# Estimating Tail Probabilities of the Ratio of the Largest Eigenvalue to the Trace of a Wishart Matrix

Yinqiu He[a], Gongjun Xu[a,*]

[a]*Department of Statistics, University of Michigan, Ann Arbor, MI, United States*

**Abstract**

This paper develops an efficient Monte Carlo method to estimate the tail probabilities of the ratio of the largest eigenvalue to the trace of the Wishart matrix, which plays an important role in multivariate data analysis. The estimator is constructed based on a change-of-measure technique and it is proved to be asymptotically efficient for both the real and complex Wishart matrices. Simulation studies further show the outperformance of the proposed method over existing approaches based on asymptotic approximations, especially when estimating probabilities of rare events.

*Keywords:* Ratio of largest eigenvalue to trace; Rare events; Wishart matrices; Tracy-Widom distribution

## 1. Introduction

Consider $n$ independent and identically distributed (i.i.d.) $p$-dimensional observations $\mathbf{x}_1, ..., \mathbf{x}_n$ from a real or complex valued Gaussian distribution with mean zero and covariance matrix $\Sigma = \sigma^2 I_p$. Here $\sigma^2$ is an unknown scaling factor and $I_p$ is the $p \times p$ identity matrix. Define the $n \times p$ data matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n]^\top$, and assume $\lambda_1 \geq \dots \geq \lambda_p$ are the ordered real eigenvalues of the sample covariance matrix $\hat{\Sigma} = n^{-1}\mathbf{X}^H\mathbf{X}$, where "$H$" denotes the conjugate transpose. Note that if $p > n$, the last $p - n$ of the $\lambda$'s are zero. Let $U_n$ be the ratio of the largest eigenvalue to the trace:

$$U_n := \frac{\lambda_1}{\min\{n, p\}^{-1} \sum_{i=1}^{p} \lambda_i}. \tag{1}$$

We are interested in estimating the following rare-event tail probability of $U_n$:

$$\alpha_n(x) := \Pr(U_n > x),$$

where $x$ is some constant such that $\alpha_n(x)$ is small. Estimating rare-event tail probabilities is often of interest in multivariate data analysis. For instance, in multiple testing problems, it is often needed to evaluate very small $p$-values for individual test statistics to control the overall false-positive error rate.

The random variable $U_n$ plays an important role in multivariate statistics on testing the covariance structure and possible mean singles. For instance, it has been popularly used to test for equality of the population covariance to a scaled identity matrix, i.e., the sphericity test [e.g., 22]: $H_0 : \Sigma = \sigma^2 I_p$ v.s. $H_1 : \Sigma \neq \sigma^2 I_p$ with $\sigma^2$ unknown. The test statistic $U_n$ does not depend on the unknown variance parameter $\sigma^2$ and has high detection power against alternative covariance matrix with a low rank perturbation of the null $\sigma^2 I_p$. In particular, under the alternative of rank-1 perturbation with $\Sigma = hh' + \sigma^2 I_p$ for some unknown $h \in \mathbb{R}^p$ and $\sigma^2$, the likelihood ratio test statistic $L_n = \sup_{h,\sigma^2} f_1(X; h, \sigma^2)/\sup_{\sigma^2} f_0(X; \sigma^2)$ can be written as a monotone function of $U_n$ and therefore $\alpha_n(x)$ corresponds to the $p$-value [e.g., 4, 22]. Please refer to [17, 22, 24] for more discussions and many other applications.

*Corresponding author

*Email addresses:* yqhe@umich.edu (Yinqiu He), gongjun@umich.edu (Gongjun Xu)

Despite its importance in multivariate analysis, the exact distribution of $U_n$ may be difficult to compute, especially when estimating the rare-event tail probabilities. Note that $\sigma^{-2}n^{-1}\mathbf{X}^H\mathbf{X}$ follows a Wishart distribution $\mathcal{W}_{\beta,p}(n, n^{-1}I_p)$ ($\beta = 1$ for real Gaussian and $\beta = 2$ for complex Gaussian), then the distribution of $U_n$ corresponds to that of the ratio of largest eigenvalue and the trace of a $\mathcal{W}_{\beta,p}(n, n^{-1}I_p)$. However, such distribution is nonstandard and the exact distribution formulas for $U_n$ involves high dimensional integrals or as inverses of certain Laplace transforms. Numerical evaluation has been studied in [6, 7, 16, 18, 25, 28]. But for high-dimensional data with large $p$, the computation becomes more challenging, which is particularly the case when $\alpha_n(x)$ is small due to the additional computational cost to control the relative estimation error of $\alpha_n(x)$.

The asymptotic distribution of $U_n$ with $p$ and $n$ both going to infinity has also been studied in the literature. It is known that $U_n$ asymptotically behaves similarly to the largest eigenvalue $\lambda_1$, whose limiting distribution has been studied in [13] and [14], and $U_n$ also asymptotically follows the Tracy-Widom distribution [e.g., 4, 23]. That is,

$$\Pr\left(\frac{U_n - \mu_{n,p}}{\sigma_{n,p}} > x\right) \to 1 - \mathcal{TW}_\beta(x), \tag{2}$$

where $\mathcal{TW}_\beta$ denotes the Tracy-Widom distribution of order $\beta$, with $\beta \in \{1, 2\}$ for real and complex valued observations respectively. In particular, for real valued observations,

$$\begin{aligned} \mu_{n,p} &= \frac{1}{n}\Big(\sqrt{n-\frac{1}{2}} + \sqrt{p-\frac{1}{2}}\Big)^2, \\ \sigma_{n,p} &= \frac{1}{n}\Big(\sqrt{n-\frac{1}{2}} + \sqrt{p-\frac{1}{2}}\Big)\Big(\frac{1}{\sqrt{n-1/2}} + \frac{1}{\sqrt{p-1/2}}\Big)^{1/3}. \end{aligned} \tag{3}$$

The convergence rate is shown to be of the order $O(\min\{n, p\}^{-2/3})$ [21]. For complex case, similar expressions can be found in [15]. [23] studied the accuracy of the Tracy-Widom approximation for finite values of $n$ and $p$, and found that the approximation may be inaccurate for small and even moderate values of $p$ when $n$ is large. A correction term was therefore proposed by [23] to improve the approximation result, which is derived using the Fredholm determinant representation, and the approximation rate is shown to be $o(\min\{n, p\}^{-2/3})$ when $X$ follows *complex* Gaussian. For *real* Gaussian case, which is of the interest in many statistics applications, [23] conjectured the result also holds. The calculation of correction term in [23] depends on the second derivative of the non-standard Tracy-Widom distribution, which usually involves numerical discretization scheme.

Another limitation of the existing methods is that they may become less efficient when estimating small tail probabilities of rare events. This paper aims to address such rare-event estimation problem. In particular, we propose an efficient Monte Carlo method to estimate the exact tail probability of $U_n$ by utilizing the importance sampling technique. Importance sampling is a commonly used tool to reduce the Monte Carlo variance and it has been used to estimated small tail probabilities, especially when the event is rare, in a wide variety of stochastic systems with both light-tailed and heavy-tailed distributions [e.g., 2, 3, 5, 11, 19, 20, 26, 29].

An importance sampling algorithm needs to construct an alternative sampling measure (a change of measure) under which the eigenvalues are sampled. Note that it is necessary to normalize our estimator with a Radon-Nikodym derivative to ensure an unbiased estimator. 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; and one common heuristic is to utilize a sampling measure that approximates the conditional distribution of $U_n$ given the event $\{U_n > x\}$. This paper proposes a change of measure $Q$ that asymptotically approximate the conditional measure $\Pr(\cdot \mid U_n > x)$. We establish rigorous analysis of proposed estimator for $U_n$ and show it is *asymptotically efficient*. Simulation studies show that the proposed method outperforms existing approximation approaches, especially when estimating probabilities of rare events.

The remainder of the paper is organized as follows. In Section 2, we propose the importance sampling estimator and provide the main result on the estimator's asymptotic efficiency in Theorem 1. Numerical results are presented in Section 3 to illustrate its performance. We discuss the possibility of generalizing the result to the ratio of the sum of the largest $k$ eigenvalues to the trace of a Wishart matrix in Section 4. The proof for Theorem 1 is given in Section 5.

## 2. Importance sampling estimation

For ease of discussion, we consider the setting of $p \leq n$, $p \to \infty$ and $n \to \infty$. When $p > n$, the algorithm and theory are essentially the same up to switching labels of $p$ and $n$, which is explained in Remark 4. We use the notation $\beta$ to denote the real Wishart Matrix ($\beta = 1$) and complex Wishart matrix ($\beta = 2$). Since $U_n = p\lambda_1/(\sum_{i=1}^p \lambda_i)$ is invariant to $\sigma^2$, the analysis does not depend on the specific values of $\sigma^2$, and we take $\sigma^2$ as follows in order to simplify the notation and unify the real and complex cases under the same representation as specified in the following equation (4).

- When $\beta = 1$, we assume that $\sigma^2 = 1$. That is, $\mathbf{X}$'s entries are i.i.d. $\mathcal{N}(0, 1)$, and $(\lambda_1, \ldots, \lambda_p)$ are ordered eigenvalues of $n^{-1}\mathbf{X}^\top\mathbf{X}$.
- When $\beta = 2$, we assume $\sigma^2 = 2$. Consider the circularly symmetric Gaussian random variable [e.g., 27], and we say $X := Y + iZ \sim \mathcal{CN}(0, \sigma^2)$ when $Y$ and $Z$ are i.i.d. $\mathcal{N}(0, \sigma^2/2)$. In the following, we assume that $\mathbf{X}$'s entries are i.i.d. $\mathcal{CN}(0, 2)$, and $(\lambda_1, \ldots, \lambda_p)$ are ordered eigenvalues of $n^{-1}\mathbf{X}^H\mathbf{X}$.

Under the two cases with $\beta = 1$ and 2, the $p$ eigenvalues $\lambda_1 \geq \ldots \geq \lambda_p \geq 0$ are distributed with the following probability density function [e.g., 9]:

$$f_{n,p,\beta}(\lambda) = C_{n,p,\beta} \prod_{i<j}^{p} \left|\lambda_i - \lambda_j\right|^\beta \prod_{i=1}^{p} \lambda_i^{\frac{\beta(n-p+1)}{2}-1} e^{-\frac{n}{2}\sum_{i=1}^p \lambda_i}, \quad \text{for } \beta = 1 \text{ and } 2, \tag{4}$$

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

$$C_{n,p,\beta} = p!\left(\frac{n}{2}\right)^{\frac{\beta np}{2}} \prod_{j=1}^{p} \frac{\Gamma(1 + \frac{\beta}{2})}{\Gamma(1 + \frac{\beta}{2}j)\Gamma\{\frac{\beta}{2}(n - p + j)\}}.$$

Then the target probability $\alpha_n(x) = P(U_n > x)$ can be written as

$$\alpha_n(x) = \int_{\lambda_1 \geq \ldots \geq \lambda_p \geq 0} 1(U_n > x) f_{n,p,\beta}(\lambda_1, \ldots, \lambda_p) d\lambda_1 \ldots d\lambda_p.$$

where $1(\cdot)$ is the indicator function. As discussed in the introduction, directly evaluating the above $p$-dimensional integral is computationally challenging, especially when $p$ is relatively large.

This work aims to design an efficient Monte Carlo method to estimate $\alpha_n(x)$. We first introduce some computational concepts in rare-event analysis literature, which helps to evaluate the computation efficiency of a Monte Carlo estimator.

Consider an estimator $L_n(x)$ of a rare-event probability $\alpha_n(x)$, which goes to 0 as $n \to \infty$. We simulate $N$ i.i.d. copies of $L_n(x)$, $\{L_n^{(j)}(x) : j = 1, ..., N\}$ and obtain the average estimator $\bar{L}_n(x) = N^{-1}\sum_{j=1}^N L_n^{(j)}(x)$. We want to control the relative error $|\bar{L}_n(x) - \alpha_n(x)|/\alpha_n(x)$ such that for some prescribed $\varepsilon, \delta > 0$,

$$\Pr\left\{|\bar{L}_n(x) - \alpha_n(x)|/\alpha_n(x) > \varepsilon\right\} < \delta.$$

Consider the direct Monte Carlo estimator for an example. The direct Monte Carlo directly generates samples from the density (4) and uses $L_n(x) = 1(U_n > x)$. So each simulation we have a Bernoulli variable with mean $\alpha_n(x)$. According to the central limit theorem, the direct Monte Carlo simulation requires $N = \Theta\{\varepsilon^{-2}\delta^{-1}\alpha_n(x)^{-1}\}$ i.i.d. replicates to achieve the above accuracy, where the notation $\Theta(\cdot)$ is defined as follows. For any $a_n$ and $b_n$ depending on $n$, $a_n = \Theta(b_n)$ denotes that $0 < \liminf_{n\to\infty} |a_n/b_n| \leq \limsup_{n\to\infty} |a_n/b_n| < \infty$. This implies that the direct Monte Carlo method becomes inefficient and even infeasible as $\alpha_n(x) \to 0$.

A more efficient estimator is the *asymptotically efficient* estimator [e.g., 3, 26]. An unbiased estimator $L_n(x)$ of $\alpha_n(x)$ is called asymptotically efficient if

$$\liminf_{n\to\infty} \frac{\ln \operatorname{Var}\{L_n(x)\}}{\ln \alpha_n(x)^2} \geq 1. \tag{5}$$

Note that (5) is equivalent to

$$\limsup_{n\to\infty} \frac{\text{Var}\{L_n(x)\}}{\alpha_n(x)^{2-\eta}} = 0, \tag{6}$$

for any $\eta > 0$. In addition, since $\text{E}(L_n^2) \geq Var\{L_n(x)\}$ and

$$\limsup_{n\to\infty} \frac{\ln \text{E}(L_n^2)}{\ln \alpha_n(x)^2} \leq 1$$

by Hölder's inequality that $\text{E}(L_p^2) \geq \{\text{E}(L_p)\}^2 = \alpha_n^2(x)$, (5) is also equivalent to

$$\lim_{n\to\infty} \frac{\ln \text{E}(L_p^2)}{\ln \alpha_n(x)^2} = 1.$$

When $L_n(x)$ is asymptotically efficient, by Chebyshevs inequality,

$$\Pr\left\{|\bar{L}_n(x) - \alpha(x)|/\alpha(x) > \varepsilon\right\} \leq \text{Var}\{L_n(x)\}/\{N\alpha(x)^2\epsilon^2\},$$

and therefore (6) implies that we only need $N = O\{\varepsilon^{-2}\delta^{-1}\alpha(x)^{-\eta}\}$, for any $\eta > 0$, i.i.d. replicates of $L_n(x)$. Compared with the direct Monte Carlo simulation, the efficient estimators substantially reduce the computational cost, especially when $\alpha_n(x)$ is small.

To construct an asymptotically efficient estimator, we use the importance sampling technique, which is a popularly used method for variance reduction of a Monte Carlo estimator. We use $P$ to denote the probability measure of the eigenvalues $(\lambda_1, \ldots, \lambda_n)$. The importance sampling estimator is constructed based on the following identity:

$$\Pr(U_n > x) = \text{E}\{1(U_n > x)\} = \text{E}_Q\left\{1(U_n > x)\frac{dP}{dQ}\right\},$$

where $Q$ is a probability measure such that the Radon-Nikodym derivative $dP/dQ$ is well defined on the set $\{U_n > x\}$, and we use E and $\text{E}_Q$ to denote the expectations under the measures $P$ and $Q$, respectively. Let $f_{n,p}^Q(\cdot)$ be the density function of the eigenvalues $(\lambda_1, \ldots, \lambda_n)$ under the change of measure $Q$. Then, the random variable defined by

$$L_n := \frac{f_{n,p}(\lambda_1, \ldots, \lambda_n)}{f_{n,p}^Q(\lambda_1, \ldots, \lambda_n)} 1(U_n > x)$$

is an unbiased estimator of $\alpha_n(x)$ under the measure $Q$. Therefore, to have $L_n$ asymptotically efficient, we only need to choose a change of measure $Q$ such that

$$\liminf_{n\to\infty} \frac{\left|\ln \text{E}_Q\left\{\frac{f_{n,p}(\lambda_1,\ldots,\lambda_n)^2}{f_{n,p}^Q(\lambda_1,\ldots,\lambda_n)^2} 1(U_n > x)\right\}\right|}{|2\ln \alpha_n(x)|} \geq 1. \tag{7}$$

To have an insight of the requirement (7), we consider some examples. First consider the direct Monte Carlo with $f_{n,p}^Q(\cdot) = f_{n,p}(\cdot)$, the right hand side of (7) then equals $1/2$ which is smaller than 1. On the other hand, consider $Q(\cdot)$ to be the conditional probability measure given $U_n > x$, i.e., $f_{n,p}^Q(\cdot) = a_n(x)^{-1} f_{n,p}(\cdot) 1(U_n > x)$; then the right hand side of (7) is exactly 1. Note that this change of measure is of no practical use since $L_n$ depends on the unknown $\alpha_n(x)$. But if we can find a measure $Q$ that is a good approximation of the conditional probability measure given $U_n > x$, we would expect (7) to hold and the corresponding estimator $L_n$ to be efficient. In other words, the asymptotic efficiency criterion requires the change of measure $Q$ is a good approximation of the conditional distribution of interest.

Following the above argument, we construct the change of measure $Q$ as follows, which is motivated by a recent study of [12]. [12] studied the tail probability of the largest eigenvalue, i.e., $\Pr(\lambda_1 > px)$ with $p > n$ and proposed a change of measure that approximates the conditional probability measure given $\lambda_1 > px$ in total variation when $p \gg n$. It is known that the asymptotic behavior $\lambda_1$ and $U_n$ are closely related. We therefore adapt the change of measure for the current problem of estimating $U_n$. However, we would like to clarify that the problem of estimating

$U_n$ is different from that in [12] in terms of both theoretical justification and computational implementation, which is further discussed in Remark 3.

Specifically, we propose the following importance sampling estimator.

**Algorithm 1.** *Every iteration in the algorithm contains three steps:*

*Step 1* *We use the matrix representation of the $\beta$-Laguerre ensemble introduced in [9], and 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*

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

*The notation $\chi_a$ denotes the square-root of the chi-square distribution with degree of freedom $a$, and the diagonal and sub diagonal elements of $\mathbf{B}_{n-1,p-1,\beta}$ are generated independently. We then calculate the corresponding ordered eigenvalues of $n^{-1}\mathbf{L}_{n-1,p-1,\beta}$, denoted by $\lambda_2 \geq \ldots \geq \lambda_p$.*

*Step 2* *Conditional on $(\lambda_2, \ldots, \lambda_p)$, sample $\lambda_1$ from an exponential distribution with density*

$$f(\lambda_1) = nre^{-nr(\lambda_1 - \tilde{x}\vee\lambda_2)} \cdot I(\lambda_1 > \tilde{x} \vee \lambda_2), \tag{8}$$

*where $r$ is a rate function such that*

$$r = \frac{1}{2} - \beta\gamma \int \frac{1}{\beta x - y} d\sigma_\beta(y) - \frac{1-\gamma}{2x} \tag{9}$$

*with $\gamma = p/n$, the $\sigma_\beta(\cdot)$ denotes the probability distribution function of the Marchenko-Pastur law such that*

$$\sigma_\beta(ds) = \frac{\sqrt{(s-s_*)(s^*-s)}}{\beta \times 2\pi\gamma s} 1_{s\in[s_*,s^*]} ds \tag{10}$$

*with $s^* = \beta\left(\sqrt{\gamma}+1\right)^2$ and $s_* = \beta\left(\sqrt{\gamma}-1\right)^2$, and $\tilde{x}$ is a constant depending on $n$, $p$, $\beta$ and $x$ such that*

$$\tilde{x} = \frac{x \operatorname{Tr}(n^{-1}\mathbf{L}_{n-1,p-1,\beta})}{p-x}.$$

*Step 3* *Based on the collected $\lambda_1 \geq \ldots \geq \lambda_p$, a corresponding importance sampling estimate can be calculated as in* (12) *below and the value of the estimate is saved.*

*The three steps above are repeated in every iteration. After the last iteration, the saved sampling estimates from all iterations are averaged to give an unbiased estimate of $\alpha(x)$.*

Now we give the details on how to compute the importance sampling estimate (12) in every iteration of the algorithm. Let $Q$ be the measure induced by combining the above two-step sampling procedure. From [9], under the change of measure $Q$, the density of $(\lambda_2^*, \ldots, \lambda_p^*) := n(n-1)^{-1}(\lambda_2, \ldots, \lambda_p)$ is

$$f_{n,p}^{Q}(\lambda_2^*, \ldots, \lambda_p^*) = C_{n-1,p-1,\beta} \prod_{2\leq i<j\leq p} |\lambda_i^* - \lambda_j^*|^\beta \cdot \prod_{i=2}^{p} (\lambda_i^*)^{\frac{\beta(n-p+1)}{2}-1} \cdot e^{-\frac{n-1}{2}\sum_{i=2}^{p}\lambda_i^*}.$$

This implies the density function of $(\lambda_2, \ldots, \lambda_p)$ under $Q$ is

$$f^Q_{n,p}(\lambda_2, \ldots, \lambda_p) = \left(\frac{n}{n-1}\right)^{\frac{\beta(n-1)(p-1)}{2}} C_{n-1,p-1,\beta} \prod_{2\le i<j\le p} |\lambda_i - \lambda_j|^\beta \cdot \prod_{i=2}^{p} \lambda_i^{\frac{\beta(n-p+1)}{2}-1} \cdot e^{-\frac{n}{2}\sum_{i=2}^p \lambda_i}. \tag{11}$$

Therefore $dQ/dP$ takes the form

$$\begin{aligned} \frac{dQ}{dP} &= \frac{f^Q_{n,p}(\lambda_2, \ldots, \lambda_p) \times nre^{-nr(\lambda_1 - \tilde{x}\vee\lambda_2)} \cdot I_{(\lambda_1 > \tilde{x}\vee\lambda_2)}}{f_{n,p}(\lambda_1, \lambda_2, \ldots, \lambda_p)} \\ &= \frac{\left(\frac{n}{n-1}\right)^{\frac{\beta(n-1)(p-1)}{2}} C_{n-1,p-1,\beta}\, nre^{-nr(\lambda_1 - \tilde{x}\vee\lambda_2)} \cdot I_{(\lambda_1 > \tilde{x}\vee\lambda_2)}}{C_{n,p,\beta} \prod_{i=2}^p (\lambda_1 - \lambda_i) \cdot \lambda_1^{\frac{\beta(n-p+1)}{2}-1} \cdot e^{-\frac{n}{2}\lambda_1}}. \end{aligned}$$

The corresponding importance sampling estimate is given by

$$L_n(x) = \frac{dP}{dQ} 1_{(U_n > x)}, \tag{12}$$

where $U_n$ is calculated with the sampled $\lambda_1, \ldots, \lambda_p$ based on the (1).

We claim that for the proposed algorithm 1, with the chosen $r$ in (9), the importance sampling estimator $L_n(x)$ is asymptotically efficient in estimating the target tail probability. This result is given in Theorem 1 below, whose proof is given in Section 5.

**Theorem 1.** *When $p/n \to \gamma \in \mathbb{R}$, the estimator $L_n(x)$ in* (12) *is asymptotically efficient in estimating $\alpha_n(x)$ for $x > \left(\sqrt{\gamma} + 1\right)^2$.*

**Remark 1.** *Our discussion on the asymptotic efficiency focuses on the case of estimating rare-event tail probability $\alpha_n(x)$, that is, when $\{U_n > x\}$ corresponds to a rare event. When $x \le \left(\sqrt{\gamma} + 1\right)^2$, $\{U_n > x\}$ is not rare, and we can still apply the importance sampling algorithm with a reasonable positive r value as the exponential distribution's rate. However, the theoretical properties of the importance sampling estimator shall be studied under a different framework and therefore is not further pursued in this study.*

**Remark 2.** *We explain the Marchenko-Pastur form of (10). When $\mathbf{X}$'s entries have mean 0 and variance 1 ($\beta = 1$ and 2), the Marchenko-Pastur law for eigenvalues of $n^{-1}\mathbf{X}^H\mathbf{X}$ takes the following standard form [e.g., Theorem 3.2 in 24]*

$$f(d\bar{s}) = \frac{\sqrt{(\bar{s}_+ - \bar{s})(\bar{s} - \bar{s}_-)}}{2\pi\gamma\bar{s}} \mathbf{1}_{[\bar{s}_-, \bar{s}_+]}(\bar{s}) d\bar{s} \tag{13}$$

*with $\bar{s}_- = (1 - \sqrt{\gamma})^2$ and $\bar{s}_+ = (1 + \sqrt{\gamma})^2$. For the considered setting of this paper, the real case ($\beta = 1$) has $\sigma^2 = 1$, so* (10) *and* (13) *are consistent. On the other hand, the complex case ($\beta = 2$) has $\sigma^2 = 2$ and therefore* (10) *and* (13) *are different up to a factor of $\beta = 2$. Specifically, let $(\bar{\lambda}_1, \ldots, \bar{\lambda}_p)$ and $(\lambda_1, \ldots, \lambda_p)$ be eigenvalues of $n^{-1}\mathbf{X}^H\mathbf{X}$ when $\mathbf{X}$ has i.i.d. entries of $\mathcal{CN}(0, 1)$ and $\mathcal{CN}(0, 2)$ respectively. Then we know $(\lambda_1, \ldots, \lambda_p) \sim 2(\bar{\lambda}_1, \ldots, \bar{\lambda}_p)$ and* (13) *implies the empirical distribution in (10).*

**Remark 3.** *We discuss the differences between the proposed method and the method in [12] on the largest eigenvalue, which also employs an importance sampling technique. First, the two methods have different targets, i.e., $\Pr(\lambda_1 > x)$ in [12] and $\Pr(U_n > x)$ here, and therefore use different change of measures to construct efficient importance sampling estimators. As discussed in Section 2, in order to achieve asymptotical efficiency, the change of measures should approximate the target conditional distribution measures, i.e., $\Pr(\cdot \mid \lambda_1 > x)$ in [12] and $\Pr(\cdot \mid U_n > x)$ in this paper. Due to the difference between the two conditional distributions, two different change of measures are constructed in the two methods. Specifically, [12] samples the largest eigenvalue $\lambda_1$ from a truncated exponential distribution depending on the second largest eigenvalue $\lambda_2$ while this work samples $\lambda_1$ from an exponential distribution depending on eigenvalues $(\lambda_2, \cdots, \lambda_p)$. Second, the proof techniques of the main asymptotic results in the two papers are also* *different. In particular, to show the asymptotic efficiency of the importance sampling estimators as defined in* (5)*, we need to derive asymptotic approximations for both the rare-event probability* $\alpha(x)$ *and the second moments of the importance sampling estimator* $E_Q\{L_n^2(x)\}$*. Even though the largest eigenvalue* $\lambda_1$ *and the ratio statistic* $U_n$ *have similar large deviation approximation results for their tail probabilities, the asymptotic approximations for the second moments of the importance sampling estimators are different due to the differences between the considered change of measures as well as the effect of the trace term in* $U_n$*. Please refer to the proof for more details.*

**Remark 4.** *The method and the theoretical results can be easily extended from the case of* $p \leq n$ *to* $p \geq n$ *by switching the labels of n and p and changing* $\gamma$ *to* $\gamma^{-1}$ *correspondingly. Note that when* $p \geq n$*, eigenvalues of* $n^{-1}\mathbf{X}^H\mathbf{X}$ *and* $p^{-1}\mathbf{X}\mathbf{X}^H$ *give the same test statistic* $U_n$ *as defined in* (1)*, which is because* $\mathbf{X}^H\mathbf{X}$ *and* $\mathbf{X}\mathbf{X}^H$ *have the same set of nonzero eigenvalues and* $U_n$ *is scale invariant. By symmetry, when* $p \geq n$*, the joint density function of the eigenvalues of* $p^{-1}\mathbf{X}\mathbf{X}^H$ *have the same form as* (4)*, except that the labels of n and p are switched. Therefore, the cases when* $p \leq n$ *and* $p \geq n$ *are equivalent up to the label switching. Note that after* $p/n$ *is changed to* $n/p$*,* $\gamma$ *becomes* $\gamma^{-1}$ *correspondingly.*

## 3. Numerical study

We conduct numerical studies to evaluate the performance of our algorithm. We first take combinations $(n, p) = (100, 10), (100, 20), (500, 20), (1000, 50)$, and $\beta = 1, 2$, respectively. Then we compare our algorithm with other methods and present the results in Table 1 and 2.

For the proposed importance sampling estimator, we repeat $N_{IS} = 10^4$ times and show the estimated probabilities ("$EST_{IS}$" column) along with the estimated standard deviations of $L_p$, i.e., $\sqrt{Var^Q(L_p)}$ ("$SD_{IS}$" column). The ratios between estimated standard deviations and estimates ("$SD_{IS}/EST_{IS}$" column) indicate the efficiency of algorithms. Note that with $N_{IS} = 10^4$ replications, the standard error of the estimate is $SD_{IS}/\sqrt{N_{IS}} = SD_{IS}/100$. In addition, three alternative methods are considered, including the direct Monte Carlo, the Tracy-Widom distribution approximation, and the corrected Tracy-Widom approximation [23]. We compute direct Monte Carlo estimates ("$EST_{DMC}$" column) with $N_{DMC} = 10^6$ independent replications. We present the standard deviation of direct Monte Carlo estimates ("$SD_{DMC}$" column) and the ratios between estimated standard deviations and estimates ("$SD_{DMC}/EST_{DMC}$"). In addition, we use the approximation of Tracy-Widom distribution ("$TW$" column) specified in equation (2). The $TW(x)$ is computed from `RMTstat` package in R. Furthermore, following [23], we compute the Tracy-Widom approximation with correction term ("$c.TW$" column):

$$\Pr\left(\frac{U-\mu_{n,p}}{\sigma_{n,p}} > x\right) \approx 1 - \mathcal{TW}_\beta(x) + \frac{1}{2}\left(\frac{2}{np}\right)\left(\frac{\mu_{n,p}}{\sigma_{n,p}}\right)^2 \mathcal{TW}_\beta''(x), \tag{14}$$

where $\mathcal{TW}''(x)$ is computed numerically via a standard central differencing scheme with $\Delta x = 10^{-3}$. When $\beta = 1$, $\mu$ and $\sigma$ is chosen according to equation (3). When $\beta = 2$, $\mu$ and $\sigma$ is chosen according to [15].

We can see from Table 1 and Table 2 that the Tracy-Widom distribution ("$TW$" column) significantly overestimates the tail probabilities for all considered settings and the finding is consistent with that in [23]. And the corrected Tracy-Widom approximation ("$c.TW$" column) underestimates the tail probability $\alpha_n(x)$ and goes to a negative number as $\alpha_n(x)$ goes small. Since the proposed importance sampling and the direct Monte Carlo method are both unbiased estimators, next we compare their computational efficiency. As discussed in Section 2, for the average estimator $\bar{L}_n(x) = N^{-1}\sum_{j=1}^N L_n^{(j)}(x)$, "$SD_{IS}/EST_{IS}$" and "$SD_{DMC}/EST_{DMC}$" can be used as a measure of the computational efficiency in terms of iteration numbers. From the results in Tables 1 and 2, as $\alpha(x)$ decreases, "$SD_{DMC}/EST_{DMC}$" goes large quickly and even becomes "NaN". On the other hand, "$SD_{IS}/EST_{IS}$" increases slowly and is generally smaller than "$SD_{DMC}/EST_{DMC}$", showing that the proposed importance sampling is more efficient than the direct Monte Carlo method.

To further illustrate, we compare the iteration numbers $N_{IS}$ and $N_{DMC}$ that would be needed to achieve the same level of relative standard errors of the estimators. Specifically, in order to have the same ratios of the standard errors to the estimates, i.e., $SE_{IS}/EST_{IS} = (SD_{IS}/\sqrt{N_{IS}})/EST_{IS}$ and $SE_{DMC}/EST_{DMC} = (SD_{DMC}/\sqrt{N_{DMC}})/EST_{DMC}$, obtained under the importance sampling and direct direct Monte Carlo, respectively, we need

$$\frac{N_{DMC}}{N_{IS}} = \frac{(SD_{DMC}/EST_{DMC})^2}{(SD_{IS}/EST_{IS})^2}. \tag{15}$$

Based on the above equation, the simulation results show that to have a similar standard error obtained under the importance sampling, the direct Monte Carlo method needs more iterations as $\alpha(x)$ goes small. For example, from Table 1, when $n = 100$, $p = 10$ and $x = 2.1$, we need $N_{DMC}$ to be approximately $4.3 \times 10^2$ times larger than $N_{IS}$; when $n = 1000$, $p = 50$ and $x = 1.62$, we need $N_{DMC}$ about $1.3 \times 10^4$ times larger.

Besides from the iteration numbers, we also compare the average time cost of each iteration under the importance sampling and the direct Monte Carlo method, respectively. For the direct Monte Carlo, two methods are considered in computing the eigenvalues. The first method directly computes the test statistic $U_n$ using the eigen-decomposition of a randomly sampled Wishart matrix. The second method computes the eigenvalues from the tridiagonal representation form as in Step 1 of Algorithm 1. We run $10^4$ iterations for all the methods and report the average time of one iteration in Table 3, where the first method of the direct Monte Carlo is denoted as $T_{DMC\_1}$, the second method is denoted as $T_{DMC\_2}$, and the importance sampling method is denoted as $T_{IS}$. The simulation results show that $T_{DMC\_1}$ has the highest time cost per iteration, while $T_{DMC\_2}$ and $T_{IS}$ are similar. We further explain the simulation results from the perspective of algorithm complexity. For each iteration, the first direct Monte Carlo method samples a $p \times p$ Wishart matrix and performs its eigen-decomposition, which usually has the cost of $O(p^3)$. The second direct Monte Carlo method and the importance sampling only need to sample $O(p)$ number of $\chi^2$ random variables and then decompose a symmetric tridiagonal matrix, which have $O(p^2)$ cost per iteration [8]. Although the importance sampling also samples from an exponential distribution in Step 2, the distribution parameters can be calculated in advance and it does not affect the overall complexity much. Therefore, the time complexity of the algorithm $T_{DMC\_1}$ is higher while $T_{DMC\_2}$ and $T_{IS}$ are similar per iteration. Together with the result in (15), we can see that the importance sampling is more efficient than the direct Monte Carlo method in terms of both the iteration number and the overall time cost.

To further check the influence of replication number $N_{IS}$ of the importance sampling algorithm, we focus on the case of $n = 100$ and $p = 10$ and compare the performance of different $N_{IS}$'s. In order to obtain accurate reference values of the tail probabilities, we use direct Monte Carlo with repeating time $N_{DMC} = 10^8$ to estimate multiple tail probabilities $\alpha_n(x)$'s ranging from $10^{-2}$ to $10^{-6}$ under $\beta = 1, 2$ respectively. Then we estimate the corresponding $\alpha_n(x)$'s using our algorithm with $N_{IS} = 10^4, 10^5, 10^6$ respectively. The results are presented in Figure 1, where the x-axis represents the reference values $\log_{10}(EST_{DMC})$. The line "*DMC with error bar*" represents the (approximated) pointwise 95% confidence intervals $\left[\log_{10}\left(EST_{DMC} - 2 \times SD_{DMC}/\sqrt{N_{DMC}}\right), \log_{10}\left(EST_{DMC} + 2 \times SD_{DMC}/\sqrt{N_{DMC}}\right)\right]$. Similarly, the line "*Importance Sampling with error bar*" represents the importance sampling estimates and pointwise 95% confidence intervals $\left[\log_{10}\left(SD_{IS} - 2 \times SD_{IS}/\sqrt{N_{IS}}\right), \log_{10}\left(EST_{IS} + 2 \times SD_{IS}/\sqrt{N_{IS}}\right)\right]$. From the figures, the proposed algorithm can well estimate the probability as small as $10^{-6}$ with $N_{IS} = 10^4$, which is more efficient than directed Monte Carlo and more accurate than Tracy-Widom approximations. Furthermore, Figure 1 shows that the algorithm improves when number of iterations increases. We also plot the Tracy-Widom approximations in (2) and (14) in Figure 1 for comparison. Figure 1 shows that without correction, the Tracy-Widom distribution in (2) is not accurate and overestimates the probabilies. The correction term in (14) improves the approximation when the probability is larger than the scale of about $10^{-2}$, which is consistent with the result in [23]. But when the probability goes smaller, the corrected approximation has larger deviation from true values (on the $\log_{10}$ scale) and even becomes negative. Note that since we cannot plot the $\log_{10}$ of negative numbers in the figures, the lines of the corrected Tracy-Widom approximations appear to be shorter. These results validate the results in Table 1 and 2.

Table 1: Estimation Results for $\beta = 1$

(a) n=100 , p=10

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ | c.TW | TW |
|---|---|---|---|---|---|---|---|---|
| 1.80 | 2.44e-2 | 1.25e-1 | 5.14 | 2.46e-2 | 1.55e-1 | 6.30 | 2.58e-2 | 5.07e-2 |
| 1.95 | 1.02e-3 | 5.00e-3 | 4.89 | 1.08e-3 | 3.28e-2 | 30.46 | 3.90e-4 | 4.37e-3 |
| 1.98 | 5.32e-4 | 3.55e-3 | 6.66 | 5.57e-4 | 2.36e-2 | 42.36 | 4.96e-6 | 2.48e-3 |
| 2.10 | 2.43e-5 | 2.48e-4 | 10.22 | 2.20e-5 | 4.69e-3 | 213.20 | -7.46e-5 | 2.07e-4 |
| 2.30 | 5.25e-8 | 7.72e-7 | 14.71 | 0 | 0 | NaN | 0 | 0 |

(b) n=100 , p=20

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ | c.TW | TW |
|---|---|---|---|---|---|---|---|---|
| 2.10 | 9.14e-2 | 3.73e-1 | 4.09 | 8.99e-2 | 2.86e-1 | 3.18 | 9.29e-2 | 1.21e-1 |
| 2.30 | 2.86e-3 | 2.04e-2 | 7.13 | 2.71e-3 | 5.20e-2 | 19.19 | 2.31e-3 | 6.09e-3 |
| 2.40 | 3.44e-4 | 2.60e-3 | 7.54 | 3.11e-4 | 1.76e-2 | 56.70 | 1.54e-4 | 9.07e-4 |
| 2.50 | 2.89e-5 | 2.01e-4 | 6.95 | 2.60e-5 | 5.10e-3 | 196.11 | -6.13e-6 | 1.05e-4 |
| 2.70 | 1.50e-7 | 1.78e-6 | 11.85 | 0 | 0 | NaN | 0 | 0 |

(c) n=500 , p=20

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ | c.TW | TW |
|---|---|---|---|---|---|---|---|---|
| 1.46 | 4.64e-2 | 2.21e-1 | 4.76 | 4.68e-2 | 2.11e-1 | 4.51 | 4.87e-2 | 6.51e-2 |
| 1.51 | 3.98e-3 | 2.16e-2 | 5.43 | 3.70e-3 | 6.07e-2 | 16.40 | 3.70e-3 | 7.03e-3 |
| 1.56 | 1.57e-4 | 7.13e-4 | 4.54 | 1.55e-4 | 1.24e-2 | 80.32 | 1.28e-4 | 4.40e-4 |
| 1.62 | 2.14e-6 | 1.49e-5 | 6.97 | 3.00e-6 | 1.73e-3 | 577.35 | -1.87e-6 | 6.71e-6 |
| 1.70 | 2.43e-9 | 2.72e-8 | 11.20 | 0 | 0 | NaN | 0 | 0 |

(d) n=1000 , p=50

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ | c.TW | TW |
|---|---|---|---|---|---|---|---|---|
| 1.52 | 2.75e-2 | 1.29e-1 | 4.70 | 2.90e-2 | 1.68e-1 | 5.78 | 2.96e-2 | 3.59e-2 |
| 1.55 | 2.51e-3 | 1.16e-2 | 4.63 | 2.57e-3 | 5.06e-2 | 19.71 | 2.53e-3 | 7.98e-4 |
| 1.60 | 1.41e-5 | 5.25e-5 | 3.72 | 2.20e-5 | 4.69e-3 | 213.20 | 1.15e-5 | 3.25e-5 |
| 1.62 | 1.40e-6 | 8.70e-6 | 6.21 | 2.00e-6 | 1.41e-3 | 707.11 | -7.93e-7 | 6.71e-6 |
| 1.66 | 7.49e-9 | 3.69e-8 | 4.93 | 0 | 0 | NaN | 0 | 0 |

Table 2: Estimation Results for for $\beta = 2$

(a) n=100 , p=10

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ | c.TW | TW |
|---|---|---|---|---|---|---|---|---|
| 1.77 | 3.72e-3 | 3.34e-2 | 8.98 | 3.79e-3 | 6.15e-2 | 16.21 | 2.20e-3 | 1.26e-2 |
| 1.81 | 9.21e-4 | 1.32e-2 | 14.34 | 8.97e-4 | 2.99e-2 | 33.37 | -1.36e-4 | 4.42e-3 |
| 1.91 | 1.89e-5 | 3.28e-4 | 17.37 | 1.70e-5 | 4.12e-3 | 242.53 | -1.22e-4 | 2.11e-4 |
| 1.93 | 6.68e-6 | 8.44e-5 | 12.64 | 4.00e-6 | 2.00e-3 | 500 | -7.44e-5 | 1.07e-4 |
| 1.99 | 2.98e-7 | 4.25e-6 | 14.27 | 0 | 0 | NaN | -1.29e-5 | 1.24e-5 |

(b) n=100 , p=20

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ | c.TW | TW |
|---|---|---|---|---|---|---|---|---|
| 2.10 | 1.20e-2 | 7.99e-2 | 6.68 | 1.45e-2 | 1.20e-1 | 8.23 | 1.41e-2 | 2.70e-2 |
| 2.18 | 1.04e-3 | 7.59e-3 | 7.28 | 1.34e-3 | 3.66e-2 | 27.29 | 8.64e-4 | 3.65e-3 |
| 2.30 | 2.18e-5 | 3.47e-4 | 15.94 | 2.30e-5 | 4.80e-3 | 208.51 | -2.06e-5 | 8.86e-5 |
| 2.38 | 6.73e-7 | 1.94e-5 | 28.86 | 1.00e-6 | 1.00e-3 | 1000 | -2.70e-6 | 4.83e-6 |
| 2.46 | 1.63e-8 | 2.83e-7 | 17.36 | 0 | 0 | NaN | -1.73e-7 | 1.93e-7 |

(c) n=500 , p=20

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ | c.TW | TW |
|---|---|---|---|---|---|---|---|---|
| 1.45 | 8.04e-3 | 5.49e-4 | 6.84 | 8.98e-3 | 9.43e-2 | 10.51 | 8.95e-3 | 1.58e-2 |
| 1.48 | 6.56e-4 | 8.02e-3 | 12.22 | 6.49e-4 | 2.55e-2 | 39.24 | 5.07e-4 | 1.59e-3 |
| 1.50 | 8.77e-5 | 1.16e-3 | 13.18 | 8.60e-5 | 9.27e-3 | 107.83 | 3.88e-5 | 2.70e-4 |
| 1.525 | 5.05e-6 | 5.37e-5 | 10.63 | 8.00e-6 | 2.83e-3 | 353.55 | -1.87e-6 | 2.28e-5 |
| 1.55 | 1.85e-7 | 1.71e-6 | 9.28 | 0 | 0 | NaN | -4.66e-7 | 1.49e-6 |

(d) n=1000 , p=50

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ | c.TW | TW |
|---|---|---|---|---|---|---|---|---|
| 1.51 | 5.85e-3 | 6.67e-2 | 11.39 | 5.20e-3 | 7.19e-2 | 13.83 | 5.31e-3 | 7.46e-3 |
| 1.53 | 2.65e-4 | 1.96e-3 | 7.39 | 3.04e-4 | 1.74e-2 | 57.35 | 2.98e-4 | 5.32e-4 |
| 1.56 | 1.72e-6 | 1.84e-5 | 10.72 | 0 | 0 | NaN | 1.33e-6 | 4.20e-6 |
| 1.58 | 3.15e-8 | 2.86e-7 | 9.10 | 0 | 0 | NaN | 1.46e-8 | 9.85e-8 |
| 1.60 | 4.24e-10 | 3.80e-9 | 8.97 | 0 | 0 | NaN | -6.21e-11 | 1.56e-9 |

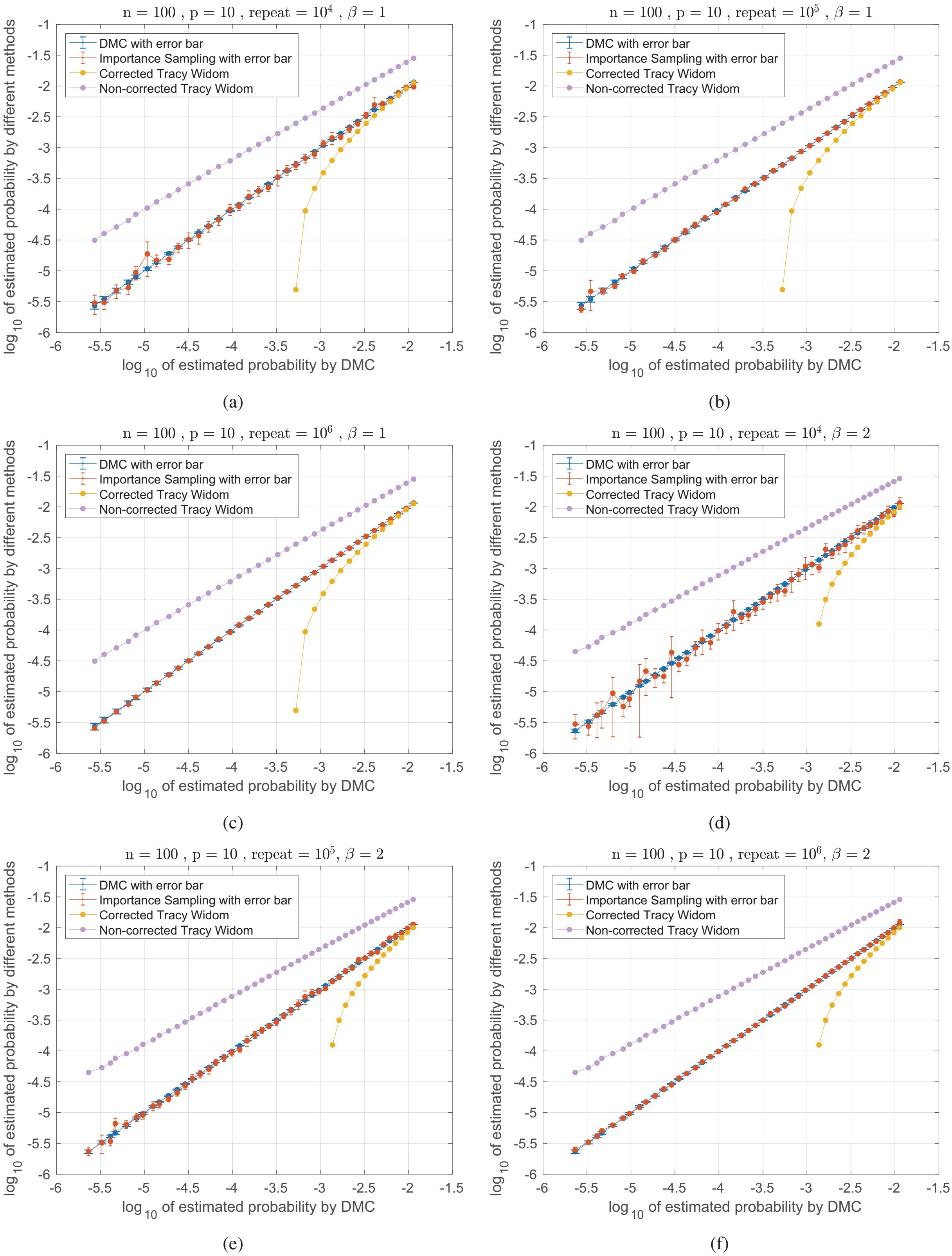


Figure 1: Estimation results for $n = 100$ and $p = 10$

Table 3: Estimation of Time

(a) $\beta = 1$

| $n$ | $p$ | $x$ | $T_{DMC\_1}$ | $T_{DMC\_2}$ | $T_{IS}$ |
|---|---|---|---|---|---|
| 100 | 10 | 1.95 | 1.28e-03 | 7.26e-04 | 8.89e-05 |
| 100 | 10 | 1.98 | 1.15e-03 | 9.23e-05 | 8.51e-05 |
| 100 | 20 | 2.3 | 1.58e-03 | 7.35e-05 | 6.84e-05 |
| 100 | 20 | 2.4 | 1.65e-03 | 1.79e-04 | 6.33e-05 |
| 500 | 20 | 1.51 | 1.27e-03 | 9.87e-05 | 9.32e-05 |
| 500 | 20 | 1.56 | 1.67e-03 | 7.39e-05 | 8.82e-05 |
| 1000 | 50 | 1.55 | 3.19e-03 | 1.05e-04 | 1.56e-04 |
| 1000 | 50 | 1.6 | 3.12e-03 | 9.76e-05 | 1.34e-04 |

(b) $\beta = 2$

| $n$ | $p$ | $x$ | $T_{DMC\_1}$ | $T_{DMC\_2}$ | $T_{IS}$ |
|---|---|---|---|---|---|
| 100 | 10 | 1.77 | 1.87e-03 | 1.75e-04 | 6.08e-05 |
| 100 | 10 | 1.81 | 1.85e-03 | 5.47e-05 | 5.90e-05 |
| 100 | 20 | 2.18 | 2.86e-03 | 8.37e-05 | 1.20e-04 |
| 100 | 20 | 2.3 | 2.69e-03 | 1.11e-04 | 6.69e-05 |
| 500 | 20 | 1.45 | 2.79e-03 | 8.46e-05 | 7.01e-05 |
| 500 | 20 | 1.48 | 3.53e-03 | 7.24e-05 | 8.90e-05 |
| 1000 | 50 | 1.53 | 8.65e-03 | 9.03e-05 | 1.53e-04 |
| 1000 | 50 | 1.56 | 8.35e-03 | 9.61e-05 | 1.49e-04 |

## 4. Conclusions and Extensions

This paper proposes an asymptotically efficient Monte Carlo method to estimate the tail probabilities of the ratio of the largest eigenvalue to the trace of the Wishart matrix. Theoretically, we prove the importance sampling estimator is asymptotic efficient. Numerically, we conduct extensive studies to evaluate the performance of the proposed algorithm compared with other existing methods in terms of estimation accuracy and computational cost in estimating the tail probabilities.

The method can be adapted to estimating tail probabilities of the ratio of the sum of the first $k$ largest eigenvalues to the trace of the Wishart matrix, which is defined as

$$U_n^k = \frac{\sum_{i=1}^k \lambda_i}{\min\{p,n\}^{-1} \sum_{i=1}^p \lambda_i},$$

where $k$ is a fixed positive integer. We consider the algorithm as follows. First, sample $\lambda_2, \ldots, \lambda_p$ from $n^{-1}\mathbf{L}_{n-1,p-1,\beta}$ using the same method in Algorithm 1. Second, conditioning on $\lambda_2, \ldots, \lambda_p$, sample $\lambda_1$ from a truncated exponential distribution with the same form as (8), but we redefine

$$\tilde{x} = \frac{x \sum_{i=2}^p \lambda_i - p \sum_{i=2}^k \lambda_i}{p - x}$$

and choose $r$ to be a small constant that depend on the large deviation result of the largest $k$ eigenvalues. We conducted a numerical study to show the validation and efficiency of the proposed method in estimating the tail probabilities of $U_n^k$. Following the design in Section 3, the sampling is repeated $10^4$ times for the importance sampling method and $10^6$ times for the direct Monte Carlo method. The $k$ is chosen to be $2, 3, 4$, $n = 100$, $p = 50$, and we take $r = 1/10$. Tables 4 and 5 summarize the results of $\beta = 1$ and $\beta = 2$, which show similar patterns as Tables 1 and 2. When the tail probability becomes smaller, $SD_{IS}/EST_{IS}$ is smaller than $SD_{DMC}/EST_{DMC}$, which indicates that the importance sampling is more efficient than the direct Monte Carlo method in estimating the tail probabilities as discussed in Section 3. It would be interesting to study the asymptotic property of this algorithm on estimating the tail probability of $U_n^k$; however, it needs development of asymptotic theory on the tail probabilities of the first $k$ largest eigenvalues, which is beyond the scope of this study, and we would like to leave it for future work.

Table 4: $U_n^k$ Results for $\beta = 1$

(a) n=100 , p=50, k=2

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ |
|---|---|---|---|---|---|---|
| 5.9 | 1.14e-03 | 6.70e-03 | 5.89 | 1.55e-03 | 3.93e-02 | 25.41 |
| 6.0 | 3.22e-04 | 3.80e-03 | 11.78 | 2.98e-04 | 1.73e-02 | 57.92 |
| 6.1 | 5.68e-05 | 9.37e-04 | 16.49 | 5.50e-05 | 7.42e-03 | 134.84 |
| 6.4 | 1.09e-07 | 3.21 e-06 | 29.50 | 0 | 0 | NaN |

(b) n=100 , p=50, k=3

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ |
|---|---|---|---|---|---|---|
| 8.4 | 1.56e-03 | 1.77e-02 | 11.36 | 1.55e-03 | 3.93e-02 | 25.41 |
| 8.5 | 4.22e-04 | 5.48e-03 | 12.98 | 4.04e-04 | 2.01e-02 | 49.74 |
| 8.7 | 1.46e-05 | 2.99e-04 | 20.44 | 1.80e-05 | 4.24e-03 | 235.70 |
| 8.9 | 7.26e-07 | 2.53e-05 | 34.83 | 0 | 0 | NaN |

(c) n=100 , p=50, k=4

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ |
|---|---|---|---|---|---|---|
| 10.6 | 7.60e-03 | 5.65e-02 | 7.43 | 8.01e-03 | 8.91e-02 | 11.13 |
| 10.8 | 6.58e-04 | 6.63e-03 | 10.08 | 8.44e-04 | 2.90e-02 | 34.41 |
| 11.0 | 5.49e-05 | 1.47e-03 | 26.73 | 6.40e-05 | 8.00e-03 | 125.00 |
| 11.3 | 1.70e-07 | 5.56e-06 | 32.77 | 0 | 0 | NaN |

## 5. Proof of Theorem 1

This section provides the proof for Theorem 1 of the estimator's asymptotic efficiency. We focus on the case when $p \le n$ and $p/n \to \gamma \in (0, 1]$. For the case of $p \ge n$ and $p/n \to \gamma \in [1, \infty)$, the proof follows from the same argument by switching the labels of $n$ and $p$, as shown in Remark 4, and therefore is skipped.

Recall the definition of $Q$, $L_p = \frac{dP}{dQ} 1(U_n > x)$ and $\alpha(x) = \Pr(U_n > x)$. To prove the asymptotic efficiency defined in (5), we only need to show $\liminf_{n\to\infty} \ln \mathrm{E}_Q(L_p^2)/\{2 \ln \alpha_n(x)\} \ge 1$ since $\mathrm{E}_Q(L_p^2)/\{2 \ln \alpha_n(x)\} \le \mathrm{Var}_Q(L_p^2)/\{2 \ln \alpha_n(x)\}$. We give an outline of the proof first.

Step 1. We give the asymptotic approximation of $\lim_{n\to\infty} n^{-1} \ln \alpha_n(x) = -\gamma I_\beta(\beta x)$, where $I_\beta(\beta x)$ is the large deviation rate function.

Step 2. By the result in Step 1, we only need to prove that

$$\liminf_{n\to\infty} \frac{\ln \mathrm{E}_Q(L_p^2)}{2 \ln \alpha_n(x)} = \liminf_{n\to\infty} \frac{\ln \mathrm{E}_Q(L_p^2)}{-2\gamma I_\beta(\beta x)} \ge 1.$$

This is established using the upper bound $I_1 + I_2 + I_3$ of $\mathrm{E}_Q(L_p^2)$ in (19) together with the limiting properties of $I_1, I_2$, and $I_3$ in (20) (21) and (22) respectively.

Table 5: $U_n^k$ Results for $\beta = 2$

(a) n=100 , p=50, k=2

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ |
|---|---|---|---|---|---|---|
| 5.6 | 4.67e-03 | 3.52e-02 | 7.54 | 5.03e-03 | 7.07e-02 | 14.07 |
| 5.7 | 5.08e-04 | 5.95e-03 | 11.72 | 4.98e-04 | 2.23e-02 | 44.80 |
| 5.8 | 4.75e-05 | 9.55e-04 | 20.12 | 3.80e-05 | 6.16e-03 | 162.23 |
| 6.0 | 7.71e-08 | 2.48e-06 | 32.18 | 0 | 0 | NaN |

(b) n=100 , p=50, k=3

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ |
|---|---|---|---|---|---|---|
| 8.1 | 1.78e-03 | 2.08e-02 | 11.67 | 2.16e-03 | 4.64e-02 | 21.50 |
| 8.2 | 3.67e-04 | 8.31e-03 | 22.67 | 2.90e-04 | 1.70e-02 | 58.71 |
| 8.3 | 1.87e-05 | 3.73e-04 | 19.96 | 2.50e-05 | 5.00e-03 | 200.00 |
| 8.5 | 1.50e-07 | 6.90e-06 | 45.95 | 0 | 0 | NaN |

(c) n=100 , p=50, k=4

| x | $EST_{IS}$ | $SD_{IS}$ | $SD_{IS}/EST_{IS}$ | $EST_{DMC}$ | $SD_{DMC}$ | $SD_{DMC}/EST_{DMC}$ |
|---|---|---|---|---|---|---|
| 10.4 | 2.49e-03 | 4.78e-02 | 19.18 | 2.73e-03 | 5.22e-02 | 19.12 |
| 10.5 | 4.27e-04 | 6.15e-03 | 14.40 | 4.42e-04 | 2.10e-02 | 47.55 |
| 10.6 | 5.47e-05 | 1.86e-03 | 34.04 | 6.90e-05 | 8.31e-03 | 120.38 |
| 10.8 | 3.17e-07 | 1.23e-05 | 38.96 | 0 | 0 | NaN |

The details of Steps 1 and 2 are given below.

**Step 1**. We first obtain the large deviation rate function for $U_n$, which gives an approximation to $n^{-1}\ln\alpha_n(x)$ as in [1]. From the argument in [4], the large deviation of $U_n$ has a similar rate function to $\lambda_1$. The explicit form of the large deviation rate function of $\lambda_1$ can be obtained from Theorem 2.6.6 in [1]. In particular, denote $(\tilde\lambda_1, \ldots, \tilde\lambda_p)$ to be the unordered eigenvalues of $n^{-1}\mathbf{X}^H\mathbf{X}$; then from (4), $(\tilde\lambda_1, \ldots, \tilde\lambda_p)$ has joint density function

$$
\begin{aligned}
f_{n,p,\beta}(\tilde\lambda_1, \ldots, \tilde\lambda_p) &= \frac{1}{p!}C_{n,p,\beta}\prod_{1\le i<j\le p}|\tilde\lambda_i - \tilde\lambda_j|^\beta \cdot \prod_{i=1}^p \tilde\lambda_i^{\frac{\beta(n-p+1)}{2}-1}\cdot e^{-\frac{n}{2}\sum_{i=1}^p\tilde\lambda_i}\\
&= \left(Z_{V,\beta}^p\right)^{-1}\left|\Delta_p(\tilde\lambda)\right|^\beta e^{-p\sum_{i=1}^p V(\tilde\lambda_i)},
\end{aligned}
$$

where the last line follows the notation of (2.6.1) in [1] with $\Delta_p(\tilde\lambda) := \prod_{1\le i<j\le p}(\tilde\lambda_i - \tilde\lambda_j)$, $Z_{V,\beta}^p := p!C_{n,p,\beta}^{-1}$ and

$$
V(x) := \frac{n}{2p}x - \frac{\beta(n-p+1)-2}{2p}\ln x \ \sim\ \frac{1}{2}\left\{\frac{x}{\gamma} - \beta\left(\frac{1}{\gamma}-1\right)\ln x\right\}.
$$

The notation "$a_n \sim b_n$" denotes $a_n = (1+o(1))b_n$. Following the definition in (2.6.3) of [1], we further define

$$
\begin{aligned}
Z_{pV/(p-1),\beta}^{p-1} &:= \int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\left|\Delta_{p-1}(\tilde\lambda)\right|^\beta e^{-(p-1)\sum_{i=1}^{p-1}\left\{\frac{p}{p-1}V(\tilde\lambda_i)\right\}}\prod_{i=1}^{p-1}d\tilde\lambda_i\\
&= \int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\prod_{1\le i<j\le(p-1)}\left|\tilde\lambda_i - \tilde\lambda_j\right|^\beta\prod_{i=1}^{p-1}\tilde\lambda_i^{\frac{\beta(n-p+1)}{2}-1}\cdot e^{-\frac{n}{2}\sum_{i=1}^{p-1}\tilde\lambda_i}\prod_{i=1}^{p-1}d\tilde\lambda_i;
\end{aligned}
$$

then the density function (11) implies that the normalization constant $Z^{p-1}_{pV/(p-1),\beta}$ equals

$$Z^{p-1}_{pV/(p-1),\beta} = \left\{ \frac{1}{(p-1)!} \left(\frac{n}{n-1}\right)^{\frac{\beta(n-1)(p-1)}{2}} C_{n-1,p-1,\beta} \right\}^{-1}.$$

With the above notation, Theorem 2.6.6 in [1] states that the large deviation approximation of $\lambda_1 = \max_{i=1}^p \tilde{\lambda}_i$ has speed $p$ and good rate function:

$$I_\beta(s) = \begin{cases} -\beta \int_{\mathbb{R}} \ln|s-t|\sigma_\beta(dt) + V(s) + \alpha_{V,\beta} & \text{if } s \geq s^* \\ \infty & \text{if } s < s^* \end{cases}$$

where $s_* = \beta(1-\sqrt{\gamma})^2$, $s^* = \beta(1+\sqrt{\gamma})^2$, $\sigma_\beta(\cdot)$ is the probability distribution function of the Marchenko-Pastur law specified in (10) and

$$\alpha_{V,\beta} := -\lim_{p\to\infty} \frac{1}{p} \ln \frac{Z^{p-1}_{pV/(p-1),\beta}}{Z^p_{V,\beta}}.$$

A direct calculation gives that for $p/n \to \gamma$,

$$\begin{aligned} \ln \frac{Z^{p-1}_{pV/(p-1),\beta}}{Z^p_{V,\beta}} &\sim \frac{\beta(n+p)}{2}\ln n - \frac{\beta p}{2}\ln p - \frac{\beta n}{2}\ln n - \frac{\beta(p+n)}{2}\cdot(\ln\beta - 1) + O(\ln n) \\ &\sim \frac{\beta}{2}\left\{\gamma \ln\left(\frac{1}{\gamma}\right) - (\gamma+1)(\ln\beta - 1)\right\} n + o(n); \end{aligned}$$

then we obtain $\alpha_{V,\beta} = (\beta/2)\cdot\{\ln\gamma + (1/\gamma+1)(\ln\beta-1)\}$. Therefore, the large deviation approximation of $\lambda_1 = \max_{i=1}^p \tilde{\lambda}_i$ has the rate function:

$$I_\beta(s) = \begin{cases} -\beta \int_{\mathbb{R}} \ln|s-t|\sigma_\beta(dt) + \frac{s}{2\gamma} - \frac{\beta}{2}\left(\frac{1}{\gamma}-1\right)\ln s \\ \quad +\frac{\beta}{2}\left\{\ln\gamma + \left(\frac{1}{\gamma}+1\right)(\ln\beta - 1)\right\} & \text{if } s \geq s^* \\ \infty & \text{if } s < s^*. \end{cases} \tag{16}$$

Recall the notation in Remark 2 and from result in [4], we know when $\mathbf{X}$ has i.i.d. entries $\mathcal{N}(0,1)$ or $\mathcal{CN}(0,1)$, largest eigenvalue $\bar{\lambda}_1$ and the ratio $U_n$ defined in (1) of $n^{-1}\mathbf{X}^H\mathbf{X}$ have the same large deviation approximation function (16). But now in our complex case, $\mathbf{X}$ has i.i.d. entries $\mathcal{CN}(0,2)$ with $\beta = 2$. Similar to argument in Remark 2, since $U_n$ is invariant to this change, we have

$$\begin{aligned} \lim_{n\to\infty} \frac{1}{n} \ln \Pr(U_n > x) &= \lim_{n\to\infty} \frac{1}{n} \ln \Pr(\bar{\lambda}_1 > x) \\ &= \lim_{p\to\infty} \frac{p}{n} \times \frac{1}{p} \ln \Pr(\lambda_1 > \beta x) = -\gamma I_\beta(\beta x). \end{aligned}$$

Therefore we have the large deviation result:

$$n^{-1} \ln \alpha_n(x) \sim -\gamma I_\beta(\beta x). \tag{17}$$

**Step 2**. We focus on the $\ln\{\mathrm{E}_Q(L_p^2)\}$ in this step. Recall that $\sigma_\beta(\cdot)$ in (10) denotes the equilibrium measure for the large deviations of the empirical distribution of eigenvalues $(\lambda_1, \ldots, \lambda_p)$ under $P$; see Lemma 2.6.2 from [1]. Define $t_1$ as a constant such that $t_1 > n/(n-1)$ but close to $n/(n-1)$. Let $B(\epsilon)$ be the ball of probability measures defined on $[0, t_1 M]$ with radius $\epsilon$ around $\sigma_\beta(\cdot)$ under the following metric $\rho$ that generates the weak convergence of probability measures on $\mathbb{R}$: for two probability measures $\mu$ and $\nu$ on $\mathbb{R}$,

$$\rho(\mu,\nu) = \sup_{\|h\|_L \leq 1} \left| \int_{\mathbb{R}} h(x)\mu(dx) - \int_{\mathbb{R}} h(x)\nu(dx) \right|, \tag{18}$$

where $h$ is a bounded Lipschitz function defined on $\mathbb{R}$ with $\|h\| = \sup_{x\in\mathbb{R}} |h(x)|$ and $\|h\|_L = \|h\| + \sup_{x\neq y} |h(x)-h(y)|/|x-y|$. Let $\mathcal{L}^Q_{p-1}$ be the empirical measure of $(\lambda^*_2, \ldots, \lambda^*_p)$ with $(\lambda_2, \ldots, \lambda_p) = \{(n-1)/n\} \times (\lambda^*_2, \ldots, \lambda^*_p)$ being constructed as in *Step 1* of Algorithm 1 under the change of measure $Q$. We know from Marchenko-Pastur law, $\mathcal{L}^Q_{p-1} \to \sigma_\beta(\cdot)$ defined in (10) a.s. Then for a big constant $M$, we have the following upper bound for $\mathrm{E}_Q(L^2_p)$

$$\begin{aligned} \mathrm{E}_Q\big(L^2_p\big) &\leq \mathrm{E}_Q\left\{\left(\frac{dP}{dQ}\right)^2; \lambda_1 > M\right\} \\ &\quad + \mathrm{E}_Q\left\{\left(\frac{dP}{dQ}\right)^2; U_n > x, M > \lambda_1, \mathcal{L}^Q_{p-1} \notin B(\epsilon)\right\} \\ &\quad + \mathrm{E}_Q\left\{\left(\frac{dP}{dQ}\right)^2; U_n > x, M > \lambda_1, \mathcal{L}^Q_{p-1} \in B(\epsilon)\right\} \\ &=: I_1 + I_2 + I_3. \end{aligned} \tag{19}$$

We will show that the first two terms of the above upper bound is ignorable, i.e.,

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln I_1 = -\infty, \tag{20}$$

$$\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln I_2 = -\infty \quad \text{for any } \epsilon > 0. \tag{21}$$

And we will show

$$\lim_{\epsilon\to 0, M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln I_3 = -2\gamma I_\beta(\beta x). \tag{22}$$

Combining (20), (21) and (22) together, we will know

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathrm{E}_Q(L^2_P) \leq -2\gamma I_\beta(\beta x).$$

Then by the result in **Step 1.** of the proof and the fact that $\log\alpha(x) < 0$, we will know

$$\liminf_{n\to\infty} \frac{\ln \mathrm{E}_Q(L^2_p)}{2\ln\alpha_n(x)} \geq 1.$$

Based on the argument above, in the following we only need to prove (20)–(22).

*Proof of* (20). Let $B_{n,p,\beta} := Z^{p-1}_{pV/(p-1),\beta}/Z^p_{V,\beta}$. From the construction of the change of measure $Q$, we can rewrite the left hand side display in (20) as

$$\begin{aligned} &\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln \mathrm{E}_Q\Bigg[\Bigg\{\frac{B_{n,p,\beta}\prod_{i=2}^p(\lambda_1-\lambda_i)^\beta \cdot \lambda_1^{\frac{\beta(n-p+1)}{2}-1} \cdot e^{-\frac{n}{2}\lambda_1}}{nre^{-nr(\lambda_1-\tilde{x}\vee\lambda_2)} \cdot I_{(\lambda_1>\tilde{x}\vee\lambda_2)}}\Bigg\}^2;\ \lambda_1 > M\Bigg] \\ \leq\ &\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln \int_{\substack{\lambda_1>M,\\ \lambda_1>\lambda_2}} r^{-2}n^{-2}B^2_{n,p,\beta}\lambda_1^{\beta(p+n-1)-2}e^{-n\lambda_1+2rn(\lambda_1-\tilde{x}\vee\lambda_2)} \\ &\qquad \times rne^{-rn(\lambda_1-\tilde{x}\vee\lambda_2)} f^Q_{n,p}(\lambda_2, \ldots, \lambda_p)d\lambda_1 d\lambda_2, \ldots, d\lambda_p \\ \leq\ &\lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln \int_{\lambda_1>M} r^{-1}n^{-1}B^2_{n,p,\beta}\lambda_1^{\beta(p+n-1)-2} \cdot e^{-n\lambda_1+rn\lambda_1-rn\tilde{x}}d\lambda_1. \end{aligned}$$

Next we change variable $\lambda_1$ to $\lambda_1 + M$ and since $(\lambda_1 + M)^{\beta(p+n-1)-2} \leq M^{\beta(p+n-1)-2} e^{\{\beta(p+n-1)-2\}\lambda_1/M}$, we obtain the following upper bound for the expectation in (20)

$$\begin{aligned}
& \lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln \int_0^\infty r^{-1} n^{-1} B_{n,p,\beta}^2 M^{\beta(p+n-1)-2} e^{\{\beta(p+n-1)-2\}\lambda_1/M-(n-rn)(\lambda_1+M)-rn\tilde{x}} d\lambda_1 \\
= & \lim_{M\to\infty} \limsup_{n\to\infty} \frac{1}{n} \ln\{B_{n,p,\beta}^2 M^{\beta(p+n-1)-2} e^{-(n-rn)M-rn\tilde{x}}\} + o(1) = -\infty,
\end{aligned}$$

where the last step follows from the approximation of $B_{n,p,\beta}$ from (16). This proves equation (20).

*Proof of* (21). Consider the expectation term in (21). Since $\lambda_1 - \lambda_i < M$ and $\lambda_2 \vee \tilde{x} \geq \tilde{x}$, the following inequality holds for any $\epsilon > 0$,

$$\begin{aligned}
\limsup_{n\to\infty} \frac{1}{n} \ln I_2 \quad \leq \quad & \limsup_{n\to\infty} \frac{1}{n} \ln \mathrm{E}_Q \Big[ \Big\{ \frac{B_{n,p,\beta} M^{\beta(p-1)} \lambda_1^{\frac{\beta(n-p+1)}{2}-1} e^{-\frac{n}{2}\lambda_1}}{rne^{-rn(\lambda_1-\tilde{x})}} \Big\}^2; \\
& U_n > x, M > \lambda_1, \mathcal{L}_{p-1}^Q \notin B(\epsilon) \Big]. \qquad (23)
\end{aligned}$$

Under the assumption that $p/n \to \gamma$, $\lambda_1 < M$ and with the result from (16), we know

$$\frac{B_{n,p,\beta} M^{\beta(p-1)} \lambda_1^{\frac{\beta(n-p+1)}{2}-1} e^{-\frac{n}{2}\lambda_1}}{rne^{-rn(\lambda_1-\tilde{x})}} = e^{O(nM)}.$$

This implies that

$$\begin{aligned}
(23) \quad & \leq \quad \limsup_{n\to\infty} \frac{1}{n} \ln[e^{O(nM)} Q\{U_n > x, M > \lambda_1, \mathcal{L}_{p-1}^Q \notin B(\epsilon)\}] \\
& \leq \quad \limsup_{n\to\infty} \Big[ O(M) + \frac{1}{n} \ln \Pr\{\mathcal{L}_{p-1}^Q \notin B(\epsilon)\} \Big].
\end{aligned}$$

The large deviation result for $\mathcal{L}_{p-1}^Q$ [Theorem 2.6.1 in 1] then gives that

$$\limsup_{n\to\infty} \frac{1}{n^2} \ln \Pr\{\mathcal{L}_{p-1}^Q \notin B(\epsilon)\} = \limsup_{n\to\infty} \frac{(p-1)^2}{n^2} \times \frac{1}{(p-1)^2} \ln \Pr\{\mathcal{L}_{p-1}^Q \notin B(\epsilon)\} < 0.$$

This proves (21).

*Proof of* (22). Define $\Omega_n := \{U_n > x, M > \lambda_1 \text{ and } \mathcal{L}_{p-1}^Q \in B(\epsilon)\}$. We can write

$$I_3 = O(1) n^{-2} B_{n,p,\beta}^2 \mathrm{E}_Q \left\{ e^{2\beta \sum_{i=2}^p \ln(\lambda_1-\lambda_i)} \lambda_1^{\beta(n-p+1)-2} e^{-n\lambda_1} e^{2nr(\lambda_1-\tilde{x}\vee\lambda_2)};\ \Omega_n \right\}.$$

Let $\Phi(z,\epsilon) = \sup_{\mu\in B(\epsilon)} \int \ln(|z-y|)\{\mu(dy) - \sigma_\beta(dy)\}$, we have

$$\begin{aligned}
\sum_{i=2}^p \ln(\lambda_1 - \lambda_i) \quad & = \quad (p-1) \int_{\mathbb{R}} \ln\Big(\frac{n\lambda_1}{n-1} - y\Big) \mathcal{L}_{p-1}^Q(dy) - (p-1) \ln \frac{n}{n-1} \\
& \leq \quad (p-1)\Phi\Big(\frac{n\lambda_1}{n-1}, \epsilon\Big) + (p-1) \int \ln\Big(\frac{n\lambda_1}{n-1} - y\Big) \sigma_\beta(dy) + O(1).
\end{aligned}$$

Under the condition that $\lambda_1 < M$, we know $n\lambda_1/(n-1) < 2M$ when $n$ is big enough. Let $G = \max\left\{\beta(1+\sqrt{\gamma})^2, 2M\right\}$ and define

$$h(x) = x 1_{x\in[0,G]}; \qquad (24)$$

then $h$ is a bounded Lipschitz function. Furthermore, given $\mathcal{L}^Q_{p-1} \in B(\epsilon)$ and under measure $Q$, we have

$$\Big|\frac{1}{p-1}\sum_{i=2}^{p}\frac{n\lambda_i}{n-1}-\beta\Big| = \Big|\int_{\mathbb{R}} h(y)\mathcal{L}^Q_{p-1}(dy) - \int_{\mathbb{R}} h(y)\sigma_\beta(dy)\Big| < O(\epsilon) = o(1),$$

for $\beta = 1$ and 2. This is because from Theorem 6.3.1 in [10], for a distribution with the same density as (13), the first moment is $\mu_{1,\gamma} = \int \bar{s}\times f(\bar{s})d\bar{s} = 1$. For density in (10), similar to Remark 2, the first moment is $\int s\times\sigma_\beta(ds) = \beta\int \bar{s}\times f(\bar{s})d\bar{s} = \beta\times\mu_{1,\gamma} = \beta$. Consider our choice of $G$ in (24),

$$\int_{\mathbb{R}} h(y)\sigma_\beta(dy) = \int_{\mathbb{R}} y\sigma_\beta(dy) = \beta\times\mu_{1,\gamma} = \beta.$$

Therefore, $U_n > x$ and $\lambda_1 > \tilde{x}$ implies that $\lambda_1 > \beta x + O(\epsilon)$ and we can write

$$\begin{aligned} I_3 \quad \leq \quad & O(1)n^{-1}B^2_{n,p,\beta}\int_{\beta x+O(\epsilon)}^{M} e^{2\beta(p-1)\Phi(\frac{n\lambda_1}{n-1},\epsilon)+2\beta(p-1)\int\ln(\frac{n\lambda_1}{n-1}-y)\sigma_\beta(dy)} \\ & \times\lambda_1^{\beta(n-p+1)-2}e^{-n\lambda_1+rn\{\lambda_1-\beta x+O(\epsilon)\}}d\lambda_1. \end{aligned}$$

Since $\beta x + O(\epsilon) < \lambda_1 < M$, we have $\Phi\Big(n\lambda_1/(n-1),\epsilon\Big) \leq \sup_{z\in[n(\beta x+O(\epsilon))/(n-1),nM/(n-1)]}\Phi(z,\epsilon)$ under the constraint $\mathcal{L}^Q_{p-1} \in B(\epsilon)$ and that

$$\begin{aligned} & \int \ln\Big(\frac{n\lambda_1}{n-1}-y\Big)\sigma_\beta(dy) \\ = \quad & \int \ln(\frac{n\beta x}{n-1}-y)\sigma_\beta(dy) + \int \ln\Big(1+\frac{n\lambda_1-n\beta x}{n\beta x-(n-1)y}\Big)\sigma_\beta(dy) \\ \leq \quad & \int \ln(\frac{n\beta x}{n-1}-y)\sigma_\beta(dy) + \int \frac{n\lambda_1-n\beta x}{n\beta x-(n-1)y}\sigma_\beta(dy). \end{aligned}$$

It follows that

$$\begin{aligned} I_3 \quad \leq \quad & O(1)n^{-1}B^2_{n,p,\beta}\times e^{2\beta(p-1)\sup_{z\in[\frac{n(\beta x+O(\epsilon))}{n-1},\frac{nM}{n-1}]}\Phi(z,\epsilon)+2\beta(p-1)\int\ln(\frac{n\beta x}{n-1}-y)\sigma_\beta(dy)} \\ & \times\int_{\beta x+O(\epsilon)}^{M} e^{2\beta(p-1)\int\frac{n\lambda_1-n\beta x}{n\beta x-(n-1)y}d\sigma_\beta(y)}\lambda_1^{\beta(n-p+1)-2}e^{-n\lambda_1+rn\{\lambda_1-x+O(\epsilon)\}}d\lambda_1 \\ = \quad & O(1)n^{-1}B^2_{n,p,\beta}\times e^{2\beta(p-1)\sup_{z\in[\frac{n(\beta x+O(\epsilon))}{n-1},\frac{nM}{n-1}]}\Phi(z,\epsilon)+2\beta(p-1)\int\ln(\frac{n\beta x}{n-1}-y)\sigma_\beta(dy)} \\ & \times\int_{O(\epsilon)}^{M-\beta x} e^{2\beta(p-1)\int\frac{n\lambda_1}{n\beta x-(n-1)y}d\sigma_\beta(y)}\cdot(\lambda_1+\beta x)^{\beta(n-p+1)-2}\cdot e^{-(1-r)n(\lambda_1+\beta x)-rn\{\beta x+O(\epsilon)\}}d\lambda_1 \\ \leq \quad & O(1)n^{-1}B^2_{n,p,\beta}\times e^{2\beta(p-1)\sup_{z\in[\frac{n(\beta x+O(\epsilon))}{n-1},\frac{nM}{n-1}]}\Phi(z,\epsilon)+2\beta(p-1)\int\ln(\frac{n\beta x}{n-1}-y)\sigma_\beta(dy)} \\ & \times(\beta x)^{\beta(n-p+1)-2}e^{-n\{\beta x+O(\epsilon)\}} \\ & \times\int_{O(\epsilon)}^{M-\beta x} e^{2\beta(p-1)\int\frac{n\lambda_1}{n\beta x-(n-1)y}d\sigma_\beta(y)+\{\beta(n-p+1)-2\}\frac{\lambda_1}{\beta x}-(1-r)n\lambda_1}d\lambda_1, \end{aligned} \tag{25}$$

where in the second step we change the variable $\lambda_1$ to $(\lambda_1+\beta x)$ for the integral and in the last step we use $(\lambda_1+\beta x)^{\beta(n-p+1)-2} \leq (\beta x)^{\beta(n-p+1)-2}e^{\{\beta(n-p+1)-2\}\lambda_1/(\beta x)}$.

Under $s^* < \beta x$, we can find a finite number $t_0$ such that $s^* < t_0 x \leq n\{\beta x+O(\epsilon)\}/(n-1)$, for small enough $\epsilon$ and big enough $n$. Recall that $t_1 M \geq nM/(n-1)$. Next we show that

$$\limsup_{\epsilon\to 0}\sup_{z\in[t_0x,t_1M]}\Phi(z,\epsilon) \leq 0. \tag{26}$$

For any $z \in [t_0 x, t_1 M]$ and $\mu \in B(\epsilon)$, let $\mathcal{S}_1(z) = \{y \in \text{supp}(\sigma_\beta) \cup \text{supp}(\mu) : |z-y| > \eta\}$ and $\mathcal{S}_2(z) = \{y \in \text{supp}(\sigma_\beta) \cup \text{supp}(\mu) : |z-y| \le \eta\}$, where $\text{supp}(\mu)$ is the support of measure $\mu$ and $\eta$ is a small constant such that $\eta < \min\{t_0 x - s^*, 1\}$ with $s^*$ defined in (10). Note that $\text{supp}(\sigma_\beta) \subset \mathcal{S}_1(z)$. Given $z \in [t_0 x, t_1 M]$, set $f_z(y) := \ln(|z-y|)$ for $y \in \mathcal{S}_1(z)$. The Lipschitz norms of the set of functions $\{f_z(\cdot); z \in [t_0 x, t_1 M]\}$ on $\mathcal{S}_1(z)$ are bounded by a constant $C < \infty$. By the definition of $\rho(\cdot,\cdot)$ in (18), we obtain

$$\begin{aligned} & \sup_{z\in[t_0x,t_1M]} \int_{\mathbb{R}} \ln(|z-y|)\{\mu(dy) - \sigma_\beta(dy)\} \\ \le\ & \sup_{z\in[t_0x,t_1M]} \int_{\mathcal{S}_1} f_z(y)\{\mu(dy) - \sigma_\beta(dy)\} + \sup_{z\in[t_0x,t_1M]} \int_{\mathcal{S}_2} f_z(y)\mu(dy) \\ \le\ & \sup_{z\in[t_0x,t_1M]} \int_{\mathcal{S}_1} f_z(y)\{\mu(dy) - \sigma_\beta(dy)\} \\ \le\ & C\rho(\mu, \sigma_\beta) < C\epsilon, \end{aligned}$$

for any $\mu \in B_\epsilon$. This implies that $\sup_{z\in[t_0x,t_1M]} \Phi(z,\epsilon) < C\epsilon$. Then (26) follows. When $r < 1 - 2\beta\gamma \int \{1/(\beta x - y)\} d\sigma_\beta(y) - \beta(1-\gamma)/(\beta x)$, we know that the integral term in (25) is $\sim e^{nO(\epsilon)}$. Therefore

$$\begin{aligned} & \lim_{\substack{\epsilon\to 0\\ M\to\infty}} \limsup_{n\to\infty} \frac{1}{n} \ln I_3 \\ =\ & 2\beta\gamma \int \ln(\beta x - y)\sigma_\beta(dy) - \beta x + \beta(1-\gamma)\ln(\beta x) - \beta\{\gamma \ln\gamma + (1+\gamma)(\ln\beta - 1)\} \\ =\ & -2\gamma I_\beta(\beta x), \end{aligned}$$

where $I_\beta(x)$ is defined as in (16). Therefore we conclude

$$\limsup_{n\to\infty} \frac{1}{n} \ln \mathrm{E}_Q(L_p^2) \le -2\gamma I_\beta(\beta x).$$

Hence, the above upper bound and the approximation in (17) imply that

$$\liminf_{n\to\infty} \frac{\ln \mathrm{E}_Q(L_p^2)}{2\ln \alpha_n(x)} \ge 1,$$

where note that $\ln \alpha_n(x) < 0$. This completes the proof.

*Acknowledgement.* The authors thank Professor Boaz Nadler for suggesting this problem. The authors also thank the editor, an associate editor, and two reviewers for many helpful and constructive comments. This work was supported in part by National Security Agency grant H98230-17-1-0308 and National Science Foundation grants DMS-1712717.

## References


[1] G. W. Anderson, A. Guionnet, and O. Zeitouni. *An Introduction to Random Matrices*. Cambridge University Press, 2010.

[2] S. Asmussen and P. Glynn. *Stochastic Simulation: Algorithms and Analysis*. Springer, New York, NY, USA, 2007.

[3] S. Asmussen and D. Kroese. Improved algorithms for rare event simulation with heavy tails. *Advances in Applied Probability*, 38:545–558, 2006.

[4] P. Bianchi, M. Debbah, M. Maïda, and J. Najim. Performance of statistical tests for single-source detection using random matrix theory. *Information Theory, IEEE Transactions on*, 57(4):2400–2419, 2011.

[5] J. Blanchet and P. Glynn. Effcient rare event simulation for the maximum of heavy-tailed random walks. *Annals of Applied Probability*, 18(4):1351–1378, 2008.

[6] M. Chiani. Distribution of the largest eigenvalue for real Wishart and Gaussian random matrices and a simple approximation for the Tracy–Widom distribution. *Journal of Multivariate Analysis*, 129:69–81, 2014.

[7] A. Davis. On the ratios of the individual latent roots to the trace of a Wishart matrix. *Journal of Multivariate Analysis*, 2(4):440–443, 1972.

[8] J. W. Demmel. *Applied Numerical Linear Algebra*. Society for Industrial and Applied Mathematics, Philadelphia, PA, USA, 1997.

[9] I. Dumitriu and A. Edelman. Matrix models for beta ensembles. *Journal of Mathematical Physics*, 43(11):5830–5847, 2002.

[10] I. Dumitriu and A. Edelman. Eigenvalue statistics for beta ensembles. *PhD thesis , Massachusetts Institute of Technology*, 2003.

[11] P. Dupuis, K. Leder, and H. Wang. Importance sampling for sums of random variables with regularly varying tails. *ACM Trans. Model. Comput. Simul.*, 17(3):1–14, 2007.

[12] T. Jiang, K. Leder, and G. Xu. Rare-event analysis for extremal eigenvalues of white Wishart matrices. *The Annals of Statistics*, 45:1609–1637., 2017.

[13] K. Johansson. Shape fluctuations and random matrices. *Communications in mathematical physics*, 209(2):437–476, 2000.

[14] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Annals of statistics*, 29:295–327, 2001.

[15] N. Karoui. A rate of convergence result for the largest eigenvalue of complex white wishart matrices. *The Annals of Probability*, 34(6):2077–2117, 2006.

[16] A. Kortun, M. Sellathurai, T. Ratnarajah, and C. Zhong. Distribution of the ratio of the largest eigenvalue to the trace of complex Wishart matrices. *IEEE Transactions on Signal Processing*, 60(10):5527–5532, 2012.

[17] W. Krzanowski. *Principles of multivariate analysis*. Oxford University Press, 2000.

[18] S. Kuriki and A. Takemura. Tail probabilities of the maxima of multilinear forms and their applications. *The Annals of Statistics*, pages 328–371, 2001.

[19] J. Liu and G. Xu. Efficient simulations for the exponential integrals of Hölder continuous Gaussian random fields. *The ACM Transactions on Modeling and Computer Simulation*, 24:9:1–9:24, 2014.

[20] J. Liu and G. Xu. On the conditional distributions and the efficient simulations of exponential integrals of Gaussian random fields. *Annals of Applied Probability*, 24:1691–1738, 2014.

[21] Z. Ma. Accuracy of the Tracy–Widom limits for the extreme eigenvalues in white Wishart matrices. *Bernoulli*, 18(1):322–359, 2012.

[22] R. J. Muirhead. *Aspects of multivariate statistical theory*, volume 197. John Wiley & Sons, 2009.

[23] B. Nadler. On the distribution of the ratio of the largest eigenvalue to the trace of a Wishart matrix. *Journal of Multivariate Analysis*, 102(2):363–371, 2011.

[24] D. Paul and A. Aue. Random matrix theory in statistics: A review. *Journal of Statistical Planning and Inference*, 150:1–29, 2014.

[25] F. Schuurmann, P. Krishnaiah, and A. Chattopadhyay. On the distributions of the ratios of the extreme roots to the trace of the Wishart matrix. *Journal of Multivariate Analysis*, 3(4):445–453, 1973.

[26] D. Siegmund. Importance sampling in the Monte Carlo study of sequential tests. *Ann. Stat.*, 4:673–684, 1976.

[27] D. Tse and P. Viswanath. *Fundamentals of wireless communication*. Cambridge university press, 2005.

[28] L. Wei, O. Tirkkonen, P. Dharmawansa, and M. McKay. On the exact distribution of the scaled largest eigenvalue. In *Communications (ICC), 2012 IEEE International Conference on*, pages 2422–2426. IEEE, 2012.

[29] G. Xu, G. Lin, and J. Liu. Rare-event simulation for the stochastic Korteweg–de Vries equation. *SIAM/ASA Journal on Uncertainty Quantification*, 2(1):698–716, 2014.