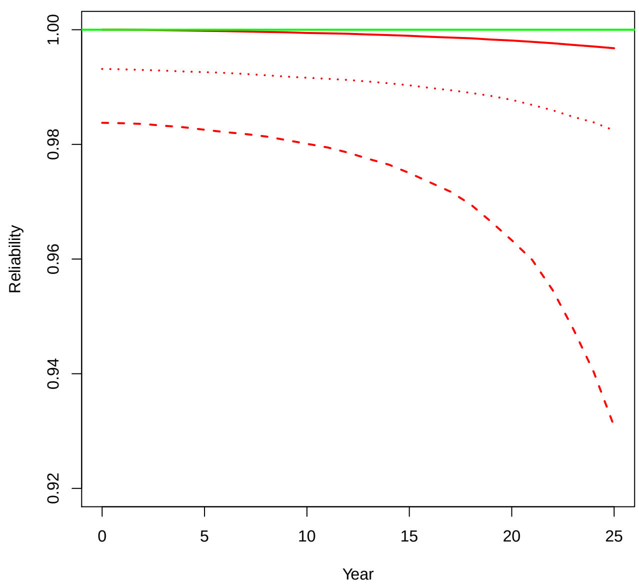
and 0.85 quantiles of the first degradation onset are 20 and 50 years, respectively. Further, they specified that the conditional probability that the first degradation onset exceeds 20 years with mean 0.85 and standard deviation 0.025 and that the conditional probability that the first degradation onset exceeds 50 years with mean 0.15 and standard deviation 0.025. Vander Wiel et al. (2001) then found the (c, d) distribution displayed in Equation (10) that satisfied this specification. When $r_t < 1$, the $Beta(99, 1)$ was chosen because the expected r_t is 0.99, i.e., a 1% reliability degradation each year once the first degradation onset occurs. If $r_t < 1$, the 0.15 and 0.85 quantiles of r_t are 0.981 and 0.998, respectively.

Figure 1a presents 100 random reliability trajectories from the RADAR method prior distribution. The first degradation onset has not occurred in many of these trajectories (i.e., the horizontal lines). If the first degradation onset has occurred, it occurs more often later than earlier. Figure 1b shows the 0.5, 0.25, 0.10 quantiles of the prior distribution of reliability π_t over 25 years as red solid, dotted and dashed lines, respectively. The green solid line corresponds to a reliability of 1.

ACCEPTED MANUSCRIPT



(a)



(b)

Figure 1: (a) 100 random reliability trajectories over 25 years from the RADAR prior distribution. (b) RADAR prior distribution of reliability π_t 0.5, 0.25, 0.10 quantiles over 25 years as red solid, dotted and dashed lines, respectively. (Green solid line shows a reliability equal to 1.)

With the data collected over time, x_t for $t = 0, \ldots, U$, the number of passes out of n_t tests, Vander Wiel et al. (2011) shows how to obtain the posterior distribution of π_t using an MCMC algorithm. Note that for the RADAR method x_t is the number of passes in n_t tests.

A Comparison of Assumptions and Properties

Logistic and probit regression are often used to fit binary (e.g., pass/fail) data as a function of a predictor other than time, where the probability of say, fail, is S-shaped as the predictor varies over its range. They are not specifically tailored to reliability analysis in that the logistic and normal distributions are not failure time distributions because they are defined on the real line, whereas failure times are non-zero. In order for reliability to decrease over time, $\beta_1 > 0$ in Equations (3) and (4). Note that both have a reliability of less than one at $t = 0$ unless $\beta_0 \approx -\infty$. That is, these models can accommodate birth defects due to manufacturing with probability $Prob_0(Fail)$. From this viewpoint, we see that only the slope β_1 controls the decrease in reliability over time.

The Weibull failure time analysis is a traditional reliability analysis that assumes the Weibull distribution as the failure time distribution. The mean of the Weibull distribution is $\psi\Gamma(\frac{\beta+1}{\beta})$, where $\Gamma(\cdot)$ is the gamma function, and the Weibull failure (hazard) rate increases, i.e., wear-out, when $\beta > 1$. Note that $R(0) = 1$ so that it cannot accommodate birth defects. Below we discuss a zero-inflated Weibull distribution to accommodate birth defects. We see that the Weibull distribution has two parameters (ψ, β) to control the reliability function for $t > 0$, where as logistic and probit regression only have one. Other failure time distributions could be considered. In fact, the reliability for the lognormal distribution is obtained by replacing t by $\log(t)$ in Equation (4). Moreover, the reliability for the loglogistic distribution is obtained by replacing t by $\log(t)$ in Equation (3).

The RADAR method is quite different from the traditional approach of assuming a single failure time distribution. First, at time 0, $R(0)$ can be less than 1 that captures reliability impacts due to birth defects; for the failure time distribution approach, $R(0) = 1$ although it can be adapted. Second, the motivation of random onsets of reliability degradation comes from

ACCEPTED MANUSCRIPT

subject-matter knowledge of failure mechanisms that initiate at various random times as the unit such as a component or system ages. Such subject-matter knowledge can be used to develop the RADAR prior model.

In order to compare the failure time analysis method and RADAR, we need to remove the difference at time 0, i.e., that reliability can be less than one. The failure time model can be adapted accordingly and is known as the zero-inflated (ZI) failure time distribution whose cdf has the following form (SAS, 2015):

$$F_{ZI}(t) = (1 - \pi_0) + \pi_0 F_{\boldsymbol{\theta}}(t),$$

where $F_{\boldsymbol{\theta}}(t)$ is the failure time distribution cdf, such as the Weibull cdf. In terms of reliability,

$$R_{ZI}(t) = \pi_0 R_{\boldsymbol{\theta}}(t),$$

where π_0 is the reliability at time 0 as it is for the RADAR model. Zero-inflated versions for the loglogistic and lognormal failure time distributions can be similarly specified.

Note that the logistic and probit regression based reliabilities in Equations (3) and (4) at time 0 can accommodate a reliability less than one, where $\beta_0 = \text{logit}(R(0))$ and $\beta_0 = \text{probit}(R(0))$, respectively. As mentioned above, logistic and probit regression only have one parameter β_1 to specify the decrease in reliability over time. The zero-inflated Weibull failure time analysis is more flexible than logistic and probit regression because it has three parameters instead of two with ψ and β specifying the decrease in reliability over time.

A Numerical Comparison

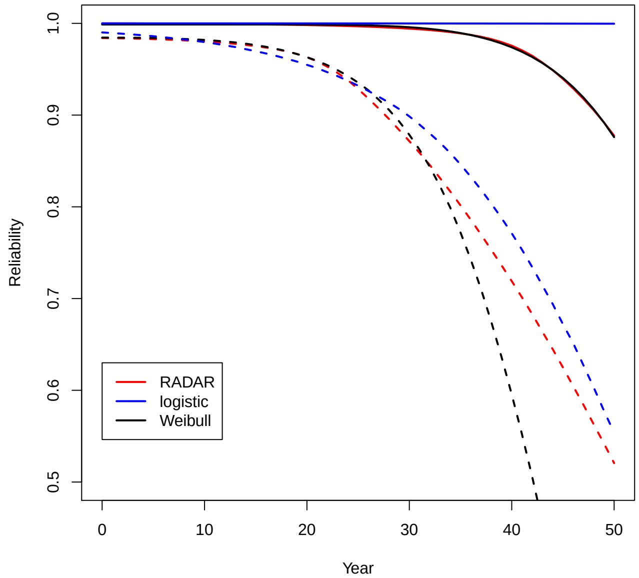
In this section, we compare logistic regression and zero-inflated Weibull failure time analysis with the RADAR method; probit regression will perform qualitatively the same as logistic regression so we do not include it in this comparison. Note that the choice of numerical quantities including the range of years, data for two scenarios and the RADAR prior model used below are chosen for illustrative purposes only. For comparability, we need to choose the prior distributions of the Weibull and logistic regression parameters so that the reliability prior distributions are similar to those in Figure 1b for the RADAR method.

Choice of Weibull Failure Time Distribution Priors

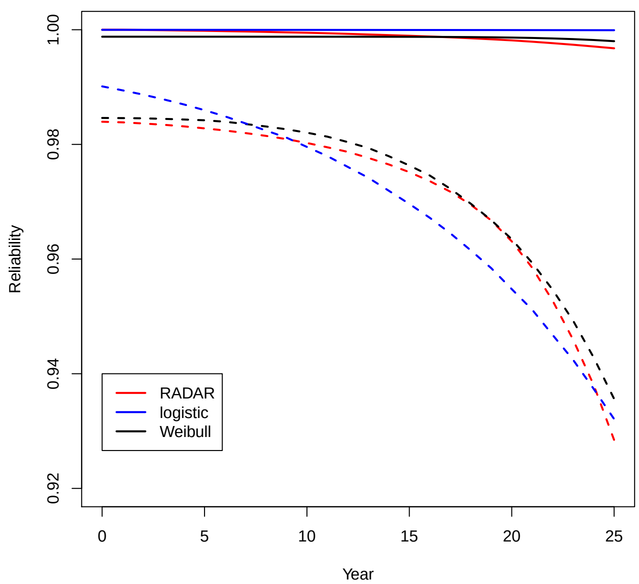
First, we identify values for the shape β, scale ψ, and π_0 so that the reliability function matches the RADAR 0.50 quantile in Figure 2a. By minimizing the sum of squared differences at times 1, 2, …, 50, we identified them as 7.3920744, 65.8089922, and 0.9987827; for non-zero failure times, the mean failure time is 61.7 years. We use these values as prior medians as follows:

- $\psi \sim Lognormal(\log(7.3920744), 0.5^2)$,
- $\beta \sim Lognormal(\log(65.8089922), 0.3^2)$, and
- $\text{logit}(\pi_0) \sim Normal(\text{logit}(0.9987827), 2^2)$.

The three variances were chosen so that the zero-inflated Weibull reliability prior 0.10 quantile matches the RADAR 0.10 quantile as best it can. See Figure 2b that shows that the 0.10 and 0.50 quantiles match well for the first 25 years. We are particularly interested in the 0.10 quantile because that is the uncertainty bound we will use for reliability.



(a)



(b)

Figure 2: (a) Prior distribution of reliability 0.50 and 0.10 quantiles over 50 years for RADAR (red), zero-inflated Weibull failure time analysis (black) and logistic regression (blue). Solid lines are medians and dashed lines are 0.10 quantiles. (b) Prior distribution of reliability 0.50 and 0.10 quantiles over 25 years for RADAR (red), zero-inflated Weibull failure time analysis (black) and logistic regression (blue). Solid lines are 0.50 quantiles and dashed lines are 0.10 quantiles.

Choice of Logistic Regression Priors

For the logistic regression prior, we chose it as best we could noting that the logistic regression is less flexible and limits how well we could match the RADAR prior. We matched the median RADAR reliability at time 0 or 0.9999838 (i.e., probability of fail is 1-0.9999838) to obtain the prior for β_0 and chose a median β_1 of 0.05 as follows:

- $\beta_0 \sim Normal(\text{logit}(1 - 0.9999838), 5^2)$.
- $\text{logit}(\beta_1) \sim Normal(\text{logit}(0.05), 0.9^2)$.

Figure 2a and 2b show that the 0.10 quantiles (dashed lines) of the three methods are similar.

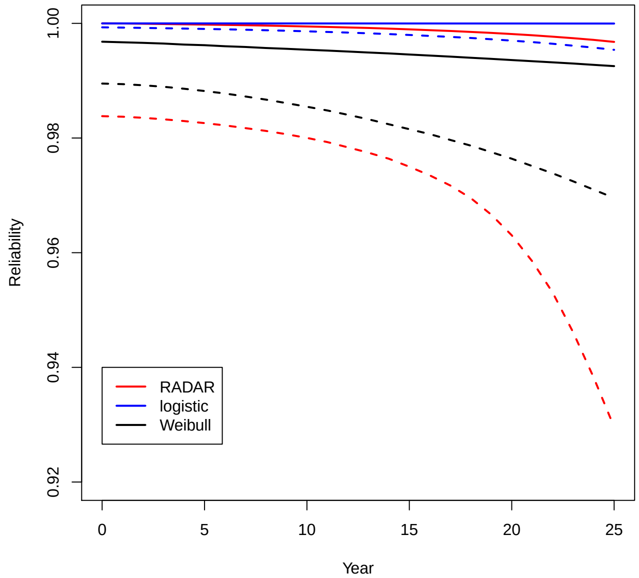
A Comparison of Reliability Estimates and Uncertainty Bounds

Now, we look at the resulting posterior distributions on reliability for the RADAR method, zero-inflated Weibull failure time analysis and logistic regression for two data scenarios. That is, for two data scenarios, we apply Bayes' theorem to update the reliability distribution. In the first scenario, we observe 11 tests at time 0 and the first 10 years with no fails. In the second scenario, we observe 11 tests at time 0 and the first 10 years with no fails like the first scenario, and in addition 11 tests are performed in year 11 with one fail out of the 11. We will focus on the 0.10 quantile as an uncertainty bound on reliability, i.e., that probability that reliability is at least that large is 0.90. In the first scenario, Figure 3a shows that the zero-inflated Weibull failure time analysis has less uncertainty (dashed line) than the RADAR method; the 0.50 quantile (solid line) is somewhat lower for the zero-inflated Weibull failure time analysis than the RADAR method. However, the logistic regression results are dramatically different with very little uncertainty; its 0.10 quantile is about the same as the 0.50 quantile of the RADAR method.

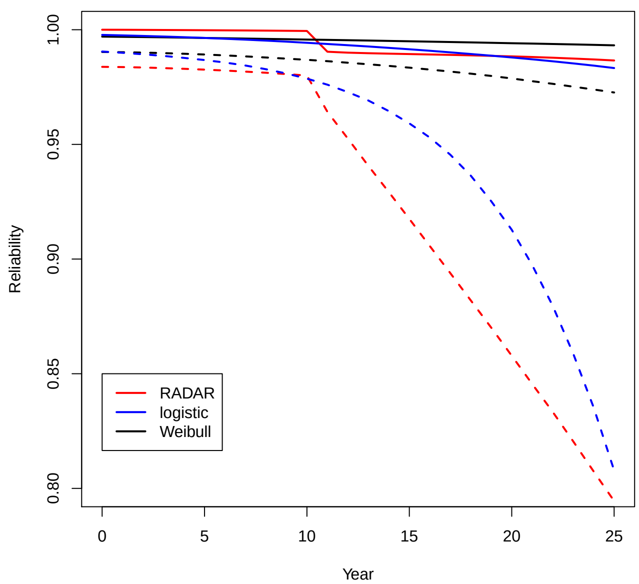
For the second scenario in which there is one fail, Figure 3b shows a dramatic difference in the uncertainty (dashed lines), with the RADAR method uncertainty much lower than that of the zero-inflated Weibull failure time analysis. We see that one fail has a dramatic effect on the RADAR method uncertainty; recall that once there is an onset of degradation it compounds over time. Note that the 0.50 quantile drops some but hardly any at all as compared with the 0.10

quantile. The second scenario shows that the nature of the zero-inflated Weibull failure time analysis and RADAR method are very different so that the appropriate choice of method for the particular application needs to be made. The logistic regression results are dramatically different from those of the first scenario; its 0.50 quantiles are near those of the RADAR method and its 0.10 quantiles have dropped dramatically and are also near those of the RADAR method. Both the logistic regression and RADAR method are sensitive to observing fail data. The sensitivity of the RADAR method is not surprising because once the first degradation onset occurs, it compounds every year. Perhaps, the sensitivity of logistic regression is somewhat surprising but recall there is only one parameter capturing the drop in reliability. If data for a 12th year are added with 9 passes in 11 tests, both the logistic regression and RADAR method curves drop even more and the logistic regression 0.10 quantiles are even lower than those of the RADAR method; the zero-inflated Weibull failure time analysis results hardly change from those of the second scenario.

ACCEPTED MANUSCRIPT



(a)



(b)

Figure 3: (a) Posterior distribution of reliability 0.50 and 0.10 quantiles over 25 years for RADAR (red), zero-inflated Weibull failure time analysis (black) and logistic regression (blue) for data with no fails from 0 to 10 years for 11 tests per year. (b) Posterior distribution of reliability 0.50 and 0.10 quantiles over 25 years for RADAR (red), zero-inflated Weibull failure time analysis (black) and logistic regression (blue)for data with no fails from 0 to 10 years and one fail in the 11th year for 11 tests per year.

ACCEPTED MANUSCRIPT

Discussion

In this article, we compared a traditional reliability approach that assumes that failures follow a Weibull distribution, whereas the RADAR method assumes that the onset of degradations to reliability occur at random. Logistic and probit regression also assumes that failures follow a logistic and normal distribution, although the logistic and normal distributions are not standard failure time distributions because they are defined on the real line; they do, however, accommodate a reliability of less than one at time 0 due to birth defects. We introduced the zero-inflated Weibull distribution that accounts for birth defects so that we could compare it with the RADAR method as well as with logistic regression.

The use of the RADAR method is motivated by leveraging subject-matter knowledge of potential mechanisms that may initiate at any time, especially as a system, component or material ages. Such subject-matter knowledge can be used to develop the RADAR method priors. We saw in the numerical comparison that the three methods, logistic regression, the zero-inflated Weibull failure time analysis and the RADAR method, provide quite different uncertainty bounds on reliability, especially when fails are observed. We also saw that logistic regression is sensitive to observed fails like the RADAR method. We have found this comparison to be useful to see how different the three methods are and how different their results can be.

One referee asked whether the results for the zero-inflated Weibull failure time analysis and logistic depend on the choice of priors. Recall that we matched the priors so the reliability distribution over time was similar to that of the RADAR method. For example, if the Weibull priors were chosen independently for the 0.10 quantile (instead of ψ) and β, then the induced prior on ψ and β results in ψ and β being correlated; yet, the results for two data scenarios using this alternative prior are nearly indistinguishable from those presented above. Given that the priors are matched, we saw that the three methods do treat the data differently especially when fails are observed.

ACCEPTED MANUSCRIPT

Acknowledgements

We thank three anonymous referees for their insightful comments that helped to improve an earlier version of this article.

References

Anderson-Cook, C.M., Klamann, R.M., and Morzinski, J. (2012). Modeling the reliability of complex systems with multiple data sources: a case study on making statistical tools accessible to engineers. *Quality Engineering*, 24, 280-291.

Casella, G. and George, E.I. (1992). Explaining the Gibbs sampler. *The American Statistician*, 46, 327-335.

Chib, S. and Greenberg, E. (1995). Understanding the Metropolis Hastings algorithm. *The American Statistician*, 49, 327-335.

Dezfuli, H., Kelly, D., Smith, C., Vedros, K., and Galyean, W. (2009). Bayesian inference for NASA probabilistic risk and reliability analysis. National Aeronautics and Space Administration Report NASA/SP-2009-569.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., and Rubin, D.B. (2013). *Bayesian Data Analysis, Third Edition*, Chapman and Hall, Boca Raton.

Hamada, M.S., Wilson, A.G., Reese, C.S., and Martz, H.F. (2008). *Bayesian Reliability*, Springer, New York.

Hosmer, Jr., D.W. and Lemshow, S. (2013). *Applied Logistic Regression, Third Edition*, John Wiley and Sons, Inc., New York.

Olwell, D.H. and Sorell, A.A. (2001). Warranty calculations for missiles with only current-status data, using Bayesian methods. *2001 Proceedings Annual Reliability and Maintainability Symposium*, 133-138.

SAS (2015). JMP®, Version 12. SAS Institute Inc., Cary.

Vander Wiel, S., Wilson, A., Graves, T., and Reese, S. (2011). A random onset model for degradation of high-reliability systems. *Technometrics*, 53, 163-172.

ACCEPTED MANUSCRIPT