Homework_02

September 10, 2019

```
In [1]: from prob140 import *
        from datascience import *
        import numpy as np
        from scipy import special

        import matplotlib.pyplot as plt
        %matplotlib inline
        import matplotlib
        matplotlib.style.use('fivethirtyeight')
```

1 Homework 2

1.1 1. Extrema and Tails

The maximum and minimum of a random sample of numbers are called the "extrema" of the sample. Distributions of extrema are best described using the left or right hand tail probabilities. In this exercise you will see how.

Fix positive integers n and N. Suppose n draws are made at random with replacement from the numbers $\{1, 2, 3, \ldots, N\}$. Let X_i be the number that appears on the ith draw.

Let $V_n = \min\{X_1, X_2, \ldots, X_n\}$ be the sample minimum and let $W_n = \max\{X_1, X_2, \ldots, X_n\}$ be the sample maximum.

a) The event that a sample maximum is "small" is straightforward to describe in terms of the individual elements of the sample. To see this, fill in the blank with an appropriate mathematical symbol or English phrase. Justify your answer.

Fix an integer k such that $1 \le k \le N$. The event "$W_n \le k$" is the same as the event "each of $X_1, X_2, \ldots, X_n$ is ______________________ k".

It might help to draw the number line, mark the integers 1 through N, and put a special mark on k. For the maximum to be at or to the left of k, where do all the X's have to be?

b) For $1 \le k \le N$, use Part **a** to find $P(W_n \le k)$.

c) Use Part **b** to find $P(W_n = k)$ for $1 \le k \le N$ and to show algebraically that $\sum_{k=1}^{N} P(W_n = k) = 1$.

d) Modify Parts **a** through **c** to find the distribution of the sample minimum, as follows. For the event that the sample minimum is "large", fill in the blank with an appropriate mathematical symbol or English phrase.

The event "$V_n \ge k$" is the same as the event "each of $X_1, X_2, \ldots, X_n$ is ______________________ k".

Use this observation and the ideas of the previous parts to find $P(V_n = k)$ for $1 \le k \le N$.

e) Let k and m be integers such that $1 \le k \le m \le N$. Find $P(V_n \ge k \mid W_n \le m)$. Compare with the unconditional probability $P(V_n \ge k)$ that you used in Part **d** and show how the comparison indicates an intuitive reasoning for the conditional probability.

1.1.1 [Solution] Extrema and Tails

a) $\le$ because $X_i \le W_n$ for each i.

b) Apply **a** and ask yourself, "Where does X_1 have to be? Then where does X_2 have to be?" And so on. $P(X_1 \le k) = \frac{k}{N}$ and the X_i's are independent, so the chance that they are all $\le k$ is $\left(\frac{k}{N}\right)^n$.

c) $P(W_n = k) = \left(\frac{k}{N}\right)^n - \left(\frac{k-1}{N}\right)^n$. See Example ??? of Sec ??.

$\sum_{k=1}^{N} P(W_n = k) = \left[\left(\frac{1}{N}\right)^n - \left(\frac{0}{N}\right)^n\right] + \left[\left(\frac{2}{N}\right)^n - \left(\frac{1}{N}\right)^n\right] + \left[\left(\frac{3}{N}\right)^n - \left(\frac{2}{N}\right)^n\right] + \cdots + \left[\left(\frac{N}{N}\right)^n - \left(\frac{N-1}{N}\right)^n\right] = \left(\frac{N}{N}\right)^n - \left(\frac{0}{N}\right)^n$ by cancellation.

d) As before, it might help to draw the number line to visualize the event "$V_n \ge k$". $\ge$ because $X_i \ge V_n$ for each i. So $P(V_n \ge k) = \left(\frac{N-k+1}{N}\right)^n$

$P(V_n = k) = \left(\frac{N-k+1}{N}\right)^n - \left(\frac{N-k}{N}\right)^n$.

e)

$$P(V_n \ge k \mid W_n \le m) = \frac{P(V_n \ge k, W_n \le m)}{P(W_n \le m)} = \frac{\left(\frac{m-k+1}{N}\right)^n}{\left(\frac{m}{N}\right)^n} = \left(\frac{m-k+1}{m}\right)^n$$

For the numerator, draw the number line yet again. For the event to occur, where does each of the X_i's have to be?

The answer has the same form as $P(V_n \ge k)$ in **d** but the upper limit N has been replaced by m. Given that all the X's were in the range $1, 2, \ldots, m$, their minimum behaves as though n independent draws are being made from $1, 2, \ldots, m$ instead of from $1, 2, \ldots, N$.

#newpage

1.1.2 2. Mirror Images

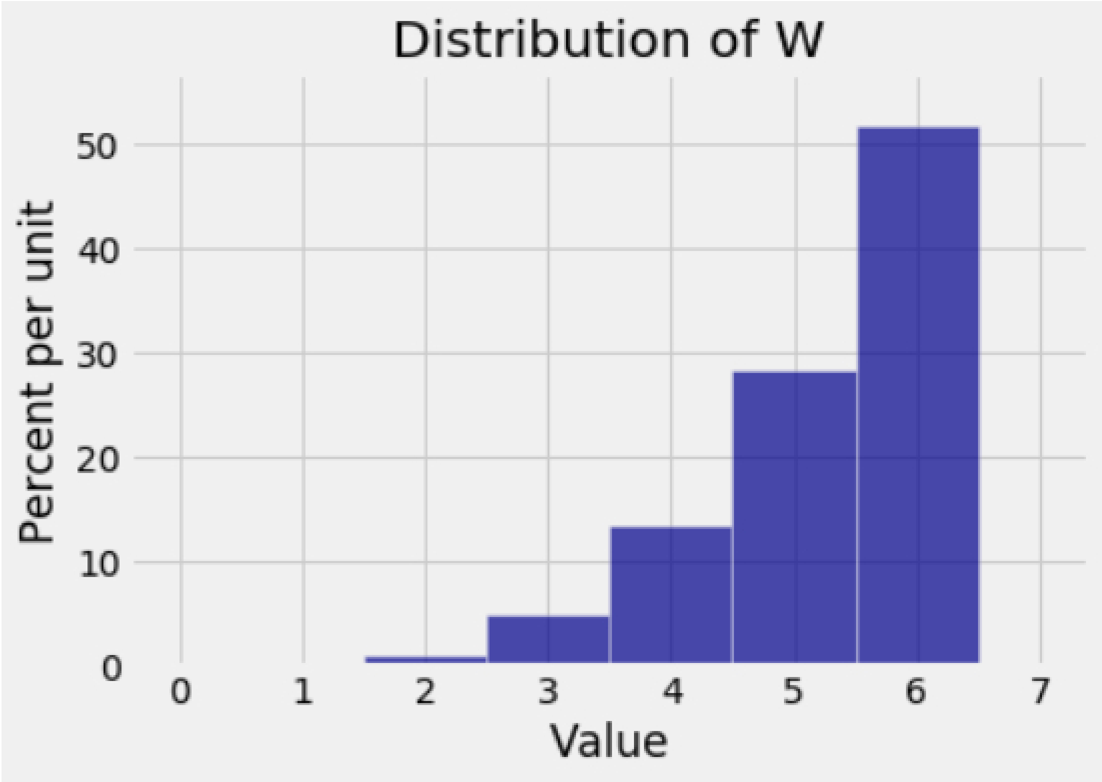
Let D_1, D_2, D_3, D_4 represent the numbers on four rolls of a die. Let $V = \min\{D_1, D_2, D_3, D_4\}$, $W = \max\{D_1, D_2, D_3, D_4\}$, $S = \sum_{i=1}^{4} D_i$.

a) If possible, use one of the symbols $\stackrel{d}{=}$ or $=$ to fill in the blank in D_1______D_2. Explain; and if both can be used, say why.

b) If possible, use one of the symbols $\stackrel{d}{=}$ or $=$ to fill in the blank in S______$4D_1$. Explain; and if both can be used, say why.

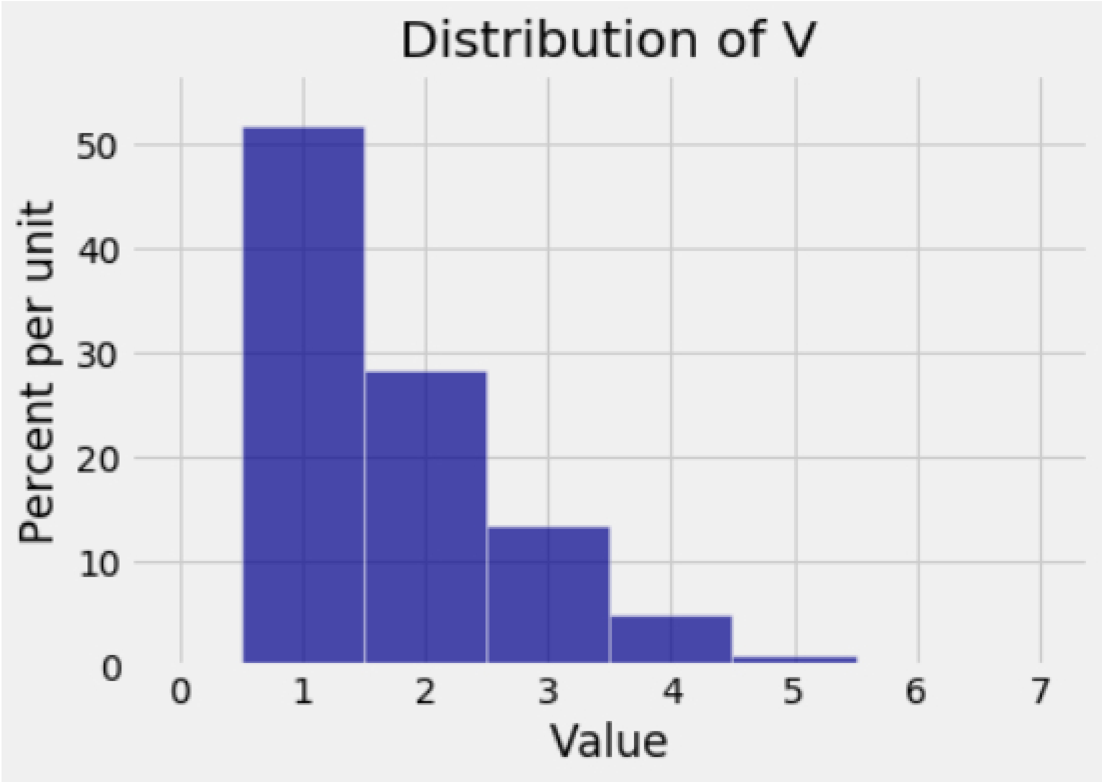
c) Use the previous exercise and the code cell below to draw the probability histogram of W. You will need some `prob140` methods which are used in Section 3.2 of the textbook.

```
In [2]: n = 4
        N = 6
        k = np.arange(1, N+1)

        # array consisting of P(W=k)
        probs_W = (k/N)**n - ((k-1)/N)**n

        dist_W = Table().values(k).probabilities(probs_W)
        Plot(dist_W)
        plt.title('Distribution of W');
```



d) Use the code cell below to draw the probability histogram of V.

In [3]:
```
# array consisting of P(V=k)
probs_V = ((N-k+1)/N)**n - ((N-k)/N)**n

dist_V = Table().values(k).probabilities(probs_V)
Plot(dist_V)
plt.title('Distribution of V');
```



e) Look at the two probability histograms above, and fill in the blanks with numbers. No explanation needed.

$P(V = 2) = P(W = \_\_\_\_)$ and $V \stackrel{d}{=} \_\_\_\_ - W$.

1.1.3 [Solution] Mirror Images

a) $\stackrel{d}{=}$ Both take values 1 through 6 with chance $1/6$ each. Can't use $=$ because the two rolls can come out differently.

b) Neither. S and $4D_1$ have different possible values. E.g. S can be 15 but $4D_1$ can't.

e) 5, 7.

#newpage

1.2 3. A Bound

Find a numerical value of k that makes the following statement true:

In k rolls of a fair die, there is at least 99% chance that all six faces will appear.

A good starting point: The complement is a union of which events? Can you bound the chance of a union and thus get an inequality to solve for k?

1.2.1 [Solution] A Bound

- Complement of "all six faces appear" is "at least one face doesn't appear"
- "at least one face doesn't appear" $= \bigcup_{i=1}^{6}$ {Face i doesn't appear}