Mathematical Probability

Some topics we understand, some we don't

slides at `www.stat.berkeley.edu/users/aldous`

Structure of talk:

- Get quickly to statement of my favourite open problem. Despite "applied" story, it really is a "pure maths" problem.

- From this standpoint, look in some different directions and survey classical and recent research areas; mention two "hot topics" where probability meets other maths fields.

(not: orderly logical progression . . . . . .)

Open problem: in the Travelling Salesman Problem on random points, is there asymptotic essential uniqueness of the optimal tour?

Travelling Salesman Problem (TSP)

- Given n points in the plane, find shortest tour through all n points.

Really an algorithmic problem, but we won't pursue algorithmic aspects.

We consider the case of n random points, distributed uniformly in some square of area n. The optimal tour T_n has some random length $L(T_n)$. Beardwood-Halton-Hammersley (1959, *Proc. Cambridge Phil. Soc.*) showed

$$\text{mean}(L(T_n)) \sim c_{\mathsf{BHH}}\, n$$

for some constant c_{BHH}. Will mention later other results/questions about $L(T_n)$. Almost nothing is know about the geometry of the actual tour T_n itself.

Elementary that, for each n:

(i) the optimal tour is unique (with prob. 1).

(ii) can create a near-optimal tour by altering the optimal tour on some small proportion of edges.

But is this the only way to get near-optimal tours? Does the random TSP have the **asymptotic essential uniqueness** property (AEU):

- for large n, every almost-optimal tour is just a small modification of the optimal tour.

I will give two equivalent ways to state a precise definition. First is unsophisticated but messy, second is sophisticated but clean.

Unsophisticated definition

For a non-optimal tour T'_n define

$$\Delta(T'_n) = \text{number of edges in } T'_n \setminus T_n.$$

Fix $\varepsilon > 0$ and consider the set of "near-optimal" tours T'_n with $L(T'_n) \leq (c_{\mathsf{BHH}} + \varepsilon)n$.

Define the AEU property as:

$\forall \varepsilon > 0$ $\exists \delta(\varepsilon)$ such that, with probability $\to 1$ as $n \to \infty$,

$\forall T'_n : \dfrac{L(T'_n)}{n} \leq (c_{\mathsf{BHH}} + \varepsilon)$ implies $\dfrac{\Delta(T'_n)}{n} \leq \delta(\varepsilon)$.

(require $\delta(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$).

[Experts on TSP algorithms suggest "false"; analogies with loosely related rigourously understood two-dimensional percolation models suggest "true"]

Sophisticated definition

Recall: n points, uniform random in the square $[-n^{1/2}/2,\ n^{1/2}/2]^2$, centred at origin.

The $n \to \infty$ limit is the well-understood Poisson point process of rate 1 on the infinite plane. Loosely

Pr(some point in A) $\sim$ area(A) as area(A) $\to 0$.

Consider an "object" consisting of
(i) an infinite set of points in R^2
(ii) a collection of infinite paths, each point from (i) being in exactly 1 path.

Now consider a probability measure μ on the set of such objects, such that
(a) the process of particles is Poisson, rate 1
(b) μ is invariant under the action of the Euclidean group.

For any such μ, can define a real number

$$c(\mu) := \text{mean length of typical edge in the path-collection.}$$

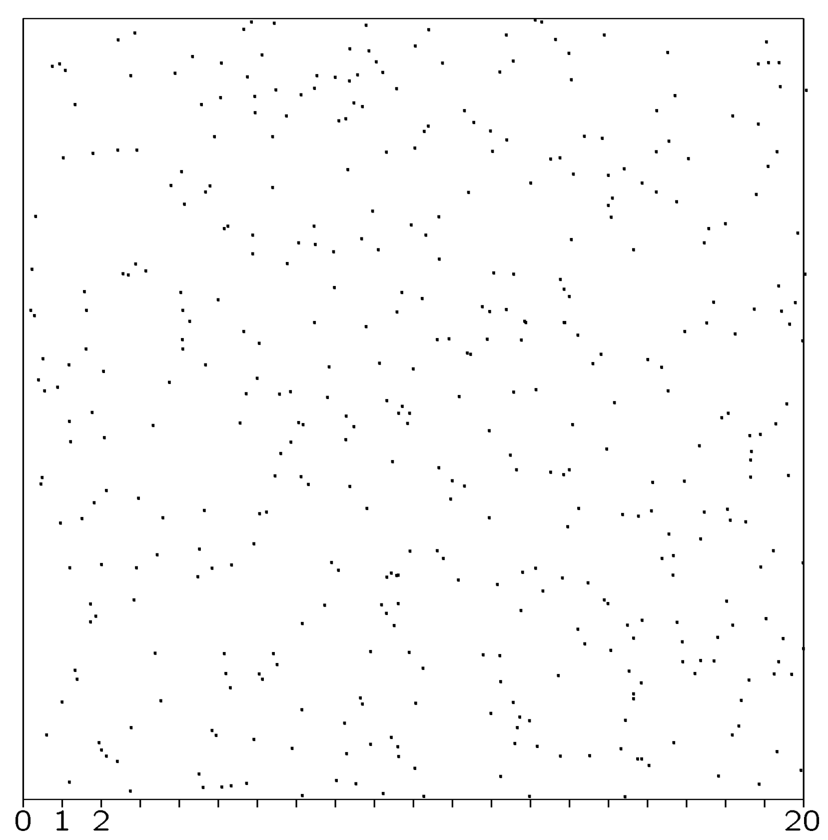
Now soft compactness/convergence arguments imply

$$\begin{aligned} c_{\mathsf{BHH}} &= \inf_\mu c(\mu) \\ \exists\mu & \text{ attaining the infimum} \\ \mathsf{AEU} &\Leftrightarrow \text{ infimum attained by unique } \mu. \end{aligned}$$

This looks like pure maths! (I hope)

No-one knows how to tackle the AEU problem for Euclidean TSP.

Why care? Interface of mathematical probability, theory of algorithms, and statistical physics of "disordered systems".



Poisson (rate 1) point process on a 20×20 square.

largest empty disc

smallest disc with 4 points

Poisson (rate 1) point process on the plane.

Elementary properties

(1) For prespecified set A of area $|A|$,

Pr(exactly k points in A) $= \exp(-|A|)|A|^k/k!$

$$k = 0, 1, 2, \ldots.$$

This is the Poisson dist. with mean $|A|$.

(2) Pr(nearest point to 0 at distance $> r$)

$$= \exp(-\pi r^2), \ r > 0.$$

Can continue with exact calculations like

Pr(typical point is its nearest neighbour's nearest neighbour).

This general style of "geometric probability" question goes back over 100 years.

Classical applied probability

Explicit, asymptotically-correct formulas for what happens inside $L \times L$ box to random quantities like

• radius of largest disc containing no points.
• radius of smallest disc containing m points.
• minimum (over all m-subsets $\{x_1, \ldots, x_m\}$ of points and all lines l) $\max_{1 \le i \le m}$ distance(x_i, l).

Feasible because they have a "finite-dimensional parametrization". See Kendall - Kendall (1980) for collinearity; Aldous (1989) for 100 back-of-envelope calculations for random extrema.

Loosely, the previous questions are solvable because they need only "purely local" calculations. Questions involving global structure are much harder.

Continuum percolation

Make each random point the centre of a disc of radius r. Then $\exists r_{\mathsf{crit}}$ such that infinite connected clusters of overlapping discs

exist for $r > r_{\mathsf{crit}}$; don't exist for $r < r_{\mathsf{crit}}$.

Known/conjectured for general percolation-type models

- At exact critical point, infinite clusters <u>do not</u> exist

- above critical point, the infinite cluster is unique.

Can view c_{BHH} as critical value for existence of TSP-paths with mean edge-length c, but for TSP we <u>do</u> have existence at the critical point.

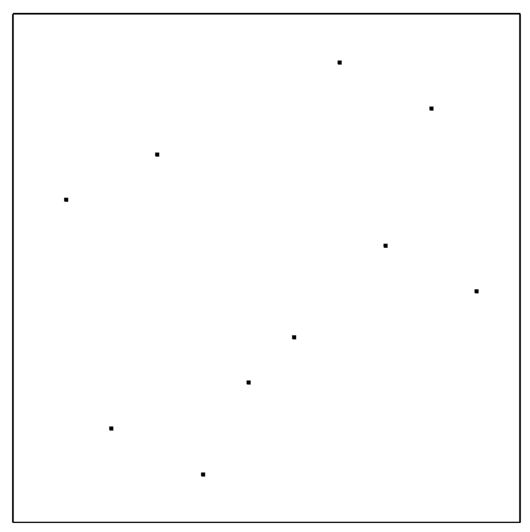
Hot Topic # 1

Percolation studied intensively on the integer lattice Z^d; simplest "bond-percolation" model has each edge present with probability p – see Grimmett (1999) monograph. Prototype of "disordered system" in statistical physics; often used as comparison in study of more complicated spatial random systems.

In $d = 2$, long-conjectured that at the critical value, if we look at percolation probabilities from one set to another, and rescale to a continuum limit, the limit probabilities should have a certain conformal invariance property. Recently proved in one model; stochastic Loewner evolution processes (derived from Loewner's (1923) slit mapping theorem) introduced and shown to describe scaling limit of boundary of critical percolation clusters.

Two of the main people in this field (Smirnov, Werner) will speak soon at the Conference on Stochastic Processes and their Applications in Cambridge, July 9 – 13.

Exactly one "join random points into paths" problem where critical value is explicitly known. n points in square determine a permutation π as follows.



7 2 8 1 3 4 10 6 9 5

Using independent uniform points gives a uniform random permutation. Hammersley (1972) observed: an increasing subsequence of π

$$7\ 2\ 8\ 1^*\ 3^*\ 4^*\ 10\ 6^*\ 9^*\ 5$$

corresponds to points which can be joined in an up-right path from the bottom-left to the top-right corners of the square.

So length $L(m)$ of longest increasing subsequence in random permutation of $m = n^2$ elements is essentially the same as

$\bar{L}_n :=$ maximum number of points of Poisson process which can be joined in an up-right path from $(0,0)$ to (n,n).

It turns out that

$$\begin{aligned} \text{mean}(\bar{L}_n) &\sim cn, \; c = 2 \\ \text{standard deviation}(\bar{L}_n) &\sim c' n^{1/3} \end{aligned}$$

the latter proved by Baik - Deift -Johansson (1999) as part of

Hot Topic # 2

Eigenvalues of random matrices, Toeplitz determinants and Painlevé equations.

The set of $m \times m$ Hermitian matrices has a natural probability measure (Gaussian Unitary Ensemble). Consider largest eigenvalue $\lambda^*(m)$. Known that as $m \to \infty$

$$\begin{aligned} \text{mean}(\lambda^*(m)) &\sim \sqrt{2m}, \\ \text{standard deviation}(\lambda^*(m)) &\sim c' m^{1/6} \end{aligned}$$

which is same behavior as the length $L(m)$ of LIS of m-permutation:

$$\begin{aligned} \text{mean}(L(m)) &\sim \sqrt{2m}, \\ \text{standard deviation}(L(m)) &\sim c' m^{1/6} \end{aligned}$$

and even the same non-Gaussian rescaled limit distribution, given in terms of the solution of a Painlevé II function.

See Tracy - Widom (1999).

If X_n is a random quantity which is a sum of n components of order(1) size, and if mean$(X_n) \sim cn$, then the "normal" behavior is

$$\text{standard deviation}(X_n) \sim c' n^{1/2}$$

corresponding to the Central Limit Theorem and the Gaussian limit distribution for sums of independent quantities.

In the longest increasing subsequence setting, we have $O(n^{1/3})$ spread instead; also it is clear that AEU fails.

On the other hand, if instead of the TSP we consider the minimal spanning tree (MST) on n random points in an area-n square, then its length $L(M_n)$ does follow the "normal" behavior and satisfies the Central Limit Theorem (technically hard); easy to show AEU holds, from greedy algorithm for construction of MST.

Intriguing whether there is any general relationship between AEU and the normal order-$n^{1/2}$ growth of standard deviation. General <u>concentration of measure</u> arguments apply to TSP to give $O(n^{1/2})$ upper bound, but ......

Final variation on TSP is to throw away geometry. Take n points and assume the $\binom{n}{2}$ inter-point distances are independent random quantities; scale so that mean distance to nearest neighbour $= 1$.

Remarkably, there exists a well-defined $n \to \infty$ limit structure, and the limit constant

$$c := \lim_n n^{-1} \text{ mean}(L(T_n))$$

has the analogous reformulation as the mean edge-length in an optimal tour on the limit structure. In contrast to the Euclidean case, this reformulation enables us to compute c.
(The ultimate calculation is simple, technical justification hard: Aldous (2001) gives details in related problem).

D.J. Aldous. Probability Approximations via the Poisson Clumping Heuristic. Springer, 1989.

D.J. Aldous. The $\zeta(2)$ limit in the random assignment problem. To appear in *Random Structures Algorithms*, 2001.

J. Baik, P.A. Deift, and K. Johansson. On the distribution of the length of the longest increasing subsequence of random permutations. *J. Amer. Math. Soc.*, 12:1119–1178, 1999.

J. Beardwood, H.J. Halton, and J.M. Hammersley. The shortest path through many points. *Proc. Cambridge Phil. Soc.*, 55:299–327, 1959.

J.J. Bentley. Fast algorithms for geometric traveling salesman problems. *ORSA J. Comput.*, 4:387–411, 1992.

G.R. Grimmett. *Percolation*. Springer-Verlag, Berlin, 2nd edition, 1999.

J.M. Hammersley. A few seedlings of research. In *Proc. Sixth Berkeley Symp. Math. Statist. and Probability, Volume 1*, pages 345–394. University of California Press, 1972.

D.G. Kendall and W. S. Kendall. Alignments in two-dimensional random sets of points. *Adv. in Appl. Probab.*, 12:380–424, 1980.

C.A. Tracy and H. Widom. Random unitary matrices, permutations and Painlevé. *Comm. Math. Phys.*, 207:665–685, 1999.