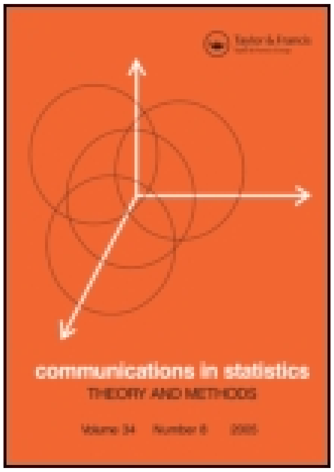

Communications in Statistics - Theory and Methods
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/lsta20

A Study of Expansions of Posterior Distributions

Ruby C. Weng [a] & Cheng-Hung Hsu [a]
[a] National Chenghai University , Taipei , Taiwan

Published online: 10 Dec 2012.

Communications in Statistics—Theory and Methods, 42: 346–364, 2013

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610926.2011.579701

A Study of Expansions of Posterior Distributions

RUBY C. WENG AND CHENG-HUNG HSU

National Chenghai University, Taipei, Taiwan

Johnson (1970) obtained expansions for marginal posterior distributions through Taylor expansions. Here, the posterior expansion is expressed in terms of the likelihood and the prior together with their derivatives. Recently, Weng (2010) used a version of Stein's identity to derive a Bayesian Edgeworth expansion, expressed by posterior moments. Since the pivots used in these two articles are the same, it is of interest to compare these two expansions. We found that our $O(t^{-1/2})$ term agrees with Johnson's arithmetically, but the $O(t^{-1})$ term does not. The simulations confirmed this finding and revealed that our $O(t^{-1})$ term gives better performance than Johnson's.

Keywords Edgeworth expansion; Marginal posterior densities; Stein's identity.

Mathematics Subject Classification 41; 62.

1. Introduction

Let $g(\theta)$ be a smooth function on the parameter space Θ. The calculation of the posterior mean of $g(\theta)$, given a sample of observations x_t, requires integration over Θ of the form

$$E_\xi^t[g(\theta)] = E_\xi[g(\theta)|x_t] = \frac{\int_\Theta g(\theta)\exp(\ell_t(\theta))\xi(\theta)d\theta}{\int_\Theta \exp(\ell_t(\theta))\xi(\theta)d\theta}, \tag{1}$$

where ℓ_t is the log-likelihood function and ξ the prior. Nowadays, modern computing techniques like Markov chain Monte Carlo and importance sampling have made many computations possible. Still, analytic approximations are simpler to compute for some models, and are useful as a starting point for more exact methods.

A conventional analytic approach to this problem (1) starts from a Taylor series expansion at the maximum likelihood estimator (or at the modes of the integrands), proceeds from there to develop expansions on both the numerator and denominator, and then obtains approximations by formal division of the two series. For example, Johnson (1967, 1970) derived expansions associated with posterior distribution of some pivotal quantity Lindley (1961, 1980) and Mosteller and

Received March 15, 2010; Accepted April 15, 2011
Address correspondence to Ruby C. Weng, Department of Statistics, National Chengchi University, Taipei, Taiwan; E-mail: chweng@nccu.edu.tw

Wallace (1964) obtained second-order approximations for the integral by applying standard Laplace method to both numerator and denominator and taking the ratio. Tierney and Kadane (1986) renewed interest in Laplace method by assuming that g is positive and expanding the integrand of the numerator in (1) at the mode of the integrand itself, rather than at the posterior mode. In fact, this is a numerical method, not requiring analytic expansions. This work was followed by Tierney et al. (1989), who proposed a numerical device to approximate the posterior expectation for a general function g (possibly non positive) and showed that this numerical approach is arithmetic equivalent to second order expansions by the standard Laplace method.

Recently, Weng (2010) used a version of Stein's identity to derive an Edgeworth series for posterior distribution of a normalized quantity Z_t. Note that an Edgeworth series is an expansion of a probability distribution in terms of its moments. In contrast, the expansion in Johnson (1970) is expressed in terms of the likelihood and the prior. Since the pivots used in these two articles are the same, it is of interest to compare these two expansions. The goals of the present article are to suitably approximate these posterior moments in terms of the likelihood and the prior, and then substitute these approximations into the Edgeworth series and compare with the result in Johnson (1970). We found that our $O(t^{-1/2})$ term is arithmetically equivalent to Johnson's, but the $O(t^{-1})$ term is not. Since the derivation is tedious and difficult to detect errors, we conducted simulation studies to further compare these expansions. The simulations confirmed that the two expressions for $O(t^{-1/2})$ term yield close results, and revealed that our $O(t^{-1})$ term gives better performance than Johnson's. Note that the emphasis here is on comparison of the two expansions, rather than the regularity conditions for the expansions.

Section 2 introduces the model, and Stein's Identity. Section 3 reviews expansions of posterior distributions in Johnson (1970) and Weng (2010). Section 4 presents approximations for posterior moments of Z_t and compare the expansions in these two articles. Section 5 gives some remarks. The appendices contain some proofs.

2. The Model and Stein's Identity

2.1. *The Model*

Let X_t be a random vector distributed according to a family of probability densities $p_t(x_t \mid \theta)$, where t is a discrete or continuous parameter and $\theta \in \Theta \subset \Re^p$. Consider a Bayesian model in which θ has a prior density ξ which is twice differentiable in $\Re^p$ and vanishes off of Θ. Assume that the log-likelihood function $\ell_t(\theta)$ is twice differentiable with respect to θ. Assume that maximum likelihood estimator $\hat{\theta}_t$ exists and satisfies $\nabla\ell_t(\hat{\theta}_t) = 0$ and $-\nabla^2\ell_t(\hat{\theta}_t)$ being positive definite, where ∇ indicates differentiation with respect to θ.

Throughout 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. Define Σ_t and Z_t as

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

$$Z_t = \Sigma_t(\theta - \hat{\theta}_t). \tag{3}$$

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

$$\zeta_t(z_t) \propto \xi_t(\theta(z_t)) \propto \exp[\ell_t(\theta) - \ell_t(\hat{\theta}_t)]\xi(\theta), \tag{4}$$

where the relation of θ and z_t is given in (3). Now define

$$u_t(\theta) = \ell_t(\theta) - \ell_t(\hat{\theta}_t) + \frac{1}{2}\|z_t\|^2. \tag{5}$$

So, (4) can be rewritten as

$$\zeta_t(z_t) \propto \phi_p(z_t)f_t(z_t), \tag{6}$$

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

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

$$\frac{\nabla f_t(Z_t)}{f_t(Z_t)} = (\Sigma_t^T)^{-1}\left[\frac{\nabla\xi(\theta)}{\xi(\theta)} + \nabla u_t(\theta)\right], \tag{7}$$

$$\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)} = (\Sigma_t^T)^{-1}\left[\frac{\nabla^2\xi}{\xi} + \frac{\nabla\xi}{\xi}\nabla u_t^T + \nabla u_t\frac{\nabla\xi^T}{\xi} + \nabla^2 u_t + \nabla u_t \nabla u_t^T\right]\Sigma_t^{-1}, \tag{8}$$

where by (5) we can derive

$$\nabla u_t(\theta) = \nabla\ell_t(\theta) - \nabla^2\ell_t(\hat{\theta}_t)(\theta - \hat{\theta}_t), \tag{9}$$

$$\nabla^2 u_t(\theta) = \nabla^2\ell_t(\theta) - \nabla^2\ell_t(\hat{\theta}_t). \tag{10}$$

2.2. *Stein's Identity*

Write $\Phi_p h = \int h d\Phi_p$ for functions h for which the integral is finite. 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 $h_0 = \Phi_p h$, $h_p = h$,

$$h_k(y_1, \ldots, y_k) = \int_{\Re^{p-k}} h(y_1, \ldots, y_k, w)\Phi_{p-k}(dw), \tag{11}$$

$$g_k(y_1, \ldots, y_p) = e^{\frac{1}{2}y_k^2}\int_{y_k}^{\infty} [h_k(y_1, \ldots, y_{k-1}, w) - h_{k-1}(y_1, \ldots, y_{k-1})]e^{-\frac{1}{2}w^2}dw, \tag{12}$$

for $-\infty < y_1, \ldots, y_p < \infty$ and $k = 1, \ldots, p$. Then let $Uh = (g_1, \ldots, g_p)^T$ and $Vh = (U^2h + U^2h^T)/2$, where U^2h is the $p \times p$ matrix whose kth column is Ug_k and g_k is as in (12). For example, for $z = (z_1, \ldots, z_p)^T \in \Re^p$, if $h(z) = z_1$, then $Uh(z) = (1, 0, \ldots, 0)^T$. Simple calculations by taking $f(z)$ in Lemma 2.1 below as z_i

$$\Phi_p(Uh) = \int_{\Re^p} zh(z)\Phi_p(dz). \tag{13}$$

Lemma 2.1 (Stein's Identity). *Let r be a nonnegative integer. Suppose that $f : \Re^p \to \Re$ is continuously differentiable on $\Re^p$, and*

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

Then,

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

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

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

then

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

for all $h \in H_r$.

The proof of Lemma 2.1 is in Woodroofe (1989, Proposition 1); see also Weng and Woodroofe (2000). Observe from (6) that the posterior distribution of Z_t is of a form suitable for Stein's Identity. So, by Lemma 2.1,

$$E_\xi^t\{h(Z_t)\} = \Phi_p h + E_\xi^t\left\{[Uh(Z_t)]^T \frac{\nabla f_t(Z_t)}{f_t(Z_t)}\right\}, \tag{14}$$

$$E_\xi^t\{h(Z_t)\} = \Phi_p h + (\Phi_p Uh)^T E_\xi^t\left[\frac{\nabla f_t(Z_t)}{f_t(Z_t)}\right] + E_\xi^t\left\{\mathrm{tr}\left[Vh(Z_t)\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}\right]\right\}. \tag{15}$$

Moreover, if $h(z) : \Re^p \to \Re$ is a function of z_p alone (that is, $h(z) = h^*(z_p)$, where $h^* : \Re \to \Re$), then by repeatedly applying this identity, Weng (2010) obtains the following equation:

$$\Phi(fh) = \Phi f \cdot \Phi h^* + \sum_{k=1}^{s-1}(\Phi U^k h^*)\int_{\Re^p} \frac{\partial^k f(z)}{\partial z_p^k}\Phi_p(dz) + \int_{\Re^p} U^s h^*(z)\frac{\partial^s f(z)}{\partial z_p^s}\Phi_p(dz),$$

provided all the integrals exist. Applying this equation to posterior distributions yield

$$E_\xi^t(h^*(Z_{tp})) = \Phi h^* + \sum_{k=1}^{s-1}(\Phi U^k h^*)E_\xi^t\left[\frac{\partial^k f_t/\partial z_{tp}^k}{f_t}(Z_t)\right] + E_\xi^t\left\{[U^s h^*(Z_{tp})]\frac{\partial^s f_t/\partial z_{tp}^s}{f_t}(Z_t)\right\}. \tag{16}$$

Equation (16) will be used in the proofs of Lemmas 4.7 and 4.9 below to obtain expressions for $E_\xi^t Z_{tp}^4$ and $E_\xi^t Z_{tp}^6$.

3. Review of Posterior Expansions

Johnson (1970) considered the posterior distribution of a centered and scaled variable (see his Eq. (2.1), p. 853) in 1-dimensional case:

$$\psi = (\theta - \hat{\theta}_t)b(\hat{\theta}_t), \tag{17}$$

where t is the sample size and

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

Denote the posterior cumulative distribution function (cdf) of $t^{1/2}\psi$ by F_t. He showed that the posterior distribution of F_t possesses an asymptotic expansion in powers of $t^{-1/2}$ (see his Theorem 2.1):

$$|F_t(w) - \Phi(w) - \sum_{j=1}^{K} \gamma_j(w, x)t^{-j/2}| \le D_1 t^{-\frac{1}{2}(K+1)}. \tag{18}$$

The forms of γ_1 and γ_2 are given in Johnson's is Sec. 2.4 (see Eq. (2.25) and (2.26), p. 858):

$$\gamma_1(w, x) = -\phi(w)c_{00}^{-1}[c_{10}(w^2+2) + c_{01}], \tag{19}$$

$$\gamma_2(w, x) = -\phi(w)c_{00}^{-1}[c_{20}w^5 + (5c_{20} + c_{11})w^3 + (15c_{20} + 3c_{11} + c_{02})w]. \tag{20}$$

Here, the c_{lm} involves the prior ξ and the likelihood together with their derivatives:

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

where

$$a_{kt}(\theta) = \frac{1}{t}\left(\frac{1}{k!}\right)\sum_{i=1}^{t} \frac{\partial^k}{\partial\theta^k} \log p(x_i, \theta). \tag{21}$$

To compare with Weng (2010), let $Z_t = (Z_{t1}, \dots, Z_{tp})^T$ be as in (3). Define $J_1 = \{1, 3\}$ and $J_i = \{3i-4, 3i-2, 3i\}$ for $i > 1$; for example, $J_2 = \{2, 4, 6\}$, $J_3 = \{5, 7, 9\}$. Weng (2010, Sec. 3) showed that the marginal posterior distribution of Z_{tp} can be expanded as

$$P_\xi^t(Z_{tp} \le w) = \Phi(w) + \sum_{i=1}^{m} R_{it}(w)\phi(w) + O\left(t^{-\frac{m+1}{2}}\right), \tag{22}$$

where

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

with q_i being Hermite polynomials, given by $q_k(z)\phi(z) = (-d/dz)^k\phi(z)$. For instance, for $k = 0, 1, 2, 3$ we have $q_0(z) = 1$, $q_1(z) = z$, $q_2(z) = z^2 - 1$, $q_3(z) = z^3 - 3z$. Moreover, the marginal posterior density for θ_p is

$$\xi_p^t(a) = [\Sigma_t]_{pp}\left\{\phi(w) + \sum_{i=1}^{m} Q_{it}(w)\phi(w) + O(t^{-\frac{m+1}{2}})\right\}, \tag{23}$$

where

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

In particular, if $m = 2$, then we have

$$\left|P_\xi^t(Z_{tp} \le w) - \Phi(w) - \phi(w)[R_{1t}(w) + R_{2t}(w)]\right| = O(t^{-3/2}), \tag{24}$$

where

$$R_{1t}(w) = q_0(w)E_\xi^t(q_1(Z_{tp})) + \frac{1}{3!}q_2(w)E_\xi^t(q_3(Z_{tp})) = O(t^{-1/2}), \tag{25}$$

$$\begin{aligned} R_{2t}(w) &= \frac{1}{2!}q_1(w)E_\xi^t(q_2(Z_{tp})) + \frac{1}{4!}q_3(w)E_\xi^t(q_4(Z_{tp})) + \frac{1}{6!}q_5(w)E_\xi^t(q_6(Z_{tp})) \\ &= O(t^{-1}). \end{aligned} \tag{26}$$

Since the normalized quantity Z_t in (3) is the multivariate version of ψ in (17), it is of interest to compare the $O(t^{-1/2})$ and $O(t^{-1})$ terms in (18) and (24); that is, the terms $-\phi R_{1t}, -\phi R_{2t}$ in (24) and $\gamma_1 t^{-1/2}$, $\gamma_2 t^{-1}$ in (18). To proceed further, first one needs to approximate the moments $E_\xi^t(q_k(Z_{tp}))$, $k = 1, 2, 3, 4, 6$ in terms of the likelihood and the prior. Then, we plug these approximations into (25) and (26) and compare with (19) and (20). The results are in the next section.

4. Main Results

We shall first obtain approximations of posterior moments of Z_t. Some notations are needed. First, denote $\ell_t^{(k)}$ and $\hat{\ell}_t^{(k)}$ as the kth partial derivative and its value at $\hat{\theta}_t$; and denote as $\ell_{i_1\cdots i_k}^{(k)}$ and $\hat{\ell}_{i_1\cdots i_k}^{(k)}$ to emphasize that the derivatives are with respect to $\theta_{i_1}, \ldots, \theta_{i_k}$. We denote similarly for derivatives of ξ. Then, for a given matrix A, its ith row is denoted as $[A]_{i\cdot}$ and its (i, j)-component is written as $[A]_{ij}$. Some matrices and vectors involving higher order derivatives of ℓ_t are needed. We denote minus the inverse Hessian of ℓ_t at $\hat{\theta}_t$ as either $(-\nabla^2\hat{\ell}_t)^{-1}$ or H. For $k, l, i, j = 1, \ldots, p$, let D_k and D_{kl} denote the $p \times p$ matrices with $[D_k]_{ij} = \hat{\ell}_{kij}^{(3)}$ and $[D_{kl}]_{ij} = \hat{\ell}_{klij}^{(4)}$. Moreover, define $V_k = (\Sigma_t^T)^{-1}D_k\Sigma_t^{-1}$. It is easy to see that H, D_k, D_{kl}, and V_k are all symmetric matrices. Now define

$$S = (\mathrm{tr}(V_1), \ldots, \mathrm{tr}(V_p))^T = (\mathrm{tr}(D_1H), \cdots, \mathrm{tr}(D_pH))^T, \tag{27}$$

$$W_{ij} = [D_i]_{j\cdot}H\left(\frac{\nabla\hat{\xi}}{\hat{\xi}} + \frac{1}{2}S\right) + \frac{1}{2}\mathrm{tr}(HD_iHD_j) + \frac{1}{2}\mathrm{tr}(D_{ij}H). \tag{28}$$

Note that for simplicity of notation, the dependency of these matrices and vectors on t will be suppressed when this leads to no ambiguity. When $p = 1$, they have simpler forms:

$$D_k = \hat{\ell}_t^{(3)}, \quad D_{kl} = \hat{\ell}_t^{(4)}, \quad H = (-\hat{\ell}_t^{(2)})^{-1},$$
$$S = V_k = \frac{\hat{\ell}_t^{(3)}}{-\hat{\ell}_t^{(2)}}, \quad W = \left(\frac{\hat{\ell}_t^{(3)}}{-\hat{\ell}_t^{(2)}}\right)\left(\frac{\hat{\xi}^{(1)}}{\hat{\xi}}\right) + \left(\frac{\hat{\ell}_t^{(3)}}{-\hat{\ell}_t^{(2)}}\right)^2 + \frac{1}{2}\left(\frac{\hat{\ell}_t^{(4)}}{-\hat{\ell}_t^{(2)}}\right), \tag{29}$$

4.1. *Moments of Z_t*

Recall that X_t is a random vector from $p_t(x_t \mid \theta)$, where θ is chosen according to the prior density ξ. Let θ_0 denote the true underlying parameter. The lemma below is well known under some regularity conditions and we state it here for later use. The proof is in, for instance, Johnson (1970).

Lemma 4.2. *Let $M_t(r; r_1, \ldots, r_p)$ denote rth joint posterior moments of Z_t with $0 < r \leq 6$; that is, $M_t(r; r_1, \ldots, r_p) = E_\xi^t h(Z_t)$, where $h(z) = \prod_{i=1}^p z_i^{r_i}$ with $\sum r_i = r$. Then*

(i) $E_\xi^t h(Z_t) = O(t^{-1/2})$ *for odd r*;
(ii) $E_\xi^t h(Z_t) = \Phi h + O(t^{-1})$ *for even r*.

Next we refine approximations for the first two moments of Z_t. Remember that 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, (14) and (15) give

$$E_\xi^t Z_t = E_\xi^t\left(\frac{\nabla f_t(Z_t)}{f_t(Z_t)}\right), \tag{30}$$

$$E_\xi^t(Z_{ti} Z_{tj}) = \delta_{ij} + E_\xi^t\left[\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}\right]_{ij}. \tag{31}$$

Note that if ξ is smooth, then by Lemma 4.2(i) we have

$$E_\xi^t\left(\frac{\nabla\xi}{\xi}\right) = \frac{\nabla\hat{\xi}}{\hat{\xi}} + O(t^{-1}) \quad \text{and} \quad E_\xi^t\left(\frac{\nabla^2\xi}{\xi}\right) = \frac{\nabla^2\hat{\xi}}{\hat{\xi}} + O(t^{-1}). \tag{32}$$

The proofs of the next two results are in Appendices A.1 and A.2, respectively.

Lemma 4.3. *Let ∇u_t and $\nabla^2 u_t$ be as in* (9) *and* (10). *Then*:

(i) $E_\xi^t[\nabla u_t(\theta)] = \frac{1}{2}S + O(t^{-1})$;
(ii) $E_\xi^t[\nabla^2 u_t(\theta)]_{ij} = \frac{1}{2}[D_i]_j.H(\frac{\nabla\hat{\xi}}{\hat{\xi}} + \frac{1}{2}S) + \frac{1}{2}\mathrm{tr}(D_{ij}H) + O(t^{-1})$;
(iii) $[E_\xi^t(\nabla u_t \nabla u_t^T)]_{ij} - [E_\xi^t(\nabla u_t)E_\xi^t(\nabla u_t^T)]_{ij} = \frac{1}{2}\mathrm{tr}(HD_kHD_l) + O(t^{-1})$.

Lemma 4.4. *Let S and W be as in* (27) *and* (28). *Then*:

(i) $E_\xi^t Z_t = (\Sigma_t^T)^{-1}(\frac{\nabla\hat{\xi}}{\hat{\xi}} + \frac{1}{2}S) + O(t^{-3/2})$;
(ii) $V_\xi^t Z_t = I_p + (\Sigma_t^T)^{-1}[(\frac{\nabla^2\hat{\xi}}{\hat{\xi}}) - (\frac{\nabla\hat{\xi}}{\hat{\xi}})(\frac{\nabla\hat{\xi}^T}{\hat{\xi}}) + W]\Sigma_t^{-1} + O(t^{-2})$.

Approximations to some higher order posterior moments of Z_t are also required, and are given in Lemmas 4.5–4.9. Let $\{e_1, \ldots, e_p\}$ denote the standard orthonormal basis of $\Re^p$.

Lemma 4.5. *Let* $1 \le i < p$ *and* $1 \le s, l \le p$.

(i) *If* $h(z) = z_p^3$, *then* $Uh(z) = (z_p^2 + 2)e_p$ *and* $[Vh(z)]_{sl} = z_p 1_{\{(s,l)=(p,p)\}}$.
(ii) *If* $h(z) = z_i z_p^2$, *then* $Uh(z) = e_i + z_i z_p e_p$ *and* $[Vh(z)]_{sl} = z_i 1_{\{(s,l)=(p,p)\}}$.
(iii) $E_\xi^t Z_{tp}^3 = 3E_\xi^t Z_{tp} + E_\xi^t(Z_{tp}[\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}]_{pp})$.
(iv) $E_\xi^t(Z_{tp}^2 Z_{ti}) = E_\xi^t Z_{ti} + E_\xi^t(Z_{ti}[\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}]_{pp})$.

Proof. (i) and (ii) follow from (11) and (12); (iii) and (iv) follow from (i), (ii), and (15). □

The proof of Lemma 4.6 is in Appendix A.3. With Lemmas 4.5(iii) and 4.6 we can express $E_\xi^t(q_3(Z_{tp}))$ in terms of the likelihood and prior.

Lemma 4.6. *Let* $1 \le i, j, k \le p$. *Then,*

$$E_\xi^t\left(Z_{ti}\left[\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}\right]_{jk}\right) = \sum_{l=1}^{p}\{[\Sigma_t^{-1}]_{li}[V_l]_{jk}\} + O(t^{-3/2}).$$

The proof of Lemma 4.7 below is in Appendix A.4.

Lemma 4.7. *Let Q be a p-dimensional vector defined by* $Q_r = \mathrm{tr}(V_r) + 2[V_r]_{pp}$; *and let* J, B_1, B_2, *and* A *be* $p \times p$ *matrices defined by*

$$\begin{aligned}
[J]_{rs} &= [\Sigma_t^{-1}]_{rp}[\Sigma_t^{-1}]_{sp},\\
[B_1]_{rs} &= [D_r]_{s\cdot}\Sigma_t^{-1}E_\xi^t(Z_{tp}^2 Z_t) + \frac{1}{2}\mathrm{tr}(D_{rs}H) + \mathrm{tr}(D_{rs}J),\\
[B_2]_{rs} &= \frac{1}{2}\mathrm{tr}(V_r V_s) + \frac{1}{4}\mathrm{tr}(V_r)\mathrm{tr}(V_s) + \frac{1}{2}[V_r]_{pp}\mathrm{tr}(V_s)\\
&\quad + \frac{1}{2}[V_s]_{pp}\mathrm{tr}(V_r) + 2\sum_{i=1}^{p}([V_r]_{ip}[V_s]_{ip}),\\
A &= (\Sigma_t^T)^{-1}\left[\frac{\nabla^2\hat{\xi}}{\hat{\xi}} + \frac{1}{2}\left(\frac{\nabla\hat{\xi}}{\hat{\xi}}Q^T + Q\frac{\nabla\hat{\xi}^T}{\hat{\xi}}\right) + B_1 + B_2\right]\Sigma_t^{-1}.
\end{aligned}$$

Then,

$$E_\xi^t Z_{tp}^4 = 3 + [A]_{pp} + 5E_\xi^t(Z_{tp}^2 - 1) + O(t^{-2}). \tag{33}$$

Note that the term $E_\xi^t(Z_{tp}^2 Z_t)$ (a $p \times 1$ vector) in B_1 can be approximated using Lemma 4.5(iii)–(iv) and Lemma 4.6.

The next two lemmas are necessary for approximating $E_\xi^t Z_{tp}^6$. The term B_1^* in Lemma 4.9 requires $E_\xi^t(Z_{tp}^4 Z_t)$ (a $p \times 1$ vector), which by Lemma 4.8 can be approximated using Lemma 4.5(iii)(iv) and Lemma 4.6. Since the proofs of Lemmas 4.8 and 4.9 are similar to that of Lemmas 4.5 and 4.7, we omit it.

Lemma 4.8. *Let* $1 \le i < p$. *Then*:

$$E_\xi^t(Z_{tp}^4 Z_{ti}) = 3E_\xi^t(Z_{ti}) + 6E_\xi^t\left(Z_{ti}\left[\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}\right]_{pp}\right) + O(t^{-3/2}),$$

$$E_\xi^t(Z_{tp}^5) = 15E_\xi^t(Z_{tp}) + 10E_\xi^t\left(Z_{tp}\left[\frac{\nabla^2 f_t(Z_t)}{f_t(Z_t)}\right]_{pp}\right) + O(t^{-3/2}).$$

Lemma 4.9. *Let* Q^* *be a p-dimensional vector defined by* $Q_r^* = 3\mathrm{tr}(V_r) + 12[V_r]_{pp}$; *and let* J, B_1^*, B_2^*, *and* A^* *be* $p \times p$ *matrices defined by*

$$\begin{aligned}
[J]_{rs} &= [\Sigma_t^{-1}]_{rp}[\Sigma_t^{-1}]_{sp},\\
[B_1^*]_{rs} &= [D_r]_{s\cdot}\Sigma_t^{-1}E_\xi^t(Z_{tp}^4 Z_t) + \frac{3}{2}\mathrm{tr}(D_{rs}H) + 6\mathrm{tr}(D_{rs}J),\\
[B_2^*]_{rs} &= \frac{3}{2}\mathrm{tr}(V_rV_s) + \frac{3}{4}\mathrm{tr}(V_r)\mathrm{tr}(V_s) + 3[V_r]_{pp}\mathrm{tr}(V_s)\\
&\quad + 3[V_s]_{pp}\mathrm{tr}(V_r) + 12\sum_{i=1}^{p}([V_r]_{ip}[V_s]_{ip}) + 6[V_r]_{pp}[V_s]_{pp},\\
A^* &= (\Sigma_t^T)^{-1}\left[3\frac{\nabla^2\hat{\xi}}{\hat{\xi}} + \frac{1}{2}\left(\frac{\nabla\hat{\xi}}{\hat{\xi}}(Q^*)^T + Q^*\frac{\nabla\hat{\xi}^T}{\hat{\xi}}\right) + B_1^* + B_2^*\right]\Sigma_t^{-1}.
\end{aligned}$$

Then,

$$E_\xi^t Z_{tp}^6 = 15 + [A^*]_{pp} + 9[A]_{pp} + 33E_\xi^t(Z_{tp}^2 - 1) + O(t^{-2}). \tag{34}$$

4.2. *Comparing γ_i and r_{it}*

Since the notation in Johnson (1970) is 1-dimensional, we take $p = 1$ and so our Z_t is 1-dimensional. When specializing Lemmas 4.4(i), 4.5, and 4.6 to $p = 1$, we have

$$E_\xi^t(q_1(Z_t)) = E_\xi^t Z_t = \left((-\hat{\ell}_t^{(2)})^{-\frac{1}{2}}\right)\left(\frac{\hat{\xi}^{(1)}}{\hat{\xi}} + \frac{1}{2}\frac{\hat{\ell}_t^{(3)}}{(-\hat{\ell}_t^{(2)})}\right) + O(t^{-\frac{3}{2}}),$$

$$E_\xi^t(q_3(Z_t)) = E_\xi^t Z_t^3 - 3E_\xi^t Z_t = \left((-\hat{\ell}_t^{(2)})^{-\frac{3}{2}}\right)\hat{\ell}_t^{(3)} + O(t^{-\frac{3}{2}}).$$

Plugging these approximations into (25) gives

$$\begin{aligned}
R_{1t}(w) &= q_0(w)E_\xi^t(q_1(Z_{tp})) + \frac{1}{3!}q_2(w)E_\xi^t(q_3(Z_{tp}))\\
&= \left((-\hat{\ell}_t^{(2)})^{-\frac{1}{2}}\right)\frac{\hat{\xi}^{(1)}}{\hat{\xi}} + \frac{w^2+2}{6}\left((-\hat{\ell}_t^{(2)})^{-\frac{3}{2}}\right)\hat{\ell}_t^{(3)}.
\end{aligned}$$

On the other hand, $\gamma_1(w, x)$ in (19) can be written as

$$\gamma_1(w, x) = -\phi(w)\left((w^2+2)b^{-3}a_{3t}(\hat{\theta}) + b^{-1}\frac{\hat{\xi}^{(1)}}{\hat{\xi}}\right),$$

noting that $a_{kt}(\theta)$ in (21) is $(1/t)(1/k!)\ell_t^{(k)}(\theta)$ in our notation. Therefore, the $O(t^{-1/2})$ terms in both expansions (i.e., $-\phi(w)R_{1t}(w)$ in (24) and $\gamma_1(w,x)t^{-1/2}$ in (18)) are equivalent.

The comparison of the $O(t^{-1})$ terms can be done similarly. First, we obtain $E_\xi^t(q_k(Z_t))$, $k=2,4,6$ from Lemmas 4.4(ii), 4.7, 4.9. For example, by Lemma 4.4 and S and W in (29) we have:

$$\begin{aligned} E_\xi^t(q_2(Z_t)) &= E_\xi^t(Z_t^2-1) = (E_\xi^t Z_t)^2 + V_\xi^t Z_t - 1 \\ &= (-\hat{\ell}_t^{(2)})^{-1}\left[\frac{\hat{\xi}^{(2)}}{\hat{\xi}} + 2\frac{\hat{\xi}^{(1)}}{\hat{\xi}}\left(\frac{\hat{\ell}_t^{(3)}}{-\hat{\ell}_t^{(2)}}\right) + \frac{5}{4}\left(\frac{\hat{\ell}_t^{(3)}}{-\hat{\ell}_t^{(2)}}\right)^2 + \frac{1}{2}\left(\frac{\hat{\ell}_t^{(4)}}{-\hat{\ell}_t^{(2)}}\right)\right] \\ &\quad + O(t^{-2}), \end{aligned} \tag{35}$$

where the leading terms are $O(t^{-1})$; and simple algebra gives

$$E_\xi^t(q_4(Z_t)) = E_\xi^t(Z_t^4 - 6Z_t^2 + 3) = E_\xi^t(Z_t^4 - 5(Z_t^2-1) - 3) - E_\xi^t(Z_t^2-1),$$

which can be approximated to $O(t^{-2})$ by (33) and (35); and similarly, simple algebra gives

$$E_\xi^t(q_6(Z_t)) = E_\xi^t(Z_t^6 - 33(Z_t^2-1) - 15) - 15E_\xi^t(q_4(Z_t)) - 12E_\xi^t(q_2(Z_t)),$$

where the first term on the right side can be approximated to $O(t^{-2})$ by (34). Then, we can plug these approximations into (26) and compare with $\gamma_2(w,x)$ in (20). We omit details of the derivations. Unfortunately, we found that the two approximations do not agree arithmetically; in fact, our (26) gives much more terms than $\gamma_2(w,x)$ in (20). Since the derivation is tedious and difficult to detect errors, in the next section we conduct simulations to further compare these two expansions.

4.3. *Examples*

Since Johnson's formulas are for the 1-dimensional case, in the first two examples we use one-parameter models to compare his results with ours. Since an approximation that includes $O(t^{-1/2})$ term has an error of order $O(t^{-1})$, throughout this section we refer to an approximation that includes $O(t^{-1/2})$ term as an *approximation to* $O(t^{-1})$; similarly, an approximation that includes $O(t^{-1})$ term is said to be an *approximation to* $O(t^{-3/2})$. The simulations show that the two analytic approximations to $O(t^{-1})$ are fairly close, and for $O(t^{-3/2})$ ours performs better than Johnson's. In the third example, we assess the accuracy of our analytic approximations for a two-parameter logistic model.

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

4.3.1. *Beta-Binomial Example.* 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.

We compare the approximate posterior density of θ by Johnson's formulas and our (23) with $m=1,2$ and posterior moments replaced by approximations derived

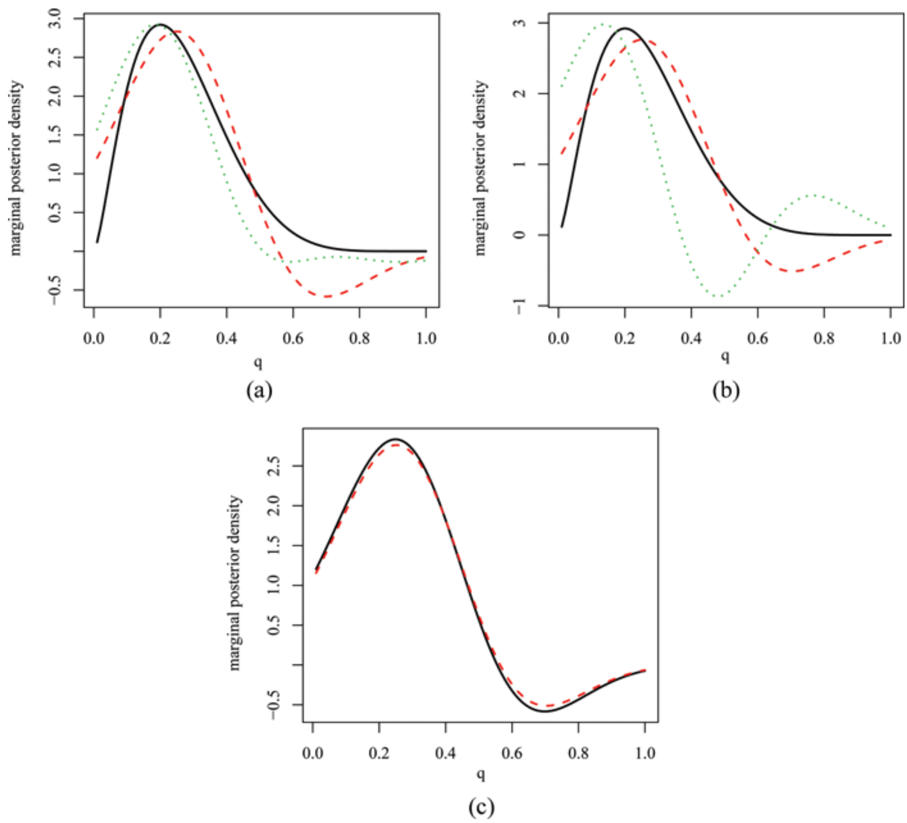

Figure 1. Marginal posterior pdf of θ. Beta-Binomial model. (a) Solid: Exact distribution; Dashed: Our $O(t^{-1})$; Dotted: Our $O(t^{-3/2})$. (b) Solid: Exact distribution; Dashed: Johnson's $O(t^{-1})$; Dotted: Johnson's $O(t^{-3/2})$. (c) Solid: Our $O(t^{-1})$; Dashed: Johnson's $O(t^{-1})$. (color figure available online)

in Sec. 4.1. Here, Johnson's approximation to $O(t^{-1})$ is obtained by taking $K = 1$ in (18):

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

and the approximation to $O(t^{-3/2})$ is by taking $K = 2$ in (18). Fig. 1(a) gives the true density and our approximations to $O(t^{-1})$ and $O(t^{-3/2})$; Fig. 1(b) gives the true density and Johnson's approximations to $O(t^{-1})$ and $O(t^{-3/2})$; and Fig. 1(c) contains the two $O(t^{-1})$ approximations.

We have some observations. First, Fig. 1(c) shows that the two $O(t^{-1})$ approximations are quite close, which agrees with our theoretical finding. Secondly, Fig. 1(a) shows that our approximation to $O(t^{-3/2})$ is closer to the true density than approximation to $O(t^{-1})$, but Fig. 1(b) reveals that Johnson's formula to $O(t^{-3/2})$ does not improve upon $O(t^{-1})$.

4.3.2. *Gamma-Poisson Example.* To further assess the accuracy of our approximations and Johnson's formulas, we consider an i.i.d. sample $y_1, \ldots, y_n$

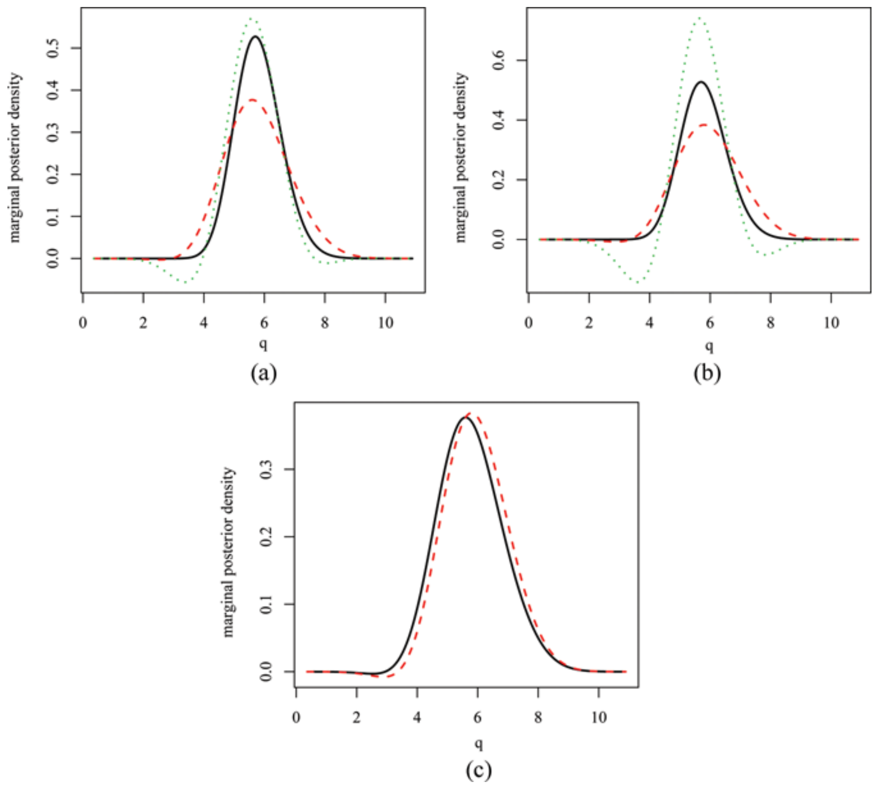

Figure 2. Marginal posterior pdf of θ. Poisson model with prior Gamma(30,5). (a) Solid: Exact distribution; Dashed: Our $O(t^{-1})$; Dotted: Our $O(t^{-3/2})$. (b) Solid: Exact distribution; Dashed: Johnson's $O(t^{-1})$; Dotted: Johnson's $O(t^{-3/2})$. (c) Solid: Our $O(t^{-1})$; Dashed: Johnson's $O(t^{-1})$. (color figure available online)

from Poisson(θ), where the prior of θ is assumed to be Gamma(a, b). Suppose that $(y_1, y_2, y_3, y_4, y_5) = (3, 5, 7, 10, 3)$ and that $(a, b) = (30, 5)$. Thus, the MLE of θ is 5.6, the prior mean of θ is 6 and the posterior distribution of θ follows $\text{Gamma}(a + \sum_{i=1}^{n} y_i, b + n) = \text{Gamma}(58, 10)$. We have similar observations as in Sec. 4.3.1: First, Fig. 2(c) indicates that the two $O(t^{-1})$ approximations are fairly close. Secondly, Fig. 2(a) shows that our approximation to $O(t^{-3/2})$ improves upon $O(t^{-1})$, but Fig. 2(b) shows that Johnson's does not.

Now we change the prior distribution to see its effect on the analytic approximations. Suppose that $(a, b) = (15, 5)$. So, the prior mean of θ is 3 and $\theta|_y \sim$ Gamma(43, 10). The results are in Fig. 3. As before, Fig. 3(c) indicates that the two $O(t^{-1})$ approximations are close. However, possibly due to the fact that the prior mean of θ is farther from the MLE, we found from Figs. 3(a) and 3(b) that both $O(t^{-3/2})$ approximations are worse than $O(t^{-1})$. A closer look at these two $O(t^{-3/2})$ curves show that Johnson's approximation (ranges between -2 and 1.5) fluctuates more widely than ours (ranges between -1 and 1).

4.3.3. *Logistic Example.* We consider a data taken from Mendenhall et al. (1989); see also Tanner (1996). The explanatory variable is the number of days of

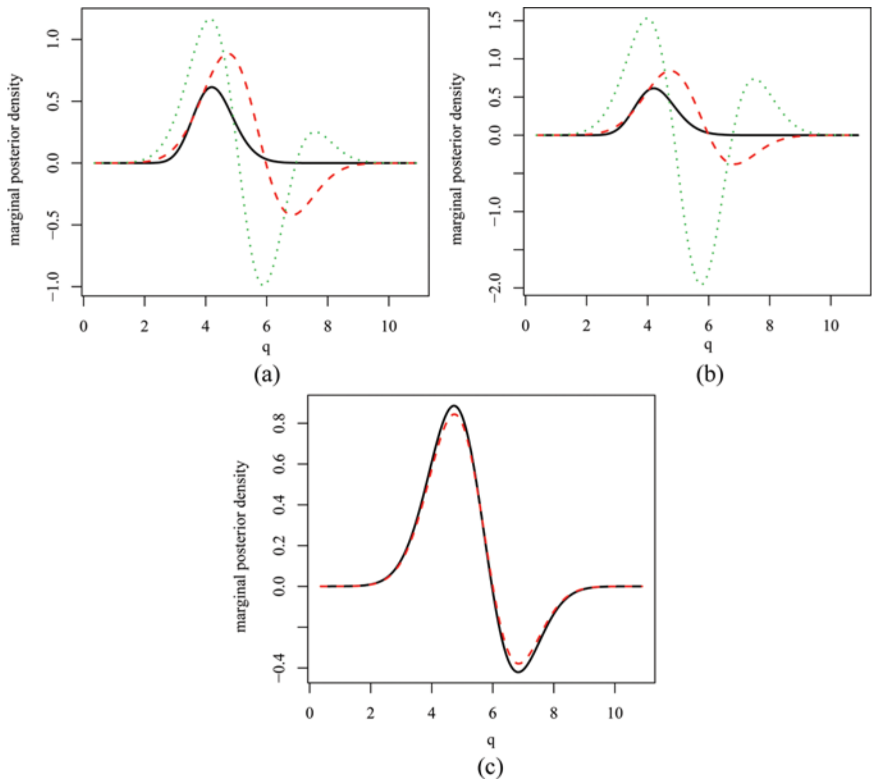

Figure 3. Marginal posterior pdf of θ. Poisson model with prior Gamma(15,5). (a) Solid: Exact distribution; Dashed: Our $O(t^{-1})$; Dotted: Our $O(t^{-3/2})$. (b) Solid: Exact distribution; Dashed: Johnson's $O(t^{-1})$; Dotted: Johnson's $O(t^{-3/2})$. (c) Solid: Our $O(t^{-1})$; Dashed: Johnson's $O(t^{-1})$. (color figure available online)

radiotherapy received by each of 24 patients, and the response variable is the absence (1) and presence (0) of disease at a site three years after treatment. A problem of interest is to use the covariate (days) to predict outcome.

We fit the data using the logistic regression model

$$\log\left(\frac{p_i}{1-p_i}\right) = \theta_1 + \theta_2 c_i,$$

where c_i is the covariate (days) for patient i and p_i is the probability of no disease. So, $p_i = \exp(\theta_1 + \theta_2 c_i)/(1 + \exp(\theta_1 + \theta_2 c_i))$. The intercept θ_1 represents the log-odds of success for zero days, while the slope θ_2 represents the change in the log-odds of success (no disease) for every unit increase in the covariate. The loglikelihood is

$$\begin{aligned}\ell_t(\theta) &= \sum_{i=1}^{t}[y_i \log p_i + (1-y_i)\log(1-p_i)] \\ &= \sum_{i=1}^{t}[y_i(\theta_1 + \theta_2 c_i) - \log(1 + \exp(\theta_1 + \theta_2 c_i))];\end{aligned}$$

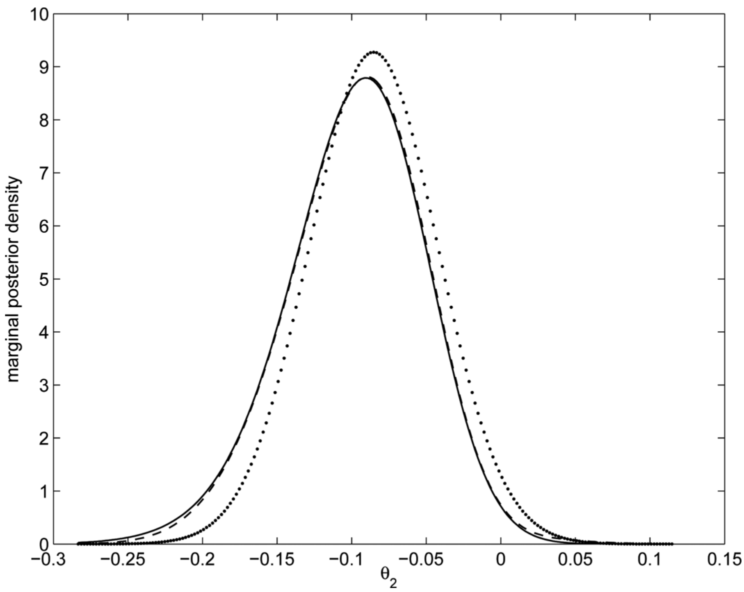

Figure 4. Marginal posterior pdf of θ_2. Logit2p-flat model. Solid line: Exact distribution by numerical integration; Dashed line: Our approximation to $O(t^{-3/2})$; Dotted: Normal approximation.

and the marginal posterior density of θ_2 involves two integrals:

$$p(\theta_2 \mid x) = \frac{\int \xi(\theta_1, \theta_2)\exp[\ell_t(\theta_1, \theta_2)]d\theta_1}{\int\int \xi(\theta_1, \theta_2)\exp[\ell_t(\theta_1, \theta_2)]d\theta_1 d\theta_2}.$$

These two integrals are intractable for commonly used priors such as flat or normal.

Now we take flat priors on both θ_1 and θ_2, and use the expansion (23) with $m = 2$ and approximate moments obtained in Sec. 4.1. Figure 4 provides the approximate marginal posterior densities of θ_2 by normal approximation, our approximation, and the exact density using numerical integration. Note that here $\hat{\theta}_{t2} = -0.0853$ and $[\Sigma_t]_{22} = 23.25$; and so the normal approximation says that the posterior density of θ_2 given data is approximately $N(-0.085, (1/23.25)^2)$. The figure shows that our approximation is quite close to the exact distribution.

To see whether the analytic approximation performs well or not with different priors, we consider three normal priors: $N(0, 1)$, $N(0, 3)$, and $N(0, 6)$. Since the posterior standard error of θ_2 is around $1/23.25 = 0.043$, the prior mean 0 is about two standard errors away from $\hat{\theta}_{t2}$. With such priors, Figure 5 showed that the less informative the prior is (i.e. larger variance), the more accurate the approximate density is.

5. Concluding Remarks

We showed how to use Stein's identity to derive approximations of posterior moments and compared our approximation with Johnson (1970). Our derivation showed that the $O(t^{-1/2})$ terms in both expressions are arithmetically equivalent; however, the $O(t^{-1})$ terms are not. Then we provided some examples to assess the

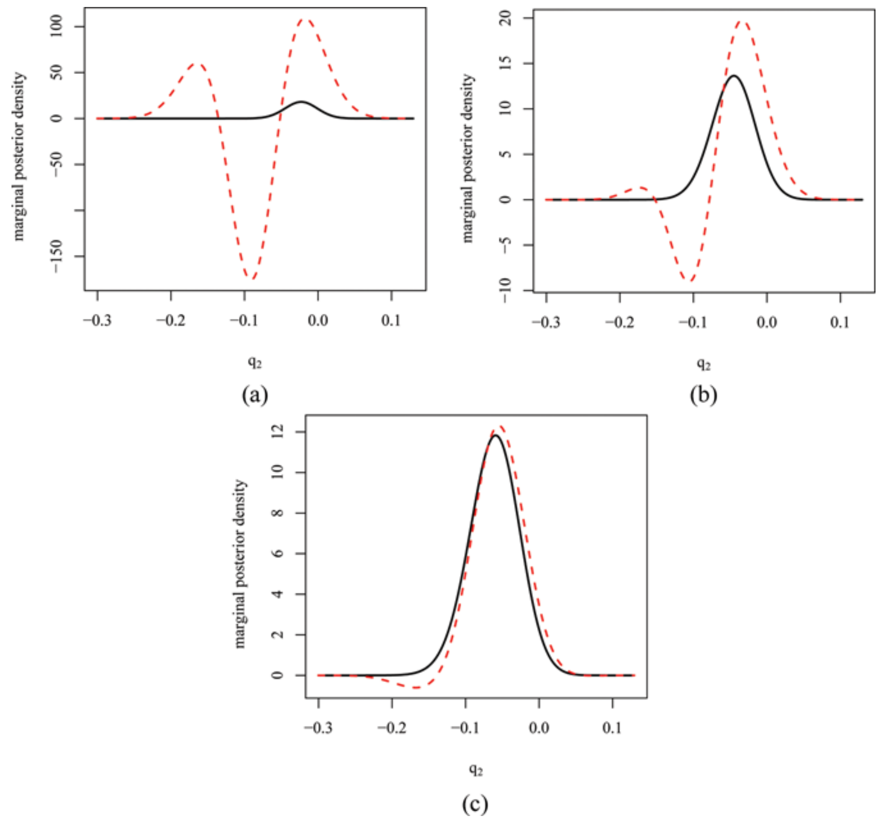

Figure 5. Marginal posterior pdf of θ_2. Logit2p-normal model. Solid line: Exact distribution by numerical integration; Dashed line: Our approximation $O(t^{-3/2})$. (a) $N(0, 1)$ prior; (b) $N(0, 3)$ prior; (c) $N(0, 6)$ prior. (color figure available online)

accuracy of these approximations. The simulation study in Secs. 4.3.1 and 4.3.2 confirmed above findings and revealed that our $O(t^{-1})$ term is slightly better than Johnson's. We also considered different priors in Sec. 4.3.3 and found that the analytic approximations performs better when the prior is less informative.

Appendix

A.1 *Proof of Lemma* 4.3

It should always be remembered that the derivatives of f_t are in (7) and (8), and ∇u_t and $\nabla^2 u_t$ are in (9) and (10). First note that if h is a polynomial of order r, Uh and Vh are of orders $r-1$ and $r-2$ (see Weng and Woodroofe, 2000, Lemma 8); and that by (13), $\Phi_p Uh = 0$ for even r. Denote $\delta_t = (\delta_{t1}, \ldots, \delta_{tp})^T = \theta - \hat{\theta}_t$. Then, by Taylor expansions,

$$[\nabla u_t(\theta)]_i = \frac{1}{2}\delta_t^T D_i \delta_t + (\mathrm{Rem}_1) = \frac{1}{2} Z_t^T V_i Z_t + (\mathrm{Rem}_1), \tag{36}$$

$$[\nabla^2 u_t(\theta)]_{ij} = [D_i]_{j.}\Sigma_t^{-1} Z_t + \frac{1}{2}\sum_{k,s} \hat{\ell}^{(4)}_{ijks}[Z_t^T(\Sigma_t^T)^{-1} e_k e_s^T \Sigma_t^{-1} Z_t] + (\mathrm{Rem}_2), \tag{37}$$

where $(\mathrm{Rem}_1) = (1/6)\sum_{jks} \hat{\ell}^{(4)}_{ijks}\delta_{tj}\delta_{tk}\delta_{ts} + (1/24)\sum_{jksq} \ell^{(5)}_{ijksq}(\tilde{\theta}_t)\delta_{tj}\delta_{tk}\delta_{ts}\delta_{tq}$, $\tilde{\theta}_t$ lies between θ and $\hat{\theta}_t$, and (Rem_2) has a similar form. So, by Lemma 4.2(i), it can be shown that $E^t_\xi(\mathrm{Rem}_i) = O(t^{-1})$ for $i = 1, 2$.

Now consider the quadratic terms in (36) and (37). By Lemma 4.2(ii) we have

$$\begin{aligned} E^t_\xi(Z_t^T V_i Z_t) &= \mathrm{tr}(V_i) + O(t^{-1}) = S_i + O(t^{-1}), \\ E^t_\xi[Z_t^T(\Sigma_t^T)^{-1} e_k e_s^T \Sigma_t^{-1} Z_t] &= \mathrm{tr}[(\Sigma_t^T)^{-1} e_k e_s^T \Sigma_t^{-1}] + O(t^{-2}) \\ &= [(-\nabla^2 \hat{\ell}_t)^{-1}]_{ks} + O(t^{-2}), \end{aligned} \tag{38}$$

where the last line follows from (2). So, (i) follows. Next, (ii) follows by taking posterior expectations on (37) and employing Lemma 4.3(i) and (38). Finally, some algebra yields

$$\begin{aligned} [E^t_\xi(\nabla u_t) E^t_\xi(\nabla u_t^T)]_{ij} &= \frac{1}{4}\sum_i([V_k]_{ii}[V_l]_{ii}) + \frac{1}{2}\sum_{i<j}([V_k]_{ii}[V_l]_{jj}) + O(t^{-1}), \\ E^t_\xi[\nabla u_t \nabla u_t^T]_{ij} &= \frac{3}{4}\sum_i([V_k]_{ii}[V_l]_{ii}) + \frac{1}{2}\sum_{i<j}([V_k]_{ii}[V_l]_{jj}) + \sum_{i<j}([V_k]_{ij}[V_l]_{ij}) + O(t^{-1}); \end{aligned}$$

and together with the fact that $\mathrm{tr}(V_k V_l) = \mathrm{tr}(HD_k HD_l)$, (iii) follows.

A.2 *Proof of Lemma* 4.4

Assertion (i) follows from (7), (30), (32) and Lemma 4.3(i). For assertion (ii), first write

$$\begin{aligned} [V^t_\xi Z_t]_{ij} &= E^t_\xi(Z_{ti}Z_{tj}) - (E^t_\xi Z_{ti})(E^t_\xi Z_{tj}) \\ &= \delta_{ij} + \{(\Sigma_t^T)^{-1} E^t_\xi\Big[\Big(\frac{\nabla^2\xi}{\xi}\Big) + \nabla^2 u_t + \nabla u_t \nabla u_t^T\Big]\Sigma_t^{-1}\}_{ij} \\ &\quad - \Big\{(\Sigma_t^T)^{-1}[E^t_\xi\Big(\frac{\nabla\xi}{\xi}\Big)E^t_\xi\Big(\frac{\nabla\xi^T}{\xi}\Big) + E^t_\xi(\nabla u_t)E^t_\xi(\nabla u_t^T)]\Sigma_t^{-1}\Big\}_{ij} + O(t^{-2}) \\ &= \delta_{ij} + \Big\{(\Sigma_t^T)^{-1} E^t_\xi\Big[\Big(\frac{\nabla^2\xi}{\xi}\Big) - E^t_\xi\Big(\frac{\nabla\xi}{\xi}\Big)E^t_\xi\Big(\frac{\nabla\xi^T}{\xi}\Big)\Big]\Sigma_t^{-1}\Big\}_{ij} \\ &\quad + \Big\{\Big(\Sigma_t^T)^{-1}[E^t_\xi(\nabla^2 u_t + \nabla u_t \nabla u_t^T) - E^t_\xi(\nabla u_t)E^t_\xi(\nabla u_t^T)]\Sigma_t^{-1}\Big\}_{ij} + O(t^{-2}), \end{aligned}$$

where the first equality follows since $E^t_\xi((\nabla\xi/\xi)\nabla u_t^T) = O(t^{-1})$. Then, together with (32) and Lemma 4.3(ii)–(iii), we obtain (ii).

A.3 *Proof of Lemma* 4.6

From Lemma 4.2, (9), and some straightforward calculations, we have

$$E^t_\xi\Big(Z_{ti}\Big[\frac{\nabla^2 f_t}{f_t}\Big]_{jk}\Big) = E^t_\xi[(\Sigma_t^T)^{-1}(Z_{ti}\nabla^2 u_t)\Sigma_t^{-1}]_{jk} + O(t^{-3/2}),$$

where by (37) the posterior expectation of the (r, s)-component of $Z_{ti}\nabla^2 u_t$ is

$$E^t_\xi([Z_{ti}\nabla^2 u_t]_{rs}) = E^t_\xi(Z_{ti}[D_r]_{s\cdot}\Sigma_t^{-1}Z_t) + O(t^{-1/2}) = \sum_{l=1}^{p}\hat{\ell}^{(3)}_{lrs}[\Sigma_t^{-1}]_{li} + O(t^{-1/2}).$$

So, and the desired result follows by writing

$$\begin{aligned} E^t_\xi(Z_{ti}\nabla^2 u_t) &= \sum_{l=1}^{p}\{[\Sigma_t^{-1}]_{li}D_l\} + O(t^{-1/2}),\\ E^t_\xi[(\Sigma_t^T)^{-1}(Z_{ti}\nabla^2 u_t)\Sigma_t^{-1}] &= \sum_{l=1}^{p}\{[\Sigma_t^{-1}]_{li}(\Sigma_t^T)^{-1}D_l\Sigma_t^{-1}\} + O(t^{-3/2})\\ &= \sum_{l=1}^{p}\{[\Sigma_t^{-1}]_{li}V_l\} + O(t^{-3/2}). \end{aligned}$$

A.4 *Proof of Lemma* 4.7

If $h^*(z_p) = z_p^4$, then $\Phi U h^* = 0$ and $U^2 h^*(z_p) = (z_p^2 + 5)$; and therefore, in (16) taking $h^*(z_p) = z_p^4$ and $s = 2$ yields

$$E^t_\xi Z^4_{tp} = 3 + E^t_\xi\left(Z^2_{tp}\left[\frac{\nabla^2 f_t}{f_t}(Z_t)\right]_{pp}\right) + 5E^t_\xi\left(\left[\frac{\nabla^2 f_t}{f_t}(Z_t)\right]_{pp}\right),$$

where $E^t_\xi([\nabla^2 f_t/f_t(Z_t)]_{pp})$ has been obtained from (31) and Lemma 4.4. So, it suffices to evaluate $E^t_\xi(Z^2_{tp}[\nabla^2 f_t/f_t(Z_t)]_{pp})$. First, taking Taylor's expansions of $\nabla\xi/\xi$ and $\nabla^2\xi/\xi$ at $\hat{\theta}_t$ and using Lemma 4.2 and (36) gives

$$\begin{aligned} E^t_\xi\left(Z^2_{tp}\frac{\nabla^2\xi}{\xi}\right) &= \frac{\nabla^2\hat{\xi}}{\hat{\xi}} + O(t^{-1})\\ E^t_\xi\left(Z^2_{tp}\left[\frac{\nabla\xi}{\xi}\nabla u_t^T\right]_{ij}\right) &= \frac{\hat{\xi}^{(1)}_i}{2\hat{\xi}}(\mathrm{tr}(V_j) + 2[V_j]_{pp}) + O(t^{-1})\\ E^t_\xi\left(Z^2_{tp}\left[\frac{\nabla\xi}{\xi}\nabla u_t^T\right]\right) &= \frac{\nabla\hat{\xi}}{2\hat{\xi}}Q^T + O(t^{-1}), \end{aligned}$$

where $Q = (Q_1, \ldots, Q_p)^T$ with $Q_j = \mathrm{diag}(V_j) + 2[V_j]_{pp}$. Next, by (37), we have

$$\begin{aligned} E^t_\xi(Z^2_{tp}[\nabla^2 u_t]_{rs}) &= E^t_\xi\Big(Z^2_{tp}[D_r]_{s\cdot}\Sigma_t^{-1}Z_t + \frac{1}{2}Z^2_{tp}\sum_{i,j}\hat{\ell}^{(4)}_{rsij}\delta_i\delta_j\Big) + O(t^{-1})\\ &= [D_r]_{s\cdot}\Sigma_t^{-1}E^t_\xi(Z^2_{tp}Z_t) + \frac{1}{2}\sum_{i,j}\hat{\ell}^{(4)}_{rsij}E^t_\xi(Z^2_{tp}\delta_i\delta_j) + O(t^{-1}), \end{aligned}$$

where straightforward calculations yield

$$E^t_\xi(Z^2_{tp}\delta_i\delta_j) = [H]_{ij} + 2[\Sigma_t^{-1}]_{ip}[\Sigma_t^{-1}]_{jp} + O(t^{-1});$$

and hence, letting J be the $p \times p$ matrix defined by $[J]_{ij} = [\Sigma_t^{-1}]_{ip}[\Sigma_t^{-1}]_{jp}$, we have

$$\frac{1}{2}\sum_{i,j} \hat{\ell}^{(4)}_{rsij} E^t_\xi(Z^2_{tp}\delta_i\delta_j) = \mathrm{tr}\left(\frac{1}{2}D_{rs}H + D_{rs}S\right) + O(t^{-1}).$$

Moreover, (36) and some calculations give

$$\begin{aligned}
&E^t_\xi(Z^2_{tp}[\nabla u_t \nabla u_t^T]_{kl}) \\
&= \frac{1}{4}E^t_\xi\left\{Z^2_{tp}\left(\sum_i [V_k]_{ii}Z^2_{ti} + 2\sum_{i<j}[V_k]_{ij}Z_{ti}Z_{tj}\right)\left(\sum_i [V_l]_{ii}Z^2_{ti} + 2\sum_{i<j}[V_l]_{ij}Z_{ti}Z_{tj}\right)\right\} + O(t^{-1}) \\
&= \frac{1}{2}\mathrm{tr}(V_k V_l) + \frac{1}{4}\mathrm{tr}(V_k)\mathrm{tr}(V_l) + \frac{1}{2}[V_k]_{pp}\mathrm{tr}(V_l) + \frac{1}{2}[V_l]_{pp}\mathrm{tr}(V_k) \\
&\quad + 2\sum_{i=1}([V_k]_{ip}[V_l]_{ip}) + O(t^{-1}).
\end{aligned}$$

Then the desired results follows.

Acknowledgments

The authors would like to thank the referees for their valuable comments on the article. The authors are partially supported by the National Science Council of Taiwan.

References

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

Johnson, R. (1970). Asymptotic expansions associated with posterior distributions. *Ann. Math. Statist.* 41:851–864.

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

Lindley, D. V. (1980). Approximate Bayesian methods. In: Bernardo, J. M., DeGroot, M. H., Lindley, D. V., A. F. M. S., eds. *Bayesian Statistics*. Valencia, Spain: University Press, pp. 223–237.

Mendenhall, W. M., Parsons, J. T., Stringer, S. P., Cassissi, N. J., Million, R. R. (1989). T2 oral tongue carcinoma treated with radiotherapy: analysis of local control and complications. *Radiother. Oncol.* 16:275–282.

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

R Development Core Team (2009). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna. URL http://www.R-project.org.

Tanner, M. A. (1996). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*. 3rd ed. New York: Springer.

Tierney, L., Kadane, J. B. (1986). Accurate approximations for posterior moments and marginal densities. *J. Amer. Statist. Assoc.* 81:82–86.

Tierney, L., Kass, R. E., Kadane, J. B. (1989). Fully exponential laplace approximations to expectations and variances of nonpositive functions. *J. Amer. Statist. Assoc.* 84:710–716.

Weng, R. C. (2010). A Bayesian Edgeworth expansion by Stein's Identity. *Bayesian Anal.* 5(4):741–764.

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

Woodroofe, M. (1989). Very weak expansions for sequentially designed experiments: linear models. *Ann. Statist.* 17:1087–1102.