Chapter 3

Continuous Random Variables

3.1 Introduction

Rather than *summing* probabilities related to discrete random variables, here for continuous random variables, the *density* curve is *integrated* to determine probability.

Exercise 3.1 (Introduction)
Patient's number of visits, X, and duration of visit, Y.

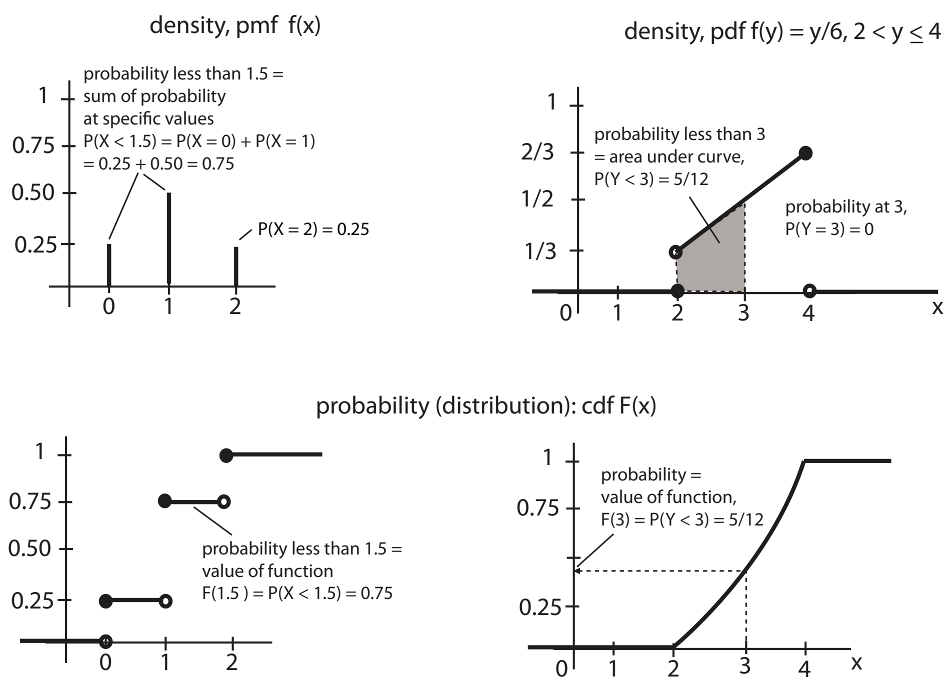

Figure 3.1: Comparing discrete and continuous distributions

1. Number of visits, X is a (i) **discrete** (ii) **continuous** random variable, and duration of visit, Y is a (i) **discrete** (ii) **continuous** random variable.

2. *Discrete*

   (a) $P(X = 2) =$ (i) **0** (ii) **0.25** (iii) **0.50** (iv) **0.75**

   (b) $P(X \leq 1.5) = P(X \leq 1) = F(1) = 0.25 + 0.50 = 0.75$
   requires (i) **summation** (ii) **integration** and is a value of a
   (i) **probability mass function** (ii) **cumulative distribution function**
   which is a (i) **stepwise** (ii) **smooth increasing** function

   (c) $E(X) =$ (i) $\boldsymbol{\sum xf(x)}$ (ii) $\boldsymbol{\int xf(x)\,dx}$

   (d) $Var(X) =$ (i) $\boldsymbol{E(X^2) - \mu^2}$ (ii) $\boldsymbol{E(Y^2) - \mu^2}$

   (e) $M(t) =$ (i) $\boldsymbol{E\left(e^{tX}\right)}$ (ii) $\boldsymbol{E\left(e^{tY}\right)}$

   (f) Examples of discrete densities (distributions) include (choose one or more)
   (i) **uniform**
   (ii) **geometric**
   (iii) **hypergeometric**
   (iv) **binomial (Bernoulli)**
   (v) **Poisson**

3. *Continuous*

   (a) $P(Y = 3) =$ (i) **0** (ii) **0.25** (iii) **0.50** (iv) **0.75**

   (b) $P(Y \leq 3) = F(3) = \int_2^3 \frac{x}{6}\,dx = \frac{x^2}{12}\Big]_{x=2}^{x=3} = \frac{3^2}{12} - \frac{2^2}{12} = \frac{5}{12}$
   requires (i) **summation** (ii) **integration** and is a value of a
   (i) **probability density function** (ii) **cumulative distribution function**
   which is a (i) **stepwise** (ii) **smooth increasing** function

   (c) $E(Y) =$ (i) $\boldsymbol{\sum yf(y)}$ (ii) $\boldsymbol{\int yf(y)\,dy}$

   (d) $Var(Y) =$ (i) $\boldsymbol{E(X^2) - \mu^2}$ (ii) $\boldsymbol{E(Y^2) - \mu^2}$

   (e) $M(t) =$ (i) $\boldsymbol{E\left(e^{tX}\right)}$ (ii) $\boldsymbol{E\left(e^{tY}\right)}$

   (f) Examples of continuous densities (distributions) include (choose one or more)
   (i) **uniform**
   (ii) **exponential**
   (iii) **normal (Gaussian)**
   (iv) **Gamma**
   (v) **chi-square**
   (vi) **student-t**
   (vii) **F**

3.2 Definitions

Random variable X is *continuous* if *probability density function* (pdf) f is continuous at all but a finite number of points and possesses the following properties:

- $f(x) \geq 0$, for all x,
- $\int_{-\infty}^{\infty} f(x)\, dx = 1$,
- $P(a < X \leq b) = \int_a^b f(x)\, dx$

The *(cumulative) distribution function* (cdf) for random variable X is

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\, dt,$$

and has properties

- $\lim_{x \to -\infty} F(x) = 0$,
- $\lim_{x \to \infty} F(x) = 1$,
- if $x_1 < x_2$, then $F(x_1) \leq F(x_2)$; that is, F is nondecreasing,
- $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a) = \int_a^b f(x)\, dx$,
- $F'(x) = \frac{d}{dx} \int_{-\infty}^{x} f(t)\, dt = f(x)$.

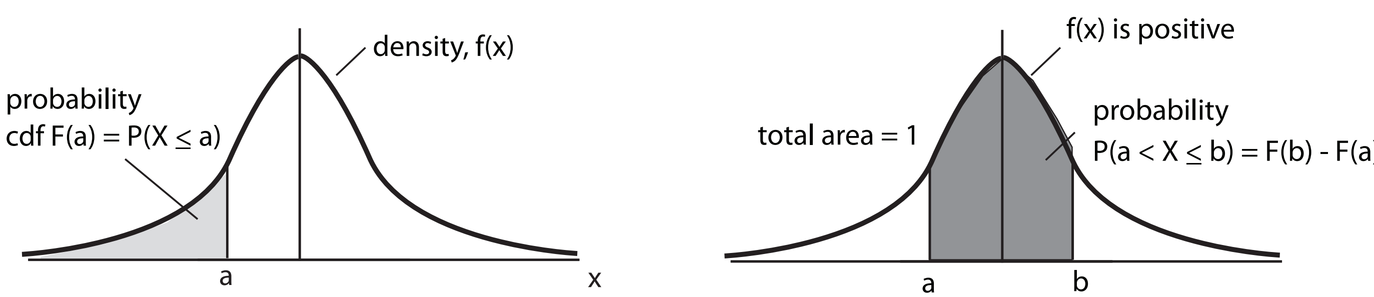

Figure 3.2: Continuous distribution

The *expected value* or *mean* of random variable X is given by

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\, dx,$$

the *variance* is

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = E(X^2) - [E(X)]^2 = E(X^2) - \mu^2$$

with associated *standard deviation*, $\sigma = \sqrt{\sigma^2}$.

The *moment-generating function* is

$$M(t) = E\left[e^{tX}\right] = \int_{-\infty}^{\infty} e^{tX} f(x)\, dx$$

for values of t for which this integral exists.

Expected value, assuming it exists, of a function u of X is

$$E[u(X)] = \int_{-\infty}^{\infty} u(x) f(x)\, dx$$

The $(100p)$th *percentile* is a value of X denoted π_p where

$$p = \int_{-\infty}^{\pi_p} f(x)\, dx = F(\pi_p)$$

and where π_p is also called the *quantile of order p*. The 25th, 50th, 75th percentiles are also called *first*, *second*, *third quartiles*, denoted $p_1 = \pi_{0.25}$, $p_2 = \pi_{0.50}$, $p_3 = \pi_{0.75}$ where also 50th percentile is called the *median* and denoted $m = p_2$. The *mode* is the value x where f is maximum.

Exercise 4.2 (Definitions)

1. *Waiting time.*
   Let the time waiting in line, in minutes, be described by the random variable X which has the following pdf,

$$f(x) = \begin{cases} \frac{1}{6}x, & 2 < x \leq 4, \\ 0, & \text{otherwise.} \end{cases}$$

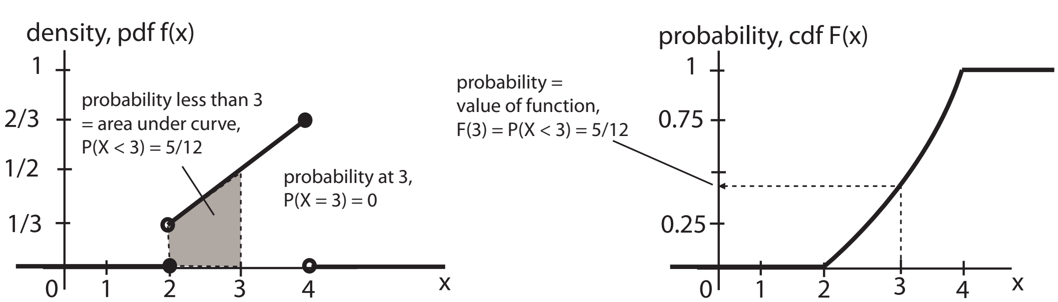

Figure 3.3: f(x) and F(x)

(a) Verify function $f(x)$ satisfies the second property of pdfs,

$$\int_{-\infty}^{\infty} f(x)\,dx = \int_2^4 \frac{1}{6}x\,dx = \left.\frac{x^2}{12}\right]_{x=2}^{x=4} = \frac{4^2}{12} - \frac{2^2}{12} = \frac{12}{12} =$$

(i) **0** (ii) **0.15** (iii) **0.5** (iv) **1**

(b) $P(2 < X \leq 3) =$

$$\int_2^3 \frac{1}{6}x\,dx = \left.\frac{x^2}{12}\right]_{x=2}^{x=3} = \frac{3^2}{12} - \frac{2^2}{12} =$$

(i) **0** (ii) $\mathbf{\frac{5}{12}}$ (iii) $\mathbf{\frac{9}{12}}$ (iv) **1**

(c) $P(X = 3) = P(3^- < X \leq 3) = \frac{3^2}{12} - \frac{3^2}{12} = 0 \neq f(3) = \frac{1}{6} \cdot 3 = 0.5$
(i) **True** (ii) **False**
So the pdf $f(x) = \frac{1}{6}x$ determined at some value of x does *not* determine probability.

(d) $P(2 < X \leq 3) = P(2 < X < 3) = P(2 \leq X \leq 3) = P(2 \leq X < 3)$
(i) **True** (ii) **False**
because $P(X = 3) = 0$ and $P(X = 2) = 0$

(e) $P(0 < X \leq 3) =$

$$\int_2^3 \frac{1}{6}x\,dx = \left.\frac{x^2}{12}\right]_{x=2}^{x=3} = \frac{3^2}{12} - \frac{2^2}{12} =$$

(i) **0** (ii) $\mathbf{\frac{5}{12}}$ (iii) $\mathbf{\frac{9}{12}}$ (iv) **1**
Why integrate from 2 to 3 and not 0 to 3?

(f) $P(X \leq 3) =$

$$\int_2^3 \frac{1}{6}x\,dx = \left.\frac{x^2}{12}\right]_{x=2}^{x=3} = \frac{3^2}{12} - \frac{2^2}{12} =$$

(i) **0** (ii) $\mathbf{\frac{5}{12}}$ (iii) $\mathbf{\frac{9}{12}}$ (iv) **1**

(g) Determine cdf (not pdf) $F(3)$.

$$F(3) = P(X \leq 3) = \int_2^3 \frac{1}{6}x\,dx = \left.\frac{x^2}{12}\right]_{x=2}^{x=3} = \frac{3^2}{12} - \frac{2^2}{12} =$$

(i) **0** (ii) $\mathbf{\frac{5}{12}}$ (iii) $\mathbf{\frac{9}{12}}$ (iv) **1**

(h) Determine $F(3) - F(2)$.

$$F(3)-F(2) = P(X \leq 3)-P(X \leq 2) = P(2 \leq X \leq 3) = \int_2^3 \frac{1}{6}x\,dx = \left.\frac{x^2}{12}\right]_{x=2}^{x=3} = \frac{3^2}{12}-\frac{2^2}{12} =$$

(i) **0** (ii) $\mathbf{\frac{5}{12}}$ (iii) $\mathbf{\frac{9}{12}}$ (iv) **1**
because everything left of (below) 3 subtract everything left of 2 equals what is between 2 and 3

(i) The *general* distribution function (cdf) is

$$F(x) = \begin{cases} \int_{-\infty}^{x} 0\,dt \;=\; 0, & x \leq 2, \\ \int_{2}^{x} \frac{t}{6}\,dt \;=\; \left.\frac{t^2}{12}\right]_{t=2}^{t=x} \;=\; \frac{x^2}{12} - \frac{4}{12}, & 2 < x \leq 4, \\ 1, & x > 4. \end{cases}$$

In other words, $F(x) = \frac{x^2}{12} - \frac{4}{12} = \frac{x^2-4}{12}$ on $(2, 4]$. Both pdf density and cdf distribution are given in the figure above.
(i) **True** (ii) **False**

(j) $F(2) = \frac{2^2}{12} - \frac{4}{12} =$ (i) **0** (ii) $\mathbf{\frac{5}{12}}$ (iii) $\mathbf{\frac{9}{12}}$ (iv) **1**.

(k) $F(3) = \frac{3^2}{12} - \frac{4}{12} =$ (i) **0** (ii) $\mathbf{\frac{5}{12}}$ (iii) $\mathbf{\frac{9}{12}}$ (iv) **1**.

(l) $P(2.5 < X < 3.5) = F(3.5) - F(2.5) = \left(\frac{3.5^2}{12} - \frac{4}{12}\right) - \left(\frac{2.5^2}{12} - \frac{4}{12}\right) =$
(i) **0** (ii) **0.25** (iii) **0.50** (iv) **1**.

(m) $P(X > 3.5) = 1 - P(X \leq 3.5) = 1 - F(3.5) = 1 - \left(\frac{3.5^2}{12} - \frac{4}{12}\right) =$
(i) **0.3125** (ii) **0.4425** (iii) **0.7650** (iv) **1**.

(n) *Expected value.* The average wait time is

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{2}^{4} x\left(\frac{1}{6}x\right)\,dx = \int_{2}^{4} \frac{x^2}{6}\,dx = \left.\frac{x^3}{18}\right]_2^4 = \frac{4^3}{18} - \frac{2^3}{18} =$$

(i) $\mathbf{\frac{23}{9}}$ (ii) $\mathbf{\frac{28}{9}}$ (iii) $\mathbf{\frac{31}{9}}$ (iv) $\mathbf{\frac{35}{9}}$.

```
library(MASS) # INSTALL once, RUN once (per session) library MASS
mean <- 4^3/18 - 2^3/18; mean; fractions(mean) # turn decimal to fraction
```

```
[1] 3.111111
[1] 28/9
```

(o) *Expected value of function $u = x^2$.*

$$E\left(X^2\right) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_{2}^{4} x^2\left(\frac{1}{6}x\right)\,dx = \int_{2}^{4} \frac{x^3}{6}\,dx = \left.\frac{x^4}{24}\right]_2^4 = \frac{4^4}{24} - \frac{2^4}{24} =$$

(i) **9** (ii) **10** (iii) **11** (iv) **12**.

(p) *Variance, method 1.* Variance in wait time is

$$\sigma^2 = Var(X) = E\left(X^2\right) - \mu^2 = 10 - \left(\frac{28}{9}\right)^2 =$$

(i) $\mathbf{\frac{23}{81}}$ (ii) $\mathbf{\frac{26}{81}}$ (iii) $\mathbf{\frac{31}{81}}$ (iv) $\mathbf{\frac{35}{81}}$.

```
library(MASS) # INSTALL once, RUN once (per session) library MASS
sigma2 <- 10 - (28/9)^2 ; fractions(sigma2) # turn decimal to fraction
```

```
[1] 26/81
```

(q) *Variance, method 2.* Variance in wait time is

$$\begin{aligned}
\sigma^2 = Var(X) = E[(X-\mu)^2] &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx \\
&= \int_2^4 \left(x - \frac{28}{9}\right)^2 \left(\frac{1}{6}x\right)\, dx \\
&= \int_2^4 \left(\frac{x^3}{6} - \frac{56x^2}{54} + \frac{784x}{486}\right)\, dx \\
&= \left.\frac{x^4}{24} - \frac{56x^3}{162} + \frac{784x^2}{972}\right]_2^4 =
\end{aligned}$$

(i) $\mathbf{\frac{23}{81}}$ (ii) $\mathbf{\frac{26}{81}}$ (iii) $\mathbf{\frac{31}{81}}$ (iv) $\mathbf{\frac{35}{81}}$.

```
library(MASS) # INSTALL once, RUN once (per session) library MASS
sigma2 <- (4^4/24 - 56*4^3/162 + 784*4^2/972) - (2^4/24 - 56*2^3/162 + 784*2^2/972)
fractions(sigma2) # turn decimal to fraction
```

```
[1] 26/81
```

(r) *Standard deviation.* Standard deviation in time spent on the call is,

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{26}{81}} \approx$$

(i) **0.57** (ii) **0.61** (iii) **0.67** (iv) **0.73**.

(s) *Moment generating function.*

$$\begin{aligned}
M(t) = E[e^{tX}] &= \int_{-\infty}^{\infty} e^{tx} f(x)\, dx \\
&= \int_2^4 e^{tx}\left(\frac{1}{6}x\right)\, dx \\
&= \frac{1}{6}\int_2^4 xe^{tx}\, dx \\
&= \left.\frac{1}{6}\, e^{tx}\left(\frac{x}{t} - \frac{1}{t^2}\right)\right]_{x=2}^4 \\
&= \frac{1}{6}e^{4t}\left(\frac{4}{t} - \frac{1}{t^2}\right) - \frac{1}{6}e^{2t}\left(\frac{2}{t} - \frac{1}{t^2}\right), \quad t \neq 0.
\end{aligned}$$

(i) **True** (ii) **False**

use integration by parts for $t \neq 0$ case: $\int u\, dv = uv - \int v\, du$ where $u = x$, $dv = e^{tx} dx$.

(t) *Median.* Since the distribution function $F(x) = \frac{x^2-4}{12}$, then median $m = \pi_{0.50}$ occurs when

$$F(m) = P(X \leq m) = \frac{m^2 - 4}{12} = \frac{1}{2}$$

so

$$m = \pi_{0.50} = \sqrt{\frac{12}{2} + 4} \approx$$

(i) **1.16** (ii) **2.16** (iii) **3.16**

negative answer $m \approx -3.16$ is not in $2 < x \leq 4$, so $m \neq -3.16$

2. *Density with unknown a.* Find a such that

$$f(x) = \begin{cases} ax^2, & 1 < x \leq 5, \\ 0, & \text{otherwise.} \end{cases}$$

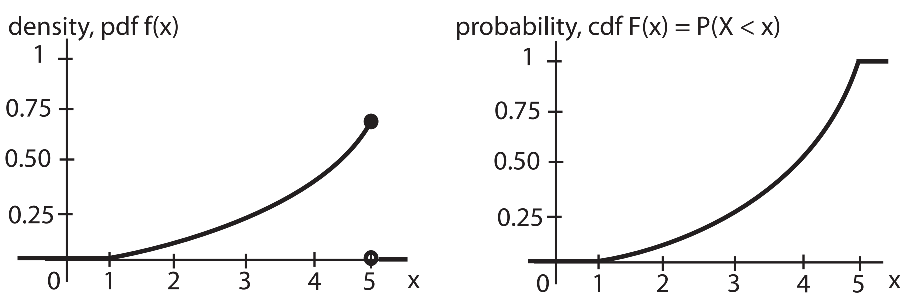

Figure 3.4: f(x) and F(x)

(a) *Find constant a.* Since the cdf $F(5) = 1$, and

$$F(x) = \int_1^x at^2 \, dt \; = \; a\frac{t^3}{3}\Big]_{t=1}^{t=x} \; = \; \frac{ax^3}{3} - \frac{a}{3} \; = \; \frac{a(x^3-1)}{3}$$

then

$$F(5) = \frac{a(5^3-1)}{3} \; = \; \frac{124a}{3} \; = \; 1,$$

so $a =$ (i) $\boldsymbol{\frac{3}{124}}$ (ii) $\boldsymbol{\frac{2}{124}}$ (iii) $\boldsymbol{\frac{1}{124}}$

(b) Determine pdf.

$$f(x) = ax^2 =$$

(i) $\boldsymbol{\frac{3}{124}x^2}$ (ii) $\boldsymbol{\frac{2}{124}x^2}$ (iii) $\boldsymbol{\frac{1}{124}x^2}$

(c) Determine cdf.

$$F(x) = \frac{a(x^3-1)}{3} = \frac{3}{124} \times \frac{x^3-1}{3} =$$

(i) $\boldsymbol{\frac{3}{124}(x^3-1)}$ (ii) $\boldsymbol{\frac{2}{124}(x^3-1)}$ (iii) $\boldsymbol{\frac{1}{124}(x^3-1)}$

(d) Determine $P(X \geq 2)$.

$$P(X \geq 2) = 1 - P(X < 2) = 1 - F(2) = 1 - \frac{1}{124}(2^3 - 1) =$$

(i) $\boldsymbol{\frac{61}{124}}$ (ii) $\boldsymbol{\frac{117}{124}}$ (iii) $\boldsymbol{\frac{61}{117}}$

(e) Determine $P(X \geq 4)$.

$$P(X \geq 4) = 1 - P(X < 4) = 1 - F(4) = 1 - \frac{1}{124}(4^3 - 1) =$$

(i) $\boldsymbol{\frac{61}{124}}$ (ii) $\boldsymbol{\frac{117}{124}}$ (iii) $\boldsymbol{\frac{61}{117}}$

(f) Determine $P(X \geq 4 | X \geq 2)$.

$$P(X \geq 4|X \geq 2) = \frac{P(X \geq 4 \cap X \geq 2)}{P(X \geq 2)} = \frac{P(X \geq 4)}{P(X \geq 2)} = \frac{61/124}{117/124} =$$

(i) $\boldsymbol{\frac{116}{124}}$ (ii) $\boldsymbol{\frac{117}{124}}$ (iii) $\boldsymbol{\frac{61}{117}}$

(g) *Expected value.*

$$\mu = E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_1^5 x\left(\frac{3}{124}x^2\right)\,dx = \int_1^5 \frac{3x^3}{124}\,dx = \left.\frac{3x^4}{496}\right]_1^5 = \frac{3\cdot 5^4}{496} - \frac{3\cdot 1^4}{496} =$$

(i) $\boldsymbol{\frac{114}{31}}$ (ii) $\boldsymbol{\frac{115}{31}}$ (iii) $\boldsymbol{\frac{116}{31}}$ (iv) $\boldsymbol{\frac{117}{31}}$.

(h) *Expected value of function $u = x^2$.*

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_1^5 x^2\left(\frac{3}{124}x^2\right)\,dx = \int_1^5 \frac{3x^4}{124}\,dx = \left.\frac{3x^5}{620}\right]_1^5 = \frac{3\cdot 5^5}{620} - \frac{3\cdot 1^5}{620} \approx$$

(i) **15.12** (ii) **16.12** (iii) **17.12** (iv) **18.12**.

(i) Determine 25th percentile (first quartile), $p_1 = \pi_{0.25}$.

Since $F(x) = \frac{1}{124}(x^3 - 1)$, then,

$$F(\pi_{0.25}) = \frac{1}{124}(\pi_{0.25}^3 - 1) = 0.25 = \frac{1}{4},$$

and so

$$\pi_{0.25} = \sqrt[3]{\frac{124}{4} + 1} \approx$$

(i) **2.38** (ii) **3.17** (iii) **4.17**

The $\pi_{0.25}$ is that value where 25% of probability is at or below (to the left of) this value.

(j) Determine $\pi_{0.01}$.

$$F(\pi_{0.01}) = \frac{1}{124}(\pi_{0.01}^3 - 1) = 0.01,$$

and so

$$\pi_{0.01} = \sqrt[3]{0.01 \cdot 124 + 1} \approx$$

(i) **1.31** (ii) **3.17** (iii) **4.17**

The $\pi_{0.01}$ is that value where 1% of probability is at or below (to the left of) this value.

3. *Piecewise pdf.* Let random variable X have pdf

$$f(x) = \begin{cases} x, & 0 < x \leq 1, \\ 2 - x, & 1 < x \leq 2, \\ 0, & \text{elsewhere.} \end{cases}$$

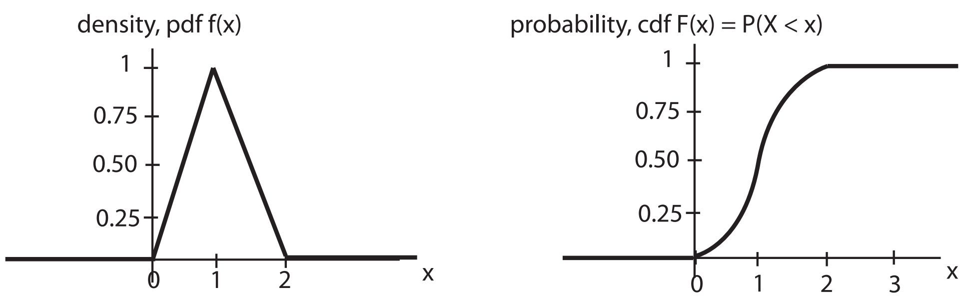

Figure 3.5: f(x) and F(x)

(a) *Expected value.*

$$\begin{aligned}
\mu \;=\; E(X) &= \int_{-\infty}^{\infty} x f(x)\, dx \\
&= \int_0^1 x\,(x)\ dx + \int_1^2 x\,(2-x)\ dx \\
&= \int_0^1 x^2\, dx + \int_1^2 \left(2x - x^2\right)\ dx \\
&= \left[\frac{x^3}{3}\right]_0^1 + \left[\frac{2x^2}{2} - \frac{x^3}{3}\right]_1^2 \\
&= \left(\frac{1^3}{3} - \frac{0^3}{3}\right) + \left(2^2 - \frac{2^3}{3}\right) - \left(1^2 - \frac{1^3}{3}\right) =
\end{aligned}$$

(i) **1** (ii) **2** (iii) **3** (iv) **4**.

(b) $E(X^2)$.

$$\begin{aligned}
E\left(X^2\right) &= \int_0^1 x^2\,(x)\;dx + \int_1^2 x^2\,(2-x)\;dx \\
&= \int_0^1 x^3\;dx + \int_1^2 \left(2x^2 - x^3\right)\;dx \\
&= \left[\frac{x^4}{4}\right]_0^1 + \left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_1^2 \\
&= \left(\frac{1^4}{4} - \frac{0^4}{4}\right) + \left(\frac{2(2)^3}{3} - \frac{2^4}{4}\right) - \left(\frac{2(1)^3}{3} - \frac{1^4}{4}\right) =
\end{aligned}$$

(i) $\mathbf{\frac{4}{6}}$ (ii) $\mathbf{\frac{5}{6}}$ (iii) $\mathbf{\frac{6}{6}}$ (iv) $\mathbf{\frac{7}{6}}$.

(c) *Variance.*

$$\sigma^2 = \text{Var}(X) = E\left(X^2\right) - \mu^2 = \frac{7}{6} - 1^2 =$$

(i) $\mathbf{\frac{1}{3}}$ (ii) $\mathbf{\frac{1}{4}}$ (iii) $\mathbf{\frac{1}{5}}$ (iv) $\mathbf{\frac{1}{6}}$.

(d) *Standard deviation.*

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{6}} \approx$$

(i) **0.27** (ii) **0.31** (iii) **0.41** (iv) **0.53**.

(e) *Median.* Since the distribution function is

$$F(x) = \begin{cases} 0 & x \leq 0 \\ \int_0^x t\;dt \;=\; \left.\frac{t^2}{2}\right]_{t=0}^{x} = \frac{x^2}{2}, & 0 < x \leq 1, \\ \int_1^x (2-t)\;dt + F(1) \;=\; 2t - \left.\frac{t^2}{2}\right]_{t=1}^{x} + \frac{1}{2} = 2x - \frac{x^2}{2} - (2(1) - \frac{1^2}{2}) + \frac{1}{2}, & 1 < x \leq 2, \\ 1, & x > 2. \end{cases}$$

$F(1) = \frac{1}{2}$, so add $\frac{1}{2}$ to $F(x)$ for $1 < x \leq 2$

then median m occurs when

$$F(x) = \begin{cases} \frac{m^2}{2} = \frac{1}{2}, & 0 < x \leq 1, \\ 2m - \frac{m^2}{2} - 1 = \frac{1}{2}, & 1 < x \leq 2, \\ 1, & x > 2. \end{cases}$$

so for both $0 < x \leq 1$ and $1 < x \leq 2$, $m =$ (i) **1** (ii) **1.5** (iii) **2**

3.3 The Uniform and Exponential Distributions

Two special probability density functions are discussed: uniform and exponential. The continuous *uniform* (rectangular) distribution of random variable X has density

$$f(x) = \begin{cases} \frac{1}{b-a}, & a \leq x \leq b, \\ 0, & \text{elsewhere,} \end{cases}$$

where "X is $U(a,b)$" means "X is uniform over [a,b]", distribution function,

$$F(x) = \begin{cases} 0 & x < a, \\ \frac{x-a}{b-a} & a \leq x \leq b, \\ 1 & x > b, \end{cases}$$

and its expected value (mean), variance and standard deviation are,

$$\mu = E(X) = \frac{a+b}{2}, \quad \sigma^2 = Var(X) = \frac{(b-a)^2}{12}, \quad \sigma = \sqrt{Var(X)},$$

and moment-generating function is

$$M(t) = E\left[e^{tX}\right] = \frac{e^{tb} - e^{ta}}{t(b-a)}, \ t \neq 0.$$

The continuous *exponential* random variable X has density

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0, \\ 0, & \text{elsewhere,} \end{cases}$$

distribution function,

$$F(x) = \begin{cases} 0 & x < 0, \\ 1 - e^{-\lambda x} & x \geq 0, \end{cases}$$

and its expected value (mean), variance and standard deviation are,

$$\mu = E(X) = \frac{1}{\lambda}, \quad \sigma^2 = V(Y) = \frac{1}{\lambda^2}, \quad \sigma = \frac{1}{\lambda},$$

and moment-generating function is

$$M(t) = E\left[e^{tX}\right] = \frac{\lambda}{\lambda - t}, \quad t < \lambda.$$

Notice if $\theta = \frac{1}{\lambda}$ in the exponential distribution,

$$f(x) = \frac{1}{\theta} e^{-\frac{x}{\theta}}, \quad F(x) = 1 - e^{-\frac{x}{\theta}}, \quad \mu = \sigma = \theta, \quad M(t) = \frac{\theta}{1 - t\theta}.$$

Exercise 3.3 (The Uniform and Exponential Distributions)

1. *Uniform: potato weights.* An automated process fills one bag after another with Idaho potatoes. Although each filled bag should weigh 50 pounds, in fact, because of the differing shapes and weights of each potato, each bag weight, X, is anywhere from 49 pounds to 51 pounds, with uniform density:

$$f(x) = \begin{cases} 0.5, & 49 \leq x \leq 51, \\ 0, & \text{elsewhere.} \end{cases}$$

   (a) Since $a = 49$ and $b = 51$, the distribution is

$$F(x) = \begin{cases} 0 & x < 49, \\ \frac{x-49}{51-49} & 49 \leq x \leq 51, \\ 1 & x > 52, \end{cases}$$

   and so graphs of density and distribution are given in the figure.
   (i) **True** (ii) **False**

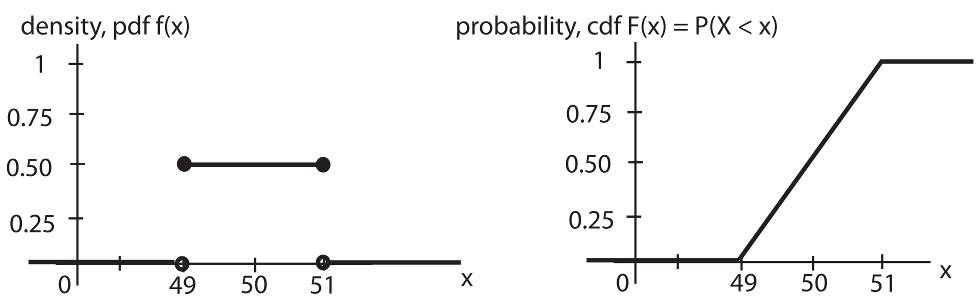

Figure 3.6: Distribution function: continuous uniform

   (b) Determine $P(49.5 < X < 51)$ by integrating pdf.

$$P(49.5 < X < 51) = \int_{49.5}^{51} \frac{1}{2}\, dx = \frac{x}{2}\Big]_{x=49.5}^{51} = \frac{51}{2} - \frac{49.5}{2} = \frac{1.5}{2}$$

   (i) **0.25** (ii) **0.50** (iii) **0.75** (iv) **1**.

   (c) Determine $P(49.5 < X < 51)$ using cdf.

$$P(49.5 < X < 51) = F(51) - F(49.5) = \frac{51-49}{51-49} - \frac{49.5-49}{51-49} =$$

   (i) **0.25** (ii) **0.50** (iii) **0.75** (iv) **1**.

```
1 - punif(49.5,49,51) # uniform P(49.5 < X < 51) = 1 - P(X < 49.5), 49 < x < 51

[1] 0.75
```

(d) The chance the bags weight more than 49.5

$$P(X > 49.5) = 1 - P(X \leq 49.5) = 1 - F(49.5) = 1 - \frac{49.5 - 49}{51 - 49} =$$

(i) **0.25** (ii) **0.50** (iii) **0.75** (iv) **1**.

```
1 - punif(49.5,49,51) # uniform P(X > 49.5) = 1 - P(X < 49.5), 49 < x < 51
```

```
[1] 0.75
```

(e) What is the mean weight of a bag of potatoes?

$$\mu = E(X) = \frac{a+b}{2} = \frac{49+51}{2} =$$

(i) **49** (ii) **50** (iii) **51** (iv) **52**.

(f) What is the standard deviation in the weight of a bag of potatoes?

$$\sigma = \sqrt{\frac{(b-a)^2}{12}} = \sqrt{\frac{(51-49)^2}{12}} \approx$$

(i) **0.44** (ii) **0.51** (iii) **0.55** (iv) **0.58**.

(g) Determine probability within 1 standard deviation of mean.

$$\begin{aligned} P(\mu - \sigma < X < \mu + \sigma) &\approx P(50 - 0.58 < X < 50 + 0.58) \\ &= P(49.42 < X < 50.58) \\ &= \int_{49.42}^{50.58} \frac{1}{2}\, dx = \left.\frac{x}{2}\right]_{x=49.42}^{50.58} = \end{aligned}$$

(i) **0.44** (ii) **0.51** (iii) **0.55** (iv) **0.58**.

(h) *Function of X.* If it costs \$0.0556 (5.56 cents) per pound of potatoes, then the cost of X pounds of potatoes is $Y = 0.0556X$. Determine the probability a bag of potatoes chosen at random costs at least \$2.78.

$$P(Y \geq 2.78) = P(0.0556X \geq 2.78) = P(X \geq 50) = \int_{50}^{51} \frac{1}{2}\, dx = \left.\frac{x}{2}\right]_{x=50}^{51} =$$

(i) **0.25** (ii) **0.50** (iii) **0.75** (iv) **1**.

2. *Exponential: atomic decay.* Assume atomic time-to-decay obeys exponential pdf with (inverse) mean rate of decay $\lambda = 3$,

$$f(x) = \begin{cases} 3e^{-3x}, & x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

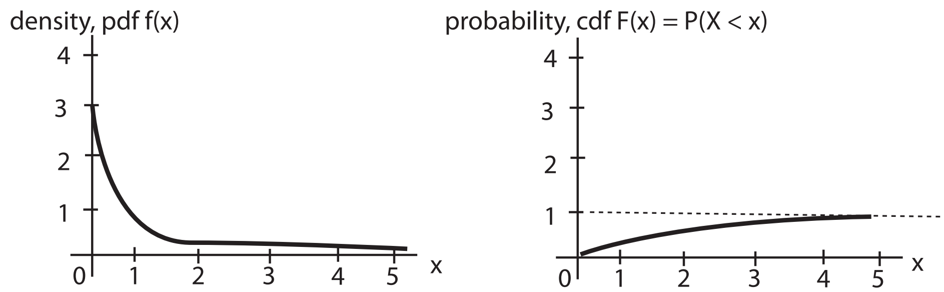

Figure 3.7: f(x) and F(x)

(a) *Find $F(x)$.* By substituting $u = -3x$, then $du = -3\,dx$ and so

$$F(x) = \int 3e^{-3x}\,dx = -\int e^{-3x}(-3)\,dx = -\int e^u\,du = -e^u =$$

(i) $\boldsymbol{-e^{-3x}}$ (ii) $\boldsymbol{e^{-3x}}$ (iii) $\boldsymbol{-3e^{-3x}}$
then

$$F(x) = \left[-e^{-3t}\right]_{t=0}^{t=x} = \left(-e^{-3x} - (-e^{-3(0)})\right) =$$

(i) $\boldsymbol{e^{-3x} - 1}$ (ii) $\boldsymbol{e^{-3x} + 1}$ (iii) $\boldsymbol{1 - e^{-3x}}$

(b) Chance atomic decay takes at least 2 microseconds, $P(X \geq 2)$, is

$$P(X \geq 2) = 1 - P(X < 2) = 1 - F(2) = 1 - \left(1 - e^{-3(2)}\right) \approx$$

(i) **0.002** (ii) **0.003** (iii) **0.004**

```
1 - pexp(2,3) # exponential P(X > 2), lambda = 3
```

```
[1] 0.002478752
```

(c) Chance atomic decay is between 1.13 and 1.62 microseconds is

$$\begin{aligned} P(1.13 < X < 1.62) &= F(1.62) - F(1.13) \\ &= \left(1 - e^{-(3)(1.62)}\right) - \left(1 - e^{-(3)(1.13)}\right) \\ &= e^{-(3)(1.13)} - e^{-(3)(1.62)} \approx \end{aligned}$$

(i) **0.014** (ii) **0.026** (iii) **0.034** (iv) **0.054**.

```
pexp(1.62,3) - pexp(1.13,3) # exponential P(1.13 < X < 1.62), lambda = 3
```

```
[1] 0.02595819
```

(d) Calculate probability atomic rate between mean and 1 standard deviation below mean. Since $\lambda = 3$, so $\mu = \sigma = \frac{1}{3}$, so

$$\begin{aligned} P(\mu - \sigma < X < \mu) &= P\left(\frac{1}{3} - \frac{1}{3} < X < \frac{1}{3}\right) \\ &= P\left(0 < X < \frac{1}{3}\right) \\ &= F\left(\frac{1}{3}\right) - F(0) \\ &= \left(1 - e^{-(3)\frac{1}{3}}\right) - \left(1 - e^{-(3)(0)}\right) \approx \end{aligned}$$

(i) **0.33** (ii) **0.43** (iii) **0.53** (iv) **0.63**.

(e) *Check if $f(x)$ is a pdf.*

$$\int_0^\infty 3e^{-3x}\,dx = \lim_{b\to\infty}\int_0^b 3e^{-3x}\,dx = \lim_{b\to\infty}\left[-e^{-3x}\right]_{x=0}^{x=b} = \lim_{b\to\infty}\left[-e^{-3b} - (-e^{x(0)})\right] = 1$$

or, equivalently,

$$\lim_{b\to\infty} F(b) = \lim_{b\to\infty} 1 - e^{-3b} = 1$$

(i) **True** (ii) **False**

(f) *Show $F'(x) = f(x)$.*

$$F'(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\left(1 - e^{-3x}\right) =$$

(i) $\boldsymbol{3e^{-3x}}$ (ii) $\boldsymbol{e^{-3x} + 1}$ (iii) $\boldsymbol{1 - e^{-3x}}$

(g) Determine μ using mgf. Since $M(t) = \frac{\lambda}{\lambda - t}$,

$$M'(t) = \frac{d}{dt}\left(\frac{\lambda}{\lambda - t}\right) = \frac{\lambda}{(\lambda - t)^2},$$

so $\mu = M'(0) =$ (i) $\boldsymbol{\lambda}$ (ii) **1** (iii) $\boldsymbol{1 - \lambda}$ (iv) $\boldsymbol{\frac{1}{\lambda}}$.

use quotient rule, since $f = \frac{u}{v} = \frac{\lambda}{\lambda - t}$, $f' = \frac{vu' - uv'}{v^2} = \frac{(\lambda - t)(0) - \lambda(-1)}{(\lambda - t)^2} = \frac{\lambda}{(\lambda - t)^2} = \frac{1}{\lambda}$ if $t = 0$

(h) *Memoryless property of exponential.* Chance atomic decay lasts at least 10 microseconds is

$$P(X > 10) = 1 - F(10) = 1 - (1 - e^{-3(10)}) =$$

(i) $\boldsymbol{e^{-10}}$ (ii) $\boldsymbol{e^{-20}}$ (iii) $\boldsymbol{e^{-30}}$ (iv) $\boldsymbol{e^{-40}}$.

and chance atomic decay lasts at least 15 microseconds, given it has already lasted at least 5 microseconds is

$$P(X > 15 | X > 5) = \frac{P(X > 15, Y > 5)}{P(X > 5)} = \frac{P(X > 15)}{P(X > 5)} = \frac{1 - (1 - e^{-3(15)})}{1 - (1 - e^{-3(5)})} =$$

(i) $\boldsymbol{e^{-10}}$ (ii) $\boldsymbol{e^{-20}}$ (iii) $\boldsymbol{e^{-30}}$ (iv) $\boldsymbol{e^{-40}}$.

so, in other words,

$$P(X > 15|X > 5) = \frac{P(X > 15)}{P(X > 5)} = \frac{P(X > 10+5)}{P(X > 5)} = P(X > 10)$$

or

$$P(X > 10+5) = P(X > 10) \cdot P(X > 5)$$

This is an example of the "memoryless" property of the exponential, it implies time intervals are independent of one another. Chance of decay after 15 microsecond, given decay after 5 microseconds, same as chance of decay after 10 seconds; it is as though first 5 microseconds "forgotten".

3. *Verifying if experimental decay is exponential.* In an atomic decay study, let random variable X be time-to-decay where $x \geq 0$. The relative frequency distribution table and histogram for a *sample* of 20 atomic decays (measured in microseconds) from this study are given below. Does this data follow an exponential distribution?

0.60, 0.07, 0.66, 0.17, 0.06, 0.14, 0.15, 0.19, 0.07, 0.36
0.85, 0.44, 0.71, 1.02, 0.07, 0.21, 0.16, 0.16, 0.01, 0.88

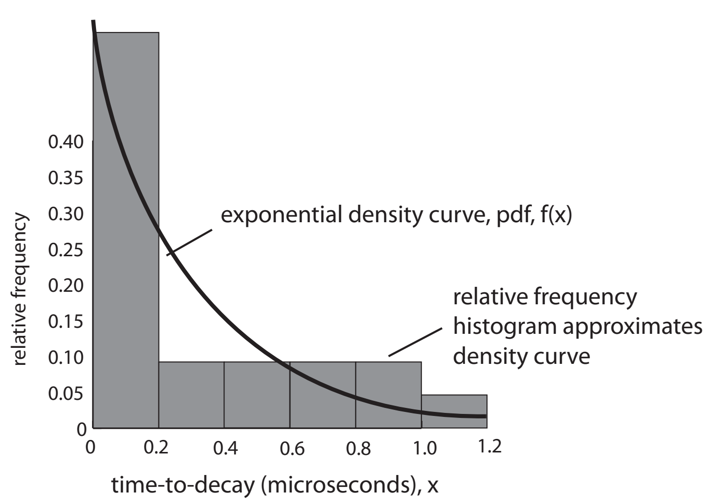

Figure 3.8: Histogram of exponential decay times

bin	frequency	relative frequency	exponential probability
(0,0.2]	11	$\frac{11}{20} = 0.55$	0.44
(0.2,0.4]	2	0.10	0.25
(0.4,0.6]	2	0.10	0.14
(0.6,0.8]	2	0.10	0.08
(0.8,1]	2	0.10	0.04
(1,1.2]	1	0.05	0.02

(a) *Sample mean time-to-decay.*

$$\bar{x} = \frac{0.60 + 0.07 + \cdots + 0.88}{20} =$$

(i) **0.330** (ii) **0.349** (iii) **0.532** (iv) **0.631**.

```
x <- c(0.60,0.07,0.66,0.17,0.06,0.14,0.15,0.19,0.07,0.36,0.85,0.44,0.71,1.02,0.07,0.21,0.16,0.16,0.01,0.88)
mean(x)
```

```
[1] 0.349
```

(b) *Approximate λ for exponential.* Since approximate mean time-to-decay is

$$\mu \approx \bar{x} = 0.349,$$

then, if time-to-decay is exponential, approximate value of λ parameter, since $\mu = \frac{1}{\lambda}$, is

$$\lambda = \frac{1}{\mu} \approx \frac{1}{\bar{x}} \approx$$

(i) **2.9** (ii) **3.0** (iii) **3.1** (iv) **3.2**

and so possible model of data might be exponential density
(i) $\boldsymbol{2.9e^{-3x}}$ (ii) $\boldsymbol{3e^{-2.9x}}$ (iii) $\boldsymbol{2.9e^{-2.9x}}$ (iv) $\boldsymbol{3e^{-3x}}$.

(c) *Approximate probabilities.*

$$P(0 \leq X \leq 0.2) = \int_0^{0.2} 2.9e^{-2.9x}\,dx \approx$$

(i) **0.44** (ii) **0.25** (iii) **0.14** (iv) **0.08**.
remaining probabilities in table calculated in a similar way

```
pexp(0.2,2.9) - pexp(0,2.9) # exponential P(0 < X < 0.2), lambda approx 2.9
```

```
[1] 0.4401016
```

(d) Since sample relative frequencies do not closely match exponential with $\lambda = 2.9$ in table, this distribution (i) **is** (ii) **is not** a good fit to the data.

3.4 The Normal Distribution

The continuous *normal* distribution of random variable X, defined on the interval $(-\infty, \infty)$, has pdf with parameters μ and σ, that is, "X is $N(\mu, \sigma^2)$",

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(x-\mu)/\sigma]^2},$$

and cdf

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} \frac{1}{\sigma\sqrt{2\pi}} e^{-(1/2)[(t-\mu)/\sigma]^2}\, dt,$$

and its expected value (mean), variance and standard deviation are,

$$E(X) = \mu, \quad Var(X) = \sigma^2, \quad \sigma = \sqrt{Var(X)},$$

and mgf is

$$M(t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}.$$

A normal random variable, X, may be transformed to a *standard* normal, Z,

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2},$$

where "Z is $N(0, 1)$" and

$$\Phi(z) = P(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt,$$

where $\mu = 0$, $\sigma = 1$ and $M(t) = e^{t^2/2}$ using the following equation,

$$Z = \frac{X - \mu}{\sigma}.$$

The *distribution* of this density does not have a closed–form expression and so must be solved using numerical integration methods. We will use both R and the tables to obtain approximate numerical answers.

Exercise 3.4 (The Normal Distribution)

1. *Nonstandard normal: IQ scores.*
   It has been found that IQ scores, Y, can be distributed by a normal distribution. Densities of IQ scores for 16 year olds, X_1, and 20 year olds, X_2, are given by

$$\begin{aligned} f(x_1) &= \frac{1}{16\sqrt{2\pi}} e^{-(1/2)[(y-100)/16]^2}, \\ f(x_2) &= \frac{1}{20\sqrt{2\pi}} e^{-(1/2)[(y-120)/20]^2}. \end{aligned}$$

   A graph of these two densities is given in the figure.

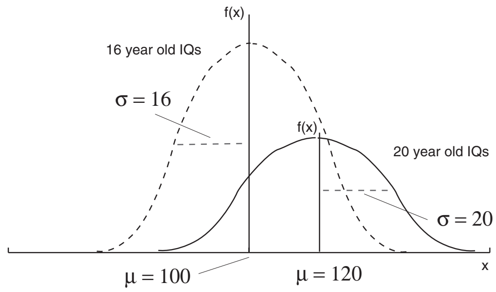

Figure 3.9: Normal distributions: IQ scores

(a) Mean IQ score for 20 year olds is
$\mu =$ (i) **100** (ii) **120** (iii) **124** (iv) **136**.

(b) Average (or mean) IQ scores for 16 year olds is
$\mu =$ (i) **100** (ii) **120** (iii) **124** (iv) **136**.

(c) Standard deviation in IQ scores for 20 year olds
$\sigma =$ (i) **16** (ii) **20** (iii) **24** (iv) **36**.

(d) Standard deviation in IQ scores for 16 year olds is
$\sigma =$ (i) **16** (ii) **20** (iii) **24** (iv) **36**.

(e) Normal density for 20 year old IQ scores is
(i) **broader than normal density for 16 year old IQ scores.**
(ii) **as wide as normal density for 16 year old IQ scores.**
(iii) **narrower than normal density for 16 year old IQ scores.**

(f) Normal density for the 20 year old IQ scores is
(i) **shorter than normal density for 16 year old IQ scores.**
(ii) **as tall as normal density for 16 year old IQ scores.**
(iii) **taller than normal density for 16 year old IQ scores.**

(g) Total area (probability) under normal density for 20's IQ scores is
(i) **smaller than area under normal density for 16's IQ scores.**
(ii) **the same as area under normal density for 16's IQ scores.**
(iii) **larger than area under normal density for 16's IQ scores.**

(h) Number of different normal densities:
(i) **one** (ii) **two** (iii) **three** (iv) **infinity**.

2. *Percentages: IQ scores.*
Densities of IQ scores for 16 year olds, X_1, and 20 year olds, X_2, are given by

$$\begin{aligned} f(x_1) &= \frac{1}{16\sqrt{2\pi}} e^{-(1/2)[(y-100)/16]^2}, \\ f(x_2) &= \frac{1}{20\sqrt{2\pi}} e^{-(1/2)[(y-120)/20]^2}. \end{aligned}$$

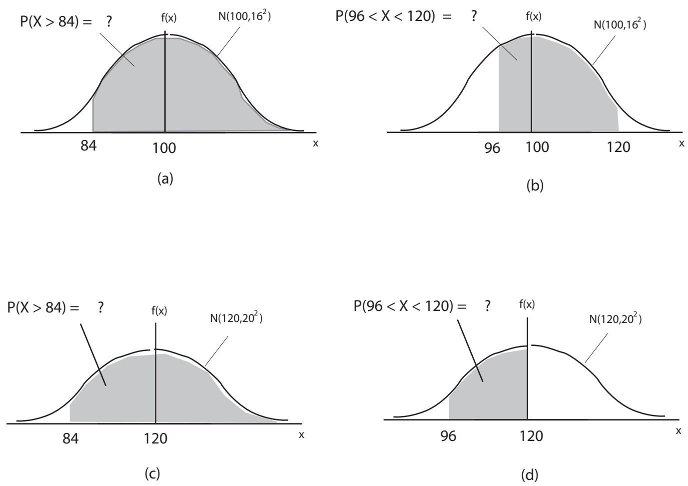

Figure 3.10: Normal probabilities: IQ scores

(a) For the sixteen year old normal distribution, where $\mu = 100$ and $\sigma = 16$,

$$P(X_1 > 84) = \int_{84}^{\infty} \frac{1}{16\sqrt{2\pi}} e^{-(1/2)[(y-100)/16]^2} \, dx_1 \approx$$

(i) **0.4931** (ii) **0.9641** (iii) **0.8413** (iv) **0.3849**.

```
1 - pnorm(84,100,16) # normal P(X > 84), mean = 100, SD = 16
```

```
[1] 0.8413447
```

(b) For sixteen year old normal distribution, where $\mu = 100$ and $\sigma = 16$, $P(96 < X_1 < 120) \approx$ (i) **0.4931** (ii) **0.9641** (iii) **0.8413** (iv) **0.3849**.

```
pnorm(120,100,16) - pnorm(96,100,16) # normal P(96 < X < 120), mean = 100, SD = 16
```

```
[1] 0.4930566
```

(c) For *twenty* year old, where $\mu = 120$ and $\sigma = 20$, $P(X_2 > 84) \approx$ (i) **0.4931** (ii) **0.9641** (iii) **0.8413** (iv) **0.3849**.

```
1 - pnorm(84,120,20) # normal P(X > 84), mean = 120, SD = 20
```

```
[1] 0.9640697
```

(d) For twenty year old, where $\mu = 120$ and $\sigma = 20$, $P(96 < X_2 < 120) \approx$ (i) **0.4931** (ii) **0.9641** (iii) **0.8413** (iv) **0.3849**.

```
pnorm(120,120,20) - pnorm(96,120,20) # normal P(96 < X < 120), mean = 120, SD = 20
```

```
[1] 0.3849303
```

(e) *Empirical Rule (68-95-99.7 rule).* If X is $N(120, 20^2)$, find probability within 1, 2 and 3 SDs of mean.

$P(\mu - \sigma < X < \mu + \sigma) = P(120 - 20 < X < 120 + 20) = P(100 < X < 140) \approx$

(i) **0.683** (ii) **0.954** (iii) **0.997** (iv) **1**.

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = P(80 < X < 160) \approx$$

(i) **0.683** (ii) **0.954** (iii) **0.997** (iv) **1**.

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = P(60 < X < 180) \approx$$

(i) **0.683** (ii) **0.954** (iii) **0.997** (iv) **1**.

Empirical rule is true for *all* X which are $N(\mu, \sigma^2)$.

```
pnorm(140,120,20) - pnorm(100,120,20);  pnorm(160,120,20) - pnorm(80,120,20); pnorm(180,120,20) - pnorm(60,120,20)
```

```
[1] 0.6826895  [1] 0.9544997  [1] 0.9973002
```

3. *Standard normal.*
Normal densities of IQ scores for 16 year olds, X_1, and 20 year olds, X_2,

$$\begin{aligned} f(x_1) &= \frac{1}{16\sqrt{2\pi}} e^{-(1/2)[(y-100)/16]^2}, \\ f(x_2) &= \frac{1}{20\sqrt{2\pi}} e^{-(1/2)[(y-120)/20]^2}. \end{aligned}$$

Both densities may be transformed to a *standard* normal with $\mu = 0$ and $\sigma = 1$ using the following equation,

$$Z = \frac{Y - \mu}{\sigma}.$$

(a) Since $\mu = 100$ and $\sigma = 16$, a 16 year old who has an IQ of 132 is $z = \frac{132-100}{16} =$ (i) **0** (ii) **1** (iii) **2** (iv) **3**
standard deviations above the mean IQ, $\mu = 100$.

(b) A 16 year old who has an IQ of 84 is
$z = \frac{84-100}{16} =$ (i) $\mathbf{-2}$ (ii) $\mathbf{-1}$ (iii) **0** (iv) **1**
standard deviations below the mean IQ, $\mu = 100$.

(c) Since $\mu = 120$ and $\sigma = 20$, a 20 year old who has an IQ of 180 is
$z = \frac{180-120}{20} =$ (i) **0** (ii) **1** (iii) **2** (iv) **3**
standard deviations above the mean IQ, $\mu = 120$.

(d) A 20 year old who has an IQ of 100 is
$z = \frac{100-120}{20} =$ (i) $\mathbf{-3}$ (ii) $\mathbf{-2}$ (iii) $\mathbf{-1}$ (iv) **0**
standard deviations below the mean IQ, $\mu = 120$.

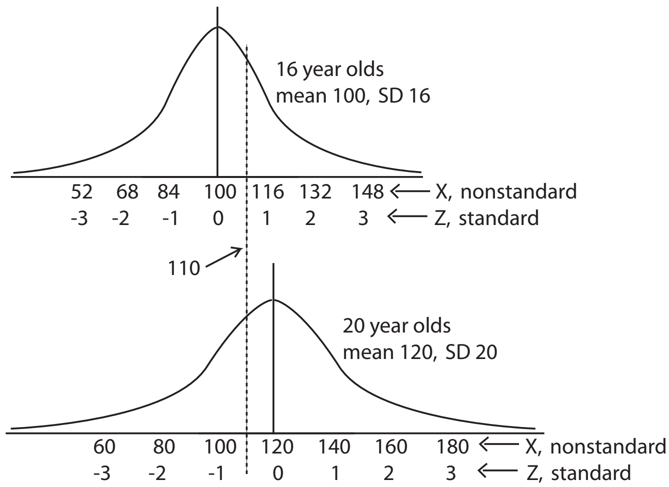

Figure 3.11: Standard normal and (nonstandard) normal

(e) Although both the 20 year old and 16 year old scored the same, 110, on an IQ test, the 16 year old is clearly brighter relative to his/her age group than is the 20 year old relative his/her age group because

$$z_1 = \frac{110 - 100}{16} = 0.625 > z_2 = \frac{110 - 120}{20} = -0.5.$$

(i) **True** (ii) **False**

(f) The probability a 20 year old has an IQ greater than 90 is

$$P(X_2 > 90) = P\left(Z_2 > \frac{90 - 120}{20}\right) = P\left(Z_2 > -1.5\right) \approx$$

(i) **0.93** (ii) **0.95** (iii) **0.97** (iv) **0.99**.

```
1 - pnorm(90,120,20) # normal P(X > 90), mean = 120, SD = 20
1 - pnorm(-1.5) # standard normal P(Z > -1.5), defaults to mean = 0, SD = 1
```

```
[1] 0.9331928
```

Or, using Table C.1 from the text,

$$P\left(Z_2 > -1.5\right) = 1 - \Phi(-1.5) \approx 1 - 0.0668 \approx$$

(i) **0.93** (ii) **0.95** (iii) **0.97** (iv) **0.99**.

Table C.1 can *only* used on *standard* normal questions, where Z is $N(0, 1)$.

(g) The probability a 20 year old has an IQ between 125 and 135 is

$$P(125 < X_2 < 135) = P\left(\frac{125 - 120}{20} < Z_2 < \frac{135 - 120}{20}\right) = P\left(0.25 < Z_2 < 0.75\right) =$$

(i) **0.13** (ii) **0.17** (iii) **0.27** (iv) **0.31**.

```
pnorm(135,120,20) - pnorm(125,120,20) # normal P(125 < X < 135), mean = 120, SD = 20
pnorm(0.75) - pnorm(0.25) # standard normal P(0.25 < Z < 0.75)
```

```
[1] 0.1746663
```

Or, using Table C.1 from the text,

$$P(0.25 < Z_2 < 0.75) = \Phi(0.75) - \Phi(0.25) \approx 0.7734 - 0.5987 \approx$$

(i) **0.13** (ii) **0.17** (iii) **0.27** (iv) **0.31**.

(h) If a normal random variable X with mean μ and standard deviation σ can be transformed to a standard one Z with mean $\mu = 0$ and standard deviation $\sigma = 1$ using

$$Z = \frac{X - \mu}{\sigma},$$

then Z can be transformed to X using

$$X = \mu + \sigma Z.$$

(i) **True** (ii) **False**

(i) A 16 year old who has an IQ which is three (3) standards above the mean IQ has an IQ of $x_1 = 100 + 3(16) =$
(i) **116** (ii) **125** (iii) **132** (iv) **148**.

(j) A 20 year old who has an IQ which is two (2) standards below the mean IQ has an IQ of $x_2 = 120 - 2(20) =$
(i) **60** (ii) **80** (iii) **100** (iv) **110**.

(k) A 20 year old who has an IQ which is 1.5 standards below the mean IQ has an IQ of $x_2 = 120 - 1.5(20) =$
(i) **60** (ii) **80** (iii) **90** (iv) **95**.

(l) The probability a 20 year old has an IQ greater than one (1) standard deviation above the mean is

$$P(Z_2 > 1) = P(X_2 > 120 + 1(20)) = P(X_2 > 140) =$$

(i) **0.11** (ii) **0.13** (iii) **0.16** (iv) **0.18**.

```
1 - pnorm(1) # standard normal P(Z > 1), defaults to mean = 0, SD = 1
1 - pnorm(140,120,20) # normal P(X > 140), mean = 120, SD = 20
```

```
[1] 0.1586553
```

(m) *Percentile.* Determine 1st percentile, $\pi_{0.01}$, when X is $N(120, 20^2)$.

$$F(\pi_{0.01}) = \int_{-\infty}^{\pi_{0.01}} \frac{1}{20\sqrt{2\pi}} e^{-(1/2)[(y-120)/20]^2} = 0.01,$$

and so $\pi_{0.01} \approx$ (i) **72.47** (ii) **73.47** (iii) **75.47**

```
qnorm(0.01,120,20) # normal 1st percentile, mean = 120, SD = 20
[1] 73.47304
```

Or, using Table C.1 "backwards", $F(\pi_{0.01}) = 0.01$ when

$$Z \approx$$

(i) $\mathbf{-2.33}$ (ii) $\mathbf{-2.34}$ (iii) $\mathbf{-2.35}$

but since X is $N(120, 20^2)$, 1st percentile

$$\pi_{0.01} = X = \mu + \sigma Z \approx 120 + 20(-2.33) =$$

(i) **73.4** (ii) **74.1** (iii) **75.2**

The $\pi_{0.01}$ is that value where 1% of probability is at or below (to the left of) this value.

3.5 Functions of Continuous Random Variables

Assuming cdf $F(x)$ is *strictly* increasing (as opposed to just *monotonically* increasing) on $a < x < b$, one method to determine the pdf of a function, $Y = U(X)$, of random variable X, requires first determining the cdf of X, $F(x)$, then using $F(X)$ to determine the cdf of Y, $F(Y)$, and finally differentiating the result,

$$f(y) = F'(x) = \frac{dF}{du}.$$

The second fundamental theorem of calculus is sometimes used in this method,

$$\frac{d}{dy} \int_{u_1(y)}^{u_2(y)} f_X(x)\, dx = f_X(u_2(y)) \frac{du_2}{dy} - f_X(u_1(y)) \frac{du_1}{dy}.$$

Related to this is an algorithm for sampling at random values from a random variable X with a desired distribution by using the uniform distribution:

- determine cdf of X, $F(x)$, desired distribution
- find $F^{-1}(y)$: set $F(x) = y$, solve for $x = F^{-1}(y)$
- generate values $y_1, y_2, \ldots, y_n$ from Y, assumed to be $U(0, 1)$,
- use $x = F^{-1}(y)$ to simulate observed $x_1, x_2, \ldots, x_n$, from desired distribution.

Exercise 3.5 (Functions of Continuous Random Variables)

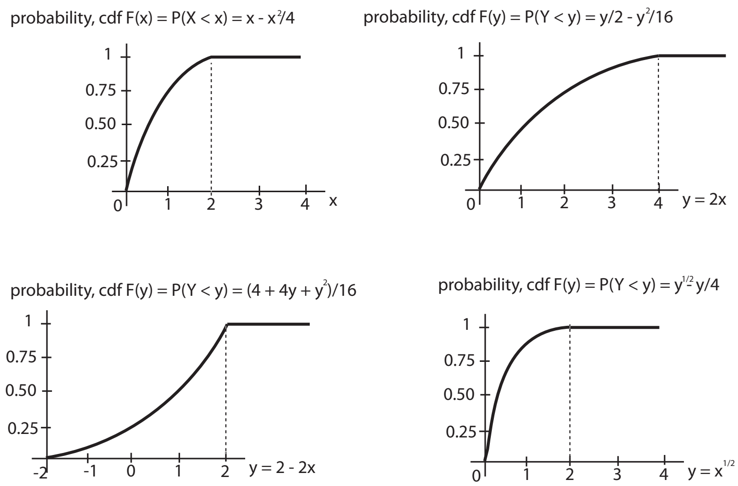

Figure 3.12: Distributions of various functions of $Y = U(X)$

1. *Determine pdf of $Y = U(X)$, given pdf of X.* Consider density

$$f_X(x) = \begin{cases} 1 - \frac{x}{2}, & 0 \leq x \leq 2 \\ 0 & \text{elsewhere} \end{cases}$$

   (a) *Calculate pdf of $Y = 2X$, version 1.*

$$\begin{aligned} F_X(x) &= \int_0^x \left(1 - \frac{t}{2}\right) dt = \left(t - \frac{t^2}{4}\right)_{t=0}^{t=x} = x - \frac{x^2}{4}, \\ F_Y(y) &= P(Y \leq y) = P(2X \leq y) = P\left(X \leq \frac{y}{2}\right) = F_X\left(\frac{y}{2}\right) \\ &= \frac{y}{2} - \frac{\left(\frac{y}{2}\right)^2}{4} = \frac{y}{2} - \frac{y^2}{16}, \\ f_Y(y) &= F'_Y(y) = \frac{dF}{dy} = \end{aligned}$$

   (i) $\frac{\mathbf{3}}{\mathbf{2}} - \frac{\mathbf{2y}}{\mathbf{32}}$ (ii) $\frac{\mathbf{1}}{\mathbf{2}} - \frac{\mathbf{2y}}{\mathbf{32}}$ (iii) $\frac{\mathbf{1}}{\mathbf{2}} - \frac{\boldsymbol{y}}{\mathbf{32}}$ (iv) $\frac{\mathbf{1}}{\mathbf{2}} - \frac{\boldsymbol{y}}{\mathbf{8}}$,

   where since $y = 2x$, $0 \leq x \leq 2$ implies $0 \leq \frac{y}{2} \leq 2$, or
   (i) $\mathbf{-1 \leq \boldsymbol{y} \leq 2}$ (ii) $\mathbf{-2 \leq \boldsymbol{y} \leq 1}$ (iii) $\mathbf{0 \leq \boldsymbol{y} \leq 3}$ (iv) $\mathbf{0 \leq \boldsymbol{y} \leq 4}$

(b) *Calculate pdf of $Y = 2X$, version 2.*

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \;=\; P(2X \leq y) \;=\; P\left(X \leq \frac{y}{2}\right) \\ &= F_X\left(\frac{y}{2}\right) \;=\; \int_0^{\frac{y}{2}} f_X(x)\,dx \\ f_Y(y) &= \frac{d}{dy}\int_0^{\frac{y}{2}} f_X(x)\,dx \;=\; f_X\left(\frac{y}{2}\right)\frac{d}{dy}\left(\frac{y}{2}\right) - 0 \;=\; \left(1 - \frac{y/2}{2}\right)\cdot\frac{1}{2} \;= \end{aligned}$$

(i) $\boldsymbol{\frac{3}{2} - \frac{2y}{32}}$ (ii) $\boldsymbol{\frac{1}{2} - \frac{2y}{32}}$ (iii) $\boldsymbol{\frac{1}{2} - \frac{y}{32}}$ (iv) $\boldsymbol{\frac{1}{2} - \frac{y}{8}}$,

Second fundamental theorem of calculus used here, notice cdf of $F_X(x)$ is not explicitly calculated.

(c) *Calculate pdf for $Y = 2 - 2X$.*

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \;=\; P(2 - 2X \leq y) \;=\; P\left(X \geq \frac{2-y}{2}\right) \\ &= \int_{\frac{2-y}{2}}^{2} f_X(x)\,dx, \\ f_Y(y) &= \frac{d}{dy}\int_{\frac{2-y}{2}}^{2} f_X(x)\,dx \;=\; 0 - f_X\left(\frac{2-y}{2}\right)\frac{d}{dy}\left(\frac{2-y}{2}\right) \\ &= -\left(1 - \frac{\frac{2-y}{2}}{2}\right)\cdot\frac{-1}{2} \;= \end{aligned}$$

(i) $\boldsymbol{\frac{3}{2} - \frac{2y}{32}}$ (ii) $\boldsymbol{\frac{1}{2} + \frac{2y}{32}}$ (iii) $\boldsymbol{\frac{1}{2} + \frac{y}{32}}$ (iv) $\boldsymbol{\frac{1}{4} + \frac{y}{8}}$,

where since $y = 2 - 2x$, $0 \leq x \leq 2$ implies $0 \leq \frac{2-y}{2} \leq 2$, or
(i) $\boldsymbol{-1 \leq y \leq 2}$ (ii) $\boldsymbol{-2 \leq y \leq 1}$ (iii) $\boldsymbol{-1 \leq y \leq 1}$ (iv) $\boldsymbol{-2 \leq y \leq 2}$

(d) *Calculate pdf for $Y = X^2$.*

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \;=\; P\left(X^2 \leq y\right) \;=\; P(X \leq \sqrt{y}) \;=\; \int_0^{\sqrt{y}} f_X(x)\,dx, \\ f_Y(y) &= \frac{d}{dy}\int_0^{\sqrt{y}} f_X(x)\,dx \;=\; f_X(\sqrt{y})\frac{d}{dy}(\sqrt{y}) - 0 \;=\; \left(1 - \frac{\sqrt{y}}{2}\right)\cdot\frac{1}{2\sqrt{y}} \;= \end{aligned}$$

(i) $\boldsymbol{\frac{3}{2} - \frac{2y}{32}}$ (ii) $\boldsymbol{\frac{1}{2\sqrt{y}} - \frac{1}{4}}$ (iii) $\boldsymbol{\frac{1}{\sqrt{y}} - \frac{1}{4}}$ (iv) $\boldsymbol{\frac{1}{2\sqrt{y}} + \frac{1}{4}}$,

Shorten up number of steps, omitted $F_X(\sqrt{y})$.

where $y = x^2$, $0 \leq x \leq 2$ implies $0 \leq \sqrt{y} \leq 2$, or
(i) $\boldsymbol{0 \leq y \leq 1}$ (ii) $\boldsymbol{0 \leq y \leq 2}$ (iii) $\boldsymbol{0 \leq y \leq 3}$ (iv) $\boldsymbol{0 \leq y \leq 4}$

Although there are two roots, $\pm\sqrt{y}$, only the positive root fits in the positive interval $0 \leq y \leq 2$.

2. *Simulate random sample from desired cdf $F(x)$ using uniform.*
Simulate a random sample of size 5 from X with desired cdf (not pdf),

$$F(x) = \begin{cases} 0, & x < 3 \\ 1 - \left(\frac{3}{x}\right)^3, & x \geq 3 \end{cases}$$

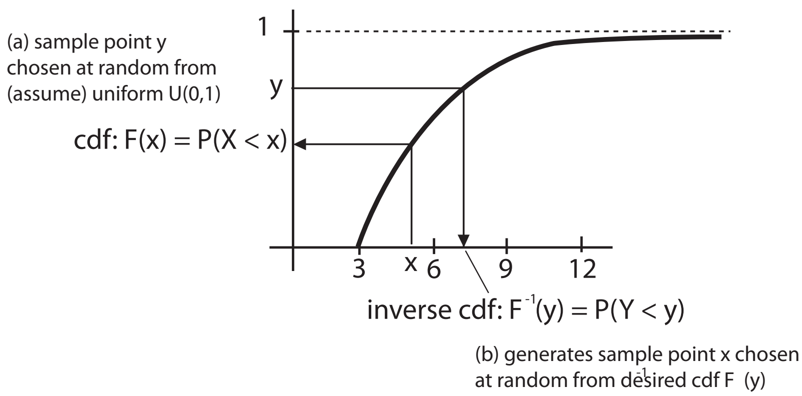

Figure 3.13: Distribution $y = F(x)$ and inverse $x = F^{-1}(y)$

(a) *Determine cdf of X, desired $F(x)$.*
It is given (does not need to be worked out in this case): $F(x) = 1 - \left(\frac{3}{x}\right)^3$

(b) *Find inverse cdf $F^{-1}(y)$.* Since $F(x) = 1 - \left(\frac{3}{x}\right)^3 = y$, then
$x = F^{-1}(y) =$ (i) $\boldsymbol{\left(\frac{3}{x}\right)^3}$ (ii) $\boldsymbol{1 - \left(\frac{9}{x}\right)^3}$ (iii) $\boldsymbol{1 - 3(1-y)^{-\frac{1}{3}}}$ (iv) $\boldsymbol{3(1-y)^{-\frac{1}{3}}}$

(c) *Generate values y_1, y_2, y_3, y_4, y_5 from Y, assumed to be $U(0,1)$.*
One possible random sample of size 5 from uniform on $U(0,1)$ gives

0.2189, 0.9913, 0.5775, 0.2846, 0.8068.

```
y <- runif(5); y
[1]  0.2189 0.9913 0.5775 0.2846 0.8068
```

(d) *Use $x = F^{-1}(y)$ to generate observed $x_1, x_2, \ldots, x_n$, from desired $F(x)$.*
Complete the following table.

y sample from $U(0,1)$	0.2189	0.9913	0.5775	0.2846	0.8068
$x = F^{-1}(y) = 3(1-y)^{-\frac{1}{3}}$	3.2575	14.586			

```
y <- c(0.2189,0.9913,0.5775,0.2846,0.8068)
x <- 3*(1-y)^{-1/3}; x

[1]  3.257510 14.586402  3.998027  3.354323  5.189421
```