

# Applied Stochastic Processes

---

Lecture notes for Math 607

MATH 607 • Applied Stochastic Processes  
Fall 2018 — Dr Peter Ralph

Printed on 2 November 2018.

DEPARTMENT OF MATHEMATICS  
FENTON HALL  
UNIVERSITY OF OREGON  
EUGENE, OR 97403-1222 USA

ABSTRACT. These are notes.

2010 *Mathematics Subject Classification.* 60G05.

*Key words and phrases.* stochastic calculus.

## Part I

# Applied Stochastic Processes

### Contents

---

§ 1	A Gaussian Process on $L^2(\mathbb{R})$ . . . . .	3
§ 2	A Different Construction of $Z(f)$ . . . . .	5
§ 3	The Brownian Bridge . . . . .	6
§ 4	An Application . . . . .	6
§ 5	Conditional Distributions . . . . .	6
§ 6	Gaussian Process Regression . . . . .	7
§ 7	Another Application . . . . .	9
§ 8	Motivation . . . . .	12
§ 9	PPP Properties . . . . .	13
§ 10	Examples . . . . .	13
	10.1) Wald's Equation . . . . .	15



## § 1 | A Gaussian Process on $L^2(\mathbb{R})$

We will construct a process on  $L^2(\mathbb{R})$  by taking a limit of a simpler process on the space of sequences

$$\ell^2(\mathbb{R}) := \left\{ (a_k)_{k \in \mathbb{Z}} : \sum_{k \in \mathbb{Z}} |a_k|^2 < \infty, a_k \in \mathbb{R} \right\}.$$

Given a set of independent random variables  $\{X_k\}_{k \in \mathbb{Z}}$  satisfying  $X_k \sim N(0, 1)$  (think of each  $X_k$  as “noise” at the integer  $k$ ), define  $Z(a) := \sum_{k \in \mathbb{Z}} a_k X_k$  for  $a = (a_k) \in \ell^2(\mathbb{R})$ . The collection  $\{Z(a)\}_{a \in \ell^2(\mathbb{R})}$  is a Gaussian process that is centered (ie,  $\mathbb{E}[Z(a)] = 0$ ) and satisfies

$$\text{var}[Z(a)] = \text{cov} \left[ \sum_{k \in \mathbb{Z}} a_k X_k, \sum_{l \in \mathbb{Z}} a_l X_l \right] = \sum_{k, l \in \mathbb{Z}} a_k a_l \text{cov}[X_k, X_l] = \sum_{k \in \mathbb{Z}} a_k^2$$

for each  $a \in \ell^2(\mathbb{R})$ .

Now recall  $L^2(\mathbb{R}) := \{f: \mathbb{R} \rightarrow \mathbb{R} : \int_{-\infty}^{\infty} |f(x)|^2 dx < \infty\}$ . Given  $f \in L^2(\mathbb{R})$ , define a sequence of random variables  $(Z^{(N)}(f))_{N \in \mathbb{N}}$  by

$$Z^{(N)}(f) := \frac{1}{\sqrt{N}} \sum_{k \in \mathbb{Z}} f\left(\frac{k}{N}\right) X_k.$$

It can be shown that  $Z^{(N)}(f)$  converges in distribution to a random variable which we will call  $Z(f)$ . It can be also shown that the collection  $\{Z(f)\}_{f \in L^2(\mathbb{R})}$  is a centered Gaussian process satisfying  $\text{cov}[Z(f), Z(g)] = \int_{-\infty}^{\infty} f(x)g(x) dx$ . It turns out that these properties (centered Gaussian and satisfying the above covariance formula) characterize  $\{Z(f)\}_{f \in L^2(\mathbb{R})}$ , so we might as well define it this way:

**DEFINITION.** The collection  $\{Z(f)\}_{f \in L^2(\mathbb{R})}$  is **the centered Gaussian process on  $L^2(\mathbb{R})$**  satisfying

$$\text{cov}[Z(f), Z(g)] := \int_{-\infty}^{\infty} f(x)g(x) dx.$$

Note that  $Z(af + bg) = aZ(f) + bZ(g)$  for all  $a, b \in \mathbb{R}$  and all  $f, g \in L^2(\mathbb{R})$ , so  $f \mapsto Z(f)$  defines a linear map from  $L^2(\mathbb{R})$  to the space of random variables. The random variable  $Z(f)$  is called the *stochastic integral* of  $f$ , denoted

$$Z(f) =: \int_{-\infty}^{\infty} f(t) dW_t,$$

and the integrator  $dW_t$  is interpreted as “white noise” that weights the value of  $f(t)$ . Assuming it is known that  $\text{cov}[dW_s, dW_t] = \delta_s(t) ds dt$ , where  $\delta_s$  is the Dirac  $\delta$ -functional centered at  $s$ , we can

recover the covariance formula using stochastic integral notation:

$$\begin{aligned}\text{cov}[Z(f), Z(g)] &= \text{cov} \left[ \int_{-\infty}^{\infty} f(s) dW_s, \int_{-\infty}^{\infty} g(t) dW_t \right] \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(s)g(t) \text{cov}[dW_s, dW_t] = \int_{-\infty}^{\infty} f(t)g(t) dt.\end{aligned}$$

The following examples show that some stochastic processes can be defined as stochastic integrals.

**1.1 Example (Brownian motion)** If we define  $B_t := Z(\mathbf{1}_{[0,t]}) = \int_0^t dW_s$ , then the process  $\{B_t\}_{t \in [0, \infty)}$  is a Brownian motion. Indeed,

- (i)  $B_t$  is Gaussian,
- (ii)  $\text{cov}[B_s, B_t] = \int_{-\infty}^{\infty} \mathbf{1}_{[0,s]}(u) \mathbf{1}_{[0,t]}(u) du = \int_0^{\min\{s,t\}} du = \min\{s, t\}$  (and hence  $\text{var}[B_t] = t$ ),
- (iii) if  $u < v \leq s < t$  then

$$\text{cov}[B_t - B_s, B_v - B_u] = \int_{-\infty}^{\infty} \mathbf{1}_{[s,t]}(x) \mathbf{1}_{[u,v]}(x) dx = \int_{-\infty}^{\infty} \mathbf{1}_{[s,t] \cap [u,v]}(x) dx = 0.$$

•

**1.2 Example (Energy)** Define a process on  $\{Y_t\}_{t \in \mathbb{R}}$  by

$$Y_t := Z \left( e^{-(t-s)} \mathbf{1}_{(-\infty, t]}(s) \right) = \int_{-\infty}^{\infty} e^{-(t-s)} dW_s.$$

Then

- (i)  $Y_t$  is Gaussian,
- (ii) for  $s \leq t$  we have

$$\text{cov}[Y_s, Y_t] = \int_{-\infty}^{\infty} e^{-(s-u)} \mathbf{1}_{(-\infty, s]}(u)^2 e^{-(t-u)} \mathbf{1}_{(-\infty, t]}(u)^2 du = e^{-s-t} \int_{-\infty}^s e^{2u} du = \frac{e^{-(t-s)}}{2},$$

and hence  $\text{var}[Y_t] = \frac{1}{2}$ .

•

§ 2 | A Different Construction of  $Z(f)$ 

DEFINITION. Define the *Haar basis*  $\{h_{n,k}\}_{n,k \in \mathbb{N}}$  for  $L^2([0,1])$  by  $h_{0,0} \equiv 1$  and

$$h_{n,k}(t) := \begin{cases} 2^{(n-1)/2} & \text{if } 2^{-n}(2^k) \leq t < 2^{-n}(2^k+1), \\ -2^{(n-1)/2} & \text{if } 2^{-n}(2^k+1) \leq t < 2^{-n}(2^k+2), \\ 0 & \text{otherwise} \end{cases}$$

if  $n \geq 1$  and  $0 \leq k \leq 2^{n-1}$  (otherwise set  $h_{n,k} \equiv 0$ ).

It is easily verified that  $\{h_{n,k}\}$  is an orthonormal set in  $L^2([0,1])$ , ie,

$$\int_0^1 h_{n,k}(t) h_{m,\ell}(t) dt = \begin{cases} 1 & \text{if } n = m \text{ and } k = \ell, \\ 0 & \text{otherwise} \end{cases}$$

Now suppose  $\{Z_{n,k}\}$  is a collection of independent  $N(0,1)$  variables and  $f \in L^2([0,1])$  expands as

$$f(t) = \sum_{n=0}^{\infty} \sum_{k=0}^{2^{n-1}} a_{n,k} h_{n,k}(t), \quad \text{where} \quad a_{n,k} = \int_0^1 f(t) h_{n,k}(t) dt.$$

The random variable  $Z(f)$  defined by

$$Z(f) := \sum_{n=0}^{\infty} \sum_{k=0}^{2^{n-1}} a_{n,k} Z_{n,k}$$

gives a Gaussian process  $\{Z(f)\}$  for which

$$\text{cov}[Z(f), Z(g)] = \sum_{n=0}^{\infty} \sum_{k=0}^{2^{n-1}} \sum_{m=0}^{\infty} \sum_{\ell=0}^{2^{m-1}} a_{n,k} b_{m,\ell} \text{cov}[Z_{n,k}, Z_{m,\ell}] = \sum_{n=0}^{\infty} \sum_{k=0}^{2^{n-1}} a_{n,k} b_{n,k}.$$

In particular we have

$$\text{var}[Z(f)] = \sum_{n=0}^{\infty} \sum_{k=0}^{2^{n-1}} a_{n,k}^2 = \int_0^1 |f(t)|^2 dt,$$

where the latter equality is given by Parseval's identity (see page 85 of Rudin's *Real & Complex Analysis*). In the following example we see that Brownian motion can again be constructed by applying  $Z$  to a particular collection of functions.

**2.1 Example (Lévy's Construction)** Setting  $B_t = Z(\mathbf{1}_{[0,t]})$  for  $t \in [0,1]$ , we have

$$a_{n,k} = \int_0^1 \mathbf{1}_{[0,t]}(s) h_{n,k}(s) ds = \int_0^t h_{n,k}(s) ds$$

and

$$B_t = \sum_{n=0}^{\infty} \sum_{k=0}^{2^{n-1}} Z_{n,k} \int_0^t h_{n,k}(s) ds.$$

Since  $Z$  is linear (with respect to  $f$ ), for  $s \leq t$  we have

$$\text{cov}[B_t - B_s] = \text{cov}[Z(\mathbf{1}_{[0,t]}) - Z(\mathbf{1}_{[0,s]})] = \text{cov}[Z(\mathbf{1}_{[s,t]})] = \|\mathbf{1}_{[s,t]}\|_2^2 = t - s.$$

### § 3 | The Brownian Bridge

Let  $U_t = (B_t \mid B_1 = 0)$  for  $0 \leq t \leq 1$ . Since  $B_1 = Z_{0,0}$ , we have

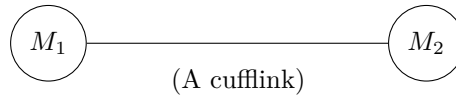
$$U_t = (B_t \mid Z_{0,0} = 0) = \sum_{n=1}^{\infty} \sum_{k=1}^{2^{n-1}} Z_{n,k} \int_0^t h_{n,k}(s) ds = B_t - tB_1.$$

Lecture 3  
03 October 2018

★ ★ ★ ★ ★ ★ ★

### § 4 | An Application

We operate a cufflink machine:



The weights of the two ends,  $M_1$  and  $M_2$  are *i.i.d*  $N(5, 10)$ . We want to only keep cufflinks that are balanced. That means  $|M_1 - M_2|$  is small.

**A.** It's easy to weigh both ends,  $M_1 + M_2$ . Does this help identify the bad cufflinks?

**No.**  $M_1 + M_2$  and  $M_1 - M_2$  are independent. This is because:

$$\text{cov}[M_1 + M_2, M_1 - M_2] = \text{var}[M_1] - \text{var}[M_2] = 0.$$

**B.** What is the distribution of  $M_1$  among cufflinks with  $M_1 + M_2 \approx 12$ ?

Let  $A = M_1 + M_2$  and  $B = M_1 - M_2$ , then  $M_1 = \frac{A+B}{2}$ . From Part A we know  $A$  and  $B$  are independent.

$$(M_1 | A = 12) \stackrel{d}{=} \frac{12 + B}{2} \sim N(6, 2)$$

This result follows from the fact that:

$$B \sim N(0, 2)$$

since  $B = M_1 - M_2$ .

### § 5 | Conditional Distributions

**5.1 Lemma** (Conditional Distributions) Let  $X_{n \times n}$  and  $Y_{m \times m}$  be jointly Gaussian with mean zero and covariance:

$$\begin{bmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{XY}^T & \Sigma_{YY} \end{bmatrix}$$

$\begin{matrix} n \times n & m \times n \\ n \times m & m \times m \end{matrix}$

NOTE:



Then,

$$\mathbb{E}[X|Y] = \Sigma_{XY} \Sigma_{YY}^{-1} Y$$

and

$$(X|Y = y) \sim N(\Sigma_{XY} \Sigma_{YY}^{-1} y, \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T)$$

Note that if  $\Sigma_{YY}$  is not invertible, we may use the Moore-Penrose inverse. That is, the above equations then remain true if the inverse used is the Moore-Penrose generalized inverse. Further, note that  $\Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T$  is the Schur complement of our covariance matrix.

this line correct?

*Proof.* Let  $A = \Sigma_{XY} \Sigma_{YY}^{-1} Y$  and  $X = A + B$ . Claim:  $B$  is independent of  $Y$ .

$$\text{cov}[X - \Sigma_{XY} \Sigma_{YY}^{-1} Y, Y] = \text{cov}[X, Y] - \Sigma_{XY} \Sigma_{YY}^{-1} \text{cov}[Y, Y] = 0.$$

Now we can say:

$$(X|Y = y) \stackrel{d}{=} \Sigma_{XY} \Sigma_{YY}^{-1} y + B$$

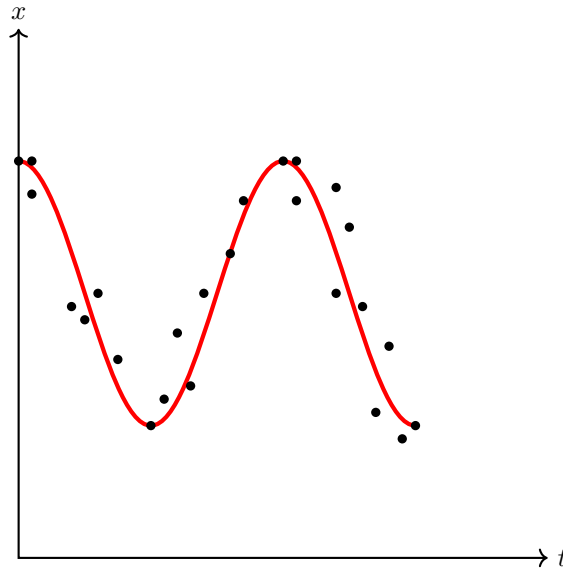
To confirm that this is the same distribution as before we need to calculate:

$$\begin{aligned} \text{cov}[B, B] &= \text{cov}[X - \Sigma_{XY} \Sigma_{YY}^{-1} Y, X - \Sigma_{XY} \Sigma_{YY}^{-1} Y] \\ &= \text{cov}[X, X] - \text{cov}[X, Y] \Sigma_{YY}^{-1} \Sigma_{XY}^T - \Sigma_{XY} \Sigma_{YY}^{-1} \text{cov}[Y, X] + \Sigma_{XY} \Sigma_{YY}^{-1} \text{cov}[Y, Y] \Sigma_{YY}^{-1} \Sigma_{XY}^T \\ &= \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T \end{aligned}$$

Thus,  $(X|Y = y) \sim N(\Sigma_{XY} \Sigma_{YY}^{-1} y, \Sigma_{XX} - \Sigma_{XY} \Sigma_{YY}^{-1} \Sigma_{XY}^T)$ .  $\square$

## § 6 | Gaussian Process Regression

Suppose we have a random curve that is drawn from a Gaussian process. We observe the value of this process at a few noisy points.



Using this information we can find out more about this curve, including confidence intervals about points.

But, first we'll look more at the Brownian Bridge. Recall (for  $0 \leq t < 1$ ),

$$U_t = (B_t | B_1 = 0) \stackrel{d}{=} B_t - tB_1$$

Check this,

$$\text{cov}[B_1, B_t - tB_1] = t - t \cdot 1 = 0$$

Recall,

$$U_t = B_t - Z_{0,0} \int_0^t h_0(s) ds = Z(\mathbf{1}_{[0,t)}) - tZ_{0,0} = \sum_{n \geq 1} \sum_{k=1}^{2^n} Z_{n,k} \int_0^t h_{n,k}(s) ds$$

What is the distribution of  $U_t - U_s$  ? (for  $s < t$ )

$$U_t - U_s = \sum_{n \geq 1} \sum_{k=1}^{2^n} Z_{n,k} \int_s^t h_{n,k}(s) ds = Z(\mathbf{1}_{[s,t)}) - (t-s)Z_{0,0}$$

So,

$$\text{var}[U_t - U_s] = \text{var}[Z(\mathbf{1}_{[s,t)})] - 2(t-s) \cdot \text{cov}[Z(\mathbf{1}_{[s,t)}), Z(\mathbf{1}_{[0,1)})] + (t-s)^2 \cdot \text{var}[Z(\mathbf{1}_{[0,1)})]$$

Note that,

$$\text{var}[Z(\mathbf{1}_{[s,t)})] = \int_0^1 \mathbf{1}_{[s,t)}(u) \cdot \mathbf{1}_{[s,t)}(u) du = t - s$$

So, we'll have

$$\text{var}[U_t - U_s] = (t-s) - 2(t-s)^2 + (t-s)^2 = (t-s)[1 - (t-s)]$$

Thus,

$$U_t - U_s \sim N\left(0, (t-s)[1 - (t-s)]\right)$$

## § 7 | Another Application

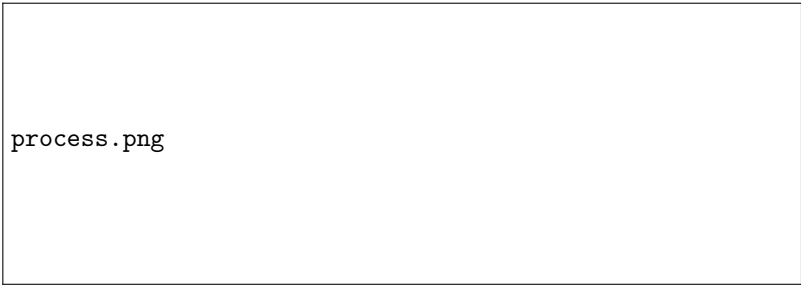
Sediment decomposition: Let  $S_k$  = (amount of sediment deposited in year  $k$ ) for  $0 \leq K \leq N = 10^4$  and we model:

$$S_{k+1} = \mu + (1 - a) \cdot [S_k - \mu] + \eta_{k+1}$$

we could also write this as:

$$[S_{k+1}|S_k \sim N((1 - a) \cdot [S_k - \mu], \sigma^2)]$$

Here  $\eta_{k+1} \sim i.i.dN(0, \sigma^2)$  where  $\sigma^2 = 10^{-4}$ , and  $a = 10^{-3}$  So,  $S_k$  with  $\mu = 0$  would look something like this:



process.png

As shown this is a process that stays distributed about a mean.

Suppose we have noisy observations of  $S_t$  for a few years Let  $Y_i$  be our measurements

$$Y_i = S_{k_i} + \varepsilon_i$$

For  $1 \leq i \leq N$  and  $\varepsilon_i \sim i.i.dN(0, \sigma_\varepsilon^2)$  Our Goal: Estimate the total amount of sediment deposited

$$S_{\text{total}} = \sum_{k=1}^N S_k$$

Let  $D_k = S_k - \mu$  we can rewrite this as:

$$D_k = \eta_k + (1 - a)D_{k-1} = \eta_k + (1 - a)\eta_{k-1} + (1 - a)^2 D_{k-2} = \sum_{j \geq 0} (1 - a)^j \eta_{k-j}$$

Thus, we know:

$$D_k \sim N\left(0, \sigma^2 \cdot \sum_{j \geq 0} (1 - a)^{2j}\right) = N\left(0, \sigma^2 \frac{1}{(2 - a)a}\right)$$

And since  $\sigma^2 = 10^{-4}$  and  $a = 10^{-3}$  we can simplify this further,  $D_k \sim N(0, \sim 0.05)$ . Let's say that  $(W_t)_{t \in \mathbb{R}}$  is a Brownian motion and  $\eta_k = W_{\frac{k+1}{N}} - W_{\frac{k}{N}} \sim N(0, \frac{1}{N} = \sigma^2)$  Then,

$$D_k = \sum_{j \geq 0} (1 - a)^j \left(W_{\frac{k-j+1}{N}} - W_{\frac{k-j}{N}}\right) \approx \sum_{j \geq 0} e^{-aj} \left(W_{\frac{k-j+1}{N}} - W_{\frac{k-j}{N}}\right) = \int_{-\infty}^t e^{-aN(t-s)} dW_s$$

Where,  $t = \frac{k}{N}$  and  $0 \leq t \leq 1$  Further simplifying we get,

$$D_k = \int_{-\infty}^t e^{-aN(t-s)} dW_s \sim N\left(0, \int_{-\infty}^t (e^{-aN(t-s)})^2 ds\right) = \frac{1}{2aN} \approx \frac{\sigma^2}{2a} + \mathcal{O}(a^2 \sigma^2)$$

Let  $U_t = \int_{-\infty}^t e^{-aN(t-s)} dW_s$  Also,  $S_{\text{total}} \simeq N\mu + \sum_{k=1}^N U_{\frac{k}{N}} \sim N \cdot (\mu + \int_0^1 U_t dt)$

New question: Let  $T = \int_0^1 U_t dt$  and  $X_i = U_{t_i} + \varepsilon_i$  where  $\varepsilon_i \sim i.i.dN(0, \sigma_\varepsilon^2)$  What is the conditional distribution of  $(T|X = X_1, \dots, X_N = X_N)$  ?

Now, we are trying to estimate the total integral, or the purple area below.



The points on the blue curve are values of  $U_t$ , and the purple area is:  $\int_0^1 U_t dt = T$  To do this we need everything inside to be jointly Gaussian. i.e. we need to use the same white noise.

Using the lemma from last lecture, we only need their covariance to calculate this area. Recall the lemma, if  $(X, Y)$  is  $N\left(0, \begin{bmatrix} \Sigma'_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{bmatrix}\right)$  then,  $(X|Y = y) \sim N(\Sigma_{XY}\Sigma_{YY}^{-1}y, \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX})$

To apply this we first need to expand  $T$

$$\begin{aligned} T &= \int_0^1 U_t dt = \int_0^1 \int_{-\infty}^t e^{-aN(t-s)} dW_s dt = \int_{-\infty}^1 \int_{\max(s, 0)}^t e^{-aN(t-s)} dt dW_s \\ &= \int_{-\infty}^1 \frac{1}{aN} e^{aN \min(0, s)} (1 - e^{-aN}) dW_s := \int_{-\infty}^1 \phi(s) dW_s \end{aligned}$$

Now, we know the following:

$$\text{var}[T] = \int_{-\infty}^1 \phi^2(s) ds$$

$$\text{var}[X_i] = \text{var}[U_{t_i}] + \sigma_\varepsilon^2 = \frac{1}{2aN} + \sigma_\varepsilon^2$$

$$\text{cov}[X_i, X_j] = \int_{-\infty}^{t_i} e^{-aN(t_i-s)} \cdot e^{-aN(t_j-s)} ds = \frac{1}{2aN} [e^{-aN(t_j-t_i)}]$$

for  $i \neq j$  and  $t_i < t_j$  Thus,

$$\text{cov}[X_i, T] = \int_{-\infty}^{t_i} e^{-aN(t_i-s)} \phi(s) ds$$

where,

$$\Sigma = \begin{bmatrix} \|\phi^2\| & 0 & \dots & \dots & 0 \\ 0 & \frac{1}{2aN} + \varepsilon^2 & 0 & \dots & 0 \\ 0 & 0 & \ddots & & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \frac{1}{2aN} + \varepsilon^2 \end{bmatrix}$$

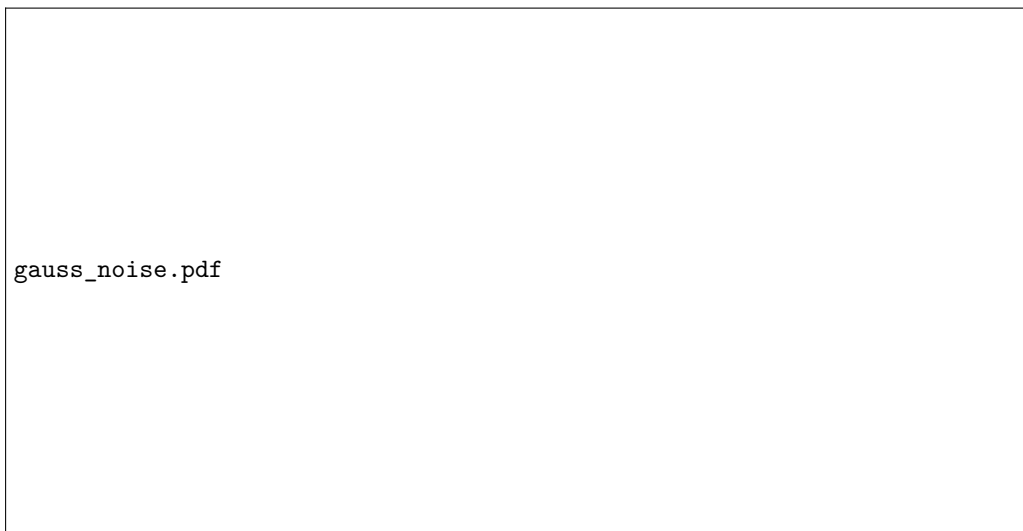


FIGURE 1. A realization of the basic noise used to construct a Gaussian process.



FIGURE 2. A realization of the basic noise used to construct a Poisson process.

Where the columns are associated with  $T$ ,  $X_1$ ,  $X_2$ , and so forth. And the rows are similarly associated with  $T$ ,  $X_1$ ,  $X_2$ , and so forth. (This means the diagonal elements are the variances of  $T$ ,  $X_1$ ,  $X_2$ ,  $\dots$ ,  $X_N$ ).

Matt

★ ★ ★ ★ ★ ★ ★

Recall a realization of the basic noise for Gaussian processes looked like that in Figure ?? . Now, arrows are either muted or (rarely) point up. See Figure ?? .

## § 8 | Motivation

Suppose in some space  $X$  we lay down a large number of LED lights, each with their own battery, with density given by a  $\sigma$ -finite measure  $\mu$ . We do this in a way so that, for each region  $A \subset X$ , we put down about  $M\mu(A)$  lights in that region, where  $M$  is some large number. Independently we turn on each light with probability  $M^{-1}$ , and leave off otherwise.

We would like to answer the following question: how many lights in  $A$  are on? To that end, let  $N(A)$  denote the number of lights on in  $A$  and compute

$$(1) \quad \mathbb{E}[N(A)] = \mathbb{E} \left[ \sum_{\text{lights in } A} \mathbb{1}_{\{\text{light on}\}} \right] = \sum_{\text{lights in } A} \mathbb{P}\{\text{light is on}\} = M\mu(A) \left( \frac{1}{M} \right) = \mu(A).$$

Thus  $\mu$  gives the expected density for the set of lights that are on in  $A$ . By construction, we know  $N(A) \sim \text{Binom}(M\mu(A), M^{-1})$ , and hence the distribution of  $N(A)$  is approximately  $\text{Pois}(\mu(A))$ . To see this, put  $L = M\mu(A)$  and observe,

$$\begin{aligned} (2) \quad \mathbb{P}\{N(A) = n\} &= \binom{L}{n} \left( \frac{1}{M} \right)^n \left( 1 - \frac{1}{M} \right)^{L-n} \\ (3) \quad &= \frac{L(L-1) \cdots (L-n+1)}{n! M^n} \left( 1 - \frac{1}{M} \right)^{L-n} \\ (4) \quad &\simeq \frac{1}{n!} \left( \frac{L}{M} \right)^n \exp \left( -\frac{L}{M} \right) + \mathcal{O} \left( \frac{1}{M} \right) \\ (5) \quad &\simeq \frac{\mu(A)^n}{n!} e^{-\mu(A)} \end{aligned}$$

This motivates the following definition.

**DEFINITION.** Let  $\mu$  be a  $\sigma$ -finite measure on some space  $X$ . A *Poisson Point Process* (PPP) on  $X$  with *mean measure* (or, *intensity*)  $\mu$  is a random point measure  $N$  such that:

(a) For any Borel set  $A \subset X$ , we have  $N(A) \in \mathbb{Z}_{\geq 0}$  and  $N(A) \sim \text{Pois}(\mu(A))$ , i.e.

$$(6) \quad \mathbb{P}\{N(A) = n\} = \frac{\mu(A)^n}{n!} e^{-\mu(A)}.$$

(b) If  $A$  and  $B$  are disjoint Borel subsets of  $X$ , then  $N(A)$  and  $N(B)$  are independent random variables.

Recall a point measure is just a measure whose mass is atomic. That is, if  $\{x_i\} \subset X$  then a point measure is of the form

$$(7) \quad \mu = \sum_i a_i \delta_{x_i}$$

where  $\delta_x$  is the unit point mass at  $x$ .

## § 9 | PPP Properties

It is sometimes useful to think of a PPP as a random collection of points. With this in mind, we list some important properties of  $N \sim \text{PPP}(\mu)$  on some space  $X$ :

- *Enumeration*: It is always possible to enumerate the points of  $N$ , i.e. there is a random collection of points  $\{x_i\} \subset X$  such that

$$(8) \quad N = \sum_i \delta_{x_i}.$$

- *Mean measure*: If  $f: X \rightarrow \mathbb{R}$  then

$$(9) \quad \mathbb{E} \left[ \int f(x) dN(x) \right] = \int f(x) d\mu(x).$$

Note: This is a more general property of point processes, as any point process has a mean measure. To see (??) holds without needing  $N$  to be a *Poisson* point process, let  $f$  be a simple function, i.e.

$$(10) \quad f(x) = \sum_{i=1}^n f_i \mathbf{1}_{A_i}(x), \quad \text{where} \quad X = \bigcup_i A_i, \quad A_i \cap A_j = \emptyset \text{ for } i \neq j.$$

Then we compute

$$(11) \quad \mathbb{E} \left[ \int_X f(x) dN(x) \right] = \mathbb{E} \left[ \sum_i f_i N(A_i) \right] = \sum_i f_i \mathbb{E} [N(A_i)] = \sum_i f_i \mu(A_i) = \int_X f(x) d\mu(x).$$

This can then be extended to arbitrary measurable functions through the standard limiting procedure.

- *Thinning*: Independently discard each point of  $N$  with probability  $1 - p(x)$  for a point at  $x \in X$ . The result is a PPP( $\nu$ ), where

$$(12) \quad \nu(A) = \int_A p(x) d\mu(x).$$

In other words, if  $N = \sum_i \delta_{x_i}$  and  $A_i = 1$  with probability  $p(x_i)$  and  $A_i = 0$  otherwise, then

$$(13) \quad \tilde{N} = \sum_i A_i \delta_{x_i} \sim \text{PPP}(\nu).$$

- *Additivity*: If  $N_1 \sim \text{PPP}(\mu_1)$  and  $N_2 \sim \text{PPP}(\mu_2)$  are independent on  $X$ , then  $N_1 + N_2 \sim \text{PPP}(\mu_1 + \mu_2)$ . In particular, if  $\mathbb{P}\{X = n\} = \frac{\lambda^n}{n!} e^{-\lambda}$  and  $\mathbb{P}\{Y = n\} = \frac{\nu^n}{n!} e^{-\nu}$  are independent, then

$$(14) \quad \mathbb{P}\{X + Y = n\} = \frac{(\lambda + \nu)^n}{n!} e^{-(\lambda + \nu)}.$$

- *Labeling*: For each point in a PPP, associate an independent label from a space  $Y$  according to some probability distribution  $\nu$ . Let  $N = \sum_i \delta_{x_i}$  for  $\{x_i\} \subset X$  and let  $G_1, G_2, \dots \in Y$  be iid with density  $\nu$ . Then

$$(15) \quad \bar{N} := \sum_i \delta_{(x_i, G_i)} \sim \text{PPP}(\mu \times \nu)$$

on  $X \times Y$ .

## § 10 | Examples

Henceforth, let  $\lambda$  denote Lebesgue measure.

**10.1 Example** Let  $N \sim \text{PPP}(\lambda)$  on  $\mathbb{R}_{\geq 0}$ , where  $\lambda$  is Lebesgue measure. As before, we think of the points of  $N$  as ‘lights’, here positioned on the positive reals.

- (a) How far until the first light?
- (b) Suppose each light is independently either red or green with probability  $\frac{1}{2}$ . How far until the first red light?

SOLUTION. Let  $N = \sum_i \delta_{x_i}$  and put  $T = \min\{x_i\}$ . Using (??) we compute

$$(16) \quad \mathbb{P}\{T > t\} = \mathbb{P}\{N([0, t]) = 0\} = e^{-t}.$$

This solves part (a). For the colorblind readers, this also solves part (b).

Now let  $\{\tilde{x}_i\} \subset \{x_i\}$  be the (random) set of red lights and define  $\tilde{N} = \sum_i \delta_{\tilde{x}_i}$ , the point process for the red lights from  $N$ . By the thinning property (??),  $\tilde{N} \sim \text{PPP}(\frac{1}{2}\lambda)$ . Similarly define  $\tilde{T} = \min\{\tilde{x}_i\}$  and observe

$$(17) \quad \mathbb{P}\{\tilde{T} > t\} = \mathbb{P}\{\tilde{N}([0, t]) = 0\} = e^{-t/2},$$

thus (b) is solved. ◆  
•

**10.2 Example** Rain falls for 10 minutes on a large patio at a rate of  $\nu = 5000$  drops per minute per square meter. Each drop splatters to a random radius  $R$  that has an Exponential distribution, with mean 1cm, independently of the other drops. Assume the drops are 1mm thick and the set of locations of the raindrops is a PPP.

- (a) What is the mean and variance of the total amount of water falling on a square with area  $1\text{m}^2$ ?
- (b) A very small ant is running around the patio. See Figure ???. What is the chance the ant gets hit?

SOLUTION. Let  $N = \sum_i \delta_{(x_i, y_i)}$  where  $(x_i, y_i)$  is the center of the  $i$ th drop. Take  $N \sim \text{PPP}(\nu\lambda)$  and let  $M$  denote the number of drops in  $[0, 1]^2$ , so that  $M = N([0, 1]^2) \sim \text{Pois}(\nu)$ . Then the total volume  $V$  is

$$(18) \quad V = \sum_{i=1}^M \frac{\pi}{10^3} R_i^2$$

where  $R_i$  is the radius of the  $i$ th drop. Note this is a sum of random variables where the number of terms is also a random variable. Thus we use Wald’s equation (??) to obtain

$$(19) \quad \mathbb{E}[V] = \frac{\pi}{10^3} \mathbb{E}[M] \mathbb{E}[R_1^2] = \frac{\pi}{10^3} \cdot \nu \cdot \frac{2}{100^2} = \frac{2\pi}{10^7} \nu$$

The second step in (??) was obtained from the fact that an exponentially distributed random variable  $X$  with mean  $\beta^{-1}$  has higher moments given by

$$(20) \quad \mathbb{E}[X^n] = \frac{n!}{\beta^n}.$$

This is proved by an iterated application of integration by parts, and the result gives rise to

$$(21) \quad \text{var}[X^n] = \mathbb{E}[X^{2n}] - \mathbb{E}[X^n]^2 = \frac{(2n)! - (n!)^2}{\beta^{2n}}.$$

The  $n = 2$  case will turn out to be useful when computing the variance of  $V$ .



Indeed, to compute the variance we utilize the variance decomposition formula. Observe,

$$\begin{aligned}
 (22) \quad \text{var}[V] &= \mathbb{E}[\text{var}[V \mid M]] + \text{var}[\mathbb{E}[V \mid M]] \\
 (23) \quad &= \mathbb{E}\left[M \left(\frac{\pi}{10^3}\right)^2 \text{var}(R^2)\right] + \text{var}\left[M \left(\frac{\pi}{10^3}\right) \mathbb{E}[R^2]\right] \\
 (24) \quad &= \nu \left(\frac{\pi}{10^3}\right)^2 \left(\frac{20}{100^4}\right) + \nu \left(\frac{\pi}{10^3}\right)^2 \left(\frac{2}{100^2}\right)^2 \\
 (25) \quad &= \left(\frac{\pi}{10^3}\right)^2 \left(\frac{24}{100^4}\right) \nu.
 \end{aligned}$$

This solves part (a).

Now, part (b) can be solved by way of the labeling property. Here, we use the radius  $R_i$  of the  $i$ th drop to label the point  $(x_i, y_i)$ . Recall the density of an Exponential random variable with mean 0.01 is  $100 \exp(-100r) dr$ . So we define a measure  $\mu$  on  $X := \mathbb{R}^2 \times [0, \infty)$  by

$$(26) \quad \mu(A) = \int_A 100\nu \exp(-100r) dx dy dr.$$

We think of  $X$  as the (closed) upper half plane in  $\mathbb{R}^3$  where the third coordinate is a realization of  $R$ . By the labeling property (??),  $\bar{N} := \sum_i \delta_{(x_i, y_i, R_i)} \sim \text{PPP}(\mu)$  on  $X$ . For the ant to remain dry, any drop with radius  $r$  must land outside the circle of radius  $r$  centered at the ant. Viewed from the space  $X$ , we want to integrate over the cone with its tip at the ant, whose horizontal cross-section at height  $r$  is a circle of radius  $r$ . From this we compute

$$(27) \quad \mathbb{P}\{\text{ant is dry}\} = \mathbb{P}\{\bar{N}(A) = 0\} = \exp(-\mu(A)) = \exp\left(-100\pi\nu \int_0^\infty r^2 e^{-100r} dr\right) = \exp\left(-\frac{\pi\nu}{5000}\right).$$

Plugging in the given value for  $\nu$  yields  $\mathbb{P}\{\text{ant is dry}\} = \exp(-\pi) \approx 0.0432$ . The ant had better grab an umbrella! ◆

•

## 10.1

### Wald's Equation

The following is the statement of Wald's equation, taken from Wikipedia <sup>\*</sup>.

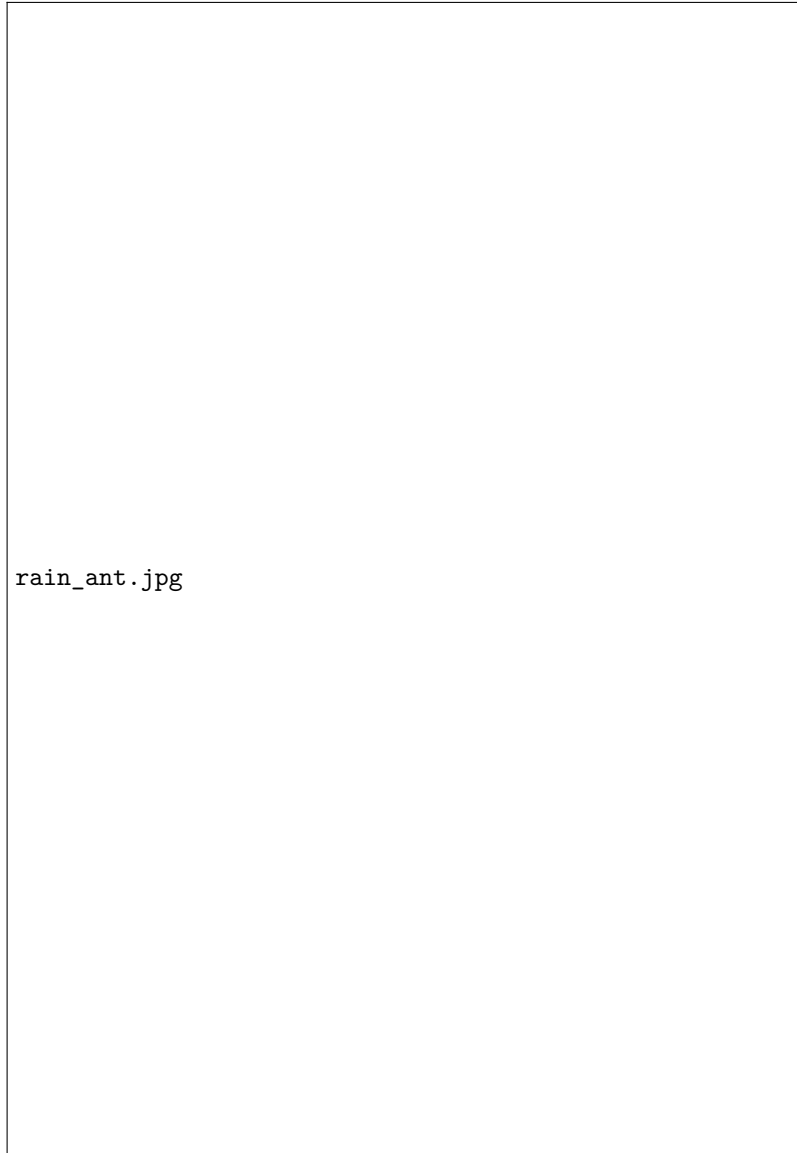
**10.3 Theorem** (Wald's Equation) Let  $(X_n)_{n \in \mathbb{N}}$  be a sequence of real-valued, independent and identically distributed random variables and let  $N$  be a nonnegative integer-valued random variable that is independent of the sequence  $(X_n)_{n \in \mathbb{N}}$ . Suppose that  $N$  and the  $X_n$  have finite expectations. Then

$$(28) \quad \mathbb{E}\left[\sum_{i=1}^N X_i\right] = \mathbb{E}[N] \mathbb{E}[X_1].$$

---

<sup>\*</sup> The proof is also on Wikipedia.

FIGURE 3. A realization of the ant from Example ???. Looks like he had an umbrella after all.



Nathan Hunter

★ ★ ★ ★ ★ ★ ★

**More rain:** a hailstorm falls on our patio for 1 hour. The rate of hailstones having mass  $x$  grams is  $\frac{1}{x}e^{-x}$  stones per hour (in total on the patio). Assume the hail is pure ice (water has density 1 gram per cubic cm).

- (1) How many hailstones have mass greater than 0.01 grams fall?
- (2) How many hailstones total fall?

- (3) Pick a random stone with mass greater than 1 gram. What is the chance it weighs more than 2 grams?  
 (4) What is the total weight of all hails?  
 (5) Line up the hailstones side-by-side. What is the total length of hail?

1). Let  $N([a, b]) = \# \{\text{hailstones with mass } a \leq x \leq b\}$ . Then  $N \sim PPP(\mu)$  on  $(0, \infty)$  with  $d\mu(x) = \frac{1}{x}e^{-x}dx$ . Thus

$$N([0.01, \infty)) \sim \text{Pois} \left( \int_{0.01}^{\infty} \frac{1}{x} e^{-x} dx \right).$$

Define  $E(x) = \int_x^{\infty} \frac{1}{y} e^{-y} dy$ . Then  $\mathbb{E}[N([0.01, \infty))] = E(0.01) = 4.03793$  (from an integral table) and  $\text{var}[N([0.01, \infty))] = E(0.01) = 4.03793$ .

2). How many hailstones total fall?  $\mathbb{E}[N([\epsilon, \infty))] = E(\epsilon) \xrightarrow{\epsilon \searrow 0} \infty$  so  $\mathbb{P}\{N(0, \infty) = \infty\} = 1$ . Picture: with the patio as the horizontal axis and  $x$  as the vertical axis, there's high density of low  $x$  values along the patio.

**Property:** (*conditional uniformity*) Let  $N \sim PPP(\mu)$  on  $X$  and  $A \subset X$  with  $\mu(A) < \infty$ . Conditioned on  $N(A)$ , the points of  $N$  that fall in  $A$  are i.i.d. with distribution proportional to  $\mu$ ; i.e. if  $B \subset A$  then

$$\mathbb{P}\{N(B) = k | N(A) = n\} = \binom{n}{k} \left( \frac{\mu(B)}{\mu(A)} \right)^k \left( 1 - \frac{\mu(B)}{\mu(A)} \right)^{n-k}.$$

3). Let  $B = [2, \infty)$  and  $A = [1, \infty)$ . The mass of stones with mass greater than 1 gram has probability density  $\frac{\frac{1}{x}e^{-x}}{E(1)}$  for  $x \geq 1$ . So

$$\mathbb{P}\{\text{stone} > 2g | \text{stone} > 1g\} = \frac{E(2)}{E(1)} = 0.2228992.$$

4). Let  $W = \int_0^{\infty} x dN(x) = \sum_i x_i = \text{total weight of all hailstones}$ . We find the mean and variance of  $W$ .

$$\mathbb{E}[W] = \int_0^{\infty} x \frac{1}{x} e^{-x} dx = 1 \text{ gram}$$

since  $\mathbb{E}[\int f dN] = \int f d\mu$ .

**Lemma:**  $\text{var}[\int f(x) dN(x)] = \int f(x)^2 d\mu(x)$ .

*Proof:* Let  $C_j^{(\epsilon)}$  be a partition of  $X$  with all  $C_j^{(\epsilon)}$  be " $\epsilon$ -small", and let  $z_j^{(\epsilon)} \in C_j^{(\epsilon)}$  be the center of each  $C_j^{(\epsilon)}$ . Then

$$\int f(x) dN(x) = \sum_i f(x_i) \approx \sum_j N(C_j^{(\epsilon)}) f(z_j^{(\epsilon)})$$

so

$$\begin{aligned}
\text{var} \left[ \int f(x) dN(x) \right] &\approx \text{var} \left[ \sum_j N(C_j^{(\epsilon)}) f(z_j^{(\epsilon)}) \right] \\
&= \sum \text{var}[N(C_j^{(\epsilon)}) f(z_j^{(\epsilon)})] \\
&= \sum f(z_j^{(\epsilon)})^2 \mu(C_j^{(\epsilon)}) \\
&\xrightarrow{\epsilon \searrow 0} \int f(x)^2 d\mu(x) \quad \blacksquare
\end{aligned}$$

Thus  $\text{var}[W] = \int_0^\infty x^2 \frac{1}{x} e^{-x} dx = 1$ .

5). Since the density of water is 1 g/cm<sup>3</sup>,  $\mathbb{E}[\text{length}] = \int_0^\infty x^{1/3} \frac{1}{x} e^{-x} dx = \Gamma(\frac{1}{3})$ .

---

**Fact:** if  $X$  and  $Y$  are random variables,  $\mathbb{E}[e^{i\alpha X}] = \mathbb{E}[e^{i\alpha Y}] \forall \alpha \in \mathbb{R}$  if and only if  $X \stackrel{d}{=} Y$  (i.e.  $\mathbb{P}\{X \in A\} = \mathbb{P}\{Y \in A\}$  for all  $A$ , or  $\mathbb{E}[f(X)] = \mathbb{E}[f(Y)]$  for all  $f$ ).

The *characteristic function* for  $X$  is  $\varphi_X(\alpha) = \mathbb{E}[e^{i\alpha X}]$ ; in other words,  $\varphi_X$  is the Fourier transform of the density function  $f_X$  of  $X$ .

Characteristic functions are convenient for calculating moments, but *not* probabilities.

**Lemma:** Let  $\psi(\alpha) = \log \varphi_X(\alpha)$  be the *cumulant generating function*. Then

$$\begin{aligned}
\frac{d}{d\alpha} \psi(0) &= i \mathbb{E}[X] \\
\frac{d^2}{d\alpha^2} \psi(0) &= -\text{var}[X]
\end{aligned}$$

*Proof:* (1)

$$\begin{aligned}
\frac{d}{d\alpha} \mathbb{E}[e^{i\alpha X}] &= \mathbb{E} \left[ \frac{d}{d\alpha} e^{i\alpha X} \right] \\
&= \mathbb{E}[iX e^{i\alpha X}]
\end{aligned}$$

$$\text{so } \frac{d}{d\alpha} \psi(0) = \frac{\mathbb{E}[iX e^{i0X}]}{\mathbb{E}[e^{i0X}]} = i\mathbb{E}[X].$$

(2) Similarly:

$$\begin{aligned}
\frac{d^2}{d\alpha^2} \psi(\alpha) &= \frac{\mathbb{E}[-X^2 e^{i\alpha X}] - \mathbb{E}[iX e^{i\alpha X}]^2}{\mathbb{E}[e^{i\alpha X}]^2} \\
\frac{d^2}{d\alpha^2} \psi(0) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \text{var}[X]
\end{aligned}$$

as desired.  $\blacksquare$

**Characteristic functions:** Let  $N \sim PPP$  on  $X$  with intensity  $\mu$  and let  $f : X \rightarrow \mathbb{R}$ . Then

$$\mathbb{E} \left[ \exp \left( i\alpha \int f(x) dN(x) \right) \right] = \exp \left( \int_X (e^{i\alpha f(x)} - 1) \mu(dx) \right).$$

*Lemma:* If  $Z \sim \text{Pois}(\gamma)$ , then

$$\mathbb{E}[e^{i\alpha Z}] = \sum_{n \geq 0} e^{i\alpha n} e^{-\gamma} \frac{\gamma^n}{n!} = e^{-\gamma} \sum_{n \geq 0} (\gamma e^{i\alpha})^n / n! = \exp(\gamma(e^{i\alpha} - 1))$$

*Proof (theorem):* Let  $f$  be piecewise constant  $f(x) = f_i$  for  $x \in A_i$  with  $\sqcup_i A_i = X$ . Then  $\int f(x) dN(x) = \sum_i N(A_i) f_i$  (and the  $N(A_i)$  are all independent) so

$$\begin{aligned} \mathbb{E}[e^{i\alpha \int f dN}] &= \mathbb{E} \left[ \prod_j e^{i\alpha N(A_j) f_j} \right] \\ &= \prod_j \mathbb{E}[e^{i\alpha N(A_j) f_j}] \\ &= \prod_j \exp(\mu(A_j)(e^{i\alpha f_j} - 1)) \\ &= \exp \left( \int_X (e^{i\alpha f(x)} - 1) \mu(dx) \right) \quad \blacksquare \end{aligned}$$

*Corollary:*  $\mathbb{E}[\int f dN] = \int f d\mu$  and  $\text{var}[\int f dN] = \int f^2 d\mu$ .

Ex: **Cauchy process:** Let  $N \sim PPP$  on  $[0, \infty) \times (\mathbb{R} \setminus \{0\})$  with mean measure  $\mu(dt, dx) = \frac{dt dx}{|x|^2}$ . Picture: with  $t$  as the horizontal and  $x$  as the vertical axes, higher density of points near the  $t$ -axis.

Let  $C_t = \int_0^t \int_{\mathbb{R}} x dN(s, x)$  = sum of  $x$ -coordinates of points in  $[0, t] \times \mathbb{R}$ .

Note:

$$\begin{aligned} \mathbb{P}\{\text{no jumps in } [t, t + \epsilon)\} &= \mathbb{P}\{C_s = C_t : s \in [t, t + \epsilon)\} \\ &= \mathbb{P}\{N([t, t + \epsilon) \times \mathbb{R}) = 0\} \\ &= \exp \left( -\epsilon \int \frac{1}{|x|^2} dx \right) = 0 \end{aligned}$$

Also:

$$\mathbb{P}\{\text{no jumps bigger than } \delta \text{ in } [t, t + \epsilon)\} = \exp \left( -2\epsilon \int_{\delta}^{\infty} \frac{dx}{x^2} \right) = e^{-2\epsilon/\delta}.$$

What is the distribution of  $C_t$ ?

$$\begin{aligned} \mathbb{E}[e^{i\alpha C_t}] &= \exp \left( \int_0^t \int_{-\infty}^{\infty} (e^{i\alpha x} - 1) \frac{1}{|x|^2} dx dt \right) \\ &= \exp(-t|\alpha|) \\ &= \int_{-\infty}^{\infty} e^{iz\alpha} \frac{dz}{\pi t(1 + (z/t)^2)} \end{aligned}$$

i.e.  $C_t \sim \text{Cauchy}(t)$  = probability density  $\frac{1}{\pi t(1 + (z/t)^2)}$ .

An interesting property:  $C_n = C_1 + (C_2 - C_1) + \dots + (C_n - C_{n-1}) = n$  i.i.d.  $\sim C_1$ . Thus  $\frac{1}{n}C_n \stackrel{d}{=} C_1$  and  $\text{var}[\frac{1}{n}C_n] = \frac{1}{n}\text{var}[C_1]$ .

Helen

★ ★ ★ ★ ★ ★ ★

## § 11 | Examples

**11.1 Example** Motivating example: Moments of the Poisson Let  $(N_t)_{t \geq 0}$  be a Poisson counting process where  $N \sim \text{PPP}(\text{constant rate } 1)$  on  $[0, \infty)$ . The fact that  $(N_t)_{t \geq 0}$  is a Poisson counting process means that  $N_t \in \mathbb{Z}$  and  $N_t$  “jumps by 1 unit at rate 1”, ie.  $N_t = \bar{N}([0, t]) + N_0$ . Note that writing  $N \sim \text{PPP}(\text{constant rate } 1)$  is the same as writing  $N \sim \text{PPP}(\lambda)$ , where  $\lambda$  is the Lebesgue measure.

We want to answer the question: What is  $\mathbb{E}[N_t^k]$ ? First we define

$$\mathbb{E}^x[f(N_t)] := \mathbb{E}[f(N_t)|N_0 = x].$$

We see that

$$\begin{aligned} \mathbb{E}^x[f(N_t)] &= f(x) + \int_0^t \frac{d}{ds} \mathbb{E}^x[f(N_s)] ds = f(x) + \int_0^t \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{E}^x[f(N_{s+\epsilon}) - f(N_s)] ds \\ &= f(x) + \int_0^t \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{E}^x[\mathbb{E}^s[f(N_{s+\epsilon}) - f(N_s)|N_s]] ds \end{aligned}$$

This motivates the following definition.

DEFINITION. Let

$$(Gf)(y) := \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{E}^y[f(N_\epsilon) - f(y)].$$

The function  $G$  is called the **generator** for  $N_t$ .

Since  $N_\epsilon - N_0 \sim \text{Poisson}(\epsilon)$ ,

$$\begin{aligned} Gf(x) &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{E}^x[f(N_\epsilon) - f(x)] \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{E}[f(N_\epsilon) - f(x)|N_0 = x] \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} (\mathbb{P}\{N_\epsilon - N_0 = 0\}(f(x) - f(x)) + \mathbb{P}\{N_\epsilon - N_0 = 1\}(f(x+1) - f(x)) \\ &\quad + \mathbb{P}\{N_\epsilon - N_0 \geq 2\} \mathbb{E}[f(N_\epsilon) - f(x)|N_\epsilon - N_0 \geq 2]) \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left( e^{-\epsilon} \cdot \epsilon \cdot (f(x+1) - f(x)) + \sum_{n \geq 2} \frac{e^{-\epsilon} \epsilon^n}{n!} \mathbb{E}[f(N_\epsilon) - f(x)|N_\epsilon - N_0 \geq 2] \right) \\ &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} (e^{-\epsilon} \cdot \epsilon \cdot (f(x+1) - f(x)) + \mathcal{O}(\epsilon^2)) \\ &= f(x+1) - f(x) \end{aligned}$$

So  $Gf(x) = f(x+1) - f(x)$  is the generator for the constant rate 1 Poisson process.

By the Markov property,

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{E}^x[\mathbb{E}^s[f(N_{s+\epsilon}) - f(N_s)|N_s]] = \mathbb{E}^x[Gf(N_s)].$$

It follows from this and (??) that

$$\mathbb{E}^x[f(N_t)] = f(x) + \int_0^t \mathbb{E}^x[f(N_s + 1) - f(N_s)] ds$$

Next we need the following notation. Define  $(x)_m := x(x-1)\cdots(x-m+1)$  and observe that

$$\begin{aligned} (x+1)_m - (x)_m &= (x+1)x(x-1)\cdots(x-m+2) - x(x-1)\cdots(x-m+1) \\ &= ((x+1) - (x-m+1))(x)_{m-1} \\ &= m(x)_{m-1}. \end{aligned}$$

Now by (??),

$$\mathbb{E}^0[(N_t)_m] = 0 + \int_0^t \mathbb{E}^0[(N_s+1)_m - (N_s)_m] ds = 0 + \int_0^t m \mathbb{E}^0[(N_s)_{m-1}] ds$$

and  $\mathbb{E}^0[(N_t)_0] = 1$ .

Let

$$g_m(t) := \int_0^t m \mathbb{E}^0[(N_s)_{m-1}] ds$$

and let  $g_0(t) = 1$ . Then  $g_m(t) = \int_0^t m g_{m-1}(s) ds$ , so

$$g_1(t) = \mathbb{E}^0[N_t] = \int_0^t 1 ds = t$$

$$g_2(t) = \mathbb{E}^0[N_t(N_t-1)] = \int_0^t 2s ds = t^2$$

and by induction,

$$\mathbb{E}^0[(N_t)_m] = g_m(t) = t^m$$

•

**11.2 Example** What is the generator for Brownian motion? Let  $(B_t)_{t \geq 0}$  be a Brownian motion, ie.  $B_{t+s} - B_t \sim \text{Normal}(0, s)$ . Then

$$\begin{aligned} \mathbb{E}^x[f(B_\epsilon)] &\approx \mathbb{E}^x[f(x) + f'(x)(B_\epsilon - x) + \frac{1}{2}f''(x)(B_\epsilon - x)^2 + \cdots] \\ &= f(x) + f'(x) \cdot 0 + \frac{1}{2}f''(x) \cdot \epsilon + \mathcal{O}(\epsilon^{3/2}) \end{aligned}$$

So

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{E}^x[f(B_\epsilon) - f(x)] = \frac{1}{2}f''(x)$$

and thus  $Gf(x) = \frac{1}{2}f''(x)$  is the generator for standard Brownian motion.

•

**11.3 Example** How can we make  $N_t$  into Brownian motion? We know that:

- (1) By the Central Limit Theorem, adding up independent noise, centering and scaling gets you the Gaussian distribution.
- (2)  $N_t$  is the number of points in a PPP on  $[0, t]$ , and so is a sum of a bunch of independent things.
- (3) Brownian motion started at zero has  $\mathbb{E}[B_t] = 0$ ,  $\mathbb{E}[B_t^2] = t$ .

$N_t$  does not have enough noise in each interval to be Brownian. So consider  $N_{Mt}$ . Since  $\mathbb{E}[N_{Mt}] = Mt$ , we can subtract  $Mt$ , and then divide by  $\sqrt{M}$  to get the correct variance. Thus we define

$$X_t^{(M)} = \frac{N_{Mt} - Mt}{\sqrt{M}}$$

Let  $G_M$  denote the generator of  $X_t^{(M)}$ . Then

$$G_M f(x) = M \left( f \left( x + \frac{1}{\sqrt{M}} \right) - f(x) \right) - \frac{1}{\sqrt{M}} f'(x).$$

This is discussed in more detail on Day 15. As  $M \rightarrow \infty$ ,  $G_M f(x) \rightarrow \frac{1}{2} f''(x)$ . Therefore,

$$(X^{(M)})_{t \geq 0} \xrightarrow[M \rightarrow \infty]{d} (B_t)_{t \geq 0}$$

•



## § 12 | Theory

Let  $(X_t)_{t \geq 0}$  be a time-homogeneous Markov process on a locally compact, separable metric space  $S$ , and define  $C_0 := C_0(S)$  to be the set of all continuous functions  $f: S \rightarrow \mathbb{R}$  vanishing at infinity, ie, given  $\epsilon > 0$ , there is a compact  $K \subset S$  such that  $f(x) < \epsilon$  for all  $x \in K$ . Note that  $C_0$  is a Banach space with the uniform norm:  $\|f\|_\infty := \sup_{x \in S} |f(x)|$ . Assume  $\mathbb{P}(\{X_t \in S\}) = 1$  for all  $t$ .

DEFINITION. Define the transition semigroup  $(P_t)_{t \geq 0}$  by  $(P_t f)(x) := \mathbb{E}^x[f(X_t)]$  for  $f \in C_0$ .

NOTE: The assumption

$$(X_t | X_0 = x) \xrightarrow[x \rightarrow y]{d} (X_t | X_0 = y)$$

implies  $P_t: C_0 \rightarrow C_0$ .

We have the following properties:

(1)  $P_0 = I$  since  $P_0 f(x) = \mathbb{E}^x[f(X_0)] = f(x)$ .

(2)  $P_s P_t = P_{s+t}$  since

$$P_s P_t f(x) = \mathbb{E}^x[P_t f(x_s)] = \mathbb{E}^x[\mathbb{E}^{x_s}[f(X_t)]] = \mathbb{E}^x[f(X_{t+s})] = P_{s+t} f(x).$$

(3) If we assume that

$$(X_t | X_0 = x) \xrightarrow[t \rightarrow 0]{d} x$$

for each  $x$  then  $P_t \rightarrow \text{id}$  as  $t \rightarrow 0$ .

DEFINITION. The generator of  $(X_t)_{t \geq 0}$  and/or  $(P_t)_{t \geq 0}$  is

$$G = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} (P_\epsilon - \text{id})$$

and

$$P_t = e^{tG} = \sum_{n \geq 0} \frac{t^n}{n!} G^n$$

if the above statements make sense.

NOTE: If  $(X_t)_{t \geq 0}$  satisfies (??) and (??) it is said to be **Feller**. The generator of a Feller process uniquely determines its distribution. This is clear from (??) when it makes sense. To prove this claim in general, use resolvents.

The generator has the following properties:

(1)  $G1 = 0$  since  $(P_t - \text{id})1(x) = \mathbb{E}^x[1 - 1] = 0$

(2)  $\pi$  is a stationary measure for  $(X_t)_{t \geq 0}$  if and only if

$$\int Gf(x) d\pi(x) = 0$$

for all  $f$ .

Next we consider some examples.

**12.1 Example** Let  $Gf(x) = f'(x)$ . Then

$$P_t f(x) = \sum_{n \geq 0} \frac{t^n}{n!} f^{(n)}(x) = \sum_{n \geq 0} \frac{t^n}{n!} f^{(n)}(x + t - t) = f(x + t).$$

So  $X_t = X_0 + t$ . Therefore,  $(d/dx)$  corresponds to “deterministic flow at rate 1.” •

**12.2 Example** Let  $Gf(x) = \frac{1}{2}f''(x)$ . Recall that this is the generator for Brownian motion. Then

$$P_t f(x) = \sum_{n \geq 0} \frac{2^{-n} t^n}{n!} f^{(2n)}(x).$$

Denote by  $\hat{f}(\xi)$  the Fourier transform of  $f$ . Then

$$\widehat{P_t f}(x) = \sum_{n \geq 0} \frac{2^{-n} t^n}{n!} (-\xi^2)^n \hat{f}(\xi) = e^{-\frac{t}{2}\xi^2} \hat{f}(\xi).$$

Because  $e^{-(t/2)\xi^2}$  is the Fourier transform of the Gaussian density with variance  $t$ , and the Fourier transform takes convolution to multiplication,

$$P_t f(x) = \sum_{n \geq 0} \frac{2^{-n} t^n}{n!} f^{(2n)}(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-\frac{(y-x)^2}{2t}} f(y) dy.$$

•

It turns out that if  $Gf(x) = f^{(k)}(x)$  for  $k > 2$ ,  $G$  is not the generator of a Feller process. One reason for this is that for  $k > 2$  there is no way to write a discrete approximation to  $f^{(k)}(x)$  as a sum over values of  $f$  with all coefficients except that of  $f(x)$  positive. More generally, we can appeal to the following theorem.

**12.3 Theorem** (Hille-Yosida Theorem) Let  $A$  be a linear operator on  $\mathcal{D} \subset C_0$ . Then  $A$  has closure that is the generator of a Feller process if and only if

- (i)  $\mathcal{D}$  is dense in  $C_0$ ;
- (ii) the range of  $\lambda - A$  is dense in  $C_0$  for some  $\lambda > 0$ ;
- (iii) (Positive Maximum Principle) if  $f(x) \leq f(x_0)$  for all  $x \in S$  and  $f(x_0) > 0$ , then  $Af(x_0) \leq 0$ .

NOTE: If all three conditions in the above theorem hold, then  $(\lambda - A)^{-1}$  exists for all  $\lambda > 0$ . To see this, suppose that  $\lambda - A$  is not invertible on  $C_0$  for some  $\lambda > 0$ . Then there exists  $f \in C_0$  such that  $Af = \lambda f$ . Then  $\mathbb{E}^x[f(X_t)] = e^{t\lambda} f(x)$ . But  $\mathbb{E}^x[f(X_t)]$  is bounded since  $f \in C_0$ , so we have reached a contradiction.

NOTE: If  $A = \lim_{t \rightarrow 0} \frac{1}{t}(P_t - \text{id})$ , then third condition automatically holds. For if

$$Af(x_0) = \lim_{t \rightarrow 0} \frac{1}{t}(P_t - \text{id})f(x_0) = \lim_{t \rightarrow 0} \frac{1}{t} \mathbb{E}^{x_0}[f(x_t) - f(x_0)].$$

Since  $f(x_t) - f(x_0) \leq 0$  by assumption,  $\mathbb{E}^{x_0}[f(x_t) - f(x_0)] \leq 0$ , and  $Af(x_0) \leq 0$ .