Lecture 11: Continuous Random Variables

CSCI2244-Randomness and Computation

March 20, 2019

Overview The definition of a random variable is the same for continuous probability spaces as for discrete ones: A random variable just associates a number to every outcome in the sample space.

Recall that for a discrete random variable X, we had two different ways to describe the distribution: The PMF

$$P_X(x) = P(X = x)$$

which gives the probability of each value of the random variable, and the CDF F_X defined by

$$F_X(x) = P(X \leq x).$$

These two functions provide exactly the same information, but in different form: You can obtain the CDF from the PMF by taking cumulative sums. You obtain the PMF from the CDF by taking differences of successive values.

For continuous random variables, the CDF still makes sense, and has the same definition, but the PMF gives no information—it typically assigns probability 0 to every real number! What replaces the PMF in the continuous case—the *probability density function* of the random variable is the big new idea in this section.

1 CDFs for some simple distributions

1.1 One spinner

We define the random variable X to be simply the number showing on the spinner. We have,

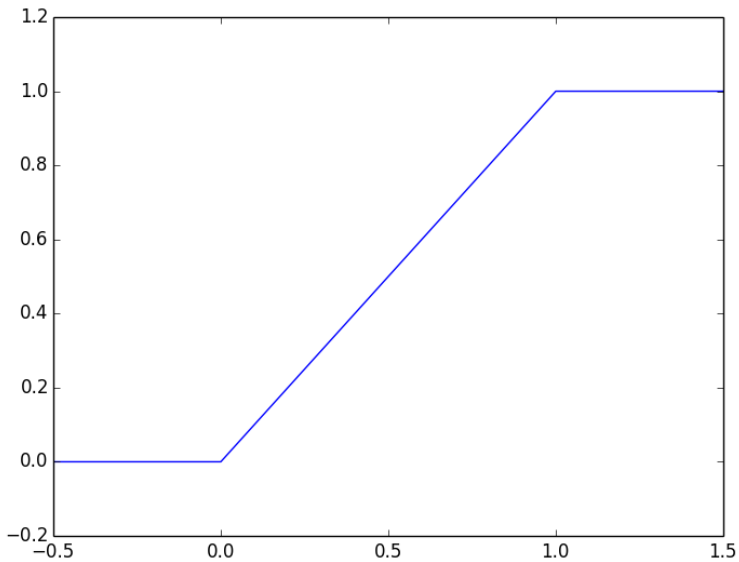

Figure 1: CDF for the outcome of a single spinner

$$F_X(a) = P(X \leq a) = \begin{cases} 0 & : a < 0 \\ a & : 0 \leq a \leq 1 \\ 1 & : 1 < a \end{cases}$$

The graph of the cumulative distribution function is shown in Figure 1. Observe that like CDFs for discrete random variables, the function is nondecreasing with values going from 0 to 1. Here, however, the function is continuous, instead of consisting of discrete steps.

1.2 Sum of two spinners.

Let Y denote the sum of the results of two spinners (or of two successive spins of the same spinner–it doesn't matter). To compute the cumulative distribution function we use our trick of depicting the underlying sample space as the unit square, and determine the area of the event $Y < a$. There are two cases to consider, depending on whether $a < 1$ or $a > 1$. The situation is depicted in Figure 2 below. If $a < 1$ then the probability of the event is the area of the small right triangle

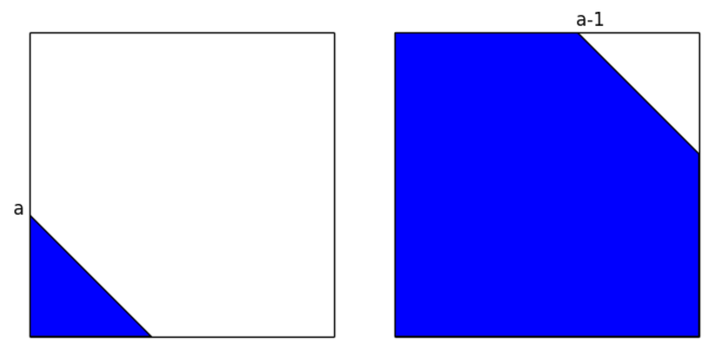

Figure 2: Sum of two spinners. Calculation of $P(Y < a)$ for $a < 1$ (left) and $a > 1$ (right).

with both legs having length a: This is $a^2/2$. If $a > 1$, then the *complementary* probability is the area of the right triangle with both legs of length $(2-a)$. Putting it all together gives:

$$F_Y(a) = \begin{cases} 0 & : a < 0 \\ \frac{a^2}{2} & : 0 \leq a \leq 1 \\ 1 - \frac{(2-a)^2}{2} & : 1 \leq a \leq 2 \\ 1 & : 2 < a \end{cases}$$

The CDF is graphed in Figure 3.

1.3 Darts

Imagine a dartboard one foot in radius. The experiment consists of throwing a dart at the board, and the outcome is the point at which it hits. We assume a uniform probability distribution–we only consider darts that hit the board, but we're not

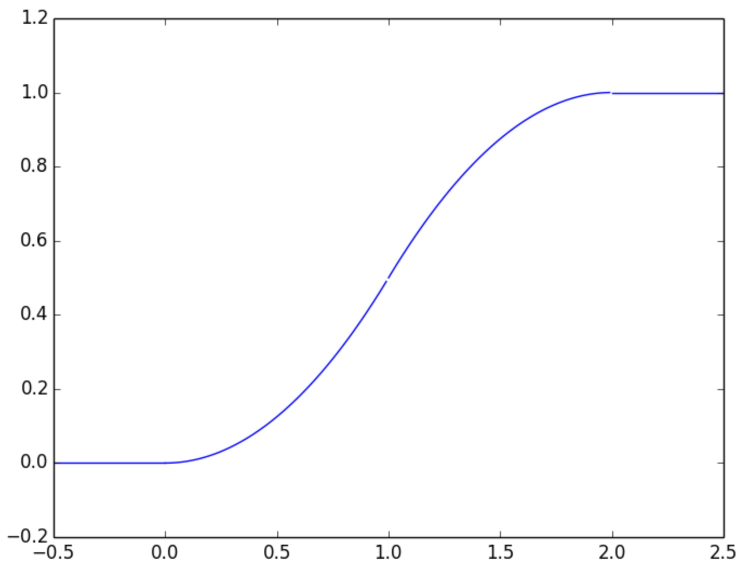

Figure 3: CDF for sum of two spinners.

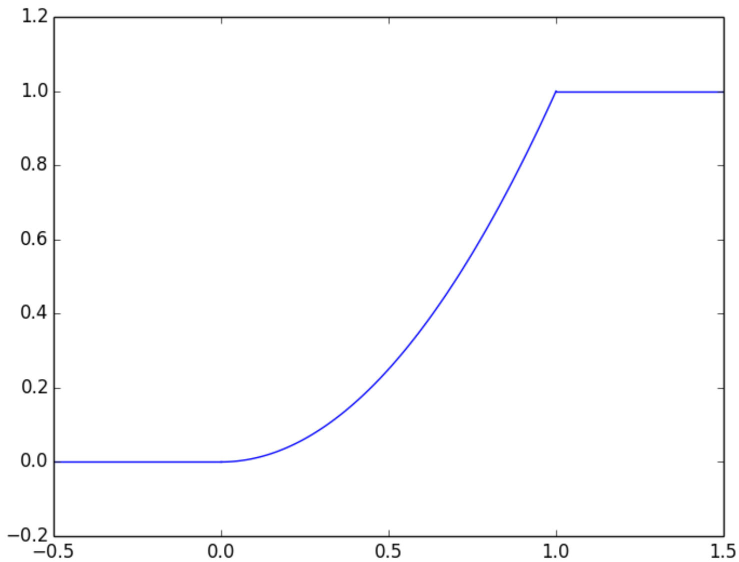

Figure 4: CDF for distance of dart from center of dartboard.

really aiming either, so no region of the board is more likely to be hit than any other equal-sized region. This means that the sample space is

$$S = \{(x, y) \in \mathbf{R}^2 : x^2 + y^2 \leq 1\},$$

and for any region $E \subseteq S$,

$$P(E) = \frac{\text{area}(E)}{\text{area}(S)} = \frac{\text{area}(E)}{\pi}.$$

Let the random variable Z be the distance of the dart from the center. If $0 \leq a \leq 1$, then the event $X \leq a$ is a disk of radius a, which has area πa^2, and thus

$$F_Z(a) = \begin{cases} 0 & : a < 0 \\ a^2 & : 0 \leq a \leq 1 \\ 1 & : 1 < a \end{cases}$$

The CDF is graphed in Figure 4.

2 Probability density function

For discrete probability spaces we studied the probability mass function, defined by $p_X(a) = P(X = a)$. For continuous distributions the function defined this way will typically have value 0 at all a, so this gives no information. The role of the probability mass function is instead played by the *probability density function,* or PDF, defined by

$$f_X(a) = F'_X(a).$$

That is, the density function is the derivative of the cumulative distribution function. If you turn things around, the CDF can be recovered by integrating the PDF:

$$F_X(a) = \int_{-\infty}^{x} f_X(t)dt.$$

Probability density functions satisfy the following properties:

- $f_X(x) \geq 0$ for all $x \in \mathbf{R}$.
- $\int_{-\infty}^{\infty} f_X(t)dt = 1$.

These are just like the properties of a PMF, except that the discrete sum is replaced by an integral. Just as we can define discrete random variables by giving the PMF, we can define a continuous random variable by giving a function that satisfies the two properties above.

2.1 Single spinner

The PDF of the random variable giving the outcome of a spinner is shown in Figure 5. Observe that the CDF is not differentiable at $x = 0$ and $x = 1$ (the sharp corners on the graph) so the PDF is not really defined at these points. (The dashed lines are not really part of the graph; they're just there to guide your eyes.) Observe that the two properties of a PDF are satisfied—the graph never drops below the x-axis, and the area between the x-axis and the graph is 1.

Compare this to the graph of the PMF of a uniform discrete random variable, for example the outcome of a single die (Figure 6.) The shape is the same—in a sense the spinner is a continuous version of the die.

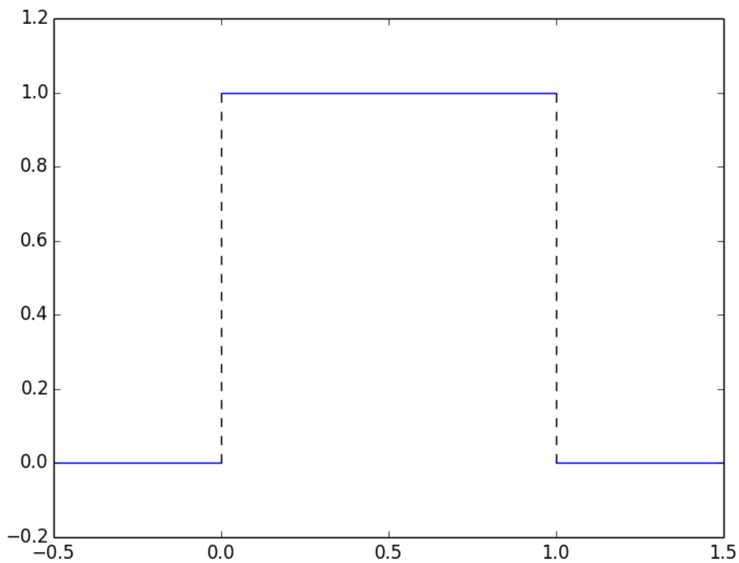

Figure 5: Probability density function for outcome of a single spinner.

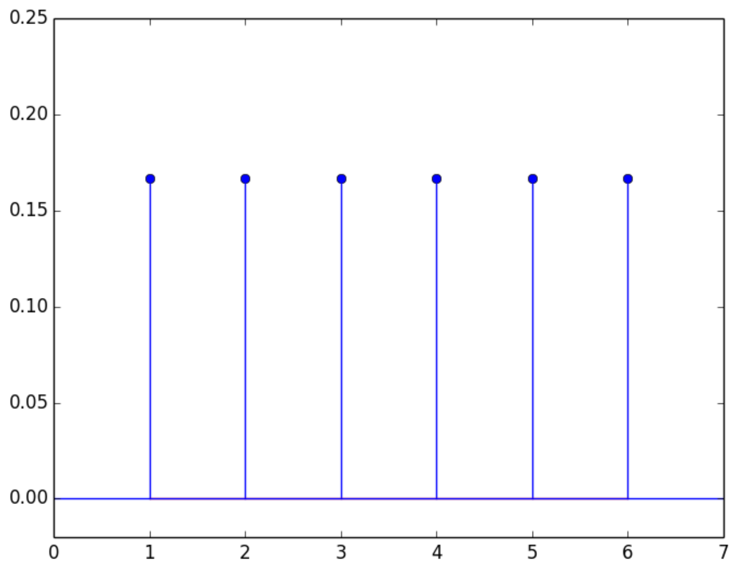

Figure 6: PMF for roll of a single die–note that the overall shape is the same as that of the PDF for a single spinner.

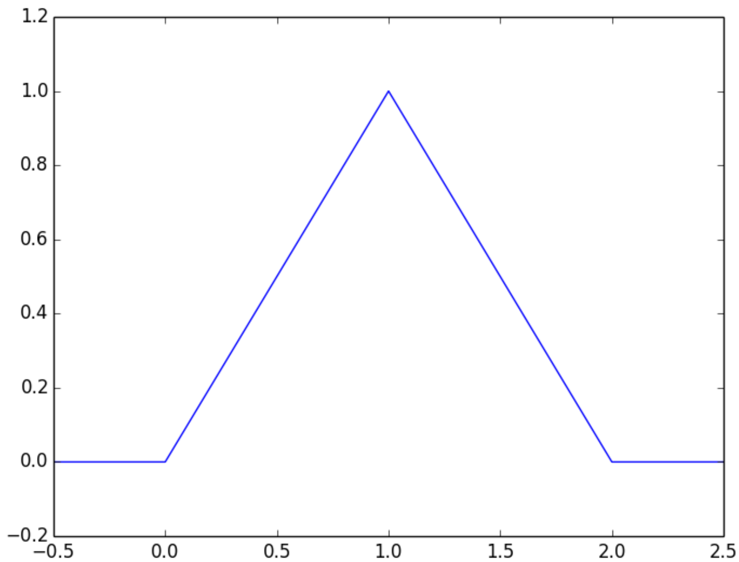

Figure 7: Probability density function for sum of two spinners.

2.2 Sum of two spinners.

The PDF is shown in Figure 7—we obtained this by differentiating the different pieces given earlier for the CDF. The shape tells us that it is more likely that the sum will be close to 1 than to either of the extremes 0 or 2. Compare this to the PMF for sum of two dice (Figure 8) which once again has the same basic shape.

3 Expected value of a continuous random variable.

The definition of expected value resembles that of the expected value of a discrete random variable, but we replace the PMF by the PDF, and summation by integration. So we have

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Once again, we have the linearity property (expected value of a sum of random variables is the sum of the expected values), although the proof is different. Here

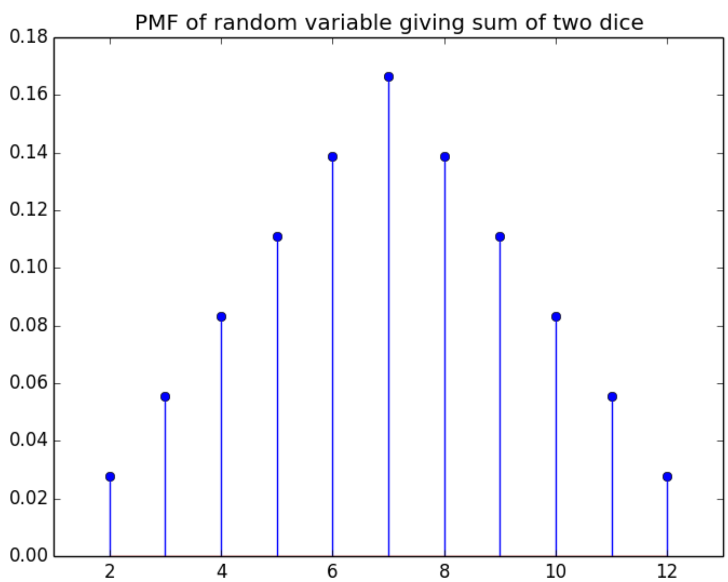

Figure 8: Probability mass function for the sum of two dice.

we compute the expected values for our examples above.

3.1 Single spinner.

Since $f_X(x) = 0$ outside the interval between 0 and 1, and $f_X(x) = 1$ if x is in the interval,we have

$$E(X) = \int_{-\infty}^{\infty} x f_X(x) dx = \int_0^1 x dx = x^2/2 \Big|_0^1 = \frac{1}{2},$$

which of course is exactly what you would expect—if you spin the spinner a bunch of times, the values of the spins should average to $\frac{1}{2}$.

3.2 Sum of two spinners.

We could repeat the calculation using the definition of expected value (I did this in class, just to show what this would look like), but it is simpler to use the fact that this random variable Y is the sum of two random variables having the distribution given above for a single spinner. So by linearity, $E(Y) = \frac{1}{2} + \frac{1}{2} = 1$. Again, exactly what you'd expect.

3.3 Dartboard.

Let's do our dartboard example. The PDF is $f_Z(x) = 2x$ on the interval $[0, 1]$ and 0 outside this interval. So

$$E(Z) = \int_0^1 2x^2 dx = \frac{2}{3}.$$

3.4 The expected value is not guaranteed to exist

We saw the same phenomenon with discrete random variables (St. Petersburg Paradox): Suppose a random variable X has density given by

$$f_X(x) = \begin{cases} 0 & : x < 1 \\ x^{-2} & : 1 \leq x \end{cases}$$

This function has nonnegative values and integrates to 1, so it satisfies the requirements to be a probability density function. However the integral

$$\int_{-\infty}^{\infty} x f_X(x) dx$$

is infinite, so the expected value is not defined.

4 Simulations

We can simulate the random experiments that give rise to these variables. For example, if we wanted to simulate the dartboard example, we can generate a large number of points uniformly distributed in the unit circle (first generate such points in the circumscribed square, which is easier to do, then throw out all the ones that are not in the circle) and return their distances from the center. A histogram of such samples will closely match the shape of the density function, and the mean of the samples should be close to the expected value of the random variable.

Alternatively, we can generate samples with the same distribution as a random variable without having to simulate the underlying experiment: For this we use the inverse of the cumulative distribution function. Both these approaches, applied to the dartboard example, are illustrated in the accompanying code.

5 Exponential distribution

Let's say that on average, a customer enters a shop once every five minutes. Or an alpha particle that is the result of radioactive decay strikes a detector on average 5 times per second. Start a clock, and let X denote the time until the next arrival.

This is a continuous analogue of the experiment where we flip a coin repeatedly until we get heads. We assume that this process is *memoryless*: The coin does not remember whether it last came up heads or tails; and even if it just came up heads ten times in a row, it is not more likely to come up tails on the next toss in light of this. By the same token, the arrival of a new customer or a new alpha particle is not the result of some buildup, making the arrival more likely the longer we have been waiting for it. We can express this memorylessness property by

$$P(X > r + s | X > r) = P(X > s).$$

In other words, the probability that we will wait more than $r + s$ seconds for the next arrival, assuming that we wait at least r seconds is the same as the probability that we wait at least s seconds. The process does not 'know' that it has already gone r seconds without an arrival.

We can rewrite the above equation as

$$P(X > s) = \frac{P(X > r + s) \cap P(X > r)}{P(X > s)} = \frac{P(X > r + s)}{P(X > s)},$$

so

$$P(X > s) \cdot P(X > r) = P(X > r + s)$$

for all positive r, s. Let $U(x) = P(X > x)$. The above equation (with $r = s = 1$) implies

$$U(2) = U(1)^2,$$

and likewise

$$U(3) = U(1)^3,$$

etc. But also, with $r = s = \frac{1}{2}$,

$$U(\frac{1}{2}) = U(1)^{\frac{1}{2}}.$$

In fact, the only continuous function satisfying the above equation is

$$U(x) = U(1)^x.$$

Thus the cumulative distribution function is

$$F_X(x) = 1 - U(x) = 1 - a^{-x},$$

for $x > 0$, where $a = 1/U(1) > 1$. We usually write this as $1 - e^{-\lambda x}$ where $\lambda > 0$. The parameter λ is related to the average wait time, as we shall see.

This is called the *exponential distribution*. The CDF, and the PDF, its derivative, which is $\lambda e^{-\lambda x}$ for positive x, are shown in Figures 8 and 9. Compare the PMF of the geometric distribution (Figure 10.) The exponential distribution is just the continuous version.

5.1 Expected value.

You need to dredge up some stuff you may have learned in calculus about integration by parts to evaluate the integral! We'll go through it in class, but here is the punchline: For an exponentially distributed random variable with parameter λ, the expected value (which represents the average wait until the next arrival) is given by

$$\int_0^\infty \lambda x e^{-\lambda x} dx = \frac{1}{\lambda}.$$

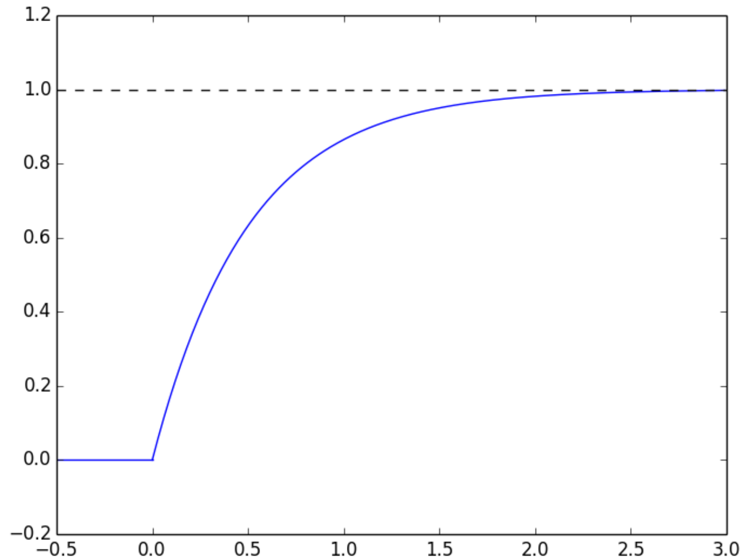

Figure 9: CDF for exponential distribution with $\lambda = 2$.

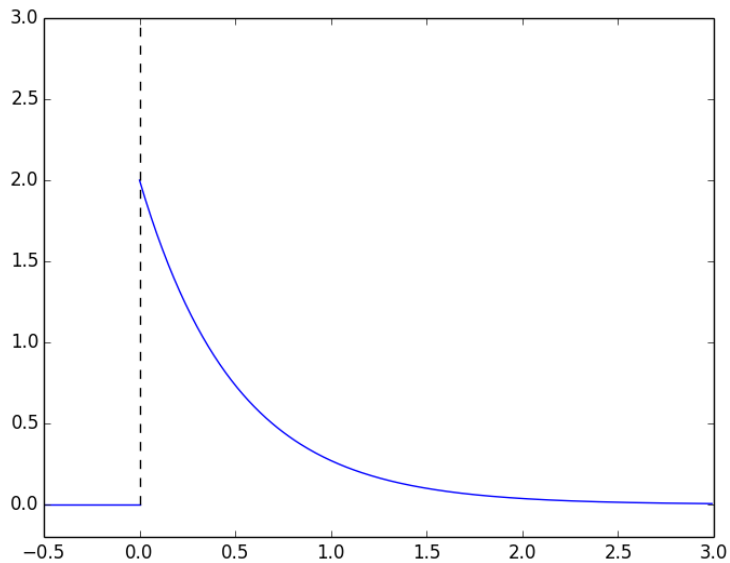

Figure 10: Probability density function exponential distribution with $\lambda = 2$.

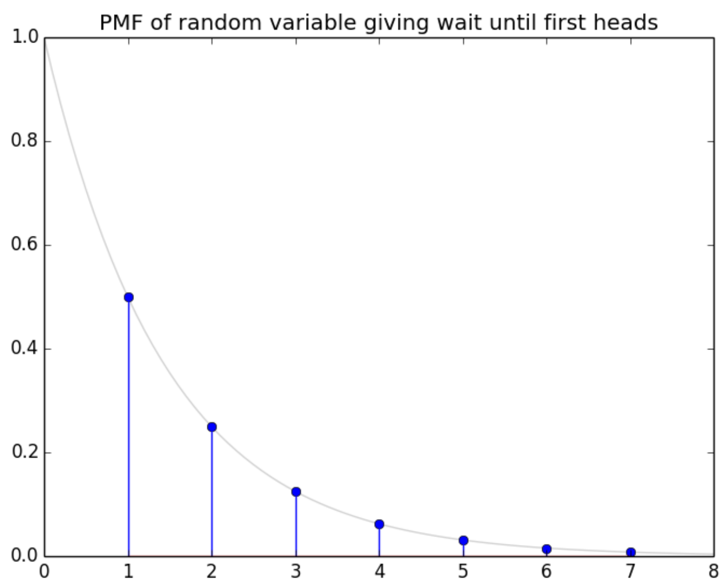

Figure 11: Probability mass function for geometric distribution.