Bayesian Analysis (2010) **5**, Number 4, pp. 741–764

A Bayesian Edgeworth expansion by Stein's Identity

Ruby C. Weng*

Abstract. The Edgeworth expansion is a series that approximates a probability distribution in terms of its cumulants. One can derive it by first expanding the probability distribution in Hermite orthogonal functions and then collecting terms in powers of the sample size. This paper derives an expansion for posterior distributions which possesses these features of an Edgeworth series. The techniques used are a version of Stein's Identity and properties of Hermite polynomials. Two examples are provided to illustrate the accuracy of our series.

Keywords: Edgeworth expansion; Hermite polynomials; Laplace method; marginal posterior distribution; Stein's identity.

1 Introduction

The Edgeworth expansion, named after F. Y. Edgeworth (1845-1926), is an expansion that approximates a probability distribution in terms of its cumulants. It is over a century old and it provides an improvement to the central limit theorem. In the past decades it has received a revival of interest in statistics; for example, see Hall (1992) on how Edgeworth expansion and bootstrap methods can help explain each other. The Edgeworth expansion has been applied to other areas as well; for example, Blinnikov and Moessner (1998) compared Gram-Charlier, Gauss-Hermite and Edgeworth expansions in problems of astrophysics, and Filho and Rosenfeld (2004) considered the problem of testing option pricing with Edgeworth expansion, among others. Actually, Blinnikov and Moessner (1998) gave a simple algorithm to calculate higher-order terms of Edgeworth expansion, and they obtained the cumulants up to 10th order in the application to peculiar velocities from cosmic strings.

Wallace (1958, Section 3) and Blinnikov and Moessner (1998) provided reviews on early developments of the series. Let F be the distribution to be approximated and $\{\kappa_r\}$ its cumulants; let γ_r be the cumulants of a standard normal distribution and D the differential operator representing differentiation with respect to x; let Φ and ϕ be the cdf and pdf of a standard normal variable. Chebyshev and Charlier considered the identity

$$F(x) = \exp \sum_{r=1}^{\infty} (\kappa_r - \gamma_r) \frac{(-D)^r}{r!} \Phi(x)$$

and proceeded by expanding and collecting terms according to the order of the derivatives. The resulting expansion is commonly known as the Gram-Charlier series (of type

*Department of Statistics, National Chengchi University, Taipei, Taiwan, mailto:chweng@nccu.edu.tw

 DOI:10.1214/10-BA526

A) and it turned out to be identical with the expansion of F in Hermite orthogonal functions; or equivalently, for a pdf $p(x)$,

$$p(x) = \sum_{k=0}^{\infty} c_k q_k(x)\phi(x), \tag{1}$$

where q_k are Hermite polynomials and, by the orthogonal property (29) below,

$$c_k = \frac{1}{k!}\int_{-\infty}^{\infty} p(x)q_k(x)dx. \tag{2}$$

Blinnikov and Moessner (1998, Section 4) also showed that the Gram-Charlier series (1) is just a Fourier expansion of $p(x)/\phi(x)$ in Hermite polynomials. Note that the sample size plays no role in this expansion, and it is known that this expansion has poor convergence properties; see Cramér (1957). Edgeworth considered the standardized sum of n independent and identically distributed random variables, and developed a similar expansion. Actually, the Edgeworth series can be obtained by collecting terms in the Gram-Charlier series according to powers of n.

The most basic result of Edgeworth expansion is for independent and identically distributed random variables $X_1, ..., X_n$ with mean θ_0 and finite variance σ^2. Let $\hat{\theta}_n$ be the sample mean of X_i's. Under regularity conditions, the distribution function of $Y \equiv n^{1/2}(\hat{\theta}_n - \theta_0)$ may be expanded as

$$P(\frac{n^{1/2}(\hat{\theta}_n - \theta_0)}{\sigma} \leq x) = \Phi(x) + n^{-1/2}p_1(x)\phi(x) + \cdots + n^{-j/2}p_j(x)\phi(x) + \cdots. \tag{3}$$

Formula (3) is termed an Edgeworth expansion. The functions p_j are polynomials with coefficients depending on cumulants of $\hat{\theta}_n - \theta_0$. In particular, p_j is a polynomial of degree at most $3j-1$ and is an odd or even function according to whether j is even or odd.

Many researchers have derived Edgeworth expansions in non-iid contexts; for example, Bickel and Ghosh (1990) considered the signed-root transformation, and Jing and Wang (2003) obtained expansions for U-statistics. There are also studies from a Bayesian perspective. Let g be a smooth function of the parameter θ. The usual approach to asymptotic posterior expansions starts from writing the posterior mean of $g(\theta)$ as a ratio of two integrals,

$$E_\xi[g(\theta)|x_t] = \frac{\int g(\theta)\exp(\ell_t(\theta))\xi(\theta)d\theta}{\int \exp(\ell_t(\theta))\xi(\theta)d\theta},$$

where ℓ_t is the loglikelihood function and ξ the prior density, next takes a Taylor series expansion at the maximum likelihood estimator and develops expansions on both the numerator and denominator, and then obtains an approximation of the posterior mean by formal division of the two series. Johnson (1967, 1970) provides a careful account of this approach. There are other papers that apply Laplace method to both numerator and denominator and then take the ratio; see, for example, Lindley (1961, 1980), Mosteller and Wallace (1964), Tierney and Kadane (1986), and references therein. However, these asymptotic expansions for posterior distributions are not in terms of the cumulants or moments.

Recently Weng (2003) and Weng and Tsai (2008) applied a version of Stein's Identity, established by Woodroofe (1989, 1992) for integrable expansions for posterior distributions, to asymptotic posterior normality; and Weng and Lin (2010) applied it for Bayesian online ranking. The idea of this identity originated from the famous Stein's lemma (Stein 1981, 1987), but the latter considers the expectations of normal distributions, while the former the expectations of distributions which are "nearly" normal (in the sense of (4) below). The application of this identity to posterior normality starts by writing the posterior density of a normalized maximum likelihood estimator Z_t in a form close to normal, next applies Stein's Identity to obtain an expansion for posterior expectations of $h(Z_t)$, and then analyzes the remainder term in the expansion. The present paper takes one step further to show that by repeatedly employing Stein's Identity, together with some properties of Hermite polynomials, one can expand the marginal posterior distribution in the form of (1); then, we proceed to obtain the orders of the c_k terms (2) and form an asymptotic series. Note that our expansion resembles the classic Edgeworth expansion in that both are directly connected to the cumulants or moments, and both can be viewed as an expansion of the probability distribution in Hermite orthogonal functions together with rearrangement of terms in powers of the sample size. These two properties are lost in existing posterior expansions in the literature. The advantage of expressing a distribution in terms of the moments is that the information about the distribution can be efficiently stored.

This paper is organized as the following. The next section introduces Stein's Identity and the model. Section 3 starts with reviews of Hermite polynomials, and then develops a Bayesian Edgeworth expansion. Section 4 provides detailed comparisons with Johnson (1970). Section 5 presents two examples for illustration. Section 6 gives concluding remarks. Appendices contain some proofs.

2 Stein's Identity and the Model

2.1 Stein's Identity

Let Φ_p denote the standard p-variate normal distribution and ϕ_p the density; let Φ be the abbreviation of Φ_1, and similarly for ϕ. Write

$$\Phi_p h = \int h d\Phi_p$$

for functions h for which the integral is finite. Next let Γ denote a finite signed measure of the form

$$d\Gamma = f d\Phi_p, \tag{4}$$

where f is a real-valued function defined on $\Re^p$ satisfying $\Phi_p|f| < \infty$. For $s > 0$, denote H_s as the collection of all measurable functions $h: \Re^p \to \Re$ for which $|h(z)|/b \leq 1+\|z\|^s$

for some $b > 0$. Given $h \in H_s$, let $h_0 = \Phi_p h$, $h_p = h$,

$$h_k(y_1, ..., y_k) = \int_{\Re^{p-k}} h(y_1, ..., y_k, w)\Phi_{p-k}(dw), \tag{5}$$

$$g_k(y_1, ..., y_p) = e^{\frac{1}{2}y_k^2} \int_{y_k}^{\infty} [h_k(y_1, ..., y_{k-1}, w) - h_{k-1}(y_1, ..., y_{k-1})]e^{-\frac{1}{2}w^2} dw, \tag{6}$$

for $-\infty < y_1, ..., y_p < \infty$ and $k = 1, ..., p$. Then let $Uh = (g_1, ..., g_p)^T$ and $Vh = (U^2h + U^2h^T)/2$, where U^2h is the $p \times p$ matrix whose k-th column is Ug_k and g_k is as in (6). For example, for $z \in \Re^p$, if $h(z) = z_1$, then $Uh(z) = (1, 0, ..., 0)^T$ and if $h(z) = \|z\|^2$, then $Uh(z) = z$. Simple calculations by taking $f(z)$ in Lemma 1 below as z_i and $z_i z_j$ yield

$$\Phi_p(Uh) = \int_{\Re^p} z h(z)\Phi_p(dz), \tag{7}$$

$$\Phi_p(U^2h) = \int_{\Re^p} \frac{1}{2}(zz^T - 1)h(z)\Phi_p(dz). \tag{8}$$

Lemma 1. *(Stein′s Identity) Let r be a nonnegative integer. Suppose that $d\Gamma = f d\Phi_p$ as above, where f is a differentiable function on $\Re^p$, and that*

$$\int_{\Re^p} |f(z)|\Phi_p(dz) + \int_{\Re^p} (1 + \|z\|^r)\|\nabla f(z)\|\Phi_p(dz) < \infty. \tag{9}$$

Then

$$\Gamma h = \Gamma 1 \cdot \Phi_p h + \int_{\Re^p} (Uh(z))^T \nabla f(z)\Phi_p(dz), \tag{10}$$

for all $h \in H_r$. If $\partial f/\partial z_j$, $j = 1, ..., p$, are differentiable, and

$$\int_{\Re^p} (1 + \|z\|^r)\|\nabla^2 f(z)\|\Phi_p(dz) < \infty, \tag{11}$$

then

$$\Gamma h = \Gamma f \cdot \Phi_p h + (\Phi_p Uh)^T \int_{\Re^p} \nabla f(z)\Phi_p(dz) + \int_{\Re^p} \mathrm{tr}[(Vh(z))\nabla^2 f(z)]\Phi_p(dz), \tag{12}$$

for all $h \in H_r$.

The proof of Lemma 1 is in Woodroofe (1989, Proposition 1); see also Weng and Woodroofe (2000, Lemma 1). Here we sketch the proof as it will be used in Proposition 2 in Section 3. For (10), it follows from an application of the interchange of orders of integration; below we borrow a few lines from Woodroofe (1989). Take $p = 1$ and let $'$ denote the differentiation. By assumptions in Lemma 1, we have $f(x) = \int_{-\infty}^{x} f'(y)dy$

and

$$\begin{aligned}\Gamma h - \Gamma 1 \cdot \Phi h &= \Phi(fh) - \Phi f \cdot \Phi h \\ &= \int_{\Re} \Big\{ \int_{-\infty}^{x} f'(y)dy \Big\} \phi(x)[h(x) - \Phi h]dx \\ &= \int_{\Re} \Big\{ \int_{y}^{\infty} \phi(x)[h(x) - \Phi h]dx \Big\} f'(y)dy \\ &= \int_{\Re} Uh(y)f'(y)\phi(y)dy,\end{aligned}$$

where the interchange of orders of integration is justified by assumed integrability conditions. For (12), it follows by writing

$$(Uh(z))'\nabla f(z) = \sum_{i=1}^{p} g_i(z)\frac{\partial f(z)}{\partial z_i}, \tag{13}$$

and then applying (10) with h and f replacing by g_i and $\partial f/\partial z_i$.

The following lemma will be used later. The proof is in Woodroofe and Coad (1997, Proposition 1); see also Weng and Woodroofe (2000, Lemma 8).

Lemma 2. *If $h(z) \in H_0$, then $Uh \in H_0$. Further, if $h(z) = \|z\|^p$, where $p \geq 1$, then*

$$\|Uh(z)\| \leq C\{1 + \|z\|^{p-1}\}.$$

2.2 The model

Let X_t be a random vector distributed according to a family of probability densities $p_t(x_t|\theta)$, where t is a discrete or continuous parameter and $\theta \in \Theta$, an open subset in $\Re^p$. Assume that the log-likelihood function $\ell_t(\theta)$ is twice differentiable with respect to θ. Assume also that the maximum likelihood estimator $\hat{\theta}_t$ exists and satisfies $\nabla\ell_t(\hat{\theta}_t) = 0$ and that $-\nabla^2\ell_t(\hat{\theta}_t)$ is positive definite, where ∇ indicates differentiation with respect to θ. Define Σ_t and Z_t as

$$\Sigma_t^T\Sigma_t = -\nabla^2\ell_t(\hat{\theta}_t), \tag{14}$$

$$Z_t = \Sigma_t(\theta - \hat{\theta}_t). \tag{15}$$

Consider a Bayesian model in which θ has a prior density ξ. Then the posterior density of θ given data x_t is $\xi_t(\theta) \propto \exp(\ell_t(\theta))\xi(\theta)$, and the posterior density of Z_t is

$$\zeta_t(z) \propto \xi_t(\theta(z)) \propto \exp[\ell_t(\theta) - \ell_t(\hat{\theta}_t)]\xi(\theta), \tag{16}$$

where the relation of θ and z is given in (15). Now define

$$u_t(\theta) = \ell_t(\theta) - \ell_t(\hat{\theta}_t) + \frac{1}{2}\|z_t\|^2. \tag{17}$$

So, (16) can be rewritten as

$$\zeta_t(z) \propto \phi_p(z) f_t(z), \tag{18}$$

where $f_t(z) = \xi(\theta(z))\exp[u_t(\theta)]$.

Observe that the posterior distribution of Z_t in (18) is of a form suitable for Stein's Identity. If ξ is twice differentiable on $\Re^p$ and vanishes off of Θ, then so does $f_t(z)(= \xi(\theta(z))\exp[u_t(\theta)])$. Moreover, if (9) and (11) hold, then by Lemma 1 we have

$$E_\xi^t\{h(Z_t)\} = \Phi_p h + E_\xi^t\Big\{[Uh(Z_t)]^T \frac{\nabla f_t(Z_t)}{f_t(Z_t)}\Big\}, \tag{19}$$

$$E_\xi^t\{h(Z_t)\} = \Phi_p h + (\Phi_p U h)^T E_\xi^t\Big[\frac{\nabla f_t(Z_t)}{f_t(Z_t)}\Big] + E_\xi^t\Big\{\mathrm{tr}\Big[Vh(Z_t)\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}\Big]\Big\}. \tag{20}$$

In particular, if $h(z) = z_i$, $Uh(z) = e_i$; and if $h(z) = z_i z_j$ and $i < j$, $Uh(z) = z_i e_j$. So, (19) and (20) give

$$E_\xi^t Z_t = E_\xi^t\Big(\frac{\nabla f_t(Z_t)}{f_t(Z_t)}\Big) \qquad \text{and} \qquad E_\xi^t(Z_{ti}Z_{tj}) = \delta_{ij} + E_\xi^t\Big[\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}\Big]_{ij}. \tag{21}$$

Throughout $\nabla\xi$ and $\nabla^2\xi$ denote the gradient and Hessian of ξ with respect to θ, ∇f and $\nabla^2 f$ the gradient and Hessian of f with respect to Z, and E_ξ^t the posterior expectation given data x_t. Some calculations are useful for later reference.

$$\frac{\nabla f_t(Z_t)}{f_t(Z_t)} = (\Sigma_t^T)^{-1}\Big[\frac{\nabla\xi(\theta)}{\xi(\theta)} + \nabla u_t(\theta)\Big], \tag{22}$$

$$\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)} = (\Sigma_t^T)^{-1}\Big[\frac{\nabla^2\xi}{\xi} + \frac{\nabla\xi}{\xi}\nabla u_t^T + \nabla u_t\frac{\nabla\xi^T}{\xi} + \nabla^2 u_t + \nabla u_t \nabla u_t^T\Big]\Sigma_t^{-1}, \tag{23}$$

where by (17) we can derive

$$\nabla u_t(\theta) = \nabla\ell_t(\theta) - \nabla^2\ell_t(\hat\theta_t)(\theta - \hat\theta_t), \tag{24}$$

$$\nabla^2 u_t(\theta) = \nabla^2\ell_t(\theta) - \nabla^2\ell_t(\hat\theta_t). \tag{25}$$

3 Edgeworth expansions

3.1 Hermite polynomials

We shall review Hermite polynomials as they are closely related to the Edgeworth expansion. Let q_k denote Hermite polynomials, given by

$$q_k(z)\phi(z) = \Big(-\frac{d}{dz}\Big)^k \phi(z). \tag{26}$$

For instance, for $k = 0, 1, ..., 5$ we have $q_0(z) = 1$, $q_1(z) = z$, $q_2(z) = z^2 - 1$, $q_3(z) = z^3 - 3z$, $q_4(z) = z^4 - 6z^2 + 3$, and $q_5(z) = z^5 - 10z^3 + 15z$. These polynomials are an orthogonal polynomial sequence in the sense of (29) below. The one in (26) is the probabilist's version, while the physicist's version is defined by

$$q_k^{\text{phy}}(z)e^{-z^2} = \Big(-\frac{d}{dz}\Big)^k e^{-z^2}.$$

It is easily seen that these two versions differ in just the scaling: $q_k^{\text{phy}}(z) = 2^{n/2} q_k(\sqrt{2}z)$. Courant and Hilbert (1953, Section 9) provided several properties of q_k^{phy}. In fact, Hermite polynomials are solutions of the simple harmonic oscillator of quantum mechanics (see Boas (2006, Section 22) and Weber and Arfken (2004, Chapter 13)) and they are integral parts of mathematical physics. We review three properties, numbered (27)-(29) below, for later use. Let $q'_k(z)$ denote the differentiation with respect to z. Then,

$$q'_k(z) = kq_{k-1}(z), \tag{27}$$

$$q_{k+1}(z) = zq_k(z) - kq_{k-1}(z), \tag{28}$$

$$\int q_k(z)q_j(z)d\Phi(z) = \begin{cases} 0 & \text{if } k \neq j, \\ k! & \text{if } k = j. \end{cases} \tag{29}$$

For the sake of being self-contained, we outline the proofs of (27)-(29). First, define a generating function

$$\psi(z,t) = e^{-\frac{t^2}{2}+tz} = e^{\frac{z^2}{2}-\frac{(t-z)^2}{2}} = \sum_{n=0}^{\infty} \frac{q_n(z)}{n!} t^n.$$

From this equation it follows that

$$q_n(z) = \Big(\frac{\partial \psi(z,t)}{\partial t^n}\Big)\Big|_{t=0} = (-1)^n e^{z^2/2} \frac{d^n e^{-z^2/2}}{dz^n}, \tag{30}$$

which is equivalent to (26). Next, the relation $\partial\psi(z,t)/\partial z = t\psi(z,t)$ gives (27); and from the relation $\partial\psi(z,t)/\partial t + (t-z)\psi(z,t) = 0$ we obtain the recursive relation (28). Finally, the orthogonal property (29) can be derived from

$$\begin{aligned} \int_{-\infty}^{\infty} q_m(z)q_n(z)e^{-\frac{z^2}{2}}dz &= (-1)^n \int_{-\infty}^{\infty} q_m(z)\frac{d^n e^{-z^2/2}}{dz^n}dz \\ &= \cdots = (-1)^{n-m} m! \int_{-\infty}^{\infty} q_0(z)\frac{d^{n-m} e^{-z^2/2}}{dz^{n-m}}dz = 0 \end{aligned}$$

for $n > m$ by repeated partial integration, keeping in mind equation (30) and the fact that $e^{-z^2/2}$ and all its derivatives vanish for infinite z.

With (27) and (28) we can prove the following proposition, which is needed in Section 3.2. We defer the proof to Appendix 6. Define $C_i^k = k!/(i!(k-i)!)$.

Proposition 1. *Let Z denote a standard normal random variable. Then, for $k = 1, 2, ...$*

$$q_k(x) = x^k - \sum_{i=0}^{k-1} C_i^k q_i(x) E(Z^{k-i}). \tag{31}$$

3.2 Bayesian Edgeworth expansion

Recall that $Uh = (g_1, ..., g_p)^T$ is defined following (6). In the lemma below, we write $g_l = (Uh)_l$.

Lemma 3. *Suppose that $h \in H_r$ and that $h(z) = h^*(z_i)$, where $i \in \{1, ..., p\}$ and $h^* : \Re \to \Re$. Then, $(Uh)_l = 0$ if $l \neq i$ and $(Uh)_i(z) = (Uh)_i(z_i) = Uh^*(z_i)$, depending only on z_i.*

Proof. Since $h(z) = h^*(z_i)$, from (5) it is not difficult to see that $h_l = \Phi_p h$ for $l = 0, ..., i-1$ and that $h_l(z) = h(z_i)$ for $l = i, ..., p$. Then, by (6), the desired results follow. □

The following result follows from Lemma 3 and Lemma 1. It is useful for developing marginal posterior distributions.

Proposition 2. *Let r and s be nonnegative integers. Suppose that $d\Gamma = f d\Phi_p$, where f is a differentiable function on $\Re^p$. Suppose also that $h \in H_r$, that $h(z) = h^*(z_i)$, where $i \in \{1, ..., p\}$ and $h^* : \Re \to \Re$, and that*

$$\int_{\Re^p} |f(z)| \Phi_p(dz) + \int_{\Re^p} (1 + |z_i|^r) \left\| \frac{\partial^k f(z)}{\partial z_i^k} \right\| \Phi_p(dz) < \infty, \tag{32}$$

for $k \leq s$. Then,

$$\Gamma h = \Gamma 1 \cdot \Phi h^* + \sum_{j=1}^{s-1} (\Phi U^j h^*) \int_{\Re^p} \frac{\partial^j f(z)}{\partial z_i^j} \Phi_p(dz) + \int_{\Re^p} U^s h^*(z_i) \frac{\partial^s f(z)}{\partial z_i^s} \Phi_p(dz). \tag{33}$$

Proof. If $h(z) = h^*(z_i)$, then by Lemma 3 and (13) we can write (10) as

$$\Gamma h = \Gamma h^* = \Gamma 1 \cdot \Phi h^* + \int_{\Re^p} U h^*(z_i) \frac{\partial f(z)}{\partial z_i} \Phi_p(dz). \tag{34}$$

Next applying (34) with h^* and f replaced by Uh^* and $\partial f / \partial z_i$ yields

$$\Gamma h = \Gamma h^* = \Gamma 1 \cdot \Phi h^* + \Phi U h^*(z_i) \int_{\Re^p} \frac{\partial f(z)}{\partial z_i} \Phi_p(dz) + \int_{\Re^p} U^2 h^*(z_i) \frac{\partial^2 f(z)}{\partial z_i^2} \Phi_p(dz).$$

Repeatedly applying (34) with h^* and f replaced by $U^j h^*$ and $\partial^j f / \partial z_i^j$ gives (33). □

To apply this proposition to the posterior distribution of Z_t, we need the integrability condition (32), which involves $\partial^k f_t(z) / \partial z_i^k$. For $k = 1, 2$, $\partial^k f_t(z) / \partial z_i^k$ can be obtained from (22) and (23). For $k \geq 3$, the forms are complicated; however, for the purpose of verifying (32), it suffices to use a 1-dimensional notation. For any function $g(\theta)$, let $g^{(k)}$ denote the kth derivative with respect to θ. Recall from (18) that $f_t(z) = \xi(\theta(z))\exp[u_t(\theta)]$. Straightforward calculations give

$$\frac{d^k f_t(z)}{dz^k} = \Big(\frac{d\theta}{dz}\Big)^k f_t(z) G_k(\theta), \tag{35}$$

where

$$G_1 = \frac{\xi^{(1)}}{\xi} + u_t^{(1)} \qquad \text{and} \qquad G_k = G_1 G_{k-1} + G_{k-1}^{(1)}. \tag{36}$$

For example,

$$\begin{aligned} G_1(\theta) &= u_t^{(1)} + \frac{\xi^{(1)}}{\xi}, \\ G_2(\theta) &= [u_t^{(1)}]^2 + u_t^{(2)} + 2u_t^{(1)}\frac{\xi^{(1)}}{\xi} - \Big(\frac{\xi^{(1)}}{\xi}\Big)^2, \\ G_3(\theta) &= [u_t^{(1)}]^3 + 3u_t^{(1)}u_t^{(2)} + u_t^{(3)} + 3[u_t^{(1)}]^2\frac{\xi^{(1)}}{\xi} + 3u_t^{(2)}\frac{\xi^{(1)}}{\xi} - u_t^{(1)}\Big(\frac{\xi^{(1)}}{\xi}\Big)^2 \\ &\quad + 2u_t^{(1)}\frac{\xi^{(2)}}{\xi} + \Big(\frac{\xi^{(1)}}{\xi}\Big)^3 - 2\Big(\frac{\xi^{(1)}}{\xi}\Big)\Big(\frac{\xi^{(2)}}{\xi}\Big), \end{aligned}$$

where G_1 and G_2 are 1-dimensional versions of (22) and (23). In general, we can show that G_k has the form:

$$G_k(\theta) = \sum_l c_{kl}\Big\{\Big(\prod_{i=1}^k [u_t^{(i)}]^{r_{ki}}\Big)\Big[\prod_{j=1}^k \Big(\frac{\xi^{(j)}}{\xi}\Big)^{s_{kj}}\Big]\Big\}, \tag{37}$$

where r_{ki} and s_{kj} satisfy

$$\sum_{i=1}^k (i r_{ki}) + \sum_{j=1}^k (j s_{kj}) = k. \tag{38}$$

Note that r_{ki} and s_{kj} depend on l, but we suppressed the dependence in the notation. The proofs of (36)-(38) are in Appendix 6.

To ensure (32), the conditions below are required.

(A1) For each $r > 0$, $E_\xi^t(\|Z_t\|^r) = O(1)$.

(A2) For any $k \geq 3$, $\ell_t^{(k)}(\theta)/t$ is uniformly bounded in t and in $\theta \in \Theta$.

(A3) $\|\xi^{(k)}\|/\xi \leq b(1 + \|\theta\|^s)$ for some $b > 0$ and $s > 0$.

Here $O(1)$ means convergence of a sequence of real numbers as $t \to \infty$. Note that condition (A3) holds for a wide class of distributions, and it implies that $\|\xi^{(k)}\|/\xi \leq$

$b(1+\|\hat\theta_t+\Sigma_t^{-1}z_t\|^s) \le b_t(1+\|z_t\|^s)$ for some $0<b_t<\infty$, where b_t may depend on the data x_t.

Now we can verify (32) using the 1-dimensional notation. First, since ζ_t in (18) is a posterior density, the integral $\int |f_t|\Phi_p(dz)$ is finite and we denote it as C_t. Next from the expression (35) we have

$$\begin{aligned}
& \int (1+|z|^r)\Big|\frac{d^k f_t(z)}{dz^k}\Big|\Phi(dz) \\
= & \int (1+|z|^r)\Big|\Big(\frac{d\theta}{dz}\Big)^k f_t(z)G_k(\theta)\Big|\Phi(dz) \\
= & C_t\Big(\frac{d\theta}{dz}\Big)^k E_\xi^t\Big((1+|Z_t|^r)|G_k(\theta)|\Big) \\
\le & b_t^* C_t\Big(\frac{d\theta}{dz}\Big)^k E_\xi^t\Big((1+|Z_t|^r)(1+|Z_t|^s)(\prod_{i=1}^k [u_t^{(i)}]^{r_i})\Big), \qquad (39)
\end{aligned}$$

where $0<b_t^*<\infty$ and the last line follows from (37) and condition (A3); moreover, from (24) and (25) and the Mean Value Theorem it follows that

$$u_t^{(1)}(\theta) = \frac{1}{2}\ell_t^{(3)}(\eta_t)\delta_t^2, \quad u_t^{(2)}(\theta) = \ell_t^{(3)}(\omega_t)\delta_t, \quad u_t^{(3)}(\theta) = \ell_t^{(3)}(\theta), \qquad (40)$$

where η_t and ω_t lie between θ and $\hat\theta_t$. Then, by (A1) and (A2) the right hand side of (39) is finite. Therefore, we have the following theorem.

Theorem 1. *Suppose that $\xi(\theta)$ and $\ell_t(\theta)$ are s times differentiable and that conditions (A1)-(A3) hold. Then, for $k\le s$*

$$\int_{\Re^p} |f_t(z)|\Phi_p(dz) + \int_{\Re^p}(1+|z_i|^r)\Big\|\frac{\partial^k f_t(z)}{\partial z_i^k}\Big\|\Phi_p(dz) < \infty;$$

and hence, for h^* as in Proposition 2 we have

$$E_\xi^t(h^*(Z_{ti})) = \Phi h^* + \sum_{j=1}^{s-1}(\Phi U^j h^*)E_\xi^t\Big[\frac{\partial^j f_t/\partial z_{ti}^j}{f_t}(Z_t)\Big] + E_\xi^t\Big\{[U^s h^*(Z_{ti})]\frac{\partial^s f_t/\partial z_{ti}^s}{f_t}(Z_t)\Big\}. \qquad (41)$$

The next two propositions connect the posterior expansion (41) with Hermite polynomials q_k (26) and the moments of Z_{ti}.

Proposition 3. *Let $h^*:\Re\to\Re$ be a measurable function. Then, for $k=1,2,...$*

$$\Phi(U^k h^*) = \frac{1}{k!}\int_\Re q_k(z)h^*(z)\Phi(dz). \qquad (42)$$

Proof. We shall prove it by induction. For $k=1,2$, (42) yields exactly (7) and (8). Now suppose that (42) holds for $k=1,...,n-1$. In Proposition 2, take $s=n+1$ and

$f(z) = z_i^n$, noting that (32) holds for this f and $\partial^{n+1} f / \partial z_i^{n+1} = 0$. With this f and using (42) for $k = 1, ..., n-1$, (33) becomes

$$\Phi U^n h^* = \frac{1}{n!} \int_{\Re} \Big[z^n - \sum_{i=1}^{n-1} C_i^n q_i(z) E(Z^{n-i}) \Big] h^*(z) d\Phi(z),$$

where Z denotes the standard normal variate. Then, by Proposition 1 the right hand side of the above line is $(1/n!) \int_{\Re} q_n(z) h^*(z) \Phi(dz)$. So, (42) holds for $k = n$. □

Proposition 4. *Suppose that $E_\xi^t(Z_{ti}^k) < \infty$. Then,*

$$E_\xi^t \Big(\frac{\partial^k f_t / \partial z_{ti}^k}{f_t} (Z_t) \Big) = E_\xi^t (q_k(Z_{ti})).$$

Proof. First, in (41) take $h^*(z) = q_k(z)$ and $s = k$; therefore, $\Phi h^* = 0$, $U^k h^*(z) = 1$, and

$$E_\xi^t(q_k(Z_{ti})) = \sum_{j=1}^{k-1} (\Phi U^j h^*) E_\xi^t \Big(\frac{\partial^j f_t / \partial z_{ti}^j}{f_t} (Z_t) \Big) + E_\xi^t \Big(\frac{\partial^k f_t / \partial z_{ti}^k}{f_t} (Z_t) \Big),$$

where from Proposition 3 and the orthogonality property (29) we have

$$\Phi U^j h^* = \frac{1}{j!} \int q_j(z) h^*(z) d\Phi = \frac{1}{j!} \int q_j(z) q_k(z) d\Phi = 0 \text{ for } j \neq k.$$

So, the desired result follows. □

Note that when $k = 1, 2$ the above proposition gives the 1-dimensional version of (21). Take h^* in (41) as the indicator function $1(z_{ti} \leq w)$, where $w \in \Re$. Then, Propositions 3 and 4 and the relation

$$\int_{-\infty}^{w} q_k(z) \phi(z) dz = -q_{k-1}(w) \phi(w) \tag{43}$$

together suggests that the marginal posterior density of Z_{ti} has the form

$$\zeta_t(z_i) = \sum_{k=0}^{\infty} c_k q_k(z_i) \phi(z_i), \tag{44}$$

where

$$c_k = \frac{1}{k!} \int_{-\infty}^{\infty} \zeta_t(z_i) q_k(z_i) dz_i = \frac{1}{k!} E_\xi^t(q_k(Z_{ti})).$$

Equation (44) is essentially (1).

Our next theorem concerns the orders of terms in (41). By (40) and (A1)-(A3) we have that in (37) the terms associated with $|u_t^{(i)}|^{r_i}$, $i \geq 3$, contribute $O(t^{r_i})$ to $E_\xi^t[G_k(\theta)]$, while $|u_t^{(1)}|^{r_1}$ contributes $O(1)$ and $|u_t^{(2)}|^{r_2}$ contributes $O(t^{r_2/2})$; for example, by (40)

$$\begin{aligned} E_\xi^t\Big\{[u_t^{(1)}]^2[u_t^{(3)}]^3\Big(\frac{\xi^{(1)}}{\xi}\Big)^2\Big\} &= E_\xi^t\Big\{\Big[\frac{1}{2}\ell_t^{(3)}(\eta_t)\delta_t^2\Big]^2[\hat{\ell}_t^{(3)}]^3\Big(\frac{\xi^{(1)}}{\xi}\Big)^2\Big\} \\ &\leq Ct^3E_\xi^t\Big(\frac{\xi^{(1)}}{\xi}\Big)^2 \\ &= O(t^3), \end{aligned}$$

where the second line follows from (A1) and (A2), and the last line from (A1) and (A3). Together with the constraint (38), it is not difficult to see that the highest order of G_k is $\lfloor k/3 \rfloor$, the greatest integer not exceeding $k/3$. Furthermore, if $-\nabla^2\hat{\ell}_t = O(t)$, then

$$E_\xi^t\Big(\frac{d^k f_t(z)/dz^k}{f_t}\Big) = E_\xi^t\Big[\Big(\frac{d\theta}{dz}\Big)^k G_k(\theta)\Big] = O(t^{-\frac{k}{2}+\lfloor\frac{k}{3}\rfloor}) = O(t^{-\frac{i}{2}}) \text{ if } k \in J_i, \tag{45}$$

where $J_1 = \{1,3\}$ and $J_i = \{3i-4, 3i-2, 3i\}$ for $i > 1$; for example, $J_2 = \{2,4,6\}$, $J_3 = \{5,7,9\}$, $J_4 = \{8,10,12\}$. So, if $h \in H_r$ and $h(z) = h^*(z_p)$, then by Lemma 2 it follows that $U^s h^*$ is in H_{r-s} if $r > s$ and in H_0 if $r \leq s$; hence,

$$\sup_{h\in H_r}\Big|E_\xi^t\Big\{[U^s h^*(Z_{ti})]\frac{\partial^s f_t/\partial z_{ti}^s}{f_t}(Z_t)\Big\}\Big| = O(t^{-\frac{s}{2}+\lfloor\frac{s}{3}\rfloor}).$$

The above arguments lead to the following theorem.

Theorem 2. *Suppose that $\xi(\theta)$ and $\ell_t(\theta)$ are $(3s+1)$ times differentiable, that conditions (A1)-(A3) hold, and that $-\nabla^2\hat{\ell}_t = O(t)$. Then,*

$$\sup_{h\in H_r}\Big|E_\xi^t(h^*(Z_{ti})) - \Phi h^* - \sum_{\substack{k\in\{1,\ldots,3s\}\\ k\neq 3s-1}}(\Phi U^k h^*)E_\xi^t\Big[\frac{\partial^k f_t/\partial z_{ti}^k}{f_t}(Z_t)\Big]\Big| = O(t^{-\frac{3s+1}{2}+s}). \tag{46}$$

Note that in (46) the summation excludes $k = 3s-1$ because by (45) this term has the same order as the remainder term. Note also that Proposition 4 and (45) together imply that

$$E_\xi^t(q_k(Z_{ti})) = O(t^{-\frac{j}{2}}) \text{ for } k \in J_j. \tag{47}$$

Now suppose that Σ_t in (14) is obtained by a Cholesky decomposition. So, it is upper triangular and Z_{tp} has a simpler form:

$$Z_{tp} = [\Sigma_t]_{pp}(\theta_p - \hat{\theta}_{tp}). \tag{48}$$

Corollary 1. *Let Σ_t in (14) be upper triangular so that Z_{tp} has the form (48). Take h^* in (41) as the indicator function $1(z_{tp} \leq w)$, where $w \in \Re$. Then, the marginal posterior distribution for the individual parameter θ_p is $P_\xi^t(\theta_p \leq a) = P_\xi^t(Z_{tp} \leq w)$ and*

$$\sup_{w\in\Re}\Big|P_\xi^t(Z_{tp}\leq w) - \Phi(w) - \sum_{\substack{i\in\{1,\ldots,3s\}\\ i\neq 3s-1}}\frac{1}{i!}q_{i-1}(w)\phi(w)E_\xi^t(q_i(Z_{tp}))\Big| = O(t^{-\frac{3s+1}{2}+s}), \tag{49}$$

where $w = [\Sigma_t]_{pp}(a - \hat{\theta}_{tp})$*. Moreover, the marginal posterior density for* θ_p *is*

$$\xi_p^t(a) = [\Sigma_t]_{pp}\{\phi(w) + \sum_{\substack{i\in\{1,\ldots,3s\}\\ i\neq 3s-1}} \frac{1}{i!} q_i(w)\phi(w)E_\xi^t(q_i(Z_{tp})) + O(t^{-\frac{3s+1}{2}+s})\}. \tag{50}$$

Proof. Equation (49) follows from (46), Propositions 3 and 4, and the relation (43). Equation (50) follows by taking derivative of (49) with respect to a and using the fact that, by (26), $(d/dw)[q_{i-1}(w)\phi(w)] = -q_i(w)\phi(w)$. □

We can rearrange (49) to be

$$P_\xi^t(Z_{tp} \le w) = \Phi(w) + \sum_{i=1}^m R_i(w)\phi(w) + O(t^{-\frac{m+1}{2}}), \tag{51}$$

where

$$R_i(w) \quad = \quad \sum_{j\in J_i} \frac{1}{j!} q_{j-1}(w)\phi(w)E_\xi^t(q_j(Z_{tp})) = O(t^{-\frac{i}{2}})$$

by (47). Moreover, the function R_i is a polynomial of degree at most $3i - 1$ and is an odd or even function according to whether i is even or odd; and the coefficients of this polynomial depend on moments of Z_{tp}. So, (51) also has the properties of the Edgeworth expansion in (3), and hence we term it a Bayesian Edgeworth expansion.

Similarly, we can rearrange (50) to be

$$\xi_p^t(a) = [\Sigma_t]_{pp}\{\phi(w) + \sum_{i=1}^m Q_i(w)\phi(w) + O(t^{-\frac{m+1}{2}})\}, \tag{52}$$

where

$$Q_i(w) = \sum_{j\in J_i} \frac{1}{j!} q_j(w)\phi(w)E_\xi^t(q_j(Z_{tp})) = O(t^{-\frac{i}{2}}).$$

In particular if $j = 2$, the approximations (51) and (52) are accurate to $O(t^{-3/2})$, which is often called a second order approximation.

4 Some comparisons

Johnson (1970) showed that the centered and scaled posterior distribution possesses an asymptotic expansion in powers of $t^{-1/2}$ (where t is the sample size) having the standard normal as a leading term. Let ψ denote the centered and scaled variable (see his Eq. (2.1), p. 853) defined by

$$\psi = (\theta - \hat{\theta}_t)b(\hat{\theta}_t), \tag{53}$$

where

$$b(\hat{\theta}_t) = \Big[-\frac{1}{t}\sum_{i=1}^t \frac{\partial^2}{\partial\theta^2}\log f(x_i,\theta)|_{\theta=\hat{\theta}_t}\Big]^{1/2}.$$

Denote the posterior cdf of $t^{1/2}\psi$ by F_t. Then, his Theorem 2.1 gives the expansion for posterior distribution F_t:

$$|F_t(w) - \Phi(w) - \sum_{j=1}^{K} \gamma_j(w,x) t^{-j/2}| \leq D_1 t^{-\frac{1}{2}(K+1)}, \tag{54}$$

and his Proposition 2.1 shows that each $\gamma_j(w,x)$ is a polynomial in w having coefficients bounded in x multiplied by the standard normal density. In particular, the forms of γ_1 and γ_2 are given in his Section 2.4 (see Eq. (2.25) and (2.26), p.858):

$$\gamma_1(w,x) = -\phi(w) c_{00}^{-1} [c_{10}(w^2+2) + c_{01}], \tag{55}$$
$$\gamma_2(w,x) = -\phi(w) c_{00}^{-1} [c_{20} w^5 + (5c_{20} + c_{11}) w^3 + (15c_{20} + 3c_{11} + c_{02}) w], \tag{56}$$

where the c_{lm} can be expressed in terms of the prior ξ and the likelihood together with their derivatives ($\xi^{(1)}$, a_{3t}, a_{4t}):

$$\begin{aligned} &c_{00} = \xi(\hat{\theta}_t); c_{01} = b^{-1}\xi^{(1)}(\hat{\theta}_t); c_{02} = b^{-2}\xi^{(2)}(\hat{\theta}_t);\\ &c_{10} = b^{-3} a_{3t}(\hat{\theta}_t)\xi(\hat{\theta}_t); c_{11} = b^{-4} a_{4t}(\hat{\theta}_t)\xi(\hat{\theta}_t) + b^{-4} a_{3t}(\hat{\theta}_t)\xi^{(1)}(\hat{\theta}_t);\\ &c_{20} = 2^{-1} b^{-6} a_{3t}^2(\hat{\theta}_t)\xi(\hat{\theta}_t). \end{aligned}$$

Since our normalized quantity Z_t in (15) is the multivariate version of ψ (53), it is of interest to compare his expansion with ours. First, we observe some similarities: terms with $i = 1, 3$ in (49) are of order $t^{-1/2}$, corresponding to Hermite polynomials q_0 and q_2, which agrees with the degrees of the polynomials in γ_1 (55); terms with $i = 2, 4, 6$ in (49) are of order t^{-1}, corresponding to Hermite polynomials q_1, q_3, q_5, which agrees with the degrees of the polynomials in γ_2 (56). In fact in our (49), if we substitute the posterior moments by asymptotic moment approximations to suitable orders, it will lead to Johnson's formula.

The main difference between these two expansions is that our expansion is in terms of moments, while Johnson's is in terms of prior and likelihood together with their derivatives (the expressions for higher order terms of γ_j may be complicated). Such difference in expression may be due to using different approaches: theirs is based on Taylor expansion, but ours is based on Stein's identity.

Ghosh et al. (1982) have also studied the expansions of the posterior distribution. Their expansion is the same as Johnson (1970), but while Johnson (1970) considers valid posterior expansions under P_{θ_0}, they study the expansion under P_ξ, where ξ is the prior.

5 Examples

We provide two examples to show that the expansion (50) has the ability to capture the shape of the posterior distribution even if it is skewed or is not unimodal; in these cases, approximations correct to $O(t^{-3/2})$ do not provide good estimates.

Remember that, with a Fourier series, one can store a function by part of its Fourier coefficients. The same thing applies to an Edgeworth expansion. For instance, in our example 5.1, we can suitably recover the posterior density by a few posterior moments.

All computations here are done in R (R Development Core Team 2009) and available at `http://www3.nccu.edu.tw/~chweng/publication.htm`

5.1 Binomial model

Consider a binomial variable $X \sim \text{Bin}(t, \theta)$, where the prior of θ is assumed to be Beta(a, b). Suppose that $a = 0.5, b = 4, t = 5, x = 2$. Thus, the sample size is small and the posterior distribution of θ, Beta(2.5,7), is skewed.

Figure 1 presents the true posterior density of θ and the estimates using (50) with $s = 2$ and 13 (corresponding to orders $O(t^{-3/2})$ and $O(t^{-7})$, respectively). Here the moments in (50) are approximated by numerical integration. The figure suggests that an approximation to order $O(t^{-3/2})$ is not satisfactory. Further, information about this density can be stored by $\hat{\theta}_t$, Σ_t, and these moments. Also included is Johnson's approximation to $O(t^{-1})$, obtained by taking $K = 1$ in (54); that is,

$$p_t(w) \equiv \frac{dF_t(w)}{dw} = \phi(w) + \frac{d\gamma_j(w, x)}{dw} t^{-1/2} + O(t^{-1}).$$

Perhaps due to small sample size, this density approximation takes negative values around $\theta = 0.7$; and the approximation to $O(t^{-1})$ by taking $K = 2$ in (54) is no better and not shown here.

Figure 2 gives the approximate posterior density of θ using (50) with 5, 6, 7, 8, 9, 10, 20, 40 moments of Z_{tp}. As expected, the curves get closer to the true density when more moments are used.

We also try a large sample case to assess Johnson's result. We take $t = 50$ and $x = 20$, and keep a and b unchanged. Figure 3 gives the exact density, normal approximation, and Johnson's approximation to $O(t^{-1})$. The figure shows that Johnson's approximation has improved upon normal approximation. The results using (50) are pretty good and omitted.

5.2 Bivariate normal model

In Section 2, $\hat{\theta}_t$ is defined to be the maximum likelihood estimate. It is, however, not assumed to be the *unique* MLE. To assess the performance of (50) when multiple MLEs exist, we consider the posterior distribution of the correlation coefficient in the bivariate normal data given by Murray (1977); see also Tanner and Wong (1987). The data set is in Table 1, where 12 observations are assumed to come from the bivariate normal distribution with $\mu_1 = \mu_2 = 0$, the correlation coefficient ρ, and variances σ_1^2 and σ_2^2. In this data, 2 pairs have correlation 1, 2 pairs have correlation -1, and there are 8 missing values. Denote the covariance matrix as Γ. As in Tanner and Wong (1987), we suppose

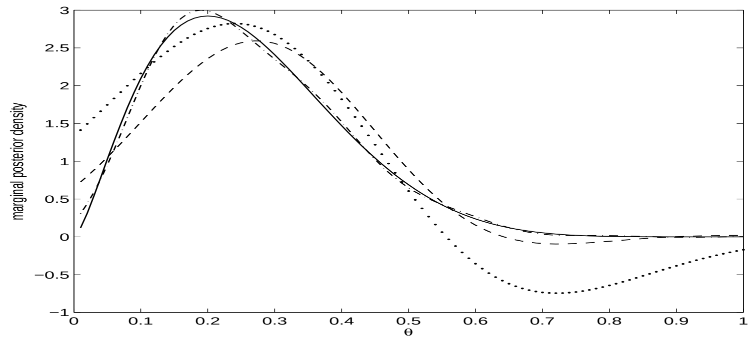

Figure 1: Bin(5,θ), $x = 2$. Marginal posterior pdf of θ. Solid: Exact distribution; Dashed: Approximation to $O(t^{-3/2})$; Dashed-Dotted: Approximation to $O(t^{-7})$; Dotted line: Johnson's approximation to $O(t^{-1})$.

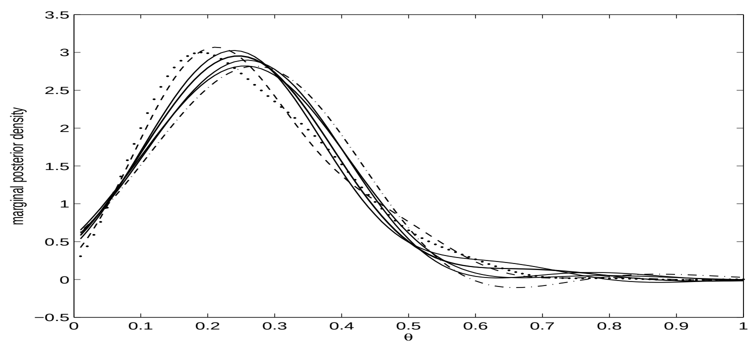

Figure 2: Bin(5,θ), $x = 2$. Marginal posterior pdf of θ. Dashed-Dotted: 5 moments; Solid: 6, 7, 8, 9, 10 moments; Dashed: 20 moments; Dotted: 40 moments.

that the prior of Γ is

$$\xi(\Gamma) \propto |\Gamma|^{-(k+1)/2},$$

where k is the dimension of the multivariate normal distribution.

The two MLEs of $\theta = (\sigma_1^2, \sigma_2^2, \rho)$ are (2.67,2.67,-0.5) and (2.67,2.67,0.5). We use the former as the $\hat{\theta}_t$ in our Z_t (15). In Figure 4 we plot the estimated posterior densities of ρ using (50) with $s = 2$ and 33 (the latter corresponds to about 100 moments of Z_{tp}, approximated by numerical integration). We also plot the true posterior density of ρ, which is proportional to $(1 - \rho^2)^{4.5}/(1.25 - \rho^2)^8$. The results show that the estimate using tens of moments performs nicely around the two modes, while an approximation to order $O(t^{-3/2})$ does not.

Finally, we tried approximations using (50) with 20, 40, 60, 80 moments of Z_{tp}. We found that the magnitude of oscillation decreases when more moments were used. The results are in Figure 5.

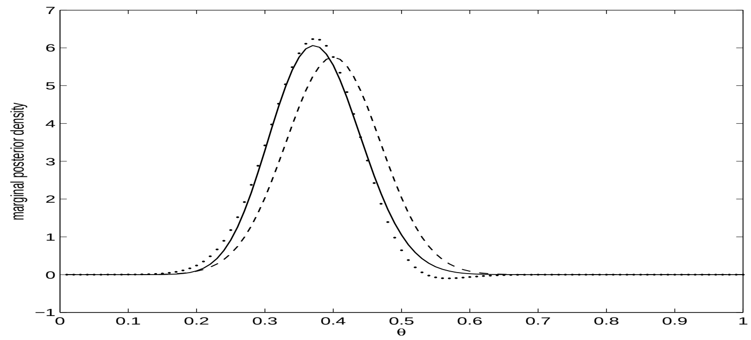

Figure 3: Bin(50,θ), $x = 20$. Marginal posterior pdf of θ. Solid: Exact distribution; Dashed: Normal approximation; Dotted: John's approximation to $O(t^{-1})$.

1	1	-1	-1	2	2	-2	-2	*	*	*	*
1	-1	1	-1	*	*	*	*	2	2	-2	-2

Table 1: Data from Bivariate Normal Distribution. (* indicates value not observed)

6 Concluding Remarks

We have obtained an Edgeworth expansion for marginal posterior densities. We have shown two examples where the incorporation of our expansion and numerical integration (for moments of Z_{tp}) produce reasonable approximations when the sample size is small or multiple modes are present.

It is worth mentioning that Z_t may be defined in different ways. For example, if Z_t is the signed-root transformation as in Bickel and Ghosh (1990), under certain regularity conditions the representation of the posterior expectation in (41) still holds. Then, together with Propositions 3 and 4 we can also obtain the expansion in the Hermite polynomials; that is, (44). However, with this new Z_t, the f_t in (18) will be different and the order of $E^t_\xi(q_k(Z_{ti}))$ needs to be re-examined.

Several questions deserve further study. First, the nonparametric density estimation has been a popular topic. It is not clear whether the results in the present paper can be extended for density estimators. One theoretical bottleneck for the extension would be whether the posterior density of the density estimation can be expressed in the form (18). Second, since the posterior expansion based on Taylor series can not be applied to the case of non-smooth priors, it is interesting to extend the current results to such problems. One possible starting point is to modify Stein's identity in Lemma 1 for piecewise smooth f. Third, we may use the expansions to validate convergence of simulation results. The idea is that if the posterior sample has converged to the true distribution, the density induced by the sample should agree with the one obtained by putting the empirical moments of the sample into (50). Finally, in the present paper

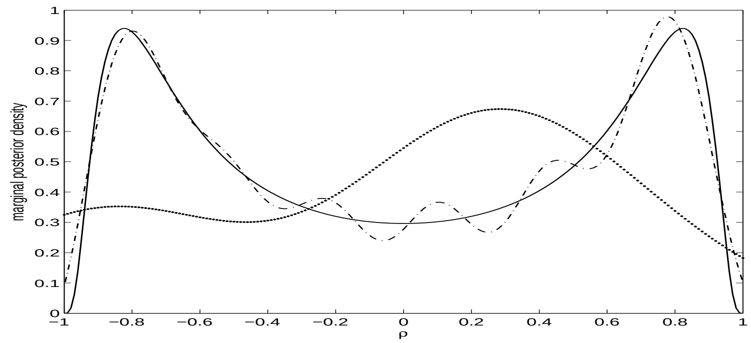

Figure 4: Marginal posterior pdf of ρ. Solid: Exact distribution; Dotted: Approximation to $O(t^{-3/2})$; Dashed-Dotted: Approximation using 100 moments.

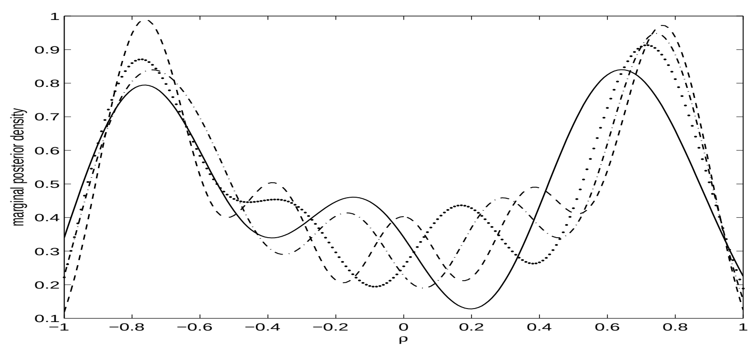

Figure 5: Marginal posterior pdf of ρ. Solid: 20 moments; Dotted: 40 moments; Dashed-Dotted: 60 moments; Dashed: 80 moments.

and Blinnikov and Moessner (1998), there is no guideline or methodologies for how many terms should be included in the expansion based on real data; the method for determining the order of expansion based on data may be a topic of future work.

Appendix

A: Proof of Proposition 1

We need one lemma.

Lemma 4. *Let Z denote a standard normal random variable. Then,*

$$\text{(a)}\qquad \sum_{i=1}^{n} C_i^n i q_{i-1}(x) E(Z^{n-i}) = \sum_{i=0}^{n} C_i^n q_i(x) E(Z^{n+1-i}), \tag{57}$$

$$\text{(b)}\qquad \sum_{i=0}^{n} C_i^{n+1} q_i(x) E(Z^{n+1-i}) = \sum_{i=0}^{n-1} C_i^n q_{i+1}(x) E(Z^{n-i}) + \sum_{i=0}^{n} C_i^n q_i(x) E(Z^{n+1-i}). \tag{58}$$

Proof. For (a), we need the fact that $E(Z^r) = (r-1)(r-3)\cdots(3)(1)$ if r is even and zero if r is odd. If n is even, there are $n/2$ nonzero terms on each side of (57). Let $m = 2j$. Then, the jth nonzero terms on left and right sides are respectively

$$C_m^n m q_{m-1} E(Z^{n-m}) = \frac{n(n-1)\cdots(n-m+1)}{m!} m q_{m-1}(n-m-1)(n-m-3)\cdots(3)(1)$$

and

$$\begin{aligned} & C_{m-1}^n q_{m-1} E(Z^{n-m+2}) \\ = \;& \frac{n(n-1)\cdots(n-m+2)}{(m-1)!} q_{m-1}(n-m+1)(n-m-1)(n-m-3)\cdots(3)(1), \end{aligned}$$

and they are equal. The proof for odd n is similar and we omit it.

For (b), we need the fact that $C_i^{n+1} = C_i^n + C_{i-1}^n$. So,

$$\begin{aligned} & \sum_{i=0}^{n} C_i^{n+1} q_i(x) E(Z^{n+1-i}) - \sum_{i=0}^{n} C_i^n q_i(x) E(Z^{n+1-i}) \\ = \;& \sum_{i=1}^{n} C_{i-1}^n q_i(x) E(Z^{n+1-i}) \\ = \;& \sum_{i=0}^{n-1} C_i^n q_{i+1}(x) E(Z^{n-i}). \end{aligned}$$

This completes the proof. □

Proof of Proposition 1. We shall prove (31) by induction. First, it is easily seen that (31) holds for $k = 1, 2$. Next, suppose that (31) holds for $k = n-1$ and n. Together with (28), we have

$$q_n(x) = x^n - \sum_{i=0}^{n-1} C_i^n q_i(x) E(Z^{n-i}) = x q_{n-1}(x) - (n-1) q_{n-2}(x).$$

Taking derivative in the equation above with respect to x gives

$$n x^{n-1} - \sum_{i=0}^{n-1} C_i^n q_i'(x) E(Z^{n-i}) = q_{n-1}(x) + x q_{n-1}'(x) - (n-1) q_{n-2}'(x);$$

and by (27) and (28) and some algebra we obtain

$$nx^{n-1} = \sum_{i=1}^{n} C_i^n i q_{i-1}(x) E(Z^{n-i}). \tag{59}$$

Now, by (28) and the fact that (31) holds for $k = n-1$ and n, it follows that

$$\begin{aligned} & q_{n+1}(x) = xq_n(x) - nq_{n-1}(x) \\ = \quad & x^{n+1} - \sum_{i=0}^{n-1} C_i^n x q_i(x) E(Z^{n-i}) - \Big[nx^{n-1} - n \sum_{i=0}^{n-2} C_i^{n-1} q_i(x) E(Z^{n-1-i}) \Big], \end{aligned} \tag{60}$$

where straightforward calculations and (28) give

$$\begin{aligned} & \sum_{i=0}^{n-1} C_i^n x q_i(x) E(Z^{n-i}) - n \sum_{i=0}^{n-2} C_i^{n-1} q_i(x) E(Z^{n-1-i}) \\ = \quad & xq_0(x) + \sum_{i=1}^{n-1} C_i^n x q_i(x) E(Z^{n-i}) - n \sum_{i=1}^{n-1} C_{i-1}^{n-1} q_{i-1}(x) E(Z^{n-i}) \\ = \quad & xq_0(x) + \sum_{i=1}^{n-1} C_i^n (xq_i(x) - iq_{i-1}(x)) E(Z^{n-i}) \\ = \quad & \sum_{i=0}^{n-1} C_i^n q_{i+1}(x) E(Z^{n-i}). \end{aligned}$$

Then, by (59) and Lemma 4(a), we can rewrite (60) as

$$\begin{aligned} q_{n+1}(x) \quad &= \quad x^{n+1} - \sum_{i=0}^{n} C_i^n q_i(x) E(Z^{n+1-i}) - \sum_{i=0}^{n-1} C_i^n q_{i+1}(x) E(Z^{n-i}) \\ &= \quad x^{n+1} - \sum_{i=0}^{n} C_i^{n+1} q_i(x) E(Z^{n+1-i}), \end{aligned}$$

where the last line follows by Lemma 4(b). Therefore, (31) holds for $k = n+1$. This completes the proof. □

B: Proofs of (36)-(38)

Since $f_t(z) = \xi(\theta(z)) \exp[u_t(\theta)]$, it is easily seen that

$$\frac{df_t(z)}{dz} = \Big(\frac{d\theta}{dz}\Big) f_t(z) \Big(\frac{\xi^{(1)}}{\xi} + u_t^{(1)}\Big) = \Big(\frac{d\theta}{dz}\Big) f_t(z) G_1(\theta).$$

So, (37) and (38) hold for G_1. Next, suppose that

$$\frac{d^{k-1} f_t(z)}{dz^{k-1}} = \Big(\frac{d\theta}{dz}\Big)^{k-1} f_t(z) G_{k-1}(\theta).$$

Then,

$$\begin{aligned}\frac{d^k f_t(z)}{dz^k} &= \left(\frac{d\theta}{dz}\right)^{k-1}\left[\frac{f_t(z)}{dz}G_{k-1}(\theta) + f_t(z)\frac{dG_{k-1}(\theta)}{d\theta}\frac{d\theta}{dz}\right] \\ &= \left(\frac{d\theta}{dz}\right)^{k} f_t(z)(G_1 G_{k-1} + G_{k-1}^{(1)}).\end{aligned}$$

Thus, we proved (36).

Now, we shall prove (37) and (38) by induction. Suppose that G_k is of the form (37) and (38) holds. It suffices to show that G_{k+1} also has these two properties. To start, write

$$G_{k+1} = G_1 G_k + G_k^{(1)}.$$

The first term on the right side is

$$G_1 G_k = \left(\frac{\xi^{(1)}}{\xi} + u_t^{(1)}\right)\sum_l c_{kl}\Big\{\Big(\prod_{i=1}^k [u_t^{(i)}]^{r_{ki}}\Big)\Big[\prod_{j=1}^k \Big(\frac{\xi^{(j)}}{\xi}\Big)^{s_{kj}}\Big]\Big\},$$

and the second term is $G_k^{(1)} = dG_k/d\theta$. So, G_{k+1} is of the form (37).

Then, we will show that (38) holds for G_{k+1}; that is,

$$\sum_{i=1}^{k+1}(ir_{k+1,i}) + \sum_{j=1}^{k+1}(js_{k+1,j}) = k+1. \tag{61}$$

As G_k is multiplied by the factor $\xi^{(1)}/\xi$, the power corresponding to this factor increases by 1 (that is, $s_{k+1,1} = s_{k1} + 1$), and the remaining powers are unchanged (that is, $r_{k+1,i} = r_{ki}$ $\forall i$ and $s_{k+1,j} = s_{kj}$ for $j \neq 1$); hence (61) holds for terms in $G_k(\xi^{(1)}/\xi)$. Similar arguments apply for terms in $G_k u_t^{(1)}$. Next, consider $G_k^{(1)}(= dG_k/d\theta)$. It involves differentiation of either $[u_t^{(i)}]^{r_{ki}}$ or $(\xi^{(j)}/\xi)^{s_{kj}}$ with respect to θ. Note that

$$\frac{d[u_t^{(i)}]^{r_{ki}}}{d\theta} = r_{ki}[u_t^{(i)}]^{r_{ki}-1}u_t^{(i+1)}.$$

So, $r_{k+1,i} = r_{ki} - 1$ and $r_{k+1,i+1} = r_{k,i+1} + 1$; and hence

$$ir_{k+1,i} + (i+1)r_{k+1,i+1} = ir_{ki} + (i+1)r_{k,i+1} + 1,$$

which satisfies (61). The treatment for $(d/d\theta)(\xi^{(j)}/\xi)^{s_{kj}}$ is similar and we omit it. This completes the proof. □

References

Bickel, P. and Ghosh, J. K. (1990). "A decomposition for the likelihood ratio statistic and the Bartlett correction–A Bayesian argument." *Ann. Statist.*, 18: 1070–1090. 742, 757

Blinnikov, S. and Moessner, R. (1998). "Expansions for nearly Gaussian distributions." *Astron. Astrophys. Suppl. Ser.*, 130: 193–205. 741, 742, 758

Boas, M. L. (2006). *Mathematical Methods in Physical Sciences*. New Jersey: John Wiley & Sons, Inc., 3rd edition. 747

Courant, R. and Hilbert, D. (1953). *Methods of Mathematical Physics*, volume 1. Interscience. 747

Cramér, H. (1957). *Mathematical Methods of Statistics*. Princeton: Princeton University Press. 742

Filho, R. G. B. and Rosenfeld, R. (2004). "Testing option pricing with the Edgeworth expansion." *Physica A*, 344: 484–490. 741

Ghosh, J. K., Sinha, B., and Joshi, S. (1982). "Expansions for posterior probability and integrated Bayes risk." In Gupta, S. and Berger, J. (eds.), *Statistical Decision Theory and Related Topics III*, volume 1, 403–456. New York: Acamedic. 754

Hall, P. (1992). *The Bootstrap and Edgeworth Expansion*. New York: Springer. 741

Jing, B.-Y. and Wang, Q. (2003). "Edgeworth expansion for U-statistics under minimal conditions." *Ann. Statist.*, 31(4): 1376–1391. 742

Johnson, R. (1967). "An asymptotic expansion for posterior distributions." *Ann. Math. Statist.*, 38: 1899–1906. 742

— (1970). "Asymptotic expansions associated with posterior distributions." *Ann. Math. Statist.*, 41: 851–864. 742, 743, 753, 754

Lindley, D. V. (1961). "The use of prior probability distributions in statistical inference and decisions." *Proc. 4th. Berkeley Symp.*, 1: 453–468. 742

— (1980). "Approximate Bayesian methods." In Bernardo, J. M., DeGroot, M. H., Lindley, D. V., and (Eds.), A. F. M. S. (eds.), *Bayesian Statistics*. University Press. 742

Mosteller, F. and Wallace, D. L. (1964). *Inference and Disputed Authorship: The Federalist Papers*. Reading, Mass.: Addison-Wesley. 742

Murray, G. D. (1977). "Comment on "Maximum likelihood from incomplete data via the EM algorithm" by A. P. Dempster, N. M. Laird and D. B. Rubin." *Journal of the Royal Statistical Society, Ser. B*, 39: 27–28. 755

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org`. 755

Stein, C. (1981). "Estimation of the mean of a multivariate normal distribution." *Ann. Statist.*, 9: 1135–1151. 743

— (1987). *Approximate Computation of Expectations*. Hayward, Calif: IMS. 743

Tanner, M. A. and Wong, W. H. (1987). "The calculation of posterior distributions by data augmentation." *Journal of the American Statistical Association*, 82: 528–540. 755

Tierney, L. and Kadane, J. B. (1986). "Accurate approximations for posterior moments and marginal densities." *Journal of the American Statistical Association*, 81: 82–86. 743

Wallace, D. L. (1958). "Asymptotic approximations to distributions." *Annals of Mathematical Statistics*, 29: 635–654. 741

Weber, H. J. and Arfken, G. B. (2004). *Essential Mathematical Methods for Physicists*. San Diego: Elsevier Academic Press. 747

Weng, R. C. (2003). "On Stein's Identity for posterior normality." *Statistica Sinica*, 13: 495–506. 743

Weng, R. C. and Lin, C.-J. (2010). "A Bayesian approximation method for online ranking." *Revision invited by Journal of Machine Learning Research*. 743

Weng, R. C. and Tsai, W.-C. (2008). "Asymptotic posterior normality for multiparameter problems." *Journal of Statistical Planning and Inference*, 138: 4068–4080. 743

Weng, R. C. and Woodroofe, M. (2000). "Integrable expansions for posterior distributions for multiparameter exponential families with applications to sequential confidence levels." *Statistica Sinica*, 10: 693–713. 744, 745

Woodroofe, M. (1989). "Very weak expansions for sequentially designed experiments: linear models." *The Annals of Statistics*, 17: 1087–1102. 743, 744

— (1992). "Integrable expansions for posterior distributions for one-parameter exponential families." *Statistica Sinica*, 2: 91–111. 743

Woodroofe, M. and Coad, D. S. (1997). "Corrected confidence sets for sequentially designed experiments." *Statistica Sinica*, 7: 53–74. 745

Acknowledgments

The author would like to thank the referees for their valuable comments on the paper. The author is partially supported by the National Science Council of Taiwan.