17 Random Variables

A *random variable* is a real-value function on the sample space, $X : \Omega \to \mathbb{R}$. An example is the total number of dots at rolling two dice, or the number of heads in a sequence of ten coin flips.

Bernoulli trial process. Recall that an independent trial process is a sequence of identical experiments in which the outcome of each experiment is independent of the preceding outcomes. A particular example is the *Bernoulli trial process* in which the probability of success is the same at each trial:

$$\begin{aligned} P(\text{success}) &= p; \\ P(\text{failure}) &= 1 - p. \end{aligned}$$

If we do a sequence of n trials, we may define X equal to the number of successes. Hence, Ω is the space of possible outcomes for a sequence of n trials or, equivalently, the set of binary strings of length n. What is the probability of getting exactly k successes? By the Independent Trial Theorem, the probability of having a sequence of k successes followed by $n-k$ failures is $p^k(1-p)^{n-k}$. Now we just have to multiply with the number of binary sequences that contain k successes.

BINOMIAL PROBABILITY LEMMA. The probability of having exactly k successes in a sequence of n trials is $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$.

As a sanity check, we make sure that the probabilities add up to one. Using the Binomial Theorem, get

$$\sum_{k=0}^{n} P(X = k) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k},$$

which is equal to $(p + (1-p))^n = 1$. Because of this connection, the probabilities in the Bernoulli trial process are called the *binomial probabilities*.

Expectation. The function that assigns to each $x_i \in \mathbb{R}$ the probability that $X = x_i$ is the *distribution function* of X, denoted as $f : \mathbb{R} \to [0,1]$; see Figure 20. More formally, $f(x_i) = P(A)$, where $A = X^{-1}(x_i)$. The *expected value* of the random variable is $E(X) = \sum_i x_i P(X = x_i)$.

As an example, consider the Bernoulli trial process in which X counts the successes in a sequence of n trials,

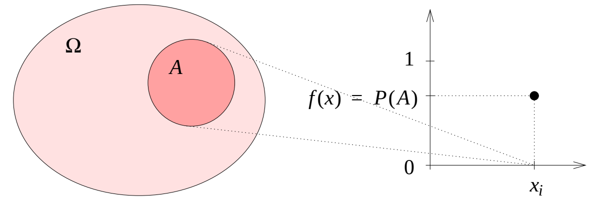

Figure 20: The distribution function of a random variable is constructed by mapping a real number, x_i, to the probability of the event that the random variable takes on the value x_i.

that is, $P(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. The corresponding distribution function maps k to the probability of having k successes, that is, $f(k) = \binom{n}{k} p^k (1-p)^{n-k}$. We get the expected number of successes by summing over all k.

$$\begin{aligned} E(X) &= \sum_{k=0}^{n} k f(k) \\ &= \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} \\ &= np \sum_{k=1}^{n} \binom{n-1}{k-1} p^{k-1} (1-p)^{n-k} \\ &= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{n-k-1}. \end{aligned}$$

The sum in the last line is equal to $(p + (1-p))^{n-1} = 1$. Hence, the expected number of successes is $X = np$.

Linearity of expectation. Note that the expected value of X can also be obtained by summing over all possible outcomes, that is,

$$E(X) = \sum_{s \in \Omega} X(s) P(s).$$

This leads to an easier way of computing the expected value. To this end, we exploit the following important property of expectations.

LINEARITY OF EXPECTATION. Let $X, Y : \Omega \to \mathbb{R}$ be two random variables. Then

(i) $E(X + Y) = E(X) + E(Y)$;
(ii) $E(cX) = cE(X)$, for every real number c.

The proof should be obvious. Is it? We use the property to recompute the expected number of successes in a Bernoulli trial process. For i from 1 to n, let X_i be the expected number of successes in the i-th trial. Since there is only one i-th trial, this is the same as the probability of having a success, that is, $E(X_i) = p$. Furthermore, $X = X_1 + X_2 + \ldots + X_n$. Repeated application of property (i) of the Linearity of Expectation gives $E(X) = \sum_{i=1}^{n} E(X_i) = np$, same as before.

Indication. The Linearity of Expectation does not depend on the independence of the trials; it is also true if X and Y are dependent. We illustrate this property by going back to our hat checking experiment. First, we introduce a definition. Given an event, the corresponding *indicator random variable* is 1 if the event happens and 0 otherwise. Thus, $E(X) = P(X = 1)$.

In the hat checking experiment, we return n hats in a random order. Let X be the number of correctly returned hats. We proved that the probability of returning at least one hat correctly is $P(X \geq 1) = 1 - e^{-1} = 0.6\ldots$ To compute the expectation from the definition, we would have to determine the probability of returning exactly k hats corrects, for each $0 \leq k \leq n$. Alternatively, we can compute the expectation by decomposing the random variable, $X = X_1 + X_2 + \ldots + X_n$, where X_i is the expected value that the i-th hat is returned correctly. Now, X_i is an indicator variable with $E(X_i) = \frac{1}{n}$. Note that the X_i are not independent. For example, if the first $n - 1$ hats are returned correctly then so is the n-th hat. In spite of the dependence, we have

$$E(X) = \sum_{i=1}^{n} E(X_i) = 1.$$

In words, the expected number of correctly returned hats is one.

Example: computing the minimum. Consider the following algorithm for computing the minimum among n items stored in a linear array.

```
min = A[1];
for i = 2 to n do
  if min > A[i] then min = A[i] endif
endif.
```

Suppose the items are distinct and the array stores them in a random sequence. By this we mean that each permutation of the n items is equally likely. Let X be the number of assignments to min. We have $X = X_1 + X_2 + \ldots + X_n$, where X_i is the expected number of assignments in the i-th step. We get $X_i = 1$ iff the i-th item, $A[i]$, is smaller than all preceding items. The chance for this to happen is one in i. Hence,

$$\begin{aligned} E(X) &= \sum_{i=1}^{n} E(X_i) \\ &= \sum_{i=1}^{n} \frac{1}{i}. \end{aligned}$$

The result of this sum is referred to as the n-*th harmonic number*, $H_n = 1 + \frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n}$. We can use $\int_{x=1}^{n} = \ln n$ to show that the n-th harmonic number is approximately the natural logarithm of n. More precisely, $\ln(n+1) \leq H_n \leq 1 + \ln n$.

Waiting for success. Suppose we have again a Bernoulli trial process, but this time we end it the first time we hit a success. Defining X equal to the index of the first success, we are interested in the expected value, $E(X)$. We have $P(X = i) = (1-p)^{i-1} p$ for each i. As a sanity check, we make sure that the probabilities add up to one. Indeed,

$$\begin{aligned} \sum_{i=1}^{\infty} P(X = i) &= \sum_{i=1}^{\infty} (1-p)^{i-1} p \\ &= p \cdot \frac{1}{1 - (1-p)}. \end{aligned}$$

Using the Linearity of Expectation, we get a similar sum for the expected number of trials. First, we note that $\sum_{j=0}^{\infty} j x^j = \frac{x}{(1-x)^2}$. There are many ways to derive this equation, for example, by index transformation. Hence,

$$\begin{aligned} E(X) &= \sum_{i=0}^{\infty} i P(X = i) \\ &= \frac{p}{1-p} \sum_{i=0}^{\infty} i (1-p)^i \\ &= \frac{p}{1-p} \cdot \frac{1-p}{(1-(1-p))^2}, \end{aligned}$$

which is equal to $\frac{1}{p}$.

Summary. Today, we have learned about random variable and their expected values. Very importantly, the expectation of a sum of random variables is equal to the sum of the expectations. We used this to analyze the Bernoulli trial process.