Language Model Setup

- **Data**: Document d= $x_1\ x_2\ \ldots\ x_{|d|}$, $x_i \in V=\{w_1, \ldots, w_M\}$ is a word
- **Model**: Unigram LM θ(=topic) : $\{\theta_i=p(w_i|\theta)\}$, i=1, …, M; $\theta_1+\ldots+\theta_M=1$
- **Likelihood** function:

$$p(d|\theta) = p(x_1|\theta)\times\ldots\times p(x_{|d|}|\theta)$$
$$= p(w_1|\theta)^{c(w_1,d)}\times\ldots\times p(w_M|\theta)^{c(w_M,d)}$$
$$= \prod_{i=1}^{M} p(w_i|\theta)^{c(w_i,d)} = \prod_{i=1}^{M}\theta_i^{c(w_i,d)}$$

- ML **estimate:** $(\hat{\theta}_1,\ldots,\hat{\theta}_M) = \arg\max_{\theta_1,\ldots,\theta_M} p(d|\theta) = \arg\max_{\theta_1,\ldots,\theta_M}\prod_{i=1}^{M}\theta_i^{c(w_i,d)}$