Problem set 1: Probability

Due October 7, 2024, at 10am

(Your name here)

NOTE: Start with the file `ps1_2024_probability.qmd` (available from the github repository at https://github.com/UChicago-pol-methods/IntroQSS-F24/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 knitted PDF via Canvas.

Problem 1: joint probability, conditional probability, independence

Let A and B be two events that could result from a random process.

Suppose the joint probabilities are:

Result	Probability
$A \cap B$	1/4
$A^C \cap B$	1/12
$A \cap B^C$	1/6
$A^C \cap B^C$	1/2

(1a) What are $P(A)$ and $P(B)$, i.e. the marginal probabilities of A and B?

Answer:

(1b) What is $P(A|B)$, i.e. the conditional probability of A given B?

Answer:

(1c) Suppose the random process is the selection of a student at random from a school. Considering the joint probability table above, what might events A and B be? Consider, for example, "the student eats carrots in his/her lunch", "the student can read", "the student has at least one sibling", "the student plays with Lego every day", "the student can tie his/her shoes", "the student wears a hat to school". You may want to specify what kind of school you have in mind.

Answer:

(1d) Suppose events A and B had the same marginal probabilities you reported in (1a), but the events were independent. What then would be the joint probabilities? Fill out the table below. Make sure the probabilities add up to 1.

Answer:

Result	Probability
$A \cap B$	
$A^C \cap B$	
$A \cap B^C$	
$A^C \cap B^C$	

Problem 2: `R` coding

First set your seed to 123 so that our answers are comparable.

```r
set.seed(123)
```

(The seed determines the output of "random" processes in R, so that if two students use the same function after setting the same seed, they will get the same answer. Set the seed once at the beginning of your code; do not set it again later in the code.)

(2a) Create a vector of length 1000 that could be a sample from the marginal distribution of event A in the previous question, where "A" indicates that event A occurred and "!A" indicates that it did not occur. Store this vector in a variable called `A_vec`. Report the proportion of times that event A occurred in this sample.

Answer:

```r
# your code goes here.
```

(2b) Do the same for event B. That is, create a vector of length 1000 that could be a sample from the marginal distribution of event B in the previous question, where "B" indicates that event B occurred and "!B" indicates that it did not occur. Store this vector in a variable called `B_vec`. Report the proportion of times that event B occurred in this sample.

Answer:

(2c) Given how `A_vec` and `B_vec` are created, A and B are independent events. This implies that $\mathrm{P}(A \mid B) = \mathrm{P}(A)$, $\mathrm{P}(B \mid A) = \mathrm{P}(B)$, and $\mathrm{P}(A \cap B) = \mathrm{P}(A) \times P(B)$. Confirm that this is approximately true in the samples `A_vec` and `B_vec`.

Answer:

(2d) Repeat (2a)-(2c) where `A_vec` and `B_vec` now have length 1 million instead of 1000. You should find that the equivalencies in (2c) are more nearly true in these larger samples. (We will later investigate this in the law of large numbers.) Is it the case here?

Answer: