Obtaining Quality-Proved Near Optimal Results for Traveling Salesman Problem

Wenhong Tian\textsuperscript{1,2}, Xinyang Wang\textsuperscript{1}, Qin Xiong\textsuperscript{1}, Yu Chen\textsuperscript{1}
\textsuperscript{1} School of Computer Science and Engineering
University of Electronic Science and Technology of China, China
\textsuperscript{2} Visiting Scholar, Department of Information and Computing Systems
University of Melbourne, Australia
Email: tian.wenhong@unimelb.edu.au Email: tian_wenhong@uestc.edu.cn

Abstract

The traveling salesman problem (TSP) is one of the most challenging NP-hard problems. It has widely applications in various disciplines such as physics, biology, computer science and so forth. The best known approximation algorithm for Symmetric TSP (STSP) whose cost matrix satisfies the triangle inequality (called $\triangle$STSP) is Christofides algorithm which was proposed in 1976 and is a $\frac{3}{2}$-approximation. Since then no proved improvement is made and improving upon this bound is a fundamental open question in combinatorial optimization. In this paper, for the first time, we propose Truncated Generalized Beta distribution (TGB) for the probability distribution of optimal tour lengths in a TSP. We then introduce an iterative TGB approach to obtain quality-proved near optimal approximation, i.e., $(1+\frac{1}{2}(\alpha+1)^{K-1})$-approximation where $K$ is the number of iterations in TGB and $\alpha(>>1)$ is the shape parameters of TGB. The result can approach the true optimum as $K$ increases.

category:F.2.2, Nonnumerical Algorithms and Problems, Computations on discrete structures
terms: Design, Algorithms, Performance

Index Terms
Symmetric traveling salesman problem (STSP); Triangle inequality; Random TSP in a unit square; TSPLIB instances; Approximation ratio; k-opt; Computational complexity

I. INTRODUCTION

The TSP is one of most researched problems in combinatorial optimization because of both its importance in both academic need and real world applications. For surveys of the TSP and its applications, the reader is referred to [Cook,2012][An et al., 2012][Vygen, 2012] and references therein.

After 38 years, Christofides’ $\frac{3}{2}$-approximation algorithm [Christofide, 1976] still keeps the best performance guarantee known for the symmetric traveling salesmen problem satisfying triangle inequality ($\triangle$STSP), and improving upon this bound is a fundamental open question in combinatorial optimization, see [Cook, 2012][Gutin et al., 2002] and references therein. [Vygen, 2012] also provides a detailed survey on new approximation algorithms for the TSP.

[Gharan et al., 2011] give a $(\frac{3}{2}-\epsilon)$-approximation for a tiny $\epsilon > 0$ using a probabilistic analysis, as [Klarreich, 2013] reports that the new progress is a 49.99...96% (totally 46 nines replaced by...
"..." ) over the optimum, a tiny margin for “graphical” traveling salesman problems. Notice that a very small percentage improvement may also be of great impact to the total length of large TSP instances. [Cook, 2012] introduces the TSP from history to the state-of-the-art.

[Johnson et al., 1998] provide a complete comparative study on the local optimization methods for TSP. David Johnson [2014] discusses the importance and applications of random TSP.

[Reiter and Rice 1966] study the cost distribution of local optima under a gradient maximizing search in 39 integer programming problems. Their results suggest that the local optima follow a Beta distribution. [Golden 1978] examines six problems from the TSPLIB archive[15]. The resulting estimates of optimal solutions are compared to the best solution found by the Lin-Kernighan algorithm [13]. The authors found that the Beta distribution is “a more appropriate distribution” than the Weibull distribution.

Recently, [Vig and Palekar 2008], apply sampling techniques similar to Golden, and use the Lin-Kernighan algorithm to find optimal tour costs. The authors estimate raw moments from the one to four of the probability distribution of optimal tour lengths. They use these estimates to fit various candidate distributions including the Beta, Weibull and Normal cases. Vig and Palekar conclude that the Beta distribution yields the best fit.

More recently, [Stuffle 2009] provide exact solutions to compute the mean, variance, the third and fourth central moment of all tour lengths. The computational complexity of computing variance, the third and fourth central moment is respectively \(O(n^2), O(n^4)\) and \((n^6)\) where \(n\) is the number of nodes in a TSP.

A typical probability distribution of all tour lengths for a random TSP in a square unit is shown in Fig. 1 where the total node number is 12. An example of TSPLIB Burma14 is shown in Fig.2. Similar results are observed for different total number of cities for which all tour lengths can be obtained.

Since there are \((n-1)!/2\) number of total tours for a STSP with \(n\) nodes, it is impractical to permute all tours for medium or large size problems. However, if the probability distribution of all tour lengths can be modelled well by a closed-form solution, then it can be applied to analyze performance of some algorithms and obtain statistical characteristics easily. Based on extensive previous and our own results, the probability density function of all tour lengths can be better approximated by Generalized Beta Distribution with four parameters, which we will introduce in the section III.

The organization of remaining parts of this paper is: our major contributions are summarized in Section II, our methods are introduced in Section III, and Conclusions and future work are discussed in Section IV.

II. Results

The main contributions of this work can be summarized as follows:

- We propose Generalized Beta (GB) distribution as the probability density function of all tour lengths distribution in a symmetric TSP in Euclidean space (STSP), the four parameters of the GB can be computed from given ESTSP data directly.
- For the first time, we introduce an iterative Truncated GB (TGB) closed-form solution to obtain \((1+\frac{1}{2}\frac{n+1}{\alpha+2})^{K-1}\)-approximation for a STSP where \(K\) is the number of total iterations in TGB, and \(\alpha (>>1)\) is the shape parameter of TGB and can be determined once the TSP instance is given. The result can approach the true optimum as \(K\) increases.
III. METHODS

Firstly, a problem formulation and some preliminaries are provided in this section.

A. Problem Formulation

Consider the \(n\)-node TSP defined in Euclidean space. This can be represented on a complete graph \(G = (V, E)\) where \(V\) is the set of vertices and \(E\) is the set of edges. The cost of an edge \((u, v)\) is the Euclidean distance \(c_{uv}\) between \(u\) and \(v\). Let the edge cost matrix be \(C = [c_{ij}]\) which satisfies the triangle inequality.

**Definition 1.** Symmetric TSP (STSP) is TSP in Euclidean distance (called ESTSP) and the edge cost matrix \(C\) is symmetric.

**Definition 2.** \(\triangle\)STSP is a STSP whose edge costs are non-negative and satisfies the triangle inequality, i.e., for any three distinct nodes (not necessary neighboring) \((i, j, k)\), \(c_{ij} + c_{jk} \geq c_{ik}\).
**Definition 3.** TSP tour. Given a symmetric graph \( G \) in 2-dimensional Euclidean distance and its distance matrix \( C \) where \( c_{ij} \) denote the distance between node \( i \) and \( j \) (symmetrically). A tour \( T \) has length

\[
L = \sum_{k=0}^{N-1} c_{T(k),T(k+1)}
\]

where \( N \) is the total number of nodes in \( G \) and \( T(N) = T(0) \) so that a feasible tour is formed.

**Definition 4.** The approximation ratio of an algorithm. The ratio is the result obtained by the algorithm over the optimum (abbreviated as OPT in this paper), the OPT is obtained by LKH codes [14] in this paper.

**Observation 1.** The probability density function of all tour lengths in a random ESTSP can be modelled by a Generalized Beta (GB) distribution. This is observed in Fig. 1 and other random TSPs for which we can obtain all tour lengths. More results are provided in next section. This is also validated and shown in [Vig 2008], where a scaled Beta distribution is applied with scaled mean and scaled variance. The author validated estimated results by Anderson-Darling (A-D) test and Kolmogorov-Smirnov (K-S) test for random TSP. They use these estimates to fit various candidate distributions including the Beta, Weibull and Normal cases, and conclude that the (scaled) Beta distribution yields the best fit.

We further propose a Generalized Beta (GB) distribution. The probability density function (pdf) of GB is defined as

\[
f(x, \alpha, \beta, A, B) = \frac{(x - A)^{\alpha-1}(B - x)^{\beta-1}}{Beta(\alpha, \beta)}
\]

where \( Beta(\alpha, \beta) \) is the beta function

\[
Beta(\alpha, \beta) = \int_0^1 t^{\alpha-1}(1 - t)^{\beta-1} dt
\]

\( A \) and \( B \) is the lower bound and upper bound respectively, \( \alpha > 0, \beta > 0 \), see [Hahn and Shapiro, page 91-98,126-128,1967]. For TSP, \( A \) and \( B \) represents the minimum and maximum tour length respectively.

The four central moments, mean \( (\mu) \), variance, skewness and kurtosis of the Generalized Beta distribution with parameters \( (\alpha, \beta, A, B) \) are given by:

\[
\mu(x) = A + (B - A) \frac{\alpha}{\alpha + \beta}
\]

\[
Var(x) = (B - A)^2 \frac{\alpha \beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}
\]

\[
Skewness(x) = \frac{2(\beta - \alpha)\sqrt{1 + \alpha + \beta}}{\sqrt{\alpha + \beta}(2 + \alpha + \beta)}
\]

\[
Kurtosis(x) = \frac{6[\alpha^3 + \alpha^2(1 - 2\beta) + \beta^2(1 + \beta) - 2\alpha\beta(2 + \beta)]}{\alpha \beta(\alpha + \beta + 2)(\alpha + \beta + 3)}
\]
The standard deviation is then given by
\[ \sigma(x) = \sqrt{Var(x)} \] (8)

Once four central moments are known, or any four parameters of \((A, B, \mu, \text{Var, skewness, kurtosis})\) are given, then four parameters of GB, i.e., \((\alpha, \beta, A, B)\) can be determined easily from the four moments match using Eqns.(4)-(7). When the problem size is not large, the four central moments can be computed exactly using methods proposed in [Stuffle 2009]. As the problem size increases, we can find any four parameters of \((A, B, \mu, \text{Var, skewness})\) firstly, then find four parameters of GB, i.e., \((\alpha, \beta, A, B)\).

For medium or large size problem, currently it is not easy to find the fourth central moment. However, the lower bound \((A)\) can be easily computed by LKH code [14]. So in the following sections, we find four values \((A, \text{mean } (\mu), \text{variance, skewness})\) firstly, and then compute other parameters \((B, \alpha, \beta)\). Firstly we introduce a method to compute maxTSP \((B)\) [Gutin et al.,2002].

**Definition 5.** maxTSP. The maximum tour length \((B)\) is obtained using LKH where each edge cost \((c_{ij})\) is replaced by a very large value \((M)\) minus the original edge cost, i.e., \((M-c_{ij})\). \(M\) can be set as the maximum edge cost plus 1.

Since the characteristics of random TSP and TSPLIB instances are differerent in Euclidean space, we introduce the GB as the probability density function for them separately.

### B. GB as the Probability Density Function for Random TSP

For medium size random TSP problems with \(n\) varying from 20 to 100, we can obtain four central moments easily [Suffle 2009] and apply four moments match to find four parameters for GB. After obtaining four parameters, we then use linear regression to find closed-form solution to \((\alpha, \beta)\) of GB for random TSP. For \(n=20\) to \(n=100\), and we find that
\[
\alpha(n) = 1.9197n - 32.166, R^2 = 0.9994 \quad (9)
\]
\[
\beta(n) = 1.1168n - 15.854, R^2 = 0.9982 \quad (10)
\]
\[
A(n) = 0.6932\sqrt{n} + 0.8029, R^2 = 0.9956 \quad (11)
\]
\[
B(n) = 0.7649n - 0.6393, R^2 = 0.998 \quad (12)
\]

where \(R^2\) is a measure of goodness-of-fit with value between 0 and 1, the larger the better. We observe that Eqns.(9)-(12) are highly accurate by extensive computation results.

Alternatively, applying results in Eqns.(4)-(7), after some math reduction, the lower bound \((A)\) for different size \((n)\) random TSP is given as:
\[
A(n) = \mu(n) - \sigma(n)\sqrt{\frac{\alpha(n) + \beta(n)}{\beta(n)}} \quad (13)
\]
and upper bound \((B)\) is
\[
B(n) = \mu(n)(1 + \frac{\beta(n)}{\alpha(n)}) - A(n)\frac{\beta(n)}{\alpha(n)} \quad (14)
\]
Observation 2. The relative errors between estimated results (maxTSPs) by GB and LKH results are within 6.5% for random TSP.

The relative error is defined as $(\text{EstimatedValue} - \text{OPT}) / \text{OPT}$. We conduct tests for $n=100$ to $n=500$. The results are shown in Fig.3 where LKH is used to obtain maxTSP (B) results. It is widely accepted that LKH works as the benchmark for TSP problems, see [Cook, 2012][Applegate et al, 2003] and references therein. We can observe that the relative error between our results and LKH are within 6.5% off the true optimums, with an average below 5%.

![Fig. 3. The relative error between estimated results by GB and OPT by LKH for Random TSP $n=200$ to 500](image-url)

C. GB as the Probability Density Function for TSPLIB Instances in Euclidean Distance

Firstly, we show an example of Burma14.tsp, which we can permute its all tours and find that minimum tour length $(A)$ is 3233 and maximum tour length $(B)$ is 9139, Mean $(\mu)=6679$, Variance=503064, Skewness=-0.0632, Kurtosis=2.7972. Fig. 4 shows exact result (in black color) by permuting all tour lengths and estimated probability density function (in green color) by four central moment match with $(A=3233, B=9579, \alpha=13.96, \beta=11.79)$. It can be observed that two results match very well.

Observation 3. The relative error between the estimated maximum tour lengths by GB and LKH results is below 7% for medium size TSPLIB instances, with an average below 5%.

We conduct tests by set $n=14$ to $n=52$ for which the four central moments can be easily computed and are given in [Stuffle 2009]. Fig. 5 shows the relative error between estimated results by GB and OPT by LKH.

In Table 1 we show four parameters of GB for TSPLIB with $n$ varying from 14 to 52.

Observing that GB models the probability density function very well of all tour lengths, we apply this closed-form solution to obtain quality-proven approximation in the following. Our results are based on Chritofides algorithm so we revisit this algorithm firstly.

D. Truncated Generalized Beta Distribution Based on Christofides Algorithm

Christofides algorithm [Christofides, 1976] performs in five steps:
Four Moments Match Parameters: $A = 3233$ $B = 9579$ $\alpha = 13.96$ $\beta = 11.79$

Fig. 4. The exact (Permute All Tours) and estimated (Four Moments Match) probability distribution of Burma14.tsp

Fig. 5. The relative error between estimated results by GB and OPT by LKH for TSPLIB instances for $n$ from 14 to 52

- (1). Finding the minimum spanning tree $MST$ of the input graph $G$;
- (2). Taking $G$ restricted to vertices of odd degrees in $MST$ as the subgraph $G^*$. This graph has an even number of nodes and is complete;
- (3). Finding a minimum weight matching $M^*$ on $G^*$;
- (4). Uniting the edges of $M^*$ with those of the $MST$ to create a graph $H$ with all vertices having even degrees;
- (5). Creating a Eulerian Tour on $H$ and reduce it to a feasible solution using the triangle inequality, a short cut is a contraction of two edges $(i,j)$ and $(j,k)$ to a single edge $(i,k)$.

It is well known that Christofides algorithm is $\frac{3}{2}$-approximation and has computational complexity $O(n^3)$, mainly because of finding minimum weight matching in the above step (3).

**Lemma 1.** Applying Christofides algorithm to a ESTSP forms a truncated GB (TGB) for the probability density function of optimal tour lengths, with expectation (average) value at most $1.5OPT-\epsilon$.  

7
TABLE I
Four parameters of GB for some TSPLIB instances

| TSPLIB  | A(OPT) | B     | α    | β    |
|---------|--------|-------|------|------|
| burma14 | 3323   | 9139  | 13.97| 11.79|
| ulysses16| 73.98  | 180.52| 10.24| 6.52 |
| gr17    | 2085   | 6160  | 19.22| 10.60|
| gr21    | 2707   | 10680 | 32.95| 19.80|
| ulysses22| 75.3   | 241.50| 17.52| 12.79|
| gr24    | 1272   | 4929  | 51.55| 27.21|
| fri26   | 937    | 3681  | 28.87| 16.91|
| bayg29  | 1610   | 6654  | 42.17| 26.42|
| bays29  | 2020   | 8442  | 45.14| 27.52|
| dantzig42| 4355  | 699   | 40.28| 22.69|
| swiss42 | 6676   | 1273  | 58.07| 32.47|
| gr48    | 30021  | 5046  | 79.96| 49.90|
| hk48    | 68691  | 11461 | 49.96| 28.57|
| att48   | 70347  | 10628 | 43.60| 25.88|
| eil51   | 2356.43| 428.85| 71.99| 45.14|
| berlin52| 39698.12| 7544.35| 57.40| 23.40|

TABLE II
Four parameters of GB for some random TSP instances

| n   | A(OPT) | B     | α    | β    |
|-----|--------|-------|------|------|
| 90  | 7.18   | 65.75 | 179.68| 96.49|
| 91  | 7.15   | 69.49 | 180.38| 96.30|
| 92  | 7.01   | 69.18 | 185.51| 108.55|
| 93  | 8.00   | 71.24 | 182.22| 100.63|
| 94  | 7.73   | 68.58 | 175.38| 94.00|
| 95  | 7.77   | 69.53 | 181.09| 105.93|
| 96  | 7.40   | 73.10 | 204.60| 115.61|
| 97  | 8.00   | 74.01 | 186.85| 107.33|
| 98  | 7.73   | 76.44 | 208.67| 123.96|
| 99  | 7.54   | 74.95 | 203.46| 109.03|

Proof: This is because that Christofides algorithm assures that its result is at most 1.5OPT so that those tours with lengths more than 1.5OPT are excluded (truncated), as shown in Fig. 6 where tour lengths larger than 1.5OPT (1.5A) are truncated in black color.

The TGB in this case is truncated from above. Set $X$ as the variate of the GB, the probability density function (pdf) of TGB is given by

$$f_t^1(x, \alpha, \beta, A, B, a, b) = \frac{f(x, \alpha, \beta, A, B)}{P_r[a \leq X \leq b]}$$

$$= \int_a^b \frac{(x-A)^{a-1}(B-x)^{\beta-1}}{\int_a^b (x-A)^{a-1}(B-x)^{\beta-1}}$$

(15)

where $a=A$ and $b=1.5A$. We know that the average of Christofides algorithm is at most 1.5OPT-$\epsilon$ (set as $\mu_1^1$) where $\epsilon$ is a very small value, this is also validated in [Blaser et al., 2012].
Definition 6. Local search with \(k\)-exchange neighborhoods, also called \(k\)-opt, is the most widely used heuristic method for the TSP. \(k\)-opt is a tour improvement algorithm, where in each step \(k\) links of the current tour are replaced by \(k\) links in such a way that a shorter tour is achieved (see [Helsgaun 2009] for detailed introduction).

In [Helsgaun 2009], a method with computational complexity of \(O(k^3 + k\sqrt{n})\) is introduced for \(k\)-opt.

Lemma 2. Applying \(k\)-opt to the result of Christofides algorithm forms another TGB for probability density function of optimal tour lengths.

Proof: Applying \(k\)-opt to the result obtained by Christofide algorithm as shown in Fig.7. The TGB in this case is truncated from above. Denote the first truncation by Christofides’ algorithm
as the first truncation (\(K=1\)). The probability density function of the second TGB is given by

\[
f_2^t(x, \alpha, \beta, A, B, a_2, b_2) = \frac{(x - A)^{\alpha-1}(B - x)^{\beta-1}}{\int_{a_2}^{b_2} (x - A)^{\alpha-1}(B - x)^{\beta-1} dx}
\]  

(16)

In this case, \(a_2 = A\), \(b_2 = 1.5A\) because the distribution is based on the result after applying Christofides algorithm which assures the upper bound is at most 1.5\(A\), see Fig.7. Setting \(\hat{x} = \frac{x - A}{B - A}\), \(\hat{a}_2 = \frac{A - A}{B - A} = 0\), \(\hat{b}_2 = \frac{1.5A - A}{B - A} = 0.5A\), we have

\[
C_0 = \int_{a_2}^{b_2} (x - A)^{\alpha-1}(B - x)^{\beta-1} dx
\]

\[
= \int_{0}^{b_2} ((B - A)\hat{x})^{\alpha-1}((B - A)(1 - \hat{x})^{\beta-1} dx
\]

\[
= (B - A)^{\alpha+\beta-1}B_2(0, \hat{b}_2, \alpha, \beta)
\]

(17)

where

\[
B_2(0, t, \alpha, \beta) = \int_{0}^{t} x^{\alpha-1}(1-x)^{\beta-1} dt
\]

(18)

By the definition of the expectation (mean) value of the \(f_2^t(x, \alpha, \beta, A, B, a_2, b_2)\), denotes as \(\mu_2^t\), we have

\[
\mu_2^t - A = \int_{a_2}^{b_2} (x - A) f_2^t(x, \alpha, \beta, A, B, a_2, b_2) dx
\]

\[
= \frac{\int_{a_2}^{b_2} (x - A)^{\alpha}(B - x)^{\beta-1} dx}{C_0}
\]

\[
= (B - A)^{\alpha+\beta}B_2(0, \hat{b}_2, \alpha + 1, \beta)
\]

\[
= (B - A)\frac{B_2(0, \hat{b}_2, \alpha + 1, \beta)}{B_2(0, \hat{b}_2, \alpha, \beta)}
\]

\[
=> \mu_2^t = A + (B - A)\frac{B_2(0, \hat{b}_2, \alpha + 1, \beta)}{B_2(0, \hat{b}_2, \alpha, \beta)}
\]

(19)

Iteratively we apply this approach, taking the expectation value of \((K-1)\)-th iteration as the upper bound (\(\hat{b}_K = \frac{\mu_{K-1}^t - A}{B - A}\)) of the \(K\)-th iteration, we have the expectation value after \(K\) iterations \((K \geq 2)\), denoted as denotes as \(\mu_K^t\),

\[
\mu_K^t = A + (B - A)\frac{B_2(0, \hat{b}_K, \alpha + 1, \beta)}{B_2(0, \hat{b}_K, \alpha, \beta)}
\]

\[
= A + (B - A)g(\hat{b}_K)
\]

We observe that \(g(\hat{b}_K)\) can be approximated by \(\hat{b}_K\) when \(n > 20\). Fig.8 shows \(g(\hat{b}_K)\) where \(n\) is varying from 20 to 200 and we can see that \(g(\hat{b}_K)\) is decreasing from 0.18 to 0.04 and \(g(\hat{b}_K)\) is close to zero as \(n\) increases.
The four parameters of GB for some random TSP and TSPLIB instances are provided in Table 1 and 2 respectively.

Next we provide the proof for our main theorem.

**THEOREM 1.** Applying TGB iteratively, we can obtain quality-proved approximation, i.e.,

\[(1 + \frac{1}{2}(\frac{\alpha + 1}{\alpha + 2}))^{K-1}\]-approximation where \(K\) is the number of iterations in TGB, \(\alpha\) is the shape parameter of TGB and can be determined once TSP instance is given.

**Proof:** Notice that the expectation value of \((K-1)\)-iteration is taken as the upper bound \((\hat{b}_K = \frac{\mu^{K-1}_K - A}{B - A})\) of \(K\)-iteration, as shown in Fig.7. Setting

\[
g(\hat{b}_K) = \frac{B_2(0, \hat{b}_K, \alpha + 1, \beta)}{B_2(0, b_K, \alpha, \beta)}
\]  

(20)

The exact expression of \(g(\hat{b}_K)\) can be stated in a hypergeometric series, and

\[
B_2(0, \hat{b}_K, \alpha, \beta) = \frac{x^\alpha}{\alpha} G(\alpha, 1 - \beta, \alpha + 1, \hat{b}_K)
\]

(21)

and

\[
G(a, b, c, x) = 1 + \frac{ab}{c} x + \frac{a(a + 1)b(b + 1)}{c(c + 1)2!} x^2 + \frac{a(a + 1)(a + 2)b(b + 1)(b + 2)}{c(c + 1)(c + 2)3!} x^3 + ...
\]

(22)

When \(x\) is small (less than 0.5 for all cases) and \(\alpha, \beta\) is large (from a few tens to a few hundreds in all cases), we can take the second term in above equation to approximate \(G(a, b, c, x)\), and set \(\hat{b}_K = \frac{\mu^{K-1}_K - A}{B - A}\) for \(K \geq 2\), this is because the expectation value of \((K - 1)\)-th iteration is taken as the upper bound of \(K\)-th truncation, we find that

\[
g(\hat{b}_2) \leq \frac{\alpha + 1}{\alpha + 2} \frac{0.5A}{B - A}
\]

(23)
\[ u_i^2 = A + (B - A)g(\hat{b}_2) \leq A + 0.5A \frac{\alpha + 1}{\alpha + 2} \] (24)

continue this for \( g(\hat{b}_3), u_i^3, g(\hat{b}_4), u_i^4, \ldots \), so forth, we have

\[
g(\hat{b}_K) = \frac{B_2(0, \hat{b}_K, \alpha + 1, \beta)}{B_2(0, b_K, \alpha, \beta)} \leq \frac{\alpha + 1}{\alpha + 2} \hat{b}_K = \frac{0.5A(\frac{\alpha + 1}{\alpha + 2})^{K-1}}{B - A}, \text{ when } n > 20 \text{ and } \hat{b}_K < 0.5 \] (25)

Therefore

\[
\mu_i^K = A + (B - A) \frac{B_2(0, \hat{b}_K, \alpha + 1, \beta)}{B_2(0, b_K, \alpha, \beta)} = A + (B - A)g(\hat{b}_K) \leq (1 + \frac{1}{2}(\frac{\alpha + 1}{\alpha + 2})^{K-1})A
\] (26)

This means that, after the \( K \)-th iterative truncation, we can obtain the expectation value of \( \mu_i^K \) which is close to the optimum (OPT=A) when \( K \) increases, and the worst case result \( \mu_i^{K-1} \) is the expectation of the \( (K - 1) \)-iteration.

Notice that the practical way to achieve this is to apply Chained LK which is reported [Applegate 2003] to have results within 0.4% off the true optimum on the average. Table 3 shows OPT, \( \alpha, \beta \), iteration numbers and the approximation ratio (Appr) for TSPLIB instances with \( n \leq 600 \), where \((\alpha, \beta)\) is obtained from \((A, \text{mean, variance, skewness})\) in Eqns (4)-(7), and \( K \) is obtained in LKH code. We observe that the TGB results are consistent with LKH OPT results in most cases, there are only a few cases where TGB results are few percentage difference from OPT, with 0.2% off the true optimum on the average. For instance, the difference is 7.8% for berlin52.tsp and 1.1% for ulysses22.tsp. Our results are consistent with [Applegate 2003]. These results validate Theorem 1. Notice that LKH code performs very fast in practice and our results are based on average performance.

IV. CONCLUSIONS

In this paper, for the first time, we proposed GB and Truncated Generalized Beta distribution (TGB) for the probability distribution of optimal tour lengths in a symmetric TSP in Euclidean space. Applying TGB iteratively, we can obtain quality-proved approximation, i.e., \((1 + \frac{1}{2}(\frac{\alpha + 1}{\alpha + 2})^{K-1})\) where \( K \) is the number of iterations in TGB and \( \alpha(>>1) \) is the shape parameters of TGB. The result can approach the true optimum as \( K \) increases. Once the TSP instance is given, \( \alpha \) can be computed readily from GB. As for applying TGB iteratively, a practical method is using Chained LK which was proposed in [Applegate 2003], combined with dynamical \( k \)-opt. Dynamical \( k \)-opt means that the program can adaptively selects the value of \( k \) so that in each iteration (truncation), less length tour can be found. Notice that our TGB results are based on expectation (average) value of probability distribution, which may be overestimated for the number of iterations. In practice, LKH algorithm performs very fast. A few possible research directions include:
| Instance   | A(OPT) | α   | β    | K-1  | $\left(1 + 0.5(\frac{\alpha + 1}{\alpha + 2})^{K-1}\right)A$ | Appr  |
|------------|--------|-----|------|------|-------------------------------------------------|-------|
| ulysses16  | 73.98  | 10.24 | 6.52 | 209  | 74.26                                           | 1.000 |
| ulysses22  | 75.30  | 17.52 | 12.78 | 91   | 82.17                                           | 1.0119|
| eil51      | 428.85 | 71.99 | 45.13 | 1127 | 428.85                                          | 1.000 |
| berlin52   | 7544.35| 57.40 | 23.41 | 101  | 9052.43                                         | 1.0784|
| st70       | 677.09 | 104.10| 65.00 | 1429 | 677.09                                           | 1.000 |
| eil76      | 544.35 | 113.35| 67.82 | 1509  | 544.35                                          | 1.000 |
| pr76       | 108159.41 | 115.54 | 68.48 | 2451  | 108159.41                                      | 1.000 |
| rat99      | 1219.21| 128.18| 90.84 | 1821  | 1219.21                                         | 1.000 |
| kroA100    | 21285.39| 137.30 | 98.17 | 3366  | 21285.39                                       | 1.000 |
| kroB100    | 22139.03| 132.53 | 91.06 | 4460  | 22139.03                                       | 1.000 |
| kroC100    | 20750.71| 136.14 | 101.14| 3763  | 20750.71                                       | 1.000 |
| kroD100    | 21294.24| 139.69 | 93.37 | 2659  | 21294.24                                       | 1.000 |
| kroE100    | 22068.71| 133.90 | 96.65 | 3729  | 22068.71                                       | 1.000 |
| rd100      | 7910.35| 157.44 | 98.22 | 4341  | 7910.35                                         | 1.000 |
| db101      | 640.18 | 156.51 | 97.84 | 1280  | 640.18                                          | 1.000 |
| lin105     | 14382.96| 141.04 | 101.03| 750   | 14420.26                                       | 1.0026|
| pr107      | 44301.66| 126.14 | 100.59| 2216  | 44301.66                                       | 1.000 |
| pr124      | 59030.71| 180.89 | 105.40| 3937  | 59030.71                                       | 1.000 |
| bier127    | 118293.46| 193.75 | 77.48 | 3895  | 118293.46                                      | 1.000 |
| ch130      | 6110.66| 220.15 | 133.87 | 3941  | 6110.66                                         | 1.000 |
| pr136      | 96770.89| 209.07 | 140.55 | 12134 | 96770.89                                       | 1.000 |
| pr144      | 58587.13| 231.32 | 139.79 | 5051  | 58587.13                                       | 1.000 |
| ch150      | 6530.83 | 252.56 | 150.72 | 9999  | 6530.83                                        | 1.000 |
| kroA150    | 26524.79| 216.22 | 153.52 | 5500  | 26524.79                                        | 1.000 |
| kroB150    | 26127.28| 209.02 | 157.42 | 11839 | 26127.28                                       | 1.000 |
| pr152      | 73821.19| 218.85 | 111.24 | 3520  | 73821.19                                        | 1.000 |
| u159       | 42075.63| 239.25 | 169.86 | 3257  | 42075.63                                        | 1.000 |
| rat195     | 2333.79 | 275.99 | 194.22 | 15816 | 2333.79                                        | 1.000 |
| d198       | 15808.58| 139.08 | 43.58  | 27297 | 15808.58                                       | 1.000 |
| kroA200    | 29369.30| 292.61 | 212.81 | 9977  | 29369.30                                       | 1.000 |
| kroB200    | 29440.32| 296.15 | 204.24 | 16149 | 29440.32                                       | 1.000 |
| ts225      | 126645.92| 413.60 | 244.05 | 25812 | 126645.92                                      | 1.000 |
| tsp225     | 3858.99 | 358.25 | 200.48 | 17616 | 3858.99                                        | 1.000 |
| pr226      | 80370.25| 359.00 | 210.84 | 10995 | 80370.25                                       | 1.000 |
| gil262     | 2385.71 | 470.04 | 282.52 | 16855 | 2385.71                                        | 1.000 |
| pr264      | 49134.99| 320.08 | 264.68 | 13798 | 49134.99                                       | 1.000 |
| a280       | 2586.75 | 443.37 | 303.37 | 34916 | 2586.75                                        | 1.000 |
| pr299      | 48194.85| 422.28 | 307.86 | 29117 | 48194.85                                       | 1.000 |
| lin318     | 42042.44| 563.15 | 343.07 | 39112 | 42042.44                                       | 1.000 |
| rd400      | 15275.79| 735.15 | 438.52 | 34936 | 15275.79                                       | 1.000 |
| fl417      | 11914.07| 535.26 | 389.80 | 120725| 11914.07                                       | 1.000 |
| pr439      | 107215.22| 636.74 | 307.91 | 63939 | 107215.22                                      | 1.000 |
| pcb442     | 50783.48| 806.62 | 466.13 | 81145 | 50783.48                                       | 1.000 |
| d493       | 35018.32| 695.93 | 357.46 | 129767| 35018.32                                       | 1.000 |
| u574       | 36934.49| 947.24 | 612.98 | 122692| 36934.49                                       | 1.000 |
| rat575     | 6796.36 | 892.03 | 609.61 | 84814 | 6796.36                                        | 1.000 |

- Improving the computational complexity. Currently the Christofides algorithm with minimum perfect matching has computational complexity $O(n^3)$. For large instances, this should be reduced.
- Find more efficient ways to compute especially the third and fourth central moments of a
given TSP instance. From our intensive computation experiences, we observe that the more accurate the four moments are, the better the estimated results by GB and TGB. Currently the computational complexity for the third and fourth central moment is $O(n^4)$ and $O(n^6)$ respectively [Stuffle 2009]. These are not practical for large size problems and more efficient ways should be researched.

- Finding more applications. With closed-form probability density function at hand, a lot of things can be done better. For instance, computing more statistical metrics, analyzing the average performance of approximation algorithms and others.

Acknowledgement
This research is partially supported by China National Science Foundation (CNSF) with project ID 61450110440 and Fundamental Research Funds for the Central Universities (No.ZYGX2012J073).

REFERENCES
[1] D. Applegate, W. Cook, A. Rohe, Chained Lin-Kernighan for large traveling salesman problems, INFORMS Journal on Computing; Winter 2003; 15, 1; ABI/INFORM Global pg. 82.
[2] H.C. An, Robert Kleinberg, David B. Shmoys, Improving Christofides Algorithm for the s-t Path TSP, 2012.
[3] M. Blaser, K. Panagiotou and B. V. R. Rao, A Probabilistic Analysis of Christofides’ Algorithm, Algorithm Theory-SWAT 2012, Lecture Notes in Computer Science Volume 7357, 2012, pp 225-236.
[4] N. Christofides, Worst-case analysis of a new heuristic for the travelling salesman problem, Report 388, Graduate School of Industrial Administration, CMU, 1976.
[5] W. Cook, In Pursuit of the Traveling Salesman, Princeton University Press, 2012.
[6] In G. Gutin and A. Punnen, editors, The Traveling Salesman Problem and its Variations, Chapter 1 The Maximum Traveling Salesman Problem, Kluwer Academic Publishers, New York, NY, USA, 2002.
[7] B. L. Golden. Closed-form statistical estimates of optimal solution values to combinatorial problems. In ACM 78: Proceedings of the 1978 annual conference, pages 467-470, New York, NY, USA, 1978. ACM.
[8] G.J. Hahn and S.S. Shpiro, Statistical Models in Engineering, John Wiley & Sons, Inc., New York, 1967.
[9] K. Helsgaun,General k-opt submoves for the Lin-Kernighan TSP heuristic. Mathematical Programming Computation, 2009, doi: 10.1007/s12532-009-0004-6.
[10] D. S. Johnson, L. A. McGeoch, The Traveling Salesman Problem: A Case Study in Local Optimization, Local search in combinatorial optimization 1, 215-310, 1998.
[11] D.S. Johnson, The Traveling Salesman Problem for random points in the unit square: An experimental and statistical study. Technical Report, 2014.
[12] E. Klarreich, Computer Scientists Find New Shortcuts for Infamous Traveling Salesman Problem, SIMONS SCIENCE NEWS, Jan. 29, 2013
[13] S. Lin, B.W. Kernighan, An effective heuristic algorithm for the traveling-salesman problem. Oper. Res.21, 498516 (1973).
[14] LKH codes, [http://www.akira.ruc.dk/~keld/research/LKH/](http://www.akira.ruc.dk/~keld/research/LKH/), last accessed Jan. 15th of 2015.
[15] G. Reinelt, TSPLIB, a traveling salesman problem library. ORSA Journal on Computing, 3(4):376-384, 1991.
[16] S. Reiter and D. B. Rice. Discrete optimizing solution procedures for linear and nonlinear integer programming problems. Management Science, 12(11):829-850, 1966.
[17] P. Stuffle, Moments over the Solution Space TSP, PhD Thesis, University of Technology, Sydney, 2009.
[18] V. Vig and U. S. Palekar. On estimating the distribution of optimal traveling salesman tour lengths using heuristics. European Journal of Operational Research, 186(1):111-119, April 2008.
[19] J. Vygen, New approximation algorithms for the TSP, OPTIMA 90 (2012), 1-12.