

## CHAPTER 1

---

### PRELIMINARIES

---

#### *Probability Theory*

1.1 Using the properties of the probability measure in **Definition 1.1.1**:

A *probability*  $\mathbb{P}$  is a rule that assigns a number  $0 \leq \mathbb{P}(A) \leq 1$  to each event  $A$ , such that  $\mathbb{P}(\Omega) = 1$ , and such that for any sequence  $A_1, A_2, \dots$  of disjoint events

$$\mathbb{P}\left(\bigcup_i A_i\right) = \sum_i \mathbb{P}(A_i),$$

prove the following results.

- (a)  $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ .
- (b)  $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$ .

1.2 Prove the product rule (1.4):

For any sequence of events  $A_1, A_2, \dots, A_n$ .

$$\mathbb{P}(A_1 \cdots A_n) = \mathbb{P}(A_1) \mathbb{P}(A_2 | A_1) \mathbb{P}(A_3 | A_1 A_2) \cdots \mathbb{P}(A_n | A_1 \cdots A_{n-1}),$$

using the abbreviation  $A_1 A_2 \cdots A_k \equiv A_1 \cap A_2 \cap \cdots \cap A_k$ .

for the case of three events.

1.3 We draw three balls consecutively from a bowl containing exactly five white and five black balls, without putting them back. What is the probability that all drawn balls will be black?

1.4 Consider the random experiment where we toss a biased coin until heads comes up. Suppose the probability of heads on any one toss is  $p$ . Let  $X$  be the number of tosses required. Show that  $X \sim G(p)$ .

1.5 In a room with many people, we ask each person his/her birthday, for example, May 5. Let  $N$  be the number of people queried until we get a "duplicate" birthday.

- Calculate  $\mathbb{P}(N > n)$ ,  $n = 0, 1, 2, \dots$
- For which  $n$  do we have  $\mathbb{P}(N \leq n) \geq 1/2$ ?
- Use a computer to calculate  $\mathbb{E}[N]$ .

1.6 Let  $X$  and  $Y$  be independent standard normal random variables, and let  $U$  and  $V$  be random variables that are derived from  $X$  and  $Y$  via the linear transformation

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} \sin \alpha & -\cos \alpha \\ \cos \alpha & \sin \alpha \end{pmatrix} \begin{pmatrix} X \\ Y \end{pmatrix}.$$

- Derive the joint pdf of  $U$  and  $V$ .
- Show that  $U$  and  $V$  are independent and standard normally distributed.

1.7 Let  $X \sim \text{Exp}(\lambda)$ . Show that the *memoryless property* holds:

$$\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s) \quad \text{for all } s, t \geq 0.$$

1.8 Let  $X_1, X_2, X_3$  be independent Bernoulli random variables with success probabilities  $1/2, 1/3$ , and  $1/4$ , respectively. Give their conditional joint pdf, given that  $X_1 + X_2 + X_3 = 2$ .

1.9 Verify the expectations and variances in Table 1.1 below.

Table 1.1 Expectations and variances for some well-known distributions.

Dist.	$\mathbb{E}[X]$	$\text{Var}(X)$	Dist.	$\mathbb{E}[X]$	$\text{Var}(X)$
$\text{Bin}(n, p)$	$np$	$np(1-p)$	$\text{Gamma}(\alpha, \lambda)$	$\frac{\alpha}{\lambda}$	$\frac{\alpha}{\lambda^2}$
$G(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$N(\mu, \sigma^2)$	$\mu$	$\sigma^2$
$\text{Poi}(\lambda)$	$\lambda$	$\lambda$	$\text{Beta}(\alpha, \beta)$	$\frac{\alpha}{\alpha+\beta}$	$\frac{\alpha\beta}{(\alpha+\beta)^2(1+\alpha+\beta)}$
$U(\alpha, \beta)$	$\frac{\alpha+\beta}{2}$	$\frac{(\beta-\alpha)^2}{12}$	$\text{Weib}(\alpha, \lambda)$	$\frac{\Gamma(1/\alpha)}{\alpha\lambda}$	$\frac{2\Gamma(2/\alpha)}{\alpha} - \left(\frac{\Gamma(1/\alpha)}{\alpha\lambda}\right)^2$
$\text{Exp}(\lambda)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$			

1.10 Let  $X$  and  $Y$  have joint density  $f$  given by

$$f(x, y) = cxy, \quad 0 \leq y \leq x, \quad 0 \leq x \leq 1.$$

(a) Determine the normalization constant  $c$ .

(b) Determine  $\mathbb{P}(X + 2Y \leq 1)$ .

1.11 Let  $X \sim \text{Exp}(\lambda)$  and  $Y \sim \text{Exp}(\mu)$  be independent. Show that

(a)  $\min(X, Y) \sim \text{Exp}(\lambda + \mu)$ ,

(b)  $\mathbb{P}(X < Y \mid \min(X, Y)) = \frac{\lambda}{\lambda + \mu}$ .

1.12 Verify the properties of variance and covariance in Table 1.2 below.

**Table 1.2** Properties of variance and covariance.

1	$\text{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$
2	$\text{Var}(aX + b) = a^2 \text{Var}(X)$
3	$\text{Cov}(X, Y) = \mathbb{E}[XY] - \mathbb{E}[X] \mathbb{E}[Y]$
4	$\text{Cov}(X, Y) = \text{Cov}(Y, X)$
5	$\text{Cov}(aX + bY, Z) = a \text{Cov}(X, Z) + b \text{Cov}(Y, Z)$
6	$\text{Cov}(X, X) = \text{Var}(X)$
7	$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \text{Cov}(X, Y)$
8	$X$ and $Y$ indep. $\implies \text{Cov}(X, Y) = 0$

1.13 Show that the correlation coefficient always lies between  $-1$  and  $1$ . (Hint: use the fact that the variance of  $aX + Y$  is always nonnegative, for any  $a$ .)

1.14 Consider **Examples 1.1–1.2**. Define  $X$  as the function that assigns the number  $x_1 + \dots + x_n$  to each outcome  $\omega = (x_1, \dots, x_n)$ . The event that there are exactly  $k$  heads in  $n$  throws can be written as

$$\{\omega \in \Omega : X(\omega) = k\}.$$

If we abbreviate this to  $\{X = k\}$  and further abbreviate  $\mathbb{P}(\{X = k\})$  to  $\mathbb{P}(X = k)$ , then we obtain exactly (1.7). Verify that one can always view random variables in this way, that is, as real-valued functions on  $\Omega$ , and that probabilities such as  $\mathbb{P}(X \leq x)$  should be interpreted as  $\mathbb{P}(\{\omega \in \Omega : X(\omega) \leq x\})$ .

1.15 Show that

$$\text{Var} \left( \sum_{i=1}^n X_i \right) = \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i < j} \text{Cov}(X_i, X_j).$$

1.16 Let  $\Sigma$  be the covariance matrix of a random column vector  $\mathbf{X}$ . Write  $\mathbf{Y} = \mathbf{X} - \boldsymbol{\mu}$ , where  $\boldsymbol{\mu}$  is the expectation vector of  $\mathbf{X}$ . Hence,  $\Sigma = \mathbb{E}[\mathbf{Y}\mathbf{Y}^T]$ . Show that  $\Sigma$  is positive semidefinite. That is, for any vector  $\mathbf{u}$ , we have  $\mathbf{u}^T \Sigma \mathbf{u} \geq 0$ .

1.17 Suppose  $Y \sim \text{Gamma}(n, \lambda)$ . Show that for all  $x \geq 0$

$$\mathbb{P}(Y \leq x) = 1 - \sum_{k=0}^{n-1} \frac{e^{-\lambda x} (\lambda x)^k}{k!}. \quad (1.1)$$

1.18 Consider the random experiment where we draw uniformly and independently  $n$  numbers,  $X_1, \dots, X_n$ , from the interval  $[0, 1]$ .

- (a) Let  $M$  be the smallest of the  $n$  numbers. Express  $M$  in terms of  $X_1, \dots, X_n$ .  
 (b) Determine the pdf of  $M$ .

1.19 Let  $Y = e^X$ , where  $X \sim N(0, 1)$ .

- (a) Determine the pdf of  $Y$ .  
 (b) Determine the expected value of  $Y$ .

1.20 We select a point  $(X, Y)$  from the triangle  $(0, 0) - (1, 0) - (1, 1)$  in such a way that  $X$  has a uniform distribution on  $(0, 1)$  and the conditional distribution of  $Y$  given  $X = x$  is uniform on  $(0, x)$ .

- (a) Determine the joint pdf of  $X$  and  $Y$ .  
 (b) Determine the pdf of  $Y$ .  
 (c) Determine the conditional pdf of  $X$  given  $Y = y$  for all  $y \in (0, 1)$ .  
 (d) Calculate  $\mathbb{E}[X \mid Y = y]$  for all  $y \in (0, 1)$ .  
 (e) Determine the expectations of  $X$  and  $Y$ .

### Poisson Processes

1.21 Let  $\{N_t, t \geq 0\}$  be a Poisson process with rate  $\lambda = 2$ . Find

- (a)  $\mathbb{P}(N_2 = 1, N_3 = 4, N_5 = 5)$ ,  
 (b)  $\mathbb{P}(N_4 = 3 \mid N_2 = 1, N_3 = 2)$ ,  
 (c)  $\mathbb{E}_i N_4 \mid N_2 = 2$ ,  
 (d)  $\mathbb{P}(N[2, 7] = 4, N[3, 8] = 6)$ ,  
 (e)  $\mathbb{E}[N[4, 6] \mid N[1, 5] = 3]$ .

1.22 Show that for any fixed  $k \in \mathbb{N}$ ,  $t > 0$ , and  $\lambda > 0$ ,

$$\lim_{n \rightarrow \infty} \binom{n}{k} \left(\frac{\lambda t}{n}\right)^k \left(1 - \frac{\lambda t}{n}\right)^{n-k} = \frac{(\lambda t)^k}{k!} e^{-\lambda t}.$$

(Hint: write out the binomial coefficient and use the fact that  $\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda t}{n}\right)^n = e^{-\lambda t}$ .)

1.23 Consider the Bernoulli approximation in Section 1.11. Let  $U_1, U_2, \dots$  denote the times of success for the Bernoulli process  $X$ .

- (a) Verify that the “intersuccess” times  $U_1, U_2 - U_1, \dots$  are independent and have a geometric distribution with parameter  $p = \lambda h$ .  
 (b) For small  $h$  and  $n = \lfloor t/h \rfloor$ , show that the relationship  $\mathbb{P}(A_1 > t) \approx \mathbb{P}(U_1 > n)$  leads in the limit, as  $n \rightarrow \infty$ , to

$$\mathbb{P}(A_1 > t) = e^{-\lambda t}.$$

1.24 If  $\{N_t, t \geq 0\}$  is a Poisson process with rate  $\lambda$ , show that for  $0 \leq u \leq t$  and  $j = 0, 1, 2, \dots, n$ ,

$$\mathbb{P}(N_u = j | N_t = n) = \binom{n}{j} \left(\frac{u}{t}\right)^j \left(1 - \frac{u}{t}\right)^{n-j},$$

that is, the conditional distribution of  $N_u$  given  $N_t = n$  is binomial with parameters  $n$  and  $u/t$ .

### Markov Processes

1.25 Determine the (discrete) pdf of each  $X_n$ ,  $n = 0, 1, 2, \dots$  for the random walk in **Example 1.10**. (That is, the random walk on the integers, with transition graph in Figure 1.1 and starting at 0.)

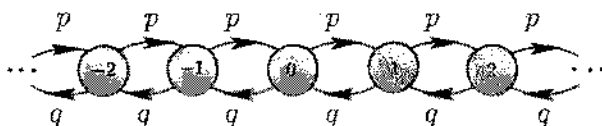


Figure 1.1 Transition graph for a random walk on  $\mathbb{Z}$ .

Also, calculate  $\mathbb{E}[X_n]$  and the variance of  $X_n$  for each  $n$ .

1.26 Let  $\{X_n, n \in \mathbb{N}\}$  be a Markov chain with state space  $\{0, 1, 2\}$ , transition matrix

$$P = \begin{pmatrix} 0.3 & 0.1 & 0.6 \\ 0.4 & 0.4 & 0.2 \\ 0.1 & 0.7 & 0.2 \end{pmatrix},$$

and initial distribution  $\pi = (0.2, 0.5, 0.3)$ . Determine

- $\mathbb{P}(X_1 = 2)$ ,
- $\mathbb{P}(X_2 = 2)$ ,
- $\mathbb{P}(X_3 = 2 | X_0 = 0)$ ,
- $\mathbb{P}(X_0 = 1 | X_1 = 2)$ ,
- $\mathbb{P}(X_1 = 1, X_3 = 1)$ .

1.27 Consider two dogs harboring a total number of  $m$  fleas. Spot initially has  $b$  fleas and Lassie has the remaining  $m - b$ . The fleas have agreed on the following immigration policy: at every time  $n = 1, 2, \dots$  a flea is selected at random from the total population and that flea will jump from one dog to the other. Describe the flea population on Spot as a Markov chain and find its stationary distribution.

1.28 Classify the states of the Markov chain with the following transition matrix:

$$P = \begin{pmatrix} 0.0 & 0.3 & 0.6 & 0.0 & 0.1 \\ 0.0 & 0.3 & 0.0 & 0.7 & 0.0 \\ 0.3 & 0.1 & 0.6 & 0.0 & 0.0 \\ 0.0 & 0.1 & 0.0 & 0.9 & 0.0 \\ 0.1 & 0.1 & 0.2 & 0.0 & 0.6 \end{pmatrix}.$$

1.29 Consider the following snakes-and-ladders game in Figure 1.2. Let  $N$  be the number of tosses required to reach the finish using a fair die. Calculate the expectation of  $N$  using a computer.

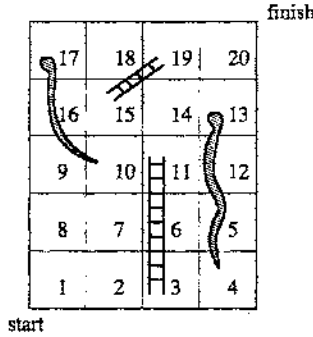


Figure 1.2 A snakes-and-ladders game.

1.30 Ms. Ella Brum walks back and forth between her home and her office every day. She owns three umbrellas, which are distributed over two umbrella stands (one at home and one at work). When it is not raining, Ms. Brum walks without umbrella. When it is raining, she takes one umbrella from the stand at the place of her departure, provided there is one available. Suppose the probability that it is raining at the time of any departure is  $p$ . Let  $X_n$  denote the number of umbrellas available at the place where Ella arrives after walk number  $n$ ;  $n = 1, 2, \dots$ , including the one that she possibly brings with her. Calculate the limiting probability that it rains and no umbrella is available.

1.31 A mouse is let loose in the maze of Figure 1.3. From each compartment the mouse chooses one of the adjacent compartments with equal probability, independent of the past. The mouse spends an exponentially distributed amount of time in each compartment. The mean time spent in each of the compartments 1, 3, and 4 is two seconds; the mean time spent in compartments 2, 5, and 6 is four seconds. Let  $\{X_t, t \geq 0\}$  be the Markov jump process that describes the position of the mouse for times  $t \geq 0$ . Assume that the mouse starts in compartment 1 at time  $t = 0$ .

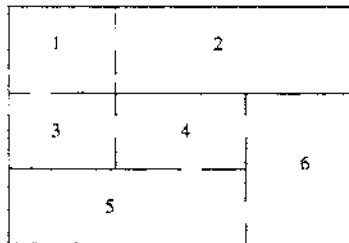


Figure 1.3 A maze.

What are the probabilities that the mouse will be found in compartments 1, 2,  $\dots$ , 6 at some time  $t$  far away in the future?

1.32 In an  $M/M/\infty$ -queueing system, customers arrive according to a Poisson process with rate  $a$ . Every customer who enters is immediately served by one of an infinite number of servers; hence, there is no queue. The service times are exponentially distributed, with mean  $1/b$ . All service and interarrival times are independent. Let  $X_t$  be the number of customers in the system at time  $t$ . Show that the limiting distribution of  $X_t$ , as  $t \rightarrow \infty$ , is Poisson with parameter  $a/b$ .

### Optimization

1.33 Let  $\mathbf{a}$  and  $\mathbf{x}$  be  $n$ -dimensional column vectors. Show that  $\nabla_{\mathbf{x}} \mathbf{a}^T \mathbf{x} = \mathbf{a}$ .

1.34 Let  $A$  be a symmetric  $n \times n$  matrix and let  $\mathbf{x}$  be an  $n$ -dimensional column vector. Show that  $\nabla_{\mathbf{x}} \frac{1}{2} \mathbf{x}^T A \mathbf{x} = A \mathbf{x}$ . What is the gradient if  $A$  is not symmetric?

1.35 Show that the optimal distribution  $\mathbf{p}^*$  in **Example 1.17** is given by the uniform distribution.

1.36 Derive the program (1.78).

1.37 Consider the MinxEnt program

$$\begin{aligned} \min_{\mathbf{p}} \quad & \sum_{i=1}^n p_i \ln \frac{p_i}{q_i} \\ \text{subject to: } & \mathbf{p} \geq \mathbf{0}, \quad A\mathbf{p} = \mathbf{b}, \quad \sum_{i=1}^n p_i = 1, \end{aligned}$$

where  $\mathbf{p}$  and  $\mathbf{q}$  are probability distribution vectors and  $A$  is an  $m \times n$  matrix.

(a) Show that the Lagrangian for this problem is of the form

$$\mathcal{L}(\mathbf{p}, \boldsymbol{\lambda}, \beta, \boldsymbol{\mu}) = \mathbf{p}^T \boldsymbol{\xi}(\mathbf{p}) - \boldsymbol{\lambda}^T (A\mathbf{p} - \mathbf{b}) - \boldsymbol{\mu}^T \mathbf{p} + \beta(1^T \mathbf{p} - 1).$$

(b) Show that  $p_i = q_i \exp(-\beta - 1 + \mu_i + \sum_{j=1}^m \lambda_j a_{ji})$  for  $i = 1, \dots, n$ .

(c) Explain why, as a result of the KKT conditions, the optimal  $\boldsymbol{\mu}^*$  must be equal to the zero vector.

(d) Show that the solution to this MinxEnt program is exactly the same as for the program where the nonnegativity constraints are omitted.

