

19 Probability Distributions

Although individual events based on probability are unpredictable, we can predict patterns when we repeat the experiment many times. Today, we will look at the pattern that emerges from independent random variables, such as flipping a coin.

Coin flipping. Suppose we have a fair coin, that is, the probability of getting head is precisely one half and the same is true for getting tail. Let X count the times we get head. If we flip the coin n times, the probability that we get k heads is

$$P(X = k) = \binom{n}{k} / 2^n.$$

Figure 22 visualizes this distribution in the form of a histogram for $n = 10$. Recall that the *distribution function* maps every possible outcome to its probability, $f(k) = P(X = k)$. This makes sense when we have a discrete domain. For a continuous domain, we consider the *cumulative distribution function* that gives the probability of the outcome to be within a particular range, that is, $\int_{x=a}^b f(x) dx = P(a \leq X \leq b)$.

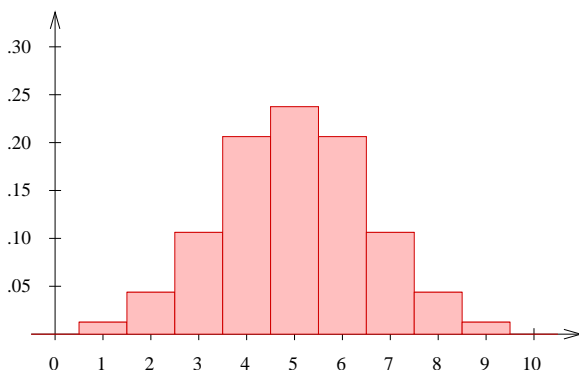


Figure 22: The histogram that shows the probability of getting 0, 1, ..., 10 heads when flipping a coin ten times.

Variance. Now that we have an idea of what a distribution function looks like, we wish to find succinct ways of describing it. First, we note that $\mu = E(X)$ is the expected value of our random variable. It is also referred to as the *mean* or the *average* of the distribution. In the example above, where X is the number of heads in ten coin flips, we have $\mu = 5$. However, we would not be surprised if we had four or six heads but we might be surprised if we had

zero or ten heads when we flip a coin ten times. To express how surprised we should be we measure the spread of the distribution. Let us first determine how close we expect a random variable to be to its expectation, $E(X - E(X))$. By linearity of expectation, we have

$$E(X - \mu) = E(X) - E(\mu) = \mu - \mu = 0.$$

Hence, this measurement is not a good description of the distribution. Instead, we use the expectation of the square of the difference to the mean. Specifically, the *variance* of a random variable X , denoted as $V(X)$, is the expectation $E((X - \mu)^2)$. The *standard deviation* is the square root of the variance, that is, $\sigma(X) = V(X)^{1/2}$. If X_4 is the number of heads we see in four coin flips, then $\mu = 2$ and

$$V(X_4) = \frac{1}{16} [(-2)^2 + 4 \cdot (-1)^2 + 4 \cdot 1^2 + 2^2],$$

which is equal to 1. For comparison, let X_1 be the number of heads that we see in one coin flip. Then $\mu = \frac{1}{2}$ and

$$V(X_1) = \frac{1}{2} \left[\left(0 - \frac{1}{2}\right)^2 + \left(1 - \frac{1}{2}\right)^2 \right],$$

which is equal to one quarter. Here, we notice that the variance of four flips is the sum of the variances for four individual flips. However, this property does not hold in general.

Variance for independent random variables. Let X and Y be independent random variables. Then, the property that we observed above is true.

ADDITIVITY OF VARIANCE. If X and Y are independent random variables then $V(X + Y) = V(X) + V(Y)$.

We first prove the following more technical result.

LEMMA. If X and Y are independent random variables then $E(XY) = E(X)E(Y)$.

PROOF. By definition of expectation, $E(X)E(Y)$ is the product of $\sum_i x_i P(X = x_i)$ and $\sum_j y_j P(Y = y_j)$. Pushing the summations to the right, we get

$$\begin{aligned} E(X)E(Y) &= \sum_i \sum_j x_i y_j P(X = x_i) P(Y = y_j) \\ &= \sum_{i,j} z_{ij} P(X = x_i) P(Y = y_j), \end{aligned}$$

where $z_{ij} = x_i y_j$. Finally, we use the independence of the random variables X and Y to see that $P(X = x_i)P(Y = y_j) = P(XY = z_{ij})$. With this, we conclude that $E(X)E(Y) = E(XY)$. \square

Now, we are ready to prove the Additivity of Variance, that is, $V(X + Y) = V(X) + V(Y)$ whenever X and Y are independent.

PROOF. By definition of variance, we have

$$V(X + Y) = E((X + Y - E(X + Y))^2).$$

The right hand side is the expectation of $(X - \mu_X)^2 + 2(X - \mu_X)(Y - \mu_Y) + (Y - \mu_Y)^2$, where μ_X and μ_Y are the expected values of the two random variables. With this, we get

$$\begin{aligned} V(X + Y) &= E((X - \mu_X)^2) + E((Y - \mu_Y)^2) \\ &= V(X) + V(Y), \end{aligned}$$

as claimed. \square

Normal distribution. If we continue to increase the number of coin flips, then the distribution function approaches the *normal distribution*,

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

This is the limit of the distribution as the number of coin flips approaches infinity. For a large number of trials, the

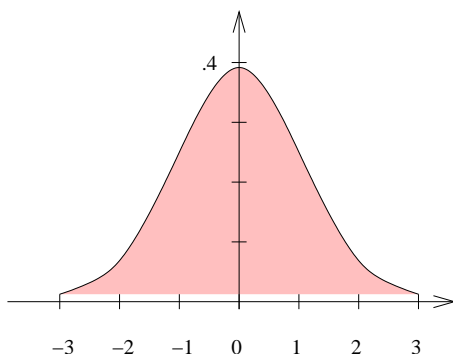


Figure 23: The normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. The probability that the random variable is between $-\sigma$ and σ is 0.68, between -2σ and 2σ is 0.955, and between -3σ and 3σ is 0.997.

normal distribution can be used to approximate the probability of the sum being between a and b standard deviations from the expected value.

STANDARD LIMIT THEOREM. The probability of the number of heads being between $a\sigma$ and $b\sigma$ from the mean goes to

$$\frac{1}{\sqrt{2\pi}} \int_{x=a}^b e^{-\frac{x^2}{2}} dx$$

as the number of flips goes to infinity.

For example, if we have 100 coin flips, then $\mu = 50$, $V(X) = 25$, and $\sigma = 5$. It follows that the probability of having between 45 and 55 heads is about 0.68.

Summary. We used a histogram to visualize the probability that we will have k heads in n flips of a coin. We also used the mean, μ , the standard deviation, σ , and the variance, $V(X)$, to describe the distribution of outcomes. As n approaches infinity, we see that this distribution approaches the normal distribution.