Lecture 02: Mathematical Basics (Probability)

Distributions, conditioning, independence,
and Bayes' rule

The puzzle: breakfast before 7?

Observed model

T is wake-up time, and $B \in \{\mathrm{T}, \mathrm{F}\}$ records whether breakfast is eaten.

$$\Pr[B = \mathrm{T} \mid T \leq 7] = ?$$

Guiding question

Which mass is the numerator? Which total mass is the denominator?

t	$\Pr[T = t, B = \mathrm{T}]$	$\Pr[T = t, B = \mathrm{F}]$
4	0.03	0
5	0.02	0
6	0.30	0.05
7	0.20	0.10

We will return to this after defining conditioning.

Roadmap

By the end of the lecture, we will be able to

1. distinguish outcomes, sample spaces, distributions, and random variables;
2. compute the distribution of a function $f(X)$;
3. read joint distributions and extract marginal distributions;
4. condition correctly and use the product rule, Bayes' rule, and the chain rule;
5. characterize independence in two equivalent ways; and
6. solve the studying–passing and wake-up–breakfast examples.

Recommended reference: [HPS14, Sec. 5.3, pp. 228–245].

Four objects to keep separate

1. Outcome

A single possible result, denoted ω.

2. Sample space

The set Ω of all possible outcomes.

3. Distribution

Nonnegative weights $p(\omega)$ satisfying $\sum_{\omega \in \Omega} p(\omega) = 1$.

4. Random variable

A function from an underlying experiment to a value space.

Convention for this lecture

Writing "X is a random variable over Ω" means that X takes values in Ω; its induced distribution is recorded by $\Pr[X = x]$.

See [HPS14, Sec. 5.3].

Probability basics: a biased coin

Let
$$\Omega = \{\text{Heads}, \text{Tails}\}.$$

Suppose
$$\Pr[X = \text{Heads}] = \frac{1}{3},$$
$$\Pr[X = \text{Tails}] = \frac{2}{3}.$$

Notation

$\Pr[X = x]$ is the probability mass at the outcome x.

A distribution p on a finite sample space satisfies

$$p(x) \geq 0 \quad (x \in \Omega), \qquad \sum_{x \in \Omega} p(x) = 1.$$

For an event $A \subseteq \Omega$,

$$\Pr[X \in A] = \sum_{x \in A} \Pr[X = x]\,.$$

Sample space: what can happen. **Distribution: how likely.**

Functions of a random variable

Let $f\colon \Omega \to \Omega'$ and let X take values in Ω. The transformed random variable $f(X)$ takes values in Ω' and

$$\boxed{\Pr[f(X) = y] = \sum_{\substack{x\in\Omega \\ f(x)=y}} \Pr[X = x]\,.}$$

Example

Let X be uniform on $\{1, 2, 3, 4\}$ and let $f(x) = x \bmod 2$. Then

$$\Pr[f(X) = 0] = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$

and similarly $\Pr[f(X) = 1] = 1/2$.

Push probability through the map

To find the mass at y, collect the probability of every input in the preimage $f^{-1}(y)$.

Suppose (X_1, X_2) takes values in $\Omega_1 \times \Omega_2$. Its joint distribution is

$$p_{X_1,X_2}(x_1, x_2) := \Pr[X_1 = x_1,\ X_2 = x_2]\,.$$

How to read a joint outcome

An outcome is an ordered pair (x_1, x_2): the first coordinate lies in Ω_1 and the second lies in Ω_2.

Temperature example

Let X_1 and X_2 be the temperatures in West Lafayette and Lafayette. Their values lie in $\mathbb{Z} \times \mathbb{Z}$, and the two coordinates can be correlated.

Joint information

The joint law contains both marginals and the dependence between the coordinates.

Marginal distributions: forget one coordinate

	a	b	row sum
u	p_{ua}	p_{ub}	p_{u*}
v	p_{va}	p_{vb}	p_{v*}
col. sum	p_{*a}	p_{*b}	1

First marginal

$$\Pr[X_1 = x_1] = \sum_{x_2 \in \Omega_2} p_{X_1,X_2}(x_1, x_2).$$

Second marginal

$$\Pr[X_2 = x_2] = \sum_{x_1 \in \Omega_1} p_{X_1,X_2}(x_1, x_2).$$

Marginalization keeps one coordinate and sums over the coordinate that is discarded.

Conditional distributions: restrict, then renormalize

For $\Pr[X_2 = x_2] > 0$, the conditional distribution of X_1 given $X_2 = x_2$ is

$$\boxed{\Pr[X_1 = x_1 \mid X_2 = x_2] = \frac{\Pr[X_1 = x_1, X_2 = x_2]}{\Pr[X_2 = x_2]}.}$$

Step 1: restrict

Retain only outcomes for which $X_2 = x_2$.

Step 2: renormalize

Divide by the total mass $\Pr[X_2 = x_2]$ of that slice.

Sanity check

For the fixed value x_2, the numbers $\Pr[X_1 = x_1 \mid X_2 = x_2]$ sum to 1 over $x_1 \in \Omega_1$.

See [HPS14, Secs. 5.3.1–5.3.2].

Three identities that are easy to conflate

Conditional probability

For $\Pr[B] > 0$,

$$\Pr[A \mid B] = \frac{\Pr[A \cap B]}{\Pr[B]}.$$

Product rule

$$\begin{aligned}\Pr[A \cap B] &= \Pr[A] \Pr[B \mid A] \\ &= \Pr[B] \Pr[A \mid B] .\end{aligned}$$

Bayes' rule

$$\Pr[A \mid B] = \frac{\Pr[B \mid A] \Pr[A]}{\Pr[B]}.$$

The product rule follows a sequential process; Bayes' rule reverses conditioning. See [HPS14, Sec. 5.3.2].

Chain rule

Theorem (Chain Rule)

Let $(X_1, \ldots, X_n)$ be jointly distributed. Whenever the displayed conditional probabilities are defined,

$$\begin{aligned}&\Pr[X_1 = x_1, \ldots, X_n = x_n] \\ &\quad = \Pr[X_1 = x_1] \prod_{i=2}^{n} \Pr[X_i = x_i \mid X_1 = x_1, \ldots, X_{i-1} = x_{i-1}] \,.\end{aligned}$$

Read the product from left to right

First-step probability $\times$ each later-step probability conditioned on the entire past. See [MU17, Ch. 1].

Sequential sampling: a two-bit example

Suppose $X_1 \in \{0, 1\}$ with $\Pr[X_1 = 0] = 1/2$. Conditioned on $X_1 = x_1$, let X_2 agree with x_1 with probability $2/3$ and disagree with probability $1/3$.

First step

$$\Pr[X_1 = 0] = \frac{1}{2}.$$

Second step

$$\Pr[X_2 = 1 \mid X_1 = 0] = \frac{1}{3}.$$

Multiply

$$\begin{aligned} \Pr[X_1 = 0, X_2 = 1] &= \frac{1}{2} \cdot \frac{1}{3} \\ &= \frac{1}{6}. \end{aligned}$$

A generative description becomes a joint probability through the product rule.

Independence of random variables

Definition

The random variables X_1 and X_2 are independent if, for every $x_1 \in \Omega_1$ and $x_2 \in \Omega_2$,

$$\Pr[X_1 = x_1, X_2 = x_2] = \Pr[X_1 = x_1] \Pr[X_2 = x_2] \,.$$

Equivalent conditional statement

Whenever $\Pr[X_1 = x_1] > 0$,

$$\Pr[X_2 = x_2 \mid X_1 = x_1] = \Pr[X_2 = x_2] \,.$$

Thus learning $X_1 = x_1$ does not change the distribution of X_2.

Independence is a property of the joint distribution, not merely of the two marginals.

First example: studying and passing

Let $S \in \{\mathrm{Y}, \mathrm{N}\}$ record whether I studied, and let $P \in \{\mathrm{Y}, \mathrm{N}\}$ record whether I passed. The joint distribution is

	$P = \mathrm{Y}$	$P = \mathrm{N}$
$S = \mathrm{Y}$	$1/2$	$1/4$
$S = \mathrm{N}$	0	$1/4$

Sample space

$$\Omega = \{\mathrm{Y}, \mathrm{N}\} \times \{\mathrm{Y}, \mathrm{N}\}.$$

Each cell gives $\Pr[S = s, P = p]$ for one pair (s, p).

First check

The four joint probabilities are nonnegative and sum to 1.

Marginals from the joint table

	$P = \mathrm{Y}$	$P = \mathrm{N}$	$\Pr[S = s]$
$S = \mathrm{Y}$	$1/2$	$1/4$	$3/4$
$S = \mathrm{N}$	0	$1/4$	$1/4$
$\Pr[P = p]$	$1/2$	$1/2$	1

The probability that I pass is

$$\begin{aligned}\Pr[P = \mathrm{Y}] &= \Pr[S = \mathrm{Y}, P = \mathrm{Y}] \\ &\quad + \Pr[S = \mathrm{N}, P = \mathrm{Y}] \\ &= \frac{1}{2} + 0 = \frac{1}{2}.\end{aligned}$$

The probability that I study is

$$\Pr[S = \mathrm{Y}] = \frac{1}{2} + \frac{1}{4} = \frac{3}{4}.$$

Direction matters

Column sums give the marginal of P; row sums give the marginal of S.

Conditioning on having studied

The probability that I pass conditioned on having studied is

$$\Pr[P = \mathrm{Y} \mid S = \mathrm{Y}] = \frac{\Pr[P = \mathrm{Y}, S = \mathrm{Y}]}{\Pr[S = \mathrm{Y}]}$$
$$= \frac{1/2}{3/4} = \frac{2}{3}.$$

Marginal probability

$$\Pr[P = \mathrm{Y}] = \frac{1}{2}.$$

Conclusion

Since $2/3 \neq 1/2$, conditioning on $S = \mathrm{Y}$ changes the probability of passing. Hence S and P are not independent.

Second example: wake-up time and breakfast

Let $T \in \{4, 5, 6, 7, 8, 9, 10\}$ be wake-up time, and let $B \in \{\mathrm{T}, \mathrm{F}\}$ indicate breakfast.

t	$\Pr[T = t, B = \mathrm{T}]$	$\Pr[T = t, B = \mathrm{F}]$	$\Pr[T = t]$
4	0.03	0	0.03
5	0.02	0	0.02
6	0.30	0.05	0.35
7	0.20	0.10	0.30
8	0.10	0.08	0.18
9	0.05	0.05	0.10
10	0	0.02	0.02
total	0.70	0.30	1

Question

Compute $\Pr[B = \mathrm{T} \mid T \leq 7]$.

Solve the puzzle: condition on $T \leq 7$

Numerator: breakfast and wake by 7

$$\begin{aligned}\Pr[B = \mathrm{T},\, T \leq 7] &= 0.03 + 0.02 + 0.30 + 0.20 \\ &= 0.55.\end{aligned}$$

Denominator: wake by 7

$$\begin{aligned}\Pr[T \leq 7] &= 0.03 + 0.02 + 0.35 + 0.30 \\ &= 0.70.\end{aligned}$$

Conditional probability

$$\begin{aligned}\Pr[B = \mathrm{T} \mid T \leq 7] &= \frac{0.55}{0.70} \\ &= \frac{11}{14} \\ &\approx 0.786.\end{aligned}$$

Discard the rows with $T > 7$, then renormalize the remaining mass.

A compact workflow for probability tables

1. **Name the joint outcome.** Identify every coordinate and its value space.
2. **Marginalize by summing.** Sum over coordinates that are not retained.
3. **Condition by restricting and normalizing.** The denominator is the probability of the conditioning event.
4. **Use sequential structure.** Multiply conditional factors with the product or chain rule.
5. **Test independence.** Compare joint mass with the product of marginals, or conditional mass with the marginal.

Two denominator checks

A conditional probability is defined only when the denominator is positive, and all conditional probabilities over the remaining outcomes must sum to 1.

Takeaways

1. **Probability model:** sample space + distribution.
2. **Function of a random variable:** collect mass over preimages.
3. **Marginal:** sum out coordinates that are forgotten.
4. **Conditional distribution:** restrict to an event and renormalize.
5. **Product, Bayes, and chain rules:** organize joint and conditional probabilities.
6. **Independence:** conditioning on one variable does not change the other.

The central calculation pattern

Identify the probability mass relevant to the question, then normalize by the correct total mass.

References

[BT08] Dimitri P. Bertsekas and John N. Tsitsiklis.
***Introduction to Probability*.**
Athena Scientific, 2 edition, 2008.

[HPS14] Jeffrey Hoffstein, Jill Pipher, and Joseph H. Silverman.
***An Introduction to Mathematical Cryptography*.**
Undergraduate Texts in Mathematics. Springer, New York, NY, 2 edition, 2014.
Recommended reference: Section 5.3, "Probability Theory," pp. 228–245.
doi:10.1007/978-1-4939-1711-2.

[MU17] Michael Mitzenmacher and Eli Upfal.
***Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*.**
Cambridge University Press, 2 edition, 2017.
doi:10.1017/CBO9781316477343.

[Ros19] Sheldon M. Ross.
***A First Course in Probability*.**
Pearson, 10 edition, 2019.