Stein's method and Poisson process approximation

Aihua Xia
Department of Mathematics and Statistics
The University of Melbourne, VIC 3010
Australia
and
Department of Statistics and Applied Probability
National University of Singapore
Singapore 117546

6 August, 2003

Lecture 1. Poisson point processes

Poisson processes on the real line

Assume at time $t = 0$ we start recording the random times of occurence $0 \leq T_1 \leq T_2 \leq \cdots \leq T_n \leq \cdots$ of a sequence of events, e.g. customers who join a queue or the sales of an item.

- $N_t = \#\{i : \ T_i \leq t\}$: the number of events that have occurred before and including time t

- $\{N_t\}_{t \geq 0}$ is called a *counting process*.

- $N_t - N_s$: the number of events occurred in time interval $(s, t]$ and is called the increment of the counting process over the interval $(s, t]$

Definition 1 The *stationary Poisson process* $\{N_t\}_{t \geq 0}$ is defined as a counting process which has

(P1) independent increments on disjoint time intervals and

(P2) for each $t \geq 0$, $N_t \sim \text{Pn}(\lambda t)$.

Equivalent ways to view a stationary Poisson process:

Proposition 2 A counting process $\{N_t : t \geq 0\}$ is a Poisson process iff

(a) $N_0 = 0$,
(b) the process has stationary, independent increments,
(c) $\mathbf{P}(N_t \geq 2) = o(t)$ as $t \to 0$,
(d) $\mathbf{P}(N_t = 1) = \lambda t + o(t)$ as $t \to 0$, λ is a constant.

Define $\tau_1 = T_1$ and $\tau_i = T_i - T_{i-1}$ for $i \geq 2$, the following proposition can be shown easily.

Proposition 3 A counting process $\{N_t\}_{t\geq 0}$ is a Poisson process with rate λ iff $\{\tau_i,\ i = 1, 2, \cdots\}$ are independent and identically distributed with mean $1/\lambda$.

This observation is heavily dependent on the structure of the real line.

Proposition 4 A counting process $\{N_t\}_{t\geq 0}$ is a stationary Poisson process with rate λ iff

(a) for each fixed t, N_t follows Poisson distribution with parameter λt;

(b) given that $N_t = n$, the n arrival times $T_1,\ T_2,\ \ldots,\ T_n$ have the same joint distribution as the order statistics corresponding to n independent uniform $[0, t]$ random variables.

Remark 5 The arrival times are naturally ordered; $T_1 < T_2 < \ldots < T_n$. Thus, given that n arrivals occurred in $[0, t]$, the unordered arrival times (a random permutation of $T_1, T_2, \ldots, T_n$) look like n independent uniform $[0, t]$ random variables.

Poisson point processes

• Γ: a locally compact complete separable metric space. Such a space is necessarily $\sigma-$compact

• $\mathcal{B}$: the Borel algebra

• $\mathcal{B}_b$: the ring consisting of all *bounded* (note that a set is bounded if its closure is compact) Borel sets

• A measure μ on $(\Gamma, \mathcal{B})$ is called locally finite if $\mu(B) < \infty$ for all $B \in \mathcal{B}_b$.

• A locally finite measure ξ is called a point measure if $\mu(B) \in \mathbf{N} := \{0, 1, 2, \cdots\}$ for all $B \in \mathcal{B}_b$

• δ_x: the Dirac measure at $x \in \Gamma$, namely, $\delta_x(B) = \begin{cases} 1 & \text{if } x \in B \\ 0 & \text{if } x \notin B \end{cases}$.

• Since Γ is $\sigma-$compact, it is possible to write a point measure ξ as

$$\xi = \sum_{i=1}^{\infty} \delta_{x_i}$$

with $\#\{i : \ x_i \in B\} < \infty$ for every $B \in \mathcal{B}_b$

• $\mathcal{H}$: the space of all point measures on $(\Gamma, \mathcal{B}_b)$

• $|\xi|$ or $\xi(\Gamma)$: the total measure (the number of points) of ξ

• we say that ξ_n converges to $\xi \in \mathcal{H}$ *vaguely* if $\int f d\xi_n \to \int f d\xi$ for all continuous functions on Γ with compact support.

Proposition 6 *The following statements are equivalent:*

(i) ξ_n *converges to* ξ *vaguely;*

(ii) $\xi_n(B) \to \xi(B)$ *for all Borel set* B *such that its boundary* ∂B *satisfies* $\xi(\partial B) = 0$*;*

(iii) $\limsup_{n\to\infty} \xi_n(F) \le \xi(F)$ *and* $\liminf_{n\to\infty} \xi_n(G) \ge \xi(G)$ *for all closed* $F \in \mathcal{B}$ *and open* $G \in \mathcal{B}$*.*

Proposition 7 *The space* $\mathcal{H}$ *is Polish in the vague topology.*

• $\mathcal{B}(\mathcal{H})$: the Borel σ-algebra generated by the vague topology.

Definition 8 [Kallenberg (1976), p. 5] A point process Ξ is a measurable mapping from a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ to $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$. The measure $\boldsymbol{\lambda}$ defined by $\boldsymbol{\lambda}(B) = \mathbf{E}\Xi(B)$, $B \in \mathcal{B}_b$ is called the mean measure of Ξ.

Definition 9 A point process Ξ with locally finite mean measure $\boldsymbol{\lambda}$ is a Poisson point process, denoted as $\mathrm{Pn}(\boldsymbol{\lambda})$, if

(PP1) for any bounded Borel set B, $\Xi(B)$ is Poisson distributed with mean $\boldsymbol{\lambda}(B)$;

(PP2) for any $k \in \mathbf{N}$, $B_1, \cdots, B_k$, $\Xi(B_1), \cdots, \Xi(B_k)$ are independent random variables.

The following Proposition is a generalization of Proposition 4.

Proposition 10 A point process Ξ is a Poisson process with mean measure $\boldsymbol{\lambda}$ iff for each bounded Borel set B, $\Xi(B) \sim \mathrm{Pn}(\boldsymbol{\lambda}(B))$ and

(PP3) given $\Xi(B) = k$, Ξ restricted to B, $\Xi|_B$, has the same distribution as $X_k = \sum_{i=1}^k \delta_{\theta_i}$, where θ_i, $1 \le i \le k$ are independent and identically distributed B-valued random elements with the common distribution $\boldsymbol{\lambda}(\cdot)/\boldsymbol{\lambda}(B)$.

Proof. Assume Ξ is a Poisson point process, then for every bounded Borel set B and every partition of B consisting (bounded) Borel sets B_1, $\cdots$, B_m, and $i_1 + \cdots + i_m = k$,

$$\begin{aligned}
&\mathbf{P}\left(\cap_{j=1}^m \left\{\Xi(B_j) = i_j\right\} \middle| \Xi(B) = k\right) \\
&= \frac{\mathbf{P}\left(\bigcap_{j=1}^m \left\{\Xi(B_j) = i_j\right\}\right)}{\mathbf{P}(\Xi(B) = k)} \\
&= \frac{k!}{\prod_{j=1}^m i_j!} \cdot \frac{\prod_{j=1}^m \boldsymbol{\lambda}(B_j)^{i_j}}{\boldsymbol{\lambda}(B)^k} \\
&= \mathbf{P}\left(\cap_{j=1}^m \{X(B_j) = i_j\}\right).
\end{aligned}$$

Conversely, assume (PP3) holds, then for any $m \in \mathbf{N}$, bounded Borel sets B_1, $\cdots$, B_m, $B :=$

$\cup_{i=1}^m B_i$ is also bounded, and, with $k = i_1 + \cdots + i_m$, we have

$$\begin{aligned}
&\mathbf{P}\left(\cap_{j=1}^m \{\Xi(B_j) = i_j\}\right) \\
&= \mathbf{P}\left(\cap_{j=1}^m \{\Xi(B_j) = i_j\} \middle| \Xi(B) = k\right) \mathbf{P}(\Xi(B) = k) \\
&= \frac{k!}{i_1! \cdots i_m!} \cdot \frac{\boldsymbol{\lambda}(B_1)^{i_1} \cdots \boldsymbol{\lambda}(B_m)^{i_m}}{\boldsymbol{\lambda}(B)^k} \cdot \frac{e^{-\boldsymbol{\lambda}(B)} \boldsymbol{\lambda}(B)^k}{k!} \\
&= \prod_{j=1}^m \frac{e^{-\boldsymbol{\lambda}(B_j)} \boldsymbol{\lambda}(B_j)^{i_j}}{i_j!}. \blacksquare
\end{aligned}$$

• To construct a Poisson point process on Γ, since Γ is $\sigma-$compact, one can partition the space Γ into at most countably many bounded subsets Γ_i with $\boldsymbol{\lambda}(\Gamma_i) > 0$. For each i, one can then independently define a Poisson process using Proposition 10, then the union of these independent Poisson processes is the desired Poisson process [see Reiss (1993)].

When the carrier space is compact

• Γ is compact with a metric $d_0 \leq 1$

• We say a sequence $\{\xi_n\} \subset \mathcal{H}$ converges to $\xi \in \mathcal{H}$ *weakly* if $\int f d\xi_n \to \int f d\xi$ for all bounded continuous functions on Γ

• The vague topology is the same as weak topology

• Metrics to quantify the weak topology

• Wasserstein metric on $\mathcal{H}$ [Barbour and Brown (1992 A)]: $\mathcal{K}$ =the set of functions $k : \Gamma \to [-1, 1]$ such that

$$s_1(k) = \sup_{y_1 \neq y_2 \in \Gamma} \frac{| k(y_1) - k(y_2) |}{d_0(y_1, y_2)} \leq 1,$$

$$d_1(\xi_1, \xi_2) = \begin{cases} 1, & \text{if } |\xi_1| \neq |\xi_2|, \\ \frac{\sup_{k \in \mathcal{K}} \{ | \int k d\xi_1 - \int k d\xi_2 | \}}{m}, & \text{if } |\xi_1| = |\xi_2| = m > 0. \end{cases}$$

• By the Kantorovich-Rubinstein duality theorem, if $\xi_1 = \sum_{i=1}^m \delta_{y_i}$, $\xi_2 = \sum_{i=1}^m \delta_{z_i}$, then $d_1(\xi_1, \xi_2)$ can be interpreted as the average distance between a best coupling of the points of ξ_1 and ξ_2 :

$$d_1(\xi_1, \xi_2) = m^{-1} \min_\pi \sum_{i=1}^m d_0(y_i, z_{\pi(i)}),$$

where π ranges over all permutations of $(1, \cdots, m)$.

• The Prohorov distance: if $\xi_1 = \sum_{i=1}^n \delta_{y_i}$, $\xi_2 = \sum_{i=1}^m \delta_{z_i}$,

$$\rho_1(\xi_1, \xi_2) = \begin{cases} 1, & \text{if } n \neq m, \\ \inf_\pi \max_{1 \leq i \leq n} d_0(y_i, z_{\pi(i)}), & \text{if } n = m > 0, \end{cases}$$

The feature of maximum appeared here makes the Prohorov metric behave in a similar fashion to the total variation metric.

- d_1': if $\xi_1 = \sum_{i=1}^n \delta_{y_i}$, $\xi_2 = \sum_{i=1}^m \delta_{z_i}$ with $n \le m$,

$$d_1'(\xi_1, \xi_2) = (m-n) + \min_\pi \sum_{i=1}^n d_0(y_i, z_{\pi(i)}),$$

where π ranges over all permutations of $(1, \cdots, m)$.

Example 11 $\Gamma = [0,1]$ with metric $d(x,y) = |x-y|$. If $\xi_1 = \sum_{i=1}^n \delta_{t_i}$ with $0 \le t_1 \le \cdots \le t_n \le 1$ and $\xi_2 = \sum_{i=1}^n \delta_{s_i}$ with $0 \le s_1 \le \cdots \le s_n \le 1$, then

$$\sum_{i=1}^n |t_i - s_i| \le \sum_{i=1}^n |t_i - s_{\pi(i)}|$$

for all permutations π of $(1, \cdots, n)$, and hence

$$d_1(\xi_1, \xi_2) = \frac{1}{n} \sum_{i=1}^n |t_i - s_i|.$$

Proof Mathematical induction on n.

For $n = 2$, wlog, $t_1 = \min\{t_1, t_2, s_1, s_2\}$. Three cases.

(i) $t_2 \ge s_2$, then

$$\begin{aligned}&|t_1 - s_1| + |t_2 - s_2| = s_1 - t_1 + t_2 - s_2 \\ &\le |t_1 - s_2| + |t_2 - s_1|\end{aligned}$$

(ii) $s_1 \le t_2 < s_2$, then

$$\begin{aligned}&|t_1 - s_1| + |t_2 - s_2| = s_1 - t_1 + s_2 - t_2 \\ &\le |t_1 - s_2| + |t_2 - s_1|\end{aligned}$$

(iii) $t_2 < s_1$, then

$$|t_1 - s_1| + |t_2 - s_2| = s_1 - t_1 + s_2 - t_2$$
$$= |t_1 - s_2| + |t_2 - s_1|$$

Suppose claim holds for $n \leq k$ with $k \geq 2$, we shall prove it holds for $n = k+1$ and all permutations π of $(1, \cdots, k+1)$.

If $\pi(k+1) = k+1$, obvious.

Assume $\pi(k+1) \neq k+1$, then it follows that

$$\begin{aligned}
&\sum_{i=1}^{k+1} |t_i - s_{\pi(i)}| \\
&= \sum_{i \neq k+1, i \neq \pi^{-1}(k+1)} |t_i - s_{\pi(i)}| \\
&\quad + [|t_{k+1} - s_{\pi(k+1)}| + |t_{\pi^{-1}(k+1)} - s_{k+1}|] \\
&\geq \sum_{i \neq k+1, i \neq \pi^{-1}(k+1)} |t_i - s_{\pi(i)}| \\
&\quad + |t_{\pi^{-1}(k+1)} - s_{\pi(k+1)}| + |t_{k+1} - s_{k+1}| \\
&\geq \sum_{i=1}^{k} |t_i - s_i| + |t_{k+1} - s_{k+1}|. \blacksquare
\end{aligned}$$

Proposition 12 *The following statements are equivalent:*

(i) ξ_n converges to ξ vaguely;

(ii) $\xi_n(B) \to \xi(B)$ *for all Borel set* B *such that its boundary* ∂B *satisfies* $\xi(\partial B) = 0$*;*

(iii) $\limsup_{n\to\infty} \xi_n(F) \leq \xi(F)$ *and* $\liminf_{n\to\infty} \xi_n(G) \geq \xi(G)$ *for all closed* $F \in \mathcal{B}$ *and open* $G \in \mathcal{B}$*;*

(iv) ξ_n *converges to* ξ *weakly;*

(v) $d_1(\xi_n, \xi) \to 0$*.*

(vi) $\rho_1(\xi_n, \xi) \to 0$*.*

(vii) $d_1'(\xi_n, \xi) \to 0$*.*

Proof. The equivalence of (i)-(iv) is well-known.

(vi) is equivalent to (v): if $\xi(\Gamma) \neq 0$,

$$d_1(\xi_n, \xi) \leq \rho_1(\xi_n, \xi) \leq \xi(\Gamma) d_1(\xi_n, \xi).$$

(v) is equivalent of (vii): similarly.

(iv) is equivalent to (v): Assume first that $d_1(\xi_n, \xi) \to 0$, then there exists an n_0 such that $\xi_n(\Gamma) = \xi(\Gamma)$ for all $n \geq n_0$. For any bounded continuous function f, as Γ is compact, f is uniformly continuous and for $n \geq n_0$,

$$\left| \int f d\xi_n - \int f d\xi \right| \\ \leq |\xi| \sup_{d_0(x,y) \leq |\xi| d_1(\xi_n, \xi)} |f(x) - f(y)| \to 0$$

Conversely, choose $f \equiv 1$, we have $\xi_n(\Gamma) \to \xi(\Gamma) := m$, so when n is sufficiently large, say, $n \geq m_0$, $\xi_n(\Gamma) = \xi(\Gamma)$. Let $\xi = \sum_{i=1}^m \delta_{z_i}$, and relabel if necessary, for $n \geq m_0$, we may write $\xi_n = \sum_{i=1}^m \delta_{y_i^n}$ so that $d_1(\xi_n, \xi) = [\sum_{i=1}^m d_0(y_i^n, z_i)]/m$. Since Γ is compact, for any sequence $\{n_k\} \subset \mathbf{N}$, there exists a subsequence $\{n_{k_j}\}$ and $\tilde{z}_i$ such that $d_0(y_i^{n_{k_j}}, \tilde{z}_i) \to 0$ for all $i = 1, \cdots, m$. Set $\tilde{\xi} = \sum_{i=1}^m \delta_{\tilde{z}_i}$, then $\int f d\tilde{\xi} = \int f d\xi$ for all bounded function f, so $\tilde{\xi} = \xi$ and $\sum_{i=1}^m d_0(y_i^{n_{k_j}}, z_i) \leq \sum_{i=1}^m d_0(y_i^{n_{k_j}}, \tilde{z}_i) \to 0$. This implies $d_1(\xi_n, \xi) \to 0$. ∎

Proposition 13 *$\mathcal{H}$ is a locally compact separable metric space.*

Proof. By Proposition 7, it suffices to show $\mathcal{H}$ is locally compact. In fact, define $U_k = \{\xi \in \mathcal{H} : \ \xi(\Gamma) = k\}$, $\mathcal{H} = \cup_{k=0}^{\infty} U_k$ and for each k, U_k is a compact, open and as well as closed set. ∎

Characterization of Poisson point processes

- Palm theory

For a non-negative integer-valued random variable X, $X \sim \text{Pn}(\lambda)$ iff

$$\frac{\mathbf{E}g(X)X}{\lambda} = \mathbf{E}g(X+1) \ \forall \text{bounded } g \text{ on } \mathbf{Z}_+$$

The equation completely characterizes the Poisson distribution and is the starting point for Stein's identity.

• Palm theory was developed after the work of Palm (1943)

Heuristicly, since a Poisson process is a point process with independent increments and one dimensional distribution is Poisson, we may imagine that a Poisson process is pieced together by lots of independent "Poisson components" (if the location is an atom, the "component" will be Poisson, but if the location is diffuse, then the "component" is either 0 or 1). Hence, to specify a Poisson process, one needs to check that "each component" is Poisson and also independent of the rest. This can be intuitively done by

$$\frac{\mathbf{E}g(\Xi)\Xi(d\alpha)}{\mathbf{E}\Xi(d\alpha)} = \mathbf{E}g(\Xi + \delta_\alpha),$$

for all bounded function g on $\mathcal{H}$ and all $\alpha \in \Gamma$.

• To make the argument rigorous, one needs the tools of Campbell measures and Radon-Nikodym derivatives [Kallenberg (1976), p. 69].

• For each point process Ξ with locally finite mean measure $\boldsymbol{\lambda}$, we may define probability measures $\{P_\alpha,\ \alpha \in \Gamma\}$ on $\mathcal{B}(\mathcal{H})$ [Kallenberg (1976), p. 69]:

$$P_\alpha(\mathbf{B}) = \frac{\mathbf{E}\left[1_{[\Xi \in \mathbf{B}]}\Xi(d\alpha)\right]}{\boldsymbol{\lambda}(d\alpha)},\ \alpha \in \Gamma\ \boldsymbol{\lambda} - a.s.,\ \mathbf{B} \in \mathcal{B}(\mathcal{H}).$$

• A point process Ξ_α on Γ is called a Palm process of Ξ at location α if it has the Palm distribution P_α of Ξ at α.

• The Palm process of a Poisson process has the same distribution as the original process except one additional point is added

• $\Xi_\alpha - \delta_\alpha$: the reduced Palm process

• For any measurable function $f : \Gamma \times \mathcal{H} \to [0, \infty)$,

$$\mathbf{E}\left(\int_B f(\alpha, \Xi)\Xi(d\alpha)\right) = \mathbf{E}\left(\int_B f(\alpha, \Xi_\alpha)\boldsymbol{\lambda}(d\alpha)\right)$$

$$\mathbf{E}\left(\int_B f(\alpha, \Xi - \delta_\alpha)\Xi(d\alpha)\right) = \mathbf{E}\left(\int_B f(\alpha, \Xi_\alpha - \delta_\alpha)\boldsymbol{\lambda}(d\alpha)\right)$$

for all Borel set $B \subset \Gamma$.

• One can use Palm equation to establish the Stein identity for Poisson process approximation

Theorem 14 A point process Ξ with locally finite mean measure $\boldsymbol{\lambda}$ is a Poisson process iff its reduced Palm processes have the same distribution as that of the original process.

• Janossy densities

Suppose Ξ is a Poisson process on Γ with a finite mean measure $\boldsymbol{\lambda}$, let $\boldsymbol{\nu}(d\alpha) = \boldsymbol{\lambda}(d\alpha)/\lambda$, then by Proposition 10, for each bounded measurable function f on $\mathcal{H}$,

$$\mathbf{E}f(\Xi) = \sum_{n=0}^{\infty} \frac{e^{\lambda}\lambda^n}{n!} \int_{\Gamma^n} f\left(\sum_{j=1}^{n} \delta_{\alpha_j}\right) \boldsymbol{\nu}^n(d\alpha_1, \cdots, d\alpha_n).$$

General point processes: Janossy densities after its first introduction in Janossy (1950)

- A few assumptions:

(JA1) the point process Ξ is finite with distribution $R_n = \mathbf{P}(|\Xi| = n)$, $\sum_{n=0}^{\infty} R_n = 1$.

(JA2) Given the total number of points equals $n \geq 1$, there is a probability distribution $\Pi_n(\cdot)$ on Γ^n which determines the joint distribution of the positions of the points of the point process Ξ.

From these assumptions, we have

$$\begin{aligned}\mathbf{E}f(\Xi) &= \sum_{n=0}^{\infty} \mathbf{E}[f(\Xi)||\Xi| = n]R_n \\ &= \sum_{n=0}^{\infty} \int_{\Gamma^n} f\left(\sum_{j=1}^{n} \delta_{\alpha_j}\right) \Pi_n(d\alpha_1, \cdots, d\alpha_n)R_n.\end{aligned}$$

• $\Pi_n^s = \frac{1}{n!}\sum_\pi \Pi_n(d\alpha_{\pi(1)}, \cdots, d\alpha_{\pi(n)})$, where the sum ranges over all permutations π of $(1, \cdots, n)$, then

$$\begin{aligned} J_n(d\alpha_1, \cdots, d\alpha_n) &= R_n \sum_\pi \Pi_n(d\alpha_{\pi(1)}, \cdots, d\alpha_{\pi(n)}) \\ &= n! R_n \Pi_n^s \end{aligned}$$

• The measures $\{J_n\}$ are called Janossy measures.

• We can find a reference measure $\boldsymbol{\nu}$ such that J_n is absolutely continuous w.r.t. $\boldsymbol{\nu}^n$ and denote its Radon-Nikodym derivative as j_n.

• For any non-negative measurable function $f : \mathcal{H} \to \mathbf{R}$,

$$\mathbf{E}(f(\Xi)) = \\ \sum_{n\geq 0} \int_{\Gamma^n} \frac{1}{n!} f\left(\sum_{i=0}^{n} \delta_{\alpha_i}\right) j_n(\alpha_1, \cdot, \alpha_n) \boldsymbol{\nu}^n(d\alpha_1, \cdot, d\alpha_n)$$

• The density of the first moment measure $\boldsymbol{\lambda}$ of the point process Ξ with respect to $\boldsymbol{\nu}$:

$$\mu(\alpha) = \sum_{n\geq 0} \int_{\Gamma^n} \frac{1}{n!} j_{n+1}(\alpha, \alpha_1, \cdot, \alpha_n) \boldsymbol{\nu}^n(d\alpha_1, \cdot, d\alpha_n).$$

• The density of a point being at α, given the configuration Ξ^α of Ξ outside N_α: Let $m \in \mathbf{N}$ be fixed and $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_m) \in (N_\alpha^c)^m$, define

$$g(\alpha, \boldsymbol{\beta}) := \frac{\sum_{r\geq 0} \int_{N_\alpha^r} j_{m+r+1}(\alpha, \boldsymbol{\beta}, \boldsymbol{\gamma})(r!)^{-1} \boldsymbol{\nu}^r(d\boldsymbol{\gamma})}{\sum_{s\geq 0} \int_{N_\alpha^s} j_{m+s}(\boldsymbol{\beta}, \boldsymbol{\eta})(s!)^{-1} \boldsymbol{\nu}^s(d\boldsymbol{\eta})},$$

where the term with $r = 0$ is interpreted as $j_{m+1}(\alpha, \boldsymbol{\beta})$ and the term with $s = 0$ similarly. Then $g(\alpha, \boldsymbol{\beta})$ is the density of a point being near α given that Ξ^α is $\sum_{i=1}^m \delta_{\beta_i}$.

Theorem 15 *Ξ is a Poisson point process with finite mean measure $\boldsymbol{\lambda}$ iff, with respect to $\boldsymbol{\lambda}$, its Janossy densities $j_n \equiv e^{-\lambda}$.*

• Compensator

$\mathcal{F} = (\mathcal{F}_s)_{s\geq 0}$: right-continuous filtration

$(N_s)_{s\geq 0}$: an $\mathcal{F}-$adapted nonnegative integer-valued increasing right continuous process

A point process $(N, \mathcal{F})$ is called simple if $\Delta N_s := N_s - N_{s^-} = 0$ or 1, $\forall s \geq 0$, almost surely.

The compensator A of $(N, \mathcal{F})$ is the unique previsible right-continuous increasing process such that $N - A$ is an $(\mathcal{F})_{s\geq 0}$ local martingale [Jacod and Shiryaev (1987) or Dellacherie and Meyer (1982)]

• The compensator of a Poisson process with respect to its intrinsic filtration is a deterministic function, and in this case the Poisson process is simple iff its compensator is continuous.

• Any simple point process with continuous compensator is locally Poisson in character, in the sense that there exists a time transformation that converts the process into a Poisson process. More precisely, the transformation is given by the inverse of the compensator A of the simple point process $(N_t)_{t\geq 0}$:

$$\sigma(t) := \inf\{s : A_s > t\}.$$

The continuity of A ensures that $\sigma(t)$ is an $(\mathcal{F}_t)_{t\geq 0}$ stopping time, and if $\lim_{t\uparrow\infty} A_t = \infty$ almost surely, then the transformed process $\bar{N}_t := N_{\sigma(t)}$ is a Poisson process with unit rate with respect to filtration $(\bar{\mathcal{F}}_t)_{t\geq 0}$, where $\bar{\mathcal{F}}_t = \mathcal{F}_{\sigma(t)}$

• If M is a Poisson process with compensator $B_t = t, 0 \leq t < \infty$, then $(N, \bar{N})$ is a natural coupling of the distributions of M and N.

Immigration-death point processes

Immigration-death processes

• X_t: the number of objects in a system at time t

• Suppose that arrivals and departures occur independently of one another, but at rates depending on the state of the process. Thus, when there are n objects in the system, new arrivals enter at an exponential rate λ_n and the n objects leave at an exponential rate μ_n; that is, when there are n objects in the system the time to the next arrival (departure) is exponentially distributed with mean $1/\lambda_n$ ($1/\mu_n$).

• $\{X_t; t \ge 0\}$: an immigration-death process and the parameters $\{\lambda_n\}$ and $\{\mu_n\}$ are the arrival (or immigration) rates and departure (or death) rates.

It is a continuous time Markov chain with states 0, 1, 2, . . .

• If

$$\sum_{j=1}^{\infty} \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1\mu_2\cdots\mu_j} < \infty,$$

then the Markov chain is ergodic with equilibrium distribution

$$\pi_0 = \left(1 + \sum_{j=1}^{\infty} \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1\mu_2\cdots\mu_j}\right)^{-1}$$
$$\pi_j = \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1\mu_2\cdots\mu_j}\pi_0, \; j = 1, 2, \cdots$$

• If $\lambda_j = \lambda$ and $\mu_j = j$ for all $j \ge 0$, then the equilibrium distribution is a $\text{Pn}(\lambda)$.

The idea of using Markov immigration-death process to interpret Stein-Chen method was initiated by Barbour (1988).

From now on, all immigration-death processes will be assumed to be ergodic.

• Z_i is an immigration-death process with immigration rates $\{\lambda_n\}$ and death rates $\{\mu_n\}$ and started in state i

Lemma 16 *If $\lambda_n = \lambda$ and $\mu_n = n$ for all n, then*

$$Z_n(t) = D_n(t) + Z_0(t),$$

where $D_n(t)$ is a pure death process with unit per capita death rate and $D_n(t)$ and $Z_0(t)$ are independent.

Proof.

• [Barbour (1988)] For an immigration-death process, if the initial distribution is the equilibrium distribution, then the immigation-death process has equilibrium distribution for all time; if the process starts from a distribution close to equilibrium distribution, then it takes little time to stablize and if the initial distribution is far away from equilibrium distribuition, then it will need a lot of time to reach stationarity.

Spatial immigration-death processes

For Poisson process approximation, instead of integer-valued Markov immigration-death process, we need to run a spatial immigration-death process [Preston (1977)].

• Γ: compact; $\boldsymbol{\lambda}$: finite measure on Γ

There are infinitely many spatial immigration-death processes with equilibrium distribution Pn($\boldsymbol{\lambda}$). The one we use is as follows. Given the process takes a configuration $\xi \in \mathcal{H}$, the process stays in state ξ for an exponentially distributed random sojourn time with mean $1/(n+\lambda)$ where $\xi(\Gamma) = n$, then with probability $\lambda/(n+\lambda)$, it takes an immigration point and the point distributes on Γ according to $\boldsymbol{\lambda}/\lambda$, independent of the existing configuration; and with probability $n/(n+\lambda)$, a point is chosen equally likely from the existing configuration ξ and deleted from the system. This is equivalent to say that each point in ξ has an exponentially distributed lifetime with mean 1 and it is called unit per capita death rate.

Such a spatial immigration-death process, denoted as $\mathbf{Z}_\xi(t)$, can be defined by its generator on $\mathcal{H}$ as follows

$$\begin{aligned}\mathcal{A}h(\xi) &= \int_{\alpha\in\Gamma}[h(\xi+\delta_\alpha)-h(\xi)]\boldsymbol{\lambda}(d\alpha) \\ &\quad + \int_{\alpha\in\Gamma}[h(\xi-\delta_\alpha)-h(\xi)]d\xi(\alpha), \forall \xi \in \mathcal{H}\end{aligned}$$

for suitable function h on $\mathcal{H}$.

Strictly speaking, we need to prove that the spatial immigration-death process constructed is the **unique** spatial Markov process

Proposition 17 *The spatial immigration-death process has a unique equilibrium distribution Pn($\boldsymbol{\lambda}$) to which it converges in distribution from any initial state.*

Proof. Need to add a proof here. ■

Poisson process approximation by coupling

The metrics for point process approximation

For any two probability measures $\mathbf{Q}_1$ and $\mathbf{Q}_2$ on $(\mathcal{H}, \mathcal{B}(\mathcal{H}))$, the total variation distance, d_{TV}, between $\mathbf{Q}_1$ and $\mathbf{Q}_1$ is defined as

$$\begin{aligned} d_{TV}(\mathbf{Q}_1, \mathbf{Q}_1) &= \sup_{D \in \mathcal{B}(\mathcal{H})} |\mathbf{Q}_1(D) - \mathbf{Q}_1(D)| \\ &= \inf \mathbf{P}(X \neq Y), \end{aligned}$$

where the infimum is taken over all couplings of (X, Y) such that $X \sim \mathbf{Q}_1$ and $Y \sim \mathbf{Q}_2$. This equality follows from the duality theorem [Rachev (1991)].

However, although the total variation metric for random variable approximation is a natural choice, it is too strong to use for process approximation.

Example. Let I_i, $1 \leq i \leq n$ be independent indicators with $\mathbf{P}(I_i = 1) = 1 - \mathbf{P}(I_i = 0) = 1 - p_i$. Define the Bernoulli process $\Xi = \sum_{i=1}^n I_i \delta_{i/n}$, and set $\lambda = \sum_i p_i$, $\boldsymbol{\mu} = \sum_{i=1}^n p_i \delta_{i/n}$ and $\boldsymbol{\lambda}(dx) = np_i dx$, for $x \in ((i-1)/n, i/n]$. What is the total variation distance between the distribution of Ξ and Pn($\boldsymbol{\mu}$)? Pn($\boldsymbol{\lambda}$)?

This inspired people to look for weaker metrics which would be small in such situations. The metric d_1 and the metric d_2 derived from d_1 serve the purpose very well.

• Ψ: the sets of functions $f : \mathcal{H} \mapsto [-1, 1]$ such that

$$\sup_{\xi_1 \neq \xi_2 \in \mathcal{H}} \frac{|f(\xi_1) - f(\xi_2)|}{d_1(\xi_1, \xi_2)} \leq 1.$$

• The second Wasserstein metric d_2 between probability measures $\mathbf{Q}_1$ and $\mathbf{Q}_2$ over $\mathcal{H}$ with respect to d_1 is defined as

$$\begin{aligned} d_2(\mathbf{Q}_1, \mathbf{Q}_2) &= \sup_{f \in \Psi} \left| \int f d\mathbf{Q}_1 - \int f d\mathbf{Q}_2 \right| \\ &= \inf \mathbf{E} d_1(X, Y) \\ &= \inf \left\{ \mathbf{P}(|X| \neq |Y|) + \mathbf{E} d_1(X, Y) \mathbf{1}_{[|X|=|Y|]} \right\} \end{aligned}$$

where, again, the infimum is taken over all couplings of (X, Y) with $X \sim \mathbf{Q}_1$ and $Y \sim \mathbf{Q}_2$ due to the duality theorem.

• Interpretation of d_2 metric: it is the total variation distance between the distributions of the total number of points, and given the two distributions have the same number of points, it then measures the average distance of mismatched pairs.

Poisson process approximation using stochastic calculus

Brown (1983):

Theorem 18 *If* $(M, \mathcal{F})$ *is a simple point process with compensator* A*,* $\boldsymbol{\mu}$ *is a measure on* $[0, \infty)$ *and* $t \geq 0$ *is non-random, then*

$$d_{TV}(\mathcal{L}(M^t), Pn(\boldsymbol{\mu}^t)) \leq \mathbf{E}|A - \boldsymbol{\mu}|_t + \mathbf{E}\left\{\sum_{s \leq t} \Delta A_s^2\right\},$$

where $\mathcal{L}(M^t)$ *is the distribution of* M *confined to* $[0, t]$*,* $Pn(\boldsymbol{\mu}^t)$ *is the distribution of a Poisson process on* $[0, t]$ *with mean measure* $\boldsymbol{\mu}$ *and* $|A - \boldsymbol{\mu}|_t$ *is the pathwise variation of the signed measure of* $A - \boldsymbol{\mu}$ *on* $[0, t]$*, i.e.* $|A - \boldsymbol{\mu}|_t = \int_0^t |dA(s) - \boldsymbol{\mu}(ds)|$.

Example 19 *Bernoulli process* Ξ *by* $Pn(\boldsymbol{\mu})$*.*

Theorem 20 *Suppose* M *is a simple point process with arbitrary compensator* A*, and* B *is a continuous increasing function with* $B(0) = 0$*. Let* $p, q \in [1, \infty]$ *satisfying* $1/p + 1/q = 1$*. Then, for any fixed* $\delta > 0$*, we have*

$$\begin{aligned} & d_2(\mathcal{L}(M^T), Pn(B^T)) \leq C_q(\lambda)\|(B, A)\|_p \\ & + \mathbf{E}\delta \wedge |A_{T^-} - B_T| \\ & + \mathbf{P}(|A_{T^-} - B_T| \geq \delta) + \mathbf{E}\Delta A_T \\ & + \mathbf{E}\left\{\sum_{s < T} \Delta A_s^2\right\}, \end{aligned}$$

where $\lambda = B_T$*,* $C_q(\lambda) = \left[\sum_{i=0}^{\infty} \frac{1}{(i+1)^q} \frac{e^{-\lambda}\lambda^i}{i!}\right]^{1/q}$,

$$\begin{aligned} & \|(B, A)\|_p \\ & = \left\{\mathbf{E}\left[\int_{[0,T)} 1 \wedge (|B^{-1} \circ A_t - t| \vee |B^{-1} \circ A_{t^-} - t|) dA_t\right]^p\right\}^{1/p} \end{aligned}$$

Example 21 Cox Process. We say that M is a Cox process with compensator A if, conditional on A, it is a Poisson process with mean measure A. If we let $\mathcal{F}_s = \sigma(M_z, z \le s) \vee \sigma(A)$, then the compensator of $(M, (\mathcal{F}_s)_{s\ge 0})$ is A. For a Cox process M to be simple, A is necessarily continuous so it follows that

$$d_2(\mathcal{L}(M^T), \text{Pn}(B^T)) \le C_q(\lambda)\|(B, A)\|_p$$
$$+\mathbf{E}\delta\wedge \mid A_T - B_T \mid + \mathbf{P}(|A_T - B_T| \ge \delta)$$

Example 22 If $T = 1$, $B_t = at, A_t = bt$ with $a \ge b > 0$, and $p = \infty, q = 1$, then

$$\begin{aligned} d_2(\text{Pn}(A^T), \text{Pn}(B^T)) &\le C_1(a)\frac{b|b-a|}{2a} + |b-a| \\ &\asymp |b-a|. \end{aligned}$$

On the other hand,

$$d_2(\text{Pn}(A^T), \text{Pn}(B^T)) \le (1 \wedge 1.65b^{-\frac{1}{2}})|b-a|,$$

by Stein's method

Poisson process approximation via Stein's method

Stein's equation

• $\Xi \sim \text{Pn}(\boldsymbol{\lambda})$ iff $\mathbf{E}\mathcal{A}h(\Xi) = 0$

• To investigate how close a point process is from $\text{Pn}(\boldsymbol{\lambda})$, estimate $\mathbf{E}\mathcal{A}h(\Xi)$

• For a bounded function f, Stein equation is

$$\mathcal{A}h(\xi) = f(\xi) - \text{Pn}(\boldsymbol{\lambda})(f).$$

• $|\mathbf{E}(f(\Xi)) - \mathsf{Pn}(\boldsymbol{\lambda})(f)| = |\mathbf{E}\mathcal{A}h(\Xi)|$

Example 23 Bernoulli process by $\mathsf{Pn}(\boldsymbol{\mu})$

• Solution? The resolvent of $\mathcal{A}$ at $\rho > 0$ is given by

$$(\rho - \mathcal{A})^{-1}g(\xi) = \int_0^\infty e^{-\rho t}\mathbf{E}g(\mathbf{Z}_\xi(t))dt$$

What we need to do is to argue that when $\rho = 0$ the equation still holds for suitable functions g.

Lemma 24 *For bounded function f, the integral*

$$\int_0^\infty [\mathbf{E}f(\mathbf{Z}_\xi(t)) - Pn(\boldsymbol{\lambda})(f)]dt$$

is well-defined and the solution to Stein equation is

$$h(\xi) = -\int_0^\infty [\mathbf{E}f(\mathbf{Z}_\xi(t)) - Pn(\boldsymbol{\lambda})(f)]dt.$$

Proof. $0 \le f \le 1$. Set $|\xi| = n$, $\boldsymbol{\tau}_{n0} = \inf\{t : \mathbf{Z}_\xi(t) = 0\} = \inf\{t : |\mathbf{Z}_\xi(t)| = 0\} = \tau_{n0}$, and $\boldsymbol{\tau}_{couple} = \inf\{t > \boldsymbol{\tau}_{n0} : |\mathbf{Z}_P(t)| = |\mathbf{Z}_\xi(t)|\}$, where $\mathbf{Z}_P$ is the immigration-death process with initial distribution the same as $\mathsf{Pn}(\boldsymbol{\lambda})$. Then we can couple in such a way that $\mathbf{Z}_P(t) = \mathbf{Z}_\xi(t)$ for $t > \boldsymbol{\tau}_{couple}$, which implies

$$\int_0^\infty |\mathbf{E}f(\mathbf{Z}_\xi(t)) - Po(\boldsymbol{\lambda})(f)|dt \le \mathbf{E}(\boldsymbol{\tau}_{n0} + \boldsymbol{\tau}_{couple}) < \infty.$$

Since the integral is absolutely convergent, we may split it at the first time of jump
$\boldsymbol{\tau} = \inf\{t : \ \mathbf{Z}_\xi(t) \neq \xi\}$:

$$\begin{aligned}
&-(\lambda + n)h(\xi)\\
&= [f(\xi) - \mathrm{Pn}(\boldsymbol{\lambda})(f)]\\
&+(\lambda + n)\mathbf{E}\int_{\boldsymbol{\tau}}^{\infty}[f(\mathbf{Z}_\xi(t)) - \mathrm{Pn}(\boldsymbol{\lambda})(f)]dt\\
&= [f(\xi) - \mathrm{Pn}(\boldsymbol{\lambda})(f)] - \int_\Gamma h(\xi + \delta_\alpha)\boldsymbol{\lambda}(d\alpha)\\
&- \int_\Gamma h(\xi - \delta_\alpha)\xi(d\alpha),
\end{aligned}$$

and some reognization gives that h satisfies Stein equation. ▪

Define

$$\begin{aligned}
&\Delta h(\xi) = \sup_{\eta - \xi \in \mathcal{H}, x \in \Gamma} |h(\eta + \delta_x) - h(\eta)|\\
&\Delta^2 h(\xi)\\
&= \sup_{\eta - \xi \in \mathcal{H}, x, y \in \Gamma} |h(\eta + \delta_x + \delta_y) - h(\eta + \delta_x) - h(\eta + \delta_y) + h(\eta)|
\end{aligned}$$

$\xi|_B$: $\xi|_B(C) = \xi(B \cap C)$ for Borel sets $C \subset \Gamma$.

• Assume that, for each α, there is a Borel set $A_\alpha \subset \Gamma$ such that the mapping

$$\Gamma \times \mathcal{H} \to \Gamma \times \mathcal{H} : (\alpha, \xi) \mapsto (\alpha, \xi|_{A_\alpha^c})$$

is product measurable.

• It is ensured by $A = \{(x,y) : y \in A_x, x \in \Gamma\}$ measurable in Γ^2

Now,

$$\begin{aligned}
&\mathbf{E}\int_\Gamma [h(\Xi - \delta_\alpha) - h(\Xi)]\Xi(d\alpha) \\
&= \mathbf{E}\int_\Gamma \{[h(\Xi - \delta_\alpha) - h(\Xi)] - [h(\Xi|_{A_\alpha^c}) - h(\Xi|_{A_\alpha^c} + \delta_\alpha)]\}\Xi(d\alpha) \\
&+\mathbf{E}\int_\Gamma [h(\Xi|_{A_\alpha^c}) - h(\Xi|_{A_\alpha^c} + \delta_\alpha)][\Xi(d\alpha) - \boldsymbol{\lambda}(d\alpha)] \\
&+\mathbf{E}\int_\Gamma \{[h(\Xi|_{A_\alpha^c}) - h(\Xi|_{A_\alpha^c} + \delta_\alpha)] - [h(\Xi) - h(\Xi + \delta_\alpha)]\}\boldsymbol{\lambda}(d\alpha) \\
&+\mathbf{E}\int_\Gamma [h(\Xi) - h(\Xi + \delta_\alpha)]\boldsymbol{\lambda}(d\alpha)
\end{aligned}$$

Theorem 25 For each bounded measurable function f on $\mathcal{H}$,

$$\begin{aligned}
&|\mathbf{E}f(\Xi) - \mathrm{Pn}(\boldsymbol{\lambda})(f)| \\
\leq\ & \mathbf{E}\int_{\alpha\in\Gamma} \Delta_2 h(\Xi|_{A_\alpha^c})(\Xi(A_\alpha) - 1)\Xi(d\alpha) \\
&+ \min\{\epsilon_1(f,\Xi), \epsilon_2(f,\Xi)\} \\
&+\mathbf{E}\int_{\alpha\in\Gamma} \Delta_2 h(\Xi|_{A_\alpha^c})\boldsymbol{\lambda}(d\alpha)\Xi(A_\alpha),
\end{aligned}$$

where

$$\epsilon_1(f,\Xi) = \mathbf{E}\int_{\alpha\in\Gamma} \Delta h(\Xi|_{A_\alpha^c})|g(\alpha, \Xi|_{A_\alpha^c}) - \mu(\alpha)|\boldsymbol{\nu}(d\alpha)$$

which is valid if Ξ is a simple point process, and

$$\begin{aligned}
&\epsilon_2(f,\Xi) \\
&= \mathbf{E}\int_{\alpha\in\Gamma} |[h(\Xi|_{A_\alpha^c}) - h(\Xi|_{A_\alpha^c} + \delta_\alpha)] - [h(\Xi_\alpha|_{A_\alpha^c}) - h(\Xi_\alpha|_{A_\alpha^c} + \delta_\alpha)]|\boldsymbol{\lambda}(d\alpha).
\end{aligned}$$

Remark How judicious $(A_\alpha; \alpha \in \Gamma)$ are chosen is reflected in the upper bound.

Poisson process approximation: total variation

Lemma 26 If $f = 1_A$, $A \in \mathcal{B}(\mathcal{H})$, then
(i) $\Delta h(\xi) \le 1$ for all ξ.
(ii) $\Delta_2 h(\xi) \le 1$ for all ξ.

Proof.

Theorem 27 We have

$$d_{TV}(\mathcal{L}\Xi, \text{Pn}(\boldsymbol{\lambda})) \le \mathbf{E}\int_{\alpha\in\Gamma}(\Xi(A_\alpha)-1)\Xi(d\alpha)$$
$$+\min\{\epsilon_1(f,\Xi),\epsilon_2(f,\Xi)\} + \mathbf{E}\int_{\alpha\in\Gamma}\boldsymbol{\lambda}(d\alpha)\Xi(A_\alpha),$$

where

$$\epsilon_1(f,\Xi) = \int_{\alpha\in\Gamma}\mathbf{E}|g(\alpha,\Xi|_{A_\alpha^c}) - \mu(\alpha)|\boldsymbol{\nu}(d\alpha)$$

which is valid if Ξ is a simple point process, and

$$\epsilon_2(f,\Xi) = \mathbf{E}\int_{\alpha\in\Gamma}\Delta_2 h(\Xi|_{A_\alpha^c}\wedge\Xi_\alpha)\|\Xi|_{A_\alpha^c} - \Xi_\alpha|_{A_\alpha^c}\|\boldsymbol{\lambda}(d\alpha).$$

Poisson process approximation: Wasserstein metric

- Uniform bounds

Lemma 28 If $f \in \Psi$, then
(i) $\Delta h(\xi) \le 1 \wedge (1.65\lambda^{-0.5}) := c_1(\lambda)$ for all ξ
(ii) $\Delta_2 h(\xi) \le 1 \wedge \frac{2[1+2\ln^+(\lambda/2)]}{\lambda} := c_2(\lambda)$ for all ξ
(iii) $|[h(\xi+\delta_\alpha) - h(\xi)] - [h(\eta+\delta_\alpha) - h(\eta)]| \le c_2(\lambda) d_1'(\xi,\eta)$

Proof.

Remark. The bounds are of the right order.

Theorem 29 We have

$$d_2(\mathcal{L}\Xi, \text{Pn}(\boldsymbol{\lambda})) \leq \mathbf{E}c_2(\lambda)\int_{\alpha\in\Gamma}(\Xi(A_\alpha)-1)\Xi(d\alpha)$$

$$+\min\{\epsilon_1(f,\Xi),\epsilon_2(f,\Xi)\}+\mathbf{E}c_2(\lambda)\int_{\alpha\in\Gamma}\boldsymbol{\lambda}(d\alpha)\Xi(A_\alpha),$$

where

$$\epsilon_1(f,\Xi)=\mathbf{E}c_1(\lambda)\int_{\alpha\in\Gamma}|g(\alpha,\Xi|_{A_\alpha^c})-\mu(\alpha)|\boldsymbol{\nu}(d\alpha)$$

which is valid if Ξ is a simple point process, and

$$\epsilon_2(f,\Xi)=\mathbf{E}c_2(\lambda)\int_{\alpha\in\Gamma}d_1'(\Xi|_{A_\alpha^c},\Xi_\alpha|_{A_\alpha^c})\boldsymbol{\lambda}(d\alpha).$$

- Non-uniform bounds

Lemma 30 For each $f\in\Psi$ and $\xi\in\mathcal{H}$ with $|\xi|=n$, the solution h satisfies

(i) $|\Delta^2 h(\xi)|\leq\frac{5}{\lambda}+\frac{3}{n+1}$.

(ii) For $\xi,\eta\in\mathcal{H}$ and $x\in\Gamma$,

$$|[h(\xi+\delta_x)-h(\xi)]-[h(\eta+\delta_x)-h(\eta)]|$$
$$\leq\left(\frac{5}{\lambda}+\frac{3}{|\eta|\wedge|\xi|+1}\right)d_1''(\xi,\eta).$$

- There is a parallel theorem with constants $c_2(\lambda)$ replaced by the non-uniform bound.

Applications

- Bernoulli process
- Matérn hard core process
- Rare words in biomolecular sequences
- 2-runs
- From Poisson process to compound Poisson approximation