RARE-EVENT ANALYSIS FOR EXTREMAL EIGENVALUES OF WHITE WISHART MATRICES

BY TIEFENG JIANG[*,1], KEVIN LEDER[*,2] AND GONGJUN XU[†,3]

University of Minnesota * *and University of Michigan* †

In this paper, we consider the extreme behavior of the extremal eigenvalues of white Wishart matrices, which plays an important role in multivariate analysis. In particular, we focus on the case when the dimension of the feature p is much larger than or comparable to the number of observations n, a common situation in modern data analysis. We provide asymptotic approximations and bounds for the tail probabilities of the extremal eigenvalues. Moreover, we construct efficient Monte Carlo simulation algorithms to compute the tail probabilities. Simulation results show that our method has the best performance among known approximation approaches, and furthermore provides an efficient and accurate way for evaluating the tail probabilities in practice.

1. Introduction. In many modern scientific settings, data sets are generated where the dimension of the samples is comparable or even larger than the sample size. Analysis on such multidimensional data frequently involves estimating rare-event probabilities, such as small tail probabilities of test statistics. For instance, in statistical hypothesis testing, consider multiple comparisons with relatively few signals of interest among a large number of null statistics. In order to control the overall false-positive error rate at a certain level, we may need to evaluate a very small marginal p-value for each individual test statistic.

This paper focus on the tail probabilities of extremal eigenvalues of white Wishart matrices, which play an important role in multivariate statistical analysis and have wide applications in many fields, such as image analysis, signal processing, and functional data analysis. A white Wishart matrix with parameters $\mathbf{\Sigma} = \mathbf{I}_p$ (the $p \times p$ identity matrix), n and $\beta = 1$ is the sample covariance matrix $\mathbf{X}^*\mathbf{X}$ where $\mathbf{X} = (x_{ij})_{n \times p}$ and x_{ij} are i.i.d. $N(0, 1)$ random variables. The most natural alternative values of β are $\beta = 2$ for x_{ij} complex valued and $\beta = 4$ for x_{ij} quaternion valued. Most data analysis in statistics focuses on the case when $\beta = 1$. In engineering and applied science applications, such as signal processing, oceanography and atmospheric sciences, it is common to use complex valued variables

to study two-dimensional signals. In these settings, the use of $\beta = 2$ is useful for the understanding of statistical properties of the data set. In physics, for quantum systems with a time reversal symmetry T, where either $T^2 = 1$ or $T^2 = -1$, the former leads to symmetric matrices ($\beta = 1, 2$) and the latter leads to symplectic matrices ($\beta = 4$).

The largest eigenvalue of a sample covariance matrix gives useful information for distinguishing a "signal subspace" of higher variance from the background noise variables [Johnstone (2001)]. In particular, for n i.i.d. p dimensional Gaussian observations following $N(\mathbf{0}, \boldsymbol{\Sigma})$, consider testing the null hypothesis that $\boldsymbol{\Sigma} = \mathbf{I}_p$, where $\mathbf{I}_p$ is the identity matrix. Following Roy's union intersection principle [Roy (1953)], one can take the largest eigenvalue of the sample covariance matrix as the test statistics and reject the null hypothesis for large values. Then the corresponding p-value is the tail probability of the largest eigenvalue under $\boldsymbol{\Sigma} = \mathbf{I}_p$. For example, see Patterson, Price and Reich (2006) for applications in SNP (single nucleotide polymorphism) data, Bianchi et al. (2011) for applications in detecting single-source with a sensor array, and Kwapień, Drożdż and Speth (2003) for applications in financial market analysis. Accurate evaluations of such tail probabilities are needed in performing the corresponding statistical analysis and this motivates our study.

1.1. *Problem setting and related studies.* For a white Wishart matrix, it is in fact possible to consider arbitrary values of $\beta > 0$. This more general class of matrices is referred to in the literature as the β-Laguerre ensemble. In this work, we primarily focus on the largest eigenvalues of the β-Laguerre ensemble in the setting of $p \geq n$, $\boldsymbol{\Sigma} = \mathbf{I}_p$ and arbitrary $\beta > 0$. For this setting, the n positive eigenvalues of the β-Laguerre ensemble $\lambda_1, \ldots, \lambda_n$ are distributed with probability density function

(1)
$$
\begin{aligned}
&f_{n,p,\beta}(\lambda_1, \ldots, \lambda_n) \\
&= c_{n,p,\beta} \prod_{1 \leq i < j \leq n} |\lambda_i - \lambda_j|^\beta \cdot \prod_{i=1}^{n} \lambda_i^{\frac{\beta(p-n+1)}{2}-1} \cdot e^{-\frac{1}{2}\sum_{i=1}^{n} \lambda_i},
\end{aligned}
$$

where $c_{n,p,\beta}$ is a normalizing constant taking the form of

(2)
$$
c_{n,p,\beta} = 2^{-\frac{\beta n p}{2}} \prod_{j=1}^{n} \frac{\Gamma(1 + \frac{\beta}{2})}{\Gamma(1 + \frac{\beta}{2}j)\Gamma(\frac{\beta}{2}(p-n+j))}.
$$

In particular, when $\beta = 1, 2$ and 4, the function $f_{n,p,\beta}(\lambda_1, \ldots, \lambda_n)$ in (1) is the density function of the n positive eigenvalues of Wishart matrix $\mathbf{X}^*\mathbf{X}$, where $\mathbf{X} = (x_{ij})_{n \times p}$ and x_{ij}'s are i.i.d. standard ($\beta = 1$), complex ($\beta = 2$) or quaternion ($\beta = 4$) Gaussian random variables (r.v.'s). See, for example, James (1964) and Muirhead (1982) for the cases of $\beta = 1$ and 2, and Macdonald (1995) and

Edelman and Rao (2005) for $\beta = 4$. See also Anderson, Guionnet and Zeitouni (2010) for further discussion and applications.

Let $\lambda_{(1)} > \cdots > \lambda_{(n)}$ be the order statistics of $\lambda_1, \ldots, \lambda_n$. The joint density function of the order statistics is

$$(3) \qquad g_{n,p,\beta}(\lambda_1, \ldots, \lambda_n) = n! f_{n,p,\beta}(\lambda_1, \ldots, \lambda_n) \times I_{(\lambda_1 > \cdots > \lambda_n)},$$

where $I_{(\cdot)}$ is the indicator function. In this paper, we focus on the asymptotic approximation and efficient simulation of tail probabilities

$$P(\lambda_{(1)} > px) \qquad \text{as } p \to \infty$$

for any $\beta > 0$ and $x > \beta$. In particular, we consider the high-dimensional settings where $p/n \to \gamma \in [1, \infty)$ or $p/n \to \infty$.

Large sample properties of the largest eigenvalue have been extensively studied in the literature, most of which focus on the asymptotic distribution of $\lambda_{(1)}$ and its large deviation principle. For the asymptotic distribution of $\lambda_{(1)}$, Johansson (2000) and Johnstone (2001) studied the cases when $p/n \to \gamma \in (0, \infty)$ and $\beta = 2$ and 1, and showed that the largest eigenvalue (with proper recentering and rescaling) follows the Tracy–Widom distribution as appeared in the study of the Gaussian unitary ensemble. El Karoui (2003) extended the asymptotic regime to the case when $p/n \to \infty$. For general $\beta > 0$, the limiting distribution of $\lambda_{(1)}$ is obtained by Ramírez, Rider and Virág (2011) for the β-Laguerre ensemble when $p/n \to \gamma \in [1, \infty)$. Recently, Jiang and Li (2014) studied the distribution of $\lambda_{(1)}$ when $p/n^3 \to \infty$. The large deviation principle for $\lambda_{(1)}$ has also been studied in the literature; see, for example, Chapter 2.6 in Anderson, Guionnet and Zeitouni (2010). Maïda (2007) investigated the large deviations for $\lambda_{(1)}$ of rank one deformations of Gaussian ensembles when $p/n \to \gamma \in [1, \infty)$, corresponding to the β-Laguerre ensemble with $\beta = 2$. Jiang and Li (2014) studied the case when $p/n \to \infty$ and derived the closed form of the large deviation rate function.

In practice, however, to estimate the tail probabilities of $\lambda_{(1)}$, especially when the probabilities are small, that is, rare events occur, approximations based on the large sample distribution and large deviation results may not be directly applicable or sufficiently precise. In particular, to our knowledge efficient estimation methods for the tail probabilities of $\lambda_{(1)}$ as well as sharp asymptotic approximations are still lacking in the literature.

1.2. *Our contributions.* The current paper deals with the efficient estimation of tail probabilities of $\lambda_{(1)}$. To do so, we study the extreme behaviors of the largest eigenvalue and describe the conditional distribution of $\lambda_{(1)}$ given the occurrence of the event $\{\lambda_{(1)} > px\}$. In particular, we use a so-called "three-step peeling" technique to approximate the tail probability (see the proofs of Theorem 1 and Lemma 7 in the Supplementary Material [Jiang, Leder and Xu (2017)]) and give

asymptotic approximations of $P(\lambda_{(1)} > px)$, which provides the necessary technical tools for the development and theoretical analysis of Monte Carlo based computational algorithms.

More importantly, from a computational point of view, we utilize the technique of *importance sampling* to develop an efficient Monte Carlo estimator of $P(\lambda_{(1)} > px)$. Importance sampling is commonly used as a numerical tool for estimating rare event probabilities in a wide variety of stochastic systems [see, e.g., Asmussen and Glynn (2007), Asmussen and Kroese (2006), Blanchet and Glynn (2008), Dupuis, Leder and Wang (2007), Liu and Xu (2014a), Siegmund (1976), Xu, Lin and Liu (2014)]. However, to the best of the authors' knowledge, this is the first use of this technique for estimating rare event probabilities in the spectrum of random matrices. In order to implement an importance sampling algorithm, it is necessary to construct an alternative sampling measure (or change of measure) under which the eigenvalues of the β-Laguerre ensemble are sampled. Ideally, one develops a sampling measure so that the event of interest is no longer rare under the sampling measure. The challenge is of course the construction of an appropriate sampling measure; one common heuristic is to utilize a sampling measure that approximates the conditional distribution of $\lambda_{(1)}$ given $\{\lambda_{(1)} > px\}$.

In this paper, we propose a change of measure denoted by Q that approximates the conditional measure $P(\cdot \mid \lambda_{(1)} > px)$ in total variation when p is much larger than n. The proposed change of measure is not of a classical exponential-tilting form commonly used in light-tailed stochastic systems [e.g., Asmussen and Glynn (2007), Siegmund (1976)] and it has features that are appealing both theoretically and computationally. Our proposed estimators are *asymptotically efficient* for all $p/n \to \gamma \in [1, \infty]$, that is, the second moments of estimators decay at the same exponential rate as the square of the first moments; see Section 2.2 for more details. Simulation studies in Section 3 show that the proposed method has the best performance among existing approximation approaches, especially when estimating probabilities of rare-events.

The proposed method can be easily generalized to the estimation of the smallest eigenvalue $\lambda_{(n)}$. With completely analogous analysis, we provide approximations of the tail probability of $\lambda_{(n)}$, that is,

$$P(\lambda_{(n)} < py) \qquad \text{as } p \to \infty$$

for any $\beta > 0$ and $0 < y < \beta$. Moreover, we construct the corresponding efficient simulation algorithms as shown in Section 2.4.

The rest of the paper is organized as follows. In Section 2, we present the main results, including asymptotic approximations of $P(\lambda_{(1)} > px)$ as well as efficient simulation algorithms. In Section 3, we illustrate the theoretical results through a simulation study and a real data example. Detailed proofs of main theorems and supporting lemmas are presented in Section 4 and the Supplementary Material, respectively.

Throughout this paper, we write $a_n = O(b_n)$ if $\limsup_{n\to\infty} |a_n|/|b_n| < \infty$, $a_n = \Theta(b_n)$ if $0 < \liminf_{n\to\infty} |a_n|/|b_n| \leq \limsup_{n\to\infty} |a_n|/|b_n| < \infty$, $a_n = o(b_n)$ if $\lim_{n\to\infty} |a_n|/|b_n| = 0$, $a_n \sim b_n$ if $\lim_{n\to\infty} |a_n|/|b_n| = 1$, $a_n \lesssim b_n$ if $\limsup_{n\to\infty} |a_n|/|b_n| \leq 1$, $a_n = O_p(b_n)$ if $a_n = O(b_n)$ in probability, and $a_n = o_p(b_n)$ if $a_n = o(b_n)$ in probability.

2. Main results. We are interested in efficiently estimating $P(\lambda_{(1)} > px)$, which converges to 0 as $p \to \infty$. In Section 2.1, we introduce some commonly used efficiency criteria in the literature; in Sections 2.2–2.4, we present the main asymptotic approximation results and the efficient simulation algorithms.

2.1. *Efficiency criteria in rare-event simulation.* In the context of rare-event simulations [e.g., Asmussen and Glynn (2007), Siegmund (1976)], it is necessary to consider the relative computational error with respect to the rare-event probability of interest. In particular, a Monte Carlo estimator L_p is said to be *asymptotically efficient* in estimating the rare-event probability $P(\lambda_{(1)} > px)$ if $E[L_p] = P(\lambda_{(1)} > px)$ and

$$(4) \qquad \lim_{p\to\infty} \frac{\log E[L_p^2]}{2 \log P(\lambda_{(1)} > px)} = 1.$$

Moreover, L_p is said to be *strongly efficient* if $E[L_p] = P(\lambda_{(1)} > px)$ and

$$(5) \qquad \limsup_{p\to\infty} \frac{E[L_p^2]}{P(\lambda_{(1)} > px)^2} < \infty.$$

There is a rich rare-event simulation literature. An incomplete list of recent works includes Adler, Blanchet and Liu (2012), Asmussen and Kroese (2006), Blanchet and Glynn (2008), Blanchet, Glynn and Leder (2012), Blanchet and Liu (2008), Dupuis, Leder and Wang (2007), Liu and Xu (2014a, 2014b), Xu, Lin and Liu (2014). It is interesting to note that the importance sampling measure we construct in this work has a similar structure to that used in Asmussen and Kroese (2006) where they were studying rare events for sums of i.i.d. heavy-tailed random variables.

REMARK 1. Suppose we plan to estimate $P(\lambda_{(1)} > px)$ with a given relative accuracy, that is, to compute an estimator Z_p such that

$$(6) \qquad P(|Z_p/P(\lambda_{(1)} > px) - 1| > \varepsilon) < \delta$$

for some prescribed $\varepsilon, \delta > 0$. For an estimator L_p, we can simulate N i.i.d. copies of L_p, $\{L_p^{(j)} : j = 1, \dots, N\}$ and obtain the final estimator $Z_p = \frac{1}{N} \sum_{j=1}^{N} L_p^{(j)}$. Then the estimation error is $|Z_p - P(\lambda_{(1)} > px)|$. When L_p is a strongly efficient estimator as defined in (5), the averaged estimator Z_p has a relative mean

squared error equal to $\text{Var}^{1/2}(L_p)/[N^{1/2}P(\lambda_{(1)} > px)]$. A simple application of Chebyshev's inequality yields that it suffices to simulate $N = \Theta(\varepsilon^{-2}\delta^{-1})$ i.i.d. replicates of L_p to achieve the accuracy in (6). When L_p is an asymptotically efficient estimator, it suffices to sample $N = \Theta(\varepsilon^{-2}\delta^{-1}P(\lambda_{(1)} > px)^{-\eta})$, for any $\eta > 0$, i.i.d. replicates of L_p. Compared with the crude Monte Carlo simulation, which requires $N = \Theta(\varepsilon^{-2}\delta^{-1}P(\lambda_{(1)} > px)^{-1})$ i.i.d. replicates, the efficient estimators substantially reduce the computational cost. See Section 3 for a simulation study and further discussion.

Importance sampling is one of the most widely used methods for variance reduction of Monte Carlo estimators. For ease of notation, we use P to denote the probability measure of the vector $(\lambda_1, \ldots, \lambda_n)$. The importance sampling estimator is constructed based on the following identity:

$$P(\lambda_{(1)} > px) = E[1_{(\lambda_{(1)} > px)}] = E^Q\left[1_{(\lambda_{(1)} > px)}\frac{dP}{dQ}\right],$$

where Q is a probability measure such that the Radon–Nikodym derivative dP/dQ is well defined on the set $\{\lambda_{(1)} > px\}$, and we use E and E^Q to denote the expectations under the measures P and Q, respectively. Then the random variable defined by

$$(7) \qquad\qquad L_p = \frac{dP}{dQ}1_{(\lambda_{(1)} > px)}$$

is an unbiased estimator of $P(\lambda_{(1)} > px)$ under the measure Q. Note that when generating the estimator (7) we sample $\lambda_{(1)}$ according to the new measure Q.

If we choose $Q(\cdot)$ to be $P_{px}^*(\cdot) := P(\cdot \mid \lambda_{(1)} > px)$, the conditional probability measure given $\lambda_{(1)} > px$, then the corresponding likelihood ratio dP/dQ is exactly $P(\lambda_{(1)} > px)$ on the set $\{\lambda_{(1)} > px\}$ and it has zero variance under Q. However, this change of measure is of no practical use since it needs the value of the target probability $P(\lambda_{(1)} > px)$. Nonetheless, this conditional measure P_{px}^* provides a guideline for constructing an efficient change of measure. If we can find a measure Q that is a good approximation of P_{px}^*, we would expect the corresponding estimator L_p defined in (7) to be efficient.

In the following, we design such change of measures for two different cases: $p/n \to \infty$ in Section 2.2 and $p/n \to \gamma \in [1, \infty)$ in Section 2.3. An analogous analysis of the smallest eigenvalue $\lambda_{(n)}$ is provided in Section 2.4.

2.2. *Efficient simulation for $P(\lambda_{(1)} > px)$ when $p/n \to \infty$.* To achieve efficient estimates as defined above, we need to approximate and bound the tail probability $P(\lambda_{(1)} > px)$ as well as the second moment of the estimator. In Section 2.2.1, we derive asymptotic approximations and bounds of $P(\lambda_{(1)} > px)$ under different conditions. We design efficient simulation algorithms in Section 2.2.2 and show that the estimate is efficient in the sense of (4) and (5).

2.2.1. *Tail probability approximation of* $\lambda_{(1)}$. We have the following approximations for $P(\lambda_{(1)} > px)$ when p is large. An exact approximation is given in Theorem 1 when $p/n^{5/3} \to \infty$. For the general case when $p/n \to \infty$, exact approximations are difficult to obtain and we provide tail approximation bounds, which are good enough to establish the efficiency of the simulation algorithm.

THEOREM 1. *Let* $x > \beta$. *When* $p/n^{5/3} \to \infty$ *as* $n \to \infty$,

$$(8) \qquad\qquad P(\lambda_{(1)} > px) \sim \exp\big(B_{n,p,\beta}(x)\big),$$

where $B_{n,p,\beta}(x)$ *is defined by*

$$B_{n,p,\beta}(x) = p\left(\frac{\beta}{2} - \frac{\beta}{2}\log\beta - \frac{x}{2} + \frac{\beta}{2}\log x\right) + \frac{\beta n}{2}\log\frac{p}{n} - \frac{\beta+1}{2}\log p$$

$$+ \beta n\left(-\frac{\log x}{2} + \log(x-\beta) - \frac{\log\beta}{2} + \frac{1}{2}\right) + \frac{1}{2}\log n - \frac{\beta^3 n^2}{2(x-\beta)^2 p}$$

$$- (\beta+1)\log(x-\beta) + \frac{\beta}{2}\log(2x) - \log(\pi) + \log\Gamma\left(1 + \frac{\beta}{2}\right).$$

More generally, if $n \to \infty$ *and* $p/n \to \infty$ *we have*

$$(9) \qquad\qquad \log P(\lambda_{(1)} > px) = B_{n,p,\beta}(x) + O(1)\frac{n^{5/2}}{p^{3/2}}.$$

REMARK 2. The tail probability approximation results in Theorem 1 provides technical support for the theoretical analysis of our importance sampling algorithm, where one needs to ensure that the exponential decay rate of the estimator variance matches that of the target tail probability. Although the term $O(1)$ in (9) in the general case is not specified, the developed approximations are sufficient enough to guarantee the asymptotical efficiency of the proposed Monte Carlo methods. Construction of the importance sampling estimator and the corresponding approximation results for the estimator variance will be provided in Section 2.2.2.

When n is fixed with $p \to \infty$, from the proof of Theorem 1, we have the same approximation results as in Theorem 1. In addition, we believe that it is possible to extend the proof to the case when $p = \Theta(n^{1+\epsilon})$ for any $\epsilon > 0$, and derive sharper asymptotic approximation results than Theorem 1. However, this involves the calculation of the expectation of $\exp\{-\sum_{i=1}^{n}(\lambda_i/p - \beta)^k\}$ for $k \geq 3$ (see the proof of Theorem 1 for more details). These extensions will be considered in future work.

2.2.2. *Efficient simulation method*. We characterize the proposed measure Q in (7) through two ways. First, we describe the simulation of the eigenvalues from Q by following a two-step procedure.

ALGORITHM 1. The algorithm goes as follows.

Step 1. Generate matrix $\mathbf{L}_{n-1,p-1,\beta} := \mathbf{B}_{n-1,p-1,\beta}\mathbf{B}_{n-1,p-1,\beta}^{\top}$, where $\mathbf{B}_{n-1,p-1,\beta}$ is a bidiagonal matrix defined by

$$\mathbf{B}_{n-1,p-1,\beta} = \begin{pmatrix} \chi_{\beta p-\beta} & & & \\ \chi_{\beta(n-2)} & \chi_{\beta p-2\beta} & & \\ & \ddots & \ddots & \\ & & \chi_{\beta} & \chi_{\beta p-\beta(n-1)} \end{pmatrix}_{(n-1)\times(n-1)}.$$

Here, all of the diagonal and sub diagonal elements are mutually independent with the distribution of χ_a, the square-root of the chi-square distribution with degree of freedom a. Calculate the corresponding eigenvalues $(\lambda_2, \ldots, \lambda_n)$ of $\mathbf{L}_{n-1,p-1,\beta}$ and the order statistics $\lambda_{(2)} > \cdots > \lambda_{(n)}$.

Step 2. Conditional on $(\lambda_{(2)}, \ldots, \lambda_{(n)})$, sample $\lambda_{(1)}$ from the exponential distribution with density

$$(10) \qquad f(\lambda_{(1)}) := \frac{x-\beta}{2x} e^{-\frac{x-\beta}{2x}(\lambda_{(1)} - px \vee \lambda_{(2)})} \cdot I_{(\lambda_{(1)} > px \vee \lambda_{(2)})},$$

where $x > \beta$ is the threshold value in $P(\lambda_{(1)} > px)$.

Let Q be the measure induced by combining the above two-step sampling procedure on $(\lambda_{(1)}, \ldots, \lambda_{(n)})$. It is defined on $[0, \infty)^n$. We next describe it using the Radon–Nikodym derivative between Q and the original measure P. From Dumitriu and Edelman (2002), we know the order statistics of the eigenvalues of $\mathbf{L}_{n-1,p-1,\beta}$ has density function

$$g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n) = (n-1)! f_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n) \times I_{(\lambda_2 > \cdots > \lambda_n)}$$

as defined in (3). Then the sampled $\lambda_{(1)}, \ldots, \lambda_{(n)}$ under Q has density function

$$(11) \qquad g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n) \cdot \frac{x-\beta}{2x} e^{-\frac{x-\beta}{2x}(\lambda_1 - px \vee \lambda_2)} I_{(\lambda_1 > px \vee \lambda_2)}.$$

The corresponding importance sampling estimator following (7) is

$$L_p = \frac{g_{n,p,\beta}(\lambda_{(1)}, \ldots, \lambda_{(n)}) 1_{(\lambda_{(1)} > px)}}{g_{n-1,p-1,\beta}(\lambda_{(2)}, \ldots, \lambda_{(n)}) \cdot \frac{x-\beta}{2x} e^{-\frac{x-\beta}{2x}(\lambda_{(1)} - px \vee \lambda_{(2)})} I_{(\lambda_{(1)} > px \vee \lambda_{(2)})}}.$$

The joint density $g_{n,p,\beta}(\lambda_1, \ldots, \lambda_n)$ equals

$$I_{(\lambda_1 > \cdots > \lambda_n)} \times n! f_{n,p,\beta}(\lambda_1, \ldots, \lambda_n)$$

$$= I_{(\lambda_1 > \cdots > \lambda_n)} \times n! c_{n,p,\beta} \prod_{1 \le i < j \le n} |\lambda_i - \lambda_j|^{\beta} \cdot \prod_{i=1}^{n} \lambda_i^{\frac{\beta(p-n+1)}{2} - 1} \cdot e^{-\frac{1}{2}\sum_{i=1}^{n} \lambda_i}$$

$$(12)$$

$$= I_{(\lambda_1 > \cdots > \lambda_n)} n A_n \prod_{i=2}^{n} (\lambda_1 - \lambda_i)^{\beta} \cdot \lambda_1^{\frac{\beta(p-n+1)}{2} - 1} \cdot e^{-\frac{1}{2}\lambda_1}$$

$$\times g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n),$$

where

(13) $$A_n = c_{n,p,\beta}/c_{n-1,p-1,\beta},$$

with $c_{n,p,\beta}$ and $c_{n-1,p-1,\beta}$ defined as in (2). Therefore, the importance sampling estimator L_p can be written as

(14) $$L_p = \frac{nA_n \prod_{i=2}^n (\lambda_{(1)} - \lambda_{(i)})^\beta \cdot \lambda_{(1)}^{\frac{\beta(p-n+1)}{2}-1} \cdot e^{-\frac{1}{2}\lambda_{(1)}}}{\frac{x-\beta}{2x} e^{-\frac{x-\beta}{2x}(\lambda_{(1)}-px\vee\lambda_{(2)})} \cdot I_{(\lambda_{(1)} > px\vee\lambda_{(2)})}} 1_{(\lambda_{(1)} > px)}.$$

Under the measure Q, $\lambda_{(1)} > px \vee \lambda_{(2)}$ and, therefore, L_p is well defined.

The measure Q is constructed such that the behavior of the eigenvalues under Q mimics the tail behavior given the rare event $\{\lambda_{(1)} > px\}$ under P. According to the proposed simulation procedure, the largest eigenvalue is generated from a truncated exponential distribution at the level about px while the other eigenvalues are generated from the original measure. We have the following theorem to show the efficiency of the proposed measure.

THEOREM 2. (i) *If* $p/n^{5/3} \to \infty$, *the measure Q approximates* P_{px}^*, *the conditional probability measure given* $\lambda_{(1)} > px$, *in the total variation sense, that is,*

$$\lim_{n\to\infty} \sup_{A\in\mathcal{F}} |Q(A) - P_{px}^*(A)| = 0,$$

where $\mathcal{F}$ is the σ-field $\mathcal{B}([0,\infty)^n)$. In addition,

$$E^Q[L_p^2] \sim P(\lambda_{(1)} > px)^2.$$

(ii) *If* $n^{5/3}/p = O(1)$, *we have*

$$E^Q[L_p^2] = O(1)P(\lambda_{(1)} > px)^2,$$

that is, the importance sampling estimate based on Q is strongly efficient.

(iii) *More generally, if* $p/n \to \infty$, *we have*

$$\frac{\log E^Q[L_p^2]}{2\log P(\lambda_{(1)} > px)} \to 1,$$

that is, the importance sampling estimate is asymptotically efficient.

REMARK 3. Theorem 2 shows that the conditional distribution of $(\lambda_1, \ldots, \lambda_n)$ given $\lambda_{(1)} > px$ essentially behaves like the proposed measure Q. When n is fixed and $p \to \infty$, a similar argument as in the proof of Theorem 2 gives that $E^Q[L_p^2] \sim P(\lambda_{(1)} > px)^2$ and the importance sampling estimate is strongly efficient.

It is conceived that the total variation distance between the proposed measure Q and the conditional distribution converges to 0 for the general case when $p = \Theta(n^{1+\epsilon})$, $\epsilon > 0$. As discussed in Remark 2, this needs the calculation of the expectation of $\exp\{-\sum_{i=1}^n (\lambda_i/p - \beta)^k\}$ for $k \geq 3$, which we would like to investigate in the future.

REMARK 4.   From the proof of Theorem 2, we can see that the estimator is still asymptotically efficient if in the second step of Algorithm 1, we sample $\lambda_{(1)}$ from an alternative exponential distribution with density

$$f(\lambda_{(1)}) := J_{\beta,x} e^{-J_{\beta,x}(\lambda_{(1)} - px \vee \lambda_{(2)})} \cdot I_{(\lambda_{(1)} > px \vee \lambda_{(2)})},$$

where the rate $J_{\beta,x}$ is some positive constant smaller than $(x - \beta)/x$. However, as shown in Theorem 2, when $J_{\beta,x} = (x - \beta)/(2x)$, the change of measure approximates the conditional distribution given $\lambda_{(1)} > px$ in total variation when p is large, and it is conceivable that this rate function yields more efficient results than others.

REMARK 5.   The above results show that the estimator L_p is asymptotically efficient. To estimate $P(\lambda_{(1)} > px)$, we simulate N i.i.d. copies of L_p, $\{L_p^{(j)} : j = 1, \ldots, N\}$ and the final estimator is $Z_p = \frac{1}{N} \sum_{j=1}^{N} L_p^{(j)}$. To achieve the accuracy in (6), by the above theorem, at most we need $N = \Theta(\varepsilon^{-2} \delta^{-1})$ if $p/n^{5/3} \to \infty$ or $N = \Theta(\varepsilon^{-2} \delta^{-1} P(\lambda_{(1)} > px)^{-\eta})$, for any $\eta > 0$ and $p/n \to \infty$.

2.3. *Efficient simulation for $P(\lambda_{(1)} > px)$ when $p/n \to \gamma \in [1, \infty)$.*   When $p/n \to \gamma \in [1, \infty)$, from a direct application of Theorem 2.6.6 [Anderson, Guionnet and Zeitouni (2010)], $\lambda_{(1)}/n$ satisfies the large deviation principle in $\mathbb{R}$ with speed n and good rate function

$$(15) \quad I_\beta(x) = \begin{cases} -\beta \int_{\mathbb{R}} \log|x - y| \sigma_\beta(dy) + \frac{1}{2} x - \frac{1}{2} \beta(\gamma - 1) \log x + \alpha_\beta, \\ \qquad \text{if } x \geq x^*; \\ \infty, \qquad \text{if } x < x^*, \end{cases}$$

where $\alpha_\beta = \frac{\beta}{2}[(\gamma + 1)(\log \beta - 1) + \gamma \log \gamma]$, and σ_β is the Marchenko–Pastur law [Marčenko and Pastur (1967)] corresponding to the empirical distribution of eigenvalues $(\lambda_1/n, \ldots, \lambda_n/n)$ with $x_* = \beta(\sqrt{\gamma} - 1)^2$, $x^* = \beta(\sqrt{\gamma} + 1)^2$; see also, for example, Hiai and Petz (1998) and Dumitriu (2003) for more details.

We now consider the tail probability $P(\lambda_{(1)}/p > x) = P(\lambda_{(1)}/n > (p/n)x)$ for $\gamma x > x^*$. From the large deviation result, we know $P(\lambda_{(1)} > px)$ converges to 0 as $n \to \infty$. To construct an efficient estimator, the proposed algorithm in Section 2.2.2 cannot be directly applied and we need to modify the change of measure accordingly.

The new algorithm is given as follows: keep Step 1 in the algorithm from Section 2.2.2. In Step 2, we sample $\lambda_{(1)}$ from the exponential distribution with density

$$f(\lambda_{(1)}) = J_{\beta,x} e^{-J_{\beta,x} \times (\lambda_{(1)} - px \vee \lambda_{(2)})} \cdot I_{(\lambda_{(1)} > px \vee \lambda_{(2)})},$$

where the rate $J_{\beta,x} > 0$ is chosen such that

$$(16) \qquad J_{\beta,x} < 1 - 2\beta \int \frac{1}{\gamma x - y} \sigma_\beta(dy) - \beta \frac{\gamma - 1}{\gamma x}.$$

The quantity in the right-hand side of (16) is the derivative of the rate function $2I_\beta$ at $\gamma x > x^*$. It is positive on (x^*, ∞) due to the fact that the rate function $I_\beta(x)$ is a convex function with positive second derivative on set (x^*, ∞) and it achieves the minimum 0 at x^* [Theorem 2.6.6, Anderson, Guionnet and Zeitouni (2010)]. Therefore, the constant $J_{\beta,x}$ is well defined.

Let $\tilde{Q}$ be the measure induced by combining the above two-step sampling procedure on $(\lambda_{(1)}, \dots, \lambda_{(n)})$. It is defined on $[0, \infty)^n$. By the same argument as in that between (10) and (12), we know the corresponding importance sampling estimate $\tilde{L}_p := \frac{dP}{d\tilde{Q}}$ is given by

$$(17) \qquad \tilde{L}_p = \frac{n A_n \prod_{i=2}^n (\lambda_{(1)} - \lambda_{(i)})^\beta \cdot \lambda_{(1)}^{\frac{\beta(p-n+1)}{2} - 1} \cdot e^{-\frac{1}{2}\lambda_{(1)}}}{J_{\beta,x} e^{-J_{\beta,x}(\lambda_{(1)} - px \vee \lambda_{(2)})} \cdot I_{(\lambda_{(1)} > px \vee \lambda_{(2)})}} 1_{(\lambda_{(1)} > px)}.$$

We have the following efficiency result for $\tilde{L}_p$.

THEOREM 3. *If $p/n \to \gamma \in [1, \infty)$, then the importance sampling estimate $\tilde{L}_p$ is asymptotically efficient.*

REMARK 6. To achieve strong efficiency results as in Theorem 2, we need to derive a more accurate approximation of the tail probability $P(\lambda_{(1)} > px)$ as in Theorem 1. However, the techniques developed in this paper may not be directly applicable, though some of the derived approximation bounds in the auxiliary lemmas (such as bounds in the proof of Lemma 5) can be generalized to the case of $p/n \to \gamma \in [1, \infty)$. We leave this as a future work.

2.4. *Efficient simulation for $\lambda_{(n)}$.* Recall that $\lambda_{(n)}$ is the smallest eigenvalue of the β-Laguerre ensemble defined as in (3). In this section, we focus on the probability $P(\lambda_{(n)} < py)$ as $p \to \infty$ for any $\beta > 0$ and $0 < y < \beta$. We have the following approximation results similar to Theorem 1. Their proofs follow from analogous arguments as in those for $\lambda_{(1)}$ and, therefore, are omitted.

THEOREM 4. *For $0 < y < \beta$ the following hold as $n \to \infty$:*

(1) *If $p/n \to \infty$, $\log P(\lambda_{(n)} < py) = B_{n,p,\beta}(y) + O(1)n^{5/2}p^{-3/2}$, where $B_{n,p,\beta}(y)$ is defined as in Theorem 1.*
(2) *If $p/n^{5/3} = \Theta(1)$, $\log P(\lambda_{(n)} < py) = B_{n,p,\beta}(y) + O(1)$.*
(3) *If $p/n^{5/3} \to \infty$, $P(\lambda_{(n)} < py) \sim \exp(B_{n,p,\beta}(y))$.*

Since $\lambda_{\max}/p$ and $\lambda_{\min}/p$ are all positive, it can be seen from Theorems 1 and 4 that the two rate functions on $(0, \infty)$ look "symmetric" with respect to the line $x = \beta$. The dominant term in the above expression of $B_{n,p,\beta}(y)$ is $p(\frac{\beta}{2} - \frac{\beta}{2} \log \beta - \frac{y}{2} + \frac{\beta}{2} \log y)$, which is negative if $0 < y < \beta$, and thus $\exp(B_{n,p,\beta}(y))$ is no more

than 1. In addition, this gives the same exponential decay rate for $\lambda_{(n)}$ as found in
Jiang and Li (2014).

To obtain an efficient Monte Carlo estimator of $P(\lambda_{(n)} < py)$, we propose an
importance sampling procedure similar to that for the largest eigenvalue $\lambda_{(1)}$.

Step 1. Generate matrix $\mathbf{L}_{n-1,p-1,\beta}$. Calculate the corresponding eigenvalues
$(\lambda_1, \ldots, \lambda_{n-1})$ of $\mathbf{L}_{n-1,p-1,\beta}$ and the order statistics $\lambda_{(1)} > \cdots > \lambda_{(n-1)}$.

Step 2. Conditional on $(\lambda_{(1)}, \ldots, \lambda_{(n-1)})$, sample $\lambda_{(n)}$ from the distribution
with density

$$f(\lambda_{(n)}) := \frac{\beta - y}{2y} e^{\frac{\beta - y}{2y}(\lambda_{(n)} - py \wedge \lambda_{(n-1)})} \cdot I_{(\lambda_{(n)} < py \wedge \lambda_{(n-1)})}.$$

The importance sampling estimator L_p can be written as

$$L_p = \frac{n A_n \prod_{1 \leq i < n} (\lambda_{(i)} - \lambda_{(n)})^\beta \cdot \lambda_{(n)}^{\frac{\beta(p-n+1)}{2} - 1} \cdot e^{-\frac{1}{2}\lambda_{(n)}}}{\frac{\beta - y}{2y} e^{\frac{\beta - y}{2y}(\lambda_{(n)} - py \wedge \lambda_{(n-1)})} \cdot I_{(\lambda_{(n)} < py \wedge \lambda_{(n-1)})}} I_{(\lambda_{(n)} < py)}.$$

The efficiency of the above importance sampling estimator is stated in the next
theorem.

THEOREM 5.   *Assume $0 < y < \beta$ and $n \to \infty$. We have*:

(1) *if $p/n^{5/3} \to \infty$, $E^Q[L_p^2] \sim P(\lambda_{(n)} < py)^2$;*
(2) *if $p/n^{5/3} = \Theta(1)$, $E^Q[L_p^2] = O(1)P(\lambda_{(n)} < py)^2$;*
(3) *if $p/n \to \infty$, L_p is asymptotically efficient.*

3. Numerical study.

3.1. *Simulation study.*   In order to evaluate the actual performance of our al-
gorithms, we conduct a numerical study over different p and n values. We take
$\beta = 1$ and choose six combinations of n and p: $(n, p) = (10, 10^2)$, $(10, 10^3)$,
$(10, 10^4)$, $(50, 10^2)$, $(50, 10^3)$ and $(50, 10^4)$. We follow Algorithm 1 to estimate
$P(\lambda_{(1)} > px)$ for different values of x's. For $(n, p) = (10, 100)$ and $(50, 100)$,
the algorithm in Section 2.3 gives similar results and, therefore, are not presented.
Based on the simulation results, we would suggest use Algorithm 1 in practice for
$p/n \to \gamma \in [1, \infty]$.

Tables 1 and 2 show estimated tail probabilities (column "Est") along with the
estimated standard deviations $\text{Std}(L_p) = \sqrt{\text{Var}^Q(L_p)}$ (column "Std"). The sim-
ulation results are based on 10^4 independent simulations and it takes just a few
seconds in the statistical software "R" for each case. Note that the standard devia-
tion of the final estimate (in the column "Est.") is the reported standard deviation
(in the column of "Std.") divided by $\sqrt{10^4} = 100$.

TABLE 1
Estimates of $P(\lambda_{(1)} > px)$ for $n = 10$ and $p = 100, 1000, 10{,}000$. The standard deviation of the estimate "Est." is Std./100

x	Est.	Std.	Std./Est.	TA	TW	DMC
			$n = 10, p = 100$			
1.9	1.01e–02	6.91e–03	0.69	5.89e–03	1.34e–02 (1.06e–02)	1.00e–02
2.0	1.71e–03	9.95e–04	0.58	1.07e–03	2.20e–03 (1.68e–03)	1.74e–03
2.1	2.31e–04	1.14e–04	0.49	1.55e–04	2.79e–04 (2.06e–04)	2.41e–04
2.5	1.64e–08	5.43e–09	0.33	1.29e–08	–	–
3	9.24e–15	2.16e–15	0.23	7.91e–15	–	–
4	2.21e–29	3.33e–30	0.15	2.03e–29	–	–
			$n = 10, p = 1000$			
1.25	1.16e–02	9.08e–03	0.78	6.81e–03	1.75e–02 (1.28e–02)	1.16e–02
1.28	1.19e–03	7.65e–04	0.64	7.70e–04	2.04e–03 (1.42e–03)	1.22e–03
1.3	2.07e–04	1.22e–04	0.59	1.43e–04	3.93e–04 (2.66e–04)	2.02e–04
1.4	3.88e–09	1.50e–09	0.39	3.11e–09	–	–
1.5	3.26e–15	9.56e–16	0.29	2.82e–15	–	–
2.0	1.88e–59	2.59e–60	0.14	1.81e–59	–	–
			$n = 10, p = 10{,}000$			
1.07	4.26e–02	4.17e–02	0.97	2.32e–02	6.18e–02 (4.58e–02)	4.29e–02
1.08	4.05e–03	3.03e–03	0.75	2.52e–03	6.98e–03 (4.84e–03)	4.02e–03
1.09	2.15e–04	1.26e–04	0.59	1.45e–04	4.69e–04 (3.10e–04)	2.18e–04
1.10	6.49e–06	3.41e–06	0.52	4.76e–06	1.84e–05 (1.11e–05)	5e–06
1.15	1.36e–16	4.19e–17	0.31	1.18e–16	–	–
1.20	8.15e–32	1.81e–32	0.22	7.49e–32	–	–

To validate our importance sampling results, we compute direct Monte Carlo estimates based on 10^6 independent simulations (column "DMC"). Note that this validation is not feasible for all probabilities considered. We also present the results from asymptotic approximation methods. The tail probability approximations from Theorem 1 are presented in the column "TA" and the approximation results based on the Tracy–Widom distribution are given in the column "TW." The tail probabilities of the Tracy–Widom distribution are calculated using R package *"RMTstat"* [Johnstone et al. (2010)]. In particular, it is known that when $\beta = 1$

$$\frac{\lambda_{(1)} - \mu_{n,p}}{\sigma_{n,p}}$$

converges to the Tracy–Widom law [El Karoui (2003), Johnstone (2001)], where

$$\mu_{n,p} = (\sqrt{n} + \sqrt{p-1})^2,$$

$$\sigma_{n,p} = (\sqrt{n} + \sqrt{p-1})\left(\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p-1}}\right)^{1/3}.$$

TABLE 2
Estimates of $P(\lambda_{(1)} > px)$ for $n = 50$ and $p = 100, 1000, 10,000$

x	Est.	Std.	Std./Est.	TA	TW	DMC
			$n = 50, p = 100$			
3	4.74e–02	1.16e–01	2.46	4.57e–02	5.07e–02 (4.84e–02)	4.77e–02
3.25	9.15e–04	1.44e–03	1.58	7.22e–04	8.94e–04 (8.36e–04)	9.01e–04
3.5	6.02e–06	5.56e–06	0.92	4.47e–06	2.38e–06 (2.03e–06)	-
4	3.09e–11	1.94e–11	0.63	2.18e–11	–	–
5	3.05e–24	1.17e–24	0.39	2.27e–24	–	–
			$n = 50, p = 1000$			
1.55	3.51e–03	6.57e–03	1.87	4.43e–03	4.30e–03 (3.60e–03)	3.46e–03
1.57	5.81e–04	8.43e–04	1.45	6.60e–04	7.20e–04 (5.91e–04)	5.77e–04
1.6	2.74e–05	3.14e–05	1.15	2.79e–05	3.25e–05 (2.56e–05)	2.0e–05
1.7	8.52e–11	6.63e–11	0.78	7.31e–11	–	–
2.0	2.32e–34	8.90e–35	0.38	1.91e–34	–	–
			$n = 50, p = 10,000$			
1.155	1.75e–02	3.68e–02	2.11	3.21e–02	2.25e–02 (1.86e–02)	1.77e–02
1.16	3.99e–03	7.37e–03	1.84	6.09e–03	5.22e–03 (4.20e–03)	3.87e–03
1.17	1.06e–04	1.50e–04	1.41	1.33e–04	1.57e–04 (1.20e–04)	1.02e–04
1.20	4.22e–11	3.62e–11	0.86	3.95e–11	–	–
1.25	2.75e–26	1.47e–26	0.53	2.36e–26	–	–

A more accurate approximation has been proposed in Johnstone and Ma (2012) and Ma (2012), where

$$\mu_{n,p} = \left(\sqrt{n - \frac{1}{2}} + \sqrt{p - \frac{1}{2}} \right)^2,$$

$$\sigma_{n,p} = \left(\sqrt{n - \frac{1}{2}} + \sqrt{p - \frac{1}{2}} \right) \left(\frac{1}{\sqrt{n - \frac{1}{2}}} + \frac{1}{\sqrt{p - \frac{1}{2}}} \right)^{1/3}.$$

We report both approximation results in the column "TW" with the second in the parentheses.

From Tables 1 and 2, we can see that the proposed importance sampling estimates ("Est") are consistent with those from direct Monte Carlo simulation ("DMC"). The ratios between the estimated standard deviations of L_p and the estimated tail probabilities ("Std/Est") stay reasonably small, indicating the efficiency of the algorithm [see equations (4) and (5)]. The ratio becomes smaller as x increases. This implies that the algorithm is more efficient for larger x's values. Moreover, the proposed method provides an efficient way to evaluate the performance of the theoretical approximation methods. In particular, we can see that the approximations based on the Tracy–Widom distribution ("TW") overestimate the tail probabilities, especially for larger x's values. In addition, larger estimation

(relative) errors can be observed for $n = 10$ than for $n = 50$. For the tail approximations ("TA"), we can see they do not give accurate estimates for smaller x's values while the performance gets better as x increases. Overall, the importance sampling method outperforms results based on the Tracy–Widom distribution and the tail probability approximations. Lastly, it should be noted that the approximations based on Tracy–Widom distribution and the direct Monte Carlo approach are not suitable for estimating the probability of extremely rare events. For these events the only possibility is to use the methods developed in the current work, that is, importance sampling or the the tail approximation. R code of the proposed importance sampling algorithm can be found at http://users.stat.umn.edu/~xuxxx360/IS.R.

3.2. *Non-Gaussian matrices.* We next investigate the behavior of the algorithm and approximations in the non-Gaussian setting. In particular, we generate the matrix $\mathbf{X}^*\mathbf{X}$ where $\mathbf{X} = (x_{ij})_{n \times p}$ with $n = 10$, $p = 100$, 1000 and 10,000, and x_{ij} are i.i.d. random variables following a standardized Binomial or t-distribution with mean zero and variance one. Table 3 presents the simulation results, where columns "Est" is the importance sampling estimates under the Gaussian assumption as in Table 1, "TW" is the Tracy–Widom estimates, and the last five columns are direct Monte Carlo results under different standardized distributions with 10^5 replications. Table 3 shows that the importance sampling estimators are generally comparable to Tracy–Widom estimators. Moreover, as the distribution of $x_{1,1}$ becomes more like the normal distribution (such as when the number of trials of a Binomial distribution increases or the degrees of freedom of a t-distribution increases), the importance sampling estimators become more accurate and outperform Tracy–Widom estimators.

Bordenave and Caputo (2014) studied the large deviations properties of the spectrum of Wigner matrices whose entries were random variables with density proportional to $e^{-|x|^\alpha}$ for all $x \in \mathbb{R}$ with parameter $\alpha \in (0, 2)$. Interestingly, they

TABLE 3
Estimates of $P(\lambda_{(1)} > px)$ for $n = 10$ and $p = 100, 1000, 10{,}000$

x	Est.	TW	$B_{(5, 0.5)}$	$B_{(10, 0.5)}$	$B_{(20, 0.5)}$	t_{50}	t_{100}
			$n = 10, p = 100$				
1.9	1.01e–2	1.34 (1.06)e–2	0.57e–2	0.81e–2	0.88e–2	1.15e–2	1.08e–2
2.0	1.71e–3	2.20 (1.68)e–3	0.72e–3	1.27e–3	1.45e–3	1.93e–3	1.76e–3
			$n = 10, p = 1000$				
1.25	1.16e–2	1.75 (1.28)e–2	0.72e–2	1.00e–2	1.10e–2	1.36e–2	1.22e–2
1.28	1.19e–3	2.04 (1.42)e–3	0.55e–3	0.94e–3	1.13e–3	1.61e–3	1.20e–3
			$n = 10, p = 10{,}000$				
1.07	4.26e–2	6.18 (4.58)e–2	3.04e–2	3.65e–2	3.96e–2	4.85e–2	4.49e–2
1.08	4.05e–3	6.98 (4.84)e–3	2.16e–3	3.20e–3	3.61e–3	4.77e–2	4.25e–3

observed that both the speed and the rate function of the large deviations principle depends on the parameter α. Although the Wigner matrices and the Wishart matrices belong to different ensembles, their large deviation principles are of similar structures; see, for examples, Anderson, Guionnet and Zeitouni (2010). Therefore, we do not expect a big universality family for our theoretical approximation results. However, as illustrated in the above simulation, the importance sampling estimator based on the Gaussian assumption will provide an adequate approximation for many cases.

3.3. *Dengue virus example.* To illustrate the use of our algorithm, we consider a real data set of immunity to the Dengue virus (DENV). The data set contains the innate immune response to DENV infection in whole blood samples of acutely infected humans in Bangkok, Thailand during the season of 2009 [Kwissa et al. (2014)]. The data set can be downloaded from http://www.ncbi.nlm.nih.gov/ geo/query/acc.cgi?acc=GSE51808. Whole blood samples were analyzed from 18 dengue fever patients and 10 dengue hemorrhagic fever patients hospitalized at the Siriraj Hospital in Bangkok, Thailand, and the samples were obtained between days 2 and 9 after onset of symptoms. Blood samples from 19 convalescence patients were also obtained at 4 weeks or later after discharge. In addition, there is a control group of 9 healthy donors, and their blood was also sampled.

We consider data from four groups: 18 dengue fever patients; 10 dengue hemorrhagic fever patients; 19 convalescence patients; and 9 healthy control patients. Each individual has $p = 54,715$ covariates and Robust Multi-array Average (RMA) normalization was performed using Expression Console software [Kwissa et al. (2014)]. We further standardize each group by their global mean and standard deviation. For each group, we consider the null hypothesis that their covariance matrix is a white Wishart matrix. We calculate the largest eigenvalues from the four groups' covariance matrices as our test statistics. The eigenvalues of the matrix $\mathbf{X}^*\mathbf{X}/p$ of each group is given in the Figure 1. We compute the corresponding p-values of the four test statistics to be $\ll 10^{-10}$ and, therefore, we reject all four null hypotheses.

4. Proof of theorems. In this section, we present the proofs of main theorems. Technical lemmas and their proofs are provided in the Supplementary Material [Jiang, Leder and Xu (2017)].

Outline of proofs. In our analysis of the tail probability $P(\lambda_{(1)} > px)$ as well as the variance of the Monte Carlo estimator, we frequently use the following factorization of the joint probability density function (p.d.f.) in (3):

$$g_{n,p,\beta}(\lambda_1, \ldots, \lambda_n)$$
$$= nA_n \prod_{i=2}^{n} (\lambda_1 - \lambda_i) \lambda_1^{\frac{\beta(p-n+1)}{2} - 1} e^{-\lambda_1/2} g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n),$$
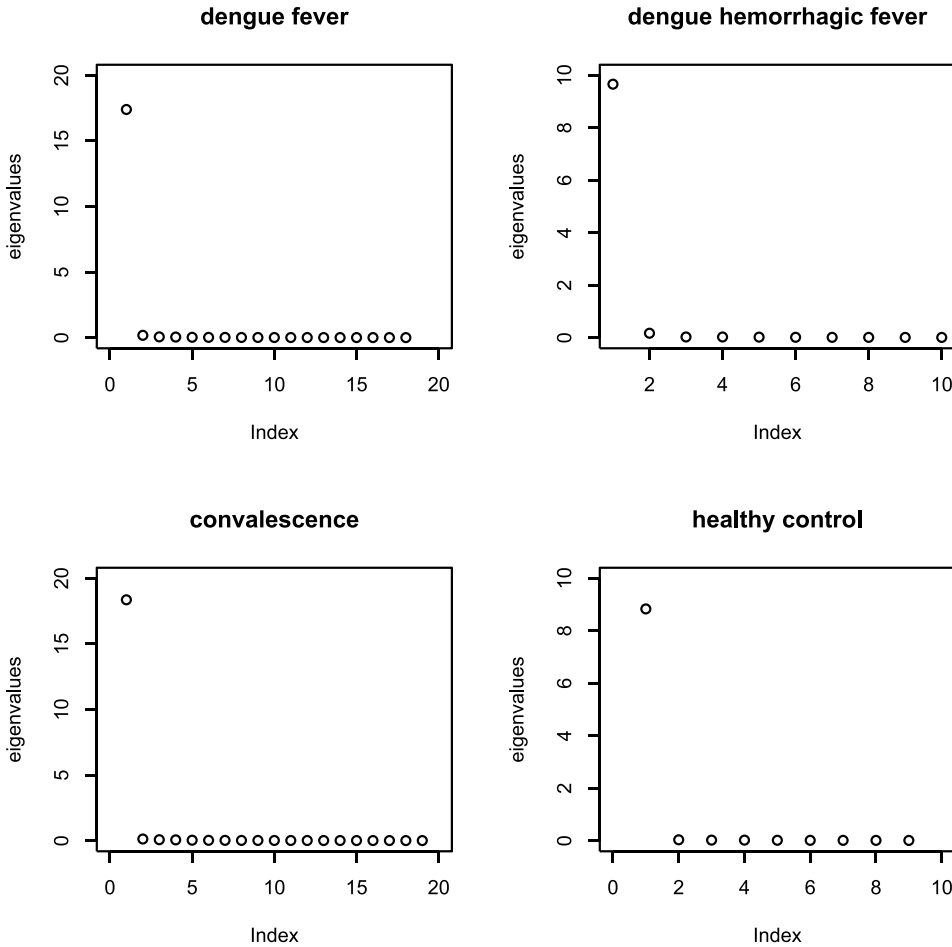
FIG. 1. *Plot of eigenvalues.*

where A_n is defined as in (13). Some key lemmas are developed to show that in the case $p/n \to \infty$ we can find a positive sequence δ_n converging to zero (in certain rate) such that we can focus on the event

$$(18) \qquad \{p(x + \delta_n) > \lambda_{(1)} > px, \lambda_{(2)} < p(\beta + \delta_n), \lambda_{(n)} > p(\beta - \delta_n)\}$$

instead of $\{\lambda_{(1)} > px\}$. Roughly speaking, we can think of this as the following: If $\lambda_{(1)} > px$, then (i) $\lambda_{(1)} \approx px$ and (ii) the remaining eigenvalues are approximately $p\beta$; see Lemmas 6 and 7 [Jiang, Leder and Xu (2017)] for further details. With the result in (18), we can approximate $\prod_{i=2}^{n}(\lambda_1 - \lambda_i)$ by

$$\psi_{n,p,\beta}(\lambda_1, \ldots, \lambda_n) = (px - p\beta)^{\beta(n-1)} e^{(n+o(n)) \frac{\lambda_1 - px}{px - p\beta} - \sum_{i=2}^{n} \frac{\lambda_i - p\beta}{px - p\beta} - \alpha \sum_{i=2}^{n} (\frac{\lambda_i - p\beta}{px - p\beta})^2},$$

with α being approximately equal to $1/2$. Consider then approximating $g_{n,p,\beta}$ with

$$n A_n \psi_{n,p,\beta}(\lambda_1, \ldots, \lambda_n) \lambda_1^{\frac{\beta(p-n+1)}{2}-1} e^{-\lambda_1/2} g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n).$$

The benefit of working with the approximation in the previous display is that when integrating we can factor our integrand into the product of terms involving λ_1 and terms involving λ_i for $i \in \{2, \ldots, n-1\}$. The term involving λ_1 are quite simple, basically boiling down to the p.d.f. of a gamma distributed random variable. The term with λ_i for $i > 1$ requires the approximation of

$$(19) \qquad E \exp\left[-\sum_{i=2}^n \frac{\lambda_i - p\beta}{px - p\beta} - \alpha \sum_{i=2}^n \left(\frac{\lambda_i - p\beta}{px - p\beta}\right)^2\right].$$

We achieve this through using a matrix representation of the β-Laguerre ensemble by Dumitriu and Edelman (2002) to express $\sum_{i=2}^n \lambda_i$ and $\sum_{i=2}^n (\lambda_i - p\beta)^2$ as sums of independent random variables. We are then able to develop approximations to the expected value in (19); see Lemma 3 in the Supplementary Material [Jiang, Leder and Xu (2017)].

PROOF OF THEOREM 1. We first focus on the case when $p/n^{5/3} \to \infty$. Set $a_n = \sqrt{np^{-1} + p^{-1}}$ and $b_n = pn^{-2}$. Choose $\delta_n = \min\{\sqrt{a_n}, \sqrt{a_n b_n}\}$. From the assumption $p/n^{5/3} \to \infty$, it is trivial to check that

$$(20) \qquad \delta_n \to 0, \qquad p\delta_n \to \infty, \qquad \frac{\delta_n^2 p}{n} \to \infty, \quad \text{and} \quad \frac{\delta_n n^2}{p} \to 0.$$

In the discussion below, whenever we need a restriction about δ_n we can always get it from the above limits.

To prove (8), we first show that

$$(21) \quad \begin{aligned} &P(\lambda_{(1)} > px) \\ &\qquad \gtrsim \frac{2x}{x-\beta} n A_n (px - p\beta)^{\beta(n-1)} (px)^{\frac{\beta(p-n+1)}{2}-1} e^{-\frac{px}{2} - \frac{\beta^3 n^2}{2(x-\beta)^2 p}}; \end{aligned}$$

$$(22) \quad \begin{aligned} &P(\lambda_{(1)} > px) \\ &\qquad \lesssim \frac{2x}{x-\beta} n A_n (px - p\beta)^{\beta(n-1)} (px)^{\frac{\beta(p-n+1)}{2}-1} e^{-\frac{px}{2} - \frac{\beta^3 n^2}{2(x-\beta)^2 p}}. \end{aligned}$$

We prove them separately.

Proof of (21). By Lemmas 6 and 7 in the Supplementary Material [Jiang, Leder and Xu (2017)],

$$P(\lambda_{(1)} > p(x + \delta_n)) = o(1) P(\lambda_{(1)} > px),$$

$$P(\lambda_{(1)} > px, \lambda_{(2)} > p(\beta + \delta_n)) = o(1) P(\lambda_{(1)} > px),$$

$$P(\lambda_{(1)} > px, \lambda_{(n)} < p(\beta - \delta_n)) = o(1) P(\lambda_{(1)} > px).$$

Therefore, $P(\lambda_{(1)} > px)$ is asymptotically equivalent to the probability of $\{px < \lambda_{(1)} < p(x + \delta_n), \lambda_{(2)} < p(\beta + \delta_n),$ and $\lambda_{(n)} > p(\beta - \delta_n)\}$. That is,

$$P(\lambda_{(1)} > px)$$

$$\sim \int_{\substack{\lambda_1 > \cdots > \lambda_n, px < \lambda_1 < p(x+\delta_n), \\ \lambda_2 < p(\beta+\delta_n), \lambda_n > p(\beta-\delta_n)}} n! f_{n,p,\beta}(\lambda_1, \ldots, \lambda_n) \, d\lambda_1 \cdots d\lambda_n$$

$$= \int_{\substack{\lambda_1 > \cdots > \lambda_n, px < \lambda_1 < p(x+\delta_n), \\ \lambda_2 < p(\beta+\delta_n), \lambda_n > p(\beta-\delta_n)}} n A_n \prod_{i=2}^{n} (\lambda_1 - \lambda_i)^\beta \cdot \lambda_1^{\frac{\beta(p-n+1)}{2}-1} \cdot e^{-\frac{1}{2}\lambda_1}$$

$$\times g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n) \, d\lambda_1 \cdots d\lambda_n$$

$$= \int_{\substack{\lambda_1 > \cdots > \lambda_n, px < \lambda_1 < p(x+\delta_n), \\ \lambda_2 < p(\beta+\delta_n), \lambda_n > p(\beta-\delta_n)}} n A_n (px - p\beta)^{\beta(n-1)}$$

$$\times \prod_{i=2}^{n} \left(1 + \frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_i - p\beta}{px - p\beta}\right)^\beta$$

$$\times \lambda_1^{\frac{\beta(p-n+1)}{2}-1} e^{-\frac{1}{2}\lambda_1} g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n) \, d\lambda_1 \cdots d\lambda_n.$$

Take z and a_2 in Lemma 1 such that

$$z = \frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_i - p\beta}{px - p\beta} \quad \text{and} \quad \alpha_2 = \frac{1}{2} - \frac{\delta_n}{x - \beta},$$

then $\alpha_2 < \frac{1}{2}$ as n is sufficiently large (we will have similar situations in the rest of the paper, the same interpretation "as n is sufficiently large" applies unless otherwise specified). Consequently, for $px < \lambda_1 < p(x + \delta_n)$ and $p(\beta - \delta_n) < \lambda_n \le \cdots \le \lambda_2 < p(\beta + \delta_n)$

$$\prod_{i=2}^{n} \left(1 + \frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_i - p\beta}{px - p\beta}\right)$$

$$= e^{\sum_{i=2}^{n} \log(1 + \frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_i - p\beta}{px - p\beta})}$$

(23)

$$\le e^{\sum_{i=2}^{n} (\frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_i - p\beta}{px - p\beta}) - \alpha_2 \sum_{i=2}^{n} (\frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_i - p\beta}{px - p\beta})^2}$$

$$\le e^{(n + o(n)) \frac{\lambda_1 - px}{px - p\beta} - \sum_{i=2}^{n} \frac{\lambda_i - p\beta}{px - p\beta} - \alpha_2 \sum_{i=2}^{n} (\frac{\lambda_i - p\beta}{px - p\beta})^2},$$

where in the last step we used $e^{-(n-1)\alpha_2 (\frac{\lambda_1 - px}{px - p\beta})^2} \le 1$ and $\alpha_2 \frac{\lambda_1 - px}{px - p\beta} \sum_{i=2}^{n} \frac{\lambda_i - p\beta}{px - p\beta} = o(n) \frac{\lambda_1 - px}{px - p\beta}$ since $\frac{\lambda_i - p\beta}{px - p\beta} \le O(\delta_n) = o(1)$ uniformly for all $2 \le i \le n$. Then we have the following upper bound:

$$P(\lambda_{(1)} > px)$$

$$\lesssim n A_n (px - p\beta)^{\beta(n-1)} \int_{px}^{p(x+\delta_n)} \lambda_1^{\frac{\beta(p-n+1)}{2}-1} e^{(\beta n + o(n)) \frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_1}{2}} \, d\lambda_1$$

$$\times \int_{\substack{\lambda_2 > \cdots > \lambda_n \\ \lambda_2 < p(\beta+\delta_n), \lambda_n > p(\beta-\delta_n)}} e^{-\beta \sum_{i=2}^n \frac{\lambda_i - p\beta}{p(x-\beta)} - \beta\alpha_2 \sum_{i=2}^n (\frac{\lambda_i - p\beta}{px - p\beta})^2}$$

(24)

$$\times g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n) \, d\lambda_2 \cdots d\lambda_n$$

$$\leq n A_n (px - p\beta)^{\beta(n-1)} \int_{px}^{p(x+\delta_n)} \lambda_1^{\frac{\beta(p-n+1)}{2}-1} e^{(\beta n + o(n))\frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_1}{2}} \, d\lambda_1$$

$$\times E\Big[e^{-\beta \sum_{i=2}^n \frac{(\lambda_i - \beta p)}{p(x-\beta)} - \beta\alpha_2 \sum_{i=2}^n \frac{(\lambda_i - \beta p)^2}{p^2(x-\beta)^2}} \Big].$$

Trivially, $\frac{n^2}{p} - \frac{(n-1)^2}{p-1} = o(1)$ since $n/p \to 0$. We then have from Lemma 3 that

(25) $$E\Big[e^{-\beta \sum_{i=2}^n \frac{(\lambda_i - \beta p)}{p(x-\beta)} - \beta\alpha_2 \sum_{i=2}^n \frac{(\lambda_i - \beta p)^2}{p^2(x-\beta)^2}} \Big] = e^{-\frac{\alpha_2 \beta^3}{(x-\beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]}.$$

This implies that

$$P(\lambda_{(1)} > px)$$

$$\lesssim n A_n (px - p\beta)^{\beta(n-1)} e^{-\frac{\alpha_2 \beta^3}{(x-\beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]}$$

$$\times \int_{px}^{p(x+\delta_n)} \lambda_1^{\frac{\beta(p-n+1)}{2}-1} e^{(\beta n + o(n))\frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_1}{2}} \, d\lambda_1$$

$$= n A_n (px - p\beta)^{\beta(n-1)} e^{-\frac{\alpha_2 \beta^3}{(x-\beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]}$$

(26)

$$\times \int_0^{p\delta_n} (\lambda_1 + px)^{\frac{\beta(p-n+1)}{2}-1} e^{(\beta n + o(n))\frac{\lambda_1}{px - p\beta} - \frac{\lambda_1 + px}{2}} \, d\lambda_1$$

$$\leq n A_n (px - p\beta)^{\beta(n-1)} e^{-\frac{px}{2} - \frac{\alpha_2 \beta^3}{(x-\beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]} (px)^{\frac{\beta(p-n+1)}{2}-1}$$

$$\times \int_0^{p\delta_n} e^{\{\frac{\beta(p-n+1)}{2}-1\}\frac{\lambda_1}{px} + (\beta n + o(n))\frac{\lambda_1}{px - p\beta} - \frac{\lambda_1}{2}} \, d\lambda_1$$

$$\sim \frac{2x}{x-\beta} n A_n (px - p\beta)^{\beta(n-1)} (px)^{\frac{\beta(p-n+1)}{2}-1} e^{-\frac{px}{2} - \frac{\alpha_2 \beta^3}{(x-\beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]},$$

where in the second step we changed the variable λ_1 to $\lambda_1 + px$; in the third step we used $(\lambda_1 + px) \leq (px) \exp\{\lambda_1/(px)\}$; the last step follows from the fact that $\{\frac{\beta(p-n+1)}{2} - 1\}\frac{\lambda_1}{px} + (\beta n + o(n))\frac{\lambda_1}{px - p\beta} - \frac{\lambda_1}{2} \sim \frac{\beta - x}{2x}\lambda_1$ by using the facts $p\delta_n \to \infty$ and $n/p \to 0$.

Finally, noticing that $\frac{\alpha_2 \beta^3}{(x-\beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}] - \frac{\beta^3 n^2}{2(x-\beta)^2 p} \to 0$ due to the fact $\frac{\delta_n n^2}{p} \to 0$ and $\frac{n^{5/3}}{p} \to 0$, we obtain (21).

Proof of (22). By the same argument as in the above derivation, take $\alpha_1 = 1/2 + \delta_n/(x - \beta)$ in Lemma 1 to have

$$\sum_{i=2}^n \log\left(1 + \frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_i - p\beta}{px - p\beta}\right)$$

$$\geq \sum_{i=2}^n \log\left(1 - \frac{\lambda_i - p\beta}{px - p\beta}\right)$$

$$\geq -\sum_{i=2}^n \frac{\lambda_i - p\beta}{px - p\beta} - \alpha_1 \sum_{i=2}^n \frac{(\lambda_i - p\beta)^2}{(px - p\beta)^2}$$

under the restriction $px < \lambda_1 < p(x + \delta_n)$, $\lambda_2 < p(\beta + \delta_n)$, and $\lambda_n > p(\beta - \delta_n)$. Therefore,

$$P(\lambda_{(1)} > px) \geq \int_{\substack{\lambda_1 > \cdots > \lambda_n, \lambda_1 > px, \\ \lambda_2 < p(\beta + \delta_n), \lambda_n > p(\beta - \delta_n)}} n! f_{n,p,\beta}(\lambda_1, \ldots, \lambda_n) \, d\lambda_1 \cdots d\lambda_n$$

$$\geq nA_n(px - p\beta)^{\beta(n-1)}$$

$$\times \int_{\substack{\lambda_2 > \cdots > \lambda_n \\ \lambda_2 < p(\beta + \delta_n), \lambda_n > p(\beta - \delta_n)}} \int_{px}^{p(x + \delta_n)} e^{-\beta \sum_{i=2}^n \frac{\lambda_i - p\beta}{p(x - \beta)} - \beta\alpha_1 \sum_{i=2}^n \frac{(\lambda_i - p\beta)^2}{p^2(x - \beta)^2}}$$

(27)

$$\times \lambda_1^{\frac{\beta(p - n + 1)}{2} - 1} e^{-\frac{\lambda_1}{2}} \times g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n) \, d\lambda_1 \, d\lambda_2 \cdots d\lambda_n$$

$$= nA_n(px - p\beta)^{\beta(n-1)} \times \int_{px}^{p(x + \delta_n)} \lambda_1^{\frac{\beta(p - n + 1)}{2} - 1} e^{-\frac{\lambda_1}{2}} \, d\lambda_1$$

$$\times \int_{\substack{\lambda_2 > \cdots > \lambda_n, \\ \lambda_2 < p(\beta + \delta_n), \lambda_n > p(\beta - \delta_n)}} e^{-\beta \sum_{i=2}^n \frac{\lambda_i - p\beta}{p(x - \beta)} - \beta\alpha_1 \sum_{i=2}^n \frac{(\lambda_i - p\beta)^2}{p^2(x - \beta)^2}}$$

$$\times g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n) \, d\lambda_2 \cdots d\lambda_n.$$

By Lemmas 3 and 8, we have the second integral in (27) is $e^{-\frac{\alpha_1 \beta^3}{(x - \beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]}$ since $\frac{\delta_n n^2}{p} \to 0$ and $p\delta_n \to \infty$. It follows that

$$P(\lambda_{(1)} > px)$$

(28) $$\gtrsim nA_n(px - p\beta)^{\beta(n-1)} e^{-\frac{\alpha_1 \beta^3}{(x - \beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]} \int_{px}^{p(x + \delta_n)} \lambda_1^{\frac{\beta(p - n + 1)}{2} - 1} e^{-\frac{\lambda_1}{2}} \, d\lambda_1$$

$$\sim \frac{2x}{x - \beta} nA_n(px - p\beta)^{\beta(n-1)}(px)^{\frac{\beta(p - n + 1)}{2} - 1} e^{-\frac{px}{2} - \frac{\alpha_1 \beta^3}{(x - \beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]},$$

where the last step follows the same argument as in (26) due to the fact $p\delta_n \to \infty$. This yields (22) by noticing that $\frac{\alpha_1 \beta^3}{(x - \beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}] - \frac{\beta^3 n^2}{2(x - \beta)^2 p} \to 0$.

Next, we prove the result (8). By the above derivations,

$$\log P(\lambda_{(1)} > px)$$

(29)
$$= \log n + \log A_n + \log \frac{2x}{x - \beta} + \beta(n - 1)\log(px - p\beta)$$

$$+ \left[\frac{\beta(p - n + 1)}{2} - 1\right]\log(px) - \frac{px}{2} - \frac{\beta^3 n^2}{2(x - \beta)^2 p} + o(1).$$

From the Stirling formula,

$$\log \Gamma(z) = z \log z - z - (\log z)/2 + \log \sqrt{2\pi} + o(1)$$

for large $|z|$; we know

$$\log A_n = \log \frac{2^{-\frac{\beta(n+p-1)}{2}}\Gamma(1 + \frac{\beta}{2})}{\Gamma(1 + \frac{\beta n}{2})\Gamma(\frac{\beta p}{2})} = \log \frac{2^{-\frac{\beta(n+p-1)}{2}}\Gamma(1 + \frac{\beta}{2})}{\frac{\beta n}{2}\Gamma(\frac{\beta n}{2})\Gamma(\frac{\beta p}{2})}$$

$$= -\frac{\beta}{2}(n + p - 1)\log 2 + \log \Gamma\left(1 + \frac{\beta}{2}\right) - \frac{\beta n}{2}\log \frac{\beta n}{2} - \frac{\beta p}{2}\log \frac{\beta p}{2}$$

$$+ \frac{\beta n}{2} + \frac{\beta p}{2} + \frac{1}{2}\log \frac{\beta p}{2} - \frac{1}{2}\log \frac{\beta n}{2} - 2\log \sqrt{2\pi} + o(1)$$

(30)
$$= -\frac{\beta p}{2}\log p - \frac{\beta p}{2}(\log \beta - 1) - \frac{\beta n}{2}\log n - \frac{\beta n}{2}(\log \beta - 1)$$

$$+ \frac{1}{2}\log \frac{\beta p}{2} - \frac{1}{2}\log \frac{\beta n}{2} + \frac{\beta}{2}\log 2 - 2\log \sqrt{2\pi} + \log \Gamma\left(1 + \frac{\beta}{2}\right)$$

$$+ o(1).$$

Therefore, plugging in the above expansion of $\log A_n$ into equation (29), we obtain

$$\log P(\lambda_{(1)} > px)$$

$$= \log n - \frac{\beta p}{2}\log p - \frac{\beta p}{2}(\log \beta - 1) - \frac{\beta n}{2}\log n - \frac{\beta n}{2}(\log \beta - 1)$$

$$+ \frac{1}{2}\log \frac{\beta p}{2} - \frac{1}{2}\log \frac{\beta n}{2} + \frac{\beta}{2}\log 2 - 2\log \sqrt{2\pi} + \log \Gamma\left(1 + \frac{\beta}{2}\right)$$

$$+ \log \frac{2x}{x - \beta} + \beta(n - 1)\log(px - p\beta)$$

$$+ \left[\frac{\beta(p - n + 1)}{2} - 1\right]\log(px) - \frac{px}{2} - \frac{\beta^3 n^2}{2(x - \beta)^2 p} + o(1)$$

$$= p\left(\frac{\beta}{2} - \frac{\beta}{2}\log \beta - \frac{x}{2} + \frac{\beta}{2}\log x\right) + \frac{\beta n}{2}\log \frac{p}{n} - \frac{\beta + 1}{2}\log p$$

$$+ \beta n \left(-\frac{\log x}{2} + \log(x - \beta) - \frac{\log \beta}{2} + \frac{1}{2} \right) + \frac{1}{2} \log n - \frac{\beta^3 n^2}{2(x - \beta)^2 p}$$

$$- (\beta + 1) \log(x - \beta) + \frac{\beta}{2} \log(2x) - \log(\pi) + \log \Gamma \left(1 + \frac{\beta}{2} \right) + o(1).$$

This completes the proof of (8).

Next, we prove the result (9) when $p/n \to \infty$. For any $l_n > 0$ such that $l_n \to \infty$ and $\sqrt{np^{-1}l_n} \to 0$, take $\delta_n = \sqrt{np^{-1}l_n}$. Then $\delta_n \to 0$ and $\frac{\delta_n^2}{np^{-1}} \to \infty$. Reviewing the proof of the asymptotic upper bound (26) and the lower bound (28), we only use the three conditions: $\delta_n \to 0$, $\frac{\delta_n^2}{np^{-1}} \to \infty$ and $p/n \to \infty$. Consequently,

$$B_{n,p,\beta}(x) + \left(\frac{1}{2} - \alpha_1 \right) \frac{\beta^3 n^2}{(x - \beta)^2 p} + \alpha_1 \frac{\beta^3 O(1) n^3}{3(x - \beta)^2 p^2}$$

$$\leq \log P(\lambda_{(1)} > px)$$

$$\leq B_{n,p,\beta}(x) + \left(\frac{1}{2} - \alpha_2 \right) \frac{\beta^3 n^2}{(x - \beta)^2 p} + \alpha_2 \frac{\beta^3 O(1) n^3}{3(x - \beta)^2 p^2},$$

where $B_{n,p,\beta}(x)$ is defined as in Theorem 1, $\alpha_1 = \frac{1}{2} + \frac{\delta_n}{x - \beta}$ and $\alpha_2 = \frac{1}{2} - \frac{\delta_n}{x - \beta}$. Replace δ_n with $\sqrt{np^{-1}l_n}$ and we have

$$B_{n,p,\beta}(x) - \Theta(1) \frac{l_n n^{5/2}}{p^{3/2}} \leq \log P(\lambda_{(1)} > px) \leq B_{n,p,\beta}(x) + \Theta(1) \frac{l_n n^{5/2}}{p^{3/2}},$$

or equivalently,

$$\left| \log P(\lambda_{(1)} > px) - B_{n,p,\beta}(x) \right| \cdot \frac{p^{3/2}}{n^{5/2}} \leq \Theta(l_n)$$

for any l_n satisfying $l_n \to \infty$ and $l_n = o(\frac{p}{n})$. Observe that the left-hand side of the above does not depend on l_n, we conclude

$$\limsup_{n \to \infty} \left| \log P(\lambda_{(1)} > px) - B_{n,p,\beta}(x) \right| \cdot \frac{p^{3/2}}{n^{5/2}} < \infty$$

by using a trivial argument of contradiction. This completes the proof. $\square$

PROOF OF THEOREM 2. First consider (i). Since $p/n^{5/3} \to \infty$, we are able to pick $\delta_n > 0$ satisfying $\delta_n \to 0$, $\delta_n^2 n^{-1} p \to \infty$ and $\delta_n n^2/p \to 0$. To show $E^Q[L_p^2] = E^Q[L_p^2; \lambda_{(1)} > px] \sim P(\lambda_{(1)} > px)^2$, by Lemma 9 in the Supplementary Material [Jiang, Leder and Xu (2017)], it suffices to show that

$$(31) \quad \begin{aligned} E^Q\big[L_p^2; \lambda_{(1)} > px, p(\beta + \delta_n) > \lambda_{(2)} > \cdots > \lambda_{(n)} > p(\beta - \delta_n)\big] \\ \sim P(\lambda_{(1)} > px)^2. \end{aligned}$$

Following (11) and (14),

and the left-hand side of (31),

$$
= E^Q\left[\left(\frac{nA_n \times \prod_{i=2}^n (\lambda_1 - \lambda_i)^\beta \cdot \lambda_1^{\frac{\beta(p-n+1)}{2}-1} \cdot e^{-\frac{1}{2}\lambda_1}}{\frac{x-\beta}{2x}e^{-\frac{x-\beta}{2x}(\lambda_1 - px)} \cdot I_{(\lambda_1 > px)}}\right)^2 ;\right.
$$

$$
\left. \lambda_1 > px,\, p(\beta + \delta_n) > \lambda_2 > \cdots > \lambda_n > p(\beta - \delta_n) \right]
$$

$$
= 2x(x-\beta)^{-1}n^2 A_n^2 e^{-\frac{x-\beta}{2x}px}
$$

$$
\times \int_{\substack{\lambda_1 > px, \\ p(\beta+\delta_n) > \lambda_2 > \cdots > \lambda_n > p(\beta-\delta_n)}} \prod_{i=2}^n (\lambda_1 - \lambda_i)^{2\beta} \cdot \lambda_1^{\beta(p-n+1)-2} \cdot e^{-(1-\frac{x-\beta}{2x})\lambda_1}\, d\lambda_1
$$

(32)

$$
\times\, g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n)\, d\lambda_2 \cdots d\lambda_n
$$

$$
= 2x(x-\beta)^{-1}n^2 A_n^2 e^{-\frac{x-\beta}{2x}px}(px - p\beta)^{2\beta(n-1)}
$$

$$
\times \int_{\substack{\lambda_2 > \cdots > \lambda_n \\ \lambda_2 < p(\beta+\delta_n),\, \lambda_n > p(\beta-\delta_n)}} \int_{px}^\infty \lambda_1^{\beta(p-n+1)-2} \cdot e^{-(1-\frac{x-\beta}{2x})\lambda_1}\, d\lambda_1
$$

$$
\times \prod_{i=2}^n \left(1 + \frac{\lambda_1 - px}{px - p\beta} - \frac{\lambda_i - p\beta}{px - p\beta}\right)^{2\beta}
$$

$$
\times\, g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n)\, d\lambda_2 \cdots d\lambda_n.
$$

Using the upper bound as in (23) and part of the arguments in (24) and (25), we have

Display (32)

$$
\lesssim 2x(x-\beta)^{-1}n^2 A_n^2 e^{-\frac{x-\beta}{2x}px}(px - p\beta)^{2\beta(n-1)}
$$

$$
\times \int_{px}^\infty e^{(2\beta n + o(n))\frac{\lambda_1 - px}{px - p\beta} - 2\beta(n-1)\alpha_2 (\frac{\lambda_1 - px}{px - p\beta})^2} \cdot \lambda_1^{\beta(p-n+1)-2}
$$

$$
\times\, e^{-(1-\frac{x-\beta}{2x})\lambda_1}\, d\lambda_1
$$

$$
\times \int_{\substack{\lambda_2 > \cdots > \lambda_n \\ \lambda_2 < p(\beta+\delta_n),\, \lambda_n > p(\beta-\delta_n)}} e^{-2\beta \sum_{i=2}^n \frac{\lambda_i - p\beta}{p(x-\beta)} - 2\beta\alpha_2 \sum_{i=2}^n (\frac{\lambda_i - p\beta}{px - p\beta})^2}
$$

$$
\times\, g_{n-1,p-1,\beta}(\lambda_2, \ldots, \lambda_n)\, d\lambda_2 \cdots d\lambda_n
$$

$$
\sim 2x(x-\beta)^{-1}n^2 A_n^2 e^{-\frac{x-\beta}{2x}px - 2\alpha_2 \frac{\beta^3}{(x-\beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]}(px - p\beta)^{2\beta(n-1)}
$$

(33)

$$
\times \int_{px}^\infty e^{(2\beta n + o(n))\frac{\lambda_1 - px}{px - p\beta} - 2\beta(n-1)\alpha_2 (\frac{\lambda_1 - px}{px - p\beta})^2} \cdot \lambda_1^{\beta(p-n+1)-2}
$$

$$\times\, e^{-(1-\frac{x-\beta}{2x})\lambda_1}\, d\lambda_1$$

$$= 2x(x-\beta)^{-1}n^2 A_n^2 e^{-\frac{x-\beta}{2x}px - 2\alpha_2 \frac{\beta^3}{(x-\beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]}(px-p\beta)^{2\beta(n-1)}$$

$$\times \int_0^\infty e^{\frac{(2\beta n + o(n))\lambda_1}{px - p\beta} - 2\beta(n-1)\alpha_2(\frac{\lambda_1}{px-p\beta})^2 - (1-\frac{x-\beta}{2x})(\lambda_1 + px)}$$

$$\times (\lambda_1 + px)^{\beta(p-n+1)-2}\, d\lambda_1$$

$$\lesssim (2x)^2(x-\beta)^{-2}n^2 A_n^2 e^{-px - 2\alpha_2 \frac{\beta^3}{(x-\beta)^2}[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]}(px-p\beta)^{2\beta(n-1)}$$

$$\times (px)^{\beta(p-n+1)-2},$$

where $\alpha_2 = \frac{1}{2} - \frac{\delta_n}{x-\beta}$; in the third step we changed variable λ_1 to $\lambda_1 + px$ for the integral; the last step follows from the inequality that

$$\int_0^\infty e^{\frac{(2\beta n + o(n))\lambda_1}{px - p\beta} - 2\beta(n-1)\alpha_2(\frac{\lambda_1}{px-p\beta})^2 - (1-\frac{x-\beta}{2x})(\lambda_1 + px)} \cdot (\lambda_1 + px)^{\beta(p-n+1)-2}\, d\lambda_1$$

$$\leq (px)^{\beta(p-n+1)-2}e^{-(1-\frac{x-\beta}{2x})px}$$

$$\times \int_0^\infty e^{\frac{(2\beta n + o(n))\lambda_1}{px - p\beta} - (1-\frac{x-\beta}{2x})\lambda_1}e^{[\beta(p-n+1)-2]\frac{\lambda_1}{px}}\, d\lambda_1$$

$$\sim (px)^{\beta(p-n+1)-2}e^{-(1-\frac{x-\beta}{2x})px}\frac{2x}{x-\beta},$$

where in the first step we used $(\lambda + px)^{\beta(p-n+1)-2} \leq (px)^{\beta(p-n+1)-2}\exp(\lambda_1/(px))$ and the second step we used $\frac{(2\beta n + o(n))\lambda_1}{px - p\beta} - (1-\frac{x-\beta}{2x})\lambda_1 + [\beta(p-n+1) - 2]\frac{\lambda_1}{px} \sim -\frac{x-\beta}{2x}\lambda_1$. Easily, $\alpha_2[\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}] - \frac{n^2}{2p} \to 0$. Based on (33) and Theorem 1, we know that

$$\frac{E^Q[(\frac{dP}{dQ})^2; \lambda_{(1)} > px]}{P(\lambda_{(1)} > px)^2} \lesssim 1$$

provided $p/n^{5/3} \to \infty$. Since

$$(34)\quad E^Q\left[\left(\frac{dP}{dQ}\right)^2; \lambda_{(1)} > px\right] \geq \left\{E^Q\left[\left(\frac{dP}{dQ}\right); \lambda_{(1)} > px\right]\right\}^2 = P(\lambda_{(1)} > px)^2$$

by Hölder's inequality, we have

$$(35)\qquad \frac{E^Q[(\frac{dP}{dQ})^2; \lambda_{(1)} > px]}{P(\lambda_{(1)} > px)^2} \sim 1.$$

So the second statement of the theorem is obtained.

Now we prove the first one. Recall $P^*_{px} = P(\cdot \mid \lambda_{(1)} > px)$ as described in the paragraph following (7). It is easy to check that

$$\frac{dP^*_{px}}{dQ} = \frac{dP}{dQ} \cdot \frac{I(\lambda_{(1)} > px)}{P(\lambda_{(1)} > px)}$$

a.s. with respect to Q defined on $\mathcal{B}([0, \infty)^n)$. For any $A \subset \mathcal{B}([0, \infty)^n)$,

$$\begin{aligned}
\left| Q(A) - P^*_{px}(A) \right| &= \left| \int_A \left(\frac{dP^*_{px}}{dQ} - 1 \right) dQ \right| \\
&\leq \left\{ E^Q \left(\frac{dP}{dQ} \cdot \frac{I(\lambda_{(1)} > px)}{P(\lambda_{(1)} > px)} - 1 \right)^2 \right\}^{1/2} \\
&= \left(\frac{E^Q[(\frac{dP}{dQ})^2; \lambda_{(1)} > px]}{P(\lambda_{(1)} > px)^2} - 1 \right)^{1/2} \to 0
\end{aligned}$$

by Hölder's inequality and (35). Thus,

$$\lim_{n \to \infty} \sup_{A \in \mathcal{F}} \left| Q(A) - P^*_{px}(A) \right| = 0$$

as $p/n^{5/3} \to \infty$. This gives the first conclusion of part (i).

We next prove the conclusion in (ii) and (iii). For general $p/n \to \infty$, by Lemma 9, we have

$$\begin{aligned}
&E^Q[L_p^2; \lambda_{(1)} > px] \\
&\quad \sim E^Q[L_p^2; \lambda_{(1)} > px, p(\beta + \delta_n) > \lambda_{(2)} > \cdots > \lambda_{(n)} > p(\beta - \delta_n)] \\
&\quad\quad + o(1) P(\lambda_{(1)} > px)^2.
\end{aligned}$$

Note that $P(\lambda_{(1)} > px)^2 \leq E^Q[L_p^2; \lambda_{(1)} > px]$. Then we have

$$\begin{aligned}
&E^Q[L_p^2; \lambda_{(1)} > px] \\
&\quad \sim E^Q[L_p^2; \lambda_{(1)} > px, p(\beta + \delta_n) > \lambda_{(2)} > \cdots > \lambda_{(n)} > p(\beta - \delta_n)].
\end{aligned}$$

Then following exactly the same argument as in (32) and (33), which requires the assumption that $p/n \to \infty$ only, we have

$$\begin{aligned}
&E^Q[L_p^2; \lambda_{(1)} > px] \\
&\quad \lesssim (2x)^2 (x - \beta)^{-2} n^2 A_n^2 e^{-px - 2\alpha_2 \frac{\beta^3}{(x-\beta)^2} [\frac{n^2}{p} - \frac{O(1)n^3}{3p^2}]} \\
&\quad\quad \times (px - p\beta)^{2\beta(n-1)} (px)^{\beta(p-n+1)-2}.
\end{aligned}$$

Similar to the proof of Theorem 1, this implies that

$$\begin{equation}
E^Q[L_p^2; \lambda_{(1)} > px] \leq \exp\left\{ 2B_{n,p,\beta}(x) + O(1) \frac{n^{5/2}}{p^{3/2}} \right\}. \tag{36}
\end{equation}$$

Recall that $L_p = \frac{dP}{dQ} 1\{\lambda_{(1)} > px\}$. Then the ratio

$$E^Q[L_p^2; \lambda_{(1)} > px] / P(\lambda_{(1)} > px)^2 = O(1)$$

provided $n^{5/3}/p = O(1)$; (36) together with (34) further imply that

$$\lim_{n \to \infty} \frac{\log E^Q[(\frac{dP}{dQ})^2; \lambda_{(1)} > px]}{2 \log P(\lambda_{(1)} > px)} = 1$$

as $p/n \to \infty$. The proof is complete. $\square$

Acknowledgments. The authors thank the Editor, an Associate Editor and an anonymous reviewer for many helpful and constructive comments.

SUPPLEMENTARY MATERIAL

Supplement to "Rare-event analysis for extremal eigenvalues of white Wishart matrices." (DOI: 10.1214/16-AOS1502SUPP; .pdf). The online Supplementary Material contains proofs of technical lemmas (Lemmas 1–9) and Theorem 3.

REFERENCES

ADLER, R. J., BLANCHET, J. H. and LIU, J. (2012). Efficient Monte Carlo for high excursions of Gaussian random fields. *Ann. Appl. Probab.* **22** 1167–1214. MR2977989

ANDERSON, G. W., GUIONNET, A. and ZEITOUNI, O. (2010). *An Introduction to Random Matrices.* Cambridge Univ. Press, Cambridge. MR2760897

ASMUSSEN, S. and GLYNN, P. W. (2007). *Stochastic Simulation*: *Algorithms and Analysis.* Springer, New York. MR2331321

ASMUSSEN, S. and KROESE, D. P. (2006). Improved algorithms for rare event simulation with heavy tails. *Adv. in Appl. Probab.* **38** 545–558. MR2264957

BIANCHI, P., DEBBAH, M., MAIDA, M. and NAJIM, J. (2011). Performance of statistical tests for single-source detection using random matrix theory. *IEEE Trans. Inform. Theory* **57** 2400–2419. MR2809098

BLANCHET, J. and GLYNN, P. (2008). Efficient rare-event simulation for the maximum of heavy-tailed random walks. *Ann. Appl. Probab.* **18** 1351–1378. MR2434174

BLANCHET, J., GLYNN, P. and LEDER, K. (2012). On Lyapunov inequalities and subsolutions for efficient importance sampling. *ACM Trans. Model. Comput. Simul.* **22** Art. 13, 27. MR2980826

BLANCHET, J. H. and LIU, J. (2008). State-dependent importance sampling for regularly varying random walks. *Adv. in Appl. Probab.* **40** 1104–1128. MR2488534

BORDENAVE, C. and CAPUTO, P. (2014). A large deviation principle for Wigner matrices without Gaussian tails. *Ann. Probab.* **42** 2454–2496. MR3265172

DUMITRIU, I. (2003). Eigenvalue Statistics for Beta-Ensembles. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

DUMITRIU, I. and EDELMAN, A. (2002). Matrix models for beta ensembles. *J. Math. Phys.* **43** 5830–5847. MR1936554

DUPUIS, P., LEDER, K. and WANG, H. (2007). Importance sampling for sums of random variables with regularly varying tails. *ACM Trans*. *Model*. *Comput*. *Simul*. **17** 1–14.

EDELMAN, A. and RAO, N. R. (2005). Random matrix theory. *Acta Numer*. **14** 233–297. MR2168344

EL KAROUI, N. (2003). On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n tend to infinity. Preprint. Available at arXiv:Math/0309355.

HIAI, F. and PETZ, D. (1998). Eigenvalue density of the Wishart matrix and large deviations. *Infin*. *Dimens*. *Anal*. *Quantum Probab*. *Relat*. *Top*. **1** 633–646. MR1665279

JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann*. *Math*. *Stat*. **35** 475–501. MR0181057

JIANG, T., LEDER, K. and XU, G. (2017). Supplement to "Rare-event analysis for extremal eigenvalues of white Wishart matrices." DOI:10.1214/16-AOS1502SUPP.

JIANG, T. and LI, D. (2014). Approximation of rectangular beta-Laguerre ensembles and large deviations. *J*. *Theoret*. *Probab*. To appear.

JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Comm*. *Math*. *Phys*. **209** 437–476. MR1737991

JOHNSTONE, I. M. (2001). On the distribution of the largest eigenvalue in principal components analysis. *Ann*. *Statist*. **29** 295–327. MR1863961

JOHNSTONE, I. M. and MA, Z. (2012). Fast approach to the Tracy–Widom law at the edge of GOE and GUE. *Ann*. *Appl*. *Probab*. **22** 1962–1988. MR3025686

JOHNSTONE, I. M., MA, Z., PERRY, P. O. and SHAHRAM, M. (2010). RMTstat: Distributions, statistics and tests derived from random matrix theory. Unpublished manuscript.

KWAPIEŃ, J., DROŻDŻ, S. and SPETH, J. (2003). Alternation of different fluctuation regimes in the stock market dynamics. *Phys*. *A* **330** 605–621. MR2044843

KWISSA, M., NAKAYA, H. I., ONLAMOON, N., WRAMMERT, J., VILLINGER, F., PERNG, G. C., YOKSAN, S., PATTANAPANYASAT, K., CHOKEPHAIBULKIT, K., AHMED, R. and PULENDRAN, B. (2014). Dengue virus infection induces expansion of a CD14+ CD16+ monocyte population that stimulates plasmablast differentiation. *Cell Host & Microbe* **16** 115–127.

LIU, J. and XU, G. (2014a). On the conditional distributions and the efficient simulations of exponential integrals of Gaussian random fields. *Ann*. *Appl*. *Probab*. **24** 1691–1738. MR3211008

LIU, J. and XU, G. (2014b). Efficient simulations for the exponential integrals of Hölder continuous Gaussian random fields. *ACM Trans*. *Model*. *Comput*. *Simul*. **24** Art. 9, 24. MR3195228

MA, Z. (2012). Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli* **18** 322–359. MR2888709

MACDONALD, I. G. (1995). *Symmetric Functions and Hall Polynomials*, 2nd ed. Oxford Univ. Press, New York. MR1354144

MAÏDA, M. (2007). Large deviations for the largest eigenvalue of rank one deformations of Gaussian ensembles. *Electron*. *J*. *Probab*. **12** 1131–1150 (electronic). MR2336602

MARČENKO, V. A. and PASTUR, L. A. (1967). Distribution of eigenvalues for some sets of random matrices. *Sb*. *Math*. **1** 457–483.

MUIRHEAD, R. J. (1982). *Aspects of Multivariate Statistical Theory*. Wiley, New York. MR0652932

PATTERSON, N., PRICE, A. L. and REICH, D. (2006). Population structure and eigenanalysis. *PLoS Genet*. **2** e190.

RAMÍREZ, J. A., RIDER, B. and VIRÁG, B. (2011). Beta ensembles, stochastic Airy spectrum, and a diffusion. *J*. *Amer*. *Math*. *Soc*. **24** 919–944. MR2813333

ROY, S. N. (1953). On a heuristic method of test construction and its use in multivariate analysis. *Ann*. *Math*. *Stat*. **24** 220–238. MR0057519

SIEGMUND, D. (1976). Importance sampling in the Monte Carlo study of sequential tests. *Ann*. *Statist*. **4** 673–684. MR0418369

Xu, G., Lin, G. and Liu, J. (2014). Rare-event simulation for the stochastic Korteweg–de Vries equation. *SIAM/ASA J. Uncertain. Quantificat.* **2** 698–716. MR3280104

T. Jiang
School of Statistics
University of Minnesota
224 Church Street SE
Minneapolis, Minnesota 55455
USA
E-mail: jiang040@umn.edu

K. Leder
Industrial and Systems Engineering
University of Minnesota
111 Church Street SE
Minneapolis, Minnesota 55455
USA
E-mail: lede0024@umn.edu

G. Xu
Department of Statistics
University of Michigan
456 West Hall
1085 South University
Ann Arbor, Michigan 48109
E-mail: gongjun@umich.edu