BAYESIAN PROBABILITY PAPERS

M. Kaminskiy
Center for Technology and Systems Management,
University of Maryland, College Park, MD 20742, USA

V. Krivtsov [2]
Office of Technical Fellow in Quality Engineering,
Ford Motor Company, Dearborn, MI 48120, USA

ABSTRACT

The paper introduces a method of Bayesian probability papers for estimating the reliability function of popular in reliability analysis location-scale life time distributions. We use simulation examples to validate the method and a real engineering data example to illustrate its practical application.

Key words: Bayesian regression, Bayesian estimation, probability paper.

1. INTRODUCTION

A Bayes' approach to reliability (survivor) function estimation is introduced. This Bayes' approach is similar to the widely used *probability papers*, which can be considered as the respective *classical* analog.

The traditional probability paper technique is applied to the distributions, whose cumulative distribution functions (or reliability functions) can be linearized in the way that the distribution parameters are estimated through the simple linear regression model $y = ax + b$. The family of such distributions includes such popular distributions as the Weibull, exponential, normal, log-normal, and log-logistic. The estimates obtained using the probability papers are considered as the initial estimates (for the subsequent nonlinear estimation), but in reliability engineering practice, they often turn out to be the final ones as well.

In this paper, the basic assumptions related to the simple normal linear regression model are discussed in the framework of the probability papers procedures, and the basic violations of these assumptions are specified.

Analogously, the probability papers procedures are considered from the standpoint of Bayesian simple linear regression model. It is shown that the Bayesian simple regression model can be applied to the probability paper procedures with approximately the same number of violations of the respective

[2] Corresponding author. *E-mail address:* VKrivtso@Ford.com

Bayesian assumptions as the classical probability papers procedures have with respect to classical simple regression model.

The discussion below is limited to the respective point estimation procedures.

2. CLASSICAL PROBABILITY PAPERS AND SIMPLE LINEAR REGRESSION

The above mentioned linearization is applicable to those lifetime (time to failure) distributions for which some transform of lifetime has a location-scale parameter distribution. The location-scale distribution for a lifetime random variable t is defined as the distribution having the probability density function (PDF), which can be written in the following form [Lawless, 2003]:

$$f(t) = \frac{1}{b} f_0\left(\frac{t-u}{b}\right) \qquad -\infty < t < \infty \qquad (1)$$

where u ($-\infty < u < \infty$) and $b > 0$ are location and scale parameters, and $f_0(x)$ is a specified PDF on $(-\infty, \infty)$.

2.1 Classical Simple Linear Regression

Consider the basic assumptions associated with the simple normal linear regression model. Let's assume that a random *response variable y* fluctuates about an unknown nonrandom *response function* $\eta(x)$ of nonrandom known *explanatory variable x*, that is $y = \eta(x) + \varepsilon$, where ε is the random fluctuation or error. In the following, we consider $\eta(x)$ in the simple linear form, so that it can be written as

$$y(x)= \beta_0 + \beta_1 x + \varepsilon, \qquad (2)$$

or as

$$y(x)= \beta_0 x_0 + \beta_1 x_1 + \varepsilon, \qquad (2\text{-}1)$$

where β_0 and β_1 are unknown parameters to be estimated; $x_0 \equiv 1$, and $x_1 \equiv x$. The data related to model (2) are the pairs composed of observations $y_i(x_i)$ and the respective values x_i (i= 1, 2, . . ., n), $n \geq 2$.

For these observations, it is assumed that

$$y_i(x_i) = \beta_0\, x_0 + \beta_1 x_i + \varepsilon_i \qquad (3)$$

where errors ε_i (i= 1, 2, . . ., n) are independent normally distributed with mean 0 and variance σ^2. In other words, the observations $y_i(x_i)$ are independent normally distributed with mean $\beta_0\, x_0 + \beta_1 x_i$ and variance σ^2.

For the following discussion, let us consider Equation (3) in its matrix form, which is given by

$$Y = XB + \varepsilon \qquad (3\text{-}1)$$

where

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{bmatrix}, \qquad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ . & . \\ . & . \\ . & . \\ 1 & x_n \end{bmatrix}, \qquad B = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ . \\ \varepsilon_n \end{bmatrix},$$

where ε is the vector of errors with zero means and the following matrix of variances

$$Var\ (\varepsilon) = \sigma^2 I,$$

and I is the (n x n) unit (identity) matrix, i.e.,

$$I = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

The matrix form (3-1) is used below in Example 1 and in the further discussion.

study

The estimates of parameters β_0 and β_1 are found as

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \tag{4}$$

$$\hat{\beta}_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2},$$

where $\bar{y} = n^{-1}\Sigma y_i$ and $\bar{x} = n^{-1}\Sigma x_i$.

The estimates (4) can be written in the matrix form as

$$\hat{B} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (X'X)^{-1} X'Y, \tag{4-1}$$

where X' is the transpose of matrix X, and $(X'X)^{-1}$ is the inverse of the matrix product of X' and X.

A more general case of model (3) is the so-called *weighted regression* when errors ε_i are still independent but have different variances σ^2_i (i= 1, 2, . . ., n). This model, which is called *weighted linear regression*, will be discussed in the following, so we need to write it here as

$$y_i(x_i) = \beta_0 x_0 + \beta_1 x_i + \varepsilon_i \tag{5}$$

where errors ε_i are independent normally distributed with mean 0 and different variances $\sigma^2_i(x_i)$ (i= 1, 2, . . ., n). In the matrix form, the model (5) can be written as

$$Y = XB + \varepsilon, \qquad (3\text{-}2)$$

where ε is the vector of independent errors with zero means and (in opposite to (3-1)) the following symmetric positively defined diagonal (n x n) matrix of variances Var (ε) =Σ,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & \sigma_n^2 \end{bmatrix}$$

The above variance matrix Σ can be represented in the form needed for the following consideration

$$\Sigma = \sigma^2 \begin{bmatrix} 1/w_1 & 0 & 0 & 0 \\ 0 & 1/w_2 & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & 1/w_2 \end{bmatrix},$$

where w_1, w_2. . . , w_n are the so-called weights. It is obvious, the greater variance, the smaller the respective weight is. The matrix of weights is defined as

$$\Sigma^{-1} = \sigma^{-2} \begin{bmatrix} w_1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 \\ . & . & . & . \\ 0 & 0 & 0 & w_2 \end{bmatrix}$$

The estimates of parameters β_0 and β_1 for model (5) can be found as

$$\hat{B} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y \qquad (4\text{-}2)$$

2.2 Classical Probability Papers

Without loss of generality, consider the Weibull probability paper estimation procedure, which is one most popular in life data analysis. Let the cumulative distribution function (CDF) of the Weibull time to failure (TTF) distribution $F(t)$ be given in the following form

$$F(t) = 1 - \exp\left(-\left(\frac{t}{\alpha}\right)^{\beta}\right) \qquad (6)$$

where t is TTF, α and β are the scale and shape parameters, respectively. Applying the logarithmic transformation twice, the above CDF is transformed to the following expression

$$\ln(-\ln(1-F(t))) = \beta \ln t - \beta \ln\alpha \qquad (6\text{-}1)$$

Introducing the following notation $y(t) = \ln(-\ln(1 - F(t))$, $\ln t = x$, $\beta_0 = \beta \ln(\alpha)$, Equation (6-1) takes on the simple linear response function form (2-1):

$$y(x)= \beta_0 x_0 + \beta_1 x_1 + \varepsilon \qquad (6\text{-}2)$$

It should be noted that there is no guarantee that the errors ε are independent and normally distributed with mean 0 and variance σ^2 anymore. Nevertheless, the simple linear regression technique is widely applied to Equation 6-1, which is known as the Weibull probability plotting. The corresponding procedure also includes estimation of CDF $F(t)$ using order statistic, which is illustrated in the framework of the following example.

Example 1. 100 identical components were put on a life test. The test data are Type II censored: the test was terminated at the time of the fifth failure. Failure times $t_{(i)}$ (in hours) of the 5 failed components were 11.96, 39.10, 71.52, 74.90, 123.14.

The traditional estimates $\hat{F}(t_{(i)})$ of CDF $F(t)$, used in the Weibull probability papers is given by the following formulae [1]:

$$\hat{F}(t_{(i)}) = \frac{i-0.5}{n} \qquad (7)$$

where $t_{(i)}$ (i = 1, 2, . . . , r; and $r \leq n$) are the ordered failure times. In our example r = 5 and n = 100.

Calculating these estimates for our data and applying double logarithmic transformation (6-1) results in the following table (vector) of observations y's

$$Y = \begin{bmatrix} -5.29581 \\ -4.19216 \\ -3.67625 \\ -3.33465 \\ -3.07816 \end{bmatrix}$$

The explanatory variable x_1 is obviously the logarithm of the failure times, so that our explanatory variable matrix X is evaluated as

$$X = \begin{bmatrix} 1 & 2.48196 \\ 1 & 3.66605 \\ 1 & 4.27004 \\ 1 & 4.31620. \\ 1 & 4.81332 \end{bmatrix}$$

Now we can find the estimates of the parameters β and $\beta_0 = \beta \ln(\alpha)$ using Equations (4) or (4-1) as $\hat{\beta} = 0.971$ (the estimate of the shape parameter) and $\hat{\beta}_0 = -7.712$, so that the estimate of the scale parameter is $\hat{\alpha} = 2809.852$.

At this point, it must be mentioned that the test data in this example are simulated from the Weibull distribution with the scale parameter $\alpha = 1000$ and the shape parameter $\beta = 1.5$. It is clear that the estimates obtained are rather biased.

3. BAYESIAN SIMPLE LINEAR REGRESSION AND BAYESIAN PROBABILITY PAPERS

3.1 Bayesian Interpretation of Classical Simple Linear Regression

Consider simple normal linear regression (3). In Bayesian context, it is assumed that the parameters of model β_0 , β_1 and $\log\sigma$ are uniformly and independently distributed, i.e.,

$$p(\beta_0 , \beta_1, \sigma) \propto 1/\sigma \tag{8}$$

Note that it is an extra assumption, i.e., the assumptions about the observations $y_i(x_i)$ are not changed.

Assumption (8) is a convenient form of the so-called, *noninformative prior distribution*.

It can be shown [2] that under the given assumptions, the conditional posterior probability density function for β_0 and β_1 has the *bivariate normal* form with mean ($\hat{\beta}_0$, $\hat{\beta}_1$), which are given by

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \tag{9}$$

$$\hat{\beta}_1 = \frac{\Sigma(x_i - \bar{x})(y_i - \bar{y})}{\Sigma(x_i - \bar{x})^2}$$

where $\bar{y} = n^{-1}\Sigma y_i$ and $\bar{x} = n^{-1}\Sigma x_i$. The above expressions for $\hat{\beta}_0$ and $\hat{\beta}_1$ are the easily recognizable classical least squares estimates (4) for the simple linear regression (3)

3.2 Including Prior Information about Model Parameters

In the framework of Bayesian linear regression analysis, the prior information can be added to one or several regression parameters. Let's begin with including prior information about a single regression parameter, say β_1. It is supposed that the information can be expressed as the normal distribution with known mean β_{1pr} and variance $\sigma^2_{\beta 1pr}$ [3], i.e.,

$$\beta_{1\,pr} \sim N(\beta_1, \sigma^2_{\beta 1pr})$$

Note that this prior distribution is similar to classical *assumptions* about observations $y_i(x_i)$ ($i = 1, 2, \ldots, n$), introduced in Section 1.

Based on this similarity, the prior information on parameter β_1 is interpreted as an additional (pseudo) "data point" in the regression data set, and the posterior point estimates are calculated using the same Equations (4) or (9). For the case considered, this "observed" value of y corresponds to $x_0 = 0$ and $x_1 = 1$.

Including prior information about a set of regression parameter is performed in the similar way. For example, the prior information about the other regression parameter, β_0 is included as a "data point" having the prior $\beta_{0\,pr} \sim N(\beta_0, \sigma^2_{\beta 0pr})$. This "observed" value of y corresponds to $x_0 = 1$ and $x_1 = 0$.

Because, for the time being, we consider the case of independent observations with equal variances, we are expand this assumption to the priors, i.e., it is assumed that the priors are independently and normally distributed with equal variances, i.e.,

$$\sigma^2_{\beta 0pr} = \sigma^2_{\beta 1pr} = \sigma^2 \qquad (10)$$

The following example illustrates the issues discussed in the given section.

Example 2. **The data from Example 1 are used. The prior information about the unknown parameters is incorporated as follows.**

Example 2.1

The prior shape parameter of the Weibull distribution $\beta_{\,pr} = 1.5$, and the prior scale parameter $\alpha_{\,pr} = 1000$. Note that we use the true values of the parameters of the Weibull distribution, from which the data were generated, so that to an extent, our prior information is ideal.

In terms of the regression model (6-2), parameter β_1 as an additional (pseudo) "data point" is 1.5 with corresponding $x_0 = 0$ and $x_1 = 1$. The parameter β_o as another additional point is $\beta\ \ln(\alpha) = -10.36$ with corresponding $x_0 = 1$ and $x_1 = 0$. The table (vector) of observations y' with these two new point now is

$$Y = \begin{vmatrix} -5.29581 \\ -4.19216 \\ -3.67625 \\ -3.33465 \\ -3.07816 \\ -10.3616 \\ 1.5000 \end{vmatrix}$$

The respective explanatory variable matrix X is now

$$X = \begin{vmatrix} 1 & 2.48196 \\ 1 & 3.66605 \\ 1 & 4.27004 \\ 1 & 4.31620 \\ 1 & 4.81332 \\ 1 & 0.00000 \\ 0 & 1.00000 \end{vmatrix}$$

As in Example 1, the estimates of the posterior estimates of parameters β and $\beta_0 = \beta \ln(\alpha)$ are calculated using Equations (4) or (4-1), which gives $\hat{\beta}_{1post} = 1.512$ (the estimate of the shape parameter) and $\hat{\beta}_0 =$ -9.915, so that the posterior estimate of the scale parameter is $\hat{\alpha}_{post} = 705.294$. See Figure 1 for a graphical interpretation of Example 2.

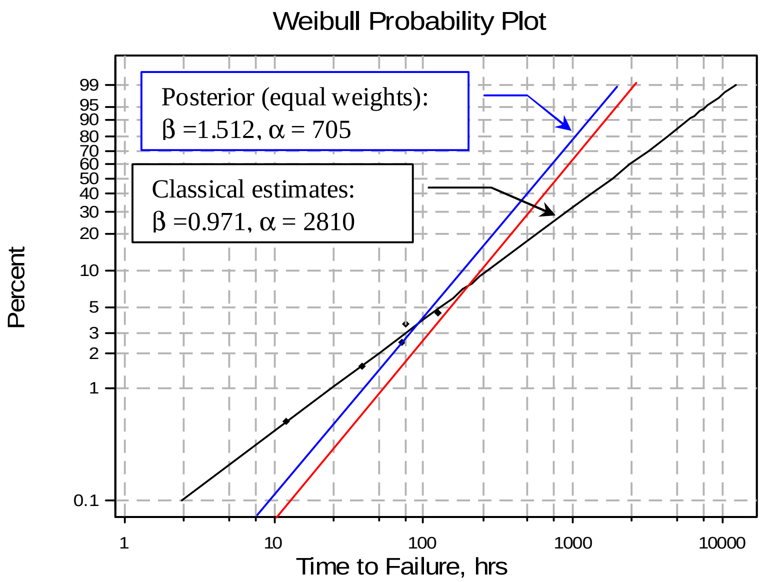

Figure 1. Weibull probability Plot of Prior & Posterior Distributions.

The above example represents a case when the pseudo data points are assumed having the same variances as the real data points (observations). In the framework of the weighted regression, this case corresponds to the equal weights situation. From Bayesian standpoint, it is the situation when the prior information has as much value as the real data.

Now consider the following two extreme cases.

Example 2.2

In the first case, the prior information has a negligible value. This case can be realized using very small weights (large variances) related to the pseudo data points on the Weibull plot. Let's consider the data of Example 2 with the following variance matrix:

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1000 \end{bmatrix}$$

Applying Equation (4-2) for the weighted linear regression, results in $\hat{\beta}_{1post}$ = 0.975 (the estimate of the shape parameter) and $\hat{\beta}_0$ = -7.726, so that the estimate of the scale parameter of the Weibull Distribution is $\hat{\alpha}_{post}$ = 2772.408. It is clear that, the posterior estimates are close to the classical ones (see Example 1). The result shows that the prior information does not play a significant role in the estimation.

Example 2.3

Now consider the opposite case. Let's select very large weights (very small variances) related to the pseudo data points on the Weibull plot. Let's consider the data of Example 2 with the following variance matrix:

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0.001 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.001 \end{bmatrix}$$

Applying the same Equation (4-2) for the weighted linear regression, gives the following estimates: $\hat{\beta}_{1post}$ = 1.509 (the estimate of the shape parameter) and $\hat{\beta}_0$ = -10.359, so that the estimate of the scale parameter is $\hat{\alpha}_{post}$ = 958.276. It is clear that, the posterior estimates are close to the true values of the Weibull distribution, from which the data were generated. This result reveals that in the considered case, the prior information does play a dominant role in estimation.

Example 2.4

Now consider the case close to real practical application of the given Bayesian procedure. One can assume that the data points on the Weibull probability plot have equal variances (standard deviations). Let's assume that they are equal to 1. A degree of belief in the prior information about the Weibull distribution parameters can be expressed in the same terms of standard deviations. It is reasonable to assume that the standard deviations related to the respective pseudo data points are, say, 3 times larger compared to the real data points, e.g., three times larger. Let's consider this case using the same example. The respective variance matrix for this case is

$$\Sigma = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 9.000 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 9.000 \end{bmatrix}$$

Applying the same Equation (4-2), gives the following estimates: the estimate of the shape parameter $\hat{\beta}_{1post}$ = 1.934, and the estimate of the scale parameter is $\hat{\alpha}$ = 1356.920. It is clear that, the posterior estimates are based on both types of data – the real observations and the prior information.

3.3 Including Prior Information about Reliability or Cumulative Distribution Function

It is clear that prior information about the reliability function or the CDF can be included in data set using a similar approach. That is, treating the prior knowledge about the reliability function at some given times as additional data points, and expressing the degree of belief in terms of standard deviations of prior reliability function estimates, which can be obtained using either expert opinion elicitation, or appropriate data (e.g., data on the predecessor product, alpha version testing etc.)

Table 1. Summary of Examples.

True values of the Weibull distribution parameters are: $\alpha = 1000, \beta = 1.5$.

Example	Estimation Procedure and Data	Estimate of α	Estimate of β
Example 1	Classical procedure. Real data only	2810	0.971
Example 2.1	Bayes' procedure with equal weights based on real data and ideal prior estimates, i.e., α_{pr} = 1000, β_{pr} = 1.5.	705	1.512
Example 2.2	Bayes' procedure with negligible prior information, i.e., prior estimates have very small weights (large variances)	2772	0.975
Example 2.3	Bayes' procedure with prior information strongly dominating real data, i.e., prior estimates have very large weights (small variances)	958	1.509
Example 2.4	Bayes' procedure with prior information comparable with real data information	1357	1.934

4. ACCELERATED FATIGUE TEST DATA

A sample of 12 induction-hardened steel ball joints underwent an accelerated fatigue life test with the following cycles to failure (in 1000s): 150, 170, 180, 200, 200, 215, 220, 220, 250, 260, 265, 300. Based on long-term history of such tests, the underlying life distribution was assumed to be lognormal. The CDF of the lognormal distribution with location parameter μ, and scale parameter σ is linearized using the following simple transformation:

$$\Phi^{-1}[F(t)] = \frac{1}{\sigma} ln(t) - \frac{\mu}{\sigma},$$

where $\Phi^{-1}[.]$ is the inverse of the standard normal cumulative distribution function. The classical least-square estimates of the location and scale parameters in this case are found to be 12.279 and 0.204, respectively. Historical data suggested that the scale parameter should be 0.160. Using the procedure similar to that outlined in Example 2.1 (equal weights), the Bayesian posterior estimate of the scale parameter was found to be 0.171. The analysis is graphically summarized in Figure 2.

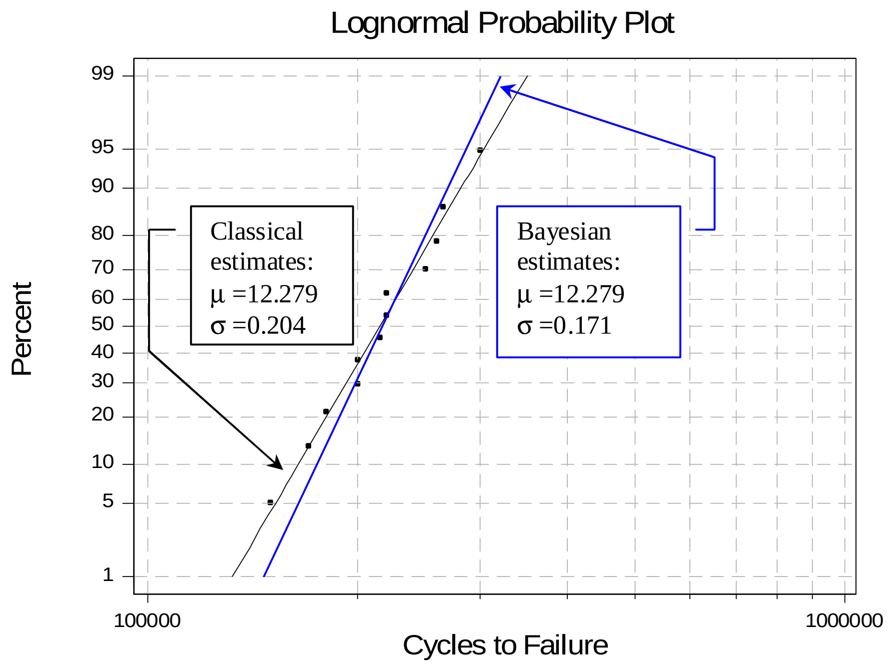

Figure 2. Lognormal Probability Plot of Ball Joint Fatigue Life Data.

study

REFERENCES

1. Lawless, J. F. *Statistical Models and Methods for Lifetime Data*, New York: Wiley; 2003.
2. Zelliner, A. *An Introduction to Bayesian Inference in Econometrics*, New York: Wiley; 1996.
3. Gelman, A. et al. *Bayesian Data Analysis*, London: Chapman & Hall; 1995.