

# Bayes ML Assignment 2

Author: Teague Henry

Fall 2024

## Part 1: Using Conjugate Priors with White-Wine Data (33 pts)

To begin, take a look at the white-wine dataset that accompanies this assignment. There are 11 continuous predictors (we are not considering quality as a continuous variable here).

- Choose 2 continuous predictors and plot density plots (using `ggplot2`).
- For each predictor, consider the density plot. What do you notice? Do they look like they are normally distributed? Skewed?

Now, for each variable please do the following:

1. Using a normal likelihood with **known** variance (the known variance for each variable is the observed variance, go ahead and calculate that), use the appropriate conjugate prior and calculate out the posterior distribution (you should be able to look up the formula for the posteriors, no need to derive them out yourself). When you calculate these posterior distributions, use two sets of hyperparameters (per variable), one where the hyperparameters specify a fairly uninformative prior, and the other where the hyperparameters are much more informative (this doesn't need to be a reasonable value either, this exercise is to demonstrate the impact of hyperparameter choice). At the end of this exercise, you should have the parameters for 4 posterior distributions.
2. What are the impacts of different hyperparameter choices on the posterior distributions? Is it possible to chose "bad" hyperparameters? If so, why? What are the consequences for inference?

Now, we are going to repeat the process, but this time using a different likelihood function.

1. Repeat the previous work, but this time use an exponential likelihood and corresponding conjugate prior (again, you can look this up and get the formula that way).
2. Again, what are the impacts of the hyperparameter choice?
3. In the previous example, it was very simple for you to interpret the parameters in the posteriors. In this case, you will need to calculate the expected value and variance of the posterior distribution. You can look this up for the distribution (use Wikipedia!). How do

these values differ from the values you found when using a normal distribution as the likelihood?

## Part 2: Multinomial Priors for Wine Quality (33 points)

The `quality` variable in this dataset is a categorical variable with values taking letter grades (A, C, F). We can consider this variable to be multinomially distributed. Multinomial distributions have a conjugate prior in the Dirichlet distribution. The Dirichlet distribution can be parameterized using either a single  $\alpha$  parameter that applies to each category, or you can specify a different  $\alpha_k$  parameter for each category. In any case, the posterior distribution for a multinomial-Dirichlet model is  $Dirichlet(\alpha + n)$ , where  $\alpha$  is a vector of either all the same number, or the hyperparameter choice per category, and  $n$  is a vector of the counts of each category.

1. Looking at the above formula for the posterior distribution, how can you interpret the meaning of  $\alpha$ ?
2. Choosing two sets of hyperparameters, one fairly uninformative and one highly informative, generate 1000 observations from the posterior distributions (using `rdirichlet` from the `dirmult` R package). At the end of this generative process, you should have two data.frames or matrices that have 1000 rows and 3 columns.
3. Plot these posterior distributions (you should end up with 2 figures of box plots, one figure per prior specification, each figure containing 3 boxplots, one for each letter grade).
4. Comment on the impact of prior choice here.

## Part 3: A Bayesian Test of Inference (34 points)

What we've been doing so far here is exploring the impact of priors using marginal distributions. While you could technically do some form of statistical inference with these, the inference isn't that interesting (is `alcohol` significantly different from 0? for example). In this part, we are going to be using conjugate priors to examine the difference in `alcohol` content between wines rated `A` and wines rated `F`.

To do this, follow these steps:

1. Using a normal distribution with known variance (again, using the variances you can calculate from the data), specify 2 hyperparameter choices, one fairly uninformative, one very informative, for the alcohol content in wines rated A and wines rated F. Note, you will need hyperparameters for each type of wine, but those hyperparameters can be the same for each type of wine.

2. Calculate out the posterior distributions for alcohol content in wines with an F rating, and wines with an A rating. Because the posterior distribution will be a normal distribution with a value for the posterior mean and variance, you will have two means and two variances (per hyperparameter set, so you'll have 4 in total.)
3. These posterior distributions are still for the marginal distributions of alcohol content, and we are interested in if the alcohol content differs between the two levels of wine quality. Fortunately, the difference between normal distributions is a normal distribution, so we can hand calculate the posterior distribution of the differences between alcohol content:
  1. The posterior mean of the difference between two normal distributions with means  $\mu_x$  and  $\mu_y$  is simply  $\mu_x - \mu_y$
  2. The posterior variance of the difference between two normal distributions (with variances  $\sigma_x^2$  and  $\sigma_y^2$ ) is simply  $\sigma_x^2 + \sigma_y^2$ .
4. Now, you should have the posterior distributions of the differences between alcohol contents for wines rated A vs F. You'll have 2 of these posterior distributions because you had two sets of priors, one uninformative, one highly informative.
  1. Calculate the 95% HDI for each of the posterior distributions. What does this interval tell you about the difference between the alcohol quantities in the two grades of wine? Would you consider the alcohol content to be 'significantly' different?
  2. How does prior choice impact this?

## Extra Credit (25 pts)

Savvy students will notice I made an important assumption in specifying the likelihoods in part 1 and 2, that they are normal likelihoods with **known** variance. This was to simplify the posterior from a normal-inverse-gamma to just a normal distribution. However, technically, it's a bad assumption to make. In this extra credit, you will be writing a small Stan analysis to test the difference between alcohol quantities in wines rated A and wines rated F, when we don't assume we know the variance of the alcohol quantities.

1. Write a Stan model that specifies normal-likelihoods with unknown means and variances. The priors for the means should be normal distributions, while the priors for variances should be a Half-Cauchy (note, this is not the conjugate prior, but we are going to be using Stan here, so we don't need to chose conjugate priors!). Then, using the transformed parameters block, calculate the difference between the means of the marginal distributions.
2. Run the Stan models using two sets of hyperparameters, the uninformative choices and a highly informative choice. Plot the posterior distributions of the differences in the means.

3. How are these posteriors different/similar to those from the original analyses where we specified the variances as known? What was the impact of priors here?