Chapter 7

Mean and Variance Of a Discrete Random Variable

References: Pruim 2.5

Probability mass functions provide a global overview of the behavior of a random variable. However, it is often helpful to focus the information contained in the PMF by summarizing certain features. In this chapter, we develop two of the most important summaries of a random variable: its mean and variance.

7.1 The Mean of a Discrete Random Variable

Example 7.1. Suppose a student has taken 10 courses and received 5 A's, 4 B's and 1 C. Using the traditional numerical scale where an A is worth 4, B is worth 3 and a C is worth 2, what is this student's grade point average (GPA)?

. . . .

The grade point average in example 7.1 is a weighted sum of the grades:

$$GPA = \sum (\text{grade})(\text{proportion of times student gets that grade}).$$

The mean of a random variable is a similar weighted sum with the probabilities serving as weights.

The expected value or expectation of a random variable is merely its average value, where we speak of "average" value as a weighted average with the probabilities as weights.

We will now look at the definition and examples of calculating the mean for discrete random variables.

Definition 7.1. Let X be a discrete random variable with PMF f. The mean (also called expected value) of X is denoted as μ or $E[X]$ and is defined by

$$E[X] = \sum_x x \cdot P(X = x) = \sum_x x \cdot f(x).$$

The sum is taken over all the possible values of X. When the possible values are infinite, we require that the sum is well defined and is finite. If $E[X] = \pm\infty$, we simply say that $E[X]$ does not exist.

. . . .

Example 7.2. Find $E[X]$ where X is the outcome when we roll a fair die.

. . . .

The mean of a random variable is not necessarily a possible value of the random variable. It represents a long-run average.

We will learn easier ways to calculate the expected value later, but for now the definition is the only tool we have. Let's use it to calculate the expected value of a binomial random variable.

Theorem 7.1. *Let $X \sim Binom(n, \pi)$. Then*

$$E[X] = n\pi.$$

Proof. Before diving into the proof, let's be mindful of a couple of useful facts: First, for any $x = 1, 2, \ldots, n$, we have the result:

$$x\binom{n}{x} = n\binom{n-1}{x-1}.$$

Second, the binomial PMF sums to 1:

$$\sum_{x=0}^{n}\binom{n}{x}\pi^x(1-\pi)^{n-x} = 1.$$

Now, applying the definition of the expected value, we have:

$$\begin{aligned}
E[X] &= \sum_{x=0}^{n} x \cdot f(x), \\
&= \sum_{x=0}^{n} x\binom{n}{x}\pi^x(1-\pi)^{n-x}, \\
&= \sum_{x=1}^{n} x\binom{n}{x}\pi^x(1-\pi)^{n-x}, \\
&= n\pi\sum_{x=1}^{n}\binom{n-1}{x-1}\pi^{x-1}(1-\pi)^{n-x}, \quad set y = x-1 \\
&= n\pi\sum_{y=0}^{n-1}\underbrace{\binom{n-1}{y}\pi^y(1-\pi)^{n-1-y}}_{\text{PMF of a } Binom(n-1,\pi)} \\
&= n\pi.
\end{aligned}$$

$\square$

Suppose in Example 7.2, we are interested in calculating $E[X+3]$. It seems plausible that the answer should be $E[X]+3$, since we are just adding 3 to the number that results from rolling the die. The definition below formalizes this idea and is stated for any arbitrary linear transformation of X.

Lemma 7.1 (Linearity of Expected Values)**.** *Let X be a discrete random variable, let a and b be constants and let $Y = aX + b$. Then Y is a discrete random variable and*

$$E[Y] = E[aX+b] = aE[X] + b.$$

Proof. Suppose X takes values $x_1, x_2, x_3, \ldots$ and $f(x_1)$, $f(x_2)$, $f(x_3)$ and so on, are the probabilities.

Then Y takes values $ax_1 + b, ax_2 + b, ax_3 + b, \ldots$ and

$$P(Y = ax_i + b) = P(X = x_i) = f(x_i).$$

Therefore, the expected value of Y can now be found as follows:

$$\begin{aligned} E[Y] &= \sum_x (ax+b) \cdot f(x), \\ &= a\sum_x x \cdot f(x) + b\sum_x f(x), \\ &= aE[X] + b. \end{aligned}$$

□

In the following example, we use the linearity of expected value in a binomial setting.

. . . .

Example 7.3. A typical day's production of a certain electronic component is twelve. The probability that one of these components needs rework is 0.11. Each component needing rework costs \$100. What is the average daily cost for defective components?

. . . .

Lemma 7.1 and example 7.3 illustrate that we do not need to find the PMF of Y in order to find it's expectation. In fact, this idea applies to any transformation of X, not just a linear one, and is stated below.

Lemma 7.2 (Law of the Unconscious Probabilist)**.** *Let X be a discrete random variable with PMF f and let $t(X)$ be a transformation of X for some function t. Then $Y = t(X)$ is a discrete random variable and*

$$E[Y] = E[t(X)] = \sum_x t(x) f(x).$$

The result in Lemma 7.2 is actually quite intuitive since $t(X)$ will equal $t(x)$ whenever $X = x$; it seems reasonable that $E[t(X)]$ should just be a weighted average of the values $t(x)$ with the weights being equal to the probability that $X = x$.

. . . .

Example 7.4. Suppose X is a discrete random variable with PMF as shown in the following table.

x	-2	-1	0	1	2
probability	0.2	0.1	0.4	0.1	0.2

Use Lemma 7.2 to calculate $E\left[X^2\right]$.

. . . .

7.1.1 Practice Problems

1. Suppose you flip two fair coins. Let $X = 3$ if both coins show the same result, and $X = 5$ otherwise. Calculate $E[X]$.
2. Tay Sachs is a rare, but fatal disease. If a couple are both carriers of the disease, then their children have a 25% chance each of being born with it, independently of the others. Suppose such a couple has three children. How many of their 3 kids can be expected to be born with Tay Sachs? (Hint: Let X denote the number of their children who are born with Tay Sachs. What is the distribution of X?)
3. Suppose X has PMF as shown in the table below. Define the new random variable: $Y = c^X$. Find $c \neq 1$ such that $E[Y] = 1$.

x	$f(x)$
1	3/4
-1	1/4

7.2 Variance of a Discrete Random Variable

The mean or expected value is a one number summary that describes a typical value of an observation of the random variable. We can think of it as our "educated" guess for the value of the random variable. However, it does not tell us anything about the variation or spread of the values.

For example, consider two random variables X and Y with PMFs as shown in the table below. Both have a mean or expected value of 3. The range of each is also the same, however, there is much less spread in the possible values of X in the sense that it takes on values around the mean with higher probabilities than does Y.

x	1	2	3	4	5
f_X	0.1	0.2	0.4	0.2	0.1
f_Y	0.3	0.1	0.2	0.1	0.3

One way to quantify the spread of a random variable is by how far away X would be from its mean, on average. The variance provides such a measure of spread.

Definition 7.2. Let X be a discrete random variable with mean μ. The **variance** of X, assuming it exists, is defined by

$$\sigma_X^2 = Var[X] = E\left[(X-\mu)^2\right] = \sum_x (x-\mu)^2 \times f(x).$$

The **standard deviation** is the positive square root of the variance.

An alternative formula for $Var[X]$ that is not as intuitive as the definition, but is often easier to use in practice is derived below:

$$\begin{aligned}
Var[X] &= E\left[(X-\mu)^2\right], \\
&= \sum_x (x-\mu)^2 \cdot f(x), \\
&= \sum_x (x^2 - 2x\mu + \mu^2) f(x), \\
&= \sum_x x^2 f(x) - \sum_x 2x\mu f(x) + \sum_x \mu^2 f(x) \\
&= E\left[X^2\right] - 2\mu \sum_x x f(x) + \mu^2 \sum_x f(x), \\
&= E\left[X^2\right] - 2\mu \cdot \mu + \mu^2 \cdot 1, \\
&= E\left[X^2\right] - \mu^2.
\end{aligned}$$

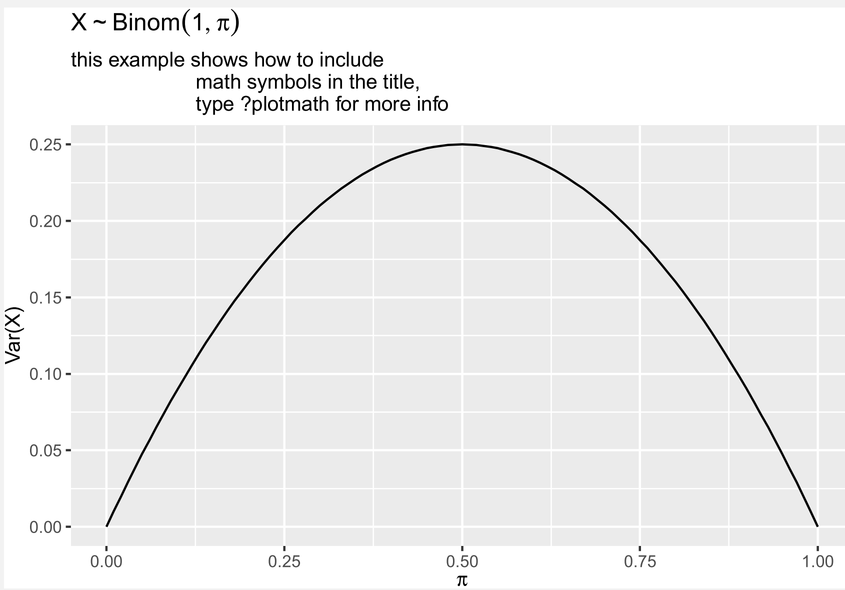
. . . .

Example 7.5. Calculate $Var[X]$ if X represents the outcome when a fair die is rolled.

Example 7.6. If $X \sim Binom(1, \pi)$, then X is called a **Bernoulli** random variable. For a Bernoulli random variable, what are $E[X]$ and $Var[X]$?

As a function of π, $Var(X)$ in example 7.6 is quadratic. Its graph is a parabola that opens downward and the largest variance occurs when $\pi = \frac{1}{2}$. The following code chunk describes how you can use ggplot to make plots of functions. In addition, it demonstrates the use of the `expression` function to add greek symbols in the axis labels and title.

```
ggplot() +
  geom_function(fun = function(x){x*(1-x)}, xlim = c(0,1) ) +
  labs(x = expression(pi),
          y = "Var(X)",
          title = expression(X %~% Binom(1, pi)),
          subtitle = "this example shows how to include
                      math symbols in the title,
                      type ?plotmath for more info")
```



. . . .

In the previous section we encountered a change of scale that can be applied to expected values. A similar result arises in connection with the variance of a linear transformation, the proof of which will be left as an exercise.

Lemma 7.3. *Let X be a discrete random variable and suppose a and b are constants. Then*

$$Var\left[aX+b\right]=a^2Var\left[X\right].$$

. . . .

Example 7.7. Suppose $E(X) = \mu$ and the standard deviation of X equals σ. Find the expected value and standard deviation of the random variable

$$Y = (X - \mu)/\sigma.$$

The fact that

$$Z = (X - \mu)/\sigma$$

has mean 0 and standard deviation 1 is the reason it is often referred to as the standardization of X.

. . . .

We conclude this section with an inequality which demonstrates the usefulness of the variance as a measure of spread.

Lemma 7.4 (Chebyshev's Inequality)**.** *For any random variable X with mean μ and standard deviation σ:*

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2},$$

where k is some positive number.

Proof Before we provide a heuristic proof of the inequality, let's make sure we understand the event whose probability we are calculating. When we say $|X - \mu|$ exceeds a cutoff of $k\sigma$, we mean that either

$$(X - \mu) \geq k\sigma \Rightarrow X \geq \mu + k\sigma$$

or

$$(X - \mu) \leq -k\sigma \Rightarrow X \leq \mu - k\sigma.$$

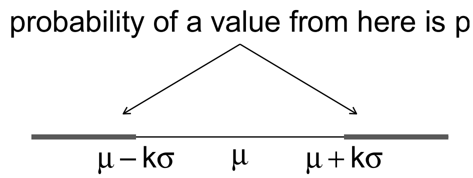
That is, Chebyshev's inequality guarantees that at most $\frac{1}{k^2}$ of the distribution can be k or more standard deviations from the mean[1]

[1] Wikipedia

Now, on to the proof. Let us denote

$$p = P(|X - \mu| \geq k\sigma). \tag{7.1}$$

For any X which satisfies equation (7.1) we have the following:



Now consider the following PMF which satisfies equation (7.1):

value	$\mu - k\sigma$	μ	$\mu + k\sigma$
probability	$\frac{p}{2}$	$1 - p$	$\frac{p}{2}$

The crux of the argument is that this PMF yields the smallest variance that is possible for any distribution that satisfies (7.1).

It is straightforward to check that the variance of this distribution is $pk^2\sigma^2$ and therefore

$$Var(X) = \sigma^2 \geq pk^2\sigma^2$$

from which it follows that

$$p = P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}.$$

. . . .

Example 7.8. What does Chebyshev's inequality say about the probability that a random variable is at least 2 standard deviations from its mean? Calculate the exact probability for the outcome, X, when a fair die is rolled.

Chebyshev's inequality only provides informative answers for values of k greater than 1. In addition, the bound $\frac{1}{k^2}$ is usually much larger than the exact probability for many distributions.

7.2.1 Practice Problems

1. Suppose X has PMF

x	2	3	4	5
$f(x)$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

a. Find $E[X]$.
b. Find $Var[X]$.

2. Suppose $E[X] = 1$ and $Var[X] = 5$. Find $E[(2+X)^2]$.

3. Suppose that X is a Bernoulli random variable, that is, $X \sim Binom(1, \pi)$. If $E[X] = 3\,Var[X]$, find π.