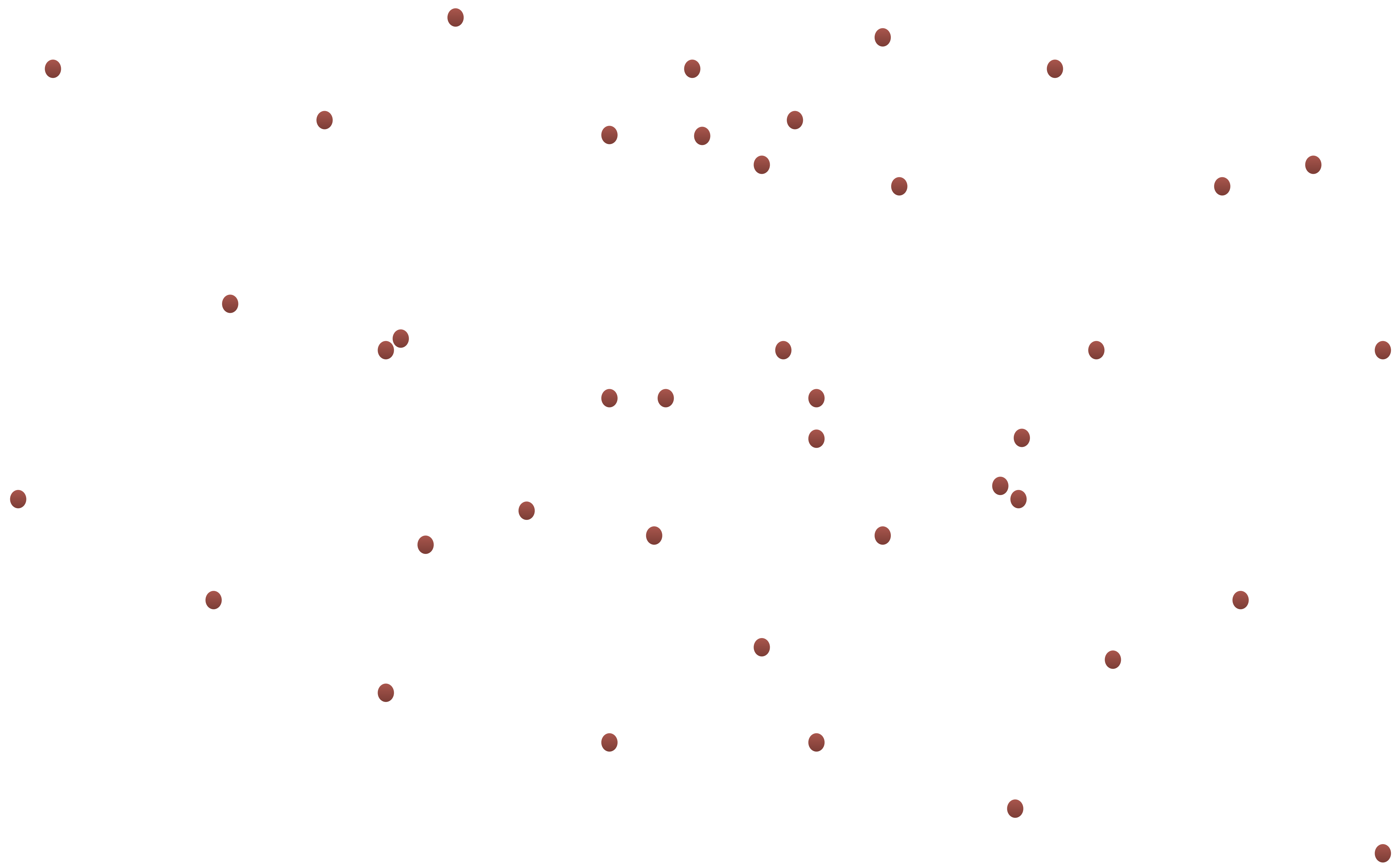
Algorithmically Random Point Processes

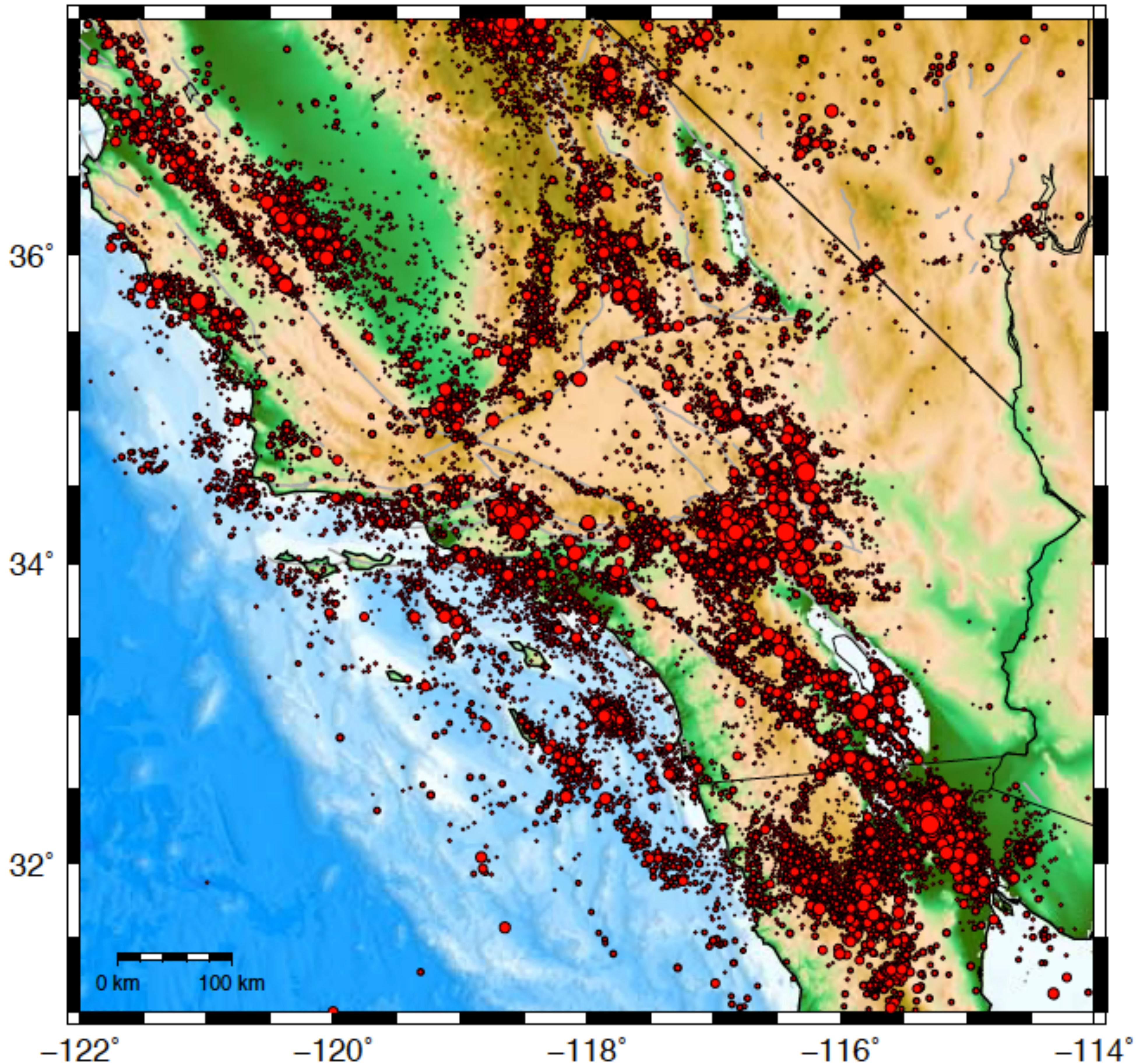
Jan Reimann



When is a countable set of points $\in \mathbb{R}^n$ random?

How "random"?

What measure?



Hauksson-Shearer-Yang catalog of southern CA earthquakes 1981-2011

Point Processes

- Random closed set (RACS): random variable

$$X : \Omega \to \mathcal{F}(E)$$

E locally compact, second countable Hausdorff space

$\mathcal{F}(E)$ closed subsets of E

- Borel structure:

  - Fell topology generated by

$$\mathcal{F}_G = \{F \in \mathcal{F} : F \cap G \neq \emptyset\} \qquad G \subseteq E \text{ open}$$

$$\mathcal{F}^K = \{F \in \mathcal{F} : F \cap K = \emptyset\} \qquad K \subseteq E \text{ compact}$$

- Simple point process: RACS that is a.s. locally finite.

Algorithmically RACS

- Find a mapping

$$f : A^{\mathbb{N}} \to \mathcal{F}(E) \qquad \text{(A finite)}$$

  and "push forward" ML-randomness [BBCDW].

- Or: Given a "nice" basis for E, one can define an effective basis for $\mathcal{F}(E)$

$$\mathcal{F}^{\overline{B_{i_1}} \cup \cdots \cup \overline{B_{i_k}}}_{B_{j_1} \cap \cdots \cap B_{j_l}}$$

  and develop ML-tests [Axon].

Choquet's Theorem

- How to specify measures on $\mathcal{F}(E)$?

- Any RACS X comes with a capacity functional $T_X : \mathcal{K}(E) \to [0, 1]$ given by

$$T_X(K) = \mathbb{P}_X(\mathcal{F}_K) = \mathbb{P}[X \cap K \neq \emptyset]$$

- Any such functional is
  - monotone
  - subadditive
  - upper semicontinuous: for $K_1 \supseteq K_2 \supseteq K_3 \supseteq, \quad T(K_i) \to T(\bigcap K_i)$
  - is completely alternating

- Choquet: Any T with these properties is the capacity of some RACS.

Examples

- Poisson process

$$T_{\Pi}(K) = 1 - e^{-\lambda m(K)} \qquad (m \text{ Lebesgue measure on } \mathbb{R}^n)$$

- Generalized Poisson process

$$T_{\mu}(K) = 1 - e^{-\mu(K)} \qquad (\mu \text{ Borel, non-atomic on } E)$$

- ETAS model: (conditional) intensity function

$$\lambda(t, x, y, m) = s(m)\left[\mu(x, y) + \sum_{\{k:\ t_k < t\}} \kappa(m_k) g(t - t_k) f(x - x_k, y - y_k; m_k)\right]$$

(non-stationary Poisson process)

Examples

- Poisson process

$$T_{\Pi}(K) = 1 - e^{-\lambda m(K)} \qquad (m \text{ Lebesgue measure on } \mathbb{R}^n)$$

- Generalized Poisson process

$$T_{\mu}(K) = 1 - e^{-\mu(K)} \qquad (\mu \text{ Borel, non-atomic on } E)$$

μ is called the intensity measure of the process

- ETAS model: (conditional) intensity function

$$\lambda(t, x, y, m) = s(m) \left[\mu(x, y) + \sum_{\{k:\ t_k < t\}} \kappa(m_k) g(t - t_k) f(x - x_k, y - y_k; m_k) \right]$$

(non-stationary Poisson process)

Questions

- Given a locally finite closed set, can we describe the RACS for which it is algorithmically random?
    - How do computational and geometric properties interact?
- What is the relation between an algorithmically RACS and the randomness of its points?
- Is there an analog of Schnorr’s Theorem?

Randomness of Sets vs Randomness of Points

- Axon: For the 1-dimensional Poisson process $\Pi_{\mathbb{R}}$ on $\mathbb{R}$, if $X \subseteq \mathbb{R}$ is algorithmically $\Pi_{\mathbb{R}}$-random, then

$$x \in X \quad \Rightarrow \quad x \text{ is ML-random.}$$

- Axon & R.; Rute: This holds for generalized Poisson processes Π_{μ}, too.

Randomness of Sets vs Randomness of Points

- Let $X = \{x_1 < x_2 < x_3 < \dots\} \subset \mathbb{R}$. Let $x_i^\circ = x_i - [x_i]$, and

$$N_k = |\{i\colon x_i \in (k, k+1]\}|$$

Then X is algorithmically $\Pi_{\mathbb{R}}$-random with intensity λ (computable) iff

  - Every finite collection of x_i° is mutually ML-random relative to (N_k), and

  - The sequence $(N_k) \in \mathbb{N}^{\mathbb{N}}$ is random with respect to the distribution induced by $$\mathbb{P}[N_k = n] = e^{-\lambda} \cdot \frac{\lambda^n}{n!}$$

A similar (albeit technically more difficult) characterization is possible for $\mathbb{R}^n$.

Dynamics and Point Processes

- Often a point process is assumed to be generated as an orbit of a dynamical system (Ω, T, μ).
- What can an (observed) orbit tell about the underlying measure?

Measures as Fractals

- Multifractal spectra try to capture these variations by studying

  (1) the local scaling behavior of a measure at a given point,

  (2) the global (average) scaling behavior of balls.

- For many measures the two aspects are closely related

  → Multifractal Formalism

Local: Pointwise Dimension

- Pointwise (local) dimension of μ at x:

Local scaling behavior at x

$$Y_\mu(x) = \lim_{\delta \to 0} \frac{\log \mu B(x, \delta)}{\log \delta}$$

(If the limit does not exist work with lim inf and lim sup.)

- Thm: For μ computable and x μ-random,

$$\dim_H x = Y_\mu(x).$$

Global: Generalized Renyi Dimensions

- Let B(x,ε) be the N-dimensional ε-ball around x.
- For $-\infty < q < \infty$, $q \neq 1$, let

$$\theta(q) = \lim_{\varepsilon \to 0} \frac{\log \left[\int (\mu B(x, \varepsilon))^{q-1} d\mu(x) \right]}{\log \varepsilon}$$

θ measures the average scaling of the q-th moment of μ(B(x,ε))

- For integer $q \geq 2$, θ(q)/q-1 is also called the correlation dimension of order q.

Multifractal Formalism: from global to local

- μ satisfies the (strong) multifractal formalism if

$$\theta(q) = \inf_y \{qy - f(y)\}$$

holds whenever $f(y) > 0$, where

$$f(y) = \dim_H\{x : Y_\mu(x) = y\}.$$

Multifractal spectrum

Multifractal Formalism: from global to local

- μ satisfies the (strong) multifractal formalism if

$$\theta(q) = \inf_y \{qy - f(y)\}$$

holds whenever $f(y) > 0$, where

Legendre transform

$$f(y) = \dim_H\{x : Y_\mu(x) = y\}.$$

Multifractal spectrum

A Universal Multifractal

- Furthermore, the multifractal spectrum of any other (computable) measure can be gauged against the spectrum of the universal semimeasure M*.

- Thm: If μ is computable, then

M* pointwise dimension

$$\dim_H f_\mu(y) = \dim_H \left\{ x \colon \frac{\overline{\dim_H x}}{\dim_\mu x} = y \right\}.$$

A Universal Multifractal

- Furthermore, the multifractal spectrum of any other (computable) measure can be gauged against the spectrum of the universal semimeasure M*.

- Thm: If μ is computable, then

M* pointwise dimension

$$\dim_H f_\mu(y) = \dim_H \left\{ x \colon \frac{\dim_H x}{\dim_\mu x} = y \right\}.$$

Billingsley dimension

Estimation and Stability of Multifractal Spectra

- Grassberger-Procaccia: $C_\mu(\varepsilon)$:= Probability two random, independent points x, y are no more than distance ε apart. By Fubini's Theorem,

$$C_\mu(\varepsilon) = \mu \times \mu\{(x,y) : \|x-y\| \leq \varepsilon\} = \int \mu B(x,\varepsilon) d\mu(x) = \theta(2)$$

- If we have only finitely many observations $x_1, \ldots, x_n$, this suggests using

$$C(n,\varepsilon) = \frac{\sum_{i=1}^{n} \sum_{j>i} 1_{\{\|x_i - x_j\| \leq \varepsilon\}}}{\binom{n}{2}}$$

as an estimator of $C_\mu(\varepsilon)$.

(Similar estimators exist for higher moments.)

- This estimator is consistent if the x_i are chosen i.i.d. according to μ.

Information Distance

- In practice, the GP estimator often suffers from a boundary effect (earthquake catalogs).
- Idea: replace the use of the Euclidean distance in the GP-algorithm by an information distance (conditional to the boundary, if applicable).
- One such distance is based on Kolmogorov complexity [Bennet et al]:

$$EC(\sigma,\tau) = C(\sigma\tau) - \min\{C(\sigma),C(\tau)\}$$

- One can prove that this is a consistent estimator, too.

Practical Issues

- For applications, we have to approximate C with
  - compressors (Lempel-Ziv etc.)
  - string complexity functions (Lempel-Ziv, Ehrenfeucht-Mycielski, Becher-Heiber)
- The consistency results still seem to go through if we work with a normal compressor (Cilibrasi-Vitanyi):

(1) Idempotency: $C(aa) = C(a)$

(2) Monotonicity: $C(ab) \geq C(a)$

(3) Symmetry: $C(ab) = C(ba)$

(4) Distributivity: $C(ab) + C(c) \leq C(ac) + C(bc)$

Practical Issues

- For applications, we have to approximate C with
  - compressors (Lempel-Ziv etc.)
  - string complexity functions (Lempel-Ziv, Ehrenfeucht-Mycielski, Becher-Heiber)
- The consistency results still seem to go through if we work with a normal compressor (Cilibrasi-Vitanyi):

(1)	Idempotency:	$C(aa) = C(a)$
(2)	Monotonicity:	$C(ab) \geq C(a)$
(3)	Symmetry:	$C(ab) = C(ba)$
(4)	Distributivity:	$C(ab) + C(c) \leq C(ac) + C(bc)$

Problem: For the popular compression algorithms it has not been formally established yet that they are normal