Chapter 4

Dependent Random Variables

4.1 Conditioning

One of the key concepts in probability theory is the notion of conditional probability and conditional expectation. Suppose that we have a probability space $(\Omega, \mathcal{F}, P)$ consisting of a space Ω, a σ-field $\mathcal{F}$ of subsets of Ω and a probability measure on the σ-field $\mathcal{F}$. If we have a set $A \in \mathcal{F}$ of positive measure then conditioning with respect to A means we restrict ourselves to the set A. Ω gets replaced by A. The σ-field $\mathcal{F}$ by the σ-field $\mathcal{F}_A$ of subsets of A that are in $\mathcal{F}$. For $B \subset A$ we define

$$P_A(B) = \frac{P(B)}{P(A)}$$

We could achieve the same thing by defining for arbitrary $B \in \mathcal{F}$

$$P_A(B) = \frac{P(A \cap B)}{P(A)}$$

in which case $P_A(\cdot)$ is a measure defined on $\mathcal{F}$ as well but one that is concentrated on A and assigning 0 probability to A^c. The definition of conditional probability is

$$P(B|A) = \frac{P(A \cap B)}{P(A)} .$$

Similarly the definition of conditional expectation of an integrable function $f(\omega)$ given a set $A \in \mathcal{F}$ of positive probability is defined to be

$$E\{f|A\} = \frac{\int_A f(\omega) dP}{P(A)} .$$

In particular if we take $f = \chi_B$ for some $B \in \mathcal{F}$ we recover the definition of conditional probability. In general if we know $P(B|A)$ and $P(A)$ we can recover $P(A \cap B) = P(A)P(B|A)$ but we cannot recover $P(B)$. But if we know $P(B|A)$ as well as $P(B|A^c)$ along with $P(A)$ and $P(A^c) = 1 - P(A)$ then

$$P(B) = P(A \cap B) + P(A^c \cap B) = P(A)P(B|A) + P(A^c)P(B|A^c).$$

More generally if $\mathcal{P}$ is a partition of Ω into a finite or even a countable number of disjoint measurable sets $A_1, \cdots, A_j, \cdots$

$$P(B) = \sum_j P(A_j)P(B|A_j).$$

If ξ is a random variable taking distinct values $\{a_j\}$ on $\{A_j\}$ then

$$P(B|\xi = a_j) = P(B|A_j)$$

or more generally

$$P(B|\xi = a) = \frac{P(B \cap \xi = a)}{P(\xi = a)}$$

provided $P(\xi = a) > 0$. One of our goals is to seek a definition that makes sense when $P(\xi = a) = 0$. This involves dividing 0 by 0 and should involve differentiation of some kind. In the countable case we may think of $P(B|\xi = a_j)$ as a function $f_B(\xi)$ which is equal to $P(B|A_j)$ on $\xi = a_j$. We can rewrite our definition of

$$f_B(a_j) = P(B|\xi = a_j)$$

as

$$\int_{\xi = a_j} f_B(\xi)dP = P(B \cap \xi = a_j) \qquad \text{for each} \quad j$$

or summing over any arbitrary collection of j's

$$\int_{\xi \in E} f_B(\xi)dP = P(B \cap \{\xi \in E\}).$$

Sets of the form $\xi \in E$ form a sub σ-field $\Sigma \subset \mathcal{F}$ and we can rewrite the definition as

$$\int_A f_B(\xi)dP = P(B \cap A)$$

for all $A \in \Sigma$. Of course in this case $A \in \Sigma$ if and only if A is a union of the atoms $\xi = a$ of the partition over a finite or countable subcollection of the possible values of a. Similar considerations apply to the conditional expectation of a random variable G given ξ. The equation becomes

$$\int_A g(\xi) dP = \int_A G(\omega) dP$$

or we can rewrite this as

$$\int_A g(\omega) dP = \int_A G(\omega) dP$$

for all $A \in \Sigma$ and instead of demanding that g be a function of ξ we demand that g be Σ measurable which is the same thing. Now the random variable ξ is out of the picture and rightly so. What is important is the information we have if we know ξ and that is the same if we replace ξ by a one-to-one function of itself. The σ-field Σ abstracts that information nicely. So it turns out that the proper notion of conditioning involves a sub σ-field $\Sigma \subset \mathcal{F}$. If G is an integrable function and $\Sigma \subset \mathcal{F}$ is given we will seek another integrable function g that is Σ measurable and satisfies

$$\int_A g(\omega) dP = \int_A G(\omega) dP$$

for all $A \in \Sigma$. We will prove existence and uniqueness of such a g and call it the conditional expectation of G given Σ and denote it by $g = E[G|\Sigma]$.

The way to prove the above result will take us on a detour. A signed measure on a measurable space $(\Omega, \mathcal{F})$ is a set function $\lambda(.)$ defined for $A \in \mathcal{F}$ which is countably additive but not necessarily nonnegative. Countable addivity is again in any of the following two equivalent senses.

$$\lambda(\cup A_n) = \sum \lambda(A_n)$$

for any countable collection of disjoint sets in $\mathcal{F}$, or

$$\lim_{n \to \infty} \lambda(A_n) = \lambda(A)$$

whenver $A_n \downarrow A$ or $A_n \uparrow A$.

Examples of such λ can be constructed by taking the difference $\mu_1 - \mu_2$ of two nonnegative measures μ_1 and μ_2.

Definition 4.1. *A set $A \in \mathcal{F}$ is totally positive (totally negative) for λ if for every subset $B \in \mathcal{F}$ with $B \subset A$ $\lambda(B) \geq 0$. (≤ 0)*

Remark 4.1. A measurable subset of a totally positive set is totally positive. Any countable union of totally positive subsets is again totally positive.

Lemma 4.1. *If λ is a countably additive signed measure on $(\Omega, \mathcal{F})$,*

$$\sup_{A \in \mathcal{F}} |\lambda(A)| < \infty$$

Proof. The key idea in the proof is that, since $\lambda(\Omega)$ is a finite number, if $\lambda(A)$ is large so is $\lambda(A^c)$ with an opposite sign. In fact, it is not hard to see that $||\lambda(A)| - |\lambda(A^c)|| \leq |\lambda(\Omega)|$ for all $A \in \mathcal{F}$. Another fact is that if $\sup_{B \subset A} |\lambda(B)|$ and $\sup_{B \subset A^c} |\lambda(B)|$ are finite, so is $\sup_B |\lambda(B)|$. Now let us complete the proof. Given a subset $A \in \mathcal{F}$ with $\sup_{B \subset A} |\lambda(B)| = \infty$, and any positive number N, there is a subset $A_1 \in \mathcal{F}$ with $A_1 \subset A$ such that $|\lambda(A_1)| \geq N$ and $\sup_{B \subset A_1} |\lambda(B)| = \infty$. This is obvious because if we pick a set $E \subset A$ with $|\lambda(E)|$ very large so will $\lambda(E^c)$ be. At least one of the two sets E, E^c will have the second property and we can call it A_1. If we proceed by induction we have a sequence A_n that is $\downarrow$ and $|\lambda(A_n)| \to \infty$ that contradicts countable additivity. □

Lemma 4.2. *Given a subset $A \in \mathcal{F}$ with $\lambda(A) = \ell > 0$ there is a subset $\bar{A} \subset A$ that is totally positive with $\lambda(\bar{A}) \geq \ell$.*

Proof. Let us define $m = \inf_{B \subset A} \lambda(B)$. Since the empty set is included, $m \leq 0$. If $m = 0$ then A is totally positive and we are done. So let us assume that $m < 0$. By the previous lemma $m > -\infty$.

Let us find $B_1 \subset A$ such that $\lambda(B_1) \leq \frac{m}{2}$. Then for $A_1 = A - B_1$ we have $A_1 \subset A$, $\lambda(A_1) \geq \ell$ and $\inf_{B \subset A_1} \lambda(B) \geq \frac{m}{2}$. By induction we can find A_k with $A \supset A_1 \supset \cdots \supset A_k \cdots$, $\lambda(A_k) \geq \ell$ for every k and $\inf_{B \subset A_k} \lambda(A_k) \geq \frac{m}{2^k}$. Clearly if we define $\bar{A} = \cap A_k$ which is the decreasing limit, $\bar{A}$ works. □

Theorem 4.3. (Hahn-Jordan Decomposition). *Given a countably additive signed measure λ on $(\Omega, \mathcal{F})$ it can be written always as $\lambda = \mu^+ - \mu^-$ the difference of two nonnegative measures. Moreover μ^+ and μ^- may be chosen to be orthogonal i.e, there are disjoint sets $\Omega_+, \Omega_- \in \mathcal{F}$ such that $\mu^+(\Omega_-) = \mu^-(\Omega_+) = 0$. In fact Ω_+ and Ω_- can be taken to be subsets of Ω that are respectively totally positive and totally negative for λ. $\mu^\pm$ then become just the restrictions of λ to $\Omega_\pm$.*

Proof. Totally positive sets are closed under countable unions, disjoint or not. Let us define $m^+ = \sup_A \lambda(A)$. If $m^+ = 0$ then $\lambda(A) \le 0$ for all A and we can take $\Omega_+ = \Phi$ and $\Omega_- = \Omega$ which works. Assume that $m^+ > 0$. There exist sets A_n with $\lambda(A) \ge m^+ - \frac{1}{n}$ and therefore totally positive subsets $\bar{A}_n$ of A_n with $\lambda(\bar{A}_n) \ge m^+ - \frac{1}{n}$. Clearly $\Omega_+ = \cup_n \bar{A}_n$ is totally positive and $\lambda(\Omega_+) = m^+$. It is easy to see that $\Omega_- = \Omega - \Omega_+$ is totally negative. $\mu^\pm$ can be taken to be the restriction of λ to $\Omega_\pm$. □

Remark 4.2. If $\lambda = \mu^+ - \mu^-$ with μ^+ and μ^- orthogonal to each other, then they have to be the restrictions of λ to the totally positive and totally negative sets for λ and such a representation for λ is unique. It is clear that in general the representation is not unique because we can add a common μ to both μ^+ and μ^- and the μ will cancel when we compute $\lambda = \mu^+ - \mu^-$.

Remark 4.3. If μ is a nonnegative measure and we define λ by

$$\lambda(A) = \int_A f(\omega)\, d\mu = \int \chi_A(\omega) f(\omega)\, d\mu$$

where f is an integrable function, then λ is a countably additive signed measure and $\Omega_+ = \{\omega : f(\omega) > 0\}$ and $\Omega_- = \{\omega : f(\omega) < 0\}$. If we define $f^\pm(\omega)$ as the positive and negative parts of f, then

$$\mu^\pm(A) = \int_A f^\pm(\omega)\, d\mu.$$

The signed measure λ that was constructed in the preceeding remark enjoys a very special relationship to μ. For any set A with $\mu(A) = 0$, $\lambda(A) = 0$ because the integrand $\chi_A(\omega) f(\omega)$ is 0 for μ-almost all ω and for all practical purposes is a function that vanishes identically.

Definition 4.2. *A signed measure λ is said to be absolutely continuous with respect to a nonnegative measure μ, $\lambda << \mu$ in symbols, if whenever $\mu(A)$ is zero for a set $A \in \mathcal{F}$ it is also true that $\lambda(A) = 0$.*

Theorem 4.4. (Radon-Nikodym Theorem). *If $\lambda << \mu$ then there is an integrable function $f(\omega)$ such that*

$$\lambda(A) = \int_A f(\omega)\, d\mu \tag{4.1}$$

for all $A \in \mathcal{F}$. The function f is uniquely determined almost everywhere and is called the Radon-Nikodym derivative of λ with respect to μ. It is denoted by

$$f(\omega) = \frac{d\lambda}{d\mu} \ .$$

Proof. The proof depends on the decomposition theorem. We saw that if the relation 4.1 holds, then $\Omega_+ = \{\omega : f(\omega) > 0\}$. If we define $\lambda_a = \lambda - a\mu$, then λ_a is a signed measure for every real number a. Let us define $\Omega(a)$ to be the totally positive subset of λ_a. These sets are only defined up to sets of measure zero, and we can only handle a countable number of sets of measure 0 at one time. So it is prudent to restrict a to the set Q of rational numbers. Roughly speaking $\Omega(a)$ will be the sets $f(\omega) > a$ and we will try to construct f from the sets $\Omega(a)$ by the definition

$$f(\omega) = [\sup a \in Q : \omega \in \Omega(a)].$$

The plan is to check that the function $f(\omega)$ defined above works. Since λ_a is getting more negative as a increases, $\Omega(a)$ is $\downarrow$ as $a \uparrow$. There is trouble with sets of measure 0 for every comparison between two rationals a_1 and a_2. Collect all such troublesome sets (only a countable number and throw them away). In other words we may assume without loss of generality that $\Omega(a_1) \subset \Omega(a_2)$ whenever $a_1 > a_2$. Clearly

$$\begin{aligned} \{\omega : f(\omega) > x\} &= \{\omega : \omega \in \Omega(y) \ \text{ for some rational } \ y > x\} \\ &= \cup_{\substack{y>x \\ y \in Q}} \Omega(y) \end{aligned}$$

and this makes f measurable. If $A \subset \cap_a \Omega(a)$ then $\lambda(A) - a\mu(A) \geq 0$ for all A. If $\mu(A) > 0$, $\lambda(A)$ has to be infinite which is not possible. Therefore $\mu(A)$ has to be zero and by absolute continuity $\lambda(A) = 0$ as well. On the other hand if $A \cap \Omega(a) = \Phi$ for all a, then $\lambda(A) - a\mu(A) \leq 0$ for all a and again if $\mu(A) > 0$, $\lambda(A) = -\infty$ which is not possible either. Therefore $\mu(A)$, and by absolute continuity, $\lambda(A)$ are zero. This proves that $f(\omega)$ is finite almost everywhere with respect to both λ and μ. Let us take two real numbers $a < b$ and consider $E_{a,b} = \{\omega : a \leq f(\omega) \leq b\}$. It is clear that the set $E_{a,b}$ is in $\Omega(a')$ and $\Omega^c(b')$ for any $a' < a$ and $b' > b$. Therefore for any set $A \subset E_{a,b}$ by letting a' and b' tend to a and b

$$a\,\mu(A) \leq \lambda(A) \leq b\,\mu(A).$$

Now we are essentially done. Let us take a grid $\{n\,h\}$ and consider $E_n = \{\omega : nh \le f(\omega) < (n+1)h\}$ for $-\infty < n < \infty$. Then for any $A \in \mathcal{F}$ and each n,

$$\begin{aligned}\lambda(A \cap E_n) - h\mu(A \cap E_n) \le n\,h\,\mu(A \cap E_n) \le \int_{A\cap E_n} f(\omega)\,d\mu \\ \le (n+1)\,h\mu(A \cap E_n) \le \lambda(A \cap E_n) + h\,\mu(A \cap E_n).\end{aligned}$$

Summing over n we have

$$\lambda(A) - h\,\mu(A) \le \int_A f(\omega)\,d\mu \le \lambda(A) + h\,\mu(A)$$

proving the integrability of f and if we let $h \to 0$ establishing

$$\lambda(A) = \int_A f(\omega)\,d\mu$$

for all $A \in \mathcal{F}$. □

Remark 4.4. **(Uniqueness).** If we have two choices of f say f_1 and f_2 their difference $g = f_1 - f_2$ satisfies

$$\int_A g(\omega)\,d\mu = 0$$

for all $A \in \mathcal{F}$. If we take $A_\epsilon = \{\omega : g(\omega) \ge \epsilon\}$, then $0 \ge \epsilon\mu(A_\epsilon)$ and this implies $\mu(A_\epsilon) = 0$ for all $\epsilon > 0$ or $g(\omega) \le 0$, almost everywhere with respect to μ. A similar argument establishes $g(\omega) \ge 0$ almost everywhere with respect to μ. Therefore $g = 0$ a.e. μ proving uniqueness.

Exercise 4.1. If f and g are two integrable functions, maesurable with respect to a σ-filed $\mathcal{B}$ and if $\int_A f(\omega)dP = \int_A g(\omega)dP$ for all sets $A \in \mathcal{B}_0$, a field that generates the σ-field $\mathcal{B}$, then $f = g$ a.e. P.

Exercise 4.2. If $\lambda(A) \ge 0$ for all $A \in \mathcal{F}$, prove that $f(\omega) \ge 0$ almost everywhere.

Exercise 4.3. If Ω is a countable set and $\mu(\{\omega\}) > 0$ for each single point set prove that any measure λ is absolutely continuous with respect to λ and calculate the Radon-Nikodym derivative.

Exercise 4.4. Let $F(x)$ be a distribution function on the line with $F(0) = 0$ and $F(1) = 1$ so that the probability measure α corresponding to it lives on the interval $[0, 1]$. If $F(x)$ satisfies a Lipschitz condition

$$|F(x) - F(y)| \leq A|x - y|$$

then prove that $\alpha << m$ where m is the Lebesgue measure on $[0, 1]$. Show also that $0 \leq \frac{d\alpha}{dm} \leq A$ almost surely.

If ν, λ, μ are three nonnegative measures such that $\nu << \lambda$ and $\lambda << \mu$ then show that $\nu << \mu$ and

$$\frac{d\nu}{d\mu} = \frac{d\nu}{d\lambda}\frac{d\lambda}{d\mu}$$

a.e.

Exercise 4.5. If λ, μ are nonnegative measures with $\lambda << \mu$ and $\frac{d\lambda}{d\mu} = f$, then show that g is integrable with respect to λ if and only if $g\, f$ is integrable with respect to μ and

$$\int g(\omega)\, d\lambda = \int g(\omega)\, f(\omega)\, d\mu.$$

Exercise 4.6. Given two nonnegative measures λ and μ, λ is said to be uniformly absolutely continuous with respect to μ on $\mathcal{F}$ if for any $\epsilon > 0$ there exists a $\delta > 0$ such that for any $A \in \mathcal{F}$ with $\mu(A) < \delta$ it is true that $\lambda(A) < \epsilon$. Use the Radon-Nikodym theorem to show that absolute continuity on a σ-field $\mathcal{F}$ implies uniform absolute continuity. If $\mathcal{F}_0$ is a field that generates the σ-field $\mathcal{F}$ show by an example that absolute continuity on $\mathcal{F}_0$ does not imply absolute continuity on $\mathcal{F}$. Show however that uniform absolute continuity on $\mathcal{F}_0$ implies uniform absolute continuity and therefore absolute continuity on $\mathcal{F}$.

Exercise 4.7. If F is a distribution function on the line show that it is absolutely continuous with respect to Lebesgue measure on the line, if and only if for any $\epsilon > 0$, there exists a $\delta > 0$ such that for arbitrary finite collection of disjoint intervals $I_j = [a_j, b_j]$ with $\sum_j |b_j - a_j| < \delta$ it follows that $\sum_j [F(b_j) - F(a_j)] \leq \epsilon$.

4.2 Conditional Expectation

In the Radon-Nikodym theorem, if $\lambda << \mu$ are two probability distributions on $(\Omega, \mathcal{F})$, we defined the Radon-Nikodym derivative $f(\omega) = \frac{d\lambda}{d\mu}$ as an $\mathcal{F}$

measurable function such that

$$\lambda(A) = \int_A f(\omega)\, d\mu \qquad \text{for all} \;\; A \in \mathcal{F}$$

If $\Sigma \subset \mathcal{F}$ is a sub σ-field, the absolute continuity of λ with respect to μ on Σ is clearly implied by the absolute continuity of λ with respect to μ on $\mathcal{F}$. We can therefore apply the Radon-Nikodym theorem on the measurable space (Ω, Σ), and we will obtain a new Radon-Nikodym derivative

$$g(\omega) = \frac{d\lambda}{d\mu} = \frac{d\lambda}{d\mu}\Big|_{\Sigma}$$

such that

$$\lambda(A) = \int_A g(\omega)\, d\mu \qquad \text{for all} \;\; A \in \Sigma$$

and g is Σ measurable. Since the old function $f(\omega)$ was only $\mathcal{F}$ measurable, in general, it cannot be used as the Radon-Nikodym derivative for the sub σ-field Σ. Now if f is an integrable function on $(\Omega, \mathcal{F}, \mu)$ and $\Sigma \subset \mathcal{F}$ is a sub σ-field we can define λ on $\mathcal{F}$ by

$$\lambda(A) = \int_A f(\omega)\, d\mu \qquad \text{for all} \;\; A \in \mathcal{F}$$

and recalculate the Radon-Nikodym derivative g for Σ and g will be a Σ measurable, integrable function such that

$$\lambda(A) = \int_A g(\omega)\, d\mu \qquad \text{for all} \;\; A \in \Sigma$$

In other words g is the perfect candidate for the conditional expectation

$$g(\omega) = E\big\{f(\cdot)|\Sigma\big\}$$

We have therefore proved the existence of the conditional expectation.

Theorem 4.5. *The conditional expectation as mapping of $f \to g$ has the following properties.*

1. *If $g = E\big\{f|\Sigma\big\}$ then $E[g] = E[f]$. $E[1|\Sigma] = 1$ a.e.*
2. *If f is nonnegative then $g = E\big\{f|\Sigma\big\}$ is almost surely nonnegative.*

3. *The map is linear. If a_1, a_2 are constants*

$$E\{a_1 f_1 + a_2 f_2 | \Sigma\} = a_1 E\{f_1|\Sigma\} + a_2 E\{f_2|\Sigma\} \quad \text{a.e.}$$

4. *If $g = E\{f|\Sigma\}$, then*

$$\int |g(\omega)|\, d\mu \leq \int |f(\omega)|\, d\mu$$

5. *If h is a bounded Σ measurable function then*

$$E\{f\, h|\Sigma\} = h\, E\{f|\Sigma\} \quad \text{a.e.}$$

6. *If $\Sigma_2 \subset \Sigma_1 \subset \mathcal{F}$, then*

7. **Jensen's Inequality.** *If $\phi(x)$ is a convex function of x, and $g = E\{f|\Sigma\}$ then*

$$E\{\phi(f(\omega))|\Sigma\} \geq \phi(g(\omega)) \quad \text{a.e.} \tag{4.2}$$

and if we take expectations

$$E[\phi(f)] \geq E[\phi(g)].$$

Proof. (i), (ii) and (iii) are obvious. For (iv) we note that if $d\lambda = f\, d\mu$

$$\int |f|\, d\mu = \sup_{A \in \mathcal{F}} \lambda(A) - \inf_{A \in \mathcal{F}} \lambda(A)$$

and if we replace $\mathcal{F}$ by a sub σ-field Σ the right hand side is decreased. (v) is obvious if h is the indicator function of a set A in Σ. To go from indicator functions to simple functions to bounded measurable functions is routine. (vi) is an easy consequence of the definition. Finally (vii) corresponds to Theorem 1.7 proved for ordinary expectations and is proved analogously. We note that if $f_1 \geq f_2$ then $E\{f_1|\Sigma\} \geq E\{f_2|\Sigma\}$ a.e. and consequently $E\{\max(f_1, f_2)|\Sigma\} \geq \max(g_1, g_2)$ a.e. where $g_i = E\{f_i|\Sigma\}$ for $i = 1, 2$. Since we can represent any convex function ϕ as $\phi(x) = \sup_a [ax - \psi(a)]$, limiting ourselves to rational a, we have only a countable set of functions to deal with, and

$$\begin{aligned} E\big\{\phi(f)|\Sigma\big\} &= E\big\{\sup_a [af - \psi(a)]|\Sigma\big\} \\ &\geq \sup_a \big[a\,E\big\{f|\Sigma\big\} - \psi(a)\big] \\ &= \sup_a \big[a\,g - \psi(a)\big] \\ &= \phi(g) \end{aligned}$$

a.e. and after taking expectations

$$E[\phi(f)] \geq E[\phi(g)].$$

□

Remark 4.5. Conditional expecation is a form of averaging, i.e. it is linear, takes constants into constants and preserves nonnegativity. Jensen's inequality is now a cosequence of convexity.

In a somewhat more familiar context if $\mu = \lambda_1 \times \lambda_2$ is a product measure on $(\Omega, \mathcal{F}) = (\Omega_1 \times \Omega_2, \mathcal{F}_1 \times \mathcal{F}_2)$ and we take $\Sigma = \big\{A \times \Omega_2 : A \in \mathcal{F}_1\big\}$ then for any function $f(\omega) = f(\omega_1, \omega_2)$, $E[f(\cdot)|\Sigma] = g(\omega)$ where $g(\omega) = g(\omega_1)$ is given by

$$g(\omega_1) = \int_{\Omega_2} f(\omega_1, \omega_2)\, d\lambda_2$$

so that the conditional expectation is just integrating the unwanted variable ω_2. We can go one step more. If $\phi(x, y)$ is the joint density on R^2 of two random variables X, Y (with respect to the Lebesgue measure on R^2), and $\psi(x)$ is the marginal density of X given by

$$\psi(x) = \int_{-\infty}^{\infty} \phi(x, y)\, dy$$

then for any integrable function $f(x, y)$

$$E[f(X, Y)|X] = E[f(\cdot, \cdot)|\Sigma] = \frac{\int_{-\infty}^{\infty} f(x, y)\, \phi(x, y)\, dy}{\psi(x)}$$

where Σ is the σ-field of vertical strips $A \times (-\infty, \infty)$ with a measurable horizontal base A.

Exercise 4.8. If f is already Σ measurable then $E[f|\Sigma] = f$. This suggests that the map $f \to g = E[f|\Sigma]$ is some sort of a projection. In fact if we consider the Hilbert space $\mathbf{H} = L_2[\Omega, \mathcal{F}, \mu]$ of all $\mathcal{F}$ measurable square integrable functions with an inner product

$$< f, g >_\mu = \int f\, g\, d\mu$$

then

$$\mathbf{H}_0 = L_2[\Omega, \Sigma, \mu] \subset \mathbf{H} = L_2[\Omega, \mathcal{F}, \mu]$$

and $f \to E[f|\Sigma]$ is seen to be the same as the orthogonal projection from $\mathbf{H}$ onto $\mathbf{H}_0$. Prove it.

Exercise 4.9. If $\mathcal{F}_1 \subset \mathcal{F}_2 \subset \mathcal{F}$ are two sub σ-fields of $\mathcal{F}$ and X is any integrable function, we can define $X_i = E[X|\mathcal{F}_i]$ for $i = 1, 2$. Show that $X_1 = E[X_2|\mathcal{F}_1]$ a.e.

Conditional expectation is then the best nonlinear predictor if the loss function is the expected (mean) square error.

4.3 Conditional Probability

We now turn our attention to conditional probability. If we take $f = \chi_B(\omega)$ then $E[f|\Sigma] = P(\omega, B)$ is called the conditional probability of B given Σ. It is characterized by the property that it is Σ measurable as a function of ω and for any $A \in \Sigma$

$$\mu(A \cap B) = \int_A P(\omega, B)\, d\mu.$$

Theorem 4.6. *$P(\cdot, \cdot)$ has the following properties.*

1. *$P(\omega, \Omega) = 1, P(\omega, \Phi) = 0$ a.e.*

2. *For any $B \in \mathcal{F}$, $0 \le P(\omega, B) \le 1$ a.e.*

3. *For any countable collection $\{B_j\}$ of disjoint sets in $\mathcal{F}$,*

$$P(\omega, \cup_j B_j) = \sum_j P(\omega, B_j) \quad \text{a.e.}$$

4. If $B \in \Sigma$, $P(\omega, B) = \chi_B(\omega)$ a.e.

Proof. All are easy consequences of properties of conditional expectations. Property (iii) perhaps needs an explanation. If $E[|f_n - f|] \to 0$ by the properties of conditional expectation $E[|E\{f_n|\Sigma\} - E\{f|\Sigma\}| \to 0$. Property (iii) is an easy consequence of this. □

The problem with the above theorem is that every property is valid only almost everywhere. There are exceptional sets of measure zero for each case. While each null set or a countable number of them can be ignored we have an uncountable number of null sets and we would like a single null set outside which all the properties hold. This means constructing a good version of the conditional probability. It may not be always possible. If possible, such a version is called a regular conditional probability. The existence of such a regular version depends on the space $(\Omega, \mathcal{F})$ and the sub σ-field Σ being nice. If Ω is a complete separable metric space and $\mathcal{F}$ are its Borel stes, and if Σ is any countably generated sub σ-field of $\mathcal{F}$, then it is nice enough. We will prove it in the special case when $\Omega = [0, 1]$ is the unit interval and $\mathcal{F}$ are the Borel subsets $\mathcal{B}$ of $[0, 1]$. Σ can be any countably generated sub σ-field of $\mathcal{F}$.

Remark 4.6. In fact the case is not so special. There is theorem [6] which states that if $(\Omega, \mathcal{F})$ is any complete separable metric space that has an uncountable number of points, then there is one-to-one measurable map with a measurable inverse between $(\Omega, \mathcal{F})$ and $([0, 1], \mathcal{B})$. There is no loss of generality in assuming that $(\Omega, \mathcal{F})$ is just $([0, 1], \mathcal{B})$.

Theorem 4.7. *Let P be a probability distribution on $([0, 1], \mathcal{B})$. Let $\Sigma \subset \mathcal{B}$ be a sub σ-field. There exists a family of probability distributions Q_x on $([0, 1], \mathcal{B})$ such that for every $A \in \mathcal{B}$, $Q_x(A)$ is Σ measurable and for every $\mathcal{B}$ measurable f,*

$$\int f(y) Q_x(dy) = E^P[f(\cdot)|\Sigma] \qquad \text{a.e. } P. \tag{4.3}$$

If in addition Σ is countably generated, i.e there is a field Σ_0 consisting of a countable number of Borel subsets of $[0, 1]$ such that the σ-field generated by Σ_0 is Σ, then

$$Q_x(A) = \mathbf{1}_A(x) \qquad \text{for all } A \in \Sigma. \tag{4.4}$$

Proof. The trick is not to be too ambitious in the first place but try to construct the conditional expectations

$$Q(\omega, B) = E\{\chi_B(\omega)|\Sigma\}$$

only for sets B given by $B = (-\infty, x)$ for rational x. We call our conditional expectation, which is in fact a conditional probability, by $F(\omega, x)$. By the properties of conditional expectations for any pair of rationals $x < y$, there is a null set $E_{x,y}$, such that for $\omega \notin E_{x,y}$

$$F(\omega, x) \leq F(\omega, y).$$

Moreover for any rational $x < 0$, there is a null set N_x outside which $F(\omega, x) = 0$ and similar null sets N_x for $x > 1$, ouside which $F(\omega, x) = 1$. If we collect all these null sets, of which there are only countably many, and take their union, we get a null set $N \in \Sigma$ such that for $\omega \notin N$, we have have a family $F(\omega, x)$ defined for rational x that satisfies

$$\begin{aligned}
F(\omega, x) \leq F(\omega, y) \quad &\textit{if} \quad x < y \quad \textit{are} \quad \textit{rational} \\
F(\omega, x) = 0 \quad &\textit{for} \quad \textit{rational} \quad x < 0 \\
F(\omega, x) = 1 \quad &\textit{for} \quad \textit{rational} \quad x > 1 \\
P(A \cap [0, x]) = \int_A F(\omega, x)\, dP \quad &\text{for} \quad \text{all} \quad A \in \Sigma.
\end{aligned}$$

For $\omega \notin N$ and real y we can define

$$G(\omega, y) = \lim_{\substack{x \downarrow y \\ x \text{ rational}}} F(\omega, x).$$

For $\omega \notin N$, G is a right continuous nondecreasing function (distribution function) with $G(\omega, y) = 0$ for $y < 0$ and $G(\omega, y) = 1$ for $y \geq 1$. There is then a probability measure $\hat{Q}(\omega, B)$ on the Borel subsets of $[0, 1]$ such that $\hat{Q}(\omega, [0, y]) = G(\omega, y)$ for all y. $\hat{Q}$ is our candidate for regular conditional probability. Clearly $\hat{Q}(\omega, I)$ is Σ measurable for all intervals I and by standard arguments will continue to be Σ measurable for all Borel sets $B \in \mathcal{F}$. If we check that

$$P(A \cap [0, x]) = \int_A G(\omega, x)\, dP \quad \text{for} \quad \text{all} \quad A \in \Sigma$$

for all $0 \le x \le 1$ then

$$P(A \cap I) = \int_A \hat{Q}(\omega, I)\, dP \quad \text{for all} \quad A \in \Sigma$$

for all intervals I and by standard arguments this will extend to finite disjoint unions of half open intervals that constitute a field and finally to the σ-field $\mathcal{F}$ generated by that field. To verify that for all real y,

$$P(A \cap [0, y]) = \int_A G(\omega, y)\, dP \quad \text{for all} \quad A \in \Sigma$$

we start from

$$P(A \cap [0, x]) = \int_A F(\omega, x)\, dP \quad \text{for all} \quad A \in \Sigma$$

valid for rational x and let $x \downarrow y$ through rationals. From the countable additivity of P the left hand side converges to $P(A \cap [0, y])$ and by the bounded convergence theorem, the right hand side converges to $\int_A G(\omega, y)\, dP$ and we are done.

Finally from the uniqueness of the conditional expectation if $A \in \Sigma$

$$\hat{Q}(\omega, A) = \chi_A(\omega)$$

provided $\omega \notin N_A$, which is a null set that depends on A. We can take a countable set Σ_0 of generators A that forms a field and get a single null set N such that if $\omega \notin N$

$$\hat{Q}(\omega, A) = \chi_A(\omega)$$

for all $A \in \Sigma_0$. Since both side are countably additive measures in A and as they agree on Σ_0 they have to agree on Σ as well. $\square$

Exercise 4.10. (Disintegration Theorem.) Let μ be a probability measure on the plane R^2 with a marginal distribution α for the first coordinate. In other words if we denote α is such that, for any f that is a bounded measurable function of x,

$$\int_{R^2} f(x)\, d\mu = \int_R f(x)\, d\alpha$$

Show that there exist measures β_x depending measurably on x such that $\beta_x[\{x\} \times R] = 1$, i.e. β_x is supported on the vertical line through $(x, y) : y \in R$ and $\mu = \int_R \beta_x\, d\alpha$. The converse is of course easier. Given α and β_x we can construct a unique μ such that μ disintegrates as expected.

4.4 Markov Chains

One of the ways of generating a sequence of dependent random variables is to think of a system evolving in time. We have time points that are discrete say $T = 0, 1, \cdots, N, \cdots$. The state of the system is described by a point x in the state space $\mathcal{X}$ of the system. The state space $\mathcal{X}$ comes with a natural σ-field of subsets $\mathcal{F}$. At time 0 the system is in a random state and its distribution is specified by a probability distribution μ_0 on $(\mathcal{X}, \mathcal{F})$. At successive times $T = 1, 2, \cdots$, the system changes its state and given the past history $(x_0, \cdots, x_k)$ of the states of the system at times $T = 0, \cdots, k-1$ the probability that system finds itself at time k in a subset $A \in \mathcal{F}$ is given by $\pi_k(x_0, \cdots, x_{k-1}\,; A\,)$. For each $(x_0, \cdots, x_{k-1})$, π_k defines a probability measure on $(\mathcal{X}, \mathcal{F})$ and for each $A \in \mathcal{F}$, $\pi_k(x_0, \cdots, x_{k-1}\,; A\,)$ is assumed to be a measurable function of $(x_0, \cdots, x_{k-1})$, on the space $(\mathcal{X}^k, \mathcal{F}^k)$ which is the product of k copies of the space $(\mathcal{X}, \mathcal{F})$ with itself. We can inductively define measures μ_k on $(\mathcal{X}^{k+1}, \mathcal{F}^{k+1})$ that describe the probability distribution of the entire history $(x_0, \cdots, x_k)$ of the system through time k. To go from μ_{k-1} to μ_k we think of $(\mathcal{X}^{k+1}, \mathcal{F}^{k+1})$ as the product of $(\mathcal{X}^k, \mathcal{F}^k)$ with $(\mathcal{X}, \mathcal{F})$ and construct on $(\mathcal{X}^{k+1}, \mathcal{F}^{k+1}\,)$ a probability measure with marginal μ_{k-1} on $(\mathcal{X}^k, \mathcal{F}^k)$ and conditionals $\pi_k(x_0, \cdots, x_{k-1}; \cdot)$ on the fibers $(x_1, \cdots, x_{k-1}) \times \mathcal{X}$. This will define μ_k and the induction can proceed. We may stop at some finite terminal time N or go on indefinitely. If we do go on indefinitely, we will have a consitent family of finite dimensional distributions $\{\mu_k\}$ on $(\mathcal{X}^{k+1}, \mathcal{F}^{k+1})$ and we may try to use Kolmogorov's theorem to construct a probability measure P on the space $(\mathcal{X}^\infty, \mathcal{F}^\infty)$ of sequences $\{x_j : j \geq 0\}$ representing the total evolution of the system for all times.

Remark 4.7. However Kolmogorov's theorem requires some assumptions on $(\mathcal{X}, \mathcal{F})$ that are satisfied if $\mathcal{X}$ is a complete separable metric space and $\mathcal{F}$ are the Borel sets. However, in the present context, there is a result known as Tulcea's theorem (see [8]) that proves the existence of a P on $(\mathcal{X}^\infty, \mathcal{F}^\infty)$ for any choice of $(\mathcal{X}, \mathcal{F})$, exploiting the fact that the consistent family of finite dimensional distributions μ_k arise from well defined successive regular conditional probability distributions.

An important subclass is generated when the transition probability depends on the past history only through the current state. In other words

$$\pi_k(x_0, \cdots, x_{k-1}\,; \cdot) = \pi_{k-1,k}\,(x_{k-1}\,; \cdot).$$

In such a case the process is called a Markov Process with transition probabilities $\pi_{k-1,k}(\cdot,\cdot)$. An even smaller subclass arises when we demand that $\pi_{k-1,k}(\cdot,\cdot)$ be the same for different values of k. A single transition probability $\pi(x,A)$ and the initial distribution μ_0 determine the entire process i.e. the measure P on $(\mathcal{X}^\infty,\mathcal{F}^\infty)$. Such processes are called time-homogeneous Markov Proceses or Markov Processes with stationary transition probabilities.

Chapman-Kolmogorov Equations:. If we have the transition probabilities $\pi_{k,k+1}$ of transition from time k to $k+1$ of a Markov Chain it is possible to obtain directly the transition probabilities from time k to $k+\ell$ for any $\ell \geq 2$. We do it by induction on ℓ. Define

$$\pi_{k,k+\ell+1}(x,A) = \int_{\mathcal{X}} \pi_{k,k+\ell}(x,dy)\,\pi_{k+\ell,k+\ell+1}(y,A) \tag{4.5}$$

or equivalently, in a more direct fashion

$$\pi_{k,k+\ell+1}(x,A) = \int_{\mathcal{X}} \cdots \int_{\mathcal{X}} \pi_{k,k+1}(x,dy_{k+1}) \cdots \pi_{k+\ell,k+\ell+1}(y_{k+\ell},A)$$

Theorem 4.8. *The transition probabilities $\pi_{k,m}(\cdot,\cdot)$ satisfy the relations*

$$\pi_{k,n}(x,A) = \int_{\mathcal{X}} \pi_{k,m}(x,dy)\,\pi_{m,n}(y,A) \tag{4.6}$$

for any $k < m < n$ and for the Markov Process defined by the one step transition probabilities $\pi_{k,k+1}(\cdot,\cdot)$, for any $n > m$

$$P[x_n \in A|\Sigma_m] = \pi_{m,n}(x_m,A) \quad \text{a.e.}$$

where Σ_m is the σ-field of past history upto time m generated by the coordinates $x_0, x_1, \cdots, x_m$.

Proof. The identity is basically algebra. The multiple integral can be carried out by iteration in any order and after enough variables are integrated we get our identity. To prove that the conditional probabilities are given by the right formula we need to establish

$$P[\{x_n \in A\} \cap B] = \int_B \pi_{m,n}(x_m,A)\,dP$$

for all $B \in \Sigma_m$ and $A \in \mathcal{F}$. We write

$$
\begin{aligned}
&P[\{x_n \in A\} \cap B] = \int_{\{x_n \in A\} \cap B} dP \\
&= \int \cdots \int_{\{x_n \in A\} \cap B} d\mu(x_0)\, \pi_{0,1}(x_0, dx_1) \cdots \pi_{m-1,m}(x_{m-1}, dx_m) \\
&\qquad\qquad \pi_{m,m+1}(x_m, dx_{m-1}) \cdots \pi_{n-1,n}(x_{n-1}, dx_n) \\
&= \int \cdots \int_B d\mu(x_0)\, \pi_{0,1}(x_0, dx_1) \cdots \pi_{m-1,m}(x_{m-1}, dx_m) \\
&\qquad\qquad \pi_{m,m+1}(x_m, dx_{m-1}) \cdots \pi_{n-1,n}(x_{n-1}, A) \\
&= \int \cdots \int_B d\mu(x_0)\, \pi_{0,1}(x_0, dx_1) \cdots \pi_{m-1,m}(x_{m-1}, dx_m)\, \pi_{m,n}(x_m, A) \\
&= \int_B \pi_{m,n}(x_m, A)\, dP
\end{aligned}
$$

and we are done. $\square$

Remark 4.8. If the chain has stationary transition probabilities then the transition probabilities $\pi_{m,n}(x, dy)$ from time m to time n depend only on the difference $k = n - m$ and are given by what are usually called the k step transition probabilities. They are defined inductively by

$$\pi^{(k+1)}(x, A) = \int_{\mathcal{X}} \pi^{(k)}(x, dy)\, \pi(y, A)$$

and satisfy the Chapman-Kolmogorov equations

$$\pi^{(k+\ell)}(x, A) = \int_{\mathcal{X}} \pi^{(k)}(x, dy) \pi^{(\ell)}(x, A) = \int_{\S} \pi^{(\ell)}(x, dy) \pi^{(k)}(y, A)$$

Suppose we have a probability measure P on the product space $X \times Y \times Z$ with the product σ-field. The Markov property in this context refers to equality

$$E^P[g(z)|\Sigma_{x,y}] = E^P[g(z)|\Sigma_y] \qquad a.e.\ P \tag{4.7}$$

for bounded measurable functions g on Z, where we have used $\Sigma_{x,y}$ to denote the σ-field generated by projection on to $X \times Y$ and Σ_y the corresponding

σ-field generated by projection on to Y. The Markov property in the reverse direction is the similar condition for bounded measurable functions f on X.

$$E^P[f(x)|\Sigma_{y,z}] = E^P[f(x)|\Sigma_y] \qquad a.e.\ P \tag{4.8}$$

They look different. But they are both equivalent to the symmetric condition

$$E^P[f(x)g(z)|\Sigma_y] = E^P[f(x)|\Sigma_y]E^P[g(z)|\Sigma_y] \qquad a.e.\ P \tag{4.9}$$

which says that given the present, the past and future are conditionally independent. In view of the symmetry it sufficient to prove the following:

Theorem 4.9. *For any P on $(X \times Y \times Z)$ the relations (4.7) and (4.9) are equivalent.*

Proof. Let us fix f and g. Let us denote the common value in (4.7) by $\hat{g}(y)$ Then

$$\begin{aligned}
E^P[f(x)g(z)|\Sigma_y] &= E^P[E^P[f(x)g(z)|\Sigma_{x,y}]|\Sigma_y] && a.e.\ P\\
&= E^P[f(x)E^P[g(z)|\Sigma_{x,y}]|\Sigma_y] && a.e.\ P\\
&= E^P[f(x)\hat{g}(y)|\Sigma_y] && a.e.\ P\ (by\ (4.5))\\
&= E^P[f(x)|\Sigma_y]\hat{g}(y) && a.e.\ P\\
&= E^P[f(x)|\Sigma_y]E^P[g(z)|\Sigma_y] && a.e.\ P
\end{aligned}$$

which is (4.9). Conversely, we assume (4.9) and denote by $\bar{g}(x,y)$ and $\hat{g}(y)$ the expressions on the left and right side of (4.7). Let $b(y)$ be a bounded measurable function on Y.

$$\begin{aligned}
E^P[f(x)b(y)\bar{g}(x,y)] &= E^P[f(x)b(y)g(z)]\\
&= E^P[b(y)E^P[f(x)g(z)|\Sigma_y]]\\
&= E^P[b(y)\big\{E^P[f(x)|\Sigma_y]\big\}\big\{E^P[g(z)|\Sigma_y]\big\}]\\
&= E^P[b(y)\big\{E^P[f(x)|\Sigma_y]\big\}\hat{g}(y)]\\
&= E^P[f(x)b(y)\hat{g}(y)].
\end{aligned}$$

Since f and b are arbitrary this implies that $\bar{g}(x,y) = \hat{g}(y)$ a.e. P. $\square$

Let us look at some examples.

1. Suppose we have an urn containg a certain number of balls (nonzero) some red and others green. A ball is drawn at random and its color is noted. Then it is returned to the urn along with an extra ball of the same color. Then a new ball is drawn at random and the process continues ad infinitum. The current state of the system can be characterized by two integers r, g such that $r + g \geq 1$. The initial state if the system is some r_0, g_0 with $r_0 + g_0 \geq 1$. The system can go from (r, g) to either $(r+1, g)$ with probability $\frac{r}{r+g}$ or to $(r, g+1)$ with probability $\frac{g}{r+g}$. This is clearly an example of a Markov Chain with stationary transition probabilities.

2. Consider a queue for service in a store. Suppose at each of the times $1, 2, \cdots$, a random number of new customers arrive and and join the queue. If the queue is non empty at some time, then exactly one customer will be served and will leave the queue at the next time point. The distribution of the number of new arrivals is specified by $\{p_j : j \geq 0\}$ where p_j is the probability that exactly j new customers arrive at a given time. The number of new arrivals at distinct times are assumed to be independent. The queue length is a Markov Chain on the state space $\mathcal{X} = \{0, 1, \cdots, \}$ of nonegative integers. The transition probabilities $\pi(i, j)$ are given by $\pi(0, j) = p_j$ because there is no service and nobody in the queue to begin with and all the new arrivals join the queue. On the other hand $\pi(i, j) = p_{j-i+1}$ if $j + 1 \geq i \geq 1$ because one person leaves the queue after being served.

3. Consider a reservoir into which water flows. The amount of additional water flowing into the reservoir on any given day is random, and has a distribution α on $[0, \infty)$. The demand is also random for any given day, with a probability distribution β on $[0, \infty)$. We may also assume that the inflows and demands on successive days are random variables ξ_n and η_n, that have α and β for their common distributions and are all mutually independent. We may wish to assume a percentage loss due to evaporation. In any case the storage level at successive days have a recurrence relation

$$S_{n+1} = [(1-p)S_n + \xi_n - \eta_n]^+$$

p is the loss and we have put the condition that the outflow is the demand unless the stored amount is less than the demand in which case the outflow is the available quantity. The current amount in storage is a Markov Process with Stationary transition probabilities.

4. Let $X_1, \cdots, X_n, \cdots$ be a sequence of independent random variables with a common distribution α. Let $S_n = Y + X_1 + \cdots + X_n$ for $n \geq 1$ with $S_0 = Y$ where Y is a random variable independent of $X_1, \ldots, X_n, \ldots$ with distribution μ. Then S_n is a Markov chain on $\mathbf{R}$ with one step transition probaility $\pi(x, A) = \alpha(A - x)$ and initial distribution μ. The n step transition probability is $\alpha^n(A - x)$ where α^n is the n-fold convolution of α. This is often referred to as a random walk.

The last two examples can be described by models of the type

$$x_n = f(x_{n-1}, \xi_n)$$

where x_n is the current state and ξ_n is some random external disturbance. ξ_n are assumed to be independent and identically distributed. They could have two components like inflow and demand. The new state is a deterministic function of the old state and the noise.

Exercise 4.11. Verify that the first two examples can be cast in the above form. In fact there is no loss of generality in assuming that ξ_j are mutually independent random variables having as common distribution the uniform distribution on the interval $[0, 1]$.

Given a Markov Chain with stationary transition probabilities $\pi(x, dy)$ on a state space $(\mathcal{X}, \mathcal{F})$, the behavior of $\pi^{(n)}(x, dy)$ for large n is an important and natural question. In the best situation of independent random variables $\pi^{(n)}(x, A) = \mu(A)$ are independent of x as well as n. Hopefully after a long time the Chain will 'forget' its origins and $\pi^{(n)}(x, \cdot) \to \mu(\cdot)$, in some suitable sense, for some μ that does not depend on x . If that happens, then from the relation

$$\pi^{(n+1)}(x, A) = \int \pi^{(n)}(x, dy)\pi(y, A),$$

we conclude

$$\mu(A) = \int \pi(y, A)\, d\mu(y) \quad \text{for all} \quad A \in \mathcal{F}$$

Measures that satisfy the above property, abbreviated as $\mu\pi = \mu$, are called invariant measures for the Markov Chain. If we start with the initial distribution μ which is inavariant then the probability measure P has μ as marginal at every time. In fact P is stationary i.e., invariant with respect to time translation, and can be extended to a stationary process where time runs from $-\infty$ to $+\infty$.

4.5 Stopping Times and Renewal Times

One of the important notions in the analysis of Markov Chains is the idea of stopping times and renewal times. A function

$$\tau(\omega) : \Omega \to \{n : n \geq 0\}$$

is a random variable defined on the set $\Omega = \mathcal{X}^\infty$ such that for every $n \geq 0$ the set $\{\omega : \tau(\omega) = n\}$ (or equivalently for each $n \geq 0$ the set $\{\omega : \tau(\omega) \leq n\}$) is measurable with respect to the σ-field $\mathcal{F}_n$ generated by $X_j : 0 \leq j \leq n$. It is not necessary that $\tau(\omega) < \infty$ for every ω. Such random variable τ are called stopping times. Examples of stopping times are, constant times $n \geq 0$, the first visit to a state x, or the second visit to a state x. The important thing is that in order to decide if $\tau \leq n$ i.e. to know if what ever is supposed to happen did happen before time n the chain need be observed only up to time n. Examples of τ that are not stopping times are easy to find. The last time a site is visited is not a stopping time nor is is the first time such that at the next time one is in a state x. An important fact is that the Markov property extends to stopping times. Just as we have σ-fields $\mathcal{F}_n$ associated with constant times, we do have a σ field $\mathcal{F}_\tau$ associated to any stopping time. This is the information we have when we observe the chain upto time τ. Formally

$$\mathcal{F}_\tau = \left\{ A : A \in \mathcal{F}^\infty \quad \text{and} \quad A \cap \{\tau \leq n\} \in \mathcal{F}_n \quad \text{for} \quad \text{each} \quad n \right\}$$

One can check from the definition that τ is $\mathcal{F}_\tau$ measurable and so is X_τ on the set $\tau < \infty$. If τ is the time of first visit to y then τ is a stopping time and the event that the chain visits a state z before visiting y is $\mathcal{F}_\tau$ measurable.

Lemma 4.10. (Strong Markov Property.) *At any stopping time τ the Markov property holds in the sense that the conditional distribution of*

$X_{\tau+1}, \cdots, X_{\tau+n}, \cdots$ conditioned on $\mathcal{F}_\tau$ is the same as the original chain starting from the state $x = X_\tau$ on the set $\tau < \infty$. In other words

$$\begin{aligned} &P_x\{X_{\tau+1} \in A_1, \cdots, X_{\tau+n} \in A_n | \mathcal{F}_\tau\} \\ &\quad = \int_{A_1} \cdots \int_{A_n} \pi(X_\tau, dx_1) \cdots \pi(x_{n-1}, dx_n) \end{aligned}$$

a.e. on $\{\tau < \infty\}$.

Proof. Let $A \in \mathcal{F}_\tau$ be given with $A \subset \{\tau < \infty\}$. Then

$$\begin{aligned} &P_x\{A \cap \{X_{\tau+1} \in A_1, \cdots, X_{\tau+n} \in A_n\}\} \\ &\qquad = \sum_k P_x\{A \cap \{\tau = k\} \cap \{X_{k+1} \in A_1, \cdots, X_{k+n} \in A_n\}\} \\ &\qquad = \sum_k \int_{A \cap \{\tau = k\}} \int_{A_1} \cdots \int_{A_n} \pi(X_k, dx_{k+1}) \cdots \pi(x_{k+n-1}, dx_{k+n})\, dP_x \\ &\qquad = \int_A \int_{A_1} \cdots \int_{A_n} \pi(X_\tau, dx_1) \cdots \pi(x_{n-1}, dx_n)\, dP_x \end{aligned}$$

We have used the fact that if $A \in \mathcal{F}_\tau$ then $A \cap \{\tau = k\} \in \mathcal{F}_k$ for every $k \geq 0$. $\square$

Remark 4.9. If $X_\tau = y$ a.e. with respect to P_x on the set $\tau < \infty$, then at time τ, when it is finite, the process starts afresh with no memory of the past and will have conditionally the same probabilities in the future as P_y. At such times the process renews itself and these times are called renewal times.

4.6 Countable State Space

From the point of view of analysis a particularly simple situation is when the state space $\mathcal{X}$ is a countable set. It can be taken as the integers $\{x : x \geq 1\}$. Many applications fall in this category and an understanding of what happens in this situation will tell us what to expect in general.

The one step transition probability is a matrix $\pi(x, y)$ with nonnegative entries such that $\sum_y \pi(x, y) = 1$ for each x. Such matrices are called stochastic matrices. The n step transition matrix is just the n-th power of the matrix defined inductively by

$$\pi^{(n+1)}(x,y) = \sum_z \pi^{(n)}(x,z)\pi(z,y)$$

To be consistent one defines $\pi^{(0)}(x,y) = \delta_{x,y}$ which is 1 if $x = y$ and 0 otherwise. The problem is to analyse the behaviour for large n of $\pi^{(n)}(x,y)$. A state x is said to communicate with a state y if $\pi^{(n)}(x,y) > 0$ for some $n \geq 1$. We will assume for simplicity that every state communicates with every other state. Such Markov Chains are called **irreducible**. Let us first limit ourselves to the study of irreducible chains. Given an irreducible Markov chain with transition probabilities $\pi(x,y)$ we define $f_n(x)$ as the probability of returning to x for the first time at the n-th step assuming that the chain starts from the state x.. Using the convention that P_x refers to the measure on sequences for the chain starting from x and $\{X_j\}$ are the successive positions of the chain

$$\begin{aligned} f_n(x) &= P_x\Big\{X_j \neq x \ \text{ for } \ 1 \leq j \leq n-1 \ \text{ and } \ X_n = x\Big\} \\ &= \sum_{\substack{y_1 \neq x \\ \cdots \\ y_{n-1} \neq x}} \pi(x,y_1)\,\pi(y_1,y_2)\cdots\pi(y_{n-1},x) \end{aligned}$$

Since $f_n(x)$ are probailities of disjoint events $\sum_n f_n(x) \leq 1$. The state x is called **transient** if $\sum_n f_n(x) < 1$ and **recurrent** if $\sum_n f_n(x) = 1$. The recurrent case is divided into two situations. If we denote by $\tau_x = \inf\{n \geq 1 : X_n = x\}$, the time of first visit to x, then recurrence is $P_x\{\tau_x < \infty\} = 1$. A recurrent state x is called **positive recurrent** if

$$E^{P_x}\{\tau_x\} = \sum_{n \geq 1} n\, f_n(x) < \infty$$

and **null recurrent** if

$$E^{P_x}\{\tau_x\} = \sum_{n \geq 1} n\, f_n(x) = \infty$$

Lemma 4.11. *If for a (not necessarily irreducible) chain starting from x, the probability of ever visiting y is positive then so is the probability of visiting y before returning to x.*

Proof. Assume that for the chain starting from x the probability of visiting y before returning to x is zero. But when it returns to x it starts afresh and so will not visit y until it returns again. This reasoning can be repeated and so the chain will have to visit x infinitely often before visiting y. But this will use up all the time and so it cannot visit y at all. $\square$

Lemma 4.12. *For an irreducible chain all states x are of the same type.*

Proof. Let x be recurrent and y be given. Since the chain is irreducible, for some k, $\pi^{(k)}(x,y) > 0$. By the previous lemma, for the chain starting from x, there is a positive probability of visiting y before returning to x. After each successive return to x, the chain starts afresh and there is a fixed positive probability of visiting y before the next return to x. Since there are infinitely many returns to x, y will be visited infinitely many times as well. Or y is also a recurrent state.

We now prove that if x is positive recurrent then so is y. We saw already that the probability $p = P_x\{\tau_y < \tau_x\}$ of visiting y before returning to x is positive. Clearly

$$E^{P_x}\{\tau_x\} \geq P_x\{\tau_y < \tau_x\}\, E^{P_y}\{\tau_x\}$$

and therefore

$$E^{P_y}\{\tau_x\} \leq \frac{1}{p} E^{P_x}\{\tau_x\} < \infty.$$

On the other hand we can write

$$\begin{aligned}
E^{P_x}\{\tau_y\} &\leq \int_{\tau_y < \tau_x} \tau_x \, dP_x + \int_{\tau_x < \tau_y} \tau_y \, dP_x \\
&= \int_{\tau_y < \tau_x} \tau_x \, dP_x + \int_{\tau_x < \tau_y} \{\tau_x + E^{P_x}\{\tau_y\}\} \, dP_x \\
&= \int_{\tau_y < \tau_x} \tau_x \, dP_x + \int_{\tau_x < \tau_y} \tau_x \, dP_x + (1-p)\, E^{P_x}\{\tau_y\} \\
&= \int \tau_x \, dP_x + (1-p)\, E^{P_x}\{\tau_y\}
\end{aligned}$$

by the renewal property at the stopping time τ_x. Therefore

$$E^{P_x}\{\tau_y\} \leq \frac{1}{p} E^{P_x}\{\tau_x\}.$$

We also have

$$E^{P_y}\{\tau_y\} \leq E^{P_y}\{\tau_x\} + E^{P_x}\{\tau_y\} \leq \frac{2}{p} E^{P_x}\{\tau_x\}$$

proving that y is positive recurrent. $\square$

Transient Case: We have the following theorem regarding transience.

Theorem 4.13. *An irreducible chain is transient if and only if*

$$G(x,y) = \sum_{n=0}^{\infty} \pi^{(n)}(x,y) < \infty \quad \text{for all} \quad x, y.$$

Moreover for any two states x and y,

$$G(x,y) = f(x,y)G(y,y)$$

and

$$G(x,x) = \frac{1}{1 - f(x,x)}$$

where $f(x,y) = P_x\{\tau_y < \infty\}$.

Proof. Each time the chain returns to x there is a probability $1 - f(x,x)$ of never returning. The number of returns has then the geometric distribution

$$P_x\{\text{ exactly } \ n \ \text{ returns to } \ x\} = (1 - f(x,x))f(x,x)^n$$

and the expected number of returns is given by

$$\sum_{k=1}^{\infty} \pi^{(k)}(x,x) = \frac{f(x,x)}{1 - f(x,x)} \ .$$

The left hand side comes from the calculation

$$E^{P_x} \sum_{k=1}^{\infty} \chi_{\{x\}}(X_k) = \sum_{k=1}^{\infty} \pi^{(k)}(x,x)$$

and the right hand side from the calculation of the mean of a Geometric distribution. Since we count the visit at time 0 as a visit to x we add 1 to both sides to get our formula. If we want to calculate the expected number of visits to y when we start from x, first we have to get to y and the probability of that is $f(x,y)$. Then by the renewal property it is exactly the same as the expected number of visits to y starting from y, including the visit at time 0 and that equals $G(y,y)$. $\square$

Before we study the recurrent behavior we need the notion of periodicity. For each state x let us define $D_x = \{n : \pi^{(n)}(x,x) > 0\}$ to be the set of times at which a return to x is possible if one starts from x. We define d_x to be the greatest common divisor of D_x.

Lemma 4.14. *For any irreducible chain $d_x = d$ for all $x \in \mathcal{X}$ and for each x, D_x contains all sufficiently large multiples of d.*

Proof. Let us define
$$D_{x,y} = \{n : \pi^{(n)}(x,y) > 0\}$$
so that $D_x = D_{x,x}$. By the Chapman-Kolmogorov equations

$$\pi^{(m+n)}(x,y) \geq \pi^{(m)}(x,z)\pi^{(n)}(z,y)$$

for every z, so that if $m \in D_{x,z}$ and $n \in D_{z,y}$, then $m+n \in D_{x,y}$. In particular if $m,n \in D_x$ it follows that $m+n \in D_x$. Since any pair of states communicate with each other, given $x,y \in \mathcal{X}$, there are positive integers n_1 and n_2 such that $n_1 \in D_{x,y}$ and $n_2 \in D_{y,x}$. This implies that with the choice of $\ell = n_1 + n_2$, $n + \ell \in D_x$ whenever $n \in D_y$; similarly $n + \ell \in D_y$ whenever $n \in D_x$. Since ℓ itself belongs to both D_x and D_y both d_x and d_y divide ℓ. Suppose $n \in D_x$. Then $n + \ell \in D_y$ and therefore d_y divides $n + \ell$. Since d_y divides ℓ, d_y must divide n. Since this is true for every $n \in D_x$ and d_x is the greatest common divisor of D_x, d_y must divide d_x. Similarly d_x must divide d_y. Hence $d_x = d_y$. We now complete the proof of the lemma. Let d be the greatest common divisor of D_x. Then it is the greatest common divisor of a finite subset $n_1, n_2, \cdots, n_q$ of D_x and there will exist integers $a_1, a_2, \cdots, a_q$ such that
$$a_1 n_1 + a_2 n_2 + \cdots + a_q n_q = d$$
Some of the a's will be positive and others negative. Seperating them out, and remembering that all the n_i are divisible by d, we find two integers $md, (m+1)d$ such that they both belong to D_x. If now $n = kd$ with $k > m^2$ we can write $k = \ell m + r$ with a large $\ell \geq m$ and the remainder r is less than m.

$$kd = (\ell m + r)d = \ell md + r(m+1)d - rmd = (\ell - r)md + r(m+1)d \in D_x$$

since $\ell \geq m > r$. □

Remark 4.10. For an irreducible chain the common value d is called the period of the chain and an irreducible chain with period $d = 1$ is called aperiodic.

The simplest example of a periodic chain is one with 2 states and the chain shuttles back and forth between the two. $\pi(x, y) = 1$ if $x \neq y$ and 0 if $x = y$. A simple calculation yields $\pi^{(n)}(x, x) = 1$ if n is even and 0 otherwise. There is oscillatory behavior in n that persists. The main theorem for irreducible, aperiodic, recurrent chains is the following.

Theorem 4.15. *Let $\pi(x, y)$ be the one step transition probability for a recurrent aperiodic Markov chain and let $\pi^{(n)}(x, y)$ be the n-step transition probabilities. If the chain is null recurrent then*

$$\lim_{n\to\infty} \pi^{(n)}(x, y) = 0 \quad \text{for all} \quad x, y$$

If the chain is positive recuurrent then of course $E^{P_x}\{\tau_x\} = m(x) < \infty$ for all x, and in that case

$$\lim_{n\to\infty} \pi^{(n)}(x, y) = q(y) = \frac{1}{m(y)}$$

exist for all x and y is independent of the starting point x and $\sum_y q(y) = 1$.

The proof is based on

Theorem 4.16. (Renewal Theorem.) *Let $\{f_n : n \geq 1\}$ be a sequence of nonnegative numbers such that*

$$\sum_n f_n = 1, \qquad \sum_n n\, f_n = m \leq \infty$$

and the greatest common divisor of $\{n : f_n > 0\}$ is 1. Suppose that $\{p_n : n \geq 0\}$ are defined by $p_0 = 1$ and recursively

$$p_n = \sum_{j=1}^{n} f_j\, p_{n-j} \tag{4.10}$$

Then

$$\lim_{n\to\infty} p_n = \frac{1}{m}$$

where if $m = \infty$ the right hand side is taken as 0.

Proof. The proof is based on several steps.

Step 1. We have inductively $p_n \leq 1$. Let $a = \limsup_{n\to\infty} p_n$. We can choose a subsequence n_k such that $p_{n_k} \to a$. We can assume without loss of generality that $p_{n_k+j} \to q_j$ as $k \to \infty$ for all positive and negative integers j as well. Of course the limit q_0 for $j = 0$ is a. In relation 4.10 we can pass to the limit along the subsequence and use the dominated convergence theorem to obtain

$$q_n = \sum_{j=1}^{\infty} f_j \, q_{n-j} \tag{4.11}$$

valid for $-\infty < n < \infty$. In particular

$$q_0 = \sum_{j=1}^{\infty} f_j \, q_{-j} \tag{4.12}$$

Step 2: Because $a = \limsup p_n$ we can conclude that $q_j \leq a$ for all j. If we denote by $S = \{n : f_n > 0\}$ then $q_{-k} = a$ for $k \in S$. We can then deduce from equation 4.11 that $q_{-k} = a$ for $k = k_1 + k_2$ with $k_1, k_2 \in S$. By repeating the same reasoning $q_{-k} = a$ for $k = k_1 + k_2 + \cdots + k_\ell$. By lemma 3.6 because the greatest common factor of the integers in S is 1, there is a k_0 such that for $k \geq k_0$,we have $q_{-k} = a$. We now apply the relation 4.11 again to conclude that $q_j = a$ for all positve as well as negative j.

Step 3: If we add up equation 4.10 for $n = 1, \cdots, N$ we get

$$\begin{aligned} p_1 + p_2 + \cdots + p_N &= (f_1 + f_2 + \cdots + f_N) + (f_1 + f_2 + \cdots + f_{N-1})p_1 \\ &+ \cdots + (f_1 + f_2 + \cdots + f_{N-k})p_k + \cdots + f_1 p_{N-1} \end{aligned}$$

If we denote by $T_j = \sum_{i=j}^{\infty} f_i$, we have $T_1 = 1$ and $\sum_{j=0}^{\infty} T_j = m$. We can now rewrite

$$\sum_{j=1}^{N} T_j p_{N-j+1} = \sum_{j=1}^{N} f_j$$

Step 4: Because $p_{N-j} \to a$ for every j along the subsequence $N = n_k$, if $\sum_j T_j = m < \infty$, we can deduce from the dominated convergence theorem that $m\,a = 1$ and we conclude that

$$\limsup_{n\to\infty} p_n = v1m$$

If $\sum_j T_j = \infty$, by Fatou's Lemma $a = 0$. Exactly the same argument applies to liminf and we conclude that

$$\liminf_{n\to\infty} p_n = \frac{1}{m}$$

This conclues the proof of the renewal theorem. $\square$

We now turn to

Proof. (of Theorem 4.15). If we take a fixed $x \in \mathcal{X}$ and consider $f_n = P_x\{\tau_x = n\}$, then f_n and $p_n = \pi^{(n)}(x,x)$ are related by (1) and $m = E^{P_x}\{\tau_x\}$. In order to apply the renewal theorem we need to establish that the greatest common divisor of $S = \{n : f_n > 0\}$ is 1. In general if $f_n > 0$ so is p_n. So the greatest common divisor of S is always larger than that of $\{n : p_n > 0\}$. That does not help us because the greatest common divisor of $\{n : p_n > 0\}$ is 1. On the other hand if $f_n = 0$ unless $n = k\,d$ for some k, the relation 4.10 can be used inductively to conclude that the same is true of p_n. Hence both sets have the same greatest common divisor. We can now conclude that

$$\lim_{n\to\infty} \pi^{(n)}(x,x) = q(x) = \frac{1}{m(x)}$$

On the other hand if $f_n(x,y) = P_x\{\tau_y = n\}$, then

$$\pi^{(n)}(x,y) = \sum_{k=1}^{n} f_k(x,y)\,\pi^{(n-k)}(y,y)$$

and recurrence implies $\sum_{k+1}^{\infty} f_k(x,y) = 1$ for all x and y. Therefore

$$\lim_{n\to\infty} \pi^{(n)}(x,y) = q(y) = \frac{1}{m(y)}$$

and is independent of x, the starting point. In order to complete the proof we have to establish that

$$Q = \sum_y q(y) = 1$$

It is clear by Fatou's lemma that

$$\sum_y q(y) = Q \leq 1$$

By letting $n \to \infty$ in the Chapman-Kolmogorov equation

$$\pi^{(n+1)}(x,y) = \sum_z \pi^n(x,z)\pi(z,y)$$

and using Fatou's lemma we get

$$q(y) \geq \sum_z \pi(z,y)q(z)$$

Summing with repect to y we obtain

$$Q \geq \sum_{z,y} \pi(z,y)q(z) = Q$$

and equality holds in this relation. Therefore

$$q(y) = \sum_z \pi(z,y)q(z)$$

for every y or $q(\cdot)$ is an invariant measure. By iteration

$$q(y) = \sum_z \pi^n(z,y)q(z)$$

and if we let $n \to \infty$ again an application of the bounded convergence theorem yields

$$q(y) = Q\, q(y)$$

implying $Q = 1$ and we are done. $\square$

Let us now consider an irreducible Markov Chain with one step transition probability $\pi(x,y)$ that is periodic with period $d > 1$. Let us choose and fix a reference point $x_0 \in \mathcal{X}$. For each $x \in \mathcal{X}$ let $D_{x_0,x} = \{n : \pi^{(n)}(x_0,x) > 0\}$.

Lemma 4.17. *If $n_1, n_2 \in D_{x_0,x}$ then d divides $n_1 - n_2$.*

Proof. Since the chain is irreducible there is an m such tha $\pi^{(m)}(x,x_0) > 0$. By the Chapman-Kolmogorov equations $\pi^{(m+n_i)}(x_0,x_0) > 0$ for $i = 1,2$. Therefore $m + n_i \in D_{x_0} = D_{x_0,x_0}$ for $i = 1,2$. This implies that d divides both $m + n_1$ as well as $m + n_2$. Thus d divides $n_1 - n_2$. $\square$

The residue modulo d of all the integers in $D_{x_0,x}$ are the same and equal some number $r(x)$, satisfying $0 \leq r(x) \leq d-1$. By definition $r(x_0) = 0$. Let us define $\mathcal{X}_j = \{x : r(x) = j\}$. Then $\{\mathcal{X}_j : 0 \leq j \leq d-1\}$ is a partition of $\mathcal{X}$ into disjoint sets with $x_0 \in \mathcal{X}_0$.

Lemma 4.18. *If $x \in \mathcal{X}$, then $\pi^{(n)}(x,y) = 0$ unless $r(x) + n = r(y) \mod d$.*

Proof. Suppose that $x \in \mathcal{X}$ and $\pi(x,y) > 0$. Then if $m \in D_{x_0,x}$ then $(m+1) \in D_{x_0,y}$. Therefore $r(x) + 1 = r(y)$ modulo d. The proof can be completed by induction. The chain marches through $\{\mathcal{X}_j\}$ in a cyclical way from a state in $\mathcal{X}_j$ to one in $\mathcal{X}_{j+1}$ □

Theorem 4.19. *Let $\mathcal{X}$ be irreducible and positive recurrent with period d. Then*

$$\lim_{\substack{n\to\infty \\ n+r(x)=r(y) \text{ modulo } d}} \pi^{(n)}(x,y) = \frac{d}{m(y)}$$

Of course

$$\pi^{(n)}(x,y) = 0$$

unless $n + r(x) = r(y)$ modulo d.

Proof. If we replace π by $\tilde{\pi}$ where $\tilde{\pi}(x,y) = \pi^{(d)}(x,y)$, then $\tilde{\pi}(x,y) = 0$ unless both x and y are in the same $\mathcal{X}_j$. The restriction of $\tilde{\pi}$ to each $\mathcal{X}_j$ defines an irreducible aperiodic Markov chain. Since each time step under $\tilde{\pi}$ is actually d units of time we can apply the earlier results and we will get for $x, y \in \mathcal{X}_j$ for some j,

$$\lim_{k\to\infty} \pi^{(k\,d)}(x,y) = \frac{d}{m(y)}$$

We note that

$$\pi^{(n)}(x,y) = \sum_{1\leq m\leq n} f_m(x,y)\,\pi^{(n-m)}(y,y)$$

$$\begin{aligned} f_m(x,y) = P_x\{\tau_y = m\} &= 0 \quad \text{unless} \quad r(x)+m = r(y) \quad \text{modulo} \quad d \\ \pi^{(n-m)}(y,y) &= 0 \quad \text{unless} \quad n-m = 0 \quad \text{modulo} \quad d \\ \sum_m f_m(x,y) &= 1. \end{aligned}$$

The theorem now follows. □

Suppose now we have a chain that is not irreducible. Let us collect all the transient states and call the set $\mathcal{X}_{tr}$. The complement consists of all the recurrent states and will be denoted by $\mathcal{X}_{re}$.

Lemma 4.20. *If $x \in \mathcal{X}_{re}$ and $y \in \mathcal{X}_{tr}$, then $\pi(x, y) = 0$.*

Proof. If x is a recuurrent state, and $\pi(x, y) > 0$, the chain will return to x infinitely often and each time there is a positive probability of visiting y. By the renewal property these are independent events and so y will be recurrent too. ☐

The set of recurrent states $\mathcal{X}_{re}$ can be divided into one or more equivalence classes accoeding to the following procedure. Two recurrent states x and y are in the same equivalence class if $f(x, y) = P_x\{\tau_y < \infty\}$, the probability of ever visiting y starting from x is positive. Because of recurrence if $f(x, y) > 0$ then $f(x, y) = f(y, x) = 1$. The restriction of the chain to a single equivalence class is irreducible and possibly periodic. Different equivalence classes could have different periods, some could be positive recurrent and others null recurrent. We can combine all our observations into the following theorem.

Theorem 4.21. *If y is transient then $\sum_n \pi^{(n)}(x, y) < \infty$ for all x. If y is null recurrent (belongs to an equivalence class that is null recurrent) then $\pi^{(n)}(x, y) \to 0$ for all x, but $\sum_n \pi^{(n)}(x, y) = \infty$ if x is in the same equivalence class or $x \in \mathcal{X}_{tr}$ with $f(x, y) > 0$. In all other cases $\pi^{(n)}(x, y) = 0$ for all $n \geq 1$. If y is positive recurrent and belongs to an equivalence class with period d with $m(y) = E^{P_y}\{\tau_y\}$, then for a nontransient x, $\pi^{(n)}(x, y) = 0$ unless x is in the same equivalence class and $r(x) + n = r(y)$ modulo d. In such a case,*

$$\lim_{\substack{n \to \infty \\ r(x)+n=r(y) \text{ modulo } d}} \pi^{(n)}(x, y) = \frac{d}{m(y)} .$$

If x is transient then

$$\lim_{\substack{n \to \infty \\ n=r \text{ modulo } d}} \pi^{(n)}(x, y) = f(r, x, y) \frac{d}{m(y)}$$

where

$$f(r, x, y) = P_x\{X_{kd+r} = y \quad \text{for some} \quad k \geq 0\}.$$

Proof. The only statement that needs an explanation is the last one. The chain starting from a transient state x may at some time get into a positive recurrent equivalence class $\mathcal{X}_j$ with period d. If it does, it never leaves that class and so gets absorbed in that class. The probability of this is $f(x,y)$ where y can be any state in $\mathcal{X}_j$. However if the period d is greater than 1, there will be cyclical subclasses $C_1, \cdots, C_d$ of $\mathcal{X}_j$. Depending on which subclass the chain enters and when, the phase of its future is determined. There are d such possible phases. For instance, if the subclasses are ordered in the correct way, getting into C_1 at time n is the same as getting into C_2 at time $n+1$ and so on. $f(r,x,y)$ is the probability of getting into the equivalence class in a phase that visits the cyclical subclass containing y at times n that are equal to r modulo d. $\square$

Example 4.1. (Simple Random Walk).

If $\mathcal{X} = Z^d$, the integral lattice in R^d, a random walk is a Markov chain with transition probability $\pi(x,y) = p(y-x)$ where $\{p(z)\}$ specifies the probability distribution of a single step. We will assume for simplicity that $p(z) = 0$ except when $z \in F$ where F consists of the $2\,d$ neighbors of 0 and $p(z) = \frac{1}{2d}$ for each $z \in F$. For $\xi \in R^d$ the characteristic function of $\hat{p}(\xi)$ of $p(\cdot)$ is given by $\frac{1}{d}(\cos\xi_1 + \cos\xi_2 + \cdots + \cos\xi_d)$. The chain is easily seen to irreducible, but periodic of period 2. Return to the starting point is possible only after an even number of steps.

$$\begin{aligned} \pi^{(2n)}(0,0) &= (\frac{1}{2\pi})^d \int_{\mathbf{T}^d} [\hat{p}(\xi)]^{2n} d\xi \\ &\simeq \frac{C}{n^{\frac{d}{2}}}. \end{aligned}$$

To see this asymptotic behavior let us first note that the integration can be restricted to the set where $|\hat{p}(\xi)| \geq 1-\delta$ or near the 2 points $(0,0,\cdots,0)$ and $(\pi,\pi,\cdots,\pi)$ where $|\hat{p}(\xi)| = 1$. Since the behaviour is similar at both points let us concentrate near the origin.

$$\frac{1}{d}\sum_{j=1}^{d} \cos\xi_j \leq 1 - c\sum_j \xi_j^2 \leq \exp[-c\sum_j \xi_j^2]$$

for some $c > 0$ and

$$[\frac{1}{d}\sum_{j=1}^{d} \cos\xi_j]^{2n} \leq \exp[-2\,n\,c\sum_j \xi_j^2]$$

and with a change of variables the upper bound is clear. We have a similar lower bound as well. The random walk is recurrent if $d = 1$ or 2 but transient if $d \geq 3$.

Exercise 4.12. If the distribution $p(\cdot)$ is arbitrary, determine when the chain is irreducible and when it is irreducible and aperiodic.

Exercise 4.13. If $\sum_z zp(z) = m \neq 0$ conclude that the chain is transient by an application of the strong law of large numbers.

Exercise 4.14. If $\sum_z zp(z) = m = 0$, and if the covariance matrix given by , $\sum_z z_i z_j p(z) = \sigma_{i,j}$, is nondegenerate show that the transience or recurrence is determined by the dimension as in the case of the nearest neighbor random walk.

Exercise 4.15. Can you make sense of the formal calculation

$$\begin{aligned}\sum_n \pi^{(n)}(0,0) &= \sum_n (\frac{1}{2\pi})^d \int_{\mathbf{T}^d} [\hat{p}(\xi)]^n d\xi \\ &= (\frac{1}{2\pi})^d \int_{\mathbf{T}^d} \frac{1}{(1-\hat{p}(\xi))} d\xi \\ &= (\frac{1}{2\pi})^d \int_{\mathbf{T}^d} \mathcal{R}eal\ \mathcal{P}art \left[\frac{1}{1-\hat{p}(\xi)}\right] d\xi\end{aligned}$$

to conclude that a necessary and sufficient condition for transience or recurrece is the convergence or divergence of the integral

$$\int_{\mathbf{T}^d} \mathcal{R}eal\ \mathcal{P}art \left[\frac{1}{1-\hat{p}(\xi)}\right] d\xi$$

with an integrand

$$\mathcal{R}eal\ \mathcal{P}art \left[\frac{1}{1-\hat{p}(\xi)}\right]$$

that is seen to be nonnegative ?

Hint: Consider instead the sum

$$\begin{aligned}\sum_{n=0}^{\infty} \rho^n \pi^{(n)}(0,0) &= \sum_n (\frac{1}{2\pi})^d \int_{\mathbf{T}^d} \rho^n [\hat{p}(\xi)]^n d\xi \\ &= (\frac{1}{2\pi})^d \int_{\mathbf{T}^d} \frac{1}{(1-\rho\hat{p}(\xi))} d\xi\end{aligned}$$

$$= (\frac{1}{2\pi})^d \int_{\mathbf{T}^d} \mathcal{R}eal\ \mathcal{P}art\left[\frac{1}{1-\rho\hat{p}(\xi)}\right] d\xi$$

for $0 < \rho < 1$ and let $\rho \to 1$.

Example 4.2. (The Queue Problem).

In the example of customers arriving, except in the trivial cases of $p_0 = 0$ or $p_0 + p_1 = 1$ the chain is irreducible and aperiodoc. Since the service rate is at most 1 if the arrival rate $m = \sum_j j\, p_j > 1$, then the queue will get longer and by an application of law of large numbers it is seen that the queue length will become infinite as time progresses. This is the transient behavior of the queue. If $m < 1$, one can expect the situation to be stable and there should be an asymptotic distribution for the queue length. If $m = 1$, it is the borderline case and one should probably expect this to be the null recurent case. The actual proofs are not hard. In time n the actual number of customers served is at most n because the queue may sometomes be empty. If $\{\xi_i : i \ge 1\}$ are the number of new customers arriving at time i and X_0 is the initial number in the queue, then the number X_n in the queue at time n satisfies $X_n \ge X_0 + (\sum_{i=1}^n \xi_i) - n$ and if $m > 1$, it follows from the law of large numbers that $\lim_{n\to\infty} X_n = +\infty$, thereby establishing transience. To prove positive recurrence when $m < 1$ it is sufficient to prove that the equations

$$\sum_x q(x)\pi(x,y) = q(y)$$

has a nontrivial nonnegative solution such that $\sum_x q(x) < \infty$. We shall proceed to show that this is indeed the case. Since the equation is linear we can alaways normalize tha solution so that $\sum_x q(x) = 1$. By iteration

$$\sum_x q(x)\pi^{(n)}(x,y) = q(y)$$

for every n. If $\lim_{n\to\infty} \pi^{(n)}(x,y) = 0$ for every x and y, because $\sum_x q(x) = 1 < \infty$, by the bounded convergence theorem the right hand side tends to 0 as $n \to \infty$. therefore $q \equiv 0$ and is trivial. This rules out the transient and the null recurrent case. In our case $\pi(0,y) = p_y$ and $\pi(x,y) = p_{y-x+1}$ if $y \ge x-1$ and $x \ge 1$. In all other cases $\pi(x,y) = 0$. The equations for $\{q_x = q(x)\}$ are then

$$q_0 p_y + \sum_{x=1}^{y+1} q_x p_{y-x+1} = q_y \quad \text{for} \quad y \ge 0. \tag{4.13}$$

Multiplying equation 4.13 by z^n and summimg from 1 to ∞, we get

$$q_0\, P(z) + \frac{1}{z}\, P(z)\, [Q(z) - q_0] = Q(z)$$

where $P(z)$ and $Q(z)$ are the generating functions

$$P(z) = \sum_{x=0}^{\infty} p_x z^x$$
$$Q(z) = \sum_{x=0}^{\infty} q_x z^x.$$

We can solve for Q to get

$$\begin{aligned}
\frac{Q(z)}{q_0} &= P(z)\left[1 - \frac{P(z)-1}{z-1}\right]^{-1} \\
&= P(z) \sum_{k=0}^{\infty} \left[\frac{P(z)-1}{z-1}\right]^k \\
&= P(z) \sum_{k=0}^{\infty} \left[\sum_{j=1}^{\infty} p_j (1 + z + \cdots + z^{j-1})\right]^k
\end{aligned}$$

is a power series in z with nonnegative coefficients. If $m < 1$, we can let $z \to 1$ to get

$$\frac{Q(1)}{q_0} = \sum_{k=0}^{\infty} \left[\sum_{j=1}^{\infty} j\, p_j\right]^k = \sum_{k=0}^{\infty} m^k = \frac{1}{1-m} < \infty$$

proving positive recurrence.

The case $m = 1$ is a little bit harder. The calculations carried out earlier are still valid and we know in this case that there exists $q(x) \geq 0$ such that each $q(x) < \infty$ for each x, $\sum_x q(x) = \infty$, and

$$\sum_x q(x)\, \pi(x, y) = q(y).$$

In other words the chain admits an infinite invariant measure. Such a chain cannot be positive recurrent. To see this we note

$$q(y) = \sum_x \pi^{(n)}(x, y) q(x)$$

and if the chain were positive recurrent

$$\lim_{n\to\infty} \pi^{(n)}(x,y) = \tilde{q}(y)$$

would exist and $\sum_y \tilde{q}(y) = 1$. By Fatou's lemma

$$q(y) \geq \sum_x \tilde{q}(y)q(x) = \infty$$

giving us a contradiction. To decide between transience and null recurrence a more detiled investigation is needed. We will outline a general procedure.

Suppose we have a state x_0 that is fixed and would like to calculate $F_{x_0}(\ell) = P_{x_0}\{\tau_{x_0} \leq \ell\}$. If we can do this, then we can answer questions about transience, recurrence etc. If $\lim_{\ell\to\infty} F_{x_0}(\ell) < 1$ then the chain is transient and otherwise recurrent. In the recurrent case the convergence or divergence of

$$E^{P_{x_0}}\{\tau_{x_0}\} = \sum_\ell [1 - F_{x_0}(\ell)]$$

determines if it is positive or null recurrent. If we can determine

$$F_y(\ell) = P_y\{\tau_{x_0} \leq \ell\}$$

for $y \neq x_0$, then for $\ell \geq 1$

$$F_{x_0}(\ell) = \pi(x_0, x_0) + \sum_{y \neq x_0} \pi(x_0, y) F_y(\ell - 1).$$

We shall outline a procedure for determining for $\lambda > 0$,

$$U(\lambda, y) = E_y\big[\exp[-\lambda\tau_{x_0}]\big].$$

Clearly $U(x) = U(\lambda, x)$ satisfies

$$U(x) = e^{-\lambda} \sum_y \pi(x,y)U(y) \quad \text{for} \quad x \neq x_0 \tag{4.14}$$

and $U(x_0) = 1$. One would hope that if we solve for these equations then we have our U. This requires uniqueness. Since our U is bounded in fact by 1, it is sufficient to prove uniqueness within the class of bounded solutions of equation 4.14. We will now establish that any bounded solution U of equation 4.14 with $U(x_0) = 1$, is given by

$$U(y) = U(\lambda, y) = E_y\big[\exp[-\lambda \tau_{x_0}]\big].$$

Let us define $E_n = \{X_1 \neq x_0, X_2 \neq x_0, \cdots, X_{n-1} \neq x_0, X_n = x_0\}$. Then we will prove, by induction, that for any solution U of equation (3.7), with $U(\lambda, x_0) = 1$,

$$U(y) = \sum_{j=1}^{n} e^{-\lambda j} P_y\{E_j\} + e^{-\lambda n} \int_{\tau_{x_0} > n} U(X_n)\, dP_y. \tag{4.15}$$

By letting $n \to \infty$ we would obtain

$$U(y) = \sum_{j=1}^{\infty} e^{-\lambda j} P_y\{E_j\} = E^{P_y}\{e^{-\lambda \tau_{x_0}}\}$$

because U is bounded and $\lambda > 0$.

$$\begin{aligned}
&\int_{\tau_{x_0} > n} U(X_n)\, dP_y \\
&\quad = e^{-\lambda} \int_{\tau_{x_0} > n} [\sum_y \pi(X_n\,, y)\, U(y)]\, dP_y \\
&\quad = e^{-\lambda} P_y\,\{E_{n+1}\} + e^{-\lambda} \int_{\tau_{x_0} > n} [\sum_{y \neq x_0} \pi(X_n\,, y)\, U(y)]\, dP_y \\
&\quad = e^{-\lambda} P_y\,\{E_{n+1}\} + e^{-\lambda} \int_{\tau_{x_0} > n+1} U(X_{n+1})\, dP_y
\end{aligned}$$

completing the induction argument. In our case, if we take $x_0 = 0$ and try $U_\sigma(x) = e^{-\sigma\, x}$ with $\sigma > 0$, for $x \geq 1$

$$\begin{aligned}
\sum_y \pi(x, y) U_\sigma(y) &= \sum_{y \geq x-1} e^{-\sigma\, y} p_{y-x+1} \\
&= \sum_{y \geq 0} e^{-\sigma\,(x+y-1)} p_y \\
&= e^{-\sigma\, x} e^{\sigma} \sum_{y \geq 0} e^{-\sigma\, y} p_y = \psi(\sigma) U_\sigma(x)
\end{aligned}$$

where

$$\psi(\sigma) = e^{\sigma} \sum_{y \geq 0} e^{-\sigma\, y} p_y.$$

Let us solve $e^{\lambda} = \psi(\sigma)$ for σ which is the same as solving $\log\psi(\sigma) = \lambda$ for $\lambda > 0$ to get a solution $\sigma = \sigma(\lambda) > 0$. Then

$$U(\lambda, x) = e^{-\sigma(\lambda)\, x} = E^{P_x}\{e^{-\lambda\tau_0}\}.$$

We see now that recurrence is equivalent to $\sigma(\lambda) \to 0$ as $\lambda \downarrow 0$ and positive recurrence to $\sigma(\lambda)$ being differentiable at $\lambda = 0$. The function $\log\psi(\sigma)$ is convex and its slope at the origin is $1 - m$. If $m > 1$ it dips below 0 initially for $\sigma > 0$ and then comes back up to 0 for some positive σ_0 before turning positive for good. In that situation $\lim_{\lambda\downarrow 0}\sigma(\lambda) = \sigma_0 > 0$ and that is transience. If $m < 1$ then $\log\psi(\sigma)$ has a positive slope at the origin and $\sigma'(0) = \frac{1}{\psi'(0)} = \frac{1}{1-m} < \infty$. If $m = 1$, then $\log\psi$ has zero slope at the origin and $\sigma'(0) = \infty$. This concludes the discussion of this problem.

Example 4.3. (The Urn Problem.)

We now turn to a discussion of the urn problem.

$$\pi(p, q\,; p+1, q) = \frac{p}{p+q} \quad \text{and} \quad \pi(p, q\,; p, q+1) = \frac{q}{p+q}$$

and π is zero otherwise. In this case the equation

$$F(p, q) = \frac{p}{p+q} F(p+1, q) + \frac{q}{p+q} F(p, q+1) \quad \text{for all } p, q$$

which will play a role later, has lots of solutions. In particular, $F(p, q) = \frac{p}{p+q}$ is one and for any $0 < x < 1$

$$F_x(p, q) = \frac{1}{\beta(p, q)} x^{p-1}(1-x)^{q-1}$$

where

$$\beta(p, q) = \frac{\Gamma(p)\Gamma(q)}{\Gamma(p+q)}$$

is a solution as well. The former is defined on $p + q > 0$ where as the latter is defined only on $p > 0, q > 0$. Actually if p or q is initially 0 it remains so for ever and there is nothing to study in that case. If f is a continuous function on $[0,1]$ then

$$F_f(p,q) = \int_0^1 F_x(p,q) f(x)\,dx$$

is a solution and if we want we can extend F_f by making it $f(1)$ on $q=0$ and $f(0)$ on $p=0$. It is a simple exercise to verify

$$\lim_{\substack{p,q\to\infty \\ \frac{p}{q}\to x}} F_f(p,q) = f(x)$$

for any continuous f on $[0,1]$. We will show that the ratio $\xi_n = \frac{p_n}{p_n+q_n}$ which is random, stabilizes asymptotically (i.e. has a limit) to a random variable ξ and if we start from p,q the distribution of ξ is the Beta distribution on $[0,1]$ with density

$$F_x(p,q) = \frac{1}{\beta(p,q)} x^{p-1}(1-x)^{q-1}$$

Suppose we have a Markov Chain on some state space $\mathcal{X}$ with transition probability $\pi(x,y)$ and $U(x)$ is a bounded function on $\mathcal{X}$ that solves

$$U(x) = \sum_y \pi(x,y)\,U(y).$$

Such functions are called (bounded) Harmonic functions for the Chain. Consider the random variables $\xi_n = U(X_n)$ for such an harmonic function. ξ_n are uniformly bounded by the bound for U. If we denote by $\eta_n = \xi_n - \xi_{n-1}$ an elementary calculation reveals

$$\begin{aligned} E^{P_x}\{\eta_{n+1}\} &= E^{P_x}\{U(X_{n+1}) - U(X_n)\} \\ &= E^{P_x}\{E^{P_x}\{U(X_{n+1}) - U(X_n)\}|\mathcal{F}_n\}\} \end{aligned}$$

where $\mathcal{F}_n$ is the σ-field generated by $X_0,\cdots,X_n$. But

$$E^{P_x}\{U(X_{n+1}) - U(X_n)\}|\mathcal{F}_n\} = \sum_y \pi(X_n,y)[U(y) - U(X_n)] = 0.$$

A similar calculation shows that

$$E^{P_x}\{\eta_n\eta_m\} = 0$$

for $m \neq n$. If we write

$$U(X_n) = U(X_0) + \eta_1 + \eta_2 + \cdots + \eta_n$$

this is an orthogonal sum in $L_2[P_x]$ and because U is bounded

$$E^{P_x}\{|U(X_n)|^2\} = |U(x)|^2 + \sum_{i=1}^{n} E^{P_x}\{|\eta_i|^2\} \leq C$$

is bounded in n. Therefore $\lim_{n\to\infty} U(X_n) = \xi$ exists in $L_2[P_x]$ and $E^{P_x}\{\xi\} = U(x)$. Actually the limit exists almost surely and we will show it when we discuss martingales later. In our example if we take $U(p,q) = \frac{p}{p+q}$, as remarked earlier, this is a harmonic function bounded by 1 and therefore

$$\lim_{n\to\infty} \frac{p_n}{p_n + q_n} = \xi$$

exists in $L_2[P_x]$. Moreover if we take $U(p,q) = F_f(p,q)$ for some continuous f on $[0,1]$, because $F_f(p,q) \to f(x)$ as $p, q \to \infty$ and $\frac{p}{q} \to x$, $U(p_n, q_n)$ has a limit as $n \to \infty$ and this limit has to be $f(\xi)$. On the other hand

$$\begin{aligned} E^{P_{p,q}}\{U(p_n, q_n)\} &= U(p_0, q_0) = F_f(p_0, q_0) \\ &= \frac{1}{\beta(p_0, q_0)} \int_0^1 f(x) x^{p_0 - 1}(1-x)^{q_0 - 1}\, dx \end{aligned}$$

giving us

$$E^{P_{p,q}}\{f(\xi)\} = \frac{1}{\beta(p,q)} \int_0^1 f(x)\, x^{p-1}(1-x)^{q-1}\, dx$$

thereby identifying the distribution of ξ under $P_{p,q}$ as the Beta distribution with the right parameters.

Example 4.4. (Branching Process). Consider a population, in which each individual member replaces itself at the beginning of each day by a random number of offsprings. Every individual has the same probability distribution, but the number of offsprings for different individuals are distibuted independently of each other. The distribution of the number N of offsprings is given by $P[N = i] = p_i$ for $i \geq 0$. If there are X_n individuals in the population on a given day, then the number of individuals X_{n+1} present on the next day, has the represenation

$$X_{n+1} = N_1 + N_2 + \cdots + N_{X_n}$$

as the sum of X_n independent random variables each having the offspring distribution $\{p_i : i \geq 0\}$. X_n is seen to be a Markov chain on the set of nonnegative integers. Note that if X_n ever becomes zero, i.e. if every member on a given day produces no offsprings, then the population remains extinct.

If one uses generating functions, then the transition probability $\pi_{i,j}$ of the chain are

$$\sum_j \pi_{i,j} z^j = \big[\sum_j p_j z^j\big]^i.$$

What is the long time behavior of the chain?

Let us denote by m the expected number of offsprings of any individual, i.e.

$$m = \sum_{i \geq 0} i p_i.$$

Then

$$E[X_{n+1}|\mathcal{F}_n] = mX_n.$$

1. If $m < 1$, then the population becomes extinct sooner or later. This is easy to see. Consider

$$E[\sum_{n\geq 0} X_n|\mathcal{F}_0] = \sum_{n\geq 0} m^n X_0 = \frac{1}{1-m} X_0 < \infty.$$

   By an application of Fubini's theorem, if $S = \sum_{n\geq 0} X_n$, then

$$E[S|X_0 = i] = \frac{i}{1-m} < \infty$$

   proving that $P[S < \infty] = 1$. In particular

$$P[\lim_{n\to\infty} X_n = 0] = 1.$$

2. If $m = 1$ and $p_1 = 1$, then $X_n \equiv X_0$ and the poulation size never changes, each individual replacing itself everytime by exactly one offspring.

3. If $m = 1$ and $p_1 < 1$, then $p_0 > 0$, and there is a positive probabiity $q(i) = q^i$ that the poulation becomes extinct, when it starts with i individuals. Here q is the probabilty of the population becoming extinct when we start with $X_0 = 1$. If we have initially i individulas each of the i family lines have to become extinct for the entire population to become extinct. The number q must therefore be a solution of the equation
$$q = P(q)$$
where $P(z)$ is the generating function
$$P(z) = \sum_{i \geq 0} p_i z^i.$$
If we show that the equation $P(z) = z$ has only the solution $z = 1$ in $0 \leq z \leq 1$, then the population becomes extinct with probability 1 although $E[S] = \infty$ in this case. If $P(1) = 1$ and $P(a) = a$ for some $0 \leq a < 1$ then by the mean value theorem applied to $P(z) - z$ we must have $P'(z) = 1$ for some $0 < z < 1$. But if $0 < z < 1$
$$P'(z) = \sum_{i \geq 1} i z^{i-1} p_i < \sum_{i \geq 1} i p_i = 1$$
a contradiction.

4. If $m > 1$ but $p_0 = 0$ the problem is trivial. There is no chance of the population becoming extinct. Let us assume that $p_0 > 0$. The equation $P(z) = z$ has another solution $z = q$ besides $z = 1$, in the range $0 < z < 1$. This is seen by considering the function $g(z) = P(z) - z$. We have $g(0) > 0, g(1) = 0, g'(1) > 0$ which implies another root. But $g(z)$ is convex and therefore ther can be atmost one more root. If we can rule out the possibility of extinction probability being equal to 1, then this root q must be the extinction probability when we start with a single individual at time 0. Let us denote by q_n the probability of extinction with in n days. Then
$$q_{n+1} = \sum_i p_i q_n^i = P(q_n)$$
and $q_1 < 1$. A simple consequence of the monotonicity of $P(z)$ and the inequalities $P(z) > z$ for $z < q$ and $P(z) < z$ for $z > q$ is that if

we start with any $a < 1$ and iterate $q_{n+1} = P(q_n)$ with $q_1 = a$, then $q_n \to q$.

If the population does not become extinct, one can show that it has to grow indefinitely. This is best done using martingales and we will revisit this example later as Example 5.6.

Example 4.5. Let $\mathcal{X}$ be the set of integers. Assume that transitions from x are possible only to $x-1, x$, and $x+1$. The transition matrix $\pi(x, y)$ appears as a tridiagonal matrix with $\pi(x, y) = 0$ unless $|x - y| \leq 1$. For simplicity let us assume that $\pi(x, x), \pi(x, x-1)$ and $\pi(x, x+1)$ are positive for all x.

The chain is then irreducible and aperiodic. Let us try to solve for

$$U(x) = P_x\{\tau_0 = \infty\}$$

that satisfies the equation

$$U(x) = \pi(x, x-1)\, U(x-1) + \pi(x, x)\, U(x) + \pi(x, x+1)\, U(x+1)$$

for $x \neq 0$ with $U(0) = 0$. The equations decouple into a set for $x > 0$ and a set for $x < 0$. If we denote by $V(x) = U(x+1) - U(x)$ for $x \geq 0$, then we always have

$$U(x) = \pi(x, x-1)\, U(x) + \pi(x, x)\, U(x) + \pi(x, x+1)\, U(x)$$

so that

$$\pi(x, x-1)\, V(x-1) - \pi(x, x+1)\, V(x) = 0$$

or

$$\frac{V(x)}{V(x-1)} = \frac{\pi(x, x-1)}{\pi(x, x+1)}$$

and therefore

$$V(x) = V(0) \prod_{i=1}^{x} \frac{\pi(i, i-1)}{\pi(i, i+1)}$$

and

$$U(x) = V(0) \left[1 + \sum_{y=1}^{x-1} \prod_{i=1}^{y} \frac{\pi(i, i-1)}{\pi(i, i+1)} \right].$$

If the chain is to be transient we must have for some choice of $V(0)$, $0 \leq U(x) \leq 1$ for all $x > 0$ and this will be possible only if

$$\sum_{y=1}^{\infty} \prod_{i=1}^{y} \frac{\pi(i, i-1)}{\pi(i, i+1)} < \infty$$

which then becomes a necessary condition for

$$P_x\{\tau_0 = \infty\} > 0$$

for $x > 0$. There is a similar condition on the negative side

$$\sum_{y=1}^{\infty} \prod_{i=1}^{y} \frac{\pi(-i, -i+1)}{\pi(-i, -i-1)} < \infty.$$

Transience needs at least one of the two series to converge. Actually the converse is also true. If, for instance the series on the positive side converges then we get a function $U(x)$ with $0 \leq U(x) \leq 1$ and $U(0) = 0$ that satisfies

$$U(x) = \pi(x, x-1)\, U(x-1) + \pi(x, x)\, U(x) + \pi(x, x+1)\, U(x+1)$$

and by iteration one can prove that for each n,

$$U(x) = \int_{\tau_0 > n} U(X_n)\, dP_x \leq P\{\tau_0 > n\}$$

so the existence of a nontrivial U implies transience.

Exercise 4.16. Determine the conditions for positive recurrence in the previous example.

Exercise 4.17. We replace the set of integers by the set of nonnegative integers and assume that $\pi(0, y) = 0$ for $y \geq 2$. Such processes are called birth and death processes. Work out the conditions in that case.

Exercise 4.18. In the special case of a birth and death process with $\pi(0, 1) = \pi(0, 0) = \frac{1}{2}$, and for $x \geq 1$, $\pi(x, x) = \frac{1}{3}$, $\pi(x, x-1) = \frac{1}{3} + a_x$, $\pi(x, x+1) = \frac{1}{3} - a_x$ with $a_x = \frac{\lambda}{x^\alpha}$ for large x, find conditions on positive α and real λ for the chain to be transient, null recurrent and positive recurrent.

Exercise 4.19. The notion of a Markov Chain makes sense for a finite chain $X_0, \cdots, X_n$. Formulate it precisely. Show that if the chain $\{X_j : 0 \leq j \leq n\}$ is Markov so is the reversed chain $\{Y_j : 0 \leq j \leq n\}$ where $Y_j = X_{n-j}$ for $0 \leq j \leq n$. Can the transition probabilities of the reversed chain be determined by the transition probabilities of the forward chain? If the forward chain has stationary transition proabilities does the same hold true for the reversed chain? What if we assume that the chain has a finte invariant probability distribution and we initialize the chain to start with an initial distribution which is the invariant distribution?

Exercise 4.20. Consider the simple chain on nonnegative integers with the following transition probailities. $\pi(0, x) = p_x$ for $x \geq 0$ with $\sum_{x=0}^{\infty} p_x = 1$. For $x > 0$, $\pi(x, x-1) = 1$ and $\pi(x, y) = 0$ for all other y. Determine conditions on $\{p_x\}$ in order that the chain may be transient, null recurrent or positive recurrent. Determine the invariant probability measure in the positive recurrent case.

Exercise 4.21. Show that any null recurrent equivalence class must necessarily contain an infinite number of states. In patricular any Markov Chain with a finite state space has only transient and positive recurrent states and moreover the set of positive recurrent states must be non empty.