University of Texas, MD Anderson Cancer Center

UT MD Anderson Cancer Center Department of Biostatistics Working Paper Series

Year 2006 *Paper* 11

Bayesian Aggregation Error?

Valen E. Johnson*

*U.T.M.D. Anderson Cancer Center, vejohnson@mdanderson.org

Bayesian Aggregation Error?

Valen E. Johnson

Abstract

Differences between inferences obtained from Bayesian reliability models using system-level data versus component-level data have recently motivated a call for a "basic restructuring of the Bayes procedure [1]." In this note, I explore the source of such differences and demonstrate that Bayesian models would be aberrant only if such differences did not exist.

Bayesian Aggregation Error?

Valen E. Johnson

Professor of Biostatistics

U.T. M.D. Anderson Cancer Center

Abstract

Differences between inferences obtained from Bayesian reliability models using system-level data versus component-level data have motivated a call for a "basic restructuring of the Bayes procedure [1]." In this note, I explore the source of such differences and demonstrate that Bayesian models would be aberrant only if such differences did not exist.

Keywords: Multi-level reliability model, Markov chain Monte Carlo, Bayesian inference

1 Introduction

Aggregation errors, or more appropriately *aggregation differences*, occur when estimates of system reliability obtained from Bayesian probability models differ depending upon the level at which data is incorporated into a model. For instance, incorporating component-level data into a Bayesian reliability model often results in a different estimate of system reliability than is obtained when the component-level failures are aggregated and then modeled as system failures. Bier [2] provided several theorems that characterize multi-level reliability models that avoid this phenomenon. Such models are said to possess *perfect aggregation* properties, and may be of some interest because they eliminate the requirement to record and model data at the component level.

In a related article [1], the phenomenon of aggregation error was heralded as a "failure of Bayes system reliability inference based on data with multi-level applicability," and it was suggested that aggregation error "represents a fundamental breakdown in the usual Bayesian methodology." In this note, I investigate the source of aggregation error in two simple examples. These examples highlight the point that inferential differences attributed to aggregation should be expected whenever information of different specificity is incorporated into multi-level reliability models. I also present a theorem that elucidates the relation between posterior distributions based aggregated and disaggregated data.

2 A beta-binomial example

Consider first a simple series system composed of two non-redundant, independent components with probabilities π_1 and π_2 of successfully functioning when operated. Suppose further that elicitation of expert opinion results in a prior beta density with parameters (α_1, β_1) for the first component, and a prior beta density with parameters (α_2, β_2) for the second component.

2.1 Aggregrated analysis

If the system is tested n times and successfully functions in x of the n trials, then the joint posterior distribution on the success probabilities π_1 and π_2 based on the aggregated data is proportional to

$$f(\pi_1, \pi_2 \,|\, x, n) \propto (\pi_1\pi_2)^x (1-\pi_1\pi_2)^{n-x} \pi_1^{\alpha_1-1}(1-\pi_1)^{\beta_1-1}\pi_2^{\alpha_2-1}(1-\pi_2)^{\beta_2-1}. \qquad (1)$$

Although this posterior distribution is not of standard form, its properties can be explored easily using Markov chain Monte Carlo (MCMC) methods (e.g., [3]) to obtain the induced posterior distribution on the system reliability, equal here to the product $\pi_1\pi_2$.

2.2 Disaggregated analysis

If success/failure data is available at the component level, it is possible to model the system more precisely. In our two component system, if the first component successfully functions in x_1 of n trials, and the second component functions in x_2 of n trials, then the joint posterior distribution on (π_1, π_2) that results from these data is proportional to

$$f(\pi_1, \pi_2 \,|\, x, n) \propto \pi_1^{x_1+\alpha_1-1}(1-\pi_1)^{n-x_1+\beta_1-1}\pi_2^{x_2+\alpha_2-1}(1-\pi_2)^{n-x_2+\beta_2-1}. \qquad (2)$$

For consistency with the aggregrated analysis, we require that the number of trials in which *at least* one of the components fails is x. As in the aggregated analysis, the posterior distribution on the system success probability $\pi_1\pi_2$ can be obtained by a transformation of variables and can be investigated numerically by examining posterior samples from an MCMC algorithm.

Bier [2] provides special conditions under which the joint posterior on the product $\pi_1\pi_2$ obtained from the aggregated and disaggregated analyses are the same. For general choices of (α_1, β_1) and (α_2, β_2), the posterior distributions do not coincide.

2.3 Numerical examples

For the aggregated analysis, suppose that $n = 10$ and $x = 2$, or that 2 out of 10 system tests are successful. Suppose further that Jeffreys' non-informative prior is assumed for both component probabilities; that is, $\alpha_1 = \alpha_2 = \beta_1 = \beta_2 = 0.5$. Then from (1), the posterior distribution on the probability that the system functions successfully can be determined numerically. In this case, the posterior mean of this probability is 0.20, and a plot of the posterior distribution for $\pi_1\pi_2$ based on these data appears in Figure 1 as the "aggregated" curve.

Now suppose that we wish to model the system more precisely and, through further investigation, learn that the first component failed in 8 of the 10 tests and the second component was successful in all 10 tests (i.e., $x_1 = 2$, $x_2 = 10$). How does this affect our inference? Based on the disaggregated data, the posterior mean that the system functions is now 0.21, and the curve labeled "1" in Figure 1 represents the posterior density on $\pi_1\pi_2$ based on this component-level data. Clearly, both the density and the posterior mean have changed now that more detailed data—and thus more detailed information—have been obtained. But is this really problematic?

To answer this question, suppose that instead of observing $x_1 = 2$ and $x_2 = 10$, we had instead observed 6 out of 10 successes for each component. That is, each component failed 4 times, with a maximum of one component failing on each system test. Should our inference regarding the successful functioning of the system be different now? Obviously it should be, since the posterior distributions on π_1 and π_2 now concentrate around 0.6, whereas they previously concentrated near 0.2 and 1.0. The posterior mean of the system failure rate in this case is 0.35, close to the maximum likelihood estimate of 0.36. To help visualize the difference, the actual posterior distribution for the system success probability for these data is also illustrated in Figure 1 and is represented by the curve labeled "2."

The two scenarios for disaggregated data demonstrate that very different inferences *should* be drawn for data that aggregate in exactly the same way. In other words, because both sets of disaggregated data yield the same aggregated data, it is clear from this example that any resolution of the phenomenon of aggregation error requires that the same the inferences be drawn for both sets of disaggregated data. But that is clearly absurd, and suggests that aggregation differences are a desirable feature of well-specified reliability models, not an indictment against them. Aggregation differences are also not unique to Bayes estimates: Similar differences are observed for the maximum likelihood estimates.

3 An exponential failure time example

My second example is borrowed from [1] and concerns the failure time distribution of a system that has two components arranged in series. Failure times for both components are assumed to follow independent exponential distributions. The prior distribution on the failure rate of the first component, λ_1, is assumed to be an exponential distribution with mean $\mu_1 = 0.01$/sec. The prior distribution for the failure rate of the second component, λ_2, is also assumed to be an exponential distribution, but with prior mean $\mu_2 = 0.001$/sec. Both components are operated for a period $t = 1000$ seconds. The number of failures observed for the first component, r_1, is 1, and the number of failures observed for the second component, r_2, is 2.

From simple probability calculus, it follows that the failure rate of the system is exponentially distributed with mean $\lambda = \lambda_1 + \lambda_2$. The joint posterior distribution on (λ_1, λ_2) based on the disaggregated data is proportional to

$$f(\lambda_1, \lambda_2 | r_1 = 1, r_2 = 2, t = 1000) \propto \lambda_1^1 \lambda_2^2 \exp[-1100\lambda_1 - 2000\lambda_2], \qquad (3)$$

while the joint posterior distribution based on the aggregrated data is proportional to

$$f(\lambda_1, \lambda_2 \,|\, r_1 + r_2 = 3, t = 1000) \propto (\lambda_1 + \lambda_2)^3 \exp[-1100\lambda_1 - 2000\lambda_2]. \qquad (4)$$

As noted in [1], the posterior mean of the failure rate based on the disaggregated data is $3.3E-3$, and the posterior mean of the system failure rate based on the aggregated data is $3.8E-3$.

Although we should expect aggregation differences in failure time data just as we should for failure data, it is illuminating to investigate the source of these differences. In this case, the prior mean of the system failure rate is .011, and the first component's contribution to this prior mean is 0.01. The second component's contribution is ten times smaller—.001. Despite these differences in prior expectations, twice as many failures occur at the second component than at the first component, and the observed failure rate at the first component is one-tenth that which is expected *a priori*. Clearly, these features of the disaggregated data were not what we expected to see when we specified our prior density.

But what about the aggregated data? That is, if we know only that 3 failures occurred, how would we assign the failures to individual components? The Bayesian paradigm provides a simple answer to this question. Given values of λ_1 and λ_2, the conditional probability that a failure occurred in component 1 (given that a failure has occurred) is equal to

$$\frac{\lambda_1}{\lambda_1 + \lambda_2}. \tag{5}$$

Averaging over the posterior distribution of λ_1 and λ_2 given in (4), it follows that the posterior distribution on the number of failures that occur at the first component has mean 2.2, and the posterior probabilities of observing 0, 1, 2, and 3 failures at the first component are $(0.08, 0.15, 0.27, 0.50)$, respectively. The probability of seeing 2 or 3 failures at the first component, given that 3 failures occurred, is equal to 0.77. In other words, it is likely that 2 or 3 of the failures occurred in the first component, and we implicitly update our prior belief about the value of λ_1 based on the prediction that the observed number of failures in the first component is more likely to be 2 or 3 than it is to be 0 or 1.

Somewhat more formally, an application of the law of total probability shows that the posterior density on (λ_1, λ_2) based on the aggregated data can be written as a weighted average of the posterior densities on (λ_1, λ_2), weighted with respect to the posterior probabilities that each configuration of disaggregated data occurred. In this weighting, $r_1 = 1$ gets only 15% weight.

More generally, the relation between posterior distributions based on aggregated and disaggregated data can be summarized by the following theorem.

Theorem. Let $\mathbf{y}$ denote a vector of (disaggregated) observations drawn from a sample space $\mathbf{Y}$ and suppose that $\mathbf{y}$ has a probability density function in a parametric family $\{f(\mathbf{y}|\theta), \theta \in \Theta \subset R^p\}$ defined with respect to a σ-finite measure μ. If $\mathbf{y}$ is continuous, then μ is Lebesgue measure. If $\mathbf{y}$ is discrete, then μ is the counting measure on $\mathbf{Y}$. Let $\pi(\theta)$ denote the prior density on θ and let $p_y(\theta|\mathbf{y})$ denote the posterior distribution on θ given $\mathbf{y}$. Let $\mathbf{x} = t(\mathbf{y})$ denote a (possibly vector-valued) function (aggregation) of the observation vector $\mathbf{y}$, and let $A = \{\mathbf{z} : \mathbf{z} \in \mathbf{Y},\ \mathbf{x} = t(\mathbf{z})\}$. Then the posterior distribution on θ given $\mathbf{x}$, say $q_x(\theta|\mathbf{x})$, can be expressed

$$q_x(\theta|\mathbf{x}) = \int_A p_z(\theta|\mathbf{z}) r(\mathbf{z}|\mathbf{x})\, d\mathbf{z}, \tag{6}$$

where

$$r(\mathbf{z}|\mathbf{x}) = \int_\Theta f(\mathbf{z}|\theta) q_x(\theta|\mathbf{x})\, d\theta. \tag{7}$$

Proof: Recognizing that $p_z(\theta|\mathbf{z}) = p_z(\theta|\mathbf{z}, \mathbf{x})$, the proof follows directly from the law of total probability (e.g., [4], page 37).

The weighting in (6) encapsulates the difference in aggregated and disaggregated estimates of the system failure rate. The aggregated estimate weights over all possible configurations of disaggregated data, with weights proportional to the posterior probability of each configuration. The disaggregated estimate assigns weight 1 to the configuration actually observed. If this configuration is relatively unlikely given the aggregated data and the assumed statistical model, the two estimates can be expected to differ sharply.

Interestingly, in this example if the observations are switched so that $r_1 = 2$ and

$r_2 = 1$, then the posterior mean of the system failure rate based on the disaggregated data becomes $3.7E-3$. In some sense, this configuration of the 3 failures is closest to the posterior probabilities of component failures based on the aggregated data, and so the Bayes estimate of the system failure rate based on this configuration of disaggregated data is very similar to the aggregated failure rate estimate. Of course, this estimate of the system failure rate also differs from the estimate obtained using the original configuration $r_1 = 1$, $r_2 = 2$, demonstrating once again that aggregated estimates cannot, in general, be consistent with all configurations of disaggregated data.

4 Summary

Aggregation error is not an error. Differences in posterior inferences based on aggregated data and disaggregated data occur because the information content of data—as reflected through likelihood functions—is generally different for aggregated and disaggregated data. Aggregation differences thereby represent a natural feature of properly specified statistical models. Aggregation errors should not be interpreted as evidence of model inadequacy.

References

[1] Philipson, L. (1996), "The Failure of Bayes System Reliability Inference Based on Data with Multi-level Applicability," *IEEE Transactions on Reliability*, bf 45(1), 66-69.

[2] Bier, V.M. (1995), "On the Concept of Perfect Aggregation in Bayesian Estimation," *Reliability Engineering and System Safety*, **46**, 271-281,

[3] Gilks, W.R., Richardson, S., and Spiegelhalter, D.J. (1996), *Markov Chain Monte Carlo in Practice,* Chapman & Hall: London.

[4] Walpole, R.E. and Myers, R.H. (1989), *Probability and Statistics for Engineers and Scientists*, Macmillan Publishers: New York.

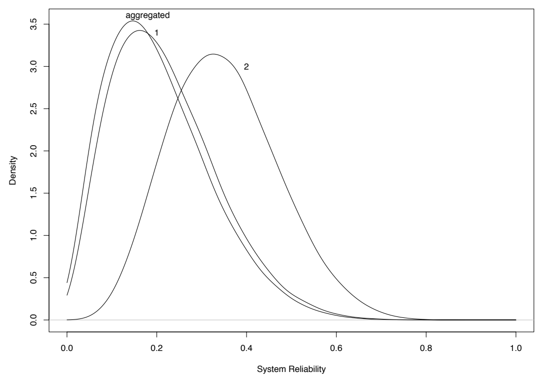

Figure 1: This plot depicts posterior densities for system failure probability based on aggregated and disaggregated data. The curve labeled "1" is based on the observation of 2 out of 10 and and 10 out of 10 successes for the first and second components, respectively, while the curve labeled "2" is based on the observation of 6 out of 10 successes for both components.