Probability and Integration

1. Simple probability: discrete variables

Dice provide a striaghtforward introduction to the basic concepts of probability. Tossing two dice, the number of spots that come up lies between 2 ans 12. We answer these questions.

What is the probability of getting 7?

Wat is the probability of getting 6 or 7?

What is the probability of getting 2, 3, or 4?

Assuming the dice are fair, each of the six outcomes of each die is equally likely, so each of the 36 combinations occurs with probability $\frac{1}{36}$. To find the probability of getting a certain outcome, count the number of combinations giving that outcome and multiply by $\frac{1}{36}$, the probability of each outcome.

For example, a 7 can be obtained from exactly these combinations

$$(\text{first die, second die}) = (1,6), (2,5), (3,4), (4,3), (5,2), (6,1)$$

that is, 6 combinations give an outcome of 7, so the probability of getting 7 is $\frac{6}{36} = \frac{1}{6}$. Letting the variable X denote the outcome of tossing two dice, we write $P(X = 7) = \frac{1}{6}$.

Another way to view this is to plot the possible outcomes, 2 through 12, along the x-axis, and above each outcome list the combinations giving that outcome. See the left side of Fig. 1. Those combinations giving a 7 are shaded on the right side of Fig. 1. The probability of getting 7 is just the number of shaded boxes divided by the total number of boxes; equivalently, *the fraction of the area occupied by those outcomes corresponding to 7.*

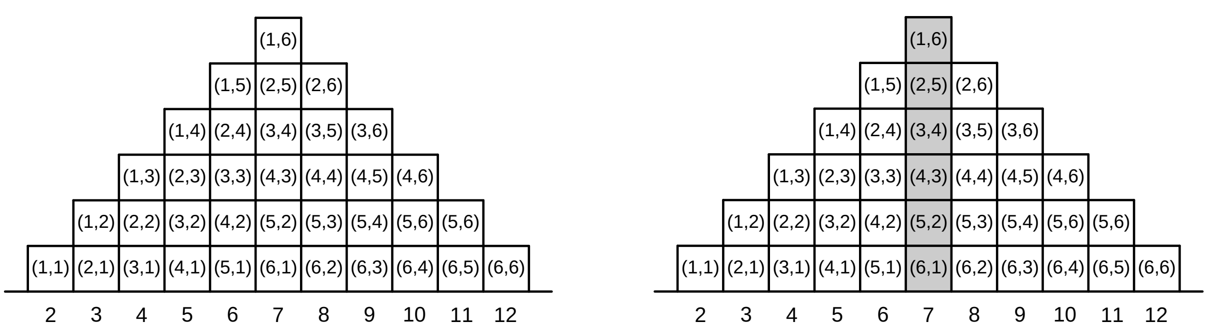

Fig. 1 Left: Dice toss combinations arranged by outcome. Right: those combinations giving a 7.

Continuing in this way, we see $P(6 \leq X \leq 7)$ and $P(X \leq 4)$ are the fractions of the area shaded in the left and right sides of Fig. 2.

The probabilities can be found by calculating what portion of the total area is shaded. We find $P(6 \leq X \leq 7) = \frac{11}{36}$ and $P(2 \leq X \leq 4) = \frac{1}{6}$.

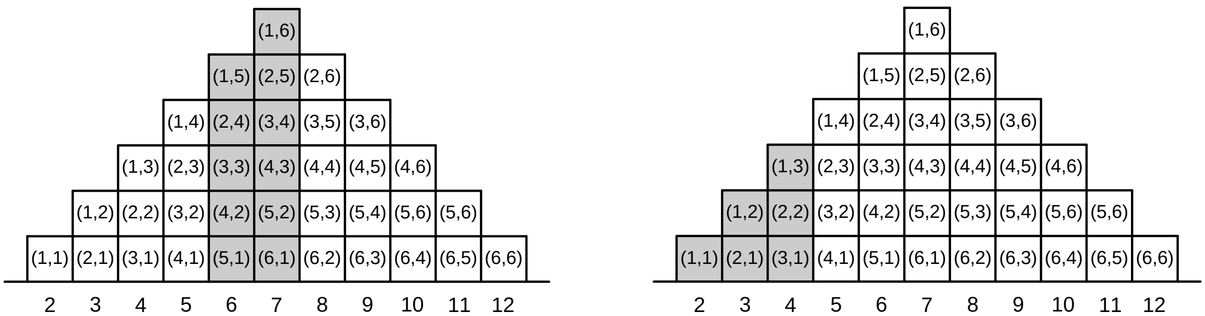

Fig. 2 Shading other combinations.

2. Continuous variables

Not all experiments have only a finite collection of possible outcomes. Measuring the heights of people is an example, if we assume lengths are continuously distributed. Physical measurements cannot be done with infinite accuracy; perhaps we measure heights to the nearest cm, or to the nearest mm, or to the nearest 0.1 mm. Any of these would give a graph looking like Fig. 1, and also the areas of regions composed of these rectangles suggest a Riemann sum approximation to an integral. The function f approximated by these Riemann sums is called a *probability density function*, and

$$P(a \leq X \leq b) = \int_a^b f(x)dx$$

The function $f(x)$ must satisfy two conditions:

$$(1)\ f(x) \geq 0 \text{ for all } x$$

and

$$(2) \int_{-\infty}^{\infty} f(x)dx = 1$$

where $\int_{-\infty}^{\infty} f(x)dx$ is the limit $\lim_{R\to-\infty} \int_R^0 f(x)dx + \lim_{S\to\infty} \int_0^S f(x)dx$.

For example, the function

$$f(x) = \begin{cases} 0 & \text{for } x < 0 \\ 6x(1-x) & \text{for } 0 \leq x \leq 1 \\ 0 & \text{for } 1 < x \end{cases}$$

pictured on the left of Fig. 3, is a probability density function.

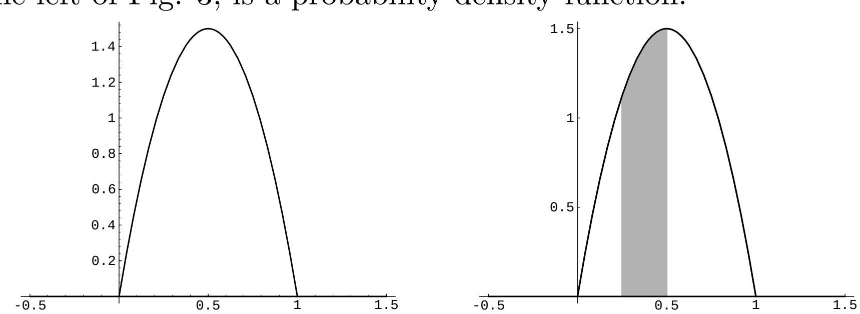

Fig. 3 A continuous probability distribution.

Then, for example, $P(.25 \leq X \leq .5) = \int_{.25}^{.5} 6x(1-x)dx = (3x^2 - 2x^3)|_{.25}^{.5} = 0.34375$, the area of the shaded region on the right side of Fig. 3.

3. Expected values and variances

Returning to the dice toss example, the average, or *expected value*, of the experiment, over many repetitions, can be deduced from the left side of Fig. 1. The value 2 occurs in about $\frac{1}{36}$ repetitions, the value 3 in about 2 36, the value 4 in about $\frac{3}{36}$, ..., the value 12 in about $\frac{1}{36}$ repetitions. Consequently, the expected value is

$$E(X) = 2 \cdot \tfrac{1}{36} + 3 \cdot \tfrac{2}{36} + 4 \cdot \tfrac{3}{36} + \cdots + 12 \cdot \tfrac{1}{36} = 7$$

Note the expected value need never occur. For example, tossing three (fair) coins yields eight equally likely outcomes:

TTT, TTH, THT, THH, HTT, HTH, HHT, HHH

Then $P(X = 0) = \frac{1}{8}$, $P(X = 1) = \frac{3}{8}$, $P(X = 2) = \frac{3}{8}$, and $P(X = 3) = \frac{1}{8}$; the expected value is

$$0 \cdot \tfrac{1}{8} + 1 \cdot \tfrac{3}{8} + 2 \cdot \tfrac{3}{8} + 3 \cdot \tfrac{1}{8} = \tfrac{3}{2}.$$

For continuous distributions the analogous computation is

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

For $f(x) = 6x(1 - x)$, the expected value is

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_{0}^{1} x f(x) dx = \int_{0}^{1} 6x^2(1 - x) dx = 2x^3 - \frac{3}{2}x^4 = \frac{1}{2}.$$

For any positive integer r, the r^{th} *moment* of X is $E(X^r)$, so the first moment is just the expected value. The second moment has an important interpretation. First, if $E(X)$ is finite, then translate the variables by $E(X)$. That is, consider the new variables $X^0 = X - E(X)$. The second moment $E((X^0)^2)$ is the *variance*, $\sigma^2(X)$, of X; the *standard deviation* of X is $\sigma(X) = \sqrt{\sigma^2(X)}$. An interesting formulation is

$$\sigma^2(X) = E(X^2) - E(X)^2$$

For a sample $\{x_1, x_2, \ldots, x_N\}$ of measurements, we can compute the *sample mean* and *sample standard deviation.* These are

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_N}{N}$$

$$\sqrt{\frac{\sum_{i=1}^{N}(x_i - \overline{x})^2}{N - 1}}$$

A natural question is how well these sample quantities approximate those of the population.

In the next section we interpret the standard deviation for a familiar probability distribution.

4. The normal distribution

Normal distributions are probability distributions characterized by two numbers, μ and σ, and have density function

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

The numbers μ and σ are the mean and standard deviation of the normal distribution. The left side of Fig. 4 shows plots of normal distributions with $mu = 0$ and $\sigma = 1, 2$. The middle picture shows the $\sigma = 1$ normal distribution, shading the area one standard deviation on either side of the mean; the right picture shades two standard deviations on either side of the mean. How are we to interpret these?

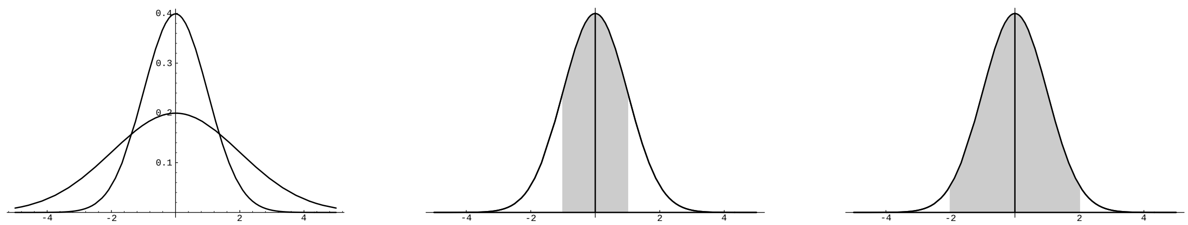

Fig. 4 Some aspects of some normal distributions

The left side of Fig. 4 illustrates how σ characterizes the spread of the normal density function. The smaller σ, the more tightly the density function is concentrated about the mean.

The normal density function has no antiderivative, but numerical integration gives $\int_{-\sigma}^{\sigma} f(x)dx \approx 0.6827$ and $\int_{-2\sigma}^{2\sigma} f(x)dx \approx 0.9545$. These results exemplify the rarity of events far from the mean. For instance, the probability of observing a value more than 10σ from the mean is about 10^{-24}. Some financial data, for example, exhibit 10σ events every few years, providing strong evidence that these do not follow the normal distribution.