random variables

A random variable is a numeric function of the outcome of an experiment, not the outcome itself.

Ex.

Let $H$ be the number of Heads when 20 coins are tossed
Let $T$ be the total of 2 dice rolls
Let $X$ be the number of coin tosses needed to see 1st head

Note: even if the underlying experiment has “equally likely outcomes,” the associated random variable may not

<table>
<thead>
<tr>
<th>Outcome</th>
<th>$X = #H$</th>
<th>$P(X)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>TT</td>
<td>0</td>
<td>$P(X=0) = 1/4$</td>
</tr>
<tr>
<td>TH</td>
<td>1</td>
<td>$P(X=1) = 1/2$</td>
</tr>
<tr>
<td>HT</td>
<td>1</td>
<td>{</td>
</tr>
<tr>
<td>HH</td>
<td>2</td>
<td>$P(X=2) = 1/4$</td>
</tr>
</tbody>
</table>

random variables

20 balls numbered 1, 2, ..., 20
Draw 3 without replacement
Let $X$ = the maximum of the numbers on those 3 balls
What is $P(X \geq 17)$

$P(X = 20) = \left(\begin{array}{c} 19 \\ 2 \end{array}\right) / \left(\begin{array}{c} 20 \\ 3 \end{array}\right) = \frac{3}{20} \approx 0.150$
$P(X = 19) = \left(\begin{array}{c} 18 \\ 2 \end{array}\right) / \left(\begin{array}{c} 20 \\ 3 \end{array}\right) = \frac{18 \cdot 17 / 2}{20 \cdot 19 / 3} \approx 0.134$

$\sum_{i=17}^{20} P(X = i) \approx 0.508$

$P(X \geq 17) = 1 - P(X < 17) = 1 - \left(\begin{array}{c} 16 \\ 3 \end{array}\right) / \left(\begin{array}{c} 20 \\ 3 \end{array}\right) \approx 0.508$

numbered balls

Flip a (biased) coin repeatedly until 1st head observed
How many flips? Let $X$ be that number.

$P(X=1) = P(H) = p$
$P(X=2) = P(TH) = (1-p)p$
$P(X=3) = P(TTH) = (1-p)^2p$

$\sum_{i=1}^{\infty} x^i = \frac{1}{1 - x}$, when $|x| < 1$

first head
A discrete random variable is one taking on a countable number of possible values.

Ex:

\( X = \text{sum of 3 dice}, \ 3 \leq X \leq 18, \ X \in \mathbb{N} \)

\( Y = \text{position of 1st head in seq of coin flips}, \ 1 \leq Y, \ Y \in \mathbb{N} \)

\( Z = \text{largest prime factor of } (1+Y), \ Z \in \{2, 3, 5, 7, 11, \ldots\} \)

**Probability mass functions**

**Definition:** If \( X \) is a discrete random variable taking on values from a countable set \( T \subseteq \mathbb{R} \), then

\[ p_X(a) = \begin{cases} P(X = a) & \text{for } a \in T, \\ 0 & \text{otherwise} \end{cases} \]

is called the probability mass function. Note: \( \sum_{a \in T} p_X(a) = 1 \)

Let \( X \) be the number of heads observed in \( n \) coin flips

\[ P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}, \text{ where } p = P(H) \]

Probability mass function (\( p = \frac{1}{2} \)):

<table>
<thead>
<tr>
<th>( k )</th>
<th>( k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.0</td>
<td>0.2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>( k )</th>
<th>( k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

The cumulative distribution function for a random variable \( X \) is the function \( F : \mathbb{R} \to [0,1] \) defined by

\[ F(a) = P[X \leq a] \]

Ex: 3 students; homework returned according to random permutation.

\( X \) is number of homeworks returned to their correct owner.

\[ \begin{array}{c|c|c|c}
\text{Prob} & \text{outcome} & \text{X} \\
\hline
1/6 & 123 & 3 \\
1/6 & 132 & 1 \\
1/6 & 213 & 1 \\
1/6 & 231 & 0 \\
1/6 & 312 & 0 \\
1/6 & 321 & 1 \\
\end{array} \]
The cumulative distribution function for a random variable \( X \) is the function \( F: \mathbb{R} \rightarrow [0,1] \) defined by
\[
F(a) = P(X \leq a)
\]
summing over \( j \in \text{Range}(X) \).

The cumulative distribution function for a random variable \( X \) is the function \( F: \mathbb{R} \rightarrow [0,1] \) defined by
\[
F(a) = P(X \leq a)
\]

Ex: if \( X \) has probability mass function given by:
\[
p(1) = \frac{1}{4} \quad p(2) = \frac{1}{2} \quad p(3) = \frac{1}{8} \quad p(4) = \frac{1}{8}
\]

\[
F(a) = \begin{cases} 
0 & a < 1 \\
\frac{1}{4} & 1 \leq a < 2 \\
\frac{3}{4} & 2 \leq a < 3 \\
\frac{7}{8} & 3 \leq a < 4 \\
1 & 4 \leq a
\end{cases}
\]

NB: for discrete random variables, be careful about “\( \leq \)” vs “\(<\)”

For a discrete r.v. \( X \) with p.m.f. \( p(\cdot) \), the expectation of \( X \), aka expected value or mean, is
\[
E[X] = \sum x p(x)
\]
average of random values, weighted by their respective probabilities

For the equally-likely outcomes case, this is just the average of the possible random values of \( X \)

For unequally-likely outcomes, it is again the average of the possible random values of \( X \), weighted by their respective probabilities

Ex 1: Let \( X = \) value seen rolling a fair die \( p(1), p(2), \ldots, p(6) = 1/6 \)
\[
E[X] = \sum_{i=1}^{6} i p(i) = \frac{1}{6} (1 + 2 + \cdots + 6) = \frac{21}{6} = 3.5
\]

Ex 2: Coin flip; \( X = +1 \) if H (win $1), -1 if T (lose $1)
\[
E[X] = (+1)\cdot p(+1) + (-1)\cdot p(-1) = 1\cdot(1/2) + (-1)\cdot(1/2) = 0
\]
For a discrete r.v. $X$ with p.m.f. $p(*)$, the expectation of $X$, aka expected value or mean, is

$$E[X] = \sum_x x p(x)$$

average of random values, weighted by their respective probabilities

Another view: A 2-person gambling game. If $X$ is how much you win playing the game once, how much would you expect to win, on average, per game, when repeatedly playing?

**Example:**

3 students; homework returned according to random permutation.

$X$ is number of homeworks returned to their correct owner.

<table>
<thead>
<tr>
<th>Prob</th>
<th>outcome</th>
<th>$X$</th>
<th>$E(X)$ = ?</th>
</tr>
</thead>
<tbody>
<tr>
<td>1/6</td>
<td>123</td>
<td>3</td>
<td>$E(X) = 0 \cdot \frac{2}{6} + 1 \cdot \frac{3}{6} + 3 \cdot \frac{1}{6}$</td>
</tr>
<tr>
<td>1/6</td>
<td>132</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1/6</td>
<td>213</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1/6</td>
<td>231</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1/6</td>
<td>312</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>1/6</td>
<td>321</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

$$= X(123) \cdot \frac{1}{6} + X(132) \cdot \frac{1}{6} + X(213) \cdot \frac{1}{6} + X(231) \cdot \frac{1}{6} + X(312) \cdot \frac{1}{6} + X(321) \cdot \frac{1}{6}$$
For a discrete r.v. $X$ with p.m.f. $p(x)$, the expectation of $X$, aka expected value or mean, is

$$E[X] = \sum_x x p(x)$$

Another view: $E[X] = \sum_{x \in S} X(x) \cdot p(s)$

20 balls numbered 1, 2, ..., 20
Draw 3 without replacement
Let $X$ = the maximum of the numbers on those 3 balls
$E(X) = ?$

$$E(X) = \sum_{k=3}^{20} k \left( \frac{k-1}{\binom{20}{3}} \right)$$