Notes on Stochastic Process

Sizhe Liu

Department of Mechanical Science and Engineering

University of Illinois at Urbana-Champaign

Version 1.0

When I get to the Warwick Avenue, please drop the past and be true...

Contents

1 Concepts Review **7**

1.1 Probability model . . . 7
1.2 Independent events and experiments . . . 8
1.3 Random variables(rv's) . . . 8
1.4 Expectations . . . 9
1.4.1 Conditional Expectation . . . 10
1.5 Indicator random variables . . . 11
1.6 Moment generating functions and other transforms . . . 12
1.7 Multiple random variables . . . 13
1.8 IID random variables . . . 13
1.9 Laws of large numbers . . . 14
1.10 Bernoulli process . . . 14
1.11 Random variables as functions of other random variables . . . 16
1.12 Basic inequality . . . 17

2 Some Examples for Review **19**

3 Markov Chains: First Steps **30**

3.1 Chapman–Kolmogorov Relationship . . . 32
3.2 Most Recent Past Property . . . 32
3.3 Joint Distribution . . . 33
3.4 Long-Term Behavior . . . 33
3.5 Simulation . . . 34

4 Markov Chains for Long Term **35**

4.1 Limiting distribution . . . 35
4.2 Stationary distribution . . . 35
4.2.1 Stationary Distribution for Random Walk on a Weighted Graph . . . 37
4.2.2 The Eigenvalue Connection . . . 37

4.3 CAN YOU FIND THEWAY TO STATE a? 38
4.3.1 Recurrence and Transience 38
4.3.2 Canonical decomposition 40
4.3.3 Irreducible Markov Chains 40
4.4 Periodicity 42
4.5 Ergodic Markov Chain 43
4.6 Time reversibility 45
4.6.1 Reversible Markov Chains and Random Walk on Weighted Graphs . 45
4.7 Absorbing Chain 46
4.7.1 Expected Hitting Times for Irreducible Chains 48
4.7.2 Patterns in sequence 48

5 Branching Process **50**

5.1 Mean generation size 50
5.2 Probability Generating Function 51
5.2.1 Sums of Independent Random Variables 52
5.2.2 Moments 52
5.3 Extinction Probability 52

6 Markov Chain Monte Carlo **55**

6.1 METROPOLIS–HASTINGS ALGORITHM 55
6.1.1 Continuous state space 56
6.2 Gibbs Sampler 57
6.2.1 Ising Model 59

7 Gentle Intro to Poisson Process **61**

7.1 First Definition of Poisson Process 61
7.1.1 Translated Poisson Process 63
7.2 Arrival, Interarrival Times 63
7.2.1 Memorylessness 64

7.3 Infinitesimal Probability . . . 67
7.4 Thinning, Superposition . . . 68
7.4.1 Embedding and the Birthday Problem . . . 70
7.5 Uniform Distribution . . . 72
7.5.1 Simulation . . . 74
7.6 Spatial Poisson Process . . . 75
7.7 Non-homogeneous Poisson Process . . . 75
7.8 Parting Paradox . . . 76

8 Continuous-Time Markov Chains 79

8.1 Transition rate and alarm clocks . . . 80
8.2 Infinitesimal Generator . . . 81
8.2.1 Forward, Backward Equations . . . 82
8.2.2 Matrix Exponential . . . 83
8.2.3 Diagonalization . . . 85
8.3 Long-term Behavior . . . 85
8.3.1 Absorbing states . . . 87
8.3.2 Global Balance . . . 89
8.4 Time reversibility . . . 89
8.4.1 Birth-and-Death Process . . . 90
8.4.2 Long-term stationary distribution and Embedding stationary distribution . . . 92
8.5 Queueing Theory . . . 92
8.5.1 M/M/c Queue . . . 93
8.6 Poisson Subordination . . . 95
8.6.1 Long-Term Behavior, Simulation, Computation . . . 97

9 Brownian Motion 99

9.1 Simulating Brownian Motion . . . 100
9.1.1 Sample Space for Brownian Motion and Continuous Paths . . . 101

9.2 Brownian Motion and Random Walk 102
9.2.1 Invariant Principle 102
9.3 Gaussian Process 103
9.3.1 Nowhere differentiable path 105
9.4 Transformation and Properties 105
9.4.1 Markov Process 106
9.4.2 First Hitting Time and Strong Markov Property 106
9.4.3 Reflection Principle and the Maximum of Brownian Motion 107
9.4.4 Zeros of Brownian Motion and Arcsine Distribution 108
9.5 Variations and Applications 110
9.5.1 Brownian Bridge 111
9.5.2 Geometric Brownian Motion 112
9.6 Martingales 114
9.6.1 Black-Scholes 117
9.6.2 Optional Stopping Theorem 119
9.7 Two Solved Exercises 122

10 A Gentle Intro to Stochastic Calculus 125

10.1 White Noise 127
10.2 Ito Integral 128
10.2.1 Ito's Lemma 129
10.3 Stochastic Differential Equations 131
10.3.1 Numerical Approximation and the Euler–Maruyama Method 134

11 Introduction for Less-dumb 137

11.1 Renewal Processes 139
11.2 Convolution 140
11.3 Calculation of an expectation of a counting process 142
11.3.1 Laplace transform 142
11.3.2 Expectation of a counting process 142

11.3.3 An example of calculating expectation of a counting process 143
11.4 Limit theorems for renewal processes 143
11.5 Solved exercises 145

12 Poisson Process for Less-dumb 147

12.1 Memoryless Property 148
12.2 Other Definition of Poisson processes 148
12.3 Non-homogeneous Poisson processes 150
12.4 Relation between renewal theory and non-homogeneous Poisson processes . . 151
12.5 Elements of Queueing theory: M/G/K systems 153
12.6 Solved Exercises 153

1 Concepts Review

1.1 Probability model

Probability models have 3 components: a sample space Ω, which is an arbitrary set of sample points; a collection of events, each of which is a subset of Ω; and a probability measure, which assigns a probability (a number in $[0,1]$) to each event. The collection of events satisfies the following axioms:

1. Ω is an event.
2. If $A_1, A_2, \ldots,$ are events, then $\cup_{n=1}^{\infty} A_n$ is an event.
3. If A is an event, the complement A^c is an event.

If all sample points are singleton events, then all finite and countable sets are events (i.e., they are finite and countable unions of singleton sets).

Δ Theorem

deMorgan's law

$$\left[\bigcup_n A_n\right]^c = \bigcap_n A_n^c$$

From the law above, countable intersections of events are events. All combinations of intersections and unions of events are also events.

The probability measure on events satisfies the following axioms:

1. $\Pr\{\Omega\} = 1$
2. If A is an event, then $\Pr\{A\} \geq 0$.
3. If $A_1, A_2, \ldots$ are disjoint events, then
   $$\Pr\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \sum_{n=1}^{\infty} \Pr\{A_n\} = \lim_{m\to\infty} \sum_{n=1}^{m} \Pr\{A_n\}$$

A few simple consequences are:

$$\begin{gathered}
\Pr\{\phi\} = 0 \\
\Pr\left\{\bigcup_{n=1}^{m} A_n\right\}^3 = \sum_{n=1}^{m} \Pr\left\{A_n\right\} \\
\Pr\left\{A^c\right\} = 1 - \Pr\{A\} \\
\Pr\{A\} \leq \Pr\{B\} \\
\Pr\left\{A_n\right\} \leq 1 \\
\sum_{n=1}^{n} \Pr\left\{A_n\right\} \leq 1 \\
\Pr\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \lim_{n\to\infty} \Pr\left\{\bigcup_{n=1}^{m} A_n\right\}\} \\
\Pr\left\{\bigcup_{n=1}^{\infty} A_n\right\} = \lim_{n\to\infty} \Pr\left\{A_n\right\}
\end{gathered} \tag{1.1}$$

1.2 Independent events and experiments

Two events A_1 and A_2 are independent if $\Pr\{A_1 A_2\} = \Pr\{A_1\}\Pr\{A_2\}$ Given two probability models, a combined model can be defined in which, first, the sample space Ω is the Cartesian product $\Omega_1 \times \Omega_2$, and, second, for every event A in model 1 and B in model 2 , $\Pr\{AB\} = \Pr\{A\}\Pr\{B\}$.

1.3 Random variables(rv's)

Δ Definition

A rv X (or $X(\omega)$) is a function from Ω to $\mathbb{R}$. This function must satisfy the constraint that $\{\omega : X(\omega) \leq a)\}$ is an event for all $a \in \mathbb{R}$. Also, if $X_1, X_2, \ldots X_n$ are each rv's, then $\{\omega : X_1(\omega) \leq a_1; \ldots, X_n(\omega) \leq \ a_n\}$ is an event for all $a_1, \ldots, a_n$ each in $\mathbb{R}$.

Every rv X has a distribution function $F_X(x) = \Pr\{X \leq x\}$. It's a non-decreasing function from 0 to 1.

If X maps only into a finite or countable set of values, it is discrete and has a probability mass function (PMF) where $\mathrm{p}_X(x) = \Pr\{X = x\}$

If $dF_X(x)/dx$ exists and is finite for all x, then X is continuous and has a density, $f_X(x) = dF_X(x)/dx$.

In general, $F_X(x) = \Pr\{X \le x\}$ always exists. Because $X = x$ is included in $X \le x$, we see that if $F_X(x)$ has a jump at x, then $F_X(x)$ is the value at the top of the jump.

Theoretical nit-pick: $F_X(x)$ must be continuous from the right, i.e.,

$$\lim_{k\to\infty} F_X(x + 1/k) = F_X(x)$$

Proof: Let $A_k = \left\{\omega : X(\omega) > x + \frac{1}{k}\right\}$. Then $A_{k-1} \subseteq A_k$ for each $k > 1$.

$$\{\omega : X(\omega) > x\} = \bigcup_{k=1}^{\infty} A_k$$

$\Pr\{X > x\} = \Pr\left\{\bigcup_k A_k\right\} = \lim_k \Pr\{A_k\} = \lim_k \Pr\left\{X > x + \frac{1}{k}\right\}$ Center step: let $B_1 = A_1; B_k = A_k - A_{k-1}$ for $k > 1$ Then $\{B_k; k \ge 1\}$ are disjoint.

$$\begin{aligned} 1 - \Pr\left\{\bigcup_k A_k\right\} &= 1 - \Pr\left\{\bigcup_k B_k\right\} = 1 - \sum_{k=1}^{\infty} \Pr\{B_k\} \\ &= 1 - \lim_{k\to\infty} \sum_{m=1}^{k} B_m = 1 - \lim_{k\to\infty} A_k \end{aligned}$$

1.4 Expectations

$$\begin{array}{ll} \mathrm{E}[X] & = \sum_i a_i \mathrm{p}_X(a_i) \quad \text{for discrete } X \\ \mathrm{E}[X] & = \int x \mathrm{f}_X(x) dx \quad \text{for continuous } X \\ \mathrm{E}[X] & = \int \mathrm{F}^{\mathrm{c}}_X(x) dx \quad \text{for arbitrary nonneg } X \\ \mathrm{E}[X] & = \int_{-\infty}^{0} \mathrm{F}_X(x) dx + \int_0^{\infty} \mathrm{F}^{\mathrm{c}}_X(x) dx \quad \text{for arbitrary } X \end{array} \tag{1.2}$$

Whether or not $X_1, X_2, \ldots, X_n$ are independent, the expected value of $S_n = X_1 + X_2 + \cdots + X_n$ satisfies

$$\mathrm{E}[S_n] = \mathrm{E}[X_1 + X_2 + \cdots + X_n] = \mathrm{E}[X_1] + \mathrm{E}[X_2] + \cdots + \mathrm{E}[X_n]$$

1.4.1 Conditional Expectation

More generally yet, let X be an arbitrary rv and let y be a sample value of a discrete rv Y with $p_Y(y) > 0$. The conditional CDF of X conditional on $Y = y$ is defined as

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

And

$$\mathrm{E}[X|Y = y] = -\int_{-\infty}^{0} \mathrm{F}_{X|Y}(x|y)dx + \int_{0}^{\infty} \mathrm{F}^{\mathrm{c}}_{X|Y}(x|y)dx$$

The conditional expectation of X conditional on a discrete rv Y can also be viewed as a rv. With the possible exception of a set of zero probability, each $\omega \in \Omega$ maps to $\{Y = y\}$ for some y with $p_Y(y) > 0$ and $E[X|Y = y]$ is defined for that y. Thus we can define $E[X|Y]$ as a rv that is a function of Y, mapping ω to a sample value, say y of Y, and mapping that y to $E[X|Y = y]$.

This can be used to express the unconditional mean of X as

$$\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]]$$

where the inner expectation is over X for each sample value of Y and the outer expectation is over the rv $\mathrm{E}[X|Y]$, which is a function of Y.

Generally, if we still assume X to be discrete, we can write out this expectation by using $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]]$ for $\mathrm{E}[X|Y = y]$:

$$\begin{aligned}\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]] &= \sum_y p_Y(y)\mathrm{E}[X|Y = y] \\ &= \sum_y \mathrm{p}_Y(y) \sum_x x\mathrm{p}_{X|Y}(x|y)\end{aligned}$$

△ Theorem

Total expectation: Let X and Y be discrete rvs. If X is non-negative, then $\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]] = \sum_y \mathrm{p}_Y(y)\mathrm{E}[X|Y = y]$ · If X has both positive and negative values, and if at most one of $\mathrm{E}\left[X^{+}\right]$ and $\mathrm{E}\left[-X^{-}\right]$ is infinite, then

$$\mathrm{E}[X] = \mathrm{E}[\mathrm{E}[X|Y]] = \sum_{y} \mathrm{p}_Y(y)\mathrm{E}[X|Y=y]$$

If X and Y are continuous, we can essentially extend these results to probability densities. In particular, defining $E[X|Y=y]$ as

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

we have

$$\mathrm{E}[X] = \int_{-\infty}^{\infty} \mathrm{f}_Y(y)\mathrm{E}[X|Y=y]dy = \int_{-\infty}^{\infty} \mathrm{f}_Y(y) \int_{-\infty}^{\infty} x\mathrm{f}_{X|Y}(x|y)dxdy$$

△ Theorem

Law of Total Possibility: Let $B_1, \ldots, B_k$ be a sequence of events that partition the sample space. That is, the B_i are mutually exclusive (disjoint) and their union is equal to Ω. Then, for any event A,

$$P(A) = \sum_{i=1}^{k} P\left(A \cap B_i\right) = \sum_{i=1}^{k} P\left(A|B_i\right) P\left(B_i\right)$$

1.5 Indicator random variables

△ Definition

For every event A in a probabilty model, an indicator $rv\mathbb{I}_A$ is defined where $\mathbb{I}_A(\omega) = 1$ for $\omega \in A$ and $\mathbb{I}_A(\omega) = 0$ otherwise. Note that $\mathbb{I}_A$ is a binary rv.

From the definition we have:

$$\mathrm{p}_{\mathrm{I}_A}(0) = 1 - \Pr\{A\}; \quad \mathrm{p}_{\mathbb{I}_A}(1) = \Pr\{A\}$$

$$\mathrm{E}\left[\mathbb{I}_A\right] = \Pr\{A\} \quad \sigma_{\mathbb{I}_A} = \sqrt{\Pr\{A\}(1-\Pr\{A\})}$$

1.6 Moment generating functions and other transforms

The MGF for a rv X is given by

$$g_X(r) = \mathrm{E}\left[e^{rX}\right] = \int_{-\infty}^{\infty} e^{rx} d\mathrm{F}_X(x)$$

△ Theorem

If $g_X(r)$ exists in an open region of r around 0 (i.e., if $r_- < 0 < r_+$) , then derivatives of all orders exist in that region. They are given by

$$\frac{d^k g_X(r)}{dr^k} = \int_{-\infty}^{\infty} x^k e^{rx} d\mathrm{F}_X(x) \quad ; \quad \left.\frac{d^k g_X(r)}{dr^k}\right|_{r=0} = \mathrm{E}\left[X^k\right]$$

△ Theorem

let $S_n = X_1 + X_2 + \cdots + X_n$

$$\begin{aligned} \mathfrak{g}_{S_n}(r) = \mathrm{E}\left[e^{rS_n}\right] &= \mathrm{E}\left[\exp\left(\sum_{i=1}^{n} rX_i\right)\right] \\ &= \mathrm{E}\left[\prod_{i=1}^{n} \exp\left(rX_i\right)\right] = \prod_{i=1}^{n} g_{X_i}(r) \end{aligned}$$

If $X_1, \ldots, X_n$ are also IID,

$$g_{S_n}(r) = \left[g_X(r)\right]^n$$

Say $i\theta$ where $i = \sqrt{-1}$ and θ is real. Then $g_X(i\theta) = E\left[e^{i\theta x}\right]$ is called the characteristic function of X. since $\left|e^{i\theta x}\right|$ is 1 for all $x, g_X(i\theta)$ exists for all rvs X and all real θ, and its magnitude is at most 1.

1.7 Multiple random variables

Is a random variable (rv) X specified by its distribution function $F_X(x)$? No, the relationship between rv's is important.

$$F_{XY}(x,y) = \Pr\{\{\omega : X(\omega) \le x\} \cap \{\omega : Y(\omega) \le y\}\}$$

The rv's $X_1, \ldots, X_n$ are independent if $\quad F_{\vec{X}}(x_1, \ldots x_n) = \prod_{m=1}^{n} F_{X_m}(x_m) \quad$ for all $x_1, \ldots, x_n$

For discrete rv's, independence is more intuitive when stated in terms of conditional probabilities:

$$\mathrm{p}_{X|Y}(x|y) = \frac{\mathrm{p}_{XY}(x,y)}{\mathrm{p}_Y(y)}$$

NitPick: If $X_1, \ldots, X_n$ are independent, then all subsets of $X_1, \ldots, X_n$ are independent. (This isn't always true for independent events).

For continuous rv's, X and Y, if they have a joint density and they are IID, we have:

$$f_{XY}(x,y) = f_X(x) f_Y(y) \qquad \text{for each } x \in \mathbb{R}, y \in \mathbb{R}$$

If $f_Y(y) > 0$, the conditional density can be defined as $f_{X|Y}(x|y) = \frac{f_{XY}(x,y)}{f_Y(y)}$. Then statistical independence can be expressed as

$$f_{X|Y}(x|y) = f_X(x) \qquad \text{where } f_Y(y) > 0$$

A set of rv's is said to be pairwise independent if each pair of rv's in the set is independent. pairwise independence does not imply that the entire set is independent.

The marginal probability of rv X_i is defined as:

$$\mathrm{F}_{X_i}(x_i) = \mathrm{F}_{X_1 \cdots X_{i-1} X_{i+1} \cdots X_n}(\infty, \ldots, \infty, x_i, \infty, \ldots, \infty)$$

1.8 IID random variables

The random variables $X_1, \ldots, X_n$ are independent and identically distributed (IID) if for all $x_1, \ldots, x_n$

$$F_{\vec{X}}(x_1, \ldots, x_n) = \prod_{k=1}^{n} F_X(x_k)$$

1.9 Laws of large numbers

Let $X_1, X_2, \ldots, X_n$ be IID rv's with mean $\bar{X}$, variance σ^2. Let $S_n = X_1 + \cdots + X_n$. Then $\sigma^2_{S_n} = n\sigma^2$. The center of the distribution varies with n and the spread (σ_{S_n}) varies with $\sqrt{n}$. The sample average is S_n/n, which is a rv of mean $\bar{X}$ and variance σ^2/n. The center of the distribution is $\bar{X}$ and the spread decreases with $1/\sqrt{n}$.

Note that $S_n - n\bar{X}$ is a zero mean rv with variance $n\sigma^2$. Thus $\frac{S_n - n\bar{X}}{\sqrt{n}\sigma}$ is zero mean, unit variance.

Central limit theorem:

$$\lim_{n\to\infty}\left[\Pr\left\{\frac{S_n - n\bar{X}}{\sqrt{n}\sigma} \le y\right\}\right] = \int_{-\infty}^{y} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right) dx$$

1.10 Bernoulli process

A stochastic process (or random process) is an infinite collection of rv's, usually indexed by an integer or a real number often interpreted as time. Thus each sample point of the probability model maps to an infinite collection of sample values of rv's. If the indexes are regarded as time, then each sample point maps to a function of time called a sample path or sample function.

As an example of sample paths that vary at only discrete times, we might be concerned with the times at which customers arrive at some facility. These 'customers' might be customers entering a store, incoming jobs for a computer system, arriving packets to a communication system, or orders for a merchandising warehouse.

The Bernoulli process is an example of how such customers could be modeled. A Bernoulli process is a sequence, $Y_1, Y_2, \ldots,$ of IID binary random vari- ables. Let $p = \Pr\{Y_i = 1\}$ and $1-p = \Pr\{Y_i = 0\}$. We usually visualize a Bernoulli process as evolving in discrete time with the event $\{Y_i = 1\}$ representing an arriving customer at time i and $\{Y_i = 0\}$ representing no arrival. Thus at most one arrival occurs at each integer time. We visualize the process as starting at time 0, with the first opportunity for an arrival at time 1.

First, consider the first interarrival time, X_1, which is defined as the time of the first arrival. If $Y_1 = 1$, then (and only then)$X_1 = 1$. Thus $\mathrm{p}_{X_1}(1) = p$. Next, $X_1 = 2$ if and only $Y_1 = 0$ and $Y_2 = 1$, so $\mathrm{p}_{X_1}(2) = pq$. Continuing, we see that X_1 has the geometric PMF,

$$\mathrm{p}_{X_1}(j) = p(1-p)^{j-1}$$

Each subsequent interarrival time X_k can be found in this same way. It has the same geometric PMF and is statistically independent of $X_1, \ldots, X_{k-1}$. Thus the sequence of interarrival times is an IID sequence of geometric rv's.

Let $S_n = \sum_{i=1}^{n} Y_i$, then S_n is simply the sum of n binary rv's and thus has the binomial distribution. $\mathrm{p}_{S_n}(k)$ is the probability that k out of n of the Y_i 's have the value 1:

$$\mathrm{p}_{S_n}(k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k q^{n-k}$$

The ratio of k to n in these bounds is denoted $\widetilde{p} = k/n$ and the bounds are given in terms of a quantity $D(\widetilde{p}\|p)$ called the binary Kullback-Liebler divergence (or relative entropy) and defined by

$$\mathrm{D}(\widetilde{p}\|p) = \widetilde{p}\ln\left(\frac{\widetilde{p}}{p}\right) + (1-\widetilde{p})\ln\left(\frac{1-\widetilde{p}}{1-p}\right) \geq 0$$

· **Lemma**

Let ps_n $(\tilde{p}n)$ be the PMF of the binomial distribution for an underlying binary PMF py $(1) = p > 0$, py $(0) = q > 0$. Then for each integer $\tilde{p}n, 1 \leq \tilde{p}n \leq n-1$

$$\begin{aligned} p_{S_n}(\widetilde{p}n) &< \sqrt{\tfrac{1}{2\pi n\widetilde{p}(1-\widetilde{p})}}\exp[-nD(\widetilde{p}\|p)] \\ \mathrm{p}_{S_n}(\widetilde{p}n) &> \left(1 - \tfrac{1}{12n\widetilde{p}(1-\widetilde{p})}\right)\sqrt{\tfrac{1}{2\pi n\widetilde{p}(1-\widetilde{p})}}\exp[-n\mathrm{D}(\widetilde{p}\|p)] \end{aligned}$$

Also, $\mathrm{D}(\widetilde{p}\|p) \geq 0$ with strict inequality for all $\widetilde{p} \neq p$

An upper and lower bound of this type is said to be asymptotically tight, and the result is denoted as

$$\mathrm{p}_{S_n}(\widetilde{p}n)) \sim \sqrt{\frac{1}{2\pi n\widetilde{p}(1-\widetilde{p})}}\exp[-n\mathrm{D}(\widetilde{p}\|p)] \quad \text{for } 0 < \widetilde{p} < 1$$

1.11 Random variables as functions of other random variables

Random variables are often defined in terms of each other. For example, if h is a function from $\mathbb{R}$ to $\mathbb{R}$ and X is a rv, then $Y = h(X)$ is the rv that maps each sample point ω to the composite function $h(X(\omega))$. And the expectation of Y is:

$$\mathrm{E}[Y] = \int_{-\infty}^{\infty} h(x)d\mathrm{F}_X(x)$$

The existence of E [X] does not guarantee the existence of E [Y]

Next suppose X and Y are rv s and consider the rv20 Z = X + Y. If we assume that X and Y are independent, then the CDF of Z = X + Y is given by:

$$\mathrm{F}_Z(z) = \int_{-\infty}^{\infty} \mathrm{F}_X(z-y)d\mathrm{F}_Y(y) = \int_{-\infty}^{\infty} \mathrm{F}_Y(z-x)d\mathrm{F}_X(x)$$

The equation above is referred as the convolution of CDF. If X and Y both have densities, this can be rewritten as

$$\mathrm{f}_Z(z) = \int_{-\infty}^{\infty} \mathrm{f}_X(z-y)\mathrm{f}_Y(y)dy = \int_{-\infty}^{\infty} \mathrm{f}_Y(z-x)\mathrm{f}_X(x)dx$$

If X and Y are non-negative rv s, then the integrands in these two equations are non-zero only between 0 and z, so we often use 0 and z as the limits.

If $X_1, X_2, \ldots, X_n$ are independent rvs, then the distribution of the rv $S_n = X_1 + X_2 + \cdots + X_n$ can be found by first convolving the CDFs of X_1 and X_2 to get the CDF of S_2 and then, for each $n \geq 2$, convolving the CDF of S_n and X_{n+1} to get the CDF of S_{n+1} The CDFs can be convolved in any order to get the same resulting CDF.

1.12 Basic inequality

This is the simplest and most basic of these inequalities. It states that if a non-negative rv Y has a mean $\mathrm{E}[Y]$, then, for every $y > 0, \Pr\{Y \geq y\}$ satisfies

$$\Pr\{Y \geq y\} \leq \frac{\mathrm{E}[Y]}{y} \quad \text{for every } y > 0 \quad \text{(Markov inequality)}$$

We now use the Markov inequality to establish the well-known Chebyshev inequality. Let Z be an arbitrary rv with finite mean $\mathrm{E}[Z]$ and finite variance σ_Z^2, and define Y as the non-negative rv $Y = (Z - \mathrm{E}[Z])^2$. Thus $\mathrm{E}[Y] = \sigma_Z^2$. Applying Markov inequality, we have:

$$\Pr\left\{(Z - \mathrm{E}[Z])^2 \geq y\right\} \leq \frac{\sigma_Z^2}{y} \quad \text{for every } y > 0$$

Replacing y with ϵ^2 and noting that the event $\{(Z - E[Z])^2 \geq \epsilon^2\}$ is the same as $|Z - \mathrm{E}[Z]| \geq \epsilon$, this becomes

$$\Pr\{|Z - \mathrm{E}[Z]| \geq \epsilon\} \leq \frac{\sigma_Z^2}{\epsilon^2} \quad \text{for every } \epsilon > 0 \quad \text{(Chebyshev inequality)}$$

For any given rv$Z$, let $I(Z)$ be the interval over which the MGF$\mathrm{g}_Z(r) = \mathrm{E}\left[e^{rZ}\right]$ exists. Letting $Y = e^{rZ}$ for any $r \in I(Z)$, the Markov inequality applied to Y is

$$\Pr\{\exp(rZ) \geq y\} \leq \frac{g_Z(r)}{y} \quad \text{for every } y > 0 \quad \text{(Chernoff bound).}$$

if y is replaced by e^{rb}. Note that $\exp(rZ) \geq \exp(rb)$ is equivalent to $Z \geq b$ for $r > 0$ and to $Z \leq b$ for $r < 0$. Thus, for any real b, we get the following two bounds, one for $r > 0$ and the other for $r < 0$:

$$\Pr\{Z \geq b\} \leq \mathfrak{g}_Z(r)\exp(-rb) \quad \text{(Chernoff bound for } r > 0, r \in I(Z))$$
$$\Pr\{Z \leq b\} \leq \mathfrak{g}_Z(r)\exp(-rb) \quad \text{(Chernoff bound for } r < 0, r \in I(Z))$$

· **Lemma**

Let $\{X_i; i \geq 1\}$ be IID rvs and let $S_n = X_1 + \cdots + X_n$ for each $n \geq 1$. Let $I(X)$ be the interval over which the semi-invariant MGF $\gamma_X(r)$ is finite and assume 0 is in the interior of $I(X)$. Then, for each $n \geq 1$ and each real number a,

$$\Pr\{S_n \geq na\} \leq \exp\left(n\mu_X(a)\right) \quad \text{where } \mu_X(a) = \inf_{r \geq 0, r \in I(X)} \gamma_X(r) - ra|$$

Furthermore, $\mu_X(a) < 0$ for $a > \bar{X}$ and $\mu_X(a) = 0$ for $a \leq \bar{X}$. And inf means the largest lower bound.

2 Some Examples for Review

· **Example**

Gambler's ruin The gambler's ruin problem was introduced in Example 1.6. A gambler starts with k dollars. On each play a fair coin is tossed and the gambler wins $1 if heads occurs, or loses $1 if tails occurs. The gambler stops when he reaches $n (n > k) or loses all his money. Find the probability that the gambler will eventually lose.

Hint: We make two observations, which are made more precise in later chapters. First, the gambler will eventually stop playing, either by reaching n or by reaching 0. One might argue that the gambler could play forever. However, it can be shown that that event occurs with probability 0. Second, assume that after, say, 100 wagers, the gambler's capital returns to $k. Then, the probability of eventually winning $n is the same as it was initially. The memoryless character of the process means that the probability of winning $n or losing all his money only depends on how much capital the gambler has, and not on how many previous wagers the gambler made.

Let p_k denote the probability of reaching n when the gambler's fortune is k. What is the gambler's status if heads is tossed? Their fortune increases to $k+1$ and the probability of winning is the same as it would be if the gambler had started the game with $k+1$. Similarly, if tails is tosed and the gambler's fortune decreases to $k-1$. Hence,

$$p_k = p_{k+1}\left(\frac{1}{2}\right) + p_{k-1}\left(\frac{1}{2}\right)$$

$$p_{k+1} - p_k = p_k - p_{k-1}, \qquad \text{for} \qquad k = 1, \ldots, n-1$$

with $p_0 = 0$ and $p_n = 1$. Unwinding the recurrence gives

$$p_k - p_{k-1} = p_{k-1} - p_{k-2} = p_{k-2} - p_{k-3} = \cdots = p_1 - p_0 = p_1$$

for $k = 1, \ldots, n$. We have that $p_2 - p_1 = p_1$, giving $p_2 = 2p_1$. Also, $p_3 - p_2 = p_3 - 2p_1 = p_1$, giving $p_3 = 3p_1$. More generally, $p_k = kp_1$, for $k = 1, \ldots, n$ Sum over suitable k to obtain

$$\sum_{k=1}^{n-1}(p_{k+1} - p_k) = \sum_{k=1}^{n-1}(p_k - p_{k-1})$$

which gives $1-p_1=p_{n-1}=(n-1)p_1$, so $p_1=1/n$. Thus

$$p_k=kp_1=\frac{k}{n},\text{ for } k=0,\ldots,n$$

The probability that the gambler eventually wins \$$n$ is k/n. Hence, the probability of the gambler's ruin is $(n-k)/n$.

Δ Theorem

(Bayes' Rule)For events A and B

$$P(B|A)=\frac{P(A|B)P(B)}{P(A|B)P(B)+P\left(A|B^c\right)P\left(B^c\right)}$$

a more general formcof Bayes' rule is

$$P\left(B_i|A\right)=\frac{P\left(A|B_i\right)P\left(B_i\right)}{\sum_j P\left(A|B_j\right)P\left(B_j\right)},\quad \text{for}\quad i=1,2,\ldots$$

Δ Theorem

Conditional PMF of Y given X=x is:

$$P(Y=y|X=x)=\frac{P(X=x,Y=y)}{P(X=x)}$$

Note that the conditional pmf *is* a probability function. For fixed x, the probabilities $P(Y=y|X=x)$ are nonnegative and sum to 1, as

$$\sum_y P(Y=y|X=x)=\sum_y\frac{P(X=x,Y=y)}{P(X=x)}=\frac{P(X=x)}{P(X=x)}=1$$

· **Example**

The joint pmf of X and Y is

$$P(X=x,Y=y)=\frac{x+y}{18},\text{ for } x,y=0,1,2$$

Find the conditional pmf of Y given $X=x$

Solution: The marginal distribution of X is

$$P(X=x)=\sum_{y=0}^{2}P(X=x,Y=y)=\frac{x}{18}+\frac{x+1}{18}+\frac{x+2}{18}=\frac{x+1}{6}$$

for $x=0,1,2$. The conditional pmf is

$$P(Y=y|X=x)=\frac{P(X=x,Y=y)}{P(X=x)}=\frac{(x+y)/18}{(x+1)/6}=\frac{x+y}{3(x+1)}$$

· **Example**

Random variables X and Y have joint density

$$f(x,y)=e^{-x},\ \text{for } 0<y<x<\infty$$

Find $P(Y<2|X=5)$

Solution:

$$P(Y<2|X=5)=\int_0^2 f_{Y|X}(y|5)dy$$

To find the conditional density function $f_{Y|X}(y|x)$, find the marginal density

$$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy=\int_0^x e^{-x}dy=xe^{-x},\qquad \text{for}\qquad x>0$$

$$f_{Y|X}(y|x)=\frac{f(x,y)}{f_X(x)}=\frac{e^{-x}}{xe^{-x}}=\frac{1}{x},\ \text{for } 0<y<x$$

$$P(Y<2|X=5)=\int_0^2 f_{Y|X}(y|5)dy=\int_0^2\frac{1}{5}dy=\frac{2}{5}$$

△ **Theorem**

Conditional densities are used to compute conditional probabilities. For $R\subseteq\mathbb{R}$,

$$P(Y\in R|X=x)=\int_R f_{Y|X}(y|x)dy$$

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

· **Example**

Tom picks a real number X uniformly distributed on (0, 1). Tom shows his number x to Marisa who then picks a number Y uniformly distributed on (0, x). Find (i) the conditional distribution of Y given X = x; (ii) the joint distribution of X and Y; and (iii) the marginal density of Y.

Solution: (i) The conditional distribution of Y given $X = x$ is given directly in the statement of the problem. The distribution is uniform on $(0, x)$. Thus,

$$f_{Y|X}(y|x) = \frac{1}{x}, \qquad \text{for} \qquad 0 < y < x$$

(ii) For the joint density,

$$f(x, y) = f_{Y|X}(y|x)f_X(x) = \frac{1}{x}(1) = \frac{1}{x}, \qquad \text{for} \qquad 0 < y < x < 1$$

(iii) To find the marginal density of Y, integrate out the x variable in the joint density function. This gives

$$f_Y(y) = \int_{-\infty}^{\infty} f(x, y)dx = \int_y^1 \frac{1}{x}dx = -\ln y, \qquad \text{for} \qquad 0 < y < 1$$

△ **Theorem**

Conditional Expectation of Y given $X = x$

$$E(Y|X = x) = \begin{cases} \sum_y yP(Y = y|X = x), & \text{discrete} \\ \int_{-\infty}^{\infty} yf_{Y|X}(y|x)dy, & \text{continuous.} \end{cases}$$

· **Example**

Let X and Y be independent Poisson random variables with respective parameters λ and μ. Find the conditional expectation of Y given $X + Y = n$.

Solution: Notice that

$$E(Y|X = x) = \begin{cases} \sum_y yP(Y = y|X = x), & \text{discrete} \\ \int_{-\infty}^{\infty} yf_{Y|X}(y|x)dy, & \text{continuous.} \end{cases}$$

We should have

$$
\begin{aligned}
P(Y=y|X+Y=n) &= \frac{P(Y=y, X+Y=n)}{P(X+Y=n)} = \frac{P(Y=y, X=n-y)}{P(X+Y=n)} \\
&= \frac{P(Y=y)P(X=n-y)}{P(X+Y=n)} \\
&= \frac{\left(e^{-\mu}\mu^y/y!\right)\left(e^{-\lambda}\lambda^{n-y}/(n-y)!\right)}{e^{-(\lambda+\mu)}(\lambda+\mu)^n/n!} \\
&= \binom{n}{y}\left(\frac{\mu}{\lambda+\mu}\right)^y\left(\frac{\lambda}{\lambda+\mu}\right)^{n-y}
\end{aligned}
$$

for $y = 0, \ldots, n$. The form of the conditional pmf shows that the conditional distribution is binomial with parameters n and $p = \mu/(\lambda+\mu)$. The desired conditional expectation is the mean of this binomial distribution. That is,

$$E(Y|X+Y=n) = np = \frac{n\mu}{\lambda+\mu}$$

· **Example**

Assume that X and Y have joint density

$$f(x,y) = \frac{2}{xy}, \text{ for } 1 < y < x < e$$

Find $E(Y|X=x)$

Solution: The marginal density of X is

$$f_X(x) = \int_1^x \frac{2}{xy}dy = \frac{2\ln x}{x}, \text{ for } 1 < x < e$$

The conditional density of Y given $X = x$ is

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{2/(xy)}{2\ln x/x} = \frac{1}{y\ln x}, \text{ for } 1 < y < x$$

with conditional expectation

$$E(Y|X=x) = \int_1^x y f_{Y|X}(y|x)dy = \int_1^x \frac{y}{y\ln x}dy = \frac{x-1}{\ln x}$$

Δ **Theorem**

Properties of Conditional Expectation

1. (Linearity) For constants a and b and random variables $X, Y,$ and Z,
   $E(aY + bZ|X = x) = aE(Y|X = x) + bE(Z|X = x)$
2. If g is a function,
$$E(g(Y)|X = x) = \begin{cases} \sum_y g(y)P(Y = y|X = x), & \text{discrete} \\ \int_{-\infty}^{\infty} g(y)f_{Y|X}(y|x)dy, & \text{continuous.} \end{cases}$$
3. (Independence) If X and Y are independent,
$$E(Y|X = x) = E(Y)$$
4. If $Y = g(X)$ is a function of X
$$E(Y|X = x) = g(x)$$

$\cdot$ **Example**

A fair coin is flipped repeatedly. Find the expected number of flips needed to get two heads in a row.

Solution: Let Y be the number of flips needed Consider three events: (i) T, the first flip is tails; (ii) HT, the first fip in in in the second flip is tails; and (iii) HH, the first two fins are heads. The events $|T, HT, HH$ partition the sample space. By the law of total expectation,

$$\begin{aligned} E(Y) &= E(Y|T)P(T) + E(Y|HT)P(HT) + E(Y|HH)P(HH) \\ &= E(Y|T)\frac{1}{2} + E(Y|HT)\frac{1}{4} + (2)\frac{1}{4} \end{aligned}$$

Consider $E(Y|T)$. Assume that the first fivis tails. Since successive flips are indepen- dent, atter the first ailuws that $E(Y/T) = 1 + E(Y)$. Similarly, after first heads and then tails we starl preragain, having used up two coin tosses. Thus, $E(Y|HT) = 2 + E(Y)$. This gives

$$E(Y) = (1 + E(Y))\frac{1}{2} + (2 + E(Y))\frac{1}{4} + (2)\frac{1}{4} = E(Y)\frac{3}{4} + \frac{3}{2}$$

Solving for $E(Y)$ gives $E(Y)(1/4) = 3/2$, or $E(Y) = 6$

· **Example**

Every day Bob goes to the pizza shop, orders a slice of pizza, and picks a topping—pepper, pepperoni, pineapple, prawns, or prosciutto—uniformly at random. On the day that Bob first picks pineapple, find the expected number of prior days in which he picked pepperoni.

Solution: Let Y be the number of days, before the day Bob first picked pineapple, in which he picks pepperoni. Let X be the number of days needed for Bob to first pick pineapple. Then, X has a geometric distribution with parameter 1/5.

If $X = x$, then on the first $x - 1$ days pineapple was not picked. And for each of these days, given that pineapple was not picked, there was a 1/4 chance of picking pepperoni. The conditional distribution of Y given $X = x$ is binomial with parameters $x - 1$ and 1/4. Thus, $E[Y|X = x] = (x - 1)/4, |$ and

$$E(Y) = \sum_{x=1}^{\infty} E(Y|X = x)P(X = x)$$

$$= \textstyle\sum_{x=1}^{\infty} \left(\frac{x-1}{4}\right)\left(\frac{4}{5}\right)^{x-1}\frac{1}{5}$$
$$= \left(\frac{1}{4}\textstyle\sum_{x=1}^{\infty} x\left(\frac{4}{5}\right)^{x-1}\frac{1}{5}\right) - \left(\frac{1}{4}\textstyle\sum_{x=1}^{\infty}\left(\frac{4}{5}\right)^{x-1}\frac{1}{5}\right)$$
$$= \tfrac{1}{4}E(X) - \tfrac{1}{4} = \tfrac{5}{4} - \tfrac{1}{4} = 1$$

· **Example**

The time that Joe spends talking on the phone is exponentially distributed with mean 5minutes.What is the expected length of his phone call if he talks for more than 2minutes?

Solution Let Y be the length of Joe's phone call. With $A = \{Y > 2\}$, the desired conditional expectation is

$$E(Y|A) = E(Y|Y > 2) = \frac{1}{P(Y > 2)}\int_2^{\infty} y\frac{1}{5}e^{-y/5}dy$$
$$= \left(\frac{1}{e^{-2/5}}\right)7e^{-2/5} = 7 \text{ minutes.}$$

Note that the solution can be obtained using the memoryless property of the exponential distribution. The conditional distribution is equal to distribution of $2 + Z$, where Z has the same distribution has Y. Thus,

$$E(Y|Y > 2) = E(2 + Z) = 2 + E(Z) = 2 + E(Y) = 2 + 5 = 7$$

Δ Theorem

Conditional Expectation of Y given X The conditional expectation $E(Y|X)$ has three defining properties.

1. $E(Y|X)$ is a random variable.
2. $E(Y|X)$ is a function of X.
3. $E(Y|X)$ is equal to $E(Y|X = x)$ whenever $X = x$. That is, if

$$E(Y|X = x) = g(x), \text{ for all } x,$$

then $E(Y|X) = g(X)$. $E(Y|X)$ **is a rv but** $E(Y|X = x)$ **is a function of x**

Δ Theorem

Law of Total Expectation For random variables X and Y

$$E(Y) = E(E(Y|X))$$

· **Example**

Angel will harvest N tomatoes in her garden, where N has a Poisson distribution with parameter λ. Each tomato is checked for defects. The chance that a tomato has diffects is p. Defects are indpendent from tomato to tomato. Find the expected number of tomatoes with defects.

Solution: Let X be the number of tomatoes with defects. The conditional distribution of X given $N = n$ is binomial with parameters n and p. This gives $E(X|N = n) = np$ since this is true for all $n, E(X|N) = Np$. By the law of total expectation,

$$E(X) = E(E(X|N)) = E(Np) = pE(N) = p\lambda$$

· **Example**

Ellen's insurance will pay for a medical expense subject to a \$100 deductible. Assume that the amount of the expense is exponentially distributed with mean \$500. Find the expectation and standard deviation of the payout.

Solution: Let M be the amount of the medical expense and let X be the insurance company's payout. Then,

$$X = \begin{cases} M - 100, & \text{if } M > 100 \\ 0, & \text{if } M \leq 100 \end{cases}$$

where M is exponentially distributed with parameter 1/500. To find the expected payment, apply the law of total expectation, giving

$$\begin{aligned} E(X) = E(E(X|M)) &= \int_0^\infty E(X|M=m)\lambda e^{-\lambda m} dm \\ &= \int_{100}^\infty E(M-100|M=m)\frac{1}{500}e^{-m/500} dm \\ &= \int_{100}^\infty (m-100)\frac{1}{500}e^{-m/500} dm \\ &= 500e^{-100/500} = \$409.365 \end{aligned}$$

$$\begin{aligned} E\left(X^2\right) = E\left(E\left(X^2|M\right)\right) &= \int_0^\infty E\left(X^2|M=m\right)\lambda e^{-\lambda m} dm \\ &= \int_{100}^\infty E\left((M-100)^2|M=m\right)\frac{1}{500}e^{-m/500} dm \\ &= \int_{100}^\infty (m-100)^2\frac{1}{500}e^{-m/500} dm \\ &= 500000e^{-1/5} = 409365 \end{aligned}$$

$$\begin{aligned} SD(X) &= \sqrt{\text{Var}(X)} = \sqrt{E\left(X^2\right) - E(X)^2} \\ &= \sqrt{409365 - (409.365)^2} = \$491.72 \end{aligned}$$

Δ Theorem

Properties of Conditional Variance 1.

$$\mathrm{Var}(Y|X=x) = E\left(Y^2|X=x\right) - (E(Y|X=x))^2$$

2. For constants a and b

$$\mathrm{Var}(aY+b|X=x) = a^2\,\mathrm{Var}(Y|X=x)$$

Δ Theorem

Law of Total Variance

$$\mathrm{Var}(Y) = E(\mathrm{Var}(Y|X)) + \mathrm{Var}(E(Y|X))$$

Δ Theorem

Random Sums of Random Variables

Let $X_1, X_2, \ldots$ be an i.d. sequence of random variables with common mean μ_X and variance σ_X^2. Let N be a positive, integer-valued random variable independent of the X_i with $E(N) = \mu_N$ and $Var(N) = \sigma_N^2$. Let $T = \sum_{i=1}^{N} X_i$. Then,

$$E(T) = \mu_X \mu_N \text{ and } \mathrm{Var}(T) = \sigma_X^2 \mu_N + \sigma_N^2 \mu_X^2$$

· Example

Let $X_1, X_2, \ldots$ be an i.i.d. sequence of random variables with common mean μ. Let $S_n = X_1 + \cdots + X_n$, for $n \geq 1$

(a) Find $E\left(S_m|S_n\right)$, for $m \leq n$

(b) Find $E\left(S_m|S_n\right)$ for $m > n$

Solution:For $m > n$,

$$\begin{aligned}
E\left(S_m|S_n\right) &= E\left(S_n + X_{n+1} + \cdots + X_m|S_n\right) \\
&= E\left(S_n|S_n\right) + E\left(X_{n+1} + \cdots + X_m|S_n\right) \\
&= S_n + \sum_{i=n+1}^{m} E\left(X_i|S_n\right) = S_n + \sum_{i=n+1}^{m} E\left(X_i\right) \\
&= S_n + (m-n)\mu
\end{aligned}$$

For $m < n$,

$$\begin{aligned}
E\left(S_m|S_n\right) &= S_n - E\left(X_{n+1} + \cdots + X_m|S_n\right) \\
&= S_n - \sum_{i=m+1}^{n} E\left(X_i|S_n\right) = S_n - \sum_{i=m+1}^{n} E\left(X_i\right) \\
&= S_n - (n-m)\mu
\end{aligned}$$

3 Markov Chains: First Steps

△ Definition

Markov Chain Let S be a discrete set. A Markov chain is a sequence of random variables $X_0, X_1, \ldots$ taking values in S with the property that

$$\begin{aligned} & P\left(X_{n+1}=j \mid X_0=x_0, \ldots, X_{n-1}=x_{n-1}, X_n=i\right) \\ & \quad=P\left(X_{n+1}=j \mid X_n=i\right) \end{aligned}$$

for all $x_0, \ldots, x_{n-1}, i, j \in \mathcal{S}$, and $n \geq 0$. The set S is the state space of the Markov chain.

A Markov chain is time-homogeneous if the probabilities in Equation above do not depend on n. That is,

$$P\left(X_{n+1}=j \mid X_n=i\right)=P\left(X_1=j \mid X_0=i\right)$$

This probability can be arranged in a transition matrix/Markov matrix, P, so that the ijth entry is $P_{ij}=P\left(X_1=j \mid X_0=i\right)$.If the state space has k elements, then the transition matrix is a square $k \times k$ matrix. If the state space is countably infinite, the transition matrix is infinite. And

$$\sum_j P_{ij}=\sum_j P\left(X_1=j \mid X_0=i\right)=\sum_j \frac{P\left(X_1=j, X_0=i\right)}{P\left(X_0=i\right)}=\frac{P\left(X_0=i\right)}{P\left(X_0=i\right)}=1$$

A nonnegative matrix whose rows sum to 1 is called a stochastic matrix.

△ Definition

Stochastic Matrix A stochastic matrix is a square matrix P, which satisfies 1. $P_{ij} \geq 0$ for all i, j 2. For each row i,

$$\sum_j P_{ij}=1$$

· Example

An independent and identically distributed sequence of random variables is trivially a Markov chain. Assume that $X_0, X_1, \ldots$ is an i.i.d. sequence that takes values in

$\{1,\ldots,k\}$ with

$$P\left(X_n = j\right) = p_j, \text{ for } j = 1,\ldots,k, \text{ and } n \geq 0$$

where $p_1 + \cdots + p_k = 1$. By independence

$$P\left(X_1 = j | X_0 = i\right) = P\left(X_1 = j\right) = p_j$$

The transition matrix P is

$$\begin{pmatrix} p_1 & p_2 & \cdots & p_k \\ p_1 & p_2 & \cdots & p_k \\ \vdots & \vdots & & \vdots \\ p_1 & p_2 & \cdots & p_k \end{pmatrix}$$

· **Example**

Transition matrix for Gambler's ruin problem is:

$$P_{ij} = \begin{cases} p, & \text{if } j = i+1, \quad 0 < i < n \\ 1-p, & \text{if } j = i-1, \quad 0 < i < n \\ 1, & \text{if } i = j = 0, \text{ or } i = j = n \\ 0, & \text{otherwise.} \end{cases}$$

Gambler's ruin is an example of simple random walk with absorbing boundaries. since $P_{00} = P_{nn} = 1$, when the chain reaches 0 or n, it stays there forever.

Δ Theorem

n-Step Transition Matrix

Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix P. The matrix P^n is the n -step transition matrix of the chain. For $n \geq 0$,

$$P_{ij}^n = P\left(X_n = j | X_0 = i\right), \text{ for all } i, j.$$

3.1 Chapman–Kolmogorov Relationship

For $m, n \geq 0$, the matrix identity $P^{m+n} = P^m P^n$ gives

$$P_{ij}^{m+n} = \sum_k P_{ik}^m P_{kj}^n, \text{ for all } i, j$$

By time-homogeneity, this gives

$$\begin{aligned} P\left(X_{n+m} = j | X_0 = i\right) &= \sum_k P\left(X_m = k | X_0 = i\right) P\left(X_n = j | X_0 = k\right) \\ &= \sum_k P\left(X_m = k | X_0 = i\right) P\left(X_{m+n} = j | X_m = k\right) \end{aligned}$$

The probabilistic interpretation is that transitioning from i to j in $m+n$ steps is equivalent to transitioning from i to some state k in m steps and then moving from that state to j in the remaining n steps. This is known as the Chapman-Kolmogorov relationship.

∆ Theorem

Distribution of X_{n}

Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix P and initial distribution α. For all $n \geq 0$, the distribution of X_n is αP^n. That is,

$$P\left(X_n = j\right) = \left(\alpha P^n\right)_j, \text{ for all } j$$

3.2 Most Recent Past Property

Let $X_0, X_1, \ldots$ be a Markov chain. Then, for all $m < n$

$$\begin{aligned} &P\left(X_{n+1} = j | X_0 = i_0, \ldots, X_{n-m-1} = i_{n-m-1}, X_{n-m} = i\right) \\ &= P\left(X_{n+1} = j | X_{n-m} = i\right) \\ &= P\left(X_{m+1} = j | X_0 = i\right) = P_{ij}^{m+1} \end{aligned}$$

for all $i, j, i_0, \ldots, i_{n-m-1}$, and $n \geq 0$.

3.3 Joint Distribution

The initial distribution and transition matrix give a complete probabilistic description of a Markov chain.

Consider an arbitrary joint probability:

$$P\left(X_5 = i, X_6 = j, X_9 = k, X_{17} = l\right), \text{ for some states } i, j, k, l$$

With conditional probability, the Markov property, and time-homogeneity, we have:

$$\begin{aligned}
&P\left(X_5 = i, X_6 = j, X_9 = k, X_{17} = l\right) \\
&\qquad = P\left(X_{17} = l | X_5 = i, X_6 = j, X_9 = k\right) P\left(X_9 = k | X_5 = i, X_6 = j\right) \\
&\qquad\quad \times P\left(X_6 = j | X_5 = i\right) P\left(X_5 = i\right) \\
&\qquad = P\left(X_{17} = l | X_9 = k\right) P\left(X_9 = k | X_6 = j\right) P\left(X_6 = j | X_5 = i\right) P\left(X_5 = i\right) \\
&\qquad = P\left(X_8 = l | X_0 = k\right) P\left(X_3 = k | X_0 = j\right) P\left(X_1 = j | X_0 = i\right) P\left(X_5 = i\right) \\
&\qquad = P_{kl}^8 P_{jk}^3 P_{ij} \left(\alpha P^5\right)_i
\end{aligned}$$

In general, we have

△ Theorem

Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix P and initial distribution α. For all $0 \leq n_1 < n_2 < \cdots < n_{k-1} < n_k$ and states $i_1, i_2, \ldots, i_{k-1}, i_k$

$$\begin{aligned}
&P\left(X_{n_1} = i_1,\ X_{n_2} = i_2, \ldots, X_{n_{k-1}} = i_{k-1}, X_{n_k} = i_k\right) \\
&\qquad = \left(\alpha P^{n_1}\right)_{i_1} \left(P_2^{n_2 - n_1}\right)_{i_1 i_2} \cdots \left(P^{n_k - n_{k-1}}\right)_{i_{k-1} i_k}
\end{aligned}$$

3.4 Long-Term Behavior

· **Example**

Changes in the distribution of wetlands in Yinchuan Plain, China are studied in Zhang et al. (2011). Wetlands are considered among the most important ecosystems on earth.A Markov model is developed to track yearly changes in wetland type. Based on imaging and satellite data from 1991, 1999, and 2006, researchers measured annual distributions of wetland type throughout the region and estimated the Markov transition matrix

	River	Lake	Pond	Paddy	Non
River	0.342	0.005	0.001	0.020	0.632
Lake	0.001	0.252	0.107	0.005	0.635
Pond	0.000	0.043	0.508	0.015	0.434
Paddy	0.001	0.002	0.004	0.665	0.328
Non	0.007	0.007	0.007	0.025	0.954

The state Non refers to nonwetland regions. Based on their model, the scientists predict that "The wetland distribution will essentially be in a steady state in Yinchuan Plain in approximately 100 years."

3.5 Simulation

A Markov chain can be simulated from an initial distribution and transition matrix. To simulate a Markov sequence $X_0, X_1, \ldots$, simulate each random variable sequentially conditional on the outcome of the previous variable. That is, first simulate X_0 according to the initial distribution. If $X_0 = i$, then simulate X_1 from the i th row of the transition matrix. If $X_1 = j$, then simulate X_2 from the j th row of the transition matrix, and so on.

Algorithm for Simulating a Markov Chain Input: (i) initial distribution α, (ii) transition matrix P, (iii) number of steps n. Output: $X_0, X_1, \ldots, X_n$

```
Algorithm:
Generate X_0 according to alpha
FOR i=1, ..., n
    Assume that X_{i-1}=j
    Set p=j th row of P
    Generate X_{i} according to p
END FOR
```

4 Markov Chains for Long Term

4.1 Limiting distribution

Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix P. A limiting distribution for the Markov chain is a probability distribution λ with the property that for all i and j,

$$\lim_{n\to\infty} P_{ij}^n = \lambda_j \ .$$

By the uniqueness of limits, if a Markov chain has a limiting distribution, then that distribution is unique.

For two stage Markov chain, we have:

$$P^n = \frac{1}{p+q}\begin{pmatrix} q+p(1-p-q)^n & p-p(1-p-q)^n \\ q-q(1-p-q)^n & p+q(1-p-q)^n \end{pmatrix}$$

If p and q are not both 0, nor both 1, then $|1-p-q|<1$ and

$$\lim_{n\to\infty} P^n = \frac{1}{p+q}\begin{pmatrix} q & p \\ q & p \end{pmatrix}$$

The limiting distribution is

$$\lambda = \left(\frac{q}{p+q}, \frac{p}{p+q}\right)$$

The limiting distribution gives the long-term probability that a Markov chain hits each state. It can also be interpreted as the long-term proportion of time that the chain visits each state.

4.2 Stationary distribution

Let $X_0, X_1, \ldots$ be a Markov chain with transition matrix P. A stationary distribution is a probability distribution π, which satisfies

$$\pi = \pi P$$

$$\pi_j = \textstyle\sum_i \pi_i P_{ij}, \text{ for all } j$$

If the initial distribution is a stationary distribution, then $X_0, X_1, X_2, \ldots$ is a sequence of identically distributed random variables.

The fact that the stationary chain is a sequence of identically distributed random variables does not mean that the random variables are independent. On the contrary, the dependency structure between successive random variables in a Markov chain is governed by the transition matrix, regardless of the initial distribution

· Lemma

Assume that π is the limiting distribution of a Markov chain with transition matrix P. Then, π is a stationary distribution. the converse is not true—stationary distributions are not necessarily limiting distributions

A matrix M is said to be positive if all the entries of M are positive. We write $M > 0$. Similarly, write $x > 0$ for a vector x with all positive entries.

△ Definition

A transition matrix P is said to be regular if some power of P is positive. That is, $P^n > 0$, for some $n \geq 1$.

△ Theorem

A Markov chain whose transition matrix P is regular has a limiting distribution, which is the unique, positive, stationary distribution of the chain. That is, there exists a unique probability vector $\pi > 0$, such that

$$\lim_{n\to\infty} P_{ij}^n = \pi_j$$

$$\sum_i \pi_i P_{ij} = \pi_j$$

Equivalently, there exists a positive stochastic matrix Π such that

$$\lim_{n\to\infty} P^n = \Pi$$

where $\boldsymbol{\Pi}$has equal rows with common row $\boldsymbol{\pi}$, and $\boldsymbol{\pi}$ is the unique probability vector, which satisfies

$$\boldsymbol{\pi P} = \boldsymbol{\pi}$$

Here is one way to tell if a stochastic matrix is not regular. If for some power n, all the 0s in P^n appear in the same locations as all the 0s in P^{n+1}, then they will appear in the same locations for all higher powers, and the matrix is not regular.

4.2.1 Stationary Distribution for Random Walk on a Weighted Graph

Let G be a weighted graph with edge weight function $w(i,j)$. For random walk on G, the stationary distribution π is proportional to the sum of the edge weights incident to each vertex. That is,

$$\pi_v = \frac{w(v)}{\sum_z w(z)}, \text{ for all vertices } v$$
$$w(v) = \sum_{z\sim v} w(v,z)$$

· **Example**

Find the stationary distribution for random walk on the hypercube.

Solution The k -hypercube graph, as described in Example 2.8, has 2^k vertices. Each vertex has degree k. The sum of the vertex degrees is $k2^k$, and the stationary distribution π is given by

$$\pi_v = \frac{k}{k2^k} = \frac{1}{2^k}, \text{ for all } v$$

The hypercube is an example of a regular graph. A graph is regular if all the vertex degrees are the same.

4.2.2 The Eigenvalue Connection

Recall that an eigenvector of M is a column vector x^T such that $Mx^T = \lambda x^T$, for some scalar λ. We call such a vector a right eigenvector of M. A left eigenvector of M is a row vector y, which satisfies $yM = \mu y$, for some scalar μ. A left eigenvector of M is simply a right eigenvector of M^T.

A matrix and its transpose have the same set of eigenvalues, with possibly different eigenvectors. It follows that $\lambda = 1$ is an eigenvalue of P^T with some corresponding right eigenvector y^T. Equivalently, y is a left eigenvector of P. That is, there exists a row vector y such that $yP = y$. If a multiple lof y can be normalized so that its components are non-negative and sumto 1, then this gives astationary distribution. However some of the entries of y might be negative, or complex-valued, and the vector might not be able to be normalized to give a probability distribution.

4.3 CAN YOU FIND THEWAY TO STATE a?

Say that state j is accessible from state i, if $P_{ij}^n > 0$, for some $n \geq 0$. That is, there is positive probability of reaching j from i in a finite number of steps. States i and j communicate if i is accessible from j and j is accessible from i.

Since communication is an equivalence relation the state space can be partitioned into equivalence classes, called communication classes. That is, the state space can be divided into disjoint subsets, each of whose states communicate with each other but do not communicate with any states outside their class.

△ Definition

Irreducibility

A Markov chain is called irreducible if it has exactly one communication class. That is, all states communicate with each other.

4.3.1 Recurrence and Transience

Given a Markov chain $X_0, X_1, \ldots$, let $T_j = \min\{n > 0 : X_n = j\}$ be the first passage time to state j. If $X_n \neq j$, for all $n > 0$, set $T_j = \infty$. Let

$$f_j = P(T_j < \infty | X_0 = j)$$

be the probability that the chain started in j eventually returns to j.

△ Definition

Recurrent and Transient States

State j is said to be recurrent if the Markov chain started in jeventually revisits j. That is, $f_j = 1$. State j is said to be transient if there is positive probability that the Markov chain started in j never returns to j. That is, $f_j < 1$.

For the chain started in i, let

$$I_n = \begin{cases} 1, & \text{if} \quad X_n = j \\ 0, & \text{otherwise} \end{cases}$$

for $n \geq 0$. Then, $\sum_{n=0}^{\infty} I_n$ is the number of visits to j. The expected number of visits to j is

$$E\left(\sum_{n=0}^{\infty} I_n\right) = \sum_{n=0}^{\infty} E\left(I_n\right) = \sum_{n=0}^{\infty} P\left(X_n = j | X_0 = i\right) = \sum_{n=0}^{\infty} P_{ij}^n$$

From j, the chain will revisit j again, with probability 1, and so on. It follows that j will be visited infinitely many times, and

$$\sum_{n=0}^{\infty} P_{jj}^n = \infty$$

$$\sum_{n=0}^{\infty} P_{ij}^n = \infty$$

If the state j is transient, we will have:

$$\sum_{n=0}^{\infty} P_{jj}^n < \infty$$

$$\sum_{n=0}^{\infty} P_{ij}^n < \infty$$

Which leads to:

$$\lim_{n \to \infty} P_{ij}^n = 0$$

△ Theorem

The states of a communication class are either all recurrent or all transient.

On the other hand, if one state is transient, the other states must be transient.

△ Theorem

For a finite irreducible Markov chain, all states are recurrent.

4.3.2 Canonical decomposition

A set of states C is said to be closed if no state outside of C is accessible from any state in C. If C is closed, then $P_{ij} = 0$ for all $i \in C$ and $j \notin C$.

△ Theorem

A communication class is closed if it consists of all recurrent states. A finite communication class is closed only if it consist of all recurrent states.

The state space S of a finite Markov chain can be partitioned into transient and recurrent states as $S = T \cup R_1 \cup \cdots \cup R_m$, where T is the set of all transient states and the R_i are closed communication classes of recurrent states. This is called the canonical decomposition. The computation of many quantities associated with Markov chains can be simplified by this decomposition.

Given a canonical decomposition, the state space can be reordered so that the Markov transition matrix has the block matrix form

$$\begin{pmatrix} * & * & \cdots & * \\ 0 & P_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & P_m \end{pmatrix}$$

Where first several lines represents probability distribution of transient points. Each submatrix $\boldsymbol{P}_1, \ldots, \boldsymbol{P}_m$ is a square stochastic matrix corresponding to a closed recurrent communication class. By itself, each of these matrices is the matrix of an irreducible Markov chain with a restricted state space.

4.3.3 Irreducible Markov Chains

△ Theorem

Limit Theorem for Finite Irreducible Markov Chains

Assume that $X_0, X_1, \ldots$ is a finite irreducible Markov chain. For each state j, let $\mu_j = E(T_j | X_0 = j)$ be the expected return time to j. Then, μ_j is finite, and there exists a unique, positive stationary distribution π such that

$$\pi_j = \frac{1}{\mu_j}, \text{ for all } j$$

Furthermore, for all states i,

$$\pi_j = \lim_{n\to\infty} \frac{1}{n} \sum_{m=0}^{n-1} P_{ij}^m$$

The theorem does not assert that π is a limiting distribution.

A recurrent state j is called positive recurrent if $E(T_j | X_0 = j) < \infty$, and null recurrent if $E(T_j | X_0 = j) = \infty$. Thus, the theorem holds for irreducible Markov chains for which all states are positive recurrent.

· **Example**

Consider a Markov chain with transition matrix

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

for states a, b, and c.From state a, find the expected return time $E(T_a | X_0 = a\ )$ using first-step analysis.

Solution Let $e_x = E(T_a | X_0 = x)$, for $x = a, b, c$. Thus, e_a is the desired expected return time, and e_b and e_c are the expected first passage times to a for the chain started in b and c, respectively.

For the chain started in a, the next state is b, with probability 1. From b, the further evolution of the chain behaves as if the original chain started at b. Thus,

$$e_a = 1 + e_b$$

From b, the chain either hits a, with probability $1/2$, or moves to c, where the chain behaves as if the original chain started at c. It follows that

$$e_b = \frac{1}{2} + \frac{1}{2}(1 + e_c)$$

Similarly,

$$e_c = \frac{1}{3} + \frac{1}{3}(1 + e_b) + \frac{1}{3}(1 + e_c)$$

Solving the three equations gives

$$e_c = \frac{8}{3}, \quad e_b = \frac{7}{3}, \quad \text{and} \quad e_a = \frac{10}{3}$$

The desired expected return time is $10/3$.

4.4 Periodicity

An example of a finite irreducible Markov chain with no limiting distribution is random walk on the n -cyule, when. The graph is regular (all vertex degrees are the same and the unique stationary distribution is uniform. But there is no limiting distribution since the chain flip-flops back and forth even and and states. The chain's position after n steps depends on the parity of the initial state. It is precisely the finite irreducible Markov chains that do not exhibit this type of periodic behavior, which have limiting distributions.

∆ Definition

Periodicity

For a Markov chain with transition matrix P, the period of state i, denoted $d(i)$, is the greatest common divisor of the set of possible return times to i. That is,

$$d(i) = \gcd \{n > 0 : P_{ii}^n > 0\}.$$

If $d(i) = 1$, state i is said to be aperiodic. If the set of return times is empty set $d(i) = +\infty$.

· **Lemma**

The states of a communication class all have the same period.

Δ **Definition**

Periodic, Aperiodic Markov Chain

A Markov chain is called periodic if it is irreducible and all states have period greater than 1.

A Markov chain is called aperiodic if it is irreducible and all states have period equal to 1.

4.5 Ergodic Markov Chain

A Markov chain is called ergodic if it is irreducible, aperiodic, and all states have finite expected return times.

Δ **Theorem**

Limiting theorem for ergodic Markov chain

Let $X_0, X_1, \ldots$ be an ergodic Markov chain. There exists a unique, positive, stationary distribution π, which is the limiting distribution of the chain. That is,

$$\pi_j = \lim_{n\to\infty} P_{ij}^n, \text{ for all } i, j$$

· **Example**

Google's **PageRank** search algorithm is based on the random surfer model, which is a random walk on the webgraph. For this graph, each vertex represents an internet page. A directed edge connects i to j if there is a hypertext link from page i to page j. When the random surfer is at page i, they move to a new page by choosing from the available links on i uniformly at random.

Let the following network matrix describe the network

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

Note that N is not a stochastic matrix, as page g has no out-link. Row g consists of all 0s. Thus, the "g" is called "dangling node".

So, how to solve dangling node problem and the problem of potentially getting stuck in small subgraphs of the webgraph?

Solution: the fix for dangling nodes is to assume that when the random surfer reaches such a page they jump to a new page in the network uniformly at random. A new matrix Q is obtained where each row in the network matrix N corresponding to a dangling node is changed to one in which all entries are $1/k$. The matrix Q is a stochastic matrix.

For the problem of potentially getting stuck in small subgraphs of the webgraph, the solution proposed in the original paper by Brin and Page (1998) was to fix a damping factor $0 < p < 1$ for modifying the Q matrix. In their model, from a given page the random surfer, with probability $1 - p$, decides to not follow any links on the page and instead navigate to a new rage on the network. On the other hand, with probability p, they follow the links on the page as usual. This defines the PageRank transition matrix

$$P = pQ + (1 - p)A$$

where A is a $k \times k$ matrix all of whose entries are $1/k$. The damping factor used by Google was originally set to $p = 0.85$.

4.6 Time reversibility

△ Definition

An irreducible Markov chain with transition matrix P and stationary distribution π is reversible, or time reversible, if

$$\pi_i P_{ij} = \pi_j P_{ji}, \text{ for all } i, j$$

Equations above are called the detailed balance equations. They say that for a chain in stationarity,

$$P(X_0 = i, X_1 = j) = P(X_0 = j, X_1 = i), \text{ for all } i, j$$

That is, the frequency of transitions from i to j is equal to the frequency of transitions from j to i. if a stationary Markov chain is reversible then:

$$P(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n) = P(X_0 = i_n, X_1 = i_{n-1}, \ldots, X_n = i_0)$$

for all $i_0, i_1, \ldots, i_n$

4.6.1 Reversible Markov Chains and Random Walk on Weighted Graphs

Random walk on a weighted graph is time reversible. In fact, every reversible Markov chain can be considered as a random walk on a weighted graph.

Given a reversible Markov chain with transition matrix P and stationary distribution π, construct a weighted graph on the state space by assigning edge weights $w(i,j) = \pi_i P_{ij}$. With these choice of weights, random walk on the weighted graph moves from i to j with probability

$$\frac{w(i,j)}{\sum_v w(i,v)} = \frac{\pi_i P_{ij}}{\sum_v \pi_i P_{iv}} = \frac{\pi_i P_{ij}}{\pi_i} = P_{ij}$$

Conversely, given a weighted graph with edge weight function $w(i,j)$, the transition matrix of the corresponding Markov chain is obtained by letting

$$P_{ij} = \frac{w(i,j)}{\sum_v w(i,v)}$$

where the sum is over all neighbors of i. The stationary distribution is

$$\pi_i = \frac{\sum_y w(i,y)}{\sum_x \sum_y w(x,y)}$$

· **Lemma**

Let P be the transition matrix of a Markov chain. If x is a probability distribution which satisfies

$$x_i P_{ij} = x_j P_{ji}, \text{ for all } i, j$$

then x is the stationary distribution, and the Markov chain is reversible.

4.7 Absorbing Chain

△ **Definition**

State i is an absorbing state if $P_{ii} = 1$. A Markov chain is called an absorbing chain if it has at least one absorbing state.

Consider an absorbing Markov chain on k states for which t states are transient and $k-t$ states are absorbing. The states can be reordered, as in the canonical decomposition, with the transition matrix written in block matrix form

$$P = \left(\begin{array}{c|c} Q & R \\ \hline 0 & I \end{array}\right)$$

where Q is a $t \times t$ matrix, R is a $t \times (k-t)$ matrix, $\mathbf{0}$ is a $(k-t) \times t$ matrix of $0s$, and I is the $(k-t) \times (k-t)$ identity matrix.

In general,

$$P^n = \left(\begin{array}{c|c} Q^n & (I + Q + \cdots + Q^{n-1})\, R \\ \hline 0 & I \end{array}\right), \text{ for } n \geq 1$$

· **Lemma**

Let A be a square matrix with the property that $A^n \to 0$, as $n \to \infty$. Then,

$$\sum_{n=0}^{\infty} A^n = (I - A)^{-1}$$

With this lemma, now we have:

$$\lim_{n\to\infty} P^n = \left(\frac{\lim_{n\to\infty} Q^n}{0} \middle| \frac{\lim_{n\to\infty} (I + Q + \cdots + Q^{n-1})\, R}{I}\right)$$

$$= \left(\begin{array}{c|c} 0 & (I-Q)^{-1}R \\ \hline 0 & I \end{array}\right)$$

Consider the interpretation of the limiting submatrix $(I - Q)^{-1}R$. The matrix is indexed by transient rows and absorbing columns. The ijth entry is the long-term probability that the chain started in transient state i is absorbed in state j.

For absorbing Markov chains, the matrix $(\boldsymbol{I} - \boldsymbol{Q})^{-1}$ is called thefundamental matrix.

△ Theorem

Consider an absorbing Markov chain with t transient states. Let F be a $t \times t$ matrix indexed by transient states, where F_{ij} is the expected number of visits to j given that the chain starts in i. Then,

$$\boldsymbol{F} = (\boldsymbol{I} - \boldsymbol{Q})^{-1}$$

· **Lemma**

1. (Absorption probability) The probability that from transient state i the chain is absorbed in state j is $(FR)_{ij}$.

2. (Absorption time) The expected number of steps from transient state i until the chain is absorbed in some absorbing state is $(F1)_i$.

4.7.1 Expected Hitting Times for Irreducible Chains

For an irreducible Markov chain, first hitting times can be analyzed as absorption times for a suitably modified chain. In particular, assume that $\mathcal{P}$ is the transition matrix of an irreducible Markov chain. To find the expected time until state i is first hit, consider a new chain in which i is an absorbing state. The transition matrix $\widetilde{P}$ for the new chain is gotten by zeroing out the i th row of the P matrix and setting $\widetilde{P}_{ii} = 1$. The resulting Q matrix is obtained from $\widetilde{P}$ by deleting the i th row and the i th column of P. The time that the original P -chain first hits i is equal to the time that the modified $\widetilde{P}$ -chain is absorbed in i.

4.7.2 Patterns in sequence

· **Example**

A biased coin comes up heads, with probability 2/3, and tails, with probability 1/3. The coin is repeatedly flipped. How many flips are needed, on average, until the pattern HTHTH first appears?

An absorbing Markov chain is constructed with transition matrix:

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

The last row is for pattern of HTHTH, the first row is for $\emptyset$.The fundamental matrix is

$$\begin{pmatrix} 2/3 & -2/3 & 0 & 0 & 0 \\ 0 & 1/3 & -1/3 & 0 & 0 \\ -1/3 & 0 & 1 & -2/3 & 0 \\ 0 & -2/3 & 0 & 1 & -1/3 \\ -1/3 & 0 & 0 & 0 & 1 \end{pmatrix}^{-1}$$

$$
= \begin{pmatrix} 45/8 & 81/4 & 27/4 & 9/2 & 3/2 \\ 33/8 & 81/4 & 27/4 & 9/2 & 3/2 \\ 33/8 & 69/4 & 27/4 & 9/2 & 3/2 \\ 27/8 & 63/4 & 21/4 & 9/2 & 3/2 \\ 15/8 & 27/4 & 9/4 & 3/2 & 3/2 \end{pmatrix}
$$

The sum of the first row of the fundamental matrix is

$$
\frac{45}{8} + \frac{81}{4} + \frac{27}{4} + \frac{9}{2} + \frac{3}{2} = \frac{309}{8} = 38.625
$$

It takes, on average, 38.625 flips before HTHTH first appears.

5 Branching Process

Applications of branching process include nuclear chain reactions and the spread of computer software viruses. Their original motivation was to study the extinction of family surnames, an issue of concern to the Victorian aristocracy in 19th century Britain.

> · **Lemma**
>
> In a branching process, all nonzero states are transient

since all nonzero states are transient and the chain has infinite state space, there are two possibilities for the long-term evolution of the process: either it gets absorbed in state 0, that is, the population eventually goes extinct, or the population grows without bound.

5.1 Mean generation size

In a branching process, the size of the n th generation is the sum of the total offspring of the individuals of the previous generation. That is,

$$Z_n = \sum_{i=1}^{Z_{n-1}} X_i$$

Let $\mu = \sum_{k=0}^{\infty} k a_k$ be the mean of the offspring distribution. To find the mean of the size of the n th generation $E(Z_n)$, condition on Z_{n-1}. By the law of total expectation,

$$\begin{aligned} E(Z_n) &= \sum_{k=0}^{\infty} E(Z_n | Z_{n-1} = k) P(Z_{n-1} = k) \\ &= \sum_{k=0}^{\infty} E\left(\sum_{i=1}^{Z_{n-1}} X_i | Z_{n-1} = k\right) P(Z_{n-1} = k) \end{aligned}$$

$$\begin{aligned} &= \textstyle\sum_{k=0}^{\infty} E\left(\sum_{i=1}^{k} X_i | Z_{n-1} = k\right) P(Z_{n-1} = k) \\ &= \textstyle\sum_{k=0}^{\infty} E\left(\sum_{i=1}^{k} X_i\right) P(Z_{n-1} = k) \\ &= \textstyle\sum_{k=0}^{\infty} k\mu P(Z_{n-1} = k) = \mu E(Z_{n-1}) \end{aligned}$$

Since $E(Z_0) = 1$, we have

$$E(Z_n) = \mu E(Z_{n-1}) = \mu^2 E(Z_{n-2}) = \cdots = \mu^n E(Z_0) = \mu^n, \text{ for } n \geq 0$$

For the long-term expected generation size,

$$\lim_{n\to\infty} E\left(Z_n\right) = \lim_{n\to\infty} \mu^n = \begin{cases} 0, & \text{if } \mu < 1 \\ 1, & \text{if } \mu = 1 \\ \infty, & \text{if } \mu > 1 \end{cases}$$

A branching process is said to be subcritical if $\mu < 1$, critical if $\mu = 1$, and supercritical if $\mu > 1$. For a subcritical branching process, mean generation size The variance of generation size is:

$$\begin{aligned} \operatorname{Var}\left(Z_n\right) &= \operatorname{Var}\left(\mu Z_{n-1}\right) + E\left(\sigma^2 Z_{n-1}\right) \\ &= \mu^2 \operatorname{Var}\left(Z_{n-1}\right) + \sigma^2 \mu^{n-1}, \text{ for } n \geq 1 \end{aligned}$$

5.2 Probability Generating Function

For a discrete random variable X taking values in $\{0, 1, \ldots\}$, the probability generating function of X is the function

$$\begin{aligned} G(s) = E\left(s^X\right) &= \sum_{k=0}^{\infty} s^k P(X = k) \\ &= P(X = 0) + sP(X = 1) + s^2 P(X = 2) + \cdots \end{aligned}$$

The function is a power series whose coefficients are probabilities. Observe that $G(1) = 1$. The series converges absolutely for $|s| \leq 1$. To emphasize the underlying random variable X, we may write $G(s) = G_X(s)$.

Probabilities for X can be obtained from the generating function by successive differentiation. We have that

$$\begin{aligned} G(0) &= P(X = 0) \\ G'(0) &= \left.\sum_{k=1}^{\infty} k s^{k-1} P(X = k)\right|_{s=0} = P(X = 1) \\ G''(0) &= \left.\sum_{k=2}^{\infty} k(k-1) s^{k-2} P(X = k)\right|_{s=0} = 2P(X = 2) \end{aligned}$$

In general, we have

$$G^{(j)}(0) = \sum_{k=j}^{\infty} k(k-1)\cdots(k-j+1)s^{k-j}P(X=j)\bigg|_{s=0} = j!P(X=j)$$

5.2.1 Sums of Independent Random Variables

Assume that $X_1, \dots, X_n$ are independent. Let $Z = X_1 + \cdots + X_n$. The probability generating function of Z is

$$\begin{aligned} G_Z(s) &= E\left(s^Z\right) = E\left(s^{X_1+\cdots+X_n}\right) \\ &= E\left(\prod_{k=1}^{n} s^{X_k}\right) = \prod_{k=1}^{n} E\left(s^{X_k}\right) \\ &= G_{X_1}(s)\cdots G_{X_n}(s) \end{aligned}$$

If X_i are also identically distributed, we have

$$G_Z(s) = G_{X_1}(s)\cdots G_{X_n}(s) = [G_X(s)]^n$$

5.2.2 Moments

The probability generating function of X can be used to find the mean, variance, and higher moments of X. Observe that

$$G'(1) = \left|E\left(Xs^{X-1}\right)\right|_{s=1} = E(X)$$

$$\begin{aligned} \mathrm{Var}(X) &= E\left(X^2\right) - E(X)^2 = \left(E\left(X^2\right) - E(X)\right) + E(X) - E(X)^2 \\ &= G''(1) + G'(1) - G'(1)^2 \end{aligned}$$

5.3 Extinction Probability

For $n \geq 0$, let

$$G_n(s) = \sum_{k=0}^{\infty} s^k P\left(Z_n = k\right)$$

Let

$$G(s) = \sum_{k=0}^{\infty} s^k a_k$$

be the generating function of the offspring distribution. We have

$$G_n(s) = E\left(s^{Z_n}\right) = E\left(s^{\sum_{k=1}^{Z_{n-1}} X_k}\right) = E\left(E\left(s^{\sum_{k=1}^{Z_{n-1}} X_k | Z_{n-1}}\right)\right)$$

From the independence of Z_{n-1} and X_k

$$\begin{aligned} E\left(s^{\sum_{k=1}^{z_{k-1}} X_k} | z_{n-1} = z\right) &= E\left(s^{\sum_{i=1}^{z} x_k} | z_{n-1} = z\right) \\ &= E\left(s^{\sum_{i=1}^{z} x_k}\right) = E\left(\prod_{k=1}^{z} s^{x_k}\right) \\ &= \prod_{k=1}^{z} E\left(s^{x_k}\right) = [G(s)]^z \end{aligned}$$

For all z, we get

$$E\left(s^{\sum_{k=1}^{Z_{n-1}} X_k} | Z_{n-1}\right) = [G(s)]^{Z_{n-1}}$$

Taking expectations,

$$G_n(s) = E\left(G(s)^{Z_{n-1}}\right) = G_{n-1}(G(s)), \quad \text{for } n \geq 1$$

Observe that $G_0(s) = s$, and $G_1(s) = G_0(G(s)) = G(s)$. From the latter we see that the distribution of Z_1 is the offspring distribution a. Continuing,

$$G_2(s) = G_1(G(s)) = G(G(s)) = G\left(G_1(s)\right)$$

In general,

$$G_n(s) = G_{n-1}(G(s)) = \underbrace{G(\cdots G(G(s)) \cdots)}_{n\text{ -fold}} = G\left(G_{n-1}(s)\right)$$

Δ Theorem

Extinction probability

Given a branching process, let G be the probability generating function of the offspring distribution. Then, the probability of eventual extinction is the smallest positive root of the equation $s = G(s)$. If $\mu \leq 1$, that is, in the subcritical and critical cases, the extinction probability is equal to 1

· **Example**

Let **a** be an offspring distribution with generating function G. Let X be a random variable with distribution a. Let Z be a random variable whose distribution is that of X conditional on $X > 0$. That is, $P(Z = k) = P(X = k | X > 0\ )$. Find the generating function of Z in terms of G.

Solution:

$$P(Z = k) = P(X = k | X > 0) = \frac{P(X = k, X > 0)}{P(X > 0)}$$

for $k > 0$, we have

$$P(Z = k) = \frac{P(X = k)}{P(X > 0)} = \frac{P(X = k)}{1 - a_0}$$

Thus, we have:

$$G_Z(s) = \frac{1}{1 - a_0}(G(s) - a_0)$$

6 Markov Chain Monte Carlo

The MCMC algorithm constructs an ergodic Markov chain whose limiting distribution is the desired π. One then runs the chain long enough for the chain to converge, or nearly converge, to its limiting distribution, and outputs the final element or elements of the Markov sequence as a sample from π.

> △ **Theorem**
>
> **Strong Law of Large Numbers for Markov Chains** Assume that $X_0, X_1, \ldots$ is an ergodic Markov chain with stationary distribution π. Let r be a bounded, real-valued function. Let X be a random variable with distribution π. Then, with probability 1 ,
>
> $$\lim_{n \to \infty} \frac{r(X_1) + \cdots + r(X_n)}{n} = E(r(X))$$
>
> where $E(r(X)) = \sum_j r(j)\pi_j$

6.1 METROPOLIS–HASTINGS ALGORITHM

Let $\pi = (\pi_1, \pi_2, \ldots)$ be a discrete probability distribution. The algorithm constructs a reversible Markov chain $X_0, X_1, \ldots$ whose stationary distribution is π.

Let T be a transition matrix for any irreducible Markov chain with the same state space as π. It is assumed that the user knows how to sample from T. The T chain will be used as a proposal chain, generating elements of a sequence that the algorithm decides whether or not to accept.

To describe the transition mechanism for $X_0, X_1, \ldots$, assume that at time n the chain is at state i. The next step of the chain X_{n+1} is determined by a two-step procedure.

1. Choose a new state according to T. That is, choose j with probability T_{ij}. State j is called the proposal state. 2. Decide whether to accept j or not. Let

$$a(i, j) = \frac{\pi_j T_{ji}}{\pi_i T_{ij}}$$

The function a is called the acceptance function. If $a(i, j) \geq 1$, then j is accepted as the next state of the chain. If $a(i, j) < 1$, then j is accepted with probability $a(i, j)$. If j is not accepted, then i is kept as the next step of the chain.

The sequence $X_0, X_1, \ldots$ constructed by the Metropolis-Hastings algorithm is a reversible Markov chain whose stationary distribution is π.

6.1.1 Continuous state space

For a continuous state space Markov process a transition function replaces the transition matrix, where P_{ij} is the value of a conditional density function given $X_0 = i$

The Metropolis-Hastings algorithm is essentially the same as in the discrete case, with densities replacing discrete distributions. Without delving into any new theory, we present the continuous case by example.

· **Example**

Using only a uniform random number generator, simulate a standard normal random variable using MCMC.

From state s, the proposal chain moves to t, where t is uniformly distributed on $(s-1, s+1)$. Hence, the conditional density given s is constant, with

$$T_{st} = \frac{1}{2}, \text{ for } s - 1 \leq t \leq s + 1$$

The acceptance function is

$$a(s,t) = \frac{\pi_t T_{ts}}{\pi_s T_{st}} = \left(\frac{e^{-t^2/2}}{\sqrt{2\pi}}\right)(1/2) / \left(\frac{e^{-s^2/2}}{\sqrt{2\pi}}\right)(1/2) = e^{-\left(t^2 - s^2\right)/2}$$

Notice that the length of the interval for the uniform proposal distribution does not affect the acceptance function.

6.2 Gibbs Sampler

In the Gibbs sampler, the target distribution π is an m -dimensional joint density

$$\pi\left(x_1, \ldots, x_m\right)$$

A multivariate Markov chain is constructed whose limiting distribution is π, and which takes values in an m -dimensional space. The algorithm generates elements of the form.

$$\begin{aligned} & X^{(0)}, X^{(1)}, X^{(2)}, \ldots \\ & \quad=\left(X_1^{(0)}, \ldots, X_m^{(0)}\right),\left(X_1^{(1)}, \ldots, X_m^{(1)}\right),\left(X_1^{(2)}, \ldots, X_m^{(2)}\right), \ldots \end{aligned}$$

by iteratively updating each component of an m-dimensional vector conditional on the other m-1 components.

· **Example**

Consider a bivariate standard normal distribution with correlation ρ

$$f(x, y)=\frac{1}{2 \pi \sqrt{1-\rho^2}} e^{-\frac{x^2-2 \rho n+y^2}{2\left(1-\rho^2\right)}}$$

The bivariate normal distribution has the property that conditional distributions are normal. If (X, Y) has a bivariate standard normal distribution, then the conditional distribution of X given $Y=y$ is normal with mean ρy and variance $1-\rho^2$. Similarly, the conditional distribution of Y given $X=x$ is normal with mean ρx and variance $1-\rho^2$.

The Gibbs sampler is implemented to simulate (X, Y) from a bivariate standard normal distribution with correlation ρ.

Solution:

1. Initialize: $\left(x_0, y_0\right) \leftarrow(0,0)$ $m \leftarrow 1$

2. Generate x_m from the conditional distribution of X given $Y=y_{m-1}$. That is, simulate from a normal distribution with mean ρy_{m-1} and variance $1-\rho^2$

3. Generate y_m from the conditional distribution of Y given $X=x_m$. That is, simulate from a normal distribution with ρx_m and variance $1-\rho^2$.

4. $m \leftarrow m + 1$

5. Return to Step 2.

· **Example**

The following implementation of the Gibbs sampler for a three-dimensional joint distribution is a classic example based on Casella and George (1992). Random variables X, P, and N have joint density:

$$\pi(x,p,n) \propto \begin{pmatrix} n \\ x \end{pmatrix} p^x(1-p)^{n-x}\frac{4^n}{n!}$$

for $x = 0, 1, \ldots, n, 0 < p < 1, n = 0, 1, \ldots$ The p variable is continuous; x and n are discrete.

Solution: The Gibbs sampler requires being able to simulate from the conditional distributions of each component given the remaining variables. The trick to identifying these conditional distributions is to treat the two conditioning variables in the joint density function as fixed constants.

1. Initialize: $(x_0, p_0, n_0) \leftarrow (1, 0.5, 2)$

2. Generate x_m from a binomial distribution with parameters n_{m-1} and p_{m-1}

3. Generate x_m from a binomial distribution with parameters n_{m-1} and p_{m-1}

3. Generate p_m from a beta distribution with parameters $x_m + 1$ and $n_{m-1} - x_m + 1$

4. Let $n_m = z + x_m$, where z is simulated from a Poisson distribution with parameter $4\,(1 - p_m)$.

5. $m \leftarrow m + 1$

6. Return to Step 2.

The Gibbs sampler is remarkably versatile, and can be applied to a large variety of complex multidimensional problems.

6.2.1 Ising Model

Consider a graph consisting of sites (vertices), in which each site v is assigned a spin of $+1$ or -1. A configuration σ is an assignment of spins to each site. That is, $\sigma_v = \pm 1$, for all v. We will assume a square $n \times n$ grid of sites, where each site is connected to four neighbors (up, down, left, and right), except at the boundary. Thus, there are n^2 sites and 2^{n^2} possible configurations.

Associated with each configuration σ is its energy, defined as

$$E(\boldsymbol{\sigma}) = -\sum_{v \sim w} \sigma_v \sigma_w$$

The Gibbs distribution is a probability distribution on the set of configurations, defined by

$$\pi_\sigma = \frac{e^{-\beta E(\sigma)}}{\sum_\tau e^{-\beta E(\tau)}}$$

The Ising model is studied, in part, because of its phase transition properties. In two dimensions, the system undergoes a radical change of behavior at the critical temperature $1/\beta = 2/\ln(1+\sqrt{2}) \approx 2.269(\beta = 0.441)$. Above that temperature, the system appears disorganized and chaotic. Below the critical temperature, the system is magnetized, and a phase transition occurs.

Denote the sites as $1, \ldots, m$. Let σ_k be the spin at site k. Let

$$\boldsymbol{\sigma_{-k}} = (\sigma_1, \ldots, \sigma_{k-1}, \sigma_{k+1}, \ldots, \sigma_m)$$

denote the other m-1 sites of the configuration. For fixed k, write

$$\boldsymbol{\sigma}^+ = (\sigma_1, \ldots, \sigma_{k-1}, +1, \sigma_{k+1}, \ldots, \sigma_m)$$
$$\boldsymbol{\sigma}^- = (\sigma_1, \ldots, \sigma_{k-1}, -1, \sigma_{k+1}, \ldots, \sigma_m)$$

$$\begin{aligned}
P\left(\sigma_k = +1 | \sigma_{-k}\right) &= \frac{P\left(\sigma^+\right)}{P\left(\sigma_{-k}\right)} = \frac{P\left(\sigma^+\right)}{P\left(\sigma^+\right) + P\left(\sigma^-\right)} \\
&= \frac{e^{-\beta E\left(\sigma^+\right)}}{e^{-\beta E(\sigma^+)} + e^{-\beta E(\sigma^-)}} \\
&= \frac{1}{1 + e^{\beta[E(\sigma^+) - E(\sigma^-)]}}
\end{aligned}$$

Observe that

$$E\left(\boldsymbol{\sigma}^{+}\right) = -\left(\sum_{\substack{i\sim j\\ ij>k}} \sigma_i\sigma_j + \sum_{i\sim k}\sigma_i\right)$$

$$E\left(\boldsymbol{\sigma}^{-}\right) = -\left(\sum_{\substack{i\sim j\\ ij<k}} \sigma_i\sigma_j - \sum_{i\sim k}\sigma_i\right)$$

This gives

$$E\left(\sigma^{+}\right) - E\left(\sigma^{-}\right) = -2\sum_{i\sim k}\sigma_i$$

$$P\left(\sigma_k = +1|\sigma_{-k}\right) = \frac{1}{1+e^{\beta[E(\sigma^{+})-E(\sigma^{-})]}} = \frac{1}{1+e^{-2\beta\sum_{i\sim k}\sigma_i}}$$

$$P\left(\sigma_k = -1|\sigma_{-k}\right) = 1 - P\left(\sigma_k = +1|\sigma_{-k}\right)$$

7 Gentle Intro to Poisson Process

A Poisson process is a special type of counting process. Given a stream of events that arrive at random times starting at $t = 0$, let N_t denote the number of arrivals that occur by time t, that is, the number of events in $[0, t]$. For instance, N_t might be the number of text messages received up to time t.

△ Definition

A counting process $(N_t)_{t \geq 0}$ is a collection of non-negative, integer-valued random variables such that if $0 \leq s \leq t$, then $N_s \leq N_t$.

7.1 First Definition of Poisson Process

△ Definition

A Poisson process with parameter λ is a counting process $(N_t)_{t \geq 0}$ with the following properties:

1. $N_0 = 0$

2. For all $t > 0, N_t$ has a Poisson distribution with parameter λt

3. (Stationary increments) For all $s, t > 0, N_{t+s} - N_s$ has the same distribution as N_t. That is,

$$P\left(N_{t+s} - N_s = k\right) = P\left(N_t = k\right) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}, \text{ for } k = 0, 1, \ldots$$

4. (Independent increments) For $0 \leq q < r \leq s < t, N_t - N_s$ and $N_r - N_q$ are independent random variables.

since N_t has a Poisson distribution, $E\left(N_t\right) = \lambda t$. That is, we expect about λt arrivals in t time units. Thus, the rate of arrivals is $E\left(N_t\right)/t = \lambda$

· **Example**

Starting at 6 a.m., customers arrive at Martha's bakery according to a Poisson process at the rate of 30 customers per hour. Find the probability that more than 65 customers arrive between 9 and 11 a.m.

Solution: Let $t = 0$ represent 6a.m. Then, the desired probability is $P(N_5 - N_3 > 65)$. By stationary increments,

$$\begin{aligned} P(N_5 - N_3 > 65) &= P(N_2 > 65) = 1 - P(N_2 \leq 65) \\ &= 1 - \sum_{k=0}^{65} P(N_2 = k) \\ &= 1 - \sum_{k=0}^{65} \frac{e^{-30(2)}(30(2))^k}{k!} = 0.2355 \end{aligned}$$

· **Example**

Joe receives text messages starting at 10 a.m. at the rate of 10 texts per hour according to a Poisson process. Find the probability that he will receive exactly 18 texts by noon and 70 texts by 5 p.m.

Solution: The desired probability is $P(N_2 = 18, N_7 = 70)$, with time as hours. If 18 texts arrive in $[0, 2]$ and 70 texts arrive in $[0, 7]$, then there are $70 - 18 = 52$ texts in (2,7]. That is,

$$\{N_2 = 18, N_7 = 70\} = \{N_2 = 18, N_7 - N_2 = 52\}$$

The intervals $[0, 2]$ and $(2, 7]$ are disjoint, which gives

$$\begin{aligned} P(N_2 = 18, N_7 = 70) &= P(N_2 = 18, N_7 - N_2 = 52) \\ &= P(N_2 = 18)\, P(N_7 - N_2 = 52) \\ &= P(N_2 = 18)\, P(N_5 = 52) \\ &= \left(\frac{e^{-2(10)}(2(10))^{18}}{18!}\right)\left(\frac{e^{-5(10)}(5(10))^{52}}{52!}\right) \\ &= 0.0045 \end{aligned}$$

7.1.1 Translated Poisson Process

Δ Theorem

Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter λ. For $s > 0$, let

$$\tilde{N}_t = N_{t+s} - N_s, \text{ for } t \geq 0$$

Then, $\left(\tilde{N}_t\right)_{t\geq 0}$ is a Poisson process with parameter λ

· **Example**

On election day, people arrive at a voting center according to a Poisson process. On average, 100 voters arrive every hour. If 150 people arrive during the first hour, what is the probability that at most 350 people arrive before the third hour?

Solution: Let N_t denote the number of arrivals in the first t hours. Then, $(N_t)_{t\geq 0}$ is a Poisson process with parameter $\lambda = 100$. Given $N_1 = 150$, the distribution of $N_3 - N_1 = N_3 - 150$ is equal to the distribution of N_2. This gives

$$\begin{aligned} P\left(N_3 \leq 350 | N_1 = 150\right) &= P\left(N_3 - 150 \leq 200 | N_1 = 150\right) \\ &= P\left(N_2 \leq 200\right) \\ &= \sum_{k=0}^{200} \frac{e^{-100(2)}(100(2))^k}{k!} \\ &= 0.519 \end{aligned}$$

7.2 Arrival, Interarrival Times

For a Poisson process with parameter λ, let X denote the time of the first arrival. Then, $X > t$ if and only if there are no arrivals in $[0, t]$. Thus,

$$P(X > t) = P\left(N_t = 0\right) = e^{-\lambda t}, \text{ for } t > 0$$

Hence, X has an exponential distribution with parameter λ. What is true for the time of the first arrival is also true for the time between the first and second arrival, and for all interarrival times.

△ Definition

Poisson process–Definition 2

Let $X_1, X_2, \ldots$ be a sequence of i.i.d. exponential random variables with parameter λ. For $t > 0$, let

$$N_t = \max\{n : X_1 + \cdots + X_n \leq t\}$$

with $N_0 = 0$. Then, $(N_t)_{T\geq 0}$ defines a Poisson process with parameter λ

Let

$$S_n = X_1 + \cdots + X_n, \text{ for } n = 1, 2, \ldots$$

We call $S_1, S_2, \ldots$ the arrival times of the process, where S_k is the time of the k th arrival. Furthermore,

$$X_k = S_k - S_{k-1}, \text{ for } k = 1, 2, \ldots$$

is the interarrival time between the $(k-1)$ th and k th arrival, with $S_0 = 0$

A benefit of Definition 2 is that it leads to a direct method for constructing, and simulating, a Poisson process:

1. Let $S_0 = 0$
2. Generate i.i.d. exponential random variables $X_1, X_2, \ldots$
3. Let $S_n = X_1 + \cdots + X_n$, for $n = 1, 2, \ldots$
4. For each $k = 0, 1, \ldots$, let $N_t = k$, for $S_k \leq t < S_{k+1}$

7.2.1 Memorylessness

△ Definition

A random variable X is memoryless if, for all $s, t > 0$, $P(X > s+t|X > s) = P(X > t)$

The exponential distribution is the only continuous distribution that is memoryless. (The geometric distribution has the honors for the discrete case.)

Δ Theorem

Minimum of Independent Exponential Random Variables

Let $X_1, \ldots, X_n$ be independent exponential random variables with respective parameters $\lambda_1, \ldots, \lambda_n$. Let $M = \min(X_1, \ldots, X_n)$

1. For $t > 0$,

$$P(M > t) = e^{-t(\lambda_1 + \cdots + \lambda_n)}$$

That is, M has an exponential distribution with parameter $\lambda_1 + \cdots + \lambda_n$.

2. For $k = 1, \ldots, n$,

$$P(M = X_k) = \frac{\lambda_k}{\lambda_1 + \cdots + \lambda_n}$$

· **Example**

A Boston subway station services the red, green, and orange lines. Subways on each line arrive at the station according to three independent Poisson processes. On average, there is one red train every 10 minutes, one green train every 15 minutes, and one orange train every 20 minutes.

(i) When you arrive at the station what is the probability that the first subway that arrives is for the green line?

(ii) How long will you wait, on average, before some train arrives?

(iii) You have been waiting 20 minutes for a red train and have watched three orange trains go by. What is the expected additional time you will wait for your subway?

Solution:

(i)Let X_G, X_R, and X_O denote, respectively, the times of the first green, red, and orange subways that arrive at the station.

$$P(\min(X_G, X_R, X_O) = X_G) = \frac{1/15}{1/10 + 1/15 + 1/20} = \frac{4}{13} = 0.31$$

(ii) The time of the first train arrival is the minimum of X_G, X_R, and X_O, which has an exponential distribution with parameter

$$\frac{1}{10}+\frac{1}{15}+\frac{1}{20}=\frac{13}{60}$$

Thus, you will wait, on average $60/13 = 4.615$ minutes.

(iii) Your waiting time is independent of the orange arrivals. By memorylessness of interarrival times, the additional waiting time for the red line has the same distribution as the original waiting time. You will wait, on average, 10 more minutes.

Δ Theorem

Arrival Times and Gamma Distribution

For $n = 1, 2, \ldots,$ let S_n be the time of the n th arrival in a Poisson process with parameter λ. Then, S_n has a gamma distribution with parameters n and λ. The density function of S_n is

$$f_{S_n}(t)=\frac{\lambda^n t^{n-1} e^{-\lambda t}}{(n-1)!}, \text{ for } t>0$$

Mean and variance are

$$E\left(S_n\right)=\frac{n}{\lambda} \quad \text{and} \quad \operatorname{Var}\left(S_n\right)=\frac{n}{\lambda^2}$$

· **Example**

The times when goals are scored in hockey aremodeled as a Poisson process in Morrison (1976). For such a process, assume that the average time between goals is 15 minutes.

(i) In a 60-minute game, find the probability that a fourth goal occurs in the last 5 minutes of the game.

(ii) Assume that at least three goals are scored in a game. What is the mean time of the third goal?

Solution: (i) The desired probability is

$$P\left(55<S_4 \leq 60\right)=\frac{1}{6} \int_{55}^{60}(1/15)^4 t^3 e^{-t/15} dt=0.068$$

(ii) The desired expectation is

$$\begin{aligned} E\left(S_3 | S_3 < 60\right) &= \frac{1}{P\left(S_3 < 60\right)} \int_0^{60} t f_{S_3}(t) dt \\ &= \frac{1}{P\left(S_3 < 60\right)} \int_0^{60} t \frac{(1/15)^3 t^2 e^{-t/15}}{2} dt \\ &= \frac{25.4938}{0.7619} = 33.461 \text{ minutes.} \end{aligned}$$

7.3 Infinitesimal Probability

A third way to define the Poisson process is based on an infinitesimal description of the distribution of points (e.g., events) in small intervals.

To state the new definition, we use little-oh notation. Write $f(h) = \mathrm{o}(h)$ to mean that

$$\lim_{h \to 0} \frac{f(h)}{h} = 0$$

More generally, say that a function f is little-oh of g, and write $f(h) = \mathrm{o}(g(h))$, to mean that

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

Little-oh notation is often used when making order of magnitude statements about a function, or in referencing the remainder term of an approximation.

· **Example**

$$e^h = 1 + h + \frac{h^2}{2} + \frac{h^3}{6} + \cdots = 1 + h + R(h)$$

where $R(h) = e^z h^2/2$, for some $z \in (-h, h)$. since $R(h)/h = e^z h/2 \to 0$, as $h \to 0$, we can write

$$e^h = 1 + h + \mathrm{o}(h)$$

Note that if two functions f and g are little-oh of h, then $f(h) + g(h) = \mathrm{o}(h)$, since $(f(h) + g(h))/h \to 0$, as $h \to 0$. Similarly, if $f(h) = \mathrm{o}(h)$, then $cf(h) = \mathrm{o}(h)$, for any constant c. If $f(h) = o(1)$, then $f(h) \to 0$, as $h \to 0$.

Δ **Definition**

Poisson Process-Definition 3 A Poisson process with parameter λ is a counting process $(N_t)_{t\geq 0}$ with the following properties: 1. $N_0 = 0$

2. The process has stationary and independent increments.

3. $P(N_h = 0) = 1 - \lambda h + o(h)$

4. $P(N_h = 1) = \lambda h + o(h)$

5. $P(N_h > 1) = o(h)$

Properties 3–5 essentially ensure that there cannot be infinitely many arrivals in a finite interval, and that in an infinitesimal interval there may occur at most one event.

7.4 Thinning, Superposition

· **Example**

Assume that babies are born on a maternity ward according to a Poisson process $(N_t)_{t\geq 0}$ with parameter λ. How can the number of male births and the number of female births be described?

Babies' sex is independent of each other. We can think of a male birth as the result of a coin fip whose heads probability is p. Assume that there are n birth as by time t. Then, the number of male births by time t is the number of heads in n i.i. d. coin flips, which has a binomial distribution with parameters n and p. Similarly the number of female births in $[0, t]$ has a binomial distribution with parameters n and $1 - p$.

Let M_t denote the number of male births by time t. Similarly define the number of female births F_t. Thus, $M_t + F_t = N_t$. The joint probability mass function of (M_t, F_t) is

$$
\begin{aligned}
P\left(M_t=m, F_t=f\right) &= P\left(M_t=m, F_t=f, N_t=m+f\right) \\
&= P\left(M_t=m, F_t=f \mid N_t=m+f\right) P\left(N_t=m+f\right) \\
&= P\left(M_t=m \mid N_t=m+f\right) P\left(N_t=m+f\right) \\
&= \frac{(m+f)!}{m!f!} p^m(1-p)^f \frac{e^{-\lambda t}(\lambda t)^{m+f}}{(m+f)!} \\
&= \frac{p^m(1-p)^f e^{-\lambda t(p+(1-p))}(\lambda t)^{m+f}}{m!f!} \\
&= \left(\frac{e^{-\lambda p t}(\lambda p t)^m}{m!}\right)\left(\frac{e^{-\lambda(1-p)t}(\lambda(1-p)t)^f}{f!}\right)
\end{aligned}
$$

for $m, f = 0, 1, \ldots$ This shows that M_t and F_t are independent Poisson random variables with parameters $\lambda p t$ and $\lambda(1-p)t$, respectively. In fact, each process $(M_t)_{t\geq 0}$ and $(F_t)_{t\geq 0}$ is a Poisson process, called a thinned process.

△ Definition

Thinned Poisson process

Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter λ. Assume that each arrival, independent of other arrivals, is marked as a type k event with probability p_k, for $k = 1, \ldots, n$, where $p_1 + \cdots + p_n = 1$. Let $N_t^{(k)}$ be the number of type k events in $[0, t]$. Then, $\left(N_t^{(k)}\right)_{t\geq 0}$ is a Poisson process with parameter λp_k. Furthermore, the processes

$$\left(N_t^{(1)}\right)_{t\geq 0}, \ldots, \left(N_t^{(n)}\right)_{t\geq 0}$$

are independent. Each process is called a thinned Poisson process.

△ Definition

Superposition Process

Assume that $\left(N_t^{(1)}\right)_{t\geq 0}, \cdots, \left(N_t^{(n)}\right)_{t\geq 0}$ are n independent Poisson processes with respective parameters $\lambda_1, \ldots, \lambda_n$. Let $N_t = N_t^{(1)} + \cdots + N_t^{(n)}$, for $t \geq 0$. Then, $(N_t)_{t\geq 0}$ is a Poisson process with parameter $\lambda = \lambda_1 + \cdots + \lambda_n$

· **Example**

In the land of Oz, sightings of lions, tigers, and bears each follow a Poisson process with respective parameters, $\lambda_L, \lambda_T, \lambda_B$, where the time unit is hours. Sightings of the three species are independent of each other.

(i) Find the probability that Dorothy will not see any animal in the first 24 hours from when she arrives in Oz.

(ii) Dorothy saw three animals one day. Find the probability that each species was seen.

Solution: (i) The process of animal sightings $(N_t)_{t\geq 0}$ is the superposition of three independent Poisson processes. Thus, it is a Poisson process with parameter $\lambda_L + \lambda_T + \lambda_B$. The desired probability is

$$P(N_{24} = 0) = e^{-24(\lambda_L+\lambda_T+\lambda_B)}$$

(ii) Let L_t, T_t, and B_t be the numbers of lions, tigers, and bears, respectively, seen by time t. The desired probability is

$$\begin{aligned}
&P(L_{24} = 1, B_{24} = 1, T_{24} = 1 | N_{24} = 3) \\
&\quad = \frac{P(L_{24} = 1, B_{24} = 1, T_{24} = 1, N_{24} = 3)}{P(N_{24} = 3)} \\
&\quad = \frac{P(L_{24} = 1, B_{24} = 1, T_{24} = 1)}{P(N_{24} = 3)} \\
&\quad = \frac{P(L_{24} = 1)\, P(B_{24} = 1)\, P(T_{24} = 1)}{P(N_{24} = 3)} \\
&\quad = \frac{6\lambda_L\lambda_B\lambda_T}{(\lambda_L + \lambda_B + \lambda_T)^3}
\end{aligned}$$

7.4.1 Embedding and the Birthday Problem

The classic birthday problem asks, "How many people must be in a room before the probability that some share a birthday, ignoring year and leap days, is at least 50%? "

The probability that two people have the same birthday is 1 minus the probability that no one shares a birthday, which is

$$p_k = 1 - \prod_{i=1}^{k} \frac{366 - i}{365} = 1 - \frac{365!}{(365-k)!365^k}$$

The product above indicates that the probability of none of "k" people shares a birthday is conditional. For example, the $P($ second people does not share B-day with first people$|$B-day of first people is $X) = 364/365$. One finds that $p_{22} = 0.476$ and $p_{23} = 0.507$. Thus, 23 people are needed.

· **Example**

Consider this variant of the birthday problem, assuming a random person's birthday is uniformly distributed on the 365 days of the year. People enter a room one by one. How many people are in the room the first time that two people share the same birthday? Let K be the desired number.

We show how to find the mean and standard deviation of K by embedding. Consider a continuous-time version of the previous question. People enter a room according to a Poisson process $(N_t)_{t\geq 0}$ with rate $\lambda = 1$. Each person is independently marked with one of 365 birthdays, where all birthdays are equally likely. The procedure creates 365 thinned Poisson processes, one for each birthday. Each of the 365 processes are independent, and their superposition gives the process of people entering the room.

Let $X_1, X_2, \ldots,$ be the interarrival sequence for the process of people entering the room. The X_i are i.i.d. exponential random variables with mean 1. Let T be the first time when two people in the room share the same birthday. Then,

$$T = \sum_{i=1}^{K} X_i$$

The X_i are independent of K. The random variable T is represented as a random sum of random variables. By results for such sums

$$E[T] = E[K]E[X_1] = E[K]$$

For each $k = 1, \ldots, 365$, let Z_k be the time when the second person marked with birthday k enters the room. Then, the first time two people in the room have the same birthday is

$T = \min_{1 \le k \le 365} Z_k$. Each Z_k, being the arrival time of the second event of a Poisson process, has a gamma distribution with parameters $n = 2$ and $\lambda = 1/365$, with density

$$f(t) = \frac{te^{-t/365}}{365^2}, \text{ for } t > 0$$

This gives

$$\begin{aligned}
P(T > t) &= P\left(\min\left(Z_1, \ldots, Z_{365}\right) > t\right) \\
&= P\left(Z_1 > t, \ldots, Z_{365} > t\right) \\
&= P\left(Z_1 > t\right)^{365} = \left(1 + \frac{t}{365}\right)^{365} e^{-t}, \text{ for } t > 0
\end{aligned}$$

The desired birthday expectation is

$$E(K) = E(T) = \int_0^\infty P(T > t) dt = \int_0^\infty \left(1 + \frac{t}{365}\right)^{365} e^{-t} dt$$

7.5 Uniform Distribution

Let $U_1, \ldots, U_n$ be an i.i.d. sequence of random variables uniformly distributed on $[0, t]$. Their joint density function is

$$f_{U_1, \ldots, U_n}\left(u_1, \ldots, u_n\right) = \frac{1}{t^n}, \text{ for } 0 \le u_1, \ldots, u_n \le t$$

Arrange the U_i in increasing order $U_{(1)} \le U_{(2)} \le \cdots \le U_{(n)}$, where $U_{(k)}$ is the k th smallest of the U_i. The ordered sequence $\left(U_{(1)}, \ldots, U_{(n)}\right)$ is called the order statistics of the original sequence.

The joint density function of the order statistics is

$$f_{U_{(1)}, \ldots, U_{(n)}}\left(u_1, \ldots, u_n\right) = \frac{n!}{t^n}, \text{ for } 0 \le u_1 < \cdots < u_n \le t$$

Proof: Consider a sample of n independent uniform random variables on $[0, t]$. There are $n!$ such samples that would give rise to these order statistic values, as there are $n!$ orderings of the n distinct numbers $u_1, \ldots, u_n$. The value of the joint density for each of these uniform samples is equal to $1/t^n$.

△ Theorem

Arrival Times and Uniform Distribution

Let $S_1, S_2, \ldots,$ be the arrival times of a Poisson process with parameter λ. Conditional on $N_t = n$, the joint distribution of $(S_1, \ldots, S_n)$ is the distribution of the order statistics of n i.i.d. uniform random variables on $[0, t]$. That is, the joint density function of $S_1, \ldots, S_n$ is

$$f(s_1, \ldots, s_n) = \frac{n!}{t^n}, \text{ for } \quad 0 < s_1 < \cdots < s_n < t$$

Equivalently, let $U_1, \ldots, U_n$ be an i.i.d. sequence of random variables uniformly distributed on $[0, t]$. Then, conditional on $N_t = n$

$$(S_1, \ldots, S_n) \text{ and } \left(U_{(1)}, \ldots, U_{(n)}\right)$$

have the same distribution.

From this, one can obtain joint distributions of any subset of the arrival times, including the marginal distributions. For instance, the last (first) arrival time, conditional on there being n arrivals in [0, t], has the same distribution as the maximum (minimum) of n independent uniform random variables on (0, t).

· Example

Concert-goers arrive at a show according to a Poisson process with parameter λ. The band starts playing at time t. The k th person to arrive in $[0, t]$ waits $t - S_k$ time units for the start of the concert, where S_k is the k th arrival time. Find the expected total waiting time of concert-goers who arrive before the band starts.

Solution: The desired expectation is $E\left(\sum_{k=1}^{N_t}(t-S_k)\right)$. Conditioning on N_t

$$\begin{aligned}
E\left(\sum_{k=1}^{N_t}(t-S_k)\right) &= \sum_{n=1}^{\infty} E\left(\sum_{k=1}^{n}(t-S_k)\,|N_t=n\right)P(N_t=n)\\
&= \sum_{n=1}^{\infty} E\left(tn-\sum_{k=1}^{n}S_k|N_t=n\right)P(N_t=n)\\
&= \sum_{n=1}^{\infty}\left(\mathrm{tn}-E\left(\sum_{k=1}^{n}S_k|N_t=n\right)\right)P(N_t=n)\\
&= \lambda t^2-\sum_{n=1}^{\infty}E\left(\sum_{k=1}^{n}S_k|N_t=n\right)P(N_t=n)
\end{aligned}$$

Conditional on n arrivals in $[0,t]$, $S_1+\cdots+S_n$ has the same distribution as the sum of the uniform order statistics. Furthermore, $\sum_{k=1}^{n}U_{(k)}=\sum_{k=1}^{n}U_k$. This gives

$$\begin{aligned}
\sum_{n=1}^{\infty}E\left(\sum_{k=1}^{n}S_k|N_t=n\right)P(N_t=n) &= \sum_{n=1}^{\infty}E\left(\sum_{k=1}^{n}U_{(k)}\right)P(N_t=n)\\
&= \sum_{n=1}^{\infty}E\left(\sum_{k=1}^{n}U_k\right)P(N_t=n)\\
&= \sum_{n=1}^{\infty}\frac{nt}{2}P(N_t=n)\\
&= \frac{\lambda t^2}{2}
\end{aligned}$$

Finally

$$E\left(\sum_{k=1}^{N_t}(t-S_k)\right)=\lambda t^2-\frac{\lambda t^2}{2}=\frac{\lambda t^2}{2}$$

7.5.1 Simulation

Results for arrival times and the uniform distribution offer a new method for simulating a Poisson process with parameter λ on an interval $[0,t]$:

1. Simulate the number of arrivals N in $[0,t]$ from a Poisson distribution with parameter λt

2. Generate N i.i.d. random variables uniformly distributed on $(0,t)$.

3. Sort the variables in increasing order to give the Poisson arrival times.

7.6 Spatial Poisson Process

For $d \geq 1$ and $A \subseteq \mathbb{R}^d$, let N_A denote the number of points in the set A. Write $|A|$ for the size of A (e.g., area in $\mathbb{R}^2$, volume in $\mathbb{R}^3$).

△ Definition

Spatial Poisson Process

A collection of random variables $(N_A)_{A\subset\mathbb{R}^d}$ is a spatial Poisson process with parameter λ if

1. for each bounded set $A \subseteq \mathbb{R}^d$, N_A has a Poisson distribution with parameter $\lambda|A|$.

2. whenever A and B are disjoint sets, N_A and N_B are independent random variables.

The uniform distribution arises for the spatial process in a similar way to how it does for the one-dimensional Poisson process. Given a bounded set $A \subseteq \mathbb{R}^d$, then conditional on there being n points in A, the locations of the points are uniformly distributed in A. For this reason, a spatial Poisson process is sometimes called a model of complete spatial randomness.

To simulate a spatial Poisson process with parameter λ on a bounded set A, first simulate the number of points N in A according to a Poisson distribution with parameter $\lambda|A|$. Then, generate N points uniformly distributed in A.

7.7 Non-homogeneous Poisson Process

△ Definition

A counting process $(N_t)_{t\geq0}$ is a nonhomogeneous Poisson process with intensity function $\lambda(t)$, if 1. $N_0 = 0$ 2. For all $t > 0$, N_t has a Poisson distribution with mean

$$E(N_t) = \int_0^t \lambda(x)dx$$

3. For $0 \leq q < r \leq s < t, \quad N_r - N_q$ and $N_t - N_s$ are independent random variables.

It can be shown that for $0 < s < t, N_t - N_s$ has a Poisson distribution with parameter

$\int_s^t \lambda(x)dx$. If $\lambda(t) = \lambda$ is constant, we obtain the regular Poisson process with parameter λ

· **Example**

In reliability engineering one is concerned with the probability that a system is working during an interval of time. A common model for failure times is a nonhomogeneous Poisson process with intensity function of the form

$$\lambda(t) = \alpha\beta t^{\beta-1}$$

where $\alpha, \beta > 0$ are parameters, and t represents the age of the system. At $\beta = 1$, the model reduces to a homogeneous Poisson process with parameter α. If the system starts at $t = 0$, the expected number of failures after t time units is

$$E(N_t) = \int_0^t \lambda(x)dx = \int_0^t \alpha\beta x^{\beta-1}dx = \alpha t^{\beta}$$

Because of the power law form of the mean failure time, the process is sometimes called a power law Poisson process.

Of interest, is the reliability $R(t)$, defined as the probability that a system, which starts at time t, is operational up through time $t + c$, for some constant c, that is, the probability of no failures in the interval $(t, t+c]$. This gives

$$R(t) = P(N_{t+c} - N_t = 0) = e^{-\int_t^{t+c}\lambda(x)dx} = e^{-\int_t^{t+c}\alpha\beta x^{\beta-1}dx} = e^{-\alpha\left((t+c)^{\beta} - t^{\beta}\right)}$$

7.8 Parting Paradox

The following classic is based on Feller (1968) . Buses arrive at a bus stop according to a Poisson process. The time between buses, on average, is 10 minutes. Lisa gets to the bus stop at time t. How long can she expect to wait for a bus? Here are two possible answers:

(i) By memorylessness, the time until the next bus is exponentially distributed with mean 10 minutes. Lisa will wait, on average, 10 minutes.

(ii) Lisa arrives at some time between two consecutive buses. The expected time between consecutive buses is 10 minutes. By symmetry, her expected waiting time should be half that, or 5 minutes.

Paradoxically, both answers have some truth to them! To explain the paradox, consider the process of bus arrivals. The rate of one arrival per 10 minutes is an average. The time between buses is random, and buses may arrive one right after the other, or there may be a long time between consecutive buses. When Lisa gets to the bus stop, she is more likely to get there during a longer interval between buses than a shorter interval.**Please refer to the code "Waiting time paradox" for details.**

This example illustrates the phenomenon of length-biased or size-biased sampling. If you reach into a bag containing pieces of string of different lengths and pick a string at random, you tend to pick a longer rather than a shorter piece. Bus interarrival times are analogous to lengths of string.

For the bus waiting problem, the expected length of an interarrival time, which contains a fixed time t, is larger-about twice as large-than the average interval length between buses.

Here is an exact analysis. Consider a fixed $t > 0$. The time of the last bus before t is S_{N_t}. The time of the next bus after t is S_{N_t+1}. The expected length of the interval containing t is

$$E\left(S_{N_i+1} - S_{N_t}\right) = E\left(S_{N_t+1}\right) - E\left(S_{N_t}\right)$$

$$E\left(S_{N_t+1} \mid N_t = k\right) = E\left(S_{k+1} \mid N_t = k\right) = E\left(S_{k+1}\right) = \frac{k+1}{\lambda}$$

The second equality is because the $(k+1)$ th arrival occurs after time t, and is thus independent of N_t. It follows that $E\left(S_{N_t+1} \mid N_t\right) = \left(N_t + 1\right)/\lambda$. By the law of total expectation,

$$E\left(S_{N_t+1}\right) = E\left(E\left(S_{N_t+1} \mid N_t\right)\right) = E\left(\frac{N_t+1}{\lambda}\right) = \frac{\lambda t + 1}{\lambda} = t + \frac{1}{\lambda}$$

For $E\left(S_{N_t}\right)$, we have that $E\left(S_{N_t} \mid N_t = k\right) = E\left(S_k \mid N_t = k\right)$. Conditional on $N_t = k$ the k th arrival time has the same distribution as the maximum of k i.i.d. uniform random variables distributed on $(0, t)$. Since for $X_1, X_2, \ldots, X_k \sim \text{Unif}(0, t)$, iid and the maximum M has CDF as:

$$\begin{aligned}
P(M \leq x) &= P\left(\max\left(X_1, X_2, \cdots, X_k\right) \leq x\right) \\
&= P\left(X_1 \leq x, X_2 \leq x, \cdots, X_k \leq x\right) \\
&\overset{ind}{=} \prod_{i=1}^{k} P\left(X_i \leq x\right) \\
&= \prod_{i=1}^{k} \frac{x}{t} \\
&= \left(\frac{x}{t}\right)^k
\end{aligned}$$

We have PDF for the maximum of k i.i.d. uniform rv as

$$f(x) = k(\frac{x^{k-1}}{t^k})$$

We get $E(M) = E(S_{N_t}|N_t = k) = tk/(k+1)$ and thus

$$E\left(S_{N_t}|N_t\right) = tN_t/\left(N_t + 1\right) = t - t/\left(N_t + 1\right)$$

$$E\left(S_{N_t}\right) = E\left(E\left(S_{N_t}|N_t\right)\right) = E\left(t - \frac{t}{N_t + 1}\right) = t - tE\left(\frac{1}{N_t + 1}\right)$$

Because

$$\begin{aligned}
E\left(\frac{1}{N_t + 1}\right) &= \sum_{k=0}^{\infty}\left(\frac{1}{k+1}\right)\frac{e^{-\lambda t}(\lambda t)^k}{k!} \\
&= \frac{e^{-\lambda t}}{\lambda t}\sum_{k=0}^{\infty}\frac{(\lambda t)^{k+1}}{(k+1)!} = \frac{e^{-\lambda t}}{\lambda t}\sum_{k=1}^{\infty}\frac{(\lambda t)^k}{k!} \\
&= \frac{e^{-\lambda t}}{\lambda t}\left(e^{\lambda t} - 1\right) = \frac{1 - e^{-\lambda t}}{\lambda t}
\end{aligned}$$

Together with $E(S_{N_t})$, we have

$$E\left(S_{N_t}\right) = t - \frac{1}{\lambda} + \frac{e^{-\lambda t}}{\lambda}$$

Finally,

$$E\left(S_{N_{t+1}} - S_{N_t}\right) = \left(t + \frac{1}{\lambda}\right) - \left(t - \frac{1}{\lambda} + \frac{e^{-\lambda t}}{\lambda}\right) = \frac{2 - e^{-\lambda t}}{\lambda} \approx \frac{2}{\lambda}$$

8 Continuous-Time Markov Chains

A continuous-time process allows one to model not only the transitions between states, but also the duration of time in each state. The central Markov property continues to hold—given the present, past and future are independent.

In a continuous-time Markov chain, when a state is visited, the process stays in that state for an exponentially distributed length of time before moving to a new state.

△ Definition

A continuous-time stochastic process $(X_t)_{t\geq 0}$ with discrete state space S is a continuous-time Markov chain if

$$P\left(X_{t+s}=j|X_s=i, X_u=x_u, 0\leq u<s\right)=P\left(X_{t+s}=j|X_s=i\right)$$

for all $s, t, \geq 0, x_u \in S,$ and $0 \leq u < s$

The process is said to be time-homogeneous if this probability does not depend on s. That is,

$$P\left(X_{t+s}=j|X_s=i\right)=P\left(X_t=j|X_0=i\right), \text{ for } s\geq 0$$

For cotinuous Markov chains, the transition function $P(t)$ has similar properties as that of the transition matrix for a discrete-time Markov chain. For instance, the Chapman-Kolmogorov equations hold.

△ Theorem

For a continuous-time Markov chain $(X_t)_{t\geq 0}$ with transition function $P(t)$

$$P(s+t)=P(s)P(t)$$

for $s, t \geq 0$

$$P_{ij}(s+t) = [P(s)P(t)]_{ij} = \sum_k P_{ik}(s)P_{kj}(t), \text{ for states } i, j, \text{ and } s, t \geq 0$$

△ Theorem

Holding times are exponentially distributed

Let T_i be the holding time at state i, that is, the length of time that a continuous-time Markov chain started in i stays in i before transitioning to a new state. Then, T_i has an exponential distribution.

The evolution of a continuous-time Markov chain which is neither absorbing nor an explosive can be described as follows. Starting from i, the process stays in i for an exponentially distributed length of time, on average $1/q_i$ time units. Then, it hits a new state $j \neq i$, with some probability p_{ij}. The process stays in j for an exponentially distributed length of time, on average $1/q_j$ time units, and so on.

The transition probabilities p_{ij} describe the discrete transitions from state to state. If we ignore time, and just watch state to state transitions, we see a sequence $Y_0, Y_1, \ldots,$ where Y_n is the n th state visited by the continuous process. The sequence $Y_0, Y_1, \ldots$ is a discrete-time Markov chain called the embedded chain.

Let $\widetilde{P}$ be the transition matrix for the embedded chain. That is, $\widetilde{P}_{ij} = p_{ij}$. Then, $\widetilde{\boldsymbol{P}}$ is a stochastic matrix whose diagonal entries are 0.

8.1 Transition rate and alarm clocks

Imagine that for each state i, there are independent alarm clocks associated with each of the states that the process can visit after i. If j can be hit from i, then the alarm clock associated with (i, j) will ring after an exponentially distributed length of time with parameter q_{ij}.

The q_{ij} are called the transition rates of the continuous-time process. From the transition rates, we can obtain the holding time parameters and the embedded chain transition probabilities.

Consider the process started in i. The clocks are started, and the first one that rings determines the next transition. The time of the first alarm is the minimum of independent exponential random variables with parameters $q_{i1}, q_{i2}, \ldots$ The minimum has an exponential distribution with parameter $\sum_k q_{ik}$. That is, the chain stays at i for an exponentially distributed amount of time with parameter $\sum_k q_{ik}$. From the discussion of holding times, the exponential length of time that the process stays in i has parameter q_i. That is,

$$q_i = \sum_k q_{ik}$$

The interpretation is that the rate that the process leaves state i is equal to the sum of the rates from i to each of the next states.

From i, the chain moves to j if the (i, j) clock rings first, which occurs with probability $q_{ij}/\sum_k q_{ik} = q_{ij}/q_i$. Thus, for the embedded chain transition probabilities

$$p_{ij} = \frac{q_{ij}}{\sum_k q_{ik}} = \frac{q_{ij}}{q_i}$$

8.2 Infinitesimal Generator

Assume that $(X_t)_{t\geq 0}$ is a continuous-time Markov chain with transition function $P(t)$. Assume the transition function is differentiable. Note that

$$P_{ij}(0) = \begin{cases} 1, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases}$$

If $X_t = i$, then the instantaneous transition rate of hitting $j \neq i$ is

$$\begin{aligned}
\lim_{h\to 0^+} \frac{E(\text{ Number of transitions to } j \text{ in } (t, t+h])}{h} &= \lim_{h\to 0^+} \frac{P\left(X_{t+h} = j | X_t = i\right)}{h} \\
&= \lim_{h\to 0^+} \frac{P\left(X_h = j | X_0 = i\right)}{h} \\
&= \lim_{h\to 0^+} \frac{P_{ij}(h)}{h} \\
&= \lim_{h\to 0^+} \frac{P_{ij}(h) - P_{ij}(0)}{h} \\
&= P'_{ij}(0)
\end{aligned}$$

Let $\boldsymbol{Q} = \boldsymbol{P}'(0)$. The off-diagonal entries of $\boldsymbol{Q}$ are the instantaneous transition rates, which are the transition rates q_{ij} introduced in the last section. That is, $Q_{ij} = q_{ij}$, for $i \neq j$. In the language of infinitesimals, if $X_t = i$, then the chance that $X_{t+dt} = j$ is $q_{ij}dt$

$$\begin{aligned}
Q_{ii} = P'_{ii}(0) = \lim_{h\to 0^+} \frac{P_{ii}(h) - P_{ii}(0)}{h} \lim_{h\to 0^+} \frac{P_{ii}(h) - 1}{h} \\
= \lim_{h\to 0^+} -\frac{\sum P_{ij}(h)}{h} = -\sum_{j\neq i} \lim_{h\to 0^+} \frac{P_{ij}(h)}{h} \\
= -\sum_{j\neq i} Q_{ij} = -\sum_{j\neq i} q_{ij} = -q_i
\end{aligned}$$

The Q matrix is called the generator or infinitesimal generator. It is the most important matrix for continuous-time Markov chains. Clearly, the generator is not a stochastic matrix. Diagonal entries are negative, entries can be greater than 1, and rows sum to 0.

△ Theorem

For a continuous-time Markov chain, let q_{ij}, q_i, and p_{ij} be defined as above. For $i \neq j$

$$q_{ij} = q_i p_{ij}$$

8.2.1 Forward, Backward Equations

△ Definition

A continuous-time Markov chain with transition function $P(t)$ and infinitesimal generator Q satisfies the forward equation:

$$P'(t) = P(t)Q$$

And the backward equation:

$$P'(t) = QP(t)$$

Equivalently, for each i and j:

$$P'_{ij}(t) = \sum_k P_{ik}(t)q_{kj} = -P_{ij}(t)q_j + \sum_{k \neq j} P_{ik}(t)q_{kj}$$

$$P'_{ij}(t) = \sum_k q_{ik}P_{kj}(t) = -q_i P_{ij}(t) + \sum_{k \neq i} q_{ik}P_{kj}(t)$$

· **Example**

(Poisson process) The transition probabilities for the Poisson process with parameter λ:

$$P_{ij}(t) = \frac{e^{-\lambda t}(\lambda t)^{j-i}}{(j-i)!}, \text{ for } j \geq i$$

They satisfy the Kolmogorov forward equations:

$$\begin{aligned} &P'_{ii}(t) = -\lambda P_{ii}(t) \\ &P'_{ij}(t) = -\lambda P_{ij}(t) + \lambda P_{i,j-1}, \text{ for } j = i+1, i+2, \ldots \end{aligned}$$

and the backward equations

$$\begin{aligned} &P'_{ii}(t) = -\lambda P_{ii}(t) \\ &P'_{ij}(t) = -\lambda P_{ij}(t) + \lambda P_{j+1,i}, \text{ for } j = i+1, i+2, \ldots \end{aligned}$$

8.2.2 Matrix Exponential

Δ Definition

Let A be a $k \times k$ matrix. The matrix exponential e^A is

$$e^A = \sum_{n=0}^{\infty} \frac{1}{n!} A^n = I + A + \frac{1}{2}A^2 + \frac{1}{6}A^3 + \cdots$$

The matrix exponential satisfies many familiar properties of the exponential function. These include

△ Theorem

1. $e^0 = I$
2. $e^A e^{-A} = I$
3. $e^{(s+t)A} = e^{sA} e^{tA}$
4. If $AB = BA$, then $e^{A+B} = e^A e^B = e^B e^A$
5. $\frac{d}{dt} e^{tA} = A e^{tA} = e^{tA} A$

For a continuous-time Markov chain with generator $\boldsymbol{Q}$, the matrix exponential e^{tQ} is the unique solution to the forward and backward equations. Letting $\boldsymbol{P}(t) = e^{tQ}$ gives

$$P'(t) = \frac{d}{dt} e^{tQ} = Q e^{tQ} = e^{tQ} Q = P(t)Q = QP(t)$$

△ Theorem

For a continuous-time Markov chain with transition function $P(t)$ and infinitessimal generator Q

$$P(t) = e^{tQ} = \sum_{n=0}^{\infty} \frac{1}{n!} (tQ)^n = I + tQ + \frac{t^2}{2} Q^2 + \frac{t^3}{6} Q^3 + \cdots$$

8.2.3 Diagonalization

If the Q matrix is diagonalizable, then so is e^{tQ}, and the transition function can be expressed in terms of the eigenvalues and eigenvectors of Q. Write $Q = SDS^{-1}$, where

$$\boldsymbol{D} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k \end{pmatrix}$$

and $\mathbf{S}$ is an invertible matrix whose columns are the corresponding eigenvectors. This gives

$$\begin{aligned} e^{tQ} &= \sum_{n=0}^{\infty} \frac{1}{n!}(tQ)^n = \sum_{n=0}^{\infty} \frac{t^n}{n!}\left(SDS^{-1}\right)^n \\ &= \sum_{n=0}^{\infty} \frac{t^n}{n!} SD^nS^{-1} = S\left(\sum_{n=0}^{\infty} \frac{t^n}{n!} D^n\right) S^{-1} \\ &= Se^{tD}S^{-1} \end{aligned}$$

And

$$\begin{aligned} e^{tD} &= \sum_{n=0}^{\infty} \frac{t^n}{n!}\boldsymbol{D}^n = \sum_{n=0}^{\infty} \frac{t^n}{n!} \begin{pmatrix} \lambda^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_k^n \end{pmatrix} \\ &= \begin{pmatrix} e^{t\lambda_1} & 0 & \cdots & 0 \\ 0 & e^{t\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & e^{i\lambda_k} \end{pmatrix} \end{aligned}$$

8.3 Long-term Behavior

△ **Definition**

Limiting distribution A probability distribution π is the limiting distribution of a continuous-time Markov chain if for all states i and j

$$\lim_{t\to\infty} P_{ij}(t) = \pi_j$$

△ Definition

Stationary distribution A probability distribution π is a stationary distribution if

$$\pi = \pi P(t), \text{ for } t \geq 0$$

That is, for all states j,

$$\pi_j = \sum_i \pi_i P_{ij}(t), \text{ for } t \geq 0$$

A continuous-time Markov chain is irreducible if for all i and j, $P_{ij}(t) > 0$ for some $t > 0$.

· Lemma

If $P_{ij}(t) > 0$, for some $t > 0$, then $P_{ij}(t) > 0$, for all $t > 0$

Therefore, All states are essentially aperiodic in continuous-time Markov chain.

A finite-state continuous-time Markov chain is irreducible if all the holding time parameters are positive. On the contrary, if $q_i = 0$ for some i, then i is an absorbing state. If we assume that all the holding time parameters are finite, then there are two possibilities: (i) the process is irreducible, all states communicate, and $P_{ij}(t) > 0$, for $t > 0$ and all i, j or (ii) the process contains one or more absorbing states.

△ Theorem

Fundamental limiting theorem

Let $(X_t)_{t \geq 0}$ be a finite, irreducible, continuous-time Markov chain with transition function $P(t)$. Then, there exists a unique stationary distribution π, which is the limiting distribution. That is, for all j

$$\lim_{t \to \infty} P_{ij}(t) = \pi_j, \text{ for all initial } i$$

$$\lim_{t \to \infty} P(t) = \Pi$$

where $\mathbf{\Pi}$ is a matrix all of whose rows are equal to π.

Δ Theorem

Stationary Distribution and Generator Matrix

A probability distribution π is a stationary distribution of a continuous-time Markov chain with generator Q if and only if

$$\pi Q = 0 \ .$$

That is:

$$\sum_i \pi_i Q_{ij} = 0, \text{ for all } j$$

The stationary probability π_j can be interpreted as the long-term proportion of time that the chain spends in state j. This is analogous to the discrete-time case in which the stationary probability represents the long-term fraction of transitions that the chain visits a given state.

8.3.1 Absorbing states

An absorbing Markov chain is one in which there is at least one absorbing state.As in the discrete case, the nonabsorbing states are transient. There is positive probability that the chain, started in a transient state, gets absorbed and never returns to that state. For transient state i, we derive an expression for a_i, the expected time until absorption.

Let $(X_t)_{t\geq 0}$ be an absorbing continuous-time Markov chain on $\{1, \ldots, k\}$, and T denote the set of transient states. Write the generator in canonical block matrix form

$$Q = \begin{pmatrix} 0 & \mathbf{0} \\ * & V \end{pmatrix}$$

where V is a $(k-1) \times (k-1)$ matrix.

Δ Theorem

For an absorbing continuous-time Markov chain, define a square matrix F on the set of transient states, where F_{ij} s the expected time, for the chain started in i, that the process is in j until absorption. Then,

$$F = -V^{-1}$$

For the chain started in i, the mean time until absorption is,

$$a_i = \sum_j F_{ij}$$

The matrix F is called the fundamental matrix.

· Example

A patient may advance into, or recover from, successively more severe stages of a disease until some terminal state. Each stage represents a state of an absorbing continuous-time Markov chain.

Bartolomeo et al. (2011) develops such a model to study the progression of liver disease among patients diagnosed with cirrhosis of the liver. The general form of the infinitesimal generator matrix for their three-parameter model is

$$\begin{array}{c} \begin{array}{ccc} \quad\quad 1 \quad\quad & \quad 2 \quad & 3 \end{array} \\ \begin{pmatrix} -(q_{12}+q_{13}) & q_{12} & q_{13} \\ 0 & -q_{23} & q_{23} \\ 0 & 0 & 0 \end{pmatrix} \end{array}$$

where state 1 represents cirrhosis, state 2 denotes liver cancer (hepatocellular carcinoma), and state 3 is death. The fundamental matrix is

$$F = -\begin{pmatrix} -(q_{12}+q_{13}) & q_{12} \\ 0 & -q_{23} \end{pmatrix}^{-1}$$

with mean absorption times

$$a_1 = \frac{1}{q_{12} + q_{13}} + \frac{q_{12}}{q_{23}\left(q_{12} + q_{13}\right)} \quad \text{and} \quad a_2 = \frac{1}{q_{23}}$$

The fundamental matrix is estimated to be

$$\begin{pmatrix} 1 & 2 \\ 45.05 & 23.95 \\ 0.00 & 35.21 \end{pmatrix}$$

From the fundamental matrix, the estimated mean time to death for patients with liver cirrhosis is $45.05 + 23.95 = 69$ months.

A snippet of code is provided to simulate the liver cancer growth.

8.3.2 Global Balance

Assume that π is the stationary distribution of a continuous-time Markov chain. From $\pi Q = 0$, we have that

$$\sum_{i \neq j} \pi_i q_{ij} = \pi_j q_j, \text{ for all } j$$

The holding time parameter q_j is the transition rate from j. Since π_j is the long-term proportion of time the process visits j, the right-hand side of the equation above is the long-term rate that the process leaves j. Also $\pi_i q_{ij}$, is the long-term rate of transitions from i to j. Thus, the left-hand side of the equation is the long-term rate that the process enters j.

8.4 Time reversibility

Intuitively, a continuous-time Markov chain is time reversible if the process in forward time is indistinguishable from the process in reversed time. A consequence is that for all states i and j, the long-term forward transition rate from i to j is equal to the long-term backward rate from j to i.

△ Definition

A continuous-time Markov chain with generator Q and unique stationary distribution π is said to be time reversible if

$$\pi_i q_{ij} = \pi_j q_{ji} \text{ , for all } i, j$$

A continuous-time Markov chain is time reversible if and only if its embedded discrete-time chain is time reversible.

$$\pi_i q_i p_{ij} = \pi_j q_j p_{ji}, \text{ for all } i, j$$

Consider an undirected transition graph for a continuous-time Markov chain where vertices i and j are joined by an edge if $q_{ij} > 0$ or $q_{ji} > 0$. Then

△ Theorem

Markov Processes on Trees are Time Reversible

Assume that the transition graph of an irreducible continuous-time Markov chain is a tree. Then, the process is time reversible.

Theorem above gives a sufficient condition for a continuous-time Markov chain to be time reversible. However, the condition is not necessary.

8.4.1 Birth-and-Death Process

For Birth-and-Death these processes, transitions only occur to neighboring states. Births occur from i to $i+1$ at the rate λ_i. Deaths occur from i to $i-1$ at the rate μ_i. The generator matrix for the general birth-and-death process on $\{0, 1, \ldots\}$ is

$$\begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & \cdots \\ \mu_1 & -(\lambda_1+\mu_1) & \lambda_1 & 0 & \cdots \\ 0 & \mu_2 & -(\lambda_2+\mu_2) & \lambda_2 & \cdots \\ 0 & 0 & \mu_3 & -(\lambda_3+\mu_3) & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

The local balance equations for a birth-and-death process are

$$\pi_i\lambda_i = \pi_{i+1}\mu_{i+1}, \text{ for } i = 0, 1, \ldots$$

$$\pi_k = \pi_0 \frac{\lambda_0 \cdots \lambda_{k-1}}{\mu_1 \cdots \mu_k} = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}, \text{ for } k = 0, 1, \ldots$$

$$1 = \sum_{k=0}^{\infty} \pi_k = \pi_0 \sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}$$

△ Theorem

Stationary Distribution for Birth-and-Death Process

$$\sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} < \infty$$

Then, the unique stationary distribution is:

$$\pi_k = \pi_0 \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i}, \text{ for } k = 1, 2, \ldots$$

where

$$\pi_0 = \left(\sum_{k=0}^{\infty} \prod_{i=1}^{k} \frac{\lambda_{i-1}}{\mu_i} \right)^{-1}$$

· **Example**

Yule Process

The Yule process arises in biology to describe the growth of a population where each individual gives birth to an offspring at a constant rate λ independently of other individuals. Let X_t denote the size of the population at time t. If $X_t = i$, then a new individual is born when one of the i members of the population gives birth, /which occurs at rate $i\lambda$. A Yule process is a birth-and-death process with birth rate $\lambda_i = i\lambda$ and death rate $\mu_i = 0$. In a Yule process, all states are transient and no limiting distribution exists.

Solution: Assume that the initial size of the population is 1. The Yule process satisfies the Kolmogorov forward equations

$$P'_{1,j}(t) = -\lambda j P_{1,j}(t) + \lambda(j-1)P_{1,j-1}(t), \text{ for } j = 1, 2, \ldots$$
$$P_{1,1}(0) = 1 \quad \text{and} \quad P_{1,j} = 0, \text{ for } j \geq 2$$

The solution to the system of differential equations

$$P_{1,j}(t) = e^{-\lambda t}\left(1 - e^{-\lambda t}\right)^{j-1}, \text{ for } j = 1, 2 \ldots$$

For the process started with i individuals, the transition function is

$$P_{ij}(t) = \begin{pmatrix} j-1 \\ i-1 \end{pmatrix} e^{-\lambda i t}\left(1 - e^{-\lambda t}\gamma^{j-i}, \text{ for } j \geq i\right.$$

8.4.2 Long-term stationary distribution and Embedding stationary distribution

The probability π_j is the long-term proportion of time that the process spends in state j. On the other hand, the embedded chain stationary probability ψ_j is the long-term proportion of transitions that the process makes into state j.Each stationary distribution can be derived from the other.

$$\psi_j = \frac{\widetilde{\psi}_j}{\sum_k \widetilde{\psi}_k} = \frac{\pi_j q_j}{\sum_k \pi_k q_k}$$
$$\pi_j = \frac{\psi_j / q_j}{\sum_k \psi_k / q_k}, \text{ for all } j$$

8.5 Queueing Theory

A standard notation of the form A/B/n is used to describe a queueing model, where A denotes the arrival time distribution, B the service time distribution, and n the number of servers.

The M/M/1 queue is a basic model. The M stands for Markov or memoryless. In this model, both arrival and service times have exponential distributions, and there is one server. The M/M/1 queue is a birth-and-death process with constant birth and death rates.

△ **Theorem**

Little's Formula

In a queueing system, let L denote the long-term average number of customers in the system, λ the rate of arrivals, and W the long-term average time that a customer is in the system. Then,

$$L = \lambda W$$

8.5.1 M/M/c Queue

An M/M/c queue has c servers. Consider the dynamics of the process. If there are $0 < k \leq c$ customers then k servers are busy. The number of customers will decrease by one, the first time one of the servers completes their service. The time until that happens is the minimum of k independent exponential random variables, which has an exponential distribution with parameter $k\mu$. If there are more than c customers in the system, then the time until the first service time is complete is the minimum of c independent exponential random variables, and thus is exponentially distributed with parameter $c\mu$

The M/M/c queue is a birth-and-death process with parameters $\lambda_i = \lambda$, for all i and

$$\mu_i = \begin{cases} i\mu, & \text{for } i = 1, \dots, c \\ c\mu, & \text{for } i = c+1, c+2, \dots \end{cases}$$

The stationary probabilities are:

$$\pi_k = \begin{cases} \frac{\pi_0}{k!}\left(\frac{\lambda}{\mu}\right)^k, & \text{for } 0 \leq k < c \\ \frac{\pi_0}{c^{k-c}c!}\left(\frac{\lambda}{\mu}\right)^k, & \text{for } k \geq c \end{cases}$$

with

$$\pi_0 = \left(\sum_{k=0}^{c-1} \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} + \frac{(\lambda/\mu)^c}{c!} \left(\frac{1}{1-\lambda/c\mu}\right) \right)^{-1}$$

· **Example**

(At the hair salon) A hair salon has five chairs. Customers arrive at the salon at the rate of 6 per hour. The hair stylists each take, on average, half an hour to service a customer, independent of arrival times.

1. Jill, the owner, wants to know the long-term probability that no customers are in the salon.

2. Danny, a potential customer, wants to know the average waiting time for a haircut.

3. Leslie, a hair stylist, wants to know the long-term expected number of customers in the salon.

Solution: The system is an M/M/5 queue with $\lambda = 6, c = 5$, and $\mu = 2$

1. The long-term probability that no customers are in the salon is

$$\pi_0 = \left(\sum_{k=0}^{4} \frac{3k}{k!} + \frac{3^5}{5!} \left(\frac{1}{1 - 6/10} \right) \right)^{-1} = \frac{16}{343} = 0.0466$$

2. The long-time average waiting time in the queue, in the notation of Litle's formula, is $W_q = L_q/\lambda$. To find L_q, the expected number of customers in the queue, observe that there are k people in the queue if and only if there are $k + c$ customers in the system. This gives

$$L_q = \sum_{k=c}^{\infty} (k - c)\pi_k = \sum_{k=c}^{\infty} (k - c) \frac{\pi_0}{c^{k-c} c!} \left(\frac{\lambda}{\mu} \right)^k$$

$$= \frac{\pi_0}{c!} \left(\frac{\lambda}{\mu} \right)^c \sum_{k=0}^{\infty} k \left(\frac{\lambda}{c\mu} \right)^k$$
$$= \frac{\pi_0}{c!} \left(\frac{\lambda}{\mu} \right)^c \frac{\lambda}{c\mu} \left(\frac{1}{1-\lambda/c\mu} \right)^2$$
$$= \frac{\pi_0 3^5}{5!} \left(\frac{6}{10} \right) \left(\frac{1}{1-6/10} \right)^2$$
$$= 0.35423$$

3. The long-term expected waiting time in the system is

$$W = W_q + W_s = W_q + \frac{1}{\mu} = 0.059 + 0.5 = 0.559$$

or about 33.54 minutes. The expected number of customers in the system, by Little's formula, is

$$L = \lambda W = 6(0.559) = 3.354$$

· **Example**

For an M/M/∞ queue with $\lambda = \mu = 1$ find the mean time until state 4 is hit for the process started in state 1.

Solution: Set state 4 as absorbing state. Let $(X_t)_{t\geq 0}$ be an absorbing continuous-time Markov chain on $\{0, \ldots, 4\}$, thus, we can write the generator $\mathbf{Q}$ in canonical block matrix form as

$$\left|\begin{array}{cccccc} r/c & s_4 & s_0 & s_1 & s_2 & s_3 \\ s_4 & 0 & 0 & 0 & 0 & 0 \\ s_0 & 0 & -1 & 1 & 0 & 0 \\ s_1 & 0 & 1 & -2 & 1 & 0 \\ s_2 & 0 & 0 & 2 & -3 & 1 \\ s_3 & 1 & 0 & 0 & 3 & -4 \end{array}\right|$$

where the transition rate from i to i-1 is $i\mu$, and λ from i to i+1. Therefore, we can find fundamental matrix as $\mathbf{F}$:

$$(-V)^{-1} = \left(- \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 2 & -3 & 1 \\ 0 & 0 & 3 & -4 \end{pmatrix} \right)^{-1} = \begin{pmatrix} 10 & 9 & 4 & 1 \\ 9 & 9 & 4 & 1 \\ 8 & 8 & 4 & 1 \\ 6 & 6 & 3 & 1 \end{pmatrix}$$

with row sums (24, 23, 21, 16). The desired mean time is 23.

8.6 Poisson Subordination

Consider a finite-state, irreducible, discrete-time Markov chain $Y_0, Y_1, \ldots$ with transition matrix R. Let $(N_t)_{t\geq 0}$ be a Poisson process with parameter λ, which is independent of the

Markov chain. Define a continuous-time process $(X_t)_{t\geq 0}$ by $X_t = Y_{N_t}$. That is, transitions for the X_t process occur at the arrival times of the Poisson process. From state i, the process holds an exponentially distributed amount of time with parameter λ and then transitions to j with probability R_{ij}.

The process $(X_t)_{\mathrm{t}>0}$ is a continuous-time Markov chain whose transition function $P(t)$ has a surprisingly simple form. By conditioning on N_t:

$$\begin{aligned}
P_{ij}(t) &= P\left(X_t = j | X_0 = i\right) \\
&= \sum_{k=0}^{\infty} P\left(X_t = j | N_t = k, X_0 = i\right) P\left(N_t = k | X_0 = i\right) \\
&= \sum_{k=0}^{\infty} P\left(Y_k = j | N_t = k, X_0 = i\right) P\left(N_t = k\right) \\
&= \textstyle\sum_{k=0}^{\infty} P\left(Y_k = j | Y_0 = i\right) P\left(N_t = k\right) \\
&= \textstyle\sum_{k=0}^{\infty} R_{ij}^k \frac{e^{-\lambda t}(\lambda t)^k}{k!}
\end{aligned}$$

We say that the $(X_t)_{t\geq 0}$ Markov chain is subordinated to a Poisson process.

Consider a continuous-time Markov chain with generator Q and holding time parameters $q_1, q_2, \ldots$ Assume the parameters are uniformly bounded. That is, there exists a constant λ such that $q_i \leq \lambda$, for all i. This will always be the case if the chain is finite, and we can take $\lambda = \max_i q_i$. Let

$$R = \frac{1}{\lambda} Q + I$$

The matrix R is a stochastic matrix. Entries are non-negative and rows sum to 1. The transition function can be given in terms of R, as

$$\begin{aligned}
P(t) = e^{tQ} = e^{-\lambda t} e^{tQ} e^{\lambda t} &= e^{-\lambda t} e^{t(Q+\lambda I)} \\
&= e^{-\lambda t} \sum_{k=0}^{\infty} \frac{1}{k!} t^k (Q + \lambda I)^k \\
&= \sum_{k=0}^{\infty} \left(\frac{1}{\lambda} Q + I\right)^k \frac{e^{-\lambda t}(\lambda t)^k}{k!} \\
&= \sum_{k=0}^{\infty} R^k \frac{e^{-\lambda t}(\lambda t)^k}{k!}
\end{aligned}$$

The entries of **R** matrix are:

$$R_{ij} = \begin{cases} q_{ij}/\lambda, & \text{for } i \neq j \\ 1 - q_i/\lambda, & \text{for } i = j \end{cases}$$

Poisson subordination can be described as follows. From a given state i, wait an exponential length of time with rate λ. Then, flip a coin whose heads probability is q_i/λ. If heads, transition to a new state according to the R matrix. If tails, stay at i and repeat. Thus, holding time parameters are constant, and transitions, or pseudo-transitions, from a state to itself are allowed.

8.6.1 Long-Term Behavior, Simulation, Computation

For a Markov chain subordinated to a Poisson process, the discrete R-chain has the same stationary distribution as the original chain.

Poisson subordination leads to an efficient method to simulate a continuous-time Markov chain, since transitions rates are constant and thus not dependent on the current state.Please refer to the Python code for simulating the following example:

· **Example**

Consider a Markov chain with generator

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -3 & 2 \\ 0 & 1 & -1 \end{pmatrix}$$

Letting $\lambda = \max\{q_1, q_2, q_3\} = 3$ gives

$$R = \frac{1}{3}Q + I = \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 0 & 2/3 \\ 0 & 1/3 & 2/3 \end{pmatrix}$$

Observe that $\pi = (1/8, 2/8, 5/8)$ is the stationary distribution for both the original chain and the discrete-time R -chain.

we can simulate the distribution of $X_{1.5}$ for the Markov chain

9 Brownian Motion

The mathematical object we call Brownian motion is a continuous-time, continuous-state stochastic process, also called the Wiener process, named after the American mathematician Norbert Wiener.

△ **Definition**

Standard Brownian Motion

A continuous-time stochastic process $(B_t)_{t \geq 0}$ is a standard Brownian motion if it satisfies the following properties:

1. $B_0 = 0$

2. (Normal distribution) For $t > 0, B_t$ has a normal distribution with mean 0 and variance t.

3. (Stationary increments) For $s, t > 0, B_{t+s} - B_s$ has the same distribution as B_t. That is,

$$P\left(B_{t+s} - B_s \leq z\right) = P\left(B_t \leq z\right) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx$$

4. (Independent increments) If $0 \leq q < r \leq s < t$, then $B_t - B_s$ and $B_r - B_q$ are independent random variables.

5. (Continuous paths) The function $t \mapsto B_t$ is continuous, with probability 1 .

Brownian motion can be thought of as the motion of a particle that diffuses randomly along a line. At each point t, the particle's position is normally distributed about the line with variance t. As t increases, the particle's position is more diffuse.

· **Example**

For $0 < s < t$, find the distribution of $B_s + B_t$

Solution: Write $B_s + B_t = 2B_s + (B_t - B_s)$. The sum of independent normal variables is normal. Thus, $B_s + B_t$ is normally distributed with mean $E\left(B_s + B_t\right) = E\left(B_s\right) + E\left(B_t\right) = 0$, and variance

$$\begin{aligned}\operatorname{Var}(B_s+B_t) &= \operatorname{Var}(2B_s+(B_t-B_s)) = \operatorname{Var}(2B_s)+\operatorname{Var}(B_t-B_s)\\ &= 4\operatorname{Var}(B_s)+\operatorname{Var}(B_{t-s}) = 4s+(t-s)\\ &= 3s+t\end{aligned}$$

· **Example**

Find the covariance of B_s and B_t

Solution: For the covariance,

$$\operatorname{Cov}(B_s,B_t) = E(B_sB_t)-E(B_s)E(B_t) = E(B_sB_t)$$

For $s<t$, write $B_t=(B_t-B_s)+B_s$,, which gives

$$\begin{aligned}E(B_sB_t) &= E(B_s(B_t-B_s+B_s))\\ &= E(B_s(B_t-B_s))+E\left(B_s^2\right)\\ &= E(B_s)E(B_t-B_s)+E\left(B_s^2\right)\\ &= 0+\operatorname{Var}(B_s)=s\end{aligned}$$

Thus, Cov $(B_s,B_t)=s$. For $t<s$, by symmetry $\operatorname{Cov}(B_s,B_t)=t$. In either case,

$$\operatorname{Cov}(B_s,B_t)=\min\{s,t\}$$

9.1 Simulating Brownian Motion

Consider simulating Brownian motion on $[0,t]$. Assume that we want to generate n variables at equally spaced time points, that is $B_{t_1},B_t,\ldots,B_{t_n}$, where $t_i=it/n$, for $i=1,2,\ldots,n$. By stationary and independent increments, with $B_{t_0}=B_0=0$

$$B_{t_i}=B_{t_{i-1}}+\left(B_{t_i}-B_{t_{i-1}}\right)\stackrel{d}{=}B_{t_{i-1}}+X_i$$

where X_i is normally distributed with mean 0 and variance $t_i-t_{i-1}=t/n$, and is independent of $B_{t_{i-1}}$. The notation $X\stackrel{d}{=}Y$ means that random variables X and Y have the same distribution.

This leads to a recursive simulation method. Let $Z_1, Z_2, \ldots, Z_n$ be independent and identically distributed standard normal random variables. Set

$$B_{t_i} = B_{t_{i-1}} + \sqrt{t/n} Z_i, \text{ for } i = 1, 2, \ldots, n$$

$$B_{t_i} = \sqrt{\frac{t}{n}} \left(Z_1 + \cdots + Z_n\right)$$

More generally, to simulate $B_{t_1}, B_{t_2}, \ldots, B_{t_n}$, for time points $t_1 < t_2 < \cdots < t_n$, set

$$B_{t_i} = B_{t_{i-1}} + \sqrt{t_i - t_{i-1}} Z_i, \text{ for } i = 1, 2, \ldots, n$$

with $t_0 = 0$.

9.1.1 Sample Space for Brownian Motion and Continuous Paths

Consider the fifth defining property of Brownian motion: the function $t \mapsto B_t$ is continuous.

In the context of stochastic processes, $X_t = X_t(\omega)$ is a function of two variables: t and ω. For fixed $t \in \mathbb{R}, X_t$ is a random variable. Letting $X_t(\omega)$ vary as $\omega \in \Omega$ generates the different values of the process at the fixed time t. On the other hand, for fixed $\omega \in \Omega, X_t(\omega)$ is a function of t. Letting t vary generates a sample path or realization. One can think of these realizations as random functions.

A probability space Ω is easy enough to identify. Let Ω be the set of all continuous functions on $[0, \infty)$. Each $\omega \in \Omega$ is a continuous function. Then, $B_t(\omega) = \omega_t$, the value of ω evaluated at t. This is the easiest way to insure that B_t has continuous sample paths. The hard part is to construct a probability function P on the set of continuous functions, which is consistent with the properties of Brownian motion. This was precisely Norbert Wiener's contribution. That probability function, introduced by Wiener in 1923, is called Wiener measure.

9.2 Brownian Motion and Random Walk

The simple random walk is a discrete-state process. To obtain a continuous-time process with continuous sample paths, values are connected by linear interpolation. Recall that $\lfloor x \rfloor$ is the floor of x, or integer part of x, which is the largest integer not greater than x. Extending the definition of S_t to real $t \geq 0$, let

$$S_t = \begin{cases} X_1 + \cdots + X_t, & \text{if } t \text{ is an integer} \\ S_{\lfloor t \rfloor} + X_{\lfloor t \rfloor + 1}(t - \lfloor t \rfloor), & \text{otherwise} \end{cases}$$

Observe that if k is a positive integer, then for $k \leq t \leq k+1$, S_t is the linear interpolation of the points (k, S_k) and $(k+1, S_{k+1})$.

The process is now scaled both vertically and horizontally. Let $S_t^{(n)} = S_{nt}/\sqrt{n}$, for $n = 1, 2, \ldots$ On any interval, the new process has n times as many steps as the original walk. And the height at each step is shrunk by a factor of $1/\sqrt{n}$.

The scaling preserves mean and variance, as $E\left(S_t^{(n)}\right) = 0$ and $\operatorname{Var}\left(S_t^{(n)}\right) = \operatorname{Var}\left(S_{nt}\right)/n \approx t$.Sample paths are continuous and for each n, the process retains independent and stationary increments. Considering the central limit theorem, it is reasonable to think that as $n \to \infty$, the process converges to Brownian motion.

9.2.1 Invariant Principle

The construction of Brownian motion from simple symmetric random walk can be generalized so that we start with any i.i.d. sequence $X_1, \ldots, X_n$ with mean 0 and variance 1. Let $S_n = X_1 + \cdots + X_n$. Then, $S_{nt}/\sqrt{n}$ converges to B_t, as $n \to \infty$. This is known as Donsker's invariance principle.

If g is a bounded, continuous function, whose domain is the set of continuous functions on $[0,1]$, then $g\left(S_{nt}/\sqrt{n}\right) \approx g\left(B_t\right)$, for large n.

Functions whose domain is a set of functions are called functionals. The invariance principle means that properties of random walk and of functionals of random walk can be derived by considering analogous properties of Brownian motion, and vice versa.

· **Example**

For a simple symmetric random walk, consider the maximum value of the walk in the first n steps. Let $g(f) = \max_{0\leq t\leq 1} f(t)$. By the invariance principle,

$$\lim_{n\to\infty} g\left(\frac{S_{tn}}{\sqrt{n}}\right) = \lim_{n\to\infty} \max_{0\leq t\leq 1} \frac{S_{tn}}{\sqrt{n}} = \lim_{n\to\infty} \max_{0\leq k\leq n} \frac{S_k}{\sqrt{n}} = g(B_t) = \max_{0\leq t\leq 1} B_t$$

This gives $\max_{0\leq k\leq n} S_k \approx \sqrt{n}\max_{0\leq t\leq 1} B_t$, for large n. It is shown that the random variable $M = \max_{0\leq t\leq 1} B_t$ has density function:

$$f_M(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}, \text{ for } x > 0$$

Mean and standard deviation are

$$E(M) = \sqrt{\frac{2}{\pi}} \approx 0.80 \quad \text{and} \quad SD(M) = \frac{\pi - 2}{\pi} \approx 0.60$$

With these results, we see that in the first n steps of simple symmetric random walk, the maximum value is about $(0.80)\sqrt{n}$ give or take $(0.60)\sqrt{n}$. In $n = 10,000$ steps, the probability that a value greater than 200 is reached is

$$\begin{aligned} P\left(\max_{0\leq k\leq n} S_k > 200\right) &= P\left(\max_{0\leq k\leq n} \frac{S_k}{100} > 2\right) \\ &= P\left(\max_{0\leq k\leq n} \frac{S_k}{\sqrt{n}} > 2\right) \\ &\approx P(M > 2) \\ &= \int_2^\infty \sqrt{\frac{2}{\pi}} e^{-x^2/2} dx = 0.0455 \end{aligned}$$

9.3 Gaussian Process

Brownian Motion is a Gaussian process.

Δ Definition

Multivariate Normal Distribution

Random variables $X_1, \ldots, X_k$ have a multivariate normal distribution if for all real numbers $a_1, \ldots, a_k$, the linear combination

$$a_1X_1 + \cdots + a_kX_k$$

has a univariate normal distribution. A multivariate normal distribution is completely determined by its mean vector

$$\mu = (\mu_1, \ldots, \mu_k) = (E(X_1), \ldots, E(X_k))$$

and covariance matrix $\boldsymbol{V}$, where

$$V_{ij} = \mathrm{Cov}(X_i, X_j), \text{ for } 1 \leq i, j \leq k$$

The joint density function of the multivariate normal distribution is

$$f(x) = \frac{1}{(2\pi)^{k/2}|V|^{1/2}} \exp\left(-\frac{1}{2}(x-\mu)^T V^{-1}(x-\mu)\right)$$

where $x = (x_1, \ldots, x_k)$ and $|V|$ is the determinant of V

A Gaussian process extends the multivariate normal distribution to stochastic processes.

Δ Definition

A Gaussian process $(X_t)_{t\geq 0}$ is a continuous-time stochastic process with the property that for all $n = 1, 2, \ldots$ and $0 \leq t_1 < \cdots < t_n$, the random variables $X_{t_1}, \ldots, X_{t_n}$ have a multivariate normal distribution.

A Gaussian process is completely determined by its mean function $E(X_t)$, for $t \geq 0$, and covariance function $\mathrm{Cov}(X_s, X_t)$, for $s, t \geq 0$

Δ Definition

Gaussian Process and Brownian Motion

A stochastic process $(B_t)_{t\geq 0}$ is a standard Brownian motion if and only if it is a Gaussian process with the following properties:

1. $B_0 = 0$
2. (Mean function) $E(B_t) = 0$, for all t
3. (Covariance function) $\operatorname{Cov}(B_s, B_t) = \min\{s, t\}$, for all s, t
4. (Continuous paths) The function $t \mapsto B_t$ is continuous, with probability 1.

9.3.1 Nowhere differentiable path

Brownian motion preserves its character after rescaling. For instance, given a standard Brownian motion on [0, 1], if we look at the process on, say, an interval of length one-trillionth then after resizing by a factor of $1/\sqrt{10^{-12}} = 1{,}000{,}000$, what we see is indistinguishable from the original Brownian motion.

This highlights the invariance, or fractal, structure of Brownian motion sample paths. It means that the jagged character of these paths remains jagged at all time scales. This leads to the remarkable fact that Brownian motion sample paths are nowhere differentiable.

9.4 Transformation and Properties

△ Definition

Transformation of Brownian motion

Let $(B_t)_{t\geq 0}$ be a standard Brownian motion. Then, each of the following transformations is a standard Brownian motion.

1. (Reflection) $(-B_t)_{t\geq 0}$.
2. (Translation) $(B_{t+s} - B_s)_{t\geq 0}$, for all $s \geq 0$.
3. (Rescaling) $\left(a^{-1/2} B_{at}\right)_{t\geq 0}$ for all $a > 0$.

4. (Inversion) The process $(X_t)_{t\geq 0}$ defined by $X_0 = 0$ and $X_t = tB_{1/t}$, for $t > 0$.

· **Example**

Let $(X_t)_{T\geq 0}$ be a Brownian motion process started at $x = 3$. Find $P(X_2 > 0)$.

Solution: Write $X_t = B_t + 3$. Then,

$$P(X_2 > 0) = P(B_2 + 3 > 0) = P(B_2 > -3) = \int_{-3}^{\infty} \frac{1}{\sqrt{4\pi}} e^{-x^2/4} dx = 0.983$$

9.4.1 Markov Process

For a continuous-state Markov process, the transition function, or transition kernel, $K_t(x, y)$ plays the role that the transition matrix plays for a discrete-state Markov chain. The function $K_t(x, \cdot)$ is the conditional density of X_t given $X_0 = x$. If $(X_t)_{t\geq 0}$ is Brownian motion started at x, then X_t is normally distributed with mean x and variance t. The transition kernel is

$$K_t(x, y) = \frac{1}{\sqrt{2\pi t}} e^{-(y-x)^2/2t}$$

The transition kernel of a Markov process satisfies the Chapman–Kolmogorov equations. For continuous-state processes, integrals replace sums. The equations are

$$K_{s+t}(x, y) = \int_{-\infty}^{\infty} K_s(x, z) K_t(z, y) dz, \text{ for all } s, t$$

9.4.2 First Hitting Time and Strong Markov Property

By the strong Markov property this also holds for some random times as well. If S is a stopping time, $(B_{t+S})_{t\geq 0}$ is a Brownian motion process.

At any time t, B_t is equally likely to be above or below the line $y = 0$. Assume that $a > 0$. For Brownian motion started at a, the process is equally likely to be above or below the line $y = a$. This gives,

$$P(B_t > a | T_a < t) = P(B_t > 0) = \frac{1}{2}$$

$$\frac{1}{2} = P(B_t > a | T_a < t) = \frac{P(B_t > a, T_a < t)}{P(T_a < t)} = \frac{P(B_t > a)}{P(T_a < t)}$$

The last equality is because the event $\{B_t > a\}$ implies $\{T_a < t\}$ by continuity of sample paths. We have that

$$\begin{aligned} P(T_a < t) &= 2P(B_t > a) \\ &= 2\int_a^\infty \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx \\ &= 2\int_{a/\sqrt{t}}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx \end{aligned}$$

If $a < 0$, the argument is similar with $1/2 = P(B_t < a | T_a < t)$. In either case,

$$P(T_a < t) = 2\int_{|a|/\sqrt{t}}^\infty \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

△ Definition

First Hitting Time Distribution

For a standard Brownian motion, let T_a be the first time the process hits level a. The density function of T_a is

$$f_{T_a}(t) = \frac{|a|}{\sqrt{2\pi t^3}} e^{-a^2/2t}, \text{ for } t > 0$$

9.4.3 Reflection Principle and the Maximum of Brownian Motion

Brownian motion reflected at a first hitting time is a Brownian motion. This property is known as the reflection principle and is a consequence of the strong Markov property.

The reflection principle is applied in the following derivation of the distribution of $M_t = \max_{0 \le s \le t} B_s$, the maximum value of Brownian motion on $[0, t]$

Let $a > 0$. If at time t, B_t exceeds a, then the maximum on $[0, t]$ is greater than a. That is, $\{B_t > a\}$ implies $\{M_t > a\}$. This gives

$$\begin{aligned} \{M_t > a\} &= \{M_t > a, B_t > a\} \cup \{M_t > a, B_t \le a\} \\ &= \{B_t > a\} \cup \{M_t > a, B_t \le a\} \end{aligned}$$

As the union is disjoint, $P(M_t > a) = P(B_t > a) + P(M_t > a, B_t \leq a)$ Consider a sample path that has crossed a by time t and is at most a at time t. By the reflection principle, the path corresponds to a reflected path that is at least a at time t. This gives that $P(M_t > a, B_t \leq a) = P\left(\tilde{B}_t \geq a\right) = P(B_t > a)$. Thus,

$$P(M_t > a) = 2P(B_t > a) = \int_a^\infty \sqrt{\frac{2}{\pi t}} e^{-x^2/2t} dx, \text{ for } a > 0$$

or

$$\begin{aligned} P(M_t > a) = P(T_a < t) &= \int_0^t \frac{a}{\sqrt{2\pi s^3}} e^{-a^2/2s} ds \\ &= \int_a^\infty \sqrt{\frac{2}{\pi t}} e^{-x^2/2t} dx, \text{ for } a > 0 \end{aligned}$$

The last equality is achieved by the change of variables $a^2/s = x^2/t$

· **Example**

A laboratory instrument takes annual temperature measurements. Measurement errors are assumed to be independent and normally distributed. As precision decreases over time, errors are modeled as standard Brownian motion. For how many years can the lab be guaranteed that there is at least 90% probability that all errors are less than 4 degrees?

Solution: The problem asks for the largest t such that $P(M_t \leq 4) \geq 0.90$. We have

$$0.90 \leq P(M_t \leq 4) = 1 - P(M_t > 4) = 1 - 2P(B_t > 4) = 2P(B_t \leq 4) - 1$$

$$0.95 \leq P(B_t \leq 4) = P\left(Z \leq \frac{4}{\sqrt{t}}\right)$$

where Z is a standard normal random variable. The 95 th percentile of the standard normal distribution is 1.645. Solving $4/\sqrt{t} = 1.645$ gives

$$t = \left(\frac{4}{1.645}\right)^2 = 5.91 \text{ years.}$$

9.4.4 Zeros of Brownian Motion and Arcsine Distribution

△ Theorem

For $0 \leq r < t$, let $z_{r,t}$ be the probability that standard Brownian motion has at least one zero in $(\ r, t)$. Then,

$$z_{r,t} = \frac{2}{\pi} \arccos(\sqrt{\frac{r}{t}})$$

With $r = 0$, the result gives that standard Brownian motion has at least one zero in $(0, \epsilon)$ with probability

$$z_{0,e} = (2/\pi) \arccos(0) = (2/\pi)(\pi/2) = 1$$

That is, $B_t = 0$, for some $0 < t < \epsilon$. By the strong Markov property, for Brownian motion restarted at t, there is at least one zero in (t, ϵ), with probability 1. Continuing in this way, there are infinitely many zeros in $(0, \epsilon)$.

The arcsine distribution arises in the proof of the Theorem above and other results related to the zeros of Brownian motion. The distribution is a special case of the beta distribution, and has cumulative distribution function

$$F(t) = \frac{2}{\pi} \arcsin(\sqrt{t}), \text{ for } 0 \leq t \leq 1$$

PDF is

$$f(t) = F'(t) = \frac{1}{\pi\sqrt{t(1-t)}}, \text{ for } 0 \leq t \leq 1$$

△ Theorem

Last Zero Standing

Let L_t be the last zero in $(0, t)$. Then,

$$P(L_t \leq x) = \frac{2}{\pi} \arcsin(\sqrt{\frac{x}{t}}),$$

for $0 < x < t$

9.5 Variations and Applications

△ **Definition**

Brownian Process with Drift For real μ and $\sigma > 0$, the process defined by

$$X_t = \mu t + \sigma B_t, \text{ for } t \geq 0$$

is called Brownian motion with drift parameter μ and variance parameter σ^2.

Brownian motion with drift is a Gaussian process with continuous sample paths and independent and stationary increments. For $s, t > 0, X_{t+s} - X_t$ is normally distributed with mean μs and variance $\sigma^2 s$.

· **Example**

Home team advantage

A novel application of Brownian motion to sports scores is given in Stern (1994). The goal is to quantify the home team advantage by finding the probability in a sports match the home team wins the game given that they lead by l points after a fraction $0 \leq t \leq 1$ of the game is completed. The model is applied to basketball where can be reasonably approximated by a continuous distribution.

For $0 \leq t \leq 1$, let X_t denote the difference in scores between the home and visiting teams after $100t$ percent of the game has been completed. The process is modeled as a Brownian motion with drift, where the mean parameter μ is a measure of home team advantage. The probability that the home team wins the game, given that they have an l point lead at time $t < 1$, is

$$\begin{aligned}
p(l,t) &= P\left(X_1 > 0 | X_t = l\right) = P\left(X_1 - X_t > -l\right) \\
&= P\left(X_{1-t} > -l\right) = P\left(\mu(1-t) + \sigma B_{1-t} > -l\right) \\
&= P\left(B_{1-t} < \frac{l + \mu(1-t)}{\sigma}\right) \\
&= P\left(B_t < \frac{\sqrt{t}[l + \mu(1-t)]}{\sigma\sqrt{1-t}}\right)
\end{aligned}$$

The last equality is because B_t has the same distribution as $\sqrt{t/(1-t)}B_{1-t}$

9.5.1 Brownian Bridge

Δ Definition

From standard Brownian motion, the conditional process $(B_t)_{0\leq t\leq 1}$ given that $B_1 = 0$ is called a Brownian bridge. The Brownian bridge is tied down to 0 at the endpoints of $[0,1]$.

For $0 \leq t \leq 1$, the distribution of X_i is equal to the conditional distribution of B given $B_1 = 0$. Since theconditional distributions for Gaussian process are Gaussian, it follows that Brownian bridge is a Gaussian process. Continuity of sample paths, and independent and stationary increments are inherited from standard Brownian motion.

To find the mean and covariance functions results are needed for bivariate normal distributions.

$$E\left(B_s|B_t=y\right)=\frac{sy}{t} \quad \text{and} \quad \operatorname{Var}\left(B_s|B_t=y\right)=\frac{s(t-s)}{t}$$

for $0<s<t$. This gives gives the mean function of Brownian bridge

$$E\left(X_t\right)=E\left(B_t|B_1=0\right)=0,\ \text{for } 0\leq t\leq 1$$

$$\begin{aligned}
E\left(X_sX_t\right) &= E\left(E\left(X_sX_t\right)|X_t\right)=E\left(X_tE\left(X_s|X_t\right)\right) \\
&= E\left(X_t\frac{sX_t}{t}\right)=\frac{s}{t}E\left(X_t^2\right)=\frac{s}{t}E\left(B_t^2|B_1=0\right) \\
&= \frac{s}{t}\operatorname{Var}\left(B_t|B_1=0\right)=\left(\frac{s}{t}\right)\frac{t(1-t)}{1}=s-st
\end{aligned}$$

By symmetry, for $t<s$, $E\left(X_sX_t\right)=t-$st. In either case, the covariance function is

$$\operatorname{Cov}\left(X_s,X_t\right)=\min\{s,t\}-st$$

· **Example**

Let $X_t = B_t - tB_1$, for $0 \le t \le 1$. Show that $(X_t)_{0\le t\le 1}$ is a Brownian bridge.

Solution: The process is a Gaussian process since $(B_t)_{t\ge 0}$ is a Gaussian process. Sample paths are continuous, with probability 1. It is suffice to show that the process has the same mean and covariance functions as a Brownian bridge.

The mean function is $E(X_t) = E(B_t - tB_1) = E(B_t) - tE(B_1) = 0$. The covariance function is

$$\begin{aligned}\operatorname{Cov}(X_s, X_t) &= E(X_s X_t) = E((B_s - sB_1)(B_t - tB_1)) \\ &= E(B_s B_t) - tE(B_s B_1) - sE(B_t B_1) + \mathrm{st}\, E\left(B_1^2\right) \\ &= \min\{s,t\} - ts - st + st = \min\{s,t\} - st\end{aligned}$$

The construction described in this example gives a direct method for simulating a Brownian bridge

9.5.2 Geometric Brownian Motion

△ **Definition**

Let $(X_t)_{t\ge 0}$ be a Brownian motion with drift parameter μ and variance parameter σ^2. The process $(G_t)_{t\ge 0}$ defined by

$$G_t = G_0 e^{X_t}, \text{ for } t \ge 0$$

where $G_0 > 0$, is called geometric Brownian motion.

Taking logarithms, we see that $\ln G_t = \ln G_0 + X_t$ is normally distributed with mean

$$E(\ln G_t) = E(\ln G_0 + X_t) = \ln G_0 + \mu t$$

and variance

$$\operatorname{Var}(\ln G_t) = \operatorname{Var}(\ln G_0 + X_t) = \operatorname{Var}(X_t) = \sigma^2 t$$

A random variable whose logarithm is normally distributed is said to have a lognormal distribution.

$$E(G_t) = G_0 e^{t\left(\mu+\sigma^2/2\right)} \text{ and } \quad \operatorname{Var}(G_t) = G_0^2 e^{2t\left(\mu+\sigma^2/2\right)}\left(e^{t\sigma^2}-1\right)$$

Geometric Brownian motion arises as a model for quantities which can be expressed as the product of independent random multipliers. consider the ratio

$$\frac{G_{t+s}}{G_t} = \frac{G_0 e^{\mu(t+s)+\sigma X_{t+s}}}{G_0 e^{\mu t+\sigma X_t}} = e^{\mu s+\sigma(X_{t+s}-X_t)}$$

which has the same distribution as $e^{\mu s+\sigma X_s} = G_s/G_0$. For $0 \leq q < r \leq s < t$

$$\frac{G_t}{G_s} = e^{\mu(t-s)+\sigma(X_t-X_s)} \text{ and } \quad \frac{G_r}{G_q} = e^{\mu(r-q)+\sigma(X_r-X_q)}$$

are independent random variables, because of independent increments for $(X_t)_{t\geq 0}$.

Let $Y_k = G_k/G_{k-1}$, for $k = 1, 2, \ldots$ Then, $Y_1, Y_2, \ldots$ is an i.i.d. sequence, and

$$G_n = \left(\frac{G_n}{G_{n-1}}\right)\left(\frac{G_{n-1}}{G_{n-2}}\right)\cdots\left(\frac{G_2}{G_1}\right)\left(\frac{G_1}{G_0}\right)G_0 = G_0 Y_1 Y_2 \cdots Y_{n-1} Y_n$$

· **Example**

Stock Price Model

Historical data for many stocks indicate long-term exponential growth or decline. Prices cannot be negative and geometric Brownian motion takes only positive values. Let Y_t denote the price of a stock after t days. Since the price on a given day is probably close to the price on the next day (assuming normal market conditions), stock prices are not independent. However, the percent changes in price from day to day Y_t/Y_{t-1}, for $t = 1, 2, \ldots$ might be reasonably modeled as independent and identically distributed. This leads to geometric Brownian motion. In the context of stock prices, the standard deviation σ is called the volatility. Assume that XYZ stock currently sells for \$80 a share and follows a geometric Brownian motion with drift parameter 0.10 and volatility 0.50. Find the probability that in 90 days the price of XYZ will rise to at least \$100.

Let Y_t denote the price of XYZ after t years. Round 90 days as 1/4 of a year, then

$$\begin{aligned} P\left(Y_{0.25} \geq 100\right) &= P\left(80e^{\mu(0.25)+\sigma B_{0.25}} \geq 100\right) \\ &= P\left((0.1)(0.25) + (0.5)B_{0.25} \geq \ln 1.25\right) \\ &= P\left(B_{0.25} \geq 0.396\right) = 0.214 \end{aligned}$$

9.6 Martingales

△ Definition

A stochastic process $(Y_t)_{t\geq 0}$ is a martingale, if for all $t \geq 0$,

1. $E\left(Y_t | Y_r, 0 \leq r \leq s\right) = Y_s$, for all $0 \leq s \leq t$
2. $E\left(|Y_t|\right) < \infty$.

A most important property of martingales is that they have constant expectation. By the law of total expectation,

$$E\left(Y_t\right) = E\left(E\left(Y_t | Y_r, 0 \leq r \leq s\right)\right) = E\left(Y_s\right)$$

for all $0 \leq s \leq t$. That is,

$$E\left(Y_t\right) = E\left(Y_0\right), \text{ for all } t$$

· Example

Show that simple symmetric random walk is a martingale.

Solution:

$$X_i = \begin{cases} +1, & \text{with probability } 1/2 \\ -1, & \text{with probability } 1/2 \end{cases}$$

for $i = 1, 2, \ldots$ For $n \geq 1$, let $S_n = X_1 + \cdots + X_n$, with $S_0 = 0$. Then,

$$\begin{aligned} E\left(S_{n+1} | S_0, \ldots, S_n\right) &= E\left(X_{n+1} + S_n | S_0, \ldots, S_n\right) \\ &= E\left(X_{n+1} | S_0, \ldots, S_n\right) + E\left(S_n | S_0, \ldots, S_n\right) \\ &= E\left(X_{n+1}\right) + S_n = S_n \end{aligned}$$

The third equality is because X_{n+1} is independent of $X_1, \ldots, X_n$, and thus independent of $S_0, S_1, \ldots, S_n$. The fact that $E\left(S_n | S_0, \ldots, S_n\right) = S_n$ is a consequence of a general property of conditional expectation, which states that if X is a random variable and g is a function, then $E(g(X)|X) = g(X)$.

The second part of the martingale definition is satisfied as

$$E\left(|S_n|\right) = E\left(\left|\sum_{i=1}^{n} X_i\right|\right) \leq E\left(\sum_{i=1}^{n} |X_i|\right) = \sum_{i=1}^{n} E\left(|X_i|\right) = n < \infty$$

· **Example**

Show that standard Brownian motion $(B_t)_{t\geq 0}$ is a martingale.

Solution:

$$\begin{aligned} E\left(B_t | B_r, 0 \leq r \leq s\right) &= E\left(B_t - B_s + B_s | B_r, 0 \leq r \leq s\right) \\ &= E\left(B_t - B_s | B_r, 0 \leq r \leq s\right) + E\left(B_s | B_r, 0 \leq r \leq s\right) \\ &= E\left(B_t - B_s\right) + B_s = B_s \end{aligned}$$

where the second equality is because of independent increments. Also,

$$E\left(|B_t|\right) = \int_{-\infty}^{\infty} |x| \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx = \int_{0}^{\infty} x \sqrt{\frac{2}{\pi t}} e^{-x^2/2t} dx = \sqrt{\frac{2t}{\pi}} < \infty$$

△ **Theorem**

Martingale with Respect to Another Process

Let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be stochastic processes. Then, $(Y_t)_{t\geq 0}$ is a martingale with respect to $(X_t)_{t\geq 0}$, if for all $t \geq 0$

1. $E\left(Y_t | X_r, 0 \leq r \leq s\right) = Y_s$, for all $0 \leq s \leq t$
2. $E\left(|Y_t|\right) < \infty$

The most common application of this is when Y_t is a function of X_t. That is, $Y_t = g\left(X_t\right)$ for some function g. It is useful to think of the conditioning random variables $(X_r)_{0\leq r\leq s}$ as representing past information, or history, of the process up to time s.

· **Example**

Quadratic martingale

Let $Y_t = B_t^2 - t$, for $t \geq 0$. Show that $(Y_t)_{t\geq 0}$ is a martingale with respect to Brownian motion. This is called the quadratic martingale.

Solution: For $0 \leq s < t$

$$\begin{aligned}
E\left(Y_t | B_r, 0 \leq r \leq s\right) &= E\left(B_t^2 - t | B_r, 0 \leq r \leq s\right) \\
&= E\left(\left(B_t - B_s + B_s\right)^2 | B_r, 0 \leq r \leq s\right) - t \\
&= E\left(\left(B_t - B_s\right)^2 + 2\left(B_t - B_s\right) B_s + B_s^2 | B_r, 0 \leq r \leq s\right) - t \\
&= E\left(\left(B_t - B_s\right)^2\right) + 2 B_s E\left(B_t - B_s\right) + B_s^2 - t \\
&= (t - s) + B_s^2 - t = B_s^2 - s = Y_s
\end{aligned}$$

Again, $E(B_s^2 | B_r) = B_s^2$ is the consequence of a general property of conditional expectation: $E(g(X)|X) = g(X)$.Furthermore,

$$E\left(|Y_t|\right) = E\left(\left|B_t^2 - t\right|\right) \leq E\left(B_t^2 + t\right) = E\left(B_t^2\right) + t = 2t < \infty$$

· **Example**

Let $G_t = G_0 e^{X_t}$ be a geometric Brownian motion, where $(X_t)_{t\geq 0}$ is Brownian motion with drift μ and variance σ^2. Let $r = \mu + \sigma^2/2$. Show that $e^{-rt}G_t$ is a martingale with respect to standard Brownian motion.

Solution: For $0 \leq s < t$

$$\begin{aligned}
E\left(e^{-rt} G_t | B_r, 0 \leq r \leq s\right) &= e^{-rt} E\left(G_0 e^{\mu t + \sigma B_t} | B_r, 0 \leq r \leq s\right) \\
&= e^{-rt} E\left(G_0 e^{\mu(t-s) + \sigma(B_t - B_s)} e^{\mu s + \sigma B_s} | B_r, 0 \leq r \leq s\right) \\
&= e^{-rt} e^{\mu s + \sigma B_s} E\left(G_0 e^{\mu(t-s) + \sigma(B_t - B_s)}\right)
\end{aligned}$$

$$\begin{aligned}
&= e^{-rt} e^{\mu s + \sigma B_s} E\left(G_{t-s}\right) \\
&= e^{-t\left(\mu + \sigma^2/2\right)} e^{\mu s + \sigma B_s} G_0 e^{(t-s)\left(\mu + \sigma^2/2\right)} \\
&= e^{-s\left(\mu + \sigma^2/2\right)} G_0 e^{\mu s + \sigma B_s} \\
&= e^{-rs} G_s
\end{aligned}$$

Also,

$$E\left(\left|e^{-rt}G_t\right|\right) = e^{-rt}E\left(G_t\right) = e^{-t\left(\mu+\sigma^2/2\right)}G_0e^{t\left(\mu+\sigma^2/2\right)} = G_0 < \infty, \text{ for all } t$$

9.6.1 Black-Scholes

There are several critical assumptions underlying the Black-Scholes formula. One is that stock prices follow a geometric Brownian motion. Another is that an investment should be *risk neutral*. What this means is that the expected return on an investment should be equal to the so-called risk-free rate of return, such as what is obtained by a short-term U.S. government bond.

Let r denote the risk-free interest rate. Starting with P dollars, because of compound interest, after t years of investing risk free your money will grow to $P(1+r)^t$ dollars. Under continuous compounding, the future is $F = Pe^{rt}$. This gives the future value of your present dollars. On the other hand, assume that t years from now you will be given F dollars. To find its present value requires discounting the future amount by a factor of e^{-rt}. That is, $P = e^{-rt}F$.

Let G_t denote the price of a stock t years from today. Then, the present value of the stock price is $e^{-rt}G_t$. The Black-Scholes risk neutral assumption means that the discounted stock price process is a fair game, that is, a martingale. For any time $0 < s < t$, the expected present value of the stock t years from now given knowledge of the stock price up until time s should be equal to the present value of the stock price s years from now. In other words,

$$E\left(e^{-rt}G_t|G_x, 0 \le x \le s\right) = e^{-rs}G_s$$

The equation above holds for Brownian motion if $r = \mu + \sigma^2/2$, or $\mu = r - \sigma^2/2$. The probability calculations for the Black-Scholes formula are obtained with this choice of μ. In the language of Black-Scholes, this gives the risk-neutral probability for computing the options price formula. The Black-Scholes formula for the price of an option is then the present value of the expected payoff of the option under the risk-neutral probability.

Now we consider a financial option. An option is a contract that gives the buyer the right to buy shares of stock sometime in the future at a fixed price. If we assumed that XYZ stock currently sells for \$80 a share. Assume that an XYZ option is selling for \$$s$. Under the terms of the option, in 90 days you may buy a share of XYZ stock for \$100.

If you decide to purchase the option, consider the payoff. Assume that in 90 days the price of XYZ is greater than \$100. Then, you can exercise the option, buy the stock for \$100, and turn around and sell XYZ for its current price. Your payoff would be $G_{90/365} - 100$, where $G_{90/365}$ is the price of XYZ in 90 days.

On the other hand, if XYZ sells for less than \$100 in 90 days, you would not exercise the option, and your payoff is nil. In either case, the payoff in 90 days is $\max\left\{G_{90/365} - 100, 0\right\}$. Your final profit would be the payoff minus the initial \$$x$ cost of the option.

We now use Black-Scholes formula to find the future expected payoff of the option, and then, the price for XYZ option s.

Let G_0 denote the current stock price. Let t be the expiration date, which is the time until the option is exercised. Let K be the strike price, which is how much you can buy the stock for if you exercise the option. For $XYZ, G_0 = 80$, $t = 90/365$ (measuring time in years), and $K = 100$.

$$\begin{aligned}
E\left(\max\left\{G_t - K, 0\right\}\right) &= E\left(\max\left\{G_0 e^{\mu t + \sigma B_t} - K, 0\right\}\right) \\
&= \int_{-\infty}^{\infty} \max\left\{G_0 e^{\mu t + \sigma x} - K, 0\right\} f(x) dx \\
&= \int_{\beta}^{\infty} \left(G_0 e^{\mu t + \sigma x} - K\right) f(x) dx \\
&= G_0 e^{\mu t} \int_{\beta}^{\infty} e^{\sigma x} f(x) dx - KP\left(Z > \frac{\beta}{\sqrt{t}}\right)
\end{aligned}$$

where $\beta = \left(\ln\left(K/G_0\right) - \mu t\right)/\sigma,$ and Z is a standard normal random variable.

By completing the square, the integral in the last expression is

$$\begin{aligned}
\int_{\beta}^{\infty} e^{\sigma x} f(x) dx &= \int_{\beta}^{\infty} e^{\sigma x} \frac{1}{\sqrt{2\pi t}} e^{-x^2/2t} dx \\
&= e^{\sigma^2 t/2} \int_{\beta}^{\infty} \frac{1}{\sqrt{2\pi t}} e^{-(x-\sigma t)^2/2t} dx
\end{aligned}$$

$$= e^{\sigma^2 t/2} \int_{(\beta-\sigma t)/\sqrt{t}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$
$$= e^{\sigma^2 t/2} P\left(Z > \frac{\beta-\sigma t}{\sqrt{t}}\right)$$

This gives

$$E\left(\max\left\{G_t - K, 0\right\}\right)$$
$$= G_0 e^{t\left(\mu+\sigma^2/2\right)} P\left(Z > \frac{\beta-\sigma t}{\sqrt{t}}\right) - KP\left(Z > \frac{\beta}{\sqrt{t}}\right)$$

The present value of the expected payoff is obtained by multiplying by the discount factor

$$e^{-rt} E\left(\max\left\{G_t - K, 0\right\}\right) = G_0 P\left(Z > \frac{\alpha - \sigma t}{\sqrt{t}}\right) - e^{-rt} KP\left(Z > \frac{\alpha}{\sqrt{t}}\right)$$

where

$$\alpha = \frac{\ln\left(K/G_0\right) - \left(r - \sigma^2/2\right) t}{\sigma}$$

For the XYZ stock example, $G_0 = 80, K = 100, \sigma^2 = 0.25$, and $t = 90/365$. Furthermore, assume $r = 0.02$ is the risk-free interest rate. Then,

$$\alpha = \frac{\ln(100/80) - (0.02 - 0.25/2)(90/365)}{0.5} = 0.498068$$

and the Black–Scholes option price is

$$80P\left(Z > \frac{\alpha - 0.50(90/365)}{0.5}\right) - e^{-0.02(90/365)}(100)P\left(Z > \frac{\alpha}{0.5}\right) = \$2.426$$

9.6.2 Optional Stopping Theorem

A martingale $\left(Y_t\right)_{t\geq 0}$ has constant expectation. For all $t \geq 0, E\left(Y_t\right) = E\left(Y_0\right)$. The property holds for all fixed, deterministic times, but not necessarily for random times. If T is a random variable, which takes values in the index set of a martingale, it is not necessarily true that $E\left(Y_T\right) = E\left(Y_0\right)$. For instance, let T be the first time that a standard Brownian motion hits level a. Then, $B_T = a = E\left(B_T\right)$. However, $E\left(B_t\right) = 0$, for all $t \geq 0$.

△ Theorem

Optional Stopping Theorem

Let $\left(Y_t\right)_{t\geq 0}$ be a martingale with respect to a stochastic process $\left(X_t\right)_{t\geq 0}$. Assume that T is a stopping time for $\left(X_t\right)_{t\geq 0}$. Then, $E\left(Y_T\right) = E\left(Y_0\right)$ if one of the following is satisfied.

1. T is bounded. That is, $T \leq c$, for some constant c.

2. $P(T < \infty) = 1$ and $E(|Y_t|) \leq c$, for some constant c, whenever $T > t$

· **Example**

Let $a, b > 0$. For a standard Brownian motion, find the probability that the process hits level a before hitting level $-b$.

Solution: Let p be the desired probability. Consider the time T that Brownian motion first hits either a or $-b$. That is, $T = \min\{t : B_t = a \text{ or } B_t = -b\}$.

The random variable T is a stopping time. Furthermore, it satisfies the conditions of the optional stopping theorem. The first hitting time T_a is finite with probability 1. By the strong Markov property, from a, the first time to hit $-b$ is also finite with probability 1. Thus, the first part of condition 2 is satisfied.

Observe that $B_T = a$, with probability p, and $B_T = -b$, with probability $1 - p$. By the optional stopping theorem

$$0 = E(B_0) = E(B_T) = pa + (1 - p)(-b)$$

Solving for p gives $p = b/(a + b)$

· **Example**

Find the expected time that standard Brownian motion first hits the boundary of the region defined by the lines $y = a$ and $y = -b$

Solution: Based on the previous example, we use quadratic martingale here

$$E\left(B_T^2 - T\right) = E\left(B_0^2 - 0\right) = 0$$

$$E(T) = E\left(B_T^2\right) = a^2\left(\frac{b}{a + b}\right) + b^2\left(\frac{a}{a + b}\right) = ab$$

· **Example**

Gambler's Ruin: restated

It was shown that discrete-time simple symmetric random walk $S_0, S_1, \ldots$ is a martingale. As with the quadratic martingale, the process $S_n^2 - n$ is a martingale. The results from the last two examples for Brownian motion can be restated for gambler's ruin on $\{-b, -b+1, \ldots, 0, \ldots, a-1, a\}$ starting at the origin. This gives the following:

1. The probability that the gambler gains a before losing b is $b/(a+b)$.

2. The expected duration of the game is ab.

· **Example**

Time to first hit the line $y = a - bt$

For $a, b > 0$, let $T = \min\{t : B_t = a - bt\}$ be the first time a standard Brownian motion hits the line $y = a - bt$. The random variable T is a stopping time, which satisfies the optional stopping theorem. This gives

$$0 = E(B_0) = E(B_T) = E(a - bT) = a - bE(T)$$

Hence, $E(T) = a/b$

· **Example**

Time to reach a for Brownian motion with drift

Assume that $(X_t)_{t \geq 0}$ is a Brownian motion process with drift parameter μ and variance parameter σ^2, where $\mu > 0$. For $a > 0$, find the expected time that the process first hits level a.

Solution: Let $T = \min\{t : X_t = a\}$ be the first time that the process hits level a. Write $X_t = \mu t + \sigma B_t$. Then, $X_t = a$ if and only if

$$B_t = \frac{a - \mu t}{\sigma} = \frac{a}{\sigma} - \left(\frac{\mu}{\sigma}\right) t$$

Applying the result of the previous Example $E(T) = (a/\sigma)/(\mu/\sigma) = a/\mu$.

· **Example**

Variance of First Hitting time

Assume that $(X_t)_{t\geq 0}$ is a Brownian motion process with drift μ and variance σ^2. Let $T = \min\{t : X_t = a\}$ be the first hitting time to reach level a. In the last example, the expectation of T was derived. Here, the variance of T is obtained using the quadratic martingale $Y_t = B_t^2 - t$

Solution: since T is a stopping time with respect to B_t,

$$0 = E(Y_0) = E(Y_T) = E\left(B_T^2 - T\right) = E\left(B_T^2\right) - E(T)$$

Thus, $E\left(B_T^2\right) = E(T) = a/\mu$. Write $X_t = \mu t + \sigma B_t$. Then, $X_T = \mu T + \sigma B_T$, giving

$$B_T = \frac{X_T - \mu T}{\sigma} = \frac{a - \mu T}{\sigma}$$

$$\begin{aligned}
\operatorname{Var}(T) &= E\left((T - E(T))^2\right) = E\left(\left(T - \frac{a}{\mu}\right)^2\right) \\
&= \frac{1}{\mu^2} E\left((\mu T - a)^2\right) = \frac{\sigma^2}{\mu^2} E\left(\left(\frac{a - \mu T}{\sigma}\right)^2\right) \\
&= \frac{\sigma^2}{\mu^2} E\left(B_T^2\right) = \frac{\sigma^2}{\mu^2}\left(\frac{a}{\mu}\right) = \frac{a\sigma^2}{\mu^3}
\end{aligned}$$

A simulation of the mean and variance of first hitting time has been coded in Python. Check the codes

9.7 Two Solved Exercises

· **Example**

Consider standard Brownian motion. Let $0 < s < t$.

(a) Find the joint density of (B_s, B_t).

(b) Show that the conditional distribution of B_s given $B_t = y$ is normal, with mean and variance

$$E\left(B_s|B_t=y\right)=\frac{sy}{t}\quad \text{and}\ \operatorname{Var}\left(B_s|B_t=y\right)=\frac{s(t-s)}{t}$$

Solution:

(a) For the joint density of B_s and B_t, since

$$\{B_s=x,B_t=y\}=\{B_s=x,B_t-B_s=y-x\}$$

it gives that

$$f_{B_s,B_t}(x,y)=f_{B_s}(x)f_{B_{t-s}}(y-x)=\frac{1}{\sqrt{2\pi s}}e^{-x^2/2s}\frac{1}{\sqrt{2\pi(t-s)}}e^{-(y-x)^2/2(t-s)}$$

(b)

$$f_{B_s|B_t=y}(x)=cf_{B_s,B_t=y}(x,y)$$

where c is the rescale constant to make sure that $\int_{-\infty}^{\infty}f_{B_s|B_t=y}(x)dx=1$. Complete the square we have:

$$f_{B_s|B_t=y}(x)=\frac{xe^{-t(x-sy/t)^2/xs(t-s)}}{2\pi\sqrt{s(t-s)/t}}e^{-\frac{(s-s^2/t)y^2}{2s(t-s)}}\times c$$

Thus, $c=e^{\frac{(s-s^2/t)y^2}{2s(t-s)}}$. Therefore, we have

$$E\left(B_s|B_t=y\right)=\frac{sy}{t}\quad \text{and}\ \operatorname{Var}\left(B_s|B_t=y\right)=\frac{s(t-s)}{t}$$

· **Example**

A standard Brownian motion crosses the t -axis at times $t=2$ and $t=5$. Find the probability that the process exceeds level $x=1$ at time $t=4$

Solution:

$$P(B_4>1|B_2=0,B_5=0)=P(B_4-B_2>1|B_5=0)=P(B_2>1|B_3=0)$$

From the previous example, we have

$$f_{B_s,B_t}(x,y)=\frac{1}{\sqrt{4\pi}}e^{-x^2/4}\frac{1}{\sqrt{2\pi}}e^{-(y-x)^2/2}$$

and when $y = 0$ we have

$$f_{B_2|B_3}(x|0) = \frac{\sqrt{3}}{\sqrt{4\pi}} e^{-\frac{3x^2}{4}}$$

finally

$$P = \int_1^{\infty} \frac{\sqrt{3}}{\sqrt{4\pi}} e^{-\frac{3x^2}{4}} dx \approx 0.11$$

10 A Gentle Intro to Stochastic Calculus

To start off, for $0 \leq a < b$, consider the integral

$$\int_a^b B_s ds$$

Let $I_t = \int_0^t B_s ds$, for $t \geq 0$. It can be shown that $(I_t)_{t\geq 0}$ is a Gaussian process with continuous sample paths. The mean function is

$$E(I_t) = E\left(\int_0^t B_s ds\right) = \int_0^t E(B_s)\, ds = 0$$

where the interchange of expectation and integral can be justified.

For $s \leq t$, the covariance function is

$$\begin{aligned}
\operatorname{Cov}(I_s, I_t) = E(I_s I_t) = E\left(\int_0^s B_x dx \int_0^t B_y dy\right) \\
&= \int_0^s \int_0^t E(B_x B_y)\, dy dx = \int_0^s \int_0^t \min\{x, y\} dy dx \\
&= \int_0^s \int_0^x y dy dx + \int_0^s \int_x^t x dy dx \\
&= \frac{s^3}{6} + \left(\frac{ts^2}{2} - \frac{s^3}{3}\right) = \frac{3ts^2 - s^3}{6}
\end{aligned}$$

Letting $s = t$ gives $\operatorname{Var}(I_t) = t^3/3$ Thus, the stochastic integral $\int_0^t B_s ds$ is a normally distributed random variable with mean 0 and variance $t^3/3$. The integral $\int_0^t B_s ds$ is called integrated Brownian motion.

In probability, if X is a continuous random variable with density function f and cumulative distribution function F, and g is a function, the expectation $E(g(X))$ can be written as

$$E(g(X)) = \int_{-\infty}^{\infty} g(x) f(x) dx = \int_{-\infty}^{\infty} g(x) F'(x) dx = \int_{-\infty}^{\infty} g(x) dF(x)$$

that is, as Riemann-Stieltlies integral of g with respect to the cumulative distribution function F.Based on the Riemann-Stielties integral, we can define the integral of a function g with respect to Brownian motion

$$I_t = \int_0^t g(s) dB_s$$

Technical conditions require that g be a bounded and continuous function, and satisfy $\int_0^\infty g^2(s)ds < \infty$. By analogy with the Riemann-Stieltjes integral, for the partition

$$0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = t$$

let

$$I_t^{(n)} = \sum_{k=1}^{n} g\left(t_k^*\right)\left(B_{t_k} - B_{t_{k-1}}\right)$$

where $t_k^* \in \left[t_{k-1}, t_k\right]$. since $B_{t_k} - B_{t_{k-1}}$ is normally distributed with mean 0 and variance $t_k - t_{k-1}$, the approximating sum $I_t^{(n)}$ is normally distributed for all n. It can be shown that in the limit, as $n \to \infty$

$$\begin{aligned} E\left(I_t\right) = E\left(\lim_{n\to\infty} I_t^{(n)}\right) &= \lim_{n\to\infty} E\left(\sum_{k=1}^{n} g\left(t_k^*\right)\left(B_{t_k} - B_{t_{k-1}}\right)\right) \\ &= \lim_{n\to\infty} \sum_{k=1}^{n} g\left(t_k^*\right) E\left(B_{t_k} - B_{t_{k-1}}\right) = 0 \end{aligned}$$

By independent increments,

$$\operatorname{Var}\left(I_t^{(n)}\right) = \sum_{k=1}^{n} g^2\left(t_k^*\right) \operatorname{Var}\left(B_{t_k} - B_{t_{k-1}}\right) = \sum_{k=1}^{n} g^2\left(t_k^*\right)\left(t_k - t_{k-1}\right)$$

The last expression is a Riemann sum whose limit, as n tends to infinity, is $\int_0^t g^2(s)ds$. In summary,

$$\int_0^t g(s)dB_s \sim \text{ Normal } \left(0, \int_0^t g^2(s)ds\right)$$

In fact, it can be shown that $\left(I_t\right)_{t\geq 0}$ is a Gaussian process with continuous sample paths, independent increments, mean function 0, and covariance function

$$\operatorname{Cov}\left(I_s, I_t\right) = E\left(\int_0^s g(x)dB_x \int_0^t g(y)dB_y\right) = \int_0^{\min(s,t)} g^2(x)dx$$

The stochastic integral

$$\int_a^b g(s)dB_s$$

has many familiar properties. Linearity holds. For functions g and h, for which the integral is defined, and constants α, β

$$\int_a^b [\alpha g(s) + \beta h(s)]dB_s = \alpha \int_a^b g(s)dB_s + \beta \int_a^b h(s)dB_s$$

For $a < c < b$,

$$\int_a^b g(s)dB_s = \int_a^c g(s)dB_s + \int_c^b g(s)dB_s$$

The integral also satisfies an integration-by-parts formula. If g is differentiable,

$$\int_0^t g(s)dB_s = g(t)B_t - \int_0^t B_s g'(s)ds$$

By letting $g(t) = 1$, we capture the identity

$$\int_0^t dB_s = B_t$$

10.1 White Noise

If Brownian motion paths were differentiable, we have

$$B_t = \int_0^t dB_s = \int_0^t \frac{dB_s}{ds}ds$$

The "process" $W_t = dB_t/dt$ is called white noise. The reason for the quotation marks is because W_t is not a stochastic process in the usual sense, as Brownian motion derivatives do not exist. Nevertheless, Brownian motion is sometimes described as integrated white noise.

Consider the following formal treatment of the distribution of white noise. Letting Δ_t represent a small incremental change in t

$$W_t \approx \frac{B_{t+\Delta_t} - B_t}{\Delta_t}$$

The random variable W_t is approximately normally distributed with mean 0 and variance $1/\Delta_t$. We can think of W_t as the result of letting $\Delta_t \to 0$. White noise can be thought of as an idealized continuous-time Gaussian process, where W_t is normally distributed with mean 0 and variance $1/dt$. Furthermore, for $s \neq t$

$$E\left(W_s W_t\right) = E\left(\frac{dB_s}{ds}\frac{dB_t}{dt}\right) = \frac{1}{dsdt}E\left(\left(B_{s+ds} - B_s\right)\left(B_{t+dt} - B_t\right)\right) = 0$$

Hence, W_s and W_t are independent, for all $s \neq t$. In a physical context, white noise refers to sound that contains all frequencies in equal amounts, the analog of white light.

10.2 Ito Integral

We are now ready to consider a more general stochastic integral of the form

$$I_t = \int_0^t X_s dB_s$$

where $(X_t)_{t \geq 0}$ is a stochastic process, and $(B_t)_{t \geq 0}$ is standard Brownian motion.

This brings us to the Ito integral, named after Kiyoshi Ito, a brilliant 20 th century Japanese mathematician whose name is most closely associated with stochastic calculus. The Ito integral is based on taking each t_k^* to be the left endpoint of the subinterval $[t_{k-1}, t_k]$. That is,

$$\int_0^t X_s dB_s = \lim_{n \to \infty} \sum_{k=1}^n X_{t_{k-1}} \left(B_{t_k} - B_{t_{k-1}}\right)$$

The Ito integral requires

1. $\int_0^t E\left(X_s^2\right) ds < \infty$

2. X_t does not depend on the values $\{B_s : s > t\}$ and only on $\{B_s : s \leq t\}$. We say that X_t is adapted to Brownian motion.

3. The limit shown above is defined in the mean-square sense. A sequence of random variables $X_0, X_1, \ldots$ is said to converge to X in mean-square if $\lim_{n \to \infty} E\left((X_n - X)^2\right) = 0$

△ Theorem

Properties of Ito Integral

1. For processes $(X_t)_{t \geq 0}$ and $(Y_t)_{t 20}$, and constants α, β

$$\int_0^t \left(\alpha X_s + \beta Y_s\right) dB_s = \alpha \int_0^t X_s dB_s + \beta \int_0^t Y_s dB_s$$

2. For $0 < r < t$

$$\int_0^t X_s dB_s = \int_0^r X_s dB_s + \int_r^t X_s dB_s$$

3.

$$E\left(I_t\right) = 0$$

4.
$$\text{Var}\left(I_t\right) = E\left(\left(\int_0^t X_s dB_s\right)^2\right) = \int_0^t E\left(X_s^2\right) ds$$

5. $(I_t)_{t\geq 0}$ is a martingale with respect to Brownian motion.

The Ito integral does not satisfy the usual integration-by-parts formula

10.2.1 Ito's Lemma

· **Lemma**

Let g be a real-valued function that is twice continuously differentiable. Then,

$$g\left(B_t\right) - g\left(B_0\right) = \int_0^t g'\left(B_s\right) dB_s + \frac{1}{2}\int_0^t g''\left(B_s\right) ds$$

This is often written in shorthand differential form

$$dg\left(B_t\right) = g'\left(B_t\right) dB_t + \frac{1}{2} g''\left(B_t\right) dt$$

· **Example**

Evaluate

$$\int_0^t B_s^2 dB_s \quad \text{and} \quad \int_0^t \left(B_s^2 - s\right) dB_s$$

Solution:

(i) Use Ito's Lemma with $g(x) = x^3$. This gives

$$B_t^3 = \int_0^t 3B_s^2 dB_s + \frac{1}{2}\int_0^t 6B_s ds$$

Rearranging gives

$$\int_0^t B_s^2 dB_s = \frac{1}{3}B_t^3 - \int_0^t B_s ds$$

(ii) By linearity of the Ito integral,

$$\begin{aligned}\int_0^t \left(B_s^2 - s\right) dB_s &= \int_0^t B_s^2 dB_s - \int_0^t s dB_s \\ &= \frac{1}{3}B_t^3 - \int_0^t B_s ds - \left(tB_t - \int_0^t B_s ds\right) \\ &= \frac{1}{3}B_t^3 - tB_t\end{aligned}$$

The second equality is by integration by parts, which is valid for stochastic integrals with deterministic integrands. since Ito integrals are martingales, the process $\left(\frac{1}{3}B_t^3 - tB_t\right)_{t\geq 0}$ is a martingale.

An extended version of Ito's Lemma allows g to be a function of both t and B_t. The extended result can be motivated by considering a second-order Taylor series expansion of g.

· **Lemma**

Extended version of Ito lemma

Let $g(t,x)$ be a real-valued function whose second-order partial derivatives are continuous. Then,

$$\begin{aligned}g\left(t, B_t\right) - g\left(0, B_0\right) = &\int_0^t \left(\frac{\partial}{\partial t} g\left(s, B_s\right) + \frac{1}{2}\frac{\partial^2}{\partial x^2} g\left(s, B_s\right)\right) ds \\ &+ \int_0^t \frac{\partial}{\partial x} g\left(s, B_s\right) dB_s\end{aligned}$$

In shorthand differential form,

$$dg = \left(\frac{\partial g}{\partial t} + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\right) dt + \frac{\partial g}{\partial x} dB_t$$

· **Example**

Use Ito's Lemma to evaluate $d\left(B_t^3\right)$ and $E\left(B_t^3\right)$

Solution: Let $g(t,x) = x^3$. By Ito's Lemma,

$$d\left(B_t^3\right) = 3B_t dt + 3B_t^2 dB_t$$

$$B_t^3 = 3\int_0^t B_s ds + 3\int_0^t B_s^2 dB_s$$

Taking expectations gives

$$E\left(B_t^3\right) = 3\int_0^t E\left(B_s\right) ds + 3E\left(\int_0^t B_s^2 dB_s\right) = 3(0) + 0 = 0$$

10.3 Stochastic Differential Equations

To motivate the discussion, consider an exponential growth process, be it the spread of a disease, the population of a city, or the number of cells in an organism. Let X_t denotes the size of the population at time t. The deterministic exponential growth model is described by an ordinary differential equation

$$\frac{dX_t}{dt} = \alpha X_t, \text{ and } X_0 = x_0$$

and the solution is

$$X_t = x_0 e^{\alpha t}, \text{ for } t \geq 0$$

The most common way to incorporate uncertainty into the model is to add a random error term, such as a multiple of white noise W_t, to the growth rate. This gives the stochastic differential equation:

$$\frac{dX_t}{dt} = \left(\alpha + \beta W_t\right) X_t = \alpha X_t + \beta X_t \frac{dB_t}{dt}$$

where α and β are parameters, and $X_0 = x_0$. This is really a shorthand for the integral form

$$X_t - X_0 = \alpha \int_0^t X_s ds + \beta \int_0^t X_s dB_s$$

For the stochastic exponential model, we show that geometric Brownian motion defined by

$$X_t = x_0 e^{\left(\alpha - \frac{\beta^2}{2}\right)t + \beta B_t}, \text{ for } t \geq 0$$

is the solution.

· **Example**

Logistic equation

Unfettered exponential growth is typically unrealistic for biological populations. The logistic model describes the growth of a self-limiting population. The standard deterministic model is described by the ordinary differential equation

$$\frac{dP_t}{dt} = rP_t\left(1 - \frac{P_t}{K}\right)$$

where P_t denotes the population size at time t, r is the growth rate, and K is the carrying capacity, the maximum population size that the environment can sustain.

The solution of the deterministic equation-obtained by separation of variables and partial fractions-is

$$P_t = \frac{KP_0}{P_0 + (K - P_0)\, e^{-rt}}, \text{ for } t \geq 0$$

Observe that $P_t \to K$, as $t \to \infty$; that is, the population size tends to the carrying capacity.

A stochastic logistic equation is described by the SDE

$$dP_t = rP_t\left(1 - \frac{P_t}{K}\right)dt + \sigma P_t dB_t$$

where $\sigma > 0$ is a parameter. Let $(X_t)_{t\geq 0}$ be the geometric Brownian motion process defined by

$$X_t = e^{\left(r - \frac{\sigma^2}{2}\right)t + \sigma B_t}$$

It can be shown that the solution to the logistic SDE is

$$P_t = \frac{P_0 K X_t}{K + P_0 r \int_0^t X_s ds}$$

Ito's Lemma is an important tool for working with stochastic differential equations. The Lemma can be extended further to include a wide class of stochastic processes, which are solutions to SDEs of the form

$$dX_t = a\,(t, X_t)\, dt + b\,(t, X_t)\, dB_t$$

where a and b are functions of t and X_t. The integral form is

$$X_t - X_0 = \int_0^t a\,(s, X_s)\, ds + \int_0^t b\,(s, X_s)\, dB_s$$

Such processes are called diffusions or Ito processes. A diffusion is a Markov process with continuous sample paths. The functions a and b are called, respectively, the drift coefficient and diffusion coefficient.

· **Lemma**

Ito's Lemma for Diffusions

Let $g(t,x)$ be a real-valued function whose second-order partial derivatives are continuous. Let $(X_t)_{t\geq 0}$ be an Ito process Then

$$\begin{aligned} g\left(t, X_t\right) - g\left(0, X_0\right) = & \int_0^t \left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\alpha\left(s, X_s\right) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\beta^2\left(s, X_s\right)\right) ds \\ & + \int_0^t \left(\frac{\partial g}{\partial x}\beta\left(s, X_s\right)\right) dB_s \end{aligned}$$

In shorthand differential form,

$$dg = \left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\alpha\left(t, X_t\right) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\beta^2\left(t, X_t\right)\right) dt + \frac{\partial g}{\partial x}\beta\left(t, X_t\right) dB_t$$

· **Example**

Ornstein–Uhlenbeck process

The Ornstein-Uhlenbeck process, called the Langevin equation in physics, arose as an attempt to model the velocity. In finance, it is known as the Vasicek model and has been used to model interest rates. The process is called mean-reverting as there is a tendency, over time, to reach an equilibrium position.

The SDE for the Ornstein-Uhlenbeck process is

$$dX_t = -r\left(X_t - \mu\right) dt + \sigma B_t$$

where r, μ, and $\sigma > 0$ are parameters. The process is a diffusion with

$$a(t,x) = -r(x-\mu) \text{ and } b(t,x) = \sigma$$

The SDE can be solved using Ito's Lemma by letting $g(t,x) = e^{rt}x$, with partial derivatives

$$\frac{\partial g}{\partial t} = re^{rt}x, \quad \frac{\partial g}{\partial x} = e^{rt}, \quad \text{and} \quad \frac{\partial^2 g}{\partial x^2} = 0$$

By Ito's Lemma,

$$\begin{aligned} d\left(e^{rt}X_t\right) &= \left(re^{rt}X_t - e^{rt}r\left(X_t-\mu\right)\right)dt + e^{rt}\sigma dB_t \\ &= r\mu e^{rt}dt + e^{rt}\sigma dB_t \end{aligned}$$

$$e^{rt}X_t - X_0 = r\mu\int_0^t e^{rs}ds + \sigma\int_0^t e^{rs}dB_s = \mu\left(e^{rt}-1\right) + \sigma\int_0^t e^{rs}dB_s$$

$$X_t = \mu + \left(X_0-\mu\right)e^{-rt} + \sigma\int_0^t e^{-r(t-s)}dB_s$$

If X_0 is constant, then by Equation (9.2), X_t is normally distributed with

$$E\left(X_t\right) = \mu + \left(X_0-\mu\right)e^{-rt}$$

$$\mathrm{Var}\left(X_t\right) = \sigma^2\int_0^t e^{-2r(t-s)}ds = \frac{\sigma^2}{2r}\left(1-e^{-2rt}\right)$$

10.3.1 Numerical Approximation and the Euler–Maruyama Method

The differential form of a stochastic differential equation lends itself to an intuitive method for simulation. Given the SDE

$$dX_t = a\left(t,X_t\right)d_t + b\left(t,X_t\right)dB_t$$

Partition the interval $[0,T]$ into n equally spaced points

$$0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T$$

where $t_i = iT/n$, for $i = 0,1,\ldots,n$. The differential dt_i is approximated by $t_i - t_{i-1} = T/n$. The stochastic differential dB_{t_i} is approximated by $B_{t_i} - B_{t_{i-1}}$, which is normally distributed with mean 0 and variance $t_i - t_{i-1} = T/n$. Thus, dB_{t_i} can be approximated by $\sqrt{T/n}Z$, where Z is a standard normal random variable. Let

$$X_{i+1} = X_i + a\left(t_i,X_i\right)T/n + b\left(t_i,X_i\right)\sqrt{T/n}Z_i, \text{ for } i = 0,1,\ldots,n-1$$

where $Z_0, Z_1, \ldots, Z_{n-1}$ are independent standard normal random variables. The sequence $X_0, X_1, \ldots, X_n$ is defined recursively and gives a discretized approximate sample path for $(X_t)_{0\le t\le T}$.

To simulate the solution of the Ornstein–Uhlenbeck SDE

$$dX_t = -r(X_t - \mu)\, dt + \sigma dB_t, \text{ for } 0 \le t \le T$$

$$X_{i+1} = X_i - r(X_i - \mu)\, T/n + \sigma\sqrt{T/n} Z_i, \text{ for } i = 0, 1, \ldots, n-1$$

An implementation in Python with With $n = 1000$, $X_0 = 2, \mu = -1, r = 0.5$, and $\sigma = 0.1$.

· **Example**

Stochastic Resonance

Stochastic resonance is a remarkable phenomenon whereby a signal, which is too weak to be detected, can be boosted by adding noise to the system. The idea is counter-intuitive, since we typically expect that noise (e.g., random error) makes signal detection more difficult. Yet the theory has found numerous applications over the past 25 years in biology, physics, and engineering, and has been demonstrated experimentally in the operation of ring lasers and in the neurons of crayfish.

The phenomenon was first introduced by Roberto Benzi in 1980 in the context of climate research, where it was proposed as a mechanism to explain how dramatic climactic events such as the almost periodic occurrence of the ice ages might be caused by minute changes in the earth's orbit around the sun. The theory has prompted discussions of whether rapid climate change is a hallmark of human impact (e.g., noise in the system). As explained in Benzi (2010), stochastic resonance can be observed by considering the SDE

$$dX_t = \left(X_t - X_t^3 + A \sin t\right) dt + \sigma dB_t$$

Think of the sinusoidal term, called a periodic forcing, as representing a weak, external signal, with amplitude A. We are interested in studying the effect of the noise parameter σ on detection of the forcing signal.

The process $(X_t)_{t \ge 0}$ has two stable points, at ± 1. For small σ (little noise), paths tend to stay near one of these values, although jumps may occur from one stable point to another.

An optimal value of σ is chosen at $\sigma = 0.8$. The hidden periodic forcing is now apparent. The added noise is sufficient for paths of the process to intersect with the range of the sine wave, which facilitates switching states. The system exhibits stochastic resonance.(Check codes for visualization)

11 Introduction for Less-dumb

△ Question

Difference between deterministic and stochastic variable?

In deterministic world, the variable is not random at a given time. In stochastic world, we have random variable at any given time.

The following two examples give single variables in deterministic and stochastic world separately:

Stochastic: A height of a randomly chosen building in the center of London.

Deterministic: An amplitude of earthquake in Tokyo every hour during first week of January,2010.(the process is stochastic, but the variable is not.)

△ Question

What is Ω sample space, sigma algebra and probability measure?

General Theory	Bernoulli Process [1, success; 0, failure] n experiments	Uniform distribution [0,1]	Note
Ω-sample space: all possible outcomes of n experiment	total number of $\Omega = 2^n$, Each Ω represents a vector of n-length $(a_1, a_2, ...a_n)$ where a_n is either 0 or 1.	[0,1]	
$\sigma-$algebra: A collection of subset of Ω space	$F-$power set number of elements: $F = 2^{\Omega} = 2^{2^n}$	Probability that a point is in interval of $[\alpha, \beta]$,$(\alpha, \beta]$, (α, β),$\{\beta\}$(Borel sigma algebra, Given a real random variable defined on a probability space, its probability distribution is a measure on the Borel algebra)	
P, probability measure	P(1)=p, P(0)=1-p	P$\{[\alpha, \beta]\}$=$\beta - \alpha$	$P\{\Omega\} = 1$ $P : F \longrightarrow [0, 1]$

Table 1: Example and definition of some important concepts

In the table above, we use Bernoulli process and uniform random distribution as two example to explain these concepts.

△ Question

Definition of a stochastic process

Given a probability space $(\Omega, \mathbb{F}, \mathbb{P})$, we first define a random variable:a random variable is a measurable function ξ from Ω to the set of real numbers,$\xi : \Omega \longrightarrow \mathbb{R}$.By the word measurable, we mean that for any borrowed subset $\forall B \in \mathbb{B}(\mathbb{R})$. We have that $\xi^{-1}(B) \in \mathbb{F}$.

Let us consider the following example. An agent flips a coin 2 times. In this model, $\Omega = \{(h, h); (h, t); (t, h); (t, t)\}$, where t means tails and h means heads. $\sigma - algebra$ $\mathbb{F}$ contains all possible combinations of those 4 elements in Ω (by the way, the number of elements in $\mathbb{F}$ is exactly 2^4).

Let, $\xi(h; h) = 1, \xi(h; t) = 2, \xi(t; h) = 3, \xi(t; t) = 4$. (Note that $P(\xi = k) = 1/4$,k=1,2,3,4 and 0 otherwise.) Clearly, if we apply ξ^{-1} to both sides of those equations, we obtain $\xi^{-1}(1) = (h; h)$ and so on. Therefore, $\forall B \in \mathbb{B}(\mathbb{R})$,$\xi^{-1}(B) \in \mathbb{F}$.

Now, we introduce the notion of **random function**. Let T be a set of any nature but we will associate the set with time.And then we will consider functions X, $X : T \times \Omega \longrightarrow \mathbb{R}$. And this function is called random function.If $\forall t \in T$ $X(t, \cdot)$ is a random variable on $(\Omega, \mathbb{F}, \mathbb{P})$. This $X(t, \cdot)$ is also referred as X_t.

There are some types of random functions:

If $T = \mathbb{R}_+$, we call this function as random process, or stochastic process. T can be discrete $\mathbb{N}(\mathbb{Z})$ or continuous.

If $T = {\mathbb{R}_+}^n$, we call this function as random field.

△ Question

An analogue to the trajectory of X_t?

The stock price versus time.

△ **Question**

What is finite-dimensional distribution of stochastic processes?

it is the distribution of a vector of random functions $(X_{t1}, X_{t2}, ...X_{tn})$, where $t1, t2...tn \in \mathbb{R}$.Note that here these variables are dependent.

11.1 Renewal Processes

Renewal process, S_n, is a discrete time process such that $S_0 = 0$, and $S_n = S_{n-1} + \xi_n$, where $\xi_1, \xi_2, ...$ is a sequence of independent identically distributed (i.i.d.) random variables with almost surely positive distribution.

The Counting process related to renewal processes is denoted by N_t and it is defined as the maximum index k such that $S_k <= t$:$N_t = argmax_k(S_k < t)$. A trajectory of N_t is shown below:

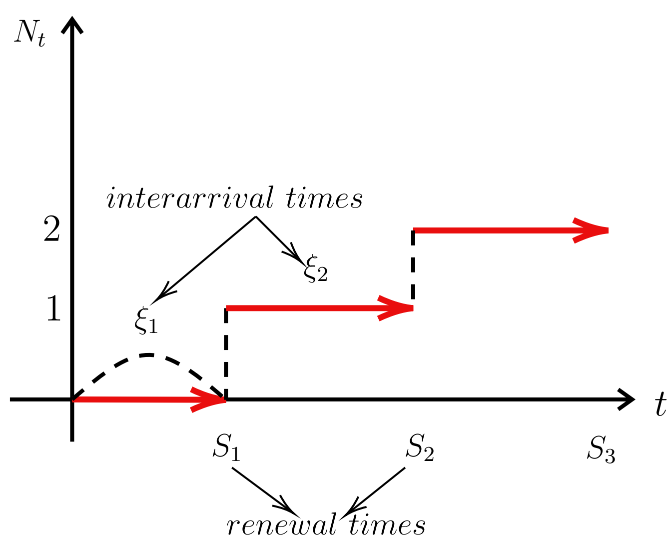

Figure 1: Trajectory for Counting process

From the figure above, we know $\{S_n > t\} = \{N_t < n\}$. In the real life, renewal process could happen for a shop where the time needed for n-th customer to come is a random variable ξ_n

11.2 Convolution

Let $X \sim \exp(\lambda)$ and $Y \sim \exp(\mu)$. Find the probability distribution function of $X+Y$. Use the following convolution formula

$$p_{X+Y}(x) = \int_{-\infty}^{+\infty} p_X(x-y)p_Y(y)dy$$

$$\begin{aligned}
&p_{X+Y}(x) = \int_{-\infty}^{+\infty} p_X(x-y)p_Y(y)dy = \\
&= \begin{cases} \int_0^{+\infty} p_X(x-y)p_Y(y)dy, & x \geq y > 0 \\ 0, & \text{else} \end{cases} \\
&= \int_0^x p_X(x-y)p_Y(y)dy = \int_0^x \lambda e^{-\lambda(x-y)}\mu e^{-\mu y}dy = \\
&= \lambda\mu e^{-\lambda x}\int_0^x e^{(\lambda-\mu)y}dy = \lambda\mu e^{-\lambda x}\frac{1}{\lambda-\mu}e^{y(\lambda-\mu)}\Big|_0^x = \\
&= \frac{\lambda\mu}{\lambda-\mu}\left(e^{-\mu x} - e^{-\lambda x}\right)
\end{aligned} \tag{11.1}$$

Convolution in terms of distribution function:$F_X * F_Y$. Convolution in terms of distribution density function:$P_X * P_Y$. Because for renewal process $S_n = \xi_1 + \xi_2 + ...\xi_n$, we have $F^{n*} = F * F * ... * F$. The first properties of F^{n*} are:

- $F^{n*}(x) <= F^n(x)$ if $F(0) = 0$
- $F^{(n+1)*}(x) <= F^{n*}(x)$

*Proof:*Let us provide a proof for $\{\xi_1 + \cdots + \xi_n \leq x\} \subset \{\xi_1 \leq x; \ldots; \xi_n \leq x\}$: Firstly, since ξ_i are all almost surely positive, i.e. $\mathbb{P}\{\xi_i \geq 0\} = 1$, it is obvious that if $\{\xi_1 + \cdots + \xi_n \leq x\}$ then $\{\xi_1 \leq x_i \ldots \xi_n \leq x\}$ also holds. Secondly, there is a vector $(\xi_1, \ldots, \xi_n)$ such that every component $\xi_i \leq x$, but the sum of all components is larger than x. Now, with $\{\xi_1 + \cdots + \xi_n \leq x\} \subset \{\xi_1 \leq x; \ldots; \xi_n \leq x\}$, we have: $P\{\xi_1 + \cdots + \xi_n \leq x\} \leq \prod_{k=1}^n P\{\xi_k \leq x\}\}$.

Some theorems related to $F^{n*}(t)$:

- $u(t) = \sum_{n=1}^{\infty} F^{n*}(t) < \infty$
- $\mathbb{E}(\mathcal{N}_t) = u(t)$

Here is an empirical understanding of the second theorem above:

Let a = P(N = 1), b = P(N = 2), and c = P(N = 3) and the only values N can take on are (1,2,3). Thus, if we use the standard definition for the expected value, we get the following expression:

$E[N] = (a) + (b + b) + (c + c + c)$

We can represent this sum as a right triangle

a

b b

c c c

Under the standard way, we are adding these up "row-wise", but now we added them "column-wise".

That would be equivalent to $E[N] = (a + b + c) + (b + c) + (c)$

$(a + b + c) = P(N >= 1)$

$(b + c) = P(N >= 2)$

$(c) = P(N >= 3)$

Which is the alternative expression for expectation we desired. Therefore, for a counting process of single random variable, we have:

$$E\left[N_t\right] = \sum_{n=1}^{\infty} P\left(N_t >= n\right)$$

However, from section 1.1 we know that $\{S_n > t\} = \{N_t < n\}$. Thus, $\{S_n <= t\} = \{N_t >= n\}$, and $P\{N_t >= n\} = P\{S_n <= t\} = F^{n*}(t)$ as $S_n = \xi_1 + \xi_2 + ... + \xi_n$. Therefore, we have $E\left[N_t\right] = \sum_{n=1}^{\infty} F^{n*}(t) = u(t)$

11.3 Calculation of an expectation of a counting process

11.3.1 Laplace transform

For a function $f : \mathbb{R}_+ \longrightarrow \mathbb{R}$, we have $\mathcal{L}_f(s) = \int_0^\infty e^{-sx} f(x)dx$. Laplace transform has following properties:

- if f is a propability density of a random variable ξ, then, $\mathcal{L}_f(s) = \mathbb{E}\left[e^{-s\xi}\right]$
- if you have two such function f_1, f_2, then $\mathcal{L}_{f_1 * f_2}(s) = \mathcal{L}_{f_2}(s) \cdot \mathcal{L}_{f_2}(s)$
- Let F as a ditribution function, and $F(0) = 0$ and $P = F'$, we have:

$$\mathcal{L}_F(s) = \frac{\mathcal{L}_P(s)}{s} = -\int_{R_+} F(x)\frac{d\left(e^{-sx}\right)}{s} = -\frac{F(x)e^{-sx}}{s}|_0^\infty + \frac{1}{s}\int_{R_+} P(x)e^{-sx}dx$$

  Because the first term on the right hand side is zero, we have the relation between $\mathcal{L}_F(s)$ and $\mathcal{L}_P(s)$.

Some examples of Laplace transform:

$$t^n(n,\ \text{a positive integer }) \longrightarrow \frac{n!}{s^{n+1}} \quad s > 0$$

$$e^{at} \longrightarrow \frac{1}{s-a} \quad s > a$$

11.3.2 Expectation of a counting process

$$\mathbb{E}(N_t) = u(t) = \sum_{n=1}^{\infty} F^{n*}(t) = F(t) + (\sum_{n=1}^{\infty} F^{n*}) * F(t)$$

Thus

$$u = F + u * F = F + u * P, (F' = P)$$

The equality above holds because

$$\int_R u(x-y)dF(y) = \int_{\mathbb{R}}^n (x-y)P(y)dy$$

Apply Laplace transform, we have

$$\mathcal{L}_y(s) = \mathcal{L}_F(s) + \mathcal{L}_u(s) \cdot \mathcal{L}_P(s)$$

Because $\mathcal{L}_F(s) = \frac{\mathcal{L}_P(s)}{s}$

$$\mathcal{L}_u(s) = \frac{\mathcal{L}_P s}{s(1 - \mathcal{L}_P(s))}$$

With equation above, we propose the following method to calculate $\mathbb{E}(N_t)$:

1) From F, we can calculate P and then $\mathcal{L}_P$

2) Using the equation above to calculate $\mathcal{L}_u$ from $\mathcal{L}_P$

3) Now, calculate u(i.e.) from $\mathcal{L}_u$

11.3.3 An example of calculating expectation of a counting process

If we have a renewal process

$$S_n = S_{n-1} + \xi_n$$

where ξ_n has a probability density function as:

$$P(x) = \frac{e^{-x}}{2} + e^{-2x}, x > 0$$

First, we find $\mathcal{L}_P(s)$:

$$\mathcal{L}_P(s) = \frac{1}{2(s+1)} + \frac{1}{s+2} = \frac{3s+4}{2(s+1)(s+2)}$$

Now, we get

$$\mathcal{L}_u(s) = \frac{\mathcal{L}_P(s)}{s(1 - \mathcal{L}_P(s))} = \frac{3s+4}{s^2(2s+3)} = \frac{4}{3s^2} + \frac{1}{9s} + \frac{2}{18s+27}$$

Finally, we perform reverse Laplace transform:

$$\mathcal{L}_u^{-1}(s) = u(t) = \frac{4}{3}t + \frac{1}{9} - \frac{1}{9} \cdot e^{-3/2t}$$

11.4 Limit theorems for renewal processes

△ Theorem

$$\mu = \mathbb{E}(\xi_1) < \infty$$

Then

$$\frac{N_t}{t} \underset{t\to\infty}{\to} \frac{1}{\mu}$$

The theorem above is an analogue to Large number theorem:$\frac{\xi_1+\ldots+\xi_n}{n} \underset{n\to\infty}{\to} \mu$

Proof: From figure1, we know that $S_{N_t} \leq t \leq S_{N_t+1}$, so we have $\frac{N_t}{S_{N_t+1}} \leqslant \frac{N_t}{t} \leqslant \frac{N_t}{S_{N_t}}$. Therefore, $\lim_{t\to\infty} \frac{N_t}{S_t} = \lim_{n\to\infty} \frac{n}{s_n} = \frac{1}{\mu}$. On the other hand, $\lim \frac{N_t}{S_{N_t+1}} = \lim \frac{N_t}{N_t+1} \cdot \lim \frac{N_t+1}{S_{N_t+1}} = \frac{1}{\mu}$, as $\lim \frac{N_t}{N_t+1} = 1$

△ Theorem

Let

$$\sigma^2 = Var(\xi_n) < 0$$

Then,

$$Z_t = \frac{N_t - t/\mu}{\sigma\sqrt{t}/\mu^{3/2}} \underset{\mathrm{t}\to\infty}{\overset{\text{distribution}}{\to}} N(0,1)$$

This theorem reflects the center limit theorem:

$$\frac{\xi_1 \ldots + \xi_n - n\cdot\mu}{\sigma\sqrt{n}} \overset{distribution}{\to} N(0,1) = \phi(x)$$

Proof: From center limit theorem, $\mathbb{P}\{S_n \leqslant n\mu + \sigma\sqrt{n}x\} \to \phi(x)$, we set $t = n\mu + \sigma\sqrt{n}x$. We know that $\mathbb{P}\{S_n \leqslant t\} = \mathbb{P}\{N_t \geqslant n\}$. Because $n\mu \approx t$ at large n, we have $n \approx t/\mu$, we now have $n = \frac{t}{\mu} - \frac{\sigma\sqrt{n}}{\mu}x = \frac{t}{\mu} - \frac{\sigma\sqrt{t}}{\mu^{3/2}}x$.Therefore, $\mathbb{P}\{N_t \geqslant n\} = \mathbb{P}\{Z_t \geqslant -x\}$. Finally, $\mathbb{P}\{Z_t \leqslant x\} = 1 - \mathbb{P}\{Z_t > -x\} \to 1 - \phi(-x) = \phi(x)$.

11.5 Solved exercises

1. Let η be a random variable with distribution function F_η. Define a stochastic process $X_t = \eta + t$. compute the distribution function of a finite-dimensional distribution $(X_{t_1}, \ldots X_{t_n})$, where $t_1, \ldots, t_n \in \mathbb{R}_+$:

$$\mathbb{F}_{\vec{X}}(\vec{x}) = \mathbb{P}(X_{t_1} \le x_1, \ldots, X_{t_n} \le x_n)$$
$$= \mathbb{P}(\eta + t_1 \le x_1, \ldots, \eta + t_n \le x_n) = \mathbb{P}(\eta \le x_1 - t_1, \ldots, \eta \le x_n - t_n)$$
$$= \mathbb{F}_\eta \{\min(x_1 - t_1, \ldots, x_n - t_n)\}$$

2.Let S_n be a renewal process such that $\xi_n = S_n - S_{n-1}$ takes the values 1 or 2 with equal probabilities $p = 1/2$. Find the mathematical expectation of the counting process N_t at $t = 3$:

$$\mathbb{E}(N_3) = \textstyle\sum_{k=0}^{\infty} k \cdot \mathbb{P}(N_3 = k)$$
$$= 0 + 1 \cdot \mathbb{P}(N_3 = 1) + 2 \cdot \mathbb{P}(N_3 = 2) + 3 \cdot \mathbb{P}(N_3 = 3) + 4 \cdot \mathbb{P}(N_3 = 4)$$
$$= 1 \cdot \mathbb{P}(\xi_1 = 2; \xi_2 = 2) + 2 \cdot (\mathbb{P}(\xi_1 = 1; \xi_2 = 2) + \mathbb{P}(\xi_1 = 2; \xi_2 = 1) +$$
$$+\mathbb{P}(\xi_1 = 1; \xi_2 = 1, \xi_3 = 2)) + 3 \cdot \mathbb{P}(\xi_1 = 1; \xi_2 = 1; \xi_3 = 1)$$
$$= 1 \cdot \tfrac{1}{4} + 2 \cdot \left(\tfrac{1}{4} + \tfrac{1}{4} + \tfrac{1}{8}\right) + 3 \cdot \tfrac{1}{8}$$
$$= \tfrac{1}{4} + 1\tfrac{1}{4} + \tfrac{3}{8} = 15/8$$

3.Let $S_n = S_{n-1} + \xi_n$ be a renewal process and $p_\xi(x) = \lambda e^{-\lambda x}$. Find the mathematical expectation of the corresponding counting process N_t :

4.Let η be a random variable with distribution function F_η. Define a stochastic process $X_t = e^\eta t^2$. What is the distribution function of $(X_{t_1}, \ldots, X_{t_n})$ for positive $t_1, \ldots, t_n$?

$$\mathbb{F}_{\vec{x}}(\vec{x}) = \mathbb{P}(X_{t_1} \le x_1, \ldots, X_{t_n} \le x_n)$$
$$= \mathbb{P}(e^\eta t_1^2 \le x_1, \ldots, e^\eta t_n^2 \le x_n) = \mathbb{P}(e^\eta \le x_1/t_1^2, \ldots, e^\eta \le x_n/t_n^2)$$
$$= \mathbb{F}_\eta \{\min(\ln(x_1/t_1^2), \ldots, \ln(x_n/t_n^2))\}$$

5.Let N_t be a counting process of a renewal process $S_n = S_{n-1} + \xi_n$ such that the i.i.d. random variables $\xi_1, \xi_2, \ldots$ have a probability density function $p_\xi(x) = \begin{cases} \frac{1}{2}e^{-x}(x+1), & x \ge 0 \\ 0, & x < 0 \end{cases}$

Find the mean of N_t :

$$\mathbb{E}N_t = u(t) = -\frac{1}{9} + \frac{2}{3}t + \frac{1}{9}e^{-(3/2)t}$$

$$\begin{aligned}
\text{Step } 1 \cdot p &\to \mathcal{L}_p(s) \\
\mathcal{L}_p(s) &= \int_0^\infty e^{-sx} \cdot \lambda e^{-\lambda x} dx \\
&= \lambda \int_0^\infty e^{-(s+\lambda)x} dx = \frac{\lambda}{s+\lambda}(0-1) = \frac{\lambda}{s+\lambda} \\
\text{Step } 2 \cdot \mathcal{L}_p(s) &\to \mathcal{L}_u(s) \\
\mathcal{L}_u &= \frac{\frac{\lambda}{s+\lambda}}{s\left(1 - \frac{\lambda}{s+\lambda}\right)} = \frac{\lambda}{s^2} \\
\text{Step } 3 . \mathcal{L}_u(s) &\to u \\
\text{since } \mathcal{L}_u &= \lambda \cdot \frac{1!}{s^{1+1}}, \text{ then } u(t) = \lambda t
\end{aligned}$$

6. Let ξ and η be 2 random variables. It is known that the distribution of η is symmetric, that is, $\mathbb{P}\{\eta > x\} = \mathbb{P}\{\eta < -x\}$ for any $x > 0$, and moreover $\mathbb{P}\{\eta = 0\} = 0$. Find the probability of the event that the trajectories of stochastic process $X_t = \xi^2 + t(\eta + t), t \geq 0$ increase:

$$\mathbb{P}\left(\frac{d}{dt}X_t > 0 \forall t \geq 0\right) = \mathbb{P}(2t + \eta > 0 \forall t \geq 0) = \mathbb{P}(\eta > 0) = \frac{1}{2}$$

12 Poisson Process for Less-dumb

△ Definition

Poisson Process is a renewal process such that:

$$S_0 = 0, S_n = S_{n-1} + \xi_n, N_t = \underset{k}{argmax}\{S_k \leq t\}$$

$$\xi_i \sim p(x) = \lambda e^{-\lambda x} \forall \{x > 0\}$$

where λ is called intensity or rate.

△ Theorem

A distribution function of S_n:

$$F_{S_n}(x) = \begin{cases} 1 - e^{-\lambda x} \sum_{k=0}^{n-1} \frac{(\lambda x)^k}{k!}, x > 0 \\ 0, x < 0 \end{cases}$$

The distribution density function:

$$P_{S_n}(x) = \lambda \frac{(\lambda x)^{n-1}}{(n-1)!} e^{-\lambda x} \cdot \forall x > 0$$

The possibility of $N_t = n$ is:

$$\mathbb{P}\{N_t = n\} = e^{-\lambda t} \cdot \frac{(\lambda t)^n}{n!}$$

In other words,

$$P(k \text{ events in interval h }) = e^{-\lambda h} \frac{(\lambda h)^k}{k!}$$

And $N_t \sim Poisson(\lambda t)$

Proof for $P_{S_n}(x)$: (i) $n = 1 : S_1 = \xi_1$, $P_{S_1}(x) = \lambda \cdot e^{-\lambda x}, x > 0$. When $n \to n+1$, $P_{S_{n+1}}(x) = \int_0^x P_{S_n}(x-y) P_{S_{n+1}}(y) dy = \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda x} \int_0^x (x-y)^{n-1} dy = \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda x} \cdot \frac{x^n}{n}$

Proof for $\mathbb{P}\{N_t = n\}$: Starting from an equality of $\mathbb{P}\{N_t = n\} = \mathbb{P}\{S_n \leq t\} - \mathbb{P}\{S_{n+1} \leqslant t\}$. This equality holds because $\{N_t = n\} = \{S_n \leqslant t\} \cap \{S_{n+1} > t\} = A \cap B$. Since $A \cap B = A \setminus B^c$ and $B^2 \subset A$, we have $\mathbb{P}\{A \cap B\} = \mathbb{P}\{A\} - \mathbb{P}\{B\}$. Thus, $\mathbb{P}\{S_n \leq t\} - \mathbb{P}\{S_{n+1} \leqslant t\} = (1 - e^{-\lambda t} \sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!}) - (1 - e^{-\lambda t} \sum_{k=0}^{n} \frac{(\lambda t)^k}{k!}) = e^{-\lambda t} \cdot \frac{(\lambda t)^n}{n!}$

12.1 Memoryless Property

A random variable (r.v.) X possesses the memoryless property if and only if $\mathbb{P}\{X > u+v\} = \mathbb{P}\{X > u\}\mathbb{P}\{X > v\}$ If $P\{X > v\} > 0$, thus:

$$\mathbb{P}\{X > u+v|X > v\} = \mathbb{P}\{X > u\}$$

△ Theorem

Let X be a r.v. with density p(x), if X is memoryless, we have $p(x) = \lambda e^{-\lambda x}$.

A real life example of not using Poisson process: Buses arrive every 20 ± 2 minute, we can not use Poisson process to model the arrival of buses. Because if $v = 19min, u = 10min, \mathbb{P}\{X > 29|X > 19\} = 0$, but $\mathbb{P}\{X > 10\} = 1$. Thus, the theorem 2.2 is violated.

△ Question

A short exercise: Let X be a random variable which with probability 0.5 is equal to -1, and with probability 0.5 has a uniform distribution in a range [0,10]. Thus, its distribution function is equal to

$$F(x) = \begin{cases} 0, & x < -1 \\ 0.5, & x \in [-1; 0] \\ 0.5 + x * 0.05, & x \in [0; 10] \\ 1, & x > 10 \end{cases}$$

Is X a memoryless random variable?

Answer: Let us check for $u = 6, v = 7$. We can see that the $\mathbb{P}\{X > 13|X > 7\} = 0$ whereas the right-hand side equals $\mathbb{P}\{X > 6\} = 0.2$. Thus X is not memoryless.

12.2 Other Definition of Poisson processes

Both renewal process and counting process are called Poisson, from now on we only call counting processes as Poisson. Poisson processes are widely used for modeling the occurrence of sum of odds.

Δ Definition

N_t is a integral-valued process such that: 0) $N_0 = 0$

1) N_t has independent increments. For any time points $t_0, t_1, ..$ the increments of the process $N_{t1} - N_{t0}, N_{t2} - N_{t1}, ..N_{t_n} - N_{t_{n-1}}$ are independent.

2) N_t has stationary increments: $N_t - N_s$ has the same distribution as N_{t-s}, $N_t - N_s \overset{d}{=} N_{t-s}$.

3) $N_t - N_s \sim Pois(\lambda(t-s))$

4)

$$\lim_{h \to 0} \frac{\mathbb{P}\left(N_{t+h} - N_t \geq 2\right)}{\mathbb{P}\left(N_{t+h} - N_t = 1\right)} = 0$$

Δ Definition

Let $f(h) = \bar{\bar{o}}(g(h)) = \lim_{h \to 0} \frac{f(h)}{g(h)} = 0$

$$1)\mathbb{P}\left\{N_{t+h} - N_t = 0\right\} = 1 - \lambda h + \bar{\bar{o}}(h), h \to 0$$

$$2)\mathbb{P}\left\{N_{t+h} - N_t = 1\right\} = \lambda h + \bar{\bar{o}}(h), h \to 0$$

$$3)\mathbb{P}\left\{N_{t+h} - N_t \geq 2\right\} = \bar{\bar{o}}(h), h \to 0$$

Proof for 1): Since $\mathbb{P}\left\{N_{t+h} - N_t = 0\right\} = \mathbb{P}\left\{N_h = 0\right\} = e^{-\lambda h}$, we have

$$\lim_{h \to 0} \frac{1 - \mathbb{P}\left\{N_{t+h} - N_t = 0\right\}}{h} = \lim_{h \to 0} \frac{1 - e^{-\lambda h}}{h} = \lim_{h \to 0} \frac{\lambda e^{-\lambda h}}{1} = \lambda$$

Thus, we arrive at: $\frac{1 - \mathrm{P}(N_t + h - N_t = 0)}{h} - \lambda = \bar{\bar{\bar{o}}}(1)$ $\mathbb{P}\left\{N_{t+h} - N_t = 0\right\} = 1 - \lambda h + \bar{\bar{o}}(1) h_t$ (the sign before $\bar{\delta}(1)h$ does not matter. $\mathbb{P}\left\{N_t + h - N_t = 0\right\} = 1 - \lambda h + \bar{\bar{o}}(h)$, because $\lim_{h \to 0} \frac{o(1)h}{h} = 0$

△ Definition

N_t is a poisson process, if:

i) $N_0 = 0$

ii) N_t has indep. increm.

iii) N_t has stationary incr.

iv)$\lim_{h\to 0} \frac{\mathbb{P}(N_{t+h}-N_t\geq 2)}{\mathbb{P}(N_{t+h}-N_t=1)} = 0$

12.3 Non-homogeneous Poisson processes

From the definitions above, we know that

$$N_t \sim Pois(\lambda t) \rightarrow \mathbb{E}(N_t) = \lambda t$$

△ Definition

Let $\Lambda(t)$ be a differentiable increasing function, $\Lambda(0) = 0$, then $X_t = N_t$ is a non-homogeneous Poisson process if

i) $N_0 = 0$

ii) N_t has indep. increments

iii) $N_t - N_s \sim Poi(\Lambda(t) - \Lambda(s))$

Notice that the process become homogeneous if $\Lambda(t) = \lambda t$

Here are some properties of this non-homogeneous poisson process (NHPP):

(1) Let $\lambda(t) = \Lambda'(t)$ be the intensity function. And the mean value for this non-homogeneous Poisson process is:

$$\mathbb{E}(N_t) = \sum_{0}^{\infty} k \frac{\Lambda(t)^k e^{-\Lambda(t)}}{k!} = \Lambda(t)$$

Notice that the $\Lambda(t)$ above can be linear or polynomial like $\Lambda(t) = \alpha t^{\beta}, \alpha, \beta > 0$.

(2)If $\lambda(t) = const \Rightarrow \Lambda(t) = const \cdot t$.

(3) since the function capital lambda of t is differentiable, then it is definitely continuous. And since we know that lambda of t is more of increasing, then we get that it has an inverse function:

$$\Lambda^{-1}(t) \exists\ \Lambda(t) \in \mathbb{R}_+$$

we have $N_{\Lambda^{-1}(t)}$ is a homogeneous PP.

12.4 Relation between renewal theory and non-homogeneous Poisson processes

We can construct renewal process from a counting process: If we know that $N_t - N_s \sim Poi(\Lambda(t) - \Lambda(s))$, we have renewal process as,

$$S_n = \underset{t}{argmin}\{N_t = n\}$$

Because $\xi_n = S_n - S_{n-1}$, we ask the following question:

△ Question

Whether it is possible to be told from non-homogeneous PP that ξ_n are identically and independent distributed (i.i.d)?

First, the possibility density of ξ_1 is

$$p_{\xi_1}(x) = \lambda(x)e^{-\Lambda(x)}$$

Proof for this is: $\mathbb{P}\{\xi_1 \leq x\} = P\{S_1 \leq x\} = \mathbb{P}\{N_x \geqslant 1\} = 1 - \mathbb{P}\{N_x = 0\}$. Now, from above, we have:

$$P_{\xi_2|\xi_1}(t|s) = \lambda(t+s)e^{-\Lambda(t+s)+\Lambda(s)}$$

Proof:

$$F_{\xi_1,\xi_2}(s,t) = P\{\xi_1 \leq s, \xi_2 \leq t\} = \int_0^s \mathbb{P}\{\xi_1 \leq s, \xi_2 \leq t | \xi_1 = y\} \cdot P_{\xi_1}(y)dy$$
$$= \int_0^s \mathbb{P}\{\xi_2 \leq t | \xi_1 = y\} \cdot P_{\xi_1}(y)dy$$
$$= \int_0^s \mathbb{P}\{N_{t+y} - N_y \geq 1 | \xi_1 = y\} \cdot P_{\xi_1}(y)dy$$

Because $N_{t+y} - N_y \geq 1$ happens after $\xi_1 = y$, we have

$$\mathbb{P}\left\{N_{y+t} - N_y \geq 1 | \xi_1 = y\right\} = \mathbb{P}\left\{N_{y+t} - N_y \geq 1 | N_y = 1\right\}$$

$= \mathbb{P}\left\{N_{y+t} - N_y \geq 1 | N_y - N_0 = 1\right\}$. And since the Poisson process has independent increments, it is equal to $\mathbb{P}\left\{N_y + t - N_y \geq 1\right\}$. Now we get:

$$F_{\xi_1,\xi_2}(s,t) = \int_0^s \mathbb{P}\{N_{t+y} - N_y \geq 1\} \cdot P_{\xi_1}(y)dy$$
$$= \int_0^s (1 - e^{-\Lambda(t+y) - \Lambda(y)})\lambda(y)e^{-\Lambda(y)}dy$$

From the distribution above, we have the density:

$$p_{(\xi_1,\xi_2)}(s,t) = \frac{\partial}{\partial t}\left(\frac{\partial}{\partial s}F_{\xi_1,\xi_2}(s,t)\right)$$
$$= \frac{\partial}{\partial t}((1 - e^{-\Lambda(t+s)+\Lambda(s)})\lambda(s)e^{-\Lambda(s)})$$
$$= \lambda(t+s)e^{-\Lambda(t+s)+\Lambda(s)}\lambda(s)e^{-\Lambda(s)}$$

Since $P_{\xi_2|\xi_1}(t|s) = P_{\xi_1,\xi_2}(s,t)/P_{\xi_1}(s)$, the equality is proved.

Now, if $\xi_1, \xi_2, ...$ are i.i.d, we have

$$P_{\xi_1}(t) = P_{\xi_2|\xi_1}(t|s) \forall t, s > 0$$

. If we now integrate the both side of this equation by $\int_0^T ...dt$, we get:

$$e^{-\Lambda 0} - e^{-\Lambda T} = e^0 - e^{-\Lambda T + s + \Lambda(s)}$$

$$\Rightarrow \Lambda T + s - \Lambda(s) = \Lambda T \forall s, T > 0$$

Because ΛT is non-decreasing, we have $\Lambda(T) = \lambda t$ So the general answer to the question above is: the non-homogeneous Poisson process can be obtained from a regular process if and only if, it is in fact a homogeneous Poisson.

12.5 Elements of Queueing theory: M/G/K systems

For NHPP, the following theorem holds

Δ Theorem

1) $\mathbb{P}\{N_t, h - N_t = 0\} = 1 - \lambda(t)h + \bar{\bar{o}}(h), h \to 0$

2) $\mathbb{P}\{N_{t+h} - N_t = 1\} = \lambda(t)h + \bar{\bar{o}}(h), h \to 0$

3) $\mathbb{P}\{N_{t+h} - N_t \geq 2\} = \bar{\bar{o}}(h), h \to 0$

Where λ is a function of time while $\bar{\bar{o}}$ goes to zero when $h \to 0$

12.6 Solved Exercises

1. Compute the mathematical expectation of a Poisson process N_t with intensity λ :

Answer: λt

Solution: This is the basic feature of the Poisson process. Keep in mind that $N_t \sim$ Pois (λt)

2. Find the probability generating function of a random variable with binomial distribution,

$$\mathbb{P}\{\xi = k\} = C_n^k p^k (1-p)^{n-k}, \quad k = 0, 1, \ldots, n, \quad p \in (0,1)$$

Answer: $\varphi(u) = (up + (1-p))^n$

Solution: $PGF = \varphi_\xi(u) = \mathbb{E}\left[u^\xi\right] = \sum_{k=0}^n u^k C_n^k p^k (1-p)^{n-k} = \sum_{k=0}^n C_n^k (up)^k (1-p)^{n-k} =$ (Newton binomial) $= (up + (1-p))^n$

3. Let N_t be a (homogeneous) Poisson process with intensity λ. Find the limit $\lim_{h \to 0} \mathbb{P}\{N_h = 0\}$:

Answer: 1

Solution: $\lim_{h \to 0} \mathbb{P}\{N_h = 0\} = \lim_{h \to 0} e^{-\lambda h} = 1$

4. Let N_t be a (homogeneous) Poisson process with intensity λ. Find the limit $\lim_{h \to 0} \mathbb{P}\{N_h = 1\}$:

Answer: 0

Solution: $\lim_{h\to 0} \mathbb{P}\{N_h = 3\} = \lim_{h\to 0} e^{-\lambda h} \frac{\lambda h}{1!} = 1 \cdot 0$

5. Let N_t be a (homogeneous) Poisson process with intensity λ. Find the limit $\lim_{h\to 0} \mathbb{P}\{N_h = 3\}$:

Answer: 0

Solution: $\lim_{h\to 0} \mathbb{P}\{N_h = 3\} = \lim_{h\to 0} e^{-\lambda h} \frac{(\lambda h)^3}{3!} = 1 \cdot 0$

6. 2 friends are chating: one has a messaging speed equal to 3 messages per minute, another -2 messages per minute. Assuming that for every person the process of writing the messages is modeled with Poisson process and these processes are independent, find the probability that there will be sent only 2 messages during the first minute:

Answer: $e^{-5}\frac{25}{2}$

Solution: $\mathbb{P}^* = \mathbb{P}\left(N^A = 2, N^B = 0\right) + \mathbb{P}\left(N_1^A = 0, N_1^B = 2\right) + \mathbb{P}\left(N_1^A = 1, N_1^B = 1\right) = e^{-5}\left(\frac{3}{2!}^2 \frac{2^n}{0!} + \frac{3^n}{1!}\frac{2^2}{2!} + \frac{3}{1}\right) = e^{-5}\frac{25}{2}$

7. Purchases in a shop are modelled by the homogeneous Poisson process: 30 purchases are made on average during an hour after the opening of the shop. Find the probability the interval between k and $k+1$ purchases will be less than or equal to 4 minutes, given that the purchase number k was in the time moment s :

Answer: $1 - e^{-2}$

Solution:

$$\begin{aligned}
\mathbb{P}\left(S_{k+1} - S_k \leq 4 | N_s = k\right) &= \mathbb{P}\left(N_{s+4} - N_s \geq 1 | N_s - N_0 = k\right) \\
&= \mathbb{P}\left(N_{s+4} - N_s > 0\right) \\
&= 1 - \mathbb{P}\left(N_{s+4} - N_s = 0\right) \\
&= 1 - \mathbb{P}\left(N_4 = 0\right) = 1 - e^{-2}
\end{aligned}$$

because $N_4 \sim$ Pois $(4 \cdot 30/60)$

8. The amount of claims to an insurance company is modelled by the Poisson process, and the claim sizes are modelled by an exponential distribution. On average there are 100

claims per day, and the mean value of 1 claim is 5000USD. Find the variance of the process X_t, which is equal to the total amount of claims till time t :

Answer: $5t \times 10^9$

Solution: $\xi \sim \exp(\mu)$ $\mathbb{E}\xi = \frac{1}{\mu} = 5\times10^3$ $\operatorname{Var}\xi = \frac{1}{\mu^2} = 25\times10^6$ $\mathbb{E}\xi^2 = \operatorname{Var}\xi + (\mathbb{E}\xi)^2 = 5\times10^7$ $\operatorname{Var}X_t = \lambda t \mathbb{E}\xi^2 = 100t \cdot 5 \times 10^7 = 5t \times 10^9$

9. The amount of claims to an insurance company is modelled by the Poisson process, and the claim sizes are modelled by an exponential distribution. On average there are 100 claims per day, and the mean value of 1 claim is 5000USD. Find the probability that the process X_t, which is equal to the total amount of claims till time t, is equal to 0 at the moment t

Answer: e^{-100t}

Solution: $\mathbb{P}(X_t = 0) = \mathbb{P}(N_t = 0) + \mathbb{P}(N_t > 0) \cdot \mathbb{P}\left(\sum_{k=1}^{N_t} \xi_k = 0\right) = e^{-100t} + 0$

10. The amount of claims to an insurance company is modelled by the Poisson process, and the claim sizes are modelled by an exponential distribution. On averesse there are 100 claims per day, and the mean value of 1 claim is 5000USD.Find the mean value of the process X_t, which is equal to the total amount of claims till time t :

Answer: 500000

Solution: According to the corollary about Compound Poisson processes: $\mathbb{E}(X_t) = \lambda t \cdot \mathbb{E}(\xi) = 100t \cdot 5000 = 500000$

11. Purchases in a shop are modelled with non-homogeneous Poisson process: $30t^{5/4}$ purchases are made on average during t hours after the opening of the shop. Find the probability that the interval between k and $k+1$ purchases will be less or equal than 2 minutes, given that the purchase number k was in the time moment s :

Answer: $1 - e^{-30(s+1/30)^{5/4}+30s^{5/4}}$

Solution:

$\mathbb{P}\left(S_{k+1}-S_k \leq 2|N_s=k\right)$

$=\mathbb{P}\left(N_{s+2}-N_s \geq 1|N_s=k\right)=\mathbb{P}\left(N_{s+2}-N_s \geq 1|N_s-N_0=k\right)$

$=\mathbb{P}\left(N_{s+2}-N_s \geq 1\right)$

$=1-\mathbb{P}\left(N_{s+2}-N_s=0\right)=1-e^{-\Lambda(t+s)+\Lambda(s)}$

$=1-e^{-30(s+1/30)^{5/4}+30s^{5/4}}$

12. Number of downloads of an app in Googgle-Play are modelled by a non-homogeneous Poisson process with intensity $\Lambda(t)=t^{13/5}$, where t is measured in hours after app's commencement time. Find the probability that the time between the 1000^{th} and 1001^{st} downloads is less than or equal to 36 seconds $\big($0.01 hour) given 1000^{th} download time being 14 hours after app's launch.

Solution: $\mathbb{P}\left(S_{1001}-S_{1000} \leq 0.01|S_{1000}=14\right)=\mathbb{P}\left(S_{1001}-S_{1000} \leq 0.01|N_{14}=1000\right)=\mathbb{P}\left(N_{14.01}-N_{14} \geq 1|N_{14}-N_0=1000\right)=\mathbb{P}\left(N_{14.01}-N_{14} \geq 1\right)=1-\mathbb{P}\left(N_{14.01}-N_{14}=0\right)=1-e^{-(\Lambda(14.01)-\Lambda(14))}\frac{(\Lambda(14.01)-\Lambda(14))^0}{0!}=1-e^{-14.01^{2.6}+14^{2.6}}=0.83$

13. Sales of a product are modelled by a non-homogeneous Poisson process with intensity $\Lambda(t), t$ in house. Find the probability that the time between the $10th$ and $25th$ purchases is less than or equal to 1 hour given $10th$ product purchase time being 18 minutes (0.3 hours) after the shop's opening.

Solution:

$\mathbb{P}\left(S_{25}-S_{10} \leq 1|S_{10}=0.3\right)$

$=\mathbb{P}\left(S_{25}-S_{10} \leq 1|N_{0.3}=10\right)=\mathbb{P}\left(N_{1.3}-N_{0.3} \geq (25-10)|N_{0.3}-N_0=10\right)$

$=\mathbb{P}\left(N_{1.3}-N_{0.3} \geq 15\right)=1-\mathbb{P}\left(N_{1.3}-N_{0.3}<15\right)$

$=1-e^{-(\Lambda(1.3)-\Lambda(0.3))}\sum_{k=0}^{14}\frac{(\Lambda(1.3)-\Lambda(0.3))^k}{k!}$

14. Find the probability generating function of the the random variable N_3 (where N_t is a homogeneous Poisson process) using the formula $PGF=\varphi_\alpha(u)=\mathbb{E}\left(u^\alpha\right)$:

Answer: $e^{-3\lambda(1-u)}$ Solution: $\mathbb{E}\left(u^{N_3}\right)=\sum_{k=0}^{\infty}u^k e^{-3\lambda}\frac{(3\lambda)^k}{k!}=e^{-3\lambda}\sum_{k=0}^{\infty}\frac{(3u\lambda)^k}{k!}=e^{-3\lambda}e^{3\lambda u}=e^{(-3\lambda(1-u))}$, because $\sum_{k=0}^{\infty}\frac{\alpha^k}{k!}=e^\alpha$

15. There is a speed limit on the street near the secondary school. To keep a lid on traffic violations, local administration decided to put a speed register. If a car violates the speed limit, the register correctly identifies its ID number with probability 80%. Assume that the number of cars passing the school and violating the speed limit is modelled by the homogeneous Poisson process N_t with intensity equal to 20. Find the probability that during 2 hours after midday there will be 1 cars registered.

Answer: 7%

Solution: Denote the number of registered cars till time t by M_t. The probability of a correct identification of a car is equal to $p = 0.8$ So, we need to calculate $\mathbb{P}\left(M_{2p.m.} - M_{12a.m.} = 16\right)$ More generally,

$$\begin{aligned}\mathbb{P}\left(M_t - M_s = m\right) =& \mathbb{P}\left(M_t - M_s = m \cap N_t - N_s = \mathbf{m}\right)\\ &+ \mathbb{P}\left(M_t - M_s = m \cap N_t - N_s = \mathbf{m+1}\right) +\\ &+ \mathbb{P}\left(M_t - M_s = m \cap N_t - N_s = \mathbf{m} + 2\right) + \cdots\end{aligned}$$

$$\begin{aligned}&= \textstyle\sum_{n=m}^{\infty} \mathbb{P}\left(M_t - M_s = m \cap N_t - N_s = \mathbf{n}\right)\\ &= \textstyle\sum_{n=m}^{\infty} \mathbb{P}\left(M_t - M_s = m | N_t - N_s = \mathbf{n}\right) \cdot \mathbb{P}\left(N_t - N_s = \mathbf{n}\right)\\ &= \textstyle\sum_{n=m}^{\infty} C_n^m p^m (1-p)^{n-m} \cdot e^{-\lambda(t-s)} \frac{(\lambda(t-s))^n}{n!}\\ &= \textstyle\sum_{n=m}^{\infty} \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} \cdot e^{-\lambda(t-s)} \frac{(\lambda(t-s))^n}{n!}\end{aligned}$$

$$\begin{aligned}&= \textstyle\sum_{n=m}^{\infty} \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m} \cdot e^{-\lambda(t-s)} \frac{(\lambda(t-s))^n}{n!}\\ &= \frac{p^m e^{-\lambda(t-s)}}{m!} \textstyle\sum_{n=m}^{\infty} \frac{n!}{(n-m)!} \frac{(1-p)^n}{(1-p)^m} \cdot \frac{(\lambda(t-s))^n}{n!}\\ &= \left(\frac{p}{1-p}\right)^m \frac{e^{-\lambda(t-s)}}{m!} \textstyle\sum_{n=m}^{\infty} \frac{n!}{(n-m)!} (1-p)^n \cdot \frac{(\lambda(t-s))^n}{n!}\\ &= \left(\frac{p}{1-p}\right)^m \frac{e^{-\lambda(t-s)}}{m!} (\lambda(t-s)(1-p))^{\mathbf{k+m}} e^{\lambda(t-s)(1-p)}\\ &= \frac{(\lambda p(t-s))^m}{m!} e^{-\lambda p(t-s)}\end{aligned}$$

Therefore,

$$\mathbb{P}\left(M_{2p,m.} - M_{12a,m.} = 16\right) = \frac{(20 \cdot 0.8(2-0))^{16}}{16!} e^{-20.8(2-0)} = 0.07\%$$