

Inferential Statistics

Lecture 1: Introduction and random variables

Jérémy Hervelin

Pforzheim University
`j.hervelin@lb.hs-pforzheim.de`

5 October 2026 13:45–15:15
Winter semester 2026/27

Adapted from the teaching materials of Prof. Dr. Thilo Klein.

Our course and the examination

- 14 lectures: 11 theory lectures and 3 practical sessions in R.
- Each teaching day: 13:45–15:15 and 15:30–17:00.
- 5 and 12 October; 2, 9, 16 and 30 November; 7 December.

Final exam: Friday, 18 December 2026, 13:00–14:00.

- We will practise calculations, method selection and interpretation.
- Exam resources and administrative instructions will be confirmed by the instructor.

What you will learn today

1. Explain why a sample does not give perfect information about a population.
2. Distinguish a random variable X from an observed value x .
3. Calculate discrete probabilities using a probability mass function and a cumulative distribution function.
4. Explain why continuous probabilities are areas under a density.

No new calculus techniques today: an integral represents accumulated area.

The course roadmap

Lectures	Main question
1–3	How do we model uncertain outcomes?
4–5	How do samples and estimates vary?
6–7	How precisely can we estimate a parameter?
8–9	How do we evaluate evidence against a hypothesis?
10	R: one-sample inference
11–12	Two-sample inference, followed by R
13–14	Categorical data, followed by R and revision

A sample can answer a population question

Illustrative question: What proportion of young job seekers find employment within six months?

- Population: the young job seekers about whom we want to learn.
- Sample: the people whose outcomes we observe.
- Descriptive statistics summarise the observed sample.
- Inferential statistics quantify what the sample tells us about the population, under explicit assumptions.

A large sample can still mislead if it systematically excludes some people.

Probability describes uncertainty

Before observing an outcome, several values are possible.

- A probability model assigns probabilities to possible outcomes.
- The model is an assumption about the process that generates data.
- Data can help us assess whether that assumption is reasonable.

Example: We may model employment as a binary outcome. This does not yet tell us whether a training programme causes employment.

A random variable assigns a number to an outcome

Let ω denote an outcome in the sample space Ω .

$$X : \Omega \longrightarrow \mathbb{R}.$$

- X : the random variable, before the outcome is observed.
- x : a particular value, or realisation, of X .

For employment, set $X = 1$ if the person is employed after six months and $X = 0$ otherwise. Observing an employed person gives $x = 1$.

Discrete and continuous outcomes

Discrete	Continuous model
Countable possible values	Values along an interval
Number of job offers	Exact waiting time until an offer
Employment indicator: 0 or 1	Exact age
Probability at individual values	Probability over intervals
Recorded age in completed years is discrete, even if exact age is modelled as continuous.	

Classroom check – Q1

Concept check

1–2 minutes

Is a count of job offers discrete or continuous? Choose one:

- A. Discrete;
- B. Continuous;
- C. Neither.

Think individually, then compare your reasoning with a neighbour.

Lecture 1, classroom quiz Q1. Same question number as the printed quiz.

A familiar discrete experiment

Toss a fair coin twice. Assume the tosses are independent.

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

Each ordered outcome has probability $1/4$.

Let X be the number of heads:

Outcome	TT	HT	TH	HH
Value of X	0	1	1	2

Notice: Two different outcomes give the same value $X = 1$.

Same probability structure as the two-child example in the original lecture; assumptions made explicit.

The probability mass function

The probability mass function (PMF) gives $f_X(x) = P(X = x)$.

$$f_X(x) = \begin{cases} 0.25 & x = 0, \\ 0.50 & x = 1, \\ 0.25 & x = 2, \\ 0 & \text{otherwise.} \end{cases}$$

The support is $\mathcal{T}_X = \{0, 1, 2\}$, the values with positive probability.

Interpretation: $P(X = 1) = P(HT) + P(TH) = 0.50$.

A valid probability mass function

Two conditions must hold:

$$f_X(x) \geq 0 \quad \text{for every } x, \quad \sum_{x \in \mathcal{T}_X} f_X(x) = 1.$$

Concept check: A proposed model assigns probabilities 0.2, 0.5 and 0.4 to three exhaustive outcomes. Is it valid? Explain.

Checking these conditions establishes mathematical validity. It does not establish that the model fits the real world.

Classroom check – Q2

Application

1–2 minutes

A PMF assigns probabilities 0.2, 0.5 and 0.3 to $X = 0, 1$ and 2. Find $P(X \geq 1)$.

Think individually, then compare your reasoning with a neighbour.

Lecture 1, classroom quiz Q2. Same question number as the printed quiz.

The employment indicator

In an **illustrative model**, let the employment probability be $p = 0.6$.

$$P(X = 1) = 0.6, \quad P(X = 0) = 0.4.$$

- Here p is a model parameter: a probability in the population.
- A sample proportion is calculated from the people we observe.
- The sample proportion need not equal p in any particular sample.

Next lecture: this model is called the Bernoulli distribution.

Illustrative probabilities, not empirical findings about any programme.

The original football example

Let X be the number of goals in a match, **conditional on at most five goals**.

x	0	1	2	3	4	5
$P(X = x)$	0.1	0.2	0.3	0.2	0.1	0.1

This proposed model has non-negative probabilities that sum to one.

Worked example: The probability of exactly four goals is $P(X = 4) = 0.1$.

Numerical teaching example retained from Thilo Klein; original data provenance not supplied.

Observed frequencies and model probabilities

The original lecture also supplies these counts for matches with at most five goals:

Goals	0	1	2	3	4	5
Count	23	35	74	69	56	23
Relative frequency	.082	.125	.264	.246	.200	.082

- The counts sum to 280. Divide each count by 280.
- Relative frequencies describe this dataset; model probabilities describe an assumed process.
- Differences can reflect random variation or a poor model.

Legacy teaching counts. Rounded frequencies need not sum to exactly one.

The cumulative distribution function

Often we need the probability of being **at or below** a threshold.

$$F_X(x) = P(X \leq x).$$

For a discrete variable, add the probabilities at all values up to x .

For the football model:

$$F_X(3) = 0.1 + 0.2 + 0.3 + 0.2 = 0.8.$$

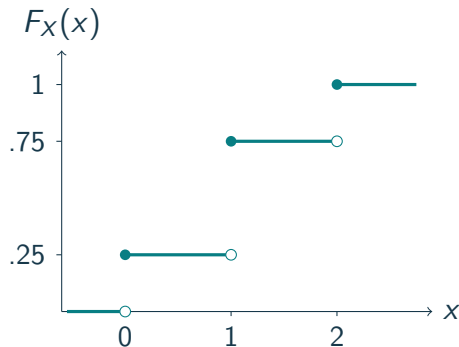
Interpretation: the model assigns probability 0.8 to at most three goals.

A discrete CDF is a step function

For the number of heads:

$$F_X(x) = \begin{cases} 0 & x < 0, \\ 0.25 & 0 \leq x < 1, \\ 0.75 & 1 \leq x < 2, \\ 1 & x \geq 2. \end{cases}$$

The jump at x equals $P(X = x)$.



Properties of every CDF

- $F_X(x)$ is non-decreasing: including more outcomes cannot reduce probability.
- $F_X(x)$ is right-continuous: the probability at x is included.
- $\lim_{x \rightarrow -\infty} F_X(x) = 0$.
- $\lim_{x \rightarrow +\infty} F_X(x) = 1$.

Check the notation: $P(X < x)$ and $P(X \leq x)$ can differ for a discrete variable.

Probabilities from a CDF

For $a < b$:

$$P(X > a) = 1 - F_X(a),$$

$$P(X < a) = F_X(a) - P(X = a),$$

$$P(a < X \leq b) = F_X(b) - F_X(a),$$

$$P(a \leq X \leq b) = F_X(b) - F_X(a) + P(X = a).$$

First translate the words into an inequality. Then decide which endpoints to include.

Classroom check – Q3

Exam-style

1–2 minutes

In the football model, the probabilities for 0, 1, 2, 3, 4 and 5 goals are 0.1, 0.2, 0.3, 0.2, 0.1 and 0.1. Calculate $P(X \leq 3)$ and $P(X > 2)$.

Think individually, then compare your reasoning with a neighbour.

Lecture 1, classroom quiz Q3. Same question number as the printed quiz.

Worked example: at least two goals

In the original football model, $P(X = 0) = 0.1$ and $P(X = 1) = 0.2$.

“At least two” means $X \geq 2$.

$$P(X \geq 2) = 1 - P(X \leq 1) = 1 - (0.1 + 0.2) = 0.7.$$

Why not $1 - F_X(2)$? That expression gives $P(X > 2)$ and excludes two goals.

Continuous outcomes: probability as area

Let T be an exact waiting time until a job offer.

- Under a continuous model, T can take any value in an interval.
- We describe probabilities using a density $f_T(t)$.
- The probability of a range is the area under the density over that range.

$$P(a < T \leq b) = \int_a^b f_T(t) dt.$$

This course considers continuous models that possess a density.

A density must have total area one

$$f_T(t) \geq 0, \quad \int_{-\infty}^{\infty} f_T(t) dt = 1.$$

- A density value is a height, not a probability at a point.
- A density can exceed one while its total area remains one.
- Density has units: if T is in months, density is per month.

Classroom check – Q4

Concept check

1–2 minutes

True or false: a continuous density of 1.4 at a point is impossible. Explain.

Think individually, then compare your reasoning with a neighbour.

Lecture 1, classroom quiz Q4. Same question number as the printed quiz.

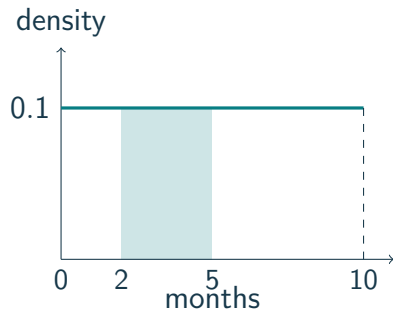
An area calculation without calculus

Suppose, illustratively, that a waiting time is uniform from 0 to 10 months.

$$f_T(t) = 0.1 \quad (0 \leq t \leq 10).$$

Then

$$P(2 < T \leq 5) = (5 - 2) \times 0.1 = 0.3.$$



Synthetic model for learning; actual job-search durations need not be uniform.

The CDF accumulates area

For a variable with a density:

$$F_T(t) = \int_{-\infty}^t f_T(u) du.$$

The symbol u is just the integration variable.

In the uniform example, $F_T(5) = 0.5$ and $F_T(2) = 0.2$.

$$P(2 < T \leq 5) = F_T(5) - F_T(2) = 0.3.$$

A single point has zero area

For a variable with a density:

$$P(T = t) = 0.$$

- An interval of zero width has zero area.
- Adding or removing an endpoint does not change an interval probability.

$$P(a < T < b) = P(a \leq T \leq b) = F_T(b) - F_T(a).$$

Measurement matters: A reported waiting time of “3 months” may represent a rounded interval, rather than the exact point $T = 3$.

A histogram needs the right scale

For a **density-scaled histogram**, the height of a bin is

$$\text{height} = \frac{\text{relative frequency in the bin}}{\text{bin width}}.$$

Its area equals the relative frequency in that bin; all bin areas sum to one.

Example: If 30% of observations fall in a bin of width 3 months, its height is $0.30/3 = 0.10$ per month.

A histogram drawn with counts on the vertical axis does not generally have total area one.

The empirical CDF counts observations

For observed values $x_1, \dots, x_n$, the empirical CDF is

$$\hat{F}_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}\{x_i \leq x\}.$$

Here $\mathbf{1}\{x_i \leq x\}$ equals one when the condition holds, and zero otherwise.

For observations 1, 2, 2, 5, we get $\hat{F}_4(2) = 3/4$.

The empirical CDF is a step function. Integrating a density histogram gives a different, bin-based approximation.

Classroom check – Q5

Exam-style

1–2 minutes

For observations 1, 2, 2, 5, what is the empirical CDF at 2? Does it estimate a population CDF or establish it exactly?

Think individually, then compare your reasoning with a neighbour.

Lecture 1, classroom quiz Q5. Same question number as the printed quiz.

The distinctions to remember

Object	Meaning
X versus x	Random variable versus a particular value
$f_X(x)$, discrete	Probability $P(X = x)$
$f_X(x)$, continuous	Density height; interval areas give probabilities
$F_X(x)$	Probability $P(X \leq x)$ in either case
$\hat{F}_n(x)$	Proportion of observed values at or below x

Preparation for Lecture 2

- Revisit the football model and translate “at most”, “more than” and “at least” into inequalities.
- Check that you can explain why $P(X = 1)$ differs from $P(X \leq 1)$.
- Next: summarising a distribution by its expected value and variance, then choosing discrete distribution models.

Reading retained from the course: Handl and Kuhlenkasper (2018), *Einführung in die Statistik: Theorie und Praxis mit R*, Springer Spektrum (in German).

Reference: endpoint conventions

For $a < b$, with $F_X(a^-) = \lim_{x \uparrow a} F_X(x)$:

$$P(X = a) = F_X(a) - F_X(a^-),$$

$$P(X < a) = F_X(a^-),$$

$$P(a \leq X < b) = F_X(b) - F_X(a) - P(X = b) + P(X = a),$$

$$P(a < X < b) = F_X(b) - F_X(a) - P(X = b).$$

For a continuous distribution, the point probabilities vanish.

Reference material retained from the original Lecture 1; optional during class.