---
title: "Distributions in-class exercise"
subtitle: "XDASI Fall 2021"
author: "Lauren Audi, Sofia Sessa, Eric Laderman"
date: "11/1/2021"
output: html_document

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

Scenario

You want to clone a gene and you have some reporter for success of transformation in *E. coli*. Positive colonies are blue because they express β-galactosidase when plated on medium with IPTG. The probability of a successful transformation is $0 \le P(success) \le 1$. For the following exercise, you may consider one plate, many plates, or a truly gargantuan number of plates.

Exercise

The class has been split into groups and each has taken one or two of the following distributions:

- Bernoulli
- Binomial
- Hypergeometric
- Geometric
- Negative binomial
- Poisson
- Normal
- Exponential

For your assigned distribution, answer the following questions.

a. What is your function? Describe it in terms of the kind of problem that it applies to. Is it discrete or continuous?

Our function is exponential. It is continuous and describes the length of time between events. It applies to problems like radioactive decay and survivorship (length of time before something happens).

b. Describe a question that you can ask with your distribution as applied to the transformation scenario given above.

Our question is how long do you have to wait to get a "good" colony (i.e. survivorship)?

c. Describe the parameterization of the equation. What parameters does it take, and what do they mean?

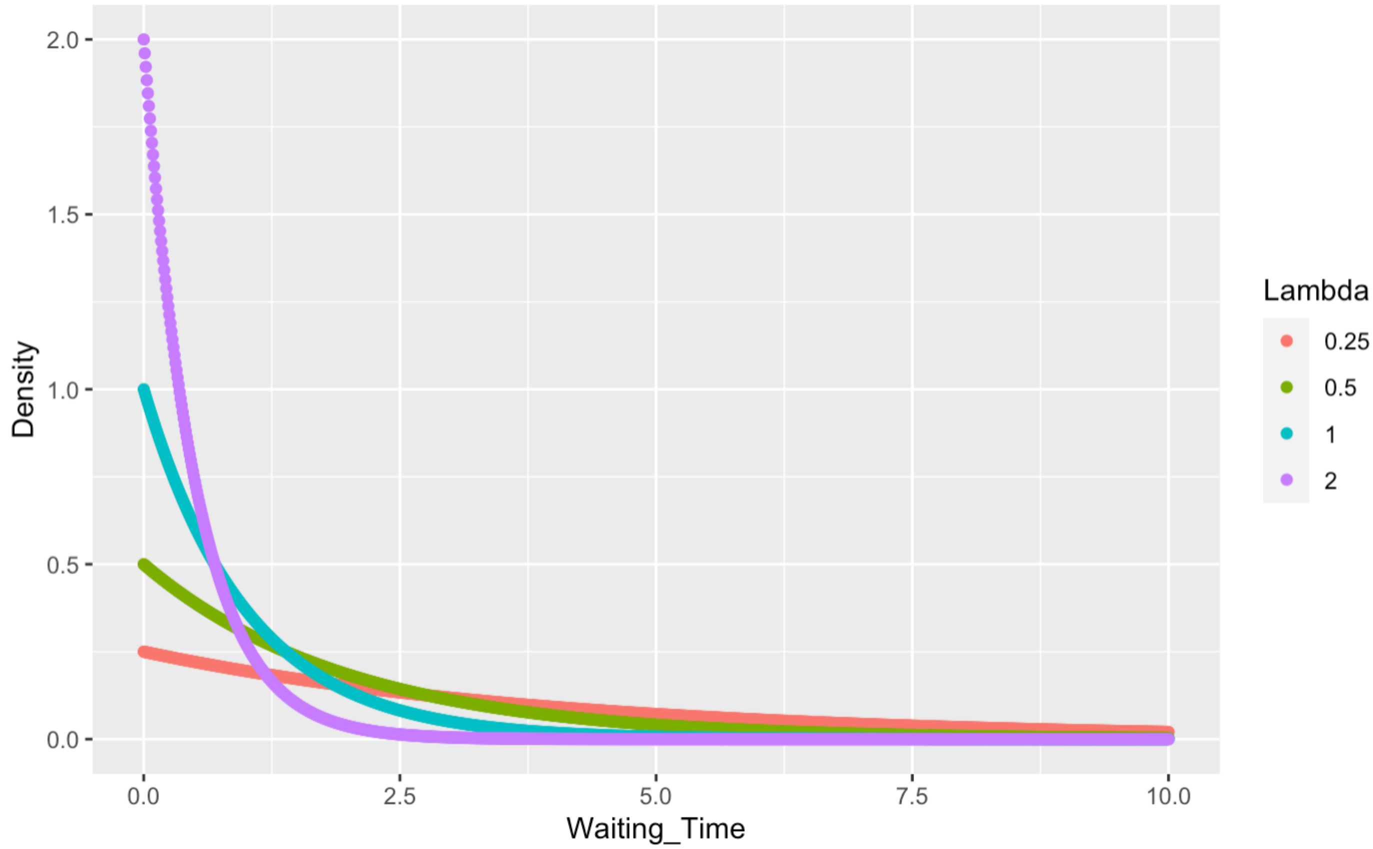
lamda = rate constant (events per unit time)
t = time
x = random variable that represents an interval of time and space

d. Pick some appropriate parameters for your distribution as applied to the transformation scenario and graph the PDF and CDF.

```
# make a vector with lambdas to plot and values to check for each lambda using dexp
lambdas = c(0.25, 0.5, 1, 2)

values_to_check = seq(0,10, by=0.01)
lambdas_for_df = c()
times = c()
probs = c()
for (lambda in lambdas) {
  lambdas_for_df = c(lambdas_for_df, rep(lambda, length(values_to_check)))
  times = c(times, values_to_check)
  probs = c(probs, dexp(values_to_check, lambda))
}

df_to_plot = data.frame('Waiting_Time' = times, 'Lambda' = as.factor(lambdas_for_df),
'Density' = probs)

ggplot(data = df_to_plot, mapping = aes(x = Waiting_Time, y = Density, color = Lambda))
+ geom_point()

time.interval<-seq(0, 10, 0.05)
```

```
plot(time.interval,dexp(time.interval,rate=0.5),type = "l",
xlab="Time(hours)",ylab="Density of Failures (i.e prob that your random var. is <= to
time(t)", main="Exponential PDF Example")

lines(interval.exp,dexp(interval.exp,rate=1),col="red")
lines(interval.exp,dexp(interval.exp,rate=2),col="blue")
lines(interval.exp,dexp(interval.exp,rate=5),col="green")
legend("topright", legend=c("lambda=0.5","lambda=1", "lambda=2","lambda=5"),
       col=c("black"," red", "blue", "green"), lty = 1, cex=0.8)
```

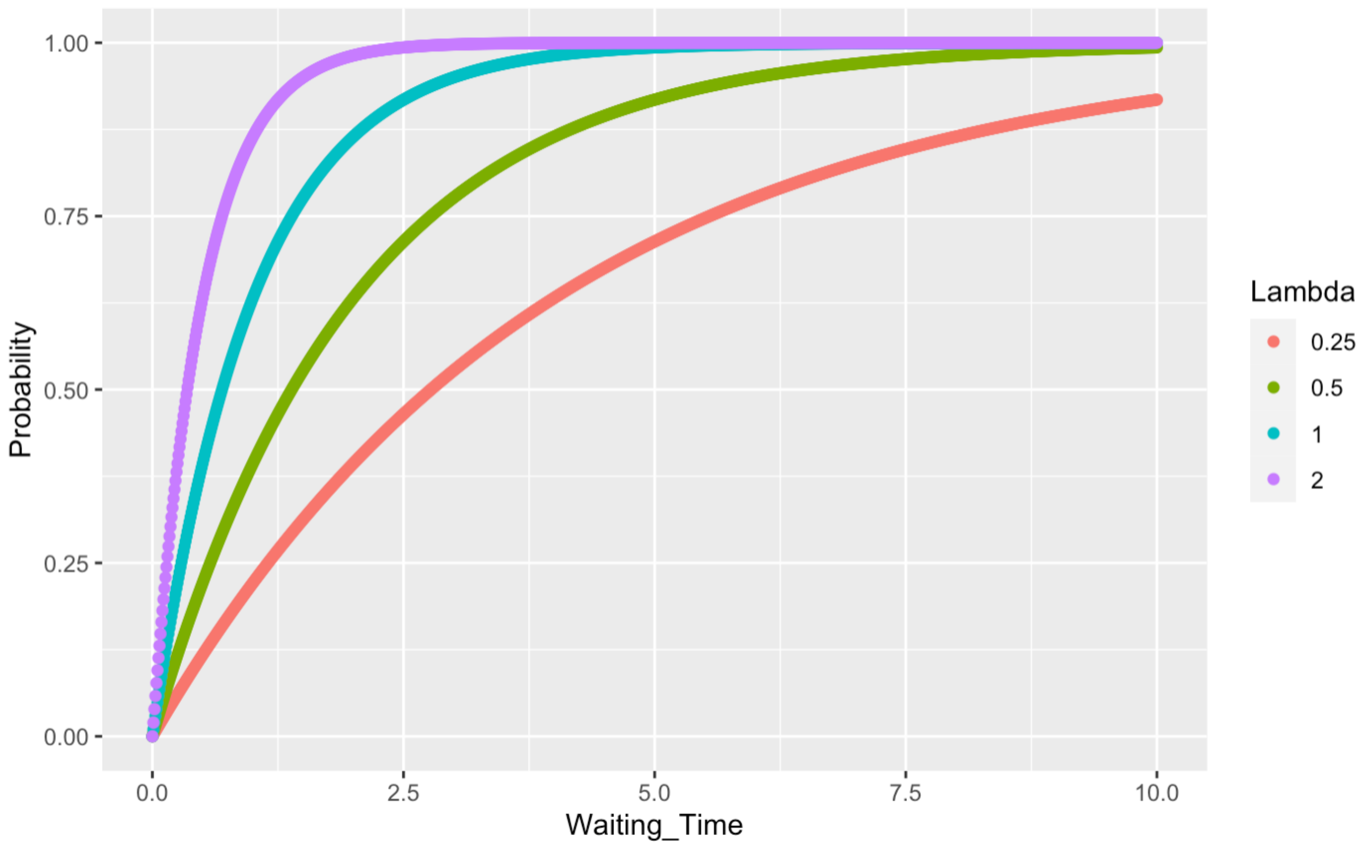
Example PDF



Exponential CDF - has a closed form

```
# make a vector with lambdas to plot and values to check for each lambda using pexp to
make a CDF
lambdas = c(0.25, 0.5, 1, 2) # events per unit time

values_to_check = seq(0,10, by=0.01)
lambdas_for_df = c()
times = c()
probs = c()
for (lambda in lambdas) {
  lambdas_for_df = c(lambdas_for_df, rep(lambda, length(values_to_check)))
  times = c(times, values_to_check)
  probs = c(probs, pexp(values_to_check, lambda))
}

df_to_plot = data.frame('Waiting_Time' = times, 'Lambda' = as.factor(lambdas_for_df),
'Probability' = probs)

ggplot(data = df_to_plot, mapping = aes(x = Waiting_Time, y = Probability, color =
Lambda)) + geom_point()
```

Example CDF




```

#### e. Compute a cumulative probability for some appropriate value and describe in words what it means.

```
lambdas = c(0.25, 0.5, 1, 2)

pexp(1, lambdas, lower.tail=T) # p of finding at least 1 good colony per different rates (lambdas)

#probability of 0.9816844 when rate is 0.25 events per waiting time
#probability of 0.8646647 when rate is 0.5 events per waiting time
#probability of 0.6321206 when rate is 1 events per waiting time
#probability of 0.3934693 when rate is 2 events per waiting time
```

#### f. Identify at least one other distribution that yours relates to and be prepared to discuss the relationship between them. This could be a special case, a limiting case, an approximation, or an inverse CDF (survival) relationship.

Exponential relates closely to Poisson and is the inverse distribution of the Poisson. Whereas Poisson is a discrete function that describes the number of events per unit time while exponential is continuous and thus describes length of time between events. For example, in exponential, the integral under the curve between those two points describes that probability.
```