Textbook: *Introduction to Probability Models* by Sheldon M. Ross, Academic Press, 2010 (10th Edition).

1.2. Sample Space and Events.

Definition. The sample space of an experiment is a collection of all possible outcomes. *Notation.* S.
Example. Flipping a coin. $S = \{H, T\}$.
Definition. An event is a subset of S. *Notation.* A, B, C.
Example. Roll a die twice. An event D is observing a double.
$D = \{(1,1),(2,2),(3,3),(4,4),(5,5),(6,6)\}$.

1.3. Probabilities Defined on Events.

Definition. The probability of an event A is a number $\mathbb{P}(A)$ satisfying

(1) $0 \leq \mathbb{P}(A) \leq 1$,
(2) $\mathbb{P}(S) = 1$,
(3) for any disjoint events $E_1, E_2, \dots$,
$\mathbb{P}\left(\cup_{n=1}^{\infty} E_n\right) = \sum_{n=1}^{\infty} \mathbb{P}(E_n)$.

Example. $\mathbb{P}(D) = 1/6$.
Useful Formulas.

$$\mathbb{P}(E^c) = 1 - \mathbb{P}(E),$$

$$\mathbb{P}(E \cup F) = \mathbb{P}(E) + \mathbb{P}(F) - \mathbb{P}(EF),$$

$$\mathbb{P}\left(\cup_{n=1}^{\infty} E_n\right) \leq \sum_{n=1}^{\infty} \mathbb{P}(E_n) \ (Boole's\ \ inequality)$$

1.4. Conditional Probabilities.

Definition. The conditional probability of an event E given that an event F occurred is

$$\mathbb{P}(E|F) = \frac{\mathbb{P}(EF)}{\mathbb{P}(F)}.$$

1.5. Independent Events.

Definition. Events E and F are independent if

$$\mathbb{P}(EF) = \mathbb{P}(E)\mathbb{P}(F)$$

or, equivalently,

$$\mathbb{P}(E|F) = \mathbb{P}(E).$$

2.1. Random Variables.

Definition. A random variable is a real-valued function defined on the sample space. *Notation.* X, Y, T, N.
Example. Flip a coin three times. Let X be the number of heads observed. Then, $X(TTT) = 0, X(HTT) = X(THT) = X(TTH) = 1, X(HHT) = X(HTH) = X(THH) = 2, X(HHH) = 3$.

2.2. Discrete Random Variables, 2.4. Expectation of a Random Variable, 2.6. Moment Generating Functions.

Definition. A discrete random variable assumes a finite or countably infinite number of values.
Definition. The probability function $p(x) = \mathbb{P}(X = x)$ satisfies

$$(1)\ \ p(x) \geq 0 \ \ \forall x,$$
$$(2)\ \ \sum_x p(x) = 1,$$
$$(3)\ \ \mathbb{P}(X \in A) = \sum_{x \in A} p(x).$$

Definition. The expectation (or expected value) of a discrete random variable is $\mathbb{E}X = \sum_x xp(x)$.
Definition. The variance of a r.v. is $\mathbb{V}arX = \mathbb{E}(X - \mathbb{E}X)^2$.
Definition. The moment generating function of a r.v. X is $\varphi(t) = \mathbb{E}[e^{tX}]$.
Useful Formula. $EX^k = \varphi^{(k)}(0)$.

Some Discrete Distributions.

(1) *Binomial(n, p)*: $p(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$, $\mathbb{E}X = np$, $\mathbb{V}arX = np(1-p)$, $\varphi(t) = (pe^t + 1 - p)^n$,
(2) *Geometric(p)*: $p(x) = p(1-p)^{x-1}$, $x = 1, 2, \ldots$, $\mathbb{E}X = \frac{1}{p}$, $\mathbb{V}arX = \frac{1-p}{p^2}$, $\varphi(t) = \frac{pe^t}{1-(1-p)e^t}$,
(3)*Poisson(λ)*: $p(x) = \frac{\lambda^x e^{-\lambda}}{x!}$, $x = 0, 1, \ldots, \mathbb{E}X = \lambda$, $\mathbb{V}arX = \lambda, \varphi(t) = e^{\lambda(e^t - 1)}$.

2.3. Continuous Random Variables, 2.4. Expectation of a Random Variable, 2.6. Moment Generating Functions.

Definition. A continuous random variable assumes values in an interval.
Definition. The probability density function (p.d.f.) $f(x)$ of a c.r.v. X satisfies

$$(1)\ \ f(x) \geq 0, \ \ \forall x,$$
$$(2)\ \ \int_{-\infty}^{\infty} f(x)\, dx = 1,$$

$$(3)\ \ \mathbb{P}(X \in A) = \int_A f(x)\,dx.$$

Definition. The cumulative distribution function (c.d.f.) $F(a) = \int_{-\infty}^{a} f(x)\,dx$.
Definition. The expectation of a c.r.v. is $\mathbb{E}X = \int_{-\infty}^{\infty} xf(x)\,dx$.

Some Continuous Distributions.

(1) *Uniform (a, b)*: $f(x) = \frac{1}{(b-a)}$, $a \le x \le b$, $F(x) = \frac{x-a}{b-a}$, $\mathbb{E}X = \frac{a+b}{2}$, $\mathbb{V}arX = \frac{(b-a)^2}{12}$, $\varphi(t) = \frac{e^{tb}-e^{ta}}{t(b-a)}$,
(2) *Exponential* (λ): $f(x) = \lambda\, e^{-\lambda\, x}$, $x \ge 0$, $F(x) = 1 - e^{-\lambda\, x}$, $\mathbb{E}X = \frac{1}{\lambda}$, $\mathbb{V}arX = \frac{1}{\lambda^2}$ $\varphi(t) = \frac{\lambda}{\lambda - t}$.
(3) *Gamma*(α, λ): $f(x) = \frac{\lambda^\alpha\, x^{\alpha-1} e^{-\lambda\, x}}{\Gamma(\alpha)}$, $x \ge 0$, where $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}\,dx$, $\Gamma(n) = (n-1)!$, $\mathbb{E}X = \frac{\alpha}{\lambda}$, $\mathbb{V}arX = \frac{\alpha}{\lambda^2}$, $\varphi(t) = (\frac{\lambda}{\lambda-t})^\alpha$.
Note. A Gamma r.v. is a sum of α independent Exponential r.v.'s.
(4) *Normal*(μ, σ): $f(x) = \frac{1}{\sqrt{2\pi}\sigma}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, $\mathbb{E}X = \mu$, $\mathbb{V}arX = \sigma^2$, $\varphi(t) = e^{\mu\, t + \frac{\sigma^2\, t^2}{2}}$.

Useful Formulas.

$$\mathbb{E}(aX + b) = a\mathbb{E}X + b,$$

$$\mathbb{V}ar(aX + b) = a^2\, \mathbb{V}arX,$$

$$\mathbb{E}g(X) = \sum_x g(x)p(x),\ \ \mathbb{E}g(X) = \int_{-\infty}^{\infty} g(x)f(x)\,dx,$$

$$\mathbb{V}arX = \mathbb{E}(X^2)\ -\ (\mathbb{E}X)^2.$$

2.5. Jointly Distributed Random Variables.

Definition. The joint probability mass function of discrete r.v.'s X and Y is $p(x, y) = \mathbb{P}(X = x,\ Y = y)$.
Definition. The marginal probability mass function $p_X(x)$ of a discrete r.v. X is obtained from $p(x, y)$ by $p_X(x) = \sum_y p(x, y)$.
Definition. The joint c.d.f. of X and Y is $F(x, y) = \mathbb{P}(X \le x,\ Y \le y)$.
Definition. The joint probability density function $f(x, y)$ of continuous r.v.'s X and Y satisfies $F(x, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f(u, v)\, dv\, du$.
Definition. The marginal density function $f_X(x)$ of a continuous r.v. X can be obtained from $f(x, y)$ by $f_X(x) = \int_{-\infty}^{\infty} f(x, y)\, dy$.
Definition. Two r.v.'s are independent if $p(x, y) = p_X(x)p_Y(y)$ (in discrete case) and $f(x, y) = f_X(x)f_Y(y)$ (in continuous case).
Definition. The covariance of X and Y is $\mathbb{C}ov(X, Y) = \mathbb{E}(X-\mathbb{E}X)(Y-\mathbb{E}Y)$.
Definition. If $\mathbb{C}ov(X, Y) = 0$, X and Y are uncorrelated.

Useful Formulas.

$$\mathbb{E}g(X, Y) = \sum_x \sum_y g(x, y)p(x, y) \ \text{ (in discrete case)},$$

$$\mathbb{E}g(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y)f(x, y)\, dx\, dy \ \text{ (in continuous case)},$$

$\forall a_1, \dots, a_n, \mathbb{E}(a_1X_1 + a_2X_2 + \cdots + a_nX_n) = a_1\mathbb{E}X_1 + a_2\mathbb{E}X_2 + \cdots + a_n\mathbb{E}X_n,$

If X and Y are independent, then $\mathbb{E}g(X)h(Y) = \mathbb{E}g(X)\mathbb{E}h(Y)$,
If X and Y are independent, then $\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t)$.

$$\mathbb{C}ov(X, Y) = \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y),$$

$$\mathbb{V}ar(X + Y) = \mathbb{V}arX + \mathbb{V}arY + 2\mathbb{C}ov(X, Y),$$

$$\mathbb{V}ar\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathbb{V}arX_i + 2\sum_{i=1}^{n} \sum_{j<i} \mathbb{C}ov(X_i, X_j).$$

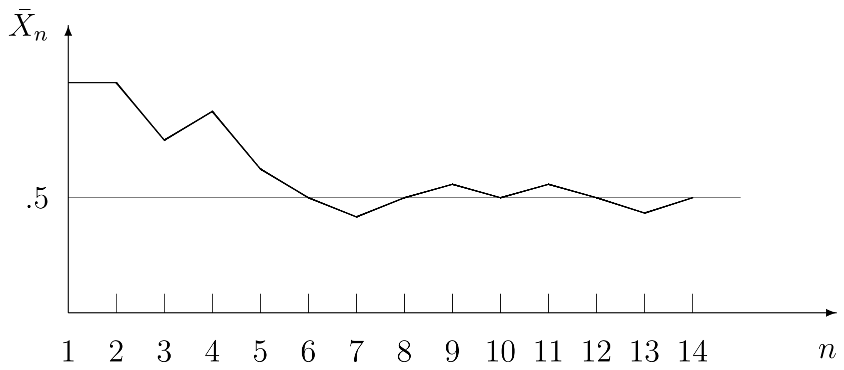
2.7. Limit Theorems.

The Strong Law of Large Numbers (LLN) Let $X_1, X_2, \dots$ be a sequence of i.i.d. r.v.'s, and let $\mu = \mathbb{E}X_i$. Then, with probability 1,

$$\bar{X}_n = \frac{X_1 + X_2 + \cdots + X_n}{n} \to \mu \ \text{ as } \ n \to \infty.$$

Example. A fair coin is flipped nine times and the following sequence is obtained $HHTHTTTHHTHTTH$. Let $X_i = 1$ if the ith flip comes up head, and 0 if it is a tail. $\mathbb{E}X_i = 0.5$.

n	1	2	3	4	5	6	7
$\bar{x}_n$	1/1=1	2/2=1	2/3≈.67	3/4=.75	3/5=.6	3/6=.5	3/7≈.43
n	8	9	10	11	12	13	14
$\bar{x}_n$	4/8=.5	5/9≈.56	5/10=.5	6/11≈.55	6/12=.5	6/13 ≈0.46	7/14=.5



The Central Limit Theorem (CLT) Let $X_1, X_2, \ldots$ be a sequence of i.i.d. r.v.'s with mean μ and variance σ^2. Then, as $n \to \infty$,

$$\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma\sqrt{n}} \to N(0,1),$$

or, equivalently,

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \to N(0,1) \text{ as } n \to \infty,$$

or, equivalently, for large n, $\bar{X}_n$ is approximately $N(\mu, \frac{\sigma^2}{n})$.

Conditional Probability and Conditional Expectation.

3.2. The Discrete Case.

Definition. The conditional probability mass function of X given $Y = y$ is

$$p_{X|Y}(x|\,y) = \mathbb{P}(X = x|\,Y = y) = \frac{\mathbb{P}(X = x, Y = y)}{\mathbb{P}(Y = y)} = \frac{p(x,y)}{p_Y(y)}.$$

Example (*). R.v.'s X and Y have the following joint probability distribution:

		x	
	p(x,y)	2	4
y	1	0.05	0.20
	3	0.20	0.30
	5	0.15	0.10

Find $p_{X|Y}(4|\,3) = \mathbb{P}(X = 4|\,Y = 3)$.
Solution: $\mathbb{P}(X = 4|\,Y = 3) = \frac{\mathbb{P}(X=4,Y=3)}{\mathbb{P}(Y=3)} = \frac{0.3}{0.2+0.3} = 0.6$.

Definition. The <u>conditional</u> expectation of X given $Y = y$ is

$$\mathbb{E}[X|\,Y = y] = \sum_x x\mathbb{P}(X = x|\,Y = y) = \sum_x x\,p_{X|Y}(x|\,y).$$

Example. In Example (*), find $\mathbb{E}[X|\,Y = 3]$.
Solution: $\mathbb{E}[X|\,Y = 3] = \sum_x x\mathbb{P}(X = x|\,Y = 3) = 2\,\mathbb{P}(X = 2|\,Y = 3) + 4\,\mathbb{P}(X = 4|\,Y = 3) = 2\,\frac{0.2}{0.5} + 4\,\frac{0.3}{0.5} = (2)(0.4) + (4)(0.6) = 3.2$.
Note that the other conditional expectations are different, $\mathbb{E}[X|\,Y = 1] = (2)(0.05/0.25) + (4)(0.20/0.25) = 3.6$ and $\mathbb{E}[X|\,Y = 5] = (2)(0.15/0.25) + (4)(0.1/0.25) = 2.8$.

3.3. The Continuous Case.

Definition. The <u>conditional</u> probability density function of X given that $Y = y$ is

$$f_{X|Y}(x|\,y) = \frac{f(x, y)}{f_Y(y)}.$$

Example. The joint p.d.f. of X and Y is

$$f(x, y) = x + y, \text{ if } 0 < x < 1,\ 0 < y < 1.$$

What is $\mathbb{P}(X \geq 0.5\,|\,Y = 0.5)$?
Solution:

$$\mathbb{P}(X \geq 1|\,Y = 0.5) = \int_{0.5}^{1} f_{X|Y}(x|\,0.5)\,dx = \int_{0.5}^{1} \frac{f(x, 0.5)}{f_Y(0.5)}\,dx$$

$$= \int_{0.5}^{1} \frac{x + 0.5}{\int_0^1 (x + 0.5)\,dx}\,dx = \int_{0.5}^{1} (x + 0.5)\,dx = 5/8.$$

Definition. The <u>conditional</u> expectation of X given that $Y = y$ is

$$\mathbb{E}[X|\,Y = y] = \int_{-\infty}^{\infty} x\,f_{X|Y}(x|\,y)\,dx.$$

Example. In the above example, compute $\mathbb{E}[X|\,Y = 0.5]$.
Solution:

$$\mathbb{E}[X|\,Y = 0.5] = \int_0^1 x\,f_{X|Y}(x|\,0.5)\,dx = \int_0^1 x(x+0.5)\,dx = 1/3+1/4 = 7/12.$$

3.4. Computing Expectations by Conditioning.

Theorem. $\mathbb{E}X = \mathbb{E}[\mathbb{E}[X|\,Y]]$.

Proof: Let $g(Y) = \mathbb{E}[X|\,Y] = \sum_x x\mathbb{P}(X = x|\,Y)$. Then,
$\mathbb{E}[g(Y)] = \sum_y g(y)\mathbb{P}(Y = y) = \sum_y \sum_x x\mathbb{P}(X = x|\,Y = y)\mathbb{P}(Y = y) = \sum_y \sum_x x \frac{\mathbb{P}(X=x,Y=y)}{\mathbb{P}(Y=y)}\mathbb{P}(Y = y) = \sum_x x \sum_y \mathbb{P}(X = x, Y = y)$
$= \sum_x x\,\mathbb{P}(X = x) = \mathbb{E}X.$ □

Example (*). The number of defects per yard in a fabric $N \sim Poisson(\lambda)$. The parameter $\lambda \sim Geom(1/3)$. Find $\mathbb{E}N$.

Solution: $\mathbb{E}N = \mathbb{E}[\mathbb{E}[N|\,\lambda]] = \mathbb{E}[\lambda] = 3$.

Definition. The r.v. $\sum_{i=1}^N X_i$, equal to the sum of a random number N of i.i.d. r.v.'s $X_i, i \ldots, N$ that are also independent of N, is called a compound r.v.

Example 3.10 on page 107 *(The Expectation of a Compound Random Variable).* Show: $\mathbb{E}[\sum_{i=1}^N X_i] = \mathbb{E}N\,\mathbb{E}X_1$.

Solution: $\mathbb{E}[\sum_{i=1}^N X_i] = \mathbb{E}[\mathbb{E}[\sum_{i=1}^N X_i|\,N]] = \mathbb{E}[\sum_{i=1}^N \mathbb{E}X_i] = \mathbb{E}[N\mathbb{E}X_1] = \mathbb{E}N\,\mathbb{E}X_1$.

Application (**): The number of students who come to instructor's office hour $N \sim Poisson(3/hour)$. The time each student takes $X_i \sim Exp$ with mean 10 minutes, independently of other students present, and the number of students present. How long, on average, the instructor can relax during her office hour?

Solution: $1 - \mathbb{E}[\sum_{i=1}^N X_i] = 1 - \mathbb{E}N\,\mathbb{E}X_1 = 1 - (3)(1/6) = 1/2$ hour.

Example 3.12 on page 109. A miner is trapped in a mine containing 3 doors. The first door leads to safety in 2 hours, the second door brings him back in 3 hours, the third – brings him back in 5 hours. If the miner chooses doors at random, what is the expected length of time until he reaches safety?

Solution: Let X be the length of time, in hours, to safety, and Y be the door he initially chooses (Y= 1,2, or 3). Then,

$$\mathbb{E}X = \mathbb{E}[\mathbb{E}[X|\,Y]] = \mathbb{E}[X|\,Y = 1]\mathbb{P}(Y = 1) + \mathbb{E}[X|\,Y = 2]\mathbb{P}(Y = 2)$$

$$+\mathbb{E}[X|\,Y = 3]\mathbb{P}(Y = 3) = \frac{1}{3}(\mathbb{E}[X|\,Y = 1] + \mathbb{E}[X|\,Y = 2] + \mathbb{E}[X|\,Y = 3])$$

$$= \frac{1}{3}(2 + 3 + \mathbb{E}X + 5 + \mathbb{E}X) \;\Rightarrow\; \mathbb{E}X = 10 \text{ hours.}$$

Computing Variances by Conditioning.

Definition. The conditional variance of X given Y is
$\mathbb{V}ar[X|\,Y] = \mathbb{E}[X^2|\,Y] - (\mathbb{E}[X|\,Y])^2$.
Theorem. $\mathbb{V}arX = \mathbb{E}[\mathbb{V}ar(X|\,Y)] + \mathbb{V}ar[\mathbb{E}(X|\,Y)]$.
Proof: see the textbook.
Example (*). Find $\mathbb{V}arN$.
Solution: $\mathbb{V}arN = \mathbb{E}[\mathbb{V}ar(N|\,\lambda)]+\mathbb{V}ar[\mathbb{E}(N|\,\lambda)] = \mathbb{E}[\lambda]+\mathbb{V}ar[\lambda] = 3+6 = 9$.
Example 3.17 on page 118 *(The Variance of a Compound Random Variable).* Show: $\mathbb{V}ar[\sum_{i=1}^{N} X_i] = (\mathbb{V}arX_1)(\mathbb{E}N) + (\mathbb{E}X_1)^2(\mathbb{V}arN)$.
Solution: $\mathbb{V}ar[\sum_{i=1}^{N} X_i] = \mathbb{E}[\mathbb{V}ar(\sum_{i=1}^{N} X_i|\,N)] + \mathbb{V}ar(\mathbb{E}[\sum_{i=1}^{N} X_i|\,N]) = \mathbb{E}[N\,\mathbb{V}arX_1] + \mathbb{V}ar(N\mathbb{E}X_1) = (\mathbb{V}arX_1)(\mathbb{E}N) + (\mathbb{E}X_1)^2(\mathbb{V}arN)$. □
Application (**): What is the variance of the time during which the instructor can relax?
Solution: $\mathbb{V}ar[1-\sum_{i=1}^{N} X_i] = \mathbb{V}ar[\sum_{i=1}^{N} X_i] = (\mathbb{V}arX_1)(\mathbb{E}N)+(\mathbb{E}X_1)^2(\mathbb{V}arN) = (1/6)^2(3) + (1/6)^2(3) = 1/6$ hour= 10 minutes.

3.5. Computing Probabilities by Conditioning.

Proposition. $p(X = x) = \sum_y \mathbb{P}(X = x|\,Y = y)\,\mathbb{P}(Y = y)$ (in discrete case), and $f_X(x) = \int_{-\infty}^{\infty} f_{X|Y}(x|\,y)f_Y(y)\,dy$ (in continuous case).
Proof:
$P(X = x) = \sum_y \mathbb{P}(X = x, Y = y) = \sum_y \mathbb{P}(X = x|\,Y = y)\,\mathbb{P}(Y = y)$. □
Example 3.18 on page 120. X and Y are independent r.v.'s, $X \sim f_X,\ Y \sim f_Y$. Find $\mathbb{P}(X < Y)$.
Solution: $\mathbb{P}(X < Y) = \int_{-\infty}^{\infty} \mathbb{P}(X < Y|\,Y = y)f_Y(y)\,dy = \int_{-\infty}^{\infty} \mathbb{P}(X < y)f_Y(y)\,dy = \int_{-\infty}^{\infty} F_X(y)f_Y(y)\,dy$. For example, if $X \sim Exp(\alpha),\ Y \sim Exp(\beta)$, then $\mathbb{P}(X < Y) = \int_0^{\infty} (1 - e^{-\alpha\,y})\beta e^{-\beta\,y}\,dy = 1 - \frac{\beta}{\alpha+\beta} = \frac{\alpha}{\alpha+\beta}$.

2.8. Stochastic Processes.

Definition. A stochastic process $\{X(t),\ t \in T\}$ is a collection of random variables, that is, for each t, $X(t)$ is a random variable.
Definition. The index t is interpreted as time, and $X(t)$ as the state of the process at time t.
Definition. The set T is the index set of the process.
Definition. If T is a countable set, the stochastic process is a discrete-time process.
Definition. If T is an interval, the stochastic process is a continuous-time process.
Definition. The state space of a stochastic process is the set of all possible values that $X(t)$ can assume.

4.1. Introduction to Discrete Markov Chains.

Definition. Consider a discrete-time stochastic process $\{X_n,\ n = 0, 1, 2, \dots\}$

with finite or countably infinite state space. Suppose that if $X_n = i$, then $X_{n+1} = j$ with fixed probability P_{ij}. The process is called a Markov chain if

$$\mathbb{P}(X_{n+1} = j | X_0 = i_0, X_1 = i_1, \dots, X_n = i) = \mathbb{P}(X_{n+1} = j | X_n = i) = P_{ij},$$

that is, the conditional distribution of any future state given the past states depends only on the present state (the Markov property).
Definition P_{ij} is called the one-step transition probability of the Markov chain.
Definition. The one-step transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} P_{00} & P_{01} & P_{02} & \dots \\ P_{10} & P_{11} & P_{12} & \dots \\ & \dots & & \\ P_{i0} & P_{i1} & P_{i2} & \dots \\ & \dots & & \end{Vmatrix}.$$

Example. If a machine is broken today, it will be fixed tomorrow with probability α. If the machine is in operating condition today, it will be broken tomorrow with probability β. Say, the process is in state 0 if the machine is broken and in state 1 if the machine is working. Then we have a two-state Markov chain with transition probability matrix

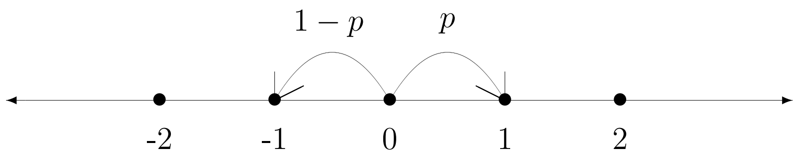
$$\mathbf{P} = \begin{Vmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{Vmatrix}.$$

Example 4.4 on page 182 – 183. Suppose if it rained for the past two days, it will rain tomorrow with probability 0.7; if it rained today but not yesterday, it will rain tomorrow with probability 0.5; if it rained yesterday but not today, it will rain tomorrow with probability 0.4; if is has not rained for the past two days, it will rain tomorrow with probability 0.2. This process is not a Markov chain (!) since whether it rains tomorrow depends not just on today's weather but on yesterday's as well. However, the process can be transformed into a Markov chain. Consider the following states:
state 0 if it rained both today and yesterday,
state 1 if it rained today but not yesterday,
state 2 if it rained yesterday but not today,
state 3 if it did not rain either today nor yesterday.
Thus, we have a four-state Markov chain with the transition probability matrix

$$\mathbf{P} = \begin{Vmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{Vmatrix}.$$

Verification: $P_{00} = \mathbb{P}(\text{rains tomorrow and today} | \text{ rains today and yesterday}) = \mathbb{P}(\text{rains tomorrow} | \text{ rains today and yesterday}) = 0.7$; $P_{01} = \mathbb{P}(\text{rains tomorrow but not today} | \text{ rains today and yesterday}) = 0$.

Example 4.5 on page 183 *(A Random Walk Model).* A Markov chain which state space is the integers $i = 0, \pm 1, \pm 2, \dots$ and which transition probabilities are $P_{i,i+1} = p = 1 - P_{i,i-1}$ is called a random walk.



The transition probability matrix of a random walk is

$$\mathbf{P} = \begin{Vmatrix} & & & & \dots & & & & \\ \dots & 0 & 1-p & 0 & p & 0 & \dots & & \\ & \dots & 0 & 1-p & 0 & p & 0 & \dots & \\ & & \dots & 0 & 1-p & 0 & p & 0 & \dots \\ & & & & \dots & & & & \end{Vmatrix}.$$

Example. A gambler either wins \$1 with probability 2/3 or loses \$1 with probability 1/3. He plays until either he is broke or he gains a fortune of \$3. The model is called a finite random walk with absorbing barriers (once entered, the states 0 and 3 are never left). The transition probability matrix is

$$\mathbf{P} = \left\| \begin{matrix} 1 & 0 & 0 & 0 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 0 & 0 & 0 & 1 \end{matrix} \right\|.$$

Remark. The probabilities in each row of a transition probability matrix must sum up to 1 !!!

4.2. Chapman – Kolmogorov Equations

Definition. The n-step transition probability P_{ij}^n is the probability that a process in state i will be in state j after n transitions, that is,

$$P_{ij}^n = \mathbb{P}(X_{k+n} = j |\, X_k = i).$$

Theorem *(The Chapman – Kolmogorov equations).* For any $n, m \geq 0$ and for any i, j,

$$P_{ij}^{n+m} = \sum_{k=0}^{\infty} P_{ik}^n \, P_{kj}^m.$$

Proof: $P_{ij}^{n+m} = \mathbb{P}(X_{n+m} = j |\, X_0 = i) =$

$$= \frac{\mathbb{P}(X_{n+m} = j,\, X_0 = i)}{\mathbb{P}(X_0 = i)} = \sum_{k=0}^{\infty} \frac{\mathbb{P}(X_{n+m} = j,\, X_n = k,\, X_0 = i)}{\mathbb{P}(X_0 = i)}$$

$$= \sum_{k=0}^{\infty} \frac{\mathbb{P}(X_{n+m} = j |\, X_n = k,\, X_0 = i)\mathbb{P}(X_n = k,\, X_0 = i)}{\mathbb{P}(X_0 = i)}$$

$$= \sum_{k=0}^{\infty} \mathbb{P}(X_{n+m} = j |\, X_n = k,\, X_0 = i)\mathbb{P}(X_n = k |\, X_0 = i)$$

$$= \sum_{k=0}^{\infty} \mathbb{P}(X_{n+m} = j |\, X_n = k)\mathbb{P}(X_n = k |\, X_0 = i) \; (by\ the\ Markov\ property)$$

$$= \sum_{k=0}^{\infty} P_{kj}^m \, P_{ik}^n. \quad \square$$

Remark. Let $\mathbf{P}^{(n)}$ denote the n-step transition probability matrix. Note that $\mathbf{P}^{(1)} = \mathbf{P}$. The Chapman – Kolmogorov equations in the matrix form are

$$\forall\, n, m, \;\; \mathbf{P}^{(n+m)} = \mathbf{P}^{(n)} \cdot \mathbf{P}^{(m)}.$$

Corollary of the theorem. $\forall\, n,\ \ \mathbf{P}^{(n)} = \mathbf{P}^n$.
Proof: $\mathbf{P}^{(2)} = \mathbf{P}^{(1+1)} = \mathbf{P}\cdot\mathbf{P} = \mathbf{P}^2$, $\mathbf{P}^{(3)} = \mathbf{P}^{(2)}\cdot\mathbf{P} = \mathbf{P}^3, \dots, \mathbf{P}^{(n)} = \mathbf{P}^{(n-1)}\cdot\mathbf{P} = \mathbf{P}^{n-1}\cdot\mathbf{P} = \mathbf{P}^n$. $\square$
Example (*). During a particular day, a machine can be in either of two states. State 0 if it is broken, or State 1 if it is working. The transition probability matrix is

$$\mathbf{P} = \begin{Vmatrix} 0.5 & 0.5 \\ 0.1 & 0.9 \end{Vmatrix}.$$

What is the conditional probability that the machine will be working four days from now given that it is working today?
Solution:

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{Vmatrix} 0.5 & 0.5 \\ 0.1 & 0.9 \end{Vmatrix}^2 = \begin{Vmatrix} 0.3 & 0.7 \\ 0.14 & 0.86 \end{Vmatrix},$$

$$\mathbf{P}^{(4)} = (\mathbf{P}^2)^2 = \begin{Vmatrix} 0.3 & 0.7 \\ 0.14 & 0.86 \end{Vmatrix}^2 = \begin{Vmatrix} 0.1880 & 0.8120 \\ 0.1624 & 0.8376 \end{Vmatrix}.$$

The desired probability is $P_{11}^4 = 0.8376$.

Example 4.9 on page 187. Given that it rained on Monday and Tuesday, what is the probability that it will rain on Thursday?
Solution: The process is initially in state 0. We want the find the probability that in two steps it will be either in state 0 or in state 1. This probability is $P_{00}^2 + P_{01}^2$. Now, the two-step transition probability matrix is

$$\mathbf{P}^{(2)} = \mathbf{P}^2 = \begin{Vmatrix} 0.7 & 0 & 0.3 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.4 & 0 & 0.6 \\ 0 & 0.2 & 0 & 0.8 \end{Vmatrix}^2 = \begin{Vmatrix} 0.49 & 0.12 & * & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{Vmatrix}.$$

Thus, $P_{00}^2 + P_{01}^2 = 0.49 + 0.12 = 0.61$.

Example. A gambler has an initial capital of \$1. He either wins \$1 with probability 1/2 or loses \$1 with probability 1/2. Whenever he goes broke his rich uncle gives him \$1, and whenever the gambler wins total of \$3, he gives \$1 back to his uncle. What is the probability that after 3 games, he has \$2?
Solution: The transition probability matrix is

$$\mathbf{P} = \begin{Vmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{Vmatrix}.$$

The model is called a <u>symmetric random walk with reflecting barriers</u> (the probabilities of losing and winning are equal – symmetric, and the states 0 and 3 "bounce" back – reflecting barriers).

The probability we are interested in is P_{12}^3. The three-step transition matrix is

$$\mathbf{P}^{(3)} = \mathbf{P}^3 = \begin{Vmatrix} 1/2 & 0 & 1/2 & 0 \\ 0 & 3/4 & 0 & 1/4 \\ 1/4 & 0 & 3/4 & 0 \\ 0 & 1/2 & 0 & 1/2 \end{Vmatrix} \cdot \begin{Vmatrix} 0 & 1 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 0 & 0 & 1 & 0 \end{Vmatrix} = \begin{Vmatrix} * & * & * & * \\ * & * & 5/8 & * \\ * & * & * & * \\ * & * & * & * \end{Vmatrix}.$$

Thus, $P_{12}^3 = 5/8$.

Proposition. Let $p_i = \mathbb{P}(X_0 = i)$ denote the probability distribution of the initial state. Then, the unconditional distribution of the state at time n is

$$\mathbb{P}(X_n = j) = \sum_{i=0}^{\infty} P_{ij}^n \, p_i.$$

Proof: $\mathbb{P}(X_n = j) = \sum_{i=0}^{\infty} \mathbb{P}(X_n = j | \, X_0 = i) \, \mathbb{P}(X_0 = i) = \sum_{i=0}^{\infty} P_{ij}^n \, p_i.$ □

Example. In Example (*), assume that $\mathbb{P}($ machine is broken today$) = 0.2$. Find the probability that four days from today the machine will be broken. *Solution:* $\mathbb{P}(X_4 = 0) = 0.2 \, P_{00}^4 + 0.8 \, P_{10}^4 = (0.2)(0.188) + (0.8)(0.1624) \approx 0.168$.

4.3. Classification of States.

Definition. State j is called accessible from state i if $P_{ij}^n > 0$ for some $n \geq 0$, or, equivalently, if, starting in i, the process will ever enter state j with positive probability. *Notation.* $i \to j$.
Proof of equivalence: $\Rightarrow$ Suppose there exists $n_0 \geq 0$ such that $P_{ij}^{n_0} > 0$. Then, $\quad \mathbb{P}(\text{ever enter } j \, | \text{ start in } i\,) = \mathbb{P}\Big(\cup_{n=0}^{\infty} \{X_n = j\} | \, X_0 = i \Big)$

$$\geq \mathbb{P}(X_{n_0} = j | \, X_0 = i) = P_{ij}^{n_0} > 0.$$

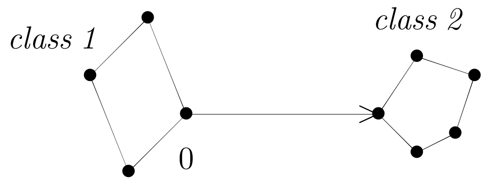
$\Leftarrow$ Will show that if state j is not accessible from state i by the first definition, it is not accessible by the second definition. Suppose that for all $n \geq 0$, $P_{ij}^n = 0$. Then,

$$\mathbb{P}(\text{ever enter } j \, | \text{start in } i\,) = \mathbb{P}\Big(\cup_{n=0}^{\infty} \{X_n = j\} | \, X_0 = i \Big)$$

$$\leq \sum_{n=0}^{\infty} \mathbb{P}(X_n = j | \, X_0 = i) \; (\textit{by Boole's inequality}) = \sum_{n=0}^{\infty} P_{ij}^n = 0. \quad \square$$

Definition. Two states i and j are said to communicate, if they are accessible to each other. *Notation.* $i \leftrightarrow j$.
Proposition. The relation of communication is an equivalence relation, that is (1) it is reflexive: state i communicates with itself, for any i ($\forall i, \; i \leftrightarrow i$); (2) it is symmetric: if state i communicates with state j, then state j communicates with state i ($i \leftrightarrow j \Rightarrow j \leftrightarrow i$); (3) it is transitive: if state i communicates with state j, and state j communicates with state k, then state i communicates with state k ($i \leftrightarrow j, \; j \leftrightarrow k \Rightarrow i \leftrightarrow k$).
Proof: (1) $P_{ii}^0 = \mathbf{P}(X_0 = i | \, X_0 = i) = 1 > 0 \Rightarrow$ by definition, $i \leftrightarrow i$; (2) $i \leftrightarrow j \Rightarrow \exists \, n, m \; P_{ij}^n > 0$, and $P_{ji}^m > 0 \Rightarrow$ by definition, $j \leftrightarrow i$; (3) $i \leftrightarrow j$ and $j \leftrightarrow k \Rightarrow \exists \, l, m, n, q \ni P_{ij}^l > 0$, $P_{ji}^m > 0$, $P_{jk}^n > 0$, $P_{kj}^q > 0 \Rightarrow P_{ik}^{l+n} = \sum_{r=0}^{\infty} P_{ir}^l \, P_{rk}^n \geq P_{ij}^l \, P_{jk}^n > 0 \Rightarrow i \to k$. Also, $P_{ki}^{q+m} = \sum_{r=0}^{\infty} P_{kr}^q \, P_{ri}^m \geq P_{kj}^q \, P_{ji}^m > 0 \Rightarrow k \to i$. □

Definition. Two states that communicate are said to belong to the same class.
Remark. By the above Proposition, the relation of communication divides the state space into separate classes. *Illustration.*



If state 0 communicated with state 1, all states in class 1 would communicate with all states in class 2, thus all the states of the chain would belong to only one class.

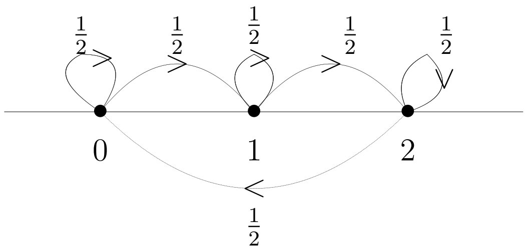
Definition. A Markov chain is irreducible if there is only one class, that is all states communicate with each other.

Example. Consider a Markov chain with the transition matrix

$$\mathbb{P} = \begin{Vmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{Vmatrix}.$$

How many classes does the Markov chain have?

Solution:



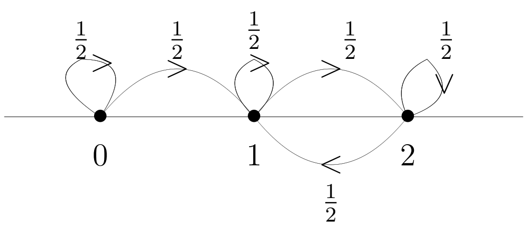
It is possible to go $0 \to 1 \to 2 \to 0$. Therefore, $P_{01} > 0$, $P_{02}^2 > 0$, $P_{10}^2 > 0$, $P_{12} > 0$, $P_{20} > 0$, $P_{21}^2 > 0$. Thus, all three states communicate with each other. The chain has one class and is irreducible.

Example. Suppose a Markov chain has the transition matrix

$$\mathbb{P} = \begin{Vmatrix} 1/2 & 1/2 & 0 \\ 0 & 1/2 & 1/2 \\ 0 & 1/2 & 1/2 \end{Vmatrix}.$$

How many classes does the Markov chain have?

Solution:



It is possible to go $0 \to 1 \to 2 \to 1$. Therefore, $1 \leftrightarrow 2$ but $0 \not\leftrightarrow 1$ and $0 \not\leftrightarrow 2$. Thus, there are two classes $\{0\}$ and $\{1, 2\}$.

Notation. Denote f_i the probability that, starting in state i, the process will ever reenter state i.
Definition. State i is recurrent if $f_i = 1$.
Definition. State i is transient if $f_i < 1$.
Note. Any state is either recurrent or transient.

Proposition 1. State i is recurrent iff $\sum_{n=1}^{\infty} P_{ii}^n = \infty$, and state i is transient iff $\sum_{n=1}^{\infty} P_{ii}^n < \infty$.
Proof: If a state is recurrent, then with probability one the process will revisit it, and then will revisit it again, and again, and again. Thus, a state is recurrent iff the expected number of times the process returns to it is infinite. On the other hand, a state is transient iff this expected value is finite. In fact, the probability that a process is in state i exactly n times is $f_i^{n-1}(1-f_i)$, which is a geometric distribution with mean $1/(1-f_i) < \infty$.
Introduce notation $I_n = \begin{cases} 1, & \text{if } X_n = i \\ 0, & \text{if } X_n \neq i \end{cases}$.
Then, $\sum_{n=0}^{\infty} I_n$ equals to the number of times the process is in state i.

$$\mathbb{E}\left[\sum_{n=0}^{\infty} I_n |\, X_0 = i\right] = \sum_{n=0}^{\infty} \mathbb{E}[I_n |\, X_0 = i] = \sum_{n=0}^{\infty} \mathbb{P}(X_n = i |\, X_0 = i) = \sum_{n=0}^{\infty} P_{ii}^n.$$

Therefore, state i is recurrent iff $\sum_{n=0}^{\infty} P_{ii}^n = \infty$. Also, state i is transient iff $\sum_{n=0}^{\infty} P_{ii}^n < \infty$. □

Proposition 2. At least one state of a finite Markov chain must be recurrent.
Proof: Suppose all states $0, 1, \dots, M$ are transient. Each state is visited finite number of times $T_0, T_1, \dots, T_M$. Thus, after time $T = max(T_0, T_1, \dots, T_M)$ no states are visited. It is impossible. □

Proposition 3. Recurrence and transience are class properties.
Proof: Need to show that if states i and j belong to the same class, that is, $i \leftrightarrow j$, and i is recurrent, then j is recurrent.
Since $i \leftrightarrow j$ there exist k and m such that $P_{ij}^k > 0$ and $P_{ji}^m > 0$. Therefore, for any n, $P_{jj}^{m+n+k} \geq P_{ji}^m P_{ii}^n P_{ij}^k$. Summing over n, we get $\sum_{n=0}^{\infty} P_{jj}^{m+n+k} \geq P_{ji}^m P_{ij}^k \sum_{n=0}^{\infty} P_{ii}^n = \infty$ since state i is recurrent. By proposition 1, state j is also recurrent. □

Example. Consider a Markov chain with transition matrix

$$\mathbf{P} = \begin{Vmatrix} 1/3 & 2/3 & 0 \\ 1/4 & 0 & 3/4 \\ 0 & 1/2 & 1/2 \end{Vmatrix}.$$

Which states are recurrent and which are transient?
Solution: Since $0 \leftrightarrow 1$ and $1 \leftrightarrow 2$, there is only one class and the chain is irreducible. By proposition 2, the chain must have at least one recurrent

state, and, since there is only one class, by proposition 3, all states must be recurrent. Thus, all threes states are recurrent.

Example. Which states are recurrent and which are transient in a Markov chain with transition matrix

$$\mathbf{P} = \begin{Vmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0 & 0.1 & 0.9 \\ 0 & 0 & 0.4 & 0.6 \end{Vmatrix}?$$

Solution: Problems like this one are easier done by definition rather than by Proposition 1. State 0 is an absorbing state, therefore, the probability of never coming back to 0 is $1 - f_0 = 0 \Rightarrow f_0 = 1 \Rightarrow$ state 0 is recurrent. Further, once the process leaves state 1, it cannot come back to it, therefore, $f_1 = 0 < 1 \Rightarrow$ state 1 is transient. States 2 and 3 communicate, so, they form a class and are either both recurrent or both transient. To see, for example, that state 2 is recurrent, write $1 - f_2 = P_{23}\mathbb{P}(stays\ in\ 3\ forever) = P_{23}(P_{33})^\infty = 0 \Rightarrow f_2 = 1$. Thus, states 2 and 3 are recurrent.

Example. Which states are recurrent and which are transient in a Markov chain with transition matrix

$$\mathbf{P} = \begin{Vmatrix} 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0.1 & 0 & 0.9 \\ 0 & 0 & 0.4 & 0.6 \end{Vmatrix}?$$

Solution: State 0 is reflecting and therefore transient ($f_0 = 0$). State 1 is absorbing and therefore recurrent ($f_1 = 1$). States 2 and 3 belong to the same class ($2 \leftrightarrow 3$), thus, they are either both recurrent or both transient. To see, for example, that state 2 is transient, write $1 - f_2 = P_{23}\mathbb{P}(stays\ in\ 3\ forever) + P_{21}\mathbb{P}(stays\ in\ 1\ forever) = P_{23}(P_{33})^\infty + P_{21} = 0 + 0.1 = 0.1 \Rightarrow f_2 = 0.9 < 1$. Thus, states 2 and 3 are transient.

4.3. Example 4.15 and further.

Proposition (*Stirling's Formula*). For large n, $n! \approx \sqrt{2\pi n}\, n^n\, e^{-n}$.
Probabilistic Proof: Let $X_1, X_2, \ldots$ be i.i.d. $Poisson(1)$ r.v.'s. Denote $S_n = \sum_{i=1}^n X_i$. Note that $S_n \sim Poisson(n)$ and therefore $\mathbb{E}S_n = n$ and $\mathbb{V}arS_n = n$. Thus, by the CLT, $(S_n - n)/\sqrt{n} \overset{approx.}{\sim} N(0,1)$. We write,

$$\mathbb{P}(S_n = n) = \mathbb{P}(n - 1 < S_n \le n) = \mathbb{P}(-1/\sqrt{n} < (S_n - n)/\sqrt{n} \le 0)$$

$$\approx \int_{-1/\sqrt{n}}^{0} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \quad \text{(by the CLT)}$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{1/\sqrt{n}} e^{-x^2/2}\, dx \quad \text{(by symmetry)}$$

$$= \frac{1}{\sqrt{2\pi}} \int_0^{1/\sqrt{n}} (1 - x^2/2 + x^4/8 - \dots)\, dx \quad \text{(by Taylor's expansion when } x \text{ is small)}$$

$$\approx \frac{1}{\sqrt{2\pi}} \int_0^{1/\sqrt{n}} 1\, dx = \frac{1}{\sqrt{2\pi n}}.$$

Further, since $S_n \sim Poisson(n)$, $\mathbb{P}(S_n = n) = \frac{n^n e^{-n}}{n!}$. Hence, for large n, $\frac{n^n e^{-n}}{n!} \approx \frac{1}{\sqrt{2\pi n}}$, or, equivalently, $n! \approx \sqrt{2\pi n}\, n^n e^{-n}$. □

Proposition. One-dimensional random walk is recurrent iff it is symmetric.
Proof: Consider a random walk on the real line, that is, consider a Markov chain with states $0, \pm 1, \pm 2, \dots$ and transition probabilities $P_{i,i+1} = p = 1 - P_{i,i-1}$. All states communicate, therefore, they are either all recurrent or all transient. Take one state, say state 0. We will use Proposition 1 (from last time) to determine when this state is recurrent and when it is transient. So, we have to determine when $\sum_{n=0}^{\infty} P_{00}^n$ is infinite or finite. The process cannot return to 0 in odd number of steps, so $P_{00}^{2k+1} = 0,\ k = 0, 1, \dots$. To come back to 0 in even number of steps $2k$, the process has to go up k steps and go down k steps, therefore, $P_{00}^{2k} = \binom{2k}{k} p^k (1-p)^k,\ k = 1, \dots$. Using Stirling's approximation, we have,

$$P_{00}^{2k} = \frac{(2k)!}{k!k!} p^k (1-p)^k \approx \frac{\sqrt{2\pi 2k}(2k)^{2k} e^{-2k}}{2\pi k\, k^{2k} e^{-2k}} (p(1-p))^k$$

$$= \frac{2^{2k}}{\sqrt{\pi k}} (p(1-p))^k = \frac{(4p(1-p))^k}{\sqrt{\pi k}}.$$

Now, $4p(1-p) \le 1$ for all p, $0 \le p \le 1$, and $4p(1-p) = 1$ if $p = 1/2$. Therefore, $\sum_{k=1}^{\infty} P_{00}^k \approx \sum_{k=1}^{\infty} \frac{(4p(1-p))^k}{\sqrt{\pi k}} \begin{cases} = \infty, & \text{if } p = 1/2 \\ < \infty, & \text{if } p \ne 1/2 \end{cases}$. Thus, we have proved that a one-dimensional random walk is recurrent iff $p = 1/2$ (symmetric RW). □

Remark. In one and two dimensions, a symmetric random walk is recurrent, and non-symmetric RW is transient, whereas in three or higher dimensions any RW is transient (without proof). If we interpret a SRW as a wandering of a drunk person, then this result asserts that "a drunk man will find his way home but a drunk bird may get lost forever" (Kakutani, UCLA colloquium).

4.4. Periodicity.

Definition. State i has period d if $P_{ii}^n = 0$ whenever n is not divisible by d, and d is the largest such integer. (If $P_{ii}^n = 0$ for all $n > 0$, then define the period of i to be infinite.)
Example. A process visits state 0 in 4, 8, 12, 16, etc. steps and doesn't visit otherwise. The period of state 0 is 4, since 4 is the greatest common divisor of 4, 8, 12, 16, etc.

Example. Find the period of any state in a one-dimensional random walk.
Solution: Any state can be visited in 2, 4, 6, etc. transitions, therefore, the period of any state is 2.
Definition. A state with period 1 is aperiodic.
Example. Suppose a state of a finite Markov chain can be revisited in 3, 5 or 7 steps only. The g.c.d.(3,5,7)=1, therefore, the state is aperiodic.
Proposition. Periodicity is a class property.
Example. Find the period of each state in the MC governed by the matrix

$$\mathbf{P} = \begin{Vmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ .5 & 0 & .5 & 0 & 0 & 0 \\ .2 & .6 & 0 & .2 & 0 & 0 \\ 0 & 0 & 0 & .5 & .5 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{Vmatrix}.$$

Solution: State 0 is an absorbing state. The process revisits state 0 in 1, 2, 3, etc. steps, so $d_0 = 1$. States 1 and 2 form a class, therefore, they have the same period. The process returns to state 1 or state 2 only in 2, 4, 6, etc. steps, thus, $d_1 = d_2 = 2$. States 3 and 4 communicate, thus they belong to the same class. The process returns to state 3 in 1, 2, 3, etc. steps, and to state 4 in 2, 3, 4, etc. steps, therefore, $d_3 = d_4 = 1$. State 5 is a reflecting state. The process never comes back to 5, therefore, state 5 has period ∞. Hence, states 0, 3, and 4 are aperiodic; states 1 and 2 have period 2, and state 5 has an infinite period.

4.4. Limiting Probabilities.

Definition. If a state i is recurrent, it is positive recurrent if the expected time until the process returns to i is finite. If the expected time is infinite, the state is called null recurrent.
Example. As was shown on the last lecture, a one-dimensional symmetric RW is recurrent with $P_{00}^{2k} \approx \frac{1}{\sqrt{\pi k}}$. Therefore, the expected number of steps until the chain returns to 0 is $\sum_{k=1}^{\infty} 2k\, P_{00}^{2k} \approx \sum_{k=1}^{\infty} \frac{2k}{\sqrt{\pi k}} = \infty$. Thus, a one-dim SRW is null recurrent.
Proposition. In a finite-state Markov chain all recurrent states are positive recurrent.
Proposition. Positive recurrence is a class property.
Theorem. For an irreducible positive recurrent Markov chain limiting probabilities $\pi_j = \lim_{n\to\infty} P_{ij}^n$ exist, are independent of the initial state i, and are the unique nonnegative solutions of

$$\pi_j = \sum_{i=0}^{\infty} \pi_i\, P_{ij}, \quad \sum_{j=0}^{\infty} \pi_j = 1.$$

Remark. The limiting probabilities π_j's can be interpreted as the long-run proportion of time that the Markov chain spends in state j.

Proof: It is obvious that π_j's must sum up to one. Remained to show that $\pi_j = \sum_{i=0}^{\infty} \pi_i P_{ij}$. Conditioning on the state at time n, we write

$$\mathbb{P}(X_{n+1} = j) = \sum_{i=0}^{\infty} \mathbb{P}(X_{n+1} = j \,|\, X_n = i)\mathbb{P}(X_n = i) = \sum_{i=0}^{\infty} P_{ij}\mathbb{P}(X_n = i).$$

Letting $n \to \infty$, and pretending we can exchange the limit and the summation signs, we get

$$\pi_j = \sum_{i=0}^{\infty} P_{ij}\pi_i. \quad \square$$

Example. In the long run, what proportion of time is the process in each of the two states if the transition matrix is

$$\mathbf{P} = \begin{Vmatrix} 0.5 & 0.5 \\ 0.3 & 0.7 \end{Vmatrix}?$$

Solution: π_0 and π_1 satisfy

$$\begin{cases} \pi_0 = P_{00}\pi_0 + P_{10}\pi_1 \\ \pi_1 = P_{01}\pi_0 + P_{11}\pi_1 \\ \pi_0 + \pi_1 = 1 \end{cases} \Leftrightarrow \begin{cases} \pi_0 = 0.5\pi_0 + 0.3\pi_1 \\ \pi_1 = 0.5\pi_0 + 0.7\pi_1 \\ \pi_0 + \pi_1 = 1 \end{cases}$$

$$\Leftrightarrow \begin{cases} 0.5\pi_0 = 0.3\pi_1 \\ \pi_0 = 1 - \pi_1 \end{cases} \Leftrightarrow \begin{cases} \pi_1 = 0.625 \\ \pi_0 = 0.375. \end{cases}$$

Notice that there is a redundancy in writing the $n + 1$ equations for the n unknowns. These equations do not contradict each other and the system can always be reduced to n equations for n unknowns. So, we could've started off by writing

$$\begin{cases} \pi_0 = P_{00}\pi_0 + P_{10}\pi_1 \\ \pi_0 + \pi_1 = 1. \end{cases}$$

Example (*Exercise 30 on page 257*). Three out of four trucks on the road are followed by a car, while only one out of every five cars is followed by a truck. What fraction of vehicles on the road are trucks?
Solution: Let state 0 = truck, state 1= car. The transition probabilities are $P_{01} = \mathbb{P}(\text{truck is followed by a car}) = 0.75,\ P_{00} = 0.25,\ P_{10} = 0.2,\ P_{11} = 0.8.$ We need to find the limiting probability π_0. The limiting probabilities satisfy $\pi_0 = 0.25\pi_0 + 0.2\pi_1,\ \pi_0 + \pi_1 = 1$. Therefore, $\pi_0 \approx 0.21$, so approximately 21% of vehicles on the road are trucks.
Example. A basketball player gets paid \$30,000 per game if he scores 12 or less points (State 0), \$50,000 if he scores between 13 and 25 points (State 1), and \$100,000 if above 25 points (State 2). The transitions between the states on two consecutive games can be modelled by a MC with transition

matrix

$$\mathbf{P} = \begin{Vmatrix} 0 & 1/3 & 2/3 \\ 1/3 & 0 & 2/3 \\ 1 & 0 & 0 \end{Vmatrix}.$$

What is the long run earning rate of the player?
Solution: The limiting probabilities π_0, π_1 and π_2 satisfy

$$\begin{cases} \pi_0 = P_{00}\pi_0 + P_{10}\pi_1 + P_{20}\pi_2 \\ \pi_1 = P_{01}\pi_0 + P_{11}\pi_1 + P_{21}\pi_2 \\ \pi_0 + \pi_1 + \pi_2 = 1 \end{cases} \Leftrightarrow \begin{cases} \pi_0 = (1/3)\pi_1 + \pi_2 \\ \pi_1 = (1/3)\pi_0 \\ \pi_0 + \pi_1 + \pi_2 = 1 \end{cases}$$

$$\Leftrightarrow \begin{cases} \pi_0 = 9/20 = 0.45 \\ \pi_1 = 3/20 = 0.15 \\ \pi_2 = 8/20 = 0.4 \end{cases}$$

Let X be the player's earning per game. Then, $\mathbb{E}X = (30,000)(\pi_0) + (50,000)(\pi_1) + (100,000)(\pi_2) = (30,000)(0.45) + (50,000)(0.15) + (100,000)(0.4) = \$61,000$.

APPLICATIONS OF MARKOV CHAINS

EXERCISE 1. A certain species of flower has three states. State 0=sustainable, State 1=endangered, and State 2=extinct. States are assigned at the beginning of each year. Transitions between states are modeled by a *non-homogenous* Markov chain with one-year transition probability matrices $\mathbf{P}_n, n = 1, 2, 3, \ldots$ given as:

$$\mathbf{P}_1 = \begin{Vmatrix} 0.85 & 0.15 & 0 \\ 0 & 0.7 & 0.3 \\ 0 & 0 & 1 \end{Vmatrix}, \quad \mathbf{P}_2 = \begin{Vmatrix} 0.9 & 0.1 & 0 \\ 0.1 & 0.7 & 0.2 \\ 0 & 0 & 1 \end{Vmatrix}, \quad \mathbf{P}_3 = \begin{Vmatrix} 0.95 & 0.05 & 0 \\ 0.2 & 0.7 & 0.1 \\ 0 & 0 & 1 \end{Vmatrix},$$

$$\mathbf{P}_n = \begin{Vmatrix} 0.95 & 0.05 & 0 \\ 0.5 & 0.5 & 0 \\ 0 & 0 & 1 \end{Vmatrix} \quad \text{for } n = 4, 5, \ldots.$$

Calculate the probability that a species endangered (State 1) at the start of year 1 will ever become extinct (State 2).

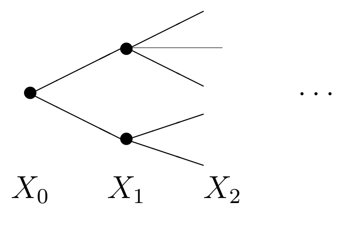
EXERCISE 2. (Inventory System). A store that sells computers is open for business Monday through Friday 8am to 5pm. It uses the following operating policy to control the inventory of computers. At 5pm on Friday, the store clerk checks to see how many computers are still in stock. If the number is less than two, then he orders enough computers to bring the total in stock up to five at the beginning of the business day on Monday. If the number in stock is two or more, no action is taken. The demand for the computers

during the week is a Poisson random variable with mean 3. Any demand that cannot be immediately satisfied is lost.
(a) Develop a Markov chain model of the computer inventory. Prove that it is a Markov chain. Specify its transition probability matrix.
(b) Suppose there are 5 computers in stock on Monday. Compute the probability that the inventory trajectory over the next four Mondays is as follows: 4, 2, 5, and 3.
(c) The weekly storage cost is \$50 per computer that is in the store at the beginning of the week. Compute the storage cost the store expects to pay weekly in a long-run.

EXERCISE 3. Rolo, a Chihuahua, spends most of the daytime sleeping in the kitchen. When a person enters the kitchen, Rolo greets him or her and wags her tail for an average time of one minute. At the end of this period Rolo is fed with probability 1/4, patted briefly with probability 5/8, or taken for a walk with probability 1/8. If fed, Rolo spends an average of two minutes eating. The walks take 15 minutes on average. After eating, being patted, or walking, she returns to sleep. Assume that people enter the kitchen on average every hour.
(a) Find a Markov chain model with four states: { sleep, greet, eat, walk }. Specify the transition probability matrix.
(b) Find the limiting probabilities of the four states.

4.7. Branching Processes.

Definition. A population of individuals is considered. Each individual lives one unit period and at the end of its life produces a random number of individuals. The offspring behaves independently of each other, each individual living one unit period, at the end of which a new offspring is produced, and so on. Denote X_0 the size of the initial population, the zeroth generation, X_1 – the size of the first generation, and, in general, X_n – the size of the nth generation. $\{X_n\}$ is called a discrete time branching process (or the Galton-Watson process).

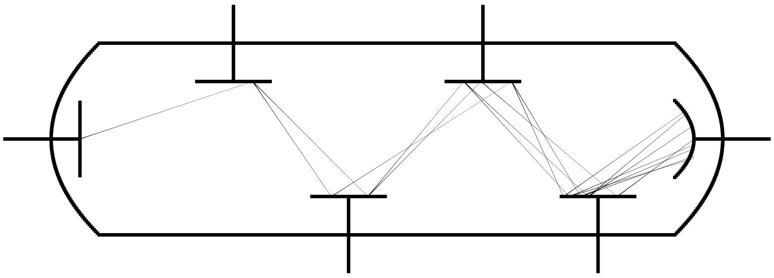


Example (*Survival of Family Names*). The family name is inherited by sons only. Suppose that each individual has some random number of male offspring. A question of interest may be how large the initial family should be to insure that the family name does not die out. Historically, this problem was introduced by Francis Galton in 1873, and solved by Henry Watson in the same year. They were concerned with the extinction of family names in Britain.

Example (*Survival of Mutant Genes*). Each individual has a chance to produce an offspring with a mutant gene. This gene may become the first in a sequence of generations of a particular mutant gene. We may inquire about the chances of survival of the mutant gene within the population of the original genes.

Example (*Smallpox Epidemics*). Smallpox can be modelled fairly well by a Galton-Watson process. Say, the disease is imported by a visitor, who transmits it to a number of residents. They, in their turn, independently infect a random number of people, and so on.

Example (*Electron Multipliers*). An electron multiplier is a device that amplifies a weak current of electrons. A series of plates are set up in the path of electrons emitted by a source. Each electron, as it strikes the first plate, generates a random number of electrons, which in turn strike the next plate and produce more electrons, etc. Let X_n be the number of electrons emitted from the nth plate. Then, $\{X_n\}$ is a branching process.



Notation. X_n is the size of the nth generation. Let Z be the number of offspring of a single individual. Denote $P_k = \mathbb{P}(Z = k)$, $\mu = \mathbb{E}Z = \sum_{k=0}^{\infty} kP_k$, $\sigma^2 = \mathbb{V}arZ = \sum_{k=0}^{\infty}(k-\mu)^2 P_k$. Note that the size of the nth generation $X_n = \sum_{i=1}^{X_{n-1}} Z_i$, where Z_i is the number of offspring of the ith individual in the $(n-1)$st generation.

Proposition. A branching process is a Markov chain.

Proof:
$$\mathbb{P}(X_n = i_n | X_0 = i_0, \ldots, X_{n-1} = i_{n-1})$$
$$= \mathbb{P}\left(\sum_{i=0}^{i_{n-1}} Z_i = i_n | X_0 = i_0, \ldots, X_{n-1} = i_{n-1}\right)$$
$$= \{Z\text{'s are independent of } X\text{'s}\}$$
$$= \mathbb{P}\left(\sum_{i=0}^{i_{n-1}} Z_i = i_n | X_{n-1} = i_{n-1}\right) = \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}). \quad \square$$

Proposition. Assume $X_0 = 1$, so the branching process starts with a single individual at time zero. Then, the mean size of the nth generation is $\mathbb{E}X_n = \mu^n$, and the variance $\mathbb{V}arX_n = \begin{cases} \sigma^2 \mu^{n-1}\left(\frac{1-\mu^n}{1-\mu}\right), & \text{if } \mu \neq 1 \\ n\sigma^2, & \text{if } \mu = 1 \end{cases}$ for $n \geq 2, \mathbb{V}arX_0 = 0$ and $\mathbb{V}arX_1 = \sigma^2$.

Proof: (a) $\mathbb{E}X_n = \mathbb{E}(\mathbb{E}[X_n | X_{n-1}]) = \mathbb{E}\left(\mathbb{E}\left[\sum_{i=0}^{X_{n-1}} Z_i | X_{n-1}\right]\right) = \mathbb{E}(X_{n-1}EZ_i) =$

$\mu\mathbb{E}(X_{n-1}) \Rightarrow \mathbb{E}X_0 = 1,\ \mathbb{E}X_1 = \mu\mathbb{E}X_0 = \mu,\ \mathbb{E}X_2 = \mu\mathbb{E}X_1 = \mu^2, \dots, \mathbb{E}X_n = \mu^n$.

(b) $\mathbb{V}ar X_n = \mathbb{E}(\mathbb{V}ar[X_n|\, X_{n-1}]) + \mathbb{V}ar(\mathbb{E}[X_n|\, X_{n-1}]) = \mathbb{E}(X_{n-1}\sigma^2) + \mathbb{V}ar(\mu X_{n-1}) = \sigma^2\mu^{n-1} + \mu^2\mathbb{V}ar(X_{n-1}) \Rightarrow \mathbb{V}ar X_0 = 0,\ \mathbb{V}ar X_1 = \sigma^2,\ \mathbb{V}ar X_2 = \sigma^2\mu + \mu^2\sigma^2, \mathbb{V}ar X_3 = \sigma^2(\mu^2+\mu^3+\mu^4), \dots, \mathbb{V}ar X_n = \sigma^2(\mu^{n-1}+\mu^n+\dots+\mu^{2n-2})$
$= \begin{cases} \sigma^2\mu^{n-1}(\frac{1-\mu^n}{1-\mu}), & \text{if } \mu \neq 1 \\ n\sigma^2, & \text{if } \mu = 1. \end{cases}$ □

Definition. If $\mu = 1$, the process is called critical. If $\mu < 1$, the process is subcritical, and if $\mu > 1$, the process is supercritical.

Definition. $\pi_0 = \lim_{n\to\infty} \mathbb{P}(X_n = 0|\, X_0 = 1)$ is the probability that the population with one initial ancestor will eventually die out.

Proposition. If $\mu \leq 1$, $\pi_0 = 1$, that is, the critical and subcritical processes die out (intuitively, too few individuals are born every generation). If $\mu > 1$, $\pi_0 < 1$ and is the smallest positive solution of the equation $\pi_0 = \sum_{j=0}^{\infty} \pi_0^j P_j$.

Proof: $\mu^n = \mathbb{E}X_n = \sum_{j=1}^{\infty} j\mathbb{P}(X_n = j) \geq \sum_{j=1}^{\infty} 1 \cdot \mathbb{P}(X_n = j) = \mathbb{P}(X_n \geq 1) \Rightarrow$ as $n \to \infty$, $\mathbb{P}(X_n \geq 1) \leq \mu^n \to 0$ if $\mu < 1$. It can be shown that if $\mu = 1$, then $\pi_0 = 1$, and if $\mu > 1$, then $\pi_0 < 1$.

To show that in the supercritical case, π_0 satisfies the above equation, condition on the size of the first generation

$$\pi_0 = \mathbb{P}(\textit{population dies out}) = \sum_{j=0}^{\infty} \mathbb{P}(\textit{population dies out}|\, X_1 = j)P_j$$
$$= \{j \textit{ independent families die out}\} = \sum_{j=0}^{\infty} \pi_0^j P_j. \quad \square$$

Example. Assume $P_0 = 0.2,\ P_1 = 0.2,\ P_2 = 0.6$. What is the probability that this branching process dies out?

Solution: $\mu = 0.2 + 1.2 = 1.4 > 1$, so the process is supercritical and π_0 is the smallest positive solution of $\pi_0 = 0.2\pi_0^0 + 0.2\pi_0^1 + 0.6\pi_0^2 \Leftrightarrow \pi_0 = 0.2 + 0.2\pi_0 + 0.6\pi_0^2 \Leftrightarrow 0.6\pi_0^2 - 0.8\pi_0 + 0.2 = 0 \Rightarrow \pi_0 = 1/3$ (the other solution is 1).

Example. What can be said about the process, if $P_0 = 1/2$ and $P_1 = 1/2$, that is, an individual can either die or continue living with probabilities 1/2?

Solution: $\mu = 1/2 < 1$, so the process is subcritical and will eventually die out with probability one.

Example. The distribution of offspring is $Bi(3, p)$. (a) Give conditions for sure extinction, (b) for $p = 1/2$ find the extinction probability, (c) compute the average size and the variance of the nth generation if $p = 1/2$.

Solution: (a) the process dies out with probability one if $\mu = 3p \leq 1$, that is, if $p \leq 1/3$,

(b) π_0 is the smallest positive solution of $\pi_0 = 1/8 + 3/8\pi_0 + 3/8\pi_0^2 + 1/8\pi_0^3$. The solutions are $1, -2 \pm \sqrt{5}$, so $\pi_0 = \sqrt{5} - 2 \approx 0.24$,

(c) $\mathbb{E}X_n = \mu^n = (3/2)^n$, $\mathbb{V}ar X_n = (3/4)(3/2)^{n-1}(\frac{1-(3/2)^n}{1-3/2}) = (3/2)^n((3/2)^n - 1)$, if $n \geq 2$, $\mathbb{V}ar X_0 = 0$, $\mathbb{V}ar X_1 = 3/4$.

5.2. The Exponential Distribution.

Definition. A continuous r.v. X has an Exponential (λ) distribution if $f_X(x) = \lambda\, e^{-\lambda x},\ x \geq 0$.

Properties of the Exponential Distribution.(Prove !!!)
(i) $\mathbb{E}X = \frac{1}{\lambda}$, (ii) $\mathbb{V}ar X = \frac{1}{\lambda^2}$, (iii) $\varphi(t) = \frac{\lambda}{\lambda - t},\ t < \lambda$, (iv) $F_X(x) = 1 - e^{-\lambda x}$, $\mathbb{P}(X > x) = 1 - F_X(x) = e^{-\lambda x}$, (v) Let $X_1, \dots, X_n$ be i.i.d. Exponential(λ) r.v.'s. Denote $S_n = \sum_{i=1}^{n} X_i$. Then, $S_n \sim Gamma(n, \lambda)$, that is, $f_{S_n}(s) = \frac{\lambda^n\, s^{n-1}}{(n-1)!}\, e^{-\lambda\, s},\quad s > 0.$, (vi) *Memoryless Property*: if $X \sim Exp(\lambda)$, then $\mathbb{P}(X > x + y |\, X > y) = \mathbb{P}(X > x)$.

Remark. The Exponential distribution is the only continuous distribution that possesses the memoryless property.

5.3. The Poisson Process.

Definition. A stochastic process $\{N(t), t \geq 0\}$ is a counting process if $N(t)$ represents the total number of events occurring by time t.

Definition. A counting process has independent increments if the number of events that occur in disjoint time intervals are independent. For example, the number of events before time 5, $N(5)$, is independent of the number of events occurring between times 5 and 10, $N(10) - N(5)$.

Definition. A counting process has stationary increments if the distribution of the number of events that occur in any time interval depends only on the length of the interval. In other words, $N(t) - N(0)$ and $N(t + s) - N(s)$ have the same distribution that depends only on t.

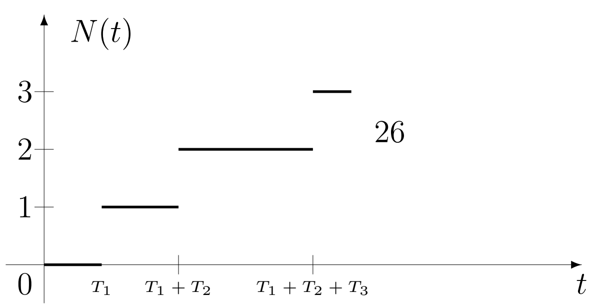
Definition. A counting process $\{N(t),\ t \geq 0\}$ is a Poisson process with rate $\lambda > 0$, if
(i) $N(0) = 0$,
(ii) the process has independent increments,
(iii) the process has stationary increments and $\mathbb{P}(N(t + s) - N(s) = n) = \mathbb{P}(N(t) = n) = \frac{(\lambda\, t)^n}{n!}\, e^{-\lambda\, t},\ n = 0, 1, \dots.$

Note. $\mathbb{E}N(t) = \mathbb{V}ar N(t) = \lambda\, t$.

Alternative Definition. A counting process $\{N(t),\ t \geq 0\}$ is a Poisson process with rate $\lambda > 0$, if
(i) $N(0) = 0$,
(ii) the process has independent stationary increments,
(iii) $\mathbb{P}(N(h) = 1) = \lambda\, h + o(h),\ h \to 0,\ \ \mathbb{P}(N(h) \geq 2) = o(h),\ h \to 0$.

Reminder. A function g is $o(h)$ if $\lim_{h \to 0} \frac{g(h)}{h} = 0$. For example, $h^2 = o(h)$ as $h \to 0$.

Definition. Let T_1 be the time of the first event, T_2 be the time between the first and the second events, etc. In general, let T_n be the time between the $(n-1)$st event and the nth event. Then, $\{T_n,\ n = 1, 2, \dots\}$ are the interarrival times.



Proposition. $T_n \overset{i.i.d.}{\sim} Exp(\lambda)$.
Proof: $\mathbb{P}(T_1 > t) = \mathbb{P}(N(t) = 0) = e^{-\lambda t} \Rightarrow T_1 \sim Exp(\lambda)$.
Further, $\mathbb{P}(T_2 > t) = \mathbb{E}[\mathbb{P}(T_2 > t | T_1)]$, $\mathbb{P}(T_2 > t | T_1 = s) = \mathbb{P}(N(s+t) - N(s) = 0 | T_1 = s) \overset{ind. inc.}{=} \mathbb{P}(N(s+t) - N(s) = 0) \overset{stat. inc.}{=} \mathbb{P}(N(t) = 0) = e^{-\lambda t} \Rightarrow \mathbb{P}(T_2 > t) = e^{-\lambda t} \Rightarrow T_2 \sim Exp(\lambda)$. Repeating the argument yields the general result. □

Remark. On intuitive level, stationary and independent increments of a Poisson process imply that the process renews itself at any given time, so the interarrival times should possess the memoryless property and thus be Exponentially distributed.
Definition. The time of the nth event, S_n, is called the waiting time until the nth event.
Proposition. $S_n \sim Gamma(n, \lambda)$.
Proof: $S_n = \sum_{i=1}^{n} T_i \Rightarrow S_n \sim Gamma(n, \lambda)$. □
Example. The following are Poisson processes:
(a) the number of people who enter a store or a bank,
(b) the number of cars that pass a certain intersection,
(c) the number of car accidents on a certain stretch of a freeway,
(d) the number of customers in a fast-food restaurant.
Proposition. Suppose that in a Poisson process $N(t)$ with rate λ an event can be either of type 1 with probability p or of type 2 with probability $1-p$. Let $N_1(t)$ and $N_2(t)$ denote the number of events of type 1 and 2 occurring by time t, respectively. Note that $N_1(t) + N_2(t) = N(t)$. Then, $N_1(t)$ and $N_2(t)$ are independent Poisson processes with rates $p\lambda$ and $(1-p)\lambda$.

Exercise 1. In *Bally's Total Fitness* treadmills are occupied exponential time with rate 5 minutes. Rob comes to the gym, but to his disappointment his favorite treadmill is taken. What is the probability that Rob has to wait at least 5 minutes to start his workout? (Note that Rob's favorite treadmill has been occupied for some unknown time).
Solution: Suppose the treadmill has been occupied for t minutes. Denote by T the random time the treadmill is occupied. $T \sim Exp(1/5)$. Then, $\mathbb{P}(T \geq t + 5 | T \geq t) \overset{memoryless}{=} \mathbb{P}(T \geq 5) = e^{-5/5} = e^{-1} \approx 0.37$.

Exercise 2. Buses arrive to *Harry's Place* carrying fifty hungry athletes each. The times that elapse between two consecutive arrivals are i.i.d. Exponential r.v.'s with mean 20 minutes. What is the total expected number of hungry athletes that are served in the restaurant within 2 hours after the opening?
Solution: Let $T_i \overset{i.i.d.}{\sim} Exp(1/20)$. Then, $50\, \mathbb{E}N(120) = 50\, \frac{120}{20} = 300$ hungry athletes.

Exercise 3. Which of the processes below can be modeled by a Poisson process?

(a) the number of migraine headaches that a boss of a large company experiences due to the stress of coping with employees, *yes*
(b) the number of migraine headaches that your instructor occasionally experiences when giving a lecture, *no*
(c) the number of shoppers in Walmart at any given time, *no*
(d) the number of unsatisfied customers who join the line to the customer service in Walmart. *yes*

EXERCISE 4. Subscribers sign up for a cable TV in accordance with a Poisson process $\{N(t),\, t \geq 0\}$ with rate 2 per minute.
(a) On average, how long does it take till the 100th subscriber signs up?
$T_i \overset{i.i.d.}{\sim} Exp(2) \Rightarrow \mathbb{E}S_{100} = \mathbb{E}\left[\sum_{i=1}^{100} T_i\right] = 100\mathbb{E}T_i = (100)(1/2) = 50$ minutes.
(b) Find $\mathbb{E}[S_{100}|\, N(30) = 50]$.
$\mathbb{E}[S_{100}|\, N(30) = 50] = 30 + \mathbb{E}[\text{time for 50 more to subscribe}] = 30 + \mathbb{E}S_{50} = 30 + 25 = 55$ minutes.
(c) Find $\mathbb{P}(N(4) = 2|\, N(1) = 0)$.
$\mathbb{P}(N(4) = 2|\, N(1) = 0) = \mathbb{P}(N(4) - N(1) = 2|\, N(1) = 0) \overset{ind.\,inc.}{=} \mathbb{P}(N(4) - N(1) = 2) \overset{stat.\,inc.}{=} \mathbb{P}(N(3) = 2) = \frac{6^2}{2!}\, e^{-6} \approx 0.04.$
(d) Compute $\mathbb{E}[N(9) - N(7)|\, N(7)]$.
$\mathbb{E}[N(9) - N(7)|\, N(7)] \overset{ind.\,inc.}{=} \mathbb{E}[N(9) - N(7)]$

$$\begin{cases} = \mathbb{E}N(9) - \mathbb{E}N(7) = 18 - 14 = 4 \\ \overset{stat.\,inc.}{=} \mathbb{E}N(2) = 4. \end{cases}$$

EXERCISE 5. Customers arriving at a department store form a Poisson process with rate 2 per 5 minutes. A customer will buy a DVD player with probability 0.3. What is the probability that during the next 15 minutes, at least two DVD players will be sold and at least two customers will walk out of the store without a DVD player?
Solution: Let $N_1(t)$ and $N_2(t)$ be the number of customers who will and who will not buy a DVD player by time t, respectively. Then, $N_1(t)$ and $N_2(t)$ are independent Poisson processes with rates (2)(0.3)=0.6 and (2)(0.7)=1.4 per 5 minutes. Thus, $\mathbb{P}(N_1(3) \geq 2,\, N_2(3) \geq 2) = \mathbb{P}(N_1(3) \geq 2)\mathbb{P}(N_2(3) \geq 2) = (1 - e^{-0.6} - 0.6e^{-0.6})(1 - e^{-1.4} - 1.4e^{-1.4}) = 0.0498.$

5.4.1 Nonhomogeneous Poisson Process.

Definition. A counting process $\{N(t),\, t \geq 0\}$ is called a nonhomogeneous (or nonstationary or time-dependent) Poisson process with intensity function $\lambda(t)$ if (i) $N(0) = 0$, (ii) the process has independent increments,
(iii) $\mathbb{P}(N(t+h) - N(t) = 1) = \lambda(t)\, h + o(h)$, $\mathbb{P}(N(t+h) - N(t) \geq 2) = o(h)$.
Proposition. Let $m(t) = \int_0^t \lambda(x)\, dx$. Then,

$$\mathbb{P}(N(t+s) - N(s) = n) = \frac{[m(t+s) - m(s)]^n}{n!}\, e^{-[m(t+s) - m(s)]},\ n \geq 0.$$

Definition. The function $m(t)$ is called the mean value function.
Example. Arrivals of 911 calls to Precinct 77 can be modelled as a time-dependent Poisson process with the per-hour intensity rate

$$\lambda(t) = \begin{cases} 11, & \text{if } 10pm < t < 6am \\ 7, & \text{if } 6am < t < 11am \text{ or } 3pm < t < 10pm \\ 4, & \text{if } 11am < t < 3pm. \end{cases}$$

5.4.2. Compound Poisson Process.

Definition. A stochastic process $\{X(t), t \geq 0\}$ is called a compound Poisson process if $X(t) = \sum_{i=1}^{N(t)} Y_i$, where $\{N(t)\}$ is a Poisson process with rate λ, and Y_i's are i.i.d. r.v.'s which are also independent of $\{N(t)\}$.
Proposition. $\mathbb{E}X(t) = \lambda\, t\, \mathbb{E}Y_1$, $\mathbb{V}arX(t) = \lambda\, t\, \mathbb{E}Y_1^2$.
Proof: $\mathbb{E}X(t) = \mathbb{E}(\mathbb{E}[X(t)|\, N(t)]) = \mathbb{E}(N(t)\mathbb{E}Y_1) = \lambda\, t\, \mathbb{E}Y_1$, $\mathbb{V}arX(t) = \mathbb{E}(\mathbb{V}ar[X(t)|\, N(t)]) + \mathbb{V}ar(\mathbb{E}[X(t)|\, N(t)]) = \mathbb{E}(N(t)\mathbb{V}arY_1) + \mathbb{V}ar(N(t)\mathbb{E}Y_1) = \lambda\, t\, (\mathbb{E}Y_1^2 - (\mathbb{E}Y_1)^2) + \lambda\, t\, (\mathbb{E}Y_1)^2 = \lambda\, t\, \mathbb{E}Y_1^2$. $\square$
Example. Buses arrive to *Harry's Place* according to a Poisson process $\{N(t),\, t \geq 0\}$. Each bus contains a random number Y_i of hungry athletes. The total number of hungry athletes that are served in the restaurant by time t is a compound Poisson process $X(t) = \sum_{i=1}^{N(t)} Y_i$.

5.4.3. Conditional Poisson Process.

Definition. Denote by L a positive r.v. A counting process $\{N(t),\, t \geq 0\}$ is called a conditional (or mixed) Poisson process if, conditioned on $L = \lambda$, $\{N(t)\}$ is a Poisson process with rate λ.
Remark. If $L \sim g$, then the marginal distribution of a conditional Poisson process $\{N(t), t \geq 0\}$ is $\mathbb{P}(N(t+s) - N(s) = n) = \int_0^\infty \mathbb{P}(N(t+s) - N(s) = n|\, L = \lambda) g(\lambda)\, d\lambda = \int_0^\infty \frac{(\lambda\, t)^n}{n!}\, e^{-\lambda\, t} g(\lambda)\, d\lambda$. This shows that a mixed Poisson process has stationary increments. It does not, in general, have independent increments since knowledge of how many events occur in a particular interval gives some information about what L might be, and so affects the distribution of the number of events on other intervals.
Proposition. $\mathbb{E}N(t) = t\, \mathbb{E}L$, $\mathbb{V}arN(t) = t\, \mathbb{E}L + t^2\, \mathbb{V}arL$, $f_{L|\, N(t)}(\lambda|\, n) = c\lambda^n g(\lambda)\, e^{-\lambda\, t}$ where c is the normalizing constant.
Proof: prove it yourself or see page 328.
Example. It has been noticed that a certain intersection has an average of 5 accidents per day if it rains, and 2 accidents per day if it is sunny. Suppose it rains tomorrow with probability 0.8. Then, $\mathbb{P}(L = 5) = 0.8 = 1 - \mathbb{P}(L = 2)$. The number of accidents occurring tomorrow is a mixed Poisson process with the marginal distribution $\mathbb{P}(N(t) = n) = \frac{(5t)^n}{n!}\, e^{-5t}(0.8) + \frac{(2t)^n}{n!}\, e^{-2t}(0.2)$.

Exercise 1. A doctor has scheduled two appointments, one at 1pm and the other at 1:30pm. The amounts of time that appointments last are in-

dependent exponential r.v.'s with mean 30 minutes. Assuming that both patients are on time, find the expected amount of time that the 1:30 appointment spends at the doctor's office.
Solution: Condition on whether the 1pm appointment is still with the doctor at 1:30, and use the fact that if the patient is still there then the remaining time is exponential with mean 30 minutes by the memoryless property. $\mathbb{E}[\text{time spent in office}] = (30)(1-e^{-30/30})+(30+30)e^{-30/30} = 30+30e^{-1} \approx 41$ minutes.

Exercise 2. People walk into a casino in Las Vegas according to a Poisson process with rate 50 per hour. Ten percent of them will not gamble at all, others will lose independently a random number of dollars which we assume has a Uniform (0, \$1000) distribution. What is the casino's expected gain during a 12-hour period? Compute the standard deviation of this gain.
Solution: The number of gamblers is Poisson process with rate $\lambda = (50)(0.9) =$ 45 per hour. The total amount of money the gamblers lose during $t = 12$ hours is a compound Poisson r.v. with mean $\lambda\, t\, \mathbb{E}Y_1 = (45)(12)(500) =$ $\$270,000$. The variance is $\lambda\, t\, \mathbb{E}Y_1^2 = (45)(12)(1000/12+(500)^2) = \$135,045,000$, hence the standard deviation equals $\sqrt{135,045,000} \approx \$11,621$.

Exercise 3. A small cafeteria in a big university serves lunch between 11:30am and 1:00pm. From 11:30am until 12noon students arrive, on average, at a linearly increasing rate that starts with the initial rate of 200 per hour and reaches the maximum of 500 per hour. After 12noon the rate of students is constant. Describe an appropriate probability model for this process. How many students, on average, are served lunch every day?
Solution: The appropriate model is a nonhomogeneous Poisson process with the intensity function

$$\lambda(t) = \begin{cases} 200 + 600t, & \text{if } 0 < t < 0.5, \\ 500, & \text{if } 0.5 < t < 1.5. \end{cases}$$

The mean value function is $m(t) = \int_0^t \lambda(x)\, dx = \begin{cases} 200\, t + 300\, t^2, & \text{if } 0 < t < 0.5, \\ 500\, t - 75, & \text{if } 0.5 < t < 1.5. \end{cases}$
The average number students served per day is $m(1.5) = 675$.

Exercise 4. An insurance company feels that each of its policyholders has a rating value and that a policyholder having rating value λ will make claims at times distributed according to a Poisson process with rate λ (in years). The firm also believes that rating values are distributed uniformly over (0,1).
(a) What is the average number of claims per person the company has to deal with every five years?
(b) What is the variance of this number of claims?
(c) What is the probability that a policyholder's rating doesn't exceed 1/4, given that he has made one claim in three years?
Solution: Let $\{N(t)\}$ denote the number of claims per person up to time t.

It is a mixed Poisson process with rate $\lambda \sim Unif(0,1)$. Hence,
(a) $\mathbb{E}N(5) = t\,\mathbb{E}\lambda = (5)(1/2) = 2.5$ claims.
(b) $\mathbb{V}arN(5) = t\mathbb{E}\lambda + t^2\mathbb{V}ar\lambda = (5)(1/2) + (25)(1/12) = 55/12$.
(c) $\mathbb{P}(\lambda \leq 1/4|\, N(3) = 1) = \frac{\int_0^{1/4} \lambda e^{-3\lambda}\, d\lambda}{\int_0^1 \lambda e^{-3\lambda}\, d\lambda} = \frac{1-(7/4)e^{-3/4}}{1-4e^{-3}} \approx 0.22.$

6.2. Continuous-Time Markov Chains.

Definition. The process $\{X(t),\, t \geq 0\}$ is a continuous-time Markov chain if for all $s, t \geq 0$, and any nonnegative integers $i, j, x(u),\ 0 \leq u < s$,

$$\mathbb{P}(X(t+s) = j \,|\, X(s) = i,\, X(u) = x(u),\ 0 \leq u < s) = \mathbb{P}(X(t+s) = j \,|\, X(s) = i).$$

Definition. If $\mathbb{P}(X(t+s) = j \,|\, X(s) = i)$ is independent of s, then the continuous-time Markov chain is said to have stationary transition probabilities.

Proposition. Let $X(t)$ be a continuous-time Markov chain with stationary transition probabilities. Denote by T_i the amount of time that the chain stays in state i before making a transition into a different state. Then, T_i has an exponential distribution.
Proof: $\mathbb{P}(T_i > t+s \,|\, T_i > s) = \mathbb{P}(X(t+s) = i \,|\, X(s) = i) \overset{stat.}{=} \mathbb{P}(X(t) = i) = \mathbb{P}(T_i > t)$. $\square$

Alternative Definition. A continuous-time Markov chain is a stochastic process such that
(1) the amount of time T_i it spends in a state i is exponential with mean $1/v_i$;
(2) when the chain leaves state i, it enters state j with probability P_{ij} such that $P_{ii} = 0$, $\sum_j P_{ij} = 1$;
(3) T_i and the next state visited are independent r.v.'s.

Remark. A continuous-time Markov chain moves from state to state as a discrete-time Markov chain, but the time spent in each state is exponentially distributed.

6.3. Birth and Death Process.

Definition. Consider a Markov process which states are represented by the number of particles. If there are n particles, the time until a new particle is born is exponentially distributed with rate λ_n, and is independent of the time until a particle dies which is itself exponentially distributed with rate μ_n. Such a process is called a birth and death process.
Proposition. (1) The state space of a birth and death process is $\{0, 1, \dots\}$.
(2) The transition probabilities are $P_{01} = 1,\ P_{i,i+1} = \frac{\lambda_i}{\lambda_i+\mu_i},\ P_{i,i-1} = \frac{\mu_i}{\lambda_i+\mu_i}$ and $P_{ij} = 0$ if $j \neq i+1$ or $i-1$ for $i > 0$.
(3) The transition rates are $v_0 = \lambda_0,\ v_i = \lambda_i + \mu_i$ for $i > 0$.

Example. A Poisson process is a birth and death process with $\mu_n = 0$ and $\lambda_n = \lambda$.
Example *(The Yule process).* Each particle independently of others gives birth to one particle with rate λ and never dies. Then, the process is a pure birth process with $\lambda_n = n\lambda$.

Proposition. Let T_i denote the time it take for the process to go from state i to state $i+1$. Then, $\mathbb{E}T_0 = \frac{1}{\lambda_0}$, and $\mathbb{E}T_i = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i}\mathbb{E}T_{i-1},\ i \geq 1$.
Proof. Let

$$I_i = \begin{cases} 1, & \text{if the first transition from } i \text{ is to } i+1 \\ 0, & \text{if the first transition from } i \text{ is to } i-1. \end{cases}$$

The time until the first transition is $\text{Exp}(\lambda_i + \mu_i)$. Therefore, $\mathbb{E}[T_i \,|\, I_i = 1] = \frac{1}{\lambda_i + \mu_i}$ and $\mathbb{E}[T_i \,|\, I_i = 0] = \frac{1}{\lambda_i + \mu_i} + \mathbb{E}[T_{i-1}] + \mathbb{E}[T_i]$. Since the probability that the first transition is a birth is $P_{i,\,i+1} = \frac{\lambda_i}{\lambda_i + \mu_i}$, we get

$$\mathbb{E}[T_i] = \mathbb{E}[T_i \,|\, I_i = 1]\, P_{i,\,i+1} + \mathbb{E}[T_i \,|\, I_i = 0]\, P_{i,\,i-1} = \frac{1}{\lambda_i + \mu_i} \cdot \frac{\lambda_i}{\lambda_i + \mu_i}$$

$$+\left(\frac{1}{\lambda_i + \mu_i} + \mathbb{E}[T_{i-1}] + \mathbb{E}[T_i]\right) \cdot \frac{\mu_i}{\lambda_i + \mu_i} = \frac{1}{\lambda_i + \mu_i} + \frac{\mu_i}{\lambda_i + \mu_i}(\mathbb{E}[T_{i-1}] + \mathbb{E}[T_i])$$

$$\Longrightarrow \mathbb{E}[T_i] = \frac{1}{\lambda_i} + \frac{\mu_i}{\lambda_i}\mathbb{E}[T_{i-1}]. \quad \square$$

Remark. The expected time to go from state i to state j where $j > i$ is $\mathbb{E}T_i + \mathbb{E}T_{i+1} + \cdots + \mathbb{E}T_{j-1}$.

Example. In a Poisson process, what is the expected time to go from state 3 to state 4? To go from state 3 to state 7?
Solution. In a Poisson process $\lambda_n = \lambda$ and $\mu_n = 0$. Therefore, $\mathbb{E}[T_3] = \frac{1}{\lambda}$. Also, $\mathbb{E}[T_3] + \mathbb{E}[T_4] + \mathbb{E}[T_5] + \mathbb{E}[T_6] = \frac{4}{\lambda}$.

7.1. Introduction to Renewal Theory.

Definition. A counting process for which the times between successive events are independent and identically distributed with an arbitrary distribution is called a renewal process.
Example. The Poisson process is a renewal process with i.i.d. exponential interarrival times.
Notation. Denote $\{N(t),\, t \geq 0\}$ a renewal process. Let X_n be the time between the $(n-1)$st and the nth event of this process. Assume $X_n \overset{i.i.d.}{\sim} F$. Define $S_n = \sum_{i=1}^{n} X_i,\ n \geq 1,\ S_0 = 0$, the time of the nth renewal. Put $\mu = \mathbb{E}X_n > 0$.
Proposition. $N(t)$, the number of renewals by time t, cannot be infinite if t is finite.
Proof: $N(t) = \max\{n : S_n \leq t\}$. By the LLN, $\frac{S_n}{n} \to \mu > 0$ as $n \to \infty$, therefore, $S_n \to \infty$ as $n \to \infty$. Hence, if t is finite, $S_n \leq t$ for n finite. $\square$

Proposition. $\mathbb{P}(N(\infty) = \infty) = 1$.

Proof: $\mathbb{P}(N(\infty) < \infty) = \mathbb{P}(X_n = \infty\ \exists n) = \mathbb{P}\left(\cup_{n=1}^{\infty} \{X_n = \infty\}\right) \overset{Boole's \neq}{\leq} \sum_{n=1}^{\infty} \mathbb{P}(X_n = \infty) = 0.$ □

7.2. Distribution of $N(t)$.

Proposition. $N(t) \geq n$ iff $S_n \leq t$.

Proposition. $\mathbb{P}(N(t) = n) = \mathbb{P}(N(t) \geq n) - \mathbb{P}(N(t) \geq n+1)$
$= \mathbb{P}(S_n \leq t) - \mathbb{P}(S_{n+1} \leq t)$.
Example. Suppose $X_n \overset{i.i.d.}{\sim} Exp(\lambda)$. Therefore, $S_n \sim Gamma(n, \lambda)$. The distribution of $N(t)$ is $\mathbb{P}(N(t) = n) = \mathbb{P}(S_n \leq t) - \mathbb{P}(S_{n+1} \leq t)$
$= \int_0^t \left[\frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} - \frac{\lambda^{n+1}}{n!} x^n e^{-\lambda x}\right] dx = \lambda^n (\frac{x^n}{n!} e^{-\lambda x})\big|_0^t = \frac{(\lambda t)^n}{n!} e^{-\lambda t}$. Thus, $N(t)$ is a Poisson process.

Definition. The mean-value (or the renewal) function of $N(t)$ is $m(t) = \mathbb{E}[N(t)]$.
Proposition. $m(t) = \mathbb{E}[N(t)] = \sum_{n=1}^{\infty} \mathbb{P}(N(t) \geq n) = \sum_{n=1}^{\infty} \mathbb{P}(S_n \leq t)$.
Example. Let $N(t)$ denote a Poisson process. Then,
$m(t) = \sum_{n=1}^{\infty} \mathbb{P}(S_n \leq t) = \sum_{n=1}^{\infty} \int_0^t \frac{\lambda^n}{(n-1)!} x^{n-1} e^{-\lambda x} dx$
$= \lambda \int_0^t e^{-\lambda x} \sum_{n=1}^{\infty} \frac{(\lambda x)^{n-1}}{(n-1)!} dx = \lambda \int_0^t e^{-\lambda x} e^{\lambda x} dx = \lambda t.$
Remark. The mean-value function $m(t)$ uniquely determines the renewal process. For example, a Poisson process is the only process with a linear mean-value function.

Theorem. *(The Renewal Equation)* The interarrival times of a renewal process $X_n \overset{i.i.d.}{\sim} F$ with density f. Then, $m(t) = F(t) + \int_0^t m(t-x) f(x)\, dx$.
Proof: Conditioning on the time of the first renewal, we get $m(t) = \mathbb{E}[N(t)] = \int_0^t \mathbb{E}[N(t) \mid X_1 = x] f(x)\, dx = \int_0^t (1 + \mathbb{E}[N(t-x)]) f(x)\, dx$
$= F(t) + \int_0^t m(t-x) f(x)\, dx.$ □
Example. Using the renewal equation, find the mean-value function of a Poisson process.
Solution: $F(t) = 1 - e^{-\lambda t}$. Therefore, $m(t)$ satisfies the integral equation $m(t) = 1 - e^{-\lambda t} + \int_0^t m(t-x) \lambda e^{-\lambda x} dx$. Substituting $u = t - x$, we have $m(t) = 1 - e^{-\lambda t} + \int_0^t m(u) \lambda e^{-\lambda (t-u)} du$, or equivalently, $m(t) = 1 - e^{-\lambda t} + \lambda e^{-\lambda t} \int_0^t m(u) e^{\lambda u} du$. This can be rewritten as $\lambda \int_0^t m(u) e^{\lambda u} du = 1 - \frac{1-m(t)}{e^{-\lambda t}}$. Taking the derivatives of both sides with respect to t, obtain $\lambda m(t) e^{\lambda t} = -\frac{-e^{-\lambda t} m'(t) + (1-m(t))\lambda e^{-\lambda t}}{e^{-2\lambda t}}$ which simplifies to $m'(t) = \lambda$ or $m(t) = \lambda t + c$. Since $m(0) = 0$, $m(t) = \lambda t$.

7.3. Limit Theorems in Renewal Theory.

We know that $N(t) \to \infty$ as $t \to \infty$, but what is the rate of convergence?
Proposition. With probability 1, $\frac{N(t)}{t} \to \frac{1}{\mu}$ as $t \to \infty$. It means that $N(t) \sim \frac{1}{\mu} t$ for large t.
Proof: The random variable $S_{N(t)}$ represents the time of the last renewal prior or at time t. Therefore, $S_{N(t)} \le t < S_{N(t)+1}$. Or $\frac{S_{N(t)}}{N(t)} \le \frac{t}{N(t)} < (\frac{S_{N(t)+1}}{N(t)+1})(\frac{N(t)+1}{N(t)})$. Since $\frac{S_{N(t)}}{N(t)} = \frac{\sum_{i=1}^{N(t)} X_i}{N(t)}$ is the average of $N(t)$ i.i.d. r.v.'s, by the LLN, $\frac{S_{N(t)}}{N(t)} \to \mathbb{E}[X_1] = \mu$ as $N(t) \to \infty$. Since $N(t) \to \infty$ as $t \to \infty$, we have $\frac{S_{N(t)}}{N(t)} \to \mu$ as $t \to \infty$. Likewise, $\frac{S_{N(t)+1}}{N(t)+1} \to \mu$ as $t \to \infty$. Also, $\frac{N(t)+1}{N(t)} \to 1$ as $t \to \infty$. The result follows. $\square$.

Theorem (*Elementary Renewal Theorem*).
(1) $\frac{m(t)}{t} \to \frac{1}{\mu}$ as $t \to \infty$,
(2) $\frac{\mathbb{V}ar(N(t))}{t} \to \frac{\sigma^2}{\mu^3}$ as $t \to \infty$ where $\sigma^2 = \mathbb{V}ar[X_1]$.

The Central Limit Theorem for Renewal Processes. As $t \to \infty$,

$$\frac{N(t) - t/\mu}{\sqrt{t\sigma^2/\mu^3}} \to N(0,1).$$

Example. The lifetime (in hours) of an electrical bulb is Uniform(600,700). If bulbs burn out independently, what is the rate at which you have to replace the bulbs? What is the probability that you have to replace at most 3 bulbs in 12 weeks?
Solution: $\mu = 650$, $\sigma^2 = 10000/12$. Let $N(t)$ denote the number of bulbs that burn out by time t. Then, $N(t)$ is a renewal process, and by Proposition, $\frac{N(t)}{t} \to \frac{1}{\mu} = \frac{1}{650}$, that is, in the long run, you have to replace one bulb every 650 hours.
By the CLT, $\mathbb{P}(N(2016) \le 3) \approx \mathbb{P}\left(Z \le \frac{3-2016/650}{\sqrt{(2016)(10000/12)/(650)^3}}\right)$
$= \mathbb{P}(Z \le -1.30) = 0.0968$.

Proposition. $\mathbb{E}[S_{N(t)+1}] = \mu[m(t)+1]$.

Remark. Let $Y(t)$ denote the time from t until the next renewal. $Y(t)$ is called the excess life at t. Then, $S_{N(t)+1}$, the time of the first renewal after t, can be written as $S_{N(t)+1} = t + Y(t)$. Therefore, the expected excess life at t is $\mathbb{E}[Y(t)] = \mu[m(t)+1] - t$.

Example. For a Poisson process with $\lambda = 3$/minute determine the expected excess life at 2 minutes.
Solution: $\mathbb{E}[Y(t)] = \frac{1}{\lambda}[\lambda t + 1] - t = \frac{1}{\lambda} = 1/3$.

8.1 — 8.3 Queueing Theory.

Definition. Suppose that customers arrive at a single-server service station (bank, gas station, shop, etc.) in accordance with a Poisson process with rate λ. Upon arrival, each customer goes into service if the server is empty, and, if not, the customer joins the queue. When the server is done with a customer, the customer leaves the system, and the next customer in line enters service. The service times are assumed to be independent exponential random variables with mean $1/\mu$. This model is called M/M/1 queue. The first M means that the arrival process is Markovian, the second M refers to the fact that the service process is Markovian too, and the 1 means that there is a single server.

Definition. Let $X(t)$ denote the number of customers in the system at time t. Define $P_n = \lim_{t\to\infty} \mathbb{P}(X(t) = n),\ n \geq 0$, that is, P_n is the limiting probability that there will be exactly n customers in the system. P_n is called the steady-state probability of exactly n customers in the system.

How to find the steady-state probabilities for M/M/1 system?

Rate-equality Principle. *The rate at which the process enters state i equals the rate at which it leaves state i.*

From the rate-equality principle, we get $\lambda\, P_0 = \mu\, P_1,\ \ (\lambda+\mu)\, P_1 = \lambda\, P_0 + \mu\, P_2$, etc. In general, we have a set of balance equations (they balance the rate at which the process enters each state with the rate at which it leaves that state): $\lambda\, P_0 = \mu\, P_1$ and for $n \geq 1,\ (\lambda+\mu)\, P_n = \lambda\, P_{n-1} + \mu\, P_{n+1}$.

To solve these equations, write $P_1 = \frac{\lambda}{\mu}\, P_0,\ P_{n+1} = \frac{\lambda}{\mu}\, P_n + (P_n - \frac{\lambda}{\mu} P_{n-1})$. Therefore, $P_2 = \frac{\lambda}{\mu}\, P_1 + (P_1 - \frac{\lambda}{\mu} P_0) = \frac{\lambda}{\mu}\, P_1 = (\frac{\lambda}{\mu})^2\, P_0,\ \ P_3 = \frac{\lambda}{\mu}\, P_2 + (P_2 - \frac{\lambda}{\mu} P_1) = \frac{\lambda}{\mu}\, P_2 = (\frac{\lambda}{\mu})^3\, P_0, \dots, P_{n+1} = (\frac{\lambda}{\mu})^{n+1}\, P_0$. To find P_0 recall that the steady-state probabilities must sum up to one. Thus,

$$1 = \sum_{n=0}^{\infty} P_n = \sum_{n=0}^{\infty} (\frac{\lambda}{\mu})^n\, P_0 = \frac{P_0}{1 - \lambda/\mu}.$$

Hence, $P_0 = 1 - \frac{\lambda}{\mu},\ P_n = (\frac{\lambda}{\mu})^n\, (1 - \frac{\lambda}{\mu}),\ n \geq 1,\ \lambda < \mu$.

Besides the steady-states probabilities, a queueing model is characterized by (1) L = the average number of customers in the system; (2) L_Q = the average number of customers waiting in queue; (3) W = the average amount of time a customer spends in the system; and (4) W_Q = the average amount of time a customer spends waiting in queue.

How to find these quantities for M/M/1 model?

$$L=\sum_{n=0}^{\infty} n\,P_n=\sum_{n=0}^{\infty} n\,(\frac{\lambda}{\mu})^n\,(1-\frac{\lambda}{\mu})=\left\{\sum_{n=0}^{\infty} n\,a^n=\frac{a}{(1-a)^2}\right\}$$
$$=(1-\frac{\lambda}{\mu})\frac{\lambda/\mu}{(1-\lambda/\mu)^2}=\frac{\lambda}{\mu-\lambda}.$$

Little's Formula. $L=\lambda\,W,\;\; L_Q=\lambda\,W_Q$.

Using Little's formula, we obtain that $W=L/\lambda=\frac{1}{\mu-\lambda}$, $W_Q=W-\mathbb{E}[\text{service time}]=W-\frac{1}{\mu}=\frac{\lambda}{\mu(\mu-\lambda)}$, and $L_Q=\lambda\,W_Q=\frac{\lambda^2}{\mu(\mu-\lambda)}$.

Example. Customers arrive to a bank with a single teller with rate 3 per 5 minutes, and depart at rate 7 per 10 minutes. What are the parameters of the model?

Solution: $\lambda=0.6$/minute, $\mu=0.7$/minute, $P_0=1/7$, $P_n=(1/7)(6/7)^n$, $L=6$ customers, $W=10$ minutes, $W_Q=60/7\approx 8.57$ minutes, $L_Q=5.14$ customers.

Proposition. The amount of time a customer spends in the system is an exponential random variable with rate $\mu-\lambda$.

Proof: Let W^* denote the amount of time a customer spends in the system. Suppose there are already n customers in the system when he arrives. The probability of this event is P_n, the steady-state probability. If $n=0$, then W^* is his service time. If $n>0$, he has to wait $Exp(\mu)$ time until one person completes the service (due to lack of memory of the exponential distribution) and additional n (including himself) Exponential (μ) random times for the others in line to go through the service. Thus, $W^*\sim Gamma(n+1,\mu\,t)$. Therefore,

$$\mathbb{P}(W^*\le w)=\sum_{n=0}^{\infty}\mathbb{P}(W^*\le w\,|\,n\text{ customers})\,P_n$$

$$=\sum_{n=0}^{\infty}\int_0^w\frac{(\mu\,t)^n}{n!}\,\mu\,e^{-\mu\,t}\,dt\,(\frac{\lambda}{\mu})^n(1-\frac{\lambda}{\mu})$$

$$\int_0^w(\mu-\lambda)\,e^{-\mu\,t}\sum_{n=0}^{\infty}\frac{(\lambda\,t)^n}{n!}\,dt=\int_0^w(\mu-\lambda)\,e^{-\mu\,t}\,e^{\lambda\,t}\,dt$$

$$=\int_0^w(\mu-\lambda)\,e^{-(\mu-\lambda)\,t}\,dt=1-e^{-(\mu-\lambda)\,w}.\quad\square$$

Remark. $\mathbb{E}[W^*]=\frac{1}{\mu-\lambda}=W$.

Example (continued...). What proportion of customers spend more than five minutes in the system?

Solution. $\mathbb{P}(W^*>5)=e^{-(0.1)(5)}\approx 0.61$. Therefore, about 61%.

8.3.2. Model M/M/1 with Finite Capacity.

Definition. Consider M/M/1 model. Suppose there cannot be more than N customers in the system at the same time. This model is called the finite capacity M/M/1 model.

To find the steady-state probabilities for the model, write the balance equations

$$\begin{cases} \lambda\, P_0 = \mu\, P_1 \\ (\lambda+\mu)\, P_n = \lambda\, P_{n-1} + \mu\, P_{n+1}, \quad 0 < n < N \\ \mu\, P_N = \lambda\, P_{N-1} \end{cases}$$

Solving the equations, we get $P_n = (\frac{\lambda}{\mu})^n\, P_0,\ 0 < n \le N$. To find P_0, write $1 = \sum_{n=0}^{N} P_n = P_0 \sum_{n=0}^{N} (\frac{\lambda}{\mu})^n = P_0\, [\frac{1-(\lambda/\mu)^{N+1}}{1-\lambda/\mu}]$. Thus, $P_0 = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{N+1}}$, and, therefore,

$$P_n = \frac{(\lambda/\mu)^n\,(1-\lambda/\mu)}{1-(\lambda/\mu)^{N+1}}, \quad 0 \le n \le N.$$

The average number of customers in the system is

$$L = \sum_{n=0}^{N} n\, P_n = \frac{1-\lambda/\mu}{1-(\lambda/\mu)^{N+1}} \sum_{n=0}^{N} n\, (\frac{\lambda}{\mu})^n = \frac{\lambda[1+N(\lambda/\mu)^{N+1} - (N+1)(\lambda/\mu)^N]}{(\mu-\lambda)(1-(\lambda/\mu)^{N+1})}.$$

To derive the formula for W, the average amount of time a customer spends in the system, we have to consider two cases: (1) a customer arrives to find the system full and immediately departs ($W = L/\lambda$); (2) a customer enters the system ($W = \frac{L}{\lambda(1-P_N)}$).

Example. Suppose that it costs $\$\,5c$ to provide service at a rate $\mu = 5$ customers per hour. Suppose also that we make a gross profit of $\$d = 10$ for each customer served. If the system has capacity 10 people and the arrival rate to the single server is $\lambda = 4$ people per hour, find the value of c that makes the system profitable.

Solution: Customers enter the system at rate $\lambda(1-P_N)$. Thus, our profit per hour is

$$\lambda(1-P_N)\, d - c\mu = \lambda\, d \left[1 - \frac{(\lambda/\mu)^N(1-\lambda/\mu)}{1-(\lambda/\mu)^{N+1}}\right] - c\mu$$

$$= \frac{\lambda\, d\,[1-(\lambda/\mu)^N]}{1-(\lambda/\mu)^{N+1}} - c\mu = \frac{(4)(10)[1-(4/5)^{10}]}{1-(4/5)^{11}} - 5c = 39.06 - 5c > 0 \implies c < \$7.8.$$

Multiserver Exponential Queueing System (Model M/M/s).

Definition. Suppose there are s servers available, each serving at rate μ.

The departure rate is

$$\mu_n = \begin{cases} n\mu, & \text{if } 1 \le n \le s \\ s\mu, & \text{if } n \ge s. \end{cases}$$

The arrival rate is λ. To find the steady-state probabilities, write the balance equations

$$\begin{cases} \lambda\, P_0 = \mu\, P_1 \\ (\lambda+\mu)\, P_1 = 2\mu\, P_2 + \lambda\, P_0 \\ (\lambda+2\mu)\, P_2 = 3\mu\, P_3 + \lambda\, P_1 \\ \dots \\ (\lambda+k\mu)\, P_k = (k+1)\, P_{k+1} + \lambda\, P_{k-1}, \quad 0 < k < s \\ (\lambda+s\mu)\, P_k = s\mu\, P_{k+1} + \lambda\, P_{k-1}, \quad k \ge s. \end{cases}$$

Solve the equations to obtain $P_k = \frac{(\lambda/\mu)^k}{k!}\, P_0, \quad 0 \le k \le s$, and $P_k = (\frac{\lambda}{s\mu})^{k-s}\, \frac{(\lambda/\mu)^s}{s!}\, P_0, \;\; k \ge s$.
Since the sum of these probabilities equals one,

$$P_0 = \left[\sum_{k=0}^{s-1} \frac{(\lambda/\mu)^k}{k!} + (\frac{s\mu}{s\mu-\lambda})\frac{(\lambda/\mu)^s}{s!}\right]^{-1}.$$

Example. For M/M/2 system with rates $\mu = 10$ and $\lambda = 6$, find the probability that the third customer has to wait.

Solution: $P_3 = (0.3)(0.18)\left[\sum_{k=0}^{1} \frac{(0.6)^k}{k!} + (\frac{20}{20-6})\frac{(0.6)^2}{2!}\right]^{-1} \approx (0.3)(0.18)(0.5385) = 0.029$.

10.1. Brownian Motion.

Definition. A stochastic process $\{B(t),\; t \ge 0\}$ is a <u>Brownian motion</u> if
(1) $B(0) = 0$,
(2) the process has independent, stationary increments,
(3) $B(t) \sim N(0, \sigma^2 t)$, $t > 0$.

PICTURE

Remark. English botanist Robert Brown in 1827 observed the movement of dust particles in liquid. Wiener (1918) described the model mathematically.

Definition. If $\sigma = 1$, the process is called a <u>standard</u> Brownian motion.
We will consider only standard Brownian motion since any other BM can be obtained from the standard BM by rescaling (multiplying by σ).

The joint density of $B(t_1), B(t_2), \dots, B(t_n)$, $0 < t_1 < t_2 < \cdots < t_n$, is given by

$$f(x_1, x_2, \dots, x_n) = f_{B(t_1)}(x_1)\, f_{B(t_2-t_1)}(x_2 - x_1) \dots f_{B(t_n-t_{n-1})}(x_n - x_{n-1}).$$

Exercise. Compute $\mathbb{P}(B(2) \le 2 \,|\, B(1) = 1)$.
Solution: $\mathbb{P}(B(2) \le 2 \,|\, B(1) = 1) = \int_{-\infty}^{2} f_{B(2)|\,B(1)}(x|\,1)\, dx = \int_{-\infty}^{2} \frac{\varphi(x-1)\varphi(1)}{\varphi(1)}\, dx = \int_{-\infty}^{1} \varphi(x)\, dx = \Phi(1)$.
Another Solution: $\mathbb{P}(B(2) \le 2 \,|\, B(1) = 1) = \mathbb{P}(B(2) - B(1) \le 1 \,|\, B(1) = 1) \stackrel{ind}{=} \mathbb{P}(B(2) - B(1) \le 1) \stackrel{stat.}{=} \mathbb{P}(B(1) \le 1) = \Phi(1)$.
Exercise. Show that $\mathbb{C}ov(B(s),\, B(t)) = \min(s, t)$.
Solution: Suppose $s < t$. Then, $\mathbb{C}ov(B(s),\, B(t)) = \mathbb{E}[B(s)B(t)] - \mathbb{E}[B(s)]\mathbb{E}[B(t)] = \mathbb{E}[B(s)(B(t) - B(s) + B(s))] = \mathbb{E}[B(s)]\mathbb{E}[B(t) - B(s)] + \mathbb{E}[B(s)]^2 = \mathbb{V}ar[B(s)] = s.$ $\square$

Exercise. Show that $X(t) = \begin{cases} tB(1/t), & \text{if } t > 0 \\ 0, & \text{if } t = 0 \end{cases}$ is a standard BM.
Solution: $\mathbb{E}[tB(1/t)] = 0,\ \mathbb{V}ar[tB(1/t)] = t^2/t = t.$

Exercise (Scaling Relation). Show that for any fixed s and $t \ge 0$, $B(st)$ and $\sqrt{s}\, B(t)$ have the same distribution.
Solution: $B(st) \sim N(0, st)$ and $\sqrt{s}B(t) \sim N(0, st)$.

Exercise. Show that for $t > s$, $\mathbb{E}[B(t) \,|\, B(s)] = B(s)$. Processes with this property are called martingales.
Solution: $\mathbb{E}[B(t) \,|\, B(s)] = \mathbb{E}[B(t) - B(s) + B(s) \,|\, B(s)] = \mathbb{E}[B(t) - B(s)] + B(s) = B(s)$.

Exercise: Show that $\exp\{B(t) - t/2\}$ is a martingale.
Solution: $\mathbb{E}[\exp\{B(t) - t/2\} \,|\, B(s)] = \mathbb{E}[\exp\{B(t) - B(s) + B(s) - t/2\} \,|\, B(s)] = \mathbb{E}[\exp\{B(t) - B(s)\}]\, \exp\{B(s) - t/2\} = \exp\{(t - s)/2 + B(s) - t/2\} = \exp\{B(s) - s/2\}$.

Exercise. Find the mean and the covariance of $X(t) = B(t) - tB(1),\ 0 < t < 1$. $X(t)$ is called a Brownian bridge.
Solution: $\mathbb{E}[B(t) - tB(1)] = t - t = 0$. Suppose $0 < s < t < 1$. Then, $\mathbb{C}ov[B(t) - tB(1), B(s) - sB(1)] = \mathbb{E}[B(t)B(s) - tB(s)B(1) - sB(t)B(1) + stB^2(1)] = s - ts - st + st = s(1 - t)$.

10.2. Hitting Times.

Definition. Let T_a denote the first time the standard BM hits a.
To compute the distribution of $T_a,\ a > 0$, write

$$\mathbb{P}(B(t) \ge a) = \mathbb{P}(B(t) \ge a \,|\, T_a \le t)\mathbb{P}(T_a \le t) + \mathbb{P}(B(t) \ge a \,|\, T_a > t)\mathbb{P}(T_a > t)$$

$$\stackrel{sym}{=} \frac{1}{2}\mathbb{P}(T_a \le t).$$

Therefore, $\mathbb{P}(T_a \le t) = 2\mathbb{P}(B(t) \ge a) = 2\bar{\Phi}(a/\sqrt{t})$.
Exercise. Consider a standard BM. Compute $\mathbb{P}(T_5 > 10)$.
Solution: $\mathbb{P}(T_5 > 10) = 1 - 2(1 - \Phi(5/\sqrt{10})) = 2\Phi(1.58) - 1$.

Exercise. Find the density of T_a.
Solution: $f_{T_a}(t) = \frac{a}{t\sqrt{t}} \frac{1}{\sqrt{2\pi}} e^{-a^2/(2t)}$.

10.2. Maximum Variable.

Definition. Let $M(t) = \max_{0 \le s \le t} B(s)$ be the maximum value that the standard BM attains on an interval $[0, t]$.
To find its distribution, notice that $M(t) \ge a$ iff $T_a \le t$. Hence, $\mathbb{P}(M(t) \le a) = 2\Phi(a/\sqrt{t}) - 1$.

10.3.1. Brownian Motion with Drift.

Definition. The process $X(t) = \mu\, t + \sigma\, B(t)$ is called a BM with drift coefficient μ and volatility σ (diffusion coefficient σ).

Exercise. Find the mean and the variance of $X(t)$.
Solution: $\mathbb{E}[X(t)] = \mu\, t,\ \mathbb{V}ar[X(t)] = \mathbb{E}[\sigma\, B(t)]^2 = \sigma^2\, t$.

Exercise. Find the distribution of $X(t)$.
Solution: $X(t) \sim N(\mu\, t,\ \sigma^2\, t)$.

Exercise. Find the distribution of $X(s) + X(t)$ where $s \le t$.
Solution: $X(s) + X(t) = 2X(s) + X(t) - X(s)$. $2X(s) \sim N(2\mu\, s,\ 4\sigma^2\, s)$ and is independent of $X(t) - X(s) \sim N(\mu(t-s),\ \sigma^2\,(t-s))$. Therefore, $X(s) + X(t) \sim N(4\mu\, s + \mu(t-s),\ 4\sigma^2\, s + \sigma^2(t-s))$.

10.3.2. Geometric Brownian Motion.

Definition. The process $Y(t) = e^{X(t)}$ where $X(t) = \mu\, t + \sigma\, B(t)$ is called the geometric BM.
Exercise. Compute the mean and variance of geometric BM.
Solution: $\mathbb{E}[Y(t)] = \mathbb{E}[e^{X(t)}] = \mathbb{E}[e^{\mu\, t+\sigma\, B(t)}] = e^{\mu\, t}\, e^{(\sigma^2\, t)/2}$.
$\mathbb{V}ar[Y(t)] = \mathbb{E}[Y(t)]^2 - (\mathbb{E}[Y(t)])^2 = \exp\{2\mu\, t + 2\sigma^2\, t\} - \exp\{2\mu\, t + \sigma^2\, t\} = \exp\{2\mu\, t + \sigma^2\, t\}(\exp\{\sigma^2\, t\} - 1)$.

PRICING STOCK OPTIONS

DEFINITION A **stock option** is a right to purchase a stock at a future time at a fixed price. Note that the option is a right, but not an obligation.
DEFINITION. An **investor's portfolio** is a collection of shares of a stock, options to buy a stock, and bonds.

EXAMPLE (OPTION PRICING) Suppose we have an option to purchase a stock at a future time at a fixed price. How much should we pay for this option now?
Suppose that the present value of a stock is \$100 per share. After one time period it will be worth either \$200 or \$50. We are given an option, at a cost of cy, to buy y shares of this stock in one time unit for \$150 per share. So, if we purchase the option, and the stock rises to \$200, we will exercise the option and get \$50 per share of a net profit. However, if the stock is worth \$50, we will not exercise the option.
Besides the option, we purchase for our portfolio x shares of the stock for \$100 each. We will assume that both x and y can be positive as well as negative (negative meaning we are selling the stock or option).
We are interested in determining c, the unit cost of the option. We will show that unless $c = 50/3$, there will be a portfolio that will result in a positive gain.

Our portfolio at time 1 will be worth \$200x+\$50y if the stock price is \$200, and \$50x if the stock price is \$50. Suppose we choose y so that $200x + 50y = 50x$, or $y = -3x$. Therefore, the cost of the original portfolio is $100x - 3xc$, and the gain is $50x - 100x + 3xc = (3c - 50)x$ which is zero if $c = 50/3$ and can be made positive otherwise. For example, if $c = 20$, then we buy one share of stock (x=1) and sell three shares of the option (y=-3) and make the profit of \$10. If $c = 15$, then we sell one share of stock (x=-1) and buy three shares of options (y=3), and attain the profit of \$5.
In this example, there is no arbitrage only if $c = 50/3$. □

THE ARBITRAGE THEOREM.

Consider an experiment with the sample space $S = \{1, 2, \dots, m\}$. Suppose that n wagers are available. A betting scheme is a vector $\mathbf{x} = (x_1, \dots, x_n)$

such that x_1 is bet on wager 1, X_2 is bet on wager 2,..., and x_n is bet on wager n. If the outcome of the experiment is j, then the return from the betting scheme is $\sum_{i=1}^{n} x_i \, r_i(j)$.

Theorem (The Arbitrage Theorem). If X is the outcome of the experiment, then there is a probability vector $\mathbf{p} = (p_1, \dots, p_m)$ for X such that for all $i = 1, \dots, n$, $\mathbb{E}[r_i(X)] = \sum_{j=1}^{m} p_j \, r_i(j) = 0$. Else, there is a betting scheme that leads to a sure win.

Example In our example, there are two possible outcomes – the values of the stock at time 1 – \$200 and \$50. There are two wagers: to buy (sell) the stock, and to buy (sell) the option. By the arbitrage theorem, there is no sure win strategy if there exists $(p, 1-p)$ such that the expected return under both wagers is 0. The return for stock is $200 - 100 = 100$ if the outcome is 200, and $50 - 100 = -50$ if the outcome is 50. Thus, $\mathbb{E}[$stock return$] = 100\,p - 50(1-p) = 0$ iff $p = 1/3$. Then, the expected return for option is $\mathbb{E}[$ option return$] = (50 - c)p - c(1-p) = 50p - c = 50/3 - c = 0$ iff $c = 50/3$. $\square$

10.4.3. The Black-Scholes Option Pricing Formula.

Definition. Suppose we will be given amount $\$v$ at time t. If we were given $\$v$ now, we could've loaned out the money with interest at a continuously compounded rate of $100\,\alpha\%$ per unit time, and get $\$v\,e^{\alpha t}$ at time t. Therefore, the present value of $\$v$ given to us at time t is $v\,e^{-\alpha t}$. The quantity α is called the discount factor. The function $e^{-\alpha t}$ is called the discount function.

Derivation of the Black-Scholes (Merton) Model (1973).

Let $X(t)$ be the price of a stock at time t. The present value of the stock is $X(0) = x_0$.
Suppose we have two wagers. One is to buy (or sell) the stock at the price $X(s)$ at time $s < t$, and then sell (or buy) this stock at time t for the price $X(t)$. The other wager is to buy (or sell) an option that gives us the right to buy stock at time t for a price K per share.
To deal with the first wager, compute the present values of the stock prices $X(s)$ and $X(t)$, so that they are on the same scale. The present value of the stock price at time s is $e^{-\alpha s}\,X(s)$, and the present value of the stock price at time t is $e^{-\alpha t}\,X(t)$.
By the arbitrage theorem, there is no sure win strategy if the expected return of this wager is zero, that is,

$$\mathbb{E}[e^{-\alpha t}\,X(t) \,|\, X(u),\, 0 \le u \le s] = e^{-\alpha s}\,X(s). \quad (*)$$

Now we go to the second wager. The option is worth $(X(t) - K)^+$ at time t, hence, and the present value of the option is $e^{-\alpha t}(X(t) - K)^+$. If c is the cost of the option, then, by the arbitrage theorem,

$$\mathbb{E}[e^{-\alpha t}(X(t)-K)^+ - c] = 0. \quad (**)$$

To find c, we need to find a probability that satisfies (*). Then, c will satisfy (**) computed under the same probability.

The Black-Scholes model assumes that $X(t)$ is a geometric Brownian motion, that is, $X(t) = x_0\, e^{Y(t)}$ where $Y(t) \sim N(\mu\, t, \sigma^2\, t)$. Therefore,

$$\mathbb{E}\Big[X(t)\,|\,X(u),\, 0 \le u \le s\Big] = \mathbb{E}\Big[x_0\, e^{Y(t)}\,|\,Y(u),\, 0 \le u \le s\Big]$$

$$= x_0\,\mathbb{E}\Big[e^{Y(t)-Y(s)+Y(s)}\,|\,Y(u),\, 0 \le u \le s\Big]$$

$$= x_0\, e^{Y(s)}\,\mathbb{E}\Big[e^{Y(t)-Y(s)}\,|\,Y(u),\, 0 \le u \le s\Big]$$

$$= X(s)\mathbb{E}\Big[e^{Y(t)-Y(s)}\Big] = X(s)\, e^{\mu(t-s)+\sigma^2\,(t-s)/2}.$$

Let $\mu + \sigma^2/2 = \alpha$. Then,

$$\mathbb{E}[e^{-\alpha\, t}\, X(t)\,|\,X(u),\, 0 \le u \le s] = e^{-\alpha\, t}\, X(s)\, e^{\alpha\,(t-s)} = e^{-\alpha\, s}\, X(s).$$

Thus, the equation (*) is satisfied if we choose the probability corresponding to $X(t) = x_0\, e^{Y(t)}$ where $Y(t) \sim N(\mu\, t, \sigma^2\, t)$, and where $\mu + \sigma^2/2 = \alpha$. Now we can compute c from equation (**). We have

$$c = \mathbb{E}\Big[e^{-\alpha\, t}(X(t)-K)^+\Big] = e^{-\alpha\, t}\int_{-\infty}^{\infty}(x_0\, e^y - K)^+\,\frac{1}{\sqrt{2\pi\,\sigma^2\, t}}\, e^{-(y-\mu\, t)^2/(2\sigma^2\, t)}\, dy$$

$$= \int_{\ln(K/x_0)}^{\infty}(x_0\, e^y - K)\,\frac{1}{\sqrt{2\pi\,\sigma^2\, t}}\, e^{-(y-\mu\, t)^2/(2\sigma^2\, t)}\, dy$$

$$\overset{\text{pg616}}{=\cdots=} x_0\,\Phi(\sigma\,\sqrt{t} + b) - Ke^{-\alpha\, t}\Phi(b)$$

where

$$b = \frac{\mu\, t - \ln(K/x_0)}{\sigma\sqrt{t}}.$$

Example. The current price of a stock is \$100. Suppose the stock price can be modelled by the Black-Scholes model with drift coefficient $\mu = -0.45$ and volatility $\sigma = 1$. Compute the cost of the option to buy the stock at time $t = 3$ for the cost of $K = \$100$.

Solution: $\alpha = \mu + \sigma^2/2 = 0.05$, $b = \frac{(-0.45)(3)-\ln(100/100)}{(1)\sqrt{3}} \approx -0.78$, $c = (100)\Phi(0.95)-(100)e^{-(0.05)(3)}\Phi(-0.78) \approx (100)(0.8289)-(100)(0.8607)(0.2177) \approx \$64.17$.

10.5 White Noise. Ito's Stochastic Integral.

Definition. Let function f be such that f' exists and is continuous in the interval $[a,b]$. Consider $a = t_0 < t_1 < t_2 < \cdots < t_n = b$, a partition of $[a,b]$. Denote by $\Delta(t) = \max(t_i - t_{i-1}),\ i = 1,\dots,n$. For $B(t)$, a standard Brownian motion, define an Ito stochastic integral as

$$\int_a^b f(t)\,dB(t) = \lim_{n\to\infty,\,\Delta(t)\to 0} \sum_{i=1}^{n} f(t_{i-1})[B(t_i) - B(t_{i-1})].$$

Definition. The process $\{dB(t),\ t \geq 0\}$ is called the white noise process. It is not a stochastic process, because its paths are everywhere discontinuous. That is, the derivative of a Brownian motion is not defined in any point.

Proposition 1. Prove that an Ito integral $\int_a^b f(t)\,dB(t)$ is a random variable with mean 0 and variance $\int_a^b f^2(t)\,dt$.
Proof:

$$\mathbb{E}\Big[\int_a^b f(t)\,dB(t)\Big] = \lim_{n\to\infty,\,\Delta(t)\to 0} \sum_{i=1}^{n} \mathbb{E}\Big[f(t_{i-1})[B(t_i) - B(t_{i-1})]\Big]$$

$$= \lim_{n\to\infty,\,\Delta(t)\to 0} \sum_{i=1}^{n} f(t_{i-1})\Big[\mathbb{E}[B(t_i)] - \mathbb{E}[B(t_{i-1})]\Big] = 0.$$

The variance is

$$\mathbb{V}ar\Big[\int_a^b f(t)\,dB(t)\Big] = \lim_{n\to\infty,\,\Delta(t)\to 0} \sum_{i=1}^{n} \mathbb{V}ar\Big[f(t_{i-1})[B(t_i) - B(t_{i-1})]\Big]$$

$$= \lim_{n\to\infty,\,\Delta(t)\to 0} \sum_{i=1}^{n} f^2(t_{i-1})[t_i - t_{i-1}] = \int_a^b f^2(t)\,dt.$$

Proposition 2. The following identity is true for an Ito integral

$$\int_0^t B(s)\,dB(s) = \frac{1}{2}B^2(t) - \frac{1}{2}\,t.$$

Proof: Denote by $\Delta B_i = B(t_i) - B(t_{i-1})$, and let $\Delta t_i = t_i - t_{i-1}$ where $0 = t_0 < t_1 < \cdots < t_n = t$. Notice that $\mathbb{E}[\sum_{i=1}^n (\Delta B_i)^2] = \sum_{i=1}^n \Delta t_i = t$. We need to show that $\mathbb{V}ar[\sum_{i=1}^n (\Delta B_i)^2] \to 0$ as $n \to \infty$. We write

$$\mathbb{V}ar[\sum_{i=1}^{n} (\Delta B_i)^2] = \mathbb{E}\Big[\sum_{i=1}^{n} (\Delta B_i)^2 - t\Big]^2 = \mathbb{E}\Big[\sum_{i=1}^{n}\Big((\Delta B_i)^2 - \Delta t_i\Big)\Big]^2$$

$$= \sum_{i=1}^{n} \mathbb{E}\Big((\Delta B_i)^4 - 2\Delta t_i(\Delta B_i)^2 + (\Delta t_i)^2\Big) + \text{cross terms}$$

$$= \sum_{i=1}^{n} \Big(3(\Delta t_i)^2 - 2(\Delta t_i)^2 + (\Delta t_i)^2\Big) + 0 = 2\sum_{i=1}^{n} (\Delta t_i)^2$$

$$\leq 2 \max \Delta t_i \sum_{i=1}^{n} \Delta t_i = 2\,\Delta(t)\,t \to 0 \text{ as } \ n \to \infty.$$

Therefore, the random variable $\sum_{i=1}^{n} (\Delta B_i)^2$ converges to a constant, which must be its mean, t.
Now, we write

$$B^2(t) = \sum_{i=1}^{n} (B^2(t_i) - B^2(t_{i-1})) = \sum_{i=1}^{n} \Big((B(t_i) - B(t_{i-1}))^2 + 2B(t_i)B(t_{i-1}) - B^2(t_{i-1}) \Big)$$

$$= \sum_{i=1}^{n} \Big((B(t_i) - B(t_{i-1}))^2 + 2B(t_{i-1})(B(t_i) - B(t_{i-1})) \Big).$$

Hence, by the definition of Ito's integral,

$$B^2(t) = \lim_{n\to\infty,\, \Delta(t)\to 0} \sum_{i=1}^{n} (\Delta B_i)^2 + 2 \int_0^t B(s)\,dB(s) = t + 2 \int_0^t B(s)\,dB(s). \quad \square$$

Remark. Proposition 2 shows that Ito's integral is not an ordinary integral from calculus. If it were so, then we would have $B^2(t)/2 = \int_0^t B(s)\,dB(s)$.