The Weight of the Shortest Path Tree

Remco van der Hofstad∗ Gerard Hooghiemstra† Piet Van Mieghem†

June 13, 2005

Abstract

The minimal weight of the shortest path tree in a complete graph $K_N$, with $N + 1$ nodes, equipped with independent and exponential (mean 1) random weights, is shown to converge to a Gaussian distribution. We prove the central limit theorem by conditioning sums of involved random indicators to be close to their respective mean values.

1 Introduction

Consider the complete graph $K_N$, with $N + 1$ nodes and $\frac{1}{2}N(N + 1)$ edges. To each edge we assign independently an exponentially distributed weight with mean 1. Take two distinct nodes, for instance node 1 and node 2. The shortest path between these two nodes is the (with probability one) unique path between node 1 and node 2, such that the sum of the exponential weights taken over the edges of the path is minimal. In this paper we consider the total weight $W_N$ of the shortest path (SPT) tree rooted at node 1 to all other nodes in the complete graph. In [6, 8], we have rephrased the shortest path problem between two arbitrary nodes in the complete graph with exponential link weights to a Markov discovery process which starts the path searching process at the source and which is a continuous time Markov chain with $N + 1$ states. Each state $n$ represents the $n$ already discovered nodes (including the source node). If at some stage in the Markov discovery process $n$ nodes are discovered, then the next node is reached with rate $\lambda_n = n(N + 1 - n)$, which is the transition rate in the continuous-time Markov chain. Since the discovery of nodes at each stage only increases $n$, the Markov discovery process is a pure birth process with birth rate $n(N + 1 - n)$. We call $\tau_n$ the inter-attachment time between the inclusion of the $n^{th}$ and $(n + 1)^{st}$ node to the SPT for $n = 1, \ldots, N$. The inter-attachment time $\tau_n$ is exponentially distributed with parameter $\lambda_n$ as follows from the theory of Markov processes. By the memoryless property of the exponential distribution, the new node is added uniformly to an already discovered node. Hence, the resulting SPT to all nodes is exactly a uniform recursive tree (URT). A URT of size $N + 1$ is a random tree rooted at some source node and where at each stage a new node is attached uniformly to one of the existing nodes until the total number of nodes is equal to $N + 1$.

∗rhofstad@win.tue.nl, Department of Mathematics and Computer Science, Eindhoven University of Technology, P. O. Box 513, 5600 MB Eindhoven, The Netherlands.
†G.Hooghiemstra@ewi.tudelft.nl and P.VanMieghem@ewi.tudelft.nl, Electrical Engineering, Mathematics and Computer Science, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands.
The average of the weight $W_N$ of the SPT equals
\[
E[W_N] = \sum_{k=1}^{N} \frac{1}{k^2},
\]
and the variance is
\[
\text{var}[W_N] = \frac{4}{(N + 1)} \sum_{k=1}^{N} \frac{1}{k^3} + 4 \sum_{j=1}^{N} \frac{1}{j^2} \sum_{k=1}^{j} \frac{1}{k} - 5 \sum_{j=1}^{N} \frac{1}{j^4}.
\]

The result for the mean (1) has been found first in [7], but it is rederived in Section 2.1 because the method is considerably simpler. The derivation for the variance (2) is in Section 2.2 while many appearing sums are computed in the Appendix. The asymptotic form of the average weight is immediate from (1) as
\[
E[W_N] = \zeta(2) + O\left(\frac{1}{N}\right),
\]
while the corresponding result for the variance, derived in Section 2.2, is
\[
\text{var}[W_N] = \frac{4\zeta(3)}{N} + o\left(\frac{1}{N}\right).
\]

The third result in this paper is that we show that the scaled weight of the SPT tends to a Gaussian. In particular,
\[
\sqrt{N} (W_N - \zeta(2)) \xrightarrow{d} N(0, \sigma^2),
\]
where $\sigma^2 = \sigma^2_{SPT} = 4\zeta(3) \simeq 4.80823$. A related result for the minimum spanning tree (MST) is worth mentioning. The average weight of the minimum spanning tree $W_{\text{MST}}$ in the complete graph with exponential with mean 1 (or uniform on $[0, 1]$) link weights has been computed earlier by Frieze [2]. For large $N$, Frieze showed that
\[
E[W_{\text{MST}}] \to \zeta(3).
\]
Janson [4] extended Frieze’s result by proving that the scaled weight of the MST tends to a Gaussian,
\[
\sqrt{N} (W_{\text{MST}} - \zeta(3)) \xrightarrow{d} N(0, \sigma^2_{\text{MST}}),
\]
where
\[
\sigma^2_{\text{MST}} = 2\zeta(4) - 2 \sum_{i=0}^{\infty} \sum_{j=1}^{\infty} \sum_{k=1}^{\infty} \frac{(i + k - 1)!k^i(i + j)^{i-2}j}{i!k!(i + j + k)^{i+k+2}} \simeq 1.6857.
\]

## 2 The weight of the shortest path tree

From the Markov discovery process briefly explained in Section 1, the discovery time to the $k$th discovered node from the root equals
\[
v_k = \sum_{n=1}^{k} \tau_n
\]
where the inter-attachment times $\tau_1, \tau_2, \ldots, \tau_k$ are independent, exponentially distributed random variables with parameter $\lambda_n = n(N + 1 - n), 1 \leq n \leq k$. An arbitrary uniform recursive tree consisting of $N + 1$ nodes and with the root labeled by zero can be represented as
\[
(0 \leftarrow 1) (N_2 \leftarrow 2) \ldots (N_N \leftarrow N)
\]
where \((N_j \leftarrow j)\) means that the \(j^{th}\) discovered node is attached to node \(N_j \in \{0, \ldots, j-1\}\). Hence, \(N_j\) is the predecessor of \(j\) and this relation is indicated by \(\leftarrow\). The weight \(W_N\) of an arbitrary SPT from the root 0 to all \(N\) other nodes is with (5) and \(v_0 = 0\) and \(N_1 = 0\),

\[
W_N = \sum_{j=1}^{N} (v_j - v_{N_j}) = \sum_{j=1}^{N} \sum_{n=N_{j}+1}^{j} \tau_n.
\]

In the URT, the integer \(N_j\), \(1 \leq j \leq N\), are independent and uniformly distributed over the interval \(\{0, \ldots, j-1\}\). It is more convenient to use a discrete uniform random variable on \(\{1, \ldots, j\}\) which we define as \(A_j = N_j + 1\). We rewrite

\[
W_N = \sum_{j=1}^{N} \sum_{n=A_j}^{j} \tau_n = \sum_{j=1}^{N} \sum_{n=1}^{j} 1\{A_j \leq n\} \tau_n = \sum_{n=1}^{N} \tau_n \left( \sum_{j=n}^{N} 1\{A_j \leq n\} \right)
\]

The set \(\{A_j\}_{1 \leq j \leq N}\) are independent random variables with \(\mathbb{P}[A_j = k] = \frac{1}{j}\) for \(k \in \{1, 2, \ldots, j\}\). In addition, we define for \(n \in \{1, \ldots, N\}\) the random variables

\[
B_n = \sum_{j=n}^{N} 1\{A_j \leq n\},
\]

(7)

to obtain

\[
W_N = \sum_{n=1}^{N} B_n \tau_n.
\]

(8)

The \(n\) random variables \(B_1, B_2, \ldots, B_n\) are dependent. The mean of the random variable \(B_n\) follows from (7) as

\[
\mathbb{E}[B_n] = \sum_{j=n}^{N} \mathbb{E}[1\{A_j \leq n\}] = \sum_{j=n}^{N} \mathbb{P}[A_j \leq n] = \sum_{j=n}^{N} \frac{n}{j}.
\]

(9)

Also, \(B_n \leq \sum_{j=n}^{N} 1 = N + 1 - j\). The variance \(\text{var}[B_n]\) and covariances \(\text{cov}[B_n, B_m]\) are given in Lemma 1 below.

### 2.1 The average weight of the SPT

It is immediate from (8) and the independence of the \(A_1, A_2, \ldots, A_N\) from the inter-attachment times \(\tau_1, \tau_2, \ldots, \tau_N\) that

\[
\mathbb{E}[W_N] = \sum_{n=1}^{N} \mathbb{E}[B_n] \mathbb{E}[\tau_n] = \sum_{n=1}^{N} \sum_{j=n}^{N} \frac{1}{jn} (N+1-n) = \sum_{n=1}^{N} \frac{1}{(N+1-n)} \sum_{j=n}^{N} \frac{1}{j} = \sum_{j=1}^{N} \frac{1}{j} \sum_{k=N+1-j}^{N} \frac{1}{k},
\]

which is by the equality (28) below equal to (1).

### 2.2 The variance of the weight of the SPT

To compute the variance of \(W_N\), we use the formula

\[
\text{var}[W_N] = \text{var}[\mathbb{E}[W_N|B_1, \ldots, B_N]] + \mathbb{E}[\text{var}[W_N|B_1, \ldots, B_N]].
\]

(10)
Since for an exponential random variable $\tau_n$ with parameter $\lambda_n = n(N + 1 - n)$, the expectation equals $1/\lambda_n$ and the variance $1/\lambda_n^2$, we have
\[
\mathbb{E}[W_N|B_1, \ldots, B_N] = \mathbb{E} \left[ \sum_{n=1}^{N} B_n \tau_n | B_1, \ldots, B_N \right] = \sum_{n=1}^{N} \lambda_n^{-1} B_n, \tag{11}
\]
\[
\text{var}[W_N|B_1, \ldots, B_N] = \text{var} \left[ \sum_{n=1}^{N} B_n \tau_n | B_1, \ldots, B_N \right] = \sum_{n=1}^{N} \lambda_n^{-2} B_n^2. \tag{12}
\]
Combining (10), (11) and (12),
\[
\text{var}[W_N] = \text{var} \left[ \sum_{n=1}^{N} \lambda_n^{-1} B_n \right] + \sum_{n=1}^{N} \lambda_n^{-2} \mathbb{E}[B_n^2]. \tag{13}
\]
To proceed, we need expressions for the covariance of $B_n$ and $B_m$, which are computed in the following lemma:

**Lemma 1** For every $n, m \geq 1$,

(i) \[
\text{var}[B_n] = \sum_{j=n}^{N} \left( \frac{n}{j} - \frac{n^2}{j^2} \right), \tag{14}
\]

(ii) \[
\text{cov}[B_n, B_m] = \sum_{j=m}^{N} \frac{n}{j} \left( 1 - \frac{m}{j} \right), \quad n \leq m. \tag{15}
\]

**Proof.** The proof of (i) follows from that of (ii) with $n = m$.

(ii) The bilinearity of the covariance yields, for $n \leq m$,
\[
\text{cov}[B_n, B_m] = \text{cov} \left[ \sum_{i=n}^{N} 1_{\{A_i \leq n\}}, \sum_{j=m}^{N} 1_{\{A_j \leq m\}} \right] = \sum_{i=n}^{N} \sum_{j=m}^{N} \text{cov} \left[ 1_{\{A_i \leq n\}}, 1_{\{A_j \leq m\}} \right].
\]
Since $A_i$ and $A_j$ are independent for $i \neq j$, we have that $\text{cov} \left[ 1_{\{A_i \leq n\}}, 1_{\{A_j \leq m\}} \right] = 0$, for $i \neq j$, such that
\[
\sum_{i=n}^{N} \sum_{j=m}^{N} \text{cov} \left[ 1_{\{A_i \leq n\}}, 1_{\{A_j \leq m\}} \right] = \sum_{j=m}^{N} \text{cov} \left[ 1_{\{A_j \leq n\}}, 1_{\{A_j \leq m\}} \right].
\]
With $\text{cov} \left[ 1_{\{A_j \leq n\}}, 1_{\{A_j \leq m\}} \right] = \mathbb{E} \left[ 1_{\{A_j \leq n\}} 1_{\{A_j \leq m\}} \right] - \mathbb{P}[A_j \leq n] \mathbb{P}[A_j \leq m]$ and $1_{\{A_j \leq n\}} 1_{\{A_j \leq m\}} = 1_{\{A_j \leq \min(n,m)\}} = 1_{\{A_j \leq n\}}$ for $n \leq m$, we obtain
\[
\sum_{j=m}^{N} \text{cov} \left[ 1_{\{A_j \leq n\}}, 1_{\{A_j \leq m\}} \right] = \sum_{j=m}^{N} \frac{n}{j} \left( 1 - \frac{m}{j} \right). \quad \square
\]
Applying Lemma 1 to the right side of (13) gives
\[
\text{var}[W_N] = \text{var} \left[ \sum_{n=1}^{N} \lambda_n^{-1} B_n \right] + \sum_{n=1}^{N} \lambda_n^{-2} \mathbb{E}[B_n^2] = 2 \sum_{n \leq m} \lambda_n^{-1} \lambda_m^{-1} \text{cov}[B_n, B_m] + \sum_{n=1}^{N} \lambda_n^{-2} (\mathbb{E}[B_n])^2
\]
\[
= 2 \sum_{n \leq m} \lambda_n^{-1} \lambda_m^{-1} \sum_{j=m}^{N} \frac{n}{j} \left( 1 - \frac{m}{j} \right) + \sum_{n=1}^{N} \lambda_n^{-2} \left( \sum_{j=n}^{N} \frac{n}{j} \right)^2 = T_2(N) + T_1(N).
\]
where the sum $T_1(N)$ is defined as

$$T_1(N) = \sum_{n=1}^{N} \frac{1}{(N + 1 - n)^2} \left( \sum_{j=n}^{N} \frac{1}{j} \right)^2 = 2 \sum_{k=1}^{N} \frac{1}{k^3} \sum_{n=1}^{k} \frac{1}{n} - \left( \sum_{k=1}^{N} \frac{1}{k^2} \right)^2,$$

where the last equality is proved in the Appendix (see (44) below), while

$$T_2(N) = 2 \sum_{n=1}^{N} \frac{1}{N+1-n} \sum_{m=n}^{N} \frac{1}{m(N+1-m)} \sum_{j=m}^{N} \frac{1}{j} \left( 1 - \frac{m}{j} \right)$$

$$= \frac{4}{N+1} \sum_{k=1}^{N} \frac{1}{k^3} - \frac{5}{N^4} \sum_{k=1}^{N} \frac{1}{k^4} + 2 \sum_{n=1}^{N} \frac{1}{n^3} \sum_{k=1}^{n} \frac{1}{k} + \left( \sum_{k=1}^{N} \frac{1}{k^2} \right)^2,$$

where the last equality is proved in the Appendix (see (42) below). Summing $T_1(N)$ and $T_2(N)$ gives the explicit form (2) of the variance for $W_N$.

We next investigate the asymtotics of the variance of $W_N$ for large $N$. We write the sum of the last two terms in (2) by

$$Q(N) = 4 \sum_{j=1}^{N} \frac{1}{j^3} \sum_{k=1}^{j} \frac{1}{k} - \frac{5}{N^4} \sum_{j=1}^{N} \frac{1}{j^4}$$

and, by summation, $Q(N) = Q + O \left( \frac{\log N}{N^2} \right)$, where the limit $Q = \lim_{N \to \infty} Q(N)$ exists, by (18). It follows from [1, Corollary 4, main theorem] that

$$Q = 4 \sum_{j=1}^{\infty} \frac{1}{j^3} \sum_{k=1}^{j} \frac{1}{k} - \frac{5}{N^4} \sum_{j=1}^{\infty} \frac{1}{j^4} = 0,$$

so that

$$Q(N) = O \left( \frac{\log N}{N^2} \right).$$

Hence, asymptotically, we arrive at (4).

3 Central limit theorem for $W_N$

In this section we prove a central limit theorem for $W_N$. We use the symbol $\overset{d}{\rightarrow}$ to denote convergence in distribution and the symbol $\overset{p}{\rightarrow}$ for convergence in probability. We denote by $\sigma(W_N)$, the standard deviation of $W_N$, so $\sigma^2(W_N) = \text{var}[W_N]$. We denote by $\mathcal{N}(0,1)$ a random variable with standard normal distribution. The main result proved in this section is the following central limit theorem for $W_N$:

**Theorem 2** As $N \to \infty$,

$$\frac{W_N - \mathbb{E}[W_N]}{\sigma(W_N)} \overset{d}{\rightarrow} \mathcal{N}(0,1).$$
We start with an outline of the proof. We wish to prove that $W_N$ is asymptotically normal, in the sense that $\sqrt{N}(W_N - \mathbb{E}[W_N])$ has an asymptotic normal distribution. We first define

$$Ns^2_N = N \sum_{j=1}^{N} \frac{B_j^2}{j^2(N+1-j)^2}. \tag{21}$$

We split $W_N = X_N + Y_N$, where

$$X_N = \sum_{j=1}^{N} \left( \tau_j - \frac{1}{j(N+1-j)} \right) B_j, \quad \text{and} \quad Y_N = \sum_{j=1}^{N} \frac{B_j}{j(N+1-j)}. \tag{22}$$

Our strategy is to prove the following steps.

1. Define an event $A_N$ with $\mathbb{P}(A_N^c) \leq N^{-\delta}$, such that, on $A_N$, we obtain $Ns^2_N - \sigma^2_{1,N} = o(1)$, where

$$\sigma^2_{1,N} = N \sum_{j=1}^{N} \frac{(\mathbb{E}[B_j])^2}{j^2(N+1-j)^2} = NT_1(N).$$

Consecutively, we show that $\sigma^2_1 = \lim_{N \to \infty} \sigma^2_{1,N}$ exists.

2. Prove the central limit theorem for $\sqrt{N}X_N$ with variance $Ns^2_N$, conditionally on $\{A_j\}_{j=1}^{N}$, when $\{A_j\}_{j=1}^{N}$ is such that $A_N$ holds.

3. Prove that $\sqrt{N}(Y_N - \mathbb{E}[Y_N])$ converges in distribution to a normal random variable with variance $\sigma^2_2 = \lim_{N \to \infty} NT_2(N)$.

These steps would prove the claim. Indeed, we compute

$$\phi(t) = \mathbb{E}[e^{it\sqrt{N}W_N}] = \mathbb{E}[e^{it\sqrt{N}W_N}1_{A_N}] + O(\mathbb{P}(A_N^c)) = \mathbb{E}\left[\mathbb{E}_A[e^{it\sqrt{N}W_N}]1_{A_N}\right] + O(N^{-\delta}),$$

where $\mathbb{E}_A$ is the conditional expectation given $\{A_j\}_{j=1}^{N}$. We split $W_N = X_N + Y_N$, and use that $Y_N$ is measurable with respect to $\{A_j\}_{j=1}^{N}$ to arrive at

$$\phi(t) = \mathbb{E}\left[\mathbb{E}_A[e^{it\sqrt{N}X_N}]\mathbb{E}_A[e^{it\sqrt{N}X_N}]1_{A_N}\right] + O(N^{-\delta}).$$

According to Step 2,

$$\mathbb{E}_A[e^{it\sqrt{N}X_N}] = e^{-t^2Ns^2_N/2} + o(1),$$

and, according to Step 1,

$$\mathbb{E}_A[e^{it\sqrt{N}X_N}] = e^{-t^2\sigma^2_{1,N}/2} + o(1).$$

Therefore, using that $\mathbb{E}[W_N] = \mathbb{E}[Y_N]$,

$$\mathbb{E}[e^{it\sqrt{N}(W_N - \mathbb{E}[W_N])}] = e^{-t^2\sigma^2_{1,N}/2}\mathbb{E}[e^{it\sqrt{N}(Y_N - \mathbb{E}[Y_N])}]1_{A_N} + o(1),$$

again by Step 1. Now, by Step 3, we have that

$$\mathbb{E}[e^{it\sqrt{N}(Y_N - \mathbb{E}[Y_N])}] = e^{-t^2\sigma^2_{2}/2} + o(1),$$

so that

$$\mathbb{E}[e^{it\sqrt{N}(W_N - \mathbb{E}[W_N])}] = e^{-t^2\sigma^2/2} + o(1),$$

where $\sigma^2 = \sigma^2_1 + \sigma^2_2$, and $\sigma^2_1 = \lim_{N \to \infty} \sigma^2_{1,N}$. We now turn to the detail of the proof. We will prove Steps 1-3 in Sections 3.1–3.3 respectively.
3.1 Step 1: The good event and convergence in probability of $N s_N^2$

Fix $a \in (0, 1)$ and an integer $n_0$, and define

$$A_N = B_N \cap C_N,$$

where

$$B_N = \bigcap_{j=N^a}^{N-N^a} \{|B_j - \mathbb{E}[B_j]| \leq \varepsilon_N \mathbb{E}[B_j]\}, \quad (23)$$

and

$$C_N = \bigcap_{j=1}^{N^a} \{B_j \leq \max(2n_0, j) \log N\}, \quad (24)$$

with

$$\varepsilon_N = N^{-(\min(a,1-a))/3}.$$

Later we will see that in fact we need $n_0$ large and $a > \frac{3}{4}$. On the event $B_N$ all random variables $B_j$, $N^a \leq j \leq N - N^a$, are close to their respective mean $\mathbb{E}[B_j]$; on the event $C_N$, we have a logarithmic bound on the random variables $B_j$, $1 \leq j \leq N^a$.

We will show two lemmas. The first shows that $A_N$ occurs with high probability, while the second proves that $N s_N^2$ is close to a constant on $A_N$. Together the lemmas imply the claims in step 1.

**Lemma 3** There exists a $\delta > 0$, such that for $N$ sufficiently large,

$$\mathbb{P}(A_N^c) \leq N^{-\delta}.$$

**Proof.** We use Boole’s inequality to obtain,

$$\mathbb{P}(A_N^c) \leq \sum_{j=1}^{N^a} \mathbb{P}(B_j > \max(2n_0, j) \log N) + \sum_{j=N^a}^{N-N^a} \mathbb{P}(|B_j - \mathbb{E}[B_j]| \geq \varepsilon_N \mathbb{E}[B_j]).$$

Note that $B_j$ is the sum of independent indicators, and, therefore, by the estimate of Janson [5] and with $0 < \epsilon < 1$,

$$\mathbb{P}(|B_j - \mathbb{E}[B_j]| \geq \epsilon \mathbb{E}[B_j]) \leq 2e^{-\frac{3}{8} \epsilon \mathbb{E}[B_j]}$$

where $\mathbb{E}[B_j]$ is given in (9) which we bound as

$$j \log \left(\frac{N}{j}\right) = j \int_j^N \frac{1}{x} \, dx \leq \mathbb{E}[B_j] \leq j \int_{j-1}^N \frac{1}{x} \, dx = j \log \left(\frac{N}{j-1}\right).$$

Therefore, we have that

$$\mathbb{P}(|B_j - \mathbb{E}[B_j]| \geq \epsilon \mathbb{E}[B_j]) \leq 2e^{-\frac{3}{8} \epsilon \log \frac{N}{j}},$$

which is $o(N^{-2})$ for all $n_0 \leq j \leq N^a$ and $n_0$ sufficiently large. On the other hand, for $j \leq n_0$,

$$\mathbb{P}(B_j \geq 2n_0 \log N) \leq \mathbb{P}(|B_j - \mathbb{E}[B_j]| \geq n_0 \mathbb{E}[B_j]) \leq N^{-2},$$

again for $n_0 > 1$ sufficiently large. We conclude that with high probability $B_j \leq \max(2n_0, j) \log N$ and all $j \leq N^a$. 


We complete the argument as follows. For \( N^a \leq j \leq N - N^a \), and \( \epsilon_N = N^{-(\min(a,1-a))/3} \), we have that
\[
\mathbb{P}(|B_j - \mathbb{E}[B_j]| \geq \epsilon_N \mathbb{E}[B_j]) \leq 2e^{-\frac{2}{N^a}} j \log N \leq 2e^{-c_N N^{\min(a,1-a)}},
\]
since, for all \( j \) such that \( N^a \leq j \leq N - N^a \), we have \( j \log N \geq N^{\min(a,1-a)}. \)

Recall that \( Ns_N^2 = N \sum_{j=1}^{N} \frac{B_j^2}{j^2(N+1-j)^2} \). We next investigate \( Ns_N^2 \) on the event \( A_N \):

**Lemma 4** On the event \( A_N \), and for \( N \to \infty \),
\[
N s_N^2 = N \sum_{j=1}^{N} \frac{(\mathbb{E}[B_j])^2}{j^2(N+1-j)^2} = o(1).
\]

**Proof.** From (7) the inequality \( B_j \leq N + 1 - j \) is obvious, hence
\[
N \sum_{j=N-N^a}^{N} \frac{B_j^2}{j^2(N+1-j)^2} \leq N \sum_{j=N-N^a}^{N} \frac{1}{j^2} \leq \frac{N^a}{N-N^a} = O(N^{a-1}),
\]
for any \( a < 1 \). Therefore, we have that
\[
N s_N^2 = N \sum_{j=1}^{N-N^a} \frac{B_j^2}{j^2(N+1-j)^2} + o(1).
\]

On the event \( C_N \),
\[
N \sum_{j=1}^{N-N^a} \frac{B_j^2}{j^2(N+1-j)^2} \leq N \sum_{j=1}^{N-N^a} \frac{(\max(2n_0,j))^2(\log N)^2}{j^2N^2} \leq O(N^{-1+a}(\log N)^2) = o(1),
\]
so that on \( C_N \),
\[
N s_N^2 = N \sum_{j=N^a}^{N-N^a} \frac{B_j^2}{j^2(N+1-j)^2} + o(1).
\]

On \( B_N \), and for \( N^a \leq j \leq N - N^a \), we can sandwich \((1 - \epsilon_N)^2(\mathbb{E}[B_j])^2 \leq B_j^2 \leq (1 + \epsilon_N)^2(\mathbb{E}[B_j])^2\), so with probability \( 1 - O(N^{-\delta}) \) we find,
\[
N s_N^2 = (1 + O(\epsilon_N)) N \sum_{j=N^a}^{N-N^a} \frac{(\mathbb{E}[B_j])^2}{j^2(N+1-j)^2} + o(1).
\]

This completes the proof of the lemma.

The argument of convergence in probability of \( N s_N^2 \) is complete when we prove that
\[
N \sum_{j=1}^{N} \frac{(\mathbb{E}[B_j])^2}{j^2(N+1-j)^2} \to \sigma_1^2.
\]

For this, we note that
\[
N \sum_{j=1}^{N} \frac{(\mathbb{E}[B_j])^2}{j^2(N+1-j)^2} = NT_1(N),
\]
and from (45), we find that \( \sigma_1^2 = 2\zeta(2) \).
3.2 Step 2: Central limit theorem for $X_n$ conditionally on $\{A_j\}_{j=1}^N$ in good event

In this section, we compute $E_A[e^{it\sqrt{N}X_n}]$, where $\{A_j\}_{j=1}^N$ is such that $A_N$ holds. For this, we note that, for any random variable $X$ with finite third moment, we have that

$$\phi_X(t) = E[e^{itX}] = e^{it\mu - t^2\sigma^2/2 + O(|t|^3 m_3)},$$ \hspace{1cm} (25)

where $\mu = E[X]$, $\sigma^2 = \text{var}(X)$ and $m_3 = E[|X|^3]$. The independence of the $\tau_j$, conditionally on $\{A_j\}_{j=1}^N$, gives that

$$E_A[e^{it\sqrt{N}X_n}] = \prod_{j=1}^N E_A[e^{it\sqrt{N}(\tau_j - \mu(j+1/N))B_j}].$$

By (25), and since $B_1, \ldots, B_N$ are measurable with respect to the $\sigma-$algebra spanned by the random variables $A_1, A_2, \ldots, A_N$ we obtain that

$$E_A[e^{it\sqrt{N}(\tau_j - \mu(j+1/N))B_j}] = \exp\left[-t^2N \frac{B_j^2}{2j^2(N+1-j)^2} + O\left(|t|^3 \frac{N\sigma_j^3}{j^3(N+1-j)^3}\right)\right].$$

Therefore,

$$E_A[e^{it\sqrt{N}X_n}] = e^{-t^2Ns^2/2}O(|t|^3 \nu_N),$$

where

$$\nu_N = N^{3/2} \sum_{j=1}^N \frac{B_j^3}{j^3(N+1-j)^3}.$$ 

We finally show that $\nu_N = o(1)$ on $A_N$. First, we note that, since $B_j \leq N + 1 - j$ and $a < 1$,

$$N^{3/2} \sum_{j=N-N^a}^N \frac{B_j^3}{j^3(N+1-j)^3} \leq N^{3/2} \sum_{j=N-N^a}^N \frac{1}{j^3} \leq N^{3/2 - 3\eta} \leq N^{-\frac{1}{2} \eta}.$$ 

When $j \leq N - N^a$, we can use the bounds provided by $A_N$. We start with the contribution due to $j \leq N^a$, for which we can bound on $C_N$,

$$N^{3/2} \sum_{j=1}^{N^a} \frac{B_j^3}{j^3(N+1-j)^3} \leq N^{3/2} \sum_{j=1}^{N^a} \frac{(\max(2j_0, j))^3 \log(N)^3}{j^3(N+1-j)^3} \leq (\log(N))^3 N^{2/3 - 3\eta} \leq N^{-1/4}.$$ 

Finally, for $N^a \leq j \leq N - N^a$, we obtain on $B_N$, using $E[B_j] \leq j \log N$,

$$N^{3/2} \sum_{j=N^a}^{N-N^a} \frac{B_j^3}{j^3(N+1-j)^3} = (1 + \epsilon_N)^3 N^{3/2} \sum_{j=N^a}^{N-N^a} \frac{(E[B_j])^3}{j^3(N+1-j)^3}$$ 

$$\leq (1 + \epsilon_N)^3 N^{3/2} \sum_{j=N^a}^{N-N^a} \frac{(\log(N))^3}{(N+1-j)^3}$$ 

$$\leq (1 + \epsilon_N)^3 N^{3/2} N^{-2\eta} \leq N^{-\eta},$$

for any $\eta < 2a - \frac{3}{2}$, and we note that $\eta > 0$ when $a > \frac{3}{4}$. This completes the proof that $\nu_N \leq N^{-\eta}$ when $a < \frac{3}{4}$, and completes the proof that, for $\{A_j\}_{j=1}^N$ such that $A_N$ holds,

$$E_A[e^{it\sqrt{N}X_n}] = e^{-t^2Ns^2/2}O(N^{-\eta}).$$
3.3 Step 3: The central limit theorem for $Y_N$

We again use convergence of characteristic functions to that of a normal random variable with mean 0. We rewrite

$$\sqrt{N}(Y_N - \mathbb{E}[Y_N]) = \sqrt{N} \sum_{j=1}^{N} \frac{1}{j(N + 1 - j)} \sum_{k=j}^{N} (1_{\{A_k \leq j\}} - \frac{j}{k}) = \sum_{k=1}^{N} (Y_{k,N} - \mathbb{E}[Y_{k,N}]),$$

where

$$Y_{k,N} = \sqrt{N} \sum_{j=1}^{k} \frac{1}{j(N + 1 - j)}. \tag{26}$$

The summands $Y_{1,N}, \ldots, Y_{N,N}$ are independent. We wish to show that $\sqrt{N}(Y_N - \mathbb{E}[Y_N])$ is asymptotically normal with asymptotic variance $N \text{var}(Y_N)$. From the independence of the summands,

$$\mathbb{E}[e^{it\sqrt{N}(Y_N - \mathbb{E}[Y_N])}] = \prod_{k=1}^{N} \mathbb{E}[e^{it(Y_{k,N} - \mathbb{E}[Y_{k,N}])}].$$

Then, we note that

$$|Y_{k,N}| \leq \sqrt{N} \sum_{j=1}^{k} \frac{1}{j(N + 1 - j)} \leq \sqrt{N} \sum_{j=1}^{N} \frac{1}{j(N + 1 - j)} \leq \frac{2 \log N}{\sqrt{N}}. \tag{27}$$

Therefore, we have that, for $N$ sufficiently large and $t > 0$,

$$\mathbb{E}[e^{it(Y_{k,N} - \mathbb{E}[Y_{k,N}])}] = e^{-\left(\frac{t^2}{2} \text{var}(Y_{k,N}) + O(|t|^3 \text{var}(Y_{k,N}))\right)},$$

where $m_{k,N} = \mathbb{E}[|Y_{k,N} - \mathbb{E}[Y_{k,N}]|^3]$ denotes the third central moment. By (27),

$$m_{k,N} \leq \frac{4 \log N}{\sqrt{N}} \text{var}(Y_{k,N}).$$

Hence

$$\mathbb{E}[e^{it\sqrt{N}(Y_N - \mathbb{E}[Y_N])}] = \prod_{k=1}^{N} \mathbb{E}[e^{it(Y_{k,N} - \mathbb{E}[Y_{k,N}])}] = \prod_{k=1}^{N} e^{-\left(\frac{t^2}{2} \text{var}(Y_{k,N}) + O(|t|^3 \text{var}(Y_{k,N}))\right)}$$

$$= e^{-t^2 \sigma_{2,N}^2 / 2} e^{O(|t|^3 \sigma_{2,N}^2)} N^{-1/2 \log N}.$$

This completes the proof because

$$\sigma_{2,N}^2 = \sum_{k=1}^{N} \text{var}(Y_{k,N}) = N \text{var}\left(\sum_{k=1}^{N} \sum_{j=1}^{k} \lambda_j^{-1} 1_{\{A_k \leq j\}}\right) = N \text{var}\left(\sum_{j=1}^{N} \sum_{k=j}^{N} \lambda_j^{-1} 1_{\{A_k \leq j\}}\right)$$

$$= N \text{var}\left(\sum_{j=1}^{N} \lambda_j^{-1} B_j\right) = NT_2(N) \rightarrow \sigma_2 = 4\zeta(3) - 2\zeta(2),$$

as shown by (43) in the appendix.

**Appendix**

In Section A we prove a couple of identities formulated as lemmas. Lemma 5 to Lemma 10 are all proven in an identical way by taking differences. We therefore leave out some of the details. We will denote the partial sums in these identities by $C(N), D(N), \ldots,$ instead of $C_N, D_N, \ldots,$ in order to distinguish them from the standard notation for random variables. In Section B, we apply these identities to obtain asymptotic expressions for the variance of $X_N$ and $Y_N$. 

10
A Identities

Lemma 5 For all \( N \geq 1 \),
\[
C(N) = \sum_{j=1}^{N} \frac{1}{j} \sum_{k=N+1-j}^{N} \frac{1}{k} = \sum_{k=1}^{N} \frac{1}{k^2}. \tag{28}
\]

The identity (28) was proved in [7] by induction. We give a new and simpler proof.

Proof: Clearly, \( C(1) = 1 \) and
\[
C(N) - C(N-1) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{k} + \sum_{j=1}^{N-1} \frac{1}{j} \left( \sum_{k=N+1-j}^{N} \frac{1}{k} - \sum_{k=N-j}^{N} \frac{1}{k} \right)
= \frac{1}{N} \sum_{k=1}^{N} \frac{1}{k} + \sum_{j=1}^{N-1} \frac{1}{j} \left( \frac{1}{N} - \frac{1}{N-j} \right) = \frac{1}{N} \sum_{k=1}^{N} \frac{1}{k} - \sum_{j=1}^{N-1} \frac{1}{j(N-j)} = \frac{1}{N^2}.
\]

Summing both sides from \( N = 2 \) to \( N = M \), using \( C(1) = 1 \) and relabeling \( M \to N \) then leads to the right hand side in (28). \( \square \)

A related sum which we will need is
\[
D(N) = \sum_{j=1}^{N} \frac{1}{j} \sum_{k=1}^{j} \frac{1}{k} = \frac{1}{2} \left( \sum_{k=1}^{N} \frac{1}{k} \right)^2 + \frac{1}{2} \sum_{k=1}^{N} \frac{1}{k^2}. \tag{29}
\]

Relation (29) is straightforward by symmetry.

Lemma 6 For all \( N \geq 1 \),
\[
F(N) = \sum_{j=1}^{N} \frac{1}{j^2} \sum_{k=N+1-j}^{N} \frac{1}{k} = 2 \sum_{k=1}^{N} \frac{1}{k^3} - \sum_{k=1}^{N} \frac{1}{k} \sum_{j=1}^{k} \frac{1}{j^2}. \tag{30}
\]

Proof: Parallel to the proof of Lemma 5
\[
F(N) - F(N-1) = \frac{1}{N^2} \sum_{k=1}^{N} \frac{1}{k} + \sum_{j=1}^{N-1} \frac{1}{j^2} \left( \sum_{k=N+1-j}^{N} \frac{1}{k} - \sum_{k=N-j}^{N} \frac{1}{k} \right)
= \frac{1}{N^2} \sum_{k=1}^{N} \frac{1}{k} - \sum_{j=1}^{N-1} \frac{1}{jN(N-j)} = \frac{1}{N^2} \sum_{k=1}^{N} \frac{1}{k} - \frac{1}{N} \sum_{j=1}^{N-1} \frac{1}{j(N-j)}.
\]

Writing \( \frac{1}{N^2} \sum_{k=1}^{N} \frac{1}{k} = \frac{1}{N^2} \sum_{k=1}^{N-1} \frac{1}{k} + \frac{1}{N^2} \), and using \( \frac{1}{j(N-j)} = \frac{1}{N} \left( \frac{1}{j} + \frac{1}{N-j} \right) \) on the second summand, we find that
\[
F(N) - F(N-1) = \frac{1}{N^3} - \frac{1}{N^2} \sum_{j=1}^{N-1} \frac{1}{j^2}.
\]

As in the proof of Lemma 5 this leads to the quoted result by iteration from \( F(1) = 1 \). \( \square \)

The next lemma states a somewhat more involved identity:

Lemma 7 For all \( N \geq 1 \),
\[
G(N) = \sum_{k=1}^{N} \frac{1}{k} \left( \sum_{m=N+1-k}^{N} \frac{1}{m} \right)^2 = \sum_{k=1}^{N} \frac{1}{k^2} \sum_{m=1}^{k} \frac{1}{m}. \tag{31}
\]

11
Lemma 8

For all $N \geq 1$, we obtain the result by iteration.

Proof:

$$G(N) - G(N - 1) = \frac{1}{N} \left( \sum_{m=1}^{N} \frac{1}{m} \right)^2 + \sum_{k=1}^{N-1} \frac{1}{k} \left( \sum_{m=N+1-k}^{N-1} \frac{1}{m} + \frac{1}{N} \right)^2 - \sum_{k=1}^{N-1} \frac{1}{k} \left( \sum_{m=N+1-k}^{N-1} \frac{1}{m} + \frac{1}{N - k} \right)^2$$

$$= \frac{1}{N} \left( \sum_{m=1}^{N} \frac{1}{m} \right)^2 + \sum_{k=1}^{N-1} \frac{1}{k} \left( \frac{2}{N} - \frac{2}{N - k} \right) \sum_{m=N+1-k}^{N-1} \frac{1}{m} + \frac{1}{N^2} - \frac{1}{(N - k)^2}$$

$$= \frac{1}{N} \left( \sum_{m=1}^{N} \frac{1}{m} \right)^2 - \frac{2}{N} \sum_{k=1}^{N-1} \frac{1}{m} + \frac{N-1}{N^2} \sum_{k=1}^{N-1} \frac{1}{k} - \frac{N-1}{k (N - k)^2}.$$ 

Now, using identity (29),

$$\sum_{k=1}^{N-1} \frac{1}{(N - k)} = \sum_{m=N+1-k}^{N-1} \frac{1}{m} = D(N-1) - \sum_{k=1}^{N-1} \frac{1}{j^2} = \frac{1}{2} \left( \sum_{n=1}^{N-1} \frac{1}{n} \right)^2 - \frac{1}{2} \sum_{n=1}^{N-1} \frac{1}{n^2},$$

and the partial fractions result:

$$\frac{1}{k(N - k)^2} = \frac{1}{N^2 k} + \frac{1}{N^2 (N - k)} + \frac{1}{N (N - k)^2},$$

we arrive at

$$G(N) - G(N - 1) = \frac{1}{N} \left( \sum_{m=1}^{N} \frac{1}{m} \right)^2 - \frac{1}{N^2} \left( \sum_{n=1}^{N-1} \frac{1}{n} \right)^2 - \frac{1}{N^2} \sum_{k=1}^{N-1} \frac{1}{k} = \frac{1}{N^2} \sum_{m=1}^{N} \frac{1}{m}.$$ 

As before we obtain the result by iteration. \qed

Lemma 8 For all $N \geq 1$,

$$L(N) = \sum_{k=1}^{N} \sum_{j=1}^{k} \frac{1}{kj} \sum_{n=N+1-j}^{N} \frac{1}{n} = \sum_{k=1}^{N} \frac{1}{k^3}. \quad (33)$$

Proof: After a tedious, but straightforward, computation we get

$$L(N) - L(N - 1) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{j^2} \sum_{n=N+1-j}^{N} \frac{1}{n} - \frac{1}{N} \sum_{k=1}^{N-1} \frac{1}{k} \sum_{j=N-k}^{N-1} \frac{1}{j^2}.$$ 

With identity (28),

$$L(N) - L(N - 1) = \frac{1}{N} \sum_{j=1}^{N} \frac{1}{j^2} - \frac{1}{N} \sum_{j=1}^{N-1} \frac{1}{j^2} = \frac{1}{N^3}.$$ 

This yields the proof. \qed

Lemma 9 For all $N \geq 1$,

$$R(N) = 2 \sum_{k=1}^{N} \sum_{n=1}^{k} \frac{1}{kn} \sum_{j=N+1-n}^{N} \frac{1}{j^2} = 5 \sum_{k=1}^{N} \frac{1}{k^4} - 2 \sum_{n=1}^{N} \frac{1}{n^3} \sum_{k=1}^{N} \frac{1}{k} - \left( \sum_{k=1}^{N} \frac{1}{k^2} \right)^2. \quad (34)$$
Proof: Once more we compute the difference

\[ R(N) - R(N-1) = \frac{2}{N} \sum_{n=1}^{N} \frac{1}{n} \sum_{j=N+1-n}^{N} \frac{1}{j^2} + 2 \sum_{k=1}^{N-1} \sum_{n=1}^{k} \frac{1}{kn} \left( \frac{1}{N^2} - \frac{1}{(N-n)^2} \right) \]

\[ = \frac{2}{N} \sum_{n=1}^{N} \frac{1}{n} \sum_{j=N+1-n}^{N} \frac{1}{j^2} + 2 \sum_{k=1}^{N-1} \frac{1}{k} \sum_{n=1}^{k} \frac{1}{n} - 2 \sum_{k=1}^{N-1} \sum_{n=1}^{k} \frac{1}{n(N-n)^2}. \]

Using the partial fraction result (32) on the last sum

\[ R(N) - R(N-1) = \frac{2}{N^2} \sum_{j=1}^{N} \frac{1}{j^2} + \frac{2}{N^2} \sum_{j=N+1-k}^{N} \frac{1}{j^2} - \frac{2}{N^2} \sum_{k=1}^{N-1} \frac{1}{k^2} \sum_{m=N-k}^{N-1} \frac{1}{m^2} \]

\[ = \frac{2}{N^2} \sum_{j=1}^{N} \frac{1}{j^2} - \frac{2}{N^2} \sum_{j=1}^{N} \frac{1}{j^2} - \frac{2}{N^2} \sum_{k=1}^{N-1} \frac{1}{k^2} \sum_{m=N-k}^{N-1} \frac{1}{m^2} \]

\[ = \frac{2}{N^2} \sum_{j=1}^{N} \frac{1}{j^2} - \frac{2}{N^2} \sum_{j=1}^{N} \frac{1}{j^2} + \frac{2}{N^3} \sum_{k=1}^{N-1} \frac{1}{k} - \frac{2}{N^2} \sum_{k=1}^{N-1} \frac{1}{k(N-k)^2}, \]

where we have used (28). Using (32) to replace the last sum on the right side, we obtain,

\[ R(N) - R(N-1) = \frac{2}{N^4} - \frac{2}{N^3} \sum_{k=1}^{N-1} \frac{1}{k} - \frac{2}{N^2} \sum_{m=1}^{N} \frac{1}{m^2} = \frac{6}{N^4} - \frac{2}{N^3} \sum_{k=1}^{N} \frac{1}{k} - \frac{2}{N^2} \sum_{m=1}^{N} \frac{1}{m^2}. \]

Summing both sides, and using \( R(1) = 2, \)

\[ R(N) = 6 \sum_{k=1}^{N} \frac{1}{k^4} - 2 \sum_{n=1}^{N} \frac{1}{n^1} \sum_{k=1}^{n} \frac{1}{k} - 2 \sum_{k=1}^{N} \frac{1}{k^2} \sum_{m=1}^{N} \frac{1}{m^2}. \]

Finally by an argument parallel to (29),

\[ \sum_{k=1}^{N} \frac{1}{k^2} \sum_{m=1}^{N} \frac{1}{m^2} = \frac{1}{2} \left( \sum_{k=1}^{N} \frac{1}{k^2} \right)^2 + \frac{1}{2} \sum_{k=1}^{N} \frac{1}{k^4}. \]

(35)

Together, this yields the proof. \( \square \)

Lemma 10 For all \( N \geq 1, \)

\[ T(N) = \sum_{k=1}^{N} \frac{1}{k^2} \left( \sum_{m=N+1-k}^{N} \frac{1}{m} \right)^2 = 2 \sum_{k=1}^{N} \frac{1}{k^3} \sum_{n=1}^{k} \frac{1}{n} - \left( \sum_{k=1}^{N} \frac{1}{k^2} \right)^2. \]

(36)
Proof: As before

\[ T(N) - T(N-1) = \frac{1}{N^2} \left( \sum_{m=1}^{N} \frac{1}{m} \right)^2 + \sum_{k=1}^{N-1} \frac{1}{k^2} \left( \sum_{m=N+1-k}^{N-1} \frac{1}{m} + \frac{1}{N} \right)^2 - \sum_{k=1}^{N-1} \frac{1}{k^2} \left( \sum_{m=N+1-k}^{N-1} \frac{1}{m} + \frac{1}{N-k} \right)^2 \]

\[ = \frac{1}{N^2} \left( \sum_{m=1}^{N} \frac{1}{m} \right)^2 + \sum_{k=1}^{N-1} \frac{1}{k^2} \left( \frac{2}{N} - \frac{2}{N-k} \right) \sum_{m=N+1-k}^{N-1} \frac{1}{m} + \frac{1}{N^2} - \frac{1}{(N-k)^2} \]

\[ = \frac{1}{N^2} \left( \sum_{m=1}^{N} \frac{1}{m} \right)^2 - \frac{2}{N} \sum_{k=1}^{N-1} \frac{1}{k(N-k)} \sum_{m=N+1-k}^{N-1} \frac{1}{m} + \frac{1}{N^2} \sum_{k=1}^{N-1} \frac{1}{k^2} - \sum_{k=1}^{N-1} \frac{1}{k^2 (N-k)^2}. \]

Now from taking partial fractions, (28) and (29),

\[ \sum_{k=1}^{N-1} \frac{1}{k(N-k)} \sum_{m=N+1-k}^{N-1} \frac{1}{m} = C(N-1) \]

\[ \sum_{k=1}^{N-1} \frac{1}{m} \sum_{k=1}^{N-1} \frac{1}{m} = C(N) - \frac{2}{N^2} \sum_{k=1}^{N-1} \frac{1}{k} + \frac{D(N-1)}{N} - \frac{1}{N} \sum_{m=1}^{N-1} \frac{1}{m^2}. \]

We can simplify this using the expressions for \( C(N) \) in (28) and \( D(N) \) in (29),

\[ \sum_{k=1}^{N-1} \frac{1}{k(N-k)} \sum_{m=N+1-k}^{N-1} \frac{1}{m} = -\frac{2}{N^2} \sum_{k=1}^{N-1} \frac{1}{k} + \frac{1}{2N} \left( \sum_{n=1}^{N-1} \frac{1}{n} \right)^2 + \frac{1}{2N} \sum_{n=1}^{N-1} \frac{1}{n^2}. \]

Using partial fraction expansion again yields

\[ \sum_{k=1}^{N-1} \frac{1}{k^2 (N-k)^2} = \frac{4}{N^3} \sum_{k=1}^{N-1} \frac{1}{k} + \frac{2}{N^2} \sum_{k=1}^{N-1} \frac{1}{k^2} \]

(37)

Combining these results then gives

\[ T(N) - T(N-1) = \frac{1}{N^2} \left( \sum_{n=1}^{N} \frac{1}{n} \right)^2 - \frac{1}{N^2} \left( \sum_{n=1}^{N-1} \frac{1}{n} \right)^2 + \frac{4}{N^3} \sum_{k=1}^{N-1} \frac{1}{k} - \frac{1}{N^2} \sum_{n=1}^{N-1} \frac{1}{n^2} + \frac{1}{N^2} \sum_{k=1}^{N-1} \frac{1}{k^2} \]

\[ - \frac{4}{N^3} \sum_{k=1}^{N-1} \frac{1}{k} \]

\[ = \frac{1}{N^2} \left( \sum_{n=1}^{N-1} \frac{1}{n} \right)^2 - \frac{1}{N^2} \left( \sum_{n=1}^{N-1} \frac{1}{n} \right)^2 \]

\[ = \frac{2}{N^3} \sum_{n=1}^{N-1} \frac{1}{n} + \frac{1}{N^4} - \frac{2}{N^2} \sum_{k=1}^{N-1} \frac{1}{k^2} \]

\[ = \frac{2}{N^3} \sum_{n=1}^{N-1} \frac{1}{n} + \frac{1}{N^4} - 2 \frac{1}{N^2} \sum_{k=1}^{N-1} \frac{1}{k^2}. \]

from which by summing both sides from \( N = 2 \) to \( N = M \) (and then \( M \to N \) again)

\[ T(N) - 1 = 2 \sum_{k=2}^{N} \frac{1}{k^3} \sum_{n=1}^{k} \frac{1}{n} + \sum_{k=2}^{N} \frac{1}{k^3} - 2 \sum_{k=2}^{N} \frac{1}{k^2} \sum_{n=1}^{k} \frac{1}{n^2}. \]

Using (35) then yields the proof. \( \square \)
B The asymptotic results for the variances

In this section we use the identities of Section A, in order to compute simplified expressions for \( T_2(N) \) and \( T_1(N) \). Consequently we use these results for the asymptotic variance of \( W_N \).

The sum \( T_2(N) \) (compare (17)) equals

\[
T_2(N) = R_1(N) + R_2(N),
\]

where

\[
R_1(N) = 2 \sum_{n=1}^{N} \sum_{m=n}^{N} \frac{1}{(N+1-n) \, m \, (N+1-m)} \sum_{j=m}^{N} \frac{1}{j},
\]

and

\[
R_2(N) = -2 \sum_{n=1}^{N} \sum_{m=n}^{N} \frac{1}{(N+1-n) \, (N+1-m)} \sum_{j=m}^{N} \frac{1}{j^2}.
\]

We start with the sum \( R_1(N) \), and interchange the sums to obtain

\[
R_1(N) = 2 \sum_{k=1}^{N} \sum_{m=1}^{N} \frac{1}{k} \sum_{n=1}^{N} \frac{1}{m} \sum_{j=n+1-k}^{N} \frac{1}{j} + \frac{2}{N+1} \sum_{k=1}^{N} \sum_{m=N+1-k}^{N} \frac{1}{m} \sum_{j=m}^{N} \frac{1}{j}. \tag{38}
\]

Splitting \((n+1-n)\)^{-1} into two parts,

\[
R_1(N) = \frac{2}{N+1} \sum_{k=1}^{N} \sum_{n=1}^{N} \frac{1}{k} \sum_{j=n+1-k}^{N} \frac{1}{j} + \frac{2}{N+1} \sum_{k=1}^{N} \sum_{m=N+1-k}^{N} \frac{1}{m} \sum_{j=m}^{N} \frac{1}{j}. \tag{39}
\]

The first sum equals \(2L(N)/(N+1)\), where \(L(N)\) was simplified in Lemma 9.

By the same method that we used in (29) to obtain \(D(N)\), we find

\[
\sum_{m=N+1-k}^{N} \frac{1}{m} \sum_{j=n+1-k}^{N} \frac{1}{j} = \frac{1}{2} \left( \sum_{j=N+1-k}^{N} \frac{1}{j} \right)^2 + \frac{1}{2} \sum_{j=N+1-k}^{N} \frac{1}{j^2}. \tag{40}
\]

This implies using (30),

\[
\sum_{k=1}^{N} \sum_{m=N+1-k}^{N} \frac{1}{m} \sum_{j=n+1-k}^{N} \frac{1}{j} = \frac{1}{2} \sum_{k=1}^{N} \left( \sum_{m=N+1-k}^{N} \frac{1}{m} \right)^2 + \frac{1}{2} \sum_{k=1}^{N} \sum_{m=N+1-k}^{N} \frac{1}{m^2} \tag{41}
\]

Combining all,

\[
R_1(N) = \frac{4}{N+1} \sum_{k=1}^{N} \frac{1}{k^3} = \frac{4\zeta(3)}{N}(1+O(N^{-2})).
\]
We now turn to the sum

\[ R_2(N) = 2 \sum_{n=1}^{N} \sum_{m=n}^{N} \frac{1}{(N+1-n)(N+1-m)} \sum_{j=m}^{N} \frac{1}{j^2} \]

\[ = 2 \sum_{k=1}^{N} \sum_{m=N+1-k}^{N} \frac{1}{k(N+1-m)} \sum_{j=m}^{N} \frac{1}{j^2} = 2 \sum_{k=1}^{N} \sum_{n=1}^{k} \frac{1}{kn} \sum_{j=N+1-n}^{N} \frac{1}{j^2} = R(N), \]

and \( R(N) \) was simplified in Lemma 9.

Together we find

\[ T_2(N) = R_1(N) - R_2(N) = R_1(N) - R(N) \]

\[ = \frac{4}{N+1} \sum_{k=1}^{N} \frac{1}{k^3} - 5 \sum_{k=1}^{N} \frac{1}{k^4} + 2 \sum_{n=1}^{N} \frac{1}{n^3} \sum_{k=1}^{n} \frac{1}{k} + \left( \sum_{k=1}^{N} \frac{1}{k^2} \right)^2. \]  

(42)

From (??) we find that \( R(N) - R(N-1) = \frac{-2\zeta(2)}{N^2} + O(\frac{\log N}{N}) \), so that, by summation,

\[ R(N) = R + \frac{2\zeta(2)}{N} + O\left( \frac{\log N}{N^2} \right), \]

where

\[ R = 5 \sum_{k=1}^{\infty} \frac{1}{k^4} - 2 \sum_{n=1}^{\infty} \sum_{k=1}^{n} \frac{1}{k} - \left( \sum_{k=1}^{\infty} \frac{1}{k^2} \right)^2 = 5 \zeta(4) - \zeta(2)^2 = 0, \]

The first equality follows by (19), while the second follows by [3, (9.542)1]. Thus, we obtain the asymptotics

\[ T_2(N) = R_1(N) - R(N) = \frac{4\zeta(3) - 2\zeta(2)}{N} + O\left( \frac{\log N}{N^2} \right). \]  

(43)

We finally turn to the second sum \( T_1(N) \) (see (16)), which sum is equal to the sum \( T(N) \) displayed in Lemma 10. Therefore,

\[ T_1(N) = \sum_{n=1}^{N} \frac{1}{(N+1-n)^2} \left( \sum_{j=n}^{N} \frac{1}{j} \right)^2 = 2 \sum_{k=1}^{N} \frac{1}{k^3} \sum_{n=1}^{k} \frac{1}{n} - \left( \sum_{k=1}^{N} \frac{1}{k^2} \right)^2. \]  

(44)

From the proof of Lemma 10, the difference

\[ T_1(N) - T_1(N-1) = \frac{2}{N^3} \sum_{n=1}^{N} \frac{1}{n} + \frac{1}{N^4} - \frac{2}{N^2} \sum_{k=1}^{N} \frac{1}{k^2} = - \frac{2}{N^2} \sum_{k=1}^{N} \frac{1}{k^2} + O(N^{-3}), \]

which shows, by summation that for large \( N \), \( T_1(N) \) behaves asymptotically as

\[ T_1(N) = T_1 + \frac{2\zeta(2)}{N} + O(N^{-2}), \]  

(45)

where we write

\[ T_1 = 2 \sum_{k=1}^{\infty} \frac{1}{k^3} \sum_{n=1}^{k} \frac{1}{n} - \left( \sum_{k=1}^{\infty} \frac{1}{k^2} \right)^2 = 5 \cdot \frac{\zeta(4) - \zeta(2)^2}{2} = 0, \]  

(46)

and, again, the first equality follows by (19), while the second follows by [3, (9.542)1].

16
Acknowledgements

The work of RvdH was supported in part by the Netherlands Organisation for Scientific Research (NWO).

References


