Probability Distributions

Tossing a coin is an experiment with exactly two outcomes: heads (“success”) with a probability of, say \( p \), and tails (“failure”) with a probability of \( 1 - p \). Such an experiment is called a Bernoulli trial. Let \( Y \) be a random variable that is 1 if the experiment succeeds and is 0 otherwise. \( Y \) is called a Bernoulli or an indicator random variable. For such a variable we have

\[
E[Y] = p \cdot 1 + (1 - p) \cdot 0 = p = \Pr[Y = 1]
\]

Thus for a fair coin if we consider heads as "success" then the expected value of the corresponding indicator random variable is \( 1/2 \).

A sequence of Bernoulli trials means that the trials are independent and each has a probability \( p \) of success. We will study two important distributions that arise from Bernoulli trials: the geometric distribution and the binomial distribution.

The Geometric Distribution

Consider the following question. Suppose we have a biased coin with heads probability \( p \) that we flip repeatedly until it lands on heads. What is the distribution of the number of flips? This is an example of a geometric distribution. It arises in situations where we perform a sequence of independent trials until the first success where each trial succeeds with a probability \( p \).

Note that the sample space \( \Omega \) consists of all sequences that end in \( H \) and have exactly one \( H \). That is

\[
\Omega = \{H, TH, TTH, TTTTH, \ldots\}
\]

For any \( \omega \in \Omega \) of length \( i \), \( \Pr[\omega] = (1 - p)^{i-1}p \).

**Definition.** A geometric random variable \( X \) with parameter \( p \) is given by the following distribution for \( i = 1, 2, \ldots \):

\[
\Pr[X = i] = (1 - p)^{i-1}p
\]

We can verify that the geometric random variable admits a valid probability distribution as follows:

\[
\sum_{i=1}^{\infty} (1 - p)^{i-1}p = p \sum_{i=1}^{\infty} (1 - p)^{i-1} = \frac{p}{1 - p} \sum_{i=1}^{\infty} (1 - p)^i = \frac{p}{1 - p} \cdot \frac{1 - p}{1 - (1 - p)} = 1
\]

Note that to obtain the second-last term we have used the fact that \( \sum_{i=1}^{\infty} c^i = \frac{c}{1 - c} \), \( |c| < 1 \).
Let’s now calculate the expectation of a geometric random variable, $X$. We can do this in several ways. One way is to use the definition of expectation.

\[
E[X] = \sum_{i=0}^{\infty} i \Pr[X = i]
\]

\[
= \sum_{i=0}^{\infty} i(1-p)^{i-1}p
\]

\[
= \frac{p}{1-p} \sum_{i=0}^{\infty} i(1-p)^i
\]

\[
= \left( \frac{p}{1-p} \right) \left( \frac{1-p}{(1-(1-p))^2} \right)
\]

\[
= \left( \frac{p}{1-p} \right) \left( \frac{1-p}{p^2} \right)
\]

\[
= \frac{1}{p}
\]

Another way to compute the expectation is to note that $X$ is a random variable that takes on non-negative values. From a theorem proved in last class we know that if $X$ takes on only non-negative values then

\[
E[X] = \sum_{i=1}^{\infty} \Pr[X \geq i]
\]

Using this result we can calculate the expectation of the geometric random variable $X$. For the geometric random variable $X$ with parameter $p$,

\[
\Pr[X \geq i] = \sum_{j=i}^{\infty} (1-p)^{j-1}p = (1-p)^{i-1}p \sum_{j=0}^{\infty} (1-p)^j = (1-p)^{i-1}p \times \frac{1}{1-(1-p)} = (1-p)^{i-1}
\]

Therefore

\[
E[X] = \sum_{i=1}^{\infty} \Pr[X \geq i] = \sum_{i=1}^{\infty} (1-p)^{i-1} = \frac{1}{1-p} \sum_{i=1}^{\infty} (1-p)^i = \frac{1}{1-p} \cdot \frac{1-p}{1-(1-p)} = \frac{1}{p}
\]

**Memoryless Property.** For a geometric random variable $X$ with parameter $p$ and for $n > 0$,

\[
\Pr[X = n + k \mid X > k] = \Pr[X = n]
\]

**Conditional Expectation.** The following is the definition of conditional expectation.

\[
E[Y \mid Z = z] = \sum_y y \Pr[Y = y \mid Z = z],
\]

where the summation is over all possible values $y$ that the random variable $Y$ can assume.
Example. For any random variables $X$ and $Y$,

$$E[X] = \sum_y \Pr[Y = y]E[X | Y = y]$$

We can also calculate the expectation of a geometric random variable $X$ using the memoryless property of the geometric random variable. Let $Y$ be a random variable that is 0, if the first flip results in tails and that is 1, if the first flip is a heads. Using conditional expectation we have

$$E[X] = \Pr[Y = 0]E[X | Y = 0] + \Pr[Y = 1]E[X | Y = 1]$$

$$= (1 - p)(E[X] + 1) + p \cdot 1 \quad \text{(using the memoryless property)}$$

$$\therefore pE[X] = 1$$

$$E[X] = \frac{1}{p}$$

**Binomial Distributions**

Consider an experiment in which we perform a sequence of $n$ coin flips in which the probability of obtaining heads is $p$. How many flips result in heads?

If $X$ denotes the number of heads that appear then

$$\Pr[X = j] = \binom{n}{j} p^j (1 - p)^{n-j}$$

**Definition.** A *binomial* random variable $X$ with parameters $n$ and $p$ is defined by the following probability distribution on $j = 0, 1, 2, \ldots, n$:

$$\Pr[X = j] = \binom{n}{j} p^j (1 - p)^{n-j}$$

We can verify that the above is a valid probability distribution using the binomial theorem as follows

$$\sum_{j=1}^{n} \binom{n}{j} p^j (1 - p)^{n-j} = (p + (1 - p))^n = 1$$

What is the expectation of a binomial random variable $X$? We can calculate $E[X]$ is two
ways. We first calculate it directly from the definition.

\[
E[X] = \sum_{j=0}^{n} j \binom{n}{j} p^j (1-p)^{n-j}
\]

\[
= \sum_{j=0}^{n} j \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j}
\]

\[
= \sum_{j=1}^{n} \frac{n!}{j!(n-j)!} p^j (1-p)^{n-j}
\]

\[
= \sum_{j=1}^{n} \frac{(j-1)!}{(n-j)!} p^j (1-p)^{n-j}
\]

\[
= np \sum_{j=1}^{n} \frac{(n-1)!}{(j-1)!((n-1)-(j-1))!} p^{j-1}(1-p)^{(n-1)-(j-1)}
\]

\[
= np \sum_{k=0}^{n-1} \frac{(n-1)!}{k!(n-1-k)!} p^k (1-p)^{(n-1)-k}
\]

\[
= np \sum_{k=0}^{n-1} \binom{n-1}{k} p^k (1-p)^{(n-1)-k}
\]

\[
= np
\]

The last equation follows from the binomial expansion of \((p + (1-p))^{n-1}\).

We can obtain the result in a much simpler way by using the linearity of expectation. Let \(X_i, 1 \leq i \leq n\) be the indicator random variable that is 1 if the \(i\)th flip results in heads and is 0 otherwise. We have \(X = \sum_{i=1}^{n} X_i\). By the linearity of expectation we have

\[
E[X] = \sum_{i=1}^{n} E[X_i] = \sum_{i=1}^{n} p = np
\]

What is the variance of the binomial random variable \(X\)? Since \(X = \sum_{i=1}^{n} X_i\), and \(X_1, X_2, \ldots, X_n\) are independent we have

\[
\text{Var}[X] = \sum_{i=1}^{n} \text{Var}[X_i]
\]

\[
= \sum_{i=1}^{n} E[X^2_i] - E[X_i]^2
\]

\[
= n(p - p^2)
\]

\[
= np(1 - p)
\]
Coupon Collector’s Problem.

We are trying to collect \( n \) different coupons that can be obtained by buying cereal boxes. The objective is to collect at least one coupon of each of the \( n \) types. Assume that each cereal box contains exactly one coupon and any of the \( n \) coupons is equally likely to occur. How many cereal boxes do we expect to buy to collect at least one coupon of each type?

**Solution.** Let the random variable \( X \) denote the number of cereal boxes bought until we have at least one coupon of each type. We want to compute \( E[X] \). Let \( X_i \) be the random variable denoting the number of boxes bought to get the \( i \)th new coupon. Clearly,

\[
X = X_1 + X_2 + X_3 + \ldots + X_n
\]

Using the linearity of expectation we have

\[
E[X] = E[X_1] + E[X_2] + E[X_3] + \ldots + E[X_n] \quad (1)
\]

What is the distribution of random variable \( X_i \)? Observe that the probability of obtaining the \( i \)th new coupon is given by

\[
p_i = \frac{n - (i - 1)}{n} = \frac{n - i + 1}{n}
\]

Thus the random variable \( X_i, 1 \leq i \leq n \) is a geometric random variable with parameter \( p_i \).

\[
E[X_i] = \frac{1}{p_i} = \frac{n}{n - i + 1}
\]

Combining this with equation (1) we get

\[
E[X] = \frac{n}{n} + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{2} + \frac{n}{1} = n \sum_{i=1}^{n} \frac{1}{i}
\]

The summation \( \sum_{i=1}^{n} \frac{1}{i} \) is known as the harmonic number \( H(n) \) and \( H(n) = \ln n + c \), for some constant \( c < 1 \).

Hence the expected number of boxes needed to collect \( n \) coupons is about \( nH(n) < n(\ln n + 1) \).

Relations

A binary relation is a set of ordered pairs. For example, let \( R = \{(1, 2), (2, 3), (5, 4)\} \). Then since \((1, 2) \in R\), we say that 1 is related to 2 by relation \( R \). We denote this by \( 1 \, R \, 2 \). Similarly, since \((4, 7) \notin R\), 4 is not related to 7 by relation \( R \), denoted by \( 4 \not\, R \, 7 \).

A binary relation \( R \) from set \( A \) to set \( B \) is a subset of the cartesian product \( A \times B \). When \( A = B \), we say that \( R \) is a relation on set \( A \).
**Example.** Let $A = \{1, 2, 3, 4\}$ and $B = \{a, b, c\}$. Consider the following relations.

$R_1 = \{(1, 1), (1, 2), (2, 2), (2, 3)\}$
$R_2 = \{(1, 2), (2, 3), (3, 4), (4, 1), (4, 4)\}$
$R_3 = \{(1, a), (2, a), (3, b), (4, c)\}$
$R_4 = \{(a, 1), (a, 3), (a, 4), (c, 1)\}$
$R_5 = \{(a, a), (a, b), (1, c)\}$

$R_1$ and $R_2$ are relations on $A$. $R_3$ is a relation from $A$ to $B$. $R_4$ is a relation from $B$ to $A$. $R_5$ is not a relation on sets $A$ and $B$ and it is neither a relation from $A$ to $B$ nor a relation from $B$ to $A$.

Below are some more examples of relations.

- If $S$ is a set then “is a subset of”, $\subseteq$ is a relation on $\mathcal{P}(S)$, the power set of $S$.
- “is a student in” is a relation from the set of students to the set of courses.
- “$=$” is a relation on $\mathbb{Z}$.
- “has a path in $G$ to” is a relation on $V(G)$, the set of vertices in $G$.

**Example.** How many relations are there on a set of $n$ elements?

**Solution.** Note that $|A \times A| = n^2$. Since any relation on $A$ is a subset of $A \times A$, the number of possible relations is the cardinality of the power set of $A \times A$, which is $2^{n^2}$. 