Day 9 - Probability III

Introduction

In this activity, we will introduce some key theorems in probability. By the end of this activity, you should understand:

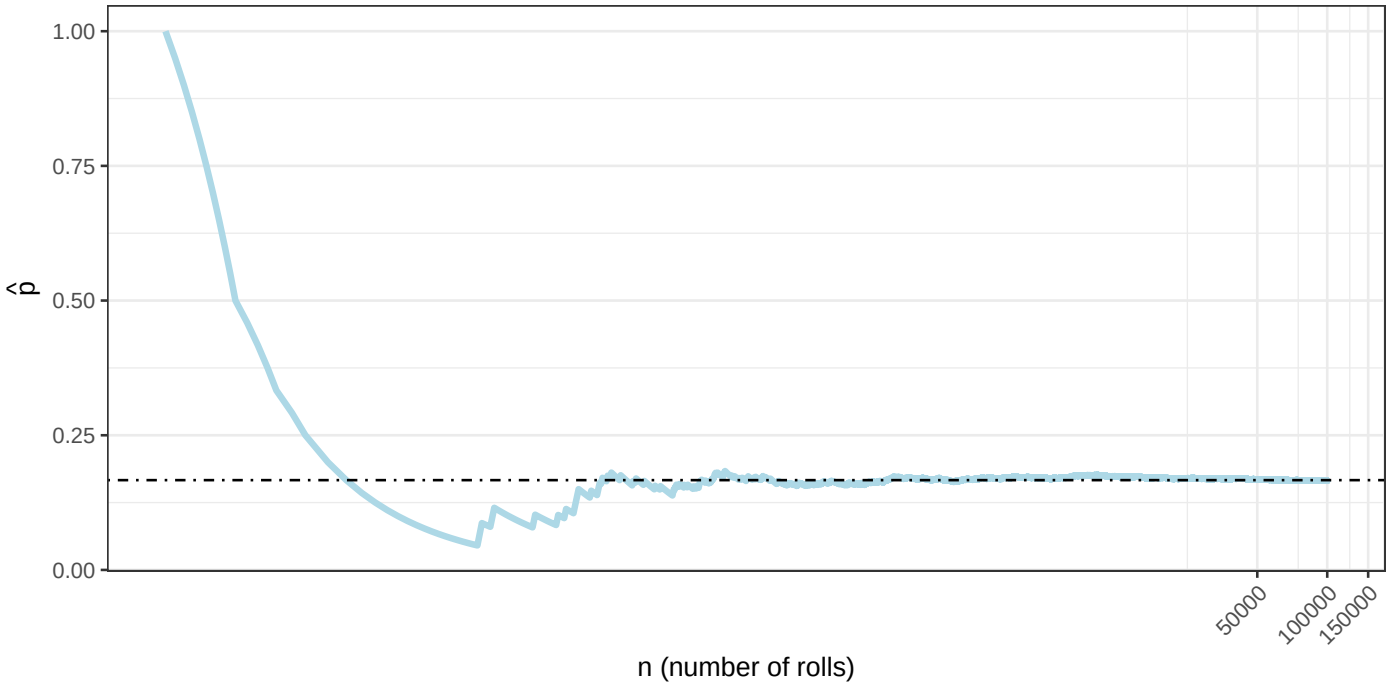
- the Law of Large Numbers
- Conditional probability
- Bayes' theorem
- partitions
- the Law of Total Probability
- collectively exhaustive events
- sensitivity
- specificity

Notable theorems in probability

Note

The _____ states that as more outcomes are observed, the observed proportion of events, $\hat{p}$, converges to the true probability of observing those events.

Example: Suppose we are interested in the probability of rolling a one on a fair die. The time plot below describes the observed proportion of rolling a one over 1,000,000 simulated rolls.



Note

_____ provides a way to calculate **conditional probability**.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Rearranging the above equation results in a new equation for joint probabilities.

$$P(A \cap B) = P(A|B)P(B)$$

Combining these two equations results in the following version of Bayes' theorem:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Note

The _____ states that an event A may be calculated as the summation of its intersection with a collection of n events, $\{B_n : n = 1, 2, 3, ...\}$, so long as $\{B_n\}$ **partitions** the sample space.

$$P(A) = \sum_n (A \cap B_n)$$

A collection of events $\{B_n\}$ **partitions** a sample space if they are mutually exclusive and **collectively exhaustive**, meaning that their union describes the entire sample space.

Example: Suppose a plot of land is occupied by three types of forest: 50% of the land is forest A, 30% of the land is forest B, and 20% of the land is forest C. Within each forest, there exist poisonous plants. Approximately 20%, 40%, and 70% of the plants are poisonous in forests A, B, and C, respectively. If we randomly enter this plot of land and pick a plant from the ground, what is the probability that it will be poisonous?

Tip

Let P define the event that a plant is poisonous, B_1 define the event that you enter forest A, B_2 define the event that you enter forest B, and B_3 define the event that you enter forest C.

Example: The **sensitivity** of a test, also called the true positive rate, is the probability that a test returns a positive result, *if* the patient actually has the disease. The **specificity** of a test, also called the true negative rate, is the probability that a test returns a negative result, *if* the patient actually does not have the disease.

Suppose that 1% of women have breast cancer, and that the sensitivity and specificity of a cancer screening procedure are 0.80 and 0.90, respectively. Suppose that a woman tests positive for breast cancer using this screening procedure. What is the probability that she actually has breast cancer?