Lecture I: Review of Probability

Goals: • Review what you should remember from undergrad prob. class
• Define notation + foundational definitions.

1 **Def:** A stochastic variable is a object X defined by

a) A set of possible values or sample space: Ω_X ($\in \mathbb{R}$ for now (sometimes))

b) A probability distribution P_X satisfying

i) $P(x) \geq 0 \quad \forall x \in \Omega_X$

ii) $\int P(x)\,dx = 1$.

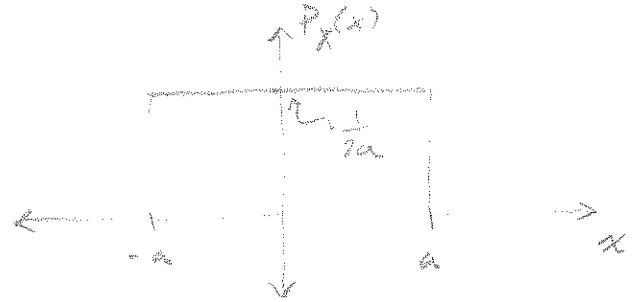
Note: the probability that $X \in (x, x+dx)$ is $P(x)dx$.

Note: Ω_X may be discrete or continuous. When discrete, we will use a sum as in $\sum P(x) = 1$.

Note: The book likes continuous spaces and employs $\delta(x) = \begin{cases} 1 & \text{if } x=0 \\ 0 & \text{else} \end{cases}$ like its going out of style.

Example: $\Omega_X = \mathbb{R}$ (the "square dist")

$$P_X(x) = \begin{cases} 0 & \text{if } |x| > a \\ \frac{1}{2a} & \text{otherwise} \end{cases}$$



Example: $\Omega_X = \{1, 2, 3, 4, 5, 6\}$ (a single die)

$P_X(x) = 1/6$

Example: $\Omega_X = \mathbb{R}$ (a single die again)

$$P_X(x) = \sum_{n=1}^{6} \frac{1}{6} \delta(x - n)$$

Example: $\Omega_X = \{ r_1 r_2 \dots r_n \mid r_i \in \{1, \dots, 6\} \}$ (n rolls of a die)

$$P_X(x) = \frac{1}{6^n}$$

Note: This is a simple stochastic process!

Q: How can I write this so that $\Omega_X \in \mathbb{R}$?
(use sextal decimals?)

Note: We are not (yet) being very careful. In a mathematical treatment, P needs to be a measure on a σ-field (see Breiman Def 1.28). Wait for Thorsley.

(2) Let $f : \mathbb{R} \to \mathbb{R}$ be a function of X. The expected value of f is

$$\langle f(X) \rangle = \int f(x) P_X(x)\, dx.$$

The m-th moment is $\langle X^m \rangle = \mu_m$ and the variance is $\langle (X - \langle X \rangle)^2 \rangle = \mu_2 - \mu_1^2$

Example: The moments of the square distribution are

$$\langle X^m \rangle = \int_{-a}^{a} x^m \frac{1}{2a}\, dx = \frac{1}{2a} \frac{1}{m+1} x^{m+1} \Big|_{-a}^{a}$$

$$= \frac{1}{2a} \frac{1}{m+1} \left(a^{m+1} - (-a)^{m+1} \right).$$

m even $\Rightarrow \langle X^m \rangle = 0$

Thus,

$$\langle X \rangle = \boxed{0}$$

$$\langle X^2 \rangle = \frac{1}{2a} \cdot \frac{1}{3} \cdot 2a^3 = \frac{a^2}{3}$$

$\vdots$

Example: For 1 toss of a die

$$\langle X \rangle = \sum_{n=1}^{6} \frac{n}{6} = \frac{1}{6} 21 = \boxed{3.5}$$

$$\langle X^2 \rangle = \sum \frac{n^2}{6} = \frac{1}{6}(1+4+9+16+25+36)$$

$$= \frac{81}{6} = \boxed{13.5}$$

$$\sigma^2 = \mu_2 - \mu_1^2 = 13.5 - (3.5)^2 = 13.5 - 12.25 = 0.25$$

The characteristic function of X whose sample space is a subset of $\mathbb{R}$ is

$$G(k) = \langle e^{ikx} \rangle = \int e^{ikx} P(x)\,dx$$

where $i = \sqrt{-1}$ and $k \in \mathbb{R}$ is some number (that doesn't really matter). Lets write the Taylor series for $G(k)$:

$$G(0) + G'(0)k + \frac{G''(0)k^2}{2!} + \frac{G'''(0)k^3}{3!} + \dots$$

Note that

$$\frac{\partial}{\partial k}\int e^{ikx} P(x)\,dx = \int \frac{\partial}{\partial k} e^{ikx} P(x)\,dx$$

$$= i \int x e^{ikx} P(x)\,dx.$$

So $G(0) + G'(0)k + \dots$

$$G(k) = \underbrace{\int P(x)dx}_{1} + (ik)\underbrace{\int xP(x)dx}_{\langle X \rangle} + \frac{(ik)^2}{2!}\underbrace{\int x^2 P(x)dx}_{\langle X^2 \rangle} + \dots$$

$$= \sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \langle X^m \rangle .$$

Thus, you can use the characteristic function to <u>generate</u> the <u>moments</u>.

Example: For the square distribution

$$G(k) = \frac{1}{2a}\int_{-a}^{a} e^{ikx}\,dx = \frac{1}{2a}\cdot\frac{1}{ik}\,e^{ikx}\Big|_{-a}^{a}$$

$\frac{1}{i} = -i$

Note $\sin(a) = \frac{ie^{-ia} - ie^{ia}}{2}$

$$= \frac{-i}{2ak}\left(e^{ika} - e^{-ika}\right) = \frac{\sin(ak)}{ak}$$

$$= \underset{\substack{\text{take}\\ \text{limit}\\ \text{as } k\to 0}}{1} + \underset{\langle X \rangle}{0k} + \underset{\langle X^2 \rangle}{\frac{a^2}{3}}\,\frac{k^2 i^2}{2!} + \dots$$

$\lim_{k\to 0} \frac{\sin(ak)}{ak} = \lim_{k\to 0} \frac{a\cos(ak)}{a} = 1$

$\frac{\partial}{\partial k}\frac{\sin(ak)}{ak} = \frac{\cos(ak)}{k} - \frac{\sin(ak)}{ak^2}$

$\frac{\partial^2}{\partial k^2} \cdot =$

Note: All odd moments are zero!

$\boxed{\text{Show Mathematica}}$

Note: One can also define the <u>cumulants</u>: by taking the Taylor expansion of

$$\log G(k) = \sum_{m=0}^{\infty} \frac{(ik)^m}{m!} \kappa_m$$

where it turns out that

$$\kappa_1 = \mu_1$$
$$\kappa_2 = \sigma^2$$
$$\vdots$$

This can be found by computing the terms of the series using the chain rule to deal with the log. (show Mathematica)

Note: In a discrete space we get

$$G(k) = \sum p_n e^{ikn}.$$

Example: One die:

$$G(k) = \sum_{n=1}^{6} \frac{1}{6} e^{ikn} = \frac{1}{6}\left(1 + e^{2ik} + e^{3ik} + \cdots\right),$$

which is easy to expand in k.

1/7

③ Multivariate Distributions

- Suppose $\bar{X} = (X_1, \ldots, X_r)$ is a vector of random variables. One has to define a <u>joint distribution</u>

$$P_r(x_1, x_2, \ldots, x_r)$$

that describes the probability that

$$X_1 = x_1 \text{ and } X_2 = x_2 \text{ and } \ldots \text{ and } X_r = x_n.$$

- Suppose $s < r$, then the <u>marginal distribution</u> is,

$$P_s = \int P_r(x_1, \ldots, x_s, x_{s+1}, \ldots, x_r)\, dx_{s+1} \cdots dx_r$$

which describes the probability that

$$X_1 = x_1, \ldots, X_s = x_2$$

regardless of what $X_{s+1}, \ldots, X_r$ are.

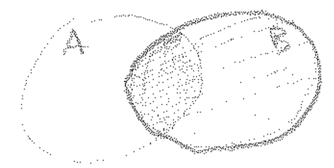
- If $X_{s+1}, \ldots, X_r$ are fixed, we get the <u>conditional distribution</u>:

$$P_{s|r-s}(x_1, \ldots x_s \mid x_{s+1}, \ldots, x_r).$$

▸ Note that

$$P_r(x_1,\dots,x_r) = P_{r-s}(x_{s+1},\dots,x_r)\,P_{s|r-s}(x_1,\dots,x_s \,|\, x_{s+1},\dots,x_r).$$

To see this another way, consider two events A and B



← want the fraction of the dark circle where a occurs.

$$P(a|b) = \frac{P(a,b)}{P(b)} \quad \begin{array}{l} \leftarrow \text{football} \\ \leftarrow \text{normalizer} \end{array}$$

which is Bayes Rule.

▸ If $P_r(x_1,\dots,x_r) = P_s(x_1,\dots,x_s)\,P_{r-s}(x_{s+1},\dots,x_r)$ then the variables $(X_1,\dots,X_s)$ and $(X_{s+1},\dots,X_r)$ are called statistically independent.

▸ Moments are defined as usual:

$$\langle X_1^{m_1} X_2^{m_2} \cdots X_r^{m_r} \rangle = \int x_1^{m_1}\cdots x_r^{m_r}\, P(x_1,\dots,x_r)\, dx_1 \cdots dx_r$$

▸ The generating function is

$$G(k_1,\dots,k_r) = \left\langle e^{i(k_1X_1+\cdots+k_rX_r)} \right\rangle$$

$$= \sum_{m_1}\cdots\sum_{m_r} \frac{(ik_1)^{m_1}\cdots(ik_r)^{m_r}}{m_1!\cdots m_r!} \langle X_1^{m_1}\cdots X_r^{m_r} \rangle.$$

and one usually defines

a) The matrix of 2^{nd} moments

$$\begin{pmatrix} X_1X_1 & \cdots & X_nX_1 \\ \vdots & & \\ X_1X_n & \cdots & X_nX_n \end{pmatrix}$$

b) The covariance matrix whose i,j-th entry is

$$\langle\langle X_iX_j\rangle\rangle = \langle (X_i - \langle X_i\rangle)(X_j - \langle X_j\rangle)\rangle$$

$$= \langle X_iX_j - X_i\langle X_j\rangle - \langle X_i\rangle X_j + \langle X_i\rangle\langle X_j\rangle\rangle$$

$$= \langle X_iX_j\rangle - \langle X_i\rangle\langle X_j\rangle.$$

▶ Independence ($r=2$ case). Say X_1 and X_2 are independent. Then the following characterize this

a) $\langle X_1^{m_1}X_2^{m_2}\rangle = \langle X_1^{m_1}\rangle\langle X_2^{m_2}\rangle$

b) $G(k_1,k_2) = G_1(k_1)G_2(k_2)$

a $\langle X_1^{m_1} X_2^{m_2} \rangle = \iint x_1^{m_1} x_2^{m_2} p(x_1, x_2)\, dx_1 dx_2$

$$= \iint x_1^{m_1} p(x_1)\, x_2^{m_2} p(x_2)\, dx_1 dx_2$$

$$= \int x_1^{m_1} p(x_1)\, dx_1 \int x_2^{m_2} p(x_2)\, dx_2 = \langle X_1^{m_1} \rangle \langle X_2^{m_2} \rangle$$

b $G(k_1, k_2) = \iint e^{ik_1 x_1} e^{ik_2 x_2} P(x_1) P(x_2)\, dx_1 dx_2$

$$= G(k_1) G(k_2).$$

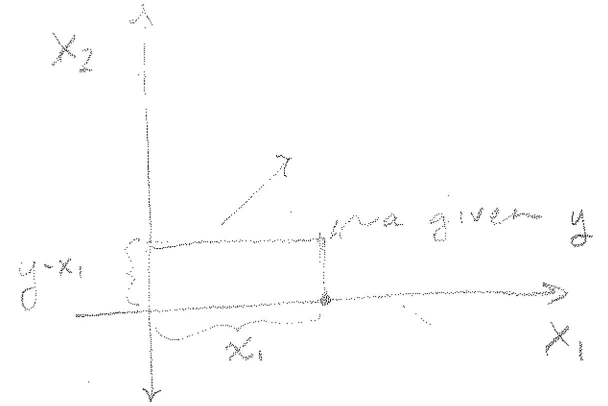
Note that if $\langle\langle X_1 X_2 \rangle\rangle = 0 \Leftrightarrow \langle X_1 X_2 \rangle = \langle X_1 \rangle \langle X_2 \rangle$ (see 3.8)) then X_1 and X_2 are merely uncorrelated, but not necessarily independent (higher order moments may differ).

4) Addition of Random Variables

Say X_1 and X_2 are random variables. Put

$$Y = X_1 + X_2.$$

What is P_Y in terms of $P(X_1, X_2)$



The probability that $Y=y$ is the integral along this line:

$$P(y) = \int P(x_1, y-x_1)\,dx_1$$

Note that if X_1 and X_2 are independent, then

$$P(y) = \int P_{X_1}(x_1)\,P_{X_2}(y-x_1)\,dx_1 \quad \left(\begin{array}{c}\text{the convolution}\\ \text{of } P_{X_1} \,\&\, P_{X_2}\end{array}\right).$$

<u>Fact</u>: $\langle Y\rangle = \langle X_1\rangle + \langle X_2\rangle$ (regardless of independence)

<u>Proof</u>: $\langle Y\rangle = \langle X_1+X_2\rangle = \iint (x_1+x_2)\,P(x_1,x_2)\,dx_1\,dx_2$

$$= \int x_1 \underbrace{\int P(x_1,x_2)\,dx_2}_{P(x_1)}\,dx_1 + \int x_2 P(x_1,x_2)\,dx_1\,dx_2$$

$$= \langle X_1\rangle + \langle X_2\rangle.$$

(Note also that $\langle aX\rangle = a\langle X\rangle$.)

Can also show $\sigma_Y^2 = \sigma_{X_1}^2 + \sigma_{X_2}^2$ if <u>uncorrelated</u>.

Finally,

$$G_Y(k) = \left\langle e^{ikY} \right\rangle = \left\langle e^{ik(X_1+X_2)} \right\rangle$$

$$= \left\langle e^{ikX_1 + ikX_2} \right\rangle = G_{X_1,X_2}(k,k).$$

If independent then

$$G_Y(k) = G_{X_1}(k)\, G_{X_2}(k).$$

The drunkard's walk

- initially position = 0
- every step is right with $p = 1/2$
  left with $p = 1/2$.

Each step i corresponds to a variable X_i with

$$P_{X_i}(x_i) = \begin{cases} 1 & \text{with prob } 1/2 \\ -1 & \text{" " "} \end{cases}.$$

After r steps, the drunkard's position is

$$Y = X_1 + \cdots + X_r.$$

- By addition of r.v.s we get

$$\langle Y \rangle = \langle X_1 \rangle + \cdots + \langle X_r \rangle$$
$$= 0 + \cdots + 0 = 0.$$

- Note that

$$\langle X_i X_j \rangle = (1)(1)\left(\frac{1}{2}\right)^2 + (1)(-1)\left(\frac{1}{2}\right)^2 + (-1)(1)\left(\frac{1}{2}\right)^2 + (-1)(-1)\left(\frac{1}{2}\right)^2$$
$$= 0.$$

thus,

$$\langle Y^2 \rangle = \langle (X_1 + \cdots + X_r)^2 \rangle$$
$$= r \langle X_i^2 \rangle = r \quad \left(\text{since } \langle X_i^2 \rangle = 1 \cdot \tfrac{1}{2} + 1 \cdot \tfrac{1}{2} = 1\right)$$

If we define the velocity as distance over time $\frac{Y}{r}$, then the variance of the velocity is

$$\left\langle \left(\frac{Y}{r}\right)^2 \right\rangle = \frac{1}{r^2}\langle Y^2 \rangle = \frac{1}{r} \longrightarrow 0 \quad \text{as } r \to \infty.$$

- When you have the mean & variance, you still don't know everything. We want the probability distribution for each r.

$$G_{Y_r}(k) = \left(G_X(k)\right)^r \quad \text{(by independence)}$$

$$= \left\langle e^{ikX} \right\rangle^r$$

$$= \left(\frac{1}{2} e^{ik} + \frac{1}{2} e^{-ik} \right)^r$$

$$= \sum_n p_n e^{ikn} \quad \text{(by def of } \langle e^{ikn} \rangle \text{)}$$

$r = 1$: $\quad \frac{1}{2} e^{ik(1)} + \frac{1}{2} e^{ik(-1)}$

$r = 2$: $\quad \frac{1}{4} \left(e^{ik}\right)^2 + \frac{1}{2} e^{ik} e^{-ik} + \frac{1}{4} \left(e^{-ik}\right)^2$

$$= \frac{1}{4} e^{ik(2)} + \frac{1}{2} e^{ik(0)} + \frac{1}{4} e^{ik(-2)}$$

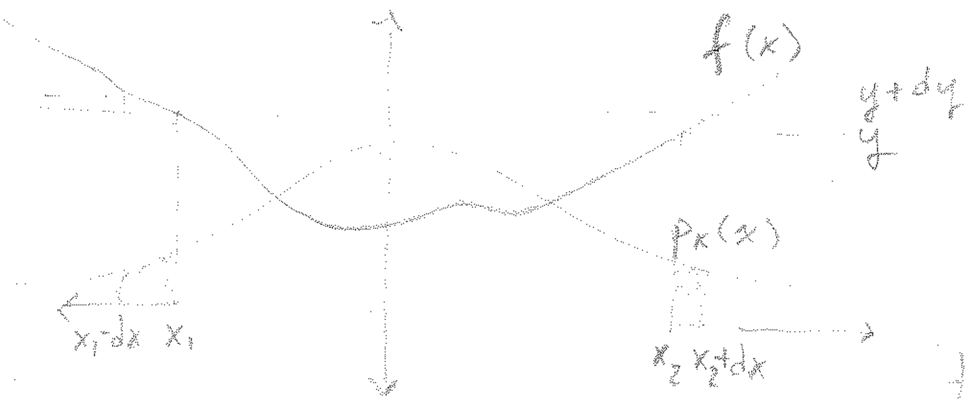
$$P_n(r) = \frac{1}{2^r} \binom{r}{\frac{1}{2}(r-n)}.$$

More generally, we don't just add random variables, we can transform them however we want to: $Y = f(X)$. This is how we will arrive at stochastic processes later.

We get

$$P_Y(y) = \int \delta(f(x) - y)\, P_X(x)\, dx$$

although this is not a very useful formula. Let's try something else:



this sum ranges over all roots of $y = g(x)$.

$$P\{y \le Y \le y+dy\} = P_Y(y)\,|dy|$$
$$= \sum_{i=1}^{n} P_X(x_i)\,|dx_i|$$

Note $\left.\frac{dy}{dx}\right|_{x_i} = f'(x_i)$

$$\Rightarrow \boxed{P_Y(y) = \sum_{i=1}^{n} \frac{P_X(x_i)}{|f'(x_i)|}}$$

this is slightly more general than van Kampen's.

Ex. Say X has the "square distribution", and $Y = X^2$. Then

$$x_1 = \sqrt{y}\ , \quad x_2 = -\sqrt{y}$$

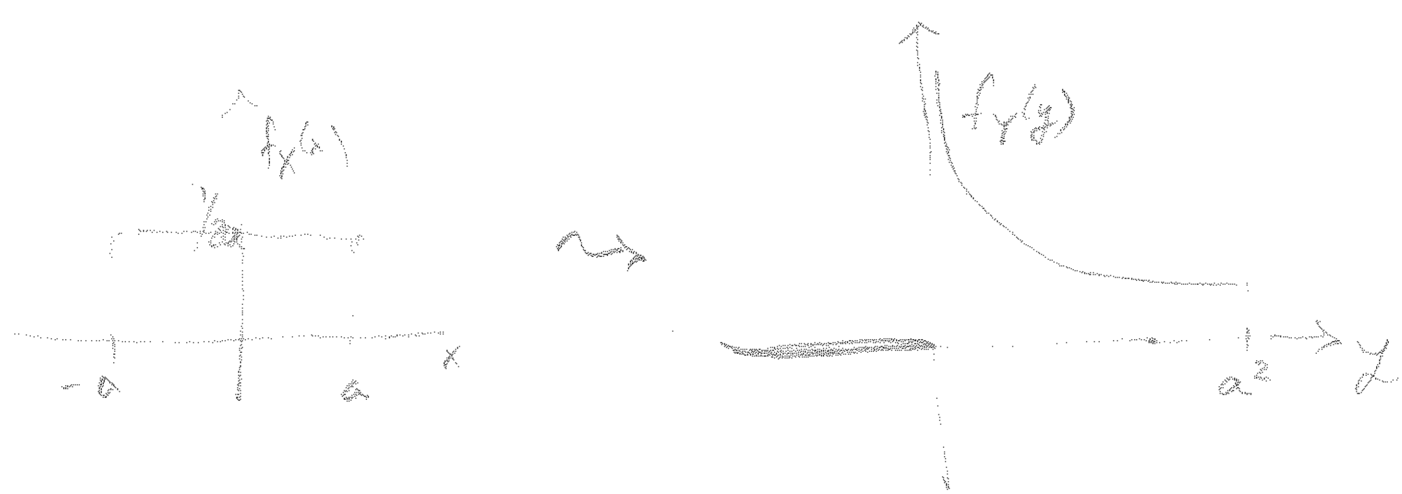
and

$$g'(\sqrt{y}) = 2\sqrt{y}\ .$$

thus

$$P_Y(y) = \frac{f_X(\sqrt{y})}{|2\sqrt{y}|} + \frac{f_X(-\sqrt{y})}{|2\sqrt{y}|}$$

$$= \begin{cases} \dfrac{1}{2a\sqrt{y}} & \text{if } \quad 0 \le y \le a^2 \\ 0 & \text{else.} \end{cases}$$



Next time:

- Gaussians
- Central Limit Θ
- Kinetic Theory + Maxwell's Distribution?

(5) The Gaussian Distribution

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

▶ Proof that $\int P(x)\,dx = 1$.

(a) Set $x = \sigma y + \mu$. Then $dx = \sigma dy$ and

$$\int P(x)\,dx = \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-\frac{y^2}{2}} \sigma\, dy$$

$$= \underbrace{\frac{1}{\sqrt{2\pi}} \int e^{-y^2/2}\, dy = 1}$$

It suffices to prove that $N(0,1)$ is a distribution.

(b)
$$\left(\frac{1}{\sqrt{2\pi}} \int e^{-y^2/2}\, dy\right)^2$$

$$= \frac{1}{2\pi} \int e^{-y_1^2/2}\, dy_1 \int e^{-y_2^2/2}\, dy_2$$

$$= \frac{1}{2\pi} \iint e^{-\frac{1}{2}(y_1^2+y_2^2)}\, dy_1\, dy_2$$

$$= \underbrace{\frac{1}{2\pi} \int_0^{2\pi} d\theta}_{1} \underbrace{\int e^{-\frac{1}{2}r^2}\, r\, dr}_{= -e^{-\frac{1}{2}r^2}\Big|_0^{\infty} = 1}$$

using $\iint f(x,y)\,dx\,dy = \iint f(r\cos\theta, r\sin\theta)\, r\, dr\, d\theta$

The moment generating function of $N(\mu, \sigma^2)$

$$G(k) = \frac{1}{\sqrt{2\pi\sigma^2}} \int e^{tx} \, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \, dx . \quad \boxed{t = ik}$$

Note:

$$-\frac{1}{2\sigma^2}\left[(x-\mu)^2 - 2\sigma^2 t x\right]$$

$$= \frac{-1}{2\sigma^2}\left[x^2 - 2(\mu + \sigma^2 t)x + \mu^2\right]$$

$$= \frac{-1}{2\sigma^2}\left[x^2 - 2(\mu+\sigma^2 t)x + \mu^2 + 2\mu\sigma^2 t + \sigma^4 t^2 - 2\mu\sigma^2 t - \sigma^4 t^2\right]$$

$$= \frac{-1}{2\sigma^2}\left[(x - (\mu + \sigma^2 t))^2\right] + \mu t + \frac{\sigma^2 t^2}{2}$$

So

$$G_p(k) = e^{\mu t + \frac{\sigma^2 t^2}{2}} \underbrace{\frac{1}{\sqrt{2\pi\sigma^2}} \int e^{-\frac{(x-(\mu+\sigma^2 t))^2}{2\sigma^2}} \, dx}_{1}$$

And

$$G(k) = e^{\mu i k - \frac{1}{2}\sigma^2 k^2}$$

Expanding $G(k)$ gives

$$G(k) = \langle e^{ikX} \rangle = \left\langle 1 + ikX + \frac{(ik)^2}{2!}X^2 + \frac{(ik)^3}{3!}X^3 + \cdots \right\rangle$$

$$= 1 + ik\langle X \rangle + \frac{(ik)^2}{2!}\langle X^2 \rangle + \frac{(ik)^3}{3!}\langle X^3 \rangle + \cdots$$

$$= G(0) + G'(0)k + \frac{G''(0)k^2}{2!} + \frac{G'''(0)}{3!} + \cdots$$

$$= 1 + (\mu i - \sigma^2 k) e^{\mu i k - \frac{1}{2}\sigma^2 k^2}\Big|_0 k + \cdots$$

$$= 1 + \underbrace{\mu}_{\text{mean}} ik + (-\sigma^2 + \mu i \mu i)\frac{k^2}{2}$$

$$= 1 + \mu ik + (\sigma^2 + \mu^2)\frac{(ik)^2}{2!}$$

recall

$$\langle\langle X \rangle\rangle = \langle (X - \mu)^2 \rangle$$

$$= \langle X^2 - 2\mu X + \mu^2 \rangle$$

$$= \langle X^2 \rangle - \mu^2 = \sigma^2 + \mu^2 - \mu^2 = \sigma^2$$

Recall the cumulants are given by

$$\log G(k) = \sum \frac{(ik)^m}{m!} \kappa_m .$$

In this case we have

$$\log G(k) = i\mu k - \frac{1}{2}\sigma^2 k^2$$

$$\Rightarrow \quad \kappa_1 = \mu_1$$
$$\kappa_2 = \sigma^2$$
$$\kappa_3 = \kappa_4 = \dots = 0 .$$

7 The Central Limit Theorem

The remarkable thing about the gaussian is this. Say $X_1, \dots, X_r$ are indep. random variables with the same distribution and 0 mean. Put $Y = \frac{1}{\sqrt{r}} \sum_{i=1}^{r} X_i$

Then

$$G_Y(k) = \left\langle e^{ik \frac{X_1 + \dots + X_r}{\sqrt{r}}} \right\rangle =$$

$$= G_{X_1}\left(\frac{k}{\sqrt{r}}\right) \times \dots \times G_{X_r}\left(\frac{k}{\sqrt{r}}\right)$$

$$\approx \left[1 - \frac{\sigma^2 k^2}{2r}\right]^r \longrightarrow e^{-\frac{\sigma^2 k^2}{2}}$$

$$G_X\left(\frac{k}{\sqrt{r}}\right) = 1 - \frac{\sigma^2 k^2}{2r} + \dots$$

$$e^x = \lim_{n\to\infty}\left(1 + \frac{x}{n}\right)^n$$

That is, as $r \to \infty$ Y approaches a zero mean gaussian with variance σ^2 !

The Multidimensional gaussian

If $X_1, \dots, X_r$ are all gaussian and independent, then

$$p(x_1, \dots, x_r) = \frac{1}{(\sqrt{2\pi})^{r/2} \sigma_1 \cdots \sigma_r} e^{-\frac{1}{2} \sum \frac{(x_i - \mu_i)^2}{\sigma_i^2}}$$

Put $K = \begin{pmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_r^2 \end{pmatrix}$, $K^{-1} = \begin{pmatrix} 1/\sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\sigma_r^2 \end{pmatrix}$

$\mu = (\mu_1, \dots, \mu_r)^T$

$x = (x_1, \dots, x_r)^T$. Then

$$p(x_1, \dots, x_r) = \frac{1}{(2\pi)^{r/2} \det(K)} e^{-\frac{1}{2}(x-\mu)^T K^{-1} (x-\mu)}$$

It turns out that this works when K is merely positive definite. Then

proved in stack + woods

$K_{ij} = \langle X_i X_j \rangle$.

Chapter II: Random Events

For random events we have samples of the form

$$s \in \mathbb{N} \quad ; \# \text{ of events}$$

$$(z_1, \ldots, z_s) \in \mathbb{R}^s \quad ;$$

That is, a sample consists of some number of "event times", possibly zero.

Symmetry: we require that, e.g. $(1,2,7)$ is the same sample as $(1,7,2)$ etc.

To specify a set of events, we need to define a family of functions

Q_0 : prob. of 0 events

$Q_1(z_1)$: prob of 1 event and that that event's time is 1

$\dfrac{Q_2(z_1, z_2)}{2}$: prob. of 2 events and that the two events occur at times z_1 & z_2 (note we divide by 2 to avoid overcounting)

Also note that we need $Q_2(z_1, z_2) = Q_2(z_2, z_1)$

⋮

$$\frac{Q_s(z_1, \dots, z_s)}{s!} : s \text{ events.}$$

Note that

$$Q_0 + \sum_{s=1}^{\infty} \underbrace{\frac{1}{s!} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} dz_1 \cdots dz_s \, Q_s(z_1, \dots, z_s)}$$

conditional prob. that there are s events regardless of when they are.

<u>Examples</u>: Pick s students from an (essentially) infinite group using distribution $\varphi(s)$ and then report their scores $\psi(z)$ on the SAT. Then, assuming these are indep. we get:

$$Q_0 = \varphi(0)$$

$$Q_1(z_1) = \varphi(1)\psi(z_1)$$

$$\frac{Q_2(z_1, z_2)}{2} = \varphi(2)\psi(z_1)\psi(z_2)$$

Then the sum above is $\varphi(0) + \varphi(1)\int \psi(z_1)\,dz_1 + \varphi(2)\int \psi(z_1)\psi(z_2)\,dz_1 dz_2 \ldots$

$= 1.$

To form an average, you need functions from the sample space:

$$A(s, (\tau_1, \dots, \tau_s)) = A_s(\tau_1, \dots, \tau_n)$$

Then

$$\langle A \rangle = A_0 Q_0 + \sum_{s=1}^{\infty} \frac{1}{s!} \int A_s(\tau_1 \dots \tau_s)\, Q_s(\tau_1, \dots, \tau_s)\, d\tau_1 \dots d\tau_s$$

Example: Say we want the number of events in the interval (t_a, t_b). then we introduce

$$\chi(\tau) = \begin{cases} 1 & \text{if } t_a < \tau < t_b \\ 0 & \text{else.} \end{cases}$$

and

$$A_s(\tau_1, \dots, \tau_s) = \chi(\tau_1) + \dots + \chi(\tau_s).$$

and

$$\langle N \rangle = \sum_{s=1}^{\infty} \frac{1}{s!} \int \left(\sum_{i=1}^{s} \chi(\tau_i) \right) Q_s(\tau_1, \dots, \tau_s)\, d\tau_1 \dots d\tau_s$$

When $Q_s(z_1, \dots, z_s) = e^{-v} q(z_1) \cdots q(z_s)$, $Q_0 = e^{-v}$

(i.e. the events are independent given s) where

$$v = \int_{-\infty}^{\infty} q(z)dz \quad \text{(normalizer)} \quad \text{then}$$

$$\langle N \rangle = \sum_{s=1}^{\infty} \frac{1}{s!} \frac{1}{e^v} \int \left(\sum \chi(z_i) \right) q(z_1) \cdots q(z_s) dz_1 \cdots dz_s$$

Pick an s. The term corresponding to s is

$$\frac{1}{s!} \frac{1}{e^v} \left(\chi(z_1) q(z_1) \cdots q(z_s) + \cdots + \chi(z_s) q(z_1) \cdots q(z_s) \right)$$

each of these s terms is essentially the same

$$= \frac{1}{s!} \frac{1}{e^v} s \underbrace{\int_{t_a}^{t_b} q(z_1) dz_1}_{B} \underbrace{\int q(z_2) dz_2}_{v} \cdots \underbrace{\int q(z_s) dz_s}_{v}$$

$$= \frac{1}{(s-1)!} \frac{1}{e^v} B v^{s-1}.$$

So

$$\langle N \rangle = \frac{1}{e^v} B \left(1 + v + \frac{v^2}{2!} + \cdots \right) = B = \int_{t_a}^{t_b} q(z_1) dz_1$$

Along the same lines, we can find the characteristic function (and ∴ the probability distribution): for N

$$G(k) = \langle e^{ikN} \rangle$$

$$= \frac{1}{e^{\nu}} \sum_{s=0}^{\infty} \frac{1}{s!} \int e^{ik(\chi(\tau_1) + \dots + \chi(\tau_s))} q(\tau_1) \dots q(\tau_s) d\tau_1 \dots d\tau_s$$

$$= \frac{1}{e^{\nu}} \sum_{s=0}^{\infty} \frac{1}{s!} \left(\int e^{ik\chi(\tau)} q(\tau) d\tau \right)^s$$

$$= \frac{1}{e^{\nu}} \exp\left(\int e^{ik\chi(\tau)} q(\tau) d\tau \right)$$

$$= \exp\left(\int \left(e^{ik\chi(\tau)} - 1 \right) q(\tau) d\tau \right)$$

$$= \exp\left(\int_{t_a}^{t_b} \left(e^{ik} - 1 \right) q(\tau) d\tau \right)$$

$$= \exp\left(\left(e^{ik} - 1 \right) \int_{t_a}^{t_b} q(\tau) d\tau \right)$$

$$= \exp\left(\left(e^{ik} - 1 \right) \langle N \rangle \right)$$

$$= \exp\left(e^{ik} \langle N \rangle \right) \exp\left(-\langle N \rangle \right)$$

$$= e^{-\langle N\rangle} \sum_{s=0}^{\infty} \frac{\left(e^{ik}\langle N\rangle\right)^s}{s!}$$

$$= e^{-\langle N\rangle} \sum_{s=0}^{\infty} \frac{\langle N\rangle^s}{s!} e^{iks}$$

$$= \sum_{s=0}^{\infty} e^{iks} p(s) \quad \text{(by definition of } e^{iks}\text{)}$$

$$\therefore \quad \boxed{p(s) = \frac{\langle N\rangle^s}{s!} e^{-\langle N\rangle}} \quad \text{(the Poisson distribution)}$$

This is the distribution of, for example raindrops in a given region in a given time, or radioactive decay.

III. Stochastic Processes.

Let X be a random variable and $f : dom(X) \times \mathbb{R} \to dom(Y)$ be a function

Then

$$Y_X(t) = f(X, t)$$

is a stochastic process.

- Putting in a specific time gives the distribution at that time:

$$Y_X(1.34) = f(X, 1.34)$$

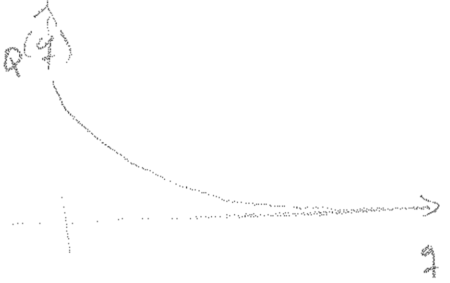
and looks just like the transformations we've encountered.

- Putting in a sample from X gives

$$Y_x(t) = f(x, t)$$

which is a plain old function of time. called a <u>sample trajectory</u>.

Ex ① Say Q is a real valued variable with pdf $P_Q = e^{-q}$, $q \geq 0$.

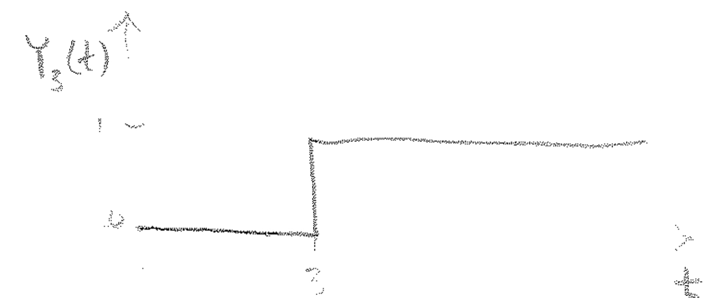


put

$$\begin{tikzcd} 0 \arrow[d, "1dt"] \\ 1 \end{tikzcd}$$

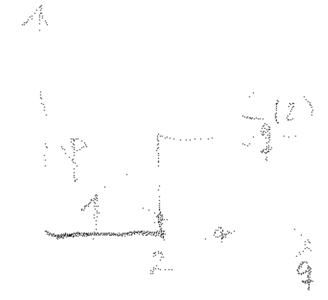
$$Y_q(t) = \begin{cases} 0 & \text{if } t < q \\ 1 & \text{if } t \geq q. \end{cases}$$

A sample trajectory looks like



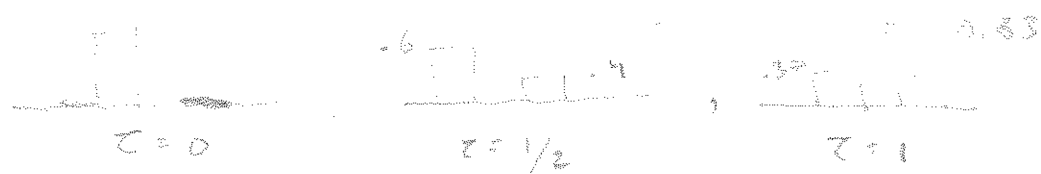
The distribution at time, for example, z is

$$Y_X(z) = \begin{cases} 0 & \text{if } z < q \\ 1 & \text{else} \end{cases}$$



$$P_{Y_X(z)}(0) = \int_z^{\infty} e^{-t}\,dt = e^{-z}$$

$$P_{Y_X(z)}(1) = \int_0^z = 1 - e^{-z}$$



Ex. 2 Say X is any r.v. and $\psi(t)$ is any function (like $\cos(t)$)

$$Y_X(t) = \psi(t)X$$

is a stochastic process.

Ex 3 Say $\Omega_X = \{$

$$(0,0,0,0,0)$$
$$(0,0,0,0,1)$$
$$\vdots$$
$$(1,1,1,1,1,1) \}$$

with $p(s_1 \cdots s_6) = \left(\frac{1}{2}\right)^6$. Put

$$Y_{s_1 \cdots s_6}(t) = s_t \quad \text{for } t \in \{1, \cdots, 6\}.$$

The idea is that the samples of X are possible <u>trajectories</u> of the stochastic process. Note that at any time t the state is either 0 or 1 with prob $1/2$.

$$P_{Y_X(3)}(0) = \sum_{s_i \in \{0,1\}} p(s_1 s_2 0 s_4 s_5 s_6)$$
$$= 2^5 \left(\frac{1}{2}\right)^6 = \boxed{\frac{1}{2}}.$$

Averages are done as functions of time:

$$\langle Y(t)\rangle = \underbrace{\int Y_x(t)\, P_X(x)\,dx}_{\substack{t \text{ is constant} \\ \text{in this integral.}}}$$

Ex ①

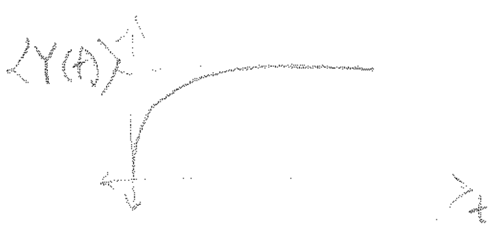
0 when $t<q$, 1 else

$$\langle Y(t)\rangle = \int Y_q(t)\, e^{-q}\,dq$$

$$= \int_0^t e^{-q}\,dq = -e^{-t} - e^{-0} = \boxed{1-e^{-t}}.$$

$0 \le q \le t$

in this range $q<t$ so $Y_q(t)=1$



Ex ②

$$\langle Y(t)\rangle = \int Y_X(t)\, P_X(x)\,dx = \int \psi(t)\, x\, p(x)\,dx = \psi(t)\langle X\rangle.$$

Ex ③

$$\langle Y(t)\rangle = \sum_{s_1 \cdots s_6} s_t \left(\frac{1}{2}\right)^6 = \frac{1}{2}.$$

Similary you can find $\langle (Y_X(t))^m \rangle$.

It is also interesting to find the nth moments:

$$\langle Y_X(t_1) Y_X(t_2) \cdots Y_X(t_n) \rangle$$

or the autocorrelation function

$$K(t_1, t_2) = \langle\langle Y_X(t_1) Y_X(t_2) \rangle\rangle$$
$$= \langle Y_X(t_1) Y_X(t_2) \rangle - \langle Y_X(t_1) \rangle \langle Y_X(t_2) \rangle$$

Ex(1) The 2^{nd} moment is

$$\langle Y(t_1) Y(t_2) \rangle = \int \underbrace{Y_q(t_1) Y_q(t_2)}_{\text{both } 1 \text{ when } t_1 \,\&\, t_2 < q} e^{-q} dq$$

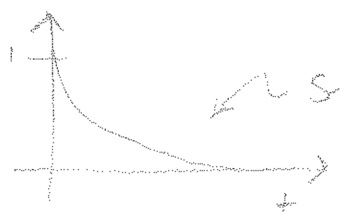
$$= \int_0^{\min(t_1, t_2)} e^{-q} \qquad \equiv \quad q < t_1 \wedge q < t_2$$

$$= 1 - e^{-\min(t_1, t_2)} = 1 - e^{-t_1}$$

(assume $t_1 \leq t_2$)

$$K(t_1, t_2) = \left(1 - e^{-t_1}\right) - e^{-t_1 - t_2}$$

$K(0,t) = e^{-t}$



state becomes uncorrelated.

compare $K(0,t)$

Ex (2) $\langle Y_X(t_1) Y_X(t_2) \rangle = \int \psi(t_1) x \, \psi(t_2) x \, P_X(x) dx$

$$= \psi(t_1)\psi(t_2)\langle X^2 \rangle.$$

$$K(t_1, t_2) = \psi(t_1)\psi(t_2)\langle X^2\rangle - \psi(t_1)\psi(t_2)\langle X\rangle^2$$

Ex (3) $\langle Y_S(t_1) Y_S(t_2) \rangle = \sum_S s_{t_1} s_{t_2} \left(\frac{1}{2}\right)^6$

$$= 2^4 \left(\frac{1}{2}\right)^6 = \frac{1}{4}.$$

$$K(t_1, t_2) = \frac{1}{4} - \frac{1}{2}\cdot\frac{1}{2} = 0.$$

Stationary: $\langle Y(t_1+\tau_1) \cdots Y(t_n+\tau_n) \rangle = \langle Y(t_1) \cdots Y(t_n) \rangle.$
Def

Only ex. (3) is stationary
(2) might be, depending on X
(1) is practically stationary
for large t.

Talk about • ensembles
• ergodic

Skip III.3

III.4 Hierarchy of distributions

Define

$P_1(y,t)$ = Probability that $Y_X(t)$ has value y at time t

$$= \int \delta(y - Y_x(t)) P_X(x) dx$$

Which is for purely definitional purposes. You would want to use the tool we developed for transformations to really figure this out.

Similarly define:

$$P_n(y_1,t_1; \dots ; y_n,t_n)$$

$$= \int \delta(y_1 - Y_x(t_1)) \cdots \delta(y_n - Y_x(t_n)) P_X(x) dx. \quad *$$

Everything we've done with stochastic processes can be done with this hierarchy: Eg

$$\langle Y(t_1) \cdots Y(t_n) \rangle = \int y_1 \cdots y_n P_n(y_1,t_1; \dots ; y_n,t_n) dy_1 \cdots dy_n$$

Properties:

i) $P_n(y_1,t_1;\dots;y_n,t_n) \geq 0$

(follows from the def. of P_X & δ)

ii) P_n does not depend on flipping (y_i,t_i) & (y_j,t_j) for any $i,j \in \{1,\dots,n\}$.

(follows from associativity of $\cdot$)

iii)

$$\int P_n(y_1,t_1;\dots;y_n,t_n)\,dy_n = P_{n-1}(y_1,t_1;\dots;y_{n-1},t_{n-1})$$

(integrate * w.r.t. y_2 - switch order and use $\int \delta(y)\,dy = 1$)

iv) $\int P_1(y,t)\,dy = 1$

(use $\int \delta(y)\,dy = 1$ & $\int P_X(x)\,dx = 1$).

What we have shown is that

A stoch proc. defined by $Y_X(t) = f(X,t)$ satisfies i — iv

Kolmogorov: Any hierarchy satisfying i — iv uniquely determine a stoch proc. $Y(t)$ [note the X is fairly bizarre].

Ex 1 Last time we determined that, for the process

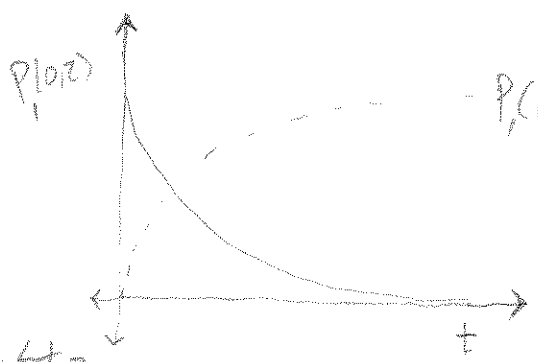
$$Y_q(t) = \begin{cases} 0 & \text{if } t \le q \\ 1 & \text{else} \end{cases}$$

with $P_Q(q) = e^{-q}$, we had

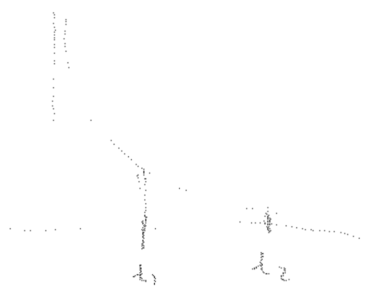
$$\left.\begin{aligned} P(0,t) &= e^{-z} \\ P(1,t) &= 1 - e^{-z} \end{aligned}\right\} \Rightarrow P_1(y,t) = (1-y)e^{-z} + y(1-e^{-z})$$

$$= e^{-z} - ye^{-z} + y - ye^{-z}$$

$$= e^{-z}(1-2y) + y$$



Say $t_1 \le t_2$



$$P_2(0,t_1;0,t_2) = e^{-t_2}$$

$$P_2(0,t_1;1,t_2) = \int_{t_1}^{t_2} e^{-q}\,dq = -e^{-q}\Big|_{t_1}^{t_2} = -e^{t_2} + e^{-t_1}$$

$$P_2(1,t_1;0,t_2) = 0$$

$$P_2(1,t_1;1,t_2) = 1 - e^{-t_1}$$

$$P_2(y_1,t_1;y_2,t_2) = (1-y_1)(1-y_2)e^{-t_2} + (1-y_1)y_2(e^{-t_2}+e^{-t_1}) + y_1y_2(1-e^{-t_1})$$

etc.

We can also define conditional probabilities:

$$P_{1|1}(y_2, t_2 \mid y_1, t_1)$$

= prob that the system has output y_2 at time t_2 given output y_1 at time t_1.

Ex: $P_{1|1}(y_2, t_2 | y_1, t_1) = \dfrac{P_2(y_2, t_2; y_1, t_1)}{P(y_1, t_1)}$

$$= \frac{(1-y_1)(1-y_2)e^{-t_2} + (1-y_1)y_1(e^{-t_2}+e^{-t_1}) + y_1 y_2(1-e^{-t_1})}{e^{t_1}(1-2y)+y}$$

For example the prob. the output is 1 at time 2 given 0 at time 1 is

$$\frac{-e^{-2}+e^{-1}}{\cancel{e^{-1}}\,1-e^{-1}}$$

And

$$P_{1|1}(1,2|1,1) = \frac{1-e^{-t_1}}{1-e^{-t_1}} = 1.$$

And of course you can define

$$P_{l|k}(y_1, t_1; \cdots; y_l, t_l \,|\, y_{l+1}, t_{l+1}; \cdots; y_k, t_k).$$

The "generating functional" is

$$G(K) = \left\langle \exp\left(i \int_{-\infty}^{\infty} k(t)\, Y(t)\, dt \right) \right\rangle,$$

here, $k(t)$ is a converging function of time.

Expanding $G(k)$ gives the moments.

$$G(k) = \left\langle 1 + i \int k(t) Y(t) dt + \frac{i^2}{2!} \left(\int k(t) Y(t) dt \right)^2 + \cdots \right\rangle$$

$$= 1 + i \iint k(t) Y_x(t) dt\, P(x) dx + \frac{i^2}{2!} \iiint k(t_1) k(t_2) Y(t_1) Y(t_2) P_X(x) dx\, dt_1 dt_2 + \cdots$$

$$= 1 + i \int k(t) \langle Y(t) \rangle dt + \frac{i^2}{2!} \int k(t_1) k(t_2) \langle Y(t_1) Y(t_2) \rangle + \cdots$$

$$= \sum_{m=0}^{\infty} \frac{i^m}{m!} \int k(t_1) \cdots k(t_m) \langle Y(t_1) \cdots Y(t_m) \rangle.$$

Note: Stationary $\Longleftrightarrow$ $\forall n \; \forall \tau$

$$P_n(y_1, t_1; \dots; y_n, t_n) = P_n(y_1, t_1+\tau; \dots; y_n, t_n+\tau).$$

Note: Gaussian $\Longleftrightarrow$

$$G(k) = \exp\Big[i \int k(t) \langle Y(t) \rangle dt$$

$$- \frac{1}{2} \iint k(t_1) k(t_2) \langle\langle Y(t_1) Y(t_2) \rangle\rangle dt_1 dt_2 \Big]$$

IV. Markov Processes

A process has the Markov Property if for any set of successive times $t_1 \le t_2 \le \dots \le t_n$ we have

$$P(y_n, t_n | y_1, t_1; \dots; y_{n-1}, t_{n-1}) = \underbrace{P(y_n, t_n | y_{n-1}, t_{n-1})}_{\text{called the transition probability}}$$

Note:

$P(y_1,t_1; y_2,t_2; y_3,t_3)$

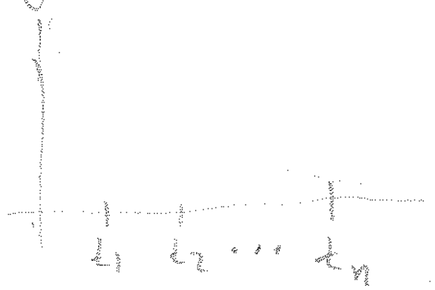
$$= P(y_1,t_1; y_2,t_2)\, P(y_3,t_3 \mid y_1,t_1; y_2,t_2)$$

$$= P(y_1,t_1)\, P(y_2,t_2 \mid y_1,t_1)\, P(y_3,t_3 \mid y_2,t_2).$$

And similarly for all n. Thus, A Markov Process is determined completely by $P(y_1,t_1)$ and $P(y_2,t_2 \mid y_1,t_1)$.

Ex. What is $P(y_n,t_n \mid y_1,t_1; \dots; y_{n-1},t_{n-1})$ $*$ for our running example?

Say all $y_i = 0$. Then



$$* = \frac{P(0,t_n; \quad ; 0,t_1)}{P(0,t_1; \dots; 0,t_{n-1})} = \frac{P(t_n < g)}{P(t_{n-1} < g)}$$

$$= P(0,t_n \mid 0,t_{n-1}). = \frac{e^{t_n}}{e^{t_{n-1}}} = e^{t_n - t_{n-1}}$$

etc.

IV.2 The Chapman-Kolmogorov Equation

Recall that a process is Markov if for all n and $t_1 \le t_2 \le \dots \le t_n$ it is the case that

$$P(y_n, t_n | y_1, t_1; \dots; y_{n-1}, t_{n-1}) = P(y_n, t_n | y_{n-1}, t_{n-1}).$$

We also found that, for a Markov Process, one only needs P_1 and $P_{1|1}$ to specify it. In particular, the rest of the hierarchy can be generated from these, as for example P_3:

$$P(y_1, t_1; y_2, t_2; y_3, t_3)$$

$$= P(y_1, t_1) P(y_2, t_2 | y_1, t_1) P(y_3, t_3 | y_2, t_2).$$

If we integrate over all y_2 we get

$$P(y_1,t_1\,;y_3,t_3) = P(y_1,t_1)\int P(y_2,t_2|y_1,t_1)\,P(y_3,t_3|y_2,t_2)\,dy_2$$

Divide by $P(y_1,t_1)$ and use Bayes Rule to get

$$\boxed{P(y_3,t_3|y_1,t_1) = \int P(y_2,t_2|y_1,t_1)\,P(y_3,t_3|y_2,t_2)\,dy_2}$$

which is valid for $t_1 \le t_2 \le t_3$. This eqn. is called the <u>Chapman-Kolmogorov Equation</u>.

<u>Theorem</u>:

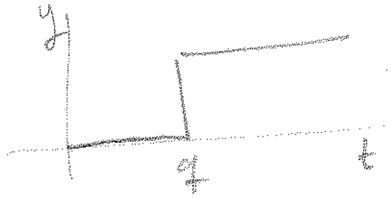
a) Any Markov Process obeys

i) The C.K.E.

ii) $P(y_2,t_2) = \int P(y_2,t_2|y_1,t_1)\,P(y_1,t_1)\,dy_1$

b) Any P_1 and $P_{1|1}$ obeying i and ii define a Markov Process.

Note: This is important: Not any P_1 and $P_{1|1}$ will specify a Markov Process.

Ex. 

Is specified by

$$P(0,t) = e^{-t}$$

$$P(1,t) = 1 - e^{-t}$$

$$P(0,t_2|0,t_1) = \frac{e^{-t_2}}{e^{-t_1}}$$

$$P(0,t_2|1,t_1) = 0$$

$$P(1,t_2|0,t_1) = \frac{-e^{-t_2} + e^{-t_1}}{e^{-t_1}}$$

$$P(1,t_2|1,t_1) = 1$$

In this case the CKE looks like

this a sum instead of an integral

$$P(0,t_3|0,t_1) = \sum_{y_2=0}^{1} P(0,t_3|y_2,t_2)\,P(y_2,t_2|0,t_1)$$

$$= \underbrace{\frac{e^{-t_3}}{e^{-t_2}} \cdot \frac{e^{-t_2}}{e^{-t_1}}}_{y_2=0} + \underbrace{0 \cdot \frac{e^{-t_2}}{e^{-t_1}}}_{y_2=1} = \frac{e^{-t_3}}{e^{-t_2}}$$

And similarly for each 00, 01, 10, 11.

This can all be compactly written as

$$A(t|t') = \begin{pmatrix} P(0,t|0,t') & P(0,t|1,t') \\ P(1,t|0,t') & P(1,t|1,t') \end{pmatrix}.$$

Then the CKE is

$$A(t_3|t_1) = A(t_3|t_2)A(t_2|t_1).$$

Note that condition ii is also satisfied:

E.g.

$$P(0,t_2) = P(0,t_2|0,t_1)P(0,t_1) + P(0,t_2|1,t_1)P(1,t_1)$$

$$= \frac{e^{-t_2}}{e^{-t_1}} e^{-t_1} + 0 \cdot (1-e^{-t_1}) = e^{-t_2}.$$

Compactly:

$$\begin{pmatrix} P(0,t_2) \\ P(1,t_2) \end{pmatrix} = A(t_2|t_1) \begin{pmatrix} P(0,t_1) \\ P(1,t_1) \end{pmatrix}.$$

Note: These matrix tricks work when Ω_Y is discrete.

Here are two other examples:

1 <u>Wiener Process</u>

$$P(y_2, t_2 | y_1, t_1) = \frac{1}{\sqrt{2\pi(t_2 - t_1)}} \exp\left[- \frac{(y_2 - y_1)^2}{2(t_2 - t_1)} \right]$$

$$P(y, 0) = \delta(y)$$

a) This obeys the CKE

$$P(y_3, t_3 | y_1, t_1) = \int \frac{1}{\sqrt{2\pi(t_3 - t_2)}\sqrt{2\pi(t_2 - t_1)}} \exp\left[- \frac{(y_3 - y_2)^2}{2(t_3 - t_2)} - \frac{(y_2 - y_1)^2}{2(t_2 - t_1)} \right] dy_2$$

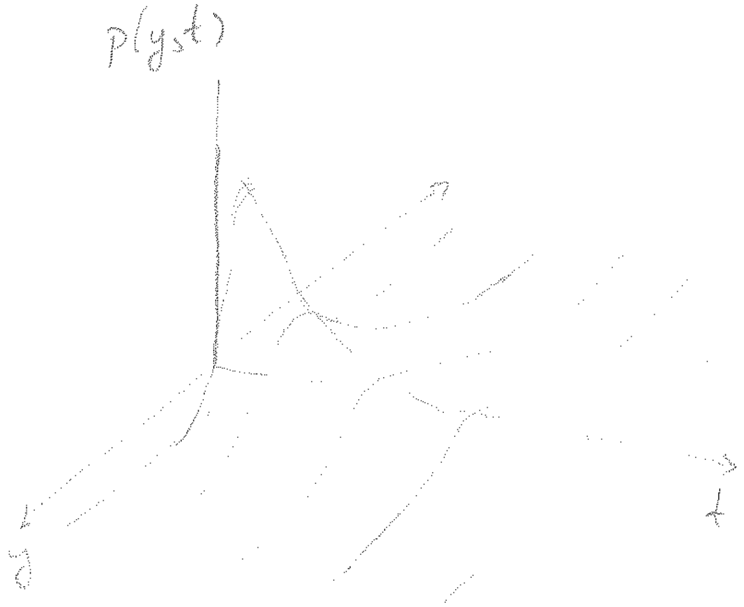
This is a pain to show.

a) Expand the exponent and complete the square

b) Factor out non-y_2 terms

c) Integrate the rest

or change coordinates



The solution can be found using (ii)

$$P(y,t) = \int P(y,t \mid y_1, 0)\, \underbrace{P(y_1, 0)}_{=\delta(y_1)}\, dy_1$$

$$= P(y, t \mid 0, 0)$$

$$= \frac{1}{\sqrt{2\pi t}}\, e^{\frac{y^2}{2t}}.$$

Note that from $P(y,t)$ we see that

$$\mu(t) = 0$$

$$\sigma^2(t) = t$$

So the process spreads out over time.

Note also that $P(y,t)$ is a solution to

$$\frac{d}{dt} P(y,t) = D \frac{\partial^2 P}{\partial y^2}.$$

This is a simple form of what's called the Fokker-Plank equation.

② Poisson Process

$$\Omega_Y = \{0, 1, 2, \dots\}$$

$$P_1(n,0) = \delta_{n,0} \quad \text{all mass at } 0$$

$$P_{1|1}(n_2,t_2 \mid n_1,t_1) = \frac{(t_2-t_1)^{n_2-n_1}}{(n_2-n_1)!} e^{-(t_2-t_1)}$$

if $n_2 \geq n_1$ and

$P_{1|1} = 0$ if $n_2 < n_1$

$\boxed{\text{Check i and ii for this.}}$

To decifer this, look at simple cases:

$$P_{1|1}(0,t\,|\,0,0) = e^{-t}$$ similar to the 2-state process we've been studying

$$P_{1|1}(1,t\,|\,0,0) = t e^{-t}$$

$$P_{1|1}(2,t+\tau\,|\,1,t) = \tau e^{-\tau}$$

$$P_{1|1}(n+1,t+\tau\,|\,n,t) = \tau e^{-\tau}$$

$$P_{1|1}(n+2,t+\tau\,|\,n,t) = \frac{\tau^2}{2} e^{-\tau}$$

$\vdots$

$$P_{1|1}(n+k,t+\tau\,|\,n,t) = \frac{\tau^k}{k!} e^{-\tau}$$ this is also the prob of getting k steps in τ seconds

$\vdots$

Another question is: What is the probability that a step occurs in the next dt seconds?

$$\frac{d}{dt} P(n+1,t\,|\,n,t) = \lim_{h\to 0} \frac{P(n+1,t+h\,|\,n,t) - P(n+1,t\,|\,n,t)}{h}$$

$$= \lim_{h\to 0} \frac{h e^{-h} - 0}{h} = 1.$$

Also,

$$\frac{d}{dt}P(n,t|n,t) = \lim_{h\to 0}\frac{P(n,t+h|n,t) - P(n,t|n,t)}{h}$$

$$= \lim_{h\to 0}\frac{e^{-h}-1}{h} = \lim_{h\to 0}\frac{-e^{-h}}{1} = -1$$

$$0 \xrightarrow{1} 1 \xrightarrow{1} 2 \xrightarrow{1} 3 \xrightarrow{1} \cdots$$

Using ii we can get P_1

$$P(n,t) = \sum_m P(n,t|m,0)\,P(m,0)$$

$$= P(n,t|0,0)$$

$$\frac{d}{dt}P(n,t) = \frac{d}{dt}P(n,t|0,0) = \lim_{h\to 0}\frac{P(n,t+h|0,0) - P(n,t|0,0)}{h}$$

$$= \lim_{h\to 0}\frac{1}{h}\left(\frac{(t+h)^n}{n!}e^{-t-h} - \frac{t^n}{n!}e^{-t}\right)$$

$$= \lim_{h\to 0}\frac{1}{n!}\left(n(t+h)^{n-1}e^{-t-h} - (t+h)^n e^{-t-h}\right)$$

$$= \frac{1}{n!}\left(nt^{n-1}e^{-t} - t^n e^{-t}\right) = \frac{t^{n-1}}{(n-1)!}e^{-t} - \frac{t^n}{n!}e^{-t}.$$

So,

$$\frac{d}{dt}P(n) = P(n-1) - P(n)$$

So we can write this:

$$\dot{p}_0 = -p_0$$
$$\dot{p}_1 = p_0 - p_1$$
$$\dot{p}_2 = p_1 - p_2$$
$$\dot{p}_3 = p_2 - p_3$$
$$\vdots$$

It is interesting to note that,

$$\frac{d}{dt}\langle n\rangle = \frac{d}{dt}\sum_{n=0}^{\infty} n p_n = \sum_{n=0}^{\infty} n \dot{p}_n$$

$$= 0(-p_0) + 1(p_0 - p_1) + 2(p_1 - p_2) + 3(p_2 - p_3) + \cdots$$

$$= p_0 - p_1 + 2p_1 - 2p_2 + 3p_2 - 3p_3 + \cdots$$

$$= p_0 + p_1 + p_2 + \cdots = 1$$

so $\frac{d}{dt}\langle n\rangle = 1 \Rightarrow \langle n\rangle = t$. This can also be checked using $P(n,t) = \frac{t^n}{n!}e^{-t}$.

IV.3 Stationary Markov Processes

See Davids Notes.

IV.4 The extraction of a subensemble

Let $Y(t)$ be a stationary Markov process given by $P_1(y_1)$ and $T_2(y_2|y_1)$.

Extracting a subensemble amounts to defining initial ~~condians~~ conditions. Fix t_0 and say that at time t_0 the values of $Y(t_0)$ are distributed according to ~~$p_0(y_0)$~~ $p(y_0)$. Then define a new process:

~~$P^*_{1,t}$~~

$$\tilde{P}_1(y_1, t_1) = \int T_{t_1 - t_0}(y_1 | y_0)\, p(y_0)\, dy_0$$

$$\tilde{P}_{1|1}(y_2, t_2 | y_1, t_1) = T_{t_2 - t_1}(y_2 | y_1)$$

Note that the transition matrix depends on Δt alone — hence it is called ~~ho~~ homogeneous in time.

The idea is that

$$\tilde{P}_1(y_1,t_1) \longrightarrow P_1(y_1) \quad \text{as} \quad t \to \infty$$

$$T_{t_1-t_0}(y_1|y_0) \longrightarrow P_1(y_1) \quad \text{"} \quad \text{"}$$

which is to say: the probability distribution stabilizes. We will prove this later.

Ex. Telegraph Revisited.

Recall the definition:

$$P_1(y) = \frac{1}{2}\delta(y-1) + \frac{1}{2}\delta(y+1)$$

$$T_\tau(y_2|y_1) = \frac{1}{2}(1+e^{-2\gamma\tau})\,\delta(y_2-y_1) + \frac{1}{2}(1-e^{-2\gamma\tau})\,\delta(y_2+y_1)$$

Because P_1 does not depend on t and T_τ doesn't depend on t_1, t_2, this process is stationary.

Note also that it spends $\frac{1}{2}$ its time in state $y=-1$ and half in $y=1$.

Extracting a subensemble amounts to setting a new initial condition. Say initially,

$$p(-1) = 1$$
$$p(1) = 0 \qquad \text{(for example)}.$$

Then, for this initial condition (when $t_0 = 0$)

$$\tilde{P}(y,t) = \int T_t(y, y_0)\, p(y_0)\, dy_0$$

$$= T_t(y, -1)$$

$$= \frac{1}{2}(1 + e^{-2\gamma t})\,\delta(y+1) + \frac{1}{2}(1 - e^{-2\gamma t})\,\delta(y-1)$$

Note that as $t \to \infty$ we get

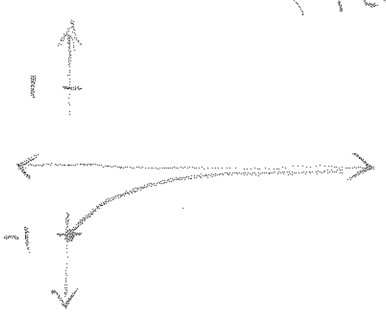
$$\lim_{t\to\infty} \tilde{P}(y,t) = \frac{1}{2}\delta(y+1) + \frac{1}{2}\delta(y-1) = P(y).$$

The book also shows how to use $\tilde{P}$ to show how the moments change

$$\langle \tilde{Y}(t) \rangle = \int y\, \tilde{P}(y,t)\, dy$$

$$= -\frac{1}{2}(1 + e^{-2\gamma t}) + \frac{1}{2}(1 - e^{-2\gamma t})$$

$$= -e^{-2\gamma t}$$



$$\langle \tilde{Y}(t)^2 \rangle = \frac{1}{2}(1+e^{-2\gamma t}) + \frac{1}{2}(1-e^{-2\gamma t})$$

$$= 1.$$

So

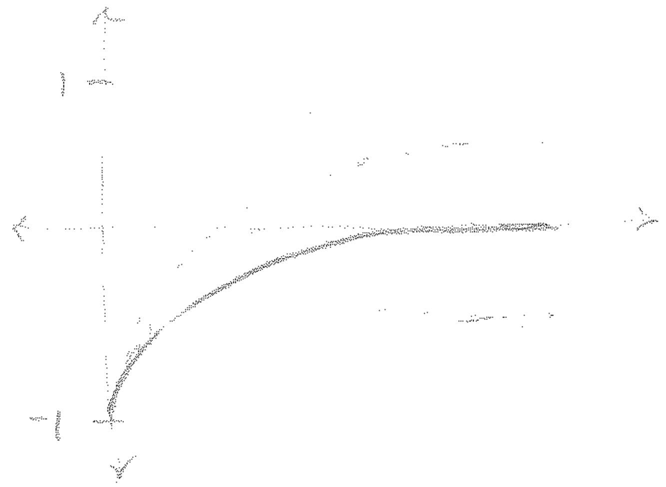
$$\langle\langle \tilde{Y}(t)^2 \rangle\rangle = \mu_2 - \mu_1^2$$

$$= 1 - (-e^{-2\gamma t})^2$$

$$= 1 + e^{-4\gamma t}$$

So when $t=0$ $\sigma^2(t) = 0$.

And as $t \to \infty$, $\sigma^2(t) \to 1$.



Note: For subensemble processes and when T only depends on $t_2 - t_1$, we call the process time-homogeneous.

Note: We will skip IV.5 for now.

IV.6 The Decay Process

- Suppose there are n_0 particles that decay, independently at some rate.

- Suppose that $w(t)$ is the probability that at time t a particular particle has not decayed.

- Then the prob. that n_1 particles have not decayed at time t_1 is

$$P_1(n_1,t_1) = \binom{n_0}{n_1} w^{n_1} (1-w)^{n_0-n_1}.$$

- If $w(t) = e^{-\gamma t}$ then

$$P_1(n_1,t_1) = \binom{n_0}{n_1} e^{-t_1 \gamma n_1} (1-e^{-\gamma t_1})^{n_0-n_1}.$$

- And

$$T_\tau(n_2|n_1) = \binom{n_1}{n_2} e^{-\gamma\tau n_2} (1-e^{-\gamma\tau})^{n_2-n_1}.$$

V.1 The Master Equation

The Chapman-Kolmogorov equation is a complex way to specify a process. It essentially assumes you know the solution $P(y,t)$ of a system. This is rarely the case. A more convenient form is the Master Equation (so named because it is the equation from which all others are derived). Here is where it comes from:

Start with the transition function time-homogeneous Markov Process:

$$T_\tau(y_2|y_1) = [\text{Prob. that no events occur}]\,\delta(y_3 - y_1)$$

$$+ [\text{Prob that 1 event occurs}]$$

$$+ \underbrace{[\text{Prob that 2 or more events occur}]}_{\text{goes to 0 as } \tau \to 0}$$

The prob. that a $y_1 \to y_2$ event occurs in time $(t, t+\tau)$ is

$$\underbrace{w(y_2|y_1)}_{\substack{\text{transition probability} \\ \text{per unit time for} \\ y_1 \to y_2.}}\,\tau$$

The probability that no events occur in time $(t, t+\tau)$ is

$$1 - \underbrace{\tau \int w(y|y_1)\,dy}_{\text{book calls this } a_0(y_1).}$$

So,

$$T_\tau(y_2|y_1) = \left[1 - \tau \int w(y|y_1)\,dy\right] \delta(y_2 - y_1)$$

$$+ \tau\, w(y_2|y_1) + o(\tau).$$

Next, we use the C.K.E.

$$T_{t+\tau}(y_3|y_1) = \int T_\tau(y_3|y_2)\,T_t(y_2|y_1)\,dy_2$$

$$= \int \left(\left[1 - \tau \int W(y|y_2)\,dy\right] \underbrace{\delta(y_3 - y_2)}_{\neq 0 \text{ when } y_2 = y_3} + \tau W(y_3|y_2) \right) T_t(y_2|y_1)\,dy_2$$

$$= \left[1 - \tau \int W(y|y_3)\,dy\right] T_t(y_3|y_1) + \tau \int W(y_3|y_2)\,T_t(y_2|y_1)\,dy_2$$

Now divide by τ:

$y \mapsto y_2$ (dummy variable)

$$\underbrace{\frac{T_{t+\tau}(y_3|y_1) - T_t(y_3|y_1)}{\tau}}_{= \frac{\partial}{\partial t} T_t(y_3|y_1)} = \int \left(W(y_3|y_2)\,T_t(y_2|y_1) - W(y_2|y_3)\,T(y_3|y_1) \right) dy_2$$

Take y_1 to be some deterministic initial value.
Then

$$T_t(y|y_1) = P(y,t).$$

Thus, we arrive at the Master Equation:

$$\boxed{\frac{\partial}{\partial t} P(y,t) = \int \Big(\underbrace{W(y|y')P(y',t)}_{\substack{\text{gain } \underline{\text{to}} \text{ } y \\ \text{from other} \\ \text{states } y'.}} - \underbrace{W(y'|y)P(y,t)}_{\substack{\text{loss } \underline{\text{from}} \\ y \text{ to other} \\ \text{states } y'.}} \Big) dy}$$

If the states are discrete then this is a sum:

$$\frac{d}{dt} p_n(t) = \sum_m \Big(W_{n,m} p_m(t) - W_{m,n} p_n(t) \Big)$$

Where $W_{n,m}$ is the rate at which probability flows from state m to state n. This can be written as

$$\dot{\vec{p}} = W \vec{p}$$

where $\vec{p} = (p_1, \ldots, p_n)^T$ and

$$W_{n,n} = -\sum_{m \neq n} W_{m,n} .$$

Example: The telegraph process:

$$\tilde{P}(y,t) = \frac{1}{2}\left(1+e^{-2\gamma t}\right)\delta(y+1) + \frac{1}{2}\left(1-e^{-2\gamma t}\right)\delta(y-1)$$

Or written differently:

$$p_{-1}(t) = \frac{1}{2}\left(1+e^{-2\gamma t}\right)$$

$$p_{1}(t) = \frac{1}{2}\left(1-e^{-2\gamma t}\right)$$

And

$$\dot{p}_{-1} = -\gamma e^{-2\gamma t} = \gamma p_1 - \gamma p_{-1}$$

$$\dot{p}_{1} = \gamma e^{-2\gamma t} = \gamma p_{-1} - \gamma p_1$$

Or

$$\dot{\vec{p}} = \begin{pmatrix} -\gamma & \gamma \\ \gamma & -\gamma \end{pmatrix} \vec{p}$$

Or

$$\begin{tikzcd} -1 \arrow[r, bend left, "\gamma"] & 1 \arrow[l, bend left, "\gamma"] \end{tikzcd}$$

Example: The $0 \to 1$ process is just

$$\textcircled{0} \xrightarrow{1} \textcircled{1} \iff \dot{\vec{p}} = \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix} \vec{p}$$

Example: The decay process

$$\textcircled{n_0} \xrightarrow{\gamma n_0} \textcircled{n_0-1} \xrightarrow{\gamma(n_0-1)} \textcircled{n_0-2} \xrightarrow{\gamma(n_0-2)} \cdots \xrightarrow{2\gamma} \textcircled{1} \xrightarrow{\gamma} \textcircled{0}$$

$$\boxed{\dot{p}_n = \gamma(n+1)p_{n+1} - \gamma n p_n}$$

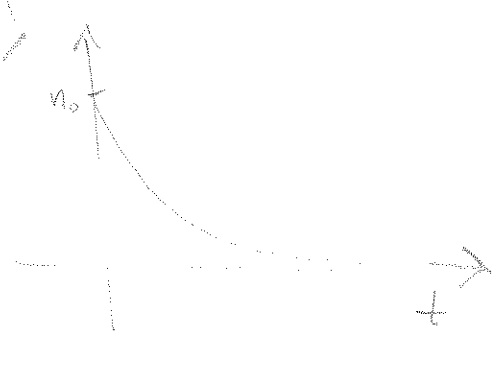
The average equations (we obtained before) can be more easily found:

$$\frac{d}{dt}\langle n \rangle = \sum n \dot{p}_n = \sum_{n=0}^{\infty} n\gamma(n+1)p_{n+1} - \gamma \sum_{n=0}^{\infty} n^2 \gamma p_n$$

$$\underset{n \mapsto n-1}{=} \gamma \sum \left((n-1) n p_n - n^2 p_n \right)$$

$$= -\gamma \sum n p_n = -\gamma \langle n \rangle$$

$$\Rightarrow \quad \langle n(t) \rangle = n_0 e^{-\gamma t}$$



V.1 Continued: Some Useful Constructions of Master Equations

Ex. 1: We have already worked out systems consisting of graphs with

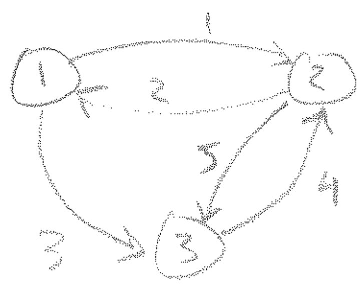
1. no self loops
2. edges with real-valued rates associated with them.

The general construction is to have

$$W_{j,i} = \text{rate from } i \text{ to } j$$

$$W_{i,i} = -\sum_{j \neq i} W_{j,i}$$

e.g.



$$W = \begin{pmatrix} -4 & 2 & 0 \\ 1 & -7 & 4 \\ 3 & 5 & -4 \end{pmatrix}$$

and

$$\dot{\vec{p}} = W\vec{p}$$

Ex 2: Continuous Space
Discrete updates

Say we have a process whose state $x \in \mathbb{R}$ is updated according to $x \mapsto f(x)$ at a rate K. What is the Master Equation? Take, for example, $f(x) = ax+b$.

Then,

$$W(x|y) = K\delta(ay+b-x)$$
$$W(y|x) = K\delta(ax+b-y)$$

and

$$\frac{d}{dt}p(x) = \int \left(W(x|y)p(y) - W(y|x)p(x) \right) dy$$

$$= K\int \delta(ay+b-x)p(y)\,dy - \underbrace{K\int \delta(ax+b-y)p(x)\,dy}_{Kp(x)}$$

$$\boxed{\text{put } u = ay+b-x \quad y = \frac{u-b+x}{a} \\ du = a\,dy}$$

$$= K\int \delta(u)\,p\left(\frac{u-b+x}{a}\right)\frac{du}{a} - Kp(x)$$

$$= \boxed{\frac{K}{a}\,p\left(\frac{x-b}{a}\right) - Kp(x)}$$

Using the Master Equation, you can get the moment dynamics:

$$\frac{d}{dt}\langle x\rangle = \int x \frac{d}{dt} p(x)\,dx$$

$$= \int x \frac{k}{a} p\left(\frac{x-b}{a}\right) dx - \overbrace{\int kx\, p(x)\, dx}^{k\langle x\rangle}$$

put $u = \frac{x-b}{a}$, $au+b = x$

$a\,du = dx$

$$= \frac{k}{a}\int (au+b)\, p(u)\, a\, du - k\langle x\rangle$$

$$= ka\langle x\rangle + kb - k\langle x\rangle$$

$$= k(a-1)\langle x\rangle + kb$$

At equilibrium, $\langle x\rangle^* = \frac{b}{1-a}$.

The system is stable when $a-1<0$

$\iff a<1$

[See Mathematica]

Similarly,

$$\langle x^2\rangle = \frac{k}{a}\int x^2\, p\left(\frac{x-b}{a}\right)dx - k\langle x^2\rangle$$

$$= k\int (au+b)^2\, p(u)\,du - k\langle x^2\rangle$$

$$= k\left\langle (ax+b)^2 - x^2\right\rangle$$

$$= k\left\langle (a^2-1)x^2 + 2abx + b^2\right\rangle$$

$$= k(a^2-1)\langle x^2\rangle + 2abk\langle x\rangle + kb^2$$

At equilibrium:

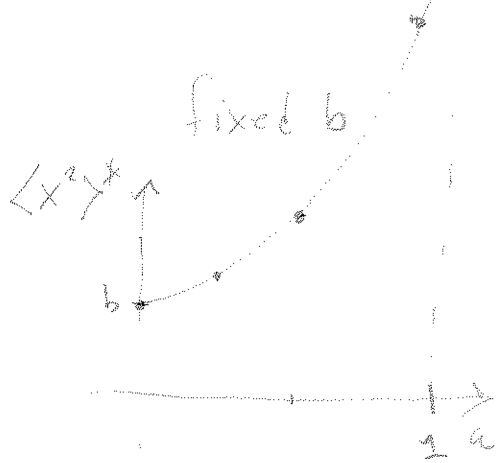
$$\langle x^2\rangle^* = \frac{2ab\,\langle x\rangle^* + b^2}{1-a^2}$$

$$= \frac{2ab\left(\frac{b}{1-a}\right) + b^2}{1-a^2}$$

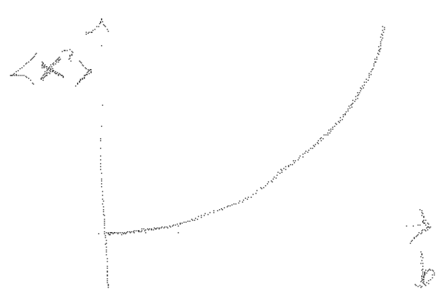
$$= \left(\frac{b}{1-a}\right)^2$$

so

$$\sigma^* = \mu_2 - \mu_1^2 = \left(\frac{b}{1-a}\right)^2 - \left(\frac{b}{1-a}\right)^2 = 0$$



fixed a



Ex. Master Eqn. for a Deterministic ODE

Say $\dot{x} = a$. Pick $\varepsilon \ll 1$.

$$\xrightarrow{\frac{a}{\varepsilon}} x-\varepsilon \xrightarrow{\frac{a}{\varepsilon}} x \xrightarrow{\frac{a}{\varepsilon}} x+\varepsilon \xrightarrow{\frac{a}{\varepsilon}}$$

The above model looks like a discrete space Markov Process. We write

$$\frac{d}{dt} p(x+\varepsilon) = \frac{a}{\varepsilon}\left(p(x) - p(x+\varepsilon) \right)$$

As $\varepsilon \to 0$ this is

$$\boxed{\frac{d}{dt} p(x) = -a \frac{\partial p(x)}{\partial x}}$$ ← master eqn for continuous ODE.

E.g. $$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-at)^2}{2}}$$

Is a solution to this.

We can find the moments as before

$$\frac{d}{dt}\langle x\rangle = -a\int x\,\frac{\partial p(x)}{\partial x}\,dx$$

$$= -a\left(x\,p(x)\Big|_{-\infty}^{\infty} - \int p(x)\,dx \right)$$

$$= a \quad \Rightarrow \quad \langle x\rangle = x_0 + at$$

$$\frac{d}{dt}\langle x^2\rangle = -a\int x^2\,\frac{\partial p(x)}{\partial x}\,dx$$

$$= -a\left(x^2 p(x)\Big|_{-\infty}^{\infty} - \int 2x\,p(x)\,dx \right)$$

$$= 2a\langle X\rangle = \langle x$$

So $$\frac{d}{dt}\sigma^2 = \frac{d}{dt}\left(\langle X^2\rangle - \langle X\rangle^2 \right)$$

$$= 2a\langle X\rangle - 2\langle X\rangle \frac{d}{dx}\langle X\rangle$$

$$= 2a\langle X\rangle - 2a\langle X\rangle = 0 \quad ✓$$

Ex. A Master Equation for a continuous system with Discrete Jumps

$$\boxed{\dot{x} = 1} \quad \frac{k}{x \mapsto x/2}$$

$$\frac{d}{dt} p(x) = -\frac{\partial p(x)}{\partial x} + 2a\, p(2x) - k\, p(x)$$

$$\frac{d}{dt}\langle x \rangle = -\int x \frac{\partial p(x)}{\partial x} dx + \int x \left(2a p(2x) - k p(x)\right) dx$$

$$= 1 - \frac{k}{2}\langle x \rangle$$

$$\frac{d}{dt}\langle x^2 \rangle = 2\langle x \rangle - \frac{3}{4}k\langle x^2 \rangle$$