Dirichlet Prior Smoothing

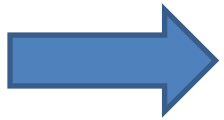
- Dirichlet distribution is a conjugate prior for multinomial sampling distribution

"pseudo" word counts $\alpha_i = \mu\, p(w_i|REF)$

$$Dir(\theta \mid \alpha_1, \ldots, \alpha_N) = \frac{\Gamma(\alpha_1 + \cdots + \alpha_N)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_N)} \prod_{i=1}^{N} \theta_i^{\alpha_i - 1}$$

X

$$p(d \mid \theta) = \frac{|d|!}{c(w_1)! \ldots c(w_N)!} \prod_{i=1}^{N} \theta_i^{c(w_i, d)}$$



$$p(\theta \mid d) = Dir(\theta \mid \alpha_1 + c(w_1), \ldots, \alpha_N + c(w_N))$$

$$= \frac{\Gamma(\alpha_1 + \cdots + \alpha_N + |d|)}{\Gamma(\alpha_1 + w_1) \cdots \Gamma(\alpha_N + w_N)} \prod_{i=1}^{N} \theta_i^{c(w_i) + \alpha_i - 1}$$