Eurisko Assignment 35

Charlie Weinberger

April 7, 2021

1 35-1

Your friend is randomly stating positive integers that are less than some upper bound (which your friend knows, but you don't know). The numbers your friend states are as follows:

$$1, 17, 8, 25, 3$$

You assume that the numbers come from a **discrete uniform distribution** $U\{1, 2, \ldots, k\}$ defined as follows:

$$p_k(x) = \begin{cases} \frac{1}{k} & x \in \{1, 2, \ldots, k\} \\ 0 & x \notin \{1, 2, \ldots, k\} \end{cases}$$

(a) Compute the likelihood $P(\{1, 17, 8, 25, 3\} \mid k)$. Remember that the likelihood is just the probability of getting the result $\{1, 17, 8, 25, 3\}$ under the assumption that the data was sampled from the distribution $p_k(x)$. Your answer should be a piecewise function expressed in terms of k.

Solution:

$$P(\{1, 17, 8, 25, 3\} \mid k) = \begin{cases} \dfrac{1}{k^5} & k \geq 25 \\ 0 & \text{otherwise} \end{cases}$$

(b) Compute the posterior distribution by normalizing the likelihood. That is to say, find the constant c such that

$$\sum_{k=1}^{\infty} c \cdot P(\{1, 17, 8, 25, 3\} \mid k) = 1$$

Then, the posterior distribution will be

$$P(k \mid \{1, 17, 8, 25, 3\}) = c \cdot P(\{1, 17, 8, 25, 3\} \mid k)$$

SUPER IMPORTANT: You won't be able to figure this out analytically (i.e. just using pen and paper). Instead, you should write a Python script in assignment-problems/assignment_35_stats.py to approximate the sum by evaluating it for a very large number of terms. You should use as many terms as you need until the result appears to converge.

Solution:

$$\sum_{k=1}^{\infty} c \cdot P(\{1, 17, 8, 25, 3\} \mid k) = 1$$

$$c \cdot \sum_{k=1}^{\infty} P(\{1, 17, 8, 25, 3\} \mid k) = 1$$

$$c = \frac{1}{\sum_{k=1}^{\infty} P(\{1, 17, 8, 25, 3\} \mid k)}$$

$$c = 1443199.783177$$

That means

$$P(k \mid \{1, 17, 8, 25, 3\}) = c \cdot P(\{1, 17, 8, 25, 3\} \mid k)$$

$$P(k \mid \{1, 17, 8, 25, 3\}) = \begin{cases} \dfrac{1443199.783177}{k^5} & k \geq 25 \\ 0 & \text{otherwise} \end{cases}$$

(c) What is the most probable value of k? You can tell this just by looking at the distribution $p(k)$, but make sure to justify your answer with an explanation.

Solution:

The most probable value of k is 25. If k was less than 25, then the probability would be 0. If k was greater than 25, then there are more numbers that can be pulled. That means the probability of getting a 1, 3, 8, 17, and 25 will be smaller than if $k = 25$. Therefore, the most probable value of k is 25.

(d) The largest number in the dataset is 25. What is the probability that 25 is actually the upper bound chosen by your friend?

Solution:

$$P(25) = \frac{1}{1443199.783177 \cdot 25^5}$$

$$= 7.09534475 \cdot 10^{-14}$$

(e) What is the probability that the upper bound is less than or equal to 30?

Solution:

$$
\begin{aligned}
P(25 \leq k \leq 30) &= \sum_{25}^{30} p(x)\,\mathrm{d}x \\
&= \sum_{25}^{30} \frac{1}{1443199.783177k^5}\,\mathrm{d}x \\
&= \frac{1}{1443199.783177 \cdot 25^5} + \frac{1}{1443199.783177 \cdot 26^5} + \frac{1}{1443199.783177 \cdot 27^5} \\
&+ \frac{1}{1443199.783177 \cdot 28^5} + \frac{1}{1443199.783177 \cdot 29^5} + \frac{1}{1443199.783177 \cdot 30^5} \\
&= 2.8011908 \cdot 10^{-13}
\end{aligned}
$$

(f) Fill in the blank: you can be 95% sure that the upper bound is less than ___.

SUPER IMPORTANT: You won't be able to figure this out analytically (i.e. just using pen and paper). Instead, you should write another Python function in assignment-problems/assignment_35_stats.py to approximate value of k needed (i.e. the number of terms needed) to have $p(K \leq k) = 0.95$.

Solution:

I can be 95% sure that the upper bound is less than 52.

2 35-2

A joint distribution is a probability distribution on two or more random variables. To work with joint distributions, you will need to use multi-dimensional integrals.

For example, given a joint distribution $p(x, y)$, the distribution must satisfy

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x, y)\,\mathrm{d}x\,\mathrm{d}y = 1.$$

The probability that $(X, Y) \in [a, b] \times [c, d]$ is given by

$$P((X, Y) \in [a, b] \times [c, d]) = \iint_{[a,b]\times[c,d]} p(x, y)\,\mathrm{d}A,$$

or equivalently,

$$P(a < X \leq b,\, c < Y \leq d) = \int_c^d \int_a^b p(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

The expectations are

$$\mathrm{E}[X] = \int_c^d \int_a^b x \cdot p(x, y)\,\mathrm{d}x\,\mathrm{d}y,$$
$$\mathrm{E}[Y] = \int_c^d \int_a^b y \cdot p(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

The joint uniform distribution $\mathcal{U}([a, b] \times [c, d])$ is a distribution such that all points (x, y) have equal probability in the region $[a, b]\times[c, d]$ and zero probability elsewhere. So, it takes the form

$$p(x, y) = \begin{cases} k & (x, y) \in [a, b] \times [c, d] \\ 0 & (x, y) \notin [a, b] \times [c, d] \end{cases}$$

for some constant k.

(a) Find the value of k such that $p(x, y)$ is a valid probability distribution. Your answer should be in terms of a, b, c, d.

Solution:

$$
\begin{aligned}
1 &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} p(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
1 &= \int_{c}^{d}\int_{a}^{b} p(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
1 &= \int_{c}^{d}\int_{a}^{b} k\,\mathrm{d}x\,\mathrm{d}y \\
1 &= \int_{c}^{d}\left(kx\Big|_{a}^{b}\right)\,\mathrm{d}y \\
1 &= \int_{c}^{d}\left((b-a)k\right)\,\mathrm{d}y \\
1 &= (b-a)\int_{c}^{d} k\,\mathrm{d}y \\
1 &= (b-a)\left(ky\Big|_{c}^{d}\right) \\
1 &= (b-a)(d-c)\cdot k \\
k &= \frac{1}{(b-a)(d-c)}
\end{aligned}
$$

(b) Given that $(X,Y)\sim p$, compute $\mathrm{E}[X]$ and $\mathrm{E}[Y]$. You should get $\mathrm{E}[X] = \frac{a+b}{2}$ and $\mathrm{E}[Y] = \frac{c+d}{2}$.

Solution:

$$
\begin{aligned}
\mathrm{E}[X] &= \int_{c}^{d}\int_{a}^{b} x\cdot p(x,y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{c}^{d}\int_{a}^{b} \frac{x}{(b-a)(d-c)}\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{c}^{d}\left(\frac{x^2}{2(b-a)(d-c)}\Big|_{a}^{b}\right)\,\mathrm{d}y \\
&= \int_{c}^{d}\left(\frac{b^2}{2(b-a)(d-c)} - \frac{a^2}{2(b-a)(d-c)}\right)\,\mathrm{d}y
\end{aligned}
$$

$$
\begin{aligned}
&= \int_c^d \frac{b^2 - a^2}{2(b-a)(d-c)} \, \mathrm{d}y \\
&= \int_c^d \frac{b+a}{2(d-c)} \, \mathrm{d}y \\
&= \left. \frac{y(b+a)}{2(d-c)} \right|_c^d \\
&= \frac{d(b+a)}{2(d-c)} - \frac{c(b+a)}{2(d-c)} \\
&= \frac{(d-c)(b+a)}{2(d-c)} \\
\mathrm{E}[X] &= \frac{a+b}{2}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{E}[Y] &= \int_c^d \int_a^b y \cdot p(x,y) \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_c^d \int_a^b \frac{y}{(b-a)(d-c)} \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_c^d \left(\left. \frac{y}{(b-a)(d-c)} x \right|_a^b \right) \mathrm{d}y \\
&= \int_c^d \frac{y(b-a)}{(b-a)(d-c)} \, \mathrm{d}y \\
&= \int_c^d \frac{y}{d-c} \, \mathrm{d}y \\
&= \left. \frac{y^2}{2(d-c)} \right|_c^d \\
&= \frac{d^2}{2(d-c)} - \frac{c^2}{2(d-c)} \\
&= \frac{d^2 - c^2}{2(d-c)} \\
&= \frac{(d-c)(d+c)}{2(d-c)} \\
\mathrm{E}[Y] &= \frac{c+d}{2}
\end{aligned}
$$

(c) Geometrically, $[a,b] \times [c,d]$ represents a rectangle bounded by $x = a$, $x = b$, $y = c$, and $y = d$. What is the geometric interpretation of the point $(E[X], E[Y])$

in this rectangle?

Solution:

E$[X]$ is the average of a and b. E$[Y]$ is the average of c and d. That means $(E[X], E[Y])$ is the center of the rectangle.