Consider the density function $f(x) = \begin{cases} 2a(-x^3 + x^2 + x + 2), & 0 < x < 2 \\ 0, & \text{else} \end{cases}$

a) First, determine $a$ to make sure $f(x)$ is a density function.

b) Compute the prob. that $x$ will be between 0 and 1.

c) Use R to plot $f(x)$. Include code and figure.

\[
a) \quad \int_0^\infty f(x) \, dx = a \int_0^2 (-x^3 + x^2 + x + 2) \, dx = a \left[ -\frac{1}{4} x^4 + \frac{1}{3} x^3 + \frac{1}{2} x^2 + 2x \right]_0^2 \\
= a \left[ -\frac{1}{4} (16) + \frac{1}{3} 8 + \frac{1}{2} 4 + 4 \right] = a \left( 4 + \frac{8}{3} + 2 + 4 \right) = \frac{14}{3} \Rightarrow a = \frac{3}{14}
\]

\[
\Rightarrow f(x) = \frac{3}{14}(-x^3 + x^2 + x + 2) \quad 0 < x < 2
\]

\[
b) \quad \text{prob } (0 < x < 1) = \int_0^1 f(x) \, dx = \int_0^1 \frac{3}{14} (-x^3 + x^2 + x + 2) \, dx
\]

\[
= \frac{3}{14} \left[ -\frac{1}{4} x^4 + \frac{1}{3} x^3 + \frac{1}{2} x^2 + 2x \right]_0^1 = \frac{3}{14} \left( -\frac{1}{4} + \frac{1}{3} + \frac{1}{2} + 2 \right) = \frac{3}{14} \cdot \frac{3}{12}
\]

\[
\text{good enough.}
\]

x = seq(0,2,by=0.01)
y = (3/14)*(-x^3 + x^2 + x + 2)
png("t.png")
plot(x,y)
dev.off()
The Bernoulli distribution discussed in the lecture does have a formula:

\[ p(x) = \pi^x (1-\pi)^{1-x}, \]

where \( \pi \) is some parameter between 0,1, and \( x = 0,1 \).

Show that it's a mass function.

\[ p(x) > 0 \text{ for } x = 0,1. \]

\[
\sum_{x=0}^{1} p(x) = p(x=0) + p(x=1) = \pi^0 (1-\pi)^0 + \pi^1 (1-\pi)^1 = 1 - \pi + \pi = 1
\]

Also recall from the chicken question:

What proportion of time do we expect to get \( x = 1 \)?

\[ \text{prop (} x=1 \text{)} = p(x=1) = \pi^1 (1-\pi)^0 = \pi. \]

This gives the parameter \( \pi \) a nice interpretation:

It's the probability of times we get \( x = 1 \).

If \( x = 0,1 \) represent heads and tails, then \( \pi \) is the probability of getting a heads.
Show That

a) \[
\int_0^\infty e^{-2x} \, dx = 1
\]
\[y = 2x \rightarrow \, dy = 2 \, dx\]
\[
\text{LHS} = \int_0^\infty e^{-2x} \, dx = \int_0^\infty e^{-y} \, dy = -e^{-y} \bigg|_0^\infty = -(0-1) = 1.
\]

b) \[
\sum_{x=0}^{\infty} \frac{e^{-a} a^x}{x!} = 1
\]
[Hint: use the Taylor series expansion for \(e^x\)]
\[
\text{LHS} = \sum_{x=0}^{\infty} \frac{e^{-a} a^x}{x!} = e^{-a} \sum_{x=0}^{\infty} \frac{a^x}{x!} = \frac{e^{a}}{x!} = 1.
\]
Taylor series exp.
\[
\exp(a) = \sum_{x=0}^{\infty} \frac{a^x}{x!} = \left(1 + a + \frac{1}{2!} a^2 + \cdots\right)
\]

\[
c) \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi \sigma^2}} \, e^{-\frac{1}{2} \left(\frac{x-M}{\sigma}\right)^2} \, dx = 1 \quad \text{[use} \int_{-\infty}^{\infty} e^{-\frac{1}{2} x^2} \, dx = \sqrt{2\pi} \text{]}\]
\[
\text{LHS} = \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left(\frac{x-M}{\sigma}\right)^2} \, dx = \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \, z^2} \, dz
\]
\[
z = \frac{x-M}{\sigma} \Rightarrow \, dz = \frac{1}{\sigma} \, dx
\]
\[
= \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} z^2} \, dz = 1.
\]

[It's harder to show \(\sum x=1\) for binomial. we'll do it later]
Suppose the density function for $x$ is given by the normal distribution with parameters $\mu$, $\sigma$.

I.e.,
$$f(x) = \frac{1}{\sqrt{2\pi \sigma^2}} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2}$$

$$f(z)$$

(a) Compute the density function for $z = \frac{x-\mu}{\sigma}$.

Hint: Start with $\int_{-\infty}^{\infty} f(x) \, dx = 1$ and derive $\int_{-\infty}^{\infty} f(z) \, dz = 1$.

$$\int_{-\infty}^{\infty} f(x) \, dx = 1 \implies \frac{1}{\sqrt{2\pi \sigma^2}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2} \, dx = 1$$

$$z = \frac{x-\mu}{\sigma} \implies dz = \frac{1}{\sigma} \, dx$$

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2} z^2} \, dz = 1$$

$$\therefore \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2} \, dz = 1 \implies f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2}$$

(b) In the place where $\mu$ and $\sigma$ appear in $f(x)$, what values find in $f(z)$?

$$f(z) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left(\frac{z-\mu}{\sigma}\right)^2} \implies \mu = 0, \sigma = 1.$$
**hw_lect6-2**

What's the 10th percentile of the uniform dist. between -1,1?

Hint: for uniform dist., integration is trivial.

\[ f(x) = \frac{1}{2} \Rightarrow x = \frac{1}{2} \]

\[ \text{10th percentile} = -(1-x) = - \frac{4}{5} \]

**Note:** This number has the units of \( x \).

It can also be negative.

---

**hw_lect6-3** Consider the power-law distribution

\[ f(x) = a x^{a-1}, \quad 0 < x < 1, \quad a > 0 \]

Find the \( n \)th percentile.

**Just FYI:** note that \( f(x) \) is a distribution:

\[
\begin{align*}
& f(x) > 0 \\
& \int_{-\infty}^{\infty} f(x) \, dx = \int_{0}^{1} a x^{a-1} \, dx = a \left. \frac{x^a}{a} \right|_0^1 = 1 - 0 = 1
\end{align*}
\]

\[ \int_{0}^{m} a x^{a-1} \, dx = \frac{n}{100} \Rightarrow a \left. \frac{x^a}{a} \right|_0^m = m^a = \frac{n}{100} \Rightarrow m = \left( \frac{n}{100} \right)^{1/a} \]
Consider one of the two continuous variables, and one of the two discrete variables, in hw.lect 1. Make comparative boxplots for the continuous variable for each level of the discrete variable. E.g. if the discrete var. has 4 levels, then you need to show 4 boxplots for the cont. var. all on the same plot, side-by-side. Interpret.

The answers will vary across students, but the code will be something like this:

```r
dat = read.table(...)  
X1 = dat[,1]  # 1st categorical variable 
X2 = dat[,2]  # 2nd " " " 
X3 = dat[,3]  # 1st continuous var. 
X4 = dat[,4]  # 2nd " " " 

boxplot(X3[X1 == A], X3[X1 == B], ..., X3[X1 == Z])
# where A, B, ..., Z are the levels of X1.
```
By R (see prelab)

One piece of info that boxplots don’t convey is sample size. Let’s “run some simulations” to explore that issue. To that end:

a) take a sample of size 20 from a normal with \( \mu = 0, \sigma = 10 \).
b) \( \ldots \) 30

c) \( \ldots \) 40

d) \( \ldots \) 50

e) \( \ldots \) 100

f) Make a comparative boxplot of the 5 samples a–e.

set.seed(123)  # This makes sure we all get the answer below, but yours will be similar anyway.
x1 = rnorm(20,0,10)
x2 = rnorm(30,0,10)
x3 = rnorm(40,0,10)
x4 = rnorm(50,0,10)
x5 = rnorm(100,0,10)
boxplot(x1,x2,x3,x4,x5)

Note that the boxplots generally don’t change as the sample size changes. This is a true (although counter-intuitive) result, quite generally.
a) Use the binomial mass function to show that the probability of getting "at least 1 head out of n tosses" is $1 - (1 - \pi)^n$, where $\pi$ is the probability of getting a head on a single toss. 

Show work! \[ x \geq 1 \text{ where } x = \# \text{ of heads out of } n. \]

\[
\therefore \ \text{prob}(x \geq 1) = 1 - \text{prob}(x = 0)
\]

\[
= 1 - p(x = 0) = 1 - \frac{n!}{0!(n-0)!} \pi^0 (1-\pi)^{n-0}
\]

\[
= 1 - (1-\pi)^n
\]

b) What is the numerical value of that probability as $n \to \infty$? Think about the answer you get; it’s interesting and counterintuitive.

As $n \to \infty$, $(1-\pi)^n \to 0$ because $(1-\pi) < 1$.

So, $\text{prob}(x \geq 1) \to 1$.

I.e. it’s nearly certain that you will get at least 1 head. This is counterintuitive, because the probability of getting 1 head is still $\pi$, regardless of $n$, and that $\pi$ could be any value between 0 and 1. But the probability of at least 1 head approaches 1.