Unit 2

Expectation, Normal distribution and CLT

Ethan Levien

Math 50

September 4, 2025

Outline

Sample averages

In statistics, we often have iid samples $Y_1, \dots, Y_n$. Our goal is to say something about the probability distribution $P_Y(y)$. For example,

- The sample mean **sample mean**

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i$$

- The probability of a given outcome:

$$\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} 1_A(Y_i) \approx P(\{Y \in A\})$$

In both cases we are using an **sample average**.

Expectation

- The sample average converges to an expectation

$$E[Y] = \sum_{y \in S} yP(\{Y = y\}) = \sum_{y \in S} y \frac{n_y}{n} = \frac{1}{n} \sum_{i=1}^{n} Y_i$$

  The function E takes a random variable and outputs a deterministic quantity.
- The sample average therefore converts are random dataset into a (approximately) deterministic number which can be used to deduce attributes of our probability model.

The central structure of classical statistics is

$$\underset{\text{(from data)}}{g(\bar{Y})} \approx E[f(Y)] \underset{\text{solving equations}}{\longrightarrow} \text{Model parameters}$$

Example

We are given n samples X_i of a Bernoulli random variable. If we want to estimate the parameter q, then

$$q = P(X = 1) = 0 \times P(X = 0) + 1 \times P(X = 1) = E[X] \tag{1}$$

Therefore

$$q \approx \bar{X}. \tag{2}$$

We say that $\bar{X}$ is an **estimator** of q and write $\hat{q}$. In general we write $\hat{\theta}$ for an estimator of some parameter θ.

Sample distribution

How “good” an approximation is this? Intuitively, it will become better when n becomes large, but how to we quantity