Probability Basics

Serena An–November 4, 2018

1 Introduction

Consider a set of (finite) possible outcomes. If all outcomes are equally likely, then the probability that A occurs is

$$P(A) = \frac{\text{number of ways A can occur}}{\text{total number of outcomes}}.$$

Problem 1.1

If a fair coin is flipped once, what is the probability that it lands on heads?

Solution 1.1: There are two equally likely possibilities for the result of the coin flip: heads or tails. The probability is $\boxed{\dfrac{1}{2}}$.

Problem 1.2

If a fair coin is flipped twice, what is the probability that exactly one head is flipped?

Solution 1.2: There are four equally likely possibilities for two coin flips: HH, HT, TH, TT. Two of these, namely HT and TH, include exactly one head, so the probability is $\frac{2}{4} = \boxed{\dfrac{1}{2}}$.

Problem 1.3

Three coins are flipped and the number of heads is recorded. What is the probability that exactly one head is recorded?

Solution 1.3: There are four possibilities for the number of heads recorded: 0, 1, 2, 3. However, these outcomes are not equally likely, so the probability is not simply $\frac{1}{4}$. There are $2^3 = 8$ total possibilities for the outcomes of three coin flips. Out of these, three have exactly one head: HTT, THT, TTH. Thus, the probability is $\boxed{\dfrac{3}{8}}$.

2 Simpson's Paradox

Problem 2.1: Simpson's Paradox

(a) Consider a black box with 5 red and 6 green balls and a white box with 3 red and 4 green balls. From which of these two boxes is the probability of drawing a red ball greater?

(b) Consider a second black box with 6 red and 3 green balls and a second white box with 9 red and 5 green balls. From which of these two boxes is the probability of drawing a red ball greater?

(c) Consider adding the contents of the second black box to the first black box and adding the contents of the second white box to the first white box. Now, from which of these two boxes is the probability of drawing a red ball greater?

Solution 2.1: **(a)** The probabilities of drawing a red ball in the black and white boxes are $\frac{5}{11} = 0.455$ and $\frac{3}{7} = 0.429$, respectively, so the probability of drawing a red ball is greater in the black box.

(b) The probabilities of drawing a red ball in the second black and second white boxes are $\frac{2}{3} = 0.667$ and $\frac{9}{14} = 0.643$, respectively, so the probability of drawing a red ball is greater in the second black box.

(c) Combining, there are now 11 red and 9 green balls in the black box and 12 red and 9 green balls in the white box. Intuitively, the probability of drawing a red ball from the combined black box should be greater than the probability of drawing a red ball from the combined white box. However, the black and white boxes probabilities are now $\frac{11}{20} = 0.55$ and $\frac{12}{21} = 0.571$, respectively, and thus the probability of drawing a red ball is greater in the combined white box.

This is an example of **Simpson's paradox**. Simpson's paradox occurs when groups of data show one particular trend, but the trend is reversed when the groups of data are combined.

Why does Simpson's paradox occur? One would expect that larger probabilities in each case leads to larger probabilities overall. However, this is only true when the case sizes are equal. In the above example, there are 11 and 7 balls in the first boxes and 9 and 14 balls in the second boxes, which are somewhat large differences in the number of balls.

One real life example is batting averages. David Justice had a higher batting average than Derek Jeter in both 1995 and 1996, but across the two years, Derek Jeter's batting average was higher. Another example is University of California admission rates. A study showed that men were accepted at higher rates than women (44% to 35%), but in each department, women were usually accepted at a rate at least of the men. This was due to more women than men applying to lower admittance departments.