For typed text, my vectors and matrices are denoted with bold font. On your homework, please explicitly denote vectors using the arrow notation!

Also, you're welcome to drop the subscripts on distributions! I provide them for clarity, but at this point you're all hopefully able to tell the difference between $f(\mathbf{y})$ and $f(\theta)$.

1. Let's now consider count data. Let Y_i represent counts from some process, $i = 1, \ldots, n$. For example, Y_i might represent number of fish caught at Lake Dunmore in an hour on the i-th fishing trip! A common distribution for count data, and the one we will use here, is the Poisson:

$$Y_i|\theta \stackrel{\text{iid}}{\sim} \text{Poisson}(\theta), \qquad i = 1 \ldots, n$$

Take a second now to look at the PMF of the Poisson and remind yourself of the support!

(a) The *rate* parameter θ of the distribution must be positive. A distribution with support on the positive real line is the Gamma distribution. Take a second now to look at the PDF of the Gamma!

We will assume the following prior distribution for θ:

$$\theta \sim \text{Gamma}(a, b)$$

Under the proposed sampling model, obtain the posterior distribution for θ by explicitly deriving/evaluating the marginal likelihood along the way. You should be able to name the posterior exactly!

(b) Clearly relate the posterior mean of θ to the sample mean and the prior mean. Can you provide some way to "interpret" the a and b in the context of this statistical model?

(c) Coming soon...

2. We will consider the general case of the Bernoulli distribution, which is the **Categorical distribution**. We have a discrete random variable Y that can take one of the values $1, 2, \ldots, J$, each with a specified probability. The PMF of this distribution is

$$f_{y|\boldsymbol{\theta}}(y|\boldsymbol{\theta}) = \Pr(Y = j|\theta_1, \ldots, \theta_J) = \begin{cases} \theta_j & j = 1, 2, \ldots J \\ 0 & o.w. \end{cases},$$

.

There is a constraint on the θ_j: they must live on the $(J-1)$-*dimensional simplex* Δ^{J-1}. This means the probabilities θ_j are such that:

- $\theta_j \geq 0$ for all $j = 1, \ldots, J$, and
- $\sum_{j=1}^{J} \theta_j = 1$.

In the context of the Categorical distribution, this ensure that the probability of obtaining category j is non-negative, and the experiment will always result in one of the J categories.

We denote this distribution as $Y \sim \text{Categorical}_J(\boldsymbol{\theta})$ or $Y \sim \text{Categorical}(\theta_1, \ldots, \theta_J)$. This is like a Bernoulli, but with J possible outcomes, rather than just 2. *Note: the Categorical distribution is equivalent to the Multinomial distribution with 1 trial.*

The θ_j are unknown, and we would like to obtain a posterior distribution for $\boldsymbol{\theta}$ based on data from a Categorical sampling model. Because $\boldsymbol{\theta}$ is a random *vector*, we will need a multivariate prior distribution. Consider the Dirichlet$(\alpha_1, \alpha_2, \ldots, \alpha_J)$ distribution, where $J \geq 2$ and each $\alpha_j > 0$ (pronounced "deer-ee-shlay").

If $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_J) \sim \text{Dirichlet}(\alpha_1, \ldots, \alpha_J)$, its PDF is

$$f_{\boldsymbol{\theta}}(\theta_1, \ldots, \theta_J) = \frac{\Gamma(\sum_{j=1}^J \alpha_j)}{\prod_{j=1}^J \Gamma(\alpha_j)} \prod_{j=1}^J \theta_j^{\alpha_j - 1} \qquad \text{if } \boldsymbol{\theta} \in \Delta^{J-1},$$

and 0 otherwise. *Please don't be scared!*

Suppose that our sampling model is $Y_1, Y_2, \ldots Y_n | \boldsymbol{\theta} \overset{\text{iid}}{\sim} \text{Categorical}(\theta_1, \ldots, \theta_J)$, and we use a Dirichlet$(\alpha_1, \alpha_2, \ldots, \alpha_J)$ prior for $\boldsymbol{\theta}$.

(a) Obtain the posterior distribution $f_{\boldsymbol{\theta}|\mathbf{y}}(\boldsymbol{\theta}|\mathbf{y})$. You may either explicitly derive the marginal likelihood along the way, or you may not. You should be able to name the posterior exactly! *It may be useful to use indicator notation at some point!*

(b) Obtain the posterior predictive distribution $f_{Y^*|\mathbf{y}}(y^*|\mathbf{y}) = \Pr(Y^* = l|\mathbf{y})$, for $l = 1, 2, \ldots, J$. Interpret the value that you obtain. *It may be useful to use indicator notation at some point!*

3. TBD