Journal Club – May 2021

Stabilizing Fctls
and
CLT w/
Appl. to β-Ms.

STABILIZING FUNCTIONALS AND CENTRAL LIMIT THEOREMS IN STOCHASTIC GEOMETRY WITH APPLICATIONS TO TDA

by

Johannes Krebs

(Heidelberg)

Zoom

31/5/2021

OUTLINE

PART I

- POISSON PROCESS, BINOMIAL PROCESS
- THE FRAMEWORK OF PENROSE AND YUKICH: "STABILIZING FUNCTIONALS"
- CENTRAL LIMIT THEOREMS

The Annals of Applied Probability
2001, Vol. 11, No. 4, 1005–1041

CENTRAL LIMIT THEOREMS FOR SOME GRAPHS IN COMPUTATIONAL GEOMETRY

BY MATHEW D. PENROSE AND J. E. YUKICH[1]

University of Durham and Lehigh University

Let (B_n) be an increasing sequence of regions in d-dimensional space with volume n and with union $\mathbb{R}^d$. We prove a general central limit theorem for functionals of point sets, obtained either by restricting a homogeneous

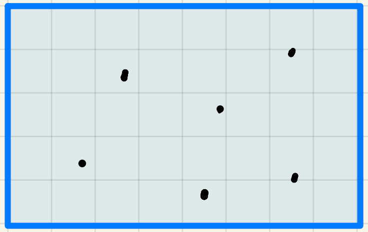
PART II

- APPLICATIONS IN TDA: CLT FOR BETTI NUMBERS
- SMOOTH BOOTSTRAP

Joint work with
- W. Polonik
- C. Hirsch
- B. Roycraft

1. DATA GENERATION

- underlying prob. space $(\Omega, \mathcal{F}, \mathbb{P})$
- **POISSON PROCESS** $\mathcal{P}$ WITH **INTENSITY** $\lambda \in \mathbb{R}_+$ on $\mathbb{R}^d$
  - **(a)** For all Borel sets $A \subseteq \mathbb{R}^d$: $\mathcal{P}(A) \sim Po(\lambda |A|)$
  - **(b)** $\forall A_1, \dots, A_k$ pairwise disjoint Borel sets $\{\mathcal{P}(A_1), \dots, \mathcal{P}(A_k)\}$ indep.



Construction: $X_1, X_2, \dots$ iid on $[0,1]^d = B$

$N \sim Po(\lambda)$

$\mathcal{P}$ on B: $X_1, X_2, \dots, X_N$

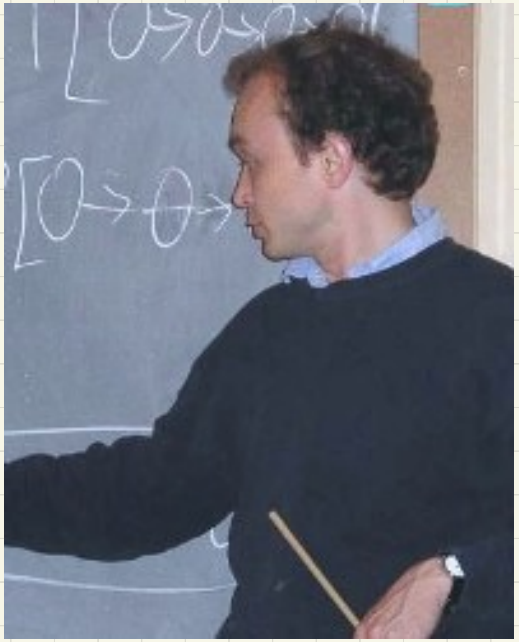
- **BINOMIAL PROCESS** $\mathcal{U}_m$ on bdd **BOREL SET** B
  - **(a)** $m \in \mathbb{N}$ $\hat{=}$ no. of realizations
  - **(b)** $X_1, \dots, X_m$ iid according to some distribution on B
- **OBSERVATION WINDOWS**: $(B_n)_{n \in \mathbb{N}}$, $B_n \subseteq \mathbb{R}^d$ s.t.
  - **(a)** $|B_n| = n/\lambda$
  - **(b)** $\bigcup_{n \in \mathbb{N}} \bigcap_{m \geq n} B_m = \mathbb{R}^d$
  - **(c)** $|\partial_r B_n| / n \to 0$ $(n \to \infty)$ "vanishing boundary"

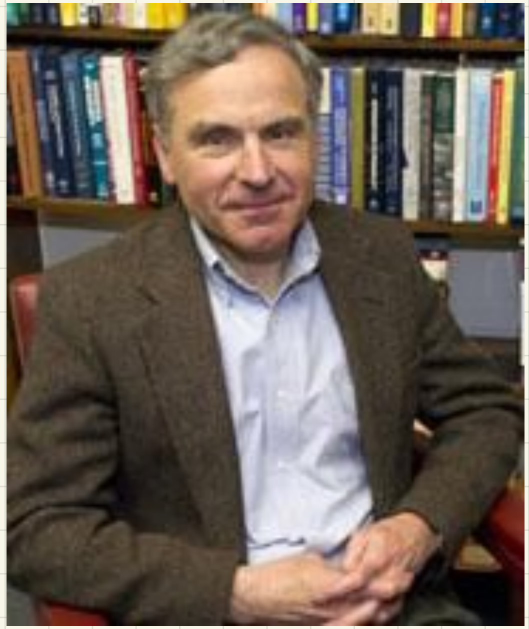
2. THE FRAMEWORK OF PENROSE & YUKICH (AOAP, 2001)

- H real-valued, defined on all finite subsets of $\mathbb{R}^d$
- H is translation invariant: $H(\mathcal{X}+x) = H(\mathcal{X}) \quad \forall \mathcal{X} \subseteq \mathbb{R}^d$ finite $\forall x \in \mathbb{R}^d$
- add-one cost: $\Delta(\mathcal{X}) = H(\mathcal{X} \cup \{0\}) - H(\mathcal{X})$

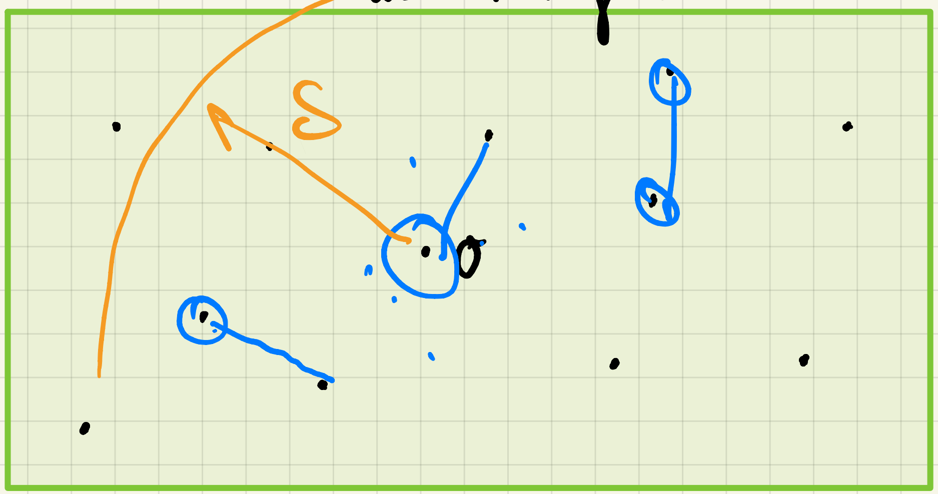
DEFINITION: H is strongly stabilizing if $\exists$ a.s. finite r.v. S ("radius of stabilization") and $\Delta(\infty)$ such that with probability 1

$$\Delta((\mathcal{P} \cap B(0,S)) \cup A) = \Delta(\infty) \quad \forall A \subseteq \mathbb{R}^d \setminus B(0,S) \text{ finite.}$$





EXAMPLE: H counts distance to nearest neighbor



3. Towards the CLT / further conditions

4

- **Uniform bounded moments condition:**

$$\sup_{\substack{A \in \mathcal{B} \\ 0 \in A}} \ \sup_{m \in \left[\frac{\lambda|A|}{2}, \frac{3\lambda|A|}{2}\right]} \mathbb{E}\left[|\Delta(\mathcal{U}_{m,A})|^4\right] < \infty,$$

where $\mathcal{B} = \{ B_n + x \mid x \in \mathbb{R}^d, n \in \mathbb{N} \}$

$\mathcal{U}_{m,A} \mathrel{\hat{=}} m$- binomial process on A w/ uniform density.

- H is **polynomially bounded**: $\exists\, \gamma \in \mathbb{R}_+$

$$|H(\mathcal{X})| \leq \gamma \left(\operatorname{diam}(\mathcal{X}) + \#\mathcal{X} \right)^{\gamma} \quad \forall \mathcal{X} \in \mathbb{R}^d \text{ finite.}$$

$\Rightarrow$ Strong stabilization, uniform bounded moments and polynomial boundedness lead to CLT for $H(\mathcal{P}_n)$ and $H(\mathcal{U}_{nn})$, where $\mathcal{P}_n = \mathcal{P}|_{B_n}$ and $\mathcal{U}_{nn}$ n-bin. process on B_n w/ unif. density.

5

THEOREM (P & Y, 2001): Suppose H is strongly stabilizing, satisfies uniform bounded moments condition on $\mathcal{B}$ and is polynomially bounded. Then there are constants $\sigma^2 \geq \tau^2 \geq 0$ such that as $n \longrightarrow \infty$

(i) $n^{-1} \operatorname{Var}(H(\mathcal{P}_n)) \to \sigma^2$,

(ii) $n^{-1/2}(H(\mathcal{P}_n) - \mathbb{E}[H(\mathcal{P}_n)]) \Rightarrow \mathcal{N}(0, \sigma^2)$,

(iii) $n^{-1} \operatorname{Var}(H(\mathcal{U}_{n,n})) \to \tau^2$,

(iv) $n^{-1/2}(H(\mathcal{U}_{n,n}) - \mathbb{E}[H(\mathcal{U}_{n,n})]) \Rightarrow \mathcal{N}(0, \tau^2)$.

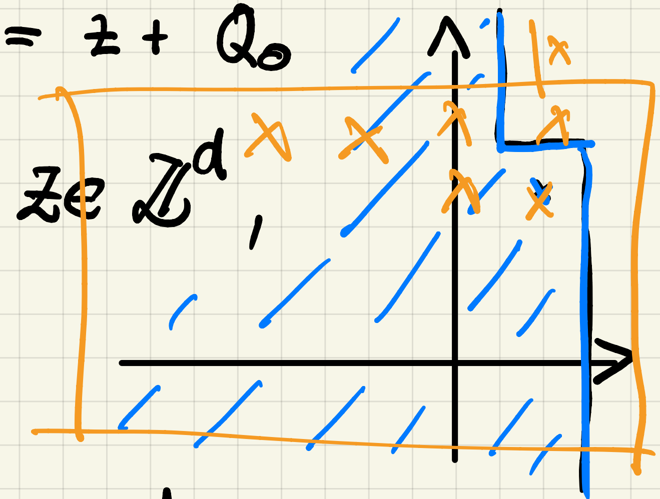
Also given λ, σ^2 and τ^2 are independent of the choice of $(B_n)_n$.
If the distribution of $\Delta(\infty)$ is non-degenerate, then $\tau^2 > 0$.

Remark: (i) One can show that $\tau^2 = \sigma^2 - (\mathbb{E}[\Delta(\infty)])^2$.

(ii) Extension to nonhomogeneous Poisson / binomial process, eg. K.D. Trinh, 2019, ECP ("$H(n^{1/d}\mathcal{P}_n)$").

Proof – main ingredients: Poisson model only

1. **Poisson filtration:** $Q_0 = [-\frac{1}{2}, \frac{1}{2}]^d$, $Q_z = z + Q_0$



$$\mathcal{F}_z = \sigma\{\mathcal{P}|_{Q_y} \mid y \preceq z,\ y \in \mathbb{Z}^d\}, \quad z \in \mathbb{Z}^d,$$

where $\preceq$ lexicographic ordering on $\mathbb{Z}^d$.

2. **Martingale differences:** $\mathbb{E}[H(\mathcal{P}_n) \mid \mathcal{F}_z] \mathrel{\hat{=}}$ cond. exp. of $H(\mathcal{P}_n)$ wrt. $\mathcal{F}_z$

$B_n' = \{z \in \mathbb{Z}^d : Q_z \cap B_n \neq \emptyset\}$; $z_1, \dots, z_{k_n}$ enumeration of B_n' wrt. lex.

$z_0 \preceq z_1$ and $Q_{z_0} \cap B_n = \emptyset$. Then

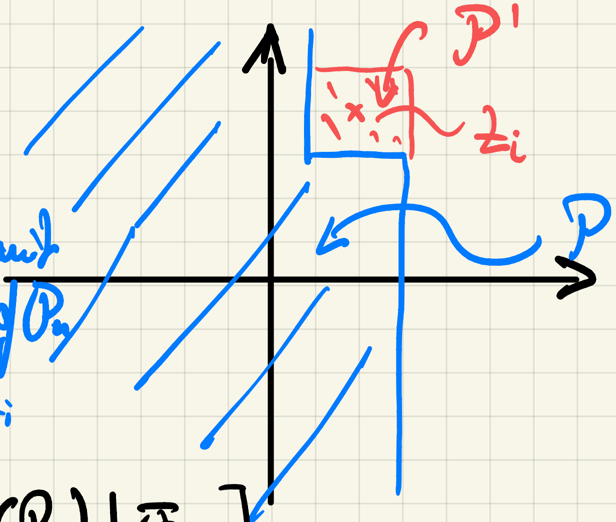
$$H(\mathcal{P}_n) - \mathbb{E}[H(\mathcal{P}_n)] = \sum_{i=1}^{k_n} \mathbb{E}[H(\mathcal{P}_n) \mid \mathcal{F}_{z_i}] - \mathbb{E}[H(\mathcal{P}_n) \mid \mathcal{F}_{z_{i-1}}]$$

3. $$\overset{(*)}{=} \sum_{i=1}^{k_n} \mathbb{E}[H(\mathcal{P}_n) - H((\mathcal{P}_n \setminus Q_{z_i}) \cap (\mathcal{P}_n' \cap Q_{z_i})) \mid \mathcal{F}_{z_i}]$$

where $\mathcal{P}'$ independent copy of $\mathcal{P}$ and $\mathcal{P}_n' = \mathcal{P}'|_{B_n}$.

(*) $\mathbb{E}[H((P_n \setminus Q_{z_i}) \cap (P_n' \cap Q_{z_i})) \mid \mathcal{F}_{z_i}]$

independent of $\mathcal{F}_{z_i}$ $\Rightarrow$ only info of P_n in Q_{z_i}



$= \mathbb{E}[H((P_n \setminus Q_{z_i}) \cap (P_n' \cap Q_{z_i}) \mid \mathcal{F}_{z_{i-1}}]$

$= \mathbb{E}[H((P_n \setminus Q_{z_i}) \cap (P_n \cap Q_{z_i})) \mid \mathcal{F}_{z_{i-1}}] = \mathbb{E}[H(P_n) \mid \mathcal{F}_{z_{i-1}}]$

7

4. Set $\Delta_{z_i}(\mathcal{B}_n) = H(P_n) - H((P_n \setminus Q_{z_i}) \cap (P_n' \cap Q_{z_i}))$, $1 \le i \le k_n$.

Then $H(P_n) - \mathbb{E}[H(P_n)] = \sum_{i=1}^{k_n} \mathbb{E}[\Delta_{z_i}(\mathcal{B}_n) \mid \mathcal{F}_{z_{i-1}}] = \sum_{i=1}^{k_n} D_{ni}$

$\Rightarrow$ CLT for martingale differences (McLeish, 1974, AoP):

unif. bdd. moment

(i) $\sup_{n \ge 1} \mathbb{E}[\max_{i \le k_n} (k_n^{-1/2} |D_{ni}|)^2] < \infty$

(ii) $k_n^{-1/2} \max_{i \le k_n} |D_{ni}| \to 0$ in probability

Strong stabilization

(iii) $k_n^{-1} \sum_{i \le k_n} D_{ni}^2 \to \lambda\sigma^2 \ge 0$ in $L^1(P)$

$(\|X\|_1 = \mathbb{E}[|X|])$

Then:

$\boxed{k_n^{-1/2} \sum_{i=1}^{k_n} D_{ni} \Rightarrow \mathcal{N}(0, \sigma^2)}$

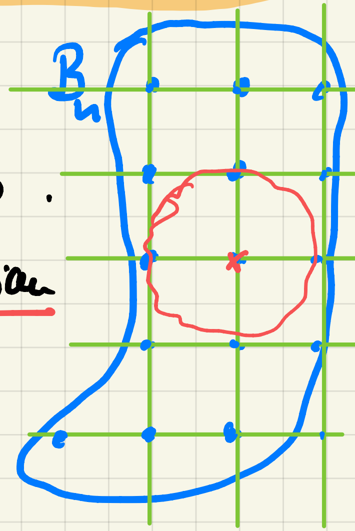
(note $\frac{k_n}{n} \to \frac{1}{\lambda}$ by constr.)

5. (i) and (ii) follow from moments condition.

8

(iii) Idea: "Each $\Delta_{z_i}(B_n)$ acts in a local neighborhood of z_i only" whp.

Indeed, using strong stabilization, have for z fixed



$$\Delta_z(B_n) \longrightarrow \Delta_z(\infty) \quad \text{a.s. as } n \to \infty.$$

Now, one can show using vanishing relative boundary condition

$$\left\| k_n^{-1} \sum_{i \le k_n} D_{ni}^2 - k_n^{-1} \sum_{i \le k_n} \mathbb{E}[\Delta_{z_i}(\infty) \mid \mathcal{F}_{z_i}]^2 \right\|_1$$

$$= \left\| k_n^{-1} \sum_{i \le k_n} \underbrace{\mathbb{E}[\Delta_{z_i}(B_n) - \Delta_{z_i}(\infty) \mid \mathcal{F}_{z_i}]}_{\text{small for many } i} \cdot \underbrace{\mathbb{E}[\Delta_{z_i}(B_n) + \Delta_{z_i}(\infty) \mid \mathcal{F}_{z_i}]}_{\text{"well behaved" due to moment condition}} \right\|_1$$

Moreover, using pointwise ergodic thm, we have

$$k_n^{-1} \sum_{i \le k_n} \mathbb{E}[\Delta_{z_i}(\infty) \mid \mathcal{F}_{z_i}]^2 \longrightarrow \mathbb{E}\left[\mathbb{E}[\Delta_0(\infty) \mid \mathcal{F}_0]^2\right] = \lambda\sigma^2$$

$$\Longrightarrow \quad k_n^{-1} \sum_{i \le k_n} D_{ni}^2 \longrightarrow \lambda\sigma^2 \quad \text{in } L^1(\mathbb{P}).$$

∎

Remark: Need to show $\lambda\sigma^2 > 0$ given the stated condition.

EXAMPLES: k-nearest neighbors graph, Voronoi graph, ...
persistent Betti numbers

→ challenging part: Given H verify, the strong stabilization!

→ problem: percolation in the underlying graph (or simplicial complex) → changing the local point configuration can have long-range effects.

Remark: seminal papers on asymptotic normality and stabilization

- first ideas of stabilization in Lee (1997, 1999)
- quantitative normal approximation (rates of convergence to $\mathcal{N}$)
  Chatterjee (2008), Lachièze-Rey and Peccati (2016)
- Score functionals $H(\mathcal{P}_n) = \sum_{x \in \mathcal{P}_n} \xi(x, \mathcal{P}_n)$
  Lachièze-Rey et al. (2019), Yogeshwaran et al. (2019).

PERSISTENT BETTI NUMBERS AND CLT

- Setting as above: $(B_n)_n$, $(\mathcal{P}_n)_n$, $(\mathcal{U}_{nn})_n$:
- **Filtration**: $\mathbb{X} \subseteq \mathbb{R}^d$ finite, $r>0$

(a) **Čech**: $\mathcal{C}_r(\mathbb{X}) = \{\sigma \subseteq \mathbb{X} \mid \bigcap_{x\in\sigma} B(x,r) \neq \emptyset\}$

(b) **Victoris-Rips**: $VR_r(\mathbb{X}) = \{\sigma \subseteq \mathbb{X} \mid \operatorname{diam}(\sigma) \leq r\}$

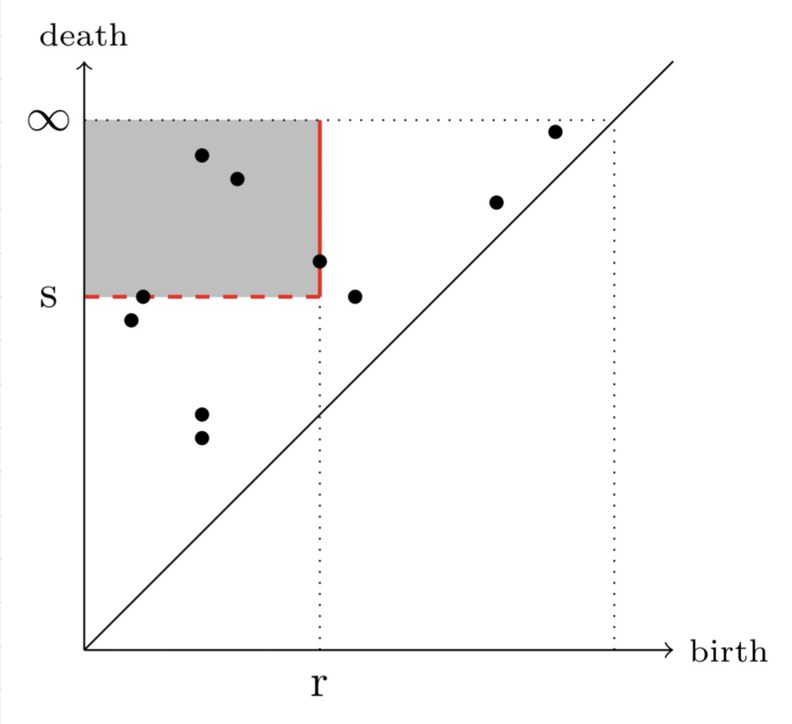
in the following $K_r(\mathbb{X})$ for both. $K(\mathbb{X}) := (K_r(\mathbb{X}) : r>0)$

- **persistent Beti numbers**:

$0 < r \leq s < \infty$, $0 \leq q \leq d-1$

$$\beta_q^{r,s}(K) = \dim(Z_q(K_r)) - \dim(B_q(K_s) \cap Z_q(K_r))$$

$(\beta_q^{r,s})_{r,s} \longleftrightarrow$ persistence diagram



11

- **Contributions:** Yogeshwaran et al. (2017, PTRF), Hiraoka et al. (2018, AOAP), K.D. Trinh (2019, ECP), K. and Polonik (2019). "CLTs with restrictions to parameter (r, s)" "Strong stabilization"
- **State of the art:** multivariate asymptotic normality in critical regime

THEOREM: Let $\ell \in \mathbb{N}$ and $(r_1, s_1), \dots, (r_\ell, s_\ell) \in \Delta$. Let k be continuous density on $[0,1]^d$. Let $\mathcal{P}_n$ Poisson process on $[0,1]^d$ w/ intensity $n \cdot k$. Then for all $q \in \{0, \dots, d-1\}$

$$\begin{pmatrix} \beta_q^{r_1, s_1}(K(n^{1/d}\mathcal{P}_n)) - \mathbb{E}[\beta_q^{r_1,s_1}(K(n^{1/d}\mathcal{P}_n))] \\ \vdots \\ \beta_q^{r_\ell, s_\ell}(K(n^{1/d}\mathcal{P}_n)) - \mathbb{E}[\beta_q^{r_\ell,s_\ell}(K(n^{1/d}\mathcal{P}_n))] \end{pmatrix} \Longrightarrow \mathcal{N}(0, \Sigma)$$

Σ nondegenerate covariance matrix.

Remark: Similar statement for binomial process.

APPLICATION: THE SMOOTH BOOTSTRAP

- Aim: Confidence interval for $\mathbb{E}[\beta_q^{r,s}(K(n^{1/d}\mathcal{P}_n))]$
- Idea: Use asymptotic normality to obtain approximate confidence interval:

$$\beta_q^{r,s}(K(n^{1/d}\mathcal{P}_n)) \pm q_{n,\alpha}$$

- Problem: $q_{n,\alpha}$ cannot be computed in practice!
- Solution: "Smooth Bootstrap" Roycraft, K., Polonik, 2021
  1. Estimate κ and draw $\mathcal{P}^*_{n,1}, \dots, \mathcal{P}^*_{n,B}$ from $\hat{\kappa}$
  2. Compute for $j = 1, \dots, B$

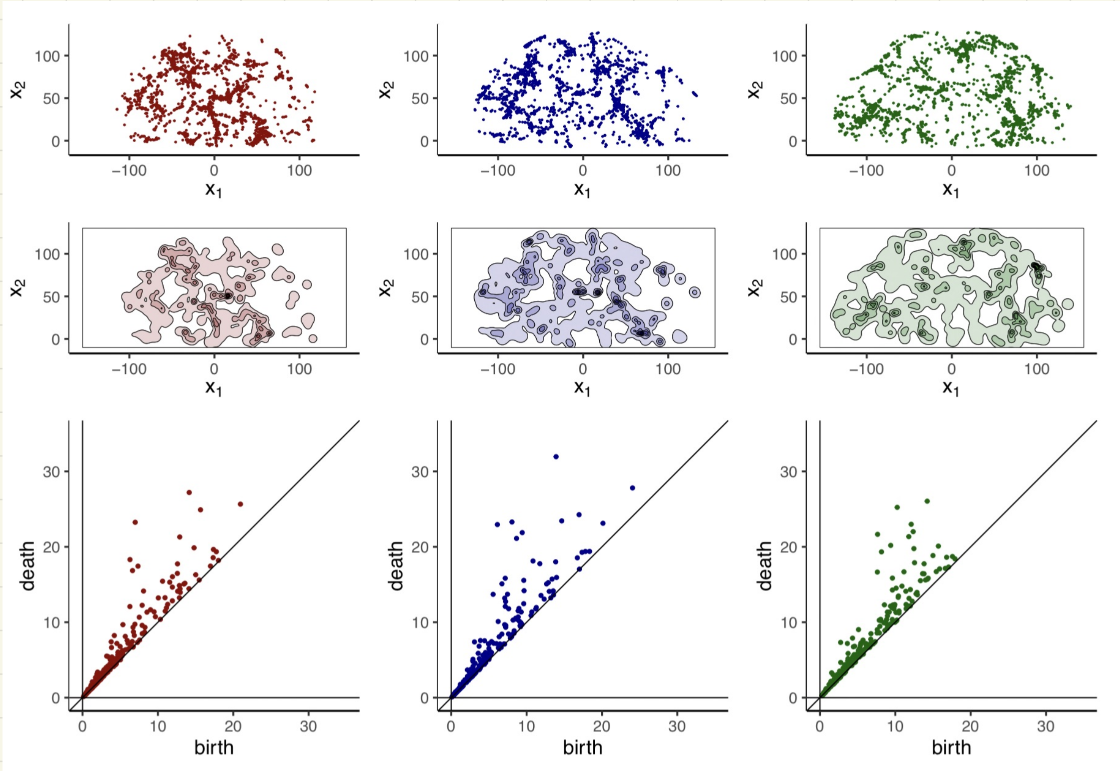
$$Z_j = \beta_q^{r,s}(K(n^{1/d}\mathcal{P}^*_{n,j})) - \mathbb{E}[\beta_q^{r,s}(K(n^{1/d}\mathcal{P}^*_{n,j}))]$$

  3. Approximate $q_{n,\alpha}$ by the distribution of $\{Z_j\}_{j=1}^{B}$: $q^*_{n,\alpha}$
  4. $[\beta_q^{r,s}(K(n^{1/d}\mathcal{P}_n)) - q^*_{n,\alpha},\ \beta_q^{r,s}(K(n^{1/d}\mathcal{P}_n)) + q^*_{n,\alpha}]$.

BETTI NO. FROM GALAXIES OF SDSS-IV

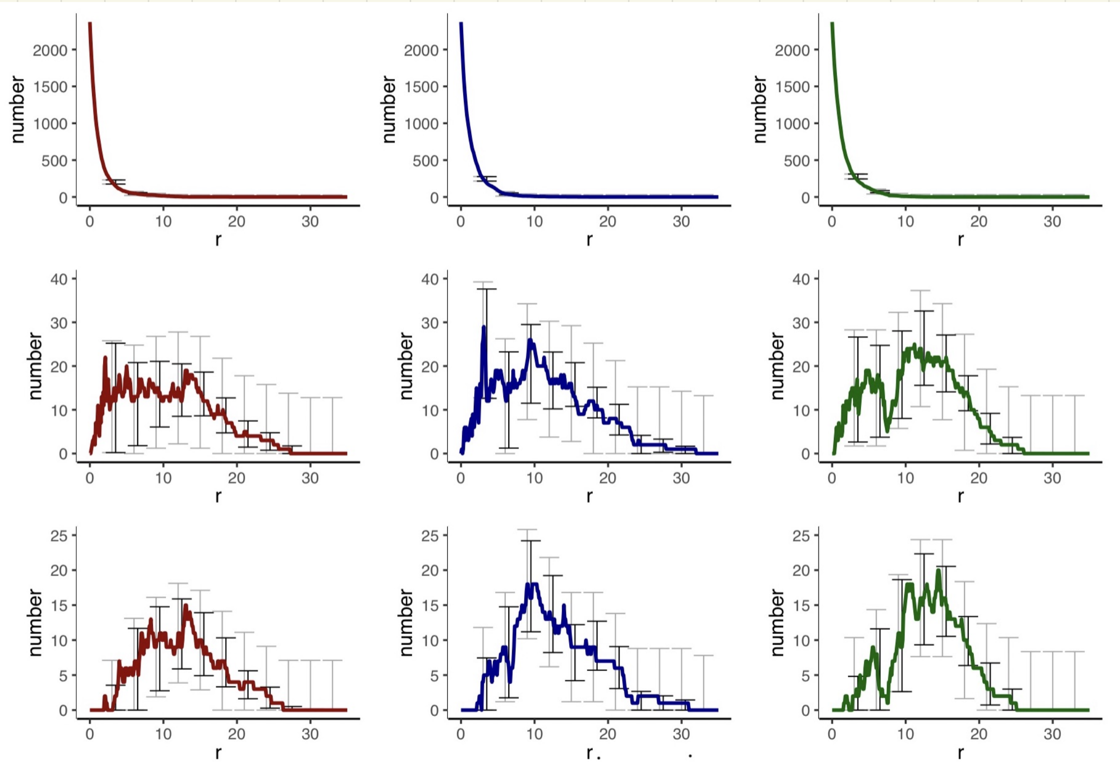
13

- Three samples of galaxies.
- Split according to red-shift.



Gibbs

R.N.-de
wrt Poi



QUESTION:

- Statistically different pattern in each sample?
- Analyze PDs for $q = 0, 1$ w/ VR-filtration
- Compare Betti-curves using smooth bootstrap.