Math 612: Single Cell Analysis **2019W Term 1**

Lecture 8: October 1

Lecturer: Geoffrey Schiebinger *Scribe: Sibyl Drissler*

8.1 Markov Processes

8.1.1 Finite State Markov Processes

Definition 8.1 *Transition Matrix: What is X_{t+1} given X_t?*

$X_t = \{\text{Eat, Sleep, Active}\}, t = 1, 2, 3$

$$M = \begin{array}{c} \\ \\ \\ \end{array}\begin{array}{cccc} \text{E} & \text{A} & \text{S} & \\ \end{array}$$

$$M = \begin{array}{cccc} \text{E} & \text{A} & \text{S} & \\ 0.5 & 0 & 0.5 & \text{E} \\ 0.5 & 0.5 & 0 & \text{A} \\ 0 & 0.5 & 0.5 & \text{S} \end{array}$$

Note: The rows sum to 1.

If we know $x_o = E$, $X_1 \sim M \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$. If we think that the state at time t is $P_t = \begin{pmatrix} 0.1 \\ 0.8 \\ 0.1 \end{pmatrix}$, then we can push P_t through M to get the distribution over states at time $t + 1$.

$P_{t+1} = MP_t$

so for $x_o = E$, $X_2 \sim MX_1 = M^2 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$.

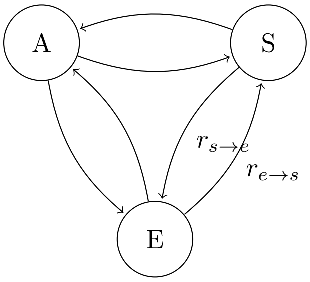
The transition from time 1 to time 3, $X_3|X_1$, is M^2. **(Compostition property of the transition matrix)**

As $t \to \infty$, $P_t \to$ top eigenvector of M, P_∞ (stationary distribution).

$P_\infty M = P_\infty$, ie $\lambda = 1$

Exercise to check: A matrix whose rows sum to 1 will always have an eigenvector with $\lambda = 1$.

8.1.2 Finite State Continuous Time Markov Processes

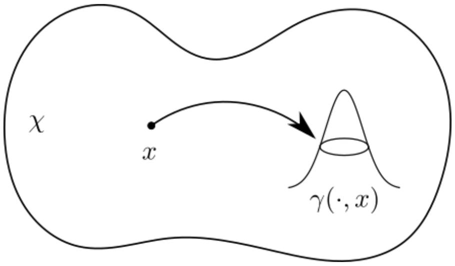


Transition rates, eg $r_{e\to s}$, $r_{s\to e}$, are exponential random variables. Whichever event happens first is the path you choose.

8.1.3 Continuous Time, Continuous State Markov Processes

Definition 8.2 *A transition kernal is a map from point x in state space χ to distributions on state space.*

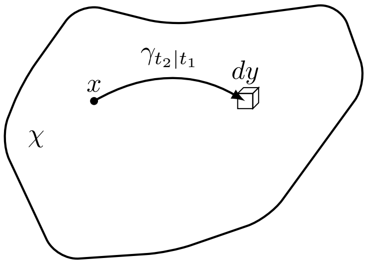
$$\gamma : \underbrace{x}_{\text{point}} \longmapsto \underbrace{\gamma(\cdot|x)}_{\text{probability measure}}$$



The probability that x transitions to somewhere in A is $\gamma(A|x)$ for any set $A \subset \chi$.

A continuous time Markov process has a transition kernel for every pair of time points $t_1 < t_2$.

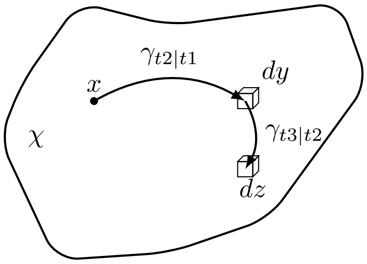
$\gamma_{t_2|t_1}(\cdot|x)$ is the probability distribution for the state at time t_2 given we start from state x at time t_1. $\gamma_{t_2|t_1}(dy|x)$ is the probability of transitioning into dy.



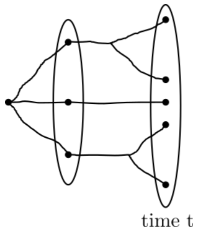
Denote the state at time t by X_t. Then $\gamma_{t_2|t_1}(\cdot|x)$ is the law of $X_{t_2}|X_{t_1} = x$.

8.1.3.1 Compositional Property

$$\begin{aligned}(\gamma_{t_2|t_1} \circ \gamma_{t_3|t_2})(dz|x) &= \int \gamma_{t_2|t_1}(dy|x)\gamma_{t_3|t_2}(dz|y) \\ &= \gamma_{t_3|t_1}(dz|x) \quad \text{definition of Markov}\end{aligned} \tag{8.1}$$



8.2 Developmental Stochastic Processes



$\mathbb{P}$ describes the population of cells at time t. $\gamma_{t2|t1}$ describes the "transitions". We interpret $\gamma_{t2|t1}(\cdot|x)$ as the distribution of descendants of x.

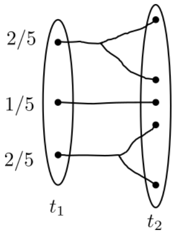
$$\gamma : \underbrace{x}_{\text{at time } t_1} \longmapsto \underbrace{\gamma_{t_2|t_1}(\cdot|x)}_{\text{descendants at time } t_2\text{, probability measure on } \chi}$$

For any set $A \subset \chi$, $\gamma(A|x)$ is the probability that the cell x has a descendant in A.

Note:

Consider a random cell X_{t_2} at time t_2. $X_{t_2} \sim \mathbb{P}_{t_2}$. Define A_{t_1} to be the unique ancestor of X_{t_2}.

$A_{t_1} = Ancest(X_{t_2})$



$A_{t_1} \sim \mathbb{Q}_{t1} \neq \mathbb{P}_{t_1}$

We can define a joint distribution $\gamma_{t_2,t_1} = (A_{t_1}, X_{t_2})$. There is a nice relationship between $\gamma_{t_2|t_1}$ and γ_{t_2,t_1}. Details next class.

8.2.1 Sampling from an Elemental Stochastic Process with a scRNA-seq Time Course

Goal: Learn about $\gamma_{t2|t1}$. But measurements kill cells. So we can't look at transitions in high dimensional χ.

Steps:

1. Prepare independent populations following the same process
2. Sample at different time points

At time t_1 get samples $X_1, X_2, ..., X_{n1} \sim \mathbb{P}_{t1}$
At time t_2 get samples $Y_1, Y_2, ..., Y_{n2} \sim \mathbb{P}_{t2}$

$\vdots$

At time t_T get samples $Z_1, Z_2, ..., Z_{nT} \sim \mathbb{P}_{tT}$

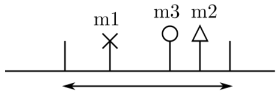
We can construct $\hat{\mathbb{P}}_{t1} = \sum_{i=1}^{n_i} \delta_{x_i}$, but how do we construct $\gamma_{t2|t1}$?

Methods to construct lineage trajectories:

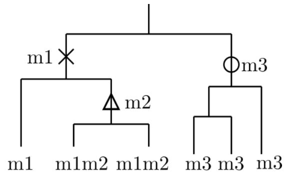
1. Computationally infer from samples $\hat{\mathbb{P}}_{t1}, \hat{\mathbb{P}}_{t2}, \hat{\mathbb{P}}_{t3}, \ldots$
2. Lineage tracing at time t_i

   Gives information on lineage tree, but not state of the ancestors.

   Use CRISPR to create mutations in an unimportant part of DNA. It is still a bit of an open problem how to merge this data with scRNA-seq.



Target site for Lineage Tracing



3. RNA velocity

   If we read enough of the RNA sequence, we can tell if it has been spliced or not. Splicing occurs with a given rate. The longer mRNA has been around, the more likely it has been spliced.