Simulating Language Assessment 2

Deadline: 8th of December at noon

Answer all the questions below. There is not a word limit, but your answers must be concise - typically only a few sentences will be necessary, and unnecessarily lengthy answers will be penalised. Some of the answers require a figure, which may include one or more graphs[1] - if you are unsure of how to save a graph from a jupyter notebook, look back at Lab 1.

Please start your answer to each question on a new page in the document you submit on turnitin, and make sure you indicate on each page what question you are answering!

Each question has an allocated number of marks to indicate its importance; the entire assignment is marked out of 42, and we will scale the resulting marks to lie on the University's Common Marking Scheme.

This work should be completed on your own. **Do not discuss your answers with other students!** Note that you should not need to include any code in your answer, only text and figures which clearly illustrate your results; however, keep a copy of the code you used to answer the questions, we may ask to see it.

Please note that we will **only be answering questions about this assignment until noon on Friday 2nd December, and all questions must be posted in the assignment channel on Teams so everyone can see the answers**.[2] This is due to past experience that answering questions for too long after the assignment is released introduces unfairness in the results.

---

[1] Also, you can check out matplotlib.org for more details of how to create various different kinds of plots if you wish.

[2] This rule does not apply if you have a specific adjustment on your student record that allows for clarification on assignment questions.

Questions:

Q1 (14 Points)
In the appendix to their paper, Kirby et al (2015) report the effects of starting their compositionality simulation off with degenerate rather than holistic languages. Using the code from lab 6,

1. **Compare and contrast** (using a **graph or graphs** with **captions explaining** the results):
    a. holistic starting points
        - with population turnover &
        - without population turnover
    b. degenerate starting points
        - with population turnover &
        - without population turnover

    (8 points)
2. Does **starting** with a compositional language guarantee that all future forms of the language will **also** be compositional? Answer using a graph or graphs and caption.
    (6 points)

Q2 (14 points)

1. Iterated learning always converges over time to the agents' prior when the sampling strategy is used for hypothesis selection. This means that varying the bottleneck size has no effect on the endpoint of cultural evolution (with this type of learner). However, it does take some time for this convergence to the prior to occur in cultural evolution. Using the code from lab 8, and **using a graph and a caption explaining the results**, demonstrate the effect of varying the bottleneck on how long it takes to converge to the prior.
    (8 points)
2. Can we tell if human languages are *now* in a stationary distribution?
    (6 points)

Q3 (14 Points)
In Lab 7, we looked at how hierarchical Bayesian learners use the data they observe to estimate the overall bias of speakers for regularity vs variability, over and above the particular variant they prefer. Learners have a prior belief about this different overall bias, which is represented in this model by a prior over α values. In the lab, the prior for α is initially uniform, i.e. all 6 values of α are given equal weight. The end of the Lab 7 answers notebook uses the `alpha_calculate_logprior_new()` function to give an example of how to create a hierarchical learner with a bias towards regularising values of α.

1. Create two new functions:
    a. `alpha_calculate_logprior_regular()` which produces a prior strongly favouring regular languages,
    b. `alpha_calculate_logprior_confused()` which produces a prior with a strong bias spread over very regular and very variable speakers, but nothing in between

    (4 points)
2. Using a graph or graphs with a caption explaining the results, demonstrate the kind of languages each type of learner above (i.e. `regular` and `confused`) would **produce** after **exposure to data from** a population which **consists of two agents**, where one agent is a very regular speaker and the other agent is a very variable speaker.
    (10 points)