

# Stochastics

Bachelor Applied Artificial Intelligence (AAI-B3)

André Herzwurm

TH Rosenheim, WiSe 2022/23

## Literature

H.-O. Georgii, *Stochastics*, DeGruyter, 2nd Edition, 2013.

A. Steland, *Basiswissen Statistik*, Springer, 4. Auflage, 2016. (German)

S.M. Ross, *Introduction to Probability and Statistics for Engineers and Scientists*, Academic Press, 2014.

# Contents

<b>I</b>	<b>Introduction</b>	<b>1</b>
<b>II</b>	<b>Probability Theory – Discrete Case</b>	<b>2</b>
1	Discrete Probability Spaces . . . . .	2
2	Probability Mass Functions . . . . .	4
3	Conditional Probability and Independence . . . . .	6
4	Random Variables . . . . .	8
5	Empirical Distribution . . . . .	12
6	Special Discrete Distributions . . . . .	14

# Chapter I

## Introduction

We consider births in Germany. Question (two variants):

- (i) What is the probability of having a girl?
- (ii) Is it more likely to have a boy than a girl?

Empirical data from statistical sample yields the gender of  $N$  births:

$k$  girls (1),  $N - k$  boys (0).

The relative frequencies are given by

$$\hat{p}(1) = \frac{k}{N}, \quad \hat{p}(0) = \frac{N - k}{N} = 1 - \hat{p}(1)$$

and serve as a metric of the data (descriptive statistics).

Naive answers to the above questions read:

- (i) The sought probability is given by  $\hat{p}(1)$ .
- (ii) Yes if and only if  $\hat{p}(0) > \hat{p}(1)$ .

Criticism:

- The answers are based on only one sample or data set.
- The sample size and the variability of the data are not taken into account.

Therefore we will consider

**inferential statistics** (see Chapters ?? and ??)

to infer properties of an underlying probability distribution. This, in turn, requires a

**mathematical model** (see Chapters II and ??)

of the underlying random mechanism.

# Chapter II

## Probability Theory – Discrete Case

In this section we aim at modelling and analyzing random experiments with at most countably infinite outcomes (discrete case).

In the sequel let  $\Omega$  be a finite or a countably infinite set. The set  $\Omega$  is a model for the possible results of a random experiment.

### 1 Discrete Probability Spaces

**Definition 1.** The set  $\Omega$  is called *sample space*. Its elements  $\omega \in \Omega$  are called *outcomes*. Any subset  $A \subseteq \Omega$  of  $\Omega$  is called an *event*.

**Example 2.** Consider a fair coin with head (1) and tail (0) that is tossed twice. The corresponding sample space is given by

$$\Omega = \{(1, 1), (1, 0), (0, 1), (0, 0)\} = \{1, 0\}^2.$$

The event  $A = \{(1, 1), (1, 0)\}$  contains all outcomes with 1 in the first toss.

Clearly, we may also use this sample space as a model for two-stage processes having two outcomes (e.g., yes/no) per stage, e.g., a two-stage production process where in each stage some tolerance level is satisfied or not.

*Terminology:* We say that the event  $A \subseteq \Omega$  occurs if  $\omega \in A$ .

*Notation:* For sets  $A, B$  we write  $A \subseteq B$  if every element of  $A$  is also an element of  $B$ , and we write  $A = B$  if  $A \subseteq B$  and  $B \subseteq A$ . The empty set containing no elements is denoted by  $\emptyset$ . The cardinality of a set  $A$  is denoted by  $|A|$ .

**Definition 3.** Let  $A, B \subseteq \Omega$  be events. The *union*, the *intersection*, and the *difference* of  $A$  and  $B$  as well as the *complement* of  $A$  are defined by

$$\begin{aligned} A \cup B &= \{\omega \in \Omega: \omega \in A \text{ or } \omega \in B\}, \\ A \cap B &= \{\omega \in \Omega: \omega \in A \text{ and } \omega \in B\}, \\ A \setminus B &= \{\omega \in \Omega: \omega \in A \text{ and } \omega \notin B\}, \\ A^c &= \Omega \setminus A = \{\omega \in \Omega: \omega \notin A\}, \end{aligned}$$

respectively.

**Remark 4.** For events  $A, B, C \subseteq \Omega$  we have

$$\begin{aligned} A \cap (B \cup C) &= (A \cap B) \cup (A \cap C), \\ A \cup (B \cap C) &= (A \cup B) \cap (A \cup C), \\ (A \cap B)^c &= A^c \cup B^c, \\ (A \cup B)^c &= A^c \cap B^c, \\ A \setminus B &= A \cap B^c. \end{aligned}$$

**Definition 5.** The (countable) union of events  $A_1, A_2, A_3, \dots \subseteq \Omega$  is defined by

$$\bigcup_{i=1}^{\infty} A_i = \{\omega \in \Omega : (\exists i \in \mathbb{N} : \omega \in A_i)\}.$$

The events  $A_1, A_2, A_3, \dots \subseteq \Omega$  are *pairwise disjoint* if  $A_i \cap A_j = \emptyset$  for all  $i, j \in \mathbb{N}$  with  $i \neq j$ .

**Remark 6.** Clearly, we can express the union of finitely many events as a countable union, i.e., for  $n \in \mathbb{N}$  and events  $A_1, \dots, A_n \subseteq \Omega$  we have

$$\bigcup_{i=1}^n A_i = A_1 \cup \dots \cup A_n = \bigcup_{i=1}^{\infty} A_i$$

if we put  $A_i = \emptyset$  for  $i > n$ .

We are now going to assign probabilities  $P(A)$  to events  $A \subseteq \Omega$ .

**Definition 7.** The set

$$\mathcal{P}(\Omega) = \{A : A \subseteq \Omega\}$$

of all subsets of  $\Omega$  is called *power set* of  $\Omega$ .

**Example 8.** The power set of  $\Omega = \{0, 1\}$  is given by

$$\mathcal{P}(\Omega) = \{\emptyset, \{0\}, \{1\}, \Omega\}.$$

Moreover, we have  $|\mathcal{P}(\Omega)| = 2^{|\Omega|}$ .

**Definition 9.** A function  $P : \mathcal{P}(\Omega) \rightarrow \mathbb{R}$  is a *probability measure (on  $\Omega$ )* (or *probability distribution (on  $\Omega$ )*) if

- (i)  $0 \leq P(A) \leq 1$  for all  $A \subseteq \Omega$ ,
- (ii)  $P(\Omega) = 1$ ,
- (iii) for all pairwise disjoint events  $A_1, A_2, \dots \subseteq \Omega$  we have

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i). \quad (\sigma\text{-additivity})$$

In such a case the triple  $(\Omega, \mathcal{P}(\Omega), P)$  is called a *discrete probability space*.

**Proposition 10.** Let  $P$  be a probability measure on  $\Omega$  and let  $A, B \in \mathcal{P}(\Omega)$ . Then we have

- (i)  $A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$  (*additivity*),
- (ii)  $A \subseteq B \Rightarrow P(B \setminus A) = P(B) - P(A)$ ,
- (iii)  $A \subseteq B \Rightarrow P(A) \leq P(B)$  (*monotonicity*),
- (iv)  $P(A^c) = 1 - P(A)$ ,
- (v)  $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ .

*Proof.* ad (i): Since  $A \cap B = \emptyset$ , we have

$$A \cup B = A \cup B \cup \emptyset \cup \emptyset \cup \dots$$

for pairwise disjoint events and thus the  $\sigma$ -additivity of  $P$  shows

$$P(A \cup B) = P(A) + P(B) + \sum_{i=3}^{\infty} P(\emptyset).$$

Since  $P(A \cup B) \in [0, 1]$ , we have  $P(\emptyset) = 0$ .

ad (ii)-(v): See **Exercise 1.1**. □

## 2 Probability Mass Functions

**Definition 1.** Let  $P$  be a probability measure on  $\Omega$ . The function  $p: \Omega \rightarrow \mathbb{R}$  defined by

$$p(\omega) = P(\{\omega\})$$

is called *probability mass function (associated to  $P$ )*.

**Proposition 2.**

- (i) For  $P$  and  $p$  according to Definition 1 and for every  $A \subseteq \Omega$  we have

$$P(A) = \sum_{\omega \in A} p(\omega).$$

In particular,  $P$  is uniquely determined by  $p$ .

- (ii) Every probability mass function  $p: \Omega \rightarrow \mathbb{R}$  satisfies

$$\forall \omega \in \Omega: 0 \leq p(\omega) \leq 1 \quad \wedge \quad \sum_{\omega \in \Omega} p(\omega) = 1. \quad (1)$$

- (iii) Every function  $p: \Omega \rightarrow \mathbb{R}$  satisfying (1) defines a probability measure  $P$  on  $\Omega$  by

$$P(A) = \sum_{\omega \in A} p(\omega)$$

for  $A \subseteq \Omega$ .

*Proof.* Since  $\Omega$  is countable, every  $A \subseteq \Omega$  is countable and can thus be expressed as a countable union  $A = \bigcup_{\omega \in A} \{\omega\}$  of pairwise disjoint sets.

ad (i): The  $\sigma$ -additivity of  $P$  yields for  $A \subseteq \Omega$

$$P(A) = P\left(\bigcup_{\omega \in A} \{\omega\}\right) = \sum_{\omega \in A} P(\{\omega\}) = \sum_{\omega \in A} p(\omega).$$

ad (ii): Use part (i) with  $A = \Omega$  and note that  $P(\Omega) = 1$ .

ad (iii): Verify the conditions of Definition 1.9. □

**Example 3.** Let  $\Omega$  be finite with  $|\Omega| = n \in \mathbb{N}$ . For  $\omega \in \Omega$  we put

$$p(\omega) = \frac{1}{n}.$$

Then  $p$  satisfies (1) and the associated probability measure  $P$  is given by

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

for every  $A \subseteq \Omega$ . Hence the calculation of probabilities is based on counting elements.

**Definition 4.** The probability measure  $P$  according to Example 3 is called the *discrete uniform distribution (on the finite set  $\Omega$ )*.

**Example 5.** Consider a fair coin that is tossed twice, cf. Example 1.2. We model this random experiment by using the discrete uniform distribution  $P$  on

$$\Omega = \{(1, 1), (1, 0), (0, 1), (0, 0)\}.$$

The events

$$\begin{aligned} A &= \{(1, 1), (1, 0)\} = \text{“1 in the first toss”}, \\ B &= \{(1, 1), (0, 1)\} = \text{“1 in the second toss”} \end{aligned}$$

satisfy

$$\begin{aligned} P(A) &= \frac{2}{4} = \frac{1}{2} = P(B), \\ P(A \cap B) &= P(\{(1, 1)\}) = \frac{1}{4} = P(A) \cdot P(B), \\ P(A \cup B) &= P(\{(1, 1), (1, 0), (0, 1)\}) = \frac{3}{4}. \end{aligned}$$

**Example 6.** Consider a two-stage production process where in each process a tolerance level is satisfied (1) or not (0). We model this by

$$\Omega = \{(1, 1), (1, 0), (0, 1), (0, 0)\}$$

and the (fictional) probability mass function

$\omega$	(1, 1)	(1, 0)	(0, 1)	(0, 0)
$p(\omega)$	0.8	0.09	0.01	0.1

The events

$$A = \{(1, 1), (1, 0)\} = \text{“tolerance level satisfied in first stage”},$$

$$B = \{(1, 1), (0, 1)\} = \text{“tolerance level satisfied in second stage”}$$

satisfy

$$P(A) = 0.89,$$

$$P(B) = 0.81,$$

$$P(A \cap B) = 0.8 \neq 0.7209 = P(A) \cdot P(B).$$

### 3 Conditional Probability and Independence

In the sequel let  $(\Omega, \mathcal{P}(\Omega), P)$  be a discrete probability space.

Question: How do we change the probability measure  $P$  if we know that a certain event  $B \subseteq \Omega$  has occurred.

**Definition 1.** For  $A, B \subseteq \Omega$  with  $P(B) > 0$  the *conditional probability of  $A$  given  $B$*  is defined by

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

**Example 2.** Let  $P$  be the discrete uniform distribution on a finite set  $\Omega$ . For  $A, B \subseteq \Omega$  with  $B \neq \emptyset$  we have  $P(B) > 0$  and

$$P(A|B) = \frac{|A \cap B|}{|\Omega|} \cdot \frac{|\Omega|}{|B|} = \frac{|A \cap B|}{|B|}.$$

In Example 2.5 (fair coin tossed twice “independently”) we thus have

$$P(A|B) = P(B|A) = \frac{1}{2}$$

for  $A = \{(1, 1), (1, 0)\}$  and  $B = \{(1, 1), (0, 1)\}$ .

**Example 3.** In Example 2.6 (two-stage production process) we have

$$P(A|B) = \frac{P(\{(1, 1)\})}{P(\{(1, 1), (0, 1)\})} = \frac{0.8}{0.81} \approx 0.9876$$

and

$$P(B|A) = \frac{P(\{(1, 1)\})}{P(\{(1, 1), (1, 0)\})} = \frac{0.8}{0.89} \approx 0.8988$$

for  $A = \{(1, 1), (1, 0)\}$  and  $B = \{(1, 1), (0, 1)\}$ .



**Remark 4.** Let  $p$  be the probability mass function associated to  $P$ , and let  $B \subseteq \Omega$  with  $P(B) > 0$ . For  $A \subseteq \Omega$  we have

$$P(A|B) = \frac{1}{P(B)} \cdot \sum_{\omega \in A \cap B} p(\omega) = \sum_{\omega \in A} q(\omega),$$

where

$$q(\omega) = \begin{cases} \frac{p(\omega)}{P(B)}, & \text{if } \omega \in B, \\ 0, & \text{else.} \end{cases}$$

Then  $Q: \mathcal{P}(\Omega) \rightarrow \mathbb{R}$  given by

$$Q(A) = P(A|B)$$

defines a probability measure on  $\Omega$  with probability mass function  $q$ .

**Definition 5.** Two events  $A, B \subseteq \Omega$  are *independent* if

$$P(A \cap B) = P(A) \cdot P(B).$$

**Remark 6.** Let  $P(B) > 0$ . Then we have

$$A, B \text{ independent} \Leftrightarrow P(A|B) = P(A).$$

**Example 7.** In Example 2.5 (fair coin tossed twice “independently”) the events  $A$  and  $B$  are independent. In Example 2.6 (two-stage production process) the events  $A$  and  $B$  are dependent.

**Proposition 8** (Bayes’s law). Let  $n \in \mathbb{N}$  and  $\{B_1, \dots, B_n\} \subseteq \mathcal{P}(\Omega)$  be a partition<sup>1</sup> of  $\Omega$  with  $P(B_i) > 0$  for all  $i = 1, \dots, n$ .

(i) For all  $A \subseteq \Omega$  we have

$$P(A) = \sum_{i=1}^n P(A|B_i) \cdot P(B_i).$$

(ii) For all  $A \subseteq \Omega$  with  $P(A) > 0$  and for all  $k = 1, \dots, n$  we have

$$P(B_k|A) = \frac{P(A|B_k) \cdot P(B_k)}{\sum_{i=1}^n P(A|B_i) \cdot P(B_i)}.$$

*Proof.* ad (i): Since  $\{B_1, \dots, B_n\}$  is a partition, we obtain for every  $A \subseteq \Omega$

$$A = A \cap \left( \bigcup_{i=1}^n B_i \right) = \bigcup_{i=1}^n (A \cap B_i)$$

with pairwise disjoint sets  $A \cap B_1, \dots, A \cap B_n \subseteq \Omega$ . Hence we get

$$P(A) = \sum_{i=1}^n P(A \cap B_i).$$

---

<sup>1</sup>A partition of  $\Omega$  is a family of non-empty pairwise disjoint sets  $B_1, B_2, \dots \subseteq \Omega$  with  $\bigcup_{i=1}^n B_i = \Omega$ .

Since  $P(B_i) > 0$  for all  $i = 1, \dots, n$ , the conditional probabilities  $P(A | B_i)$  are well-defined such that  $P(A \cap B_i) = P(A | B_i) \cdot P(B_i)$ .

ad (ii): If  $P(A) > 0$ , the conditional probability  $P(B_k | A)$  is well-defined and we have

$$P(B_k | A) \cdot P(A) = P(B_k \cap A) = P(A \cap B_k) = P(A | B_k) \cdot P(B_k)$$

for all  $k = 1, \dots, n$ . Use part (i). □

## 4 Random Variables

In the sequel let  $(\Omega, \mathcal{P}(\Omega), P)$  be a discrete probability space.

Question: How can we describe particular aspects of a random experiment (rather than the full experiment)?

**Example 1.** Consider a fair die that is rolled twice independently, i.e., we consider a discrete uniform distribution  $P$  on  $\Omega = \{1, \dots, 6\}^2$ . Particular aspects could be:

- (i) “number of pips in first die roll”,
- (ii) “number of pips in second die roll”,
- (iii) “sum of pips”.

In the sequel let  $\mathfrak{X} \subseteq \mathbb{R}$  be finite or countably infinite. We will typically consider  $\mathfrak{X} \subseteq \mathbb{N}_0$ .

**Definition 2.** A function  $X: \Omega \rightarrow \mathfrak{X}$  is called a *random variable (with values in  $\mathfrak{X}$ )*. Its function values  $x = X(\omega) \in \mathfrak{X}$  are called *realizations* of  $X$ .

**Example 3** (Continuation of Example 1). The first two aspects are described by  $\mathfrak{X} = \{1, \dots, 6\}$  and the random variables  $X_1, X_2: \Omega \rightarrow \mathfrak{X}$  given by

$$X_1(\omega) = \omega_1, \quad X_2(\omega) = \omega_2$$

for  $\omega = (\omega_1, \omega_2) \in \Omega$ . The third aspect is described by  $\mathfrak{S} = \{2, \dots, 12\}$  and the random variable  $S: \Omega \rightarrow \mathfrak{S}$  given by

$$S(\omega) = X_1(\omega) + X_2(\omega) = \omega_1 + \omega_2.$$

In the sequel let  $X: \Omega \rightarrow \mathfrak{X}$  be a random variable. In many cases one is just interested in the probabilities

$$P_X(A) = P(\{\omega \in \Omega: X(\omega) \in A\})$$

for  $A \subseteq \mathfrak{X}$  and in particular in

$$p_X(x) = P_X(\{x\}) = P(\{\omega \in \Omega: X(\omega) = x\})$$

for  $x \in \mathfrak{X}$ .

**Example 4** (Continuation of Example 3). For  $x \in \{1, \dots, 6\}$  we have

$$p_{X_1}(x) = p_{X_2}(x) = \frac{1}{6},$$

and  $p_S$  is given by

$s$	2	3	4	5	6	7	8	9	10	11	12
$p_S(s)$	$\frac{1}{36}$	$\frac{2}{36}$	$\frac{3}{36}$	$\frac{4}{36}$	$\frac{5}{36}$	$\frac{6}{36}$	$\frac{5}{36}$	$\frac{4}{36}$	$\frac{3}{36}$	$\frac{2}{36}$	$\frac{1}{36}$

**Proposition 5.** The following holds true:

- (i)  $p_X: \mathfrak{X} \rightarrow \mathbb{R}$  is a probability mass function,
- (ii)  $P_X$  is the associated probability measure on  $\mathfrak{X}$ , i.e., for all  $A \subseteq \mathfrak{X}$  we have

$$P_X(A) = \sum_{x \in A} p_X(x).$$

**Definition 6.**  $P_X$  and  $p_X$  are called *distribution* and *probability mass function* of  $X$ , respectively.

**Example 7** (Continuation of Example 4). A *rod graph* can be used to illustrate the probability mass function  $p_S$ , see Figure 4.1.

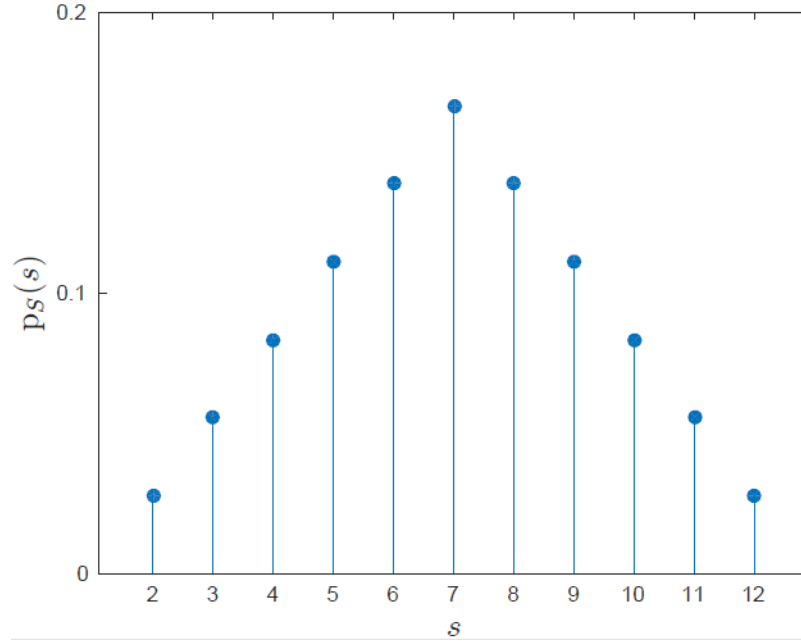


Figure 4.1:

**Definition 8.** Two random variables  $X_1, X_2: \Omega \rightarrow \mathfrak{X}$  are *identically distributed* if

$$P_{X_1}(A) = P_{X_2}(A)$$

for all  $A \subseteq \mathfrak{X}$ .

**Example 9** (Continuation of Example 4).  $P_{X_1}$  and  $P_{X_2}$  are both the discrete uniform distribution on  $\{1, \dots, 6\}$ . Hence  $X_1$  and  $X_2$  are identically distributed.

**Proposition 10** (Criterion for identical distributions).  $X_1, X_2: \Omega \rightarrow \mathfrak{X}$  are identically distributed if and only if

$$p_{X_1}(x) = p_{X_2}(x)$$

for all  $x \in \mathfrak{X}$ .

*Proof.* Use Proposition 5. □

**Remark 11.** The case of a multivariate random variable (random vector)

$$X = (X_1, \dots, X_n): \Omega \rightarrow \mathfrak{X}$$

with an at most countably infinite set  $\mathfrak{X} \subseteq \mathbb{R}^n$  is treated analogously. The components  $X_i$  of  $X$  are random variables.

**Example 12** (Continuation of Example 3).

(i) For  $\mathfrak{X} = \{1, \dots, 6\}^2$  and  $x = (x_1, x_2) \in \mathfrak{X}$  we have

$$p_{(X_1, X_2)}(x) = P(\{\omega \in \Omega: \omega = x\}) = P(\{x\}) = \frac{1}{36}.$$

Hence  $P_{(X_1, X_2)}$  is the discrete uniform distribution on  $\mathfrak{X}$ .

(ii) For  $\mathfrak{X} = \{(x_1, s) \in \mathbb{N}^2: 1 \leq x_1 \leq 6 \wedge x_1 + 1 \leq s \leq x_1 + 6\}$  and  $(x_1, s) \in \mathfrak{X}$  we have

$$p_{(X_1, S)}(x_1, s) = P(\{\omega \in \Omega: \omega_1 = x_1 \wedge \omega_1 + \omega_2 = s\}) = \frac{1}{36}.$$

Hence  $P_{(X_1, S)}$  is the discrete uniform distribution on  $\mathfrak{X}$ .

In the sequel we use

$$\begin{aligned} \{X = x\} &= \{\omega \in \Omega: X(\omega) = x\}, \\ \{X \in A\} &= \{\omega \in \Omega: X(\omega) \in A\}, \end{aligned}$$

and we consider random variables  $X_1, \dots, X_n$  on  $(\Omega, \mathcal{P}(\Omega), P)$  taking values in an at most countably infinite set  $\mathfrak{X}$ .

**Definition 13.**  $X_1, \dots, X_n$  are *independent* if for all  $A_1, \dots, A_n \subseteq \mathfrak{X}$  we have

$$P\left(\bigcap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n P(\{X_i \in A_i\}).$$

**Example 14** (Continuation of Example 3).

(i) For  $\mathfrak{X} = \{1, \dots, 6\}$  and  $A_1, A_2 \subseteq \mathfrak{X}$  we have

$$\begin{aligned} P(\{X_1 \in A_1\} \cap \{X_2 \in A_2\}) &= P(\{\omega \in \Omega: \omega_1 \in A_1 \wedge \omega_2 \in A_2\}) \\ &= \frac{|A_1| \cdot |A_2|}{|\Omega|} = \frac{|A_1| \cdot |A_2|}{36} \end{aligned}$$

and according to Example 9 for  $i = 1, 2$

$$P(\{X_i \in A_i\}) = \frac{|A_i|}{6}.$$

Hence  $X_1$  and  $X_2$  are independent.

(ii) For  $\mathfrak{X} = \{1, \dots, 12\}$ ,  $A_1 = \{6\}$  and  $B = \{2\}$  we have

$$\begin{aligned} P(\{X_1 \in A_1\} \cap \{S \in B\}) &= P(\{\omega \in \Omega: \omega_1 = 6 \wedge \omega_1 + \omega_2 = 2\}) \\ &= P(\emptyset) = 0 \end{aligned}$$

as well as  $P(\{X_1 \in A_1\}) > 0$  and  $P(\{S \in B\}) > 0$ . Hence  $X_1$  and  $S$  are not independent.

**Proposition 15** (Criterion for independence).  $X_1, \dots, X_n$  are independent if and only if for all  $x_1, \dots, x_n \in \mathfrak{X}$  we have

$$P\left(\bigcap_{i=1}^n \{X_i = x_i\}\right) = \prod_{i=1}^n p_{X_i}(x_i). \quad (2)$$

*Proof.* “ $\Rightarrow$ ”: Take  $A_i = \{x_i\}$ .

“ $\Leftarrow$ ”: Since  $\mathfrak{X}$  is countable, every  $A_i \subseteq \mathfrak{X}$  can be expressed as a countable union  $A_i = \bigcup_{x_i \in A_i} \{x_i\}$  of pairwise disjoint sets. Use the  $\sigma$ -additivity of  $P$  and (2). [...]  $\square$

**Remark 16.** Consider the special case of  $n = 2$  and  $\mathfrak{X} = \{0, \dots, k\}$  for some  $k \in \mathbb{N}$ . Put

$$p_{i,j} = P(\{X_1 = i\} \cap \{X_2 = j\})$$

and

$$p_{i,\bullet} = P(\{X_1 = i\}) \quad \text{and} \quad p_{\bullet,j} = P(\{X_2 = j\})$$

for  $i, j \in \{0, \dots, k\}$ . Clearly, we have

$$p_{i,\bullet} = \sum_{j=0}^k p_{i,j} \quad \text{and} \quad p_{\bullet,j} = \sum_{i=0}^k p_{i,j}$$

for all  $i, j \in \{0, \dots, k\}$ . The corresponding *contingency table* is given by Table II.1. Proposition 15 shows that

$$X_1, X_2 \text{ independent} \Leftrightarrow \forall i, j \in \{0, \dots, k\}: p_{i,j} = p_{i,\bullet} \cdot p_{\bullet,j}.$$

$X_1 \backslash X_2$	0	$\dots$	$k$	$\Sigma$
0	$p_{0,0}$	$\dots$	$p_{0,k}$	$p_{0,\bullet}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$k$	$p_{k,0}$	$\dots$	$p_{k,k}$	$p_{k,\bullet}$
$\Sigma$	$p_{\bullet,0}$	$\dots$	$p_{\bullet,k}$	1

Table II.1: Contingency table from Remark 16.

## 5 Empirical Distribution

In the sequel let  $(\Omega, \mathcal{P}(\Omega), P)$  be a discrete probability space and let  $X: \Omega \rightarrow \mathfrak{X}$  be a random variable with a countable set  $\mathfrak{X} \subseteq \mathbb{R}$ .

We consider the case where the set  $\mathfrak{X}$  of possible realizations is known but the distribution  $P_X$  is unknown. Instead, a *sample* (data set)

$$x_1, \dots, x_N \in \mathfrak{X} \quad (3)$$

consisting of realizations of  $N$  independent repetitions of the random experiments is available.

Question: Can we approximately determine the probability mass function  $p_X$  of  $X$ ?

**Definition 1.** Given the sample (3) the *relative frequency* of  $x \in \mathfrak{X}$  is given by

$$\begin{aligned} \hat{p}(x) &= \frac{|\{\ell \in \{1, \dots, N\} : x_\ell = x\}|}{N} \\ &= \frac{\text{number of elements of the sample equal to } x}{N}. \end{aligned}$$

**Remark 2.** For  $p = \hat{p}$  we have (1) in Proposition 2.2(ii). According to Proposition 2.2(iii), see also Proposition 4.5, we obtain a probability distribution  $\hat{P}$  on  $\mathfrak{X}$  satisfying

$$\begin{aligned} \hat{P}(A) &= \sum_{x \in A} \hat{p}(x) = \frac{|\{\ell \in \{1, \dots, N\} : x_\ell \in A\}|}{N} \\ &= \frac{\text{number of elements of the sample with values in } A}{N} \end{aligned}$$

for  $A \subseteq \mathfrak{X}$ .

**Definition 3.**  $\hat{P}$  is called *empirical distribution* of the sample (3).

**Remark 4.** A rod graph can be used to illustrate empirical distributions in terms of relative frequencies. Based on the scale of measure we distinguish the following types:

- (i) Nominal scale: Elements of  $\mathfrak{X}$  indicate a name (operations  $=, \neq$ ). Example: gender (male (0), female (1)).

$X_1 \backslash X_2$	0	$\dots$	$k$	$\Sigma$
0	$\hat{p}_{0,0}$	$\dots$	$\hat{p}_{0,k}$	$\hat{p}_{0,\bullet}$
$\vdots$	$\vdots$	$\ddots$	$\vdots$	$\vdots$
$k$	$\hat{p}_{k,0}$	$\dots$	$\hat{p}_{k,k}$	$\hat{p}_{k,\bullet}$
$\Sigma$	$\hat{p}_{\bullet,0}$	$\dots$	$\hat{p}_{\bullet,k}$	1

Table II.2: Contingency table for empirical distribution.

- (ii) Ordinal scale: Elements of  $\mathfrak{X}$  allow for a ranking (operations  $<, >, =$ ). Example: rank order.
- (iii) Metric (cardinal) scale: Elements of  $\mathfrak{X}$  are numeric and allow for arithmetic operations. Examples: time, temperature, sales.

**Remark 5.** The case of a multivariate random variable (random vector)

$$X = (X_1, \dots, X_n): \Omega \rightarrow \mathfrak{X}$$

with an at most countably infinite set  $\mathfrak{X} \subseteq \mathbb{R}^n$  is treated analogously.

Example (rod graph): Simulation of two independent fair coins, sample size  $N = 1000$ .

**Remark 6** (Counterpart of Remark 4.16). Consider the special case of two random variables  $(X_1, X_2): \Omega \rightarrow \mathfrak{X}$  and  $\mathfrak{X} = \{0, \dots, k\}^2$  with  $k \in \mathbb{N}$ . Put

$$\hat{p}_{i,j} = \frac{|\{\ell \in \{1, \dots, N\}: x_\ell = (i, j)\}|}{N}$$

and

$$\hat{p}_{i,\bullet} = \frac{|\{\ell \in \{1, \dots, N\}: x_{\ell,1} = i\}|}{N}, \quad \hat{p}_{\bullet,j} = \frac{|\{\ell \in \{1, \dots, N\}: x_{\ell,2} = j\}|}{N}$$

for  $i, j \in \{0, \dots, k\}$ . We clearly have

$$\hat{p}_{i,\bullet} = \sum_{j=0}^k \hat{p}_{i,j} \quad \text{and} \quad \hat{p}_{\bullet,j} = \sum_{i=0}^k \hat{p}_{i,j}$$

for all  $i, j \in \{0, \dots, k\}$ . The corresponding contingency table is given by Table II.2. The random variables  $X_1$  and  $X_2$  are assumed to be independent if and only if

$$\hat{p}_{i,j} \approx \hat{p}_{i,\bullet} \cdot \hat{p}_{\bullet,j}$$

for all  $i, j \in \{0, \dots, k\}$ .

Example (crosstab): Simulation of two independent fair coins, sample size  $N = 1000$ .

Example (crosstab): Simulation of two-stage production process, sample size  $N = 1000$ .

## 6 Special Discrete Distributions

In this section we discuss several discrete probability distributions that serve as standard models for special random experiments.

### Binomial distribution

Consider a random experiment with the two outcomes 1 (success) and 0 (failure) that is repeated  $n$  times independently. This can be modeled by

- (i) parameters  $n \in \mathbb{N}$  (number of repetitions) and  $p \in [0, 1]$  (probability of success),
- (ii) independent and identically distributed (*i.i.d.*) random variables  $X_1, \dots, X_n$  satisfying

$$p = P(\{X_i = 1\}) = 1 - P(\{X_i = 0\})$$

for all  $i = 1, \dots, n$ .

Put

$$S = \sum_{i=1}^n X_i$$

and note that for  $\omega \in \Omega$  we have

$$S(\omega) = |\{i \in \{1, \dots, n\} : X_i(\omega) = 1\}|,$$

i.e., the number of successes is given by  $S$ .

**Proposition 1.** For  $k \in \{0, \dots, n\}$  we have

$$P(\{S = k\}) = \binom{n}{k} \cdot p^k \cdot (1 - p)^{n-k}.$$

**Definition 2.** A random variable  $X$  is *binomially distributed* with parameters  $n \in \mathbb{N}$  und  $p \in [0, 1]$  if

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

for all  $k \in \{0, \dots, n\}$ .

*Notation:*  $X \sim B(n, p)$ .

Statistical problem: Given  $n$  and  $k$  (by a sample), estimate  $p$ .

**Example 3.** pmf of  $X \sim B(10, p)$  with  $p = 1/2$  and  $p = 1/4$ .

**Example 4.** pmf of  $X \sim B(100, p)$  with  $p = 1/2$  and  $p = 1/20$ .

**Proposition 5.** Let  $X$  and  $Y$  be independent with  $X \sim B(n, p)$  and  $Y \sim B(m, p)$  for  $m, n \in \mathbb{N}$  and  $p \in [0, 1]$ . Then we have

$$X + Y \sim B(n + m, p).$$



## Multinomial distribution

Consider a random experiment with outcomes  $0, \dots, m-1$  that is repeated  $n$  times independently. This can be modeled by

- (i) parameters  $n \in \mathbb{N}$  (number of repetitions),  $m \in \mathbb{N} \setminus \{1\}$  (number of outcomes), and  $p_0, \dots, p_{m-1} \in [0, 1]$  such that  $\sum_{j=0}^{m-1} p_j = 1$  (probabilities of  $m$  outcomes),
- (ii) i.i.d. random variables  $X_1, \dots, X_n$  satisfying

$$P(\{X_i = j\}) = p_j$$

for all  $i = 1, \dots, n$  and  $j = 0, \dots, m-1$ .

Put

$$S_j(\omega) = |\{i \in \{1, \dots, n\} : X_i(\omega) = j\}|$$

for  $\omega \in \Omega$  and  $j \in \{0, \dots, m-1\}$ , i.e.,  $S_j$  is the number of random experiments with outcome  $j$ , and put

$$S = (S_0, \dots, S_{m-1}).$$

**Remark 6.** (i) For  $j = 0, \dots, m-1$  we have  $S_j \sim B(n, p_j)$ .

(ii) The random variables  $S_0, \dots, S_{m-1}$  are not independent in general. We have

$$\sum_{j=0}^{m-1} S_j(\omega) = n$$

for all  $\omega \in \Omega$ .

**Proposition 7.** For  $k = (k_0, \dots, k_{m-1}) \in \mathbb{N}_0^m$  with  $\sum_{j=0}^{m-1} k_j = n$  we have

$$P(\{S = k\}) = \frac{n!}{k_0! \cdots k_{m-1}!} \cdot p_0^{k_0} \cdots p_{m-1}^{k_{m-1}}.$$

**Definition 8.** A random vector  $X$  follows a *multinomial distribution* with parameters  $n \in \mathbb{N}$  and  $p_0, \dots, p_{m-1} \in [0, 1]$  such that  $\sum_{j=0}^{m-1} p_j = 1$  if

$$P(\{X = k\}) = \frac{n!}{k_0! \cdots k_{m-1}!} \cdot p_0^{k_0} \cdots p_{m-1}^{k_{m-1}}$$

for all  $k = (k_0, \dots, k_{m-1}) \in \mathbb{N}_0^m$  with  $\sum_{j=0}^{m-1} k_j = n$ .

*Notation:*  $X \sim M(n, p_0, \dots, p_{m-1})$ .