Random Variables

Let \((S, P)\) be a sample space. A random variable is a function \(X\) from \(S\) to \(V\) where \(V\) is some set.

Usually \(V\) is a set of numbers, and \(X\) is a function that measures some attribute of the outcome.

Example: Let \(S = \{\text{result of tossing a red die and a blue die}\} = \{(1,1), \ldots, (6,6)\}\)

\[
\text{Let } X(s) = \text{ the sum of the two values in } s
\]

So \(X((3,2)) = 5, \quad X((4,6)) = 10, \quad \text{ etc.}\)

Another Example: Let \(S = \{\text{result of tossing a coin 5 times}\} = \{\text{HHHHH, HHHHT, \ldots, TTTTT}\}\)

\[
\text{Let } X(s) = \text{ the number of Heads in } s
\]

So \(X(\text{HHHHH}) = 5, \quad X(\text{THHTT}) = 2, \quad \text{ etc.}\)
Yet Another Example: Let $S = \{10,11,12,13,14,15\}$ and let $P(i) = \frac{1}{6}$ $\forall i \in S$

Let $X(s) = s \mod 4$ (where $\%$ means “mod”) 

<table>
<thead>
<tr>
<th>$s$</th>
<th>$X(s)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>15</td>
<td>3</td>
</tr>
</tbody>
</table>

Let’s pursue this example a bit further. We can define events based on $X$. For example, “$X = 2$” is an event, as is “$X \in \{1,2\}$”, “$X \geq 2$”, etc. And since we have events, it makes sense to determine the probability of these events. So what is $P(X = 2)$?

Looking at the table above, we can see that $X(s) = 2$ occurs when $s \in \{10,14\}$, and since 10 and 14 have combined probabilities $= \frac{1}{3}$, we can see that $P(X = 2) = \frac{1}{3}$

Now let’s look at the coin flipping example described previously:

Let $S = \{\text{result of tossing a coin 5 times}\} = \{\text{HHHHH, HHHHT, ..., TTTTT}\}$

Let $X(s)$ = the number of Heads in $s$

So $X(\text{HHHHH}) = 5$, $X(\text{THHTT}) = 2$, etc.

Now we can ask questions such as “What is the probability that $X = 2$?” or equivalently, “What is $P(X = 2)$?”

At this point, many people (who should know better) make a crucial mistake: they assume without any justification that the coin we are flipping is balanced (i.e. H and T have equal probability of coming up on each toss). We can’t make that assumption, so we deal with the
more general case. Suppose the coin has probability $p$ of coming up H, and therefore probability $(1-p)$ of coming up T.

What is $P(X = 2)$ with $X(s)$ defined as the number of Heads in $s$?

The coin tosses are independent (the coin has no memory), so the probability of tossing the outcome HHTTT (2 heads and then 3 tails) is

$$p \cdot p \cdot (1-p) \cdot (1-p) \cdot (1-p)$$

This is one of the outcomes that gives $X = 2$

But so is TTHTH, which has probability $(1-p) \cdot (1-p) \cdot p \cdot (1-p) \cdot p$

and in fact we can see that every outcome containing exactly 2 H’s will have probability $p^2 \cdot (1-p)^3$, and $P(X=2)$ will just be the sum of all of these. How many are there?

The two heads can occur in any of the five positions, so a simple counting argument tells us there are $\binom{5}{2}$ different outcomes with exactly 2 H’s. The final answer is:

$$P(X = 2) = \binom{5}{2} \cdot p^2 \cdot (1-p)^3$$

Similarly we can compute $P(X = 4) = \binom{5}{4} \cdot p^4 \cdot (1-p)$

And of course this generalizes completely: when we toss the coin $n$ times, the probability that $X = k$ is given by

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

which we all know and love as the Binomial Theorem or Binomial Formula. What’s it doing here in our discussion of probability? It’s showing us (again!) that probability theory is a branch of discrete math.

Exercise: Suppose we are tossing a coin with $P(H) = 1/3$, $P(T) = 2/3$. If we toss the coin four times and $X(s)$ is the number of Heads we see in outcome $s$, what is $P(X=2)$?