Distributions In-Class Exercise

XDAS 2019 group effort

September 30, 2019

Contents

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 $0 \leq P(success) \leq 1$. For the following exercise, you may consider one plate, many plates, or a truly gargantuan number of plates.

Exercise

The class will split into 8 groups and each one will take one of the following distributions:

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

For your assigned distribution, answer the following questions. At the end of the class, each group will present their distribution and how it applies to the scenario above to the rest of the class.

Bernoulli

Cassandra and Yen

Function Summary

$$PDF: \quad f(x) = P(X = x) = \pi^x (1 - \pi)^{1-x}, \quad x \in \{0, 1\}$$

where π represents the probability of "success", and ranges from zero to one: $0 \leq \pi \leq 1$.

Since x can only take on values of 0 or 1, and there is only one trial, $f(x)$ can take on only one of two values:

$$f(1) = P(X = 1) = \pi \quad or \quad f(0) = P(X = 0) = 1 - \pi$$

$$CDF: \quad F(X) = \begin{cases} 1 - \pi & x = 0 \\ 1 & x = 1 \end{cases}$$

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

The function we choose is Bernoulli experiment. It's a special case of binomial distribution where n = 1. It's describing a single trial of a binary random variable. It applies to many of the experiments that we want to conceptualize the success rate (like cloning, transfection, gene editting, and etc.). It's the basis of how we can estimate the probability of having a specific number of success (binomial), how many experiments we need before getting a or x success (negative binomial / geometric).

Bernoulli distribution is discrete.

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

With preliminary experiments to estimate the success rate p, in the transformation with reporterexperiment, we can use Bernoulli experiment to ask ***how likely each bacterium will take in the plasmid and be transformed****, thus expressing the reporter gene LacZ or GFP.*

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

The PDF and CDF of Bernoulli distribution only takes p (the success rate) and x (the number of success cases, in which can only be 0 or 1), where the PDF is:

$$P(x) = \begin{cases} p \quad if \quad x = 1 \\ \\ 1 - p \quad if \quad x = 0 \end{cases}$$

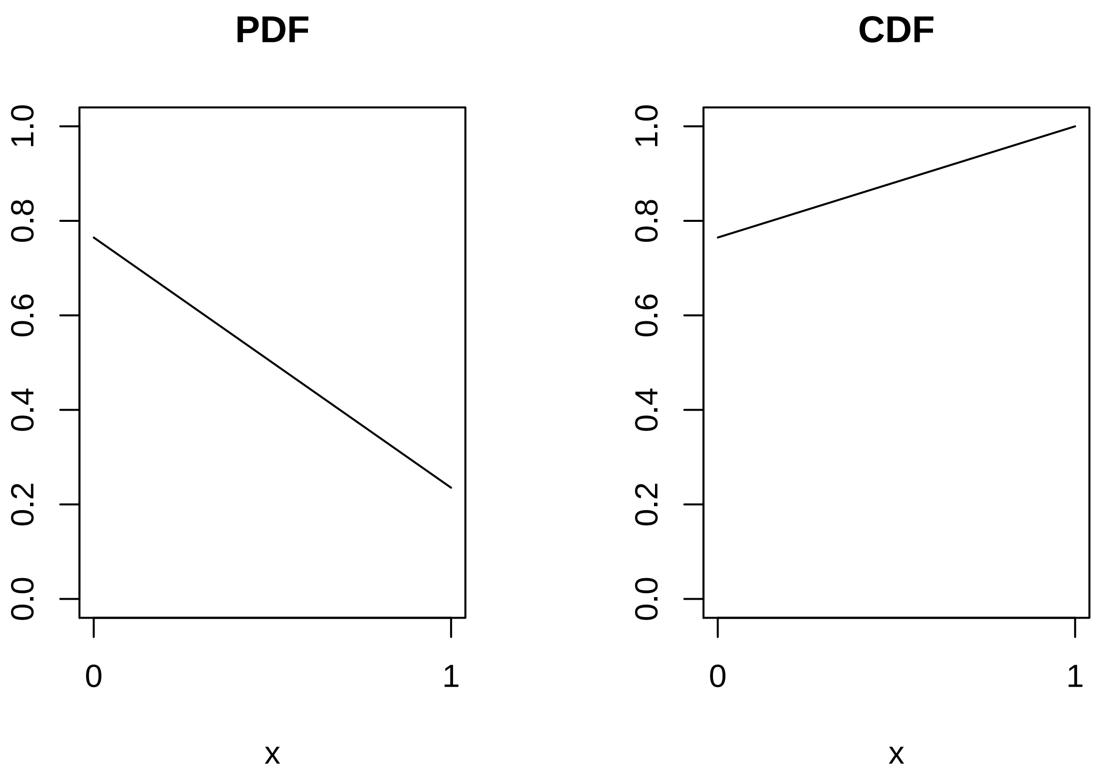
and the CDF is:

$$F(x) = \begin{cases} 1 \quad if \quad x = 1 \\ \\ 1 - p \quad if \quad x = 0 \end{cases}$$

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

The transformation rate is ~23% ($p = \frac{4}{17}$).

```r
set.seed(20190930)
par(mfrow = c(1,2))
plot(x= c(0, 1), y = dbinom(c(0, 1), 1, 4/17), type = "l",
     main = "PDF", xlab = "x", ylab = "", ylim = c(0, 1), xaxt='n')
axis(side = 1, at = c(0, 1))
plot(x= c(0, 1), y = pbinom(c(0, 1), 1, 4/17), type = "l",
     main = "CDF", xlab = "x", ylab = "", ylim = c(0, 1), xaxt='n')
axis(side = 1, at = c(0, 1))
```



e. Compute a cumulative probability for some appropriate value and describe in words what it means. Be prepared to explain your answer to the class.

By definition, Bernoulli distribution only exist for $x \in \{0, 1\}$, so and it has a CDF:

$$F(x) = \begin{cases} 1 \quad if \quad x = 1 \\ \\ 1 - p \quad if \quad x = 0 \end{cases}$$

It means that the cumulative chance to getting no success ($x = 0$) is $1 - p$, while the chance of getting equal or less than 1 success is the sum of every possible case, thus $P(x = 1) = 1$.

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.

Binomial distribution can be seen as repeating Bernoulli trial for n times.

Binomial

Rebecca and Adhiti

Function Summary

$$X \sim BIN(n, \pi),\ with\ \mu = E(X) = np\ and\ \sigma^2 = V(X) = np(1-p)$$

$$PDF: \quad f(x) = \binom{n}{x} \pi^x (1-\pi)^{n-x}, \quad x \in \{0, 1, ..., n\}$$

$$CDF: \quad F(X) = \sum_{x=0}^{n} \binom{n}{x} \pi^x (1-\pi)^{n-x} = 1$$

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

We have the binomial distribution, which is the binom function in R. This is a statistical method for describing the probability of successes and failures. It is asking a yes/no question. It is a discreet probability

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

Out of 10 experiments (n or number of experiments), how many positively transformed (via blue color) colonies are observed on a given plate (x or number of successes)?

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

X ~ BIN(n,pi). It uses n for the number of trials and x for the overall number of successes

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

```r
dbinom(1,10, (1/3))
```

```
## [1] 0.08670765
```

```r
# What is the probability of 1 success out of 10 trials given the probability of success is 1/3 in a si

dbinom(0:10, 10, (1/3))
```

```
##  [1] 1.734153e-02 8.670765e-02 1.950922e-01 2.601229e-01 2.276076e-01
##  [6] 1.365645e-01 5.690190e-02 1.625768e-02 3.048316e-03 3.387018e-04
## [11] 1.693509e-05
```

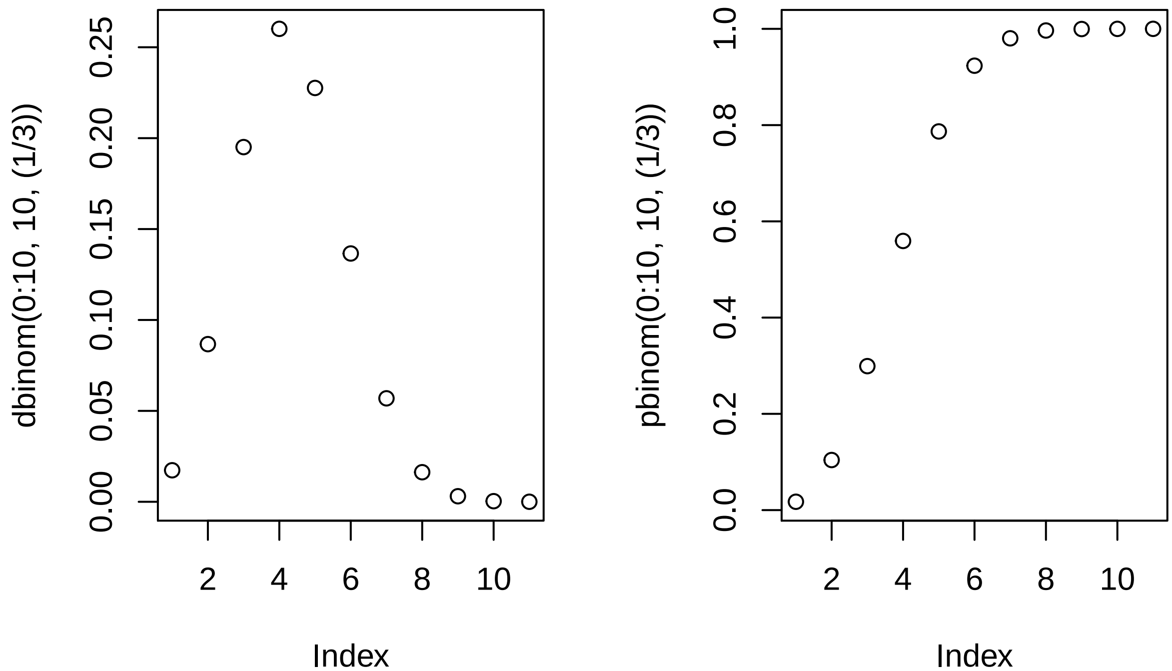
```r
#Then we wanted to know what the discrete probabilities would be for 0-10 trials given the observed pro
```

e. Compute a cumulative probability for some appropriate value and describe in words what it means. Be prepared to explain your answer to the class.

```
pbinom(0:10, 10, (1/3))
```

```
##  [1] 0.01734153 0.10404918 0.29914139 0.55926434 0.78687192 0.92343647
##  [7] 0.98033836 0.99659605 0.99964436 0.99998306 1.00000000
```

```
par(mfrow = c(1,2))
plot(dbinom(0:10,10, (1/3)))
plot(pbinom(0:10,10, (1/3)))
```



Here we have computed the cumultive probability or CDF of getting 0-10 successes out of 10 trials (adding the individual probabilities for each of the 10 trials).

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.

We could use a negative binomial distribution in order to check the faliure rate and see if it matches up

Geometric

Daniel and Larissa

Function Summary

$$X \sim GEO(p), \quad E(x) = 1/p \;\; and \;\; Var(X) = (1-p)/p^2$$

The geometric distribution is a special case of the NB, where $r = 1$: $X \sim GEO(p)$ is equivalent to $X \sim NB(1, p)$.

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

Geometric function: It gives the probability of getting x number of failures before a success

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

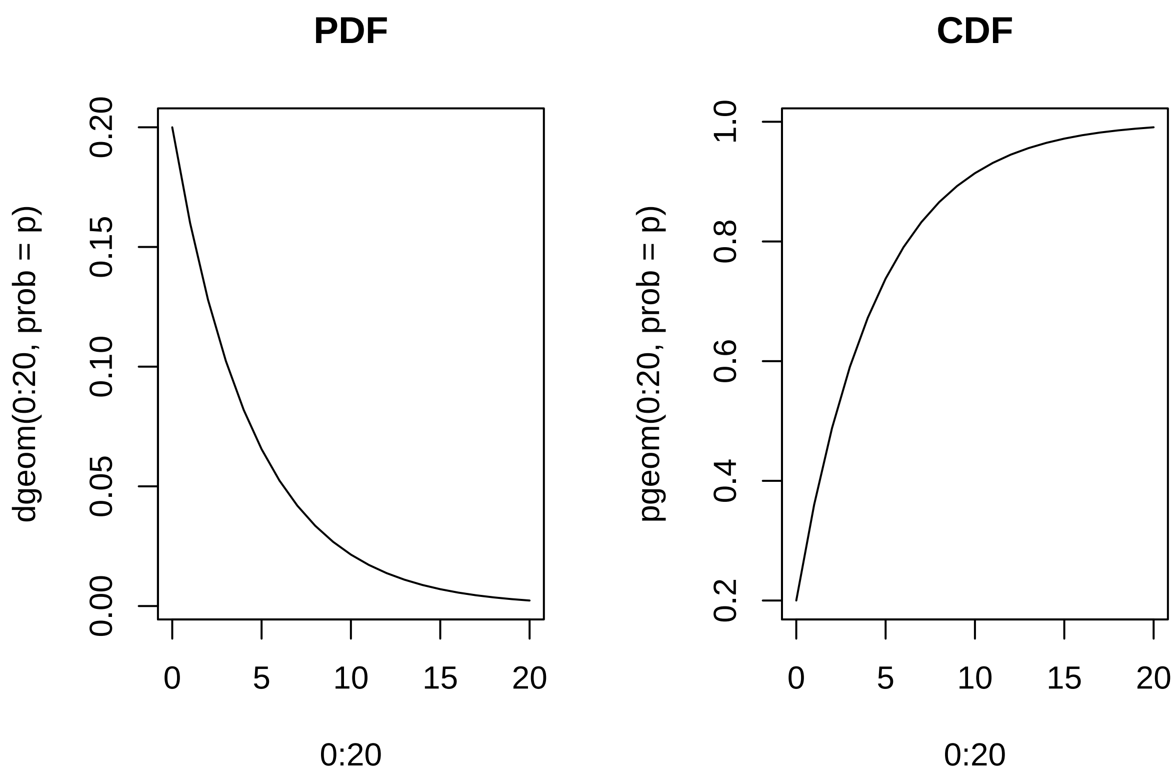
When randomly picking colonies from yor plate, what is the probability of getting 6 white colonies before getting a green one?

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

The equation takes two parameters: x - number of failures p - probability of success

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

```
p = 0.2
par(mfrow=c(1,2))
plot(0:20, dgeom(0:20,prob = p), type = "l", main = "PDF")
plot(0:20, pgeom(0:20,prob = p), type = "l", main = "CDF")
```



e. Compute a cumulative probability for some appropriate value and describe in words what it means. Be prepared to explain your answer to the class.

```
pgeom(6,prob = p)

## [1] 0.7902848
```

The probability of getting 6 and less white colonies before getting a green colony is 0.79.

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.

Negative binomial. The geometric function is a specific case of the negative binomial where r;number of successes is 1.

Hypergeometric

Jackie and Soobeom

Function Summary

$$X \sim HYP(n, M, N)$$

$$E(X) = nM/N, \quad Var(X) = n\frac{M}{N}\left(1 - \frac{M}{N}\right)\frac{N-n}{N-1}$$

$$PDF: \quad f(x) = P(X = x) = \binom{M}{x}\binom{N-M}{n-x} \Big/ \binom{N}{n}$$

where:

- $x \in \{0, 1, 2, ..., n\}$ the number of successful trials
- $N \in \{1, 2, ...\}$ is the total number of selectable items
- $M \in \{0, 1, 2, ..., N\}$ is the total number of possible successful outcomes (i.e. th number of items in the group of interest)
- $n \in \{0, 1, 2, ..., N\}$ is the number of items sampled

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

Hyper Geometric: It is a discrete function and you sample without replacement from a universal value. Each trial outcome affects the outcome of the next trial.

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

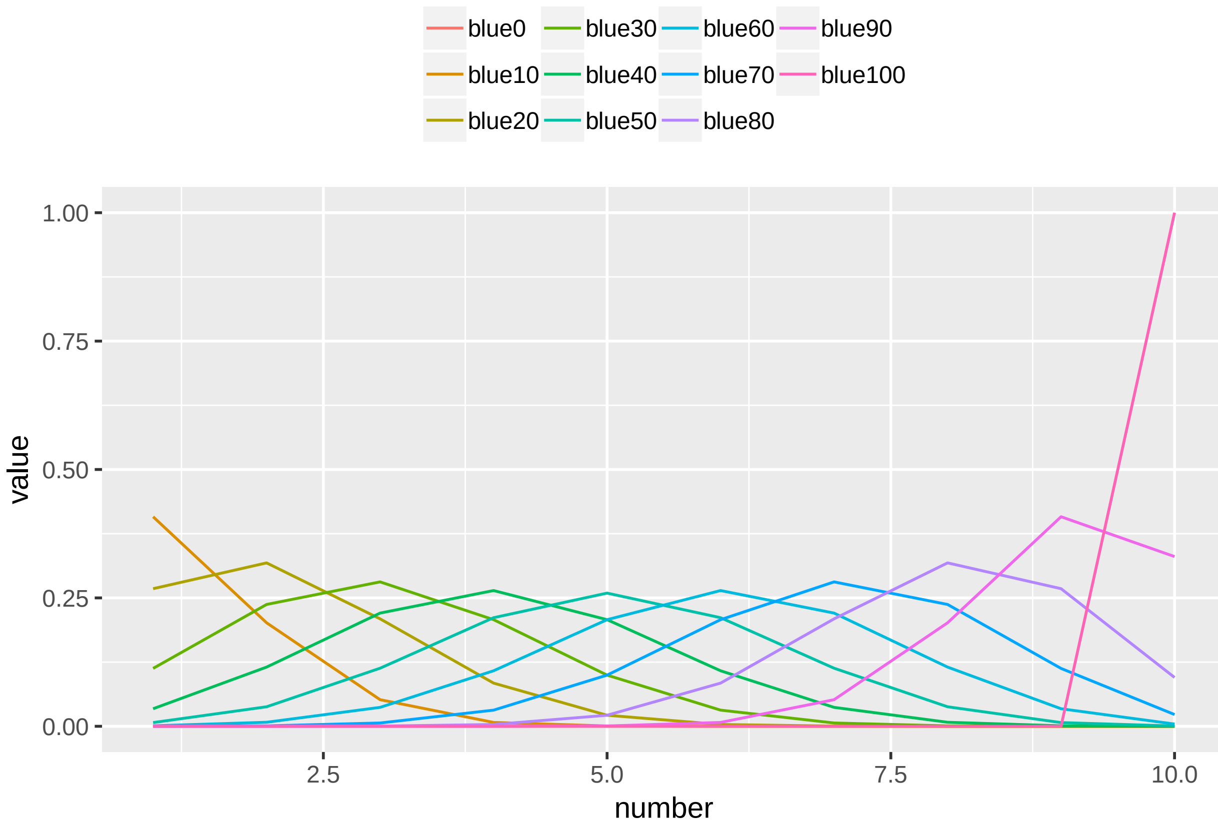
There are 100 bacterial colonies on a plate. 25 colonies are blue and 75 colonies are white. If you take a sample 10 colonies, whats the probability of getting 10 blue transformed colonies after 10 trials?

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

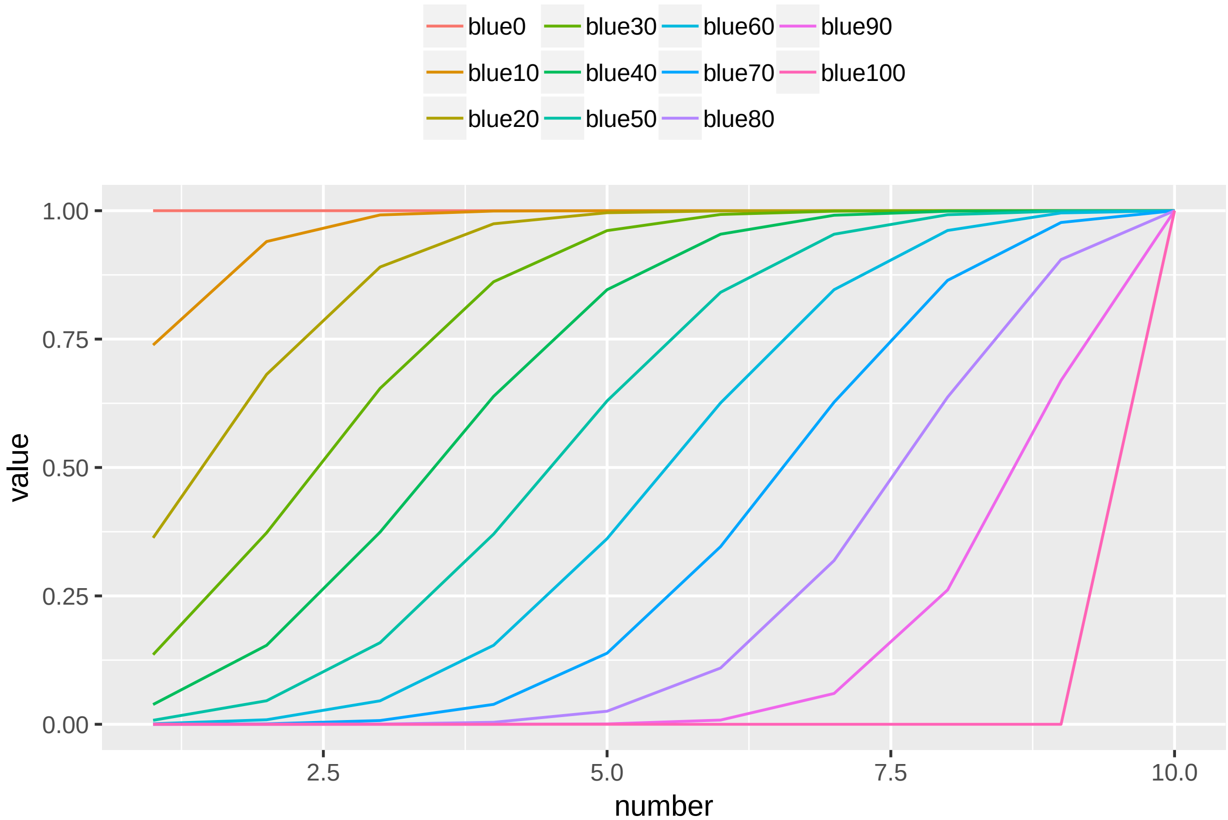
Hyper Geometric parameters: N = 100 n = 10 x = 10 M = 25 <- blue M = 75 <- white

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

```r
library(ggplot2)
library(reshape2)

blue <- c(0,10,20,30,40,50,60,70,80,90,100) # our parameter

# make PDF
pdf <- NULL
for (i in blue){
  xx <- dhyper(1:10, i, 100-i, 10)
  pdf <- cbind(pdf, xx)
}

colnames(pdf) <- paste("blue", blue, sep = "")

# make CDF
cdf <- NULL
for (i in blue){
  xx <- phyper(1:10, i, 100-i, 10)
  cdf <- cbind(cdf, xx)
}

colnames(cdf) <- paste("blue", blue, sep = "")

# make dataframe for ggplot
pdf2 <- data.frame(pdf, number=1:10)
pdf2 <- melt(pdf2, "number")
cdf2 <- data.frame(cdf, number=1:10)
cdf2 <- melt(cdf2, "number")

# plot PDF with 10 different parameters
ggplot()+
  geom_line(aes(y = value, x = number, colour = variable),
                   data = pdf2, stat="identity") +
  theme(legend.title = element_blank(),
        legend.position="top")
```

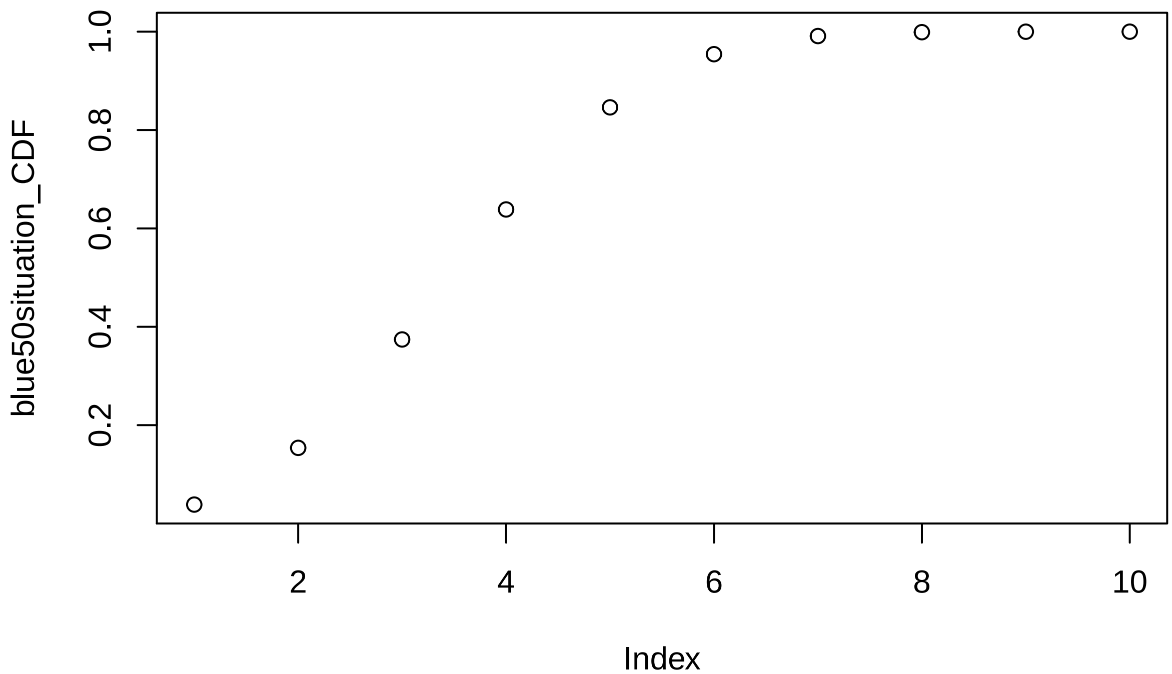



```r
# plot CDF with 10 different parameters
ggplot()+
  geom_line(aes(y = value, x = number, colour = variable),
            data = cdf2, stat="identity") +
  theme(legend.title = element_blank(),
        legend.position="top")
```



e. Compute a cumulative probability for some appropriate value and describe in words what it means. Be prepared to explain your answer to the class.

```
# when there are 50 blue colonies among 100 colonies, we will sample 10 colonies.
blue50situation_PDF <- pdf[,5]
blue50situation_CDF <- cdf[,5]

plot(blue50situation_CDF)
```




```
# according to the plot of CDF, we can obtain at least 6 colonies by 95.42% when sampling 10 colonies.
```

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.

When the number of colonies is gargantuan, we can consider binominal distribution. This is because whether replacing or not doesn't affect outcomes after a certain of trials.

Negative Binomial

Ornob and Savita

Function Summary

$$X \sim NB(r,p), \quad with \;\; E(X) = \frac{r}{p} \; and \; Var(X) = \frac{r(1-p)}{p^2}$$

where:

- trials are independent and only two outcomes are possible
- the random variable X_r represents the number of *unsuccessful trials*
- r is the number of successful trials
- x is the observed number of *failures* preceding r successes in n independent trials (note that $x + r = n$)
- the probability of success p in each trial is constant

$$PDF: \quad f(x) = P(X_r = x) = \binom{x+r-1}{r-1} p^r (1-p)^x = \binom{x+r-1}{x} p^r (1-p)^x$$
$$= \binom{n-1}{r-1} p^r (1-p)^x = \binom{n-1}{x} p^r (1-p)^x$$

Since the last (nth) Bernoulli trial is the rth success, the binomial coefficient gives the number of ways to obtain x failures in the preceding $x + r - 1 = n - 1$ trials, which is equivalent to the number of ways to obtain $r - 1$ successes preceding the rth success. The binomial coefficients above are equivalent, since

$$\binom{a}{b} = \binom{a}{a-b} \;\; for \;\; 0 \le b \le a.$$

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

The negative binomial distribution:

$$P(X = x) = \binom{x + r - 1}{r - 1} p^r (1 - p)^x$$

It is discrete. It's often applied to situations where you have binary outcomes, and you want to stop after getting a certain number of one of those outcomes.

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

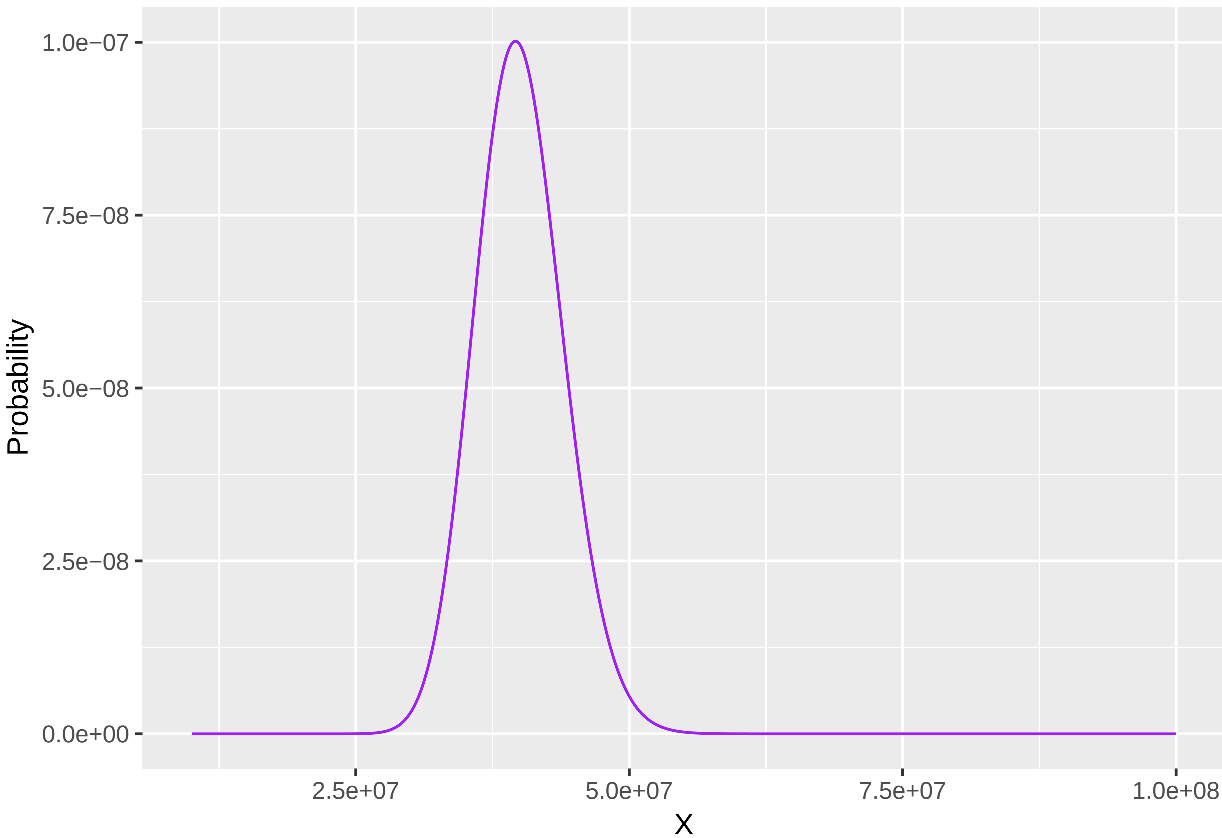
What is the probability of getting a certain number of transformants after transforming and plating a certain number of cells?

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

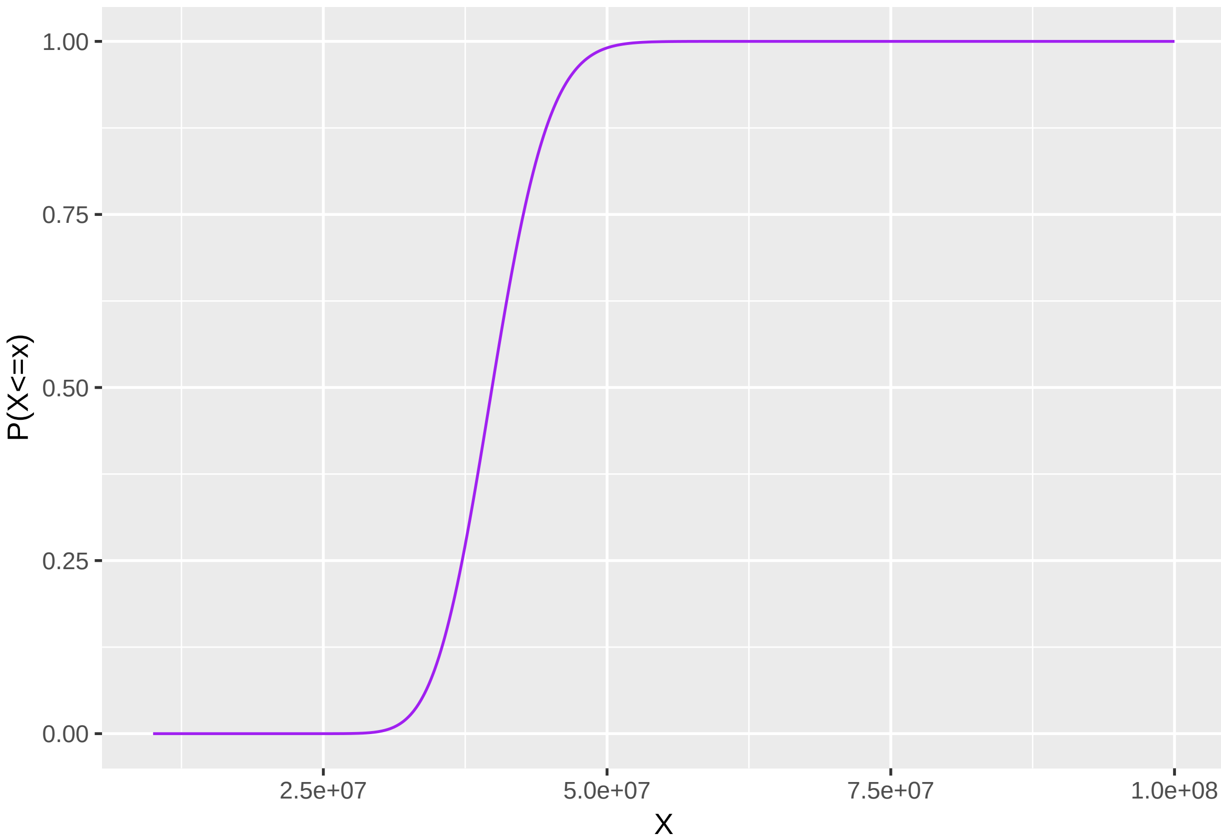
x is the number of failures. r is the number of successes. p is the probability of success. The equation multiplies the probabilities of all successes by the probabilities of all failures.

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

```r
require(ggplot2)

r=100
p=0.0000025
nmax=10^8
nmin=10^7
xmax=nmax-r
xmin=nmin-r
xrange=seq(xmin,xmax,by=100000)

pmf <- dnbinom(xrange,r,p)
pmf.data <- as.data.frame(pmf)
ggplot(pmf.data,aes(x=xrange,y=pmf))+
  geom_line(color="purple")+
  xlab("X")+
  ylab("Probability")
```




```
cdf <- pnbinom(xrange,r,p)
cdf.data <- as.data.frame(cdf)
ggplot(cdf.data,aes(x=xrange,y=cdf))+
  geom_line(color="purple")+
  xlab("X")+
  ylab("P(X<=x)")
```




```
#There is another useful way to use this function. Like what is the #probability of getting 100 or less
```

e. Compute a cumulative probability for some appropriate value and describe in words what it means. Be prepared to explain your answer to the class.

```
cdf2 <- pnbinom(5*10^7,r,p)
cdf2
```

```
## [1] 0.9906215
```

It means that in the range from 0 to the specified number of cells, the total probability of getting 100 transformants is very close to 1.

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.

The geometric distribution is identical to the negative binomial distribution when r=1.

Poisson

Avrami and Marcel

Function Summary

$$X \sim POI(\lambda), \quad E(x) = Var(X) = \lambda$$

$$PDF: \quad f(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

where X is a Poisson random variable and λ is the rate at which events occur in a given interval of time (space), i.e. for $t = 1$.

The Poisson makes the following assumptions:

- The number of observed events $x \in \{0, 1, 2, ...\}$ is independent in any interval
- Events occur at a constant (random) rate over a set of intervals of time (or space)
- The rate of events is $\lambda > 0$, the *rate constant* per unit interval
- The probability of observing two or more events in the same interval will approach 0 as intervals become smaller

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 the prpbability, p, of geting a number, x, of blue cells per plate. A blue cells means the plasmid harboring the gene for beta galactosidase was transformed, and is considered a success. IOur distirbution is discrete, and there are only two possibilites, while cells or blue cells (0 or 1).

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

Knowing the average number of blue cells per plate, what is the PMF of blue cells per plate within the range of 0:100. Cells that are blue have been succesfully transformed.

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

Lamba, which is a rate of events/time or events/space. We chose events/space because blue white screening works better with space. In other words, you either have blue or white cells (events) on the plate (space). This is the only parameter we need. The possion distirubution will give us the probability of each number of blue cells within on our range based on the lambda parameter.

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

```
lambda <- 40 #blue cells per plate
range.vector <- 0:100

# We chose our rate to be 40 blue cells/plate because based on our lab experience, that is usually how
```

e. Compute a cumulative probability for some appropriate value and describe in words what it means. Be prepared to explain your answer to the class.

```
blue.cells.per.plate <- 40
range.vector <- 0:100

par(mfrow = c(1,3))

PDF.Poisson <- dpois(range.vector, lambda = blue.cells.per.plate)
plot(PDF.Poisson, type = "l", main = "Marcel's Swag", xlab = "# Blue cells/plate", ylab = "PDF")
```

```r
CDF.Poisson <- ppois(range.vector, lambda = blue.cells.per.plate)
plot(CDF.Poisson, type = "l", main = "Avrami's Swag", xlab = "# Blue cells/plate", ylab = "CDF")

PDF.Exp <- dexp(range.vector, rate = blue.cells.per.plate)
plot(PDF.Exp, type = "l")
```

Between 0 blue cells and 100 blue cells, we have the highest probability of getting 40 blue cells. That probability is still low, lying at 0.06.

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.

One other distribution that relates for Poisson is exponential, as it can also account for series of events that account for space & time. Exponential can be used for approximations or survival relationships.

Normal

Zarina and Conor

Function Summary

$$X \sim N(\mu, \sigma^2), \quad where \quad E(X) = \mu \quad and \quad V(X) = \sigma^2$$

$$PDF: \quad f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$CDF: \quad F(x) = \int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} dx = 1$$

where

1. Outcomes x are continuous and independent.
2. $x \in \mathbb{R}$
3. $\mu \in \mathbb{R}$
4. $\sigma > 0$

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

It is the normal distribution, which applies to continuous data. It can describe and model distributions of quantitative traits, such as height and weight.

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

Depending on the parameters of the normal distribution, you can figure out the likelest number of successes (number of cells transformed if the transformation had a 25% success rate).

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

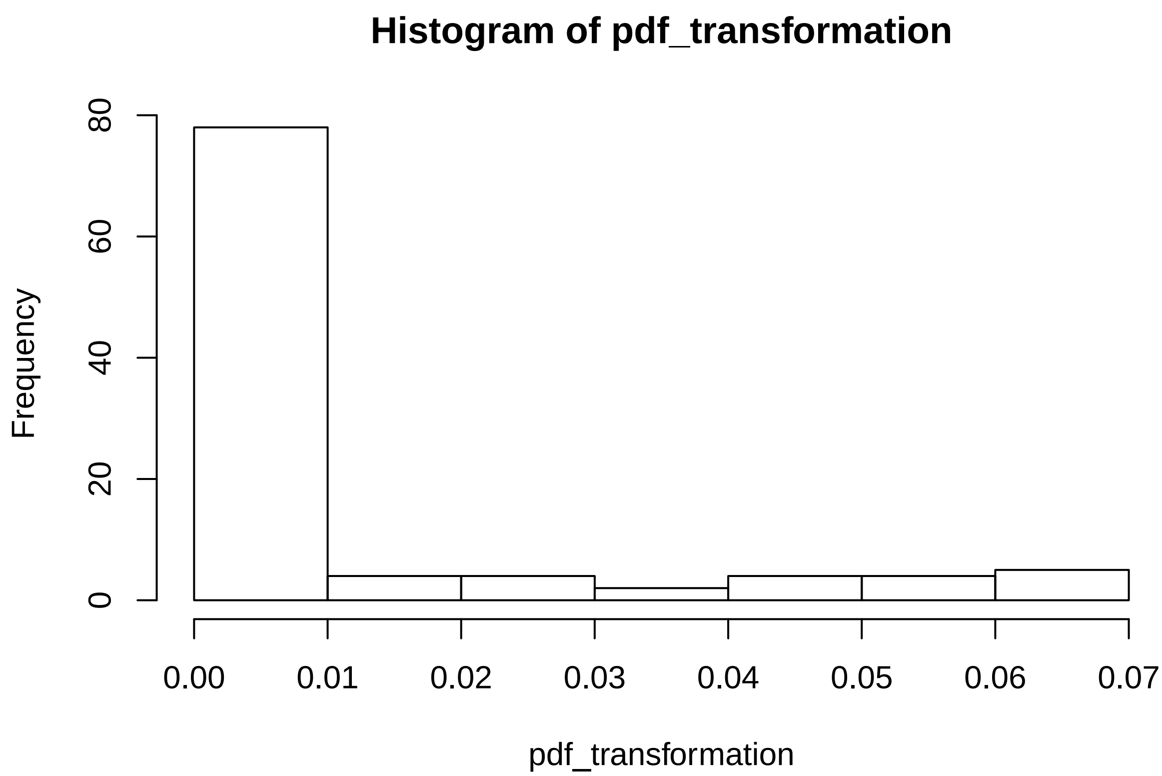
- x is the number of outcomes (which are continuous, independent, and real).
- μ the population average
- σ is the standard deviation, which is the square root of the variance.

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

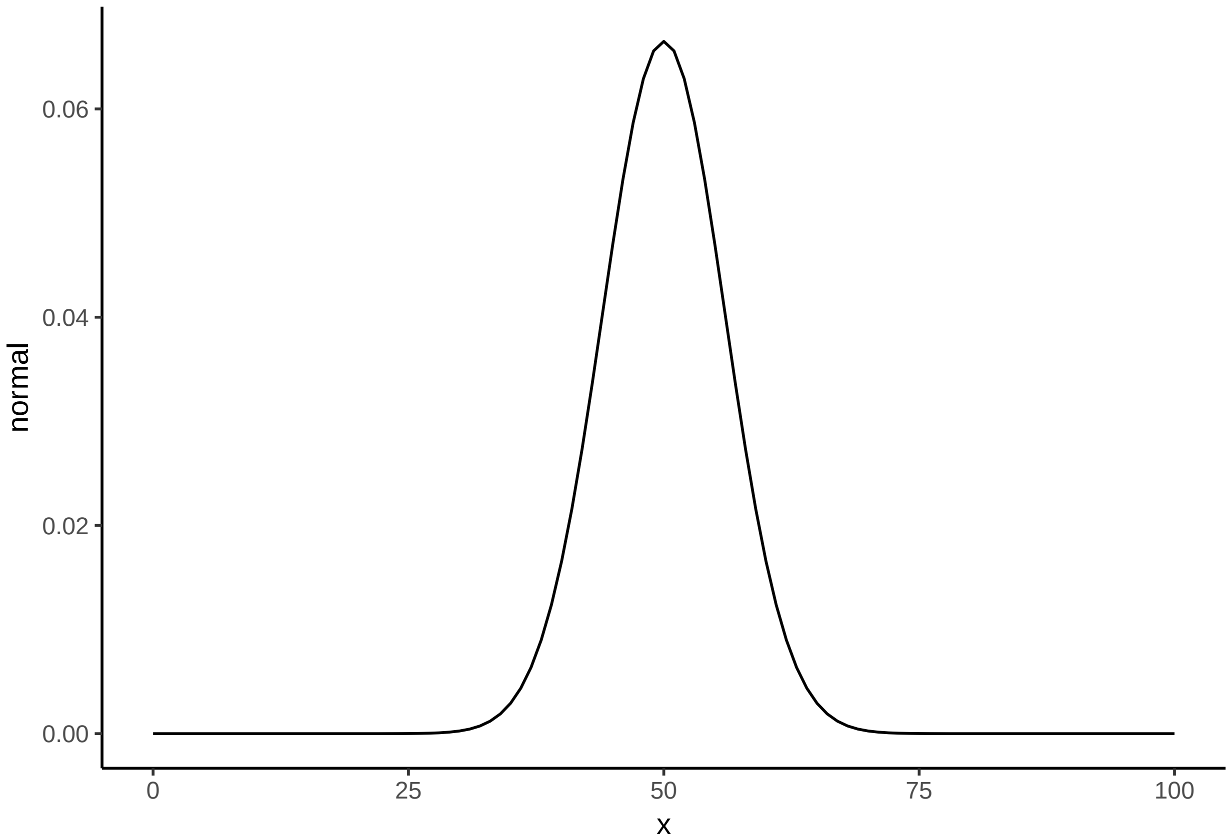
```r
pdf_transformation <- dnorm(x = 0:100, mean = 50, sd = 6)
head(pdf_transformation)
```

```
## [1] 5.534639e-17 2.188992e-16 8.420452e-16 3.150379e-15 1.146375e-14
## [6] 4.057201e-14
```

```r
hist(pdf_transformation)
```

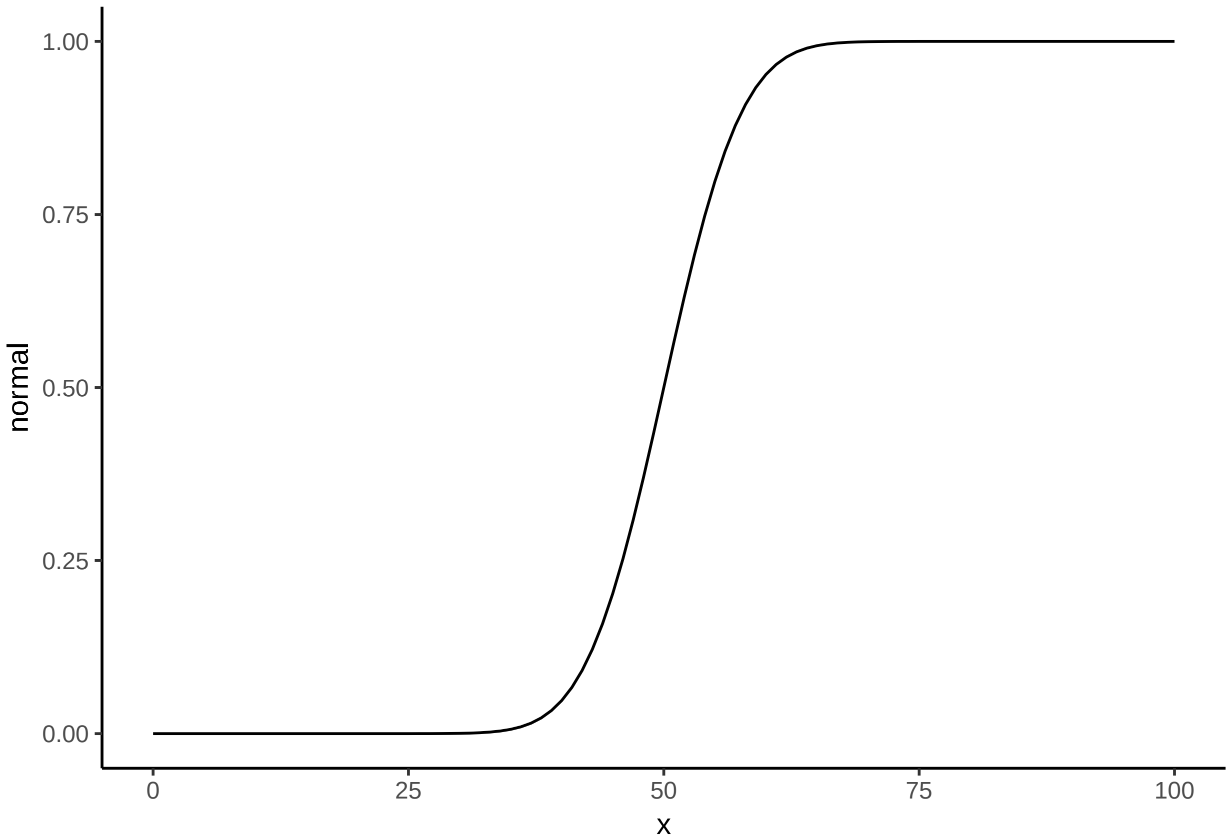



```
pdf_df <- data.frame(normal = pdf_transformation,
                     X = 0:100)

ggplot(pdf_df,
       aes_string(x= 0:100, y="normal")) +
  theme_classic() +
  geom_line()
```



e. Compute a cumulative probability for some appropriate value and describe in words what it means. Be prepared to explain your answer to the class.

```
cdf_transformation <- pnorm(q = 0:100, mean = 50, sd = 6)

cdf_df <- data.frame(normal = cdf_transformation,
                     X = 0:100)

ggplot(cdf_df,
       aes_string(x= 0:100, y="normal")) +
  theme_classic() +
  geom_line()
```



The x axis represents the number of cells successfully transformed, and the y axis is the cumulative probability. For instance, if you are looking for the x number of successfully transformed cells, the cumulative probability is the probability the 0 to x cells are transformed.If you look at x= 50, there is a 50% chance that 50 or fewer cells were transformed.

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.

Although binomial distribution is discrete, given a large enough n, the distribution of data approximates a normal distribution.

Exponential

Yingzhen and Veronica

Function Summary

$$T \sim EXP(\lambda), \quad where \quad E(T) = SD(T) = 1/\lambda, \quad Var(T) = (1/\lambda)^2$$

for $\lambda > 0$ and $t \geq 0$.

$$PDF: \quad f(t) = \lambda e^{-t\lambda}$$

The CDF has a closed form that does not require integration:

$$CDF: \quad F(t) = \int_0^\infty \lambda e^{-\lambda t} dt = 1 - e^{-t\lambda}$$

Yinghzen version

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

PDF of the exponential distribution is:

$$f(x) = \lambda e^{-\lambda x}$$

Problem: time or space interval between two poisson events. Given λ the probability and x the time. It is continuous.

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

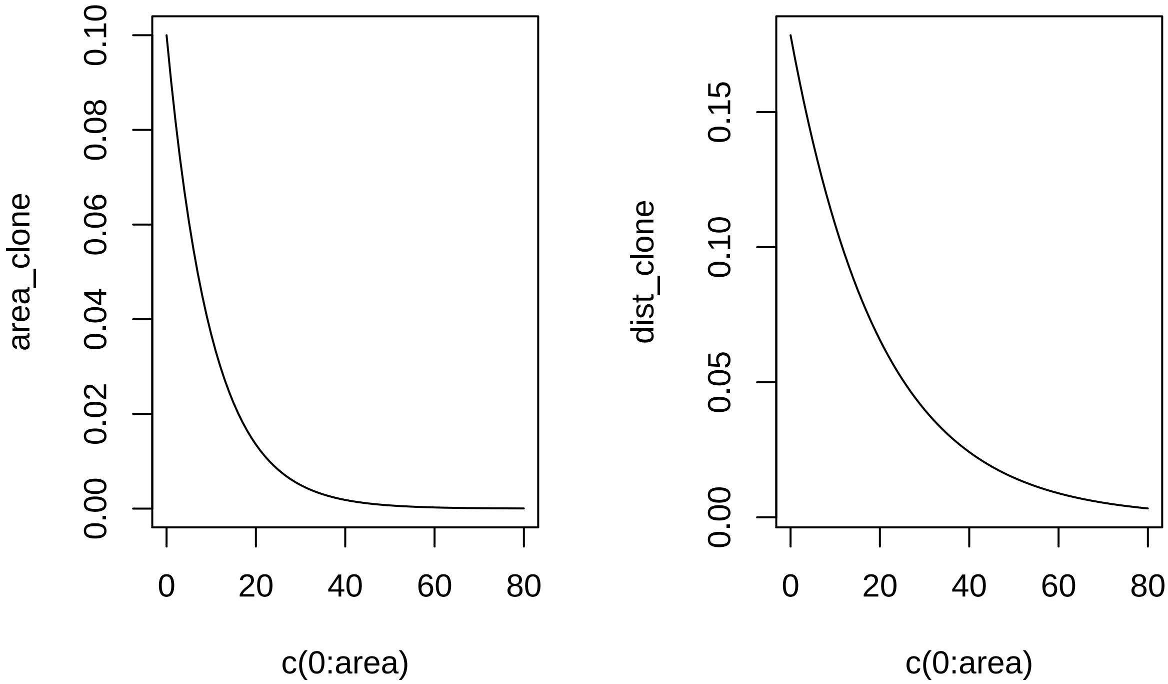
Distance or of closest pair of clones.

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

λ is the rate of finding another transformed clone within an area of one square centermeter of one clone. For example, $\lambda = 0.1$ means you can find 0.1 transformed clones within one square centimeter from any clone on average. x the area you need to find another clone.

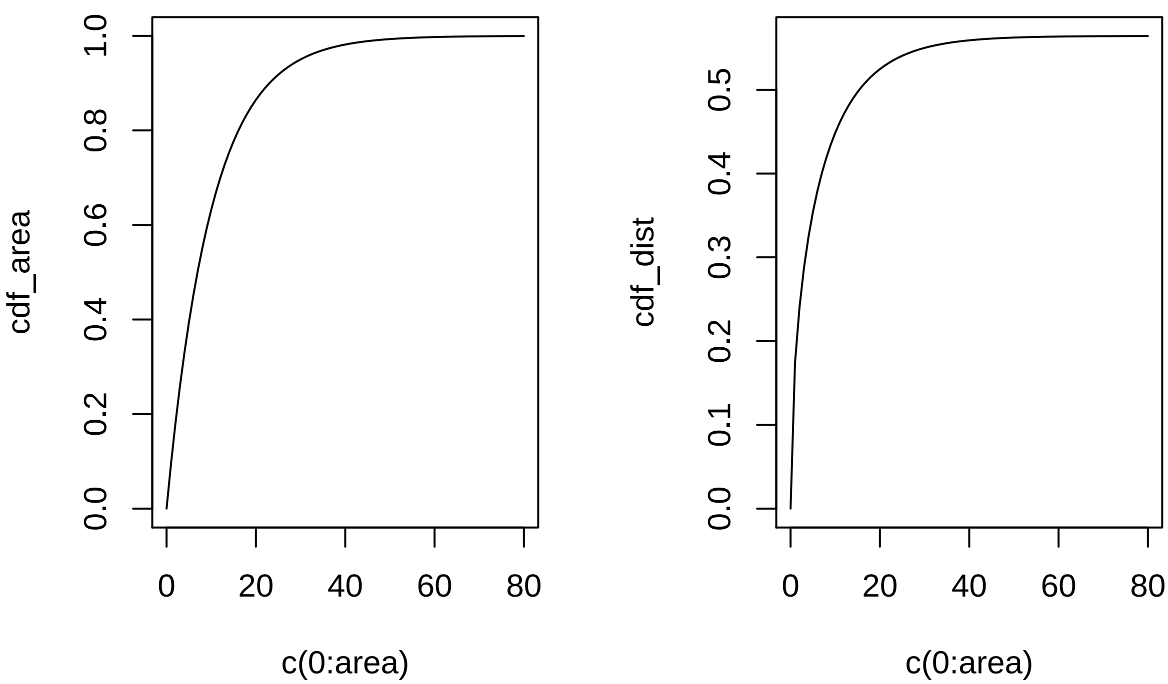
d. Pick some appropriate parameters for your distribution as applied to the transformation scenario and graph the PDF and CDF.

```
# given a circular dish, with the total number of clones we can know the distribution of transformed cl
area = 80
rate = 0.1  # rate is lambda
area_clone = dexp(c(0:area), rate)
dist_clone = (area_clone / 3.14) ** 0.5

par(mfrow = c(1,2))
plot(c(0:area), area_clone, type = "l")
plot(c(0:area), dist_clone, type = "l")
```



e. Compute a cumulative probability for some appropriate value and describe in words what it means. Be prepared to explain your answer to the class.

```
par(mfrow = c(1,2))
cdf_area = pexp(c(0:area), rate)
cdf_dist = (cdf_area / 3.14) ** 0.5
plot(c(0:area), cdf_area, type = "l")
plot(c(0:area), cdf_dist, type = "l")
```



Within certain area or distance, the chance to find next transformed clone. This is depend mainly on the total number of clones per dish, or concentration.

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.

Poisson distribution. Poisson describes the number of clones per unit area, while exponential describes the area to find next clone.

Veronica version

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

$$f(t) = \lambda e^{(-\lambda t)}, \quad where \ \ \lambda > 0 \ \ and \ \ t > 0$$

continuous describes the amount of time elapsed between events in a poisson point process memoryless

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

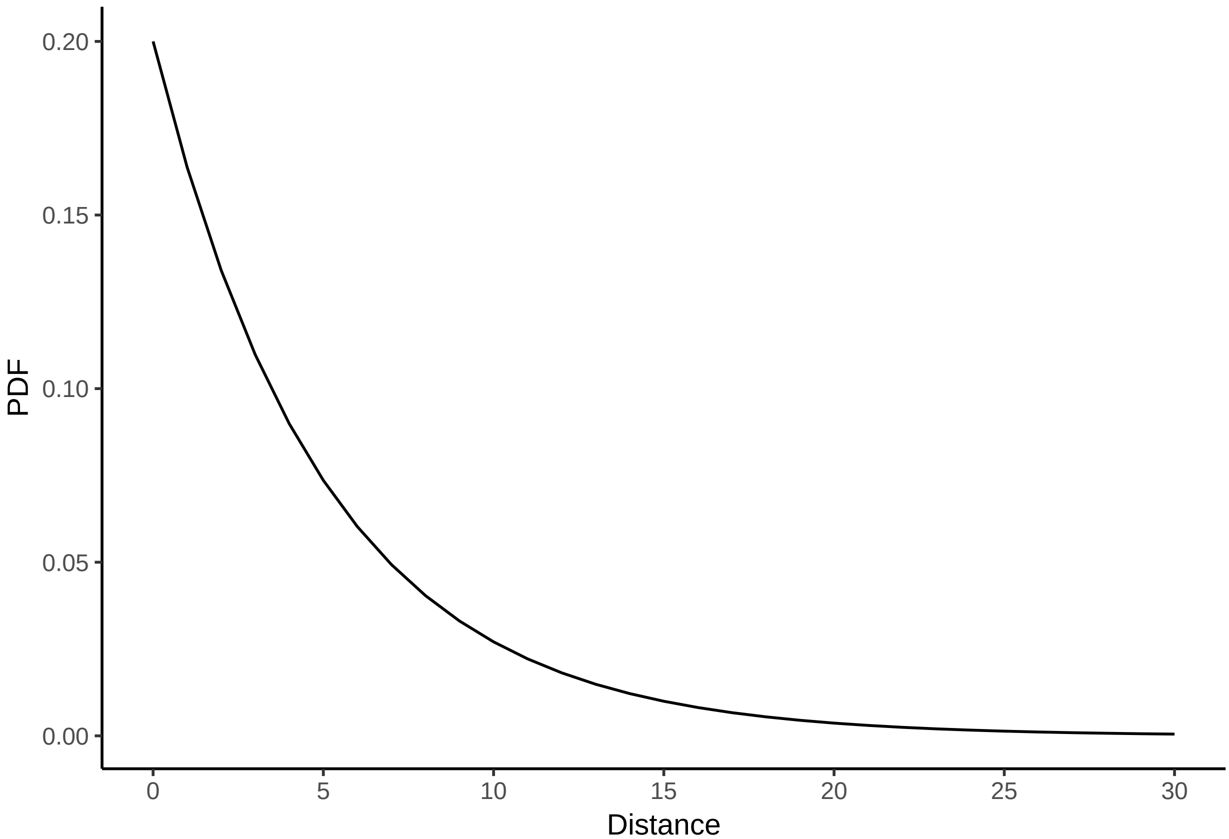
distance between transformed colonies to their closest transformed neighbor

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

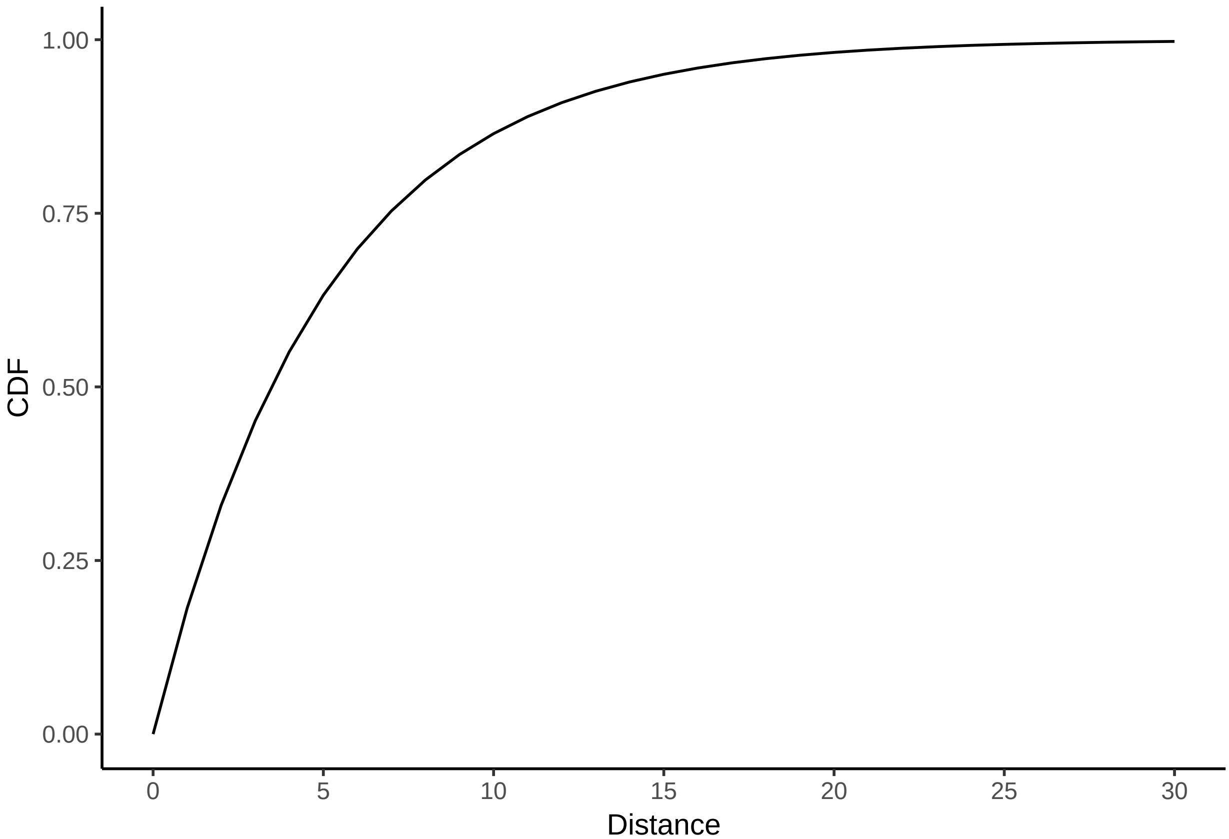
Takes lambda and t Lambda is the rate or density of transformed colonies, t is time or distance between colonies

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

```r
library(ggplot2)
library(scales) # need this for pretty_breaks()

lambda=.2
colony.distance.pdf=dexp(0:30, lambda)
colony.distance.cdf=pexp(0:30, lambda)
colony.distance = data.frame(PDF = colony.distance.pdf, CDF=colony.distance.cdf, Distance=0:30)
ggplot( colony.distance ,aes( x = Distance ,y = PDF )) +
  theme_classic() +
  geom_line() +
  scale_x_continuous(breaks=pretty_breaks())
```




```r
ggplot( colony.distance ,aes( x = Distance ,y = CDF )) +
  theme_classic() +
  geom_line() +
  scale_x_continuous(breaks=pretty_breaks())
```



e. Compute a cumulative probability for some appropriate value and describe in words what it means. Be prepared to explain your answer to the class.

```
colony.distance.cdf
```

```
##  [1] 0.0000000 0.1812692 0.3296800 0.4511884 0.5506710 0.6321206 0.6988058
##  [8] 0.7534030 0.7981035 0.8347011 0.8646647 0.8891968 0.9092820 0.9257264
## [15] 0.9391899 0.9502129 0.9592378 0.9666267 0.9726763 0.9776292 0.9816844
## [22] 0.9850044 0.9877227 0.9899482 0.9917703 0.9932621 0.9944834 0.9954834
## [29] 0.9963021 0.9969724 0.9975212
```

Given a density of one transformed colony per unit distance, here is a 100% chance of finding a transformed neighbor within 17 units of distance.

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.

continuous version of geometric distribution