This writeup was originally created when we were using a different textbook in Math 1710. It’s optional, but the slightly different style may help you better understand the challenging parts of Chapters 16 and 17.

1. Random Variables

A random variable is a variable describing a number arising from a random experiment. Such variables have probability distributions associated with them. You can think of such a distribution as a table giving the probability of each value showing up. Here are some examples:

Example 1 Suppose you roll two fair dice. There are 36 equally possible outcomes. But if we just pay attention to the sum of the numbers on the dice, we have a random variable, say $T$, that can take on the values 2, 3, 4, ..., 10, 11, or 12. Not all values of $T$ are equally likely; we can describe the probability distribution of $T$ by the following table: Of course the logic behind a typical entry, e.g. the probability $\frac{1}{36}$ of a 4 showing up is based on the fact that there are 3 ways (1, 3), (2, 2), and (3, 1) to obtain a 4. And each of these has probability $\frac{1}{36}$. One writes this more formally as

$$P(T = 4) = \frac{3}{36}$$

Actually the random variable $T$ can be thought of as the sum of two independent random variables $U$ and $V$, each describing the number from 1 to 6 which shows up on one of the die, $U$ describes one of the die, and $V$ the other. We call these random variables independent because knowing which value of $U$ shows up (e.g. $U = 3$) doesn’t affect the probability of any value of $V$ showing up. ($V$ keeps track of the other die.)
Even though they are independent, $U$ and $V$ have the same probability distribution, namely

**Example 2**

<table>
<thead>
<tr>
<th>U or V</th>
<th>Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1/6</td>
</tr>
<tr>
<td>2</td>
<td>1/6</td>
</tr>
<tr>
<td>3</td>
<td>1/6</td>
</tr>
<tr>
<td>4</td>
<td>1/6</td>
</tr>
<tr>
<td>5</td>
<td>1/6</td>
</tr>
<tr>
<td>6</td>
<td>1/6</td>
</tr>
</tbody>
</table>

So we can think of $T = U + V$ as random variables. What does that get us?

2. **MEAN AND VARIANCE OF RANDOM VARIABLES**

Random variables have a *mean* and a *variance*. If the probability distribution of $X$ is as given in the following table then the mean $\mu_X$ (or expectation $E(X)$) is defined by $\mu_X = \sum p_i x_i$. When only one random variable is present, we may drop the subscript $X$ and just denote the mean as $\mu$.

The variance of $X$ is defined by $Var(X) = \sum p_i (x_i - \mu_X)^2$. As with data, there is also a standard deviation $\sigma_X$ whose square is $Var(X)$.

As a simple example, consider the random variable $U$ of Example 2 above. $U$ describes the experiment of rolling one fair die. Then

$$
\mu_U = \frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3.5
$$

and

$$
Var(U) = \frac{1}{6}((1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2) = 2.917
$$

And the standard deviation $\sigma_U = \sqrt{2.917} = 1.708$.

What is the significance of these three numbers for random variables? Basically they predict the behavior we will typically see if we roll a die many times. For example the following table summarizes one experiment of rolling a die 1000 times: For this data set, the sample mean $\bar{x} = 3.597$. Not too far from the random variable mean of 3.5. Similarly the standard deviation for the data set is $s = 1.709$ vs. a standard deviation of 1.708 for the random variable. Very close! Of course the variances also match very well.

In fact, it is a consequence of the Law of Large Numbers, that as the number of simulations gets large, with probability 1, the data collected from the simulations will have a *sample mean* and *sample standard deviation* that approach the *random variable mean* and *random variable standard deviation* respectively.
### 3. Operations on Random Variables

Given random variables $X$ and $Y$ on the same sample space, and a constant $c$, we can form new random variables $X + Y$, $cX$, and $X + c$.

A typical example of adding random variables arises from the random variables $U$, $V$, and $X$ discussed in examples 1 and 2 above. The sample space consists of 36 equally probable outcomes for the rolling of two dice. $X$ describes the sum, $U$ the result on one die, and $V$ the result on the other. So $X = U + V$.

Means behave very simply with respect to these operations:

\[
\begin{align*}
\mu_{X+Y} &= \mu_X + \mu_Y \\
\mu_{X+c} &= \mu_X + c \\
\mu_{cX} &= c\mu_X
\end{align*}
\]

For example, this says that because the mean of rolling one die is 3.5, it follows that the mean of the sum in rolling two dice is $3.5 + 3.5 = 7$.

For the operation of adding or multiplying by a constant $c$, means and variances also behave simply.

\[
\begin{align*}
\sigma_{cX} &= |c|\sigma_X \\
\text{Var}(cX) &= c^2\text{Var}(X) \\
\sigma_{c+X} &= \sigma_X \\
\text{Var}(c + X) &= \text{Var}(X)
\end{align*}
\]

What does an operation like multiplication by $c$ represent? Consider two gambling games. Both involve the rolling of one die. In game 1, there is a payoff in dollars of the number which appears on the roll. The random variable $U$ describes this payoff. In game 2, imagine the payoff is $10$ times the roll. For example $20$ if a 2 is rolled. This game is described by the random variable $10U$. No surprise that the second game has 10 times the mean and standard deviation of the first.

The operation of addition of a constant is also natural. Imagine game 1 above is altered so that you have to pay $4$ to play it. You’ll then be paid 1, 2, or … 6 dollars depending on your roll. Counting your $4$ admission ticket, your overall return is described by the random variable $U - 4$. Its mean is $4$ less than that of $U$, but its standard deviation is unchanged.

The operation of addition of random variables turns out to be very important for statistical reasoning. That’s because operations like averaging a bunch of results have addition as key components.
The variance of the sum of two random variables is much more complicated than the others we have discussed in this section. For example \( \text{Var}(X + X) = \text{Var}(2X) = 4\text{Var}(X) \). But in general, unless we introduce a new definition (the covariance of two random variables usually), there is no real formula for expressing \( \text{Var}(X + Y) \) in terms of \( \text{Var}(X) \) and \( \text{Var}(Y) \).

But in one case there is. Suppose \( X \) and \( Y \) are independent random variables. Recall this means that knowledge of what value \( X \) takes does not affect the probability distribution of \( Y \). For independent random variables
\[
\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)
\]

This is for example the key to deriving the formula for the standard deviation of a random variable following the binomial distribution \( \text{binomial}(n,p) \). Such a random variable \( X \) can be thought of as the sum of \( n \) independent Bernoulli random variables \( W_1, W_2, \ldots, W_n \) where each \( W_i \) has the probability distribution below:

<table>
<thead>
<tr>
<th>( W_i )</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1-p</td>
</tr>
<tr>
<td>1</td>
<td>p</td>
</tr>
</tbody>
</table>

It is easily seen that such a random variable has mean \( p \) and variance \( p(1-p) \). Hence if you add \( n \) independent copies of these together (as in counting the number of heads in tossing \( n \) potentially unfair coins), the mean number of heads becomes \( np \), and the variance \( np(1-p) \). The square root of the latter is the standard deviation of a binomial random variable.

For example these formulas for the mean and standard deviation of a binomial distribution tell us which normal distribution \( \mathcal{N}(\mu, \sigma) \) we can often use as an approximation.

4. Exercises

These are not to be handed in unless specifically assigned.

1) Suppose we roll 3 fair dice. What would the mean and standard deviation of the random variable describing the sum of the values on the 3 dice be?

2) Verify the calculation of the mean and standard deviation of the Bernoulli random variables \( W_i \) above.

3) Same question as 1) but with 10 dice. We’ll learn later that often values more than 3 standard deviations from the mean are very uncommon. If that applied to the case of 10 dice (in fact it is not clear that it should . . . ), how big a sum would be the cutoff for “very uncommon” in this sense?

4) Suppose two fair dice (one black, one white) are rolled and you are to receive an amount of money in dollars equal to four times the black number plus three times the white number. What would be a “fair” price to pay for this gambling game? What would the standard deviation of your payment after one play be?
5. Random Variables — More Advanced

In probability, a random variable is defined as a number valued function with domain a sample space \( S \). Assuming that the sample space has a probability assignment on it, we can figure out the probability of each value showing up as the value of the random variable, and so obtain the probability distribution of the random variable.

For example, the random phenomenon of rolling two dice might be described by the sample space

\[
S = \{(1, 1), (1, 2), (1, 3), (1, 4), (1, 5), (1, 6),
(2, 1), (2, 2), (2, 3), (2, 4), (2, 5), (2, 6),
(3, 1), (3, 2), (3, 3), (3, 4), (3, 5), (3, 6),
(4, 1), (4, 2), (4, 3), (4, 4), (4, 5), (4, 6),
(5, 1), (5, 2), (5, 3), (5, 4), (5, 5), (5, 6),
(6, 1), (6, 2), (6, 3), (6, 4), (6, 5), (6, 6)\}
\]

where e.g. the outcome \((3, 4)\) means the first die came out to 3 and the second to 4.

In the notation of our introductory examples, for the outcome \((a, b)\), \(U_1 = a\), \(U_2 = b\), and \(T = a + b\). This is just what viewing the function \(T\) as the sum of the functions \(U_1\) and \(U_2\) would suggest.

Comparing the relatively complicated probability distributions of \(T\) to the simpler ones of \(U_1\) and \(U_2\) shows us that it is a bit complicated to work out the probability distribution for the sum of even independent random variables. But for the other two operations we study, the distributions are much simpler. Below are the probability distributions of \(10U\) and \(U - 4\) respectively. The probabilities are the same as those in the probability distribution of \(U\) but each value has been multiplied by 10 or reduced by 4 respectively.

**The Probability Distribution of 10\(U\)**

<table>
<thead>
<tr>
<th>10(U)</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1/36</td>
</tr>
<tr>
<td>20</td>
<td>1/36</td>
</tr>
<tr>
<td>30</td>
<td>1/36</td>
</tr>
<tr>
<td>40</td>
<td>1/36</td>
</tr>
<tr>
<td>50</td>
<td>1/36</td>
</tr>
<tr>
<td>60</td>
<td>1/36</td>
</tr>
</tbody>
</table>

**The Probability Distribution of \(U - 4\)**
<table>
<thead>
<tr>
<th>$U - 4$</th>
<th>probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3</td>
<td>1/3</td>
</tr>
<tr>
<td>-2</td>
<td>1/3</td>
</tr>
<tr>
<td>-1</td>
<td>1/3</td>
</tr>
<tr>
<td>0</td>
<td>1/9</td>
</tr>
<tr>
<td>1</td>
<td>1/9</td>
</tr>
<tr>
<td>2</td>
<td>1/9</td>
</tr>
</tbody>
</table>