Random directed and on-line neighbour networks

Mathew D. Penrose

University of Bath, UK

Workshop on Mathematical Aspects of Large Networks,
Centre de Recerca Matemàtica
November 2009

STOCHASTIC GEOMETRY

Let $\mathcal{B}_n := \{X_1, \ldots, X_n\}$, with

$\{X_i\}$ independent uniform random points in $[0, 1]^2$, and consider some function $F(\mathcal{B}_n)$. Can we give criteria under which the classical limit laws hold for F?

Motivation: Spatial/multivariate statistics

Analysis of algorithms.

Example: If $F(\mathcal{B})$ is the length of the optimal traveling salesman tour on $\mathcal{B}$, then $F(\mathcal{B}_n)$ is known to satisfy a Law of Large Numbers but not known to satisfy a Central Limit Theorem.

Many functionals F on $\mathcal{B}_n$, however, DO satisfy the Central Limit Theorem, i.e. for $Z \sim N(0, \sigma^2)$, some σ:

$$P\left[F(\mathcal{B}_n) - \mathbf{E}F(\mathcal{B}_n) \leq t\right] \to P[Z \leq t]$$

Examples: Central Limit Theorem holds when F is the total length

for any of the following graphs:

- **Nearest neighbour graph**
- **Minimal spanning tree** (connected graph of min. total length)
- **Delaunay graph** (X_i, X_j adjacent if they have adjacent Voronoi cells)

MORE EXAMPLES SATISFYING CENTRAL LIMIT THEOREM:

- Number of components of $G(\mathcal{B}_n, r)$.
- Size of largest component of $G(\mathcal{B}_n, rn^{-1/2})$ (r supercritical).

Here, $G(\mathcal{B}, \rho)$ is the geometric graph with vertex set $\mathcal{B}$ and $xy \in E$ iff $|x - y| \leq \rho$.

See Penrose (2003).

In all of these these examples, a notion of **stabilization** (local dependence) holds.

We can write

$$F(\mathcal{B}) = \sum_{x \in \mathcal{B}} \xi(x, \mathcal{B})$$

with $\xi(x, \mathcal{P} \cup \{x\})$ unaffected by changes to a Poisson process $\mathcal{P}$ beyond a random but finite radius with well-behaved tail.

General results (e.g. P. and Yukich 01, 03) say that

Stabilization + Moments condition $\Longrightarrow$ CLT, LLN

EXTENSIONS OF THE GENERAL THEORY

For $F(\mathcal{B}) = \sum_x \xi(x, \mathcal{B})$ (stabilizing, moments), the CLT has been extended to:

- Non-uniform distributions. Previously had $\mathcal{B}_n = \{n^{1/2} X_i\} : 1 \leq i \leq n\}$ with $\{X_i\}$ i.i.d. uniform in $[0,1]^2$. Relax uniform condition.

- Convergence of measures. Sums of contributions from disjoint regions asymptotically independent normal.

DIRECTED NEAREST NEIGHBOUR GRAPH

(Bhatt and Roy 98–04)

Use 'coordinate-wise' partial ordering $\prec$ on $\mathbf{R^2}$.
($x \prec y$ if x south-west of y)

Connect each $X_i \in \mathcal{B}_n$ to nearest $Y \in \mathcal{B}_n \cup \{(0,0)\}$ with $Y \prec X_i$.

Resulting **directed NNG** on $\mathcal{B}_n \cup \{(0,0)\}$ is also a tree, the **Minimal Directed Spanning Tree.**
Motivation: Communications. Drainage.

QUANTITIES OF INTEREST for DNNG.

- $L_n :=$ **total length** of edges of DNNG $(\mathcal{B}_n)$.
- $L_n^0 :=$ total length of edges **incident to** $(0,0)$.
- $M_n :=$ **maximum** edge length.

P. and Wade 04–06: Identify the limiting distribution for L_n^0, M_n and for $L_n - EL_n$ as $n \to \infty$

LENGTH OF EDGES AT $(0,0)$ in DNNG, L_n^0.

Say X_i is **minimal** in $\mathcal{B}_n$ if no $X_j \prec X_i$ exists.

$$L_n^0 = \sum_{i: X_i \text{ minimal}} |X_i|$$

Take strips S_x, S_y, of width $n^{-1/3}$ along x and y axes respectively, contributing L_n^x (resp. L_n^y) to L_n^0. Then with high probability,

$$L_n^0 \approx L_n^x + L_n^y$$

List the **minimal** points of $\mathcal{B}_n$ in order of increasing y-coordinate, as $\xi_1, \xi_2, \xi_3, \ldots$. In coordinates,

$$\xi_1 = (U_1, V_1),\ \xi_2 = (U_1U_2, V_2),\ \xi_3 = (U_1U_2U_3, V_3), \cdots$$

Then the U_i are indep. $U(0,1)$, and if we restrict to points in S_x the V_i are all small. So summing suggests that

$$L_n^x \Longrightarrow U_1 + U_1U_2 + U_1U_2U_3 + \cdots := Z$$

and sim. for L_n^y, so that

$$L_t^0 \Longrightarrow Z + Z' = Z_2$$

DICKMAN-TYPE DISTRIBUTIONS.

With $U_1, U_2, \ldots$ indep. Uniform(0,1), let

$$Z := U_1 + U_1U_2 + U_1U_2U_3 + \cdots$$

Z has a **Dickman** dist., with **fixed-point eqn**:

$$Z = U_1(1 + Z') \quad \text{in distribution}$$

from which one can derive the MGF of Z. Also, Z_2 satisfies the fixed-point equation

$$Z_2 = U^{1/2}(1 + Z_2).$$

MAXIMUM EDGE LENGTH IN DNNG, M_n

Set M_n^x = max length of DNNG edges from pts near the x-axis. Let U_i be the x-coordinate of the ith lowest point.

Asymptotically, M_n^x is the maximum over i of the distance from U_i to its nearest neighbour to the left in the set $\{0, U_1, U_2, \ldots, U_{i-1}\}$. Denote this maxmimum by M. Then

$$M \stackrel{\mathcal{D}}{=} \max(U, (1-U)M').$$

and

$$M_n \Longrightarrow \max(M, M')$$

A LINEAR FRAGMENTATION PROCESS

Given $U_1, U_2, U_3, \ldots$ (iid $U(0,1)$), let D_n be the distance from U_n to its nearest neighbour to the left in the linear point process $\{0, U_1, \ldots, U_{n-1}\}$. Let $H_n = \sum_{i=1}^{n} D_i$. Then:

$$H_n = U + UH'_M + (1-U)H''_{n-1-M}$$

where $M \sim \mathrm{Bin}(n-1, U)$.

In fact, $H_n - EH_n$ **converges** in L^2 to some H, with

$$H \stackrel{\mathcal{D}}{=} UH + (1-U)H' + U + U \ln U + (1-U)\ln(1-U)$$

TOTAL LENGTH of DNNG T_t.

Take strips S_x, S_y of width $n^{-0.6}$, and $\bar{S}_x, \bar{S}_y$ of width $n^{-0.5}$.

Let T_t^x = contribution to T_t from S_x:
T_t^y, $\bar{T}_t^x$, $\bar{T}_t^y$ similar. Then

- $T_t^x - ET_t^x \Longrightarrow H$, with H satisfying the fixed-pt eqn

$$H \stackrel{\mathcal{D}}{=} UH + (1-U)H' + U + U\ln U + (1-U)\ln(1-U)$$

- $\mathrm{Var}(\bar{T}_t^x - T_t^x) \to 0$ and $\mathrm{Var}(\bar{T}_t^y - T_t^y) \to 0$.

- Defining the ‘interior’ contribution

$$T_t^{\mathrm{i}} = T_t - (\bar{T}_t^x + \bar{T}_t^y)$$

by stabilization etc. can show for some σ^2,

$$T_t^{\mathrm{i}} - ET_t^{\mathrm{i}} \Longrightarrow W \sim N(0, \sigma^2).$$

- Also, T_t^x, T_t^y, T_t^{i} asymptotically independent, yielding the **result**

$$T_t - ET_t \Longrightarrow H + H' + W.$$

VARIANTS AND COMMENTS

Power weighted edges. Suppose edge e has weight $|e|^\alpha$. Redefine L_t, L_t^0 accordingly.

- Now L_t^0 is asymptotically GD$(2/\alpha)$.
- If $\alpha < 1$ then L_t satisfies a CLT.
- If $\alpha > 1$ then $L_t - EL_t \Longrightarrow H_\alpha + H'_\alpha$ with

$$H_\alpha \stackrel{\mathcal{D}}{=} U^\alpha H_\alpha + (1-U)^\alpha H'_\alpha + U^\alpha$$

.

Unrooted DNNG. $L_t - L_t^0 - E[\cdot] \Longrightarrow W + J + J'$:

$$J \overset{\mathcal{D}}{=} UJ + (1-U)H + U\log U + (1-U)\log(1-U) + (1-U)$$

'SOUTH' VERSION: $x \prec y$ if x south of y.
Join each point to its nearest neighbour below.
The limit has a similar form.

Quicksort. The limit distribution for the time needed to sort a random permutation of $\{1, \ldots, n\}$ satisfies a similar fix-point equation to H [Rösler 91]

On-line NNG (ONG):

Suppose $X_1, \ldots, X_n$ are independent and uniform over $[0,1]^d$.

For $i \geq 2$, connect X_i to the nearest point in $\{X_1, \ldots, X_{i-1}\}$.

- Arises from boundary behaviour of 'South' DNG in $[0,1]^{d+1}$. (P. and Wade 2009)
- Also seen in network modelling (BBBCR 2003)

Let T_n be the total length of the ONG in d dimensions.

If $d > 4$ then T_n satisfies a CLT (P. 2005)

If $d = 1$ then $T_n - ET_n$ converges to a limit variable described in terms of a distributional fixed-point equation (P. and Wade 2008).

For $d = 2, 3, 4$, the limiting distribution for T_n remains an open problem. Wade (2008) has provided a variance upper bound of $O(\max(n^{1(2/d)}, \log n))$.

Power weighted edges are also considered in these papers.

References

[1] M. Penrose, *Random Geometric Graphs*, Oxford University Press, Oxford (2003).

[2] M.D. Penrose and J.E. Yukich, *Central limit theorems for some graphs in computational geometry*, Ann. Appl. Probab. **11** (2001), 1005-1041.

[3] M. D. Penrose and J. E. Yukich, *Weak laws of large numbers in geometric probability.* Ann. Appl. Probab. **13** (2003), no. 1, 277–303.

[4] M. D. Penrose and A. R. Wade. Random and directed on-line networks. In *New perspectives in Stochastic Geometry*, eds. W. S. Kendall and I. Molchanov. Oxford University Press, December 2009.