Module C: Understanding the Data-Generating Process

Slide Deck C8:

The Exponential Distribution

The section in which we learn about our second named continuous distribution. The Exponential distribution is used to model the time until some event occurs for the first time.

Today's Objectives

By the end of this slidedeck, you should

1. Determine what random variables may follow an Exponential distribution
2. Specify the characteristics of an Exponential distribution
3. Compare and contrast the Exponential with the Uniform distribution
4. Calculate probabilities using the CDF
5. Calculate the quantiles using the CDF

Characteristics of the Exponential

The Exponential distribution is a continuous distribution that describes random variables whose probability of occurring decreases with time. It is frequently used to model "time until" something happens when there is no upper bound.

$$X \sim \mathcal{E}xp(\lambda)$$

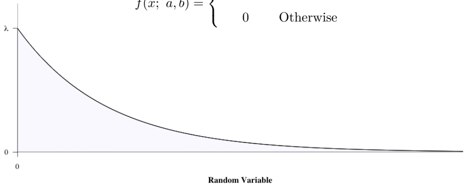
It is characterized by these features

- One parameter: λ, the rate of the thing happening
- Sample space is $\mathcal{S} = (0, \infty)$
- $\mathbb{E}\left[\, X \,\right] = \frac{1}{\lambda}$
- $\mathbb{V}\left[\, X \,\right] = \frac{1}{\lambda^2}$

Probability Density Function

The probability density function for the Exponential distribution is given by

$$f(x;\ a, b) = \begin{cases} \lambda\ \mathrm{e}^{-\lambda x} & x > 0 \\ 0 & \text{Otherwise} \end{cases}$$



Purposes of pdf

Again, note that the probability density function has two purposes:

1. To help the researcher understand probability for a continuous distribution
2. To help the researcher calculate probabilities of a continuous random variable

Important!!

Probabilities are **areas** under the *density* curve.

Example I: Chunky the Squirrel

Example

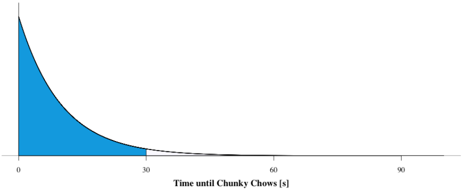
The time between when I fill by bird feeder with seed and when Chunky the Squirrel starts eating from the feeder follows an Exponential distribution with an average time of 10s. In other words, if we define T as the time in seconds until Chunky raids the bird feeder,

$$T \sim \mathcal{E}xp(\lambda = 1/10)$$

What is the probability that he waits at most a half-minute before chowing down on some awesome seed?

Example I: Chunky the Squirrel

We are asked to calculate $\mathbb{P}[\ T \leq 30\ ]$. The dark color is the area we are to calculate:



Example I: Chunky the Squirrel

Because we are asked to calculate $\mathbb{P}[\ X \leq x\ ]$, we can use the CDF for the Exponential distribution.

The Exponential CDF

$$\mathbb{P}[\ X \leq x\ ] = 1 - e^{-\lambda x}$$

The calculation required to obtain this formula requires Integral Calculus. As such, it will be ignored. I have included it in the appendix to these slides, however.

Example I: Chunky the Squirrel

Thus, the probability that it takes Chunky no more than 30s to start feeding on the expensive bird food is

$$\begin{aligned}
\mathbb{P}[\ T \leq 30\ ] &= F(30; \lambda = 0.10) \\
&= 1 - e^{-\lambda x} \\
&= 1 - e^{-(0.10)(30)} \\
&= 1 - e^{-3} \\
&= 1 - 0.049787068367864 \\
&= 0.950212931632136
\end{aligned}$$

Thus, there is a 95% probability that Chunky will wait no more than 30s to raid the feeder:

```
pexp(30, rate=1/10)
```


Example II: The Gold Express

Example

The time I wait until the Gold Express bus comes follows an Exponential distribution. My average wait time is 5 minutes. In the cold-of-winter, it takes 10 minutes for frostbite to start. Given this, what is the probability that I will have frostbite before the bus arrives?

Example II: The Gold Express

Solution
Given that T is the time I wait for the bus (in minutes), we are given

$$T \sim \mathcal{E}xp(\lambda = 1/5)$$

We are asked to calculate $\mathbb{P}[\ T > 10\ ]$.

This is not in the accepted form of a CDF, which is $\mathbb{P}[\ X \leq x\ ]$.

However, we can easily convert it to such a form. Recall the concept of *complements* from Slidedeck C2

$$\mathbb{P}[\ T > 10\ ] = 1 - \mathbb{P}[\ T \leq 10\ ]$$

Example II: The Gold Express

With this, we have

$$\begin{aligned}
\mathbb{P}[\ T > 10\ ] &= 1 - \mathbb{P}[\ T \leq 10\ ] \\
&= 1 - F(10; \lambda = 0.20) \\
&= 1 - \left(1 - \mathrm{e}^{-(0.20)10}\right) \\
&= \mathrm{e}^{-(0.20)10} \\
&= \mathrm{e}^{-2} \\
&= 0.135335283236613
\end{aligned}$$

Thus, the probability that I wait more than 10 minutes is only 13.5%. In R, this is

```
1-pexp(10, rate=0.20)
```

Example II: The Gold Express

From the problem statement, we know the mean time until the bus arrives is 5 minutes. However, we can also calculate the median time until the bus arrives, $\tilde{x}$.

$$\begin{aligned}
\mathbb{P}[\ X \leq \tilde{x}\ ] &= 0.500 \\
1 - e^{-(0.20)\tilde{x}} &= 0.500 \\
e^{-(0.20)\tilde{x}} &= 0.500 \\
-(0.20)\tilde{x} &= \ln(0.500) = \ln\left(\frac{1}{2}\right) = -\ln 2 \\
\tilde{x} &= \frac{\ln 2}{0.20} = 5 \ln 2 = \mu \ln 2 \\
&= 3.466 \text{ minutes}
\end{aligned}$$

In R, this is `qexp(0.5, rate=0.20)`

Example II: The Gold Express

Aside: In the solution, we determined

$$\tilde{x} = \mu \ln 2$$

Who cares? This proves that the Exponential distribution is skewed positive according to the Hildebrand ratio:

$$\begin{aligned}
H &= \frac{\mu - \tilde{x}}{\sigma} \\
&= \frac{\mu - \mu \ln 2}{\mu} \\
&= 1 - \ln 2 \\
&= 0.306\,852
\end{aligned}$$

Example II: The Gold Express

We can also calculate things like the 90th percentile (quantile). Why? Perhaps we want to perform some sort of robust analysis, wondering how frequently the bus driver has to get the bus by my stop.

$$\begin{aligned}
\mathbb{P}[\ X \leq x^* \ ] &= 0.90 \\
1 - e^{-(0.20)x^*} &= 0.90 \\
e^{-(0.20)x^*} &= 0.100 \\
-(0.20)x^* &= \ln(0.100) \\
x^* &= -\frac{\ln(0.100)}{0.20} \\
&= 11.51 \text{ minutes}
\end{aligned}$$

In R: `qexp(0.9, rate=0.20)`

Today's Objectives

Now that we have concluded this lecture, you should be able to

1. Probability is the area under the pdf curve
2. The cumulative distribution function (CDF) is the probability $\mathbb{P}[\ X \leq x\ ]$
3. Sometimes, calculus is needed to find the CDF
4. The Exponential distribution describes a “wait time” random variable
5. The mean and variance of an Exponential random variable are $\frac{1}{\lambda}$ and $\frac{1}{\lambda^2}$, respectively

Useful R Functions

In this slide deck, we covered two R functions. This is in addition to one we have already experienced and one we will experience:

- `dexp(x, rate)` is the density function, $f(x)$
- `pexp(x, rate)` is the CDF $F(x) = \mathbb{P}[\ X \leq x\ ]$
- `qexp(p, rate)` is the quantile function $\mathbb{P}[\ X \leq x\ ] = p$
- `rexp(n, rate)` gives a random sample from this distribution

Please do not forget to use the `allProbabilities` file that lists all of the probability functions in R.

Supplemental Activities

The following are supplements for the topics covered today.

- SCA 6a is for continuous distributions like the Exponential.

Note that you can access all Statistical Computing Activities here:
`https://www.kvasaheim.com/courses/stat200/sca/`

In addition to the SCA, **Laboratory Activity C** is helpful for learning how to handle some continuous distributions (including the Exponential distribution). The lab actually illustrates the Central Limit Theorem.
`https://www.kvasaheim.com/courses/stat200/labs/`

Supplemental Readings

The following are some readings that may be of interest to you in terms of understanding continuous distributions, including the Exponential:

- Hawkes Learning: None
- Intro to Modern Statistics: None
- R for Starters: Appendix B.5 (and B.6, perhaps)
- Wikipedia: Exponential Distribution
  Gamma Distribution

Calculus Extra

Here is a proof of the formula for the Exponential CDF:

$$\begin{aligned}
\mathbb{P}[\ X \leq x\ ] &:= \int_0^x f(t)\ dt \\
&= \int_0^x \lambda\ e^{-\lambda t}\ dt \\
&= \lambda \int_0^x e^{-\lambda t}\ dt \\
&= \lambda \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_{t=0}^{x} \\
&= \left[-e^{-\lambda t} \right]_{t=0}^{x} \\
&= (-e^{-\lambda x}) - (-e^{-\lambda 0}) \\
&= (-e^{-\lambda x}) + 1 \\
\implies \quad \mathbb{P}[\ X \leq x\ ] &= 1 - e^{-\lambda x}
\end{aligned}$$