3 Transformation of random variables

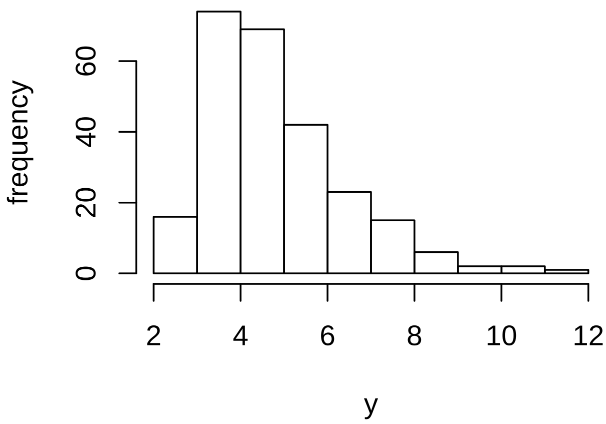
3.1 Transforming a normal random variable

The normal distribution is very widely used to model data. However, normal random variables take values on the entirety of $\mathbb{R}$ and they are symmetric around the mean. This is unrealistic in many situations – for example, if we want to model people's heights or weights. One way to obtain a positive random variable is to define

$$Y = e^X, \quad \text{where } X \sim N(\mu, \sigma^2).$$

Then Y is a random variable which is a **transformation** of X. What does the distribution of Y look like? We can use sampling to find out, at least approximately:

```r
mu = 1.5
sigma = 0.3
n = 250
y = exp(rnorm(n, mean = mu, sd = sigma))
hist(y, xlab = 'y', ylab = 'frequency', main = '')
```



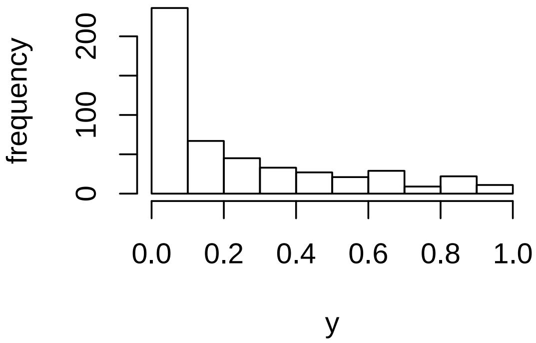
The random variable Y can only adopt positive values, and the distribution is positively skewed (long tail on the right). This kind of distribution is useful to model quantities like height – but what is the PDF of Y? A first guess is $f_Y(y) = f_X(\log(y))$ since log maps Y back to X, but this is **not** the correct answer.

3.2 Computing PDFs for transformed random variables

Example 3.1: If $X \sim U(0,1)$, what is the PDF of $Y = X^3$?

To confirm this result, we can simulate Y. Does the histogram look right?

```
y = (runif(500))^3
hist(y, xlab = 'y', ylab = 'frequency', main = '')
```



Notes:

Example 3.2: If $X \sim N(0,1)$, what is the PDF of $Y = X^2$?

Example 3.3: If $X \sim N(\mu, \sigma^2)$, what is the PDF of $Y = e^X$? This is the example used at the start of this section. We say Y has the **log-normal** distribution and write $Y \sim LN(\mu, \sigma^2)$.

3.3 The transformation method

These examples are all special cases of the following result.

If X has PDF $f_X(x)$ and $h(x)$ is a monotonically increasing or monotonically decreasing function then h is invertible and the PDF of $Y = h(X)$ is

$$f_Y(y) = f_X(h^{-1}(y)) \times \left| \frac{d\left(h^{-1}(y)\right)}{dy} \right|$$

when y is in the range of $h(x)$ and zero otherwise.

This result allows direct calculation of $f_Y(y)$ and is called the **univariate transformation method**. We will see this result again in semester 2, in MAS2903.

Proof:

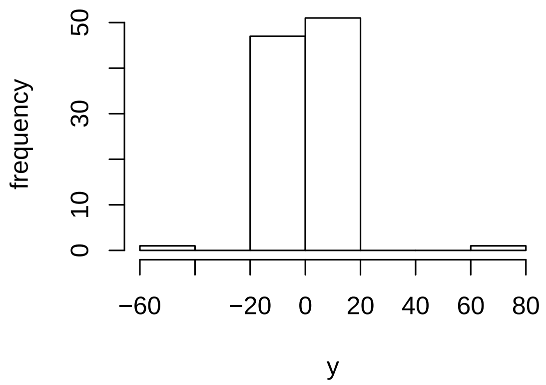
Example 3.4: When $X \sim U(0,1)$, use the transformation method to find the PDF of $Y = X^3$.

Example 3.5: When $X \sim U(-1,1)$ and $Y = e^X$, use the transformation method to find the PDF of Y.

Example 3.6: Use the transformation method to find the PDF of the log-normal distribution.

Example 3.7: Suppose $X \sim U(-\frac{\pi}{2}, \frac{\pi}{2})$ and $Y = \tan(X)$. Use simulation to study the distribution of Y and explicitly compute the PDF.

```
y = tan(runif(100, -pi/2, pi/2))
hist(y, xlab = 'y', ylab = 'frequency', main = '')
```



This distribution looks odd! Calculate the PDF:-