3.7 Functions of random variables
Unit 3: Discrete Random Variables

Adapted from Blitzstein-Hwang Chapter 3.

In this section we will discuss what it means to take a function of a random variable, and we will build understanding for why *a function of a random variable is a random variable*. That is, if X is a random variable, then X^2, e^X, and $\sin(X)$ are also random variables, as is $g(X)$ for any function $g : \mathbb{R} \to \mathbb{R}$.

To see how to define functions of an r.v. formally, let's rewind a bit. At the beginning of this chapter, we considered a random variable X defined on a sample space with 6 elements. Figure 3.1.1 used arrows to illustrate how X maps each pebble in the sample space to a real number, and the left half of Figure 3.1.4 showed how we can equivalently imagine X writing a real number inside each pebble. Now we can, if we want, apply the same function g to all the numbers inside the pebbles. Instead of the numbers $X(s_1)$ through $X(s_6)$, we now have the numbers $g(X(s_1))$ through $g(X(s_6))$, which gives a new mapping from sample outcomes to real numbers---we've created a new random variable, $g(X)$.

> **Definition 3.7.1 (Function of an r.v.).**
>
> For an experiment with sample space S, an r.v. X, and a function $g : \mathbb{R} \to \mathbb{R}$, $g(X)$ is the r.v. that maps s to $g(X(s))$ for all $s \in S$.

Taking $g(x) = \sqrt{x}$ for concreteness, Figure 3.7.2 represents $g(X)$ by directly labeling the sample outcomes. If X crystallizes to 4, then $g(X)$ crystallizes to 2.

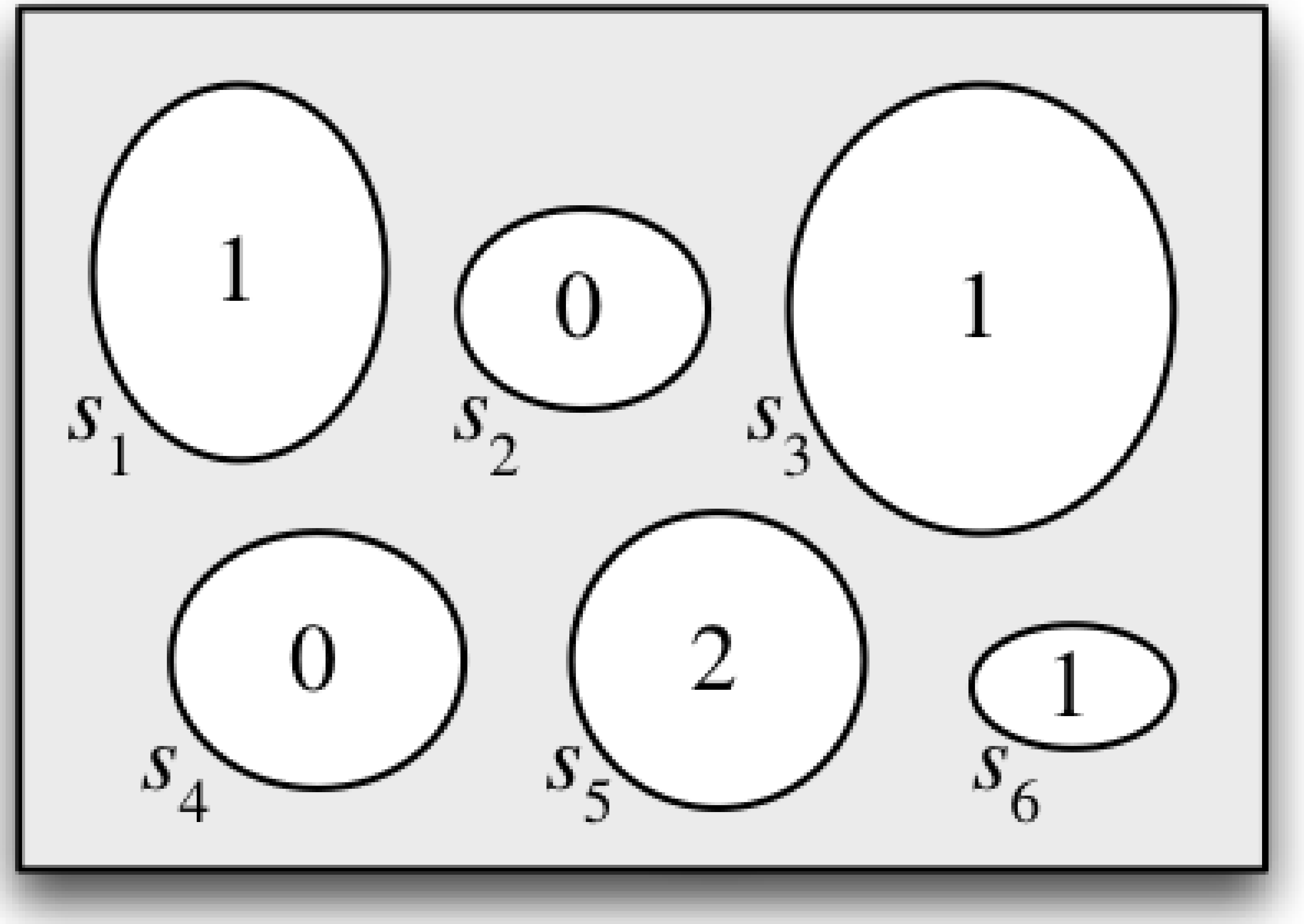

Figure 3.7.2: Since $g(X) = \sqrt{X}$ labels each pebble with a number, it is an r.v.

View Larger Image
Image Description

☣Warning 3.7.3 (Category errors).

Many common mistakes in probability can be traced to confusing two of the following fundamental objects with each other: distributions, random variables, events, and numbers. Such mistakes are examples of *category errors*. In general, a category error is a mistake that doesn't just happen to be wrong, but in fact is necessarily wrong since it is based on the wrong category of object.

For example, answering the question "How many people live in Boston?" with "-42" or "π" or "pink elephants" would be a category error - we may not know the population size of a city, but we do know that it is a nonnegative integer at any point in time. To help avoid being categorically wrong, always think about what category an answer should have.

An especially common category error is to confuse a random variable with its distribution. The following saying sheds light on the distinction between a random variable and its distribution:

''The word is not the thing; the map is not the territory.'' - Alfred Korzybski

We can think of the distribution of a random variable as a map or *blueprint* describing the r.v. Just as different houses can share the same blueprint, different r.v.s can have the same distribution, even if the *experiments* they summarize, and the *sample spaces* they map from, are not the same.