Applied Stochastic Processes

Lecture notes for Math 607

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

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

Chapter 1

Gaussian point processes and Random Fields

Maya

Lecture 1
24 September 2018

Let X, Y be random variables on a probability space and recall the following definitions:

- the **covariance** is given by
$$\mathrm{cov}[X,Y] \coloneqq \mathbb{E}[(X-\mathbb{E}[X])(Y-\mathbb{E}[Y])] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]\,;$$
- the **variance** is given by
$$\mathrm{var}[X] \coloneqq \mathrm{cov}[X,X]$$

The covariance is bilinear operator: for $a, b \in \mathbb{R}$
$$\mathrm{cov}[aX+bY,Z] = a\,\mathrm{cov}[X,Z] + b\,\mathrm{cov}[Y,Z]\,;$$
and thus the variance is 2-homogeneous,
$$\mathrm{var}[aX] = a^2\,\mathrm{var}[X]\,.$$

Motivation: Additive noise. Consider the random variables
$$X_k = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ -1 & \text{with probability } \frac{1}{2} \end{cases}$$
for $k \in \mathbb{Z}$, and suppose they are independent. A basic thing we might want to do with these values is add them up. So, let $S_{k,n}$ be the sum of the n adjacent values starting with the kth value; that is,
$$S_{k,n} = \sum_{j=k}^{k+n-1} X_j\,.$$
Note that $\mathbb{E}[X_k] = 0$ and $\mathrm{var}[X_k] = 1$, and so $\mathbb{E}[S_{k,n}] = 0$, $\mathrm{var}[S_{k,n}] = n$.

Recall the Central Limit Theorem, which essentially says “(well-enough behaved) additive noise makes Gaussian distributions". That is, adding up a bunch of small things that make the same-size contribution and rescaling yields basically a Gaussian distribution.

In our case, this says that
$$\frac{1}{\sqrt{n}} S_{k,n} \xrightarrow[d]{n\to\infty} N(0,1)\,.$$

ie. for any $a < b$,

$$\mathbb{P}\left\{a \leq \frac{1}{\sqrt{n}} S_{k,n} \leq b\right\} \xrightarrow{n\to\infty} \int_a^b \frac{1}{2\pi e^{-x^2/2}}\,dx$$

and, for "any" f,

$$\mathbb{E}\left[f\left(\frac{1}{\sqrt{n}} S_{k,n}\right)\right] \xrightarrow{n\to\infty} \mathbb{E}[f(Z)] = \int_{-\infty}^{\infty} f(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}\,dx\,, \tag{1}$$

where $Z \sim N(0,1)$.

§ 1 | Facts About Gaussian Processes

Say $X \sim N(\mu, \sigma^2)$, ie.

$$\mathbb{E}[f(X)] = \int_{-\infty}^{\infty} f(X)\frac{1}{\sqrt{2\pi\sigma^2}}e^{-(X-\mu)^2/2\sigma^2}\,dx\,.$$

Then

(1) (scaling) if $a \in \mathbb{R}$, then $aX \sim N(a\mu, a^2\sigma^2)$; and
(2) (linearity) if $X \sim N(\mu_X, \sigma_X^2)$ and $Y \sim N(\mu_Y, \sigma_Y^2)$, then $X + Y \sim N(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$.

Let X_k and $S_{k,n}$ be defined as above. We can visualize these variables with the following picture where the location of the tip of each arrow is (k, X_k).

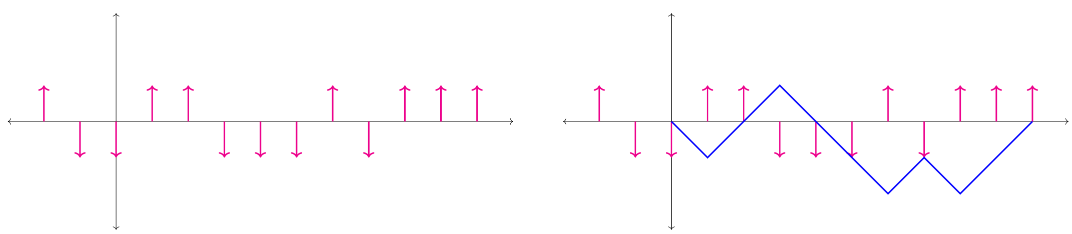


FIGURE 1. Plots of the random variables $\{X_k\}$ and $\{S_{0,n}\}$

Using this picture, if we suppose m and n are the values on the horizontal axis marked below, then $S_{0,m}$ and $S_{0,n}$ are the height of the points marked below. Additionally, $S_{m,n-m}$ is the signed distance indicated by the green arrow below, which gives the vertical distance from the first blue point to the second.

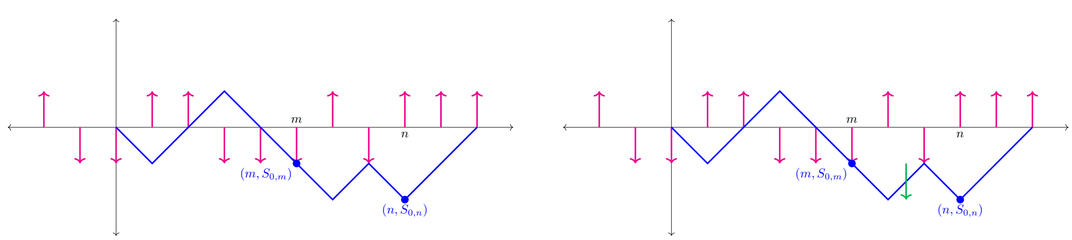


FIGURE 2. Illustrations of $S_{0,m}$, $S_{0,n}$ and $S_{m,n-m}$

This gives a visualization of the fact that for $n \geq m$,

$$S_{0,n} = S_{0,m} + S_{m,n-m} \,.$$

Similarly, we can visualize $\{S_{-2,n}\}_{n=1,2,\dots}$. The portion of the blue curve that's on the right half of the vertical axis is the same as in the pictures above, since $S_{-2,2} = 0$.

Observe that $\mathrm{var}[S_{0,n}] = n$, which can be used to show the more general statement that for $m \leq n$, $\mathrm{cov}[S_{0,m}, S_{0,n}] = m$. This comes from the following chain of equalities:

$$\begin{aligned}
\mathrm{cov}[S_{0,m}, S_{0,n}] &= \mathrm{cov}[S_{0,m}, S_{0,m} + S_{m,n-m}] \\
&= \mathrm{cov}[S_{0,m}, S_{0,m}] + \mathrm{cov}[S_{0,m}, S_{m,n-m}] && \text{since cov is bilinear} \\
&= m + 0 && \text{since } S_{0,m} \text{ and } S_{m,n-m} \text{ are independent} \\
&= m.
\end{aligned}$$

Note that Equation 1 holds in the more general case that $\mathbb{E}[Z] = 0$ and $\mathrm{var}[Z] = 1$. Add reference

§ 2 | Brownian Motion

Let $S_{k,n}$ be defined as above and

$$B_t^{(N)} = \frac{1}{\sqrt{N}} S_{0,\lfloor tN \rfloor} \,.$$

Then let

$$B_t = \lim_{N \to \infty} \frac{1}{\sqrt{N}} S_{0,\lfloor tN \rfloor}$$

Then $\{B_t\}$ is a **Brownian motion**. We can visualize Brownian motion by considering the graphs yielded by the maps $t \mapsto B^t$ and $t \mapsto B_t^{(N)}$.

For example, if we let $N = 25$, the function $t \mapsto B_t^{(N)}$ yields the following graph for $t = 0$ to $t = 10$, for a particular sequence $\{X_k\}_{k \in \mathbb{Z}_{\geq 0}}$.

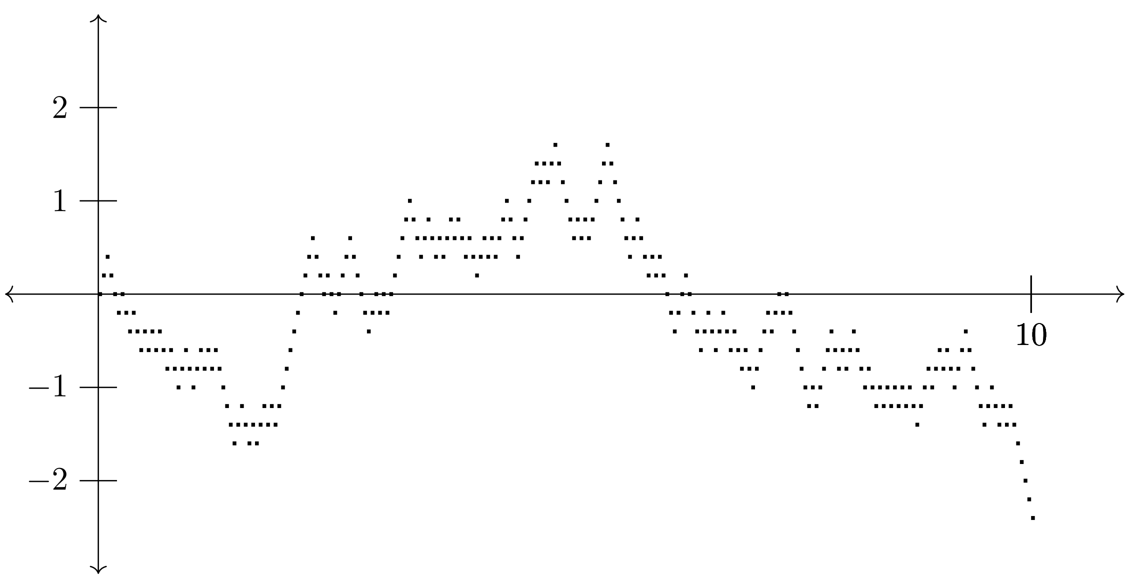


FIGURE 3. Plot of a Brownian motion with $N = 25$ and $0 \leq t \leq 10$

The Central Limit Theorem tells us that $B_t - B_s \sim N(0, t-s)$. Additionally,

$$\mathrm{var}[B_t] = \lim_{N\to\infty} \frac{\lfloor tN \rfloor}{N} = t$$

and

$$\mathrm{cov}[B_s, B_t] = s$$

for $s \le t$.

Definition. A **standard Brownian motion** is a stochastic process $\{B_t\}_{t\ge 0}$ such that

(i) $B_0 = 0$
(ii) $B_t - B_s \sim N(0, t-s)$—that is, the variance of an increment is proportional to the time difference
(iii) $B_t - B_s$ is independent of $B_v - B_u$ for $u < v \le s < t$

That is, Brownian motion is the stochastic process with independent Gaussian increments, ie. how it moves in an interval just depends on the length of that interval, not the interval itself.

1.1 Example Suppose a stream of energetic particles is absorbed by some object and the energy is slowly released from that object. Suppose at time t, the proportion of energy that remains in the object from a particle absorbed t time units ago is e^{-t}. Assuming that the object only absorbs energy at times $t = 0, 1, 2, \ldots$, let X_t be the amount of energy absorbed at time t. Let Z_n be the total energy contained in the object at time n. Then

$$Z_n = \sum_{k=0}^{\infty} e^{-k} X_{n-k}\,.$$

We can scale the x and y axes: let

$$Z_{\lfloor tN \rfloor} = \frac{1}{\sqrt{N}} \sum_{k=0}^{\infty} e^{-k/N} X_{\lfloor tN \rfloor - k}.$$

Then

$$Z_{\lfloor tN \rfloor} \xrightarrow{N\to\infty} Z_t\,.$$

•

Lecture 2
26 September 2018

A note on the previous example: If we start with Brownian motion and incorporate exponential decay, a more general central limit theorem applies (for more information, look up Lindeberg-Feller condition). Maya

Definition. Random variables $Z_1, \ldots, Z_n$ (which we can think of as a random vector of length n, $(Z_1, \ldots, Z_n)$) are *jointly Gaussian* if for any $a_1, \ldots, a_n \in \mathbb{R}$, we have

$$\sum_{k=1}^{n} a_k Z_k \sim N(m, \sigma^2)$$

for some m, σ^2. That is, these random variables are jointly Gaussian if and only if any linear combination of them is univariate Gaussian.

Definition. Suppose that the random variables $(Z_1, \ldots, Z_n)$ are jointly Gaussian. Let $Z = (Z_1, \ldots, Z_n)$. Define a vector of means

$$\mu = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix},$$

where μ_i is the mean of Z_i. Define the $n \times n$ covariance matrix $\Sigma = \{\Sigma_{i,j}\}$, where $\Sigma_{i,j} = \mathrm{cov}[Z_i, Z_j]$. Then we write $Z \sim N(\mu, \Sigma)$.

Now, suppose that we have Z, μ, and Σ as in the above definition, and let $a_1, \ldots, a_n \in \mathbb{R}$. Set

$$a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Then we know that $\sum_{k=1}^{n} a_k Z_k \sim N(m, \sigma^2)$ for some m, σ^2. A small amount of algebraic manipulation yields the following equations:

$$m = \sum_{k=1}^{n} a_k \mu_k = a^T \mu$$

and

$$\sigma^2 = \sum_{1 \le i,j \le n} a_i \Sigma_{i,j} a_j = a^T \Sigma a\,.$$

1.2 Example We can have random variables that are marginally Gaussian (that is, each of them have Gaussian distributions themselves) but that are not jointly Gaussian. Let $X \sim N(0,1)$ and $Y \sim N(0,1)$ be independent random variables. Let $Z = \mathrm{sign}(X)|Y|$. So X and Z are "almost" independent, but have the same sign as each other. Then X and Z both have Gaussian distributions, but X and Z are not jointly Gaussian. •

Definition. A *Gaussian process* on an index set T is a collection of random variables $\{X_t\}_{t \in T}$ such that for any $n \in \mathbb{N}$, for any $(t_1, \ldots, t_n) \in T^n$, $(X_{t_1}, \ldots, X_{t_n})$ is jointly Gaussian. It is *centered* if $\mathbb{E}[X_t] = 0$ for all t.

1.3 Example Let $Z \sim N(\mu, \Sigma)$ as above, then it is a Gaussian process on $\{1, \ldots, n\}$. •

1.4 Example Let $\{B_t\}_{t \ge 0}$ be a Brownian motion; then it is a Gaussian process on $[0, \infty)$. •

§ 3 | Facts about jointly Gaussian variables and Gaussian processes

(1) If $Z \in \mathbb{R}^n$ and $Z \sim N(\mu, \Sigma)$ (that is, Z is an n-dimensional multivariate Gaussian), then

$$\mathbb{E}[f(Z)] = \int_{\mathbb{R}^n} f(x) \frac{1}{\sqrt{(2\pi)^n \det(\Sigma)}} \exp\left(\frac{-(x-\mu)^T \Sigma^{-1} (x-\mu)}{2} \right) dx.$$

(2) The distribution of a Gaussian process is determined by its mean, $\mu(t) = \mathbb{E}[X_t]$ and covariance (also called covariance kernel) $\sigma^2(s,t) = \operatorname{cov}[X_s, X_t]$.

(3) Covariance kernels are positive semidefinite. Given a Gaussian process on T, where T is a measure space, we can use this to define an inner product on some subset of the set of functions $T \to \mathbb{R}$ (I don't know what the subset is that we need to take in order for this definition to make sense). Given $f, g\colon T \to \mathbb{R}$, define

$$\langle f, g\rangle_\sigma := \sum_{s\in T}\sum_{t\in T} f(s)g(t)\sigma^2(s,t)$$

and $\|f\|_\sigma^2 = \langle f, f\rangle_\sigma \geq 0$.

Example For Brownian motion, $\mathbb{E}[B_t] = 0$, $\operatorname{cov}[B_s, B_t] = \sigma^2(s,t) = \min(s,t)$. So, for $f, g\colon [0,\infty) \to \mathbb{R}$,

$$\langle f, g\rangle_\sigma = \int_0^\infty \int_0^\infty f(s)g(t)\min(s,t)\, ds\, dt\,.$$

(4) If you have a linear space V, with a symmetric positive definite inner product $\langle\cdot,\cdot\rangle$, and a countable orthonormal basis $\{\varphi_k\}_{k=1}^\infty$, then you can define a centered isomorphic Gaussian process, ie. a Gaussian process on V such that $\mathbb{E}[X_t] = 0$ for all $t \in V$ and $\operatorname{cov}[X_s, X_t] = \langle s, t\rangle$. That is, you can construct random variables and index them using V in such a way that the covariance of two random variables is equal to the inner product of their indices.

This can be done in the following way. Let $\{Z_k\}_{k=1}^\infty$ be independent, identically distributed (iid) variables each with distribution $N(0,1)$. For $t \in V$, which we can write as $t = \sum_{k=1}^\infty \langle t, \phi_k\rangle \phi_k$, define $X_t = \sum_{k=1}^\infty \langle t, \phi_k\rangle Z_k \in \mathbb{R}$. Because Z_k are independent, and each have variance 1,

$$\operatorname{cov}[Z_i, Z_j] = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

and so

$$\begin{aligned} \operatorname{cov}[X_s, X_t] &= \operatorname{cov}\left[\sum_k \langle s, \phi_k\rangle Z_k, \sum_j \langle t, \phi_j\rangle Z_j\right] \\ &= \sum_{j,k} \langle s, \phi_k\rangle \langle t, \phi_j\rangle \operatorname{cov}[Z_k, Z_j] \\ &= \sum_j \langle s, \phi_j\rangle \langle t, \phi_j\rangle \\ &= \langle s, t\rangle\,. \end{aligned}$$

1.5 Example Suppose we wanted to construct a Gaussian process with covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \frac{1}{2} & 0 \\ \frac{1}{2} & 1 & \frac{1}{2} \\ 0 & \frac{1}{2} & 1 \end{pmatrix}.$$

That is, we want to construct random variables X_{v_1}, X_{v_2}, X_{v_3}, where $v_1, v_2, v_3 \in \mathbb{R}^3$, such that for any pair (X_{v_i}, X_{v_j}), we have $\operatorname{cov}[X_{v_i}, X_{v_j}] = \langle v_i, v_j\rangle = \Sigma_{i,j}$, where $\Sigma_{i,j}$ is the entry in the ith row and jth column of our covariance matrix.

If we start with three independent random variables, each with distribution $N(0,1)$, we can use the process above to construct $X_{v_1}, X_{v_2}, X_{v_3}$. In order to do this, however, we first need to determine what v_1, v_2, and v_3 are. To do this, we will use the fact that we need $\langle v_i, v_j\rangle = \Sigma_{i,j}$.

(I wasn't able to write down anything coherent about the discussion about how to do this/why it exists, so this is a gap.)

Once we have v_1, v_2, v_3, then we can start with $Z_{v_1}, Z_{v_2}, Z_{v_3}$, independent and each with distribution $N(0,1)$, we can construct $X_{v_1}, X_{v_2}, X_{v_3}$ using the process described above. •

(5) Linear transformations of Gaussians are Gaussian: if $Z = (Z_1, \dots, Z_n) \sim N(\mu, \Sigma)$ and $A \in \mathbb{R}^{k\times n}$, then $AZ \sim N(A\mu, A\Sigma A^T)$.

1.6 Example Let $\Sigma = kk^T$ be the Cholesky decomposition of Σ, and let $Z = (Z_1, \dots, Z_n)$ independent, each with distribution $N(0,1)$. Then if we let $(X_1, X_2, X_3) = X = kZ$, then $X \sim N(0, kIk^T = \Sigma)$. •

If X, Y are jointly Gaussian, then X and Y are independent if and only if $\mathrm{cov}[X, Y] = 0$.

Joe

Lecture 3
28 September 2018

§ 4 | 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

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

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

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

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

Now recall $L^2(\mathbb{R}) := \{f \colon \mathbb{R} \to \mathbb{R} \colon \int_{-\infty}^{\infty} |f(x)|^2\, dx < \infty\}$. 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 $\mathrm{cov}[Z(f), Z(g)] = \int_{-\infty}^{\infty} f(x)g(x)\, dx$. 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

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

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 $\operatorname{cov}[dW_s, dW_t] = \delta_s(t)\, ds\, dt$, where δ_s is the Dirac δ-functional centered at s, we can recover the covariance formula using stochastic integral notation:

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

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

1.7 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) $\operatorname{cov}[B_s, B_t] = \int_{-\infty}^{\infty} \mathbf{1}_{[0,s)}(u)\mathbf{1}_{[0,t)}(u)\, du = \int_0^{\min\{s,t\}} du = \min\{s,t\}$ (and hence $\operatorname{var}[B_t] = t$),

(iii) if $u < v \le s < t$ then

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

•

1.8 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 $s \le t$ we have

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

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

•

Lecture 4
01 October 2018

Joe

§ 5 | 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

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

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

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

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

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

In particular we have

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

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.

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

$$a_{n,k} = \int_0^1 mathbf1_{[0,t)} h_{n,k}(s)\,ds = \int_0^t h_{n,k}(s)\,ds$$

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

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

•

§ 6 | 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

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

§ 7 | An Application

We operate a cufflink machine:

(A cufflink)

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:

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

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 $B = M_1 - M_2$.

§ 8 | Conditional Distributions

1.10 Lemma (Conditional Distributions) Let $\underset{n\times n}{X}$ and $\underset{m\times m}{Y}$ be jointly Gaussian with mean zero and covariance:

$$\begin{bmatrix} \underset{n\times n}{\Sigma_{XX}} & \underset{m\times n}{\Sigma_{XY}} \\ \underset{n\times m}{\Sigma_{XY}^{\mathrm{T}}} & \underset{m\times m}{\Sigma_{YY}} \end{bmatrix}$$

Note:

Then,

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

and

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

Note that if Σ_{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 $\Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{XY}^{\mathrm{T}}$ is the Schur complement of our covariance matrix.

Is this line correct?

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

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

Now we can say:

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

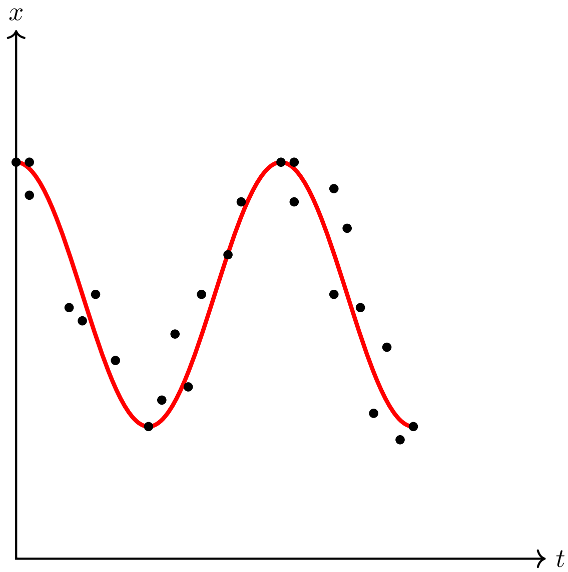
To confirm that this is the same distribution as before we need to calculate:

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

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

§ 9 | 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 $0\le t<1$),

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

Check this,

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

Recall,

$$U_t=B_t-Z_{0,0}\int_0^t h_0(s)ds=Z(\mathbf{1}_{[0,t)}))-tZ_{0,0}=\sum_{n\ge1}\sum_{k=1}^{2^n}Z_{n,k}\int_0^t h_{n,k}(s)ds$$

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

$$U_t-U_s=\sum_{n\ge1}\sum_{k=1}^{2^n}Z_{n,k}\int_s^t h_{n,k}(s)ds=Z(\mathbf{1}_{[s,t)})-(t-s)Z_{0,0}$$

So,

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

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,

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

Lecture 6
05 October 2018

Chandler

§ 10 | Another Application

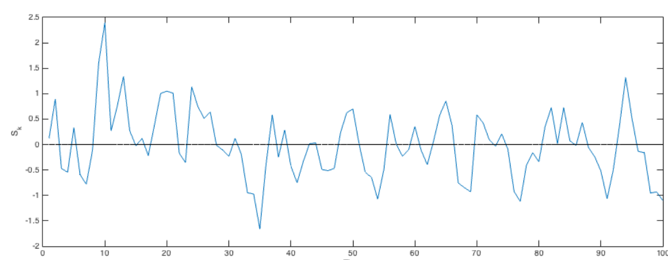
Sediment decomposition: Let $S_k =$ (amount of sediment deposited in year k) for $0 \leq K \leq N = 10^4$ and we model:

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

we could also write this as:

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

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



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 $\varepsilon_i \sim i.i.d N(0, \sigma_\varepsilon^2)$ 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:

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

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 $\eta_k = W_{\frac{k+1}{N}} - W_{\frac{k}{N}} \sim N\big(0, \frac{1}{N} = \sigma^2\big)$ Then,

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

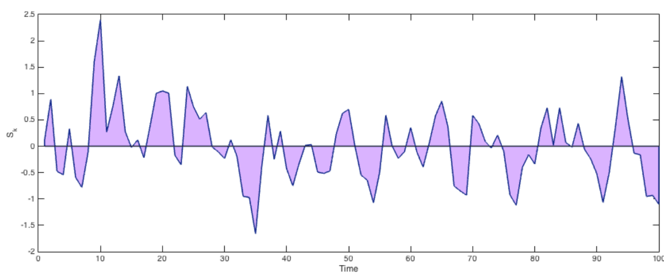
Where, $t = \frac{k}{N}$ and $0 \leq t \leq 1$ Further simplifying we get,

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

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 $\varepsilon_i \sim i.i.d N(0, \sigma_\varepsilon^2)$ 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, $(X|Y = y) \sim N(\Sigma_{XY}\Sigma_{YY}^{-1}y, \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{XY}^{\intercal})$

To apply this we first need to expand T

$$T = \int_0^1 U_t dt = \int_0^1 \int_{-\infty}^{t} e^{-aN(t-s)} dW_s dt = \int_{-\infty}^{1} \int_{\max(s,0)}^{t} e^{-aN(t-s)} dt dW_s$$

$$= \int_{-\infty}^{1} \frac{1}{aN} e^{aN\min(0,s)}\big(1 - e^{-aN}\big) dW_s \coloneqq \int_{-\infty}^{1} \phi(s) dW_s$$

Now, we know the following:

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

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

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

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

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

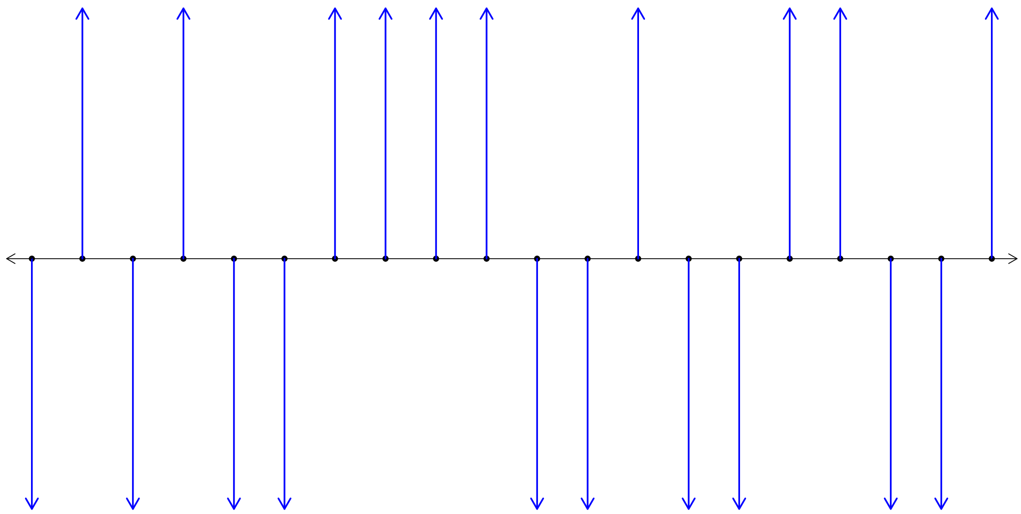


FIGURE 4. A realization of the basic noise used to construct a Gaussian process.

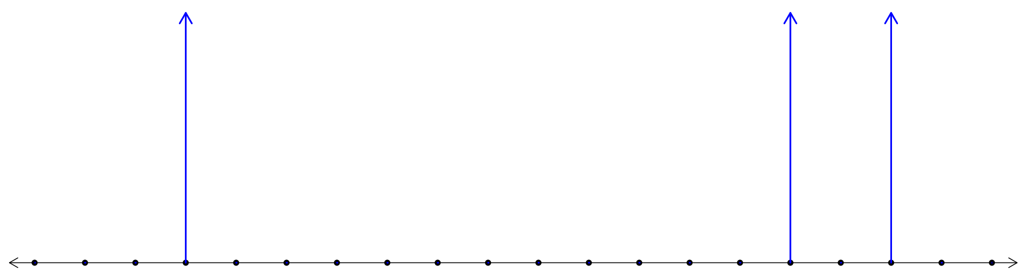


FIGURE 5. A realization of the basic noise used to construct a Poisson process.

where,

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

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).

Lecture 7
08 October 2018

Recall a realization of the basic noise for Gaussian processes looked like that in Figure 4. Now, arrows are either muted or (rarely) point up. See Figure 5. Matt

§ 11 | 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 σ-finite measure μ. 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

$$\mathbb{E}\left[N(A)\right] = \mathbb{E}\left[\sum_{\text{lights in } A} \mathbb{1}_{\{\text{light on}\}}\right] = \sum_{\text{lights in } A} \mathbb{P}\{\text{light is on}\} = M\mu(A)\left(\frac{1}{M}\right) = \mu(A). \tag{2}$$

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

$$\mathbb{P}\{N(A) = n\} = \binom{L}{n}\left(\frac{1}{M}\right)^n\left(1 - \frac{1}{M}\right)^{L-n} \tag{3}$$

$$= \frac{L(L-1)\cdots(L-n+1)}{n!M^n}\left(1 - \frac{1}{M}\right)^{L-n} \tag{4}$$

$$\simeq \frac{1}{n!}\left(\frac{L}{M}\right)^n \exp\left(-\frac{L}{M}\right) + \mathcal{O}\left(\frac{1}{M}\right) \tag{5}$$

$$\simeq \frac{\mu(A)^n}{n!}e^{-\mu(A)} \tag{6}$$

This motivates the following definition.

Definition. Let μ be a σ-finite measure on some space X. A *Poisson Point Process* (PPP) on X with *mean measure* (or, *intensity*) μ 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.

$$\mathbb{P}\{N(A) = n\} = \frac{\mu(A)^n}{n!}e^{-\mu(A)}. \tag{7}$$

(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

$$\mu = \sum_i a_i \delta_{x_i} \tag{8}$$

where δ_x is the unit point mass at x.

§ 12 | 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

$$N = \sum_i \delta_{x_i}. \tag{9}$$

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

$$\mathbb{E}\left[\int f(x)\, dN(x)\right] = \int f(x) d\mu(x). \tag{10}$$

  Note: This is a more general property of point processes, as any point process has a mean measure. To see (10) holds without needing N to be a *Poisson* point process, let f be a simple function, i.e.

$$f(x) = \sum_{i=1}^{n} f_i \mathbb{1}_{A_i}(x), \qquad \text{where} \qquad X = \bigcup_i A_i, \quad A_i \cap A_j = \varnothing \text{ for } i \neq j. \tag{11}$$

  Then we compute

$$\mathbb{E}\left[\int_X f(x)\, dN(x)\right] = \mathbb{E}\left[\sum_i f_i N(A_i)\right] = \sum_i f_i \mathbb{E}\left[N(A_i)\right] = \sum_i f_i \mu(A_i) = \int_X f(x)\, d\mu(x). \tag{12}$$

  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(ν), where

$$\nu(A) = \int_A p(x)\, d\mu(x). \tag{13}$$

  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

$$\widetilde{N} = \sum_i A_i \delta_{x_i} \sim \mathrm{PPP}(\nu). \tag{14}$$

- *Additivity:* If $N_1 \sim \mathrm{PPP}(\mu_1)$ and $N_2 \sim \mathrm{PPP}(\mu_2)$ are independent on X, then $N_1 + N_2 \sim \mathrm{PPP}(\mu_1 + \mu_2)$. 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

$$\mathbb{P}\{X + Y = n\} = \frac{(\lambda + \nu)^n}{n!} e^{-(\lambda+\nu)}. \tag{15}$$

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

$$\overline{N} := \sum_i \delta_{(x_i, G_i)} \sim \mathrm{PPP}(\mu \times \nu) \tag{16}$$

  on $X \times Y$.

§ 13 | Examples

Henceforth, let λ denote Lebesgue measure.

1.11 Example Let $N \sim \mathrm{PPP}(\lambda)$ on $\mathbb{R}_{\geq 0}$, where λ 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 (7) we compute

$$\mathbb{P}\{T > t\} = \mathbb{P}\{N([0,t]) = 0\} = e^{-t}. \tag{17}$$

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 $\widetilde{N} = \sum_i \delta_{\tilde{x}_i}$, the point process for the red lights from N. By the thinning property (13), $\widetilde{N} \sim \text{PPP}\left(\frac{1}{2}\lambda\right)$. Similarly define $\widetilde{T} = \min\{\tilde{x}_i\}$ and observe

$$\mathbb{P}\left\{\widetilde{T} > t\right\} = \mathbb{P}\left\{\widetilde{N}([0,t]) = 0\right\} = e^{-t/2}, \tag{18}$$

thus (b) is solved. ♦

•

1.12 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 1m^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 ith 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

$$V = \sum_{i=1}^{M} \frac{\pi}{10^3} R_i^2 \tag{19}$$

where R_i is the radius of the ith 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 (29) to obtain

$$\mathbb{E}[V] = \frac{\pi}{10^3}\mathbb{E}[M]\mathbb{E}[R_1^2] = \frac{\pi}{10^3} \cdot \nu \cdot \frac{2}{100^2} = \frac{2\pi}{10^7}\nu \tag{20}$$

The second step in (20) was obtained from the fact that an exponentially distributed random variable X with mean β^{-1} has higher moments given by

$$\mathbb{E}[X^n] = \frac{n!}{\beta^n}. \tag{21}$$

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

$$\text{var}[X^n] = \mathbb{E}\left[X^{2n}\right] - \mathbb{E}[X^n]^2 = \frac{(2n)! - (n!)^2}{\beta^{2n}}. \tag{22}$$

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,

$$\begin{aligned}
\operatorname{var}[V] &= \mathbb{E}\left[\operatorname{var}[V \mid M]\right] + \operatorname{var}\left[\mathbb{E}[V \mid M]\right] && (23)\\
&= \mathbb{E}\left[M\left(\frac{\pi}{10^3}\right)^2 \operatorname{var}(R^2)\right] + \operatorname{var}\left[M\left(\frac{\pi}{10^3}\right)\mathbb{E}[R^2]\right] && (24)\\
&= \nu\left(\frac{\pi}{10^3}\right)^2\left(\frac{20}{100^4}\right) + \nu\left(\frac{\pi}{10^3}\right)^2\left(\frac{2}{100^2}\right)^2 && (25)\\
&= \left(\frac{\pi}{10^3}\right)^2\left(\frac{24}{100^4}\right)\nu. && (26)
\end{aligned}$$

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 ith 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 μ on $X := \mathbb{R}^2 \times [0,\infty)$ by

$$\mu(A) = \int_A 100\nu\exp(-100r)\,dxdydr. \tag{27}$$

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 (16), $\overline{N} := \sum_i \delta_{(x_i,y_i,R_i)} \sim \mathrm{PPP}(\mu)$ 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

(28)

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

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

•

13.1 Wald's Equation

The following is the statement of Wald's equation, taken from Wikipedia [†].

1.13 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

$$\mathbb{E}\left[\sum_{i=1}^N X_i\right] = \mathbb{E}[N]\,\mathbb{E}[X_1]. \tag{29}$$

Lecture 8
12 October 2018

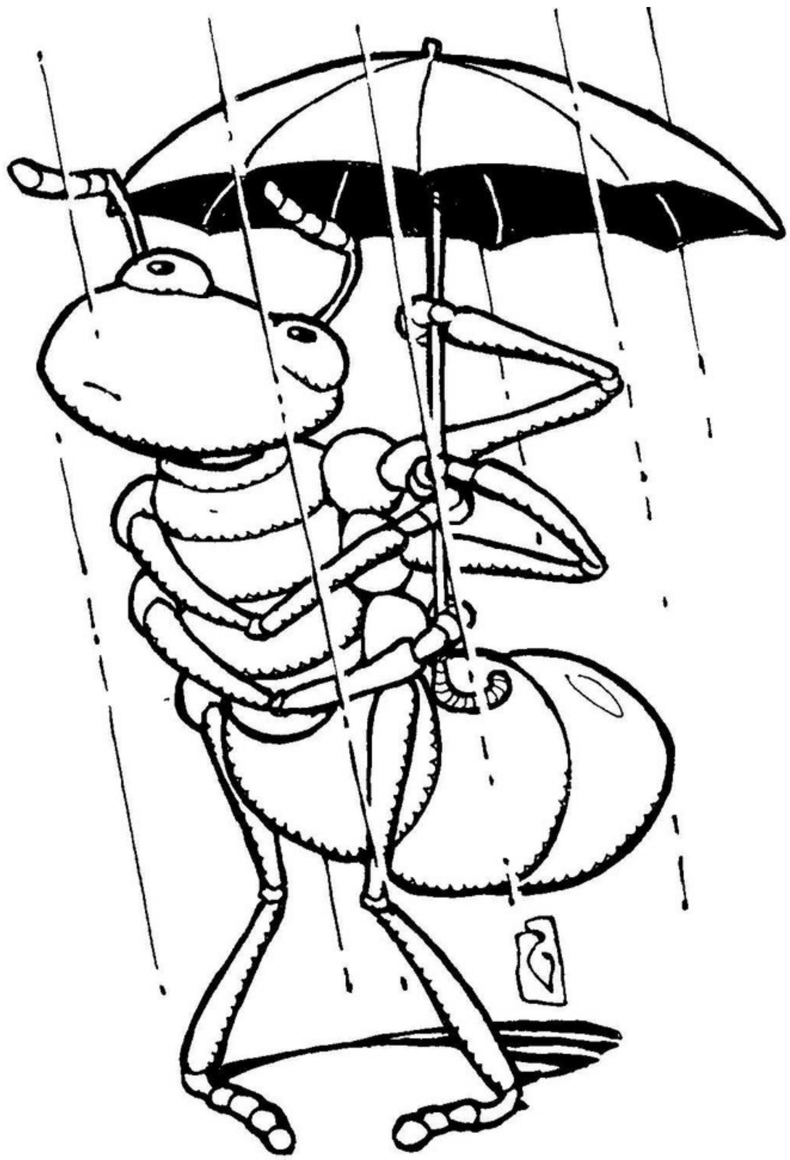
Nathan Hunte

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?

[†] The proof is also on Wikipedia.

FIGURE 6. A realization of the ant from Example 1.12. Looks like he had an umbrella after all.



(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 $N([a,b]) = \#\{$hailstones with mass $a \leq x \leq b$. Then $N \sim PPP(\mu)$ on $(0,\infty)$ with $d\mu(x) = \dfrac{1}{x}e^{-x}dx$. 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 μ; 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 $\dfrac{\frac{1}{x}e^{-x}}{E(1)}$ for $x \geq 1$. So

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

4). Let $W = \int_0^\infty x\, dN(x) = \sum_i x_i =$ total wieght 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 "ϵ-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

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

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

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

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 $X \overset{d}{=} Y$ (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, φ_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

$$\frac{d}{d\alpha}\psi(0) = i\,\mathbb{E}[X]$$

$$\frac{d^2}{d\alpha^2}\psi(0) = -\mathrm{var}[X]$$

Proof: (1)

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

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

(2) Similarly:

$$\frac{d^2}{d\alpha^2}\psi(\alpha) = \frac{\mathbb{E}[-X^2e^{i\alpha X}] - \mathbb{E}[iXe^{i\alpha X}]^2}{\mathbb{E}[ie^{i\alpha X}]^2}$$

$$\frac{d^2}{d\alpha^2}\psi(0) = \mathbb{E}[X^2] = \mathbb{E}[X]^2 = \mathrm{var}[X]$$

as desired. ■

Characteristic functions: Let $N \sim PPP$ on X with intensity μ and let $f : X \to \mathbb{R}$. 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 $Z \sim \mathrm{Pois}(\gamma)$, 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

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

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

Ex: **Cauchy process:** Let $N \sim PPP$ on $[0,\infty)\times(\mathbb{R}\backslash\{0\})$ with mean measure $\mu(dt,dx) = \dfrac{dt\,dx}{|x|^2}$. 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 $\dfrac{1}{\pi t(1+(z/t)^2)}$.

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

Lecture 9
22 October 2018

Helen

§ 14 | Examples

1.14 Example Motivating example: Moments of the Poisson Let $(N_t)_{t\geq 0}$ be a Poisson counting process where $N \sim$ PPP(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. $N_t = N([0,t]) + N_0$. Note that writing $N \sim$ PPP(constant rate 1) is the same as writing $N \sim$ PPP(λ), where λ is the Lebesgue measure.

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

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

We see that

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

This motivates the following definition.

Definition. Let

$$(Gf)(y) \coloneqq \lim_{\epsilon\to 0}\frac{1}{\epsilon}\mathbb{E}^y[f(N_\epsilon) - f(y)]\,.$$

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

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

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

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

By the Markov property,

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

It follows from this and (1.14) 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 $(x)_m \coloneqq x(x-1)\cdots(x-m+1)$ 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 (1.14),

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

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,

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

•

1.15 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

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

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

1.16 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 $M\to\infty$, $G_M f(x) \to \dfrac{1}{2}f''(x)$. Therefore,

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

•

Lecture 10
24 October 2018

Helen

§ 15 | Theory

Let $(X_t)_{t\geq 0}$ be a time-homogeneous Markov process on a locally compact, separable metric space S, and define $C_0 \coloneqq C_0(S)$ to be the set of all continuous functions $f\colon S \to \mathbb{R}$ 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: $\|f\|_\infty \coloneqq \sup_{x\in S}|f(x)|$. Assume $\mathbb{P}(\{X_t \in S\}) = 1$ for all t.

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

Note: The assumption

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

implies $P_t\colon C_0 \to C_0$.

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

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

for each x then $P_t \to \mathrm{id}$ as $t \to 0$.

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

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

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 (2) 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 $(P_t - \mathrm{id})1(x) = \mathbb{E}^x[1 - 1] = 0$
(2) π 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.

1.17 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." •

1.18 Example Let $Gf(x) = \dfrac{1}{2}f''(x)$. 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 $\widehat{f}(\xi)$ the Fourier transform of f. Then

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

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,

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

•

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.

1.19 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 $A = \lim_{t\to 0} \frac{1}{t}(P_t - \mathrm{id})$, then third condition automatically holds. For if

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

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

Lecture 11
31 October 2018

Andrew

10/31

Cell Proliferation and Cancer

The following setup is the motivation for a later theorem on Hitting Probabilities. Mutations can affect the rates of cell proliferation (splitting) and death. Suppose we counted the number of occurences of a particular mutation in each of many tissue samples. We wonder if the distribution of mutation numbers is consistent with a <u>driver mutation</u>, i.e. one with a higher than normal rate of proliferation [†] .

Model

Let X_t be the number of mutated cells at time t, so $X_t \in \mathbb{N}_{\geq 0}$, and assume that $X_0 = 1$ (so we start with one mutated cell).

Assume that each cell divides at rate λ and dies at rate μ. This means that

$$\begin{aligned} X_t &\to X_t + 1 \quad \text{at rate } \lambda X_t \\ X_t &\to X_t - 1 \quad \text{at rate } \mu X_t. \end{aligned}$$

Questions:

(1) What is the probability that the mutation "grows?" That is, for some large N, what is

$$\mathbb{P}\{X_t = N \text{ before } X_t \text{ hits } 0\} = ?$$

(2) If the mutation is not a driver but present due to an ongoing mutation of rate a, what is the expected distribution of the number of mutations per tissue sample?

Hitting Probabilities

To answer question 1 we need a bit more theory. Suppose $(X_t)_{t\geq 0}$ is a Continuous Time Markov Chain with generator G on a state space $\mathcal{X}$. For any subset $A \subset \mathcal{X}$, we define

$$\tau_A = \inf\{t \geq 0 \mid X_t \in A\},$$

which is the first time that the Markov chain lands in A. If $A = \{x\}$ we just write τ_x, which is the first time the Markov chain hits x. We now state the theorem:

> **1.20 Theorem** (Harmonic Functions) Suppose that $A, B \subset X$ are disjoint subsets with the property that
>
> $$\imath\{\tau_{A\cup B} < \infty\} = 1.$$
>
> For $x \in \mathcal{X}$ we define
>
> $$h(x) = \imath^x\{\tau_A < \tau_B\} := \imath\{\tau_A < \tau_B \mid X_0 = x\}.$$
>
> Then the function h is the unique function satisfying the following two conditions:
>
> (1) $h(x) = \begin{cases} 1 & \text{for } x \in A \\ 0 & \text{for} x \in B \end{cases}$
>
> (2) $Gh(x) = 0$ for $x \notin A \cup B$.

To prove the theorem we need two lemmas.

> **1.21 Lemma** Let $T_1, \ldots, T_n$ be independent random variables with $T_i \sim \text{Exp}(\lambda_k)$. Let $T := \min\{T_1, \ldots, T_n\}$. Then $T \sim \text{Exp}\left(\sum_{k=1}^{n} \lambda_k\right)$ and
>
> $$\imath\{T_k = T\} = \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}.$$

[†] Here, normal would mean that the rate of cell proliferation equals the rate of cell death.

Proof of Lemma 1.

$$\begin{aligned}
\mathbb{P}\{T > t\} &= \mathbb{P}\{T_k > t \quad \forall k\} \\
&= \prod_{k=1}^{n} \mathbb{P}\{T_k > t\} && \text{(by independence)} \\
&= \prod_{k=1}^{n} e^{-\lambda_k t} \\
&= e^{-t\sum_k \lambda_k}
\end{aligned}$$

which establishes the first claim. For the second claim, first note that

$$\begin{aligned}
\mathbb{P}\{T_k \in dt, T_j > t \quad \text{for } j \neq k\} &= \lambda_k e^{-\lambda_k t} \prod_{j \neq k} e^{-\lambda_j t} \\
&= \lambda_k e^{-t\sum_j \lambda_j},
\end{aligned}$$

so we now integrate over t to get

$$\mathbb{P}\{T_k = T\} = \lambda_k \int e^{-t\sum_j \lambda_j}\, dt = \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}.$$

$\square$

Lecture 12
2 November 2018

Andrew

1.22 Lemma Let $\tau_+ = \inf\{t \geq 0 \mid X_t \neq X_0\}$ be the time of the first jump. Suppose $\mathbb{P}\{\tau_+ > 0\} > 0$. Then $\tau_+ \sim \operatorname{Exp}(\lambda)$ for some $\lambda > 0$.

Proof. Let $F(t) = \mathbb{P}\{\tau_+ > t\}$ be the distribution function for τ_+. By the Markov property on X_t, we have

$$\mathbb{P}\{\tau_+ > t + s \mid \tau_+ > t\} = \mathbb{P}\{\tau_+ > s\} = F(s).$$

Unraveling the conditional probability on the left, we find

$$F(s) = \frac{\mathbb{P}\{\tau_+ > t + s \text{ and } \tau_+ > t\}}{\mathbb{P}\{\tau_+ > t\}} = \frac{F(t+s)}{F(t)}.$$

We then have that $F(t+s) = F(t)F(s)$, and the only continuous solution to this equation is $F(t) = e^{-\lambda t}$ for some $\lambda > 0$.

$\square$

To solve our problem with splitting and dying cells, we model it with a continuous time Markov chain with state space $\{1, 2, \ldots, N\}$ and generator G given by

$$G = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & \cdots & 0 \\
\mu & -\mu - \lambda & \lambda & 0 & 0 & \cdots & 0 \\
0 & 2\mu & -2\mu - 2\lambda & 2\lambda & 0 & \cdots & 0 \\
0 & 0 & 3\mu & -3\mu - 3\lambda & 3\lambda & \cdots & 0 \\
\vdots & & & & & & \vdots
\end{pmatrix}$$

Let $h(x) = \mathbb{P}^x\{\tau_N < \tau_0\}$ be the hitting probability. From the theorem on hitting probabilities, we know that

$$\begin{aligned}
h(0) &= 0 \\
h(N) &= 1 \\
Gh(x) &= 0 \quad \forall x \neq 0, N.
\end{aligned}$$

The last of these equations says

$$x\lambda(h(x+1)-h(x)) + x\mu(h(x-1)-h(x)) = 0$$

which rearranges into

$$h(x+1)-h(x) = \frac{\mu}{\lambda}(h(x)-h(x-1)).$$

Since $h(1)-h(0)=h(1)$ by our boundary condition we can use this as the seed for this recurrence relation. We get

$$h(x+1)-h(x) = \left(\frac{\mu}{\lambda}\right)^x h(1),$$

so

$$\begin{aligned} h(x) &= h(1) + \sum_{y=1}^{x-1} h(y+1)-h(y) \\ &= h(1)\left\{1 + \sum_{y=1}^{x-1}\left(\frac{\mu}{\lambda}\right)^y\right\} \\ &= h(1)\left(\frac{1-\left(\frac{\mu}{\lambda}\right)^x}{1-\frac{\mu}{\lambda}}\right). \end{aligned}$$

Next we use the other boundary condition:

$$1 = h(N) = h(1)\left(\frac{1-\left(\frac{\mu}{\lambda}\right)^N}{1-\frac{\mu}{\lambda}}\right)$$

which gives

$$h(1) = \left\{\frac{1-\left(\frac{\mu}{\lambda}\right)^N}{1-\frac{\mu}{\lambda}}\right\}^{-1},$$

and

$$h(x) = \frac{1-\left(\frac{\mu}{\lambda}\right)^x}{1-\left(\frac{\mu}{\lambda}\right)^N}.$$

This gives us the answer to our first question: what is the probability that a single mutation "grows?" This is answered by the value

$$h(1) = \frac{1-\left(\frac{\mu}{\lambda}\right)}{1-\left(\frac{\mu}{\lambda}\right)^N} \approx 1-\frac{\mu}{\lambda} \text{ for large } N.$$

Stationary Distributions

1.23 Proposition Let $(X_t)_{t\geq 0}$ be a continuous time Markov chain with generator G on state space $\mathcal{X}$. Suppose that $G^T\pi = 0$ for some probability distribution π on $\mathcal{X}$. Then

$$\mathbb{P}^\pi\{X_t = y\} = \pi(y).$$

The conditions on π above mean

$$\sum_{y\in\mathcal{X}} \pi(y) = 1, \quad \pi(y)\geq 0, \quad \sum_{y\in\mathcal{X}} \pi(y)G_{yx} = 0.$$

The notation $\mathbb{P}^\pi$ means

$$\mathbb{P}^\pi\{X_t = y\} = \sum_{x\in\mathcal{X}} \pi(x)\mathbb{P}^x\{X_t = y\}.$$

Such a distribution π is called a *stationary distribution.*

Proof. Let $f : \mathcal{X} \to \mathbb{R}$ be a function (thought of as a column vector) with $\sum_y \pi(y)f(y) < \infty$ (if we wished to generalize to a countable state space). Recall that

$$\mathbb{E}^x_{f(X_t)} = P_t f(x) = e^{tG} f(x).$$

Since $\pi G = 0$,

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

So,

$$\mathbb{E}^\pi_{f(X_t)} = \pi P_t f = \pi f = \sum_x \pi(x) f(x).$$

The statement of the proposition now follows from taking f to be the indicator of y:

$$f(x) = \begin{cases} 1 & x = y \\ 0 & x \neq y. \end{cases}$$

$\square$

Fact. If π is unique, then $P_t \to \pi$ as $t \to \infty$. That is, for all f, x,

$$\mathbb{E}^x_{f(X_t)} \to \sum_y \pi(y) f(y) \quad \text{as } t \to \infty.$$

This can also be phrased as saying that X_t converges in distribution to π.

Lecture 13
5 November 2018

Recall: We're interested in probability measures π which are solutions to $\pi G = 0$.

Eli

Assume we have a countable (or finite) state space X. $g_{i,j}$ represents the transition rate from i to j, and $g_i = \sum_{j \neq i} g_{i,j}$ is the rate of leaving state i. Suppose $\max_i g_i = g < \infty$ (uniformly bounded rates).

Let $P_{i,j} = \frac{g_{i,j}}{g}$ for $j \neq i$. Then $\sum_{j \neq i} P_{i,j} = \frac{g_i}{g} \leq 1$. Let $P_{i,i} = 1 - \frac{g_i}{g}$ so that $\sum_j P_{i,j} = 1$. P is a transition matrix with $G = g(P - I)$.

Note, $\pi G = 0$ iff $\pi P = \pi$. In continuous time, $P_t = e^{tG}$ is the operator corresponding to moving forward t units of time. $\pi P_t = \pi \sum_{k=0}^\infty \frac{t^k G^k}{k!} = \pi$ since nonzero powers of G vanish under π. πP_t is the distribution at time t, starting at π, so this says we will remain in π for all times t.

Existence: When is there a solution to $\pi P = \pi$?

Assume P is irreducible so that for all i, j, there exists t such that $P^t(i,j) > 0$ (meaning every state is reachable by any other state eventually.) Then there exists a solution to $\pi P = \pi$ but π need not be a finite measure.

Example: Consider the simple random walk on the integers with $P_{i,j} = \frac{1}{2}$ if $j = i \pm 1$ and $P_{i,j} = 0$ otherwise. The solution is $\pi = 1$.

Definition: A chain with transition matrix P is called positive recurrent if $\pi P = \pi$ has a finite measure solution π. A state i is called recurrent if $\mathbb{P}_i(\text{Return to } i) = 1$. Otherwise, the state is called transient.

If P is irreducible, then either all states are recurrent or all states are transient.

Theorem: A state i is recurrent iff $\mathbb{E}_i[\tau_i^+] < \infty$ where $\tau_i^+ = \min\{t \geq 1 : X_t = i\}$. Equivalently, $\mathbb{P}_i(\tau_i^+ < \infty) = 1$. (Obvious when the state space X is finite and irreducible.)

Example: A random walk on the non-negative integers where the state increases by one with probability p and decreases by one with probability $q = 1 - p$. (The transition from 0 to 0 takes probability q.)

$p > q$ implies the chain is not recurrent. If $q > p$, then the chain might be recurrent. For $k \neq 0$ we have $\pi(k) = \pi(k-1)p + \pi(k+1)q$ since there are two different ways to arrive at state k. As the special case, we have $\pi(0) = \pi(0)q + \pi(1)q$.

We want to solve $\pi P = \pi$. We get $\pi(1) = \frac{p}{q}\pi(0)$ and $\pi(k) = \left(\frac{p}{q}\right)^k \cdot \pi(0)$ as a solution. Thus, $\sum_k \pi(k) = \pi(0)\left(1 - \frac{p}{q}\right)$ and $\pi(0) = 1 - \frac{p}{q}$. We have a probability distribution when $\pi(k) = \left(1 - \frac{p}{q}\right)\left(\frac{p}{q}\right)^k$. Thus, positive recurrence.

Definition: A chain is aperiodic if $\gcd\{t : P^t(x,x) > 0\} = 1$. In the case of a bipartite graph, we can only return at even times, and $P^t(x,x) = 0$ if t is odd. (Not needed in continuous time.)

If the chain is positive recurrent and irreducible (with aperiodicity in the discrete time case), then there exists a unique solution π which solves $\pi P = \pi$ (equivalently $\pi G = 0$).

(Convergence) Theorem: The distribution starting at x after t steps converges to π as $t \to \infty$ in the sense that $d(P^t(x,\cdot), \pi) \to 0$ for some metric d.

Example: $d_1(P^t(x,\cdot), \pi) = \sum_y |P^t(x,y) - \pi(y)| \to 0$ and $t \to \infty$. Thus, $P^t(x,y) \to \pi(y)$ for all x.

Now take μ to be any distribution. We have $\mu P^t \to \pi$.

Application: Tilings of a square by dominoes. We want to select a tiling uniformly at random. The naive way to do this is list (enumerate) them all and then select one using a random number generator. In practice, this is very hard to do because there are so many possibilities. Another way might be to choose positions and orientations of dominoes in sequence until the whole space is filled, but not all tilings will occur will equal probability.

Alternatively, come up with a way to transition between random tilings such as rotating a square of two dominoes or modifying other such sub-tilings. We can construct a Markov chain with the uniform distribution as the stationary distribution π. We start by choosing the starting position according to any distribution μ and then proceed by running the chain for sufficiently large t. The closer μ is to be uniform, the faster the convergence, but μ could theoretically be very far from uniform.

Coupling: Let X_t be a chain started from x, and let $\tilde{Y}_t$ be a chain started from π. Let $\tau = \inf(t > 0 : X_t = \tilde{Y}_t)$. When the state space X is finite, then $\tau < \infty$. Define $Y_t = \tilde{Y}_t$ for $t \leq \tau$ and $Y_t = X_t$ for $t \geq \tau$.

Check: Y_t is a Markov chain with transition matrix P still started with π (comes from $\tilde{Y}_t$), and Y_t always has distribution π.

$$\begin{aligned}
|P^t(x,z) - \pi(z)| &= |\mathbb{P}(X_t = z) - \mathbb{P}(Y_t = z)| \\
&= |\mathbb{P}(X_t = z, \tau \leq t) + \mathbb{P}(X_t = z, \tau > t) - \mathbb{P}(Y_t = z, \tau \leq t) - \mathbb{P}(Y_t = z, \tau > t)| \\
&= |\mathbb{P}(X_t = z, \tau > t) - \mathbb{P}(Y_t = z, \tau > t)| \\
&\leq \mathbb{P}(X_t = z, \tau > t) + \mathbb{P}(Y_t = z, \tau > t) \\
&\leq 2\mathbb{P}(\tau > t)
\end{aligned}$$

Using $\mathbb{P}(\tau < \infty) = 1$, we have $2\mathbb{P}(\tau > t) \to 0$ and thus $d_1(P^t(x,\cdot), \pi) = \sum_z |P^t(x,z) - \pi(z)| \to 0$ as $t \to \infty$.

In general, solving for the stationary distribution takes work.

Definition: P (or G) is reversible with respect to π if $\pi(x)P(x,y) = \pi(y)P(y,x)$ (or $\pi(x)G(x,y) = \pi(y)G(y,x)$). These are called the detailed balance equations which say the "flow" from x to y is the same as from y to x.

Example: Weighted random walk with $P(x,y) = \frac{w_{x,y}}{w_x}$ where $w_x = \sum_z w_{x,z}$. P is reversible with $\pi(x) = w_x / \left(\sum_z w_z\right)$.

If P is reversible, then P need not be symmetric but can be diagonalized. $A(x,y) = \sqrt{\pi(x)/\pi(y)}P(x,y)$ is symmetric. By the Spectral Theorem, A has real eigenvalues $\lambda_1 > \cdots > \lambda_n$ and the eigenfunctions $\varphi_1, \ldots, \varphi_n$ form an orthonormal basis for $\mathbb{R}^n$.

$A = D_\pi^{1/2} P D_\pi^{-1/2}$ where D_π is diagonal with $D_\pi(i,i) = \pi(i)$. $f_j = D_\pi^{1/2}\varphi_j$ are eigenfunctions for P.

Define $< f, g >_\pi = \sum_x f(x)g(x)\pi(x)$, then $\{f_j\}$ are orthonormal with respect to this inner product. $P^t(x,y) = \sum_j f_j(y)\lambda_j^t f_j(x)\pi(y)$.

$$\left| \frac{P^t(x,y)}{\pi(y)} - 1 \right| \leq \sum_{j=2}^{n} |f_j(x)f_j(y)|\lambda_j^t$$

where λ_j^t's tell us how fast the convergence is.

Daniel **Homework Problem:** Spiders live at the points of a PPP on a long wire with uniform intensity of 0.1 per centimeter. Each claims the area withint 2 cm on either side, so a fly landing on the wire is attacked by all spiders within 2 cm of it.

Lecture 14
9 November 2018

(1) A fly lands on the wire independent of the spiders. What is the probability that the fly is safe?
(2) We have a PPP of flies (independent of the spiders) with uniform intensity 0.1 per cm landing on a 1 meter section of wire. Let $F_k = \#\{\text{flies attacked by } k \text{ spiders}\}$. What is $\mathbb{P}\{F_k = n\}$?
(3) *Bonus:* compute the mean and variance of the total length of wire in a 1 meter section *not* claimed by spiders.

More rain: a hailstorm falls on our patio for 1 hour. The rate of hailstones having mass x grams is $\dfrac{1}{x}e^{-x}$ 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 $N([a,b]) = \#\{\text{hailstones with mass } a \le x \le b$. Then $N \sim PPP(\mu)$ on $(0,\infty)$ with $d\mu(x) = \dfrac{1}{x}e^{-x}dx$. 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 μ; 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 $\dfrac{\frac{1}{x}e^{-x}}{E(1)}$ for $x \geq 1$. So

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

4). Let $W = \int_0^\infty x\,dN(x) = \sum_i x_i$ = total wieght 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 "ϵ-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

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

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^3, $\mathbb{E}[\text{length}] = \int_0^\infty x^{1/3}\frac{1}{x}e^{-x}\,dx = \Gamma(\frac{1}{3})$.

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 $X \overset{d}{=} Y$ (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, φ_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

$$\frac{d}{d\alpha}\psi(0) = i\,\mathbb{E}[X]$$

$$\frac{d^2}{d\alpha^2}\psi(0) = -\text{var}[X]$$

Proof: (1)

$$\frac{d}{d\alpha}\mathbb{E}[e^{i\alpha X}] = \mathbb{E}\left[\frac{d}{d\alpha}e^{i\alpha X}\right]$$
$$= \mathbb{E}[iXe^{i\alpha X}]$$

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

(2) Similarly:

$$\frac{d^2}{d\alpha^2}\psi(\alpha) = \frac{\mathbb{E}[-X^2e^{i\alpha X}] - \mathbb{E}[iXe^{i\alpha X}]^2}{\mathbb{E}[ie^{i\alpha X}]^2}$$

$$\frac{d^2}{d\alpha^2}\psi(0) = \mathbb{E}[X^2] = \mathbb{E}[X]^2 = \mathrm{var}[X]$$

as desired. ■

Characteristic functions: Let $N \sim PPP$ on X with intensity μ and let $f : X \to \mathbb{R}$. 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 $Z \sim \mathrm{Pois}(\gamma)$, 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

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

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

Ex: **Cauchy process:** Let $N \sim PPP$ on $[0,\infty)\times(\mathbb{R}\backslash\{0\})$ with mean measure $\mu(dt,dx) = \dfrac{dt\,dx}{|x|^2}$. 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:

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

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

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

Lecture 15
9 November 2018

Homework Problem: Spiders live at the points of a PPP on a long wire with uniform intensity of 0.1 per centimeter. Each claims the area withint 2 cm on either side, so a fly landing on the wire is attacked by all spiders within 2 cm of it. Daniel

(1) A fly lands on the wire independent of the spiders. What is the probability that the fly is safe?
(2) We have a PPP of flies (independent of the spiders) with uniform intensity 0.1 per cm landing on a 1 meter section of wire. Let $F_k = \#\{\text{flies attacked by } k \text{ spiders}\}$. What is $\mathbb{P}\{F_k = n\}$?
(3) *Bonus:* compute the mean and variance of the total length of wire in a 1 meter section *not* claimed by spiders.

More rain: a hailstorm falls on our patio for 1 hour. The rate of hailstones having mass x grams is $\dfrac{1}{x}e^{-x}$ 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 $d\mu(x) = \dfrac{1}{x}e^{-x}dx$. 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 μ; 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 $\dfrac{\frac{1}{x}e^{-x}}{E(1)}$ for $x \geq 1$. So

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

4). Let $W = \int_0^\infty x\, dN(x) = \sum_i x_i =$ total wieght 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 "ϵ-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

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

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^3, $\mathbb{E}[\text{length}] = \int_0^\infty x^{1/3} \frac{1}{x} e^{-x}\, dx = \Gamma(\frac{1}{3})$.

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, φ_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

$$\frac{d}{d\alpha}\psi(0) = i\,\mathbb{E}[X]$$
$$\frac{d^2}{d\alpha^2}\psi(0) = -\mathrm{var}[X]$$

Proof: (1)

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

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

(2) Similarly:

$$\frac{d^2}{d\alpha^2}\psi(\alpha) = \frac{\mathbb{E}[-X^2e^{i\alpha X}] - \mathbb{E}[iXe^{i\alpha X}]^2}{\mathbb{E}[ie^{i\alpha X}]^2}$$
$$\frac{d^2}{d\alpha^2}\psi(0) = \mathbb{E}[X^2] = \mathbb{E}[X]^2 = \mathrm{var}[X]$$

as desired. ■

Characteristic functions: Let $N \sim PPP$ on X with intensity μ and let $f : X \to \mathbb{R}$. 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 $Z \sim \mathrm{Pois}(\gamma)$, 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

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

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

Ex: **Cauchy process:** Let $N \sim PPP$ on $[0,\infty)\times(\mathbb{R}\backslash\{0\})$ with mean measure $\mu(dt,dx) = \dfrac{dt\,dx}{|x|^2}$. 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. $C_t \sim \text{Cauchy}(t) = \text{probability density}\dfrac{1}{\pi t(1+(z/t)^2)}$.

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

Martin

Lecture 16
14 November 2018

§ 16 | Levy Processes

Definition. A Levy process with drift rate α, diffusion rate σ, and jump kernel ν is a stochastic process with distribution $X_0 = 0$

$$X_t = \alpha t + \sigma B_t + \int_0^t\int_{-\infty}^\infty xN(ds,dx) \tag{30}$$

where B_t is Brownian motion, so that $\mathbb{E}[B_t] = 0$ and $\text{var}[B_t] = t$, and where N is a Poisson point process, independent from B_t, on $[0,\infty)\times\mathbb{R}$ with intensity measure $dt\nu(dx)$.

Fact: (Levy-Khinchine) Any Markov process on $\mathbb{R}$ with $X_0 = 0$ and stationary independent increments is Levy.

i.e. a Markov process such that

(1) the distribution of the increment $X_{t+h} - X_t$ depends only on h.
(2) if $a < b \le c < d$ then $(X_d - X_c)$ is independent of $(X_b - X_a)$.

Note that (30) makes sense if the jumps are absolutely summable, i.e. let $N = \sum_i \delta_{(t_i, x_i)}$. Then the "jump component" of X is

$$\begin{aligned} J_t &= \int_0^t \int_{-\infty}^{\infty} xN(ds, dx) \\ &= \sum_{i: t_i \leq t} x_i \end{aligned}$$

and we want $\sum |x_i|, \infty$.

Since jumps with $|x_i| > 1$ happen only finitely many times, absolute summability can be determined by looking at jumpes with $|x_i| < 1$. Checking the expected value of this sum we have

$$\begin{aligned} \mathbb{E}\left[\sum_{\substack{i:|x_i|<1, \\ t_i<t}} |x_i|\right] &= \mathbb{E}\left[\int_0^t \int_{-1}^1 |x| N(ds, dx)\right] \\ &= t\int_{-1}^1 |x| \nu dx \end{aligned}$$

which is finite if $\int_{-\infty}^{\infty} \min(|x|, 1)\nu(dx) < \infty$.

Properties of X_t :

(1) X has generator

(31) $$Gf(x) = \alpha f'(x) + \frac{\sigma^2}{2} f''(x) + \int_{-\infty}^{\infty} \left(f(x+y) - f(x)\right)\nu(dy)$$

(2) (Levy-Khinchine formula) The characteristic function of X_t is given by

(32) $$\phi(t, u) = \mathbb{E}\left[e^{iuX_t}\right] = e^{t\psi(u)}$$

where $\psi(u) = i\alpha u - \frac{\sigma^2}{2}u^2 + \int_{-\infty}^{\infty}\left(e^{iux} - 1\right)\nu(dx)$ since

(33) $$\mathbb{E}\left[e^{iuX_t}\right] = \mathbb{E}\left[e^{iu\alpha t} e^{iu\sigma B_t} e^{iu\int_0^t \int_{-\infty}^{\infty} xN(ds,dx)}\right]$$

1.24 Example Stable Subordinator of Index 1/2

Let $(B_t)_{t\geq 0}$ be Brownian motion and define $\tau_r = \inf\{t \geq 0 : B_t \geq r\}$. Then $(\tau_r)_{r\geq 0}$ Levy with $\alpha = 0, \sigma = 0, \nu(dx) = x^{-3/2}dx$. •

§ 17 | Levy Processes continued

Definition. A subordinator is a nondecreasing Levy process.

Let $(B_t)_{t\geq 0}$ be Brownian motion. Set $t_x = \inf\{t \geq 0 : B_t \geq x\}$. Then $(\tau_x)_{x\geq 0}$ is a nondecreasing Markov process with stationary, independent increments and so is a subordinator.

For $x, y \geq 0$, we have, setting $\tilde{B}_t = B_{\tau_x + t} - x$,

$$\begin{aligned} tau_{x+y} - \tau_x &= \inf\{t \geq \tau_x : B_t = x + y\} - \inf\{t \geq 0 : B_t = x\} \\ &= \inf\left\{t \geq 0 : \tilde{B}_t = y\right\} - 0 \\ &\overset{d}{=} \tau_y \end{aligned}$$

which implies that τ has stationary increments.

Moreover, $(\tau_{x+y})_{y\geq 0}$ only depends on $(\tilde{B}_t)_{t\geq 0}$ which is independent of $(B_s)_{0\leq s\leq \tau_x}$ so τ is Markov.

Scaling: Since $\overline{B}_s := \frac{1}{c}B_{c^2 s})_{s\geq 0} \overset{d}{=} (B_s)_{s\geq 0}$ and

$$\begin{aligned}
\tau_x &= \inf\{t \geq 0 : B_t \geq x\} \\
&\overset{d}{=} \inf\left\{t\geq 0 : \frac{1}{c}B_{c^2 t} \geq x\right\} \\
&= \frac{1}{c^2}\inf\{t\geq 0 : B_t \geq cx\} \\
&= \frac{\tau_{cx}}{c^2}
\end{aligned}$$

we have $\tau_x \overset{d}{=} x^2\tau_1$.

To find the distribution of τ we look at

$$\mathbb{P}\{\tau_1 < t\} = \mathbb{P}\left\{\sup_{0\leq s\leq t} B_s \geq 1\right\} = \mathbb{P}\{B_t \geq 1\} + \mathbb{P}\{b_t < 1, M_t \geq 1\} \tag{34}$$

where $M_t = \sup\{B_s : 0 \leq s \leq t\}$.

1.25 Lemma Reflection Principle:
$\mathbb{P}\{B_t < 1, M_t \geq 1\} = \mathbb{P}\{B_t \geq 1\}$ i.e.

$$\left(W_s = \begin{cases} B_s & \text{if } s \leq \tau_1 \\ 1-(B_s-1) & \text{if } s \geq \tau_1 \end{cases}\right) \overset{d}{=} (B_s)_{s\geq 0}$$

Combining this Lemma with (34) we have

$$\begin{aligned}
\mathbb{P}\{\tau_1 < 1\} &= 2\mathbb{P}\{B_t \geq 1\} \\
&= 2\mathbb{P}\left\{B_1 \geq \frac{1}{\sqrt{t}}\right\} \\
&= \frac{2}{\sqrt{2\pi}}\int_{1/\sqrt{t}}^{\infty} e^{-x^2/2}dx \\
&= \frac{1}{\sqrt{2\pi}}\int_0^t s^{-3/2}e^{-1/2s}ds
\end{aligned}$$

where the last equality follows from setting $s = \frac{1}{x^2}$. Thus, $\mathbb{P}\{\tau_x < t\} = \mathbb{P}\left\{x^2\tau_1 < t\right\} = \mathbb{P}\left\{\tau_1 < t/x^2\right\}$ so the density of τ_x is given by

$$\frac{1}{\sqrt{2\pi}}\frac{1}{x^2}\left(\frac{\tau}{x^2}\right)^{-3/2}e^{-x^2/2t}dt.$$

Recall that τ is a Levy process.

$$\begin{aligned}
\mathbb{E}\left[e^{-\lambda\tau}\right] &= \int_0^\infty e^{-\lambda t}\frac{1}{\sqrt{2\pi}}t^{-3/2}e^{-1/2t}dt \\
&= e^{-\sqrt{2\lambda}} \\
&= \exp\left(\int_0^\infty (e^{-\lambda t}-1)\frac{x^{-3/2}}{\sqrt{2\pi}}dx\right)
\end{aligned}$$

and at the same time,

$$\mathbb{E}\left[e^{-\lambda\tau}\right] = \exp\left(\int_0^\infty (1-e^{-\lambda t})\nu(dx)\right)$$

where ν is the jump measure for τ. It follows that $\nu(dx) = \frac{x^{-3/2}}{\sqrt{2\pi}}dx$. So, $\tau_t = \int_0^t \int_0^\infty xN(dt, dx)$ where N is a Poisson point process on $(0,\infty)^2$ with intensity $\frac{x^{-3/2}}{\sqrt{2\pi}}dtdx$. Also,

$$\mathbb{E}\left[\tau_1\right] = -\frac{d}{d\lambda}\mathbb{E}\left[e^{-\lambda t_1}\right]\big|_{\lambda=0} = \frac{1}{\sqrt{2\pi}}e^{-\sqrt{2\lambda}}\big|_{\lambda=0} = \infty$$

Lecture 17
19 November 2018

Isaac

1.26 Theorem A [Markov] Levy process is (strictly) *stable* if

$$(X_t)_{t\geq 0} \stackrel{\mathrm{d}}{=} (\tfrac{X_{r_t}}{r^{1/p}})_{t\geq 0},$$

which happens iff [it is Levy and]

(1) $p = 2$ and X is BM.
(2) $0 < p < 2$ and X is pure jump † with $\nu(dx) \propto x^{-1-p}dx$

Proof. Let $X \sim Levy(b, \sigma, \nu)$ have drift, diffusion, and jump kernel coefficients. (i.e. with generator

$$G(f) = b\frac{d}{dx}f + \frac{\sigma^2}{2}\frac{d^2}{dx^2}f + \int \nu(dy)(f(x+y) - f(x)).)$$

Note, we will write $s = r^p$ in the following for notational simplicity. s is the coefficient 'r' from the theorem statement (i.e., $X_{r_t}/r^{1/p} = X_{s_t}/s^{1/p}$). Then, if we examine $X(r^p t)$, the generator associated with the drift is

$$b\tfrac{d}{dx} \to X_t = bt,$$

so $X_{r^p t} = br^p t$. Similarly, $\frac{\sigma^2}{2}\frac{d^2}{dx^2} \to X_t = \sigma B_t \sim \mathbb{N}(0, \sigma^2 t)$. So, $X_{r^p t} \sim \mathbb{N}(0, \sigma^2 r^p t) \stackrel{\mathrm{d}}{=} (\sigma r^{p/2})B_t$. And for the jumps, similar logic shows that $X(r^p t) \sim Levy(r^p b, r^{p/2}\sigma, r^p\nu)$. This is equal in distribution to

$$rX(t) \sim Levy(rb, r\sigma, \tilde{\nu})$$

where $\tilde{\nu}([a,b]) = \nu([\frac{a}{r}, \frac{b}{r}])$. If $(X(r^p t))_{t\geq 0} \stackrel{\mathrm{d}}{=} (rX(t))_{t\geq 0}$ then

(1) either $b = 0$ or $p = 1$
(2) either $\sigma = 0$ or $p = 2$
(3) either $\nu = 0$ or $r^p\nu = \tilde{\nu}$.

If $r^p\nu = \tilde{\nu}$ then $\nu([x,\infty)) = r^{-p}\nu([\frac{x}{r},\infty))$ so setting $r = x$ implies that

$$\nu([x,\infty)) = x^{-p}\nu([1,\infty)) \implies \nu([x,\infty)) \propto x^{-p}.$$

This also holds for $\nu((-\infty,-x])$, and so if ν is absolutely continuous wrt Lebesgue measure then

$$\nu(dx) = \begin{cases} C_+ x^{-p-1}dx & x \geq 0 \\ C_- x^{-p-1}dx & x < 0 \end{cases}$$

for some $C_\pm \geq 0$. (this is what is meant when writing *pure jump* with the corresponding measure in the theorem statement. ∎

Definition (Stable limits). Let $\{Y_i\}_i$ be a family of iid (nondegenerate) Random Variables in $\mathbb{R}$ and let $S_n = Y_1 + Y_2 + \cdots + Y_n$. Suppose $S_n/n^\alpha \xrightarrow{d} X$ converges in distribution for some distribution X. Then X is (strictly) stable, with index $p = 1/\alpha$ i.e., having the distribution of $X(1)$ from the previous theorem.

Proof. $\frac{S_{nk}}{(nk)^\alpha} \xrightarrow{d} X$ but also converges to $\frac{X_1+\cdots+X_k}{k^\alpha}$ where X_i are iid and $\sim X$, all of this because $S_{nk} \overset{\text{d}}{=} S_n^{(1)} + \cdots + S_n^{(k)}$ are iid copies of S_n. That is, we see that the distribution X can be rescaled into a k-fold convolution of k copies of itself: $X \overset{\text{d}}{=} \frac{X_1+\cdots+X_k}{k^\alpha}$.

Now let Z be the Poissonization of S, i.e. $Z_t = S_{N_t}$ where $N_t = PP(1)$ (poisson process) on $[0,\infty)$. Then, $Z \sim Levy(0,0,\nu)$ where ν is the probability distribution of Y. and $S_n/n^\alpha \approx Z_n/n^\alpha$ since $N_n = n + \mathcal{O}(\sqrt{n})$, and $Z_n/n^\alpha \sim Levy(0,0,\nu^{(n)})$ where $\nu^{(n)}([x,\infty)) = \nu([n^\alpha x,\infty))$

Let $Z_t^{(n)} = \frac{1}{n^\alpha} Z_{nt}$ The assumption $X_n = Y_1 + \cdots + Y_n$ implies that $Z_1^{(n)}$ converges in distribution to X, and so by the fact that $X \overset{\text{d}}{=} \frac{X_1+\cdots+X_k}{k^\alpha}$, then X_t as a process is stable if the condition holds for time $t \neq 1$. but we could do the same argument to show that in fact the quoted assumption above implies $(Z_t^{(n)})_{t\geq 0} \xrightarrow{d} (U_t)_{t\geq 0}$ where U is a stable process whose increments are determined by X: $U_1 \overset{\text{d}}{=} X$.

□

Note: $\tau_1 = \inf\{t \geq 0 : B_t = 1\} \sim Stable(\frac{1}{2})$ on $[0,\infty)$. So, $\tau_n \overset{\text{d}}{=} \tau_1^{(n)} + \cdots + \tau_1^{(n)}$ decomposes into a sum of n iid copies of τ_1. Then

$$\tau_n/n^2 \overset{\text{d}}{=} \frac{\tau_1^{(n)} + \cdots + \tau_1^{(n)}}{n^2} \overset{\text{d}}{=} \tau_1,$$

(but also converges in distribution to τ_1 ($\xrightarrow{d} \tau_1$).)

Isaac

Lecture 18
21 November 2018

17.1 Critical Branching

Definition. Let $(V_k)_{k=1}^\infty$ be defined by $(V_{k+1}|V_k) = \sum_{j=1}^{V_k} M_{kj}$ for $V_k \in \mathbb{Z}$, where

$$\{M_{kj} = \#(\text{offspring of indiv } j \text{ at time } k)\}$$

are iid with distribution

$$M = \begin{cases} 0 & \text{with probability } p \\ 1 & 1-2p \\ 2 & p \end{cases}$$

Notice the *branching property*, that

$$((V_k)_{k=1}^\infty | V_0 = n) \overset{\text{d}}{=} ((V_k^{(1)} \underset{\text{iid}}{+ \cdots +} V_k^{(n)}) | V_0^{(1)} = \cdots = V_0^{(n)} = 1).$$

We are interested in the total # of "individuals", that is, in

$$T = \sum_{k\geq 0} \sum_{j=1}^{V_k} \mathbf{1}_{\{M_{kj}=0\}}$$

[INSERT branching diagram]

1.27 Lemma $T < \infty$ almost surely.

More generally, for a branching process with offspring distribution M, what is the *extinction probability*

$$P_e = \mathbb{P}\{\lim_{k\to\infty} V_k = 0 | V_0 = 1\}$$

that is that V dies out. We can introduce some recursion here, as V dies out iff the families of all first-generation offspring die out. Then, by the branching property,

$$P_e = \mathbb{E}[P_e^M] =: \phi(p_e)$$

That is, define

$$\phi(u) = \mathbb{E}[u^M] = \sum_{n\geq 0} u^n \mathbb{P}\{M=n\},$$

a concave function as $\mathbb{P}\{M=n\} \geq 0 \implies \phi''(u) \geq 0$ for $u \in [0,1]$. P_e is a fixed point of $\phi(u)$, and we can compute

$$\phi(0) = \mathbb{P}\{M=0\}, \quad \phi(1) = 1,$$

and $\phi'(1) = \mathbb{E}[M] = \mu$. Assume this is finite, and recalling the concavity condition on ϕ, this leads to classifying conditions:

(1) $\mu < 1$ (subcritical): $P_e = 1$ is the only solution.
(2) $\mu = 1$ critical: $P_e = 1$ is the only solution.
(3) $\mu > 1$ (supercritical): If $\mu > 1$ then $P_e < 1$.

[INSERT criticality plots]

Let

$$\psi(u) = \mathbb{E}[u^T|V_0=1] = \sum_{n\geq 0} u^n \mathbb{P}\{T=n|V_0=1\}.$$

and

$$u^T \overset{\mathrm{d}}{=} \begin{cases} u & M_p = 0 \\ u^T & M_p = 1 \\ u^{T^{(1)}} \times u^{T^{(2)}} = (u^T)^2 & M_p = 2 \end{cases}$$

where we have an assumption that $T^{(1)}, T^{(2)}$ are iid $\sim T$. Then, since $\mathbb{E}[u^{T^{(1)}} u^{T^{(2)}}] = \mathbb{E}[u^{T^{(1)}}]\mathbb{E}[u^{T^{(2)}}] = \psi(u)^2$,

$$\psi(u) = pu + (1-2p)\psi(u) + p\psi(u)^2$$

that is, solving the quadratic,

$$\begin{aligned}\psi(u) &= 1 - \sqrt{1-u} \\ &= \sum_{n\geq 1} \frac{1}{2\sqrt{\pi}} \frac{\Gamma(n-\frac{1}{2})}{n!} u^n \text{ by Binomial Series} \\ \implies \mathbb{P}\{T=n|V_0=1\} &= \frac{1}{\sqrt{2\pi}} \frac{\Gamma(n-\frac{1}{2})}{n!}\end{aligned}$$

Stirling's formula gives the approximation $\log\Gamma(z) = z\log z - z + O(\log z)$ and

$$\begin{aligned}\log\frac{\Gamma(n-\frac{1}{2})}{\Gamma(n+1)} &= (n-\tfrac{1}{2})\log(n-\tfrac{1}{2}) - (n-\tfrac{1}{2}) - (n+1)\log(n+1) + (n+1) + O(\log n) \\ &\approx \tfrac{3}{2}\log(n)\end{aligned}$$

giving the approximation

$$\mathbb{P}\{T=n|V_0=1\} \approx \frac{1}{\sqrt{2\pi}} n^{-3/2}.$$

That is, $\mathbb{P}\{T>n\} \approx n^{-1/2}$.

Let $H_n = T^{(1)} + \cdots + T^{(n)}$. For instance, if T_n is the number of berries on a fully grown branch of a bush, then H_n is the total number of berries after n time steps (fully grown branch "generations").

What will make this stabilize? $n^{3/2}$, $n^{5/2}$, or $n^{1/2}$? Well, using the stability property we can see that $\frac{H_n}{n^2} \to \text{Stable}(\frac{1}{2})$.

Reminder: Stability Property says that if Y_i are iid, $\mathbb{P}\{Y_i > y\} \sim y^{-p}$ as $y \to \infty$, then $\frac{\Sigma Y_i}{n^{1/p}} \to$ Stable process of order (p).

Then, $\left(X_t = H_{n_t}/n^2\right)_{t\geq 0} \to (\tau)t)_{t\geq 0}$, BM hit time, which is a stable(1/2) subordinator. That is, there will probably be a bush with $\approx n^2$ berries (after n "generations"):

$$\mathbb{E}[\#k : T^{(k)} > x] = n\mathbb{P}\{T > x\} \sim nx^{-1/2} = 1$$

if $x \propto n^2$, so $\mathbb{P}\{T > n^2\} \sim n^{-1}$.