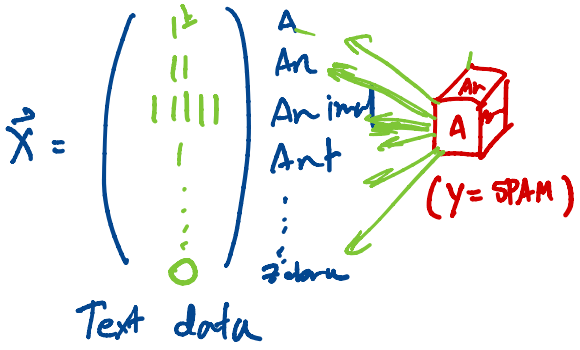

Our job is to estimate the probability of selecting the word w_d

$$P(w_d \mid y = \text{SPAM})$$

Remark: $P([X]_d = c \mid y = \text{SPAM}) = \binom{m}{c} \left(P(w_d \mid y = \text{SPAM})\right)^c$

where m is the total number of all words across X

Modeling distributions

- $P_\theta(w_d \mid y = \text{SPAM}) = \theta_{d,\text{SPAM}}$
- $P_\theta([X]_d = c \mid y = \text{SPAM}) = \binom{m}{c} \left(\theta_{d,\text{SPAM}}\right)^c$
- $P_\theta(\vec{X} = \vec{x} \mid y = \text{SPAM}) = \prod_{d=1}^{d} \binom{m}{x_d} \left(\theta_{d,\text{SPAM}}\right)^{x_d}$

$$= \binom{m}{x_1}\binom{m-x_1}{x_2}\binom{m-x_1-x_2}{x_3}\cdots\binom{m-(x_1+\cdots+x_{d-1})}{0} \cdot \prod_{d=1}^{d} \left(\theta_{d,\text{SPAM}}\right)^{x_d}$$

$$= \frac{m!}{x_1!(m-x_1)!} \cdot \frac{(m-x_1)!}{x_2!(m-x_1-x_2)!} \cdot \frac{(m-x_1-x_2)!}{x_3!(m-x_1-x_2-x_3)!} \cdot \ldots \cdot \frac{(m-(x_1+\cdots+x_{d-1}))!}{x_d!(m-(x_1+\cdots+x_d))!} \times \prod_{d=1}^{d} \left(\theta_{d,\text{SPAM}}\right)^{x_d}$$

$$= \frac{m!}{x_1!\,x_2!\,x_3!\cdots x_d!} \times \prod_{d=1}^{d} \left(\theta_{d,\text{SPAM}}\right)^{x_d}$$

Multinomial distribution

Notice that we only need to estimate $\theta_{d,\text{SPAM}}$

Estimate of $\theta_{\alpha, \text{SPAM}}$

- Intuitively $\theta_{\alpha, \text{SPAM}} \approx \frac{\#\text{ of times } w_\alpha \text{ appears in all spam emails}}{\#\text{ of words in all spam emails combine}}$

- Formally, $$\theta_{\alpha, \text{SPAM}} = \frac{\sum_{i=1}^{n} I(Y_i = \text{SPAM})[X_i]_\alpha}{\sum_{i=1}^{n} I(Y_i = \text{SPAM}) \left(\sum_{\alpha=1}^{d} [X_i]_\alpha \right)}$$