

Problem set 2: More probability

Due October 9, 2023, at 9pm

(Your name here)

NOTE: Start with the file `ps2_2023_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 $\{A_1, A_2, A_3, \dots\}$ is a partition of Ω and $B \in \mathcal{S}$ with $P(B) > 0$,

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

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

Step 1:

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

Step 2:

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

Step 3:

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

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:

Step 1 is stating the *multiplicative law of probability* (sometimes known as the *product rule*), which is just a rearrangement of the *definition of conditional probability*. For any two events A and B , you can state the law either way: $P(A \cap B) = P(A | B)P(B)$ or $P(A \cap B) = P(B | A)P(A)$, so the two products must be equal.

Step 2 is rearranging Step 1 by dividing the second and third expressions by $P(B)$. This relies on the condition that $P(B) > 0$.

Step 3 is restating $P(B)$ using the *law of total probability*. This relies on the condition that $\{A_1, A_2, A_3, \dots\}$ is a partition of Ω .

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:

Call the event of rejecting at least one null hypothesis A . That's the probability we want to know. There are many ways for A to occur – e.g. reject every time, reject the first null hypothesis but not thereafter, etc – and it may be tedious to list them and compute the probability of each one. But there is only one way for A^C , the complement of A , to occur: on every test, the light turns green. The **complement rule** says that $P(A) = 1 - P(A^C)$. So what is $P(A^C)$? Given that the null hypothesis is true, the probability of the light turning green a single test is given as $19/20$. The tests are independent, so by the **definition of independence** the probability of ten green lights in a row is $(19/20)^{10}$. Therefore the probability of the light turning red at least once is $1 - (19/20)^{10} \approx .401$.

$$1 - (1 - .05)^{10}$$

[1] 0.4012631

(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.

```
# using the rbinom function, which gives us n results of the experiment,
# i.e. n counts of red lights in each experiment
tests <- rbinom(n = 1000000, size = 10, prob = .05)
mean(tests > 0)
```

[1] 0.4013101

```
# can also do it using sample
# here is a large number of tests -- 1 means red light when null is true
red_lights <- sample(x = c(1, 0), size = 1000000, prob = c(.05, .95), replace = T)
# we organize them in a big matrix with 10 columns,
# so that each row is a set of 10 tests
reshaped_red_lights <- matrix(red_lights, ncol = 10)
# we use apply() to get the sum in each row --
# the number of red lights in 10 tests
reds_per_row <- apply(reshaped_red_lights, 1, sum)
mean(reds_per_row > 0)
```

[1] 0.40321

(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: Let R denote the event that the light turns red, and let H_0 denote the event that the null hypothesis is true (with H_0^C denoting the event that the null hypothesis is false). Using the **law of total probability**, we have $P(R) = P(R \mid H_0)P(H_0) + P(R \mid H_0^C)P(H_0^C) = \frac{1}{20} \cdot \frac{9}{10} + \frac{4}{5} \cdot \frac{1}{10} = \frac{9}{200} + \frac{4}{50} = \frac{9}{200} + \frac{16}{200} = \frac{25}{200} = \frac{1}{8}$.

(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: Here we use **Bayes Rule**, which can be derived from the definition of **conditional probability** or the **product rule**:

$$P(H_0^C | R) = \frac{P(R | H_0^C)P(H_0^C)}{P(R)}$$

Using our answer for $P(R)$ above, we have $(4/5)(1/10)/(1/8) = 32/50 = 16/25$.

(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).

```
# first a version with no randomness -- we make a joint distribution
# that exactly matches the conditions in the problem
ntrue <- 9000
nfalse <- 1000
null_is_true <- c(rep(T, ntrue), rep(F, nfalse))
red_light <- c(rep(T, ntrue*(1/20)), rep(F, ntrue*(19/20)),
               rep(T, nfalse*(4/5)), rep(F, nfalse*(1/5)))
# confirming 1c
mean(red_light)
```

```
[1] 0.125
```

```
# confirming 1d
mean(!null_is_true[red_light])
```

```
[1] 0.64
```

```
# now a version with randomness
n <- 100000
null_is_true <- sample(c(F, T), size = n, replace = T, prob = c(.1, .9))
red_light <- rep(NA, n)
red_light[null_is_true] <- sample(c(F, T), size = sum(null_is_true),
                                replace = T, prob = c(.95, .05))
red_light[!null_is_true] <- sample(c(F, T), size = sum(!null_is_true),
                                replace = T, prob = c(.2, .8))
# 1c
mean(red_light)
```

```
[1] 0.12407
```

```
# 1d
mean(!null_is_true[red_light])
```

```
[1] 0.6330297
```

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:

$$\Omega = \{HHHH, HHHT, HHTH, HHTT, HTHH, HTHT, HTTH, HTTT, \\ THHH, THHT, THTH, THTT, TTHH, TTHT, TTTH, TTTT\}$$

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

Answer: They are equally likely to occur if H and T are equally likely (i.e. if it's a fair coin, i.e. $P(H) = .5$). Otherwise they would have different probabilities of occurring.

(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!), $\sum_{x=0}^4 f(x) = 1$.

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

Answer:

Generally, the probability of specific outcome with k heads and $4 - k$ tails is $q^k(1 - q)^{4-k}$. This is because the tosses are independent (so that the probability of any combination of results is the product of the probabilities of each result, and the order doesn't matter). But unless the outcome we're talking about is 4 heads or 4 tails, there is more than one way to get the

outcome. For example, you get one head (and therefore $X = 1$) with $HTTT$, $THTT$, $TTHT$, or $TTTH$. So the probability of exactly one head is $4q(1-q)^3$ – there are four outcomes that produce that event, each of which occurs with probability $q(1-q)^3$.

The PMF is therefore:

$$f(x) = \begin{cases} (1-q)^4 & x = 0 \\ 4(1-q)^3q & x = 1 \\ 6(1-q)^2q^2 & x = 2 \\ 4q^3(1-q) & x = 3 \\ q^4 & x = 4 \\ 0 & \text{otherwise} \end{cases}$$

Here we confirm that, for $q = 2/3$, the PMF sums to 1:

```
q <- 2/3
(1-q)^4 + 4*(1-q)^3*q + 6*(1-q)^2*q^2 + 4*q^3*(1-q) + q^4
```

[1] 1

(2d) Write the CDF $F(x)$ by replacing the “?”’s in the expression below.

Answer:

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

$$\Pr[X \leq x] = F(x) = \begin{cases} 0 & x < 0 \\ (1-q)^4 & 0 \leq x < 1 \\ (1-q)^4 + 4q(1-q)^3 & 1 \leq x < 2 \\ (1-q)^4 + 4q(1-q)^3 + 6q^2(1-q)^2 & 2 \leq x < 3 \\ (1-q)^4 + 4q(1-q)^3 + 6q^2(1-q)^2 + 4q^3(1-q) & 3 \leq x < 4 \\ 1 & x \geq 4 \end{cases}$$

You could also use the complement rule and write this as

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

(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:

```
q <- 2/3
draws <- sample(x = 0:4, size = 100000,
                prob = c((1-q)^4, 4*(1-q)^3 * q, 6*(1-q)^2*q^2, 4*q^3*(1-q), q^4),
                replace = T)
# this is our estimate of F(3)
mean(draws <= 3)
```

```
[1] 0.80146
```

```
# should be close to this
1 - (2/3)^4
```

```
[1] 0.8024691
```

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: The uniform density is a rectangle. In this case it goes between -5 and 3 so its length is 8. Its area must be 1 (this is true of any density), so its height must be $1/8$.

$\Pr[X < 1]$ is the area of this rectangle below 1, so its length is 6; its area is therefore $6 \times 1/8 = 3/4$. Confirmation in R:

```
punif(1, min = -5, max = 3)
```

```
[1] 0.75
```

$\Pr[-3 < X < 1/2]$ is the area of this rectangle between -3 and 1/2, so its length is 7/2; its area is therefore $7/2 \times 1/8 = 7/16$. Confirmation in R:

```
punif(1/2, min = -5, max = 3) - punif(-3, min = -5, max = 3)
```

```
[1] 0.4375
```

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

```
pnorm(1.96) - pnorm(-1.96)
```

```
[1] 0.9500042
```

(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: $\Pr[-.1 < X < 0]$ is the area under the pdf between -.1 and 0. By the fundamental theorem of calculus, the exact value is:

```
truth <- pnorm(0, sd = 2) - pnorm(-.1, sd = 2)
truth
```

```
[1] 0.01993881
```

We can approximate that area as the rectangle with the same width as the desired area (.1) and height equal to the density at -.05:

```
rect_ans <- dnorm(-.05, sd = 2)*.1
rect_ans
```

```
[1] 0.01994088
```


This approximation would be exact if the density between $x = 0$ and $x = -.1$ were a straight line, but the normal density is never a straight line, so the approximation is not exact.

To see how close it is:

```
(rect_ans - truth)/truth
```

```
[1] 0.0001041026
```

We could get even closer if we divided the interval from $-.1$ to 0 into equal segments, computed rectangles for each segment, and added up the areas. That's the basic principle of integral calculus.