24 2025-03-27 | Week 10 | Lecture 24

24.1 Statistics: a general framework

Chapters 5.3 and 5.4

The objective of statistics to is to make an inference about a population based on information contained in a sample from that population and to provide an associated measure of goodness for the inference. Here we introduce a conceptual framework for mathematical statistics.

A ***population*** is a large body of data that is the target of our interest. The subset collected from it is our ***sample***.

Example 109 (Populations)**.** A population can be real or theoretical. Here are some examples

- The set of people in Hawai'i (real, finite)
- The set of voters in the 2026 Midterm elections (hypothetical, finite)
- The decimal expansion of π (countably infinite)
- The infinitely many observations that could be made during a laboratory experiment if the experiment were repeated over and over again (hypothetical)
- The lifetimes of light bulbs produced by a factory

Importantly, a population can also be a *probability distribution*, specified by a pdf, pmf, or cdf

- Observations made from an exponential distribution with rate $\lambda > 0$ (i.e., the distribution with pdf $f(x) = \lambda e^{-\lambda x}\mathbf{1}_{[x>0]}$.)

End of Example 109. □

More commonly in mathematical statistics, we think of the population as a distribution.

Definition 110 (Random sample from a distribution)**.** Consider a given probability distribution on $\mathbb{R}$ that can be represented by a pdf or pmf f. We say that the random variables $X_1, \ldots, X_n$ form a ***random sample*** from this distribution if these random variables are independent and the distribution of each is given by f. Such random variables are also said to be ***independent and identically distributed*** (iid). The number of random variables n is the ***sample size***.

24.2 What is a statistic?

Definition 111 (Statistic)**.** A ***statistic*** is any quantity whose value can be calculated from sample data.

Here are two important examples of statistics:

Example 112. Let $X_1, \ldots, X_n$ be random sample.

- The ***sample mean***, denoted $\overline{X}$, is
$$\overline{X} = \frac{X_1 + \ldots + X_n}{n}.$$
- The ***sample variance***, denoted S^2, is
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

End of Example 112. □

Because a statistic is a *function* of random variables, it is itself a random variable. The probability distribution of a statistic is called its ***sampling distribution***.

Theorem 113. *Let $X_1, \ldots, X_n$ be a random sample from a distribution with mean μ and standard deviation $\sigma > 0$. Then*

- $\mu_{\overline{X}} := \mathbb{E}\left[\overline{X}\right] = \mu$
- $\sigma^2_{\overline{X}} := \mathrm{Var}(\overline{X}) = \sigma^2/n$ *(and $\sigma_{\overline{X}} = \sigma/\sqrt{n}$)*

Proof. First we compute $\mathbb{E}\left[\overline{X}\right]$:

$$\begin{aligned}
\mathbb{E}\left[\overline{X}\right] &= \mathbb{E}\left[\frac{X_1 + \ldots + X_n}{n}\right] && \text{by definition of } \overline{X} \\
&= \frac{1}{n}\left(\mathbb{E}\left[X_1\right] + \ldots + \mathbb{E}\left[X_n\right]\right) && \text{by linearity of expectation} \\
&= \frac{1}{n}\left(n\mu\right) && \text{since } \mathbb{E}\left[X_i\right] = \mu \text{ for all } i \\
&= \mu.
\end{aligned}$$

Next we compute $\mathrm{Var}(\overline{X})$:

$$\begin{aligned}
\mathrm{Var}(\overline{X}) &= \mathrm{Var}\left(\frac{X_1 + \ldots + X_n}{n}\right) && \text{by defintion of } \overline{X} \\
&= \frac{1}{n^2}\mathrm{Var}\left(X_1 + \ldots + X_n\right) && \text{by Eq. (7)} \\
&= \frac{1}{n^2}\left(\mathrm{Var}(X_1) + \ldots + \mathrm{Var}(X_n)\right) && \text{by additivity (see Eq. (8))} \\
&= \frac{1}{n^2}\left(n\sigma^2\right) && \text{since } \mathrm{Var}(X_i) = \sigma^2 \text{ for all } i \\
&= \frac{\sigma^2}{n}.
\end{aligned}$$

□

Theorem 114. *Let $X_1, \ldots, X_n$ be a random sample from a normal distribution with mean μ and standard deviation σ. Then for any n, $\overline{X} \sim \mathcal{N}(\mu, \sigma^2/n)$.*