Dirichlet-Multinomial Model (WIP)

General Principles

To model the relationship between a categorical outcome variable with more than two categories and one or more independent variables with overdispersion , we can use a *Dirichlet* model.

Considerations

> 🔥 Caution
>
> - We have the same considerations as for the Multinomial model.
>
> - One major difference from the multinomial model is that the Dirichlet model doesn't require a simplex but rather strictly positive values. We can thus exponentiate the outputs from the categorical regressions instead of using the softmax function. This example is based on McElreath (2018).
>
> -

Example

```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
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.halfnormal(0.5, shape=(1,), name='b')
    s_1 = alpha[0] + beta * income[0]
    s_2 = alpha[1] + beta * income[1]
    s_3 = alpha[0] + beta * income[0]
    p = jax.nn.exp(jnp.stack([s_1[0], s_2[0], s_3[0]]))
    # Likelihood
    m.dirichletmultinomial(p[career], lambda_, obs=career)

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

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

Mathematical Details

Formula

Bayesian Model

In the Bayesian formulation, we define each parameter with priors . We can express the Bayesian regression model accounting for prior distributions as follows:

Reference(s)

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