Random assignment with integer costs

Robert Parviainen
Department of Mathematics, Uppsala University
P.O. Box 480, SE-75106 Uppsala, Sweden
robert.parviainen@math.uu.se

June 14, 2001

Abstract

The random assignment problem is to minimize the cost of an assignment in a $n \times n$ matrix of random costs. In this paper we study this problem for some integer valued cost distributions. We consider both uniform distributions on $1, 2, \ldots, m$, for $m = n$ or n^2, and random permutations of $1, 2, \ldots, n$ for each row, or of $1, 2, \ldots, n^2$ for the whole matrix. We find the limit of the expected cost for the n^2 cases, and prove bounds for the n cases. This is done by simple coupling arguments together with Aldous recent results for the continuous case. We also present a simulation study of these cases.

1 Introduction

In the assignment problem we are to choose n elements from a $n \times n$ matrix $\boldsymbol{C}$ of costs, one element from each row and each column, in such a way that the total cost is minimized. In other words, we are looking for a permutation π, that minimizes

$$Z = \sum_{i=1}^{n} C_{i\pi(i)}.$$

If we let the elements of $\boldsymbol{C}$ be random variables, we have the random assignment problem. Traditionally, the random costs have been independent, identically distributed, with the exponential or the uniform distribution.

When the costs are i.i.d. exponential (mean 1) there are strong conjectures for the more general case of k-assignment from a $m \times n$ cost matrix. Let $Z^*(k,m,n)$ denote the minimal cost. Mézard and Parisi [8], [9], conjectured that

$$\lim_{n\to\infty} E(Z^*(n,n,n)) = \pi^2/6.$$

This was proven by Aldous [1]. Parisi [11] has also conjectured that

$$E(Z^*(n,n,n)) = \sum_{i=1}^{n} \frac{1}{i^2},$$

which was improved by Coppersmith and Sorkin [3] to

$$E(Z^*(k,m,n)) = \sum_{i+j<k} \frac{1}{(m-i)(n-j)}.$$

The last conjecture was proven by Alm and Sorkin [2] for $k \leq 4, k = m = 5$, and $k = m = n = 6$. Linusson and Wästlund [7] extended this to $k \leq 6$, and $k = m = n = 7$.

1.1 Discrete variants

We will study four discrete variants of the random assignment problem.

Case I Each row in $\boldsymbol{C}$ is an independent random permutation of $\{1, 2, \ldots, n\}$, chosen uniformly from the set of all permutations.

Case II Each element in $\boldsymbol{C}$ is an independent random number, chosen uniformly from $\{1, 2, \ldots, n\}$.

Case III $\boldsymbol{C}$ is a random permutation of $\{1, 2, \ldots, n^2\}$ chosen uniformly.

Case IV Each element in $\boldsymbol{C}$ is an independent random number, chosen uniformly from $\{1, 2, \ldots, n^2\}$.

In the first two cases we normalize by n, and in cases III and IV by n^2, thus considering the problem of minimizing

$$Z = \frac{1}{n} \sum_i C_{i\pi(i)} \text{ or } Z = \frac{1}{n^2} \sum_i C_{i\pi(i)}.$$

The (random) minimal costs will be denoted by $Z_i^*(n)$, for the four discrete cases, and by $Z_c^*(n)$ in the case of continuous costs.

In [1], Aldous proves the following theorems, valid for any non-negative continuous distribution, such that the density of the independent costs have value 1 at 0. Let π denote the permutation giving an optimal assignment.

Theorem 1.1.

$$\lim_{n \to \infty} EZ_c^*(n) = \frac{\pi^2}{6}.$$

Theorem 1.2. *$nC_{i\pi(i)}$ converges in distribution. The limit distribution has density*

$$h(x) = \frac{e^{-x}(e^{-x} - 1 + x)}{(1 - e^{-x})^2}, \quad 0 \leq x < \infty.$$

Theorem 1.3.

$$\lim_{n \to \infty} P(C_{i\pi(i)} \text{ is the } k\text{th smallest element of the } i\text{th row in } \mathbf{C}) = 2^{-k}.$$

Remark. In a simulation study in [10], Olin noted that, even for as small dimensions as $n = 50$, the row rank distribution is surprisingly close to the above.

2 Coupling arguments

In this section we will prove the following theorem.

Theorem 2.1. *Let* $EZ_i^* = \lim_{n\to\infty} EZ_i^*(n)$*. Then*

$$\frac{\pi^2}{6} \le EZ_1^* \le 2, \quad \frac{\pi^2}{6} + \frac{12}{24} \le EZ_2^* \le \frac{\pi^2}{6} + \frac{13}{24},$$
$$EZ_3^* = \frac{\pi^2}{6}, \quad EZ_4^* = \frac{\pi^2}{6}.$$

The idea is to compare a discrete case of the problem with the case of (continuous) uniform costs. We want to generate matrices for both cases simultaneously, such that an optimal assignment for one matrix is close to optimal for the other.

When we say that π is optimal for the matrix $\boldsymbol{C}$, we mean that π is a permutation giving an optimal assignment for the random assignment problem, with cost matrix $\boldsymbol{C}$.

2.1 Case I

Let $\boldsymbol{U}$ be a $n \times n$ matrix of i.i.d. $U(0,n)$ (uniform on $(0,n)$) random variables. It will be convenient to denote the rows of $\boldsymbol{U}$ by $\boldsymbol{U}^{(i)}$. We want to use $\boldsymbol{U}$ to get an independent random permutation for each row. To achieve this, we can use the row ranks of the matrix $\boldsymbol{U}$. If we let

$$P_i(j) = \text{rank } U_j^{(i)},$$

each P_i will be an independent random permutation, chosen uniformly from the set of all permutations. By Theorem 1.3 we have,

$$\lim_{n\to\infty} P(\text{rank } U_j^{(i)} = k) = 2^{-k}.$$

This gives, if π is the optimal assignment for $\boldsymbol{U}$,

$$\lim_{n\to\infty} EZ_1^*(n) \le \lim_{n\to\infty} E\left(\frac{1}{n}\sum_{i=1}^{n} P_i(\pi(i))\right) = \lim_{n\to\infty} E(\text{rank } U_j^{(i)}) = 2.$$

For a lower bound, assume that, for $1 \le i \le n$, P_i is a random permutation of $\{1,2,\ldots,n\}$, and that $\boldsymbol{V}$ is a $n \times n$ matrix with i.i.d. $U(0,n)$ random variables as elements. We will now use the permutations P_i to rearrange the rows of $\boldsymbol{V}$. This will give us another matrix, $\boldsymbol{U}$, also with i.i.d. $U(0,n)$ elements, such that U_{ij} is close to $P_i(j)$. To be precise, let

$$U_j^{(i)} = V_{(P_i(j))}^{(i)} = \text{ the } P_i(j)\text{th smallest element in row } i \text{ of } \boldsymbol{V},$$

and note that $E(k - V_{(k)}^{(i)}) = k - nk/(n+1) = k/(n+1)$, since $V_{(k)}^{(i)}/n$ is $Beta(n+k+1,k)$. We therefore have, for all permutations π,

$$E(P_i(\pi(i)) - U_{\pi(i)}^{(i)}) > 0.$$

Now assume that π is an optimal assignment for the discrete problem. The cost can then be bounded below by the cost of the problem with cost matrix $\boldsymbol{U}$:

$$EZ_1^*(n) = E\left(\frac{1}{n}\sum_{i=1}^{n} P_i(\pi(i))\right) = E\left(\frac{1}{n}\sum_{i=1}^{n}\left(U_{\pi(i)}^{(i)} + P_i(\pi(i)) - U_{\pi(i)}^{(i)}\right)\right) >$$

$$> E\left(\frac{1}{n}\sum_{i=1}^{n} U_{\pi(i)}^{(i)}\right) \geq EZ_c^*(n) \to \frac{\pi^2}{6}.$$

2.2 Case II

Let U_{ij} be i.i.d. $U(0,n)$. To get i.i.d. random variables from the discrete uniform distribution on $\{1,2,\ldots,n\}$, we can simply take the integer part of U_{ij} and add 1. Let

$$Y_{ij} = [U_{ij}] + 1,$$

where $[x]$ denotes the integer part of x. Then Y_{ij} are i.i.d. with the desired distribution, and the differences $Y_{ij} - U_{ij}$ are uniform on $(0,1)$. Assume that π is an optimal assignment for $\boldsymbol{Y}$. Then we still have

$$Y_{i\pi(i)} - U_{i\pi(i)} \in U(0,1) \text{ and } E(Y_{i\pi(i)} - U_{i\pi(i)}) = 1/2,$$

and for the lower bound of $EZ_2^*(n)$,

$$EZ_2^*(n) = E\left(\frac{1}{n}\sum_{i=1}^{n} Y_{i\pi(i)}\right) = E\left(\frac{1}{n}\sum_{i=1}^{n} U_{i\pi(i)}\right) + \frac{1}{2} \geq EZ_c^*(n) + \frac{1}{2}.$$

Now for the other direction. Assume that π is the optimal assignment for $\boldsymbol{U}$. Svante Janson [5] has calculated the expectation of the fractional part of one element in the optimal assignment, $\{U_{i\pi(i)}\} = U_{i\pi(i)} - [U_{i\pi(i)}]$, with respect to the limit distribution, given by Theorem 1.2.

$$\lim_{n\to\infty} E(U_{i\pi(i)} - [U_{i\pi(i)}]) = \int_0^\infty \{x\}h(x)dx = \frac{1}{2} - \frac{1}{24} + \sum_{k=1}^{\infty} \frac{\pi^2}{\sinh^2(2\pi^2 k)} = \frac{11}{24} + c,$$

where $c \approx 2.83 \cdot 10^{-16}$. Let Z_2^π be the cost of $\boldsymbol{Y}$ given by the assignment π.

$$\lim_{n\to\infty} E(Z_c^* - Z_2^\pi) = \lim_{n\to\infty} E\left(\frac{1}{n}\sum_{i=1}^{n} U_{i\pi(i)} - Y_{i\pi(i)}\right)$$

$$= \lim_{n\to\infty} E(U_{1\pi(1)} - [U_{1\pi(1)}]) - 1 = -\frac{13}{24} + c.$$

Since $Z_2^* \leq Z_2^\pi$, we get the upper bound

$$EZ_2^* \leq \frac{\pi^2}{6} + \frac{13}{24}.$$

2.3 Case III

This is similar to the first case, but for ease of notation we consider a vector of n^2 elements instead of a $n \times n$ matrix.

Given a random permutation P of $\{1,2,\ldots,n^2\}$, and a vector $\boldsymbol{V}$ of n^2 i.i.d. $U(0,n^2)$ random variables, let U_i be the $P(i)$th smallest element of $\boldsymbol{V}$, that is, $U_i = V_{(P(i))}$.

Conversely, given random variables U_i, $1 \leq i \leq n^2$, i.i.d. $U(0,n^2)$, define the random permutation by $P(k) = \text{rank } U_k$.

This gives our desired relations between $\boldsymbol{U}$ and P. By noting that, since $V_{(k)}/n^2$ is $Beta(n^2+k+1,k)$ distributed,

$$E(k - V_{(k)}) = \frac{k}{n^2+1},$$

we also have for i in the optimal assignment for either case

$$\frac{1}{n^2+1} \leq E(P(i) - U_i) \leq \frac{n^2}{n^2+1}.$$

Now, if π is optimal for P,

$$EZ_3^*(n) = E\left(\frac{1}{n^2}\sum_{i\in\pi} P(i) - U(i) + U(i)\right) \geq EZ_c^*(n) + \frac{1}{n(n^2+1)},$$

and if π is optimal for $\boldsymbol{U}$,

$$EZ_c^*(n) = E\left(\frac{1}{n^2}\sum_{i\in\pi} U(i) - P(i) + P(i)\right) \geq EZ_3^*(n) - \frac{n}{n^2+1}.$$

And by letting n tend to infinity, we get the limit

$$EZ_3^* = \lim_{n\to\infty} EZ_3^*(n) = \frac{\pi^2}{6}.$$

2.4 Case IV

As in the second case, given the i.i.d. uniform $(0, n^2)$ variables U_{ij}, define X_{ij} and Y_{ij} by

$$X_{ij} = [U_{ij}], \quad Y_{ij} = X_{ij} + 1.$$

If π is optimal for $\boldsymbol{Y}$,

$$Z_4^*(n) = \frac{1}{n^2}(Y_{1\pi(1)} + \cdots + Y_{n\pi(n)}) \geq \frac{1}{n^2}(U_{1\pi(1)} + \cdots + U_{n\pi(n)}) \geq Z_c^*(n).$$

If π is optimal for $\boldsymbol{U}$,

$$Z_c^*(n) = \frac{1}{n^2}(U_{1\pi(1)} + \cdots + U_{n\pi(n)}) \geq \frac{1}{n^2}(X_{1\pi(1)} + \cdots + X_{n\pi(n)}) \geq Z_4^*(n) - \frac{1}{n}.$$

Combining this, we get by letting n tend to infinity

$$EZ_4^* = \lim_{n\to\infty} EZ_4^*(n) = \frac{\pi^2}{6}.$$

3 Simulation

The primary purpose of the simulation study is of course to estimate the expected minimal cost. Besides that, we look at the variance of the expected minimal cost, as well as the row rank distribution.

To solve the realizations, we used an algorithm by Jonker and Volgenant [6]. In a recent survey [4], it came out as one of the fastest available algorithms for

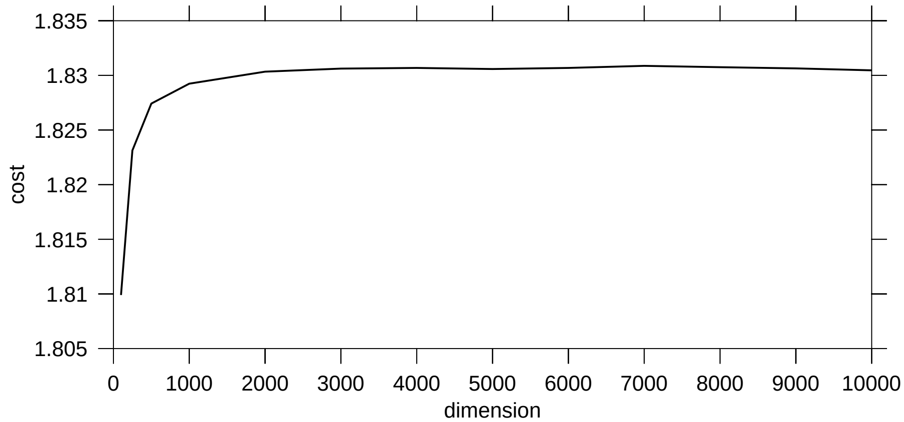

Figure 1: Simulation results, case I

problems like ours. Source code written by Jonker is available on the Internet[1], and a C++ version was used for these simulations. The algorithm has time complexity $O(n^3)$. Beside the dimension, the time also depends on the size of the matrix elements, which makes the simulations of cases III and IV more time consuming.

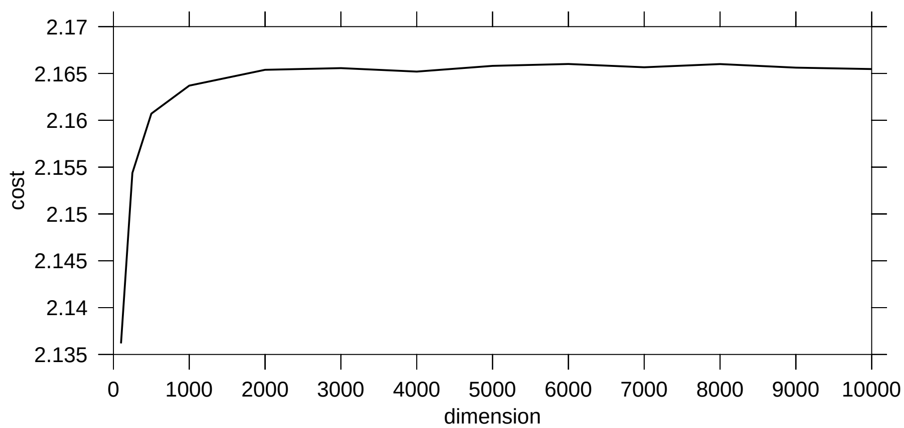

Figure 2: Simulation results, case II

As an indication of how fast the implemented algorithm really is, we note that in the permutation cases, the generation of the matrices takes about the same time as solving the assignment problem. In the independent cases the proportion of the time, spent generating the matrices, is about 0.25–0.4, depending on the dimension. An instance of dimension 1000 is solved in less than a second for all cases. For cases I and II it takes about 75–95 seconds to solve the problem with dimension 10000, and 30 seconds to generate the matrix. Almost 400 MB of RAM is needed for this dimension. The high dimension cases was run on a computer with two 1000 MHz Pentium III processors and 2 GB of RAM.

[1] http://www.magiclogic.com/assignment.html

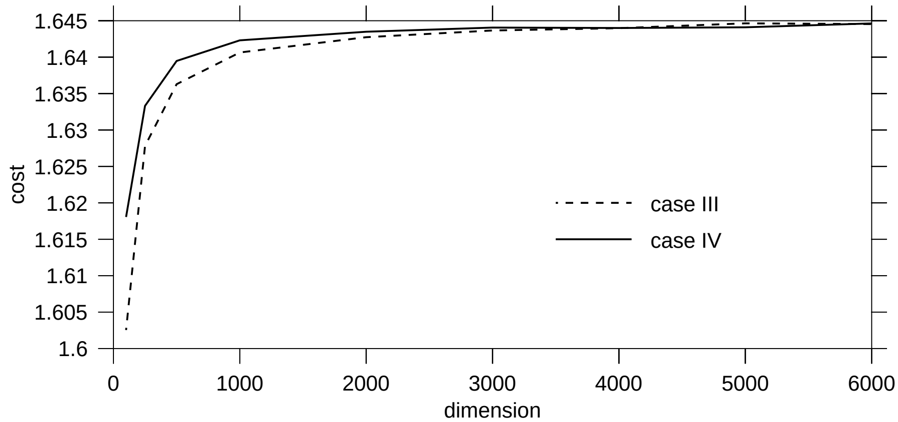

Figure 3: Simulation results, cases III and IV

3.1 Results

3.1.1 Mean

The results are summarized in Tables 1–4 and Figures 1–3. Note that n in the tables is the number of realizations.

For case I and case II we simulated problems with dimensions up to 10000. The number of realizations varies between 40000 and 4000. We see that the estimated means stabilize quite fast. The difference between dimensions 2000 and 10000 is of order 10^{-4}, the same order as the standard error.

The n^2 cases III and IV behaves as expected. The mean increases nicely towards $\pi^2/6$, with case IV slightly ahead. Since these cases are more time consuming, and the limit is known to be $\pi^2/6$, we was content with simulations up to dimension 6000.

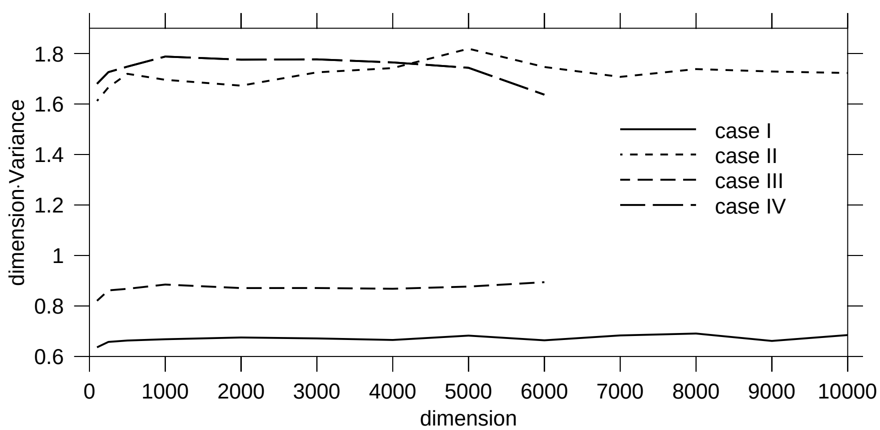

Figure 4: Estimated variance

3.1.2 Variance

Alm and Sorkin [2] conjectures that the variance in the exponential case is $2/n + O(\log n/n^2)$. It is natural to suspect the same behavior in all our four cases. Figure 4 shows n times the estimated variance plotted against n. It is interesting to note is that the variance in the permutation cases is about half of that in the independent cases.

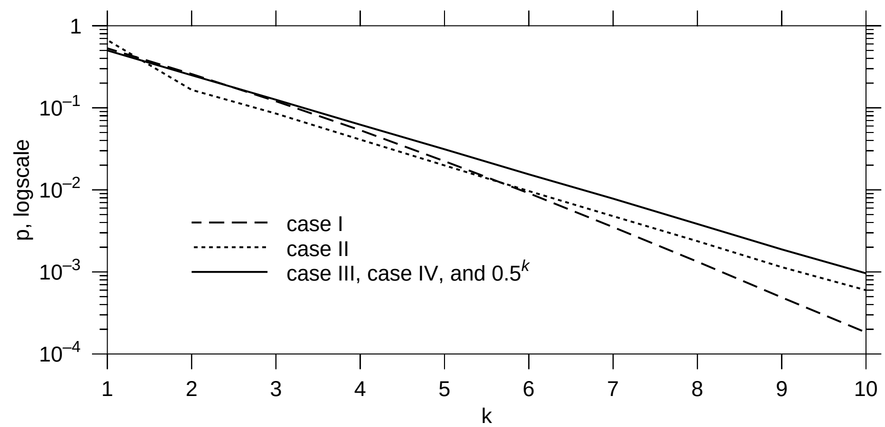

Figure 5: Estimated rank distribution, log-scale

3.1.3 The rank distribution

In the continuous cases, the limiting rank distribution is geometric, with parameter 1/2. For comparison, we generated 1000 matrices of dimension 2000 for each discrete case, and determined the rank of every element in the optimal assignment given by the program. (Optimal assignments are not necessarily unique.) In the case of ties, we gave the element the lowest rank.

As suspected, cases III and IV seems to have the same limiting distribution as in the continuous case. Also in case II a geometric distribution, but with extra weight on 1, fit the data very well. For case I the picture looks a bit different. When plotted on a logarithmic scale, (Figure 5) we no longer get a straight line, but a slightly concave curve. (In this scale, a polynomial in k of degree 2 fit the data well.)

Acknowledgement

I would like to thank Svante Janson for suggesting both the problem and the coupling approach.

References

[1] D. J. Aldous. The $\zeta(2)$ limit in the random assignment problem. To appear in *Random Structures and Algorithms*.

[2] S. E. Alm and G. B. Sorkin. Exact expectations and distributions in the random assignment problem. To appear in *Combinatorics, Probability and Computing*.

[3] D. Coppersmith and G. B. Sorkin. Constructive bounds and exact expectations for the random assignment problem. *Random Structures and Algorithms*, 15(2):113–144, 1999.

[4] M. Dell'Amico and P. Toth. Algorithms and codes for dense assignment problems: the state of the art. *Discrete Appl. Math.*, 100(1-2):17–48, 2000.

[5] S. Janson, 2001. Personal communication.

[6] R. Jonker and A. Volgenant. A shortest augmenting path algorithm for dense and sparse linear assignment problems. *Computing*, 38(4):325–340, 1987.

[7] S. Linusson and J. Wästlund. A generalization of the random assignment problem. Preprint, 2000.

[8] M. Mézard and G. Parisi. Replicas and optimization. *Journal de Physiques Lettres*, 46:771–778, 1985.

[9] M. Mézard and G. Parisi. On the solution of the random link matching problem. *Journal de Physiques Lettres*, 48:1451–1459, 1987.

[10] B. Olin. *Asymptotic properties of random assignment problems*. PhD thesis, Kungliga Tekniska Högskolan, Stockholm, Sweden, 1992.

[11] G. Parisi. A conjecture on random bipartite matching. Preprint, 1998.

Table 1: Simulation results, case I.

dimension	n	mean	std. dev.	s.e. mean
100	40000	1.80990	0.0797616	0.000398808
250	40000	1.82313	0.0512889	0.000256445
500	20000	1.82742	0.0364174	0.000257510
1000	20000	1.82924	0.0258506	0.000182791
2000	10000	1.83034	0.0183741	0.000183741
3000	10000	1.83062	0.0149616	0.000149616
4000	5000	1.83068	0.0128971	0.000182393
5000	5000	1.83058	0.0116841	0.000165238
6000	4000	1.83068	0.0105195	0.000166328
7000	4000	1.83087	0.0098804	0.000156223
8000	4000	1.83075	0.0092924	0.000146927
9000	4000	1.83064	0.0085737	0.000135563
10000	4000	1.83046	0.0082750	0.000130840

Table 2: Simulation results, case II.

dimension	n	mean	std. dev.	s.e. mean
100	40000	2.13618	0.1269710	0.000634857
250	40000	2.15438	0.0816278	0.000408139
500	20000	2.16071	0.0586444	0.000414678
1000	20000	2.16370	0.0411801	0.000291188
2000	10000	2.16539	0.0289196	0.000289196
3000	10000	2.16557	0.0239816	0.000239816
4000	5000	2.16520	0.0208703	0.000295150
5000	5000	2.16581	0.0190736	0.000269742
6000	4000	2.16601	0.0170615	0.000269765
7000	4000	2.16566	0.0156181	0.000246944
8000	4000	2.16600	0.0147405	0.000233067
9000	4000	2.16562	0.0138590	0.000219130
10000	4000	2.16547	0.0131255	0.000207532

Table 3: Simulation results, case III.

dimension	n	mean	std. dev.	s.e. mean
100	40000	1.60254	0.0905816	0.000452908
250	40000	1.62781	0.0587116	0.000293558
500	20000	1.63629	0.0416650	0.000294616
1000	20000	1.64064	0.0297464	0.000210339
2000	10000	1.64274	0.0208671	0.000208671
3000	10000	1.64366	0.0170409	0.000170409
4000	2000	1.64398	0.0147322	0.000329422
5000	2000	1.64464	0.0132429	0.000296119
6000	2000	1.64454	0.0122086	0.000272994

Table 4: Simulation results, case IV.

dimension	n	mean	std. dev.	s.e. mean
100	40000	1.61806	0.1296140	0.000648068
250	40000	1.63331	0.0830974	0.000415487
500	20000	1.63948	0.0591263	0.000418086
1000	20000	1.64231	0.0422843	0.000298995
2000	10000	1.64349	0.0297958	0.000297958
3000	10000	1.64406	0.0243364	0.000243364
4000	2000	1.64400	0.0210041	0.000469665
5000	2000	1.64410	0.0186728	0.000417537
6000	2000	1.64463	0.0165169	0.000369328