

# Modeling Cables Events

Allison J.B. Chaney

August 27, 2015

## 1 Introduction

Our objective is to detect historical events.

## 2 Related Work

- BBVI, DEFs; this deals with finding updates for non-conjugate models
- non uniform sampling (MMSB in PNAS); we want events to be represented appropriately. Also: how to update global stuff from non-uniform document sampling—it's a bit of challenge)
- spike and slab (e.g., Empirical study of SVI for Beta Bernoulli Process; also IBP papers) because we have variables that fit this pattern
- SSVI; this helps us with the event variables—they are bound together (is this the real purpose here?)

## 3 Generative model

We start with a fitted LDA model where documents are represented in terms of topics ( $\theta$ , a  $D \times K$  matrix), and topics are represented as a distribution over words ( $\beta$ , 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  $\eta_\epsilon$  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  $\alpha_0$  and  $\beta_0$  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

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

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

## 4 Inference

For now, we assume that we know the LDA topics  $\beta$  and only observe the documents in terms of their topics  $\theta$ ; breaking this assumption makes inference a little more complicated as the updates for  $\theta$  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:  $\mathcal{L}(q) = \mathbb{E}_{q(\epsilon, \pi, \phi)}[\log p(\theta, \epsilon, \pi, \phi) - \log q(\epsilon, \pi, \phi)]$
- we define the approximating distribution  $q$  using the mean field assumption:  $q(\epsilon, \pi, \phi) = \prod_i q(\epsilon_i) \prod_k [q(\phi_{0k}) \prod_i q(\pi_{ik})]$
- $q(\pi)$  and  $q(\phi)$  are both gamma-distributed with variational parameters  $\lambda^\pi$  and  $\lambda^\phi$ , respectively, where we use the softmax function  $\mathcal{M}(x) = \log(1 + \exp(x))$  to constrain the free variational parameters.  $q(\epsilon)$  is Poisson-distributed with variational parameter  $\lambda^\epsilon$ ; 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

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

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

and

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

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

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

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

and

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

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

**For Reference** The gamma distribution and derivatives:

$$\log \text{Gamma}(x | a, b) = a b \log b - \log \Gamma(ab) + (ab - 1) \log x - bx \quad (1)$$

$$\nabla_a \log \text{Gamma}(x | \mathcal{M}(a), \mathcal{M}(b)) = \mathcal{M}'(a) \mathcal{M}(b) [\log \mathcal{M}(b) - \Psi(\mathcal{M}(a) \mathcal{M}(b)) + \log x] \quad (2)$$

$$\nabla_b \log \text{Gamma}(x | \mathcal{M}(a), \mathcal{M}(b)) = \mathcal{M}'(b) [\mathcal{M}(a) ((\log \mathcal{M}(b) + 1) - \Psi(\mathcal{M}(a) \mathcal{M}(b)) + \log x) - x] \quad (3)$$

The Poisson distribution and derivative:

$$\log \text{Poisson}(x | \lambda) = x \log \lambda - \log(x!) - \lambda \quad (4)$$

$$\nabla_\lambda \log \text{Poisson}(x | \mathcal{M}(\lambda)) = \mathcal{M}'(\lambda) \left[ \frac{x}{\mathcal{M}(\lambda)} - 1 \right]. \quad (5)$$

The softmax function and derivative:

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

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

---

**Algorithm 1:** Inference for Cables Model

---

**Input:** document topics  $\theta$

**Output:** estimates of latent parameters entity topics  $\pi$ , event topics  $\phi$ , and event occurrences  $\epsilon$

**Initialize**  $\lambda^\pi$ ,  $\lambda^\phi$ , and  $\lambda^\epsilon$  to respective priors

**Initialize** iteration count  $t = 0$

**while** *change in validation likelihood*  $< \delta$  **do**

$t += 1$

    initialize  $\hat{\lambda}^\pi$ ,  $\hat{\lambda}^\phi$ , and  $\hat{\lambda}^\epsilon$  to 0 matrices

**for**  $j = 1, \dots, B$  (*batch size, or # of document samples*) **do**

        Sample a document  $j$  that is sent on date  $c_j$  and has topics  $\theta_j$

**for**  $s = 1, \dots, S$  **do**

            test

**end**

**end**

    set  $\rho = (t + \tau)^\kappa$

    update event content for each event  $i$  and topic  $k$ :  $\lambda_{ik}^\pi = \lambda_{ik}^\pi + \rho \hat{\lambda}_{ik}^\pi$

    update general entity topics for each topic  $k$ :  $\lambda_k^\phi = \lambda_k^\phi + \rho \hat{\lambda}_k^\phi$

    update event occurrences for each event  $i$ :  $\lambda_i^\epsilon = \lambda_i^\epsilon + \rho \hat{\lambda}_i^\epsilon$

    set  $\mathbb{E}[\pi] = \lambda^{\pi,a}$

    set  $\mathbb{E}[\phi] = \lambda^{\phi,a}$

    set  $\mathbb{E}[\epsilon] = \lambda^\epsilon$

**end**

**return**  $\mathbb{E}[\pi]$ ,  $\mathbb{E}[\phi]$ ,  $\mathbb{E}[\epsilon]$

---