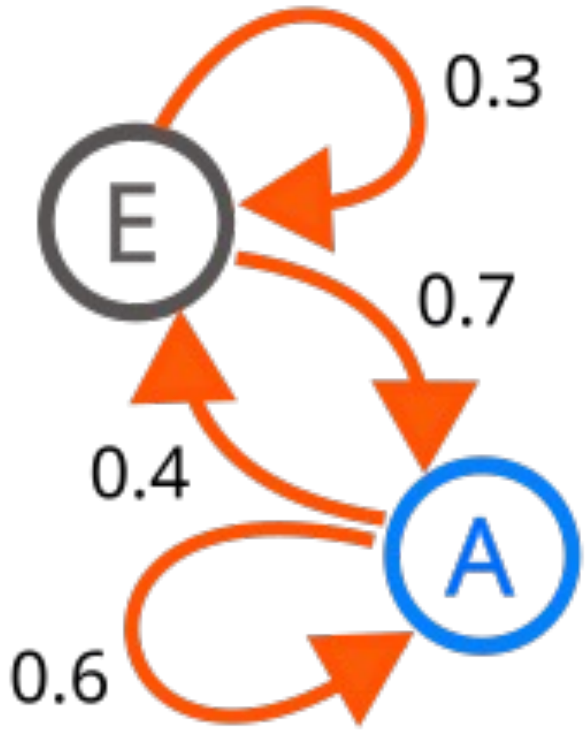
Markov chains: a refresher

Key materials: https://doi.org/10.1007/978-3-319-08488-6
https://www.statslab.cam.ac.uk/~rrw1/markov/M.pdf

What is a Markov Chain?

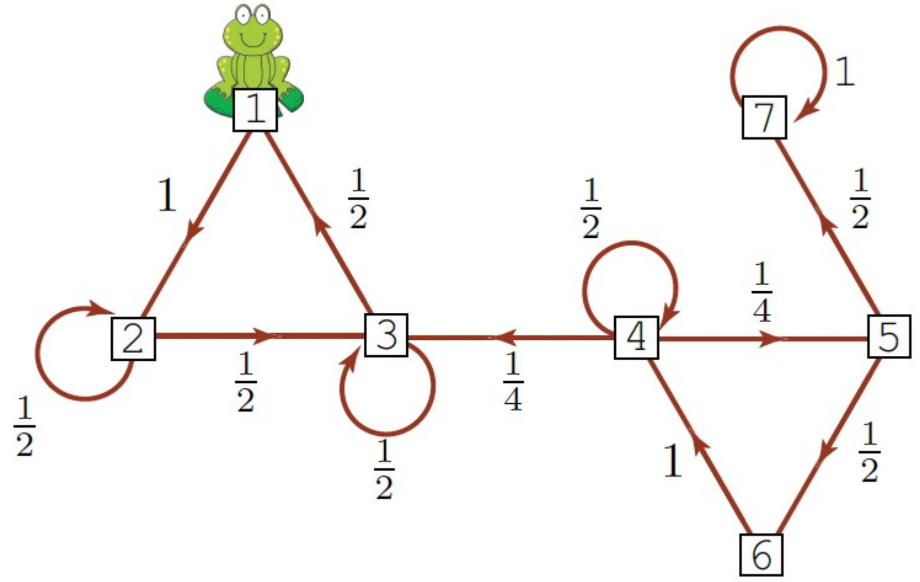
- stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.

- **conditional** on the present state of the system
- its future and past states are **independent**



Introduction

Example 1.1. A frog hops about on 7 lily pads. The numbers next to arrows show the probabilities with which, at the next jump, he jumps to a neighbouring lily pad (and when out-going probabilities sum to less than 1 he stays where he is with the remaining probability).



$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{1}{2} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Starting in state 4, what is the probability that we ever reach state 7?

1/3

From frogs to formal definitions:

I be a countable set, $\{i, j, k, \ldots\}$.

We also have a **transition matrix** $P = (p_{ij} : i, j \in I)$ with $p_{ij} \geq 0$ for all i, j.

It is a **stochastic matrix**, meaning that $p_{ij} \geq 0$ for all $i, j \in I$ and $\sum_{j \in I} p_{ij} = 1$

Study: sequence of random variables $X_0, X_1, \ldots$

Definition 1.2. We say that $(X_n)_{n \geq 0}$ is a **Markov chain** with initial distribution λ and transition matrix P if for all $n \geq 0$ and $i_0, \ldots, i_{n+1} \in I$,

(i) $P(X_0 = i_0) = \lambda_{i_0}$;

(ii) $P(X_{n+1} = i_{n+1} \mid X_0 = i_0, \ldots, X_n = i_n) = P(X_{n+1} = i_{n+1} \mid X_n = i_n) = p_{i_n i_{n+1}}$.

Example: two-state Markov chain

1.6 $P^{(n)}$ for a two-state Markov chain

Example 1.4 (A two-state Markov chain).

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

For some A and B:

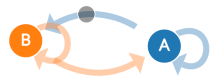
$$p_{11}^{(n)} = A + B(1-\alpha-\beta)^n$$

Or!

https://setosa.io/ev/markov-chains/

A simple, two-state Markov chain is shown below.

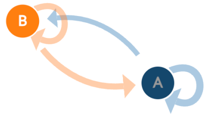
speed



With two states (A and B) in our state space, there are 4 possible transitions (not 2, because a state can transition back into itself). If we're at 'A' we could transition to 'B' or stay at 'A'. If we're at 'B' we could transition to 'A' or stay at 'B'. In this two state diagram, the probability of transitioning from any state to any other state is 0.5.

Of course, real modelers don't always draw out Markov chain diagrams. Instead they use a "transition matrix" to tally the transition probabilities. Every state in the state space is included once as a row and again as a column, and each cell in the matrix tells you the probability of transitioning from its row's state to its column's state. So, in the matrix, the cells do the same job that the arrows do in the diagram.

speed



	A	B
A	P(A|A): 0.67	P(B|A): 0.33
B	P(A|B): 0.50	P(B|B): 0.50

Calculation of n-step transition probabilities

- Use of symmetry
- Markov property

Theorem 2.3 (Markov property). *Let $(X_n)_{n\geq 0}$ be Markov(λ, P). Then conditional on $X_m = i$, $(X_{m+n})_{n\geq 0}$ is Markov(δ_i, P) and is independent of the random variables $X_0, X_1, \ldots, X_m$.*

- Class structure: break a Markov chain into smaller pieces (communicating classes: I leads to i-> j)

Theorem 2.4. *For distinct states i and j the following are equivalent.*

(i) $i \to j$;

(ii) $p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_{n-1} i_n} > 0$ *for some states* $i_0, i_1, \ldots, i_n$, *where* $n \geq 1$, $i_0 = i$ *and* $i_n = j$;

(iii) $p_{ij}^{(n)} > 0$ *for some* $n \geq 1$.

- Close cases:

It is clear from (ii) that $i \leftrightarrow j$ and $j \leftrightarrow k$ imply $i \leftrightarrow k$, and also that $i \leftrightarrow i$. So $\leftrightarrow$ satisfies the conditions for an equivalence relation on I and so partitions I into **communicating classes**. We say a C is a **closed class** if

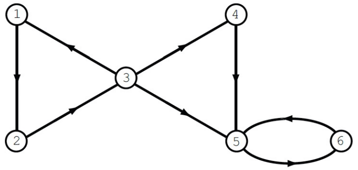
$$i \in C,\ i \to j \implies j \in C.$$

A closed class is one from which there is no escape. A state i is **absorbing** if $\{i\}$ is a closed class.

If C is not closed then it is **open** and there exist $i \in C$ and $j \notin C$ with $i \to j$ (you can escape).

Example 2.5. Find the classes in P and say whether they are open or closed.

$$P = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ \frac{1}{3} & 0 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$



Absorbtion probabillities and hitting times

Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P. The first **hitting time** of a subset A of I is the random variable $H^A : \Omega \to \{0, 1, \ldots\} \cup \{\infty\}$ given by

$$H^A(\omega) = \inf\{n \geq 0 : X_n(\omega) \in A\}$$

where we agree that the infimum of the empty set is ∞. The probability starting from i that $(X_n)_{n\geq 0}$ ever hits A is

$$h_i^A = P_i(H^A < \infty).$$

When A is a closed class, h_i^A is called an **absorption probability**. The **mean hitting time** for $(X_n)_{n\geq 0}$ reaching A is given by

$$k_i^A = E_i(H^A) = \sum_{n<\infty} nP_i(H^A = n) + \infty P(H^A = \infty).$$

Informally,

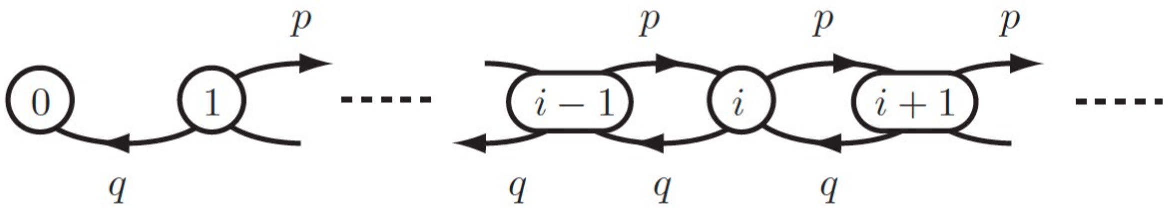
$$h_i^A = P_i(\text{hit } A), \quad k_i^A = E_i(\text{time to hit } A).$$

Theorem 3.4. *The vector of hitting probabilities $h^A = (h_i^A : i \in I)$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} h_i^A = 1 & \text{for } i \in A \\ h_i^A = \sum_j p_{ij} h_j^A & \text{for } i \notin A \end{cases} \tag{3.1}$$

(Minimality means that if $x = (x_i : i \in I)$ is another solution with $x_i \geq 0$ then $x_i \geq h_i$ for all i.)

Example: gamblers ruin



The transition probabilities are

$$p_{00} = 1,$$
$$p_{i,i-1} = q, \quad p_{i,i+1} = p \quad \text{for } i = 1, 2, \ldots$$

Set $h_i = P_i(\text{hit } 0)$, then it is the minimal non-negative solution to

$$h_0 = 1,$$
$$h_i = ph_{i+1} + qh_{i-1} \quad \text{for } i = 1, 2, \ldots$$

If $p \neq q$ this recurrence has general solution

$$h_i = 1 - A + A\,(q/p)^i\,.$$

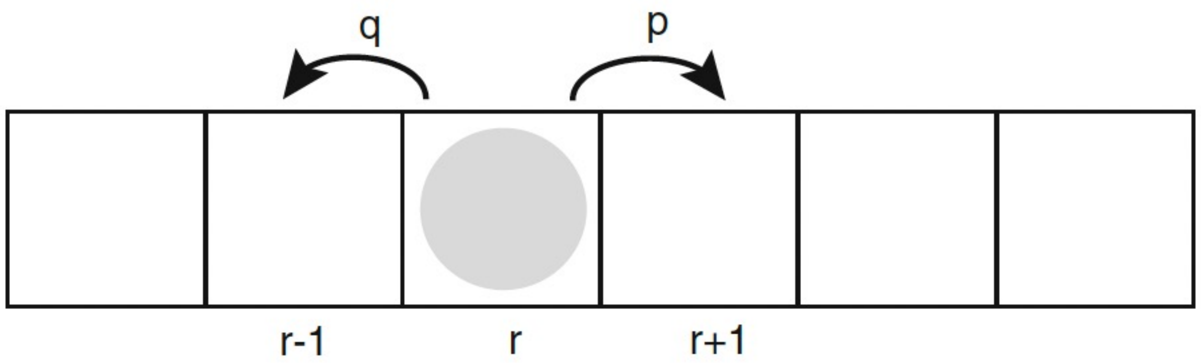
We know by Theorem 3.4 that the minimal solution is h, a distribution on $\{0, 1, \ldots\}$.

If $p < q$ then the fact that $0 \leq h_i \leq 1$ for all i forces $A = 0$. So $h_i = 1$ for all i.

If $p > q$ then in seeking a minimal solution we will wish to take A as large as possible, consistent with $h_i \geq 0$ for all i. So $A = 1$, and $h_i = (q/p)^i$.

Finally, if $p = q = 1/2$ the recurrence relation has a general solution $h_i = 1 + Bi$, and the restriction $0 \leq h_i \leq 1$ forces $B = 0$. Thus $h_i = 1$ for all i. So even if you find a fair casino you are certain to end up broke. This apparent paradox is called **gambler's ruin**.

Discrete-Time Random Walk



$$P_N(r) = pP_{N-1}(r-1) + qP_{N-1}(r+1), \quad r \in \mathbb{Z}, \quad N \geq 1.$$

Characteristic function (fixed N):

$$G_N(k) = \sum_{r=-\infty}^{\infty} e^{ikr} P_N(r), \quad k \in [-\pi, \pi].$$

generates moments of the random displacement variable r

$$\left(-i\frac{d}{dk}\right)^m G_N(k)\bigg|_{k=0} = \sum_{r=-\infty}^{\infty} r^m P_N(r) = \langle r^m \rangle,$$

Multiply by e^(-ik) & sum over r

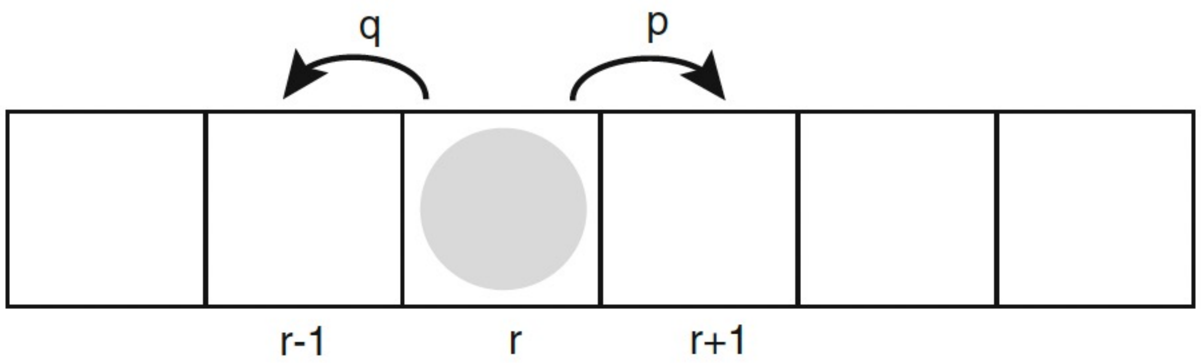
$$G_N(k) = (pe^{ik} + qe^{-ik})G_{N-1}(k).$$

$P_0(r) = \delta_{r,0}$ and $G_0(k) = 1$,

$$G_N(k) = u(k)^N \quad u(k) = pe^{ik} + qe^{-ik}.$$

Discrete Fourier Transform

Discrete-Time Random Walk



$$P_N(r) = pP_{N-1}(r-1) + qP_{N-1}(r+1), \quad r \in \mathbb{Z}, \quad N \geq 1.$$

$$G_N(k) = u(k)^N \quad u(k) = pe^{ik} + qe^{-ik}.$$

Discrete Fourier Transform

$$\begin{aligned} P_N(r) &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikr} u(k)^N dk \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ikr} \sum_{m=0}^{N} \binom{N}{m} p^m q^{N-m} e^{-ik(N-2m)} dk \\ &= \frac{N!}{\left(\frac{N+r}{2}\right)!\left(\frac{N-r}{2}\right)!} p^{(N+r)/2} q^{(N-r)/2} \end{aligned}$$

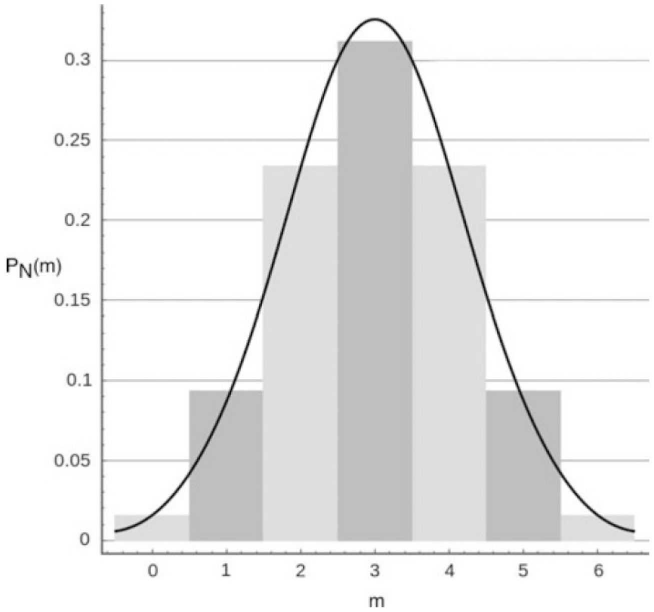
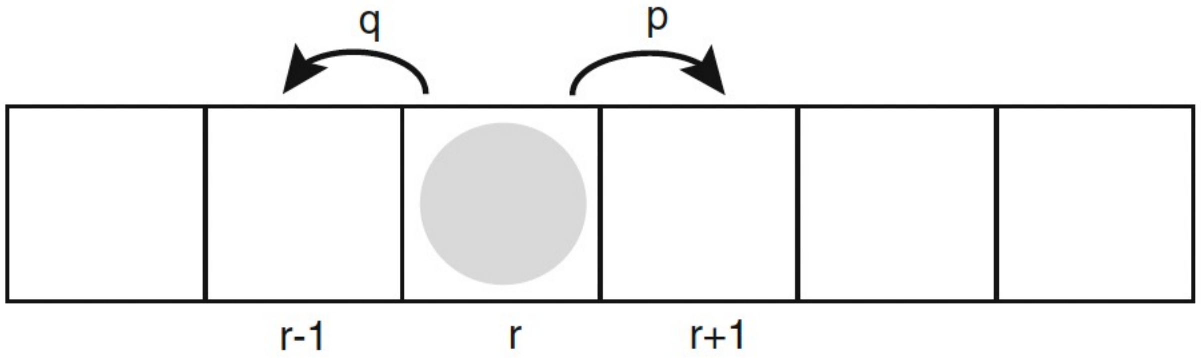


Fig. 2.2: Binomial distribution for $N = 6$ and $p = q = 1/2$. Also shown is a Gaussian fit of the binomial distribution

Discrete-Time Random Walk

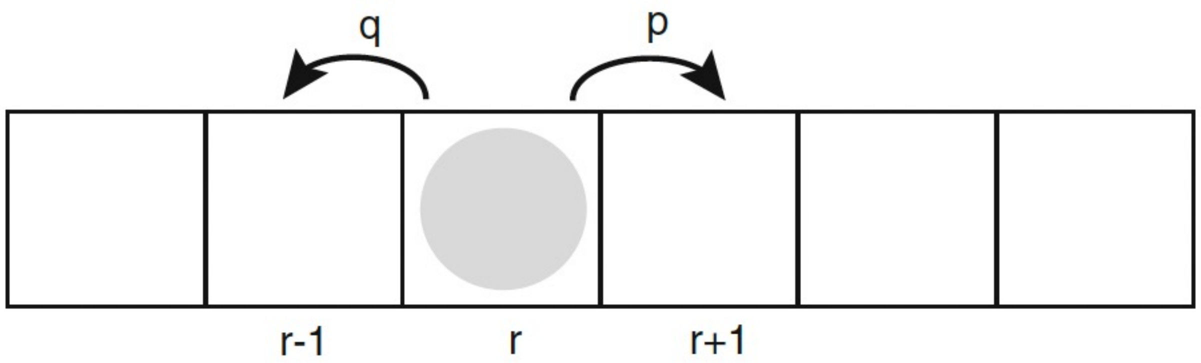


$$P_N(r) = pP_{N-1}(r-1) + qP_{N-1}(r+1), \quad r \in \mathbb{Z}, \quad N \geq 1.$$

q > p -> biased to the left
p > q -> biased to the right
p = q = ½ -> unbiased or Bernoulli distribution!

$$P_N(r) \sim \frac{1}{\sqrt{2\pi N}} \mathrm{e}^{-[r-N(p-q)]^2/2N}.$$

Discrete-Time Random Walk



$$P_N(r) = pP_{N-1}(r-1) + qP_{N-1}(r+1), \quad r \in \mathbb{Z}, \quad N \geq 1.$$

q > p -> biased to the left
p > q -> biased to the right
p = q = ½ -> unbiased or Bernoulli distribution!

$$P_N(r) \sim \frac{1}{\sqrt{2\pi N}} \mathrm{e}^{-[r-N(p-q)]^2/2N}.$$

Discrete-Time Random Walk – Generating function

$$\Gamma(r,z) = \sum_{N=0}^{\infty} z^N P_N(r).$$

discrete Laplace transform

$$\hat{\Gamma}(k,z) \equiv \sum_{r=-\infty}^{\infty} \mathrm{e}^{ikr}\Gamma(r,z) = \sum_{N=0}^{\infty} z^N G_N(k),$$

Fourier space

$$\hat{\Gamma}(k,z) = \frac{1}{1-zu(k)}.$$

Summed

Discrete-Time Random Walk – 2D

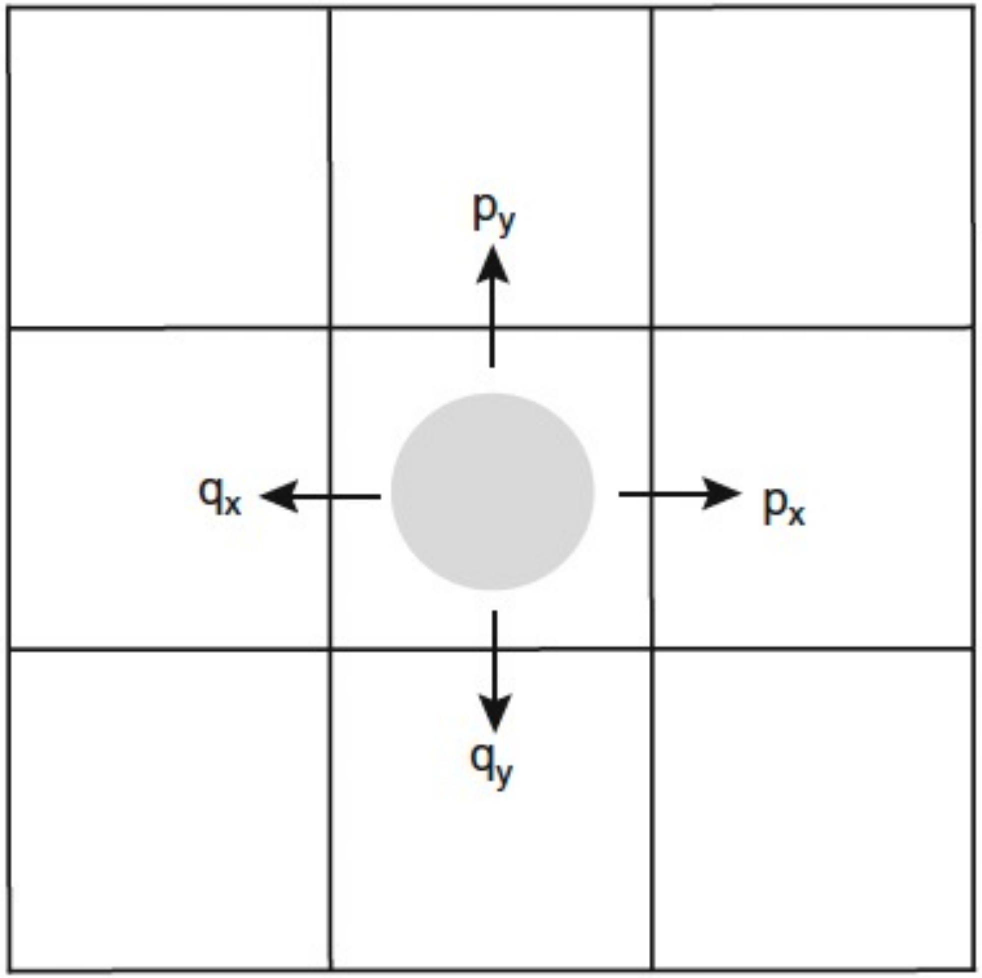


Fig. 2.3: A random walk on a 2D square lattice with $p_x + q_x + p_y + q_y = 1$

Continuous random walks (unbiased)

step lengths δx and δt for space and time $\quad P_N(r) = \rho(x,t)\delta x$ with $x = r\delta x, t = N\delta t$.

$$\rho(x,t) = p\rho(x-\delta x, t-\delta t) + q\rho(x+\delta x, t-\delta t)$$
$$\approx (p+q)\left[\rho(x,t) - \frac{\partial \rho}{\partial t}\delta t\right] - (p-q)\frac{\partial \rho}{\partial x}\delta x + \frac{(p+q)}{2}\frac{\partial^2 \rho}{\partial x^2}\delta x^2$$

drift

$$V = \lim_{\delta x, \delta t \to 0} (p-q)\frac{\delta x}{\delta t},$$

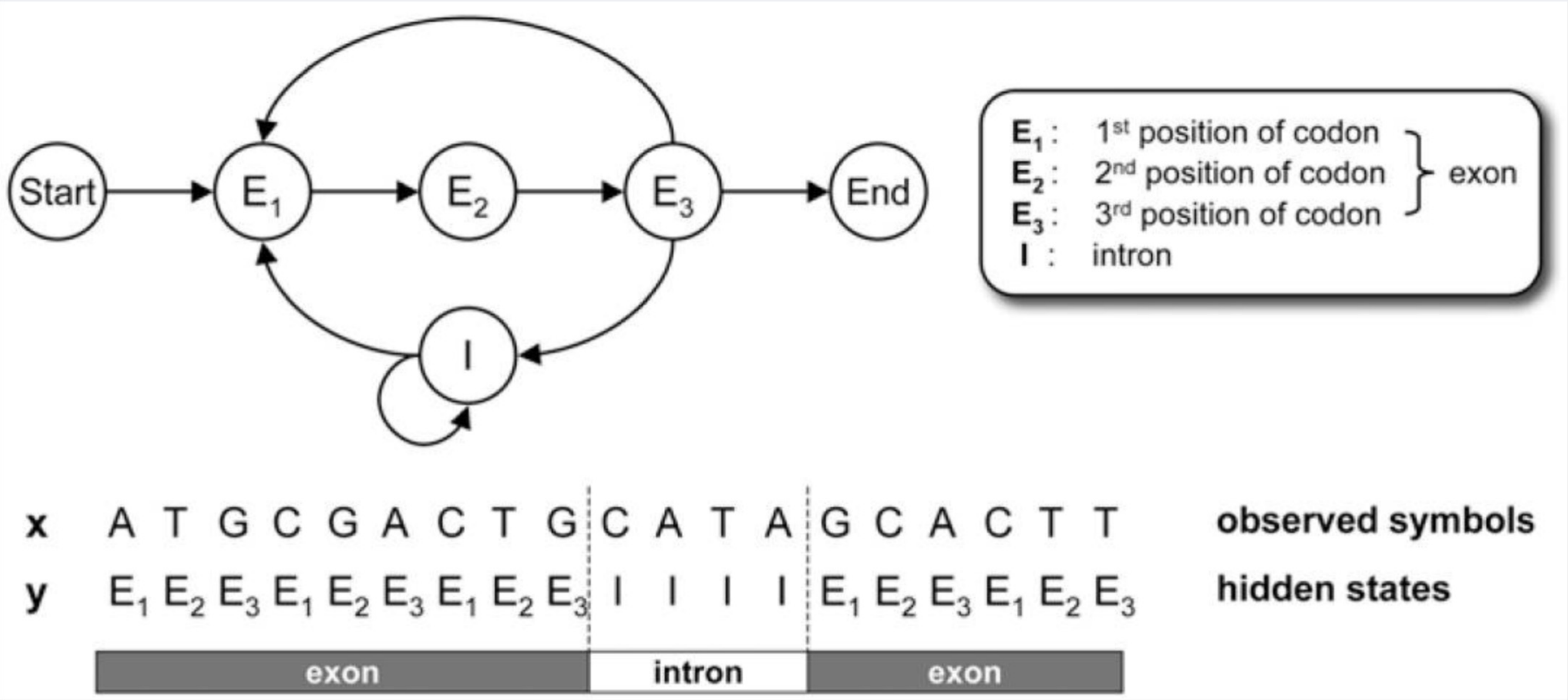
diffusivity

$$D = \lim_{\delta x, \delta t \to 0} \frac{\delta x^2}{2\delta t},$$

$$\frac{\partial \rho(x,t)}{\partial t} = -V\frac{\partial [\rho(x,t)]}{\partial x} + D\frac{\partial^2 \rho(x,t)}{\partial x^2}.$$

Bonus: HMM used for DNA sequence analysis

hidden Markov model is one where the rules for producing the chain are unknown or "hidden."



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2766791/