MAT1830

Lecture 23: Random variables

In a game, three standard dice will be rolled and the number of sixes will be recorded. We could let X stand for the number of sixes rolled. Then X is a special kind of variable whose value is based on a random process. These are called *random variables*.

Because the value of X is random, it doesn't make sense to ask whether $X = 0$, for example. But we can ask what the *probability is* that $X = 0$ or that $X \geqslant 2$. This is because "$X = 0$" and "$X \geqslant 2$" correspond to events from our sample space.

23.1 Formal definition

Formally, a random variable is defined as a function from the sample space to $\mathbb{R}$. In the example above, X is a function from the process's sample space that maps every outcome to the number of sixes in that outcome.

Example. Let X be the number of 1s in a binary string of length 2 chosen uniformly at random. Formally, X is a function from $\{00, 01, 10, 11\}$ to $\{0, 1, 2\}$ such that

$$X(00) = 0, \ \ X(01) = 1, \ \ X(10) = 1, \ \ X(11) = 2.$$

For most purposes, however, we can think of X as simply a special kind of variable.

Example A word is chosen uniformly at random from the set $\{$I, feed, buns, to, the, elephant$\}$

Let X be the number of e's in the word chosen.
Let Y be the number of t's in the word chosen.

$\Pr(X = 1) = \frac{1}{6}$ (just 'the').

$\Pr(X = 2) = \frac{2}{6} = \frac{1}{3}$ ('feed' and 'elephant').

$\Pr(Y = 1) = \frac{3}{6} = \frac{1}{2}$ ('to', 'the' and 'elephant').

$\Pr(X = 1 \wedge Y = 1) = \frac{1}{6}$ (just 'the').

23.2 Probability distribution

We can describe the behaviour of a random variable X by listing, for each value x that X can take, the probability that $X = x$. This gives the *probability distribution* of the random variable. Again, formally this listing is a function from the values of X to their probabilities.

Example. Continuing with the last example, the probability distribution of X is given by

$$\Pr(X = x) = \begin{cases} \frac{1}{4} & \text{if } x = 0 \\ \frac{1}{2} & \text{if } x = 1 \\ \frac{1}{4} & \text{if } x = 2. \end{cases}$$

It can be convenient to give this as a table:

x	0	1	2
$\Pr(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

.

Example. A standard die is rolled three times. Let X be the number of sixes rolled. What is the probability distribution of X? Obviously X can only take values in $\{0, 1, 2, 3\}$. Each roll there is a six with probability $\frac{1}{6}$ and not a six with probability $\frac{5}{6}$. The rolls are independent.

$\Pr(X = 0) = \frac{5}{6} \times \frac{5}{6} \times \frac{5}{6}$

$\Pr(X = 1) = (\frac{1}{6})(\frac{5}{6})(\frac{5}{6}) + (\frac{5}{6})(\frac{1}{6})(\frac{5}{6}) + (\frac{5}{6})(\frac{5}{6})(\frac{1}{6})$

$\Pr(X = 2) = (\frac{1}{6})(\frac{1}{6})(\frac{5}{6}) + (\frac{1}{6})(\frac{5}{6})(\frac{1}{6}) + (\frac{5}{6})(\frac{1}{6})(\frac{1}{6})$

$\Pr(X = 3) = \frac{1}{6} \times \frac{1}{6} \times \frac{1}{6}$

So the probability distribution of X is

x	0	1	2	3
$\Pr(X = x)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

.

Question

23.1 An elevator is malfunctioning. Every minute it is equally likely to ascend one floor (U), descend one floor (D), or stay where it is (S). When it begins malfunctioning it is on level 5. Let X be the level it is on three minutes later. Find the probability distribution for X.

Answer The elevator must end up on a floor between 2 and 8 inclusive, so the possible values for X are $\{2, 3, 4, 5, 6, 7, 8\}$.

We can record the elevator's journey as a string of length 3 over the alphabet $\{U, D, S\}$. Each string occurs with probability $= \frac{1}{27}$.

$\Pr(X = 2) = \Pr(DDD) = \frac{1}{27}$.

$\Pr(X = 3) = \Pr(SDD) + \Pr(DDS) + \Pr(DSD) = \frac{3}{27}$.

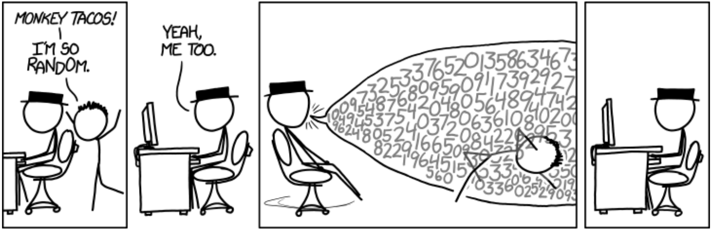
$\Pr(X = 4) = \Pr(SDS) + \Pr(SSD) + \Pr(DSS) + \Pr(DDU) + \Pr(DUD) + \Pr(UDD)$
$= \frac{6}{27}$.

$\Pr(X = 5) = \Pr(SSS) + \Pr(SDU) + \Pr(SUD) + \Pr(DSU) +$
$\Pr(DUS) + \Pr(USD) + \Pr(UDS) = \frac{7}{27}$.

By symmetry, the probability distribution of X is

x	2	3	4	5	6	7	8
$\Pr(X = x)$	$\frac{1}{27}$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{7}{27}$	$\frac{2}{9}$	$\frac{1}{9}$	$\frac{1}{27}$

.


MONKEY TACOS!
I'M SO RANDOM.
YEAH, ME TOO.

23.3 Independence

We have seen that two events are independent when the occurrence or non-occurrence of one event does not affect the likelihood of the other occurring. Similarly two random variables are *independent* if the value of one does not affect the likelihood that the other will take a certain value.

> Random variables X and Y are *independent* if, for all x and y,
>
> $$\Pr(X = x \wedge Y = y) = \Pr(X = x)\Pr(Y = y).$$

Example. An integer is generated uniformly at random from the set $\{10, 11, \ldots, 29\}$. Let X and Y be its first and second (decimal) digit. Then X and Y are independent random variables because, for $x \in \{1, 2\}$ and $\{0, 1, \ldots, 9\}$,

$$\begin{aligned}
\Pr(X = x \wedge Y = y) &= \tfrac{1}{20} \\
&= \tfrac{1}{2} \times \tfrac{1}{10} \\
&= \Pr(X = x)\Pr(Y = y).
\end{aligned}$$

Flux Exercise (LQMTZZ)

An integer is generated uniformly at random from the set $\{11, 12, \ldots, 30\}$. Let X and Y be its first and second (decimal) digit. Are the random variables X and Y independent?

A. Yes

B. No

Answer No.

X can take values in $\{1, 2, 3\}$
Y can take values in $\{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$ (each with probability $1/10$).

The probability distribution of X is

x	1	2	3
$\Pr(X = x)$	$\frac{9}{20}$	$\frac{10}{20}$	$\frac{1}{20}$

.

$\Pr(X = 3 \wedge Y = 1) = 0$
$\Pr(X = 3)\Pr(Y = 1) = (\frac{1}{20})(\frac{1}{10}) = \frac{1}{200}$.

So $\Pr(X = 3 \wedge Y = 1) \neq \Pr(X = 3)\Pr(Y = 1)$ and X and Y are not independent.

23.4 Operations

From a random variable X, we can create new random variables such as $X+1, 2X$ and X^2. These variables work as you would expect them to.

Example. If X is the random variable with distribution

x	-1	0	1
$\Pr(X=x)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

,

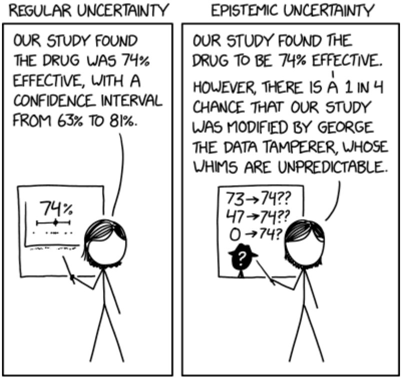
then the distributions of $X+1, 2X$ and X^2 are

y	0	1	2
$\Pr(X+1=y)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

y	-2	0	2
$\Pr(2X=y)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{2}$

y	0	1
$\Pr(X^2=y)$	$\frac{1}{3}$	$\frac{2}{3}$

.


REGULAR UNCERTAINTY
EPISTEMIC UNCERTAINTY
OUR STUDY FOUND THE DRUG WAS 74% EFFECTIVE, WITH A CONFIDENCE INTERVAL FROM 63% TO 81%.
74%
OUR STUDY FOUND THE DRUG TO BE 74% EFFECTIVE.
HOWEVER, THERE IS A 1 IN 4 CHANCE THAT OUR STUDY WAS MODIFIED BY GEORGE THE DATA TAMPERER, WHOSE WHIMS ARE UNPREDICTABLE.
73→74??
47→74??
0→74?
?

23.5 Sums and products

From random variables X and Y we can define a new random variable $Z = X + Y$. Working out the distribution of Z can be complicated, however. We give an example below of doing this when X and Y are independent.

Example. Let X and Y be independent random variables with

x	0	1	2
$\Pr(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

y	0	1	2	3
$\Pr(Y = y)$	$\frac{1}{6}$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{6}$

Let $Z = X + Y$. To find $\Pr(Z = z)$ for some value of z, we must consider all the ways that $X + Y$ could equal z. For example, $X + Y = 3$ could occur as $(X, Y) = (0, 3)$, $(X, Y) = (1, 2)$ or $(X, Y) = (2, 1)$. Because X and Y are independent, we can find the probability that each of these occur

$$\begin{aligned}\Pr(X = 0 \wedge Y = 3) &= \tfrac{1}{4} \times \tfrac{1}{6} = \tfrac{1}{24},\\ \Pr(X = 1 \wedge Y = 2) &= \tfrac{1}{2} \times \tfrac{1}{3} = \tfrac{1}{6},\\ \Pr(X = 2 \wedge Y = 1) &= \tfrac{1}{4} \times \tfrac{1}{3} = \tfrac{1}{12}.\end{aligned}$$

So, because the three are mutually exclusive,

$$\Pr(X = 3) = \frac{1}{24} + \frac{1}{6} + \frac{1}{12} = \frac{7}{24}.$$

Doing similar calculations for each possible value, we see that the distribution of Z is

z	0	1	2	3	4	5
$\Pr(Z = z)$	$\frac{1}{24}$	$\frac{1}{6}$	$\frac{7}{24}$	$\frac{7}{24}$	$\frac{1}{6}$	$\frac{1}{24}$

.

The distribution of a product of two independent random variables can be found in a similar way.

Finding the distribution of sums or products of dependent random variables is even more complicated. In general, this requires knowing the probability of each possible combination of values the variables can take.

Flux Exercise (LQMTZZ)

Let X and Y be independent random variables with distributions

x	0	2
$\Pr(X = x)$	$\frac{1}{4}$	$\frac{3}{4}$

y	0	1	2	3
$\Pr(Y = y)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

.

What is $\Pr(X + Y = 3)$?

A. 1/4

B. 1/16

C. 3/16

D. 3/4

Answer

$X + Y = 3$ could happen in two ways: $(X = 0,\ Y = 3)$ OR $(X = 2,\ Y = 1)$

$\Pr(X = 0 \wedge Y = 3) = \frac{1}{4} \times \frac{1}{4} = \frac{1}{16}$

$\Pr(X = 2 \wedge Y = 1) = \frac{3}{4} \times \frac{1}{4} = \frac{3}{16}$

$\Pr(X + Y = 3) = \Pr(X = 0 \wedge Y = 3) + \Pr(X = 2 \wedge Y = 1) = \frac{1}{16} + \frac{3}{16} = \frac{1}{4}$.

So A.

Question

23.3 Let X and Y be independent random variables with distributions

x	0	2
$\Pr(X = x)$	$\frac{1}{4}$	$\frac{3}{4}$

y	0	1	2	3
$\Pr(Y = y)$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

Find the probability distribution of $Z = X + Y$.

Answer Since X and Y are independent, and Y is uniform, for each y,

$\Pr(X = 0 \wedge Y = y) = \Pr(X = 0)\Pr(Y = y) = (\frac{1}{4})(\frac{1}{4}) = \frac{1}{16}$.

$\Pr(X = 2 \wedge Y = y) = \Pr(X = 2)\Pr(Y = y) = (\frac{3}{4})(\frac{1}{4}) = \frac{3}{16}$.

$\Pr(X + Y = 0) = \Pr(X = 0 \wedge Y = 0) = \frac{1}{16}$.

$\Pr(X + Y = 1) = \Pr(X = 0 \wedge Y = 1) = \frac{1}{16}$.

$\Pr(X + Y = 2) = \Pr(X = 0 \wedge Y = 2) + \Pr(X = 2 \wedge Y = 0) = \frac{4}{16}$.

$\Pr(X + Y = 3) = \Pr(X = 0 \wedge Y = 3) + \Pr(X = 2 \wedge Y = 1) = \frac{4}{16}$.

$\Pr(X + Y = 4) = \Pr(X = 2 \wedge Y = 2) = \frac{3}{16}$.

$\Pr(X + Y = 5) = \Pr(X = 2 \wedge Y = 3) = \frac{3}{16}$.

The distribution of Z is

z	0	1	2	3	4	5
$\Pr(Z = z)$	$\frac{1}{16}$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{3}{16}$	$\frac{3}{16}$

Flux Exercise (LQMTZZ)

I toss 100 coins and count the number H of heads: I think about the events (E1) $H = 50$, (E2) $H > 60$, (E3) $H = n^2$ for some integer n, and (E4) $H \equiv 0 \mod 4$. Ordering these events from most likely to least likely I get:

A. E1, E2, E3, E4.

B. E1, E4, E3, E2.

C. E2, E4, E1, E3.

D. E3, E2, E4, E1.

E. E4, E3, E1, E2.

F. E4, E1, E3, E2.

Answer E. In fact
$\Pr(E1) \approx 7.96\%$, $\Pr(E2) \approx 1.76\%$, $\Pr(E3) \approx 8.11\%$ and $\Pr(E4) = 25.00\%$.