Biometrika (1994), **81**, 3, *pp.* 624–9
Printed in Great Britain

A note on Gauss–Hermite quadrature

BY QING LIU
Department of Biostatistics, St. Jude Children's Research Hospital, 332 North Lauderdale, P.O. Box 318, Memphis, Tennessee 38101, U.S.A.

AND DONALD A. PIERCE
Department of Statistics, Oregon State University, Corvallis, Oregon 97331, U.S.A.

SUMMARY

For Gauss–Hermite quadrature, we consider a systematic method for transforming the variable of integration so that the integrand is sampled in an appropriate region. The effectiveness of the quadrature then depends on the ratio of the integrand to some Gaussian density being a smooth function, well approximated by a low-order polynomial. It is pointed out that, in this approach, order one Gauss–Hermite quadrature becomes the Laplace approximation. Thus the quadrature as implemented here can be thought of as a higher-order Laplace approximation.

Some key words: Asymptotic approximation; Bayesian inference; Generalized linear mixed models; Integrated likelihood; Measurement errors in covariables; Numerical integration.

1. INTRODUCTION

Gauss–Hermite quadrature is often used for numerical integration in statistics, because of its relation to Gaussian densities, but it seems that there is often inadequate thought given to its implementation. Such quadrature is defined in terms of integrals of the form

$$\int_{-\infty}^{\infty} f(x) \exp(-x^2)\, dx. \tag{1}$$

In many statistical applications a Gaussian density is an explicit factor of the integrand. Of course a linear transformation can then be made so that this factor takes the form $\exp(-x^2)$, and this seems often to be the approach taken (SERC, 1989; Crouch & Spiegelman, 1990). When a Gaussian density is not a factor in the integrand the integral is sometimes put into the form (1) by dividing and multiplying the original integrand by $\exp(-x^2)$ (Davies & Rabinowitz, 1975, Ch. 2), or by some other Gaussian density.

In Gauss–Hermite quadrature the integral (1) is approximated by

$$\int_{-\infty}^{\infty} f(x) \exp(-x^2)\, dx \doteq \sum_{i=1}^{m} w_i f(x_i), \tag{2}$$

where the nodes x_i are zeros of the mth order Hermite polynomial and the w_i are suitably corresponding weights; see for example, Davis & Rabinowitz (1975, Ch. 2). Tables of (x_i, w_i) for $m = 1, 2, \ldots, 10, 12, 16, 20$ are given by Abramowitz & Stegun (1972, p. 924). The x_i are symmetric about zero, ranging when $m = 20$, for example, from $-5{\cdot}3875$ to $5{\cdot}3875$. Thus, without further considerations, the function $f(x)$ is sampled at points irrespective of the range where it is 'interesting'. As indicated above, the $f(x)$ used here is usually not the original integrand.

Clearly, for good results some transformation must ordinarily be made, in order that the original integrand be sampled in a reasonable range. The possible transformation mentioned above, simply to put an integrand with a Gaussian density factor into form (1), is not at all directed towards

this aim. This is indeed confusing, and thus we consider here a systematic way to apply Gauss–Hermite quadrature to integrals of the form

$$\int_{-\infty}^{\infty} g(t)\,dt, \tag{3}$$

where $g(t) > 0$. One needs to consider at the outset the class of functions $g(t)$ for which the Gauss–Hermite quadrature is suitable. As we shall see, the requirement for effective results is that the ratio of $g(t)$ to some Gaussian curve be a moderately smooth function. This arises frequently, for example when $g(t)$ is a likelihood function, the product of a likelihood function and a Gaussian density, and the product of several likelihood functions, etc. We assume in the following that $g(t)$ has such characteristics, and in particular, that it is unimodal.

2. Gauss–Hermite quadrature

The goal is a transformation on t so that the integrand $g(t)$ will be sampled in a suitable range. One can think of this in various ways, but we find the following the clearest. First, one can trivially re-express Gauss–Hermite quadrature (1) as for integrals of form

$$\int_{-\infty}^{\infty} f(t)\phi(t;\mu,\sigma)\,dt, \tag{4}$$

where $\phi(t;\mu,\sigma)$ is an arbitrary Gaussian density (Naylor & Smith, 1982). The sampling nodes are then at $t_i = \mu + 2^{\frac{1}{2}}\sigma x_i$, and the weights are modified to $w_i/\sqrt{\pi}$. We will then choose μ and σ so that $g(t)$ will be sampled in an appropriate region. In particular, we take $\hat{\mu}$ to be the mode of $g(t)$, and $\hat{\sigma} = 1/\sqrt{\hat{j}}$, where

$$\hat{j} = -\frac{\partial^2}{\partial t^2}\log g(t)\bigg|_{t=\hat{\mu}}.$$

This gives a Gaussian density $\phi(t;\hat{\mu},\hat{\sigma})$ having the same logarithmic derivatives to second order, at the mode, as the integrand $g(t)$. Define

$$h(t) = \frac{g(t)}{\phi(t;\hat{\mu},\hat{\sigma})} \tag{5}$$

so that one can write

$$\int_{-\infty}^{\infty} g(t)\,dt = \int_{-\infty}^{\infty} h(t)\phi(t;\hat{\mu},\hat{\sigma})\,dt. \tag{6}$$

If we now apply Gauss–Hermite quadrature in the form (4) using $\phi(t;\hat{\mu},\hat{\sigma})$, the function $h(t)$, and hence $g(t)$, will be sampled in the relevant range, giving

$$\begin{aligned}\int_{-\infty}^{\infty} g(t)\,dt &\simeq \sum_{i=1}^{m} \frac{w_i}{\sqrt{\pi}} h(\hat{\mu} + 2^{\frac{1}{2}}\hat{\sigma}x_i) \\ &= 2^{\frac{1}{2}}\hat{\sigma}\sum_{i=1}^{m} w_i^* g(\hat{\mu} + 2^{\frac{1}{2}}\hat{\sigma}x_i),\end{aligned} \tag{7}$$

where $w_i^* = w_i \exp(x_i^2)$.

To consider the effectiveness of the Gauss–Hermite quadrature as implemented here, suppose that $h(t)$ in (5) has an expansion of the form

$$h(t) = h(\hat{\mu})\left\{1 + \sum_{k=1}^{\infty} c_k(t-\hat{\mu})^k\right\}. \tag{8}$$

The coefficients c_1 and c_2 are zero due to the choice of $\phi(t; \hat{\mu}, \hat{\sigma})$. In the integral (6) the odd terms involving $c_3, c_5, \ldots$ are annihilated. Each increase by one in the order of the quadrature picks up exactly the contribution to the integral of one additional even term. That is the m-order Gauss–Hermite quadrature would be exact if terms in (8) beyond $c_{2(m+1)}$ were zero.

It should be noted, however, that these statements would also be true if $\phi(t; \hat{\mu}, \hat{\sigma})$ in (5) and (6) were replaced by $\exp(-t^2)$, that is if the quadrature were done in the naive way without the transformation. But then for low-order quadrature to be effective, $h(t)$ would have to be well approximated by a low-order polynomial in a rather global sense, that is over a typically wide region including both the range of the standard sampling points and the relevant range regarding the integrand $g(t)$. In the approach taken above, it is only necessary that $h(t)$, as defined by (5), be well approximated by a low-order polynomial in the relevant region for $g(t)$, since this is also where the sampling nodes are taken.

Thus, simply put, the m-order Gauss–Hermite quadrature as implemented by (7) will be highly effective if the ratio of $g(t)$ to the Gaussian density $\phi(t; \hat{\mu}, \hat{\sigma})$ can be approximated well by a polynomial of order $2m+1$, in the region where $g(t)$ is substantial. Some other reasons for the effectiveness of (7) in many applications will be given in § 4 and the Appendix.

3. RELATION TO LAPLACE APPROXIMATION

When (7) is applied with only one node the result is

$$\int_{-\infty}^{\infty} g(t)\,dt \doteq h(\hat{\mu}) = (2\pi)^{\frac{1}{2}}\hat{\sigma}g(\hat{\mu}), \tag{9}$$

which is the Laplace approximation to the integral (De Bruijn, 1961, p. 60; Barndorff-Nielsen & Cox, 1989, p. 59). Thus, the m-order Gauss–Hermite quadrature as implemented in (7) can be thought of alternatively as the form of 'm-order Laplace approximation'. Indeed, computation of a few low-order quadratures in this way may often be preferable, in applied work, to use of the standard asymptotic error term for the Laplace approximation. As discussed above for the general case, the Laplace approximation would be exact if the even coefficients $c_4, c_6, \ldots$ in the expansion (8) were zero. It often performs even better than that might suggest, perhaps because in approximating, rather than expanding, $h(t)$ in form (8) there is substantial latitude in choosing the coefficients c_i.

The asymptotic accuracy of Laplace approximations for Bayesian inference was studied by Tierney & Kadane (1986). Wong & Li (1992) considered improvements to Laplace approximations for statistical applications by incorporating a higher order correction term. Their method involves calculations of third and fourth derivatives of the integrand $g(t)$. The Gauss–Hermite quadrature (7) involves derivatives of $g(t)$ only to second order, replacing in essence the use of higher derivatives by sampling of the function $g(t)$.

4. RATIOS OF INTEGRALS

Statistical applications often involve a parametric family of integrals

$$I(\beta) = \int_{-\infty}^{\infty} g(t; \beta)\,dt,$$

where the function $I(\beta)$ need only be approximated up to a constant of proportionality. In such cases all that matters about the accuracy of quadrature or Laplace approximation is that the relative error varies slowly with β.

For example, in generalized linear models with random effects and in models with random errors in covariables, the likelihood function is often defined by

$$L(\beta) = \int_{-\infty}^{\infty} L(t; \beta)p(t)\,dt, \tag{10}$$

where $L(t; \beta)$ is a likelihood function in either of its arguments, and $p(t)$ is a probability density function, often taken as Gaussian. For the case of generalized linear models with random effects, Liu & Pierce (1993) found that the Laplace approximation is often virtually exact.

This same issue also arises in Bayesian inference when the integral of interest defines a functional of the posterior density. For example, if $g(t)$ is only proportional to a posterior density, then the posterior mean for a parametric function $a(t)$ would be computed numerically as

$$\int_{-\infty}^{\infty} a(t)g(t)\,dt \Big/ \int_{-\infty}^{\infty} g(t)\,dt.$$

Tierney & Kadane (1986) investigated the improved error rate for the ratio of Laplace approximations to two integrals, when $a(t)$ is positive.

5. Examples

There are many practical settings in the literature that require very few nodes for (7) to work extremely well. For example, in problems considered by Liu & Pierce (1993) the Laplace approximation is extremely good. From the viewpoint of this paper, this means that the 'rescaling' of t works so well that only the order 1 Gauss–Hermite quadrature is needed.

Example 1. We consider the logistic regression model for binary data where the explanatory variable is measured with error. The data consist of n independent pairs (Y_i, Z_i) where Y_i is a binary response variable, and Z_i is a measurement of the unobserved explanatory variable x_i. Suppose that (i) $\text{pr}(Y_i = 1 \mid x_i) = \text{logit}^{-1}(\beta_0 + \beta_1 x_i)$, (ii) the distribution of Z_i given x_i is Gaussian, and (iii) the marginal distribution of x_i is Gaussian. Assuming the parameters involved in (ii) and (iii) are known, the likelihood $L(\beta_0, \beta_1)$ is given by the product of integrals of form

$$\int_{-\infty}^{\infty} L(\beta_0, \beta_1; y_i \mid x) p(x \mid z_i)\,dx.$$

Here

$$\log L(\beta_0, \beta_1; y \mid x) = y(\beta_0 + \beta_1 x) - \log(1 + e^{\beta_0 + \beta_1 x})$$

is the log likelihood function when x is known, and $p(x \mid z)$, the conditional density for X given $Z = z$, is a Gaussian density. For illustration, we generated a data set of n binary observations with parameters $\beta_0 = -5$, $\beta_1 = 0{\cdot}1$, $E(X) = 60$, $\text{var}(X) = 100$ and $\text{var}(Z \mid X = x) = 50$. The interest here

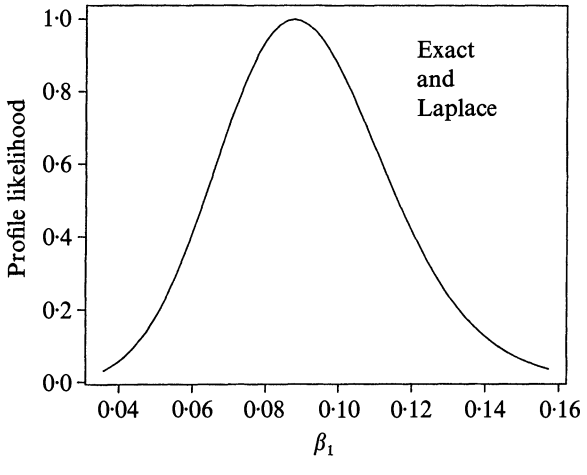

Fig. 1. Profile likelihoods for β_1 in Example 1.

is in computing the profile likelihood function for β_1, and we compare the Laplace approximation to use of the Gauss–Hermite quadrature (7). The sample size n was taken as 300. Note that the quadratures are applied for each binary observation. The accuracy of the Laplace approximation seen here does not decrease for small n, for example $n = 10$; on the contrary adequacy of dealing with the measurement error becomes more critical as n increases, which is the reason for choosing $n = 300$. The profile likelihood for β_1 is calculated by evaluating $L(\beta_0, \beta_1)$ on a fine grid for (β_0, β_1) and then maximizing with respect to β_0 for each fixed β_1. In Fig. 1 we see that the Laplace approximation to the profile likelihood function is very accurate, indistinguishable from the Gauss–Hermite quadrature (7) with 9 nodes.

Example 2. This example provides a setting where the Laplace approximation does not work so well and therefore the Gauss–Hermite quadrature is needed. It differs from Example 1 in that X as well as $Z \mid X = x$ are log-normal. To apply the Gauss–Hermite quadrature, define the likelihood function as the product of integrals with respect to $u = \log(x)$. The underlying integrals are of the form

$$\int_{-\infty}^{\infty} L(\beta; y \mid u) p(u \mid z)\, du,$$

where

$$L(\beta; y \mid u) = y(\beta_0 + \beta_1 e^u) - \log\{1 + \exp(\beta_0 + \beta_1 e^u)\},$$

and $p(u \mid z)$ is a Gaussian density. The main difference from the previous example is that $x = e^u$ is used in the likelihood function in the integrand. A data set was simulated by considering parameter values similar to those of Carroll, Gail & Lubin (1993), and again $n = 300$. We chose $\beta_0 = -3$, $\beta_1 = 0{\cdot}5$; $\log(Z) \mid X = x$ as being Gaussian with mean $\log(x)$ and variance 0·25; and $\log(X)$ as being Gaussian with mean $-0{\cdot}5$ and variance 1. Figure 2 demonstrates the convergence to the likelihood for β_1 as the number of nodes increases from 1 to 10, 12, 16 and 20. The solid curve is the Laplace approximation. In this example the parameter β_0 is taken as fixed at the maximum likelihood estimator, in order not to confuse inadequacies of the quadrature with computation of the profile likelihood.

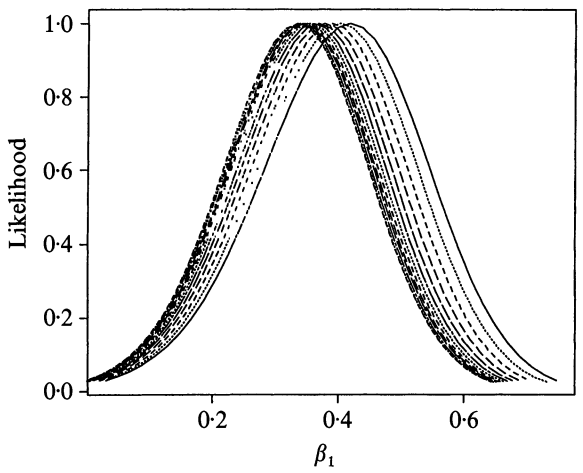

Fig. 2. Likelihoods for β_1 in Example 2.

ACKNOWLEDGEMENT

Q. Liu's research was supported by the National Cancer Institute. D. A. Pierce's work was supported by the National Science Foundation.

Appendix

Asymptotic behaviour of Gauss–Hermite quadrature

The standard asymptotic analysis of the Laplace approximation considers integrals of form

$$\int_{-\infty}^{\infty} \exp\{nl(t)\}\, dt \tag{11}$$

as $n \to \infty$, where $l(t)$ is a unimodal function. (De Bruijn, 1961, p. 63; Barndorff-Nielsen & Cox, 1989, p. 59). Here we consider the asymptotic behaviour of the Gauss–Hermite quadrature (7) for integral (11). The first $2m+1$ terms in (8) are picked up exactly by m-order Gauss–Hermite quadrature (7) and thus the error is of the same order as the integral of the term involving $c_{2(m+1)}$. Let $\hat{\mu}$ be the mode of $l(t)$ and $\ddot{l}(\hat{\mu}) = (d/dt)^2 l(\hat{\mu})$. Define $\pi(t)$ such that $h(t) = h(\hat{\mu})\pi(t)$. Then

$$\pi(t) = \exp\left(nl(t) - nl(\hat{\mu}) + \frac{n}{2}\left[\frac{t-\hat{\mu}}{\sqrt{\{-\ddot{l}(\hat{\mu})\}}} \right]^2 \right). \tag{12}$$

The coefficient $c_{2(m+1)}$ in (8) is

$$c_{2(m+1)} = \frac{1}{\{2(m+1)\}!} \left\{ \frac{d^{2(m+1)}}{dt^{2(m+1)}} \pi(t) \Big|_{t=\hat{\mu}} \right\}.$$

One can show that

$$\frac{d^{2(m+1)}}{dt^{2(m+1)}} \pi(t) \Big|_{t=\hat{\mu}} = O(n^{[2(m+1)/3]})$$

where notation $[r]$ indicates the largest integer not exceeding r; the proof for this can be obtained from the authors. The integral involving $c_{2(m+1)}$ is proportional to

$$\{-n\ddot{l}(\hat{\mu})\}^{-(m+1)} c_{2(m+1)},$$

and thus the error of the Gauss–Hermite quadrature (7) is of order $O(n^{-[m/3+1]})$. For $m = 1$, we obtain the standard result that the error of the Laplace approximation is of order $O(n^{-1})$ (Tierney & Kadane, 1986; Barndorff-Nielsen & Cox, 1989, p. 59).

References

Abramowitz, M. & Stegun, I. (Eds) (1972). *Handbook of Mathematical Functions.* New York: Dover.

Barndorff-Nielsen, O. E. & Cox, D. R. (1989). *Asymptotic Techniques for Use in Statistics.* London: Chapman and Hall.

Carroll, R. J., Gail, M. H. & Lubin, J. H. (1993). Case-control studies with errors in covariates. *J. Am. Statist. Assoc.* **88**, 185–99.

Crouch, E. A. C. & Spiegelman, D. (1990). The evaluation of integrals of the form $\int f(t) \exp(-t^2)\, dt$: application to logistic normal models. *J. Am. Statist. Assoc.* **85**, 464–9.

Davis, P. J. & Rabinowitz, P. (1975). *Methods of Numerical Integration.* New York: Academic Press.

De Bruijn, N. G. (1961). *Asymptotic Methods in Analysis.* New York: Dover.

Liu, Q. & Pierce, D. A. (1993). Heterogeneity in Mantel–Haenszel-type models. *Biometrika* **80**, 543–56.

Naylor, J. C. & Smith, A. F. M. (1982). Applications of a method for the efficient computation of posterior distributions. *Appl. Statist.* **31**, 214–25.

SERC (1989). *EGRET Users' Manual.* Seattle: Statistics and Epidemiology Research Corporation.

Tierney, L. & Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Am. Statist. Assoc.* **81**, 82–6.

Wong, H. W. & Li, B. (1992). Laplace expansion for posterior densities of nonlinear functions of parameters. *Biometrika* **79**, 393–8.

[*Received August* 1993]