

Approximate Marginal Densities of Nonlinear Functions

By

Luke Tierney, Robert E. Kass and Joseph B. Kadane ¹

Technical Report No. 513

School of Statistics

University of Minnesota

February 12, 1988

¹Research of the authors was supported in part by National Science Foundation Grant DMS-8705646

Abstract

This paper presents an asymptotic approximation for the marginal density of a nonlinear function $g(\theta)$ that is applicable when the joint density of θ is dominated by a single mode and the gradient of g is nonzero or the Jacobian of g is of full rank near that mode. The approximation is based on Laplace's method and its asymptotic properties are similar to those of the saddlepoint approximation. The approximation is applied to the computation of a marginal posterior density, a marginal sampling density and a marginal density based on a multivariate saddlepoint approximation to a joint density.

1 Introduction and summary

We consider a probability density function on \mathfrak{R}^m of the form

$$p(\theta) = cb(\theta) \exp\{-H(\theta)\} \quad (1)$$

where $H(\theta) = nh(\theta)$ and n is assumed to be large. The function H is assumed to have a local minimum at $\hat{\theta}$ that is dominant in the sense that the probability assigned by p to the complement of any neighborhood of $\hat{\theta}$ is assumed to be of exponentially small order in n . In the derivations given below we will take the functions b and h to be fixed. However the arguments remain valid if they are replaced by well-behaved sequences of functions b_n and h_n .

Leonard (1982), Phillips (1983), and Tierney and Kadane (1986) have considered the following approximation for marginal densities of coordinate functions for joint densities of the form (1): If θ is partitioned as (θ_1, θ_2) , with θ_1 a k -dimensional vector and θ_2 an $(m - k)$ -dimensional vector, then the approximate marginal density for θ_1 is given by

$$\hat{p}_1(\theta_1) = (2\pi)^{-k/2} \left(\frac{\det \Sigma(\theta_1)}{\det \Sigma} \right)^{1/2} \frac{p(\theta_1, \hat{\theta}_2(\theta_1))}{p(\hat{\theta})} \quad (2)$$

where Σ is the inverse of the Hessian of H evaluated at $\hat{\theta}$, the vector $\hat{\theta}_2(\theta_1)$ minimizes the function $f(\cdot) = H(\theta_1, \cdot)$, which is the function H with θ_1 held fixed, and $\Sigma(\theta_1)$ is the inverse of the Hessian of the function $H(\theta_1, \cdot)$ evaluated at $\hat{\theta}_2(\theta_1)$. Leonard and Tierney and Kadane are concerned with the approximation of marginal posterior densities while Phillips is primarily concerned with approximating marginal sampling distributions. As Sweeting (1987) points out, approximation (2) used as an approximate marginal posterior density is also closely related to the conditional profile likelihood of Cox and Reid (1987).

Approximation (2) has what might be called *saddlepoint accuracy*. That is, its asymptotic properties are very similar to the properties of the saddlepoint approximation (see, for example, Reid, 1988, and the references cited therein). In particular, (i) the relative errors of the approximation are uniformly of order $O(n^{-1})$ on a fixed neighborhood of the joint mode; (ii) the approximations generally do not integrate exactly to one, but (iii) as a result of the uniformity in the error term numerical renormalization of the approximation produces an approximation with relative errors of order $O(n^{-3/2})$ in neighborhoods of the joint mode that shrink in diameter at the rate $O(n^{-1/2})$. Thus the local behavior of the

renormalized approximations is comparable to the behavior of an Edgeworth expansion with two terms in addition to the leading normal density term. Approximation (2) is, however, more accurate than an Edgeworth expansion in the tails since relative errors of Edgeworth expansions generally only tend to zero on $n^{-1/2}$ -neighborhoods of the mode.

In this note we are concerned with approximating the marginal density of more general functions $g(\theta)$ than simple coordinate functions. The function g can be real-valued or k -dimensional vector-valued and is assumed to be smooth and have a gradient that does not vanish, or a Jacobian that is of rank k , near $\hat{\theta}$. In some cases it may be possible to reduce this problem to the coordinate margin case by finding an explicit global transformation $\phi = \psi(\theta)$ such that ϕ can be partitioned as $\phi = (g(\theta), \phi_2)$. This is the approach taken in Phillips (1983). However, in many cases it may be difficult or inconvenient to produce such an explicit reparameterization. We propose an approximation that does not depend on the explicit specification of a transformation. For a k -dimensional function g the approximation is given by

$$\hat{p}_g(\gamma) = (2\pi)^{-k/2} \left(\frac{\det(\Sigma(\gamma))}{\det(\Sigma) \det((Dg)^T \Sigma(\gamma) (Dg))} \right)^{1/2} \frac{p(\hat{\theta}(\gamma))}{p(\hat{\theta})} \quad (3)$$

where Σ is the inverse of the Hessian of H evaluated at $\hat{\theta}$, $\hat{\theta}(\gamma)$ minimizes the function $H(\cdot)$ subject to the constraint $g(\theta) = \gamma$, $\Sigma(\gamma)$ is the inverse of the Hessian of H evaluated at $\hat{\theta}(\gamma)$, and Dg is the gradient or Jacobian of g evaluated at $\hat{\theta}(\gamma)$. That is, if g is one-dimensional then Dg is a column vector with components $\frac{\partial}{\partial \theta_i} g(\hat{\theta}(\gamma))$ for $i = 1, \dots, m$, and if g is k -dimensional then Dg is an m by k matrix with elements $\frac{\partial}{\partial \theta_i} g_j(\hat{\theta}(\gamma))$ for $i = 1, \dots, m$ and $j = 1, \dots, k$. If g is one-dimensional then the argument to the second determinant in the denominator of (3) is a scalar. This approximation was stated for a one dimensional g , without proof, in Kass, Tierney and Kadane (1987a).

Approximation (3) requires that the gradient of g be non-zero, or the Jacobian be of full rank, over the support of the joint density p . Note that, as in (2), the normalizing constant c in (1) need not be evaluated since it cancels from the numerator and denominator in (3). Also note that (3) is equivariant under invertible transformations of g . That is, if ϕ is a differentiable, invertible transformation of \mathfrak{R}^k and $f(\theta) = \phi(g(\theta))$ then the approximate marginal densities for $f(\theta)$ and $g(\theta)$ satisfy $\hat{p}_f(t) = \hat{p}_g(\phi^{-1}(t)) |\det(D\phi(\phi^{-1}(t)))|^{-1}$. In general (3) is not invariant to transformations of the joint density (1).

In the next section we derive this approximation and show that it has *saddlepoint ac-*

curacy. In the third section we apply the approximation to the calculation of a marginal posterior density, a marginal sampling density, and a marginal density based on a multivariate saddlepoint approximation to a joint density. We conclude with a brief discussion of a related approximation of Hsu, Leonard and Tsui (1987) and some general remarks.

2 Derivation and Asymptotic Properties

The basic idea behind the derivation of (3) is this. Asymptotically the joint density p is dominated by its behavior near $\hat{\theta}$. The density of $g(\theta)$ can therefore be computed from p restricted to a neighborhood of $\hat{\theta}$. On a sufficiently small neighborhood there exists a local transformation $\phi = (\phi_1, \phi_2)$ such that $\phi = (g(\theta), \phi_2)$. Approximation (3) is then obtained by applying Laplace's method in this parameterization in a form that allows the Jacobians of the transformation to cancel. In this section we will treat the functions b and H in (1) as fixed. The approach of Kass, Tierney and Kadane (1987b) can be used to adapt the arguments given here to the case where b , H and $\hat{\theta}$ are replaced by well-behaved sequences.

To complete the derivation we need the following three lemmas. The first lemma is a simple determinant identity for partitioned matrices, the second gives a sufficient condition for the existence of a suitable local reparameterization, and the third states Laplace's method in the form we require.

Lemma 1 *Let A be a matrix partitioned as*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad (4)$$

and assume that A , A_{11} and A_{22} are nonsingular. Then the determinant of A_{22} satisfies the identity

$$\det(A_{22}) = \det(A) \det((A^{-1})_{11}), \quad (5)$$

where $(A^{-1})_{11}$ is the upper lefthand corner of A^{-1} .

This result follows from equations (5) and (6) of Morrison (1976). Note that if A_{11} is a one by one matrix then (5) follows from the adjoint form of the inverse of A . This lemma also shows that (2) is equivalent to (3) when g is a simple coordinate mapping.

Lemma 2 *Let g be a function from a set $\Theta \subseteq \mathbb{R}^m$ into \mathbb{R}^k and let $\hat{\theta}$ be an interior point of Θ . Assume that g is r -times continuously differentiable on an open neighborhood of $\hat{\theta}$, and that the derivative of g at $\hat{\theta}$ is of rank k . Then there exists an open neighborhood U of $\hat{\theta}$ and a one-to-one mapping $\psi : U \rightarrow \Gamma \subseteq \mathbb{R}^m$ such that $\psi = (\psi_1, \psi_2)$ is r -times continuously differentiable, of full rank (its Jacobian is nonsingular), and ψ_1 , the vector of the first k components of ψ , satisfies $\psi_1 = g(\theta)$ on U .*

Proof Without loss of generality we assume that the first k rows of $Dg(\hat{\theta})$, the Jacobian of g at $\hat{\theta}$, are linearly independent (if not we simply introduce an appropriate permutation). Let $\psi(\theta) = (g(\theta), \theta_2)$. Then the Jacobian of ψ is nonsingular at $\hat{\theta}$ and is continuous near $\hat{\theta}$. Thus there exists a neighborhood U of $\hat{\theta}$ on which ψ is of full rank. \square

Lemma 3 *Let C be a compact subset of \mathbb{R}^m and let \hat{x} be an interior point of C . Let f and q be real valued functions defined on C and six times continuously differentiable on the interior of C . Assume that \hat{x} uniquely minimizes q over C and let q'' denote the Hessian of q . Then*

$$\int_C f(x) e^{-nq(x)} dx = (2\pi)^{m/2} (\det(q''(\hat{x})))^{1/2} f(\hat{x}) e^{-nq(\hat{x})} \{1 + O(n^{-1})\}$$

as n tends to infinity. If the functions f and q and their derivatives depend in a continuously differentiable way on a parameter taking values in a compact set then the error term and its derivative or gradient with respect to the parameter are uniformly of order $O(n^{-1})$.

Versions of this result, known as Laplace's method, can be found in most texts on asymptotic analysis; see for example Chapter 4 of De Bruijn (1970).

We now return to the derivation of (3). The following regularity conditions are assumed to hold:

- (i) The functions b , h and g are defined on a subset Θ of \mathbb{R}^m . There exists a point $\hat{\theta}$ in the interior of Θ such that $\hat{\theta}$ is a strict local minimum of h , $b(\hat{\theta}) > 0$, and b , h and g are six times continuously differentiable on a neighborhood of $\hat{\theta}$.
- (ii) For every neighborhood U of $\hat{\theta}$ the probability assigned by the joint density p to the complement of U is of order $O(n^{-r})$ for every $r > 0$.

(iii) Let U be a neighborhood of $\hat{\theta}$, let $\hat{\gamma} = g(\hat{\theta})$, let P_n denote the probability measure on Θ corresponding to the density (1) and let $\mu(\cdot)$ denote Lebesgue measure on \mathfrak{R}^k . There exist an $n_0 > 0$, a $\delta > 0$ and a sequence of constants c_n with $c_n = O(n^{-r})$ for every $r > 0$ such that for every measurable $A \subseteq (\hat{\gamma} - \delta, \hat{\gamma} + \delta)$

$$P_n\{g(\theta) \in A, \theta \notin U\} \leq c_n \mu(A)$$

for all $n > n_0$.

Note that (iii) will hold if b is bounded, H has a unique global minimum at $\hat{\theta}$, and for some n the density of $g(\theta)$ under P_n is bounded in a neighborhood of $\hat{\gamma}$.

Theorem 1 *Let $p_g(\gamma)$ denote the marginal density of $g(\theta)$, and let $\hat{\gamma} = g(\hat{\theta})$. Under the regularity conditions above there exists a $\delta > 0$ such that for all γ in the open ball $\mathcal{B} = \{\gamma : \|\gamma - \hat{\gamma}\| < \delta\}$ the posterior density approximation (3) satisfies*

$$p_g(\gamma) = \hat{p}_g(\gamma) \left\{ 1 + \frac{e_n(\gamma)}{n} + O_\gamma(n^{-r}) \right\} \quad (6)$$

where the functions e_n are continuously differentiable on \mathcal{B} , the sequence $\{e_n\}$ and the sequence of derivatives or gradients $\{e'_n\}$ are uniformly bounded on \mathcal{B} , and $O_\gamma(n^{-r})$ is uniformly of order $O(n^{-r})$ on \mathcal{B} for every $r > 0$.

Proof Let U and $\psi = (\psi_1, \psi_2)$ be as in Lemma 2 and let $U(\gamma) = \{\theta \in U : g(\theta) = \gamma\}$. Let $V(\gamma) = \psi_2(U(\gamma))$ and let $J = J(\theta)$ be the Jacobian of ψ . Let $\delta > 0$ be small enough to satisfy assumption (iii), to insure that for any $\gamma \in \mathcal{B}$ there exists a strict minimizer $\hat{\theta}(\gamma)$ of H over the set $U(\gamma)$, and to insure that $\hat{\theta}(\gamma)$ is continuous on \mathcal{B} . Then for $\gamma \in \mathcal{B}$

$$p_g(\gamma) = \int_{V(\gamma)} \frac{p(\psi^{-1}(\gamma, \phi_2))}{\det(J(\psi^{-1}(\gamma, \phi_2)))} d\phi_2 \{1 + O_\gamma(n^{-r})\} \quad (7)$$

for every $r > 0$. By Lemma 3, with $f = b$ and $nq = nh = H$, we incur a relative error of order $O(n^{-1})$ if we replace the constant c in (1) by $(2\pi)^{-m/2} (\det \Sigma)^{-1/2} b(\hat{\theta})^{-1} \exp(H(\hat{\theta}))$, where Σ is the inverse of the Hessian of H (based on the θ parameterization). Thus

$$p_g(\gamma) = \int_{V(\gamma)} \frac{p(\psi^{-1}(\gamma, \phi_2))}{(2\pi)^{m/2} (\det \Sigma)^{1/2} p(\hat{\theta}) \det(J(\psi^{-1}(\gamma, \phi_2)))} d\phi_2 \{1 + O(n^{-1}) + O_\gamma(n^{-r})\} \quad (8)$$

with the $O(n^{-1})$ term independent of γ .

Let $\Sigma_{\phi 22}(\gamma)$ denote the inverse of the Hessian of the function $H(\psi^{-1}(\gamma, \cdot))$ evaluated at $\hat{\theta}(\gamma)$. Applying Lemma 3 once more, with $e^{-nq(\cdot)} = p(\psi^{-1}(\gamma, \cdot))$ and f equal to the remaining factor in (8), we have

$$p_g(\gamma) = \frac{(\det \Sigma_{\phi 22}(\gamma))^{1/2}}{(2\pi)^{k/2}(\det \Sigma)^{1/2} \det(J(\hat{\theta}(\gamma)))} \frac{p(\hat{\theta}(\gamma))}{p(\hat{\theta})} \left\{1 + \frac{e_n(\gamma)}{n} + O_\gamma(n^{-r})\right\} \quad (9)$$

where the e_n are continuously differentiable on \mathcal{B} and $\{e_n\}$ and $\{e'_n\}$ are uniformly bounded on \mathcal{B} . Let $\Sigma_\phi(\gamma)$ denote the inverse of the Hessian of the function $H(\psi^{-1}(\cdot))$ evaluated at $\hat{\theta}(\gamma)$. Then $\det(\Sigma_\phi(\gamma)) = \det(\Sigma(\gamma))(\det(J(\hat{\theta}(\gamma))))^2$, and thus, by Lemma 1,

$$\frac{\det(\Sigma_{\phi 22}(\gamma))}{(\det(J(\hat{\theta}(\gamma))))^2} = \frac{\det(\Sigma(\gamma))}{\det((Dg)^T \Sigma(\gamma) (Dg))},$$

which completes the proof. \square

To capture the relative error of approximation (3) on neighborhoods of $\hat{\gamma}$ that shrink in diameter at the rate $O(n^{-1/2})$ we compare the ratio of the actual marginal density at $\hat{\gamma}$ and points near $\hat{\gamma}$ to the corresponding ratio for the approximations.

Corollary *Fix $u \in \mathfrak{R}^k$ and set $\gamma_n = \hat{\gamma} + n^{-1/2}u$. Then under the assumptions of the preceding theorem*

$$\frac{p_g(\gamma_n)}{p_g(\hat{\gamma})} = \frac{\hat{p}_g(\gamma_n)}{\hat{p}_g(\hat{\gamma})} \{1 + O_u(n^{-3/2})\},$$

with $O_u(n^{-3/2})$ uniformly of order $O(n^{-3/2})$ for bounded values of u .

This result follows immediately from the form of the error given in the preceding theorem.

3 Applications

In the first two subsections of this section we use (3) to approximate a marginal posterior density and a marginal sampling density. We have chosen simple, two dimensional problems in order to be able to verify the approximations using numerical integration and an exact expression for the marginal density, respectively. In the final subsection we use (3) to derive an approximation for a marginal density based on the saddlepoint approximation for the joint density of the sufficient statistic in a multidimensional exponential family given in Barndorff-Nielsen and Cox (1979).

3.1 A Marginal Posterior Density

Feigl and Zelen (1965) examine the relationship between Leukemia survival and a concomitant variable, the white blood cell count of a patient. They consider a model in which survival times are assumed to be exponentially distributed with a mean survival time of the form $\theta_1 \exp\{\theta_2 x\}$, where x is the natural logarithm of the white blood cell count measured in units of 10,000. Thus θ_1 represents the mean survival time for a patient with a white blood cell count of 10,000 and θ_2 represents the approximate percentage change in mean survival time corresponding to a one percent increase in the white blood cell count. Their sample consisted of patients classified as AG positive or AG negative based on examination of the leukemic cells.

We consider the sample of 17 AG positive patients and are particularly interested in the two year survival probability of patients at a white blood cell count of 50,000, a value slightly above the third quartile of the white blood cell counts for this sample of patients. We computed the approximate marginal posterior density (3) for this probability based on an improper uniform prior distribution on θ_1 and θ_2 . For this computation we chose b in equation (1) to be a constant and H equal to minus the log likelihood. The exact and approximate marginal densities for this survival probability are shown in Figure 1. Figure 2 shows the same densities transformed to the logit scale. The exact marginal density was calculated by reparameterizing the joint density in terms of the survival probability and θ_2 and numerically integrating over θ_2 using a 200 point trapezoidal rule on the interval $[-2, 0]$. This interval contains essentially all the support of the posterior distribution on θ_2 ; use of a larger interval or more points did not change the graphs of the marginal density of the survival probability.

3.2 A Marginal Sampling Distribution

Suppose we would like to estimate the proportion of a normal population that falls below a point x_0 . If the mean and variance of the population are unknown then a reasonable estimator might be

$$\hat{P} = \Phi \left(\frac{x_0 - \bar{X}}{S} \right) \quad (10)$$

where \bar{X} is the sample mean, S the sample standard deviation, and Φ the cumulative distribution function of the standard normal distribution.

The marginal density of this estimator can be approximated using (3). Taking b to be a constant and H to be minus the logarithm of the joint density of \bar{X} and S the approximate marginal density at p is given by

$$\left(\frac{n(n-1)^{n-1}}{\pi(n-2)^{n-2}(n-2(n-1)s^2\Phi^{-1}(p)^2)} \right)^{1/2} \frac{s^{n-1}}{\varphi(\Phi^{-1}(p))} \times \exp\left\{-\frac{1}{2}(n(x_0 - s\Phi^{-1}(p))^2 + n - 2 - (n-1)s^2)\right\} \quad (11)$$

where

$$s = \frac{n\Phi^{-1}(p)x_0 + \sqrt{n^2\Phi^{-1}(p)^2x_0^2 + 4(n-2)(n-1+n\Phi^{-1}(p)^2)}}{2(n-1+n\Phi^{-1}(p)^2)} \quad (12)$$

and φ is the standard normal density.

The exact density of \hat{P} can be computed by noting that \hat{P} is a function of a noncentral t random variable with $n-1$ degrees of freedom and noncentrality parameter $\sqrt{n}x_0$. The exact and approximate densities for $x_0 = 1$ and $n = 5$ are shown in Figure 3. Figure 3 also shows the approximate densities based on the joint density of \bar{X} and $\log(S)$ and the joint density of \bar{X} and S^2 . Figure 4 shows the same densities transformed to the probit scale.

All three approximations appear to perform reasonably. The choice of parameterization does have an effect; in this case the parameterization based on S appears to perform better than the parameterization based on $\log(S)$, which is in turn superior to the parameterization based on S^2 . Note that the graphs for the approximation based on S are almost indistinguishable from the graphs of the exact density.

3.3 Marginalizing a Saddlepoint Approximation

As pointed out by Reid (1988), interest in saddlepoint methods has increased considerably in recent years, in particular since the publication of Barndorff-Nielsen and Cox (1979). Approximation (3) can be used to approximate marginal densities from saddlepoint approximations to joint densities.

Suppose X_1, \dots, X_n are independent and identically distributed random variables from an m -dimensional exponential family with density $\exp\{-x^T\phi\}f(x)/M(\phi)$. Let $T_n = \frac{1}{n}\sum X_i$; denote the sufficient statistic for the family, let $K(\phi) = \log(M(\phi))$, and let $K'(\phi)$ and $K''(\phi)$

denote the gradient and Hessian of K , respectively. Let $\hat{\phi}$ denote the maximum likelihood estimator of ϕ , which is the solution to the equation $T_n = -K'(\phi)$. Then the saddlepoint approximation to the joint density of T_n at t is

$$(2\pi n)^{-m/2} \frac{\exp\{nK(\hat{\phi}) - nK(\phi) + nt^T(\hat{\phi} - \phi)\}}{(\det(K''(\hat{\phi})))^{1/2}} \quad (13)$$

Here $\hat{\phi}$ represents the maximum likelihood value corresponding to an observed value of t for T_n . The marginal density of a function $g(T_n)$ of the sufficient statistic can be approximated by applying (3) to (13). In this case it is convenient to choose $b = (\det(K''(\hat{\phi})))^{-1}$. Then $H = -n(K(\hat{\phi}) - K(\phi) + t^T(\hat{\phi} - \phi))$. Let $\hat{t}(\gamma)$ minimize H subject to the constraint $g(t) = \gamma$ and let $\hat{\phi}(\gamma)$ be the corresponding maximum likelihood point. Then $\Sigma(\gamma) = K''(\hat{\phi}(\gamma))$ and the approximate marginal density of $g(T_n)$ at γ is proportional to

$$(2\pi n)^{-k/2} \frac{\exp\{nK(\hat{\phi}(\gamma)) - nK(\phi) + n\hat{t}(\gamma)^T(\hat{\phi}(\gamma) - \phi)\}}{(\det((Dg)^T K''(\hat{\phi}(\gamma))(Dg)))^{1/2}} \quad (14)$$

4 A Related Approximation

A related approximation to the marginal posterior density of a nonlinear function is described in Hsu, Leonard and Tsui (1987). To facilitate comparison with their approximation, we write (3) without the normalizing constant as

$$\hat{p}_g(\gamma) \propto \left(\frac{\det \Sigma(\gamma)}{\det((Dg)^T \Sigma(\gamma)(Dg))} \right)^{1/2} p(\hat{\theta}(\gamma)) \quad (15)$$

Let ℓ_γ denote the gradient of H at $\hat{\theta}(\gamma)$, let $\tilde{\theta}(\gamma) = \hat{\theta}(\gamma) - \Sigma(\gamma)\ell_\gamma$, and let $f(\gamma|\mu, C)$ denote the density of $g(\theta)$ at γ when θ has a normal distribution with mean vector μ and covariance matrix C . By expanding the logarithm of the joint density p about the point $\hat{\theta}(\gamma)$ up to second order, Hsu, Leonard and Tsui derive the intermediate approximation

$$\tilde{p}_g(\gamma) \propto (\det \Sigma(\gamma))^{1/2} p(\hat{\theta}(\gamma)) \exp\{\ell_\gamma^T \Sigma \ell_\gamma\} f(\gamma|\tilde{\theta}(\gamma), \Sigma(\gamma)) \quad (16)$$

This approximation is an intermediate step to a final approximation since the density f is generally not available in closed form and needs to be approximated.

The derivation of (16) does not assume that g is well behaved. In particular g need not be differentiable on a neighborhood of $\hat{\theta}$, though the asymptotic accuracy of the approximation

will depend on the degree of smoothness of g . As long as f can be computed, (16) can thus be used to approximate the marginal density of quite general functions g . Since samples from normal distributions are easy to generate, a general strategy for approximating f might be to simulate observations from the density f . Hsu, Leonard and Tsui are particularly concerned with certain quadratic g that are used as measures of association in contingency tables. For that application they propose a particular gamma distribution as an approximation to f .

If the function g is differentiable and its Jacobian is of full rank near $\hat{\theta}$ then the error in using the intermediate approximation (16) will be of order $O(n^{-1})$ on a fixed neighborhood for $\hat{\theta}$. In this case one can then use (15) to approximate f for the special case of a joint normal distribution of θ . This will reproduce approximation (15). Hsu, Leonard and Tsui suggest the possibility of approximating $f(\gamma|\tilde{\theta}(\gamma), \Sigma(\gamma))$ by the density of a normal distribution with mean $g(\tilde{\theta}(\gamma))$ and variance $(Dg)(\tilde{\theta}(\gamma))^T \Sigma(\gamma) (Dg)(\tilde{\theta}(\gamma))$ evaluated at γ . If g is not linear this would generally result in an approximation with a relative error of order $O(n^{-1/2})$ on neighborhoods of $\hat{\theta}$ that shrink in diameter at the rate $O(n^{-1/2})$.

If g is not smooth or if its Jacobian is not of full rank near $\hat{\theta}$ then using (16) with a specialized approximation to f may produce better results than (15). In some of the examples considered by Hsu, Leonard and Tsui an independence model is close to the support of the joint posterior distribution p . The gradient of the quadratic measure of association they consider vanishes at an independence model, and as a result in these cases approximation (15) does not perform as well as their approximation, based on (16) with a gamma approximation to f .

5 Discussion

The approximations described in this paper are most useful for numerically evaluating and plotting approximate marginal densities. Though there are some exceptions, they do not in general produce simple closed form expressions that can be manipulated analytically.

The computational requirements for these approximations are rather modest. If routines for determining the minimum and the Hessian of the function H are available then they can easily be adapted for computing approximation (3). For example, if a one dimensional approximate density is to be computed on a grid of points then one can start at $\hat{\gamma}$ and work

outward along the grid and taking one or two Newton steps for the gradient and constraint equations, using the $\hat{\theta}(\gamma)$ value for the each grid point as the starting value for the next grid point. A set of programs that can be used for these computations from within the S system are described in Tierney, Kass and Kadane (1987).

A weakness of these approximations is that they generally do not integrate to one. Numerical integration has to be used to renormalize the approximations. This is a nuisance for higher dimensional margins but is not a serious problem for one and two dimensional margins. The approximations do integrate approximately to one, and if they are to be used primarily to obtain graphical representations of the marginal densities then the constant of integration is less important as it will affect only the scale on the density axis, not the shapes of the approximate densities. The implication of the corollary in Section 2 above is that the shape of the marginal density is approximated more accurately than the constant of integration.

Several practical issues require further investigation. The parameterization used in the joint density clearly affects the quality of the approximation. A parameterization in which the joint density is closer to normal will usually produce a better approximation. At present only heuristic guidelines on the choice of a parameterization are available; it would be useful to place these on a firmer footing. Another issue requiring further investigation is how close a zero of the gradient or a point of degeneracy of the Jacobian of g can be to the center of the joint distribution before the approximation breaks down. The computation of probabilities also needs to be explored further. Numerical integration of the approximate density appears to be the only alternative available at present. It may be possible to adapt some of the techniques for the evaluation of tail areas from saddlepoint approximations to the present setting.

As a final remark we would like to point out the relation between (3) and modified or conditional profile likelihoods studied, for example, in Cox and Reid (1987). The formula for the conditional profile likelihood given by Cox and Reid is essentially the same as (2) if the joint density p is replaced by the likelihood. It may be possible to give (3) a similar interpretation for problems where there are many parameters, a function of the parameters is of interest, but an explicit decomposition of the parameter vector into the quantity of interest and a nuisance parameter vector is either difficult or impossible.

Acknowledgements

We would like to thank Tom Leonard for bringing the results described in Hsu, Leonard and Tsui (1987) to our attention. We would also like to thank Kathryn Chaloner for helpful comments on earlier drafts of this manuscript.

References

- BARNDORFF-NIELSEN, O. AND COX, D. R. (1979), Edgeworth and saddlepoint approximations with statistical applications (with discussion). *J. R. Statist. Soc. B*, 41, 279-312.
- COX, D. R., AND REID, N. (1987), Parameter orthogonality and approximate conditional inference (with discussion). *J. R. Statist. Soc. B*, 81, 1-39.
- DE BRUIJN, N. G. (1970), *Asymptotic Methods in Analysis*, North-Holland, Amsterdam.
- FEIGL, P. AND ZELEN, M. (1965), Estimation of exponential survival probabilities with concomitant information, *Biometrics*, 21, 826-837.
- HSU, J. S. J., LEONARD, T. AND TSUI, K. (1987), Bayesian and likelihood marginal inference, with applications to contingency table analysis, Technical Report, Department of Statistics, University of Wisconsin.
- KASS, R. E., TIERNEY, L. AND KADANE, J. B. (1987a), Asymptotics in Bayesian computation, to appear in *Bayesian Statistics 3*, J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith, eds.
- KASS, R. E., TIERNEY, L. AND KADANE, J. B. (1987b), The validity of posterior expansions based on Laplace's method. Technical Report, Department of Statistics, Carnegie-Mellon University.
- MORRISON, D. F. (1976), *Multivariate Statistical Methods*, 2nd edition, New York: McGraw-Hill.
- PHILLIPS, P. C. B. (1983), Marginal densities of instrumental variables estimators in the general single equation case. Cowles Foundation Paper No. 582.
- REID, N. (1988), Saddlepoint methods and statistical inference. *Statistical Science*, to appear.
- SWEETING, T. (1987), discussion to Cox and Reid (1987).
- TIERNEY, L. AND KADANE, J. B. (1986), Accurate approximations for posterior moments and marginal densities, *JASA*, 81, 82-86.
- TIERNEY, L., KASS, R. E. AND KADANE, J. B. (1986), Approximation of expectations and variances using Laplace's Method, Technical Report, Department of Statistics, Carnegie-Mellon University.
- TIERNEY, L., KASS, R. E. AND KADANE, J. B. (1987), Interactive Bayesian analysis using accurate asymptotic approximations, *Proc. of 19th Symp. on the Interface*, 15-21, R. M. Heiberger, ed., ASA, Alexandria, Va.

Figure 1: Posterior Density of Survival Probability, Probability Scale

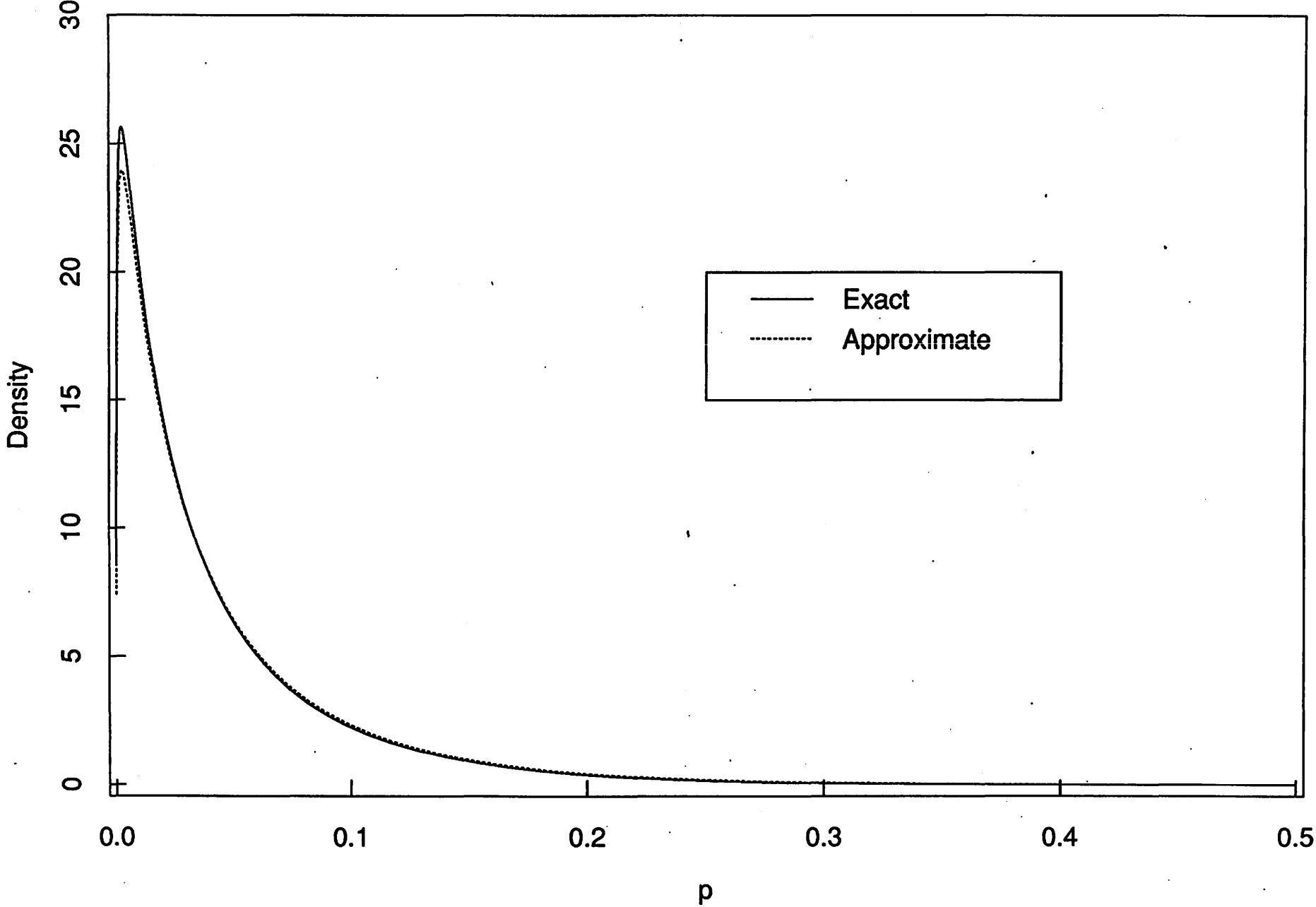


Figure 2: Posterior Density of Survival Probability, Logit Scale

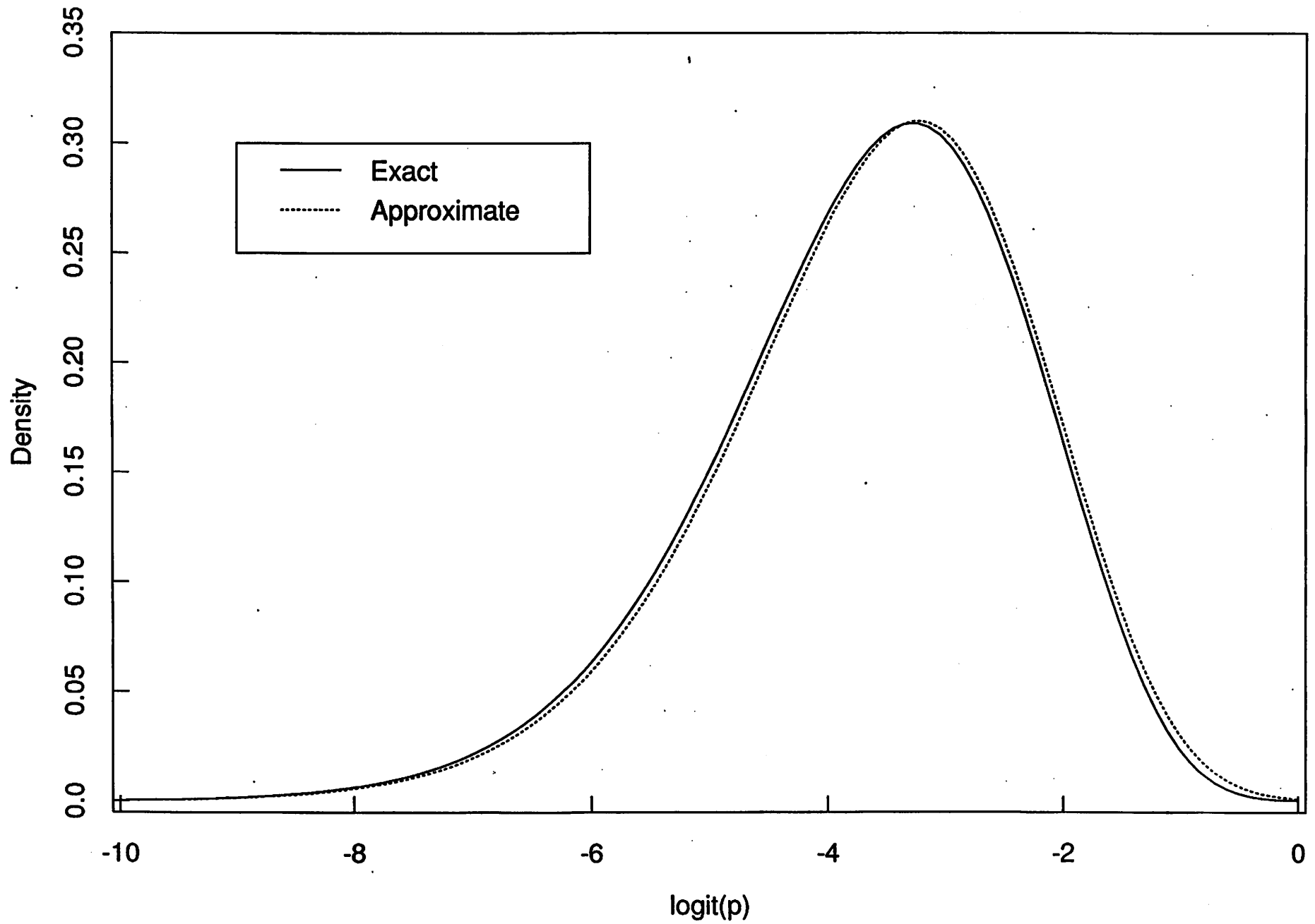


Figure 3: Density of Normal Proportion Estimator, Probability Scale

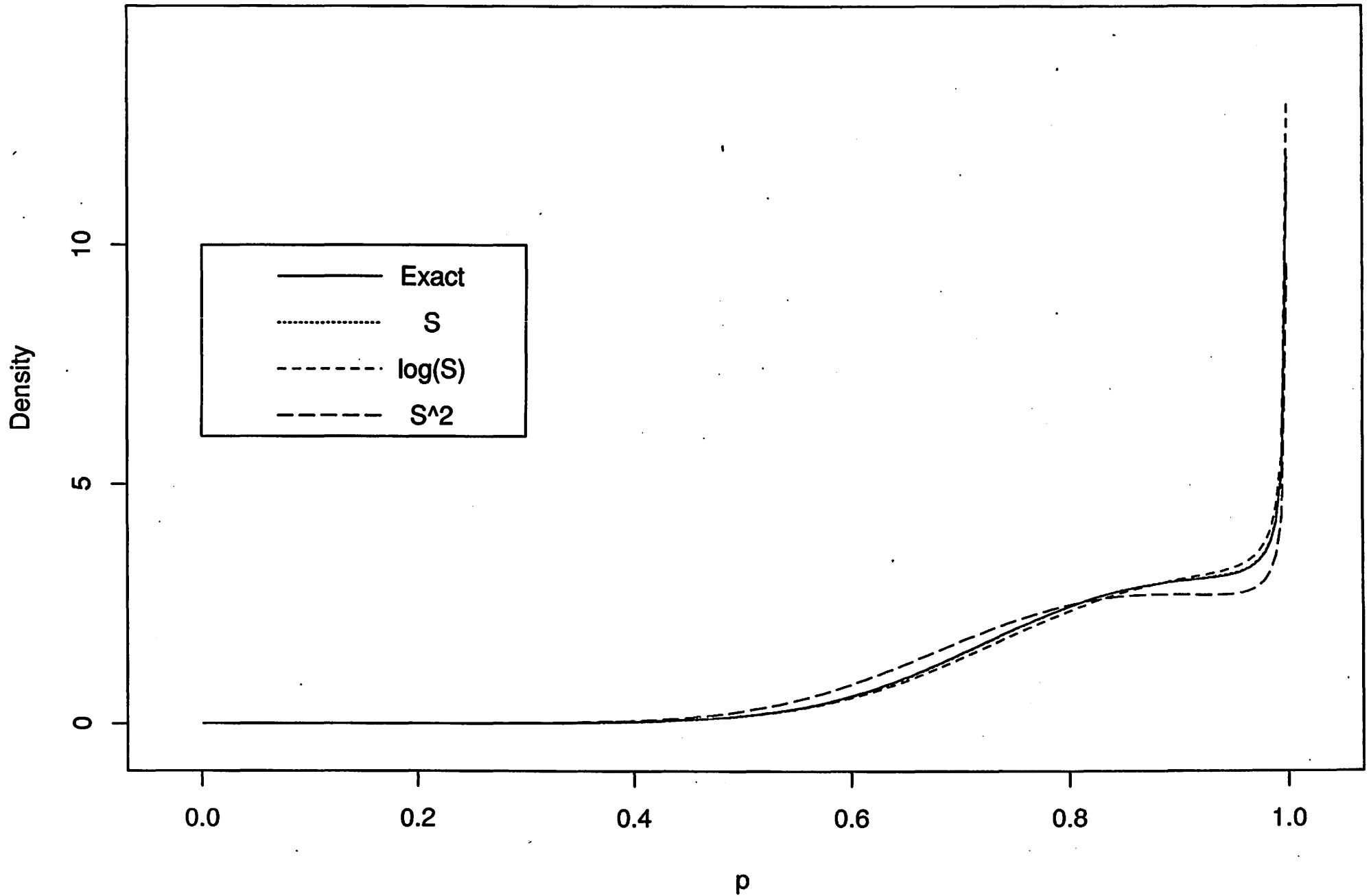


Figure 4: Density of Normal Proportion Estimator, Probit Scale

