Contents

1 **1**

2 **1**
2.1 . . . 1
2.2 . . . 1
2.3 . . . 1
2.4 Gambler's Ruin . . . 2
2.5 Simple Branching Processes . . . 2

3 Long time behavior for Markov Chains **4**
3.1 Classification of States . . . 4
3.2 Classification of Chains . . . 5
3.3 Stationary distributions and the limit theorem . . . 7
3.4 Limit Theorems . . . 10
3.5 Reversibility . . . 11
3.6 Chains with finitely many states . . . 12
3.7 Branching Processes . . . 12
3.8 Review (in notes) . . . 13

4 Continuous Time Markov Chains **13**
4.1 The Poisson Process . . . 13
4.2 Birth Processes . . . 13
4.3 More on Poisson Processes . . . 15
4.4 The Death Process . . . 16
4.5 Birth-Death Processes . . . 17
4.6 Limiting Behavior of Birth-Death Processes . . . 18

5 Markov Chain Monte Carlo **19**
5.1 Background . . . 19
5.2 The Hastings-Metropolis Algorithm . . . 19
5.3 The Gibbs Sampler . . . 20
5.4 Simulated Annealing . . . 21

1

2

2.1

2.2

2.3

X_n: S = 0, 1, ..., r-1, r, ..., N (0 - r-1 transient, r - N absorbing).

Recall from last time: g(i) = "rate" to transient state i. Let $W_i = E(\Sigma_{n=0}^{T-1} g(X_n)|X_0 = i)$. g(i)=1 $\Rightarrow \Sigma_{n=0}^{T-1} g(X_n) = T$.

ex 2.3.4 - if g(i) = 1 if i=k, 0 i not = k, k = transient state, gives $W_i = W_{ik}$, mean number of visits to state k before being absorbed.

Note: $\Sigma_{n=0}^{T-1} g(X_n)$ always includes $g(X_0) = g(i)$.

If a transition is made from state i to a transient state j, the sum will include some future terms. The Markov property implies the future sum proceeding from state j has value W_j.

Using the total law of probability, (2.3.3) $W_i = g(i) + \Sigma_{j=0}^{r-1} P_{ij} W_j, i = 0, 1, 2, ..., r-1$.

exercise: explain this.

Example 2.3.5 - In ex 2.3.3, g(i) = 1 for all i. This implies that $\nu_i = E(T|X_0 = i) = W_i$, and (2.3.4) $\nu_i = 1 + \Sigma_{j=0}^{r-1} P_{ij} V_j, i = 0, 1, ..., r-1$.

Ex 2.3.6 - in ex 2.3.4, (2.3.5) $W_{ik} = \delta_{ik} + \Sigma_{j=0}^{r-1} P_{ij} W_{jk}, i = 0, 1, ..., r-1$.

Example 2.3.7 - A Markov chain has P= (1, 0, 0, 0 & .1, .4, .4, .1 & .2, .1, .6, .1 & 0, 0, 0, 1), S=0, 1, 2, 3. Start in state 1. Determine probability of being absorbed into state 0, and mean time until absorption. 0, 3 absorbing states. 1, 2 transient. With U_{i0} = P(absorption into 0 $|X_0 = i$). (2.3.2) gives: $U_{10} = P_{10} + P_{11}U_{10} + P_{12}U_{20}.U_{20} = P_{20} + P_{21}U_{10} + P_{22}U_{20}.U_{10} = .1 + .4U_{10} + .1U_{20}.U_{20} = .2 + .1U_{10} + .6U_{20}. \Rightarrow U_{10} = 6/23, U_{20} = 13/23.(2.3.4)\nu_1 = 1 + .4\nu_1 + .1\nu_2.\nu_2 = 1 + .1\nu_1 + .6\nu_2. \Rightarrow \nu_1 = 50/23, \nu_2 = 70/23$.

2.4 Gambler's Ruin

ex 2.1.2, 2.1.9 introduced random walks. X_n = location. Def 2.4.1 - if $r_i = 0, p_i = p, q_i = q(q = 1-p)$, this is a simple random walk.

ex 2.4.1 - gambler's ruin - We have a game for two people A, B. Total fortune of A and B is \$N. At each step i, A has probability p_i of winning \$1, q_i of losing \$1, and r_i of drawing. $0 < p_i, q_i < 1, 0 \leq r_i < 1, p_i + q_i + r_i = 1$. If A's fortune reaches 0 or N, game stops. Let X_n = fortune of A at time n. X_n is a Markov chain. P = (1, 0, 0, ... & q_1, r_1, p_1, 0, ... & 0, q_2, r_2, p_2, 0, ... & ... & 0, ..., $q_{N-1}, r_{N-1}, p_{N-1}$ & 0, ..., 0, 1). States k=0, N are absorbing.

We address some intermediate time questions using (2.3.2), (2.3.4).

ex 2.4.3 - the probability of ruin for player A starting with \$i in ex 2.4.1 is $U_i = U_{i0}$ in (2.3.2), (2.4.1) $U_i = P_i U_{i+1} + r_i U_i + q_i U_{i-1}, i = 1, 2, ..., N-1$, with boundary conditions $U_0 = 1, U_N = 0$. We can find an explicit formula for the solution in some cases.

ex 2.4.4 - assume $r_i = 0, p_i = p, q_i = q, q = 1-p$. (2.4.1) becomes (2.4.2) $U_i = pU_{i+1} + qU_{i-1} for 1 \leq i \leq N-1, U_0 = 1, U_N = 0$. We look for a solution in the form $U_i = \Theta^i$. $\Theta^i = p\Theta^{i+1} + q\Theta^{i-1}, \Theta \neq 0 \Rightarrow \Theta = p\Theta^2 + q$. This has roots $\Theta_1 = 1, \Theta_2 = q/p$. If $p \neq 1/2, \Theta_1 \neq \Theta_2$. The general solution is a linear combination $U_i = A_1\Theta_1^i + A_2\Theta_2^i$. A_1, A_2 are constants. Using the boundary conditions, $U_0 = 1 = A_1 + A_2, U_N = 0 = A_1 + A_2 \frac{p}{q}^N \Rightarrow$ (2.3.4) $U_i = \frac{(q/p)^i - (q/p)^N}{1-(p/q)^N}, p \neq 1/2, 0 < i < N$. If $p = 1/2, \Theta_1 = \Theta_2 = 1$, (2.4.4) $U_i = 1 - i/N, 0 < i < N$. For mean time to absorption, (2.3.4) becomes (2.4.5) $\nu_i = 1 + p\nu_{i+1} + q\nu_{i-1}, \nu_0 = \nu_N = 0 \Rightarrow$ (2.3.6) $\nu_i = \frac{1}{q-p}(i - N(\frac{1-(q/p)^i}{1-(q/p)^N}))when p \neq 1/2, i(N-i) when p = 1/2$.

ex 2.4.5 - Suppose in ex 2.4.1 that A has a backer that guarantees A's losses. There is no ruin when A's fortune reaches 0. We can let P = (q_0, p_0, 0, ... & q_1, r_1, p_1, ..., & ...). If $p_i = p, q_i = q$, for all i, the absorbing times satisfy (2.4.5) again, but $\nu_N = 0, \nu_0 : p\nu_0 = 1 + p\nu_1$. $q_0 = 0, r_0 + p_0 = 1$.

2.5 Simple Branching Processes

Model for evolution of a population. We start at time 0 with a progenitor. At the first time, the progenitor splits into k offspring with probability p_k, where p_k is a pmf, and then dies immediately. We assume the offspring reproduce in the same way. Process continues until extinction - when a generation produces no offspring. Let X_n = population at time n.

Def 2.5.1, X_n is a branching process.

Theorem 2.5.1, X_n is a Markov chain.

Example 2.5.1 - Neutron Chain Reaction - A nucleus is split by a chance collision with a neutron and it releases a random number of new neutrons. These may hit other nuclei and cause further fission.

Definition 2.5.2 - Branching Process - Let $\{Z_{nj} n \geq 1, j \geq 1\}$ be iid non-negative integer-value random variables with pmf p_k. Below, if a random sum has zero summands we give it the value zero. The branching process X_n is defined (2.5.1) $X_0 = 1, X_1 = Z_{1,X_0}, X_2 = Z_{21} + Z_{22} + \cdots + Z_{2X_1}, \ldots, X_n = Z_{n1} + Z_{n2} + \cdots + Z_{nX_{n-1}}$. Z_{nj} = number of members in nth generation who are offspring of the jth member of the n-1st generation. Note: if $X_n = 0$, then $X_{n+1} = 0$, so 0 is absorbing.

Theorem 2.5.2 - X_n is a Markov chain. Proof: excercise.

The branching process is defined in terms of random sums. We develop a few more useful facts.

Definition 2.5.3 - Random Sums - Let $Y_1, Y_2, \ldots$ be a sequence of iid. r.v. Let N be a non-negative, integer-valued r.v. independent of $\{Y_j\}$. Let N have pmf $p_N(n) = p(N = n), n = 0, 1, 2, \ldots$. Set $X = 0 for N = 0, Y_1 + \cdots + Y_N for N > 0$. X is a random sum.

Theorem 2.5.3 - Assume that $\{Y_i\}$ and N have finite moments. $E(Y_i) = \mu, Var(Y_i) = \sigma^2, E(N) = \nu, Var(N) = \tau^2$. Then (2.5.2) $E(X) = \mu\nu$. (2.5.3) $Var(X) = \nu\sigma^2 + \mu^2\tau^2$. Proof: $E(X) = \Sigma_{n=0}^{\infty} E(X|N = n)p_N(n) = \Sigma_{n=0}^{\infty} E(Y_1 + \cdots + Y_N|N = n)p_N(n) = \Sigma_{n=0}^{\infty} E(Y_1 + \cdots + Y_n)p_N(n) = \mu\Sigma_{n=0}^{\infty} np_N(n)$. (2.5.3) is excercise.

Little notation: if Y is a random variable with pmf $\{p_k\}$, we define $\mu = E(Y), \sigma^2 = Var(Y)$.

Theorem 2.5.4 - Let X_n be a branching process with pmf $\{P_k\}$ and assume μ, σ^2 are finitie. Let $M_{(n)}, V_{(n)}$ be the mean and variance of X_n conditioned on $X_0 = 1$. (2.5.4) $M(n) = \mu^n$. (2.5.5) $V(n) = \sigma^2\mu^{n-1} * \{n for \mu = 1, \frac{1-\mu^n}{1-\mu} for \mu \neq 1$. Proof: using (2.5.2) in (2.5.1) gives (2.5.6) $\{M(n+1) = \mu M(n), V(n+1) = \sigma^2 M(n) + \mu^2 V(n). X_0 = 1, M(0) = 1, V(0) = 0$. Now iterate. $M(n+1) = \mu M(n) = \mu\mu M(n-1) = \mu^3 M(n-2) = \ldots$. This gives the result.

The mean population increases geometrically if $\mu > 1$, decreases geometrically if $\mu < 1$, and is fixed if $\mu = 1$. The variance increases or decreases geometrically if $\mu > 1$ or $\mu < 1$. It increases linearly if $\mu = 1$.

The probability of extinction:

Definition 2.5.4 - The random time of extinction N is the first time for which $X_N = 0$. This is an absorption time. We let (2.5.7) $U_n = P(N \leq n|X_0 = 1) = P(X_n = 0|X_0 = 1)$ be the probability of extinction at or prior to the nth generation, conditioned on $X_0 = 1$.

Theorem 2.5.5 - We have (2.5.8) $U_0 = 0, U_1 = P_0, U_n = \Sigma_{k=0}^{\infty} P_k (U_{n-1})^k, n \geq 2$. Proof: the single parent X_0 has $Z_{1,X_0} = k$ offspring. These offspring in turn have more offspring. If the original population dies out in n generations, then each of these k offspring lines die out in $n-1$ generations, or less. The k offspring lines are independent of each other and have the same statistics as the original generation. The probability that one of the k offspring lines dies out is U_{n-1}. So the probability that they all die out is $(U_{n-1})^k$. The total law of probability gives (2.5.8).

Example 2.5.3 - $P_0(1/4), P_1(1/8), P_2(1/2), P_3(1/8), P_4(0), \ldots$.

Recall §1.5. Recall Thm 1.5.3: if $Y_1, \ldots, Y_n$ are indep r.v. having prob. generating functions $P_{Y_1}, \ldots, P_{Y_n}$, then the generating function for $X = Y_1 + \cdots + Y_n$ is (2.5.9) $P_X(s) = P_{Y_1}(s) \cdots P_{Y_n}(s)$.

Thm 1.5.1: if a r.v. Y has pmf $\{P_k\}$ and prob. generating function P_Y, (2.5.10) $\frac{dP_Y(1)}{ds} = P_1 + 2P_2 + 3P_3 + \cdots = E(Y)$. (2.5.11) $Var(Y) = \frac{d^2 P_Y(s)}{ds^2}|_{s=1} + \frac{dP_Y(s)}{ds}|_{s=1} - (\frac{dP_Y(s)}{ds}|_{s=1})^2$. See § 9.2 in text.

Theorem 2.5.6 - If $Z_1, Z_2, \ldots$ is a sequence of iid r.v. with common generating function P_Z and if $N > 0$ is an integer-valued non-negative r.v. indep of $\{Z_i\}$ with prob generating function P_N, then $X = Z_1 + \cdots + Z_N$ has prob. generating function (2.5.12) $P_X(s) = P_N(P_Z(s))$. Proof: $P_X(s) = E(s^X) = E(E(s^X|N)) = \Sigma_n E(s^X|N = n)P(N = n) = \Sigma_n E(s^{Z_1 + \cdots + Z_n})P(N = n) = \Sigma_n E(S^{Z_1}) \cdots E(s^{Z_n})P(N = n) = \Sigma_n (P_Z(s))^n P(N = n) = P_N(P_Z(s))$.

Returning to a branching process X_n, we assume the offspring pmf $\{P_k\}$ has generating function $\phi(s) = E(s^{X_1}) = \Sigma_k P_k s^k$. We want the prob. generating function ϕ_n for X_n.

Theorem 2.5.7 - We have (2.5.13) $\phi_{m+n}(s) = \phi_m(\phi_n(s)) = \phi_n(\phi_m(s))$. (2.5.14) $\phi_n(s) = \phi(\phi(\phi(\ldots\phi(s)))\ldots)$ (n compositions). Proof: Every member of the $(m+n)$th generation has a unique ancestor in the mth generation. So $X_{m+n} = Z_1 + \cdots + Z_{X_m}, Z_i =$ number of members of the (m+n)th generation that descend from the ith member of the mth generation. The Z_i are iid r.v. with the same distribution as X_n, by the Markov property. By theorem 2.5.6, $\phi_{m+n}(s) = \phi_m(\phi_{Z_1}(s)), \phi_{Z_1}(s) = \phi_n(s)$.

X_n branching process. $\{P_k\}$ offspring distribution. $\phi(s) = E(s^{X_1}) = \Sigma_{k=0}^{\infty} P_k s^k$. We are interested in $\phi_n(s) =$ prob. gen. function of X_n.

Fixed point theory (aside, not in notes). Solve the problem given a function g, determine x such that g(x) = x [fixed point problem, x = fixed point].

Define f(s) = g(s)-s then f(x) = 0 iff g(x) = x. root problem connected to fixed point problems.

Fixed point iteration. Given an initial guess X_0, let $X_1 = g(X_0), X_2 = g(X_1), \ldots$. Under certain conditions (Banach, ...), $X_n \to X$.

Example 2.5.3 - Suppose $0 \leq p \leq 1$ and pmf for the offspring is $\{qp^k\}_{k \geq 0}, q = 1 - p$. The prob. gen. function is $\phi(s) = q(1 - ps)^{-1}$. $\Sigma_{k=0}^{\infty} qp^k s^k = q\Sigma_{k=0}^{\infty} (ps)^k, |ps| < 1$. Each family size is one less than a geometric variable. By induction, $\phi_n(s) = \{\frac{n-(n-1)s}{n+1-ns} for p = q = 1/2, \frac{q(p^n - q^n - ps(p^{n-1} - q^{n-1}))}{p^{n+1} - q^{n+1} - ps(p^n - q^n)} for p \neq q$.

We have {ultimate extinction} $= U_n\{X_n = 0\}$. Moreover, $A_n = \{X_n = 0\}$, satisfies $A_n < A_{n+1} < \ldots$. Thus, $A = U_{n=1}^{\infty} A_n = \lim_n A_n$. $P(A) = \lim_{n \to \infty} P(A_n)$. $P(X_n = 0) \to_{n \to \infty} P(ultimate extinction) = \{1 for p \leq q, q/p for p > q$.

Extinction occurs with probability 1 if $\mu = E(X_1) = p/q = \frac{1}{(\frac{q}{p})}$ satisfies $\mu < 1$. Makes sense: if $E(X_1) < 1$ then $X_n = 0$ at some point.

Theorem 2.5.8 - let ϕ be the prob. gen. function of the offspring. Then P(ultimate extinction) = $\lim_{n\to\infty} P(X_n = 0) = \eta$, where η is the smallest nonnegative fixed point of the equation (2.5.15) $\phi(s) = s$. (1) If $|\phi'(1)| < 1, \eta > 1$. (2) If $|\phi'(1)| > 1, \eta < 1$. (3) If $|\phi'(1)| = 1, \eta = 1$ as long as the offspring distribution has positivie variance.

Proof: Let $\eta_n = P(X_n = 0)$. Theorem 2.5.7 implies (2.5.16) $\eta_n = \phi_n(0) = \phi(\phi_{n-1}(0)) = \phi(\eta_{n-1})$ (fixed point iteration). From the discussion above, $\eta_n \uparrow \eta$. Since ϕ is continuous, take the limit in (2.5.16) gives $\eta = \phi(\eta), \eta_n = \phi(\eta_{n-1}), \lim_{n\to\infty} \eta_n = \eta = \phi(\lim_{n\to\infty} \eta_{n-1}) = \phi(\eta)$.

Suppose ψ is a nonnegative root of $s = \phi(s)$. We show $\eta \le \psi$. ϕ is nondecreasing on [0, 1]. $\phi(s) = \Sigma_{k=0}^{\infty} P_k s^k$. $\eta_1 = \phi(0) \le \phi(\psi) = \psi, \eta_2 = \phi(\eta_1) \le \phi(\psi) = \psi, \ldots, \eta \le \psi$.

To proove the rest, we use the fact that ϕ is convex. $\phi''(s) = E(X_1(X_1 - 1)s^{X_1 - 2}) \ge 0$ if $s \ge 0$. ϕ is convex and nondecreasing on [0, 1] and $\phi(1) = 1$. The two curves $Y = s$ and $Y = \phi(s)$ can have two intersection points in [0, 1], η and 1. If $|\phi'(1)| < 1, \eta = 1$. If $|\phi'(1)| > 1, \eta \ne 1$. (Consult graphs in notes.)

Example 2.5.4 - An individual has a disease. Transmits the disease to some fraction of people that are contacted. Assume: a person is infectious and has the disease for a brief period. Each infected person contacts a number of people with a Poisson distribution with mean 10, and each contact has probability p of being infected. Model the infected population X_n. For $p = .2$, what is the prob. that the disease will die out. If $p < .1$, the disease will die out (ultimate extinction). prob gen func of a Poisson process with mean μ is $e^{\mu(s-1)}$, so we solve $s = e^{\cdot 2*10(s-1)} = e^{2(s-1)} \Rightarrow s \approx .2$.

Example 2.5.5 - Suppose $p_0 = 1/5, p_1 = 1/2, p_2 = 3/10, p_i = 0, i \ge 3. \mu = 1/2 + 2 * 3/10 = 11/10 > 1. \phi(s) = 1/5 + 1/2s + 3/10s^2$. Verify $\phi'(1) = \mu$. Prob of ultimate extinction: $1/5 + t/2 + \frac{3t^2}{10} = t \Rightarrow \eta = 2/3, \eta = 1$ (2/3 was circled, 1 not circled). See ex 2.5.6 in notes.

3 Long time behavior for Markov Chains

3.1 Classification of States

Viewpoint here: think of the Markov chain as describing the motion of a particle in the state space. Does a particle return to its starting state within some time? It suffices to consider the distribution of the length of time until the particle returns the first time. Let X_n be a Markov chain with state space S.

Definition 3.1.1 - A state i is persistent or recurrent if $P(X_n = i$ for some $n \ge 1 | X_0 = i) = 1$. If $P(X_n = i$ for some $n \ge 1 | X_0 = i) < 1$, i is transient.

Recall - Long time analysis - Classification of states. Recurrent state: $P(X_n = i$ for some $n \ge 1 | X_0 = i) = 1$. Transient state: $P(X_n = i$ for some $n \ge 1 | X_0 = i) < 1$. First passage time: the smallest time it takes to go from state i to state j. We are interested in mean first passage time.

Theorem 3.1.2 - (1) j is recurrent if $\Sigma_n P_{jj}^n = \infty$. (2) j is transient if $\Sigma_n P_{jj}^n < \infty$. (3) If j is transient, then $P_{ij}^n \to 0$ as $n \to \infty$ for all i.

Example 3.1.6 - Random Walk. We consider the simple random walk in ex 2.2.2. $X_n = X_0 + \Sigma_{k=1}^n B_k$, $\{B_k\}$ are i.i.d. Bernoulli variables. $P(B_k = 1) = p, P(B_k = -1) = 1 - p = q$. Consider state j. $P_{jj}^{2n-1} = 0$ for n = 1, 2, 3, ... (2n-1 = odd numbers, n = 1, 2, 3, ...) To return in 2n steps: we must take n steps in one direction and then n in other direction. This has probability (3.1.3) $P_{jj}^{2n} = \binom{2n}{n} p^n (1-p)^n = \frac{(2n)!}{n!n!}(p(1-p))^n$. We approximate these terms by an expression valid for n large then consider the sum of the approximations. Deciding if a series converges or not is not affected if we drop a finite number of terms from the beginning of the series. We use an asympotic expression for n! valid for n large.

Stirling formula - (3.1.4) $n! \sim n^n \sqrt{n} e^{-n} \sqrt{2\pi}$ n large, which means $\lim_{n\to\infty} \frac{n!}{n^n \sqrt{n} e^{-n} \sqrt{2\pi}} = 1$. We can substitute (3.1.4) into the series $\Sigma_n P_{jj}^{2n}$ without affecting convergence/divergence. $P_{jj}^{2n} \sim \frac{(4p(1-p))^n}{\sqrt{\pi n}}$. When $p = 1/2, P_{jj}^{2n} \sim \frac{1}{\sqrt{\pi n}}, \Sigma_n P_{jj}^n = \infty$. Any state is recurrent when $p = 1/2$. If $p \ne 1/2, 4p(1-p) < 1, \Sigma_n P_{jj}^n < \infty$. Any state is transient when $p \ne 1/2$. Note: thereom 3.1.3 implies that any state is either recurrent or transient.

Theorem 3.1.4 - the number of times $N(i)$ that a Markov chain visits its starting point i satisfies $P(N(i) = \infty) = \{$1 if i is recurrent, 0 if i is transient. Proof - After any return to i, a subsequent return is guaranteed

iff $f_{ii} = 1$.

Another classification -

Definition 3.1.5 - Let $T_j = min\{n \geq i : X_n = j\}$ be the time of the first visit to state j where $T_j = \infty$ if X_n never visits j. (T_j depends on X_0.)

Theorem 3.1.5 - $P(T_i = \infty | X_0 = i) > 0$ iff i is transient. When i is transient, $E(T_i | X_0 = i) = \infty$. What about reurrent states?

Definition 3.1.6 - The mean recurrence time μ_i of a state i is: $\mu_i = E(T_i | X_0 = i) = \{\Sigma_{n=1}^{\infty} f_{ii}(n)$ for i recurrent, ∞ for i transient. μ_i may be infinite when i is recurrent.

Definition 3.1.7 - A recurrent state i is null if $\mu_i = \infty$ and positive if $\mu_i < \infty$.

Theorem 3.1.6 - A recurrent state is null iff $P_{ii}^n \to 0$ as $n \to \infty$ and if this holds, $P_{ji}^n \to 0$ for all j. Proof later.

Example 3.1.7 - Consider the genotype example 3.1.4. AA and aa are recurrent (0 = "aa"). $f_{00}(1) = 1, f_{00}(n) = 0, n > 1, \Rightarrow f_{00} = 1$. These states are positive.

Example 3.1.8 - For simple random walk, ex 3.1.6, when $p = 1/2, P_{jj}^n \approx \frac{1}{\sqrt{\pi n}} \to 0$ as $n \to \infty$. So any state in a simple random walk with $p = 1/2$ is null recurrent.

The last classification of states we discuss: recall in the simple random walk, the chain can return only with an even number of steps, 2, 4, 6, /dots all divisible by 2.

Definition 3.1.8 - The greatest common divisor of a set of integers $\{n_1, n_2, \dots\}$ written g.c.d.$(n_1, n_2, \dots)$ is the largest integer m such that m divides $n_1, n_2, \dots$ all without remainder.

Example 3.1.9 - gcd(2, 4, 6, 8) = 2. gcd(2, 3, 5) = 1.

Definition 3.1.9 - The period d(i) of state i is $d(i) = \text{gcd}\{n : P_{ii}^n > 0\}$. If d(i) = 1, i is called aperiodic. If d(i) > 1, i is called periodic.

Ex 3.1.10 - Consider the OFF/ON system in ex 3.1.5. If $0 < p < 1, 0 < q < 1$, then $P_{00}, P_{01}, P_{10}, P_{11}$ are all strictly between 0 and 1. Hence $d(i) = 1$ for $i = 0, 1$. Suppose $p = q = 1$, then $P_{00}^n > 0$ for n even, $P_{00}^n = 0$ for n odd. d(0) = 2.

Example 3.1.11 - Simple random walk is periodic with d(i) = 2 when $p = 1/2$.

Ex 3.1.12 - Consider gambler's ruin in §2.4, modified so A has \$1 initially, A has a backer that guarantees A's losses (ex 2.4.5), B is infinitely wealthy. We assume $r_1 = r_2 = \cdots = 0, p_0 = p_1 = \cdots = p, q_1 = q_2 = \cdots = q$. Exercise: P = (q p 0 ... & 0 q 0 p 0 ... & 0 0 q 0 p 0 ...). $P_{11}^1 = 0, P_{11}^2 > 0, P_{11}^3 > 0, d(i) = 1$, single gcd(2, 3) = 1.

Definition 3.1.10 - If all the states of a Markov chain are aperiodic, we call the chain aperiodic.

Definition 3.1.11 - A state is ergodic if it is recurrent, positive, and aperiodic.

Ex 3.1.13 - Consider a branching process. 0 is absorbing and once there a chain never leaves. So $P_{00}^n = 1$ for all n, and 0 is recurrent. Using the formulas for $f_{ii}, \mu_0 = 1$, 0 is positive, 0 is aperiodic, so 0 is ergodic. All other states are transient.

3.2 Classification of Chains

Definition 3.2.1 - State i communicates with state j $i \to j$ if the chain may visit j with positive probability having started in i. $i \to j \Leftrightarrow P_{ij}^m > 0$ some m. If $i \to j$ and $j \to i$, then i and j intercommunicate, $i \leftrightarrow j$.

Example 3.2.1 - In the roulette wheel in ex 2.1.6, all nonzero states intercommunicate, and 0 only communicates with itself.

Theorem 3.2.1 - If $i \neq j$ then $i \to j$ iff $f_{ij} > 0$. Proof: exercise. Note, $i \leftrightarrow i$ since $P_{ii}^0 = 1$. If we fix i, we could search for all states j that intercommunicate with it.

Def 3.2.2 - All the states that intercommunicate with a given state form a communication class. $i \leftrightarrow j, j \leftrightarrow k$, so $i \leftrightarrow k$.

Definition 3.2.3 - An equivalence relation $\sim$ on a set S is an operation on pairs of elements satisfying (1) $a \sim a, a \in S$, (2) $a \sim b \Rightarrow b \sim a, a, b \in S$, (3) $a \sim b, b \sim c \Rightarrow a \sim c$.

Theorem 3.2.2 - $\leftrightarrow$ is an equivalence relation.

Example 3.2.2 - In the roulette wheel, ex 3.2.1, there are two communication classes $\{0\}, \{1, 2, \dots, 38\}$.

Example 3.2.3 - In the genotype example, ex 2.1.7, each state forms its own class, $\{AA\}, \{aa\}, \{Aa\}$.

Example 3.2.4 - In the ON/OFF system, ex 2.2.3, with $0 < p < 1, 0 < q < 1$, there is one class: $\{ON, OFF\}$.

Theorem 3.2.3 - If $i \leftrightarrow j$, (1) i is transient iff j is transient, (2) i and j have the same period, (3) i is null recurrent iff j is null recurrent.

Proof - (1) if $i \leftrightarrow j$, there are $m, n \geq 0$ such that $\alpha = P_{ij}^m P_{ji}^n > 0$. By the Chapman-Kolmogorov equations (2.2.1) $P_{ii}^{m+r+n} \geq P_{ij}^m P_{jj}^r P_{ji}^n$ for any inteteg $r > 0$. Summing over r, $\Sigma_{r=0}^{\infty} P_{ii}^r < \infty \Rightarrow \Sigma_{r=0}^{\infty} P_{jj}^r < \infty$. The argument holds with j and i reversed. By thm 3.1.2, (1) holds. (2) excercise. (3) will be proved below.

Definition 3.2.4 - A set C of states in the state space S is closed if $P_{ij} = 0$ for all $i \in C$ and $j \notin C$. A closed set with one element is called absorbing. Once a Markov chain takes a value in a closed set, it never leaves.

Example 3.2.5 - Consider the chain with $S = \{0, 1, 2\}$. $\{0, 2\}$ forms a closed set.

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

.

Definition 3.2.5 - A set C of states in the state space S is irreducible if $i \leftrightarrow j$ for all $i, j \in C$. The communication classes of a Markov chain are irreducible.

Because of theorem 3.2.3, Definition 3.2.6 - An irreducible set C is periodic, transient, or null recurrent if all or any of the states in C have these properties.

Definition 3.2.7 - If the entire state space is irreducible, we say the Markov chain is irreducible.

Theorem 3.2.4 - In an irreducible Markov chain, either all states are transient or all states are recurrent.

Theorem 3.2.5 - Decomposition Theorem - The state space S can be partitioned uniquely as $S = T \cup C_1 \cup C_2 \cup \ldots, T = \{\text{transient states}\}, \{C_i\} =$ irreducible, closed sets of recurrent states. Lots of words in notes.

Proof: Let $\{C_j\}$ be the recurrent equivalence classes of intercommunication ($\leftrightarrow$). We only need to show that each C_r is closed. Suppose on the contrary that $i \in C_r, j \notin C_r$, and $P_{ij} > 0, j \not\leftarrow i$, so $P(X_n$ never returns to $i) \geq P(X_n \text{reaches } j), P(X_n \neq i \text{ for } n \geq 1 | X_0 = i) \geq P(X_1 = j | X_0 = i) > 0$. This contradicts the assumption that i is recurrent.

Markov chains with finite state spaces are special. For example, it is impossible to stay in transient states for all time.

Theorem 3.2.6 - If the state space is finite, then at least one state is recurrent and all recurrent states are positive.

Proof - assume all states are transient. $1 = \Sigma_{j \in S} P_{ij}^n$ (finite sum). So, $1 = \lim_{n\to\infty} \Sigma_{j \in S} P_{ij}^n = \Sigma_{j \in S} \sim_{n\to\infty} P_{ij}^n = 0$ (by thm 3.1.2(3)). (in general $\lim_{n\to\infty} \Sigma_{i=0}^{\infty} a_i(n) \neq \Sigma_{i=0}^{\infty} \lim_{n\to\infty} a_i(n)$). That is a contradiction. The same argument works for the closed set of all null recurrent states (exercise).

Theorem 3.2.7 - Suppose the state space is finite. i is transient $\iff$ there is a state j with $i \to j$ but $j \not\to i$.

Example 3.2.6 - Consider a Markov chain with $S = \{0, 1, 2, 3, 4\}$.

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 \\ 1/4 & 1/4 & 0 & 0 & 0 \\ 0 & 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/8 & 3/8 \end{pmatrix}$$

$C_1 = \{0, 1\}, C_2 = \{2, 3, 4\}$ are both closed. They are both irreducible. They must contain positive recurrent states. $S = C_1 \cup C_2$.

Example 3.2.7 - Consider random walk with (1 0 0 ... & q 0 p 0 ... & 0 q 0 p 0 ... & 0 0 q 0 p 0 ... & ... & ... 0 q 0 p & 0 ... 0 0 1), states 0, ..., N.

Three classes $C_1 = \{0\}, T = \{1, 2, \ldots, N-1\}, C_2 = \{N\}.\{1, 2, \ldots, N-1\} \to \{0\}$, but $\{0\} \not\to \{1, 2, \ldots, N-1\}.\{1, 2, \ldots, N-1\} \to \{N\}$, but $\{N\} \not\to \{1, 2, \ldots, N-1\}.\{0\}$ and $\{N\}$ are absorbing.

State i is recurrent: $P(X_n = i \text{ some } n \geq 1 | X_0 = i) = 1$. i is transient: $P(X_n = i \text{ some } n \geq 1 | X_0 = i) < 1$. $f_{ij}(n) = P(X_1 \neq j, X_2 \neq j, \ldots, X_n \neq j, X_n = j | X_0 = i).f_{ij} = \Sigma_{n=1}^{\infty} f_{ij}(n)$. j is recurrent $\iff f_{jj} = 1$.

Theorem - j is recurrent if $\Sigma_n P_{jj}^n = \infty$. j is transient if $\Sigma_n P_{jj}^n < \infty$.

T_j =min$\{n \geq 1 : X_n = j\}$. time of first visit to $j(X_0 = i).P(T_i = \infty | X_0 = i) > 0 \iff i$ is transient. $\mu_i = E(T_i|X_0 = i) = \{\Sigma_{n=1}^{\infty} f_{ii}(n)$ for i recurrent, ∞ for i transient. Recurrent state is null if $\mu_i = \infty$. Recurrent state is positive if $\mu_i < \infty$.

Theorem - Recurrent state is null iff $P_{ii}^n \to 0$ as $n \to \infty$.

Example 3.2.8 - $S = \{0, 1, 2, 3, 4, 5\}$.

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 & 0 & 0 & 0 \\ 1/4 & 3/4 & 0 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 & 0 & 0 \\ 1/4 & 0 & 1/4 & 1/4 & 0 & 1/4 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 1/2 & 1/2 \end{pmatrix}$$

. $\{0, 1\}, \{4, 5\}$ irreducible and closed. They canta(?) in positive recurrent states. 2, 3 are transient: $2 \to 3 \to 5$. But return to 2 or 3 from 5 is impossible. $T = \{2, 3\}, C_1 = \{0, 1\}, C_2 = \{4, 5\}$. All states have period 1 since $P_{ii} > 0$ for all i (all entries on diagonal > 0). 0, 1, 4, 5 are ergodic. We can compute $f_0(1) = P_{00} = 1/2, f_{00}(n) = P_{01}(P_{11})^{n-2}P_{10} = 1/2(3/4)^{n-2}1/4, n \geq 2.\mu_0 = \Sigma_n f_{00}(n) * n = 3$.

Example 3.2.9 - Success Runs - $S = \{0, 1, \dots\}$. P = (q0 p0 0 ... & q1 0 p1 0 ... & q2 0 0 p2 0 ... & ...). $q_{ii}p_i \geq 0, q_i + p_i = 1$ for all i. This is a success run chain. Intuition: assume $p_i = p$ for all i. We attempt independent Bernoulli trials with probability p of success. We count the number of successful trials in a row. If we had n successes in a row, we can extend the run to $n+1$ if we have success on the next trial or we start over with a run of 0 if we fail in the next trial. This gives the row (q(0th) 0 ... p((n+1)st) 0 ...). We assume $0 < p_i < 1$ for all i so the chain is irreducible. This means state i is recurrent iff state 0 is recurrent. We have $f_{00}(1) = q_0$, and for $n \geq 2, f_{00}(n) = P(X_1 = 1, X_2 = 2, \dots, X_{n-1} = n-1, X_n = 0|X_0 = 0) = P_0P_1P_2 \dots P_{n-2} * q_{n-1}$. Set $U_n = \Pi_{i=0}^n P_i, n \geq 0$ since $q_{n-1} = 1 - P_{n-1}, f_{00}(n) = U_{n-2} - U_{n-1} = \Pi_{i=0}^{n-2} P_i(1 - P_{n-1})$. So $\Sigma_{n=1}^{N+1} f_{00}(n) = q_0 + (U_0 - U_1) + \dots + (U_{N-1} - U_N) = q_0 + U_0 - U_N = 1 - U_N$. 0 is recurrent iff $U_N = \Pi_{i=0}^N Pi \to 0$ as $N \to \infty$. L'Hopital's rule implies that if $0 < P_i < 1$ for all $i, U_N = \Pi_{i=0}^N P_i \to 0 \iff \Sigma_{i=0}^{\infty}(1 - P_i) = \infty.\Pi_{i=0}^{\infty} P_i > 0 \iff \Sigma_{i=0}^{\infty}(1 - P_i) < \infty$. 0 is recurrent iff $\Sigma_{i=0}^{\infty}(1 - P_i) = \infty$, or the P_i's cannot be too close to 1. If $P_i = 1 - (1/2)^i$, not recurrent. P_i constant then recurrent. (Chapter IV, section §3 in text.)

3.3 Stationary distributions and the limit theorem

We consider behavior as $n \to \infty$. Does the distribution of X_n converge to something?

Example 3.3.1 - ON/OFF system - ex 2.2.3 - P = $(1 - p$, p & q, $1 - q)$. P^n as before. $0 < p < 1, 0 < q < 1, P^n \to \frac{1}{p+q}*$(q p & q p) [as before]. We choose the initial state X_0 according to the probabilities $P(X_0 = 0) = \nu_0, P(X_0 = 1) = \nu_1 = 1 - \nu_0$.

Definition 3.3.1 - An initial distribution is a probability distribution for the initial state of a Markov chain. The probability distribution of X_1, conditioned on X_0 is $P(X_1 = j|X_0) = P_{0j}\nu_0 + P_{ij}\nu_1, j = 0, 1$. Matrix notation $(P(X_1 = 0|X_0)P(X_1 = 1|X_0)) = \nu p$. Suppose we take $\nu_0 = \frac{q}{q+p}, \nu_1 = \frac{p}{q+p}$. If we compute, $P(X_1 = 0) = (1 - p)\frac{q}{q+p} + q\frac{p}{q+p} = \frac{q}{p+q} = \nu_0$ and $P(X_1 = 1) = \nu_1$. In matrix notation $\nu = \nu p$. That particular initial distribution does not change over time.

Definition 3.3.2 - Let S = state space, The vector Π is a stationary distribution if $\Pi = (\Pi_i)_{i \in S}$ satisfies (1) $\Pi_i \geq 0$ all i, $\Sigma_{i \in S}\Pi_i = 1$, (2) $\Pi = \Pi P(\Pi_j = \Sigma_{i \in S}\Pi_i P_{ij}$ all $j \in S)$. P = probability transition matrix. These are also called invariant distributions and equilibrium distributions.

Theorem 3.3.1 - If Π is a stationary distribution, (3.3.1) $\Pi P^n = \Pi$ for all $n \geq 0$. If X_0 has distribution Π, then so does X_n for $n \geq 0$. Proof: exercise.

Aside: long time behavior of ODE's $\dot{y} = f(y)$. Stead-state/equilibrium solutions $f(y_s) = 0 \Rightarrow y_s$ constant, $\dot{y}_s = f(y_s) = 0$.

We assume the chain is irreducible and explore the existence of stationary distributions.

Example 3.3.2 - Consider ex 3.3.1 (ON/OFF), $\Pi = \Pi P = (\Pi_0\Pi_1)(1 - p$, p & q, $1 - q)$= $(\Pi_0\Pi_1) \Rightarrow \Pi_1 = p/q\Pi_0$ (1st equation), $\Pi_1 = p/q\Pi_0$ (2nd equation), $\Pi_0 + \Pi_1 = 1 \Rightarrow \Pi = (\frac{q}{p+q}, \frac{p}{p+q})$.

The assumption that the chain is irreducible is important.

Example 3.3.3 - Consider the genotype example, ex 3.1.4: P = (1 0 0 & 1/4 1/2 1/4 & 0 0 1). The chain is not irreducible. We try to find a stationary distribution anyway. $\Pi = \Pi P$ becomes $\Pi_0\Pi_1\Pi_2)$ (1 0 0 &

1/4 1/2 1/4 & 0 0 1)= $(\Pi_0\Pi_1\Pi_2)$. 1st: $\Pi_0 + 1/4\Pi_1 = \Pi_0 \Rightarrow \Pi_1 = 0$. 2nd: $0 = 0$. 3rd: $\Pi_2 = \Pi_2$. Any distribution of the form $\Pi = (\alpha, 0, 1-\alpha, 0 \leq \alpha \leq 1$ is a stationary distribution.

In the finite state space case, we later prove: Theorem 3.3.2 - If the state space has r states, then $\Pi = \Pi P$ has at most $r-1$ linearly independent equations. Including the equation $\Sigma_{j\in S}\Pi_j = 1$, we obtain at most r linearly independent equations. There is always atleast one solution. If the chain is not irreducible there may be more than one solution. The situation with an infinite state space case is more complicated.

Example 3.3.4 - Consider the gambler's ruin model, ex 3.1.12, where A has a backer and B is infinitely rich. Assume $r_i = 0$, and $q_i = p_i = 1/2$, and A starts with \$1. P = (1/2 1/2 0 ... & 1/2 0 1/2 0 ... & 0 1/2 0 1/2 0 ... & ...). $\Pi = \Pi P$. 1st: $1/2\Pi_0 + 1/2\Pi_1 = \Pi_0 \Rightarrow \Pi_1 = \Pi_0$. 2nd: $1/2\Pi_0 + 1/2\Pi_2 = \Pi_1 = \Pi_0 \Rightarrow \Pi_2 = \Pi_0$. Continuing, we find that IF the distribution exists, then $\Pi_n = \Pi_0$ for all n. $\Pi = (\Pi_0, \Pi_0, \dots) \Rightarrow \Sigma\Pi_0 = 1$, which is impossible. So there are no such stationary distributions.

Suppose $r_i = 0$, but $P_0 = P_1 = P_2 = \cdots = p < 1/2$. P = (1-p p 0 ... & 1-p 0 p 0 ... & 0 1-p 0 p 0 ... & ...). $\Pi P = \Pi$. 1st: $(1-p)\Pi_0 + (1-p)\Pi_1 = \Pi_0 \Rightarrow \Pi_1 = \frac{p}{1-p}\Pi_0$. 2nd: $\Pi_2 = (\frac{p}{1-p})^2\Pi_0, \dots, \Pi_n = (\frac{p}{1-p})^n\Pi_0, n \geq 1, \Sigma\Pi_i = 1 \Rightarrow 1 = \Pi_0 * \Sigma_{n=0}^{\infty}(\frac{p}{1-p})^n = \Pi_0\frac{1-p}{1-2p}, \Pi_n = \frac{1-2p}{1-p}(\frac{p}{1-p})^n, n \geq 0$ for the stationary distribution.

In both cases, the chain is irreducible and recurrent. In the first example, it is null, and in the second it is positive.

Theorem 3.3.3 - An irreducible chain has the stationary distribution Π iff all the states are positive recurrent. In this case there is a unique stationary distribution. The solution is $\Pi_i = \frac{1}{\mu_i}, i \in S$, where μ_i is the mean recurrence time. The proof has several parts.

Definition 3.3.3 - Fix state k. Define $\rho_i(k)$ = mean number if visits to state i between two successive visits to state k. $T_k = \min\{n \geq 1 : X_n = k | X_0 = k\}$ = time of first return to state k. $N_i = \Sigma_{n=1}^{\infty} 1_{\{X_n = i\}\cap\{T_k \geq n\}}$ (count 1s). $\{X_n = i\} \cap \{T_k \geq n\}$ is the event where X_n is in state i and the time to return to state k is larger than n. N_i counts the number of visits to state i between successive visits to state k. Hence, $\rho_i(k) = E(N_i | X_0 = k)$. Now $N_k = 1$, so $\rho_k(k) = 1$. We also have (3.3.2) $\rho_i(k) = \Sigma_{n=1}^{\infty} P(X_n = i, T_k \geq n | X_0 = k)$. $\rho(k) = (\rho_1(k)\rho_2(k)\dots)$.

Theorem 3.3.4 - The mean recurrence time μ_k satisfies (3.3.3) $\mu_k = \Sigma_{i\in S}\rho_i(k)$. Proof: The time between visits to state k must be spent in some state. $T_k = \Sigma_{i\in S} N_i$. Take expectations.

Theorem 3.3.5 - For any state k of an irreducible, recurrent chain, $\rho(k)$ satisfies $\rho_i(k) < \infty$ for all i and $\rho(k) = \rho(k)P$. Proof: first, $\rho_i(k) < \infty$. Set $l_{ki}(n) = P(X_n = i, T_k \geq n | X_0 = k)$ = probability that the chain reaches state i in n steps, with no intermediate returns to state k. Observation: the first return time to k is $m+n$ if (1) $X_m = i$, (2) there is no return to k before m steps, and (3) the next visit to state k occurs after n more steps. This implies $f_{kk}(m+n) \geq l_{ki}(m)f_{ik}(n)$. Since the chain is irreducible, there is an n with $f_{ik}(n) > 0$. Using this n, $l_{ki}(m) \leq \frac{f_{kk}(m+n)}{f_{ik}(n)}$. We have $\rho_i(k) = \Sigma_{n=1}^{\infty} l_{ki}(m)$ (important, exercise: explain) $\leq \frac{1}{f_{ik}(n)}\Sigma_{m=1}^{\infty} f_{kk}(m+n) \leq \frac{1}{f_{ik}(n)} < \infty$. To prove the second claim, start with $\rho_i(k) = \Sigma_{n=1}^{\infty} l_{ki}(n)$. Now $l_{ki}(n) = P_{ki}, n \geq 2, l_{ki}(n) = \Sigma_{j\in S, j\neq k} P(X_n = i | X_{n-1} = j, T_k \geq n | X_0 = k) = \Sigma_{j\in S, j\neq k} l_{kj}(n-1)P_{ji}$. Summing in n, $\rho_j(k) = P_{ki} + \Sigma_{j\in S, j\neq k}(\Sigma_{n\geq 2} l_{kj}(n-1))P_{ji} = \rho_k(k)P_{ki} + \Sigma_{j\in S, j\neq k}\rho_j(k)P_{ji}$.

Theorem 3.3.6 - Every positive recurrent irreducible chain has a stationary distribution. Proof: We just proved $\rho(k) = \rho(k)P$. The components of $\rho(k)$ are nonnegative and sum to μ_k. If $\mu_k < \infty, \Pi : \Pi_i = \rho_i(k)/\mu_k \Rightarrow \Pi = \Pi P, \Pi$ is a p.m.f.

Decomposition Theorem - S = state space. Π is a stationary distribution. If (1) $\Pi_j \geq 0, \Sigma_{j\in S}\Pi_j = 1$, (!!!) (2) $\Pi = \Pi P, \Pi$ is a LEFT eigenvector of the prob transition matrix. This implies $\Pi = \Pi P^n, n \geq 1$. Does Π exist? (Thm 3.3.2:) Finite dimension state space $\Rightarrow$ theory of eigenvectors.

Theorem 3.3.3 (no assumption of finite state) - An irreducible chain has a stationary distribution $\Pi \iff$ all the states are positive recurrent. In this case, Π is unique, $\Pi_i = \frac{1}{\mu_i}, i \in S$. Let $\rho_i(k)$ = mean number of visits to state i between successive visits to state k.

Theorem (3.3.3) - $\mu_k = \Sigma_{i\in S}\rho_i(k)$.

Theorem - If $\rho(k) = (\rho_i(k))$, if the communication class of k is irreducible and recurrent, then $\rho_i(k) < \infty$ for all i and $\rho(k) = \rho(k)P$.

Theorem - Every positive recurrent irreducible Markov chain has a stationary distribution.

Theorem 3.3.7 - If the chain is irreducible and recurrent there is a solution of $x = xP$ with strictly positive entries that is unique up to a multiplicative factor. The chain is positive if $\Sigma X_i < \infty$ and null if $\Sigma X_i = \infty$. Now we are ready to prove theorem 3.3.3.

Proof of Theorem 3.3.3 - Suppose Π is a stationary distribution. If all the states are transient then $P_{ij}^n \to 0$ as $n \to \infty$ by Thm 3.1.2. By (3.3.1), (3.3.4) $\Pi_j = \Sigma_i \Pi_i P_{ij}^n \to 0, n \to \infty$ for all j. This contradicts defn 3.3.2(1), hence all states are recurrent. (More argument in notes.) We next show that all states are positive and $\Pi_i = \mu_i^{-1}$ for all i. Suppose X_0 has distribution Π. Exercise: $\Pi_j \mu_j = \Sigma_{n=1}^\infty P(T_j \geq n | X_0 = j) P(X_0 = j) = \Sigma_{n=1}^\infty P(T_j \geq n, X_0 = j)$. $P(T_j \geq 1, X_0 = j) = P(X_0 = j), n \geq 2, P(T_j \geq n | X_0 = j) = P(X_0 = j, X_m \neq j, 1 \leq m \leq n-1) = P(X_m \neq j, 1 \leq m \leq n-1) - P(X_m \neq j, 0 \leq m \leq n-1) = P(X_m \neq j, 0 \leq m \leq n-2) - P(X_m \neq j, 0 \leq m \leq n-1) = a_{n-2} - a_{n-1}, a_n = P(X_m \neq j, 0 \leq m \leq n)$. Sum over $n, \Pi_j \mu_j = P(X_0 = j) + P(X_0 \neq j) - \lim_{n\to\infty} a_n = 1 - \lim_{n\to\infty} a_n$. $a_n \to P(X_m \neq j$ for all $m) = 0$ since j is recurrent. We have shown that $\mu_j \Pi_j = 1$ so $\mu_j = \Pi_j^{-1} < \infty$ if $\Pi_j < 0$. To show $\Pi_j > 0$ for all j, assume some $\Pi_j = 0$. $0 = \Pi_j = \Sigma_{j(\text{mightbeiinsteadofj})} \Pi_i P_{ij}^n \geq \Pi_i P_{ij}^n$ for all i, n. This means $\Pi_i = 0$ when $j \to i$. The chain is irreducible so $\Pi_i = 0$ for all i. This contradicts $\Sigma_i \Pi_i = 1$. Hence $\mu_j < \infty$ for all j, and all the states are positive. The other direction follows from the intermediate steps.

Proof of Thm 3.2.3 (3) is given on page 163.

Example 3.3.5 - OFF/ON system: ex 3.3.2 - P = (1/2 1/2 & 1/4 3/4). We can compute $\Pi = (1/3, 2/3) \Rightarrow \mu_1 = 3, \mu_2 = 3/2$. (What is new here is that we can now compute the mean recurrence times, μ_i.)

Example 3.3.6 - Consider Gambler's Ruin in ex 3.3.4 where $P < 1/2$. P = (1-p p 0 ...& 1-p 0 p 0 ...& 0 1-p 0 p 0 ...& ...). Using the formula for Π_n computed there, $\mu_n = \Pi_n^{-1} = \frac{1-p}{1-2p}(\frac{1-p}{p})^n, n \geq 0, p = 1/4 \Rightarrow \mu_n = \frac{3}{2}3^n, n \geq 0$.

Theorem 3.3.3 can be used to determine if an irreducible chain is positive recurrent.

Theorem 3.3.8 - Let $s \in S$ be a state of an irreducible chain. The chain is transient iff there is a nonzero solution $\{Y_i, i \in S\}$ of the equations (3.3.5) $Y_i = \Sigma_{j \in S, j \neq s} P_{ij} Y_j, i \neq s$, with $|Y_j| \leq 1$ for all j.

Theorem 3.3.8 - Let $s \in S$ be a state of an irreducible chain. The chain is transient $\iff$ there is a nonzero solution$\{Y_i, i \in S\}$ of (3.3.5) $Y_i = \Sigma_{j \in S, j \neq s} P_{ij} Y_j, i \neq s$, with $|Y_i| \leq 1$ for all i.

Proof - The chain is transient iff s is transient. Suppose s is transient. Define (3.3.6) $\tau_i(n) = P($no visit to s in the first n steps$|X_0 = i) = P(X_m \neq s, 1 \leq m \leq n | X_0 = i)$. $\tau_i(1) = \Sigma_{j \neq s} P_{ij}$, which is $X_1 \neq s$. $\tau_i(n+1) = \Sigma_{j \neq s} P_{ij} \tau_i(n)$. Furthermore, $\tau_i(n) \geq \tau_i(n+1), \tau_i = \lim_{n\to\infty} \tau_i(n) = P($no visit to s$|X_0 = i) = 1 - f_{is}$. Exercise: τ_i satisfies (3.3.5). Also $\tau_i > 0$ for some i. Otherwise $f_{is} = 1$ for all $i \neq s$. This implies (condition on X_1) $f_{ss} = P_{ss}[X_1 = s] + \Sigma_{i \neq s} P_{si} f_{is} [X_1 \neq s] = \Sigma_i P_{si} = 1$ which contradicts the transiency of s. Let Y satisfy (3.3.5) with $|Y_i| \leq 1$. $|Y_i| \leq \Sigma_{j \neq s} P_{ij} |Y_j| \leq \Sigma_{j \neq s} P_{ij} = \tau_i(1)$. Going back to (3.3.5) $|Y_i| \leq \Sigma_{j \neq s} P_{ij} \tau_j(1) = \tau_i(2) \ldots$. So, $|Y_i| \leq \tau_i(n)$ for all n. Exercise: Let $n \to \infty$ to show that $\tau_i = \lim_{n\to\infty} \tau_i(n) > 0$ for some i ($Y_i \neq 0$ for some i by assumption) which implies s is transient.

Theorem 3.3.9 - An irreducible chain is recurrent iff the only bounded solution of (3.3.5) is the zero solution.

Example 3.3.7 - Gambler's Ruin ex 3.1.2 - P = (q p 0 & q 0 p 0 & 0 q 0 p 0 & ...), $q + p = 1$. Set $\gamma = \frac{p}{q}$. (1) If $q < p(\gamma > 1)$, choose $s = 0$ to test thm 3.3.8. (3.3.5) read: $Y_0 = P_{01} Y_1 = pY_1, Y_1 = P_{02} Y_2 = pY_2, Y_2 = qY_1 + pY_3$(this is what was written on the board, should the + be there?), ... Exercise: if $Y_j = 1 - \gamma^{-j}$, then Y solves the equations and the chain is transient. (2) We can solve $\Pi = \Pi P$ to find a stationary solution with $\Pi_j = \gamma^j (1 - \gamma) \iff q > p$. The chain is positive recurrent $\iff q > p$.

Example 3.3.8 - Consider discrete queuing (ex 2.1.10). Customers arrive at a service place and take a place in a line (queue). In each period of time, 1 customer is served and a random number arrive. C_n =number of customers that arrive during nth period. $P(C_n = k) = a_k$, where a = p.m.f. X_n = number of customers waiting in line at time n. $X_{n+1} = \max\{X_n - 1, 0\} + C_n$. P = (a0 a1 a2 ...& a0 a1 a2 ...& 0 a0 a1 a2 ...& ...). $\Pi = \Pi P, \Pi_0 = \Pi_0 a_0 + \Pi_1 a_0, \Pi_1 = \Pi_0 a_1 + \Pi_1 a_1 + \Pi_2 a_2, \Pi_2 = \Pi_0 a_2 + \Pi_1 a_2 + \Pi_2 a_1 + \Pi_3 a_0, \ldots$. (3.3.7) $\Pi_i = a_0 \Pi_{i+1} + \Sigma_{j=1}^{i+1} \Pi_j a_{i+1-j}$. We use generating functions: $\Pi(t) = \Sigma_{i=0}^\infty \Pi_i t^i$. We multiply (3.3.7) by t^i and sum. (3.3.8) $\Pi(t) = \Pi_0 * \Sigma_{i=0}^\infty a_i t^i + \Sigma_{i=0}^\infty \Sigma_{j=1}^{i+1} \Pi_j a_{i+1-j} t^i . 1 \leq j \leq i+1 \Rightarrow i \geq j-1, j \geq 1, (A(t) =$ p.g.f. for a)$A(t) = \Sigma_{i=0}^\infty a_i t^i$, the right-hand side of (3.3.8) is $\Pi_0 A(t) + \Sigma_{j=1}^\infty \Pi_j t^{j-1} * \Sigma_{i=j-1}^\infty a_{i-j+1} t^{i-j+1} = \Pi_0 A(t) = t^{-1}(\Sigma_{j=1}^\infty \Pi_j t^j) A(t) = \Pi_0 A(t) + t^{-1}(\Pi(t) - \Pi_0) A(t)$ or $\Pi(t) = \Pi_0 A(t)(1 - t^{-1} + t^{-1} \Pi(t) A(t) \Rightarrow$ (3.3.9)$\Pi(t) = \frac{\Pi_0 A(t)}{1 - \frac{1 - A(t)}{1-t}}$. The question is: when is it possible to specify Π_0 so $\Pi(1) = \Sigma_k \Pi_k = 1$. This implies a stationary distribution exists. Since $\{a_k\}$ is a pmf, $A(1) = 1$. We want to let $t \uparrow 1$ in (3.3.9). We let $\lim_{t\uparrow 1} \frac{1 - A(t)}{1-t} = A(1) = \gamma = \Sigma_{k=0}^\infty k a_k, \gamma =$ mean number of arrivals per service interval. [Something illegible] $t \uparrow 1$ in (3.3.9), $\Pi(1) = \frac{\Pi(0)}{1-\gamma}$. We can choose Π_0 so $\Pi(1) = 1 \iff 0 < \gamma < 1$ and then $\Pi_0 = 1 - \gamma$. The queuing chain is positive recurrent $\iff \gamma < 1$, which says that the mean number of arrivals does overwhelm

the facility.

Next consider $\gamma > 1$. We show that (3.3.5) has a nonzero solution Y with $0 \leq Y_i \leq 1$ for all i. We choose $s = 0$, use (3.3.5) to get (3.3.10) $Y_1 = \Sigma_{i=1}^{\infty} a_i Y_i, \ldots, Y_n = \Sigma_{i=0}^{\infty} a_i Y_{i+n-1}, n \geq 2$. Guessing based on branching processes, we try $Y_i = 1 - t^i, 0 < t < 1$. $Y_n : 1 - t^n = \Sigma_{i=0}^{\infty} a_i (1 - t^{i+n-1}) = 1 - (\Sigma_{i=0}^{\infty} a_i t^i) t^{n-1}$, with $A(t) = \Sigma_{i=0}^{\infty} a_i t^i, t^n = A(t) t^{n-1} \Rightarrow t = A(t)$. The branching process (fixed point) analysis $\Rightarrow t = A(t)$ has a solution with $0 < t < 1$ when $\gamma > 1$. $\gamma > 1 \Rightarrow$ chain is transient. We argue that if the chain is transient, (3.3.10) has a nonzero solution. If the chain is transient, then for each $j = 0, 1, 2, \ldots$, there is a last visit. There is therefore a last visit to any finite set $\{0, 1, 2, \ldots, M\}$. So there is an $n_0 = n_0(M)$ such that for $n > n_0, X_n > M$. Hence, $X_n \to \infty$ as $n \to \infty$. A_{n+1} = number of arrivals in $(n, n+1)$.$P(A_{n+1} = k) = a_k, E(A_{n+1}) = p, X_{n+1} = \max\{X_n - 1, 0\} + A_{n+1}, n \geq n_0, X_{n+1} = X_{n-1} + A_{n+1}$. (This is a way to get out of a low customer state when we have a low number in the queue.) If $N \geq n_0, \Sigma_{n=n_0}^{N} (X_{n+1} - X_n) = -(N - n_0) + \Sigma_{n=n_0}^{\infty} A_{n+1}, X_{N+1} - X_{n_0} = -(N - n_0) + \Sigma_{n=n_0+1}^{N+1} A_n, X_{N+1} - \Sigma_{n=1}^{N+1} (A_{n-1}) = X_{n_0} + n_0 - \Sigma_{n=1}^{n_0} A_n$ (entire last term up until = is constant, doesn't depend on N). Therefore $X_{N+1} \to \infty$ implies $\Sigma_{n=1}^{N+1} (A_{n-1}) \to \infty$. Exercise: a sum of iid rv with mean μ converges to $\infty \iff \mu > 0$ equivalently $\rho > 1$.

Review from a question in class: why did we need a PGF last time? 1) Basic problem: finding $\Pi = \Pi P, \Pi$ pmf. Π exists sometimes. Relation to being positive recurrent. 2) Basic problem: find a bounded solution of $Y = PY$ with a row/column "missing", related to transient. Then we got to ex 3.3.8 - classic example: queuing theory. This example had several points: 1) Consider $\Pi = \Pi P$. Used generating functions. Came to the conclusion that Π exists $\iff \gamma = \Sigma_{k=0}^{\infty} k a_k < 1$, but only if chain is positive recurrent. So we needed to proved that the chain is positive recurrent so we can solve the problem. 2) We consider solution of $Y = PY, |Y_i| \leq 1$, with missing row/column (not full P). It this case using the fixed point theory we prove Y exists, but this holds iff chain is transient, which holds iff $\gamma > 1$. So, theorem is: positive recurrent if $\gamma < 1$, transient if $\gamma > 1$.

3.4 Limit Theorems

We explore the link between a stationary distribution and the behavior of P_{ij}^n as $n \to \infty$.

Example 3.4.1 - Consider ON/OFF system in ex 2.2.3 with P = (1-p p & q 1-q) $0 \leq p \leq 1, 0 \leq q \leq 1$. When $0 < p, q < 1, P^n \to \frac{1}{p+q}$(q p & q p) and we know there is a stationary distribution. Now suppose $p = q = 1$, the system changes states at every step. The stationary distribution satisfies $(\Pi_0, \Pi_1) = (\Pi_0, \Pi_1)$(0 1 & 1 0) $\Rightarrow \Pi_0 = \Pi_1 = 1/2$. We can compute, e.g. $P_{00}^n = \{$ 0 if n even, 1 if n odd. There is no limiting behavior in this case. However the states are periodic with period 2.

Theorem 3.4.1 - For an irreducible, aperiodic Markov chain, (3.4.1) $P_{ij}^n \to \frac{1}{\mu_j}$ as $n \to \infty$ for all i, j (μ_j is the mean recurrence time). The limiting value is the same for all states i. $P^n \to$ (1/mu0 1/mu0 ... & 1/mu1 1/mu1 ... & ...)^T = (1/mu0 1/mu1 ... & 1/mu0 1/mu1 ... & ...).

Definition 3.4.1 - If there is a probability distribution q on the state space S such that $P_{ij}^n \to q_j$ for all $i, j \in S$ then q is a limit distribution of the chain.

Intuition: q_j describes the probability that the chain is in state j at some "late" time and by this time the chain has "forgotten" where it started. $P(X_n = j) = \Sigma_i P(X_0 = i) P_{ij}^n \to q_j$ regardless of the initial distribution of X_0. Consequences:

Theorem 3.4.2 - (a) If the chain is transient or null recurrent, then $P_{ij}^n \to 0$ for all i, j. (b) If the chain is positive recurrent, then $P_{ij}^n \to \Pi_j = \mu_j^{-1}$, where Π is the unique stationary distribution.

Theorem 3.4.3 - If X_n is an irreducible chain with period d, then $Y_n = X_{nd}, n \geq 0$, is an aperiodic, irreducible chain, $P_{jj}^{nd} = P(Y_n = j | Y_0 = j) \to \frac{d}{\mu_j}$ as $n \to \infty$. Immediately from this follows the proof of theorem 3.1.6 (see notes).

Connection between limiting and stationary distributions: consider a Markov chain at some "late" time n. The stationary distribution gives the proportion of time spent in the different states up to time n.

The limit distribution gives the proportion of "time" spent in the various states at the large time, where we count by considering many realizations.

Example 3.4.2 - Consider ON/OFF system, ex 3.4.1 again. The stationary distribution $(1/2, 1/2)$ says that equal amounts of time are spent in each state up to some large time, say $n = 1000$. If we look precisely at $n = 1000$, the chain must return to its initial state. Multiple realizations give probability 1 to be in the initial state and 0 to be in the other.

Theorem 3.4.4 - An ergodic Markov chain has the property that it has both stationary and limiting distributions and these are equal.

Proof of theorem 3.4.1 - We treat different cases. The simplest case is a transient chain, because theorem 3.1.2 (3) implies $P_{ij}^n \to 0$ as $n \to \infty$ for all i, j. The recurrent cases are treated with "coupling".

Definition - 3.4.2 - Let X_n, Y_n be independent Markov chains with common state space S and common probability transition matrix P. The coupled chain $Z_n = (X_n, Y_n)$ taking values in $S \times S$.

Theorem 3.4.5 - Z_n is a Markov chain with $P_{ij,kl} = P_{ik}P_{jl}$. If X_n, Y_n are irreducible and aperiodic, then Z_n is irreducible. Proof: $P_{ij,kl} = P(Z_{n+1} = (k,l)|Z_n = (i,j)) = P(X_{n+1} = k|X_n = i) \times P(X_{n+1} = l|Y_n = j).X_n, Y_n$ aperiodic, irreducible $\Rightarrow$ for any i, j, k, l there is an $N = N(i, j, k, l)$ such that $P_{ik}^n P_{jl}^n > 0, n \leq N$. Exercise: this implies Z_n is irreducible.

Comment: this is the only place we use the assumption X_n is aperiodic.

We assume X_n (in thm 3.4.1) is positive recurrent, so it has unique stationary distribution Π. (Consider $Y = X$ in the construction of Z.) Exercise: $Z_n = (X_n, Y_n)$ has a stationary distribution $\nu = (\nu_{ij}, i, j \in S), \nu_{ij} = \Pi_i \Pi_j$. This implies Z_n is also positive recurrent (due to stationary distribution). Choose $X_0 = i, Y_0 = j, Z_0 = (i, j)$. Choose $s \in S$. Set $T = \min(n \geq 1 : Z_n = (s, s))$. The recurrence of Z_n implies that $P(T < \infty) = 1$ (exercise).

(Review) Theorem 3.4.1 - irreducible, aperiodic MC. $P_{ij}^n \to_{n\to\infty} \frac{1}{\mu_j}$ for all i, j. μ_j is mean recurrent time.

Theorem 3.4.4 - An ergodic MC has both stationary and limiting distributions and these are equal.

Recurrent cases in proof: X_n, Y_n have common S, P. $Z_n = (X_n, Y_n) \in S \times S, Z_n$ M.C. $P_{ij,kl} = P_{ik}P_{jl}$. Assume X_n positive recurrent. Π is unique stationary distribution. (With Y_n "=" X_n), Z_n has stationary distribution $\nu = (\nu_{ij}, i, j \in S), \nu_{ij} = \Pi_i \Pi_j$. Z_n is positive recurrent. Choose $X_0 = i, Y_0 = j, Z_0 = (i, j)$. Choose $s \in S$. $T = \min\{n \geq 1 : Z_n = (s, s,)\}$. Z_n recurrent $\Rightarrow P(T < \infty) = 1$.

Observation: Suppose $m \leq n$ and $X_m = Y_m$. Then X_n and Y_n are identically distributed. Thus conditional on $\{T \leq n\}, X_n$ and Y_n have the same distribution. We use this observation and the fact that T is finite to prove that in large time, the distributions of X_n and Y_n are independent of the initial values.

Computation: Start from $Z_0 = (i, j).P_{ik}^n = P(X_n = k) = P(X_n = k, T \leq n) + P(X_n = k, T > n) = P(Y_n = k, T \leq n) + P(Y_n = k, T > n) \leq P(Y_n = k) + P(T > n) = P_{jk}^n + P(T > n)$. The symmetric argument implies $P_{jk}^n \leq P_{ik}^n + P(T > n)$. Hence, $|P_{ik}^n - P_{jk}^n| \leq P(T > n) \to_{n\to\infty} 0$ (T finite!), for all $i, j, k \in S$. So (3.4.2) $P_{ik}^n - P_{jk}^n \to 0$ as $n \to \infty$ for all $i, j, k \in S$. If $\lim_{n\to\infty} P_{ik}^n$ exists, then it is independent of i. We show the limit exists. We write $\Pi_k - P_{jk}^n = \Sigma_i \Pi_i (P_{ik}^n - P_{jk}^n)$ (we can write this because Π is stationary (Π_k, P_{ik}^n) and a pmf (P_{jk}^n). For any finite set $F \in S, \Sigma_{i\in S} \Pi_i |P_{ik}^n - P_{jk}^n| \leq \Sigma_{i\in F} |P_{ik}^n - P_{jk}^n| + 2\Sigma_{i\notin F} \Pi_i$. As $n \to \infty$, this converges to $2\Sigma_{i\notin F} \Pi_i$. This converges to 0 as $F \uparrow S$. So (3.4.3) $\Pi_k - P_{jk}^n = \Sigma_i \Pi_i (P_{ik}^n - P_{jk}^n) \to_{n\to\infty} 0$. Read in notes about when X_n is null recurrent, 185-187.

More general version of theorem 3.4.1 (no proof given) that drops irreducibility. Theorem 3.4.6 - For any aperiodic state j of a Markov chain, $P_{jj}^n \to \frac{1}{\mu_j}$ as $n \to \infty$. If i is any other state, $P_{ij}^n \to \frac{f_{ij}}{\mu_j}$ as $n \to \infty$. More to the theorem in notes.

3.5 Reversibility

Some physical situations have the property that observations of the system taken at some times look the same if time runs forward or backward. Let X_n be a Markov chain, $\{X_n, 0 \leq n \leq N\}$ irreducible, positive recurrent Markov chain, prob transition matrix P and stationary distribution Π.

Definition 3.5.1 - The reversed chain or time reversal Y_n is $Y_n = X_{N-n}, 0 \leq n \leq N$.

Theorem 3.5.1 - Y_n is a Markov chain with $P(Y_{n+1} = j|Y_n = i) = \frac{\Pi_j}{\Pi_i} P_{ji}$. Proof: $P(Y_{n+1} = i_{n+1}|Y_n = i_n, \ldots, Y_0 = i_0) = \frac{P(Y_k = i_j, 0 \leq k \leq n+1)}{P(Y_k = i_k, 0 \leq k \leq n)} = \frac{P(X_{N-n-1} = i_{n+1}, X_{N-n} = i_n, \ldots, X_N = i_0)}{P(X_{N-n} = i_n, \ldots, X_N = i_0)} = \frac{\Pi_{i_{n+1}} P_{i_{n+1}, i_n} P_{i_n, i_{n-1}}, \ldots, P_{i_1, i_0}}{\ldots} = \frac{\Pi_{i_{n+1}} P_{i_{n+1}, i_n}}{\Pi_{i_n}}$.

Definition 3.5.2 - The chain is reversible if the probability transition matricies of X_n and its time reversal Y_n are the same, (3.5.1) $\Pi_i P_{ij} = \Pi_j P_{ji}$ for all i, j. (3.5.1) are called the detailed balance equations. A transition matrix P and a probability distribution λ are in detailed balance if $\lambda_i P_{ij} = \lambda_j P_{ji}$ for all i, j. An irreducible chain X_n with a stationary distribution Π is reversible in equilibrium if its probability transition matrix is in detailed balance with Π.

Theorem 3.5.1 - Let P be the probability transition matrix of an irreducible chain X_n and suppose there is a distribution Π with $\Pi_i P_{ij} = \Pi_j P_{ji}$ for all $i, j \in S$. Then Π is the stationary distribution of X_n and X_n

is reversible in equilibrium. Proof: $\Sigma_i \Pi_i P_{ij} = \Sigma_i \Pi_j P_{ji} = \Pi_j \Sigma_i P_{ji} = \Pi_j$ or $\Pi = \Pi P$.

X_n irreducible positive recurrent M.C. Stationary distribution Π. Time reversal $Y_n = X_{N-n}, 0 \le n \le N$. Y_n is a M.C. $P(Y_{n+1} = j | Y_n = i) = \frac{\Pi_j}{\Pi_i} P_{ji}$. X_n is reversible if $\Pi_i P_{ij} = \Pi_j P_{ji}$ for all i, j.

If there is a distribution Π such that (3.5.1) holds, then Π is the stationary distribution and X_n is reversible.

Example 3.5.1 - Consider ON/OFF system in ex 2.2.3, P = (1-p, p & q, 1-q), $0 \le p \le 1, 0 \le q \le 1$. (3.5.1) $\Rightarrow \Pi_0 P_{00} = \Pi_0 P_{00}, \ldots$ (see notes). $\Rightarrow \Pi_0 P = \Pi_1 q$. Recall the stationary distribution is $\Pi_0 = \frac{q}{p+q}, \Pi_1 = \frac{p}{p+q}$. These equations hold, so the chain is reversible.

Example 3.5.2 - Ehrenfest Model of Diffusion - (See notes for drawing.) Two containers (A, B), stopped at top, connected with a small tube at the neck. Filled with M total molecules of a gas. At each time, one molecule picked at random moves between the containers. X_n = number of molecules in A at time n. X_n is a M.C. P = (0, 1, 0, ... & 1/m, 0, 1-(1/m), 0 ... & 0, 2/m, 0, 1-(2/m), 0, ... & ...). $P_{i,i+1} = 1 - \frac{i}{m}, P_{i,i-1} = \frac{i}{m}, 0 \le i \le m$. If we look for solutions of (3.5.1) we find $\Pi_i = \binom{m}{i}(\frac{1}{2})^m$, which is a stationary distribution and the chain is reversible.

3.6 Chains with finitely many states

As discussed before, the theory is much simpler when there is a finite state space. Theorem 3.2.6 implies If X_n is irreducible and has a finite state space, then it is positive recurrent.

Theorem 3.6.1 - An irreducible aperiodic chain with a finite state space has a stationary distribution which is also a limiting distribution.

Theorem 3.6.2 - Perron-Fröbenius - If P is the transition probability matrix of an irreducible chain with period d and a finite state space, (1) $\lambda_0 = 1$ is an eigenvalue of P. (2) the d complex roots: $\lambda_1 = e^{2\pi i/d}, \lambda_2 = e^{2\pi i 2/d}, \ldots, \lambda_d = e^{2\pi i \frac{d-1}{d}}, i = \sqrt{-1}, (\lambda_m)^d = 1$ are eigenvalues of P (look in notes for cool drawing of the unit circle with d roots drawn as spokes). (3) the remaining eigenvalues $\lambda_{d+1}, \ldots, \lambda_m$ satisfy $|\lambda_m| \le 1$.

What this means: the power method. Let P be a matrix. Suppose $\{V_1, \ldots, V_N\}$ is a basis of eigenvectors. $PV_i = \lambda_i V_i$. Let V be arbitrary, then $V = \Sigma_{j=1}^N \alpha_j V_j$ for some $\{\alpha_1, \ldots, \alpha_N\}$. Think of V as being the initial condition of a markov process (X_0). So [some stuff that I don't get and isn't in the notes.] $P^n V = \Sigma_j \alpha_j \lambda_j^n V_j$. Suppose λ_1 is the largest $|\lambda_1| > |\lambda_j|, j \ne 1$. More stuff I don't understand.

Example 3.6.1 - Suppose the chain is aperiodic and P has distinct eigenvalues. There is a B with $P = B^{-1} \wedge B = B^{-1}$ (diagonal matrix with λ_i in diagonals, $0 \le i \le N$, 0 else) $B, P^n = B^{-1} \wedge^n B$ and $|\lambda_i|^n \to 0$ for $i \ge 1$. $|\lambda_0| = 1$.

3.7 Branching Processes

Unfortunately the preceeding theory is not directly useful. Consider the branching process: $\{Z_{nj}, n \ge 1, j \ge 1\}$ i.i.d. nonnegative i.v., pmf $\{P_k\}$. $X_0 = 1, X_1 = Z_{1,X_0}, X_2 = Z_{21} + Z_{22} + \cdots + Z_{2,X_1}, \ldots, X_n = Z_{n,1} + \cdots + Z_{n,X_{n-1}}$. If there is a strictly positive probability that each family is empty, $P(X_n = 0) > 0$ then 0 is an absorbing state. Hence, 0 is a positive recurrent state and all other states are transient. The chain is not irreducible, but there is a unique stationary distribution: $\Pi_0 = 1, \Pi_i = 0$ for $i \ge 1$. This gives no interesting information. The difficulty is that the process can behave in a number of different ways depending on, for example, whether or not it becomes extinct. We study the behavior conditional on some event, e.g. extinction. Recall $\rho_k = P(X_1 = k), \phi(s) = E(s^{X_1}), N = \inf\{n : X_n = 0\}$ = time of extinction [inf = min]. $N = \infty$ if $X_n \ne 0$ for all n. If $N = \infty$, the process grows without bound and if $N < \infty$, the process is limited in size and eventually becomes extinct. Recall the probability $\eta = P(N < \infty)$ of extinction is the smallest nonnegative root of $S = \phi(s)$. We let $E_n = \{n < N < \infty\}$ be the event that extinction occurs at some time larger than n. We studied the distribution of the X_n conditional on E_n. We let $P_{0,j}(n) = P(X_n = j | E_n)$ be the conditional probability that $X_n = j$ given the future extinction. We try to compute $\Pi_{0,j} = \lim_{n\to\infty} P_{0,j}(n)$ (if the limit exists). To avoid trivial cases, we assume $0 < \rho_0 + \rho_1 < 1, \rho_0 > 0$ (some chance we will have more than one offspring). This implies $0 < P(E_n) < 1$ and $0 < \eta \le 1$.

Theorem 3.7.1 - If $E(X_1) < \infty$, then $\Pi_{0,j} = \lim_{n\to\infty} P_{0,j}(n)$ exists. The generating function $G(s) = \Sigma_{j=0}^\infty \Pi_{0,j} s^j$ satisfies (3.7.1) $G(\eta^{-1}\phi(s\eta)) = \alpha G(s) + 1 - \alpha, \alpha = \phi'(\eta)$. If $\mu = E(X_1) \le 1$, then $\eta \equiv 1$ and $\alpha = \mu$, so (3.7.2) $G(\phi(s)) = \mu G(s) + 1 - \mu$ for all $\mu, G'(\eta) \le 1$ and $G'(\eta) = 1 \iff \mu = 1$. See outline of proof in notes.

Theorem 3.7.2 - (1) If $E(X_1) \neq 1, \Sigma_j \Pi_{0,j} = 1$, (2) If $E(X_1) = 1, \Pi_{0,j} = 0$ for all j.

3.8 Review (in notes)

4 Continuous Time Markov Chains

Continuous time chains stay in each state a random time that is a continuous random variable that may depend on the state. $X(t)$ = state at time $t, t \geq 0$. $X(t)$ may or may not be a Markov process.

4.1 The Poisson Process

Example 4.1.1 - We use a Geiger counter to observe the emission of particles from a radioactive source. If we switch on the counter at time zero, the count $N(t)$ is the outcome of an apparently random process. Observations: (a) $N(0) = 0, N(t) \in \{0, 1, 2, 3, \dots\}$, (b) If $s < t, N(s) \leq N(t)$. We conjecture a continuity assumption: in a time period $(t, t+h)$ the probability of an emission is proportional to h for small h.

Definition 4.1.1 - A Poisson process with intensity λ is a process $N = \{N(t), t \geq 0\}$ taking values in $S = \{0, 1, 2, 3, \dots\}$ such that (a) $N(0) = 0$, (b) $S < t \Rightarrow N(s) \leq N(t)$, (c) $P(N(t+h) = n+m | N(t) = n) = \{\lambda h + O(h)$ for $m = 1, O(h)$ for $m > 1, 1 - \lambda h + O(h)$ for $m = 0$. $O(h)$ means an expression $A(h)$ such that $\lim_{h \to 0} \frac{|A(h)|}{h} \to 0$, (d) If $s < t$, the number $N(t) - N(s)$ of emissions in $(s, t]$ is independent of the times of emissions in $(0, s]$ (this is pretty much the Markov condition).

Definition 4.1.2 - $N(t)$ = the number of arrivals or occurences or events or emissions at time t. N is called a counting process.

Theorem 4.1.1 - $N(t)$ has the Poisson distribution with parameter λt, i.e., (4.1.1) $P(N(t) = j) = \frac{(\lambda t)^j}{j!} e^{-\lambda t}, j = 0, 1, 2, \dots$.

Proof: We condition $N(t+h)$ on $N(t)$. $P(N(t+h) = j) = \Sigma_i P(N(t) = i) P(N(t+h) = j | N(t) = j) = \Sigma_i P(N(t) = i) P((j-i)$ arrivals in $(t, t+h]) = P(N+t) = j-1) P(1$ arrival in $(t, t+h =) + P(N(t) = j) P($no arrivals in $(t, t+h]) + O(h)$. Set $P_j(t) = P(N(t) = j), P_j(t+h) = \lambda h P_{j-1}(t) + (1-\lambda h) P_j(t) + O(h), P_0(t+h) = (1-\lambda h) P_0(t) + O(h), j \neq 0$. Assuming $P_j(t)$ is a smooth function, we subtract $P_j(t)$ from each side, divide by h and let $h \downarrow 0$. $\lim_{h \to 0} \frac{P_j(t+h) - P_j(t)}{h}$, etc. (4.1.2) $P_j'(t) = \lambda P_{j-1}(t) - \lambda P_j(t), j \neq 0$. (4.1.3) $P_0'(t) = -\lambda P_0(t)$. The initial condition is (4.1.4) $P_j(0) = \delta_{j0} = \{1$ for $j = 0, 0$ for $j \neq 0$. This is a big system of equations that we need to solve to go further.

Two approaches: (4.1.3) + (4.1.4) together yield $P_0(t) = e^{-\lambda t}$. We substitute into (4.1.2) with $j = 1, P_1'(t) = \lambda e^{-\lambda t} - \lambda P_1(t). P_1'(t) + \lambda P_1(t) = \lambda e^{-\lambda t}, e^{\lambda t} P_1'(t) + e^{\lambda t} \lambda P_1(t) = e^{\lambda t} \lambda e^{-\lambda t} \Rightarrow \frac{d}{dt}(e^{\lambda t} P_1(t) = \lambda, e^{\lambda t} P_1(t) = \lambda t + c \Rightarrow P_1(t) = \lambda t e^{-\lambda t}$. Iteration yields $P_j(t) = \frac{(\lambda t)^j}{j!} e^{-\lambda t}$.

Second approach: Define a generating function $G(s, t) = \Sigma_{j=0}^{\infty} P_j(t) s^j$. Multiply (4.1.2) by s^j and sum (some details) $\Rightarrow \frac{\nabla G}{\nabla t} = \lambda(s-1) G, G(s, 0) = 1$. The solution is (4.1.5) $G(s, t) = e^{\lambda(s-1)t} = e^{-\lambda t} \Sigma_{j=0}^{\infty} \frac{(\lambda t)^j}{j!} s^j$. Reformulation of the process: important for computational purposes.

Definition 4.1.3 - Let $T_0, T_1, T_2, \dots$ be given by (4.1.6) $T_0 = 0, T_n = \inf_t \{N(t) = n\}$ (inf = min). T_n is the arrival or waiting time for the nth event. The interarrival or sojourn times $X_1, X_2, \dots$ are given by (4.1.7) $X_n = T_n - T_{n-1}$. If we know N, we can compute $X_1, X_2, \dots$. Vice versa if we know the entire collection of sojourn times $\{X_n\}$ then (4.1.8) $T_n = \Sigma_{i=1}^n X_i, N(t) = \max_{T_n \leq t} n$. See picture in notes.

Theorem 4.1.2 - The random variables $X_1, X_2, \dots$ are i.i.d. with exponential distribution with parameter λ.

4.2 Birth Processes

Definition 4.2.1 - $N(t)$ (continuous-time birth process). Assume: (a) $N(0) \geq 0$, (b) $s < t \Rightarrow N(s) \leq N(t)$, (c) $P(N(t+h) = n+m | N(t) = m) = \{\lambda_n h + O(h)$ for $m = 1, O(h)$ for $m > 1, 1 - \lambda_n h + O(h)$ for $m = 0$, (d) If $s < t$ then (conditional on the value of $N(s)$, the increment $N(t) - N(s)$ is independent of the times of arrivals prior to s.

Transition probabilities: $P_{ij}(t) = P(N(t) = j | N(0) = i) = P(N(s+t) = j | N(s) = i), s \leq t$.

Forward System - Assume $\lambda_{-1} = 0, P_{ij}(0) = \delta_{ij}$, then (4.2.1) $P_{ij}'(t) = \lambda_{j-1} P_{ij-1}(t) - \lambda_j P_{ij}(t)$.

Backward System - (4.2.2) $P'_{ij}(t) = \lambda_i, P_{i+1j}(t) - \lambda_i P_{ij}(t), j \geq i$. (Yes, that is P_{i+1j}. I asked in class and verified it. I don't understand why...)

Theorem 4.2.3 - The forward system has a unique solution that also satisfies the backward system. Proof: First note that (4.2.3) $P_{ij}(t) = 0$ if $j < i$. We solve the forward problem with $j = i$, so $P'_{ii}(t) = \lambda_{i-1}P_{ii-1}(t) - \lambda_i P_{ii}(t)$ (first term after equals goes to zero), so (4.2.4) $P_{ii}(t) = e^{-\lambda_i t}$. We substitute into the forward equation with $j = i+1$ to find that $P_{ii+1}(t)$ exists (using standard ODE theory). By induction, we conclude the solution of the forward system exists and is unique. We use the Laplace transform: $\widehat{P_{ij}}(\theta) = \int_0^\infty e^{-\theta t} P_{ij}(t)dt, P_{ij}(t)\widehat{\rightarrow} = \widehat{P_{ij}}(\theta)$. This transforsm derivatives with respect to t to products in the θ variable domain. If we transform both sides of the forward equation, we get $(\theta + \lambda_j)\widehat{P_{ij}}(\theta) = \delta_{ij} + \lambda_{j-1}\widehat{P_{ij-1}}(\theta)$. This difference equation can be solved: (4.2.5) $\widehat{P_{ij}}(\theta) = \frac{1}{\lambda_j}\frac{\lambda_i}{\theta+\lambda_i}\frac{\lambda_{i+1}}{\theta+\lambda_{i+1}}\cdots\frac{\lambda_j}{\theta+\lambda_j}, j \geq i$. Using the inverse Laplace transform gives $P_{ij}(t)$.

To show the claim about the backward equations, we also take the Laplace transform in the same way but now with the backward equation to find that any solution (call it Π—it may be P, may not be) with $\widehat{\Pi_{ij}}(\theta) = \int_0^\infty e^{-\theta t}\Pi_{ij}(t)dt$ satisfies $(\theta + \lambda_j)\widehat{\Pi_{ij}}(\theta) = \delta_{ij} + \lambda_i\widehat{\Pi_{i+1j}}(\theta)$. Note that $\widehat{P}_{ij}$ satisfies this equation, but so can other functions. (TeX note: the widehat change here is meaningless.)

Theorem 4.2.4 - If $\{P_{ij}(t)\}$ is the unique solution of the forward system then any solution $\{\Pi_{ij}\}$ of the backwards system satisfies $P_{ij}(t) \leq \Pi_{ij}(t)$ for all t. Proof not given.

Observe: if (4.2.6) $\Sigma_j P_{iJ}(t) \equiv 1$ (for all t), then Thm 4.2.4 would imply that $\{P_{ij}\}$ is the unique solution of the backward system that is a probability distribution. However (4.2.6) may not hold.

Definition 4.2.5 - An explosion occurs if the birth rates λ_n increase sufficiently quickly that there is a positive probability that the process N can pass through all finite states in finite (bounded) time.

Definition 4.2.6 - Let $T_\infty = \lim_{n\to\infty} T_n$ be the limit of the arrival times. We say N is honest if $P(T_\infty = \infty) = 1$ and dishonest otherwise.

Theorem 4.2.5 - (4.2.6) holds (for each $i\{P_{ij}(t)\}$ is a probability distribution in j) if and only if N is honest. Proof: (4.2.6) is equivalent to $P(T_\infty > t) = 1$, any t, why? Exercise.

Theorem 4.2.6 - N is honest $\iff \Sigma_n\lambda_n^{-1} = \infty$. This says that if the birth rates are sufficiently small (increase sufficiently slowly) then N is honest. $\Sigma a_n = \infty, a_n$ decreases sufficiently slowly, e.g. $\Sigma_n \frac{1}{n} = \infty$. If the λ_n increase sufficiently quickly that $\Sigma_n\lambda_n^{-1}$ converges, then N is dishonest. We can think of the deficit as $1 - \Sigma_j P_{ij}(t)$ as the probability $P(T_\infty \leq t)$ of escaping to infinity at time t starting from state i.

Theorem 4.2.6 follows from:

Theorem 4.2.7 - Let $X_1, X_2, \ldots$ be independent random variables with X_n having the exponential distribution with parameter λ_{n-1} and let $T_\infty = \Sigma_n X_n$. Then $P(T_\infty < \infty) = \{0$ for $\Sigma_n\lambda_n^{-1} = \infty, 1$ for $\Sigma_n\lambda_n^{-1} < \infty$. Proof in notes. When the rates vary, the situation becomes more complicated.

Last topic in this section: what does the condition (d) mean? Recall that a sequence of random variables $\{X_n, n \geq 0\}$ satisfies the Markov property if, conditional on the event $\{X_n = i\}$ events related to the collection $\{X_m, m > n\}$ are independent of events related to the collection $\{X_m, m < n\}$.

Theorem 4.2.8 - Weak Markov Property - Let $N(t)$ be a birth process and T a fixed time. Conditional on the event $\{N(T) = i\}$ the evolution of the process after T is independent of the evolution before T. Proof: definition 4.2.1(a).

Theorem 4.2.8 - Weak Markov Property - $N(t)$ is a birth process. T is a fixed time. Conditional on the event $\{N(T) = i\}$ the evolution of the process after time T is independent of the evolution prior to T. It is useful to let T be a random variable. This theorem as stated cannot hold for all possilbe random times T. For example, if T "looks into the future" the conclusion cannot hold.

Definition 4.2.7 - A random time T is a stopping time for the process $N(t)$ if for all $t \geq 0$, the indicator function of the event $\{T \leq t\}$ is a function of the values $\{N(s), s \leq t\}$ of the process up to time t. We can decide if T has occured by t by examining the values of N up to time t.

Example 4.2.5 - The arrival times $T_1, \ldots,$ are stopping times. $I_{T_n=t} = \{1$ for $N(s) \neq n, N(t) = n, s < t, 0$ else. For fixed k, the kth time a process N visits a state n is a stopping time. The last time a process visits a state n is not a stopping time.

Example 4.2.6 - $T_4 - 2$ and $\frac{1}{2}(T_1 + T_2)$ are not stopping times (why?).

For stopping times we can prove:

Theorem 4.2.9 - Strong Markov Property - Let $N(t)$ be a birth process and T a stopping time. Let A be an event that depends on $\{N(s), s > t\}$ and B an event that depends on $\{N(s), s < t\}$. Then (4.2.9)

$P(A|N\Pi) = i, B) = P(A|N(T) = i)$. Read outline of proof in notes.

4.3 More on Poisson Processes

While Poisson processes are special among continuous time processes, they also occur frequently in "nature". Part of the reason is the Law of Rare Events.

Law of Rare Events - This says that in a situation in which a certain event can occur in any of a large number of possibilities but where the probability it occurs in any given possibility is small, then the total number of occurrences follows approximately a Poisson distribution.

Consider a large number N of Bernoulli trials. The probability of success in each trial is p. Let $X_{N,p} =$ number of successes in the N trials. (4.3.1) $P(X_{N,p} = k) = \frac{N!}{k!(N-k)!}p^k(1-p)^{N-k}, k = 0, 1, \ldots, N$. We consider the limit as $N \to \infty$ and $p \to 0$ so that $\mu = Np$ is fixed. $P(X_{N,p} = k) = N(N-1)\cdots(N-k+1)\frac{p^k(1-p)^N}{k!(1-p)^k}$. Substitute $p = \frac{\mu}{N} \Rightarrow P(X_{N,p} = k) = 1 * (1-\frac{1}{N})\cdots(1-\frac{k-1}{N})\frac{\mu^k(1-\frac{\mu}{N})^N}{k!(1-\frac{\mu}{N})^k}$. So $N \to \infty (p \to 0).1 * (1-\frac{1}{N})\cdots(1-\frac{k-1}{N}) \to 1, (1-\frac{\mu}{N})^N \to e^{-\mu}, (1-\frac{\mu}{N}^k) \to 1$.

Theorem 4.3.1 - Law of Rare Events - If $X_{N,p}$ is the number of successes in N Bernoulli trials with probability of success p, then (4.3.2) $\lim_{N\to\infty, pN=\mu} P(X_{N,p} = k) = \frac{\mu^k e^{-\mu}}{k!}$. (Poisson distribution with parameter μ.) (4.3.2) can be used to compute Binomial probabilities.

Suppose the probability of success varies with each trial. Let $Y_1, Y_2, \ldots,$ be independent Bernoulli r.v. with $P(Y_i = 1) = P_i, P(Y_i = 0) = 1 - P_i$. $S_n = Y_1 + \cdots + Y_N =$ number of successes in N trials. $P(S_N = k) = \Sigma_{(k)}\Pi_{i=1}^N P_i^{Y_i}(1-P_i)^{N-Y_i}$ where $\Sigma_{(k)}$ is the sum over all 0, 1 valued Y_is that sum $Y_1 + \cdots + Y_N = k$.

Theorem 4.3.2 - Law of Rare Events, v.2 - (4.3.3) $|P(S_N = k) - \frac{\mu^k e^{-\mu}}{k!}| \le \Sigma_{i=1}^N P_i^2$ with $\mu = P_1 + \cdots + P_N$. If the probabilities of success in the trials are small, we obtain a good approximation. Proof: see text, pg. 285.

Recall Theorem 4.1.1 - Let $N(t)$ be a Poisson process. Then (4.1.1) $P(N(t) = j) = \frac{(\lambda t)^j}{j!}e^{-\lambda t}, j = 0, 1, \ldots$ ($\lambda =$ rate of the process).

Alternate Proof (using the Law of Rare Events. Divide $[0, t]$ in to n subintervals of equal length $h = \frac{t}{n}$, setting $Y_i = \{1$ if there is an event in $((i-1)\frac{t}{n}, i\frac{t}{n}], 0$ otherwise. $S_n = Y_1 + \cdots + Y_n$ counts the number of subintervals that contain at least one event and $P_i = P(Y_i = 1) = \lambda\frac{t}{n} + o(\frac{t}{n})$ by definition of a Poisson process. Using (4.3.) $|P(S_n = k) - \frac{\mu^k e^{-\mu}}{k!}| \le n(\frac{\lambda t}{n} + o(\frac{t}{n}))^2 = \frac{(\lambda t)^2}{n} + 2\lambda + o(\frac{t}{n}) + no((\frac{t}{n})^2)$. Now $no(\frac{t}{n}) = t\frac{o(\frac{t}{n})}{\frac{t}{n}} = t\frac{o(h)}{h} \to_{n\to\infty, h\to 0} 0$. S_n differs from $N(t)$ only if one of the subintervals has more than one event. $P(N(t) \ne S_n) = \Sigma_{i=1}^n P(\text{number of events in } ((i-1)\frac{t}{n}, i\frac{t}{n}] \ge 2) \le n * o(\frac{t}{n})$ by assumption. So $\lim_{n\to\infty} P(N(t) \ne S_n) = 0$.

Theorem 4.3.3 - The arrival time T_n has the gamma distribution with pdf (4.3.4) $f_{T_n}(t) = \frac{\lambda^n t^{n-1}}{(n-1)!}e^{-\lambda t}, n = 1, 2, \ldots, t \ge 0$. Proof: the event $\{T_n \le t\}$ occurs iff there are atleast n events in $(0, t]$. $N(t)$ is a Poisson distribution with parameter mean λt. So the cdf of T_n is $F_{T_n}(t) = P(T_n \le t) = P(N(t) \ge n) = \Sigma_{k=n}^\infty \frac{(\lambda t)^k e^{-\lambda t}}{k!} = 1 - \Sigma_{k=0}^{n-1}\frac{(\lambda t)^k e^{-\lambda t}}{k!}$. Differentiating, [some long messy computation, see notes] $= \frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}$.

Theorem 4.3.4 - For $0 < s < t, 0 \le k \le n$, (4.3.5) $P(N(s) = k|N(t) = n) = \frac{n!}{k!(n-k)!}(\frac{s}{t})^k(1-\frac{s}{t})^{n-k}$. Proof in notes.

The next result says that conditioned on a fixed total number of events in an interval, the times of occurence of the events is uniformly distributed in a certain way.

Related computation: consider an interval $(0, t]$. Choose a fixed number n of points uniformly distributed on the interval. Positions $= \{U_1, U_2, \ldots, U_n\}$. The pdf is $f(s) = \{\frac{1}{t}$ for $0 \le s \le t, 0$ else. Set $\{W_1, W_2, \ldots, W_n\}$ be the positions $\{U_1, U_2, \ldots, U_n\}$ ordered from left to right. (See drawing in notes.) The joint pdf for $W_1, \ldots, W_n$ is (4.3.6) $f_{W_1,W_2,\ldots,W_n}(w_1, w_2, \ldots, w_n) = n!t^{-n}, 0 < w_1 < \cdots < w_n \le t$. Arguing by induction, $f_{W_1,W_2}(w_1, w_2)\Delta w_1\Delta w_2 = P(w_1 < U_1 \le w_1 + \Delta w_1, w_2 \le U_2 \le w_2 + \Delta w_2) + P(w_1 < U_2 \le w_1 + \Delta w_1, w_2 \le u_2 \le w_2 + \Delta w_2) = 2(\frac{\Delta w_1}{t})(\frac{\Delta w_2}{t}) = 2t^{-2}\Delta w_1\Delta w_2$. Divide by $\Delta w_1\Delta w_2$, take the limit as $\Delta w_1\Delta w_2 \to 0$, then (4.3.6) is proved. In general there are $n!$ arrangements of the U_is.

Theorem 4.3.5 - Let $T_1, \ldots, T_n$ be arrival times for a Poisson process with rate $\lambda > 0$. Conditional on $N(t) = n$, the random variables $T_1, \ldots, T_n$ have joint pdf (4.3.7) $f_{T_1,\ldots,T_n}(t_1, \ldots, t_n) = n!t^{-n}, 0 < t_1 < t_2 < \cdots < t_n \le t$. Proof: We create some subintervals from $(0, t]$, using a set of times $t_1, \ldots, t_n$ in increments $\Delta t_1, dots, \Delta t_n$. (See drawing in notes.) The event $\{t_i < T_i < t_i + \Delta t_i, i = 1, 2, \ldots, n, N(t) = n\}$. This

means no events in $(0,t_1),(t_1+\Delta t_1,t_2),\ldots,(t_n+\Delta t_n,t]$ and exactly one event in each interval $(t_1,t_1+\Delta t_1],(t_2,t_2+\Delta t_2],\ldots,(t_n,t_n+\Delta t_n]$. The intervals are disjoint and $P(N(t_1)=0,\ldots,N(t)-N(t_n+\Delta t_n)=0)=e^{-\lambda t_1}e^{-\lambda(t_2-t_1-\Delta t_1)}\cdots e^{-\lambda(t_n-t_{n-1}-\Delta t_{n-1})}\cdots e^{-\lambda(t-t_n-\Delta t_n)}=e^{-\lambda t}(1+o(\max_i\Delta t_i)), P(N(t_1+\Delta t_1)-N(t_1)=1,\ldots,N(t_n+\Delta t_n)-N(t_n)=1)=\lambda(\Delta t_1)\lambda(\Delta t_2)\cdots\lambda(\Delta t_n)(1+o(\max_i\Delta t_i))$ [$o(\max_i\Delta t_i)$ goes to zero faster than $C\Delta t_i$]. Hence, more stuff in notes. Get to $=n!t^{-n}\Delta t_1\cdots\Delta t_n(1+o(\max_i\Delta t_i))$. Divide by $\Delta t_1\cdots\Delta t_n$, take limit as $\Delta t_1\cdots\Delta t_n\to 0$.

Example 4.3.2 - Customers arrive at a facility according to a Poisson process with rate λ. Each customer pays \$1 when they arrive. We want to evaluate (compute) the expected value of the total sum collected during some interval $(0,t]$ when the amounts are "discounted" back to time 0 at a rate β: If a customer pays \$1 at event time T_k, the discounted amount is $e^{-\beta T_k}*\$1$. $M=E(\Sigma_{k=1}^{N(t)}e^{-\beta T_k})$. We condition on $N(t)=n$!. $M=\Sigma_{n=1}^{\infty}E(\Sigma_{k=1}^{n}e^{-\beta T_k}|N(t)=n)P(N(t)=n)$. Let $U_1,\ldots,U_n$ be uniformly distributed iid variables in $(0,t]$. By theorem 4.3.5, $E(\Sigma_{k=1}^{n}e^{-\beta T_k}|N(t)=n)=E(\Sigma_{k=1}^{n}e^{-\beta U_k})$. Each term is treated in the same way: $=nE(e^{-\beta U_1})=nt^{-1}\int_0^t e^{-\beta u}du=\frac{n}{\beta t}(1-e^{-\beta t}), M=\frac{1}{\beta t}(1-e^{-\beta t}\Sigma_{n=1}^{\infty}nP(N(t)=n)=\frac{\lambda}{\beta}(1-e^{-\beta t})$.

The proof of the strong Markov property, theorem 4.2.9 for birth processes, uses the weak Markov property and temporal homogeneity. The strong Markov proprety is very important for the analysis of continuous processes. When applied to a birth process, it implies that the new process: $\widetilde{N}(t)=N(t+T)-N(t), t>0$, conditional on $N(T)=i$, is also a birth process when T is a stopping time. Exercise: show the rates are $\lambda_i,\lambda_{i+1},\ldots$.

Definition 4.3.1 - A Poisson process N has stationary, independent increments in the sense that (1) the distribution of $N(t)-N(s)$ depends only on $t-s$, (2) any finite set of increments $\{N(t_i)-N(s_i), i=1,2,\ldots,n\}$ are independent if $0<s_1\le t_1\le s_2\le\cdots\le t_n$. (Exercise: verify.)

Theorem 4.3.6 - Suppose that $M(t), t\ge 0$, is a nondecreasing, right-continuous, integer valued stochastic process with (1) $M(0)=0$, (2) stationary independent increments, (3) M only has jump discontinuities of size 1. So M is a Poisson process. Outline of proof: for $u,v\ge 0, E(M(u+v))=E(M(u))+E(M(u+v)-M(u))=E(M(u))+E(M(v))$ [stationary increments]. $E(M(u))$ is nondecreasing in u, so there is a λ such that (4.3.8) $E(M(u))=\lambda u, u\ge 0$. Let $T=\sup\{t: M(t)=0\}$ be the time of the first jump (sup is like max). By right continuity almost surely, $M(T)=1$ and T is a stopping time for M (check!). Now (4.3.9) $E(M(s))=E(E(M(s)|T)), E(M(s)|T)=0, s<T$. For $s\ge T$, (4.3.10) $E(M(s)|T=t)=E(M(t)|T=t)+E(M(s)-M(t)|T=t)=1+E(M(s-t))$. If $F(t)$ is the distribution function for T, then (4.3.10) into (4.3.9) yields $E(M(s))=\int_0^s(1+E(M(s-t))dF(t)$. $E(M(s))=\lambda s\Rightarrow$ (4.3.11) $\lambda s=F(s)+\lambda\int_0^s(s-t)dF(t)$. This is called an integral equation for the unknown function F. This could be solved with Laplace transforms, $F(t)=1-e^{-\lambda t}, t\ge 0$. T has the exponential distribution. We can make the same argument used for sojourn times to argue the inter-event times are independent and exponentially distributed. Conclude that M is a Poisson process with rate λ.

4.4 The Death Process

This runs through states $N,N-1,\ldots,0$, where it is absorbed (extinction).

Definition 4.4.1 - A death process $X(t)$ with death parameters $\mu_1,\ldots,\mu_N,\mu_i>0$, is a process with state space $S=\{0,1,\ldots,N\}$ such that (1) $X(0)=N$, (2) $s<t\Rightarrow X(s)\ge X(t)$, (3) $P(X(t+h)=k-m|X(t)=k)=\{\mu_k h+o(h)$ for $m=1, 1-\mu_k h+o(h)$ for $m=0, o(h)$ for $m\ge 2$, (4) if $s<t$, then conditional on $X(s)$, the increment $X(t)-X(s)$ is independent of all deaths prior to s.

Theorem 4.4.1 - The sojourn time in state k is exponentially distributed with parameter μ_k and the sojourn times are independent. Proof: HW. (See notes for graph of realization.)

Theorem 4.4.2 - If the $\{\mu_i\}$ are distinct, then the transition probabilities $P_n(t)=P(X(t)=n|X(0)=N)$ satisfy $P_N(t)=e^{-\mu_N t}$, $P_n(t)=\mu_{n+1}\ldots\mu_N(A_{n,n}e^{-\mu_n t}+\cdots+A_{N,n}e^{-\mu_N t})$ where $A_{k,n}=\frac{1}{(\mu_N-\mu_k)(\mu_{N-1}-\mu_k)\cdots(\mu_{k+1}-\mu_k)(\mu_{k-1}-\mu_k)\cdots(\mu_1-\mu_k)}$ [note that $(\mu_k-\mu_k)$ is skipped].

Example 4.4.1 - Linear death process - We assume $\mu_k=\alpha k, \alpha=$ individual death rate. Can compute $A_{k,n}$, and $P_n(t)=\frac{N!}{n!(N-n)!}e^{-n\alpha t}(1-e^{-\alpha t})^{N-n}, n=0,1,\ldots,N$. $T=$ time of extinction $=\min_{t\ge 0}X(t)=0, T\le t, X(t)=0$. cdf of T: $F_T(t)=P(T\le t)=P(X(t)=0)=P_0(t)=(1-e^{-\alpha t})^N, t\ge 0$.

Example 4.4.2 - We consider the failure of a cable made up of parallel fibers under tension. The model is based on what happens with one fiber. The cable is required to hold 1000Kg for 100 years. How many fibers are required? A single fiber is subject to failure after some time when placed under load. (See notes

for graph.) log(load) vs log(time) is linear with slope $-\beta = -40$. $log_{10}(\mu_T) = 2 - 40 log_{10}(I), \mu_T =$ mean life, $I =$ load. Naive (wrong) analysis: Rationalization using average properties $\Rightarrow$ 100 year design for 1000Kg requires about 1000 fibers. (This is wrong.) Because of variation around the mean properties. Based on observation, the probability distribution for the failure time T of a single fiber under a time varying load $l(t) : P(T \leq t) = 1 - e^{-\int_0^t K(l(s))ds}, r(t) = K(l(t)) =$ failure rate, which determines the probability that a fiber under load $l(t)$ will fail in $(t, t + \Delta t]$ with probability $P(t < T < t + \Delta t | T > t) = K(l(t))\Delta t + o(\Delta t)$. Assume a power law relation $K(l) = l^\beta / A, A, \beta > 0$ constant. Under a constant load, $l(t) = l$, the single fiber failure time is exponentially distributed with mean $\mu_T = E(T|l) = \frac{1}{K(l)} = Al^{-\beta} \Rightarrow$ gives the linear log-log plot above. We put N fibers in a cable, and subject the cable to lead NL, where L is the load [initially] per fiber. The cable failure time is the failure of the last fiber. The number of intact fibers $X(t)$ is a death process with $\mu_k = kK(NL/k), k = 1, 2, \ldots, N$. (See argument.) Cable failure time si W_N. Using (4.4.1) $P(W_N \leq t) = P_0(t)$, (4.4.2) $E(W_N) = AL^{-\beta}\Sigma_{k=1}^N (\frac{k}{N})^{\beta-1}\frac{1}{N} = E(W_N) \approx \frac{A}{\beta L^\beta}$. The average fiber life is $\frac{A}{L^\beta}$. If we use 1000 strands, expected life is 2.5 years. To last 100 years, (see notes) we need 1097 fibers.

4.5 Birth-Death Processes

We let $X(t)$ increase and decrease. If at time t, the process is in state n, after some random sojourn time it can move to either $n - 1$ or $n + 1$. This is a continuous time extension of a random walk.

Definition 4.5.1 - A birth-death process $X(t)$ is a Markov process on the state space $S = \{0, 1, 2, 3, \ldots\}$ with stationary transition probabilities described by the transition probability function (4.5.1) $P_{ij}(t) = P(X(t + s) = j | X(s) = i), s \geq 0, t \geq 0$ where (4.5.2) $P_{i,i+1}(h) = \lambda_i h + o(h), h \downarrow 0, i \geq 0$. (4.5.3) $P_{i,i-1}(h) = \mu_i h + o(h), h \downarrow 0, i \geq 0$. (4.5.4) $P_{ii}(h) = (1 - (\lambda_i + \mu_i)h) + o(h), h \downarrow 0, i \geq 0$. (4.5.5) $P_{ij}(0) = \delta_{ij}$. (4.5.6) $\mu_0 = 0, \lambda_0 > 0, \mu_i, \lambda_i > 0, i = 1, 2, \ldots$. The matrix (not the probability transition matrix), called the infinitesimal generator:

$$A = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & \ldots & & \\ \mu_1 & -(\mu_1 + \lambda_1) & \lambda_1 & 0 & \ldots & \\ 0 & \mu_2 & -(\mu_2 + \lambda_2) & \lambda_2 & 0 & \ldots \\ \ddots & & & & & \end{pmatrix}$$

Recall from ODEs: $\dot{y} = Ay, y = y_0 e^{At}$. Since the $P_{ij}(t)$ are probabilities, (4.5.7) $P_{ij}(t) \geq 0, \Sigma_{j=0}^\infty P_{ij}(t) \leq 1$. We can also show:

Theorem 4.5.1 - Chapman-Kolmogorov - (4.5.8) $P_{ij}(t + s) = \Sigma_{k=0}^\infty P_{ik}(t) P_{kj}(s)$.

Definition 4.5.2 - Assume $q_i = P(X(0) = i), i = 0, 1, 2, \ldots$.

Theorem 4.5.2 - (4.5.9) $P(X(t) = n) = \Sigma_{i=0}^\infty q_i P_{in}(t)$.

We now calculate the distribution of the sojourn times $\{S_i\}, S_i =$ sojourn time of $X(t)$ in state i. We set $G_i(t) = P(S_i \geq t)$. By the Markov property, as $h \downarrow 0, G_i(t + h) = G_i(t)G_i(h) = G_i(t)(P_{ii}(h) + o(h)) = G_i(t)(1 - (\lambda_i + \mu_i)h + o(h) \Rightarrow \frac{G_i(t+h) - G_i(t)}{h} = -(\lambda_i + \mu_i)G_i(t) + o(1), h \to 0 \Rightarrow$ (4.5.10) $G_i'(t) = -(\lambda_i + \mu_i)G_i(t), G_i(0) = 1 \Rightarrow G_i(t) = e^{-(\lambda_i + \mu_i)t}$.

Theorem 4.5.3 - The sojourn times are exponentially distributed with parameter $(\lambda_i + \mu_i)^{-1}$.

We use another intuitive fact:

Theorem 4.5.4 - Given a transition occurs at time t, the probability that the transition is $i \to i + 1$ is $\frac{\lambda_i}{\mu_i + \lambda_i}$ and $i \to i - 1$ is $\frac{\mu_i}{\lambda_i + \mu_i}$.

Picture: $X(t)$ sojourns in state i for a time exponentially distributed with parameter $(\lambda_i + \mu_i)$, then enters state $i + 1$ with probability $\frac{\lambda_i}{\lambda_i + \mu_i}$ or state $i - 1$ with probability $\frac{\mu_i}{\lambda_i + \mu_i}$. This suggests a method to compute realizations. (1) Choose t_i from an exponential distribution with parameter $(\lambda_i + \mu_i)$ ($t_i =$ sojourn time in state i). (2) Toss a coin with probability of heads $p_i = \frac{\lambda_i}{\lambda_i + \mu_i}$, if we get heads move to state $i + 1$, otherwise move to state $i - 1$. Does this give a valid realization? It turns out that we can show this works. However, there are actually several stochastic processes that satisfy (4.5.2) - (4.5.6), (4.5.7), (4.5.8)! There are several Markov processes with the same infinitesimal generator.

Theorem 4.5.5 - For birth-death processes with $\lambda_0 > 0$, a sufficient condition that there is a unique Markov process with transition probability function $P_{ij}(t)$ satisfying (4.5.7) and (4.5.8) is (1) (4.5.10) $\Sigma_{n=0}^\infty \frac{1}{\lambda_n \theta_n} \Sigma_{k=0}^n \theta_k = \infty$ where $\theta_0 = 1, \theta_n = \frac{\lambda_0 \lambda_1 \cdots \lambda_{n-1}}{\mu_1 \mu_2 \cdots \mu_n}, n = 1, 2, 3, \ldots$. As with pure birth processes, $P_{ij}(t)$

satisfy some differential equations. Starting from (4.5.8) $P_{ij}(t+h) = \Sigma_k P_{ik}(h)P_{kj}(t)$. Using (4.5.2) - (4.5.4), $\Sigma_{k\neq i-1,i,i+1}P_{ik}(t)P_{kj}(t) \leq \Sigma_{k\neq i-1,i,i+1}P_{ik}(h) \leq 1-(P_{ii}(h)+P_{i,i-1}(h)+P_{i,i+1}(h))+o(h) = o(h) \Rightarrow P_{ij}(t+h) = \mu_i h P_{i-1,j}(t) + (1-(\lambda_i+\mu_i)h)P_{ij}(t) + \lambda_i h P_{i+1,j}(t) + o(h)$.

Theorem 4.5.6 - Under appropriate assumptions, $P_{ij}(t)$ satisfy the backward Kolmogorov equations. (4.5.12) =

$$\left\{\begin{array}{l} P'_{0j}(t) = -\lambda_0 P_{0j}(t) + \lambda_0 P_{1j}(t) \\ P'_{ij}(t) = \mu_i P_{i-1,j}(t) - (\lambda_i+\mu_i)P_{ij}(t) + \lambda_i P_{i+1,j}(t) \\ P_{ij}(0) = \delta_{ij} \end{array}\right\}, i \geq 0$$

If instead, we decompose $(0,t+h) \to (0,t),(t,t+h)$ and we assume in addition that (4.5.13) $\frac{P_{kj}(h)}{h} = o(1), k \neq j-1, j, j+1$ where $o(1)$ means tends to zero as $h \to 0$ and is uniformly bounded with respect to k for fixed j so we can prove $\Sigma_{k\neq i-1,i,i+1}P_{ik}(t)P_{kj}(h) = o(h)$.

Theorem 4.5.7 - Under appropriate assumptions, $P_{ij}(t)$ satisfy the forward Kolmogorov equations, (4.5.14) =

$$\left\{\begin{array}{l} P'_{i0}(t) = -\lambda_0 P_{i0}(t) + mu_1 P_{i1}(t) \\ P'_{ij}(t) = \lambda_{j-1} P_{i,j-1}(t) - (\lambda_j+\mu_j)P_{ij}(t) + \mu_{j+1} P_{i,j+1}(t) \\ P_{ij}(0) = \delta_{ij} \end{array}\right\}, j \geq 1$$

Example 4.5.1 - A linear growth process has $\lambda_n = \lambda \cdot n + a, \lambda > 0, \mu_n = \mu \cdot n, \mu > 0$. a describes infinitesimal "immigration". Rates grow linearly in population (λ, μ). Forward equations: $P'_{i0}(t) = -aP_{i0}(t) + \mu P_{i1}(t), P'_{ij}(t) = (\lambda(j-1)+a)P_{i,j-1}(t) - ((\lambda+\mu)j+1)P_{ij}(t) + \mu(j+1)P_{i,j+1}(t), j \geq 1$. If we multiply by j and sum, the expected value $m(t) = E(X(t)) = \Sigma_{j=1}^{\infty} jP_{ij}(t)$. (4.5.15) $= \{m'(t) = a + (\lambda-\mu)m(t), m(0) = i$ when $X(0) = i$. (4.5.16) $m(t) = at + i, \lambda = \mu$, (4.5.17) $m(t) = \frac{a}{\lambda-\mu}(e^{(\lambda-\mu)t}-1) + ie^{(\lambda-\mu)t}, \lambda \neq \mu$. So $m(t) \to \{\infty$ for $\lambda \geq \mu, \frac{a}{\mu-\lambda}$ for $\lambda < \mu$.

Example 4.5.2 - Consider a Markov Process where $S = \{0,1\}$, (on/off system). The infinitesimal matrix is $A = \binom{-\alpha,\alpha}{\beta,-\beta}$. The process varies between 0 and 1 with sojourn times in 0 iid with exponential distribution with parameter α and sojourn times in 1 iid with exponential distribution with parameter β. This is a finite-state birth-death process: $\lambda_0 = \alpha, \lambda_1 = 0, \mu_0 = 0, \mu_1 = \beta$. 1st Kol. equation: $P'_{00}(t) = -\alpha P_{00}(t) + \beta P_{01}(t) = \beta - (\alpha+\beta)P_{00}(t) \Rightarrow P_{00}(t) = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}e^{-(\alpha+\beta)t}, P_{01}(t) = \frac{\alpha}{\alpha+\beta} - \frac{\alpha}{\alpha+\beta}e^{-(\alpha+\beta)t}, P_{11}(t) = \frac{\alpha}{\alpha+\beta} + \frac{\beta}{\alpha+\beta}e^{-(\alpha+\beta)t}, P_{10}(t) = \ldots$. $\lim_{t\to\infty} P_{00}(t) = \lim_{t\to\infty} P_{10}(t) = \frac{\beta}{\alpha+\beta}, \lim_{t\to\infty} P_{11}(t) = \lim_{t\to\infty} P_{01}(t) = \frac{\alpha}{\alpha+\beta}$. There is a limiting distribution.

4.6 Limiting Behavior of Birth-Death Processes

General result:

Theorem 4.6.1 - If $X(t)$ is a birth-death process with no absorbing states. Then $\lim_{t\to\infty} P_{ij}(t) = \Pi_j \geq 0$ exists for all i. However, it is possible that $\Pi_j = 0$ for all j.

Definition 4.6.1 - If the $\{\Pi_j\}$ in Theorem 4.6.1 are strictly positive and satisfy $\Sigma_j \Pi_j = 1$, then $\{\Pi_j\}$ is the limiting probability distribution for the process.

Theorem 4.6.2 - If $X(t)$ is a birth-death process with no absorbing states and with limiting distribution $\{\Pi j\}$ then $\{\Pi j\}$ is also a stationary distribution: $\Pi_j = \Sigma_{i=0}^{\infty} \Pi_i P_{ij}(t), t \geq 0$. Read proof in notes.

There is an explicit formula for determining the limiting distribution.

Theorem 4.6.3 - Let $X(t)$ be a birth-death process with no absorbing states. Let (4.6.1) $\{\theta_0 = 1, \theta_j = \frac{\lambda_0\lambda_1\cdots\lambda_{j-1}}{\mu_1\mu_2\cdots\mu_j}, j \geq 1$. If $\Sigma_j\theta_j < \infty$, (4.6.3) $\{\Pi_0 = \frac{1}{\Sigma_{k=0}^{\infty}\theta_k}, \Pi_j = \theta_j\Pi_0 = \frac{\theta_j}{\Sigma_{k=0}^{\infty}\theta_k}$. If $\Sigma_j\theta_j = \infty, \Pi_j = 0$ for all j. Proof: Kol. forward equations: (4.6.3) $\{P'_{i0}(t) = -\lambda_0 P_{i0}(t) + \mu_1 P_{i1}(t), P'_{ij}(t) = \lambda_{j-1}P_{ij-1}(t) - (\lambda_j + \mu_j)P_{ij}(t) + \mu_{j+1}P_{ij+1}(t), j \geq 1, P_{ij}(0) = \delta_{ij}$. As $t \to \infty, P'_{i0}(t) \to$ limit, and $P'_{ij}(t) \to$ limit. $P_{ij}(t) \to$ constant $\Rightarrow \lim_{t\to\infty} P'_{ij}(t) = 0 \Longrightarrow 0 = -\lambda_0\Pi_0 + \mu_1\Pi_1, 0 = \lambda_{j-1}\Pi_{j-1} - (\lambda_j+\mu_j)\Pi_j + \mu_{j+1}\Pi_{j+1}, j \geq 1$. By induction $\Pi_0 = (\Sigma_k\theta_k)^{-1}$, and so on.

Example 4.6.1 - Linear growth with immigration, (birth) $\lambda_n = a + n \cdot \lambda$, (death) $\mu_n = n \cdot \mu$. If $\lambda < \mu$, the population mean converges $m(t) \to \frac{a}{\mu-\lambda}, t \to \infty$. We can compute as above when $\lambda < \mu, \theta_0 = 1, \theta_1 = \frac{a}{\mu}, \theta_2 = \cdots, \Rightarrow \Pi_0 = (1-\frac{\lambda}{\mu})^{a/\lambda}, \Pi_k = (\frac{\lambda}{\mu})^k \frac{(a/\lambda)(a/\lambda+1)\cdots(a/\lambda+k-1)}{k!}(1-\frac{\lambda}{\mu})^{a/k}, k > 1$.

Example 4.6.2 - Logistic process - $S = \{N, N+1, N+2, \ldots, M\}$. Details in notes.

5 Markov Chain Monte Carlo

Difficult Problem: Compute realizations of a random vector X whose component random variables are dependent. We describe a way to do this.

5.1 Background

We consider a finite state Markov chain $\{X_n\}$ with state space $S = \{1, 2, \ldots, N\}$, P_{ij} = transition probability matrix, X_0 = initial condition. We assume X_n is irreducible and aperiodic. We let $\{\Pi_j\}$ be the stationary distribution of X_n, (5.1.1) $\{\Pi = \Pi P, \Sigma_{j=1}^N \Pi_j = 1$, $\{\Pi_j\}$ is also the limiting distribution. (5.1.2) $\Pi_j = \lim_{n\to\infty} P(X_n = j), j = 1, \ldots, N$.

Sometimes the following method can be used to solve (5.1.1). Suppose there are positive numbers $\theta_j, j = 1, \ldots, N$ such that (5.1.2) $\{\theta_i P_{ij} = \theta_j P_{ji}, i \neq j, \Sigma_{j=1}^N \theta_j = 1$. Summing $\Sigma_{i=1}^N \theta_j P_{ij} = \theta_j \Sigma_{i=1}^N P_{ji} = \theta_j$. So $\theta_j = \Pi_{ji}, j = 1, \ldots, N$.

We also solve

Theorem 5.1.1 - Let h be a function on S, then with probability 1, (5.1.3) $\lim_{n\to\infty} \frac{1}{h} \Sigma_{i=1}^n h(X_j) = \Sigma_{j=1}^n \Pi_j h(j)$. Proof: $\frac{1}{n}\Sigma_{i=1}^n h(X_i)$ = expected value of $\{h(X_i)\}$. $P_j(n)$ = proportion of time the chain is in state j up to time n, then $\frac{1}{n}\Sigma_{i=1}^n h(X_i) = \Sigma_{j=1}^n h(j) P_j(n)$ as $n \to \infty, P_j(n) \to \Pi_j$.

A Markov chain is time reversible if $\Pi_i P_{ij} = \Pi_j P_{ji}, j \neq i$. If the initial state is chosen according to $\{\Pi_j\}$, then the sequence of states considered backwards is a Markov chain with transition probability matrix P_{ij}.

Suppose we want to generate the value of a random variable X having pmf $P(X = j) = P_j, j = 1, \ldots, N$. If we can generate an irreducible aperiodic Markov chain with limiting distribution $\{P_j\}$ then we can approximately generate X by running the chain for n steps to obtain X_n for n large.

If the goal is to generate many random variables with distribution $\{P_j\}$ so as to estimate (5.1.4) $E(h(X)) = \Sigma_{j=1}^N h(j) P_j$ then we estimate $E(h(X))$ using (5.1.5) $\frac{1}{n}\Sigma_{i=1}^n h(X_i)$. Since early states can be influenced by the initial condition, $\frac{1}{n-k}\Sigma_{i=k+1}^n h(X_i)$.

Recall: the problem is to generate the value of a random variable (vector) X with pmf $P(X = j) = P_j, j = 1, \ldots, N$. Idea: generate an irreducible, aperiodoc Markov chain with limiting distribution $\{P_i\}$ then we can approximately generate X by running the chain for a long time and using the state. If we want to estimate (5.1.4) $E(h(X)) = \Sigma_{j=1}^N h(j) P_j$ then we use (5.1.5) $\frac{1}{n}\Sigma_{i=1}^n h(X_n)$. People also use (5.1.6) $\widehat{\theta} = \frac{1}{n-k}\Sigma_{i=k}^n h(X_n)$.

5.2 The Hastings-Metropolis Algorithm

Let $\{b(j)\}_{j=1}^m$ be positive numbers and $B = \Sigma_{j=1}^m b(j)$. Assume m is large and B is difficult to compute. We want to simulate a random variable (or sequence of random variables) with pmf $\Pi(j) = \frac{b(j)}{B}, j = 1, 2, \ldots, m$. We find a Markov chain whose limiting distribution is $\{\Pi_j\}$ and which is easy to comupte. This is the Hastings-Metropolis algorithm. In addition, the chain is also reversible.

Let Q be the transition probability matrix for an irreducible Markov chain with state space $S = \{1, 2, \ldots, m\}$ and entries $q(i, j)$. We define a Markov chain $\{X_n\}_{n\geq 0}$ as follows: when $X_n = i$, a random variable X is generated such that $P(X = j) = q(i, j), j = 1, \ldots, m$. If $X = j$, then X_{n+1} is set equal to j with probability $\alpha(i, j)$ and set equal to i with probability $1 - \alpha(i, j)$. The sequence of states constitutes a Markov chain with transition probabilities $P_{ij} = q(i, j)\alpha(i, j), j \neq i, P_{ii} = q(i, i) + \Sigma_{k\neq 1} q(i, k)(1 - \alpha(i, k))$. The Markov chain is reversible, and the stationary probabilities satisfy $\Pi_i P_{ij} = \Pi_j P_{ji}, j \neq i$ or $\Pi_i q(i, j)\alpha(i, j) = \Pi_j q(j, i)\alpha(j, i)$. These equations are satisfied if (5.2.1) $\alpha(i, j) = \min(\frac{\Pi_j q(i,j)}{\Pi_i q(i,j)}, 1) = \min(\frac{b_j q(j,i)}{b_i q(i,j)}, 1)$. Note: we do not use B to define the chain. Note: generally (if $P_{ii} > 0$ for some i), $\{\Pi_i\}$ is also the limiting distribution.

Hastings-Metropolis Algorithm: generate a time reversible Markov chain with limiting distribution $\Pi_j = \frac{b(j)}{B}, j = 1, 2, \ldots, m, B = \Sigma_{j=1}^m b(j)$.

1. Choose an irreducible Markov chain on $\{1, \ldots, m\}$ (that is easy to comupte) with transition probability matrix $Q = (q(i, j))$. Choose $k \in \{1, 2, \ldots, m\}$.

2. Let $n = 0$ and $X_0 = k$.

3. Generate a random variable X with $P(X = j) = q(X_n, j)$ and a uniform random varible $U \in [0, 1]$.

4. If $U < \frac{b(X)q(X,X_n)}{b(X_n)q(X_n,X)}$ then $NS = X$ or else $NS = X_n$.

5. $n = n + 1, X_n = NS$.

6. Go to (3).

Example 5.2.1 - Suppose we want to generate a random element from a set ℓ of all permutations $(X_1, \ldots, X_n)$ of the numbers $\{1, 2, \ldots, n\}$ for which $\Sigma_{j=1}^n jX_j > a$, a is fixed. We define two such permutations to be neighbors if one results from the other by an interchange of two of the positions of the other. So, e.g., (1, 2, 3, 4) and (1, 2, 4, 3) are potentially neighbors if $1 + 2 \cdot 2 + 3 \cdot 4 + 4 \cdot 3 > a$. But, (1, 2, 3, 4) and (1, 3, 4, 2) certainly are not. We define the q transition probability function as follows: $N(s)$ is the set of neighbors of permutation s. $|N(s)|$ is the number of members in $N(s)$. Then we set $q(s, t) = \{\frac{1}{|N(s)|}$ for $t \in N(s), 0$ else. Thus, the target next state from s is equally likely to be any of its neighbors. Since the desired limiting probabilities are uniform and constant: $\Pi(s) = \Pi(t) = c, \alpha(s, t) = \min(\frac{|N(s)|}{|N(t)|}, 1)$. If the present state is s, one of its neighbors t is chosen at random. If t is a state with fewer neighbors than s, then t is the next state. If t does not have fewer neighbors, a random number U uniform on $[0, 1]$ is chosen and the next state is t if $U < \frac{|N(s)|}{|N(t)|}$ and s otherwise. The limiting probabilities are $\Pi_i = \frac{1}{|\ell|}$.

5.3 The Gibbs Sampler

Let $\vec{X} = \{X_1, \ldots, X_n\}$ be a random vector with pmf $P(\vec{x})$, that need only be specified up to a multiplicative constant. Suppose we want to generate a random vector having pmf $P(\vec{x}) = cg(\vec{x})$ where g is known but c is not. We assume that for any i and values $x_j, j \neq i$, we can generate a random variable X having pmf (5.3.1) $P(X = x) = P(X_i = x|X_j = x_j, j \neq i)$. We use the Hastings-Metropolis algorithm on a Markov chain with states $\vec{x} = (x_1, \ldots, x_n)$ and transition probabilities defined as follows: when the present state is $\vec{x}$ a coordinate i that is equally likely to be any of $1, 2, \ldots, n$ is chosen, then a random variable X whose pmf is given by (5.3.1) is generated, and if the value of X is x, the state $\vec{y} = (x_1, x_2, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_n)$ is considered as the candidate for the next state. In other words, with $\vec{x}$ and $\vec{y}$ given, we use the Hastings-Metropolis algorithm with $q(\vec{x}, \vec{y}) = \frac{1}{n}P(X_i = x|X_j = x_j, j \neq i) = \frac{P(\vec{y})}{nP(X_j=x_j,j\neq i)}$. (5.2.1) (given that the limiting distribution should be P), implies $\vec{y}$ is accepted as the new state with probability $\alpha(\vec{x}, \vec{y}) = \min(\frac{P(\vec{y})q(\vec{y},\vec{x})}{P(\vec{x})q(\vec{x},\vec{y})}, 1) = \min(\frac{P(\vec{y})P(\vec{x})}{P(\vec{x})P(\vec{y})}, 1) = 1$.

Example 5.3.1 - We want to generate n random points in the unit circle conditional on the event that no two points are within a distance d of each other, where $\beta = P$(no two points are within a distance d from each other) is assumed to be small and positive. Brute force sampling will work, but there will be lots of rejections in general. We use a Gibbs Sampler. We start with n points $x_1, \ldots, x_n$ in the circle such that no two are within distance d of each other. We generate a random number U, uniformly distributed on $[0, 1]$, and set $I =$ integer part of $(nU) + 1$ (I is discretely distributed on the integers 1, 2, 3, up to n). We also generate 1 random point in the unit circle. If this point is not within d of any of the $n - 1$ points excluding x_I, then replace x_I by this generated point, otherwise generate a new point and continue.

Example 5.3.2 - Let $X_i, i = 1, \ldots, n$ be independent random variables with X_i having exponential distribution with parameter $\lambda_i, i = i, \ldots, n$. Let $S = \Sigma_{i=1}^n X_i$. Suppose we want to generate the random vector $\vec{X} = (X_1, \ldots, X_n)$ conditional on the event that $S > c$ for some constant $c > 0$ (in general a large constant which makes this difficult). We want to generate the value of a random vector with density function $f(x_1, \ldots, x_n) = \frac{1}{P(S>c)}\Pi_{i=1}^n \lambda_i e^{-\lambda_i x_i}$, where $\Sigma_{i=1}^n x_i > c$. We start with an initial vector $\vec{x} = (x_1, \ldots, x_n)$ satisfying $x_i > 0, i = 1, \ldots, n$, and $\Sigma_{i=1}^n x_i > c$. We generate $U \sim \text{unif}(0, 1), I =$ integer part of $(nU) + 1$. Suppose $I = i$. We want to generate an exponential random variable X with rate λ_i conditioned on the event that $X + \Sigma_{k\neq i} x_i > c$. Or generate X conditional on the event $X > c - \Sigma_{k\neq i} x_i$. Fact: an exponential conditioned to be greater than a positive constant is distributed as the constant plus the exponential. We generate the exponential random variable Y with rate λ_i ($Y = -\frac{1}{\lambda_i}\log U, U \sim \text{unif}(0, 1)$) and set $X = Y + (c - \Sigma_{j\neq i} x_j)^+$ where $b^+ = \{b$ for $b > 0, 0$ for $b \leq 0$.

5.4 Simulated Annealing

We consider the following problem: let A be a finite set of vectors. Let $V(\vec{x})$ be a non negative funciton an A. Let $V^* = \max_{\vec{x}\in A} V(x), M = \{\vec{x} \in A : V(\vec{x}) = V^*\}$. We want to find V^* and at least one element in M. We let $\lambda > 0$, consider the pmf on A, $P_\lambda(\vec{x}) = \frac{e^{\lambda V(\vec{x})}}{\Sigma_{\vec{x}\in A} e^{\lambda V(\vec{x})}}$. Let $|M|$ = number of elements in M. Then $P_\lambda(\vec{x})$ = [big equation]. For $x \notin M, P_\lambda(\vec{x}) \to_{\lambda\to\infty} \frac{\delta(\vec{x},M)}{|M|}, \sigma(\vec{x}, M) = \{1$ for $\vec{x} \in M, 0$ else. We let λ be large and generate a Markov chain whose limiting distribution is $P_\lambda(\vec{x})$ so that most of the mass of the limiting distribution is on vectors in M. Use the Hastings-Metropolis algorithm. $\vec{x}, \vec{y} \in A$ are neighbors if (1) they differ only in a single coordinate or (2) one is obtained from the other by switching two components. The target next state from a given state $vecx$ is equally likely to be any of its neighbors. If this is $\vec{y}$, $\vec{y}$ is chosen as the next state with probability $\min(1, \frac{e^{\lambda V(\vec{y})}/|N(\vec{y})|}{e^{\lambda V(\vec{x})}/|N(\vec{x})|})$ or stays $\vec{x}$ otherwise.

Simulated annealing, we let λ_n vary, so if the nth stage of the Markov chain is $\vec{x}$, and $\vec{y}$ is a candidate, then $\vec{y}$ is chosen with probability $\min(1, \frac{e^{\lambda_n V(\vec{y})}/|N(\vec{y})|}{e^{\lambda_n V(\vec{x})}/|N(\vec{x})|})$. λ_n starts small then becomes large. $\lambda_n = c\log(1+n)$.