Outline

Probability Histograms
Probability Histograms

Consider the box

\[ 1 \ 1 \ 1 \ 1 \ 3 \ 4 \ 4 \]
Then, the chances of obtaining a ticket with a 1 are $4/7$, the chances of a 3 are $1/7$ and the chances of a 4 are $2/7$. We can display that information graphically in a probability histogram.
Each box is centered at a number and its area corresponds to the probability of that number.

The sum of the areas of the boxes is equal to one. This is because the areas are associated with probabilities or chances.
Probability histograms

Probability histograms are used to represent chance and not the frequency of data.

Histograms based on sampled data are used to represent how the data are distributed over their range. Probability histograms correspond to the chances that a random variable take some specific values.
Empirical histograms based on the frequencies of observed outcomes of an experiment converge to the corresponding probability histograms, as can be seen by the example of rolling two dice.
In the previous example consider taking the product of the two dice.

The convergence is also true when considering the product of the two dice. In this case we notice that the probability histogram is much more irregular than the one obtained for the sum.
The regularity is a general feature related to the sum.
Consider the problem of tossing a fair coin a certain number of times $n$. We can obtain the probability histogram of the number of heads for each $n$.

We observe that the probability histogram of the number of heads converges to a very regular curve as the number of tosses is increased.

This curve is the well known probability density named **Gaussian** curve.
The Sum of Draws

We can approximate the probability histogram of the sum of heads in a large number of coin tosses using the normal curve.

Q: A coin is tossed 100 times, what is the probability of getting exactly 50 heads?

A: We can look at the probability histogram for this case. We observe that the chances corresponding to 50 are equal to the area of the box that has a base from 49.5 to 50.5. The area of this box is 7.96%.
Q: What about an approximation using the normal curve?

A: First step is to calculate the mean and standard deviation. Consider a box model where there is a zero for the tail and 1 for the head,

\[ \begin{array}{c}
0 \\
1 
\end{array} \]
Average of the Box: $\frac{1}{2}$. SD of the Box: $\frac{1}{2}$

When drawing a ticket from this box 100 times with replacement, the expected value of the sum of the draws is

$$100 \times \frac{1}{2} = 50$$

**Expected Value**

In general, the expected value of the sum of the draws is given by

$$(\text{number of draws}) \times (\text{average of box})$$