

4. Random Variables

Chris Piech and Mehran Sahami

Oct 2017

A Random Variable is a variable that probabilistically takes on different values. You can think of an RV as being like a variable in a programming language. They take on values, have types and have domains over which they are applicable. We can define events that occur if the random variable takes one values that satisfy a numerical test (eg does the variable equal 5, is the variable less than 8). We often think of the probabilities of such events.

As an example, let's say we flip three fair coins. We can define a random variable Y to be the total number of "heads" on the three coins. We can ask about the probability of Y taking on different values using the following notation:

- $P(Y = 0) = 1/8$ (T, T, T)
- $P(Y = 1) = 3/8$ (H, T, T), (T, H, T), (T, T, H)
- $P(Y = 2) = 3/8$ (H, H, T), (H, T, H), (T, H, H)
- $P(Y = 3) = 1/8$ (H, H, H)
- $P(Y \geq 4) = 0$

Even though we use the same notation for random variables and for events (both use capitol letters) they are distinct concepts. An event is a scenario, a random variable is an object. The scenario where a random variable takes on a particular value (or range of values) is an event. When possible, I will try and use letters E, F, G for events and X, Y, Z for random variables.

Using random variables is a convenient notation technique that assists in decomposing problems. There are many different types of random variables (indicator, binary, choice, Bernoulli, etc). The two main families of random variable types are discrete and continuous. For now we are going to develop intuition around discrete random variables.

Probability Mass Function

For a discrete random variable, the most important thing to know is a mapping between the values that the random variable could take on and the probability of the random variable taking on said value. In mathematics, we call associations functions.

The probability mass functions (PMF) is a function that maps possible outcomes of a random variable to the corresponding probabilities. Because it is a function, we can plot PMF graphs where the x -axis are the values that the random variable could take on and the y -axis is the probability of the random variable taking on said value:

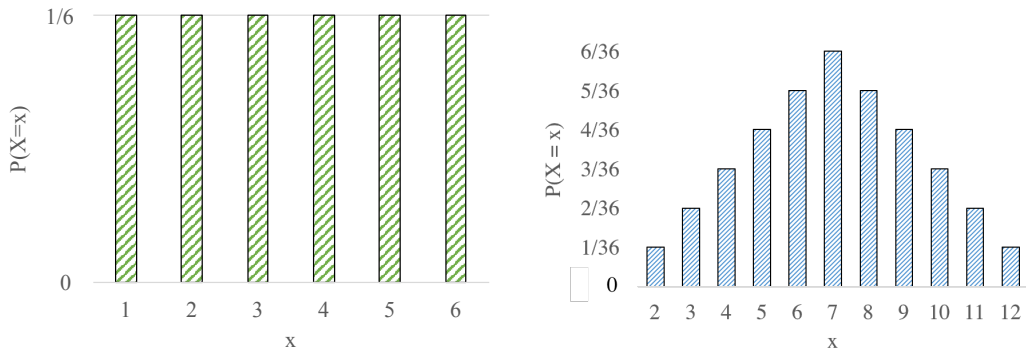


Figure: On the left, the PMF of a single 6 sided die roll. On the right, the PMF of the sum of two dice rolls.

There are many ways that these Probability Mass Functions can be specified. We could draw a graph. We could have a table (or for you CS folks, a Map) that lists out all the probabilities for all possible events. Or we could write out a mathematical expression.

For example lets consider the random variable X which is the sum of two dice rolls. The probability mass function can be defined by the graph on the right of figure . It could have also been defined using the equation:

$$p_X(x) = \begin{cases} \frac{x}{36} & \text{if } x \in \mathbb{R}, 0 \leq x \leq 6 \\ \frac{12-x}{36} & \text{if } x \in \mathbb{R}, x \leq 7 \\ 0 & \text{else} \end{cases}$$

The probability mass function, $p_X(x)$, defines the probability of X taking on the value x . The new notation $p_X(x)$ is simply different notation for writing $P(X = x)$. Using this new notation makes it more apparent that we are specifying a function. Try a few values of x , and compare the value of $p_X(x)$ to the graph in figure . They should be the same.

Expected Value

A relevant statistic for a random variable is the average value of the random variable over many repetitions of the experiment it represents. This average is called the Expected Value.

The Expected Value for a discrete random variable X is defined as:

$$E[X] = \sum_{x:P(x)>0} xP(x)$$

It goes by many other names: Mean, Expectation, Weighted Average, Center of Mass, 1st Moment.

Example 1

Lets say you roll a 6-Sided Die and that a random variable X represents the outcome of the roll. What is the $E[X]$? This is the same as asking what is the average value.

$$E[X] = 1(1/6) + 2(1/6) + 3(1/6) + 4(1/6) + 5(1/6) + 6(1/6) = 7/2$$

Example 2

Lets say a school has 3 classes with 5, 10, and 150 students. If we randomly choose a class with equal probability and let X = size of the chosen class:

$$\begin{aligned} E[Y] &= 5(1/3) + 10(1/3) + 150(1/3) \\ &= 165/3 = 55 \end{aligned}$$

If instead we randomly choose a student with equal probability and let Y = size of the class the student is in

$$\begin{aligned} E[X] &= 5(5/165) + 10(10/165) + 150(150/165) \\ &= 22635/165 = 137 \end{aligned}$$

Example 3

Consider a game played with a fair coin which comes up heads with $p = 0.5$. Let n = the number of coin flips before the first “tails”. In this game you win $\$2^n$. How many dollars do you expect to win? Let X be a random variable which represents your winnings.

$$\begin{aligned} E[X] &= \left(\frac{1}{2}\right)^1 2^0 + \left(\frac{1}{2}\right)^2 2^1 + \left(\frac{1}{2}\right)^3 2^2 + \left(\frac{1}{2}\right)^4 2^3 + \dots = \sum_{i=0}^{\infty} \left(\frac{1}{2}\right)^{i+1} 2^i \\ &= \sum_{i=0}^{\infty} \frac{1}{2} = \infty \end{aligned}$$

Properties of Expectation

Expectations preserve linearity which means that

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

It also holds in the case where you are adding random variables. Regardless of the relationship between random variables, the expectation of the sum is equal to the sum of the expectation. For random variables A and B :

$$E[A + B] = E[A] + E[B] \tag{1}$$

There is a wonderful law called the Law of the Unconscious Statistician that is used to calculate the expected value of a function $g(X)$ of a random variable X when one knows the probability distribution of X but one does not explicitly know the distribution of $g(X)$.

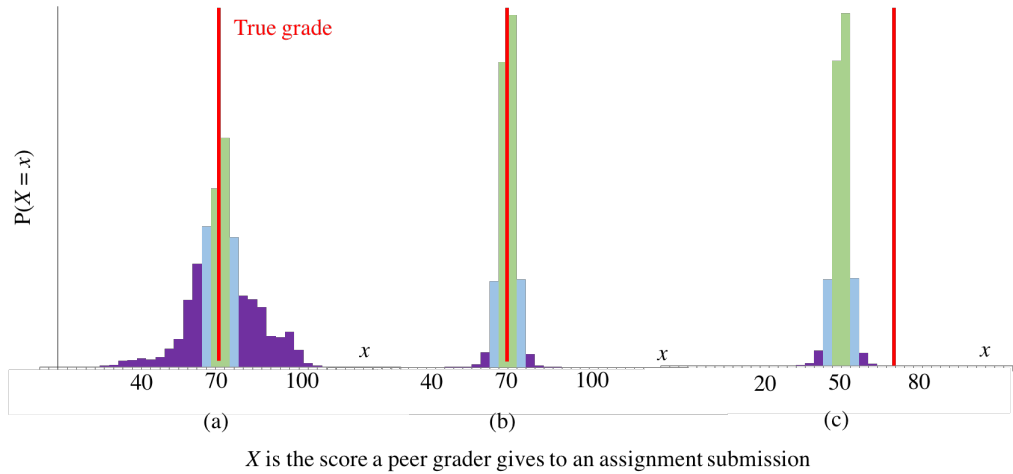
$$E[g(X)] = \sum_x g(x) \cdot p_X(x)$$

For example, lets apply the law of the unconscious statistician to compute the expectation of the square of a random variable (called the second moment).

$$\begin{aligned} E[X^2] &= E[g(X)] && \text{where } g(X) = X^2 \\ &= \sum_x g(x) \cdot p_X(x) && \text{by the unconscious statistician} \\ &= \sum_x x^2 \cdot p_X(x) && \text{by the unconscious statistician} \end{aligned}$$

Variance

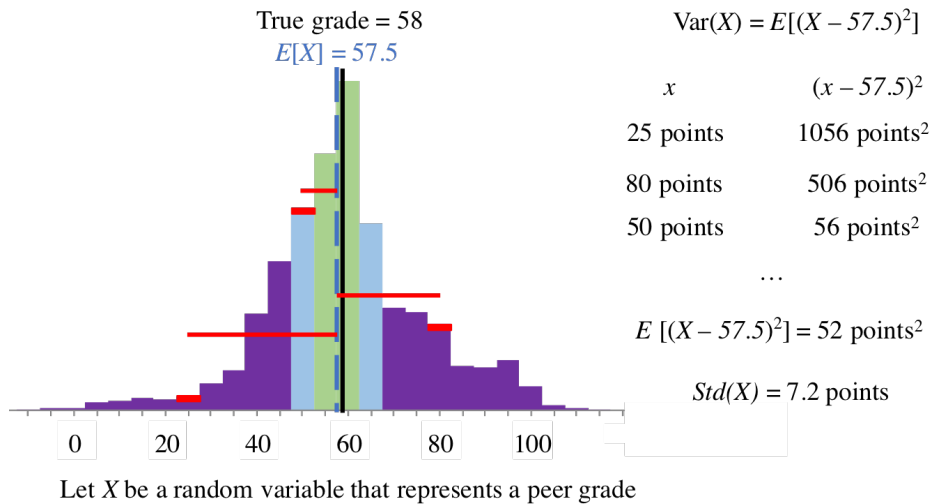
Expectation is a truly useful statistic, but it does not give a detailed view of the probability mass function. Consider the following 3 distributions (PMFs) which describe the distribution of grades given to student assignments by three different graders:



For each distribution we show the “true grade” that the student deserves using a red line. Which of the three graders would you like to be giving you a grade on your assignment? We can start by eliminating option (c). This grader, on average, gives a grade that is about 22 percentage points less than the student deserves.

How about grader (a) vs grader (b)? Both distributions have the same expected value but the “spread” of the two distributions are quite different. Expectation is a useful statistic to know about a random variable, the next most useful statistic to know is variance, a measure of the spread of a distribution.

How should we turn the idea of spread into a number? The formula for variance asks the question: What is the expected Euclidean distance between a given grade and the expected grade? In the figure bellow for a single grader’s accuracy random variable, we show three sample grades that could have been given by the grader and calculate the Euclidean distance between that grade and the expectation of the distribution (57.5). The Expectation is simply the weighted average of those distances:



The variance of a discrete random variable, X , with expected value μ is:

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

When computing the variance often we use a different form of the same equation:

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

Intuitively this is the weighted average distance of a sample to the mean.

Variance is especially useful for *comparing* the spread of two distributions. A larger variance means that there is more deviation around the mean. However, if you look at the example above, the units of variance are the square of points. This makes it hard to interpret the numerical value. What does it mean that the spread is 52 points²? A more interpretable measure of spread is the square root of Variance, which we call the Standard Deviation $SD(X) = \sqrt{\text{Var}(X)}$. The standard deviation of our grader is 7.2 points. Under certain assumptions (that we will get to later in the book) we can make interpretable, mathematical claims about the standard deviation.

There are many measures that statisticians could have used to quantify spread, and I encourage you at this point to think of other ways that you could turn random variable spread into a number. Some ideas you may have come up with could include, how far can you get from the expectation such that there is a 50% chance that the random variable takes on a value in that range. Variance, as defined above, has been the most common measure (perhaps because it is easy to calculate), but in another universe we could have developed an alternate statistic.

Why is it true that $E[(X - \mu)] = E[X^2] - E[X]^2$?

$\begin{aligned} \text{Var}(X) &= E[(X - \mu)^2] \\ &= \sum_x (x - \mu)^2 p_x(x) \\ &= \sum_x (x^2 - 2\mu x + \mu^2) p_x(x) \\ &= \sum_x x^2 p_x(x) - 2\mu \sum_x x p_x(x) + \mu^2 \sum_x p_x(x) \\ &= E[X^2] - 2\mu E[X] + \mu^2 \cdot 1 \\ &= E[X^2] - 2\mu^2 + \mu^2 \\ &= E[X^2] - \mu^2 \\ &= E[X^2] - (E[X])^2 \end{aligned}$	<p>μ is defined to be expectation. Ie $\mu \equiv E[X]$</p> <p>Law of unconscious statistician</p> <p>By algebraic expansion</p> <p>Breaking apart the sum</p> <p>Law of unconscious statistician</p> <p>Recall that: $\mu \equiv E[X]$</p> <p>$\mu \equiv E[X]$</p>
---	--

Identities of Variance

A provable identity about variance is that:

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

This (provable) equation claims that if you add a scalar value, b , to a random variable, it will shift the expectation of the distribution, but the spread around the expectation will be the same. However, if you multiply your random variable with a scalar, a , it will elongate the spread and the resulting distribution will have a variance that is a^2 times greater.

Example 1

Let X = value on roll of a 6 sided die. Recall that $E[X] = 7/2$. First lets calculate $E[X^2]$

$$E[X^2] = (1^2)\frac{1}{6} + (2^2)\frac{1}{6} + (3^2)\frac{1}{6} + (4^2)\frac{1}{6} + (5^2)\frac{1}{6} + (6^2)\frac{1}{6} = \frac{91}{6}$$

Which we can use to compute the variance:

$$\begin{aligned} \text{Var}(X) &= E[X^2] - (E[X])^2 \\ &= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} \end{aligned}$$