9.S918: Statistical Inference for Brain and Cognitive Sciences, Pset 1
due 14 Feb 2025

7 Feb 2025

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(Possessor is **animate**). That is, your plots will have the prior probability P(Possessor is **animate**) on the x-axis, and the posterior probability P(Possessor is **animate**|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 $\mathrm{Var}[X+Y|I] = \mathrm{Var}[X|I] + \mathrm{Var}[Y|I]$. (Note that it is common not to specify exactly what is conditioned on in discussions of this topic, so one will often see $\mathrm{Var}[X+Y| = \mathrm{Var}[X] + \mathrm{Var}[Y]$.)

1. Use the definition of variance ($Var[X] = E[(X - E[X])^2]$) to show that the variance of a Bernoulli random variable is $p(1-p)$. For what values of p is the variance the greatest, and the least?

2. **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.

3. Suppose we apply a linear transformation to a random variable X: $X' = bX + c$. Then $\mathrm{Var}[X'] = b^2X$. **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 a sample of size n is $\mathrm{Var}[\hat{p}_{\mathrm{MLE}}] = \frac{p(1-p)}{n}$.

3 Sample mean and sample variance

The **sample mean** and **sample variance** of a sample $x_1, x_2, \ldots, x_n$ is closely related to the mean and variance of probability distributions, which we covered in the first week of class. The sample mean is defined as:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad \text{(sample mean)}$$

The term "sample variance" is most commonly used to refer to the following, and we will use it this way:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} \qquad \text{(sample variance)}$$

but occasionally one might see it used to refer to this quantity:

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$$

1. The sample mean is **unbiased**. Provide an example in code that illustrates this lack of bias, using any probability distribution you like (so long as the distribution has variance above 0).

2. One of the two formulations of sample variance is unbiased. Using simulations, show which one it is. (Want a challenge? Provide logical reasoning for which one it must be, and/or prove it.)

3. Which of the two formulations of the sample variance estimator has lower variance?

4 Data manipulation and plotting

The dataset made available at https://github.com/Lakens/Stroop/blob/master/stroop.csv by Daniel Laakens is a simple Stroop dataset, two experimental trials (observations) per participant. If you're not already familiar with it, you can read about the Stroop task and effect on its Wikipedia page; also, you can do theexperiment on yourself by visiting http://faculty.washington.edu/chudler/java/ready.html. The four columns of the dataset are:

`PPTT`	Participant number
`Congruent`	Response time for the congruent-condition trial
`Incongruent`	RT (for response time, or reaction time) for the incongruent-condition trial
`Year`	Calendar year in which the data were collected

Somewhere in the below there is a trick question. Identify the trick question, explain why it is a trick question, and do the rest of the questions.

1. Compute the sample mean and the sample variance for RTs in each of the Congruent and Incongruent conditions. Which has a higher sample mean? Which has a higher sample variance?

2. Analogously, compute the sample mean and the sample variance for each of the two years in which data were collected. Which year had a higher mean? Which had a higher sample variance?

3. Create a "spaghetti plot" (look this up if you're not familiar) of participant-specific RTs in the two conditions. Does this information reveal anything that was not already evident from the condition-specific sample means and variances?

4. Create an analogous "spaghetti plot" for the two years instead of the two conditions. Does this information reveal anything that was not already evident from the year-specific sample means and variances?

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.

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