

# Inferential Statistics

## Lecture 2: Distribution parameters and discrete models

Jérémy Hervelin

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

5 October 2026 15:30–17:00  
Winter semester 2026/27

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

# What you will learn today

- Calculate and interpret expected value, variance and standard deviation.
- Choose between uniform, Bernoulli, binomial, geometric, hypergeometric and Poisson models.
- State the assumptions that make each model useful.
- Check the units and the event before calculating a probability.

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

# Expected value: a probability-weighted average

Imagine observing many independent outcomes from the same model.

The expected value describes their long-run average, provided the expectation exists.

$$\mu = E(X) = \sum_{x \in \mathcal{T}_X} x f_X(x).$$

Multiply each possible value by its probability, then add.

For the number of heads in two independent fair tosses:

$$E(X) = 0(0.25) + 1(0.50) + 2(0.25) = 1.$$

# Classroom check – Q1

## Concept check

1–2 minutes

Can  $E(X)$  be a value that  $X$  never takes? Give an example.

*Think individually, then compare your reasoning with a neighbour.*

Lecture 2, classroom quiz Q1. Same question number as the printed quiz.

# An average need not be a possible outcome

A fair die has expected value

$$E(X) = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = 3.5.$$

You cannot roll 3.5. The expectation is not a prediction of a single roll.

For an illustrative employment indicator with  $P(X = 1) = 0.6$ :

$$E(X) = 0(0.4) + 1(0.6) = 0.6.$$

The mean of a binary outcome equals its success probability.

# The original choice under uncertainty

| Option A               | Option B                                                     |
|------------------------|--------------------------------------------------------------|
| Receive \$240 for sure | Receive \$1,000 with probability 0.25; otherwise receive \$0 |

$$E(A) = 240, \quad E(B) = 0(0.75) + 1000(0.25) = 250.$$

Option B has a higher expected monetary payoff.

A person may prefer the certainty of A. Expected payoff alone does not describe risk preferences.

Payoffs retained from Klein's Tversky–Kahneman example.

# Variance measures spread around the mean

Squared deviations give positive weight to departures in either direction.

$$\sigma^2 = \text{Var}(X) = E[(X - \mu)^2].$$

For a discrete variable:

$$\text{Var}(X) = \sum_{x \in \mathcal{T}_X} (x - \mu)^2 f_X(x).$$

Variance is in squared units. Standard deviation  $\sigma = \sqrt{\text{Var}(X)}$  is in the same units as  $X$ .

## Worked example: the risk in the two options

Option A always pays \$240, so  $\text{Var}(A) = 0$ .

For B, the mean is \$250:

$$\begin{aligned}\text{Var}(B) &= (0 - 250)^2(0.75) + (1000 - 250)^2(0.25) \\ &= 187,500 \quad \text{dollars squared,}\end{aligned}$$

$$\text{SD}(B) = \sqrt{187,500} \approx 433.01 \quad \text{dollars.}$$

The higher expected payoff comes with considerable variability.

# Continuous moments and sample summaries

For a variable with density  $f_X$ , when the moments exist:

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx, \quad \text{Var}(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx.$$

These are population quantities defined by a probability model.

For observed data,  $\bar{x} = \frac{1}{n} \sum_i x_i$  is a sample mean.

The descriptive average squared deviation uses  $1/n$ . Later, the usual unbiased sample variance uses  $1/(n-1)$ .

# Choosing a discrete model

| Question                                          | Candidate model                     |
|---------------------------------------------------|-------------------------------------|
| One binary outcome                                | Bernoulli                           |
| Successes in a fixed number of independent trials | Binomial                            |
| Failures before the first success                 | Geometric                           |
| Successes in a sample without replacement         | Hypergeometric                      |
| Events in a fixed time interval                   | Poisson, under suitable assumptions |
| Equally likely values $1, \dots, N$               | Discrete uniform                    |

# Discrete uniform distribution

If the integers  $1, \dots, N$  are equally likely:

$$P(X = x) = \frac{1}{N}, \quad x = 1, \dots, N.$$

$$E(X) = \frac{N+1}{2}, \quad \text{Var}(X) = \frac{N^2-1}{12}.$$

**Original die example:** With  $N = 6$ ,  $E(X) = 3.5$  and  $\text{Var}(X) = 35/12 \approx 2.917$ .

These formulas assume consecutive integer values starting at one.

# Bernoulli outcomes and repeated trials

For one binary outcome  $Y$ :

$$P(Y = 1) = p, \quad P(Y = 0) = 1 - p, \quad E(Y) = p, \quad \text{Var}(Y) = p(1 - p).$$

A Bernoulli process repeats trials that are:

- independent of one another;
- governed by the same success probability  $p$ .

“Success” simply labels the event being counted. It need not mean a desirable outcome.

# Classroom check – Q2

## Application

1–2 minutes

For an illustrative employment indicator with  $p = 0.6$ , find  $E(X)$  and  $\text{Var}(X)$ .

*Think individually, then compare your reasoning with a neighbour.*

Lecture 2, classroom quiz Q2. Same question number as the printed quiz.

# The binomial distribution counts successes

Let  $X$  count successes in  $n$  independent Bernoulli trials with common probability  $p$ .

$$X \sim \text{Bin}(n, p), \quad P(X = x) = \binom{n}{x} p^x (1 - p)^{n-x}, \quad x = 0, \dots, n.$$

- $p^x(1 - p)^{n-x}$ : probability of one particular sequence.
- $\binom{n}{x} = \frac{n!}{x!(n-x)!}$ : number of sequences with  $x$  successes.

A Bernoulli variable is the special case  $n = 1$ .

# Classroom check – Q3

## Exam-style

1–2 minutes

Five independent applicants each receive an offer with probability 0.6. Find the probability of exactly four offers.

*Think individually, then compare your reasoning with a neighbour.*

Lecture 2, classroom quiz Q3. Same question number as the printed quiz.

# An employment-count example

In a **synthetic model**, five young people each have employment probability  $p = 0.6$ , independently.

What is the probability that exactly four are employed?

$$P(X = 4) = \binom{5}{4} (0.6)^4 (0.4) = 5(0.1296)(0.4) = 0.2592.$$

The five sequences correspond to the five possible people who are not employed.

In an actual programme, shared employers or local shocks may make outcomes dependent.

# The original five-trial calculation

With five independent fair coin tosses:

$$P(\text{exactly four heads}) = \binom{5}{4} (0.5)^4 (0.5) = 0.15625.$$

Exactly four heads also means exactly one tail:

$$P(\text{exactly one tail}) = \binom{5}{1} (0.5) (0.5)^4 = 0.15625.$$

The events are identical, not counterexamples to each other.

Numerical exercise adapted from Klein.

# Binomial location and spread

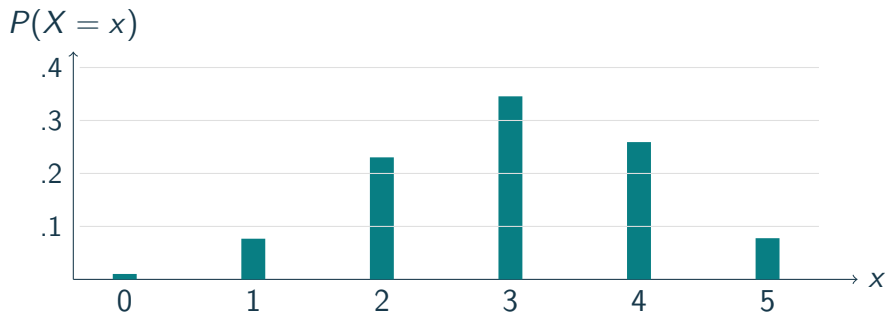
$$E(X) = np, \quad \text{Var}(X) = np(1 - p).$$

For five people with  $p = 0.6$ :

$$E(X) = 3, \quad \text{Var}(X) = 1.2, \quad \text{SD}(X) \approx 1.095.$$

- $n$  fixes the number of trials;  $X$  is the random count.
- For  $p = 0.5$ , the PMF is symmetric.
- For  $0 < p < 0.5$  it is right-skewed; for  $0.5 < p < 1$  it is left-skewed.

# A binomial probability plot



**Model:**  $n = 5$ ,  $p = 0.6$ . The heights sum to one.

The most likely count is three, but other counts remain possible.

# The geometric distribution counts failures

Repeat independent trials with constant success probability  $0 < p \leq 1$  until the first success.

Let  $X$  count failures **before** that success:

$$P(X = x) = (1 - p)^x p, \quad x = 0, 1, 2, \dots$$

$$E(X) = \frac{1 - p}{p}, \quad \text{Var}(X) = \frac{1 - p}{p^2}.$$

In a simplified job-search model, a trial could be an application. Independence and a constant offer probability are strong assumptions.

# Trials until success: a different convention

Let  $Y$  count all trials, including the successful one. Then  $Y = X + 1$ .

$$P(Y = y) = (1 - p)^{y-1}p, \quad y = 1, 2, \dots$$

$$E(Y) = \frac{1}{p}, \quad \text{Var}(Y) = \frac{1 - p}{p^2}.$$

**Original die exercise:** The probability that the *first* six occurs on roll eight is

$$P(Y = 8) = (5/6)^7(1/6) \approx 0.04651.$$

Getting a six on roll eight, regardless of earlier rolls, has probability  $1/6$ .

# Sampling without replacement changes the model

The original urn contains  $N = 10$  balls, of which  $M = 4$  are white.

Draw  $n = 3$  balls and count the white balls.

- With replacement:  $X \sim \text{Bin}(3, 0.4)$ .
- Without replacement: draws are dependent and the number of white balls remaining changes.

The second count follows a hypergeometric distribution.

Applied interpretation: select three distinct files from ten applicants, four of whom have vocational qualifications.

# The hypergeometric distribution

From  $N$  objects containing  $M$  successes, select  $n$  objects uniformly without replacement.

$$P(X = x) = \frac{\binom{M}{x} \binom{N-M}{n-x}}{\binom{N}{n}}.$$

$$\max\{0, n - (N - M)\} \leq x \leq \min\{n, M\}, \quad x \text{ integer.}$$

- Numerator: selections with  $x$  successes and  $n - x$  failures.
- Denominator: all equally likely selections of  $n$  objects.

# Classroom check – Q4

## Concept check

1–2 minutes

Four out of ten applicants have vocational qualifications. You randomly select three distinct applicants. Which model describes the number with vocational qualifications?

- A. Binomial;
- B. Hypergeometric;
- C. Geometric. Explain.

*Think individually, then compare your reasoning with a neighbour.*

Lecture 2, classroom quiz Q4. Same question number as the printed quiz.

# Worked comparison: two white balls

With replacement:

$$P(X = 2) = \binom{3}{2} (0.4)^2 (0.6) = 0.288.$$

Without replacement:

$$P(X = 2) = \frac{\binom{4}{2} \binom{6}{1}}{\binom{10}{3}} = \frac{6 \times 6}{120} = 0.3.$$

The fraction of successes in the population is the same. The dependence between draws changes the distribution.

# Hypergeometric moments and approximation

Write  $p = M/N$ , with  $N > 1$ . Then

$$E(X) = np, \quad \text{Var}(X) = np(1-p)\frac{N-n}{N-1}.$$

The finite population factor  $(N-n)/(N-1)$  reduces the variance relative to binomial sampling with replacement.

For  $N = 10$ ,  $M = 4$ ,  $n = 3$ , the variance is  $0.72(7/9) = 0.56$ .

A binomial approximation is useful when the sampling fraction  $n/N$  is small, not simply whenever  $N$  is large.

# The original used-book exercise

Six books are available. Four have coffee stains. You select four books uniformly at random, without replacement.

What is the probability that all four selected books have stains?

$$P(X = 4) = \frac{\binom{4}{4} \binom{2}{0}}{\binom{6}{4}} = \frac{1}{15} \approx 0.06667.$$

The random-selection assumption matters. Choosing books after inspecting them would change the model.

# Poisson counts over a fixed interval

A Poisson model describes an event count  $X$  with expected count  $\lambda > 0$ :

$$P(X = x) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

$$E(X) = \lambda, \quad \text{Var}(X) = \lambda.$$

For a homogeneous Poisson process, counts in disjoint time intervals are independent and the event rate is constant.

Examples include customer arrivals or requests to a job-search helpdesk, when these assumptions are reasonable.

# The original shop example: align the time units

A shop opens ten hours a day and averages one customer per day. Assume a constant Poisson arrival rate.

$$\lambda_{\text{hour}} = 1/10 = 0.1.$$

For the number  $K$  of customers in one hour:

$$\begin{aligned} P(K > 1) &= 1 - P(K = 0) - P(K = 1) \\ &= 1 - e^{-0.1}(1 + 0.1) \\ &\approx 0.004679. \end{aligned}$$

That is about 0.468%, not 0.468 as a probability.

Retain full precision until the last step.

# Classroom check – Q5

## Exam-style

1–2 minutes

A homogeneous Poisson helpdesk receives an average of two calls per hour. What is the probability of zero calls in half an hour?

*Think individually, then compare your reasoning with a neighbour.*

Lecture 2, classroom quiz Q5. Same question number as the printed quiz.

# How Poisson connects to binomial

A binomial model approaches a Poisson model when

$$n \longrightarrow \infty, \quad p \longrightarrow 0, \quad np \longrightarrow \lambda.$$

Many opportunities, each with a small success probability, can produce a count with a moderate expected value.

An event count is not automatically Poisson. Clustering, changing rates or unequal opportunities can make its variance differ substantially from its mean.

# Returning to the football counts

|       |    |    |    |    |    |    |    |   |
|-------|----|----|----|----|----|----|----|---|
| Goals | 0  | 1  | 2  | 3  | 4  | 5  | 6  | 7 |
| Count | 23 | 35 | 74 | 69 | 56 | 23 | 19 | 7 |

These original counts sum to 306 and give

$$\bar{x} = \frac{\sum_x x n_x}{306} = \frac{892}{306} \approx 2.9150.$$

Unlike Lecture 1, these counts include matches with six or seven goals.

The original Poisson model uses  $\lambda = 2.9$ , a one-decimal rounding of the sample mean. Visual resemblance alone does not validate a model.

Legacy teaching counts from Klein. Their underlying empirical source is not supplied.

# Preparation for the next lecture

- Choose the model from the sampling process, not from the topic of the story.
- Distinguish a single Bernoulli outcome from a binomial count.
- For a geometric variable, state whether you count failures or all trials.
- For Poisson counts, match the time unit of the parameter and question.

Next lecture: continuous uniform, exponential and normal models, followed by distributions used in inference.