Problem 1

A pair of fair dice is thrown. Let the random variable X denote the sum of the outcomes.

Table 1: Problem 1 Calculations for parts (a), (b), and (d).

Calculations											
X	2	3	4	5	6	7	8	9	10	11	12
pmf $p(x)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
cdf $F(x)$	0.028	0.083	0.167	0.278	0.417	0.583	0.722	0.833	0.917	0.972	1.000
$xp(x)$	0.056	0.167	0.333	0.556	0.833	1.167	1.111	1.000	0.833	0.611	0.333

(a) Graph the probability mass function $p(x)$.

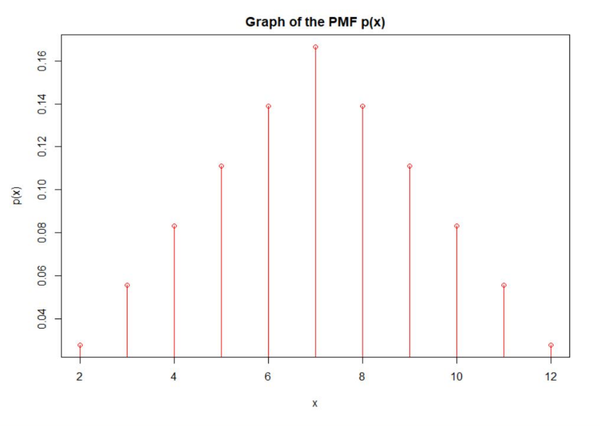


Figure 1: The probability mass function graph

R Code:

```
# Assign variable for all the possible outcomes from 2 to 12
outcomes <- 2:12
# pmf values based on the possible outcomes
pmf_x <- c(1/36, 2/36, 3/36, 4/36, 5/36, 6/36, 5/36, 4/36, 3/36, 2/36, 1/36)

# Compute cdf
cumm_sum <- cumsum(pmf_x)

require(graphics)
plot(outcomes, pmf_x, type="h", col=2, main="Graph of the PMF p(x)",
xlab="x", ylab="p(x)")
points(outcomes, pmf_x, col=2); abline(h=0,col=3)
```

(b) Graph the cumulative distribution function $F(x)$.

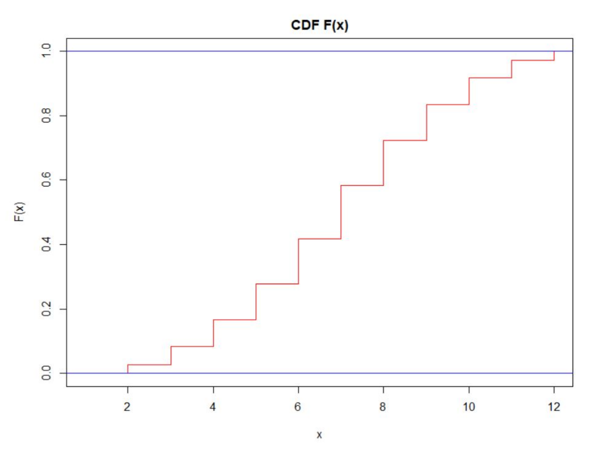


Figure 2: The cumulative distribution function graph

R Code:

```
plot(c(1, outcomes), c(0, cumm_sum), type="s", ylab="F(x)", col=3,
xlab="x", main="CDF F(x)"); abline(h = 0:1, col = 4)
```

(c) Generate four variates corresponding to the following $U(0,1)$'s (use the inverse-cdf technique):

Table 2: Generating variates using inverse-cdf technique.

u	0.495	0.762	0.927	0.002
X	7	9	11	2

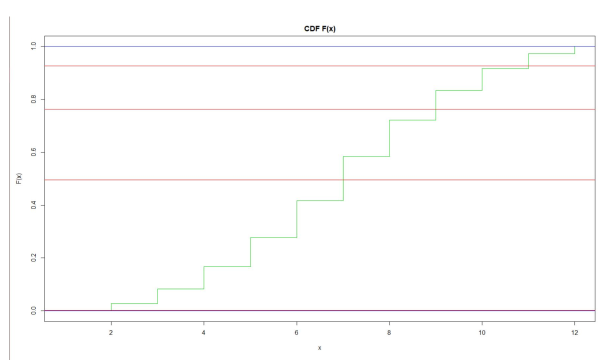


Figure 3: The cdf graph with horisontal lines for values of u

R Code:

```
plot(c(1, outcomes), c(0, cumm_sum), type="s", ylab="F(x)", col=3, xlab="x",
    main="CDF F(x)"); abline(h = 0:1, col = 4);
abline(h=c(0.002, 0.927, 0.762, 0.495), col=2)
```

(d) Find the population mean $\mu = E[X]$.
Using the calculations from the Table 1:

$$\mu = E[X] = \sum_{allx} xp(x) = 7.$$

Problem 2

Write an algorithm for generating a variate from a distribution with probability density function

$$f(x) = \begin{cases} x^3, & \text{for } 0 \leq x < 1 \\ 3/4, & \text{for } 1 \leq x \leq 2. \end{cases}$$

When $0 \leq x < 1$, $F(x) = \int_0^x m^3 dm = \frac{1}{4}x^4$.
Then, to find $F^{-1}(x)$ we need to solve for x:

$$Y = \frac{1}{4}x^4$$
$$X = (4Y)^{1/4}, \qquad 0 \leq y < \frac{1}{4}.$$

When $1 \leq x \leq 2$,

$$F(x) = F(1) + \int_1^x \frac{3}{4} dm = \frac{1}{4} + \frac{3}{4}m\Big|_1^x = \frac{1}{4} + \frac{3}{4}x - \frac{3}{4} = \frac{3}{4}x - \frac{1}{2} = \frac{3x-2}{4}.$$

Then, to find $F^{-1}(x)$ we need to solve for x:

$$Y = \frac{3x-2}{4}$$
$$X = \frac{4Y-2}{3}, \qquad \frac{1}{4} \leq y \leq 1.$$

Once we have the equations for X we can develop and algorithm for generating a variate from the given distribution.

1. Generate random number (y) from $U(0, 1)$
2. Check which range this number falls into:
   if the random number is in $[0, \frac{1}{4})$ range, then use $X = (4Y)^{1/4}$;
   else the random number is in the $[\frac{1}{4}, 1]$ range and need to use $X = \frac{4Y-2}{3}$.

Problem 3

The Pareto distribution is often used to model the distribution of incomes. Its probability density function is

$$f(x) = \frac{\alpha^{\beta}\beta}{x^{\beta+1}} \qquad x \geq \alpha,$$

where α and β are two positive parameters.

(a) Graph the probability density function $f(x)$ when $\alpha = 5$ and $\beta = 1.1$.

$$f(x) = \frac{5^1.1 * 1.1}{x^{1.1+1}} = 6.46/x^{2.1}$$

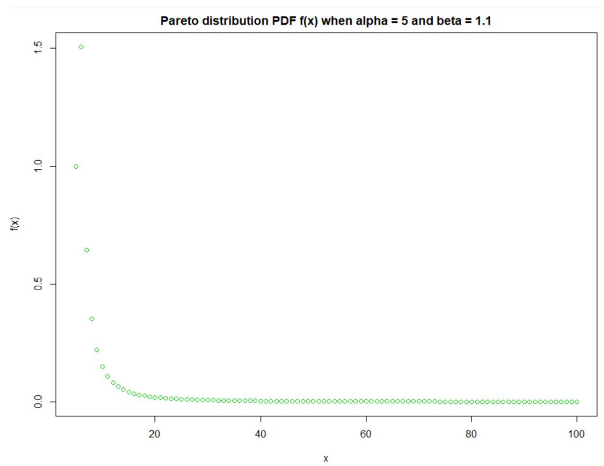


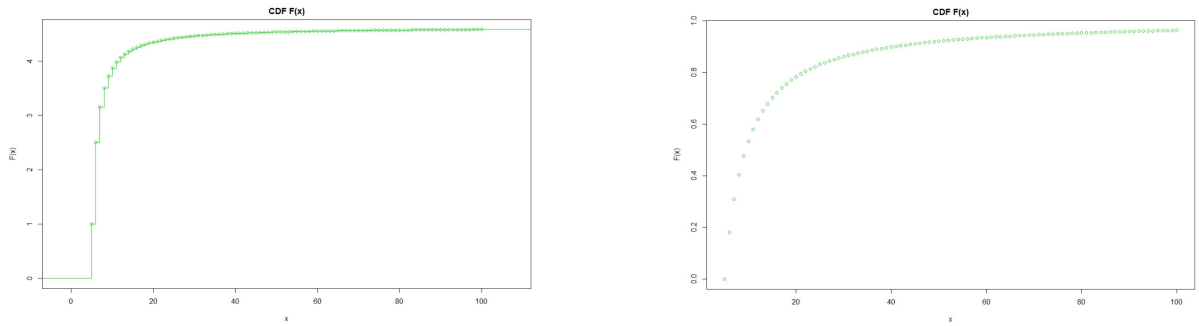
Figure 4: The Pareto pdf when $\alpha = 5$ and $\beta = 1.1$

R Code:

```
x <- 5:100
p <- rep(1, 96)
for (i in 2:96) p[i] = 6.46/i^2.1
plot(x, p,  ylab="f(x)", col=3, xlab="x", main="Pareto distribution PDF f(x)
    when alpha = 5 and beta = 1.1");
```

(b) Graph the cumulative distribution function $F(x)$ when $\alpha = 5$ and $\beta = 1.1$.

$$F(x) = \int_{\alpha}^{x} \frac{6.46}{m^{2.1}} dm = \frac{-6.46}{1.1} m^{-1.1} \Big|_{\alpha}^{x} = -\frac{6.46}{1.1}\left(x^{-1.1} - \alpha^{-1.1}\right) = -5.873\left(x^{-1.1} - \alpha^{-1.1}\right)$$
$$F(x) = -5.873\left(x^{-1.1} - \alpha^{-1.1}\right).$$



(a) CDF graph using the cumsum(p) from part (a) (b) CDF graph using the F(x) formula

Figure 5: CDF graph of Pareto distribution using two different approaches.

R Code using the cumsum function in R and p(x) from Problem 3 part (a):

```
F_x = c(0,cumsum(p))
plot(stepfun(x, F_x), ylab="F(x)", col=3, xlab="x", main="CDF F(x)")
```

R Code using the F(x) formula derived in Problem 3 part (b):

```
cdf_x <- -5.873*(x^(-1.1) - 5^(-1.1))
plot(x,cdf_x, ylab="F(x)", col=3, xlab="x", main="CDF F(x)")
```

(c) Show that this is a valid probability density function for any $\alpha > 0$ and $\beta > 0$.
Need to show: $\int f(x)dx = 1 => \int_\alpha^\infty \frac{\alpha^\beta \beta}{x^{\beta+1}}dx = 1$.

$$\int_\alpha^\infty \frac{\alpha^\beta \beta}{x^{\beta+1}}dx = \frac{\alpha^\beta \beta}{-\beta} x^{-\beta}\bigg|_\alpha^\infty =$$
$$= -\alpha^\beta \left(0 - \frac{1}{\alpha^\beta}\right) =$$
$$= 1.$$

$\therefore$ This is a valid pdf for any $\alpha > 0$ and $\beta > 0$.

(d) Find the cumulative distribution function $F(x)$.

$$F(x) = \int_\alpha^x \frac{\alpha^\beta \beta}{t^{\beta+1}}dt =$$
$$= \alpha^\beta \beta \left(-\frac{1}{\beta}\right)\left(t^{-\beta-1+1}\big|_\alpha^x\right) =$$
$$= -\alpha^\beta \left(x^{-\beta} - \alpha^{-1}\right) =$$
$$= 1 - \frac{\alpha^\beta}{x^\beta}, \qquad x \geq \alpha$$

Then,

$$F(x) = 1 - \frac{\alpha^\beta}{x^\beta}, \qquad x \geq \alpha$$

(e) Find $\mu = E[X]$ for any α and β.

$$\begin{aligned}\mu = E[X] = \int_\alpha^\infty xp(x)dx = \int_\alpha^\infty x\frac{\alpha^\beta \beta}{x^{\beta+1}}dx &= \\ = \alpha^\beta \beta \int_\alpha^\infty x^{-\beta}dx &= \\ = \frac{\alpha^\beta\beta}{1-\beta}\left(x^{1-\beta}\Big|_\alpha^\infty\right)&\end{aligned}$$

From here we can have two outcomes based on the values of β:

1. For $\beta < 1$ and $\alpha > 0$, $\mu = E[X] \to \infty$;
2. For $\beta \geq 1$ and $\alpha > 0$,

$$\begin{aligned}\mu = E[X] = &= \frac{\alpha^\beta\beta}{1-\beta}\left(x^{1-\beta}\Big|_\alpha^\infty\right) \\ &= \frac{\alpha^\beta\beta}{1-\beta}\left(-\alpha^{1-\beta}\right) = \\ &= \frac{\alpha^\beta\beta\alpha}{(\beta-1)\alpha^\beta} = \\ &= \frac{\alpha\beta}{\beta-1}.\end{aligned}$$

Then,

$$\mu = E[X] = \frac{\alpha\beta}{\beta-1}.$$

(f) Find $\sigma^2 = V[X]$ for any α and β.

$$\sigma^2 = V[X] = E(Y^2) - (E(Y))^2 = E(Y^2) - \left(\frac{\alpha\beta}{\beta-1}\right)^2$$

Step 1. Find $E(Y^2)$:

$$\begin{aligned}E(X^2) = \int_\alpha^\infty x^2p(x)dx = \int_\alpha^\infty x^2\frac{\alpha^\beta \beta}{x^{\beta+1}}dx &= \\ = \alpha^\beta \beta \int_\alpha^\infty \frac{x^2}{x^\beta x}dx &= \\ = \alpha^\beta \beta \int_\alpha^\infty x^{1-\beta}dx &= \\ = \frac{\alpha^\beta\beta}{2-\beta}\left(x^{2-\beta}\Big|_\alpha^\infty\right)&\end{aligned}$$

From here we can have two outcomes based on the values of β:

1. For $\beta < 2$ and $\alpha > 0$, $E[X^2] \to \infty$;

2. For $\beta \geq 2$ and $\alpha > 0$,

$$\begin{aligned}
E[X^2] &= \frac{\alpha^\beta \beta}{2-\beta}\left(x^{2-\beta}\Big|_\alpha^\infty\right) = \\
&= \frac{\alpha^\beta \beta}{2-\beta}\left(0-\alpha^{2-\beta}\right) = \\
&= \frac{\alpha^\beta \beta \alpha^2}{(\beta-2)\alpha^\beta} = \\
&= \frac{\alpha^2\beta}{\beta-2}.
\end{aligned}$$

Step 2. Find $\sigma^2 = V[X]$ for $\beta \geq 2$ and $\alpha > 0$:

$$\begin{aligned}
V(X) &= \frac{\alpha^2\beta}{\beta-2} - \left(\frac{\alpha\beta}{\beta-1}\right)^2 = \\
&= \frac{\alpha^2\beta(\beta-1)^2 - \alpha^2\beta^2(\beta-2)}{(\beta-2)(\beta-1)^2} = \\
&= \frac{\alpha^2\beta(\beta^2-2\beta+1) - \alpha^2\beta^2(\beta-2)}{(\beta-2)(\beta-1)^2} = \\
&= \frac{\alpha^2\beta^3 - 2\alpha^2\beta^2 + \alpha^2\beta - \alpha^2\beta^3 + 2\alpha^2\beta^2}{(\beta-2)(\beta-1)^2} = \\
&= \frac{\alpha^2\beta}{(\beta-2)(\beta-1)^2}.
\end{aligned}$$

Then,

$$\sigma^2 = V[X] = \frac{\alpha^2\beta}{(\beta-2)(\beta-1)^2}.$$

(g) Find the median of the distribution for any α and β.
We need to solve $\int_\alpha^m f(x)dx = \frac{1}{2}$ for m.
Step 1. Calculate the integral:

$$\begin{aligned}
\int_\alpha^m f(x)dx &= \int_\alpha^m \frac{\alpha^\beta\beta}{x^{\beta+1}}dx = \\
&= \frac{\alpha^\beta\beta}{-\beta}\left(x^{-\beta}\Big|_\alpha^m\right) = \\
&= -\alpha^\beta(m^{-\beta} - \alpha^{-\beta}) = \\
&= 1 - \frac{\alpha^\beta}{m^\beta}
\end{aligned}$$

Step 2. Solve for m :

$$\begin{aligned}
1 - \frac{\alpha^\beta}{m^\beta} &= \frac{1}{2} \\
\frac{\alpha^\beta}{m^\beta} &= \frac{1}{2} \\
m &= 2^{1/\beta}\alpha
\end{aligned}$$

Then,

$$m = \alpha 2^{1/\beta}$$

(h) Generate four incomes corresponding to the following $U(0, 1)$'s(use the inverse-cdf technique):
0.558 0.775 0.936 0.008
assuming that $\alpha = 8500$ and $\beta = 1.1$.

We will use the cdf function from Problem 3 part (d) to find its inverse $F^{-1}(x)$ and solve for x and then plug in the random variables in the inverse function:

$$F(x) = 1 - \frac{\alpha^\beta}{x^\beta}, \qquad x \ge \alpha$$

$$\begin{aligned} Y &= 1 - \frac{\alpha^\beta}{x^\beta} \\ \frac{\alpha^\beta}{x^\beta} &= 1 - Y \\ x^\beta &= \frac{\alpha^\beta}{1-y} \\ x &= \frac{\alpha}{(1-y)^{1/\beta}} \end{aligned}$$

R Code to calculate the x values:

```
alpha = 8500
beta = 1.1
y = c(0.558, 0.775, 0.936, 0.008)

# use the inverse-cdf formula from h)
for (i in y){
  x = alpha / (1- y)^(1/beta)
}

print(x)
# 17855.101  32987.083 103445.084   8562.294
```

Table 3: Generating incomes using inverse-cdf technique.

u	0.558	0.775	0.936	0.008
X	17855.101	32987.083	103445.084	8562.294

(i) Write an R program to generate 1001 Pareto random variates using the inverse-cdf technique, and estimate the median of the distribution. Compare your estimate with the theoretical value. Assume that $\alpha = 8500$ and $\beta = 1.1$.

R-code:

```
alpha = 8500
beta = 1.1
y = runif(1001)

# use the inverse-cdf formula from h)
x = alpha / (1- y)^(1/beta)
med = median(x)

# median formula from (g)
m = alpha * 2^(1/beta)
m   # 15961.83
med # 15926.95
```

Based on the output the two approaches give approximately same results. The median, obtained by generating random numbers and using the inverse-cdf formula obtained in Problem 3 part (h), will vary every time we re-generate new set of Pareto random variables. On the other hand, the median formula obtained in Problem 3 part (g) will not be affected by variations in variates.