Math 561 Assignment 3*

Due date: 24 Feb 2023.

	Brazilian	Not
Likes soccer	9	1
Not	1	9

We are interested in determining whether the random variable X=whether the person is Brazilian and the random variable Y=whether the person likes soccer are independent. We have a dataset of $n = 20$ people classified according to the 2×2 table above. Let u be the matrix of counts given above: $u = \begin{pmatrix} 9 & 1 \\ 1 & 9 \end{pmatrix}$.

The 2×2 independence model is $\mathcal{M}_{X \perp\!\!\!\perp Y} = \{p = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \in \Delta_3 : p_{ij} = \alpha_i \beta_j, (\alpha, \beta) \in \Delta_1 \times \Delta_1\}$.

1. Likelihood inference.

- Write down the likelihood function $L(\theta|D) = \prod p_\theta(j)^{u_j}$.
- Write down the log-likelihood function $l(\theta|D) = \log L(\theta|D)$.
- The maximum likelihood estimate $\hat{\theta}$ is the maximizer of the log-likelihood function:

$$\hat{\theta} = \text{argmax}_{\theta \in \Theta} l(\theta|D).$$

  How would you find the maximum likelihood estimate for the model $\mathcal{M}_{X \perp\!\!\!\perp Y}$ and the matrix of counts u using the log-likelihood function? (You do not have to explicitly write down the solution, but discuss what are the steps.)
  - Write down the solution explicitly using any software of your choice; of course, you need to show your work.
- Alternatively, the model $\mathcal{M}_{X \perp\!\!\!\perp Y}$ can be given implicitly as $\{P = \begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \in \Delta_3 : p_{11}p_{22} - p_{12}p_{21} = 0\}$.

How would you find the maximum likelihood estimate for the implicit model? (No need to compute this explicitly, unless you want extra credit **5 points extra for complete solution with explanation.**)

2. Evidence for or against independence of X and Y.

- What is the estimated probability talbe from the data table u?
  - In other words, write down the MLEs for each p_{ij} in 2×2 table format.
- What is the expected table $\hat{u}$?
  - In other words, write down the what the independence model predicts for the value of each cell u_{ij}.
- *Define* the p-value of the data table u
- Compute the p-value of u using the chi-sauare test of independence in `R` or `python` or by hand. [Show your code/work in your HW submission.]
- Compute the p-value of u using Fisher's exact test of independence in `R` or `python` or by hand. [Show your code/work in your HW submission.]
- What do you conclude?

*Algebraic and Geometric Methods in Statistics, Spring 2023