Probability Generating Functional

Ryan McCorvie[*]

November 1st, 2017

This is a study of the probability generating functional (p.g.fl.) and related objects, which are to random measures as the probability generating function is to random variables. The p.g.fl. uniquely specifies the distribution of a random measure, but its much more amenable to algebraic and analytical techniques than finite dimensional (fidi) distributions.

1 Laplace transform

A quick review of some Laplace transform facts.

1. Definition (Laplace transform): The *Laplace transform* of a nonnegative random variable X for $s > 0$

$$\mathcal{L}_X(s) = \mathbb{E}(e^{-sX}) = \int_0^\infty e^{-sx}\,\mu(dx)$$

where μ is the law of X, and $\mu[0,x] = \mathbb{P}(X \leq x)$. For a nonnegative random vector, we generalize the definition to $s \in \mathbb{R}_+^d$ to be $\mathcal{L}_X(s) = \mathbb{E}(e^{-s\cdot X})$

For example, for $X \sim \mathsf{Poisson}(\lambda)$,

$$\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-sk} e^{-\lambda} = \exp(\lambda e^{-s} - \lambda) = \exp(\lambda(e^{-s} - 1))$$

The Laplace transform is essentially the same as the characteristic function $\Phi_X(s) = \mathbb{E}(e^{isX})$ since $\mathcal{L}_X(s) = \Phi(-is))$. However, for nonnegative random variables, Laplace transforms have the advantage of being positive, monotone, convex and analytic. We'll content ourselves with proving $\mathcal{L}_X$ is smooth.

2. Claim $\mathcal{L}_X \in C^\infty(0,\infty)$. *Moreover for* $n \in \mathbb{N}$ *and* $s > 0$

$$\frac{d^n}{dx^n}\mathcal{L}_X(s) = \mathbb{E}(X^n e^{-sX}) = \int_0^\infty x^n e^{-sx}\,\mu(dx)$$

Proof. Calculate $\frac{\mathcal{L}(s+h) - \mathcal{L}(s)}{h} = \int_0^\infty \frac{e^{-hx}-1}{h} e^{-sx}\,\mu(dx)$. However since $xe^{-xs/4} \to 0$ it has a maximum $C > 0$, so using $|e^x - 1| \leq |x|e^{|x|}$

$$\left|\frac{e^{-hx}-1}{h}\right| \leq \frac{e^{xs/4}x|h|}{|h|} \leq Ce^{xs/2} \text{ when } h \leq s/4$$

Therefore dominated convergence yields $\mathcal{L}'(s) = \int_0^\infty \frac{d}{ds} e^{-sx}\,\mu(dx) = -\int_0^\infty xe^{-sx}\,\mu(dx)$. Repeating the argument for higher derivatives gives the desired formula. $\square$

[*]mccorvie@berkeley.edu

I learned the following argument from [Fel66], not sure what the original paper is. A similar argument using exponential distributions instead of Poisson distributions is an exercise in [Wil61]

3. Theorem (Inversion formula) *If μ is the law of X then*

$$\mu[0,x] - \frac{1}{2}\mu(\{x\}) = \lim_{n\to\infty} \sum_{k\le nx} (-1)^k \frac{n^k}{k!} \mathcal{L}_X^{(k)}(n)$$

Proof. For any $x > 0$ and $n \in \mathbb{N}$

$$\sum_{k\le nx} (-1)^k \frac{n^k}{k!} \mathcal{L}_X^{(k)}(n) = \int_0^\infty \sum_{k\le nx} \frac{(nu)^k}{k!} e^{-nu}\, \mu(du)$$

However the expression $\sum_{k\le nx} \frac{(nx)^k}{k!} e^{-nx}$ is $\mathbb{P}(Y \le nx)$ where $Y \sim \mathsf{Poisson}(nu)$. When $x < u$ then

$$\mathbb{P}(Y \le nx) \le \mathbb{P}(|Y - nu| \ge n(u-x)) \le \frac{\mathrm{Var}(Y)}{n^2(u-x)^2} = \frac{u}{n(u-x)^2} \to 0$$

A similar argument shows $\mathbb{P}(Y \le nx) \to 1$ when $x > u$.

For the case $u = x$, since Y is infinitely divisible write $Y = Y_1 + \cdots + Y_n$ where $Y_n \sim \mathsf{Poisson}(u)$ are i.i.d.. Let $\hat{Y}_i := Y_i - u$ be the centered random variable, so that by the central limit theorem

$$\mathbb{P}(Y \le nx) = \mathbb{P}\left(\frac{\hat{Y}_1 + \cdots + \hat{Y}_n}{\sqrt{xn}} \le 0\right) \to \frac{1}{2}$$

So pointwise the integrand converges to $\mathbb{1}_{u<x} + \frac{1}{2}\mathbb{1}_{\{x\}}$. Therefore by dominated convergence (dominating by 1) the integral becomes the inversion formula. ☐

4. Corollary (Uniqueness of Laplace transforms) *For non-negative random variables or vectors X, Y we have $\mathcal{L}_X = \mathcal{L}_Y$ if and only if $X \sim Y$*

Proof. If $X \sim Y$, the Laplace transforms are equal by definition. For random variables, if $\mathcal{L}_X = \mathcal{L}_Y$ then the inversion formula shows they have the same law. For random vectors, we can use the one-dimensional inversion formula and Wold's device that $s \cdot X \sim s \cdot Y$ for all s if and only if $X \sim Y$. ☐

2 Linear Functionals

Basically this follows [DVJ07] chapter 9.4, a lot of which comes from [Wes72]. Some of the same material is also covered in [Kal17] chapter 2.1.

Let $(S, \mathcal{S})$ be a measurable space and let $\mathrm{BM}(S)$ be the set of all bounded measurable functions with bounded support in S. First we state a sort of probabilistic version of the Reisz representation theorem, which says that random measures are equivalent to random continuous linear functionals.

5. Claim *Let $\{\xi_f\}$ be a family of random variables indexed by the elements f of $\mathrm{BM}(S)$. There exists a unique random measure ξ such that*

$$(*) \qquad \xi_f = \int f\, d\xi$$

if and only if

(i) $\xi_{\alpha f + \beta g} = \alpha\xi_f + \beta\xi_g$ a.s. for all scalars α, β and $f, g \in \mathrm{BM}(S)$

(ii) $\xi_{f_n} \to \xi_f$ *a.s. as* $n \to \infty$ *for all sequences* $f_n \uparrow f$ *pointwise where* $f_n \in \mathrm{BM}_+(S)$ *for all* n

Proof. Its clear that $(*)$ implies the conditions from the linearity and monotone convergence properties of integrals.

Conversely, for $A \in \mathcal{S}$ define $\xi(A) = \xi_{\mathbb{1}(A)}$, the element of the family corresponding to the indicator function $\mathbb{1}_A$. Then (i) shows ξ is additive since $\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B$ for disjoint $A, B \in \mathcal{S}$. And (ii) shows ξ is countably additive. For simple functions $s = \sum_i a_i \mathbb{1}_{A_i}$, we can apply (i) to get $\xi_s = \sum_i a_i \xi(A_i) = \int s\, d\xi$. For arbitrary measurable functions, $(*)$ follows from (ii) and approximating by simple functions. Uniqueness follows from lemma 3 in "point process zoo"– finite dimensional distributions are uniquely specified by the joint distribution of integrals $(\xi f_1, \ldots, \xi f_n) = (\xi_{f_1} \ \ldots, \xi_{f_n})$

Don't we need some assumption about positivity to ensure this is a positive measure? Like $\xi_f \geq 0$ a.s. whenever $f \geq 0$. □

Now we use functionals to define a generalization of Laplace transforms suitble for analysis of random measures.

6. Definition (Laplace functional): The *Laplace transform* of a random measure ξ is given for any $f \in \mathrm{BM}_+(S)$ by

$$\mathcal{L}_\xi[f] = \mathbb{E}[e^{-\int f\, d\xi}] = \mathbb{E}[e^{-\xi f}]$$

Sometimes its useful to employ the conditional Laplace transform with repect to some σ-algebra $\mathcal{F}$

$$\mathcal{L}_\xi[f \mid \mathcal{F}] = \mathbb{E}(e^{-\xi f} \mid \mathcal{F})$$

in which case $\mathcal{L}_\xi[f] = \mathbb{E}(\mathcal{L}_\xi[f \mid \mathcal{F}])$

7. Lemma (Continuity) *Consider a sequence* f_n *with* $\sup_{x\in X} |f_n(x) - f(x)| \to 0$ *(in other words,* f_n *converges pointwise uniformly to* f *in* X*). Then* $\mathcal{L}[f_n] \to \mathcal{L}[f]$ *if any of the following hold*

(i) $f_n \uparrow f$ *pointwise*

(ii) ξ *is totally bounded*

(iii) there is a bounded Borel set containing the support of every f_n

Proof. By condition (i) and monotone convergence, $\xi f_n \to \xi f$ for each realization of ξ and hence almost surely. Using condition (ii) and bounded convergence we can make a similar argument since the f_n are uniformly bound by some constant (for large enough n, the functions f_n are close to f which is bounded). If bounded $B \supset \operatorname{supp} f_n$ for all n Using condition (iii) we can consider the random measure defined by $\hat{\xi}(A) = \xi(B \cap A)$. This random measure is totally bounded, and also $\hat{\xi} f_n = \xi f_n$ since $f_n = 0$ outside of B. So this reduces to case (ii).

In any case, given $\xi f_n \to \xi f$ alomst surely, bounded convergence implies $\mathbb{E}(\exp(-\xi f_n)) \to \mathbb{E}(\exp(\xi f))$, which is the desired result. □

8. Theorem (Laplace Functional Uniqueness) *Let* $\Lambda[f]$ *be defined for all* $f \in \mathrm{BM}_+(S)$*. Then* $\Lambda = \mathcal{L}_\xi$ *is the Laplace transform of a random measure* ξ *on* S *if and only if*

(i) For every finite family $f_1, \ldots, f_n \in \mathrm{BM}_+(S)$ *the function* $\Lambda[s_1 f_1 + \cdots + s_n f_n]$ *is the Laplace transform of some proper random vector* $(\xi_{f_1}, \ldots, \xi_{f_n})$

(ii) For every sequence $f_1, f_2, \cdots \in \mathrm{BM}_+(S)$ *with* $f_n \uparrow f$ *pointwise,* $\Lambda[f_n] \to \Lambda[f]$

(iii) $\Lambda(0) = 1$

Moreover when the conditions are satisfied, the functional $\mathcal{L}$ *uniquely determines the distribution of* ξ

Proof. If Λ is the Laplace transform of a random measure ξ so that $\Lambda = \mathcal{L}_\xi$, then let $\xi_f := \xi f$, and note this is a random variable for all $f \in \mathrm{BM}_+(S)$. Then (i) follows from the properties of of Laplace transforms of random variables, and (ii) follows from lemma 7 and (iii) follows from the definition. Furthermore, if there is a measure ξ' such that $\mathcal{L}_{\xi'}[f] = \Lambda[f] = \mathcal{L}_\xi[f]$ for all $f \in \mathrm{BM}_+(S)$ then the fidi distributions $(\xi f_1, \ldots, \xi f_n) \sim (\xi' f_1, \ldots, \xi' f_n)$ are the same, by the uniqueness of Laplace transforms of random vectors (theorem 4). But this is a necessary and sufficient condition that $\xi \sim \xi'$ as random measures.

What remains to show a random measure ξ exists so that $\Lambda = \mathcal{L}_\xi$ whenever (i)-(iii) are satisfied. Condition (i) implies that the Kolmogorov consistency conditions are satisfied. Specifically, the sum $\sum s_i f_i = \sum s_{\pi i} f_{\pi i}$ is invariant under any permutation π. The marginals are consistent because $\Lambda(s_1 f_1 + \cdots + s_n f_n)$ corresponds to the Laplace transform of a random vector $(\xi_{f_1}, \ldots, \xi_{f_n})$, which we write as $\mathcal{L}_{(\xi_{f_1}, \ldots, \xi_{f_n})}(s_1, \ldots, s_n)$. For consistency we wish to show that

$$\mathcal{L}_{(\xi_{f_1}, \ldots, \xi_{f_n})}(s_1, \ldots, s_{n-1}, 0) = \mathcal{L}_{(\xi_{f_1}, \ldots, \xi_{f_{n-1}})}(s_1, \ldots, s_{n-1})$$

But this is clearly true because $\Lambda(s_1 f_1 + \cdots + s_{n-1} f_{n-1} + 0 \cdot f_n) = \Lambda(s_1 f_1 + \cdots + s_{n-1} f_{n-1})$. Therefore by Komolgorov's extension theorem, there is a common probability space where the ξ_f are random variables for all $f \in \mathrm{BM}_+(S)$.

Next we will show that $f \mapsto \xi_f$ is a continuous linear functional, which, by claim 5, implies that $\xi_f = \xi f$ for some measure ξ. For linearity, $\xi_{f_3} = \alpha \xi_{f_1} + \beta \xi_{f_2}$ iff for all $s_1, s_2, s_3 \in \mathbb{R}_+$ we have $s_1 \xi_{f_1} + s_2 \xi_{f_2} + s_3 \xi_{f_3} \sim (s_1 + \alpha s_3) \xi_{f_1} + (s_2 + \beta s_3) \xi_{f_2}$. But this is equivalent to the Laplace transforms being equal, which occurs only when

$$\Lambda(s_1 f_1 + s_2 f_2 + s_3 f_3) = \Lambda((s_1 + \alpha s_3) f_1 + (s_2 + \beta s_3) f_2)$$

Now this is certainly true whenever $f_3 = \alpha f_1 + \beta f_2$ which verifies condition (i) of claim 5.

To show condition (ii) in claim 5, note that when $f_n \uparrow f$ then $\xi_{f - f_n}$ converges in distribution to 0 and hence in probability. But then by linearity, $\xi_{f_n} \to \xi_f$ in probability. The linearity of $f \mapsto \xi_f$ and the positivity of the measure implies that ξ_{f_n} is increasing when f_n is increasing. This means that ξ_{f_n} converges almost surely to some limit in $[0, \infty]$. But since ξ_{f_n} converges in probability the limit must be ξ_f. $\square$

For point processes, it is sometimes more convenient to consider the following functional which is an analogue to the probability generating function.

9. Definition (Probability generating functional): For a point process N which is almost surely finite, and for measurable $g \geq 0$, the *probability generating functional* is given by

$$G_\xi[h] = \mathbb{E} \exp(\xi \log h) = \mathbb{E}_\xi \prod_{k \geq 1} h(x_k)$$

where $\{x_k\}_{k \geq 1}$ represents a realization of the point process $\xi = \sum_k \delta_{x_k}$, and the expectation is taken over all realizations.

One way to ensure the product is well defined is to restrict h to the class of measurable functions $h : X \to (0, 1]$ where h is 1 outside of some compact set. Then any realization of ξ given by $\{x_k\}_{k \geq 1}$ has only finitely points where $h(x_k) \neq 1$, and also the product is always positive. Note that $G_\xi[h] = \mathcal{L}_\xi[-\log h]$, and the condition that $-\log h \in \mathrm{BM}_+(S)$ is the same as the conditions above on the range of h. An alternative way to ensure the product exists is that $\int |\log h(x)| \, M(dx)$ where M is the intensity measure of ξ.

From the definition its evident that the product $h(x_1) \cdots h(x_k)$ is positive, monotonic and convex in g for every realization of ξ, so the p.g.fl. must also be positive, monotonic and convex. There's a variation of theorem 8 for p.g.fl.

10. Theorem *Let the functional $G[h]$ be real-valued and defined for all continuous measurable $h : X \to (0, 1]$ which is 1 outside of a compact set. Then G is a p.g.fl. of a point process ξ if and only if*

(i) For every h of the form

$$1 - h(x) = \sum_{k=1}^{n} (1 - z_k) \mathbb{1}_{A_k}(x)$$

for bounded disjoint Borel sets $A_1, \dots, A_n$ and $z_i \in (0, 1]$ the p.g.fl. reduces to the joint p.g.f. for the integer valued random variables ξA_k

(ii) for every sequence $h_n \downarrow h$ poitwise, $G[h_n] \to G[h]$ whenever $1 - h$ has bounded support

(iii) $G[1] = 1$

When these conditions are satisfied, the functional G uniquely determines the distribution of ξ.

3 Moment Measures and Generating Functional Expansions

This follows the development in [LP17] chapter 4 (with the theorems in [Kal17] chapter 1.2 providing context). This material is developed in a much more confusing way in [DVJ03] chapters 5.3, 5.4 and 5.5.

Given a point process ξ on S, its easy to define a point process on S^n by $\xi^n = \xi \otimes \cdots \otimes \xi$. Given a realization of $\xi = \sum_{k \in I} \delta_{s_k}$ then a realization of $\xi^n = \sum_{(k_1, \dots, k_n) \in I^n} \delta_{(s_{k_1}, \dots, s_{k_n})}$. However, often it is more convenient to exclude the diagonal sets.

So given a realization $\xi = \sum_{i \in I} \delta_{s_i}$ where the points range over some (countable) index set I then

$$\xi^{(n)} = \sum_{\substack{i_1, \dots, i_n \in I \\ i_k \text{ distinct}}} \delta_{(s_{i_1}, \dots, s_{i_n})}$$

Note the s_k themselves need not be distinct. However, in the case that ξ is simple then $\xi^{(n)} = \mathbb{1}_{\{s_1, \dots, s_n \text{ distinct}\}} \xi^n$. Kallenberg calls $\mu^{(n)}$ the *factorial measure* [Kal17]. Note that when $A_1, \dots, A_n$ are disjoint then

$$\xi^n(A_1 \times \dots A_n) = \xi^{(n)}(A_1 \times \cdots \times A_n) = \xi(A_1) \cdots \xi(A_n)$$

When $B = A^n$

$$\xi^n(A^n) = \xi(A)^n \qquad \text{but} \qquad \xi^{(n)}(A^n) = \xi(A)(\xi(A) - 1) \cdots (\xi(A) - n + 1) = \xi(A)^{(n)}$$

In particular $\xi^{(n)}(A^n) = 0$ unless A has at least n points in it. As another example, for arbitrary $A, B \in \mathcal{S}$, $\xi^2(A \times B) = \xi(A)\xi(B)$ but $\xi^{(2)}(A \times B) = \xi(A)\xi(B) - \xi(A \cap B)$.

11. Definition (Moment measures): For measurable $B \in \mathcal{S}^n$, define

$$\begin{aligned} M_n(B) &= \mathbb{E}(\xi^n(B)) && \text{moment measure} \\ M_{(n)}(B) &= \mathbb{E}(\xi^{(n)}(B)) && \text{factorial moment measure} \\ J_n(B) &= \mathbb{E}(\mathbb{1}_{\{\xi(S)=n\}} \xi^{(n)}(B)) && \text{Janossy measure} \end{aligned}$$

These are measures by the linearity and continuity of expectations. When $n = 1$, $M_{(1)}(B) = M_1(B) = \mathbb{E}(\xi(B))$ is just the intensity measure, which we abbreviate as M.

For a general point process and bounded set $B \in \mathcal{S}$ we can define the local Janossy measure $J_{n,B}$ as the Janossy measure with respect to the restricted process $\xi_B(B') = \xi(B \cap B')$. For $B' \supset B$, the local measures $J_{n,B}$ and $J_{n,B'}$ agree on subsets of B^n, which show that the Janossy measure really is localized. (This follows from the trivial observation that $A \cap B = A \cap B'$ if $A \subset B \subset B'$).

The Janossy measure also provides a sort of probability density function for the point process. In particualr, if J_n is absolutely continuous with respect to the Lesbegue measure, then its derivative $j_n(x_1, \ldots, x_k)\, dx_1 \ldots dx_k$ is the probability density there is exactly one point at each of $\{x_1, \ldots, x_k\}$. More generally we have

12. Lemma *For a partition $A_1, \ldots, A_k$ of S we have*

$$\mathbb{P}(\xi A_1 = n_1, \ldots, \xi A_k = n_k) = n_1! \cdots n_k!\, J_n(A_1^{n_1} \times \cdots \times A_k^{n_k})$$

Proof. This follows from taking expectations of the following identity, where $n = n_1 + \cdots + n_k$

$$\int \mathbb{1}_{\{\xi S = n\}} \mathbb{1}_{A_1^{n_1} \times \cdots \times A_k^{n_k}}\, d\xi^{(n)} = n_1! \cdots n_k!\, \mathbb{1}_{\{\xi A_1 = n_1, \ldots, \xi A_k = n_k\}}$$

Observe that the integrand on the left is non-zero if and only if the first n_1 coordinates in a realization of $\xi^{(n)}$ are in A_1, the next n_2 in A_2 and so forth. But, since the A_k are a partition of S, this happens if and only if the realization of ξ consists of exactly n points, where A_k contains n_k points. This realization of ξ corresponds to $n_1! \cdots n_k!$ points in the realization of $\xi^{(n)}$, permuting each block of n_k coordinates. □

13. Corollary *If ξ and ξ' are point processes on X whose corresponding local Janossy measures are equal $J_{n,B} = J'_{n,B}$ for each $n \in \mathbb{N}$ then the distributions are equal conditional on being finite, or $\mathbb{P}(\eta(B) < \infty, \eta \in \cdot) = \mathbb{P}(\eta(B) < \infty, \eta' \in \cdot)$. In particular if η and η' are finite almost surely, they have the same distribution iff they have the same Janossy measures.*

14. Corollary *The Janossy measures satisfy $\sum_{n=0}^{\infty} \frac{J_n(S^n)}{n!} = 1$*

For a random variable, the Taylor expansion of the p.g.f. at 0 is related to the probabilities, and the Taylor expansion of the p.g.f. at 1 is related to the moments. We have a similar relationship of the p.g.fl. to Janossy measures and factorial moment measures.

15. Claim (Expansions of the PGFL) *We have the following expansions of the p.g.fl. G of a point process ξ*

$$G[f] = \sum_{n=0}^{\infty} \frac{1}{n!} \int_{S^n} \prod_{i=1}^{n} f(s_i)\, J_n(ds_1 \ldots ds_n)$$
$$G[1-g] = 1 + \sum_{n=1}^{\infty} (-1)^n \frac{1}{n!} \int_{S^n} \prod_{i=1}^{n} g(s_i)\, M_{(n)}(ds_1 \ldots ds_n)$$

Proof. Suppose the realization of $\xi = \sum_{i=1}^{n} \delta_{x_i}$ with n points. Then $\xi^{(n)} = \sum_{\text{perm } \pi} \delta_{(x_{\pi 1}, \ldots, x_{\pi n})}$, where the sum is over the set of all permutations. Since the product $f(x_1) \cdots f(x_n)$ symmetric with respect to permutations of the coordinates we have

$$\int_{S^n} \prod_{i=1}^{n} f(s_i)\, d\xi^{(n)} = \sum_{\text{perm. } \pi} \prod_{i=1}^{n} f(x_{\sigma i}) = n! \prod_{i=1}^{n} f(x_i)$$

and therefore

$$\exp \xi \log f = \sum_{n=0}^{\infty} \mathbb{1}_{\{\xi(S)=n\}} \frac{1}{n!} \xi^{(n)} \prod_i f(s_i)$$

Taking expectations gives the first expansion.

For the second first note that

$$\prod_{i=1}^{n} (1 - g(x_i)) = 1 + \sum_{k=1}^{n} (-1)^n \sum_{1 \le i_1 < \cdots < i_k \le n} \xi^{(n)} \prod_i g(s_{i_k})$$

Like in the previous case, owing to the symmetry of the product $g(s_1) \cdots g(s_k)$ under permutations, when $\xi = \sum \delta_{x_i}$ we can rewrite the sum in terms of the factorial measure

$$k! \sum_{1 \le i_1 < \cdots < i_k \le n} \xi^{(n)} \prod_k g(x_{i_k}) = \sum_{\substack{1 \le i_1, \ldots, i_k \le n \\ i_1, \ldots, i_k \text{ distinct}}} \prod_k g(x_{i_k}) = \xi^{(n)} \prod_{i=1}^{k} g(s_i)$$

Using this and the fact that $\xi^{(n')} = 0$ whenever the realization of ξ only has $n < n'$ points, we get the identity

$$\exp \xi \log(1 - g) = 1 + \sum_{k=1}^{\infty} (-1)^n \frac{1}{k!} \xi^{(k)} \prod_{i=1}^{k} g(s_i)$$

Taking expectations of both sides gives the second expansion. □

16. Corollary *If the moment measures grow slowly enough, say $M_{(n)}(S^n) \le n! c^k$, then the moment measures uniquely specify the distribution of ξ*

Proof. The expansion converges for $g < c^{-1}/2$ and therefore the p.g.fl. is uniquely determined given the moment measures. □

17. Claim *For a Poisson process directed by λ the factorial moment measures are given by $M_{(n)} = \lambda^n$*

Proof. By inspection from the expansion of the p.g.fl. or using the Mecke equation as in [LP17] □

4 Calculations

18. Claim *(i) For a Cox process ξ driven by η*

$$G_\xi[f \mid \eta] = \exp(\eta(f - 1))$$

(ii) For a Poisson process with intensity μ

$$G_\xi[f] = \exp(\mu(f - 1))$$

(iii) For a mixed binomial process

$$G_\xi[f \mid \kappa] = (\mu f)^\kappa$$

Proof. (i) As shown earlier, if $\kappa \sim \mathsf{Poisson}(m)$ then $\mathbb{E}\, s^\kappa = \exp(m(s-1))$. Therefore if $f = s_1 \mathbb{1}_{A_1} + \dots + s_n \mathbb{1}_{A_n}$ is a simple function for disjoint A_k and constants $s_k \geq 0$ and ξ is a Poisson process driven by η

$$\mathbb{E} \exp(\xi \log f) = \mathbb{E} \exp \sum_{k \leq n} \log s_k \xi A_k = \prod_{k \leq n} \mathbb{E}\, s_k^{\xi A_k} = \prod_{k \leq n} \exp(\eta A_k (s_k - 1)) = \exp(\eta(f-1))$$

For general $f \in \mathrm{BM}_+(X)$ approximate by simple functions and use monotone convergence. Conditioning on η gives the general result.

(ii) This is a special case of (i)

(iii) Let $\xi = \sum_{k \geq n} \delta_{s_k}$ where s_k are iid chosen with law μ and $n \in \mathbb{Z}_+$. For any $f \in \mathrm{BM}_+(X)$

$$\mathbb{E} \exp(\xi \log f) = \mathbb{E} \prod_{k \leq n} f(s_k) = \prod_{k \leq n} \mathbb{E}\, f(s_k) = (\mu f)^n$$

So conditioning on the value of κ gives the result.

□

Does this mean we've proved the existence of a Poisson or even Cox process? The standard construction is just mixed binomial where the number of points has a Poisson distribution, which is intuitive as a finite number of random variables.

So let's calculate the p.g.fl. of a shot noise process. In this case ξ is a Cox process on $\mathbb{R}$ where the intensity measure is given by $\sum_i Y_i g(x - x_i)\, dx$ where the x_i are Poisson with intensity ν and the Y_i are i.i.d. random variables. Therefore by (i)

$$G_\xi[f \mid Y_i, x_i] = \exp\left(\sum_i -Y_i \int_0^\infty (1 - f(u + x_i)) g(u)\, du \right)$$

Taking expectations over the Y_i and recognizing the p.g.fl. is the product of Laplace transforms of i.i.d. Y

$$G_\xi[f \mid x_i] = \prod_i \mathcal{L}_Y \left[\int_0^\infty (1 - f(u + x_i)) g(u)\, du \right]$$

Taking expectations over the x_i again using part (i) above

$$(**) \quad G_\xi[f] = \exp\left(\int 1 - \mathcal{L}_Y \left[\int_0^\infty (1 - f(u+x)) g(u)\, du \right] \nu(dx) \right)$$

Now let's turn to cluster processes. See [Wes71] for more examples.

19. Claim *(i) For a cluster process ξ with center process ξ_c on S and cluster members ζ on T indexed by S (with a slight abuse of notation) let $G_c := G_{\xi_c}$ be the probability generating functional of ξ_c and (again slightly abusing notation, conflating kernels with conditionals) let $G[\cdot \mid s] := G_{\zeta(s,\cdot)}$ be the probability generating functional of $\eta(s, \cdot)$. Then*

$$G_\xi[f] = G_c[G[f \mid s]]$$

(ii) If $M^c_{(n)}$ are the moment measures of the center process and $N^x_{(n)}$ are the moment measures of

the clusters, then the first two moment measures of the cluster process are given by

$$M(A) = \int_S N^x(A)\, M^c(dx)$$
$$M_{(2)}(A \times B) = \int_S N^x_{(2)}(A \times B)\, M^c(dx) + \int_{S^2} N^x(A) N^y(B)\, M^c_{(2)}(dx\, dy)$$

(iii) *For a Poisson cluster process ξ with cluster center intensity λ and cluster members ζ, let $G[f \mid s] := G_{\zeta(s,\cdot)}[f]$ be the the probability generating functional of $\zeta(s,\cdot)$. Then*

$$G_\xi[f] = \exp(\lambda(G[f \mid s] - 1)) = \exp\left(\int (G[f \mid s] - 1)\, \lambda(ds) \right)$$

(iv) *For a Neyman-Scott cluster process ξ with Poisson cluster process given by intensity λ and with i.i.d clusters members given by a mixed binomial process with parameters κ and μ. Let $g(s) = \mathbb{E}\, s^\kappa$ be the p.g.f. of κ and let τ_u be the translation operator so that $\tau_u f(s) = f(s-u)$. Then*

$$G_\xi[f] = \exp(\lambda(g(\mu(\tau_s f) - 1))) = \exp\left(- \int 1 - g\left(\int f(t+s)\, \mu(dt) \right)\, \lambda(dx) \right)$$

Proof. (i) For a realization of of the cluster process $\xi_c = \sum_k \delta_{s_k}$ we get a realization $\xi = \sum_k \xi_{s_k}$ where each ξ_{s_k} is chosen according to the law $\zeta(s_k, \cdot)$. So conditioning on the s_k calculate

$$\mathbb{E}(\exp \xi \log f \mid s_1, s_2, \ldots) = \mathbb{E}\left(\exp \sum_k \xi_{s_k} \log f \right) = \prod_k \mathbb{E}(\xi_{s_k} \log f) = \prod_k G[f \mid s_k]$$

However the right hand side equals $\exp \xi \log G[f|s]$. Taking expectations over the realizations of ξ yields the desired result. Note the conditional independence is the key to the above argument.

(ii) Use the formula for (i), and the moment expansion of the cluster center p.g.fl. G_c to get

$$G_\xi[f] = 1 + \int_S (G[f \mid x] - 1)\, M^c(dx) + \frac{1}{2} \int_{S^2} (G[f \mid x] - 1)(G[f \mid y] - 1)\, M^c_{(2)}(dx\, dy) + \ldots$$

Let $f = 1 + g$ and insert the moment expansion for the cluster member p.g.fl.

$$G[1 + g \mid x] - 1 = \int_S g(s)\, N^x(ds) + \frac{1}{2} \int_{S^2} g(s_1) g(s_2)\, N^x_{(2)} + \ldots$$

Combine terms in powers of g and use Fubini's theorem to switch the order of integration. Then we can equate terms in the moment expansion for $G_\xi[1 + g \mid x]$ to get the above formulas.

An alternative approach is to note condition on the realization of $\xi_c = \sum_k \delta_{s_k}$ to find

$$\mathbb{E}(\xi \mid s_1, s_2, \ldots) = \mathbb{E} \sum_k \eta_{s_k} = \sum_k N^{x_i} = \int N^x \xi_c(dx)$$

Taking expecations of both sides yields the desired formula. The second moment measure can be found similarly, where one term correspond to the case that the two points come from a single cluster and the other term the case they come from two different clusters.

(iii) This is just case (i) with $G_c[f] = \exp(\lambda(f-1))$.

(iv) This is just case (iii) where $G[f \mid s] = \mathbb{E}(\mu\tau_s f)^\kappa = g(\mu\tau_s f)$ by claim 18 part (iii) and the definition of the p.g.f. of κ.

$\square$

Comparing (iii) to equation ($**$) shows that shot noise process also has a cluster representation, since its p.g.fl. matches the p.g.fl. of a cluster process. Simplifying this example a little, given a point process ξ and integrable f, the Cox process with intensity given by $\int f(t-u)\xi(du)$ is equivalent to a cluster process with center ξ and clusters given by a Poisson process with mean $(f(t))$. I think this is a little surprising!

5 Branching Processes

20. Claim *Let ξ be a general branching process and let $G[f \mid x]$ represent the p.g.fl. of the branching kernel.*

(i) Let $G_n[f \mid x]$ p.g.fl. for the point process which represents members of the nth generation of the branching process starting with a single member at x. Then

$$G_{n+1}[f \mid x] = G[G_n[f \mid \cdot] \mid x]$$

and hence by induction $G_n[f \mid x] = G^{(n)}[f \mid x]$, the nth functional iterate of $G[f \mid x]$.

(ii) Let $\rho(x)$ represent the probability that a branching process starting at x goes extinct. Then $\rho(x) = \lim_{n\to\infty} G_n[0 \mid x]$ and its the smallest non-negative solution which satisfies the functional equation $\rho(x) = G[\rho \mid x]$.

(iii) Let $H_n[f \mid x]$ represent the p.g.fl. for the point process which contains the points in all the generations of the branching process up to n starting from a single individual at x. Then

$$H_{n+1}[f \mid x] = f(x)G[H_n[f \mid \cdot] \mid x]$$

If extinction occurs with probability 1, then $\lim_{n\to\infty} H_n[f \mid x]$ exists and is the p.g.fl. H for the point process consisting of the points in all the generations. The p.g.fl. H satisfies the functional equation

$$H[f \mid x] = f(x)G[H[f \mid \cdot] \mid x]$$

(iv) Let the factorial moment measures associated with $G[\cdot \mid x]$ be given by $N^x_{(n)}$ and the factorial moment measures associated with $H[\cdot \mid x]$ be given by $M^x_{(n)}$. Then the first two factorial moments satisfy

$$\begin{aligned}
M^x(A) &= \delta_x(A) + \int_S M^s(A)\, N^x(ds) \\
M^x_{(2)}(A\times B) &= \delta_x(A)\int_S M^s(B)\, N^x(ds) + \delta_x(B)\int_S M^s(A)\, N^x(ds) \\
&\qquad + \int_S M^s_{(2)}(A\times B)\, N^x(ds) + \int_{S^2} M^{s_1}(A) M^{s_2}(B)\, N^x_{(2)}(ds_1\, ds_2)
\end{aligned}$$

Proof. (i) This follows from claim 19 part (i) and the recursive definition of a cluster process.

(ii) If $f = 0$ then $\prod_i f(x_i) = \mathbb{1}_{\xi(S)=0}$. Therefore $G_n[0 \mid x] = \mathbb{P}(\text{extinction before } n\text{th generation})$. Now $\rho_1(x) = G[0 \mid x] \geq 0$ and, so because p.g.fl.'s are monotonic, applying $G_n[\cdot \mid x]$ to both sides shows $\rho_{n+1}(s) \geq \rho_n(x)$ for all n. Since $\rho_n \leq 1$ it must have a pointwise limit $\rho_n \uparrow \rho$. Taking limits of the equation in (i) the function must satisfy $\rho(x) = G[\rho \mid x]$. Why is it the smallest solution? Something to do with convexity I think, but I can't think of the generalization of Galton–Watson. Maybe there are at most two solutions, like in the multitype Galton–Watson case.

(iii) The nth generation process consists of a (non-random) point at x, plus $n-1$ generation processes rooted at $x_1, x_2, \ldots$ where the x_i are given by the point process associated with $G[\cdot \mid x]$. By independence, conditional on the x_i, the p.g.fl. $H_n[f \mid x, x_1, x_2 \ldots] = f(x) \prod_i H_{n-1}[f \mid x_i]$ and therefore, taking expectations, $H_n[f \mid x] = f(x) G[H_{n-1}[f \mid \cdot] \mid x]$ as desired. Let $\zeta_{x,n}$ represent the nth generation of the point process started at x. Then if extinction happens almost surely, then almost surely there is an n such that $\zeta_{x,n} = \zeta_{x,n+1} = \zeta_{x,n+2} = \ldots$ and $\lim H_n[f \mid x]$ converges almost surely. Call the limit $H[f \mid x]$. Taking limits of the functional equation for H_n gives the functional equation for H.

(iv) Use the formula in (iii) and the factorial moment expansions as in claim 19 part (ii)

□

Let ξ be a *stationary* Poisson branching process whose branching kernel is driven by an intensity given by $\eta(x, A) = \lambda(A - x)$. Assume that $m = \lambda(S) < 1$. Then by the Galton–Watson theory, exinction happens almost surely since the average number of offspring of each individual is $m < 1$.

Because the branching kernel is stationary, the factorial moment measures are stationary, and $M_{(n)}(A \mid x) = M_{(n)}(A - (x, x, \ldots, x) \mid 0)$. Let $M_{(n)} := M_{(n)}(\cdot \mid 0)$. Using part (iv) of claim 20 combined and the fact Poisson factorial moments have the form $N^0_{(n)} = \lambda^n$, we get the following recursive formulas for the moments.

$$
\begin{aligned}
M(A) &= \delta_0(A) + \int_S M(A - s)\,\lambda(ds) \\
&= \delta_0(A) + \lambda(A) + \lambda * \lambda(A) + \lambda * \lambda * \lambda(A) + \ldots \\
M_{(2)}(A \times B) &= M(A)M(B) + \int_S M_{(2)}(A - s, B - s)\lambda(ds) - \delta_0(A)\delta_0(B)
\end{aligned}
$$

The second expression for $M(A)$ comes by interatively substituting $M(A)$ into its recursive formula.

We can further specialize to the branching process in the Hawkes process, where $S = \mathbb{R}$ and $\lambda((\infty, 0]) = 0$, so there are no offspring before time 0. Let $\tilde{M}_{(n)}(u) = \int_{\mathbb{R}^n_+} e^{-u \cdot x} M_{(n)}(dx)$ be the Laplace transform of the factorial moment measures $M^x_{(n)}$. Then the previous formulas imply

$$
\widetilde{M}(u) = \frac{1}{1 - \widetilde{\lambda}(u)} \qquad \widetilde{M}_{(2)}(u, v) = \frac{\widetilde{M}(u)\widetilde{M}(v) - 1}{1 - \widetilde{\lambda}(u + v)}
$$

6 Hawkes calculations

A Hawkes process is a cluster process where the cluster center process is a Poisson process whose intensity is ν times the Lesbegue measure, and the clusters the translations of a branching process with Poisson branching kernel. The branching kernel intensity is given by $\eta(x, A) = \int_A \gamma(u - x)\,du$ where $\gamma : \mathbb{R} \to \mathbb{R}_+$ satisfies $\gamma(x) = 0$ for $x \leq 0$ and $m = \int_{\mathbb{R}} \gamma(x)\,dx < 0$. Its evident from the definition that the process is stationary. Here are a calculations related to the Hawkes process, primarily from [HO74].

First note, that since $m < 1$, Galton–Watson theory says the clusters are almost surely finite. By lemma 12 in "zoo" this means the process is locally finite. We can use claim 19 and claim 20 and 18 to write the p.g.fl. in terms of the translation operator $\tau_t f(u) = f(u - t)$

$$
G[f] = \exp\left(\int_{-\infty}^{\infty} \nu(H[\tau_t f]] - 1)\,dt \right)
$$

where H is the p.g.fl. of a cluster starting from 0 which satisfies

$$
\begin{aligned}
H_{n+1}[f] &= f(0) \exp\left\{ \int_0^\infty H_n[\tau_t f] \gamma(t)\, dt \right\} \\
H[f] &= \lim_{n\to\infty} H_n[f]
\end{aligned} \tag{1}
$$

and also

$$
H[f] = f(0) \exp\left\{ \int_0^\infty H[\tau_t f] \gamma(t)\, dt \right\} \tag{2}
$$

The equations are hard to solve, but we can try to tease some information out. When $f(x) = s$ is a constant, then H becomes the p.g.f. for the total progeny in the Poisson Galton–Watson process. In theory we can get the interval distributions by solving (2). For example, let $\pi_{[a,b]}(z)$ be the p.g.f. for the number of points in the interval $[a, b]$. This corresponds to $H[f]$ where $f = 1 + (z-1)\mathbb{1}_{[a,b]}$ and therefore satisfies the equation

$$
\pi_{[a,b]}(z) = \begin{cases} \exp\left\{ \int_0^b (\pi_{[a-t,b-t]}(z) - 1)\, \gamma(t)\, dt \right\} & a > 0 \\ z \exp\left\{ \int_0^b (\pi_{[a-t,b-t]}(z) - 1)\, \gamma(t)\, dt \right\} & a \le 0 \le b \\ 1 & b < 0 \end{cases} \tag{3}
$$

Specializing (3) to $a = 0$ we can study the cluster distribution itself. This p.g.f. satisfies

$$
\pi_{[0,u]}(z) = z \exp\left(\int_0^b [\pi_{[0,u-t]}(z) - 1]\, \gamma(t)\, dt \right) \qquad u \ge 0
$$

In the limit $u \to \infty$ this becomes the p.g.f. for the total cluster size

$$
\pi(z) = z \exp\left(m(\pi(z) - 1) \right)
$$

From this it follows that the mean and variance of the cluster size are $1/(1-m)$ and $1/(1-m)^3$. According to [Lew69] theorem of Lewis for Poisson branching processes says that the number of points suitably scaled has a normal limit. (I think, I can't access this paper). In the case of the equilibrium Hawkes process ξ

$$
\lim_{u\to\infty} \frac{\xi[0,u] - u\nu/(1-m)}{\sqrt{\nu/(1-m)^3}} \sim \mathsf{Norm}(0,1)
$$

Starting again from (3), the p.g.f. for the equilibrium Hawkes process for the number of events in $[0, l]$ (or any length l interval since the process is stationary) is given by

$$
Q_l(z) = \exp\left\{ \int_{-l}^\infty [\pi_{[t,t+l]}(z) - 1] \nu\, dt \right\}
$$

By setting $z = 0$ we can calculate the survival function for the forward recurrence time of the Hawkes process.

$$
\begin{aligned}
\mathbb{P}(R > l) &= \mathbb{P}\{\text{no events in } [0, l]\} = Q_l(0) \\
&= \exp\left(-vl - v \int_0^\infty [1 - \phi(t, l)]\, dt \right)
\end{aligned}
$$

where $\phi(y,l) = \pi_{[y,y+l]}(0)$ so

$$\phi(t,l) = \begin{cases} \exp\left(\int_0^{y+l} [\phi(y-t,l)-1]\,\gamma(t)dt\right) & y > 0 \\ 0 & y \le 0 \le y+l \\ 1 & y < -l \end{cases}$$

Next let $f(x) = \mathbb{1}_{[0,x]}(u)$ then $H[f]$ is the p.g.f. for the total cluster size having length at most x. In this case (2) becomes

$$\begin{aligned} D(x) &= \mathbb{P}\{\text{cluster length} \le x\} \\ &= \begin{cases} \exp\left(-m + \int_0^x D(x-u)\,\gamma(u)\,du\right) & x \ge 0 \\ 0 & x < 0 \end{cases} \end{aligned}$$

As a final calculation we turn to the moment measures. So using claim 19 and the discussion after 20

$$\begin{aligned} M(A) &= \int_S \int_A \nu(\delta_0 + \gamma + \gamma * \gamma + \dots)(x+s)\,dsdx \\ &= \int_A \nu(1 + m + m^2 + \dots)\,dx = |A|\nu/(1-m) \end{aligned}$$

So the first moment measure is the Lesbegue measure times $\nu/(1-m)$. (The fact the equilibrium Hawkes process is stationary implies the intensity must be a multiple of the Lesbegue measure). Its possible to work through an expression for the second moment, but we'll wait for the next set of notes on Barlett spectra to do this.

[DVJ03] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume I*. Springer Science & Business Media, 2003.

[DVJ07] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II*. Springer Science & Business Media, 2007.

[Fel66] William Feller. *An introduction to probability theory and its applications. Volume II*. John Wiley & Sons, 1966.

[HO74] Alan G Hawkes and David Oakes. A cluster process representation of a self-exciting process. *Journal of Applied Probability*, 11(3):493–503, 1974.

[Kal17] Olav Kallenberg. *Random Measures, Theory and Applications*, volume 77 of *Probability Theory and Stochastic Modelling*. Springer, 2017.

[Lew69] PAW Lewis. Asymptotic properties and equilibrium conditions for branching poisson processes. *Journal of Applied Probability*, 6(2):355–371, 1969.

[LP17] Günter Last and Mathew Penrose. *Lectures on the Poisson process*. Cambridge University Press, 2017.

[Wes71] M Westcott. On existence and mixing results for cluster point processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 290–300, 1971.

[Wes72] M Westcott. The probability generating functional. *Journal of the Australian Mathematical Society*, 14(4):448–466, 1972.

[Wil61] David Williams. *Probability with Martingales*. Cambridge University Press, 1961.