Modeling Cables Events

Allison J.B. Chaney

August 5, 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

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

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

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: $\mathcal{L}(q) = \text{E}_{q(\pi,\phi)}[\log p(\theta, \pi, \phi) - \log q(\pi, \phi)]$
- 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

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

and

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

We can also write these in their log forms:

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

and

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

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

$$\nabla_{\lambda^{\pi}_{ik}} \mathcal{L} = \mathrm{E}_q \left[\nabla_{\lambda^{\pi}_{ik}} \log q(\pi_{ik} \mid \lambda^{\pi}_{ik}) \left(\log p^{\pi}_{ik}(\theta, \pi, \phi) - \log q(\pi_{ik} \mid \lambda^{\pi}_{ik}) \right) \right]$$

and

$$\nabla_{\lambda^{\phi}_{k}} \mathcal{L} = \mathrm{E}_q \left[\nabla_{\lambda^{\phi}_{k}} \log q(\phi_{0k} \mid \lambda^{\phi}_{k}) \left(\log p^{\phi}_{k}(\theta, \pi, \phi) - \log q(\phi_{0k} \mid \lambda^{\phi}_{k}) \right) \right]$$

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

Input: document topics θ
Initialize λ^{π} and λ^{ϕ} randomly
Initialize $t_i^{\pi} = 0$ for all days i and $t^{\phi} = 0$
repeat
  Sample a document j that is sent on date c_j and has topics θ_j
  t^{ϕ} += 1
  for $s = 1, \ldots, S$ **do**
    for $k = 1, \ldots, K$ **do**
      draw sample $\phi_{0k}[s] \sim \text{Gamma}(\lambda_k^{\phi})$
      for each event i on date $a_i \in (c_j - d, c_j]$ **do**
        draw sample $\pi_{ik}[s] \sim \text{Gamma}(\lambda_{ik}^{\pi})$
        t_i^{π} += 1
      set $\phi_{jk}[s] = \phi_{0k}[s] + \sum_{i:a_i \in (c_j - d, c_j]} \pi_{ik}[s] f(a_i, c_j)$
      set $p_k^{\theta}[s] = \log p(\theta_j \mid \phi_{jk}[s], \beta_c) = \beta_c \phi_{jk}[s] \log \beta_c - \log \Gamma(\beta_c \phi_{jk}[s]) + (\beta_c \phi_{jk}[s] - 1) \log \theta_k - \beta_c \theta_k$
      set $p_k^{\phi}[s] = \log p(\phi_{0k} \mid \alpha, \beta) = \alpha \log \beta - \log \Gamma(\alpha) + (\alpha - 1) \log \phi_{0k}[s] - \beta \phi_{0k}[s]$
      set (for each relevant i) $p_{ik}^{\pi}[s] = \log p(\pi_{ik} \mid \alpha_0, \beta_0) = \alpha_0 \log \beta_0 - \log \Gamma(\alpha_0) + (\alpha_0 - 1) \log \pi_k[s] - \beta_0 \pi_{ik}[s]$
      set $q_k^{\phi}[s] = \log q(\phi_{0k} \mid \lambda_k^{\phi}) = \lambda_k^{\phi,\alpha} \log \lambda_k^{\phi,\beta} - \log \Gamma(\lambda_k^{\phi,\alpha}) + (\lambda_k^{\phi,\alpha} - 1) \log \phi_{0k}[s] - \lambda_k^{\phi,\beta} \phi_{0k}[s]$
      set (for each relevant i) $q_{ik}^{\pi}[s] = \log q(\pi_{ik} \mid \lambda_{ik}^{\pi}) = \lambda_{ik}^{\pi,\alpha} \log \lambda_{ik}^{\pi,\beta} - \log \Gamma(\lambda_{ik}^{\pi,\alpha}) + (\lambda_{ik}^{\pi,\alpha} - 1) \log \pi_{ik}[s] - \lambda_{ik}^{\pi,\beta} \pi_{ik}[s]$
      set $g_k^{\phi}[s] = \nabla_{\lambda_k^{\phi}} \log q(\phi_{0k} \mid \lambda_k^{\phi}) = \langle \log \lambda_k^{\phi,\beta} - \Psi(\lambda_k^{\phi,\alpha}) + \log \phi_{0k}[s], \lambda_k^{\phi,\alpha} / \lambda_k^{\phi,\beta} - \phi_{0k}[s] \rangle$
      set (for each relevant i) $g_{ik}^{\pi}[s] = \nabla_{\lambda_{ik}^{\pi}} \log q(\pi_{ik} \mid \lambda_{ik}^{\pi}) = \langle \log \lambda_{ik}^{\pi,\beta} - \Psi(\lambda_{ik}^{\pi,\alpha}) + \log \pi_{ik}[s], \lambda_{ik}^{\pi,\alpha} / \lambda_{ik}^{\pi,\beta} - \pi_{ik}[s] \rangle$
  for each event i on date $a_i \in (c_j - d, c_j]$ **do**
    set $\hat{\lambda}_{ik}^{\pi} = \frac{1}{S} \sum_s g_{ik}^{\pi}[s] (p_{ik}^{\pi}[s] + J p_k^{\theta}[s] - q_{ik}^{\pi}[s])$
    set $\rho_i^{\pi} = (t_i^{\pi} + \tau)^{\kappa}$
    update $\lambda_{ik}^{\pi} = \lambda_{ik}^{\pi} + \rho_i^{\pi} \hat{\lambda}_{ik}^{\pi}$
  set $\hat{\lambda}_k^{\phi} = \frac{1}{S} \sum_s g_{ik}^{\phi}[s] (p_k^{\phi}[s] + J p_k^{\theta}[s] - q_k^{\phi}[s])$
  set $\rho^{\phi} = (t^{\phi} + \tau)^{\kappa}$
  update $\lambda_k^{\phi} = \lambda_k^{\phi} + \rho^{\phi} \hat{\lambda}_k^{\phi}$
until change in validation likelihood $< \delta$

2.2 Inference

- Minimizing the KL divergence is equivalent to maximizing the ELBO: $\mathcal{L}(q) = \text{E}_{q(\pi,\phi,\epsilon)}[\log p(\theta, \pi, \phi, \epsilon) - \log q(\pi, \phi, \epsilon)]$
- 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

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

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

and

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

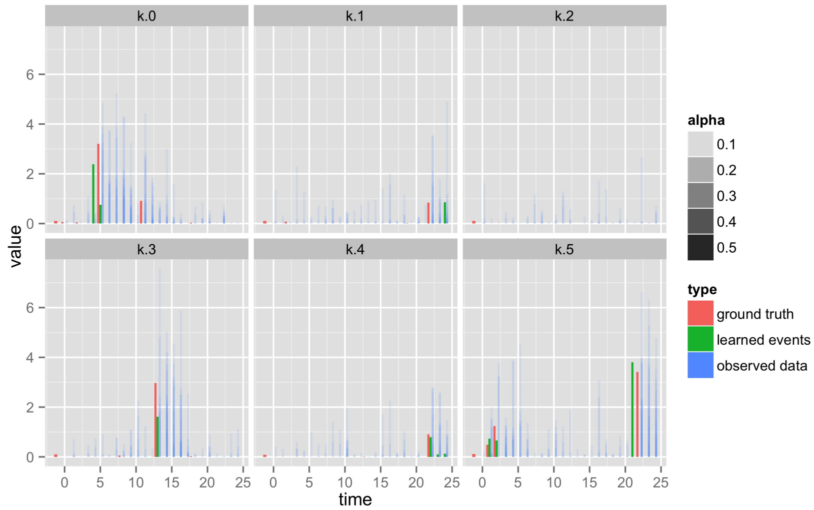

Figure 1: Identifying simulated events.

We can also write these in their log forms:

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

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

and

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

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

$$\nabla_{\lambda^{\pi}_{ik}} \mathcal{L} = \mathrm{E}_q \left[\nabla_{\lambda^{\pi}_{ik}} \log q(\pi_{ik} \mid \lambda^{\pi}_{ik}) \left(\log p^{\pi}_{ik}(\theta, \pi, \phi, \epsilon) - \log q(\pi_{ik} \mid \lambda^{\pi}_{ik}) \right) \right],$$

$$\nabla_{\lambda^{\phi}_{k}} \mathcal{L} = \mathrm{E}_q \left[\nabla_{\lambda^{\phi}_{k}} \log q(\phi_{0k} \mid \lambda^{\phi}_{k}) \left(\log p^{\phi}_{k}(\theta, \pi, \phi, \epsilon) - \log q(\phi_{0k} \mid \lambda^{\phi}_{k}) \right) \right],$$

and

$$\nabla_{\lambda^{\epsilon}_{i}} \mathcal{L} = \mathrm{E}_q \left[\nabla_{\lambda^{\epsilon}_{i}} \log q(\epsilon_i \mid \lambda^{\epsilon}_{i}) \left(\log p^{\epsilon}_{i}(\theta, \pi, \phi, \epsilon) - \log q(\epsilon_i \mid \lambda^{\epsilon}_{i}) \right) \right].$$

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.

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

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

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

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

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

Specific equations used in algorithm, fully expanded Prior probabilities.

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

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

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

$$\log p(\epsilon_i[s] \mid \eta) = \epsilon_i[s] \log \eta + (1 - \epsilon_i[s]) \log(1 - \eta) \tag{4}$$

Probabilities of the latent variables given the free variational parameters.

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

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

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

Gradients of variational distributions.

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

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

$$\nabla_{\lambda_{ik}^{\epsilon}} \log q(\epsilon_i \mid \lambda_i^{\epsilon}) = \epsilon_i \frac{\mathcal{S}'(\lambda_i^{\epsilon})}{\mathcal{S}(\lambda_i^{\epsilon})} + (1 - \epsilon_i) \frac{\mathcal{S}'(\lambda_i^{\epsilon})}{1 - \mathcal{S}(\lambda_i^{\epsilon})} \tag{10}$$

$$\tag{11}$$

Algorithm 2 Black Box variational inference for Cables Model v2

Input: document topics θ
Initialize λ^{π}, λ^{ϕ}, and λ^{ϵ} to respective priors
Initialize iteration count $t = 0$
repeat
  $t \mathrel{+}= 1$
  initialize $\hat{\lambda}^{\pi}$, $\hat{\lambda}^{\phi}$, and $\hat{\lambda}^{\epsilon}$ to 0 matrices
  for $s = 1, \ldots, S$ **do**
    for $k = 1, \ldots, K$ **do**
      draw sample $\phi_{0k}[s] \sim \text{Gamma}(\mathcal{M}(\lambda_k^{\phi}))$
      set $p_k^{\phi}[s] = \log p(\phi_{0k}[s] \mid \alpha, \beta)$ ▷ see Eqn. 2
      set $q_k^{\phi}[s] = \log q(\phi_{0k} \mid \lambda_k^{\phi})$ ▷ see Eqn. 5
      set $g_k^{\phi}[s] = \nabla_{\lambda_k^{\phi}} \log q(\phi_{0k}[s] \mid \lambda_k^{\phi})$ ▷ see Eqn. 8
      for each event i on date $a_i \in (c_j - d, c_j]$ **do**
        draw sample $\pi_{ik}[s] \sim \text{Gamma}(\mathcal{M}(\lambda_{ik}^{\pi}))$
        draw sample $\epsilon_i[s] \sim \text{Bernoulli}(\mathcal{S}(\lambda_i^{\epsilon}))$
        set $p_{ik}^{\pi}[s] = \log p(\pi_{ik}[s] \mid \alpha_0, \beta_0)$ ▷ see Eqn. 3
        set $q_{ik}^{\pi}[s] = \log q(\pi_{ik}[s] \mid \lambda_{ik}^{\pi})$ ▷ see Eqn. 6
        set $g_{ik}^{\pi}[s] = \nabla_{\lambda_{ik}^{\pi}} \log q(\pi_{ik}[s] \mid \lambda_{ik}^{\pi})$ ▷ see Eqn. 9
    for each event i **do**
      set $p_i^{\epsilon}[s] = \log p(\epsilon_i[s] \mid \eta)$ ▷ see Eqn. 4
      set $q_i^{\epsilon}[s] = \log q(\epsilon_i \mid \lambda_i^{\epsilon})$ ▷ see Eqn. 7
      set $g_i^{\pi}[s] = \nabla_{\lambda_i^{\epsilon}} \log q(\epsilon_i[s] \mid \lambda_i^{\epsilon})$ ▷ see Eqn. 10
    for $j = 1, \ldots, B$ (batch size, or # of document samples) **do**
      for $k = 1, \ldots, K$ **do**
        Sample a document j that is sent on date c_j and has topics θ_j
        set $\phi_{jk}[s] = \phi_{0k}[s] + \sum_{i:a_i \in (c_j - d, c_j]} \epsilon_i[s]\pi_{ik}[s] f(a_i, c_j)$
        set $p_k^{\theta}[s] = \log p(\theta_j \mid \phi_{jk}[s], \beta_c)$ ▷ see Eqn. 1
        for each event i **do**
          update $\hat{\lambda}_{ik}^{\pi} \mathrel{+}= \frac{1}{SB} g_{ik}^{\pi}[s](p_{ik}^{\pi}[s] + Jp_k^{\theta}[s] - q_{ik}^{\pi}[s])$
        update $\hat{\lambda}_k^{\phi} \mathrel{+}= \frac{1}{SB} g_{ik}^{\phi}[s](p_k^{\phi}[s] + Jp_k^{\theta}[s] - q_k^{\phi}[s])$
      for each event i **do**
        update $\hat{\lambda}_i^{\epsilon} \mathrel{+}= \frac{1}{SB} g_i^{\epsilon}[s](p_i^{\epsilon}[s] + J \sum_k p_k^{\theta}[s] - q_i^{\epsilon}[s])$
  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: $\lambda_k^{\epsilon} = \lambda_i^{\epsilon} + \rho\hat{\lambda}_i^{\epsilon}$
until change in validation likelihood $< \delta$