École Polytechnique
Fédérale de Lausanne

Advanced Information, Computation and Communication II

Notes taken during Prof. Christoph Gastpar's lectures at EPFL in Spring 2026.

Elie Baier (397222)
Computer Science Bachelor
February 17, 2026
Version: `8715dfc`

Contents

1 Entropy and Data Compression **2**

1.1 Probability Review . . . 2

1.1.1 Definitions . . . 2

1.1.2 Conditional Probability and Independence . . . 3

1.1.3 Law of Total Probability and Bayes' Theorem . . . 4

1.1.4 Random Variables and Probability Distributions . . . 6

1.1.5 Expectation and Independence of Random Variables . . . 8

Chapter 1

Entropy and Data Compression

1.1 Probability Review

This section will review some basic concepts of probability theory that were covered in the previous course CS-101.

1.1.1 Definitions

Definition 1.1.1 (Sample Space).
A sample space is a set Ω of possible outcomes of a random experiment. Each outcome $\omega \in \Omega$ is called an elementary event.

Definition 1.1.2 (Event).
An event is a subset $E \subseteq \Omega$ of the sample space Ω. The probability of an event E is denoted:

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

Note that this definition of probability assumes that all outcomes in the sample space are equally likely, which is often the case in simple experiments.

Example. Consider the experiment of tossing a fair coin three times. The sample space is:

$$\Omega = \{HHH, HHT, HTH, HTT, THH, THT, TTH, TTT\}$$

Each outcome has a probability of $\frac{1}{8}$. An example of an event is:

$$E = \{\text{at least two heads}\} = \{HHH, HHT, HTH, THH\}$$

which has a probability of $P(E) = \frac{4}{8} = \frac{1}{2}$.

Definition 1.1.3 (Probability Mass Function).
A probability mass function (PMF) is a function $P : \Omega \to [0, 1]$ that assigns a probability to each outcome in the sample space, such that:

$$\sum_{\omega\in\Omega} P(\omega) = 1$$

Thus we can also define the probability of an event E as:

$$P(E) = \sum_{\omega\in E} P(\omega)$$

1.1.2 Conditional Probability and Independence

Definition 1.1.4 (Conditional Probability).
The conditional probability of an event E given an event F is defined as:

$$P(E|F) = \frac{P(E\cap F)}{P(F)}$$

provided that $P(F) > 0$.

Is it often useful to see F as the "new" sample space, and $E \cap F$ as the "new" event we are interested in. It can also be translated into english as: the probability of E given that F has occurred.

Definition 1.1.5 (Independence).
Two events E and F are said to be independent if:

$$P(E \cap F) = P(E)P(F)$$

Equivalently, if $P(F) > 0$, we can also say that E and F are independent if:

$$P(E|F) = P(E)$$

Example. Let's consider the experiment of tossing two fair coins. The sample space is:

$$\Omega = \{HH, HT, TH, TT\}$$

Let E be the event "the first coin is heads" and F be the event "the second coin is heads". We have:

$$P(E) = \frac{2}{4} = \frac{1}{2}, \quad P(F) = \frac{2}{4} = \frac{1}{2}$$

and

$$P(E \cap F) = P(\{HH\}) = \frac{1}{4}$$

Since $P(E)P(F) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, we see that $P(E \cap F) = P(E)P(F)$, so E and F are independent.

Example. Consider the experiment of rolling a fair six-sided die. Let E be the event "the outcome is even" and F be the event "the outcome is greater than 3". We have:

$$P(E) = \frac{3}{6} = \frac{1}{2}, \quad P(F) = \frac{3}{6} = \frac{1}{2}$$

and

$$P(E \cap F) = P(\{4, 6\}) = \frac{2}{6} = \frac{1}{3}$$

Since $P(E)P(F) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}$, we see that $P(E \cap F) \neq P(E)P(F)$, so E and F are not independent.

Definition 1.1.6 (Disjoint Events).
Two events E and F are said to be disjoint (or mutually exclusive) if:

$$E \cap F = \emptyset$$

In this case, we have:

$$P(E \cup F) = P(E) + P(F)$$

1.1.3 Law of Total Probability and Bayes' Theorem

Definition 1.1.7 (Partition of the Sample Space).
A collection of events $\{B_1, B_2, \ldots, B_n\}$ is called a partition of the sample space Ω if:

- $B_i \cap B_j = \emptyset$ for all $i \neq j$ (the events are mutually disjoint).
- $\bigcup_{i=1}^{n} B_i = \Omega$ (the events cover the entire sample space).

Theorem 1.1.1 (Law of Total Probability). Let $\{B_1, B_2, \ldots, B_n\}$ be a partition of the sample space Ω, meaning that $B_i \cap B_j = \emptyset$ for $i \neq j$ and $\bigcup_{i=1}^{n} B_i = \Omega$. Then for any event E, we have:

$$P(E) = \sum_{i=1}^{n} P(E|B_i)P(B_i)$$

Proof. Since $\{B_1, B_2, \ldots, B_n\}$ is a partition of Ω, we can express E as the union of its intersections with the B_i:

$$E = (E \cap B_1) \cup (E \cap B_2) \cup \cdots \cup (E \cap B_n)$$

Since the B_i are mutually disjoint, the sets $E \cap B_i$ are also mutually disjoint. Therefore, we can write:

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

Now, using the definition of conditional probability, we have:

$$P(E|B_i) = \frac{P(E \cap B_i)}{P(B_i)} \quad \text{for } P(B_i) > 0$$

Rearranging this gives:

$$P(E \cap B_i) = P(E|B_i)P(B_i)$$

Substituting this back into our expression for $P(E)$, we get:

$$P(E) = \sum_{i=1}^{n} P(E|B_i)P(B_i)$$

□

Example. Let's consider two factories that produce light bulbs.

- Factory A's bulbs work for over 5000 hours in 99% of the cases.
- Factory B's bulbs work for over 5000 hours in 95% of the cases.
- It is known that factory A produces 60% of the bulbs.

Let's compute the probability that a randomly chosen bulb works for over 5000 hours. We can use the law of total probability with the partition $\{B_A, B_B\}$, where B_A is the event "the bulb is produced by factory A" and B_B is the event "the bulb is produced by factory B". We have:

$$P(\text{works}) = P(\text{works}|B_A)P(B_A) + P(\text{works}|B_B)P(B_B)$$

Substituting the given probabilities, we get:

$$P(\text{works}) = (0.99)(0.60) + (0.95)(0.40) = 0.594 + 0.38 = 0.974$$

Sometimes we are interested in the reverse problem: given that a bulb works for over 5000 hours, what is the probability that it was produced by factory A? This is where Bayes' theorem comes into play.

Theorem 1.1.2 (Bayes' Theorem)**.** Let $F, E \in \Omega$ be two events such that $P(E) > 0$. Then we have:

$$P(F|E) = \frac{P(E|F)P(F)}{P(E)}$$

Proof. By the definition of conditional probability, we have:

$$P(F|E) = \frac{P(F \cap E)}{P(E)}$$

Now, we can express $P(F \cap E)$ using the definition of conditional probability again:

$$P(F \cap E) = P(E|F)P(F)$$

Substituting this back into our expression for $P(F|E)$, we get:

$$P(F|E) = \frac{P(E|F)P(F)}{P(E)}$$

□

Example. Using the same example of the light bulbs, we want to compute the probability that a randomly chosen bulb was produced by factory A given that it works for over 5000 hours. We can use Bayes' theorem with $F = B_A$ and $E = \text{works}$:

$$P(B_A|\text{works}) = \frac{P(\text{works}|B_A)P(B_A)}{P(\text{works})}$$

Substituting the values we computed earlier, we get:

$$P(B_A|\text{works}) = \frac{(0.99)(0.60)}{0.974} \approx 0.609$$

1.1.4 Random Variables and Probability Distributions

Definition 1.1.8 (Random Variable).
A random variable is a function $X : \Omega \to \mathbb{R}$ that assigns a real number to each outcome in the sample space.

Definition 1.1.9 (Probability Distribution).
The probability distribution $P_X(x)$ of a random variable X is the probability that X takes the value x i.e. the probability of the event:

$$E = \{\omega \in \Omega : X(\omega) = x\}$$

Hence:

$$P_X(x) = \sum_{\omega \in E} P(\omega)$$

Note that it is often written as:

$$P(X = x) = P_X(x)$$

Example. Let's consider the game of rolling a fair six-sided die and if the outcome is 6, you receive 10$ otherwise you pay 1$. We can define a random variable X that represents the amount of money you win (or lose) in this game. The sample space is:

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

and the random variable X is defined as:

$$X(\omega) = \begin{cases} -1 & \text{if } \omega \in \{1, 2, 3, 4, 5\} \\ 10 & \text{if } \omega = 6 \end{cases}$$

The probability distribution of X is:

$$P_X(-1) = P(X = -1) = \frac{5}{6}, \quad P_X(10) = P(X = 10) = \frac{1}{6}$$

Instead of computing the probability that X takes a specific value, we can compute, for example, the probability that $X \in [a, b]$ for some $a, b \in \mathbb{R}$. This can be done in two ways:

- using P:
$$\sum_{\omega \in G} P(\omega)$$
where $G = \{\omega \in \Omega : X(\omega) \in [a, b]\}$.
- using P_X:
$$\sum_{x \in [a,b]} P_X(x)$$

Example. Lisa rolls two dices and annouces the sum L. We want to compute the distribution of L. The sample space is:

$$\Omega = \{(i, j) : i, j \in \{1, 2, 3, 4, 5, 6\}\}$$

and the random variable L is defined as:

$$L(i, j) = i + j$$

The probability distribution of L is:

$$\begin{aligned}
P_L(2) &= P(L = 2) = P(\{(1, 1)\}) = \frac{1}{36} \\
P_L(3) &= P(L = 3) = P(\{(1, 2), (2, 1)\}) = \frac{2}{36} \\
P_L(4) &= P(L = 4) = P(\{(1, 3), (2, 2), (3, 1)\}) = \frac{3}{36} \\
&\vdots \\
P_L(7) &= P(L = 7) = P(\{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}) = \frac{6}{36} \\
&\vdots \\
P_L(12) &= P(L = 12) = P(\{(6, 6)\}) = \frac{1}{36}
\end{aligned}$$

Definition 1.1.10 (Joint Distribution).
Let X and Y be two random variables defined on the same sample space Ω. The joint distribution of X and Y is the probability distribution of the pair (X, Y), which assigns a probability to each possible pair of values (x, y) that X and Y can take. Formally, it is defined as:

$$P_{X,Y}(x, y) = P(X = x, Y = y) = P(\{\omega \in \Omega : X(\omega) = x \text{ and } Y(\omega) = y\})$$

Definition 1.1.11 (Marginal Distribution).
Let X and Y be two random variables defined on the same sample space Ω. The marginal distribution of X is the probability distribution of X obtained by summing over all possible

values of Y. Formally, it is defined as:

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

where $P_{X,Y}(x,y)$ is the joint distribution of X and Y.

Example. Continuing with the example of Lisa rolling two dice, where $L = L_1L_2$ is the sum written as two digit number. The alphabet of L is:

$$L = \{02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12\}$$

Thus, the alphabet of L_1 is $\{0, 1\}$ and the alphabet of L_2 is $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$. Let's compute the probability P_{L_1}, knowing P_L. We marginalize:

$$P_{L_1}(1) = \sum_x P_{L_1,L_2}(1,x) = P_L(10) + P_L(11) + P_L(12) = \frac{3}{36} + \frac{2}{36} + \frac{1}{36} = \frac{6}{36} = \frac{1}{6}$$

1.1.5 Expectation and Independence of Random Variables

Definition 1.1.12 (Expected Value).
The expected value (or expectation) of a random variable X is defined as:

$$\mathbb{E}[X] = \sum_\omega X(\omega)P(\omega) = \sum_x xP_X(x)$$

where the sum is taken either over all outcomes ω in the sample space or over all values x in the alphabet of X.

Let's show that these two definitions are equivalent. We have:

$$\begin{aligned}\sum_\omega X(\omega)P(\omega) &= \sum_x \sum_{\omega:X(\omega)=x} X(\omega)P(\omega)\\ &= \sum_x x \sum_{\omega:X(\omega)=x} P(\omega)\\ &= \sum_x xP_X(x)\end{aligned}$$

Example. Let's consider the game of rolling a fair six-sided die and if the outcome is 6, you receive 10\$ otherwise you pay 1\$. We can define a random variable X that represents the amount of money you win (or lose) in this game. The expected value of X is:

$$\mathbb{E}[X] = (-1)\cdot\frac{5}{6} + 10\cdot\frac{1}{6} = -\frac{5}{6} + \frac{10}{6} = \frac{5}{6} \approx 0.83$$

This means that, on average, you can expect to win about 0.83\$ per game in the long run.

Theorem 1.1.3 (Linearity of Expectation)**.** Let $X_1, X_2, \ldots, X_n$ be n random variables and let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be real numbers. Then we have:

$$\mathbb{E}[\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n] = \alpha_1 \mathbb{E}[X_1] + \alpha_2 \mathbb{E}[X_2] + \cdots + \alpha_n \mathbb{E}[X_n]$$

Proof. We can write the expected value of the linear combination of the random variables as:

$$\mathbb{E}[\alpha_1 X_1 + \alpha_2 X_2 + \cdots + \alpha_n X_n] = \sum_{\omega} (\alpha_1 X_1(\omega) + \alpha_2 X_2(\omega) + \cdots + \alpha_n X_n(\omega)) P(\omega)$$

We can distribute the sum over the linear combination:

$$= \alpha_1 \sum_{\omega} X_1(\omega) P(\omega) + \alpha_2 \sum_{\omega} X_2(\omega) P(\omega) + \cdots + \alpha_n \sum_{\omega} X_n(\omega) P(\omega)$$

Recognizing that each sum is the expected value of the corresponding random variable, we get:

$$= \alpha_1 \mathbb{E}[X_1] + \alpha_2 \mathbb{E}[X_2] + \cdots + \alpha_n \mathbb{E}[X_n]$$

□

Definition 1.1.13 (Independence of Random Variables)**.**
Two random variables X and Y are said to be independent if for all values x and y, we have:

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

This definition can be generalized to n random variables $X_1, X_2, \ldots, X_n$ by requiring that for all values $x_1, x_2, \ldots, x_n$, we have:

$$P_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = P_{X_1}(x_1) P_{X_2}(x_2) \cdots P_{X_n}(x_n)$$