Lecturers: Torben Ott, Robert Gütig

Causal inference in multisensory perception

The report for this project (no more than 4 pages of text with discussion and interpretation as a PDF file) as well as the source code should be handed in no later than **July 31 2025, 11:59pm** through the Moodle interface. You can discuss your advances and hurdles with the project instructor until **July 15** during the weekly supervision meetings.

1 Summary

The goal of this project is to explore a model of causal inference for a multisensory perception task [Körding et al., 2007]. The model was designed to explain the behaviour of subjects in an experiment where they are asked to identify the location of both auditory and visual stimuli. Depending on the proximity of the two stimuli, they may infer a common cause. This belief influences the location they report for the stimuli. You will first implement and explore the behaviour of the model in response to changes in parameters. You will then use the model to generate "experimental" data. This experimental data is then used to fit the model using two different approaches: a brute-force parameter space search and MCMC-sampling.

2 Detailed instructions

1. Understand and implement the model:
   a) Read the paper, with a particular focus on the Materials and Methods.
   b) Use default parameters of $p_{common} = 0.8$, $\sigma_v = 0.6$, $\sigma_a = 3.1$, $\sigma_p = 15$, unless instructed otherwise. Set $\mu_p = 0$ and leave it that way for all tasks.
   c) Write a function that implements Eq. 2 from the paper, i.e., the probability of a common cause given noisy stimuli $p(C = 1|x_v, x_a)$ and plot this probability as a function of $x_v - x_a$ (assume $x_a = 0$). Vary p_{common}, σ_v, σ_a, and σ_p separately and explain why the shape of the curve changes the way it does.
   d) Write functions that implement Eqs. 9 & 10, i.e. the estimated stimulus positions $\hat{s}_v$ & $\hat{s}_a$ and plot both of the estimates as a function of x_v assuming $x_a = 0$. Once again see what effect each model parameter has. Plot and explain your results.
   e) Consider the integral in Equation 13, how might you evaluate it? Within the integral appears $p(\hat{s}_v|x_v, x_a)$: What kind of function is this? How does it affect the analytic evaluation of the integral?
   f) The nature of $p(\hat{s}_v|x_v, x_a)$ also suggests a numerical method for obtaining estimates of $p(\hat{s}_a|s_v, s_a)$ and $p(\hat{s}_v|s_v, s_a)$. Use that method to plot both as a function of s_v assuming $s_a = 0$ (use a heat-map for these plots).
2. Fit the computer model to simulated "experimental" data. To this end, we will first generate data from a simulated subject with a given set of parameters:

- Simulate an experiment in which you show a set of s_a, s_v values. Use regularly spaced values $s_{v,a} \in \{-12, -6, 0, 6, 12\}$, and show each of the 25 pairs of values for s_v and s_a many times.
- For each trial, generate noisy observations x_a, x_v based on the parameters σ_v, σ_a.
- Based on these observations, perform the inference of $\hat{s}_{a/v}$ for the given set of parameters ($p_{common}, \sigma_v, \sigma_a$ and σ_p). These parameters characterize the simulated "subject". In the following, we will pretend that we don't know them – as for a real subject – and that we want to find out what they are.
- Translate those inferences into behavioral responses. Here, we will use button presses, with 5 buttons corresponding to position estimates $(\hat{s}_v, \hat{s}_a)$ in the ranges $(-\infty,\ -9), [-9,\ -3), [-3,\ 3), [\,3,\ 9), [\,9,\ \infty)$. This way, we get two button presses for every trial, one for the auditory domain – based on $\hat{s}_a$ – and one for the visual domain – based on $\hat{s}_v$.
- The goal is now to find out the parameters by identifying the posterior distribution of those parameters given only the button presses of the simulated subject.

But now step by step:

a) Write a function *make_button_presses* that takes the number of trials N and parameters for the generative model and produces a histogram of button presses for every possible stimulus pair (s_v, s_a).

b) Using that function, generate button presses for $N = 10^4$ trials, covering all the stimulus pairs evenly. This will be your "experimental data". An efficient implementation of this function should run in ≤ 50 ms for 10^5 trials.

c) With our experimental data in hand, we now want to investigate how well we can infer the parameters of the generative model. Write a function that computes the log-likelihood (Eq. 16) of a model whose parameters are passed as arguments (*Hint:* you can re-use *make_button_presses* within this new function). Use at least 10 times as many trials as there are in the experimental data. Why is this important?

d) Test this new function using a set of parameters different from those used to generate the experimental data. Look at numerical aspects, e.g. NaN & Infty. Why do they appear in the calculation of the likelihood? What does this tell you about the shape of the likelihood in the parameter space? Does this have implications for using gradient-based approaches to fit the model? Consider adding a tiny ϵ to p_i in Eq. 16. What effect does this have and why is this desirable?

e) We will now attempt a naïve approach to fit the generative model to the experimental data. Compute the log-likelihood for all models within a reasonable parameter space ($0 \leq p_{common} \leq 1$, $0 < \sigma_{v,a,p} < 20$). Sample the parameter space at least 10 times per axis, yielding $\geq 10^4$ models to test. Find the global maximum likelihood value for the parameters. Given the particular values that you tested, and the true values, did you find the best possible estimate? Plot the marginal likelihoods with respect to each parameter, and indicate the global max-likelihood values. Would it have been possible to identify those maximum values using only the marginals? Explain why.

f) Having completed a brute force search of the parameter space, let's now try something a bit smarter: Use MCMC sampling (using the Affine Invariant method provided in the package *emcee*) to estimate the posterior probability distribution of the parameters. You will need to implement a function to calculate the log-probability;

it will call the log-likelihood function you implemented as well as a log-prior function (use a rectangular prior that is finite for some reasonable parameter range). Plot a corner plot of the result using the package *corner*.

g) Compute the mean value of the sampled points in the parameter space and compare its accuracy to that of the brute-force approach. Discuss the advantages and disadvantages of the two methods.

h) Generate a new experimental data set using two "subjects", one with $\sigma_a = 9$, and the other with $\sigma_a = 2$ (the other parameters should be $p_{common} = 0.3$, $\sigma_v = 2.1$, $\sigma_p = 12$ for both). Combine their data into a single dataset. How well does the MCMC sampler deal with this new data? What implications does this have for fitting real experimental data?

3. Write a report that presents the results of your model analysis. Marks will be based on the clarity of the presentation and on the thoroughness of the analysis.

References

[Körding et al., 2007] Körding, K. P., Beierholm, U., Ma, W. J., Quartz, S., Tenenbaum, J. B., and Shams, L. (2007). Causal inference in multisensory perception. *PLoS one*, 2(9):e943.