Problem set 2: More probability

Due October 9, 2023, at 9pm

(Your name here)

Note: Start with the file `ps2_2023_more_probability.qmd` (available from the github repository at https://github.com/UChicago-pol-methods/IntroQSS-F23/tree/main/assignments). Modify that file to include your answers. Make sure you can "render" the file (e.g. in RStudio by clicking on the `Render` button). Submit both the qmd file and the PDF via Canvas.

Problem 0: Bayes' Rule

One formulation of Bayes' Rule states that, if $\{A_1, A_2, A_3, ...\}$ is a partition of Ω and $B \in S$ with $P(B) > 0$,

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_i P(B \mid A_i)P(A_i)}$$

Here is a proof that is missing explanations for the steps:

Step 1:

$$P(A_i \cap B) = P(A_i \mid B)P(B) = P(B \mid A_i)P(A_i)$$

Step 2:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{P(B)}$$

Step 3:

$$P(A_i \mid B) = \frac{P(B \mid A_i)P(A_i)}{\sum_i P(B \mid A_i)P(A_i)}$$

Explain each step in the proof: what definition(s)/rule(s)/law(s)/axiom(s)/condition(s)/mathematical operation(s) is the proof relying on?

Answer:

Problem 1: Error rates in hypothesis testing

You have a fancy device that tests null hypotheses. Null hypotheses are statements about the world that can be either true or false. The device is designed to turn red when a null hypothesis is false and green when it is true, but it doesn't work perfectly: when a null hypothesis is false it turns red with probability 4/5 (i.e. it mistakenly turns green with probability 1/5), and when a null hypothesis is true it turns green with probability 19/20 (i.e. it mistakenly turns red with probability 1/20). Tests of different null hypotheses are independent, and 9/10 of the null hypotheses you test are true.

(1a) If you test 10 true null hypotheses in a row, what is the probability that the alarm turns red at least once? Explain your solution with reference to any axioms/definitions/rules/laws of probability you use.

Answer:

(1b) Write a simulation to check your answer to (1a). That is, use `R` to generate many draws according to the random process described (testing ten true null hypotheses in a row), and confirm that the proportion of draws with at least one red light is approximately the same as in your answer above.

```
# your code here
```

(1c) What is the probability of getting a red light in any given test (i.e. when you don't know if the null hypothesis is true)? Explain your solution with reference to any axioms/definitions/rules/laws of probability you use.

Answer:

(1d) If the light turns red in a given test, what is the probability that the null hypothesis is false? Explain your solution with reference to any axioms/definitions/rules/laws of probability you use.

Answer:

(1e) Write a simulation to check your answer to (1c) and (1d). That is, use `R` to generate many draws according to the random process described (testing null hypotheses), and confirm your answer about the proportion of red-light-producing draws (1c) and the proportion of red-light-producing draws in which the null hypothesis is false (1d).

```
# your code here
```

Problem 2: discrete random variables

Consider tossing a coin four times. The results of the coin flips are assumed to be independent. Let H and T denote the outcome that it produces a head or a tail, respectively.

(2a) Write the sample space Ω for this random process.

Answer:

$$\Omega = \{list, outcomes, \\ here\}$$

(2b) Are each of the outcomes in Ω equally likely to occur? Why or why not?

Answer:

(2c) Let the random variable X denote the number of heads in four tosses of the coin. In each toss, the probability of getting a heads is q.

Write the PMF $f(x)$ by replacing the "?"'s in the expression below. Use `R` to confirm that for a given value of q (choose one!), $\sum_{x=0}^{4} f(x) = 1$.

Answer:

$$f(x) = \begin{cases} ? & x = 0 \\ ? & x = 1 \\ ? & x = 2 \\ ? & x = 3 \\ ? & x = 4 \\ 0 & \text{otherwise} \end{cases}$$

```
# your code confirming that it sums to 1 for a value of q
```

(2d) Write the CDF $F(x)$ by replacing the "?"'s in the expression below.

Answer:

$$\Pr[X \leq x] = F(x) = \begin{cases} ? & x < 0 \\ ? & 0 \leq x < 1 \\ ? & 1 \leq x < 2 \\ ? & 2 \leq x < 3 \\ ? & 3 \leq x < 4 \\ ? & x \geq 4 \end{cases}$$

(2e) Assume $q = 2/3$. Use the `sample()` function in `R` to draw a large number of samples from the PMF you specified above, and confirm that $F(3)$ agrees with your answer from (2d).

Answer:

```
# your code here
```

Problem 3: continuous random variables

(3a) Let X be uniformly distributed between -5 and 3. Compute $\Pr[X < 1]$ and $\Pr[-3 < X < 1/2]$ analytically (e.g. by computing the length and height of the area to be integrated) and confirm your results using `R`.

Answer:

Confirmation in `R`:

```
# your code here
```

(3b) Let X be normally distributed with mean 0 and standard deviation 1. Using `R`, compute $\Pr[-1.96 < X < 1.96]$.

Answer:

```
# your code here
```

(3c) Let X be normally distributed with mean 0 and standard deviation 2. Using `R`, use `dnorm()` to approximate $\Pr[-.1 < X < 0]$, and use `pnorm()` to see how close your approximation is. Explain why the two values should be close but not exactly the same.

Answer:

```
# your code here
```