

Introduction

What Statistics Is, and What This Course Covers

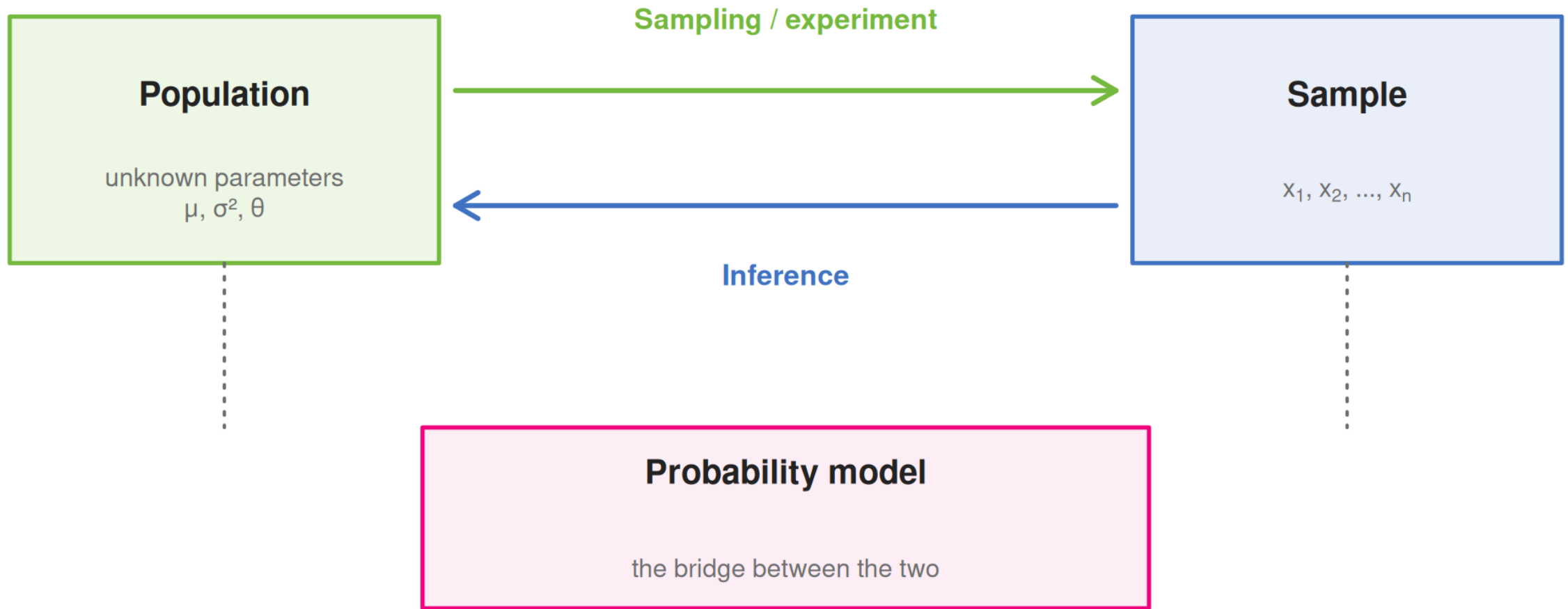
Course Overview

- Statistical inference and decision theory
- Point estimation: methods and properties
- Interval estimation and confidence intervals
- Hypothesis testing: theory and applications
- Asymptotic theory and limit theorems
- Bayesian inference

The thread running through all of it

We observe a **sample**. We want to say something about the **population** it came from — and quantify how sure we are.

Statistics



Sampling goes one way; the interesting (and hard) direction is the other one.

Statistics

- Statistics can generally be divided into two parts:
 - **Descriptive statistics** — summarising the data you actually have
 - **Inferential statistics** — drawing conclusions beyond the data you have
- Statistics typically involves analysing a **sample** taken from a **population**.
- From the sample, we generally make statements about the population.
- It is therefore important that the sample is **representative** of the population.

Working assumption

For most of this course we will simply assume that the samples *are* representative. When that assumption fails, no amount of clever mathematics repairs it.

Sampling, Live

Draw a sample from a known population and watch how well $\bar{x}$ recovers μ .

Population

Skewed (income-like) ▾

Sample size n

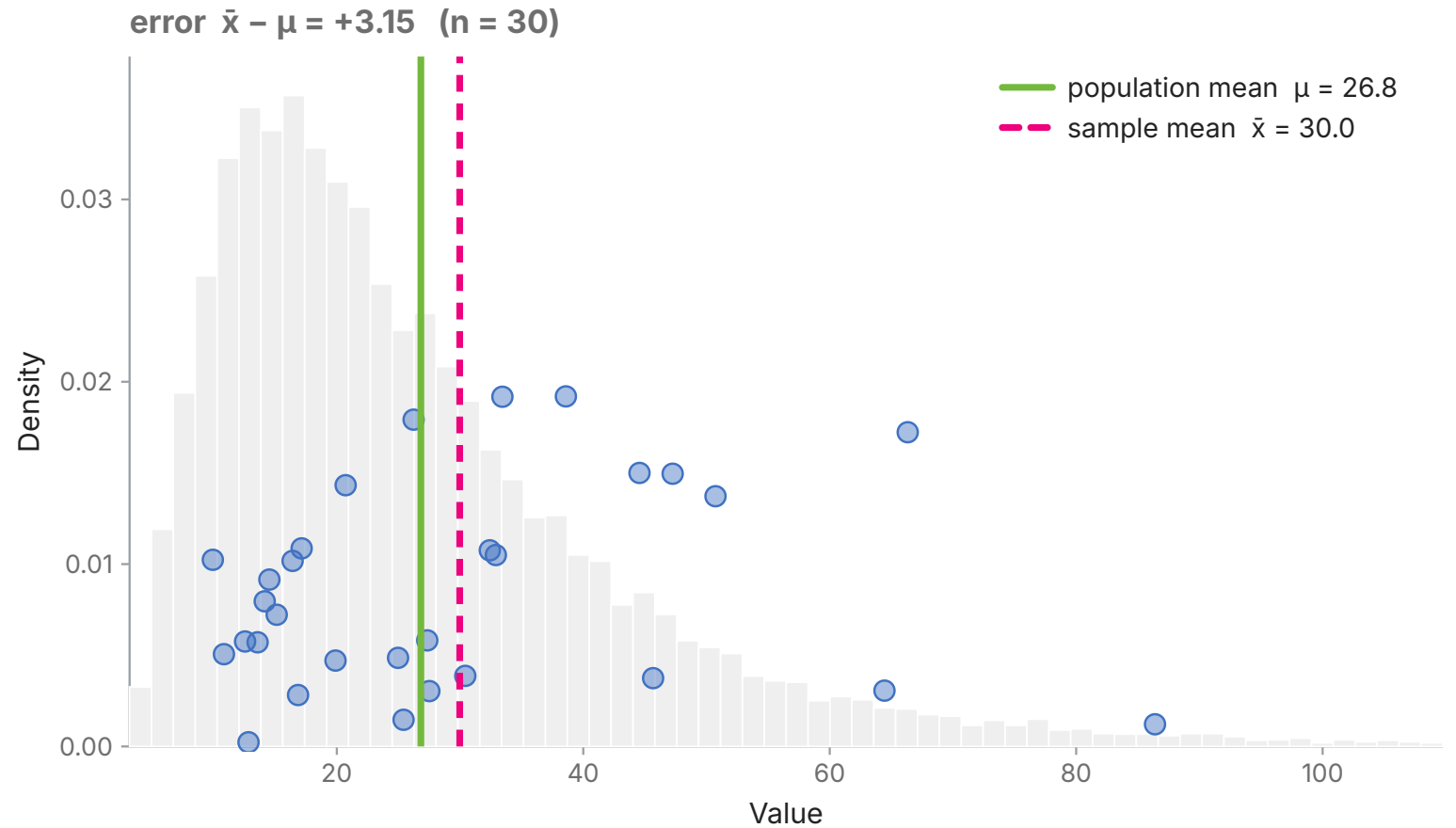
30



Sampling

Random (representative) ▾

Draw a new sample



What the Simulation Shows

- With **random sampling**, larger n shrinks the error — and the error is *centred on zero*: press “draw” repeatedly and it wanders either side of μ .
- With a **biased scheme**, larger n does **not** help. The error stays stubbornly on one side, no matter how much data you collect.

! The point

Sample size fixes **noise**. It never fixes **bias**. Everything in this course assumes the sampling itself was done honestly.

Where We Go From Here



Next: Random Variables

- Turning outcomes into numbers
- CDFs, PMFs, PDFs
- Moments and common distributions
- Laws of large numbers, CLT



Then: Inference

- Multivariate random variables
- Estimation theory
- Hypothesis testing