Finding a Minimally Informative Dirichlet Prior Using Least Squares

ANS PSA 2011

Dana Kelly
Corwin Atwood

March 2011

This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint should not be cited or reproduced without permission of the author. This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, or any of their employees, makes any warranty, expressed or implied, or assumes any legal liability or responsibility for any third party’s use, or the results of such use, of any information, apparatus, product or process disclosed in this report, or represents that its use by such third party would not infringe privately owned rights. The views expressed in this paper are not necessarily those of the United States Government or the sponsoring agency.
FINDING A MINIMALLY INFORMATIVE DIRICHLET PRIOR USING LEAST SQUARES

Dana Kelly
Idaho National Laboratory
P. O. Box 1625, Idaho Falls, ID 83415
Dana.Kelly@inl.gov

Corwin Atwood
Statwood Consulting
2905 Covington Rd., Silver Spring, MD 20910
cory@statwoodconsulting.com

ABSTRACT

Abstract In a Bayesian framework, the Dirichlet distribution is the conjugate distribution to the multinomial likelihood function, and so the analyst is required to develop a Dirichlet prior that incorporates available information. However, as it is a multiparameter distribution, choosing the Dirichlet parameters is less straightforward than choosing a prior distribution for a single parameter, such as p in the binomial distribution. In particular, one may wish to incorporate limited information into the prior, resulting in a minimally informative prior distribution that is responsive to updates with sparse data. In the case of Poisson \( \lambda \), the principle of maximum entropy can be employed to obtain a so-called constrained noninformative prior. However, even in the case of \( p \), such a distribution cannot be written down in the form of a standard distribution (e.g., beta, gamma), and so a beta distribution is used as an approximation in the case of \( p \). In the case of the multinomial model with parametric constraints, the approach of maximum entropy does not appear tractable. This paper presents an alternative approach, based on constrained minimization of a least-squares objective function, which leads to a minimally informative Dirichlet prior distribution. The alpha-factor model for common-cause failure, which is widely used in the United States, is the motivation for this approach, and is used to illustrate the method. In this approach to modeling common-cause failure, the alpha-factors, which are the parameters in the underlying multinomial model for common-cause failure, must be estimated from data that are often quite sparse, because common-cause failures tend to be rare, especially failures of more than two or three components, and so a prior distribution that is responsive to updates with sparse data is needed.

Key Words: Keywords: alpha-factor, common-cause failures, constrained noninformative prior, Jeffreys prior

1 INTRODUCTION AND BACKGROUND

This work was motivated by the need to specify a prior distribution for common-cause failure parameters in a model of failure of a support system. In such a model, where the common-cause component group size is large (e.g., four or more redundant pumps), failure of the system is dominated by common-cause failure, and because the components of the system are highly reliable, failures are rare, and so estimates of the common-cause parameters are dominated by the prior distribution. In a Bayesian framework, the Dirichlet distribution is the conjugate distribution to the multinomial likelihood function, and so the analyst is required to
develop a Dirichlet prior that incorporates available information. However, as it is a multiparameter distribution, choosing the Dirichlet parameters is less straightforward than choosing a prior distribution for a single parameter, such as $p$ in the binomial distribution. In particular, one may wish to incorporate limited information into the prior, resulting in a minimally informative prior distribution that is responsive to updates with sparse data.

1.1 Review of Constrained Noninformative Prior

To provide some background, we first review the case of $p$, the unknown parameter in the binomial distribution, which is used as a stochastic model for failure to change state. Bayesian inference is used in PRA to estimate the values of parameters in stochastic models, as described in [1]. Two possible noninformative prior distributions for $p$ are the uniform distribution over the interval [0, 1], and the Jeffreys prior, which is a beta(0.5, 0.5) distribution, as described in [1]. With observed data in the form of $x$ failures in $n$ demands, the posterior means for $p$ are $(x + 1)/(n + 2)$ with the uniform distribution, and $(x + 0.5)/(n + 1)$ with the Jeffreys prior. In both cases, if $p$ is small and the data are sparse, the posterior mean may be overly conservative, as it is pulled considerably toward the prior mean of 0.5.

To alleviate this problem, [2] suggested a generalization of the Jeffreys prior that is constrained to have a specified mean value ($< 0.5$ typically), but which is as noninformative as possible otherwise. The resulting distribution cannot be expressed in the form of a standard distribution (e.g., beta, gamma), but can be approximated well by a beta distribution with parameters $a$ and $b$, as discussed in [2]. The beta distribution that approximates the CNI prior has its first (second) parameter approximately equal to 0.5 when the specified prior mean of $p$ is $< (>) 0.5$. Note that in the work below we are using a simplified approximation of the CNI prior, which uses 0.5 regardless of the exact assumed mean. The same method could be used with the values from Table 1 of [2] instead of 0.5. However, because the prior means are not known with certainty, and we are ignoring uncertainty in these constraints, as did [2], using the value from Table 1 of [2] instead of the approximate value of 0.5 only appears to be more precise. With the value of 0.5 as one of the parameters, the remaining beta distribution parameter is then obtained using the specified mean, $E(p)$, as follows. If $E(p) < 0.5$, the first parameter is 0.5 and the second parameter is

$$b = \frac{0.5[1 - E(p)]}{E(p)} \tag{1}$$

If $E(p) > 0.5$, the second parameter is 0.5 and the first parameter is

$$a = \frac{0.5E(p)}{[1 - E(p)]} \tag{2}$$

The posterior mean obtained by updating the approximate CNI prior with $x$ failures in $n$ demands is then $(x + a)/(n + a + b)$. 


1.2 Multinomial Model for CCF and Dirichlet Prior Distribution

The most commonly encountered aleatory model for CCF in the U.S. is the multinomial distribution. The observed random variable is the vector of failure counts for a common-cause component group (CCCG) of size \( m, n = (n_1, n_2, ..., n_m) \). This model has a vector of unknown parameters, known as alpha-factors: \( \alpha = (\alpha_1, \alpha_2, ..., \alpha_m) \), with \( \alpha_k \) representing the probability that a specific group of \( k \) components in a CCCG fail due to a shared cause. In equation form, the density function can be written as

\[
f(n | \alpha) \sim \prod_{k=1}^{m} \alpha_k^{n_k}
\]

The alpha-factors must sum to unity, a constraint we will employ later.

The conjugate prior to the multinomial model is the Dirichlet distribution, a multiparameter extension of the beta distribution. Analogous to the single-parameter case, where the beta distribution was placed on \( p \), the unknown parameter of the binomial aleatory model, the Dirichlet distribution represents uncertainty in the vector of alpha-factors. The Dirichlet density function is given by

\[
\pi_0 (\alpha | \theta) \sim \prod_{k=1}^{m} \alpha_k^{\theta_k-1}
\]

The Dirichlet distribution has a vector of parameters, \( \theta \), of size \( m \). The marginal distribution of \( \alpha_k \) is a beta distribution with parameters \( \theta_k \) and \( \theta_t - \theta_k \), where

\[
\theta_t = \sum_{i=1}^{m} \theta_i
\]

Below we will also make use of the marginal mean and variance of \( \alpha_k \), which are given by

\[
E(\alpha_k) = \frac{\theta_k}{\theta_t}
\]

and

\[
Var(\alpha_k) = \frac{\theta_k (\theta_t - \theta_k)}{\theta_t^2 (\theta_t + 1)}
\]

Because of its conjugate nature, the posterior distribution of the alpha-factors will be a Dirichlet distribution, with parameters \( \theta + n \). Thus, each prior Dirichlet parameter, \( \theta_k \), is updated to \( \theta_k + n_k \), and \( \theta_t \) can be interpreted as a prior number of pseudo-events. The larger \( \theta_t \) is, the more prior pseudo-events have been observed and the more informative is the prior.
1.2.1 Noninformative Dirichlet Priors

Most common PRA formulas used to estimate alpha-factor values are either maximum likelihood estimates or posterior means from updating a noninformative Dirichlet prior. The maximum likelihood estimate is

\[ \hat{\alpha}_k = \frac{n_k}{\sum n_i} \]  

(8)

The noninformative prior distribution most commonly used in PRA for alpha-factors is the uniform Dirichlet distribution, with \( \theta_k = 1 \) for all \( k \). With this prior, the posterior mean of \( \alpha_k \) becomes

\[ \frac{n_k + 1}{\sum_{j=1}^{m} n_j + m} \]  

(9)

For CCCGs of three or more components, the failure data are typically quite sparse, meaning that \( n_k \) is often 0 for \( k > 2 \). In such cases, the posterior mean from updating the noninformative Dirichlet prior can be overly conservative with respect to the mean obtain using an informative prior, especially in applications in which CCF of all components in the CCCG is the dominant contributor to the desired figure of merit in the PRA. As discussed above, it was such a problem that led to the development of the constrained noninformative (CNI) prior for \( p \) by [2]. As pointed out by [1], the CNI prior is a type of maximum entropy prior, and relies upon a definition of entropy that requires the analyst to specify a “natural” noninformative prior distribution. [2] took the Jeffreys prior as the natural noninformative prior in the single-parameter case. For the analogous multiparameter case considered here for the alpha-factors, the approach of maximizing entropy appears somewhat intractable. However, the idea of having a prior distribution that incorporates available information without overly influencing the resulting posterior in updates with sparse data is attractive. We will refer to such a prior as minimally informative.

2 DEVELOPING A MINIMALLY INFORMATIVE DIRICHLET PRIOR

Our goal then is to specify a Dirichlet prior distribution for the alpha-factors that incorporates past data or engineering judgment, and which is specified in terms of expected values of alpha-factors, while avoiding a prior that is overly informative (i.e., too narrow). An overly informative prior is undesirable because it could have undue influence on the posterior distribution in updates with relatively sparse data.

2.1 Using Information about \( \alpha_1 \) Only

A first approach to specifying such a minimally informative prior proceeds as follows. Let's say the analyst is confident that \( E(\alpha_1) \) is near unity (because multiple failures are rare in well maintained engineered systems, we expect that \( \alpha_1 \) will be near unity), but cannot provide detailed
information about the relative values of the higher-order alpha-factors. Marginally, each $\alpha_k$ is analogous to $p$, and thus a CNI marginal distribution for $\alpha_k$ could be approximated with a beta distribution, as described above. For example, suppose that there are four alpha-factors in all, $m = 4$. If $E(\alpha_1) = 0.95$, the approximate CNI prior for $\alpha_1$ is beta(9.5, 0.5). This corresponds to a total of 10 prior events. We do not have other constraints except that the means of the alpha-factors sum to 1. With $E(\alpha_1) = 0.95$, this leaves 0.05 to be divided equally among the three remaining alpha-factors. An appeal to the symmetry of the constraints with respect to $\alpha_2$ through $\alpha_4$ shows that $\alpha_2$ through $\alpha_4$ should be treated identically, each with mean equal to 0.05. The Dirichlet prior that gives the above beta distribution for $\alpha_1$ and the above means for the remaining parameters is Dirichlet(9.5, 0.1667, 0.1667, 0.1667). The resulting marginal distribution for each $\alpha_k$, $k = 2, ..., 4$, is beta(0.1667, 9.8333). This is equivalent to 10 pseudo-event counts.

2.2 Using Information about Higher-Order Alpha Factors

We can extend this approach to cases where we have more prior information with which to estimate the means of the higher-order alpha-factors. Thus we might be able to specify a mean value for each $\alpha_k$: $E(\alpha_k) = \mu_{spec,k}$. These will be taken as the mean values for the marginal beta distributions. To achieve the requirement of a minimally informative prior, we take the marginal beta distribution for $\alpha_k$ to be the approximate CNI prior described above.

Thus, independent of the constraint that the alpha factors sum to unity, which makes the alpha-factors dependent, each $\alpha_k$ will have a marginal beta distribution with specified marginal means and variances given in terms of the parameters of the approximate CNI beta prior distributions by

$$\mu_{spec,k} = \frac{a_k}{a_k + b_k}$$

(10)

and

$$\sigma_{spec,k}^2 = \frac{\mu_{spec,k} \left(1 - \mu_{spec,k}\right)}{a_k + b_k + 1}$$

(11)

This gives $2m$ constraints in a CCCG of size $m$, but there are only $m$ Dirichlet parameters, so the problem is over-specified. However, we can use least-squares optimization to find a Dirichlet distribution whose marginal means and variances are close to the values specified. This forces the alpha-factors to have (nearly) the specified means and guarantees that they sum to unity. Under those constraints it makes the distribution minimally informative, in the sense that the variances of the individual alpha-factors are close to those of the marginal CNI priors.

The objective function to be minimized is

$$f(\theta) = \sum_{i=1}^{m} \left(\mu_{spec,i} - \mu_{est,i}\right)^2 + \left(\sigma_{spec,i}^2 - \sigma_{est,i}^2\right)^2$$

(12)
where $\mu_{est}$ and $\sigma_{est}^2$ are given by Eqs. 8 and 9. The desired Dirichlet parameters are those that minimize Eq. 12, subject to the additional constraint that each $\theta_k > 0$. Numerical optimization can be employed via a variety of software packages; the R Package [3] was used for the calculations which follow. Note that one could use standard deviations instead of variances in the objective function, and weights could be used to emphasize some means and variances more than others. These are straightforward extensions to the approach we have taken here. For the example considered below, using standard deviations instead of variances in the objective function did not change the results significantly.

### 3 NUMERICAL EXAMPLE

Consider a CCCG size of four, so in the alpha-factor model for CCF we need estimates for $\alpha_1$ through $\alpha_4$. We have information from past experience that suggests $\alpha_1$ is near unity, and that $\alpha_2$ through $\alpha_4$ are much smaller. For this example, we will specify the vector of alpha-factor means as $\mu_{spec} = (0.95, 0.03, 0.015, 0.005)$. The parameters of the beta distributions that approximate the marginal CNI priors are $a_{prior} = (9.5, 0.5, 0.5, 0.5)$ and $b_{prior} = (0.5, 16.17, 32.83, 99.50)$. We will take $a_{prior}$ as the vector of initial values for the Dirichlet parameter vector $\theta$ in the numerical optimization.

The vector of Dirichlet parameters that minimizes Eq. 12 is found to be $\theta = (9.52, 0.30, 0.15, 0.05)$. The prior marginal means of the alpha-factors with these parameters are $(0.95, 0.03, 0.015, 0.005)$, as desired. The marginal standard deviations of the alpha-factors are close to those of the approximate marginal CNI priors with the exception of $\alpha_4$, where the marginal standard deviation is 0.02, much larger than the desired standard deviation of 0.007. In this example, the information in the minimally informative Dirichlet prior with the means of all alpha-factors specified is equivalent to about 10 events, a relatively small amount of information, but significantly more than a noninformative Dirichlet prior, which would be equivalent to 4 events.

#### 3.1 Comparison of Updates

For this example, let $n = (35, 1, 0, 0)$ be the vector of observed failure counts. Updating our minimally informative Dirichlet prior with this vector of failure counts gives a Dirichlet posterior with parameters $(44.52, 1.30, 0.15, 0.05)$. The posterior means of the alpha-factors are $(0.97, 0.03, 0.003, 0.001)$. For comparison, updating a uniform Dirichlet prior (all parameters equal to 1) would give $(0.90, 0.05, 0.025, 0.025)$ as the posterior alpha-factor means.

Uncertainty in the alpha-factors can be summarized by 90% credible intervals. Table I shows the 90% posterior credible intervals that result from updating the two priors we have considered with the example vector of failure counts. Note that the extremely small 5th percentiles obtained with the minimally informative prior should not be taken literally, as they are an artifact of the posterior Dirichlet parameters $\theta_3$ and $\theta_4$ being $<< 1$. This is analogous to what happens when using a beta distribution for binomial $p$, with the first parameter of the beta distribution $<< 1$.

<table>
<thead>
<tr>
<th>Prior</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$\alpha_3$</th>
<th>$\alpha_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Minimally</td>
<td>$(0.92, 0.003)$</td>
<td>$(8E-11, 4E-27)$</td>
<td>$(8E-11, 4E-27)$</td>
<td>$(8E-11, 4E-27)$</td>
</tr>
</tbody>
</table>
4 CONCLUSIONS

This paper has presented an approach to developing a Dirichlet prior that incorporates information on the marginal mean values of the alpha factors of the conjugate multinomial aleatory model for CCF, but which is otherwise minimally informative. Because the notion of maximum entropy does not appear tractable in the case of a multinomial likelihood, we adopted a pragmatic approach based on least-squares optimization. The marginal means and variances of the corresponding marginal CNI priors for the alpha factors are used in the objective function, and this function is optimized numerically to give the least-squares estimates of the parameters of the Dirichlet prior distribution.

An illustrative example of this approach was presented. Because in nuclear plant applications failures of more than one component in a CCCG are rare, the marginal mean of \( \alpha_1 \) is near unity, making \( \theta_1 \) the dominant parameter in the Dirichlet prior. This paper has presented a pragmatic solution to the problem of finding a minimally informative Dirichlet prior with specified marginal means, but it has not solved the difficult problem of obtaining a robust posterior lower bound when the specified prior mean for a cell is small and the observed data contain no observations in that cell. Without actual observations of multiple failures, the posterior results are unavoidably influenced strongly by the prior distribution.

5 ACKNOWLEDGMENTS

The authors would like to acknowledge the many helpful discussions we have had with John Schroeder of INL over the years. It was John's insistent questioning that led us to develop this approach.

This manuscript has been authored by Battelle Energy Alliance, LLC under Contract No. DE-AC07-05ID14517 with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a nonexclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.

6 REFERENCES