

Lecture: 10am, Monday 28th October

MAS2906: Computational Probability and Statistics with R

Newcastle University
`lee.fawcett@ncl.ac.uk`

Semester 1, 2024/25

Chapter 4

The Inverse CDF Method

4 The Inverse CDF Method

Let's pretend that the built-in function in R to simulate from an exponential distribution (`rexp`) does not exist.

How can we simulate values from this distribution?

What does R do when we use the function `rexp`?

We can use the **Inverse CDF Method** to generate random values from the exponential distribution (and indeed *any* distribution with a valid CDF).

4 The Inverse CDF Method

Let's pretend that the built-in function in R to simulate from an exponential distribution (`rexp`) does not exist.

How can we simulate values from this distribution?

What does R do when we use the function `rexp`?

We can use the **Inverse CDF Method** to generate random values from the exponential distribution (and indeed *any* distribution with a valid CDF).

4 The Inverse CDF Method

Let's pretend that the built-in function in R to simulate from an exponential distribution (`rexp`) does not exist.

How can we simulate values from this distribution?

What does R do when we use the function `rexp`?

We can use the *Inverse CDF Method* to generate random values from the exponential distribution (and indeed any distribution with a valid CDF).

4 The Inverse CDF Method

Let's pretend that the built-in function in R to simulate from an exponential distribution (`rexp`) does not exist.

How can we simulate values from this distribution?

What does R do when we use the function `rexp`?

We can use the **Inverse CDF Method** to generate random values from the exponential distribution (and indeed *any* distribution with a valid CDF!).

4 The Inverse CDF Method

Suppose $X \sim \text{Exp}(\lambda)$. Then we know that the probability density function (PDF) is given by

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

The following code can be used to produce a plot of the PDF for X when $\lambda = 2$.

```
#Set up a sequence of values for the x co-ordinates:
x = seq(0, 4, 0.01)
#Now generate the corresponding y co-ordinates:
lambda = 2
y = lambda*exp(-lambda*x)
plot(x, y, type='l', col='purple', lwd=2, main='PDF for
      X, where X~exp(2)', ylab='Density')
```

4 The Inverse CDF Method

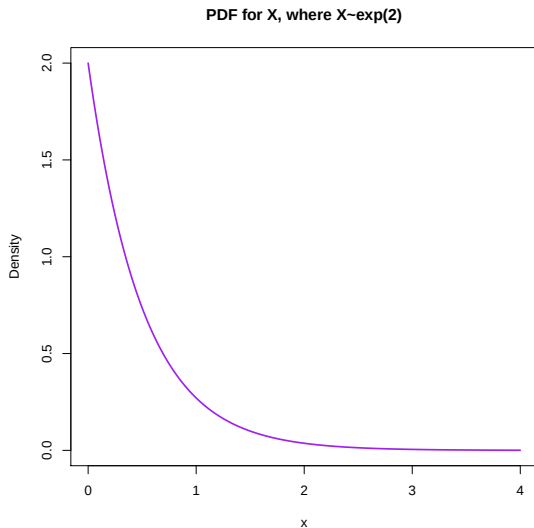
Suppose $X \sim \text{Exp}(\lambda)$. Then we know that the probability density function (PDF) is given by

$$f_X(x) = \lambda e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

The following code can be used to produce a plot of the PDF for X when $\lambda = 2$.

```
1 #Set up a sequence of values for the x co-ordinates:
2 x = seq(0, 4, 0.01)
3 #Now generate the corresponding y co-ordinates:
4 lambda = 2
5 y = lambda*exp(-lambda*x)
6 plot(x, y, type="l", col="purple", lwd=2, main="PDF for
   X, where X~exp(2)", ylab="Density")
```

4 The Inverse CDF Method



4 The Inverse CDF Method

Now we want to be able to produce a sample of values from X , where $X \sim \text{Exp}(2)$, and we can do this using the inverse CDF method.

It works like this:

- Find the inverse of the CDF (cumulative distribution function) for our target distribution
- Apply the inverse CDF to $U \sim U(0, 1)$

Let's see why this works.

4 The Inverse CDF Method

Now we want to be able to produce a sample of values from X , where $X \sim \text{Exp}(2)$, and we can do this using the inverse CDF method.

It works like this:

- Find the inverse of the CDF (cumulative distribution function) for our target distribution
- Apply the inverse CDF to $U \sim U(0, 1)$

Let's see why this works.

4 The Inverse CDF Method

Now we want to be able to produce a sample of values from X , where $X \sim \text{Exp}(2)$, and we can do this using the inverse CDF method.

It works like this:

- Find the inverse of the CDF (cumulative distribution function) for our target distribution
- Apply the inverse CDF to $U \sim U(0, 1)$

Let's see **why** this works.

4 The Inverse CDF Method

First, we need to find the CDF.

You might know from other courses that the CDF for the exponential distribution is

$$F_X(x) = 1 - e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

But what if you didn't know this? Then you'd need to integrate over the PDF. Thus,

$$\begin{aligned} F_X(x) &= \int_0^x \lambda e^{-\lambda t} dt \\ &= \lambda \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^x \\ &= \left[-e^{-\lambda t} \right]_0^x \\ &= \left(-e^{-\lambda x} \right) - \left(-e^{-0} \right) = 1 - e^{-\lambda x}, \quad \text{as required.} \end{aligned}$$

4 The Inverse CDF Method

First, we need to find the CDF.

You might know from other courses that the CDF for the exponential distribution is

$$F_X(x) = 1 - e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

But what if you didn't know this? Then you'd need to integrate over the PDF. Thus,

$$\begin{aligned} F_X(x) &= \int_0^x \lambda e^{-\lambda t} dt \\ &= \lambda \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^x \\ &= \left[-e^{-\lambda t} \right]_0^x \\ &= \left(-e^{-\lambda x} \right) - \left(-e^{-0} \right) = 1 - e^{-\lambda x}, \quad \text{as required.} \end{aligned}$$

4 The Inverse CDF Method

First, we need to find the CDF.

You might know from other courses that the CDF for the exponential distribution is

$$F_X(x) = 1 - e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

But what if you didn't know this? Then you'd need to integrate over the PDF. Thus,

$$\begin{aligned} F_X(x) &= \int_0^x \lambda e^{-\lambda t} dt \\ &= \lambda \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^x \\ &= \left[-e^{-\lambda t} \right]_0^x \\ &= \left(-e^{-\lambda x} \right) - \left(-e^{-0} \right) = 1 - e^{-\lambda x}, \quad \text{as required.} \end{aligned}$$

4 The Inverse CDF Method

First, we need to find the CDF.

You might know from other courses that the CDF for the exponential distribution is

$$F_X(x) = 1 - e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

But what if you didn't know this? Then you'd need to integrate over the PDF. Thus,

$$F_X(x) = \int_0^x \lambda e^{-\lambda t} dt$$

$$= \lambda \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^x$$

$$= \left[-e^{-\lambda t} \right]_0^x$$

$$= \left(-e^{-\lambda x} \right) - \left(-e^{-0} \right) = 1 - e^{-\lambda x}, \quad \text{as required.}$$

4 The Inverse CDF Method

First, we need to find the CDF.

You might know from other courses that the CDF for the exponential distribution is

$$F_X(x) = 1 - e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

But what if you didn't know this? Then you'd need to integrate over the PDF. Thus,

$$\begin{aligned} F_X(x) &= \int_0^x \lambda e^{-\lambda t} dt \\ &= \lambda \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^x \\ &= \left[-e^{-\lambda t} \right]_0^x \\ &= (-e^{-\lambda x}) - (-e^{-0}) = 1 - e^{-\lambda x}, \quad \text{as required.} \end{aligned}$$

4 The Inverse CDF Method

First, we need to find the CDF.

You might know from other courses that the CDF for the exponential distribution is

$$F_X(x) = 1 - e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

But what if you didn't know this? Then you'd need to integrate over the PDF. Thus,

$$\begin{aligned} F_X(x) &= \int_0^x \lambda e^{-\lambda t} dt \\ &= \lambda \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^x \\ &= \left[-e^{-\lambda t} \right]_0^x \\ &= (-e^{-\lambda x}) - (-e^{-0}) = 1 - e^{-\lambda x}, \quad \text{as required.} \end{aligned}$$

4 The Inverse CDF Method

First, we need to find the CDF.

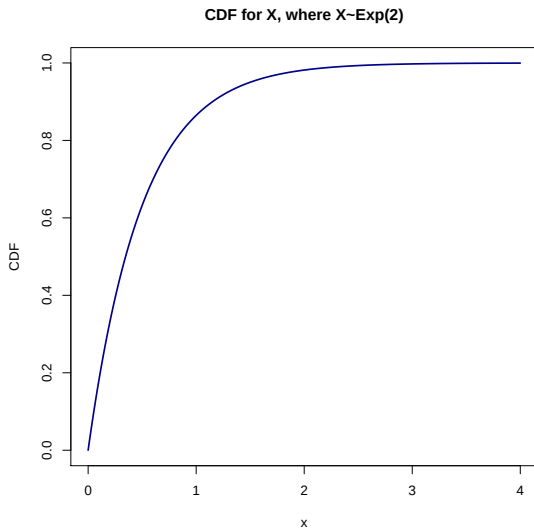
You might know from other courses that the CDF for the exponential distribution is

$$F_X(x) = 1 - e^{-\lambda x}, \quad x > 0, \quad \lambda > 0.$$

But what if you didn't know this? Then you'd need to integrate over the PDF. Thus,

$$\begin{aligned} F_X(x) &= \int_0^x \lambda e^{-\lambda t} dt \\ &= \lambda \left[-\frac{1}{\lambda} e^{-\lambda t} \right]_0^x \\ &= \left[-e^{-\lambda t} \right]_0^x \\ &= \left(-e^{-\lambda x} \right) - \left(-e^{-0} \right) = 1 - e^{-\lambda x}, \quad \text{as required.} \end{aligned}$$

4 The Inverse CDF Method



4 The Inverse CDF Method

Look carefully at the plot of the CDF.

- If we want to simulate values from $X \sim \text{Exp}(2)$, then we want values lying on the x-axis
- Values from the CDF, shown on the y-axis, are between 0 and 1 (as they represent (cumulative) probabilities)
- If we simulate values u from $U \sim U(0, 1)$, this in effect gives us random values from the y-axis in our plot...
- To then obtain corresponding values from X , on the x-axis, we need to feed our values from U through the *inverse* of the CDF

4 The Inverse CDF Method

Look carefully at the plot of the CDF.

- If we want to simulate values from $X \sim \text{Exp}(2)$, then we want values lying on the x-axis
 - Values from the CDF, shown on the y-axis, are between 0 and 1 (as they represent (cumulative) probabilities)
 - If we simulate values u from $U \sim U(0, 1)$, this in effect gives us random values from the y-axis in our plot...
 - To then obtain corresponding values from X , on the x-axis, we need to feed our values from U through the *inverse* of the CDF

4 The Inverse CDF Method

Look carefully at the plot of the CDF.

- If we want to simulate values from $X \sim \text{Exp}(2)$, then we want values lying on the x-axis
- Values from the CDF, shown on the y-axis, are between 0 and 1 (as they represent (cumulative) probabilities)
- If we simulate values u from $U \sim U(0, 1)$, this in effect gives us random values from the y-axis in our plot...
- To then obtain corresponding values from X , on the x-axis, we need to feed our values from U through the *inverse* of the CDF

4 The Inverse CDF Method

Look carefully at the plot of the CDF.

- If we want to simulate values from $X \sim \text{Exp}(2)$, then we want values lying on the x-axis
- Values from the CDF, shown on the y-axis, are between 0 and 1 (as they represent (cumulative) probabilities)
- If we simulate values u from $U \sim U(0, 1)$, this in effect gives us random values from the y-axis in our plot...

■ To then obtain corresponding values from X , on the x-axis, we need to feed our values from U through the *inverse* of the CDF

4 The Inverse CDF Method

Look carefully at the plot of the CDF.

- If we want to simulate values from $X \sim \text{Exp}(2)$, then we want values lying on the x-axis
- Values from the CDF, shown on the y-axis, are between 0 and 1 (as they represent (cumulative) probabilities)
- If we simulate values u from $U \sim U(0, 1)$, this in effect gives us random values from the y-axis in our plot...
- To then obtain corresponding values from X , on the x-axis, we need to feed our values from U through the *inverse* of the CDF

4 The Inverse CDF Method

So, we need the inverse of the CDF (hence the name of this method!). Recall we have

$$F_X(x) = 1 - e^{-\lambda x}.$$

Setting $y = 1 - e^{-\lambda x}$ and solving for x , gives:

$$\begin{aligned} 1 - y &= e^{-\lambda x} \\ \log(1 - y) &= -\lambda x \\ x &= -\frac{1}{\lambda} \log(1 - y). \end{aligned}$$

and so

$$F_X^{-1}(x) = -\frac{1}{\lambda} \log(1 - x).$$

4 The Inverse CDF Method

So, we need the inverse of the CDF (hence the name of this method!). Recall we have

$$F_X(x) = 1 - e^{-\lambda x}.$$

Setting $y = 1 - e^{-\lambda x}$ and solving for x , gives:

$$1 - y = e^{-\lambda x}$$

$$\log(1 - y) = -\lambda x$$

$$x = -\frac{1}{\lambda} \log(1 - y).$$

and so

$$F_X^{-1}(x) = -\frac{1}{\lambda} \log(1 - x).$$

4 The Inverse CDF Method

So, we need the inverse of the CDF (hence the name of this method!). Recall we have

$$F_X(x) = 1 - e^{-\lambda x}.$$

Setting $y = 1 - e^{-\lambda x}$ and solving for x , gives:

$$1 - y = e^{-\lambda x}$$

$$\log(1 - y) = -\lambda x$$

$$x = -\frac{1}{\lambda} \log(1 - y).$$

and so

$$F_X^{-1}(x) = -\frac{1}{\lambda} \log(1 - x).$$

4 The Inverse CDF Method

So, we need the inverse of the CDF (hence the name of this method!). Recall we have

$$F_X(x) = 1 - e^{-\lambda x}.$$

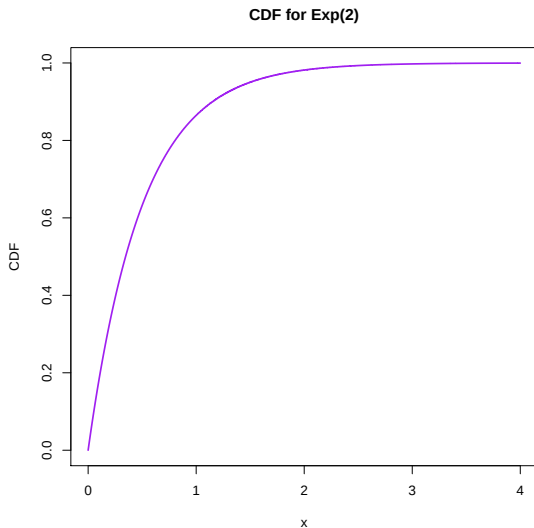
Setting $y = 1 - e^{-\lambda x}$ and solving for x , gives:

$$\begin{aligned} 1 - y &= e^{-\lambda x} \\ \log(1 - y) &= -\lambda x \\ x &= -\frac{1}{\lambda} \log(1 - y), \end{aligned}$$

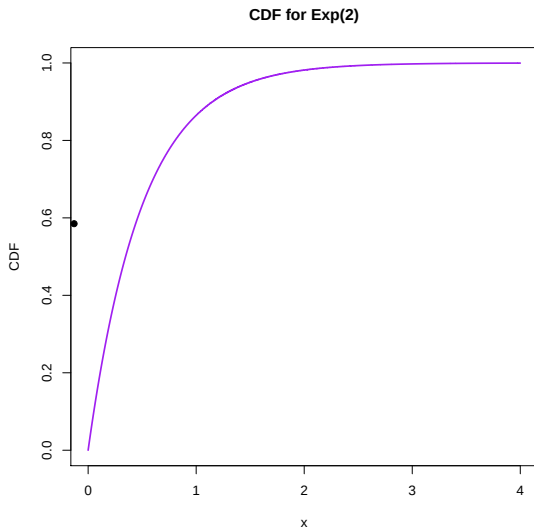
and so

$$F_X^{-1}(x) = -\frac{1}{\lambda} \log(1 - x).$$

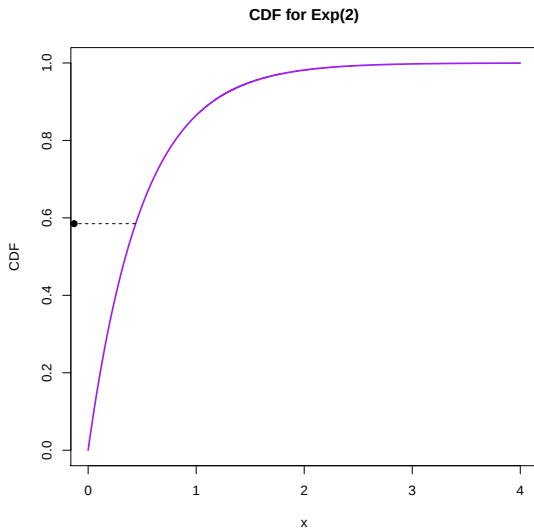
4 The Inverse CDF Method



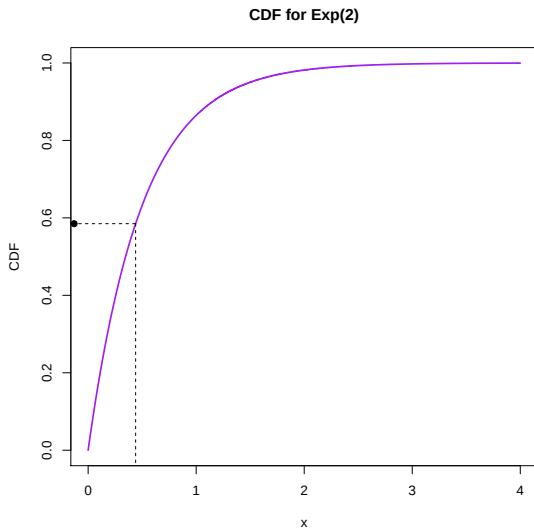
4 The Inverse CDF Method



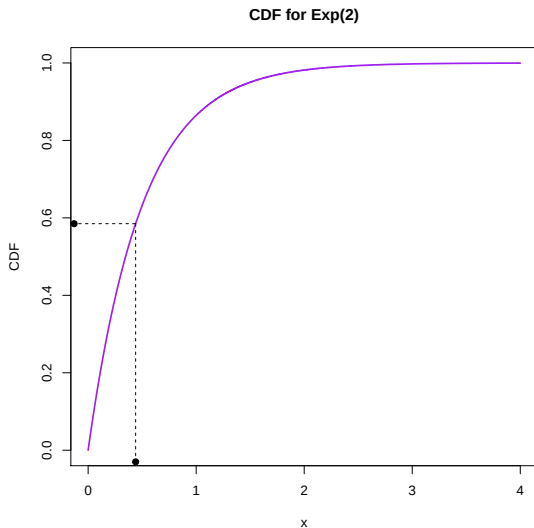
4 The Inverse CDF Method



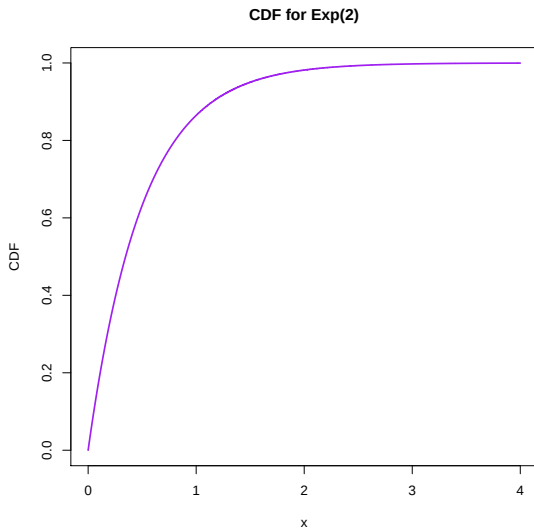
4 The Inverse CDF Method



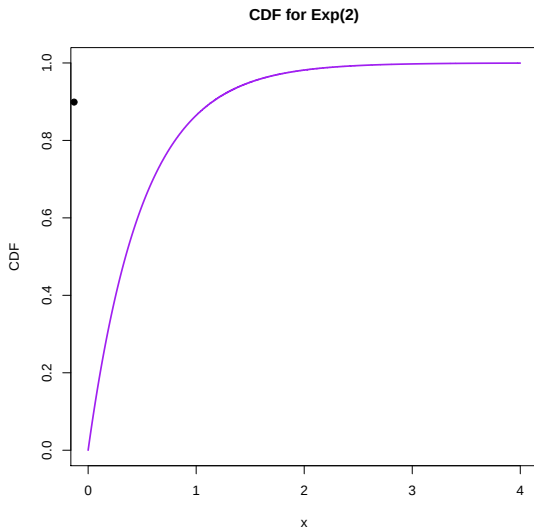
4 The Inverse CDF Method



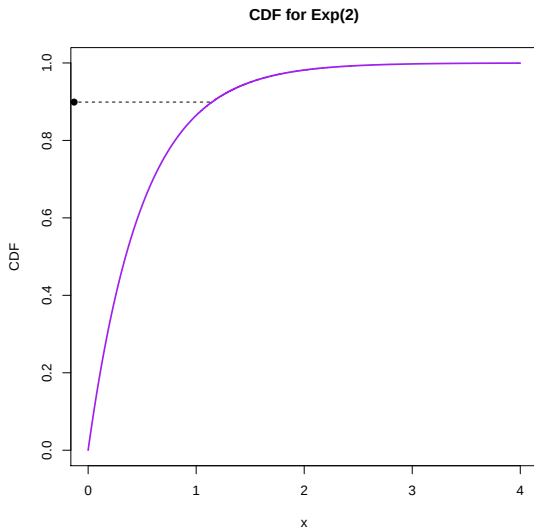
4 The Inverse CDF Method



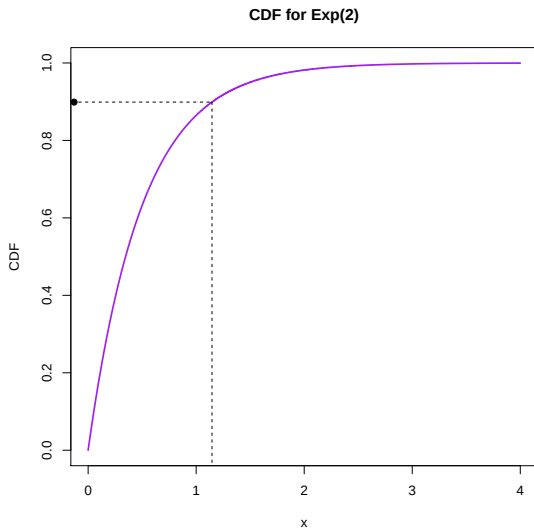
4 The Inverse CDF Method



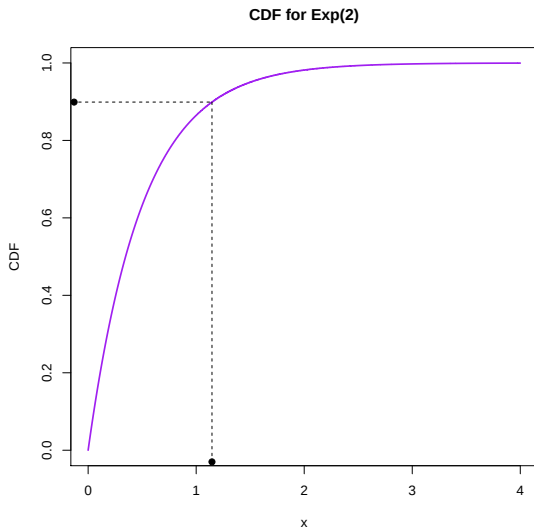
4 The Inverse CDF Method



4 The Inverse CDF Method



4 The Inverse CDF Method



4 The Inverse CDF Method

$$X = F^{-1}(U).$$

So

$$\begin{aligned}\Pr(X \leq x) &= \Pr(F^{-1}(U) \leq x) \\ &= \Pr(F(F^{-1}(U)) \leq F(x)) \\ &= \Pr(U \leq F(x)) \\ &= F(x),\end{aligned}$$

since $0 \leq F(x) \leq 1$ (see Lee's sketches too from the lecture).

4 The Inverse CDF Method

$$X = F^{-1}(U).$$

So

$$\begin{aligned}\Pr(X \leq x) &= \Pr(F^{-1}(U) \leq x) \\ &= \Pr(F(F^{-1}(U)) \leq F(x)) \\ &= \Pr(U \leq F(x)) \\ &= F(x),\end{aligned}$$

since $0 \leq F(x) \leq 1$ (see Lee's sketches too from the lecture).

4 The Inverse CDF Method

$$X = F^{-1}(U).$$

So

$$\begin{aligned}\Pr(X \leq x) &= \Pr(F^{-1}(U) \leq x) \\ &= \Pr(F(F^{-1}(U)) \leq F(x)) \\ &= \Pr(U \leq F(x)) \\ &= F(x),\end{aligned}$$

since $0 \leq F(x) \leq 1$ (see Lee's sketches too from the lecture).

4 The Inverse CDF Method

$$X = F^{-1}(U).$$

So

$$\begin{aligned}\Pr(X \leq x) &= \Pr(F^{-1}(U) \leq x) \\ &= \Pr(F(F^{-1}(U)) \leq F(x)) \\ &= \Pr(U \leq F(x)) \\ &= F(x),\end{aligned}$$

since $0 \leq F(x) \leq 1$ (see Lee's sketches too from the lecture).

4 The Inverse CDF Method

$$X = F^{-1}(U).$$

So

$$\begin{aligned}\Pr(X \leq x) &= \Pr(F^{-1}(U) \leq x) \\ &= \Pr(F(F^{-1}(U)) \leq F(x)) \\ &= \Pr(U \leq F(x)) \\ &= F(x),\end{aligned}$$

since $0 \leq F(x) \leq 1$ (see Lee's sketches too from the lecture).

Example 4.1

Example

Suppose we obtain two observations from a $U(0, 1)$ distribution: 0.1 and 0.85. Use these values to obtain two observations from $X \sim \text{Exp}(2)$.

Example 4.1: Solution

From the illustrative example at the start, we already know that

$$F_X^{-1}(x) = -\frac{1}{2} \log(1 - x).$$

We also know that to generate random values from X , we 'push' our observations from $U(0,1)$ through the inverse CDF, in this case giving:

$$\begin{aligned}x_1 &= -\frac{1}{2} \log(1 - 0.1) = 0.0527 \quad \text{and} \\x_2 &= -\frac{1}{2} \log(1 - 0.85) = 0.9486.\end{aligned}$$

Example 4.1: Solution

From the illustrative example at the start, we already know that

$$F_X^{-1}(x) = -\frac{1}{2} \log(1 - x).$$

We also know that to generate random values from X , we ‘push’ our observations from $U(0, 1)$ through the inverse CDF, in this case giving:

$$x_1 = -\frac{1}{2} \log(1 - 0.1) = 0.0527 \quad \text{and}$$

$$x_2 = -\frac{1}{2} \log(1 - 0.85) = 0.9486.$$

Example 4.1: Solution

From the illustrative example at the start, we already know that

$$F_X^{-1}(x) = -\frac{1}{2} \log(1 - x).$$

We also know that to generate random values from X , we ‘push’ our observations from $U(0, 1)$ through the inverse CDF, in this case giving:

$$\begin{aligned} x_1 &= -\frac{1}{2} \log(1 - 0.1) = 0.0527 \quad \text{and} \\ x_2 &= -\frac{1}{2} \log(1 - 0.85) = 0.9486. \end{aligned}$$

Example 4.2

Example

Write some R code to sample n observations from an $Exp(2)$ distribution.

Do not use the built-in `rexp` function.

Check your code by comparing a histogram of some simulated realisations against the PDF for an $Exp(2)$ distribution.

Also check your code against results from the built-in `rexp` function.

Example 4.2: Solution

The following code applies the inverse CDF for the exponential distribution (derived in the illustrative example) to simulated values u from $U \sim U(0, 1)$ – in effect getting us from y-axis to x-axis in the plot we looked at earlier:

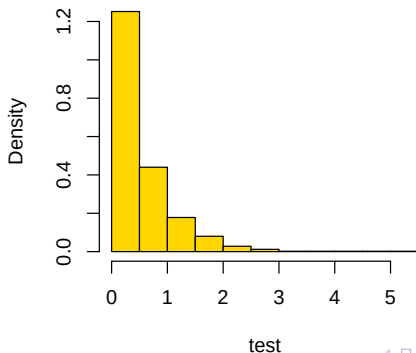
```
1 exp.sim = function(n, lambda=2){  
2   U = runif(n)  
3   answer = -1/lambda*log(1-U)  
4   return(answer)}
```

Example 4.2: Solution

Putting the function to good use:

```
1 test = exp.sim(1000)
2 hist(test, col="gold", main="Simulated Exp(2) random
  variates", probability=TRUE)
```

Simulated Exp(2) random variates

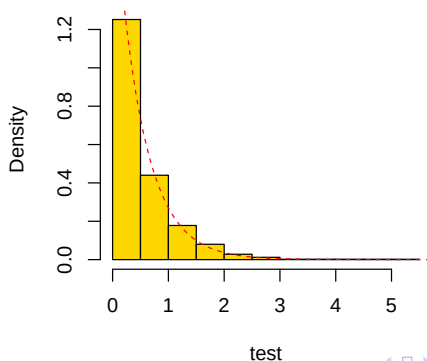


Example 4.2: Solution

Now to check it has worked!

```
1 x = seq(0, 6, 0.001)
2 y = 2*exp(-2*x)
3 lines(x, y, type="l", col="red", lty=2)
```

Simulated Exp(2) random variates



Example 4.2: Solution

Checking against R's built-in function:

```
test2 = rexp(1000, 2)
summary(test)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000064 0.147899 0.344216 0.507016 0.697876 5.889635
summary(test2)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000017 0.143081 0.346082 0.501171 0.693844 4.527852
```

Look very similar – we should be happy our function has worked!
Of course, we could always compare to the *true* values from $X \sim \text{Exp}(2)$, just to make sure!

- $E(X) = 1/\lambda = 1/2$
- $q(p) = -[\log(1-p)]/\lambda$, so
 - $q(0.25) = -[\log(1-0.25)]/2 = 0.1438$
 - $q(0.5) = -[\log(1-0.5)]/2 = 0.3466$
 - $q(0.75) = -[\log(1-0.75)]/2 = 0.6932$

Example 4.2: Solution

Checking against R's built-in function:

```
1 test2 = rexp(1000, 2)
2 summary(test)
3      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
4 0.000064 0.147899 0.344216 0.507016 0.697876 5.889635
5 summary(test2)
6      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
7 0.000017 0.143081 0.346082 0.501171 0.693844 4.527852
```

Look very similar – we should be happy our function has worked!

Of course, we could always compare to the true values from $X \sim \text{Exp}(2)$, just to make sure!

- $E(X) = 1/\lambda = 1/2$

- $q(p) = -\lceil \log(1-p) \rceil / \lambda$, so

- $q(0.25) = -\lceil \log(1-0.25) \rceil / 2 = 0.1438$

- $q(0.5) = -\lceil \log(1-0.5) \rceil / 2 = 0.3466$

- $q(0.75) = -\lceil \log(1-0.75) \rceil / 2 = 0.6932$

Example 4.2: Solution

Checking against R's built-in function:

```
1 test2 = rexp(1000, 2)
2 summary(test)
3      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
4 0.000064 0.147899 0.344216 0.507016 0.697876 5.889635
5 summary(test2)
6      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
7 0.000017 0.143081 0.346082 0.501171 0.693844 4.527852
```

Look very similar – we should be happy our function has worked!
Of course, we could always compare to the *true* values from $X \sim \text{Exp}(2)$, just to make sure!

■ $E(X) = 1/\lambda = 1/2$

■ $q(p) = -\log(1-p)/\lambda$, so

■ $q(0.25) = -\log(1-0.25)/2 = 0.1438$

■ $q(0.5) = -\log(1-0.5)/2 = 0.3466$

■ $q(0.75) = -\log(1-0.75)/2 = 0.6932$

Example 4.2: Solution

Checking against R's built-in function:

```
1 test2 = rexp(1000, 2)
2 summary(test)
3      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
4 0.000064 0.147899 0.344216 0.507016 0.697876 5.889635
5 summary(test2)
6      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
7 0.000017 0.143081 0.346082 0.501171 0.693844 4.527852
```

Look very similar – we should be happy our function has worked!
Of course, we could always compare to the *true* values from $X \sim \text{Exp}(2)$, just to make sure!

- $E(X) = 1/\lambda = 1/2$
- $q(p) = -[\log(1 - p)]/\lambda$, so

$$q(0.25) = -[\log(1 - 0.25)]/2 = 0.1438$$

$$q(0.5) = -[\log(1 - 0.5)]/2 = 0.3466$$

$$q(0.75) = -[\log(1 - 0.75)]/2 = 0.6932$$

Example 4.2: Solution

Checking against R's built-in function:

```
1 test2 = rexp(1000, 2)
2 summary(test)
3      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
4 0.000064 0.147899 0.344216 0.507016 0.697876 5.889635
5 summary(test2)
6      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
7 0.000017 0.143081 0.346082 0.501171 0.693844 4.527852
```

Look very similar – we should be happy our function has worked!
Of course, we could always compare to the *true* values from $X \sim \text{Exp}(2)$, just to make sure!

- $E(X) = 1/\lambda = 1/2$
- $q(p) = -[\log(1 - p)]/\lambda$, so
 - $q(0.25) = -[\log(1 - 0.25)]/2 = 0.1438$

$$q(0.5) = -[\log(1 - 0.5)]/2 = 0.3466$$

$$q(0.75) = -[\log(1 - 0.75)]/2 = 0.6932$$

Example 4.2: Solution

Checking against R's built-in function:

```
1 test2 = rexp(1000, 2)
2 summary(test)
3      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
4 0.000064 0.147899 0.344216 0.507016 0.697876 5.889635
5 summary(test2)
6      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
7 0.000017 0.143081 0.346082 0.501171 0.693844 4.527852
```

Look very similar – we should be happy our function has worked!
Of course, we could always compare to the *true* values from $X \sim \text{Exp}(2)$, just to make sure!

- $E(X) = 1/\lambda = 1/2$
- $q(p) = -[\log(1 - p)]/\lambda$, so
 - $q(0.25) = -[\log(1 - 0.25)]/2 = 0.1438$
 - $q(0.5) = -[\log(1 - 0.5)]/2 = 0.3466$

$q(0.75) = -[\log(1 - 0.75)]/2 = 0.6932$

Example 4.2: Solution

Checking against R's built-in function:

```
1 test2 = rexp(1000, 2)
2 summary(test)
3      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
4 0.000064 0.147899 0.344216 0.507016 0.697876 5.889635
5 summary(test2)
6      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.
7 0.000017 0.143081 0.346082 0.501171 0.693844 4.527852
```

Look very similar – we should be happy our function has worked!
Of course, we could always compare to the *true* values from $X \sim \text{Exp}(2)$, just to make sure!

- $E(X) = 1/\lambda = 1/2$
- $q(p) = -[\log(1 - p)]/\lambda$, so
 - $q(0.25) = -[\log(1 - 0.25)]/2 = 0.1438$
 - $q(0.5) = -[\log(1 - 0.5)]/2 = 0.3466$
 - $q(0.75) = -[\log(1 - 0.75)]/2 = 0.6932$

Example 4.3

Example

The **Pareto distribution**, named after the Italian civil engineer, economist, and sociologist Vilfredo Pareto, is a power-law probability distribution that is used in the description of social, scientific, geophysical, actuarial, and many other types of observable phenomena.

It was originally used to describe the distribution of wealth in a society, fitting the trend that a large portion of wealth is held by a small fraction of the population.

If $X \sim \text{Pareto}(\sigma)$, then it has cumulative distribution function

$$F_X(x; \sigma) = 1 - (\sigma/x)^{\log_e 5}, \quad x \geq \sigma, \sigma > 0.$$

Write a function in R to produce realisations from this distribution, and test your function out for $\sigma = \{1, 2, 5\}$.

Example 4.3

Example

The **Pareto distribution**, named after the Italian civil engineer, economist, and sociologist Vilfredo Pareto, is a power-law probability distribution that is used in the description of social, scientific, geophysical, actuarial, and many other types of observable phenomena.

It was originally used to describe the distribution of wealth in a society, fitting the trend that a large portion of wealth is held by a small fraction of the population.

If $X \sim \text{Pareto}(\sigma)$, then it has cumulative distribution function

$$F_X(x; \sigma) = 1 - (\sigma/x)^{\sigma}, \quad x \geq \sigma, \sigma > 0.$$

Write a function in R to produce realisations from this distribution, and test your function out for $\sigma = \{1, 2, 5\}$.

Example 4.3

Example

The **Pareto distribution**, named after the Italian civil engineer, economist, and sociologist Vilfredo Pareto, is a power-law probability distribution that is used in the description of social, scientific, geophysical, actuarial, and many other types of observable phenomena.

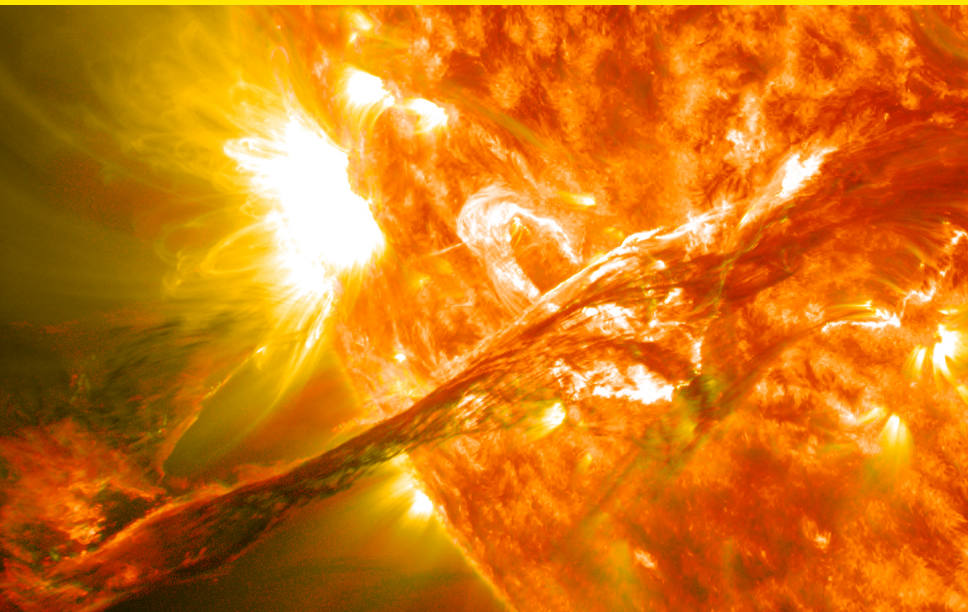
It was originally used to describe the distribution of wealth in a society, fitting the trend that a large portion of wealth is held by a small fraction of the population.

If $X \sim \text{Pareto}(\sigma)$, then it has cumulative distribution function

$$F_X(x; \sigma) = 1 - (\sigma/x)^{\log_4 5}, \quad x \geq \sigma, \sigma > 0.$$

Write a function in R to produce realisations from this distribution, and test your function out for $\sigma = \{1, 2, 5\}$.

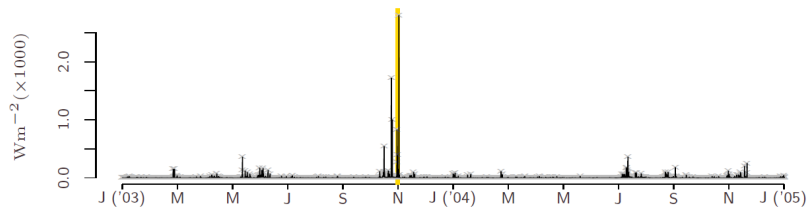
Example from my own research: Extreme solar storms



What's at stake?

- Disruption to **global positioning systems**. GPS industry estimated to be a \$1 trillion business
- **Satellite communications** at high risk
- Huge disruption to the **electrical grid**: areas such as the eastern US particularly vulnerable as power infrastructure highly interconnected, large cities without power for weeks
- Key questions: When might we see the next Carrington-like solar storm? What is the **return period** of a Carrington-like event? What is the strength of solar flare we might expect to see once per century?

Data



Example 4.3: Solution

Find the inverse CDF.

If $y = 1 - (\sigma/x)^{\log_4 5}$, then

$$\left(\frac{\sigma}{x}\right)^{\log_4 5} = 1 - y$$

$$\frac{\sigma}{x} = (1 - y)^{\frac{1}{\log_4 5}}$$

$$x = \frac{\sigma}{(1 - y)^{\frac{1}{\log_4 5}}},$$

and so

$$F_X^{-1}(x) = \frac{\sigma}{(1 - x)^{\frac{1}{\log_4 5}}}.$$

We'll now try this out in R, for $\sigma = \{1, 2, 5\}$.

Example 4.3: Solution

Find the inverse CDF.

If $y = 1 - (\sigma/x)^{\log_4 5}$, then

$$\begin{aligned}\left(\frac{\sigma}{x}\right)^{\log_4 5} &= 1 - y \\ \frac{\sigma}{x} &= (1 - y)^{\frac{1}{\log_4 5}} \\ x &= \frac{\sigma}{(1 - y)^{\frac{1}{\log_4 5}}},\end{aligned}$$

and so

$$F_X^{-1}(x) = \frac{\sigma}{(1 - x)^{\frac{1}{\log_4 5}}}.$$

We'll now try this out in R, for $\sigma = \{1, 2, 5\}$.

Example 4.3: Solution

Find the inverse CDF.

If $y = 1 - (\sigma/x)^{\log_4 5}$, then

$$\left(\frac{\sigma}{x}\right)^{\log_4 5} = 1 - y$$

$$\frac{\sigma}{x} = (1 - y)^{\frac{1}{\log_4 5}}$$

$$x = \frac{\sigma}{(1 - y)^{\frac{1}{\log_4 5}}},$$

and so

$$F_X^{-1}(x) = \frac{\sigma}{(1 - x)^{\frac{1}{\log_4 5}}}.$$

We'll now try this out in R, for $\sigma = \{1, 2, 5\}$.

Example 4.3: Solution

Find the inverse CDF.

If $y = 1 - (\sigma/x)^{\log_4 5}$, then

$$\begin{aligned}\left(\frac{\sigma}{x}\right)^{\log_4 5} &= 1 - y \\ \frac{\sigma}{x} &= (1 - y)^{\frac{1}{\log_4 5}} \\ x &= \frac{\sigma}{(1 - y)^{\frac{1}{\log_4 5}}},\end{aligned}$$

and so

$$F_X^{-1}(x) = \frac{\sigma}{(1 - x)^{\frac{1}{\log_4 5}}}.$$

We'll now try this out in R, for $\sigma = \{1, 2, 5\}$.

Example 4.3: Solution

Find the inverse CDF.

If $y = 1 - (\sigma/x)^{\log_4 5}$, then

$$\begin{aligned}\left(\frac{\sigma}{x}\right)^{\log_4 5} &= 1 - y \\ \frac{\sigma}{x} &= (1 - y)^{\frac{1}{\log_4 5}} \\ x &= \frac{\sigma}{(1 - y)^{\frac{1}{\log_4 5}}},\end{aligned}$$

and so

$$F_X^{-1}(x) = \frac{\sigma}{(1 - x)^{\frac{1}{\log_4 5}}}.$$

We'll now try this out in R, for $\sigma = \{1, 2, 5\}$.

Example 4.3: Solution

Find the inverse CDF.

If $y = 1 - (\sigma/x)^{\log_4 5}$, then

$$\begin{aligned}\left(\frac{\sigma}{x}\right)^{\log_4 5} &= 1 - y \\ \frac{\sigma}{x} &= (1 - y)^{\frac{1}{\log_4 5}} \\ x &= \frac{\sigma}{(1 - y)^{\frac{1}{\log_4 5}}},\end{aligned}$$

and so

$$F_X^{-1}(x) = \frac{\sigma}{(1 - x)^{\frac{1}{\log_4 5}}}.$$

We'll now try this out in R, for $\sigma = \{1, 2, 5\}$.

Example 4.3: Solution

Find the inverse CDF.

If $y = 1 - (\sigma/x)^{\log_4 5}$, then

$$\begin{aligned}\left(\frac{\sigma}{x}\right)^{\log_4 5} &= 1 - y \\ \frac{\sigma}{x} &= (1 - y)^{\frac{1}{\log_4 5}} \\ x &= \frac{\sigma}{(1 - y)^{\frac{1}{\log_4 5}}},\end{aligned}$$

and so

$$F_X^{-1}(x) = \frac{\sigma}{(1 - x)^{\frac{1}{\log_4 5}}}.$$

We'll now try this out in R, for $\sigma = \{1, 2, 5\}$.

Example 4.3: Solution

```
1 pareto.sim = function(n, sigma){  
2   U = runif(n)  
3   answer = sigma/(((1-U)^(1/(log(5, base=4))))))  
4   return(answer)}
```

We can then test this out:

```
test_sigma1 = pareto.sim(100, 1)  
test_sigma2 = pareto.sim(100, 2)  
test_sigma5 = pareto.sim(100, 5)  
  
par(mfrow=c(1,3))  
hist(test_sigma1, main="Sigma=1", col="darkgoldenrod1")  
hist(test_sigma2, main="Sigma=2", col="deepskyblue2")  
hist(test_sigma5, main="Sigma=5", col="yellowgreen")
```

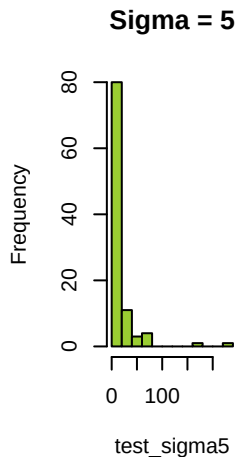
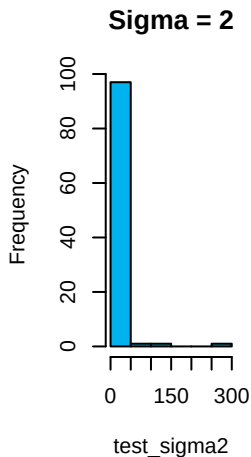
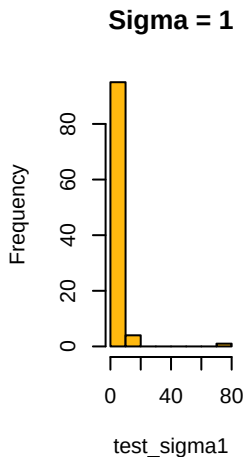
Example 4.3: Solution

```
1 pareto.sim = function(n, sigma){  
2   U = runif(n)  
3   answer = sigma/(((1-U)^(1/(log(5, base=4))))))  
4   return(answer)}
```

We can then test this out:

```
1 test_sigma1 = pareto.sim(100, 1)  
2 test_sigma2 = pareto.sim(100, 2)  
3 test_sigma5 = pareto.sim(100, 5)  
4  
5 par(mfrow=c(1,3))  
6 hist(test_sigma1, main="Sigma=1", col="darkgoldenrod1")  
7 hist(test_sigma2, main="Sigma=2", col="deepskyblue2")  
8 hist(test_sigma5, main="Sigma=5", col="yellowgreen")
```

Example 4.3: Solution



Example 4.4

Example

The **Weibull distribution** is often used in survival data analysis to model lifetimes (e.g. the lifetimes of patients receiving a particular drug).

If $X \sim \text{Weibull}(\lambda, \kappa)$, then it has cumulative distribution function

$$F_X(x; \lambda, \kappa) = 1 - e^{-(x/\lambda)^\kappa}, \quad x \geq 0, \lambda, \kappa > 0.$$

Write a function in R to produce realisations from this distribution, and test your function out.

(A recent study suggested that, for a particular cancer drug, $\lambda = 2$ and $\kappa = 1.5$ for survival times in months after starting this drug).

Example 4.4

Example

The **Weibull distribution** is often used in survival data analysis to model lifetimes (e.g. the lifetimes of patients receiving a particular drug).

If $X \sim \text{Weibull}(\lambda, \kappa)$, then it has cumulative distribution function

$$F_X(x; \lambda, \kappa) = 1 - e^{-(x/\lambda)^\kappa}, \quad x \geq 0, \lambda, \kappa > 0.$$

Write a function in R to produce realisations from this distribution, and test your function out.

(A recent study suggested that, for a particular cancer drug, $\lambda = 2$ and $\kappa = 1.5$ for survival times in months after starting this drug).

Example 4.4: Solution

First, find the inverse CDF.

If $y = 1 - e^{-(x/\lambda)^\kappa}$, then

$$e^{-(x/\lambda)^\kappa} = 1 - y$$

$$-(x/\lambda)^\kappa = \log(1 - y)$$

$$x/\lambda = [-\log(1 - y)]^{1/\kappa}$$

$$x = \lambda [-\log(1 - y)]^{1/\kappa}$$

and so

$$F_X^{-1}(x) = \lambda [-\log(1 - x)]^{1/\kappa}$$

Example 4.4: Solution

First, find the inverse CDF.

If $y = 1 - e^{-(x/\lambda)^\kappa}$, then

$$e^{-(x/\lambda)^\kappa} = 1 - y$$

$$-(x/\lambda)^\kappa = \log(1 - y)$$

$$x/\lambda = [-\log(1 - y)]^{1/\kappa}$$

$$x = \lambda [-\log(1 - y)]^{1/\kappa}$$

and so

$$F_X^{-1}(x) = \lambda [-\log(1 - x)]^{1/\kappa}$$

Example 4.4: Solution

First, find the inverse CDF.

If $y = 1 - e^{-(x/\lambda)^\kappa}$, then

$$e^{-(x/\lambda)^\kappa} = 1 - y$$

$$-(x/\lambda)^\kappa = \log(1 - y)$$

$$x/\lambda = [-\log(1 - y)]^{1/\kappa}$$

$$x = \lambda [-\log(1 - y)]^{1/\kappa}$$

and so

$$F_X^{-1}(x) = \lambda [-\log(1 - x)]^{1/\kappa}$$

Example 4.4: Solution

First, find the inverse CDF.

If $y = 1 - e^{-(x/\lambda)^\kappa}$, then

$$e^{-(x/\lambda)^\kappa} = 1 - y$$

$$-(x/\lambda)^\kappa = \log(1 - y)$$

$$x/\lambda = [-\log(1 - y)]^{1/\kappa}$$

$$x = \lambda [-\log(1 - y)]^{1/\kappa}$$

and so

$$F_X^{-1}(x) = \lambda [-\log(1 - x)]^{1/\kappa}$$

Example 4.4: Solution

First, find the inverse CDF.

If $y = 1 - e^{-(x/\lambda)^\kappa}$, then

$$e^{-(x/\lambda)^\kappa} = 1 - y$$

$$-(x/\lambda)^\kappa = \log(1 - y)$$

$$x/\lambda = [-\log(1 - y)]^{1/\kappa}$$

$$x = \lambda [-\log(1 - y)]^{1/\kappa}$$

and so

$$F_X^{-1}(x) = \lambda [-\log(1 - x)]^{1/\kappa}$$

Example 4.4: Solution

First, find the inverse CDF.

If $y = 1 - e^{-(x/\lambda)^\kappa}$, then

$$e^{-(x/\lambda)^\kappa} = 1 - y$$

$$-(x/\lambda)^\kappa = \log(1 - y)$$

$$x/\lambda = [-\log(1 - y)]^{1/\kappa}$$

$$x = \lambda [-\log(1 - y)]^{1/\kappa}$$

and so

$$F_X^{-1}(x) = \lambda [-\log(1 - x)]^{1/\kappa}$$

Example 4.4: Solution

First, find the inverse CDF.

If $y = 1 - e^{-(x/\lambda)^\kappa}$, then

$$e^{-(x/\lambda)^\kappa} = 1 - y$$

$$-(x/\lambda)^\kappa = \log(1 - y)$$

$$x/\lambda = [-\log(1 - y)]^{1/\kappa}$$

$$x = \lambda [-\log(1 - y)]^{1/\kappa}$$

and so

$$F_X^{-1}(x) = \lambda [-\log(1 - x)]^{1/\kappa}$$

Example 4.4: Solution

In R:

```
1 weibull.sim = function(n, lambda, kappa){  
2   U = runif(n)  
3   answer = lambda*((-log(1-U))^(1/kappa))  
4   return(answer)}
```

Then, to test it out using the provided parameters:

```
test = weibull.sim(1000, 2, 1.5)  
summary(test)  
      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.     
0.009993 0.879331 1.590474 1.779364 2.451390 6.215579  
  
hist(test, main='Weibull simulations', freq=FALSE, col=  
      'mediumvioletred')
```

Example 4.4: Solution

In R:

```
1 weibull.sim = function(n, lambda, kappa){  
2   U = runif(n)  
3   answer = lambda*((-log(1-U))^(1/kappa))  
4   return(answer)}
```

Then, to test it out using the provided parameters:

```
test = weibull.sim(1000, 2, 1.5)  
summary(test)  
      Min.   1st Qu.   Median     Mean   3rd Qu.    Max.     
0.009993 0.879331 1.590474 1.779364 2.451390 6.215579  
  
hist(test, main='Weibull simulations', freq=FALSE, col=  
      'mediumslateblue')
```

Example 4.4: Solution

In R:

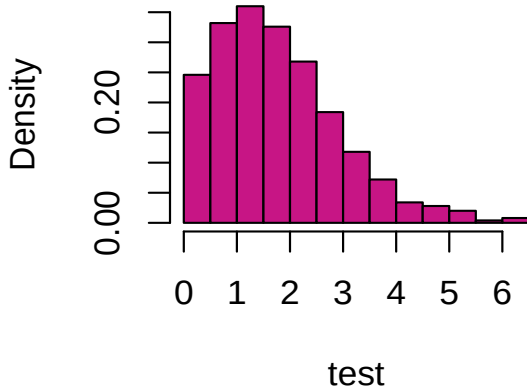
```
1 weibull.sim = function(n, lambda, kappa){  
2   U = runif(n)  
3   answer = lambda*((-log(1-U))^(1/kappa))  
4   return(answer)}
```

Then, to test it out using the provided parameters:

```
1 test = weibull.sim(1000, 2, 1.5)  
2 summary(test)  
3      Min.   1st Qu.   Median     Mean   3rd Qu.   Max.  
4 0.009993 0.879331 1.590474 1.779364 2.451390 6.215579  
5  
6 hist(test, main='Weibull simulations', freq=FALSE, col=  
   'mediumvioletred')
```

Example 4.4: Solution

Weibull simulations



Example 4.4: Solution

We can check our work, as there is a built-in function in R to simulate from a Weibull!

We must note that $\lambda = 2$ is the scale parameter and $\kappa = 1.5$ is the shape parameter:

```
test2 = rweibull(1000, scale=2, shape=1.5)
summary(test2)
```

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
	0.02692	0.86352	1.56831	1.78488	2.50828	6.57937

Note that these are similar to the summaries from our own function. However, a more conclusive check might be to overlay the PDF.

Example 4.4: Solution

We can check our work, as there is a built-in function in R to simulate from a Weibull!

We must note that $\lambda = 2$ is the **scale parameter** and $\kappa = 1.5$ is the **shape parameter**:

```
test2 = rweibull(1000, scale=2, shape=1.5)
summary(test2)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.02692	0.86352	1.56831	1.78488	2.50828	6.57937

Note that these are similar to the summaries from our own function. However, a more conclusive check might be to overlay the PDF.

Example 4.4: Solution

We can check our work, as there is a built-in function in R to simulate from a Weibull!

We must note that $\lambda = 2$ is the **scale parameter** and $\kappa = 1.5$ is the **shape parameter**:

```
1 test2 = rweibull(1000, scale=2, shape=1.5)
2 summary(test2)
3      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
4 0.02692 0.86352 1.56831 1.78488 2.50828 6.57937
```

Note that these are similar to the summaries from our own function. However, a more conclusive check might be to overlay the PDF.

Example 4.4: Solution

We can check our work, as there is a built-in function in R to simulate from a Weibull!

We must note that $\lambda = 2$ is the **scale parameter** and $\kappa = 1.5$ is the **shape parameter**:

```
1 test2 = rweibull(1000, scale=2, shape=1.5)
2 summary(test2)
3      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
4 0.02692 0.86352 1.56831 1.78488 2.50828 6.57937
```

Note that these are **similar** to the summaries from our own function. However, a more conclusive check might be to overlay the PDF.

Example 4.4: Solution

Recall that to get from the CDF to the PDF, we differentiate; thus,

$$\begin{aligned}f_X(x) &= \frac{d}{dx} F_X(x) \\&= \frac{d}{dx} \left(1 - e^{-(x/\lambda)^\kappa} \right).\end{aligned}$$

After careful (and multiple!) use of the chain rule, you should find that

$$f_X(x) = \frac{\kappa}{\lambda} \left(\frac{x}{\lambda} \right)^{\kappa-1} e^{-(x/\lambda)^\kappa}.$$

Example 4.4: Solution

Recall that to get from the CDF to the PDF, we differentiate; thus,

$$\begin{aligned}f_X(x) &= \frac{d}{dx} F_X(x) \\&= \frac{d}{dx} \left(1 - e^{-(x/\lambda)^\kappa} \right).\end{aligned}$$

After careful (and multiple!) use of the chain rule, you should find that

$$f_X(x) = \frac{\kappa}{\lambda} \left(\frac{x}{\lambda} \right)^{\kappa-1} e^{-(x/\lambda)^\kappa}.$$

Example 4.4: Solution

So, we can add this to our plot in R:

```
1 weibull.pdf = function(x, lambda, kappa){  
2   answer = (kappa/lambda)*((x/lambda)^(kappa-1))*exp  
3     (-((x/lambda)^kappa))  
4   return(answer)}  
5 x = seq(0, 7, 0.001)  
6 y = weibull.pdf(x, 2, 1.5)  
7 lines(x, y, type='l', lwd=2, lty=2)
```

Example 4.4: Solution

Weibull simulations

