

Lecture: 2pm, Monday 11th November

MAS2906: Computational Probability and Statistics with R

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

Semester 1, 2024/25

Where we're at...

■ Weeks 1-3: R Refresher

- The basics – including syntax, vectors and dataframes
- Data summaries – graphical and numerical
- Functions, for loops and if statements, with four case studies

■ Weeks 4-5: Computation from probability models

- Standard distributions: Binomial, Poisson, geometric; Normal, exponential, uniform
- More complex models: An arrival times model; a Normal mixture model

■ Week 6: Simulation using the Inverse CDF method

- Practical motivation for how it works
- Formal mathematical proof
- Lots of examples!

Where we're at...

■ Weeks 1-3: R Refresher

- The basics – including syntax, vectors and dataframes
- Data summaries – graphical and numerical
- Functions, for loops and if statements, with four case studies

■ Weeks 4-5: Computation from probability models

- Standard distributions: Binomial, Poisson, geometric; Normal, exponential, uniform
- More complex models: An arrival times model; a Normal mixture model

■ Week 6: Simulation using the Inverse CDF method

- Practical motivation for how it works
- Formal mathematical proof
- Lots of examples!

Where we're at...

- Weeks 1-3: R Refresher
 - The basics – including syntax, vectors and dataframes
 - Data summaries – graphical and numerical
 - Functions, for loops and if statements, with four case studies
- Weeks 4-5: Computation from probability models
 - Standard distributions: Binomial, Poisson, geometric; Normal, exponential, uniform
 - More complex models: An arrival times model; a Normal mixture model
- Week 6: Simulation using the Inverse CDF method
 - Practical motivation for how it works
 - Formal mathematical proof
 - **Lots of examples!**

This week

- TWO HOUR lecture (and every week for the rest of term)
 - We will complete Chapter 4
 - We will review your last assignment (Numbas) and then start the next assignment (mini project)
 - Computer practical back to the usual Tuesday/Wednesday slots): We will make a start on the mini project

This week

- TWO HOUR lecture (and every week for the rest of term)
- We will complete Chapter 4
 - We will review your last assignment (Numbas) and then start the next assignment (mini project)
 - Computer practical back to the usual Tuesday/Wednesday slots): We will make a start on the mini project

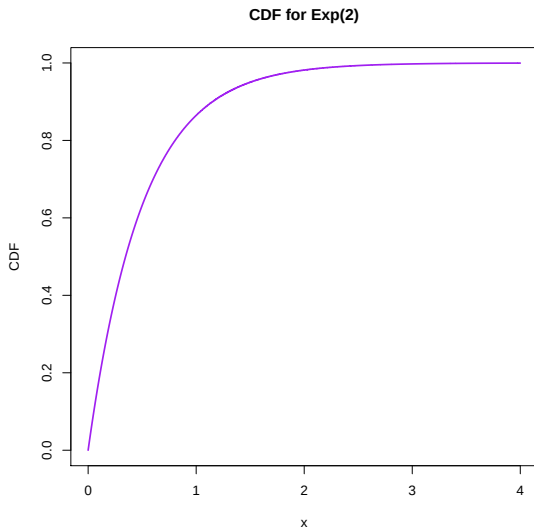
This week

- TWO HOUR lecture (and every week for the rest of term)
- We will complete Chapter 4
- We will review your last assignment (Numbas) and then start the next assignment (mini project)
- Computer practical back to the usual Tuesday/Wednesday slots): We will make a start on the mini project

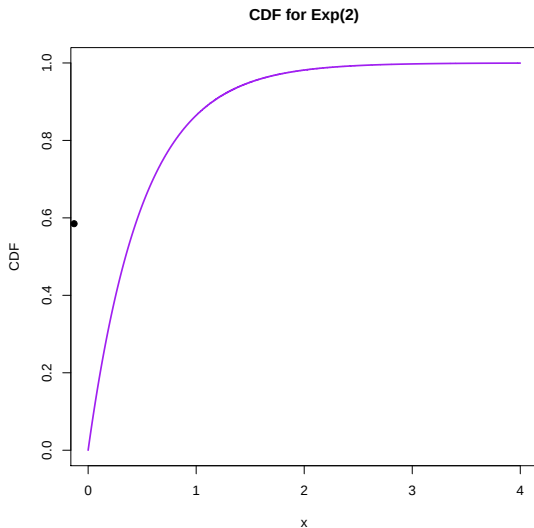
This week

- TWO HOUR lecture (and every week for the rest of term)
- We will complete Chapter 4
- We will review your last assignment (Numbas) and then start the next assignment (mini project)
- Computer practical back to the usual Tuesday/Wednesday slots): We will make a start on the mini project

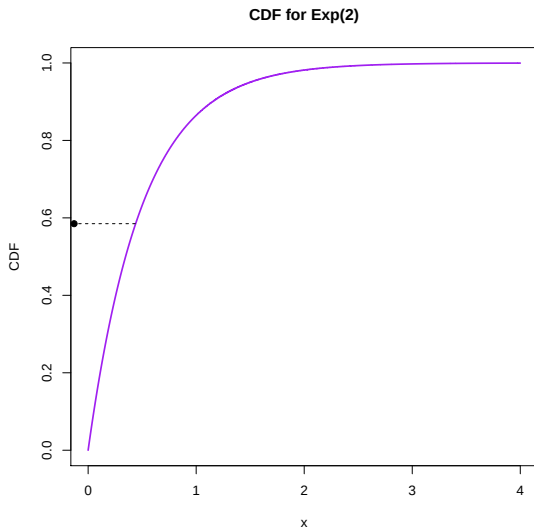
4 The Inverse CDF Method



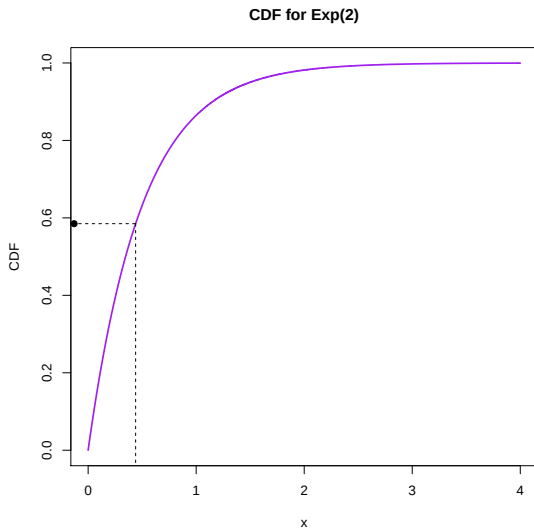
4 The Inverse CDF Method



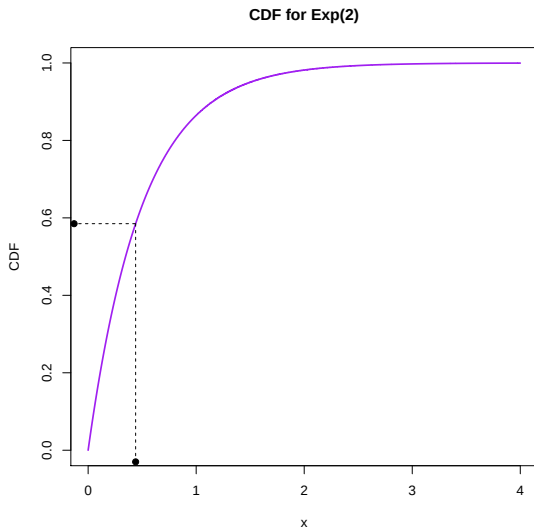
4 The Inverse CDF Method



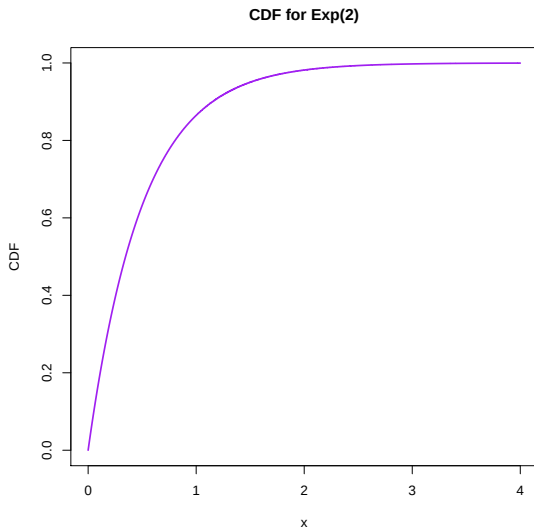
4 The Inverse CDF Method



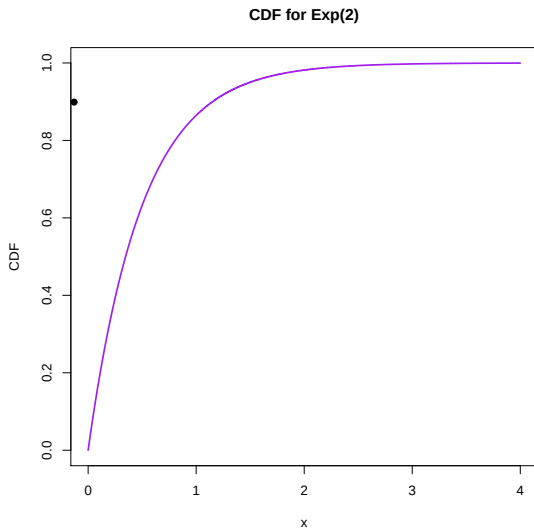
4 The Inverse CDF Method



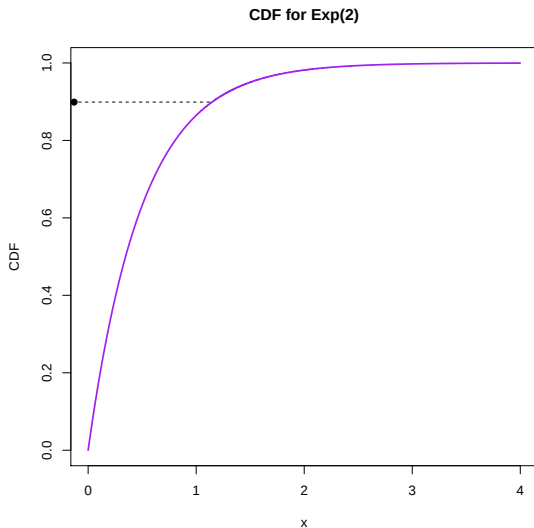
4 The Inverse CDF Method



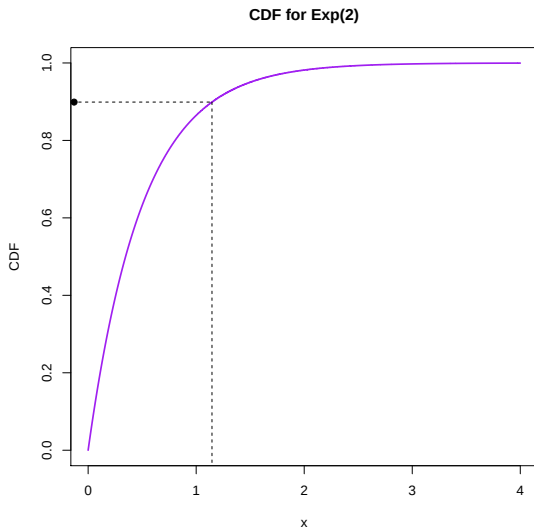
4 The Inverse CDF Method



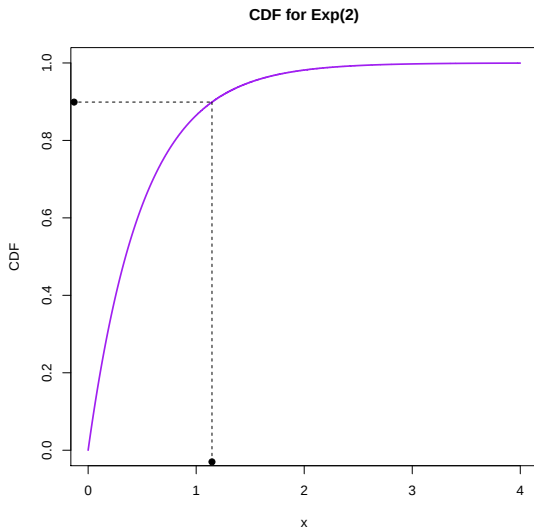
4 The Inverse CDF Method



4 The Inverse CDF Method



4 The Inverse CDF Method



Recap Quiz

1. Write down R code that simulates 100 observations from the $U(0, 1)$ distribution and stores the output in the vector `U`.
2. Suppose X has CDF

$$F_X(x) = 1 - \frac{1}{x}, \quad x \geq 1.$$

Find $F_X^{-1}(x)$.

3. Write an R function called `inv.cdf` to generate values from X .
4. Produce a histogram of your realizations of X , coloured yellow and with a suitable title.

Recap Quiz

1. Write down R code that simulates 100 observations from the $U(0, 1)$ distribution and stores the output in the vector `U`.
2. Suppose X has CDF

$$F_X(x) = 1 - \frac{1}{x}, \quad x \geq 1.$$

Find $F_X^{-1}(x)$.

3. Write an R function called `inv.cdf` to generate values from X .
4. Produce a histogram of your realizations of X , coloured yellow and with a suitable title.

Recap Quiz

1. Write down R code that simulates 100 observations from the $U(0, 1)$ distribution and stores the output in the vector `U`.
2. Suppose X has CDF

$$F_X(x) = 1 - \frac{1}{x}, \quad x \geq 1.$$

Find $F_X^{-1}(x)$.

3. Write an R function called `inv.cdf` to generate values from X .
4. Produce a histogram of your realizations of X , coloured yellow and with a suitable title.

Recap Quiz

1. Write down R code that simulates 100 observations from the $U(0, 1)$ distribution and stores the output in the vector `U`.
2. Suppose X has CDF

$$F_X(x) = 1 - \frac{1}{x}, \quad x \geq 1.$$

Find $F_X^{-1}(x)$.

3. Write an R function called `inv.cdf` to generate values from X .
4. Produce a histogram of your realizations of X , coloured yellow and with a suitable title.

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:

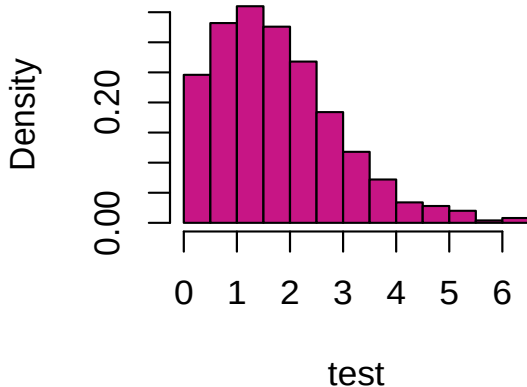
```
1 weibull.sim = function(n, lambda, kappa){  
2   U = runif(n)  
3   answer = lambda*((-log(1-U))^(1/kappa))  
4   return(answer)}
```

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

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

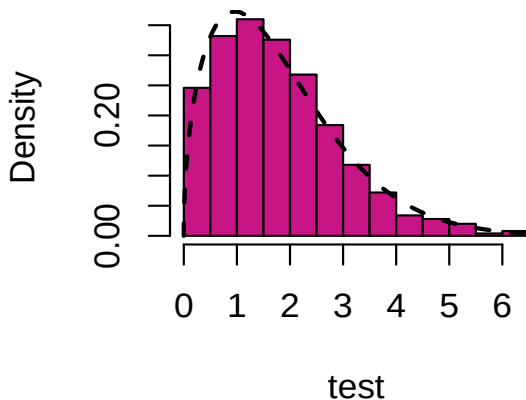
Example 4.4: Solution

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



Example 4.5 (pages 34–35)

- Patients arriving at a doctor's surgery have appointment times 15 minutes apart.
- The random variable L measures how late a patient is for his/her appointment, with negative values denoting early arrivals.

The PDF of L is

$$f_L(\ell) = \begin{cases} 0.0025(\ell + 20), & -20 \leq \ell \leq 0, \\ 0.05e^{-\ell/10}, & \ell > 0. \end{cases}$$

Example 4.5: Questions

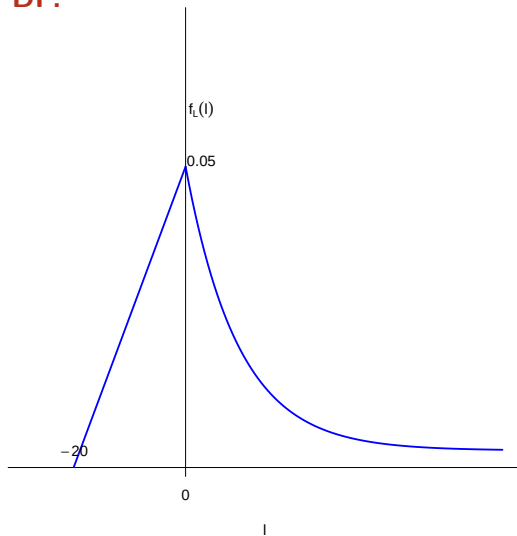
- a. Sketch $f_L(\ell)$.
- b. Show that the CDF of L is

$$F_L(\ell) = \begin{cases} 0, & \ell < -20, \\ 0.00125(\ell + 20)^2, & -20 \leq \ell \leq 0, \\ 1 - \frac{e^{-\ell/10}}{2}, & \ell > 0. \end{cases}$$

- c. Suppose that 0.205 and 0.713 are two realized values of $U \sim U(0, 1)$. Show that the corresponding realized values of L are -7.19 and 5.55 under the inverse CDF simulation method.

Example 4.5: Solution

First, the PDF:



Example 4.5: Solution

Now to find the CDF: L lies between -20 and $+\infty$, so
 $F_L(\ell) = 0$ when $\ell < -20$.

For $-20 \leq \ell \leq 0$:

$$\begin{aligned}F_L(\ell) &= \int_{-20}^{\ell} f_L(s) ds \\&= \int_{-20}^{\ell} 0.0025(s+20) ds \\&= \left[\frac{0.0025}{2} (s+20)^2 \right]_{-20}^{\ell} \\&= 0.00125(\ell+20)^2 \quad \text{for } -20 \leq \ell \leq 0.\end{aligned}$$

Also, $F_L(0) = 0.5$.

Example 4.5: Solution

Now to find the CDF: L lies between -20 and $+\infty$, so $F_L(\ell) = 0$ when $\ell < -20$.

For $-20 \leq \ell \leq 0$:

$$\begin{aligned} F_L(\ell) &= \int_{-20}^{\ell} f_L(s) ds \\ &= \int_{-20}^{\ell} 0.0025(s+20) ds \\ &= \left[\frac{0.0025}{2} (s+20)^2 \right]_{-20}^{\ell} \\ &= 0.00125(\ell+20)^2 \quad \text{for } -20 \leq \ell \leq 0. \end{aligned}$$

Also, $F_L(0) = 0.5$.

Example 4.5: Solution

Now to find the CDF: L lies between -20 and $+\infty$, so $F_L(\ell) = 0$ when $\ell < -20$.

For $-20 \leq \ell \leq 0$:

$$\begin{aligned} F_L(\ell) &= \int_{-20}^{\ell} f_L(s) ds \\ &= \int_{-20}^{\ell} 0.0025(s+20) ds \\ &= \left[\frac{0.0025}{2} (s+20)^2 \right]_{-20}^{\ell} \\ &= 0.00125(\ell+20)^2 \quad \text{for } -20 \leq \ell \leq 0. \end{aligned}$$

Also, $F_L(0) = 0.5$.

Example 4.5: Solution

Now to find the CDF: L lies between -20 and $+\infty$, so $F_L(\ell) = 0$ when $\ell < -20$.

For $-20 \leq \ell \leq 0$:

$$\begin{aligned} F_L(\ell) &= \int_{-20}^{\ell} f_L(s) ds \\ &= \int_{-20}^{\ell} 0.0025(s+20) ds \\ &= \left[\frac{0.0025}{2} (s+20)^2 \right]_{-20}^{\ell} \\ &= 0.00125(\ell+20)^2 \quad \text{for } -20 \leq \ell \leq 0. \end{aligned}$$

Also, $F_L(0) = 0.5$.

Example 4.5: Solution

Now to find the CDF: L lies between -20 and $+\infty$, so $F_L(\ell) = 0$ when $\ell < -20$.

For $-20 \leq \ell \leq 0$:

$$\begin{aligned} F_L(\ell) &= \int_{-20}^{\ell} f_L(s) ds \\ &= \int_{-20}^{\ell} 0.0025(s+20) ds \end{aligned}$$

$$= \left[\frac{0.0025}{2} (s+20)^2 \right]_{-20}^{\ell}$$

$$= 0.00125(\ell+20)^2 \quad \text{for } -20 \leq \ell \leq 0.$$

Also, $F(0) = 0.5$.

Example 4.5: Solution

Now to find the CDF: L lies between -20 and $+\infty$, so $F_L(\ell) = 0$ when $\ell < -20$.

For $-20 \leq \ell \leq 0$:

$$\begin{aligned}F_L(\ell) &= \int_{-20}^{\ell} f_L(s) ds \\&= \int_{-20}^{\ell} 0.0025(s + 20) ds \\&= \left[\frac{0.0025}{2} (s + 20)^2 \right]_{-20}^{\ell}\end{aligned}$$

$$= 0.00125(\ell + 20)^2 \quad \text{for } -20 \leq \ell \leq 0.$$

Also, $F(0) = 0.5$.

Example 4.5: Solution

Now to find the CDF: L lies between -20 and $+\infty$, so $F_L(\ell) = 0$ when $\ell < -20$.

For $-20 \leq \ell \leq 0$:

$$\begin{aligned}F_L(\ell) &= \int_{-20}^{\ell} f_L(s) ds \\&= \int_{-20}^{\ell} 0.0025(s + 20) ds \\&= \left[\frac{0.0025}{2} (s + 20)^2 \right]_{-20}^{\ell} \\&= 0.00125(\ell + 20)^2 \quad \text{for } -20 \leq \ell \leq 0.\end{aligned}$$

Also, $F_L(0) = 0.5$.

Example 4.5: Solution

Now

$$F_L(\ell) = \int_{-20}^0 f_L(s)ds + \int_0^\ell f_L(s)ds \quad \text{for } \ell > 0$$

$$= 0.5 + \int_0^\ell 0.05e^{-s/10} ds$$

$$= 0.5 + \left[0.05 \times -10e^{-s/10} \right]_0^\ell$$

$$= 0.5 + 0.5 - 0.5e^{-\ell/10}$$

$$= 1 - \frac{e^{-\ell/10}}{2}, \quad \text{as required.}$$

Example 4.5: Solution

Now

$$F_L(\ell) = \int_{-20}^0 f_L(s)ds + \int_0^\ell f_L(s)ds \quad \text{for } \ell > 0$$

$$= 0.5 + \int_0^\ell 0.05e^{-s/10}ds$$

$$= 0.5 + \left[0.05 \times -10e^{-s/10} \right]_0^\ell$$

$$= 0.5 + 0.5 - 0.5e^{-\ell/10}$$

$$= 1 - \frac{e^{-\ell/10}}{2}, \quad \text{as required.}$$

Example 4.5: Solution

Now

$$F_L(\ell) = \int_{-20}^0 f_L(s)ds + \int_0^{\ell} f_L(s)ds \quad \text{for } \ell > 0$$

$$= 0.5 + \int_0^{\ell} 0.05e^{-s/10}ds$$

$$= 0.5 + \left[0.05 \times -10e^{-s/10} \right]_0^{\ell}$$

$$= 0.5 + 0.5 - 0.5e^{-\ell/10}$$

$$= 1 - \frac{e^{-\ell/10}}{2}, \quad \text{as required.}$$

Example 4.5: Solution

Now

$$\begin{aligned}F_L(\ell) &= \int_{-20}^0 f_L(s)ds + \int_0^{\ell} f_L(s)ds \quad \text{for } \ell > 0 \\&= 0.5 + \int_0^{\ell} 0.05e^{-s/10}ds \\&= 0.5 + \left[0.05 \times -10e^{-s/10}\right]_0^{\ell} \\&= 0.5 + 0.5 - 0.5e^{-\ell/10} \\&= 1 - \frac{e^{-\ell/10}}{2}, \quad \text{as required.}\end{aligned}$$

Example 4.5: Solution

Now

$$\begin{aligned}F_L(\ell) &= \int_{-20}^0 f_L(s)ds + \int_0^{\ell} f_L(s)ds \quad \text{for } \ell > 0 \\&= 0.5 + \int_0^{\ell} 0.05e^{-s/10}ds \\&= 0.5 + \left[0.05 \times -10e^{-s/10}\right]_0^{\ell} \\&= 0.5 + 0.5 - 0.5e^{-\ell/10} \\&= 1 - \frac{e^{-\ell/10}}{2}, \quad \text{as required.}\end{aligned}$$

Example 4.5: Solution

Now for some realized values:

$u = 0.205 < 0.5$, so realized value ℓ lies between -20 and 0 . Thus

$$F_L(\ell) = 0.00125(\ell + 20)^2.$$

Find the inverse, and evaluate at $u = 0.205$:

$$y = 0.00125(\ell + 20)^2, \quad \text{so}$$

$$\ell = \pm \sqrt{\frac{y}{0.00125}} - 20,$$

giving

$$\ell = \sqrt{\frac{0.205}{0.00125}} - 20 = -7.2.$$

Example 4.5: Solution

Now for some realized values:

$u = 0.205 < 0.5$, so realized value ℓ lies between -20 and 0 . Thus

$$F_L(\ell) = 0.00125(\ell + 20)^2.$$

Find the inverse, and evaluate at $u = 0.205$:

$$y = 0.00125(\ell + 20)^2, \quad \text{so}$$

$$\ell = \pm \sqrt{\frac{y}{0.00125}} - 20,$$

giving

$$\ell = \sqrt{\frac{0.205}{0.00125}} - 20 = -7.2.$$

Example 4.5: Solution

Now for some realized values:

$u = 0.205 < 0.5$, so realized value ℓ lies between -20 and 0 . Thus

$$F_L(\ell) = 0.00125(\ell + 20)^2.$$

Find the inverse, and evaluate at $u = 0.205$:

$$y = 0.00125(\ell + 20)^2, \quad \text{so}$$

$$\ell = \pm \sqrt{\frac{y}{0.00125}} - 20,$$

giving

$$\ell = \sqrt{\frac{0.205}{0.00125}} - 20 = -7.2.$$

Example 4.5: Solution

Now for some realized values:

$u = 0.205 < 0.5$, so realized value ℓ lies between -20 and 0 . Thus

$$F_L(\ell) = 0.00125(\ell + 20)^2.$$

Find the inverse, and evaluate at $u = 0.205$:

$$y = 0.00125(\ell + 20)^2, \quad \text{so}$$

$$\ell = \pm \sqrt{\frac{y}{0.00125}} - 20,$$

giving

$$\ell = \sqrt{\frac{0.205}{0.00125}} - 20 = -7.2.$$

Example 4.5: Solution

Now for some realized values:

$u = 0.205 < 0.5$, so realized value ℓ lies between -20 and 0 . Thus

$$F_L(\ell) = 0.00125(\ell + 20)^2.$$

Find the inverse, and evaluate at $u = 0.205$:

$$y = 0.00125(\ell + 20)^2, \quad \text{so}$$

$$\ell = \pm \sqrt{\frac{y}{0.00125}} - 20,$$

giving

$$\ell = \sqrt{\frac{0.205}{0.00125}} - 20 = -7.2.$$

Example 4.5: Solution

Now for some realized values:

$u = 0.205 < 0.5$, so realized value ℓ lies between -20 and 0 . Thus

$$F_L(\ell) = 0.00125(\ell + 20)^2.$$

Find the inverse, and evaluate at $u = 0.205$:

$$y = 0.00125(\ell + 20)^2, \quad \text{so}$$

$$\ell = \pm \sqrt{\frac{y}{0.00125}} - 20,$$

giving

$$\ell = \sqrt{\frac{0.205}{0.00125}} - 20 = -7.2.$$

Example 4.5: Solution

Now for some realized values:

$u = 0.205 < 0.5$, so realized value ℓ lies between -20 and 0 . Thus

$$F_L(\ell) = 0.00125(\ell + 20)^2.$$

Find the inverse, and evaluate at $u = 0.205$:

$$y = 0.00125(\ell + 20)^2, \quad \text{so}$$

$$\ell = \pm \sqrt{\frac{y}{0.00125}} - 20,$$

giving

$$\ell = \sqrt{\frac{0.205}{0.00125}} - 20 = -7.2.$$

Example 4.5: Solution

Similarly, $u = 0.713 > 0.5$ so realized value > 0 . Thus

$$F_L(\ell) = 1 - \frac{e^{-\ell/10}}{2}.$$

Find the inverse, and evaluate at $u = 0.713$:

$$\begin{aligned} y &= 1 - \frac{e^{-\ell/10}}{2} \quad \text{so} \\ \ell &= -10 \log(2 - 2y), \end{aligned}$$

giving

$$\ell = -10 \log(2 - 2 \times 0.713) = 5.55.$$

Example 4.5: Solution

Similarly, $u = 0.713 > 0.5$ so realized value > 0 . Thus

$$F_L(\ell) = 1 - \frac{e^{-\ell/10}}{2}.$$

Find the inverse, and evaluate at $u = 0.713$:

$$\begin{aligned}y &= 1 - \frac{e^{-\ell/10}}{2} \quad \text{so} \\ \ell &= -10 \log(2 - 2y),\end{aligned}$$

giving

$$\ell = -10 \log(2 - 2 \times 0.713) = 5.55.$$

Example 4.5: Solution

Similarly, $u = 0.713 > 0.5$ so realized value > 0 . Thus

$$F_L(\ell) = 1 - \frac{e^{-\ell/10}}{2}.$$

Find the inverse, and evaluate at $u = 0.713$:

$$\begin{aligned} y &= 1 - \frac{e^{-\ell/10}}{2} \quad \text{so} \\ \ell &= -10 \log(2 - 2y), \end{aligned}$$

giving

$$\ell = -10 \log(2 - 2 \times 0.713) = 5.55.$$

Example 4.5: Solution

Similarly, $u = 0.713 > 0.5$ so realized value > 0 . Thus

$$F_L(\ell) = 1 - \frac{e^{-\ell/10}}{2}.$$

Find the inverse, and evaluate at $u = 0.713$:

$$y = 1 - \frac{e^{-\ell/10}}{2} \quad \text{so}$$

$$\ell = -10 \log(2 - 2y),$$

giving

$$\ell = -10 \log(2 - 2 \times 0.713) = 5.55.$$

Example 4.5: Solution

Similarly, $u = 0.713 > 0.5$ so realized value > 0 . Thus

$$F_L(\ell) = 1 - \frac{e^{-\ell/10}}{2}.$$

Find the inverse, and evaluate at $u = 0.713$:

$$\begin{aligned} y &= 1 - \frac{e^{-\ell/10}}{2} \quad \text{so} \\ \ell &= -10 \log(2 - 2y), \end{aligned}$$

giving

$$\ell = -10 \log(2 - 2 \times 0.713) = 5.55.$$

Chapter 5

Transformations

5.1 Transforming a normal random variable

- Consider how we might model a **continuous positive** random quantity e.g. blood pressure or height
- The Normal distribution is not suitable!

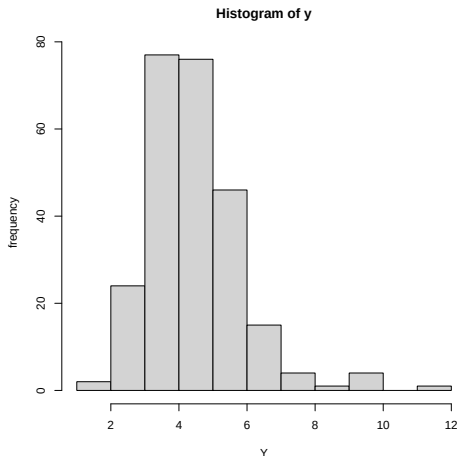
One way to obtain a positive random variable is to define

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

Then Y is a random variable which is a **transformation** of X .

5.1 Simulation to explore transformation

```
1 y = exp(rnorm(250, mean = 1.5, sd = 0.3))  
2 hist(y, xlab = 'Y', ylab = 'frequency')
```



5.1 Calculating PDFs

- This looks like a useful distribution – what is its PDF?
- $Y = \exp(X)$ so $X = \log(Y)$
- So $f_Y(y) = f_X(\log(y))$ is a good guess, but is **NOT** the right answer!

Example

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

Example 5.1: Solution

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^3 \leq y) \\&= \Pr(X \leq y^{1/3}) \\&= F_X(y^{1/3}).\end{aligned}$$

Therefore

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} F_X(y^{1/3}) \\&= f_X(y^{1/3}) \times \frac{1}{3} y^{-2/3} \quad \text{by chain rule} \\&= \frac{1}{3} y^{-2/3}, \quad 0 \leq y \leq 1 \text{ (and zero otherwise).}\end{aligned}$$

Example 5.1: Solution

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^3 \leq y) \\&= \Pr(X \leq y^{1/3}) \\&= F_X(y^{1/3}).\end{aligned}$$

Therefore

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} F_X(y^{1/3}) \\&= f_X(y^{1/3}) \times \frac{1}{3} y^{-2/3} \quad \text{by chain rule} \\&= \frac{1}{3} y^{-2/3}, \quad 0 \leq y \leq 1 \text{ (and zero otherwise).}\end{aligned}$$

Example 5.1: Solution

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^3 \leq y) \\&= \Pr(X \leq y^{1/3}) \\&= F_X(y^{1/3}).\end{aligned}$$

Therefore

$$\begin{aligned}f_Y(y) &= \frac{d}{dy}F_X(y^{1/3}) \\&= f_X(y^{1/3}) \times \frac{1}{3}y^{-2/3} \quad \text{by chain rule} \\&= \frac{1}{3}y^{-2/3}, \quad 0 \leq y \leq 1 \text{ (and zero otherwise).}\end{aligned}$$

Example 5.1: Solution

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^3 \leq y) \\&= \Pr(X \leq y^{1/3}) \\&= F_X(y^{1/3}).\end{aligned}$$

Therefore

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} F_X(y^{1/3}) \\&= f_X(y^{1/3}) \times \frac{1}{3} y^{-2/3} \quad \text{by chain rule} \\&= \frac{1}{3} y^{-2/3}, \quad 0 \leq y \leq 1 \text{ (and zero otherwise).}\end{aligned}$$

Example 5.1: Solution

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^3 \leq y) \\&= \Pr(X \leq y^{1/3}) \\&= F_X(y^{1/3}).\end{aligned}$$

Therefore

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} F_X(y^{1/3}) \\&= f_X(y^{1/3}) \times \frac{1}{3} y^{-2/3} \quad \text{by chain rule} \\&= \frac{1}{3} y^{-2/3}, \quad 0 \leq y \leq 1 \text{ (and zero otherwise).}\end{aligned}$$

Example 5.1: Solution

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^3 \leq y) \\&= \Pr(X \leq y^{1/3}) \\&= F_X(y^{1/3}).\end{aligned}$$

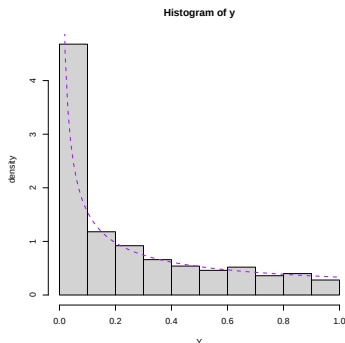
Therefore

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} F_X(y^{1/3}) \\&= f_X(y^{1/3}) \times \frac{1}{3} y^{-2/3} \quad \text{by chain rule} \\&= \frac{1}{3} y^{-2/3}, \quad 0 \leq y \leq 1 \text{ (and zero otherwise).}\end{aligned}$$

Example 5.1: Checking via simulation

To confirm this result, we can simulate Y :

```
1 y = (runif(500))^3
2 hist(y, xlab = 'Y', ylab = 'density', prob=TRUE)
3 curve((1/3)*(x^(-2/3)), from=0, to=1, col="purple", lty
      =2, add=TRUE)
```



Example 5.2

Example

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

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^2 \leq y) \\&= \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \Pr(X \leq -\sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \{1 - \Pr(X \leq \sqrt{y})\} \\&= 2 \cdot \Pr(X \leq \sqrt{y}) - 1 \\&= 2 \cdot F_X(\sqrt{y}) - 1.\end{aligned}$$

Example 5.2

Example

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

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^2 \leq y) \\&= \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \Pr(X \leq -\sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \{1 - \Pr(X \leq \sqrt{y})\} \\&= 2 \cdot \Pr(X \leq \sqrt{y}) - 1 \\&= 2 \cdot F_X(\sqrt{y}) - 1.\end{aligned}$$

Example 5.2

Example

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

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^2 \leq y) \\&= \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \Pr(X \leq -\sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \{1 - \Pr(X \leq \sqrt{y})\} \\&= 2 \cdot \Pr(X \leq \sqrt{y}) - 1 \\&= 2 \cdot F_X(\sqrt{y}) - 1.\end{aligned}$$

Example 5.2

Example

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

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^2 \leq y) \\&= \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \Pr(X \leq -\sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \{1 - \Pr(X \leq \sqrt{y})\} \\&= 2 \cdot \Pr(X \leq \sqrt{y}) - 1 \\&= 2 \cdot F_X(\sqrt{y}) - 1.\end{aligned}$$

Example 5.2

Example

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

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^2 \leq y) \\&= \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \Pr(X \leq -\sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \{1 - \Pr(X \leq \sqrt{y})\} \\&= 2 \cdot \Pr(X \leq \sqrt{y}) - 1 \\&= 2 \cdot F_X(\sqrt{y}) - 1.\end{aligned}$$

Example 5.2

Example

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

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^2 \leq y) \\&= \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \Pr(X \leq -\sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \{1 - \Pr(X \leq \sqrt{y})\} \\&= 2 \cdot \Pr(X \leq \sqrt{y}) - 1 \\&= 2 \cdot F_X(\sqrt{y}) - 1.\end{aligned}$$

Example 5.2

Example

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

$$\begin{aligned}F_Y(y) &= \Pr(Y \leq y) \\&= \Pr(X^2 \leq y) \\&= \Pr(-\sqrt{y} \leq X \leq \sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \Pr(X \leq -\sqrt{y}) \\&= \Pr(X \leq \sqrt{y}) - \{1 - \Pr(X \leq \sqrt{y})\} \\&= 2 \cdot \Pr(X \leq \sqrt{y}) - 1 \\&= 2 \cdot F_X(\sqrt{y}) - 1.\end{aligned}$$

Example 5.2

Then

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} \left(2 \cdot F_X(y^{1/2}) - 1 \right) \\&= 2f_X(y^{1/2}) \times \frac{1}{2}y^{-1/2} \quad \text{by chain rule} \\&= \frac{1}{\sqrt{y}} f_X(\sqrt{y}) \\&= \frac{1}{\sqrt{y}} \times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(\sqrt{y})^2 \right\} \\&= \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad \text{for } y \geq 0, \text{ zero otherwise.}\end{aligned}$$

This is the PDF for a **chi-squared distribution on 1 degree of freedom**.

Example 5.2

Then

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} \left(2 \cdot F_X(y^{1/2}) - 1 \right) \\&= 2f_X(y^{1/2}) \times \frac{1}{2}y^{-1/2} \quad \text{by chain rule} \\&= \frac{1}{\sqrt{y}} f_X(\sqrt{y}) \\&= \frac{1}{\sqrt{y}} \times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(\sqrt{y})^2 \right\} \\&= \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad \text{for } y \geq 0, \text{ zero otherwise.}\end{aligned}$$

This is the PDF for a **chi-squared distribution on 1 degree of freedom**.

Example 5.2

Then

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} \left(2 \cdot F_X(y^{1/2}) - 1 \right) \\&= 2f_X(y^{1/2}) \times \frac{1}{2}y^{-1/2} \quad \text{by chain rule} \\&= \frac{1}{\sqrt{y}} f_X(\sqrt{y}) \\&= \frac{1}{\sqrt{y}} \times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(\sqrt{y})^2 \right\} \\&= \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad \text{for } y \geq 0, \text{ zero otherwise.}\end{aligned}$$

This is the PDF for a **chi-squared distribution on 1 degree of freedom**.

Example 5.2

Then

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} \left(2 \cdot F_X(y^{1/2}) - 1 \right) \\&= 2f_X(y^{1/2}) \times \frac{1}{2}y^{-1/2} \quad \text{by chain rule} \\&= \frac{1}{\sqrt{y}} f_X(\sqrt{y}) \\&= \frac{1}{\sqrt{y}} \times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(\sqrt{y})^2 \right\} \\&= \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad \text{for } y \geq 0, \text{ zero otherwise.}\end{aligned}$$

This is the PDF for a **chi-squared distribution on 1 degree of freedom**.

Example 5.2

Then

$$\begin{aligned}f_Y(y) &= \frac{d}{dy} \left(2 \cdot F_X(y^{1/2}) - 1 \right) \\&= 2f_X(y^{1/2}) \times \frac{1}{2}y^{-1/2} \quad \text{by chain rule} \\&= \frac{1}{\sqrt{y}} f_X(\sqrt{y}) \\&= \frac{1}{\sqrt{y}} \times \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{1}{2}(\sqrt{y})^2 \right\} \\&= \frac{1}{\sqrt{2\pi y}} e^{-y/2}, \quad \text{for } y \geq 0, \text{ zero otherwise.}\end{aligned}$$

This is the PDF for a **chi-squared distribution on 1 degree of freedom**.

Example 5.2

To confirm this result we can simulate Y :

```
1 y = (rnorm(1000))^2
2 hist(y, xlab="Y", ylab="density", prob=TRUE)
3 curve((1/sqrt(2*pi*x))*exp(-x/2), from=0, to=10, col="
  purple", lty=2, add=TRUE)
```

