Introduction to Probability

Arazim©*

Contents

*Based on Eilon Solan's lecture notes with additional material from Nattalie Tamam and Dudu Lagziels recitation summary.

1 Probability spaces

Definition 1. A probability space is a pair $(\Omega, \mathbb{P})$ such that:

- Ω is a finite set.
- $\mathbb{P} : \Omega \mapsto [0,1]$ is a function which fulfils the requirement $\sum_{\omega\in\Omega)} \mathbb{P}(\omega) = 1$.

Ω is sometimes referred to as a probability space and $\mathbb{P}$ is called a probability function.

1.1 Events

Definition 2. An event is a subset A of Ω.

Definition 3. Let $(\Omega, \mathbb{P})$ be a probability space. For any event $A \subseteq \Omega$, the **probability** of A is defined as

$$P(A) \coloneqq \sum_{\omega\in A} \mathbb{P}(\omega)$$

Theorem 1. *Let $(\Omega, \mathbb{P})$ be a probability space. The following always occur:*

1. $\mathbb{P}(\emptyset) = 0$*: the probability of the empty event is 0.*
2. $P(\Omega)$*: The probability of the whole space is 1.*
3. $P(A^c) = 1 - P(A)$*: The probability of the complement of an event is equal to 1 minus the probability of the event.*
4. $0 < \mathbb{P} \leq 1$ *for any event* $A \subseteq \Omega$*: The probability of* A *is between 1 and 0.*
5. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$ *for any two disjoint events* $A, B \subseteq \Omega$*.*

Definition 4. The events $A_1, A_2, \ldots, A_n$ are **pairwise disjoint** if $A_i \cap A_j = \emptyset$ for all $1 \leq i < j \leq n$.

Theorem 2. *Let $A_1, A_2, \ldots, A_n$ be a set of pairwise disjoint events, then:*

$$\mathbb{P}\left(\cup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

1.2 Uniform distribution

Definition 5. Let $(\Omega, \mathbb{P})$ be a probability space. The distribution is called a **uniform distribution** if:

$$\forall\omega \in \Omega . \mathbb{P}(\omega) = \frac{|1|}{|\Omega|}$$

Theorem 3. *If $\mathbb{P}$ is a uniform distribution then for any $A \subseteq \Omega$*

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|}$$

1.3 Unions of events

Theorem 4. *Let A and B be two events, then*

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Theorem 5 (Union bound)**.** *Let $A_1, A_2, \ldots, A_n$ be events. Then:*

$$\mathbb{P}\left(\cup_{i=1}^{n} A_i\right) \leq \sum_{i=1}^{n} \mathbb{P}(A_i)$$

Theorem 6 (General Inclusion-Exclusion Principle)**.** *Let $A_1, A_2, \ldots, A_n$ be events. Then:*

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_n) = \sum_{k=1}^{n} \sum_{1\leq i_1<i_2<\ldots i_k} (-1)^{k+1} \mathbb{P}\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = \sum_{I\subseteq\{1,2,\ldots,n\}} (-1)^{|I|+1} \mathbb{P}\left(\cap_{i\in I} A_i\right)$$

2 Conditional probability and independence

2.1 Conditional probability

Definition 6. Let B be an event with positive probability ($\mathbb{P}(B) > 0$). For any $\omega \in \Omega$ we will define

$$\mathbb{P}\left(\omega|B\right) := \begin{cases} 0 & \omega \notin B \\ \frac{\mathbb{P}(\omega)}{\mathbb{P}(B)} & \omega \in B \end{cases}$$

The weight $\mathbb{P}\left(\omega|B\right)$ is called the **conditional probability** of the result ω given B.

Theorem 7. *For any event B with a positive probability, the function $\mathbb{P}\left(\cdot|B\right) : \Omega \to \mathbb{R}$ in the previous definition defines a probability function.*

Theorem 8. *Let B be an event such that $\mathbb{P}(B) > 0$ then for all events A:*

$$\mathbb{P}\left(A|B\right) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Theorem 9 (General product rule)**.** *Let $A_1, A_2, \ldots, A_n$ be events for which $\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) > 0$. Then:*

$$\mathbb{P}(A_1 \cap A_2 \cap \ldots A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}\left(A_2|A_1\right) \cdot \mathbb{P}\left(A_3|A_1 \cap A_2\right) \cdot \cdots \cdot \mathbb{P}\left(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}\right)$$

2.2 Independence

Definition 7. Two events are called independent if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Theorem 10. *Let A, B be events.*

1. *If A and B are independent, then so are A and B^c.*
2. *If A and B are independent, then so are A^c and B.*
3. *If A and B are independent, then so are A^c and B^c.*
4. *A and $\emptyset$ are independent.*
5. *A and Ω are independent.*

Theorem 11. *Let A be an event and B be an event with positive probability. The events A, B are disjoint iff $\mathbb{P}(A) = \mathbb{P}\left(A|B\right)$*

2.3 Product spaces

Definition 8. Let $(\Omega_1, \mathbb{P}_1)$ and $(\Omega_2, \mathbb{P}_2)$ be probability spaces. Their **product space** is defined as follows:

1. $\Omega = \Omega_1 \times \Omega_2 = \left\{(\omega_1, \omega_2) : \omega_1 \in \Omega_1 \wedge \omega_2 \in \Omega_2\right\}$
2. $\mathbb{P}(\omega_1, \omega_2) = \mathbb{P}_1(\omega_1) \cdot \mathbb{P}(\omega_2)$

Theorem 12. *The space $(\Omega, \mathbb{P})$ is a probability space.*

Definition 9. Let $(\Omega_1, \mathbb{P}_1)$ and $(\Omega_2, \mathbb{P}_2)$ be probability spaces and let $(\Omega, \mathbb{P})$ be their product space. We will say that the event $A \subseteq \Omega$ refers to the first space if there exists a $B_1 \subseteq \Omega_1$ such that $A = B_1 \times \Omega_2$ and we We will say that the event $A \subseteq \Omega$ refers to the second space if there exists a $B_2 \subseteq \Omega_2$ such that $A = \Omega_1 \times B_2$.

Theorem 13. *Let* $(\Omega_1, \mathbb{P}_1)$ *and* $(\Omega_2, \mathbb{P}_2)$ *be probability spaces, let* $(\Omega, \mathbb{P})$ *be their product space and let* $A_1, A_2 \subseteq \Omega$*. If* A_1 *refers to the first space and* A_2 *refers to the second then* A_1 *and* A_2 *are independent.*

Definition 10. The events $A_1, A_2, \ldots, A_n$ are called independent if for all $k \leq n$

$$\mathbb{P}(A_1 \cap A_2 \cap \cdots \cap A_n) = \mathbb{P}(A_1) \cdot \mathbb{P}(A_2) \cdot \cdots \cdot \mathbb{P}(A_n)$$

Theorem 14. *If the events* $A_1, A_2, \ldots, A_n$ *are independent then so are* $A_1, A_2, \ldots A_{n-1}$*.*

Definition 11. The events $A_1, A_2, \ldots, A_n$ are called **pairwise independent** if for all i, j such that $1 \leq i < j \leq n$, the events A_i, A_j are independent.

Theorem 15. *Let* $A_1, A_2, \ldots, A_n$ *be independent events. The events* $B_1, B_2, \ldots, B_n$ *are also independent where for all* $i \in \{1, 2, \ldots, n\}$ *we have that* B_i *equals to either* A_i *or* A_i^c*.*

Definition 12. Let $n \in \mathbb{N}$ and $(\Omega, \mathbb{P})$ be a probability space in which $\Omega = \{1, 2, \ldots, n\}$. The distribution function $\mathbb{P}$ is called a **binomial distribution function** with the parameters n and pj if:

$$\mathbb{P}(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

Note 1. Sometimes, we will mark binomial distribution with the parameters n and p with $B(n, p)$ or $Bin(n, p)$

2.4 Coupling

Definition 13. Let $(\Omega_1, \mathbb{P}_1)$ and $(\Omega_2, \mathbb{P}_2)$ be probability spaces and let $f : \Omega_1 \mapsto \Omega_2$ be a function. For every $\omega_2 \in \Omega_2$ we will define

$$f^{-1}(\omega_2) = \left\{\omega_1 \in \Omega_1 : f(\omega_1) = \omega_2\right\}$$

The set $f^{-1}(\omega_2) \subseteq \Omega_1$ is called the **inverse image** of ω_1.

Theorem 16. *Let* $(\Omega_1, \mathbb{P}_1)$ *and* $(\Omega_2, \mathbb{P}_2)$ *be probability spaces and let* $f : \Omega_1 \mapsto \Omega_2$ *be a function. If* $\omega_2, \omega_2' \in \Omega_2$ *are two different events then* $f^{-1}(\omega_2) \cap f^{-1}\left(\omega_2'\right) = \emptyset$*.*

Definition 14. Let $(\Omega_1, \mathbb{P}_1)$ and $(\Omega_2, \mathbb{P}_2)$ be probability spaces. $f : \Omega_1 \mapsto \Omega_2$ is called a **probability preserving function** if for all $\omega_2 \in \Omega_2$ we have

$$\mathbb{P}_2(\omega_2) = \mathbb{P}_1\left(f^{-1}(\omega_2)\right) = \sum_{\omega_1 \in f^{-1}(\omega_2)} \mathbb{P}_1(\omega_1)$$

Theorem 17. *Let* $(\Omega_1, \mathbb{P}_1)$ *and* $(\Omega_2, \mathbb{P}_2)$ *be probability spaces,* $f : \Omega_1 \mapsto \Omega_2$ *is a probability preserving function and* $A_2 \subseteq \Omega_2$ *be an event then we will define*

$$A_1 \coloneqq \cup_{\omega_2 \in A_2} f^{-1}(\omega_2) \;\; \textit{then } \mathbb{P}_1(A_1) = \mathbb{P}_2(A_2)$$

Definition 15. Let $(\Omega_1, \mathbb{P}_1)$ and $(\Omega_2, \mathbb{P}_2)$ be probability spaces. A **coupling** between the two spaces is composed of:

- A probability space $(\Omega, \mathbb{P})$
- A probability preserving function $f_1 : \Omega_1 \mapsto \Omega$
- and a probability preserving function $f_2 : \Omega_2 \mapsto \Omega$

Theorem 18. *Let* $(\Omega_1, \mathbb{P}_1)$ *and* $(\Omega_2, \mathbb{P}_2)$ *be probability spaces and let* $(\Omega, \mathbb{P}, f_1, f_2)$ *be a coupling between them where* $A_1 \subseteq \Omega_1$ *and* $A_2 \subseteq \Omega_2$ *be events. If* $\cup_{\omega_1 \in A_1} f_1^{-1}(\omega_1) \subseteq \cup_{\omega_2 \in A_2} f_2^{-1}(\omega_2)$ *then* $\mathbb{P}_1(A_1) \leq \mathbb{P}_2(A_2)$*.*

Theorem 19 (Total probability formula)**.** *Let A be an event and $\{B_1, B_2, \dots, B_n\}$ is a partition of Ω to disjoint sets with positive probability. Which means that $\cup_{i=1}^{n} B_i = \Omega$ and $\mathbb{P}(B_i) > 0$ for all $i \in \{1, 2, \dots, n\}$ then:*

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(B_i)\, \mathbb{P}\left(A|B_i\right)$$

Theorem 20 (Bayes rule)**.** *Let A be an event with positive probability and B is an event such that $0 < \mathbb{P}(B) < 1$ then:*

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

2.5 Conditional Independence

Theorem 21. *Let C be an event with positive probability. We will mark with Q the conditional probability given C:*

$$Q(A) = \mathbb{P}\left(A|C\right), \forall A \subseteq \Omega$$

Let B be an event for which $\mathbb{P}(B \cap C) > 0$ and let A be an event. Then $Q\left(A|B\right) = \mathbb{P}(B \cap C)$

Definition 16. Let A, B and C be three events and let us assume that $\mathbb{P}(C) > 0$. The events A and B are called **independent given the event C** if they are independent in the probability space $\left(\Omega, \mathbb{P}\left(\cdot|C\right)\right)$

Theorem 22. *If the events A, B and C are independent and if $\mathbb{P}(C) > 0$ then the events A and B are independent given C*

3 Countable Probability Spaces

Definition 17. A set Ω is called **countable** if there exists a one-to-one and onto function $f : \Omega \to \mathbb{N}$

Definition 18. A **countable probability space** is a pair $(\Omega, \mathbb{P})$ in which Ω is a countable set and $\mathbb{P} : \Omega \to [0,1]$ is a function for which $\sum_{\omega \in \Omega} \mathbb{P}(\omega) = 1$

3.1 Geometric Distribution

Definition 19. Let $p \in [0,1)$. Let $(\Omega, \mathbb{P})$ be a probability space where the sample space is $\Omega = \mathbb{N}$ and the probability function $\mathbb{P}$ is given by

$$\mathbb{P}(n) = p(1-p)^{n-1}, \forall n \in \Omega$$

The probability function $\mathbb{P}$ is called a **geometric distribution with the parameter** p.

Theorem 23. *Let $\mathbb{P}$ be geometric distribution with the parameter p on the sample space $\Omega \coloneqq \mathbb{N}$ and define the set $B_n \coloneqq \{n, n+1, n+2, \dots\}$then for all $n \in \mathbb{N}$ we have*

$$\mathbb{P}\left(\{n\}|B_n\right) = p$$

3.2 Poisson distribution

Definition 20. Let $(\Omega, \mathbb{P})$ be a probability space where the sample space is $\Omega = \{0, 1, 2, 3, \dots\}$ and the probability function is given by

$$\mathbb{P}(n) = \frac{e^{-\lambda}\lambda^n}{n!}, \forall n \in \Omega$$

Theorem 24. *Let the sequence of random variables $X_1, X_2, \ldots$ such that $X_n \sim \mathrm{Bin}(n, \mathbb{P}_n)$. Given that $n \cdot \mathbb{P}_n \to \lambda$ for some constant λ, then for any constant k we have:*

$$\mathbb{P}(X_n = k) \xrightarrow{n\to\infty} e^{-\lambda}\frac{\lambda^k}{k!}$$

This theorem has a generalization where there is no need to look at a random variable X_n which consists of n independent experiments. Conversely, we can look at the sum of n Bernoulli independent random variables such that every event has the value 1 with the probability $\mathbb{P}_{n,k}$ for $k = 1, \ldots, n$ and $\sum_{k=1}^{n} \mathbb{P}_{n,k} \to \lambda$.

3.3 Additivity of independent random variables

Lemma 1 (Addition of two Poisson random variables)**.** *Let $X_1 \sim P(\lambda_1), X_2 \sim P(\lambda_2)$ be two random variables, then $X_1 + X_2 \sim P(\lambda_1 + \lambda_2)$*

Lemma 2. *Let* $\mathrm{Bin}(n, p)$ *and* $X_2 \sim \mathrm{Bin}(m, p)$ *be two random variables, then* $X_1 + X_2 \sim \mathrm{Bin}(n + m, p)$

4 Random Variables

Definition 21. A **random variable** is a function $X : \Omega \to \mathbb{R}$

Definition 22. Let $(\Omega, \mathbb{P})$ be a probability space and $A \subseteq \Omega$ an event. The **indicator random variable** of A is the random variable $\mathbf{1}_A$ which is defined as follows:

$$\mathbf{1}_A(\omega) := \begin{cases} 1 & \omega \in A \\ 0 & \omega \notin A \end{cases}$$

Definition 23. Let X be a random variable. A distribution function of X is a function which matches for every x_1 that the variable can attain the probability $\mathbb{P}(X = x_i)$.

Definition 24. Let $p \in [0, 1]$. A random variable X is called a **random bernoulli variable with the parameter** p if it attains only two values, 0 and 1 and if $\mathbb{P}(X = 1) = p$

Definition 25. Let n be a natural number and $p \in [0, 1]$. A random variable X is called a **binomial random variable with the parameters** n **and** p if its possible values are $0, 1, \ldots, n$ and

$$\mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \forall k \in \{0, 1, \ldots, n\}$$

Definition 26. Let r be a natural number and $p \in [0, 1]$. A random variable X is called a **negative binomial variable with the parametes** r **and** p if its possible values are $0, 1, \ldots, n$ and

$$\mathbb{P}(X = k) = \binom{k+r-1}{k} p^r (1-p)^k, \forall k \in \{0, 1, \ldots, n\}$$

Definition 27. Let r be a natural number and $p \in [0, 1]$. A random variable X is called a **hypergeometric random variable with the parameters** N **and** D if its possible values are $0, 1, \ldots, \min\{N, D\}$ and:

$$\mathbb{P}(X = k) = \frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$$

Definition 28. Let $(\Omega, \mathbb{P})$ be a probability space and let X and Y be two random variables defined on this probability space and get the values $(x_i)_{i=1}^n$ and $(y_j)_{i=1}^n$ respectively. The **joint probability** of the random variables X and Y is a function which matches for every pair (x_i, y_i) the probability $\mathbb{P}(X = x_i, Y = y_j)$.

Theorem 25. *For every pair of random variables X and Y which get the values $(x_i)_i$ and $(y_j)_j$ respectively, the following takes place:*

$$\sum_i \sum_j \mathbb{P}(X = x_i, Y = y_j) = 1$$

4.1 Random walks and the reflection principle

Definition 29. A **random walk** is a probability space $(\Omega, \mathbb{P})$ where the sample space is $\Omega = \{-1, 1\}^n$ and the distribution is $\mathbb{P}(\omega) = \frac{1}{2^n}$ for all $\omega \in \Omega$: the distribution is uniform.

Note 2 (The reflection principle). Let $A = (0, a)$ and $B = (n, b)$ be two points with $a, b > 0$. Let $A' = (0, -a)$ be the reflection of A through the x-axis. The number of paths from A to B that touch the x-axis is the same as the number of paths from A' to B

4.2 Independent random variables

Definition 30. Let X and Y be random variables which get the values $(x_i)_i$ and $\left(y_j\right)_j$ respectively. The random variables X and Y are called **independent** if for all i and j the events $\{X = x_i\}$ and $\left\{Y = y_j\right\}$ are independent events.

Theorem 26. *The events A and B are independent iff the indicator random variables $\mathbf{1}_A$ and $\not\Vdash_B$ are independent random variables.*

Theorem 27. *Let X and Y be two independent random variables. Then for all subsets $C, D \subseteq \mathbb{R}$ we have*

$$\mathbb{P}\left(\left\{\omega \in \Omega : X(\omega) \in C, Y(\omega \in D)\right\}\right) = \mathbb{P}\left(\left\{\omega \in \Omega : X(\omega) \in C\right\}\right) \times \mathbb{P}\left(\left\{\omega \in \Omega : Y(\omega \in D)\right\}\right)$$

4.3 Functions of random variables

Definition 31. Let $X : \Omega \to \mathbb{R}$ be a random variable and $f : \mathbb{R} \to \mathbb{R}$ be a function. The random variable $f(X)$ is a function from Ω to $\mathbb{R}$ defined in the following way:

$$\left(f(X)\right)(\omega) \coloneqq f\left(X(\omega)\right)$$

Theorem 28. *Let $(\Omega, \mathbb{P})$ be a probability space and X and Y be two variables. Let $f, g : \mathbb{R} \to \mathbb{R}$ be two functions. We will define two random variables Z and W by*

$$\begin{aligned} \forall \omega \in \Omega \quad Z(\omega) &\coloneqq f\left(X(\omega)\right) \\ \forall \omega \in \Omega \quad W(\omega) &\coloneqq g\left(Y(\omega)\right) \end{aligned}$$

If the random variables X and Y are independent then so are Z and W.

Definition 32. Let $X_1, X_2, \ldots, X_n : \Omega \to \mathbb{R}$ be random variables and let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. The random variable $f(X_1, X_2, \ldots, X_n)$ is a function from Ω to $\mathbb{R}$ which is defined in the following way:

$$\left(f(X_1, X_2, \ldots, X_n)\right)(\omega) \coloneqq f\left(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)\right)$$

5 Expectation and variance of random variables

Definition 33. Let X be a random variable which is defined on a finite probability space $(\Omega, \mathbb{P})$. The **expectation** of X is defined by:

$$\mathbb{E}[X] \coloneqq \sum_{\omega \in \Omega} \mathbb{P}(\omega) X(\omega)$$

Definition 34. Let X be a random variable defined on a countable probability space $(\Omega, \mathbb{P})$. If the series $\sum_{\omega \in \Omega} \mathbb{P}(\omega) X(\omega)$ is an absolutely convergent series then the **expectation** of X is defined by:

$$\mathbb{E}[X] \coloneqq \sum_{\omega \in \Omega} \mathbb{P}(\omega) X(\omega)$$

5.1 Expectation

Theorem 29. *Let X be a random variable. For every real number a we will define*

$$f(a) := \mathbb{E}\left[(x-a)^2\right]$$

Then the number $a \in \mathbb{R}$ for which f has a minimum is $a = \mathbb{E}[X]$

Theorem 30. *Let $a \in \mathbb{R}$ and X be a constant random variable that is equal to a - $\forall \omega \in \Omega, X(\omega) = a$ then the random variable X has an expectation and*

$$\mathbb{E}[X] = a$$

Theorem 31. *Let X be random variable that has an expectation which is defined over $(\Omega, \mathbb{P})$ and let $a \in \mathbb{R}$. We will define a random variable Z by*

$$Z(\omega) := aX(\omega)$$

Then Z has an expectation too and

$$\mathbb{E}[Z] = a\mathbb{E}[X]$$

Theorem 32. *Let X be a random variable which attains the values $(x_i)_{i=1}^n$. If the expectation of X exists then*

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i \mathbb{P}(X = x_i)$$

Theorem 33. *Let X and Y be two random variables with an expectation defined over $(\Omega, \mathbb{P})$. We will define a random variable Z by*

$$Z(\omega) := X(\omega) + Y(\omega)$$

Then Z has an expectation and

$$\mathbb{E}[Z] = \mathbb{E}[X] + \mathbb{E}[Y]$$

Theorem 34. *Let $X_1, X_2, \ldots, X_n$ be random variables defined on the same probability space $(\Omega, \mathbb{P})$. Then*

$$\mathbb{E}\left[\sum_{i=1}^{n} X_i\right] = \sum_{i=1}^{n} \mathbb{E}[X_i]$$

Theorem 35. *Let X and Y be two random variables with expectations. We will define the random variable Z by*

$$Z(\omega) := X(\omega) \times Y(\omega)$$

If X and Y are independent, then Z has an expectation for which

$$\mathbb{E}[Z] = \mathbb{E}[X] \times \mathbb{E}[Y]$$

5.2 Covariance

Definition 35. Let X and Y be two random variables with expectations for which XY also has an expectation. The **covariance** of X and Y is defined by

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

Theorem 36. *Let X be a random variable. If the random variable X^2 has an expectation then also the random variable X has an expectation.*

Definition 36. Let X be a random variable and n is a natural number. If $\mathbb{E}[X^n]$ is defined we will call it the **moment of order** n of X and we will say that X has a moment of order n.

Theorem 37. *Let X be a random variable and n is a natural number. If the Moment of order $n+1$ exists then so does the moment of order n*

Theorem 38. *Let X and Y be two variables with a second moment, then the random variable XY has an expectation.*

Definition 37. A pair of random variables are called **uncorrelated** if $\mathrm{Cov}(X,Y)=0$.

Theorem 39. *Let X and Y be two random variables for which the random variable XY has an expectation. We will mark the expectation of these random variables with $\mu_X \coloneqq \mathbb{E}[X]$ and $\mu_Y = \mathbb{E}[Y]$. Then:*

$$\mathrm{Cov}(X,Y) = \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]$$

Theorem 40. *Let X and Y be two random variables for which the random variable XY has an expectation. Then $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$.*

Theorem 41. *Let X and Y be two random variables which have an expectation and let a and b be two real numbers. We will define the random variable Z by $Z = aX + b$ as in $Z(\omega) = aX(\omega) + b$ for all $\omega \in \Omega$. If the random variable XY has an expectation then so does ZY and*

$$\mathrm{Cov}(Z,Y) = a\mathrm{Cov}(X,Y)$$

Theorem 42. *Let $(X_i)_{i=1}^n$ and Y be random variables with expectations such that for all $i \in \{1,2,\dots,n\}$ the random variable X_iY has an expectation. Then:*

$$\mathrm{Cov}\left(\sum_{i=1}^n X_i, Y\right) = \sum_{i=1}^n \mathrm{Cov}(X_i, Y)$$

Corollary 1. *Let $(X_i)_{i=1}^n$ and $\left(Y_j\right)_{j=1}^m$ be random variables such that for all $i \in \{1,2,\dots,n\}$ and for all $j \in \{1,2,\dots,m\}$ the random variable X_iY_j has an expectation. Then:*

$$\mathrm{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \mathrm{Cov}\left(X_i, Y_j\right)$$

5.3 Variance

Definition 38. Let X be a random variable with a second moment. The **variance** of X is the size:

$$\mathrm{Var}(X) \coloneqq \mathrm{Cov}(X,X) = \mathbb{E}\left[X^2\right] - \left(\mathbb{E}[X]\right)^2$$

Theorem 43. *Let X be a random variable with a second moment. We will mark the expectation of this random variable with $\mu_X \coloneqq \mathbb{E}[X]$. Then:*

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

Theorem 44. *The following occur for any random variable with a second moment:*

1. $\mathrm{Var}(X) \geq 0$
2. $\mathrm{Var}(X) = 0$ *iff* $\mathbb{P}\left(X = \mathbb{E}[X]\right) = 1$*, if and only if the random variable is constant.*

Theorem 45. *Let X be a random variable with a second moment and $a \in \mathbb{R}$. We will define a random variable Y with $Y(\omega) = aX(\omega)$ for all $\omega \in \Omega$. Then the random variable Y has a second moment and*

$$\mathrm{Var}(Y) = a^2\mathrm{Var}(X)$$

Theorem 46. *Let X and Y be two random variables for which the random variable $|X| + |Y|$ has a second moment. Then:*

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) + 2 \cdot \operatorname{Cov}(X,Y)$$
$$\operatorname{Var}(X-Y) = \operatorname{Var}(X) + \operatorname{Var}(Y) - 2 \cdot \operatorname{Cov}(X,Y)$$

Corollary 2. *Let X and Y be two random variables with independent variances. Then:*

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + \operatorname{Var}(Y)$$

Theorem 47. *Let $X_1, X_2, \ldots, X_n$ be independent random variables, then*

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

If the random variables are pairwise independent then

$$\operatorname{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \operatorname{Var}(X_i)$$

Definition 39. Let X be a random variable with a second moment. The **standard deviation** of the random variable X is defined by

$$\sigma_X := \sqrt{\operatorname{Var}(X)}$$

Theorem 48. *Let $X : \Omega \to \mathbb{R}$ be a random variable with a second moment. Then:*

1. *The standard deviation is non-negative: $\sigma_X \ge 0$*
2. *$\sigma_{aX} = |a| \sigma_X$ for all $a \in \mathbb{R}$*
3. *$\sigma_{X+b} = \sigma_X$ for all $b \in \mathbb{R}$*

6 Correlation coefficient and inequalities

Definition 40. Let $X, Y : \Omega \to \mathbb{R}$ be two random variables which aren't constant. The **correlation coefficient** between X and Y, marked with $\rho(X,Y)$ is the number defined by

$$\rho(X,Y) := \frac{\operatorname{Cov}(X,Y)}{\sigma_x \sigma_y}$$

Theorem 49. *For any pair of random variables where neither are constant, the following occur:*

1. $-1 \le \rho(X,Y) \le 1$
2. *$\rho(X,Y) = 1$ iff there exist $a > 0$and $b \in \mathbb{R}$ such that $Y = aX + b$.*
3. *$\rho(X,Y) = 1$ iff there exist $a < 0$and $b \in \mathbb{R}$ such that $Y = aX + b$.*

6.1 Chebyshev's and Markov's inequalities

Theorem 50 (Markov Inequality)**.** *Let X be a non-negative random variable which is defined on the probability space $(\Omega, \mathbb{P})$. Then for all $c > 0$ we have*

$$\mathbb{P}(X \ge c) \le \frac{\mathbb{E}[X]}{c}$$

Theorem 51 (Chebyshev's inequality)**.** *Let X be a random variable with a second moment. For all $c > 0$ we have*

$$\mathbb{P}\left(|X - \mu_X| \ge c\right) \le \frac{\operatorname{Var}(X)}{c^2}$$

Theorem 52 (One tailed Chebyshev's inequality)**.** *Let X be a random variable with a second moment. For all $c > 0$ we have*

$$\mathbb{P}(X - \mu_X \ge c) \le \frac{\operatorname{Var}(X)}{\operatorname{Var}(X) + c^2}$$

7 Conditional expectation and variance

7.1 Conditional expectation

Theorem 53. *Let B be an event with positive probability and X a random variable with an expected value. The series $\sum_{\omega\in\Omega}\mathbb{P}\left(\omega|B\right)X\left(\omega\right)$ absolutely converges.*

Definition 41. Let B be an event with positive probability and X a random variable with an expected value. The **conditional expectation** of the random variable given B is

$$\mathbb{E}\left[X|B\right]=\sum_{\omega\in\Omega}\mathbb{P}\left(\omega|B\right)X\left(\omega\right)=\sum_{\omega\in\Omega}\frac{\mathbb{P}\left(\omega\right)}{\mathbb{P}\left(B\right)}X\left(\omega\right)$$

Theorem 54. *Let $X_1,X_2,\dots,X_n$ be random variables with expected values which are defined on the same probability space $(\Omega,\mathbb{P})$ and Let B be an event with positive probability. Then:*

$$\mathbb{E}\left[\sum_{i=1}^{n}X_i\middle|B\right]=\sum_{i=1}^{n}\mathbb{E}\left[X_i|B\right]$$

Theorem 55. *Let B be an event with positive probability and X is a random variable with positive probability.*

$$\mathbb{E}\left[X|B\right]=\sum_{i=1}^{n}x_i\mathbb{P}\left(X=x_i|B\right)$$

Definition 42. Let X be a random variable with an expectation and Y is a random variable. We will define the random variable $Z:\Omega\to\mathbb{R}$ as follows:

$$Z\left(\omega\right):=\begin{cases}\mathbb{E}\left[X|Y=Y\left(\omega\right)\right] & \mathbb{P}\left(Y=Y\left(\omega\right)\right)>0\\ 0 & \mathbb{P}\left(Y=Y\left(\omega\right)\right)=0\end{cases}$$

The random variable Z is called the **conditional probability of X given Y** and it is marked with $\mathbb{E}\left[X|Y\right]$

Theorem 56. *Let $X_1,X_2,\dots,X_n$ be random variables with expected values and Y be a random variable. Then:*

$$\mathbb{E}\left[\sum_{i=1}^{n}X_i|Y\right]=\sum_{i=1}^{n}\mathbb{E}\left[X_i|Y\right]$$

Theorem 57. *Let X and Y be two independent random variables. If X has an expected value then for all $\omega\in\Omega$ such that $\mathbb{P}\left(Y=Y\left(\omega\right)\right)>0$ then $\mathbb{E}\left[X|Y\right]\left(\omega\right)=\mathbb{E}\left[X\right]$*

Theorem 58 (Law of total expectation)**.** *Let X be a random variable with an expectation and Y be a random variable. The random variable $\mathbb{E}\left[X|Y\right]$ has an expected value and it is equal to*

$$\mathbb{E}\left[\mathbb{E}\left[X|Y\right]\right]=\mathbb{E}\left[X\right]$$

7.2 Conditional variance

Definition 43. Let X be a random variable with a second moment and let Y be a random variable. The **conditional variance** of X given Y is the random variable $\mathrm{Var}\left(X|Y\right)$ which matches for every $\omega\in\Omega$ the variance of X according to the distribution $\mathbb{P}\left(\cdot|Y=Y\left(\omega\right)\right)=Y\left(\omega\right)$ which means that:

$$\mathrm{Var}\left(X|Y\right)\left(\omega\right):=\begin{cases}\mathbb{E}\left[X^2|Y=Y\left(\omega\right)\right]-\left(\mathbb{E}\left[X|Y=Y\left(\omega\right)\right]\right)^2 & P\left(Y=Y\left(\omega\right)\right)>0\\ 0 & P\left(Y=Y\left(\omega\right)\right)=0\end{cases}$$

Theorem 59. *Let X, Y and Z be three independent variables. If X and Y have expected values then*

$$\mathbb{E}\left[XY|Z\right] = \mathbb{E}\left[X|Z\right]\mathbb{E}\left[Y|Z\right]|$$

Theorem 60. *Let $X_1, X_2, \ldots, X_n, Y$ be independent random variables. If the $(X_i)_{i=1}^{n}$ has a second moment then*

$$\text{Var}\left(\sum_{i=1}^{n} X_i|Y\right) = \sum_{i=1}^{n}\mathbb{E}\left[(X_i)^2|Y\right] - \sum_{i=1}^{n}\left(\mathbb{E}\left[X_i|Y\right]\right)^2$$

Theorem 61 (Law of total variance)**.** *Let X be a random variable with a second moment and Y be a random variable.*

$$\text{Var}\left(X\right) = \mathbb{E}\left[\text{Var}\left(X|Y\right)\right] + \text{Var}\left(\mathbb{E}\left[X|Y\right]\right)$$

8 Limit laws

Theorem 62 (The weak law of large numbers)**.** *Let X be a random variable with a second moment with an expected value of μ_X and let $\varepsilon > 0$. There exists an $N \in \mathbb{N}$ such that for all $n \geq N$ and for all n independent random variables $X_1, X_2, \ldots, X_n$ with the same distribution as X we have*

$$\mathbb{P}\left(\left|\frac{\sum_{i=1}^{n} X_i}{n} - \mu_X\right| \geq \varepsilon\right) \leq \varepsilon$$

Theorem 63 (The strong law of large numbers)**.** *Let X be a random variable with a fourth moment and an expectation μ_X and let $\varepsilon > 0$.Then there exists an $N \in \mathbb{N}$ such that for all $n \geq N$ and for all n random variables $X_1, X_2, \ldots, X_n$ with the same distributions as X then:*

$$\forall k \in \{N, N+1, \ldots, n\}.\mathbb{P}\left(\left|\frac{\sum_{i=1}^{n} X_i}{k} - \mu_X\right| \geq \varepsilon\right) \leq \varepsilon$$

Theorem 64 (Central limit theorem)**.** *Let $X_1, X_2, \ldots, X_3$ with the expectation μ and the variance σ^2. We will mark $S_n = \sum_{i=1}^{n} X_i$. Then:*

$$\mathbb{P}\left(\frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq b\right) \sim \varphi(b)$$

and in particular

$$\mathbb{P}\left(a \leq \frac{S_n - n\mu}{\sqrt{n\sigma^2}} \leq b\right) \sim \varphi(b) - \varphi(a)$$

where

$$\varphi(z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-t^2}{2}\right) dt$$

9 Markov Chains

In order to define Markov chins, we first need to define three mathematical objects:

1. A finite set of states S - is a set that contains all possible states which we can reach.
2. A matrix $\mathbf{P}$ of probabilities between the states of S - for all $x, y \in S$ let $\mathbf{P}(x, y) = \mathbf{P}(x \to y)$ be the probability of going from the state y to the state x in 1 step. The condition for the transition matrix:

$$\sum_{y \in S} \mathbf{P}(x, y) = 1 \qquad \forall x \in S$$

3. The starting distribution μ on S - therefore $\mu : S \to [0,1]$ is a function on S such that

$$\sum_{x \in S} \mu(x) = 1$$

Definition 44. The Markov chain until the time n is the collection of random variables $X_0, X_1, \ldots, X_n$ which get values in S such that

$$\mathbf{P}(X_0 = 0) = \mu(x) \qquad \forall x \in S$$

And for any subset $x_0, x_1, \ldots, x_n \in S$ we have

$$\mathbf{P}(X_0 = x_0, \ldots, x_n = x_n) = \mu(x_0)\mathbf{P}(x_0, x_1) \cdots \mathbf{P}(x_{n-1}, x_n)$$

9.1 Basic properties

Definition 45. An event A is dependent only on $X_0, \ldots, X_t$ is there is a function $f : S^t \to \{0,1\}$ such that

$$\mathbf{1}_A = f(X_0, \ldots, X_t)$$

Note 3. We will use $\mathbf{P}_\mu$ for probabilities where the starting probability is μ in order to emphasize this distribution. If $\mu(x) = 1$ for some state $x \in S$, then we will mark the probability with $\mathbf{P}_x$.

Lemma 3 (Markov property)**.** *For all $t \geq 0$ and events A, B such that A is determined by $X_0, \ldots, X_{t-1}$ (past), $A \subseteq \{X_t = x\}$ (present) and B is determined by $X_{t+1}, \ldots, X_n$ (future) then $\mathbf{P}_\mu(B|A) = \mathbf{P}(B')$ where $\mathbf{1}_B = f(X_{t+1}, \ldots, X_n)$ then $\mathbf{1}_{B'} = f(X_{t+1}, \ldots, X_{n-t})$*

Lemma 4 (Advancing distributions)**.** *We will mark the distribution of X_t with μ_t e.g. $\mu_t(x) = \mathbf{P}_\mu(X_t = x)$ for all $x \in S$, then*

$$\mu_t = \mu \cdot \mathbf{P}^t$$

Definition 46 (Stationary distribution)**.** A distribution π on S is called **stationary** if $\pi = \pi \cdot \mathbf{P}$.

Theorem 65 (Existence of stationary distributions)**.** *For all Markov chains there exists a stationary distribution. Which means that there exists a π on S such that*

$$\pi = \pi \cdot \mathbf{P}$$

9.2 Reducibility and periodicity

Definition 47 (Irreducible chains)**.** A Markov chain is called irreducible if for all pairs of states $x, y \in S$ there exists a stage t such that $\mathbf{P}^t(x, y) > 0$, or in other words $\mathbf{P}_x(X_t = y) > 0$.

Lemma 5. *If the chain is irreducible, then the period of all of the states is called the* ***period of the chain****.*

Note 4. If for all $x \in S$ the period is 1 then the chain will be called **aperiodic**, if not then it is called **periodic**.

Note 5. The fact that there exists a stationary distribution for a transformation matrix does not mean that the distribution of a stage t, μ_t converges to π. For example: if the chain is irreducible and periodic, then the distribution μ_t does not converge.

Theorem 66. *If we have a Markov chain with a starting distribution μ and a time $t \in \mathbb{N}$. We will define a function π_t on S in the following way:*

$$\pi_t(X) = \frac{1}{t}\mathbb{E}_\mu[\#\ \textit{of visits to } x \textit{ by stage } t-1]$$

Then

$$\pi_t \xrightarrow{t \to \infty} \pi$$

and π is the stationary distribution of the Markov chain.

Theorem 67 (Ergodic theorem of Markov chains)**.** *Given a irreducible aperiodic Markov chain with a stationary distribution* π*. There exist constants* $C > 0$ *and* $0 < \alpha < 1$ *such that for all starting distributions* μ *and consequently for any state* $x \in S$

$$\left|\mathbf{P}_{\mu}\left(X_t = x\right) - \pi\left(x\right)\right| < C\alpha^t$$

Definition 48. A Markov chain is called **bi-stochiastic** if every column sums to 1 (e.g. $\sum_i \mathbf{P}\left(i, j\right) = 1$ for all $j \in S$)

Claim 1. If a chain is irreducible with a period of p then one can divide the state space to p disjoint sets.

10 Random variable properties

Variable	Distribution	Expected Value	Variance	Marking
Geometric	$(1-p)^{k-1}p$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$X \sim G\left(p\right)$
Uniform	$\frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n(n+2)}{12}$	$X \sim U\left(n\right)$
Negative Binomial	$\binom{k-1}{r-1}p^r\left(1-p\right)k-r$	$\frac{r}{p}$	$\frac{r(1-p)}{p^2}$	$X \sim NB\left(r, p\right)$
Binomial	$\binom{n}{k}p^k \cdot (1-p)^{n-k}$	np	$np\left(1-p\right)$	$X \sim B\left(n, p\right)$
Hypergeometric	$\frac{\binom{D}{k}\binom{N-D}{n-k}}{\binom{N}{n}}$	$n\frac{D}{N}$	$n\frac{D}{N}\left(1-\frac{D}{n}\right)\frac{N-n}{N-1}$	$X \sim H\left(N, D, n\right)$
Poisson	$e^{-\lambda}\frac{\lambda^k}{k!}$	λ	λ	$X \sim P\left(\lambda\right)$

- Geometric - A special case of a negative binomial variable where $n = 1$.
- Uniform - When choosing a random number between a and b, $x \in \{a, a+1, \ldots, b-1, b\}$.
- Negative binomial - The number of experiments needed to reach the n-th (inclusive) success where the experiments are independent and the chance of success is p.
- Binomial - The odds of succeeding on n independent experiments, each with a chance of succeeding p.
- Hypergeometric - The number of "special" items in a sample of size n from a population of size N with D special items.
- Poisson - The number of successes in a binomial process where $p \to 0$ and $n \to \infty$ however $\lambda = np$ is constant. This is used for cases where the chance of success is small but the population is very large.