APSTA Week 05 Exercises

1. Octrahedral Die (T)

Let T be random variable characterising the outcome of rolling a fair, octahedral die.
(a) Write down the probability mass function of T.
Solution:
The PMF is the function p that takes inputs $1, 2, ..., 7, 8$ and probability $\frac{1}{8}$ for all inputs.

(b) Determine the expected value and variance of T.
Solution:
Expected value:

$$\begin{aligned}
E[T] &= \sum_i a_i p(a_i) \\
&= 1 \cdot \frac{1}{8} + 2 \cdot \frac{1}{8} + 3 \cdot \frac{1}{8} + 4 \cdot \frac{1}{8} + 5 \cdot \frac{1}{8} + 6 \cdot \frac{1}{8} + 7 \cdot \frac{1}{8} + 8 \cdot \frac{1}{8} \\
&= \frac{1+2+3+4+5+6+7+8}{8} \\
&= \frac{36}{8} \\
&= 4.5
\end{aligned}$$

Variance:

$$\begin{aligned}
\mathrm{Var}(T) &= E[T^2] - (E[T])^2 \\
&= E[T^2] - (4.5)^2 \\
&= \left(\sum_i a_i^2 p(a_i) \right) - (4.5)^2 \\
&= \frac{1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2 + 7^2 + 8^2}{8} - (4.5)^2 \\
&= \frac{1 + 4 + 9 + 16 + 25 + 36 + 49 + 64}{8} - (4.5)^2 \\
&= \frac{204}{8} - (4.5)^2 \\
&= 25.5 - 20.25 \\
&= 5.25
\end{aligned}$$

2. Expected Value and Variance (T)

Let $X \sim U(1, 2)$.
(a) Compute the expected value of X.

Solution:

$$\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} xf(x)\ \mathrm{d}x \\
&= \int_1^2 x \cdot \frac{1}{2-1}\ \mathrm{d}x \\
&= \int_1^2 x\ \mathrm{d}x \\
&= \left[\frac{x^2}{2}\right]_1^2 \\
&= \frac{2^2}{2} - \frac{1^2}{2} \\
&= \frac{2^2 - 1^2}{2} \\
&= \frac{3}{2} \\
&= 1.5
\end{aligned}$$

(b) Compute the variance of X.
Solution:

$$\begin{aligned}
\mathrm{Var}(X) &= E[X^2] - (E[X])^2 \\
&= E[X^2] - \left(\frac{3}{2}\right)^2 \\
&= \int_1^2 x^2\ \mathrm{d}x - \left(\frac{3}{2}\right)^2 \\
&= \left[\frac{x^3}{3}\right]_1^2\ \mathrm{d}x - \left(\frac{3}{2}\right)^2 \\
&= \left(\frac{2^3}{3} - \frac{1^3}{3}\right) - \left(\frac{3}{2}\right)^2 \\
&= \frac{7}{3} - \left(\frac{3}{2}\right)^2 \\
&= \frac{7}{3} - \frac{3^2}{2^2} \\
&= \frac{7}{3} - \frac{9}{4} \\
&= \frac{1}{12}
\end{aligned}$$

(c) Let $f(x) = \mathrm{e}^x$. Compute the expected value of $f(X)$.
Solution:

$$\begin{aligned}
E[\mathrm{e}^X] &= \int_1^2 \mathrm{e}^x\ \mathrm{d}x \\
&= [\mathrm{e}^x]_1^2 \\
&= \mathrm{e}^2 - \mathrm{e}^1 \\
&= \mathrm{e}^2 - \mathrm{e}
\end{aligned}$$

3. Transformaing a Random Variable (T)

Given is a random variable X with the probability density function f given by $f(x) = 4x - 4x^3$ for $0 \leq x \leq 1$, and $f(x) = 0$, otherwise.
(a) Determine the distribution function F_X.
Solution:

$$\begin{aligned} F_X(b) &= \int_{-\infty}^{b} f(x)\ \mathrm{d}x \\ &= \int_{0}^{b} 4x - 4x^3\ \mathrm{d}x \\ &= [2x^2 - x^4]_0^b \\ &= (2b^2 - b^4) - (2 \cdot 0^2 - 0^4) \\ &= 2b^2 - b^4 \end{aligned}$$

(b) Let $Y = \sqrt{X}$. Determine the distribution function F_Y.
Solution:
Not sure how exactly to compute the changed variable. I've only seen a formula for computing the *expected value* of a changed variable, not the actual distribution function, and $\sqrt{4x - 4x^3}$ doesn't have an integral.

(c) Determine the probability density function of Y.
Solution:
I imagine it's just $\sqrt{f(x)} = \sqrt{4x - 4x^3}$.

4. Accessing Data and Numeric Summaries (R)

(a) Take `Cars93` (MASS) data set. What is the type of the Cylinders variable? What does the summary command do for the Cylinders variable? Get the names of the cars having 8 cylinders. What is the mean horsepower of the cars having 8 cylinders, how about standard deviation? How about those for the cars having 6 cylinders? Is the result what you expect?
Solution:

```
library(MASS)
cars <- Cars93
sprintf("Unique Cylinders: %s", paste(unique(cars$Cylinders), collapse = ', '))
```

```
## [1] "Unique Cylinders: 4, 6, 8, 3, rotary, 5"
```

```
print(summary(cars$Cylinders))
```

```
##      3      4      5      6      8 rotary
##      3     49      2     31      7      1
```

```
eight <- cars[cars$Cylinders == 8,]
sprintf("Mean HP of 8 cylinder cars is %s, and SD is %s", mean(eight$Horsepower), sd(eight$Horsepower))
```

```
## [1] "Mean HP of 8 cylinder cars is 234.714285714286, and SD is 54.4264466526899"
```

```
six <- cars[cars$Cylinders == 6,]
sprintf("Mean HP of 6 cylinder cars is %s, and SD is %s", mean(six$Horsepower), sd(six$Horsepower))
```

```
## [1] "Mean HP of 6 cylinder cars is 175.58064516129, and SD is 32.3334441855987"
```

(b) For the `precip` data set, find the mean and standard deviation of the rain fall over cities. Find all the cities with the average annual rain fall exceeding 50 inches. Which cities are the dryest? Does this match your expectation?
Solution:

```
rain <- precip
sprintf("mean: %s, sd: %s", mean(rain), sd(rain))
```

```
## [1] "mean: 34.8857142857143, sd: 13.7066500914256"
```

```
sprintf("big boy cities: %s", paste(names(rain[rain > 50]), collapse = ", "))
```

```
## [1] "big boy cities: Mobile, Juneau, Jacksonville, Miami, New Orleans, San Juan"
```

```
sprintf("cringe dry cities: %s", paste(names(sort(rain)[1:5]), collapse = ", "))
```

```
## [1] "cringe dry cities: Phoenix, Reno, Albuquerque, El Paso, Boise"
```

(c) The `rivers` contains the lengths of the 141 major rivers in North America. Compare the mean and 25% trimmed mean on the data set. What does the result tell you? How big is the standard deviation?
Solution:

```
walter <- rivers
mn <- mean(walter)
mn25 <- mean(walter, trim = 0.25)
sprintf("Mean: %s. 25%% trimmed mean: %s", mn, mn25)
```

```
## [1] "Mean: 591.184397163121. 25% trimmed mean: 449.915492957746"
```

```
print("Trimmed mean is way lower, which means we have some outliers bringing the mean really high up.")
```

```
## [1] "Trimmed mean is way lower, which means we have some outliers bringing the mean really high up."
```

```
print(tail(sort(walter), 10))
```

```
##  [1] 1270 1306 1450 1459 1770 1885 2315 2348 2533 3710
```

```
print("look at these long ass rivers")
```

```
## [1] "look at these long ass rivers"
```

```r
sprintf("standard deviation is %s", sd(walter))
```

```
## [1] "standard deviation is 493.87084203459"
```

```r
print("thas crazy, its almost as higher than the damn trimmed mean")
```

```
## [1] "thas crazy, its almost as higher than the damn trimmed mean"
```

5. Simulation (R)

Let X be a continuous random variable with the probability density function

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

a) Derive the distribution function of $f(x)$.
Solution:

$$\begin{aligned} F_X(b) &= \int_{-\infty}^{b} f(x)\ \mathrm{d}x \\ &= \int_{0}^{b} 2x\ \mathrm{d}x \\ &= [x^2]_0^b \\ &= b^2 \end{aligned}$$

(b) Describe the principle how you can simulate the random variables following the probability density $f(x)$ when you have an access to a random number generator of uniformly distributed random numbers.
Solution:
First we find the inverse, F^{inv}.

$$\begin{aligned} F(x) = u &\Leftrightarrow x^2 = u \\ &\Leftrightarrow \sqrt{x^2} = \sqrt{u} \\ &\Leftrightarrow x = \sqrt{u} \end{aligned}$$

That means that the random variable X is defined by

$$X = F^{\text{inv}}(U) = \sqrt{U}$$

(c) Write a program that uses stochastic simulation to draw 100 independent samples from $f(x)$.

```r
samples <- sqrt(runif(100))
# they did not tell me to plot them.
```