Transition P

$P_{ij} = \mathbb{P}(X_{t+1} = j \mid X_t = i)$

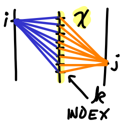
$P = \begin{bmatrix} & & \\ & P_{ij} & \\ & & \end{bmatrix} \to j$ (row i)

Remember: down (i), then right (j)

Chapman-Kolmogorov $\Longleftrightarrow$ Matrix Form

$$P^{n+m}_{ij} = \sum_{k \in \mathcal{X}} P^n_{ik} P^m_{kj}$$

like the law of total probability



Recursive

$P^t = \underbrace{P_0 P_0 \dots P}_{t \times \text{composition}}$

Application to Marginals

$\alpha^{(t)}_i := \mathbb{P}(X_t = i)$

$\alpha^{(t)} = \alpha^{(0)} P^{(t)}$

States

Accessible $\longrightarrow$ Communicate $(i \leftrightarrow j)$

$(i \to j$, j is accessible from $i)$

$(\exists t \geq 0)$ s.t. $\mathbb{P}(X_t = j \mid X_0 = i) > 0$

$\Longleftrightarrow P^t_{ij} > 0$

$\Longleftrightarrow \exists$ path $(i, \dots, j) \in \mathcal{X}^t$ s.t. $P_{i k_1} \cdot P \cdots P_{k_{t-1} j} > 0$

You will use comm. classes to form a partition of communicating areas of the graph *

Divide in classes, then either all in it are recurrent or transient

$f_i := \mathbb{P}\left(\bigcup_{t \geq 1} \{X_t = i\} \mid X_0 = i\right)$ — P of eventually being at i again in the future — does not require leaving

- Recurrence $\to f_i = 1$
- Transience $\to f_i < 1$

Return Time $\to S_k = \begin{cases} \text{time of } k_{th}\text{-return to } i \text{ of } (X_t)_{t \geq 0} \\ \infty \quad \text{if last return was at } j < k \\ 1 \quad k = 0 \end{cases}$

Note $S_k = \min\{t \geq S_{k-1} + 1 : X_t = i\}$

Time between returns $\to T_k := S_k - S_{k-1}$ $\left(\begin{matrix} T_k = \infty \\ \text{if } S_k = \infty \end{matrix}\right)$

Note T_i iid RV, result of deterministic function on X_k

Assuming $S_{k-1} < \infty$: $\operatorname{supp}(T_i) = \mathbb{N} \cup \{\infty\}$

Total number of visits to i over all t: $V_i = \sum_{t=0}^{\infty} \mathbb{1}(X_t = i)$, V_i is RV on $\mathbb{N} \cup \{\infty\}$

Probability of returning at least r times to r:

$\mathbb{P}(V_i > r \mid X_0 = i) = (f_i)^r, \quad r \geq 0$

- Recurrent $\Rightarrow \mathbb{P}(V_i = \infty) = 1$ (almost sure)
- Transient $\Rightarrow V_i \sim \text{Geom}(1 - f_i)$

Equivalent, easier checks:

Recurrent $\Longleftrightarrow \sum_{t=0}^{\infty} P^t_{ii} = \infty$

Transient $\Longleftrightarrow \sum_{t=0}^{\infty} P^t_{ii} < \infty$

$i \leftrightarrow j \Longrightarrow$ Both (i, j) are either transient or recurrent *

You can also find a converging upper bound or a diverging lower bound

- Irreducible $\to$ All states communicate with each other
- Closed $\to$ Accessible but you cannot leave (a dead end)

$0 \rightleftarrows 0 \xrightarrow{\ p\ } 0$

Proof

$$\mathbb{E}[V_i \mid X_0 = i] = \mathbb{E}\left[\sum_{t=0}^{\infty} \mathbb{I}(X_t = i) \mid X_0 = i\right] = \sum_{t=0}^{\infty} \mathbb{E}[\mathbb{I}(X_t = i) \mid X_0 = i] = \sum_{t=0}^{\infty} P^t_{ii}$$

Limiting Distribution

$$\pi_j = \lim_{t \to \infty} P^t_{kj} \overset{\text{Chapman Kolmogorov}}{=} \lim_{t \to \infty} \sum_{i \in \mathcal{X}} P^{t-1}_{ki} P_{ij} = \sum_{i \in \mathcal{X}} \left(\lim_{t \to \infty} P^{t-1}_{ki}\right) P_{ij}$$

$\pi_j = \sum_{i \in \mathcal{X}} \pi_i P_{ij}$ (Stationary) → limiting ⇓ stationary

$\pi P = \pi$

- $|\mathcal{X}| < \infty \Rightarrow$ stationary distribution exists
- $(X_t)_{t \geq 0}$ irreducible $\Rightarrow$ at most one stationary distribution

Procedure:

1. Find stationary $\pi = \pi P$
2. Show convergence:
   - → irreducible
   - → irreducible + aperiodic

π_i = mean proportion of $(X_t)_{t \geq 0}$ at i (end state IP)

Assuming $(X_t)_{t \geq 0}$ irreducible

visits to i by X_t before n: $V_i(n) = \sum_{t=0}^{n-1} \mathbb{1}(X_t = i)$

Mean time it takes to return: $m_i = \mathbb{E}[T_1 \mid X_0 = i]$ — same as $V_i(n) = \frac{n}{m_i}$

- $X_0 \sim$ any distribution $\Rightarrow \mathbb{P}\left(\frac{V_i(n)}{n} \to \frac{1}{m_i}\right) = 1$ (a.s. convergence)
- $X_0 \sim \pi \Longrightarrow \frac{1}{m_i} = \pi_i \quad \forall x \in \mathcal{X}$ | $\frac{1}{m_i}$ = mean permanence at that state in one t (like $f = \frac{1}{T}$)

Aperiodic: $\forall i \in \mathcal{X}$ the $\gcd(\{t \in \mathbb{N}^+ : P^t_{ii} > 0\}) = 1$

($\gcd$ = biggest loop size explaining all returns; t where $P(\text{return}) > 0$)

⊕ Irreducible = **Convergence Thm**: $P^t_{ij} \to \pi_{ij}$ as $t \to \infty$

AKA: if aperiodic, stationary, stationary $\Longleftrightarrow X_t \rightsquigarrow \pi$

Ergodic Theorem: (sample $\mathbb{E}$ goes to stationary)

$(X_t)_{t \geq 0}$ irreducible MC, aperiodic

π stationary. $\forall g : \mathcal{X} \to \mathbb{R}$ bounded

$\frac{1}{n} \sum_{t=0}^{n-1} g(x_t) \to \mathbb{E}_\pi[g]$

LLN for MC, no need for aperiodicity

$|\mathcal{X}| < \infty \to \exists \pi : \pi = \pi P$

↳ irreducible $\to \exists! \pi$, ergodic thm.

↳ aperiodic $\to$ conv. thm

Counter-ex to convergence thm:

- Gambler's ruin

$\to$ 3x class: $\{0\}\{\dots\}\{N\}$

$|\mathcal{X}| = \mathbb{Z}$, no $\pi = P\pi$

$P_{ij} = \begin{cases} j = i+1 & p \\ j = i-1 & 1-p \\ & 0 \end{cases}$ — there is no limiting distribution (dispersion of probabilities)

$\to$ hitting time

$h(i) = \lim_{t \to \infty} P^t_{iN} = \mathbb{P}(X_t \text{ hits } N \text{ before } 0 \mid X_0 = i)$

$\lim_{t \to \infty} P^t_{ij} = 0$ — not a distribution

Reversibility is stronger than stationarity:

Reversible MC wrt some π: $\pi_i P_{ij} = \pi_j P_{ji}$ — Detailed Balance

⇓

π-stationary

$0 \rightleftarrows 0$

$P(\text{being there}) P(\text{coming here from there}) = (\dots)$

$\pi_i \sum_{j \setminus i} P_{ij} = \sum_{j \setminus i} \pi_j P_{ji}$ — Global Balance

π reversible ⇕ time reversible π stationary

Let $Y_i = X_{n-i} \Longleftrightarrow \pi_i P_{ij} = \pi_j P_{ji}$

$Q_{ij} = P_{ij}$

Compute h_i

$h(i) = P(\text{win} \mid X_0 = i)$

$= P(\text{win} \mid X_1 = i+1) \underbrace{P(X_1 = i+1 \mid X_0 = i)}_{= p} + P(\text{win} \mid X_1 = i-1) \underbrace{P(X_1 = i-1 \mid X_0 = i)}_{= 1-p}$

$= p h(i+1) + (1-p) h(i-1)$ → you include time homogeneity: $h_t(i) = h_{t+1}(i)$

⊕ $\begin{cases} h(0) = 0 \\ h(N) = 1 \end{cases}$ → boundary conditions

$\begin{cases} h(0) = 0, \quad h(N) = 1 \\ h(i) = p h(i+1) + (1-p) h(i-1) \end{cases}$ (apply def to RHS)

$P_N = \lim_{t \to \infty} P^t_{iN} = \frac{1 - \vartheta^i}{1 - \vartheta^N}$ | $\vartheta = \frac{1-p}{p}$ | $p = P_{i, i+1}$

i = starting point, $i \in \{0, N\}$

Global $\to \pi_i = \sum_{j \in \mathcal{X}} \pi_j P_{ji}$

(or) $\pi_i \sum_{j \setminus i} P_{ij} = \sum_{j \setminus i} \pi_j P_{ji}$

Detailed $\to \pi_j P_{ji} = \pi_i P_{ij}$

General computation: ($|\mathcal{X}|$ support is finite) → in general, not unique ⇓ add constraints

Hit time $\to \tau_C = \min\{t \geq 0 : X_t \in C\}$

Hit P $\to h(i) = P(\tau_C < \infty \mid X_0 = i)$

$\begin{cases} h(i) = \sum_{j \in \mathcal{X}} P_{ij} h(j) & i \in \mathcal{X} \setminus C \\ h(i) = 1 & i \in C \end{cases}$ → Chapman-Kolm. equation

• Random Walk on a Directed Graph

$$P^{(RW)}_{ij} = \begin{cases} \frac{1}{d_i} & \text{if } \exists e_{ij} \\ 0 & \text{else} \end{cases} \quad \Big| \; d_i = \text{out degree}$$

→ Page Rank: rank websites with $P(\text{stay at website})$ in the RW, using π stationary.

π must exist and be unique $\Rightarrow P^{(PR)}_{ij} = \alpha P^{RW}_{ij} + (1-\alpha)\frac{1}{n}$ (ensure ergodicity)

π_i gives importance to connected / neighbours to prominent nodes

Sparsity is used to optimise $\pi = \lim_{t\to\infty} \alpha \cdot P^t$

• Sequential Testing:

Two drugs, $P_1 \neq P_2$ cure rate $\begin{cases} \text{1 and 2 have the same} \\ \text{\# patients} \end{cases}$, the test is a race between drugs

Decide if $P_1 > P_2$

$$D_t = \sum_i^N \mathbb{I}(i^{th} \text{ patient of 1 gets cured}) - \sum_i \mathbb{I}(i^{th} \text{ patient of 1 gets cured}) \to \Delta D = \{0, \pm 1\}$$

$\overset{iid}{\sim} Bern(P_1)$; $\overset{iid}{\sim} Bern(P_2)$

$$P(\text{test says } P_1 > P_2) = P(D_t \text{ hits } A \text{ before } -B \mid D_0 = 0)$$

$$h(i) = P_{i,i+1}h(i+1) + P_{i,i-1}h(i-1) + P_{ii}h(i) \quad i \in \{-B \dots A-1\}$$

$$= p \cdot h(i+1) + (1-p)h(i-1) \longrightarrow \oplus \; h(-B) = 0, \; h(A) = 1$$

Solve equation, larger A,B means more powerful test

$X_0 \to X_1 \to X_2 \to X_3 \to \dots$

$\downarrow \quad \downarrow \quad \downarrow \quad \downarrow$

$S_0 \quad S_1 \quad S_2 \quad S_3$

Hidden Markov Model → X_t evolves, but only $y_t \sim \mathbb{P}_{X_t}$ is observed

⇓ Problem to infer X_t from y_t

Note: $S_t \perp (X_i, S_i)_{i \neq t} \mid X_t$ } → arrows give the direction of information

(conditional on X_t, S_t is indep. of rest) → knowing X_t, the value of S_t gives no extra information

---

Exponential Distribution

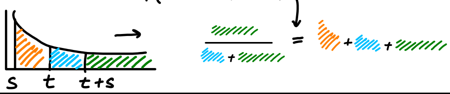
$$f(x) = \lambda e^{-\lambda x} \mathbb{I}(x>0) \longrightarrow \begin{cases} \mathbb{E}(X) = \frac{1}{\lambda} \\ Var(X) = \frac{1}{\lambda^2} \end{cases}$$

$$F(x) = (1 - e^{-\lambda x})$$

• X = time before event happens

If event wasn't happened yet, it has the same P as it happening at the start

Memoryless → $P(X > s+t \mid X > t) = P(X > s)$



$X \sim Poiss(\lambda)$

$\lambda = \mu = \sigma^2$

$X_i \sim Exp(\lambda_i) \to r_i(x) = \lambda_i$

then $\min(\{X_i\}) \sim exp(\sum \lambda_i)$

$X_i \sim Poiss(\lambda_i)$

then $\sum X_i \sim Poiss(\sum \lambda_i)$

(Hazard) Rates (for continuous r.v.)

$$\boxed{r(t) = \frac{f(t)}{1-F(t)}} \to \mathbb{P}(\text{happens at } t \mid \text{hasn't yet happened})$$

Fully characterises a distribution

$\bar{F}(t)$ — [Survival function → $\mathbb{P}$(has not happened by t)]

Note → $P(X \in [t, t+dt] \mid X > t) = r(t)dt + o(dt)$

$$\bar{F}(t) = e^{-\int_0^t r(s)ds} \quad \text{changing rate}$$

$r(t)$ is rate of change of $P(X \mid X > t)$

$\mathcal{X} = (0, \infty)$ is memoryless

iff $X \sim exp(\lambda)$. Then, $r(t) \equiv \lambda$

$(X_i)_{i=1..n}$ iid, $\mathcal{X} = (0,\infty)$ with r_i

Then $\begin{cases} r_{min}(t) = \sum_{i=1}^n r_i(t) \to \text{sum the rates to get the rate of min} \\ P(I = i \mid X = t) = \frac{r_i(t)}{\sum_{j=1}^n r_j(t)} \end{cases}$

P of the index

Jump Rates → necessary to construct a single matrix that determines the whole trajectory

Jump rate is only defined for $i \neq j$

For $i \neq j$

$$Q_{ij} = \lim_{h\to 0} \frac{P_{ij}(h)}{h} \longrightarrow P_{ij}(h) = P_{ij}(X(t+h) = j \mid X(t) = i)$$

$$= Q_{ij} h + o(h)$$

For each i, it's a hazard race to each j — hazard rate

You would need to know $P(t)$ in an interval:

$P(t) = P(\varepsilon)^n P(t - n\varepsilon), \; n = \lfloor \frac{t}{\varepsilon} \rfloor$

} derivative corresponds to no loss of info because it comes from an open interval

---

Continuous-time Markov chain $(|\mathcal{X}| < \infty)$ — memoryless has to ∞ with last bit of information

$$P(X(t+s) = j \mid X(s) = i, X(u) = x(u)) = P(X(t+s) = j \mid X(s) = i)$$

↳ Time-homogeneity:

$P_{ij}(t) = P(X(t+s) = j \mid X(s) = i) \to$ does not depend on s

(aka. const tran. prob)

Note $P_{ij}(s+t) = \sum_{k\in\mathcal{X}} P_{ik}(s)P_{kj}(t) \longrightarrow P(t+s) = P(t)P(s)$

Poisson Process → $\begin{cases} Q_{i,i+1} = \lambda & \forall i \in \mathcal{X} \\ Q_{i,j} = 0 & \text{else} \end{cases}$

If $\# = i$ events have rate λ, $i\lambda$ is the rate of first event:

$$Q_{ij} = \begin{cases} i\lambda & j = i+1 \\ 0 & \text{else} \end{cases}$$

(since it happens in continuous time, there can be no overlaps)

Generator $(Q \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{X}|})$

$\cdot \; Q_{ii} = -\sum_{j\neq i} Q_{ij}$ } completes the Q by setting the diagonal

Note: $\sum_{j\in\mathcal{X}} Q_{ij} = 0$ (sum over row)

$Q = P'(0)$

Kolm. Back/Forw. Equat.: $\begin{cases} \text{KFE} \to P'(t) = Q\,P(t) \\ \text{KBE} \to P'(t) = P(t)\,Q \end{cases}$ → the solution is the same: $P(t) = e^{tQ}$, exists but may not be in closed form

Construction → trajectories are piecewise constant

CAD-LAG assumption → $\begin{cases} \lim_{h\to 0^+} X(t+h) = x(t) \\ \lim_{h\to 0^-} X(t+h) \text{ exists} \end{cases}$ every point is an accumulation point

Under CADLAG, $X(t)$ is characterised by (T_n, y_n) pairs:

$$\begin{cases} T_0 \mid y_0 = i \sim exp(\nu_i), \; \nu_i > 0 \\ K_{ij} = P(y_1 = j \mid y_0 = i) \end{cases}$$

• exponential duration
• transition probability

Birth and Death Process

$$Q_{ij} = \begin{cases} \lambda_i & j = i+1 \\ \mu_i & j = i-1 \\ 0 \end{cases} \to \lambda, \mu \text{ are rates, and only one event can happen at a time (CADLAG)}$$

Simulation:

A. Sample $T_\mu \sim exp(\mu)$, $T_\nu \sim exp(\nu)$

$T = \min(T_\mu, T_\nu)$ is time at i

Move to $j = i+1$ if μ, else ν

B. Sample $T \sim exp(\mu_i + \lambda_i)$

Move to $j = i+1$ with $p = \frac{\lambda_i}{\lambda_i + \mu_i}$,

otherwise go to $j = i-1$

Note: if you are taking the first event (min) among n, μ_i will become $n\mu$

Counting Process → for $s < t$,

$(N(t))_{t\in[0,\infty)}$, $\mathcal{X} = \mathbb{Z}$ $\quad N(t) - N(s) = \#\text{events in } (s,t]$

T_i inter-arrival times are related

$N(t) > n \iff \sum_{i}^{n} T_i \le t$

• Time homogeneity ⇝ stationary increments: $N(t+s) - N(t)$ has no dep. in distribution wrt t

• Independent increments ⇝ $N(t_{n+1}) - N(t_n)$ are indep.

Non-homogeneous

$\lambda(t)$ counting function

$(P(E \in (t, t+h]) \approx \lambda(t)h$

PP(λ) is same as to say:

independent increments, and

• $N(t+s) - N(t) \sim Poiss\left(\int_s^{s+t} \lambda(y)dy\right)$

... using $\lambda(t)$

Equivalent to $N \sim PP(\lambda)$:

• $N(t+s) - N(s) \sim Poiss(\lambda t)$

$$\begin{cases} P(N(t+h) - N(h) = 1) = \lambda h + o(h) \\ P(N(t+h) - N(t) \ge 2) = o(h) \end{cases} \quad t \ge 0, \; \lambda \ge 0$$

• $T_i \overset{iid}{\sim} exp(\lambda)$

$$\begin{cases} \nu_i = \sum_{j\neq i} Q_{ij} \\ K_{ij} = \frac{Q_{ij}}{\nu_i} \end{cases} \to$$

1. Simulate independently $T_{i\to j} \sim exp(Q_{ij}) \; \forall j \in \mathcal{X} \setminus \{i\}$
2. $T_n := \min_{j\in\mathcal{X}\setminus\{i\}} T_{i\to j}$, go to $y_{n+1} = \arg\min_{j\in\mathcal{X}\setminus\{i\}} T_{i\to j}$

$\square \xrightarrow{i} \square \to \square \to \square \to \dots$

$T \downarrow \quad \downarrow \quad \downarrow \quad \downarrow$

$\square \quad \square \quad \square \quad \square$ HMM model equivalence

Superposition

N_1, N_2 be independent PP with λ_1, λ_2

Then $\begin{cases} N = (N(A))_{A\in\mathcal{O}} \\ N(A) = N_1(A) + N_2(A) \end{cases}$ satisfies $N \sim PP(\lambda_1 + \lambda_2)$

Limiting Distribution (Birth/Death Process)

Global Balance $\pi_i Q_{ij} = \pi_j Q_{ji}$

⇑

Local Balance $\pi Q = 0$

$p = 1 - \frac{\lambda}{\mu}$

$$\begin{cases} \pi_j = (1-p)^j p \\ X(t) \rightsquigarrow Geom(p) \end{cases}$$

$$\mathbb{E}(X(t)) \approx \frac{1-p}{p} = \frac{\lambda}{\mu - \lambda}$$

BD Server $(s = \#\text{servers})$

$$\begin{cases} \mu_n = \min(s,n)\cdot\mu & n \ge 1 \\ \lambda_n = \lambda \end{cases}$$

General Case

$D \longrightarrow \bar{D} = \{\text{finite subsets of } D\}$

$N: \Omega \to \bar{D}$ measurable,

For example, $N(A) = |A|$

+ Axiom for $A \cap B = \emptyset$

$N(A \cup B) = N(A) + N(B)$

Point Process

$N = (N(A))_{A\in\mathcal{O}}$ with

$\lambda: \bar{D} \to \mathbb{R}$ intensity measure

• $N(A) \perp N(B) \quad \forall A,B$ disjoint (indep. incren)

• $N(A) \sim Poiss(\lambda(A)) \; \forall A$ (Poiss. incren)

Notable Consequence

$N(A \cup B) = (N(A) + N(B) \sim Poiss(\lambda(A) + \lambda(B)) = Poiss(\lambda(A \cup B))$

Thinning

$N \sim PP(\lambda)$ and set each point to $\{1, 2\}$ using a Bernoulli(p). Then the sub-processes N_1, N_2 are $PP(\lambda_1 = p\lambda)$ and $PP(\lambda_2 = (1-p)\lambda)$ and they are independent

FINITE-DIM DISTRIBUTION → MARGINALS HAVE A KNOWN IP

$X = (X_t)_{t\in T}$ REGULAR ENOUGH PROCESSES MAY BE DESCRIBED COMPLETELY

THE DISTRIBUTIONS OF $X_{1:k} = (X_{t_1} \dots X_{t_k})$, GIVEN $(t_1, \dots t_k) \in T^k$

BROWNIAN MOTION

- $(Y_n)_{n\in\mathbb{N}}$ SYMMETRIC RW ON $\mathbb{Z}$ $\begin{cases} +1 & 1/2 \\ -1 & 1/2 \end{cases}$
- $\delta_x, \delta_t > 0$, $X_t = \delta_x Y_{\lceil t/\delta t\rceil}$, $t \geq 0$
- $Y_n = \sum z_i$, z_i iid

IF $\delta x = O(\sqrt{\delta t})$, $\lim_{\delta t\downarrow 0} \mathbb{V}(X_t) = \lim_{\delta t\downarrow 0} \sigma^2 \delta t \lceil t/\delta t\rceil = \sigma^2 t$ — CONDITION IS REQUIRED FOR NON-TRIV. LIMIT

z_i iid, SO BY CLT $\boxed{X_t} \xrightarrow{d} \boxed{\forall t>0 \; N(0,\sigma^2 t) \text{ AS } \delta x \downarrow 0}$

- $\mathbb{E}[Y_n] = n\mathbb{E}[z_1] = 0 \rightarrow \mathbb{E}[X_t] = \delta x\, \mathbb{E}[Y_{\lceil t/\delta t\rceil}] = 0$
- $\mathbb{V}(Y_n) = n\cdot\mathbb{V}(z_1) = n$; $\mathbb{V}(X_t) = (\delta x)^2 \mathbb{V}(Y_{\lceil t/\delta t\rceil}) = (\delta x)^2 \lceil t/\delta t\rceil$

YOU DON'T NEED MORE THAN z_i iid $\mu=0, \sigma^2$ $\longrightarrow X_t = \sigma\sqrt{\delta t}\sum_{j=1}^{\lceil t/\delta t\rceil} z_j$ HAS THE SAME DISTRIBUTION

BROWNIAN MOTION

$X = (X_t)_{t\in[0,\infty)}$ ON $\mathcal{X} = \mathbb{R}$

$X \sim BM(\sigma^2)$ IF ALL HOLD:

- $\mathbb{P}(X_0 = 0) = 1$
- $\mathbb{P}(t \mapsto X_t \in C^0 \;\forall t\in[0,\infty)) = 1$
- STATIONARY ⊕ INDEPENDENT INCREMENTS $\Rightarrow X_t - X_s \overset{d}{=} X_{t-s} - X_0$, $X_t - X_s \sim N(0, (t-s)\sigma^2)$
- $X_t \sim N(0, \sigma^2 t) \;\forall t$

$\sigma = 1 \rightarrow$ STANDARD BM

BM COVERS: $O(\sqrt{t})$ SPACE IN $O(t)$ TIME

INDEP INCREMENTS ⇓ MARKOV PROPERTY

SQUARE ROOT SCALING

α-QUANTILES OF X_t ARE $q_\alpha\sqrt{t}$ | q_α FROM $N(0,\sigma^2)$

INVARIANCES

$a, b > 0 \rightarrow Y_t = a X_{t/b} \sim BM\left(\frac{a\sigma}{\sqrt{b}}\right)$

$Y_t = X_{1-t} - X_1 \sim$ BM OF X_t

$Y_0 = 0$, $Y_t = t X_{1/t} \sim BM(\sigma)$

TRAJECTORIES

- a.s. CONTINUOUS
- a.s. NOWHERE DIFFERENTIABLE
- RECURRENCE: a.s. CROSS 0 INFINITELY MANY TIMES ↓ ALSO FOR A NEIGHBOURHOOD OF 0

⟶ THERE'S TWO, EQUALLY INTERESTING WAYS TO GENERALISE BM: GAUSSIAN PROCESSES AND DIFFUSION PROCESSES

GAUSSIAN PROCESS

$(X_t)_{t\in T}$, $\mathcal{X} = \mathbb{R}$, $T \subseteq \mathbb{R}^d$ GENERAL

THE FINITE DIM. DISTRIBUTIONS ARE MULTIVARIATE GAUSSIANS:

$\forall k\in\mathbb{N}$, $X_{t_1\dots t_k} \sim N(\mu, \Sigma)$, $\mu\in\mathbb{R}^k$, $\Sigma\in\mathbb{R}^{k\times k}$

THEN $X\sim BM(\sigma)$ IS GAUSSIAN PROC. ⟵ BM(σ) IS UNIQUE CONTINUOUS GP S.T. $\begin{cases} m(t) = 0 \\ C(t,s) = \sigma^2 \min(t,s) \end{cases}$

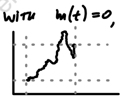
$m(t) = 0$

$C(t,s) = \mathbb{E}[X_t X_s] = \mathbb{E}[X_t^2] + \mathbb{E}[X_t(X_s - X_t)] = \sigma^2 t$ — POSITIVE DEFINITE FUNCTION

BROWNIAN BRIDGE

$X \sim BM(\sigma)$, $Y_t := (X_t | X_1 = 0)$

THEN $Y \sim GP(m, C)$ WITH $m(t) = 0$, $C(t,t') = \sigma^2 t(1-t')$



BB WITH σ^2, END AT $Y_s = y$ IS A GP WITH $T = [0,s]$,

$\begin{cases} m(t) = \frac{t}{s} y & t\in[0,s] \\ C(t,t') = \sigma^2 t(s-t') & 0\leq t\leq t'\leq s \end{cases}$

ROLE OF $m(t)$:

$X \sim GP(0,C) \Longrightarrow Y_t = X_t + m(t) \sim GP(m, C)$

JUST FOCUS ON $X \sim GP(0,C)$ WLOG

ROLE OF C:

$C: T\times T \rightarrow \mathbb{R}$

IE $\text{cov}(X_t, X_s)$

SYMMETRIC, POS. DEF. FUNCTION $\Rightarrow$ EXISTS VALID $GP(0,C)$ (USUALLY HAS THE FORM $K(x,y) = f(x-y)$)

THE FUNCTION IS CALLED KERNEL

- $a\cdot\exp(-b|t-s|) \rightarrow$ CONTINUOUS TRAJECTORIES $\Big] \rightarrow \begin{cases} Y_t = \sqrt{a} X_t \\ Y_t = X_{bt} \end{cases}$
- $a\cdot\exp(-b|t-s|^2) \rightarrow$ DIFFERENTIABLE TRAJECTORIES

THE REGULARITY OF THE TRAJECTORY DEPENDS ON REGULARITY AT $x=0$

K CAN BE

$P^t(x,y)$ = DENSITY OF IP THAT THE PROCESS WILL START AT X AND BE AT Y WITHIN t TIME

YOU CAN STUDY THE EVOLUTION OF P_t DIRECTLY

INITIAL CONDITIONS $P^0(x,y) = \delta(x-y)$

THIS IS THE HEAT EQUATION

DIFFUSION PROCESS

STOCHASTIC DE

BM DEFN. IMPLIES $\frac{\partial}{\partial t}P^t(x,y) = \frac{\sigma^2}{2}\frac{\partial^2 P^t}{\partial y^2}(x,y)$

BM IS THE ONLY MC:

- $\mathbb{E}[X_{t+h} - X_t | X_t = x] = o(h)$] → NO UP/DOWN DRIFT
- $\mathbb{E}[(X_{t+h} - X_t)^2 | X_t = x] = \sigma^2 h + o(h)$] → CONSTANT VOLATILITY
- $\mathbb{P}(|X_{t+h} - X_t| > \varepsilon | X_t = x) = o(h)$] → NO JUMPS

⟺

X_t CONTINUOUS TIME MP, CONTINUOUS TRAJECTORIES, $\mu(x,t)$ DRIFT FUNCTION, $\sigma^2(x,t)$ INSTANT. σ^2

IF

- $\mathbb{E}[X_{t+h} - X_t | X_t = x] = \mu(x,t)h + o(h)$
- $\mathbb{E}[(X_{t+h} - X_t)^2 | X_t = x] = \sigma^2(x,t)h + o(h)$
- $\mathbb{P}(|X_{t+h} - X_t| > \varepsilon | X_t = x) = o(h)$] → IMPOSES CONTINUITY $\mu_{2\varepsilon}$

μ EQUALS THE DERIVATIVE (IN $\mathbb{E}$), σ^2 IS THE INFINITESIMAL VARIANCE

DETERMINISTIC TERM

(ITO FORM) $dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$

$X_t = X_0 + \int_0^t \mu(X_s, s)ds + \int_0^t \sigma(X_s, s)dB_s$

DERIVATIVE INTERPRETATION: $\lim_{h\downarrow 0} \mathbb{E}\left[\frac{X_{t+h} - X_t}{h} \Big| X_t = x\right] = \mu(x,t)$

AKA. $dX_t = f(X_t, t)dt$, $\lim_{h\downarrow 0}\frac{X_{t+h} - X_t}{h} = f(X_t)$

$\lim_{h\downarrow 0} \mathbb{E}\left[\frac{(X_{t+h} - X_t)^2}{h} \Big| X_t = x\right] = \sigma^2(x,t)$

IN THIS CASE, THE SOLUTION IS SUM OF SOLUTIONS:

$\begin{cases} dX_t = \mu dt \rightarrow X_t = \mu t \\ dX_t = \sigma dB_t \rightarrow X_t = \sigma B_t \end{cases}$

$B_t \sim BM(\mu, \sigma)$

$X_t = \mu t + \sigma B_t$

X_t IS A DIFFUSION PROCESS $\begin{cases} \mu(x,t) = \mu \\ \sigma^2(x,t) = \sigma^2 \end{cases}$ (ZERO DRIFT, CONST. VOLATILITY)

DETERMINISTIC $t \mapsto X_t \Rightarrow \begin{cases} X_{t+h} - X_t = \mu h + o(h) \\ (X_{t+h} - X_t)^2 = \sigma^2 h + o(h) \end{cases}$

BOTH CANNOT HOLD TOGETHER (IN SDE, THEY HOLD IN $\mathbb{E}$)

IN BM WITH CONST. DRIFT, YOU CAN SOLVE FOR THE TWO COMPONENTS OF THE SOLUTION, THEN ADD THEM. IN GENERAL YOU CAN'T DUE TO INTERACTION EFFECTS

TRANSITION KERNEL

$X = (X_t)_{[0,\infty)}$ → TIME-HOM, CONT. TIME, MP., $\mathcal{X} = \mathbb{R}$

TREATED AS AN OPERATOR: $(P^t f)(x) = \int f(y)P^t(x,y)dy = \mathbb{E}[f(X_t)|X_0 = x]$

THE PMF IS $P^t(x,A) = \mathbb{P}(X_t\in A|X_0 = x) = \int_A P^t(x,y)dy$

$P^{t+s}_{(x,y)} = \int P^t_{(x,z)} P^s_{(z,y)} dz$ WITH $P^t P^s_f = P^{t+s}_f$ — CHAPMAN KOLMOGOROV EQUATIONS

$X \sim BM(\sigma) \longrightarrow$ TRANSITION KERNEL $P^t(x,y) = \frac{\exp\left(-\frac{(x-y)^2}{2t\sigma^2}\right)}{\sqrt{2\pi t\sigma^2}}$

INTEGRAL REPRESENTATION

dB_t NOTATION MEANS YOU CAN GET X BY INTEGRATING $\mu(x,t)dt$, $\sigma(x,t)dB_t$

$X = \lim_{h\downarrow 0}$ OF $X_{t+h} = X_t + \mu(X_t, t)h + \sigma(X_t, t)\cdot(B_{t+h} - B_t)$

$X_t = X_0 + \int_0^t \mu(X_s, s)ds + \int_0^t \sigma(X_s, s)dB_s$

FOKKER-PLANCK (KOLMOGOROV EQUATIONS FOR DIFFUSION) → SEE CONNECTION TO SCHRÖDINGER EQUATION

IF $X_t = \mu(X_t)dt + \sigma(X_t)dB_t$

THE GENERATOR SHOULD EXTEND → $Q = \frac{d}{dt}P^t\Big|_{t=0}$ $\begin{cases} f \mapsto Qf, \\ \text{BOTH } \mathbb{R}\rightarrow\mathbb{R} \end{cases}$

$\boxed{Q = \mu\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2\frac{\partial^2}{\partial x^2}}$

ACTS ON A FUNCTION AND PLAYS THE ROLE OF Q-MATRIX OF CTMC

IN BM, $Q = \frac{\sigma^2}{2}\frac{d^2}{dx^2}$ (AKA. LAPLACIAN) SEE HEAT EQ.

YOU CAN ESTIMATE X_t DISTRIBUTION FROM X_0 DISTR. USING FOKK-PLANCK:

1. $h(y,t)$ = PDF OF X_t
2. BY CK EQUATION, $h(y,t) = \int P^t_{(x,y)} h(x,0)dx$
3. $\frac{d}{dt}h(y,t) = -\frac{\partial}{\partial y}(\mu(y)h(y,t)) + \frac{1}{2}\frac{\partial^2}{\partial y^2}(\sigma^2(y)h(y,t))$ (NOTE y IS VALUE YOU GO TO: $x\rightarrow y$)

Q ACTS ON $f(x)$, GIVING THE RATE OF CHANGE OF $\mathbb{E}$

$Qf = \mu(x)f'(x) + \frac{\sigma^2(x)}{2}f''(x)$

Q^* AKA. Q ADJOINT IS THE RHS OF THE FOKKER-PLANCK

FOKKER-PLANCK EQUATION

A SUFF. REGULAR DIFFUSION PROCESS HAS:

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

IN THE GENERATOR,

$\mu\frac{\partial}{\partial x}$ → DETERMINISTIC PART (DRIFT) $\sim O(h)$

$\frac{\sigma^2}{2}\frac{\partial^2}{\partial x^2}$ → STOCASTIC PART (VOLATILITY) $\sim O(\sqrt{h})$

JUST WHAT YOU GET BY PLUGGING IN THE TERMS

STATIONARY DISTRIBUTIONS

$h(y,0) = \pi(y) \Rightarrow \forall t, y \;\; \frac{\partial}{\partial t}h(y,t) = 0$

BY FK, $(\mu\pi)' = \frac{(\sigma^2\pi)''}{2}$ $\Big\}$ $\mu\pi = \frac{1}{2}(\sigma^2\pi)' + \text{CONST}$ HAS ∞ SOLUTIONS

ALSO IMPOSING $\sigma^2(x) = \sigma^2$, CONST = 0 — THEN ∃ BIJECTION $\pi \leftrightarrow \mu$

$\mu(x) = \frac{1}{2}\sigma^2\frac{\pi'(x)}{\pi(x)} = \frac{\sigma^2}{2}(\log\pi)'(x)$

AKA. $dX_t = \frac{\sigma^2}{2}\nabla\log\pi(X_t)dt + \sigma dB_t$ IS π-INVARIANT DIFFUSION PROCESS (ALSO REVERSIBLE)

→ NOISE AND MOVEMENT SCALE EQUALLY.

INTRODUCES NOISE TO DETERMINISTIC $m\ddot{x} = \dots + \varepsilon$, $\varepsilon \sim N(\dots)$

P^t MARCHES IN TIME:

- $\mathbb{E}[f(X_t)|X_0 = x] = \int P^t_{(x,y)} f(y)dt$
- $h(y,t) = \int P^t_{(x,y)} h(x,0)dt$

P^t IS THE SOLUTION TO THE FOKKER-PLANCK PDE:

$\begin{cases} \frac{\partial P^t_{(x,y)}}{\partial y} = -\frac{\partial}{\partial y}\left[\mu(y,t)P^t_{(x,y)}\right] + \frac{1}{2}\frac{\partial^2}{\partial y^2}\left[\sigma^2(y,t)P^t_{(x,y)}\right] \\ P^0_{(x,y)} = \delta(x-y) \end{cases}$

MONTE CARLO METHODS

ALGORITHMS THAT PERFORM REPEATED RANDOM SAMPLING TO COMPUTE QUANTITIES OF INTEREST

LINEAR CONGRUENTIAL GENERATOR

$x_{i+1} = (a x_i + c) \% m$

m MODULUS

$0 \le a < m$ MULTIPLIER

$0 \le c < m$ INCREMENT

x_0 SEED

HAS A PERIOD OF m, SO $m >> \#$ SAMPLES NEEDED

POISSON PROCESS

- HOMOGENOUS: EVENTS ARE PLACED IN TIME BY SAMPLING ΔT BETWEEN THEM.

$$P(T) = \lambda \exp(-\lambda T)$$

- INHOMOGENOUS: $\lambda(t)$ RATE FUNCTION

$$P(N(t+dt) - N(t)) = \lambda(t)dt$$

$$N(t+s) - N(t) \sim \mathrm{Poiss}\left(\int_t^{t+s} \lambda(u)du\right)$$

$$\boxed{x \sim F^{-1}(y) \mid y \sim U([0,1]) \quad F(y) = 1 - \exp\left(-\int_t^{t+x} \lambda(u)du\right) \quad \text{(TO SAMPLE WAITING TIME)}}$$

IF THE INTEGRAL IS NOT EASY TO INVERT, THINNING:
- $\mu(t)$ EASY TO INVERT, s.t. $\forall t \; \mu(t) \ge \lambda(t)$
- ACCEPT EVENT WITH $\mathbb{P} = \lambda(t)/\mu(t)$

PSEUDO-RANDOM NUMBER GENERATORS

PRODUCE A DETERMINISTIC SEQUENCE $(a_n) \in [0,1]$ THAT IMITATES iid $\sim U([0,1])$

START FROM SEED x_0, COMPUTE THE NEXT RV

TEST BY PERFORMING HYPOTHESIS OF $x_i \overset{iid}{\sim} U(0,1)$

INTEGRALS IN $\dim >> 1$

- RIEMANN SUMS: $[x_0, x_N]$ HAS (EQUIDISTANT POINT GRID) $x_i = x_0 + \left(\frac{x_N - x_0}{N}\right) i$
- MONTE CARLO

$$I(A) \approx \frac{1}{N} \sum^N f(x_i) \mid x_i \overset{iid}{\sim} \mathrm{UNIF}(A)$$

BY CLT, ERROR SCALES AS $\frac{1}{\sqrt{N}}$ AND $N \to \infty$ PRODUCES THE TRUE VALUE

SAMPLING FROM A KNOWN CDF

$$F(x) = \int_{-\infty}^{x} f(z)dz$$

$y \sim U([0,1])$

$x = F^{-1}(y) \sim F_X$

BOX-MULLER GENERATES PAIRS OF GAUSSIAN RV (FASTER)

MONTE CARLO EXPECTATION

$$\langle f(x) \rangle_{\mathbb{P}} = \frac{1}{N_s} \sum_{i=1}^{N_s} f(x_i) \mid x_i \overset{iid}{\sim} \mathbb{P}$$

IN STATISTICAL PHYSICS,

$$P_B(x) = \frac{1}{Z} \exp\left(-\frac{E(x)}{kT}\right)$$ BOLTZMANN FUNCTION

Z → PARTITION FUNCTION

$$\underset{\text{FREE ENERGY}}{F(E)} = \underset{\text{ENERGY}}{E} - T \underbrace{S(E)}_{\text{ENTROPY}}$$

METROPOLIS HASTINGS

BUILD x_t MP s.t. P_B is STATIONARY

1. x_0 INITIALISATION
2. DEFINE $x \to x'$ TRANSITIONS WITH SOME $\mathbb{P} = q(x \to x')$
3. DRAW A PROPOSED MOVE
4. ACCEPT THE PROPOSAL IF $P_B(x') q(x' \to x) \ge P_B(x) q(x \to x')$
5. OTHERWISE, ACCEPT $\sim p = \frac{P_B(x') q(x' \to x)}{P_B(x) q(x \to x')}$

CONCISELY,

$$\boxed{P_{Acc} = \min\left\{1, \frac{P_B(x') q(x' \to x)}{P_B(x) q(x \to x')}\right\}}$$

DEPENDS ONLY IN P_B ENERGY DIFFERENCE $\Rightarrow$ NO NEED TO CALCULATE Z

DETAILED BALANCE $\Rightarrow$ PREREQ. TO CONVERGENCE

IF $q(x \to x') = q(x' \to x)$,

$$W(x \to x') \propto \min\left(1, \frac{P_B(x')}{P_B(x)}\right)$$

MOVES:
- INCREASE $P_B \Rightarrow$ ACCEPT
- DECREASE $P_B \Rightarrow$ ACCEPT WITH p

IRREDUCIBLE ⊕ APERIODIC $\Rightarrow$ MC METHOD CONVERGES

$D_{KL}(t)$ DECREASES ALWAYS, MAYBE TOO SLOW

IN SYSTEMS WITH MULTIPLE MINIMA, MC CONVERGES TO NEAREST E_{min}

SIMULATING CONTINUOUS TIME MARKOV CHAINS

GILLESPIE:
- INITIALISE SYSTEM
- COMPUTE REACTION RATE $a_i(t)$ AND $a_{tot}(t) = \sum a_i(t)$
- DRAW TIME OF NEXT STATE $\tau \sim \mathrm{Exp}(a_{tot}(t))$
- DRAW STATE $i \sim \frac{a_i(t)}{a_{tot}(t)}$
- UPDATE $t \mapsto t + \tau$

ISING MODEL

$S \in \{\pm 1\}^N$

- D-DIM SQUARE GRID
- $E(S) = -J \sum_{(i,j)} S_i S_j$ — SUM OVER PAIRS OF NEIGHBOURS
- $\langle E \rangle_{P_B} \to$ ENERGY
- $\langle S \rangle_{P_B} \to$ MAGNETISATION
- $E(S)$ HAS $2 \times \min$ (ALL $+1$, ALL -1)
- AS $T \to \infty$, $\langle E \rangle = 0$ (P_B BECOMES UNIFORM)

POLARISATION

- 1D, 2D → EXACT SOLUTION
- D ≥ 3 → NO ANALYTICAL SOLUTION
- 2D, 3D PHASE TRANSITION FOR T_c, T_c'

SIMULATION

TO COMPUTE STATISTICS, YOU NEED A HIGH NUMBER OF SAMPLES

ISSUES WITH MH ALGORITHM

- VERY SLOW NEAR PHASE TRANSITIONS (2D ISING → CORR. TIME $\propto L^4$)
- GETS STUCK AT LOW E IN MODELS WITH RUGGED LANDSCAPE

CLUSTER ALGORITHMS

CORR. TIME DIVERGES AT HIGH TEMP, SO IN LARGE NEIGHBOURHOODS SPINS ARE IDENTICAL.

FLIP AN ENTIRE CLUSTER

WOLFF ALGORITHM

- PICK A SPIN TO FLIP
- FOR EACH SAME-SPIN NEIGHBOUR, ADD IT TO THE LIST TO FLIP BY (P_{add} PROBABILITY)

$q(x \to x') = Q(1 - P_{add})^m$

$q(x' \to x) = Q(1 - P_{add})^n$

DETAILED BALANCE: GUARANTEES CONVERGENCE

$P_{add} = p$ s.t. DETAILED BALANCE HOLDS

SIMULATED ANNEALING

- $T = T_{max}$, $X_0 = x_0$ WITH $E(x)$
- PROPOSE $x \to x'$, COMPUTE $E(x')$
- ACCEPT WITH $p = \min(1, \exp(-\frac{\Delta E}{T}))$
- AFTER n_T STEPS, $T \mapsto T/\alpha$

(TEMPERATURE COOLING SCHEDULE DEPENDS ON PROBLEM)

SIMULATED TEMPERING

- ADD AS VARIABLE $\beta = 1/kT \in \{\beta_1, \dots, \beta_{N_T}\}$
- ADD TRANSITIONS FOR β IN MH

SINCE YOU CARE ABOUT $\langle f(x) \rangle_\beta$, YOU RESTRICT SAMPLES TO β

EDWARDS-ANDERSON

- $S \in \{\pm 1\}^N$
- $J_{ij} \sim \mathcal{N}(0, J)$
- $E(S) = -\sum_{(i,j)} J_{ij} S_i S_j$

SIMPLIFIED SPIN GLASS → MAGNETIZATION IS CLOSE TO 0 AT ALL TEMPERATURES

$$q_{\alpha,\beta} = \frac{1}{N} \sum_i S_i^\alpha S_i^\beta$$ (OVERLAP BETWEEN STATES)

FRUSTRATED SYSTEM (NO WAY TO SATISFY ALL BONDS) $\Rightarrow$ SLOW CONV, RUGGED LANDSCAPE

PARALLEL TEMPERING

USE MULTIPLE SYSTEMS WITH DIFFERENT β.

β_{N_T} IS max, ABOVE T_c

AFTER N_x STEPS, THE REPLICAS CAN SWITCH WITH SOME $P(h \leftrightarrow h')$

LIKELY EXCHANGES HAVE CLOSE ENERGIES

N_T LARGE SO THE DISTRIBUTIONS OF ENERGIES OVERLAP