STAT 345 ⋄ UNIFORM, EXPONENTIAL AND WEIBULL DISTRIBUTIONS

KELLIN RUMSEY ⋄ SPRING 2020

Definition: A continuous RV X is said to have a *continuous uniform* distribution if its probability density function (PDF) is

$$f(x) = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & \text{otherwise} \end{cases}$$

We often refer to this distribution simply as a *uniform* distribution, and we write $X \sim U(a, b)$. The corresponding CDF is

$$F(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & a < x < b \\ 1, & x > b \end{cases}$$

The mean and variance of X are given by:

$$E(X) = \frac{a+b}{2} \qquad \text{Var}(X) = \frac{(b-a)^2}{12}$$

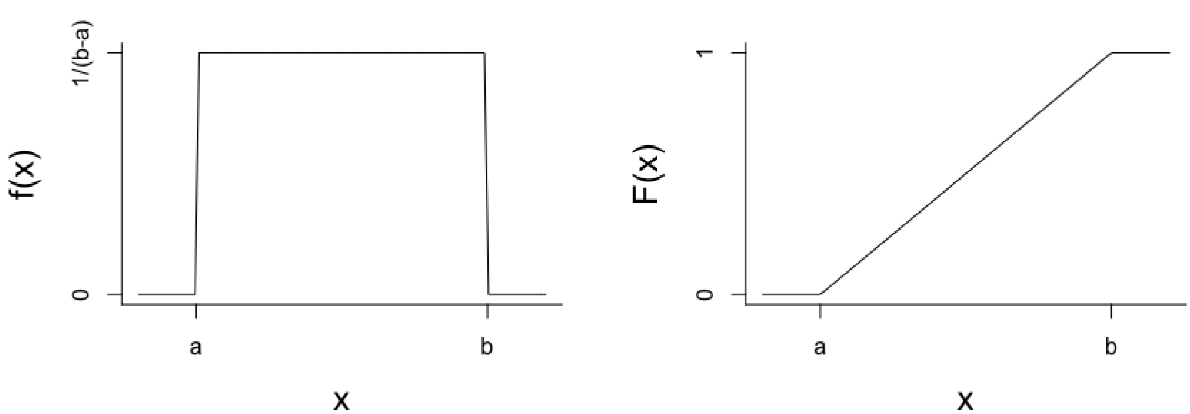

FIGURE 1. Probability density function $f(x)$ and cumulative distribution function $F(x)$ for the uniform distribution between a and b.

Example: Professor Halfbrain has an appointment with his dentist at 9 : 00am. The time it takes him to drive to the appointment is a random variable $X \sim U(20, 45)$.

i) If Professor Halfbrain leaves for his appointment at 9 : 30, what is the probability he will be on time?

$$P(\text{on time}) = P(X \leq 30) = F(30) = \frac{30 - 20}{45 - 20} = 0.4$$

ii) What time should Professor Halfbrain leave if he wants to be 90% sure he arrives on time to his appointment?

We need to find x such that $P(X < x) = 0.9$.

$$\begin{aligned} F(x) &= 0.9 && \Leftrightarrow \\ \frac{x - 20}{25} &= 0.9 && \Leftrightarrow \\ x &= 20 + 25(0.9) && \Leftrightarrow \\ x &= 42.5 \end{aligned}$$

So Prof. Halfbrain needs to leave 42.5 *minutes before his appointment, i.e. at* $8:17:30am$.

iii) **Challenge.** Suppose that Prof. Halfbrain is charged \$5 for every minute that he is late for the appointment. His time is valuable, so he loses \$2 for every minute that he is early. What time should he leave, to minimize his expected cost.

Assume that Prof. Halfbrain leaves $t \in (20, 45)$ minutes before his appointment. Let Y be the amount of money that Prof. Halfbrain will lose.

$$Y = g(X) = \begin{cases} 2(t - X), & X < t \\ 5(X - t), & X \geq t \end{cases}$$

Note that when $X < t$ he is early, and when $X \geq t$ he is late. Prof Halfbrains expected loss is

$$\begin{aligned} E(Y) = E(g(X)) &= \int_{20}^{t} 2(t - x)\frac{1}{25}dx + \int_{t}^{45} 5(x - t)\frac{1}{25}dx \\ &= \frac{2}{25}\left(tx - \frac{1}{2}x^2\right)\Big|_{20}^{t} + \frac{5}{25}\left(\frac{1}{2}x^2 - tx\right)\Big|_{t}^{45} \\ &= \frac{2}{25}\left(\frac{1}{2}t^2 - (20t - 20^2/2)\right) + \frac{5}{25}\left((45^2/2 - 45t) + \frac{1}{2}t^2\right) \\ &= \frac{7}{50}t^2 - \frac{265}{25}t + \frac{10925}{50} \end{aligned}$$

So for example, if he leaves $t = 30$ minutes before his appointment, his expected loss is

$$E(Y) = \frac{7}{50}30^2 - \frac{265}{25}30 + \frac{10925}{50} = \$26.5.$$

The final step, is to choose a value of t which minimizes his expected loss. This comes down to finding the minimum of the polynomial $E(Y) = h(t)$. We take the derivative

$$\frac{dE(Y)}{dt} = \frac{dh(t)}{dt} = \frac{14}{50}t - \frac{265}{25}.$$

Setting the derivative equal to zero and solving for t gives us the answer, $t_{min} = \frac{530}{14} = 37.857$. He should therefore leave 37 minutes and 51 seconds before his appointment, for an expected loss of \$17.86.

Definition: A continuous RV X is said to have an *exponential* distribution, with rate parameter $\lambda > 0$, if its probability density function (PDF) is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x < 0 \\ 0, & \text{otherwise} \end{cases}$$

Notationally, we write $X \sim Exp(\lambda)$. The corresponding CDF is

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\lambda x}, & x > 0 \end{cases}$$

The mean and variance of X are given by:

$$E(X) = \frac{1}{\lambda} \qquad \text{Var}(X) = \frac{1}{\lambda^2}$$

The exponential distribution is defined in terms of its rate parameter λ, due to a connection between the exponential distribution and the discrete Poisson distribution. Note that it is sometimes convenient to define this distribution in terms of a "scale" parameter $\mu = \frac{1}{\lambda}$, so that the mean becomes $E(X) = \mu$ and the variance becomes $\text{Var}(X) = \mu^2$. We will use this "alternate parameterization" a few times in the future, but for the purpose of these notes we stick with the rate parameter λ.

Example. Suppose that the lifetime of a particular brand of lightbulb is an exponential RV X with a mean of 250 days.

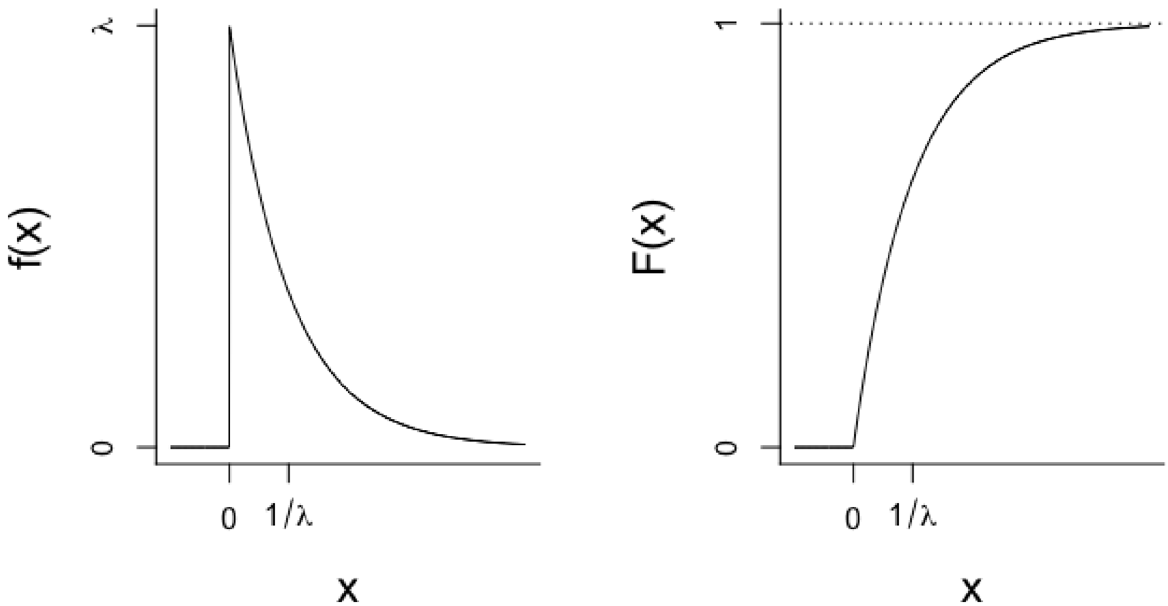

FIGURE 2. Probability density function ($f(x)$) and cumulative distribution function ($F(x)$) for the exponential distribution with rate parameter λ.

i) Write out the PDF and CDF of X.

$$f(x) = \frac{1}{250}e^{-x/250}, \; x > 0$$

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-x/250}, & x > 0 \end{cases}$$

ii) What is the probability that a lightbulb last exactly 100 days?

This is a bit of a trick question. Since we are measuring time continuously (i.e. X is a continuous RV), we know that $P(X = x) = 0$ for any value of x. So $P(X = 100) = 0$.

iii) What is the probability that a lightbulb lasts more than 100 days?

$$P(X > 100) = 1 - P(X \leq 100) = 1 - F(100) = 1 - (1 - e^{-100/250}) = e^{-100/250} = 0.67$$

iv) What is the probability that a lightbulb lasts between 100 and 400 days?

$$P(100 < X < 400) = F(400) - F(100) = 1 - e^{-400/250} - (1 - 0.67) = 0.47$$

v) What is the probability that a lightbulb lasts more than 300 days, given that it will last between 100 and 400 days?

$$P(X > 300|100 < X < 400) = \frac{P(X > 300 \cap 100 < X < 400)}{P(100 < X < 400)} = \frac{P(300 < X < 400)}{0.47} = \frac{F(400) - F(300)}{0.47} = 0.211$$

vi) Suppose that we purchase 5 lightbulbs, and the lifetime of all lightbulbs are independent of each other. Let T be the time before the first lightbulb fails. What is the distribution of T?

It is easy to find the CDF of T, because if $T \leq x$, then at least 1 of the individual lightbulbs must also have a lifetime shorter than x (and vice-versa).

$$\begin{aligned}
P(T \leq x) &= P(\{X_1 \leq x\} \cup \{X_2 \leq x\} \cup \cdots \{X_5 \leq x\}) \\
&= 1 - P(\{X_1 > x\} \cap \{X_2 > x\} \cap \cdots \{X_5 > x\}) && \textit{(DeMorgan)} \\
&= 1 - P(X_1 > x)P(X_2 > x) \cdots P(X_5 > x) && \textit{(independence)} \\
&= 1 - [1 - F(x)]^5 \\
&= 1 - \left[e^{-x/250}\right]^5 && = 1 - e^{-x/50}
\end{aligned}$$

So for instance, we know that the probability of at least one lightbulb going out in the first 100 days is $P(T \leq 100) = 1 - e^{-100/50} = 0.86$. Note that the form of the CDF tells us that $T \sim Exp(1/50)$. So we also know that the expected time before a bulb goes out is 50 days and the variance is 50^2.

Connection to Poisson. Suppose that you are counting the number of events that happen in 1 "unit". Starting at time 0, you have to wait X_1 units for the first event to occur. Then you must wait X_2 units for the second event to occur and so on. These random variables $X_1, X_2, X_3, X_4 \cdots$ are called *arrival times*. In general, arrival times can have any distribution with the extra requirement that they should be non-negative random variables.

Now let Y be the number of events that occur during the one unit of time. For example, suppose that

$$x_1 = 0.3 \quad x_2 = 0.05 \quad x_3 = 0.45 \quad x_4 = 0.15 \quad x_5 = 0.2 \quad \cdots$$

in this case, Y would be equal to 4, because the first 4 events occur in the first 1.0 unit of time, but the fifth event does not occur time 1.15. The following theorem shows the relationship between the exponential and Poisson distributions. We will prove this (hopefully) in a few weeks.

Theorem: The distribution of Y is $Poiss(\lambda)$ if (and only if) the arrival times are independent exponential distributions, i.e. $X_i \overset{ind}{\sim} Exp(\lambda)$.

The Memorylessness Property: Let $X \sim Exp(\lambda)$ and let s and t be constants, such that $s > t$. Then

$$P(X > s | X > t) = P(X > s - t)$$

Proof: The proof is straightforward. First, note that the RHS of the above equation can be written as

$$P(X > s - t) = 1 - F(s - t) = e^{-\lambda(s-t)}.$$

Now, using the definition of conditional probability

$$P(X > s | X > t) = \frac{P(X > s \cap X > t)}{P(X > t)} = \frac{P(X > s)}{P(X > t)} = \frac{e^{-\lambda s}}{e^{-\lambda t}} = e^{-\lambda(s-t)} \qquad \square$$

Not counting Professor Halfbrain, the Exponential distribution is the only continuous distribution that has this property for all $s > t$. It's an interesting property that basically says "X has no memory of how long it has been waiting to occur". For example, if $X \sim Exp(1)$, then we expect to wait just one minute for the event to occur. If we have already waited 10 minutes, we still "expect" to wait one more minute. Interestingly, this seems to be the exact behavior observed for things like the decay of radioactive particles.

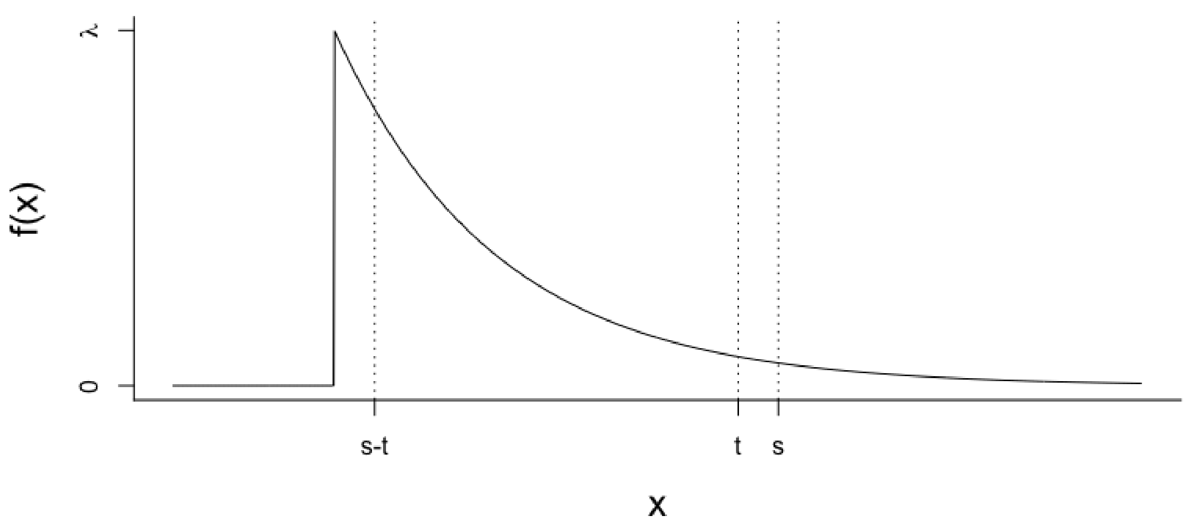

FIGURE 3. The memorylessness property.

Skewness. We have defined the expected value of a random variable X for both the continuous and discrete case, i.e. $\mu = E(X)$. The variance is another parameter which describes the distribution and can be defined as

$$\sigma^2 = \text{Var}(X) = E\left((X-\mu)^2\right) = E(X^2) - \mu^2$$

The *skewness* of a RV X can be defined mathematically as

$$\gamma = \text{Skew}(X) = E\left(\left(\frac{X-\mu}{\sigma}\right)^3\right) = \frac{E(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3},$$

where μ and σ^2 are the expected value (i.e. mean) and variance of a RV respectively. When $\gamma > 0$ we say that a distribution is right (or positive) skewed and when $\gamma < 0$ we say that a distribution is left (or negative) skewed. When $\gamma = 0$, the distribution is usually (but not always) *symmetric*.

For an exponential distribution, we know that $\mu = 1/\lambda$ and $\sigma^2 = 1/\lambda^2$. To find the skew, we just need to find $E(X^3)$.

$$E(X^3) = \int_0^\infty x^3 \lambda e^{-\lambda x} dx = \cdots = \frac{6}{\lambda^3}$$

It is good for you to do this integral on your own for practice, but it can be easily done with integration by parts or tabular integration. Now to find $\gamma = Skew(X)$, we simply plug $E(X^3)$ into the equation

$$Skew(X) = \frac{6\lambda^{-3} - 3\lambda^{-1}\lambda^{-2} - \lambda^{-3}}{\lambda^{-3}} = \frac{2\lambda^{-3}}{\lambda^{-3}} = 2.$$

Thus the exponential distribution is right skewed with a skewness of (positive) 2 for all values of λ.

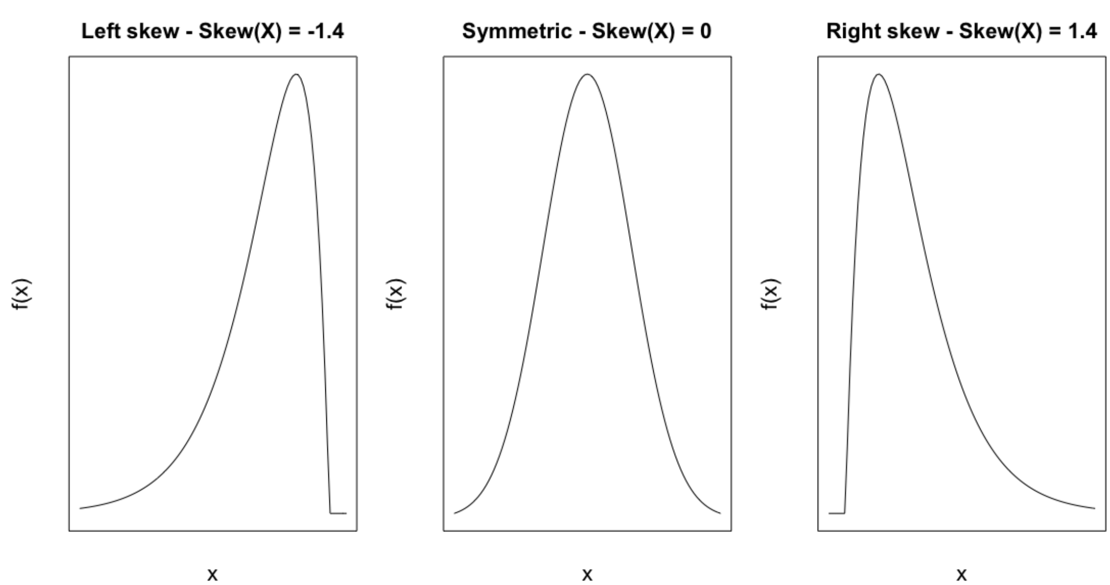

FIGURE 4. Skewness of a distribution.

We finish these notes with two more simple results.

Claim: If $X \sim Exp(\lambda)$ and $Y = \frac{1}{a}X$, then Y also has an exponential distribution with rate parameter $a\lambda$.

Proof: We prove this by finding the CDF of Y and showing that it has the form of an exponential CDF. Let $F_X(x) = 1 - e^{-\lambda x}$ be the CDF of X.

$$\begin{aligned} F_Y(Y) &= P(Y \leq y) \\ &= P\left(\frac{X}{a} \leq y\right) \\ &= P(X \leq ay) \\ &= F_X(ay) \\ &= 1 - e^{-\lambda(ay)} \\ &= 1 - e^{-(a\lambda)y} \end{aligned}$$

$\square$

We will now use a similar approach to derive an entirely new distribution, known as the *Weibull*.

Definition: Let $X \sim Exp(1)$ and let $\beta > 0$ and $\kappa > 0$ be constants. Then the RV $Y = \left(\frac{X}{\beta}\right)^{1/\kappa}$ has a Weibull distribution with parameters β and κ, i.e. $Y \sim \text{Weib}(\beta, \kappa)$.

We will derive its CDF using the same technique as in the previous example. The CDF of X is $F_X(x) = 1 - e^{-x}$.

$$\begin{aligned} F_Y(Y) &= P(Y \leq y) \\ &= P\left(\left(\frac{X}{\beta}\right)^{1/\kappa} \leq y\right) \\ &= P(X \leq \beta y^\kappa) \\ &= F_X(\beta y^\kappa) \\ &= \begin{cases} 0, & y \leq 0 \\ 1 - e^{-\beta y^\kappa}, & y > 0 \end{cases} \end{aligned}$$

To find the PDF of Y, we must take the derivative of the CDF (with respect to y).

$$f_Y(y) = \kappa\beta y^{\kappa-1} e^{-\beta y^\kappa},\ y > 0.$$

The Weibull distribution is very flexible, in the sense that it can be left skewed, right skewed or symmetric depending on the chosen parameters. We will calculate the mean and variance of this distribution after spring break.