

Dirichlet Prior Smoothing (Bayesian Smoothing)

- **Bayesian estimator** of multinomial distribution (unigram LM)
 - First consider posterior of parameters: $p(\theta | d) \propto p(d | \theta)p(\theta)$
 - Then, consider the mean or mode of the posterior distribution
- **Sampling distribution**(of data): $p(d | \theta)$
- **Prior** (on model parameters): $P(\theta)=p(\theta_1, \dots, \theta_N)$, where θ_i is probability of the i -th word in the vocabulary
- **Conjugate Prior**: intuitive & mathematically convenient
 - “encode” the prior as “**extra pseudo counts,**” which can be **conveniently combined with the observed actual counts**
 - $p(d | \theta)$ and $p(\theta)$ have the **same functional form**