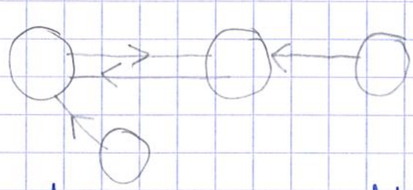
PCS Master - M2

STOCHASTIC PROCESSES

G^al introduction

A stochastic process is a collection of random variables, indexed by a q^ty that can be viewed usually as "time" (discrete, or continuous).

Ex : - position of a macromolecule, or an ensemble of ~, as a function of time, in a solution
- intensity of a light bulb
- traffic on a motorway, stockprice, rainfall...
- PageRank algorithm (random walk on a network), the first that was used by google. The web is viewed as an oriented network of pages, with links between them



A random walk is used to assign a weight (importance) to every page, containing a given set of words → ranks pages by relevance.
A random walker starts from an arbitrary page (not important), and jumps with proba $d = 82\%$ to another page by choosing a random link from the page it is currently visiting; with proba $1-d = 18\%$, the walker jumps to a random page in the network. d is called the damping factor.
After many iterations (jumps), the walker's probability vector to sit on each page reaches a constant → steady state. The page with largest proba is ranked as the most relevant one. Larry Page → 6^th richest in world (120 Billion$ 2021)

Applications in many disciplines : φ, χ, bio, economy, social sciences, computer sciences ⊕ this is a branch of mathematics, since 1920s

GOAL : characterize the process $x(t)$, ie specify the fluctuation properties either at fixed time, or through time → $p(x_1, x_2, \cdots, x_n)$
↳ $p(x,t)$, $\langle x(t)\rangle$, $V(x_t) = \langle x^2_t\rangle - \langle x_t\rangle^2$

time t_1, t_2, time t_n

$\langle x(t_1)x(t_2)\rangle - \langle x(t_1)\rangle\langle x(t_2)\rangle$

→ introduce the tools to understand phenomena ruled by randomness.

Many pbs have an element of randomness, coming from an ignorance, or chosen from the outset (as in PageRank) → best analyzed with a probabilistic model

→ establish bridges with statistical physics, information theory (computer science) and math: fluctuation dissipation, entropy ⇔ information ⇔ large deviations

Outline

- Reminder proba: CLT and its breakdown, extreme value statistics
- Large deviations
- Langevin eq / Brownian motion
- Fluctuations close to equil., FDT
- Markov processes, Fokker Planck equat, Ito, Stratonovich, Döblin, Kramers
- Fluctuation theorems / stochastic thermod
- First passage properties
- Martingale theory → link to math
- Functionals of Brownian motion

This is not a mathematical course: we shall try to be exact though, when we are not rigorous

Prerequisites:
- eq stat mech
- basic proba, key laws (Bernoulli, Gaussian, Poisson)
- honest math: complex calculus, linear algebra

See web for prerequisites & exercises

Notations: $=$, $\equiv$, $\simeq$

$\sim$ for mathematical equivalent ($f \sim g$ means $f/g \to 1$ in some limit)

$\propto$ proportional

$\stackrel{.}{=}$ equal to leading exponential order: $e^{x+\sqrt{x}} \stackrel{.}{=} e^{x}$ for $x \to \infty$

i.e. $f(x) \stackrel{.}{=} g(x)$ for $\frac{\log f(x)}{\log g(x)} \longrightarrow 1$.

For n even $\binom{n}{n/2} \stackrel{.}{=} 2^{n}$

$\propto 2^{n}/\sqrt{n}$

$\sim 2^{n}\sqrt{\frac{2}{\pi n}}$

I) A PROBABILITY BOOTSTRAP

Goal: when does the bell shape prevail (central limit theorem)?
what can go wrong?

1°/ Moments, cumulants, and their generating functions

a) Def

X is a random variable of density $p_X(x)$: $\Pr[X \text{ in } [a,b]] = \int_a^b p_X(x)\,dx$

$$\langle f(X)\rangle = \int_{-\infty}^{+\infty} f(x)\, p_X(x)\, dx = E[f(X)] \quad ; f \text{ is some function}$$

$$\hat{p}_X(k) \equiv \langle e^{ikX}\rangle$$

characteristic function, Fourier transform

$$|\hat{p}_X(k)| \leq \int |e^{ikx}|\, p_X(x)\,dx \leq 1 = \hat{p}_X(0)$$

Here, $k \in \mathbb{R}$; can be complex...

Take X, density p_X; Y density p_Y: $Z \equiv X + Y$

$$p_Z(z) = \int dx\,dy\; p_X(x)\, p_Y(y)\; \delta(z = x+y)$$

$$\Rightarrow \hat{p}_Z(k) = \langle e^{ikZ}\rangle = \int e^{ikz}\, \delta(z-x-y)\, p_X(x)\, p_Y(y)\, dx\,dy\,dz$$

$$= \langle e^{ikX}\rangle\langle e^{ikY}\rangle$$

$$= \hat{p}_X(k)\, \hat{p}_Y(k)$$

$\hat{p}(k)$ is related to the moment generating function, usually defined as

$$\langle e^{tX}\rangle = \hat{p}_X(-it) = 1 + t\langle X\rangle + \frac{t^2}{2}\langle X^2\rangle + \frac{t^3}{3!}\langle X^3\rangle \dots$$

$$\langle X^n\rangle = \frac{d^n}{dt^n}\langle e^{tX}\rangle\Big|_{t=0} \quad \text{are the moments}$$

From this, the cumulant generating function is defined as

$$K_X(t) = \log\langle e^{tX}\rangle = t\, C_1 + \frac{t^2}{2} C_2 + \frac{t^3}{3!} C_3 + \dots$$

↳ cumulants

If the moments exist:

$$C_1 = \langle X\rangle$$

$$C_2 = \langle X^2\rangle - \langle X\rangle^2 = \langle (X - \langle X\rangle)^2\rangle \equiv V(X) \quad \text{variance}$$

$$C_3 = \langle X^3\rangle - 3\langle X^2\rangle\langle X\rangle + 2\langle X\rangle^3 = \langle (X-\langle X\rangle)^3\rangle$$

$$C_4 = \langle X^4\rangle - 4\langle X^3\rangle\langle X\rangle - 3\langle X^2\rangle^2 + 12\langle X^2\rangle\langle X\rangle^2 - 6\langle X\rangle^4$$

$$\neq \langle (X-\langle X\rangle)^4\rangle \Rightarrow \text{the "shifted variable" pattern breaks}$$

$$= \langle (X-\langle X\rangle)^4\rangle - 3\langle (X-\langle X\rangle)^2\rangle^2$$

Sum rule: $\sum$ coef $= 0$ since for a $\delta(x-x_0)$, all cumulants $= 0$ and moments all in x_0^n

f) Examples and use

→ Take Poisson distribution of param λ

$$\boxed{\Pr[N=n] = e^{-\lambda}\frac{\lambda^n}{n!}}, \quad n = 0, 1, \dots$$

$$\langle e^{tN}\rangle = \sum_{n\geq 0} e^{-\lambda}\frac{\lambda^n}{n!}e^{tn} = \exp\left[\lambda\left(e^t - 1\right)\right]$$

$$= \lambda\left(t + \frac{t^2}{2} + \frac{t^3}{3!} + \dots\right)$$

⇒ all cumulants $c_n = \lambda$.

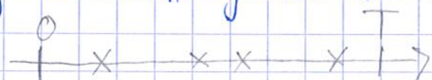
The sum of 2 Poisson variables $N \to \mathcal{P}(\lambda)$ and $M \rightsquigarrow \mathcal{P}(\mu)$, independent is $\mathcal{P}(\lambda+\mu)$, visible from charact function / moment generating function

Indeed, $\mathcal{P}(\lambda)$ holds for a counting statistics

* independent events
* occuring at random ; the # of events between t and $t+T$ only depends on T
* 2 events cannot happen at same time

denoting r the rate, and counting N, the # of events in time duration T:

$n \rightsquigarrow \mathcal{P}(rT)$



$P_n(t+\delta t)$: proba to count n events in time $t+\delta t$

$$= P_n(t)\left[1 - r\,\delta t\right] + P_{n-1}(t)\, r\,\delta t$$

$$\Rightarrow \begin{cases} \dfrac{dP_n}{dt} = r\left(P_{n-1} - P_n\right) \\[2mm] \dfrac{dP_0}{dt} = -r P_0 \end{cases}$$

show by recurrence that $P_n(\tau) = e^{-\lambda}\dfrac{\lambda^n}{n!}$

$\lambda = r\tau$

Thus, it becomes clear that $\mathcal{P}(\lambda) + \mathcal{P}(\mu) = \mathcal{P}(\lambda+\mu)$: we observe the same phenomenon over time τ_1 $(\lambda = r\tau_1)$, then time τ_2 $(\mu = r\tau_2)$ and finally over time $\tau_1+\tau_2$ (thus param $\lambda+\mu$).

Also : the time between 2 events, t_i, has pdf $\quad p(t) = r e^{-rt} \quad (t\geq 0)$

denoting $T_n = \sum_{i=1}^{n} t_i$; t_i IID exponential

T_n has density $\quad p_n(\tau) = e^{-r\tau}\dfrac{(r\tau)^{n-1}}{(n-1)!}\, r$

Indeed, $\underset{\text{in a time } t}{\Pr(N \geq n)} = \underbrace{\int_0^t p_n(\tau_n)\, d\tau_n}_{\Pr(T_n \leq t)} = \sum_{N=n}^{\infty} e^{-rt}\dfrac{(rt)^N}{N!}$

→ cumulants are useful for their gaussian properties

Marcinkiewicz theorem (1939), says that a cumulant generating function (in the multivariate case $\langle \exp(i\vec{k}\cdot\vec{X})\rangle$) cannot be a polynomial of degree larger than 2 in the $\vec{k}$ components. Thus, either all cumulants but the first 2 vanish (and we deal with a gaussian), or there is an ∞ number of non vanishing cumulants

→ cumulants also useful because they characterize the shape of the pdf

c_2 → variance, dispersion around the mean

Ex: Levothyrox (levothyroxine) used to treat thyroid hormone deficiency (hypothyroidism), sold by Merk: new drug had same mean response of patients but ↗ variance ⇒ huge pb

c_3 → skewness (asymmetry)

c_4 → "flatness"

c_3 and c_4 can be put in a form that does not depend on a rescaling of the variable X. Note that they do not depend on a shift of X since $X - \langle X\rangle$ does not. Indeed, from a variable X that is time independent, we define

$$Y_t = b(t)\,X + a(t) \quad ; \quad a(t) \text{ and } b(t) \text{ arbitrary function}$$

then p_Y depends on t, and so do the cumulants

$$V(Y_t) = V(b(t)X) = b^2 V(X) = b^2 c_2^X = c_2^Y$$

$$c_3^Y = \langle (Y - \langle Y\rangle)^3\rangle = b^3 \langle (X - \langle X\rangle)^3\rangle = b^3 c_3^X$$

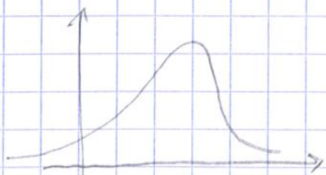
$$c_4^Y = b^4 c_4^X$$

Thus $$\frac{c_3^Y}{(c_2^Y)^{3/2}} = \frac{\langle (Y - \langle Y\rangle)^3\rangle}{\langle (Y - \langle Y\rangle)^2\rangle^{3/2}} = \frac{c_3^X}{(c_2^X)^{3/2}} \equiv \text{skewness}$$

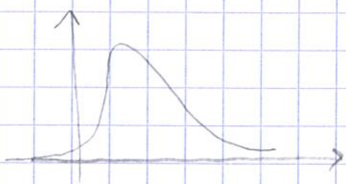
$$\frac{c_4^Y}{(c_2^Y)^2} = \frac{c_4^X}{(c_2^X)^2} \equiv \underbrace{\text{kurtosis}}_{= \frac{\langle (X - \langle X\rangle)^4\rangle}{[V(X)]^2}} - 3$$

do not depend on shift/rescaling; depend only on shape

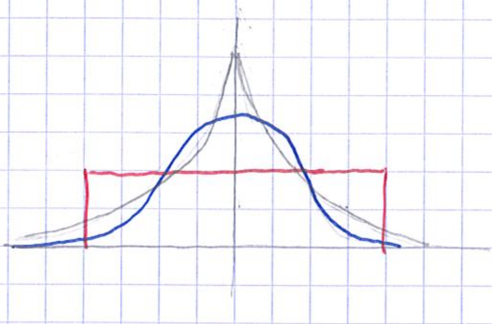
↳ intrinsic



skewness < 0



skewness > 0



	Skewness	Kurtosis	kurt − 3
$p(x) \propto e^{-\lvert x\rvert}$	0	6	3
Gaussian	0	3	0
window	0	$+\frac{9}{5}$	$-\frac{6}{5}$

c) From distributions to moments / cumulants ... and back ?

In general, the moments are not sufficient to characterize the distribution. Think about r.v. for which moments do diverge, possibly all ... there are many. Yet, given the moments, the pdf is uniquely determined if the series

$$1 + t\langle X\rangle + \frac{t^2}{2}\langle X^2\rangle + \frac{t^3}{3!}\langle X^3\rangle + \cdots$$ has a finite radius of convergence

In this case, the characteristic function is analytic and thus its log also is → the cumulant generating function

d) Broad distributions ("fat tails")

If the proba of large events decreases fast enough (exp, gaussian, window...) then all moments exist. But ∃ many cases where decrease is slow, as power-law

$$p(x) \propto \frac{1}{|x|^{1+\mu}} \qquad \text{for } |x| > x^*, \text{ some threshold}$$

μ is called the Lévy index ; $\mu > 0$ to ensure normalization

These power-law tails are called Pareto tails. Pareto was studying income and wealth distribution, and found $\mu \simeq 3/2$

Problem : only moments of order $< \mu$ do exist.

Ex Student distribution

$$p(x) = \mathcal{N}\frac{1}{\left(1+\frac{x^2}{\mu}\right)^{\frac{1+\mu}{2}}} \qquad \mu = 1 \text{ is Cauchy distribution}$$

Important in hypothesis testing, and statistical significance estimation

Note that for $\mu \to \infty$, $\left(1+\frac{x^2}{\mu}\right)^{-\mu/2} \propto \exp\left(-x^2/2\right)$: Gaussian recovered