Limit Theorems in Stochastic Geometry with Applications

Mathew Penrose
(University of Bath)

Workshop on Geometry and Stochastics of
Nonlinear, Functional and Graph Data
Bornholm, August 2016

Locally determined functionals

Let $d \in \mathbf{N}$. Suppose $\xi(x, F) \in \mathbf{R}$ is defined for $F \subset \mathbf{R}^d$ finite, $x \in F$, with $\xi(x, F)$ determined either by $F \cap B_1(x)$ [here $B_r(x)$ is a ball], or by $F \cap B_{N_k(x,F)}(x)$, with $N_k(x, F)$ the k-nearest neighbour dist., k fixed.

Examples include $\xi(x, F) = N_1(x, F)$, or [with $G(F, r)$ a geometric graph] $\xi(x, F) =$ the number of triangles in $G(F, 1)$ that include x.
(Our methods apply to other ξ...)

Interested in limit theorems (LLN, CLT) for $\sum_{x \in F_n} \xi_n(x, F_n)$ for empirical pt. processes F_n (sample of size n from some density),

where $\xi_n(x, F) = \xi(n^{1/d}x, n^{1/d}F)$, assuming translation invariance.

Some point processes in $\mathbf{R}^d$

Let $X_1, X_2, \dots$ be independent random d-vectors

with common density f in $\mathbf{R}^d$ with support $\mathcal{K} \subseteq \mathbf{R}^d$ (e.g. $\mathcal{K} = [0,1]^d$).

Let $F_n := \{X_1, \dots, X_n\}$.

For $a > 0$, let $\mathcal{H}_a$ be a homogeneous Poisson process in $\mathbf{R}^d$ with intensity a.

Main interest is in $\sum_{i=1}^n \xi_n(X_i, F_n)$

Laws of Large Numbers (P.-Yukich 2003, P. 2007a)

Let $\varepsilon > 0$. If $\sup_n E[|\xi_n(X_1, F_n)|^{1+\varepsilon}] < \infty$, then

$n^{-1} \sum_{i=1}^n \xi_n(X_i, F_n) \to \int E\xi(0, \mathcal{H}_{f(x)}) f(x) dx$ in L^1,

Idea of proof. Locally $n^{1/d}(-X_i + F_n)$ resembles $\mathcal{H}_{f(X_i)}$.

Can improve to L^2 convergence under $2 + \varepsilon$ moments condition.

Can improve to a.s. convergence under stronger moments and smoothness.

If ξ is *homogeneous*, i.e. $\xi(ax, aF) = a^\beta \xi(x, F)\ \forall x, F$ (some β), then

RHS simplifies to $E\xi(0, \mathcal{H}_1) I_{1-\beta/d}(f)$ [where $I_\alpha(f) = \int_{\mathcal{K}} f(x)^\alpha dx$.]

Example: Entropy estimators (see P.-Yukich, 2011, 2013)

Given $\rho \in (0,1) \cup (1,\infty)$, the *Renyi ρ-entropy of f* is computed in terms of $I_\rho(\alpha)$ (see Leonenko et al. *Ann. Stat.* 2008)

Put $\xi(x,F) = N_1(x,F)^\alpha$. Assuming moment condition, preceding LLN gives [with π_d = vol. of unit ball in $\mathbf{R}^d$]:

$$n^{-1} \sum_{i=1}^{n} (n^{1/d} N_1(X_i, F_n))^\alpha \to \pi_d^{-\alpha/d} \Gamma(1 + \frac{\alpha}{d}) I_{1-\alpha/d}(f) \quad \text{in } L^1$$

providing a consistent estimator for $(1-\alpha/d)$-entropy of (unknown) f.

Put $\xi(x,F) = \log(\pi_d N_1(x,F)^d)$. Can show $E\xi(0,\mathcal{H}_a) = -\gamma - \log a$ (Euler const.) so given the moment condition,

$$n^{-1} \sum_i \log(n\pi_d N_1(X_i,F_n)^d) \to I_0(f) - \gamma \quad \text{in } L^1$$

with $I_0(f) = -\int f \log f$ the Shannon entropy of f.

When do the moments conditions hold in the preceding examples?

A sufficient condition for the $(1+\varepsilon)$ moments condition [and hence L^1 LLN] for $\xi(x, F) = N_1(x, F)^\alpha$ is any of

• $\alpha > 0$ and $\mathcal{K}$ a finite union of convex compact sets with f bounded away from 0 and ∞ on $\mathcal{K}$.

• $-d < \alpha < 0$ and f bounded

• $0 < \alpha < d$ and $I_{1-\alpha/d}(f) < \infty$ and $E[|X_1|^r] < \infty$, some $r > d/(d-\alpha)$.

Sufficient for the L^2 LLN for $\xi(x, F) = \log N_1(x, F)$ is either

• f and $\mathcal{K}$ both bounded, or

• $E[|X_1|^r] < \infty$, some $r > 0$.

Extending the general theory to manifolds (P. and Yukich 2013)

Now suppose the points X_i lie on an m-dimensional submanifold $\mathcal{M}$ of $\mathbf{R}^d$ with $m \le d$. Each $x \in \mathcal{M}$ has a neighbourhood $g(U)$, some open $U \subset \mathbf{R}^m$ and smooth $g : U \to \mathcal{M}$. Integration over $\mathcal{M}$ is defined locally on $g(U)$ by

$$\int_{g(U)} h(x)dx = \int_U h(g(x))D_g(x)dx$$

with D_g a Jacobian. Now f is the density on $\mathcal{M}$, so

$$P[X_i \in A] = \int_A f(x)dx, \qquad A \subseteq \mathcal{M}.$$

Given ξ, set $\xi_n(x,F) = \xi(n^{1/m}x, n^{1/m}F)$, and let $\mathcal{H}_a$ be a homogeneous Poisson process in $\mathbf{R}^m$ (embedded in $\mathbf{R}^d$).

Law of large numbers in manifolds

The general LLN carries through to manifolds if ξ is (i) translation *and rotation* invariant and (ii) continuous, in the sense that $\forall k \in \mathbf{N}$, Lebesgue-almost all $(x_1, \ldots, x_k) \in (\mathbf{R}^m)^k$ lie at a continuity point of the mapping on $\mathbf{R}^{mk} \to \mathbf{R}$ given by

$$(x_1, \ldots, x_k) \mapsto \xi(0, \{x_1, \ldots, x_k\}).$$

The result says that under a $(1+\varepsilon)$-moment condition we have

$$n^{-1} \sum_{i=1}^{n} (\xi_n(X_i, F_n)) \to \int_{\mathcal{M}} E[\xi(0, \mathcal{H}_{f(y)})] f(y) dy$$

The idea is similar to before: the rescaled point process $n^{1/m}(-X_i + F_n)$ approximates to $\mathcal{H}_{f(X_i)}$ after rotation. There is an extension the non-RI case.

The Levina-Bickel dimension estimator

Want to estimate m from data in $\mathbf{R}^d$. Let $k \in \mathbb{N}$. Consider

$$\zeta(x,F) = (k-2)\left(\sum_{j=1}^{k-1} \log \frac{N_k(x,F)}{N_j(x,F)}\right)^{-1}$$

This is homogeneous of order 0, ie $\zeta(ax,aF) = \zeta(x,F)$. Also $\{(N_j(0,\mathcal{H}_a)/N_k(0,\mathcal{H}_a))^m\}_{j=1}^{k-1}$ are a sample from the $U(0,1)$ distribution so

$$\mathbb{E}\zeta(0,\mathcal{H}_a) = (k-2)m\mathbb{E}[(\sum_{j=1}^{k-1}\log(U_j^{-1}))^{-1}] = m$$

where U_j are independent $U(0,1)$.

Consistency of Levina-Bickel (P. and Yukich, 2013)

Suppose $\mathcal{K}$ is a compact m-dim. submanifold-with-boundary of $\mathcal{M}$, and f is bounded away from 0 and ∞ on $\mathcal{K}$, and $k \geq 11$. Recall $\zeta(x, F) = (k-2)/\sum_{j=1}^{k-1} \log \frac{N_k(x,F)}{N_j(x,F)}$. Then a.s.

$$\lim_{n\to\infty} n^{-1} \sum_{i=1}^{n} \zeta(X_i, F_n) = m$$

Moments condition might fail! If $m = 1, d = 3$ and $\mathcal{M}$ includes part of z-axis and part of unit circle in (x, y)-plane, then $P[\zeta(X_1, F_n) = \infty] > 0$.

Consistency result proved via truncation.

Central Limit theorem in flat space (P., 2007b)

Under a $(2+\varepsilon)$-moment condition on $\xi_n(x, F_n)$ and $\xi_n(x, F_n \cup \{y\})$, $x, y \in \mathcal{K}$ and similar moment conditions for F_{M_λ} (M_λ an indep. Poisson (λ) variable with $\lambda \sim n$)

$n^{-1} \mathrm{Var} \sum_{i=1}^{n} \xi_n(X_i, F_n) \to \int V^\xi(f(x)) f(x) dx - (\int \delta^\xi(f(x)) f(x) dx)^2$

$$V^\xi(a) = E\xi(0, \mathcal{H}_a)^2 + a \int ([E\xi(0, \mathcal{H}_a^u)\xi(u, \mathcal{H}_a^0) - (E\xi(0, \mathcal{H}_a))^2]) du$$

$$\delta^\xi(a) = E\xi(0, \mathcal{H}_a) + a \int E[\xi(0, \mathcal{H}_a^u - \xi(0, \mathcal{H}_a)] du$$

where $\mathcal{H}_a^u = \mathcal{H}_a \cup \{u\}$. Also we have an associated CLT. Moreover, we have similar results in manifolds!

Sketch proof of variance convergence and CLT

(i) Poissonize. First consider $\sum_{i=1}^{M_n} \xi_n(X_i, F_{M_n})$.

(ii) Mecke-type moment formulae. Express variance in terms of integrals.

(iii) Variance limit. Rescaled point process near x locally resembles $\mathcal{H}_{f(x)}$.

(iv) CLT via local spatial dependency (Stein's method).

(v) De-Poissonize: approximate linearity of $\sum_{i=1}^{m} \xi_n(X_i, F_m)$ w.r.t. $o(n)$ added/removed points.

Examples where the general CLT applies

Assume f bounded away from 0 and ∞ on $\mathcal{K}$ and $\mathcal{K}$ is compact convex (in $\mathbf{R}^m$) or a compact submanifold-with-boundary of $\mathcal{M}$ (eg if $\mathcal{M}$ is a sphere and $\mathcal{K} = \mathcal{M}$). Then the general CLT applies if

e.g. $\xi(x, F) = h(N_1(x, F))$ with h bounded

e.g. $\xi(x, F) = N_1(x, F)^\alpha$ with $\alpha > 0$.

e.g. $\xi(x, F) =$ number of triangles in $G(F, 1)$ including x.

e.g. $\xi_n(x, F) = \zeta(x, F)\mathbf{1}\{N_1(x, F) \leq \rho\}$, for some fixed $\rho > 0$, depending on $\mathcal{M}$. Can get a CLT for the modified Levina-Bickel statistic which ignores terms with $N_1(x, F) > \rho$.

Examples where the moment condition fails

The $(2+\varepsilon)$ moment condition for $\xi_n(x, F_n \cup \{y\})$ fails e.g. when

$$\xi(x,F) = N_1(x,F)^\alpha, \qquad -m/2 < \alpha < 0$$

$$\xi(x,F) = \log N_1(x,F),$$

Nevertheless, can obtain CLTs for these examples, using truncation $\xi^\varepsilon = N_1^\alpha(x,F)\mathbf{1}_{\{N_1(x,F)>\varepsilon\}}$, and Efron-Stein inequality to control $\mathrm{Var} \sum_i (\xi_n - \xi_n^\varepsilon)(X_i, F_n)$.

Efron-Stein bounds this variance in terms of the 'add one costs'.

Example: Spacings, ϕ-divergence (Baryshnikov, P. and Yukich 2009)

Consider another density g with same support $\mathcal{K}$ as f. Let $\phi : \mathbf{R}^+ \to \mathbf{R}$ satisfy appropriate growth bounds on $|\phi|$ at 0 and ∞, e.g. $\phi(x) = -\log x$ (or $x \log x$ or x^r, $r > 0$). The ϕ-divergence of g from f is

$$\int_{\mathcal{K}} \phi(\frac{g(x)}{f(x)}) f(x) dx$$

and an empirical version (used in eg goodness of fit test) is given by

$$\sum_{i=1}^{n} \phi(n \int_{B_{N_1(X_i,F_n)}(X_i)} g(y)dy) \approx \sum_{i=1}^{n} \phi(n \pi_d N_1(X_i, F_n)^d g(x))$$

corresponding to (non translation invariant)

$$\xi(x, F) = \phi(g(x) \pi_d N_1(x, F)^d)$$

Assume f, g, bounded away from 0 and ∞ on convex compact support $\mathcal{K}$.

Similar methods to before, adapted to the non-TI invariant case by setting

$$\xi_n(x, F) = \xi(x, -x + n^{1/d}(-x + F)),$$

can be used to show that the empirical ϕ-divergence

$$\sum_{i=1}^{n} \phi(n\pi_d N_1(X_i, F_n)^d g(x))$$

converges to the $\hat{\phi}$-divergence

$$\int_{\mathcal{K}} \hat{\phi}(\frac{g(x)}{f(x)}) f(x) dx$$

where $\hat{\phi}(t) = E[\phi(te_1)]$ and e_1 is exponential with mean 1.

Associated CLTs are available.

References

- M.D. Penrose and J.E. Yukich (2003) Weak laws of large numbers in geometric probability. *Ann. Appl. Probab.* **13**, 277-303.
- M.D. Penrose (2007a) Laws of large numbers in stochastic geometry with statistical applications. *Bernoulli* **13**, 1124-1150.
- M.D. Penrose (2007b) Gaussian limits for random geometric measures. *Electronic Journal of Probability* **12** (paper 35), 989-1035.
- Y. Baryshnikov, M.D. Penrose and J.E. Yukich (2009). Gaussian limits for generalized spacings, *Ann. Appl. Probab.* **19**, 158-185.
- M.D. Penrose and J. E. Yukich (2011) Laws of large numbers and nearest neighbor distances. *Advances in Directional and Linear Statistics*, eds. M.T. Wells, and A. SenGupta, pp. 189–199 Physica-Verlag HD Berlin.
- M.D. Penrose and J.E. Yukich (2013) Limit theory for point processes in manifolds. *Annals of Applied Probability* **23**, 2161-2211.