Consider the Bernoulli dist. with parameter $\pi$:

$$
 p(x) = \pi^x (1-\pi)^{1-x}, \quad x = 0, 1 \quad 0 < \pi < 1
$$

a) Show that it's a distribution (prob. mass function).

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

b) Find the prob that $x=1$.

\[p(y=x=1) = p(x=1) = \pi \checkmark
\]

This gives the param $\pi$ a nice interpretation. It's the prob of times we got $x=1$.

If $x=0, 1$ represent heads and tails, then $\pi$ is the prob of getting a heads.

hw-lect4-2:

Based on data, we have observed that $x$ is between 0 and 1/2 about 25% of the time. Which of the following is the more reasonable distribution from which our data may have come? Show work (always!)

A) $f(x) = 2x \quad 0 < x < 1$

B) $f(x) = e^{-x} \quad 0 < x < \infty$

A) $f(x) = 2x \quad 0 < x < 1$

B) $f(x) = e^{-x} \quad 0 < x < \infty$

A) $\int_0^{0.5} 2x \, dx = 2 \left[ \frac{1}{2} x^2 \right]_0^{0.5} = \frac{1}{4} = \frac{1}{4} = \frac{\sqrt{2}}{4} \quad \leftarrow \text{It's more likely our data have come from this distr./pop.}$

B) $\int_0^{0.5} e^{-x} \, dx = -e^{-x} \bigg|_0^{0.5} = 1 - e^{-0.5} \approx 0.40$
The number of boys in a sample of size 10 has all the characteristics of being a Binomial r.v. with n = 10.

Since the prop. of boys in the pop is 0.5, then the prop of drawing a boy is 0.5, and that's the p parameter of Binomial.

\[ P(X = 1) = \frac{10!}{9!1!} \cdot (0.5)^1 \cdot (1 - 0.5)^9 = \frac{10 \cdot 0.5}{2} \cdot \left(\frac{1}{2}\right)^{10} = 10 \left(\frac{1}{2}\right)^{10} \]

\[ P(X = 2) = \frac{10!}{2!8!} \cdot (0.5)^2 \cdot (1 - 0.5)^8 = \frac{10 \cdot 0.5^2}{2 \cdot 8!} \cdot \left(\frac{1}{2}\right)^{10} = 45 \left(\frac{1}{2}\right)^{10} \]

\[ P(X = 1 \text{ or } X = 2) = (10 + 45) \left(\frac{1}{2}\right)^{10} = 55 \left(\frac{1}{2}\right)^{10} \]
Let $x$ have a normal dist. with params $\mu, \sigma$, i.e. $x \sim \mathcal{N}(\mu, \sigma)$.

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

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

**Hint:** Start with $\int f(x) \, dx = 1$, do not do the integral, but instead, do a change of variables until you get $\int [\ldots] \, dz = 1$. Then $[\ldots] \overset{?}{=} f(z)$.

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

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

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

$$\therefore \int \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2} \, dz = 1 \quad \Rightarrow \quad f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} z^2}$$

b) In $f(z)$, in the place where you would expect to find $\mu$ and $\sigma$, what numbers do you see?

$$f(z) = \frac{1}{\sqrt{2\pi(1)}} e^{-\frac{1}{2} \left(\frac{z-0}{1}\right)^2} \quad \Rightarrow \quad \mu = 0, \quad \sigma = 1$$

**Moral 1:** if $x \sim \mathcal{N}(\mu, \sigma)$, then $z = \frac{x-\mu}{\sigma} \sim \mathcal{N}(0, 1)$.

**Moral 2:** if $x \sim$ any distr., then this method can be used to find the distr. of any function of $x$ (e.g. $\frac{x-\mu}{\sigma}$).
Show That

a) \( \int_0^\infty e^{-x} \, dx = 1 \)

LHS: \( \int_0^\infty e^{-x} \, dx = \int_0^\infty e^{-y} \, dy = -e^{-y}\bigg|_0^\infty = -(0-1) = 1 \).

b) \( \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = 1 \) [Hint: use the Taylor series expansion for \( e^{\frac{\lambda}{\lambda}} \)]

LHS: \( \sum_{x=0}^{\infty} \frac{e^{-\lambda} \lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} \xrightarrow{x!} e^\lambda = 1 \).

Taylor series exp. \( e^\lambda = \sum_{x=0}^{\infty} \frac{\lambda^x}{x!} = (1 + \lambda + \frac{1}{2!}\lambda^2 + \ldots) \)

c) \( \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-M)^2} \, dx = 1 \) [Use \( \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2} \, dx = \sqrt{2\pi} \)]

LHS: \( \frac{1}{\sqrt{2\pi}\sigma^2} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-M)^2} \, dx = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}z^2} \, dz \)

\( z = \frac{x-M}{\sigma} \Rightarrow \, dz = \frac{1}{\sigma} \, dx \)

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

The \( \text{p.c.} = 1 \) for binomial will be proved, later, (and in a different way)
Find the prob. of $x \leq 2$ if

a) $X \sim \text{Bernoulli} (p = \text{arbitrary})$

b) $X \sim \text{exp} (\lambda = 3)$

c) $X \sim \text{poiss} (\lambda = 3)$

d) $X \sim \text{Binom} (n = 10, p = \frac{1}{4})$

e) $X \sim \mathcal{N}(0, 1)$

\[ a) \sum_{x=0}^{2} p(x) = \sum_{x=0}^{2} p(x) (1-p)^{1-x} = p^0 (1-p)^1 + p (1-p)^0 = 1 \]

\[ b) \int_{0}^{2} 3e^{-3x} \, dx = \frac{3e^{-3x}}{-3} \bigg|^{2}_{0} = -1(e^{-6} - 1) = 1 - e^{-6} \]

c) $\sum_{x=0}^{2} e^{-3x} \frac{x^n}{x!} = e^{-3} \left( \frac{3^0}{0!} + \frac{3^1}{1!} + \frac{3^2}{2!} \right) = e^{-3} \left( 1 + 3 + \frac{9}{2} \right) = \frac{17}{2} e^{-3}$

\[ d) \sum_{x=0}^{10} \frac{10!}{x! (10-x)!} \left( \frac{1}{4} \right)^x \left( \frac{3}{4} \right)^{10-x} = \frac{10!}{0! 10!} \left( \frac{3}{4} \right)^{10} + \frac{10!}{1! 9!} \left( \frac{3}{4} \right)^9 + \frac{10!}{2! 8!} \left( \frac{3}{4} \right)^8 \]

\[ = \frac{1}{4^{10}} \left( \frac{3}{4} \right)^{10} + \frac{3}{4^{10}} \left( \frac{3}{4} \right)^9 + \frac{45}{4^{10}} \left( \frac{3}{4} \right)^8 \]

\[ = \frac{3^8}{4^{10}} \left( 3^2 + 10 \cdot 3 + 45 \right) = \frac{3^8 (86)}{4^{10}} \]

\[ \text{Any of these answers is acceptable.} \]

\[ e) \int_{-\infty}^{2} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} x^2} \, dx = 0.9772 \]

\[ \text{Table I.} \]
If \( x \) follows \( N(\mu, \sigma) \), what's the prob. of \( x \) being within 1 \( \sigma \) of \( \mu \)?

\[
\frac{x - \mu}{\sigma} = z = 1.00 \quad \Rightarrow \quad z = \frac{x - \mu}{\sigma} = 1
\]

Standardization is important in finding probs. Although it almost always refers to the change of variable \( z = \frac{x - \mu}{\sigma} \), taking \( N(\mu, \sigma) \) to \( N(0,1) \), sometimes a different change of variable is required to obtain something that has \( N(0,1) \) distr.

Find the prob. \( \text{Pr}(x < 2) \) if \( \text{log}(x) \) has a Std. Normal distr.

\[ \text{Pr}(x < 2) = \text{Pr}(\text{log}(x) < \text{log}(2)) = \text{Pr}(z < 0.69) = 0.7549 \]

Find the \( n^{th} \) percentile of the uniform distribution over \((a, b)\). Hint, answer will depend, on \( a \) and \( b \).

Uniform was defined in problem 1.19, 1.20.

\[
\int_{a}^{n^{th} \text{percentile}} \frac{1}{b-a} \, dx = \frac{n}{100}
\]

\[
\frac{n}{100} = \frac{1}{b-a} (n^{th} \text{percentile} - a) \quad \Rightarrow \quad n^{th} \text{percentile} = a + \frac{n(b-a)}{100}
\]

Note: This number has the units of \( x \). It can also be negative. Also note that for \( \text{Unif}(a,b) \), integration is trivial, i.e. areas can be found by multiplying height by width of rectangles.
Consider the boxplot of an exponential distribution with parameter \( \lambda \). How long is the box portion of the boxplot?

\[
\int_{0}^{x_1} \lambda e^{-\lambda x} \, dx = \frac{25}{100} \Rightarrow \lambda e^{-\lambda x_1} = \frac{1}{4} \quad \Rightarrow \quad -\lambda x_1 = \log \frac{4}{3} \quad \Rightarrow \quad x_1 = \frac{1}{\lambda} \log \frac{4}{3}
\]

\[
\int_{0}^{x_2} \lambda e^{-\lambda x} \, dx = \frac{75}{100} \Rightarrow \quad (1 - e^{-\lambda x_2}) = \frac{3}{4} \quad \Rightarrow \quad x_2 = -\frac{1}{\lambda} \log \frac{3}{4}
\]

\[
x_2 - x_1 = \frac{1}{\lambda} \log \frac{4}{3} - \frac{1}{\lambda} \log \frac{3}{4} = \frac{1}{\lambda} \log \left( \frac{4}{3} / \frac{3}{4} \right) = \frac{1}{\lambda} \log 3
\]
Consider ONE of the 2 continuous random vars, and ONE of the 2 discrete (categorical variables in the data you collected. 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-variable, all on the same plot, side-by-side. Interpret/discuss them. By R.

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

dat = read.table( "..." ).

X_1 = dat [, 1]  # 1st categorical/discrete var.
X_2 = dat [, 2]  # 2nd...
X_3 = dat [, 3]  # 3rd continuous var.
X_4 = dat [, 4]  # 4th...

boxplot( X_3[ X_1 == A ], X_3[ X_1 == B ], ..., X_3[ X_1 == Z ] )

# where A, B, ..., Z are the levels of X_1.

Interpretations include the “shape” of the boxplot (e.g. is there a skew), the width (which ones are wider, why?), and the relative position of the boxplots (is there a difference between the groups?)