Computation of Maximum Likelihood Estimate

Maximize p(d|θ) $(\hat{\theta}_1,\ldots,\hat{\theta}_M) = \arg\max_{\theta_1,\ldots,\theta_M} p(d\mid\theta) = \arg\max_{\theta_1,\ldots,\theta_M} \prod_{i=1}^{M} \theta_i^{c(w_i,d)}$

Max. Log-Likelihood $(\hat{\theta}_1,\ldots,\hat{\theta}_M) = \arg\max_{\theta_1,\ldots,\theta_M} \log[p(d\mid\theta)] = \arg\max_{\theta_1,\ldots,\theta_M} \sum_{i=1}^{M} c(w_i,d)\log\theta_i$

Subject to constraint: $\sum_{i=1}^{M}\theta_i = 1$

Use Lagrange multiplier approach

Lagrange function: $f(\theta\mid d) = \sum_{i=1}^{M} c(w_i,d)\log\theta_i + \lambda(\sum_{i=1}^{M}\theta_i - 1)$

$$\frac{\partial f(\theta\mid d)}{\partial\theta_i} = \frac{c(w_i,d)}{\theta_i} + \lambda = 0 \quad\longrightarrow\quad \theta_i = -\frac{c(w_i,d)}{\lambda}$$

$$\sum_{i=1}^{M} -\frac{c(w_i,d)}{\lambda} = 1 \longrightarrow \lambda = -\sum_{i=1}^{N} c(w_i,d) \longrightarrow \hat{\theta}_i = p(w_i\mid\hat{\theta}) = \frac{c(w_i,d)}{\sum_{i=1}^{M} c(w_i,d)} = \frac{c(w_i,d)}{|d|}$$

Normalized Counts