Simulation-Assisted Saddlepoint Approximation

R. W. Butler*, R.K. Sutton, J. G. Booth and P. Ohman Strickland

Revised: September 2006

Abstract

A general saddlepoint/Monte Carlo method to approximate (conditional) multivariate probabilities is presented. The method requires a tractable joint moment generating function (m.g.f.) but does not require a tractable distribution or density. The method is easy to program and has third order accuracy with respect to increasing sample size in contrast with standard asymptotic approximations, that are typically only accurate to first order.

The method is most easily described in the context of a continuous regular exponential family. Here inferences can be formulated as probabilities with respect to the joint density of the sufficient statistics or the conditional density of some sufficient statistics given the others. Analytical expressions for these densities are not generally available and it is often not possible to simulate exactly from the conditional distributions to obtain a direct Monte Carlo approximation of the required integral. A solution to the first of these problems is to replace the intractable density by a highly accurate saddlepoint approximation. The second problem can be addressed via importance sampling; that is, an indirect Monte Carlo approximation involving simulation from a crude approximation to the true density. Asymptotic normality of the sufficient statistics suggests an obvious candidate for an importance distribution.

The more general problem considers the computation of a joint probability for a subvector of random T, given its complementary subvector, when its distribution is intractable but its joint m.g.f. is computable. For such settings the distribution may be tilted, maintaining T as the sufficient statistic. Within this tilted family, the computation of such multivariate probabilities proceeds as described for the exponential family setting.

Key words and phrases: Conditional inference, Importance sampling, Monte Carlo error, Multivariate t-distribution, Nuisance parameters, p^*-formula, Saddlepoint density.

*Ronald W. Butler is C.F. Frensley Professor of Mathematical Sciences, Department of Statistical Sciences, Southern Methodist University, Dallas, TX 75205. Richard K. Sutton is Resident Physician, Johns-Hopkins University Medical School. James G. Booth is Professor, Department of Biological Statistics and Computational Biology, Cornell University. Pamela Ohman Strickland is Associate Professor, Department of Biostatistics, School of Public Health, University Medicine and Dentistry of New Jersey. Professors Butler and Booth were partially supported by NSF-DMS awards 0604318 and 0405543 respectively.

1 Introduction

Suppose that data X comes from a regular exponential family $\{f(x;\theta) : \theta \in \Theta\}$ in which

$$f(x;\theta) = \exp\left\{\theta^T t - b(\theta) - d(x)\right\}, \tag{1}$$

where $t = t(x)$ is a k-dimensional sufficient statistic. The density of the random variable $T = t(X)$ has exponential form

$$p(t;\theta) = \exp\left\{\theta^T t - b(\theta) - d_T(t)\right\}, \tag{2}$$

where $d_T(t) = \int_{t(x)=t} \exp\{-d(x)\}\, dx$, an integral that is often intractable. The plausibility of a hypothesized value θ_0 is typically measured by a probability of the form,

$$P\{s(T) > s_0; \theta_0\} = \int 1\{s(t) > s_0\} p(t;\theta_0) dt\,. \tag{3}$$

where $s(T)$ is, for example, the likelihood ratio test statistic. Exact analytical calculation of this probability is not possible if the density of T is intractable. However, it is usually straightforward to simulate data from (1) and hence a brute force Monte Carlo estimate of (3) can be calculated.

More generally, let $\theta = (\psi, \lambda)$ be a partition of θ and let $T = (U, V)$ be the corresponding partition of T. Inferences concerning ψ alone can be made free of the nuisance parameter λ using the conditional distribution of U given V. This density is proportional to the joint density in (2), being the ratio of the joint density of $T = (U, V)$ to the marginal density of V. In contrast with the unconditional case, it is often not possible to simulate from this conditional density. Thus, brute force Monte Carlo estimation of conditional probabilities is not an option.

In this article we propose a simple but general saddlepoint/Monte Carlo method for approximating both unconditional and conditional expectations like (3). The more important approximations are for conditional expectations which cannot be estimated by brute force Monte Carlo. Like the brute force methods, we obtain confidence limits on the Monte Carlo aspect of the method, which is generally the dominant error. A very small bias results from the saddlepoint aspect and cannot be removed, however this is generally negligible.

Apart from exponential family settings, the method more generally provides approximate conditional multivariate probability computations. Suppose that T has an intractable multivariate distribution with known joint moment generating function (m.g.f.). The unknown multivariate density of $T = (U, V)$ can be tilted and nested within an exponential family. Then, any multivariate conditional probability for U given V may be computed with this method so long as the joint m.g.f. of T is computable. The determination of analytical approximations for the (untilted) cumulative distribution function of U given V or other conditional multivariate probability computations have been long-standing problems in saddlepoint methods. Our solution is to apply an efficient Monte Carlo method to integrate the joint saddlepoint density of T, thus providing approximate conditional probabilities along with confidence intervals for the Monte Carlo aspect.

The first step in the method is to replace the intractable density for the sufficient statistic T by the normalized saddlepoint density approximation derived by Daniels (1954). Integration with respect to the saddlepoint density rather than the intractable true density typically results in an approximation which is essentially exact for many practical purposes. A partial explanation is that the normalized saddlepoint density approximation has an error of $O(n^{-3/2})$ (that is, third-order accuracy) on normal deviation regions with respect to increasing sample size. In contrast, the asymptotic normal approximation is only accurate to first-order or $O(n^{-1/2})$. However, saddlepoint approximations often have an accuracy with small sample sizes that defy any asymptotic explanation (see, for example, Butler et al., 1992).

In many applications the sufficient statistic is of high dimension and so exact numerical evaluation of integrals, even with respect to the tractable saddlepoint density, is not feasible. Thus, we propose instead a Monte Carlo integration procedure in which realizations of the sufficient statistics are generated from a crude approximate density based on asymptotic normality. A Monte Carlo estimate is then constructed using importance weights that are the ratio of the saddlepoint density to the crude approximate density. The use of an initial crude approximation to calculate integrals with respect to the highly accurate saddlepoint density motivates the name "simulation-assisted saddlepoint" (SAS) approximation used in the title.

We apply the SAS method in the context of inference based on independent samples from univariate gamma distributions using the aircraft air conditioner data from Cox & Snell (1981), and the cancer survival data of Cameron & Pauling (1978). Even though inference based on sampling from gamma distributions is clearly of practical relevance, there is surprisingly very little literature on exact computation in this setting. A second application uses the exponential tilting idea to approximate a complex conditional distribution arising in the context of the multivariate gamma distribution. The examples illustrate that a) the SAS method can be more efficient computationally even when direct simulation is possible; b) parameter transformation can dramatically improve the computational efficiency; c) first order asymptotic approximations can be inaccurate; and d) the SAS method can solve previously intractable problems. Specifically, all of the conditional probability computations presented are intractable by other methods.

The remainder of the article is organized as follows. Daniels' saddlepoint density formula and its relationship to Barndorff-Nielsen's p^*-formula are discussed in Section 2. Monte Carlo estimation via importance sampling is discussed in Section 3. The SAS approximation method is illustrated in Section 4 for inferences concerning gamma and normal populations. The use of exponential tilting to address the more general context is discussed in Section 5 and illustrated using the multivariate gamma distribution. We conclude in Section 6 with a discussion of an alternative Markov chain Monte Carlo (MCMC) integration method which could be used as a substitute for importance sampling when it is inefficient or numerically unstable.

Throughout the article a star superscript denotes a saddlepoint approximation and a tilde overscore denotes a Monte Carlo estimate. The letter p is used as a generic notation for a probability density. Furthermore, we allow densities to be expressed in terms of any parameterization. Thus, for example, $p(t;\theta)$ and $p(t;\mu)$ denote the density of t, using canonical and mean parameterizations respectively, where $\mu = E(T) = \partial b(\theta)/\partial\theta$.

2 Saddlepoint density formulas

Let $\hat\theta$ be the maximum likelihood estimate (m.l.e.) of the canonical parameter which satisfies the one-to-one relationship $\partial b(\hat\theta)/\partial\hat\theta = t$. Daniels derives a saddlepoint approximation to (2) given by

$$p^*(t;\theta) = (2\pi)^{-k/2}\left|\frac{\partial^2 b(\hat\theta)}{\partial\hat\theta\partial\hat\theta^T}\right|^{-1/2}\frac{\exp\{\theta^T t - b(\theta)\}}{\exp\{\hat\theta^T t - b(\hat\theta)\}}\,, \tag{4}$$

where the dependence on t is also implicit in $\hat\theta$. It is generally advisable to normalize the saddlepoint density if possible so that it integrates to one. An approximation to the conditional density of U given $V = v$ is obtained by fixing v in (4) and normalizing so that the resulting function is a probability density. This procedure is equivalent to normalizing the double saddlepoint density approximation proposed by Barndorff-Nielsen (1983).

2.1 Reparameterization

The sufficient statistic t is the m.l.e. for the mean parameter, $\mu = E(T)$. A saddlepoint formula for the m.l.e. of any one-to-one reparameterization of μ can be obtained via jacobian transformation of (4). In particular, since $\partial t/\partial\hat\theta = \partial^2 b(\hat\theta)/\partial\hat\theta\partial\hat\theta^T$, it follows that

$$p^*(\hat\theta;\theta) = (2\pi)^{-k/2}\left|\frac{\partial^2 b(\hat\theta)}{\partial\hat\theta\partial\hat\theta^T}\right|^{1/2}\frac{\exp\{\theta^T t - b(\theta)\}}{\exp\{\hat\theta^T t - b(\hat\theta)\}}\,, \tag{5}$$

where the dependence on $\hat\theta$ is also implicit in t. More generally, let $\tau = g(\theta)$ be an arbitrary one-to-one reparameterization of the canonical parameter. Then jacobian transformation of (5) reveals

$$p^*(\hat\tau;\theta) = (2\pi)^{-k/2}\left|\frac{\partial^2 b(\hat\theta)}{\partial\hat\theta\partial\hat\theta^T}\right|^{1/2}\left|\frac{\partial\hat\theta}{\partial\hat\tau}\right|\frac{\exp\{\theta^T t - b(\theta)\}}{\exp\{\hat\theta^T t - b(\hat\theta)\}}\,. \tag{6}$$

Let $j(\tau)$ denote the expected information with respect to τ; that is, the expected value of minus the second derivative of the loglikelihood function with respect to τ. Then the relation

$$j(\tau) = \frac{\partial \theta^T}{\partial \tau} j(\theta) \frac{\partial \theta}{\partial \tau^T} \tag{7}$$

implies that (6) can be re-expressed more concisely as,

$$p^*(\hat{\tau}; \tau) = (2\pi)^{-k/2} \, |j(\hat{\tau})|^{1/2} \, \frac{f(x; \tau)}{f(x; \hat{\tau})} \,, \tag{8}$$

which is Barndorff-Nielsen's p^*-formula in the exponential family context.

2.2 p^* probabilities

An approximation to the probability $P(\hat{\tau} \in A; \tau)$, where A is an arbitrary (measurable) set in R^k is given by

$$P^*(\hat{\tau} \in A; \tau) = \frac{\int 1\{\hat{\tau} \in A\} p^*(\hat{\tau}; \tau) d\hat{\tau}}{\int p^*(\hat{\tau}; \tau) d\hat{\tau}} \,. \tag{9}$$

Now, let $\tau = (\chi, \omega)$ and, for each fixed ω, let A_ω denote the set $\{\chi : (\chi, \omega) \in A\}$. Then, since the saddlepoint approximation to the conditional density of $\hat{\chi}$ given $\hat{\omega}$ is proportional to (8), an approximation to the conditional probability $P(\hat{\tau} \in A | \hat{\omega}; \tau)$ is given by

$$P^*(\hat{\chi} \in A_{\hat{\omega}} | \hat{\omega}; \tau) = \frac{\int 1\{\hat{\chi} \in A_{\hat{\omega}}\} p^*(\hat{\chi}, \hat{\omega}; \tau) d\hat{\chi}}{\int p^*(\hat{\chi}, \hat{\omega}; \tau) d\hat{\chi}} \,. \tag{10}$$

Note that the approximation given in (9) presumes integrability of the p^*-formula.

3 Monte Carlo Integration

If it is possible to simulate values, $\hat{\tau}_1, \ldots, \hat{\tau}_N$, directly from (8), then a Monte Carlo estimate of (9) is given by

$$\tilde{P}^* = N^{-1} \sum_{r=1}^{N} 1\{\hat{\tau}_r \in A\} \,.$$

3.1 Importance sampling approximation

If direct simulation from (8) is not possible, we can construct an indirect Monte Carlo estimate using the asymptotic normal distribution as an importance distribution as follows.

Simulate an i.i.d. sample, $\hat{\tau}_1, \ldots, \hat{\tau}_N$, from the asymptotic normal density $\phi(\hat{\tau}; \tau, j(\tau)^{-1})$. Then

$$\tilde{P}^* = \frac{\sum_{r=1}^N 1\{\hat{\tau}_r \in A\} w_r}{\sum_{r=1}^N w_r}, \tag{11}$$

where $w_r = p^*(\hat{\tau}_r; \tau)/\phi(\hat{\tau}_r; \tau, j(\tau)^{-1})$ is the rth importance weight. The ratio estimate $\tilde{P}^*$ given in (11) converges to the saddlepoint approximation P^* given in (9), as N increases, by the Strong Law of Large Numbers, provided the p^*-density is integrable. This condition is equivalent to the importance weights having finite expectation.

Monte Carlo estimates of conditional probabilities can be constructed in a similar manner to those just described for unconditional probabilities. The analog of (11) for approximating $P(\hat{\chi} \in A_{\hat{\omega}} | \hat{\omega}; \tau)$ involves simulating an i.i.d. sample $\hat{\chi}_1, \ldots, \hat{\chi}_N$ from the asymptotic normal approximation to the conditional density of $\hat{\chi}$ given $\hat{\omega}$. This approximating conditional density is derived from the normal approximation to the joint density of $\hat{\tau}$ using the standard formulas (see e.g. Mardia et al., 1979, p. 63). Let

$$j(\chi, \omega) = \begin{bmatrix} j_{\chi\chi} & j_{\chi\omega} \\ j_{\omega\chi} & j_{\omega\omega} \end{bmatrix}$$

be a block partition of the expected information matrix for τ and denote the corresponding partition of j^{-1} using superscripts. Then, conditional on $\hat{\omega}$, $\hat{\chi}$ is approximately normal with mean vector and variance-covariance matrix given by $\mu_{\hat{\chi}|\hat{\omega}}$ and $\Sigma_{\hat{\chi}|\hat{\omega}}$ respectively, where

$$\mu_{\hat{\chi}|\hat{\omega}} = \chi + j^{\chi\omega}(j^{\omega\omega})^{-1}(\hat{\omega} - \omega) \tag{12}$$

and

$$\Sigma_{\hat{\chi}|\hat{\omega}} = j^{\chi\chi} - j^{\chi\omega}(j^{\omega\omega})^{-1} j^{\omega\chi}. \tag{13}$$

A ratio estimate of $P(\hat{\chi} \in A_{\hat{\omega}} | \hat{\omega}; \tau)$ then has the same form as (11) with rth importance weight, $w_r = p^*(\hat{\chi}_r, \hat{\omega}; \tau)/\phi(\hat{\chi}_r; \mu_{\hat{\chi}|\hat{\omega}}, \Sigma_{\hat{\chi}|\hat{\omega}})$.

There is considerable simplification in (12) and (13) when inference concerns a portion of the canonical parameter. Specifically, the canonical parameter, ψ, is orthogonal to its complementary mean parameter, $\nu = E(V)$, in the sense that $j_{\psi\nu} = 0$. This can be verified directly using (7). In fact, in terms of the block partition in the canonical parameterization $\theta = (\psi, \lambda)$, we have

$$j(\psi, \nu) = \begin{bmatrix} j_{\psi\psi\cdot\lambda} & 0 \\ 0 & j_{\lambda\lambda}^{-1} \end{bmatrix}$$

where $j_{\psi\psi\cdot\lambda} = j_{\psi\psi} - j_{\psi\lambda} j_{\lambda\lambda}^{-1} j_{\lambda\psi}$. It follows that the mean and variance formulas used in the normal approximation to the conditional density simplify to $\mu_{\hat{\psi}|v} = \psi$ and $\Sigma_{\hat{\psi}|v} = j_{\psi\psi\cdot\lambda}^{-1}$.

3.2 Monte Carlo error assessment

The Monte Carlo error in (11) can be assessed using the asymptotic variance formula derived using the delta method,

$$\mathrm{var}(\tilde{P}^*) \approx \frac{1}{NE(w)^2} E\left\{[w(I - P^*)]^2\right\} = \frac{\sigma^2}{N}\,, \tag{14}$$

say, where $I = 1\{\hat{\tau} \in A\}$. An estimate of σ based on the N simulations is

$$\tilde{\sigma} = \frac{1}{\bar{w}} \sqrt{\frac{1}{N} \sum_{r=1}^{N} \left[w_r(I_r - \tilde{P}^*)\right]^2}\,.$$

This estimate can be updated after every simulation by keeping track of the relevant sums. To control the Monte Carlo error in practice we can continue the simulations until the absolute error is estimated to be less than a prespecified level, ϵ, with $100(1-\alpha)\%$ confidence; i.e. until $\widetilde{AE} = |z_{\alpha/2}|\tilde{\sigma}/\sqrt{N} \leq \epsilon$, where z_α denotes the α-quantile of the standard normal distribution. Alternatively, one could use a relative error criterion, $\widetilde{RE} = \widetilde{AE}/\tilde{P}^* \leq \epsilon$.

We mentioned in Subsection 3.1 that the importance weights must have finite expectation for $\tilde{P}^*$ to converge to P^* almost surely as the simulation size, N, increases. Existence of a second moment is required for the Central Limit Theorem to also hold and for $\tilde{P}^*$ to converge at rate $N^{-1/2}$. In the example considered in Subsection 4.1, $E(w^2) = \infty$ with a normal importance distribution and so the variance formula (14) is not valid. In practice, this translates in erratic convergence properties. The situation is often rectified by using a heavier-tailed multivariate-t candidate, with the same mean and scale matrix and corresponding importance weight $w_r \propto p^*(\hat{\tau}_r; \tau)/t_f(\hat{\tau}_r; \tau, j(\tau)^{-1})$, where t_f denotes a multivariate-t density with f degrees of freedom. We have found that a low value, such as $f = 5$, works well in a wide range of examples.

3.3 Choice of Parameterization

The efficiency of all numerical integration procedures, including importance sampling, vary with the parameter of integration $\hat{\tau}$. Roughly speaking, importance sampling achieves its greatest efficiency when the importance distribution t_f for $\hat{\tau}$ is an accurate approximation for its saddlepoint density p^*. An efficient reparametrization to $\hat{\tau}$ from canonical $\hat{\theta}$ should lead to $p^*(\hat{\tau}; \tau) \approx \phi\left\{\hat{\tau}; \tau, j(\tau)^{-1}\right\}$ over a range of $\hat{\tau}$ and τ-values. Our experience is that the choice of τ as a variance-stabilized parameterization works very well. Fraser (1988) provides partial confirmation to this choice for a one-dimensional parameter. He notes that the variance-stabilized parameter $\hat{\tau}$ can have a saddlepoint density $p^*(\hat{\tau}; \tau)$ with the shape of a normal density $\phi\left\{\hat{\tau}; \tau, 1\right\}$ over a range of τ-values. Accordingly, variance-stabilized parameter integration is considered in our approach whenever possible. Exact variance stabilization is sometimes difficult to implement so approximate stabilization may be used in practice. Multivariate covariance stabilization is generally difficult, so for $k \geq 2$, one might settle for variance stabilization of the individual components marginally. Inevitably the choice of parameterization in practical applications is a compromise between the efficiency of the parametrization and its difficulty of implementation using the importance sampling scheme.

4 Examples

In Subsections 4.1-4.3 we illustrate the SAS method for conducting likelihood based inferences when sampling from several independent gamma distributions. We use the notation $Ga(\alpha, \beta)$ for a gamma distribution with canonical shape and scale parameters, α and β respectively; that is,

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-\beta x\}, \quad x > 0. \tag{15}$$

The statistics, $U = \log \prod X_k$ and $V = \sum X_k$, based on a random sample from this distribution, are jointly sufficient for α and β.

Inferences concerning the shape parameter can be carried out without knowledge of the scale if they are based on the conditional distribution of U given V using the SAS methods described in Sections 2 and 3. For example, a P-value for a test of $H_0 : \alpha = \alpha_0$ based on the likelihood-ratio statistics can be obtained via integration of the p^* formula for the joint density of $(\hat{\alpha}, V)$ given in equation (10). Similarly, inference concerning β can be conducted via integration of the p^* formula for the joint density of $(\hat{\beta}, U)$. An extension of the latter is illustrated in Subsection 4.3, where we test homogeneity of scale parameters for several gamma distributions with the same shape.

Clearly, inferences based on the conditional distribution of U given V are equivalent to those based on the conditional distribution of $Z = U - n\log(V/n) = n\log(g/a)$ given V, where g and a are the geometric and arithmetic sample means respectively. However, note that since $(X_1/V, \ldots, X_n/V)$ has a symmetric Dirichlet density independent of V, the conditional distribution of Z given V is the same as its marginal distribution. This distribution has exponential form

$$p(z; \alpha) = \exp\{z\alpha - b(\alpha) - d(z)\}, \tag{16}$$

where the function $d(z)$ is intractable but $b(\alpha)$ is easily derived as

$$b(\alpha) = n\log\Gamma(\alpha) - \log\Gamma(n\alpha) + n\alpha\log n, \tag{17}$$

(see Jensen, 1986, Section 2).

These observations suggest an alternative approach to inference concerning α. Rather than constructing a test using the *full* likelihood, obtained as a product of gamma densities of the form given in (15), one can use the *conditional* likelihood obtained from (16). In the one-sample case, likelihood ratio statistics constructed from the full and conditional likelihoods are both monotone functions of z and hence lead to equivalent one-sided tests. However, the two approaches do not yield equivalent two-sided tests in the one-sample case and they are generally different in the multi-sample case discussed in Subsections 4.1 and 4.2 below.

The conditional m.l.e. for α is the solution of the equation $z = b'(\alpha)$. The density (16) and hence that of the conditional m.l.e. can be approximated directly by the p^*-formula in (5) with α and z in place of θ and t. Integrability of the p^*-formula for the conditional m.l.e. is shown by Booth et al. (1999, Appendix II). In contrast, the *full* m.l.e. is a solution of $z = c'(\alpha)$, where

$$c(\alpha) = n\left\{\alpha + \log\Gamma(\alpha) - \alpha\log\alpha\right\} . \tag{18}$$

As noted earlier, approximate conditional inference using the full m.l.e. involves the use of the p^*-formula for the joint density of $(\hat{\alpha}, V)$. Integrability of the joint p^*-formula can be shown using arguments similar to those in Booth et al.. Implementation of the SAS method is straightforward whether the full or conditional m.l.e. for α is used for inference. However, the conditional m.l.e. for α is known have superior repeated sampling properties to the full m.l.e. for small and moderate sample sizes (Yanagimoto, 1988) and hence we will emphasize methods based on the conditional m.l.e. in what follows.

Before proceeding to the numerical examples, we note that the conditional likelihood approach described above for inference concerning α is not feasible for inference about β because the conditional density of V given U is not tractable.

4.1 Testing simultaneous exponentiality

Suppose that data consist of independent random samples from m gamma distributions and consider the hypothesis that the distributions are all exponential; i.e. $\alpha_1 = \cdots = \alpha_m = 1$. As indicated above, it is appropriate to base inferences concerning the shape parameters on the marginal distribution of the statistics, $Z_i = n_i \log(g_i/a_i)$, $i = 1, \ldots, m$, with densities of the form given in (16). The likelihood ratio test based on the product of the densities for the Z_i's is given by

$$-2\log\Lambda = 2\sum_{i=1}^{m}\left\{z_i(\hat{\alpha}_i - 1) - b_i(\hat{\alpha}_i) + b_i(1)\right\} , \tag{19}$$

where the m.l.e.'s satisfy the nonlinear equations, $z_i = b_i'(\hat{\alpha}_i)$, $i = 1, \ldots, m$, which must be solved iteratively. The exact P-value for this test, $P = P(\Lambda \leq \Lambda_{obs})$, can be approximated by brute force Monte Carlo involving simulation of exponential samples and calculation of the m.l.e.'s for each simulated data set. Alternatively, a SAS approximation can be constructed using the p^*-density for each $\hat{\alpha}_i$ combined with a Student-t importance distribution. Booth

Test Statistic	$-2\log\Lambda$	DF	Chisquared	Brute Force	SAS
Conditional	15.56	10	.113	.126	.128
Full	15.59	10	.112	.165	.168

Table 1: Simulation-assisted saddlepoint and direct Monte Carlo approximations of exact conditional P-values for testing simultaneous exponentially of the aircraft, air-conditioner data

et al. show that the use of a normal importance density in this setting results in importance weights with infinite second moments. They then derive a rejection sampler using a Student-t candidate for sampling directly from this particular p^*-density. The degrees of freedom used in this context is the largest integer value for which the importance weights have a finite second moment.

For a numerical illustration of the methods consider the aircraft air conditioner data in Cox & Snell. The data consist of the times between failures of air conditioners in $m = 10$ different aircraft. The sample sizes are not particularly small, varying from 9 to 30, with a total sample size of 199 and so one might expect the usual chisquared approximation to work well. The results are summarized in Table 1.

The brute force Monte Carlo P-value was based on 18,500 simulated datasets, the number required for 5% relative error accuracy with 99% confidence (i.e. $\epsilon = .05$ and $\alpha = .01$). Simulation-assisted saddlepoint approximation using ten independent Student-t distributions with 3 degrees of freedom required 79,000 simulated sets of conditional m.l.e.'s to attain the same relative accuracy. Note that the SAS value is extremely accurate, although the chisquared approximation would probably be adequate for practical purposes in this case. Also, even though the SAS method required more simulated values of the conditional m.l.e.'s, the total computational effort was about half that of brute force simulation because in the latter case the m.l.e.'s have to be computed for each data set.

One might expect normal approximation to the null distribution of the conditional m.l.e. to be inaccurate in this setting because the support of the distribution is bounded below at zero. In particular, this means that negative simulated values of the $\hat{\alpha}_i$'s must be discarded. Problems of this nature can often be alleviated by transforming to a different parameter scale. In the present setting, use of a log transformation, $\eta_i = \log \alpha_i$, results in a marked increase in the efficiency of the SAS method. Specifically, the 5% relative error convergence criterion was met with only 25,000 simulated sets of conditional m.l.e.'s with a final value of $\tilde{P}^* = .121$. Note that use of the log transformation requires a jacobian adjustment to the p^*-formula of $|\partial\hat{\alpha}/\partial\hat{\eta}| = \prod_i \hat{\alpha}_i$, as indicated in equation (6).

The corresponding results based on the full likelihood-ratio statistic (which has the same form as in (19) with $c_i(\alpha_i)$ in place of $b_i(\alpha_i)$) are also given in Table 1. Similar comments concerning efficiency apply in this case. Note that, although the full and conditional loglikelihood-ratio statistics are almost identical, the exact conditional P-values are quite different.

The next two subsections consider tests of hypotheses involving nuisance parameters that cannot be handled by brute force simulation.

4.2 Testing homogeneity of gamma shape parameters

Now consider the hypothesis, $\alpha_1 = \cdots = \alpha_m$, of homogeneity of shape parameters for m gamma populations. As in the previous subsection, inference is based on the conditional likelihood obtained as the joint distribution of the statistics, $Z_1, \ldots, Z_m$. However, in this case the unspecified common shape is a nuisance parameter. To frame the problem in the general notation established in the earlier sections, let $\alpha_i = \lambda + \psi_i$ for $i = 1, \ldots, m$, with the identifiability constraint, $\psi_1 = 0$ so that $\lambda = \alpha_1$. Then, we wish to assess the hypothesis of homogeneity, $\psi_2 = \cdots = \psi_m = 0$, using the likelihood-ratio test with null probabilities computed from the conditional distribution of $\hat{\psi}_2, \ldots, \hat{\psi}_m$ given $z. = \sum z_i$, the sufficient statistic corresponding to the nuisance parameter λ. Note that $z.$ is the m.l.e. of the mean parameter complementary to the canonical parameter ψ and hence the simplifications described at the end of Subsection 3.1 apply. Since $\hat{\psi}_i = \hat{\alpha}_i - \hat{\alpha}_1$ and $z. = \sum b_i'(\hat{\alpha}_i)$, it follows that the jacobian adjustment to the p^*-formula in (6) due to the transformation $(\hat{\alpha}_1, \ldots, \hat{\alpha}_m) \to (\hat{\psi}_2, \ldots, \hat{\psi}_m, z.)$ is given by $\{\sum b_i''(\hat{\alpha}_i)\}^{-1}$. The likelihood ratio statistic has the same form as in the test of exponentiality except that the common unit shape is replaced by the (conditional) m.l.e. of the unspecified nuisance parameter λ. This value must be calculated for each simulated value of $\hat{\psi}_2, \ldots, \hat{\psi}_m$ and fixed $z.$ using an iterative procedure and so the SAS method involves more computation than in the previous case. The observed likelihood ratio statistic is $-2 \log \Lambda = 15.44$ in this problem resulting in a chisquared P-value approximation of .080 based on 9 degrees of freedom.

As in the test of exponentiality, log transformation of the shape parameter increases the efficiency of the SAS approximation. In fact, the method failed to converge using the canonical parameterization even after five million simulations. The value of $\tilde{P}^*$ after each 10,000 simulations is plotted in Figure 1. The plot indicates occasionally large adjustments to $\tilde{P}^*$ corresponding to extreme importance weights, which explains the slow convergence or failure of the method. In contrast, the SAS approximation utilizing a log transformation, $(\hat{\alpha}_1, \ldots, \hat{\alpha}_m) \to (\hat{\eta}_2, \ldots, \hat{\eta}_m, z.)$, is well behaved, converging after 1.6 million simulations to $\tilde{P}^* = .131$. This suggests that the chisquared approximation is not very accurate, although that approximation is for the unconditional P-value and so is not directly comparable. Another reason for skepticism about the chisquared approximation is that the data should be more compatible with the homogeneity hypothesis (as indicated by a larger P-value) than the more restrictive exponentiality hypothesis discussed above.

Calculation of the SAS approximation in this problem requires the jacobian of the transformation from $\theta = (\alpha_1, \ldots, \alpha_m)$ to $\tau = (\eta_2, \ldots, \eta_m, \nu)$, where $\nu = E(z.) = \sum b_i'(\alpha_i)$. The inverse of this jacobian is given by

$$\frac{\partial \tau}{\partial \theta^T} = \begin{bmatrix} 1/\alpha_2 & 0 & 0 & \cdots & 0 \\ 0 & 1/\alpha_3 & 0 & \cdots & 0 \\ & & \vdots & & \\ 0 & \cdots & 0 & 1/\alpha_m & 0 \\ b_2''(\alpha_2) & b_3''(\alpha_3) & \cdots & b_m''(\alpha_m) & b_1''(\alpha_1) \end{bmatrix}$$

from which we obtain $|\partial\hat{\theta}/\partial\hat{\tau}| = \prod_{i=2}^m \hat{\alpha}_i / b_1''(\hat{\alpha}_1)$ for use in (6). The mean of the importance distribution is given by the asymptotic approximation $E(\hat{\eta}_i|z.) \approx \log \hat{\alpha}_0$, where $\hat{\alpha}_0$ is the estimated common shape under the null hypothesis. The $\hat{\alpha}_i$'s are independent with asymptotic variance $1/b_i''(\alpha)$ and hence the asymptotic variance of $\hat{\tau}$ is given by

$$\left(\frac{\partial \tau^T}{\partial \theta}\right) \mathrm{diag}\left(\frac{1}{b_i''(\alpha_i)}\right) \left(\frac{\partial \tau}{\partial \theta^T}\right)_{\alpha_i = \alpha} = \begin{bmatrix} 1/\alpha^2 b_2''(\alpha) & 0 & 0 & \cdots & 1/\alpha \\ 0 & 1/\alpha^2 b_3''(\alpha) & 0 & \cdots & 1/\alpha \\ & & \vdots & & \\ 0 & \cdots & 0 & 1/\alpha^2 b_m''(\alpha) & 1/\alpha \\ 1/\alpha & 1/\alpha & \cdots & 1/\alpha & \sum b_i''(\alpha) \end{bmatrix}.$$

This implies an approximate conditional variance for $\hat{\eta} = (\hat{\eta}_2, \ldots, \hat{\eta}_m)$ given by

$$\mathrm{var}(\hat{\eta}|z.) \approx \mathrm{diag}\left(\frac{1}{\alpha^2 b_i''(\alpha)}\right)_{i=2}^m - \frac{1}{\alpha^2 \sum b_i''(\alpha)} 11^T.$$

The covariance of the importance distribution replaces α with $\hat{\alpha}_0$.

4.3 Testing homogeneity of gamma scale parameters

Assuming a common shape parameter α for all m groups, consider testing $H_0 : \beta_1 = \cdots = \beta_m$. With an alternative parameterization $(\gamma_2, \ldots, \gamma_m, \beta, \alpha)$, such that $\beta_1 = \beta$ and $\beta_i = \beta + \gamma_i$, $i = 2, \ldots, m$, this is equivalent to testing $H_0 : \gamma_2 = \ldots = \gamma_m = 0$. Under the null hypothesis,

$$v. = \sum_{i=1}^m v_i = \sum_{i=1}^m \sum_{k=1}^{n_i} x_{ki} \qquad \text{and} \qquad u = \log \prod_{i=1}^m \prod_{k=1}^{n_i} x_{ki}$$

are sufficient for α and β. Allowing the scale parameter to differ between groups introduces the additional sufficient statistics, v_i, $i = 2, \ldots, m$. In this case, it is more convenient to sample the v_i than the $\hat{\gamma}_i$ due to their simpler form. However, we note the two are related as

$$\hat{\gamma}_i = \hat{\alpha}\left(\frac{n_i}{v_i} - \frac{n_1}{v_1}\right), \; i = 2, \ldots, m. \tag{20}$$

Thus, the conditional P-value is calculated based on the approximation to the conditional density of $(v_2, \ldots, v_m)$ given $v.$ and u. Since these are the sufficient statistics, we may use

Daniels saddlepoint formula directly as given in (4), and thus avoid further computations of jacobians which are needed for transformations. Note however that it is necessary to calculate the m.l.e. for the common shape, α, for each simulated set of sufficient statistics. Then, $\hat{\beta} = (n_1\hat{\alpha})/v_1$ and the $\hat{\gamma}_i$'s can be determined from (20).

A SAS approximation to the conditional P-value can be obtained by simulating from the multivariate normal approximation to the conditional distribution which has mean components, $n_i\hat{\alpha}_0/\hat{\beta}_0 = n_i v./n.$, $i = 2, \ldots, m$ with $n. = \sum_{i=1}^m n_i$, and variance-covariance matrix with (i,j)th element

$$\frac{\hat{\alpha}_0}{\hat{\beta}_0^2}\left[n_i\mathbf{1}\{i=j\} - \frac{n_i n_j}{n.}\right] = \frac{v_.^2}{n.^2\hat{\alpha}_0}\left[n_i\mathbf{1}\{i=j\} - \frac{n_i n_j}{n.}\right] \qquad i,j = 2,\ldots,m\,.$$

For the aircraft air-conditioner data discussed above, the observed loglikelihood ratio statistic for testing homogeneity of scales is $-2\log\Lambda_{obs} = 19.169$, resulting in an asymptotic P-value of 0.024 based on the chisquared approximation with 9 degrees of freedom. The unconstrained m.l.e.'s for the scale parameters in this problem are given in Table 2. The common shape estimate is $\hat{\alpha} = 1.007$. Calculation of the SAS approximation to the conditional P-value required 141,000 importance samples to obtain a 5% relative error with 99% confidence and resulted in the value 0.029.

Aircraft number	1	2	3	4	5
Sample size	23	29	15	14	30
$\hat{\beta}_i$	0.0105	0.0121	0.0083	0.0077	0.0169

Aircraft number	6	7	8	9	10
Sample size	27	24	9	12	16
$\hat{\beta}_i$	0.0131	0.0157	0.0051	0.0097	0.0123

Table 2: Sample sizes and m.l.e.s of scale for each aircraft

For a second illustration, consider data from Cameron & Pauling which examines the effect of supplemental vitamin C on the survival of ovarian cancer patients. The data consist of survival times, in days, of six women who were treated with vitamin C. Each woman was matched with ten control patients. The average of these 10 survival times is also provided. For the purposes of this paper, we have ignored the possible correlation due to matching.

Assuming that the distribution of survival times is gamma and that the individual patients share a common shape parameter, a significant difference in the scale parameters will demonstrate a treatment effect. In particular, assume that the scale parameter for an individual control patient is β and for the treated patient is $\beta + \gamma$. The distribution of the survival time for a treated patient is then gamma with parameters α and $\beta + \gamma$, while the distribution for the averages of the controls' survival times is gamma with parameters 10α and 10β. Under the alternative hypothesis, the estimated mean survival time of a woman taking supplemental vitamin C is 884.36 days while the estimated mean survival time of an individual not taking vitamin C is 377.50 days. The likelihood ratio statistic is 3.84, yielding an approximate unconditional P-value of .050, found using the chisquared distribution with 1 degree of freedom. In contrast, the SAS approximation of the conditional P-value was 0.010 requiring 42,000 importance samples to obtain a 5% relative accuracy with 99% confidence.

4.4 One-way ANOVA under normality assumptions

It is interesting to examine how well the SAS method works in the familiar one-way ANOVA setting in which the samples are assumed to be drawn from normal populations, $N(\mu_i, \sigma_i^2)$, $i = 1, \ldots, k$. Here, the canonical parameters are $(\psi_i, \lambda_i) = (1/\sigma_i^2, \mu_i/\sigma_i^2)$, with corresponding sufficient statistics $(\sum_{j=1}^{n_i} x_{ij}^2, \sum_{j=1}^{n_i} x_{ij})$. The likelihood ratio statistic for testing homogeneity of variances is proportional to $\prod_{i=1}^{k}(W_i/W)^{n_i/2}$, where W_i is the sum of square deviations from the mean for the ith sample and $W = \sum_i W_i$ is the total within groups sum of squares (Mardia et al., 1979, page 140). Since, the random vector $(W_i/W, \ldots, W_i/W)$ has a Dirichlet distribution independent of the sample sums, $x_{i\cdot}$, $i = 1, \ldots, k$, the conditional distribution of the likelihood ratio statistic, Λ, is the same as its unconditional distribution. The SAS method therefore leads to an extremely accurate approximation to the exact null distribution of the likelihood ratio statistic. However, in this case a more direct approach is possible which does not involve any simulation. Since the m.g.f. of $\log \Lambda$ is known, the null distribution of Λ can be approximated using the Lugannani & Rice saddlepoint distribution function approximation (see Booth et al., 1996).

The hypothesis that is typically of most interest in this setting is that of homogeneity of means assuming equal variances. The likelihood ratio test is a monotone function of the ratio, B/W, of between to within groups sums of squares and therefore equivalent to the usual F-test. For $i = 1, \ldots, k$, let $\mu_i = \mu + \delta_i$ with the identifiability constraint $\delta_1 = 0$. The canonical parameters of interest in this case are then $\psi_i = \delta_i/\sigma^2$, $i = 2, \ldots, k$, while $\lambda_1 = \mu/\sigma^2$ and $\lambda_2 = 1/\sigma^2$ are nuisance parameters. The relevant conditional distribution for inference about the ψ_i's is that of $(x_{2\cdot}, \ldots, x_{k\cdot})$ given $x_{\cdot\cdot}$ and $\sum_{ij} x_{ij}^2$ or equivalently given $x_{\cdot\cdot}$ and the total corrected sum of squares, $T = B + W$. The saddlepoint density approximation for $(x_{1\cdot}, \ldots, x_{k\cdot}, T)$ and hence for $(x_{2\cdot}, \ldots, x_{k\cdot}, x_{\cdot\cdot}, \sum_{ij} x_{ij}^2)$ is exact up to a renormalization constant under the null hypotheses of homogeneity of means. Since, B/W is jointly independent of $x_{\cdot\cdot}$ and T under the null hypotheses, it follows that the SAS method is exact in this setting except for Monte Carlo error.

5 Tilting

In this section, we calculate unconditional and the more difficult conditional probabilities for a random vector T whose multivariate density is either unknown or intractable and not necessarily from an exponential family. We alternatively suppose that its joint cumulant generating function $b(\theta) = \log E\{\exp(\theta^T T)\}$ exists in an open neighborhood of the origin. If $f_T(t)$ is the unknown continuous multivariate density, then it may be tilted and nested within an exponential family

$$f_T(t;\theta) = f_T(t)\exp\{\theta^T t - b(\theta)\}\,, \tag{21}$$

which is of the form given in (2) with $d_T(t) = -\log f_T(t)$. Thus, computation of unconditional and conditional probabilities with respect to the unknown density f_T can be carried out using the SAS method.

To illustrate the power of this method we consider the computation of probabilities for previously intractable conditional distributions arising in the context of a multivariate gamma distribution. Let W be a Wishart$_k(n,\Sigma)$ variate, where the scale matrix $\Sigma = (\sigma_{ij})$ is positive definite. Then, the diagonal elements of W have a multivariate Gamma$_k(\frac{n}{2},\Sigma)$ distribution as described in Johnson & Kotz (1972, Chapter 40). The marginal densities are Gamma$(\frac{n}{2},\sigma_{ii})$, $i = 1,\ldots,k$ and the components are independent when Σ is diagonal. Jensen (1985) lists some applications of this distribution. The square roots of the components form a multivariate Rayleigh distribution that arises in signal detection as discussed in Miller (1975). The distribution has an intractable density which has prohibited its extensive use. However, in reliability there aren't many reasonable distributions on $(0,\infty)^k$ that can be used to model dependencies among the failure times of system components. This distribution becomes an attractive choice to model such dependencies when its probability computations are made tractable through use of the SAS method.

The cumulant generating function of $T = \frac{1}{2}\text{diag}W$ is given by

$$b(\theta) = -\frac{n}{2}\left\{\log|\Sigma^{-1} - \Theta_d| + \log|\Sigma|\right\}\,,$$

where $\theta = (\theta_1,\ldots,\theta_k)$ and Θ_d is the diagonal matrix with ith diagonal entry equal to θ_i. There is no loss of generality in assuming Σ is a correlation matrix. Also, for simplicity, we restrict attention to the exchangeable setting in which $\Sigma = (1-\rho)I + \rho 1 1^T$, where $|\rho| < 1$.

Let $t = (u,v)$ and consider computation of $P\{U \in C|v\}$ for some set C. We calculate this probability by nesting the multivariate gamma density within the exponential family given by (21). Then let $\theta = (\psi,\lambda)$ represent the canonical parameter of $t = (u,v)$. We can now calculate a Monte Carlo estimate of the probability

$$P\{U \in C|v\} = P\{U \in C|v;\psi = 0\}$$

according to (10). The unconstrained saddlepoints can be shown to satisfy the system of equations

$$\hat{\zeta}_i + q(1 - q\hat{\zeta}_.)^{-1}\hat{\zeta}_i^2 = \frac{2t_i}{n} \tag{22}$$

for $i = 1, \ldots, k$, where

$$\hat{\zeta}_i = \left[(1-\rho)^{-1} - \hat{\theta}_i\right]^{-1} ,$$

$\hat{\zeta}_. = \sum_{i=1}^{k} \hat{\zeta}_i$ and

$$q = \rho(1-\rho)^{-1}\left[1 + (k-1)\rho\right]^{-1} .$$

Instead of simulating the u_i or the corresponding saddlepoints $\hat{\psi}_i = \hat{\theta}_i$, $i = 1, \ldots, l$, directly, it is more efficient to simulate

$$\hat{\eta}_i = \ln\left\{(1-\rho)^{-1} - q - \hat{\psi}_i\right\} . \tag{23}$$

The p^*-density for $\{\hat{\eta}_i\}$ is obtained as in (8). The importance sampling distribution is

$$N_l\left[\ln\left\{(1-\rho)^{-1} - q\right\} 1, j_{\eta\eta\cdot\lambda}^{-1}(\psi = 0, \hat{\lambda}_0)\right] ,$$

where

$$j_{\eta\eta\cdot\lambda}(\psi = 0, \hat{\lambda}_0) = \left[(1-\rho)^{-1} - q\right]^2 j_{\psi\psi\cdot\lambda}(\psi = 0, \hat{\lambda}_0)$$

and $\hat{\lambda}_0$ is the value of the saddlepoint corresponding to v obtained from (22) for $i = l{+}1, \ldots, k$ with $\psi = 0$.

Determination of the $\hat{\psi}_i$ from the simulated $\hat{\eta}_i$ follows directly from (23). The remaining saddlepoints $\hat{\lambda}_i$ are again determined from (22) for $i = l+1, \ldots, k$ with $\psi = \hat{\psi}$. Efficient solution to these equations is usually achieved within 5 iterations of the Gauss-Newton algorithm. Then the corresponding simulated values of u can be obtained by solving the same equations (22) for $i = 1, \ldots, l$.

In general, there is no simple way to validate the accuracy of the SAS method in approximating conditional probabilities. However, for the univariate conditional probabilities in which $l = 1$, such that we are conditioning on $k-1$ variables, the Skovgaard approximation may be used as a partial check. We consider SAS estimates for the conditional probability that U is one standard deviation above its mean, i.e., $P\left\{U > \frac{n}{2} + \sqrt{\frac{n}{2}} | v\right\}$. Results are presented in Table 3 with $n = 20$ and the given values of ρ, $k-1$ and v. The SAS computations were computed to have 95% assurance of 1% relative error.

For $l > 1$, the Skovgaard approximation is no longer applicable. Table 4 shows SAS approximations with $n = 20$ to give 95% assurance of 1% relative error in the computation of $P\left\{U > \underline{1}\left(\frac{n}{2} + \sqrt{\frac{n}{2}}\right) | v\right\}$.

$k-1$	ρ	v	SAS	Skovgaard
1	.3	13	0.186	0.177
1	.6	13	0.267	0.240
1	.9	13	0.378	0.349
2	.3	10, 13	0.188	0.180
2	.6	10, 13	0.230	0.206
2	.9	10, 13	0.167	0.149
3	.3	10, 12, 14	0.214	0.194
3	.6	10, 12, 14	0.288	0.251
3	.9	10, 12, 14	0.236	0.216

Table 3: Comparison of the SAS (with $l = 1$) and Skovgaard's approximation

l	$k-l$	ρ	v	SAS
2	1	.6	13	0.101
2	1	.9	13	0.205
4	1	.6	13	0.091
4	1	.9	13	0.347
2	2	.6	10, 13	0.076
2	2	.9	10, 13	0.049
4	2	.6	10, 13	0.637
4	2	.9	10, 13	0.109

Table 4: Simulation-assisted saddlepoint approximations for conditional multivariate Gamma probabilities with $l > 1$

6 Discussion

Generally speaking, the efficiency of the SAS method we have described depends on the similarity between the importance density and the saddlepoint or p^*-density. In many settings the situation can be improved by transformation, as in the case of inferences concerning the shape parameters of several gamma distributions discussed in Subsections 4.1 and 4.2. Often it is a variance stabilizing or some similar transformation that succeeds in achieving this efficiency. Such efficiency, however, can be expected to decline as the dimension of the conditional distribution being integrated increases. In high dimensional problems the method may not ever converge in practice even if theoretical convergence is assured. In such problems an alternative Markov chain Monte Carlo (MCMC) integration method may be more fruitful.

Suppose, for example, one wishes to approximate an integral with respect to the conditional distribution of $\hat{\psi} = (\hat{\psi}_1, \ldots, \psi_l)$ given the sufficient statistic $V = v$. The basic idea of the importance sampling method is to utilize the known asymptotic normal approximation for this conditional distribution. More generally, let $\hat{\psi}_{-i}$ denote the vector $\hat{\psi}$ with its ith element removed. Then, the normal approximation to the distribution of $\hat{\psi}_i$ given $\hat{\psi}_{-i}$ and v (or the corresponding t-approximation) could be used as a candidate in a Hastings-Metropolis algorithm in which i is either chosen at random or according to a systematic sequence at each successive epoch in the Markov chain (see e.g. Besag et al., 1995). More generally, a block version of the algorithm could be employed in which subsets of $\hat{\psi}$ are updated at each stage. We have not investigated this approach since importance sampling was successful in the examples we considered. Also, the issue of assessing Monte Carlo error is more complicated for dependent samplers.

References

Barndorff-Nielsen, O. (1983). On a formula for the distribution of the maximum likelihood estimator. *Biometrika* **70**, 343–365.

Besag, J., Green, P., Higdon, D. & Mengersen, K. (1995). Bayesian computation and stochastic systems (with discussion. *Statist. Sci.* **10**, 3–41.

Booth, J. G., Butler, R. W., Huzurbazar, S. & Wood, A. T. A. (1996). Saddlepoint approximations for P-values of some tests of covariance matrices. *Journal of Statistical Computation and Simulation* **53**, 165–180.

Booth, J. G., Hobert, J. P. & Ohman, P. A. (1999). On the probable error of the ratio of two gamma means. *Biometrika* **86**, 439–452.

Butler, R. W., Huzurbazar, S. & Booth, J. G. (1992). Saddlepoint approximation for the Bartlett-Nanda-Pillai trace statistic in multivariate analysis. *Biometrika* **79**, 705–715.

Cameron, E. & Pauling, L. (1978). Supplemental ascorbate in the supportive treatment of cancer: Reevaluation of prolongation of survival times in terminal human cancer. *Proc. Nat. Acad. Sci. USA* **75**, 4538–42.

Cox, D. R. & Snell, E. J. (1981). *Applied Statistics: Principles and Examples.* Chapman and Hall.

Daniels, H. E. (1954). Saddlepoint approximations in statistics. *The Annals of Mathematical Statistics* **25**, 631–650.

Jensen, D. R. (1985). Multivariate distributions. In *Encyclopedia of Statistical Sciences*, N. L. Johnson & S. Kotz, eds., vol. 6. New York: Wiley.

Jensen, J. L. (1986). Inference for the mean of a gamma distribution with unknown shape parameter. *Scandinavian Journal of Statistics* **13**, 135–151.

Johnson, N. L. & Kotz, S. (1972). *Continuous Multivariate Distributions.* New York: Wiley.

Lugannani, R. & Rice, S. O. (1980). Saddlepoint approximation for the distribution of the sum of independent random variables. *Advances in Applied Probability* **12**, 475–490.

Mardia, K. V., Kent, J. T. & Bibby, J. M. (1979). *Multivariate Analysis.* Academic Press.

Miller, K. S. (1975). *Multivariate Distributions.* New York: Krieger.

Skovgaard, I. M. (1987). Saddlepoint expansions for conditional distributions. *Journal of Applied Probability* **24**, 875–887.

Yanagimoto, T. (1988). The conditional maximum likelihood estimator of the shape parameter in the gamma distribution. *Metrika* **35**, 161–175.

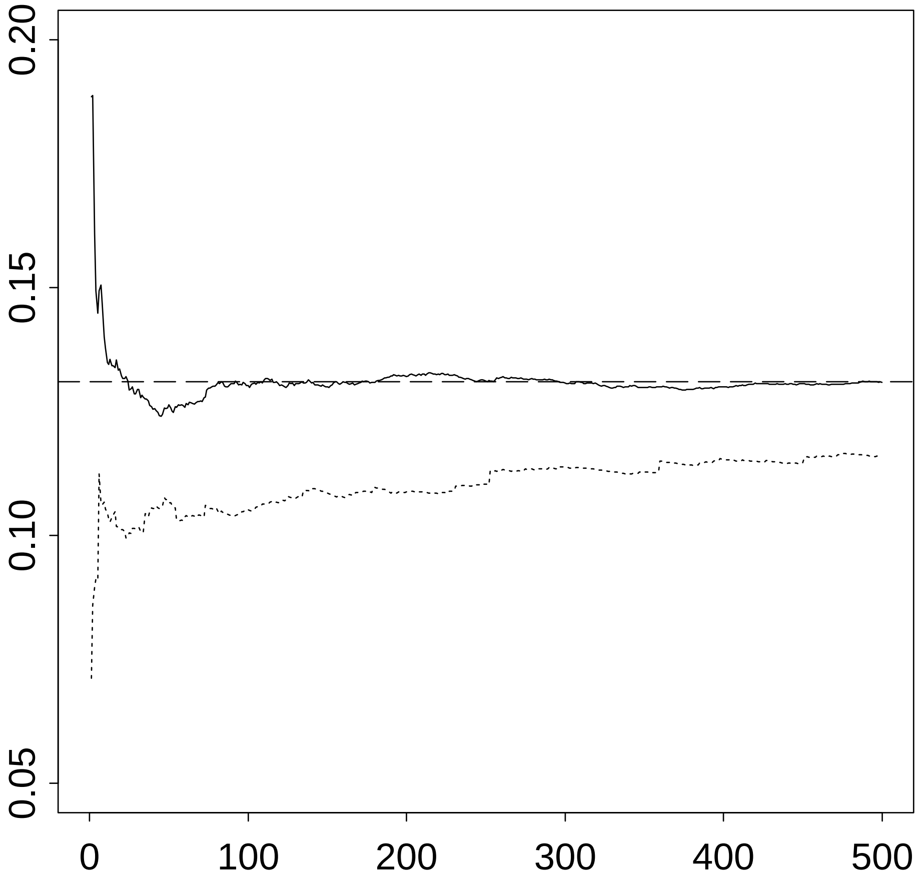

Figure 1: Simulation-assisted saddlepoint approximations of the exact P-value for testing homogeneity of shape parameters. The dashed line tracks the approximation using the canonical parameterization. The solid line tracks the corresponding approximation utilizing a log transformation.