

Correcting misinterpretations on the distribution of feasible solution lengths in the traveling salesman problem

Jerry L. Shaw^a, Donovan Fuqua^b, Hansuk Sohn^a, Manuel Ivan Rodriguez-Borbon^a & Victor Pimentel^c

^a Department of Industrial Engineering, New Mexico State University, Las Cruces, New Mexico, United States. jshaw@nmsu.edu, hsohn@nmsu.edu, ivanrodr@nmsu.edu

^b Department of Accounting and Information Systems, New Mexico State University, Las Cruces, New Mexico, United States. dfuqua@nmsu.edu

^c Department of Management, New Mexico State University, Las Cruces, New Mexico, United States. vip@nmsu.edu

Received: July 31th, 2023. Received in revised form: December 13th, 2023. Accepted: January 16th, 2024

Abstract

The traveling salesman problem (TSP) is the canonical combinatorial optimization problem famous throughout literature. There exists an objective function associated with every feasible solution. However, the increase in the number of possible solutions makes this an NP-Hard problem. We show that the central limit theorem (CLT) applies to the problem. We then conduct extensive computational testing to show that the cycle lengths tend to a normal distribution as the problem grows large. When the size of the TSP problem exceeds computational power, better understanding solution distributions allows us to save resources. This is a non-trivial result as understanding solution distributions in huge TSP problems helps us to minimize computational effort that may not lead to significantly better results.

Keywords: traveling salesman problem; distribution fitting; graph theory

Corrección de interpretaciones erróneas sobre la distribución de longitudes de solución factibles en el problema del viajante

Resumen

El problema del agente viajero ("Traveling Salesman Problem" o TSP) es el problema canónico de optimización combinatoria usando en la literatura. Existe una función objetivo asociada con cada solución factible. Sin embargo, la tasa de aumento en el número de soluciones posibles hace que este sea un problema NP Difícil. Mostramos que el teorema del límite central (Central Limit Theorem o CLT) se aplica al problema y luego realizamos pruebas computacionales extensas para mostrar que las longitudes de los ciclos tienden a una distribución normal a medida que el tamaño del problema crece. Cuando el tamaño del problema TSP excede el poder computacional, una mejor comprensión de las distribuciones de soluciones nos permite ahorrar recursos. Este es un resultado no trivial, ya que comprender las distribuciones de soluciones en problemas TSP enormes nos ayuda a minimizar el esfuerzo computacional que puede no conducir a resultados significativamente mejores.

Palabras clave: problema del viajante de comercio; accesorios de distribución; teoría de grafos.

1 Introduction

Our contribution is to dispel two widely held but incorrect perceptions regarding the Traveling Salesman Problem (TSP) in the Operations Research field. The first is that the set of feasible TSP solution objective function values is proven to not be Normally distributed. This assumption derives from the failure to prove that the cycle lengths tend

to normality in the general case using analytic means. A second misnomer is that feasible objective function values have a Weibull distribution. That inference derives from a misreading of Golden and a misinterpretation of extremal value theory [1].

The Weibull does work well to model the distribution of extremal values, but not necessarily entire populations. Derigs recounts Golden's work and then extends the

How to cite: Shaw, J.L., Fuqua, D., Sohn, H., Rodriguez-Borbon, M.I., and Pimentel, V. Correcting misinterpretations on the distribution of feasible solution lengths in the traveling salesman problem. DYNA, 91(231), pp. 57-62, January - March, 2024.

application of the Weibull to estimate minima in problems with few locations [2]. A careful application of analytic methods enables us to construct the argument that the lengths of randomly generated feasible TSP solutions do indeed tend to normal distributions, as would be expected in light of the Central Limit Theorem.

This research corrects at least two widely held but incorrect perceptions in the operations research field. One is that the set of feasible traveling salesman problem (TSP) solutions objective function values is proven to not be normally distributed. This assumption derives from the failure to prove that the cycle lengths tend to normality in the general (including one-dimensional) case using analytic means. A second misnomer is that feasible objective function values have a Weibull distribution. That inference derives from a misreading of Golden as well as a misinterpretation of extremal value theory [1]. The Weibull does work well to model the distribution of extreme values, but not necessarily entire populations. Derigs recounts Golden's work and then extends the application of the Weibull to estimate minima in problems with few locations [2]. A careful application of analytic methods enables us to construct the argument that the lengths of randomly generated feasible TSP solutions do indeed tend to normal distributions as would be expected in light of the Central Limit Theorem.

The TSP is the best-known combinatorial optimization problem. While there are several versions of this problem, we consider the simplest. Here, the objective of the problem is to create a minimum length round trip through a set of locations. In its most common form, (1) the locations exist in a bounded space in R^2 ; and (2) the distances between locations are calculated using a Euclidean metric.

The problem has historical precedent predating its popularization among academics [3–6]. Extensive histories of the problem are found in literature [7, 8]. The problem continues to influence researchers in fields ranging from computational geometry [9,10] to statistical mechanics [11,12]

Beardwood et al. is a seminal work that strongly influenced later researchers' understanding of the problem [13,14]. Their main contribution was to apply analytic rigor to the TSP. They showed that there exists a limit to the rate of increase in the length (in the sense of Lebesgue measure) of cycles as the number of vertices in a problem increases. This relationship is usually reported as $\lim_{N \rightarrow \infty} \frac{L_E}{N^{1-1/d}} = \beta_E(d)$; the rate of increase in length of a cycle through "many" points approaches a constant value.

Beardwood asserts that the lengths of random cycles through a point set in R^1 are not asymptotically normal in all cases [13]. However, the behavior in R^1 is not the same as in R^k where $k > 1$. On account of this, they declared that they failed to prove that cycle lengths are not provably normal in the general case. They stated that conditions for convergence to the Normal distribution exist in the special case where $C = E$ while E is the set of points and C is a closed region of unit measure. The problems we address exist in R^2 .

Golden noted that an estimate on lower bounds on TSP solution values could be obtained using a Weibull distribution [1]. Derigs extended that work to place bounds

on a broader set of combinatorial optimization problems [15]. Other researchers reported the influence of Golden's work [16–18]. The Weibull distribution does have properties that would make it attractive for use in modeling feasible TSP solution values. First, it can be used to model distributions with specific lower values. Second, its parametrization supports multiple distribution shapes, such as the "bell shape" that we find during our computational experimentation.

Beardwood et al. cited that the probabilistic nature of locations in the data set could cause cycle lengths to not converge [13]. For instance, a point to be visited could exist at some position with some probability. We assert, however, that this concern does not apply to the set of locations in the test TSP data sets. The locations for the problem sets are given. Once the set of locations to be visited is fixed, there is no longer a question of the probable location of any given location. While most applications focus on exact solutions to symmetric TSP instances, recent work exists on "good enough" solutions [19]. Similarly, research has been published on finding maximum-length TSP cycles or as Barvinok [2007] writes, "...known informally as the "taxicab ripoff problem" [20].

We show, both through analysis and computational results, that cycle lengths for feasible TSP solutions tend to follow a Normal distribution as the problem size grows large. Other authors have alluded to this behavior but have not published computational evidence to support or disprove the tendency [7,21]. Consequently, research on potential opportunities to exploit the statistical properties of cycle lengths has gone undone.

2 Research Objectives and Analytic Basis

Our purpose is to show that (1) TSP objective function values can be expressed as random variables; (2) the Central Limit Theorem can be applied to TSP cycle lengths; and (3) computational experimentation supports the hypothesis that cycle lengths tend to a normal distribution as the problem size increases. Research on the TSP is central to the fields of operations research, theoretical computer science, combinatorics, and graph theory.

While there is a great deal of commonality in the terminology used in these fields, there are also some differences. We have tried to reduce any confusion that may arise from the differing backgrounds of readers. We use both the terms "vertex" and "location." In general, "vertex" is used when the context of the passage derives from graph theoretic sources. The terms "distance" and "length" refer to a Lebesgue metric rather than the number of edges transited as is common in graph theory.

Definition 2.1 A walk is a nonempty sequence of edges (uv, vw, \dots) such that the tail of an edge is the head of the next edge if the sequence contains more than one edge.

Definition 2.2 A closed walk is a walk such that the number of edges equals the number of vertices and the degree of each vertex is 2.

Definition 2.3 A cycle is a closed walk $(v_1, v_2, \dots, v_n, \forall n \geq 3)$, in which then points v_i are distinct [22].

Definition 2.4 A permutation is "The arrangement of

different objects into a linear order using each object exactly once” [23].

Lemma 2.1 *A cycle can be constructed from a permutation of the vertices of K_n .*

A random variable is a variable quantity whose values depend on chance and for which a distribution function of probabilities has been defined [24]. Similarly, a function defined on a sample space is called a random variable [2]. Thus, the distance between randomly generated locations is itself a random variable. The Central Limit Theorem states that sums of Independent Identically Distributed (IID) random variables with finite variance approach a Normal distribution as the number of random variables (RVs) grows large [25]. As Hogg and Tanis state: “Given a random experiment with an outcome space S , a function X that assigns one and only one real number $X(s) = x$ to each element s in S is called a random variable [26]. The space of X is the set of real numbers, where $S\{x : X(s) = x, s \in S\}$, where $s \in S$ means that the element s belongs to the set S .”

The position of each vertex in the tour is dictated by a random number associated with that particular vertex as will be shown in Section 3. As the traveler transits the cycle, the distance between two successive vertices (say the first and second vertices visited) is a function on two random variables (the vertices). In this case, the distance between the each successive pair of vertices is a random variable. The total distance traveled is the sum along the respective edges the traveler transits in a cycle.

Lemma 2.2 *The distance between two locations having finite coordinates is finite. If we assume the lemma is false, then one of the end points must be infinitely far from the other. But each location has finite coordinates. Since the difference between finite numbers is finite, the distance between the points cannot be infinite. This is a contradiction. Thus the lemma is proven.*

Lemma 2.3 *The variance of the distances between locations having finite coordinates is finite. The sum of finitely many finite quantities is finite. Here we define variance as $\sigma^2 = \frac{\sum_i (\mu - x_i)^2}{n}$ where, $\mu = \frac{\sum_{i=1}^n x_i}{n}$. Thus, μ must be finite. The definition of variance is a squared sum of finitely many finite quantities. Therefore, the variance of the distances between locations, σ^2 , is finite*

3 Methodology

While the usual objective of the problem is to find the minimum length cycle through the locations, a feasible solution can be derived from any ordered list of locations [27]. The cycle length is the sum of the distances between each location in the round trip that constitutes the cycle. Assuming that the locations are chosen randomly, the cycle length is the sum of random variables.

We now show how to generate a cycle whose edges are constructed from IID random variables. The methods used coincidental similarity to that of Tercariol et al. but were developed independently [27]. The weight for any arc in this case is the Euclidean distance between the arc endpoints.

To generate a random cycle: let there be a cycle along the elements of the newly created ordering of P_r .

The Euclidean distance between successive vertices in the

new list is a random variable. Each element in v exists at a specified location and is addressed using Cartesian coordinates. Therefore, the distance between any two elements is finite. Since every element in the set of distances is finite, the variance of the distances must also be finite in the strictly numerical sense. It is well to remember that the TSPs we consider exist in the real world and that we treat the subject in an engineering context rather than in a solely abstract mathematical space. In the engineering sense it makes no sense to consider vertices located infinitely distant.

Lemma 3.1 *A cycle constructed as shown above is a sequence of IIDs with finite mean and variance. There is a bijection between each location and a uniformly distributed random number. That random number is a random variable. In each instance the generated random comes from the same distribution without regard to any possible previous trial. Therefore, these random variables are IID. The locations are then sorted in order by the generated random. The resulting ordering is a random variable. This ordered list constitutes a permutation and thus defines a cycle.*

Algorithm 1 Random Cycle Generation

- 1: Where, P is a set of indices mapped to locations (A)
 - 2: $R : P \rightarrow A \ni A \cup (0, 1)$. R is a bijection on P
 - 3: Input: Generate R^{-1} as a mapping of the elements of A
 - 4: $P \ni |A| = |P|$
 - 5: Sort: A to produce an ordered list, A_r
 - 6: Condition: let $P_r \subset P : A$
 - 7: Output: Generate A as a randomized ordering of locations where A maps to P
 - 8: Result: The position in which every element of P appears in P_r is the result of a random process.
-

The distances between any pair of locations is finite by Lemma 2.2. The variance of the distances is finite by Lemma 2.3. Thus is constructed a cycle composed of IID random variables with finite mean and variance.

Theorem 3.1 The length of cycles created as above tend to a Normal distribution as the number of locations grows large without bound.

Proof 3.1 The Central Limit Theorem states that sums of Independent Identically Distributed (IID) random variables with finite variance approach a Normal distribution as the number of random variables (RVs) grows large [25]. Lemma 4 showed that cycle lengths are composed of the sum of IID random variables with finite mean and variance. Therefore, the cycles so constructed satisfy the requirement of the Central Limit Theorem.

To illustrate a short example of cycle generation, we assign random deviates (r,s) generated for five locations labeled a through e . The resulting sequence of locations is then a random variable. Since the distances between successive locations are dictated by random variables, those distances are also random variables. The total cycle length is the sum of random variables. See Table 1 for an example of five data points using this method.

Table 1.

Example of five randomly generated locations and mappings

Location	r_v	v	f_v	v
1	0.997	a	0.284	d
2	0.825	b	0.312	c
3	0.312	c	0.825	b
4	0.284	d	0.941	e
5	0.941	e	0.997	a

Source: the authors.

The resulting sequence of locations is then a random variable. Since the distances between successive locations are dictated by random variables, those distances are also random variables. The total cycle length is the sum of random variables.

The remainder of the paper describes our computational methodology, results, conclusions, and recommendations for further work. Computations were executed on an AMD Ryzen 7 4800H 8-core CPU with 32 GB RAM using code written by the authors in Python 3.5.

4 Results and Discussion

Standard data sets for the TSP can be found in the TSP Library [28]. Those problems are labeled T in the Problem Source column in Table 2. Data sets motivated by human population centers (denoted as N) and Very Large-Scale Integrated circuits (denoted as V) are available from the Waterloo University, Department of Computer Science site. Results for problems containing as few as 25 locations are displayed in Table 2. For each of the test problems the null hypothesis is that the distribution of cycle lengths is Normal. To test that hypothesis, we generated closed paths containing each location in the problem. The resulting cycle lengths were then subjected to three tests to check for normality.

Those tests include: D’Agostino, Jarque-Bera, and Anderson-Darling. Our use of these tests are consistent with current literature regarding tests for normality [29–32]. Similarly, we did not use Kolmogorov- Smirnov (KS) tests or similar tests due to sensitivity to large data sets [32]. In all instances the confidence level was 99%. Anderson-Darling test statistics that do not exceed the 99%certainty critical value (1.092) result in rejection of the null hypothesis. We highlight the results of our tests in Table 2.

In all instances the confidence level was 99%. Anderson-Darling test statistics that do not exceed the 99%certainty critical value (1.092) result in rejection of the null hypothesis. We highlight the results of our tests in Tables 2 and 3 in the Appendices.

Table 2.

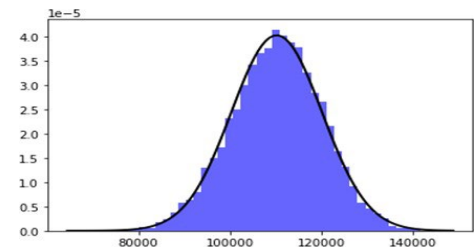
Summary of datasets tested

Vertices	Datasets Tested	Jarque Failures	Anderson Failures	D’Agostino Failures
>5000	28	0	0	0
4999-1000	45	0	0	0
994-226	45	0	0	0
<226	96	56	42	55

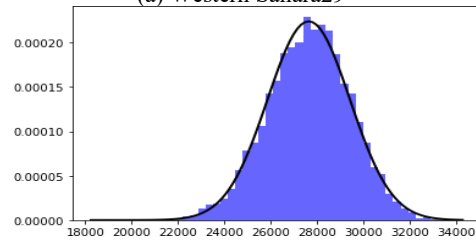
Source: the authors.

Results from a subset of the tested data sets are displayed in the table below. Statistics for each of the three the tests are shown for each data set. The test statistic or p-value is displayed in bold for instances in which the null hypothesis is rejected. For example, the null hypothesis for the CH150 data set is rejected on account of Jarque-Bera and D’Agostino tests. In Table 2 we see no failures to reject the null hypothesis for problems exceeding 1000 vertices. Consistent with our theory, we see multiple rejection of the null hypothesis where the number of vertices drop below 225.

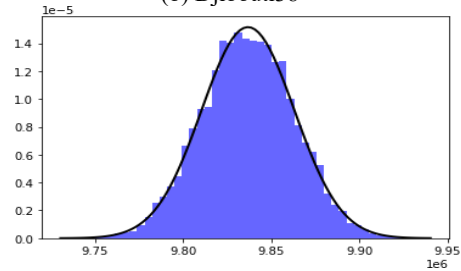
Fig. 1 display histograms of the cycle lengths produced through experimentation. While these histograms “look” Normal, statistical testing shows this is not always the case. Results from testing shows that the null hypothesis is rejected for the test problems. This is more easily shown in the Q-Q plots depicted in Fig. 2.



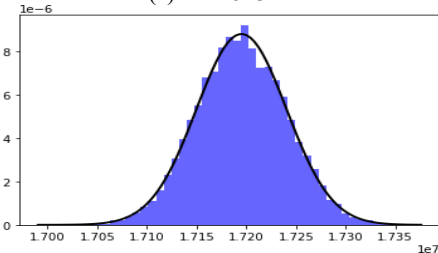
(a) Western Sahara29



(b) Djibouti38



(c) Bbz25234



(d) Xib32892

Figure 1. Histogram plots for small (1a, 1b) and large (1c, 1d) TSP problems
Source: the authors.

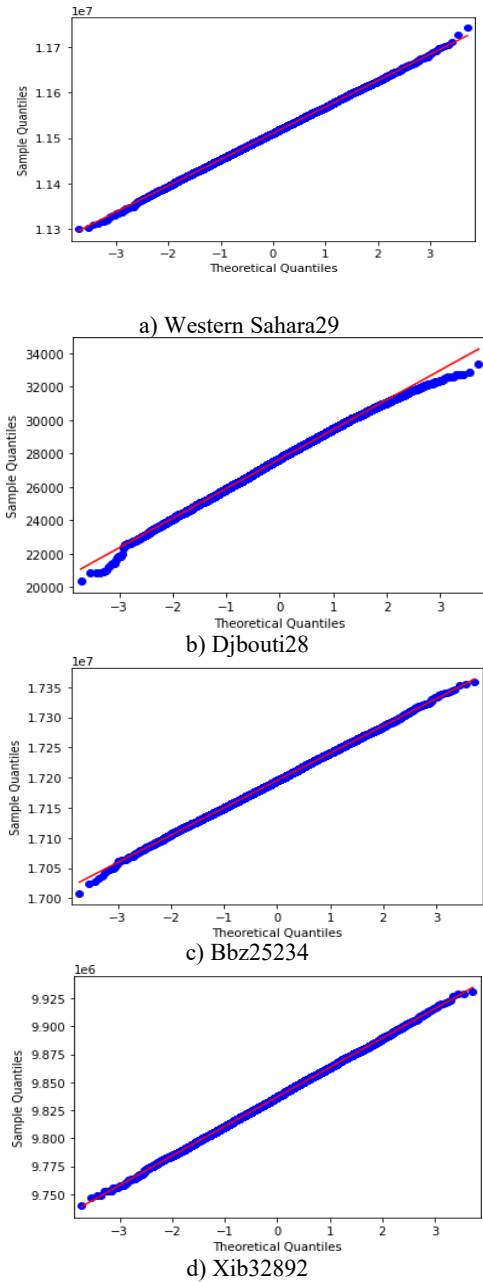


Figure 2. Q-Q plots for small(2a, 2b) and large (2c, 2d) TSP problems. Source: the authors.

Problem sizes varied from 25 to 85,900 locations. In general, the null hypothesis was rejected more frequently in problems having fewer vertices. The null hypothesis was not rejected for any problem having over 225 vertices.

In this research paper, we did the following:

1. We showed a generative mechanism to generate IID random vertex orderings that create cycles that produce feasible TSP solutions. Since the resulting inter-vertex distances are the products of a random process, they are random variables. The overall lengths of the resulting cycles are, thus, the sums of random variables.
2. The CLT tells us that the sums of RVs tend to a Normal distribution as the number of summands grows large.

Consequently, we can say that as the problem size (in terms of vertices) grows large, the lengths of the so-generated feasible solutions tend to a Normal distribution.

3. Hypothesis testing on the experimentally generated routes failed to disprove the null hypothesis – that route lengths of feasible solutions do not tend to a Normal distribution.

The experimental results support our claim that the objective function values for feasible TSP solutions tends to a Normal distribution as the number of vertices increases. We are not aware of testing of this type in existing reported literature.

While only a relatively small portion of the total number of cycle lengths are represented by in the tails of the distribution, this still accounts for a very large number in absolute terms. The experimental results support our claim that the objective function values for feasible TSP solutions tends to a Normal distribution as the number of vertices increases. Testing of this type has not been reported in the literature.

The overall probabilistic nature of feasible TSP cycle lengths has considered and then largely ignored. Carling noted that Golden’s work generated great interest but research on related topics went largely dormant [16]. Doubtless the restrictions on computational experimentation at the time hampered widespread study of the topic. The technology now exists to enable much more experimentation.

While a cursory examination of the CLT suggested our results, some hurdles had to be overcome to produce these results. The highly influential paper from Beardwood et al. noted a failure to prove asymptotic convergence of cycle lengths to a Normal distribution. However, that research relied primarily on hand calculations and analytic methods. Beardwood’s assertion that there exists an upper limit to the rate of increase in solutions has been positively received over the last six decades.

This research used some well-known tools and modern, if modest, computational resources to illustrate an interesting linkage between some combinatorial objects and probability theory. Knowing that the set of feasible solutions for a given problem yields paths whose lengths tend to a normal distribution should provide insight into the underlying structure of some combinatorial optimization problems. Since the TSP is NP Complete, we may be able to apply related analytic methods to other NP Complete problems. A natural direction of future work is to consider whether there might be statistical properties that optimal or good solutions share. Additional research might explore the use of this work to compute some relatively hard bounds on optimal solution objective values. We also recognize that while a two-dimensional TSP generates univariate random lengths, a higher dimensional TSP would require a multivariate test such as Mardia’s Test. [33]. Because of the large size of the tests, we discounted the Kolmogorov-Smirnov test and focused our analysis with the Anderson-Darling, D’Agostino, and Jarque-Bera tests. In future studies it would be of interest to determine the robustness of each test with respect to random cycle lengths.

This work has significant potential to affect commercial and public logistics activities. One interesting question is whether problems with similar vertex counts covering similar spatial areas would have similar cycle lengths. Likewise, can we relate mean cycle lengths in problems having similar spatial properties and vertex counts. A positive answer to this question could be coupled with Beardwood’s growth rate limit for length to produce cycle length estimates for different

data sets. This would be useful in capacity planning for commercial delivery companies as well as utility companies.

References

- [1] Golden, B.L., A statistical approach to the TSP. *Networks*. 7(3), pp. 209-225, 1977. DOI: <https://doi.org/10.1002/net.3230070303>.
- [2] Derigs, U., Using confidence limits for the global optimum in combinatorial optimization. *Operations Research*. 33(5), pp. 1024-1049, 1985. DOI: <https://doi.org/10.1287/opre.33.5.1024>.
- [3] Flood, M.M., The traveling-salesman problem. *Operations Research*. 4(1), pp. 61-75, 1956. DOI: <https://doi.org/10.1287/opre.4.1.61>.
- [4] Jessen, R.J., The master sample project and its use in agricultural economics. *Journal of Farm Economics*. 29(2), pp. 531-540, 1947. DOI: <https://doi.org/10.2307/1232392>.
- [5] Mahalanobis, P.C., A sample survey of the acreage under jute in Bengal. *Sankhyā: The Indian Journal of Statistics*. [online]. 4(4), pp. 511-530, 1940. Available at: <http://www.jstor.org/stable/40383954>.
- [6] Dantzig, G., Fulkerson, R., and Johnson, S., Solution of a large-scale traveling-salesman problem. *Journal of the Operations Research Society of America*. 2(4), pp. 393-410, 1954. DOI: <https://doi.org/10.1287/opre.2.4.393>
- [7] Cook, W.J., Applegate, D.L., Bixby, R.E., and Chvatal, V., *The traveling salesman problem: a computational study*. Princeton University Press, Princeton, NJ, USA, 2011.
- [8] Bellmore, M., and Nemhauser, G.L., The traveling salesman problem: a survey. *Operations Research*. 16(3), pp. 538-558, 1968. DOI: <https://doi.org/10.1287/opre.16.3.538>
- [9] Hougard, S., and Zhong, X., Hard to solve instances of the Euclidean Traveling Salesman Problem. *Mathematical Programming Computation*. 13, pp. 51-74, 2021. DOI: <https://doi.org/10.1007/s12532-020-00184-5>
- [10] Zhong, X., Slightly improved upper bound on the integrality ratio for the $s-t$ Path TSP. *Operations Research Letters*. 48(5), pp. 627-629, 2020. DOI: <https://doi.org/10.1016/j.orl.2020.07.015>.
- [11] Caracciolo, S., Di Gioacchino, A., Gherardi, M., and Malatesta, E.M., Solution for a bipartite Euclidean traveling-salesman problem in one dimension. *Physical Review E*. 97(5), art. 052109, 2018. DOI: <https://doi.org/10.1103/PhysRevE.97.052109>
- [12] Vannimenus, J., and Mézard, M., On the statistical mechanics of optimization problems of the travelling salesman type. *Journal de Physique Lettres*. 45(24), pp. 1145-1153, 1984. DOI: <https://doi.org/10.1051/jphyslet:0198400450240114500>.
- [13] Beardwood, J., Halton, J.H., and Hammersley, J.M., The shortest path through many points. In *Mathematical Proceedings of the Cambridge Philosophical Society*, Cambridge University Press. 55(4), pp. 299-327, 1959. DOI: <https://doi.org/10.1017/S0305004100034095>.
- [14] Steele, J.M., Probability and problems in Euclidean combinatorial optimization. *Statistical Science*. 1, pp. 48-56, 1993. DOI: <https://doi.org/10.1214/ss/1177011083>.
- [15] Derigs, U., Using confidence limits for the global optimum in combinatorial optimization. *Operations Research*. 33(5), pp. 1024-1049, 1985. DOI: <https://doi.org/10.1287/opre.33.5.1024>.
- [16] Carling, K., and Meng, X., On statistical bounds of heuristic solutions to location problems. *Journal of Combinatorial Optimization*. 31, pp. 1518-1549, 2016. DOI: <https://doi.org/10.1007/s10878-015-9839-0>.
- [17] Giddings, A.P., Rardin, R.L., and Uzsoy, R., Statistical optimum estimation techniques for combinatorial optimization problems: a review and critique. *Journal of Heuristics*. 20, pp. 329-358, 2014. DOI: <https://doi.org/10.1007/s10732-014-9243-4>.
- [18] Marin, A., and Salmeron, J., Tactical design of rail freight networks. Part II: local search methods with statistical analysis. *European Journal of Operational Research*. 94(1), pp. 43-53, 1996. DOI: [https://doi.org/10.1016/0377-2217\(95\)00193-X](https://doi.org/10.1016/0377-2217(95)00193-X).
- [19] Di Placido, A., Archetti, C., Cerrone, C., and Golden, B., The generalized close enough traveling salesman problem. *European Journal of Operational Research*. 310(3), pp. 974-991, 2023. DOI: <https://doi.org/10.1016/j.ejor.2023.03.010>
- [20] Barvinok, A., Gimadi, E.K., and Serdyukov, A.I., The maximum TSP. In: Gutin, G., and Punnen, A.P. (eds), *The traveling salesman problem and its variations*. Combinatorial Optimization, vol 12. Springer, Boston, MA, -USA, 2007. DOI: https://doi.org/10.1007/0-306-48213-4_12
- [21] Cerf, N.J., De Monvel, J.B., Bohigas O., Martin, O.C., Percus, A.G., The random link approximation for the Euclidean traveling salesman problem. *Journal de Physique I*. 7(1), pp. 117-136, 1997. DOI: <https://doi.org/10.1051/jpl1:1997129>.
- [22] Harary, F., Beineke, L.W., *A seminar on graph theory*, Dover Edition, Dover Publications, Inc, Mineola, New York, USA, 2015.
- [23] Bóna, M., *A walk through combinatorics: an introduction to enumeration and graph theory*. World Scientific, Singapur, 2006. DOI: <https://doi.org/10.1142/10258>.
- [24] Gnedenko, B.V., *Theory of Probability*, 6th Edition, Routledge, Amsterdam, The Netherlands, 2018. DOI: <https://doi.org/10.1201/9780203718964>.
- [25] Bickel, P.I., and Doksum, K.A., *Mathematical statistics: basic ideas and selected topics*, Volumes I-II Package, 0th Edition, Chapman and Hall/CRC, 2015. DOI: <https://doi.org/10.1201/9781315369266>.
- [26] Hogg, R.V., Tanis, E.A., and Zimmerman, D.L., *Probability and statistical inference*, Vol. 993, Macmillan, New York, USA, 1977.
- [27] Terçariol, C.A., and Martinez, A.S., An efficient algorithm to generate random uncorrelated Euclidean distances: the random link model. *Brazilian Journal of Physics*. 36, pp. 232-236, 2006. DOI: <https://doi.org/10.1590/S0103-97332006000200017>.
- [28] Reinelt, G., *TSPLIB95. Interdisziplinäres Zentrum für Wissenschaftliches Rechnen (IWR), Heidelberg*. 338, pp. 1-6, 1995.
- [29] D'Agostino, R.B., Belanger, A., D'Agostino Jr., R.B., A suggestion for using powerful and informative tests of normality. *The American Statistician*. 44(4), pp. 316-321, 1990. DOI: <https://doi.org/10.2307/2684359>.
- [30] Thadewald, T., and Büning, H., Jarque-Bera test and its competitors for testing normality—a power comparison. *Journal of applied statistics*. 34(1), pp. 87-105, 2007. DOI: <https://doi.org/10.1080/02664760600994539>.
- [31] Das, K.R., and Imon, A.H., A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics*. 5(1), pp. 5-12, 2016. DOI: <https://doi.org/10.11648/j.ajtas.20160501.12>.
- [32] Yazici, B., and Yolacan, S., A comparison of various tests of normality. *Journal of Statistical Computation and Simulation*. 77(2), pp. 175-183, 2017. DOI: <https://doi.org/10.1080/10629360600678310>.
- [33] Enomoto, R., Hanusz, Z., Hara, A., and Seo, T., Multivariate normality test using normalizing transformation for Mardia's multivariate kurtosis. *Communications in Statistics-Simulation and Computation*. 49(3), pp. 684-698, 2020. DOI: <https://doi.org/10.1080/03610918.2019.1661476>

J.L. Shaw, is a recent PhD graduate in Industrial Engineering from the New Mexico State University, United States. He is retired analytics professional at UPS and has over 30 years of work experience in the Operations Research field.
ORCID: 0000-0003-0063-3927

D. Fuqua, received his PhD in Industrial Engineering in 2019 and serves as an assistant professor of information systems at New Mexico State University, United States. His research interests are applying predictive analytics to supply chain and manufacturing problems in industry.
ORCID: 0000-0002-6163-1741

H. Sohn, received his PhD in Industrial Engineering in 2005 and currently serves as the department head of the Department of Industrial Engineering at New Mexico State University, United States. His research interests are in operations research.
ORCID: 0000-0002-8126-730X

M.I. Rodriguez-Borbon, received his PhD in Industrial Engineering in 2012 and serves as a visiting professor of Industrial Engineering at New Mexico State University, United States. His research interests are in quality engineering, operations research, and Industry 4.0.
ORCID: 0000-0001-8405-4599

V. Pimentel, received his PhD in Operations in 2016 from Washington State University, United States. He serves as an assistant professor of management at New Mexico State University, United States. His research interests are in supply chain optimization.
ORCID: 0000-0002-7637-1268