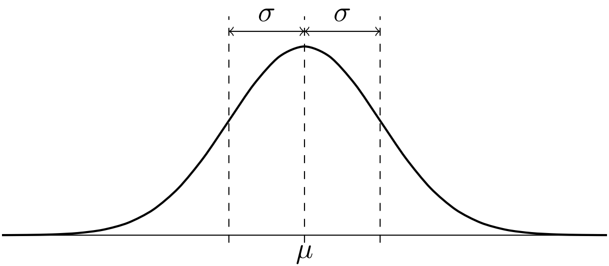
Central Limit Theorems in Stochastic Geometry

Martin Raič

Rogla

19 May 2013



the normal (Gaussian) distribution $\mathrm{N}(\mu, \sigma)$

$$\mathbb{P}(X \leq x) = \Phi\left(\frac{x-\mu}{\sigma}\right), \qquad \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-t^2/2}\, dt$$

CENTRAL LIMIT THEOREM (CLASSICAL)

Suppose that:

- $X_1, X_2, \ldots$ are independent and identically distributed random variables.
- $\mathbb{E}(X_i^2) < \infty$
- $S_n := X_1 + X_2 + \cdots + X_n$

Then the distribution of the normalized sum:

$$\frac{S_n - \mathbb{E}(S_n)}{\sqrt{\operatorname{var}(S_n)}}$$

converges weakly to the standard normal $\mathrm{N}(0, 1)$ as $n \to \infty$.

ESTIMATION OF THE ERROR

Without loss of generality, $\mathbb{E}(X_1) = 0$ and $\text{var}(X_1) = 1$, so that $\mathbb{E}(S_n) = 0$ and $\text{var}(S_n) = n$.

Uniform bound (Berry (1941), Esséen (1942), ...):

$$\sup_{z \in \mathbb{R}} \left| \mathbb{P}\left(\frac{S_n}{\sqrt{n}} \le z \right) - \Phi(z) \right| \le 0.6 \frac{\mathbb{E}\,|X_1|^3}{\sqrt{n}}$$

Large deviations (Cramér (1938), Petrov (1953), Statulevičius (1965)):
If $\mathbb{E}\, e^{HX_1} < \infty$ for some $H > 0$, then, for all $0 \le x \le C_0 \sqrt{n}$,

$$\frac{\mathbb{P}(S_n/\sqrt{n} \ge x)}{1 - \Phi(x)} = \exp\left(\frac{x^3}{\sqrt{n}} \lambda\left(\frac{x}{\sqrt{n}} \right) \right) \left(1 + \frac{C_1 \theta (1 + x)}{\sqrt{n}} \right)$$

for some $\theta \in [-1, 1]$.
C_0 and C_1 depend only on H and $\mathbb{E}\, e^{HX_1}$.

RELAXATION OF CONDITIONS

- The summands need not be identically distributed.
  The Berry–Esséen bound can be expressed in terms of $\sum_{i=1}^{n} \mathbb{E}\,|X_i|^3$.
  For large deviations, uniform boundedness of exponential moments satisfies, but can be relaxed.
- The summands may be indexed by an infinite, even random set I.
- The summands need not be independent.
  There are important extensions to *local dependence*.
  A *dependence graph* is a graph with vertex set I, such that for any disjoint $J, K \subseteq K$, such that there is no edge with one endpoint in I and the other in J, $\{X_j \,;\, j \in J\}$ and $\{X_k \,;\, k \in K\}$ are independent.

STABILIZING GEOMETRIC FUNCTIONALS

- A *geometric functional* is a measurable function ξ defined on pairs $(x, \mathcal{X})$, where $\mathcal{X} \subset \mathbb{R}^d$ is a finite set and $x \in \mathcal{X}$. For $x \notin \mathcal{X}$, set $\xi(x, \mathcal{X}) := \xi(x, \mathcal{X} \cup \{x\})$.
- ξ *stabilizes* at x with respect to $\mathcal{X}$ within radius R if $\xi(x, \mathcal{Y}) = \xi(x, \mathcal{X} \cap B_R(x))$ for all $\mathcal{Y}$ with $\mathcal{Y} \cap B_R(x) = \mathcal{X} \cap B_R(x)$. Here, $B_R(x)$ denotes the closed ball with radius R centered at x.
- We also consider *marked Euclidean spaces* $\mathbb{R}^d \times \mathcal{M}$.
- *Dependence graph* for stabilizing functionals: if ξ stabilizes at x_1 within radius R_1 and at x_2 within radius R_2, x_1 and x_2 make up an edge if $\|x_1 - x_2\| \leq R_1 + R_2$.

EXAMPLES OF STABILIZING FUNCTIONALS (1)

- ***k*-nearest-neighbor graph**: the directed edge from x to y is present if y is among the k nearest neighbors of x. Take $\xi(x, \mathcal{X}) := \sum_y f(x, y)$, where the sum runs over all y which are adjacent to x. We can consider adjacency in either direction.
- **Voronoi tesselations**: the *Voronoi cell* $V(x, \mathcal{X})$ at x is the set of points which are closer to x than to any other point in $\mathcal{X}$. The *Delaunay graph* on $\mathcal{X}$ is a graph with vertex set $\mathcal{X}$, such that two points are adjacent if the intersection of the underlying Voronoi cells is not negligible. There are many other related graphs.

EXAMPLES OF STABILIZING FUNCTIONALS (2)

- **Packing**: $\breve{\mathcal{X}} \subset \mathbb{R}^d \times [0,1]$. For $\breve{x} = (x,t) \in \breve{\mathcal{X}}$, t is the *time stamp*. Assuming that all time stamps are distinct, order $\breve{\mathcal{X}}$ accordingly: $\breve{x}_1 < \breve{x}_2 < \cdots < \breve{x}_n$. Define a new set $A := A(\breve{x}, \breve{\mathcal{X}})$ as follows: take $x_1 \in A$. Inductively, take $x_k \in A$ if $B_r(x_k) \cap \{x_1, \ldots x_{k-1}\} \cap A = \emptyset$. Then one might consider $\xi(\breve{x}, \breve{\mathcal{X}}) := \mathbf{1}(x \in A(\breve{x}, \breve{\mathcal{X}}))$ or related functionals.

CLT FOR STABILIZING FUNCTIONALS

Denote by $\mathcal{P}_g$ a Poisson point process on $\mathbb{R}^d$ with intensity function g. For simplicity, omit marks.
Take a probability density function $\kappa\colon \mathbb{R}^d \to [0, \infty)$ and another function $f\colon \mathbb{R}^d \to \mathbb{R}$.

Consider $S_\lambda := \sum_{x \in \mathcal{P}_{\lambda\kappa}} f(x)\, \xi\big(\lambda^{1/d} x, \lambda^{1/d} \mathcal{X}\big)$.

Under suitable conditions, the distribution of $\dfrac{S_\lambda}{\sigma\sqrt{\lambda}}$ converges to $N(0, 1)$ for some $\sigma > 0$ as $\lambda \to \infty$.

- Penrose, Yukich, Baryshnikov (2001, 2002, 2005): convergence
- Penrose, Yukich (2005, 2007): uniform bounds
- Baryshnikov, Eichelsbacher, Schreiber, Yukich (2008); Eichelsbacher, Raič, Schreiber (2013): large deviation results

IDEA OF PROOF (1)

Method of moments: if $\mathbb{E}(X^k) = \mathbb{E}(Y^k)$ for all $k \in \mathbb{N}$, then X and Y have the same distribution.

Instead of moments $m_k = \mathbb{E}(X^k)$, one can consider cumulants c_k.

Moment generating function: $\mathbb{E}(e^{tX}) = \sum_{k=0}^{\infty} m_k \frac{t^k}{k!}$

Moment generating function: $\log \mathbb{E}(e^{tX}) = \sum_{k=0}^{\infty} c_k \frac{t^k}{k!}$

For $X \sim \mathrm{N}(\mu, \sigma)$, we have $\mathbb{E}(e^{tX}) = \exp\left(\mu t + \frac{\sigma^2 t^2}{2}\right)$, so that $c_1 = \mu$, $c_2 = \sigma^2$ and $c_k = 0$ for $k \geq 3$.

Rudzkis, Saulis, Statulevičius (1979): bounds on the cumulants $\longrightarrow$ bounds on the relative error in the normal approximation

IDEA OF PROOF (2)

The connection between moments and cumulants can be expressed in terms of *Faà di Bruno*'s formula:

$$f\left(\sum_{k=0}^{\infty} a_k \frac{x^k}{k!}\right) = \sum_{k=0}^{\infty}\left(\sum_{L_1,\dots L_p} f^{(p)}(0)\, a_{|L_1|}\cdots a_{|L_p|}\right)\frac{x^k}{k!}$$

where the sum ranges over all unordered partitions of $\{1,\dots k\}$. That is,

$$c_k = (-1)^{p-1}(p-1)! \sum_{L_1,\dots L_p} m_{|L_1|}\cdots m_{|L_p|}$$

It turns out that the cumulants of sums $\sum_{i\in I} X_i$ can be expressed in terms of the covariances:

$$\mathbb{E}(X_{i_1},\dots X_{i_k} X_{j_1},\dots X_{j_l}) - \mathbb{E}(X_{i_1},\dots X_{i_k})\,\mathbb{E}(X_{j_1},\dots X_{j_l})$$

which vanish if $(X_{i_1},\dots X_{i_k})$ and $(X_{j_1},\dots X_{j_l})$ are independent.

THANK YOU FOR YOUR ATTENTION!