# Cognitive Modeling: Homework Assignment 3
# Discrete Models and Model Criticism

March 28, 2024

All answers and solutions to non-programming questions should be submitted to LMS as a **legible** write-up (either fully digital or a scan). All code should be committed to and merged into the `main` branch of your team's GitHub repository.

## Problem 1: True-False Questions (5 points)

Mark all statements which are **FALSE**.

1. Direct $K$-fold cross-validation requires $K$ model re-fits, which may be computationally demanding, especially when inverse inference is costly.

2. Bayes factors (BFs) are *relative measures*, that is, they cannot differentiate between "equally good" and "equally bad" models.

3. Marginal likelihoods and, by extension, Bayes factors (BFs) cannot be used to compare models with different likelihoods.

4. Both the Binomial and the Dirichlet distribution can be formulated as special cases of the Multinomial distribution.

5. Bayesian leave-one-out cross-validation (LOO-CV) relies on the posterior predictive distribution of left-out data points.

6. The Akaike Information Criterion (AIC) penalizes model complexity indirectly through the variance of a model's marginal likelihood.

7. The log-predictive density (LPD) is a relative metric of model complexity.

8. The LPD can be approximated by evaluating the likelihood of each posterior draw (e.g., as provided by an MCMC sampler) and taking the average of all resulting likelihood values.

9. Bayes factors do not depend on the prior odds, that is, the ratio of prior model probabilities $p(\mathcal{M}_1)/p(\mathcal{M}_2)$.

10. You should always prefer information criteria to cross-validation in terms of estimation predictive performance.

## Problem 2: Simple Multinomial Processing Trees (MPTs) (10 points)

Collect some data (e.g., from a friend or your teammate) on the recognition memory task from the slides (or construct your own task) and fit the following two models using `Stan`:

- The One-High-Threshold Model (1HT)
- The Two-High-Threshold Model (2HT)

The models are depcited on Slide 12 (MPT models). As usual, inspect the convergence of the MCMC samplers and report the estimation results. Do the two models suggest different estimates for the two key parameters? Describe and interpret the results.

## Problem 3: A More Complicated MPT Model (10 points)

Write down the model equations for the MPT model depicted in **Figure 1** in the paper (avaliable on LMS):

- Walk1er, G. M., Hickok, G., & Fridriksson, J. (2018). A cognitive psychometric model for assessment of picture naming abilities in aphasia. Psychological assessment, 30(6), 809.

Then, write a `Stan` program for the MPT model featuring the following five blocks: `data` – for passing the hypothetical categorical data; `parameters` – for defining the latent model parameters; `transformed parameters` – for transforming latent model parameters into probabilities; `model` – for formulating the Bayesian joint model; `generated quantities` – for sampling new frequency data given the posterior draws (generative performance).

**Bonus (6 points)**: Simulate a data set according to the forward model and inspect parameter recovery.

## Problem 4: A Discrete Conjugate Model (6 points)

Derive the analytic posterior for the conjugate Dirichlet-Multinomial model (no ChatGPT):

$$\theta \sim \text{Dirichlet}(\alpha) \tag{1}$$
$$y \sim \text{Multinomial}(y \mid \theta; N) \tag{2}$$

## Problem 5: Multiple Regression (8 points)

Extend your simple Bayesian regression model from the previous exercise into a multiple regression model:

$$\sigma \sim \text{Inv-Gamma}(\tau_0, \tau_1) \tag{3}$$
$$\alpha \sim \text{Normal}(0, \sigma_\alpha) \tag{4}$$
$$\beta \sim \text{Multivariate-Normal}(0, \sigma_\beta \mathbb{I}) \tag{5}$$
$$y_n \sim \text{Normal}(\alpha + \beta x_n^T, \sigma) \quad \text{for} \quad n = 1, \ldots, N, \tag{6}$$

where you need to set the hyperparameters of the prior (i.e., $\tau_0, \tau_1, \sigma_\beta, \sigma_\alpha$) to some reasonable values. Next, use your `Stan` program to fit a Bayesian multiple regression model for the Insurance Costs data set: `https://github.com/stedy/Machine-Learning-with-R-datasets/blob/`

`master/insurance.csv`.

Your goal is to predict the insurance charges (`charges`) from a patient's BMI (`bmi`), age (`age`), and number of children (`children`). Thus, you need to estimate three regression weights ($\beta_1$, $\beta_2$, and $\beta_3$), along with the intercept ($\alpha$), and the noise parameter ($\sigma$). It is also recommended that you standardize your predictors (i.e., subtract the means from the input variables and divide by their standard deviations) in order to bring them to a common scale. Split the data into a training set and a test set and fit the model only to the training set. Perform the usual convergence checks and describe your results. Which of the three variables is the best predictor of Insurance Charges?

**Alternative**: Use the Bayesian Ridge regression implementation from `scikit-learn`: `https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.BayesianRidge.html`, also for the next task. If you go this path, include a small description on how the Bayesian ridge differs from the model implementation suggested above.

## Problem 6: Predictive Distribution (5 points)

Use the `generated quantities` block in the `Stan` program to also pass the test data and sample from the predictive distribution. Extract the samples from the predictive distribution, compute the means predictive means from the samples, calculate the root-mean-squared error (RMSE) between the predictive means and the actual charges in the test set:

$$\text{RMSE} = \sqrt{\frac{1}{M}\sum_{m=1}^{M}(\hat{y}_m - y_m)^2}, \tag{7}$$

where $M$ denotes the number of test instances and $\hat{y}_m$ denotes the predictive means. How good are your predictions? What information did you lose by computing the predictive means? How could you possibly propagate the uncertainty information encoded in the predictive distribution to obtain a distribution over the test RMSE values?