Modeling Cables Events

Allison J.B. Chaney

August 25, 2015

1 Generative model

We start with a fitted LDA model where documents are represented in terms of topics (θ, a $D \times K$ matrix), and topics are represented as a distribution over words (β, a $K \times V$ matrix). This model fit, along with document metadata, are our observations. Alternatively, we can tack on the LDA generative process to the model below.

- for each day i with date a_i:
  - generate event occurrence/strength $\epsilon \sim \text{Poisson}(\eta_\epsilon)$, where η_ϵ is a fixed, non-negative hyperparameter for the mean event strength
  - generate the day/event's description in terms of each topic k: $\pi_{ik} \sim \text{Gamma}(\alpha_0, \beta_0)$, where α_0 and β_0 are fixed hyperparameters.
- draw the entity's base topics: $\phi_{0k} \sim \text{Gamma}(\alpha, \beta)$ (eventually for each entity, but for now, just limit data to only one entity)
- For each cable j on date c_j:
  - set cable topic parameter: $\phi_{jk} = \phi_{0k} + \sum_i f(a_i, c_j)\pi_{ik}\epsilon_i$, where f is defined below.
  - draw cable topic: $\theta_{jk} \sim \text{Gamma}(\beta_c \phi_{jk}, \beta_c)$

Note that

$$f(a,c) = \begin{cases} 1 - \frac{c-a}{d}, & \text{if } a \le c < a + d \\ 0, & \text{otherwise,} \end{cases}$$

where d is the time distance (in days) after event a at which point the event is no longer relevant.

2 Inference

For now, we assume that we know the LDA topics β and only observe the documents in terms of their topics θ; breaking this assumption makes inference a little more complicated as the updates for θ would have new dependencies. Pending the results, we should explore that vein.

Here, we follow the structure of the DEF paper to explain inference for this model.

- As usual, inference is the central computational problem.
- Variational inference minimizes the KL divergence from an approximating distribution q to the true posterior p.

- This is equivalent to maximizing the ELBO: $\mathscr{L}(q) = \mathbb{E}_{q(\epsilon,\pi,\phi)}[\log p(\theta,\epsilon,\pi,\phi) - \log q(\epsilon,\pi,\phi)]$
- we define the approximating distribution q using the mean field assumption: $q(\epsilon,\pi,\phi) = \prod_i q(\epsilon_i)\prod_k \left[q(\phi_{0k})\prod_i q(\pi_{ik})\right]$
- $q(\pi)$ and $q(\phi)$ are both gamma-distributed with variational parameters λ^π and λ^ϕ, respectively, where we use the softmax function $\mathscr{M}(x) = \log(1+\exp(x))$ to constrain the free variational parameters. $q(\epsilon)$ is Poisson-distributed with variational parameter λ^ϵ; the free parameter is also constrained by the softmax function.
- the expectations under q (needed to maximize the ELBO) do not have a simple analytic form, so we use "black box" VI techniques
- for each variable, we can write the log probability of all terms containing that variable, giving us

$$\log p_i^\epsilon(\theta,\epsilon,\pi,\phi) = \log p(\epsilon_i \mid \eta_\epsilon) + \sum_{j:f(a_i,c_j)\neq 0}\sum_k \log p(\theta_{jk} \mid \phi_{0k}, c_j, a_i, d, \beta_c, \pi_{ik}, \epsilon_i),$$

$$\log p_{ik}^\pi(\theta,\epsilon,\pi,\phi) = \log p(\pi_{ik} \mid \alpha_0, \beta_0) + \sum_{j:f(a_i,c_j)\neq 0} \log p(\theta_{jk} \mid \phi_{0k}, c_j, a_i, d, \beta_c, \pi_{ik}, \epsilon_i),$$

and

$$p_k^\phi(\theta,\epsilon,\pi,\phi) = \log p(\phi_{0k} \mid \alpha, \beta) + \sum_i\sum_j \log p(\theta_{jk} \mid \phi_{0k}, c_j, a_i d, \beta_c, \pi_{ik}, \epsilon_i).$$

- Then we can write the gradients with respect to the variational parameters as:

$$\nabla_{\lambda_i^\epsilon}\mathscr{L} = \mathbb{E}_q\left[\nabla_{\lambda_i^\epsilon}\log q(\epsilon_i \mid \lambda_i^\epsilon)\left(\log p_i^\epsilon(\theta,\epsilon,\pi,\phi) - \log q(\epsilon_i \mid \lambda_i^\epsilon)\right)\right],$$

$$\nabla_{\lambda_{ik}^\pi}\mathscr{L} = \mathbb{E}_q\left[\nabla_{\lambda_{ik}^\pi}\log q(\pi_{ik} \mid \lambda_{ik}^\pi)\left(\log p_{ik}^\pi(\theta,\epsilon,\pi,\phi) - \log q(\pi_{ik} \mid \lambda_{ik}^\pi)\right)\right],$$

and

$$\nabla_{\lambda_k^\phi}\mathscr{L} = \mathbb{E}_q\left[\nabla_{\lambda_k^\phi}\log q(\phi_{0k} \mid \lambda_k^\phi)\left(\log p_k^\phi(\theta,\epsilon,\pi,\phi) - \log q(\phi_{0k} \mid \lambda_k^\phi)\right)\right].$$

Using this framework, we construct our black box algorithm below.

For Reference The gamma distribution and derivatives:

$$\log \mathrm{Gamma}(x \mid a, b) = ab\log b - \log\Gamma(ab) + (ab-1)\log x - bx \tag{1}$$

$$\nabla_a \log \mathrm{Gamma}(x \mid \mathscr{M}(a), \mathscr{M}(b)) = \mathscr{M}'(a)\mathscr{M}(b)[\log\mathscr{M}(b) - \Psi(\mathscr{M}(a)\mathscr{M}(b)) + \log x] \tag{2}$$

$$\nabla_b \log \mathrm{Gamma}(x \mid \mathscr{M}(a), \mathscr{M}(b)) = \mathscr{M}'(b)[\mathscr{M}(a)((\log\mathscr{M}(b)+1) - \Psi(\mathscr{M}(a)\mathscr{M}(b)) + \log x) - x] \tag{3}$$

The Poisson distribution and derivative:

$$\log \mathrm{Poisson}(x \mid \lambda) = x\log\lambda - \log(x!) - \lambda \tag{4}$$

$$\nabla_\lambda \log \mathrm{Poisson}(x \mid \mathscr{M}(\lambda)) = \mathscr{M}'(\lambda)\left[\frac{x}{\mathscr{M}(\lambda)} - 1\right]. \tag{5}$$

The softmax function and derivative:

$$\mathscr{M}(x) = \log(1+e^x)$$

$$\mathscr{M}'(x) = \frac{e^x}{1+e^x}$$