

Modeling Cables Events

Allison J.B. Chaney

August 10, 2015

1 Model v1

1.1 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 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: $\phi_{jk} = \phi_{0k} + \sum_i f(a_i, c_j) \pi_{ik}$, 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.

1.2 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: $\mathcal{L}(q) = \mathbb{E}_{q(\pi, \phi)} [\log p(\theta, \pi, \phi) - \log q(\pi, \phi)]$
- we define the approximating distribution q using the mean field assumption: $q(\pi, \phi) = \prod_k [q(\phi_{0k}) \prod_i q(\pi_{ik})]$
- $q(\pi)$ and $q(\phi)$ are both gamma-distributed, with variational parameters λ^π and λ^ϕ , respectively

- 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 probability of all terms containing that variable, giving us

$$p_{ik}^{\pi}(\theta, \pi, \phi) = p(\pi_{ik} | \alpha_0, \beta_0) \prod_j p(\theta_{jk} | \phi_{0k}, c, a, d, \beta_c, \pi)$$

and

$$p_k^{\phi}(\theta, \pi, \phi) = p(\phi_{0k} | \alpha, \beta) \prod_j p(\theta_{jk} | \phi_{0k}, c, a, d, \beta_c, \pi).$$

We can also write these in their log forms:

$$\log p_{ik}^{\pi}(\theta, \pi, \phi) = \log p(\pi_{ik} | \alpha_0, \beta_0) + \sum_j \log p(\theta_{jk} | \phi_{0k}, c, a, d, \beta_c, \pi)$$

and

$$\log p_k^{\phi}(\theta, \pi, \phi) = \log p(\phi_{0k} | \alpha, \beta) + \sum_j \log p(\theta_{jk} | \phi_{0k}, c, a, d, \beta_c, \pi).$$

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

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

and

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

Using this framework, we construct our black box algorithm below. The messiness comes from the full expansions of p , q , and $\nabla_{\lambda} q$.

1.3 results

Able to identify some simulated data.

2 Model v2

2.1 Generative model

We want to make sure that the event occurrence can be modeled by a Poisson process. Except since only one event can occur per date, we model it with a Bernoulli Process.

- for each day i with date a_i :
 - generate whether or not an event occurs $\epsilon \sim \text{Bernoulli}(\eta_{\epsilon})$, where η_{ϵ} is the probability of an event, which should be about the same as the rate for a Poisson process (but capped at 1).
 - 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.
- same as before, but using softmax: $\mathcal{M}(x) = \log(1 + \exp(x))$ to transform ϕ_{jk} when drawing cable topic. Note the the derivative of the softmax function is $\mathcal{M}'(x) = \exp(x)/(1 + \exp(x))$. Similarly, we use the sigmoid function $\mathcal{S}(x) = 1/(1 + e^{-x})$ when drawing the Bernoulli from its variational parameters; there derivative of this function is $\mathcal{S}'(x) = -e^{-x}/(1 + e^{-x})^2$.
- one other change: for each cable j , set cable topic parameter: $\phi_{jk} = \phi_{0k} + \sum_i \epsilon_i f(a_i, c_j) \pi_{ik}$.

Algorithm 1 Black Box variational inference for Cables Model v1

```
1: Input: document topics  $\theta$ 
2: Initialize  $\lambda^\pi$  and  $\lambda^\phi$  randomly
3: Initialize  $t_i^\pi = 0$  for all days  $i$  and  $t^\phi = 0$ 
4: repeat
5:   Sample a document  $j$  that is sent on date  $c_j$  and has topics  $\theta_j$ 
6:    $t^\phi += 1$ 
7:   for  $s = 1, \dots, S$  do
8:     for  $k = 1, \dots, K$  do
9:       draw sample  $\phi_{0k}[s] \sim \text{Gamma}(\lambda_k^\phi)$ 
10:      for each event  $i$  on date  $a_i \in (c_j - d, c_j]$  do
11:        draw sample  $\pi_{ik}[s] \sim \text{Gamma}(\lambda_{ik}^\pi)$ 
12:         $t_i^\pi += 1$ 
13:      set  $\phi_{jk}[s] = \phi_{0k}[s] + \sum_{i: a_i \in (c_j - d, c_j]} \pi_{ik}[s] f(a_i, c_j)$ 
14:      set  $p_k^\theta[s] = \log p(\theta_j | \phi_{jk}[s], \beta_c) = \beta_c \phi_{jk}[s] \log \beta_c - \log \Gamma(\beta_c \phi_{jk}[s]) + (\beta_c \phi_{jk}[s] - 1) \log \theta_k - \beta_c \theta_k$ 
15:      set  $p_k^\phi[s] = \log p(\phi_{0k} | \alpha, \beta) = \alpha \log \beta - \log \Gamma(\alpha) + (\alpha - 1) \log \phi_{0k}[s] - \beta \phi_{0k}[s]$ 
16:      set (for each relevant  $i$ )  $p_{ik}^\pi[s] = \log p(\pi_{ik} | \alpha_0, \beta_0) = \alpha_0 \log \beta_0 - \log \Gamma(\alpha_0) + (\alpha_0 - 1) \log \pi_k[s] - \beta_0 \pi_{ik}[s]$ 
17:      set  $q_k^\phi[s] = \log q(\phi_{0k} | \lambda_k^\phi) = \lambda_k^{\phi, \alpha} \log \lambda_k^{\phi, \beta} - \log \Gamma(\lambda_k^{\phi, \alpha}) + (\lambda_k^{\phi, \alpha} - 1) \log \phi_{0k}[s] - \lambda_k^{\phi, \beta} \phi_{0k}[s]$ 
18:      set (for each relevant  $i$ )  $q_{ik}^\pi[s] = \log q(\pi_{ik} | \lambda_{ik}^\pi) = \lambda_{ik}^{\pi, \alpha} \log \lambda_{ik}^{\pi, \beta} - \log \Gamma(\lambda_{ik}^{\pi, \alpha}) + (\lambda_{ik}^{\pi, \alpha} - 1) \log \pi_{ik}[s] - \lambda_{ik}^{\pi, \beta} \pi_{ik}[s]$ 
19:      set  $g_k^\phi[s] = \nabla_{\lambda_k^\phi} \log q(\phi_{0k} | \lambda_k^\phi) = \langle \log \lambda_k^{\phi, \beta} - \Psi(\lambda_k^{\phi, \alpha}) + \log \phi_{0k}[s], \lambda_k^{\phi, \alpha} / \lambda_k^{\phi, \beta} - \phi_{0k}[s] \rangle$ 
20:      set (for each relevant  $i$ )  $g_{ik}^\pi[s] = \nabla_{\lambda_{ik}^\pi} \log q(\pi_{ik} | \lambda_{ik}^\pi) = \langle \log \lambda_{ik}^{\pi, \beta} - \Psi(\lambda_{ik}^{\pi, \alpha}) + \log \pi_{ik}[s], \lambda_{ik}^{\pi, \alpha} / \lambda_{ik}^{\pi, \beta} - \pi_{ik}[s] \rangle$ 
21:      for each event  $i$  on date  $a_i \in (c_j - d, c_j]$  do
22:        set  $\hat{\lambda}_{ik}^\pi = \frac{1}{S} \sum_s g_{ik}^\pi[s] (p_{ik}^\pi[s] + J p_k^\theta[s] - q_{ik}^\pi[s])$ 
23:        set  $\rho_i^\pi = (t_i^\pi + \tau)^\kappa$ 
24:        update  $\lambda_{ik}^\pi = \lambda_{ik}^\pi + \rho_i^\pi \hat{\lambda}_{ik}^\pi$ 
25:      set  $\hat{\lambda}_k^\phi = \frac{1}{S} \sum_s g_k^\phi[s] (p_k^\phi[s] + J p_k^\theta[s] - q_k^\phi[s])$ 
26:      set  $\rho^\phi = (t^\phi + \tau)^\kappa$ 
27:      update  $\lambda_k^\phi = \lambda_k^\phi + \rho^\phi \hat{\lambda}_k^\phi$ 
28: until change in validation likelihood  $< \delta$ 
```

2.2 Inference

- Minimizing the KL divergence is equivalent to maximizing the ELBO: $\mathcal{L}(q) = \mathbb{E}_{q(\pi, \phi, \epsilon)} [\log p(\theta, \pi, \phi, \epsilon) - \log q(\pi, \phi, \epsilon)]$
- we define the approximating distribution q using the mean field assumption: $q(\pi, \phi, \epsilon) = \prod_k [q(\phi_{0k}) \prod_i q(\pi_{ik})] \prod_i q(\epsilon_i)$
- $q(\pi)$ and $q(\phi)$ are both gamma-distributed, with variational parameters λ^π and λ^ϕ , respectively. $q(\epsilon)$ is Bernoulli-distributed and is parameterized by λ_ϵ
- 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 probability of all terms containing that variable, giving us

$$p_{ik}^\pi(\theta, \pi, \phi, \epsilon) = p(\pi_{ik} | \alpha_0, \beta_0) \prod_j p(\theta_{jk} | \phi_{0k}, c, a, d, \beta_c, \pi, \epsilon),$$

$$p_k^\phi(\theta, \pi, \phi, \epsilon) = p(\phi_{0k} | \alpha, \beta) \prod_j p(\theta_{jk} | \phi_{0k}, c, a, d, \beta_c, \pi, \epsilon),$$

and

$$p_i^\epsilon(\theta, \pi, \phi, \epsilon) = p(\epsilon_i | \eta_\epsilon) \prod_j p(\theta_{jk} | \phi_{0k}, c, a, d, \beta_c, \pi, \epsilon).$$

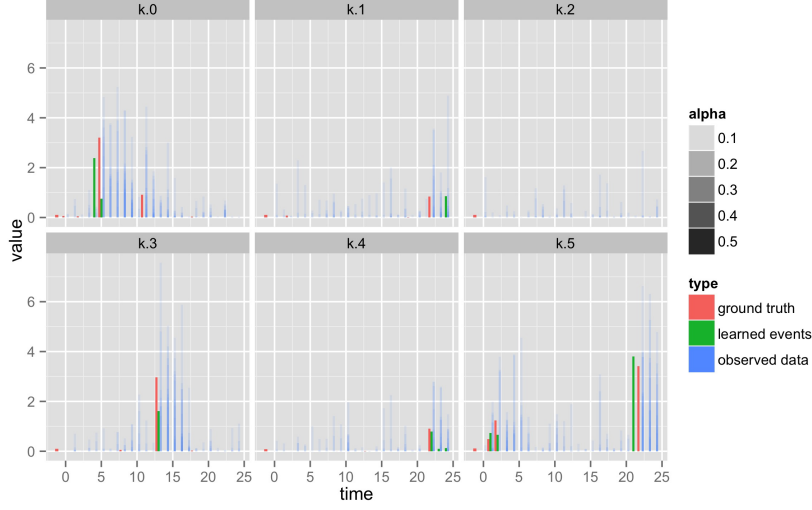


Figure 1: Identifying simulated events.

We can also write these in their log forms:

$$\log p_{ik}^{\pi}(\theta, \pi, \phi, \epsilon) = \log p(\pi_{ik} | \alpha_0, \beta_0) + \sum_j p(\theta_{jk} | \phi_{0k}, c, a, d, \beta_c, \pi, \epsilon),$$

$$\log p_k^{\phi}(\theta, \pi, \phi, \epsilon) = \log p(\phi_{0k} | \alpha, \beta) + \sum_j \log p(\theta_{jk} | \phi_{0k}, c, a, d, \beta_c, \pi, \epsilon),$$

and

$$\log p_i^{\epsilon}(\theta, \pi, \phi, \epsilon) = \log p(\epsilon_i | \eta_{\epsilon}) + \sum_j \sum_k \log p(\theta_{jk} | \phi_{0k}, c, a, d, \beta_c, \pi, \epsilon).$$

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

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

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

and

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

Using this framework, we construct our black box algorithm below. The messiness comes from the full expansions of p , q , and $\nabla_{\lambda} q$.

Recall the following for reference.

$$\log \text{Gamma}(x | \mathcal{M}(a), \mathcal{M}(b)) = \mathcal{M}(a) \log \mathcal{M}(b) - \log \Gamma \mathcal{M}(a) + (\mathcal{M}(a) - 1) \log x - \mathcal{M}(b)x$$

$$\frac{d}{da} \log \text{Gamma}(x | \mathcal{M}(a), \mathcal{M}(b)) = \mathcal{M}'(a) \log \mathcal{M}(b) - \Psi \mathcal{M}(a) \mathcal{M}'(a) + \mathcal{M}'(a) \log x$$

$$\frac{d}{db} \log \text{Gamma}(x | \mathcal{M}(a), \mathcal{M}(b)) = \mathcal{M}(a) \frac{\mathcal{M}'(b)}{\mathcal{M}(b)} - \mathcal{M}'(b)x$$

$$\log \text{Bernoulli}(x | \mathcal{S}(\lambda)) = x \log \mathcal{S}(\lambda) + (1 - x) \log(1 - \mathcal{S}(\lambda))$$

$$\frac{d}{d\lambda} \log \text{Bernoulli}(x | \mathcal{S}(\lambda)) = x \frac{\mathcal{S}'(\lambda)}{\mathcal{S}(\lambda)} + (1 - x) \frac{\mathcal{S}'(\lambda)}{1 - \mathcal{S}(\lambda)}$$

Specific equations used in algorithm, fully expanded Prior probabilities.

$$\log p(\theta_j | \phi_{jk}[s], \beta_c) = \beta_c \phi_{jk}[s] \log \beta_c - \log \Gamma(\beta_c \phi_{jk}[s]) + (\beta_c \phi_{jk}[s] - 1) \log \theta_k - \beta_c \theta_k \quad (1)$$

$$\log p(\phi_{0k}[s] | \alpha, \beta) = \alpha \log \beta - \log \Gamma(\alpha) + (\alpha - 1) \log \phi_{0k}[s] - \beta \phi_{0k}[s] \quad (2)$$

$$\log p(\pi_{ik}[s] | \alpha_0, \beta_0) = \alpha_0 \log \beta_0 - \log \Gamma(\alpha_0) + (\alpha_0 - 1) \log \pi_{ik}[s] - \beta_0 \pi_{ik}[s] \quad (3)$$

$$\log p(\epsilon_i[s] | \eta) = \epsilon_i[s] \log \eta + (1 - \epsilon_i[s]) \log(1 - \eta) \quad (4)$$

Probabilities of the latent variables given the free variational parameters.

$$\log q(\phi_{0k}[s] | \lambda_k^\phi) = \mathcal{M}(\lambda_k^{\phi, \alpha}) \log \mathcal{M}(\lambda_k^{\phi, \beta}) - \log \Gamma(\mathcal{M}(\lambda_k^{\phi, \alpha})) + (\mathcal{M}(\lambda_k^{\phi, \alpha}) - 1) \log \phi_{0k}[s] - \mathcal{M}(\lambda_k^{\phi, \beta}) \phi_{0k}[s] \quad (5)$$

$$\log q(\pi_{ik}[s] | \lambda_{ik}^\pi) = \mathcal{M}(\lambda_{ik}^{\pi, \alpha}) \log \mathcal{M}(\lambda_{ik}^{\pi, \beta}) - \log \Gamma(\mathcal{M}(\lambda_{ik}^{\pi, \alpha})) + (\mathcal{M}(\lambda_{ik}^{\pi, \alpha}) - 1) \log \pi_{ik}[s] - \mathcal{M}(\lambda_{ik}^{\pi, \beta}) \pi_{ik}[s] \quad (6)$$

$$\log q(\epsilon_i[s] | \lambda_i^\epsilon) = \epsilon_i[s] \log \mathcal{S}(\lambda_i^\epsilon) + (1 - \epsilon_i[s]) \log(1 - \mathcal{S}(\lambda_i^\epsilon)) \quad (7)$$

Gradients of variational distributions.

$$\nabla_{\lambda_k^\phi} \log q(\phi_{0k}[s] | \lambda_k^\phi) = \left\langle \mathcal{M}'(\lambda_k^{\phi, \alpha}) \left(\log \mathcal{M}(\lambda_k^{\phi, \beta}) - \Psi(\mathcal{M}(\lambda_k^{\phi, \alpha})) + \log \phi_{0k}[s] \right), \mathcal{M}'(\lambda_k^{\phi, \beta}) \left(\frac{\mathcal{M}(\lambda_k^{\phi, \alpha})}{\mathcal{M}(\lambda_k^{\phi, \beta})} - \phi_{0k}[s] \right) \right\rangle \quad (8)$$

$$\nabla_{\lambda_{ik}^\pi} \log q(\pi_{ik}[s] | \lambda_{ik}^\pi) = \left\langle \mathcal{M}'(\lambda_{ik}^{\pi, \alpha}) \left(\log \mathcal{M}(\lambda_{ik}^{\pi, \beta}) - \Psi(\mathcal{M}(\lambda_{ik}^{\pi, \alpha})) + \log \pi_{ik}[s] \right), \mathcal{M}'(\lambda_{ik}^{\pi, \beta}) \left(\frac{\mathcal{M}(\lambda_{ik}^{\pi, \alpha})}{\mathcal{M}(\lambda_{ik}^{\pi, \beta})} - \pi_{ik}[s] \right) \right\rangle \quad (9)$$

$$\nabla_{\lambda_i^\epsilon} \log q(\epsilon_i | \lambda_i^\epsilon) = \epsilon_i \frac{\mathcal{S}'(\lambda_i^\epsilon)}{\mathcal{S}(\lambda_i^\epsilon)} + (1 - \epsilon_i) \frac{\mathcal{S}'(\lambda_i^\epsilon)}{1 - \mathcal{S}(\lambda_i^\epsilon)} \quad (10)$$

$$(11)$$

3 Model v3

Changes: different gamma parameterization and adding control variates.

$$\text{Gamma}(\alpha\beta, \beta)$$

3.1 inference

We fix the gamma distribution's β variational parameter for π and ϕ . Then:

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

$$\frac{d}{da} \log \text{Gamma}(x | \mathcal{M}(a), b) = \mathcal{M}'(a)b \log b - \Psi(\mathcal{M}(a)b) \mathcal{M}'(a)b + \mathcal{M}'(a)b \log x$$

Specific equations used in algorithm, fully expanded Prior probabilities.

$$\log p(\theta_j | \phi_{jk}[s], \beta_c) = \beta_c \phi_{jk}[s] \log \beta_c - \log \Gamma(\beta_c \phi_{jk}[s]) + (\beta_c \phi_{jk}[s] - 1) \log \theta_k - \beta_c \theta_k \quad (12)$$

$$\log p(\phi_{0k}[s] | \alpha, \beta) = \alpha\beta \log \beta - \log \Gamma(\alpha\beta) + (\alpha\beta - 1) \log \phi_{0k}[s] - \beta \phi_{0k}[s] \quad (13)$$

$$\log p(\pi_{ik}[s] | \alpha_0, \beta_0) = \alpha_0 \beta_0 \log \beta_0 - \log \Gamma(\alpha_0 \beta_0) + (\alpha_0 \beta_0 - 1) \log \pi_{ik}[s] - \beta_0 \pi_{ik}[s] \quad (14)$$

$$\log p(\epsilon_i[s] | \eta) = \epsilon_i[s] \log \eta + (1 - \epsilon_i[s]) \log(1 - \eta) \quad (15)$$

Probabilities of the latent variables given the free variational parameters.

$$\log q(\phi_{0k}[s] | \lambda_k^\phi) = \mathcal{M}(\lambda_k^\phi)\beta \log \beta - \log \Gamma(\mathcal{M}(\lambda_k^\phi)\beta) + (\mathcal{M}(\lambda_k^\phi)\beta - 1) \log \phi_{0k}[s] - \beta \phi_{0k}[s] \quad (16)$$

$$\log q(\pi_{ik}[s] | \lambda_{ik}^\pi) = \mathcal{M}(\lambda_{ik}^\pi)\beta_0 \log \beta_0 - \log \Gamma(\mathcal{M}(\lambda_{ik}^\pi)\beta_0) + (\mathcal{M}(\lambda_{ik}^\pi)\beta_0 - 1) \log \pi_{ik}[s] - \beta_0 \pi_{ik}[s] \quad (17)$$

$$\log q(\epsilon_i[s] | \lambda_i^\epsilon) = \epsilon_i[s] \log \mathcal{S}(\lambda_i^\epsilon) + (1 + \epsilon_i[s]) \log(1 - \mathcal{S}(\lambda_i^\epsilon)) \quad (18)$$

Gradients of variational distributions.

$$\nabla_{\lambda_k^\phi} \log q(\phi_{0k}[s] | \lambda_k^\phi) = \mathcal{M}'(\lambda_k^\phi)\beta \left(\log \beta - \Psi(\mathcal{M}(\lambda_k^\phi)\beta) + \log \phi_{0k}[s] \right) \quad (19)$$

$$\nabla_{\lambda_{ik}^\pi} \log q(\pi_{ik}[s] | \lambda_{ik}^\pi) = \mathcal{M}'(\lambda_{ik}^\pi)\beta_0 (\log \beta_0 - \Psi(\mathcal{M}(\lambda_{ik}^\pi)\beta_0) + \log \pi_{ik}[s]) \quad (20)$$

$$\nabla_{\lambda_{ik}^\epsilon} \log q(\epsilon_i[s] | \lambda_i^\epsilon) = \epsilon_i \frac{\mathcal{S}'(\lambda_i^\epsilon)}{\mathcal{S}(\lambda_i^\epsilon)} + (1 - \epsilon_i) \frac{\mathcal{S}'(\lambda_i^\epsilon)}{1 - \mathcal{S}(\lambda_i^\epsilon)} \quad (21)$$

$$(22)$$

Algorithm 2 Black Box variational inference for Cables Model v2

```

1: Input: document topics  $\theta$ 
2: Initialize  $\lambda^\pi$ ,  $\lambda^\phi$ , and  $\lambda^\epsilon$  to respective priors
3: Initialize iteration count  $t = 0$ 
4: repeat
5:    $t += 1$ 
6:   initialize  $\hat{\lambda}^\pi$ ,  $\hat{\lambda}^\phi$ , and  $\hat{\lambda}^\epsilon$  to 0 matrices
7:   for  $s = 1, \dots, S$  do
8:     for  $k = 1, \dots, K$  do
9:       draw sample  $\phi_{0k}[s] \sim \text{Gamma}(\mathcal{M}(\lambda_k^\phi))$ 
10:      set  $p_k^\phi[s] = \log p(\phi_{0k}[s] | \alpha, \beta)$  ▷ see Eqn. 2
11:      set  $q_k^\phi[s] = \log q(\phi_{0k} | \lambda_k^\phi)$  ▷ see Eqn. 5
12:      set  $g_k^\phi[s] = \nabla_{\lambda_k^\phi} \log q(\phi_{0k}[s] | \lambda_k^\phi)$  ▷ see Eqn. 8
13:      for each event  $i$  on date  $a_i \in (c_j - d, c_j]$  do
14:        draw sample  $\pi_{ik}[s] \sim \text{Gamma}(\mathcal{M}(\lambda_{ik}^\pi))$ 
15:        draw sample  $\epsilon_i[s] \sim \text{Bernoulli}(\mathcal{S}(\lambda_i^\epsilon))$ 
16:        set  $p_{ik}^\pi[s] = \log p(\pi_{ik}[s] | \alpha_0, \beta_0)$  ▷ see Eqn. 3
17:        set  $q_{ik}^\pi[s] = \log q(\pi_{ik}[s] | \lambda_{ik}^\pi)$  ▷ see Eqn. 6
18:        set  $g_{ik}^\pi[s] = \nabla_{\lambda_{ik}^\pi} \log q(\pi_{ik}[s] | \lambda_{ik}^\pi)$  ▷ see Eqn. 9
19:      for each event  $i$  do
20:        set  $p_i^\epsilon[s] = \log p(\epsilon_i[s] | \eta)$  ▷ see Eqn. 4
21:        set  $q_i^\epsilon[s] = \log q(\epsilon_i | \lambda_i^\epsilon)$  ▷ see Eqn. 7
22:        set  $g_i^\epsilon[s] = \nabla_{\lambda_i^\epsilon} \log q(\epsilon_i[s] | \lambda_i^\epsilon)$  ▷ see Eqn. 10
23:      for  $j = 1, \dots, B$  (batch size, or # of document samples) do
24:        for  $k = 1, \dots, K$  do
25:          Sample a document  $j$  that is sent on date  $c_j$  and has topics  $\theta_j$ 
26:          set  $\phi_{jk}[s] = \phi_{0k}[s] + \sum_{i: a_i \in (c_j - d, c_j]} \epsilon_i[s] \pi_{ik}[s] f(a_i, c_j)$ 
27:          set  $p_k^\theta[s] = \log p(\theta_j | \phi_{jk}[s], \beta_c)$  ▷ see Eqn. 1
28:          for each event  $i$  do
29:            update  $\hat{\lambda}_{ik}^\pi += \frac{1}{SB} g_{ik}^\pi[s] (p_{ik}^\pi[s] + J p_k^\theta[s] - q_{ik}^\pi[s])$ 
30:            update  $\hat{\lambda}_k^\phi += \frac{1}{SB} g_{ik}^\phi[s] (p_k^\phi[s] + J p_k^\theta[s] - q_k^\phi[s])$ 
31:          for each event  $i$  do
32:            update  $\hat{\lambda}_i^\epsilon += \frac{1}{SB} g_i^\epsilon[s] (p_i^\epsilon[s] + J \sum_k p_k^\theta[s] - q_i^\epsilon[s])$ 
33:        set  $\rho = (t + \tau)^\kappa$ 
34:        update event content for each event  $i$  and topic  $k$ :  $\lambda_{ik}^\pi = \lambda_{ik}^\pi + \rho \hat{\lambda}_{ik}^\pi$ 
35:        update general entity topics for each topic  $k$ :  $\lambda_k^\phi = \lambda_k^\phi + \rho \hat{\lambda}_k^\phi$ 
36:        update event occurrences for each event  $i$ :  $\lambda_i^\epsilon = \lambda_i^\epsilon + \rho \hat{\lambda}_i^\epsilon$ 
37:      until change in validation likelihood  $< \delta$ 

```

Algorithm 3 Black Box variational inference for Cables Model v2

```

1: Input: document topics  $\theta$ 
2: Initialize  $\lambda^\pi$ ,  $\lambda^\phi$ , and  $\lambda^\epsilon$  to respective priors
3: Initialize iteration count  $t = 0$ 
4: repeat
5:    $t += 1$ 
6:   initialize  $\hat{\lambda}^\pi$ ,  $\hat{\lambda}^\phi$ , and  $\hat{\lambda}^\epsilon$  to 0 matrices
7:   for  $s = 1, \dots, S$  do
8:     for  $k = 1, \dots, K$  do
9:       draw sample  $\phi_{0k}[s] \sim \text{Gamma}(\mathcal{M}(\lambda_k^\phi))$ 
10:      set  $p_k^\phi[s] = \log p(\phi_{0k}[s] | \alpha, \beta)$  ▷ see Eqn. 13
11:      set  $q_k^\phi[s] = \log q(\phi_{0k} | \lambda_k^\phi)$  ▷ see Eqn. 16
12:      set  $g_k^\phi[s] = \nabla_{\lambda_k^\phi} \log q(\phi_{0k}[s] | \lambda_k^\phi)$  ▷ see Eqn. 19
13:      for each event  $i$  on date  $a_i \in (c_j - d, c_j]$  do
14:        draw sample  $\pi_{ik}[s] \sim \text{Gamma}(\mathcal{M}(\lambda_{ik}^\pi))$ 
15:        draw sample  $\epsilon_i[s] \sim \text{Bernoulli}(\mathcal{S}(\lambda_i^\epsilon))$ 
16:        set  $p_{ik}^\pi[s] = \log p(\pi_{ik}[s] | \alpha_0, \beta_0)$  ▷ see Eqn. 14
17:        set  $q_{ik}^\pi[s] = \log q(\pi_{ik}[s] | \lambda_{ik}^\pi)$  ▷ see Eqn. 17
18:        set  $g_{ik}^\pi[s] = \nabla_{\lambda_{ik}^\pi} \log q(\pi_{ik}[s] | \lambda_{ik}^\pi)$  ▷ see Eqn. 20
19:      for each event  $i$  do
20:        set  $p_i^\epsilon[s] = \log p(\epsilon_i[s] | \eta)$  ▷ see Eqn. 15
21:        set  $q_i^\epsilon[s] = \log q(\epsilon_i | \lambda_i^\epsilon)$  ▷ see Eqn. 18
22:        set  $g_i^\epsilon[s] = \nabla_{\lambda_i^\epsilon} \log q(\epsilon_i[s] | \lambda_i^\epsilon)$  ▷ see Eqn. 21
23:      for  $j = 1, \dots, B$  (batch size, or # of document samples) do
24:        for  $k = 1, \dots, K$  do
25:          Sample a document  $j$  that is sent on date  $c_j$  and has topics  $\theta_j$ 
26:          set  $\phi_{jk}[s] = \phi_{0k}[s] + \sum_{i: a_i \in (c_j - d, c_j]} \epsilon_i[s] \pi_{ik}[s] f(a_i, c_j)$ 
27:          set  $p_k^\theta[s] = \log p(\theta_j | \phi_{jk}[s], \beta_c)$  ▷ see Eqn. 12
28:          for each event  $i$  do
29:            update  $\hat{\lambda}_{ik}^\pi += \frac{1}{SB} g_{ik}^\pi[s] (p_{ik}^\pi[s] + J p_k^\theta[s] - q_{ik}^\pi[s])$ 
30:            update  $\hat{\lambda}_k^\phi += \frac{1}{SB} g_k^\phi[s] (p_k^\phi[s] + J p_k^\theta[s] - q_k^\phi[s])$ 
31:          for each event  $i$  do
32:            update  $\hat{\lambda}_i^\epsilon += \frac{1}{SB} g_i^\epsilon[s] (p_i^\epsilon[s] + J \sum_k p_k^\theta[s] - q_i^\epsilon[s])$ 
33:        set  $\rho = (t + \tau)^\kappa$ 
34:        update event content for each event  $i$  and topic  $k$ :  $\lambda_{ik}^\pi = \lambda_{ik}^\pi + \rho \hat{\lambda}_{ik}^\pi$ 
35:        update general entity topics for each topic  $k$ :  $\lambda_k^\phi = \lambda_k^\phi + \rho \hat{\lambda}_k^\phi$ 
36:        update event occurrences for each event  $i$ :  $\lambda_i^\epsilon = \lambda_i^\epsilon + \rho \hat{\lambda}_i^\epsilon$ 
37: until change in validation likelihood  $< \delta$ 

```
