

応用計量分析2 (第9, 10, 11回)

トピックモデルと変分ベイズ法

担当教員: 梶野 洸 (かじの ひろし)

概要

1. トピックモデル（特に、Latent Dirichlet Allocation (LDA)）
2. 変分ベイズ法による推論
3. 実装

トピックモデルでやりたいこと

文書集合が与えられた時に、トピックに応じて文書分類をしたい。

- 文書1= 「台風24号が関東に接近します。最大風速は...」
- 文書2= 「台風24号の影響で野球が中止に。」
- 文書3= 「テニス選手の〇〇が全英オープンで初優勝しました。」
- 文書4= 「〇〇教授がノーベル生理学・医学賞受賞。」

トピックモデルの重要な仮定 (1/3)

トピックが異なると使われる単語が異なる。

- 例1. 「テニス」という単語は、スポーツ記事以外ではあまり出現しない。
- 例2. 「台風」という単語は、気象の記事以外ではあまり出現しない。
- 例3. どのトピックでも出現する単語もある（「が」とか）

トピックモデルの重要な仮定 (2/3)

ひとつの文書には複数のトピックが含まれる

- 文書1 = 天候100%
- 文書2 = 天候50%、スポーツ50%
- 文書3 = スポーツ100%
- 文書4 = 科学100%

トピックモデルの重要な仮定 (3/3)

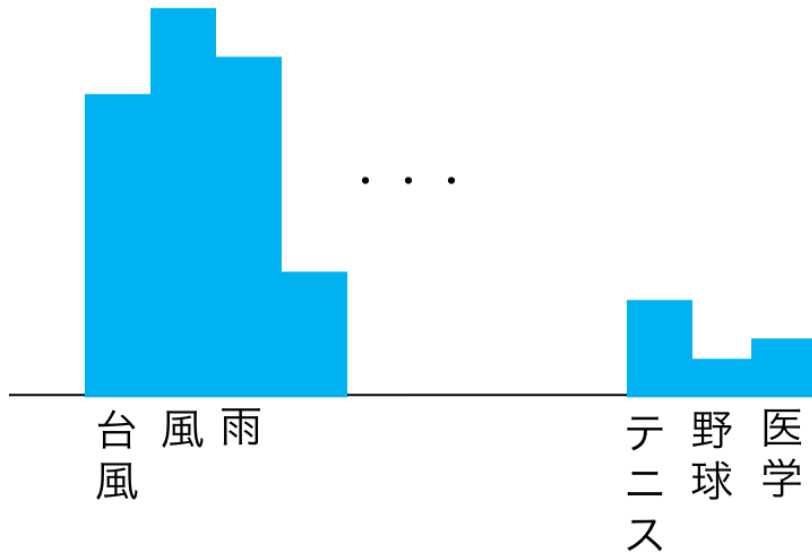
単語の順番は無視する

- 単語の順序まで考えると難しすぎる
- トピックを知る上では順番はそこまで重要じゃない

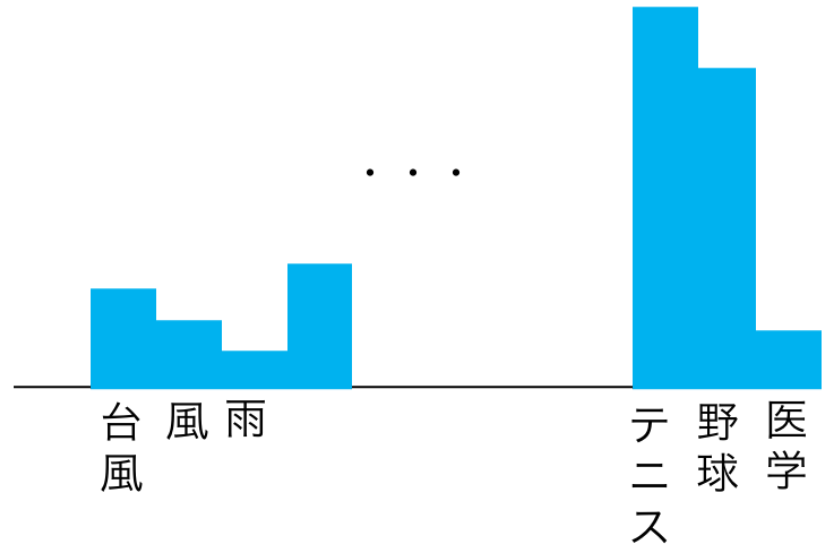
トピックモデルのなんとなくの説明

各トピックごとに異なる単語分布を持っている

トピック1の単語分布



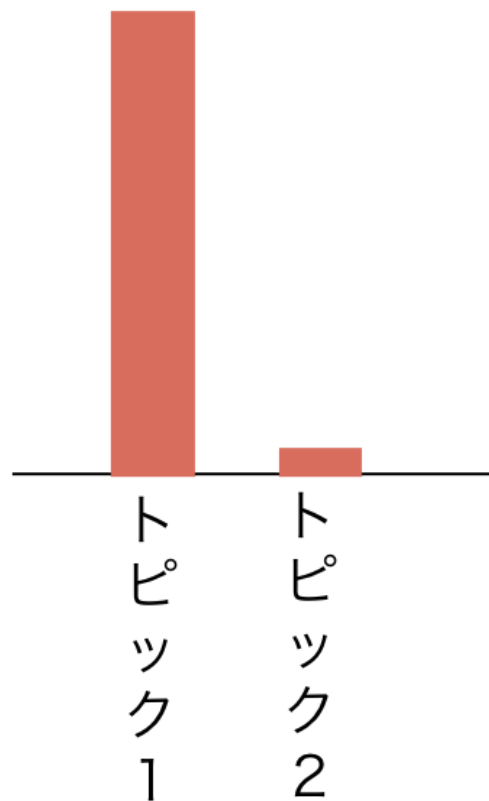
トピック2の単語分布



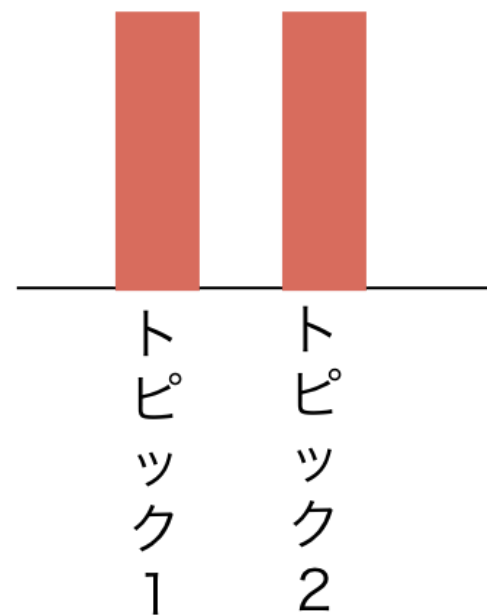
トピックモデルのなんとなくの説明

各文書はトピック分布で特徴付けられる

文書1のトピック分布

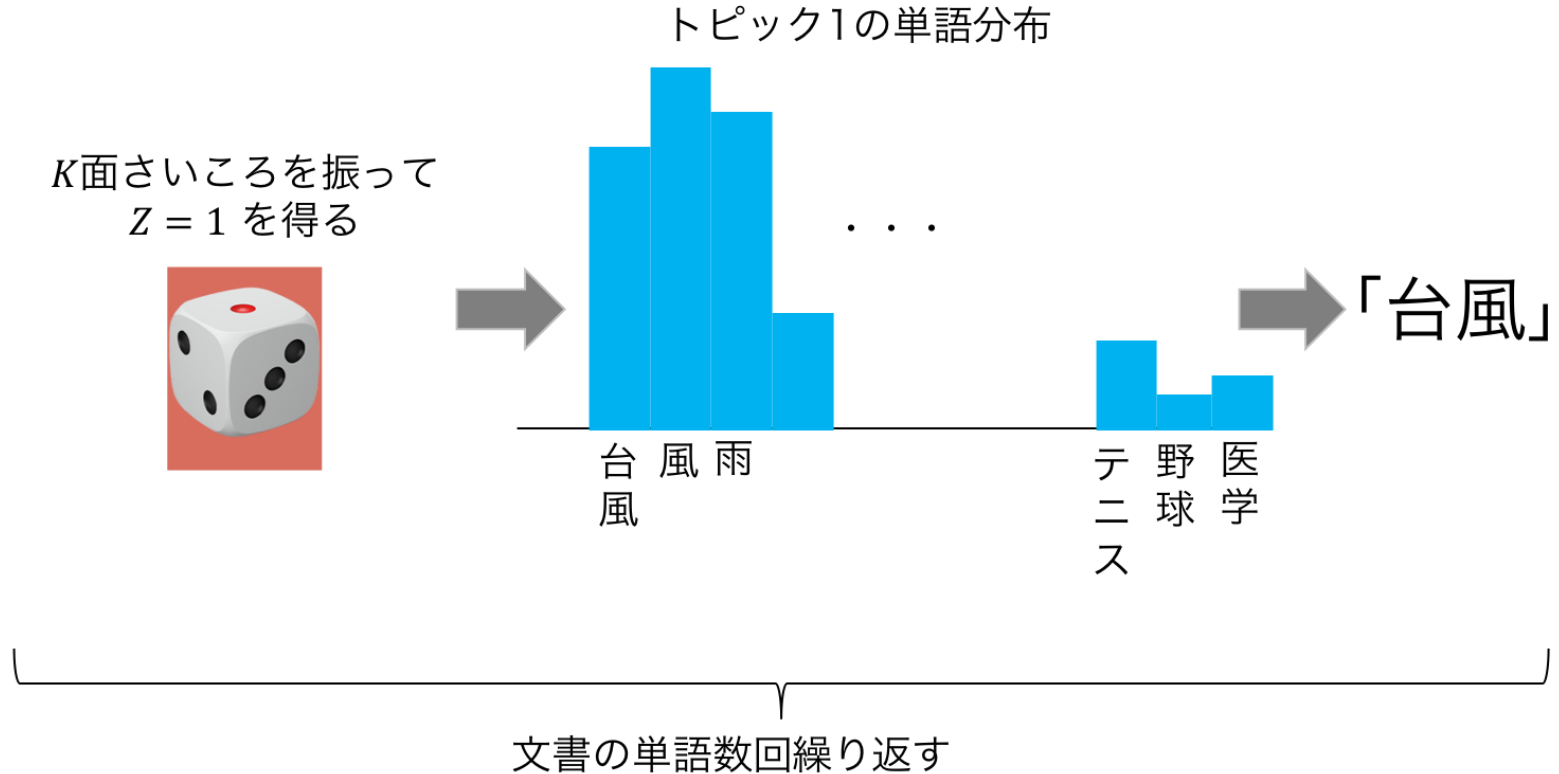


文書2のトピック分布



トピックモデルのなんとなくの説明

トピック分布からトピックを決め、そのトピックの単語分布から単語をサンプリングすることを繰り返して文書を作る。



数学的な表現への準備

データの表現

- 文書集合 $w = \{w_1, w_2, \dots, w_M\}$
- 文書 $w_m = [w_{m,1}, \dots, w_{m,N_m}] (m = 1, \dots, M)$
- 単語 $w_{m,n} \in \{1, \dots, V\}$

以下用いる記法

添字が省略されている際には、その添字の範囲を全て含むベクトル・行列を表すものとする

- $\beta_{k,v}$ は $k = 1, \dots, K, v = 1, \dots, V$ の範囲で定義されている
 - $\beta_k = [\beta_{k,1} \quad \beta_{k,2} \quad \dots \quad \beta_{k,V}]$
 - $\beta = \begin{bmatrix} \beta_{1,1} & \beta_{1,2} & \dots & \beta_{1,V} \\ \beta_{2,1} & \beta_{2,2} & \dots & \beta_{2,V} \\ & & \vdots & \\ \beta_{K,1} & \beta_{K,2} & \dots & \beta_{K,V} \end{bmatrix}$ など

数学的な表現への準備（例）

- 単語集合 $\{1, \dots, 6\} = \{\text{a, is, dog, cat, walking, sleeping}\}$
- 文書1 $w_1 = [\text{a, dog, is, walking}] = [1, 3, 2, 5]$
- 文書2 $w_2 = [\text{a, cat, is, sleeping}] = [1, 4, 2, 6]$
- 文書集合 $w = \{w_1, w_2\}$

モデル (1/3)

- 文書 m のトピック分布のパラメタ $\theta_m = [\theta_{m,1} \quad \dots \quad \theta_{m,K}] \in \Delta^{K-1}$
 - ディリクレ分布に従う
 - $\theta_{m,k} \geq 0, \sum_{k=1}^K \theta_{m,k} = 1$ ($m = 1, \dots, M$) が必ず成り立つ

$$p(\theta_m | \alpha) = \text{Dir}(\theta_m; \alpha) = \frac{\Gamma(\sum_{k=1}^K \alpha_k)}{\prod_{k=1}^K \Gamma(\alpha_k)} \prod_{k=1}^K \theta_{m,k}^{\alpha_k - 1} \quad (\alpha_k > 0)$$

モデル (2/3)

- 文書 m の n 個目の単語に紐づくトピック $z_{m,n} \in \{1, \dots, K\}$
 - パラメタ θ_m のCategorical distributionに従う

$$p(z_{m,n} = k \mid \theta_m) = \theta_{m,k}$$

$$\text{where } \sum_{k=1}^K \theta_{m,k} = 1 \quad (m = 1, \dots, M)$$

モデル (3/3)

- 単語数 = V , トピック数 = K .
- 文書 m の n 個目の単語 $w_{m,n} \in \{1, \dots, V\}$
 - その単語に紐づくトピック $z_{m,n} \in \{1, \dots, K\}$ で定まる、パラメタ $\beta_{z_{m,n}}$ の Categorical distribution に従う

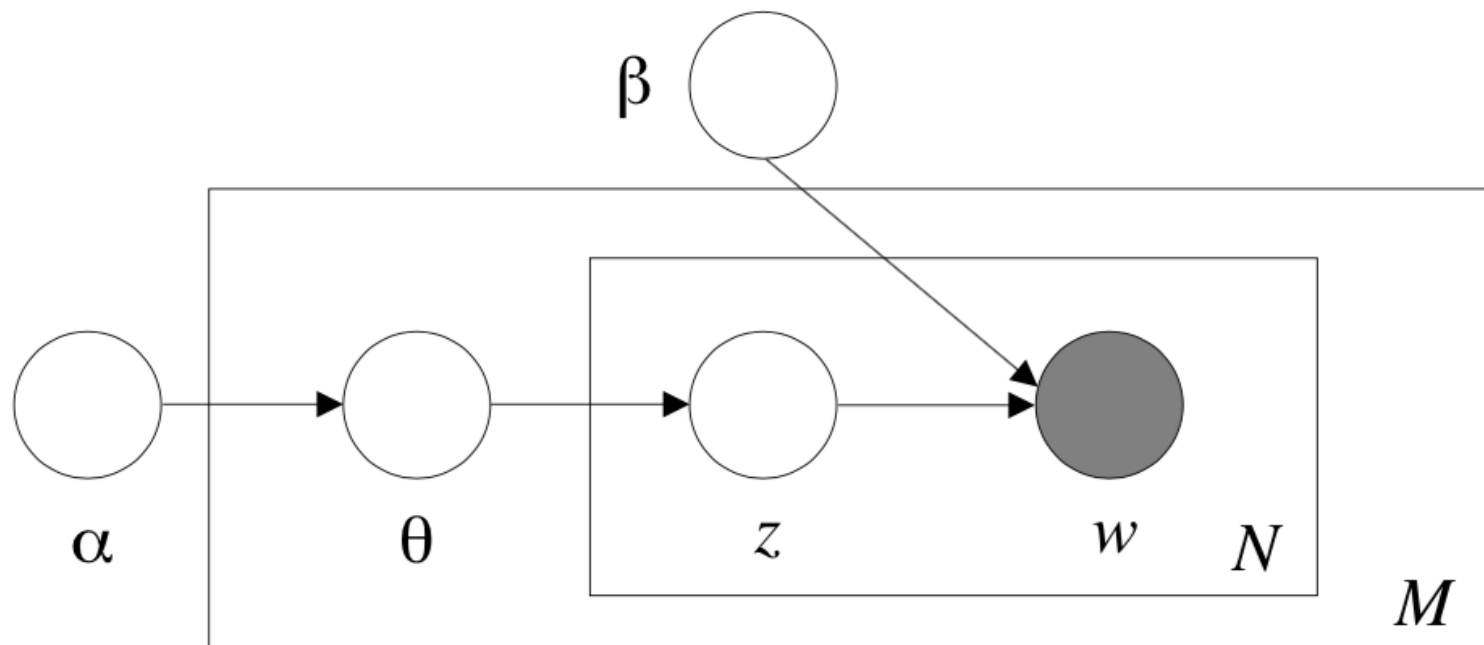
$$p(w_{m,n} = v \mid z_{m,n} = k, \beta) = \beta_{k,v}$$

$$\text{where } \sum_{v=1}^V \beta_{k,v} = 1 \quad (k = 1, \dots, K)$$

生成モデル的な説明

1. トピックごとの単語分布 $\beta_{k,v} = p(w_{m,n} = v \mid z_{m,n} = k)$
($k = 1, \dots, K, v = 1, \dots, V$) を決める
 - トピックが k だった時に単語 v が出る確率
 - V 次元の多項分布が K 個ある
2. 文書 m に対し、トピック分布のパラメタ $\theta_m \sim \text{Dir}(\alpha)$ を決める
 - $\theta_m \in \Delta^{K-1}$ は K 次元ベクトルで、多項分布のパラメタ
 - 文書 m でのトピックの出やすさを決める
3. 文書 m の単語 n のトピック $z_{m,n} \sim \text{Categorical}(\theta_m)$ を決める
4. 文書 m の単語 n の単語 $w_{m,n} \sim \text{Categorical}(\beta_{z_{m,n}})$ を決める

グラフィカルモデル



学習

データ $w = \{w_1, \dots, w_M\}$ が与えられた時、

1. $\alpha, \beta_{k,v}$ ($k = 1, \dots, K, v = 1, \dots, V$) をデータから推定する (最尤推定)
2. α, β 固定の上で潜在変数 $\theta_m, z_{m,n}$ ($m = 1, \dots, M, n = 1, \dots, N_m$) を推定する (inference)

Inference

- (α, β) は固定
- データ w を与えた元での潜在変数 (θ, z) の事後分布を推定したい

$$p(\theta, z \mid w, \alpha, \beta)$$

事後分布は直接計算できない

$$p(\theta, z | w, \alpha, \beta) = \frac{p(\theta, z, w | \alpha, \beta)}{p(w | \alpha, \beta)}$$

分子は計算できる

$$p(\theta, z, w \mid \alpha, \beta) = \prod_{m=1}^M \left[p(\theta_m \mid \alpha) \prod_{n=1}^{N_m} [p(z_{m,n} \mid \theta_m) p(w_{m,n} \mid z_{m,n}, \beta)] \right]$$

分母は計算できない

$$\begin{aligned} p(w \mid \alpha, \beta) &= \prod_{m=1}^M \int \sum_z p(\theta_m, z_m, w_m \mid \alpha, \beta) d\theta \\ &= \prod_{m=1}^M \frac{\Gamma(K\alpha)}{\Gamma(\alpha)^K} \int \left(\prod_{k=1}^K \theta_k^{\alpha-1} \right) \left(\prod_{n=1}^{N_m} \sum_{k=1}^K \prod_{v=1}^V (\theta_{m,k} \beta_{k,v})^{w_{m,n}} \right) d\theta \end{aligned}$$

事後分布の変分近似

事後分布を以下の分布で近似する

$$q(\theta, z \mid \gamma, \phi) = \prod_{m=1}^M q(\theta_m, z_m \mid \gamma_m, \phi_m)$$

$$q(\theta_m, z_m \mid \gamma_m, \phi_m) = q(\theta_m \mid \gamma_m) \prod_{n=1}^{N_m} q(z_{m,n} \mid \phi_{m,n})$$

- $q(\theta_m \mid \gamma_m) = \text{Dir}(\theta_m; \gamma_m)$
- $q(z_{m,n} \mid \phi_{m,n}) = \text{Categorical}(\phi_{m,n})$

事後分布の変分近似

- 元のモデル: θ_m と z_m が従属
- 変分分布: θ_m と z_m が独立
 - 独立な場合は計算できることが多い

事後分布の変分近似

事後分布との距離が近くなるように γ と ϕ を学習する

$$(\gamma^*, \phi^*) = \arg \min_{\gamma, \phi} \text{KL}(q(\theta, z | \gamma, \phi) || p(\theta, z | w, \alpha, \beta))$$

ここで $q(\theta, z | \gamma, \phi) = \prod_{m=1}^M q(\theta_m, z_m | \gamma_m, \phi_m)$

事後分布の変分近似（文書 m ごとに）

事後分布との距離が近くなるように γ_m と ϕ_m を学習する

$$(\gamma_m^*, \phi_m^*) = \arg \min_{\gamma_m, \phi_m} \text{KL}(q(\theta_m, z_m | \gamma_m, \phi_m) || p(\theta_m, z_m | w_m, \alpha, \beta))$$

補題1

以下の等式が成り立つ:

$$\begin{aligned} \text{KL}(q(\theta_m, z_m | \gamma_m, \phi_m) || p(\theta_m, z_m | w_m, \alpha, \beta)) &= \log p(w_m | \alpha, \beta) \\ &- \mathbb{E}_{q_m} [\log p(w_m, z_m, \theta_m | \alpha, \beta)] + \mathbb{E}_{q_m} [\log q(\theta_m, z_m | \gamma_m, \phi_m)] \end{aligned}$$

左辺は事後分布 $p(\theta_m, z_m | w_m, \alpha, \beta)$ があって計算できないが、右辺は事後分布がないため計算できることが嬉しい

証明

$$\begin{aligned}\log p(w_m \mid \alpha, \beta) &= \log \int \sum_{z_m} p(w_m, \theta_m, z_m \mid \alpha, \beta) d\theta_m \\ &= \log \int \sum_{z_m} \frac{p(w_m, \theta_m, z_m \mid \alpha, \beta)}{q(\theta_m, z_m \mid \gamma_m, \phi_m)} q(\theta_m, z_m \mid \gamma_m, \phi_m) d\theta_m \\ &\geq \int \sum_{z_m} q(\theta_m, z_m \mid \gamma_m, \phi_m) \log \left(\frac{p(w_m, \theta_m, z_m \mid \alpha, \beta)}{q(\theta_m, z_m \mid \gamma_m, \phi_m)} \right) d\theta_m\end{aligned}$$

証明

また

$$\log p(w_m | \alpha, \beta) = \int \sum_{z_m} q(\theta_m, z_m | \gamma_m, \phi_m) \log p(w_m | \alpha, \beta) d\theta_m$$

なので、(左辺) - (右辺) は、

$$\begin{aligned} & \int \sum_{z_m} q(\theta_m, z_m | \gamma_m, \phi_m) \log \left(\frac{q(\theta_m, z_m | \gamma_m, \phi_m) p(w_m | \alpha, \beta)}{p(w_m, \theta_m, z_m | \alpha, \beta)} \right) d\theta_m \\ &= \int \sum_{z_m} q(\theta_m, z_m | \gamma_m, \phi_m) \log \left(\frac{q(\theta_m, z_m | \gamma_m, \phi_m)}{p(\theta_m, z_m | w_m, \alpha, \beta)} \right) d\theta_m \\ &= \text{KL}(q(\theta_m, z_m | \gamma_m, \phi_m) || p(\theta_m, z_m | w_m, \alpha, \beta)) \end{aligned}$$

事後分布の変分近似 (文書 m ごとに)

事後分布との距離が近くなるように γ_m と ϕ_m を学習する

$$\begin{aligned}(\gamma_m^*, \phi_m^*) &= \arg \min_{\gamma_m, \phi_m} \text{KL}(q(\theta_m, z_m | \gamma_m, \phi_m) || p(\theta_m, z_m | w_m, \alpha, \beta)) \\ &= \arg \max_{\gamma_m, \phi_m} (\mathbb{E}_{q_m} [\log p(w_m, z_m, \theta_m | \alpha, \beta)] - \mathbb{E}_{q_m} [\log q(z_m, \theta_m | \gamma_m, \phi_m)])\end{aligned}$$

目的関数を分解する

$$\begin{aligned} & \mathbb{E}_{q_m} [\log p(w_m, z_m, \theta_m \mid \alpha, \beta)] - \mathbb{E}_{q_m} [\log q(z_m, \theta_m \mid \gamma_m, \phi_m)] \\ = & \mathbb{E}_{q_m} [\log p(\theta_m \mid \alpha)] + \mathbb{E}_{q_m} [\log p(z_m \mid \theta_m)] + \mathbb{E}_{q_m} [\log p(w_m \mid z_m, \beta)] \\ & - \mathbb{E}_{q_m} [\log q(\theta_m \mid \gamma_m)] - \mathbb{E}_{q_m} [\log q(z_m \mid \phi_m)] \end{aligned}$$

それぞれの項を計算してみる

第一項

$$\begin{aligned}\mathbb{E}_{q_m} [\log p(\theta_m | \alpha)] &= \mathbb{E}_{q(\theta_m | \gamma_m)} \left[\log \Gamma \left(\sum_{k=1}^K \alpha_k \right) - \sum_{k=1}^K \log \Gamma(\alpha_k) \right. \\ &\quad \left. + \sum_{k=1}^K (\alpha_k - 1) \log \theta_{m,k} \right] \\ &= \log \Gamma \left(\sum_{k=1}^K \alpha_k \right) - \sum_{k=1}^K \log \Gamma(\alpha_k) \\ &\quad + \sum_{k=1}^K (\alpha_k - 1) \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right)\end{aligned}$$

- Ψ は digamma 関数 (gamma関数を一階微分したもの)
- $\mathbb{E}_{q(\theta_m | \gamma_m)} [\log \theta_{m,k}] = \Psi(\alpha_k) - \Psi \left(\sum_{k=1}^K \alpha_k \right)$ を利用

参考

- Dirichlet 分布は指数分布族に属する
 - 確率密度関数が $p(x | \eta) = h(x) \exp[\eta^\top T(x) - A(\eta)]$ とかけること
 - $T(x) = [\log \theta_1 \quad \dots \quad \log \theta_K]$
 - $A(\eta) = \sum_{k=1}^K \log \Gamma(\alpha_k) - \log \Gamma(\sum_{k=1}^K \alpha_k)$
- 指数分布族では以下が成り立つことを利用

$$\begin{aligned} & \mathbb{E}_{p(x|\eta)} [T(x)] \\ &= \int T(x) h(x) \exp(\eta^\top T(x)) \exp(-A(\eta)) dx \\ &= \int \left(\frac{\partial}{\partial \eta} h(x) \exp(\eta^\top T(x)) \right) \exp(-A(\eta)) dx \\ &= \int \frac{\partial}{\partial \eta} \left((h(x) \exp(\eta^\top T(x))) \exp(-A(\eta)) \right) \\ & \quad - (h(x) \exp(\eta^\top T(x))) \frac{\partial}{\partial \eta} \exp(-A(\eta)) dx \\ &= 0 - \int (h(x) \exp(\eta^\top T(x))) \frac{\partial}{\partial \eta} \exp(-A(\eta)) dx = \frac{\partial}{\partial \eta} A(\eta) \end{aligned}$$

それぞれの項を計算してみる

第二項

$$\begin{aligned} & \mathbb{E}_{q_m} [\log p(z_m | \theta_m)] \\ &= \mathbb{E}_{q_m} \left[\sum_{n=1}^{N_m} \log p(z_{m,n} | \theta_m) \right] \\ &= \mathbb{E}_{q(\theta_m | \gamma_m)} \left[\sum_{n=1}^{N_m} \sum_{k=1}^K q(z_{m,n} = k | \phi_{m,n}) \log p(z_{m,n} = k | \theta_m) \right] \\ &= \mathbb{E}_{q(\theta_m | \gamma_m)} \left[\sum_{n=1}^{N_m} \sum_{k=1}^K \phi_{m,n,k} \log \theta_{m,k} \right] \\ &= \sum_{n=1}^{N_m} \sum_{k=1}^K \phi_{m,n,k} \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \end{aligned}$$

それぞれの項を計算してみる

第三項

$$\begin{aligned} & \mathbb{E}_{q_m} [\log p(w_m \mid z_m, \beta)] \\ &= \sum_{n=1}^{N_m} \mathbb{E}_{q(z_{m,n} | \phi_{m,n})} [\log p(w_{m,n} \mid z_{m,n}, \beta)] \\ &= \sum_{n=1}^{N_m} \sum_{k=1}^K \phi_{m,n,k} \log p(w_{m,n} \mid z_{m,n} = k, \beta) \\ &= \sum_{n=1}^{N_m} \sum_{k=1}^K \sum_{v=1}^V \phi_{m,n,k} \mathbb{1}[w_{m,n} = v] \log \beta_{k,v} \end{aligned}$$

それぞれの項を計算してみる

第四項

$$\begin{aligned} & \mathbb{E}_{q_m} [\log q(\theta_m | \gamma_m)] \\ &= \mathbb{E}_{q(\theta_m | \gamma_m)} \left[\log \Gamma \left(\sum_{k=1}^K \gamma_{m,k} \right) - \sum_{k=1}^K \log \Gamma(\gamma_{m,k}) + \sum_{k=1}^K (\gamma_{m,k} - 1) \log \theta_{m,k} \right] \\ &= \log \Gamma \left(\sum_{k=1}^K \gamma_{m,k} \right) - \sum_{k=1}^K \log \Gamma(\gamma_{m,k}) + \sum_{k=1}^K (\gamma_{m,k} - 1) \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \end{aligned}$$

それぞれの項を計算してみる

第五項

$$\begin{aligned}\mathbb{E}_{q_m} [\log q(z_m | \phi_m)] &= \sum_{n=1}^{N_m} \mathbb{E}_{q(z_{m,n}|\phi_{m,n})} [\log q(z_{m,n} | \phi_{m,n})] \\ &= \sum_{n=1}^{N_m} \sum_{k=1}^K \phi_{m,n,k} \log \phi_{m,n,k}\end{aligned}$$

まとめると(文書mに関する)目的関数は...

$$\begin{aligned} & L_m(\gamma, \phi; \alpha, \beta) \\ = & \log \Gamma \left(\sum_{k=1}^K \alpha_k \right) - \sum_{k=1}^K \log \Gamma(\alpha_k) + \sum_{k=1}^K (\alpha_k - 1) \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \\ & + \sum_{n=1}^{N_m} \sum_{k=1}^K \phi_{m,n,k} \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \\ & + \sum_{n=1}^{N_m} \sum_{k=1}^K \sum_{v=1}^V \phi_{m,n,k} \mathbb{1}[w_{m,n} = v] \log \beta_{k,v} \\ & - \left(\log \Gamma \left(\sum_{k=1}^K \gamma_{m,k} \right) - \sum_{k=1}^K \log \Gamma(\gamma_{m,k}) + \sum_{k=1}^K (\gamma_{m,k} - 1) \left(\Psi(\gamma_k) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \right) \\ & - \sum_{n=1}^{N_m} \sum_{k=1}^K \phi_{m,n,k} \log \phi_{m,n,k} \end{aligned}$$

ϕ に関する最大化 (γ は固定)

各 m, n について $\sum_{k=1}^K \phi_{m,n,k} = 1$ のもとで最大化

ラグランジアンは

$$\begin{aligned} L_m(\phi) &= \sum_{n=1}^{N_m} \sum_{k=1}^K \phi_{m,n,k} \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \\ &+ \sum_{n=1}^{N_m} \sum_{k=1}^K \sum_{v=1}^V \phi_{m,n,k} \mathbb{1}[w_{m,n} = v] \log \beta_{k,v} \\ &- \sum_{n=1}^{N_m} \sum_{k=1}^K \phi_{m,n,k} \log \phi_{m,n,k} + \sum_{n=1}^{N_m} \lambda_n \left(\sum_{k=1}^K \phi_{m,n,k} - 1 \right) \end{aligned}$$

ϕ に関する最大化 (γ は固定)

$\phi_{m,n,k}$ で微分すると

$$\begin{aligned} & \frac{\partial}{\partial \phi_{m,n,k}} L_m(\phi) \\ &= \Psi(\gamma_{m,k}) - \Psi\left(\sum_{k=1}^K \gamma_{m,k}\right) + \sum_{v=1}^V \mathbb{1}[w_{m,n} = v] \log \beta_{k,v} - \log \phi_{m,n,k} - 1 + \lambda_n = 0 \end{aligned}$$

ϕ に関する最大化 (γ は固定)

整理すると

$$\phi_{m,n,k} \propto \exp \left[\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) + \sum_{v=1}^V \mathbb{1}[w_{m,n} = v] \log \beta_{k,v} \right]$$

λ_n は正規化に使われる

γ に関する最大化 (ϕ は固定)

$\gamma_{m,k}$ に関する項は

$$\begin{aligned} & L_m(\gamma) \\ &= \sum_{k=1}^K (\alpha_k - 1) \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \\ &+ \sum_{n=1}^{N_m} \sum_{k=1}^K \phi_{m,n,k} \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \\ &- \left(\log \Gamma \left(\sum_{k=1}^K \gamma_{m,k} \right) - \sum_{k=1}^K \log \Gamma(\gamma_{m,k}) + \sum_{k=1}^K (\gamma_{m,k} - 1) \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \right) \end{aligned}$$

γ に関する最大化 (ϕ は固定)

整理すると

$$\begin{aligned} & L_m(\gamma) \\ &= \sum_{k=1}^K \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k'=1}^K \gamma_{m,k'} \right) \right) \left(\alpha_k - \gamma_{m,k} + \sum_{n=1}^{N_m} \phi_{m,n,k} \right) \\ & \quad - \left(\log \Gamma \left(\sum_{k=1}^K \gamma_{m,k} \right) - \sum_{k=1}^K \log \Gamma(\gamma_{m,k}) \right) \end{aligned}$$

γ に関する最大化 (ϕ は固定)

$\gamma_{m,k}$ で微分すると

$$\begin{aligned}
& \frac{\partial}{\partial \gamma_{m,k}} L_m(\gamma) \\
&= -\Psi(\gamma_{m,k}) + \Psi\left(\sum_{k'=1}^K \gamma_{m,k'}\right) \\
&\quad + \Psi'(\gamma_{m,k}) \left(\alpha_k - \gamma_{m,k} + \sum_{n=1}^{N_m} \phi_{m,n,k}\right) \\
&\quad - \Psi'\left(\sum_{k'=1}^K \gamma_{m,k'}\right) \sum_{k=1}^K \left(\alpha_k - \gamma_{m,k} + \sum_{n=1}^{N_m} \phi_{m,n,k}\right) \\
&\quad - \Psi\left(\sum_{k'=1}^K \gamma_{m,k'}\right) + \Psi(\gamma_{m,k}) \\
&= \Psi'(\gamma_{m,k}) \left(\alpha_k - \gamma_{m,k} + \sum_{n=1}^{N_m} \phi_{m,n,k}\right) \\
&\quad - \Psi'\left(\sum_{k'=1}^K \gamma_{m,k'}\right) \sum_{k=1}^K \left(\alpha_k - \gamma_{m,k} + \sum_{n=1}^{N_m} \phi_{m,n,k}\right)
\end{aligned}$$

γ に関する最大化 (ϕ は固定)

Ψ' の中身は忘れて係数を見ると、以下の γ で勾配が0になる

$$\gamma_{m,k} = \alpha_k + \sum_{n=1}^{N_m} \phi_{m,n,k}$$

Inference アルゴリズム

1. $\phi_{m,n,k}, \gamma_{m,k}$ を初期化 ($m = 1, \dots, M, n = 1, \dots, N_m, k = 1, \dots, K$).
2. 以下を収束するまで繰り返す

A.
$$\phi_{m,n,k} \propto \exp \left[\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) + \sum_{v=1}^V \mathbb{1}[w_{m,n} = v] \log \beta_{k,v} \right]$$

B.
$$\gamma_{m,k} = \alpha_k + \sum_{n=1}^{N_m} \phi_{m,n,k}$$

$\alpha, \beta_{k,v}$ ($k = 1, \dots, K, v = 1, \dots, V$) をデータから推定

- 上の Inference アルゴリズムを使うと、対数尤度の最もタイトな下界が求まる
- 下の式の右辺を ϕ, γ について最大化していた
- 対数尤度の下界を α, β について最大化すれば尤度を大きくできそう？

$$\log p(w_m | \alpha, \beta) \geq \int \sum_{z_m} q(\theta_m, z_m | \gamma_m, \phi_m) \log \left(\frac{p(w_m, \theta_m, z_m | \alpha, \beta)}{q(\theta_m, z_m | \gamma_m, \phi_m)} \right) d\theta_m$$

β に関する最大化

下界の β に関する項 + ラグランジアンは、

$$L(\beta) = \sum_{m=1}^M \sum_{n=1}^{N_m} \sum_{k=1}^K \sum_{v=1}^V \phi_{m,n,k} \mathbb{1}[w_{m,n} = v] \log \beta_{k,v} + \sum_{k=1}^K \lambda_k \left(\sum_{v=1}^V \beta_{k,v} - 1 \right)$$

β に関する最大化

$\beta_{k,v}$ で微分すると

$$\frac{\partial}{\partial \beta_{k,v}} L(\beta) = \sum_{m=1}^M \sum_{n=1}^{N_m} \frac{1}{\beta_{k,v}} \phi_{m,n,k} \mathbb{1}[w_{m,n} = v] + \lambda_k$$

β に関する最大化

$\beta_{k,v}$ で微分すると

$$\beta_{k,v} \propto \sum_{m=1}^M \sum_{n=1}^{N_m} \phi_{m,n,k} \mathbb{1}[w_{m,n} = v]$$

α に関する最大化

下界の α に関する項は、

$$L(\alpha) = \sum_{m=1}^M \left[\log \Gamma \left(\sum_{k=1}^K \alpha_k \right) - \sum_{k=1}^K \log \Gamma(\alpha_k) + \sum_{k=1}^K (\alpha_k - 1) \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \right]$$

α に関する最大化

下界の α_k に関する勾配は、

$$\begin{aligned} & \frac{\partial}{\partial \alpha_k} L(\alpha) \\ &= M \left(\Psi \left(\sum_{k=1}^K \alpha_k \right) - \Psi(\alpha_k) \right) + \sum_{m=1}^M \left(\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) \right) \end{aligned}$$

α に関する最大化

勾配=0は解けないので、勾配法で最大化する

α, β の推定アルゴリズム

1. α_k ($k = 1, \dots, K$), $\beta_{k,v}$ ($k = 1, \dots, K, v = 1, \dots, V$) を初期化
2. $\phi_{m,n,k}, \gamma_{m,k}$ を初期化 ($m = 1, \dots, M, n = 1, \dots, N_m, k = 1, \dots, K$).
3. 以下を収束するまで繰り返す

A. 変分下界の更新 (E-step)

$$\phi_{m,n,k} \propto \exp \left[\Psi(\gamma_{m,k}) - \Psi \left(\sum_{k=1}^K \gamma_{m,k} \right) + \sum_{v=1}^V \mathbb{1}[w_{m,n} = v] \log \beta_{k,v} \right]$$

$$\gamma_{m,k} = \alpha_k + \sum_{n=1}^{N_m} \phi_{m,n,k}$$

B. β の更新 (M-step 1)

$$\beta_{k,v} \propto \sum_{m=1}^M \sum_{n=1}^{N_m} \phi_{m,n,k} \mathbb{1}[w_{m,n} = v]$$

C. α の更新 (M-step 2)

$$\alpha_k \leftarrow \alpha_k + \eta \frac{\partial}{\partial \alpha_k} L(\alpha)$$

を収束するまで繰り返す。