Handout 05: Expectation and Variance

This class does not require that you have had any prior probability theory and we certainly do not have time to discuss and derive all of the details that would be covered in a semester of MATH329. However, it will be very useful to complete our study of linear regression and to motivate the next steps if we can use some probability theory. The notes here lay out the basic points about random variables, expected values, and variance.

Random variables

A random variable provides, conceptually, a way of representing uncertainty in a value. Instead of the fixed value (albeit often unknown) we typically denote with a variable such as x, a random variable has information about all of the possible outcomes it might take on and the chance that it takes on one of those values. It is possible to derive a more general theory, but today we will restrict ourselves to the case of random variables that take on real valued quantities.

The mathematical definition of a random variable Z (in our restricted case) is given by a function $f(z)$ – the *density function* – that maps the real line into the unit interval and has an integral of 1:

$$f : \mathbb{R} \to [0, 1] \tag{5.1}$$

$$\int_{-\infty}^{+\infty} f(z)dz = 1. \tag{5.2}$$

How does this object represent randomness? We define the probability that the random variable is between two value a and b as the integral of the function between these two values.

$$Pr(a \leq Z \leq b) = \int_a^b f(z)dz \tag{5.3}$$

You can see now why we force the total integral to be one: the random variable has to be somewhere.

The expected value of a random variable describes the equivalent of the mean of the random variable. We can define the expected value as:

$$\mathbb{E}Z = \int_{-\infty}^{+\infty} z \cdot f(z)dz. \tag{5.4}$$

You should convince yourself that this makes sense, for example by considering a random variable with a function f that is symmetric around the origin (it should have an expected value of 0, at least as long as the integral converges).

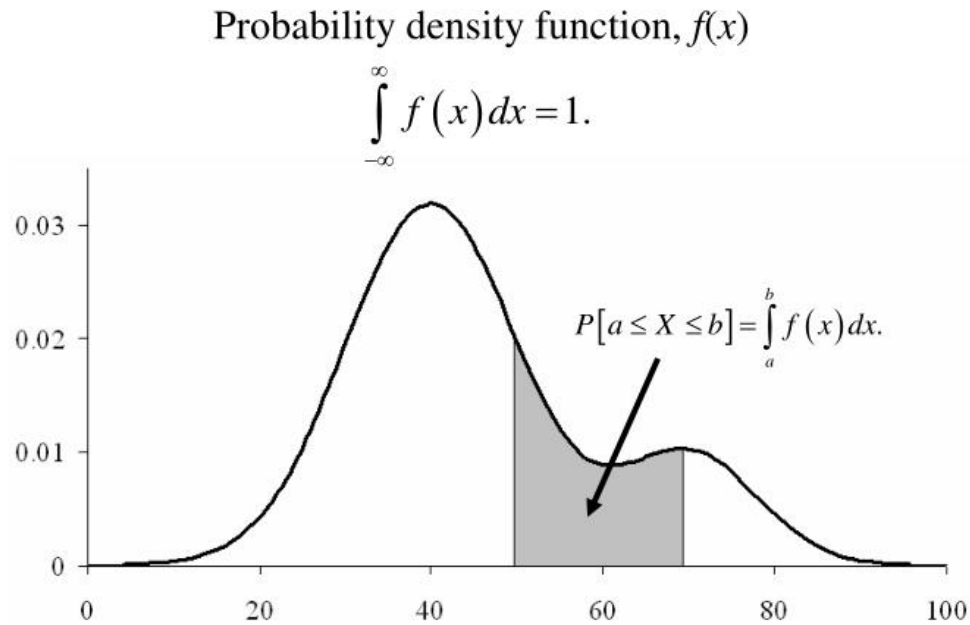

Figure 1: Example of a probability density function.

The expected value tells us where the center of the random variable is. The variance tells us how much a random variable varies around this center. Specifically, the variance is the expected squared value of the distance a random variable will be away from its mean. Or, symbolically:

$$\mathbb{V}ar(Z) = \mathbb{E}\left[(Z - \mathbb{E}Z)^2\right] \tag{5.5}$$

Whenever possible, when working with expected values and variances, you will want to avoid go back to the original definitions. We will ultimately need vector version of these two operators, but to avoid moving too quickly here are some basic properties of the expected value and variance. Let Z be a random variable and a be a non-random variables. Then:

$$\mathbb{E}(aZ) = a \cdot \mathbb{E}(Z) \tag{5.6}$$
$$\mathbb{E}(a + Z) = a + \mathbb{E}(Z) \tag{5.7}$$
$$\mathbb{V}ar(aZ) = a^2 \cdot \mathbb{V}ar(Z) \tag{5.8}$$
$$\mathbb{V}ar(a + Z) = \mathbb{V}ar(Z) \tag{5.9}$$

It is much better—conceptually and theoretically—to treat these as operators that have particular properties.

Random vectors

To actually make use of the probability theory given here in the theory of linear regression, we need to extend our definitions to random vectors. A random vector ϵ is just a vector where each component ϵ_i is itself a random variable:

$$\epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix}. \tag{5.10}$$

And the expected value of a random vector is the vector of expected values of each component:

$$\mathbb{E}\epsilon = \begin{bmatrix} \mathbb{E}\epsilon_1 \\ \vdots \\ \mathbb{E}\epsilon_n \end{bmatrix}. \tag{5.11}$$

The variance of a random variable is a bit more involved. The variance is given by an n-by-n matrix, the diagonal of which is a variance of each component. The off-diagonal quantities give the *covariance* of the respective components. Specifically:

$$\mathbb{V}ar(\epsilon)_{i,j} = \mathbb{E}\left[(\epsilon_i - \mathbb{E}\epsilon_i) \cdot (\epsilon_j - \mathbb{E}\epsilon_j)\right]. \tag{5.12}$$

Two variables that have a covariance of zero are said to be *uncorrelated*. The entire variance matrix can be written compactly using the transpose operator:

$$\mathbb{V}ar(\epsilon) = \mathbb{E}\left[(\epsilon - \mathbb{E}\epsilon) \cdot (\epsilon - \mathbb{E}\epsilon)^t\right]. \tag{5.13}$$

It is the formula that will be most helpful for us next week as we investigate statistical learning extensions of the linear model.

It should be straightforward to see that the expected value of a vector equation behaves similarly to the scalar version. Let ϵ be a random vector, v a non-random vector, and A a non-random matrix. Then:

$$\mathbb{E}(A \cdot \epsilon) = A \cdot \mathbb{E}\epsilon \tag{5.14}$$

$$\mathbb{E}(v + \epsilon) = v + \mathbb{E}\epsilon \tag{5.15}$$

There are also two equivalent properties for the Variance operator:

$$\mathbb{V}ar(v + \epsilon) = \mathbb{V}ar(\epsilon) \tag{5.16}$$

$$\mathbb{V}ar(A\epsilon) = A \cdot \mathbb{V}ar(\epsilon) \cdot A^t \tag{5.17}$$

You will derive the last equations in today's lab.

LAB QUESTIONS

1. Assume that Z is a random variable that only takes values between 0 and 5. Any value in this range is equally likely to occur. Write down the formula for the density function f that corresponds to this random variable.

2. Without doing any calculations, what do you expect to be the expected value of Z in the previous question?

3. Using the formulae in the notes, show that your guess matches the mathematical definition of $\mathbb{E}Z$.

4. Prove the equation given for $\mathbb{V}ar(AZ)$.

5. Assume that we have a random variable y defined in terms of a random variable ϵ such that:

$$y = Xb + \epsilon. \tag{5.18}$$

   Further assume that the expected value of the ϵ term is the zero vector (a vector with zeros in every component). Show that the expected value of the ordinary least squares equation for the estimate β is equal to b. This means that β is an *unbiased* estimator of b.

6. Using the same set-up as above, further assume that:

$$\mathbb{V}ar(\epsilon) = \sigma^2 \cdot I_n. \tag{5.19}$$

   For some fixed value $\sigma^2 > 0$. What is the variance of y?

7. Finally, derive a formula for the variance of the ordinary least squares estimate β.