INSTRUCTOR:
Xiaoping Shi

Irving K. Barber Faculty of Science
UBC, Okanagan Campus

This exercise set has an additional 8 questions, for a total of 58 points. These questions, in addition to those on other practice sets, are representative of what might be seen on Quiz 2.

READ THE QUESTIONS CAREFULLY

Notes and electronic devices are allowed, but they MUST be kept in airplane mode. You may use the back of a page if you run out of room on the front.

SURNAME, GIVEN NAME (print) ____________________

STUDENT NUMBER. ____________________

Signature: ____________________

```
set.seed(39133)  # do this if you want to replicate the results below.
```

[1] 1. (a) Simulate 1000 standard normal random variables and exponentiate to obtain 1000 lognormal random variables. Assign the result to `L1`. Repeat and assign the result to `L2`

```
n <- 1000
L1 <- exp(rnorm(n))
L2 <- exp(rnorm(n))
```

[1] (b) Assign `L1 + L2` to `X1` and `L1 - L2` to `X2`.

```
X1 <- L1 + L2
X2 <- L1 - L2
```

[2] (c) Compute the correlation of `X1` and `X2`. Are X_1 and X_2 correlated?

```
cor(X1, X2)
## [1] 0.00597
```

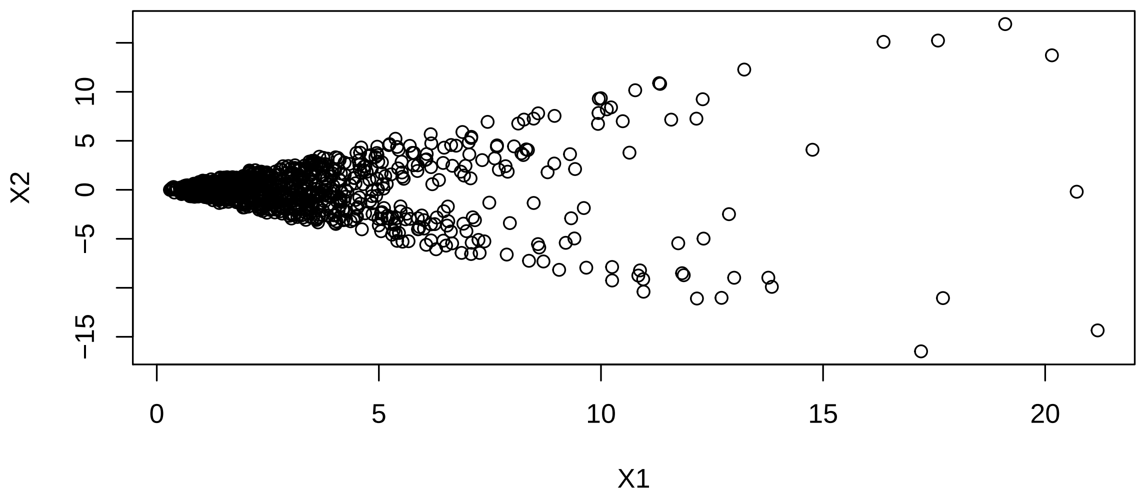
They are not correlated. The sample correlation is very small. It is also possible to show mathematically that they are not correlated: Since L_1 and L_2 have the same distribution, they have the same mean and the expected values of their squares are the same. Therefore,

$$\mathrm{Cov}(X_1, X_2) = E[X_1X_2] - E[X_1]E[X_2] = E[L_1^2 - L_2^2] - E[L_1 + L_2]E[L_1 - L_2] = E[L_1^2] - E[L_2^2] - 0 = 0.$$

The correlation is just the covariance divided by some constants, so the result follows.

[2] (d) Construct a scatterplot of `X2` and `X1` to decide if X_1 and X_2 are dependent?

```
plot(X2 ~ X1)
```



The distribution of X_2 depends on X_1 in an interesting way. They are dependent.

2. Simulate 1000 samples of size $n = 5$ from a normal distribution with mean $\mu = 3.0$ and standard deviation $\sigma = 7.0$. (Use the `matrix()` trick from the lecture notes.)

```
N <- 1000; n <- 5; mu <- 3; sigma <- 7
normalsamples <- matrix(rnorm(N*n, mean = mu, sd = sigma), nrow=n)
```

[2] (a) For all of the samples, compute sample means and standard deviations, assigning them to objects `xbar` and `std`. (Use `apply()` as in the lecture notes.)

```
xbar <- apply(normalsamples, 2, mean)
std <- apply(normalsamples, 2, sd)
```

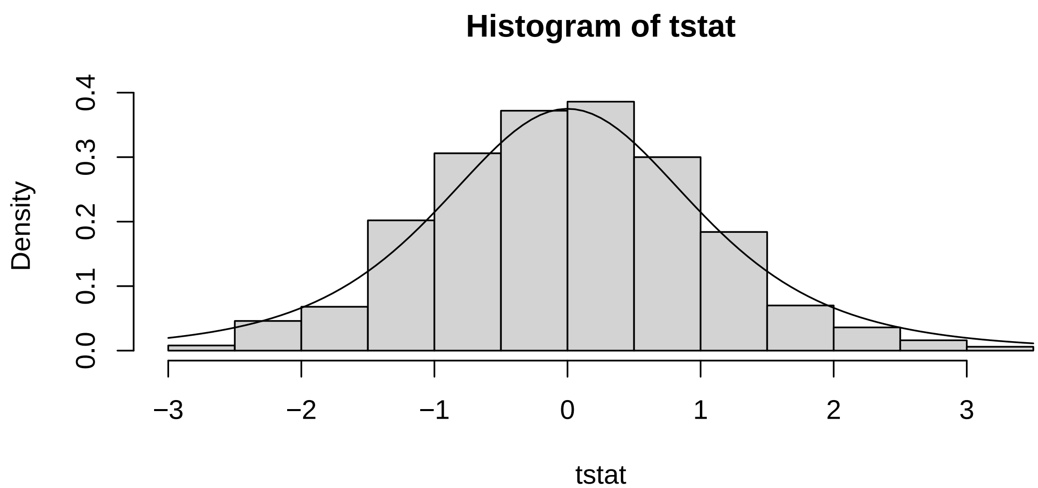
1 (b) Calculate t statistics from **xbar** and **std** using the formula:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

```
tstat <- (xbar - mu)/(sigma/sqrt(n))
```

1 (c) Plot a histogram of the t values and overlay with a curve of the t pdf on $n - 1 = 4$ degrees of freedom. Does the curve fit the histogram?

```
hist(tstat, freq = FALSE)
curve(dt(x, df = n - 1), add = TRUE)
```



The curve fits the histogram. The t statistic defined in the way can be shown - using the techniques discussed in an earlier lecture to exactly have a t distribution on $n - 1$ degrees of freedom.

3. Repeat the previous question, using samples of size $n = 8$.

```
N <- 1000; n <- 8; mu <- 3; sigma <- 7
normalsamples <- matrix(rnorm(N*n, mean = mu, sd = sigma), nrow=n)
```

2 (a) For all of the samples, compute sample means and standard deviations, assigning them to objects **xbar** and **std**.

```
xbar <- apply(normalsamples, 2, mean)
std <- apply(normalsamples, 2, sd)
```

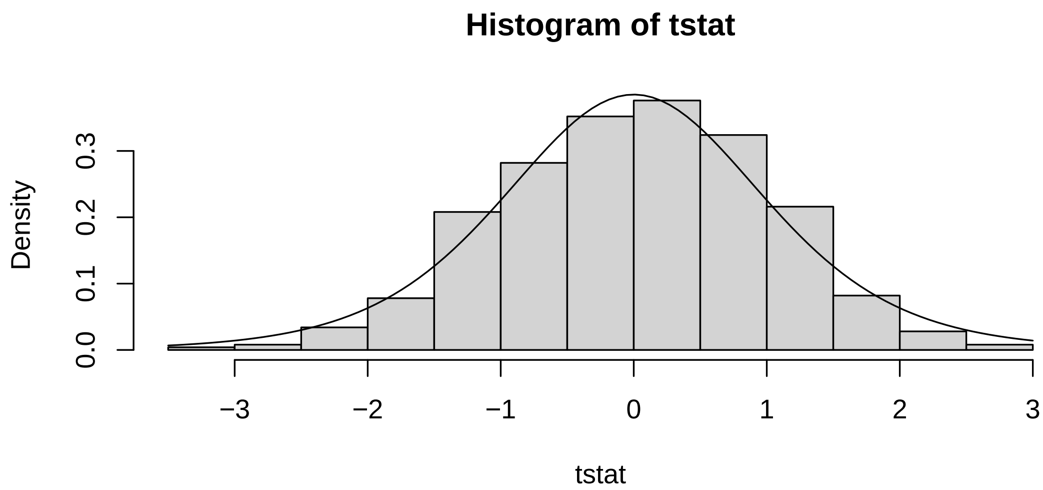
1 (b) Calculate t statistics from **xbar** and **std** using the formula:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

```
tstat <- (xbar - mu)/(sigma/sqrt(n))
```

1 (c) Plot a histogram of the t values and overlay with a curve of the t pdf on $n - 1 = 4$ degrees of freedom. Does the curve fit the histogram?

```
hist(tstat, freq = FALSE)
curve(dt(x, df = n - 1), add = TRUE)
```



The curve fits the histogram. The t statistic defined in the way can be shown - using the techniques discussed in an earlier lecture to exactly have a t distribution on $n-1$ degrees of freedom.

4. Repeat the previous question, using $\sigma = 20$.

```
N <- 1000; n <- 8; mu <- 3; sigma <- 20
normalsamples <- matrix(rnorm(N*n, mean = mu, sd = sigma), nrow=n)
```

2 (a) For all of the samples, compute sample means and standard deviations, assigning them to objects `xbar` and `std`.

```
xbar <- apply(normalsamples, 2, mean)
std <- apply(normalsamples, 2, sd)
```

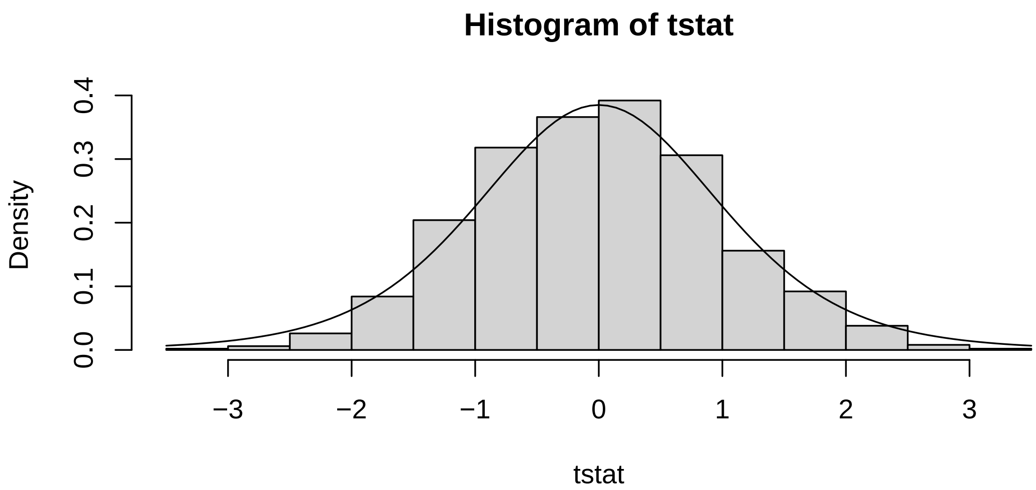
1 (b) Calculate t statistics from `xbar` and `std` using the formula:

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}}.$$

```
tstat <- (xbar - mu)/(sigma/sqrt(n))
```

1 (c) Plot a histogram of the t values and overlay with a curve of the t pdf on $n-1 = 4$ degrees of freedom. Does the curve fit the histogram?

```
hist(tstat, freq = FALSE)
curve(dt(x, df = n - 1), add = TRUE)
```



The curve fits the histogram. The t statistic defined in the way can be shown - using the techniques discussed in an earlier lecture to exactly have a t distribution on $n-1$ degrees of freedom.

5. Suppose the joint density function for random variables X and Y is given by $f_{X,Y}(x,y) = kye^{x-y^2}$, for $0 < x < 1$ and $0 < y < 1$, and 0, otherwise, for some constant k.

1 (a) Find the marginal pdf of X.

$$f_X(x) = \int_0^1 ke^{x-y^2} ydy = \frac{k}{2}e^x(1-e^{-1}), \quad 0 < x < 1.$$

In order for this to be true pdf, the value of k must be $\frac{2}{(e-1)(1-e^{-1})}$.

1 (b) Find the marginal pdf of Y.

$$f_Y(y) = \frac{2ye^{-y^2}}{1-e^{-1}}.$$

1 (c) Are X and Y independent? Explain briefly.

They are independent, because the product of $f_X(x)$ and $f_Y(y)$ is $f(x,y)$.

6. Suppose X and Y have joint pdf given by $f_{X,Y}(x,y) = 1.5(x^2+y^2)$ for $x \in [0,1]$ and $y \in [0,1]$, and 0, otherwise.

1 (a) Are X and Y independent? Explain.

The joint pdf cannot be factored into a product of functions of x alone and y alone, so X and Y cannot be independent.

2 (b) Find the marginal pdf of X

$$f_X(x) = \int_0^1 f_{X,Y}(x,y)dy = 1.5x^2 + 1/2.$$

2 (c) Find the conditional pdf of Y, given X.

$$f_{Y|X}(y;x) = \frac{x^2+y^2}{x^2+1/3}.$$

2 (d) Find the conditional cdf of Y, given $X = x$.

$$F_{Y|X}(y;x) = \frac{3x^2y + y^3}{3x^2 + 1}.$$

7. Suppose X and Y have joint pdf given by $f_{X,Y}(x, y) = (x + y)$ for $x \in [0, 1]$ and $y \in [0, 1]$, and 0, otherwise.

1 (a) Are X and Y independent? Explain.

They are not independent, because it is not possible to factor the joint pdf into a product of functions of x alone and y alone. (You could also integrate to find the marginal pdf of X and the marginal pdf of Y and observe that their product is not the joint pdf.

2 (b) Find the marginal pdf of X.

$$f_X(x) = \int_0^1 (x + y)dy = \frac{1}{2} + x, \qquad x \in [0, 1].$$

2 (c) Show that the cdf of X is $F_X(x) = x(1 + x)/2$.

$$F_X(x) = \int_0^x \left(\frac{1}{2} + z\right) dz = \frac{x}{2} + \frac{x^2}{2}.$$

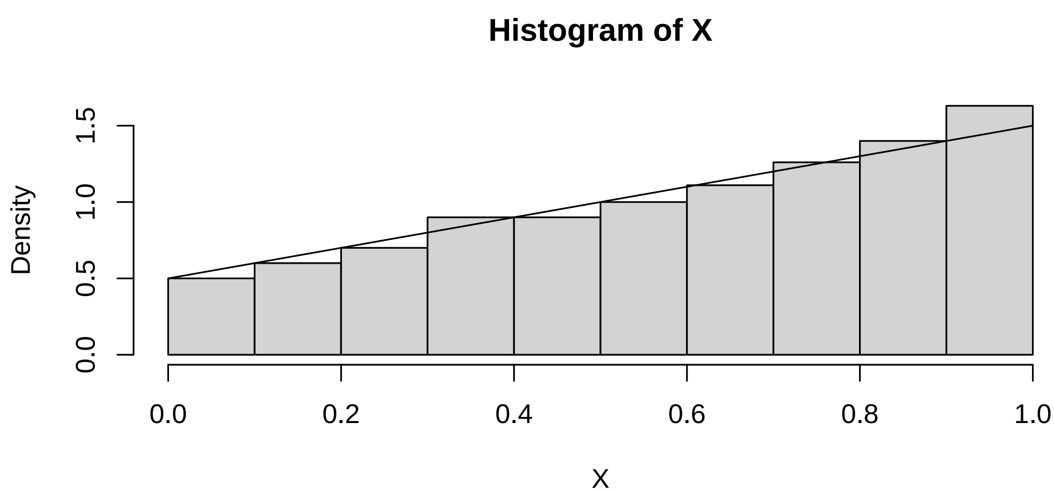
3 (d) Show that the inverse of the cdf is $F^{-1}(U) = (\sqrt{8U + 1} - 1)/2$, and use this to simulate 1000 variates from the marginal distribution of X. Plot the histogram of the values.

You can find the inverse function by using the quadratic formula to solve for x in

$$U = x(1 + x)/2.$$

There are two roots, but only the positive one makes sense.

```
U <- runif(1000)
X <- (sqrt(8*U+1) - 1)/2
hist(X, freq = FALSE)
curve(1/2 + x, add = TRUE) # not required
```



$\boxed{2}$ (e) Find the conditional pdf of Y, given X.

$$f_{Y|X}(y;x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{x+y}{1/2+x}.$$

$\boxed{2}$ (f) Find the conditional cdf of Y, given $X = x$.

$$F_{Y|X}(y;x) = \int_0^y \left(\frac{x+y}{1/2+x}\right) dy = \frac{xy}{1/2+x} + \frac{y^2}{1+2x}.$$

$\boxed{4}$ (g) Find the inverse function for the cdf and use this to simulate 1000 values of Y, assuming $X = .25$. Plot the histogram of the result.

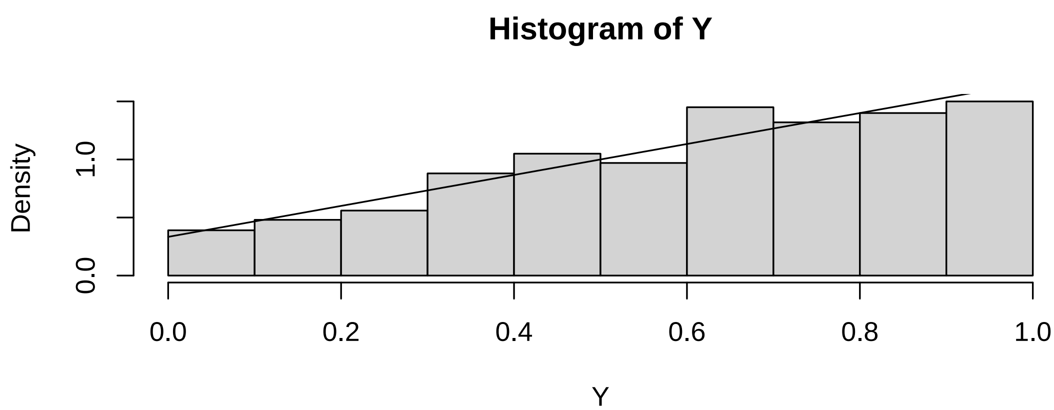
Again, use the quadratic formula on

$$Y^2 + 2xY - (2x+1)U = 0$$

and choose the positive root to see that

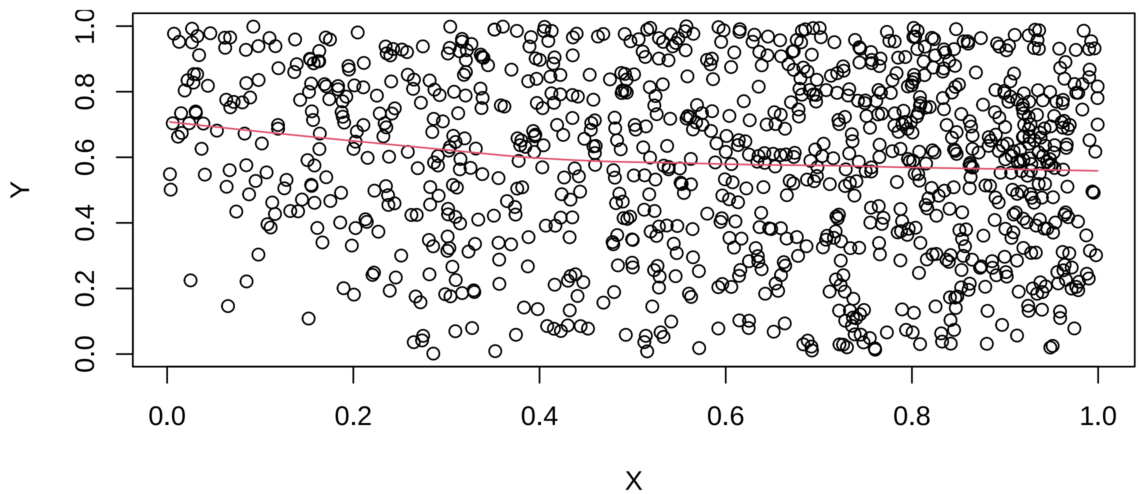
$$Y = -x + \sqrt{x^2 + U(2x+1)}.$$

```
x <- .25
U <- runif(1000)
Y <- -x + sqrt(x^2 + U*(2*x + 1))
hist(Y, freq=FALSE)
curve((.25 + x)/(1/2 + .25), add = TRUE)   # not required
```



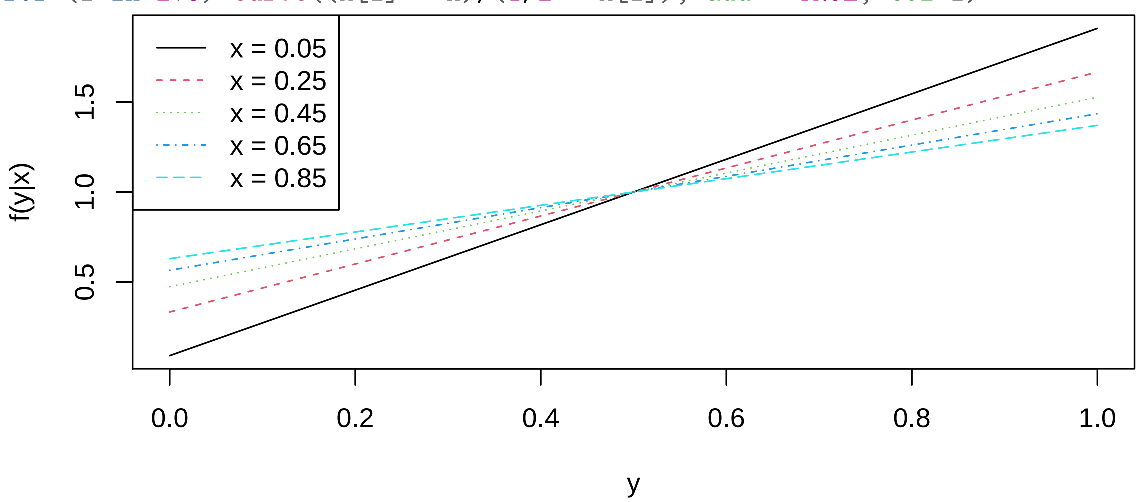
$\boxed{3}$ (h) Now, simulate 1000 values of Y, given the 1000 values of X simulated in part (d). Obtain a scatterplot of Y versus X and see if you see evidence of dependence.

```
Y <-  -X + sqrt(X^2 + U*(2*X + 1))
plot(Y ~ X)
```



A way to more clearly see the dependence of Y on X is to plot the conditional cdf for various values of X, say .05, .25, .45, .65, .85:

```
X <- seq(.05, .85, .2)
curve((X[1] + x)/(1/2 + X[1]))
for (i in 2:5) curve((X[i] + x)/(1/2 + X[i]), add = TRUE, col=i)
```



The different colored and dashed lines correspond to the different values of x. This means that the probability distribution of Y is changing according to the value of x.

8. Suppose X and Y have joint pdf given by $f_{X,Y}(x,y) = 0.5\cos(x-y)$ for $x \in [-\pi/4, \pi/4]$ and $y \in [-\pi/4, \pi/4]$, and 0, otherwise.

$\boxed{1}$ (a) Are X and Y independent? Explain.

The joint pdf cannot be factored into products of functions of x alone and y alone. Therefore, X and Y cannot be independent.

$\boxed{2}$ (b) Find the marginal pdf of Y.

$$f_Y(y) = .5(\sin(\pi/4 - y) + .5\sin(\pi/4 + y), \qquad y \in [-\pi/4, \pi/4].$$

$\boxed{2}$ (c) Show that the conditional distribution of X, given Y is

$$f_{X|Y}(x,y) = \frac{\cos(x-y)}{\sin(\pi/4 - y) + \sin(\pi/4 + y)}.$$

This follows from the fact that

$$f_{X|Y}(x;y) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

$\boxed{2}$ (d) Show that the conditional cdf of X, given $Y = y$ is

$$P(X \le x|Y = y) = \frac{\sin(x-y) + \sin(\pi/4 + y)}{\sin(\pi/4 - y) + \sin(\pi/4 + y)}.$$

Use the fact that the indefinite integral of $\cos(x-y)$ *with respect to* x *is* $\sin(x-y)$*, and note that* $\sin(-\pi/4 - y) = -\sin(\pi/4 + y)$*.*

$\boxed{4}$ (e) Find the inverse function for the cdf and use this to simulate 1000 values of X, assuming $Y = \pi/8$. Plot the histogram of the result.

Solve $U = F_{X|Y}(x;y)$ *for* X*:*

$$X = \arcsin(U(\sin(\pi/4 - y) + \sin(\pi/4 + y)) - \sin(\pi/4 + y)) + y$$

```r
y <- pi/8
U <- runif(1000)
X <- asin(U*(sin(pi/4 - y) + sin(pi/4 + y)) - sin(pi/4 + y)) + y
hist(X, freq=FALSE)
curve(cos(x - y)/(sin(pi/4 - y) + sin(pi/4 + y)), add = TRUE)
  # not required
```