ISSN 1066-5307, Mathematical Methods of Statistics, 2019, Vol. 28, No. 1, pp. 66–73. © Allerton Press, Inc., 2019.

A Large Deviation Approximation for Multivariate Density Functions

C. Joutard[1*]

[1] Univ. Paul-Valéry Montpellier 3 and IMAG, Univ. Montpellier, CNRS, Montpellier, France
Received August 16, 2018; in final form, November 5, 2018

Abstract—We establish a large deviation approximation for the density of an arbitrary sequence of random vectors, by assuming several assumptions on the normalized cumulant generating function and its derivatives. We give two statistical applications to illustrate the result, the first one dealing with a vector of independent sample variances and the second one with a Gaussian multiple linear regression model. Numerical comparisons are eventually provided for these two examples.

Keywords: large deviations, density function, sample variance, Gaussian multiple linear regression model.

AMS 2010 Subject Classification: primary 60F10, secondary 62E20, 62H10.

DOI: 10.3103/S1066530719010058

1. INTRODUCTION

Let (Z_n) be a sequence of random vectors taking values in $\mathbb{R}^p$ and let b_n be a sequence of real positive numbers going to infinity as $n \to \infty$. In this paper, we obtain a large deviation approximation with speed b_n for the probability density function of Z_n. This generalizes our earlier result in [6] which deals with real random variables.

The large deviation approximation is obtained by imposing several assumptions on the normalized cumulant generating function (c.g.f.) of $b_n Z_n$, in particular, asymptotic expansions of the c.g.f and its derivatives. The proof follows the same lines as in [5], where large deviation theorems with speed n were derived for the probability density function of an arbitrary sequence of random vectors $n^{-1}T_n$.

Unlike those results, the asymptotic expansions of the c.g.f. and its derivatives allow us to obtain an explicit asymptotic expression for the density function of Z_n that depends on n only through the speed b_n. This can be useful, for example, when the c.g.f. of $b_n Z_n$ is intractable and we have to use its asymptotic function to calculate the rate function and make the exponential change of measure (see, e.g., [2] for an example in the one-dimensional case).

We illustrate our theorem with two statistical applications, namely, a vector of independent sample variances and the least squares estimator of a Gaussian multiple linear regression model. Some numerical results are also presented.

Several large deviation results dealing with the sum of random vectors can be found in the literature. In particular, regarding the large deviation approximations for densities, one can refer to [11], [3], [9], [12], [10], [8], [1], and [4].

The paper is organized as follows. In Section 2, we introduce the framework and assumptions, before giving the main result and its proof. The statistical applications, along with the numerical comparisons are then discussed in Section 3.

[*] E-mail: cyrille.joutard@univ-montp3.fr

2. MAIN RESULT

2.1. Notation and Assumptions

Let (Z_n) be a sequence of absolutely continuous random variables taking values in $\mathbb{R}^p$, and let (b_n) be a sequence of real positive numbers such that $\lim_{n\to\infty} b_n = \infty$. In what follows, we will use the following notation for the vector products, norms, etc. For $t, s \in \mathbb{R}^p$ and γ a p-dimensional vector with nonnegative integer components, we define

$$\langle t, s\rangle = t_1 s_1 + \ldots + t_p s_p, \quad \|t\| = \max_{1\le j\le p} |t_j|, \quad |t| = |t_1| + \ldots + |t_p|, \quad t^\gamma = t_1^{\gamma_1} t_2^{\gamma_2} \cdots t_p^{\gamma_p}.$$

Moreover, the determinant of a matrix A is denoted by $|A|$. We will also use the following big O notation for p-vectors and $p\times p$ matrices: $O_p(1)$ and $O_{p\times p}(1)$ as $n \to \infty$. Regarding the matrix norm, we could consider, for example, the following: for a $p \times p$ matrix A, $\|A\| = \max_{j=1,\ldots,p} \sum_{i=1}^{p} |A_{ij}|$.

Let ϕ_n be the moment generating function (m.g.f.) of $b_n Z_n$,

$$\phi_n(t) = \mathbb{E}\{\exp(\langle t, b_n Z_n\rangle)\}, \qquad t \in \mathbb{R}^p,$$

and let φ_n be the normalized c.g.f. of $b_n Z_n$,

$$\varphi_n(t) = b_n^{-1} \log \mathbb{E}\{\exp(\langle t, b_n Z_n\rangle)\}.$$

Assume that there exists a differentiable function φ defined on the open set $U_\alpha = \{t \in \mathbb{R}^p : \|t\| < \alpha\}$, $\alpha > 0$, such that $\lim_{n\to\infty} \varphi_n(t) = \varphi(t)$ for all $t \in U_\alpha$. Let $a \in \mathbb{R}^p$ be such that there exists $\tau \in U_\alpha$ satisfying $\nabla\varphi(\tau) = a$ ($\nabla\varphi$ is the vector of first order partial derivatives). The vector τ, known in the literature as a saddle point, is used to make an exponential change of measure, which plays a key role in large deviation problems.

In order to prove the asymptotic approximation, several assumptions, in particular, on the normalized c.g.f. φ_n and on the m.g.f. ϕ_n, are considered below.

$(A.1)$ φ_n is a holomorphic function in D_α^p, where $D_\alpha = \{z \in \mathbb{C} : |z| < \alpha\}$, and there exists $M > 0$ such that $|\varphi_n(z)| < M$ for all $z \in D_\alpha^p$ and n large enough.

$(A.2)$ There exists a function H continuous at τ such that for n large enough,

$$\varphi_n(\tau) = \varphi(\tau) + b_n^{-1} H(\tau) + O(b_n^{-2}), \tag{1}$$

$$\nabla\varphi_n(\tau) = \nabla\varphi(\tau) + O_p(b_n^{-1}), \tag{2}$$

$$\nabla^2\varphi_n(\tau) = \nabla^2\varphi(\tau) + O_{p\times p}(b_n^{-1}), \tag{3}$$

where the function φ is twice differentiable at τ. Furthermore, the eigenvalues of the Hessian matrix $\nabla^2\varphi(\tau)$ are positive.

$(A.3)$ Given $\delta > 0$, there exists $0 < \eta < 1$ such that

$$\limsup_{n\to\infty} \sup_{|t|\ge\delta} \left|\frac{\phi_n(\tau + it)}{\phi_n(\tau)}\right|^{1/b_n} \le \eta. \tag{4}$$

$(A.4)$ There exist $q \in (0, 1)$, $l > 0$ such that

$$\int_{\mathbb{R}^p} \left|\frac{\phi_n(\tau + it)}{\phi_n(\tau)}\right|^{l/b_n} dt = O(e^{b_n^q}).$$

Assumption $(A.1)$ is used to bound the remainder term of the expansion of G_n defined in (6).

Assumption $(A.2)$ plays an important role in the proof of the theorem, in particular, in dealing with the function I_n (which appears in the first asymptotic approximation (9) of the density function of Z_n).

Assumptions $(A.3)$–$(A.4)$ ensure that the term I_{n1} (coming from the decomposition of I_n) goes exponentially fast to zero.

Assumption $(A.4)$ is also used to apply the inversion formula.

2.2. Large Deviation Approximation

Below is the large deviation approximation for the density of Z_n.

Theorem 1. *Assume that (Z_n) is a sequence of absolutely continuous random variables taking values in $\mathbb{R}^p$. Let $a \in \mathbb{R}^p$ be such that there exists $\tau \in U_\alpha$ satisfying $\nabla\varphi(\tau) = a$. Let assumptions $(A.1)-(A.4)$ hold. Then, for n large enough, we have the following asymptotic approximation for the probability density function k_n of Z_n,*

$$k_n(a) = \frac{b_n^{p/2}}{(2\pi)^{p/2}|\nabla^2\varphi(\tau)|^{1/2}} \exp(-b_n I(a) + H(\tau))\big[1 + O(b_n^{-1})\big], \tag{5}$$

where τ is such that $\nabla\varphi(\tau) = a$ and $I(a) = \langle\tau, a\rangle - \varphi(\tau)$.

Proof. As mentioned in the Introduction, the proof follows the same lines as in [5]. First, let us set

$$G_n(t) = -\varphi_n(\tau + it) + \varphi_n(\tau) + i\langle t, a\rangle. \tag{6}$$

Then, since φ_n is a holomorphic function in D_α^p (Assumption $(A.1)$), using (2) and recalling that $\nabla\varphi(\tau) = a$ one can find a small enough $\eta_0 > 0$ such that for $|t| < \eta_0$,

$$G_n(t) = \frac{1}{2}t^T\nabla^2\varphi_n(\tau)t + i\sum_{|\gamma|=3} a_\gamma^{(n)}t^\gamma - i\langle t, O_p(1)\rangle b_n^{-1} - R_n(\tau + it), \tag{7}$$

where $a_\gamma^{(n)}$ are the coefficients of the expansion of $\varphi_n(\tau + it)$.

Since $|\varphi_n(z)| < M$ for all $z \in D_\alpha^p$ (Assumption $(A.1)$), there exists a constant $M_0 > 0$ such that the remainder term R_n satisfies

$$|R_n(\tau + it)| \le M_0|t|^4 \tag{8}$$

for $|t| < \eta_0$. Using an exponential change of measure, let us define

$$dH_n(y) = \exp(\langle\tau, y\rangle - b_n\varphi_n(\tau))\, dK_n(y),$$

where K_n is the distribution function of b_nZ_n. By $(A.4)$, using the multivariate inversion formula, the density function of H_n is then given by

$$\frac{1}{(2\pi)^p}\int_{\mathbb{R}^p} \phi_n(\tau + it)e^{-\langle\tau+it,y\rangle}\frac{dH_n(y)}{dK_n(y)}\, dt.$$

Hence the density function k_n of Z_n is

$$k_n(y) = \left(\frac{b_n}{2\pi}\right)^p \int_{\mathbb{R}^p} \phi_n(\tau + it)e^{-b_n\langle\tau+it,y\rangle}\, dt.$$

Next, noting that (3) (where the matrix $\nabla^2\varphi(\tau)$ is invertible) implies that $|\nabla^2\varphi_n(\tau)| = |\nabla^2\varphi(\tau)|(1 + O(b_n^{-1}))$ (because $|Id + \epsilon A| = 1 + \mathrm{tr}(A)\epsilon + O(\epsilon^2)$, as $\epsilon \to 0$, for any bounded square matrix A, and Id is the identity matrix), we have

$$k_n(a) = \left[\frac{b_n^p}{(2\pi)^p|\nabla^2\varphi(\tau)|}\right]^{1/2} \exp(-b_n\varphi^*(a) + H(\tau))I_n[1 + O(b_n^{-1})], \tag{9}$$

where $\varphi^*(a) = \langle\tau, a\rangle - \varphi(\tau)$. Besides, by (1),

$$\begin{aligned} I_n &= \left(\frac{b_n}{2\pi}\right)^{p/2} |\nabla^2\varphi_n(\tau)|^{1/2} \exp(-H(\tau)) \\ &\quad \times \int_{\mathbb{R}^p} \exp(b_n[\varphi^*(a) - \langle\tau + it, a\rangle])\phi_n(\tau + it)\, dt \\ &= \left(\frac{b_n}{2\pi}\right)^{p/2} |\nabla^2\varphi_n(\tau)|^{1/2} \exp(-H(\tau)) \end{aligned}$$

$$\times \int_{\mathbb{R}^p} \exp(b_n[\varphi_n(\tau+it)-\varphi_n(\tau)-i\langle t,a\rangle+b_n^{-1}H(\tau)+O(b_n^{-2})])\,dt$$
$$=(I_{n1}+I_{n2})[1+O(b_n^{-1})]$$

with

$$I_{n1}=\left(\frac{b_n}{2\pi}\right)^{p/2}|\nabla^2\varphi_n(\tau)|^{1/2}\int_{|t|\geq\delta}\exp(-b_nG_n(t))\,dt$$

and

$$I_{n2}=\left(\frac{b_n}{2\pi}\right)^{p/2}|\nabla^2\varphi_n(\tau)|^{1/2}\int_{|t|<\delta}\exp(-b_nG_n(t))\,dt,$$

where $\delta\in(0,\eta_0)$ has to be chosen small enough. Recall that the expression of G_n is given in (6). As in [5], by using assumptions $(A.3)-(A.4)$ we can show that I_{n1} goes exponentially fast to zero.

Now, we make the change of variable $s=\sqrt{b_n}t$ to get

$$I_{n2}=\sqrt{\frac{|\nabla^2\varphi_n(\tau)|}{(2\pi)^p}}\int_{|s|<\delta\sqrt{b_n}}\exp\left(-\frac{s^T\nabla^2\varphi_n(\tau)s}{2}\right)\big[1+Z_n(s)+L_n(s)\big]\,ds,$$

where $L_n(s)=e^{Z_n(s)}-1-Z_n(s)$ and $Z_n(s)=-b_nG_n(b_n^{-1/2}s)+2^{-1}s^T\nabla^2\varphi_n(\tau)s$. Set

$$I_{n3}=\sqrt{\frac{|\nabla^2\varphi_n(\tau)|}{(2\pi)^p}}\int_{|s|<\delta\sqrt{b_n}}\exp\left(-\frac{s^T\nabla^2\varphi_n(\tau)s}{2}\right)\big[1+Z_n(s)\big]\,ds.$$

Then, by (7),

$$\begin{aligned}I_{n3}&=\sqrt{\frac{|\nabla^2\varphi_n(\tau)|}{(2\pi)^p}}\int_{|s|<\delta\sqrt{b_n}}\exp\left(-\frac{s^T\nabla^2\varphi_n(\tau)s}{2}\right)ds\\&\quad-\frac{i}{\sqrt{b_n}}\sqrt{\frac{|\nabla^2\varphi_n(\tau)|}{(2\pi)^p}}\sum_{|\gamma|=3}a_\gamma^{(n)}\int_{|s|<\delta\sqrt{b_n}}\exp\left(-\frac{s^T\nabla^2\varphi_n(\tau)s}{2}\right)s^\gamma\,ds\\&\quad+\frac{i}{\sqrt{b_n}}\sqrt{\frac{|\nabla^2\varphi_n(\tau)|}{(2\pi)^p}}\int_{|s|<\delta\sqrt{b_n}}\exp\left(-\frac{s^T\nabla^2\varphi_n(\tau)s}{2}\right)\langle s,O_p(1)\rangle\,ds\\&\quad+b_n\sqrt{\frac{|\nabla^2\varphi_n(\tau)|}{(2\pi)^p}}\int_{|s|<\delta\sqrt{b_n}}\exp\left(-\frac{s^T\nabla^2\varphi_n(\tau)s}{2}\right)R_n\Big(\tau+i\frac{s}{\sqrt{b_n}}\Big)\,ds.\end{aligned}$$

Since $\exp\big(-\frac{s^T\nabla^2\varphi_n(\tau)s}{2}\big)$ goes exponentially fast to zero as $|s|\to\infty$, the first term of the right-hand side is equal to $1+O(b_n^{-2})$. The second and third terms are equal to zero since the integrands are odd functions. Using (8), one can easily see that the fourth term is $O(b_n^{-1})$.

Finally, by Assumption $(A.2)$ there exists $\beta_2>0$ such that $s^T\nabla^2\varphi_n(\tau)s\geq\beta_2s^Ts$ for n large enough. Hence, by choosing $\delta>0$ small enough, we can show as in [5] that

$$\sqrt{\frac{|\nabla^2\varphi_n(\tau)|}{(2\pi)^p}}\int_{|s|<\delta\sqrt{b_n}}\exp\left(-\frac{s^T\nabla^2\varphi_n(\tau)s}{2}\right)L_n(s)\,ds=O(b_n^{-1}).$$

It therefore follows that $I_{n2}=1+O(b_n^{-1})$. This completes the proof of Theorem 1. □

Remark. Note that we can also obtain a similar result to Theorem 1 for lattice-valued random vectors. In this case we have to slightly change Assumption $(A.3)$ to consider the lattice case (see [5] for further details).

3. APPLICATION EXAMPLES

3.1. Example 1: Vector of Independent Sample Variances

Assume that the random vector $X = (X_1, \ldots, X_p)$ follows a multivariate normal distribution with mean vector $\mu = (\mu_1, \ldots, \mu_p)$ and covariance matrix $\Sigma = \operatorname{diag}(\sigma_1^2, \ldots, \sigma_p^2)$, where $\sigma_k^2 > 0$ for each $k = 1, \ldots, p$. We consider the vector of independent sample variances

$$Z_{n,p} = \Big((n-1)^{-1}\sum_{i=1}^{n}(X_{i1} - \overline{X}_1)^2, \ldots, (n-1)^{-1}\sum_{i=1}^{n}(X_{ip} - \overline{X}_p)^2\Big).$$

This is an unbiased estimator of the vector of variances $(\sigma_1^2, \ldots, \sigma_p^2)$.

For each $k = 1, \ldots, p$, the statistic $\sigma_k^{-2}\sum_{i=1}^{n}(X_{ik} - \overline{X}_k)^2$ follows a chi-square distribution with $n-1$ degrees of freedom. This will allow us to get an explicit expression for the moment generating function of nZ_n. The following result follows then from the application of Theorem 1 with $b_n = n$.

Corollary 1. *Let $Z_{n,p}$ be defined as above. Then for $a \in \mathbb{R}^p$ such that $a_k > 0$, for each $k = 1, \ldots, p$ and n large enough, we have the following approximation for the density function of $Z_{n,p}$,*

$$f_{Z_{n,p}}(a) = \Big(\frac{n}{4\pi}\Big)^{p/2}\frac{1}{\prod_{k=1}^{p} a_k}\exp\big(-(n-1)I(a)\big)[1 + O(n^{-1})], \tag{10}$$

where

$$I(a) = \frac{1}{2}\sum_{k=1}^{p}\left[\frac{a_k}{\sigma_k^2} - \log\Big(\frac{a_k}{\sigma_k^2}\Big) - 1\right].$$

Proof. Assumptions $(A.1)$–$(A.2)$ are easily verified (see [7]). For Assumption $(A.3)$, we have, for n large enough,

$$\begin{aligned}\left|\frac{\phi_n(\tau + it)}{\phi_n(\tau)}\right|^{1/n} &= \exp\left(-\frac{n-1}{4n}\sum_{k=1}^{p}\log\left(1 + \frac{4\sigma_k^4 t_k^2(1 + \frac{1}{n-1})^2}{(a_k^{-1}\sigma_k^2 - \frac{1}{n-1}(1 - a_k^{-1}\sigma_k^2))^2}\right)\right)\\ &\le \exp\left(-\frac{1}{5}\sum_{k=1}^{p}\log(1 + 4a_k^{-2}t_k^2)\right).\end{aligned}$$

Then, for a given $\delta > 0$, there exists $0 < \eta < 1$ such that

$$\sup_{|t|\ge\delta}\exp\left(-\frac{1}{5}\sum_{k=1}^{p}\log(1 + 4a_k^2 t_k^2)\right) = \exp\left(-\frac{1}{5}\sum_{k=1}^{p}\log(1 + 4a_k^2 t_{0,k}^2)\right) \le \eta,$$

where t_0 is such that $|t_0| = \delta$ and $(A.3)$ is verified.

Now, for $l = 4$, we have

$$\begin{aligned}\int_{\mathbb{R}^p}\left|\frac{\phi_n(\tau + it)}{\phi_n(\tau)}\right|^{l/n} dt &= \int_{\mathbb{R}^p}\prod_{k=1}^{p}\left(1 + \frac{4\sigma_k^4 t_k^2(\frac{n}{n-1})^2}{(1 - 2\sigma_k^2\tau_k\frac{n}{n-1})^2}\right)^{-\frac{(n-1)}{n}} dt\\ &= \prod_{k=1}^{p}\int_{\mathbb{R}}\left(1 + \frac{4\sigma_k^4 t_k^2(\frac{n}{n-1})^2}{(1 - 2\sigma_k^2\tau_k\frac{n}{n-1})^2}\right)^{-\frac{(n-1)}{n}} dt_k = O(1),\end{aligned}$$

which implies $(A.4)$. □

For the numerical comparisons, we consider an example with $p = 3$, $n = 100$, $\sigma_i^2 = 1$, for all $i = 1, 2, 3$ and we compare, for different values of $a \in \mathbb{R}^3$, the true densities $f_{Z_{n,3}}(a)$ and the approximations given by (5). The results are displayed in the table below.

As one can see in Table 1, the large deviation approximation gives good results everywhere.

Table 1. Approximations of $f_{Z_{n,3}}(a_1, a_2, a_3)$ for $p = 3$.

a	$(0.5, 0.6, 0.7)$	$(0.6, 0.7, 0.8)$	$(0.7, 0.8, 0.9)$	$(0.8, 0.9, 1.0)$	$(1.0, 1.0, 1.0)$
Exact	**1.85e-06**	**0.00522**	**0.6440**	**7.453**	**22.00**
Approx.	1.89e-06	0.00533	0.6571	7.605	22.45
a	$(1.0, 1.1, 1.2)$	$(1.1, 1.2, 1.3)$	$(1.2, 1.3, 1.4)$	$(1.3, 1.4, 1.5)$	$(1.4, 1.5, 1.6)$
Exact	**5.51**	**0.658**	**0.0281**	**0.00050**	**4.20e-06**
Approx.	5.62	0.671	0.0286	0.00051	4.29e-06

3.2. Example 2: Gaussian Multiple Linear Regression Model with Deterministic Regressors

Let $Y_1, Y_2, \ldots$ be a sequence of random variables and let us consider the linear regression model

$$Y_i = x_i^T \beta + \xi_i, \quad i = 1, .., n, \tag{11}$$

where ξ_i are i.i.d. random variables following an $\mathcal{N}(0, \sigma^2)$ distribution with variance $\sigma^2 > 0$, $x_i^T = (x_{i1}, \ldots, x_{ip})$ is a p-dimensional vector of deterministic regressors that depend on n and $\beta^T = (\beta_1, \ldots, \beta_p)$ is the unknown parameter vector (T stands for transpose operation).

Let $\mathbf{Y}_n^T = (Y_1, \ldots, Y_n)$ and $X_n = (x_{ij})_{1 \le i \le n, 1 \le j \le p}$, $n \ge 1$. Assuming that $\mathrm{Rank}(X_n) = p$ for some $n \ge 1$, the least squares estimator (l.s.e.) is

$$\hat{\beta}_n = (X_n^T X_n)^{-1} X_n^T \mathbf{Y}_n. \tag{12}$$

We also assume that there exist three $p \times p$ matrices A, B and C_n such that for n large enough,

$$n^{-1} X_n^T X_n = A + \frac{B}{n} + \frac{C_n}{n^2}, \tag{13}$$

where A is symmetric positive definite and C_n is a bounded matrix, that is $C_n = O_{p\times p}(1)$ (see Section 2.1 for an example of matrix norm which could be used here). Assume that the Y_i's follow the linear model (11) with the true values of the parameter vector $\beta = \beta_0$ and of the variance $\sigma^2 = \sigma_0^2$. Then, applying Theorem 1 with $b_n = n$, we have the following asymptotic approximation for the density of $\hat{\beta}_n$.

Corollary 2. *Let $\hat{\beta}_n$ be the l.s.e. defined in* (12). *Under the above assumptions, for $a \in \mathbb{R}^p$ and n large enough, we have the following approximation for the density function of $\hat{\beta}_n$,*

$$f_{\hat{\beta}_n}(a) = \frac{n^{p/2} \exp\left(-\frac{n}{2\sigma_0^2}(a-\beta_0)^T A (a-\beta_0) - \frac{1}{2\sigma_0^2}(a-\beta_0)^T B (a-\beta_0)\right)}{(2\pi)^{p/2} \sigma_0^p |A|^{-1/2}} [1 + O(n^{-1})]. \tag{14}$$

Proof. Here we will use several arguments of the proof of Corollary 2 in [7]. Assumption $(A.1)$ is easily verified, so we turn to $(A.2)$. First, let us set $\varphi(t) = \langle t, \beta_0 \rangle + \frac{\sigma_0^2}{2} t^T A^{-1} t$ and $H(t) = -\frac{\sigma_0^2}{2} t^T A^{-1} B A^{-1} t$. The equation $\nabla\varphi(t) = a$ has a closed-form solution, that is,

$$\tau = \sigma_0^{-2} A (a - \beta_0).$$

Next, the expansions (1), (2) and (3) follow from

$$\varphi_n(\tau) = \langle \tau, \beta_0 \rangle + \frac{\sigma_0^2}{2} \tau^T A^{-1} \tau - \tau^T \frac{\sigma_0^2 A^{-1} B A^{-1}}{2n} \tau + O\Big(\frac{1}{n^2}\Big) = \varphi(\tau) + \frac{H(\tau)}{n} + O\Big(\frac{1}{n^2}\Big),$$

$$\nabla\varphi_n(\tau) = \beta_0 + \sigma_0^2 A^{-1} \tau - \frac{\sigma_0^2 A^{-1} B A^{-1}}{n} \tau + O_p\Big(\frac{1}{n^2}\Big) = \nabla\varphi(\tau) + O_p\Big(\frac{1}{n}\Big)$$

and

$$\nabla^2\varphi_n(\tau) = \sigma_0^2 A^{-1} - \frac{\sigma_0^2 A^{-1} B A^{-1}}{n} + O_{p\times p}\Big(\frac{1}{n^2}\Big) = \nabla^2\varphi(\tau) + O_{p\times p}\Big(\frac{1}{n}\Big).$$

The eigenvalues of the positive definite matrix A^{-1} being positive, $(A.2)$ is verified.

For Assumption $(A.3)$, denoting $\Gamma_n = n^{-1}X_n^T X_n$, we have

$$\sup_{|t|\geq\delta}\left|\frac{\phi_n(\tau+it)}{\phi_n(\tau)}\right|^{1/n} = \sup_{|t|\geq\delta} e^{-\frac{\sigma_0^2}{2}t^T\Gamma_n^{-1}t} = e^{-\frac{\sigma_0^2}{2}t_0^T\Gamma_n^{-1}t_0},$$

where t_0 is such that $|t_0| = \delta$, $\Gamma_n^{-1} = A^{-1} - \frac{A^{-1}BA^{-1}}{n} + \frac{R_nA^{-1}}{n^2}$, and R_n is a bounded matrix (see [7]). Since A^{-1} is positive definite and R_n is bounded, one can find $0 < \eta < 1$ such that

$$\limsup_{n\to\infty} e^{-\frac{\sigma_0^2}{2}t_0^T\Gamma_n^{-1}t_0} \leq \eta$$

and $(A.3)$ is then satisfied.

Now, regarding Assumption $(A.4)$, we have for $l > 0$

$$\int_{\mathbb{R}^p}\left|\frac{\phi_n(\tau+it)}{\phi_n(\tau)}\right|^{l/n} dt = \int_{\mathbb{R}^p} e^{-l\frac{\sigma_0^2}{2}t^T\Gamma_n^{-1}t}\, dt.$$

The assumption is easily verified by noting that $\Gamma_n^{-1} \to A^{-1}$ as $n \to \infty$ and $\int_{\mathbb{R}^p} e^{-l\frac{\sigma_0^2}{2}t^TA^{-1}t}\, dt < \infty$. □

For the numerical comparisons, we consider two examples, the first one with $p = 2$ and the second one with $p = 3$. In both examples we compare the true densities $f_{\hat{\beta}_n}(a)$ and the approximations given by (14) for different values of $a \in \mathbb{R}^2$ (or $a \in \mathbb{R}^3$).

(1) For this first example, we take $p = 2$, $n = 50$, $\beta_0 = (1,1)^T$ and $\sigma_0^2 = 3$. For the regressors, we choose $x_{i1} = 1$, $x_{i2} = \frac{i}{n}$, for all $i = 1,\ldots,n$. Therefore

$$X_n^TX_n = \begin{pmatrix} n & \frac{n}{2}+\frac{1}{2} \\ \frac{n}{2}+\frac{1}{2} & \frac{n}{3}+\frac{1}{2}+\frac{1}{6n} \end{pmatrix}.$$

Then (13) holds with $A = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix}$, $B = \begin{pmatrix} 0 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}$, $C_n = \begin{pmatrix} 0 & 0 \\ 0 & \frac{1}{6} \end{pmatrix}$ and A is positive definite. The numerical results are given in Table 2.

Table 2. Approximations of $f_{\hat{\beta}_n}(a_1,a_2)$ for $p = 2$.

a	$(0,0.2)$	$(0.2,0.4)$	$(0.4,0.6)$	$(0.6,0.8)$	$(0.8,1.0)$	$(1.0,1.0)$
Exact	**3.28e-08**	**2.23e-05**	**0.00314**	**0.09118**	**0.5486**	**0.7656**
Approx.	3.28e-08	2.23e-05	0.00314	0.09120	0.5487	0.7657
a	$(1.0,1.2)$	$(1.2,1.4)$	$(1.4,1.6)$	$(1.6,1.8)$	$(1.8,2.0)$	$(2.0,2.2)$
Exact	**0.6828**	**0.1758**	**0.00937**	**0.000103**	**2.35e-07**	**1.11e-10**
Approx.	0.6829	0.1759	0.00937	0.000103	2.36e-07	1.11e-10

(2) For the second example, we take $p = 3$, $n = 60$, $\beta_0 = (1,1,1)^T$ and $\sigma_0^2 = 5$. For the regressors, we choose $x_{i1} = 1$, for all $i = 1,\ldots,n$, $x_{i2} = (-1)^i$, if $i \leq n-2 = 58$ and $x_{i2} = 0$ otherwise, $x_{i3} = 0$, if $i \leq n/2$ and $x_{i3} = 1$ otherwise.

Therefore

$$X_n^T X_n = \begin{pmatrix} n & 0 & \frac{n}{2} \\ 0 & n-2 & 0 \\ \frac{n}{2} & 0 & \frac{n}{2} \end{pmatrix}.$$

Then (13) holds with $A = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$, $B = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $C_n = 0_{3\times 3}$ and A is positive definite. The numerical results are shown in Table 3.

Table 3. Approximations of $f_{\hat{\beta}_n}(a_1, a_2, a_3)$ for $p = 3$.

a	$(0, 0.2, 0.4)$	$(0.2, 0.4, 0.6)$	$(0.4, 0.6, 0.8)$	$(0.8, 1.0, 1.2)$	$(1.0, 1.0, 1.0)$
Exact	**7.29e-07**	**0.000314**	**0.02554**	**1.1508**	**1.2975**
Approx.	7.41e-07	0.000319	0.02598	1.1705	1.3197
a	$(1.0, 1.2, 1.4)$	$(1.2, 1.4, 1.6)$	$(1.4, 1.6, 1.8)$	$(1.6, 1.8, 2.0)$	$(1.8, 2.0, 2.2)$
Exact	**0.6366**	**0.06670**	**0.00132**	**4.97e-06**	**3.54e-09**
Approx.	0.6475	0.06784	0.00135	5.06e-06	3.60e-09

As in the case of the sample variances, for both examples, the large deviation approximation turns out to be very accurate for all values, in particular, in Table 2.

REFERENCES

1. C. Andriani and P. Baldi, “Sharp Estimates of Deviations of the Sample Mean in Many Dimensions”, Ann. Inst. Henri Poincaré **33** (3), 371–385 (1997).
2. B. Bercu, F. Gamboa, and M. Lavielle, “Sharp Large Deviations for Gaussian Quadratic Forms with Applications”, ESAIM:P&S **3**, 1–29 (2000).
3. A. A. Borovkov and B. A. Rogozin, “On the Multidimensional Central Limit Theorem”, Theory Probab. Appl. **10**, 55–62 (1965).
4. A. A. Borovkov and A. A. Mogulskii, “Integro-Local Limit Theorems Including Large Deviations for Sums of Random Vectors”, I, Theory Probab. Appl. **43**, 1–12 (1998).
5. N. R. Chaganty and J. Sethuraman, “Multidimensional Large Deviation Local Limit Theorems”, J. Multivar. Anal. **20**, 190–204 (1986).
6. C. Joutard, “Asymptotic Approximation for the Probability Density Function of an Arbitrary Sequence of Random Variables”, Statist. Probab. Lett. **90**, 100–107 (2013).
7. C. Joutard, “Multidimensional Strong Large Deviation Results”, Metrika **80**, 663–683 (2017).
8. D. McDonald, “A Local Limit Theorem for Large Deviation in Sums of Independent, Nonidentically Distributed Random Variables”, Ann. Probab. **3**, 526–531 (1979).
9. D. A. Moskvin, “A Local Limit Theorem for Large Deviations in the Case of Differently Distributed Lattice Summands”, Theory Probab. Appl. **4**, 678–684 (1972).
10. P. C. B. Phillips, “A Large Deviation Limit Theorem for Multivariate Distributions”, J. Multivar. Anal. **7**, 50–62 (1977).
11. W. Richter, “Multi-Dimensional Local Limit Theorems for Large Deviations”, Theory Probab. Appl. **3**, 100–106 (1958).
12. L. Saulis and V. Statulevichus, “On Large Deviations for a Distribution Density Function”, Math. Nachrichten **70**, 111–132 (1976).