The Annals of Statistics
1998, Vol. 26, No. 1, 398–433

MCMC CONVERGENCE DIAGNOSIS VIA MULTIVARIATE BOUNDS ON LOG-CONCAVE DENSITIES

BY STEPHEN P. BROOKS

University of Bristol

We begin by showing how piecewise linear bounds may be devised, which bound both above and below any concave log-density in general dimensions. We then show how these bounds may be used to gain an upper bound to the volume in the tails outside the convex hull of the sample path in order to assess how well the sampler has explored the target distribution. This method can be used as a stand-alone diagnostic to determine when the sampler output provides a reliable basis for inference on the stationary density, or in conjunction with existing convergence diagnostics to ensure that they are based upon *good* sampler output. We provide an example and briefly discuss possible extensions to the method and alternative applications of the bounds.

1. Introduction. The use of Markov chain Monte Carlo (MCMC) methods has increased dramatically within the statistical community since their introduction by Gelfand and Smith (1990). Essentially, MCMC methods involve simulating Markov chains with particular stationary densities in order to sample indirectly from densities which are impossible to sample from directly. See, for example, Smith and Roberts (1993). There are many important implementational issues to consider for MCMC. These include (among others) the choice of sampler, the number of independent replications to be run and the choice of starting values. Arguably, the most difficult problem is that of deciding when to stop the algorithm with some degree of confidence that a state of equilibrium has been reached or convergence achieved. A number of authors have proposed methods for diagnosing convergence of a particular sampler, based upon the output they produce. See Cowles and Carlin (1996) for a review of recent methods.

However, Asmussen, Glynn and Thorisson (1992) show that there can exist no universally effective means of detecting stationarity, applicable to all stochastic simulations. They suggest that, in order to construct effective methods for detecting stationarity, we need to concentrate on particular classes of problems, and take explicit advantage of the unique characteristics of those classes. This result is directly applicable to Markov chain Monte Carlo, and it can be shown that no diagnostic, based solely upon the output from however many chains, may work for all problems.

Received February 1997; revised September 1997.

AMS 1991 *subject classifications*. Primary 00A72, 65C05; secondary 62F15, 65C10.

Key words and phrases. Bounds, estimation, Markov chain Monte Carlo, simulation, tail probability.

One useful subclass of problems is the set of all those where the target density is log-concave. The set of log-concave densities comprises a large number of those that occur in practice; see Dellaportas and Smith (1993) and Gilks and Wild (1992), for example. We begin with a formal definition of log-concavity.

DEFINITION 1. Given a positive function π, with support $E \subset \mathbb{R}^n$, we say that π is log-concave if $\log \pi$ is concave, that is,

$$\lambda \log \pi(\mathbf{x}) + (1-\lambda)\log \pi(\mathbf{y}) \leq \log \pi(\lambda \mathbf{x} + [1-\lambda]\mathbf{y}) \qquad \forall\ x, y \in E \text{ and } \lambda \in [0,1].$$

An alternative definition can be made in terms of the positive definiteness of the Hessian matrix for $\log \pi$, if this can be assumed to exist.

It is possible to show that, subject to certain regularity conditions, the information matrix for a *generalized linear model* is positive definite. Thus, if we take the canonical link function, where the Hessian and information matrix coincide [McCullagh and Nelder (1989)], the likelihood will be log-concave. Therefore, commonly used models such as the classical linear regression, log-linear Poisson, logistic, probit and complementary log-log models can all be shown to be log-concave. If we choose not to use a canonical link, Wedderburn (1976) provides a number of alternative link functions, which also lead to log-concave likelihoods. Additionally, Polson (1996) shows how the introduction of latent variables to a variety of (possibly multimodal) densities can lead to an augmented density with the log-concavity property. Thus, despite the restriction to unimodal distributions, our method may still be applied to a wide variety of models.

We also note that the class of log-concave models are known to have good convergence properties. For example, Mengersen and Tweedie (1995) show that, on the real line, the Metropolis–Hastings algorithm, under very general conditions, is geometrically ergodic if the target density is log-concave. Thus, the assumption of log-concavity may even be desirable in terms of the convergence rate of the associated Markov chain.

Restricting our attention to the class of problems for which the posterior density is log-concave, we will show how upper and lower bounds to the target density, π, can be obtained from the sampler output. Having obtained these bounds, it is then possible to obtain an upper bound to the probability in the tails (under π), lying outside the convex hull of the sample path. Thus, we might make the decision to stop the sampler when the area in the tails becomes sufficiently small. Note that our approach is quite different from others in that, instead of trying to estimate the length of the *burn-in* period, we try to ascertain whether or not the chain has explored a sufficient proportion of the parameter space for inference based upon the sampler output to adequately reflect characteristics of the underlying distribution.

We begin by showing how, given the output from our sampler, we can obtain both upper and lower bounds to the target density in the one-dimensional case, before progressing to the more general (and useful) higher-dimensional cases.

2. The method in one dimension. Given a sample $T_k = \{x_1, \dots, x_k\}$ from a univariate log-concave density with domain D, given by

$$\pi(x) = \frac{f(x)}{\int_D f(x)\,dx}$$

and for which the normalization constant is unknown, we can obtain bounds for $\pi(x)$ as follows.

Assume that the domain D of $f(x)$ is an interval on $\mathbb{R}$, that $f(x)$ is continuous and differentiable everywhere in D and that $h(x) = \log f(x)$ is concave everywhere in D. Suppose that $h(x)$ and $h'(x)$ have been evaluated at the k points in D, $x_1 \leq \cdots \leq x_k$; then we can define a piecewise linear upper bound for $h(x)$ formed from the tangents to $h(x)$ at the points in T_k in a manner similar to that described by Gilks and Wild (1992) in the context of adaptive rejection sampling.

For $j = 1, \dots, k-1$, the tangents to the points x_j and x_{j+1} intersect at the point

$$z_j = \frac{h(x_{j+1}) - h(x_j) - x_{j+1}h'(x_{j+1}) + x_j h'(x_j)}{h'(x_j) - h'(x_{j+1})};$$

see Figure 1. Therefore, if we let z_0 be the (greatest) lower bound of D (or $-\infty$ if D is unbounded below) and let z_k be the (least) upper bound of D (or $+\infty$ if D is unbounded above), then an upper bound for $h(x)$ in the region $x \in [z_{j-1}, z_j]$ can be given by

$$u_j(x) = h(x_j) + (x - x_j)h'(x_j), \qquad j = 1, \dots, k.$$

We can also define a piecewise linear lower bound to $h(x)$ for $x \in [x_j, x_{j+1}]$, formed from the chords between adjacent points in T_k by

$$l_j(x) = \frac{(x_{j+1} - x)h(x_j) + (x - x_j)h(x_{j+1})}{x_{j+1} - x_j}, \qquad j = 1, \dots, k-1.$$

Finally, if we define

$$u(x) = \begin{cases} u_j(x), & x \in [z_{j-1}, z_j], \\ 0, & \text{else}, \end{cases} \qquad j = 1, \dots, k$$

and

$$l(x) = \begin{cases} l_j(x), & x \in [x_j, x_{j+1}], \\ -\infty, & \text{else}, \end{cases} \qquad j = 1, \dots k-1,$$

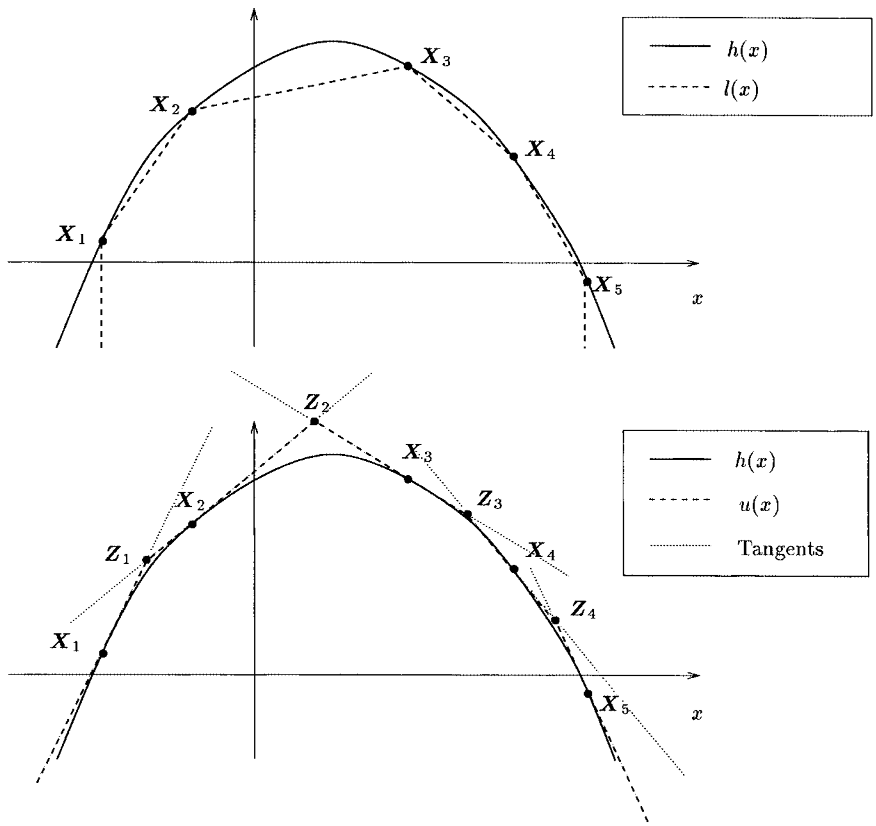

FIG. 1. *Illustrating the tangent-based bounds for five points in the one-dimensional case.*

then the concavity of $h(x)$ ensures that

$$l(x) \leq h(x) \leq u(x), \qquad \forall\, x \in D.$$

Clearly, given these bounds on h, a lower bound for the normalization constant for π is given by

$$\begin{aligned} \int_D f(x)\,dx &= \int_D \exp h(x)\,dx \\ &\geq \int_D \exp l(x)\,dx \\ &= c \quad \text{say.} \end{aligned}$$

Now, if $x \geq x_k$, then

$$\begin{aligned} h(x) &\leq u_k(x) \\ &= h(x_k) + (x - x_k)h'(x_k). \end{aligned}$$

Therefore,

$$\begin{aligned} f(x) &\le \exp[h(x_k)]\exp[(x-x_k)h'(x_k)] \\ &= f(x_k)\exp\left(\frac{-x_k f'(x_k)}{f(x_k)}\right)\exp\left(\frac{xf'(x_k)}{f(x_k)}\right) \\ &= a_k \exp(-b_k x) \quad \text{say.} \end{aligned}$$

Thus, for $x \ge x_k$, we gain an upper bound for $\pi(x)$ of the form,

$$\pi(x) \le \frac{f(x)}{c} \le \frac{a_k \exp(-b_k x)}{c}$$

and therefore we can obtain an upper bound to the proportion of the parameter space contained within the tail beyond x_k given by

$$\begin{aligned} \mathbb{P}(x \ge x_k) &= \int_{x_k}^{\infty} \pi(x)\, dx \\ &\le \int_{x_k}^{\infty} \frac{a_k \exp(-b_k x)}{c}\, dx \\ &= \frac{a_k \exp(-b_k x_k)}{cb_k}. \end{aligned}$$

Similarly, we can obtain an upper bound to the proportion in the lower tail,

$$\mathbb{P}(x \le x_1) \le \frac{a_1 \exp(-b_1 x_1)}{cb_1},$$

where

$$a_1 = f(x_1)\exp(b_1 x_1)$$

and

$$b_1 = \frac{-f'(x_1)}{f(x_1)}.$$

Thus, if our set T_k were the output from an MCMC sampler that we wanted to decide when to stop, we might continue to run the chain until the set T_k has largest and smallest elements x_k and x_1, respectively, which satisfy the criterion

$$\frac{a_1 \exp(-b_1 x_1)}{cb_1} + \frac{a_k \exp(-b_k x_k)}{cb_k} \le \varepsilon \quad \text{for some } \varepsilon > 0,$$

at which point, the convex hull of the sample path covers at last $100(1-\varepsilon)\%$ of the parameter space. We argue that this is a necessary condition for convergence, since the chain cannot possibly have converged until "the entire parameter space has been explored."

Similar results can be obtained when the derivatives are unknown (or incalculable), by introducing an extra point x_0. In this case the lower bound becomes tighter by forming the piecewise linear lower bound to h from the line segments between the ordered points in the sample, including the extra point x_0. The upper bound is formed from the straight line segments between each of the x_i in the sample and the point x_0 which is (ideally) somewhere near the mode; see Figure 2.

We now discuss how the method may be extended to higher dimensions in the case where the derivatives are known and, more importantly in the more general (and practically useful) case, where they are not. In Sections 3–5, we limit ourselves to obtaining bounds on the proportion of the parameter space explored, given a convex hull consisting of exactly $n + 1$ points in n-dimensional space. We discuss the generalization of this method to the case where the convex hull consists of $k > n + 1$ points in n-space in Section 6, before comparing the two methods in Section 7.

We begin, in Section 3, by extending bounds to h in the one-dimensional method to general dimensions in the case where derivatives are known (or calculable). In Section 4 we develop bounds on h in the case where the

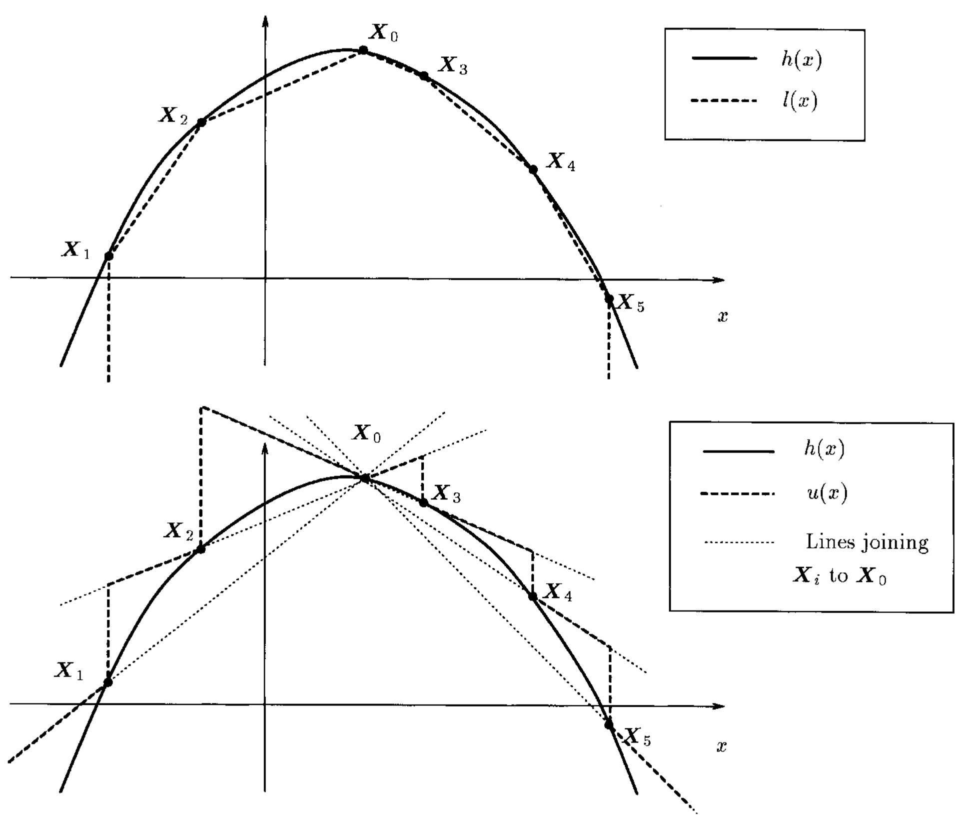

FIG. 2. *Illustrating the derivative-free bounds for six points in the one-dimensional case.*

derivatives are not available and finally, in Section 5, we show how the derivative-free bounds on h may be practically implemented to bound the proportion in the tails.

3. The tangent-based method. We begin by looking at how upper and lower bounds to the surface h may be found in the two-dimensional case before extending the tangent-based method to general dimensions.

For clarity we shall adopt the following notation. We denote a general point in the two-dimensional parameter space by a lower case letter $\mathbf{x}$, say, and let the corresponding upper case letter $\mathbf{X} = (\mathbf{x}', h)$ denote a point in three-dimensional space, where the third component of $\mathbf{X}$ corresponds to the log-density at $\mathbf{x}$. Finally, we let x_i denote the ith component of $\mathbf{x}$.

3.1. *The method in two dimensions*. Take three points on a concave surface $h = \log f$ given by $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ where $\mathbf{X}'_i = (\mathbf{x}'_i, h(\mathbf{x}_i))$ and $\mathbf{x}'_i = (x_{i1}, x_{i2})$, $i = 1, 2, 3$. The normal to the tangent plane is given by ∇H, where

$$\nabla H = \left(\frac{\partial h}{\partial x_1}, \frac{\partial h}{\partial x_2}, -1 \right).$$

If we define

$$\mathbf{N} \triangleq \begin{pmatrix} \dfrac{\partial h}{\partial x_1}(\mathbf{x}_1) & \dfrac{\partial h}{\partial x_2}(\mathbf{x}_1) & -1 \\ \dfrac{\partial h}{\partial x_1}(\mathbf{x}_2) & \dfrac{\partial h}{\partial x_2}(\mathbf{x}_2) & -1 \\ \dfrac{\partial h}{\partial x_1}(\mathbf{x}_3) & \dfrac{\partial h}{\partial x_2}(\mathbf{x}_3) & -1 \end{pmatrix} \triangleq \begin{pmatrix} \mathbf{N}'_1 \\ \mathbf{N}'_2 \\ \mathbf{N}'_3 \end{pmatrix}, \tag{1}$$

then the single point at which all three planes coincide is given by

$$\mathbf{A}_0 = [\mathbf{a}_0, h(\mathbf{a}_0)] = \mathbf{N}^{-1} \begin{pmatrix} \mathbf{N}'_1 \mathbf{X}_1 \\ \mathbf{N}'_2 \mathbf{X}_2 \\ \mathbf{N}'_3 \mathbf{X}_3 \end{pmatrix}, \tag{2}$$

since $\mathbf{A}_0$ satisfies

$$(\mathbf{X}_i - \mathbf{A}_0) \cdot \mathbf{N}_i = 0 \Rightarrow \mathbf{N}'_i \mathbf{A}_0 = \mathbf{N}'_i \mathbf{X}_i \qquad \forall\ i = 1, 2, 3.$$

Similarly, if

$$\mathbf{N}_{ij} = \begin{pmatrix} \mathbf{n}'_i \\ \mathbf{n}'_j \end{pmatrix}, \qquad i \neq j, \tag{3}$$

then the single point at which the tangent planes to $\mathbf{X}_i$ and $\mathbf{X}_j$ coincide with the $h = 0$ plane is given by

$$\mathbf{A}_{ij} = [\mathbf{a}_{ij}, h(\mathbf{a}_{ij})] = \mathbf{N}_{ij}^{-1} \begin{pmatrix} \mathbf{N}'_i \mathbf{X}_i \\ \mathbf{N}'_j \mathbf{X}_j \end{pmatrix}, \qquad i \neq j,$$

where $h(\mathbf{a}_{ij}) = 0\ \forall\ i, j$.

Thus, we have a three-dimensional hypertriangle (a triangular-based pyramid) bounding $h(\mathbf{x})$ above with vertices $\{\mathbf{A}_0, \mathbf{A}_{12}, \mathbf{A}_{13}, \mathbf{A}_{23}\}$ (Figure 3). Note that the bounds continue beyond the region $\{\mathbf{a}_{12}, \mathbf{a}_{23}, \mathbf{a}_{13}\}$, but it is convenient to visualize the hypertriangle to have a base arbitrarily at $h = 0$, say.

Straight lines between the two-dimensional vectors $\{\mathbf{a}_0, \mathbf{a}_{12}, \mathbf{a}_{13}, \mathbf{a}_{23}\}$ in the $h = 0$ plane split the parameter space into three subregions, R_i, corresponding to the open-ended triangle with vertex $\mathbf{a}_0$ and edges passing through $\mathbf{a}_{ij}$ and $\mathbf{a}_{ik}$ $j, k \neq i$, and $\forall\ i = 1, 2, 3$; see Figure 4. Then we can define an upper bound $u(\mathbf{x})$ to $h(\mathbf{x})$ by

$$u(\mathbf{x}) = \min_i \{u_i(\mathbf{x})\}, \tag{4}$$

where u_i corresponds to the bounding plane in region R_i, that is,

$$u_i(\mathbf{x}) \triangleq h(\mathbf{x}_i) - (\mathbf{x}_i - \mathbf{x}) \cdot \mathbf{n}_i, \tag{5}$$

since $[\mathbf{X}_i - (\mathbf{x}, u_i(\mathbf{x}))] \cdot \mathbf{N}_i = 0$ and $N_{ik} = -1$; see Figure 3.

Thus, $\exp u(\mathbf{x})$ is an upper bound for $f(\mathbf{x})$. Now let R_0 be the region defined by the triangle with vertices $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3\}$; then the upper bound to the volume in the tails is given by

$$\int_{R_0^c} \exp u(\mathbf{x})\, d\mathbf{x}.$$

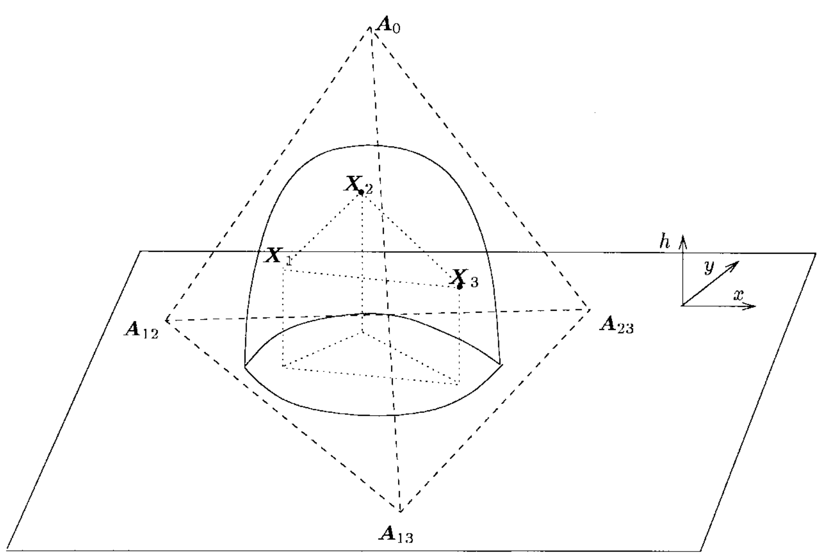

FIG. 3. *Illustrating the tangent-based bounds in the two-dimensional case. The solid line denotes the surface h while the dotted and dashed lines represent the lower and upper bounds, respectively.*

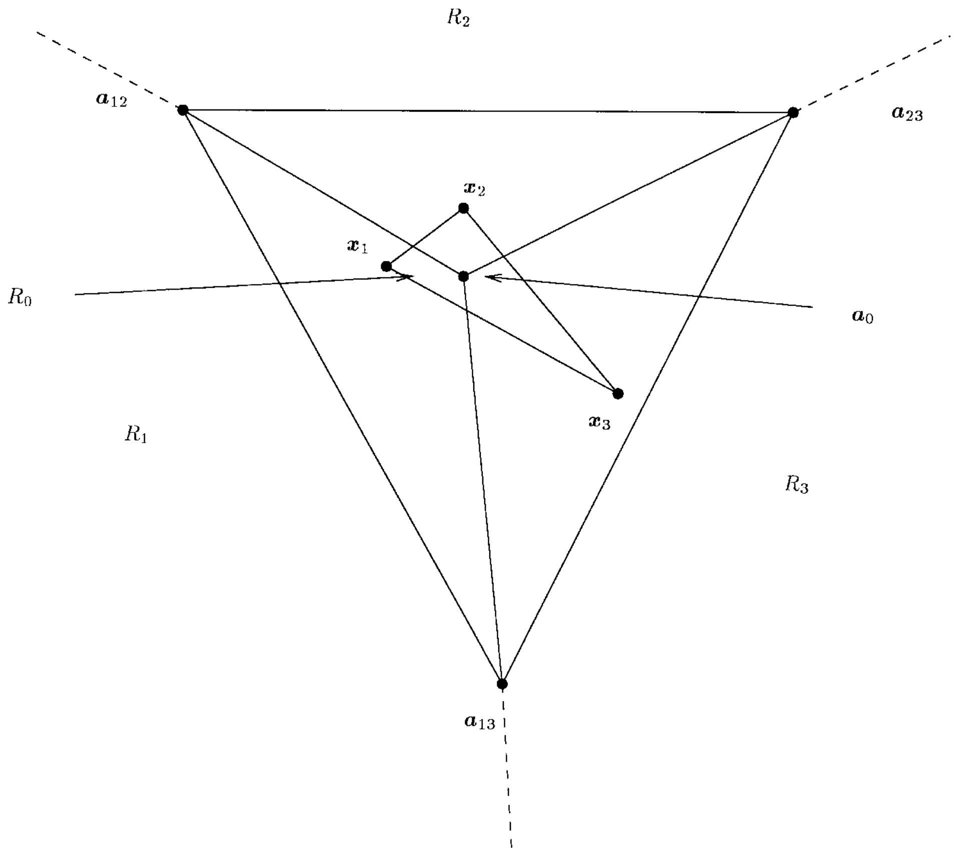

FIG. 4. *Plan view corresponding to Figure* 3, *illustrating the subregions in the two-dimensional case.*

To find the lower bound, we proceed as follows. Define the normal to the plane containing the three points $\mathbf{X}_1$, $\mathbf{X}_2$ and $\mathbf{X}_3$ by

$$\mathbf{N}_0 = (\mathbf{X}_1 - \mathbf{X}_2) \times (\mathbf{X}_1 - \mathbf{X}_3),$$

then we can proceed as before to define a lower bound $l(\mathbf{x})$, to $h(\mathbf{x})$ by

$$l(x) = \begin{cases} l_0(\mathbf{x}), & \mathbf{x} \in R_0, \\ -\infty, & \mathbf{x} \in R_0^c, \end{cases} \tag{6}$$

where we arbitrarily define $l_0(\mathbf{x})$ in terms of $\mathbf{x}_1$, by

$$l_0(\mathbf{x}) = (\mathbf{x} - \mathbf{x}_1) \cdot \mathbf{n}_0 + h(\mathbf{x}_1). \tag{7}$$

Thus, the required lower bound to the value of the density function integrated over the parameter space contained within the convex hull R_0, is given by

$$c = \int_{R_0} \exp l(\mathbf{x})\, d\mathbf{x}. \tag{8}$$

3.2. *The method in general dimensions.* The tangent-based method can be generalized to higher dimensions. First let us alter our notation and let an upper case vector $\mathbf{X}$ denote a point in $\mathbb{R}^{n+1}$ and a lower case vector $\mathbf{x}$ denote the n-dimensional projection of $\mathbf{X}$ onto the $h = 0$ plane. If we take points in $\mathbb{R}^n$, with $h(\mathbf{x})$ forming the $(n + 1)$th component, then we will require $n + 1$ points to form a hyperplane covering the corresponding $(n + 1)$-dimensional surface, h. Thus, we obtain $n + 2$ vertices for this hyperplane, one where all $n + 1$ planes meet and a further $n + 1$ where n of the planes meet with the $h = 0$ plane.

Let $\{\mathbf{X}_i = [\mathbf{x}_i, h(\mathbf{x}_i)],\ i = 1, \ldots, n + 1\}$ be the set of points on the concave surface, and let $\mathbf{N}_i$ be the normal to the tangent hyperplane to h at point $\mathbf{X}_i$. Then we can generalize the method of Section 3.1 to calculate the vertices of the tangent hyperplane. We generalize (1) to define

$$\mathbf{N}' = (\mathbf{N}_1, \ldots, \mathbf{N}_{n+1}),$$

where the $n + 1$-dimensional normal vector is given by

$$\mathbf{N}_i = \begin{pmatrix} \dfrac{\partial h}{\partial \mathbf{x}}(\mathbf{x}_i) \\ -1 \end{pmatrix} \tag{9}$$

and generalize (2) similarly to give us the point $\mathbf{A}_0$, where all $n + 1$ tangent hyperplanes meet. We then redefine (3) by

$$\mathbf{N}'_{(i)} = (\mathbf{n}_1, \ldots, \mathbf{n}_{i-1}, \mathbf{n}_{i+1}, \ldots, \mathbf{n}_{n+1}).$$

Then the point where all hyperplanes, except the one tangential to $\mathbf{X}_i$, meet the $h = 0$ plane is given by

$$\mathbf{A}_{(i)} = \mathbf{N}_{(i)}^{-1} \begin{pmatrix} \mathbf{N}'_1 \mathbf{X}_1 \\ \cdots \\ \mathbf{N}'_{i-1} \mathbf{X}_{i-1} \\ \mathbf{N}'_{i+1} \mathbf{X}_{i+1} \\ \cdots \\ \mathbf{N}'_{n+1} \mathbf{X}_{n+1} \end{pmatrix}.$$

Thus, we obtain an $n + 1$-dimensional hypertriangle with vertices $\{\mathbf{A}_0, \mathbf{A}_{(1)}, \ldots, \mathbf{A}_{(n+1)}\}$ with respect to the $h = 0$ plane.

Now, if we define $u_i(\mathbf{x})$ as in (5), then the upper bound to $h(\mathbf{x})$ is given by $u(\mathbf{x})$ as defined in (4). Finally, if we let R_0 be the hypertriangular region with vertices $\{\mathbf{x}_i\colon i = 1, \ldots, n + 1\}$ then the required (nonnormalized) upper bound to the hypervolume in the tails is given by

$$\int_{R_0^c} \exp u(\mathbf{x})\, d\mathbf{x}$$

We now proceed to define a lower bound, but first we require the following definition and lemmas.

DEFINITION 2. Following Spivak (1965), we define the $n+1$-dimensional vector product between n vectors $\mathbf{Z}_1, \dots, \mathbf{Z}_n \in \mathbb{R}^{n+1}$ to be

$$\mathbf{W} = \mathbf{Z}_1 \times \cdots \times \mathbf{Z}_n,$$

where $\mathbf{W}$ is such that the determinant of the square matrix with rows given by the $\mathbf{Z}_i$ and another vector $\mathbf{Y}$, denoted by $|\mathbf{Z}_1, \dots, \mathbf{Z}_n, \mathbf{Y}|$ is given by

$$|\mathbf{Z}_1, \dots, \mathbf{Z}_n, \mathbf{Y}| = \mathbf{W} \cdot \mathbf{Y}, \qquad \forall\, \mathbf{Y} \in \mathbb{R}^{n+1}.$$

LEMMA 1. *Given $\mathbf{Z}_1, \dots, \mathbf{Z}_n \in \mathbb{R}^{n+1}$, the $n+1$-dimensional vector product exists and is unique.*

PROOF. Let $\mathbf{Z}^i$ denote the $n \times n$ matrix with the jth row given by the vector $\mathbf{Z}_j$ with the ith component removed, that is, $Z_{j,1}, \dots, Z_{j,(i-1)}, Z_{j,(i+1)}, \dots, Z_{j,(n+1)}$; then

$$\begin{aligned}|\mathbf{Z}_1, \dots, \mathbf{Z}_n, \mathbf{Y}| &= Y_1|\mathbf{Z}^1| - Y_2|\mathbf{Z}^2| + \cdots + (-1)^n Y_{n+1}|\mathbf{Z}^{n+1}| \\ &= \mathbf{Y} \cdot \left(|\mathbf{Z}^1|, -|\mathbf{Z}^2|, \dots, (-1)^n|\mathbf{Z}^{n+1}|\right) \\ &= \mathbf{Y} \cdot \mathbf{W} \quad \text{say.}\end{aligned}$$

The result follows. □

LEMMA 2. *The $n+1$-dimensional vector product $\mathbf{W} = \mathbf{Z}_1 \times \cdots \times \mathbf{Z}_n$ represents the normal to the hyperplane containing points $\mathbf{Z}_1, \dots, \mathbf{Z}_n$.*

PROOF.

$$\begin{aligned}\mathbf{W} \cdot \mathbf{Z}_i &= Z_{i1}|\mathbf{Z}^1| + \cdots + (-1)^n Z_{i(n+1)}|\mathbf{Z}^{n+1}| \\ &= |\mathbf{Z}_1, \dots, \mathbf{Z}_n, \mathbf{Z}_i| \\ &= 0, \qquad \forall\, i = 1, \dots, n.\end{aligned}$$

The result follows. □

Thus, to define the lower bound, we let the normal to the hyperplane containing the points $\{\mathbf{X}_i : i = 1, \dots, n+1\}$ be denoted by $\mathbf{N}_0$, where

$$\mathbf{N}_0 = (\mathbf{X}_1 - \mathbf{X}_2) \times \cdots \times (\mathbf{X}_1 - \mathbf{X}_{n+1}),$$

by Lemma 2. Then we can define a lower bound $l(\mathbf{x})$ to $h(\mathbf{x})$ using (6) and (7) so that the required lower bound to the volume within the convex hull R_0 is the same as that for the lower-dimensional case, and is given by (8).

4. The derivative-free method. In practice, the derivatives will not generally be known and they will need to be estimated numerically. This can be computationally expensive, since the derivatives may need to be calculated for a large number of points. This is particularly true for slowly mixing samplers, where diagnostic methods are most needed. In this section we propose an alternative method, which does not require the calculation of derivatives.

As with the tangent-based method, we begin by looking at the derivative-free method in the two-dimensional case before extending the method to general dimensions.

4.1. *The method in two dimensions.* The tangent-based method for the two-dimensional case, can be adapted to remove the need for calculating derivatives as follows. If, instead of taking three points, we take four points on the log-surface h, $\{\mathbf{X}_i = (\mathbf{x}_i, 0)\colon\ i = 0, \dots, 3\}$, such that $\mathbf{x}_0$ lies strictly within the convex hull formed from the other three points and $h(\mathbf{x}_0) > h(\mathbf{x}_i)$ $\forall\ i = 1, 2, 3$. Then we can define three nontangent planes, each of which contains both $\mathbf{X}_0$ and two other points and has normal given by

$$\mathbf{N}_{ij} = (\mathbf{X}_0 - \mathbf{X}_i) \times (\mathbf{X}_0 - \mathbf{X}_j), \qquad i \neq j. \tag{10}$$

The lower bound to h could be formed in a similar manner to that of the tangent-based method, but a tighter lower bound may be formed by introducing the fourth point $\mathbf{X}_0$; see Figure 5. We then define the lower bound $l(\mathbf{x})$ in a similar way to that of Section 3, using the normals given in (10). Thus, a

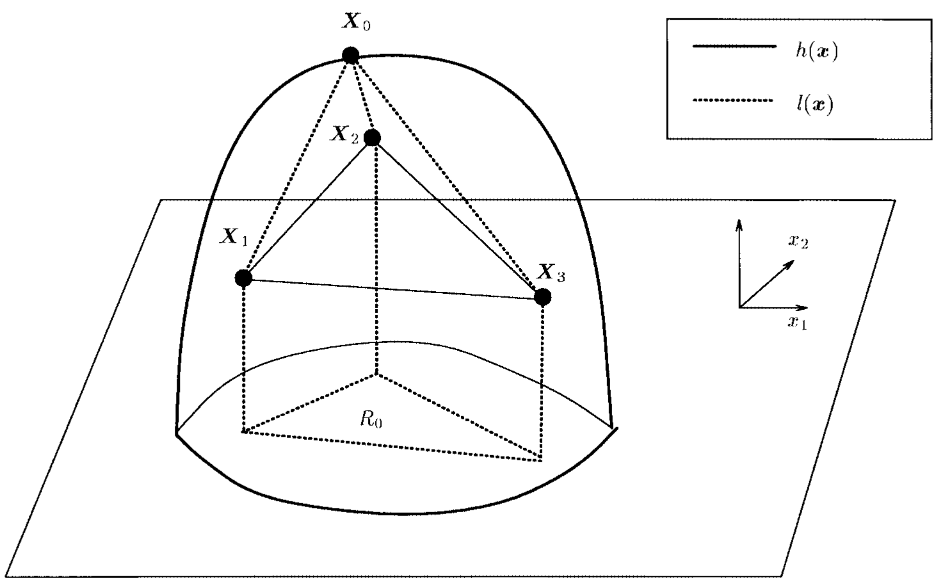

FIG. 5. *Illustrating the derivative-free lower bound in the two-dimensional case.*

lower bound to the volume under the surface is given by

$$c = \int_{R_0} \exp l(\mathbf{x})\, d\mathbf{x},$$

where

$$l(\mathbf{x}) = \begin{cases} \min_{i \neq j} l_{ij}(\mathbf{x}), & \mathbf{x} \in R_0, \\ -\infty, & \mathbf{x} \in R_0^c \end{cases} \tag{11}$$

and

$$l_{ij}(\mathbf{x}) \triangleq \frac{(\mathbf{x}_0 - \mathbf{x}) \cdot \mathbf{n}_{ij}}{N_{ij3}} + h(\mathbf{x}_0),$$

where N_{ijk} denotes the kth component of the normal vector $\mathbf{N}_{ij}$. Note $N_{ij3} \neq 0$, since $\mathbf{x}_0$ lies strictly within the convex hull.

In order to obtain an upper bound to h, beyond the convex hull, we note that each of the planes described above intersects the surface at three points $\mathbf{X}_0$, $\mathbf{X}_i$ and $\mathbf{X}_j$, say, but lies below the surface for a sectioned paraboloid region beyond the line joining $\mathbf{X}_i$ to $\mathbf{X}_j$. Thus, we cannot use this plane as an upper bound to the surface in the tails; see Figure 6. Instead, we “tilt” this plane upwards about $\mathbf{X}_0$ until all points beyond the line joining $\mathbf{x}_i$ and $\mathbf{x}_j$ lie below the plane. In effect, we drag the line between $(\mathbf{x}_i, 0)$ and $(\mathbf{x}_j, 0)$ vertically upwards until it touches the surface at only a single point, $\mathbf{X}^*_{ij}$. This point can be found as follows.

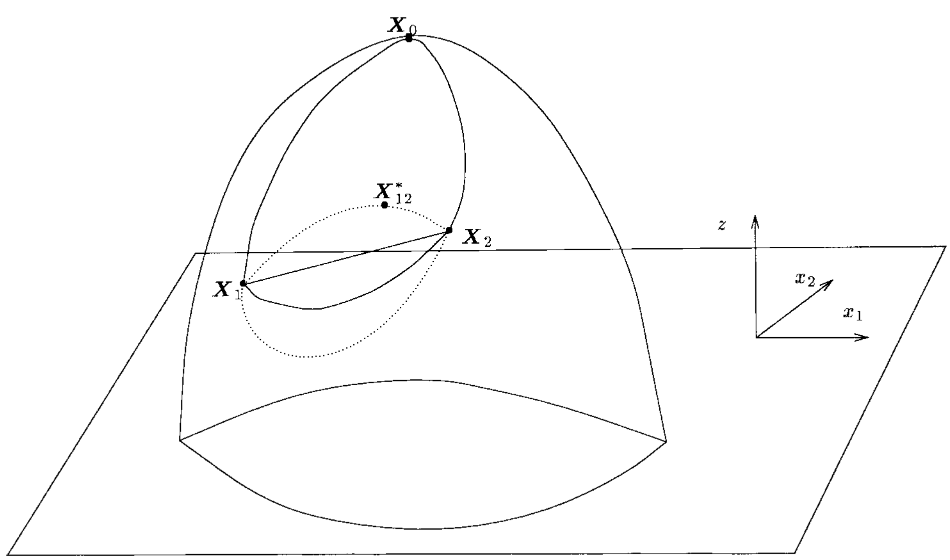

FIG. 6. *Illustrating how the derivative-free bounding (hyper) plane lies below the surface for some points in the tails and how the point* $\mathbf{X}^*_{12}$ *may be found.*

Let

$$\mathbf{x}_{ij}(\lambda) = \mathbf{x}_i + \lambda(\mathbf{x}_j - \mathbf{x}_i);$$

then $h[\mathbf{x}_{ij}(\lambda)]$ denotes the value of h for a general point on the line between $\mathbf{x}_i$ and $\mathbf{x}_j$ on the $h = 0$ plane. Now, let λ^*_{ij} denote the value of $0 < \lambda < 1$ which maximizes $h[\mathbf{x}_{ij}(\lambda)]$, then

$$\mathbf{X}^*_{ij} = \left(\mathbf{x}_{ij}(\lambda^*), h\left[\mathbf{x}_{ij}(\lambda^*)\right]\right)$$

is the single point described above. Once we have our point $\mathbf{X}^*_{ij}$, then our required (tilted) plane has normal

$$\mathbf{N}^*_{ij} = (\mathbf{X}_0 - \mathbf{X}^*_{ij}) \times \left[(\mathbf{x}_i, 0) - (\mathbf{x}_j, 0)\right]. \tag{12}$$

Before continuing, we provide a result that shows that such a method for "tilting" the original plane provides us with a plane bounding h above for all points beyond the line $\mathbf{x}_i - \mathbf{x}_j$.

THEOREM 1. *For any point in* $\mathbb{R}^2$ *of the general form*

$$\mathbf{x}_{ij}(\mu, \lambda) = \mathbf{x}^*_{ij} + \lambda(\mathbf{x}_i - \mathbf{x}_j) + \mu(\mathbf{x}^*_{ij} - \mathbf{x}_0),$$

$h[\mathbf{x}_{ij}(\mu, \lambda)]$ *is bounded above by the tilted plane* $u_{ij}[\mathbf{x}_{ij}(\mu, \lambda)]$, *for all* $\lambda \in \mathbb{R}$ *and* $\mu > 0$.

For the proof, see the Appendix.

Thus, we can define three such planes bounding the surface above (beyond the points $\mathbf{X}_{12}$, $\mathbf{X}_{13}$ and $\mathbf{X}_{23}$), by splitting the state space into three subregions as before and, in order to obtain an upper bound on the area in the tails, we can proceed as for the tangent-based method with the matrix of tilted normals,

$$\mathbf{N}^* = \begin{pmatrix} \mathbf{N}^{*\prime}_{12} \\ \mathbf{N}^{*\prime}_{13} \\ \mathbf{N}^{*\prime}_{23} \end{pmatrix} \tag{13}$$

substituted in (1). Note that we have alternative formulas for the $\mathbf{A}_{ij}$, since $\mathbf{A}_0 \equiv \mathbf{X}_0$ and we define

$$\mathbf{A}_{ij} = \mathbf{N}^{*-1}_{ij} \begin{pmatrix} \mathbf{n}^*_i \cdot \mathbf{x}_k \\ \mathbf{n}^*_j \cdot \mathbf{x}_k \end{pmatrix}, \qquad k \neq i, j.$$

Thus, in essence, we replace the computational expense of the estimation of derivatives from the first method by the optimization of the function h over all possible λ-values in the second. However, there are several advantages to the derivative-free method over the original. First, the optimization is extremely easy and can be performed by a simple downhill search mechanism and hence, the computational expense is decreased, particularly in higher dimensions. Second, we no longer require that f be differentiable everywhere on D and finally, the practical implementation of the diagnostic necessarily becomes less problem dependent.

4.2. *The method in general dimensions*. The derivative-free method can be extended to general dimensions as follows. Let

$$\mathbf{x}_{(i)}(\boldsymbol{\lambda}) = \mathbf{x}_k + \sum_{j \neq k, i} \lambda_j (\mathbf{x}_j - \mathbf{x}_k) \quad \text{for any } k \neq i.$$

Then, as before, we find $\boldsymbol{\lambda}^*$ which maximizes $h[\mathbf{x}_{(i)}(\boldsymbol{\lambda})]$ and set

$$\mathbf{X}^*_{(i)} = \left[\mathbf{x}_{(i)}(\boldsymbol{\lambda}^*), h\left(\mathbf{x}_{(i)}(\boldsymbol{\lambda}^*)\right)\right].$$

Then, our tilted plane bounding h above beyond the hyperplane with points $\{\mathbf{x}_j \colon j \neq i\}$ has normal

$$\mathbf{N}^*_{(i)} = \left(\mathbf{X}_0 - \mathbf{X}^*_{(i)}\right) \times \prod_{\substack{j=1 \\ j \neq i, k}}^{n+1} \left[(\mathbf{x}_k, 0) - (\mathbf{x}_j, 0)\right] \quad \text{for any } k \neq i.$$

If we define

$$u_{(i)}(\mathbf{x}) = \frac{(\mathbf{x}^*_{(i)} - \mathbf{x}) \cdot \mathbf{n}^*_{(i)}}{N^*_{(i)(n+1)}} + h(\mathbf{x}^*_{(i)}),$$

then we gain the following theorem.

THEOREM 2. *For any point in* $\mathbb{R}^n$ *of the general form*

$$\mathbf{x}(\mu, \boldsymbol{\lambda}) = \mathbf{x}^*_{(i)} + \mu(\mathbf{x}^*_{(i)} - \mathbf{x}_0) + \sum_{\substack{j=1 \\ j \neq i, k}}^{n+1} (\mathbf{x}_k - \mathbf{x}_j)\lambda_j \quad \textit{for any } k \neq i,$$

$h[\mathbf{x}(\mu, \boldsymbol{\lambda})]$ *is bounded above by* $u_{(i)}[\mathbf{x}(\mu, \boldsymbol{\lambda})]$ *for all* $\mu > 0$ *and* $\boldsymbol{\lambda} \in \mathbb{R}^{n-1}$.

The proof is a direct extension of the proof of Theorem 1.

Once we have calculated the $\mathbf{N}^*_{(i)}$ for all values of $i = 1, \dots, n+1$ we can define $\mathbf{N}^*$ as a direct generalization of the two-dimensional case in (13) and proceed in the same manner as the tangent-based method in general dimensions, by substituting $\mathbf{N}^*_{(i)}$ for the $\mathbf{N}_i$ to obtain the relevant upper bound. Similarly, a lower bound can be found by the obvious extension to that of the two-dimensional derivative-free method of (11), given by

$$l(\mathbf{x}) = \begin{cases} \min\limits_i l_{(i)}(\mathbf{x}), & \mathbf{x} \in R_0, \\ -\infty, & \mathbf{x} \in R_0^c, \end{cases}$$

where

$$l_{(i)}(\mathbf{x}) \triangleq \frac{(\mathbf{x}_0 - \mathbf{x}) \cdot \mathbf{n}_i}{N_{(i)(n+1)}} + h(\mathbf{x}_0)$$

and the $\mathbf{N}_i$ are the normals to the original (nontilted) planes given in (9).

Thus, we have lower bounds on h within a convex hull of $n + 1$ points in n-dimensions and an upper bound on h beyond the convex hull. We now show how the derivative-free bounds can be used in practice to gain an upper bound on the proportion in the tails beyond the convex hull under the distribution π.

5. A derivative-free bound on the tail volume. We begin by generating a *starting set* of $n + 2$ points $\{\mathbf{x}_0, \dots, \mathbf{x}_{n+1}\}$ in $\mathbb{R}^n$ from the output, such that the points $\{\mathbf{x}_1, \dots, \mathbf{x}_{n+1}\}$ are vertices of an n-dimensional hypertriangle and $\mathbf{x}_0$ is both contained within this hypertriangle, and is such that $h(\mathbf{x}_0) > h(\mathbf{x}_i)$ $\forall$ $i = 1, \dots, n + 1$. Having gained this starting set, we continue to monitor the sampler output, updating our set of points so as to maximize the volume of their convex hull, until our set covers a sufficient proportion of the parameter space.

Thus, the implementation of our method can be split into two algorithms. We shall begin by defining these two algorithms and then return to individual steps to explain how they might be performed.

Algorithm 1. *Generating a starting set*.

STEP 1. Generate $n + 1$ points $\mathbf{x}_1, \dots, \mathbf{x}_{n+1}$.
STEP 2. Define the region R, contained within the hypertriangle with vertices $\mathbf{x}_1, \dots, \mathbf{x}_{n+1}$.
STEP 3. Generate a new point $\mathbf{x}$.
STEP 4. If $\mathbf{x} \in R$, then if $h(\mathbf{x}) > h(\mathbf{x}_i)$ $\forall$ i, set $\mathbf{x}_0 = \mathbf{x}$ and stop. Otherwise return to Step 3.
STEP 5. If $\mathbf{x} \notin R$, then check to see if the hypervolume of R can be increased by replacing $\mathbf{x}_i$ by $\mathbf{x}$. If so, then replace $\mathbf{x}_i$ by $\mathbf{x}$ and return to Step 2. Otherwise return to Step 3.

This algorithm produces our *starting set*, which is simply the first set of $n + 2$ points $\{\mathbf{x}_0, \dots, \mathbf{x}_{n+1}\}$ from the sampler output such that $h(\mathbf{x}_0) > h(\mathbf{x}_i)$ $\forall$ $i = 1, \dots, n + 1$ and $\mathbf{x}_0$ lies within the convex hull of the other $n + 1$ points, R. The second algorithm monitors the output, updating our set of points whenever the volume in the region R can be increased and stopping once the probability mass outside R falls below some critical value ε, say.

Algorithm 2. *Diagnosing convergence*. Convergence can be diagnosed by using the following algorithm, which may be run several times in succession to ensure that the "best" points are used:

STEP 1. Given a starting set $\{\mathbf{x}_0, \dots, \mathbf{x}_{n+1}\}$, define R and calculate the upper bound to the probability mass beyond the convex hull of the sample path, P. If $P < \varepsilon$, then stop.
STEP 2. Generate a new point, $\mathbf{x}$.

STEP 3. If $\mathbf{x} \in R$, then if $h(\mathbf{x}) > h(\mathbf{x}_0)$, set $\mathbf{x}_0 = \mathbf{x}$ and return to Step 1. Otherwise return to Step 2.

STEP 4. If $\mathbf{x} \notin R$ and $h(\mathbf{x}) < h(\mathbf{x}_0)$, then check to see if the hyper-volume of R can be increased by replacing $\mathbf{x}_i$ by $\mathbf{x}$. If so, then replace it (so long as $\mathbf{x}_0$ remains within R) and return to Step 1. Otherwise return to Step 2.

It is not immediately obvious how these steps can be accomplished. In particular, it is not clear how we might perform any of the following:

1. Calculate the upper bound for P.
2. Define the region R and decide whether or not some point $\mathbf{x}$ is an element of R.
3. Calculate the hypervolume of R, given the $n + 1$ vertices.

We shall now discuss in detail how each of these tasks can be performed in practice.

5.1. *Calculating an upper bound for P.* If we let T denote the volume in the tails (under f) beyond the convex hull R, and M denotes the volume within the convex hull, then the proportion in the tails is given by

$$P = \frac{T}{T + M}.$$

However, we can bound T and M by

$$T \leq \int_{R_0^c} \exp[u(\mathbf{x})]\, d\mathbf{x} \tag{14}$$

and

$$M \geq \int_{R_0} \exp[l(\mathbf{x})]\, d\mathbf{x} \tag{15}$$

For general u and l, the bounds to T and M given in (14) and (15) are not easily found, but the piecewise linear nature of both u and l, as we define them, allow us to bound T above and M below via the following theorem.

THEOREM 3. *Given an $(n + 1)$-dimensional hypertriangle with vertices $\{\mathbf{X}_0, \ldots, \mathbf{X}_{n+1}\}$ the volume T can be bounded above by U and the volume M can be bounded below by L, so that P is bounded above by*

$$P_b = \frac{U}{U + L},$$

where

(16)
$$
\begin{aligned}
L &= \sum_{i=1}^{n+1} L_{(i)}, \\
L_{(i)} &= |\mathbf{J}_{(i)}| e^{b} \left(\sum_{j \neq i} \frac{\exp(a_{ij})}{\prod_{k \neq i,j} (a_{ij} - a_{ik}) a_{ij}} + \frac{(-1)^n}{\prod_{j \neq i} a_{ij}} \right), \\
U &= \sum_{i=1}^{n+1} U_{(i)}, \\
U_{(i)} &= |\mathbf{J}_{(i)}| e^{b} \left[- \sum_{j \neq i} \frac{\exp(a^*_{ij})}{\prod_{k \neq i,j} (a^*_{ij} - a^*_{ik}) a^*_{ij}} \right]
\end{aligned}
$$

(17)
$$
\begin{aligned}
\mathbf{J}'_{(i)} &= \left[(\mathbf{x}_0 - \mathbf{x}_1), \dots, (\mathbf{x}_0 - \mathbf{x}_{i-1}), (\mathbf{x}_0 - \mathbf{x}_{i+1}), \dots, (\mathbf{x}_0 - \mathbf{x}_{n+1}) \right] \\
a_{ij} &= -(\mathbf{x}_j - \mathbf{x}_0) \cdot \mathbf{n}_{(i)}, \qquad j \neq i, \\
b &= h(\mathbf{x}_0), \\
a^*_{ij} &= -\frac{(\mathbf{x}_0 - \mathbf{x}_j) \cdot \mathbf{n}^*_{(i)}}{N^*_{(i)(n+1)}}, \qquad j \neq i
\end{aligned}
$$

and where $\mathbf{N}_{(i)} = (\mathbf{n}_{(i)}, -1)$ *is the normal to the hyperplane formed from the set of vectors* $\{(\mathbf{X}_0 - \mathbf{X}_j)\colon j = 1, \dots, n+1,\ j \neq i\}$, *and* $\mathbf{N}^*_{(i)}$ *denotes the corresponding tilted normals as defined in* (12).

For the proof, see the Appendix.

Thus, given a set of $n + 2$ points in $\mathbb{R}^n$, we can bound the volume beyond the convex hull of these points by $P_b = U/(U + L)$.

5.2. *Definition of R and its elements*. Here we show how we can decide whether or not a vector $\mathbf{x}$ is within the convex hull $R \subseteq \mathbb{R}^n$, formed from the points $\mathbf{x}_1, \dots, \mathbf{x}_{n+1}$.

If we take $n = 2$, for example, and points $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$, then R is simply a triangle on the plane containing these points and we can define three vectors $\mathbf{l}_{ij}$, $i = 1, 2, 3$, $j > i$ corresponding to the three edges of R between vertices $\mathbf{x}_i$ and $\mathbf{x}_j$. A point $\mathbf{x}$ lies within the triangle R, if and only if that point lies on the *correct* side of each of the three edges. We define the edge joining vertices $\mathbf{x}_i$ and $\mathbf{x}_j$ by

$$\mathbf{l}_{ij} = \mathbf{x}_i + \lambda(\mathbf{x}_i - \mathbf{x}_j), \qquad \lambda \in \mathbb{R}$$

and let $\mathbf{n}_{ij}$ denote the perpendicular to $\mathbf{l}_{ij}$.

We may decide upon which side of $\mathbf{l}_{ij}$ some general point $x \in \mathbb{R}^2$ must lie for it to be an element of R by taking the vector $(\mathbf{x} - \mathbf{x}_i)$ and determining whether it has a positive or negative component in the $\mathbf{n}_{ij}$ direction. We know that for all points $\mathbf{x} \in R$, the sign of this component will be the same, and that no points for which this component has the opposite sign can be elements of R. Thus, to determine which side of the edge $\mathbf{l}_{ij}$ a point must lie for it to be an element of R, we need only decide what sign the dot product $(\mathbf{x}-\mathbf{x}_i)\cdot\mathbf{n}_{ij}$ should take. Now, we know that the third point $\mathbf{x}_k$, $k \neq i, j$ is an element of R so the *correct* sign is simply the sign of $(\mathbf{x}_k-\mathbf{x}_i)\cdot\mathbf{n}_{ij}$. Thus, if we define

$$g_k(\mathbf{x}) = \begin{cases} (\mathbf{x}-\mathbf{x}_i)\cdot\mathbf{n}_{ij}, & (\mathbf{x}_k-\mathbf{x}_i)\cdot\mathbf{n}_{ij} > 0 \\ -(\mathbf{x}-\mathbf{x}_i)\cdot\mathbf{n}_{ij}, & (\mathbf{x}_k-\mathbf{x}_i)\cdot\mathbf{n}_{ij} < 0 \end{cases}$$

for all k, and for any $i, j \neq k$ such that $i \neq j$. Then, $\mathbf{x} \in R$ if and only if $g_k(\mathbf{x}) > 0\ \forall\ k = 1, 2, 3$.

Clearly, we can generalize this to general dimensions as follows. If we take points $\mathbf{x}_1, \dots, \mathbf{x}_{n+1}$ and define the plane containing all points other than $\mathbf{x}_i$ to have normal

$$\mathbf{n}_{(i)} = \prod_{j \neq i, l} (\mathbf{x}_l - \mathbf{x}_j), \qquad l \neq i,$$

where $\prod$ denotes the vector cross-product as given in Definition 2. Then we can define

$$g_k(\mathbf{x}) = \begin{cases} \mathbf{n}_{(k)}\cdot(\mathbf{x}-\mathbf{x}_i), & \mathbf{n}_{(k)}\cdot(\mathbf{x}_k-\mathbf{x}_i) > 0, \\ -\mathbf{n}_{(k)}\cdot(\mathbf{x}-\mathbf{x}_i), & \mathbf{n}_{(k)}\cdot(\mathbf{x}_k-\mathbf{x}_i) < 0, \end{cases} \quad \text{for any } i \neq k$$

and $\mathbf{x} \in R$ if and only if $g_k(\mathbf{x}) > 0\ \forall\ k = 1, \dots, n+1$.

Thus in practice, for a given region R, with vertices $\{\mathbf{x}_1, \dots, \mathbf{x}_{n+1}\}$, we can define the functions $g_k(\mathbf{x})$, $k = 1, \dots, n+1$ and then $\mathbf{x} \in R$ if and only if $g_k(\mathbf{x}) > 0\ \forall\ k$.

5.3. *Calculating the hypervolume of R.* Finally, we show how we can calculate the hypervolume of R, given only the vertices $\{\mathbf{x}_1, \dots, \mathbf{x}_{n+1}\}$. We begin with the case $n = 2$. If we denote the vector $(\mathbf{x}_1 - \mathbf{x}_i)$ by $\mathbf{a}_i$, $i = 2, 3$, then R is a triangular region, with area given by

$$\begin{aligned} \tfrac{1}{2}\text{base} \times \text{height} &= \tfrac{1}{2}|\mathbf{a}_2 \times \mathbf{a}_3| \\ &= \tfrac{1}{2}\left|\det(\mathbf{a}_2\mathbf{a}_3)\right|, \end{aligned} \tag{18}$$

where θ is the angle between the vectors $\mathbf{a}_2$ and $\mathbf{a}_3$. We can generalize this result to higher dimensions via the following theorem.

THEOREM 4. *Given $n+1$ points in n-space, $\mathbf{x}_1, \dots, \mathbf{x}_{n+1}$, if $\mathbf{a}_i = \mathbf{x}_1 - \mathbf{x}_i$, $i \neq 1$, then the hypervolume of the n-dimensional hypertriangle contained within those points is given by*

$$V_n = \frac{1}{n!}\left|\det(\mathbf{a}_2, \mathbf{a}_3, \dots, \mathbf{a}_{n+1})\right|. \tag{19}$$

The proof is in the Appendix.

Thus in practice, (19) may be used to calculate the hypervolume of the region R in the final steps of Algorithms 1 and 2.

Note that this entire section deals with the use of the derivative-free bounds on h to gain an upper bound on P. Unfortunately, there is no similar procedure for using the tangent-based bounds on h to bound P, since the region beyond a particular "face" of the convex hull is bounded above by a collection of hyperplanes, rather than just one. Thus, the tangent-based equivalent of Theorem 3 requires two lemmas providing simple, computable values of

$$J_n = \int_0^\infty \cdots \int_0^\infty \exp\Big[\min_i (a_i x_i + b)\Big]\, dx_1 \cdots dx_n$$

and

$$I_n = \int_0^{s_1} \cdots \int_0^{s_n} \exp\Big[\min_i (a_i \lambda_i)\Big]\, d\lambda_1 \cdots d\lambda_n$$

to replace Lemmas A.5 and A.6 (see the Appendix).

In general, some form of numerical approach would be necessary to approximate these integrals. However, the whole point of forming these piecewise linear bounds to h is to avoid such integration, and thus the tangent-based method is implementationally no less complex than calculating P directly, by the same numerical methods. Thus, only the derivative-free bounds are of any practical value in terms of forming bounds on the proportion, P.

6. Extending the implementation. Elekes (1986) shows that the volume of the convex hull formed from any m points on an n-dimensional hypersphere with volume V, is at most $mV/2^n$. This suggests that the number of points necessary to gain a reasonable estimate of V will increase exponentially with n and so the number of points used to form our upper bound on the area in the tails should increase at least exponentially with dimension and not simply be limited to $n+1$ points. Thus, we may choose an alternative implementation of our diagnostic where, instead of looking at hypertriangles formed from $n+2$ points, we take the convex hull comprising of a set of $k \geq n+2$ points and allow k to increase exponentially with n. In this case we require the following implementational algorithm.

Algorithm 3. *Implementation for general hulls.*

STEP 1. Given output from any sampler, calculate the convex hull of the sample path, recording the points $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ which form the hull and the faces $\gamma_1, \ldots, \gamma_d$, where each face i, is composed of points $\gamma_i = \{\mathbf{x}_{\gamma_i(1)}, \mathbf{x}_{\gamma_i(2)}, \ldots, \mathbf{x}_{\gamma_i(n)}\}$. Record also a point $\mathbf{x}_0$, such that $\mathbf{x}_0$ lies within the hull formed by points $\{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$ and $h(\mathbf{x}_0) > h(\mathbf{x}_i)$ $\forall\ i = 1, \ldots, k$.

STEP 2. Calculate an upper bound U to the volume under h beyond the convex hull, and a lower bound L, to the volume within the hull.

STEP 3. Calculate $P_b' = U/(U + L)$.

If $P'_b < \varepsilon$, then we may consider the sample path to have adequately covered the parameter space. If not, then we may wish to continue the simulation or consider some form of reparameterization in order to improve the sampler's performance properties.

In terms of practical implementation, Step 1 can be performed using any standard package for computing convex hulls. In particular, the Quickhull program [Barber, Dobkin and Huhdanpaa (1993)] can be used to provide the required vertex and face information, given only the sampler output. In Step 2, the bounds U and L are computed in a similar manner to the previous method and are stated explicitly in the following corollary to Theorem 3.

COROLLARY 1. *Given a convex hull with vertices* $\{\mathbf{x}_1, \dots, \mathbf{x}_k\}$, *arranged into* d *faces* $\gamma_1, \dots, \gamma_d$, *where face* i, *is composed of points* $\gamma_i = \{\mathbf{x}_{\gamma_i(1)}, \mathbf{x}_{\gamma_i(2)}, \dots, \mathbf{x}_{\gamma_i(n)}\}$, *we can form an upper bound* U, *to the volume beyond the convex hull*, *and a lower bound* L, *to the volume within the convex hull*, *so that the proportion in the tails beyond the convex hull* P *is bounded above by*

$$P'_b = \frac{U}{U+L},$$

where

$$\begin{aligned}
L &= \sum_{i=1}^{d} L_i,\\
L_i &= |\mathbf{J}_i| e^b \left(\sum_{j \in \gamma_i} \frac{\exp(a_{ij})}{\prod_{k \in \gamma_i,\, k \neq j} (a_{ij} - a_{ik}) a_{ij}} + \frac{(-1)^n}{\prod_{j \in \gamma_i} a_{ij}} \right),
\end{aligned} \tag{20}$$

$$\begin{aligned}
U &= \sum_{i=1}^{n+1} U_i\\
U_i &= |\mathbf{J}_i| e^b \left[-\sum_{j \in \gamma_i} \frac{\exp(a^*_{ij})}{\prod_{k \in \gamma_i,\, k \neq j} (a^*_{ij} - a^*_{ik}) a^*_{ij}} \right],
\end{aligned}$$

$$\begin{aligned}
\mathbf{J}'_i &= \left[(\mathbf{x}_0 - \mathbf{x}_{\gamma_i(1)}), \dots, (\mathbf{x}_0 - \mathbf{x}_{\gamma_i(n)}) \right],\\
a_{ij} &= (\mathbf{x}_0 - \mathbf{x}_j) \cdot \mathbf{n}_i, \qquad j \in \gamma_i,\\
b &= h(\mathbf{x}_0),\\
a^*_{ij} &= -\frac{(\mathbf{x}_0 - \mathbf{x}_j) \cdot \mathbf{n}^*_i}{N^*_{i(n+1)}}, \qquad j \in \gamma_i,
\end{aligned} \tag{21}$$

where $\mathbf{N}_i = (\mathbf{n}_i, -1)$ *is the normal to the hyperplane formed from the set of vectors* $\{(\mathbf{X}_0 - \mathbf{X}_j)\colon j \in \gamma_i\}$, *and* $\mathbf{N}^*_i$ *denotes the corresponding tilted normals as defined in* (12).

The proof is identical to that of Theorem 3, allowing for the new notation.

The advantage of this method, in terms of the tightness of the bound on P, is obvious. We would expect that, by adding more points to the hull, we get much tighter bounds on h and therefore on both U and L. The disadvantage of this method is that we might expect that, for high-dimensional problems, the number of faces forming the convex hull may become quite large, leading to increased computational expense. Both of these points are addressed in the next section, which discusses the two derivative-free approaches in terms of their practical utility.

7. A practical examination of the methods. In order to examine how the two proposed methods might be expected to perform in practice, we present an example based upon the multivariate normal density. Here we take the target distribution π to be a d-dimensional normal, and we sample n points, uniformly distributed over the surface of the d-dimensional hypersphere of radius r, and centered about the origin. We examine the performance of the two methods for various values of d, n and r.

In order to calculate the convex hulls of a particular sample, we use the Quickhull algorithm of Barber, Dobkin and Huhdanpaa (1993). However, we find ourselves restricted to problems of five dimensions or less, due to the huge computational storage demands of problems in higher dimensions, primarily because of the large number of faces associated with even a small number of points in high dimensions.

Table 1 presents the number of faces associated with differing numbers of vertices on the convex hull and for different dimensions. We can see that the number of faces increases with the number of vertices in the hull and that the ratio of faces to vertices increases with dimension, so that a five-dimensional hull with 1500 points possesses over 40,000 faces. Thus, we restrict ourselves to $d \leq 5$.

Figure 7 plots the true proportion P, beyond the hypersphere of radius $r \in (0, 10]$, under the d-variate normal and for dimensions $d = 1, \dots, 5$. For each r, we obtain a sample of $n = 500$ observations, uniformly distributed over the surface of a hypersphere of radius r, and add a single point $\mathbf{x}_0$, at the origin. Given this sample, we plot the two bounds, P_b and P'_b, on P. Note

TABLE 1
Number of faces of the convex hull with v vertices of dimension, d

	Number of vertices, v						
d	**50**	**100**	**200**	**500**	**750**	**1000**	**1500**
2	50	100	200	500	750	1000	1500
3	96	196	396	996	1496	1996	2996
4	254	564	1238	3253	4853	6548	9895
5	772	1936	4536	12842	20204	27392	42060

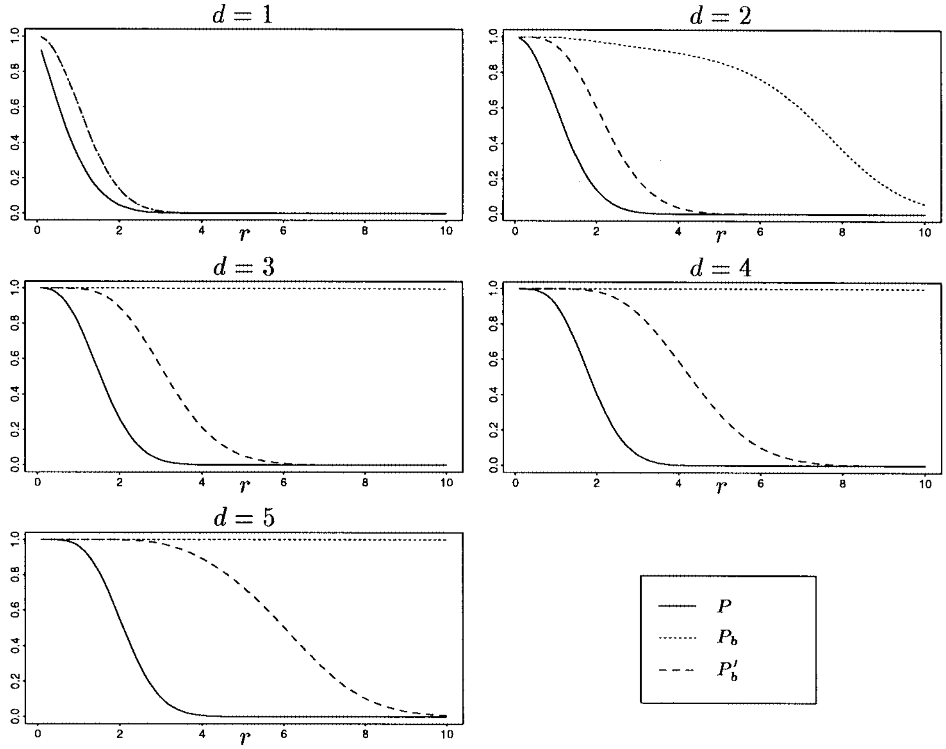

FIG. 7. *True proportion P, and bounds P_b and P'_b calculated for $n = 500$ points on a hypersphere of radius r in d dimensions, for $d = 1, 2, 3, 4, 5$.*

that since the points are all on the surface of a hypersphere, all n points will lie on the convex hull.

We can see that for $d = 1$, the two methods provide identical bounds, since there exists only two distinct points on the one-dimensional hypersphere of radius r and thus both methods use the same three points (two on the hypersphere, one at the origin). The bound in one dimension is actually quite good and would generally be acceptable in most applications. However, in one dimension there are considerably better methods for estimating (or even calculating) P, so the one-dimensional case is not a realistic one.

In higher dimensions, we observe that P_b is a very poor bound on P and, for $d \geq 3$, lies permanently close to 1, with no substantial improvement observed by taking $n > 500$. Table 2 provides the value of P'_b for r-values corresponding to $P = 0.05$, 0.1, 0.2 and 0.5, for $d = 1, \ldots, 5$. From Table 2 and Figure 7, we can see that the alternative bound P'_b performs reasonably well but, as we might expect for fixed $n = 500$, the bound weakens considerably as d increases.

TABLE 2
Value of P_b' corresponding o each value of P, for various dimensions

	Dimension				
P	**1**	**2**	**3**	**4**	**5**
0.05	0.14	0.39	0.63	0.84	0.95
0.1	0.25	0.53	0.75	0.90	0.97
0.2	0.43	0.69	0.85	0.95	0.99
0.5	0.8	0.92	0.96	0.99	1.00

Table 3 provides the execution times for the code to produce the plots of Figure 7. It should be noted that each plot is based upon 100 different r-values, so execution times are for the calculation of 100 bounds in each case. Clearly, the calculation of P_b is considerably faster than that of P_b', in dimensions $d > 1$, with each P_b' bound taking approximately two minutes for $n = 500$ and $d = 5$. In practice, we would only calculate the bounds once, so two minutes is perhaps not unreasonable, but in order to get a tighter bound in higher dimensions, we need to increase n, which leads to a substantial increase in the computational expense, as we shall see.

Figure 8 provides the plots of P_b' for $r \in (0, 10]$ and for different values of n and d. From these plots, it is clear that increasing n improves the bound, but that n must increase at least exponentially in order for the improvement to continue. This suggests that even if computational storage were not a problem, the number of points required to produce an acceptably "tight" bound may be prohibitive in higher dimensions.

Table 4 provides the execution times for the code to produce the output plotted in Figure 8 for different values of d and n. We can see that for fixed d, the execution time increases exponentially with n and similarly for fixed n, the execution time increases exponentially with d. This is due to the exponentially increasing number of faces to the convex hull, as d and n increase. You can see that for 1500 points in five dimensions, the calculation of the bound for a single value of r takes approximately 16 minutes and yet the bound on P remains very weak. In order to gain a tighter bound on P,

TABLE 3
Average execution times for the computation of the two bounds, for 100 different r-values[a]

	Dimension				
	1	**2**	**3**	**4**	**5**
P_b	0.05	0.42	1.05	2.07	3.48
P_b'	0.05	205	637	2661	13316

[a]Times are in seconds.

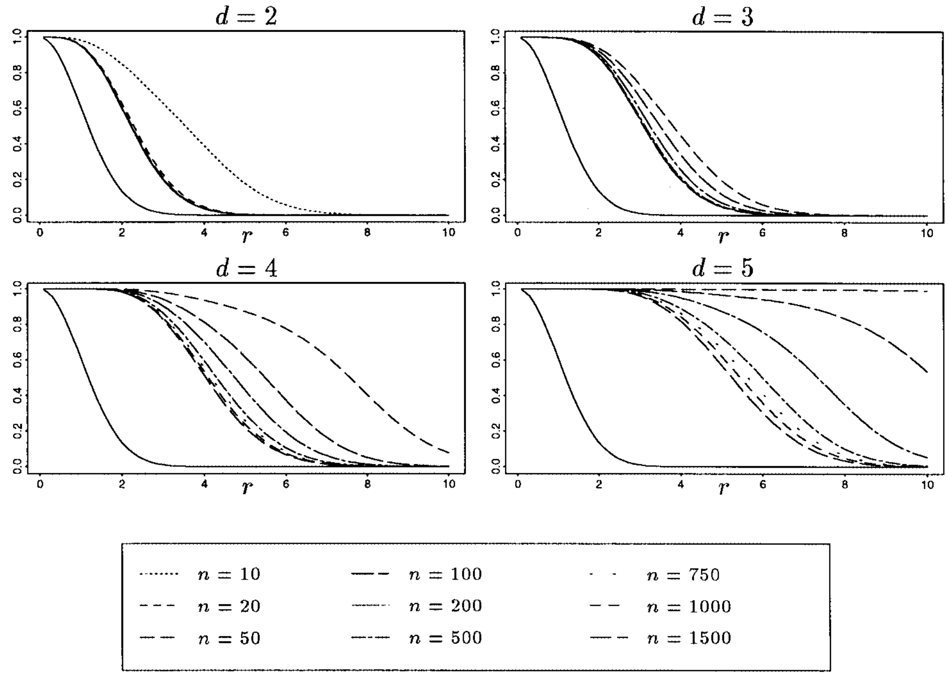

FIG. 8. *True proportion P (solid line) and bound P_b' calculated for $n \in \{10, 20, 50, 100, 200, 500, 750, 1000, 1500\}$ points on a hypersphere of radius r in d dimensions, for $d = 2, 3, 4, 5$.*

the necessary increase in the number of points would make the calculation of P_b' prohibitively time-consuming.

In summary, the results of this section indicate that the bound P_b is generally too weak and is of little practical value. The bound P_b' performs considerably better, providing a reasonable bound that improves with the number of points used. The computation time appears to grow exponentially with dimension, as does the number of points required to obtain an accurate

TABLE 4
Average execution time for the calculation of P_b' for 100 different r-values[a]

	Number of vertices, *n*						
d	**50**	**100**	**200**	**500**	**750**	**1000**	**1500**
2	5.50	14.3	42.2	210	446	765	1681
3	23.6	61.0	147	629	1288	2115	4589
4	89.9	246	639	2693	5218	8558	18058
5	405	1096	3107	13366	26806	44714	94808

[a]Times are in seconds.

bound. Thus, there is a trade-off between the strength of the bound and the computational expense incurred in higher dimensions.

8. Discussion. We have shown how the piecewise linear upper and lower bounds can be used to determine how well the output from an MCMC sampler explores the target density, in the case of log-concavity.

We provide a bound to the proportion of the sample space outside the convex hull of a sample of points and show how close this bound can be to the true value, given an "optimal" sample of points. Since, in general, the sampler output will not include points on the boundary maximizing the region R, nor the optimal point $\mathbf{x}_0$ located at the mode, we investigated how well the bound performs given a set of points sampled directly from the target density. We showed that, given such a sample, the bound provides a strong indication that the target is well explored, given a sample of sufficient size.

The method is a useful diagnostic in its own right, in that it determines when the sampler output may be a reliable basis for inference upon the stationary density. If P_{bound} is small, then the parameter space has been largely explored and there is little chance that significant portions of the space have not been visited by the sampler. Thus the method may also be useful if applied in conjunction with convergence diagnostics, such as that of Gelman and Rubin (1992). Many diagnostic methods based upon the sampler output rely on the assumption that the sampled points cover the entire sample space. There are many examples where diagnostics can spuriously indicate convergence, while large portions of the sample space remain unexplored. Thus, this method may be applied in conjunction with existing convergence diagnostics in order to ensure that such diagnostics are not being misled by "poor" samples. There may also be an additional application of these piecewise linear bounds.

Gilks and Wild (1992) use piecewise exponential bounds on a univariate density π, to provide upper and lower bounds, which can be used as *envelope* and *squeezing* functions to produce an adaptive rejection sampling method for sampling from π. The multivariate bounds described in the early sections of this chapter may be similarly used to provide a *multivariate* adaptive rejection sampling algorithm. Clearly, in high dimensions, computational expense will remain a problem, but for moderate dimensions, approximate hulls may be used to bound the true density π, with little loss in efficiency, due to the associated increase in the rejection rate. The application of these bounds to produce such an algorithm is the focus of current work and beyond the scope of this paper.

APPENDIX

Proofs of theorems. We begin with the following lemma, essentially proving that the upper bound described for the one-dimensional derivative-free method is indeed an upper bound for all points in the tail.

LEMMA A.1. *Take two points $x_0 < x_1 \in \mathbb{R}$ and a concave function h. Then, if $x(\lambda) = x_0 + \lambda(x_1 - x_0)$ and $u(x)$ is such that $\mathbf{X} = [(x, u(x)]$ lies on the chord passing through points $\mathbf{X}_0$ and $\mathbf{X}_1$, that is,*

$$u[x(\lambda)] = h(x_0) + \lambda[h(x_0) - h(x_1)] \qquad \forall \lambda \in \mathbb{R},$$

then $u[x(\lambda)] \geq h[x(\lambda)]\ \forall\ \lambda \notin (0,1)$, that is, the chord between the points $\mathbf{X}_0$ and $\mathbf{X}_1$ lies above the curve h for all x beyond the region (x_0, x_1).

PROOF. From the mean value theorem, $\exists\ \tilde{x}$ such that $x_0 < \tilde{x} < x_1$, and

$$h'(\tilde{x}) = \frac{h(x_0) - h(x_1)}{x_0 - x_1},$$

where h' denotes either the left or right derivative, both of which must exist and be nonincreasing by concavity of h. Consider the case $\lambda \geq 1$. We have that $h'(x) \leq h'(\tilde{x})\ \forall\ x > \tilde{x}$, by concavity, and $h'(\tilde{x}) = u'(x)\ \forall\ x \in \mathbb{R}$, by definition. Therefore, $h'(x) \leq u'(x)\ \forall\ x > \tilde{x}$, and hence

$$h[x(\lambda)] - h(x_1) \leq u[x(\lambda)] - u(x_1) \qquad \forall \lambda \geq 1,$$

by definition of h' and u'. Finally, since $h(x_1) = u(x_1)$,

$$h[x(\lambda)] \leq u[x(\lambda)] \qquad \forall \lambda \geq 1.$$

The proof follows similarly for the case $\lambda \leq 0$. □

We extend this result to the two-dimensional case, showing that the above method does indeed produce a plane, bounding h above for all points "beyond" the line $x_i - x_j$.

If we let

$$u_{ij}(\mathbf{x}) = \frac{(\mathbf{x}^*_{ij} - \mathbf{x}) \cdot \mathbf{n}^*_{ij}}{N_{(ij)(n+1)}} + h(\mathbf{x}^*_{ij})$$

denote the upper bound to h given by the tilted plane with normal $\mathbf{N}^*_{ij}$, then we can extend Lemma A.1 to the two-dimensional case, via Theorem 1, which requires the following (trivial) lemmas.

LEMMA A.2. *If $\mathbf{v}(\mu, \lambda) = \mu\mathbf{v}_1 + \lambda\mathbf{v}_2$ denotes a general direction vector, where $\mathbf{v}_1 = \mathbf{x}_i - \mathbf{x}_j$ and $\mathbf{v}_2 = \mathbf{x}^*_{ij} - \mathbf{x}_0$ for some i and j, then*

$$\dot{u}_{ij,\mathbf{v}(\mu,\lambda)}(\mathbf{x}^*_{ij}) = -\lambda \frac{\mathbf{v}_2 \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}}, \tag{22}$$

where $\dot{u}_{ij,\mathbf{v}}(\mathbf{x})$ denotes the derivative of u_{ij} in the direction $\mathbf{v}$ at the point $\mathbf{x}$.

PROOF. Given a function g, and the general direction vector $\mathbf{v}$, we define the derivative of g at some point $\mathbf{x}$ in the direction $\mathbf{v}(\mu, \lambda)$ by

$$\dot{g}_{\mathbf{v}} = \lim_{\tau \to 0} \frac{g(\mathbf{x} + \tau\mathbf{v}) - g(\mathbf{x})}{\tau}.$$

Then, the derivative of u_{ij} in direction $\mathbf{v}(\mu, \lambda)$ is defined by

$$\begin{aligned}
&\dot{u}_{ij,\mathbf{v}(\mu,\lambda)}(\mathbf{x})\\
&\quad = \lim_{\tau \to 0} \frac{u_{ij}[\mathbf{x} + \tau \mathbf{v}(\mu, \lambda)] - u_{ij}(\mathbf{x})}{\tau}\\
&\quad = \lim_{\tau \to 0} \frac{\dfrac{\left[\mathbf{x}^*_{ij} - \mathbf{x} - \tau \mathbf{v}(\mu, \lambda)\right] \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} + h(\mathbf{x}^*_{ij}) - \dfrac{(\mathbf{x}^*_{ij} - \mathbf{x}) \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} - h(\mathbf{x}^*_{ij})}{\tau}\\
&\quad = \frac{-\mathbf{v}(\mu, \lambda) \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}}\\
&\quad = -\mu \frac{\mathbf{v}_1 \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} - \lambda \frac{\mathbf{v}_2 \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}}.
\end{aligned}$$

However, $\mathbf{v}_1 \cdot \mathbf{n}^*_{ij} = 0$, since $\mathbf{v}_1$ is in the plane to which $\mathbf{n}^*_{ij}$ is perpendicular. Hence

$$\dot{u}_{ij,\mathbf{v}(\mu,\lambda)}(\mathbf{x}^*_{ij}) = -\lambda \frac{\mathbf{v}_2 \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}}. \qquad \square$$

LEMMA A.3. *With* $\mathbf{v}_1$ *and* $\mathbf{v}_2$ *as defined above, we obtain the following results for the derivatives of* h *and* u:

$$\dot{h}_{\mathbf{v}_1}(\mathbf{x}^*_{ij}) = \dot{u}_{ij}, \qquad \mathbf{v}_1(\mathbf{x}^*_{ij}) = 0 \tag{23}$$

and

$$0 \geq \dot{u}_{ij,\mathbf{v}_2}(\mathbf{x}^*_{ij}) \geq \dot{h}_{\mathbf{v}_2}(\mathbf{x}^*_{ij}). \tag{24}$$

PROOF. Assume that h is differentiable everywhere on D, then $\dot{h}_{\mathbf{v}_1}(\mathbf{x}^*_{ij}) = 0$, by definition of $\mathbf{x}^*_{ij}$ as the point maximizing $h(\cdot)$ over points on the line between $\mathbf{x}_i$ and $\mathbf{x}_j$. Clearly, $\dot{h}_{\mathbf{v}_1}(\mathbf{x}^*_{ij})$ can be obtained by setting $\lambda = 0$ in (22). Thus, we have

$$\dot{h}_{\mathbf{v}_1}(\mathbf{x}^*_{ij}) = \dot{u}_{ij,\mathbf{v}_1}(\mathbf{x}^*_{ij}) = 0.$$

The line $\mathbf{X}^*_{ij} + \lambda \mathbf{V}_2$ (where $\mathbf{V}_2 = \mathbf{X}^*_{ij} - \mathbf{X}_0$), crosses the curve h at $\lambda = -1$ (corresponding to $\mathbf{X}_0$) and $\lambda = 0$ (corresponding to $\mathbf{X}^*_{ij}$), by definition of $\mathbf{X}_0$ and $\mathbf{X}^*_{ij}$. However, a single line of the form $\mathbf{X} + \lambda \mathbf{V}$ can cross the surface h, at most twice, by concavity of h, and will lie below h in between the points of intersection and above h for all other points, by Lemma A.1. Thus, Lemma A.1 ensures that

$$u_{ij}\left(\mathbf{x}^*_{ij} + \lambda \mathbf{v}_2\right) \geq h\left(\mathbf{x}^*_{ij} + \lambda \mathbf{v}_2\right) \qquad \forall \lambda \geq 0 \text{ and } \lambda \leq -1$$

and, since $u_{ij}(\mathbf{x}^*_{ij}) = h(\mathbf{x}^*_{ij})$,

$$0 \geq u_{ij}\left(\mathbf{x}^*_{ij} + \lambda \mathbf{v}_2\right) - u_{ij}(\mathbf{x}^*_{ij}) \geq h\left(\mathbf{x}^*_{ij} + \lambda \mathbf{v}_2\right) - h(\mathbf{x}^*_{ij}) \qquad \forall \lambda \geq 0.$$

This implies that

$$0 \geq \dot{u}_{ij,\mathbf{v}_2}(\mathbf{x}^*_{ij}) \geq \dot{h}_{\mathbf{v}_2}(\mathbf{x}^*_{ij}),$$

by definition of $\dot{u}_{ij,\mathbf{v}}$ and $\dot{h}_{\mathbf{v}}$.

Similar proofs may be constructed for the case where h is not differentiable everywhere on D, due to the fact that left and right derivatives will be so defined, by the concavity of h. □

LEMMA A.4. *Let $m_{ij}(\mathbf{x})$ denote the "height" of the tangent plane to h at $\mathbf{x}^*_{ij}$, evaluated at the point $\mathbf{x}$, that is,*

$$m_{ij}(\mathbf{x}) = (\mathbf{x} - \mathbf{x}^*_{ij}) \cdot \mathbf{n}_m + h(\mathbf{x}^*_{ij}),$$

where

$$\mathbf{n}_m = \frac{\partial h}{\partial \mathbf{x}}(\mathbf{x}^*_{ij}).$$

Then

$$\dot{m}_{ij,\mathbf{v}(\mu,\lambda)}(\mathbf{x}) = \mu \mathbf{v}_1 \cdot \mathbf{n}_m + \lambda \mathbf{v}_2 \cdot \mathbf{n}_m \qquad \forall\, \mathbf{x} \in \mathbb{R}, \tag{25}$$

with the special cases that

$$\dot{m}_{ij,\mathbf{v}_1}(\mathbf{x}^*_{ij}) = 0,$$

and

$$\dot{u}_{ij,\mathbf{v}_2}(\mathbf{x}^*_{ij}) \geq \dot{m}_{ij,\mathbf{v}_2}(\mathbf{x}^*_{ij}), \tag{26}$$

where $\dot{m}_{ij,\mathbf{v}}(\mathbf{x})$ denotes the derivative of m_{ij} in the direction $\mathbf{v}$, at the point $\mathbf{x}$.

PROOF. By definition,

$$\begin{aligned}
&\dot{m}_{ij,\mathbf{v}(\mu,\lambda)}(\mathbf{x}) \\
&\quad = \lim_{\tau \to 0} \frac{\left[\mathbf{x} - \mathbf{x}^*_{ij} + \tau \mathbf{v}(\mu,\lambda)\right] \cdot \mathbf{n}_m + h(\mathbf{x}^*_{ij}) - (\mathbf{x} - \mathbf{x}^*_{ij}) \cdot \mathbf{n}_m - h(\mathbf{x}^*_{ij})}{\tau} \\
&\quad = \mathbf{v}(\mu,\lambda) \cdot \mathbf{n}_m \\
&\quad = \mu \mathbf{v}_1 \cdot \mathbf{n}_m + \lambda \mathbf{v}_2 \cdot \mathbf{n}_m, \qquad \forall\, \mathbf{x} \in \mathbb{R}.
\end{aligned}$$

Clearly, $\dot{m}_{ij,\mathbf{v}}(\mathbf{x}^*_{ij}) = \dot{h}_{\mathbf{v}}(\mathbf{x}^*_{ij})\ \forall\, \mathbf{v} \in \mathbb{R}^2$, by definition of m as the tangent to h at $\mathbf{x}^*_{ij}$. Therefore, (23) of Lemma A.3 implies that

$$\dot{m}_{ij,\mathbf{v}_1}(\mathbf{x}^*_{ij}) = 0$$

and (24) of Lemma A.3 implies that

$$0 \geq \dot{u}_{ij,\mathbf{v}_2}(\mathbf{x}^*_{ij}) \geq \dot{h}_{\mathbf{v}_2}(\mathbf{x}^*_{ij}) = \dot{m}_{ij,\mathbf{v}_2}(\mathbf{x}^*_{ij})$$

and the result follows. □

PROOF OF THEOREM 1. Since

$$m_{ij}\left[\mathbf{x}_{ij}(\mu,\lambda)\right] \geq h\left[\mathbf{x}_{ij}(\mu,\lambda)\right] \qquad \forall\, \mu, \lambda \in \mathbb{R},$$

by definition of m_{ij} as the tangent to h at $\mathbf{x}^*_{ij}$ and the concavity of h. Thus, it is sufficient to prove that

$$m_{ij}\left[\mathbf{x}_{ij}(\mu, \lambda)\right] \leq u_{ij}\left[\mathbf{x}_{ij}(\mu, \lambda)\right] \qquad \forall\, \mu > 0 \text{ and } \lambda \in \mathbb{R}.$$

Now,

$$\begin{aligned}
&u_{ij}\left[\mathbf{x}_{ij}(\mu, \lambda)\right] - m_{ij}\left[\mathbf{x}_{ij}(\mu, \lambda)\right] \\
&\quad = \frac{\left[\mathbf{x}^*_{ij} - \mathbf{x}_{ij}(\mu, \lambda)\right] \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} + h(\mathbf{x}^*_{ij}) - \left[\mathbf{x}_{ij}(\mu, \lambda) - \mathbf{x}^*_{ij}\right] \cdot \mathbf{n}_m - h(\mathbf{x}^*_{ij}) \\
&\quad = \left[\mathbf{x}^*_{ij} - x_{ij}(\mu, \lambda)\right] \cdot \left(\frac{\mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} + \mathbf{n}_m \right) \\
&\quad = \left(\mathbf{x}^*_{ij} - \mathbf{x}^*_{ij} - \lambda \mathbf{v}_1 - \mu \mathbf{v}_2\right) \cdot \left(\frac{\mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} + \mathbf{n}_m \right) \\
&\quad = -\lambda \frac{\mathbf{v}_1 \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} - \mu \frac{\mathbf{v}_2 \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} - \lambda \mathbf{v}_1 \cdot \mathbf{n}_m - \mu \mathbf{v}_2 \cdot \mathbf{n}_m,
\end{aligned}$$

where $\mathbf{v}_1$ and $\mathbf{v}_2$ are the vectors defined in Lemma A.2.

However, $\mathbf{v}_1 \cdot \mathbf{n}^*_{ij} = 0$, by definition of $\mathbf{n}^*_{ij}$, and $\mathbf{v}_1 \cdot \mathbf{n}_m = 0$, since $\dot{m}_{ij,\mathbf{v}_1}(\mathbf{x}^*_{ij}) = \lambda \mathbf{v}_1 \cdot \mathbf{n}_m = 0$, by Lemma A.4. Thus,

$$u_{ij}\left[\mathbf{x}_{ij}(\mu, \lambda)\right] - m_{ij}\left[\mathbf{x}_{ij}(\mu, \lambda)\right] = -\mu \frac{\mathbf{v}_2 \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} - \mu \mathbf{v}_2 \cdot \mathbf{n}_m,$$

but

$$-\frac{\mathbf{v}_2 \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} \geq \mathbf{v}_2 \cdot \mathbf{n}_m$$

by substituting the definitions of $\dot{u}_{ij,\mathbf{v}_2}(\cdot)$ and $\dot{m}_{ij,\mathbf{v}_2}(\mathbf{x}^*_{ij})$ from (22) and (25) into (26) of Lemma A.4. Hence,

$$\begin{aligned}
u_{ij}\left[\mathbf{x}_{ij}(\mu, \lambda)\right] - m_{ij}\left[\mathbf{x}_{ij}(\mu, \lambda)\right] &= -\mu \frac{\mathbf{v}_2 \cdot \mathbf{n}^*_{ij}}{N^*_{(ij)(n+1)}} - \mu \mathbf{v}_2 \cdot \mathbf{n}_m \\
&> 0 \quad \text{iff } \mu > 0.
\end{aligned}$$

Hence $u_{ij}[\mathbf{x}_{ij}(\mu, \lambda)] > m_{ij}[\mathbf{x}_{ij}(\mu, \lambda)]\ \forall\, \mu > 0$, and the result follows. □

Before we prove Theorem 3, we require the following lemmas.

LEMMA A.5.

$$\begin{aligned}
J_n &= \int_0^\infty \exp\left(\sum_{i=1}^n a_i x_i + b \right) dx_1 \cdots dx_n \\
&= (-1)^n e^b \prod_{i=1}^n \frac{1}{a_i}, \qquad a_i < 0, \forall\, i
\end{aligned}$$

The proof is trivial.

LEMMA A.6. *If the integral* I_n *is defined as*

$$I_n = \int_0^{s_1} \cdots \int_0^{s_n} \exp\left(\sum_{j=1}^{n} a_j \lambda_j \right) d\lambda_1 \cdots d\lambda_n, \qquad n \geq 2,$$

where $s_1 = 1$ *and* $s_i = 1 - \sum_{j<i} \lambda_j$, $i = 2, \dots, n$, *then*

$$I_n = \sum_{i=1}^{n} \frac{\exp(a_i)}{\prod_{j \neq i}(a_i - a_j) a_i} + \frac{(-1)^n}{\prod_{j=1}^{n} a_i}$$

PROOF. By induction, clearly,

$$\begin{aligned} I_2 &= \int_0^1 \int_0^{1-\lambda_1} \exp(a_1 \lambda_1 + a_2 \lambda_2)\, d\lambda_2\, d\lambda_1 \\ &= \int_0^1 \frac{1}{a_2} \exp((a_1 - a_2)\lambda_1 + a_2) - \frac{1}{a_2} \exp(a_1 \lambda_1)\, d\lambda_1 \\ &= \frac{\exp(a_1)}{a_1(a_1 - a_2)} + \frac{\exp(a_2)}{a_2(a_2 - a_1)} + \frac{1}{a_1 a_2}. \end{aligned}$$

Hence, the result is true for $n = 2$. Now, assume it is true for $n = k$, say, then

$$\begin{aligned} &\int_0^{s_1} \cdots \int_0^{s_{k+1}} \exp\left(\sum_{j=1}^{k+1} a_j \lambda_j \right) d\lambda_1 \cdots d\lambda_{k+1} \\ &\quad = \int_0^{s_1} \cdots \int_0^{s_k} \exp\left(\sum_{j=1}^{k} a_j \lambda_j \right) \int_0^{s_{k+1}} \exp(a_{k+1} \lambda_{k+1})\, d\lambda_1 \cdots d\lambda_{k+1} \\ &\quad = \int_0^{s_1} \cdots \int_0^{s_k} \exp\left(\sum_{j=1}^{k} a_j \lambda_j \right) \left[\frac{\exp(a_{k+1} s_{k+1})}{a_{k+1}} - \frac{1}{a_{k+1}} \right] d\lambda_1 \cdots d\lambda_k \\ &\quad = \frac{1}{a_{k+1}} \int_0^{s_1} \cdots \int_0^{s_k} \exp\left(a_{k+1} + \sum_{j=1}^{k} (a_j - a_{k+1}) \lambda_j \right) d\lambda_1 \cdots d\lambda_k - \frac{1}{a_{k+1}} I_k \\ &\quad = \frac{\exp(a_{k+1})}{a_{k+1}} \sum_{i=1}^{k} \frac{\exp(a_i - a_{k+1})}{\prod_{j \neq i}^{k}(a_i - a_j)(a_i - a_{k+1})} \\ &\qquad + \frac{(-1)^k}{\prod_{i=1}^{k}(a_i - a_{k+1})} \frac{\exp(a_{k+1})}{a_{k+1}} \\ &\qquad - \frac{1}{a_{k+1}} \sum_{i=1}^{k} \frac{\exp(a_i)}{\prod_{j \neq i}^{k}(a_i - a_j) a_i} - \frac{1}{a_{k+1}} \frac{(-1)^k}{\prod_{j=1}^{k} a_i} \end{aligned}$$

since it is true for $n = k$, and replacing a_i by $a_i - a_{k+1}$ in the definition of I_n,

$$
\begin{aligned}
&= \frac{1}{a_{k+1}} \sum_{i=1}^{k} \frac{\exp(a_i)}{\prod_{j \neq i}^{k}(a_i - a_j)} \left[\frac{1}{a_i - a_{k+1}} - \frac{1}{a_i} \right] \\
&\quad + \frac{(-1)^k \exp(a_{k+1})}{\prod_{j \neq k+1}(a_j - a_{k+1}) a_{k+1}} + \frac{(-1)^{k+1}}{\prod_{i=1}^{k+1} a_i} \\
&= \sum_{i=1}^{k} \frac{\exp(a_i)}{\prod_{j \neq i}^{k+1}(a_i - a_j) a_i} + \frac{(-1)^{2k} \exp(a_{k+1})}{\prod_{j \neq k+1}(a_{k+1} - a_j) a_{k+1}} + \frac{(-1)^{k+1}}{\prod_{i=1}^{k+1} a_i} \\
&= \sum_{i=1}^{k+1} \frac{\exp(a_i)}{\prod_{j \neq i}^{k+1}(a_i - a_j) a_i} + \frac{(-1)^{k+1}}{\prod_{j=1}^{k+1} a_i} \\
&= I_{k+1}.
\end{aligned}
$$

Hence, if the result holds for $n = k$, it also holds for $n = k + 1$, and hence for all $n \geq 2$, by induction. □

Given these lemmas, we can now proceed to prove Theorem 3 by detailing the analytic bound on P.

Proof of Theorem 3. To gain the lower bound L to M, split the hull into $n + 1$ regions by taking for each $i = 1, \dots, n + 1$, the n-dimensional hypertriangle formed from the vertices $\{\mathbf{x}_0, \dots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_{n+1}\}$. Now let $\{(\mathbf{x}_j - \mathbf{x}_0): j \neq i\}$ be a set of coordinate axes for points within this hypertriangle, then a general point on the hypertriangular plane is given by

$$
\begin{aligned}
\mathbf{x}_{(i)}(\boldsymbol{\lambda}) &= \mathbf{x}_0 + \sum_{j \neq i} \lambda_j (\mathbf{x}_j - \mathbf{x}_0) \\
&= \mathbf{x}_0 + \boldsymbol{\lambda}'_{(i)} \mathbf{J}_{(i)},
\end{aligned}
\tag{27}
$$

where $\boldsymbol{\lambda}'_{(i)} = (\lambda_1, \dots, \lambda_{i-1}, \lambda_{i+1}, \dots, \lambda_{n+1})$.

Now the height $l[\mathbf{x}_{(i)}(\boldsymbol{\lambda})]$, of this general point will be such that

$$\left[\mathbf{X}_0 - \left(\mathbf{x}_{(i)}(\boldsymbol{\lambda}), l\left[\mathbf{x}_{(i)}(\boldsymbol{\lambda})\right]\right)\right] \cdot \mathbf{N}_{(i)} = 0.$$

Therefore,

$$\mathbf{X}_0 \cdot \mathbf{N}_{(i)} - \mathbf{x}_{(i)}(\boldsymbol{\lambda}) \cdot \mathbf{n}_{(i)} - l\left[\mathbf{x}_{(i)}(\boldsymbol{\lambda})\right] N_{(i)(n+1)} = 0.$$

This implies that

$$
\begin{aligned}
l\left[\mathbf{x}_{(i)}(\boldsymbol{\lambda})\right] &= -\mathbf{X}_0 \cdot \mathbf{N}_{(i)} + \mathbf{x}_{(i)}(\boldsymbol{\lambda}) \cdot \mathbf{n}_{(i)} \quad \text{since } N_{(i)(n+1)} = -1 \\
&= -\mathbf{X}_0 \cdot \mathbf{N}_{(i)} + \mathbf{x}_0 \cdot \mathbf{n}_{(i)} - \sum_{j \neq i} (\mathbf{x}_j - \mathbf{x}_0) \cdot \mathbf{n}_{(i)} \lambda_j \quad \text{by (27)} \\
&= h(\mathbf{x}_0) - \sum_{j \neq i} (\mathbf{x}_j - \mathbf{x}_0) \cdot \mathbf{n}_{(i)} \lambda_j \\
&= b + \sum_{j \neq i} a_{ij} \lambda_j \quad \text{say}
\end{aligned}
$$

Now, within the convex hull given by points $\{\mathbf{x}_0, \ldots, \mathbf{x}_{i-1}, \mathbf{x}_{i+1}, \ldots, \mathbf{x}_{n+1}\}$, l lies below the surface h. Therefore, the hypervolume of $f = e^h$ within this convex hull is given by

$$(28)\quad \begin{aligned} I &= \int_0^{s_1} \cdots \int_0^{s_{i-1}} \int_0^{s_{i+1}} \cdots \int_0^{s_{n+1}} \exp\Big[l\big(\mathbf{x}_{(i)}(\boldsymbol{\lambda})\big)\Big] |\mathbf{J}_{(i)}| \, d\lambda_1 \\ &\qquad \cdots d\lambda_{i-1}\, d\lambda_{i+1} \cdots d\lambda_{n+1} \\ &= |\mathbf{J}_{(i)}| \int_0^{s_1} \cdots \int_0^{s_{i-1}} \int_0^{s_{i+1}} \cdots \int_0^{s_{n+1}} \exp\Big(b + \sum_{j \neq i} a_{ij} \lambda_j\Big) d\lambda_1 \\ &\qquad \cdots d\lambda_{i-1}\, d\lambda_{i+1} \cdots d\lambda_{n+1}, \end{aligned}$$

where

$$s_1 = 1, \qquad s_j = 1 - \sum_{\substack{l \neq i \\ l=1}}^{j-1} \lambda_l, \qquad j = 2, \ldots, n+1,$$

and $|\mathbf{J}_{(i)}|$ is the Jacobian term. Thus, Lemma A.6 gives us that

$$\begin{aligned} I &= |\mathbf{J}_{(i)}| e^b \left(\sum_{j \neq i} \frac{\exp(a_{ij})}{\prod_{k \neq i, j} (a_{ij} - a_{ik}) a_{ij}} + \frac{(-1)^n}{\prod_{j \neq i} a_{ij}} \right) \\ &= L_{(i)} \quad \text{say}. \end{aligned}$$

Now a lower bound to f within the convex hull formed by all of the points $\{\mathbf{x}_0, \ldots, \mathbf{x}_{n+1}\}$ is given by the sum of the lower bounds in each of the $n+1$ regions into which this hull was split. Thus, a lower bound to M, the hypervolume of f within the convex hull, is given by

$$L = \sum_{i=1}^{n+1} L_{(i)}.$$

To gain the upper bound U to T, we begin by splitting the convex hull as before and then use the tilted planes as described in Section 4 to define an upper bound to h beyond this convex hull, denoted by

$$\begin{aligned} u\big[\mathbf{x}_{(i)}(\boldsymbol{\lambda})\big] &= \frac{\mathbf{X}_0 \cdot \mathbf{N}^*_{(i)} - \mathbf{x}_{(i)}(\boldsymbol{\lambda}) \cdot \mathbf{n}^*_{(i)}}{N^*_{(i)(n+1)}} \\ &= h(\mathbf{x}_0) - \sum_{j \neq i} \frac{(\mathbf{x}_0 - \mathbf{x}_j) \cdot \mathbf{n}^*_{(i)}}{N^*_{(i)(n+1)}} \lambda_j \\ &= b + \sum_{j \neq i} a^*_{ij} \lambda_j \quad \text{say}. \end{aligned}$$

Then, the required upper bound for f beyond the convex hull formed from the set of points $\{x_0, \dots, x_{i-1}, x_{i+1}, \dots, x_{n+1}\}$ is given by

$$\begin{aligned}
I &= \int_{\{\boldsymbol{\lambda}:\, \sum \lambda_i \ge 1\}}^{\infty} \exp\big[u\big(\mathbf{x}_{(i)}(\boldsymbol{\lambda})\big)\big] |\mathbf{J}_{(i)}|\, d\lambda_1 \cdots d\lambda_{i-1}\, d\lambda_{i+1} \cdots d\lambda_{n+1} \\
&= |\mathbf{J}_{(i)}| \int_{\{\boldsymbol{\lambda}:\, \sum \lambda_i \ge 1\}}^{\infty} \exp\Big(b + \sum_{j \ne i} a_{ij}^* \lambda_j\Big)\, d\boldsymbol{\lambda}_{(i)} \\
&= |\mathbf{J}_{(i)}| \Bigg[\int_0^{\infty} \cdots \int_0^{\infty} \exp\Big(b + \sum_{j \ne i} a_{ij}^* \lambda_j\Big)\, d\boldsymbol{\lambda}_{(i)} \\
&\qquad - \int_0^{s_1} \cdots \int_0^{s_{i-1}} \int_0^{s_{i+1}} \cdots \int_0^{s_{n+1}} \exp\Big(b + \sum_{j \ne i} a_{ij}^* \lambda_j\Big)\, d\boldsymbol{\lambda}_{(i)} \Bigg] \\
&= |\mathbf{J}_{(i)}| e^b \Bigg[- \sum_{j \ne i} \frac{\exp(a_{ij}^*)}{\prod_{k \ne i, j} (a_{ij}^* - a_{ik}^*) a_{ij}^*} \Bigg] \quad \text{by Lemmas A.5 and A.6} \\
&= U_{(i)} \quad \text{say.}
\end{aligned}$$

Now an upper bound to f within the convex hull formed by all of the points $\{\mathbf{x}_0, \dots, \mathbf{x}_{n+1}\}$ is given by the sum of the upper bounds in each of the $n+1$ regions into which this hull was split. Thus an upper bound to the hypervolume of f in the tails is given by

$$U = \sum_{i=1}^{n+1} U_{(i)}$$

and, since $U \ge T$ and $L \le M$, then

$$P = \frac{T}{T+M} \le \frac{U}{U+L} = P_b. \qquad \square$$

Finally, in order to prove Theorem 4, we require the following lemma.

LEMMA A.7. *Let V_n be the hypervolume of the hypertriangle with vertices $\{\mathbf{x}_1, \dots, \mathbf{x}_{n+1}\}$ then*

$$V_n = \frac{V_{n-1} h}{n}, \tag{29}$$

where V_{n-1} is the $(n-1)$-dimensional hypervolume of the base formed from vertices $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ and h is the perpendicular height of $\mathbf{x}_{n+1}$ from the base.

PROOF. Let A_l be the $(n-1)$-dimensional hypervolume of the section of the n-dimensional hypertriangle truncated parallel to the base at height l. Then

$$A_l = V_{n-1} \frac{(h-l)^{n-1}}{h^{n-1}},$$

since at height l, each side has been reduced by a factor $(h - l)/h$ and thus the volume is reduced by a factor $((h - l)/h)^{n-1}$. Thus,

$$\begin{aligned} V_n &= \int_{l=0}^{h} A_l \, dl \\ &= \int_0^h V_{n-1} \frac{(h-l)^{n-1}}{h^{n-1}} \, dl \\ &= \left[\frac{V_{n-1}}{h^{n-1}} \frac{(h-l)^n}{(-n)} \right]_0^h \\ &= \frac{V_{n-1} h}{n}. \end{aligned}$$ □

Given this lemma, we can prove the result of Theorem 4 for the hypervolume of an n-dimensional hypertriangle.

Proof of Theorem 4. By induction, if $n = 2$, then (19) becomes

$$V_2 = \tfrac{1}{2} |\det(\mathbf{a}_2 \mathbf{a}_3)|,$$

which is simply the area of a triangle, as given in (18). Thus the theorem holds for $n = 2$.

Assume that the theorem holds for $n = k$, say, then

$$V_k = \frac{1}{k!} |\det(\mathbf{a}_2 \cdots \mathbf{a}_{k+1})|.$$

Now introduce a point $\mathbf{A}_{k+2} \in \mathbb{R}^{k+1}$ and, without loss of generality, set $\mathbf{A}_i = (\mathbf{a}_i, 0)\ \forall\ i = 2, \ldots, k+1$, since the set of points $\{\mathbf{A}_i\colon i = 2, \ldots, k+1\}$ will lie on a k-plane in $\mathbb{R}^{k+1}$, which can be arbitrarily transformed to the $\mathbf{x}_{k+1} = 0$ plane by an orthogonal transformation $\mathbf{T}$, such that $\det(\mathbf{T}) = 1$. Then

$$V_{k+1} = \frac{V_k h}{k+1},$$

by Lemma A.7, where h is simply the $(k+1)$th component of $\mathbf{A}_{k+2}$; that is, $h = A_{k+2}^{k+1}$. Therefore,

$$\begin{aligned} V_{k+1} &= \frac{1}{k!} \frac{1}{k+1} |\det(\mathbf{a}_2 \cdots \mathbf{a}_{k+1})| A_{k+2}^{k+1} \\ &= \frac{1}{(k+1)!} |\det(\mathbf{a}_2 \cdots \mathbf{a}_{k+1})| A_{k+2}^{k+1}. \end{aligned}$$

However,

$$\det(\mathbf{A}_2 \cdots \mathbf{A}_{k+2}) = (-1)^{k+1} \det(\mathbf{a}_2 \cdots \mathbf{a}_{k+1}) \mathbf{A}_{k+2}^{k+1},$$

since $\mathbf{A}_j^{k+1} = 0 \ \forall \ j = 2, \dots, k+1$. Thus,

$$V_{k+1} = \frac{1}{(k+1)!} \left| \det(\mathbf{A}_2 \ \cdots \mathbf{A}_{k+2}) \right|$$

so that the result is also true for $n = k+1$, and hence is true for all $n \geq 2$, by induction. □

Acknowledgments. The author expresses his gratitude to an Associate Editor and two anonymous referees for their encouragement and constructive criticism of an earlier draft of this paper. Gratitude is also owed to Neil Boston, Persi Diaconis, Tim Goodman, Yves Tourigny and Gareth Roberts for their invaluable help and suggestions.

REFERENCES

ASMUSSEN, S., GLYNN, P. W. and THORISSON, H. (1992). Stationarity detection in the initial transient problem. *ACM Transactions on Modelling and Computer Simulation* **2** 130–157.

BARBER, C. B., DOBKIN, D. P. and HUHDANPAA, H. (1993). The Quickhull algorithm for convex hull. Technical report, Geometry Center, Univ. Minnesota.

COWLES, M. K. and CARLIN, B. P. (1996). Markov chain Monte Carlo convergence diagnostics: a comparative review. *J. Amer. Statist. Assoc.* **91** 883–904.

DELLAPORTAS, P. and SMITH, A. F. M. (1993). Bayesian inference for generalized linear and proportional hazards models via Gibbs sampling. *J. Roy. Statist. Soc. Ser. C* **42** 443–460.

ELEKES, G. (1986). A geometric inequality and the complexity of computing volume. *Discrete Comput. Geom.* **1** 289–292.

GELFAND, A. E. and SMITH, A. F. M. (1990). Sampling based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85** 398–409.

GELMAN, A. and RUBIN, D. (1992). Inference from iterative simulation using multiple sequences. *Statist. Sci.* **7** 457–511.

GILKS, W. R. and WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *J. Roy. Statist. Soc. Ser. C* **41** 337–348.

MCCULLAGH, P. and NELDER, J. A. (1989). *Generalized Linear Models*. Chapman and Hall, London.

MENGERSEN, K. L. and TWEEDIE, R. L. (1996). Rates of convergence of the Hastings and Metropolis algorithms. *Ann. Statist.* **24** 101–121.

POLSON, N. G. (1996). Convergence of Markov chain Monte Carlo algorithms. In *Bayesian Statistics 5* (J. M. Bernardo, A. F. M. Smith, A. P. Dawid and J. O. Berger, eds.). Oxford Univ. Press.

SMITH, A. F. M. and ROBERTS, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo Methods. *J. Roy. Statist. Soc. Ser. B* **55** 3–24 (with discussion).

SPIVAK, M. (1965). *Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus*. Benjamin, New York.

WEDDERBURN, R. W. M. (1976). On the existence and uniqueness of the maximum likelihood estimates for certain generalised linear models. *Biometrika* **63** 27–32.

SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
UNIVERSITY WALK
BRISTOL BS8 1TW
UNITED KINGDOM
E-MAIL: steve.brooks@bristol.ac.uk