9.S918: Statistical Inference for Brain and Cognitive Sciences, Pset 1
due 16 April 2024

9 April 2024

1 Incremental inference about possessor animacy

English has two CONSTRUCTIONS for grammatically expressing possession within a noun phrase, as exemplified in (1)–(2) below:

(1) the queen's crown (PRENOMINAL or 'S GENITIVE: possessor comes before the possessed noun)

(2) the crown of the queen (POSTNOMINAL or *of* GENITIVE: possessor comes after the possessed noun)

There is a correlation between the ANIMACY of the possessor and the preferred construction: animate possessors, as above, tend to be preferred prenominally relative to inanimate possessors, as in (3)– (4) below (Futrell & Levy, 2019; Rosenbach, 2005):

(3) the book's cover (Prenominal)

(4) the cover of the book (Postnominal)

Here is a pair of conditional probabilities that reflects this correlation:

$$P(\text{Possessor is } \textbf{prenominal} | \text{Possessor is } \textbf{animate}) = 0.9$$
$$P(\text{Possessor is } \textbf{prenominal} | \text{Possessor is } \textbf{inanimate}) = 0.25$$

Now consider the cognitive state of language comprehenders mid-sentence who have heard each of the three respective example sentence fragments, where the nouns that have been uttered are unfamiliar words:

1) the sneg of…
2) a…
3) a tufa's dax…

Task: Based on the knowledge encoded in the probabilities above, plot the probability in each of these three cases that the comprehender should assign to the possessor being animate, as a function of the prior probability $P(\text{Possessor is } \textbf{animate})$. That is, your plots will have the prior probability $P(\text{Possessor is } \textbf{animate})$ on the x-axis, and the posterior probability $P(\text{Possessor is } \textbf{animate}|\text{the provided sentence fragment})$ on the y-axis. Show your work in setting up the computations.

2 Variance of linear combinations of random variables, and of the MLE of a Bernoulli random variable

If random variables X and Y are conditionally independent given your information state I, then the variance of the sum $X + Y$ conditional on I is $\text{Var}[X + Y|I] = \text{Var}[X|I] + \text{Var}[Y|I]$. (Note that it is common not to specify exactly what is conditioned on in discussions of this topic, so one will often see $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$.)

1. **Task:** Suppose X and Y are not conditionally independent given I. If they are positively correlated, will the variance of their sum be greater than or less than if they were independent? What if they are negatively correlated? Give an argument based on your intuitions. Then, look up the general formula for the variance of the sum of two random variables. (**Want a challenge?** Instead of looking up the formula, derive it yourself.) Based on that formula, explain whether your intuitive argument is correct.

2. Suppose we apply a linear transformation to a random variable X: $X' = bX + c$. Then $\text{Var}[X'] = b^2 X$. **Task:** use the information provided thus far in this problem to show that the variance of the maximum likelihood estimate of the success parameter p of a Bernoulli random variable from n is $\text{Var}[\hat{p}_{\text{MLE}}] = \frac{p(1-p)}{n}$.

3 Conditional Independence

In our coverage of exchangeability and conditional exchangeability, we made use of the fact that for random variables X, Y, Z, if $X \perp Y|Z$ then $P(Y|Z, X) = P(Y|Z)$. **Task:** Prove that this is true based on the definition of conditional independence we gave in class, namely that if $X \perp Y|Z$ then $P(X, Y|Z) = P(X|Z)P(Y|Z)$.

4 Computing counterfactual risks from observational data

You set up a simple experiment in which each participant logs on to your experiment's website, reads a brief article, and then reports whether they would share the article on social media. (Regardless of what they report, your website doesn't let them actually share it.)

Your experimental materials are two articles with similar content but different framings—one positive-valence and one negative-valence, the two experimental CONDITIONS—and your scientific question is whether the positive-valence or the negative-valence article is more likely to be shared. You intended to randomize the assignment of experimental condition to participant, but you accidentally deployed an alpha-testing version of your experiment, in which the participant is first presented with two options—"For optimists" and "For pessimists". Based on their selection they are then presented with the positive or negative version of the article, and then after reading it report whether they would share the article.

You ran the experiment on a lot of participants—4000!—and here is what the results look like:

Article Valence	Participant response: share the article?	Count
Positive	Yes	672
Positive	No	608
Negative	Yes	1040
Negative	No	1680

You are concerned that because the assignment of article to participant was not randomized, it may affect your ability to draw sound scientific inferences from these data. But when your PI finds out the situation, they point out that the effect size is substantial—a 62% rate of sharing positive-valence articles, versus only a 48% rate of sharing negative-valence articles—and the sample size is so big that the estimate of this effect size is fairly precise. Surely these data will be usable! **Task 1:** use the logic and definitions provided by the potential-outcomes framework (e.g., counterfactual risks, average causal effect, consistency, exchangeability, positivity, identifiability) to explain to your PI the problems in using the 62% and 48% rates observed in the data directly to characterize the causal effect of article valence on participants' self-reported proclivities to share the article.

After you have made this case to your PI, they respond, "Is there nothing to be done?" You then remember that you included in the experiment a survey that, though brief, allows you to determine with essentially perfect accuracy whether each participant is an optimist or a pessimist. When you add the results of this survey to your data, you wind up with the resulting counts (you can download this table asa CSV file at https://:

Participant Outlook	Article Valence	Participant response: Share the article?	Count
Optimist	Positive	Yes	1664
Optimist	Positive	No	896
Optimist	Negative	Yes	480
Optimist	Negative	No	160
Pessimist	Positive	Yes	16
Pessimist	Positive	No	144
Pessimist	Negative	Yes	128
Pessimist	Negative	No	512

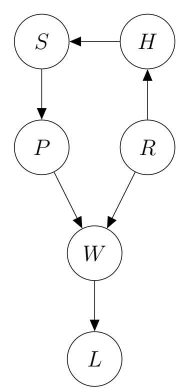
Most participants are optimists, and a strong correlation between participant outlook and choice of article valence is evident—optimists tend to choose positive articles, and pessimists tend to choose negative articles. **Task 2:** Let us provisionally assume that your measurement of participant outlook "fully captures" each participant's proclivity to choose the positively- vs. negatively-valenced article: specifically, that once outlook is taken into account, the participant's counterfactual proclivities to share each of the articles provides no further information as to which article they would choose to read. What is this notion of "fully captures" in the terminology of the potential outcomes framework that we have covered? Does participant outlook, under this assumption, help us use these data to answer the scientific question we posed? If so, explain how, and then answer the scientific question, computing the relevant quantities. If not, explain why not.

Task 3: based on common sense and whatever expertise you may have, critique this assumption that participant outlook "fully cpatures" article choice proclivity.

5 Identifying conditional independencies in (causal) Bayes nets using d-separation

Below is a modification of a classic example of a (causal) Bayes net widely used in the AI literature (e.g., Pearl, 2009; Russell & Norvig, 2016), involving the sprinklers, rain, and whether one's lawn gets wet. In this version of the scenario, the sprinklers are automatically controlled by a humidity-sensitive sensor: twice a week, they will go off if the average humidity over the preceding three days has dropped below a certain threshold set by the homeowner. Rain not only increases the humidity but also gets the lawn and footpath wet. If the footpath gets wet, it becomes slippery. For good measure, the homeowner has a separate hygrometer that measures and logs the humidity near the sensor, to make sure the sensor is working properly.

Variable	**Meaning**
R	Whether it **R**ained recently
H	Recent **H**umidity
S	Whether the sprinkers' humidity **S**ensor detected a need to water the lawn
P	Whether the s**P**rinklers went off
W	Whether the lawn and footpath to the front door got **W**et
L	Whether the footpath is s**L**ippery



Task: We will be interested in what conditional independencies among variables in the network hold, given various types of information. For this exercise, when I ask, "Are A and B conditionally independent given C?", you may find that a useful way to think about that question is: suppose I already know C, and based on it I have some beliefs about A. If I additionally observe B, can it further change my beliefs about A? If yes, then A and B are **not** conditionally independent given C.

1. Are R and P conditionally independent, given no additional information? Answer both based on your intuition regarding the scenario and on the semantics of the Bayes net.

2. The homeowner is out of town but reads the hygrometer's log for the past three days (the hygrometer logs to the cloud). Are R and P conditionally independent now?

3. After reading the hygrometer's log, the homeowner gets a text message from the cat sitter, who reports slipping on the footpath on the way to the house. Are R and P conditionally independent now?

4. For each of the preceding three questions, suppose the sensor has broken so that it randomly sets off the sprinkler with 50% probability every three days (and the homeowner knows it). How would you represent this as an INTERVENTION in the language of causal Bayes nets? What does your post-intervention causal Bayes net look like? Does this intervention change the answers to any of the above three questions? Why?

5. One thing that you may find dissatisfying about the structure specified in our causal Bayes net is the directionality of the edge between H and R: although rain can cause increases in humidity, it is surely also the case that factors leading to an increase in humidity on the ground can also lead to rain. How might you modify the causal Bayes net to take this into account? Does this modification change the answer to any of the previous three questions?

References

Futrell, R., & Levy, R. P. (2019). Do RNNs learn human-like abstract word order preferences? *Proceedings of the Society for Computation in Linguistics (SCiL) 2019*, *2*, 50–59.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (Second). Cambridge.

Rosenbach, A. (2005). Animacy versus weight as determinants of grammatical variation in English. *Language*, *81*(3), 613–644.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: A modern approach*. Pearson.