Probability For Statistics - Concise Notes

MATH50010

Arnav Singh

Colour Code - **Definitions** are **green** in these notes, **Consequences** are **red** and **Causes** are **blue**

Content from MATH40005 assumed to be known.

Mathematics
Imperial College London
United Kingdom
March 1, 2022

Contents

1 Probability Review

Definition 1.1 - Experiment
Any fixed procedure with variable outcome

Definition 1.2 - Sample space Ω
Set of all possible outcomes of an experiment

Definition 1.4 - σ-algebra (Sigma-algebra)
$\mathcal{F}$ a collection of subsets of Ω
$\mathcal{F}$ an algebra if

(i) $\emptyset \in \mathcal{F}$

(ii) $A \in \mathcal{F} \implies A^c \in \mathcal{F}$

(iii) $A, B \in \mathcal{F} \implies A \cup B \in \mathcal{F}$ (closed under finite union)

$\mathcal{F}$ a σ-algebra if closed under countable union.

Definition 1.13 - Borel sigma algebra
Let $\mathcal{F}_i, i \in \mathcal{I}$, the collection of all σ-algebras that contain all open intervals of $\mathbb{R}$
$\{\mathcal{F}_i\}$ clearly non-empty, since power set of $\mathbb{R}$ is such a sigma algebra. **Borel sigma algebra,** $\mathcal{B} := \cap_{i \in \mathcal{I}} \mathcal{F}_i$

Remarks

(i) $\mathcal{B}$ contains all open intervals, their complements, countable unions and countable intersections.

(ii) $\mathcal{F}$ a sigma algebra containing all intervals of the form above $\implies \mathcal{B} \in \mathcal{F}.\mathcal{B}$ thought of as the smallest sigma algebra containing all intervals

(iii) $B \subset \mathcal{B}$ said to be a **Borel set**

Definition 1.16 - Kolmogorov Axioms
Given Ω and a σ-algebra $\mathcal{F}$ on Ω
A **Probability function/ Probability Measure** is a function $Pr : \mathcal{F} \to [0, 1]$ satisfying:

(i) $Pr(A) \geq 0, \forall A \in F$

(ii) $Pr(\Omega) = 1$

(iii) $\{A_i\} \in \mathcal{F}$ **are pairwise disjoint then**

$$Pr(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} Pr(A_i)$$

Definition 1.17 - Probability Space
Defined as the triple $(\Omega, \mathcal{F}, Pr(\cdot))$

Properties of $Pr(\cdot)$

- $Pr(\emptyset) = 0$
- $Pr(A) \leq 1$, $Pr(A^c) = 1 - Pr(A)$
- $Pr(A \cup B) = Pr(A) + Pr(B) - Pr(A \cap B)$
- $A \subset B \implies Pr(A) \leq Pr(B)$
- $Pr(A) = \sum_{i=1}^{n=\infty} Pr(A \cap C_i$ for $\{C_i\}$ a partition of Ω

Proposition 1.18 - Continuity Property
Let $(\Omega, \mathcal{F}, Pr(\cdot))$, and $A_1, A_2, \dots \in \mathcal{F}$ an increasing sequence of events, $(A_1 \subseteq A_2 \subseteq \dots)$

$$A = \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$$

$\mathcal{F}$ a sigma algebra $\implies$

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

Definition 1.20 - Conditional Probability
$A, B \in \mathcal{F}, Pr(B) > 0$, Conditional Probability of A given B is

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

Definition 1.21 - Independence
2 events are independent if

$$Pr(A \cap B) = Pr(A)Pr(B)$$

Definition 1.22 - Mutually independent
$\{A_i\} \in \mathcal{F}$ mutually independent if for any subcollection $\{A_{i_j}\}_{j=1,\dots,k}$

$$Pr(\bigcap_{j=1}^{k} A_{i_j}) = \prod_{j=1}^{k} Pr(A_{i_j}$$

2 Random Variables

2.0 Definitions

Definition 2.1 - Random Variable
A random variable on $(\Omega, \mathcal{F}, Pr)$ a function

$$X : \Omega \to \mathbb{R}$$

such that, $\forall$ Borel set $B \in \mathcal{B}, X^{-1}(B) \in \mathcal{F}$
Random vectors defined analogously, $X : \Omega \to \mathbb{R}^n$ and Complex Random Variables $X : \Omega \to \mathbb{C}$
Definition 2.3 - Distribution
$\forall$ Borel sets $B \in \mathcal{B}$

$$Pr_X(B) = Pr(X^{-1}(B)) = Pr(\{\omega \in \Omega : X(\omega) \in B\})$$

Pr_X the **distribution** of X. Written $Pr(X \in B)$

We say X and Y **identically distributed** if $Pr(X \in B) = Pr(Y \in B) \forall B \in \mathcal{B}$
Definition 2.10 - Cumulative Distributive Function (CDF)
CDF of a random variable X a function $F_x : \mathbb{R} \to [0, 1]$

$$F_x = Pr(X \leq x)$$

Notation - Monotone limits

- Write $x_n \downarrow x$ for (x_n) a seq. (weakly) monotonically decreasing to limit x
- Write $x_n \uparrow x$ for (x_n) a seq. (weakly) monotonically increasing to limit x

Properties of the CDF

- $F_X(x)$ is non-decreasing
- $\lim_{x\to-\infty} F_X(x) = 0; \lim_{x\to+\infty} F_X(x) = 1$
- $\lim_{x\downarrow x_0} F_X(x) = F_X(x_0)$, F is continuous from the right.

Definition - Point mass CDF

The constant random variable the most trivial RV. For $a \in \mathbb{R}$ define point mass CDF as

$$\delta_a(x) = \begin{cases} 0 & x < a \\ 1 & x \geq a \end{cases}$$

Definition 2.14 - Probability Mass Function PMF

1. **if** $\exists (a_n)_{n\geq 1}$ **and** $(b_n)_{n\geq 1}$ **where** $b_i > 0$ **with** $\sum_i b_i = 1$ **with** $F_X(x)$ **s.t**

$$F_X(x) = \sum_{i=1}^{\infty} b_i \delta_{\alpha_i}(x)$$

Then X **a discrete random variable, with PMF** $f_X(x) = Pr(X = x)$

2. **if** $F_X(x)$ **continuous** $\implies$ X **a continuous random variable**

3. **if** X **a continuous random variable s.t** $\exists f_X : \mathbb{R} \to \mathbb{R}$

$$F_x(x) = \int_{-\infty}^{x} f_X(t)dt, \forall x \in \mathbb{R}$$

Then X **an absolutely continuous random variable with probability density function (PDF)** $f_X(x)$

2.1 Transformations of Random Variables

Suppose X an absolutely continuous random variable with pdf f_X and $g : \mathbb{R} \to \mathbb{R}$ a strictly monotonic and differentiable

$$Y = g(X) \implies f_Y(y) = f_X(g^{-1}(y)) \left| \frac{dg^{-1}y}{dy} \right|$$

$$f_Y(y) = f_X(x)\frac{dx}{dy}$$

Families of distributions
Scale Family
For $\sigma > 0$, we have $Y = \sigma Z$ which has pdf

$$f(y|\sigma) = \frac{1}{\sigma} f_Z(\frac{y}{\sigma})$$

Location-Scale Family
Define $W = \mu + \sigma Z$, with pdf

$$f(w|\mu, \sigma) = \frac{1}{\sigma} f_Z(\frac{w - \mu}{\sigma})$$

Probability Integral Transform
Let $U \sim Unif[0, 1]$ with $X = F^{-1}(U)$ s.t F a strictly increasing CDF $\implies$ X a random variable with CDF F

Expectation
For discrete r.v X

$$E(X) = \sum_x xPr(X = x)$$

Similary for continuous r.v

$$E(X) = \int_{-\infty}^{+\infty} xf_X(x)dx$$

Properties of Expectation

1. $E(aX + bY) = aE(x) + bE(Y), \forall a, b \in \mathbb{R}$
2. If $Pr(X \geq 0) = 1 \implies E(X) \geq 0$
3. If A an event, $E(1_A) = Pr(A)$

3 Multivariate Random Variables

3.0 Definitions

Definition 3.1 - Joint Cumulative Distribution Function (Joint CDF)
Given by

$$F_{XY}(x,y) = Pr(X \leq x, Y \leq y)$$

Jointly absolutely continuous case:

$$F_{XY}(x,y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{XY}(s,t)dsdt$$

Definition - Marginal Density Function (MDF)

$$f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)dy$$

Definition 3.3 - Independence
Finite set of r.v $\{X_i\}$ said to be **independent** if

$$Pr(X_1 \in B_1, \ldots, X_n \in B_n) = \prod_{i=1}^{n} Pr(X_i \in B)$$

$\forall$ Borel sets B_i
Corollary.
Any collection (X_i) independent if every finite subcollection independent.

Definition 3.4 - Covariance
For r.v X, Y with finite $E(X) = \mu_X, E(Y) = \mu_Y$

$$Cov(X,Y) = E((X - \mu_X)(Y - \mu_Y))$$

Definition - Correlation

$$Cor(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)Var(Y)}}$$

Change of variables for pdfs
If $(U,V) = T(X,Y)$ a function of pair of r.v (X,Y) with joint pdf f_{XY} a joint pdf for (U,V) given by

$$f_{UV}(u,v) = f_{XY}(x(u,v), y(u,v))|J(u,v)|$$

Where

$$J(u,v) = det \begin{pmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{pmatrix}$$

Remark 3.9 - (Factorisable independence)
X, Y independent if $\exists g, h : \mathbb{R} to \mathbb{R}$ such that the joint mass/density function factorises as

$$f_{XY}(x,u) = g(x)h(y), \quad \forall x, y \in R$$

Definition - Conditioning
For X a r.v, conditional CDF of X given A

$$F_{X|A}(x) = \frac{Pr(\{X \leq x\} \cap A)}{Pr(A)}$$

$$f_{X|A}(x) = \frac{d}{dx} F_{X|A}(x)$$

Definition - Conditional Probability Density Function

$$f_{Y|X}(y|x) = \frac{d}{dy} F_{Y|X}(y|x) = \frac{f_{XY}(x,y)}{f_X(x)}$$

3.1 Bivariate Normal Distribution

Definition - Standard bivariate normal distribution
Has pdf for $-1 < \rho < 1$

$$f(x,y|\rho) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{1}{2(1-\rho^2)(x^2 - 2\rho xy + y^2)}\right) \qquad (x,y) \in \mathbb{R}^2$$

Properties:
$E(X) = E(Y) = 0, E(XY) = \rho$
$Var(X) = Var(Y) = 1, Cov(X,Y) = \rho$

Vector form
$\mathbf{x} = (x,y)\ \underline{\mu} = (\mu_x, \mu_y), \Sigma = \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_y \\ \rho\sigma_x\sigma_y & \sigma_2^2 \end{pmatrix}$

$$f_{\mathbf{X}}(\mathbf{x}|\underline{\mu}, \Sigma) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} exp\left(-\frac{1}{2}(\mathbf{x} - \underline{\mu})^T(\Sigma^{-1}(\mathbf{x} - \underline{\mu})\right)$$

Extend this to **Multivariate normal distribution:**

$$f_{\mathbf{X}}(\mathbf{x}|\underline{\mu}, \Sigma) = \frac{1}{(2\pi)^{d/2}(det\Sigma)^{1/2})} exp\left(-\frac{1}{2}(\mathbf{x} - \underline{\mu})^T(\Sigma^{-1}(\mathbf{x} - \underline{\mu})\right)$$

Where $\mathbf{x} \in \mathbb{R}^d$, $(\Sigma_{ij}) = Cov(X_i, X_j)$
Remarks.

- Σ symmetric: $Cov(X,Y) = Cov(Y,X)$
- $diag(\Sigma) = \{Var(X_i)\}$
- constant $\mathbf{a} \in \mathbb{R}^d$ $Var(\mathbf{a^T x}) = \mathbf{a}^T \Sigma a$

Proposition 3.16.

$X \sim MVN_d(\underline{\mu}, \Sigma)$**,** A **invertible** $d \times d$ **matrix**

$$\implies Y = AX \sim MVN_d(A\mu, A\Sigma A^T)$$

Proposition 3.17.

Can always find linear transform Q of $\mathbf{X}$
s.t entries of $Z = Q\mathbf{X}$ uncorrelated and independent random variable.

3.2 Order statistic

Consider random sample $(X_1, \ldots, X_n)$ with cdf F_X and pdf f_X
with Y_1 smallest, Y_2 next smallest etc.
$(Y_1, \ldots, Y_n)$ the **vector of order statistics of X**

$$f(n) = \begin{cases} n! \prod_{i=1}^n f_X(y_i) & y_1 < y_2 < \cdots < y_n \\ 0, \text{otherwise} \end{cases}$$

$$f_k(y) = k\binom{n}{k} f_X(y) F_X(y)^{k-1}(1 - F_X(y))^{n-k}$$

$$F_k(y) = Pr(N_y \geq k) = \sum_{j=k}^{n} \binom{n}{j} F_X(y)^j (1 - F_X(y))^{n-j}$$

4 Convergence of Random Variables

4.1 Convergence

Definition 4.1. Sequence (X_i) of random variables said to **converge in probability** to X

$$X_n \xrightarrow{P} X \quad \text{if} \quad \forall \epsilon > 0 \lim_{n\to\infty} Pr(|X_n - X| \geq \epsilon) = 0$$

Proposition 4.4. - *(Markov's inequality)*
X a random variable taking non-negative values only.
$a > 0$ constant

$$Pr(X \geq a) \leq \frac{E(X)}{a}$$

Proposition 4.5.
Take non-negative random variable $Y = (X - \mu)^2$

$$Pr(|X - \mu| \geq \epsilon) = Pr((X - \mu)^2 \geq \epsilon^2) = P(Y \geq \epsilon^2)$$

$$Pr(Y \geq \epsilon^2) \leq \frac{E(X-\mu)^2}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2}$$

Definition 4.6.
$X_1, X_2, \ldots$

$$\underbrace{\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i}_{\textbf{sample mean}}$$

Proposition 4.7. - *(Weak law of large numbers)*
$X_1, X_2, \ldots$ sequence of iid random variable with finite μ, σ^2

$$\implies \bar{X}_n \xrightarrow{P} \mu$$

Definition 4.9.
$X_1, X_2, \ldots$ with cdfs $F_1, F_2, \ldots$
Converge in distribution to random variable X with cdf X_n

$$X_n \xrightarrow{D} X \quad \text{if} \quad \lim_{n\to\infty} F_n(X) = F_X(x)$$

$\forall x \in \mathbb{R}$ for which F_X continuous.

Proposition 4.12.
Converge in probability $\implies$ **converge in distribution**

Proposition 4.14.
Suppose $(X_n)_{n\geq 1}$ sequence of random variables

$$X_n \xrightarrow{D} c \in \mathbb{R} \implies X_n \xrightarrow{P} c$$

4.2 Limit events

Definition

$A_1, A_2, \ldots$ sequence of events

- $\{A_n\ i.o\} = A_n$ infinitely often
- $\{A_n\ a.a\} = A_n$ almost always (finitely many A_n dont occur)

$\{A_n\ a.a\} \subset \{A_n\ i.o\}$

Proposition 4.15.
Sequence of sets (A_n) We define

$$\underbrace{B_n = \bigcap_{m=n}^{\infty} A_m}_{\text{increasing sequence}} \qquad \underbrace{C_n = \bigcup_{m=n}^{\infty} A_m}_{\text{decreasing sequence}}$$

And further

$$\underbrace{\liminf_{n\to\infty} A_N = \bigcup_{n=1}^{\infty}\bigcap_{n=N}^{\infty} A_n}_{=\{A_n a.a\}} \quad \underbrace{\limsup_{n\to\infty} A_N = \bigcap_{n=1}^{\infty}\bigcup_{n=N}^{\infty} A_n}_{=\{A_n i.o\}}$$

Remark 4.16

$\{A_n\ i.o\}^C$ - only finitely many A_n occur

$$\{A_n\ i.o\}^C = \bigcup_{N=1}^{\infty}\bigcap_{n=N}^{\infty} A_n^C = \{A_n^C\ a.a\}$$

Proposition 4.17

$A_1, A_2, \dots$ sequence of events

(i) if $\sum_{n=1}^{\infty} Pr(A_n) < \infty \implies Pr(\{A_n\ i.o\}) = 0$

(ii) if $\sum_{n=1}^{\infty} Pr(A_n) = \infty$ **and** $\{A_i\}$ **independent** $\implies Pr(\{A_n i.o\}) = 1$

5 Central Limit Theorem

5.1 Moment generating functions

Definition 5.1 - ***(Moment generating functions MGFs)***

$$M_X(t) = E\left[\exp\left(tX\right)\right]$$

Proposition 5.2.
$Y = aX + b \implies M_Y(t) = \exp(bt)M_X(at)$

Proposition 5.3.
X, Y independent random variables

$$Z = X + Y \implies M_Z(t) = M_X(t)M_Y(t)$$

Proposition 5.4.
Suppose $\exists t_0 > 0$ s.t $M_X(t) < \infty$ for $|t| < t_0$

$$M_X(t) = \sum_{k=0}^{\infty} E(X^k)\frac{t^k}{k!} \quad \implies \quad \forall k > 0\ \frac{d^k}{dt^k}M_X(t)|_{t=0}E(X^k)$$

Proposition 5.5.
(Uniqueness)
Suppose X, Y random variables with common MGF finite for $|t| < t_0$ for some $t_0 > 0$

$$X, Y \text{ identically distributed}$$

(Continuity)
Suppose X a random variable with $M_X(t)$
$(X_n)_{n\geq 1}$ sequence of random variables with respective $M_{X_i}(t)$

$$\text{if } M_{X_i}(t) \xrightarrow[i\to\infty]{} M_X(t) < \infty \quad (\forall |t| < t_0, t_0 > 0$$

$$\implies X_n \xrightarrow{D} X$$

Definition 5.11

Say $f(x) = o(g(x))$ **in** $\lim_{x\to\infty}$ if

$$\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$$

Similarly defined for case $x \to 0$ limit.

Proposition 5.12.
$X_1, X_2, \ldots$ sequence of iid random variables with common MGF; $M(t)$ exists in open interval containing 0

$$\text{if } E(X_i) = \mu \;\forall i \implies \bar{X}_n \xrightarrow{P} \mu$$

5.2 The Central Limit Theorem

Proposition 5.14.
$X_1, X_2, \ldots$ sequence of iid random variables with **common MGF** $M(t)$ (existing in open interval containing 0)
$E(X_i) = \mu,\; Var(X_i) = \sigma^2 \forall i$

$$\implies \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu) \xrightarrow{D} Z \sim N(0,1)$$

6 Stochastic Processes

Definition
$\mathcal{E}-$ the state space (finite or countably infinite)

Random process - sequence of $\mathcal{E}$ valued random variables $X_0, X_1, \ldots$

6.1 Time Homogeneous Markov Chains

Definition 6.1
Stochastic Process on state space $\mathcal{E}$, collection of $\mathcal{E}$-valued r.v $(X_t)_{t\in T}$ indexed by set T; often $T = \mathbb{N}_0$

Definition 6.2
Discrete time stochastic process $(X_n)_{n\in\mathbb{N}_0}$ on $\mathcal{E}$ a **Markov chain** if

$$P(X_n = x_n | X_{n-1} = x_{n-1}, \ldots, X_0 = x_0) = P(X_n = x_n | X_{n-1} = x_{n-1}) \quad \forall n \in \mathbb{N}, \forall x_n, \ldots, x_0 \in \mathcal{E}$$

Definition 6.3
Markov chains **Time homegenous** if

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i) \quad \forall n \in \mathbb{N}_0, \forall i, j \in \mathcal{E}$$

Definition 6.4
Matrix $P = (p_{ij})_{i,j\in\mathcal{E}}$ of **transition probability** $p_{ij} = Pr(X_1 = j | X_0 = i)$
Called the **Transition Matrix** for the time homogeneous Markov chain. (X_n)

6.2 Initial distribution

Definition.
Initial distribution and Transition Matrix specify stochastic process fully i.e.
$\lambda = (\lambda_j)_{j\in\mathcal{E}}; \lambda_j = Pr(X_0 = j)$

- **Marginal distribution -** $P(X_1 = j) = \sum_{i\in\mathcal{E}} P(X_1 = j | X_0 = i) P(X_0 = i) = \sum_{i\in\mathcal{E}} p_{ij}\lambda_i$
- **Joint distribution** - $P(X_2 = k, X_1 = j) = P(X_2 = k | X_1 = j) P(X_1 = j) = p_{jk} \sum_{i\in\mathcal{E}} p_{ij}\lambda_i$

Definition 6.7
Given Markov chain $(X_n)_{n\in\mathbb{N}_0}$.

$$\textbf{N-step transition prob matrix } P(n) = p_{ij}(n) = P(X_n = j | X_0 = 1)$$

Proposition 6.9. - *(Chapman-Kolmogorov equations)*
Suppose $m \geq 0$ and $n \geq 1$

$$p_{ij}(m+n) = \sum_{l\in\mathcal{E}} p_{il}(m) p_{lj}(n)$$

$$P(m+n) = P(m)P(n) \quad \text{(Matrix form)} \implies P(m) = P^m$$

6.3 Class Structure

6.3.1 Definitions

Definition 6.11
State j **accessible** from state i; $i \to j$ if $\exists n \geq 0$ **s.t** $p_{ij}(n) > 0$

Definition 6.13
States i, j **communicate**; $i \longleftrightarrow j$ if $i \to j$ **and** $j \to i$

Proposition 6.15.
Binary relation $i \Longleftrightarrow j$ an equivalence relation on $\mathcal{E}$, partitioning $\mathcal{E}$ into communicating classes.

Definition 6.17
Set of states C **closed** if $p_{ij} = 0, \forall i \in C, j \notin C$

Definition 6.19
Set of states C **irreducible** if $i \longleftrightarrow j, \forall i, j \in C$

6.3.2 Periodicity

Definition 6.20
Period of state i; $d(i) = \{n > 0 : p_{ii}(n) > 0\}$

- $d(i) = 1$ say state is **aperiodic**
- $d(i) > 1$ say state is **periodic**

Proposition 6.22.
All states in same communicating class have same periodicity

6.4 Classifcation of states

Definition 6.24.

$i \in \mathcal{E}$ for Markov Chain X_n

- **Recurrent** if
$$P(X_n = i, n \geq 1 | X_0 = i) = P\left(\bigcup_{n=1}^{\infty} \{X_n = i\} | X_0 = i\right) = 1$$
- **Transient** if
$$P\left(\bigcup_{n=1}^{\infty} \{X_n = i\} | X_0 = i\right) < 1$$

Definition 6.25.
First passage time of state $j \in \mathcal{E}$
$$T_j = \min\{n \geq 1 : X_n = j\}$$
First n s.t $X_n = j$
Say $T_j = \infty$ if never visits state $j \implies T_j$ not a random variable since its not real valued.

Definition.
$$\{T_j = n\} = \{X_n = j, X_i \neq j : i < n\}$$

Remark 6.27
$$\begin{aligned} f_{ij}(n) &= Pr(T_j = n | X_0 = i) \\ f_{ij} &= Pr(T_j < \infty | X_0 = i) \\ &= Pr(\bigcup_{n=1}^{\infty} \{T_j = n\} | X_0 = i) \\ &= \sum_{n=1}^{\infty} f_{ij}(n) \end{aligned}$$

Remark 6.28

$$\text{State } i\text{: } \begin{cases} \text{recurrent} & \iff f_i i = 1 \iff \sum_{n=1}^{\infty} p_{ii}(n) = \infty \\ \text{transient} & \iff f_i i < 1 \iff \sum_{n=1}^{\infty} p_{ii}(n) < \infty \end{cases}$$

Proposition 6.29
$i, j \in \mathcal{E},\ n \geq 1$

$$p_{ij}(n) = \sum_{l=1}^{n} f_{ij}(l) p_{ij}(n-l)$$

$p_{ij} = p_{ij}(1) = f_{ij}(1)$

Proposition 6.32
$i \longleftrightarrow j \implies$ **either i, j both recurrent or both transient**

Proposition 6.33
C **a recurrent communicating class**
$\implies$ C **closed:** $i \in C, j \notin C$ **we have** $p_{ij} = 0$

Proposition 6.34
State space decomposes

$$\mathcal{E} = \underbrace{T}_{\text{Transient states}} \cup \underbrace{C_1 \cup C_2 \cup \dots}_{\text{irreducible closed sets of recurrent states}}$$

Definition 6.36
Mean recurrence time of state $i \in \mathcal{E}$

$$\mu_i = E(T_i | X_0 = i)$$

Remark 6.37

- **Transient States**: $\mu_i = \infty$ since $P(T_i = \infty | X_0 = i) > 0$
- **Recurrent States**: $\mu_i = \sum_{n=1}^{\infty}$ can be finite or infinite

Definition 6.38
$i \in \mathcal{E}$

- **null recurrent** if $\mu_i = \infty$
- **positive recurrent** if $\mu_i < \infty$

Definition 6.39
(X_n) a markov chain on $\mathcal{E}$ **Hitting time** of set $A \subseteq \mathcal{E}$ a random variable

$$H^A = \min\{n \geq 0 : X_n \in A\}$$

We take $\min \emptyset = \infty$
Hitting probability starting at $i \in \mathcal{E}$

$$h_i^A = Pr(H^A < \infty | X_0 = i)$$

in the case $A = \{j\}$ we write h_i^j

Proposition 6.43
$A \subseteq \mathcal{E}$ take vector $h^A = (h_i^A)_{i \in \mathcal{E}}$ solves system

$$h_i^A = \begin{cases} 1 & i \in A \\ \sum_{j \in \mathcal{E}} p_{ij} h_j^A & i \notin A \end{cases}$$

6.5 Stationary Distributions

Definition 6.44
Vector $\pi = (\pi_j)_[j \in \mathcal{E}$ a **stationary distribution** for (X_n) if

(i) $\pi_j \geq 0\ \forall j \in \mathcal{E}$ and $\sum_{j \in \mathcal{E}} \pi_j = 1$ (π a probability distribution on $\mathcal{E}$

(ii) $\pi P = \pi$

Proposition 6.45
X_n has distribution π with π stationary for (X_n)
$\implies$ **X_{n+1} has distribution π**

Proposition 6.46
irreducible chain has stationary distribution
$\iff$ all states positive recurrent
$\implies \pi_j = \frac{1}{\mu_j}$ for μ_j the mean recurrence time
$\implies$ stationary distribution is unique

Proposition 6.47
$(X_n)_{n\in\mathbb{N}_0}$ irreducible aperiodic Markov Chain with stationary distribution π
$\implies \forall$ initial distribution λ, $\forall j \in \mathcal{E}$

$$\lim_{n\to\infty} Pr(X_n = j) = \pi_j$$

$$\forall i \in \mathcal{E} : \lim_{n\to\infty} Pr(X_n = j | X_0 = i) = \pi_j \quad \text{(independent of } i\text{)}$$

Proposition 6.48 - *(Ergodic Theorem)*
$(X_n)_{n\in\mathbb{N}_0}$ irreducible Markov Chain

$$\forall i \in \mathcal{E} \text{ let } V(i) = \sum_{r=0}^{n} I(X_r = i)$$

Counts the number of visits to i before time n
$\implies \forall$ initial distributions, $i \in \mathcal{E}$ we have $Pr(\frac{V(i)}{n} \xrightarrow[n\to\infty]{} \pi_i) = 1$

Proposition
Symmetrical random walk on finite graph
$i \in \mathcal{E}$ connected to d_i other states

$$\implies \pi_i = \frac{d_i}{\sum_{j\in\mathcal{E}} d_j}$$