Applied Probability

Michael Li

March 25, 2017

Finite-state continuous-time Markov chains: basic properties. Q-matrix, backward and forward equations. The homogeneous Poisson process and its properties (thinning, superposition). Birth and death processes. [6]

General continuous-time Markov chains. Jump chains. Explosion. Minimal Chains. Communicating classes. Hitting times and probabilities. Recurrence and transience. Positive and null recurrence. Convergence to equilibrium. Reversibility. [6]

Applications: the M/M/1 and M/M/fk queues. Burkefis theorem. Jacksonfis theorem for queueing networks. The M/G/1 queue. [4]

Renewal theory: renewal theorems, equilibrium theory (proof of convergence only in discrete time). Renewal-reward processes. Littlefis formula. [4]

Moran and Wright-Fisher models. Kingmansfis coalescent. Infinite sites and infinite alleles models. Ewensfi sampling formula. [4]

Contents

1 Introduction to Continuous Time Markov Chain

We start by reviewing the definition of a discrete time Markov Chain.

Definition. The process is called an *discrete-time Markov Chain* with state space S iff for all $x_0, x_n \in S$ we have:

$$P(X_n = x_n | X_0 = x_0, \cdots X_{n-1} = x_{n-1}) = P(X_n = x_n | X_{n-1} = x_{n-1})$$

In the similar spirit, we define a continuous-time one:

Definition. The process X_t is called a *continuous-time Markov Chain* if for all $x_1, \cdots x_n \in S$ and all times $0 \leq t_1 \leq t_2, \cdots \leq t_n$ we have:

$$\mathbb{P}(X_{t_n} = x_n | X_{t_{n-1}} = x_{n-1} \cdots X_{t_1} = x_1) = \mathbb{P}(X_{t_n} = x_n | X_{t_{n-1}} = X_{n-1})$$

A *homogeneous* Markov chain is one where the righthand side only depends on $t_n - t_{n-1}$.
Similarly to discrete-time Markov chain, we define $P(t)_{xy} = P(X_t = y | X_0 = x)$ and call it the *transition semigroup* of this Markov Chain.

This semigroup is a stochastic matric for all t, and satisfies $P(t+s) = P(t)P(s)$ with $P(0) = I$.

1.1 What's Different

We define the process to be *right-continuous* to avoid theoretical measure theory process, which means the process always stay cosntant for an interval after t (that interval is stochastic). And we define the following terms:

Definition. The *jump times* $J_0, J_1 \cdots$ are defined as $J_{n+1} = \inf\{t \geq J_n, X_t \neq X_{J_n}\}$, with $J_0 = 0$. We can see that defining the *jump chain* with $Y_n = X_{J_n}$ returns us to the discrete space. The *holding time* is $S_n = J_n - J_{n-1}$, which is > 0 by right-continuity. We also define the *explosion time* to be:

$$\eta = \sup_n J_n$$

Note η can be finite! And we don't want to consider what happens after η, so we set $X_t = \infty$ after that.

Theorem. The holding time S satisfies the memoryless property and thus has the exponential distribution.

Proof.

$$\begin{aligned} P(S_x > t+s | S_x > s) &= P(X_u = x \forall u \in [0, t+s] | X_u = x \forall u \in [0, s]) \\ &= P(X_u = x \forall u \in [s, t+s] | X_u = x \forall u \in [0, s]) \\ &= P(X_u = x \forall u \in [0, t] | X_0 = x) = P(S_x > t) \end{aligned}$$

This proves that S has the memoryless property. Set $F(t) = P(S > t)$. Then we know:

$$F(t+s) = F(t)F(s)$$

So $F(q) = F(\frac{1}{n})^n$, and there exists n large enough so that $F(\frac{1}{n}) > 0$ by right-continuity, so we can set $F(1) = e^{-\lambda}$ for some $\lambda \geq 0$. Then we thus have $F(k) = e^{-\lambda k}$ for all positive integers and also rationals. As the function is decreasing and rational numbers are dense in reals, we achieve the same result on the real numbers. □

1.2 Poisson Process

We give multiple equivalent definitions of the Poisson process:

Definition. The following definitions of a Poisson process with rate λ are equivalent:

- The *holding times* are iid exponentially distributed with parameter λ.
- X has independent increments and as $h \to 0$ uniformly in t we have:

$$P(X_{t+h} - X_t = 1) = \lambda h + o(h)$$

$$P(X_{t+h} - X_t = 0) = 1 - \lambda h + o(h)$$

- X has independent and stationary incrementsand for all $t \geq 0$ we have $X_t \sim Poi(\lambda t)$.

Similar to the discrete case, we can prove this process satisfies the Markov and Strong Markov property, which is stated below:

Theorem (Markov Property). Let (X_t) be a Poisson process of rate λ. Then for all $s \geq 0$ the process $(X_{s+t} - X_s)_{t \geq 0}$ is also a Poisson process of rate λ and is independent of X_r for all $r \leq s$.

Theorem (Strong Markov Property). Let (X_t) be a poisson process with rate λ and let T be a stopping time. Then conditional on T, the process $(X_{T+s} - X_T)$ for $s \geq 0$ is also a poisson process of rate λ and is independent from X_s with $s \leq T$.

Note. Don't remember what's a stopping time? It is just an event such that for all t the event $T \leq t$ only depends on X_s with $s \leq t$.

To avoid repeating this over and over again, we will note that *all* continuous-time Markov chains satisfy the two properties above.

Now we are ready to prove that the three definitions are actually equivalent:

Proof.

$(i) \to (ii)$ if (i) holds then by the Markov property the increments are independent and stationary. By stationarity we have:

$$P(X_{t+h} - X_t = 0) = P(X_h = 0) = P(S_1 > h) = e^{-\lambda h} = 1 - \lambda h + o(h)$$

And we can similarly prove for the ≥ 1 and ≥ 2 case to reach the $= 1$ case as required.

$(ii) \to (iii)$ Note that if X satisfies (ii) then $X_{t+s} - X_t$ satisfies (ii) so the increments are stationary. So we will now prove that X_t has the Poisson distribution. Set $p_j(t) = P(X_t = j)$.

$$p_j(t+h) = p_j(t)(1 - \lambda h + o(h)) + p_{j-1}(t)(\lambda h + o(h)) + o(h)$$

From the infinitesimal definition (ii). Then rearranging and setting $s = t + h$ we have:

$$\frac{p_j(s) - p_j(s-h)}{h} = -\lambda p_j(s-h) + \lambda p_{j-1}(s-h) + o(1)$$

From this equation above we see that $p_j(t)$ is a differentiable function as we take $H \to 0$ with:

$$p_j'(t) = -\lambda p_j(t) + \lambda p_{j-1}(t)$$

Then we differentiate $e^{\lambda t} p_k(t)$ with respect to t to get that:

$$(e^{\lambda t} p_k(t))' = \lambda e^{\lambda t} p_{k-1}(t)$$

For $k = 0$ we have $p_0'(t) = -\lambda p_0(t)$ (Since if we are at 0 at $t + h$ then we are most definitely at 0 at time t), so we have $p_0(t) = e^{-\lambda t}$. Then indudctively from the equation inductively we have:

$$p_n(t) = e^{\lambda t} \frac{(\lambda t)^n}{n!}$$

Which is the Poisson distribution!

$(iii) \to (i)$ Note that a Poisson process satisfies this and this completely characterizes all the marginals of the process immediately. So (c) implies (a) and (b).

□

The next theorem is *important*.

Theorem (Thinning Property). Let X be a Poisson process of parameter λ and Z_i be iid bernoulli random varaibles with probability p. Then let Y be a process which jumps at time t iff X jumps and $Z_{X_t} = 1$. So we basically keep each jump with probability p. Then Y is a Poisson process of parameter λp and $X - Y$ is an independent Poisson process of parameter $\lambda(1-p)$.

We can use the infinitesimal definition of the Poisson process to prove it, but it is long, boring and easy so we would ignore it here.

Theorem. Let X be a Poisson process. Conditional on the event $\{X_t = n\}$ the jump times $J_1 \cdots J_n$ have joint density functions:

$$f(t_1, \cdots t_n) = \frac{n!}{t^n} 1(0 \le t_1 \le \cdots \le t_n < t)$$

Proof. We outline the main ideas of the proof here and leave out the boring check of algebra.

$S_1, \cdots, S_n$ are indeepndent, so their joint density is:

$$\lambda^{n+1} e^{-\lambda(s_1, \cdots + s_{n+1})} 1(s_1, \cdots, s_{n+1} \ge 0)$$

Then the jump times as defined by $J_1 = S_1, J_2 = S_1 + S_2, \cdots$. Then either by a variable transformation (and using Jacobian) or through induction we can see that $J_1, \cdots, J_{n+1}$ have the joint distribution:

$$\lambda^{n+1} e^{-\lambda t_{n+1}} 1(0 \leq t_1 \leq \cdots \leq t_{n+1})$$

where t_{n+1} is the time of J_{n+1}. □

1.3 Birth Process

This is just the generalization of the Poisson process in which the rate of change depends on the current state, so the rate from state i to $i+1$ is not λ but some q_i fk(so the waiting time S_i is exponentially distributed with q_i).

Most of the stuff follows *stragiht* from the Poisson process, such as the Markov property.

To do important stuff, we need the following extension fo a theorem in Statistics IB:

Proposition. Let (T_k) be a sequence of independent random variables with $T_k \sim Exp(q_k)$ and $0 < q = \sum_k q_k < \infty$. Then set $T = \inf_k T_k$, then this infimum is attained at unique K with probability 1, and with T, K independent, $T \sim Exp(q)$ and $P(K = k) = \frac{q_k}{q}$.

Proposition. With X a birth process, rates (q_i) and $X_0 = 1$, then $\sum \frac{1}{q_i} < \infty$ iff $P(\eta < \infty) = 1$.

Proof.

- $\Rightarrow$ If $\sum \frac{1}{q_n} < \infty$, then $E[\sum_n S_n] = \sum_n \frac{1}{q_n} < \infty$m so $P(\sum_n S_n < \infty) = 1$.
- $\Leftarrow$ if $\sum \frac{1}{q_i} = \infty$, then $\prod_n (1 + \frac{1}{q_n}) = \infty$. By monotone convergence and independence:

  $$E[\exp\left(-\sum S_n\right)] = \prod E[\exp(-S_n)] = \prod \left(1 + \frac{1}{q_n}\right)^{-1} = 0$$

  Then since $e^{-\sum S_n} \geq 0$ with $= 0$ iff $\sum S_n = \infty$, then $\sum_n S_n = \infty$ with probability 1.

□

Note. The clever part in this proof is to look at e^{-X} when we want to prove that $P(X = \infty) = 1$.

We also have two equivalent definitions of the birth chain:

Definition. The following equivalent definitions define a *birth chain*:

- Condition on $X_0 = i$, the holding times $S_i \sim \exp(q_i)$ and the jump chain is given by $Y_n = i + n$.
- For all $t, h \geq 0$, conditional on $X_t = i$, the process $(X_{t+h})_{h \geq 0}$ is independent of X_s for all $s \geq t$, and as $H \downarrow 0$ uniformly in t we have:

  $$P(X_{t+h} = i | X_t = i) = 1 - q_i h + o(h) \qquad P(X_{t+h} = i+1 | X_t = i) = q_i h + o(h)$$

1.4 Three Constructions of Continuous-time Markov Chains

Definition. let S be a countable set. Then a Q*-matrix* on S is a matrix that satisfies $0 \leq -q_{ii} < \infty$, $q_{ij} \geq 0$ for all $I \neq j$, and $\sum_j q_{ij} = 0$ for all i.

We then define $q_i = -q_{ii}$ and given a Q-matrix we define the *jump stochastic matrix* as followed:

For $q_x \neq 0$:

$$p_{xy} = \frac{q_{xy}}{q_{xx}} \qquad p_{xx} = 0$$

If $q_x = 0$, then $p_{xy} = \mathbf{1}_{\{x=y\}}$.

Now we can finally formally define a continuous=time Markov chain:

Definition. A markov chain X with $X_0 = \lambda$, generator Q, a Q-matrix, is a stochastic process with jump chain $Y_n = X_{J_n}$ being a discrete time Markoc chain with $Y_0 = \lambda$ and transition matrix P. Moreover, conditional on $Y_0, \cdots Y_n$, the holding times $S_i \sim \exp(q_{Y_{i-1}})$.

We now give three constructions:

- Take a discrete time Markov chain Y with initial distribution λ and transition matrix P. Then take a sequence of iid $S_i \sim \exp(q_{Y_{i-1}})$ and let $J_n = \sum_{i=1}^n S_i$. Then set $X_t = Y_n$ if $J_n \leq t < J_{n+1}$ and $X_t = \infty$ otherwise.
- Let $(T_n^y)_{n\geq 1, y \in S}$ be iid $\exp(1)$ random variables. Inductively define Y_n, S_n as follows: $Y_0 \sim \lambda$ and inductively if $Y_n = x$ then we set for $y \neq x$

$$S_{n+1}^y = \frac{T_{n+1}^y}{q_{xy}} \qquad S_{n+1} = \int_{y \neq x} S_{n+1}^y$$

- Consider independent Poisson processes N_t^{xy} for each pair of points x, y with $x \neq y$ and parameter q_{xy}. Define Y_n, J_n inductively: Let $Y_0 \sim \lambda$ and set $J_0 = 0$. If $Y_n = x$, then we have:

$$J_{n+1} = \inf\{t > J_n, N_t^{Y_n y} \neq N_{J_n}^{Y_n y} \text{ for some } y \neq Y_n\}$$

$$Y_{n+1} = \begin{cases} y & \text{if } J_{n+1} < \infty \text{ and } N_{J_{n+1}}^{Y_n y} \neq N_{J_n}^{J_n y} \\ x & \text{if } J_{n+1} = \infty \end{cases}$$

Note. These are not important and are just different ways of looking at the same thing.

We have the following theorem:

Theorem. Let X be a continuous time Markov chain with generator Q and initial distribution λ. [We will write $\text{Markov}(Q, \lambda)$ below].Then $P(\eta = \infty) = 1$ if we have a finite state space, $\sup_x q_x < \infty$, or $X_0 = x$ and x is recurrent for the jump chain.

Proof. We would only prove the case for $\sup_x q_x < \infty$. Let $q = \sup_x q_x$. Then the holding times satisfy $S_n \geq \frac{T_n}{q}$ where $T_n \sim \exp(1)$ and iid. Then:

$$\eta = \sum S_n \geq \frac{1}{q} \sum T_n = \infty$$

with probability 1 from the Strong Law of Large numbers. □

1.5 Kolmogorov's Forward and Backward Equations

We have the following important theorem that lay the foundation of continuous-time markov chains.

Theorem (Kolmogorov)**.** Let X be a minimal right continuous process with values in a countable set S and let Q be a Q-matrix with jump matrix P. Then the following conditions are equivalent:

- X is a continuous time Markov chain with generator Q.
- For all $n \geq 0$, $0 \leq t_0 \leq \cdots \leq t_n$ and all states $x_0, \cdots, x_{n+1}$:

  $$\mathbb{P}(X_{t_{n+1}} = x_{n+1} | X_{t_n} = x_n, \cdots, X_{t_0} = x_0) = p_{x_n x_{n+1}}(t_{n+1} - t_n)$$

  With $p_{xy}(t)$ the minimal non-negative solution to the *backward equation*:

  $$P'(t) = QP(t) \qquad P(0) = I$$

- Same as above, but $p_{xy}(t)$ is the minimal non-negative solution to the *forward equation*:

  $$P'(t) = P(t)Q \qquad P(0) = I$$

Only works for finite state space Conditional on $X_s = x$ the process $(X_{s+t})_{t \geq 0}$ is independent of $(X_r)_{r \leq s}$ and uniformly in t as $h \downarrow 0$ for all x, y:

$$P(X_{t+h} = y | X_t = x) = \mathbf{1}_{x=y} + q_{xy}h + o(h)$$

For a finite state space, we have $P(t) = e^{tQ}$, and the solution to the forward and backward equations are *unique*.

Remark. This is a lot to take in, but really the important parts are really the two equations, and the form in the finite state space case.

Proof.

$(i) \to (ii)$ Define $P(t)$ by setting $p_{xy}(t) = P(X_t = y | X_0 = x) = P_x(X_t = y)$. We show that $P(t)$ is the minimal non-negative solution to the backward equation.

Using (J_n) denote the jump times of the chain, we have:

$$P_x(X_t = y, J_1 > t) = e^{-q_x t}\mathbf{1}_{x=y}$$

For $J_1 \leq t$, we have:

$$P_x(X_t = y, J_1 \leq t, X_{J_1} = z) = \int_0^t \underbrace{q_x e^{-q_x s}}_{\text{PDF of transition}} \underbrace{\frac{q_{xz}}{q_x}}_{\text{Prob. of going to } z} \underbrace{p_{t-s}(z, y)}_{\text{Prob. of going from } z \text{ to } y} \mathrm{d}s$$

Then taking the sum over all $z \neq x$ and multiplying by $e^{-q_x t}$ we have:

$$e^{q_x t} p_{xy}(t) = \mathbf{1}(x = y) + \int_0^t \sum_{z \neq x} e^{q_x u} q_{xz} p_u(z, y) \mathrm{d}u$$

Thus $p_{xy}(t)$ is differentiable in t (Note that the sum has to be a uniformly convergent series of continuous functions), so then we take the derivative and get that:

$$p'_{xy}(t) = \sum_z q_{xz} p_t(z, y)$$

After arranging, which is $P'(t) = QP(t)$. Then let $\tilde{P}$ be another non-negative solution of the backward equation. First, note that the "integral-sum" expression above can be equivalently written as:

$$P_x(X_t = y, t < J_{n+1}) = e^{-q_x t}\mathbf{1}(x = y) + \sum_{z \neq x} \int_0^t e^{-q_x s} q_{xz} p_z(X_{t-s} = y, t-s < J_n) \mathrm{d}s$$

Then we reverse the steps above to see that: we have:

$$\tilde{p}_{xy}(t) = e^{-q_x t}\mathbf{1}(x = y) + \int_0^t \sum_{z \neq x} e^{-q_x s} q_{xz} \tilde{p}_{t-s}(z, y) \mathrm{d}s$$

We then prove $P_x(X_t = y, J_n < t) \leq \tilde{p}_{xy}(t)$ for all n inductively. $n = 1$ follows trivially from the expression above, and the indcution step is also trivial by the revised integral-sum expression we derived above. Then:

$$p_{xy}(t) = P_x(X_t = y, t < \eta) = \lim_{n \to \infty} p_x(X_t = y, t < J_n) \leq \tilde{p}_{xy}(t)$$

$(ii) \to (i)$ Since this uniquely determines all the marginals, then this must be a continuous time Markov chain with generator Q.

$(ii) \leftrightarrow (iii)$ This is obvious in the finite state-space case where $P(t) = e^{tQ}$ as we can differentiate term by term. Moreover, $P(t)$ is unique as suppose $\tilde{P}$ is another solution to the backward equation, then some calculus would show that:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\tilde{P}(t)e^{-tQ}) = 0$$

And with $\tilde{P}(0) = I$ we have $\tilde{P}(t) = e^{tQ}$. *The infinite case is not proven and not important.*

$(i) \to (iv)$ On a finite state space, we have $P(t) = e^{tQ}$, so as $t \downarrow 0$, we have $P(t) = I + tQ + O(t^2)$, giving the result trivially.

$(iv) \to (iii)$ Now using the infinitesimal definition, we have:

$$p_{xy}(t + h) = \sum_z (\mathbf{1}_{z=y} + q_{zy}h + o(h)) p_{xz}(t)$$

Rearranging gives:

$$\frac{p_{xy}(t + h) - p_{xy}(t)}{h} = \sum_z p_{zy} p_{xz}(t) + O(h)$$

Uniformly for all t. Taking $h \to 0$ gives $p'_{xy}(t) = \sum_z p_{xz}(t) q_{zy}$, which is the forward equation.

□

Note. In actual questions, if you must calculate e^{tQ}, it is easier to do so through diagonalizing Q. If $Q = UDU^{-1}$ where $D = \text{diag}(k_1, \cdot k_n)$, then:

$$e^{tQ} = U \, \text{diag}(e^{k_1 t}, \cdots, e^{k_n t}) U^{-1}$$

2 Properties of Continuous-time Markov Chains

Note. In this section, we would see that almost everything copies from the jump chain structure. *Almost.*

2.1 Hitting Times, Recurrence, and Invariant Distributions

This is literally a recap of the discrete case. Let $T_A = \inf\{t > 0 : X_t \in A\}$ for some $A \subseteq S$. Then let $h_A(x) = P_x(T_A < \infty)$ and $k_A(x) = E_x[T_A]$. Then we have the following theorem:

Theorem. Let $(h_A(x))_{x \in S}$ and $(k_A(x))_{x \in S}$ be the vector of *hitting probabilities* and *hitting times* respectively. They are the minimal non-negative solution to the following two sets of equations respecitvely:

$$\begin{cases} h_A(x) = 1 & \forall x \in A \\ Qh_A(x) = \sum q_{xy} h_A(y) = 0 & \forall x \notin A \end{cases} \quad \begin{cases} k_A(x) = 0 & \forall x \in A \\ Qk_A(x) = \sum q_{xy} k_A(y) = -1 & \forall x \notin A \end{cases}$$

Note. Proofs follow that of the discrete case and are not important. But do note that the values for the hitting time differ fromt he discrete case as we need to consider the exponential waiting time.

Next up is recurrence:

Definition. We call a state *recurrent* if $P_x(\{t : X_t = x\} \text{ is unbounded}) = 1$. It is *transient* if this probability is 0.

Theorem. Let X be a continuous-time Markov Chain and Y its jump chain. Then:

- x is recurrent/transient for $Y \Rightarrow x$ is recurrent/transient for X.
- Every state is either recurrent or transient.
- x is recurrent iff $\int_0^\infty p_{xx}(t) \mathrm{d}t = \infty$.

Proof.

- Suppose x is recurrent and $X_0 = x$. Then X cannot explode, and $P_x(\eta = \infty) = 1$, so $J_n \to \infty$ with probability 1. Then since X visits x infintiely many times, the set $\{t \geq 0 : X_t = x\}$ is unbounded.

  If x is transient then let N be $\sup\{n : Y_n = x\}$. Then we see that Y_n for $n \leq N$ cannot have an abosrobing state, so $J_{N+1} < \infty$, so $\{t \geq 0 : X_t = x\}$ is bounded by J_{N+1}.

- This follows from the discrete case.

Important!

$$\int_0^\infty p_{xx}(t)\mathrm{d}t = E_x\left[\int_0^\infty \mathbf{1}(X_t = x)\mathrm{d}t\right] = E_x\left[\sum_n S_{n+1}\mathbf{1}(Y_n = x)\right]$$
$$= \sum P_x(Y_n = x)E_x[S_{n+1}|Y_n = x] = \frac{1}{q_x}\cdot\sum p_{xx}(n)$$

So then the third assertion follows from the discrete case.

□

Next up we have invariant distributions.

Definition. let Q be the generator of a continuous time Markov chain and let λ be a measure. It is *invariant* if $\lambda Q = 0$.

Theorem. Let X be an irreducible continuous time Markov chain with generator Q. Then:

- Some state is positive recurrent $\Leftrightarrow$ Every state is positive recurrent $\Leftrightarrow$ X is not explosive and has invariant distribution $\lambda(x) = (q_x m_x)^{-1}$, where $m_x = E_x[T_x]$, the mean expected time to return to state x.
- $\pi Q = 0 \Leftrightarrow \pi P(s) = \pi \forall s > 0 \Leftrightarrow \pi$ is invariant for $X \Leftrightarrow \mu_x = q_x\pi_x$ is invariant for the jump chain Y.

Proof. Once again, we provide an outline, because the method is *much* more important than the actual proof.

- We define the two following measures, using T_x and H_x to denote the first return time to state x in X and jump chain Y:

$$\nu(y) = E_x\left[\sum_{n=0}^{h_x-1}\mathbf{1}(Y_n = y)\right] \qquad \mu(y) = E_x\left[\int_0^{T_x}\mathbf{1}(X_s = y)\mathrm{d}s\right]$$

now exactly same as the *important* proof above, we can show that $\mu(y) = \frac{\nu(y)}{q_y}$. And recall that from IB Markov chains, $\nu(y)$ is exactly the invariant measure we constructed! Thus using the second theorem, $\mu(y)$ is invariant for X.

Now take x to be a recurrent state. If you look closely enough at the definition of $\mu(y)$, you would see that:

$$\sum \mu(y) = E_x[T_x] = m_x$$

But this is finite by recurrence. Then μ can be normalized by m_x to give λ, and it satisfies the equation $\lambda(x) = (q_x m_x)^{-1}$. Thus we proved the assertion one way.

Now assume that we already have an invariant distribution. Then $\beta(y) = \lambda(y)q_y$ is invariant for the jump chain Y. Then we introduce a lemma that is not proved:

Lemma. Let λ be another invariant measure with $\lambda(x) = 1$ and ν defined above. Then $\lambda(y) \geq \nu(y) \forall y$, with equality everywhere when Y is recurrent.

Then define $a(y) = \frac{\beta(y)}{\lambda(x)q_x}$. Then $a(x) = 1$ and using the lemma we have $a(y) \geq \nu(y)$. So:

$$m_x = \sum \mu(y) = \sum_y \frac{1}{q_y}\nu(y) \leq \sum \frac{a(y)}{q_y} = \frac{1}{q_x \lambda(x)} < \infty$$

for all x. thus we are done.

- We first prove the last equivalence:

$$\sum \pi(x) q_{xy} = \sum_x \pi(x) q_x (p_{xy} - \mathbf{1}(x = y)) = \sum_x \mu(x)(p_{xy} - \mathbf{1}(x = y))$$

So $\mu P = \mu \Leftrightarrow \pi Q = 0$.

Then we use the first theorem [Note this is not circular as the part of the second theorem the proof above needs is proven above], and note that $\mu Q = 0$ with μ defined above. Now:

$$\mu(y) = E_x\left[\int_0^{T_x} \mathbf{1}(X_s = y)\mathrm{d}s\right] = E_x\left[\int_s^{T_x+s} \mathbf{1}(X_t = y)\mathrm{d}t\right]$$
$$= \int_0^{\infty} \sum_z P_x(X_t = z, X_{t+s} = y, t < T_x)\mathrm{d}t = \sum_z p_{zy}(s)\mu(z)$$

Where the second equality follows from splitting $[0, T_x]$ to $[0, s]$ and $[s, T_x]$, and then using Strong Markov property to change the first range to $[T_x, T_x + s]$.

□

2.2 Equilibrium, Reversibility, and Ergodic Theorem

Note. In this section, we will start to phase out the terms irreducible and non-explosive and assume if not said otherwise, it is always irreducible and non-explosive. Also, unless specified otherwise, Q is the generator and π is the invariant distribution.

First, a small lemma:

Lemma. Let Q be a Q-matrix with semigroup $P(t)$. Then:

$$|p_{xy}(t+h) - p_{xy}(t)| = |\sum_{z \neq x} p_{xz}(h) p_{zy}(t) - p_{xy}(t)(1 - p_{xx}(h))| \leq 1 - p_{xx}(h) \leq 1 - e^{-q_x h}$$

The last inequality follows fromt he fsecond inequality follows from the fact that $1 - p_{xx}(h)$ is less than the probability of the waiting time being less than h.

Theorem. Let X be irreducible, non-explosive, and continuous Markov with generator Q. Suppose λ is invariant, then:

$$p_{xy}(t) \to \lambda(y) \mathrm{as} t \to \infty$$

Outline. We pick h small enough and let $Z_n = X_{nh}$ be a discrete chain. Then we know that $p_{xy}(nh) \to \lambda(y)$ as $n \to \infty$ by discrete Markov. Then by the lemma above, we can limit how much $p_{xy}(t)$ can vary between these $nh \leq t \leq (n+1)h$ points. Standard analysis techniques tell us we are done. $\square$

Ok, this is *really* boring. What about reversibility?

Theorem. Let X be a continuous Markov chain. Suppose $X_0 \sim \pi$. Fix $T > 0$ and set $\hat{X}_t = X_{T-t}$. Then $\hat{X}$ is Markov, irreducible, non explosive, with generator $\hat{Q}$ and invariant distribution π, where $\hat{q}_{xy} = \frac{\pi(y)q_{yx}}{\pi(x)}$.

Proof. It is easy to check that $\pi\hat{Q} = 0$ and Q is a Q matrix.

Set $0 = t_0 \leq t_1 \leq \cdots \leq t_n = T$ and $x_1, \cdots, x_n \in S$. Set $s_i = t_i - t_{i-1}$. Then:

$$\begin{aligned} P(\hat{X}_{t_0} = x_0, \cdots \hat{X}_{t_n} = x_n) &= \pi(x_n)p_{x_n x_{n-1}}(s_n) \cdots p_{x_1 x_0}(s_1) \\ &= \pi(x_0)\hat{p}_{x_0 x_1}(s_1) \cdots \hat{p}_{x_{n-1} x_n}(s_n) \end{aligned}$$

Then we can check that $\hat{P}(t)$ is the minimal non-negative solution to Kolmogorov's *backward* equations with generator $\hat{Q}$. It is quite easy to check that it is a solution (as $P'(t) = P(t)Q$, and $P(t)$ minimal), but minimality is key here.

let $R(t)$ be another solution. The trick is to define $\hat{R}_{xy}(t) = \frac{\pi(y)}{\pi(x)} R_{yx}(t)$. We can then check that $\hat{R}(t)$ satisfies Kolmogorov's forward equations, but $P(t)$ is minimal so $\hat{R} \geq P$. So $R \geq \hat{P}$.

Irreducibility follows trivially from Q. For non-explosiveness, note that $\sum_y \hat{p}_{xy}(t) = 1$ for all t as Q does not explode, so $\hat{\eta} = \infty$ almost surely. $\square$

Definition. Let X be a Markov chain with generator Q. It is called *reversible* if for all $T > 0$, the processes $(X_t)_{0 \leq t \leq T}$ and $(X_{T-t})_{0 \leq t \leq T}$ have the same distribution.

λ and Q are in detailed balance iff $\forall x, y$, we have:

$$\lambda(x)q_{xy} = \lambda(y)q_{yx}$$

Lemma.
- If λ and Q are in detailed balance, $\lambda Q = 0$. Moreover, if we have a birth and death chain (each state can only go to $+1$ or -1 states), then π is invariant iff it solves the detailed balance equations.
- Let X be continuous-Markov and $X_0 \sim \pi$. Then π and Q are in detailed balance $\Leftrightarrow$ $(X_t)_{t \geq 0}$ is reversible.

These proofs follow trivially. Then we state the ergodic theorem (not in syllabus):

Theorem. Let X be continuous Markov. Then almost surely we have:

$$\frac{1}{t}\int_0^t 1(X_s = x)\mathrm{d}s \to \frac{1}{q_x m_x} \qquad t\ to \infty$$

If X is positive recurrent, then for $f : S \to \mathbb{R}$ a bounded function:

$$\frac{1}{t}\int_0^t f(X_s)\mathrm{d}s \to \sum_x f(x)\pi(x)$$

Where π is the unique invariant distribution.

Note. This is saying the long term average percentage of time spent in x is the number in the invariant distribution.

3 Queueing Theory

3.1 Introduction

We have people coming in. We then have k servers serving the people (and after that they leave). This, obviously is called a *queue*. We characterize queues in the following way:

$$\underbrace{M}_{\text{The method customers arrive in queue}} / \underbrace{G}_{\text{The method customers are being served}} / \underbrace{k}_{\text{Number of Servers}}$$

where M means Markovian, G means general. Typically $k = 1$ or ∞.

3.2 $M/M/1$ Queue and $M/M/\infty$ Queue

Definition. A $M/M/1$ queue is one that customers arrive in a poisson process of rate $\lambda > 0$, and the single server service customer with service time exponential with parameter μ. Equivalently, $q_{i,i+1} = \lambda$, $q_{i,i-1} = \mu$.

Note. Note this jump chain is just a biased random walk with reflection at 0, and we know when is that transient/recurrent. This gives all the results below.

Theorem. Let $\rho = \frac{\lambda}{\mu}$. Then X, a $M/M/1$ queue is transient iff $\rho > 1$, recurrent iff $\rho \leq 1$, and positive recurrent iff $\rho < 1$. In the last case X has equilibrium distribution $\pi(n) = (1-\rho)\rho^n$. The waiting time $W \sim \exp(\lambda - \mu)$ at equilibrium.

Definition. A $M/M/\infty$ queue is same as a $M/M/1$ queue but with infinitely many servers, so everyone gets served immediately.

Theorem. The queue length X_t is positive recurrent for all $\lambda, \mu > 0$. Furthermore the invariant distribution is Poisson with rate λ/μ.

Proof. The quantitative results follows immediately after solving the detailed balance equation.

The problem is to prove that X is not explosive. Here the rates do go to infinity, so the summing of q_i trick jut won't work. But we have another one from IB Markov Chains. Define:

$$\gamma_i = \frac{p_{i,i-1}p_{i-1,i-2}, \cdots p_{1,0}}{p_{i,i+1} \cdots, p_{1,2}}$$

Then $\sum_i \gamma_i = \infty$ implies that the chain is recurrent. Here pick k big enough so that $p_{i,i+1} \leq \frac{1}{3}$ for all $i \geq k$. This is possible as $p_{i,i+1} = \frac{\lambda}{i\mu + \lambda}$. Then:

$$\sum \gamma_i \geq A \sum_{i \geq k} 2^{i-k+2} = \infty$$

As $p_{i,i-1} \geq 2p_{i,i+1}$ for $i \geq k$. So the jump chain is recurrent, so X is not explosive. □

3.3 Burke's Theorem

Note. This is important. *Bigly.*

Theorem. Let D_t be the number of customers who have departed the queue up to time t in a $M/M/1$ queue. At equilibrium, D_t is a poisson process with rate λ, independent of μ as long as $\mu > \lambda$. Furthermore, X_t, the number of people in the queue at time t, is independnet from $D_s, s \leq t$.

Proof. Recall that the invariant distribution of a birth and death chain satisfies the detailed balance equation. Thus it is time-reversible at equilibrium. let $\hat{X}$ be the reversed process. Note that $\hat{X}$ +1 exactly when people leave the queue, so departures become arrivals in $\hat{X}$, and since X and $\hat{X}$ have the same distribution, arrivals in $\hat{X}$ is $Poi(\lambda)$, so departures in X is $Poi(\lambda)$.

For the last part, note that X_0 is independent from arrivals between 0 and T trivially. Reversing time shows that X_T is independent from departures between 0 and T. □

3.4 Queues in Tandem and Jackson Networks

Let's now consider a joint $M/M/1$ queue: people enter queue 2 after leaving the first $M/M/1$, and each queue has service rates μ_1, μ_2. We have the following result:

Theorem. Let X_t, Y_t denote queue lengths in first and second queue respectively. The process $(X, Y)$is positive recurrent iff $\lambda < \min \mu_1, \mu_2$. And we have:

$$\pi(m, n) = (1 - \rho_1)\rho_1^m(1 - \rho_2)\rho_2^n$$

where $\rho_1 = \frac{\lambda}{\mu_1}$ and $\rho_2 = \frac{\lambda}{\mu_2}$. Thus X, Y are independent at equilibrium.

Proof. We use Burke! I told you it is important. now we know for the first queue, at equilibeium, the distribution is $\pi(m) = (1 - \rho_1)\rho_1^m$ and the departure process into Queue 2 (arrival in Queue 2) is a Poisson process with rate λ. Thus, when the first queue is in equilibrium, the second queue is a $M/M/1$ queue! And there equilibrium occurs when $\pi(n) = (1-\rho_2)\rho_2^n$. Also by Burke, Y_t, the number of people in the queue only depends on Y_0 and the departure process from X by Burke, but by Burke *again* this departure process from X is independent from X_t. So X and Y are independent and irreducible, so the equilibrium distribution is just the product. The λ inequality also follows trivially from Burke. □

Now we make this concept go even further:

Definition. Let us have an irreducible network of N single-server queues. The arrival rate is λ_i, $1 \leq i \leq N$, and service rate is μ_i. When the service is ompleted, each customer can move to queue j with probability p_{ij} or exit the system with probability p_{i0}. We assume p_{i0} is positive for all i and $p_{ii} = 0$. This is a*Jackson Network.*

What if I tell you this has an analytic solution at equilibrium?

Definition. A vector $(\bar{\lambda}_1, \cdots, \bar{\lambda}_n)$ satisfies the *traffic equations* if for all $i \leq N$ we have:

$$\bar{\lambda}_i = \lambda_i + \sum_{j \neq i} \bar{\lambda}_j p_{ji}$$

If you think about it, the equilibrium rates going in and out of queue i must be $\bar{\lambda}_i$. Now the theorem:

Theorem (Jackson's Theorem)**.** There exists a unique solution to the traffic equations. Furthermore, if $\bar{\lambda}_i < \mu_i$ for all i, then the Jackson network is positive recurrent, with:

$$\pi(n) = \prod_{i=1}^{N} (1 - \bar{\rho}_i)\bar{\rho}_i^{n_i}$$

An invariant distribution, where $\bar{\rho}_i = \frac{\bar{\lambda}_i}{\mu_i}$. At equilibrium, departure processes to outside from each queue is independently poisson with rates $\bar{\lambda}_i p_{i0}$.

To prove this,we need a lemma.

Lemma (Partial Balance Equations)**.** Suppose that X_t is a Markov chain on some state space S, and $\pi(x) \geq 0$ for all $x \in S$. Assume for each $x \in S$ we can find partition of $S\{x\}$, (S_i^x), such that for all i:

$$\sum_{y \in S_i^x} \pi(x)q(x, y) = \sum_{y \in S_i^x} \pi(y)q(y, x)$$

This is literally a weaker version of the detailed balance equations, and the proof is done by taking the πQ sum over each S_i^y, and interchanging x and y in there to give that $\pi Q = 0$.

Proof. Let $\pi(n) = \prod \bar{\rho}_i^{n_i}$. Define $\tilde{q}(n.m) = \frac{\pi(m)}{\pi(n)} q(m, n)$. now let $A = \{e_i; 1 \leq i \leq N\}$fiff Then if $n \in S$ is a state and $m \in A$, then $n + m$ indicates a possible state after an arrival of a customer in the system at some queue. Similarly we define $D_j = \{e_i - e_j, i \neq j\} \cup \{-e_j\}$ (departure from queue j) and these sets cover everything. We first show that:

$$\sum_{m \in D_j} q(n, n + m) = \sum_{m \in D_j} \tilde{q}(n, n + m)$$

The left side is just $\mu_j p_{j0} + \sum \mu_j p_{ji} = \mu_j$. Now expanding the definitions give:

$$\tilde{q}(n, n + e_i - e_j) = \frac{\bar{\lambda}_i p_{ij}}{\bar{\rho}_j} \qquad \tilde{q}(n, n - e_j) = \frac{\lambda j}{\bar{\rho}_j}$$

So the right side gives:

$$\frac{\lambda j}{\bar{\rho}_j} + \sum_{i \neq j} \frac{\bar{\lambda}_i p_{ij}}{\bar{\rho}_j} = \frac{\bar{\lambda}_j}{\bar{\rho}_j} = \mu_j$$

By the traffic equations. Now we show:

$$\sum_{m \in A} q(n, n + m) = \sum_{m \in A} \tilde{q}(n, n + m)$$

The left side is just $\sum_i \lambda_i$, and we can expand definition to see that $\tilde{q}(n, n + e_i) = \bar{\lambda}_i p_{i0}$. Then we have the right side as:

$$\sum_i \bar{\lambda}_i p_{i0} = \sum_i \bar{\lambda}_i - \sum_j \sum_i \bar{\lambda}_i p_{ij} = \sum_i \bar{\lambda}_i - \sum_j (\bar{\lambda}_j - \lambda_j) = \sum_j \lambda_j$$

Where the second last equality follows from the traffic equation.

So we proved that π is invariant! Since rates are bounded, there is no explosion, so it is positive recurrent (provided that $\bar{\rho}_i < 1$ so that it is a distribution).

The independence of departures follow trivially from Burke, *again.* □

3.5 Non-Markov $M/G/1$ Queue

Assume we have a $Poi(\lambda)$ arrival but only ϵ_i iid service times. What can we do?

Proposition. The process $X(D_n)$ is a Markov chain with transition probability $p_{i,i+k-1} = p_{0k} = E[e^{-\lambda\epsilon}(\lambda\epsilon)^k/k!]$ for all $i > 0$, $k \geq 0$, and 0 otherwise.

Proof. If we look at the process between departures, there are only arrivals. The probability that k customers arrive (and then one leave) is:

$$P(A_n = k) = E[P(A_n = k|\epsilon_n)] = E[e^{-\lambda\epsilon}(\lambda\epsilon)^k/k!]$$

Done. The 0 case accounts for the fact that no one can be leaving if no one is there. □

We introduce a lemma:

Lemma. Let ϵ_i be integer valued random variables, and let $S_n = \sum_{i=1}^n \epsilon_i$ by the corresponding random walk. Then S is recurrent iff $E[\epsilon_1] = 0$.

We then introduce the following theorem: Now write $E[\epsilon] = \frac{1}{\mu}$ and denote $\rho = \frac{\lambda}{\mu}$. Then $E[A_n] = \rho$. We have the followign theorem:

Theorem. $\rho \leq 1 \Leftrightarrow$ the queue $X(D_n)$ is recurrent.

Proof.

- We see that while $X(D_n) > 0$, it is a random walk on Z with step distribution $A_n - 1$. So it is transient iff $E[A_n - 1] > 0$ as if it is < 0 it always hits back to 0, and this gives $\rho > 1$, so we result follows.
- Now call a customer C_2 an offspring of C_2 if C_2 arrives during service of C_1. Then the number of offsprings of each customer is A_n and the family tree is a branching process. The queue empties out is equivalent to saying the branching process is finite, whish from Markov IB we know is equivalent to saying $E[A_n] \leq 1$. Done.

□

4 Renewal Theory

4.1 Introduction and Elementary Results

Definition. Let (ϵ_i) be iid non-negative random variables with $P(\epsilon > 0) > 0$. Then Set $T_n = \sum_{i=1}^n \epsilon_i$ and $N_t = \max\{n \geq 0 : T_n \leq t\}$.

Intuitively, this counts how many ϵ variables have ended at time t. We have the result:

Theorem. If $\frac{1}{\lambda} = E[\epsilon] < \infty$ then we have as $t \to \infty$:

$$\frac{N_t}{t} \to \lambda \text{ almost surely} \qquad \frac{E[N_t]}{t} \to \lambda$$

Proof. This is easy. $T_{N(t)} \leq t \leq T_{N(t)+1}$, and $\frac{T_{N(t)}}{N(t)}$ converges to $\frac{1}{\lambda}$ by law of large numbers. Similarly with $\frac{T_{N(t)+1}}{N(t)}$, so by sandwich theorem:

$$\frac{t}{N(t)} \to \frac{1}{\lambda}$$

Almost surely. The second assertion is in fact harder than it looks so we will skip. □

We then introduce an important concept:

Definition. Let X be a nonnegative random variable with law $|mu$, and suppose $E[X] = m < \infty$. Then the *size-biased distirbution* is given by:

$$\hat{\mu}(dy) = \frac{y}{m}\mu(dy)$$

The use of this is shown in the next section.

4.2 Equilibrium Theory of Renewal Processes

Definition. Define $A(t) = t - T_{N(t)}$, the time since the last renewal, and $E(t) = T_{N(t)+1} - t$ the time until the next renewal. Finally ley $L(t)$ be the length of the current renewal.

Definition. A random variable ϵ is *arithmetic* if $P(\epsilon \in kZ) = 1$ for some k maximal with this property.

Then, we have our grand theorem:

Theorem. Suppose ϵ is non-arithmetic, and let $E[\epsilon] = \frac{1}{\lambda}$. Then:

$$L(t) \to \hat{\epsilon}$$

in distribution as $t \to \infty$. Moreover:

$$P(E(t) \leq y) \to \lambda \int_0^y P(\epsilon > x)\mathrm{d}x \qquad (L(t), E(t)) \to (\hat{\epsilon}, U\hat{\epsilon})$$

where U is uniform on $(0, 1)$ and independent from $\hat{\epsilon}$.

Discrete Case. Note we only need to prove the "Moreover" part.

Let ϵ be a discrete random variable taking values in $\{1, 2, \cdots\}$. Then $E(t)$ is a discrete time Markov chain with $p_{i,i-1} = 1$ fir $i \geq 1$, $p_{0n} = p(\epsilon = n + 1)$. It is clearly irreducible and recurrent, with invariant measure:

$$\pi_n = \sum_{m \geq n+1} P(\epsilon = m)$$

If $E[\epsilon] < \infty$, then this is normalizable to $\pi_n = \lambda P(\epsilon > n)$.Then we see that the non-arithmetic condition guarantees that the Markov chain is aperiodic, so by convergence to equilibrium we proved the $E(t)$ limit.

The second limit follows by considering $(L(t), E(t))$ as a Markov chain in $\mathbb{N} \times \mathbb{N}$, and its transition probabilities give a equilibrium measure of:

$$\pi(n,k) = nP(\epsilon = n) \times \frac{1}{n}\mathbf{1}_{\{0 \le k \le n-1\}}$$

After normalization, the first factor becomes $P(\hat{\epsilon} = n)$, and the second factor gives the uniform distribution. □

Remark. Now we understand the significance of the the size-biased distribution: When t is large, we are more likely to fall into larger intervals of ϵ, and the size-biased distribution corrects for that.

4.3 Renewal-Reward Processes

Now we extend this a bit further. At the end of each ϵ_i, we associate a reward R_i to the process, and let $R_t = \sum_{i=1}^{N_t} R_i$ denote the total reward collected up to time t. We have the following theorem extending the results above:

Theorem. If $E[|R|] < \infty$, then as $t \to \infty$:

$$\frac{R_t}{t} \to \lambda E[R] \quad \frac{E[R_t]}{t} \to \lambda E[R] \quad E[R_{N(t)+1}] \to \lambda E[R\epsilon]$$

This is very useful.

Example. SUppose a machine is on for a time X_i, then off for time Y_i. Then denote $\epsilon_i = X_i + Y_i$ as a full cycle length. What is the fraction of time the machine is on in the long run?

If we denote the amount fo time on as rewarded to the process after each cycle ends, then we see that:

$$\frac{R_t}{t} \to \frac{E[X]}{E[X] + E[Y]}$$

Which is intuitive. Now note that if we consider the probability $p(t)$ the meachine is on at time t, no size-biasing effect needs to be considered, so we have $p(t) \to \frac{E[X]}{E[X]+E[Y]}$.

4.4 Little's Formula

This is the last important formula we learn for renewals.

Definition. A process (X_t) is called regenerative is there exists random times τ_n such that the process regenerates, in the sense that the law of the process $(X_{t+\tau_n})_{t\ge 0}$ is identical to $(X_t)_{t\ge 0}$ and independent of $(X_t)_{t\le \tau_n}$.

Theorem (Little's Formula)**.** Let queue X be regenerative with regeneration times τ_n, N the arrival process, and W_i the waiting time of the ith customer. Then if $E[\tau_1] < \infty$

and $E[\sum_{i=1}^{n_{\tau_1}} W_i] < \infty$, then almost surely the long-run mean queue size L, average waiting time W, and average arrival rate λ exist and are deterministic:

$$L = \lim_{t\to\infty} \frac{1}{t}\int_0^t X_s \mathrm{d}s \qquad W = \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} W_i \qquad \lambda = \lim_{t\to\infty} \frac{n_t}{t}$$

Moreover, $L = \lambda W$.

Remark. The useful part of the theorem is the last sentence.

Proof. Define $Y_n = \sum_{i=1}^{N_{\tau_n}} W_i$. Then for $\tau_n \le t < \tau_{n+1}$, we have:

$$\frac{Y_n}{\tau_{n+1}} \le \frac{1}{t}\int_0^t X_s \mathrm{d}s \le \frac{Y_{n+1}}{\tau_n}$$

As the left hand-size is the lower limit of the average queue size, similarly with the right side. Note that:

$$Y_n = \sum (Y_i - Y_{i-1})$$

where $Y_i - Y_{i-1}$ is iid, and $E[Y_1] < \infty$, so by law of large numbers, both the left hand side and the right hand side converges to $\frac{E[Y_1]}{E[\tau_1]}$. This can be similarly donw for $\frac{N_t}{t}$. Now we also have:

$$\frac{Y_{\tau_n}}{N_{\tau_{n+1}}} \le \frac{1}{k}\sum_{i=1}^{k} W_i \le \frac{Y_{\tau_{n+1}}}{N_{\tau_n}}$$

Using the proved results above, and taking $n \to \infty$, we have that $W = \frac{L}{\lambda}$. □

5 Population Genetics

5.1 What is genetics?

Well, we are really talking about DNA. Here is a quick Biology terminology 101:

Definition. The genome is the collextion of all genetic information of an individual. This information is stored on a number of chromosomes, each consisting of many genes. A *gene* is a piece of coding for one specific protein, and is formed from 4 base pairs: ATCG. An *allele* is one of a number of alternative forms of the same gene, and the location of a base is called a *site*.

5.2 Moran Model and Fixation

Definition. The *Moran Model* has a constant population size. With rate 1, every individual dies, and when they die, an individual chosen uniformly (including the person dying) gives birth to a new individual identical to the person giving birth

I know, this is a shit model.

Definition. Suppose at time $t = 0$ a number of $X_0 = i$ individuals carry a specific allele a, while $N - i$ carry A. Then let X_t be the number of individuals carrying allele a at time $t \ge 0$. Define $\tau = \int\{t \ge 9 : X_t = 0 \text{ or } n\}$. We say a *fixates* if $X_\tau = N$, and there is no fixation if $X_\tau = 0$, provided $\tau < \infty$.

Theorem. $\tau < \infty$ almost surely, and $P(X_\tau = N | X_0 = i) = \frac{i}{N}$. Moreover:

$$E[\tau | X_0 = i] = \sum_{j=1}^{i-1} \frac{N-i}{N-j} + \sum_{j=i}^{N-1} \frac{i}{j}$$

Proof. Note that X_t is a markov chain with $q_{i,i+1} = \frac{(N-i)i}{N} = q_{i,i-1}$. The jump chain is thus just a simple random walk absorbed at 0 and N. Thus the result of the probability being $\frac{i}{N}$ follows from IB Markov chains.

now write τ_j as the total time spent at j. Note that $E_j[\tau_j] = \frac{1}{q_j} E_j[\text{\# visits to j}]$. And the expected number of visits is just the inverse of the probability of not returning to j (Geometric distribution!),

To not return at j, it needs to go to $j-1$ and be absorbed at 0, which has probability $\frac{1}{j}$ by using the proved result above. Or it goes to $j+1$ and gets absorbed at N with probability $\frac{1}{N-j}$. So

$$E_j[\tau_j] = \frac{1}{q_j} \frac{1}{\frac{1}{2}\frac{1}{j} + \frac{1}{2}\frac{1}{N-j}} = 1$$

Then:

$$E_i[\tau] = \sum_j E_i[\tau_j] = \sum_i P_i(X_t = j \text{ for some } t \geq 0) \times E_j[\tau_j] = \sum_{j \geq i} \frac{i}{j} + \sum_{j<i} \frac{N-i}{N-j}$$

Where the last equality follows from Simple Random Walk results in IB. □

5.3 Infinite Sites Models of Mutation

Now on top of this, we consider that each individual has mutations occurring at rate u. We assume there are infinite sites so no two mutations affect the same site (so everyone with the same mutation descended from one individual). Denote $M_j(n)$ as the number of sites where exactly j individuals carry a base at a certain site which differs from everyone else in a sample of n individuals.

Theorem. Let $\theta = uN$ and N be the population size. Then:

$$E[M_j(N)] = \frac{\theta}{j}$$

where we assume mutation of population started from $t = -\infty$, and the expectation is taken at $t = 0$.

Proof. We look at each mutation. Suppose it happened at time $-t$. Let X_s denote the number of indivuals carrying this mutation at time $-t+s$. We know from above that this is just a simple Markov chain, and the probability that we have j individuals at time 0 is just $p_t(1, j)$. Now muations occur at rate of uN, so:

$$E[M_j(N)] \int_0^\infty uN \mathrm{d}t p_t(1,j) = \theta E_1[\tau_j] = \frac{\theta}{j}$$

□

5.4 Kingman's n-coalescent and infinite coalescent

Now we consider another viewpoint. Every individual has a unique parent, and we consider partition $\prod_t$ of the sample such that i and j are in the same set iff i and j have the same ancestor at time $-t$. We have the following theorem:

Theorem. $(\prod_{Nt/2})$ is a Markov chain on P_n with:

$$q_{\pi,\pi'} = \begin{cases} 1 & \text{if } \pi\text{' can be obtained from } \pi \text{ by coagulating two of its blocks} \\ -\binom{k}{2} & \text{if } \pi' = \pi \text{ and } \pi \text{ has } k \text{ blocks} \\ 0 & \text{else} \end{cases}$$

This is *kingman's n-coalescent.*

Proof. Everything here is quite intuitive, except for the factor $\frac{N}{2}$. Now each block of $\prod_t$ is associated to an ancestor at time $-t$. Consider two blocks which have ancestors i and j. The rate at which i and j coalesces is the sum of the rates that i combines into j's family tree and j combines into i's family tree.

Equivalently we can think of the Moran model as biinfinite Poisson processes $N_t^{i,j}$ with rates $\frac{1}{N}$ where a jump at $N_t^{i,j}$ indicates j has been killed off and is replaced by offspring of i. Now thus these two coalesce at the rate of $\frac{2}{N}$, because either (going in positive time) we had the old j die and offspring of i replacing j (with rate $\frac{1}{N}$) or we had the old i die and offspring of j replacing j (with same rate). Thus the total rate is $\frac{2}{N}$. The rest follows immediately. □

Now we have an interesting property, which is that if we have Kingman's n coalescent, then a subsample of size $n-1$ has the law of kingman's $n-1$ coalescent. This proof is on the 4th example sheet.

Now with this property, we can deduce a process $\Pi_t t \geq 0$ taking values in partitions P of $\mathbb{N}$ such that for every $n \geq 1$, the process Π restricted to a n-set has the law of Kingman's n-coalescent. This is called *Kingman's infinite coalescent.*

Theorem. With probability 1, the number of blocks of Π_t is finite at any time $t > 0$ for Kingman's coalescent.

Proof. Write $|\Pi_t|$ for the number of blocks. Since the events $\{|\Pi_t^n \geq M\}$ where Π_t^n means Π limited to $\{1, \cdots, n\}$ are increasing in n. So:

$$P(|\Pi_t| \geq M) = \lim_{n\to\infty} P(|\Pi_t^n| \geq M) = \lim_{n\to\infty} P\left(\sum_{j=M+1}^{n} \tau_j \geq t\right)$$

where τ_j is the time for Π^n to drop from j to $j-1$ blocks. And we know from the n coalescent that this is exponential with rate $\binom{j}{2}$. Then by Markov's inequality:

$$P(|\Pi_t| \geq M) \leq \frac{1}{t} E\left[\sum_{j=M+1}^{\infty} \tau_j\right] \leq \frac{1}{t} \sum_{j=M+1}^{\infty} \frac{1}{\binom{j}{2}} = \frac{2}{t} \sum_{j=M+1}^{\infty} \frac{1}{j(j-1)}$$

This tends to 0 as $M \to \infty$, so the result follows. □

5.5 Infinite Alleles Model and Ewens Sampling Formula

This model is exactly the same as the infinite site model, but with the only difference is that every time it mutates, it mutates to a *completely new allele* and we partition into blocks with the same allelic type. And it turns out we can describe the distribution of Π_n explicitly, interms of A_j, the number of distinct alleles which are carried by exactly j individuals.

Theorem (Ewens Sampling Formula)**.** Let a_j be such that $\sum ja_j = n$. Then:

$$P(A_1 = a_1, \cdots, A_n = a_n) = \frac{n!}{\theta(\theta+1)\cdots(\theta+n-1)} \prod_{j=1}^{n} \frac{(\theta/j)^{a_j}}{a_j!}$$

To prove this, we introduce *Hoppe's urn*:

Definition. *Hoppe's Urn* is an urn with balls of different colors of mass 1 and a single black ball of mass θ. At each step, we draw from the urn with probability proportional to the mass of the ball. If it is colored, we put back the ball along with a ball of the same color. If it is black we put it back along with a ball of new color.

Proof. Now $(A_1, \cdots, A_n)$ depend only on the mutations which are in the genealogical tree of the sample. So we assume the tree is given by Kingman's n-coalescent and mutations fall at rate $\frac{\theta}{2}$ per unit length on each branch.

- We see mutations as killings, where a new color is eliminated in the branch. Let T_i denote successive times at which number of branches drops from i to $i-1$, either by coalescence or killing by mutation.

  Between T_m and T_{m+1} there are m branches. The $m+1$ branch can be added by a mutation (which we have $m+1$ exponential clocks with rate $\frac{\theta}{2}$), or coalescence (which we have $m(m+1)/2$ clocks with rate 1). Thus the probability of a new allelic group is:

  $$P(\text{new block}) = \frac{\theta}{\theta+m}$$

  Probability of joining existing group of size n_i is $\frac{n_i}{m+\theta}$.

- Then we observe that this is Hoppe's urn as the mutation event is drawing the black ball, and event of joining of a group of size n_i is identical to drawing a ball from that color. We inductively prove that:

  $$P(\Pi_n = \pi) = \frac{\theta^k}{\theta\cdots(\theta+n-1)} \prod (n_i - 1)!$$

  Consider the induction step on n, and let π be of size n. Let π' be π limited to $\{1, \cdots, n-1\}$. Then either n is a singleton in π or (b) n is in a block of size n_j in π. These two cases correspond to mutation and coalescence events above, and it is easy to see that the formula does hold.

□

5.6 Chinese Restaurant Process

Consider there are infinite number of tables in a restaurant and they are numbered $1, 2, \cdots$. Each table has infinite capacity. Customers arrive intro the restaurant and customer 1 occupies table 1. Cusomter $n+1$ chooses with equal probbaility either to sit to the left of a customer already sitting in a table or to start a new table.

So this is the process above with $\theta = 1$. now define $\sigma_n[n] \to [n]$ as followed: if customer i is sitting to left of customer j, then let $\sigma_n(i) = j$ and if i is sitting by himself, then $\sigma_n(i) = i$.

From induction we can see that σ_i is a random permutation. (And any subset formed by deleting one element is a random $n-1$ permutation). We can use this along with the Ewens sampling formula to get properties of random permuations.

Corollary. Let K_n be the number of occupied tables when n customers have arriced. This satisfies:

$$\frac{K_n - \log n}{\sqrt{\log n}} \to N(0, 1)$$

As $N \to \infty$.

Proof. To prove this, we see that $K_n = Z_1 + \cdots + Z_n$ where Z_i is the indicator that ith customer occupies a new table. But these are independent, so $E[K_n] = \sum \frac{1}{j} \sim \log n$. So we have the formula as required. □