Combining Random Variables

Reading

- Section 2.4.3, 2.4.4

Practice Problems

2.6.4 (Page 123) 2.41, 2.42

There are two main ways we will look at of creating new random variables from existing ones: linear transformations and addition of variables.

Linear Transformations

A linear transformation of random variables is similar to the corresponding transformations we had seen earlier, it boils down to an equation like so:

$$Y = bX + a$$

Let us look at an example: In a certain theater show the ticket price is $\$5$ plus twice the roll of a die. So if you roll a 3, then you pay $5 + 3 \cdot 3 = \$14$.

We can represent this more generally as follows: Denote by X the result of rolling a die, and by Y the price of the ticket. Then we can write:

$$Y = 2X + 5$$

So the question is how the two variables relate. That formula is a start: It means that for any particular value of X we can compute the corresponding value for Y. The key thing with linear transformations is that the probabilities stay the same, only the values change. So the probability tables for our variables would look as follows:

X	1	2	3	4	5	6
Y	7	9	11	13	15	17
p	1/6	1/6	1/6	1/6	1/6	1/6

> When performing a linear transformation of random variables, the probabilities of the different values remain the same, only the values change.
>
> You can compute the mean and standard deviation of the new variable based on the original variable:
>
> $$\mu_Y = b\mu_X + a$$
>
> $$\sigma_Y = b\sigma_X$$

For instance let us consider the above example. The average price ticket is:

$$\mu_X = \frac{1}{6} \cdot 1 + \cdots + \frac{1}{6} \cdot 6 = \frac{21}{6} = 3.5$$

And the standard deviation is:

$$\sigma_X = \sqrt{\frac{1}{6} \cdot (1 - 3.5)^2 + \cdots + \frac{1}{6} \cdot (6 - 3.5)^2} = \sqrt{\frac{17.5}{6}} = 1.708$$

If we wanted to compute the same quantities for the variable Y, we could try to do the corresponding computations from the table values for Y, or we could instead use the relation with X:

$$\mu_Y = 2\mu_X + 5 = 2 \cdot 3.5 + 5 = 12$$
$$\sigma_Y = 2\sigma_X = 3.416$$

Adding random variables

Another key operation is when we want to add two random variables. This is a more complicated operation, which actually depends on how the two variables relate to each other. Let us consider the following two problems:

> **Case A**
>
> We roll a six-sided die. If the result is 4 or more then Bob wins $\$2$. If it is a 3 or a 5, then Alice wins $\$3$. Denote by X Bob's winnings, by Y Alice's winnings, and by Z their total winnings, which we could write symbolically as $Z = X + Y$.
>
> **Case B**
>
> We roll a six-sided die. If the result is 4 or more then Bob wins $\$2$. Then we roll another six-sided die. If the result is 3, 4 or 5, then Alice wins $\$3$. Denote by X Bob's winnings, by Y Alice's winnings, and by Z their total winnings, so again we could write $Z = X + Y$.

In both of these cases, the tables for X and Y$ are the same:

X	0	2
p	1/2	1/2

Y	0	3
p	1/2	1/2

However, their "sum" is a complicated question. First off, what are the possible values for $Z = X + Y$? It could be 0, 2, 3 or 5. Let us compute the table for Z in the first case. these would correspond respectively to the sets $\{1, 2\}$, $\{6\}$, $\{3\}$ and $\{4, 5\}$ respectively. So the table for Z would be:

Z	0	2	3	5
p	1/3	1/6	1/6	1/3

Now in the second example, case B, the two rolls are independent. For instance $Z = 0$, we must have $X = 0$ and $Y = 0$, and those two are independent so we multiply them to find $P(Z = 0) = 1/4$. In general the table for Z in this case would be:

Z	0	2	3	5
p	1/4	1/4	1/4	1/4

Notice the two tables are different! The reason is that the interaction between the two variables matters a great deal. In the case B they are independent of each other, in case A they are not.

> To determine the probability distribution of the sum of two variables $Z = X + Y$ we need to know the **joint probability** distribution of X and Y, i.e. the chances of combined events like $P(X = 0 \text{ and } Y = 3)$.

The joint probability distribution is a 2-way table showing the probabilities of all combined outcomes. For case A this would look as follows:

Y\X	0	2
0	1/3	1/6
3	1/6	1/3

The corresponding table for case B will be different.

> Even though the actual distribution of a sum is harder to compute, the mean, and some times the standard deviation, have simple formulas:
>
> $$\mu_{X+Y} = \mu_X + \mu_Y$$
>
> and if the variables X, Y are independent of each other, then
>
> $$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$

Verify these results in the above example:

1. Compute μ_X and μ_Y from their tables.

2. Compute μ_Z for case A and case B directly from the computed tables.
3. Verify that $\mu_Z = \mu_X + \mu_Y$.
4. Compute σ_X^2 and σ_Y^2 from their tables.
5. Compute σ_Z^2 for case A and case B directly from the computed tables. Notice how they differ.
6. Verify that the second case matches with $\sigma_X^2 + \sigma_Y^2$.

Let us consider one more example of this:

> We roll a six-sided die. If the result is 4 or more then Bob wins $\$2$. If it is a 5 or 6, then Alice wins $\$3$. Denote by X Bob's winnings, by Y Alice's winnings, and by Z their total winnings, which we could write symbolically as $Z = X + Y$.

1. Compute the tables for X, Y and Z, as well as the joint probability table for X and Y combined.
2. Compute μ_X, σ_X^2, μ_Y, σ_Y^2 from their tables.
3. Compute μ_Z, σ_Z^2 from its table. Compute $\sigma_X^2 + \sigma_Y^2$ and notice that it differs from σ_Z^2.