Statistica Sinica **10**(2000), 693-713

INTEGRABLE EXPANSIONS FOR POSTERIOR DISTRIBUTIONS FOR MULTIPARAMETER EXPONENTIAL FAMILIES WITH APPLICATIONS TO SEQUENTIAL CONFIDENCE LEVELS

Ruby C. Weng and Michael Woodroofe

University of Michigan

Abstract: Integrable expansions for posterior distributions are obtained for sequential samples from a multiparameter exponential family. A data dependent transformation is used to convert the likelihood function to the form of a standard multivariate normal density. Then a version of Stein's Identity is applied. This leaves an expression from which an asymptotic expansion is easily obtained. The results are applied to find confidence intervals for the ratio of two Poisson means after a sequential test and compare well with simulations.

Key words and phrases: Asymptotic expansions, multiparameter exponential family, sequential confidence levels, Stein's Identity, very weak expansion.

1. Introduction

Asymptotic expansions for posterior distributions have been studied since the time of Laplace, and interest in them continues to this day. See, for example, Kass, Tierney and Kadane (1987). A conventional approach to such problems starts from a Taylor series expansion of the log-likelihood function around the maximum likelihood estimator and proceeds from there to develop expansions that hold almost surely, given the data. Johnson (1967, 1970) provides a careful account of this approach. Two recent developments are central to this paper. One of these is interest in integrable expansions, expansions that can be integrated with respect to the marginal distribution of the data. Integrable expansions provide higher order approximations to the overall Bayes' risk and, so, are useful for design considerations. In addition, they may be used to form very weak expansions for (frequentist) confidence levels, as described in Sections 6 and 7 below. Ghosh, Sinha and Joshi (1982) were the first to consider integrable expansions (to the best of the author's knowledge). They provided conditions on the prior and model under which the almost sure expansions could be integrated termwise. This work was followed by Woodroofe (1986) and Bickel and Ghosh (1990) among others. In related work, Woodroofe (1989, 1992) showed how a version of Stein's (1987) identity could be used to write posterior expectations in a form from which asymptotic expansions could be readily guessed. Moreover, there is martingale structure inherent in this approach, and the latter is useful in obtaining integrable expansions. Woodroofe (1989, 1992) considered two cases, linear models with normal errors and one-parameter exponential families. Here the approach is extended to multi-parameter exponential families. Multiparameter exponential families have also been considered by Sun (1994) and by Coad and Woodroofe (1996) in special cases, both involving just two parameters. By way of contrast, our formulation is quite general, requiring only that the family be minimal and that the natural parameter space be open.

The expansions derived here differ from those derived by Takahashi (1987), Woodroofe and Keener (1987), Woodroofe (1988) and Lai and Wang (1994), who obtain asymptotic expansions for fixed values of the parameter θ. First, the scaling is different. The latter authors derive expansions for normalized estimation error, and the first three only consider normalized sums; here we use the signed root transformation. Even in the normal case, where the scalings agree, the expansions are different. The fixed θ expansions contain ladder height distributions and oscillatory terms, like the fractional part of $N\theta$, where N is approaching infinity rapidly. The coefficients in our expansions are continuous functions of the parameter that do not involve ladder height distributions and may be estimated quite easily. The price that we pay for the more tractable coefficients is to use a weaker form of convergence, one that effectively smooths out the oscillations in the fixed θ expansions.

The model and the application of Stein's Identity to posterior distributions are reviewed in Sections 2 and 3. A key observation here is that a suitable parameter transformation, called Z_n, converts the likelihood function into a normal form. In Section 4, asymptotic expansions are derived for the posterior expectation of $h(Z_t)$ for bounded functions h and suitable families of stopping times t. In Section 6, the results are specialized to the two-parameter case, and the asymptotic expansions of Section 4 are used to develop very weak expansions for (frequentist) sequential confidence levels. The process is illustrated in Section 7 by applying it to find confidence intervals for the ratio of two Poisson means after a sequential test. Simulation experiments indicate that the approximations are very accurate. Section 5 contains some bounds that are useful for dealing with unbounded functions h. In addition to its increased generality, the paper is novel in two other ways. Even when specialized to the one-parameter case, the conditions required of the prior here are weaker than those imposed by Woodroofe (1992), at least for bounded h. Moreover, the approach is applied to a problem involving group sequential testing.

2. The Model

A k-parameter exponential family is a family of distributions defined by probability densities of the form

$$p_\omega(y) = C(\omega)\exp\{\sum_{j=1}^{k} Q_j(\omega)T_j(y)\}h(y)$$

with respect to a σ-finite measure ν on an Euclidean space $\mathcal{X}$. Employing suitable reductions by reparametrization and sufficiency, together with proper choice of the dominating measure, leads to the so-called *standard exponential family*

$$p_\theta(x) = e^{\theta' x - \psi(\theta)} \tag{1}$$

with respect to a σ-finite measure μ, where $x = T(y)$, $\theta = Q(\omega)$, and $\Omega = \{\theta : \int e^{\theta' x}\mu(dx) < \infty\}$ is the natural parameter space. Further, the family is called *minimal* if $\dim(\Omega) = \dim(\mathcal{X}) = k$, where $\mathcal{X}$ is the convex support of μ. This is equivalent to requiring p_θ not being reducible to a $(k-1)$-parameter exponential family. For references, see Brown (1986) and Lehmann (1983, 1986).

Throughout this paper $\{p_\theta : \theta \in \Omega\}$ is assumed to be a k-parameter minimal standard exponential family of the form in (1), and Ω is assumed to be open. Suppose that θ_1 is the parameter of interest and the others are nuisance parameters. Let $X_1, \ldots, X_n$ be i.i.d. from p_θ. The log likelihood function based on $x_1, \ldots, x_n$ is $L_n(\theta) = \theta' S_n - n\psi(\theta)$, $\theta \in \Omega$, where $S_n = \sum_{i=1}^n x_i$. Let $\bar{x}_n = S_n/n$ and suppose for the present that $\bar{x}_n \in \nabla\psi(\Omega)$. Then the maximum likelihood estimator solves the equation $\nabla\psi(\hat{\theta}_n) = \bar{x}_n$, and

$$L_n(\theta) = n\Psi(\hat{\theta}_n, \theta), \tag{2}$$

where $\Psi(\omega, \theta) = \theta'\nabla\psi(\omega) - \psi(\theta)$. Consider the signed-root transformation Z_n, as in Barndorff-Nielsen (1986): for $i = 1, \ldots, k$, define

$$Z_{ni} = Z_{ni}(\theta) = \Big(2[L_n(\hat{\theta}_n^{i-1}) - L_n(\hat{\theta}_n^{i})]\Big)^{1/2}\operatorname{sgn}(\theta_i - \hat{\theta}_{ni}^{i-1}), \tag{3}$$

where $\hat{\theta}_n^0 = \hat{\theta}_n$, the maximum likelihood estimator, for $i = 1, \ldots, k-1$, $\hat{\theta}_n^i$ is the restricted maximum likelihood estimator for fixed $(\theta_1, \ldots, \theta_i)$, and $\hat{\theta}_n^k$ is exactly θ. Then $L_n(\theta) = L_n(\hat{\theta}_n) - \frac{1}{2}||Z_n||^2$.

Consider a Bayesian model in which θ has a continuously differentiable prior density ξ with compact support $K \subseteq \Omega$. Then the posterior density of θ given $x_1, \ldots, x_n$ is $\xi_n(\theta) \propto e^{L_n(\theta)}\xi(\theta)$. So, the posterior density of Z_n is

$$\zeta_n(z) \propto J(\hat{\theta}_n, \theta)\xi_n(\theta) \propto J(\hat{\theta}_n, \theta)\xi(\theta)e^{-\frac{1}{2}||z||^2}, \tag{4}$$

where z and θ are related by (3) and J is a Jacobian term. Using (2) and (3), it is easily seen that the Jacobian of the transformation is $J(\hat{\theta}_n, \theta)/n^{k/2}$, where $J(\hat{\theta}_n, \theta) = \prod_{l=1}^k J^l(\hat{\theta}_n, \theta)$, with

$$\frac{1}{J^l(\hat{\theta}_n, \theta)} = \Big|\frac{\partial}{\partial \theta_l}\Big(2[\Psi(\hat{\theta}_n, \hat{\theta}_n^{l-1}) - \Psi(\hat{\theta}_n, \hat{\theta}_n^l)]\Big)^{1/2}\Big|. \tag{5}$$

The term $n^{-k/2}$ may be absorbed into the proportionality constant in (4), but reappears later. From (4),

$$\zeta_n(z) = f_n(z)\phi_k(z), \quad z \in \Re^k, \tag{6}$$

where ϕ_k denotes the standard k-variate normal density and $f_n(z) \propto J(\hat{\theta}_n, \theta)\xi(\theta)$.

3. Stein's Identity

Let Φ_k denote the standard k-variate normal distribution and write

$$\Phi_k h = \int h d\Phi_k$$

for functions h for which the integral is finite. Next let Γ denote a finite signed measure of the form $d\Gamma = f d\Phi_k$, where f is a real-valued function defined on $\Re^k$ satisfying $\Phi_k|f| = \int |f| d\Phi_k < \infty$. For $p > 0$, denote H_p as the collection of all measurable functions $h: \Re^k \to \Re$ for which $|h(z)| \le 1 + ||z||^p$. Then, define $\tilde{H}_p = \{h: |h(z)|/b \in H_p, \text{ for some } b > 0\}$ and $H = \cup_{p\ge 0}\tilde{H}_p$. Given $h \in \tilde{H}_p$, let $h_0 = \Phi_k h$, $h_k = h$,

$$h_j(y_1, \ldots, y_j) = \int_{\Re^{k-j}} h(y_1, \ldots, y_j, w)\Phi_{k-j}(dw), \tag{7}$$

and

$$g_j(y_1, \ldots, y_k) = e^{\frac{1}{2}y_j^2}\int_{y_j}^{\infty} [h_j(y_1, \ldots, y_{j-1}, w) - h_{j-1}(y_1, \ldots, y_{j-1})]e^{-\frac{1}{2}w^2} dw, \tag{8}$$

for $-\infty < y_1, \ldots, y_k < \infty$ and $j = 1, \ldots, k$. Then let $Uh = (g_1, \ldots, g_k)'$. Note that U may be iterated. Let $Vh = (U^2h + U^2h')/2$, where U^2h is the $k \times k$ matrix whose jth column is Ug_j and g_j is as in (8). Then Vh is a symmetric matrix. Simple calculations show that

$$\Phi_k(Uh) = \int_{\Re^k} zh(z)\Phi_k(dz) \tag{9}$$

and

$$\Phi_k(Vh) = \frac{1}{2}\int_{\Re^k} (zz' - I_k)h(z)\Phi_k(dz) \tag{10}$$

for all $h \in H$. When $k = 1$, these formulas simplify. Then

$$Uh(z) = e^{\frac{1}{2}z^2} \int_z^\infty (h(y) - \Phi h) e^{-\frac{1}{2}w^2} dw$$

and U^2 is the composition of U with itself. It is easily seen that if $h \in \tilde{H}_p$, then $||Uh|| \in \tilde{H}_p$.

Lemma 1. (*Stein's Identity*) *Let r be a nonnegative integer. Suppose that $d\Gamma = f d\Phi_k$ as above, where f is a differentiable function on $\Re^k$, for which*

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

Then

$$\Gamma h = \Gamma 1 \cdot \Phi_k h + \int_{\Re^k} (Uh(z))' \nabla f(z) \Phi_k(dz) \tag{11}$$

for all $h \in \tilde{H}_r$. If $\partial f / \partial z_j,\ j = 1, \ldots, k$, are differentiable, and

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

then

$$\Gamma h = \Gamma 1 \cdot \Phi_k h + \Phi_k(Uh)' \int_{\Re^k} \nabla f(z) \Phi_k(dz) + \int_{\Re^k} tr[(Vh)\nabla^2 f] d\Phi_k$$

for all $h \in \tilde{H}_r$.

Proof. The first assertion follows from Woodroofe (1989, Proposition 1). For the second assertion, write

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

and then apply (11) with h and f replaced by g_i and $\partial f / \partial z_i$.

From (6), the posterior distributions are of a form appropriate for Stein's Identity. Let

$$\Gamma_1^\xi(\hat{\theta}_n, \theta) = n^{1/2} \frac{\nabla f_n(Z_n)}{f_n(Z_n)} \tag{12}$$

and

$$\Gamma_2^\xi(\hat{\theta}_n, \theta) = n \frac{\nabla^2 f_n(Z_n)}{f_n(Z_n)}. \tag{13}$$

To understand the structure of these terms, let $\nabla_\theta f_n(Z_n)$ denote the vector of partial derivatives of $f_n(Z_n)$ by θ and $D_\theta z$ denote the matrix of partial derivatives of Z_{ni} by θ_j. So, $\nabla_\theta f_n(Z_n) = (D_\theta z)' \nabla_z f_n(Z_n)$. From this, we obtain

$$\nabla_z f_n(Z_n) = [(D_\theta z)']^{-1} \nabla_\theta f_n(Z_n) \tag{14}$$

and

$$||\Gamma_1^{\xi}(\hat{\theta}_n,\theta)|| = ||n^{1/2}(D_\theta z)^{-1}(\frac{\xi(\theta)\nabla_\theta J(\hat{\theta}_n,\theta)+J(\hat{\theta}_n,\theta)\nabla_\theta\xi(\theta)}{J(\hat{\theta}_n,\theta)\xi(\theta)})||$$
$$\leq C_1(\hat{\theta}_n,\theta)\{1+\frac{||\nabla_\theta\xi(\theta)||}{\xi(\theta)}\}, \tag{15}$$

where $C_1(\hat{\theta}_n,\theta)$ is jointly continuous and $J(\hat{\theta}_n,\theta)=\Pi_{l=1}^k J^l(\hat{\theta}_n,\theta)$ is as in (5). Similarly,

$$||\Gamma_2^{\xi}(\hat{\theta}_n,\theta)|| = ||n\frac{\nabla_z\{(D_\theta z)^{-1}[\xi(\theta)\nabla_\theta J(\hat{\theta}_n,\theta)+J(\hat{\theta}_n,\theta)\nabla_\theta\xi(\theta)]\}}{J(\hat{\theta}_n,\theta)\xi(\theta)}||$$
$$\leq C_2(\hat{\theta}_n,\theta)\{1+\frac{||\nabla_\theta\xi(\theta)||}{\xi(\theta)}+\frac{||\nabla_\theta^2\xi(\theta)||}{\xi(\theta)}\}. \tag{16}$$

In the Proposition below, let B_n denote the event $\{\hat{\theta}_n \in \nabla\psi(\Omega)\}$.

Proposition 2. *Suppose that $\nabla\xi$ is continuous. Then*

$$E_\xi^n\{h(Z_n)\} = \Phi_k h + \frac{1}{\sqrt{n}}E_\xi^n\{[Uh(Z_n)]'\Gamma_1^{\xi}(\hat{\theta}_n,\theta)\}, \tag{17}$$

a.e. on B_n, for all $h \in H$. If also $\nabla^2\xi$ is continuous, then

$$E_\xi^n\{h(Z_n)\} = \Phi_k h + \frac{1}{\sqrt{n}}(\Phi_k Uh)'E_\xi^n\{\Gamma_1^{\xi}(\hat{\theta}_n,\theta)\} + \frac{1}{n}tr\{E_\xi^n\{Vh(Z_n)\Gamma_2^{\xi}(\hat{\theta}_n,\theta)\}\} \tag{18}$$

a.e. on B_n, for all $h \in H$.

Proof. We verify (17). The proof of (18) is similar. Fix an $h \in H$, so that $h \in H_r$ for some r. Then, by Lemma 1,

$$E_\xi^n\{h(Z_n)\} = \int_{\Re^k} h(z)f_n(z)\Phi_k(dz)$$
$$= \Phi_k h + \int_{\Re^k} Uh(z)'\frac{\nabla f_n(z)}{f_n(z)}f_n(z)\Phi_k(dz) = \Phi_k h + E_\xi^n\{Uh(Z_n)'\frac{\nabla f_n(Z_n)}{f_n(Z_n)}\} \tag{19}$$

provided

$$\int_{\Re^k}(1+||z||^r)||\nabla f(z)||\Phi_k(dz) < \infty. \tag{20}$$

So, it suffices to show that (20) holds. Note that from (12), we have $\nabla f_n(Z_n)/f_n(Z_n) = \Gamma_1^{\xi}(\hat{\theta}_n,\theta)/n^{1/2}$. To verify (20), it suffices to show that $E_\xi^n\{(1+||Z_n||^r)||\Gamma_1^{\xi}(\hat{\theta}_n,\theta)||\} < \infty$, *a.e.* on B_n.

Let K denote the support of ξ, so that K is a compact subset of Ω. For fixed $x_1,\ldots,x_n$, Z_n is a continuous function of θ and hence bounded on K. Similarly,

$C_1(\hat{\theta}_n, \theta)$ is bounded on K. So, there is a constant C, depending on $x_1, \ldots, x_n$ but not on θ, for which

$$(1 + ||Z_n||^r)||\Gamma_1^\xi(\hat{\theta}_n, \theta)|| \leq C(1 + \frac{||\nabla\xi(\theta)||}{\xi(\theta)})$$

for all $\theta \in K$ and, therefore,

$$E_\xi^n\{(1 + ||Z_n||^r)||\Gamma_1^\xi(\hat{\theta}_n, \theta)||\} \leq C' \int_K (1 + \frac{||\nabla\xi(\theta)||}{\xi(\theta)})e^{L_n(\theta)}\xi(\theta)d\theta < \infty.$$

Relation (17) follows by writing $\nabla f_n(Z_n)/f_n(Z_n) = \Gamma_1^\xi(\hat{\theta}_n, \theta)/n^{1/2}$ in (19).

Corollary 3. *Suppose that* $h(Z_n) = h_0(Z_{n1})$, *where* $h_0 : \mathcal{R} \to \mathcal{R}$. *Then* (17) *and* (18) *reduce to*

$$E_\xi^n\{h_0(Z_{n1})\} = \Phi h_0 + \frac{1}{n^{1/2}} E_\xi^n\{Uh_0(Z_{n1})\Gamma_{1,1}^\xi(\hat{\theta}_n, \theta)\}$$

and

$$E_\xi^n\{h_0(Z_{n1})\} = \Phi h_0 + \frac{1}{n^{1/2}}\Phi U h_0 E_\xi^n\{\Gamma_{1,1}^\xi(\hat{\theta}_n, \theta)\} + \frac{1}{n}E_\xi^n\{Vh_0(Z_{n1})\Gamma_{2,11}^\xi(\hat{\theta}_n, \theta)\}.$$

4. Asymptotic Expansions for Bounded h

In this section we establish the first and second order expansions for the posterior expectation of $h(Z_t)$ when h is a bounded function and t is a stopping time. Let $t = t_a$ be a family of stopping times depending on a parameter $a \geq 1$. Suppose that

$$\frac{a}{t_a} \to \rho^2(\theta)$$

in P_θ-probability for almost every $\theta \in \Omega$, where ρ is a continuous function on Ω. Suppose also that for every compact $K \subseteq \Omega$ there is an $\eta > 0$ such that

$$P_\theta\{t_a \leq \eta a\} = o(a^{-q}), \tag{21}$$

uniformly with respect to $\theta \in K$ as $a \to \infty$, for some $q > 1/2$ (q may depend on K). In the theorem below, let $h : \Re^k \to \Re$ be a bounded measurable function, $R_{0,a}(h) = E_\xi^t[h(Z_t) - \Phi_k h]$, and $\bar{R}_{0,a} = \text{essup}_{h \in H_0}|R_{0,a}(h)|$.

Lemma 4. $\lim_{a\to\infty} E_\xi\{\bar{R}_{0,a}\} = 0.$

Proof. By Proposition 2, $\bar{R}_{0,a} \leq CE_\xi^t(||\Gamma_1^\xi(\hat{\theta}_t, \theta)||)/t^{1/2}$ for some constant C. So $\bar{R}_{0,a} \to 0$ in P_ξ-probability and the result follows from the Bounded Convergence Theorem, since $\bar{R}_{0,a}$ is bounded.

Lemma 5. *Let K and K_1 be compact sets for which $K \subseteq K_1^0 \subseteq K_1 \subseteq \Omega$, where K_1^0 denotes the interior of K_1. Then there are constants C and $\delta > 0$ for which*

$$\sup_{\theta \in K} P_\theta\{t > \eta a, \hat{\theta}_t \notin K_1\} \leq Ce^{-\delta \eta a},$$

for all a.

Proof. By Bernstein's Inequality applied to the coordinates of $\bar{X}_n$, there exist $\epsilon_\theta > 0$ and $\delta_\theta > 0$ such that $P_\theta(\hat{\theta}_n \notin K_1) \leq P_\theta(||\bar{X}_n - \nabla\psi(\theta)|| > \epsilon_\theta) \leq e^{-n\delta_\theta}$, for every $\theta \in K$. Let $\delta = \inf_{\theta \in K} \delta_\theta$. Then $\delta > 0$ by compactness of K and hence

$$P_\theta(t > \eta a, \hat{\theta}_t \notin K_1) \leq \sum_{n > \eta a} P_\theta(\hat{\theta}_n \notin K_1) \leq \sum_{n > \eta a} e^{-n\delta} \leq Ce^{-\delta \eta a}.$$

Now let

$$R_{1,a}(h) = a^{1/2}\{E_\xi^t(h(Z_t)) - \Phi_k h - \frac{1}{a^{1/2}} E_\xi^t[\rho(\theta)(\Phi_k Uh)' \Gamma_1^\xi(\theta, \theta)]\}.$$

By (17), $R_{1,a}(h) = R_{1,a}^1(h) + R_{1,a}^2(h) + R_{1,a}^3(h)$, where

$$R_{1,a}^1(h) = (\frac{a}{t})^{1/2} E_\xi^t\{(Uh(Z_t))'[\Gamma_1^\xi(\hat{\theta}_t, \theta) - E_\xi^t(\Gamma_1^\xi(\theta, \theta))]\},$$
$$R_{1,a}^2(h) = (\frac{a}{t})^{1/2} E_\xi^t\{(Uh(Z_t) - \Phi_k Uh)'\} E_\xi^t\{\Gamma_1^\xi(\theta, \theta)\},$$
$$R_{1,a}^3(h) = (\Phi_k Uh)' E_\xi^t\{[(\frac{a}{t})^{1/2} - \rho(\theta)]\Gamma_1^\xi(\theta, \theta)\}.$$

Then let $\bar{R}_{1,a} = \text{essup}_{h \in H_0} |R_{1,a}(h)|$, and $\bar{R}_{1,a}^i = \text{essup}_{h \in H_0} |R_{1,a}^i(h)|$.

Theorem 6. *If (21) holds for some $q > 1/2$ and $\nabla\xi$ is continuous, then $\lim_{a \to \infty} E_\xi\{\bar{R}_{1,a}\} = 0$.*

Proof. Let K denote the compact support of ξ; let K_1 be another compact set for which $K \subseteq K_1^0 \subseteq K_1 \subseteq \Omega$; and let B_a be the event $\{t > \eta a, \hat{\theta}_t \in K_1\}$. Then

$$E_\xi\{\bar{R}_{1,a}\} = \int_{\{t_a \leq \eta a\}} \bar{R}_{1,a} dP_\xi + \int_{\{t_a > \eta a, \hat{\theta}_t \notin K_1\}} \bar{R}_{1,a} dP_\xi + \int_{B_a} \bar{R}_{1,a} dP_\xi.$$

Here

$$\begin{aligned}\int_{\{t_a \leq \eta a\}} \bar{R}_{1,a} dP_\xi &\leq Ca^{1/2} \int_{\{t_a \leq \eta a\}} \bar{R}_{0,a} dP_\xi + C \int_K \rho(\theta)||\Gamma_1^\xi(\theta, \theta)|| P_\theta(t_a \leq \eta a)\xi(\theta) d\theta \\ &\leq C' a^{1/2} P_\xi(t_a \leq \eta a) + C' \sup_{\theta \in K} P_\theta(t_a \leq \eta a),\end{aligned}$$

for some constants C and C', since $\rho(\theta)\Gamma_1^\xi(\theta, \theta)\xi(\theta)$ is continuous on Ω and, therefore, bounded on K, and the right side approaches zero by (21). Similarly,

$$\int_{\{t_a > \eta a, \hat{\theta}_t \notin K_1\}} \bar{R}_{1,a} dP_\xi \leq C' a^{1/2} \sup_{\theta \in K} P_\theta(t_a > \eta a, \hat{\theta}_t \notin K_1) \to 0,$$

by Lemma 5.

For the integral over B_a, we consider $\bar{R}^i_{1,a}$ separately. Start with $R^1_{1,a}$. Observe that $\lim_{a\to\infty}\Gamma^\xi_1(\hat{\theta}_t,\theta)=\Gamma^\xi_1(\theta,\theta)=\lim_{a\to\infty}E^t_\xi(\Gamma^\xi_1(\theta,\theta))$, by the consistency of the maximum likelihood estimator and the Martingale Convergence Theorem. Moreover, from (15) and the continuity of $\Gamma^\xi_1(\theta,\theta)$, there is a constant C, depending on K_1, for which

$$||\Gamma^\xi_1(\hat{\theta}_t,\theta)||+||\Gamma^\xi_1(\theta,\theta)||\le C(1+||\frac{\nabla\xi}{\xi}(\theta)||)$$

a.e. on B_a, and the right side is integrable with respect to P_ξ, since $\nabla\xi$ is continuous on K. So

$$\begin{aligned}\int_{B_a}\bar{R}^1_{1,a}dP_\xi &\le \frac{C}{\eta^{1/2}}\int_{B_a}||\Gamma^\xi_1(\hat{\theta}_t,\theta)-E^t_\xi(\Gamma^\xi_1(\theta,\theta))||dP_\xi\\ &\le \frac{C}{\eta^{1/2}}\int_{B_a}\{||\Gamma^\xi_1(\hat{\theta}_t,\theta)-\Gamma^\xi_1(\theta,\theta)||+||\Gamma^\xi_1(\theta,\theta)-E^t_\xi(\Gamma^\xi_1(\theta,\theta))||\}dP_\xi,\end{aligned}$$

where the second term approaches zero as $a\to\infty$ because of uniformly integrability, and the first term by the Dominated Convergence Theorem. Next,

$$|\bar{R}^2_{1,a}|\le\frac{1}{\eta^{1/2}}\mathrm{essup}_{h\in H_0}|E^t_\xi\{(Uh(Z_t)-\Phi_kUh)'\}E^t_\xi\{\Gamma^\xi_1(\theta,\theta)\}|$$

on B_a. So,

$$\int_{B_a}\bar{R}^2_{1,a}dP_\xi\le CE_\xi\{\bar{R}_{0,a}E^t_\xi||\Gamma^\xi_1(\theta,\theta)||\}$$

where the right side approaches zero, since $\bar{R}_{0,a}\to 0$, $\bar{R}_{0,a}\le 2$, and $E^t_\xi||\Gamma^\xi_1(\theta,\theta)||$ is uniformly integrable. For $\bar{R}^3_{1,a}$,

$$\begin{aligned}\int_{B_a}\bar{R}^3_{1,a}dP_\xi &\le C\int_K||\Gamma^\xi_1(\theta,\theta)||E_\theta\{|(\frac{a}{t})^{1/2}-\rho(\theta)|1_{\{t>\eta a\}}\}\xi(\theta)d\theta\\ &\le C'\int_K E_\theta\{|(\frac{a}{t})^{1/2}-\rho(\theta)|1_{\{t>\eta a\}}\}d\theta\to 0,\end{aligned}$$

since $\Gamma^\xi_1(\hat{\theta}_t,\theta)\xi(\theta)$ is bounded on $\hat{\theta}_t\in K_1$ and $\theta\in K$, $(a/t_a)^{1/2}-\rho(\theta)\to 0$ in P_θ-probability for almost every $\theta\in\Omega$, and a/t_a is bounded over $\{t>\eta a\}$.

For the second order approximation, let

$$\begin{aligned}R_{2,a}(h)=a\{&E^t_\xi(h(Z_t))-\Phi_kh-t^{-1/2}(\Phi_kUh)'E^t_\xi(\Gamma^\xi_1(\theta,\theta))\\ &-\frac{1}{a}tr\{\Phi_kVhE^t_\xi[\rho^2(\theta)\Gamma^\xi_2(\theta,\theta)]\}\},\end{aligned}$$

and

$$\bar{R}^{(s)}_{2,a} = \text{essup}_{h\in H_0^s}|R_{2,a}(h)|,$$

where H_0^s denotes the set of bounded symmetric functions in H_0.

Theorem 7. *If* (21) *holds for some* $q > 1$ *and* $\nabla^2\xi$ *is continuous, then* $\lim_{a\to\infty} E_\xi\{\bar{R}^{(s)}_{2,a}\} = 0$.

Proof. The analysis for $\{t_a \le \eta a\} \cup \{\hat{\theta}_t \notin K_1\}$ is similar to that in Theorem 6. For $\{t_a > \eta a\} \cap \{\hat{\theta}_t \in K_1\}$, let h be a bounded symmetric measurable function and write

$$R_{2,a}(h) = \frac{a}{t} tr\{E^t_\xi[Vh(Z_t)\Gamma^\xi_2(\hat{\theta}_t,\theta)]\} - tr\{(\Phi_k Vh)E^t_\xi[\rho^2(\theta)\Gamma^\xi_2(\theta,\theta)]\},$$

by (18), since $\Phi_k Uh = 0$ for symmetric h.

Then, decompose $R_{2,a}(h)$ as $R_{2,a}(h) = R^1_{2,a}(h) + R^2_{2,a}(h) + R^3_{2,a}(h)$, where

$$\begin{aligned}
R^1_{2,a}(h) &= \frac{a}{t} tr\{E^t_\xi([\Gamma^\xi_2(\hat{\theta}_t,\theta) - E^t_\xi(\Gamma^\xi_2(\theta,\theta))]Vh(Z_t))\},\\
R^2_{2,a}(h) &= \frac{a}{t} tr\{E^t_\xi[Vh(Z_t) - (\Phi_k Vh)]E^t_\xi[\Gamma^\xi_2(\theta,\theta)]\},\\
R^3_{2,a}(h) &= tr\{(\Phi_k Vh)E^t_\xi\{[\frac{a}{t} - \rho^2(\theta)]\Gamma^\xi_2(\theta,\theta)\}\},
\end{aligned}$$

and, for $i = 1,2,3$, define $\bar{R}^{(s),i}_{2,a} = \text{essup}_{h\in H_0^s}|R^i_{2,a}(h)|$. Then, the analyses of $\bar{R}^{(s),1}_{2,a}$, $\bar{R}^{(s),2}_{2,a}$, and $\bar{R}^{(s),3}_{2,a}$ are similar to those in Theorem 6.

5. Some Bounds

For unbounded h, it is necessary to establish some uniform integrability of powers of $||Z_t||$. Let Ξ be the collection of all twice continuously differentiable prior densities ξ with compact support $K_\xi \in \Omega$, and let $\mathcal{T}$ be any collection of stopping times.

Lemma 8. *If* $h(z) = ||z||^p$, *where* $p \ge 1$, *then* $||Uh(z)|| \le C_1\{1 + ||z||^{p-1}\}$ *for all* $z \in \Re^k$.

Proof. The details of the proof are slightly different for even and odd p. They are given here for even p only. Let $p = 2v$, $v \ge 1$. It will be shown that

$$g_j(y_1,\ldots,y_k) \le C\{1 + ||y||^{2v-1}\} \tag{22}$$

for positive y_j for each $j = 1,\ldots,k$. A similar result may be obtained for negative y_j and the Lemma then follows. Note that $||z||^{2v} = (\sum_{i=1}^k z_i^2)^v$ is a polynomial of degree $2v$. From (7) and (8),

$$h_j(y_1,\ldots,y_j) = \sum_{l=0}^{v} \binom{v}{l} (y_1^2 + \cdots + y_j^2)^l \alpha_{k-j,v-l}$$

where α_{ij} is the jth moment of χ_i^2. Thus

$$
\begin{aligned}
g_j(y_1,\ldots,y_k) &= e^{\frac{1}{2}y_j^2}\int_{y_j}^{\infty}[h_j(y_1,\ldots,y_{j-1},w)-h_{j-1}(y_1,\ldots,y_{j-1})]e^{-\frac{1}{2}w^2}dw\\
&= e^{\frac{1}{2}y_j^2}\int_{y_j}^{\infty}[b_0(y_1,\ldots,y_{j-1})+w^2b_1(y_1,\ldots,y_{j-1})+\cdots\\
&\quad +w^{2v-2}b_{v-1}(y_1,\ldots,y_{j-1})+w^{2v}]e^{-\frac{1}{2}w^2}dw
\end{aligned}
$$

where $b_0(y_1,\ldots,y_{j-1})$ is a polynomial of degree $2v-2$ and $b_i(y_1,\ldots,y_{j-1})$ is a polynomial of degree $2(v-i)$, for $i=1,\ldots,v-1$. As in Woodroofe (1992), it is easily seen that

$$e^{\frac{1}{2}y^2}\int_y^{\infty}e^{-\frac{1}{2}w^2}dw \le C$$

and

$$e^{\frac{1}{2}y^2}\int_y^{\infty}w^{2q}e^{-\frac{1}{2}w^2}dw \le C(1+w^{2q-1})$$

for all $q\ge 0$, and (22) follows easily.

Proposition 9. *For every $\xi\in\Xi$, every compact $J\subset\Omega$, and every $p\ge 1$,*

$$\sup_{t\in\mathcal{T}}\int_{\{\hat\theta_t\in J\}}||Z_t||^p dP_\xi<\infty. \tag{23}$$

Proof. We first verify (23) when $p=2$. Let $h(z)=||z||^2$, $z\in\Re^k$. Then $Vh(z)$ is the identity matrix for all z. In view of (16), (18) and the symmetry of h,

$$
\begin{aligned}
E_\xi^n(||Z_n||^2) &= k+\frac{1}{n}E_\xi^n\{tr[Vh(Z_n)\Gamma_2^\xi(\hat\theta_n,\theta)]\}\\
&\le C\{1+E_\xi^n[||\frac{\nabla\xi}{\xi}(\theta)||]+E_\xi^n[||\frac{\nabla^2\xi}{\xi}(\theta)||]\}
\end{aligned}
$$

for some constant C depending on J, provided $\hat\theta_n\in J$ for each $n\ge 1$. So, if t is any stopping time, then

$$
\begin{aligned}
\int_{\{\hat\theta_t\in J\}}||Z_t||^2dP_\xi &\le C\{1+E_\xi[E_\xi^t(||\frac{\nabla\xi}{\xi}(\theta)||)]+E_\xi[E_\xi^t(||\frac{\nabla^2\xi}{\xi}(\theta)||)]\}\\
&= C\{1+\int_{K_\xi}||\nabla\xi(\theta)||d\theta+\int_{K_\xi}||\nabla^2\xi(\theta)||d\theta\}
\end{aligned}
$$

for all $t\in\mathcal{T}$. Since the right side is finite and does not depend on t, this establishes (23) for $p=2$. Now suppose that (23) holds for all $\xi\in\Xi$, for a given $p\ge 1$. Let $h(z)=||z||^{p+1}$, for $z\in\Re^k$. Then $||Uh(z)||\le C(1+||z||^p)$ for $z\in\Re^k$

by Lemma 8. So,

$$E_\xi^n(||Z_n||^{p+1}) = \Phi_k h + n^{-1/2} E_\xi^n\{Uh(Z_n)'\Gamma_1^\xi(\hat{\theta}_n, \theta)\}$$
$$\leq \Phi_k h + C' E_\xi^n\{||Z_n||^p[1 + ||\frac{\nabla \xi}{\xi}(\theta)||]\}$$

for some constant C' depending on J, provided $\hat{\theta}_n \in J$. Let $\tilde{\xi}$ be a twice continuously differentiable compactly supported density that is positive on K_ξ. Then $[\xi(\theta) + ||\nabla\xi(\theta)||]/\tilde{\xi}(\theta)$ is bounded on the support of $\tilde{\xi}$, and

$$E_\xi^t(||Z_t||^{p+1}) \leq C''\{1 + \int E_\theta(||Z_t||^p 1_{\{\hat{\theta}_t \in J\}})[\xi(\theta) + ||\nabla\xi(\theta)||]d\theta\}$$
$$\leq C'''\{1 + \int_{\{\hat{\theta}_t \in J\}} ||Z_t||^p dP_{\tilde{\xi}}\},$$

which is bounded with respect to $t \in \mathcal{T}$ by the induction hypothesis.

We wish to investigate some global properties of Z_n and will start it out at a particular point θ_0. From (1), letting $\lambda = \lambda(\theta) = \theta - \theta_0$, $y = x - \nabla\psi(\theta_0)$, and

$$\psi^*(\lambda) = \psi(\lambda + \theta_0) - \psi(\theta_0) - \lambda'\nabla\psi(\theta_0), \tag{24}$$

we have $p_\lambda^*(y) = e^{\lambda' y - \psi^*(\lambda)}$, with respect to some σ-finite dominating measure μ^*. Observe that $\lambda_0 = \lambda(\theta_0) = 0$, $\psi^*(\lambda_0) = 0$, and $\nabla\psi^*(\lambda_0) = 0$. Let L_n^* be the corresponding log-likelihood function. So, under θ_0, $L_n^*(\lambda_0) = 0$ and

$$||Z_n||^2 = 2(L_n^*(\hat{\lambda}_n) - L_n^*(\lambda_0)) = 2n(\hat{\lambda}_n'\bar{Y}_n - \psi^*(\hat{\lambda}_n)) \leq 2n||\hat{\lambda}_n||||\bar{Y}_n||, \tag{25}$$

where $\hat{\lambda}_n = \hat{\theta}_n - \theta_0$ and $\bar{Y}_n = \bar{X}_n - \nabla\psi(\theta_0)$.

Assumption (1): For $||\theta||$ sufficiently large, $\psi(\theta) \geq c||\theta||^{1+\alpha}$, for some $c > 0$, $\alpha > 0$.

Lemma 10. *Suppose that Assumption* (1) *holds. Then* $\psi^*(\lambda) \geq c^*||\lambda||^{1+\alpha}$, *for all* λ, *for some* $c^* > 0$.

Proof. It follows directly from (24).

Proposition 11. *Suppose that Assumption* (1) *holds. If* $K_1 \in \Omega$ *is compact, then*

$$\sup_{\theta_0 \in K_1} \sup_n \int_{\{\hat{\theta}_n \notin K_1\}} ||Z_n||^p dP_{\theta_0} < \infty.$$

Proof. By reparametrization and transformation, as described in a previous paragraph, we have (25). Further, observe that $||\hat{\lambda}_n||||\bar{Y}_n|| \geq \psi^*(\hat{\lambda}_n) \geq c^*||\hat{\lambda}_n||^{1+\alpha}$, where the first inequality follows from the fact that $L_n^*(\hat{\lambda}_n) > 0$, and the second from Lemma 10. So, $||\hat{\lambda}_n|| \le (1/c^*)||\bar{Y}_n||^{1/\alpha}$ and $||Z_n||^p \le Cn^{p/2}||\bar{Y}_n||^{(1+1/\alpha)p/2}$, for some $C > 0$. By Hölder's inequality and Lemma 5 we have

$$\int_{\{\hat{\theta}_n \notin K_1\}} ||Z_n||^p dP_{\theta_0} \le Cn^{p/2}\{E_{\theta_0}(||\bar{Y}_n||^{(1+1/\alpha)p})P_{\theta_0}(\hat{\theta}_n \notin K_1)\}^{1/2}$$
$$\le Cn^{p/2}e^{-n\delta/2}\{E_{\theta_0}(||\bar{Y}_n||^{(1+1/\alpha)p})\}^{1/2}.$$

The Proposition follows since $E_{\theta_0}(||\bar{Y}_n||^q)$ is bounded with respect to n and $\theta_0 \in K_1$ for any $q > 0$.

Theorem 12. *Suppose that Assumption* (1) *holds. Then* $\sup_{t\in\mathcal{T}} E_\xi\{||Z_t||^p\} < \infty$.

Proof. Let K be the compact support of ξ and let K_1 be another compact set for which $K \subseteq K_1^0 \subseteq K_1 \subseteq \Omega$. By Proposition 9, it suffices to show that

$$\sup_{t\in\mathcal{T}} \int_{\{\hat{\theta}_t \notin K_1\}} ||Z_t||^p dP_\xi < \infty.$$

Observe that

$$\int_{\{\hat{\theta}_t \notin K_1\}} ||Z_t||^p dP_\xi \le \sum_{n=1}^{\infty} \int_{\{\hat{\theta}_n \notin K_1\}} ||Z_n||^p dP_\xi$$
$$\le \sum_{n=1}^{\infty} \{P_\xi(\hat{\theta}_n \notin K_1) \int_{\{\hat{\theta}_n \notin K_1\}} ||Z_n||^{2p} dP_\xi\}^{1/2},$$

by Hölder's inequality. So, the result follows by Lemma 5 and Proposition 11.

6. Two-Parameter Case

The results for two-parameter case will be stated in greater detail. In this section we suppose that θ_1 is the parameter of primary interest and that θ_2 is a nuisance parameter. Throughout it is assumed that ρ is almost differentiable with respect to θ_1 and θ_2. Denote ψ_{ij} as the partial derivatives, $\psi_{ij}(\theta) = \partial^{i+j}\psi(\theta)/\partial\theta_1^i\partial\theta_2^j$, and similarly for ξ_{ij}. Next denote $\Gamma_{1,1}^\xi$ as the first component of Γ_1^ξ and denote $B(\hat{\theta}, \theta) = n^{-1/2}[\partial Z_{n2}]/[\partial\theta_1]$. Then let $B_{ij}(\hat{\theta}, \theta) = [\partial^{i+j}B(\hat{\theta}, \theta)]/[\partial\theta_1^i\partial\theta_2^j]$ and $J_{ij}^l(\hat{\theta}_n, \theta) = [\partial^{i+j}J^l(\hat{\theta}_n, \theta)]/[\partial\theta_1^i\partial\theta_2^j]$, for $l = 1, 2$. In view of (12), (14), the relation $f_n \propto \xi J^1 J^2$, and $J_{01}^1(\hat{\theta}_n, \theta) = 0$, we have

$$\Gamma_{1,1}^\xi(\hat{\theta}_n, \theta) = \frac{\xi_{10}(\theta)}{\xi(\theta)}J^1 + J_{10}^1 + \frac{J^1}{J^2}J_{10}^2 - [\frac{\xi_{01}(\theta)}{\xi(\theta)}J^1J^2 + J^1J_{01}^2]B, \qquad (26)$$

where B, J^l, and J_{ij}^l are abbreviations for $B(\hat{\theta}_n, \theta)$, $J^l(\hat{\theta}_n, \theta)$, and $J_{ij}^l(\hat{\theta}_n, \theta)$. Now, let $g_1(\theta) = (\psi_{20} - \psi_{11}^2/\psi_{02})(\theta)$ and $g_2(\theta) = \psi_{02}(\theta)$. Employing L'Hospital's rule one can obtain that

$$\begin{aligned}
J^1(\theta,\theta) &= \lim_{\omega\to\theta} J^1(\omega,\theta) = \{g_1(\theta)\}^{-1/2},\\
J^2(\theta,\theta) &= \lim_{\omega\to\theta} J^2(\omega,\theta) = \{g_2(\theta)\}^{-1/2},\\
J^1_{10}(\theta,\theta) &= \lim_{\omega\to\theta} J^1_{10}(\omega,\theta) = \frac{(-\psi_{02},\psi_{11})\cdot\nabla g_1}{3\psi_{02}g_1^{3/2}},\\
J^2_{10}(\theta,\theta) &= \lim_{\omega\to\theta} J^2_{10}(\omega,\theta) = \frac{(-3\psi_{02},\psi_{11})\cdot\nabla g_2}{6\psi_{02}g_2^{3/2}},\\
J^2_{01}(\theta,\theta) &= \lim_{\omega\to\theta} J^2_{01}(\omega,\theta) = \frac{-\psi_{03}}{3g_2^{3/2}},\\
B(\theta,\theta) &= \lim_{\omega\to\theta} B(\omega,\theta) = \frac{\psi_{11}}{g_2^{1/2}}.
\end{aligned}$$

Let $\Gamma^{\xi}_{1,1}(\theta,\theta)=\lim_{\omega\to\theta}\Gamma^{\xi}_{1,1}(\omega,\theta)$. Observe that $E^t_{\xi}\{\rho(\theta)\Gamma^{\xi}_{1,1}(\theta,\theta)\}\to\rho(\theta)\Gamma^{\xi}_{1,1}(\theta,\theta)$, w.p.1 P_{ξ}, by the Martingale Convergence Theorem. Since ρ is assumed to be almost differentiable with respect to θ_1 and θ_2, an integration by parts yields

$$\begin{aligned}
&E_{\xi}(\rho(\theta)\Gamma^{\xi}_{1,1}(\theta,\theta))\\
&=\int\int \xi(\theta)\{-\frac{\partial}{\partial\theta_1}[J^1(\theta,\theta)\rho(\theta)] + \frac{\partial}{\partial\theta_2}[(J^1J^2B)(\theta,\theta)\rho(\theta)]\\
&\quad +\rho(\theta)[J^1_{10}(\theta,\theta) + (\frac{J^1}{J^2}J^2_{10})(\theta,\theta) - (J^2_{01}J^1B)(\theta,\theta)]\}d\theta_1 d\theta_2\\
&=\int\int \xi(\theta)\kappa_1(\theta)d\theta_1 d\theta_2\\
&=\bar{\kappa}_1(\xi),
\end{aligned}$$

where

$$\begin{aligned}
\kappa_1(\theta) &= -\frac{\partial}{\partial\theta_1}[J^1(\theta,\theta)\rho(\theta)] + \frac{\partial}{\partial\theta_2}[(J^1J^2B)(\theta,\theta)\rho(\theta)]\\
&\quad +\rho(\theta)[J^1_{10}(\theta,\theta) + (\frac{J^1}{J^2}J^2_{10})(\theta,\theta) - (J^2_{01}J^1B)(\theta,\theta)] \qquad (27)\\
&= -\frac{\partial}{\partial\theta_1}[\frac{\rho}{g_1^{1/2}}(\theta)] + \frac{\partial}{\partial\theta_2}[\frac{\rho\psi_{11}}{g_1^{1/2}\psi_{02}}(\theta)] + \rho(\theta)[\frac{(-\psi_{02},\psi_{11})\cdot\nabla g_1}{3\psi_{02}g_1^{3/2}}(\theta)\\
&\quad +\frac{(-3\psi_{02},\psi_{11})\cdot\nabla g_2}{6\psi_{02}^2 g_1^{1/2}}(\theta) - \frac{(0,-\psi_{11})\cdot\nabla g_2}{3\psi_{02}^2 g_1^{1/2}}(\theta)]\\
&= \frac{(-\psi_{02},\psi_{11})\cdot\nabla\rho}{\psi_{02}g_1^{1/2}}(\theta) + \rho(\theta)[\frac{(\psi_{02},-\psi_{11})\cdot\nabla g_1}{6\psi_{02}g_1^{3/2}}(\theta) + \frac{(\psi_{02},-\psi_{11})\cdot\nabla g_2}{2\psi_{02}^2 g_1^{1/2}}(\theta)].
\end{aligned}$$

Particularly, for the normal translation family $N(\theta, \Sigma)$ both $J^1(\theta,\theta)$ and $J^2(\theta,\theta)$ are constants. In such cases, $\kappa_1(\theta)$ has a simpler form

$$\kappa_1(\theta) = \frac{(-\psi_{02}, \psi_{11}) \cdot \nabla\rho}{\psi_{02} g_1^{1/2}}(\theta),$$

which vanishes if no stopping rule is adopted.

Now let $h : \Re \to \Re$ be a bounded measurable function. Then from Theorem 6 and Corollary 3,

$$E_\xi\{h(Z_{t1})\} = \Phi h + a^{-1/2}(\Phi U h)\bar{\kappa}_1(\xi) + o(a^{-1/2}), \tag{28}$$

for all twice continuously differentiable compactly supported densities ξ. Recalling the definition of $\bar{\kappa}_1$, Woodroofe (1986) writes relations in (28) as $E_\theta\{h(Z_{t1})\} = \Phi h + a^{-1/2}(\Phi U h)\kappa_1(\theta) + o(a^{-1/2})$ *very weakly*.

The next two paragraphs include assertions that will not be proved. These are used to motivate the definition of Z_{t1}^* in Theorem 14, which will be proved. Note that if $h(z) = z$, then $\Phi h = 0$ and $Uh(z) = 1 = \Phi U h$. Formally applying (28) to this h suggests $E_\theta(Z_{t1}) \approx a^{-1/2}\kappa_1(\theta)\;\; v.w.$ Let

$$\hat{\mu}_a = \begin{cases} \hat{\kappa}_1^{a^{-1/2}} & \text{if } |\hat{\kappa}_1| \le a^{1/6}(\log(a))^{-1}, \\ a^{-1/3}(\log(a))^{-1} & \text{if } \hat{\kappa}_1 \;\; > a^{1/6}(\log(a))^{-1}, \\ -a^{-1/3}(\log(a))^{-1} & \text{if } \hat{\kappa}_1 \;\; < -a^{1/6}(\log(a))^{-1}, \end{cases} \tag{29}$$

where $\hat{\kappa}_1 = \kappa_1(\hat{\theta}_t)$ and consider $(Z_{t1} - \hat{\kappa_1})a^{-1/2}$. We have

$$E_\xi^t\{(Z_{t1} - \hat{\kappa_1}^{a^{-1/2}})^2\} = E_\xi^t(Z_{t1}^2) - 2\hat{\kappa}_1^{a^{-1/2}} E_\xi^t(Z_{t1}) + \frac{\hat{\kappa}_1^2}{a}. \tag{30}$$

If $h(z) = z^2$, we have $\Phi h = 1$, $\Phi U h = 0$, and $Vh(z) = 1$. Specializing (18) to h leads to $E_\xi^t(Z_{t1}^2) = 1 + t^{-1} E_\xi^t\{\Gamma_{2,11}^\xi(\hat{\theta}_t, \theta)\}$, where

$$\Gamma_{2,11}^\xi(\hat{\theta}_n, \theta) = n\frac{\partial^2 f_n(Z_n)/\partial Z_{n1}^2}{f_n(Z_n)}.$$

Now, we show how to obtain $\partial^2 f_n(Z_n)/\partial Z_{n1}^2$. First, from (14), we have

$$\frac{\partial f_n}{\partial z_1} = \frac{\partial \theta_1}{\partial z_1}\frac{\partial f_n}{\partial \theta_1} + \frac{\partial \theta_2}{\partial z_1}\frac{\partial f_n}{\partial \theta_2}. \tag{31}$$

Then we can derive that

$$\begin{aligned} \frac{\partial^2 f_n}{\partial z_1^2} &= \frac{\partial}{\partial z_1}(\frac{\partial f_n}{\partial z_1}) \\ &= \frac{\partial}{\partial z_1}[\frac{\partial \theta_1}{\partial z_1}\frac{\partial f_n}{\partial \theta_1} + \frac{\partial \theta_2}{\partial z_1}\frac{\partial f_n}{\partial \theta_2}] \end{aligned}$$

$$= \frac{\partial\theta_1}{\partial z_1}[\frac{\partial}{\partial\theta_1}(\frac{\partial\theta_1}{\partial z_1}\frac{\partial f_n}{\partial\theta_1})] + \frac{\partial\theta_2}{\partial z_1}[\frac{\partial}{\partial\theta_2}(\frac{\partial\theta_1}{\partial z_1}\frac{\partial f_n}{\partial\theta_1})]$$
$$+\frac{\partial\theta_1}{\partial z_1}[\frac{\partial}{\partial\theta_1}(\frac{\partial\theta_2}{\partial z_1}\frac{\partial f_n}{\partial\theta_2})] + \frac{\partial\theta_2}{\partial z_1}[\frac{\partial}{\partial\theta_2}(\frac{\partial\theta_2}{\partial z_1}\frac{\partial f_n}{\partial\theta_2})],$$

where the third equation follows by replacing f_n in (31) by $\partial f_n/\partial z_1$. So,

$$\begin{aligned}\Gamma^{\xi}_{2,11}(\hat\theta_n,\theta) = {}& J^1 J^1_{10}\frac{1}{f_n}\frac{\partial f_n}{\partial\theta_1} + (J^1)^2\frac{1}{f_n}\frac{\partial^2 f_n}{\partial\theta_1^2} - 2(J^1)^2J^2B\frac{1}{f_n}\frac{\partial^2 f_n}{\partial\theta_1\partial\theta_2}\\ &-J^1[J^1_{10}J^2B + J^1J^2_{10}B + J^1J^2B_{10}]\frac{1}{f_n}\frac{\partial f_n}{\partial\theta_2}\\ &+J^1J^2B[J^1J^2_{01}B + J^1J^2B_{01}]\frac{1}{f_n}\frac{\partial f_n}{\partial\theta_2}\\ &+(J^1J^2B)^2\frac{1}{f_n}\frac{\partial^2 f_n}{\partial\theta_2^2}, \end{aligned} \tag{32}$$

where B, J^l, J^l_{ij}, $\partial f_n/\partial\theta_i$, and $\partial^2 f_n/\partial\theta_i\partial\theta_j$ are abbreviations for $B(\hat\theta_n,\theta)$, $J^l(\hat\theta_n,\theta)$, etc. So, (30) can be expressed as $E^t_\xi\{(Z_{t1}-a^{-1/2}\hat\kappa_1)^2\}=1+a^{-1}E^t_\xi\{M_a\}$, where

$$\begin{aligned} M_a &= \frac{a}{t}\Gamma^{\xi}_{2,11}(\hat\theta_t,\theta) - 2(\frac{a}{t})^{1/2}\hat\kappa_1\Gamma^{\xi}_{1,1}(\hat\theta_t,\theta) + \hat\kappa_1^2\\ &\to \rho^2(\theta)\Gamma^{\xi}_{2,11}(\theta,\theta) - 2\rho(\theta)\kappa_1(\theta)\Gamma^{\xi}_{1,1}(\theta,\theta) + \kappa_1^2(\theta) = M^{\xi}(\theta), \end{aligned} \tag{33}$$

as $a\to\infty$ in P_ξ-probability. Assuming only that ρ is almost differentiable with respect to θ_1 and θ_2, one can derive from (26), (27), (32), and integration by parts, that

$$E_\xi\{M^\xi(\theta)\} = \int_K m(\theta)\xi(\theta)d\theta, \tag{34}$$

where $m(\theta)$ has a rather complicated form (we omit the expression). Again, ignoring the interchangeability of the limit and integral, (28) suggests $E_\theta\{(Z_{t1} - a^{-1/2}\hat\kappa_1)^2\} \approx 1 + a^{-1}m(\theta)$ *v.w.*

Let $\hat m = m(\hat\theta_t)$ and consider the renormalized pivotal quantity $Z^*_{t1} = (Z_{t1} - \hat\mu_a)/\hat\sigma_a$, where

$$\hat\sigma_a^2 = \begin{cases} 1 + \hat m/a & \text{if } |\hat m| \le a^{1/2}/[\log(a)]^{-1},\\ 1 & \text{otherwise.}\end{cases} \tag{35}$$

Lemma 13. *Let h be a bounded symmetric function and let*

$$H_0(\sigma,\mu) = \int h(\frac{z-\mu}{\sigma})\phi(z)dz$$

and

$$H_1(\sigma,\mu)=\int zh(\frac{z-\mu}{\sigma})\phi(z)dz$$

for $\sigma>0$ and $-\infty<\mu<\infty$. Then H_0 and H_1 have continuous derivatives of all orders. Further at $\mu=0$ and $\sigma=1$ we have $H_0=\Phi h$, $\frac{\partial}{\partial\mu}H_0=0$, $\frac{\partial}{\partial\sigma}H_0=-2\Phi Vh$, $\frac{\partial^2}{\partial\mu^2}H_0=2\Phi Vh$, $H_1=0$, $\frac{\partial}{\partial\mu}H_1=-2\Phi Vh$, $\frac{\partial}{\partial\sigma}H_1=0$, and $\frac{\partial^2}{\partial\mu^2}H_1=0$.

Proof. The first assertion follows by the changes of variables:

$$\int h(\frac{z-\mu}{\sigma})\phi(z)dz=\int\sigma h(y)\phi(\sigma y+\mu)dy,$$

and

$$\int zh(\frac{z-\mu}{\sigma})\phi(z)dz=\int\sigma(\sigma y+\mu)h(y)\phi(\sigma y+\mu)dy,$$

by setting $\mu=0$ and $\sigma=1$. Then simple calculations along with (9) and (10) yield the remaining assertions.

Theorem 14. *Let h be a bounded symmetric function. Suppose that $\rho(\theta)$ is almost differentiable with respect to θ_1 and θ_2. If (21) holds for some $q>1$ and $\nabla^2\xi$ is continuous, then*

$$E_\xi\{h(Z_{t1}^*)\}=\Phi h+o(\frac{1}{a}).$$

Proof. Write $h(Z_{t1}^*)=h_a(Z_{t1})$. Then

$$\begin{aligned}E_\xi^t\{h(Z_{t1}^*)\}&=E_\xi^t\{h_a(Z_{t1})\}\\&=\Phi h_a+a^{-1/2}(\Phi Uh_a)E_\xi^t\{\rho(\theta)\Gamma_{1,1}^\xi(\theta,\theta)\}\\&\quad+\frac{1}{a}(\Phi Vh_a)E_\xi^t\{\rho^2(\theta)\Gamma_{2,11}^\xi(\theta,\theta)\}+\frac{1}{a}R_{2,a}(h_a),\end{aligned}$$

where $R_{2,a}(h_a)$ is as in Theorem 7 and, therefore, $E_\xi|R_{2,a}(h_a)|\to0$ as $a\to\infty$. By (29), (35), and Lemma 13, $\Phi h_a-\{\Phi h+a^{-1}(\Phi Vh)(\hat\kappa_1^2-\hat m)\}=o(a^{-1})$ and $\Phi Uh_a+2a^{-1/2}(\Phi Vh)\hat\kappa_1=o(a^{-1})$ uniformly w.r.t. $\hat\theta_t$. So,

$$\begin{aligned}E_\xi\{h(Z_{t1}^*)\}&=E_\xi\{\Phi h+\frac{1}{a}(\Phi Vh)(\hat\kappa_1^2-\hat m)-\frac{2}{a}(\Phi Vh)\hat\kappa_1E_\xi^t[\rho(\theta)\Gamma_{1,1}^\xi(\theta,\theta)]\\&\quad+\frac{1}{a}(\Phi Vh)E_\xi^t[\rho^2(\theta)\Gamma_{2,11}^\xi(\theta,\theta)]\}+o(\frac{1}{a})\\&=\Phi h+\frac{1}{a}(\Phi Vh)E_\xi[G(\theta)]+o(\frac{1}{a}),\end{aligned}$$

where $G(\theta)=\kappa_1^2(\theta)-m(\theta)-2\rho(\theta)\Gamma_{1,1}^\xi(\theta,\theta)\kappa_1(\theta)+\rho^2(\theta)\Gamma_{2,11}^\xi(\theta,\theta)$ and $E_\xi[G(\theta)]=E_\xi\{M^\xi(\theta)-m(\theta)\}=0$, by (33) and (34). Hence the proof.

Consequently we have the approximation to a higher order, $P_\theta(|Z^*_{t1}| \le z) = 2\Phi(z) - 1 + o(a^{-1})$ $v.w.$ This forms the basis for setting confidence intervals for θ_1.

7. An Example

This section presents applications of Theorem 14 to group sequential testing problems. Let $Y_{11}, Y_{12}, \dots$ be i.i.d. Poisson(λ_1) and $Y_{21}, Y_{22}, \dots$ be i.i.d. Poisson(λ_2), where $0 < \lambda_1, \lambda_2 < \infty$ are unknown. Suppose that interest lies in ratio of the rates, λ_1/λ_2, and experiments are run in a group sequential manner with group size m_a, possibly depending on $a > 0$, and stopping times $t = t_a = \inf\{n \ge 1 : m_a | n, |\sum_{i=1}^n (Y_{1i} - Y_{2i})| \ge a\}$. Here $m_a|n$ means that m_a divides n. This is a two-sample sequential testing problem where reparametrization to a two-parameter standard exponential family is possible. To see why, write down the joint density function of y_{1i} and y_{2i} and reparametrize by $\theta_1 = \log(\lambda_1/\lambda_2)$ and $\theta_2 = \log(\lambda_1\lambda_2)$. Then, with proper choice of the dominating measure we can derive that $p_\theta(x) \propto \exp\{\theta_1 x_1 + \theta_2 x_2 - \psi(\theta)\}$, where $x_1 = (y_1 - y_2)/2$, $x_2 = (y_1 + y_2)/2$, and $\psi(\theta) = e^{\frac{\theta_2}{2}}(e^{\frac{\theta_1}{2}} + e^{-\frac{\theta_1}{2}})$.

It is easily seen from the specified stopping rule that $a/t_a \le |\bar{Y}_{1t} - \bar{Y}_{2t}|$, and $a/(t_a - m_a) \ge |\bar{Y}_{1,t-m} - \bar{Y}_{2,t-m}|$.

So $a/t_a \to \rho^2(\theta) = e^{\frac{\theta_2}{2}}|e^{\frac{\theta_1}{2}} - e^{-\frac{\theta_1}{2}}|$, provided $m_a = o(a)$. Then (28) suggests the approximation

$$E_\theta(Z_{t1}) \approx a^{-1/2}\kappa_1(\theta) = \frac{|e^{\frac{\theta_1}{2}} - e^{-\frac{\theta_1}{2}}|^{1/2}}{a^{1/2}}\{\frac{(e^{\frac{\theta_1}{2}} - e^{-\frac{\theta_1}{2}})}{12(e^{\frac{\theta_1}{2}} + e^{-\frac{\theta_1}{2}})^{1/2}} - \frac{(e^{\frac{\theta_1}{2}} + e^{-\frac{\theta_1}{2}})^{3/2}}{4(e^{\frac{\theta_1}{2}} - e^{-\frac{\theta_1}{2}})}\}$$

and $E_\theta\{(Z_{t1} - a^{-1/2}\hat{\kappa}_1)^2\} \approx 1 + a^{-1}m(\theta)$.

Monte Carlo simulations are conducted for $a = 50$ and $(\lambda_1, \lambda_2) = (i, j)$, $i, j = 1, 2, 3, 4$, with $m_a = 1$ (fully sequential), 3 (group sequential with group size 3), and 5 (group sequential with group size 5). Table 1 gives the estimated probability for $P_\theta(|Z^*_{t1}| \le 1.96)$. Tables 2-4 show the Monte Carlo estimates of $E(Z_{t1})$, $E(Z^*_{t1})$, $E(Z^2_{t1})$, and $E((Z^*_{t1})^2)$ for fully sequential, group sequential with group size 3, and group sequential with group size 5, respectively. From the simulation, the magnitude of the mean is considerably reduced for renormalized pivotal quantity Z^*_{t1}.

Table 1. $P(|Z^*_{t1}| \le 1.96)$ (replicates=10,000 $a = 50$).

λ_1	1.0			2.0			3.0			4.0		
λ_2	fully	size3	size5	fully	size3	size5	fully	size3	size5	fully	size3	size5
1.0				0.949	0.951	0.949	0.948	0.949	0.950	0.951	0.950	0.949
2.0	0.951	0.952	0.951				0.951	0.950	0.952	0.949	0.946	0.946
3.0	0.952	0.951	0.950	0.954	0.955	0.953				0.949	0.950	0.952
4.0	0.953	0.951	0.952	0.952	0.953	0.950	0.951	0.952	0.952			

Table 2. Fully sequential (replicates=10,000 $a = 50$).

λ_1	1.0		2.0		3.0		4.0	
λ_2	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)
1.0			-0.125 (1.007)	-0.001 (0.997)	-0.101 (1.009)	0.003 (1.002)	-0.118 (1.011)	-0.018 (1.001)
2.0	0.114 (0.989)	-0.010 (0.974)			-0.141 (0.981)	0.012 (0.981)	-0.111 (0.979)	0.010 (0.980)
3.0	0.090 (1.007)	-0.017 (0.984)	0.150 (0.980)	-0.003 (0.974)			-0.197 (1.025)	-0.019 (1.014)
4.0	0.068 (1.013)	-0.036 (0.984)	0.119 (0.986)	-0.004 (0.979)	0.158 (1.002)	-0.020 (1.003)		

Table 3. Group sequential with size 3 (replicates=10,000 $a = 50$).

λ_1	1.0		2.0		3.0		4.0	
λ_2	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)
1.0			-0.123 (1.011)	0.000 (1.001)	-0.093 (1.007)	0.011 (1.001)	-0.108 (1.017)	-0.009 (1.010)
2.0	0.111 (1.003)	-0.013 (0.988)			-0.139 (0.983)	0.014 (0.982)	-0.113 (0.995)	0.008 (0.996)
3.0	0.091 (1.015)	-0.016 (0.992)	0.142 (0.962)	-0.011 (0.957)			-0.177 (1.006)	0.001 (1.001)
4.0	0.067 (1.019)	-0.037 (0.991)	0.114 (0.994)	-0.009 (0.988)	0.166 (0.994)	-0.012 (0.991)		

Table 4. Group sequential with size 5 (replicates=10,000 $a = 50$).

λ_1	1.0		2.0		3.0		4.0	
λ_2	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)	EZ_{t1} (EZ_{t1}^2)	EZ_{t1}^* ($E(Z_{t1}^*)^2$)
1.0			-0.121 (1.018)	0.002 (1.009)	-0.101 (1.023)	0.003 (1.016)	-0.088 (1.010)	0.010 (1.006)
2.0	0.109 (0.999)	-0.015 (0.984)			-0.136 (0.971)	0.017 (0.971)	-0.107 (1.019)	0.014 (1.021)
3.0	0.075 (0.991)	-0.032 (0.972)	0.138 (0.971)	-0.015 (0.967)			-0.175 (0.989)	0.003 (0.983)
4.0	0.068 (1.002)	-0.035 (0.976)	0.114 (1.003)	-0.009 (0.997)	0.155 (0.991)	-0.024 (0.992)		

Figures 1 and 2 show the cumulative distribution functions for $(\lambda_1, \lambda_2) = (2, 1)$ and $(\lambda_1, \lambda_2) = (1, 2)$. The cumulative distribution functions of Z_{t1} are over-estimated for the former and under-estimated for the latter, for both fully sequential and group sequential. The renormalized quantity performs much better.

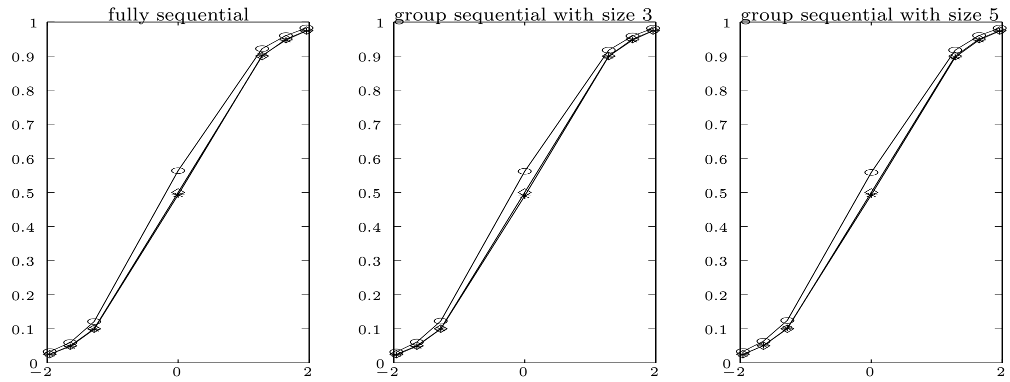

Figure 1. Cumulative distribution for $\lambda_1 = 2$, $\lambda_2 = 1$.
$\circ$ $- - -$ Z_{t1}, $*$ $- - -$ Z^*_{t1}, $\diamond$ $- - -$ standard normal

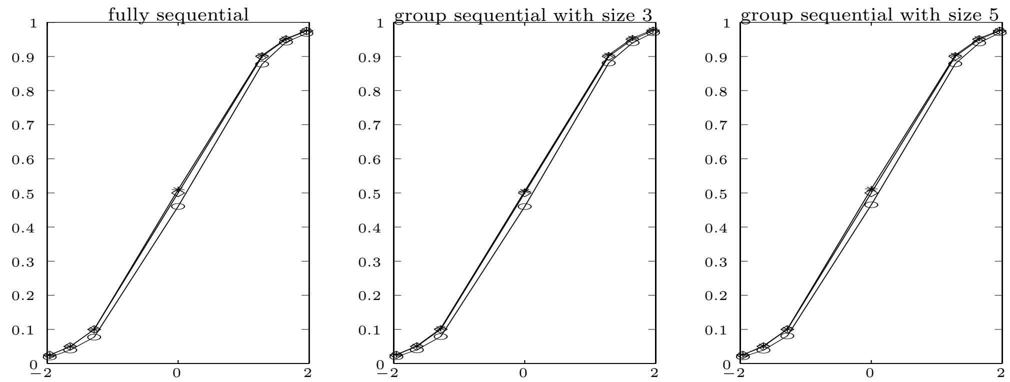

Figure 2. Cumulative distribution for $\lambda_1 = 1$, $\lambda_2 = 2$.
$\circ$ $- - -$ Z_{t1}, $*$ $- - -$ Z^*_{t1}, $\diamond$ $- - -$ standard normal

Acknowledgements

The authors would like to thank referees and an Associate Editor for their comments on the paper and for the Associate Editor's careful reading, which improved the quality of the paper. Both authors are supported by the National Science Foundation grant DMS-9626347, and the second by the U.S. Army Research Grant 37239-MA.

References

Barndorff-Nielsen, O. (1986). Inference for full or partial parameters based on the standardized signed log likelihood ratio. *Biometrika* **73**, 307-322.

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.

Brown, L. D. (1986). *Fundamentals of Statistical Exponential Families.* IMS, Hayward, CA.

Coad, D. S. and Woodroofe, M. (1996). Corrected confidence intervals after sequential testing with applications to survival analysis. *Biometrika* **83**, 763-777.

Ghosh, J. K., Sinha, B. and Joshi, S. (1982). Expansions for posterior probability and integrated Bayes risk. In *Statistical Decision Theory and Related Topics III*, Volume 1 (Edited by S. Gupta and J. Berger), 403-456. Acamedic Press, New York.

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.

Kass, R. E., Tierney, L. and Kadane, J. B. (1987). Asymptotics in Bayesian computation. In *Bayesian Statistics*, Volume 3, 261-278. Oxford University Press.

Lai, T. L. and Wang, J. Q. (1994). Asymptotic expansions for the distributions of stopped random walks and first passage times. *Ann. Probab.* **22**, 1957-1992.

Lehmann, E. L. (1983). *Theory of Point Estimation.* Wiley, New York.

Lehmann, E. L. (1986). *Testing Statistical Hypothese.* Wiley, New York.

Stein, C. (1987). *Approximate Computation of Expectations.* IMS, Hayward, CA.

Sun, D. (1994). Integrable expansions for posterior distributions for a two-parameter exponential family. *Ann. Statist.* **22**, 1808-1830.

Takahashi, H. (1987). Asymptotic expansions in Anscombe theorem for repeated significance tests and estimation after sequential testing. *Ann. Statist.* **15**, 278-295.

Woodroofe, M. (1986). Very weak expansions for sequential confidence levels. *Ann. Statist.* **14**, 1049-1067.

Woodroofe. M. (1988). Asymptotic expansions for the first passage times. *Stochastic Processes and Their Applications* **28**, 301-315.

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

Woodroofe, M. (1992). Integrable expansions for posterior distributions for one-parameter exponential families. *Statist. Sinica* **2**, 91-111.

Woodroofe, M. and Keener, R. (1987). Asymptotic expansions in boundary crossing probabilities. *Ann. Probab.* **15**, 102-114.

National Chengchi University, Taipei 11623, Taiwan.

E-mail: chweng@nccu.edu.tw

University of Michigan, Ann Arbor, MI 48109-1027, U.S.A.

E-mail:

(Received June 1998; accepted September 1999)