Exact asymptotics of sample-mean related rare-event probabilities

Julia Kuhn, Michel Mandjes, Thomas Taimre

July 19, 2016

Abstract

Relying only on the classical Bahadur-Rao approximation for large deviations of univariate sample means, we derive strong large deviation approximations for probabilities involving two sets of sample means. The main result concerns the exact asymptotics (as $n \to \infty$) of

$$\mathbb{P}\left(\max_{i \in \{1,\ldots,d_X\}} \bar{X}_{i,n} \leqslant \min_{i \in \{1,\ldots,d_Y\}} \bar{Y}_{i,n} \right),$$

with the $\bar{X}_{i,n}$s ($\bar{Y}_{i,n}$s, respectively) denoting d_X (d_Y) independent copies of sample means associated with the random variable X (Y). Assuming $\mathbb{E}\, X > \mathbb{E}\, Y$, this is a rare event probability that vanishes essentially exponentially, but with an additional polynomial term. We also point out how the probability of interest can be estimated using importance sampling in a logarithmically efficient way. To demonstrate the usefulness of the result, we show how it can be applied to compare the order statistics of the sample means of the two populations. This has various applications, for instance in queuing or packing problems.

J. Kuhn and M. Mandjes are with Korteweg-de Vries Institute for Mathematics, University of Amsterdam, Science Park 105, 1098 XH Amsterdam, The Netherlands. J. Kuhn and T. Taimre are with School of Mathematics and Physics, The University of Queensland, St. Lucia, QLD, 407, Australia. *Email*: {J.Kuhn|M.R.H.Mandjes}@uva.nl, t.taimre@uq.edu.au.

1 Introduction

Let the sequence $(X_i)_{i=1}^n$ ($(Y_i)_{i=1}^n$, respectively) consist of i.i.d. samples, all of them distributed as the random variable X (Y, respectively); in addition, the sequences are assumed to be mutually independent. In a broad range of applications including, for example, queuing theory and finance, it is relevant to quantify the behaviour of the probability, for $n \in \mathbb{N}$,

$$\alpha_1(n) := \mathbb{P}\left(\bar{X}_n \leqslant \bar{Y}_n\right),$$

with $\bar{X}_n$ and $\bar{Y}_n$ denoting the sample averages

$$\bar{X}_n := \frac{1}{n}\sum_{j=1}^{n} X_j \quad , \qquad \bar{Y}_n := \frac{1}{n}\sum_{j=1}^{n} Y_j .$$

We assume throughout that $\mathbb{E}\, X < \mathbb{E}\, Y$, which entails that $\alpha_1(n)$ corresponds to a rare event, and therefore vanishes as n grows large. Large deviation (LD) theorems such as Cramér's theorem [6] provide asymptotic expressions that capture the rough (logarithmic, that is) asymptotics of probabilities of this form; in the situation described above they identify a number $I > 0$, usually referred to as the *decay rate*, such that

$$\lim_{n\to\infty} \frac{1}{n} \log \alpha_1(n) = -I. \tag{1}$$

While such results are often useful, they may also turn out to be inaccurate; much information on the asymptotic behaviour of the probability can get lost when the logarithm of the probability is considered rather than the probability itself. To illustrate, note that (1) is valid if $\alpha_1(n)$ behaves as (i) $10^9 \cdot e^{-nI}$, (ii) $n^{-100}e^{-nI}$, or (iii) $e^{\sqrt{n}}e^{-nI}$, but obviously in none of these cases the 'naïve' approximation e^{-nI} is accurate; see e.g. [11, p. 40] for a brief exposition on this. Approximations to the probability itself, rather than its logarithm, are more scarce in the literature, and usually referred to as *strong*, *sharp* or *exact* LD results. An important result on exact LD asymptotics is due to Bahadur and Rao [3]; under some conditions (including the requirement that $X - Y$ has a finite moment generating function in a neighbourhood of the origin), it states that $\alpha_1(n)$ decays as a product of a polynomial and an exponential factor, in that, as $n \to \infty$,

$$\alpha_1(n) \sim \frac{C}{\sqrt{n}} e^{-nI},$$

for positive constants C and I (where $f(n) \sim g(n)$ denotes that $f(n)/g(n) \to 1$ as $n \to \infty$).
A natural next question concerns the context in which there are d independent copies of each of the sample means. More specifically, with $\bar{X}_{1,n}$ up to $\bar{X}_{d,n}$ ($\bar{Y}_{1,n}$ up to $\bar{Y}_{d,n}$, resp.) being i.i.d. copies of $\bar{X}_n$ ($\bar{Y}_n$, resp.), we wish to identify the exact asymptotics of

$$\alpha_d(n) := \mathbb{P}\left(\mathcal{E}_n\right), \quad \text{with } \mathcal{E}_n := \left\{ \max_{i\in\{1,\ldots,d\}} \bar{X}_{i,n} \leqslant \min_{i\in\{1,\ldots,d\}} \bar{Y}_{i,n} \right\}.$$

Some straightforward bounds on $\alpha_d(n)$ can easily be found. It is for instance clear that a necessary condition for $\mathcal{E}_n$ is that $\bar{X}_{i,n} \leqslant \bar{Y}_{i,n}$ for all $i \in \{1, \ldots, d\}$, and hence the independence of the individual sample means implies the following obvious asymptotic upper bound (in self-evident notation):

$$\alpha_d(n) \lesssim \frac{C^d}{n^{d/2}} e^{-ndI}, \tag{2}$$

as $n \to \infty$ (with C and I as above). The main result of the present paper is that we show that (2) is not tight: we prove that, for some $\widetilde{C}_d > 0$, as $n \to \infty$,

$$\alpha_d(n) \sim \frac{\widetilde{C}_d}{n^{d-\frac{1}{2}}} e^{-ndI} \tag{3}$$

(where obviously $\widetilde{C}_1 = C$). The proof relies on careful use of the Bahadur-Rao approximation for all sample means involved.

The exact asymptotics of $\alpha_d(n)$ do not follow from results that have appeared in the literature before, as we point out now. We first observe that the setting introduced above can be cast in a more general framework, involving d^2 sample means. Indeed, with

$$\bar{\boldsymbol{Z}}_n = (\bar{X}_{1,n}, \ldots, \bar{X}_{d,n}, \bar{Y}_{1,n}, \ldots, \bar{Y}_{d,n})^{\mathrm{T}} \in \mathbb{R}^{2d},$$

we can write $\alpha_d(n) = \mathbb{P}(A\bar{\boldsymbol{Z}}_n \geqslant \boldsymbol{0})$, for an appropriately chosen $d^2 \times 2d$ matrix A. Asymptotics of probabilities of the type $\mathbb{P}(A\bar{\boldsymbol{Z}}_n \geqslant \boldsymbol{b})$ are derived (for $\boldsymbol{b} \in \mathbb{R}^{d^2}$), under specific conditions, by Chaganty and Sethuraman in [5]; they typically have the form of a product of a constant, the polynomial function $n^{-d^2/2}$, and a function that decays exponentially in n. Later on in this paper, however, we will verify that for the event of our interest the conditions imposed in [5] are *not* met. (Indeed, the polynomial decay term in our asymptotic form (3) is $n^{-d+1/2}$, rather than the $n^{-d^2/2}$ that one would obtain in the setting of [5].)

We extend the asymptotics of $\alpha_d(n)$ in several ways. In the first place, in Thm. 1 we actually establish a slightly more general version of the above asymptotic equivalence, in which the number of sample means $\bar{X}_{i,n}$, say d_X, does not necessarily coincide with the number of sample means $\bar{Y}_{i,n}$, say d_Y. This result is then easily extended to the case where we consider sample means $\bar{X}_{i,p_i n}$ and $\bar{Y}_{j,q_j n}$ where $p_i\, n$, $q_j\, n \in \mathbb{N}$, see Equation (20). We also provide an importance sampling procedure for estimating such probabilities fast and accurately, and we prove the underlying algorithm is optimal in the sense that it is asymptotically efficient.
In addition, we apply our main result to derive probabilities of practical relevance. More concretely, we obtain an asymptotic expression for the false rejection probability in log-likelihood ratio testing, as well as for the probability of observing at least $k \in \{1, \dots, d\}$ unordered pairs (where the pair (i, j) is said to be unordered if $\bar{X}_{i,n} < \bar{Y}_{j,n}$). The latter can be formulated in terms of a comparison of order statistics, and may, for example, be understood as the probability that at least k jobs cannot be served, or that at least k items cannot be packed.

The paper is organised as follows. In Section 2 we recall some preliminaries (in particular the Bahadur-Rao result) and introduce our notation. Section 3 provides the decay rate of $\alpha_d(n)$, and we explain why this decay rate cannot be obtained from [5]. The result is illustrated by numerical examples, and in this context we also devise an efficient simulation procedure. In Section 4 we apply our main result to compare the order statistics of the sample means, again illustrated by an example. We conclude in Section 5.

2 Preliminaries and notation

In this section we first recall the Bahadur-Rao result and its assumptions. We then describe our setting, as was discussed in the introduction, more formally.

2.1 Bahadur-Rao result

Let $(Z_i)_{i=1}^n$ be a sequence of i.i.d. random variables, distributed as a generic random variable Z. Our results correspond to the light-tailed regime, as formalised in the following assumption.

A1 The moment generating function (mgf) $M_Z(\theta) = \mathbb{E}\, e^{\theta Z}$ is finite in an open set containing the origin.

We now define the *Legendre transform* (also referred to as the Legendre-Fenchel transform, or the convex conjugate) of the logarithm of the mgf. With $\Lambda_Z(\theta) := \log M_Z(\theta)$ denoting the logarithmic mgf (cumulant generating function), we define

$$I_Z(a) := \sup_{\theta \in \mathbb{R}} \left[\theta a - \Lambda_Z(\theta)\right].$$

A2 The optimising θ in the definition of $I_Z(a)$ exists (and is denoted by $\theta_Z(a)$).

It is well-known [6, Lemma 2.2.5] that if $a > \mathbb{E}\, Z$, then $\theta_Z(a) > 0$; likewise, if $a < \mathbb{E}\, Z$, then $\theta_Z(a) < 0$. Furthermore, the optimizing $\theta_Z(a)$ is easily seen to satisfy $I'_Z(a) = \theta_Z(a)$ as well as $\Lambda'(\theta) = a$. These facts we use repeatedly later on.

We now consider the sample mean $\bar{Z}_n := n^{-1}\sum_{i=1}^n Z_i$. We fix an $a > \mathbb{E}\,Z$. The Bahadur-Rao result states that under assumptions **A1–A2**, for some positive and finite constant $C_Z(a)$, as $n \to \infty$,

$$\mathbb{P}(\bar{Z}_n \geqslant na)\,\sqrt{n}\,e^{nI_Z(a)} \to C_Z(a); \tag{4}$$

see e.g. [6, Thm. 3.7.4]. The precise form of $C_Z(a)$ depends on whether Z corresponds to a non-lattice or a lattice random variable. In this paper we focus on the non-lattice case, in which

$$C_Z(a) = \frac{1}{\theta_Z(a)\sqrt{2\pi\Lambda_Z''(\theta_Z(a))}}.$$

The original result of Bahadur and Rao [3] on deviations of the sample mean has been extended in several ways. Notably, there are local versions of it by Petrov [12], as well as results on the uniformity of the convergence by Höglund [9]. A version not necessarily requiring the i.i.d. assumption has been proven by Chaganty and Sethuraman in [4]. This result was further extended into a multi-dimensional context in [5]: there exact asymptotics are established of the probability that a vector of sample means is in a given rectangular set. Further extensions of [5] are found in e.g. [1, 10]; there the set of interest is not necessarily rectangular but can have a more general shape.

As pointed out in the introduction, the rare event studied in this paper can be rewritten in terms of a vector of sample means attaining a value in a given rectangular set, and it may therefore seem that we can use the results from [5]. In Section 3.2, however, we show that in our setting the assumptions imposed in [5] are not satisfied.

2.2 Our model

We now define the setup considered in our main result (stated and proved in the next section). We let $(X_{i,j})_{j=1}^n$ (with $i \in \{1,\ldots,d_X\}$) be independent sequences of i.i.d. random variables $X_{i,j}$, all of them distributed as the generic random variable X. Similarly, for $i \in \{1,\ldots,d_Y\}$ we define the i.i.d. sequences $(Y_{i,j})_{j=1}^n$ with $Y_{i,j} \sim Y$. All sequences are assumed to be mutually independent. Define the sample averages

$$\bar{X}_{i,n} := \frac{1}{n}\sum_{j=1}^n X_{i,j},\ \ i \in \{1,\ldots,d_X\}, \qquad \bar{Y}_{i,n} := \frac{1}{n}\sum_{j=1}^n Y_{i,j},\ \ j \in \{1,\ldots,d_Y\},$$

where we assume $\mathbb{E}X > \mathbb{E}Y$. Let $M_X(\theta) := \mathbb{E}e^{\theta X}$ and $M_Y(\theta) := \mathbb{E}e^{\theta Y}$ denote the moment generating functions of X and Y, respectively, and $\Lambda_X(\theta) := \log M_X(\theta)$ and $\Lambda_Y(\theta) := \log M_Y(\theta)$ the corresponding logarithmic moment generating functions. Assume that **A1–A2** are fulfilled for X and Y.

As indicated previously, in this paper we focus on the non-lattice case, and more specifically, on continuous random variables. We discuss this assumption in Section 3.1.

A3 The distributions of the random variables X and Y are continuous.

We now introduce a number of functions and quantities that are useful in Section 3. In the first place it turns out to be convenient to define

$$a_{d_X,d_Y} := \arg\min_{a\in\mathbb{R}} J_{d_X,d_Y}(a) =: a^\star, \quad J_{d_X,d_Y}(a) := d_X\, I_X(a) + d_Y\, I_Y(a) =: J(a). \tag{5}$$

Note that a_{d_X,d_Y} is guaranteed to exist due to the strong convexity of the Legendre transforms [6, Exercise 2.2.24], and can be seen to lie between $\mathbb{E}Y$ and $\mathbb{E}X$. Note that since $a^\star$ minimizes $J(a^\star)$ it satisfies

$$d_Y\theta_Y(a^\star) = d_Y I_Y'(a^\star) = -d_X I_X'(a^\star) = -d_X\theta_X(a^\star)\,, \tag{6}$$

where $I_X'(a^\star) < 0$ and $I_Y'(a^\star) > 0$, as a consequence of $\mathbb{E}Y < a^\star < \mathbb{E}X$; this 'symmetry' will be useful, particularly in Section 3.2. In addition we will need the function

$$K_{d_X,d_Y}(a) := \big(-C_X(a)\big)^{d_X} C_Y(a)^{d_Y}\, d_Y I_Y'(a) := K(a)\,,$$

with $C_X(a)$ and $C_Y(a)$ as defined in Section 2.1. Note here that $C_X(a) < 0$ and $C_Y(a) > 0$; to see this, bear in mind that $\theta_{-X}(-a) = -\theta_X(a)$.
For our exact asymptotics to hold, we further impose the following regularity condition.

A4 $K_{d_X,d_Y}(a)$ is continuous in $a^\star$, and C_X or C_Y are differentiable in a neighbourhood of $a^\star$.

3 Exact asymptotics

In this section we provide in Thm. 1 the *strong large deviations approximation* of

$$\alpha_{d_X,d_Y}(n) := \mathbb{P}\left(\mathcal{E}_n\right), \quad \text{with } \mathcal{E}_n := \left\{\max_{i\in\{1,\ldots,d_X\}} \bar{X}_{i,n} \leqslant \min_{i\in\{1,\ldots,d_Y\}} \bar{Y}_{i,n}\right\}.$$

This means that our objective is to identify an explicit function $f(n)$ such that $\alpha_{d_X,d_Y}(n) \sim f(n)$ as $n\to\infty$; we say that we thus find the *exact asymptotics* of $\alpha_{d_X,d_Y}(n)$.
The result and its proof are presented in Section 3.1. In Section 3.2 we explain why this result cannot be obtained using the seemingly sufficiently general result [5, Thm. 3.4]. In Section 3.3 we provide two numerical examples featuring normal and Poisson random variables, and point out how these could be estimated efficiently relying on the importance sampling simulation methodology.

3.1 Main result

We first state the main result in Thm 1. It says that $\alpha_{d_X,d_Y}(n)$ decays (roughly) exponentially, where the decay rate is given by $J(a^\star)$ (with $a^\star$ as defined in (5)). The polynomial term is of the power $-(d_X+d_Y)/2+1/2$.

Theorem 1. *Suppose that X and Y fulfil* **A1**–**A3**, *and in addition* **A4** *applies. Then,*

$$\lim_{n\to\infty} \alpha_{d_X,d_Y}(n)\, e^{nJ(a^\star)}\, n^{(d_X+d_Y)/2-1/2} = K(a^\star)\sqrt{\frac{2\pi}{J''(a^\star)}}\,. \tag{7}$$

Proof. Assume first that C_Y is differentiable (which we can do, due to **A4**). Then our starting point is the obvious identity (that is due to conditional independence)

$$\alpha_{d_X,d_Y}(n) = \int_{-\infty}^{\infty} \left(\mathbb{P}\left(\bar{Y}_{1,n} \geqslant a\right)\right)^{d_Y} \mathbb{P}\left(\max_{i\in\{1,\ldots,d_X\}} \bar{X}_{i,n} \in \mathrm{d}a\right). \tag{8}$$

If instead C_X is differentiable, we can start from

$$\alpha_{d_X,d_Y}(n) = \int_{-\infty}^{\infty} \left(\mathbb{P}\left(\bar{X}_{1,n} \leqslant a\right)\right)^{d_X} \mathbb{P}\left(\min_{i\in\{1,\ldots,d_Y\}} \bar{Y}_{i,n} \in \mathrm{d}a\right),$$

then proceed analogously We prove a lower and an upper bound of (8), which asymptotically coincide.

Lower bound: The first step is to just consider the contribution of $a \in (a^\star - \varepsilon, a^\star + \varepsilon)$ in (8), where we choose ε such that $(a^\star - \varepsilon, a^\star + \varepsilon)$ is fully covered in the interval $(\mathbb{E}\, Y, \mathbb{E}\, X)$. The Bahadur-Rao result [3], which holds due to **A1–A2**, entails that for any $\delta > 0$ there is an n_0 such that $\alpha_{d_X,d_Y}(n)$ majorizes for any $n \geq n_0$,

$$(1-\delta)\int_{a^\star-\varepsilon}^{a^\star+\varepsilon}\left(\frac{C_Y(a)}{\sqrt{n}}e^{-nI_Y(a)}\right)^{d_Y}\mathbb{P}\left(\max_{i\in\{1,\dots,d_X\}}\bar{X}_{i,n}\in \mathrm{d}a\right)\;; \tag{9}$$

recall that the convergence in the Bahadur-Rao result holds uniformly [9, 12]. We proceed by applying integration by parts. To this end, first define

$$\begin{aligned} g(a,n) &:= (1-\delta)\left(\frac{C_Y(a)}{\sqrt{n}}e^{-nI_Y(a)}\right)^{d_Y}\mathbb{P}\left(\max_{i\in\{1,\dots,d_X\}}\bar{X}_{i,n}\leqslant a\right) \\ &\sim (1-\delta)\left(\frac{C_Y(a)}{\sqrt{n}}e^{-nI_Y(a)}\right)^{d_Y}\left(-\frac{C_X(a)}{\sqrt{n}}e^{-nI_X(a)}\right)^{d_X}\;; \end{aligned}$$

where the asymptotic equality '$\sim$' again follows from the Bahadur-Rao result.
Applying integration by parts, we find that Expression (9) asymptotically equals the sum of three terms:

$$\begin{aligned} -(1-\delta)\int_{a^\star-\varepsilon}^{a^\star+\varepsilon}\left(-\frac{C_X(a)}{\sqrt{n}}e^{-nI_X(a)}\right)^{d_X} d_Y\left(\frac{C_Y(a)^{d_Y-1}C_Y'(a)-nC_Y(a)^{d_Y}I_Y'(a)}{n^{d_Y/2}}\right)e^{-nd_YI_Y(a)}\mathrm{d}a \\ +\, g(a^\star+\varepsilon,n)-g(a^\star-\varepsilon,n). \end{aligned} \tag{10}$$

Recall that by e.g. [6, Lemma 1.2.15] the decay rate of the sum of three terms equals the largest of the decay rates that correspond to the individual terms. By definition of $a^\star$ and the function $J(\cdot)$, for any $\varepsilon > 0$,

$$\lim_{n\to\infty}\frac{1}{n}\log g(a^\star\pm\varepsilon,n) < -J(a^\star); \tag{11}$$

later on it turns out that the first term in (10) has decay rate $-J(a^\star)$, and hence this means that the second and third term can be asymptotically neglected.
We therefore focus on the first term in (10), which can be checked to be asymptotically equal to

$$(1-\delta)\int_{a^\star-\varepsilon}^{a^\star+\varepsilon}\frac{\big(-C_X(a)\big)^{d_X}C_Y(a)^{d_Y}}{n^{(d_X+d_Y)/2-1}}\,d_YI_Y'(a)\,e^{-nJ(a)}\mathrm{d}a.$$

Now define the convex functions $h_X(a) := I_X(a) - I_X(a^\star)$ and $h_Y(a) := I_Y(a) - I_Y(a^\star)$, which both equal 0 at $a^\star$. We thus find, for n sufficiently large,

$$\alpha_{d_X,d_Y}(n)\,e^{nJ(a^\star)} \geqslant (1-\delta)\int_{a^\star-\varepsilon}^{a^\star+\varepsilon}\frac{\big(-C_X(a)\big)^{d_X}C_Y(a)^{d_Y}}{n^{(d_X+d_Y)/2-1}}\,d_YI_Y'(a)\,e^{-n\left[d_Xh_X(a)+d_Yh_Y(a)\right]}\mathrm{d}a\,. \tag{12}$$

We now study $d_Xh_X(a) + d_Yh_Y(a)$ around $a = a^\star$. Setting up a Taylor expansion of $J(a)$ around $a^\star$, we can find a positive function $\psi(a) = o(a^2)$ such that

$$d_Xh_X(a)+d_Yh_Y(a) \leqslant \frac{1}{2}J''(a^\star)(a-a^\star)^2+\psi(a-a^\star),\quad J''(a^\star) := \frac{\mathrm{d}^2}{\mathrm{d}a^2}J(a)\bigg|_{a=a^\star} > 0\,, \tag{13}$$

where we used that $J(a)$ is convex and minimal at $a^\star$. Defining

$$\kappa(a^\star,\varepsilon) := \inf_{a\in(a^\star-\varepsilon,a^\star+\varepsilon)}K_{d_X,d_Y}(a) = \inf_{a\in(a^\star-\varepsilon,a^\star+\varepsilon)}\big(-C_X(a)\big)^{d_X}C_Y(a)^{d_Y}\,d_YI_Y'(a).$$

and applying the above upper bound (13) on $d_X h_X(a) + d_Y h_Y(a)$, it follows that the right-hand side of (12) majorizes

$$\frac{1-\delta}{n^{(d_X+d_Y)/2-1}}\kappa(a^\star,\varepsilon)\int_{-\varepsilon}^{\varepsilon} e^{-n[\frac{1}{2}J''(a^\star)a^2+\psi(a)]}\mathrm{d}a. \tag{14}$$

To further evaluate the integral in (14), we now apply the transformation $b=\sqrt{nJ''(a^\star)}\,a$ (such that $\mathrm{d}b=\sqrt{nJ''(a^\star)}\,\mathrm{d}a$), so that Expression (14) reads

$$\frac{1-\delta}{n^{(d_X+d_Y)/2-1/2}}\frac{\kappa(a^\star,\varepsilon)}{\sqrt{J''(a^\star)}}\int_{-\varepsilon\sqrt{nJ''(a^\star)}}^{\varepsilon\sqrt{nJ''(a^\star)}} e^{-b^2/2-n\psi(b/\sqrt{nJ''(a^\star)})}\mathrm{d}b.$$

As $n\to\infty$, relying on 'dominated convergence', and recalling that $\psi(a)=o(a^2)$, the integral in the previous display converges to a constant:

$$\int_{-\varepsilon\sqrt{nJ''(a^\star)}}^{\varepsilon\sqrt{nJ''(a^\star)}} e^{-b^2/2-n\psi(b/\sqrt{nJ''(a^\star)})}\mathrm{d}b \to \int_{-\infty}^{\infty} e^{-b^2/2}\mathrm{d}b=\sqrt{2\pi}.$$

Combining this with (12), we have thus found the asymptotic lower bound, as $n\to\infty$,

$$\liminf_{n\to\infty}\alpha_{d_X,d_Y}(n)\,e^{nJ(a^\star)}\,n^{(d_X+d_Y-1)/2}\geqslant(1-\delta)\kappa(a^\star,\varepsilon)\sqrt{\frac{2\pi}{J''(a^\star)}}.$$

Recall that $\delta>0$ and $\varepsilon>0$ were chosen arbitrarily. We thus obtain the lower bound: by letting $\delta\downarrow 0$ and $\varepsilon\downarrow 0$,

$$\liminf_{n\to\infty}\alpha_{d_X,d_Y}(n)\,e^{nJ(a^\star)}\,n^{(d_X+d_Y-1)/2}\geqslant K(a^\star)\sqrt{\frac{2\pi}{J''(a^\star)}}\,, \tag{15}$$

where $K(a^\star):=\lim_{\varepsilon\to 0}\kappa(a^\star,\varepsilon)$ (where we use **A4**).

Upper bound: The upper bound follows by showing that in (8) the contributions corresponding to $a\leqslant a^\star-\varepsilon$ (say $\alpha^-_{d_X,d_Y}(n)$) and $a\geqslant a^\star+\varepsilon$ (say $\alpha^+_{d_X,d_Y}(n)$) are asymptotically negligible; the contribution corresponding to the interval $(a^\star-\varepsilon,a^\star+\varepsilon)$ (say $\alpha^\circ_{d_X,d_Y}(n)$) can be analyzed as in the lower bound, in that it can be verified that, under the assumptions imposed,

$$\limsup_{n\to\infty}\alpha^\circ_{d_X,d_Y}(n)\,e^{nJ(a^\star)}\,n^{(d_X+d_Y-1)/2}\leqslant K(a^\star)\sqrt{\frac{2\pi}{J''(a^\star)}}.$$

Let us focus on $\pi^-_{d_X,d_Y}(n)$, i.e., the contribution corresponding to $(-\infty,a^\star-\varepsilon]$ (as the contribution due to the interval $[a^\star+\varepsilon,\infty)$ can be dealt with precisely analogously); our objective is to prove that its exponential decay rate is strictly smaller than $-J(a^\star)$. For all $\delta>0$ we can find an n_0 such that for $n\geqslant n_0$, $\pi^-_{d_X,d_Y}(n)$ is majorized by

$$(1+\delta)\int_{-\infty}^{\mathbb{E}Y}\mathbb{P}\left(\max_{i\in\{1,\ldots,d_X\}}\bar{X}_{i,n}\in\mathrm{d}a\right)+(1+\delta)\int_{\mathbb{E}Y}^{a^\star-\varepsilon}e^{-nd_YI_Y(a)}\,\mathbb{P}\left(\max_{i\in\{1,\ldots,d_X\}}\bar{X}_{i,n}\in\mathrm{d}a\right); \tag{16}$$

here a Chernoff bound argument is used in the second probability.
We start by considering the first term in (16). Suppressing the factor $(1+\delta)$ for the moment, it can be written as

$$\mathbb{P}\left(\max_{i\in\{1,\ldots,d_X\}}\bar{X}_{i,n}\leqslant\mathbb{E}Y\right)=\left(\mathbb{P}\left(\bar{X}_{i,n}\leqslant\mathbb{E}Y\right)\right)^{d_X}\leqslant e^{-nd_XI_X(\mathbb{E}Y)}.$$

Now observe that

$$d_X I_X(\mathbb{E}Y) = d_X I_X(\mathbb{E}Y) + d_Y I_Y(\mathbb{E}Y) > d_X I_X(a^\star) + d_Y I_Y(a^\star) = J(a^\star).$$

We conclude that the decay rate of the first term of (16) is strictly smaller than $-J(a^\star)$.
We now focus on the second term in (16). Using integration by parts, we obtain that this is smaller than

$$(1+\delta)\left[e^{-nd_Y I_Y(a^\star-\varepsilon)}\mathbb{P}\left(\bar{X}_{1,n} \leqslant a^\star-\varepsilon\right)^{d_X} + \int_{\mathbb{E}Y}^{a^\star-\varepsilon} nd_Y I_Y'(a)e^{-nd_Y I_Y(a)}\mathbb{P}\left(\bar{X}_{1,n} \leqslant a\right)^{d_X}\,\mathrm{d}a\right]. \quad (17)$$

Since the event $\{\bar{X}_{1,n} \leqslant a\}$ is rare for $a \leqslant a^\star - \varepsilon < \mathbb{E}X$, we can apply the Bahadur-Rao result to $\mathbb{P}\left(\bar{X}_{1,n} \leqslant a^\star-\varepsilon\right)^{d_X}$. Then, for large n, the first term in (17) behaves as

$$e^{-nd_Y I_Y(a^\star-\varepsilon)}\left(\frac{-C_X(a^\star-\varepsilon)}{\sqrt{n}}e^{-nI_X(a^\star-\varepsilon)}\right)^{d_X} = e^{-nJ(a^\star-\varepsilon)}n^{d_X/2}(-C_X(a^\star-\varepsilon))^{d_X}\,.$$

Taking the logarithm and dividing by n we see that for large n the decay rate is $-J(a^*-\varepsilon)$, which is smaller than $-J(a^\star)$.
Now consider the second term in (17), which is asymptotically equal to

$$\int_{\mathbb{E}Y}^{a^\star-\varepsilon} n^{1-d_X/2} d_Y I_Y'(a)\big(-C_X(a)\big)^{d_X} e^{-nJ(a)}\mathrm{d}a\,. \quad (18)$$

Since the Legendre transform $J(\cdot)$ is convex, it follows that $J(a) \geqslant J'(a^\star-\varepsilon)(a-a^\star+\varepsilon)+J(a^\star-\varepsilon)$ for any a, and thus (18) is at most

$$e^{-nJ(a^\star-\varepsilon)}\int_{\mathbb{E}Y}^{a^\star-\varepsilon} n^{1-d_X/2} d_Y I_Y'(a)\big(-C_X(a)\big)^{d_X} e^{-nJ'(a^\star-\varepsilon)(a-a^\star+\varepsilon)}\mathrm{d}a\,.$$

Taking the logarithm and dividing by n, we obtain that the decay rate of the second term in (17) is majorised by

$$-J(a^\star-\varepsilon) + \limsup_{n\to\infty}\frac{1}{n}\log\int_{\mathbb{E}Y}^{a^\star-\varepsilon} d_Y I_Y'(a)\big(-C_X(a)\big)^{d_X} e^{-nJ'(a^*-\varepsilon)(a-a^\star+\varepsilon)}\mathrm{d}a\,.$$

Since J is convex and takes its minimum at $a^\star$, the derivate at $a^\star-\varepsilon$ is negative: $J'(a^\star-\varepsilon) < 0$. On $(-\infty, a^\star-\varepsilon]$ we also have $a - a^\star + \varepsilon \leqslant 0$, and hence the exponential is at most 1. Hence,

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{n}\log\int_{\mathbb{E}Y}^{a^\star-\varepsilon} & d_Y I_Y'(a)\big(-C_X(a)\big)^{d_X} e^{-nJ'(a^\star-\varepsilon)(a-a^\star+\varepsilon)}\mathrm{d}a \\ &\leqslant \limsup_{n\to\infty}\frac{1}{n}\log\int_{\mathbb{E}Y}^{a^\star-\varepsilon} d_Y I_Y'(a)\big(-C_X(a)\big)^{d_X}\mathrm{d}a = 0\,.\end{aligned}$$

We conclude that the decay rate of the second term in (17) is smaller than $-J(a^\star-\varepsilon)$.
Combining the above findings, we have established that the asymptotic exponential decay rate of $\alpha^-_{d_X,d_Y}(n)$ is strictly smaller than $-J(a^\star)$ (i.e., the decay rate of $\alpha^\circ_{d_X,d_Y}(n)$).
As we mentioned above, an analogous procedure can be followed for the probability $\alpha^+_{d_X,d_Y}(n)$.
Combining all the above elements, it now follows that an asymptotic upper bound on $\alpha_{d_X,d_Y}(n)$ is given by

$$\limsup_{n\to\infty}\alpha_{d_X,d_Y}(n)\,e^{nJ(a^\star)}\,n^{(d_X+d_Y-1)/2} \leqslant K(a^\star)\sqrt{\frac{2\pi}{J''(a^\star)}}\,. \quad (19)$$

The lower bound (15) and the upper bound (19) together yield the desired result (7). □

We focus here on continuous random variables (see **A3**) because it is otherwise not clear that $a^\star$ as defined in (5) is in the support of $\bar{X}_{i,n}$ and $\bar{Y}_{j,n}$, where the latter quantities depend on n. This, however, is needed to ensure that the lower bound (9) is non-trivial. Note that the result can be expected to hold more generally; clearly the random variables do not need to be continuous *everywhere*, it is sufficient to assume that $\bar{X}_{i,n}$ and $\bar{Y}_{i,n}$ are continuous in a neighbourhood of $a^\star$. Moreover, the result can easily be adapted to the situation in which the individual sample means correspond to different numbers of samples. We find that, for $p_i n,\ q_i n \in \mathbb{N}$, as $n \to \infty$,

$$\begin{aligned}&\mathbb{P}\left(\max_{i\in\{1,\dots,d_X\}} \bar{X}_{i,np_i} \leqslant \min_{i\in\{1,\dots,d_Y\}} \bar{Y}_{i,nq_i}\right) \\ &\quad\sim \frac{(-C_X(a^\star))^{d_X}\, C_Y(a^\star)^{d_Y}}{n^{(d_X+d_Y-1)/2}} \frac{\bar{q}}{\prod_{i=1}^{d_X}\sqrt{p_i}\,\prod_{j=1}^{d_Y}\sqrt{q_j}} I_Y'(a^\star)\sqrt{\frac{\pi}{J''_{\bar{p},\bar{q}}(a^\star)}}\, e^{-nJ_{\bar{p},\bar{q}}(a^\star)}\,,\end{aligned} \tag{20}$$

where now $a^\star = \arg\min_a J_{\bar{p},\bar{q}}(a)$ with $\bar{p} := \sum_{i=1}^{d_X} p_i$, and $\bar{q} := \sum_{i=1}^{d_Y} q_i$. This more general asymptotic relation may be useful in applications, for example those we mention in Section 4.

3.2 Comparison with earlier results

In this subsection we compare the main result, as derived in the previous section, with related results from the literature. If $d_X = d_Y = 1$, then the asymptotics of (3) could also be obtained by applying the Bahadur-Rao result from [3] directly. Therefore, we first verify that indeed our expression coincides with that of Bahadur and Rao in this case.
As mentioned earlier, the event of interest can be written in terms of $d_X d_Y$ inequalities involving the sample means $\bar{X}_{i,n}$ and $\bar{Y}_{j,n}$, which suggests that we can analyse the probability $\alpha_{d_X,d_Y}(n)$ using the results from [5]. In case $d_X > 1$ and $d_Y > 1$, however, we show that one of the conditions imposed in [5] is not fulfilled, entailing that our result is thus new for this case. (If either $d_X = 1$ or $d_Y = 1$, then the result from [5] *does* apply.)

The case $d_X = d_Y = 1$. Consider first the case where $d := d_X = d_Y = 1$. Define the sample mean $\bar{Z}_n := n^{-1}\sum_{j=1}^n (Y_j - X_j)$, and note that from the Bahadur-Rao approximation stated in (4) we have

$$\alpha_{1,1}(n) = \mathbb{P}\left(\bar{Z}_n \geqslant 0\right) \sim \frac{C_Z(0)}{\sqrt{n}} e^{-nI_Z(0)}\,. \tag{21}$$

In order to compare this with (7), we first check that $\theta_Z(0) = \theta_Y(a^\star) = -\theta_X(a^\star)$, where the latter equality holds by (6). We thus have that $\theta_Y(a^\star)$ solves $a^\star - \Lambda_Y'(\theta) = 0$ as well as $-a^\star + \Lambda_X'(-\theta) = 0$. In conclusion, $\theta_Y(a^\star)$ is the unique solution to $\Lambda_Z'(\theta) = \Lambda_Y'(\theta) - \Lambda_X'(-\theta) = 0$, and hence $\theta_Y(a^\star) = \theta_Z(0)$. With this relationship it is now readily checked that $J(a^\star) = I_Z(0)$. Note that

$$\begin{aligned}J''(a) &= a\left[\theta_X''(a) + \theta_Y''(a)\right] + 2\left[\theta_X'(a) + \theta_Y'(a)\right] \\ &\quad - \left[\theta_X''(a)\Lambda_X''\big(\theta_X(a)\big) + \theta_Y''(a)\Lambda_Y''\big(\theta_Y(a)\big) + \theta_X'(a)^2\Lambda_X''\big(\theta_X(a)\big) + \theta_Y'(a)^2\Lambda_Y\big(\theta_Y(a)\big)\right].\end{aligned}$$

Because $\theta'(a) = 1/\Lambda''\big(\theta(a)\big)$ and $\Lambda'\big(\theta(a)\big) = a$, this reduces to

$$J''(a) = \frac{1}{\Lambda_X''\big(\theta_X(a)\big)} + \frac{1}{\Lambda_Y''\big(\theta_Y(a)\big)}\,.$$

We then obtain

$$K(a^\star)\sqrt{\frac{2\pi}{J''(a^\star)}} = -\frac{1}{\theta_X(a^\star)\sqrt{2\pi\Lambda_X''\big(\theta_X(a^\star)\big)}}\frac{1}{\theta_Y(a^\star)\sqrt{2\pi\Lambda_Y''\big(\theta_Y(a^\star)\big)}}\theta_Y(a^\star)\sqrt{\frac{2\pi}{J''(a^\star)}}$$
$$= \frac{1}{\theta_Y(a^\star)\sqrt{2\pi\left[\Lambda_X''\big(\theta_X(a^\star)\big)+\Lambda_Y''\big(\theta_Y(a^\star)\big)\right]}} = C_Z(0)\,.$$

Thus, we conclude that (7) reduces to (21) if $d=1$.

The case $d_X > 1$ and $d_Y > 1$. Now we consider the case that both $d_X > 1$ and $d_Y > 1$ and show that our result does not fall in the framework of [5]. As was already briefly pointed out in the introduction, we can rewrite $\alpha_{d_X,d_Y}(n)$ as $\mathbb{P}(A\bar{\boldsymbol{Z}}_n \geqslant \mathbf{0})$, where

$$\bar{\boldsymbol{Z}}_n = (\bar{X}_{1,n},\ldots,\bar{X}_{d_X,n},\bar{Y}_{1,n},\ldots,\bar{Y}_{d_Y,n})^{\mathrm{T}},$$

and A an appropriately chosen matrix of dimension $d_X d_Y \times (d_X + d_Y)$. In [5, Thm. 3.4] it is proved that, conditional on certain assumptions being satisfied, for positive constants C and I,

$$\mathbb{P}(A\bar{\boldsymbol{Z}}_n \geqslant \mathbf{0}) \sim \frac{C}{n^{(d_X\,d_Y)/2}}e^{-nI}.$$

In Thm. 1 we showed that the polynomial factor in the asymptotics is of the form $n^{-(d_X+d_Y)/2+1/2}$ rather than $n^{-(d_X\,d_Y)/2}$; in this section we show that this seeming inconsistency is due to the fact that [5, Condition (B)] is not met. Observe that if $d_X=1$ or $d_Y=1$ the powers match; we therefore consider the situation that both d_X and d_Y are strictly larger than 1.

Let us first define the multivariate cumulant function. To this end, we write $\overline{W}_{ij,n} = \bar{Y}_{j,n} - \bar{X}_{i,n}$, with $i \in \{1,\ldots,d_X\}$ and $j \in \{1,\ldots,d_Y\}$; observe that the probability of our interest equals $\mathbb{P}(\overline{\boldsymbol{W}}_n \geqslant \mathbf{0})$, where $\overline{\boldsymbol{W}}_n$ is the $d_X d_Y$-vector with entries $W_{ij,n}$. Then the corresponding multivariate moment generating function is given by

$$M(\boldsymbol{\theta}) := \prod_{j=1}^{d_Y}\mathbb{E}\left[e^{Y_j\sum_{i=1}^{d_X}\theta_{i,j}}\right]\prod_{i=1}^{d_X}\mathbb{E}\left[e^{-X_i\sum_{j=1}^{d_Y}\theta_{i,j}}\right],$$

and hence the multivariate cumulant function equals

$$\Lambda(\boldsymbol{\theta}) := \log M(\boldsymbol{\theta}) = \sum_{j=1}^{d_Y}\Lambda_Y\left(\sum_{i=1}^{d_X}\theta_{i,j}\right) + \sum_{i=1}^{d_X}\Lambda_X\left(-\sum_{i=1}^{d_Y}\theta_{i,j}\right).$$

Let $\boldsymbol{\theta}^\star$ solve $\Lambda'(\boldsymbol{\theta}) = \mathbf{0}$; it is readily checked that all $d_X\,d_Y$ entries of $\boldsymbol{\theta}^\star$ are equal (say, have value τ), and solve the equation $\Lambda_Y'\big(d_X\tau\big) = \Lambda_X'\big(-d_Y\tau\big)$. Then [5, Condition (B)] states that the determinant of the Hessian of $\Lambda(\boldsymbol{\theta}^\star)$ should be different from 0. An elementary computation yields that the elements of this Hessian are given by, with $k,\bar{k} \in \{1,\ldots,d_X\}$ and $\ell,\bar{\ell} \in \{1,\ldots,d_Y\}$,

$$\frac{\partial^2\Lambda(\boldsymbol{\theta})}{\partial\theta_{k,\ell}\partial\theta_{\bar{k},\bar{\ell}}} = r_\ell\mathbb{1}\{\ell=\bar{\ell}\} + s_k\mathbb{1}\{k=\bar{k}\},\quad r_\ell := \Lambda_Y''\left(\sum_{i=1}^{d_X}\theta_{i,\ell}\right),\quad s_k := \Lambda_X''\left(-\sum_{j=1}^{d_Y}\theta_{k,j}\right).$$

Let $R(\boldsymbol{\theta}) := \mathrm{diag}\{\boldsymbol{r}\}$ and $S(\boldsymbol{\theta}) := \mathrm{diag}\{\boldsymbol{s}\}$; in addition, E is a $d_X \times d_X$ all-ones matrix, and F a $d_Y \times d_Y$ all-ones matrix. Then we can write the Hessian compactly by

$$H(\boldsymbol{\theta}) = R(\boldsymbol{\theta}) \otimes E + F \otimes S(\boldsymbol{\theta})\,,$$

where $\otimes$ denotes the Kronecker product. Let $\boldsymbol{e}_k$ be the k-th d_X-dimensional unit row vector (i.e., $\boldsymbol{e}_k \in \mathbb{R}^{d_X}$ such that the k-th entry is 1 and all other entries 0). Likewise, $\boldsymbol{f}_\ell$ denotes the ℓ-th d_Y-dimensional unit row vector. Then define, for arbitrary $k \neq \bar{k}$ and $\ell \neq \bar{\ell}$ (which is possible as $d_X \geqslant 2$ and $d_Y \geqslant 2$),

$$\boldsymbol{v} := (\boldsymbol{e}_k \otimes \boldsymbol{f}_\ell) - (\boldsymbol{e}_k \otimes \boldsymbol{f}_{\bar{\ell}}) - (\boldsymbol{e}_{\bar{k}} \otimes \boldsymbol{f}_\ell) + (\boldsymbol{e}_{\bar{k}} \otimes \boldsymbol{f}_{\bar{\ell}}).$$

It is then an elementary computation to conclude that $\boldsymbol{v}H(\boldsymbol{\theta}^\star) = \boldsymbol{0}$, and hence $H(\boldsymbol{\theta}^\star)$ is singular. We conclude that [5, Condition (B)] does not apply.
The intuitive reason for the violation of the condition is that some of the $d_X\, d_Y$ restrictions are essentially redundant. For example, if $d_X = d_Y = 2$, then $\bar{Y}_{1,n} - \bar{X}_{1,n} > 0$ will usually occur by a realisation in which $\bar{Y}_{1,n} \approx \bar{X}_{1,n}$, and similarly for $\bar{Y}_{1,n} - \bar{X}_{2,n} > 0$ and $\bar{Y}_{2,n} - \bar{X}_{1,n} > 0$. Thus, informally speaking, these three conditions boil down to requiring that $\bar{Y}_{1,n} \approx \bar{X}_{1,n} \approx \bar{Y}_{2,n} \approx \bar{X}_{2,n}$. As a consequence, the fourth constraint, i.e., $\bar{Y}_{2,n} - \bar{X}_{2,n} > 0$, is already ensured to hold by the first three conditions with high likelihood. With this line of reasoning it also becomes intuitively clear that we should have $n^{-(d_X+d_Y)/2+1/2}$ as a pre-factor, as we obtained in (7). Informally, [5, Condition (B)] ensures that none of the restrictions imposed by $\overline{\boldsymbol{W}}_n \geqslant \boldsymbol{0}$ is redundant.

The case $d_X = 1$ or $d_Y = 1$. We finally show that in case $d_X = 1$ or $d_Y = 1$ the result from [5] does apply. This can be seen as follows. Let us assume that $d_X = d \geq 1$ and $d_Y = 1$ (the opposite case works analogously). Then by Sylvester's theorem it follows that

$$|H(\boldsymbol{\theta})| = |S(\boldsymbol{\theta})| \left|I + S(\boldsymbol{\theta})^{-1} rE\right| .$$

Note that $S(\boldsymbol{\theta})^{-1}rE$ is a matrix with rows $(r/s_k, \ldots, r/s_k)$. Furthermore, $|S(\boldsymbol{\theta})| = \prod_{k=1}^d s_k$. It can then be checked that

$$|H(\boldsymbol{\theta})| = \sum_{x \in \chi} \prod_{i=1}^{d} x_i \,,$$

where χ denotes the set of all combinations of length d from $\{r, s_1, \ldots, s_d\}$ (hence, $|\chi| = d+1$). Now, inserting $r = \Lambda_Y''\big(d\tau\big)$ and $s_k = \Lambda_X''\big(-\tau\big)$, we obtain that the determinant of $H(\boldsymbol{\theta}^\star)$ is non-zero:

$$|H(\boldsymbol{\theta}^\star)| = d\Lambda_X''\big(-\tau\big)^{d-1}\Lambda_Y''\big(d\tau\big) + \Lambda_X''\big(-\tau\big)^{d} .$$

Invoking (6) we note that $d\tau = \theta_Y(a^\star) = -d\theta_X(a^\star)$. Thus, the result from [5] states that

$$\alpha_{d,1}(n) \sim \frac{1}{(2\pi n)^{d/2}} \left(d\Lambda_X''\big(\theta_X(a^\star)\big)^{d-1}\Lambda_Y''\big(\theta_Y(a^\star)\big) + \Lambda_X''\big(\theta_X(a^\star)\big)^{d}\right)^{-1/2} e^{n\left[d\Lambda_X\left(\theta_X(a^\star)\right)+\Lambda_Y\left(\theta_Y(a^\star)\right)\right]} .$$

This can be checked to be equivalent to the expression given in Thm. 1, using that

$$J_{d,1}''(a^\star) = \frac{d}{\Lambda_X''(\theta_X(a^\star))} + \frac{1}{\Lambda_Y''(\theta_Y(a^\star))} .$$

3.3 Examples and importance sampling

In this subsection we work out two examples with Gaussian and exponentially distributed random variables, respectively. In addition, we point out how to set up a provably asymptotically efficient importance sampling procedure for general random variables satisfying **A1-A2**.

1. *Gaussian.* In this example we let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$, $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$ and fix n. We have

$$\Lambda_{-X}(\theta) = -\theta\mu_X + \frac{1}{2}\sigma_X^2\theta^2\,,$$

so that $\theta_{-X}(-a) = -(a-\mu_X)/\sigma_X^2 = -\theta_X(a)$. It follows directly that

$$I_{-X}(-a) = \frac{1}{2}\left(\frac{a-\mu_X}{\sigma_X}\right)^2 = I_X(a)\,.$$

A similar procedure can be followed for Y. Furthermore, note that $J(a) = d_X I_X(a) + d_Y I_Y(a)$ is minimized by

$$a^\star = \frac{d_X\mu_X\sigma_Y^2 + d_Y\mu_Y\sigma_X^2}{d_X\sigma_Y^2 + d_Y\sigma_X^2}\,;$$

indeed, as we remarked earlier, this quantity lies in the interval (μ_Y, μ_X). We thus arrive at the following expression for the decay rate of $\alpha_{d_X,d_Y}(n)$:

$$J(a^\star) = \frac{d_X d_Y}{2}\frac{(\mu_Y-\mu_X)^2}{d_Y\sigma_X^2 + d_X\sigma_Y^2}\,.$$

For $d_X = d_Y = 1$ and $\sigma_X = \sigma_Y$ this is just the Kullback-Leibler divergence between X and Y. Moreover, note that

$$\theta_{-X}(-a^\star) = -\frac{d_Y(\mu_Y-\mu_X)}{d_X\sigma_Y^2 + d_Y\sigma_X^2}\,,\quad \theta_Y(a^\star) = \frac{d_X(\mu_X-\mu_Y)}{d_X\sigma_Y^2 + d_Y\sigma_X^2}\,,$$

and hence

$$-C_X(a^\star) = -\frac{d_X\sigma_Y^2 + d_Y\sigma_X^2}{d_Y(\mu_Y-\mu_X)\sqrt{2\pi\sigma_X^2}}\,,\quad C_Y(a^\star) = \frac{d_X\sigma_Y^2 + d_Y\sigma_X^2}{d_X(\mu_X-\mu_Y)\sqrt{2\pi\sigma_Y^2}}$$

(which can both be checked to be positive). With $I_Y'(a^\star) = \theta_Y(a^\star)$, we can then compute $K(a^\star)$.

2. *Exponential.* The logarithmic mgf of an exponential random variable with parameter λ is, for $\theta < \lambda$, given by $\Lambda(\theta) = \log\lambda - \log(\lambda-\theta)$, so that (with $\theta(a) = \lambda - 1/a$, assuming $a \neq 0$), $I(a) = \lambda a - 1 - \log(\lambda a)$. For exponential X and Y with $\lambda_X < \lambda_Y$ we thus have

$$J(a) = a\,(d_X\lambda_X + d_Y\lambda_Y) - (d_X + d_Y)\big(\log(a) + 1\big) - d_X\log(\lambda_X) - d_Y\log(\lambda_Y)\,,$$

which is minimal at

$$a^\star = \frac{d_X + d_Y}{d_X\lambda_X + d_Y\lambda_Y}\,.$$

We obtain

$$\theta_{-X}(-a^\star) = -\frac{d_Y(\lambda_X-\lambda_Y)}{d_X + d_Y}\,,\quad \theta_Y(a^\star) = \frac{d_X(\lambda_Y-\lambda_X)}{d_X + d_Y}$$

(thus, indeed $\theta_{-X}(-a^\star) < \lambda_X$ and $\theta_Y(a^\star) < \lambda_Y$, and hence the mgf s are defined at these points). We have

$$C_X(a^\star) = \frac{d_X\lambda_X + d_Y\lambda_Y}{\sqrt{2\pi}d_Y(\lambda_X-\lambda_Y)}\,,\quad C_Y(a^\star) = \frac{d_X\lambda_X + d_Y\lambda_Y}{\sqrt{2\pi}d_X(\lambda_Y-\lambda_X)}$$

and $J(a^\star) = -(d_X + d_Y)\log(a^\star) - d_X\log(\lambda_X) - d_Y\log(\lambda_Y)$.

Our asymptotic results describe how $\alpha_{d_X,d_Y}(n)$ behaves as $n \to \infty$, but do not provide any error bound for a given $n_0 \in \mathbb{N}$. This explains the interest in devising efficient simulation procedures. As is known from the literature [2], direct (naïve) procedures do not work for small probabilities, as the number of experiments needed to obtain an estimate with a given precision (defined as the ratio of the standard error of the estimate and the estimate itself) is roughly inverse proportional to the probability to be estimated. We describe here an importance sampling algorithm that resolves this issue.

Let $f_X(\cdot)$ be the density of X, and $f_Y(\cdot)$ the density of Y. Now associate the alternative probability measure $\mathbb{Q}$ with the system in which the $X_{i,k}$ and $Y_{j,\ell}$ are sampled according to the densities

$$g_X(x) = \frac{e^{\theta_X(a^\star)x}}{M_X(\theta_X(a^\star))} f_X(x), \quad g_Y(y) = \frac{e^{\theta_Y(a^\star)y}}{M_Y(\theta_Y(a^\star))} f_Y(y) .$$

Recall that $a^\star$ minimizes $J(a)$, and therefore solves $-d_X\theta_X(a) = d_Y\theta_Y(a)$ (where it is used that $I'_X(a) = \theta_X(a)$ and $I'_Y(a) = \theta_Y(a)$). It is readily checked that $\mathbb{E}\,X > \mathbb{E}\,Y$ implies that $\theta_X(a) < 0$ and $\theta_Y(a) > 0$.

The idea is to sample all $X_{i,k}$ and $Y_{j,\ell}$ under the newly constructed measure $\mathbb{Q}$, but to weight the simulated output by a likelihood ratio (which can by interpreted as a Radon-Nikodym derivative). We now point out how a single unbiased sample is drawn; to estimate the probability of interest reliably, the average of a number of such samples needs to be taken. The usual change-of-measure argument entails that, in self-evident notation,

$$\alpha_{d_X,d_Y}(n) = \mathbb{E}_{\mathbb{Q}}\left[L\,\mathbb{1}_{\{\mathcal{E}_n\}}\right], \quad \text{where} \quad L := \left(\prod_{i=1}^{d_X}\prod_{k=1}^{n} L_X(X_{i,k})\right)\left(\prod_{j=1}^{d_Y}\prod_{\ell=1}^{n} L_Y(Y_{j,\ell})\right),$$

with the 'per-sample likelihood ratios' defined by

$$L_X(x) = M_X(\theta_X(a^\star))e^{-\theta_X(a^\star)x}, \quad L_Y(y) = M_Y(\theta_Y(a^\star))e^{-\theta_Y(a^\star)y} .$$

We now analyse the variance performance of the resulting estimator. It is said [2, 11, 13] that the latter is asymptotically efficient if it satisfies

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{E}_{\mathbb{Q}}(L^2\,\mathbb{1}_{\{\mathcal{E}_n\}}) \leqslant \limsup_{n\to\infty} \frac{2}{n} \log \mathbb{E}_{\mathbb{Q}}(L\,\mathbb{1}_{\{\mathcal{E}_n\}}) = -2J(a^\star) .$$

To this end, we first rewrite $\mathbb{E}_{\mathbb{Q}}(L^2\,\mathbb{1}_{\{\mathcal{E}_n\}})$ as

$$\big(M_X(\theta(a^\star))\big)^{2nd_X}\big(M_Y(\theta(a^\star))\big)^{2nd_Y}\mathbb{E}_{\mathbb{Q}}\left[e^{-2\theta_X(a^\star)\sum_{i=1}^{d_X}\sum_{k=1}^{n}X_{i,k}}e^{-2\theta_Y(a^\star)\sum_{j=1}^{d_Y}\sum_{\ell=1}^{n}Y_{j,\ell}}\mathbb{1}_{\{\mathcal{E}_n\}}\right] .$$

The next step is to bound, on the event $\mathcal{E}_n$, the exponential term. To this end, note that, on $\mathcal{E}_n$, for all $i \in \{1,\dots,d_X\}$ and $j \in \{1,\dots,d_Y\}$, we have that $\sum_{k=1}^{n} Y_{i,k} \geqslant \sum_{\ell=1}^{n} X_{j,\ell}$. Summing this inequality over all i and j and dividing by d_Xd_Y we obtain, on $\mathcal{E}_n$,

$$\frac{1}{d_X}\sum_{i=1}^{d_X}\sum_{k=1}^{n} X_{i,k} \leqslant \frac{1}{d_Y}\sum_{j=1}^{d_Y}\sum_{\ell=1}^{n} Y_{j,\ell} .$$

It now follows that, recalling that $-d_X\theta_X(a^\star) = d_Y\theta_Y(a^\star)$,

$$
\begin{aligned}
-\theta_Y(a^\star)\sum_{i=1}^{d_Y}\sum_{k=1}^{n} Y_{i,k} &= -d_Y\theta_Y(a^\star)\frac{1}{d_Y}\sum_{i=1}^{d_Y}\sum_{k=1}^{n} Y_{i,k} \\
&\leqslant -d_Y\theta_Y(a^\star)\frac{1}{d_X}\sum_{j=1}^{d_X}\sum_{\ell=1}^{n} X_{j,\ell} \\
&= d_X\theta_X(a^\star)\frac{1}{d_X}\sum_{j=1}^{d_X}\sum_{\ell=1}^{n} X_{j,\ell} = \theta_X(a^\star)\sum_{j=1}^{d_X}\sum_{\ell=1}^{n} X_{j,\ell}\,,
\end{aligned}
$$

from which we conclude that, for any $n \in \mathbb{N}$,

$$
\mathbb{E}_{\mathbb{Q}}\left[e^{-2\theta_X(a^\star)\sum_{i=1}^{d_X}\sum_{k=1}^{n} X_{i,k}}e^{-2\theta_Y(a^\star)\sum_{j=1}^{d_Y}\sum_{\ell=1}^{n} Y_{j,\ell}}\mathbb{1}_{\{\mathcal{E}_n\}}\right] \leqslant 1\,.
$$

This yields the desired inequality:

$$
\begin{aligned}
\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}_{\mathbb{Q}}\left[L^2\,\mathbb{1}_{\{\mathcal{E}_n\}}\right] &\leqslant 2d_X\,\Lambda_X(\theta_X(a^\star)) + 2d_Y\,\Lambda_Y(\theta_Y(a^\star)) \\
&= -2a^\star\left[d_X\theta_X(a^\star) + d_Y\theta_Y(a^\star)\right] + 2d_X\,\Lambda_X(\theta_X(a^\star)) + 2d_Y\,\Lambda_Y(\theta_Y(a^\star)) \\
&= -2J(a^\star)\,.
\end{aligned}
$$

We have thus found the following result.

Proposition 1. *The measure $\mathbb{Q}$ yields an asymptotically efficient procedure for estimating $\alpha_{d_X,d_Y}(n)$.*

In the remainder of this section we examine the accuracy of approximation by the exact asymptotics of $\alpha_{d_X,d_Y}(n)$. With the proposed importance sampling procedure, and inserting the explicit expressions we found for Gaussian and Exponential random variables, we can compare the asymptotic formula given by Thm. 1 to the probabilities as estimated by simulation. Some examples are provided in Fig. 1. The two examples indicate that the approximation tends to be more accurate if (i) d_X and d_Y are smaller or (ii) if the means of X and Y differ more. The former could be a consequence of the additional approximation steps we used compared to Bahadur and Rao in order to extend their result. The latter may be due to the fact that in this case the event is more rare so that the applied LD approximations are more accurate.

4 Applications and further refinements

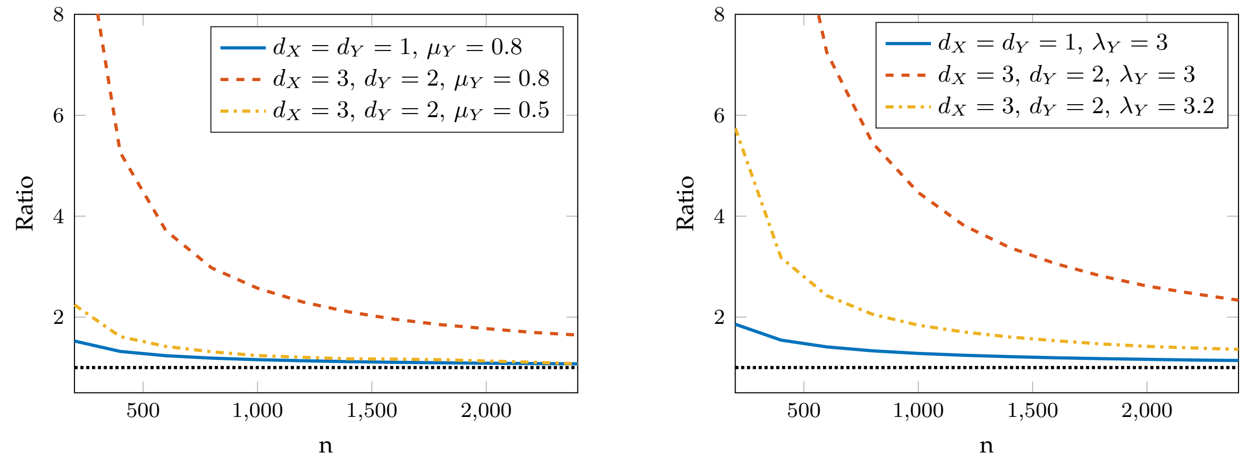
Motivated by specific practical applications, we now study two variants of our main result.

4.1 Probability of at least one sample mean pair not being ordered

It is directly seen that Thm. 1 allows us to conclude that

$$
\mathbb{P}\left(\min_{i\in\{1,\dots,d_X\}}\bar{X}_{i,n} \leqslant \max_{i\in\{1,\dots,d_Y\}}\bar{Y}_{i,n}\right) \sim \sum_{i=1}^{d_X}\sum_{j=1}^{d_Y}\mathbb{P}\left(\bar{Y}_{j,n} - \bar{X}_{i,n} \geqslant 0\right) \tag{22}
$$

because the decay rate corresponding to events $\left\{\bar{Y}_{j,n} - \bar{X}_{i,n} \geqslant 0\right\}$ is $-\inf_a J_{1,1}(a)$, which is larger than the rate functions corresponding to any number of intersections of such events given that



(a) Gaussian random variables, $\mu_Y = 1, \sigma_X = \sigma_Y = 2$.

(b) Exponential random variables, $\lambda_X = 2.8$.

Figure 1: Ratio of the asymptotic expression (7) and simulated probabilities $\alpha_{d_X,d_Y}(n)$. The dotted horizontal line indicates a ratio of 1.

those correspond to $\inf_a J_{i,j}(a)$ for $i + j > 2$. Then the asymptotic relation (22) follows from the inclusion-exclusion principle. It is thus evident that

$$\begin{aligned}\mathbb{P}\left(\min_{i\in\{1,\dots,d_X\}} \bar{X}_{i,n} \leqslant \max_{i\in\{1,\dots,d_Y\}} \bar{Y}_{i,n}\right) &\sim d_X d_Y \mathbb{P}\left(\bar{X}_{1,n} \leqslant \bar{Y}_{1,n}\right) \\ &\sim e^{-nJ_{1,1}(a_{1,1})} \frac{1}{\sqrt{n}} d_X d_Y K_{1,1}(a_{1,1}) \sqrt{\frac{2\pi}{J''_{1,1}(a_{1,1})}}\,.\end{aligned}$$

This probability has applications in log-likelihood ratio (LLR) testing. Note that LLR test statistics take the form of a sample mean. Thus, the probability (22) may be understood as the false classification probability for the problem of discriminating between two populations X and Y.
For a more specific example, suppose d_X signals are sent from an echo sounding system, and in return $d_X + d_Y$ echoes are received, d_Y of which have to thus to be identified as noise. If this echo sounding experiment is carried out n times, the probability of wrongly discarding a signal as noise can be evaluated as a probability of the form (22).
If we relax the assumption that the distributions of X and Y are known (for example, replace the mgf s of X and Y by their maximum likelihood estimators), one may also think of applications in ordinal optimization problems such as stochastic bandit problems, see e.g. [7, 8].

4.2 Existence of at least k unordered sample mean pairs

Denote the order statistics of the sample means of X and Y by $\bar{X}_{(i),n}$ and $\bar{Y}_{(j),n}$; we assume that these order statistics have been put in decreasing order. We here focus on the evaluation of the probability $1 - \mathbb{P}\left(\bar{X}_{(1),n} > \bar{Y}_{(1),n}, \dots, \bar{X}_{(d),n} > \bar{Y}_{(d),n}\right)$, or, more generally (as we can put $k = 1$)

$$\beta_{d,k}(n) := \mathbb{P}\left(\exists i \in \{1, \dots, d-k+1\} : \bar{X}_{(i),n} \leqslant \bar{Y}_{(k+i-1),n}\right), \tag{23}$$

which is the probability that for every bijection mapping the set of indices of $\bar{X}_n$ to the set of indices of $\bar{Y}_n$ there exist at least k unordered pairs (the pair (i, j) is unordered if $\bar{X}_{i,n} < \bar{Y}_{j,n}$).
For a potential application of this type of probability, think of the following packing problem. We have d ships with n containers, and dnc items that need to be packed onto these ships. We assume that the items are separated into d loads (for example, they came from d trucks) of n

batches of c items. The expected capacity of each container is μ_X – the actual capacity is random (for example, it might be that the containers arrive more or less empty than expected). The total observed capacity of ship i is $n\bar{X}_{i,n}$. The items have an expected size of μ_Y/c, so that each batch of c items has an expected size of μ_Y. The total size of load j is $n\bar{Y}_{j,n}$. After observing $n\bar{X}_{i,n}$ and $n\bar{Y}_{j,n}$ each of the d loads needs to be brought to a ship and packed into the containers. In this case the question of whether the full load can be packed (if the batches are assigned carefully) boils down to whether or not there exists a perfect matching of order statistics. More generally we can ask for the probability that at least k loads cannot be packed, which is given by (23).

For another application, suppose that we want to assign memory space/server capacities to serve d batches/queues of jobs. Suppose there are $np \in \mathbb{N}$ jobs in each batch. (As we remarked in (20) it is easy to adapt our results for the case where one of the populations has sample size pn instead of n.) The expected job size/duration is μ_Y. The size of the jobs in batch i amounts to $np\bar{Y}_{i,np}$. Each batch has to be assigned to one of d server pools, each with n servers with expected capacity $\mathbb{E}X$. The actual service capacity of server pool j amounts to $n\bar{X}_{j,n}$. Clearly a quantity of interest is of the form (23), which can be interpreted as the probability that at least k job batches cannot be served.

The main result of this subsection is as follows. It states that the asymptotics of $\beta_{d,k}(n)$ are essentially determined by those of $\alpha_{d_X,d_Y}(n)$.

Proposition 2. *Assume that* **A1-A4** *hold, and in addition that $i^\star$ defined by*

$$i^\star := \arg\min_{i\in\{1,\dots,d-k+1\}} J_{d-i+1,k+i-1}(A_i), \tag{24}$$

with $A_i := a_{d-i+1,k+i-1}$, is unique. Then,

$$\beta_{d,k}(n) \sim \begin{pmatrix} d \\ k+i^\star-1 \end{pmatrix} \begin{pmatrix} d \\ d-i^\star+1 \end{pmatrix} \alpha_{d-i^\star+1,k+i^\star-1}(n)\,. \tag{25}$$

Proof. Define $a^\star := A_{i^\star}$. First, we note that we can write

$$\begin{aligned}&\mathbb{P}\left(X_{(i),n} \leqslant Y_{(i+k-1),n}\right)\\ &\quad= \begin{pmatrix} d \\ k+i-1 \end{pmatrix} \begin{pmatrix} d \\ d-i+1 \end{pmatrix} \mathbb{P}\Bigg(\min_{j\in\{1,\dots,k+i-1\}} \bar{Y}_{j,n} > \max_{j\in\{i,\dots,d\}} \bar{X}_{j,n},\\ &\qquad\qquad \min_{j\in\{1,\dots,k+i-1\}} \bar{Y}_{j,n} \geqslant \max_{j\in\{k+i,\dots,d\}} \bar{Y}_{j,n},\ \max_{j\in\{i,\dots,d\}} \bar{X}_{j,n} \leqslant \min_{j\in\{1,\dots,i-1\}} \bar{X}_{j,n}\Bigg).\end{aligned}$$

The probability on the right-hand side can be computed as

$$\begin{aligned}&\int_{-\infty}^{\infty}\int_{a}^{\infty} \mathbb{P}\left(\max_{j\in\{k+i,\dots,d\}} \bar{Y}_{j,n} \leqslant b\right) \mathbb{P}\left(\min_{j\in\{1,\dots,k+i-1\}} \bar{Y}_{j,n} \in \mathrm{d}b\right)\\ &\qquad\qquad \mathbb{P}\left(\max_{j\in\{1,\dots,i-1\}} \bar{X}_{j,n} \geqslant a\right) \mathbb{P}\left(\max_{j\in\{i,\dots,d\}} \bar{X}_{j,n} \in \mathrm{d}a\right)\,.\end{aligned} \tag{26}$$

We again prove a lower and an upper bound which asymptotically coincide.

Lower bound: A lower bound for (26) is given by

$$\begin{aligned}&\int_{a^\star-\varepsilon}^{a^\star+\varepsilon}\int_{a}^{a^\star+\varepsilon} \mathbb{P}\left(\max_{j\in\{k+i,\dots,d\}} \bar{Y}_{j,n} \leqslant a^\star-\varepsilon\right) \mathbb{P}\left(\min_{j\in\{1,\dots,k+i-1\}} \bar{Y}_{j,n} \in \mathrm{d}b\right)\\ &\qquad\qquad \mathbb{P}\left(\max_{j\in\{1,\dots,i-1\}} \bar{X}_{j,n} \geqslant a^\star+\varepsilon\right) \mathbb{P}\left(\max_{j\in\{i,\dots,d\}} \bar{X}_{j,n} \in \mathrm{d}a\right)\,,\end{aligned}$$

which asymptotically equals

$$\int_{a^\star-\varepsilon}^{a^\star+\varepsilon}\int_a^{a^\star+\varepsilon} \mathbb{P}\left(\min_{j\in\{1,\dots,k+i-1\}} \bar{Y}_{j,n}\in \mathrm{d}b\right)\mathbb{P}\left(\max_{j\in\{i,\dots,d\}} \bar{X}_{j,n}\in \mathrm{d}a\right).$$

This can be rewritten as

$$\int_{a^\star-\varepsilon}^{a^\star+\varepsilon} \mathbb{P}\left(\min_{j\in\{1,\dots,k+i-1\}} \bar{Y}_{j,n}\geqslant a\right)\mathbb{P}\left(\max_{j\in\{i,\dots,d\}} \bar{X}_{j,n}\in \mathrm{d}a\right)$$
$$- \int_{a^\star-\varepsilon}^{a^\star+\varepsilon} \mathbb{P}\left(\min_{j\in\{1,\dots,k+i-1\}} \bar{Y}_{j,n}\geqslant a^\star+\varepsilon\right)\mathbb{P}\left(\max_{j\in\{i,\dots,d\}} \bar{X}_{j,n}\in \mathrm{d}a\right).$$

This lower bound holds for any i; we pick $i=i^\star$. The above expression is asymptotically equal to, with $\bar{d}:=d-i^\star+1$ and $\bar{k}:=k+i^\star-1$,

$$\begin{aligned}
&\alpha_{\bar{d},\bar{k}}(n)-\frac{C_Y(a^\star+\varepsilon)^{\bar{k}}}{n^{\bar{k}/2}}e^{-n\bar{k}I_Y(a^\star+\varepsilon)}\\
&\qquad\times\left[\frac{C_X(a^\star-\varepsilon)^{\bar{d}}}{n^{\bar{d}/2}}e^{-n\bar{d}I_X(a^\star-\varepsilon)}-\frac{C_X(a^\star+\varepsilon)^{\bar{d}}}{n^{\bar{d}/2}}e^{-n\bar{d}I_X(a^\star+\varepsilon)}\right]\\
&=\alpha_{\bar{d},\bar{k}}(n)-\frac{C_Y(a^\star+\varepsilon)^{\bar{k}}}{n^{\bar{k}/2}}e^{-n\,J_{\bar{d},\bar{k}}(a^\star+\varepsilon)}\\
&\qquad\times\left[\frac{C_X(a^\star-\varepsilon)^{\bar{d}}}{n^{\bar{d}/2}}e^{n\bar{d}\left[I_X(a^\star+\varepsilon)-I_X(a^\star-\varepsilon)\right]}-\frac{C_X(a^\star+\varepsilon)^{\bar{d}}}{n^{\bar{d}/2}}\right]
\end{aligned}$$

where $\alpha_{\bar{d},\bar{k}}(n)$ is as in Section 3.
Recall that the exponential term in $\alpha_{\bar{d},\bar{k}}(n)$ is $J_{\bar{d},\bar{k}}(a^\star)$. Since $a^\star$ minimizes $J_{\bar{d},\bar{k}}(a)$, we have that $\exp(-nJ_{\bar{d},\bar{k}}(a^\star))$ asymptotically dominates $\exp(-nJ_{\bar{d},\bar{k}}(a^\star+\varepsilon))$. Furthermore, recall that $a^\star<\mathbb{E}X$, and therefore $I_X(a^\star+\varepsilon)-I_X(a^\star-\varepsilon)<0$. We thus conclude that the lower bound is asymptotically equal to $\alpha_{\bar{d},\bar{k}}(n)$.

Upper bound: We can simply replace probabilities in (26) by one to obtain that

$$\int_{-\infty}^{\infty}\int_a^{\infty} \mathbb{P}\left(\min_{j\in\{1,\dots,k+i-1\}} \bar{Y}_{j,n}\in \mathrm{d}b\right)\mathbb{P}\left(\max_{j\in\{i,\dots,d\}} \bar{X}_{j,n}\in \mathrm{d}a\right)$$

is an upper bound for (26). Since this is equal to

$$\int_{-\infty}^{\infty} \mathbb{P}\left(\min_{j\in\{1,\dots,k+i-1\}} \bar{Y}_{j,n}\geqslant a\right)\mathbb{P}\left(\max_{j\in\{i,\dots,d\}} \bar{X}_{j,n}\in \mathrm{d}a\right),$$

the results of Section 3 state that an upper bound is given by $\alpha_{d-i+1,k+i-1}(n)$, which thus coincides with the lower bound. We thus find, asymptotically,

$$\mathbb{P}\left(X_{(i),n}\leqslant Y_{(k+i-1),n}\right)\sim\binom{d}{k+i-1}\binom{d}{d-i+1}\alpha_{d-i+1,k+i-1}(n)\,. \tag{27}$$

It now follows that

$$\beta_{d,k}(n)\lesssim\sum_{i=1}^{d-k+1}\binom{d}{k+i-1}\binom{d}{d-i+1}\alpha_{d-i+1,k+i-1}(n)\,.$$

Asymptotically what matters is the dominating summand given by $i^\star$ as defined in (24); as $n\to\infty$ the other summands are asymptotically negligible (under the uniqueness assumption that we imposed). Since every single summand is a lower bound for $\beta_{d,k}(n)$, we then have the asymptotic relation (25). $\square$

If there is no unique optimiser $i^\star$, we have proven the asymptotic upper bound

$$\beta_{d,k}(n) \leqslant \sum_{i\in\mathcal{I}^*} \begin{pmatrix} d \\ k+i-1 \end{pmatrix} \begin{pmatrix} d \\ d-i+1 \end{pmatrix} \alpha_{d-i+1,k+i-1}(n)\,, \tag{28}$$

where $\mathcal{I}^*$ denotes the set of optimising $i \in \{1,\ldots,d-k+1\}$. Furthermore, every summand of the right-hand side is an asymptotic lower bound.

One may now wonder whether the upper bound in (28) is asymptotically tight. Observe that the inequality in (28) is essentially a Bonferroni inequality, and one might expect that probabilities of intersections of the corresponding events are asymptotically negligible, in which case by the inclusion-exclusion principle the upper bound would be asymptotically tight (similar to the argument we gave in Section 4.1). The following heuristic argument indicates, however, that this reasoning is not valid in this case, and this is confirmed numerically in the example provided below.

In our example we consider the simplest case possible: we suppose that $d = 2$ and $k = 1$. We define the events

$$E_{i,j} := \left\{\bar{X}_{i,n} \leqslant \bar{Y}_{j,n}\right\}, \quad F_{i,j} := \left\{\bar{X}_{(i),n} \leqslant \bar{Y}_{(j),n}\right\},$$

where $i, j \in \{1, 2\}$. We have

$$\beta_{2,1}(n) = \mathbb{P}\left(F_{1,1} \cup F_{2,2}\right) = \mathbb{P}\left(F_{1,1}\right) + \mathbb{P}\left(F_{2,2}\right) - \mathbb{P}\left(F_{1,1} \cap F_{2,2}\right)\,.$$

It is directly verified that

$$\mathbb{P}(E_{1,1} \cap E_{1,2}) = \mathbb{P}(E_{2,1} \cap E_{2,2}),\;\; \mathbb{P}(E_{1,1} \cap E_{2,1}) = \mathbb{P}(E_{1,2} \cap E_{2,2}),\;\; \mathbb{P}(E_{1,1} \cap E_{2,2}) = \mathbb{P}(E_{1,2} \cap E_{2,1})\,.$$

Furthermore, relying on arguments similar to those used in Section 3.2, it follows that some events essentially imply each other, in that

$$\mathbb{P}(E_{1,1} \cap E_{2,1} \cap E_{2,2}) \approx \mathbb{P}(E_{1,1} \cap E_{2,2})\,, \quad \mathbb{P}(E_{1,2} \cap E_{2,1} \cap E_{2,2}) \approx \mathbb{P}(E_{1,2} \cap E_{2,1})\,, \ldots$$

and analogously for other probabilities of this form. Based on these findings, and applying elementary set theory, we have that $\mathbb{P}\left(F_{1,1}\right) + \mathbb{P}\left(F_{2,2}\right)$ behaves as

$$\left[2\,\mathbb{P}(E_{1,1} \cap E_{1,2}) - \mathbb{P}(E_{1,1} \cap E_{2,2})\right] + \left[2\,\mathbb{P}(E_{1,1} \cap E_{2,1}) - \mathbb{P}(E_{1,1} \cap E_{2,2})\right], \tag{29}$$

whereas $\mathbb{P}\left(F_{1,1} \cap F_{2,2}\right) \approx \mathbb{P}(E_{1,1} \cap E_{2,2})$. We conclude that this probability is thus not negligible compared to (29), and as a consequence (28) is not asymptotically tight.

Gaussian example. We consider again the example with X and Y both being normally distributed, as introduced in Section 3.3. First, assume that $\sigma_X \neq \sigma_Y$. Define a differentiable function $h : \mathbb{R} \to \mathbb{R}$ by

$$h(x) := \frac{1}{2}\frac{(d-x+1)(k+x-1)(\mu_Y - \mu_X)^2}{(k+x-1)\sigma_X^2 + (d-x+1)\sigma_Y^2}\,.$$

As can be checked by an explicit calculation, we have that $h''(x) < 0$, and hence $h(\cdot)$ is concave. Note that for $i \in \{1,\ldots,d-k+1\}$ we have $J_{d-i+1,k+i-1}(A_i) = h(i)$. We conclude that $J_{d-i+1,k+i-1}(A_i)$ is concave as a function of $i \in \{1,\ldots,d-k+1\}$, and thus takes its minima at the boundaries, that is, for $i \in \{1, d-k+1\}$. A straightforward calculation reveals that $J_{d-i+1,k+i-1}(A_i)$ is minimized at $i^\star = 1$ if $\sigma_Y > \sigma_X$, and at $i^\star = d-k+1$ otherwise.

Now consider the case $\sigma_X = \sigma_Y$. Then the function $h(\cdot)$ simplifies:

$$J_{d-i+1,k+i-1}(A_i) = \frac{(d-i+1)(k+i-1)}{2\sigma^2}\frac{(\mu_Y-\mu_X)^2}{d+k}\,.$$

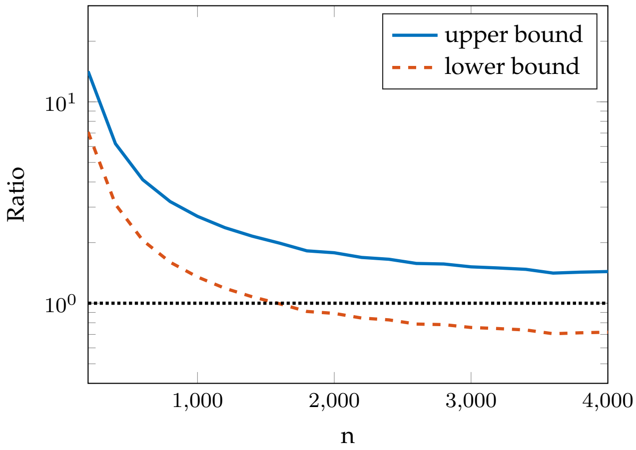


Figure 2: Ratio of the asymptotic bounds (30) and simulated probabilities $\beta_{d,k}(n)$, where $d = 3$, $k = 2$, for Gaussian random variables with $\mu_X = 1$, $\mu_Y = 0.8$, $\sigma = 2$. The dotted horizontal line indicates a ratio of 1.

As before, this concave function can attain its minimum value only at the boundary points $i \in \{1, d-k+1\}$, but note that at these points the function value is the same. Hence, from (28) we have

$$\binom{d}{k} \alpha_{d,k}(n) \leqslant \beta_{d,k}(n) \leqslant \binom{d}{k} \left[\alpha_{d,k}(n) + \alpha_{k,d}(n)\right]. \tag{30}$$

Numerical experiments such as Fig. 2 seem to confirm that these bounds are not tight, as was argued earlier in this subsection.

5 Concluding remarks

We have derived exact asymptotics for the rare event probability that all sample means of a population Y exceed all sample means of an independent population X while $\mathbb{E}X > \mathbb{E}Y$. The proof heavily relies on Bahadur-Rao type asymptotics that describe the tail distribution of the sample mean of i.i.d. random variables. Our result is new: it seemingly fits in the framework of [5], but careful inspection shows that the conditions imposed in [5] are not met in our situation (and we do obtain a different asymptotic form than that suggested in [5], with the polynomial factor being $n^{-d+1/2}$, rather than $n^{-d^2/2}$). We also provide an asymptotically efficient importance sampling procedure for estimating the probability of our interest.

We then showed that this result yields an expression for the exact asymptotics of the probability that there exists a sample mean from Y that exceeds a sample mean from X, and pointed out the relevance in log-likelihood ratio testing. We also used our result to derive the probability that there are at least k unordered sample means in every possible matching of sample means between X and Y; we explained that this probability may be of practical interest for example in particular queuing or packing problems.

Acknowledgments

Michel Mandjes' research is partly funded by the NWO Gravitation Project NETWORKS, grant 024002003. Julia Kuhn is supported by the Australian Research Council (ARC), grant DP130100156.

References

[1] C. Andriani and P. Baldi. Sharp estimates of deviations of the sample mean in many dimensions. *Annales de l'Institut Henri Poincare (B) Probability and Statistics*, 33(3):371 – 385, 1997.

[2] S. Asmussen and P. W. Glynn. *Stochastic simulation: algorithms and analysis*. Stochastic Modelling and Applied Probability. Springer, New York, 2007.

[3] R. R. Bahadur and R. R. Rao. On deviations of the sample mean. *Ann. Math. Statist*, 31(4):1015–1027, 1960.

[4] N. R. Chaganty and J. Sethuraman. Strong large deviation and local limit theorems. *The Annals of Probability*, 21(3):1671–1690, 1993.

[5] N. R. Chaganty and J. Sethuraman. Multidimensional strong large deviation theorems. *Journal of Statistical Planning and Inference*, 55(3):265 – 280, 1996.

[6] A. Dembo and O. Zeitouni. *Large Deviations Techniques and Applications*. Springer-Verlag, New York, 2 edition, 1998.

[7] P. Glynn and S. Juneja. A large deviations perspective on ordinal optimization. In *Proceedings of the 2004 Winter Simulation Conference*, volume 1. IEEE, 2004.

[8] P. Glynn and S. Juneja. Ordinal optimization-empirical large deviations rate estimators, and stochastic multi-armed bandits. *arXiv preprint arXiv:1507.04564*, 2015.

[9] T. Höglund. A unified formulation of the central limit theorem for small and large deviations from the mean. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 49(1):105–117, 1979.

[10] M. Iltis. Sharp asymptotics of large deviations in $\mathbb{R}^d$. *Journal of Theoretical Probability*, 8(3):501–522, 1995.

[11] M. Mandjes. *Large Deviations for Gaussian Queues*. John Wiley & Sons, Ltd, Chichester, UK, 2007.

[12] V. V. Petrov. *Sums of Independent Random Variables*. Springer Verlag, New York, 1975.

[13] J. S. Sadowsky and J. A. Bucklew. On large deviations theory and asymptotically efficient Monte Carlo estimation. *IEEE Transactions on Information Theory*, 36(3):579–588, 1990.