DISCRETE RANDOM VARIABLES

We have a probability space $(S, \Pr)$.
A *random variable* is a function $X : S \to V(X)$ for some set $V(X)$. In this discussion, we must have $V(X)$ is the real numbers
X induces a partition of S: for a value x of X we define

$$X = x \quad = \quad X^{-1}(x) \quad = \quad \{s \in S : X(s) = x\}$$

$X = x$ is an event, and so we know what $Pr(X = x)$ means.
We get an *expectation* of the random variable X:

$$E(X) \quad = \quad \sum_{x} x \Pr(X = x) \quad = \quad \sum_{s \in S} X(s) \Pr(s).$$

Example

Name	Age	Prob
John	20	.4
Mary	30	.3
Jean	40	.3

$$E(\text{Age}) \quad = \quad (.4)(20) + (.3)(30) + (.4)(40) \quad = \quad 29.$$

We also can add and multiply random variables.

Name	.3Age	Prob
John	6	.4
Mary	9	.3
Jean	12	.3

$E(.3\text{Age}) = .3E(\text{Age}) = 8.7.$

Another Example

Flip a coin 100 times.
The space S is the set of 100-tuples of H and T's. Each tuple is equally likely.
$X_1 = 1$ if the first flip is H, 0 otherwise.
$X_2 = 1$ if the second flip is H, 0 otherwise.
$X_{41} = 1$ if the 41^{st} flip is H, 0 otherwise.
$E(X_i) = 1/2$.
We can add random variables.
$X + Y$ is a new random variable, with $(X + Y)(s) = X(s) + Y(s)$.
The expectation always adds:
$E(X_{12} + X_{45}) = E(X_{12}) + E(X_{45}) = .5 + .5 = 1$.

Why does expectation add?

$$\begin{aligned} E(X+Y) &= \textstyle\sum_{s\in S}(X+Y)(s)\Pr(s) \\ &= \textstyle\sum_{s\in S}(X(s)+Y(s))\cdot\Pr(s) \\ &= \textstyle\sum_{s\in S}(X(s)\cdot\Pr(s)+Y(s)\cdot\Pr(s)) \\ &= \textstyle\sum_{s\in S}X(s)\Pr(s) \;+\; \sum_{s\in S}Y(s)\Pr(s) \\ &= E(X)+E(Y) \end{aligned}$$

Recall also that we multiply random variables by numbers. So cX is a random variable with $(cX)(s) = c(X(s))$.

You might similarly show that $E(cX) = cE(X)$, where c is a constant.

How about multiplication?

As it happens, it's only ok to multiply expectations when the random variables are *independent*. So suppose X and Y are independent.

$$
\begin{aligned}
E(XY) &= \sum_{x,y} \Pr(X = x, Y = y)xy \\
&= \sum_{x,y} \Pr(X = x) \Pr(Y = y)xy \\
&= \sum_{x} \sum_{y} \Pr(X = x) \Pr(Y = y)xy \\
&= \sum_{x} x \Pr(X = x) \sum_{y} \Pr(Y = y)y \\
&= \sum_{x} x \Pr(X = x)E(Y) \\
&= E(Y) \sum_{x} x \Pr(X = x) \\
&= E(Y)E(X) \\
&= E(X)E(Y)
\end{aligned}
$$

Where was independence used?
And why is the first line correct in the first place?

Why We Add Random Variables

Going back to the example with S = the 100-tuples of H, T, let

$$Y = X_1 + X_2 + \cdots + X_{100}.$$

Then $Y(s)$ gives the number of heads in the tuple s.

$$\begin{aligned} E(Y) &= E(X_1 + X_2 + \cdots + X_{100}) \\ &= E(X_1) + E(X_2) + \cdots + E(X_{100}) \\ &= (.5) + (.5) + \cdots + (.5) \\ &= 50 \end{aligned}$$

We also recall the formula

$$\Pr(Y = k) = \binom{100}{k} (.5)^{100}$$

Constant Random Variables

A random variable can also be constant, such as $X(s) = 3$ always.
In this case, $E(X) = 3$ as well.
We often will have random variables like $\text{Age} - 2$.
We think of this as the sum of the random variable Age and the random variable -2.

New expectations from old

Often $E(X)$ is called the mean of X, and is written μ.
This hides the random variable, so it would be better to write it as μ_X when we need it.

New expectations from old

Often $E(X)$ is called the mean of X, and is written μ.
This hides the random variable, so it would be better to write it as μ_X when we need it.

$$\mu_{aX} = a\mu_X$$

$$\mu_{X+Y} = \mu_X + \mu_Y$$

$$\mu_c = c$$

Make sure you understand what these mean.

VARIANCE

The variance $V(X)$ of a random variable measures "how spread out X is around its mean."

$$V(X) = E((X-\mu)^2).$$

That is, the expectation of the new random variable $(X-\mu)^2$.
In our first example,

Name	Age	Prob
John	20	.4
Mary	30	.3
Jean	40	.3

the mean is 29 and

$$V(X) = (.4)(20-29)^2 + (.3)(30-29)^2 + (.3)(40-29)^2.$$

A FORMULA

Fact: $V(X) = E(X^2) - (E(X))^2$.
This will be clearer if we write μ for $E(X)$.
So $V(X) = E((X-\mu)^2)$.
We now we prove our fact:

$$\begin{aligned} V(X) &= E((X-\mu)(X-\mu)) \\ &= E(X^2 - 2\mu X + \mu^2) \\ &= E(X^2) - E(2\mu X) + E(\mu^2)) \\ &= E(X^2) - 2\mu E(X) + \mu^2 \\ &= E(X^2) - 2\mu \cdot \mu + \mu^2 \\ &= E(X^2) - \mu^2 \end{aligned}$$

Example

From the coin flipping example, with $X_1, X_2, \ldots$.
$E(X_1) = 1/2$.
Each X_i^2 is just like X_i, since when we square 0 and 1 nothing happens.
So $E(X_1^2) = 1/2$.
Then $V(X_1) = 1/2 - (1/2)^2 = 1/2 - 1/4 = 1/4$.
More generally, suppose that X is any random variable with values 0 or 1, and suppose that $\Pr(X = 1) = p$. Then $E(X) = p$, and

$$V(X) \quad = \quad p - p^2 \quad = \quad p(1-p).$$

A random variable like this is called a Bernoulli random variable.

More on the coin flipping example

If we flip a coin 100 times, the *expected* number of heads is 50.
But the actual probability of this is very small, about 0.08.
We defined Y to be $X_1 + \cdots X_{100}$.
We might like to know $\Pr[40 \leq Y \leq 60]$, for example.
We'll get to this a little later.
We really would be interested in the variance of Y. This would tell us something related to what we want.
What would $E(Y - 50)$ tell us?
What would $E(|Y - 50|)$ tell us?
What would $E((Y - 50)^2)$ tell us?
We need a general fact: the variances of *independent* random variables add up:

$$V(X + Y) \quad = \quad V(X) + V(Y).$$

ADDING THE VARIANCES OF INDEPENDENT VARIABLES

Let's write μ_X for $E(X)$, μ_Y for $E(Y)$.
As we know $E(X+Y) = E(X) + E(Y) = \mu_X + \mu_Y$.

$$
\begin{aligned}
& V(X) + V(Y) \\
= \quad & E((X+Y)^2 - (E(X+Y))^2) \\
= \quad & E(X^2 + 2XY + Y^2 - (\mu_X + \mu_Y)^2) \\
= \quad & E(X^2) + 2E(XY) + E(Y^2) - (\mu_X + \mu_Y)^2) \\
= \quad & E(X^2) + 2E(X)E(Y) + E(Y^2) - (\mu_X + \mu_Y)^2) \quad \text{the key!} \\
= \quad & E(X^2) - 2\mu_X\mu_Y + E(Y^2) \\
& \quad -(\mu_X^2 + 2\mu_X\mu_Y + \mu_Y^2) \\
= \quad & E(X^2) - \mu_X^2 + E(Y^2) - \mu_Y^2 \\
= \quad & V(X) + V(Y)
\end{aligned}
$$

Old Variances from New

We just saw $V(X+Y) = V(X) + V(Y)$ for X, Y independent.
There are two more important laws. We'll try to find them together.
First, if a is a constant, try to get $V(aX)$.
Second, if a is again a constant, what is $V(a)$?

Back to the coin flipping example

All the X_i are independent.
So $V(Y) = \sum_i V(X_i) = 100 \cdot .25 = 25$.
Recall that $V(X) = E((X - \mu)^2))$. Usually one wants the square root of this, and this is called the standard deviation.

$$\sigma = \sqrt{V(X)}.$$

In the example that we are working with σ is $\sqrt{25} = 5$.
Please be aware that the notations μ and σ hide the random variable under discussion. Sometimes this is confusing!

BACK TO THE COIN FLIPPING EXAMPLE

Suppose that $X_1, \ldots, X_n$ are independent Bernoulli random variables with the same probability p.
Let $Y = X_1 + \cdots + X_n$.
Then $E(Y) = np$, and $V(Y) = np(1-p)$.
So $\sigma(Y) = \sqrt{np(1-p)}$.

Formulas For Sums of Bernoulli Variables

Suppose that $X_1, \ldots, X_n$ are Bernoulli rv's with mean p and variance $p(1-p)$.
Let $S = X_1 + \cdots + X_n$.
Then we have the following formula:

$$\Pr(S = k) \quad = \quad \binom{n}{k} p^k (1-p)^{n-k}.$$

Here (and elsewhere)

$$\binom{n}{k} \quad = \quad \frac{n!}{k!(n-k)!}.$$

We can simplify this a little by remembering that $n! = 1 \times 2 \times \cdots \times n$, and then

$$\binom{n}{k} \quad = \quad \frac{(n-k+1) \times (n-k+2) \times \cdots \times n}{1 \times 2 \times \cdots \times (n-k)}$$

Example

Suppose we roll a fair die 20 times. What is the probability of exactly five 3's?
Our space is the set of 20-tuples of numbers from 1 to 6.
The random variable X_1 is 1 if the first roll was a 3, 0 otherwise.
Similarly for the others.
$\Pr(X_i = 1) = 1/6$ for all i.
Again, we let $Y = X_1 + \cdots + X_{20}$.
We want to know $\Pr(Y = 5)$.
This is

$$\binom{20}{5}(1/6)^5(5/6)^{15}.$$

Example continued

$$\binom{20}{5} = \frac{1 \times 2 \times \cdots 14 \times 15 \times \cdots \times 20}{(1 \times 2 \times \cdots 14 \times 15)(1 \times 2 \times \cdots \times 5)} = \frac{16 \times \cdots \times 20}{1 \times 2 \times \cdots \times 5} = 15504$$

So we get $15004(1/6)^5(5/6)^{15}$.
This is going to be a very small number.
In case n is even bigger, the formula is difficult to evaluate exactly. And so one can use *approximations*. This is especially valuable when we want to calculate things like "What is the probability that when we roll a fair die 600 times, the number of 3's is between 90 and 110?"
"What is the probability that when we flip a fair coin 100 times, the number of heads is between 40 and 60?"

Approximations of probabilities using tables/web sites

Suppose we have independent, identically distributed random variables $X_1, \ldots, X_n$. Suppose that $\Pr(X_i = 1) = p$.
Let $Y = X_1 + \cdots X_n$.
Then for Y, the mean μ is np.
The variance σ^2 is $np(1-p)$.
(Often one sees npq, where $q = 1 - p$.)
The standard deviation σ is $\sqrt{np(1-p)}$.

Approximations of probabilities using tables/web sites

One is often interested in probabilities like $\Pr(a \leq Y \leq b)$.
Here is how to estimate them.
First, calculate μ and σ as numbers.
Second, take $a \leq Y \leq b$. Subtract μ and divide by σ.
We get

$$\frac{a-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{b-\mu}{\sigma}.$$

On the left and right, you'll have exact numbers.
Now, I would like you to use what we did before to get μ and σ for the new random variable $Y - \mu/\sigma$.

Approximations of probabilities using tables/web sites

Old random variable: $Y - \mu/\sigma$.
New one: cal it Z
The new random variable Z has mean 0 and standard deviation 1.
It can be shown that for Z obtained this way from a "large" sum of independent Bernoulli variables, the probabilities of Z are nicely approximated by the areas under standard normal curve.

Example, again

What's the probability that when we flip a fair coin 100 times, the sum is between 40 and 60?

Here $n = 100$, $p = .5$, $\mu = E(Y) = 50$,
$\sigma = \sqrt{np(1-p)} = \sqrt{25} = 5$.
We want $40 \leq Y \leq 60$, and so
this is like

$$\frac{40-50}{5} \leq Z \leq \frac{60-50}{5}.$$

That is, $-2 \leq Z \leq 2$.
The probability is 95%.

Approximations of probabilities using tables/web sites

You can look up the approximation of

$$\Pr\left(x \leq \frac{Y-\mu}{\sigma}\right)$$

for various values of x in a table. Usually a table would only list values between 0 and around 3.

This is because the negatives come for free by symmetry, and 99.7% of the probability is within three standard deviations of the mean.

If you keep in mind the picture of the bell curve, you'll understand how the approximations work.

An Example

"What is the probability that when we roll a fair die 600 times, the number of 3's is between 90 and 110?"
Here $p = .16$, $n = 600$, $\mu = 100$, $\sigma = \sqrt{(600)(1/6)(5/6)} = 9.13$.
We want $\Pr(90 \leq Y \leq 110)$.
Now $(90 - 100)/9.13 = -1.1$ and $(110 - 100)/9.13 = 1.1$ So we want $\Pr(-1.1 \leq (Y - 100)/9.13 \leq 1.1)$.
The tables give $\Phi(1.1)$ to be about 0.86. See below:

By some work with the graph, we get an approximate answer of about .72.