Solution Bayesian Network

---

see http://deeploria.gforge.inria.fr/cours/solutions/exo2TP.py

---

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)}$$

so

$$\max_\theta p(\theta|x) = \max_\theta p(x|\theta)p(\theta)$$

Likelihood term:

$$p(x|\theta) = \prod_t^T p(x_t|\theta) = \prod_t^T \theta_{x_t}$$

We group together all occurrences of the same word:

$$p(x|\theta) = \prod_i^K \theta_i^{n_i}$$

with K the vocabulary size.

Now the Dirichlet prior:

$$p(\theta) = \frac{1}{B(\alpha)} \prod_i^K \theta_i^{\alpha_i - 1}$$

Combining both:

$$p(\theta|x) \propto \frac{1}{B(\alpha)} \prod_i^K \theta_i^{n_i + \alpha_i - 1}$$

$$p(\theta|x) \propto \frac{1}{B(\alpha)} \prod_i^K \theta_i^{n_i + \alpha_i - 1}$$

We know this is a Dirichlet, so

$$p(\theta|x) \sim \text{Dirichlet}([\alpha_i + n_i]_{1 \leq i \leq K})$$

To compute the proba of an unknown work:

$$p(x) = \int p(x|\theta)p(\theta)d\theta$$

see the derivation in https://people.eecs.berkeley.edu/~stephentu/writeups/dirichlet-conjugate-prior.pdf