SCHOOL OF OPERATIONS RESEARCH
AND INDUSTRIAL ENGINEERING
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NY 14853-3801

TECHNICAL REPORT NO. 1188

March 1997

**Assessing Multivariate Nonnormality
Using Univariate Distributions¹**

by

Elizabeth H. Slate

¹This research was supported by a grant from the National Science Foundation.

Assessing multivariate nonnormality using univariate distributions

Elizabeth H. Slate

School of Operations Research
Cornell University, Ithaca, NY 14853

Summary

As computational methods improve and enable more complicated models to be investigated, the need increases for simple diagnostics that can aid the exploration of multivariate likelihood functions and posterior distributions. This paper proposes diagnostics that can indicate regions where multivariate distributions are poorly behaved in the sense that they are far from normal. The measure of nonnormality developed by Slate (1994) for univariate distributions is extended to multivariate parametric models by application to univariate marginal and conditional distributions. The approach is illustrated for univariate conditional distributions using dynamic graphics in XLISP-STAT (Tierney, 1990). Examples show that the exploratory look at “slices” through the multidimensional distribution provided by the software yields considerable insight into the multivariate shape.

Some key words: Bayesian inference; multivariate likelihood function; multivariate posterior distribution; nonnormality; reparameterization; tail probability; third derivative

1 Introduction

The recent developments in computationally intensive techniques for fitting statistical models (such as Monte Carlo Markov chain methods) and rapid advances in computing speed are enabling statisticians to fit increasingly more complicated statistical models. This trend toward more complicated, high dimensional parametric models is further fueled by the extensive data sets that are becoming ever more routinely available (as from manufacturing and retail industries, for example). Understanding the behavior of statistical models, however, is clearly more difficult as the parameter dimension grows. It is hard to anticipate regions where the likelihood function or posterior distribution will be poorly behaved and cause computational or inferential techniques to fail. Indeed, even recognizing that these failures have occurred can be problematic.

A relatively simple example, which will be studied in greater detail in Section 4, is the growth data for orange tree three given in Draper and Smith (1981, p. 524, exercise N). The trunk circumference was measured at seven times within a five-year period. One model proposed for these data is the Gompertz nonlinear regression model, which stipulates that the circumference measurements, y_i , are independent normal random variables with variance σ^2 and means $\mu_i = \theta^1 \exp(-e^{\theta^2 - \theta^3 x_i})$, where x_i is the time of the i th measurement. Under what circumstances is the distribution of $\theta = (\theta^1, \theta^2, \theta^3)$ near normal? When is inference based on joint normality of θ unreliable, and what regions of the parameter space are most problematic? These questions are already difficult to answer for a three-dimensional parameter space and the difficulty will increase as the dimensionality of the problem grows. This observation highlights the need for diagnostics that can facilitate the exploration of multivariate likelihood functions and distributions.

For both computational and inferential purposes, “well-behaved” likelihood functions and posterior distributions are typically those that are close to normal. Thus the diagnostics proposed in this paper are measures of nonnormality and are cali-

brated according to the reliability of normality-based inferences. More specifically, the diagnostics indicate when inference based on asymptotic normality may be unreliable, and, consequently, when posterior distributions (or likelihood functions) may be problematic.

For a scalar parameter θ with log-posterior distribution $\ell(\theta)$ and mode $\tilde{\theta}$, the modal normal approximation to the posterior distribution is the normal distribution with mean $\tilde{\theta}$ and precision $\{-\ell''(\tilde{\theta})\}$. Slate (1994) considers a nonnormality diagnostic based on the third derivative defined by

$$F = \ell'''(\tilde{\theta}) \{-\ell''(\tilde{\theta})\}^{-3/2},$$

which arises as the first term beyond the modal normal approximation in a Taylor expansion of the log-posterior density. This measure has also been proposed by Hills (1989) and Albert (1989) in the Bayesian context and by Sprott (1973) and Anscombe (1964) for indicating departures from normality in the likelihood (replace ℓ by the log likelihood and $\tilde{\theta}$ by the maximum likelihood estimate).

Slate (1994) uses a tail probability approximation to show that the tail regions with probability 0.05 under the modal normal approximation have coverage between 0.025 and 0.10 when $|F| \leq 0.38$. This cutoff of 0.38 provides a reference for determining when the diagnostic $|F|$ indicates that inference based on the normal approximation can be misleading. By considering tail regions of arbitrary probability, this calibration of $|F|$ can be extended to a desired precision.

This paper extends the measure $|F|$ to univariate marginal and conditional distributions to aid in exploring a multidimensional posterior distribution or likelihood function by indicating the degree of nonnormality. Section 2 shows that, to the same order of approximation, the measure $|F|$ is equivalent to a linear combination of the first and third cumulants. The appropriate linear combination of approximate values of these cumulants is used to assess nonnormality of marginal distributions. For univariate conditional distributions, discussed in Section 3, the cumulant-based diagnostic of Section 2 provides the same information as the measure $|F|$. Section 3 also

describes dynamic graphics software for exploring the conditional posterior “slices.” The dynamic graphics software is illustrated in Section 4, where a number of examples are presented. Additional discussion is in Section 5.

Notation. Let $Y = (Y_1, \dots, Y_n)$ be an observed random vector that has a probability distribution depending on an unknown m -dimensional parameter $\theta = (\theta^1, \dots, \theta^m)$. Let the posterior density for θ be $f_\theta(\theta) \propto \exp\{\ell(\theta)\}$, so that $\ell(\theta)$ is the log-posterior density. Note that the dependence on Y has been suppressed in this notation. Furthermore, let $\tilde{\theta} = (\tilde{\theta}^1, \dots, \tilde{\theta}^m)$ be the joint posterior mode, and let $\ell_i = \frac{\partial \ell(\theta)}{\partial \theta^i}$, $\ell_{ij} = \frac{\partial^2 \ell(\theta)}{\partial \theta^i \partial \theta^j}$, etc., $I = (I_{ij}) = (-\ell_{ij}(\tilde{\theta}))$, $I^{-1} = (I^{ij})$, $I_{ijk} = \ell_{ijk}(\tilde{\theta})$, etc. Here, the subscripted I s indicate derivatives of ℓ evaluated at the posterior mode.

This paper focuses on assessing the nearness of the posterior distribution to the modal normal approximation, given by the normal distribution with mean $\tilde{\theta}$ and variance-covariance matrix I^{-1} . The techniques can also be used to assess nonnormality of the likelihood by replacing ℓ with the log-likelihood function and $\tilde{\theta}$ with the maximum likelihood estimate.

2 Univariate marginal distributions

Computation of $|F|$ for a marginal distribution requires derivatives of the log-marginal density and calculation of the appropriate marginal mode. Often, however, only the form of ℓ is known so that these computations require substantial numerical resources. Because the calibration of $|F|$ described in the Introduction is based on a tail probability approximation, an alternative approach is to approximate tail probabilities for the marginal distribution directly. Let Z be a random variable of interest depending on the parameter θ . Then an Edgeworth expansion gives the following approximation to tail probabilities for Z :

$$\text{pr}(Z \leq z) = \Phi \left\{ z \left(1 - \frac{1}{6} z \kappa_3 \right) - \kappa_1 + \frac{1}{6} \kappa_3 \right\} + O(n^{-1}), \quad (1)$$

where κ_1 and κ_3 are the first and third cumulants of Z . When θ is univariate ($m = 1$) and $Z = (\theta - \tilde{\theta})/\tilde{\sigma}$ with $\tilde{\sigma}^2 = -1/\ell''(\tilde{\theta})$, $\kappa_1(Z) = F/2 + O(n^{-3/2})$ and $\kappa_3(Z) = F + O(n^{-3/2})$. Equation (1) then becomes

$$\text{pr} \left\{ (\theta - \tilde{\theta})/\tilde{\sigma} \leq z \right\} = \Phi \left\{ z \left(1 - \frac{1}{6}zF \right) - \frac{1}{3}F \right\} + O(n^{-1}), \quad (2)$$

which is equivalent to $O(n^{-1})$ to the approximation used by Slate (1994) to derive the bound on $|F|$ for adequate normality. Using (2) for the 0.05 tails gives the condition $|F| \leq 0.40$ for adequate normality. (That is, $F = -0.46$ gives the lower nominal 0.05 tail coverage 0.10 while $F = 0.40$ gives this region coverage 0.025; $F = 0.46$ gives the upper nominal 0.05 tail coverage 0.10 and $F = -0.40$ gives this region coverage 0.025.) The condition $|F| \leq 0.40$ (rather than $|F| \leq 0.38$) is adopted for use in this paper.

Equation (1), however, is more general than (2) and can be used to generate conditions on κ_1 and κ_3 so that the distribution of Z is adequately normal in the 0.05 tails. For $p, q \in (0, 1)$, (1) implies that $\text{pr}\{Z \leq \Phi^{-1}(p)\} = q + O(n^{-1})$ if and only if

$$\kappa_3 \left[1 - \left\{ \Phi^{-1}(p) \right\}^2 \right] - 6\kappa_1 = 6 \left\{ \Phi^{-1}(q) - \Phi^{-1}(p) \right\}.$$

Application of this condition with $(p, q) = (0.05, 0.025)$ and $(p, q) = (0.95, 0.975)$ gives $|6\kappa_1 + 1.71\kappa_3| = 1.89$; $(p, q) = (0.05, 0.100)$ and $(p, q) = (0.95, 0.900)$ give $|6\kappa_1 + 1.71\kappa_3| = 2.18$. Thus, being conservative, adequate normality of the 0.05 tails of the distribution for Z is indicated when $|6\kappa_1 + 1.71\kappa_3| \leq 1.89$.

Let D denote the quantity $6\kappa_1 + 1.71\kappa_3$. Then $|D|$ is called the cumulant-based diagnostic for assessing nonnormality of the distribution of a quantity of interest Z . Inference based on approximate normality of Z may be unreliable when $|D| > 1.89$. Often κ_1 and κ_3 are not known precisely for the random variable Z . In these cases the cumulants in D may be replaced by $O(n^{-1})$ approximations and the reference value of 1.89 remains valid. Let $\hat{\kappa}_1$ and $\hat{\kappa}_3$ be $O(n^{-1})$ (or better) approximations for κ_1 and κ_3 and denote the resulting cumulant-based diagnostic by $|\hat{D}| = |6\hat{\kappa}_1 + 1.71\hat{\kappa}_3|$. The remainder of this section discusses the calculation of $|\hat{D}|$ when Z is one component of

the m -dimensional parameter θ , the use of this setting for assessing joint nonnormality of θ , and also the more general choice of Z as a scalar function of θ .

2.1 One component of interest

Suppose the nearness to normality of the marginal distribution of θ^1 is of interest. DiCiccio, Field and Fraser (1990) gave expressions for the first four cumulants of $Z^1 = (\theta^1 - \tilde{\theta}^1)(I^{11})^{-1/2}$. The first and third of these cumulants are given by $\kappa_1(Z^1) = \frac{1}{2}F_{3,2} + O(n^{-3/2})$ and $\kappa_3(Z^1) = F_{3,1} + O(n^{-3/2})$, where

$$\begin{aligned} F_{3,1} &= I_{ijk} I^{i1} I^{j1} I^{k1} (I^{11})^{-3/2}, \\ F_{3,2} &= I_{ijk} I^{ij} I^{k1} (I^{11})^{-1/2}, \end{aligned}$$

and the convention is that summation occurs over all indices that appear as both subscripts and superscripts (as in McCullagh, 1987, for example). Thus, taking $\hat{\kappa}_1(Z^1) = \frac{1}{2}F_{3,2}$ and $\hat{\kappa}_3(Z^1) = F_{3,1}$, adequate normality of the marginal distribution of Z^1 is indicated when $|6\hat{\kappa}_1 + 1.71\hat{\kappa}_3| \leq 1.89$.

2.2 Assessing joint nonnormality

One may assess the nearness to normality of the marginal distribution of θ^r by observing the value of $|3F_{3,2}^r + 1.71F_{3,1}^r|$, where

$$\begin{aligned} F_{3,1}^r &= I_{ijk} I^{ir} I^{jr} I^{kr} (I^{rr})^{-3/2} \\ F_{3,2}^r &= I_{ijk} I^{ij} I^{kr} (I^{rr})^{-1/2}. \end{aligned}$$

Adequate normality in the 0.05 tails is indicated when this quantity is no more than 1.89. Note that setting $F_{3,2}^r = F_{3,1}^r = 0.40$ gives $|3F_{3,2}^r + 1.71F_{3,1}^r| = 1.89$. Note also, however, that requiring that $|F_{3,2}^r| \leq 0.40$ and $|F_{3,1}^r| \leq 0.40$ is more strict than $|3F_{3,2}^r + 1.71F_{3,1}^r| \leq 1.89$. The latter condition permits a trade-off between $F_{3,1}^r$ and $F_{3,2}^r$.

One may assess the joint normality of the distribution of θ by combining the m marginal measures. An unreliable normal approximation to the joint distribution is indicated when

$$\sum_{r=1}^m |3F_{3,2}^r + 1.71F_{3,1}^r| > m(1.89). \quad (3)$$

However, this procedure of simply summing the nonnormality measures for the component marginal distributions is not wholly satisfactory for two reasons. First, the marginal distributions may be nearly normal despite a highly nonnormal joint distribution. This behavior can be explained at least in part by the central limit type effect of the averaging required to compute marginal distributions and suggests that conditional distributions, rather than marginal distributions, will be more useful in assessing the nonnormality of the joint distribution; this approach is discussed in Section 3. The second reason that (3) is not satisfactory as an indication of unreliable joint normality is that the left side of the inequality is not invariant under affine transformations of the m -dimensional parameter θ . This concern for affine invariance motivates the use of the measures proposed by Kass (1989):

$$\begin{aligned} m^2 \bar{B}^2 &= I^{ab} I^{de} I^{cf} I_{abc} I_{def} \\ B^2 &= I^{ad} I^{be} I^{cf} I_{abc} I_{def}, \end{aligned}$$

which are affine-invariant. Interestingly, defining

$$F_{3,1}^{rst} = I_{ijk} I^{kr} (I^{rr})^{-1/2} I^{js} (I^{ss})^{-1/2} I^{it} (I^{tt})^{-1/2},$$

there is

$$\sum_{r=1}^m (F_{3,2}^r)^2 = m^2 \bar{B}^2 \quad (4)$$

$$\sum_{r,s,t=1}^m (F_{3,1}^{rst})^2 = B^2. \quad (5)$$

Equations (4) and (5) suggest heuristics for bounds for $m^2 \bar{B}^2$ and B^2 for adequate approximate joint normality based on $|F_{3,1}^r| \leq 0.40$ and $|F_{3,2}^r| \leq 0.40$. Specifically

$m^2 \bar{B}^2 \leq m(0.40)^2$ and $B^2 \leq m^3(0.40)^2$. The first condition follows directly from $|F_{3,2}^r| \leq 0.40$ while the second requires the extension that the bound of 0.40 is also appropriate for the “off-diagonal” measures $F_{3,1}^{rst}$, r, s, t not all equal.

2.3 Scalar function of interest

One may be interested in the nonnormality of the marginal distribution of a more general scalar function $g(\theta)$ rather than a component θ^r . The approach using the Edgeworth approximation stipulates that $|6\hat{\kappa}_1 + 1.71\hat{\kappa}_3| \leq 1.89$ indicates adequate approximate normality, where $\hat{\kappa}_1$ and $\hat{\kappa}_3$ are estimates of the first and third cumulants of g . Assume that g is a smooth function of constant order and let $g = g(\theta)$, $\tilde{g} = g(\tilde{\theta})$, $\tilde{g}_i = \frac{\partial g}{\partial \theta_i}|_{\tilde{\theta}}$, $\tilde{g}_{ij} = \frac{\partial^2 g}{\partial \theta_i \partial \theta_j}|_{\tilde{\theta}}$, etc., and $\tilde{\sigma}_g^2 = g_i g_j I^{ij}$. Then (see Appendix A)

$$\begin{aligned}\kappa_1 \{(g - \tilde{g})/\tilde{\sigma}_g\} &= \frac{1}{2}\tilde{\sigma}_g^{-1}\tilde{g}_i I_{abc} I^{ia} I^{bc} + \frac{1}{2}\tilde{\sigma}_g^{-1}\tilde{g}_{ij} I^{ij} + O(n^{-1}) \\ \kappa_3 \{(g - \tilde{g})/\tilde{\sigma}_g\} &= \tilde{\sigma}_g^{-3}\tilde{g}_i \tilde{g}_j \tilde{g}_k I^{ai} I^{bj} I^{ck} I_{abc} + 3\tilde{\sigma}_g^{-3}\tilde{g}_i \tilde{g}_j \tilde{g}_{kl} I^{ik} I^{jl} + O(n^{-1}).\end{aligned}$$

These approximate cumulants are invariant under affine transformations of θ so that computation may be simplified by first transforming to the orthogonal parameterization $\phi = I^{1/2}\theta$.

3 Univariate conditional distributions

An alternative to considering the univariate marginal distributions is to use “slices” through the posterior to assess nearness to normality. For the m -dimensional parameter θ , consider the slice in the direction $\rho \in \mathbb{R}^m$ given by the conditional distribution with log density $w(t) \propto \ell(\tilde{\theta} + \rho t)$. This section gives expressions for the nonnormality diagnostics F_ρ and $|\hat{D}| = |6\hat{\kappa}_1 + 1.71\hat{\kappa}_2|$ for the univariate slice in the direction ρ .

3.1 The diagnostics

It is straightforward to show that the third derivative measure for the conditional distribution in the direction ρ , $F_\rho = w'''(0) \{-w''(0)\}^{-3/2}$, is given by

$$F_\rho = (\rho^i \rho^j \rho^k I_{ijk}) (\rho^i \rho^j I_{ij})^{-3/2}.$$

Adequate normality of the conditional slice in the direction ρ is indicated if $F_\rho^2 \leq 0.16$.

The combination $|6\hat{\kappa}_1 + 1.71\hat{\kappa}_3|$ derived from the Edgeworth approximation may also be used as a nonnormality diagnostic. For the slice in the direction ρ , the first and third conditional cumulants are given by

$$\begin{aligned}\kappa_1(t/\hat{\sigma}_t) &= 3\eta\alpha^{-3/2} + O(n^{-1}) \\ \kappa_3(t/\hat{\sigma}_t) &= 6\eta\alpha^{-3/2} + O(n^{-1}),\end{aligned}$$

where $\alpha = \rho^i \rho^j I_{ij}$, $\hat{\sigma}_t = \alpha^{-1/2}$, and $\eta = \frac{1}{6}\rho^i \rho^j \rho^k I_{ijk}$. Details are given in Appendix B. Here, $\hat{\kappa}_1 = \frac{1}{2}F_\rho$ and $\hat{\kappa}_3 = F_\rho$ so that the condition $|6\hat{\kappa}_1 + 1.71\hat{\kappa}_3| \leq 1.89$ is equivalent to $|F_\rho| \leq 0.40$ or $F_\rho^2 \leq 0.16$. Thus, the cumulant-based diagnostic provides the same information as the third derivative measure F_ρ .

3.2 Dynamic graphics for exploring the slices

A graph of the univariate slice through the multivariate posterior distribution or likelihood can aid visual calibration of F_ρ^2 immensely. A useful tool is the dynamic visualization of the conditional slices and F_ρ^2 values as the direction of the slice changes. Routines have been developed in XLISP-STAT (Tierney, 1990) that implement this dynamic graphics.

An example of the conditional F^2 plot is shown in Figure 1. The graph shows the standardized conditional distribution along the chosen slice (the solid line) and the corresponding standard normal approximation (dotted line), scaled so that they match at the mode. The differences between the curves are highlighted by vertical lines, which, on color monitors, are blue when the normal approximation is beneath

the conditional distribution and green when the normal tail is above that for the slice. In Figure 1, the slice is highly right skew so that the difference in the right tail is shaded blue while that in the left tail is green. The shading in the “thermometer” along the left side of the graph depicts the value of the diagnostic F_ρ^2 . In Figure 1, the maximum value of F_ρ^2 that can be shown is 1.0, which is indicated at the top of the thermometer. There is a thin horizontal line across this thermometer at the value 0.16, the recommended cutoff for adequate approximate normality. On color monitors, the portion of the value of F_ρ^2 beneath this cutoff is represented by black shading while any excess of 0.16 is shown in red. The value of F_ρ^2 in Figure 1 is 0.77.

The direction of the slice can be controlled using sliders. There are $m - 1$ sliders for a univariate slice through an m -dimensional distribution. For the example shown in Figure 1, the dimension is $m = 2$; the one slider labeled “Phi 0” specifies the angle in radians of the slice in the parameter plane. The slider permits values of the angle from 0 to 2π , although the slice through $\pi + \phi$ is the slice through ϕ flipped about zero and so has the same F_ρ^2 value. The slider dialog appears as a separate window in XLISP-STAT and has been positioned atop the plot in Figure 1. With each new direction chosen using the slider(s), the value of F_ρ^2 is recomputed and the plot is redrawn. For the m -dimensional model, the space is parameterized as

$$\begin{aligned}\theta^1 &= \cos \phi_0 \\ \theta^2 &= \sin \phi_0 \cos \phi_1 \\ \theta^3 &= \sin \phi_0 \sin \phi_1 \cos \phi_2 \\ &\vdots \\ \theta^m &= \sin \phi_0 \sin \phi_1 \dots \sin \phi_{m-2},\end{aligned}$$

where $\phi_0, \dots, \phi_{m-3} \in (0, \pi)$ and $\phi_{m-2} \in (0, 2\pi)$. Figure 2 shows the conditional F^2 plot and direction sliders for a four-dimensional model.

There is also a facility for a grand-tour-type exploration of the $m - 1$ dimensional direction space. This feature is particularly useful when m is large. A menu item

on the plot called “Run” begins the dynamic exploration. The tour proceeds by randomly selecting two points on the m -dimensional unit sphere and then stepping a random number of times from the first toward the second. The number of steps is generated uniformly on the integers between 0 and $\pi/(2\alpha)$, where α may be set by the user. When the generated number of steps has been taken, two new points are selected on the sphere and the process repeats. With each step of the tour, the value of F_ρ^2 is computed and the conditional F^2 plot is redrawn. As the tour progresses, the largest value of F_ρ^2 encountered and the direction of this value are stored. The tour provides a way to randomly explore the large space of possible univariate slices while the direction sliders permit systematic exploration.

4 Examples

4.1 Two-parameter survival model

Consider the two-parameter model used by Feigl and Zelen (1965) to represent the relationship between survival time for leukemia patients and white blood cell count (WBC). Let Y_i and WBC_i be the survival time in weeks and the white blood cell count for the i -th patient. The model is

$$\begin{aligned} Y_i &\sim \text{Exponential with mean } \mu_i, \text{ independently} \\ \mu_i &= \theta^1 \exp(-\theta^2 x_i), \end{aligned}$$

where $\theta = (\theta^1, \theta^2)$, and $x_i = \text{WBC}_i - \overline{\text{WBC}}$, the centered white blood cell count for patient i . $i = 1, \dots, n$. Here the data for the $n = 17$ AG-positive patients are used. In addition to θ , the parameterizations λ and β may be studied, where $\mu_i = (\beta^1 x_i^*)^{\beta^2}$, with $x_i^* = \exp(-x_i)$, and $\mu_i = \exp(\lambda^1 - \lambda^2)$. The prior distribution is uniform on λ .

Maximization of the posterior density yields modes and first-order standard errors of $\tilde{\theta}^1 = 48.27$ (11.4), $\tilde{\theta}^2 = 1.11$ (0.40) with $\text{corr}(\theta^1, \theta^2) = 0.00$, $\tilde{\lambda}^1 = 3.93$ (0.24), $\tilde{\lambda}^2 = 1.11$ (0.41) with $\text{corr}(\lambda^1, \lambda^2) = 0.00$, and $\tilde{\beta}^1 = 13.64$ (7.81), $\tilde{\beta}^2 = 1.50$ (0.34)

with $\text{corr}(\beta^1, \beta^2) = -0.96$. Because this is only a two dimensional parameter, the contours of the posterior distribution are informative and easy to examine; these are shown in Figure 3. The dotted line is the 95% central region based on the modal normal approximation and, if the normal approximation is reasonable, should match the outermost contour shown. The contours, as well as other nonnormality diagnostics (see Kass and Slate, 1994), indicate that β is very poorly behaved, θ is considerably better and λ is the best of the three parameterizations.

Table 1 presents the nonnormality diagnostics for the component marginal distributions for the three parameterizations. The first two rows of this table give the values for the approximate first and third cumulants, $2\hat{\kappa}_1 = F_{3,1}$ and $\hat{\kappa}_3 = F_{3,2}$. Because $F_{3,1} = F_{3,2} = F$ when $m = 1$, one may consider advising caution for the use of a normal approximation if either $|F_{3,1}|$ or $|F_{3,2}|$ exceeds 0.40. This rule results in only the marginal distributions of λ^2 and θ^2 being deemed adequately near normal. The result of Section 2.1, however, indicates that the appropriate nonnormality diagnostic for component marginal distributions is $|\hat{D}| = |6\hat{\kappa}_1 + 1.71\hat{\kappa}_3|$. By comparing $|\hat{D}|$ to 1.89, the marginal distributions for λ^1 , λ^2 , and θ^2 are well-approximated by the normal, while neither component margin is adequately normal for β . Recall that the cutoff of 1.89 for $|\hat{D}|$ was chosen so that the coverage of the lower and upper 0.05 tails under the normal approximation is between 0.025 and 0.10. The rows labeled $\hat{p}_{.05}$ and $\hat{p}_{.95}$ give the coverages of the 0.05 and 0.95 tail regions under normal approximation, computed using the tail probability approximation of DiCiccio, Field and Fraser (1990). (The entry $--$ for β^1 indicates numerical difficulties due to the extremely poor behavior of this parameterization.) Only the margins for λ^1 , λ^2 and θ^2 have $0.025 \leq \hat{p}_{.05} \leq 0.10$ and $0.90 \leq \hat{p}_{.95} \leq 0.975$, supporting the information obtained from the diagnostic $|\hat{D}|$.

Table 1 also presents the affine-invariant diagnostics $m^2\bar{B}^2$ and B^2 discussed in Section 2.2. Because $m = 2$ here, the heuristic following equations (4) and (5) contends that the joint posterior distribution is near normal if $m^2\bar{B}^2 \leq 0.32$ and $B^2 \leq$

1.28. The conclusion confirms the information gained from the marginal measures: the posterior distribution for λ is well-approximated by a normal distribution while a normal approximation is less reasonable for θ and very poor for β .

To illustrate the ideas of Section 2.3, consider the survival function $g = \text{pr}(Y_i > t)$ for a specified survival time t and white blood cell count. Table 2 shows the diagnostics for the six-month ($t = 26$ weeks) and one-year ($t = 52$) survival probabilities for patients with white blood cell counts of 15000 and 25000. As the white blood cell count and t increase, the survival probability decreases and so its distribution becomes more highly right skew. The cumulant diagnostic $|\hat{D}|$ indicates that the survival probability has a distribution that is close to normal only for the six-month survival, both white blood cell counts. The approximate coverages $\hat{p}_{.05}$ and $\hat{p}_{.95}$ support this observation. In each case where $|\hat{D}| > 1.89$, at least one of the coverages of the nominal 0.05 tail regions is not between 0.025 and 0.10. These approximate coverages were computed using importance sampling with draws from a t_3 distribution and have standard errors less than 0.009. The coverages are not shown for the β parameterization because of the numerical difficulties associated with this parameterization.

Figure 4 shows one slice through the posterior distribution for β , the worst of the three parameterizations. This slice is at an angle of 3.12 radians from zero, a line through the joint mode $\tilde{\beta}$ with slope -0.022 . The corresponding value of F_ρ^2 is 10.51, extremely large (though not the largest value for this parameterization). In contrast, the largest value of F_ρ^2 observed for the λ parameterization on a grid of 500 angles between 0 and π is 0.12, considerably less than the suggested upper bound of 0.16 for a reliable normal approximation.

4.2 Orange tree data

Draper and Smith (1981, p. 524, exercise N) provide measurements of the trunk circumference of orange trees at seven time points for five trees. This example considers the data for tree three only. One model proposed for the relationship between the

trunk circumference (in millimeters), y_i , and time of measurement (in days/1000), x_i , is the Gompertz nonlinear regression model:

$$\begin{aligned} y_i &\sim N(\mu_i, \sigma^2), \text{ independently} \\ \mu_i &= \theta^1 \exp\left(-e^{\theta^2 - \theta^3 x_i}\right) \end{aligned}$$

for $i = 1, \dots, 7$. A flat prior distribution is used for the parameter $\theta = (\theta^1, \theta^2, \theta^3)$, so that the diagnostics are evaluating nonnormality of the likelihood function. The maximum likelihood estimate and approximate standard errors are $\hat{\theta}^1 = 188.4$ (25.87), $\hat{\theta}^2 = 0.83$ (0.076), and $\hat{\theta}^3 = 1.35$ (0.29). Approximate posterior correlations under the flat prior are $\text{corr}(\theta^1, \theta^2) = -0.21$, $\text{corr}(\theta^1, \theta^3) = -0.96$, and $\text{corr}(\theta^2, \theta^3) = 0.44$.

Table 3 presents the nonnormality diagnostics for the component marginal distributions for θ . The cumulant-based diagnostic, $|\hat{D}| = |6\hat{\kappa}_1 + 1.71\hat{\kappa}_3|$ is larger than the suggested bound of 1.89 for adequate normality for all components but θ^3 , although its value of 2.02 for θ^2 is near the bound. Thus inferences concerning θ^1 based on approximate normality may not be reliable, while the situation is less clear for θ^2 and there is no cause for concern for θ^3 . These observations are supported by the marginal density plots shown in Figure 5: the normal approximation is far from accurate for θ^1 and most accurate (especially in the tails) for θ^3 . The approximate coverages of the tail regions with probability 0.05 under the normal approximation are given in Table 3 as $\hat{p}_{.05}$ and $\hat{p}_{.95}$. These coverages were computed using the tail probability approximation of DiCiccio, Field and Fraser (1990). The approximate tail coverages for θ^2 and θ^3 satisfy the conditions for adequate normality, while for θ^1 both tails violate these conditions. Thus the tail behavior of the marginal distributions of θ^1 and θ^3 support the choice of 1.89 as the value for $|\hat{D}|$ above which at least one of the nominal 0.05 tail regions no longer has coverage between 0.05 and 0.10. The tail coverages for θ^2 , however, illustrate the approximate nature of this interpretation.

The affine-invariant measures $m^2\bar{B}^2$ and B^2 are also shown in Table 3. For $m = 3$, the heuristic following equations (4) and (5) indicates that inference based on a normal approximation may not be reliable if $B^2 > 4.32$ or $m^2\bar{B}^2 > 0.48$. The values of

$m^2 \bar{B}^2$ and B^2 in this example far exceed these bounds, indicating that a normal approximation to the joint posterior distribution for $\theta = (\theta^1, \theta^2, \theta^3)$ is not reliable.

Examination of univariate conditional slices through the posterior for θ supports the information gained from the marginal diagnostics. The maximum value of F_ρ^2 is approximately 16.2, which occurs in the direction $(-0.99996, 0, 0.0094)$. The nonzero values in the θ^1 and θ^3 directions reflect the correlation of -0.96 between these components. A less extreme slice is shown in Figure 6. This slice is in the direction $\rho = (-0.99998, -0.0007, 0.0069)$ and $F_\rho^2 = 1.66$.

4.3 A two-component reliability model

The two-component reliability model discussed by Hills (1989, Section 3.3) has likelihood given by

$$p(y_1, y_2 \mid \lambda_1, \lambda_2, \psi) = \prod_{i=1}^n \lambda_1 \lambda_2 \exp\{-(\lambda_1 y_{1i} + \lambda_2 y_{2i})\} \\ \times \left[1 + \frac{e^\psi - 1}{e^\psi + 1} \{2 \exp(-\lambda_1 y_{1i}) - 1\} \{2 \exp(-\lambda_2 y_{2i}) - 1\} \right]$$

and prior distribution given by

$$p(\lambda_1, \lambda_2, \psi) = \begin{cases} \frac{\exp(\psi)}{\lambda_1 \lambda_2 \{1 + \exp(\psi)\}^2} & \lambda_1, \lambda_2 > 0 \\ 0 & \text{otherwise} \end{cases}.$$

In addition to the parameterization $\lambda = (\lambda_1, \lambda_2, \psi)$, one may consider the parameterization $\phi = (\phi_1, \phi_2, \psi)$ where $\phi_i = \log \lambda_i$, $i = 1, 2$. Because there is no restriction on the range of the ϕ parameterization, it may be expected to produce a posterior distribution that is more accurately approximated by the normal than the λ parameterization. The data modeled here are those studied by Hills (1989) and Lawless (1982, Example 10.1.2, uncensored observations), consisting of 17 observations for each of y_1 and y_2 .

Maximization of the posterior density for λ yields a mode and approximate standard deviations of $\tilde{\lambda}_1 = 0.002$ (0.0005), $\tilde{\lambda}_2 = 0.022$ (0.005), and $\tilde{\psi} = 1.43$ (1.15).

First order posterior correlations are $\text{corr}(\lambda_1, \lambda_2) = 0.18$, $\text{corr}(\lambda_1, \psi) = 0.10$, and $\text{corr}(\lambda_2, \psi) = 0.07$. The ϕ parameterization yields $\tilde{\phi}_1 = -6.13$ (0.24), $\tilde{\phi}_2 = -3.76$ (0.24), $\tilde{\psi} = 1.48$ (1.15), $\text{corr}(\phi_1, \phi_2) = 0.18$, $\text{corr}(\phi_1, \psi) = 0.11$, $\text{corr}(\phi_2, \psi) = 0.08$.

Table 4 gives the nonnormality diagnostics for the component marginal distributions for the λ and ϕ parameterizations. The cumulant-based diagnostic, $|\hat{D}| = |6\hat{\kappa}_1 + 1.71\hat{\kappa}_3|$ is larger than the suggested bound of 1.89 for adequate normality for all of the component marginal distributions associated with the λ parameterization. Thus inference based on normal approximations to the marginal posterior distributions for λ_1 , λ_2 , and ψ may not be reliable. The values of $|\hat{D}|$ for the ϕ parameterization, however, indicate that the marginal posterior distributions for ϕ_1 and ϕ_2 would be well-approximated by a normal distribution, while that for ψ would not. Note that ψ is the same quantity in both parameterizations, but that the values of the diagnostics differ. This is because the diagnostics depend on the underlying full m -dimensional parameterization; it is this dependence that makes these measures useful for comparing parameterizations.

The approximate coverages of the tail regions with probability 0.05 under the normal approximation are given in Table 4 as $\hat{p}_{.05}$ and $\hat{p}_{.95}$. The probabilities were computed using importance sampling with draws from a t_2 distribution and have standard errors ranging from 0.001 to 0.005, except for $\hat{p}_{.95}$ for ψ , which has standard errors of 0.01 and 0.006 for the λ and ϕ parameterizations, respectively. As in the previous examples, these coverages support the choice of 1.89 as the value for $|\hat{D}|$ above which at least one of the nominal 0.05 tail regions no longer has coverage between 0.05 and 0.10. Also shown in Table 4 are the affine-invariant measures $m^2\bar{B}^2$ and B^2 . For $m = 3$, the heuristic following equations (4) and (5) indicates that inference based on a normal approximation may not be reliable if $B^2 > 4.32$ or $m^2\bar{B}^2 > 0.48$. The diagnostic $m^2\bar{B}^2$ indicates that a normal approximation to the posterior distribution for λ may not be reasonable, while these diagnostics do not give cause for concern in the ϕ parameterization.

A scalar function of interest is the survival function $g = \text{pr}(Y_1 > t_1, Y_2 > t_2)$ for specified t_1 and t_2 . Table 5 gives the diagnostics for the distribution of g for $(t_1, t_2) = (300, 30)$, $(500, 50)$, and $(700, 70)$. As t_1 and t_2 increase, the mode of g , $\tilde{g}$, becomes smaller and the distribution of g becomes more right skew (and hence more highly nonnormal). The cumulant diagnostic $|\hat{D}|$ indicates that the normal approximation to the distribution of g is not reliable for the ϕ parameterization when $(t_1, t_2) = (500, 50)$ and $(700, 70)$. That these distributions are not near normal is supported by the approximate coverages $\hat{p}_{.05}$ and $\hat{p}_{.95}$, because both $\hat{p}_{.05}$ and $\hat{p}_{.95}$ for $(t_1, t_2) = (700, 70)$ and $\hat{p}_{.05}$ for $(t_1, t_2) = (500, 50)$ fail to be within the guideline. Additionally, $\hat{p}_{.05}$ for $(t_1, t_2) = (700, 70)$ for the λ parameterization is too small. The approximate coverages were computed using importance sampling from a t_2 distribution and have standard errors of at most 0.004, except the value of $\hat{p}_{.95} = .896$ for the ϕ parameterization, which has standard error 0.007.

Examination of univariate conditional slices through the posterior distribution in the λ and ϕ parameterizations supports the information gained from the marginal diagnostics, that ψ is the problematic parameter. The maximum value of F_ρ^2 for both the λ and ϕ parameterizations occurs in the direction of ψ , $(0, 0, 1)$, and exceeds 0.35; this slice is shown for the λ parameterization in Figure 7. Unlike those for the λ parameterization, however, “typical” values of F_ρ^2 for the ϕ parameterization are well under the cutoff of 0.16.

5 Discussion

As computational methods improve and enable more complicated models to be investigated, the need increases for simple diagnostics that can provide insight into the shape of multivariate likelihood functions and posterior distributions. The diagnostics discussed in this paper provide a means for assessing nonnormality of multivariate distributions through univariate marginal and conditional distributions. Although these

measures do not capture all features of the multivariate behavior, examination of the value of the diagnostics for many univariate distributions (many “slices” through the joint distribution, for example) can reveal important deviations from normality, and regions that are potentially problematic for computation and inference.

If multivariate nonnormality is to be assessed through dimension reduction, as in this paper, it is likely that a procedure based on conditional distributions is preferable to one using marginal distributions. The averaging that takes place when computing marginal distributions suggests that margins will tend to be more nearly normal than conditional distributions, and hence mask the joint nonnormality. Section 2.2, however, showed that the marginal measures provide a basis for determining “large” values of the scalar summaries $m^2\bar{B}^2$ and B^2 of the full third derivative arrays.

Recently, sampling techniques such as Monte Carlo Markov chain are being used for inference, particularly in the Bayesian paradigm. Diagnostics of nonnormality like those proposed in this paper remain important in this setting because many sampling techniques can be made more efficient when parameters are orthogonal, which is easily accomplished when sampling distributions are close to normal. See, for example, Gilks and Roberts *et al.* (1996) and Hills (1989) for discussions of improvements to sampling techniques by orthogonalization.

An interesting extension of the use of the diagnostics $|F|$ and $|\hat{D}|$ is the study of their behavior as the data vary. Such investigations will reveal the effect of the design on computational and inferential procedures that are more efficient when approximate normality is reliable.

Acknowledgment

This research is supported by a grant from the National Science Foundation. Helpful discussions with Thomas J. DiCiccio, School of Industrial and Labor Relations, Cornell University, aided the development of this work.

References

- Albert, J. H. (1989). On the Use of Power Transformations to Simplify Posterior Distributions. Technical Report, Bowling Green State University.
- Anscombe, F. J. (1964). Normal Likelihood Functions, *Annals of the Institute of Statistical Mathematics*, **16**, 1-19.
- DiCiccio, T. J. & Stern, S. E. (1993). Bartlett adjustments for Bayesian inference, *Biometrika*, **80**, 731-740.
- DiCiccio, T. J., Field, C. A. & Fraser, D. A. S. (1990). Approximations of marginal tail probabilities and inference for scalar parameters, *Biometrika*, **77**, 77-95.
- Feigl, P. & Zelen, M. (1965). Estimation of exponential survival probabilities with concomitant information, *Biometrics*, **21**, 826-838.
- Gillks, W. R. & Roberts, G. O. (1996). Strategies for improving MCMC, in *Markov Chain Monte Carlo in Practice*, W. R. Gilks, S. Richardson, and D. J. Spiegelhalter. Ed., Chapman and Hall, New York.
- Hills, S. (1989). *The Parameterisation of Statistical Models*, Ph.D. Thesis, University of Nottingham.
- Kass, R. E. (1989). The Geometry of Asymptotic Inference, *Statistical Science*, **4**, 188-234.
- Kass, R. E. & Slate, E. H. (1994). Some diagnostics of maximum likelihood and posterior non-Normality, *The Annals of Statistics*, **22**, 668-695.
- Kass, R. E. & Slate, E. H. (1993). Curvature measures in data analysis, *Statistical Sciences and Data Analysis: Proceedings of the Third Pacific Area Statistical Conference*. eds. K. Matusita, M. L. Puri. T. Hayakawa, et al., VSP International Science Publishers, Zeist, Netherlands, 277-290.
- Kass, R. E. & Slate, E. H. (1992). Reparameterization and diagnostics of posterior non-normality, *Bayesian Statistics 4*, edited by J. O. Berger, J. M. Bernardo, A.

- P. Dawid, D. V. Lindley and A. F. M. Smith, Oxford University Press, 289-305.
- Lawless, (1982). *Statistical models and methods for lifetime data*, Wiley, New York.
- McCullagh, P. (1987). *Tensor Methods in Statistics*, Chapman and Hall, Cambridge.
- Slate, E. H. (1994). Parameterizations for natural exponential families with quadratic variance functions *Journal of the American Statistical Association*, **89**, 1471-1482.
- Sprott, D. A. (1973). Normal likelihoods and their relation to large sample theory of estimation, *Biometrika*, **60**, 457-465.
- Tierney, L. (1990). *Lisp-Stat: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics*, John Wiley & Sons, New York.
- Tierney, L., & Kadane, J. B. (1986). Accurate Approximations for Posterior Moments and Marginal Densities, *Journal of the American Statistical Association*, **81**, 82-86.
- Tierney, L., Kass, R. E. and Kadane, J. B. (1989) Approximate marginal densities of nonlinear functions, *Biometrika*, **76**, 425-433. Amendment (1991), **78**, 233-234.

A The cumulants of g

Using the form of the posterior density function $f(\theta) = cb(\theta)\exp\{h(\theta)\}$, DiCiccio and Stern (1993, equation 8) provide an expansion for the moment generating function of the ratio statistic $2\{h(\tilde{\theta}) - h(\theta)\}$. When evaluated at zero, the expansion for the moment generating function enables the normalizing constant c to be approximated and yields an approximation for the posterior distribution of θ . In the case that $b(\theta) = 1$, the first order approximation is

$$f_\delta(\delta) = N(\delta; 0, I^{-1}) \left\{ 1 + \frac{1}{6} \delta^1 \delta^b \delta^c I_{abc} + O(n^{-1}) \right\}, \quad (6)$$

where $\delta = (\delta^1, \dots, \delta^m)$ with $\delta^i = \theta^i - \tilde{\theta}^i$ and $N(\delta; 0, I^{-1})$ is the m -dimensional normal density with mean zero and variance-covariance matrix I^{-1} . Moments of δ can be computed using (6), producing

$$\begin{aligned} E(n^{1/2} \delta^i) &= \frac{1}{2} n^{1/2} I_{abc} I^{ia} I^{bc} + O(n^{-1}) \\ E(n \delta^i \delta^j) &= n I^{ij} + O(n^{-1}) \\ E(n^{3/2} \delta^i \delta^j \delta^k) &= \frac{1}{6} n^{3/2} I_{abc} (9 I^{ij} I^{ka} I^{bc} + 6 I^{ia} I^{jb} I^{kc}) + O(n^{-1}) \\ E(n^2 \delta^i \delta^j \delta^k \delta^l) &= n^2 (I^{ij} I^{kl} + I^{ik} I^{jl} + I^{il} I^{jk}) + O(n^{-1}). \end{aligned}$$

Taylor series expansion of g gives

$$\tilde{\sigma}_g^{-1}(g - \tilde{g}) = \tilde{\sigma}_g^{-1} \tilde{g}_i \delta^i + \frac{1}{2} \tilde{\sigma}_g^{-1} \tilde{g}_{ij} \delta^i \delta^j + O(n^{-1})$$

so that

$$\begin{aligned} E\{\tilde{\sigma}_g^{-1}(g - \tilde{g})\} &= \frac{1}{2} \tilde{\sigma}_g^{-1} \tilde{g}_{ij} I^{ij} + O(n^{-1}) \\ E\left[\{\tilde{\sigma}_g^{-1}(g - \tilde{g})\}^2\right] &= \tilde{\sigma}_g^{-2} \tilde{g}_i \tilde{g}_j I^{ij} + O(n^{-1}) \\ E\left[\{\tilde{\sigma}_g^{-1}(g - \tilde{g})\}^3\right] &= \tilde{\sigma}_g^{-3} \tilde{g}_i \tilde{g}_j \tilde{g}_k I_{abc} \left(\frac{3}{2} I^{ij} I^{ka} I^{bc} + I^{ia} I^{jb} I^{kc}\right) + \frac{3}{2} \tilde{\sigma}_g^{-3} \tilde{g}_i \tilde{g}_j \tilde{g}_{kl} (I^{ij} I^{kl} + 2 I^{ik} I^{jl}) \\ &\quad + O(n^{-1}). \end{aligned}$$

These moments yield the cumulants reported in Section 2.3.

B The cumulants of the conditional slice

The conditional density function along the slice in the direction ρ is given by

$$f(t) = \frac{f_\delta(\rho t)}{\int f_\delta(\rho t) dt},$$

where $\delta = (\delta^1, \dots, \delta^m)$, $\delta^i = \theta^i - \tilde{\theta}^i$, and $f_\delta(\cdot)$ is the joint posterior density for δ . Using (6) for f_δ , derived from DiCiccio and Stern (1993), the conditional density along the slice is

$$f(t) = (2\pi)^{-1/2} \alpha^{1/2} \exp(-\frac{1}{2} \alpha t^2) \left\{ 1 + \eta t^3 + O(n^{-1}) \right\}, \quad (7)$$

where $\alpha = \rho^i \rho^j I_{ij}$ and $\eta = \frac{1}{6} \rho^i \rho^j \rho^k I_{ijk}$. Standardizing, the moments of (7) are

$$\begin{aligned} E(T/\hat{\sigma}_t) &= 3\eta\alpha^{-3/2} + O(n^{-1}) \\ E\{(T/\hat{\sigma}_t)^2\} &= 1 + O(n^{-1}) \\ E\{(T/\hat{\sigma}_t)^3\} &= 15\eta\alpha^{-1/2} + O(n^{-1}), \end{aligned}$$

with $\hat{\sigma}_t = \alpha^{-1/2}$. These moments yield the cumulants given in Section 3.1.

	λ^1	λ^2	θ^1	θ^2	β^1	β^2
$2\hat{\kappa}_1$	.49	-.04	1.18	-.04	2.48	-1.33
$\hat{\kappa}_3$	.24	-.04	.94	-.04	2.95	.21
$\hat{D}$	1.87	-.19	5.15	-.19	12.47	-3.63
$\hat{p}_{.05}$	.025	.055	.004	.060	--	.025
$\hat{p}_{.95}$	.908	.951	.802	.947	.480	.984
B^2		.237		1.057		46.987
$m^2\bar{B}^2$		.237		1.390		20.548

Table 1: Diagnostics for the component marginal distributions for the Feigl-Zelen example.

WBC = 15000						
	λ parameterization		θ parameterization		β parameterization	
	6 month	1 year	6 month	1 year	6 month	1 year
$\tilde{g}$	0.57	0.33	0.56	0.31	0.56	0.31
$\hat{\kappa}_1$	0.20	0.27	0.43	0.50	-39.6	-39.6
$\hat{\kappa}_3$	0.01	0.38	0.03	0.42	-48.0	-47.6
$\hat{D}$	1.24	2.25	2.66	3.71	-319.7	-318.7
$\hat{p}_{.05}$	0.032	0.019	0.021	0.010		
$\hat{p}_{.95}$	0.933	0.913	0.895	0.864		

WBC = 25000						
	λ parameterization		θ parameterization		β parameterization	
	6 month	1 year	6 month	1 year	6 month	1 year
$\tilde{g}$	0.49	0.24	0.47	0.22	0.44	0.20
$\hat{\kappa}_1$	0.23	0.30	0.43	0.51	-148.6	-148.5
$\hat{\kappa}_3$	0.26	0.74	0.28	0.77	-570.5	-570.1
$\hat{D}$	1.80	3.07	3.08	4.38	-1867	-1866
$\hat{p}_{.05}$	0.028	0.009	0.019	0.005		
$\hat{p}_{.95}$	0.929	0.898	0.900	0.850		

Table 2: Diagnostics for the distribution of the six-month and one-year survival probability for patients with white blood cell counts of 15000 and 25000.

	θ^1	θ^2	θ^3
$2\hat{\kappa}_1$	0.61	0.47	-0.03
$\hat{\kappa}_3$	1.57	0.36	0.13
$\hat{D}$	4.51	2.02	0.12
$\hat{p}_{.05}$	0.003	0.028	0.058
$\hat{p}_{.95}$	0.855	0.917	0.941
B^2		33.54	
$m^2\bar{B}^2$		10.83	

Table 3: Diagnostics for the component marginal distributions for the the orange tree example.

	λ_1	λ_2	ψ	ϕ_1	ϕ_2	ψ
$2\hat{\kappa}_1$	0.51	0.51	0.57	-0.31	-0.30	0.49
$\hat{\kappa}_3$	0.48	0.48	0.57	-0.25	-0.25	0.58
$\hat{D}$	2.34	2.36	2.69	-1.35	-1.32	2.46
$\hat{p}_{.05}$	.017	.017	.041	.080	.081	.044
$\hat{p}_{.95}$	.904	.910	.865	.970	.970	.874
B^2		.868			.530	
$m^2\bar{B}^2$		.691			.455	

Table 4: Diagnostics for the component marginal distributions for reliability model.

	λ parameterization			ϕ parameterization		
(t_1, t_2)	(300, 30)	(500, 50)	(700, 70)	(300, 30)	(500, 50)	(700, 70)
$\tilde{g}$	0.32	0.15	0.07	0.30	0.14	0.06
$\hat{\kappa}_1$	-0.26	-0.20	-0.13	0.09	0.15	0.24
$\hat{\kappa}_3$	0.25	0.69	1.17	0.28	0.73	1.23
$\hat{D}$	-1.13	-0.05	1.22	1.03	2.17	3.52
$\hat{p}_{.05}$	0.065	0.032	0.004	0.032	0.014	0.0006
$\hat{p}_{.95}$	0.970	0.959	0.948	0.933	0.911	0.896

Table 5: Diagnostics for the survival function $g = \text{pr}(Y_1 > t_1, Y_2 > t_2)$ for the reliability example.

Figure 1: An example of the conditional F^2 plot and direction slider. This is the β parameterization for the two parameter Feigl-Zelen model discussed in Section 4. The direction of the slice is 3.04 radians and $F_\rho^2 = 0.77$.

Figure 2: The conditional F^2 plot and direction sliders for a four-dimensional model.

Figure 3: The log-posterior contours for the Feigl-Zelen example.

Figure 4: Slice through 3.12 radians for the β parameterization of the Feigl-Zelen example. Here, $F_\rho^2 = 10.51$.

Figure 5: The component marginal distributions (solid lines) and corresponding modal normal approximations (dotted lines) for the orange tree example.

Figure 6: The conditional F^2 plot and direction sliders for the orange tree example in the direction $(-0.99998, -0.0007, 0.0069)$. The diagnostic is $F^2_\rho = 1.66$.

Figure 7: The conditional F^2 plot and direction sliders for the λ parameterization of the two-component reliability model in the direction $(0, 0, 1)$. The diagnostic is $F_\rho^2 = 0.35$.