A Traveling Salesman in a Stationary World with *Sometimes* Too Many Twin Cities (Depending on the Scale)

J. Michael Steele, Wharton School, U. Pennsylvania,
(joint work with Alessandro Arlotto, Fuqua School, Duke U.)

INFORMS San Francisco 2014

Probability Theory of the Traveling Salesman Problem

Probability Theory of the Traveling Salesman Problem

- 1: The Classical Theorem of Beardwood, Halton, and Hammersley

Probability Theory of the Traveling Salesman Problem

- 1: The Classical Theorem of Beardwood, Halton, and Hammersley
  - (Few's Estimate, Basic BHH, Computational Connections and Comparison to MST)

Probability Theory of the Traveling Salesman Problem

- 1: The Classical Theorem of Beardwood, Halton, and Hammersley
  - (Few's Estimate, Basic BHH, Computational Connections and Comparison to MST)
- II: A Bookend Conjecture

Probability Theory of the Traveling Salesman Problem

- 1: The Classical Theorem of Beardwood, Halton, and Hammersley
  - (Few's Estimate, Basic BHH, Computational Connections and Comparison to MST)
- II: A Bookend Conjecture
  - (Progress with Uniformity — Old and New, Stationarity as the Frontier)

Probability Theory of the Traveling Salesman Problem

- 1: The Classical Theorem of Beardwood, Halton, and Hammersley
  - (Few's Estimate, Basic BHH, Computational Connections and Comparison to MST)
- II: A Bookend Conjecture
  - (Progress with Uniformity — Old and New, Stationarity as the Frontier)
- III: A Construction You Can Use

Probability Theory of the Traveling Salesman Problem

- 1: The Classical Theorem of Beardwood, Halton, and Hammersley
  - (Few's Estimate, Basic BHH, Computational Connections and Comparison to MST)
- II: A Bookend Conjecture
  - (Progress with Uniformity — Old and New, Stationarity as the Frontier)
- III: A Construction You Can Use
  - (Blocks and Twin Cities at Scale, Infinitely many transformations, Coup de graĉe via extreme point theory)

Probability Theory of the Traveling Salesman Problem

- 1: The Classical Theorem of Beardwood, Halton, and Hammersley
  - (Few's Estimate, Basic BHH, Computational Connections and Comparison to MST)
- II: A Bookend Conjecture
  - (Progress with Uniformity — Old and New, Stationarity as the Frontier)
- III: A Construction You Can Use
  - (Blocks and Twin Cities at Scale, Infinitely many transformations, Coup de graĉe via extreme point theory)
- IV: Where Else Does This Lead?

Probability Theory of the Traveling Salesman Problem

- 1: The Classical Theorem of Beardwood, Halton, and Hammersley
  - (Few's Estimate, Basic BHH, Computational Connections and Comparison to MST)
- II: A Bookend Conjecture
  - (Progress with Uniformity — Old and New, Stationarity as the Frontier)
- III: A Construction You Can Use
  - (Blocks and Twin Cities at Scale, Infinitely many transformations, Coup de graĉe via extreme point theory)
- IV: Where Else Does This Lead?
  - ("Places where Scale Matters" — Examples, known and unknown)

Probability Theory of the TSP, MST, and More

Probability Theory of the TSP, MST, and More

- Make life simple and consider the case of the TSP and MST for sets of n points chosen at random from a density f with compact support in $\mathbb{R}^2$.

Probability Theory of the TSP, MST, and More

- Make life simple and consider the case of the TSP and MST for sets of n points chosen at random from a density f with compact support in $\mathbb{R}^2$.
- Let L_n^{MST} and L_n^{TSP} denote the lengths of the optimal tree and optimal path.

Probability Theory of the TSP, MST, and More

- Make life simple and consider the case of the TSP and MST for sets of n points chosen at random from a density f with compact support in $\mathbb{R}^2$.
- Let L_n^{MST} and L_n^{TSP} denote the lengths of the optimal tree and optimal path.
- A moments thought suggests that these should be $O(\sqrt{n})$. Few (1955) gives a worst case upper bound of $3\sqrt{n}$ for the TSP for n points in $[0, 1]^2$.

Probability Theory of the TSP, MST, and More

- Make life simple and consider the case of the TSP and MST for sets of n points chosen at random from a density f with compact support in $\mathbb{R}^2$.
- Let L_n^{MST} and L_n^{TSP} denote the lengths of the optimal tree and optimal path.
- A moments thought suggests that these should be $O(\sqrt{n})$. Few (1955) gives a worst case upper bound of $3\sqrt{n}$ for the TSP for n points in $[0,1]^2$.
- We actually have a precise limit theorem in each case:

Probability Theory of the TSP, MST, and More

- Make life simple and consider the case of the TSP and MST for sets of n points chosen at random from a density f with compact support in $\mathbb{R}^2$.
- Let L_n^{MST} and L_n^{TSP} denote the lengths of the optimal tree and optimal path.
- A moments thought suggests that these should be $O(\sqrt{n})$. Few (1955) gives a worst case upper bound of $3\sqrt{n}$ for the TSP for n points in $[0,1]^2$.
- We actually have a precise limit theorem in each case:
-
$$lim_{n\to\infty}\frac{L_n}{\sqrt{n}} = C\int_{R^2}\sqrt{f(x)}\,dx. \qquad \text{with probability one.}$$

Probability Theory of the TSP, MST, and More

- Make life simple and consider the case of the TSP and MST for sets of n points chosen at random from a density f with compact support in $\mathbb{R}^2$.
- Let L_n^{MST} and L_n^{TSP} denote the lengths of the optimal tree and optimal path.
- A moments thought suggests that these should be $O(\sqrt{n})$. Few (1955) gives a worst case upper bound of $3\sqrt{n}$ for the TSP for n points in $[0,1]^2$.
- We actually have a precise limit theorem in each case:
- $$lim_{n\to\infty}\frac{L_n}{\sqrt{n}} = C\int_{R^2}\sqrt{f(x)}\,dx. \qquad \text{with probability one.}$$
- For the TSP this is the famous Beardwood-Halton-Hammersly theorem of 1959. For the MST the result is from Steele (1988). The constants C_{TSP} and C_{MST} are *not* known exactly.

Computational Theory of the TSP

Computational Theory of the TSP

- The TSP Problem is NP-Complete.

Computational Theory of the TSP

- The TSP Problem is NP-Complete.
- People now solve large — but not super large — TSPs with clever LP code.

Computational Theory of the TSP

- The TSP Problem is NP-Complete.
- People now solve large — but not super large — TSPs with clever LP code.
- For about 20 years there have been polynomial time ϵ-approximation algorithms, but even in theory these are $O(n^{p(1/\epsilon)})$ algorithms.

Computational Theory of the TSP

- The TSP Problem is NP-Complete.
- People now solve large — but not super large — TSPs with clever LP code.
- For about 20 years there have been polynomial time ϵ-approximation algorithms, but even in theory these are $O(n^{p(1/\epsilon)})$ algorithms.
- Bottom Line:
  The guaranteed ϵ-approximations to the TSP are not particularly practical compared to LP, but perhaps the "coding effort" conditions are not yet balanced.

Computational Theory of the TSP

- The TSP Problem is NP-Complete.
- People now solve large — but not super large — TSPs with clever LP code.
- For about 20 years there have been polynomial time ϵ-approximation algorithms, but even in theory these are $O(n^{p(1/\epsilon)})$ algorithms.
- Bottom Line:
  The guaranteed ϵ-approximations to the TSP are not particularly practical compared to LP, but perhaps the "coding effort" conditions are not yet balanced.
- On the Other Hand:
  There is a simple $O(n)$ time, ϵ-approximation algorithm (Karp-Steele (1985)) that you can use *if you assume a probability model for the points.*

Computational Theory of the MST

Computational Theory of the MST

- The MST problem is computationally easy!

Computational Theory of the MST

- The MST problem is computationally easy!
- Greedy algorithms of several kinds give $O(n^2)$ algorithms.

Computational Theory of the MST

- The MST problem is computationally easy!
- Greedy algorithms of several kinds give $O(n^2)$ algorithms.
- With fancy data structures you can even do better.

Computational Theory of the MST

- The MST problem is computationally easy!
- Greedy algorithms of several kinds give $O(n^2)$ algorithms.
- With fancy data structures you can even do better.
- Current records are faster than $O(n \log^{\epsilon} n)$.

Computational Theory of the MST

- The MST problem is computationally easy!
- Greedy algorithms of several kinds give $O(n^2)$ algorithms.
- With fancy data structures you can even do better.
- Current records are faster than $O(n \log^{\epsilon} n)$.
- Bottom Line:
  The MST and TSP may LOOK like similar problems ....

Computational Theory of the MST

- The MST problem is computationally easy!
- Greedy algorithms of several kinds give $O(n^2)$ algorithms.
- With fancy data structures you can even do better.
- Current records are faster than $O(n \log^{\epsilon} n)$.
- Bottom Line:
  The MST and TSP may LOOK like similar problems ....
- BUT:
  Their computational theory tells us that they are wildly different.

Does one Need Independence for the BHH to Hold?

Does one Need Independence for the BHH to Hold?

- The natural "book end" conjecture is that the BHH might hold for a stationary ergodic process with values in $[0,1]^2$, and there are reasons to "think positive"...

Does one Need Independence for the BHH to Hold?

- The natural "book end" conjecture is that the BHH might hold for a stationary ergodic process with values in $[0,1]^2$, and there are reasons to "think positive"...
  - There is a "weak BHH" that only requires that asymptotically each sub-square gets its "fair number" of point.

Does one Need Independence for the BHH to Hold?

- The natural "book end" conjecture is that the BHH might hold for a stationary ergodic process with values in $[0,1]^2$, and there are reasons to "think positive"...
  - There is a "weak BHH" that only requires that asymptotically each sub-square gets its "fair number" of point.
  - The result in that case is

$$\lim_{n\to\infty} \frac{\log L_n}{\log n} = \frac{1}{2}$$

Does one Need Independence for the BHH to Hold?

- The natural "book end" conjecture is that the BHH might hold for a stationary ergodic process with values in $[0,1]^2$, and there are reasons to "think positive"...
  - There is a "weak BHH" that only requires that asymptotically each sub-square gets its "fair number" of point.
  - The result in that case is

$$\lim_{n\to\infty} \frac{\log L_n}{\log n} = \frac{1}{2}$$

  - Adams and Nobel (2010) established VC theory for Stationary Ergodic Processes; so sub-squares *uniformly* get "fair share."

Does one Need Independence for the BHH to Hold?

- The natural "book end" conjecture is that the BHH might hold for a stationary ergodic process with values in $[0,1]^2$, and there are reasons to "think positive"...
  - There is a "weak BHH" that only requires that asymptotically each sub-square gets its "fair number" of point.
  - The result in that case is

$$\lim_{n\to\infty} \frac{\log L_n}{\log n} = \frac{1}{2}$$

  - Adams and Nobel (2010) established VC theory for Stationary Ergodic Processes; so sub-squares *uniformly* get "fair share."
- Reasons to "think negative"?

Does one Need Independence for the BHH to Hold?

- The natural "book end" conjecture is that the BHH might hold for a stationary ergodic process with values in $[0,1]^2$, and there are reasons to "think positive"...
  - There is a "weak BHH" that only requires that asymptotically each sub-square gets its "fair number" of point.
  - The result in that case is

$$\lim_{n\to\infty} \frac{\log L_n}{\log n} = \frac{1}{2}$$

  - Adams and Nobel (2010) established VC theory for Stationary Ergodic Processes; so sub-squares *uniformly* get "fair share."
- Reasons to "think negative"?
- The class of Stationary Ergodic Process is very rich. They look probabilistically the "same" in any block of n observations —but a priori they *may* reveal different kinds of "granularity" over larger block sizes.

Construction of a Counter-Example

Construction of a Counter-Example

- The basic idea is to construct a stationary ergodic process such that certain special times $\tau_1, \tau_2, \ldots, \tau_k \ldots$ most cities are part of a pair of twin cities, and the TSP cost is only a fraction what the TSP cost would be for an independent sequence.

Construction of a Counter-Example

- The basic idea is to construct a stationary ergodic process such that certain special times $\tau_1, \tau_2, \ldots, \tau_k \ldots$ most cities are part of a pair of twin cities, and the TSP cost is only a fraction what the TSP cost would be for an independent sequence.
- Yet for the long periods between these times (think Permian or Pleistocene Eras) one has very normal behavior where the TSP cost is comparable to that of a independent sample.

Construction of a Counter-Example

- The basic idea is to construct a stationary ergodic process such that certain special times $\tau_1, \tau_2, \ldots, \tau_k \ldots$ most cities are part of a pair of twin cities, and the TSP cost is only a fraction what the TSP cost would be for an independent sequence.
- Yet for the long periods between these times (think Permian or Pleistocene Eras) one has very normal behavior where the TSP cost is comparable to that of a independent sample.
- To make this pathology appear once is not hard. To make it appear infinitely many times requires that one over come several concrete challenges.

A Transformation to Create Twin Cities: $T_{N,\epsilon}$

A Transformation to Create Twin Cities: $T_{N,\epsilon}$

- Start with a doubly infinite stationary process $\mathcal{X} = \{X_k : k \in \mathbb{Z}\}$ with the marginal uniform distribution on $[0,1]^2$.

A Transformation to Create Twin Cities: $T_{N,\epsilon}$

- Start with a doubly infinite stationary process $\mathcal{X} = \{X_k : k \in \mathbb{Z}\}$ with the marginal uniform distribution on $[0,1]^2$.
- We break this into blocks of size N: For $-\infty < k < \infty$ we set

$$B_k = X_{kN}, X_{kN-1}, \dots, X_{(k+1)N-1}$$

A Transformation to Create Twin Cities: $T_{N,\epsilon}$

- Start with a doubly infinite stationary process $\mathcal{X} = \{X_k : k \in \mathbb{Z}\}$ with the marginal uniform distribution on $[0, 1]^2$.
- We break this into blocks of size N: For $-\infty < k < \infty$ we set

$$B_k = X_{kN}, X_{kN-1}, \ldots, X_{(k+1)N-1}$$

- We then introduce "shifted blocks" of size N:

$$B_k(\epsilon) = X_{kN} +' \epsilon, X_{kN-1} +' \epsilon, \ldots, X_{(k+1)N-1} +' \epsilon$$

A Transformation to Create Twin Cities: $T_{N,\epsilon}$

- Start with a doubly infinite stationary process $\mathcal{X} = \{X_k : k \in \mathbb{Z}\}$ with the marginal uniform distribution on $[0,1]^2$.
- We break this into blocks of size N: For $-\infty < k < \infty$ we set

$$B_k = X_{kN}, X_{kN-1}, \ldots, X_{(k+1)N-1}$$

- We then introduce "shifted blocks" of size N:

$$B_k(\epsilon) = X_{kN} +' \epsilon, X_{kN-1} +' \epsilon, \ldots, X_{(k+1)N-1} +' \epsilon$$

- We lay down the blocks and the "shifted blocks" to create a sequence $\hat{\mathcal{X}} \in ([0,1])^\infty$:

$$\ldots B_{-1}, B_{-1}(\epsilon), B_0, B_0(\epsilon), \ldots B_k, B_k(\epsilon) \ldots$$

A Transformation to Create Twin Cities: $T_{N,\epsilon}$

- Start with a doubly infinite stationary process $\mathcal{X} = \{X_k : k \in \mathbb{Z}\}$ with the marginal uniform distribution on $[0,1]^2$.
- We break this into blocks of size N: For $-\infty < k < \infty$ we set

$$B_k = X_{kN}, X_{kN-1}, \ldots, X_{(k+1)N-1}$$

- We then introduce "shifted blocks" of size N:

$$B_k(\epsilon) = X_{kN} +' \epsilon, X_{kN-1} +' \epsilon, \ldots, X_{(k+1)N-1} +' \epsilon$$

- We lay down the blocks and the "shifted blocks" to create a sequence $\hat{\mathcal{X}} \in ([0,1])^\infty$:

$$\ldots B_{-1}, B_{-1}(\epsilon), B_0, B_0(\epsilon), \ldots B_k, B_k(\epsilon) \ldots$$

- We shift $\hat{\mathcal{X}}$ by $I \sim U\{0,1,\ldots,2N-1\}$ to get $\widetilde{\mathcal{X}} = T_{N,\epsilon}(\mathcal{X})$.

Confronting the Hard Part: Infinite Repetition

Confronting the Hard Part: Infinite Repetition

- Start with an i.i.d uniform sequence on $[0,1]^2$. Call it $\mathcal{X}_0$. Let N_n, ϵ_n be given, $n = 1, 2, \ldots$

Confronting the Hard Part: Infinite Repetition

- Start with an i.i.d uniform sequence on $[0,1]^2$. Call it $\mathcal{X}_0$. Let N_n, ϵ_n be given, $n = 1, 2, \ldots$
- For $n = 1, 2, \ldots$ define $\mathcal{X}_n = T_{N_n, \epsilon_n}(\mathcal{X}_{n-1})$

Confronting the Hard Part: Infinite Repetition

- Start with an i.i.d uniform sequence on $[0,1]^2$. Call it $\mathcal{X}_0$. Let N_n, ϵ_n be given, $n = 1, 2, \ldots$
- For $n = 1, 2, \ldots$ define $\mathcal{X}_n = T_{N_n,\epsilon_n}(\mathcal{X}_{n-1})$
- Show that if N_n and ϵ_n are wisely chosen then the processes $\{\mathcal{X}_n : n = 1, 2, \ldots\}$ converge in distribution to a stationary process $\mathcal{X}$ with uniform marginal distributions.

Confronting the Hard Part: Infinite Repetition

- Start with an i.i.d uniform sequence on $[0,1]^2$. Call it $\mathcal{X}_0$. Let N_n, ϵ_n be given, $n = 1, 2, \ldots$
- For $n = 1, 2, \ldots$ define $\mathcal{X}_n = T_{N_n, \epsilon_n}(\mathcal{X}_{n-1})$
- Show that if N_n and ϵ_n are wisely chosen then the processes $\{\mathcal{X}_n : n = 1, 2, \ldots\}$ converge in distribution to a stationary process $\mathcal{X}$ with uniform marginal distributions.
- This gives us our *feed stock*. This is the guy you want to use if you face a similar problem.

Confronting the Hard Part: Infinite Repetition

- Start with an i.i.d uniform sequence on $[0,1]^2$. Call it $\mathcal{X}_0$. Let N_n, ϵ_n be given, $n = 1, 2, \ldots$
- For $n = 1, 2, \ldots$ define $\mathcal{X}_n = T_{N_n,\epsilon_n}(\mathcal{X}_{n-1})$
- Show that if N_n and ϵ_n are wisely chosen then the processes $\{\mathcal{X}_n : n = 1, 2, \ldots\}$ converge in distribution to a stationary process $\mathcal{X}$ with uniform marginal distributions.
- This gives us our *feed stock*. This is the guy you want to use if you face a similar problem.
- To extract what you need from $\mathcal{X}$ there are several issues to be addressed.

Springing the Trap

Springing the Trap

- It is easier to deal with the means $EL_n(\mathcal{X})$ than to deal with the issues of almost sure convergence. For the purpose of a counter-example, this is sufficient.

Springing the Trap

- It is easier to deal with the means $EL_n(\mathcal{X})$ than to deal with the issues of almost sure convergence. For the purpose of a counter-example, this is sufficient.
- One shows

$$\liminf_{n\to\infty} \frac{EL_n(\mathcal{X})}{\sqrt{n}} \leq c_1 < c_2 \leq \limsup_{n\to\infty} \frac{EL_n(\mathcal{X})}{\sqrt{n}}$$

Springing the Trap

- It is easier to deal with the means $EL_n(\mathcal{X})$ than to deal with the issues of almost sure convergence. For the purpose of a counter-example, this is sufficient.
- One shows

$$\liminf_{n\to\infty} \frac{EL_n(\mathcal{X})}{\sqrt{n}} \leq c_1 < c_2 \leq \limsup_{n\to\infty} \frac{EL_n(\mathcal{X})}{\sqrt{n}}$$

- If $\mathcal{M}_e$ denotes the set of ergodic shift-invariant measures on $([0,1]^2)^\infty$ and if μ is the measure determined by $\mathcal{X}$, then Dynkin (1978) tells us there is a $D_\mu(d\nu)$ such that

$$\mu(\mathcal{A}) = \int_{\mathcal{M}_e} \nu(\mathcal{A})\, D_\mu(d\nu).$$

Springing the Trap

- It is easier to deal with the means $EL_n(\mathcal{X})$ than to deal with the issues of almost sure convergence. For the purpose of a counter-example, this is sufficient.
- One shows

$$\liminf_{n\to\infty} \frac{EL_n(\mathcal{X})}{\sqrt{n}} \le c_1 < c_2 \le \limsup_{n\to\infty} \frac{EL_n(\mathcal{X})}{\sqrt{n}}$$

- If $\mathcal{M}_e$ denotes the set of ergodic shift-invariant measures on $([0,1]^2)^\infty$ and if μ is the measure determined by $\mathcal{X}$, then Dynkin (1978) tells us there is a $D_\mu(d\nu)$ such that

$$\mu(\mathcal{A}) = \int_{\mathcal{M}_e} \nu(\mathcal{A})\, D_\mu(d\nu).$$

- We then argue (pretty easily) that some ν in $\mathcal{M}_e$ must inherit the bad behavior of the μ determined by $\mathcal{X}$

How Does This Help You?

How Does This Help You?

- The BHH is a wonderful theorem with a large literature and many applications.

How Does This Help You?

- The BHH is a wonderful theorem with a large literature and many applications.
- Still, there are no meaningful extensions to dependent random variables.

How Does This Help You?

- The BHH is a wonderful theorem with a large literature and many applications.
- Still, there are no meaningful extensions to dependent random variables.
- The example described here explains the obstruction, at least in part.

How Does This Help You?

- The BHH is a wonderful theorem with a large literature and many applications.
- Still, there are no meaningful extensions to dependent random variables.
- The example described here explains the obstruction, at least in part.
- Mixtures of independent sequences satisfy the BHH and the technical part of the present analysis isolates a special class of "locally uniform processes" that satisfy the BHH. This is *all* that is know in the positive direction.

How Does This Help You?

- The BHH is a wonderful theorem with a large literature and many applications.
- Still, there are no meaningful extensions to dependent random variables.
- The example described here explains the obstruction, at least in part.
- Mixtures of independent sequences satisfy the BHH and the technical part of the present analysis isolates a special class of "locally uniform processes" that satisfy the BHH. This is *all* that is know in the positive direction.
- **Two Natural Questions:**

How Does This Help You?

- The BHH is a wonderful theorem with a large literature and many applications.
- Still, there are no meaningful extensions to dependent random variables.
- The example described here explains the obstruction, at least in part.
- Mixtures of independent sequences satisfy the BHH and the technical part of the present analysis isolates a special class of "locally uniform processes" that satisfy the BHH. This is *all* that is know in the positive direction.
- **Two Natural Questions:**
- Does the BHH hold for a stationary, uniform process if it is *strongly mixing*?

How Does This Help You?

- The BHH is a wonderful theorem with a large literature and many applications.
- Still, there are no meaningful extensions to dependent random variables.
- The example described here explains the obstruction, at least in part.
- Mixtures of independent sequences satisfy the BHH and the technical part of the present analysis isolates a special class of "locally uniform processes" that satisfy the BHH. This is *all* that is know in the positive direction.
- **Two Natural Questions:**
- Does the BHH hold for a stationary, uniform process if it is *strongly mixing*?
- What does the iterated $T_{N,\epsilon}$ construction tell us about *other spacial limit theorems*?