3.4 The Expected Value of a Function of a Random Variable

Often times in the real world, we wish to look at a function of our variables – or transform them. Changing feet into meters, Fahrenheit into centigrade, etc. In this section we seek to determine the expected value of this new variable. If we transform a random variable X by some function $Y = g(X)$, it should be clear that Y itself is a random variable.

Suppose that we have a discrete random variable X with pmf given below. We now let $Y = 2X + 5$. What is the expected value of Y? By definition $E[Y] = \sum y f_Y(y)$. So to determine the $E[Y]$, we will find $f_Y(y)$ and then compute $E[Y] = \sum y f_Y(y)$.

x	$f_X(x)$
-2	1/8
-1	1/8
0	2/8
1	1/8
2	1/8
3	1/8
4	1/8

$Y = 2X + 5$ $\longrightarrow$

y	$f_Y(y)$
1	1/8
3	1/8
5	2/8
7	1/8
9	1/8
11	1/8
13	1/8

So, $E[Y] = \sum y f_Y(y) = (1+3+7+9+11+13)\left(\frac{1}{8}\right) + 5\left(\frac{2}{8}\right) = \frac{54}{8}$.

We also note that,

$$E[Y] = \sum y f_Y(y) = \sum_{x=-2}^{4} (2x+5) f_X(x) = \sum_{x \in Support(X)} g(x) f_X(x)$$

> **Theorem:** *Given a discrete random variable X and the linear transformation $Y = g(X) = aX + b$,*
>
> $$E[Y] = E[g(X)] = \sum_{x \in Support(X)} g(x) f_X(x).$$

Proof: *Since $Y = g(X) = aX + b$ is a 1 to 1 function, this will be rather straight forward.*

$$E[Y] = \sum y f_Y(y) = \sum (ax+b) P(Y = ax+b) = \sum (ax+b) P(X = x) = \sum g(x) f_X(x)$$

Note that the argument above would work with any function $Y = g(X)$ that is 1 to 1.

> **Theorem:** *Given a discrete random variable X and the linear transformation $Y = g(X) = aX + b$,*
> $E[Y] = E[aX + b] = aE[X] + b$.

Proof: $E[aX + b] = \sum (ax+b) f(x) = \sum (ax) f(x) + \sum (b) f(x)$

$$a \sum x f(x) + b \sum f(x) = aE[X] + b = a\mu_X + b$$

Thus, the expected value of a random variable is a **linear operator**.

Example: *If a random variable X has mean $\mu = 4$, determine the mean of Y, where $Y = 2X + 3$*

$$\mu_Y = 2\mu_X + 3 = 11 \text{ or } E[Y] = E[2X+3] = 2E[X] + 3 = 11$$

If we think about this last example, it seems like a no brainer that it is true. Suppose that the average score is 4. I now double everybody's score. The new average is 8. I now add 3 to everybody's already doubled score. The new average is 11.

In this linear case, notice if we consider $Y = g(X) = 2X + 3$, then $E[Y] = E[g(X)] = 2X + 3 = g(E[X])$. We would naturally wonder if that statement always holds. That is, will it always be true that If $Y = g(X)$ then $E[Y] = g(E[X])$? The answer is no, this is not always true. **How will we show this?**

Suppose that we have a discrete random variable X with pmf given below. The expected value of X can be determined by $\sum xf(x) = 7/8$. We now let $Y = X^2$. What is the expected value of Y? By definition $E[Y] = \sum y f_Y(y)$. So to determine the $E[Y]$, we will find $f_Y(y)$ and then compute $E[Y] = \sum y f_Y(y)$.

x	$f_X(x)$	$xf_X(x)$
-2	1/8	-2/8
-1	1/8	-1/8
0	2/8	0
1	1/8	1/8
2	1/8	2/8
3	1/8	3/8
4	1/8	4/8

$\xrightarrow{Y = X^2}$

y	$f_Y(y)$	$yf_Y(y)$
4	1/8	4/8
1	1/8	1/8
0	2/8	0
1	1/8	1/8
4	1/8	4/8
9	1/8	9/8
16	1/8	16/8

$=$

y	$f_Y(y)$	$yf_Y(y)$
0	2/8	0
1	2/8	2/8
4	2/8	8/8
9	1/8	9/8
16	1/8	16/8

$E[X] = 7/8$

$E[Y] = 35/8$

We note that $E^2[X] = (7/8)^2 \neq 35/8 = E[Y] = E[X^2]$. So, we have our counter example and in general, it is not true that $E[Y] = g(E[X])$. We do however see something helpful when looking at the first two groupings in the above table. We see that the probabilities are identical and that each y-value is the square of the x-value. This leads us to the following theorem that will help us determine $E[Y]$:

Theorem: *Given a discrete random variable X and a function $Y = g(X)$, $E[Y] = E[g(X)] = \sum g(x) \cdot f_X(x)$.*

The theorem is much bigger than it seems. It allows us to determine the expected value of a transformation of a random variable without determining the distribution of the new random variable. We easily found $f_Y(y)$ in the above problem. With continuous random variables, it can often be difficult to determine $f_Y(y)$, so this will be huge when we get to continuous random variables.

3.5 Some Important Functions of a Random Variable (Moments and Variance)

Definition: *The **Variance** of a discrete random variable X is defined by the formula* $\sigma^2 = \sum_{x \in Support} (x-\mu)^2 f(x)$ *(provided the sum converges). Alternatively,* $\sigma^2 = E[(x-\mu)^2]$.

Definition: *The Standard Deviation of a random variable, denoted by σ, is defined by* $\sigma = \sqrt{\sigma^2}$

Note that we can consider the function $Y = g(X) = (x-\mu)^2$. Then the variance is just $E[g(X)] = E[Y]$.

In a data set, we use the standard deviation of the data set to denote a measure of the variation of the data. The standard deviation of a random variable does the same thing. It gives us a measure of how spread out the support is with respect to the associated probabilities.

Example: *Determine the standard deviation for rolling a single die.*

$$\sigma^2 = \sum_{x=1}^{6}(x-3.5)^2 \frac{1}{6} = \frac{1}{6}\left((1-3.5)^2 + (2-3.5)^2 + \cdots + (6-3.5)^2\right) = \frac{17.5}{6} \approx 2.916667$$. So $\sigma = 1.7078$.

Theorem: $\sigma^2 = E[(x-\mu)^2] = E[X^2] - E^2[X]$

Proof: $E[(x-\mu)^2] = \sum (x-\mu)^2 f(x) = \sum (x^2 - 2x\mu + \mu^2) f(x)$

$$= \sum x^2 f(x) - 2\mu \sum x f(x) + \mu^2 \sum f(x)$$

$$= E[X^2] - 2\mu^2 + \mu^2 = E[X^2] - \mu^2 = E[X^2] - E^2[X]$$

Example: *Recalculate the variance of a single die using the theorem.*

$$\sigma^2 = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) - 3.5^2 = \frac{91}{6} - 3.5^2 \approx 2.916667$$

In general, this theorem allows a much easier calculation than the definition. Soon, we will see problems where this formula is a necessity.

Not all random variables have a finite variance. We saw earlier that we can have a random variable that does not have a finite mean. We can also have a random variable with finite mean, but not finite variance.

$f(x)=\frac{k}{x^3}$ for $x=1,2,3,\cdots$, where k is the value that makes $f(x)$ a pmf. That is, so that $\sum_{x=1}^{\infty}\frac{k}{x^3}=1$. This random variable has finite mean since $E[X]=\sum_{x=1}^{\infty}x\frac{k}{x^3}=\sum_{x=1}^{\infty}\frac{k}{x^2}=\frac{k\pi^2}{6}$, which is finite. To determine the variance of our random variable, we need to determine $E[X^2]$. $E[X^2]=\sum_{x=1}^{\infty}x^2\frac{k}{x^3}=\sum_{x=1}^{\infty}\frac{k}{x}$ which diverges. So, this random variable has finite mean but not finite variance.

Theorem: *Given a discrete random variable X and the linear transformation $Y=g(X)=aX+b$,*
$Var[Y]=Var[aX+b]=a^2Var[X]$. Or, we can write $\sigma^2_{aX+b}=a^2\sigma^2_X$

Proof: $Var[(aX+b)]=E[\{(aX+b)-E[aX+b]\}^2]=E[\{(aX+b)-aE[X]-b]\}^2]$

$$
\begin{aligned}
&=E[(aX-aE[X])^2]=E[a^2(X-E[X])^2]\\
&=a^2E[(X-E[X])^2]\\
&=a^2Var[X]
\end{aligned}
$$

Exercise: Given a random variable X, with variance $\sigma^2_X=8$, determine the variance of $Y=3X-5$.

Exercise: Given a random variable X, with variance $\sigma^2_X=8$, determine the variance of $Y=5-3X$.

The $E[X]$ is often called the **first moment** of X and $E[X^2]$ is called the **second moment** of X.

Definition: *The* ***kth moment*** *of X is defined to be* $E[X^k]$

These moments are important features of a random variable. Sometimes they are called the moments about the origin and then moments about the mean would be defined as $E[(X-\mu)^k]$.