Bayesian estimation of mixed Weibull distributions

Anduin E. Touw *

Boeing Space and Intelligence Systems, PO Box 92919 MC W-S13-G354, Los Angeles, CA 90009-2919, USA

A R T I C L E   I N F O
Article history:
Received 2 December 2007
Accepted 10 May 2008
Available online 21 May 2008

Keywords:
Mixed Weibull distribution
Bayesian analysis
Maximum likelihood estimation

A B S T R A C T
Estimation of mixed Weibull distribution by maximum likelihood estimation and other methods is frequently difficult due to unstable estimates arising from limited data. Bayesian techniques can stabilize these estimates through the priors, but there is no closed-form conjugate family for the Weibull distribution. This paper reduces the number of numeric integrations required for using Bayesian estimation on mixed Weibull situations from five to two, thus making it a more feasible approach to the typical user. It also examines the robustness of the Bayesian estimates under a variety of different prior distributions. It is found that Bayesian estimation can improve accuracy over the MLE for situations with low mixture ratios so long as the prior on the weak subpopulation’s characteristic life has an expected value less than or equal to the true characteristic life.

1. Introduction

Bayesian estimation has long been a popular technique for parameter estimation, particularly when data are sparse and other parameter estimation techniques can be unstable. By only using observed data to update the prior distribution on the parameters, the Bayesian method avoids the problem of wildly different parameter estimates fitting the data almost equally well, something commonly found in maximum likelihood estimation when estimates are close to the edge of the parameter boundary space [2]. However, if the prior distribution on the parameters is inaccurate, giving little or no probability to the actual values, it may take a large amount of data to update the prior accordingly. Thus, the accuracy of the Bayesian technique may depend heavily on the accuracy of the priors.

In reliability analysis, there are commonly situations where only limited failure data are available, but it is believed that there is some understanding of the likely range of values for the parameters. This situation is particularly likely to arise when mixed distribution situations are suspected. Under most reliability model assumptions, all failures in a population are considered to come from the same distribution. Each individual failure time is assumed to provide information about the likely failure times of all other devices in the population. However, from time-to-time, process variation or an unexpected event will lead to the development of a fragile sub-population while some percentage of the devices remained durable [1,2]. In this paper, estimation techniques for this situation are explored.

Ideally, when a mixed distribution is suspected, the weak sub-population could be identified through determination of root cause, inspecting, and sequestering of impacted devices. But many times, for practical reasons, when the overall population is a mixture of the fragile and durable populations; there may be no non-destructive way to identify the fragile devices. For example, a section of a wafer might be contaminated but the problem not discovered until after traceability to row and column is lost for individual die. It is unknown what percentage of the die was contaminated, and if the die is already packaged, inspection may not be possible. Or, for example, early “freak” failures might be observed in qualification or characterization life tests, after infant mortals were already removed. Root cause may be traced to uninspectable manufacturing defects. The numbers of and causes for the “freaks” may be ignored for characterizing wear-out for the overall population, but the customer and supplier may want to quantify the number of early failures that can be expected in a production run [1].

In these cases, statistical estimation methods must be used, either with or without root cause information, to quantify the reliability risk to the population [2]. The accuracy of these estimates may be critical to the decision making regarding release of the product, but estimation in these circumstances is difficult. There may be very few failures, with the majority of devices simply censored. There may be no way of determining what percentage of the censored devices is in the weak subpopulation, or even verifying that the fragile devices represent a sub-populations instead of an early sign that the entire population is weaker than expected. These data limitations can lead to instabilities in the model parameters when techniques such as maximum likelihood estimation (MLE) are used and create challenges for developing accurate estimation methods for
situations with potentially fragile subpopulations [7]. The lack of data suggests application of Bayesian techniques to stabilize the estimates.

Perlstein et al. [3] proposed a Bayesian method for estimating parameters for mixed exponential populations. By reparameterizing the likelihood equation, they were able to develop closed-form solutions for expected values of the parameters and the posterior distribution. However, information about the processing problem or other root cause information may not be consistent with a constant failure rate assumption for the fragile subpopulation. It may be known that the mechanism is either an infant mortality or wear-out situation. In both cases, the distribution of failures in the overall population will appear to have a decreasing failure rate, as the limited percent of the total population is weeded out, but assuming a constant failure rate on the fragile subpopulation can be misleading. In situations where the failure mechanism may not have a constant failure rate, a Weibull distribution is frequently assumed, and modifications to [3] would be required.

In the case of mixed Weibull distributions, Bayesian estimation is difficult because of the complexity of the likelihood equation. It has been shown that there is no conjugate family of continuous joint prior distributions on Weibull parameters (for a single Weibull distribution) [4,6]. Thus, there is no hope of a conjugate family for a mixed Weibull situation. With five parameters in the mixed Weibull formulation, numeric integration could be very time consuming, making Bayesian estimation impractical.

This paper combines the work on Bayesian estimation of single Weibulls with the work on Bayesian estimation of mixed exponentials in [3] to reduce the number of numeric integrations needed. Once establishing a practical means of applying Bayesian estimation to mixed Weibulls, it goes on to examine the robustness of the method under different prior distribution assumptions.

2. Bayesian formulation

It is assumed that a product comes from either a weak or strong population with a mixture ratio \( p \)\(^1\),

\[
R(t) = p \times \exp \left[ - \left( \frac{t}{\theta_W} \right)^{\beta_W} \right] + (1 - p) \times \exp \left[ - \left( \frac{t}{\beta_S} \right) \right]
\]

where \( 0 < p < 1 \) and \( \theta_W, \theta_S, \beta_W, \beta_S > 0 \)

As with all Bayesian analyses, assumptions must be made about the prior knowledge of the parameters. It is assumed that the mixture and characteristic life parameters are all independent of each other. It is also assumed here that the shape parameters are independent from the other variables. This is perhaps a more difficult assumption, but is common in the Bayesian reliability literature [4,5].

One commonly used set of priors on the Weibull is the inverted gamma distribution on \( \theta \) and the uniform distribution on \( \beta \) [4,5]. This is similar to the assumption made in [3], who assumed a \( \beta \) prior on the mixture parameter \( p \) and gamma priors on each of the failure rates. Since for the Weibull distribution, the standard parameterization uses characteristic life, the parameter of interest is similar to the inverse of the failure rate parameter in the exponential case. If \( \Lambda \) is a random variable having a gamma

distribution with parameters \((a, b^{-1})\), then \( \Theta = 1/\Lambda \) has an inverted gamma distribution with pdf

\[
g(\Theta) = \frac{b^a}{\Gamma(a)} \left( \frac{1}{\Theta} \right)^{a+1} \exp \left( - \frac{b}{\Theta} \right) \quad a, b > 0
\]

Thus, for consistency with the previous literature, it will be assumed that the characteristic life parameters \( \Theta_W, \Theta_S \) follow inverted gamma priors, the shape parameter \( \beta \) follows a uniform distribution, and the mixture parameter \( p \) follows a \( \beta \) prior.

Using a generalization of the likelihood parameterization proposed in [3],

\[
L(p, \theta_W, \beta_S, \beta_W, \beta_S) = \frac{\Gamma(a)}{ \Gamma(a_n) } \frac{1}{b^{a_n}} \prod_{m=1}^{a_n} \left[ \left( \frac{t_m}{\beta_S} \right)^{\beta_S} \right] \left( \frac{t_m}{\beta_W} \right)^{\beta_W} \left( \frac{t_m}{\beta_S} \right)^{\beta_S} \left( \frac{t_m}{\beta_W} \right)^{\beta_W}
\]

where \( \sum_{m=1}^{a_n} \) is the sum of the elements of the \( m \)th permutation of vectors of size \( \ell \) from \( \ell = \{1, \ldots, n\} \).

The joint posterior distribution has the form

\[
g(p, \theta_W, \beta_S, \theta_W, \beta_S | D) = \frac{g(p, \beta_S, \theta_W, \beta_W, \beta_S | D) \times L(p, \beta_S, \theta_W, \beta_W, \beta_S | D)}{C(n, u)}
\]

For simplicity, we will define a constant for this denominator:

\[
C(n, u) = \frac{\sum_{m=1}^{u} \sum_{m=1}^{n} \sum_{m=1}^{u} \left\{ \left( \frac{t_m}{\beta_S} \right)^{\beta_S} \left( \frac{t_m}{\beta_W} \right)^{\beta_W} \right\}}{\frac{\Gamma(a_n)}{ \Gamma(a) } \frac{b}{ \Gamma(a_n) } \prod_{m=1}^{a_n} \left[ \left( \frac{t_m}{\beta_S} \right)^{\beta_S} \left( \frac{t_m}{\beta_W} \right)^{\beta_W} \right]}
\]

One commonly used set of priors on the Weibull is the inverted gamma distribution on \( \theta \) and the uniform distribution on \( \beta \) [4,5]. This is similar to the assumption made in [3], who assumed a \( \beta \) prior on the mixture parameter \( p \) and gamma priors on each of the failure rates. Since for the Weibull distribution, the standard parameterization uses characteristic life, the parameter of interest is similar to the inverse of the failure rate parameter in the exponential case. If \( \Lambda \) is a random variable having a gamma

\[
E(\theta_W) = \frac{1}{C(n, u)} \sum_{m=1}^{n} \sum_{m=1}^{u} \left\{ \left( \frac{t_m}{\beta_S} \right)^{\beta_S} \left( \frac{t_m}{\beta_W} \right)^{\beta_W} \right\}
\]

\[
E(p) = \frac{1}{C(n, u)} \sum_{m=1}^{n} \sum_{m=1}^{u} \left\{ \left( \frac{t_m}{\beta_S} \right)^{\beta_S} \left( \frac{t_m}{\beta_W} \right)^{\beta_W} \right\}
\]

\[
E(\beta_S) = \frac{1}{C(n, u)} \sum_{m=1}^{n} \sum_{m=1}^{u} \left\{ \left( \frac{t_m}{\beta_S} \right)^{\beta_S} \left( \frac{t_m}{\beta_W} \right)^{\beta_W} \right\}
\]

\[
E(\beta_W) = \frac{1}{C(n, u)} \sum_{m=1}^{n} \sum_{m=1}^{u} \left\{ \left( \frac{t_m}{\beta_S} \right)^{\beta_S} \left( \frac{t_m}{\beta_W} \right)^{\beta_W} \right\}
\]

\[
E(\beta_S) = \frac{1}{C(n, u)} \sum_{m=1}^{n} \sum_{m=1}^{u} \left\{ \left( \frac{t_m}{\beta_S} \right)^{\beta_S} \left( \frac{t_m}{\beta_W} \right)^{\beta_W} \right\}
\]

\[
E(\beta_W) = \frac{1}{C(n, u)} \sum_{m=1}^{n} \sum_{m=1}^{u} \left\{ \left( \frac{t_m}{\beta_S} \right)^{\beta_S} \left( \frac{t_m}{\beta_W} \right)^{\beta_W} \right\}
\]

Note that this is a slightly different parameterization of the Weibull than is typically used: the shape parameter \( \beta \) is only applied to the random variable \( t \). This is to allow easier separation of the shape and characteristic life parameters in the integrations needed to calculate the posterior distribution. It is recommended in Martz and Waller [4].
greater sampling under difficult estimation situations, more in the simulation were: \{0.05, 0.10, 0.15, 0.20, 0.25, 0.50, 0.75, 0.80, 1.0\}. Values of the mixture ratio, $p$, were varied. In order to allow greater sampling under difficult estimation situations, more values of $p$ were examined close to 0 and 1. Values of $p$ included in the simulation were: (0.05, 0.10, 0.15, 0.20, 0.25, 0.50, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00). The characteristic life was taken to be close to 1 for all cases.\(^2\) Both a constant failure rate case ($\beta = 1$) and increasing failure rate ($\beta = 3$) case were examined.

For each case, 200 datasets, each containing 100 observations, were generated. The estimation method was applied to each dataset. Bias and random error in the parameter estimates were considered. In addition, the impact of these parameter estimates on the estimates of reliability was also examined. The bias was calculated as the average difference between the estimated value and the true value; the random error was calculated as the standard deviation of the difference.

### 4. Results for Bayesian method

In the baseline case, $p$ was taken to have prior distribution $\beta(0.5, 0.5)$ with likely values either close to 0 or 1, with less likelihood of middle values; $\theta$ was taken to have the prior $G(4, 3)$ an expected value of 1; $\gamma$ was given a uniform distribution prior between 0 and 6. These prior distributions are illustrated in Fig. 1.

As illustrated in Figs. 2 and 3, the Bayesian method performed very well. It had less bias and random error in its estimates of the mixture ratio than the MLE method did in almost all cases, particularly for small values of $p$. For example, for $p = 0.05$, the bias and random error in a were reduced by an order of magnitude; this translated to a reduction in the reliability estimate at time 10 from $-0.02$ to 0 in bias and 0.07 to 0.03 in random error. However, given that the prior gave high weight to very small and very large values of $p$, the strong performance of the Bayesian method under these conditions is not unexpected. The performance of the Bayesian method at $p = 0.5$, with a lower likelihood under the prior, is a harder test. Under this condition, the Bayesian method performed comparably to the mixture method. The bias in its estimate of $p$ was a little higher (0.06 versus 0.04 for mixture) but its random error was smaller (0.09 versus 0.15). The difference in the estimate of the reliability at time 10 was also very small, $(-0.05$ versus $-0.03$ bias and 0.09 versus 0.12 random error). These results suggest that the Bayesian method would be a better choice than the mixture MLE method, since when the prior is close to correct the Bayesian method produces more accurate estimates of the mixture ratio and even when the prior gives low likelihood the Bayesian method performs about as well.

Next, the priors were left as in the baseline case, but an increasing failure rate was considered. In this case, the shape parameter $\gamma$ was increased to 3. The characteristic life $\theta$ was also increased, to 1.6, in order to maintain similar numbers of observed failures at the censor time. The results for the reliability estimate at $t = 10$ are shown in Fig. 4. The method seemed very insensitive to the shape of the Weibull, with errors being virtually identical for the constant failure rate and increasing failure rate cases. This result was expected since a uniform prior was placed on the shape parameter.

Of greater concern was whether the Bayesian method would be robust under inaccurate priors on the mixture ratio $p$ and the characteristic life $\theta$. First a prior distribution that gave higher weight to low values of the mixture ratio was selected ($p \sim G(0.5, 2)$). As expected, the performance of the Bayesian method improved for low and even medium values of $p$, as illustrated in Fig. 5. The performance only got worse at $p \geq 0.8$. At $p = 1$, the $p \sim G(0.5, 2)$ case had bias of 0.1 and random error of 0.03 for $R(10)$ as compared with 0.05 and 0.03 for the baseline $p \sim G(0.5, 0.5)$ case. Similarly, as shown in Fig. 6, when a prior was selected to give higher likelihood to middle values ($p \sim G(2, 2)$), bias in the estimates of $p$ and the reliability at 10 got 2–3 times worse under this prior for true mixture ratios close to 0 or close to 1.

Even though the Bayesian method is sensitive to the choice of prior on $p$, it is still apparent to be more robust than the mixture MLE method. None of these increases in error from choices of prior on $p$ are comparable to the order of magnitude reduction in the errors in $p$ between the mixture MLE and any of the Bayesian cases for small mixture ratios.

The robustness of the Bayesian method with regards to the prior on the characteristic life $\theta$ as also examined. In the first case, shown in Fig. 7, the prior was adjusted to give an expected value of

\[^2\] For the case of increasing failure rate, the characteristics life was adjusted from 1 to maintain a consistent number of observed failures.
Fig. 1. Prior distributions on $p$ (left), $\theta$ (middle) and $\beta$ (right) for baseline case.

Fig. 2. Comparison of errors in $p$ for the mixture MLE estimates and Bayesian estimates (baseline case: $p \sim \beta(0.5, 0.5)$; $\theta \sim \Gamma(4,3)$; $\beta \sim U(0,6)$).

Fig. 3. Comparison of error in $R(10)$, 5 times the censor time, for MLE and Bayesian estimates (baseline case: $p \sim \beta(0.5, 0.5)$; $\theta \sim \Gamma(4,3)$; $\beta \sim U(0,6)$).
0.5 (instead of 1). In this case the probability of \( y \) being greater than or equal to 1 (the true value) was about 0.09. This case produced almost no change in the errors from the baseline case where the expected value of \( y \) was equal to 1. In the second case, shown in Fig. 8, the prior was adjusted to give an expected value of 4.3, with the chance of \( y \) being less than or equal to 1 being about 0.09 for symmetry. In this case, the Bayesian estimate produced much worst estimates of the reliability than did the baseline case or the MLE method. Reducing the expected value to 2 did not substantially correct the problem, as illustrated in Fig. 9, or did more than doubling the chance of \( y \) being less than or equal to 1, as illustrated in Fig. 10.

5. Discussion

These for the variants in the prior on \( \theta \) suggest that the Bayesian model is more sensitive to over estimating \( \theta \) with the prior than to under estimating \( \theta \) with the prior. This lack of symmetry seems at first counter-intuitive. The cause is related to the correlation between estimates of the mixture ratio \( p \) and estimates of the characteristic life \( \gamma \). Estimates of \( p \) and \( \gamma \) tend to be correlated with higher estimates of \( \gamma \) being associated with higher estimates of \( p \). The likelihood function for the mixed Weibull cases considered in this paper tend to be relatively flat, particularly with low true mixture ratios because there is only
limited observed failures. These flat likelihoods are heavily influenced by the prior distribution, as shown in Fig. 11. If the prior distribution on $\theta$ has a larger weight on high values, then it is likely that a high estimate of $p$ will also be selected. This result is shown by the smaller error in the reliability estimates when the true mixture ratio is close to 1 and the higher $\theta$ prior is selected in Figs. 8 and 9. However, when the true mixture ratio and characteristic life are small, the higher priors dominate the flat likelihood and create more error in the posterior.

On the other hand, if the prior on $\theta$ has a larger weight on small values, the estimates of $p$ will also tend to be small. But if the true mixture ratio $p$ is large, more failures will have been observed than suggested by the prior. If more failures are observed, the likelihood function is less flat and the priors have less influence on the estimates from the posterior distribution, as shown in Fig. 12.

This theory was further substantiated by looking at a case where the prior on $p$ gave preference to low values and the $\theta$ prior continued to over-predict $\theta$. As shown in Fig. 13, the errors in

Fig. 6. Comparison of errors in $R(10)$ between baseline Bayesian case ($p \sim \beta(0.5, 0.5), \theta \sim \Gamma(4, 3)$, and $\beta \sim U(0, 6)$) and alternate case where $p$ has high likelihood of being a middle value ($p \sim \beta(2, 2), \theta \sim \Gamma(4, 3), \beta \sim U(0,6)$).

Fig. 7. Comparison of errors in $R(10)$ between baseline Bayesian case ($p \sim \beta(0.5, 0.5), \theta \sim \Gamma(4, 3)$, and $\beta \sim U(0, 6)$) and alternate case where $\theta$ is likely to be small ($p \sim \beta(0.5, 0.5), \theta \sim \Gamma(2.5, 0.75), \beta \sim U(0,6)$). For the alternate case, the expected value of $\theta$ is 0.5 and the chance of $\theta$ being greater or equal to 1 is about 0.09. The dots represent the baseline case and the stars the alternate case. There is almost no difference in the results.
reliability were substantially smaller as long as the true $p$ was less than about 0.8.

6. Conclusions

As expected, the Bayesian method is sensitive to the choice of prior. However, it generally performs well, when compared with MLE, unless poor choices of prior are made on both the mixture ratio $p$ and the characteristic life $\theta$. It is particularly sensitive to situations where the prior distribution is selected to give large probability to high values of the characteristic life and mixture ratio simultaneously. It seems that it is better to select a prior with an expected value less than or equal to the true $\theta$ than to possibly overestimate $\theta$.

When a low mixture ratio is suspected, the Bayesian method can greatly reduce the errors in the estimates over a mixture MLE method so long as a good prior is selected for $\theta$. It is
suspected this is because at low mixture ratios, particularly few failures will be observed. The mixture MLE then has little data for its parameter estimates, while the Bayesian method gains information from its priors. If the prior distribution does not out-weigh the data into producing results with higher than truth mixture ratios and characteristics lives, the Bayesian method produces good results.

If a poor prior distribution is chosen with the assumption that the mixture ratio is very low (i.e. gives too much probability to mixture ratio values lower than truth), the Bayesian method will perform worse than the mixture MLE method. However, the increase in error under these circumstances is smaller than the decrease in error in circumstances where the mixture ratio really is very small. Based on these results, it is recommended that Bayesian methods be considered when a low mixture ratio is suspected.

Appendix A: Detailed calculations for posterior

Perlstein et al. [3] demonstrates that it is easier to manipulate the expanded form of the likelihood. First, some notation must be established.
Let \( P_m \) the sum of the elements of the \( m \)th permutation of vectors of size \( t \). So for example, if we let \( n = 4 \) there are \( \binom{4}{2} = 6 \) permutations of vectors of size 2. So for \( t = 2 \), \( m \) takes on values from 1 to 6:

\[
\begin{align*}
\sum_{i=1}^{4} t_m^1 &= t_1 + t_2 \\
\sum_{i=1}^{4} t_m^2 &= t_1 + t_3 \\
\sum_{i=1}^{4} t_m^3 &= t_1 + t_4 \\
\sum_{i=1}^{4} t_m^4 &= t_2 + t_3 \\
\sum_{i=1}^{4} t_m^5 &= t_2 + t_4 \\
\sum_{i=1}^{4} t_m^6 &= t_3 + t_4
\end{align*}
\]

Similarly, \( \prod t_m^\ell = \) the product of the elements of the \( m \)th permutation of vectors of size \( \ell \).

Using this notation, the likelihood equation can be rewritten as

\[
L(p, \theta_W, \beta_3, \beta_W, \beta_5)
= \sum_{\ell=0}^{n} \sum_{m=1}^{\ell} \sum_{u=1}^{n} \sum_{v=1}^{n} \left( \frac{u}{v} \right) p^{\ell+v}(1-p)^{n-\ell-u-v} \\
\times \left( \frac{\beta_3}{\theta_3} \right) \prod t_m^{\ell(0)} p^{\beta_5} \exp \left( -\frac{\sum t_m^{\ell(0)} p^{\beta_5} + u \cdot v t_m^{\ell(0)}}{\beta_5} \right) \right) (A.2)
\]
The posterior distribution is calculated using Bayes theorem

\[
g(p, \theta_W, \theta_5, \beta_W, \beta_5 | \text{Data}) = \frac{g(p, \theta_W, \theta_5, \beta_W, \beta_5 | \text{Data}) g(p, \theta_W, \theta_5, \beta_W, \beta_5)}{p^{(1-p)^{\alpha_W}} (1/\theta_W)^{\alpha_5 + 1} \exp(-b_W/\theta_W) (1/\theta_W)^{\alpha_5 + 1} \exp(-(b_5)/\theta_5) (1/\theta_5)^{\alpha_5 + 1} \exp(-(b_5)/\theta_5) (1/\theta_5)^{\alpha_5 + 1}}
\]

(A.3)

Let us start by considering the integral in the denominator.

\[
\int_0^\infty \frac{n!}{(n+\ell)!} (1-p)^{\alpha_W} (1/\theta_W)^{\alpha_5 + 1} \exp\left(-\frac{b_W}{\theta_W} + \sum_{m=1}^n \frac{u_m}{v_m} \right) \frac{1}{\theta_W} \prod_{m=1}^n \left( \frac{b_m}{\theta_W} \right)^{\beta_W - 1} \prod_{m=1}^n \left( \frac{b_m}{\theta_5} \right)^{\beta_5 - 1} \prod_{m=1}^n \left( \frac{u_m}{v_m} \right)^{\beta_5 - 1}
\]

(A.4)

greatly simplifies the problem and decreases the computational time.

For simplicity, we will define a constant for this denominator

\[
C(n, u) = \sum_{\ell=0}^n \sum_{m=1}^n \frac{u_m}{v_m} \Gamma(a_5) \Gamma(b_5) \prod_{m=1}^n \Gamma(t_m)^{\beta_5 - 1}
\]

(A.5)

Using the joint posterior distribution, the posterior expected values of the parameters and the reliability can be calculated.

\[
E(p) = \frac{1}{C(n, u)} \sum_{\ell=0}^n \sum_{m=1}^n \frac{u_m}{v_m} \frac{\Gamma(a_5 + 1) \Gamma(b_5)}{\Gamma(a_5 + b_5 + 1)} \times \Gamma(a_5) \prod_{m=1}^n \Gamma(t_m)^{\beta_5 - 1}
\]

(A.6)

\[
E(\theta_W) = \frac{1}{C(n, u)} \sum_{\ell=0}^n \sum_{m=1}^n \frac{u_m}{v_m} \frac{\Gamma(a_5 + 1) \Gamma(b_5)}{\Gamma(a_5 + b_5 + 1)} \times \Gamma(a_5) \prod_{m=1}^n \Gamma(t_m)^{\beta_5 - 1}
\]

(A.7)

\[
E(\theta_5) = \frac{1}{C(n, u)} \sum_{\ell=0}^n \sum_{m=1}^n \frac{u_m}{v_m} \frac{\Gamma(a_5 + 1) \Gamma(b_5)}{\Gamma(a_5 + b_5 + 1)} \times \Gamma(a_5 - 1) \prod_{m=1}^n \Gamma(t_m)^{\beta_5 - 1}
\]

(A.8)

\[
E(\beta_W) = \frac{1}{C(n, u)} \sum_{\ell=0}^n \sum_{m=1}^n \frac{u_m}{v_m} \frac{\Gamma(a_5 + 1) \Gamma(b_5)}{\Gamma(a_5 + b_5 + 1)} \times \Gamma(a_5) \prod_{m=1}^n \Gamma(t_m)^{\beta_5 - 1}
\]

(A.9)

where

\[
\alpha' = \alpha + \ell + v \quad \beta' = \beta + n + 1 + u \quad v
\]

\[
\alpha_5 = \alpha + \ell \quad \beta_5 = \beta_5 + \sum_{m=1}^n (t_m - 1) \beta_5 - 1 + (u - v) \beta_5
\]

\[
\alpha'_W = \alpha_W + \ell \quad \beta'_W = \beta_W + \sum_{m=1}^n (t_m - 1) \beta_5 - 1 + v \beta_5
\]

The remaining integrals with respect to \( \beta_5 \) and \( \beta_W \) do not have closed form solutions and must be numerically integrated. However, limiting the numeric integration to these two integrals, as opposed to numerically integrating over all five parameters,
\[ E(\beta_2) = \frac{1}{C(n,u)} \sum_{\ell=0}^{n} \sum_{m=1}^{u} \sum_{v=0}^{\ell} \left\{ \left( \frac{u}{v} \right) \Gamma(\alpha^+ \beta^+) / \Gamma(\alpha^+ + \beta^+) \right\} \times \Gamma(a_n^+) \times \Gamma(a_n^+ + \beta_n^+ - \ell - 1) \times \prod_{m=1}^{\ell} (\Gamma(\alpha^+ \beta^+) / \Gamma(\alpha^+ + \beta^+ - 1) \times \prod_{m=1}^{\ell} (\Gamma(\alpha^+ \beta^+) / \Gamma(\alpha^+ + \beta^+ - 1) \times \prod_{m=1}^{\ell} \left( \frac{u}{v} \right) \Gamma(\alpha^+ \beta^+) / \Gamma(\alpha^+ + \beta^+) \right\} \times \left( b_n^+ \right)^{\beta_n^+} \times \beta_n^+ \right\} \right\} \] (A.10)

References