Contents

Chapter 1

Introduction

Exercises: 1.5.

Chapter 2

Set theory, probability and combinatorics

Definition 2.1. The result of a random experiment is called an *outcome*. The collection of all possible outcomes, denoted by Ω is called a *sample space*. An *event* is a subset of the sample space.

Definition 2.2. A collection $\mathcal{F}$ of events is a *sigma-algebra* on sample space if the following is satisfied:

1. $\Omega \in \mathcal{F}$.
2. If an event $E \in \mathcal{F}$ then $E^c \in \mathcal{F}$.
3. If $E_1, E_2, \ldots \in \mathcal{F}$ then $E_1 \cup E_2 \cup \ldots \in \mathcal{F}$.

Example 2.1. Let $\Omega = \{1, 2, 3\}$, the set $A = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$ is a sigma-algebra on Ω but $B = \{\emptyset, \{1\}, \{1, 2, 3\}\}$ is not since $\{1\}^c = \{2, 3\}$ is not contained in B.

Definition 2.3. A set S is *countable* if there exists a bijection $b : S \to \mathbb{N}$.

Example 2.2. The set $\mathbb{N}$ is a countable set. Also the set $\mathbb{Z} = \{\ldots, -1, 0, 1, \ldots\}$ is countable. The set $\mathbb{R}$ is an uncountable set.

Definition 2.4. Given a sample space and a sigma algebra $\mathcal{F}$ on it. A function $\mathbb{P} : \mathcal{F} \to [0, 1]$ satisfying the following:

1. $\mathbb{P}(\Omega) = 1$.
2. For any finite or countable collection of mutually disjoint events $E_1, E_2, \ldots \in \mathcal{F}$ we have $\mathbb{P}(E_1 \cup E_2 \cup \ldots) = \mathbb{P}(E_1) + \mathbb{P}(E_2) + \ldots$

is called a *probability* function.

Definition 2.5. The *inclusion-exclusion principle* for three sets A, B, C is given by $|A \cup B \cup C| = |A| + |B| + |C| - |A \cap B| - |A \cap C| - |B \cap C| + |A \cap B \cap C|$.

Theorem 2.1. *Let E_1 and E_2 be two arbitrary events of the sample space Ω. The following is true for a probability measure:*

1. $\mathbb{P}(E_1^c) = 1 - \mathbb{P}(E_1)$.
2. $\mathbb{P}(\emptyset) = 0$.
3. $\mathbb{P}(E_1 \cup E_2) = \mathbb{P}(E_1) + \mathbb{P}(E_2) - \mathbb{P}(E_1 \cap E_2)$.

The last claim can be generalized to more than two events by using the inclusion-exclusion principle.

Definition 2.6. Events $E_1, E_2, \ldots, E_n$ are *independent* if $\mathbb{P}(E_1 \cap E_2 \cap \ldots \cap E_n) = \mathbb{P}(E_1)\mathbb{P}(E_2)\ldots\mathbb{P}(E_n)$.

Theorem 2.2. *The following table gives formulas for how to count favorable outcomes, depending on how we select our outcomes-ordered and unordered, with and without repetition-of k things from an n-set.*

	Ordered	*Unordered*
Without repetition	$n(n-1)\ldots(n-k+1)$	$\binom{n}{k}$
With repetition	n^k	$\binom{n+k-1}{k}$

Definition 2.7. Let E_1 and E_2 be two events where $\mathbb{P}(E_1) > 0$. The *conditional probability* for the event E_2, given that E_1 occured, denoted by $\mathbb{P}(E_2|E_1)$ is defined as

$$\mathbb{P}(E_2|E_1) = \frac{\mathbb{P}(E_2 \cap E_1)}{\mathbb{P}(E_1)}$$

Theorem 2.3. *Let E_1 and E_2 be two events. The following holds:*

$$\mathbb{P}(E_2|E_1) = \mathbb{P}(E_2)$$

if and only if E_1 and E_2 are independent.

Theorem 2.4 (Bayes theorem)**.** *Let E_1 and E_2 be two events. The following hold:*

$$\mathbb{P}(E_2|E_1) = \frac{\mathbb{P}(E_1|E_2)\mathbb{P}(E_2)}{\mathbb{P}(E_1)}$$

Definition 2.8. A *partition* of a set X is a set $B = \{B_1, B_2, \ldots, B_k\}$ of nonempty subsets of X such that every element x in X is contained in one and only one of the subsets. That is:

1. $\emptyset \notin B$.
2. For $B_i, B_j \in B$, $B_i \cap B_j = \emptyset$ whenever $i \neq j$.

3. $B_1 \cup B_2 \cup \ldots \cup B_k = X$

Theorem 2.5 (The law of total probability)**.** *Let $E_1, E_2, \ldots, E_k$ be a partition of the sample space and let A be an event, we then have*

$$\mathbb{P}(A) = \sum_{i=1}^{k} \mathbb{P}(A|E_i)\mathbb{P}(E_i)$$

Exercises: $2.1, \mathbf{2.4}, 2.5, 2.7, 2.11, \mathbf{2.14}, \mathbf{2.16}, 2.17, 2.20, 2.25, 2.28, 2.30, \mathbf{2.32}$.

Chapter 3

Discrete random variables

Definition 3.1. A *random variable* is a function $X : \Omega \to \mathbb{E}$ for some set E. Once an experiment is completed, and an outcome ω is known, the value of the random variable $X(\omega)$ becomes determined. We call the value of the function an *observation* of the random variable.

Definition 3.2. A random variable where the range E is a finite or countable set is called a *discrete random variable*. A random variable where the range E is an uncountable set is called a *continuous random variable*.

Definition 3.3. The *probability mass function* is defined for a discrete random variable as

$$P(x) = \mathbb{P}(\omega : X(\omega) = x)$$

and the *cumulative distribution function* for a discrete random variable is defined as

$$F(x) = \sum_{\omega : X(\omega) \leq x} \mathbb{P}(\omega)$$

Definition 3.4. Let X and Y be two random variables on the same *probability space* $(\Omega, \mathcal{F}, \mathbb{P})$, the pair (X, Y) is a *random vector*. The *joint distribution* of X and Y is defined as $P_{X,Y}(x, y) := \mathbb{P}(X = x \cap Y = y)$. The *marginal distributions* for X and Y are defined as $P_X(x) = \mathbb{P}(X = x)$ and $P_Y(y) = \mathbb{P}(Y = y)$.

Theorem 3.1. *Let (X, Y) be a random vector with the joint probability distribution $P_{X,Y}(x, y)$, we have that:*

1. $P_X(x) = \sum_y P_{X,Y}(x, y)$
2. $P_Y(y) = \sum_x P_{X,Y}(x, y)$

Definition 3.5. The random variables X and Y are *independent* if

$$P_{X,Y}(x, y) = P_X(x)P_Y(y)$$

for all values of x and y.

Definition 3.6. The *expectation* of a discrete random variable is defined as

$$\mu = \mathbb{E}[X] = \sum_x xP(x)$$

Definition 3.7. The *expectation* of a function of a discrete random variable is defined as

$$\mathbb{E}[g(X)] = \sum_x g(x)P(x)$$

Theorem 3.2. *Let X and Y be two discrete random variables and a, b, c non-random numbers. The following hold:*

$$\mathbb{E}[aX + bY + c] = a\mathbb{E}[X] + b\mathbb{E}[Y] + c$$

Theorem 3.3. *Let X and Y be two independent random variables, then*

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$$

Definition 3.8. The *variance* of a discrete random variable X is defined as

$$\sigma^2 = Var(X) = \mathbb{E}[X - \mathbb{E}[X]]$$

Theorem 3.4. *For a random variable X, the following hold for the variance:*

$$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$

Definition 3.9. The *standard deviation* of a random variable is defined as

$$\sigma = Std(X) = \sqrt{Var(X)}$$

Definition 3.10. Let X and Y be to random variables. The *covariance* is defined as

$$\sigma_{XY} = Cov(X, Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$

Definition 3.11. Let X and Y be to random variables. The *correlation coefficient* is defined as

$$\rho = \frac{Cov(X, Y)}{Std(X)Std(Y)}$$

Theorem 3.5. *The* Cauchy Schwartz inequality *for random variables X and Y:*

$$|\mathbb{E}[XY]| \leq \sqrt{\mathbb{E}[X^2]}\sqrt{\mathbb{E}[Y^2]}$$

Note: Cauchy Schwartz inequality implies that $-1 \leq \rho \leq 1$.

Theorem 3.6. *For random variables X, Y, Z and W, and for non-random numbers a, b, c and d, the following hold:*

1. $Var(aX + bY + c) = a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)$.
2. $Cov(aX+bY, cZ+dW) = acCov(X, Z)+adCov(X, W)+bcCov(Y, Z)+ bdCov(Y, W)$.
3. $Cov(X, Y) = Cov(Y, X)$.

Theorem 3.7. *For independent random variable X and Y, the following hold:*

1. $Cov(X, Y) = 0$.
2. $Var(X, Y) = Var(X) + Var(Y)$.

Theorem 3.8. *The* Chebyshev's inequality *for a random variable X with expectation μ, variance σ and $\epsilon > 0$:*

$$\mathbb{P}(|X - \mu| > \epsilon) \leq (\frac{\sigma}{\epsilon})^2$$

Definition 3.12. A random variable X with two possible values is called a *Bernoulli variable* and its distribution is *Bernoulli distribution.* Any experiment with a binary outcome is called *Bernoulli trial.* The distribution is given by

$$P(x) = \begin{cases} p & \text{if } x = 1 \\ q = 1 - p & \text{if } x = 0 \end{cases}$$

Theorem 3.9. *Let X be a Bernoulli variable and let p denote the probability of a success in the Bernoulli trial, then*

1. $\mathbb{E}[X] = p$.
2. $Var(X) = pq$

Definition 3.13. A random variable X, counting the number of successes in a sequence of independent Bernoulli trials has a *Binomial distribution.* Its parameters are the number of trials, n and the probability of success, p. The distribution is given by

$$P(x) = \binom{n}{x} p^x q^{n-x}$$

Theorem 3.10. *Let X be a Binomial variable. Let p denote the probability of a success in a Bernoulli trial and n the number trials, then*

1. $\mathbb{E}[X] = np$.
2. $Var(X) = npq$.

Definition 3.14. The random variable X, which counts the number of independent Bernoulli trials needed to get the first success has *Geometric distribution.* The distribution is given by

$$P(x) = q^{x-1}p$$

Theorem 3.11. *Let X be a Geometric variable and let p denote the probability of a success in the Bernoulli trial, then*

1. $\mathbb{E}[X] = \frac{1}{p}$.
2. $Var(X) = \frac{q}{p2}$.

Definition 3.15. The random variable X, which counts the number of independent Bernoulli trials needed to obtain k successes has *Negative Binomial distribution.* The distribution is given by

$$P(x) = \binom{x-1}{k-1} p^k q^{x-k}$$

Theorem 3.12. *Let X be a Negative binomial variable. Let p denote the probability of a success in the Bernoulli trial and let k denote the number of successes, then*

1. $\mathbb{E}[X] = \frac{k}{p}$.
2. $Var(X) = \frac{kq}{p2}$.

Definition 3.16. The random variable X, which counts the number of rare events occurring within a fixed period of time has *Poisson distribution.* The distribution is given by

$$P(x) = e^{-\lambda}\frac{\lambda^x}{x!}$$

where λ denotes the frequency of the events.

Theorem 3.13. *Let X be a Poisson variable. Let λ denote the frequency/rate of the events, then*

1. $\mathbb{E}[X] = \lambda$.
2. $Var(X) = \lambda$.

Exercises: 3.1, **3.2**, 3.5, 3.9, 3.11, **3.12**, 3.13, 3.16, 3.22, **3.24**, 3.27, 3.28, **3.30**, 3.37, **3.38**.