① Categorical Distribution

$\{1, \ldots, k\}$

$\text{Categorical}(\theta_1, \ldots, \theta_k)$

or

$\text{Categorical}(\theta_{1:k})$

parameters

Probability mass function (pmf)

$0 \leq \theta_j \leq 1$ for any $j \in [k]$

$\sum_{j=1}^{k} \theta_j = 1.0$

$C \sim \text{Categorical}(\theta_{1:k})$

$P(C = j) = \theta_j$

$P_C(j) = \theta_j$

$\theta_{1:k}$ as a vector

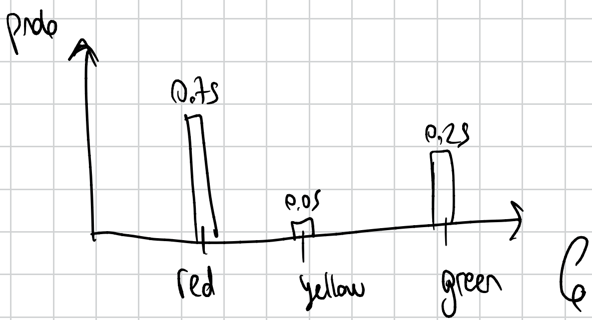
probability simplex

$\vec{\theta} = (\theta_1, \ldots, \theta_k)^T \in \Delta_{k-1} \subset \mathbb{R}^k$

We can visualise the pmf as a bar plot

$$C \sim \text{Categorical}\left(\overset{\text{red}}{0.7}, \overset{\text{yellow}}{0.05}, \overset{\text{green}}{0.25}\right)$$

to represent a traffic light



$$\theta_{red} = \theta_1$$

$$\theta_{yellow} = \theta_2$$

$$\theta_{green} = \theta_3$$

② Categorical distribution in NLP

categories
- words
- document classes
- topics
- Syntactic classes

$$C \sim \text{Categorical}\left(\overset{\text{negative}}{1/3}, \overset{\text{neutral}}{1/3}, \overset{\text{positive}}{1/3}\right)$$

for sentiment

$$W \sim \text{Categorical}\left(\overset{\text{the}}{\theta_1}, \dots \overset{\text{dog}}{\theta_{102}}, \theta_V\right)$$

for words

Conditional probability distribution (cpd)

$$W \mid C = \text{positive} \sim \text{Categorical}\left(\theta_1^{(pos)}, \dots, \theta_V^{(pos)}\right)$$

$$\theta^{(pos)} = \left(\theta_1^{(pos)}, \dots, \theta_V^{(pos)}\right)$$

$$W \mid C = \text{neutral} \sim \text{Categorical}\left(\theta_1^{(neu)}, \dots, \theta_V^{(neu)}\right)$$

$$W \mid C = \text{negative} \sim \text{Categorical}\left(\theta_1^{(neg)}, \dots, \theta_V^{(neg)}\right)$$

③ Tabular CPD

$P_{W|C}$ as a table

	w / c	1 a	2 an	3 the	4 some	... ...	101 cat	... ...	202 dog	... ...	V	
1	negative											← $\theta^{(neg)}$
2	neutral											← $\theta^{(neu)}$
3	positive											← $\theta^{(pos)}$

$\theta^{(neg)}_{some}$

$0 \leq \theta^{(pos)}_{dog} \leq 1$

(4) ESTIMATION

Maximum Likelihood Estimation (MLE)

i) gather data

$$D = \{ (c_n, w_n)_{n=1}^{N} \}$$

$$W \mid C = c \sim \text{Categorical}\left(\theta^{(c)}_{1:V}\right)$$

$$P(W = w \mid C = c) = \theta^{(c)}_{w}$$

$$P(W = \text{dog} \mid C = \text{pos}) = \theta^{(\text{pos})}_{\text{dog}}$$

$$P(W = \text{dog} \mid C = \text{neutral}) = \theta^{(\text{neu})}_{\text{dog}}$$

ii) log likelihood function

$$\mathcal{L}(\boldsymbol{\theta} \mid D) = \sum_{n=1}^{N} \log \theta^{(c_n)}_{w_n}$$

We look for a (local/global) optimum of $\mathcal{L}(\theta \mid D)$

$$\theta^{(c)}_{w} = \frac{\text{count}_{CW}(c, w)}{\sum_{\theta=1}^{V} \text{count}_{CW}(c, \theta)}$$

③ Complete Example

$$\mathcal{C} = \{\text{negative}, \text{neutral}, \text{positive}\} \qquad V=3$$

$$N=9$$

$$\mathcal{W} = \{\text{dog}, \text{cat}, \text{bird}\}$$

$$\theta_w^{(c)} = \frac{\text{count}_{CW}(c, w)}{\sum_{\theta=1}^{V} \text{count}_{CW}(c, \theta)}$$

$$D = \{ (\text{pos}, \text{dog}), (\text{pos}, \text{dog}), (\text{pos}, \text{cat}), (\text{neg}, \text{cat}), (\text{neg}, \text{cat}), (\text{neg}, \text{dog}), (\text{neutral}, \text{dog}), (\text{neu}, \text{bird}), (\text{neu}, \text{bird}) \}$$

$$W \mid C=c \sim \text{Categorical}\left(\theta^{(c)}_{1:V}\right)$$

c \ w	1 dog	2 cat	3 bird
negative	$\theta^{(neg)}_{dog}$	$\theta^{(neg)}_{cat}$	$\theta^{(neg)}_{bird}$
neutral	$\theta^{(neu)}_{dog}$	$\theta^{(neu)}_{cat}$	$\theta^{(neu)}_{bird}$
positive	$\theta^{(pos)}_{dog}$	$\theta^{(pos)}_{cat}$	$\theta^{(pos)}_{bird}$

$$\begin{array}{cccc} & \text{dog} & \text{cat} & \text{bird} \\ \theta^{(neg)}_{1:V} = & \left(1/3 \right. & 2/3 & \left. 0/3 \right) \end{array}$$

$$\theta^{(neu)}_{1:V} = \left(1/3 \quad 0/3 \quad 2/3 \right)$$

$$\theta^{(pos)}_{1:V} = \left(2/3 \quad 1/3 \quad 0/3 \right)$$

$$\begin{array}{ccc} \text{dog} & \text{cat} & \text{bird} \\ 1 & 2 & 0 \end{array}$$

$$\left(1 \quad 0 \quad 2 \right)$$

$$\left(2 \quad 1 \quad 0 \right)$$

$$\frac{1+0.1}{(1+0.1)+(2+0.1)+(0+0.1)} = \frac{1.1}{3+3\times 0.1} = \frac{1.1}{3.3}$$

$$\frac{0+0.1}{3.3} = \frac{0.1}{3.3}$$

$$\boxed{0.1}$$