§8.1 Density of a Transformed Random Variable

Example Suppose $R \sim \text{Unif}(0,2)$ is the radius of a circle. Let $A = \pi R^2$ be its area. Find the density of A. Note the range of A is $(0, 4\pi)$.

Let $F_A(x) = P(A \le x)$ be the CDF of A. Then

$$F_A(x) = P(A \le x) = P(\pi R^2 \le x)$$

$$= P\left(R^2 \le \frac{x}{\pi}\right)$$

$$= P\left(R \le \sqrt{\frac{x}{\pi}}\right)$$

$$= F_R\left(\sqrt{\frac{x}{\pi}}\right)$$

Therefore,

$$f_A(x) = F_A'(x) = F_R'\left(\sqrt{\frac{x}{\pi}}\right) \cdot \frac{d}{dx}\left(\sqrt{\frac{x}{\pi}}\right) \quad \text{(chain rule)}$$

(note the density of R is

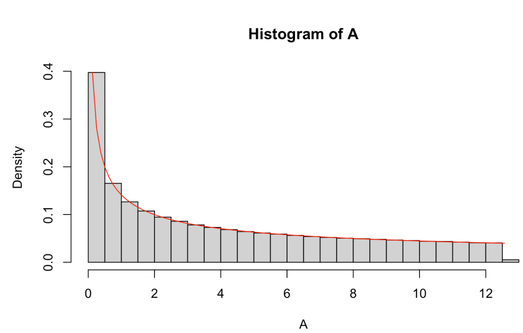
$$f_R(x) = \begin{cases} 1/2 & 0 < x < 2 \\ 0 & \text{else} \end{cases} \text{)}$$

$$= f_R\left(\sqrt{\frac{x}{\pi}}\right) \cdot \frac{1}{2}\left(\frac{x}{\pi}\right)^{-1/2} \cdot \frac{1}{\pi}$$

$$= \frac{1}{2} \cdot \frac{1}{2}\left(\frac{x}{\pi}\right)^{-1/2} \frac{1}{\pi}$$

$$= \frac{1}{4\sqrt{\pi x}} \qquad 0 < x < 4\pi$$

```{r}
R = runif(1e6,0,2)
A = pi*R^2
hist(A, probability = TRUE)
curve(1/(4*sqrt(pi*x)), from = 0, to = 4*pi, col = "red", add = TRUE)
```



Example Suppose $X \sim \mathrm{Unif}(0,1)$ and let $Y = \ln(X+1)$.

Find the range of Y and its density.

Since the range of X is the interval $(0,1)$, the range of Y is the interval $(\ln(1), \ln(2)) = (0, \ln 2)$.

The CDF of Y is

$$F_Y(x) = P(Y \le x) = P(\ln(X+1) \le x)$$

$$= P(X+1 \le e^x)$$

$$= P(X \le e^x - 1)$$

$$= F_X(e^x - 1)$$

Therefore, the density of Y is

$$f_Y(x) = F_Y'(x)$$

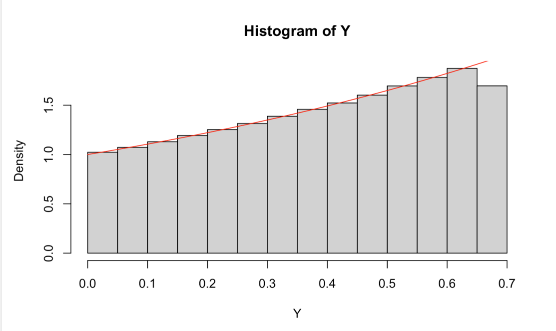
$$= F_X'(e^x - 1) \cdot \frac{d}{dx}(e^x - 1) \quad \text{(chain rule)}$$

$$= f_X(e^x - 1) \cdot e^x$$

$$= e^x \quad \text{when } 0 < x < \ln 2.$$

Note the density of X is $f_X(x) = \begin{cases} 1 & \text{if } 0<x<1 \\ 0 & \text{else} \end{cases}$

```{r}
X = runif(1e6,0,1)
Y = log(X+1)
hist(Y, probability = TRUE)
curve(exp(x), from = 0, to = log(2), col = "red", add = TRUE)
```



Theorem (Inverse Transform Method) Let $U \sim Unif(0,1)$, let F be an invertible function, and let $Y = F^{-1}(U)$. Then the CDF of Y is F.

Proof

$$F_Y(x) = P(Y \le x) = P(F^{-1}(U) \le x)$$
$$= P(U \le F(x))$$
$$= F(x)$$

Practical meaning Most programming languages can simulate random numbers from $Unif(0,1)$ distribution. This method lets you simulate random numbers from a distribution with CDF given by F, as long as F is invertible, by simulating $Unif(0,1)$ random numbers and plugging them into F^{-1}!

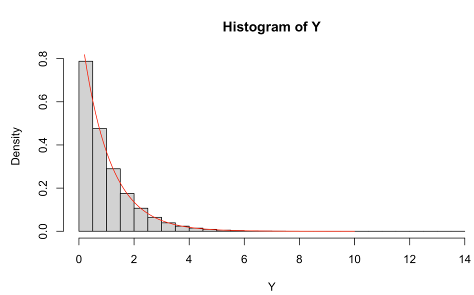
Example Use the runif command in R to simulate 1 million samples from the $Exp(1)$ distribution.

Let $Y \sim Exp(1)$. Then the CDF of Y is $F(x) = 1 - e^{-x}$ when $x > 0$. To find F^{-1} note:

$$x = 1 - e^{-y}$$
$$e^{-y} = 1 - x$$
$$-y = \ln(1-x)$$
$$y = -\ln(1-x), \text{ so } F^{-1}(x) = -\ln(1-x).$$

Therefore $Y = -\ln(1 - X)$ where $X \sim Unif(0,1)$.

```{r}
X = runif(1e6,0,1)
Y = -log(1-X) # has the Exp(1) distribution
hist(Y, probability = TRUE)
curve(exp(-x), from = 0, to = 10, col = "red", add = TRUE)
```



Problem 1. Let $X \sim \text{Exp}(\lambda)$ and let $Y = cX$. Find the density f_Y of Y and state its distribution.

$$F_Y(x) = P(Y \le x) = P(cX \le x)$$
$$= P\left(X \le \frac{x}{c}\right)$$
$$= F_X\left(\frac{x}{c}\right)$$

$$f_Y(x) = F_Y'(x) = F_X'\left(\frac{x}{c}\right) \cdot \frac{d}{dx}\left(\frac{x}{c}\right)$$
$$= f_X\left(\frac{x}{c}\right) \cdot \frac{1}{c}$$
$$= \lambda e^{-\lambda\left(\frac{x}{c}\right)} \cdot \frac{1}{c}$$
$$= \frac{\lambda}{c} e^{-\left(\frac{\lambda}{c}\right)x} \quad \text{when } x > 0$$

$$Y \sim Exp\left(\frac{1}{c}\lambda\right)$$

Problem 2. Let $X \sim \text{Exp}(1)$ and let $Y = \sqrt{X}$. Find the density f_Y of Y.

$$F_Y(x) = P(Y \le x) = P(\sqrt{X} \le x)$$
$$= P(X \le x^2)$$
$$= F_X(x^2)$$

$$f_Y(x) = F_Y'(x) = F_X'(x^2) \cdot 2x$$
$$= f_X(x^2) \cdot 2x$$
$$= 2x e^{-x^2} \quad \text{when } x > 0.$$

Problem 3. The density, and hence distribution, of Y in Problem 2 is probably not familiar to you, but we can check our answer with simulation. The following steps will show you how make a large (say 1 million element) i.i.d. sample of Y random variables, plot the histogram of the sample, and overlay the density curve you found to see if matches the histogram. In an RMarkdown code chunk do the following:

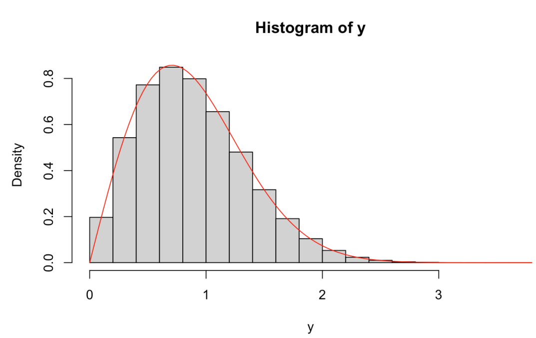
a. Use the command `x = rexp(1e6, 1)` to generate 1 million random numbers sampled from the Exp(1) distribution.

b. Use the command `y = sqrt(x)` to take the square root of each element of our sample.

c. Use the command `hist(y, prob = T)` to make a histogram of our sample.

d. Use the command `curve(..., add = T, col = ''red'')` to overlay your proposed density. For example, if you believe the density is $f_Y(x) = 5x^2e^x$ then in place of `...` you will put `5*x^2*exp(x)`.

```
```{r}
x = rexp(1e6,1)
y = sqrt(x)
hist(y, prob = T)
curve(2*x*exp(-x^2), add = T, col = "red")
```
```



Problem 4. Let

$$f(x) = \begin{cases} 3x^2 & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

be a given density function and let

$$F(x) = \begin{cases} 0 & x \leq 0 \\ x^3 & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}$$

be its corresponding cumulative distribution function.

a. State $F^{-1}(x)$ for $0 < x < 1$.

b. Let $U \sim \text{Unif}(0,1)$ and $Y = F^{-1}(U)$. Find the density of Y.

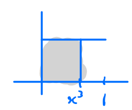
c. Use R to simulate 1 million samples of Y. Note that the R command `runif(1e6, 0, 1)` will let you simulate 1 million samples of U.

d. Make a histogram of your samples of Y and overlay its density.

(a) $F^{-1}(x) = x^{1/3}$

(b) $Y = U^{1/3}$.

$$F_Y(x) = P(Y \leq x) = P(U^{1/3} \leq x)$$
$$= P(U \leq x^3)$$
$$= x^3$$



$$f_Y(x) = F_Y'(x) = 3x^2 \quad \text{when} \quad 0 < x < 1$$

````
```{r}
x = runif(1e6,0, 1)
y = x^(1/3)
hist(y, prob = T)
curve(3*x^2, from = 0, to = 1, add = T, col = "red")
```
````