ON THE NEAREST-NEIGHBOR ALGORITHM FOR THE MEAN-FIELD TRAVELING SALESMAN PROBLEM

ANTAR BANDYOPADHYAY,* ** *Indian Statistical Institute, Delhi and Kolkata*

FARKHONDEH SAJADI,* *** *Indian Statistical Institute, Delhi*

Abstract

In this work we consider the *mean-field traveling salesman problem*, where the intercity distances are taken to be independent and identically distributed with some distribution F. We consider the simplest approximation algorithm, namely, the *nearest-neighbor algorithm*, where the rule is to move to the nearest nonvisited city. We show that the limiting behavior of the total length of the nearest-neighbor tour depends on the scaling properties of the density of F at 0 and derive the limits for all possible cases of general F.

Keywords: Nearest-neighbor algorithm; mean-field setup; traveling salesman problem

2010 Mathematics Subject Classification: Primary 60K37
Secondary 05C85; 68Q87; 68W25

1. Introduction

The *traveling salesman problem* (TSP) is one of the most well-known combinatorial optimization problems that has attracted attention in various scientific fields. The aim is to find the shortest tour, connecting a number of cities visited by a traveling salesman on his/her sales route, such that he/she visits each city exactly once and finally returns to the starting city. Formally, we are given a set of *cities* labeled $\{1, 2, \ldots, n\}$, say, and, for each $\{i, j\}$ pair of distinct cities, distance L_{ij} is assumed to be symmetric. The goal is to find a permutation π of the cities that solves the minimization problem

$$\min_{\pi} \sum_{i=1}^{n} L_{\pi(i), \pi(i+1)},$$

where $\pi(n+1) = 1$. This minimum value is called the *optimal tour length* which we denote by T_n^{opt}.

It is well known in the literature on algorithms [4] that the TSP in general is an *NP-complete* problem. So there are several approximate algorithms which try to approximate the optimal tour with polynomial running time. Among them, one of the simplest is the *nearest-neighbor (NN) algorithm* [1], which is also known as the *next best method*; see [3]. The algorithm starts at one of the cities and then always adds to the tour the nearest not yet visited city, and terminates when every city has been added. Since the NN algorithm is designed to move to the nearest

nonvisited city, having starting from say the first city $v_1 = 1$, we need to find the next city v_2 by solving
$$\arg \min\{L_{12}, L_{13}, \ldots, L_{1n}\}.$$

Then from city v_2 we find the next nonvisited nearest city v_3 by solving
$$\arg \min\{L_{v_2 u} \mid u \in \{2, 3, \ldots, n\} \setminus \{v_2\}\}.$$

We continue the algorithm until all n cities have been visited and the last edge to be added to the tour is the edge between the cities v_1 and v_n. Define T_n^{NN} to be the length of the NN tour among n cities in the TSP. Then
$$T_n^{\text{NN}} = \sum_{i=1}^{n} L_{v_i v_{i+1}},$$

where $v_1 = 1 = v_{n+1}$.

There are several randomized versions of this problem where the distances are taken to be random. We note that the version in which the distances between pairs of cities are assumed to be independent and identically distributed (i.i.d.) random variables with a given distribution F is of special interest in statistical physics. Such a model is known as the *mean-field TSP*; see [6], [8], and [7]. Asymptotic results for the optimal solution of the mean-field TSP have been obtained. Wästlund [8] proved that if $(L_{ij})_{1 \leq i, j \leq n}$ are i.i.d. positively supported random variables with $\lim_{t \to 0^+} \mathbb{P}(L_{ij} < t)/t = 1$ then
$$T_n^{\text{opt}} \xrightarrow{\mathbb{P}} \frac{1}{2} \int_0^{\infty} h(x) \, \mathrm{d}x, \tag{1}$$

where h is implicitly defined through the equation
$$\left(1 + \frac{x}{2}\right)\mathrm{e}^{-x} + \left(1 + \frac{h(x)}{2}\right)\mathrm{e}^{-h(x)} = 1.$$

Further generalizations of this can be found in [7] where the assumption has been relaxed to include i.i.d. Weibull-distributed distances.

In this paper we study the limiting behavior of the total length of the tour, obtained by the NN algorithm for the mean-field TSP. Our aim is to compare the apparent 'loss' accrued (that is, the increased distance traversed) with respect to the optimal solution when using the NN algorithm. Owing to (1), it is enough to consider the limiting behavior of T_n^{NN}. Similar studies have been conducted for the TSP when the distances are taken to be constant but defined through a metric. For cities placed in a metric space, where the intercity distances are given by the metric, it was been proved in [5] that
$$\frac{T_n^{\text{NN}}}{T_n^{\text{opt}}} \leq \frac{1}{2} \log_2 n + \frac{1}{2}.$$

We show that if F, the distribution of the distance between cities, has a density which is continuous at 0 with $F'(0+) > 0$, then the total length of the NN tour, for the mean field TSP, scales as $\log n$. This parallels the conclusions drawn in [5]. Moreover, we also consider a general distribution function F with nonnegative support and show that the asymptotic behaviors for T_n^{NN} depend on the scaling properties of the density near 0. It is worth noting that in [6] the authors considered a similar study and derived asymptotic results for $\mathbb{E}[T_n^{\text{NN}}]$, under the assumption that the intercity distances are i.i.d. with gamma distribution. Our work is more general and includes the results derived in [6] as special cases.

The rest of the paper is organized as follows. In Section 2 we state our main results, whose proofs are presented in Section 5. Section 3 is devoted to the asymptotic property of the first and the last edges of the NN tour. In Section 4 we present three auxiliary results and their proofs. Finally, in Section 6 we discuss the possible relaxation of the assumptions on the distribution F.

2. Main results

We will assume that the mean and the variance of F are finite, and that F has a density f. Throughout this paper, we use the notation $a_n \sim b_n$ if $\lim_{n\to\infty} a_n/b_n = 1$. Our first result shows that T_n^{NN} is 'close' to its expected value.

Theorem 1. *Assume that, as $t \to 0^+$, $f(t)/t^\alpha \to C$, where $C \in (0, \infty)$ is constant and $-1 < \alpha < 1$. Then, as $n \to \infty$,*

$$\{T_n^{\mathrm{NN}} - \mathbb{E}[T_n^{\mathrm{NN}}]\}_{n \geq 1} \quad \text{converges weakly.}$$

The next three results address three different cases of the behavior of f near 0.

Theorem 2. *Assume that, as $t \to 0^+$, $f(t) \to f(0)$, where $f(0) \in (0, \infty)$. Then, as $n \to \infty$,*

$$\frac{T_n^{\mathrm{NN}}}{\log n} \xrightarrow{\mathbb{P}} \frac{1}{f(0)} \tag{2}$$

and

$$\mathbb{E}[T_n^{\mathrm{NN}}] \sim \frac{1}{f(0)} \log n. \tag{3}$$

Moreover, the convergence in (2) happens in $\mathcal{L}_2$.

The following corollary is a finer result for i.i.d. exponential distances, which was also observed in [6].

Corollary 1. *In the mean-field TSP, suppose that F is the exponential distribution with mean 1. Then $T_n^{\mathrm{NN}} - \log n$ converges weakly.*

Theorem 3. *Assume that, as $t \to 0+$, $f(t)/t^\alpha \to C$, where $C > 0$ is constant and $0 < \alpha < 1$. Then, as $n \to \infty$,*

$$\frac{T_n^{\mathrm{NN}}}{n^{1-1/(1+\alpha)}} \xrightarrow{\mathbb{P}} K_\alpha, \tag{4}$$

where $K_\alpha := ((1+\alpha)/C)^{1/(1+\alpha)}(1/\alpha)\Gamma(1/(1+\alpha))$ and

$$\mathbb{E}[T_n^{\mathrm{NN}}] \sim K_\alpha n^{1-1/(1+\alpha)}. \tag{5}$$

Moreover, the convergence in (4) happens in $\mathcal{L}_2$.

In [6], (5) was derived with the same constant and for any $\alpha > 0$.

Theorem 4. *Let $-1 < \alpha < 0$, and assume that, as $t \to 0^+$, $f(t)/t^\alpha \to C$, where $C > 0$ is constant. Then the sequence $\{\mathbb{E}[T_n^{\mathrm{NN}}]\}_{n \geq 1}$ is a convergent sequence and T_n^{NN} converges weakly.*

The above results cover the cases where $|\alpha| < 1$. Note that the case $\alpha \leq -1$ cannot happen, since f is a density function. For $\alpha \geq 1$, we first consider the case when F is the Weibull distribution with shape parameter $1 + \alpha$ and scale parameter 1, which was also considered in [7]. In this case we show that not only does $\mathbb{E}[T_n^{\mathrm{NN}}]$ scale as $n^{1-1/(1+\alpha)}$ but, in fact, after centering and proper scaling, T_n^{NN} converges weakly to a normal distribution.

Theorem 5. *Let $\alpha \geq 1$, and, for $1 \leq i \leq n-1$, let the intercity distances $\{L_{ij}\}_{i<j\leq n}$, in the mean-field TSP, be i.i.d. Weibull distributed with shape parameter $1 + \alpha$ and scale parameter 1, that is, the density function is given by*

$$f(t) = (1 + \alpha)t^{\alpha}e^{-t^{(1+\alpha)}}\mathbf{1}[t > 0].$$

Then, as $n \to \infty$,

$$\frac{T_n^{\mathrm{NN}}}{n^{1-1/(1+\alpha)}} \xrightarrow{\mathbb{P}} \frac{1}{\alpha}\Gamma\left(\frac{1}{1+\alpha}\right) \tag{6}$$

and

$$\mathbb{E}[T_n^{\mathrm{NN}}] \sim \frac{1}{\alpha}\Gamma\left(\frac{1}{1+\alpha}\right)n^{1-1/(1+\alpha)}. \tag{7}$$

Furthermore, as $n \to \infty$, for $\alpha > 1$,

$$\frac{T_n^{\mathrm{NN}} - \mathbb{E}[T_n^{\mathrm{NN}}]}{n^{1/2-1/(1+\alpha)}} \xrightarrow{\mathrm{D}} N\left(0, \frac{\alpha+1}{\alpha-1}\sigma^2(\alpha)\right) \tag{8}$$

and, for $\alpha = 1$,

$$\frac{T_n^{\mathrm{NN}} - \mathbb{E}[T_n^{\mathrm{NN}}]}{\sqrt{\log n}} \xrightarrow{\mathrm{D}} N(0, \sigma^2(\alpha)), \tag{9}$$

where $\sigma^2(\alpha) = \Gamma(2/(1+\alpha) + 1) - \Gamma^2(1 + 1/(1+\alpha))$.

Finally, we have the following result for the general case $\alpha \geq 1$.

Theorem 6. *Assume that, as $t \to 0^+$, $f(t)/t^{\alpha} \to C$, where $C > 0$ is constant and $\alpha \geq 1$. Then, as $n \to \infty$, (4) holds.*

3. The last and the first edges of the NN tour

Before we prove the main results, in this section we show that the first and the last edges of the NN tour are asymptotically constant. Let the distances between cities be denoted by $\{(L_{ij})_{i<j\leq n}\}_{1\leq i\leq n-1}$, which are i.i.d. with distribution F supported on $[0, \infty)$ and have density f. Let L_n^{last} be the length of the last edge, which joins the last visited city to the first city, and let $L_n^{\mathrm{first}} := \min_{1<j\leq n} L_{1j}$ be the length of the first edge. Then the length of the NN tour, T_n^{NN}, can be written as

$$T_n^{\mathrm{NN}} \stackrel{\mathrm{D}}{=} \sum_{i=2}^{n-1} \min_{i<j\leq n} L_{ij} + L_n^{\mathrm{first}} + L_n^{\mathrm{last}}. \tag{10}$$

The following proposition shows that the sum of the lengths of the last and first edges in the NN tour do not play an important role.

Proposition 1. *In the NN tour for the mean-field TSP, the distribution function of $L_n^{\mathrm{first}} + L_n^{\mathrm{last}}$ converges to F as $n \to \infty$ and $\sum_{i=2}^{n-1} \min_{i<j\leq n} L_{ij}$ is independent of $L_n^{\mathrm{first}} + L_n^{\mathrm{last}}$. Moreover, as $n \to \infty$,*

$$\mathbb{E}[L_n^{\mathrm{first}} + L_n^{\mathrm{last}}] \to \mu \quad and \quad \mathbb{E}[(L_n^{\mathrm{first}} + L_n^{\mathrm{last}})^2] \to \mu^2 + \sigma^2,$$

where μ and σ^2 are the mean and the variance of F.

Proof. For $k = 1, 2, \ldots, n-1$, let $X_k := L_{1k+1}$ and $X_{(k)}$ be the kth order statistic of $X_1, X_2, \ldots, X_{n-1}$. By construction, the successive vertices $1 = v_1, v_2, v_3, \ldots, v_n$ have the property that, for every $2 \leq k \leq n$, given $\{v_2, v_3, \ldots, v_{k-1}\}$, the vertex v_k is uniformly distributed on the set $\{1, 2, \ldots, n\} \setminus \{1, v_2, v_3, \ldots, v_{k-1}\}$. So, in particular, the last vertex of the tour v_n is also uniformly distributed on the set $\{2, 3, \ldots, n\} \setminus \{v_2\}$. Hence, given $X_1, X_2, \ldots, X_{n-1}$, the length of the last edge is uniform on $\{X_{(2)}, X_{(3)}, \ldots, X_{(n-1)}\}$. So, for any bounded continuous function h, we have

$$\mathbb{E}[h(L_n^{\text{last}})] = \frac{1}{n-2} \sum_{k=2}^{n-1} \mathbb{E}[h(X_{(k)})] = \frac{n-1}{n-2} \mathbb{E}[h(X_1)] - \frac{\mathbb{E}[h(X_{(1)})]}{n-2}.$$

Therefore,

$$\lim_{n \to \infty} \mathbb{E}[h(L_n^{\text{last}})] = \mathbb{E}[h(X_1)].$$

This proves that the distribution function of L_n^{last} converges to F as $n \to \infty$. Now observe that $L_n^{\text{first}} \to 0$ almost surely (a.s.); therefore, by Slutsky's theorem, the distribution function of $L_n^{\text{first}} + L_n^{\text{last}}$ converges to F as $n \to \infty$.

Now observe that, by similar calculations as above,

$$\mathbb{E}[L_n^{\text{first}} + L_n^{\text{last}}] = \frac{n-1}{n-2} \mathbb{E}[X_1] + \frac{n-3}{n-2} \mathbb{E}[X_{(1)}] \to \mu.$$

The last limit follows from the dominated convergence theorem since $X_{(1)} \to 0$ a.s. and $0 \leq X_{(1)} \leq X_1$.

Furthermore,

$$\mathbb{E}[(L_n^{\text{last}})^2] = \frac{n-1}{n-2} \mathbb{E}[X_1^2] - \frac{\mathbb{E}[X_{(1)}^2]}{n-2} \to \mu^2 + \sigma^2$$

and

$$\mathbb{E}[(L_n^{\text{first}})^2] = \mathbb{E}[X_{(1)}^2] \to 0.$$

Finally,

$$\mathbb{E}[L_n^{\text{first}} L_n^{\text{last}}] = \frac{n-1}{n-2} \mathbb{E}[X_{(1)} \bar{X}_{n-1}] - \frac{\mathbb{E}[X_{(1)}^2]}{n-2} \leq \sqrt{\mathbb{E}[X_{(1)}^2] \mathbb{E}[\bar{X}_{n-1}^2]} - \frac{\mathbb{E}[X_{(1)}^2]}{n-2} \to 0,$$

where $\bar{X}_{n-1} := (1/(n-1)) \sum_{k=1}^{n-1} X_k$ and the inequality follows by the Cauchy–Schwarz inequality. Combining all these we have

$$\mathbb{E}[(L_n^{\text{first}} + L_n^{\text{last}})^2] \to \mu^2 + \sigma^2.$$

4. Auxiliary results

For the distribution function F, we define $F^{-1} \colon (0, 1) \to [0, \infty)$ by $F^{-1}(u) := \inf\{x \in \mathbb{R} \mid F(x) \geq u\}$, $0 < u < 1$. It is then a standard fact that $F^{-1}(U) \sim F$ when $U \sim \text{uniform}[0, 1]$. We start with a lemma which will give a useful representation of T_n^{NN}.

Lemma 1. *Let the distances between cities, $(L_{ij})_{i<j\leq n}$ for $i = 1, \ldots, n-1$, be i.i.d. with F denoting its common distribution function. Define the random variable $W_i := F^{-1}(1 - \exp(-Y_i/i))$, where $\{Y_i\}_{1 \leq i \leq n-1}$ are i.i.d. exponential random variables each with mean 1.*

Then

$$T_n^{\text{NN}} \overset{\text{D}}{=} \sum_{i=1}^{n-2} W_i + R_n,\tag{11}$$

where $R_n \overset{\text{D}}{=} L_n^{\text{first}} + L_n^{\text{last}}$ and is independent of $\{W_i\}_{i=1}^{n-2}$.

The proof of the lemma is straightforward, so we omit it, but we note that this lemma translates the problem of studying the asymptotic of T_n^{NN} to the problem of studying the asymptotic of the sum of successive minimums of independent sets of random variables. Our next result is the most important result for deriving all the $\mathcal{L}_2$-convergences.

Lemma 2. *Assume that F has density f and that, as $t \to 0^+$, $f(t)/t^\alpha \to C$, where $C \in (0, \infty)$ is constant and $-1 < \alpha < 1$. Then, as $n \to \infty$, $\{\sum_{i=1}^{n-2}(W_i - \mathbb{E}[W_i])\}_{n\geq 1}$ converges a.s. and in $\mathcal{L}_2$.*

Proof. By assumption, as $t \to 0^+$, $f(t)/t^\alpha \to C$; therefore, given $\varepsilon > 0$, there exists $\delta > 0$, such that, for all $0 < t < \delta$, we have $(C - \varepsilon)t^\alpha < f(t) < (C + \varepsilon)t^\alpha$. Hence, for $0 < x < \delta$, $((C - \varepsilon)/(1 + \alpha))x^{1+\alpha} < F(x) < ((C + \varepsilon)/(1 + \alpha))x^{1+\alpha}$. So

$$\left(\frac{1+\alpha}{C+\varepsilon}\right)^{1/(1+\alpha)} x^{1/(1+\alpha)} < F^{-1}(x) < \left(\frac{1+\alpha}{C-\varepsilon}\right)^{1/(1+\alpha)} x^{1/(1+\alpha)}.\tag{12}$$

Put $\delta_1 := -\ln(1 - \delta)$. If $Y_i/i < \delta_1$ (which ensures that $1 - \exp(-Y_i/i) < \delta$) then we have

$$W_i \mathbf{1}\left[\frac{Y_i}{i} < \delta_1\right] < \left(\frac{1+\alpha}{C-\varepsilon}\right)^{1/(1+\alpha)}\left(1 - \exp\left(-\frac{Y_i}{i}\right)\right)^{1/(1+\alpha)} \mathbf{1}\left[\frac{Y_i}{i} < \delta_1\right].$$

Observe that, for $\beta > 0$,

$$\mathbb{E}\left[\left(1 - \exp\left(-\frac{Y_i}{i}\right)\right)^\beta\right] = i\int_0^1 u^\beta (1-u)^{i-1}\,\mathrm{d}u = \frac{\Gamma(1+\beta)\Gamma(i+1)}{\Gamma(i+1+\beta)} \leq \frac{\Gamma(2+\beta)}{(i+1+\beta)^\beta}.$$

The last inequality follows from Wendel's double inequality [9], which states that, for real $x > 0$ and $0 < s < 1$,

$$\frac{x}{(x+s)^{1-s}}\Gamma(x) \leq \Gamma(x+s) \leq x^s \Gamma(x).$$

Therefore,

$$\mathbb{E}\left[W_i^2 \mathbf{1}\left[\frac{Y_i}{i} < \delta_1\right]\right] < \left(\frac{1+\alpha}{C-\varepsilon}\right)^{2/(1+\alpha)} \Gamma\left(2 + \frac{2}{1+\alpha}\right)\frac{1}{(i+1+2/(1+\alpha))^{2/(1+\alpha)}}.\tag{13}$$

Now, since $\mathbb{E}[Y_1] < \infty$, we have, as $i \to \infty$, $Y_i/i \to 0$ a.s. Define

$$I_0(\omega) := \min\left\{i \,\Big|\, \frac{Y_j(\omega)}{j} < \delta_1 \text{ for all } j \geq i\right\}.$$

Fix $m > 1$. Then

$$[I_0 = m] = \left[\frac{Y_i}{i} < \delta_1 \text{ for all } i \geq m \text{ and } \frac{Y_{m-1}}{m-1} > \delta_1\right].$$

Hence,

$$\mathbb{P}(I_0 = m) \leq e^{-(m-1)\delta_1}.$$

Now

$$\sum_{i=1}^{\infty} \mathbb{E}[W_i^2] = \sum_{m=1}^{\infty} \mathbb{E}\left[\sum_{i=1}^{m-1} W_i^2 \mathbf{1}[I_0 = m]\right] + \sum_{m=1}^{\infty} \mathbb{E}\left[\sum_{i=m}^{\infty} W_i^2 \mathbf{1}[I_0 = m]\right].$$

But,

$$\mathbb{E}\left[\sum_{i=1}^{m-1} W_i^2 \mathbf{1}[I_0 = m]\right] = \mathbb{E}\left[\sum_{i=1}^{m-2} W_i^2 \mathbf{1}[I_0 = m]\right] + \mathbb{E}[W_{m-1}^2 \mathbf{1}[I_0 = m]].$$

Since $[I_0 = m]$ depends on the random variables $Y_{m-1}, Y_m, Y_{m+1}, \ldots$, for $1 \leq i \leq m-2$, W_i is independent of $[I_0 = m]$. Hence,

$$\mathbb{E}\left[\sum_{i=1}^{m-2} W_i^2 \, \mathbf{1}[I_0 = m]\right] \leq e^{-(m-1)\delta_1} \sum_{i=1}^{m-2} \mathbb{E}[W_i^2].$$

Since $\mathbb{E}[W_i^2]$ is a decreasing sequence, we have $\sum_{i=1}^{m-2} \mathbb{E}[W_i^2] \leq (m-2)\mathbb{E}[W_1^2]$. Therefore,

$$\mathbb{E}\left[\sum_{i=1}^{m-2} W_i^2 \, \mathbf{1}[I_0 = m]\right] \leq (m-2)e^{-(m-1)\delta_1}\mathbb{E}[W_1^2]. \tag{14}$$

Now,

$$\mathbb{E}[W_{m-1}^2 \mathbf{1}[I_0 = m]] \leq \sqrt{\mathbb{E}[W_{m-1}^4]\mathbb{P}(I_0 = m)} \leq \mu^2 e^{-(m-1)\delta_1/2}, \tag{15}$$

where the first inequality follows by the Cauchy–Schwarz inequality and the second follows from $\mathbb{E}[W_{m-1}^4] \leq \mathbb{E}[W_4^4] \leq \mu^4$ for $m > 4$. From (14) and (15), we have

$$\sum_{m=1}^{\infty} \mathbb{E}\left[\sum_{i=1}^{m-1} W_i^2 \, \mathbf{1}[I_0 = m]\right] < \infty. \tag{16}$$

Now, by assumption, since $|\alpha| < 1$, we have $2/(1 + \alpha) > 1$; therefore, for $i \geq m$, from inequality (13) we have

$$\mathbb{E}\left[\sum_{i=m}^{\infty} W_i^2 \, \mathbf{1}[I_0 = m]\right] < K e^{-(m-1)\delta_1}, \tag{17}$$

where K is a positive constant. Hence, from (16) and (17), we conclude that

$$\sum_{i=1}^{\infty} \mathbb{E}[W_i^2] < \infty.$$

Therefore, $\mathrm{var}[\sum_{i=1}^{n} W_i]$ is bounded for all n. This shows that $\sum_{i=1}^{n-2}(W_i - \mathbb{E}[W_i])$, as a martingale, converges a.s. and in $\mathcal{L}_2$. This completes the proof.

Our next lemma gives an expression for the mean of T_n^{NN} in terms of the distribution function F. As the proof of this result is immediate from representation (11), we omit the details.

Lemma 3. *Consider a mean-field TSP with i.i.d. edge weights with distribution F which is supported on* $[0, \infty)$. *Then*

$$\mathbb{E}[T_n^{\mathrm{NN}}] = \int_0^\infty \frac{[\bar{F}(t)]^2[1 - (\bar{F}(t))^{n-2}]}{F(t)} \, \mathrm{d}t + \mathbb{E}[L_n^{\mathrm{first}} + L_n^{\mathrm{last}}].$$

5. Proofs of the main results

5.1. Proof of Theorem 1

From (10) we have

$$T_n^{\mathrm{NN}} - \mathbb{E}[T_n^{\mathrm{NN}}] \stackrel{\mathrm{D}}{=} \sum_{i=2}^{n-1} \min_{i < j \leq n} L_{ij} - \mathbb{E}\left[\sum_{i=2}^{n-1} \min_{i < j \leq n} L_{ij}\right] + L_n^{\mathrm{first}} + L_n^{\mathrm{last}} - \mathbb{E}[L_n^{\mathrm{first}} + L_n^{\mathrm{last}}].$$

But, by Lemma 2 and Lemma 1,

$$\left\{\sum_{i=2}^{n-1} \min_{i < j \leq n} L_{ij} - \mathbb{E}\left[\sum_{i=2}^{n-1} \min_{i < j \leq n} L_{ij}\right]\right\}_{n > 1}$$

converges in $\mathcal{L}_2$ and, hence, by Proposition 1, $\{T_n^{\mathrm{NN}} - \mathbb{E}[T_n^{\mathrm{NN}}]\}_{n>1}$ converges weakly.

5.2. Proofs of Theorem 2 and Theorem 3

By the assumptions of both of the theorems, $f(t)/t^\alpha \to C$ as $t \to 0+$, where $0 \leq \alpha < 1$. Hence, using inequality (12) we have, as $i \to \infty$,

$$\frac{(C/(1+\alpha))^{1/(1+\alpha)} W_i}{(Y_i/i)^{1/(1+\alpha)}} \to 1 \quad \mathrm{a.s.},$$

where the Y_i are i.i.d. exponential random variables each with mean 1 and $W_i = F^{-1}(1 - \exp(-Y_i/i))$. Therefore, as $n \to \infty$,

$$\frac{(C/(1+\alpha))^{1/(1+\alpha)} \sum_{i=1}^{n-2} W_i}{\sum_{i=1}^{n-2} (Y_i/i)^{1/(1+\alpha)}} \to 1 \quad \mathrm{a.s.} \tag{18}$$

Note that the above equation still holds when $\alpha \geq 1$.

Since $0 \leq \alpha < 1$, we have $2/(1+\alpha) > 1$; thus, $\mathrm{var}(\sum_{i=1}^{n-2}(Y_i/i)^{1/(1+\alpha)})$ is uniformly bounded. Therefore, by the martingale convergence theorem, $\sum_{i=1}^{n-2}(Y_i/i)^{1/(1+\alpha)} - \mathbb{E}[\sum_{i=1}^{n-2}(Y_i/i)^{1/(1+\alpha)}]$ converges a.s. But,

$$\mathbb{E}\left[\sum_{i=1}^{n-2}\left(\frac{Y_i}{i}\right)^{1/(1+\alpha)}\right] = \Gamma\left(1 + \frac{1}{1+\alpha}\right) \sum_{i=1}^{n-2}\left(\frac{1}{i}\right)^{1/(1+\alpha)}.$$

Thus,

$$\frac{\sum_{i=1}^{n-2} W_i}{(1/f(0)) \log n} \to 1 \quad \mathrm{a.s. \ when} \ \alpha = 0$$

and

$$\frac{\sum_{i=1}^{n-2} W_i}{K_\alpha n^{1-1/(1+\alpha)}} \to 1 \quad \mathrm{a.s. \ when} \ 0 < \alpha < 1, \tag{19}$$

where

$$K_\alpha := \left(\frac{1+\alpha}{C}\right)^{1/(1+\alpha)} \frac{1}{\alpha} \Gamma\left(\frac{1}{1+\alpha}\right).$$

Let $a(n, 0) := (1/f(0)) \log n$ and $a(n, \alpha) := K_\alpha n^{1-1/(1+\alpha)}$ when $0 < \alpha < 1$. Then

$$\sum_{i=1}^{n-2} W_i - a(n, \alpha) = \sum_{i=1}^{n-2} W_i - \mathbb{E}\left[\sum_{i=1}^{n-2} W_i\right] + \mathbb{E}\left[\sum_{i=1}^{n-2} W_i\right] - a(n, \alpha).$$

Recall that, by Lemma 2, $\sum_{i=1}^{n-2} W_i - \mathbb{E}[\sum_{i=1}^{n-2} W_i]$ has an almost-sure limit, so, using (19), we obtain, for $0 \leq \alpha < 1$,

$$\lim_{n \to \infty} \frac{\mathbb{E}[\sum_{i=1}^{n-2} W_i]}{a(n, \alpha)} = 1$$

and, hence, by Lemma 2, Lemma 1, and (11),

$$\mathbb{E}[T_n^{\mathrm{NN}}] \sim a(n, \alpha),$$

proving (2) and (4). Finally,

$$\mathbb{E}\left[\left(\frac{T_n^{\mathrm{NN}}}{a(n, \alpha)} - 1\right)^2\right] = \frac{\mathrm{var}[\sum_{i=2}^{n-1} \min_{i<j\leq n} L_{ij}]}{(a(n, \alpha))^2} + \frac{\mathrm{var}[L_n^{\mathrm{first}} + L_n^{\mathrm{last}}]}{(a(n, \alpha))^2} + \left[\frac{\mathbb{E}[T_n^{\mathrm{NN}}]}{n^{a(n,\alpha)}} - 1\right]^2$$

converges to 0 as $n \to \infty$. Hence,

$$\frac{T_n^{\mathrm{NN}}}{a(n, \alpha)} \xrightarrow{\mathbb{P}} 1$$

and in $\mathcal{L}_2$, proving (3) and (5). This completes the proofs of both Theorem 2 and Theorem 3.

5.3. Proof of Corollary 1

Since F has an exponential distribution with mean 1, from Lemma 1, we obtain

$$T_n^{\mathrm{NN}} \overset{\mathrm{D}}{=} \sum_{i=1}^{n-2} \frac{Y_i}{i} + R_n,$$

where $R_n \overset{\mathrm{D}}{=} L_n^{\mathrm{first}} + L_n^{\mathrm{last}}$ and is independent of the $\{Y_i\}_{i=1}^{n-2}$ which are i.i.d. exponential random variables with mean 1. But then, by Réyni's representation (see Theorem 6.5 of [2]) we have

$$T_n^{\mathrm{NN}} \overset{\mathrm{D}}{=} \max_{1 \leq i \leq n-2} Y_i + R_n.$$

Thus, the proof follows from standard asymptotic theory for the maximum of n i.i.d. exponential variables with mean 1 and using Proposition 1.

5.4. Proof of Theorem 4

As mentioned in the proof of Lemma 2, since $1/(1 + \alpha) > 1$, we have

$$\sup_{n \geq 1} \mathrm{var}\left(\sum_{i=1}^{n-2} W_i\right) < \infty.$$

Therefore, $\sum_{i=1}^{n-2} W_i - \mathbb{E}[\sum_{i=1}^{n-2} W_i]$, as a martingale, converges a.s. and in $\mathcal{L}_2$. So, by (11) and Proposition 1, $T_n^{\mathrm{NN}} - \mathbb{E}[T_n^{\mathrm{NN}}]$ converges weakly.

Now, to complete the proof, it is enough to show that $\{\mathbb{E}[T_n^{\text{NN}}]\}_{n \geq 1}$ is a convergent sequence. To this end, we apply Lemma 3 to obtain

$$\mathbb{E}[T_n^{\text{NN}}] = \int_0^\infty \frac{[\bar{F}(t)]^2[1 - (\bar{F}(t))^{n-2}]}{F(t)} \, \mathrm{d}t + \mathbb{E}[L_n^{\text{first}} + L_n^{\text{last}}].$$

Now fix $\varepsilon > 0$ and choose $\delta > 0$ such that the equations leading to the double inequality (12) hold. Also, find $M > 0$ such that $F(M) \geq \frac{1}{2}$. Consider the function $G : [0, \infty) \to [0, \infty)$ defined by

$$G(t) := \frac{1}{F(t)} \mathbf{1}[0 < t < \delta] + \frac{1}{F(\delta)} \mathbf{1}[\delta \leq t \leq M] + 2\bar{F}(t) \mathbf{1}[t > M].$$

Then, for any $n > 1$ and $t > 0$, we have $[\bar{F}(t)]^2[1 - (\bar{F}(t))^{n-2}]/F(t) \leq G(t)$. Also, note that $\int_M^\infty G(t) \, \mathrm{d}t \leq 2 \int_0^\infty \bar{F}(t) \, \mathrm{d}t < \infty$ since F is positively supported and has finite first moment. Furthermore, by the choice of δ, the density f on $(0, \delta)$ is strictly positive and F is strictly increasing. So

$$\int_0^\delta G(t) \, \mathrm{d}t = \int_0^\delta \frac{\mathrm{d}t}{F(t)} = \int_0^{F(\delta)} \frac{\mathrm{d}w}{w f(F^{-1}(w))} \leq \kappa \int_0^1 \frac{1}{w^{1+\alpha/(1+\alpha)}} \, \mathrm{d}w < \infty$$

with $\kappa > 0$ some constant, where the penultimate inequality follows by using the double inequality (12) and the final inequality holds because $-1 < \alpha < 0$. Thus, we have $\int_0^\infty G(t) \, \mathrm{d}t < \infty$. Therefore, by the dominated convergence theorem, we conclude that

$$\lim_{n \to \infty} \int_0^\infty \frac{[\bar{F}(t)]^2[1 - (\bar{F}(t))^{n-2}]}{F(t)} \, \mathrm{d}t$$

exists. This along with Proposition 1 proves that $\{\mathbb{E}[T_n^{\text{NN}}]\}_{n \geq 1}$ is a convergent sequence, which completes the proof of the theorem.

5.5. Proof of Theorem 5

By the assumption that F is a Weibull distribution with shape parameter $1 + \alpha$ and scale parameter 1, we have $F(x) = 1 - \mathrm{e}^{-x^{1+\alpha}}$, $x \geq 0$. Therefore, $F^{-1}(t) = [-\log(1-t)]^{1/(1+\alpha)}$, where $0 < t < 1$. Hence,

$$\sum_{i=2}^{n-1} \min_{i < j \leq n} L_{ij} \overset{\mathrm{D}}{=} \sum_{i=1}^{n-2} W_i = \sum_{i=1}^{n-2} [-\log(\mathrm{e}^{-Y_i/i})]^{1/(1+\alpha)} = \sum_{i=1}^{n-2} \left(\frac{Y_i}{i}\right)^{1/(1+\alpha)},$$

where the Y_i are i.i.d. exponential random variables each with mean 1. Now, using Lemma 1, we obtain

$$\mathbb{E}[T_n^{\text{NN}}] = \Gamma\left(1 + \frac{1}{1+\alpha}\right) \sum_{i=1}^{n-2} \left(\frac{1}{i}\right)^{1/(1+\alpha)} + \mathbb{E}[L_n^{\text{first}} + L_n^{\text{last}}].$$

This proves (7).

We define $\sigma^2(\alpha) := \mathrm{var}[Y_i^{1/(1+\alpha)}] = \Gamma(2/(1+\alpha) + 1) - \Gamma^2(1 + 1/(1+\alpha))$. Let $Z_n(\alpha) = \sum_{i=1}^{n-2} V_i(\alpha)$, where

$$V_i(\alpha) := \frac{Y_i^{1/(1+\alpha)} - \mathbb{E}[Y_i^{1/(1+\alpha)}]}{\sigma(\alpha) i^{1/(1+\alpha)} \sum_{i=1}^{n-2}(1/i)^{2/(1+\alpha)})^{1/2}}.$$

Observe that $\mathbb{E}[V_i(\alpha)] = 0$ and $\sum_{i=1}^{n-2} \mathrm{var}[V_i(\alpha)] = 1$. Choose $\delta > \alpha - 1 \geq 0$. So, for some $M > 0$,

$$\sum_{i=1}^{n-2} \mathbb{E}[|V_i(\alpha)|^{2+\delta}] \leq \frac{M}{\sigma(\alpha)^{2+\delta}} \frac{1}{[\sum_{i=1}^{n-2}(1/i)^{2/(1+\alpha)}]^{(2+\delta)/2}} \sum_{i=1}^{n-2}\left(\frac{1}{i}\right)^{(2+\delta)/(1+\alpha)}.$$

Since $2/(1+\alpha) \leq 1$ and $(2+\delta)/(1+\alpha) > 1$, we have

$$\lim_{n\to\infty} \sum_{i=1}^{n-2} \mathbb{E}[|V_i(\alpha)|^{2+\delta}] = 0.$$

Hence, Lyapunov's condition is satisfied for $\alpha \geq 1$ and so $Z_n(\alpha)$ converges in distribution to a standard normal random variable. Now observe that, by Lemma 1 and Proposition 1,

$$\frac{T_n^{\mathrm{NN}} - \mathbb{E}[T_n^{\mathrm{NN}}]}{\sigma(\alpha)(\sum_{i=1}^{n-2}(1/i)^{2/(1+\alpha)})^{1/2}} = Z_n(\alpha) + o_{\mathbb{P}}(1).$$

Thus, (8) and (9) hold because $\sum_{i=1}^{n-2}(1/i)^{2/(1+\alpha)} \sim ((\alpha+1)/(\alpha-1))n^{1/2-1/(1+\alpha)}$ when $\alpha > 1$ and $\sum_{i=1}^{n-2}(1/i)^{2/(1+\alpha)} \sim \log n$ when $\alpha = 1$.

Finally, we note that, for any $\alpha \geq 1$,

$$\frac{T_n^{\mathrm{NN}} - \mathbb{E}[T_n^{\mathrm{NN}}]}{n^{1-1/(1+\alpha)}} \xrightarrow{\mathbb{P}} 0.$$

So, using (7), we conclude that (6) also holds.

5.6. Proof of Theorem 6

From the above proof, it follows that, for $(Y_i)_{i\geq 1}$ i.i.d. exponential random variables with mean 1 and $\alpha \geq 1$, we have

$$\frac{\sum_{i=1}^{n-2}(Y_i/i)^{1/(1+\alpha)}}{n^{1-1/(1+\alpha)}} \xrightarrow{\mathbb{P}} \frac{1}{\alpha}\Gamma\left(\frac{1}{1+\alpha}\right).$$

So, using (18), we conclude that

$$\frac{\sum_{i=1}^{n-2} W_i}{n^{1-1/(1+\alpha)}} \xrightarrow{\mathbb{P}} K_\alpha.$$

Therefore, the proof of the theorem follows from Proposition 1.

6. Discussion

In our main results we have assumed that the second moment of F exists. This assumption is not needed. The following lemma says that if F is a positively supported distribution with finite βth moment then, for any $k > 2/\beta$, we must have $\mathbb{E}[(\min_{1\leq i\leq k} Z_i)^2] < \infty$, where $Z_1, Z_2, \ldots$ are i.i.d. random variables with distribution F. The proof of this lemma follows easily from Markov's inequality, so we omit it.

Lemma 4. *Suppose that Z is a nonnegative random variable such that, for some $\beta > 0$, $\mathbb{E}[Z^\beta] < \infty$. Then, for any $k > 2/\beta$, we have*

$$\int_0^\infty t\{\mathbb{P}(Z > t)\}^k \, dt < \infty.$$

Now, as in Lemma 1, let random variable $W_i = F^{-1}(1 - \exp(-Y_i/i))$, where the Y_i are exponential with mean 1. We have assumed that F has finite first moment, and so, by taking $k = 3$ in Lemma 4, we can conclude that W_i has finite second moment for $i \geq 3$. Thus, under the assumptions of Lemma 2 and by following its proof, we can conclude that $\sum_{i=3}^{n-2}(W_i - \mathbb{E}[W_i])$ converges a.s. and in $\mathcal{L}_2$. Thus, all the results stated in Section 2 hold except those on the $\mathcal{L}_2$ convergence.

Acknowledgements

References

[1] BELLMORE, M. AND NEMHAUSER, G. L. (1968). The traveling salesman problem: a survey. *Operat. Res.* **16,** 538–558.

[2] DASGUPTA, A. (2011). *Probability for Statistics and Machine Learning*. Springer, New York.

[3] GAVETT, J. W. (1965). Three heuristic rules for sequencing jobs to a single production facility. *Manag. Sci.* **11,** B-166–B-176.

[4] PAPADIMITRIOU, C. H. AND STEIGLITZ, K. (1998). *Combinatorial Optimization: Algorithms and Complexity*. Dover, Mineola, NY.

[5] ROSENKRANTZ, D. J., STEARNS, R. E. AND LEWIS, P. M., II (1977). An analysis of several heuristics for the traveling salesman problem. *SIAM J. Comput.* **6,** 563–581.

[6] VANNIMENUS, J. AND MÉZARD, M. (1984). On the statistical mechanics of optimization problems of the travelling salesman type. *J. Physique Lett.* **45,** 1145–1153.

[7] WÄSTLUND, J. (2012). Replica symmetry of the minimum matching. *Ann. Math.* **175,** 1061–1091.

[8] WÄSTLUND, J. (2010). The mean field traveling salesman and related problems. *Acta Math.* **204,** 91–150.

[9] WENDEL, J. G. (1948). Note on the gamma function. *Amer. Math. Monthly* **55,** 563–564.