Variance of Discrete Random Variables
Class 5, 18.05
Jeremy Orloff and Jonathan Bloom

1 Learning Goals

1. Be able to compute the variance and standard deviation of a random variable.
2. Understand that standard deviation is a measure of scale or spread.
3. Be able to compute variance using the properties of scaling and linearity.

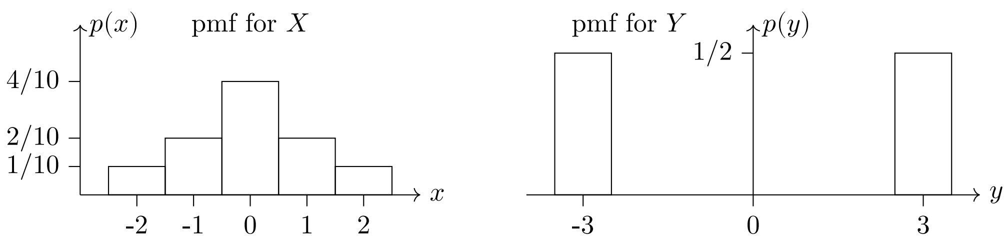
2 Spread

The expected value (mean) of a random variable is a measure of location or central tendency. If you had to summarize a random variable with a single number, the mean would be a good choice. Still, the mean leaves out a good deal of information. For example, the random variables X and Y below both have mean 0, but their probability mass is spread out about the mean quite differently.

values X	-2	-1	0	1	2
pmf $p(x)$	1/10	2/10	4/10	2/10	1/10

values Y	-3	3
pmf $p(y)$	1/2	1/2

It's probably a little easier to see the different spreads in plots of the probability mass functions. We use bars instead of dots to give a better sense of the mass.



pmf's for two different distributions both with mean 0

In the next section, we will learn how to quantify this spread.

3 Variance and standard deviation

Taking the mean as the center of a random variable's probability distribution, the variance is a measure of how much the probability mass is spread out around this center. We'll start with the formal definition of variance and then unpack its meaning.

Definition: If X is a random variable with mean $E[X] = \mu$, then the variance of X is defined by

$$\text{Var}(X) = E[(X-\mu)^2].$$

The standard deviation σ of X is defined by

$$\sigma = \sqrt{\mathrm{Var}(X)}.$$

If the relevant random variable is clear from context, then the variance and standard deviation are often denoted by σ^2 and σ ('sigma'), just as the mean is μ ('mu').

What does this mean? First, let's rewrite the definition explicitly as a sum. If X takes values $x_1, x_2, \ldots, x_n$ with probability mass function $p(x_i)$ then

$$\mathrm{Var}(X) = E[(X-\mu)^2] = \sum_{i=1}^{n} p(x_i)(x_i-\mu)^2.$$

In words, the formula for $\mathrm{Var}(X)$ says to take a weighted average of the squared distance to the mean. By squaring, we make sure we are averaging only non-negative values, so that the spread to the right of the mean won't cancel that to the left. By using expectation, we are weighting high probability values more than low probability values. (See Example 2 below.)

Note on units:
1. σ has the same units as X.
2. $\mathrm{Var}(X)$ has the same units as the square of X. So if X is in meters, then $\mathrm{Var}(X)$ is in meters squared.
Because σ and X have the same units, the standard deviation is a natural measure of spread.

Let's work some examples to make the notion of variance clear.

Example 1. Compute the mean, variance and standard deviation of the random variable X with the following table of values and probabilities.

value x	1	3	5
pmf $p(x)$	1/4	1/4	1/2

Solution: First we compute $E[X] = 7/2$. Then we extend the table to include $(X-7/2)^2$.

value x	1	3	5
$p(x)$	1/4	1/4	1/2
$(x-7/2)^2$	25/4	1/4	9/4

Now the computation of the variance is similar to that of expectation:

$$\mathrm{Var}(X) = \frac{25}{4}\cdot\frac{1}{4} + \frac{1}{4}\cdot\frac{1}{4} + \frac{9}{4}\cdot\frac{1}{2} = \frac{11}{4}.$$

Taking the square root we have the standard deviation $\sigma = \sqrt{11/4}$.

Example 2. For each random variable X, Y, Z, and W plot the pmf and compute the mean and variance.

(i)

value x	1	2	3	4	5
pmf $p(x)$	1/5	1/5	1/5	1/5	1/5

(ii)

value y	1	2	3	4	5
pmf $p(y)$	1/10	2/10	4/10	2/10	1/10

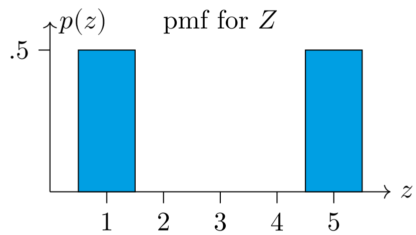
(iii)

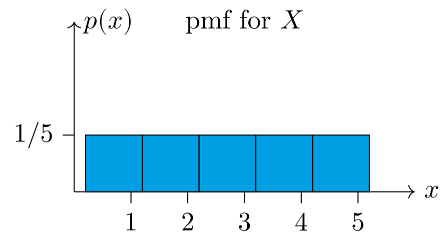
value z	1	2	3	4	5
pmf $p(z)$	5/10	0	0	0	5/10

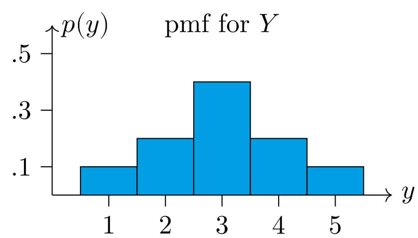
(iv)

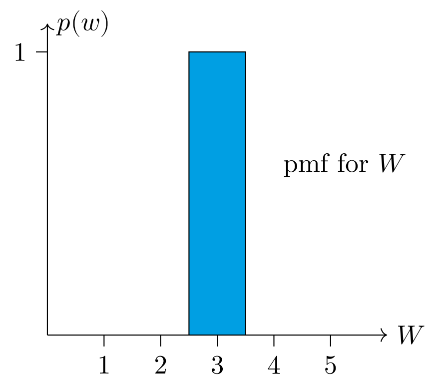
value w	1	2	3	4	5
pmf $p(w)$	0	0	1	0	0

Solution: Each random variable has the same mean 3, but the probability is spread out differently. In the plots below, we order the pmf's from largest to smallest variance: Z, X, Y, W.









Next we'll verify our visual intuition by computing the variance of each of the variables. All of them have mean $\mu = 3$. Since the variance is defined as an expected value, we can compute it using the tables.

(i)

value x	1	2	3	4	5
pmf $p(x)$	1/5	1/5	1/5	1/5	1/5
$(X-\mu)^2$	4	1	0	1	4

$\text{Var}(X) = E[(X-\mu)^2] = \frac{4}{5} + \frac{1}{5} + \frac{0}{5} + \frac{1}{5} + \frac{4}{5} = \boxed{2}$.

(ii)

value y	1	2	3	4	5
$p(y)$	1/10	2/10	4/10	2/10	1/10
$(Y-\mu)^2$	4	1	0	1	4

$\text{Var}(Y) = E[(Y-\mu)^2] = \frac{4}{10} + \frac{2}{10} + \frac{0}{10} + \frac{2}{10} + \frac{4}{10} = \boxed{1.2}$.

(iii)

value z	1	2	3	4	5
pmf $p(z)$	5/10	0	0	0	5/10
$(Z-\mu)^2$	4	1	0	1	4

$\text{Var}(Z) = E[(Z-\mu)^2] = \frac{20}{10} + \frac{20}{10} = \boxed{4}$.

(iv)

value w	1	2	3	4	5
pmf $p(w)$	0	0	1	0	0
$(W-\mu)^2$	4	1	0	1	4

$\text{Var}(W) = \boxed{0}$. Note that W doesn't vary, so it has variance 0!

3.1 The variance of a Bernoulli(p) random variable.

Bernoulli random variables are fundamental, so we should know their variance.

If $X \sim \text{Bernoulli}(p)$ then

$$\text{Var}(X) = p(1-p).$$

Proof: We know that $E[X] = p$. We compute $\text{Var}(X)$ using a table.

values X	0	1
pmf $p(x)$	$1-p$	p
$(X-\mu)^2$	$(0-p)^2$	$(1-p)^2$

$\text{Var}(X) = (1-p)p^2 + p(1-p)^2 = (1-p)p(1-p+p) = \boxed{(1-p)p.}$

As with all things Bernoulli, you should remember this formula.

Think: For what value of p does Bernoulli(p) have the highest variance? Try to answer this by plotting the PMF for various p.

3.2 A word about independence

So far we have been using the notion of independent random variable without ever carefully defining it. For example, a binomial distribution is the sum of independent Bernoulli trials. This may (should?) have bothered you. Of course, we have an intuitive sense of what independence means for experimental trials. We also have the probabilistic sense that random variables X and Y are independent if knowing the value of X gives you no information about the value of Y.

In a few classes we will work with continuous random variables and joint probability functions. After that we will be ready for a full definition of independence. For now we can use the following definition, which is exactly what you expect and is valid for discrete random variables.

Definition: The discrete random variables X and Y are independent if

$$P(X = a, Y = b) \; = \; P(X = a)P(Y = b)$$

for any values a, b. That is, the probabilities multiply.

3.3 Properties of variance

The three most useful properties for computing variance are:

1. If X and Y are independent then $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

2. For constants a and b, $\text{Var}(aX + b) = a^2\text{Var}(X)$.

3. $\text{Var}(X) = E[X^2] - E[X]^2$.

For Property 1, note carefully the requirement that X and Y are independent. We will return to the proof of Property 1 in a later class.

Property 3 gives a formula for $\text{Var}(X)$ that is often easier to use in hand calculations. The computer is happy to use the definition! We'll prove Properties 2 and 3 after some examples.

Example 3. Suppose X and Y are independent and $\text{Var}(X) = 3$ and $\text{Var}(Y) = 5$. Find:
(i) $\text{Var}(X + Y)$, (ii) $\text{Var}(3X + 4)$, (iii) $\text{Var}(X + X)$, (iv) $\text{Var}(X + 3Y)$.

Solution: To compute these variances we make use of Properties 1 and 2.
(i) Since X and Y are independent, $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) = 8$.

(ii) Using Property 2, $\text{Var}(3X + 4) = 9 \cdot \text{Var}(X) = 27$.

(iii) Don't be fooled! Property 1 fails since X is certainly not independent of itself. We can use Property 2: $\text{Var}(X + X) = \text{Var}(2X) = 4 \cdot \text{Var}(X) = 12$. (Note: if we mistakenly used Property 1, we would the wrong answer of 6.)

(iv) We use both Properties 1 and 2.

$$\text{Var}(X + 3Y) = \text{Var}(X) + \text{Var}(3Y) = 3 + 9 \cdot 5 = 48.$$

Example 4. Use Property 3 to compute the variance of $X \sim \text{Bernoulli}(p)$.

Solution: From the table

X	0	1
$p(x)$	$1-p$	p
X^2	0	1

we have $E[X^2] = p$. So Property 3 gives

$$\text{Var}(X) = E[X^2] - E[X]^2 = p - p^2 = p(1 - p).$$

This agrees with our earlier calculation.

Example 5. Redo Example 1 using Property 3.

Solution: From the table

X	1	3	5
$p(x)$	1/4	1/4	1/2
X^2	1	9	25

we have $E[X] = 7/2$ and

$$E[X^2] = 1^2 \cdot \frac{1}{4} + 3^2 \cdot \frac{1}{4} + 5^2 \cdot \frac{1}{2} = \frac{60}{4} = 15.$$

So $\text{Var}(X) = 15 - (7/2)^2 = 11/4$ –as before in Example 1.

3.4 Variance of binomial(n,p)

Suppose $X \sim \text{binomial}(n, p)$. Since X is the sum of *independent* Bernoulli(p) variables and each Bernoulli variable has variance $p(1 - p)$ we have

$$X \sim \text{binomial}(n, p) \ \Rightarrow\ \text{Var}(X) = np(1 - p).$$

3.5 Proof of properties 2 and 3

Proof of Property 2: This follows from the properties of $E[X]$ and some algebra.

Let $\mu = E[X]$. Then $E[aX + b] = a\mu + b$ and

$$\mathrm{Var}(aX+b) = E[(aX+b-(a\mu+b))^2] = E[(aX-a\mu)^2] = E[a^2(X-\mu)^2] = a^2E[(X-\mu)^2] = a^2\mathrm{Var}(X).$$

Proof of Property 3: We use the properties of $E[X]$ and a bit of algebra. Remember that μ is a constant and that $E[X] = \mu$.

$$\begin{aligned} E[(X-\mu)^2] &= E[X^2 - 2\mu X + \mu^2] \\ &= E[X^2] - 2\mu E[X] + \mu^2 \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \\ &= E[X^2] - E[X]^2. \qquad \text{QED} \end{aligned}$$

4 Tables of Distributions and Properties

Distribution	range X	pmf $p(x)$	mean $E[X]$	variance $\mathrm{Var}(X)$
Bernoulli(p)	0, 1	$p(0) = 1-p, \quad p(1) = p$	p	$p(1-p)$
Binomial(n, p)	0, 1,..., n	$p(k) = \binom{n}{k} p^k(1-p)^{n-k}$	np	$np(1-p)$
Uniform(n)	1, 2, ..., n	$p(k) = \frac{1}{n}$	$\frac{n+1}{2}$	$\frac{n^2-1}{12}$
Geometric(p)	0, 1, 2,...	$p(k) = p(1-p)^k$	$\frac{1-p}{p}$	$\frac{1-p}{p^2}$

Let X be a discrete random variable with range x_1, x_2, ...and pmf $p(x_j)$.

Expected Value:		**Variance**:
Synonyms:	mean, average	
Notation:	$E[X]$, μ	$\mathrm{Var}(X)$, σ^2
Definition:	$E[X] = \sum_j p(x_j)x_j$	$E[(X-\mu)^2] = \sum_j p(x_j)(x_j-\mu)^2$
Scale and shift:	$E[aX+b] = aE[X] + b$	$\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$
Linearity:	(for any X, Y) $E[X+Y] = E[X] + E[Y]$	(for X, Y independent) $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$
Functions of X:	$E[h(X)] = \sum p(x_j)\, h(x_j)$	
Alternative formula:		$\mathrm{Var}(X) = E[X^2] - E[X]^2 = E[X^2] - \mu^2$