# Dirichlet Model

## General Principles

To model the relationship between a vector outcome variable in which each element of the vector is a frequency from a set of more than two categories and one or more independent variables, we can use a *Dirichlet* model.

## Considerations

> **Note**
>
> - We have the same considerations as for the Multinomial model.

## Example

## Python

### R

## Mathematical Details

We can model a vector of frequencies using a Dirichlet distribution. For an outcome variable $Y_i$ with $K$ categories, the *Dirichlet* likelihood function is:

$$Y_i \sim \text{Dirichlet}(\theta_i \kappa) \theta_i = \text{Softmax}(\phi_i) \phi_{[i,1]} = \alpha_1 + \beta_1 X_i \phi_{[i,2]} = \alpha_2 + \beta_2 X_i ... \phi_{[i,k]} = 0 \kappa \sim \text{Exponential}(1) \alpha_k \sim \text{Norma}$$

Where:

- $Y_i$ is the outcome simplex for observation $i$.
- $\kappa$ is the concentration parameter, it controls the prior weight on each category.
- $\theta_i$ is a vector unique to each observation, $i$, which gives the probability of observing $i$ in category $k$.
- $\phi_i$ give the linear model for each of the $k$ categories. Note that we use the softmax function to ensure that that the probabilities $\theta_i$ form a simplex .
- Each element of $\phi_i$ is obtained by applying a linear regression model with its own respective intercept $\alpha_k$ and slope coefficient $\beta_k$. To ensure the model is identifiable, one category, $K$, is arbitrarily chosen as a reference or baseline category. The linear predictor for this reference category is set to zero. The coefficients for the other categories then represent the change in the log-odds of being in that category versus the reference category.

## Reference(s)