Homework 1: Mathematical Statistics (MATH-UA 234)

Due 09/15 at the beginning of class on Gradescope

Problem 1 (Wasserman 1.19 (Bayes theorem))**.** *Suppose that 30 percent of computer owners use a Macintosh, 50 percent use Windows, and 20 percent use Linux. Suppose that 65 percent of the Mac users have succumbed to a computer virus, 82 percent of the Windows users get the virus, and 50 percent of the Linux users get the virus. We select a person at random and learn that her system was infected with the virus. What is the probability that they are a Windows user?*

Solution. Let M, W, L be the events that a user uses Macintosh, Windows, and Linux respectively, and let V be the event that a user has a virus.

From the problem we have

$$\mathbb{P}[M] = 0.3, \quad \mathbb{P}[W] = 0.5, \quad \mathbb{P}[L] = 0.2$$

and

$$\mathbb{P}[V|M] = 0.65, \quad \mathbb{P}[V|W] = 0.82, \quad \mathbb{P}[V|L] = 0.5.$$

Our goal is to compute $\mathbb{P}[W|V]$. By Bayes' theorem,

$$\mathbb{P}[W|V] = \frac{\mathbb{P}[V|W]\mathbb{P}[W]}{\mathbb{P}[V]}.$$

We already know the quantities for the numerator, and using the law of total probability,

$$\mathbb{P}[V] = \mathbb{P}[V|M]\mathbb{P}[M] + \mathbb{P}[V|W]\mathbb{P}[W] + \mathbb{P}[V|L]\mathbb{P}[L].$$

Putting everything together we find

$$\mathbb{P}[W|V] = \frac{(.82)(.5)}{(0.82)(0.5) + (0.5)(0.2) + (0.3)(0.65)} = 0.582.$$

Problem 2. *Suppose $\mathbb{P}[A] = 2/3$ and $\mathbb{P}[B^c] = 1/4$. Can A and B be disjoint (only one can occur)? Why or why not?*

Solution. Suppose A and B are disjoint. Then, by axiom 2,

$$\mathbb{P}[A \cup B] = \mathbb{P}[A] + \mathbb{P}[B] = \mathbb{P}[A] + (1 - \mathbb{P}[B^c]) = \frac{2}{3} + 1 - \frac{1/4}{=}\frac{17}{12} > 1.$$

We know that the probability of any event is at most one, so this is a contradiction.

Thus, A and B cannot be disjoint.

Problem 3. *Use the axioms of a probability distribution to show or answer the following:*

(a) *Show* $\mathbb{P}[\emptyset] = 0$.

---

problems with a textbook reference are based on, but not identical to, the given reference

(b) *Show* $A \subseteq B \implies \mathbb{P}[A] \leq \mathbb{P}[B]$.

(c) *Does* $A \subsetneq B$ *imply* $\mathbb{P}[A] < \mathbb{P}[B]$*?*

(d) *Show* $0 \leq \mathbb{P}[A] \leq 1$ *for all* A.

Solution.

(a) We have $\Omega = \Omega \cup \emptyset$ and $\Omega \cap \emptyset = \emptyset$. In other words, Ω and $\emptyset$ are disjoint. Thus, from axiom 3, $\mathbb{P}[\Omega \cup \emptyset] = \mathbb{P}[\Omega] + \mathbb{P}[\emptyset]$. But since $\Omega \cup \emptyset = \Omega$, we get $\mathbb{P}[\Omega] = \mathbb{P}[\Omega] + \mathbb{P}[\emptyset]$. In other words, $\mathbb{P}[\emptyset] = 0$.

(b) We can write $B = B \cap A + B \cap A^c$. Since $B \cap A$ and $B \cap A^c$ are disjoint, $\mathbb{P}[B] = \mathbb{P}[B \cap A] + \mathbb{P}[B \cap A^c]$. Note that $B \cap A = A$ since $A \subset B$. Thus, we get that $\mathbb{P}[B] = \mathbb{P}[A] + \mathbb{P}[B \cap A^c] \geq \mathbb{P}[A]$, where we have used that $\mathbb{P}[B \cap A^c] \geq 0$ (axiom 1).

(c) No. For instance, consider $\mathbb{P}$ corresponding to placing a dice on 1 every time. Then $\mathbb{P}[\{1\}] = \mathbb{P}[\{1, 2, 3\}] = 1$.

(d) Since $A \subseteq \Omega$, $\mathbb{P}[A] \leq \mathbb{P}[\Omega] = 1$. By axiom 1, we also have $\mathbb{P}[A] \geq 0$.

Problem 4. *Use the axioms of a probability distribution to answer the following:*

(a) *Describe a situation where we might have a sample space* Ω *such that* $\mathbb{P}[\{\omega\}] = 0$ *for all* $\omega \in \Omega$.

(b) *Explain what is wrong with the following proof that* $1 = 0$.

Proof. *By definition* $\cup_{\omega \in \Omega}\{\omega\} = \{\omega : \exists \omega' \in \Omega \text{ with } \omega \in \{\omega'\}\} = \Omega$. *Thus,*

$$\begin{aligned}
1 &= \mathbb{P}[\Omega] && \text{(Axiom 2)} \\
&= \mathbb{P}\Big[\bigcup_{\omega \in \Omega} \{\omega\}\Big] && \text{(definition of union)} \\
&= \sum_{\omega \in \Omega} \mathbb{P}[\{\omega\}] && \text{(Axiom 3)} \\
&= \sum_{\omega \in \Omega} 0 && \text{(assumption)} \\
&= 0 && \square
\end{aligned}$$

Solution.

(a) One situation is a uniform distribution on $[0, 1]$.

(b) Let $A_\omega = \{\omega\}$. Then $A_\omega \cap A_{\omega'} = 0$ if $\omega \neq \omega'$; i.e. the events are pairwise disjoint. Moreover,

$$\bigcup_{\omega \in \Omega} A_\omega = \{\omega' \in \Omega : \exists \omega : \omega \in A_{\omega'}\} = \Omega.$$

However, axiom 3 requires the events A_i to be countable. In the case above, the A_ω are not countable since Ω is not countable.

Note that the issue is not that Ω is uncountable. Indeed, if we had defined different events $A_1, A_2, \ldots$, then we can apply axiom 3, even though Ω is uncountable.

Problem 5 (Wasserman 1.22 (simulate coin fipping)). *Suppose we flip a coin n times and let p denote the probability of heads. Let X be the number of heads. We call X a binomial random variable, which is discussed in the next chapter. Intuition suggests that X will be close to np. To see if this is true, we can repeat this experiment many times and average the X values. Carry out a simulation and compare the average of the X's to np.*

Try this for $p = .3$ *and* $n = 10$, $n = 100$, *and* $n = 1000$ *repeating each experiment 100 and then 1000 times. Make a histogram of your results, or report the sample mean and variance.*

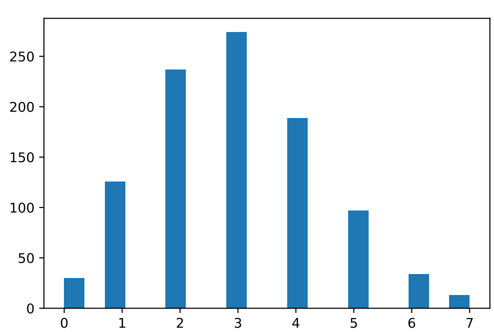
Solution.

Here is some sample code:

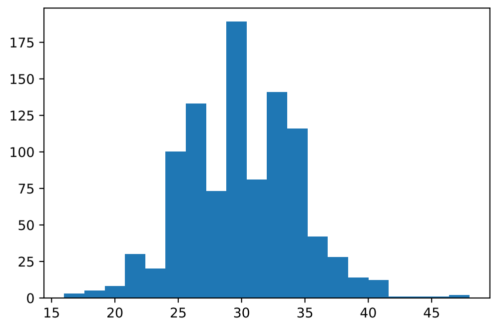
```
import numpy as np
import matplotlib.pyplot as plt

n = 100 # number of times we flip the coin
k = 1000 # numer of times we run experiment

Xs = [] # list of each X we generate

for j in range(k):
  X = 0
  for i in range(n):
    if np.random.rand() < .3:
      X = X+1

  Xs.append(X)

plt.hist(Xs)
```

In all cases, the value of X was near np. As n increased, the histogram become more like a normal random variable, and the variance was bigger too.

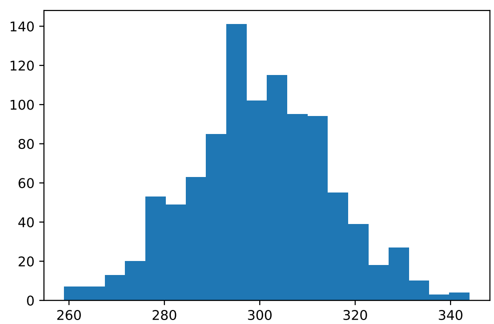
$n = 10$



$n = 100$



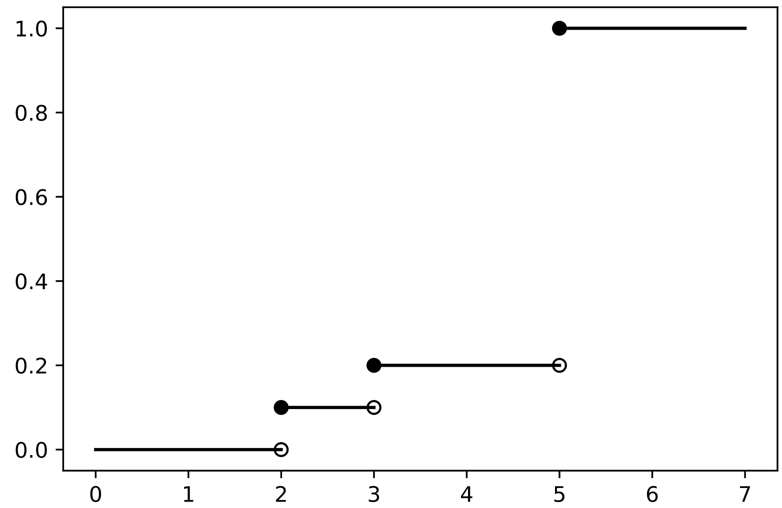
$n = 1000$



Problem 6 (Wasserman 2.2 (computing probabilities from distributions))**.** *Let X be such that* $\mathbb{P}[X = 2] = \mathbb{P}[X = 3] = 1/10$ *and* $\mathbb{P}[X = 5] = 8/10$.

Plot the CDF F for X and use the CDF to find $\mathbb{P}[2 < X < 4.8]$ *and* $\mathbb{P}[2 \leq X \leq 4.8]$.

Solution. We can make our CDF from the definition: $F(x) = \mathbb{P}[X \leq x]$.



From this, we find that $\mathbb{P}[2 < X < 4.8] = 1/10$ and $\mathbb{P}[2 \leq X \leq 4.8] = 2/10$.

Problem 7 (Wasserman 2.17 (conditional probabilities from distributions))**.** *Let*

$$f_{X,Y}(x,y) = \begin{cases} c(x+y^2) & 0 \leq x \leq 1 \textit{ and } 0 \leq y \leq 1 \\ 0 & \textit{otherwise.} \end{cases}$$

Find $\mathbb{P}[X < 1/2 | Y = 1/2]$*. Here c is a normalizing constant so that* $f_{X,Y}$ *is a probability density function.*

Solution.

This problem is asking us to compute $\mathbb{P}[X < 1/2 | Y = 1/2]$ given that we know the joint distribution $f_{X,Y}(x,y)$. This is pretty standard calculation but requires us to remember how to figure out the conditional density $f_{X|Y}(x|y)$. From the book, we see

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}.$$

We don't yet know what the marginal f_Y is but we can calculate this pretty easily by integrating out x:

$$f_Y(y) = \int f_{X,Y}(x,y)\mathrm{d}x$$

Thankfully our integration region is just a square $x \in [0,1]$ and $y \in [0,1]$ so these integrals are pretty easy. Doing the last one:

$$f_Y(y) = \int_0^1 f_{X,Y}\,\mathrm{d}x = \int_0^1 c(x+y^2)\,\mathrm{d}x = c/2 + cy^2.$$

Now is a good time to figure out c, since we know

$$\iint f_{X,Y}\,\mathrm{d}x\,\mathrm{d}y = 1$$

so

$$1 = \int f_Y(y)\,\mathrm{d}y = \int_0^1 (c/2 + cy^2)\,\mathrm{d}y = 5c/6$$

and therefore $c = 6/5$.

However, returning back to the original problem, we now know

$$f_{X|Y=y}(x) = \frac{f_{X,Y}}{f_Y} = \frac{c(x+y^2)}{c/2+cy^2}, \qquad x, y \in [0,1] \times [0,1]$$

so you'll note that the c's cancel out anyway. Now, we want $y = 1/2$, so we can plug that in and find

$$f_{X|Y=1/2}(x) = \frac{x + (1/2)^2}{1/2 + (1/2)^2}$$

and now we just integrate this to find the probability we want,

$$\mathbb{P}[X < 1/2 | Y = 1/2] = \int_0^{1/2} f_{X|Y=1/2}(x)\,\mathrm{d}x = \int_0^{1/2} \frac{x + (1/2)^2}{1/2 + (1/2)^2}\,\mathrm{d}x = 1/3.$$

Problem 8 (Wasserman 2.21 (important trick about iid maxes)). *Let $X_1, \dots, X_n \sim \mathrm{Exp}(\eta)$ be iid. Let $Y = \max\{X_1, \dots, X_n\}$. Find the PDF of Y. Hint: $Y \le y$ if and only if $X_i \le y$ for $i = 1, \dots, n$.*

Solution.

We have $Y_n = \max\{X_1, \dots, X_n\}$

Call $F_Y(y)$ the CDF of Y_n, so

$$\mathbb{P}[Y_n \le y] = F_Y(y).$$

If we knew this function, we are done, because the PDF is $F_Y'(y) = f_Y(y)$ and let's just exploit the independence of each of the X_i's by the hint in the problem

$$\begin{aligned}
F_Y(y) &= \mathbb{P}[Y_n \le y] \\
&= \mathbb{P}[\max\{X_1, \dots X_n\} \le y] \\
&= \mathbb{P}[X_1, \dots, X_n \le y]y && \text{(because all must be less than)} \\
&= \mathbb{P}[X_1 \le y]\mathbb{P}[X_2 \le y] \cdots \mathbb{P}[X_n \le y] && \text{(by independence)} \\
&= \mathbb{P}[X_1 \le y]^n && \text{(because all the } X_i \text{ are identical)} \\
&= F_{X_1}(y)^n
\end{aligned}$$

So the PDF we are after is, by the chain rule

$$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y}\left\{F_{X_1}(y)^n\right\} = n\left[F_{X_1}(y)\right]^{n-1} f_{X_i}(y).$$

We know $X_i \sim \exp(\beta)$ which has PDF and CDF

$$f_{X_i}(y) = \frac{1}{\beta}\exp\{-y/\beta\}, \qquad F_{X_i}(y) = \int_{-\infty}^{\infty} \frac{1}{\beta} e^{-s/\beta}\,\mathrm{d}s = 1 - e^{-y/\beta},$$

note that I'm using the book parameterization of the exponential distribution rather than the one Wikipedia uses. So plugging this in, we arrive at

$$f_Y(y) = n\left[F_{X_1}(y)\right]^{n-1} f_{X_i}(y) = n\left(1 - e^{-y/\beta}\right)^{n-1} \frac{1}{\beta} e^{-y/\beta}.$$

Of course, we can simplify this a bit, but it's not necessary.

Problem 9. *Look for an instance of one of the probability concepts we've seen in the course which you noticed in a different part of your life (e.g. in other classes, on the subway, at the park, on TV, etc.).*

In a few sentences, explain this instance and how it could be described mathematically.

Solution.

There are many possible answers you could put here. The first main goal of this problem is to get you thinking about how the concepts in the class appear throughout your daily life. I'm not looking for theoretical examples of where probability could be useful in real life. Instead, I want examples where you observed directly something relating to the class. The second main goal is to build practice with using words to describe mathematical ideas.