Multinomial 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 *Multinomial* model.

Considerations

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

Example

Below is an example code snippet demonstrating a Bayesian multinomial model using the Bayesian Inference (BI) package. This example is based on McElreath (2018).

Python

```python
from BI import bi
import jax.numpy as jnp
import pandas as pd
import jax
# Setup device ----------------------------------------------
m = bi('cpu')

# Import Data & Data Manipulation ---------------------------------------------------
# Import
from importlib.resources import files
```

```python
data_path = files('BI.resources.data') / 'Sim data multinomial.csv'
m.data(data_path, sep=',')

# Define model ----------------------------------------------
def model(income, career):
    # Parameter prior distributions
    alpha = m.dist.normal(0, 1, shape=(2,), name='a')
    beta = m.dist.half_normal(0.5, shape=(1,), name='b')
    s_1 = alpha[0] + beta * income[0]
    s_2 = alpha[1] + beta * income[1]
    s_3 = 0
    p = jax.nn.exp(jnp.stack([s_1[0], s_2[0], s_3[0]]))
    # Likelihood
    m.dist.multinomial(p[career], obs=career)

# Run sampler ----------------------------------------------
m.fit(model)

# Summary ----------------------------------------------
m.summary()
```

R

```r
library(BI)
m=importbi(platform='cpu')
```

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:

$$Y_i \sim \text{Multinomial}(\theta_i)\theta_i = \text{Softmax}(\phi_i)\phi_{[i,1]} = \alpha_1 + \beta_1 X_i \phi_{[i,2]} = \alpha_2 + \beta_2 X_i...\phi_{[i,k]} = 0\alpha_k \sim \text{Normal}(0,1)\beta_k \sim \text{Norm}$$

Where:

- Y_i is the outcome (i.e. the vector of frequencies for each k categories) for observation i.

- θ_i is a vector unique to each observation, i, which gives the probability of observing i in category k.

- ϕ_i give the linear model for each of the k categories. Note that we use the softmax function to ensure that that the probabilities θ_i form a simplex .

- Each element of ϕ_i is obtained by applying a linear regression model with its own respective intercept α_k and slope coefficient β_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)

McElreath, Richard. 2018. *Statistical Rethinking: A Bayesian Course with Examples in r and Stan.* Chapman; Hall/CRC.