Estimating a Service-Life Distribution Based on Production Counts and a Failure Database

Ryan, Kenneth J.
Hamada, Michael Scott
Vardeman, Stephen B.
Estimating a Service Life Distribution Based on Production Counts and a Failure Database

Kenneth J. Ryan  
West Virginia University, Morgantown, WV 26506  

Michael S. Hamada  
Los Alamos National Laboratory, Los Alamos, NM 87545  

Stephen B. Vardeman  
Iowa State University, Ames, IA 50011

Problem: A manufacturer wanted to compare the service life distributions of two similar products. These concern product lifetimes after installation (not manufacture). For each product there were available production counts and an imperfect database providing information on failing units. In the real case, these units were expensive repairable units warranted against repairs. Failure (of interest here) was relatively rare and driven by a different mode/mechanism than ordinary repair events (not of interest here).

Approach: Data models for the service life based on a standard parametric lifetime distribution and a related limited failure population were developed. These models were used to develop expressions for the likelihood of the available data that properly accounts for information missing in the failure database.

Results: A Bayesian approach was employed to obtain estimates of model parameters (with associated uncertainty) in order to investigate characteristics of the service life distribution. Custom software was developed and is included as supplemental material to this case study. One part of a responsible approach to the original case was a simulation experiment used to validate the correctness of the software and the behavior of the statistical methodology before using its results in the application, and an example of such an experiment is included here.

Because of confidentiality issues that prevent use of the original data, simulated data with characteristics like the manufacturer’s proprietary data are used to illustrate some aspects of our real analyses. We note also that although this case focuses on rare and complete product failure, the statistical methodology provided is directly applicable to more standard warranty data problems involving typically much larger warranty databases where entries are warranty claims (often for repairs) rather than reports of complete failures.

Key Words: dynamic parameters, failure date, installation date, lifetime distribution, limited failure population, log-likelihood, manufacture date, Markov chain Monte Carlo, simulation experiment, warranty period
1. Process description

A real manufacturer was interested in the service life distributions (i.e., distributions of times from installation to complete failure) of two similar products. Weekly production counts were available. The manufacturer also had records in imperfect databases of failure reports made between two particular dates. The first date was substantially after the beginning of production of one of the products, and the last was somewhat before the date of the requested data analysis. We thus refer to the available data as “limited window” data. There were other types of missing information as well, because some installations times and exact failure dates for units in the failure databases were not recorded.

This article addresses the analysis of such data (and even some similar situations slightly more complicated than the motivating example). The details of the motivating case are proprietary and cannot be disclosed. However, in broad terms, the methodology discussed in this article has been successfully applied where production over a number of years involved many hundreds of thousands of units, the size of the failure databases analyzed was a substantial but a very small fraction of the total production over the period studied, and a relatively small fraction of the cases analyzed were missing some information. (The methods of modeling and inference presented here could be used in other contexts and for a variety of purposes. These include the analysis of more comprehensive warranty databases and the consideration of the economic implications of particular warranty policies to specific types of product events.)

A number of considerations presented in this article make analyzing (warranty-database-type) data a subtle and challenging enterprise, requiring careful modeling and development of inference methodology. An important review article in the area is Lawless (1998), and Wu (2012) is a more recent review. Qiu, Nordman, and Vardeman (2014) illustrate related considerations where indicators of unit usage are available. In this article we make use of Bayesian inference, and are unaware of related work in the literature that has done so.

2. Data Collection

The goal of the real study was the comparison of service life distributions for two products. This was accomplished by separately estimating the characteristics of each. The following exposition describes the modeling and methodology separately applied to the information available concerning each product. We will call events in the life of a unit (i.e., product) manufacture, installation, failure, and retirement. We are interested in the probability that failure occurs (after installation and before retirement) in \( x \) or fewer units of service time (hours, days, weeks, months, etc.).

The production records consist of counts for units in \( n \) periods of length \( \tau \), i.e.,
\[ (0, \tau], (\tau, 2\tau], (2\tau, 3\tau], \ldots, ((n - 1)\tau, n\tau]. \]

A corresponding database covers failures occurring from (a beginning) time \( t = c_b \) to (an end) time \( t = c_e \). Complete cases in the database consist of values of the three variables

\[
\begin{align*}
m & = \text{Manufacture date} \\
I & = \text{Installation date} \\
Y & = \text{Failure date}
\end{align*}
\]

all stated in terms of elapsed time \( t \geq 0 \). Dates \( I \) and \( Y \) are capitalized because they will be indirectly modeled as random variables; these variables are also missing for some units in the database. (While in conventional warranty applications \( I \) and \( Y \) may be rarely if ever missing, whether or not this information was available had no impact on the company’s remedial response in the motivating real failure application. A nontrivial number of records were missing \( I \) or \( Y \), so this could not be ignored. Every reported failure was recorded and remediated regardless of whether complete information was known.) Our basic interest is in the distribution of random variable

\[ X = Y - I = \text{Lifetime}, \]

the service life (lifetime) of a unit, and we will in Section 3 directly define a probability model for \( X \) to facilitate data analysis. With a probability model for \( X \) in place, a model for random variables \( I \) and \( Y \) is induced if a model is directly defined for

\[ L = I - m = \text{Lag}, \]

i.e., the time lag between manufacture and installation for the failed units, and this will also be done in Section 3. In some contexts, a length of time

\[ w = \text{Warranty period} \]

might be established, and only unit failures with \( X \leq w \) appear in the database. Figure 1 is a concise pictorial reference for the timeline for a unit from this case study in terms of scalars \( m, w \) and random variables \( I, Y, X, L \).

\textbf{Figure 1: Timeline of a Hypothetical Unit that Failed After the Warranty Period.}
Let $N$ be the number of units manufactured by time $n\tau$. There are 6 possible cases for each unit, i.e., unit $i$ is

1. not in the database but manufactured before the beginning of the database ($m_i < c_b$),
2. not in the database but manufactured after the beginning of the database ($m_i \geq c_b$),
3. in the database with a complete record ($m_i, I_i, Y_i$),
4. in the database but its installation date $I_i$ is missing,
5. in the database but its failure date $Y_i$ is missing,
6. in the database but both its installation and failure dates $(I_i, Y_i)$ are missing.

The simulation study to come in Section 3.5 includes failure databases with $c_b = 0, c_e = 375$, and $w = 500$. In this context, case 1 cannot occur. An example of case 2 is $m = 100$ and $Y > 500$. An example of case 3 is $m = 8, I = 12$, and $Y = 29.5$. An example of case 4 is $m = 90$ and $Y = 159.6$ with $I$ missing. An example of case 5 is $m = 292$ and $I = 299$ with $Y$ missing. An example of case 6 is $m = 350$ with $(I, Y)$ both missing.

It is possible to also consider

$$R = \text{retirement date},$$

a variable that will typically be unknown, except that for units in the database $R > Y$. In the situation that we will consider, it is plausible to assume that $R - I$ is a known constant and for all units under consideration (both in the database and not in the database) $R > c_e$. But in other contexts, $R - I$ might be a random variable representing a length of service at which a competing risk removes an item from service (and ends the possibility that it shows up in the database as a failure). In those contexts, additional probability modeling beyond what we present here would be required.

3. Analysis and Results

We developed two data models and used Bayesian inference as described in Sections 3.1-3.3. We illustrate the inference methodology with a simulated data example in Section 3.4. We illustrate how to explore the behavior of the inference methodology via simulation in Section 3.5.

3.1 A Simple Model

Examination of the failure database (or other sources of additional information like the repair warranty registration information used in the real case) provides a relative frequency distribution for the lag times $L$ between manufacture and installation for the failed units; the relative frequencies can be used sensibly to characterize a discrete distribution of this lag for all units (through the fitting of a discrete parametric probability model, the fitting and
discretization of a continuous parametric probability distribution, or direct use of the relative frequency distribution or some “binned” version of it in modeling). From this, we obtain

\[ g_L(l) = P[L = l], \]

the distribution of time lags. If such extensive information on the lag distribution \( L \) were not readily available, then the uncertainty regarding this distribution would also have to be reflected into the data analysis. An estimation process for the parameters of the distribution of lags \( L \) would be expected to propagate more uncertainty into the parameter estimates for the lifetime distribution \( X \) as well. The current capability of our code is tied to our application in this sense; it can only be easily adapted to use any fixed discrete distribution for \( L \) as this was determined to be sufficient in our case study.

For the service life \( x \) of a unit, let

\[ f(x|\theta), x > 0 \]

be some parametric probability density and

\[ F_\theta(x) = P_\theta(X \leq x) = \int_0^x f(u|\theta) \, du \]

be the corresponding cumulative distribution function (cdf), where \( \theta \) is a parameter or vector of parameters; for example, we use a standard lifetime distribution such as the lognormal distribution with parameters \( \mu = \mu_{\log \text{lifetime}}, \sigma = \sigma_{\log \text{lifetime}} \). We might consider estimating the full cdf (a function of \( x \)). Or various summaries for the distribution of \( X \) can be considered, like the probability of failure

\[ F_\theta(s_0) \]

by time \( s_0 \) for \( s_0 \) some specified life of a unit (like, for example, an industry standard service time before retirement), the median unit life

\[ F_\theta^{-1}(0.5), \]

or the mean unit life

\[ \mathbb{E}_\theta[X] = \int_0^\infty xf(x|\theta) \, dx. \]

Next, consider the contribution to the likelihood for a unit of each of the 6 cases listed in Section 2. First, we indirectly handle claims not in the database (i.e., cases 1 and 2) by finding the probability a record is in the database (\( IDB \)). This event is simply

\[ IDB = \{c_b \leq Y \leq c_e\} \cap \{X < w\}, \]
so given an IDB unit, the support for the conditional distribution of $L$ is

$$[c_b - (m + w), c_e - (m - 0)]$$

and follows from solving inequality $c_b \leq Y \leq c_e$ for $L$ after substituting $Y = m + X + L$ and plugging in the boundary values of $w$ and $0$ for $X$ to get the largest possible lower and smallest possible upper bounds. If lags and lifetimes are independent, then

$$P_\theta(IDB) = \sum_{l=c_b-(m+w)}^{c_e-m} P_\theta([c_b - (m + l) \leq X \leq c_e - (m + l)] \cap \{0 < X < w\})g_L(l)$$

$$= \sum_{l=c_b-(m+w)}^{c_e-m} P_\theta([\max{0, c_b - (m + l)} \leq X \leq \min{w, c_e - (m + l)})]g_L(l)$$

by the law of total probability. The max and min in this expression provide bounds on the lifetime $X$ of a unit IDB given the lag time. When these bounds are added to manufacture date $m$, they account for the two possible starting and ending dates for the failure date $Y$ of an IDB unit (see Figure 2).

**Figure 2: Ranges for Failure Dates Y in the Database Given Installation Dates $I_i$ for $i = 1, 2$ and Ending Dates to the Warranty $I_i + w_{ij} = (I + w)_{ij}$ for $i = 1, 2, j = 1, 2$.**

In the list of likelihood contributions below, cases 1 and 2 both have a likelihood contribution of $1 - P(IDB)$, but are separated and simplified to promote computational efficiency. Cases 3-6 are more straightforward to derive from basic principles.

1. Based on simplifying $1 - P(IDB)$ when $m < c_b$, a term

$$L_1(\theta, m) = 1 - \sum_{l=c_b-(m+w)}^{c_e-m} [F_\theta(\min{w, c_e - (m + l)}) - F_\theta(c_b - (m + l))]g_L(l)$$

appears in the likelihood for every unit with $m < c_b$ not in the database.

2. Based on simplifying $1 - P(IDB)$ when $m \geq c_b$, a term

$$L_2(\theta, m) = 1 - \sum_{l=0}^{c_e-m} F_\theta(\min{w, c_e - (m + l)})g_L(l)$$
appears in the likelihood for every unit with \( m \geq c_b \) not in the database.

3. Then for a unit appearing in the database with no missing values, the corresponding likelihood term is

\[
L_3(\theta, l, Y) = f(Y - l | \theta).
\]

4. For a unit appearing in the database with only the installation date missing, the corresponding likelihood term is

\[
L_4(\theta, m, Y) = \sum_{l=0}^{\lfloor Y-m \rfloor} f(Y - (m + l) | \theta) g_L(l).
\]

5. For a unit in the database with only the failure date missing, the corresponding likelihood term is

\[
L_5(\theta, l) = F_\theta(\min\{w, c_e - l\}) - F_\theta(c_b - l)
\]

so that the failure time is interval censored.

6. For a unit in the database with both installation and failure dates missing, the corresponding likelihood term is

\[
L_6(\theta, m) = \sum_{l=\min\{0, c_e - (m + w)\}}^{c_e - m} [F_\theta(\min\{w, c_e - (m + l)\}) - F_\theta(c_b - (m + l))] g_L(l)
\]

so that the failure time is interval censored.

We write \( LL \) in place of \( \ln L \) for each of these term types. Then letting \( j(i) \) indicate the case (1-6) for the \( i \)th unit, the log-likelihood for the \( N \) (\( i = 1, \ldots, N \)) units is

\[
LL(\theta) = \sum_{i \text{ s.t. } j(i)=1} LL_1(\theta, m_i) + \sum_{i \text{ s.t. } j(i)=2} LL_2(\theta, m_i) + \sum_{i \text{ s.t. } j(i)=3} LL_3(\theta, l_i, Y_i) + \sum_{i \text{ s.t. } j(i)=4} LL_4(\theta, m_i, Y_i) + \sum_{i \text{ s.t. } j(i)=5} LL_5(\theta, l_i) + \sum_{i \text{ s.t. } j(i)=6} LL_6(\theta, m_i).
\]

### 3.2 A Limited Failure Population Model

A possible shortcoming of the simple model is that in most cases where it would be applied and observed failure rates are low, one is in the position of really using only data from the extreme left tail of a life distribution specified by \( f(x | \theta) \) to describe failure times that are
observed, and extrapolating the shape of the rest of the distribution from only this extreme tail behavior. (This is a problem both for effectiveness of inference and for the fact that the set of shapes available for describing a distribution of observed service times is quite limited if the overall observed failure rate is small.) A more flexible modeling possibility to some degree ameliorates this difficulty. Accordingly, we next consider use of a “limited failure population” model in this context. See Meeker (1987) for an early use of this kind of model and corresponding terminology.

To the simple model from Section 3.1, we add the possibility that a unit will never fail. Let

$$1 - p = \text{the probability the unit under consideration will never fail.}$$

To date we have assumed that $p = 1$ and are implicitly planning on using a standard lifetime model (such as lognormal) for a unit service time. If we now suppose that $p \in [0,1]$ and that conditioned on failing, a unit has lifetime distribution specified by $f(x|\theta)$, we are essentially putting probability $1 - p$ on a lifetime of $X = \infty$ and using the shape of $f(x|\theta)$ to describe those lifetimes that are not infinite. This modeling is more flexible than the earlier modeling, in that one more parameter is involved. With $F_\theta(x)$ as before and finite $x$, the values of the (sub-) cdf for unit service times are

$$P_{p,\theta}(X \leq x) = p F_\theta(x).$$

Consequently, the probability of failure by time $s_0$ of particular interest is $p F_\theta(s_0)$.

For the limited failure population model, the probability of each of the 6 cases listed in Section 2 follow.

1. For a unit with $m < c_b$, the probability that it fails to appear in the database is

$$L_1(p, \theta, m) = 1 - p \sum_{l = c_p - (m + w)}^{c_e - m} \left[ F_\theta(\min\{w, c_e - (m + l)\}) - F_\theta(\min\{c_b, (m + l)\}) \right] g_L(l),$$

and a term like this appears in the likelihood for every unit with $m < c_b$ not in the database.

2. For a unit with $m \geq c_b$, the probability that it fails to appear in the database is

$$L_2(p, \theta, m) = 1 - p \sum_{l = 0}^{c_e - m} F_\theta(\min\{w, c_e - (m + l)\}) g_L(l),$$
and a term like this appears in the likelihood for every unit with \( m \geq c_b \) not in the database.

3. Then for a unit appearing in the database with no missing values, the corresponding likelihood term is

\[
L_3(p, \theta, i, Y) = pf(Y - l | \theta).
\]

4. For a unit appearing in the database with only the installation date missing, the corresponding likelihood term is

\[
L_4(p, \theta, m, Y) = p \sum_{l=0}^{\lfloor Y - m \rfloor} f(Y - (m + l) | \theta)g_L(l).
\]

5. For a unit in the database with only the failure date missing, the corresponding likelihood term is

\[
L_5(p, \theta, l) = p[F_\theta(\min\{w, c_e - l\}) - F_\theta(c - l)].
\]

6. For a unit in the database with both installation and failure dates missing, the likelihood term is

\[
L_6(p, \theta, m) = p \sum_{l=\min\{0, c_b - (m+w)\}}^{c_e - m} [F_\theta(\min\{w, c_e - (m + l)\}) - F_\theta(c_b - (m + l))]g_L(l).
\]

We again write \( LL \) in place of \( \ln L \) for each of these term types. The log-likelihood is then

\[
LL(p, \theta) = \sum_{i \text{ s.t. } j(i) = 1} LL_1(p, \theta, m_i) + \sum_{i \text{ s.t. } j(i) = 2} LL_2(p, \theta, m_i) + \sum_{i \text{ s.t. } j(i) = 3} LL_3(p, \theta, i, Y_i) + \sum_{i \text{ s.t. } j(i) = 4} LL_4(p, \theta, m_i, Y_i) + \sum_{i \text{ s.t. } j(i) = 5} LL_5(p, \theta, l_i) + \sum_{i \text{ s.t. } j(i) = 6} LL_6(p, \theta, m_i).
\]

### 3.3 Bayes Analysis

We used Bayesian inference not only so that we could potentially incorporate any available information (i.e., a prior distribution) about the model parameters, but also because it directly provides the uncertainty about the model parameters. Uncertainty about any function of the model parameters is easily obtained using samples from the posterior distribution.

Consider a Bayes analysis for the simple model. We let

\[
h(\theta)
\]
be the prior probability density (prior) for the parameter vector $\theta$. For each unit $i = 1, \ldots, N$ under consideration we assume $m_i$ is known (units not in the database can usually with little approximation error be assumed to have been made in the middle of their respective production periods). We implemented this type of Bayes analysis by sampling from a posterior probability density proportional to $\exp(LL(\theta))h(\theta)$ with a Metropolis-Hastings-within-Gibbs Markov chain Monte Carlo (MCMC) algorithm (Gelman et al., 2013).

For the limited failure population model, let

$$h(p, \theta)$$

be the prior probability density for the parameter vector $(p, \theta)$. A natural form for the prior is a product form

$$h(p, \theta) = h_1(p) \times h_2(\theta)$$

although nothing here requires that. For each unit $i = 1, \ldots, N$ under consideration we will continue to suppose that $m_i$ is known. We also implemented this type of Bayes analysis with an MCMC sample from a probability density proportional to $\exp(LL(p, \theta))h(p, \theta)$.

### 3.4 Analysis of an Illustrative Simulated Failure Dataset

We cannot present results from our proprietary application. What we can and will do here is first show details of an illustrative analysis of a single simulated dataset using the Bayes methods of Section 3.3.

Consider analysis of the simulated failure database in Table 1. (This is one of the many datasets treated in the simulation study of Section 3.5.) It was generated using

- $p = 10^{-4}$ (a limited failure population after the assumptions of Section 3.2),
- a lognormal density $f(x|\mu, \sigma)$ with parameters $\mu = \mu_{\text{loglifetime}} = 4.58$ and $\sigma = \sigma_{\text{loglifetime}} = 1$ together with the value of $p$ giving a small failure probability of 0.0001 by $t = 1,000$,
- $c_b = 0$, $c_e = 375$, $w = 500$,
- installation lags $L$ as a random sample from the Geometric(0.1) distribution,
- per period production of 1,000 units, and
- a 0.10 probability for missing values for cases in the failure database.

(Manufacturing periods $m = 1, \ldots, 374$ all produced units not in the database that must contribute to the likelihood. For example, all 1,000 units manufactured with $m = 1$ are not represented in the database, while only 998 of the units with $m = 90$ fail to appear.)
Table 1: A Simulated Failure Database.

<table>
<thead>
<tr>
<th></th>
<th>I</th>
<th>Y</th>
<th>j(i)</th>
<th></th>
<th>I</th>
<th>Y</th>
<th>j(i)</th>
<th></th>
<th>I</th>
<th>Y</th>
<th>j(i)</th>
</tr>
</thead>
<tbody>
<tr>
<td>8</td>
<td>12</td>
<td>29.25</td>
<td>3</td>
<td>81</td>
<td>87</td>
<td>122.79</td>
<td>3</td>
<td>202</td>
<td>203</td>
<td>307.88</td>
<td>3</td>
</tr>
<tr>
<td>10</td>
<td>29</td>
<td>101.72</td>
<td>3</td>
<td>90</td>
<td>-</td>
<td>159.66</td>
<td>4</td>
<td>208</td>
<td>-</td>
<td>277.55</td>
<td>4</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
<td>57.98</td>
<td>3</td>
<td>90</td>
<td>-</td>
<td>354.17</td>
<td>4</td>
<td>222</td>
<td>226</td>
<td>327.61</td>
<td>3</td>
</tr>
<tr>
<td>17</td>
<td>21</td>
<td>206.40</td>
<td>3</td>
<td>137</td>
<td>142</td>
<td>-</td>
<td>5</td>
<td>255</td>
<td>257</td>
<td>319.52</td>
<td>3</td>
</tr>
<tr>
<td>34</td>
<td>35</td>
<td>78.79</td>
<td>3</td>
<td>162</td>
<td>168</td>
<td>202.56</td>
<td>3</td>
<td>255</td>
<td>-</td>
<td>367.45</td>
<td>4</td>
</tr>
<tr>
<td>47</td>
<td>58</td>
<td>101.33</td>
<td>3</td>
<td>165</td>
<td>175</td>
<td>275.32</td>
<td>3</td>
<td>264</td>
<td>267</td>
<td>307.93</td>
<td>3</td>
</tr>
<tr>
<td>47</td>
<td>52</td>
<td>73.40</td>
<td>3</td>
<td>169</td>
<td>-</td>
<td>190.06</td>
<td>4</td>
<td>266</td>
<td>272</td>
<td>297.08</td>
<td>3</td>
</tr>
<tr>
<td>51</td>
<td>51</td>
<td>227.63</td>
<td>3</td>
<td>178</td>
<td>-</td>
<td>205.97</td>
<td>4</td>
<td>268</td>
<td>280</td>
<td>346.80</td>
<td>3</td>
</tr>
<tr>
<td>54</td>
<td>72</td>
<td>206.30</td>
<td>3</td>
<td>181</td>
<td>187</td>
<td>330.98</td>
<td>3</td>
<td>292</td>
<td>299</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>72</td>
<td>122</td>
<td>147.06</td>
<td>3</td>
<td>194</td>
<td>216</td>
<td>244.90</td>
<td>3</td>
<td>542</td>
<td>544</td>
<td>-</td>
<td>5</td>
</tr>
<tr>
<td>80</td>
<td>101</td>
<td>122.34</td>
<td>3</td>
<td>196</td>
<td>202</td>
<td>328.21</td>
<td>3</td>
<td>652</td>
<td>654</td>
<td>-</td>
<td>5</td>
</tr>
</tbody>
</table>

Priors with the likelihood from Section 3.2 were independent with

\[ \mu \sim \text{N}(0, 100^2) \]

\[ \log(\sigma) \sim \text{N}(0, 10^2) \]

\[ \text{logit}(p) \sim \text{N}(0, 10^2). \]

MCMC (for details see the next subsection) then produces 95% credible intervals for the parameters and (parametric functions) given in Table 2. It is evident that for this particular simulation every interval covers the corresponding parameter. The robustness of our Bayesian approach with this choice of diffuse prior is discussed further in Section 3.5.

Table 2: Bayes 95% Credible Intervals for the Limited Failure Population Model of Section 3.2 Based on Analysis of the Data of Table 1.

<table>
<thead>
<tr>
<th>Parameter/Value</th>
<th>Credible Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>( p = 10^{-4} )</td>
<td>( (8.3 \times 10^{-5}, 3.1 \times 10^{-4}) )</td>
</tr>
<tr>
<td>( \mu = 4.58 )</td>
<td>( (3.96, 6.14) )</td>
</tr>
<tr>
<td>( \sigma = 1 )</td>
<td>( (0.70, 1.97) )</td>
</tr>
<tr>
<td>( pF(200) = 7.7 \times 10^{-5} )</td>
<td>( (7.0 \times 10^{-5}, 1.4 \times 10^{-4}) )</td>
</tr>
<tr>
<td>( pF(1,000) = 10^{-4} )</td>
<td>( (8.3 \times 10^{-5}, 1.4 \times 10^{-4}) )</td>
</tr>
</tbody>
</table>
To illustrate the important extra modeling flexibility provided by the limited failure population modeling, Figure 3 plots the generating model's cumulative failure probability for \( t = 10, 20, \ldots, 10^4 \) along with 95% credible limits from Bayes analyses based on the data of Table 1 and the models of both Sections 3.1 and 3.2. (We effectively changed the prior on \( p \) to put probability 1 on \( p = 1 \) while using the 2-parameter model from Section 3.1.) The analysis based on the misspecified 2-parameter model substantially over-estimates the cumulative failure probabilities for times beyond the longest lifetime in the failure database, while every interval based on the (correct) limited failure population model covers the underlying cumulative failure probability.

**Figure 3: True Population Lifetime CDF (solid) and 95% Individual Bayes Credible Limits for the 2- (dashed) and 3-Parameter (dotted) Models Given the Data of Table 1.**

3.5 Exploring the Behavior of the Inference Methodology

It is important to have a sound basis for believing that the operating characteristics of the statistical methodology are adequate for the intended use of a new inference method. This need can be addressed in an application through simulation across a spectrum of situations including ones that could yield data more or less like the real ones. To illustrate the kind of thinking involved in establishing the effectiveness of our methods, we used the factorial design summarized in Table 3.

We wanted the code of our Bayesian implementation (which is included as supplementary material) to be numerically robust, computationally efficient, and have some level of generality that might make it useful in future consulting encounters. But mainly we wanted to be sure it was coded correctly. In this last regard, we considered it to be a matter of best-practice to subject custom code to extensive testing before using it for its intended consulting purpose. So, for example, we considered the information in Table 3 expecting to see wider interval estimates, i.e., less precision, when the warranty factor was set to its low level. We also wanted to establish the credibility of our Bayesian implementation before presenting results to our clients. This justification from the simulation results to come in this section is twofold. First, the choice of our diffuse prior from Section 3.4 did not affect the frequentist
coverage probabilities of our methods, so our analysis was primarily likelihood driven. Second, maximum likelihood interval estimation resulted in numerical instabilities and was not a viable option. This second point is discussed in more detail in the final paragraph of this section after presenting the results from our simulation study, which is done now.

Table 3: Factors and Levels in the Simulation Study.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Levels</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Generating Model</td>
<td>Simple (1) vs. Limited Failure Population (2)</td>
</tr>
<tr>
<td>Early Failure Probability</td>
<td>Large (0.001 by ( t = 1000 )) (1) vs.</td>
</tr>
<tr>
<td></td>
<td>Small (0.0001 by ( t = 1000 )) (2)</td>
</tr>
<tr>
<td>Data Window</td>
<td>( c_b = 0 ) and ( n = c_e = 375 ) (1) vs.</td>
</tr>
<tr>
<td></td>
<td>( c_b = 125 ) and ( n = c_e = 500 ) (2) vs.</td>
</tr>
<tr>
<td></td>
<td>( c_b = 0 ) and ( n = c_e = 500 ) (3)</td>
</tr>
<tr>
<td>Warranty Period</td>
<td>( w = 100 ) (1) vs. ( w = 500 ) (2)</td>
</tr>
<tr>
<td>Lag Distribution</td>
<td>Geometric(0.04) (1) vs. Geometric(0.1) (2)</td>
</tr>
<tr>
<td>Per Period Production</td>
<td>Low (500) (1) vs. High (1000) (2)</td>
</tr>
<tr>
<td>Missingness in Database Records</td>
<td>Low (0.01) (1) vs. High (0.1) (2)</td>
</tr>
</tbody>
</table>

To be more specific than Table 3, the combinations of data generating population and early failure probability employed were produced using lognormal densities \( f(x|\mu, \sigma) \) and were

**Population 1.** \( p = 1, \mu = 37.81, \sigma = 10 \) for the simple population with large early failure probability,

**Population 2.** \( p = 1, \mu = 44.10, \sigma = 10 \) for the simple population with small early failure probability,

**Population 3.** \( p = 10^{-3}, \mu = 4.58, \sigma = 1 \) for the limited failure population with large early failure probability, and

**Population 4.** \( p = 10^{-4}, \mu = 4.58, \sigma = 1 \) for the limited failure population with small early failure probability.

Figure 4 shows the limited failure cdfs for Populations 3 and 4 and lognormal cdfs chosen to approximate them for relatively small \( x \) (values up to \( w \)). (The approximations were made to minimize the integral of the squared difference between the limited failure cdf and the approximating lognormal cdf over the interval \((0, w)\).) When \( w = 100 \), there are huge differences in the cdfs for even somewhat larger \( x \) (that would be important for extrapolation from available data beyond a warranty value), and plots like Figure 4 in the context of our case
study helped to differentiate between minor and major extrapolations (the latter case being that where the 2- and 3-parameters model might produce very different results).

Figure 4: The CDFs (Solid) for Lifetime Under Populations 3 (Top) and 4 (Bottom) with the Best Lognormal Approximations (Dashed) Through Times w = 100 (Left) and w = 500 (Right).

There are $2^6 \times 3 = 192$ different data generating “situations,” i.e., level combinations of the factors listed in Table 3. Ten replicate datasets were generated at each situation, and all 1,920 databases were generated in 8 minutes on a 3GHz processor. Missingness in database records was created by deleting each value of installation and failure time independently with either low (0.01) or high (0.1) probability. The mean database sizes were 16.2 and 167.5 for small and large early failure probability populations. The 2-parameter model of Section 3.1 and the 3-parameter model of Section 3.2 were used to analyze each of the 1,920 simulated databases with the Bayes methods from Section 3.3 as well as maximum likelihood. Like our motivating case study, the lag distributions were taken as known and employed as inputs in the data analyses. Notice that the simple modeling of Section 3.1 is a special case of the modeling of Section 3.2, so there is model-misspecification only when the data generated according to the limited failure population of Section 3.2 were analyzed with methods based on the 2-parameter model.

Priors for the Bayes analyses were as in the example of Section 3.4. Optimal Metropolis-Hastings step sizes (MHSS) for Metropolis-within-Gibbs algorithms having 40% accept rates were estimated by situation and model using the method of Graves (2011) within 30 hours.
Maximum likelihood estimates (MLEs) were first obtained in 87 hours, and all MCMC chains were started at the corresponding dataset MLEs. With MLE starting points and MHSSs, 2- and 3-parameter models were fit to each database via MCMC. Initial $10^4$-iteration burn-ins for all chains were discarded and subsequently $10^4$ MCMC sample iterations were saved as representing the posterior distributions. These MCMCs for all 1,920 databases finished in 540 and 610 hours for the 2- and 3-parameter models. Acceptance rates were good across the board (typically quite close to 40%). And time series plots of $\mu, \sigma, p, pF(200), pF(1,000)$ all indicated adequate mixing of the chains.

Next, we consider the coverage and then the length properties of Bayes posterior 95% credible intervals for the probability $pF(s_0)$ of a lifetime of no more than $s_0$ with $s_0 = 200$ or 1,000. First for the Bayes intervals from the 3-parameter model and $s_0 = 200$, the estimated average coverage probability for the truth $pF(200)$ was 97.7% over the 1,920 databases. Coverage rates ranged from 7/10 to 10/10 across the 192 situations, and a binomial regression of the number covered (out of 10 trials) had no significant effects (each $p$-value > 0.05) for the model with main effects and 2-factor interactions, allowing us to keep the null hypothesis of a constant coverage rate across the situations if the 3-parameter model is used for data analysis.

Table 4 lists average coverage rates. For the 2-parameter model Bayes intervals, there is a clear (and expected) drop in average coverage probability due to model misspecification in the bottom row. The average coverage rates for the 3-parameter model Bayes intervals are generally quite close to a nominal 95%. We attribute this robustness to our choice of proper prior in Section 3.4; with high prior variances, the resulting Bayesian data analysis is expected to be primarily likelihood driven. However, the average coverage of 85.5% in the lower right cell deserves some special attention. This dip in average coverage probability was due to a bimodality in the posterior of $p$ that occurred in the 12 situations with population 4, the short warranty period, and high production; see Figure 5 for an example. If there is such uncertainty as to whether or not the population is a limited failure population (i.e., does $p = 1$ or does $p = \epsilon$ for some small $\epsilon > 0$?), then there is more uncertainty in parametric functions such as $pF(1,000)$. This follows because two very different estimates for $pF(1,000)$ are plausible when extrapolating to $s_0 = 1,000$ as previously seen in Figure 4. In these examples of non-coverage where the posterior put sizeable posterior probability on $p \approx 1$, the lower endpoint of the 95% Bayes credible interval always exceeded the true $pF(1,000)$. 
Table 4: Average Coverage Probabilities of Bayes 95% Credible Intervals. The Populations are the Data Generating Model from Table 3, whereas the Data Analysis Models are the 2-(Section 3.1) or 3-Parameters (Section 3.2) Models with the Priors from Section 3.4.

<table>
<thead>
<tr>
<th>Population</th>
<th>Data Analysis Model</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2-Parameter Model</td>
</tr>
<tr>
<td></td>
<td>( p_F(200) )</td>
</tr>
<tr>
<td>2-Parameters</td>
<td>0.964</td>
</tr>
<tr>
<td>3-Parameters</td>
<td>0.646</td>
</tr>
</tbody>
</table>

Figure 5: Posterior Distribution of \( p \) (Left) and \( p_F(1,000) \) (Right) Given a Population 4 Database having the Short Warranty Period and High Production. Vertical Lines mark the Truth (Solid) and 95% Bayes Limits (Dashed).

Figure 6 shows the pattern of Bayes credible interval lengths for estimating early failure probabilities \( 10^4 p_F(200) \). As expected, these lengths decrease with increased warranty period, production level, and across levels 1-3 of window configuration. They are typically smaller when data are generated according to a simple lifetime model and larger for limited failure population models. The effects of warranty length are especially pronounced when data are generated by a limited failure population, i.e., filled-in diamonds versus circles. There are \( 2^2 = 4 \) symbols of each type in each subpanel, corresponding to the level combinations of the missingness and lag distribution factors from Table 3. Our choices for the levels of missingness and lag distribution in this simulation study had little effect on interval length, except in situations with window 1, low production, the short warranty period, and a 2-parameter data generating population. To study these effects further, 4 filled-in circles in a bottom panel were labeled (i)-(iv) by descending log length. Points (i) and (ii) [points (iii) and (iv)] correspond to the high [low] mean lag distribution, so the lag distribution had a larger effect than missingness on log-mean-length. Points (i) and (iii) [points (ii) and (iv)] correspond to the high [low] levels of missingness, so as might be expected, the higher level of missingness resulted in larger log-mean-lengths at a given lag distribution.
Figure 6: Plots of Log-Mean-Length of 95% Credible Intervals for $10^4 pF(200)$ Versus Mean Database Size. The Top Rows Label the Panels by Early Failure Probability, Production, and Window (See Table 3). The Data Analysis Model has: 2-Parameters (Top) and 3-Parameters (Bottom). Data Generating Populations 1 or 2 (with 2-Parameters) are Indicated by Open Markers, and Populations 3 or 4 (with 3-Parameters) by Filled-in Markers. Short and Long Warranty Periods are Indicated by Circles and Triangles. Sizes of Markers Indicate Coverage Relative Frequencies for Parameter $10^4 pF(200)$ Out of 10 Simulated Trials.

Figure 7 shows no appreciable increase in mean interval length using the 3-parameter instead of the 2-parameter-model Bayes methodology even when data are generated by the simple 2-parameter Populations 1 and 2. In fact, a majority of 56 out of 96 of the points are
below the reference line, so the 2-parameter methodology produces wider credible intervals on average. It may at first seem counterintuitive that the model with more parameters results in shorter interval estimates, but there is a plausible explanation. In the context of the Table 3 parameters, the parametric function of interest \( pF(200) \) is a relative frequency that is close to zero, and the 3-parameter model suggests smaller values for \( pF(x) \) in Figure 4, when for example \( w = 100 \) and \( x = 200 \) is an extrapolation. This was especially pronounced in the credible intervals for \( pF(1,000) \). We made versions of Figures 6 and 7 for \( pF(1,000) \), but excluded them because the results were comparable.

**Figure 7: Mean Lengths of Bayes 95% Credible Intervals for \( 10^4pF(200) \) (3-Versus 2-Parameter Methodologies for 2-Parameter Populations 1 and 2) with a 45° Reference Line.**

We attempted to apply maximum likelihood methods to each simulated database, but ran into numerical instabilities. If \( p = 1 \) is on the boundary of the parameter space and the 3-parameter model is fit, then standard regularity conditions are not satisfied, and these types of situations created numerical instabilities when we attempted to compute maximum likelihood based asymptotic confidence intervals for many of the simulated databases. So our Bayesian framework with its diffuse, proper prior from Section 3.4 added a robust, numerical stability.

4. Discussion

In the motivating case, the methods of this paper were applied to both products. There were some (not unexpected) discernable differences in the absolute magnitudes of estimated failure probabilities of interest across variants of the models treated here (parametric families of distributions and “ordinary” versus “limited failure population” cases). But the central overall conclusion was that there were no important discernable differences between the two
products across a suite of model cases. This allowed the company to go forward with confidence in what was the newer product.

In this article, we have addressed the need for inference methodology for a type of failure data met in a real industrial problem. The tools developed for the case have additional applications to other situations, notably to ones involving more standard (larger/more complete in terms of the fraction of units covered) warranty databases (e.g., problems concerned with the setting of economically feasible warranty periods). Both standard parametric lifetime distributions and limited failure population models have been considered, and a substantial simulation experiment has both 1) demonstrated that the Bayes methodology can be extremely effective in making inferences for failure probabilities based on incomplete and otherwise limited (failure or warranty) databases and 2) provided a template for the kind of checking that provides assurance that the operating characteristics of the methodology are adequate for an application.

We note that there are a variety of ways in which the work discussed here might be extended. For one thing, other warranty-related information might be available and used to advantage. For example, installation dates for units not producing failures might be known and used in the analysis. (Such units are known to have lifetimes exceeding \( \min\{c_e - I, w\} \).) And as mentioned earlier, there is also the possibility of more detailed accounting for retirement times \( R \) that may not be observed but cannot necessarily be assumed to exceed \( c_e \).

A situation where all failures are represented in the database can be handled by simply using \( w = \infty \). It is equally possible to modify our development by allowing warranty periods to be unit-specific (allowing, for example, consideration of situations where it is possible to purchase extended warranties). The Appendix also considers a more complex generalization, a dynamic lifetime model. This generalization was not offered to the client in the present case study. But this Appendix provides an extension that might be needed when details of a product or use conditions are rapidly evolving.

While these possibilities remain, the Bayes version of the Sections 3.1 and 3.2 analyses provided important real answers in the motivating context, lending engineering insight beyond that available before the development of the methodology of this article. R code (see R Development Core Team (2008)) for implementing the Bayes analyses of Sections 3.1-3.3 is available as Supplementary Material at http://www.asq.org/pub/jqt/.

Acknowledgments

The authors thank two anonymous referees for their insightful comments on earlier versions that helped improve this article.
Appendix: A Dynamic Lifetime Model

It is possible that units manufactured in different epochs have different lifetime characteristics and that allowing for and tracking the behavior of those characteristics over time is of importance. A dynamic generalization of the foregoing modeling and inference provides a natural way to address this problem.

Suppose that it is reasonable to identify manufacturing dates

$$0 = m_0 < m_1 < \cdots < m_K = c_e$$

for which it is safe to assume that the $K$ corresponding cohorts of intervening manufacturing dates

$$c_k = \{m: m_{k-1} < m \leq m_k \}$$

for $k = 1, \ldots, K$ define epochs, during each of which parameter vectors $(p, \theta)$ (or just $\theta$) are constant, but that parameters may change between consecutive epochs. Let
\((p_k, \theta_k)\) = a parameter vector for a model describing units manufactured in epoch \(k\).

One might consider methods of modeling and analysis for the series

\[ (p_1, \theta_1), (p_2, \theta_2), ..., (p_K, \theta_K). \]

The most obvious approach to this problem is to simply treat the epochs separately and apply some of the foregoing methodology an epoch at a time. One might then (more or less "after the fact") consider quantifying trends in series of estimated values of parametric functions like

\[ \hat{p}_1 F_{\hat{\theta}_1}(s_0), \hat{p}_2 F_{\hat{\theta}_2}(s_0), ..., \hat{p}_K F_{\hat{\theta}_K}(s_0). \]

(Fitting of trends and smoothing might be employed.)

A more formal methodology would be to adopt some kind of state space modeling, i.e., a probability model for the evolution of the parameter vectors. For sake of example, consider use of the simple version of the model (generalization to the limited failure population is straightforward) where one describes the successive parameters \(\theta_k\) using a normal random walk with independent coordinates. That is, for \(\theta\) a \(q\)-dimensional parameter let

\[ h_\theta(\cdot | \theta) \]

be a \(q\)-variate normal density with mean \(\theta\) and covariance matrix \(\Delta = \text{diag}(d_1, d_2, ..., d_q)\). A possible prior distribution for the sequence of lifetime parameters

\[ \theta_1, \theta_2, ..., \theta_K \]

has probability density

\[ \prod_{k=1}^{K} h_\Delta(\theta_k | \theta_{k-1}) \]

(for \(\theta_0\) some "initialization/starting value" for the sequence). (One can either take \(\theta_0\) as a user-chosen parameter specifying a prior density \(h_\Delta(\cdot | \theta_0)\) for \(\theta_1\) as above, or include some term specifying a prior for it in the product above.) Then for \(LL(\theta_k)\) the log-likelihood for \(\theta_k\) based on manufacturing epoch/cohort \(k\) and fixed/known \(\Delta\), the posterior for the parameters has density proportional to

\[ \prod_{k=1}^{K} h_\Delta(\theta_k | \theta_{k-1}) \exp(LL(\theta_k)) \]

From this, successive substitution MCMC algorithms that update in turn the \(\theta_k\), with the conditional probability density for \(\theta_k\) given all other parameters proportional to
\[ h_\Delta(\theta_k | \theta_{k-1}) \exp(LL(\theta_k)) h_\Delta(\theta_{k+1} | \theta_k) \]

(so the distribution of the update of \( \theta_k \) depends only upon the current value of any immediately preceding or succeeding parameter vector and the current likelihood) are obvious.

One could go so far as to treat the random walk variances in \( \Delta \) as hyperparameters and employ some joint prior density for them, say \( h(\Delta) \). A joint density for \( \Delta \) and the sequence of lifetime parameters is then proportional to

\[ h(\Delta) \prod_{k=1}^{K} h_\Delta(\theta_k | \theta_{k-1}) \exp(LL(\theta_k)) \]

and MCMC for generating observations from this probability density are then again clear.

In either case (\( \Delta \) fixed or not) samples from the posterior of the parameter sequence then produce samples from the posterior of

\[ F_{\theta_1}(s_0), F_{\theta_2}(s_0), ..., F_{\theta_K}(s_0) \]

and inferences of practical interest about the evolution of the lifetime distributions.