

Dirichlet Prior Smoothing (cont.)

Posterior distribution of parameters:

$$p(\theta | d) = \text{Dir}(\theta | c(w_1) + \alpha_1, \dots, c(w_N) + \alpha_N)$$

Property: If $\theta \sim \text{Dir}(\theta | \alpha)$, then $E(\theta) = \left\{ \frac{\alpha_i}{\sum \alpha_i} \right\}$

The predictive distribution is the same as the mean:

$$\begin{aligned} \hat{\theta}_i &= p(w_i | \hat{\theta}) = \int p(w_i | \theta) \text{Dir}(\theta | d) d\theta \\ &= \frac{c(w_i) + \alpha_i}{|d| + \sum_{i=1}^N \alpha_i} = \frac{c(w_i) + \mu p(w_i | REF)}{|d| + \mu} \end{aligned}$$



Dirichlet prior smoothing