

# Workshop 7 Homework

Los Angeles County  
ISAB

November 15, 2017

These exercises are intended to help you review basic identities related to the variance of random variables.

Print this out and take it with you to your meetings. Whenever the meeting becomes a waste of your time, discreetly take these pages out and start filling in the gaps of these derivations. Meeting hand-outs are a great source of scratch paper. Feign interest by asking for extra copies.

Why this obsession with the variance of a distribution? When we sample from a population in order to estimate some quantity (like an average or a percentage), the confidence in our estimate (the so-called confidence interval) is directly related to the variance of the population. This makes variance computations a big deal.

## Variance

Recall the identity we derived for the variance of a random variable  $X$  last Spring.

$$\text{Var}[X] = E[X^2] - E[X]^2$$

We often write  $\mu_X$  for the population mean, or  $\mu$  when  $X$  is understood.

**Problem 1** Derive the above identity from the following definition.

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

**Problem 2** How does the formula for the variance simplify when the population mean is zero?

---

## Covariance

Related to the variance is the *covariance* of two random variables. This is defined as

$$\text{cov}[X, Y] = E[(X - \mu_x)(Y - \mu_y)]$$

**Problem 3** Derive the following identity commonly used for the covariance from the above definition.

$$\text{cov}[X, Y] = E[XY] - E[X]E[Y]$$

Two random variables  $X$  and  $Y$  are said to be *independent* when  $\text{cov}[X, Y] = 0$ .

**Problem 4** Using the covariance identity above, what can we say about the expectation of a product of two independent random variables?

## Linear Combinations

**Problem 5** For a positive constant  $a$  show that  $\text{Var}[aX] = a^2\text{Var}[X]$ .

Just start with the definition and use the linearity of expectation.

$$\text{Var}[aX] = E[(aX)^2] - E[(aX)]^2$$

**Problem 6** For two random variables  $X$  and  $Y$  show that

$$\text{Var}(X + Y) = \text{Var}[X] + \text{Var}[Y] + 2 \cdot \text{cov}[X, Y]$$

**Problem 7** Show that for two random variables  $X$  and  $Y$  that **are independent** we have that

$$\text{Var}(aX + bY) = a^2\text{Var}[X] + b^2\text{Var}[Y]$$

Using the sigma notation for standard deviation where  $\sigma_X^2 = \text{Var}[X]$ , we have

$$\sigma_{aX+bY} = \sqrt{a^2\sigma_X^2 + b^2\sigma_Y^2}$$

We can see that variances are nicer for calculation. But don't forget that at the end of the day, it's the standard deviation that we use for our confidence intervals.

---

## Sample Average

Let's say we're sampling from a distribution represented by the random variable  $X$ . If we sample  $n$  times, then we represent the set as  $X_1, \dots, X_n$ . We'll stick to the case where each sample is independent of the others. Each of the  $X_i$  is a random variable with same distribution as the others. We can create a new random variable as a function of these  $n$  random variables.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i \quad (1)$$

This function should be familiar. It represents the average. But at this stage, we just want to think of it as another random variable. Like other random variables, an experiment or event occurs and a number pops out. The experiment in this case is to sample the  $X$  random variable  $n$  times and take the average. Equation (1) is called the **sample average**.

It's important not to confuse this with the **population mean**. (It's easy to do when this is new!)

$$\bar{X} \neq \mu_X = E[X]$$

The left side is a random variable; the right side is a constant. The left side yields a number with each sampling. Sometime this number may be close to  $\mu_X$ ; other times it may not be so close. What you expect the random variable to be (i.e.  $E[\bar{X}]$ ) and the value that actually pops out are two different things. The expectation is fixed; the actual value varies with each sampling. And since I just brought it up, what is this  $E[\bar{X}]$ ?

**Problem 8** Show that  $E[\bar{X}] = \mu_X$ .

This is a simple application linearity of the expected value. So let's make this a little more interesting. Apply what you derived in the **Linear Combinations** section to the following.

**Problem 9** Show that

$$\text{Var}[\bar{X}] = \frac{\text{Var}[X]}{n} \quad (2)$$

To get things started:

$$\text{Var}[\bar{X}] = \text{Var}\left[\frac{1}{n} \sum_{i=1}^n X_i\right] = \frac{1}{n^2} \text{Var}\left[\sum_{i=1}^n X_i\right] = \frac{1}{n^2} \sum_{i=1}^n \text{Var}[X_i]$$

Remember to use the fact that the  $X_i$  are independent of each other.