Topic Models(AO):Chapter1-4

@anemptyarchive

2019/07/13 - 2019/08/26

Contents

はじめに

このレジュメは、MLP シリーズの『トピックモデル』での学習を助けるためのものです。ぜひ本も併せてお読みください。

コンセプトは、「数式」と「プログラム処理」からトピックモデルを深く理解する、です。対象として、高校数学を説明されれば何となく思い出せるくらいのレベル (半年前の自分) を想定して解説しています。R については、入門書を 1 冊やったことあるくらいを想定しています。

このレジュメを通して、R 力を高めることは目的としていません。むしろ逆に、極力基本的な (処理をイメージしやすい) 関数を使って推定を行うことで、数式の意味をプログラムから理解していくことが目的です。`{tidyverse}` は基本です。

そのため、テキストクリーニングの処理は、品詞を絞り、小文字アルファベットの単語を取り除く以外のをことを行いません。従って、このレジュメで掲載している分析結果の例にも特に拘っておりません (ブラーックバターフライ)。(いえこれただのは言い訳です。5 章以降をやる頃にはちゃんとします…)

ちなみに、一連の推定に用いるテキストは、**ハロー！プロジェクト**のグループ `Juice=Juice` の楽曲の歌詞です！

それでは、

楽しんで書きました。

楽しんで読んでください。

R 元年の夏、アーーツイツイツイ！。

1 確率の基礎

1.1 確率

1.1.1 確率と確率分布

例えば、サイコロを 1 回振ると 1~6 の目 (事象の集合) からそれぞれ $\frac{1}{6}$ の確率で 1 つの目 (事象) が出る。コインの場合なら表を 1、裏を 0 として、コインを投げると 0 か 1(事象の集合) からそれぞれ $\frac{1}{2}$ の確率で 0 か 1 か (事象) が決まる。この様に、事象ごとの確率によってランダムに値をとるものを確率変数という。この例では、1~6 や 0・1 から起こり得る値を確率変数 X とする (この本では基本的に、確率変数 X ととある事象 x とを区別していない)。

このサイコロの出目という事象の集合、つまり確率変数 X が取り得る値の集合を $\mathcal{X} = \{1, 2, 3, 4, 5, 6\}$ のように表記する。コイン場合であれば $\mathcal{X} = \{0, 1\}$ である。事象の集合 $\mathcal{X}$ から事象 x 起きる (観測される) ことを $x \in \mathcal{X}$ とし、その確率を $p(X = x)$ あるいは $p(x)$ と書く。確率変数の値とその確率を表すものを確率分布と呼ぶ。(確率分布については 1.2 節にて)

サイコロのように 1 の次は 2 と連続していない値のことを離散値という (連続しているというのは 1 の次は 1.0000...のように区切れない値で、そのような値をとる場合は連続値という)。離散値の集合 $\mathcal{X}$ から 1 つの値をとる離散確率変数の場合、確率は

$$0 \leq p(x) \leq 1 \tag{1.1}$$

の値をとる。また、全ての事象の確率を足し合わせると

$$\sum_{x \in \mathcal{X}} p(x) = 1 \tag{1.2}$$

である。

この本では、主に文書中の単語の出現回数や出現確率を扱う。単語の数も離散値である。また、出現確率を ϕ を使って表す。語彙 v の出現確率を $p(v) = \phi_v$ のような形式で用いられる。また

$$\sum_{v=1}^{V} \phi_v = 1$$

である。これは、語彙 1 から V までの出現確率を全て足し合わせると 1 となることを示している。(そのため 1 章でも、確率全般に ϕ を使っている)

1.1.2 同時確率

事象 A と B が同時に起こる確率を同時確率と呼び、$p(A, B)$ と書く (A かつ B の確率 $p(A \cap B)$ のこと)。事象 A が起こった状態で事象 B が起こる確率を条件付き確率と呼び、$p(B|A)$ と書く。

この本では、例えば単語 w の出現確率が ϕ であるときに、単語 w が出現する確率を条件付き確率 $p(w|\phi)$ と表現する。

同時確率と条件付き確率には、乗法定理

$$p(A,B) = p(B|A)p(A) = p(A|B)p(B)$$

が成り立つ。また更にここから

$$p(B|A) = \frac{p(A,B)}{p(A)}$$

も成り立つ成り立つ。これらの関係は、ベイズの定理でも用いられる。

この本では、1 番目から N 番目の単語 $w_1, w_2, \cdots, w_N$ が同時に文書に出現することを $p(w_1, w_2, \cdots, w_N)$ と表現する。

ちなみに、主に w で表す単語は、文書中の単語を順番に並べた表現で、例えば 1 番目の単語と 5 番目の同じ単語だったとしても w_1, w_5 と別々に扱う。一方 v で表す語彙は、同じ単語をまとめて表現したもので、まとめる代わりに語彙 v の出現回数を N_v としている。(詳しくは、2 章以降にて)

1.1.3 周辺化

同時確率 $p(A,B)$ の 1 つの確率変数 B について全事象の和をとると $p(A)$ になる。

$$p(A) = \sum_B p(A,B) = \sum_B p(A|B)p(B)$$

この様な操作を周辺化と呼ぶ。周辺化することで、他の事象の影響を排除して考えることができる。

また、連続確率変数の場合は積分して周辺化する。

$$p(A) = \int p(A,B)dB = \int p(A|B)p(B)dB$$

積分によって変数を周辺化することを積分消去とも呼ぶ。

この本では、パラメータを積分消去することで推定すべきパラメータの数を減らす操作などで用いられている。

$$p(w|z,\beta) = \int p(w|z,\phi)p(\phi|\beta)d\phi$$

(3 章まで進めば出会えるので、今はふーんで大丈夫。) 単語の出現確率 ϕ はパラメータ β の影響を受けて決まる。その ϕ と文書 (あるいは単語) が持つトピック z の影響を受けて、単語 w が生成されると定義している。

ϕ を周辺化することで、w は β と z からなるとして推定を進めていく。

1.1.4 ベイズの定理

加法定理 (1.3) を用いて、ベイズの定理

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)} = \frac{p(A|B)p(B)}{\sum_B p(A,B)}$$

を導出できる。3 つ目の分母は、B を周辺化しており、つまり $\sum_B p(A,B) = p(A)$ のことである。また、連続値をとる場合は、$\int p(A,B)dB$ である。

ベイズの定理を用いると、B という条件の下での A の確率 $p(A|B)$ の情報があるとき、A の確率 $p(A)$ と B の確率 $p(B)$ を使って、A という条件下の B の確率 $p(A|B)$ を求めることができる。この様に、データと条件との関係が逆の確率を求められることから、ベイズの定理は逆確率とも呼ばれる。

この本では、ベイズの定理は (事後分布)=(尤度)$*$(事前分布)/(周辺尤度) の形式で用いられる。例えば、文書のデータがあり、ここから各単語がどのような確率で生成されるのかを知りたいとき

$$p(\phi|w,\beta) = \frac{p(w|\phi)p(\phi|\beta)}{p(w|\beta)}$$

の式で求められる。

実際に観測した単語の情報から求めた尤度 (尤もらしさ)$p(w|\phi)$ に対して、推定前に想定しておく単語の出現確率 (事前分布)$p(\phi|\beta)$ と、ϕ を周辺化した周辺尤度 $p(w|\beta)$ から、単語の出現確率 (事後分布) を推定する。(詳細は 3 章以降をお楽しみに)

1.1.5 独立

サイコロ A, B を振るとき、2 つのサイコロは互いに影響することなく出目が決まる。このような場合、2 つは独立であるという。独立であるとき

$$p(A,B) = p(A)p(B)$$

となる。また、これが成り立つとき A, B は独立であると言える。

C が与えられている下で A, B が独立しているとき、条件付き独立と言い

$$p(A,B|C) = p(A|C)p(B|C)$$

が成り立つ。

この本では、各単語は独立して出現するとしているので

$$p(w_1, w_2, \cdots, w_N) = \prod_{n=1}^{N} p(w_n)$$

と書き換えることができる。

1.1.6 計算例

ベイズの定理などについて、具体的な数値で確認できる。(参考書を参照)

1.1.7 期待値

・期待値

確率 $p(x)$ に従う x の関数 $f(x)$ の加重平均をとった値が期待値となる。

$$\mathbb{E}[f] = \sum_{x \in \mathcal{X}} p(x) f(x) \tag{1.5}$$

また、$f(x) = x$ であるときの期待値を平均と呼ぶ。

$$\mathbb{E}[x] = \sum_{x \in \mathcal{X}} p(x) x$$

・最頻値

最も確率が高くなる値のことを最頻値という。

$$\mathrm{mode}[x] = \operatorname*{argmax}_{x \in \mathcal{X}} p(x)$$

$\operatorname*{argmax}_x$ は、$p(x)$ を最大化 (max) する引数 (argument)x(の集合) を求めることである。

・分散

分散とは、偏差 (平均との差) の 2 乗の期待値のことで、分布の散らばり具合を表す。

$$\begin{aligned}
\mathrm{Var}[x] &= \mathbb{E}\Big[(x - \mathbb{E}[x])^2\Big] && \text{(i)} \\
&= \mathbb{E}[x^2] - \mathbb{E}\Big[\mathbb{E}[x]^2\Big] && \text{(ii)} \\
&= \mathbb{E}[x^2] - \mathbb{E}[x]^2
\end{aligned}$$

【途中式の途中式】

- (0)：変数 x とその期待値 $\mathbb{E}[x]$ との差を 2 乗して、期待値を求める。
- (i)：$\mathbb{E}[A - B] = \mathbb{E}[A] - \mathbb{E}[A]$ であるため、分割する。
- (ii)：$\mathbb{E}[x]$ は定数なので、$\mathbb{E}[\mathbb{E}[x]] = \mathbb{E}[x]$ である。

変数の「2 乗の期待値」から「期待値の 2 乗」を引いた値が分散となることが分かる。

・共分散

共分散とは、2 つの変数の偏差の積の期待値のことで、2 変数の相関を表す。

$$
\begin{aligned}
\mathrm{Cov}[x,y] &= \mathbb{E}\Big[(x-\mathbb{E}[x])(y-\mathbb{E}[y])\Big] && \text{(i)}\\
&= \mathbb{E}\Big[xy-\mathbb{E}[y]x-\mathbb{E}[x]y+\mathbb{E}[x]\mathbb{E}[y]\Big] && \text{(ii)}\\
&= \mathbb{E}[xy]-\mathbb{E}\Big[x\mathbb{E}[y]\Big]-\mathbb{E}\Big[y\mathbb{E}[x]\Big]+\mathbb{E}\Big[\mathbb{E}[x]\mathbb{E}[y]\Big] && \text{(iii)}\\
&= \mathbb{E}[xy]-\mathbb{E}[x]\mathbb{E}[y]-\mathbb{E}[x]\mathbb{E}[y]+\mathbb{E}[x]\mathbb{E}[y] && \text{(iv)}\\
&= \mathbb{E}[xy]-\mathbb{E}[x]\mathbb{E}[y]
\end{aligned}
$$

【途中式の途中式】

- (0)：各変数 x, y の偏差を掛け合わせて、期待値を求める。
- (i)：括弧を展開する。
- (ii)：$\mathbb{E}[A+B]=\mathbb{E}[A]+\mathbb{E}[B]$ であるため、分割する。
- (iii)：$\mathbb{E}[x], \mathbb{E}[y]$ は定数なので、括弧の外に出せる。

---

2 つの変数の「積の期待値」から「期待値の積」を引いた値が共分散となることが分かる。

1.1.8 カルバック・ライブラー・ダイバージェンス

カルバック・ライブラー・ダイバージェンス (KL ダイバージェンス・KL 情報量) とは、確率分布間の非類似度 (「どれくらい似ていないか」=「違いの程度」はつまりどれくらい似ているのかを表す) として用いる。相対エントロピーとも呼ばれる。

まずは、情報量という概念から確認していく。

・情報量

情報を定量的に扱うことで、足し引きや大小比較をできるようにする。そこで、事象 x の情報量 $I(x)$ を生起確率 $p(x)$ を基に

$$I(x) = -\log_2 p(x)$$

と定義する。$0 \leq p(x) \leq 1$ より $\log p(x)$ は負の値になるので、マイナスして正の値にしている。(ここでの log の底が 2 なのは、IT 分野において bit(0, 1) との相性からよく使われるためである。この本の内容では、

自然対数を用いる。)

これに従うと、確率が低いほど情報量が大きくなる。これは、分かり切った事柄の情報よりも不確実な事柄の情報の方が価値が高いことを表現できるため都合がよい。

例えば、$x = \frac{1}{10}, y = \frac{1}{4}$ としたときの情報量は

$$
\begin{aligned}
I(x) &= -\log_2 p(0.1) = 3.321928 \\
I(y) &= -\log_2 p(0.25) = 2
\end{aligned}
$$

となる。4 択の答えよりもより予測のしづらい 10 択の答えを教えてもらえる方が、情報の価値が高いことを表している。

また、2 つの事象 x, y に関する情報の価値を考えるとき、その確率は (独立していれば) 同時確率 $p(x, y) = p(x)p(y)$ として、積で求められる。この確率の対数をとった値を情報量とすることで、$I(x, y) = I(x) + I(y)$ として、和で求められる。これは、複数の情報を得たとき、情報量が加算的に増えていくことを表現できるため、感覚的な「情報量」の扱いと合致するので都合がよい。

これは

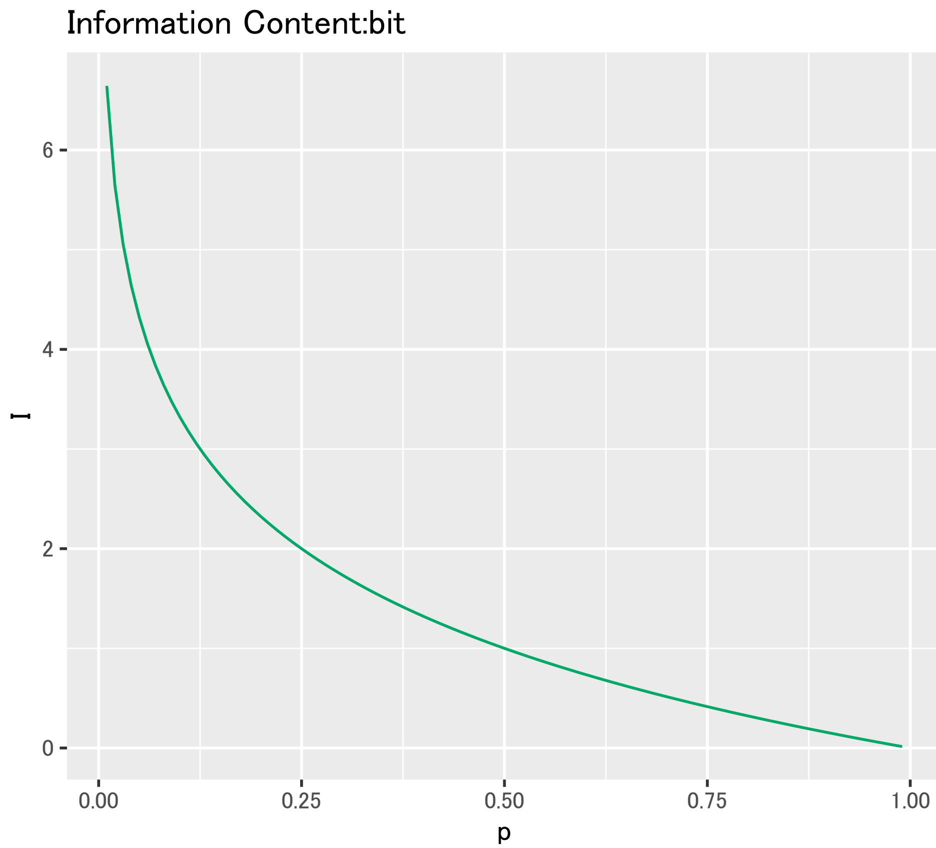
$$
\begin{aligned}
p(x, y) &= 0.1 * 0.25 = 0.025 \\
I(xy) &= -\log_2 p(0.025) = 5.321928 \\
I(x) + I(y) &= 3.321928 + 2 = 5.321928
\end{aligned}
$$

のようにも確認できる。これを定式化すると次のようになる。

$$
\begin{aligned}
I(xy) &= -\log_2 \Big(p(x)p(y) \Big) \\
&= -\log_2 p(x) + \Big(-\log_2 p(y) \Big) = I(x) + I(y)
\end{aligned}
$$

続いて、確率 $p(x)$ と情報量 $I(x)$ の関係をグラフで見る。

```r
# 利用パッケージ
library(tidyverse)

# 作図
tibble(p = seq(0.01, 0.99, 0.01),  # 確率
       I = - log2(p)) %>%          # 情報量
  ggplot(mapping = aes(x = p, y = I)) +     # データ
    geom_line(color = "#00A968") +          # 折れ線グラフ
    labs(title = "Information Content:bit") # タイトル
```



確率が低くなるほど情報量が大きくなるのを確認できる。

・平均情報量

各事象の生起確率からそれぞれの情報量が求まるということは、それらの加重平均をとることで期待値が求まる。それを平均情報量 (エントロピー)$H(x)$ という。

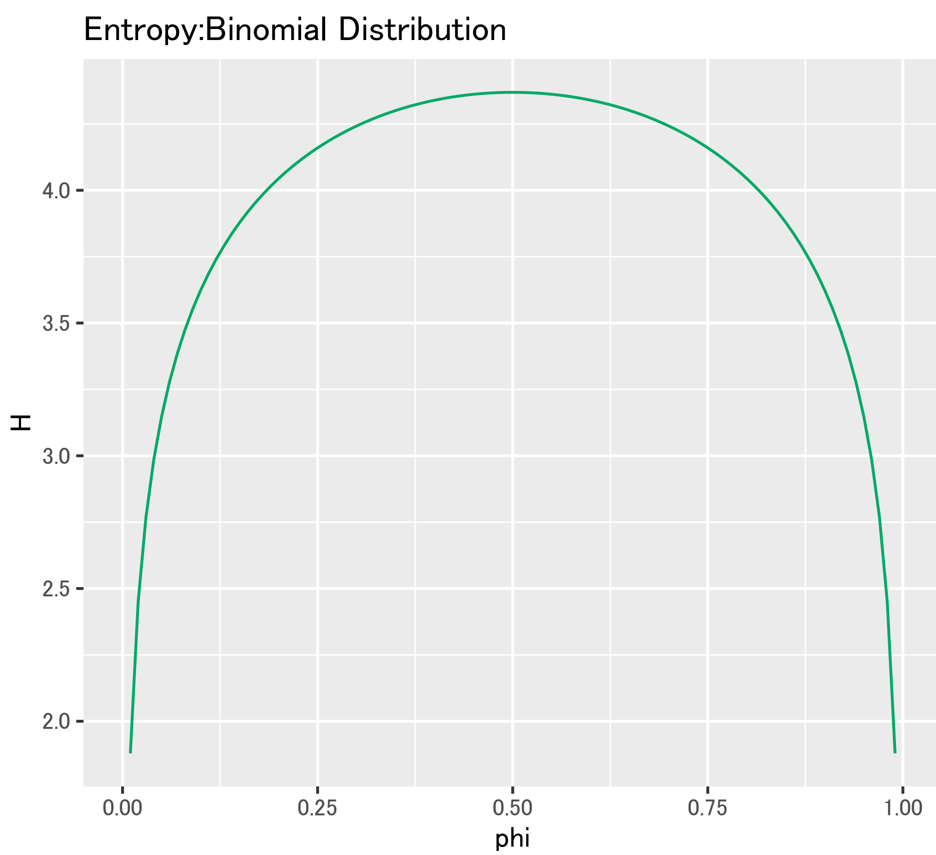
$$
\begin{aligned}
H(x) &= \sum_{x=1}^{n} \Big(-\log_2 p(x) \Big) p(x) \\
&= -\sum_{x=1}^{n} p(x) \log_2 p(x)
\end{aligned}
$$

このとき確率 $p(x)$ と情報量 $-\log_2 p(x)$ は、共に常にプラスの値をとるので、掛け合わせてもプラスの値になる。

$$
0 \leq p(x) \leq 1 \quad (x = 1, 2, \cdots, n) \ \Rightarrow -p(x) \log_2 p(x) \geq 0
$$

続いて、二項分布のパラメータ ϕ と平均情報量 $H(x)$ の関係をグラフで見る。

```r
# パラメータを指定
N <- 100   # 試行回数
x <- 0:N   # 表の回数
phi <- seq(0, 1, 0.01) # パラメータ（コインの歪み具合）

# 平均情報量を算出
H <- NULL
for(i in seq_along(phi)) {
  p <- dbinom(x, N, phi[i]) # 表の回数ごとの確率
  I <- - log2(p)            # 情報量
  H <- c(H, sum(p * I))     # 平均情報量
}

# 作図用のデータフレームを作成
entropy_df <- data.frame(phi = phi,
                         H = H)

# 描画
ggplot(data = entropy_df, mapping = aes(x = phi, y = H)) +  # データ
  geom_line(color = "#00A968") +                # 折れ線グラフ
  labs(title = "Entropy:Binomial Distribution") # タイトル
```



二項分布のパラメータ ϕ によって、平均情報量 H が変化している。ϕ はコインの歪み具合のことで、0 に近いほど裏が出やすく、1 に近いほど表が出やすくなる。グラフを見ると $\phi = 0.5$ のときに平均情報量 H が高くなっている。つまり、裏表の出やすさが等しい (どちらになるのかに偏りがなく不確実性が一番高い) とき、平均情報量が一番高くなることが確認できる。(二項分布についての詳細は 1.2.1 節にて)

・KL ダイバージェンス

KL ダイバージェンスは、$p(x)$ と $q(x)$ の情報量の差 $-\log p(x) - (-\log q(x))$ の期待値である。

$$\begin{aligned} KL(p,q) &= \sum_x p(x)\Big\{-\log p(x) - \Big(-\log q(x)\Big)\Big\} \\ &= \sum_x p(x) \log \frac{p(x)}{q(x)} \end{aligned} \tag{1.6}$$

変数 x が連続値である場合は次となる。(離散値であればそのまま和をとればよいので $\sum$ 足し合わせ、連続値であれば積分して求める。)

$$KL(p,q) = \int p(x) \log \frac{p(x)}{q(x)} dx$$

2 つの確率分布に非類似度が $KL(p,q) = 0$(差がない) ということは、同じ分布 $(p = q)$ であるということである。また、$p = q$ のとき $KL(p,q) = 0$ となる。

$$KL(p,q) = 0 \Leftrightarrow p = q$$

なお、非類似度を距離のように扱われることもあるが、$KL(p,q) \neq KL(q,p)$ であるため、数学的な距離の公理を満たさない。

真の分布と計算を簡単にした近似分布との KL ダイバージェンスの最小化が、尤度 (尤もらしさ) の最大化となる。この本では、推定処理の終了条件として用いる。

1.1.9 連続確率変数

これまではサイコロの目のように $1, 2, \cdots$ と値をとる、離散値を扱ってきた。ここでは、各値が切れ目なく続く連続値 (1 の次の値は 1.000...) について確認していく。

1~6 の値の中から 1 が出る確率を、離散値で考えると $\frac{1}{6}$ だが、連続値だと 1~6 の値は無限にあるので、$\frac{1}{\infty} = 0$ となる。
そこで確率密度という考え方を使い、相対的な出やすさで表す。(箱の中に $1, 2, \cdots, \infty$ までのカード (これ自体は離散値) が 1 枚ずつ入っている。ここから 1 枚引きそれが、1 のカード・1~10 どれかのカードである確率は $\frac{1}{\infty} = \frac{10}{\infty} = 0$ なのだが、「1」と「1~10」の 2 つを比較すると「1~10」の方が 10 倍出やすいよねという考え方)
これを座標平面上で考えると、x や y の値 (点) ではなく範囲 (面積) として表される (面積を求める、つまり積分を使って求める)。面積が大きいほど確率が高くなる。

連続値をとる確率変数 X が a から b の値をとる確率は

$$p(a \leq X \leq b) = \int_a^b p(x)dx$$

になる。これを確率密度関数と呼ぶ (離散値の場合は確率質量関数と呼ぶ)。連続値の場合でも確率なので、正の値をとり、また全ての可能性を考慮すると値は 1 となる。

$$0 \leq p(x) \leq 1, \int_{-\infty}^{\infty} p(x)dx = 1$$

離散値の場合は $\sum$ を使って和をとったが、連続値の場合は積分して期待値を求める。

$$\mathbb{E}[f] = \int_{x \in \mathcal{X}} p(x)f(x)dx$$

周辺化や KL ダイバージェンスの計算でも同様に、和を積分に置き換えることで計算できる。

1.1.10 イェンゼンの不等式

上向きの凸関数 (窪みのないグラフになる関数) に対して、イェンゼンの不等式を用いることでその関数の下限を求める事ができる。

簡単な例として単語が 2 種類のみ存在するとする。このとき、単語 $v = (1, 2)$ の出現数を x_1, x_2 とし、出現確率を $\boldsymbol{\phi} = \{\phi_1, \phi_2\}, \phi_1 + \phi_2 = 1$ とする。単語の出現数を出現確率で加重平均した値 $\phi_1 x_1 + (1 - \phi_1)x_2 = \phi_1 x_1 + \phi_2 x_2$ は単語数の期待値である。
またこれは､曲線 $f(x)$ 上の 2 点からなる線分 $x_1 x_2$ を $\phi_1 : (1 - \phi_1)$ で内分した点になる。$\phi_1 f(x_1) + \phi_2 f(x_2)$ は、同様に縦軸方向に内分する点のことである。曲線を曲線上の 2 点を結ぶ線分で分割したとき、その線分上のどの点よりも大きい曲線上の点が必ず存在することを示している。(図 1.2)

同様に単語数が V の場合 $\boldsymbol{\phi} = \{\phi_1, \phi_2, \cdots, \phi_V\}, \sum_{v=1}^V \phi_v = 1$ について考えると、確率質量関数 $f\left(\sum_{v=1}^V \phi_v x_v\right)$ と $\sum_{v=1}^V \phi_v f(x_v)$ は常に以下の関係が成り立つ。

$$f\left(\sum_{v=1}^V \phi_v x_v\right) \geq \sum_{v=1}^V \phi_v f(x_v) \tag{1.7}$$

この式をイェンゼンの不等式と呼ぶ。この式の関係性から、右辺 $\sum_{v=1}^V \phi_v f(x_v)$ を最大化するとき、左辺も最大化されていることが分かる。また、右辺は左辺の下限の式といえる。
つまり、求めたい関数の計算が困難であるとき、イェンゼンの不等式を用いてその関数の下限となる関数を求めて、それを最大化することができる。

この本では、主に変分ベイズ推定において次のように

$$
\begin{aligned}
\log p(\boldsymbol{W}) &= \log \int \sum_{\boldsymbol{z}} q(\boldsymbol{z}, \boldsymbol{\Psi}) \frac{p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi})}{q(\boldsymbol{z}, \boldsymbol{\Psi})} d\boldsymbol{\Psi} \\
&\geq \int \sum_{\boldsymbol{z}} q(\boldsymbol{z}, \boldsymbol{\Psi}) \log \frac{p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi})}{q(\boldsymbol{z}, \boldsymbol{\Psi})} d\boldsymbol{\Psi} \qquad (3.10)
\end{aligned}
$$

log の中に $\int$ や $\sum$ の計算を含み、真の分布 (1 行目) を推定することが難しいときに、イェンゼンの不等式を用いて近似分布 (2 行目) を用いて推定していく。

1.1.11 ラグランジュの未定乗数法

ラグランジュの未定乗数法を用いると、ある関数の最大値・最小値を制約条件のもとで求めることができる。(よく分からない (未定の) 値 λ を制約条件の式に掛ける (乗数とする) ことで、制約付きの最適化問題を簡単に解くことができる、というラグランジュさんが考えた方法)

$f(\boldsymbol{\phi})$ を最大化する $\boldsymbol{\phi}$ を求める。ただし、I 個の制約 $g_i(\boldsymbol{\phi}) = 0 \quad (i = 1, 2, \cdots, I)$ があるとする。このような問題を等式制約付き最適化問題といい、ラグランジュの未定乗数法を用いて解くことができる。(要は g の式を満たす範囲で f の式が最大化する値を求めるということ。そして、ラグランジュの未定乗数法を使うと簡単に解けてしまう。)

まずは

$$
\max_{\boldsymbol{\phi}} f(\boldsymbol{\phi}),\ \boldsymbol{\phi} = \{\phi_1, \phi_2, \cdots, \phi_v\}, \quad \text{s.t.} \quad g_i(\boldsymbol{\phi}) = 0 \quad (i = 1, 2, \cdots, I)
$$

について考える。未定の $\boldsymbol{\lambda} = \{\lambda_1, \cdots, \lambda_I\}$ を制約関数 $g_i(\boldsymbol{\phi})$ の乗数として $L(\boldsymbol{\phi}, \boldsymbol{\lambda})$ を作る。(かなりざっくりいうと $f(\boldsymbol{\phi})$ と制約曲線 $g_i(\boldsymbol{\phi})$ との接点 (等しい点) を見付ける感じの操作。つまり $f(\boldsymbol{\phi}) = \lambda_i g_i(\phi)$ となるような関数 $F(\boldsymbol{\phi}, \boldsymbol{\lambda})$ を新たに作る。ちなみに交点も等しい点なのだが、極値ではない)　複数 $(i = 1, \cdots, I)$ の場合には、それぞれの結果を足し合わせる。

$$
L(\boldsymbol{\phi}, \boldsymbol{\lambda}) = f(\boldsymbol{\phi}) + \sum_{i=1}^{I} \lambda_i g_i(\boldsymbol{\phi})
$$

この式を偏微分して解くことで

$$
\frac{\partial L(\boldsymbol{\phi})}{\partial \boldsymbol{\phi}} = 0,\ \frac{\partial L(\boldsymbol{\lambda})}{\partial \boldsymbol{\lambda}} = 0
$$

制約を満たす範囲内で最大化する値が求まる。

・Tips：例題

簡単な例として、長方形の面積の最大化について考える。長方形の 2 辺 x, y の合計が 10 としたとき、長方形の面積が最大となる x, y を求めるとする。つまり、制約 $g(x, y) = 10$ の下で、面積 $f(x, y)$ を最大化する 2 辺 x, y を、ラグランジュの未定乗数法を用いて求める。

それぞれの関数は

$$f(x,y) = xy,\ g(x,y) = x + y = 10$$

である。これに未定乗数 λ を使って、$L(x, y, \lambda)$ を立てる。

$$L(x,y,\lambda) = xy - \lambda(x + y - 10)$$

この式を各変数について偏微分する。

$$\begin{aligned}
L(x,y,\lambda) &= xy - \lambda(x + y - 10) \\
&= (y - \lambda)x - (\lambda y - 10\lambda) \\
\frac{\partial F}{\partial x} &= y - \lambda = 0 \qquad (1)
\end{aligned}$$

偏微分とは、複数の変数を持つ関数を微分する際に、1 つの変数のみを微分し、他の変数は定数として扱うこと。他の変数についても同様に解く。

$$\begin{aligned}
\frac{\partial L}{\partial y} &= x - \lambda = 0 \qquad (2) \\
\frac{\partial L}{\partial \lambda} &= x + y + 10 = 0 \qquad (3)
\end{aligned}$$

(1)、(2)、(3) の連立方程式を解くと

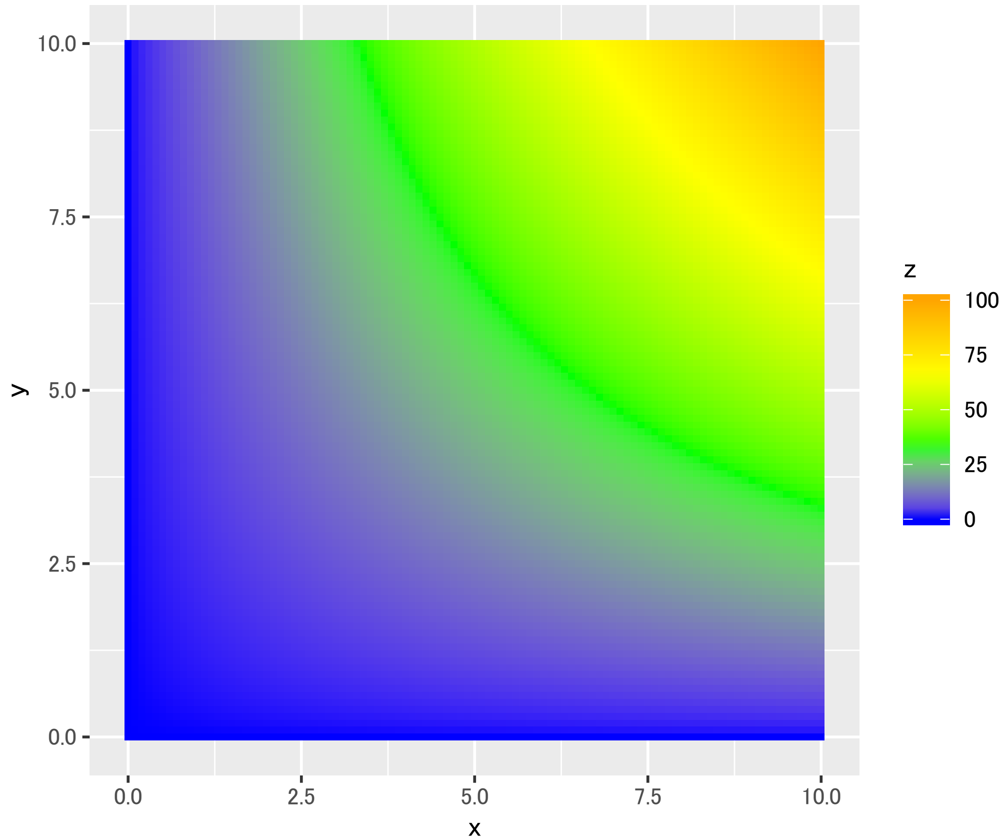
$$x = 5,\ y = 5,\ \lambda = 5$$

が求まる。$g(x, y) = x + y = 10$ の制約を満たす面積の最大値は $f(x, y) = xy = 25$ であり、またそのときの 2 辺は $x = y = 5$ であることが分かった。

続いて、この関係を図で確認していく。まずは、2 辺と面積の関係をみる。

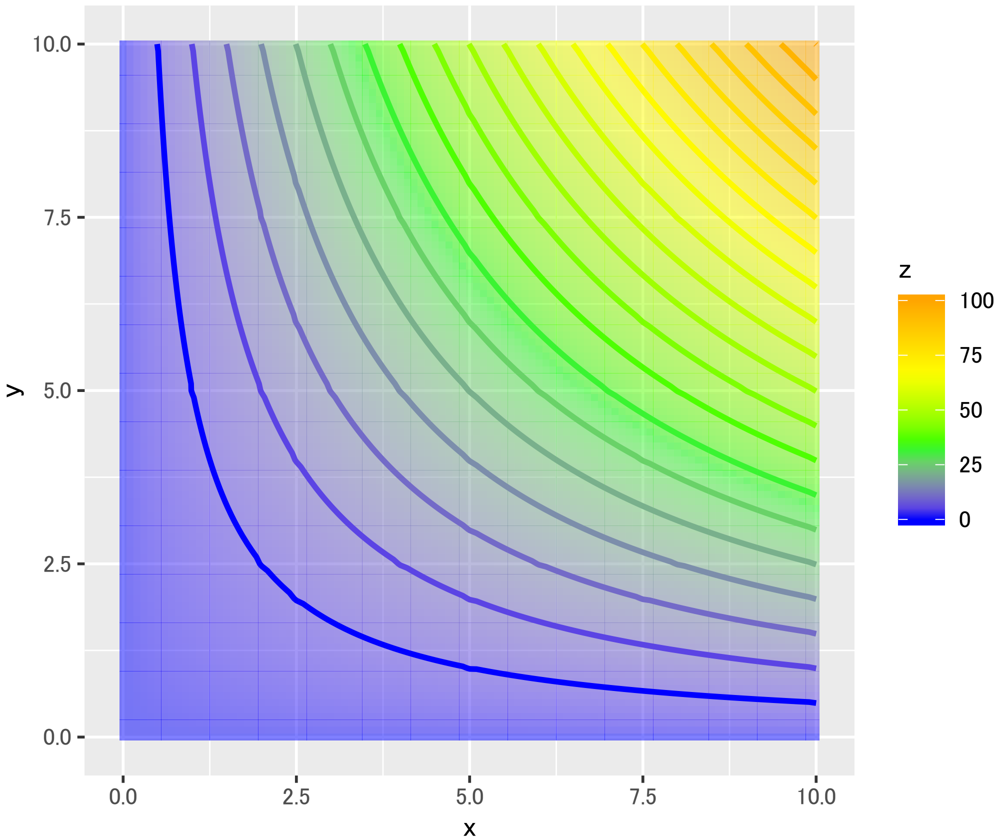
```r
# 利用パッケージ
library(tidyverse)

# 面積のデータフレームを作成
f_xy <- tibble(x = rep(seq(0, 10, 0.1), 101),
               y = rep(seq(0, 10, 0.1), each = 101),
               z = x * y)
```

```r
# 辺と面積の関係を描画
ggplot(data = f_xy, mapping = aes(x = x, y = y, z = z, fill = z)) +       # データ
  geom_tile() +                                                           # ヒートマップ
  scale_fill_gradientn(colors = c("blue", "green", "yellow", "orange")) + # タイルの色
  coord_fixed(ratio = 1) # 縦横比
```



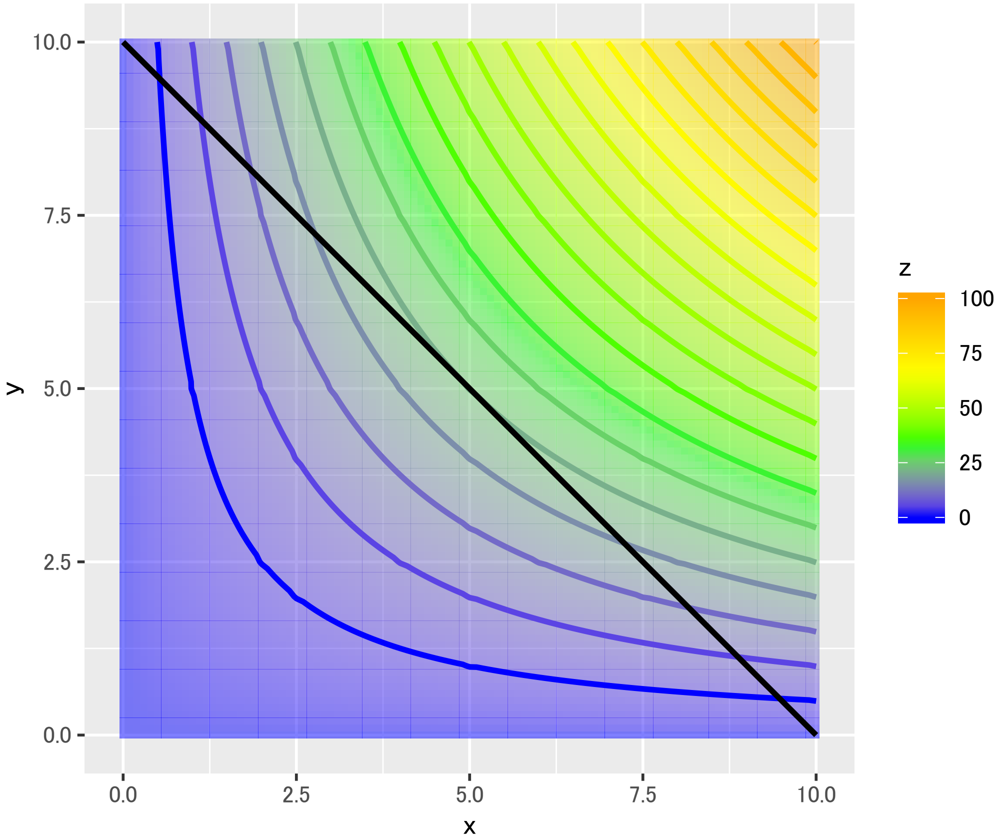
長方形の 2 辺を x 軸、y 軸とし、面積を z としている。面積 z の値が大きくなるほど、色が青からオレンジになる (各辺を縦横、面積を高さとした三次元のグラフトの代用)。このままでは制約との関係が分かりにくいため、面積 $f(x, y) = z$ が等しい 2 辺の組み合わせを結ぶ曲線を 5 刻みで引く。

```r
# 辺と面積（等高線）の関係を描画
ggplot(data = f_xy, mapping = aes(x = x, y = y, z = z)) +  # データ
  geom_tile(aes(fill = z), alpha = 0.5) +                                 # ヒートマップ
  scale_fill_gradientn(colors = c("blue", "green", "yellow", "orange")) + # タイルの色
  geom_contour(aes(color = stat(level)), binwidth = 5, size = 1) +         # 等高線
  scale_color_gradientn(colors = c("blue", "green", "yellow", "orange")) + # 線の色
  guides(color = FALSE) +  # 凡例
  coord_fixed(ratio = 1)   # 縦横比
```



一番左側の青色の曲線が $f(x, y) = 5$ となる 2 辺の組み合わせを表す。ここに、制約曲線 $g(x, y)$ を加える (ここでは直線になる)。

```
# 制約のデータフレームを作成
g_xy <- tibble(x = seq(0, 10, 0.1),
               y = 10 - x)

# 辺・面積 (等高線) と制約の関係を描画
ggplot() +
  geom_tile(data = f_xy, mapping = aes(x = x, y = y, fill = z), alpha = 0.5) +  # ヒートマップ
  scale_fill_gradientn(colors = c("blue", "green", "yellow", "orange")) +       # タイルの色
  geom_contour(data = f_xy, aes(x = x, y = y, z = z, color = stat(level)),
               binwidth = 5, size = 1) +                                      # 等高線
  scale_color_gradientn(colors = c("blue", "green", "yellow", "orange")) + # 線の色
  geom_line(data = g_xy, mapping = aes(x = x, y = y), size = 1) +  # 折れ線グラフ
  guides(color = FALSE) +  # 凡例
  coord_fixed(ratio = 1)   # 縦横比
```



制約曲線 $g(x,y)$ 上の点が、制約 $x+y=10$ を満たす 2 辺の組み合わせである。この中で面積 $f(x,y)$ を最大化する点は、$f(x,y)=25$ の曲線との接点 $(x=5,y=5)$ であることが分かる。

1.2 確率分布

1.2.1 ベルヌーイ分布

・ベルヌーイ分布

コインの裏表やくじの当たり外れのように、2 値をとる変数の確率分布のことをベルヌーイ分布という。(パラメータ等を太字で書くときは、それが集合 $\boldsymbol{\phi}=(\phi_1,\cdots,\phi_V)$ であることを表している。ベルヌーイ分布では 2 値しかとらないため、それに対応する確率は $\phi, 1-\phi$ である。つまり ϕ は 1 つの値であるため、ここでは ϕ が太字ではない。)

例えば、コインを投げて表なら 1、裏なら 0 である変数を x とするとき、次のように記す。

$$x \in \{0,1\}$$

また、x が取り得る値の確率を ϕ を使って以下とする。

$$p(x=1)=\phi,\ p(x=0)=1-\phi,\ 0 \le \phi \le 1$$

このときの確率分布がベルヌーイ分布である。

$$\mathrm{Bernoulli}(x|\phi) = \phi^x(1-\phi)^{1-x}$$

この式は、コインが表、つまり $x = 1$ のとき

$$\begin{aligned}\mathrm{Bernoulli}(x=1|\phi) &= \phi^1(1-\phi)^{1-1} \\ &= \phi * 1 \\ &= \phi\end{aligned}$$

となり、またコインが裏、つまり $x = 0$ のときは

$$\begin{aligned}\mathrm{Bernoulli}(x=0|\phi) &= \phi^0(1-\phi)^{1-0} \\ &= 1 * (1-\phi) \\ &= 1-\phi\end{aligned}$$

となる。このように、x の値に対応した確率となる式によって確率分布が表現されている。

次からは、この分布の期待値と分散を求めていく。

・期待値の導出

確率分布が取り得る値 x とその値となる確率 $p(x)$ とを掛け合わせて、和をとった値が期待値となる。

$$\begin{aligned}\mathbb{E}[x] &= \sum_{x=0}^{1} xp(x) && \text{(i)} \\ &= \sum_{x=0}^{1} x\phi^x(1-\phi)^{1-x} && \text{(ii)} \\ &= \{0 * \phi^0 * (1-\phi)^{1-0}\} + \{1 * \phi^1 * (1-\phi)^{1-1}\} && \text{(iii)} \\ &= 0 + \phi \\ &= \phi\end{aligned}$$

【途中式の途中式】

- (0)：x と $p(x)$ とを掛け合わせ、また x は 0 の場合と 1 の場合を足し合わせるために $\sum_{x=0}^{1}$ を使って式を立てる。
- (i)：$p(x) = \mathrm{Bernoulli}(x|\phi)$ より、置き換える
- (ii)：$\sum_{x=0}^{1}$ を展開する。
- (iii)：式を整理する。

(期待値が確率であるという意味ではなく、期待値が確率の値と等しくなるという意味である。)

・分散の導出

「x の 2 乗の期待値」と「x の期待値の 2 乗」との差が分散となる。そこでまずは、x の 2 乗の期待値を求める。

$$
\begin{aligned}
\mathbb{E}[x^2] &= \sum_{x=0}^{1} x^2 p(x) \\
&= \sum_{x=0}^{1} x^2 \phi^x (1-\phi)^{1-x} \\
&= \{0^2 * \phi^0 * (1-\phi)^{1-0}\} + \{1^2 * \phi^1 * (1-\phi)^{1-1}\} \\
&= 0 + \phi \\
&= \phi
\end{aligned}
$$

x の 2 乗の期待値と x の期待値の 2 乗との差を求める。

$$
\begin{aligned}
\mathrm{Var}[x] &= \mathbb{E}[x^2] - (\mathbb{E}[x])^2 \\
&= \phi - \phi^2 \\
&= \phi(1-\phi)
\end{aligned}
$$

ベルヌーイ分布とは、2 値をとる確率変数を 1 回試行したときの確率分布であった。次に、N 回行う場合の分布について考えていく。

・二項分布

表 (当たり) の出る確率が ϕ であるコイン (くじ) を N 回試行するとき、表 (当たり) の出る回数 x の確率分布を二項分布という。

x は、コインを N 回投げて表が出た回数のことである。つまり、x は 0(全て裏) から N(全て表) の値をとる。

$$
x = \{0, 1, \cdots, N\}
$$

ベルヌーイ分布と同様に、表となる確率を ϕ とすると二項分布は次の式となる。

$$
\mathrm{Binomial}(x|N,\phi) = \binom{N}{x} \phi^x (1-\phi)^{N-x}
$$

ここで

$$
\binom{N}{x} = {}_N\mathrm{C}_x = \frac{N!}{(N-x)!x!}
$$

は試行回数が N 回のときに、x 回表となる組み合わせを求めるための項である。

例えば、表が出る確率が ϕ のコインを 3 回投げた $(N=3)$ とき、表が 2 回出る $(x=2$ となる$)$ 確率は

$$\begin{aligned}\mathrm{Binomial}(x=2|N=3,\phi) &= \frac{3!}{(3-2)!2!}\phi^2(1-\phi)^{3-2} \\ &= 3*\phi^2*(1-\phi) \\ &= 3\phi^2(1-\phi)\end{aligned}$$

になる。このとき、正規化項は 3 回中 2 回表となる組み合わせを表し、$\{\mathrm{omote},\mathrm{omote},\mathrm{ura}\}$、$\{\mathrm{omote},\mathrm{ura},\mathrm{omote}\}$、$\{\mathrm{ura},\mathrm{omote},\mathrm{omote}\}$ の 3 通りであることを求めている。

また、試行回数が 1 回 $(N=1)$ のとき、この式 (分布) は

$$\begin{aligned}\mathrm{Binomial}(x|1,\phi) &= \frac{1!}{(1-x)!x!}\phi^x(1-\phi)^{1-x} \\ &= \frac{1!}{1!0!}\phi^x(1-\phi)^{1-x} \\ &= 1*\phi^x(1-\phi)^{1-x} = \mathrm{Bernoulli}(x|\phi)\end{aligned}$$

【途中式の途中式】

- $0 \le x \le N = 1$ の離散値なので、x の値は 0 か 1 である。また $0! = 1! = 1$ であるため、x がどちらであっても $\frac{1!}{(1-x)!x!} = 1$ になる。

ベルヌーイ分布と等しくなることが分かる。

次からは、この分布の期待値と分散を求めていく。

・期待値の導出

確率変数が取り得る値 x とその値となる確率 $p(x)$ とを掛け合わせて、和をとった値が期待値となる。

$$
\begin{aligned}
\mathbb{E}[x] &= \sum_{x=0}^{N} x p(X=x) && \text{(i)}\\
&= \sum_{x=0}^{N} x \binom{N}{x} \phi^x (1-\phi)^{N-x} && \text{(ii)}\\
&= \sum_{x=0}^{N} x \frac{N!}{(N-x)!x!} \phi^x (1-\phi)^{N-x} && \text{(iii)}\\
&= 0 \frac{N!}{(N-0)!0!} \phi^0 (1-\phi)^{N-0} + \sum_{x=1}^{N} x \frac{N!}{(N-x)!x!} \phi^x (1-\phi)^{N-x} && \text{(iv)}\\
&= \sum_{x=1}^{N} x \left\{ N * \frac{1}{x} * \frac{(N-1)!}{(N-x)!(x-1)!} \right\} (\phi * \phi^{x-1})(1-\phi)^{N-x} && \text{(v)}\\
&= N\phi \sum_{x=1}^{N} \frac{(N-1)!}{(N-x)!(x-1)!} \phi^{x-1} (1-\phi)^{N-x} && \text{(vi)}\\
&= N\phi \sum_{x=0}^{N-1} p(X=x-1) && \text{(vii)}\\
&= N\phi * 1\\
&= N\phi
\end{aligned}
$$

【途中式の途中式】

- (0)：表となる回数 x とその回数となる確率 $p(x)$ とを掛け合わせて、0 回 $(x=0)$ から N 回 $(x=N)$ までを全て足し合わせる $(\sum_{x=0}^{N})$。
- (i)：$p(x) = \mathrm{Binomial}(x|N,\phi)$ より、置き換える。
- (ii)：$\binom{N}{x} = \frac{N!}{(N-x)!x!}$ より、置き換える。
- (iii)：$x=0$ に関する項だけを取り出す。また、この項は 0 となるため消える。
- (iv)：$N! = N(N-1)!$、$\frac{1}{x!} = \frac{1}{x(x-1)!}$、$\phi^x = \phi\phi^{x-1}$ に、それぞれの項を分割する。
- (v)：N と ϕ を $\sum$ の外に出して、式を整理する。
- (vi)：

$$
\sum_{x=1}^{N} \frac{(N-1)!}{(N-x)!(x-1)!} \phi^{x-1} (1-\phi)^{N-x} = \sum_{x-1=0}^{N-1} \frac{(N-1)!}{\{(N-1)-(x-1)\}!(x-1)!} \phi^{x-1} (1-\phi)^{(N-1)-(x-1)}
$$

に式変形して、$N' = N-1,\ x' = x-1$ で置き換えると

$$
= \sum_{x'=0}^{N'} \frac{N'!}{(N'-x')!x'!} \phi^{x'} (1-\phi)^{N'-x'} = \sum_{x'=0}^{N'} p(x')
$$

となる。

- (vi)：試行回数が N' 回のときに、表の回数 x' について全ての可能性 (確率) を足し合わせたことになるため

$$
\sum_{x'=0}^{N'} p(x') = 1
$$

となる。

・分散の導出

「x の 2 乗の期待値」と「x の期待値の 2 乗」との差が分散となる。そこでまずは、x の 2 乗の期待値を求める。

$$
\begin{aligned}
\mathbb{E}[x^2] &= \sum_{x=0}^{N} x^2 p(x) && \text{(i)}\\
&= \sum_{x=0}^{N} x^2 \binom{N}{x} \phi^x (1-\phi)^{N-x} && \text{(ii)}\\
&= \sum_{x=0}^{N} x^2 \frac{N!}{(N-x)!x!} \phi^x (1-\phi)^{N-x} && \text{(iii)}\\
&= \sum_{x=0}^{N} \{x(x-1)+x\} \frac{N!}{(N-x)!x!} \phi^x (1-\phi)^{N-x} && \text{(iv)}\\
&= \sum_{x=0}^{N} x(x-1) \frac{N!}{(N-x)!x!} \phi^x (1-\phi)^{N-x} + \sum_{x=0}^{N} x \frac{N!}{(N-x)!x!} \phi^x (1-\phi)^{N-x} && \text{(v)}\\
&= \sum_{x=0}^{N} x(x-1) \left\{ N(N-1) * \frac{1}{x(x-1)} * \frac{(N-2)!}{(N-x)!(x-2)!} \right\} (\phi^2 * \phi^{x-2})(1-\phi)^{N-x} + N\phi && \text{(vi)}\\
&= N(N-1)\phi^2 \sum_{x=2}^{N} \frac{(N-2)!}{(N-x)!(x-2)!} \phi^{x-2} (1-\phi)^{N-x} + N\phi && \text{(vii)}\\
&= N(N-1)\phi^2 \sum_{x=0}^{N-2} p(x-2) + N\phi && \text{(viii)}\\
&= N(N-1)\phi^2 * 1 + N\phi\\
&= N(N-1)\phi^2 + N\phi
\end{aligned}
$$

【途中式の途中式】

- (0)：先ほどと同様に、x を 2 乗した値に確率を掛け合わせたものの和をとる。
- (i)：$p(x) = \mathrm{Binomial}(x|N, \phi)$ より、置き換える。
- (ii)：$\binom{N}{x} = \frac{N!}{(N-x)!x!}$ より、置き換える。
- (iii)：$x^2 = x(x-1) + x$ に、項を分割する。
- (iv)：$\sum(A+B) = \sum A + \sum B$ であることから、式を分割する。
- (v)：
  - $N! = N(N-1)(N-2)!$、$\frac{1}{x!} = \frac{1}{x(x-1)(x-2)!}$、$\phi^x = \phi^2\phi^{x-2}$ に、それぞれ項を分割する。
  - $\mathbb{E}[X] = \sum_{x=0}^{N} x \frac{N!}{(N-x)!x!} \phi^x (1-\phi)^{N-x} = N\phi$ より、置き換える。
- (vi)：$N(N-1)$ と ϕ^2 を $\sum$ の外に出して、式を整理する。
- (vii)：

$$
\sum_{x=2}^{N} \frac{(N-2)!}{(N-x)!(x-2)!} \phi^{x-2} (1-\phi)^{N-x} = \sum_{x-2=0}^{N-2} \frac{(N-2)!}{\{(N-2)-(x-2)\}!(x-2)!} \phi^{x-2} (1-\phi)^{(N-2)-(x-2)}
$$

に式変形して、$N' = N - 2,\ x' = x - 2$ で置き換えると

$$
\sum_{x'=0}^{N'} \frac{N'!}{(N'-x')!x'!} \phi^{x'} (1-\phi)^{N'-x'} = \sum_{x'=0}^{N'} p(x')
$$

となる。

- (viii)：試行回数が N' 回のときに、表の回数 x' について全ての可能性 (確率) を足し合わせたことになるため

$$
\sum_{x'=0}^{N'} p(x') = 1
$$

となる。

---

x の 2 乗の期待値と x の期待値の 2 乗との差を求める。

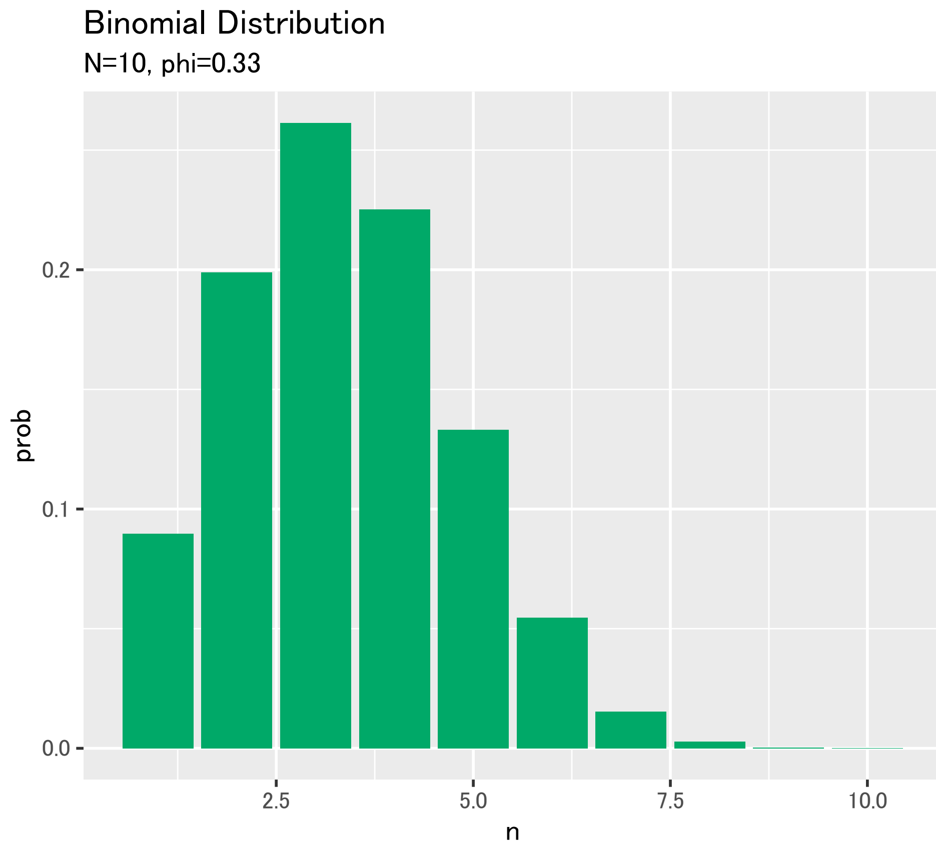
$$
\begin{aligned}
\mathrm{Var}[x] &= \mathbb{E}[x^2] - (\mathbb{E}[x])^2 \\
&= N(N-1)\phi^2 + N\phi - (N\phi)^2 \\
&= N^2\phi^2 - N\phi^2 + N\phi - N^2\phi^2 \\
&= N\phi(1-\phi)
\end{aligned}
$$

$N = 1$(試行回数が 1) のとき、二項分布の期待値は $N\phi = \phi$、分散は $N\phi(1-\phi) = \phi(1-\phi)$ となり、ベルヌーイ分布の期待値と分散とそれぞれ等しくになることも確認できる。

・分布

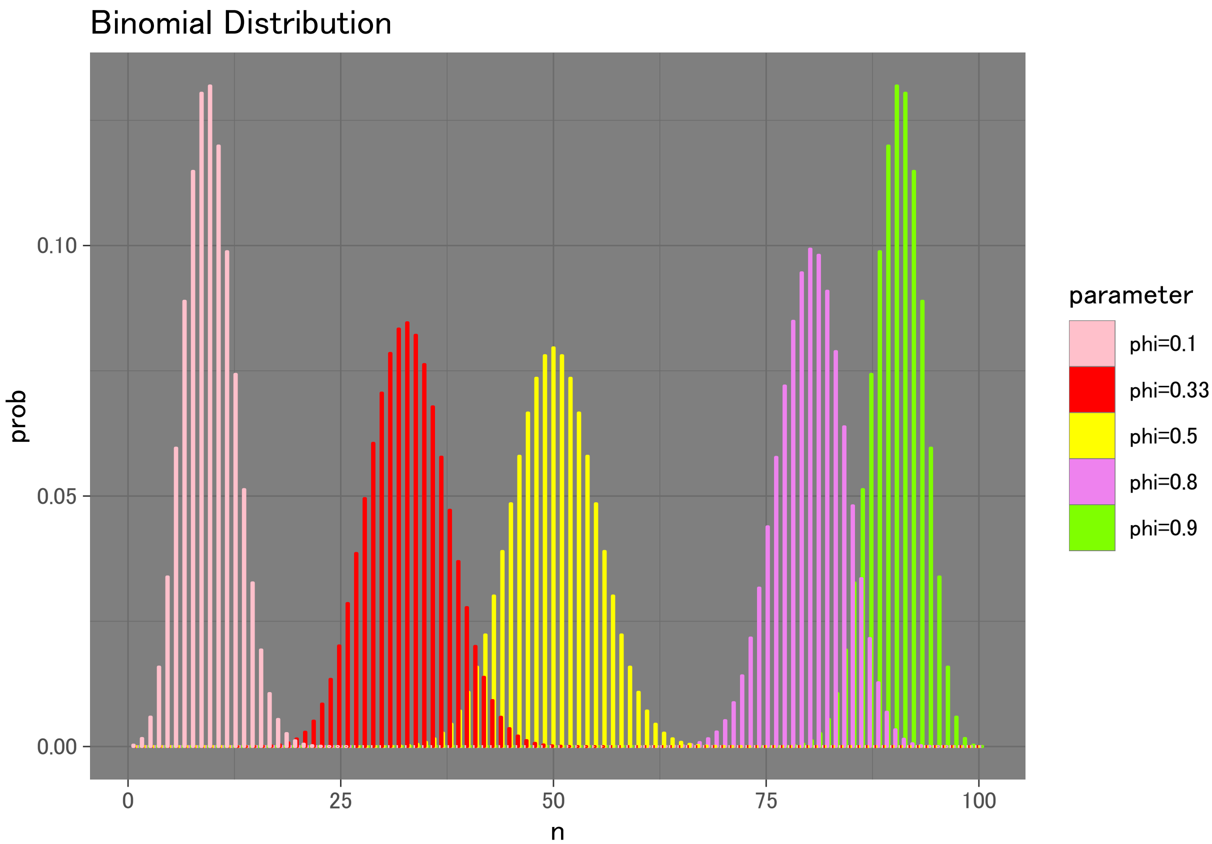
表になる確率が 0.2 である歪んだコインを 10 回投げたときに、表が出る回数の分布を確認する。

```
# 利用パッケージ
library(tidyverse)

# パラメーターを指定
N   <- 10     # 任意の値を指定する
phi <- 0.33   # 任意の値を指定する

# 作図
tibble(n = 1:N,                             # 表の回数
       prob = dbinom(n, N, phi)) %>%   # 確率の算出
  ggplot(mapping = aes(x = n, y = prob)) +           # データ
    geom_bar(stat = "identity", position = "dodge", fill = "#00A968") + # 棒グラフ
    labs(title = "Binomial Distribution",            # タイトル
         subtitle = paste0("N=", N, ", phi=", phi)) # サブタイトル
```



・確率の違いによる分布の変化

```
# パラメーターを指定
N   <- 100    # 任意の値を指定する
phi <- c(0.1, 0.33, 0.5, 0.8, 0.9) # 任意の値を指定する

# 作図用のデータフレームを作成
res_df <- data.frame()
for(i in seq_along(phi)) {
  # 計算結果のデータフレームを作成
  tmp_df <- tibble(n = 1:N,                             # 表の回数
                   prob = dbinom(n, N, phi[i]),         # 確率の算出
                   parameter = paste0("phi=", phi[i])) # 作図用のラベル

  # データフレームを結合
  res_df <- rbind(res_df, tmp_df)
}

# 描画
ggplot(data = res_df, mapping = aes(x = n, y = prob,
                                     fill = parameter, color = parameter)) + # データ
  geom_bar(stat = "identity", position = "dodge") +  # 棒グラフ
  theme_dark() +                                     # 背景色
  labs(title = "Binomial Distribution")              # タイトル
```



1.2.2 カテゴリ分布

・カテゴリ分布

ベルヌーイ分布・二項分布は、変数が 2 値 (コインに例えられる) である確率分布であった。この節では、$\{1, 2, \cdots, V\}$ から 1 つの値をとる (試行回数が 1 回) カテゴリ分布について考える。

値 v となる回数を x_v とする。

$$\boldsymbol{x} = (x_1, x_2, \cdots, x_V),\ x_v \in \{0, 1\}$$

試行回数が 1 回なので、x_v は 0 か 1 しかとらない。また、$\boldsymbol{x}$ の内、1 つだけが 1 となる。例えば、サイコロ ($V = 6$) を 1 回振って 3 が出たとき、$x_3 = 1$ であり、それ以外の x_1 から x_6 は 0 となる。従って

$$\sum_{v=1}^{V} x_v = 1$$

である。(ϕ_v は各値が 0 から 1 までの値をとり、全てを足し合わせて 1 となるのに対して、x_v は 1 つの値が 1 でありそれ以外の値が全て 0 であるため、総和が 1 である。)

このとき、v となる ($x_v = 1$ となる) 確率を

$$p(v) = \phi_v,\ 0 \leq \phi_v \leq 1,\ \sum_{v=1}^{V} \phi_v = 1$$

とする。このような確率分布をカテゴリ分布

$$\mathrm{Categorical}(\boldsymbol{x}|\boldsymbol{\phi}) = \prod_{v=1}^{V} \phi_v^{x_v}$$

と呼ぶ。ここで $\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots, \phi_V)$ は、x が 1 から V の値をとるそれぞれの確率の集合である。

カテゴリ分布が $V = 2$ のとき、$\boldsymbol{\phi} = (\phi_1, \phi_2)$ であるため

$$\begin{aligned}
\sum_{v=1}^{2} \phi_v &= 1 \\
\phi_1 + \phi_2 &= 1 \\
\phi_2 &= 1 - \phi_1
\end{aligned}$$

となる。同様に

$$\begin{aligned}
\sum_{v=1}^{2} x_v &= 1 \\
x_1 + x_2 &= 1 \\
x_2 &= 1 - x_1
\end{aligned}$$

であるため、$V = 2$ のときのカテゴリ分布は

$$\begin{aligned}
\mathrm{Categorical}(x|\phi_1, \phi_2) &= \prod_{v=1}^{2} \phi_v^{x_v} \\
&= \phi_1^{x_1} * \phi_2^{x_2} \\
&= \phi_1^{x_1}(1 - \phi_1)^{1-x_1} = \mathrm{Bernoulli}(x|\phi_1)
\end{aligned}$$

ベルヌーイ分布と等しくなることが確認できる。

次からは、カテゴリ分布の期待値と分散を求めていく。

・期待値の導出

確率変数が取り得る値 $\boldsymbol{x}$ とその値となる確率 $p(\boldsymbol{x})$ とを掛け合わせて、和をとった値が期待値となる。

$$
\begin{aligned}
\mathbb{E}[x_v] &= \sum_{v=1}^{V} x_v p(x_v) && \text{(i)} \\
&= \sum_{v=1}^{V} x_v \prod_{v'=1}^{V} \phi_{v'}^{x'_v} && \text{(ii)} \\
&= \sum_{v=1}^{V} x_v * \phi_1^{x_1} * \cdots * \phi_v^{x_v} * \cdots * \phi_V^{x_V} && \text{(iii)} \\
&= (x_1 * \phi_1^{x_1} * \cdots * \phi_v^{x_v} * \cdots * \phi_V^{x_V}) + \cdots + (x_v * \phi_1^{x_1} * \cdots * \phi_v^{x_v} * \cdots * \phi_V^{x_V}) \\
&\qquad + \cdots + (x_V * \phi_1^{x_1} * \cdots * \phi_v^{x_v} * \cdots * \phi_V^{x_V}) && \text{(iv)} \\
&= (0 * \phi_1^{0} * \cdots * \phi_v^{1} * \cdots * \phi_V^{0}) + \cdots + (1 * \phi_1^{0} * \cdots * \phi_v^{1} * \cdots * \phi_V^{0}) \\
&\qquad + \cdots + (0 * \phi_1^{0} * \cdots * \phi_v^{1} * \cdots * \phi_V^{0}) \\
&= 0 + \cdots + \phi_v + \cdots + 0 \\
&= \phi_v
\end{aligned}
$$

【途中式の途中式】

- (0)：1 から V までの、値 x_v と確率 $p(x_v)$ とを掛け合わせて和をとる。
- (i)：$p(x_v) = \mathrm{Categorical}(\boldsymbol{x}|\boldsymbol{\phi})$ より、置き換える。
- (ii)：$\prod_{v'=1}^{V}$ を展開する。(x_v の v と $\phi_{V'}^{x_{v'}}$ の v' は別物。$v' = 1, \cdots, v, \cdots, V$ の中の v が x_v の v)
- (iii)：$\sum_{v=1}^{V}$ を展開する。
- (iv)：$x_1, \cdots, x_V$ の内、x_v が 1 でそれ以外は 0 である。それぞれ代入する。

(期待値が確率になるという意味ではなく、期待値が確率の値と等しくなるという意味である。)

・分散の導出

「x_v の 2 乗の期待値」と「x_v の期待値の 2 乗」との差が分散となる。そこでまずは、x_v の 2 乗の期待値を求める。

$$
\begin{aligned}
\mathbb{E}[x_v^2] &= \sum_{v=1}^{V} x_v^2 p(x_v) \\
&= \sum_{v=1}^{V} x_v^2 \prod_{v'=1}^{V} \phi_{v'}^{x_{v'}} \\
&= \sum_{v=1}^{V} x_v^2 * \phi_1^{x_1} * \cdots * \phi_v^{x_v} * \cdots * \phi_V^{x_V} \\
&= (x_1^2 * \phi_1^{x_1} * \cdots * \phi_v^{x_v} * \cdots * \phi_V^{x_V}) + \cdots + (x_v^2 * \phi_1^{x_1} * \cdots * \phi_v^{x_v} * \cdots * \phi_V^{x_V}) \\
&\quad + \cdots + (x_V^2 * \phi_1^{x_1} * \cdots * \phi_v^{x_v} * \cdots * \phi_V^{x_v}) \\
&= (0^2 * \phi_1^0 * \cdots * \phi_v^1 * \cdots * \phi_V^0) + \cdots + (1^2 * \phi_1^0 * \cdots * \phi_v^1 * \cdots * \phi_V^0) \\
&\quad + \cdots + (0^2 * \phi_1^0 * \cdots * \phi_v^1 * \cdots * \phi_V^0) \\
&= 0 + \cdots + \phi_v + \cdots + 0 \\
&= \phi_v
\end{aligned}
$$

x_v の 2 乗の期待値と x_v の期待値の 2 乗との差を求める。

$$
\begin{aligned}
\mathrm{Var}[x_v] &= \mathbb{E}[x_v^2] - (\mathbb{E}[x_v])^2 \\
&= \phi_v - \phi_v^2 \\
&= \phi_v(1 - \phi_v)
\end{aligned}
$$

カテゴリ分布は試行回数が 1 回であった、次は試行回数が複数回の場合を考える。

・多項分布

$1, 2, \cdots, V$ からそれぞれに対応する確率 $\phi_1, \phi_2, \cdots, \phi_V$ に従って値 v をとる。それを N 回試行する。このとき、$1, 2, \cdots, V$ の値が出た回数をそれぞれ $x_1, x_2, \cdots, x_V$ とする。試行回数が N なので、x_v は $1, 2, \cdots, N$ 回の範囲となり、x_1(1 が出た回数) から X_V(V が出た回数) までを足し合わせると N 回である。

$$
\boldsymbol{x} = (x_1, x_2, \cdots, x_V),\ x_v \in \{1, 2, \cdots, N\},\ \sum_{v=1}^{N} x_v = N
$$

このときの確率分布を多項分布と呼ぶ。

$$
\mathrm{Multinomial}(\boldsymbol{x}|N, \boldsymbol{\phi}) = \frac{N!}{x_1! x_2! \cdots x_V!} \prod_{v=1}^{V} \phi_v^{x_v}
$$

多項分布が $V = 2, N = 1$ のとき、$\phi_2 = 1 - \phi_1, x_2 = 1 - x_1$ より

$$
\begin{aligned}
\mathrm{Multinomial}(\boldsymbol{x}|N = 2, \phi_1, \phi_2) &= \frac{1!}{x_1! x_2!} \prod_{v=1}^{2} \phi_v^{x_v} \\
&= \frac{1!}{x_1!(1 - x_1)!} \phi_1^{x_1} (1 - \phi_1)^{1 - x_1} \\
&= 1 * \phi_1^{x_1} (1 - \phi_1)^{1 - x_1} = \mathrm{Bernoulli}(x_1|\phi_1)
\end{aligned}
$$

ベルヌーイ分布と等しいことが分かる。また、$V=2$ のとき、$x_2=N-x_1, \phi_2=1-\phi_1$ より

$$
\begin{aligned}
\mathrm{Multinomial}(\boldsymbol{x}|N,\phi_1,\phi_2) &= \frac{N!}{x_1!x_2!}\prod_{v=1}^{2}\phi_v^{x_v} \\
&= \frac{N!}{x_1!(N-x_1)!}\phi_1^{x_1}(1-\phi_1)^{N-x_1} \\
&= \binom{N}{x_1}\phi_1^{x_1}(1-\phi_1)^{N-x_1} = \mathrm{Binomial}(x_1|N,\phi_1)
\end{aligned}
$$

二項分布と等しいことが分かる。続いて、$N=1$ のときは

$$
\begin{aligned}
\mathrm{Multinomial}(\boldsymbol{x}|N=1,\boldsymbol{\phi}) &= \frac{1!}{x_1!x_2!\cdots x_V!}\prod_{v=1}^{V}\phi_v^{x_v} \\
&= \frac{1!}{0!\cdots 1!\cdots 0!}\prod_{v=1}^{V}\phi_v^{x_v} \\
&= 1 * \prod_{v=1}^{V}\phi_v^{x_v} = \mathrm{Categorical}(\boldsymbol{x}|\boldsymbol{\phi})
\end{aligned}
$$

カテゴリ分布と等しくなる。つまりベルヌーイ分布、二項分布、カテゴリ分布は、多項分布の特殊な場合である。あるいは多項分布が、ベルヌーイ分布、二項分布、カテゴリ分布を拡張した形であると確認できた。

次からは、多項分布の期待値と分散を求めていく。

・期待値の導出

確率変数が取り得る値 $\boldsymbol{x}$ とその値となる確率 $p(\boldsymbol{x})$ とを掛け合わせて、和をとった値が期待値となる。

$$
\begin{aligned}
\mathbb{E}[x_v] &= \sum_{v=1}^{V} x_v p(x_v) && \text{(i)} \\
&= \sum_{v=1}^{V} x_v \frac{N!}{x_1!x_2!\cdots x_v!\cdots x_V!}\phi_1^{x_1}\phi_2^{x_2}\cdots\phi_v^{x_v}\cdots\phi_V^{x_V} && \text{(ii)} \\
&= \sum_{v=1}^{V} x_v \frac{1}{x_v} * \frac{N*(N-1)!}{x_1!x_2!\cdots(x_v-1)!\cdots x_V!}\phi_1^{x_1}\phi_2^{x_2}\cdots\phi_v*\phi_v^{x_v-1}\cdots\phi_V^{x_V} && \text{(iii)} \\
&= N\phi_v \sum_{v=1}^{V} \frac{(N-1)!}{x_1!x_2!\cdots(x_v-1)!\cdots x_V!}\phi_1^{x_1}\phi_2^{x_2}\cdots\phi_v^{x_v-1}\cdots\phi_V^{x_V} && \text{(iv)} \\
&= N\phi_v * 1 \\
&= N\phi_v
\end{aligned}
$$

【途中式の途中式】

- (0)：1 から V までの、値 x_v と確率 $p(x_v)$ とを掛け合わせて和をとる。
- (i)：$p(x_v) = \mathrm{Multinomial}(\boldsymbol{x}|N, \boldsymbol{\phi})$ より、置き換える。
- (ii)：$N! = N * (N-1)!$、$\frac{1}{x_v!} = \frac{1}{x_v*(x_v-1)!}$、$\phi_v^{x_v} = \phi_v * \phi_v^{x_v-1}$ に、それぞれ項を分割する。
- (iii)：N と ϕ_v を $\sum$ の外に出して、式を整理する。
- (iv)：試行回数が $N-1$ のときの全ての可能性 (確率) を足し合わせた値であるため

$$\sum_{v=1}^{V} \frac{(N-1)!}{x_1! \cdots (x_v-1)! \cdots x_V!} \phi_1 \cdots \phi_v^{x_v-1} \cdots \phi_V^{x_V} = 1$$

となる。

・分散の導出

「x_v の 2 乗の期待値」と「x_v の期待値の 2 乗」との差が分散となる。そこでまずは、x_v の 2 乗の期待値を求める。

$$\begin{aligned}
\mathbb{E}[x_v^2] &= \sum_{v=1}^{V} x_v^2 p(x_v) && \text{(i)}\\
&= \sum_{v=1}^{V} x_v^2 \frac{N!}{x_1! \cdots x_v! \cdots x_V!} \phi_1^{x_1} \cdots \phi_v^{x_v} \cdots \phi_V^{x_V} && \text{(ii)}\\
&= \sum_{v=1}^{V} \{x_v(x_v-1) + x_v\} \frac{N!}{x_1! \cdots x_v! \cdots x_V!} \phi_1^{x_1} \cdots \phi_v^{x_v} \cdots \phi_V^{x_V} && \text{(iii)}\\
&= \sum_{v=1}^{V} x_v(x_v-1) \frac{N!}{x_1! \cdots x_v! \cdots x_V!} \phi_1^{x_1} \cdots \phi_v^{x_v} \cdots \phi_V^{x_V} + \sum_{v=1}^{V} x_v \frac{N!}{x_1! \cdots x_v! \cdots x_V!} \phi_1^{x_1} \cdots \phi_v^{x_v} \cdots \phi_V^{x_V} && \text{(iv)}\\
&= \sum_{v=1}^{V} x_v(x_v-1) \frac{1}{x_v(x_v-1)} * \frac{N(N-1)*(N-2)!}{x_1! \cdots (x_v-2)! \cdots x_V!} \phi_1^{x_1} \cdots \phi_v^2 * \phi_v^{x_v-2} \cdots \phi_V^{x_V} + N\phi_v && \text{(v)}\\
&= N(N-1)\phi_v^2 \sum_{v=1}^{V} \frac{(N-2)!}{x_1! \cdots (x_v-2)! \cdots x_V!} \phi_1^{x_1} \cdots \phi_v^{x_v-2} \cdots \phi_V^{x_V} + N\phi && \text{(vi)}\\
&= N(N-1)\phi_v^2 * 1 + N\phi_v\\
&= N(N-1)\phi_v^2 + N\phi_v
\end{aligned}$$

【途中式の途中式】

- (0)：1 から V までの、値 x_v の 2 乗と確率 $p(x_v)$ とを掛け合わせて和をとる。
- (i)：$p(x_v) = \mathrm{Multinomial}(\boldsymbol{x}|N, \boldsymbol{\phi})$ より、置き換える。
- (ii)：$x^2 = x(x-1) + x$ に、項を分割する。
- (iii)：$\sum(A+B) = \sum A + \sum B$ であることから、式を分割する。
- (iv)：
  - $N! = N(N-1)(N-2)!$、$\frac{1}{x_v!} = \frac{1}{x_v(x_v-1)(x_v-2)!}$、$\phi_v^{x_v} = \phi_v^2 \phi_v^{x_v-2}$ にそれぞれ分割する。
  - $\sum_{v=1}^{V} x_v \frac{N!}{x_1! x_2! \cdots x_V!} \phi_1^{x_1} \phi_2^{x_2} \cdots \phi_V^{x_V} = \mathbb{E}[x_v] = N\phi_v$ より、置き換える。
- (v)：$N(N-1)$ と ϕ_v^2 を $\sum$ の外に出して、式を整理する。
- (vi)：試行回数が $N-2$ のときの全ての可能性 (確率) を足し合わせた値であるため

$$\sum_{v=1}^{V} \frac{(N-2)!}{x_1! \cdots (x_v-2)! \cdots x_V!} \phi_1^{x_1} \cdots \phi_v^{x_v-2} \cdots \phi_V^{x_V} = 1$$

となる。

---

x_v の 2 乗の期待値と x_v の期待値の 2 乗との差を求める。

$$\begin{aligned}
\mathrm{Var}[x_v] &= \mathbb{E}[x_v^2] - (\mathbb{E}[x_v])^2 \\
&= N(N-1)\phi_v^2 + N\phi_v - (N\phi_v)^2 \\
&= N^2\phi_v^2 - N\phi_v^2 + N\phi - N^2\phi_v^2 \\
&= N\phi_v(1-\phi_v)
\end{aligned}$$

$N = 1$(試行回数が 1) のとき、多項分布の期待値は $N\phi_v = \phi_v$、分散は $N\phi_v(1-\phi_v) = \phi_v(1-\phi_v)$ となり、カテゴリ分布の期待値と分散とそれぞれ等しくになることが確認できる。

1.2.3 ベータ分布

ベータ分布とは、ベルヌーイ分布・二項分布のパラメータ $0 \leq \phi \leq 1$ を生成するための確率分布で、共役事前分布とも呼ばれる。

ベータ分布はパラメータ

$$\alpha > 0,\ \beta > 0$$

を持つ。共役事前分布のパラメータのことをハイパーパラメータとも呼ぶ。ベータ分布

$$\mathrm{Beta}(\phi|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \phi^{\alpha-1}(1-\phi)^{\beta-1} \tag{1.8}$$

によって、ベルヌーイ分布・二項分布のパラメータ $\phi, 1-\phi$ が求まる。

まずは、確率関数として $\phi^{\alpha-1}(1-\phi)^{\beta-1}$ について考える。

これを確率分布として扱うためには、全ての事象を考慮した際の値が 1 となるようにしたい。しかしこの式のままでは、そうはならない。そこで、正規化項 (この式の逆数となる値)$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ を掛け合わせることで対応する。

・ガンマ関数

正規化項の導出の前に、ガンマ関数

$$\Gamma(x) = \int_0^\infty u^{x-1} \exp(-u) du \tag{1.9}$$

について確認する。($\exp(-u)$ は指数関数で e^{-u} のこと)

ガンマ関数には次のような性質がある。

・性質 1

$$\begin{aligned}
\Gamma(x+1) &= \int_0^\infty u^x \exp(-u) du && \text{(i)} \\
&= \int_0^\infty u^x \Big(-\exp(-u) \Big)' du && \text{(ii)} \\
&= \Big[u^x \Big(-\exp(-u) \Big) \Big]_0^\infty - \int_0^\infty x u^{x-1} \Big(-\exp(-u) \Big) du && \text{(iii)} \\
&= 0 + x \int_0^\infty u^{x-1} \exp(-u) du && \text{(iv)} \\
&= x\Gamma(x) && (1.10)
\end{aligned}$$

【途中式の途中式】

- (0):$\Gamma(x)$ のとき u^{x-1} なのだから、$\Gamma(x+1)$ のときは $u^{(x+1)-1} = u^x$ になる。
- (i):指数関数の微分 (と合成関数の微分)$(e^t)' = e^t * t$ を行うと、$(\exp(-u))' = \exp(-u) * (-1)$ となる。
- (ii):部分積分 $\int_a^b f'(x)g(x)dx = [f(x)g(x)]_a^b - \int_a^b f(x)g'(x)dx$ を、$-\exp(-u)$ を $f(x)$、u^x を $g(x)$ として行う。このとき、u^x を微分する。
- (iii):
  - $u \to \infty$ のとき $\exp(-u) = 0$、また $u = 0$ のとき $u^x = 0$ なので、$[u^x(-\exp(-u))]_0^\infty = u^x * 0 - 0 * (-\exp(-u)) = 0$ となる。
  - x を $\int$ の外に出す。
- (iv):ガンマ関数の定義より、$\Gamma(x) = \int_0^\infty u^{x-1} \exp(-u) du$ で置き換える。

・性質 2

$$\begin{aligned}
\Gamma(1) &= \int_0^\infty u^{1-1} \exp(-u) du && \text{(i)} \\
&= \int_0^\infty 1 * \exp(-u) du && \text{(ii)} \\
&= [-\exp(-u)]_0^\infty && \text{(iii)} \\
&= 0 - (-1) = 1 && (1.11)
\end{aligned}$$

【途中式の途中式】

- (i)：$x^0 = 1$ である。
- (ii)：指数関数の積分 $\int e^{ax}dx = \frac{1}{a}e^{ax}$ を行うと、$\int \exp(-u)du = \frac{1}{-1}\exp(-u)$ となる。
- (iii)：$u \to \infty$ のとき $\exp(-u) = 0$、また $u = 0$ のとき $\exp(-u) = e^0 = 1$ となる。

・**性質 3**

x が自然数のとき。

$$\begin{aligned}\Gamma(x+1) &= x\Gamma(x) &\text{(i)}\\ &= x(x-1)\Gamma(x-1) &\text{(ii)}\\ &= x(x-1)(x-2)\cdots*2*1*\Gamma(1) &\text{(iii)}\\ &= x!\end{aligned}$$

【途中式の途中式】

- (i)：$\Gamma(x+1) = x\Gamma(x)$ なのだから、$\Gamma(x) = (x-1)\Gamma(x-1)$ になる。
- (ii)：同様に、$\Gamma(x-1)$ も $(x-2)\Gamma(x-2)$ と置き換えられる。これを繰り返す。
- (iii)：$\Gamma(1) = 1$ なので、x から 1 までの自然数を掛け合わせていることになるため、$x!$ に置き換える。

・**ベータ関数**

・**正規化項の導出**

続いて、ベータ関数

$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$$

を用いて正規化項を導出する。

$m = \alpha - 1, n = \beta - 1$ とおき、$I(m,n) = \int_0^1 x^m(1-x)^n dx$ を解く (表記を分かりやすくするため、$B(\alpha,\beta) = B(m+1,n+1) = I(m,n)$ としておく)。

$$
\begin{aligned}
I(m,n) &= \int_0^1 x^m(1-x)^n dx && \text{(i)}\\
&= \left[\frac{1}{m+1}x^{m+1}(1-x)^n\right]_0^1 - \int_0^1 -\frac{1}{m+1}x^{m+1}n(1-x)^{n-1}dx && \text{(ii)}\\
&= 0 + \frac{n}{m+1}\int_0^1 x^{m+1}n(1-x)^{n-1}dx && \text{(iii)}\\
&= \frac{n}{m+1}I(m+1,n-1) && \text{(iv)}\\
&= \frac{n}{m+1}\frac{n-1}{m+2}I(m+2,n-2) && \text{(v)}\\
&= \frac{n}{m+1}\frac{n-1}{m+2}\cdots\frac{2}{m+n-1}\frac{1}{m+n}I(m+n,0) && \text{(vi)}\\
&= \frac{n!}{(m+1)(m+2)\cdots(m+n)}I(m+n,0) && \text{(vii)}\\
&= \frac{m!n!}{(m+n)!}I(m+n,0) && \text{(viii)}\\
&= \frac{m!n!}{(m+n)!}\int_0^1 x^{m+n}dx && \text{(ix)}\\
&= \frac{m!n!}{(m+n)!}\frac{1}{(m+n+1)} && \text{(x)}\\
&= \frac{m!n!}{(m+n+1)!} && \text{(xi)}\\
&= \frac{(\alpha-1)!(\beta-1)!}{\{(\alpha-1)+(\beta-1)+1\}!} && \text{(xii)}\\
&= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)} = B(\alpha,\beta)
\end{aligned}
$$

【途中式の途中式】

- (i)：部分積分 $\int_a^b f'(x)g(x)dx = [f(x)g(x)]_a^b - \int_a^b f'(x)g'(x)dx$ をする。
  - $f'(x)$ にあたる x^m を積分して、$\frac{1}{m+1}x^{m+1}$ となる。
  - $g(x)$ にあたる $(1-x)^n$ を合成関数の微分 $\{f(g(x))\}' = f'(g(x)) * g'(x)$ をして、$((1-x)^n)' = n(1-x)^{n-1} * (-x^0) = -n(1-x)^{n-1}$ となる。
- (ii)：
  - $x=0$ のとき $\frac{1}{m+1}x^{m+1} = 0$、また $x=1$ のとき $(1-x)^n = 0$ なので、$[\frac{1}{m+1}x^{m+1}]_0^1 = 0 - 0 = 0$ となる。
  - $-n$ と $\frac{1}{m+1}$ を $\int$ の外に出して、式を整理する。
- (iii)：$I(m,n) = \int_0^1 x^m(1-x)^n dx$ なのだから、$I(m+1,n-1) = \int_0^1 x^{m+1}n(1-x)^{n-1}dx$ になる。
- (iv)：$I(m,n) = \frac{n}{m+1}I(m+1,n-1)$ であることが分かったので、$I(m+1,n-1) = \frac{n-1}{m+2}I(m+2,n-2)$ になる。
- (v)：同様に、$I(m+2,n-2)$ は $\frac{n-2}{m+3}I(m+3,n-3)$ と置き換えられる。これを $\frac{1}{m+n}I(m+n,1)$ となるまで繰り返す。
- (vi)：分子の $n(n-1)\cdots 1$ は、n から 1 までの自然数を掛け合わせているので $n!$ に置き換える。
- (vii)

- 分母の $(m+n)\cdots(m+2)(m+1)$ は、$(m+n)$ から $(m+1)$ までの自然数を掛け合わせているので、ここに $m! = m(m-1)\cdots 1$ を掛けることで $m+n$ から 1 までを掛け合わせたことになる。
- つまり、分母分子に $m!$ を掛けることで $\frac{m!n!}{(m+n)!}$ となる。

- (viii)：$I(m+n,0) = \int_0^1 x^{m+n}(1-x)^0 dx$ で置き換える。
- (ix)：積分を行い、$\int_0^1 x^{m+n}dx = \frac{1}{m+n+1}1^{m+n} - \frac{1}{m+n+1}0^{m+n} = \frac{1}{m+n+1}$ となる。
- (x)：$(m+n)!$ に $(m+n)$ より 1 大きい値 $(m+n+1)$ を掛けているので $(m+n+1)!$ となる。
- (xi)：$m=\alpha-1, n=\beta-1$ を戻す。
- (xii)：ガンマ関数の性質 $(x+1)! = \Gamma(x)$ より、各項を置き換える。

ベータ関数 $B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ が求まった。この式の逆数 $\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ を正規化項として用いる。

・ベータ分布

ここからはベータ分布の平均、再頻値、分散を求めていく。

・期待値の導出

分布が取り得る値 ϕ と確率密度関数 $f(\phi)$ を掛け合わせたものを、0 から 1 の範囲で積分して期待値を求める。

$$
\begin{aligned}
\mathbb{E}[\phi] &= \int_0^1 \phi f(\phi) d\phi && \text{(i)} \\
&= \int_0^1 \phi \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \phi^{\alpha-1}(1-\phi)^{\beta-1} d\phi && \text{(ii)} \\
&= \int_0^1 \frac{\phi^{\alpha}(1-\phi)^{\beta-1}}{B(\alpha,\beta)} d\phi && \text{(iii)} \\
&= \int_0^1 \frac{\phi^{\alpha'-1}(1-\phi)^{\beta-1}}{B(\alpha'-1,\beta)} d\phi && \text{(iv)} \\
&= \int_0^1 \frac{\alpha'-1}{\alpha'+\beta-1} \frac{\phi^{\alpha'-1}(1-\phi)^{\beta-1}}{B(\alpha',\beta)} d\phi && \text{(v)} \\
&= \frac{\alpha'-1}{\alpha'+\beta-1} \int_0^1 \frac{\Gamma(\alpha'+\beta)}{\Gamma(\alpha')\Gamma(\beta)} \phi^{\alpha'-1}(1-\phi)^{\beta-1} d\phi && \text{(vi)} \\
&= \frac{\alpha'-1}{\alpha'+\beta-1} * 1 && \text{(vii)} \\
&= \frac{\alpha}{\alpha+\beta} && (1.12)
\end{aligned}
$$

【途中式の途中式】

- (i)：$f(\phi) = \mathrm{Beta}(\phi|\alpha,\beta)$ より、置き換える。

- (ii)：正規化項はベータ関数の逆数であるため、$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} = \frac{1}{B(\alpha,\beta)}$ で置き換える。また $\phi * \phi^{\alpha-1} = \phi^{\alpha}$ に、ϕ の項をまとめる。
- (iii)：$\alpha = \alpha' - 1$ で置き換える。
- (iv)：ベータ関数の性質 $B(\alpha,\beta) = \frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$ より

$$\begin{aligned}
B(\alpha-1,\beta) &= \frac{(\alpha-2)!(\beta-1)!}{(\alpha+\beta-2)!} \\
\Rightarrow \frac{\alpha-1}{\alpha+\beta-1}B(\alpha-1,\beta) &= \frac{(\alpha-1)(\alpha-2)!(\beta-1)!}{(\alpha+\beta-1)(\alpha+\beta-2)!} = B(\alpha,\beta)
\end{aligned}$$

であることが分かる。これを利用して

$$\frac{1}{B(\alpha'-1,\beta)} = \frac{\alpha'-1}{\alpha'+\beta-1}\frac{1}{B(\alpha',\beta)}$$

で置き換える。

- (v)：定数の項を $\int$ の外に出す。また、$\frac{1}{B(\alpha'-1,\beta)}$ を正規化項の形に戻す。
- (vi)：パラメータが α', β のベータ分布について、取り得る全ての値を足し合わせたものであるため、$\int_0^1 \frac{\Gamma(\alpha'+\beta)}{\Gamma(\alpha')\Gamma(\beta)}\phi^{\alpha'-1}(1-\phi)^{\beta-1}d\phi = 1$ である。
- (vii)：$\alpha' = \alpha + 1$ に戻す。

・分散の導出

「ϕ の 2 乗の期待値」と「ϕ の期待値の 2 乗」との差が分散となる。そこでまずは、ϕ の 2 乗の期待値を求める。

$$\begin{aligned}
\mathbb{E}[\phi^2] &= \int_0^1 \phi^2 f(\phi) d\phi && \text{(i)} \\
&= \int_0^1 \phi^2 \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\phi^{\alpha-1}(1-\phi)^{\beta-1}d\phi && \text{(ii)} \\
&= \int_0^1 \frac{\phi^{\alpha+1}(1-\phi)^{\beta-1}}{B(\alpha,\beta)}d\phi && \text{(iii)} \\
&= \int_0^1 \frac{\phi^{\alpha'-1}(1-\phi)^{\beta-1}}{B(\alpha'-2,\beta)}d\phi && \text{(iv)} \\
&= \int_0^1 \frac{(\alpha'-2)(\alpha'-1)}{(\alpha'+\beta-2)(\alpha'+\beta-1)}\frac{\phi^{\alpha'-1}(1-\phi)^{\beta-1}}{B(\alpha',\beta)}d\phi && \text{(v)} \\
&= \frac{(\alpha'-2)(\alpha'-1)}{(\alpha'+\beta-2)(\alpha'+\beta-1)}\int_0^1 \frac{\Gamma(\alpha'+\beta)}{\Gamma(\alpha')\Gamma(\beta)}\phi^{\alpha'-1}(1-\phi)^{\beta-1}d\phi && \text{(vi)} \\
&= \frac{(\alpha'-2)(\alpha'-1)}{(\alpha'+\beta-2)(\alpha'+\beta-1)} * 1 && \text{(vii)} \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}
\end{aligned}$$

【途中式の途中式】

- (i)：$f(\phi) = \text{Beta}(\phi|\alpha,\beta)$ より、置き換える。

- (ii)：正規化項はベータ関数の逆数であるため、$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}=\frac{1}{B(\alpha,\beta)}$ で置き換える。また、$\phi^2 * \phi^{\alpha-1}=\phi^{\alpha+1}$ となる。
- (iii)：$\alpha+1=\alpha'-1$ で置き換える。
- (iv)：ベータ関数の性質 $B(\alpha,\beta)=\frac{(\alpha-1)!(\beta-1)!}{(\alpha+\beta-1)!}$ より

$$\frac{(\alpha-1)(\alpha-2)}{(\alpha+\beta-1)(\alpha+\beta-2)}B(\alpha-2,\beta)=\frac{(\alpha-1)(\alpha-2)}{(\alpha+\beta-1)(\alpha+\beta-2)}\frac{(\alpha-3)!(\beta-1)!}{(\alpha+\beta-3)!}=B(\alpha,\beta)$$

を用いて置き換える。

- (v)：定数の項を $\int$ の外に出す。また、$\frac{1}{B(\alpha',\beta)}$ を正規化項の形に戻す。
- (vi)：パラメータが α',β のベータ分布について、取り得る全ての値を足し合わせたものであるため、$\int_0^1 \frac{\Gamma(\alpha'+\beta)}{\Gamma(\alpha')\Gamma(\beta)}\phi^{\alpha'-1}(1-\phi)^{\beta-1}d\phi=1$ である。
- (vii)：$\alpha'=\alpha+2$ に戻す。

ϕ の 2 乗の期待値と ϕ の期待値の 2 乗との差を求める。

$$\begin{aligned}
\mathrm{Var}[\phi] &= \mathbb{E}[\phi^2]-(\mathbb{E}[\phi])^2 \\
&= \frac{\alpha(\alpha+1)}{(\alpha+\beta)(\alpha+\beta+1)}-\left(\frac{\alpha}{\alpha+\beta}\right)^2 \\
&= \frac{\alpha(\alpha+1)*(\alpha+\beta)}{(\alpha+\beta)(\alpha+\beta+1)*(\alpha+\beta)}-\frac{\alpha^2*(\alpha+\beta+1)}{(\alpha+\beta)^2*(\alpha+\beta+1)} \\
&= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}
\end{aligned}$$

・最頻値の導出

最頻値は曲線 (グラフ) の頂点の値であることから、$\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\phi^{\alpha-1}(1-\phi)^{\beta-1}$ を微分して、$\frac{\partial f(\phi)}{\partial\phi}=0$ となる点である。

$$\begin{aligned}
f(\phi) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\phi^{\alpha-1}(1-\phi)^{\beta-1} && \text{(i)} \\
\frac{\partial f(\phi)}{\partial\phi} &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\{(\alpha-1)\phi^{\alpha-2}(1-\phi)^{\beta-1}-\phi^{\alpha-1}(\beta-1)(1-\phi)^{\beta-2}\} && \text{(ii)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\{(\alpha-1)\phi^{\alpha-2}(1-\phi)*(1-\phi)^{\beta-2}-\phi*\phi^{\alpha-2}(\beta-1)(1-\phi)^{\beta-2}\} && \text{(iii)} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\phi^{\alpha-2}(1-\phi)^{\beta-2}\{(\phi-1)(\alpha-1)-\phi(\beta-1)\}
\end{aligned}$$

【途中式の途中式】

- (i)：微分する。
  - $\phi^{\alpha-1}$ を $f(\phi)$、$(1-\phi)^{\beta-1}$ を $g(\phi)$ として、積の微分 $\{f(\phi)g(\phi)\}'=f'(\phi)g(\phi)+f(\phi)g'(\phi)$ をする。
  - ただし $(1-\phi)^{\beta-1}$ の微分は $(1-\phi)$ を $g(\phi)$ として、合成関数の微分 $\{f(g(\phi))\}'=f'(g(\phi))*g'(\phi)$ をする。

$$
\begin{aligned}
\{\phi^{\alpha-1}(1-\phi)^{\beta-1}\}' &= (\phi^{\alpha-1})'(1-\phi)^{\beta-1} + \phi^{\alpha-1}\{(1-\phi)^{\beta-1}\}' \\
&= (\alpha-1)\phi^{\alpha-2}(1-\phi)^{\beta-1} + \phi^{\alpha-1}(\beta-1)(1-\phi)^{\beta-2} * (1-\phi)' \\
&= (\alpha-1)\phi^{\alpha-2}(1-\phi)^{\beta-1} + \phi^{\alpha-1}(\beta-1)(1-\phi)^{\beta-2} * (-1) \\
&= (\alpha-1)\phi^{\alpha-2}(1-\phi)^{\beta-1} - \phi^{\alpha-1}(\beta-1)(1-\phi)^{\beta-2}
\end{aligned}
$$

- (ii)：$(1-\phi)^{\beta-1} = (1-\phi) * (1-\phi)^{\beta-2}$、$\phi^{\alpha-1} = \phi * \phi^{\alpha-2}$ に、それぞれ項を分割する。
- (iii)：$\phi^{\alpha-2}$ と $(1-\phi)^{\beta-2}$ を取り出して、式を整理する。

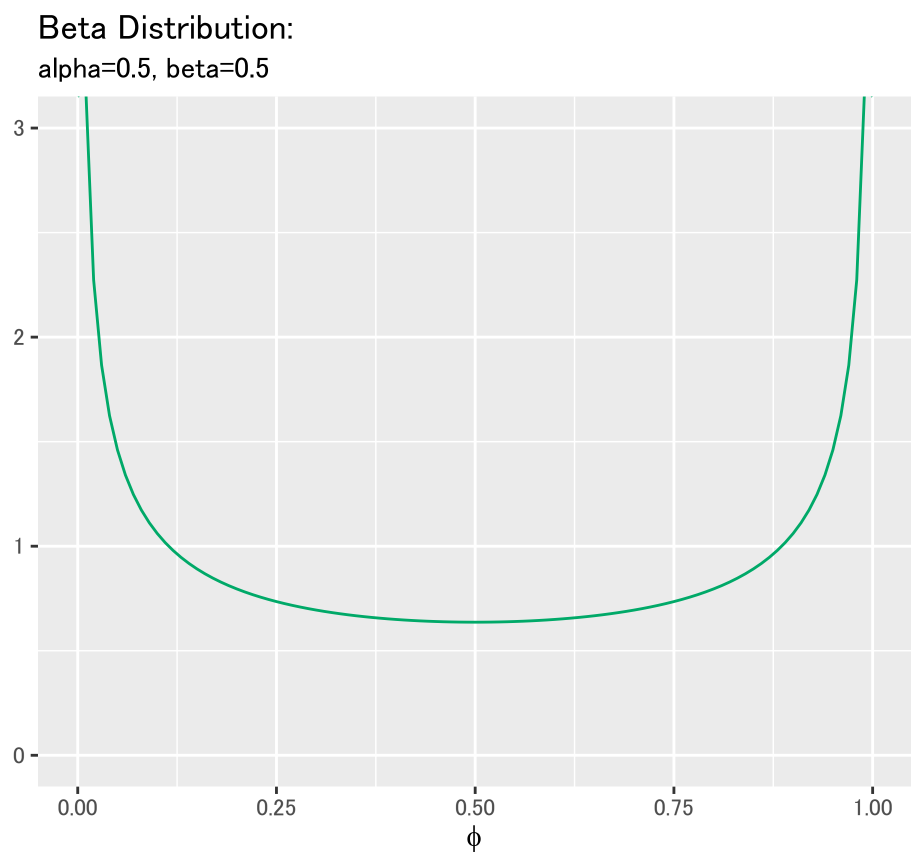
$\frac{\partial f(\phi)}{\partial \phi} = 0$ となる ϕ を求める。

$$
\begin{aligned}
\frac{\partial f(\phi)}{\partial \phi} = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\phi^{\alpha-2}(1-\phi)^{\beta-2}\{(\phi-1)(\alpha-1) - \phi(\beta-1)\} &= 0 \\
(1-\phi)(\alpha-1) - \phi(\beta-1) &= 0 \\
\alpha - 1 - \alpha\phi + \phi - \beta\phi + \phi &= 0 \\
-(\alpha+\beta-2)\phi &= -(\alpha-1) \\
\phi &= \frac{\alpha-1}{\alpha+\beta-2} = \mathrm{mode}[\phi] \qquad (\alpha > 1, \beta > 1)
\end{aligned}
$$

・分布の確認

・$\alpha = 0.5, \beta = 0.5$ のときの分布

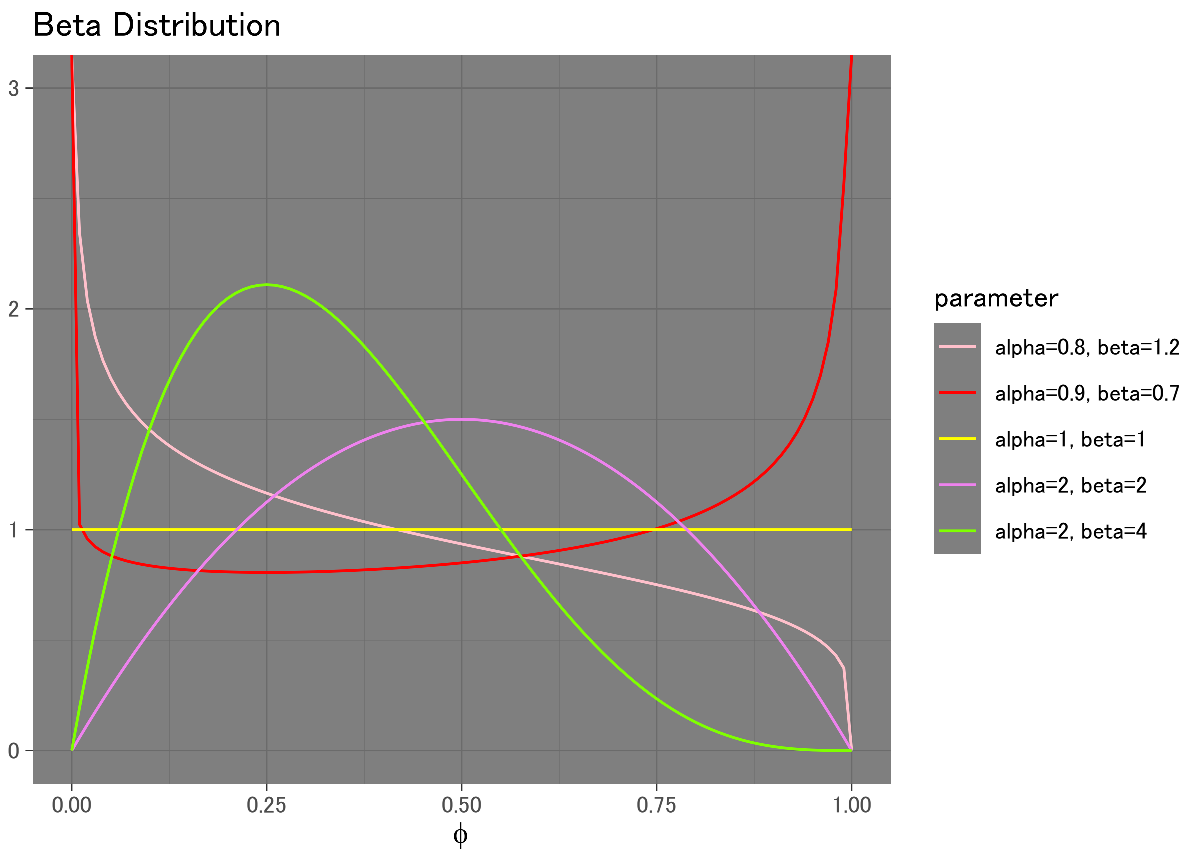
```
# 利用パッケージ
library(tidyverse)

# パラメーターを指定
val_alpha <- 0.5   # 任意の値を指定する
val_beta  <- 0.5   # 任意の値を指定する

# 作図
tibble(
  phi = seq(0, 1, 0.01),  # phi(x軸)の値の設定
  y = (1 / beta(val_alpha, val_beta)) * phi ^ (val_alpha - 1) * (1 - phi) ^ (val_beta - 1)
) %>%  # y軸の値を算出
  ggplot(mapping = aes(x = phi, y = y)) +  # データ
    geom_line(color = "#00A968") +         # 折れ線グラフ
    coord_cartesian(ylim = c(0, 3)) +      # y軸の範囲
    labs(title = "Beta Distribution:",     # タイトル
         subtitle = paste0("alpha=", val_alpha, ", beta=", val_beta),  # サブタイトル
         x = expression(phi), y = "")      # 軸ラベル
```



・パラメータの違いによる分布の変化

```
# パラメーターの指定
val_alpha <- c(1, 2, 2, 0.9, 0.8) # 任意の値を指定する
val_beta  <- c(1, 2, 4, 0.7, 1.2) # 任意の値を指定する

# 数値計算
beta_df <- data.frame()
for(i in seq_along(val_alpha)) {
  # 数値計算
  tmp_df <- tibble(
    phi = seq(0, 1, 0.01),  # phi(x軸)の値の設定
    y = (1 / beta(val_alpha[i], val_beta[i])) * phi ^ (val_alpha[i] - 1) * (1 - phi) ^ (val_beta[i] - 1
    parameter = paste0("alpha=", val_alpha[i], ", beta=", val_beta[i])
  )

  # 計算結果を結合
  beta_df <- rbind(beta_df, tmp_df)
}

# 描画
ggplot(data = beta_df, mapping = aes(x = phi, y = y, color = parameter)) + # データ
  geom_line() +                        # 折れ線グラフ
  theme_dark() +                       # 背景色
  ylim(c(0, 3)) +                      # y軸の範囲
  labs(title = "Beta Distribution",    # タイトル
```

```r
    x = expression(phi), y = "")  # 軸ラベル
```



1.2.4 ディリクレ分布

・ディリクレ分布

ディリクレ分布は、カテゴリ分布・多項分布のパラメータ $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_V), \phi_v \geq 0, \sum_{v=0}^{V} \phi_v = 1$ を生成するための確率分布で、共役事前分布とも呼ばれる。

ディリクレ分布はパラメータ

$$\boldsymbol{\beta} = (\beta_1, \beta_2, \cdots, \beta_V),\ \beta_v > 0$$

を持つ。共役事前分布のパラメータのことをハイパーパラメータとも呼ぶ。ディリクレ分布

$$\mathrm{Dirichlet}(\boldsymbol{\phi}|\boldsymbol{\beta}) = \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \prod_{v=1}^{V} \phi_v^{\beta_v - 1}$$

によって、カテゴリ分布・多項分布のパラメータ $\boldsymbol{\phi}$ が求まる。

またパラメータ β が一様であるとき、ディリクレ分布は

$$\text{Dirichlet}(\boldsymbol{\phi}|\beta\cdots\beta) = \frac{\Gamma(\beta V)}{\Gamma(\beta)^V}\prod_{v=1}^{V}\phi_v^{\beta-1}$$

となる。

ディリクレ分布が $V=2$ のとき $\phi_2 = 1-\phi_1$ であり、また $\alpha=\beta_1, \beta=\beta_2$ とおくと

$$\begin{aligned}
\text{Dirichlet}(\boldsymbol{\phi}|\beta_1,\beta_2) &= \frac{\Gamma(\sum_{v=1}^{2}\beta_v)}{\prod_{v=1}^{2}\Gamma(\beta_v)}\prod_{v=1}^{2}\phi_v^{\beta_v-1} \\
&= \frac{\Gamma(\beta_1+\beta_2)}{\Gamma(\beta_1)*\Gamma(\beta_2)}\phi_1^{\beta_1-1}*\phi_2^{\beta_2-1} \\
&= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}\phi_1^{\alpha-1}(1-\phi_1)^{\beta-1} = \text{Beta}(\phi_1|\alpha,\beta)
\end{aligned}$$

ベータ分布と等しくなることが確認できる。

次は、この式の正規化項について確認する。

・正規化項の導出

ベータ分布のときと同様に、ベータ関数 $B(\alpha,\beta)=\int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx$ を $I(\alpha,\beta)=\int_0^p x^{\alpha-1}(p-x)^{\beta-1}dx$ として利用し、正規化項を導出していく。

$$
\begin{aligned}
I(\alpha,\beta) &= \int_0^p x^{\alpha-1}(p-x)^{\beta-1}dx && \text{(i)}\\
&= \left[\frac{1}{\alpha}x^{\alpha}(p-x)^{\beta-1}\right]_0^p - \int_0^p -\frac{1}{\alpha}x^{\alpha}(\beta-1)(p-x)^{\beta-2}dx && \text{(ii)}\\
&= 0 + \frac{\beta-1}{\alpha}\int_0^p x^{\alpha}(p-x)^{\beta-2}dx && \text{(iii)}\\
&= \frac{\beta-1}{\alpha}I(\alpha+1,\beta-1) && \text{(iv)}\\
&= \frac{\beta-1}{\alpha}\frac{\beta-2}{\alpha+1}I(\alpha+2,\beta-2) && \text{(v)}\\
&= \frac{\beta-1}{\alpha}\frac{\beta-2}{\alpha+1}\frac{\beta-3}{\alpha+2}\cdots\frac{2}{\alpha+\beta-3}\frac{1}{\alpha+\beta-2}I(\alpha+\beta-1,1) && \text{(vi)}\\
&= \frac{\Gamma(\beta)}{\alpha(\alpha+1)\cdots(\alpha+\beta-2)}I(\alpha+\beta-1,1) && \text{(vii)}\\
&= \frac{\Gamma(\beta)}{\alpha(\alpha+1)\cdots(\alpha+\beta-2)}\int_0^p x^{\alpha+\beta-2}dx && \text{(viii)}\\
&= \frac{\Gamma(\beta)}{\alpha(\alpha+1)\cdots(\alpha+\beta-2)}\frac{1}{\alpha+\beta-1}p^{\alpha+\beta-1} - 0 && \text{(ix)}\\
&= \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}p^{\alpha+\beta-1} && (1)
\end{aligned}
$$

【途中式の途中式】

- (i)：部分積分 $\int_a^b f'(x)g(x)dx = [f(x)g(x)]_a^b - \int_a^b f(x)g'(x)dx$ する。
  - $f'(x)$ にあたる $x^{\alpha-1}$ を積分して、$\frac{1}{\alpha}x^{\alpha}$ となる。
  - $g(x)$ にあたる $(p-x)^{\beta-1}$ は合成関数の微分 $\{f(g(x))\}' = f'(g(x)) * g'(x)$ をして、$(\beta-1)(p-x)^{\beta-2} * (-x^0) = -(\beta-1)(p-x)^{\beta-2}$ となる。
- (ii)：
  - $x=0$ のとき $x^{\alpha}=0$、また $x=p$ のとき $(p-x)=0$ なので、$[\frac{1}{\alpha}x^{\alpha}(p-x)^{\beta-1}]_0^p = 0-0=0$ になる。
  - $-\frac{1}{\alpha}$ と $\beta-1$ を $\int$ の外に出して、式を整理する。
- (iii)：$I(\alpha,\beta) = \int_0^p x^{\alpha-1}(p-x)^{\beta-1}dx$ なのだから、$I(\alpha+1,\beta-1) = \int_0^p x^{\alpha}(p-x)^{\beta-2}dx$ になる。
- (iv)：$I(\alpha,\beta) = \frac{\beta-1}{\alpha}I(\alpha+1,\beta-1)$ であることが分かったので、$I(\alpha+1,\beta-1) = \frac{\beta-2}{\alpha+1}I(\alpha+2,\beta-2)$ になる。
- (v)：同様に、$\frac{1}{\alpha+\beta-1}I(\alpha+\beta-1,1)$ となるまで ($\beta-1$ 回) 繰り返す。
- (vi)：分子をみると、$\beta-1$ から 1 までの自然数を掛け合わせているので、$(\beta-1)!$ である。また $(\beta-1)! = \Gamma(\beta)$ であるため、置き換える。
- (vii)：$I(\alpha+\beta-1,1) = \int_0^p x^{\alpha+\beta-2}(p-x)^0dx$ である。
- (viii)：積分すると、$\int_0^p x^{\alpha+\beta-2}dx = \frac{1}{\alpha+\beta-1}p^{\alpha+\beta-1} - \frac{1}{\alpha+\beta-1}0^{\alpha+\beta-1}$ である。
- (ix)：
  - 分母をみると、α から $\alpha+\beta-1$ までの自然数を掛け合わせている。ここに $(\alpha-1)!$ を掛けることで、$\alpha+\beta-1$ から 1 までの自然数を掛け合わせることになる。
  - つまり、分母分子に $\Gamma(\alpha) = (\alpha-1)!$ を掛けることで $\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$ となる。

・正規化項の導出

$$
\begin{aligned}
&\int_0^1 \prod_{v=1}^{V} \phi_v^{\beta_v - 1} d\boldsymbol{\phi} && \text{(i)}\\
&= \int_0^1 \phi_1^{\beta_1-1} d\phi_1 * \int_0^{1-\phi_1} \phi_2^{\beta_2-1} d\phi_2 * \int_0^{1-\phi_1-\phi_2} \phi_3^{\beta_3-1} d\phi_3 \\
&\quad * \cdots * \int_0^{1-\phi_1-\cdots-\phi_{V-3}} \phi_{V-2}^{\beta_{V-2}-1} d\phi_{V-2} * \int_0^{1-\phi_1-\cdots-\phi_{V-2}} \phi_{V-1}^{\beta_{V-1}-1} * \phi_V^{\beta_V-1} d\phi_{V-1}\phi_V && \text{(ii)}\\
&= \int_0^1 \phi_1^{\beta_1-1} d\phi_1 * \int_0^{1-\phi_1} \phi_2^{\beta_2-1} d\phi_2 * \int_0^{1-\phi_1-\phi_2} \phi_3^{\beta_3-1} d\phi_3 \\
&\quad * \cdots * \int_0^{1-\phi_1-\cdots-\phi_{V-3}} \phi_{V-2}^{\beta_{V-2}-1} d\phi_{V-1} * \int_0^{1-\phi_1-\cdots-\phi_{V-2}} \phi_{V-1}^{\beta_{V-1}-1} * (1-\phi_1-\cdots-\phi_{V-2}-\phi_{V-1})^{\beta_V-1} d\phi_{V-1}
\end{aligned}
$$

【途中式の途中式】

- (i)：式を分割する。このとき積分する範囲が ϕ_v ずつ減っていく。
- (ii)：$\phi_V = 1 - \phi_1 - \cdots - \phi_{V-2} - \phi_{V-1}$ なので、置き換える。

最後の項は、$x = \phi_{V-1}$、$p = 1 - \phi_1 - \phi_2 - \cdots - \phi_{V-2}$、$\alpha = \beta_{V-1}$、$\beta = \beta_V$ として、(1) より

$$
\begin{aligned}
&\int_0^{1-\phi_1-\cdots-\phi_{V-2}} \phi_{V-1}^{\beta_{V-1}-1} (1-\phi_1-\cdots-\phi_{V-2}-\phi_{V-1})^{\beta_V-1} d\phi_{V-1} \\
&= \frac{\Gamma(\beta_V)\Gamma(\beta_{V-1})}{\Gamma(\beta_V+\beta_{V-1})} (1-\phi_1-\cdots-\phi_{V-2})^{\beta_V+\beta_{V-1}-1}
\end{aligned}
$$

と変形できる。これをさらに最後の第 2 項と合わせて

$$
\begin{aligned}
&\int_0^{1-\phi_1-\cdots-\phi_{V-3}} \phi_{V-2}^{\beta_{V-2}-1} * \frac{\Gamma(\beta_V)\Gamma(\beta_{V-1})}{\Gamma(\beta_V+\beta_{V-1})} (1-\phi_1-\cdots-\phi_{V-3}-\phi_{V-2})^{\beta_V+\beta_{V-1}-1} d\phi_{V-2} && \text{(i)}\\
&= \frac{\Gamma(\beta_V)\Gamma(\beta_{V-1})}{\Gamma(\beta_V+\beta_{V-1})} \int_0^{1-\phi_1-\cdots-\phi_{V-2}} \phi_{V-1}^{\beta_{V-1}-1} (1-\phi_1-\cdots-\phi_{V-3}-\phi_{V-2})^{\beta_V+\beta_{V-1}-1} d\phi_{V-2} && \text{(ii)}\\
&= \frac{\Gamma(\beta_V)\Gamma(\beta_{V-1})}{\Gamma(\beta_V+\beta_{V-1})} \frac{\Gamma(\beta_V+\beta_{V-1})\Gamma(\beta_{V-2})}{\Gamma(\beta_V+\beta_{V-1}+\beta_{V-2})} (1-\phi_1-\cdots-\phi_{V-3})^{\beta_V+\beta_{V-1}+\beta_{V-2}-1} && \text{(iii)}\\
&= \frac{\Gamma(\beta_V)\Gamma(\beta_{V-1})\Gamma(\beta_{V-2})}{\Gamma(\beta_V+\beta_{V-1}+\beta_{V-2})} (1-\phi_1-\cdots-\phi_{V-3})^{\beta_V+\beta_{V-1}+\beta_{V-2}-1}
\end{aligned}
$$

【途中式の途中式】

- (i)：定数の項を $\int$ の外に出す。
- (ii)：$x = \phi_{V-2}$、$p = 1 - \phi_1 - \cdots - \phi_{V-3}$、$\alpha = \beta_{V-2}$、$\beta = \beta_V + \beta_{V-1}$ として、(1) より変形する。
- (iii)：約分して、式を整理する。

となる。これを繰り返すと

$$\int_0^1 \prod_{v=1}^{V} \phi_v^{\beta_v - 1} d\boldsymbol{\phi} = \frac{\Gamma(\beta_V)\Gamma(\beta_{V-1})\Gamma(\beta_{V-2})\cdots\Gamma(\beta_2)\Gamma(\beta_1)}{\Gamma(\beta_V + \beta_{V-1} + \beta_{V-2} + \cdots + \beta_2 + \beta_1)}$$
$$= \frac{\prod_{v=1}^{V} \Gamma(\beta_v)}{\Gamma(\sum_{v=1}^{V} \beta_v)}$$

が得られる。この逆数を正規化項として用いることで、全ての事象を考慮した際の値が 1 となるため、確率分布として扱うことができるようになる。

次からは、ディリクレ分布の基本的な統計量を求めていく。

・期待値の導出

分布が取り得る値 ϕ_v と確率密度関数 $f(\phi_v)$ を掛け合わせたものを、0 から 1 の範囲で積分して期待値を求める。

$$\begin{aligned}
\mathbb{E}[\phi_v] &= \int_0^1 \phi f(\phi_v) d\phi_v && \text{(i)} \\
&= \int_0^1 \phi_v \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \prod_{v=1}^{V} \phi_v^{\beta_v - 1} d\phi_v && \text{(ii)} \\
&= \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \int_0^1 \phi_1^{\beta_1 - 1} \cdots \phi_v * \phi_v^{\beta_v - 1} \cdots \phi_V^{\beta_V - 1} d\phi_v && \text{(iii)} \\
&= \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \frac{\Gamma(\beta_1)\cdots\Gamma(\beta_v + 1)\cdots\Gamma(\beta_V)}{\Gamma(\beta_1 + \cdots + (\beta_v + 1) + \cdots + \beta_V)} && \text{(iv)} \\
&= \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \frac{\Gamma(\beta_1)\cdots\beta_v\Gamma(\beta_v)\cdots\Gamma(\beta_V)}{\Gamma(\beta_1 + \cdots + \beta_v + \cdots + \beta_V + 1)} && \text{(v)} \\
&= \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \frac{\beta_v \prod_{v=1}^{V} \Gamma(\beta_v)}{\Gamma(\sum_{v=1}^{V} \beta_v + 1)} && \text{(vi)} \\
&= \frac{\Gamma(\sum_{v=1}^{V} \beta_v)}{\prod_{v=1}^{V} \Gamma(\beta_v)} \frac{\beta_v \prod_{v=1}^{V} \Gamma(\beta_v)}{(\sum_{v=1}^{V} \beta_v)\Gamma(\sum_{v=1}^{V} \beta_v)} && \text{(vii)} \\
&= \frac{\beta_v}{\hat{\beta}} && (1.15)
\end{aligned}$$

【途中式の途中式】

- (i)：$f(\phi) = \text{Dirichlet}(\boldsymbol{\phi}|\boldsymbol{\beta})$ より、置き換える。
- (ii)：定数の項を $\int$ の外に出し、$\prod_{v=1}^{V}$ を展開して、ϕ_v の項をまとめる。
- (iii)：正規化項の導出より、変形する。ただし β_v の項は $\phi_v^{\beta_v}$ なので、$\beta_v + 1$ になる。
- (iv)：ガンマ関数の性質 $\Gamma(x + 1) = (x)\Gamma(x)$ より、$\Gamma(\beta_v + 1) = \beta_v\Gamma(\beta_v)$ であるため、置き換える。
- (v)：$\prod_{v=1}^{V}, \sum_{v=1}^{V}$ に、それぞれ項をまとめる。
- (vi)：(iii) と同様に、$\Gamma(\sum_{v=1}^{V} \beta_v + 1) = (\sum_{v=1}^{V} \beta_v)\Gamma(\sum_{v=1}^{V} \beta_v)$ であるため、置き換える。
- (vii)：$\hat{\beta} = \sum_{v=1}^{V} \beta_v$ とおき、約分して式を整理する。

・分散の導出

「ϕ_v の 2 乗の期待値」と「ϕ_v の期待値の 2 乗」との差が分散となる。そこでまずは、ϕ_v の 2 乗の期待値を求める。

$$
\begin{aligned}
\mathbb{E}[\phi_v^2] &= \int \phi_v^2 f(\phi) d\phi_v && \text{(i)} \\
&= \int_0^1 \phi_v^2 \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \prod_{v=1}^V \phi_v^{\beta_v - 1} d\phi_v && \text{(ii)} \\
&= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \int_0^1 \phi_1^{\beta_1 - 1} \cdots \phi_v^2 * \phi_v^{\beta_v - 1} \cdots \phi_V^{\beta_V - 1} d\phi_v && \text{(iii)} \\
&= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \frac{\Gamma(\beta_1) \cdots \Gamma(\beta_v + 2) \cdots \Gamma(\beta_V)}{\Gamma(\beta_1 + \cdots + (\beta_v + 2) + \cdots + \beta_V)} && \text{(iv)} \\
&= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \frac{\Gamma(\beta_1) \cdots \beta_v(\beta_v + 1)\Gamma(\beta_v) \cdots \Gamma(\beta_V)}{\Gamma(\beta_1 + \cdots + \beta_v + \cdots + \beta_V + 2)} && \text{(v)} \\
&= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \frac{\beta_v(\beta_v + 1) \prod_{v=1}^V \Gamma(\beta_v)}{\Gamma(\sum_{v=1}^V \beta_v + 2)} && \text{(vi)} \\
&= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \frac{\beta_v(\beta_v + 1) \prod_{v=1}^V \Gamma(\beta_v)}{(\sum_{v=1}^V \beta_v)(\sum_{v=1}^V \beta_v + 1)\Gamma(\sum_{v=1}^V \beta_v)} && \text{(vii)} \\
&= \frac{\beta_v(\beta_v + 1)}{\hat{\beta}(\hat{\beta} + 1)}
\end{aligned}
$$

【途中式の途中式】

- (i)：$f(\phi) = \mathrm{Dirichlet}(\boldsymbol{\phi}|\boldsymbol{\beta})$ より、置き換える。
- (ii)：定数の項を $\int$ の外に出し、$\prod_{v=1}^V$ を展開して、ϕ_v の項をまとめる。
- (iii)：正規化項の導出より、変形する。ただし β_v の項は $\phi_v^{\beta_v+1}$ なので、$\beta_v + 2$ になる。
- (iv)：ガンマ関数の性質より、$\Gamma(\beta_v + 2) = (\beta_v + 1)\Gamma(\beta_v + 1) = (\beta_v + 1)\beta_v\Gamma(\beta_v)$ であるため、置き換える。
- (v)：$\prod_{v=1}^V, \sum_{v=1}^V$ に、それぞれ項をまとめる。
- (vi)：(iv) と同様に、$\Gamma(\sum_{v=1}^V \beta_v + 2) = (\sum_{v=1}^V \beta_v + 1)\Gamma(\sum_{v=1}^V \beta_v + 1) = (\sum_{v=1}^V \beta_v + 1)(\sum_{v=1}^V \beta_v)\Gamma(\sum_{v=1}^V \beta_v)$ であるため、置き換える。
- (vii)：$\hat{\beta} = \sum_{v=1}^V \beta_v$ とおき、約分して式を整理する。

ϕ_v の 2 乗の期待値と ϕ_v の期待値の 2 乗との差を求める。

$$
\begin{aligned}
\mathrm{Var}[\phi_v] &= \mathbb{E}[\phi_v^2] - \mathbb{E}[\phi_v]^2 \\
&= \frac{\beta_v(\beta_v+1)}{\hat{\beta}(\hat{\beta}+1)} - \left(\frac{\beta_v}{\hat{\beta}}\right)^2 \\
&= \frac{\beta_v(\beta_v+1) * \hat{\beta}}{\hat{\beta}(\hat{\beta}+1) * \hat{\beta}} - \frac{\beta_v^2 * (\hat{\beta}+1)}{\hat{\beta}^2 * (\hat{\beta}+1)} \\
&= \frac{\beta_v^2\hat{\beta} + \beta_v\hat{\beta} - \beta_v^2\hat{\beta} - \beta_v^2}{\hat{\beta}^2(\hat{\beta}+1)} \\
&= \frac{\beta_v(\hat{\beta}-\beta_v)}{\hat{\beta}^2(\hat{\beta}+1)}
\end{aligned} \tag{1.16}
$$

$V = 2$ のとき $\alpha = \beta_1, \beta = \beta_2$ とすると、ディリクレ分布の期待値 $\frac{\beta_1}{\beta_1+\beta_2} = \frac{\alpha}{\alpha+\beta}$、分散は $\frac{\beta_1(\beta_1+\beta_2-\beta_1)}{(\beta_1+\beta_2)^2(\beta_1+\beta_2+1)} = \frac{\alpha\beta}{(\alpha+\beta)^2\alpha+\beta+1}$ となり、ベータ分布の期待値と分散とそれぞれ等しくなることが確認できる。

・共分散の導出

$v \neq v'$ である $\phi_v, \phi_{v'}$ の共分散を、分散と同様の手順で求めていく。

$$
\begin{aligned}
\mathbb{E}[\phi_v\phi_{v'}] &= \int_0^1 \phi_v\phi_v' \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \prod_{v=1}^V \phi_v^{\beta_v-1} d\phi_v \\
&= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \int_0^1 \phi_1^{\beta_1-1} \cdots \phi_v * \phi_v^{\beta_v-1} \cdots \phi_{v'} * \phi_{v'}^{\beta_{v'}-1} \cdots \phi_V^{\beta_V-1} d\phi_v \\
&= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \frac{\Gamma(\beta_1)\cdots\Gamma(\beta_v+1)\cdots\Gamma(\beta_{v'}+1)\cdots\Gamma(\beta_V)}{\Gamma(\beta_1+\cdots+(\beta_v+1)+\cdots+(\beta_{v'}+1)+\cdots+\beta_V)} \\
&= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \frac{\Gamma(\beta_1)\cdots\beta_v\Gamma(\beta_v)\cdots\beta_{v'}\Gamma(\beta_{v'})\cdots\Gamma(\beta_V)}{\Gamma(\beta_1+\cdots+\beta_v+\cdots+\beta_{v'}+\cdots+\beta_V+2)} \\
&= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \frac{\beta_v\beta_{v'}\prod_{v=1}^V \Gamma(\beta_v)}{\Gamma(\sum_{v=1}^V \beta_v+2)} \\
&= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \frac{\beta_v\beta_{v'}\prod_{v=1}^V \Gamma(\beta_v)}{(\sum_{v=1}^V \beta_v)(\sum_{v=1}^V \beta_v+1)\Gamma(\sum_{v=1}^V \beta_v)} \\
&= \frac{\beta_v\beta_{v'}}{\hat{\beta}(\hat{\beta}+1)}
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Cov}[\phi_v, \phi_{v'}] &= \mathbb{E}[\phi_v \phi_{v'}] - \mathbb{E}[\phi_v]\mathbb{E}[\phi_{v'}] \\
&= \frac{\beta_v \beta_{v'}}{\hat{\beta}(\hat{\beta}+1)} - \frac{\beta_v}{\hat{\beta}} \frac{\beta_{v'}}{\hat{\beta}} \\
&= \frac{\beta_v \beta_{v'} * \hat{\beta}}{\hat{\beta}(\hat{\beta}+1) * \hat{\beta}} - \frac{\beta_v \beta_{v'} * (\hat{\beta}+1)}{\hat{\beta}^2 * (\hat{\beta}+1)} \\
&= \frac{\hat{\beta}\beta_v \beta_{v'} - \hat{\beta}\beta_v \beta_{v'} - \beta_v \beta_{v'}}{\hat{\beta}^2(\hat{\beta}+1)} \\
&= -\frac{\beta_v \beta_{v'}}{\hat{\beta}^2(\hat{\beta}+1)}
\end{aligned}
$$

・最頻値の導出

最頻値は曲線 (グラフ) の頂点の値であることから、$\frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \prod_{v=1}^V \phi_v^{\beta_v - 1}$ を微分して、$\frac{\partial f(\phi)}{\partial \phi_v} = 0$ となる点である。と思うんだが解けんかった…

$$
\begin{aligned}
f(\phi_v) &= \frac{\Gamma(\sum_{v=1}^V \beta_v)}{\prod_{v=1}^V \Gamma(\beta_v)} \prod_{v=1}^V \phi_v^{\beta_v - 1} \\
\frac{\partial f(\phi)}{\partial \phi_v} &= \\
&= \cdots
\end{aligned}
$$

$\frac{\partial f(\phi_v)}{\partial \phi_v} = 0$ となる ϕ_v を求める。

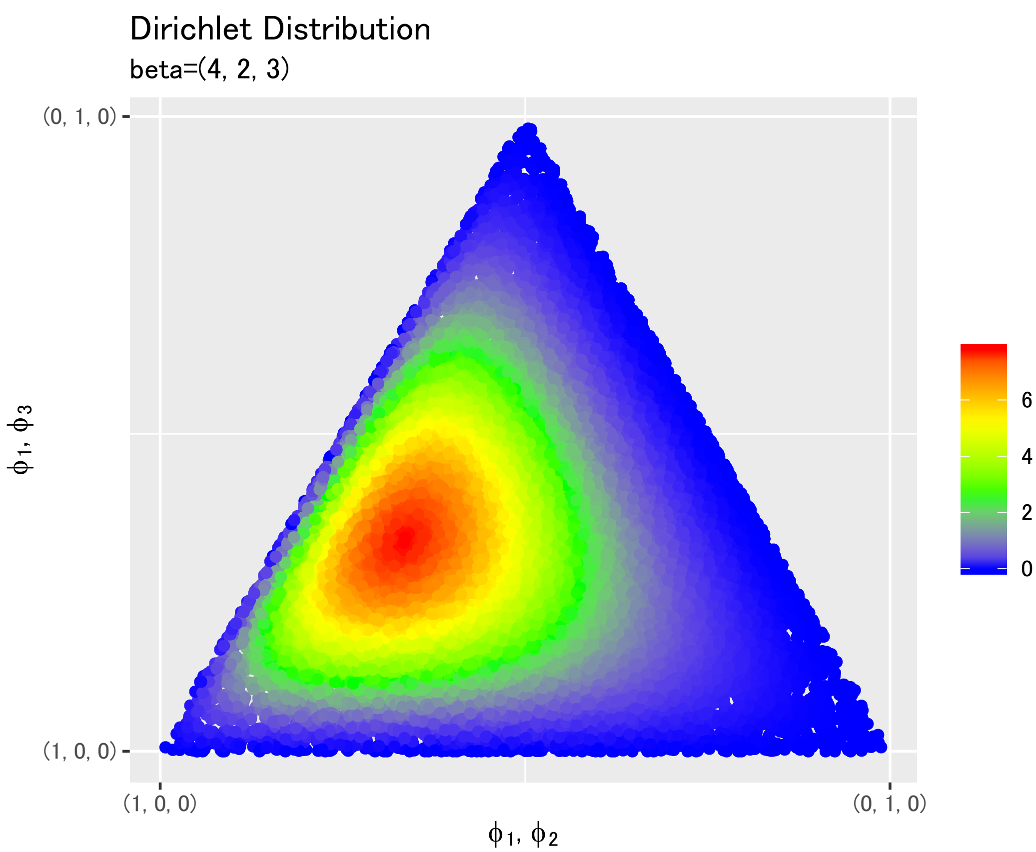
$$
\begin{aligned}
\frac{\partial f(\phi_v)}{\partial \phi_v} &= \cdots = 0 \\
&= \\
\phi_v(\beta_1 - 1 + \beta_2 - 1 + \cdots + \beta_V - 1) &= \beta_v - 1 \\
\phi_v &= \frac{\beta_v - 1}{\hat{\beta} - V} = \mathrm{mode}[\phi_v] \qquad (\beta_v > 1)
\end{aligned}
$$

・分布の確認

・$\beta_1 = 4, \beta_2 = 2, \beta_3 = 3$ のときの分布

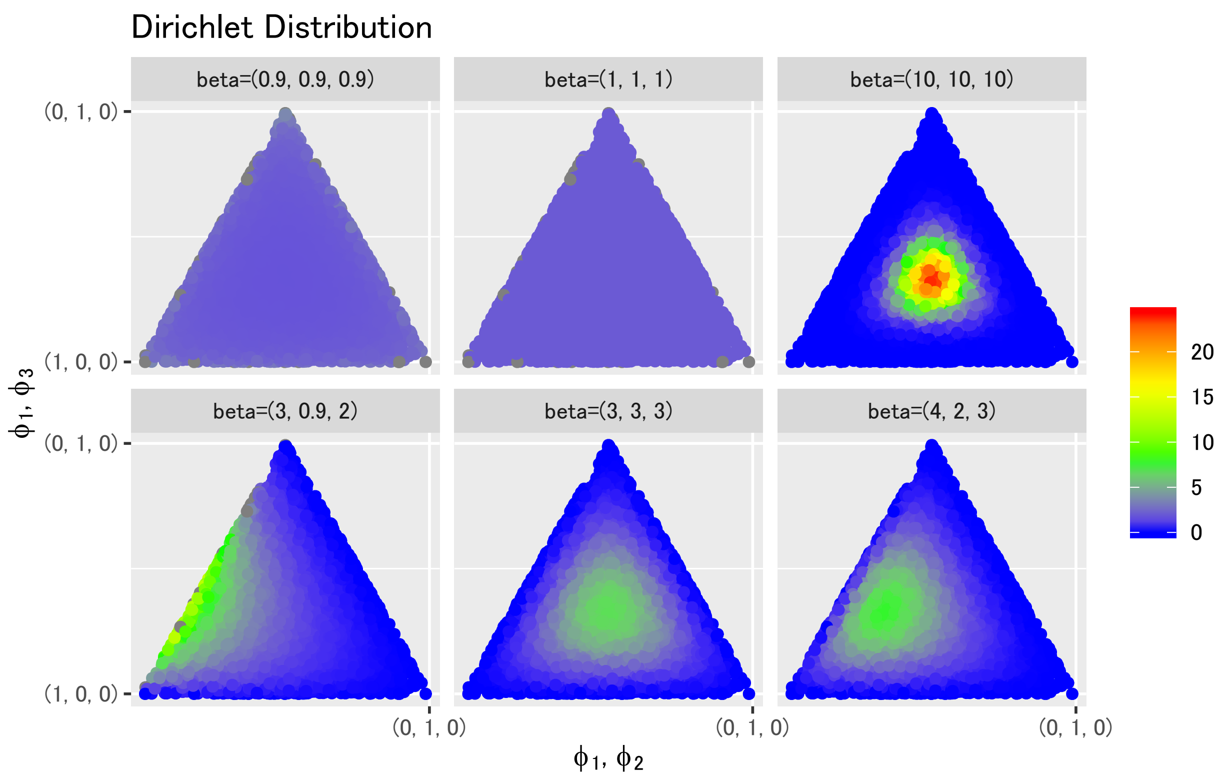
```r
# 利用パッケージ
library(tidyverse)

# パラメーターを指定
beta_v <- c(4, 2, 3) # 任意の値を指定する

# ランダムな値を生成
phi <- matrix(sample(seq(0, 1, 0.001), 90000, replace = TRUE), nrow = 3)

# 正規化
for(i in 1:ncol(phi)) {
  phi[, i] <- phi[, i] / sum(phi[, i])
}
```

```
# 正規化項を計算（対数）
tmp_beta <- lgamma(sum(beta_v)) - sum(lgamma(beta_v))

# 分布（対数）
tmp_phi  <- apply((beta_v - 1) * log(phi), 2, sum)

# 計算結果のデータフレームを作成
resu_df <- data.frame(x = phi[2, ] + (phi[3, ] / 2),  # 三角座標への変換
                      y = sqrt(3) * (phi[3, ] / 2),    # 三角座標への変換
                      z = exp(tmp_beta + tmp_phi))     # phi を算出

# 描画
ggplot(data = resu_df, mapping = aes(x = x, y = y, color = z)) +        # データ
  geom_point() +                                                        # 散布図
  scale_color_gradientn(colors = c("blue", "green", "yellow", "red")) + # プロットの色
  scale_x_continuous(breaks = c(0, 1),
                     labels = c("(1, 0, 0)", "(0, 1, 0)")) +            # x 軸の範囲
  scale_y_continuous(breaks = c(0, 0.87),
                     labels = c("(1, 0, 0)", "(0, 1, 0)")) +            # y 軸の範囲
  coord_fixed(ratio = 1) +                                              # グラフの縦横比
  labs(title = "Dirichlet Distribution",                                # タイトル
       subtitle = paste0("beta=(", beta_v[1], ", ", beta_v[2], ", ", beta_v[3], ")"),
       x = expression(paste(phi[1], ", ", phi[2], sep = "")),           # x 軸ラベル
       y = expression(paste(phi[1], ", ", phi[3], sep = "")),           # y 軸ラベル
       color = "")
```



・パラメータの違いによる分布の変化

```r
# パラメーターを指定
beta_df <- data.frame(beta1 = c(1, 1, 1),
                      beta2 = c(0.9, 0.9, 0.9),
                      beta3 = c(3, 3, 3),
                      beta4 = c(10, 10, 10),
                      beta5 = c(4, 2, 3),
                      beta6 = c(3, 0.9, 2)) # 任意の値を指定する

# ランダムな値を生成
phi <- matrix(sample(seq(0, 1, 0.001), 30000, replace = TRUE), nrow = 3)

# 正規化
for(i in 1:ncol(phi)) {
  phi[, i] <- phi[, i] / sum(phi[, i])
}

# 数値計算
resu_df <- data.frame()
for(i in 1:ncol(beta_df)){
  # 正規化項を計算（対数）
  tmp_beta <- lgamma(sum(beta_df[, i])) - sum(lgamma(beta_df[, i]))

  # 分布（対数）
  tmp_phi  <- apply((beta_df[, i] - 1) * log(phi), 2, sum)

  # 計算結果のデータフレームを作成
  tmp_df <- tibble(
    x = phi[2, ] + (phi[3, ] / 2),  # 三角座標への変換
    y = sqrt(3) * (phi[3, ] / 2),  # 三角座標への変換
    z = exp(tmp_beta + tmp_phi),    # phi を算出
    parameter = paste0("beta=(", beta_df[1, i], ", ", beta_df[2, i], ", ", beta_df[3, i], ")")
  )

  # 計算結果を結合
  resu_df <- rbind(resu_df, tmp_df)
}

# 描画
ggplot(data = resu_df, mapping = aes(x = x, y = y, color = z)) +        # データ
  geom_point() +                                                        # 散布図
  scale_color_gradientn(colors = c("blue", "green", "yellow", "red")) + # プロットの色
  scale_x_continuous(breaks = 1, labels = "(0, 1, 0)") +                # x 軸の範囲
  scale_y_continuous(breaks = c(0, 0.87),
                     labels = c("(1, 0, 0)", "(0, 1, 0)")) +            # y 軸の範囲
  coord_fixed(ratio = 1) +                                              # グラフの縦横比
  facet_wrap(~ parameter, ncol = 3) +                                   # グラフの分割
  labs(title = "Dirichlet Distribution",                                # タイトル
       x = expression(paste(phi[1], ", ", phi[2], sep = "")),           # x 軸ラベル
       y = expression(paste(phi[1], ", ", phi[3], sep = "")),           # y 軸ラベル
       color = "")
```



・対数の期待値

ディガンマ関数

$$\Psi(x) = \frac{d}{dx} \log \Gamma(x)$$

を用いて、対数をとった変数の期待値を導出する。

対数をとった変数 $\log \phi_v$ とその生成確率 $p(\boldsymbol{\phi}|\boldsymbol{\beta})$ とを掛け合わせたものを、積分して (総和をとって) 期待値を求める。

$$
\begin{aligned}
\int p(\boldsymbol{\phi}|\boldsymbol{\beta}) \log \phi_v d\boldsymbol{\phi} &= \int \frac{\Gamma(\hat{\beta})}{\prod_{v'=1}^V \Gamma(\beta_v)} \prod_{v'=1}^V \phi_{v'}^{\beta_{v'}-1} \log \phi_v d\boldsymbol{\phi} \qquad \text{(i)} \\
&= \frac{\Gamma(\hat{\beta})}{\prod_{v'=1}^V \Gamma(\beta_{v'})} \int \frac{\partial \prod_{v'=1}^V \phi_{v'}^{\beta_{v'}-1}}{\partial \beta_v} d\boldsymbol{\phi} \qquad \text{(ii)} \\
&= \frac{\Gamma(\hat{\beta})}{\prod_{v'=1}^V \Gamma(\beta_{v'})} \frac{\partial}{\partial \beta_v} \int \prod_{v'=1}^V \phi_{v'}^{\beta_{v'}-1} d\boldsymbol{\phi} \qquad \text{(iii)} \\
&= \frac{\Gamma(\hat{\beta})}{\prod_{v'=1}^V \Gamma(\beta_{v'})} \frac{\partial}{\partial \beta_v} \frac{\prod_{v'=1}^V \Gamma(\beta_{v'})}{\Gamma(\hat{\beta})} \qquad \text{(iv)} \\
&= \frac{\partial}{\partial \beta_v} \log \frac{\prod_{v'=1}^V \Gamma(\beta_{v'})}{\Gamma(\hat{\beta})} \qquad \text{(v)} \\
&= \frac{\partial \log \prod_{v'=1}^V \Gamma(\beta_{v'})}{\partial \beta_v} - \frac{\partial \log \Gamma(\hat{\beta})}{\partial \beta_v} \qquad \text{(vi)} \\
&= \frac{\partial \log \Gamma(\beta_v)}{\partial \beta_v} - \frac{\partial \log \Gamma(\hat{\beta})}{\partial \hat{\beta}} \qquad \text{(vii)} \\
&= \Psi(\beta_v) - \Psi(\hat{\beta})
\end{aligned}
$$

【途中式の途中式】

- (i)：ϕ_v を a、$\beta_v - 1$ を x としたとき、$\prod_{v'=1}^V \phi_{v'}^{\beta_{v'}-1} \log \phi_v$ が指数関数の微分 $\frac{\partial a^x}{\partial x} = a^x \log a$ 後の形をしていることから、微分前の形に置き換える。
- (ii)：$\frac{\partial \prod_{v'=1}^V \phi_{v'}^{\beta_{v'}-1}}{\partial \beta_v}$ を $\frac{\partial}{\partial \beta_v} \prod_{v'=1}^V \phi_{v'}^{\beta_{v'}-1}$ に、表記を変える。
- (iii)：ディリクレ分布の正規化項 (1.13) より、置き換える。
- (iv)：$\frac{\prod_{v'=1}^V \Gamma(\beta_{v'})}{\Gamma(\hat{\beta})}$ を $f(x)$ としたとき、式全体が対数関数の微分 $\frac{\partial \log f(x)}{\partial x} = \frac{1}{f(x)} \frac{\partial f(x)}{\partial x}$ 後の形をしていることから、微分前の形に置き換える。
- (v)：$\log \frac{A}{B} = \log A - \log B$ であることから、分割する。
- (vi)：

前の項のは、β_v に関して微分すると β_v 以外の項は消える。

$$
\begin{aligned}
\frac{\partial \log \prod_{v'=1}^V \Gamma(\beta_{v'})}{\partial \beta_v} &= \frac{\partial \sum_{v'=1}^V \log \Gamma(\beta_{v'})}{\partial \beta_v} \\
&= \frac{\partial \log \Gamma(\beta_1)}{\partial \beta_v} + \frac{\partial \log \Gamma(\beta_2)}{\partial \beta_v} + \cdots + \frac{\partial \log \Gamma(\beta_v)}{\partial \beta_v} + \cdots + \frac{\partial \log \Gamma(\beta_V)}{\partial \beta_v} \\
&= 0 + 0 + \cdots + \frac{\partial \log \Gamma(\beta_v)}{\partial \beta_v} + \cdots + 0 \\
&= \frac{\partial \log \Gamma(\beta_v)}{\partial \beta_v}
\end{aligned}
$$

後の項は、合成関数の微分 $\frac{\partial f}{\partial x} = \frac{\partial f}{\partial u} \frac{\partial u}{\partial x}$ より

$$
\frac{\partial \log \Gamma(\hat{\beta})}{\partial \beta_v} = \frac{\partial \log \Gamma(\hat{\beta})}{\partial \hat{\beta}} \frac{\partial \hat{\beta}}{\partial \beta_v}
$$

となり、また

$$
\begin{aligned}
\frac{\partial \hat{\beta}}{\partial \beta_v} &= \frac{\partial \beta_1}{\partial \beta_v} + \frac{\partial \beta_2}{\partial \beta_v} + \cdots + \frac{\partial \beta_v}{\partial \beta_v} \cdots + \frac{\partial \beta_V}{\partial \beta_v} \\
&= 0 + 0 + \cdots + 1 + \cdots + 0 \\
&= 1
\end{aligned}
$$

であるため

$$
\frac{\partial \log \Gamma(\hat{\beta})}{\partial \beta_v} = \frac{\partial \log \Gamma(\hat{\beta})}{\partial \hat{\beta}} \frac{\partial \hat{\beta}}{\partial \beta_v} = \frac{\partial \log \Gamma(\hat{\beta})}{\partial \hat{\beta}}
$$

となる。

- (vii)：ディガンマ関数の定義より、置き換える。

---

2 ユニグラムモデル

2.1 文書表現

トピックモデルでは、文書中の単語の出現順は無視して、どのような単語が使われているのかを重視する。そのため、文書を単語の多重集合 BOW(bag-fo-words) として扱う。語順の情報がなくなるため、主語述語などの係り受け関係は考慮されない。多重集合とは重複を許す集合のこと。

・記号一覧

記号	意味	関係性
D	文書数	
$d \in \{1, 2, \cdots, D\}$	文書インデックス	
N	文書全体での単語数	$N = \sum_{d=1}^{D} N_d = \sum_{v=1}^{V} N_v$
$n \in \{1, 2, \cdots, N_d\}$	単語インデックス	
N_d	文書 d の単語数 (文書長)	$N_d = \sum_{v=1}^{V} N_{dv}$
N_v	文書全体での語彙 v の出現回数	$N_v = \sum_{d=1}^{D} N_{dv}$
N_{dv}	文書 d での語彙 v の出現回数	
V	文書全体での単語の種類数 (語彙数)	$N_d \geq V$
$v \in \{1, 2, \cdots, V\}$	語彙インデックス	
$\boldsymbol{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_d, \cdots, \boldsymbol{w}_D)$	文書集合	
$\boldsymbol{w}_d = (w_{d1}, \cdots, w_{dn}, \cdots, w_{dN_d})$	文書 d の単語集合	
$w_{dn} \in \{1, 2, \cdots, V\}$	文書 d の n 番目の単語	$w_{dn} \sim \text{Categorical}(\boldsymbol{\phi})$
$\boldsymbol{\phi} = (\phi_1, \cdots, \phi_v, \cdots, \phi_V)$	単語分布	$\boldsymbol{\phi} \sim \text{Dirichlet}(\beta)$
ϕ_v	語彙 v の出現確率	$\phi \geq 1,\ \sum_{v=1}^{V} \phi_v = 1$
β	単語分布のパラメータ (ハイパーパラメータ)	

2.2 ユニグラムモデル

データを生成する確率モデルを生成モデルと呼ぶ。トピックモデルも BOW 表現された文書集合を生成する確率モデルである。

ユニグラムモデルでは、全ての単語はある 1 つのカテゴリ分布 Categorical($\boldsymbol{\phi}$) に従って生成されると仮定した文書集合 (BOW) の生成モデルである。

単語 v が出現する確率を ϕ_v とする。

$$\boldsymbol{\phi} = (\phi_1, \phi_2, \cdots, \phi_V),\ \phi_v \geq 0,\ \sum_{v=1}^{V} \phi_v = 1$$

パラメータ $\boldsymbol{\phi}$ が与えられたときの文書集合 $\boldsymbol{W}$ の確率は次のようになる。

$$
\begin{aligned}
p(\boldsymbol{W}|\boldsymbol{\phi}) &= \prod_{d=1}^{D} p(\boldsymbol{w}_d|\boldsymbol{\phi}) & \text{(i)} \\
&= \prod_{d=1}^{D} \prod_{n=1}^{N_d} p(w_{dn}|\boldsymbol{\phi}) & \text{(ii)} \\
&= \prod_{d=1}^{D} \prod_{n=1}^{N_d} \phi_{w_{dn}} & \text{(iii)} \\
&= \prod_{v=1}^{V} \phi_v^{N_v} & (2.1)
\end{aligned}
$$

【途中式の途中式】

- (0)：$\boldsymbol{\phi}$ が与えられたとき、各文書 $(\boldsymbol{w}_1, \boldsymbol{w}_2, \cdots, \boldsymbol{w}_d)$ は影響し合うことなく独立に生成されることから。

$$
p(\boldsymbol{W}|\boldsymbol{\phi}) = p(\boldsymbol{w}_1, \cdots, \boldsymbol{w}_D|\boldsymbol{\phi}) = p(\boldsymbol{w}_1|\boldsymbol{\phi}) * \cdots * p(\boldsymbol{w}_D|\boldsymbol{\phi}) = \prod_{d=1}^{D} p(\boldsymbol{w}_d|\boldsymbol{\phi})
$$

- (i)：$\boldsymbol{\phi}$ が与えられたとき、文書 d の各単語 $(w_{d1}, w_{d2}, \cdots, w_{dN_d})$ は影響し合うことなく独立に生成されることから。

$$
p(\boldsymbol{w}_d|\boldsymbol{\phi}) = p(w_{d1}, \cdots, w_{dN_d}|\boldsymbol{\phi}) = p(w_{d1}|\boldsymbol{\phi}) * \cdots * p(w_{dN_d}|\boldsymbol{\phi}) = \prod_{n=1}^{N_d} p(w_{dn}|\boldsymbol{\phi})
$$

- (ii)：$\phi_{w_{dn}}$ は単語 w_{dn} が出現する確率であることから。

$$
p(w_{dn}|\boldsymbol{\phi}) = \phi_{w_{dn}}
$$

- (iii)：語彙 v と各文書中の出現回数 N_v を用いて置き換える。

文書 d の n 番目の単語 $w_{d,n}$ は重複を許す形式である。そのため例えば、文書 1 の 1 番目と 8 番目の単語 $w_{1,1}, w_{1,8}$ と文書 3 の 5 番目の単語 $w_{3,5}$ が同じ単語であるとする。一方、語彙 v は重複しない形式であり、また文書全体での語彙 v の出現回数は N_v である。なので先ほどの例を表現すると、1 番目の語彙は全文書中に 3 回出現すること $(N_1 = 3)$ になる。

従って、単語 $w_{1,1}, w_{1,8}, w_{3,5}$ の 3 語が同時に出現する確率は

$$
\phi_{w_{1,1}} * \phi_{w_{1,8}} * \phi_{w_{3,5}} = \phi_1 * \phi_1 * \phi_1 = \phi_1^3
$$

になる。

つまり、文書全体に含まれる単語の出現確率 $\phi_{w_{dn}}$ を各文書 $(1, \cdots, D)$ と各単語 $(1, \cdots, N_d)$ までを掛け合わせたものと、語彙の出現確率 $\phi_v^{N_v}$ を各語彙 $1, \cdots, V$ までを掛け合わせたものはどちらも文書全体 $\boldsymbol{W}$ の出現確率を表している。

パラメータ $\boldsymbol{\phi}$ が与えられたときに、文書集合 $\boldsymbol{W}$ を生成するユニグラムモデルの定義を確認した。

次節からはその応用として、文書集合 $\boldsymbol{W}$ が与えられたときに、未知のパラメータ $\boldsymbol{\phi}$ を推定する。

2.3 最尤推定

最尤推定では、尤度を最大化するパラメータ $\boldsymbol{\phi}$ を推定する。

確率モデルのパラメータを $\boldsymbol{\phi}$、データを $\boldsymbol{W}$ としたとき、尤度は条件付き確率 $p(\boldsymbol{W}|\boldsymbol{\phi})$ を $\boldsymbol{\phi}$ の関数とみたものになる。

計算を簡単にするため、対数をとった尤度である対数尤度を最大化する。$\operatorname*{argmax}_{\boldsymbol{\phi}} f(x)$ は、関数 $f(x)$ を最大化させる $\boldsymbol{\phi}$ を求めることである。

$$
\operatorname*{argmax}_{\boldsymbol{\phi}} \log p(\boldsymbol{W}|\boldsymbol{\phi})
$$

ラグランジュの未定乗数法を用いて、$\sum_{v=1}^V \phi_v = 1$ の制約下で、$\log p(\boldsymbol{W}|\boldsymbol{\phi})$ が最大となるパラメータ $\boldsymbol{\phi}$ を求める。

$$
\begin{aligned}
F &= \log p(\boldsymbol{W}|\boldsymbol{\phi}) + \lambda \left(\sum_{v=1}^V \phi_v - 1 \right) && \text{(i)} \\
&= \log \left(\prod_{v=1}^V \phi_v^{N_v} \right) + \lambda \left(\sum_{v=1}^V \phi_v - 1 \right) && \text{(ii)} \\
&= \sum_{v=1}^V N_v \log \phi_v + \lambda \left(\sum_{v=1}^V \phi_v - 1 \right)
\end{aligned}
$$

【途中式の途中式】

- (0)：ラグランジュの未定乗数 λ を用いて、関数 $F = f(x) + \lambda g(x)$ をつくる。
- (i)：式 (2.1) より、置き換える。
- (ii)：
  - $\log(A * B) = \log A + \log B$ で、また $\log x^a = a * \log x$ であることから

$$
\begin{aligned}
\log \left(\prod_{v=1}^V \phi_v^{N_v} \right) &= \log(\phi_1^{N_1} * \cdots * \phi_v^{N_v} * \cdots * \phi_V^{N_V}) \\
&= \log \phi_1^{N_1} + \cdots + \log \phi_v^{N_v} + \cdots + \log \phi_V^{N_V} \\
&= \sum_{v=1}^V \log \phi_v^{N_v} \\
&= \sum_{v=1}^V N_v \log \phi_v
\end{aligned}
$$

となる。

これを ϕ_v について微分して $\frac{\partial F}{\partial \phi_v} = 0$ となる ϕ_v を求める。

$$
\begin{aligned}
\frac{\partial F}{\partial \phi_v} = \frac{N_v}{\phi_v} + \lambda = 0 \\
\phi_v = -\frac{N_v}{\lambda}
\end{aligned}
\tag{2.3}
$$

【途中式の途中式】

- 対数を微分すると $(\log x)' = \frac{1}{x}$ である。
- また、ϕ_v についての微分なので、それ以外の ϕ_1 から ϕ_V に関しては 0 となる。

従って、前の項については

$$
\begin{aligned}
\sum_{v=1}^{V} \log \phi_v &= \log \phi_1 + \cdots + \log \phi_v + \cdots + \log \phi_V \\
\left(\sum_{v=1}^{V} \log \phi_v \right)' &= 0 + \cdots + \frac{1}{\phi_v} + \cdots + 0 \\
&= \frac{1}{\phi_v}
\end{aligned}
$$

なので、$N_v \log \phi_v$ の項のみが残り、$N_v * \frac{1}{\phi_v} = \frac{N_v}{\phi_v}$ になる。同様に、後ろの項は

$$
\begin{aligned}
\sum_{v=1}^{V} \phi_v &= \phi_1 + \cdots + \phi_v + \cdots + \phi_V \\
\left(\sum_{v=1}^{V} \phi_v \right)' &= 0 + \cdots + 1 + \cdots + 0 \\
&= 1
\end{aligned}
$$

なので、$\lambda * \phi_v - 1$ のみが残り、これを微分した $\lambda * 1 - 0 = \lambda$ になる。

これを両辺で v に関して和をとると

$$
\begin{aligned}
\sum_{v=1}^{V} \phi_v &= -\sum_{v=1}^{V} \frac{N_v}{\lambda} \\
1 &= -\frac{N}{\lambda} \\
\lambda &= -N
\end{aligned}
$$

【途中式の途中式】

- $\sum_{v=1}^{V} \phi_v = 1$ なので置き換える。
- $\sum_{v=1}^{V} N_v$ は各語彙の出現回数 N_v を全て足し合わせたものなので、総単語数 N である。

となる。これを式 (2.3) に代入すると

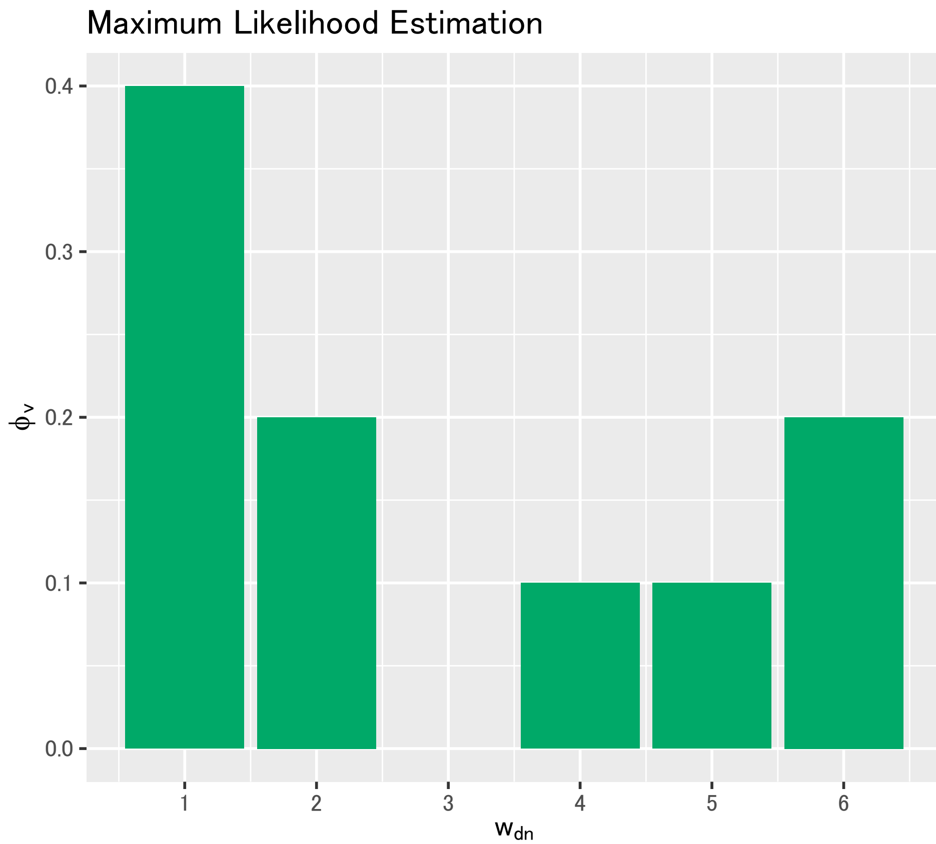
$$\phi_v = \frac{N_v}{N} \tag{2.4}$$

が得られる。これは語彙 v の出現回数 N_v を総単語数 N で割ったものである。つまり、文書全体で語彙 v が現れた割合が最尤推定値となる。

・推定結果の確認

$\beta = 2$ として、サイコロを 10 回振ったときの分布から推定を行う。

```
# 利用パッケージ
library(tidyverse)

# パラメーターの設定
beta  <- 2  # 任意の値を指定する
shake <- 10 # 任意の試行回数を指定する

# サイコロを振る
shake_result1 <- data.frame(w_dn = sample(x = 1:6, size = shake, replace = TRUE)) %>%
                 group_by(w_dn) %>%   # 出目でグループ化
                 summarise(N_v = n()) # 出目ごとにカウント

# 出ない目があったとき用の対策
shake_result2 <- left_join(data.frame(w_dn = 1:6), shake_result1, by = "w_dn")
shake_result2$N_v[is.na(shake_result2$N_v)] <- 0 # NA を 0 に置換

# 最尤推定
mle_result <- shake_result2 %>%
              mutate(phi_v = N_v / sum(N_v))

# 描画
ggplot(data = mle_result, mapping = aes(x = w_dn, y = phi_v)) + # データの設定
  geom_bar(stat = "identity", fill = "#00A968") +     # 棒グラフ
  scale_x_continuous(breaks = 1:6, labels = 1:6) +    # 軸目盛
  labs(title = "Maximum Likelihood Estimation",       # 図のタイトル
       x = expression(w[dn]), y = expression(phi[v])) # 軸ラベル
```



この方法だと試行回数が少ないと過学習のおそれがある。例えば、観測されなかった要素については確率を 0 と推定する。あるいは、サイコロを 1 回振って 4 が出たという観測データから推定すると、4 の確率を 100% と推定することになる。このように、事実と異なる推定結果となってしまう。
そこで、次節の最大事後確率推定を用いることでこの問題を緩和できる。

2.4 最大事後確率推定

最大事後確率推定 (MAP 推定) では、事後確率 $p(\boldsymbol{\phi}|\boldsymbol{W},\beta)$ が最大となるパラメータを求める。

データを観測する前の確率 $\boldsymbol{\phi}$ を表す事前確率を $p(\boldsymbol{\phi}|\beta)$ とする。β は、事前確率のパラメータである。事前確率と尤度 $p(\boldsymbol{W}|\boldsymbol{\phi})$ にベイズの定理を用いて、事後確率を求める。

$$p(\boldsymbol{\phi}|\boldsymbol{W},\beta) = \frac{p(\boldsymbol{\phi}|\beta)p(\boldsymbol{W}|\boldsymbol{\phi})}{p(\boldsymbol{W}|\beta)} \tag{2.5}$$

分母の $p(\boldsymbol{W}|\beta)$ は、$\boldsymbol{\phi}$ に依存しない。そのため、最大化したい $\boldsymbol{\phi}$ と無関係なので無視できる。従って、分子を最大化するパラメータが MAP 推定値になる。最尤推定のときと同様に、計算を簡単にするため対数をとって求めていく。

$$\mathop{\mathrm{argmax}}_{\boldsymbol{\phi}} p(\boldsymbol{\phi}|\boldsymbol{W},\beta) = \mathop{\mathrm{argmax}}_{\boldsymbol{\phi}} \left[\log p(\boldsymbol{W}|\boldsymbol{\phi}) + \log p(\boldsymbol{\phi}|\beta)\right] \tag{2.6}$$

MAP 推定では、事前分布が必要である。事後分布が事前分布と同じ形の分布になる共役事前分布を事前分布として用いると計算が簡単になる。

カテゴリ分布の共役事前分布はディリクレ分布なので、事前確率 $p(\boldsymbol{\phi}|\beta)$ をパラメータ $\beta, \cdots, \beta$ を持つディリクレ分布

$$p(\boldsymbol{\phi}|\beta) = \mathrm{Dirichlet}(\boldsymbol{\phi}|\beta, \cdots, \beta) = \frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \prod_{v=1}^{V} \phi_v^{\beta-1} \tag{2.7}$$

とする。尤度 (2.1) と事前分布 (2.7) を用いて、式 (2.6) を解く。

$$\begin{aligned}
\underset{\boldsymbol{\phi}}{\mathrm{argmax}}\, p(\boldsymbol{\phi}|\boldsymbol{W}, \beta) &= \underset{\boldsymbol{\phi}}{\mathrm{argmax}} \left[\log p(\boldsymbol{W}|\boldsymbol{\phi}) + \log p(\boldsymbol{\phi}|\beta)\right] && \text{(i)} \\
&= \underset{\boldsymbol{\phi}}{\mathrm{argmax}} \left[\log \prod_{v=1}^{V} \phi_v^{N_v} + \log \left(\frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \prod_{v=1}^{V} \phi_v^{\beta-1}\right)\right] && \text{(ii)} \\
&= \underset{\boldsymbol{\phi}}{\mathrm{argmax}} \left[\sum_{v=1}^{V} \log \phi_v^{N_v} + \log \Gamma(\beta V) - V \log \Gamma(\beta) + \sum_{v=1}^{V} \log \phi_v^{\beta-1}\right] && \text{(iii)} \\
&= \underset{\boldsymbol{\phi}}{\mathrm{argmax}} \left[\sum_{v=1}^{V} N_v \log \phi_v + \sum_{v=1}^{V} (\beta - 1) \log \phi_v\right] && \text{(iv)} \\
&= \underset{\boldsymbol{\phi}}{\mathrm{argmax}} \sum_{v=1}^{V} N_v \log \phi_v + (\beta - 1) \log \phi_v && \text{(v)} \\
&= \underset{\boldsymbol{\phi}}{\mathrm{argmax}} \sum_{v=1}^{V} (N_v + \beta - 1) \log \phi_v
\end{aligned}$$

【途中式の途中式】

- (i)：尤度 (2.1)、事前分布 (2.7) より、置き換える。
- (ii)：
  - $\log\left(\frac{A}{B}C\right) = \log A - \log B + \log C$ である。

$$\begin{aligned}
\log\left(\prod_{v=1}^{V} \phi_v^{N_v}\right) &= \log(\phi_1^{N_v} * \cdots * \phi_v^{N_v} * \cdots * \phi_V^{N_v}) \\
&= \log \phi_1^{N_v} + \cdots + \log \phi_v^{N_v} + \cdots + \log \phi_V^{N_v} \\
&= \sum_{v=1}^{V} \log \phi_v^{N_v}
\end{aligned}$$

- (iii)：
  - $\log \Gamma(\beta V) - V \log \Gamma(\beta)$ は、最大化したい $\boldsymbol{\phi}$ に影響しないので省く。
  - $\log x^a = a \log x$ である。
- (iv,v)：$\log \phi_v$ の項を 1 つにまとめる。

最尤推定と同様に、ラグランジュの未定乗数法を用いて、$\sum_{v=1}^{V} \phi_v = 1$ の制約の下で $\sum_{v=1}^{V}(N_v + \beta - 1) \log \phi_v$ を最大化する $\boldsymbol{\phi}$ を求める。

$$F = \sum_{v=1}^{V} (N_v + \beta - 1) \log \phi_v + \lambda \left(\sum_{v=1}^{V} \phi_v - 1\right)$$

ϕ_v について微分して、$\frac{\partial F}{\partial \phi_v} = 0$ になるときに最大化される。

$$\frac{\partial F}{\partial \phi_v} = \frac{N_v + \beta - 1}{\phi_v} + \lambda = 0$$
$$\phi_v = -\frac{N_v + \beta - 1}{\lambda}$$

両辺で v に関して和をとる。

$$\begin{aligned}\sum_{v=1}^V \phi_v &= -\sum_{v=1}^V \frac{N_v + \beta - 1}{\lambda} \\ 1 &= -\frac{N + (\beta - 1)V}{\lambda} \\ -\lambda &= N + (\beta - 1)V\end{aligned}$$

【途中式の途中式】

- $\sum_{v=1}^V \phi_v = 1$ なので置き換える。
- $\sum_{v=1}^V N_v$ は、各単語の出現回数 N_v を全て足し合わせたものなので、総単語数 N である。

これを $\phi_v = \frac{N_v+\beta-1}{-\lambda}$ に代入すると

$$\phi_v = \frac{N_v + \beta - 1}{N + (\beta - 1)V}$$

が求まる。これが MAP 推定値になる。

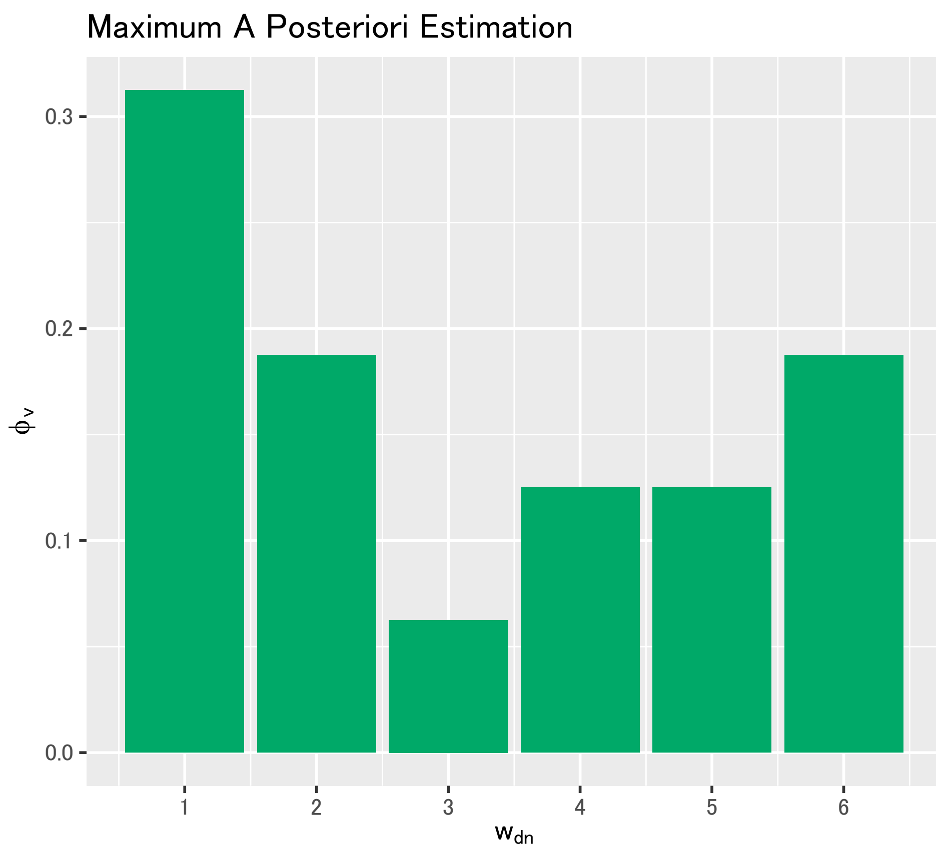
観測がない語彙に対して確率が負にならないためには、ディリクレ分布のパラメータを $\beta \geq 1$ で設定する必要があるまた、$\beta = 2$ のときラプラススムージングと呼ばれる。

・推定結果の確認

$\beta = 2$ として、サイコロを 10 回振ったときの分布から推定を行う。

```
# 利用パッケージ
library(tidyverse)

# パラメーターの設定
beta  <- 2  # 任意の値を指定する
shake <- 10 # 任意の試行回数を指定する

# サイコロを振る
shake_result1 <- data.frame(w_dn = sample(x = 1:6, size = shake, replace = TRUE)) %>%
                 group_by(w_dn) %>%   # 出目でグループ化
                 summarise(N_v = n()) # 出目ごとにカウント
```

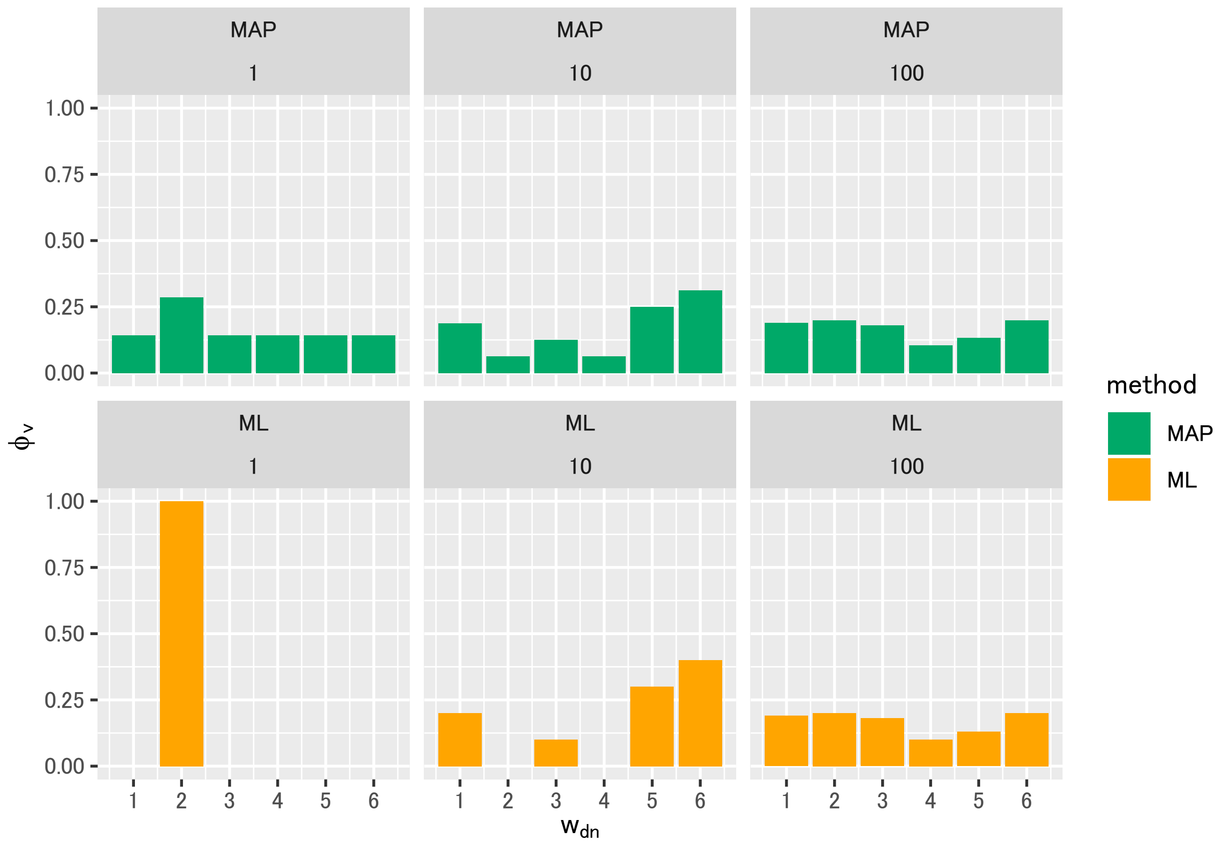
```
# 出ない目があったとき用の対策
shake_result2 <- left_join(data.frame(w_dn = 1:6), shake_result1, by = "w_dn")
shake_result2$N_v[is.na(shake_result2$N_v)] <- 0 # NA を 0 に置換

# 最大事後確率推定
map_result <- shake_result2 %>%
              mutate(phi_v = (N_v + beta - 1) / (sum(N_v) + (beta - 1) * length(w_dn)))

# 描画
ggplot(data = map_result, mapping = aes(x = w_dn, y = phi_v)) + # データの設定
  geom_bar(stat = "identity", fill = "#00A968") +      # 棒グラフ
  scale_x_continuous(breaks = 1:6, labels = 1:6) +     # 軸目盛
  labs(title = "Maximum Likelihood Estimation",        # 図のタイトル
       x = expression(w[dn]), y = expression(phi[v])) # 軸ラベル
```



最尤推定値と比較すると、語彙の出現回数 N_v に $\beta - 1$ を加えることで、観測値に極端に依存しない推定値になっている。

・**最尤推定と MAP 推定との比較**

試行回数が 1 回・10 回・100 回のときの最尤推定と MAP 推定との推定結果を比較する。



観測データ数が多くなるに従って、N_v が β に比べて大きくなるため、最尤推定値と MAP 推定値が近づいていくことが確認できる。

2.5 ベイズ推定

最尤推定と MAP 推定では、それぞれの手法で最も良いパラメータを 1 つ求めた。ベイズ推定では、ハイパーパラメータを推定することでパラメータの分布を求める。分布を推定することで、確からしさを表現でき、過学習の問題を緩和できる。

ディリクレ事前分布を $p(\phi|\beta)$、尤度を $p(\boldsymbol{W}|\phi)$ として、ユニグラムモデルのパラメータの事後分布 $p(\phi|\boldsymbol{W},\beta)$ をベイズの定理を用いて求める。

$$
\begin{aligned}
p(\boldsymbol{\phi}|\boldsymbol{W},\beta) &= \frac{p(\boldsymbol{\phi}|\beta)p(\boldsymbol{W}|\boldsymbol{\phi})}{\int p(\boldsymbol{\phi}|\beta)p(\boldsymbol{W}|\boldsymbol{\phi})d\boldsymbol{\phi}} \qquad \text{(i)} \\
&= \frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \prod_{v=1}^V \phi_v^{\beta-1} \prod_{v=1}^V \phi_v^{N_v} \int \frac{\Gamma(\beta)^V}{\Gamma(\beta V)} \frac{1}{\prod_{v=1}^V \phi_v^{\beta-1}} \frac{1}{\prod_{v=1}^V \phi_v^{N_v}} d\boldsymbol{\phi} \qquad \text{(ii)} \\
&= \frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \prod_{v=1}^V \phi_v^{\beta-1} \phi_v^{N_v} \frac{\Gamma(\beta)^V}{\Gamma(\beta V)} \frac{1}{\int \prod_{v=1}^V \phi_v^{\beta-1} \phi_v^{N_v} d\boldsymbol{\phi}} \qquad \text{(iii)} \\
&= \frac{\prod_{v=1}^V \phi_v^{N_v+\beta-1}}{\int \prod_{v=1}^V \phi_v^{N_v+\beta-1} d\boldsymbol{\phi}} \qquad \text{(iv)} \\
&= \frac{\Gamma(N+\beta V)}{\prod_{v=1}^V \Gamma(N_v+\beta)} \prod_{v=1}^V \phi_v^{N_v+\beta-1} \qquad \text{(v)} \\
&= \mathrm{Direchlet}(\boldsymbol{\phi}|N_1+\beta,\cdots,N_V+\beta)
\end{aligned}
$$

【途中式の途中式】

- (0)：MAP 推定の式 (2.5) の分母を $\boldsymbol{\phi}$ について周辺化したものである。
- (i)：
  - 式 (2.7) より $p(\boldsymbol{\phi}|\beta,\cdots,\beta) = \mathrm{Direchlet}(\boldsymbol{\phi}|\beta)$ なので、1.2.4 節のディリクレ分布の定義を用いて置き換える。
  - 式 (2.1) より、$p(\boldsymbol{W}|\boldsymbol{\phi}) = \prod_{v=1}^V \phi_v^{N_v}$ で置き換える。
- (ii,iii)：式を整理する。
- (iv)：
  - ディリクレ分布の正規化項 (1.13) より、$\frac{1}{\int \prod_{v=1}^V \phi_v^{N_v+\beta-1} d\boldsymbol{\phi}} = \frac{\Gamma(\sum_{v=1}^V N_v+\beta)}{\prod_{v=1}^V \Gamma(N_v+\beta)}$ で置き換える。
  - また、語彙 v の出現回数 N_v の総和は単語数 N なので、$\sum_{v=1}^V N_v = N$ であり、ここでの β は定数のため $\sum_{v=1}^V \beta = \beta V$ となる。
- (v)：式がディリクレ分布の形になっていることから。

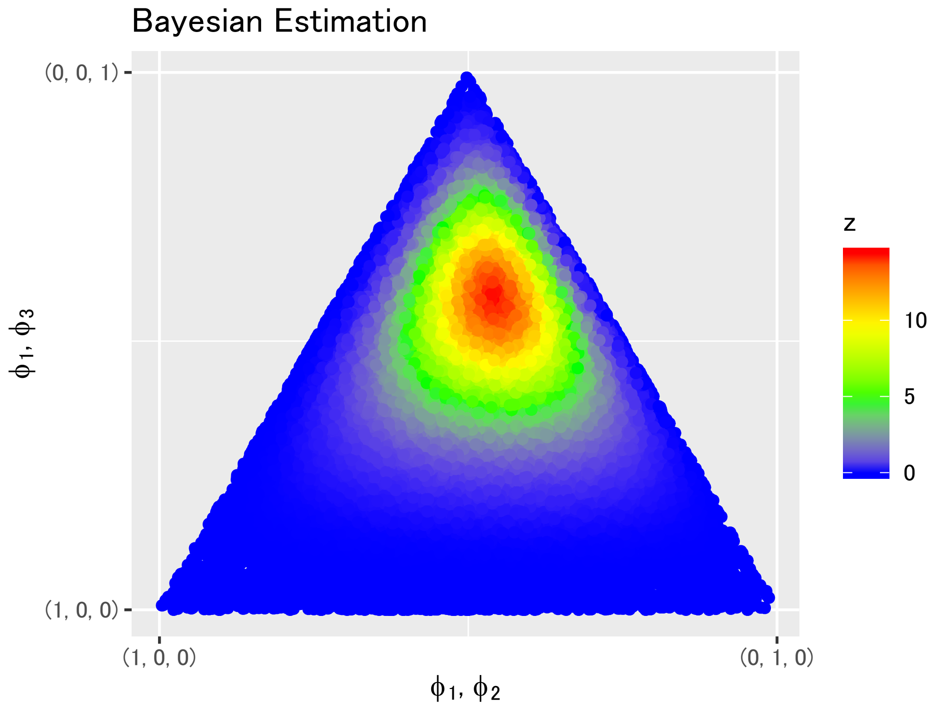
事後分布も事前分布と同様にディリクレ分布になっている。

・推定結果の確認

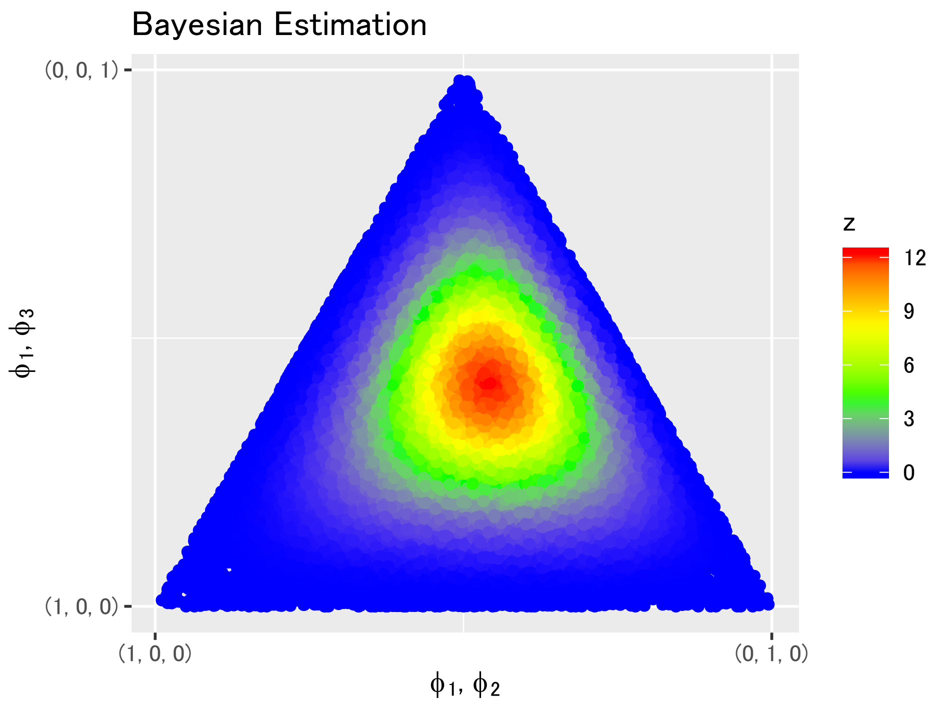
2 次元の図に落とし込めるように 3 変数 (サイコロの目が 1~3) の場合について推定する。ここでは、$\beta = 2$ として 9 回試行したときと 64 回試行したときのパラメータの分布の推定結果をみる。

```
# 利用パッケージ
library(tidyverse)

# パラメーターの設定
beta  <- 2 # 任意の値を指定する
shake <- 9 # 任意の試行回数を指定する

# サイコロを振る
shake_result1 <- data.frame(w_dn = sample(x = 1:3, size = shake, replace = TRUE)) %>%
                 group_by(w_dn) %>%    # 出目でグループ化
                 summarise(N_v = n()) # 出目ごとにカウント
```

```r
# 出ない目があったとき用の対策
shake_result2 <- left_join(data.frame(w_dn = 1:3), shake_result1, by = "w_dn")
shake_result2$N_v[is.na(shake_result2$N_v)] <- 0 # NA を 0 に置換

# 散布図用のランダムな値を用意
n_1 <- sample(x = seq(0, 1, 0.001), size = 40000, replace = TRUE)
n_2 <- sample(x = seq(0, 1, 0.001), size = 40000, replace = TRUE)
n_3 <- sample(x = seq(0, 1, 0.001), size = 40000, replace = TRUE)
n <- n_1 + n_2 + n_3 # 和が 1 となるための処理

# パラメータ：Φ
phi_1 <- n_1 / n
phi_2 <- n_2 / n
phi_3 <- n_3 / n

## ベイズ推定
# Gamma(N + beta * V)
numerator <- lgamma(sum(shake_result2$N_v) + beta * nrow(shake_result2))

# prod_{V=1}^V Gamma(N_v + beta)
denominator <- sum(lgamma(shake_result2$N_v + beta))

# N_v + beta - 1
exponent <- shake_result2$N_v + beta - 1

# phi_v^{`exponent`}
phi_v <- exponent[1] * log(phi_1) + exponent[2] * log(phi_2) + exponent[3] * log(phi_3)

# 作図用の df を作成
bayesian_result <- data.frame(x = phi_2 + (phi_3 / 2),   # 三角座標への変換処理
                              y = sqrt(3) * (phi_3 / 2), # 三角座標への変換処理
                              z = exp(numerator - denominator + phi_v)) # 推定値

# 描画
ggplot(data = bayesian_result, mapping = aes(x = x, y = y, color = z)) + # データの指定
  geom_point() +                                                        # 散布図
  scale_color_gradientn(colors = c("blue", "green", "yellow", "red")) + # 色
  scale_x_continuous(breaks = c(0, 1), labels = c("(1, 0, 0)", "(0, 1, 0)")) +    # x 軸目盛
  scale_y_continuous(breaks = c(0, 0.87), labels = c("(1, 0, 0)", "(0, 0, 1)")) + # y 軸目盛
  coord_fixed(ratio = 1) + # 縦横の比率
  labs(title = "Bayesian Estimation",                             # タイトル
       x = expression(paste(phi[1], ", ", phi[2], sep = "")), # x 軸ラベル
       y = expression(paste(phi[1], ", ", phi[3], sep = ""))) # y 軸ラベル
```



・$\beta = 2$、試行回数：64 回



試行回数が増えるとばらつきが小さくなることが確認できる。

ベイズ推定によってパラメータ $\boldsymbol{\phi}$ の分布 (事後分布) を推定した。次節では、ここから任意の単語 w^* の分布を予測していく。

2.6 ベイズ予測分布

ベイズ推定によって得たパラメータ $\boldsymbol{\phi}$ の分布を用いることで、推定のばらつきを考慮した予測が可能になる。ベイズ予測では、パラメータの事後分布を用いて、文書集合 $\boldsymbol{W}$ が与えられたときの単語 w^* が v である分布 $p(w^* = v|\boldsymbol{W})$ を求める。

ここでは、v はある特定の単語 w_v を示すもので、語彙インデックスは v' によって示している。

$$
\begin{aligned}
p(w^* = v|\boldsymbol{W}) &= \int p(w^* = v|\boldsymbol{\phi}) p(\boldsymbol{\phi}|\boldsymbol{W}) d\boldsymbol{\phi} && \text{(i)} \\
&= \int \phi_v \frac{\Gamma(N+\beta V)}{\prod_{v'=1}^{V} \Gamma(N_{v'}+\beta)} \prod_{v'=1}^{V} \phi_{v'}^{N_{v'}+\beta-1} d\boldsymbol{\phi} && \text{(ii)} \\
&= \frac{\Gamma(N+\beta V)}{\prod_{v'=1}^{V} \Gamma(N_{v'}+\beta)} \int \phi_v \phi_v^{N_v+\beta-1} \prod_{v' \neq v} \phi_{v'}^{N_{v'}+\beta-1} d\boldsymbol{\phi} && \text{(iii)} \\
&= \frac{\Gamma(N+\beta V)}{\prod_{v'=1}^{V} \Gamma(N_{v'}+\beta)} \int \phi_v^{N_v+\beta} \prod_{v' \neq v} \phi_{v'}^{N_{v'}+\beta-1} d\boldsymbol{\phi} && \text{(iv)} \\
&= \frac{\Gamma(N+\beta V)}{\prod_{v'=1}^{V} \Gamma(N_{v'}+\beta)} \frac{\Gamma(N_v+1+\beta) \prod_{v' \neq v} \Gamma(N_{v'}+\beta)}{\Gamma(N_v+1+\beta+\sum_{v' \neq v}(N_{v'}+\beta))} && \text{(v)} \\
&= \frac{\Gamma(N+\beta V)}{\prod_{v'=1}^{V} \Gamma(N_{v'}+\beta)} \frac{\Gamma(N_v+1+\beta) \prod_{v' \neq v} \Gamma(N_{v'}+\beta)}{\Gamma(N+1+\beta V)} && \text{(vi)} \\
&= \frac{\Gamma(N+\beta V) * (N_v+\beta)\Gamma(N_v+\beta) * \prod_{v' \neq v} \Gamma(N_{v'}+\beta)}{\prod_{v'=1}^{V} \Gamma(N_{v'}+\beta) * (N+\beta V)\Gamma(N+\beta V)} && \text{(vii)} \\
&= \frac{N_v+\beta}{N+\beta V}
\end{aligned}
$$

【途中式の途中式】

- (0) :??
- (i)：
  - $p(w^* = v|\boldsymbol{\phi})$ は、単語 w_v の出現確率のことなので ϕ_v である。
  - ベイズ推定で求めた事後分布より、置き換える。
- (ii,iii)：$\phi_v^{N_v+\beta-1}\phi_{v'}^{N_{v'}+\beta-1}$ から、$v' = v$ に関する項だけ取り出して

$$
\begin{aligned}
\prod_{v'=1}^{V} \phi_{v'}^{N_{v'}+\beta-1} &= \phi_1^{N_1+\beta-1} * \cdots * \phi_{v'}^{N_{v'}+\beta-1} * \cdots * \phi_V^{N_V+\beta-1} \\
&= \phi_v^{N_v+\beta-1} * \phi_1^{N_1+\beta-1} * \cdots * \phi_{v-1}^{N_{v-1}+\beta-1} * \phi_{v+1}^{N_{v+1}+\beta-1} * \cdots * \phi_V^{N_V+\beta-1} \\
&= \phi_v^{N_v+\beta-1} * \prod_{v' \neq v} \phi_{v'}^{N_{v'}+\beta-1}
\end{aligned}
$$

ϕ_v をまとめる。

- (iv)：
  - ディリクレ分布の正規化項 (1.13) より、置き換える。
  - ただし v の項については、$\phi_v^{N_v+\beta}$ と指数が他の項より 1 大きいため、式変形後の値が $N_v+1+\beta$ になる。
  - そのため、分母分子でそれぞれ v の項を除いてまとめている。
- (v)：$\sum_{v' \neq v} N_{v'}$ と $\sum_{v' \neq v} \beta = \beta(V-1)$ はそれぞれ v 以外の項を足し合わせたものなので、そこに v に関する N_v, β を加えることで、それぞれ $N, \beta V$ となる。
- (vi)：ガンマ関数の性質 $\Gamma(x+1) = x\Gamma(x)$ より、それぞれ

$$
\begin{aligned}
\Gamma(N_v+1+\beta) &= (N_v+\beta)\Gamma(N_v+\beta) \\
\Gamma(N+1+\beta V) &= (N+\beta V)\Gamma(N+\beta V)
\end{aligned}
$$

である。

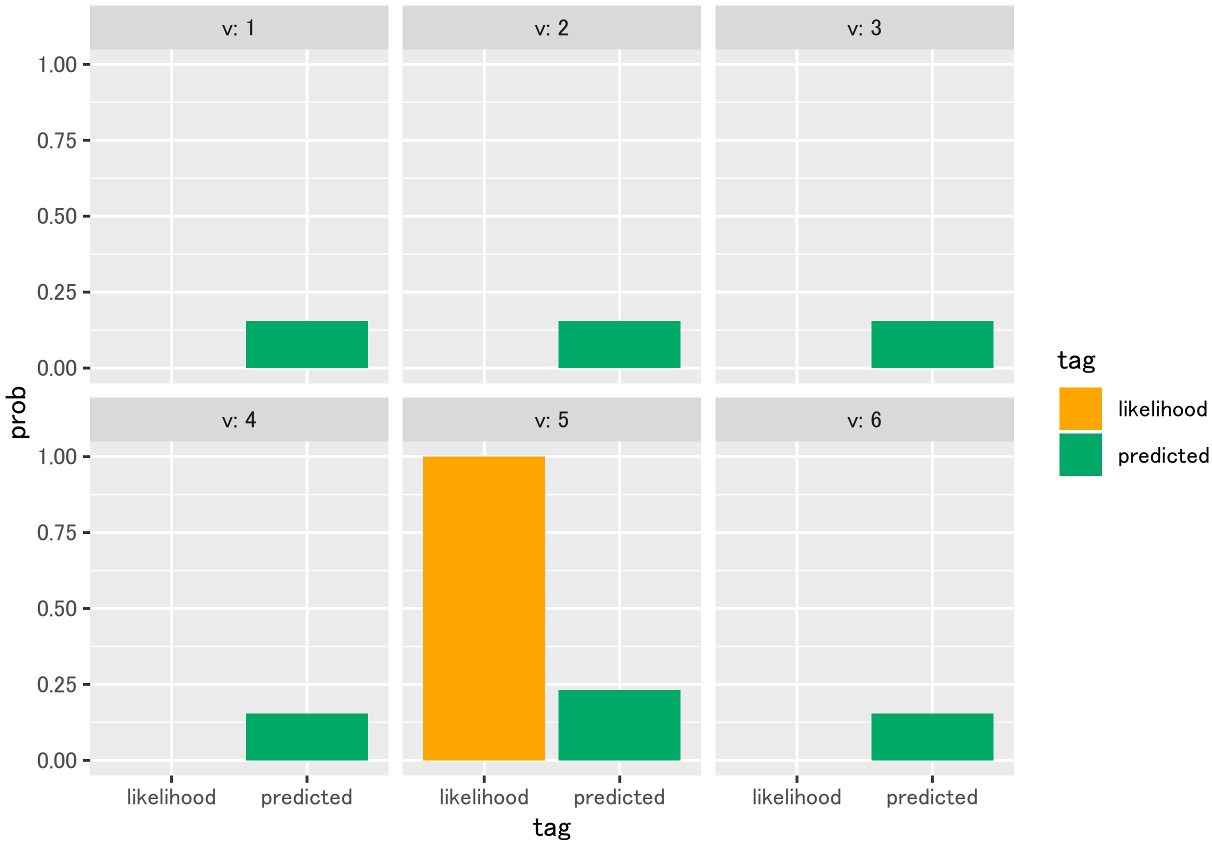
- (vii)：約分して式を整理する。ただし、$\Gamma(N_v + \beta) * \prod_{v' \neq v} \Gamma(N_{v'} + \beta) = \prod_{v'=1}^{V} \Gamma(N_{v'} + \beta)$ である。

・予測結果の確認

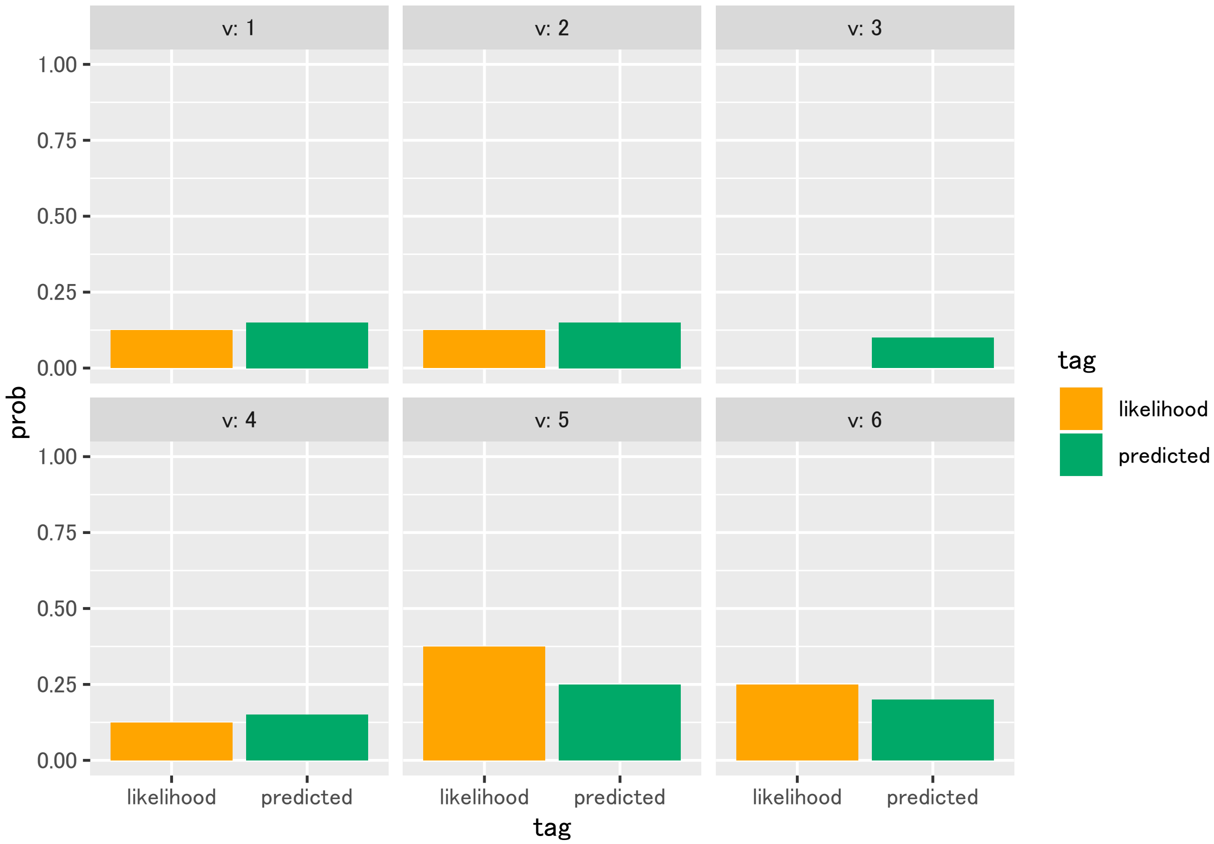
サイコロを 1 回・8 回・64 回振ったときの出目を観測データとする。$\beta = 2$ として、それぞれの観測値を基にベイズ予測を行う。試行回数ごとの予測分布と尤度、また試行回による分布の変化を確認する。

・試行回数：1

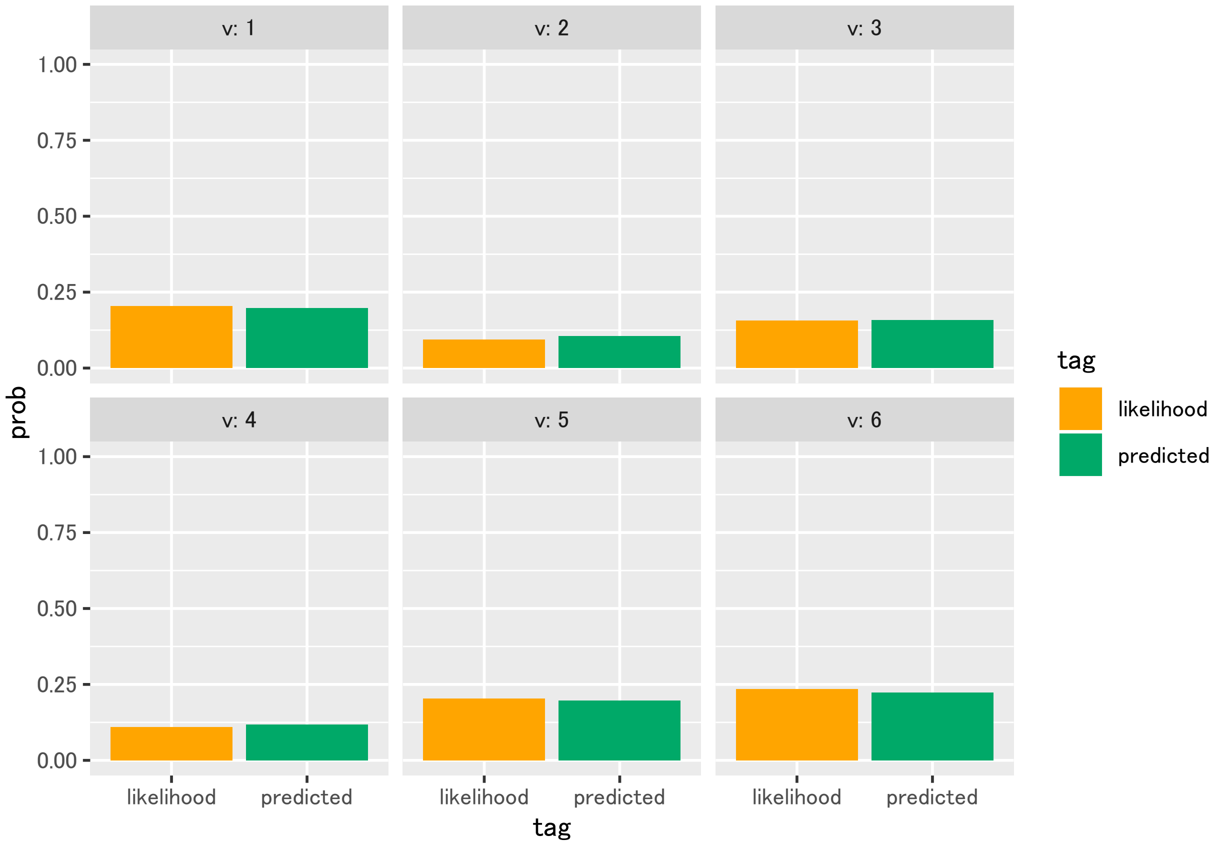
```
# パラメータの指定
beta  <- 2 # 任意の値を指定する
shake <- 1 # 任意の試行回数の指定

# 語彙インデックス：v
v <- 1:6

# サイコロを振る
shake_result1 <- data.frame(w_dn = sample(x = v, size = shake, replace = TRUE)) %>%
                 group_by(w_dn) %>%   # 出目でグループ化
                 summarise(N_v = n()) # 出目ごとにカウント

# 出ない目があったとき用の対策
shake_result2 <- left_join(data.frame(w_dn = v), shake_result1, by = "w_dn")
shake_result2$N_v[is.na(shake_result2$N_v)] <- 0 # NA を 0 に置換

# 総単語数：N
N <- sum(shake_result2$N_v)

# 総語彙数：V
V <- length(v)

# ベイズ予測
p_w <- (shake_result2$N_v + beta) / (N + beta * V)

# 作図用の df を作成
bpd_result <- data.frame(v = v,
                         likelihood = shake_result2$N_v / N,
                         predicted = p_w)
bpd_long <- gather(bpd_result, key = "tag", value = "prob", -v)

# 描画
ggplot(data = bpd_long, mapping = aes(x = tag, y = prob, fill = tag)) + # データの指定
  geom_bar(stat = "identity", position = "dodge") +    # 棒グラフ
  ylim(c(0, 1)) +                                      # y 軸の最小値・最大値
  scale_fill_manual(values = c("Orange", "#00A968")) + # 色
  facet_wrap( ~ v, labeller = label_both)              # グラフの分割
```


v: 1
v: 2
v: 3
v: 4
v: 5
v: 6
1.00
0.75
0.50
0.25
0.00
prob
likelihood
predicted
tag
tag
likelihood
predicted

・試行回数：8


v: 1
v: 2
v: 3
v: 4
v: 5
v: 6
prob
1.00
0.75
0.50
0.25
0.00
likelihood
predicted
tag
tag
likelihood
predicted

・試行回数：64



観測値から求める尤度よりも予測値の方が、真の値 1.667 に近い値となっていることが確認できる。また、試行回数が増えるにつれて尤度も予測分布も真の値に近づいていることも確認できる。

2.7 ハイパーパラメータ推定

MAP 推定とベイズ推定では、ハイパーパラメータと呼ばれる事前分布のパラメータが必要である。前節までは、ハイパーパラメータの値を $\beta = 2$ として任意に決めていた。この節では、経験ベイズ推定を用いてデータからハイパーパラメータを推定する。

・経験ベイズ推定

経験ベイズ推定では、パラメータ $\boldsymbol{\phi}$ を周辺化した周辺尤度 $p(\boldsymbol{W}|\beta)$ を最大化するハイパーパラメータ β を求める。

ユニグラムモデルの周辺尤度は

$$
\begin{aligned}
p(\boldsymbol{W}|\beta) &= \int p(\boldsymbol{W}|\boldsymbol{\phi})p(\boldsymbol{\phi}|\beta)d\boldsymbol{\phi} && \text{(i)} \\
&= \int \prod_{v=1}^{V} \phi_v^{N_v} \frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \prod_{v=1}^{V} \phi_v^{\beta-1} d\boldsymbol{\phi} && \text{(ii)} \\
&= \frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \int \prod_{v=1}^{V} \phi_v^{N_v+\beta-1} d\boldsymbol{\phi} && \text{(iii)} \\
&= \frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \prod_{v=1}^{V} \frac{\Gamma(N_v+\beta)}{\Gamma(N+\beta V)} && \text{(iv, 2.8)} \\
&= \frac{\Gamma(\beta V)}{\Gamma(N+\beta V)} \prod_{v=1}^{V} \frac{\Gamma(N_v+\beta)}{\Gamma(\beta)} && \text{(2.8')}
\end{aligned}
$$

【途中式の途中式】

- (i)：
  - 式 (2.1) より、$p(\boldsymbol{W}|\boldsymbol{\phi}) = \prod_{v=1}^{V} \phi_v^{N_v}$ で置き換える。
  - 式 (2.7) より、$p(\boldsymbol{\phi}|\beta) = \mathrm{Dirichlet}(\boldsymbol{\phi}|\beta)$ である。ディリクレ分布の式は 1.2.4 節を参照のこと。
- (ii)：$\boldsymbol{\phi}$ と関係のない項を $\int$ の外に出し、ϕ_v の項を 1 つにまとめて、式を整理する。
- (iii)：ディリクレ分布の正規化項 (1.13) より、置き換える。
- (iv)：分母分子の $\beta, \beta V$ を揃えるために、分母の項を入れ替える。

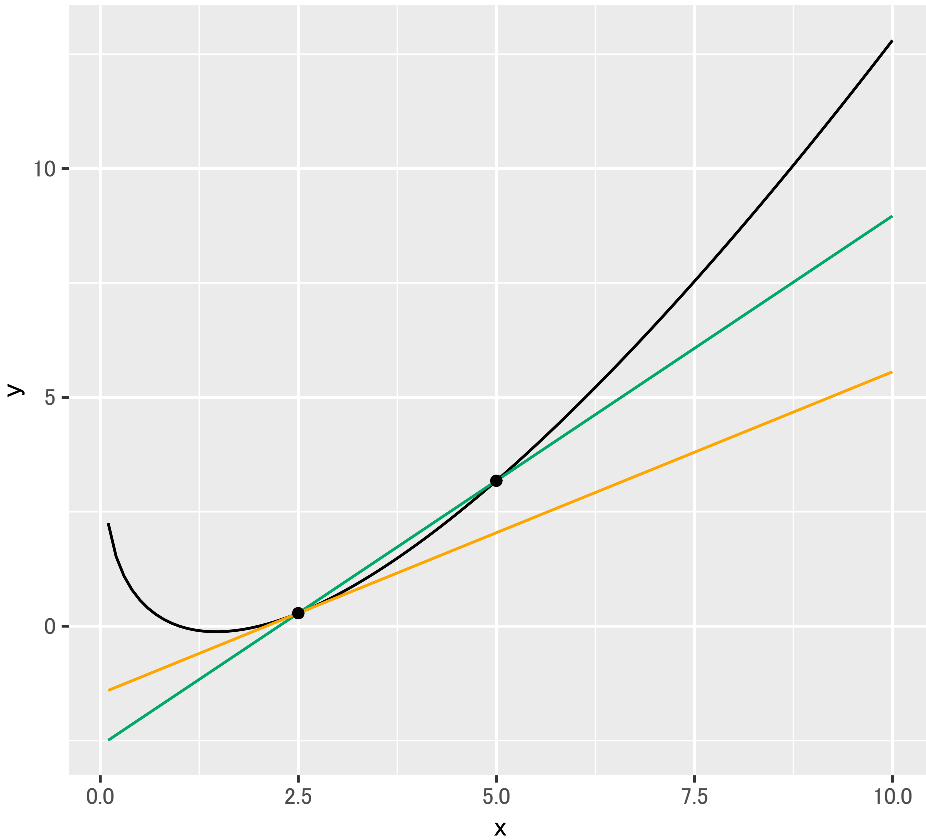
この分布はポリヤ分布と呼ばれる。

この式に不動点反復法を用いて更新を繰り返すことで、β に関して最大化できる。そこで、不動点反復法を行えるように式を変形する。

・不動点反復法のための式変形

まずは、ディガンマ関数 $\Psi(x)$ について確認する。

$$
\Psi(x) = \frac{d \log \Gamma(x)}{dx}
$$



ディガンマ関数の単調増加性から、対数をとったガンマ関数の曲線 $f(x) = \log\Gamma(x)$ 上の 2 点 $(x_1, y_1), (x_2, y_2)$ を通る直線の傾き $\frac{\log\Gamma(x_2)-\log\Gamma(x_1)}{x_2-x_1}$ と曲線上の点 (x_1, y_1) を通る接線の傾き $\Psi(x_1)$ との関係は

$$\frac{\log\Gamma(x_2) - \log\Gamma(x_1)}{x_2 - x_1} > \Psi(x_1)$$

である。ここから、x_2 を x、x_2 の 1 期前の値である x_1 を $\hat{x}$ として

$$\begin{aligned}
\frac{\log\Gamma(x) - \log\Gamma(\hat{x})}{x - \hat{x}} &\geq \Psi(\hat{x}) \\
\log\Gamma(x) - \log\Gamma(\hat{x}) &\geq \Psi(\hat{x})(x - \hat{x}) \\
\log\Gamma(x) &\geq \log\Gamma(\hat{x}) + (x - \hat{x})\Psi(\hat{x}) \qquad (1) \\
\Gamma(x) &\geq \Gamma(\hat{x}) + \exp\{(x - \hat{x})\Psi(\hat{x})\}
\end{aligned}$$

が得られる。続いて、x が $n + x$ だった場合は

$$\begin{aligned}
\Gamma(n + x) &\geq \Gamma(n + \hat{x}) + \exp[\{n + x - (n + \hat{x})\}\Psi(n + \hat{x})] \\
&= \Gamma(n + \hat{x}) + \exp\{(x - \hat{x})\Psi(n + \hat{x})\} \\
\log\Gamma(n + x) &\geq \log\Gamma(n + \hat{x}) + (x - \hat{x})\Psi(n + \hat{x}) \qquad (2)
\end{aligned}$$

である。式 (1) から式 (2) を引くと

$$
\begin{aligned}
\log\Gamma(x) - \log\Gamma(n+x) &\geq \log\Gamma(\hat{x}) + (x-\hat{x})\Psi(\hat{x}) - \log\Gamma(n+\hat{x}) - (x-\hat{x})\Psi(n+\hat{x}) \\
&= \log\Gamma(\hat{x}) - \log\Gamma(n+\hat{x}) + (x-\hat{x})\Big(\Psi(\hat{x}) - \Psi(n+\hat{x})\Big) \\
&= \log\Gamma(\hat{x}) - \log\Gamma(n+\hat{x}) + (\hat{x}-x)\Big(\Psi(n+\hat{x}) - \Psi(\hat{x})\Big) \\
\frac{\Gamma(x)}{\Gamma(n+x)} &\geq \frac{\Gamma(\hat{x})\exp\Big((\hat{x}-x)b\Big)}{\Gamma(n+\hat{x})}
\end{aligned}
$$

最後に指数をとって、この式が得られる。ここで、$b = \Psi(n+\hat{x}) - \Psi(\hat{x})$ と置いている。

同様に、(2) から (1) を引いた式を求める。

$$
\begin{aligned}
\log\Gamma(n+x) - \log\Gamma(x) &\geq \log\Gamma(n+\hat{x}) + (x-\hat{x})\Psi(n+\hat{x}) - \log\Gamma(\hat{x}) - (x-\hat{x})\Psi(\hat{x}) && \text{(i)} \\
&= \log\Gamma(n+\hat{x}) - \log\Gamma(\hat{x}) + (x-\hat{x})\Big(\Psi(n+\hat{x}) - \Psi(\hat{x})\Big) && \text{(ii)} \\
&= \log\Gamma(n+\hat{x}) - \log\Gamma(\hat{x}) + \frac{x-\hat{x}}{\hat{x}}\Big(\Psi(n+\hat{x}) - \Psi(\hat{x})\Big)\hat{x} && \text{(iii)} \\
&\fallingdotseq \log\Gamma(n+\hat{x}) - \log\Gamma(\hat{x}) + (\log x - \log\hat{x})\Big(\Psi(n+\hat{x}) - \Psi(\hat{x})\Big)\hat{x} && \text{(iv)} \\
\frac{\Gamma(n+x)}{\Gamma(x)} &\geq \frac{\Gamma(n+\hat{x})}{\Gamma(\hat{x})}\left(\frac{x}{\hat{x}}\right)^a && \text{(v)} \\
&= \frac{\Gamma(n+\hat{x})}{\Gamma(\hat{x})}\hat{x}^{-a}x^a && \text{(vi)} \\
&= cx^a
\end{aligned}
$$

【途中式の途中式】

- (i)：式を整理する。
- (ii)：$\frac{\hat{x}}{\hat{x}} = 1$ を分割して掛け合わせる。
- (iii)：対数変化率より、$\frac{x-\hat{x}}{\hat{x}} \fallingdotseq \log x - \log\hat{x}$ で置き換える。
- (iv)：
  - 両辺の指数をとる。
  - $a = \Big(\Psi(n+\hat{x}) - \Psi(\hat{x})\Big)\hat{x}$ と置く。
- (v)：$\frac{1}{x} = x^{-1}$ であり、$(x^n)^m = x^{n*m}$ であることから、$\left(\frac{x}{\hat{x}}\right)^a$ を分割する。
- (vi)：$c = \frac{\Gamma(n+\hat{x})}{\Gamma(\hat{x})}\hat{x}^{-a}$ と置く。

・まとめ

$$
\frac{\Gamma(x)}{\Gamma(n+x)} \geq \frac{\Gamma(\hat{x})\exp\{(\hat{x}-x)b\}}{\Gamma(n+\hat{x})}
$$

ここで

$$
b = \Psi(n+\hat{x}) - \Psi(\hat{x})
$$

また

$$\frac{\Gamma(n+x)}{\Gamma(x)} \geq cx^a \qquad \text{if } n \geq 1, x \geq 0$$

ここで

$$a = \Big(\Psi(n+\hat{x}) - \Psi(\hat{x})\Big)\hat{x}$$
$$c = \frac{\Gamma(n+\hat{x})}{\Gamma(\hat{x})}\hat{x}^{-a}$$

不動点反復法を用いるための式変形に、この関係性を用いる。

・ハイパーパラメータの推定

式 (2.8) に対して、不動点反復法を用いて以下の更新を収束するまで、繰り返すことで β に関して最大化する。

まずは、そのための式変形を行う。前節の関係性を用いて、式 (2.8) の前の項は

$$\frac{\Gamma(\beta V)}{\Gamma(N+\beta V)} \geq \frac{\Gamma(\hat{\beta}V)\exp\{(\hat{\beta}V - \beta V)b\}}{\Gamma(N+\hat{\beta}V)}$$

であり、後ろの項は

$$\prod_{v=1}^{V}\frac{\Gamma(N_v+\beta)}{\Gamma(\beta)} \geq \prod_{v=1}^{V}\frac{(N_v+\hat{\beta})}{\Gamma(\hat{\beta})}\hat{\beta}^{-a}\beta^{a}$$

であることから、これらをまとめて式 (2.8) は

$$\frac{\Gamma(\beta V)}{\Gamma(N+\beta V)}\prod_{v=1}^{V}\frac{\Gamma(N_v+\beta)}{\Gamma(\beta)} \geq \frac{\Gamma(\hat{\beta}V)\exp\{(\hat{\beta}V - \beta V)b\}}{\Gamma(N+\hat{\beta}V)}\prod_{v=1}^{V}\frac{(N_v+\hat{\beta})}{\Gamma(\hat{\beta})}\hat{\beta}^{-a}\beta^{a}$$

ここで

$$a = \Big(\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}$$
$$b = \Psi(n+\hat{\beta}V) - \Psi(\hat{\beta}V)$$

a, b をそれぞれ代入すると

$$\frac{\Gamma(\hat{\beta}V)\exp\Big\{(\hat{\beta}V - \beta V)\Big(\Psi(n+\hat{\beta}V) - \Psi(\hat{\beta}V)\Big)\Big\}}{\Gamma(N+\hat{\beta}V)}\prod_{v=1}^{V}\frac{(N_v+\hat{\beta})}{\Gamma(\hat{\beta})}\hat{\beta}^{-(\Psi(N_v+\hat{\beta})-\Psi(\hat{\beta}))\hat{\beta}}\beta^{(\Psi(N_v+\hat{\beta})-\Psi(\hat{\beta}))\hat{\beta}}$$

である。この式の対数をとり

$$
\begin{aligned}
G = &\log\Gamma(\hat{\beta}V) + (\hat{\beta}V - \beta V)\Big(\Psi(n+\hat{\beta}V) - \Psi(\hat{\beta}V)\Big) - \log\Gamma(N+\hat{\beta}V) \\
&+ \sum_{v=1}^{V}\log(N_v+\hat{\beta}) - V\log\Gamma(\hat{\beta}) - \sum_{v=1}^{V}\Big(\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\log\hat{\beta} + \sum_{v=1}^{V}\Big(\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\log\beta
\end{aligned}
$$

β について微分して 0 となる β を求める。

$$
\begin{aligned}
\frac{\partial G}{\partial \beta} &= -V\Big(\Psi(n+\hat{\beta}) - \Psi(\hat{\beta}V)\Big) + \sum_{v=1}^{V}\Big(\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\frac{1}{\beta} = 0 \\
\beta &= \hat{\beta}\frac{\sum_{v=1}^{V}\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})}{V\Psi(n+\hat{\beta}) - V\Psi(\hat{\beta}V)}
\end{aligned}
$$

$\hat{\beta}$ を β の 1 期前の値としていたので、これを $\hat{\beta}$ を現ステップの β、左辺の β を更新後 (次ステップ) のパラメータ β^{new} とした次の式が更新式となる。

$$
\beta^{\text{new}} = \beta\frac{\sum_{v=1}^{V}\Psi(N_v+\beta) - \Psi(\beta)}{V\Psi(n+\beta) - V\Psi(\beta V)} \tag{2.9}
$$

・最大事後確率推定

続いて、ハイパーパラメータ β の事前分布としてガンマ分布を仮定して、パラメータ $\boldsymbol{\phi}$ を周辺化したときのハイパーパラメータの事後確率を最大化する更新式を求めていく。

周辺尤度 $p(\boldsymbol{W}|\beta)$ にベイズの定理 (1.4) を用いて、両辺の対数をとる。

$$
\begin{aligned}
p(\beta|\boldsymbol{W}) &= \frac{p(\boldsymbol{W}|\beta)p(\beta)}{p(\boldsymbol{W})} \\
\log p(\beta|\boldsymbol{W}) &= \log p(\boldsymbol{W}|\beta) + \log p(\beta) - \log p(\boldsymbol{W})
\end{aligned}
$$

これを最大化する β を求める $(\underset{\beta}{\text{argmax}})$。ただし、$\log p(\boldsymbol{W})$ は β に影響しないので省く。

$$
\begin{aligned}
\underset{\beta}{\operatorname{argmax}}\, p(\beta|\boldsymbol{W}) =& \underset{\beta}{\operatorname{argmax}}[\log p(\boldsymbol{W}|\beta) + \log p(\beta)] \quad &\text{(i)}\\
=& \underset{\beta}{\operatorname{argmax}}\Bigg[\log\Gamma(\hat{\beta}V) + (\hat{\beta}V - \beta V)\Big(\Psi(n+\hat{\beta}V) - \Psi(\hat{\beta}V)\Big) - \log\Gamma(N+\hat{\beta}V)\\
&+ \sum_{v=1}^{V}\log(N_v+\hat{\beta}) - V\log\Gamma(\hat{\beta}) - \sum_{v=1}^{V}\Big(\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\log\hat{\beta}\\
&+ \sum_{v=1}^{V}\Big(\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\log\beta + \log\left(\frac{d^c}{\Gamma(c)}\beta^{c-1}\exp(-d\beta)\right)\Bigg] \quad &\text{(ii)}\\
=& \underset{\beta}{\operatorname{argmax}}\Big[\log\Gamma(\hat{\beta}V) + (\hat{\beta}V - \beta V)\Big(\Psi(n+\hat{\beta}V) - \Psi(\hat{\beta}V)\Big) - \log\Gamma(N+\hat{\beta}V)\\
&+ \sum_{v=1}^{V}\log(N_v+\hat{\beta}) - V\log\Gamma(\hat{\beta}) - \sum_{v=1}^{V}\Big(\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\log\hat{\beta}\\
&+ \sum_{v=1}^{V}\Big(\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\log\beta + c\log d - \log\Gamma(c) + (c-1)\log\beta - d\beta] \quad &\text{(iii)}\\
\frac{\partial}{\partial\beta} =& -V\Big(\Psi(n+\hat{\beta}V) - \Psi(\hat{\beta}V)\Big) + \sum_{v=1}^{V}\Big(\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\frac{1}{\beta} + (c-1)\frac{1}{\beta} - d = 0 \quad &\text{(iv)}\\
\beta =& \frac{c-1+\hat{\beta}\sum_{v=1}^{V}\Big(\Psi(N_v+\hat{\beta}) - \Psi(\hat{\beta})\Big)}{d + V\Psi(n+\hat{\beta}V) - V\Psi(\hat{\beta}V)}
\end{aligned}
$$

【途中式の途中式】

- (i)：
  - 対数周辺尤度 $\log p(\boldsymbol{W}|\beta)$ を変形したものが関数 G なので置き換える。
  - $p(\beta) = \mathrm{Gamma}(\beta|c,d)$ である。?? 何これどこから出てきた…
- (ii)：式を展開する。
- (iii)：β について微分して 0 となる β を求める。
- (iv)：β について解く。

$\hat{\beta}$ を今ステップの β、左辺の β を更新後のハイパーパラメータ β^{new} として

$$
\beta^{\mathrm{new}} = \frac{c-1+\beta\sum_{v=1}^{V}\Big(\Psi(N_v+\beta) - \Psi(\beta)\Big)}{d + V\Psi(n+\beta V) - V\Psi(\beta V)}
$$

が更新式となる。

ディリクレ事前分布を用いたユニグラムモデル場合、事後確率を最大化するハイパーパラメータは常に $\beta = 1$ であり、事前確率を用いない最尤推定と同じ結果になる。

Chapter2.8 モデル選択はパスしました。

3 混合ユニグラムモデル

・記号一覧

記号	意味	制約・関係性
D	文書数	$D = \sum_{k=1}^{K} D_k$
$d \in \{1, 2, \cdots, D\}$	文書インデックス	
D_k	トピックが k の文書数	
$\boldsymbol{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_d, \cdots, \boldsymbol{w}_D)$	文書集合	
$\boldsymbol{w}_d = (w_{d1}, \cdots, w_{dn}, \cdots, w_{dN_D})$	文書 d の単語集合	
$w_{dn} \in \{1, 2, \cdots, V\}$	文書 d の n 番目の単語	$w_{dn} \sim \text{Categorical}(\boldsymbol{\phi})$
N	文書集合全体での単語数	$N = \sum_{d=1}^{D} N_d = \sum_{v=1}^{V} N_v = \sum_{k=1}^{K} N_k$
$n \in \{1, 2, \cdots, N_{dv}\}$	単語インデックス	$N \geq V$
V	語彙数	
$v \in \{1, 2, \cdots, V\}$	語彙インデックス	
N_d	文書 d の単語数	$N_d = \sum_{v=1}^{V} N_{dv}$
N_v	文書全体での語彙 v の出現回数	$N_v = \sum_{d=1}^{D} N_{dv}$
N_{dv}	文書 d での語彙 v の出現回数	
N_k	トピックが k の文書における総単語数	$N_k = \sum_{v=1}^{V} N_{kv}$
N_{kv}	トピックが k の文書における語彙 v の出現回数	
K	トピック数	
$k \in \{1, 2, \cdots, K\}$	トピック	
$\boldsymbol{z} = (z_1, \cdots, z_d, \cdots, z_D)$	トピック集合	
$z_d \in \{1, \cdots, K\}$	文書 d に割り振られたトピック	$z_d \sim \text{Categorical}(\boldsymbol{\theta})$
$\boldsymbol{\theta} = (\theta_1, \cdots, \theta_k, \cdots, \theta_K)$	トピック分布	$\boldsymbol{\theta} \sim \text{Dirichlet}(\alpha)$
θ_k	各文書にトピック k が割り振られる確率	$p(k\|\boldsymbol{\theta}) = \theta_k \geq 0,\ \sum_{k=1}^{K} \theta_k = 1$
$\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_k, \cdots, \boldsymbol{\phi}_K)$	単語分布集合	$\boldsymbol{\phi} \sim \text{Dirichlet}(\beta)$
$\boldsymbol{\phi}_k = (\phi_{k1}, \cdots, \phi_{kv}, \cdots, \phi_{kV})$	トピック k の単語分布	
ϕ_{kv}	トピック k において語彙 v が出現する確率	$p(v\|\boldsymbol{\phi}_k) = \phi_{kv} \geq 0,\ \sum_{v=1}^{V} \phi_{kv} = 1$
$\boldsymbol{\Psi} = \{\boldsymbol{\theta}, \boldsymbol{\Phi}\}$	パラメータをまとめたもの	
α	トピック分布のパラメータ	
β	単語分布のパラメータ	
$q_{dk} = (q_{11}, \cdots, q_{DK})$	負担率	$q_{dk} \geq 0,\ \sum_{k=1}^{K} q_{dk} = 1$

3.1 混合ユニグラムモデル

混合ユニグラムモデルでは、複数のトピックの中から各文書は 1 つのトピックを持ち、またそのトピックが持つ単語分布に従って各単語が生成されているとする。

文書 d のトピック z_d の単語分布 $\boldsymbol{\phi}_{z_d}$ に従って単語が生成されると仮定する。また、パラメータ $\boldsymbol{\theta}, \boldsymbol{\Phi}$ は共役事前分布であるディリクレ分布から生成されると仮定する。

パラメータ $\boldsymbol{\theta}, \boldsymbol{\Phi}$ が与えられたときの文書 $\boldsymbol{w}_d$ の確率は以下の通りとなる。

$$
\begin{aligned}
p(\boldsymbol{w}_d|\boldsymbol{\theta},\boldsymbol{\Phi}) &= \sum_{k=1}^{K} p(\boldsymbol{w}_d, z_d = k|\boldsymbol{\theta},\boldsymbol{\Phi}) && \text{(i)} \\
&= \sum_{k=1}^{K} p(z_d = k|\boldsymbol{\theta}) p(\boldsymbol{w}_d|\boldsymbol{\phi}_k) && \text{(ii)} \\
&= \sum_{k=1}^{K} \theta_k \prod_{n=1}^{N_d} \phi_{kw_{dn}} && \text{(iii)} \\
&= \sum_{k=1}^{K} \theta_k \prod_{n=1}^{V} \phi_{kv}^{N_{dv}}
\end{aligned}
$$

【途中式の途中式】

- (i)：
  - $p(\boldsymbol{w}_d, z_d|\boldsymbol{\theta},\boldsymbol{\Phi})$ は、文書 $\boldsymbol{w}_d$ とそのトピック k の同時分布
  - これをトピック z_d に関して周辺化したものが $\sum_{k=1}^{K} p(\boldsymbol{w}_d, z_d = k|\boldsymbol{\theta},\boldsymbol{\Phi})$
  - (トピック 1 から K によって文書 $\boldsymbol{w}_d, z_d$ が決まるので、その全てについて加重平均をとるような操作)
- (ii)：
  - トピック分布 $\boldsymbol{\theta}$ によって文書 d のトピック z_d が 1~K の中から決まる。
  - また、トピック k が持つ単語分布 $\boldsymbol{\phi}_k$ によって文書 $\boldsymbol{w}_d$ が生成されるという過程から式を分割する。
- (iii)：
  - 1 つ目の項はトピック k となる確率なので、$p(k|\boldsymbol{\theta}) = \theta_k$ であるため、置き換える。
  - 2 つ目の項は単語分布に従って各文書が独立に生成されることから、$p(w_{d1}, \cdots, w_{dN_d}|\boldsymbol{\phi}_k) = \prod_{n=1}^{N_d} \phi_{kw_{dn}}$ であるため、置き換える。
- (iv)：単語から語彙に変換する。

単語集合 $\boldsymbol{w}_d$ は重複を許す集合である。一方、語彙は重複しない。そこで、語彙単位で w_{d1} から w_{dN_D} までを全て掛け合わせたものを語彙単位に変換する場合は、重複する語彙について累乗の形でまとめることができる。例えば、文書 d における単語 1 と 5 が語彙 1 であった場合の出現確率は $\phi_{kw_{d1}} * \phi_{kw_{d5}} = \phi_{k1} * \phi_{k1} = \phi_{k1}^2$ となる。また、文書 d における語彙 v の出現回数は N_{dv} であることから、$\phi_{kv}^{N_{dv}}$ となる

3.2 混合モデル

混合モデルとは、複数ある確率モデルの中から 1 つのモデルに従ってデータが生成されると仮定するモデルのことである。

混合ユニグラムモデルでは、複数のトピック (単語を生成する確率モデル) の中から、各文書は 1 つのトピックを持ち、単語を生成すると仮定している。それを式で表すと以下となる。

$$
p(\boldsymbol{w}) = \sum_{k=1}^{K} p(k) p(\boldsymbol{w}|k)
$$

$p(k)$ は混合比と呼ばれ、ここではトピック k が選ばれる確率を表す。$p(\boldsymbol{w}|k)$ は k が選ばれたときの観測 $\boldsymbol{w}$ が生成される確率を表す。ここでは、トピック k が選ばれたときに生成される文書 $\boldsymbol{w}$ の確率を表す。

また、混合ユニグラムモデルの場合は、$p(\boldsymbol{w}|k)$ はユニグラムモデルである。

3.3 EM アルゴリズム

ユニグラムモデルでは、対数尤度が最大となるパラメータを解析的に求めることができた。しかし、混合モデルでは解析的に得ることができない。そこで、EM アルゴリズムを用いて最尤推定を行う。
EM アルゴリズムとは、E(期待値) ステップと M(最大化) ステップを収束するまで繰り返すことでパラメータを推定する手法のことである。ここでは、対数尤度の下限を最大化することでパラメータの局所最適値を求める。

3.1 節より、混合ユニグラムモデルの対数尤度は次の式となる。

$$L = \sum_{d=1}^{D} \log p(\boldsymbol{w_d}|\boldsymbol{\theta}, \boldsymbol{\Phi}) = \sum_{d=1}^{D} \log \sum_{k=1}^{K} p(k|\boldsymbol{\theta}) p(\boldsymbol{w_d}|\boldsymbol{\phi}_k) \tag{3.1}$$

これに、文書 d がトピック k に属する確率である負担率 q_{dk} を掛け合わせる。

$$\begin{aligned} L &= \sum_{d=1}^{D} \log \sum_{k=1}^{K} q_{dk} \frac{p(k|\boldsymbol{\theta}) p(\boldsymbol{w_d}|\boldsymbol{\phi}_k)}{q_{dk}} && \text{(i)} \\ &\geq \sum_{d=1}^{D} \sum_{k=1}^{K} q_{dk} \log \frac{p(k|\boldsymbol{\theta}) p(\boldsymbol{w_d}|\boldsymbol{\phi}_k)}{q_{dk}} && \text{(ii)} \\ &= \sum_{d=1}^{D} \sum_{k=1}^{K} q_{dk} \log \frac{\theta_k \prod_{v=1}^{V} \phi_{kv}^{N_{dv}}}{q_{dk}} && \text{(iii)} \\ &= \sum_{d=1}^{D} \sum_{k=1}^{K} q_{dk} \left(\log \theta_k + \sum_{v=1}^{V} N_{dv} \log \phi_{kv} - \log q_{dk} \right) \equiv F && (3.2) \end{aligned}$$

【途中式の途中式】

- (0)：式 (3.1) に、負担率 q_{dk} を $\frac{q_{dk}}{q_{dk}} = 1$ として、これを分割して掛け合わせる。
- (i)：$\log \sum_{k=1}^{K} q_{dk} \frac{p(k|\boldsymbol{\theta}) p(\boldsymbol{W_d}|\boldsymbol{\phi}_k)}{q_{dk}}$ が上に凸な関数であることから、イェンゼンの不等式 (1.7) を用いて下限みる。
- (ii)：
  - トピック k が割り振られる確率は $p(k|\boldsymbol{\theta}) = \theta_k$ であるため、置き換える。
  - $p(\boldsymbol{w_d}|\boldsymbol{\phi}_k) = \prod_{v=1}^{V} \phi_{kv}^{N_{dv}}$ であるため、置き換える。
- (iii)：$\log \frac{A*B}{C} = \log A + \log B - \log C$ であり、また $\log x^a = a \log x$ であることから式を変形する。

対数尤度の下限が得られた。次からは、この下限 F を最大化するパラメータを推定していく。

・E ステップ

E ステップでは、下限 F が最大となる負担率 q_{dk} を求める。

$\sum_{k=1}^{K} q_{dk} = 1$ の制約の下で、式 (3.2) が最大となる q_{dk} をラグランジュの未定乗数法を用いて解いていく。

$$F(q_{dk}) = q_{dk}\left(\log\theta_k \prod_{v=1}^{V}\phi_{kv}^{N_{dv}} - \log q_{dk}\right) + \lambda\left(\sum_{k=1}^{K} q_{dk} - 1\right) \tag{3.3}$$

これを q_{dk} について微分して、$\frac{\partial F(q_{dk})}{\partial q_{dk}} = 0$ となる q_{dk} を求める。

$$\frac{\partial F(q_{dk})}{\partial q_{dk}} = \log\theta_k \prod_{v=1}^{V}\phi_{kv}^{N_{dv}} - \log q_{dk} - 1 + \lambda = 0 \tag{i}$$

$$\log q_{dk} = \log\theta_k \prod_{v=1}^{V}\phi_{kv}^{N_{dv}} + \lambda - 1 \tag{ii}$$

$$q_{dk} = \exp(\lambda - 1)\theta_k \prod_{v=1}^{V}\phi_{kv}^{N_{dv}} \tag{3.5}$$

【途中式の途中式】

- (i)：

式 (3.3) を展開すると

$$F(q_{dk}) = q_{dk}\left(\log\theta_k \prod_{v=1}^{V}\phi_{kv}^{N_{dv}}\right) - q_{dk}\log q_{dk} + (\lambda q_{d1} + \cdots + \lambda q_{dk} + \cdots + \lambda q_{dK}) - \lambda$$

である。

$-q_{dk}\log q_{dk}$ の項を q_{dk} について微分すると、積の微分 $(f(x)g(x))' = f'(x)g(x) + f(x)g'(x)$ と対数関数の微分 $(\log x)' = \frac{1}{x}$ より

$$\begin{aligned}(-q_{dk}\log q_{dk})' &= (-q_{dk})'\log q_{dk} + (-q_{dk})(\log q_{dk})' \\ &= (-1) * \log q_{dk} - q_{dk} * \frac{1}{q_{dk}} \\ &= -\log q_{dk} - 1\end{aligned}$$

になる。

$\lambda\left(\sum_{k=1}^{K} q_{dk} - 1\right)$ の部分は、q_{dk} 以外の λq_{d1} から λq_{dK} は定数扱いとなり、$\sum_{k=1}^{K}\lambda q_{dk}$ は $0 + \cdots + (\lambda * 1) + \cdots + 0 = \lambda$ になる。

- (ii)：両辺で指数をとる。

両辺で $k=1$ から K まで和をとると、$\sum_{k=1}^{K} q_{dk} = 1$ より

$$\begin{aligned}
\sum_{k=1}^{K} q_{dk} &= \sum_{k=1}^{K} \exp(\lambda-1)\theta_k \prod_{v=1}^{V} \phi_{kv}^{N_{dv}} \\
1 &= \exp(\lambda-1) \sum_{k=1}^{K} \theta_k \prod_{v=1}^{V} \phi_{kv}^{N_{dv}} \\
\exp(\lambda-1) &= \frac{1}{\sum_{k=1}^{K} \theta_k \prod_{v=1}^{V} \phi_{kv}^{N_{dv}}} \qquad (3.6)
\end{aligned}$$

となる。これを式 (3.5) に代入すると、

$$q_{dk} = \frac{\theta_k \prod_{v=1}^{V} \phi_{kv}^{N_{dv}}}{\sum_{k'=1}^{K} \theta_{k'} \prod_{v=1}^{V} \phi_{k'v}^{N_{dv}}} \qquad (3.3)$$

が得られる。この式が負担率 q_{dk} の更新式となる。

次では、更新した負担率を使いパラメータを更新する。

・M ステップ

M ステップでは、負担率が与えられた下での下限 F が最大になるようにパラメータ $\boldsymbol{\theta}, \boldsymbol{\phi}_k$ を更新する。

E ステップと同様に、$\sum_{k=1}^{K} \theta_k = \sum_{k=1}^{K} \phi_{kv} = 1$ の制約の下で、式 (3.2) が最大となるパラメータ $\boldsymbol{\theta}$, $\boldsymbol{\phi}_k$ を、ラグランジュの未定乗数法を用いて求めていく。

$$F(\theta_k) = \sum_{d=1}^{D} q_{dk} \left(\log \theta_k + \sum_{v=1}^{V} N_{dv} \log \phi_{kv} - \log q_{dk} \right) + \lambda \left(\sum_{k=1}^{K} \theta_k - 1 \right)$$

θ_k に関して微分して、$\frac{\partial F(\theta_k)}{\partial \theta_k} = 0$ となる θ_k を求める。

$$\begin{aligned}
\frac{\partial F(\theta_k)}{\partial \theta_k} &= \frac{\sum_{d=1}^{D} q_{dk}}{\theta_k} + \lambda = 0 \\
\theta_k &= -\sum_{d=1}^{D} \frac{q_{dk}}{\lambda} \qquad (1)
\end{aligned}$$

両辺で $k=1$ から K まで和をとると、$\sum_{k=1}^{K} \theta_k = 1$ より

$$
\begin{aligned}
\sum_{k=1}^{K} \theta_k &= -\sum_{k=1}^{K}\sum_{d=1}^{D} \frac{q_{dk}}{\lambda} \\
1 &= -\sum_{k=1}^{K}\sum_{d=1}^{D} \frac{q_{dk}}{\lambda} \\
\lambda &= \sum_{k=1}^{K}\sum_{d=1}^{D} q_{dk}
\end{aligned}
$$

となる。これを (1) に代入すると

$$
\begin{aligned}
\theta_k &= -\sum_{d=1}^{D} \frac{q_{dk}}{-\sum_{k=1}^{K}\sum_{d=1}^{D} q_{dk}} \\
&= \frac{\sum_{d=1}^{D} q_{dk}}{\sum_{k=1}^{K}\sum_{d=1}^{D} q_{dk}} \qquad (3.7)
\end{aligned}
$$

が得られる。この式がパラメータ θ_k の更新式となる。

ϕ_{kv} についても同様に求めていく。

$$
F(\phi_{kv}) = \sum_{d=1}^{D} \left(\log \theta_k + \sum_{v=1}^{V} N_{dv} \log \phi_{kv} - \log q_{dk} \right) + \lambda \left(\sum_{k=1}^{K} \phi_{kv} - 1 \right)
$$

ϕ_{kv} について微分して、$\frac{\partial F(\phi_{kv})}{\partial \theta_k} = 0$ となる θ_k を求める。

$$
\begin{aligned}
\frac{\partial F(\phi_{kv})}{\partial \phi_{kv}} = \frac{\sum_{d=1}^{D} q_{dk} N_{dv}}{\phi_{kv}} + \lambda = 0 \\
\phi_{kv} = -\sum_{d=1}^{D} \frac{q_{dk} N_{dv}}{\lambda} \qquad (2)
\end{aligned}
$$

両辺で $v = 1$ から V まで和をとると、$\sum_{v=1}^{V} \phi_{kv} = 1$ より

$$
\begin{aligned}
\sum_{v=1}^{V} \phi_{kv} &= -\sum_{v=1}^{V}\sum_{d=1}^{D} \frac{q_{dk} N_{dv}}{\lambda} \\
1 &= -\sum_{v=1}^{V}\sum_{d=1}^{D} \frac{q_{dk} N_{dv}}{\lambda} \\
\lambda &= -\sum_{v=1}^{V}\sum_{d=1}^{D} q_{dk} N_{dv}
\end{aligned}
$$

となる。これを (3) に代入すると

$$
\begin{aligned}
\phi_{kv} &= -\frac{\sum_{d=1}^{D} q_{dk} N_{dv}}{-\sum_{v=1}^{V}\sum_{d=1}^{D} q_{dk} N_{dv}} \\
&= \frac{\sum_{d=1}^{D} q_{dk} N_{dv}}{\sum_{v=1}^{V}\sum_{d=1}^{D} q_{dk} N_{dv}} \qquad (3.8)
\end{aligned}
$$

が求まる。この式がパラメータ ϕ_{kv} の更新式となる。

具体的な最大化のための更新手順については、参考書のアルゴリズム 3.1 を参照のこと。

・対数尤度 L と下限 F の関係性

対数尤度 L と下限 F との差を求めることで関係をみる。

$$
\begin{aligned}
L - F &= \sum_{d=1}^{D}\sum_{k=1}^{K} q_{dk} \log p(\boldsymbol{w}_d|\boldsymbol{\theta},\boldsymbol{\Phi}) - \sum_{d=1}^{D}\sum_{k=1}^{K} q_{dk} \log \frac{p(k|\boldsymbol{\theta})p(\boldsymbol{w}_d|k,\boldsymbol{\Phi})}{q_{dk}} && \text{(i)} \\
&= \sum_{d=1}^{D}\sum_{k=1}^{K} q_{dk} \left(\log \frac{p(k|\boldsymbol{\theta})p(\boldsymbol{w}_d|k,\boldsymbol{\Phi})}{p(\boldsymbol{w}_d|\boldsymbol{\theta},\boldsymbol{\Phi})} - \log \frac{p(k|\boldsymbol{\theta})p(\boldsymbol{w}_d|k,\boldsymbol{\Phi})}{q_{dk}} \right) && \text{(ii)} \\
&= \sum_{d=1}^{D}\sum_{k=1}^{K} q_{dk} \log \left(\frac{p(k|\boldsymbol{\theta})p(\boldsymbol{w}_d|k,\boldsymbol{\Phi})}{p(\boldsymbol{w}_d|\boldsymbol{\theta},\boldsymbol{\Phi})} \frac{q_{dk}}{p(k|\boldsymbol{\theta})p(\boldsymbol{w}_d|k,\boldsymbol{\Phi})} \right) && \text{(iii)} \\
&= \sum_{d=1}^{D}\sum_{k=1}^{K} q_{dk} \log \frac{q_{dk}}{p(k|\boldsymbol{w}_d,\boldsymbol{\theta},\boldsymbol{\Phi})} && \text{(iv)} \\
&= \sum_{d=1}^{D} \mathrm{KL}(\boldsymbol{q}_d, p(z_d|\boldsymbol{w}_d,\boldsymbol{\theta},\boldsymbol{\Phi})) && (3.9)
\end{aligned}
$$

【途中式の途中式】

- (0)：両式の形を揃えるために、対数尤度 L に $\sum_{k=1}^{K} q_{dk} = 1$ を掛け合わせている。
- (i)：ベイズの定理を用いて、$p(\boldsymbol{w}_d|\boldsymbol{\theta},\boldsymbol{\Phi}) = \frac{p(k|\boldsymbol{\theta})p(\boldsymbol{w}_d|k,\boldsymbol{\Phi})}{p(\boldsymbol{w}_d|\boldsymbol{\theta},\boldsymbol{\Phi})}$ であるため、置き換える。
- (ii)：$\log A - \log B = \log \frac{A}{B} = \log A \frac{1}{B}$ であることから、式をを変形する。
- (iii)：約分して式を整理する。
- (iv)：
  - $k = 1$ から K まで和をとる。
  - KL ダイバージェンス (1.6) より。

以上のことから、下限 F の最大化は負担率とトピックの事後分布の KL ダイバージェンスの最小化と言える。

・R で組んでみる

・コード全体

・テキスト処理

```
# 利用パッケージ
library(RMeCab)
library(tidyverse)

### テキスト処理

# 抽出しない単語を指定
stop_words <- "[a-z]" # 小文字のアルファベットを含む語

# 形態素解析
mecab_df <- docDF("フォルダ名", type = 1) # テキストファイルの保存先を指定する

# 文書 d の語彙 v の出現回数 (N_dv) の集合
N_dv <- mecab_df %>%
        filter(grepl("名詞|形容詞|^動詞", POS1)) %>% # 抽出する品詞 (大分類) を指定する
        filter(grepl("一般|^自立", POS2)) %>%         # 抽出する品詞 (細分類) を指定する
        filter(!grepl(stop_words, TERM)) %>%         # ストップワードを除く
        select(-c(TERM, POS1, POS2)) %>%             # 数値列のみを残す
        filter(apply(., 1, sum) >= 5) %>%            # 抽出する総出現回数を指定する
        t()                                          # 転置

# 確認用の行列名
dimnames(N_dv) <- list(paste0("d=", 1:nrow(N_dv)), # 行名
                       paste0("v=", 1:ncol(N_dv))) # 列名

# 文書 d の単語数 (N_d) のベクトル
N_d <- apply(N_dv, 1, sum) # 行方向に和をとる

# 文書数 (D)
D <- nrow(N_dv)

# 総語彙数 (V)
V <- ncol(N_dv)
```

・パラメータの初期設定

```
### パラメータの初期設定

# トピック数 (K)
K <- 4   # 任意の値を指定する

# 負担率 (q_dk) の集合
q_dk <- matrix(0, nrow = D, ncol = K,
               dimnames = list(paste0("d=", 1:D),  # 行名
                               paste0("k=", 1:K))) # 列名

## トピック分布 (theta)
# 初期値をランダムに設定
theta_k <- sample(seq(0.1, 10, 0.1), size = K, replace = TRUE)
```

```r
# 正規化：sum_{k=1}^K theta_k = 1
theta_k <- theta_k / sum(theta_k)
names(theta_k) <- paste0("k=", 1:K) # 確認用の要素名

## 単語分布 (phi)
# 初期値をランダムに設定
phi_kv <- sample(seq(0.1, 10, 0.1), size = K * V, replace = TRUE) %>%
          matrix(nrow = K, ncol = V,
                 dimnames = list(paste0("k=", 1:K),  # 行名
                                 paste0("v=", 1:V))) # 列名

# 正規化：sum_{v=1}^V phi_{kv} = 1
for(k in 1:K) {
  phi_kv[k, ] <- phi_kv[k, ] / sum(phi_kv[k, ])
}
```

・最尤推定

```r
### 最尤推定

# 推移の確認用
trace_theta <- as.matrix(theta_k)
trace_phi   <- as.matrix(phi_kv[1, ])

for(i in 1:50){ # 任意の回数を指定する

  # 次ステップのパラメータを初期化
  new_theta_k <- rep(0, K)
  names(new_theta_k) <- paste0("k=", 1:K)

  new_phi_kv <- matrix(0, nrow = K, ncol = V,
                       dimnames = list(paste0("k=", 1:K),
                                       paste0("v=", 1:V)))

  for(d in 1:D){ ## （各文書）

    for(k in 1:K){ ## （各トピック）

      # 負担率の計算
      tmp_phi_k <- t(phi_kv) ^ N_dv[d, ] %>%
                   apply(2, prod)
      q_dk[d, k] <- theta_k[k] * tmp_phi_k[k] / sum(theta_k * tmp_phi_k)

      # トピック分布 (theta_k) を更新
      new_theta_k[k] <- new_theta_k[k] + q_dk[d, k]

      for(v in 1:V){ ## （各語彙）

        # 単語分布 (phi_kv) を更新
        new_phi_kv[k, v] <- new_phi_kv[k, v] + q_dk[d, k] * N_dv[d, v]

      } ## （/各語彙）
    } ## （/各トピック）
  } ## （/各文書）
```

```r
  # パラメータの更新
  theta_k <- new_theta_k
  phi_kv  <- new_phi_kv

  # パラメータの正規化
  theta_k <- theta_k / sum(theta_k)
  for(k2 in 1:K) {
    phi_kv[k2, ] <- phi_kv[k2, ] / sum(phi_kv[k2, ])
  }

  # 推移の確認用
  trace_theta <- cbind(trace_theta, as.matrix(theta_k))
  trace_phi   <- cbind(trace_phi, as.matrix(phi_kv[1, ]))
}
```

・コードの解説

・利用パッケージ

```r
# 利用パッケージ
library(RMeCab)
library(tidyverse)
```

形態素解析のための `RMeCab::docDF()` と、グラフ作成時の `tidyr::gather()`、`{ggplot2}` 関連の関数以外にライブラリの読み込みが必要なのは `dplyr` のみ。(面倒なので `tidyvers` パッケージを読み込んでいる)

・テキスト処理

・形態素解析

```r
# 形態素解析
mecab_df <- docDF(" フォルダ名", type = 1) # テキストファイルの保存先を指定する
```

`RMeCab::docDF()` で、形態素解析を行う。第 1 引数に解析を行うテキストファイルがあるディレクトリのパスを指定する。引数に `type = 1` を指定すると、テキストを単語に分割する。デフォルトの設定は 0 で、`type = 0` だと文字に分割する。

```r
head(mecab_df[, 1:7])
```

```
##   TERM POS1     POS2 juice_01.txt juice_02.txt juice_03.txt juice_04.txt
## 1    ! 名詞 サ変接続            0            0            0            0
## 2   !? 名詞 サ変接続            0            0            0            0
## 3    & 名詞 サ変接続            0            0            0            0
## 4    ' 名詞 サ変接続            0            0            0            0
## 5    ( 名詞 サ変接続            0            0            0            0
## 6    ) 名詞 サ変接続            0            0            0            0
```

解析結果はこのように、単語とその品詞、文書ごとの頻度の列からなるデータフレームで出力される。(表示は 7 列目までに省略している)

・特徴語の抽出

```
# 抽出しない単語を指定
stop_words <- "[a-z]" # 小文字のアルファベットを含む語

# 文書 d の語彙 v の出現回数 (N_dv) の集合
N_dv <- mecab_df %>%
        filter(grepl("名詞|形容詞|^動詞", POS1)) %>% # 抽出する品詞 (大分類) を指定する
        filter(grepl("一般|^自立", POS2)) %>%        # 抽出する品詞 (細分類) を指定する
        filter(!grepl(stop_words, TERM)) %>%        # ストップワードを除く
        select(-c(TERM, POS1, POS2)) %>%            # 数値列のみを残す
        filter(apply(., 1, sum) >= 5) %>%           # 抽出する総出現回数を指定する
        t()                                         # 転置

# 確認用の行列名
dimnames(N_dv) <- list(paste0("d=", 1:nrow(N_dv)), # 行名
                       paste0("v=", 1:ncol(N_dv))) # 列名
```

形態素解析の結果から必要な情報を抽出する。

POS1 列 (品詞大分類) と POS2 列 (品詞小分類) の情報を用いて、filter(grepl()) に品詞を指定し必要な語彙 (行) を取り出す。ここでは、助動詞を含まないようにするために**動詞**の前に文字列の先頭であることを示す正規表現の記号の ^ を付けている。同様に非自立を除くため、**自立**の頭にも付けている。

grepl() の頭に ! を付けることで、grepl() に指定した条件以外の行を取り出すことができる。ここでは、小文字のアルファベットの a から z までを表す正規表現 [a-z] を使って、アルファベットを含む語を除く。

apply() 関数は、apply(**データ，向き，関数名**) とすることで、行列等のデータに対して指定した向き (1 なら行、2 なら列) に指定した関数を適応する。第 1 引数の . はパイプ (%>%) で渡している元のデータを指定する省略記号である。つまり、mecab_df の各行の要素に対して sum() を行うことで、文書全体での各語彙の総出現回数 N_v を求めている。その結果が、指定した回数以上である行 (語彙) を filter() で抽出する。

語彙が行、文書が列となっているので、t() で転置して N_{dv} とする。

dimnames() で行名と列名を付ける。各次元の名前の要素は list 形式で渡す。ただし、これは目で確認するためのもので、付けなくても処理に影響しない。他の変数についても同様である。

```
head(N_dv[, 1:15])
```

```
##     v=1 v=2 v=3 v=4 v=5 v=6 v=7 v=8 v=9 v=10 v=11 v=12 v=13 v=14 v=15
## d=1   0   0   0   0   0   0   0   0   0    0    0    0    0    4    0
## d=2   0   0   0   0   0   0   1   0   0    0    0    0    0    1    0
## d=3   0   0   0   0   0   0   0   0   0    1    0    0    0    7    0
## d=4   0   0   0   0   0   0   0   1   0    5    0    0    0    2    4
## d=5   0   0   0   0   0   0   0   0   0    1    0    0    0    0    0
## d=6   0   1   0   0   0   0   0   1   2    0    0    0    0    1    0
```

15 列目までに省略して表示している。

```r
# 文書dの単語数(N_d)のベクトル
N_d <- apply(N_dv, 1, sum) # 行方向に和をとる
```

apply(N_dv, 1, sum) で、N_dv の各行の要素に対して sum() を行い $N_d = \sum_{v=1}^{V} N_{dv}$ の計算をして、各文書の単語数 N_d を求める。(この節では利用しない)

```r
N_d
```

```
##  d=1  d=2  d=3  d=4  d=5  d=6  d=7  d=8  d=9 d=10 d=11 d=12 d=13 d=14 d=15
##   45   28   34   34   28   47    8   60   36   59   37   49   24   41   61
## d=16 d=17 d=18 d=19 d=20 d=21 d=22 d=23 d=24 d=25 d=26 d=27 d=28 d=29 d=30
##   36   27   25   38   34   28   54   48   47   39   23   32   40   35   20
## d=31 d=32 d=33 d=34 d=35 d=36 d=37 d=38 d=39 d=40 d=41 d=42 d=43
##   24   30   33   40   48   45   80   16   47   30   89   45   36
```

要素が 0 になった文書があると、推定時にエラーが起きることがあるので、語彙を抽出する条件を緩めるかその文書を除いて次に進む。

```r
# 文書数(D)
D <- nrow(N_dv)

# 総語彙数(V)
V <- ncol(N_dv)
```

N_dv の行数が文書数 D に、列数が語彙数 V に相当する。

・パラメータの初期設定

・トピック数

```r
# トピック数(K)
K <- 4   # 任意の値を指定する
```

任意のトピック数を指定する。

```r
# 負担率(q_dk)の集合
q_dk <- matrix(0, nrow = D, ncol = K,
               dimnames = list(paste0("d=", 1:D),  # 行名
                               paste0("k=", 1:K))) # 列名
```

全ての要素が 0 である D 行 K 列の行列を作成する。

行列 (マトリクス) は、matrix() を使って作る。第 1 引数に 0 を指定し、行数を指定する引数 nrow に文書数の D を、列数を指定する引数 ncol にトピック数の K を指定する。
行名・列名は dimneames = 引数に list 形式で指定する。

```r
## トピック分布 (theta)
# 初期値をランダムに設定
theta_k <- sample(seq(0.1, 10, 0.1), size = K, replace = TRUE)

# 正規化：sum_{k=1}^K theta_k = 1
theta_k <- theta_k / sum(theta_k)
names(theta_k) <- paste0("k=", 1:K) # 確認用の要素名
```

初期値をランダムに設定する。

`seq(最小値, 最大値, 間隔)` でランダムに取り得る範囲を指定する。そこから、`sample()` を使って、ランダムに値を取る。第 2 引数の `size` に `K`(トピック数) を指定し、要素が K 個のベクトルを作成する。`replace = TRUE` を指定すると、ランダムに取り得るに重複を許す

ただし、$\sum_{k=1}^K \theta_k = 1$ の条件を満たすために、合計値で割って正規化する。(`apply(dat, 1, sum)` 当たりで上手いことできると思うのだけどとりあえず `for()` で…)

```r
## 単語分布 (phi)
# 初期値をランダムに設定
phi_kv <- sample(seq(0.1, 10, 0.1), size = K * V, replace = TRUE) %>%
          matrix(nrow = K, ncol = V,
                 dimnames = list(paste0("k=", 1:K),  # 行名
                                 paste0("v=", 1:V))) # 列名

# 正規化：sum_{v=1}^V phi_{kv} = 1
for(k in 1:K) {
  phi_kv[k, ] <- phi_kv[k, ] / sum(phi_kv[k, ])
}
```

トピック分布と同様に、`sample()` で $K * V$ 個のランダムな値を持つベクトルを作り、`matrix()` に `nrow = K`(行数)、`ncol = V`(列数) を指定して、K 行 V 列の行列を作る。

$\sum_{v=1}^V \phi_{kv} = 1$ の条件を満たすために、各行に対して合計値で割って正規化する。

・最尤推定

以下、`for()` ループ内の処理の内容。

・パラメータの初期化

```r
# 次ステップのパラメータを初期化
new_theta_k <- rep(0, K)
names(new_theta_k) <- paste0("k=", 1:K)

new_phi_kv <- matrix(0, nrow = K, ncol = V,
                     dimnames = list(paste0("k=", 1:K),
                                     paste0("v=", 1:V)))
```

この記事の例では、変数の要素は随時置き換えるので初期化する必要はない。しかし初期化しなくとも、ループ処理の前に受け皿としてこの操作で変数を用意しなければならないので、どうせならば参考書のアルゴリズムの説明の通りのこの位置で処理しておく。あるいは、アルゴリズムの理解のために。

・負担率の計算

```
# 負担率の計算
tmp_phi_k <- t(phi_kv) ^ N_dv[d, ] %>%
              apply(2, prod)
q_dk[d, k] <- theta_k[k] * tmp_phi_k[k] / sum(theta_k * tmp_phi_k)
```

負担率は

$$q_{dk} = \frac{\theta_k \prod_{v=1}^V \phi_{kv}^{N_{dv}}}{\sum_{k'=1}^K \theta_{k'} \prod_{v=1}^V \phi_{k'v}^{N_{dv}}} \tag{3.3}$$

の計算式で求める。

まず $\prod_{v=1}^V \phi_{kv}^{N_{dv}}$ の部分の計算を行い、tmp_phi_k としておく。マトリクスとベクトルの計算は、ベクトルの要素をマトリクスの 1 列目の要素の上から 1 つずつ行われる。ベクトルの要素数よりもマトリクスの要素数の方が多い場合は、ベクトルの要素を繰り返し計算に再利用される。N_dv[d,] は V 個のベクトルになるため、phi_kv を t() で転置して列の要素が V 個のマトリクスとすることで、$\phi_{kv}^{N_{dv}}$ の計算ができる。tmp_phi_k は K 個のベクトルとなる。
続いて tmp_phi_k を、分子は特定のトピック k に関する要素のみ、分母は各トピックごとに theta_k と掛け合わせる。分母については全体を足し合わせるため sum() している。

・トピック分布の更新

```
# トピック分布 (theta_k) を更新
new_theta_k[k] <- new_theta_k[k] + q_dk[d, k]
```

θ_k に、今求めた負担率を加えて更新する。繰り返し足して更新していき、全文書分を終えた状態で最後に正規化を行う。

・単語分布の更新

```
# 単語分布 (phi_kv) を更新
new_phi_kv[k, v] <- new_phi_kv[k, v] + q_dk[d, k] * N_dv[d, v]
```

同様に、単語分布 ϕ_{kv} も各要素について負担率を加えていき更新する。全文書・語彙分を加えた状態、で最後に正規化を行う。
ただし、参考書では語彙 v(単語を重複せずに扱う集合で出現回数を考慮する) ではなく単語 w_{dn}(重複を許す単語集合で全ての出現回数が 1 回となる) を用いて $\phi_{kw_{dn}}$ としているが、(~~そっちの方が面倒なので~~) 語彙ごとに計算している。
単語から語彙に変換するために負担率に語彙の出現回数を掛けている。

ここまでの計算を各文書・各トピックに対して繰り返し行う。また、単語分布の更新は各語彙更新するため V 回繰り返し処理する。

・パラメータ更新

```
# パラメータの更新
theta_k <- new_theta_k
phi_kv  <- new_phi_kv

# パラメータの正規化
theta_k <- theta_k / sum(theta_k)
for(k2 in 1:K) {
  phi_kv[k2, ] <- phi_kv[k2, ] / sum(phi_kv[k2, ])
}
```

次の更新に用いるために、new_theta_k と new_phi_kv をそれぞれ theta_k と phi_kv に移す。theta_k, phi_kv は次の負担率の計算に用いる。new_theta_k, new_phi_kv は初期化して次の更新式の受け皿となる。

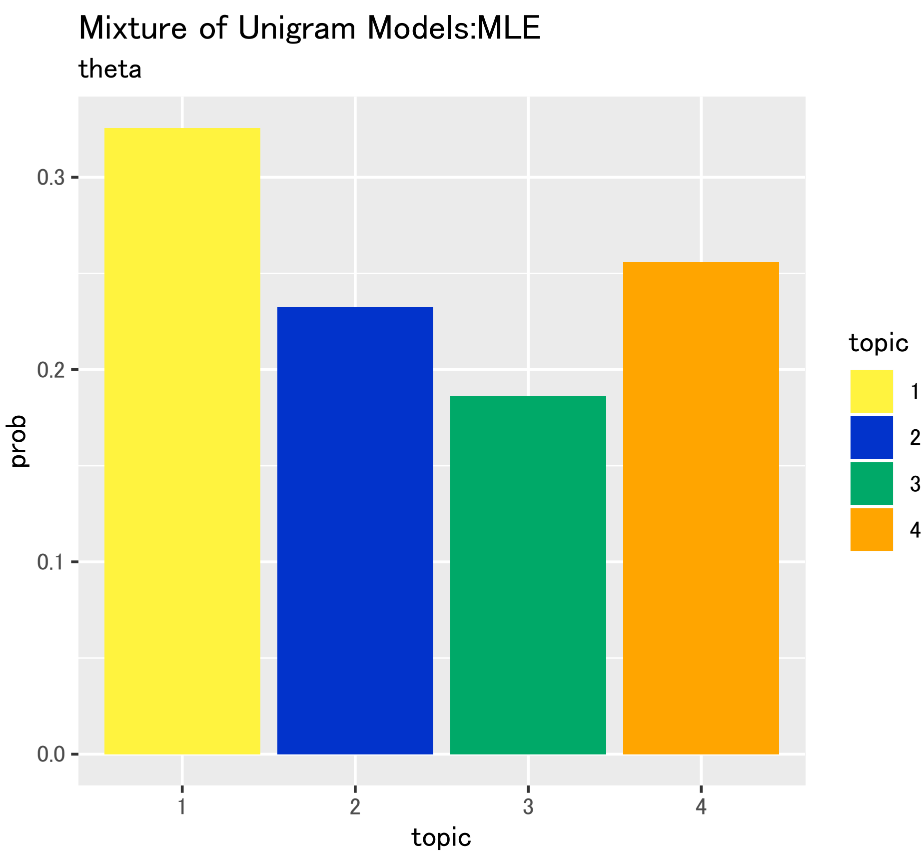
theta_k と phi_kv をそれぞれ $\sum_{k=1}^{K} \theta_k = \sum_{v=1}^{V} \phi_{kv} = 1$ となるように正規化する。

・推定結果の確認

・パラメータの推定結果の確認

・トピック分布

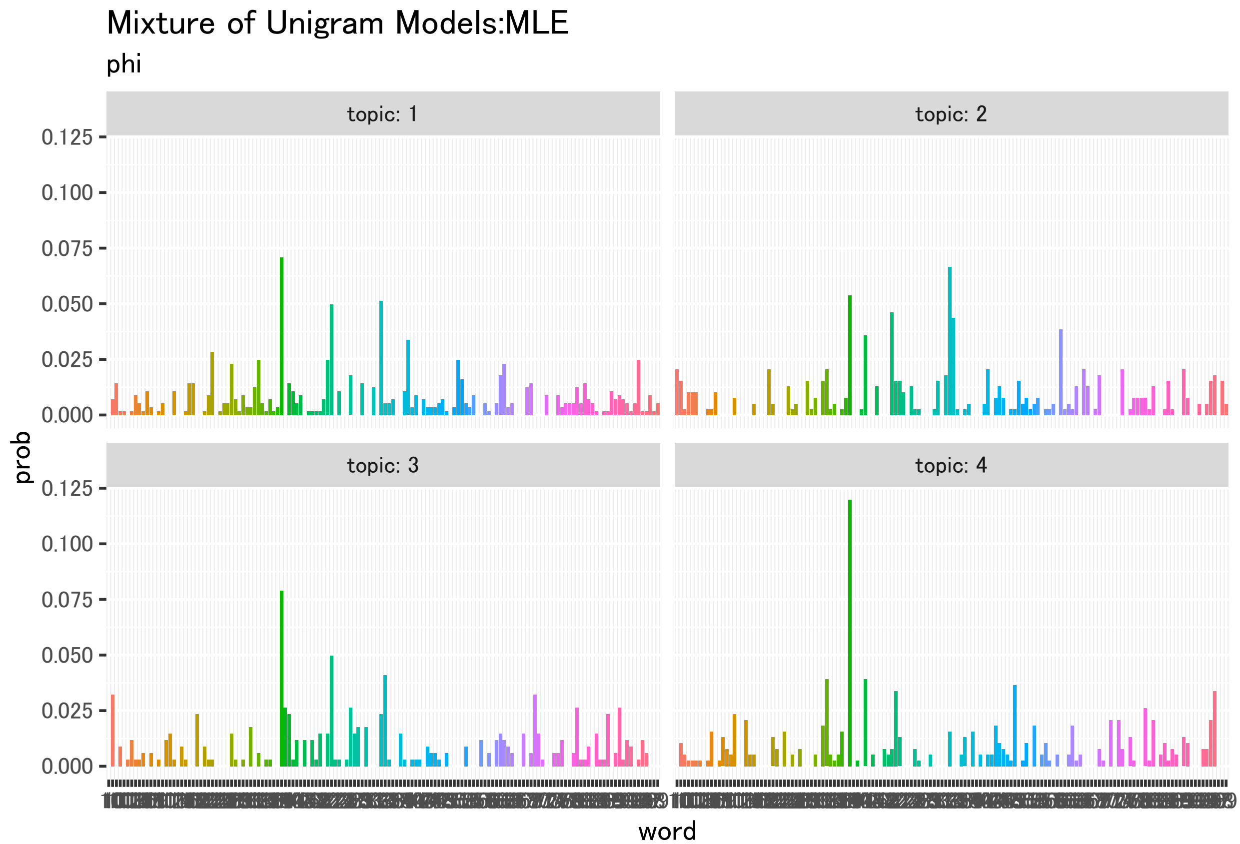
```
## トピック分布の推定結果を確認
# データフレームを作成
theta_df <- data.frame(topic = as.factor(1:K),
                       prob = theta_k)

# 描画
ggplot(data = theta_df, mapping = aes(x = topic, y = prob, fill = topic)) + # データ
  geom_bar(stat = "identity", position = "dodge") +                 # 棒グラフ
  labs(title = "Mixture of Unigram Models:MLE", subtitle = "theta") # タイトル
```



どのトピックになる確率が高いのかを確認できる。

・単語分布

```
## 単語分布の推定結果を確認
# データフレームを作成
phi_df_wide <- cbind(as.data.frame(phi_kv),
                     as.factor(1:K))

# データフレームを long 型に変換
colnames(phi_df_wide) <- c(1:V, "topic")  # key 用の行名を付与
phi_df_long <- gather(phi_df_wide, key = "word", value = "prob", -topic) # 変換
phi_df_long$term <- phi_df_long$word %>%  # 文字列になるため因子に変換
                    as.numeric() %>%
                    as.factor()

# 描画
ggplot(data = phi_df_long, mapping = aes(x = word,  y = prob, fill = word)) + # データ
  geom_bar(stat = "identity", position = "dodge") +  # 棒グラフ
  facet_wrap(~ topic, labeller = label_both) +       # グラフの分割
  theme(legend.position = "none") +                  # 凡例
  labs(title = "Mixture of Unigram Models:MLE", subtitle = "phi") # タイトル
```



トピックごとの語彙の出現確率を確認できる。

・推定推移の確認

・推移の記録

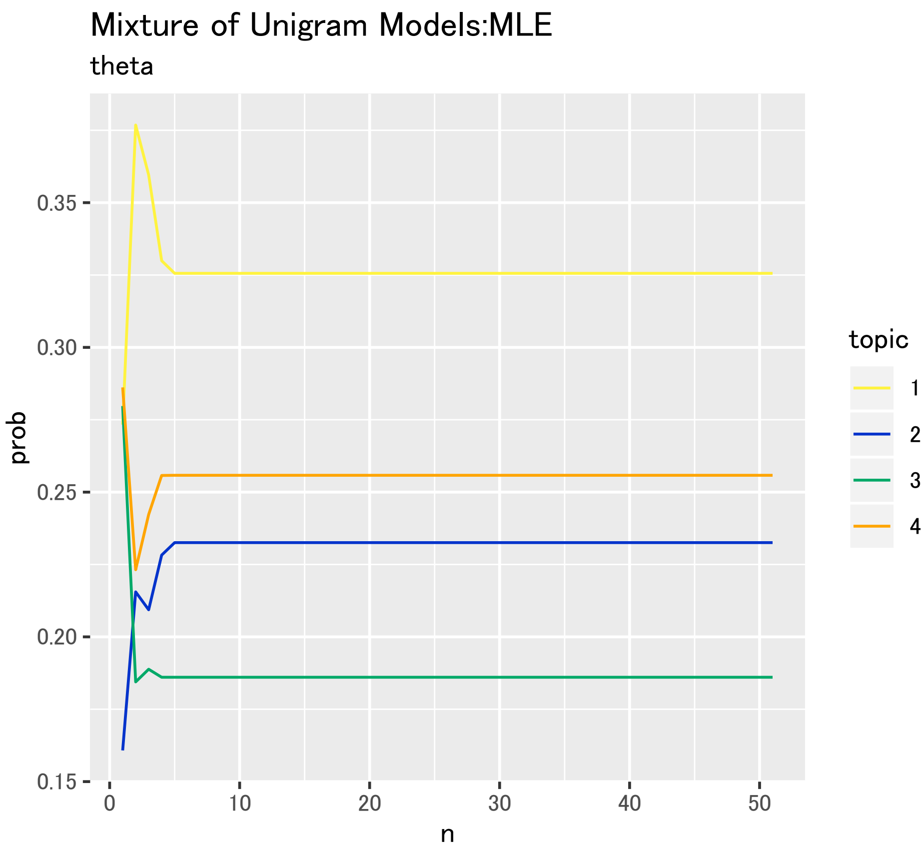
```
# 推移の確認用
trace_theta <- as.matrix(theta_k)
trace_phi   <- as.matrix(phi_kv[1, ])

# 推移の確認用
trace_theta <- cbind(trace_theta, as.matrix(theta_k))
trace_phi   <- cbind(trace_phi, as.matrix(phi_kv[1, ]))
```

推移を確認するために、パラメータを更新する度に記録しておく。単語分布については、トピック 1 のみ確認する。

・トピック分布

```
## トピック分布の推定値の推移を確認
# データフレームを作成
trace_theta_df_wide <- cbind(t(trace_theta),
                             1:ncol(trace_theta)) %>% # 推定回数
```

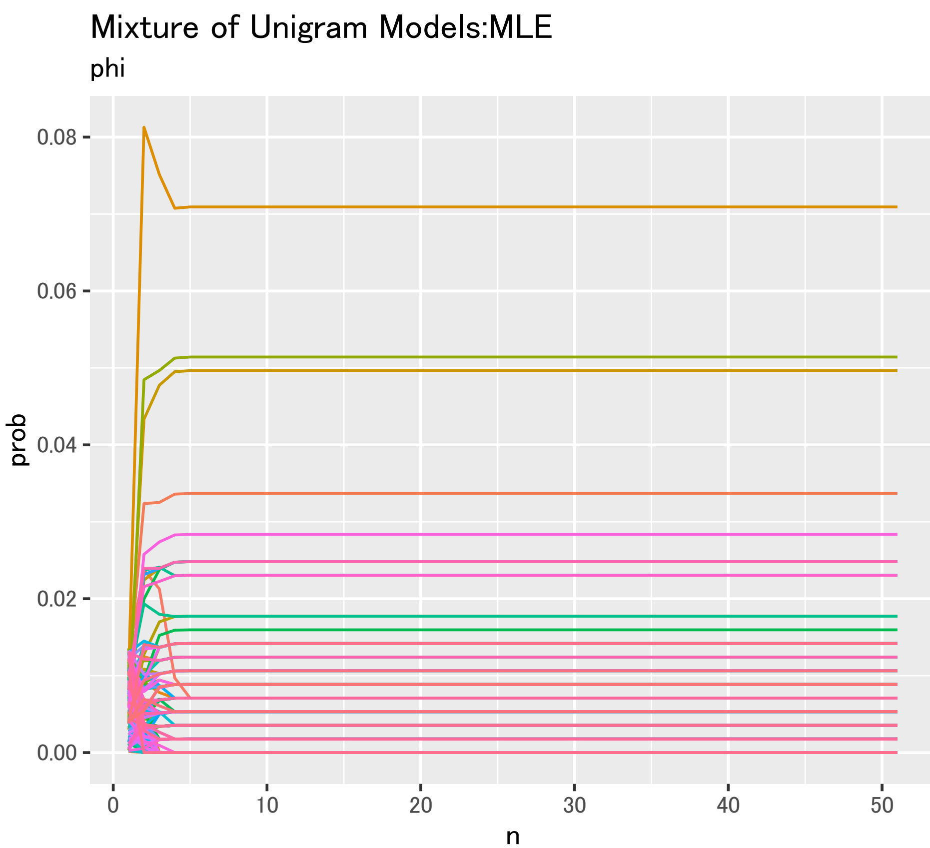
```r
                         as.data.frame()

# データフレームを long 型に変換
colnames(trace_theta_df_wide) <- c(1:K, "n") # key 用の行名を付与
trace_theta_df_long <- gather(trace_theta_df_wide, key = "topic", value = "prob", -n) # 変換

# 描画
ggplot(data = trace_theta_df_long, mapping = aes(x = n, y = prob, color = topic)) + # データ
  geom_line() +  # 折れ線グラフ
  labs(title = "Mixture of Unigram Models:MLE", subtitle = "theta") # タイトル
```



・単語分布

```r
## 単語分布の推定値の推移を確認
# データフレームを作成
trace_phi_df_wide <- cbind(t(trace_phi),
                           1:ncol(trace_phi)) %>%
                     as.data.frame()

# データフレームを long 型に変換
colnames(trace_phi_df_wide) <- c(1:V, "n")              # key 用の行名を付与
trace_phi_df_long <- gather(trace_phi_df_wide, key = "word", value = "prob", -n) # 変換
trace_phi_df_long$word <- trace_phi_df_long$word %>%  # 文字列になるため因子に変換
                          as.numeric() %>%
                          as.factor()
```

```
# 描画
ggplot(data = trace_phi_df_long, mapping = aes(x = n, y = prob, color = word)) + # データ
  geom_line() +                                # 折れ線グラフ
  theme(legend.position = "none") +   # 凡例
  labs(title = "Mixture of Unigram Models:MLE", subtitle = "phi") # タイトル
```

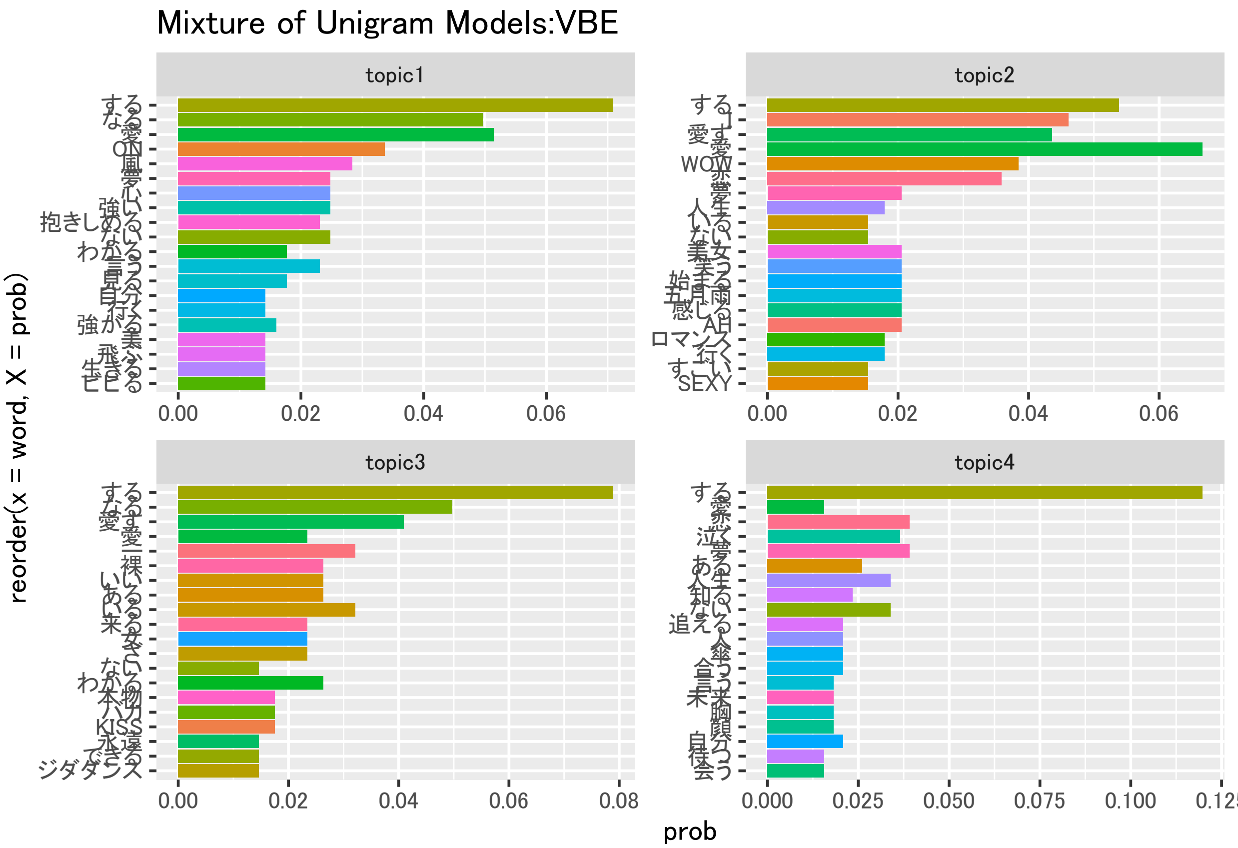


各パラメータが収束するまでの推移を確認できる。

・各トピックの出現確率の上位語の確認

```
### 各トピックの出現確率の上位語

## 語彙インデックス (v)
# 指定した出現回数以上の単語の行番号
num <- mecab_df %>%
       select(-c(TERM, POS1, POS2)) %>%
       apply(1, sum) >= 5 # 抽出する総出現回数を指定する

v_index <- mecab_df[num, ] %>%
           filter(grepl("名詞|形容詞|^動詞", POS1)) %>%  # 抽出する品詞を指定する
           filter(grepl("一般|^自立", POS2)) %>%
           filter(!grepl(stop_words, TERM)) %>%
```

```
          .[, 1]  # 単語の列を抽出する

# データフレームを作成
phi_df_wide2 <- cbind(as.data.frame(t(phi_kv)),
                      v_index) %>%
                as.data.frame()
colnames(phi_df_wide2) <- c(paste0("topic", 1:K), "word") # key 用の行名を付与

# データフレームの整形
phi_df_long2 <- data.frame()
for(i in 1:K) {
  tmp_df <- phi_df_wide2 %>%
    select(paste0("topic", i), word) %>%
    arrange(-.[, 1]) %>%  # 降順に並べ替え
    head(20) %>%          # 任意で指定した上位 n 語を抽出
    gather(key = "topic", value = "prob", -word) # long 型に変換

  phi_df_long2 <- rbind(phi_df_long2, tmp_df)
}

# 描画
ggplot(data = phi_df_long2,
       mapping = aes(x = reorder(x = word, X = prob), y = prob, fill = word)) +  # データ
  geom_bar(stat = "identity", position = "dodge") + # 棒グラフ
  coord_flip() +                                    # グラフの向き
  facet_wrap( ~ topic, scales = "free") +           # グラフの分割
  theme(legend.position = "none") +                 # 凡例
  labs(title = "Mixture of Unigram Models:VBE")     # タイトル
```



3.4 変分ベイズ

3.4.1 周辺尤度の最大化

混合ユニグラムモデルをベイズ推定することを考える。ユニグラムモデルでは、共役事前分布であるディリクレ分布を事前分布として用いることで、解析的に事後分布を計算できた。しかし混合ユニグラムモデルでは、解析的に事後分布を計算できない。そこで、変分ベイズ推定と呼ばれる方法により、事後分布の近似を計算する。

変分ベイズ推定では、周辺尤度

$$p(\boldsymbol{W}) = \int \sum_{\boldsymbol{z}} p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) d\boldsymbol{\Psi}$$

を最大化するトピックとパラメータの事後分布 $p(\boldsymbol{z}|\boldsymbol{W}), p(\boldsymbol{\Psi}|\boldsymbol{z})$ の近似を求める。ここで、$\sum_{\boldsymbol{z}} \equiv \sum_{z_1=1}^{K} \cdots \sum_{z_d=1}^{K} \cdots \sum_{z_D=1}^{K}$ はトピック集合の全ての組み合わせに関する和を表す。

この式の対数をとった対数尤度 $\log p(\boldsymbol{W})$ は、log の中に $\sum$ があり簡単に計算できない。そこで、イェンゼンの不等式を用いて、対数尤度の下限を求めて、下限を最大化する。

$$
\begin{aligned}
\log p(\boldsymbol{W}) &= \log \int \sum_{\boldsymbol{z}} p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) d\boldsymbol{\Psi} && \text{(i)} \\
&= \log \int \sum_{\boldsymbol{z}} q(\boldsymbol{z}, \boldsymbol{\Psi}) \frac{p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi})}{q(\boldsymbol{z}, \boldsymbol{\Psi})} d\boldsymbol{\Psi} && \text{(ii)} \\
&\geq \int \sum_{\boldsymbol{z}} q(\boldsymbol{z}, \boldsymbol{\Psi}) \log \frac{p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi})}{q(\boldsymbol{z}, \boldsymbol{\Psi})} d\boldsymbol{\Psi} \equiv F && (3.10)
\end{aligned}
$$

【途中式の途中式】

- (i)：求めたいトピックとパラメータの同時分布 $q(\boldsymbol{z}, \boldsymbol{\Psi})$ を $\frac{q(\boldsymbol{z}, \boldsymbol{\Psi})}{q(\boldsymbol{z}, \boldsymbol{\Psi})} = 1$ として、分割して掛け合わせる。
- (ii)：イェンゼンの不等式を用いて、下限を求める。

対数尤度の下限 F は変分下限と呼ばれる。変分ベイズ推定では、変分下限 F を最大化する事後分布の近似 $q(\boldsymbol{z}, \boldsymbol{\Psi})$ を求める。また、この事後分布の近似 $q(\boldsymbol{z}, \boldsymbol{\Psi})$ は変分事後分布と呼ばれる。

計算を簡単にするために、変分事後分布 $q(\boldsymbol{z}, \boldsymbol{\Psi})$ をトピックの変分事後分布 $q(\boldsymbol{z})$ とパラメータの変分事後分布 $q(\boldsymbol{\Psi})$ の積に分解できると仮定する。

$$
q(\boldsymbol{z}, \boldsymbol{\Psi}) = q(\boldsymbol{z}) q(\boldsymbol{\Psi}) \tag{3.11}
$$

対数周辺尤度と変分下限 F の差は

$$
\begin{aligned}
\log p(\boldsymbol{W}) - F &= \int \sum_{\boldsymbol{z}} q(\boldsymbol{z}, \boldsymbol{\Psi}) \log p(\boldsymbol{W}) d\boldsymbol{\Psi} - \int \sum_{\boldsymbol{z}} q(\boldsymbol{z}, \boldsymbol{\Psi}) \log \frac{p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi})}{q(\boldsymbol{z}, \boldsymbol{\Psi})} d\boldsymbol{\Psi} && \text{(i)} \\
&= \int \sum_{\boldsymbol{z}} q(\boldsymbol{z}, \boldsymbol{\Psi}) \log \frac{p(\boldsymbol{W}) q(\boldsymbol{z}, \boldsymbol{\Psi})}{p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi})} d\boldsymbol{\Psi} && \text{(ii)} \\
&= \int \sum_{\boldsymbol{z}} q(\boldsymbol{z}, \boldsymbol{\Psi}) \log \frac{q(\boldsymbol{z}, \boldsymbol{\Psi})}{p(\boldsymbol{z}, \boldsymbol{\Psi} | \boldsymbol{W})} d\boldsymbol{\Psi} && \text{(iii)} \\
&= \mathrm{KL}\Big(q(\boldsymbol{z}, \boldsymbol{\Psi}), p(\boldsymbol{z}, \boldsymbol{\Psi} | \boldsymbol{W}) \Big)
\end{aligned}
$$

【途中式の途中式】

- (0)：前の項は、両式の形を揃えるために対数周辺尤度 $\log p(\boldsymbol{W})$ に $\sum_{\boldsymbol{z}} q(\boldsymbol{z}) = 1$ と $\int q(\boldsymbol{\Psi}) d\boldsymbol{\Psi} = 1$ を掛け合わせたものである。
- (i)：$\log A - \log \frac{B}{C} = \log \frac{AC}{B}$ であることから、式を整理する。
- (ii)：乗法定理 (1.3) より、$\frac{p(A)}{p(A,B)} = \frac{1}{p(B|A)}$ であることから、式を変形する。
- (iii)：KL ダイバージェンス (1.6) より。

となる。

対数周辺尤度 $\log p(\boldsymbol{W})$ は変分事後分布 $q(\boldsymbol{z}, \boldsymbol{\Psi})$ に関して定数であるため、変分下限 F を最大にする $q(\boldsymbol{z}, \boldsymbol{\Psi})$ は、真の事後分布 $p(\boldsymbol{z}, \boldsymbol{\Psi} | \boldsymbol{W})$ との KL ダイバージェンスが最小となる。つまり、変分ベイズ推定において、変分下限を最大化することで、真の事後分布との KL ダイバージェンスが最小となる近似事後

分布が得られる。

3.4.2 変分事後分布の推定

トピックの変分事後分布 $q(\boldsymbol{z})$ の推定を行う。$\sum_{\boldsymbol{z}} q(\boldsymbol{z}) = 1$ の制約の下での変分下限 F の最大化をラグランジュの未定乗数法を用いて求めていく。

$$F(q(\boldsymbol{z})) = \int \sum_{\boldsymbol{z}} q(\boldsymbol{z}) q(\boldsymbol{\Psi}) \Big(\log p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) - \log q(\boldsymbol{z}) - \log q(\boldsymbol{\Psi}) \Big) d\boldsymbol{\Psi} + \lambda \left(\sum_{\boldsymbol{z}} q(\boldsymbol{z}) - 1 \right)$$

$F(q(\boldsymbol{z}))$ を $q(\boldsymbol{z})$ に関して変分する。

$$\begin{aligned}
\frac{\partial F(q(\boldsymbol{z}))}{\partial q(\boldsymbol{z})} &= \int q(\boldsymbol{\Psi}) \Big(\log p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) - \log q(\boldsymbol{z}) - \log q(\boldsymbol{\Psi}) \Big) + q(\boldsymbol{z}) q(\boldsymbol{\Psi}) \left(-\frac{1}{q(\boldsymbol{z})} \right) d\boldsymbol{\Psi} + \lambda && \text{(i)} \\
&= \int q(\boldsymbol{\Psi}) \Big(\log p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) - \log q(\boldsymbol{z}) - \log q(\boldsymbol{\Psi}) - 1 \Big) d\boldsymbol{\Psi} + \lambda && \text{(ii)} \\
&= \int q(\boldsymbol{\Psi}) \log p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) d\boldsymbol{\Psi} - \int q(\boldsymbol{\Psi}) \log q(\boldsymbol{z}) d\boldsymbol{\Psi} - \int q(\boldsymbol{\Psi}) \log q(\boldsymbol{\Psi}) d\boldsymbol{\Psi} - \int q(\boldsymbol{\Psi}) d\boldsymbol{\Psi} + \lambda && \text{(iii)} \\
&= \mathbb{E}_{q(\boldsymbol{\Psi})} \left[\log p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) \right] - \log q(\boldsymbol{z}) + \lambda - C && (3.12)
\end{aligned}$$

【途中式の途中式】

- (i)：$q(\boldsymbol{\Psi})$ でまとめて、式を整理する。
- (ii)：式を展開する。
- (iii)：
  - $\mathbb{E}_{q(x)}[f(x)] = \int q(x) f(x) dx$ は、分布 $q(x)$ を用いた場合の $f(x)$ の期待値である。
  - 連続確率変数の積分より、$\int q(\boldsymbol{\Psi}) d\boldsymbol{\Psi} = 1$ であるため、その項が他の $\boldsymbol{\Psi}$ に依存していなければ消える。
  - $q(\boldsymbol{z})$ と関係のない $\int q(\boldsymbol{\Psi}) \log q(\boldsymbol{\Psi}) d\boldsymbol{\Psi} - 1$ を定数 C と置く。

$\frac{\partial F(q(\boldsymbol{z}))}{\partial q(\boldsymbol{z})} = 0$ となる $q(\boldsymbol{z})$ を解く。

$$\begin{aligned}
\frac{\partial F(q(\boldsymbol{z}))}{\partial q(\boldsymbol{z})} =& \mathbb{E}_{q(\boldsymbol{\Psi})} \left[\log p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) \right] - \log q(\boldsymbol{z}) + \lambda - C = 0 && \text{(i)} \\
\log q(\boldsymbol{z}) &= \mathbb{E}_{q(\boldsymbol{\Psi})} \left[\log p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) \right] + \lambda - C && \text{(ii)} \\
q(\boldsymbol{z}) &= \exp \Big(\mathbb{E}_{q(\boldsymbol{\Psi})} \left[\log p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) \right] + \lambda - C \Big) && \text{(iii)} \\
&\propto \exp \Big(\mathbb{E}_{q(\boldsymbol{\Psi})} \left[\log p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) \right] \Big) && (3.13)
\end{aligned}$$

【途中式の途中式】

- (i)：式を整理する。
- (ii)：両辺の指数をとる。
- (iii)：$q(\boldsymbol{z})$ と関係のない項を省く。

同様に、パラメータの変分事後分布 $q(\boldsymbol{\Psi})$ の推定を行っていく。$\int q(\boldsymbol{\Psi})d\boldsymbol{\Psi}=1$ の制約の下での変分下限 F の最大化をラグランジュの未定乗数法を用いて求めていく。

$$F(q(\boldsymbol{\Psi}))=\int\sum_{\boldsymbol{z}}q(\boldsymbol{z})q(\boldsymbol{\Psi})\Big(\log p(\boldsymbol{W},\boldsymbol{z},\boldsymbol{\Psi})-\log q(\boldsymbol{z})-\log q(\boldsymbol{\Psi})\Big)d\boldsymbol{\Psi}+\lambda\left(\int q(\boldsymbol{\Psi})d\boldsymbol{\Psi}-1\right)$$

$F(q(\boldsymbol{\Psi}))$ を $q(\boldsymbol{\Psi})$ に関して変分する。

$$\begin{aligned}
\frac{\partial F(q(\boldsymbol{\Psi}))}{\partial q(\boldsymbol{\Psi})}&=\sum_{\boldsymbol{z}}q(\boldsymbol{z})\Big(\log p(\boldsymbol{W},\boldsymbol{z},\boldsymbol{\Psi})-\log q(\boldsymbol{z})-\log q(\boldsymbol{\Psi})\Big)+q(\boldsymbol{z})q(\boldsymbol{\Psi})\frac{1}{q(\boldsymbol{\Psi})}+\lambda && \text{(i)}\\
&=\sum_{\boldsymbol{z}}q(\boldsymbol{z})\Big(\log p(\boldsymbol{W},\boldsymbol{z},\boldsymbol{\Psi})-\log q(\boldsymbol{z})-\log q(\boldsymbol{\Psi})-1\Big)+\lambda && \text{(ii)}\\
&=\sum_{\boldsymbol{z}}q(\boldsymbol{z})\log p(\boldsymbol{W},\boldsymbol{z},\boldsymbol{\Psi})-\sum_{\boldsymbol{z}}q(\boldsymbol{z})\log q(\boldsymbol{z})-\sum_{\boldsymbol{z}}q(\boldsymbol{z})\log q(\boldsymbol{\Psi})-\sum_{\boldsymbol{z}}q(\boldsymbol{z})+\lambda && \text{(ii)}\\
&=\mathbb{E}_{q(\boldsymbol{z})}\left[\log p(\boldsymbol{W},\boldsymbol{z},\boldsymbol{\Psi})\right]-\log q(\boldsymbol{\Psi})-\lambda-C && (3.14)
\end{aligned}$$

【途中式の途中式】

- (i)：$q(\boldsymbol{z})$ でまとめて、式を整理する。
- (ii)：式を展開する。
- (iii)：
  - $\mathbb{E}_{q(x)}[f(x)]=\sum q(x)f(x)$ は、分布 $q(x)$ を用いた場合の $f(x)$ の期待値である。
  - 定義より、$\sum_{\boldsymbol{z}}q(\boldsymbol{z})=1$ であるため、その項が他の $\boldsymbol{z}$ に依存していなければ消える。
  - $q(\boldsymbol{\Psi})$ と関係のない $\log q(\boldsymbol{z})-1$ を定数 C と置く。

$\frac{\partial F(q(\boldsymbol{\Psi}))}{\partial q(\boldsymbol{\Psi})}=0$ となる $q(\boldsymbol{\Psi})$ を解く。

$$\begin{aligned}
\frac{\partial F(q(\boldsymbol{\Psi}))}{\partial q(\boldsymbol{\Psi})}=&\mathbb{E}_{q(\boldsymbol{z})}\left[\log p(\boldsymbol{W},\boldsymbol{z},\boldsymbol{\Psi})\right]-\log q(\boldsymbol{\Psi})-\lambda-C=0 && \text{(i)}\\
\log q(\boldsymbol{\Psi})&=\mathbb{E}_{q(\boldsymbol{z})}\left[\log p(\boldsymbol{W},\boldsymbol{z},\boldsymbol{\Psi})\right]-\log q(\boldsymbol{\Psi})-\lambda-C && \text{(ii)}\\
q(\boldsymbol{\Psi})&=\exp\Big(\mathbb{E}_{q(\boldsymbol{z})}\left[\log p(\boldsymbol{W},\boldsymbol{z},\boldsymbol{\Psi})\right]-\log q(\boldsymbol{\Psi})-\lambda-C\Big) && \text{(iii)}\\
&\propto\exp\Big(\mathbb{E}_{q(\boldsymbol{z})}\left[\log p(\boldsymbol{W},\boldsymbol{z},\boldsymbol{\Psi})\right]\Big) && (3.15)
\end{aligned}$$

【途中式の途中式】

- (i)：式を整理する。
- (ii)：両辺の指数をとる。
- (iii)：$q(\boldsymbol{\Psi})$ と関係のない項を省く。

・ハイパーパラメータの変分事後分布

混合ユニグラムモデルの場合、生成過程より同時確率は以下のように分解できる。

$$p(\boldsymbol{W}, \boldsymbol{z}, \boldsymbol{\Psi}) = p(\boldsymbol{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\alpha)p(\boldsymbol{W}|\boldsymbol{z}, \boldsymbol{\Psi})p(\boldsymbol{\Psi}|\beta) \tag{3.16}$$

また $\boldsymbol{\Psi} = \{\boldsymbol{\theta}, \boldsymbol{\Phi}\}$ であることから、式 (3.15) は

$$\begin{aligned} q(\boldsymbol{\theta}, \boldsymbol{\Phi}) &\propto \exp\Big(\mathbb{E}_{q(\boldsymbol{z})}\left[\log p(\boldsymbol{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\alpha)p(\boldsymbol{W}|\boldsymbol{z}, \boldsymbol{\Phi})p(\boldsymbol{\Phi}|\beta)\right]\Big) \\ &= \exp\Big(\mathbb{E}_{q(\boldsymbol{z})}\left[\log p(\boldsymbol{z}|\boldsymbol{\theta})\right] + \log p(\boldsymbol{\theta}|\alpha)\Big)\exp\Big(\mathbb{E}_{q(\boldsymbol{z})}\left[\log p(\boldsymbol{W}|\boldsymbol{z}, \boldsymbol{\Phi})\right] + \log p(\boldsymbol{\Phi}|\beta)\Big) \end{aligned} \tag{3.17}$$

と分解できる。さらに、$q(\boldsymbol{\theta}, \boldsymbol{\Phi}) = q(\boldsymbol{\theta})q(\boldsymbol{\Phi})$ に分解できることから、$\boldsymbol{\theta}$ に関する因子を取り出すと

$$\begin{aligned} q(\boldsymbol{\theta}) &\propto \exp\Big(\mathbb{E}_{q(\boldsymbol{z})}\left[\log p(\boldsymbol{z}|\boldsymbol{\theta})\right] + \log p(\boldsymbol{\theta}|\alpha)\Big) && \text{(i)} \\ &= \exp\left\{\mathbb{E}_{q(\boldsymbol{z})}\left[\sum_{d=1}^{D}\log\theta_{z_d}\right] + \log\left(\frac{\Gamma(\sum_{k=1}^{K}\alpha)}{\prod_{k=1}^{K}\Gamma(\alpha)}\prod_{k=1}^{K}\theta_k^{\alpha-1}\right)\right\} && \text{(ii)} \\ &= \exp\left(\mathbb{E}_{q(\boldsymbol{z})}\left[\sum_{d=1}^{D}\log\theta_{z_d}\right] + \log\frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} + \sum_{k=1}^{K}\log\theta_k^{\alpha-1}\right) && \text{(iii)} \\ &= \exp\left(\sum_{d=1}^{D}\sum_{k=1}^{K}q_{dk}\log\theta_k + \log\frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} +]\sum_{k=1}^{K}\log\theta_k^{\alpha-1}\right) && \text{(iv)} \\ &\propto \exp\left(\sum_{d=1}^{D}\sum_{k=1}^{K}q_{dk}\log\theta_k + \sum_{k=1}^{K}\log\theta_k^{\alpha-1}\right) && \text{(v)} \\ &= \exp\left(\sum_{k=1}^{K}\log\theta_k^{\sum_{d=1}^{D}q_{dk}} + \sum_{k=1}^{K}\log\theta_k^{\alpha-1}\right) && \text{(vi)} \\ &= \exp\left(\sum_{k=1}^{K}\log\theta_k^{\alpha+\sum_{d=1}^{D}q_{dk}-1}\right) && \text{(vii)} \\ &= \prod_{k=1}^{K}\theta_k^{\alpha+\sum_{d=1}^{D}q_{dk}-1} && \text{(viii)} \\ &\propto \mathrm{Dirichlet}(\boldsymbol{\theta}|\alpha_1, \cdots, \alpha_K) && (3.19) \end{aligned}$$

【途中式の途中式】

- (i)：
  - $p(\boldsymbol{z}|\boldsymbol{\theta}) = \prod_{d=1}^{D}\theta_{z_d}$ であるため、置き換える。
  - トピック分布 $\boldsymbol{\theta}$ の事前分布はディリクレ分布を仮定しているので、$p(\boldsymbol{\theta}|\alpha) = \mathrm{Dirichlet}(\boldsymbol{\theta}|\alpha)$ であることから、1.2.4 節より置き換える。
- (ii)：α は 1 値であるためそれぞれ $\sum_{k=1}^{K}\alpha = \alpha K$、$\prod_{k=1}^{K}\Gamma(\alpha) = \Gamma(\alpha)^K$ であり、また $\log\prod_{k=1}^{K}\theta_k^{\alpha-1} = \sum_{k=1}^{K}\log\theta_k^{\alpha-1}$ である。
- (iii)：$q_{dk} \equiv q(z_d = k)$ を文書 d のトピック z_d が k になる変分事後分布とすると

$$
\begin{aligned}
\mathbb{E}_{q(\boldsymbol{z})}\left[\sum_{d=1}^{D}\log\theta_{z_d}\right] &= \mathbb{E}_{q(\boldsymbol{z})}\left[\log\theta_{z_1}+\log\theta_{z_2}+\cdots+\log\theta_{z_D}\right] \\
&= \sum_{k=1}^{K}\Big(q(z_1=k)\log\theta_{z_1}+q(z_2=k)\log\theta_{z_2}+\cdots+q(z_D=k)\log\theta_{z_D}\Big) \\
&= \sum_{k=1}^{K}(q_{1k}\log\theta_{z_1}+q_{2k}\log\theta_{z_2}+\cdots+q_{Dk}\log\theta_{z_D}) \\
&= \sum_{d=1}^{D}\sum_{k=1}^{K}q_{dk}\log\theta_k
\end{aligned}
$$

になることから置き換える。

- (iv)：$\log\frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K}$ は、$\boldsymbol{\theta}$ に依存しないことから省く。
- (v)：$a\log x=\log x^a$ であることから、2 つの θ_k の項の形を揃える。
- (vi)：$\log x^a+\log x^b=\log x^{a+b}$ であることから、θ_k の項をまとめる。
- (vii)：指数をとる。$\exp(\sum\log x)=\prod x$ である。
- (viii)：$\alpha_k=\alpha+\sum_{d=1}^{D}q_{dk}$ としたとき、パラメータ $(\alpha_1,\alpha_2,\cdots,\alpha_K)$ を持つ正規化項のないディリクレ分布の形であることから。

変分事後分布 $q(\boldsymbol{\theta})$ がディリクレ分布であることが分かった。また

$$
\alpha_k=\alpha+\sum_{d=1}^{D}q_{dk}
$$

がハイパーパラメータ α の更新式となる。

変分事後分布 $q(\boldsymbol{\Phi})$ についても同様に、式 (3.17) から $\boldsymbol{\Phi}$ に関する因子を取り出すと

$$
\begin{aligned}
q(\mathbf{\Phi}) &\propto \exp\Big(\mathbb{E}_{q(\boldsymbol{z})}\left[\log p(\boldsymbol{W}|\boldsymbol{z},\mathbf{\Phi})\right] + \log p(\mathbf{\Phi}|\beta)\Big) && \text{(i)}\\
&= \exp\left(\mathbb{E}_{q(\boldsymbol{z})}\left[\sum_{d=1}^{D}\sum_{v=1}^{V} N_{dv}\log\phi_{kv}\right] + \log\frac{\Gamma\Big(\beta(K+V)\Big)}{\Gamma(\beta)^{K+V}} + \log\phi_{kv}^{\beta-1}\right) && \text{(ii)}\\
&= \exp\left(\sum_{k=1}^{K}\sum_{v=1}^{V}\sum_{d=1}^{D} q_{dk}N_{dv}\log\phi_{kv} + \log\frac{\Gamma\Big(\beta(K+V)\Big)}{\Gamma(\beta)^{K+V}} + \log\phi_{kv}^{\beta-1}\right) && \text{(iii)}\\
&\propto \exp\left(\sum_{k=1}^{K}\sum_{v=1}^{V}\sum_{d=1}^{D} q_{dk}N_{dv}\log\phi_{kv} + \log\phi_{kv}^{\beta-1}\right) && \text{(iv)}\\
&= \exp\left(\sum_{k=1}^{K}\sum_{v=1}^{V}\log\phi_{kv}^{\sum_{d=1}^{D} q_{dk}N_{dv}} + \log\phi_{kv}^{\beta-1}\right) && \text{(v)}\\
&= \exp\left(\sum_{k=1}^{K}\sum_{v=1}^{V}\log\phi_{kv}^{\beta+\sum_{d=1}^{D} q_{dk}N_{dv}-1}\right) && \text{(vi)}\\
&= \prod_{k=1}^{K}\prod_{v=1}^{V}\phi_{kv}^{\beta+\sum_{d=1}^{D} q_{dk}N_{dv}-1} && \text{(vii)}\\
&= \prod_{k=1}^{K}\mathrm{Dirichlet}(\boldsymbol{\phi}_k|\beta_{k1},\cdots,\beta_{kV}) && \text{(3.20)}
\end{aligned}
$$

【途中式の途中式】

- (i)：
  - $p(\mathbf{\Phi}|\beta)=\mathrm{Dirichlet}(\mathbf{\Phi}|\beta)$ を仮定しているため、1.2.4 節より置き換える。

$$
\begin{aligned}
p(\boldsymbol{W}|\boldsymbol{z},\mathbf{\Phi}) &= \prod_{d=1}^{D} p(\boldsymbol{w}_d|z_d,\mathbf{\Phi})\\
&= \prod_{d=1}^{D} p(w_{d1},w_{d2},\cdots,w_{dn_d}|z_d=k,\boldsymbol{\phi}_k)\\
&= \prod_{d=1}^{D}\prod_{d=1}^{N_d}\phi_{kw_{dn}}\\
&= \prod_{d=1}^{D}\prod_{v=1}^{V}\phi_{kv}^{N_{dv}}\\
\log p(\boldsymbol{W}|\boldsymbol{z},\mathbf{\Phi}) &= \sum_{d=1}^{D}\sum_{v=1}^{V} N_{dv}\log\phi_{kv} && (1)
\end{aligned}
$$

- (ii)：

$$
\begin{aligned}
&\mathbb{E}_{q(\boldsymbol{z})}\left[\sum_{d=1}^{D}\sum_{v=1}^{V} N_{dv}\log\phi_{kv}\right] \\
&= \mathbb{E}_{q(\boldsymbol{z})}\Big[N_{11}\log\phi_{k1} + N_{12}\log\phi_{k2} + \cdots + N_{1V}\log\phi_{kV} \\
&\qquad + N_{21}\log\phi_{k1} + N_{22}\log\phi_{k2} + \cdots + N_{2V}\log\phi_{kV} \\
&\qquad\qquad\vdots \\
&\qquad + N_{D1}\log\phi_{k1} + N_{D2}\log\phi_{k2} + \cdots + N_{DV}\log\phi_{kV}\Big] \\
&= \sum_{k=1}^{K}\Big(q(z_1=k)N_{11}\log\phi_{k1} + q(z_1=k)N_{12}\log\phi_{k2} + \cdots + q(z_1=k)N_{1V}\log\phi_{kV} \\
&\qquad + q(z_2=k)N_{21}\log\phi_{k1} + q(z_2=k)N_{22}\log\phi_{k2} + \cdots + q(z_2=k)N_{2V}\log\phi_{kV} \\
&\qquad\qquad\vdots \\
&\qquad + q(z_D=k)N_{D1}\log\phi_{k1} + q(z_D=k)N_{D2}\log\phi_{k2} + \cdots + q(z_D=k)N_{DV}\log\phi_{kV}\Big) \\
&= \sum_{k=1}^{K}(q_{1k}N_{11}\log\phi_{k1} + q_{1k}N_{12}\log\phi_{k2} + \cdots + q_{1k}N_{1V}\log\phi_{kV} \\
&\qquad + q_{2k}N_{21}\log\phi_{k1} + q_{2k}N_{22}\log\phi_{k2} + \cdots + q_{2k}N_{2V}\log\phi_{kV} \\
&\qquad\qquad\vdots \\
&\qquad + q_{Dk}N_{D1}\log\phi_{k1} + q_{Dk}N_{D2}\log\phi_{k2} + \cdots + q_{Dk}N_{DV}\log\phi_{kV}) \\
&= \sum_{k=1}^{K}\sum_{v=1}^{V}\sum_{d=1}^{D} q_{dk}N_{dv}\log\phi_{kv}
\end{aligned}
$$

- (iii)：$\log\frac{\Gamma(\beta(K+V))}{\Gamma(\alpha)^{K+V}}$ は、$\boldsymbol{\Phi}$ に依存しないことから省く。
- (iv)：$a\log x = \log x^a$ であることから、2 つの ϕ_{kv} の項を省く。
- (v)：$\log x^a + \log x^b = \log x^{a+b}$ であることから、ϕ_{kv} の項をまとめる。
- (vi)：指数をとる $\exp(\sum\log x) = \prod x$ である。
- (vii)：$\beta_{kv} = \beta + \sum_{d=1}^{D} q_{dk}N_{dv}$ としたとき、パラメータ $(\beta_{k1}, \beta_{k2}, \cdots, \beta_{kV})$ を持つ正規化項のないディリクレ分布の形であることから。

変分事後分布 $q(\boldsymbol{\Phi})$ がディリクレ分布であることが分かった。また

$$
\beta_{kv} = \beta + \sum_{d=1}^{D} q_{dk}N_{dv}
$$

がハイパーパラメータ β の更新式となる。

・トピックの変分事後分布

続いて、トピックの変分事後分布 $q(\boldsymbol{z})$ について考える。式 (3.13),(3.16) より

$$
\begin{aligned}
q(\boldsymbol{z}) &\propto \exp\Bigl(\mathbb{E}_{q(\boldsymbol{\theta},\boldsymbol{\Phi})}\left[\log p(\boldsymbol{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\alpha)p(\boldsymbol{W}|\boldsymbol{z},\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\beta)\right]\Bigr) && \text{(i)}\\
&\propto \exp\Bigl(\mathbb{E}_{q(\boldsymbol{\theta})}\left[\log p(\boldsymbol{z}|\boldsymbol{\theta})\right] + \mathbb{E}_{q(\boldsymbol{\Phi})}\left[p(\boldsymbol{W}|\boldsymbol{z},\boldsymbol{\Phi})\right]\Bigr) && \text{(ii)}\\
&= \exp\left(\mathbb{E}_{q(\boldsymbol{\theta})}\left[\sum_{d=1}^{D}\log\theta_{z_d}\right] + \mathbb{E}_{q(\boldsymbol{\Phi})}\left[\sum_{d=1}^{D}\sum_{v=1}^{V}N_{dv}\log\phi_{z_d v}\right]\right) && \text{(iii)}\\
&= \exp\left\{\sum_{d=1}^{D}\left(\Psi(\alpha_{z_d}) - \Psi\Bigl(\sum_{k'=1}^{K}\alpha_{k'}\Bigr)\right) + \sum_{d=1}^{D}\left(\sum_{v=1}^{V}N_{dv}\Psi(\beta_{z_d v}) - N_d\Psi\Bigl(\sum_{v=1}^{V}\beta_{z_d v}\Bigr)\right)\right\} && \text{(iv)}\\
&= \prod_{d=1}^{D}\exp\left(\Psi(\alpha_{z_d}) - \Psi\Bigl(\sum_{k'=1}^{K}\alpha_{k'}\Bigr) + \sum_{v=1}^{V}N_{dv}\Psi(\beta_{z_d v}) + N_d\Psi\Bigl(\sum_{v=1}^{V}\beta_{z_d v}\Bigr)\right) && (3.21)
\end{aligned}
$$

【途中式の途中式】

- (i)：$\boldsymbol{z}$ に関する項のみを取り出す。
- (ii)：
  - $p(\boldsymbol{z}|\boldsymbol{\theta}) = \prod_{d=1}^{D}\theta_{z_d}$ であるため置き換える。
  - 式 (1) より、置き換える。
- (iii)：ディリクレ分布に従う変数の対数の期待値 (1.17) を用いる。

前の因子は

$$
\begin{aligned}
\mathbb{E}_{q(\boldsymbol{\theta})}\left[\sum_{d=1}^{D}\log\theta_{z_d}\right] &= \mathbb{E}_{q(\boldsymbol{\theta})}\left[\log\theta_{z_1} + \log\theta_{z_2} + \cdots + \log\theta_{z_D}\right]\\
&= \sum_{k'=1}^{K}q(\theta_{z_1=k'}|\alpha_{z_1})\log\theta_{z_1=k'} + \sum_{k'=1}^{K}q(\theta_{z_2=k'}|\alpha_{z_2})\log\theta_{z_2=k'}\\
&\qquad + \cdots + \sum_{k'=1}^{K}q(\theta_{z_D=k'}|\alpha_{z_D})\log\theta_{z_D=k'}\\
&= \Psi(\alpha_{z_1}) - \Psi\Bigl(\sum_{k'=1}^{K}\alpha_{k'}\Bigr) + \Psi(\alpha_{z_2}) - \Psi\Bigl(\sum_{k'=1}^{K}\alpha_{k'}\Bigr) + \cdots + \Psi(\alpha_{z_D}) - \Psi\Bigl(\sum_{k'=1}^{K}\alpha_{k'}\Bigr)\\
&= \sum_{d=1}^{D}\left(\Psi(\alpha_{z_d}) - \Psi\Bigl(\sum_{k'=1}^{K}\alpha_{k'}\Bigr)\right)
\end{aligned}
$$

となる。同様に後の因子は

$$
\begin{aligned}
\mathbb{E}_{q(\boldsymbol{\Phi})}\left[\sum_{d=1}^{D}\sum_{v=1}^{V} N_{dv}\log\phi_{z_d v}\right] &= \mathbb{E}_{q(\boldsymbol{\Phi})}\left[\sum_{v=1}^{V} N_{1v}\log\phi_{z_1 v} + \sum_{v=1}^{V} N_{2v}\log\phi_{z_2 v} + \cdots + \sum_{v=1}^{V} N_{Dv}\log\phi_{z_D v}\right] \\
&= \sum_{v=1}^{V} N_{1v} q(\boldsymbol{\phi}_{z_1 v}|\beta_{z_1 v})\log\phi_{z_1 v} + \sum_{v=1}^{V} N_{2v} q(\boldsymbol{\phi}_{z_2 v}|\beta_{z_2 v})\log\phi_{z_2 v} \\
&\quad + \cdots + \sum_{v=1}^{V} N_{Dv} q(\boldsymbol{\phi}_{z_D v}|\beta_{z_D v})\log\phi_{z_D v} \\
&= \sum_{v=1}^{V} N_{1v}\Psi(\beta_{z_1 v}) - N_1\Psi\Big(\sum_{v=1}^{V}\beta_{z_1 v}\Big) + \sum_{v=1}^{V} N_{2v}\Psi(\beta_{z_2 v}) - N_2\Psi\Big(\sum_{v=1}^{V}\beta_{z_2 v}\Big) \\
&\quad + \cdots + \sum_{v=1}^{V} N_{Dv}\Psi(\beta_{z_D v}) - N_D\Psi\Big(\sum_{v=1}^{V}\beta_{z_D v}\Big) \\
&= \sum_{d=1}^{D}\left\{\sum_{v=1}^{V} N_{dv}\Psi(\beta_{z_d v}) - N_d\Psi\Big(\sum_{v=1}^{V}\beta_{z_d v}\Big)\right\}
\end{aligned}
$$

となる。

- (iv)：$\sum_{d=1}^{D}$ を exp() の外に出す。

この式は、文書 d のトピックが k である確率が

$$
q_{dk} \propto \exp\left(\Psi(\alpha_{z_d}) - \Psi\Big(\sum_{k'=1}^{K}\alpha_{k'}\Big) + \sum_{v=1}^{V} N_{dv}\Psi(\beta_{z_d v}) + N_d\Psi\Big(\sum_{v=1}^{V}\beta_{z_d v}\Big)\right) \tag{3.22}
$$

であることを示している。

EM アルゴリズムを用いた最尤推定では、各文書のトピックは分布 (負担率)$\boldsymbol{q}_d = (q_{d1}, \cdots, q_{dK})$ として求めていたが、パラメータ $\boldsymbol{\theta}, \boldsymbol{\Phi}$ は点推定していた。変分ベイズ推定ではハイパーパラメータを推定することで、トピック $q(\boldsymbol{z})$、パラメータ $q(\boldsymbol{\theta}), q(\boldsymbol{\Phi})$ の全てを分布推定している。

3.4.3 変分下限

変分事後分布を更新する際に変分下限が減少することはない。それを利用してアルゴリズムが適切に動作しているかを確認できる。

混合ユニグラムモデルの変分下限は式 (3.10),(3.11),(3.16),(3.18) より

$$
\begin{aligned}
F =& \int\int\sum_{\boldsymbol{z}} q(\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\Phi}) \log p(\boldsymbol{W},\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\Phi}|\alpha,\beta) d\boldsymbol{\theta} d\boldsymbol{\Phi} \\
&- \int\int\sum_{\boldsymbol{z}} q(\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\Phi}) \log q(\boldsymbol{z},\boldsymbol{\theta},\boldsymbol{\Phi}) d\boldsymbol{\theta} d\boldsymbol{\Phi} \\
=& \overbrace{\int\sum_{\boldsymbol{z}} q(\boldsymbol{z}) q(\boldsymbol{\theta}) \log p(\boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{\theta}}^{\text{1st term}} + \overbrace{\int q(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}|\alpha) d\boldsymbol{\theta}}^{\text{2nd term}} + \overbrace{\int\sum_{\boldsymbol{z}} q(\boldsymbol{z}) q(\boldsymbol{\Phi}) \log p(\boldsymbol{W}|\boldsymbol{z},\boldsymbol{\Phi}) d\boldsymbol{\Phi}}^{\text{3rd term}} \\
&+ \overbrace{\int q(\boldsymbol{\Phi}) \log p(\boldsymbol{\Phi}|\beta) d\boldsymbol{\Phi}}^{\text{4th term}} - \overbrace{\sum_{\boldsymbol{z}} q(\boldsymbol{z}) \log q(\boldsymbol{z})}^{\text{5th term}} - \overbrace{\int q(\boldsymbol{\theta}) \log q(\boldsymbol{\theta}) d\boldsymbol{\theta}}^{\text{6th term}} - \overbrace{\int q(\boldsymbol{\Phi}) \log q(\boldsymbol{\Phi}) d\boldsymbol{\Phi}}^{\text{7th term}}
\end{aligned}
$$

となる。それぞれの項は以下のように計算できる。

・第 1 項

$$
\begin{aligned}
& \int\sum_{\boldsymbol{z}} q(\boldsymbol{z}) q(\boldsymbol{\theta}) \log p(\boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{\theta} && \text{(i)} \\
&= \sum_{d=1}^{D}\sum_{k=1}^{K} q_{dk} \int\sum_{\boldsymbol{z}} q(\boldsymbol{\theta}) \log p(\boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{\theta} && \text{(ii)} \\
&\propto \sum_{d=1}^{D}\sum_{k=1}^{K} q_{dk} \int\sum_{\boldsymbol{z}} q(\boldsymbol{\theta}|\alpha_1,\cdots,\alpha_K) \log p(\boldsymbol{z}|\boldsymbol{\theta}) d\boldsymbol{\theta} && \text{(iii)} \\
&= \sum_{d=1}^{D}\sum_{k=1}^{K} q_{dk} \int q(\boldsymbol{\theta}|\alpha_1,\cdots,\alpha_K) \log \theta_k d\boldsymbol{\theta} && \text{(iv)} \\
&= \sum_{d=1}^{D}\sum_{k=1}^{K} q_{dk} \left(\Psi(\alpha_k) - \Psi\Big(\sum_{k'=1}^{K}\alpha_{k'}\Big)\right)
\end{aligned}
$$

【途中式の途中式】

- (i)：次の式変形より、置き換える。

$$
\begin{aligned}
q(\boldsymbol{z}) &= \sum_{d=1}^{D} q(z_d) \\
\sum_{\boldsymbol{z}} q(\boldsymbol{z}) &= \sum_{d=1}^{D}\sum_{k=1}^{K} q(z_d = k) \\
&= \sum_{d=1}^{D}\sum_{k=1}^{K} q_{dk} \qquad (2)
\end{aligned}
$$

- (ii)：式 (3.19) より、置き換える。
- (iii)：次の式変形より、置き換える。

$$
\begin{aligned}
p(\boldsymbol{z}|\boldsymbol{\theta}) &= \prod_{d=1}^{D} p(z_d|\boldsymbol{\theta}) \\
\log p(\boldsymbol{z}|\boldsymbol{\theta}) &= \sum_{d=1}^{D} p(z_d|\boldsymbol{\theta}) \\
\sum_{\boldsymbol{z}} \log p(\boldsymbol{z}|\boldsymbol{\theta}) &= \sum_{d=1}^{D} \sum_{k=1}^{K} \log p(z_d = k|\theta_k) \\
&= \sum_{k=1}^{K} \log \theta_k
\end{aligned}
$$

- (iv)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・第 2 項

$$
\begin{aligned}
&\int q(\boldsymbol{\theta}) \log p(\boldsymbol{\theta}|\alpha) d\boldsymbol{\theta} && \text{(i)} \\
&\propto \int q(\boldsymbol{\theta}|\alpha_1, \cdots, \alpha_K) \log p(\boldsymbol{\theta}|\alpha) d\boldsymbol{\theta} && \text{(ii)} \\
&= \int q(\boldsymbol{\theta}|\alpha_1, \cdots, \alpha_K) \log \left(\frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \prod_{k=1}^{K} \theta_k^{\alpha-1} \right) d\boldsymbol{\theta} && \text{(iii)} \\
&= \log \Gamma(\alpha K) - K \log \Gamma(\alpha) + (\alpha - 1) \sum_{k=1}^{K} \int q(\boldsymbol{\theta}|\alpha_1, \cdots, \alpha_K) \log \theta_k d\boldsymbol{\theta} && \text{(iv)} \\
&= \log \Gamma(\alpha K) - K \log \Gamma(\alpha) + (\alpha - 1) \sum_{k=1}^{K} \left(\Psi(\alpha_k) - \Psi\Big(\sum_{k'=1}^{K} \alpha_{k'}\Big) \right)
\end{aligned}
$$

【途中式の途中式】

- (i)：式 (3.19) より、置き換える。
- (ii,iii)：次の式変形より、置き換える。

$$
\begin{aligned}
p(\boldsymbol{\theta}|\alpha) &= \mathrm{Dirichlet}(\boldsymbol{\theta}|\alpha) \\
&= \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \prod_{k=1}^{K} \theta_k^{\alpha-1} \\
\log p(\boldsymbol{\theta}|\alpha) &= \log \Gamma(\alpha K) - K \log \Gamma(\alpha) + (\alpha - 1) \sum_{k=1}^{K} \log \theta_k
\end{aligned}
$$

- (iv)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・第 3 項

$$
\begin{aligned}
&\int \sum_{\boldsymbol{z}} q(\boldsymbol{z}) q(\boldsymbol{\Phi}) \log p(\boldsymbol{W}|\boldsymbol{z}, \boldsymbol{\Phi}) d\boldsymbol{\Phi} && \text{(i)} \\
&= \sum_{d=1}^{D} \sum_{k=1}^{K} q_{dk} \int q(\boldsymbol{\Phi}) \log p(\boldsymbol{W}|\boldsymbol{z}, \boldsymbol{\Phi}) d\boldsymbol{\Phi} && \text{(ii)} \\
&\propto \sum_{d=1}^{D} \sum_{k=1}^{K} q_{dk} \int q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV}) \log p(\boldsymbol{W}|\boldsymbol{z}, \boldsymbol{\Phi}) d\boldsymbol{\Phi} && \text{(iii)} \\
&= \sum_{d=1}^{D} \sum_{k=1}^{K} q_{dk} \sum_{v=1}^{V} N_{dv} \int q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV}) \log \phi_{kv} d\boldsymbol{\phi}_k && \text{(iv)} \\
&= \sum_{d=1}^{D} \sum_{k=1}^{K} q_{dk} \sum_{v=1}^{V} N_{dv} \left(\Psi(\beta_{kv}) - \Psi\Big(\sum_{v'=1}^{V} \beta_{kv'}\Big) \right)
\end{aligned}
$$

【途中式の途中式】

- (i)：式 (2) より、置き換える。
- (ii)：式 (3.20) より、置き換える。
- (iii)：式 (1) より、置き換える。
- (iv)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・第 4 項

$$
\begin{aligned}
&\int q(\boldsymbol{\Phi}) \log p(\boldsymbol{\Phi}|\beta) d\boldsymbol{\Phi} && \text{(i)} \\
&\propto \sum_{k=1}^{K} \int q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV}) \log p(\boldsymbol{\Phi}|\beta) d\boldsymbol{\Phi} && \text{(ii)} \\
&= \sum_{k=1}^{K} \int q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV}) \log \left(\frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \prod_{v=1}^{V} \phi_v^{\beta-1} \right) d\boldsymbol{\phi}_k && \text{(iii)} \\
&= K \log \Gamma(\beta V) - KV \log \Gamma(\beta) + (\beta - 1) \sum_{k=1}^{K} \sum_{v=1}^{V} \int q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV}) \log \phi_{kv} d\boldsymbol{\phi}_k && \text{(iv)} \\
&= K \log \Gamma(\beta V) - KV \log \Gamma(\beta) + (\beta - 1) \sum_{k=1}^{K} \sum_{v=1}^{V} \left(\Psi(\beta_{kv}) - \Psi\Big(\sum_{v'=1}^{V} \beta_{kv'}\Big) \right)
\end{aligned}
$$

【途中式の途中式】

- (i)：式 (3.20) より、置き換える。
- (ii,iii)：次の式変形より、置き換える。

$$
\begin{aligned}
p(\boldsymbol{\Phi}|\beta) &= \mathrm{Dirichlet}(\boldsymbol{\Phi}|\beta) \\
&= \frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \prod_{v=1}^{V} \phi_v^{\beta-1}
\end{aligned}
$$

$$
\log p(\boldsymbol{\Phi}|\beta) = \log \Gamma(\beta V) - V \log \Gamma(\beta) + (\beta - 1) \sum_{v=1}^{V} \log \phi_v
$$

- (iv)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・第 5 項

$$
\begin{aligned}
&\sum_{\boldsymbol{z}} q(\boldsymbol{z}) \log q(\boldsymbol{z}) && \text{(i)} \\
&= \sum_{d=1}^{D} \sum_{k=1}^{K} q_{dk} \log q_{dk}
\end{aligned}
$$

【途中式の途中式】

- (i)：式 (2) より、置き換える。

・第 6 項

$$
\begin{aligned}
&\int q(\boldsymbol{\theta}) \log q(\boldsymbol{\theta}) d\boldsymbol{\theta} && \text{(i)} \\
&\propto \int q(\boldsymbol{\theta}|\alpha_1, \cdots, \alpha_K) \log q(\boldsymbol{\theta}) d\boldsymbol{\theta} && \text{(ii)} \\
&\propto \int q(\boldsymbol{\theta}|\alpha_1, \cdots, \alpha_K) \log \left(\frac{\Gamma\Big(\sum_{k=1}^{K} \alpha_k\Big)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \right) d\boldsymbol{\theta} && \text{(iii)} \\
&= \log \Gamma\Big(\sum_{k=1}^{K} \alpha_k\Big) - \sum_{k=1}^{K} \log \Gamma(\alpha_k) + \sum_{k=1}^{K} (\alpha_k - 1) \int q(\boldsymbol{\theta}|\alpha_1, \cdots, \alpha_K) \log \theta_k d\boldsymbol{\theta} && \text{(iv)} \\
&= \log \Gamma \left(\sum_{k=1}^{K} \alpha_k \right) - \sum_{k=1}^{K} \log \Gamma\left(\alpha_k\right) + \sum_{k=1}^{K} (\alpha_k - 1) \left(\Psi(\alpha_k) - \Psi\Big(\sum_{k'=1}^{K} \alpha_{k'}\Big) \right)
\end{aligned}
$$

【途中式の途中式】

- (i)：式 (3.19) より、前の項を置き換える。
- (ii,iii)：式 (3.19) を用いて、次の式変形より後ろの項を置き換える。ただし、式 (3.19) では正規化項がなかったが、確率分布とするために正規化項を付け加えている。

$$
\begin{aligned}
q(\boldsymbol{\theta}) &\propto \mathrm{Dirichlet}(\boldsymbol{\theta}|\alpha_1, \cdots, \alpha_K) \\
&= \frac{\Gamma\Big(\sum_{k=1}^{K} \alpha_k\Big)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} \theta_k^{\alpha_k - 1} \\
\log q(\boldsymbol{\theta}) &= \log \Gamma\Big(\sum_{k=1}^{K} \alpha_k\Big) - \sum_{k=1}^{K} \log \Gamma(\alpha_k) + (\alpha_k - 1) \sum_{k=1}^{K} \log \theta_k
\end{aligned}
$$

- (iv)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・第 7 項

$$
\begin{aligned}
&\int q(\boldsymbol{\Phi}) \log q(\boldsymbol{\Phi}) d\boldsymbol{\Phi} && \text{(i)} \\
&\propto \sum_{k=1}^{K} \int q(\boldsymbol{\phi}_k | \beta_{k1}, \cdots, \beta_{kV}) \log q(\boldsymbol{\Phi}) d\boldsymbol{\Phi} && \text{(ii)} \\
&\propto \sum_{k=1}^{K} \int q(\boldsymbol{\phi}_k | \beta_{k1}, \cdots, \beta_{kV}) \log \left(\frac{\Gamma\left(\sum_{v=1}^{V} \beta_{kv}\right)}{\prod_{v=1}^{V} \Gamma(\beta_{kv})} \prod_{v=1}^{V} \phi_{kv}^{\beta_{kv}-1} \right) d\boldsymbol{\Phi} && \text{(iii)} \\
&= \sum_{k=1}^{K} \left(\log \Gamma\left(\sum_{v=1}^{V} \beta_{kv} \right) - \sum_{v=1}^{V} \log \Gamma\left(\beta_{kv} \right) + \sum_{v=1}^{V} (\beta_{kv} - 1) \int q(\boldsymbol{\phi}_k | \beta_{k1}, \cdots, \beta_{kV}) \log \phi_{kv} d\boldsymbol{\phi}_k \right) && \text{(iv)} \\
&= \sum_{k=1}^{K} \left\{ \log \Gamma\left(\sum_{v=1}^{V} \beta_{kv} \right) - \sum_{v=1}^{V} \log \Gamma\left(\beta_{kv} \right) + \sum_{v=1}^{V} (\beta_{kv} - 1) \left(\Psi(\beta_{kv}) - \Psi\left(\sum_{v'=}^{V} \beta_{kv'} \right) \right) \right\}
\end{aligned}
$$

【途中式の途中式】

- (i)：式 (3.20) より、前の項を置き換える。
- (ii,iii)：式 (3.20) を用いて、次の式変形より後ろの項を置き換える。ただし、式 (3.20) では正規化項がなかったたが、確率分布とするために正規化項を付け加えている。

$$
\begin{aligned}
q(\boldsymbol{\Phi}) &\propto \prod_{k=1}^{K} \mathrm{Dirichlet}(\boldsymbol{\phi}_k | \beta_{k1}, \cdots, \beta_{kV}) \\
&= \prod_{k=1}^{K} \frac{\Gamma\left(\sum_{v=1}^{V} \beta_{kv}\right)}{\prod_{v=1}^{V} \Gamma(\beta_{kv})} \prod_{v=1}^{V} \phi_{kv}^{\beta_{kv}-1} \\
\log q(\boldsymbol{\Phi}) &= \sum_{k=1}^{K} \left(\log \Gamma\left(\sum_{v=1}^{V} \beta_{kv} \right) - \sum_{v=1}^{V} \log \Gamma(\beta_{kv}) + (\beta_{kv} - 1) \sum_{v=1}^{V} \log \phi_{kv} \right)
\end{aligned}
$$

- (iv)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・**R** で組んでみる

・コード全体

・テキスト処理

```
# 利用パッケージ
library(RMeCab)
library(tidyverse)

### テキスト処理

# 抽出しない単語を指定
stop_words <- "[a-z]" # 小文字のアルファベットを含む語

# 形態素解析
mecab_df <- docDF(" フォルダ名", type = 1) # テキストファイルの保存先を指定する
```

```r
# 文書 d の語彙 v の出現回数 (N_dv) の集合
N_dv <- mecab_df %>%
        filter(grepl("名詞|形容詞|^動詞", POS1)) %>% # 抽出する品詞（大分類）を指定する
        filter(grepl("一般|^自立", POS2)) %>%        # 抽出する品詞（細分類）を指定する
        filter(!grepl(stop_words, TERM)) %>%         # ストップワードを除く
        select(-c(TERM, POS1, POS2)) %>%             # 数値列のみを残す
        filter(apply(., 1, sum) >= 5) %>%            # 抽出する総出現回数を指定する
        t()                                          # 転置

# 確認用の行列名
dimnames(N_dv) <- list(paste0("d=", 1:nrow(N_dv)), # 行名
                       paste0("v=", 1:ncol(N_dv))) # 列名

# 文書 d の単語数 (N_d) のベクトル
N_d <- apply(N_dv, 1, sum) # 行方向に和をとる

# 文書数 (D)
D <- nrow(N_dv)

# 総語彙数 (V)
V <- ncol(N_dv)
```

・パラメータの初期設定

```r
### パラメータの初期設定

# トピック数 (K)
K <- 4   # 任意の値を指定する

# 負担率 (q_dk) の集合
q_dk <- matrix(0, nrow = D, ncol = K,
               dimnames = list(paste0("d=", 1:D),  # 行名
                               paste0("k=", 1:K))) # 列名

# 事前分布のパラメータ (alpha, beta)
alpha <- 2 # 任意の値を指定する
beta  <- 2 # 任意の値を指定する

# 変分事後分布のパラメータ (alpha_k) のベクトル
alpha_k <- seq(1, 3, 0.1) %>%               # 任意の範囲を指定する
           sample(size = K, replace = TRUE) # 値をランダムに割り振る
names(alpha_k) <- paste0("k=", 1:K)         # 確認用の要素名

# 変分事後分布のパラメータ (beta_kv) の集合
beta_kv <- seq(1, 3, 0.1) %>%                       # 任意の範囲を指定する
           sample(size = K * V, replace = TRUE) %>% # 値をランダムに割り振る
           matrix(nrow = K, ncol = V,
                  dimnames = list(paste0("k=", 1:K),  # 行名
                                  paste0("v=", 1:V))) # 列名
```

・変分ベイズ推定

```r
### 変分ベイズ推定
```

```r
# 推移の確認用
trace_alpha <- as.matrix(alpha_k)
trace_beta  <- beta_kv[1, ]

for(i in 1:50) { # 任意の回数を指定する

  # 次ステップのハイパーパラメータ (alpha, beta) を初期化
  new_alpha_k <- rep(alpha, K)
  new_beta_kv <- matrix(beta, nrow = K, ncol = V,
                        dimnames = list(paste0("k=", 1:K),
                                        paste0("v=", 1:V)))

  for(d in 1:D){ ## (各文書)

    for(k in 1:K){ ## (各トピック)

      # 負担率 (q_dk) を計算
      tmp_q_alpha <- digamma(alpha_k[k]) - digamma(sum(alpha_k))
      tmp_q_beta  <- sum(N_dv[d, ] * digamma(beta_kv[k, ])) - N_d[d] * digamma(sum(beta_kv[k, ]))
      q_dk[d, k]  <- exp(tmp_q_alpha + tmp_q_beta)

    }

    # 負担率の正規化：sum_{k=1}^K q_dk = 1
    if(sum(q_dk[d, ]) > 0) { ## (全ての値が 0 の場合は回避)
      q_dk[d, ] <- q_dk[d, ] / sum(q_dk[d, ])
    } else if(sum(q_dk[d, ]) == 0) {
      q_dk[d, ] <- 1 / K
    }

    for(k in 1:K) { ## (各トピック (続き))

      # ハイパーパラメータ (alpha_k) を更新
      new_alpha_k[k] <- new_alpha_k[k] + q_dk[d, k]

      for(v in 1:V){ ## (各語彙)

        # ハイパーパラメータ (beta_kv) を更新
        new_beta_kv[k, v] <- new_beta_kv[k, v] + q_dk[d, k] * N_dv[d, v]

      } ## (/各語彙)
    } ## (/各トピック)
  } ## (/各文書)

  # ハイパーパラメータの (alpha, beta) を更新
  alpha_k <- new_alpha_k
  beta_kv <- new_beta_kv

  # 推移の確認用
  trace_alpha <- cbind(trace_alpha, alpha_k)
  trace_beta  <- cbind(trace_beta, beta_kv[1, ])
}
```

・コードの解説

・利用パッケージ

```r
# 利用パッケージ
library(RMeCab)
library(tidyverse)
```

　形態素解析のための `RMeCab::docDF()` と、グラフ作成時の `tidyr::gather()`、`ggplot2` 関連の関数以外にライブラリの読み込みが必要なのは `dplyr` のみ。(面倒なので `tidyvers` パッケージを読み込んでいる)

　その他テキスト処理の部分は 3.3 節を参照のこと。

・パラメータの初期設定

・トピック数

```r
# トピック数 (K)
K <- 4   # 任意の値を指定する
```

　任意のトピック数を指定する。

・負担率

```r
# 負担率 (q_dk) の集合
q_dk <- matrix(0, nrow = D, ncol = K,
               dimnames = list(paste0("d=", 1:D),  # 行名
                               paste0("k=", 1:K))) # 列名
```

　全ての要素が 0 である D 行 K 列の行列を作成する。

・事前分布のハイパーパラメータ

```r
# 事前分布のパラメータ (alpha, beta)
alpha <- 2 # 任意の値を指定する
beta  <- 2 # 任意の値を指定する
```

　事前分布 $p(\boldsymbol{\theta}|\alpha), p(\boldsymbol{\Phi}|\beta)$ のパラメータ α, β を任意の値で設定する。
　この値は更新パラメータの初期化時の値として用いる。

・変分事後分布のハイパーパラメータ

```r
# 変分事後分布のパラメータ (alpha_k) のベクトル
alpha_k <- seq(1, 3, 0.1) %>%                # 任意の範囲を指定する
           sample(size = K, replace = TRUE) # 値をランダムに割り振る
names(alpha_k) <- paste0("k=", 1:K)         # 確認用の要素名

# 変分事後分布のパラメータ (beta_kv) の集合
```

```r
beta_kv <- seq(1, 3, 0.1) %>%                          # 任意の範囲を指定する
           sample(size = K * V, replace = TRUE) %>%  # 値をランダムに割り振る
           matrix(nrow = K, ncol = V,
                  dimnames = list(paste0("k=", 1:K),  # 行名
                                  paste0("v=", 1:V))) # 列名
```

変分事後分布

$$q(\boldsymbol{\theta}) \propto \mathrm{Dicichlet}(\boldsymbol{\theta}|\beta_1, \cdots, \beta_K) \tag{3.19}$$

$$q(\boldsymbol{\Phi}) \propto \prod_{k=1}^{K} \mathrm{Dicichlet}(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV}) \tag{3.20}$$

のハイパーパラメータの初期値をランダムに設定する。ただし、ランダムに取り得る範囲は `seq()` に任意の値を指定する。

・変分ベイズ推定

以下、for() ループ内の処理の内容。

・ハイパーパラメータの初期化

```r
# 次ステップのハイパーパラメータ（alpha, beta）を初期化
new_alpha_k <- rep(alpha, K)
new_beta_kv <- matrix(beta, nrow = K, ncol = V,
                      dimnames = list(paste0("k=", 1:K),
                                      paste0("v=", 1:V)))
```

各パラメーターの更新時に用いる変数を作成する。
また、次ステップに入る前に事前分布のパラメータ α, β で、それぞれ初期化する。

・負担率の計算

```r
for(k in 1:K){ ##（各トピック）

  # 負担率（q_dk）を計算
  tmp_q_alpha <- digamma(alpha_k[k]) - digamma(sum(alpha_k))
  tmp_q_beta  <- sum(N_dv[d, ] * digamma(beta_kv[k, ])) - N_d[d] * digamma(sum(beta_kv[k, ]))
  q_dk[d, k]  <- exp(tmp_q_alpha + tmp_q_beta)

}

# 負担率の正規化：sum_{k=1}^K q_dk = 1
if(sum(q_dk[d, ]) > 0) { ##（全ての値が 0 の場合は回避）
  q_dk[d, ] <- q_dk[d, ] / sum(q_dk[d, ])
} else if(sum(q_dk[d, ]) == 0) {
  q_dk[d, ] <- 1 / K
}
```

　負担率を

$$q_{dk} \propto \exp\left(\Psi(\alpha_{z_d}) - \Psi\Bigl(\sum_{k'=1}^K \alpha_{k'}\Bigr) + \sum_{v=1}^V N_{dv}\Psi(\beta_{z_d v}) + N_d \Psi\Bigl(\sum_{v=1}^V \beta_{z_d v}\Bigr)\right) \tag{3.22}$$

で行う。

　ただしこの式がイコールの関係でないことから、そのままの値を用いずに $\sum_{k=1}^K q_{dk} = 1$ となるように正規化した値を用いる。正規化処理の時点で 1~K までの要素が必要であるため、`for(k in 1:K)` のループ処理を先に回し切り、正規化後のパラメータの更新では、もう一度 `for(k in 1:K)` のループ処理を行う。

　`N_d`, `N_dv` の値によって `q_dk` の値が 0 となることがある。文書 d に関する負担率の全ての値が 0 になると、正規化時の分母 `sum(q_dk[d, ])` も 0 になってしまう。そこでエラー回避のため、`if()` を使って全ての値が 0 のときは別の値で計算を行う。

・ハイパーパラメータ：α_k の更新

```
# ハイパーパラメータ (alpha_k) を更新
new_alpha_k[k] <- new_alpha_k[k] + q_dk[d, k]
```

　更新式

$$\alpha_k = \alpha + \sum_{d=1}^D q_{dk}$$

の計算を行い値を求める。$\sum_{d=1}^D$ については、`for(d in 1:D)` によって繰り返し足していくことがこれに相当する。

・ハイパーパラメータ：β_{kv} の更新

```
# ハイパーパラメータ (beta_kv) を更新
new_beta_kv[k, v] <- new_beta_kv[k, v] + q_dk[d, k] * N_dv[d, v]
```

　β_{kv} も同様に、更新式

$$\beta_{kv} = \beta + \sum_{d=1}^D q_{dk} N_{dv}$$

の計算を行い値を求める。

・ハイパーパラメータの更新

```
# ハイパーパラメータの (alpha, beta) を更新
alpha_k <- new_alpha_k
beta_kv <- new_beta_kv
```

各パラメータを次の計算に用いるために移す。

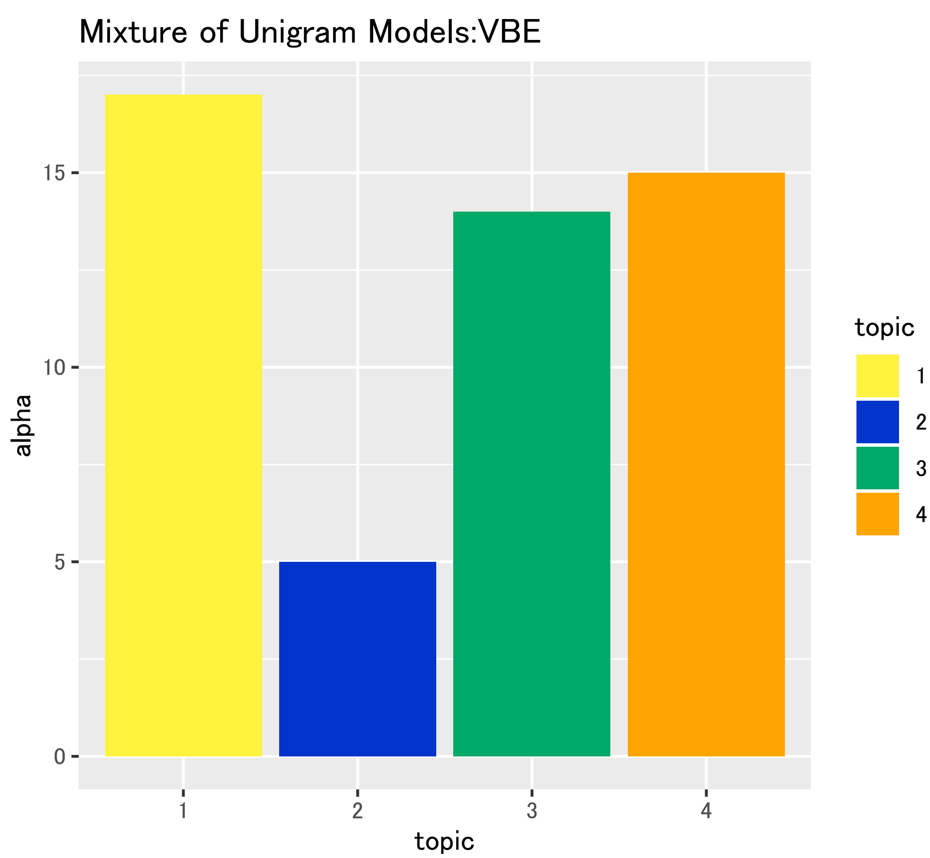
この推定処理を指定した回数繰り返す。(本当は収束条件を満たすまで繰り返すようにしたい…)

・推定結果の確認

・ハイパーパラメータの推定結果の確認

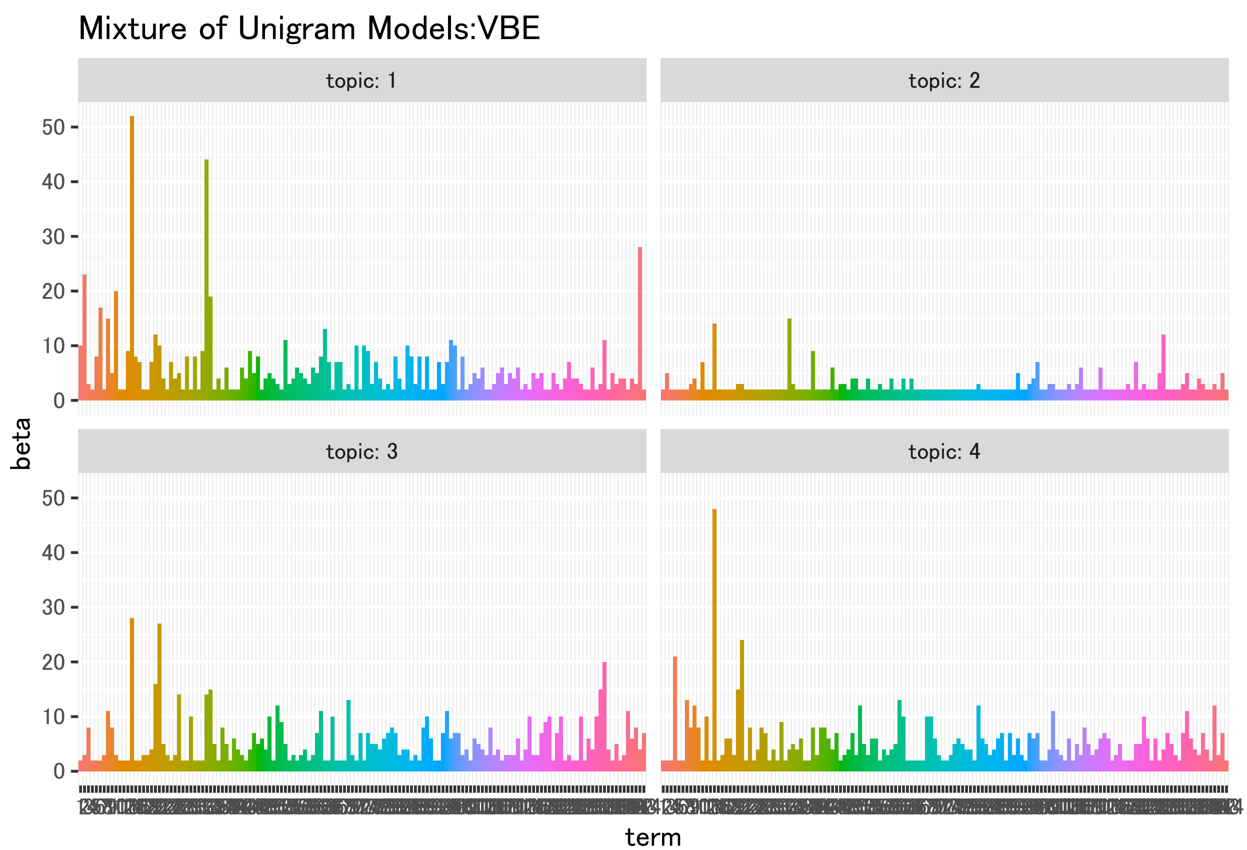
・トピック分布のパラメータ：α_k

```
## ハイパーパラメータ (alpha) の推定結果を確認
# データフレームを作成
alpha_df <- data.frame(topic = as.factor(1:K),
                       alpha = alpha_k)

# 作図
ggplot(data = alpha_df, mapping = aes(x = topic, y = alpha, fill = topic)) + # データ
  geom_bar(stat = "identity", position = "dodge") +  # 棒グラフ
  labs(title = "Mixture of Unigram Models:VBE")      # タイトル
```



・単語分布のパラメータ：β_{kv}

```
## ハイパーパラメータ (beta) の推定結果を確認
# データフレームを作成
beta_df_wide <- cbind(as.data.frame(beta_kv), as.factor(1:K))
colnames(beta_df_wide) <- c(1:V, "topic")
beta_df_long <- gather(beta_df_wide, key = "term", value = "beta", -topic)
beta_df_long$term <- as.factor(as.numeric(beta_df_long$term))

# 作図
ggplot(data = beta_df_long, mapping = aes(x = term, y = beta, fill = term)) + # データ
  geom_bar(stat = "identity", position = "dodge") +  # 棒グラフ
  facet_wrap(~topic, labeller = label_both) +         # グラフの分割
  theme(legend.position = "none") +                   # 凡例
  labs(title = "Mixture of Unigram Models:VBE")       # タイトル
```



最尤推定ではパラメータを点推定したが、変分ベイズ推定ではハイパーパラメータを推定することでパラメータを分布推定している。

・パラメータの推定結果 (平均値) の確認

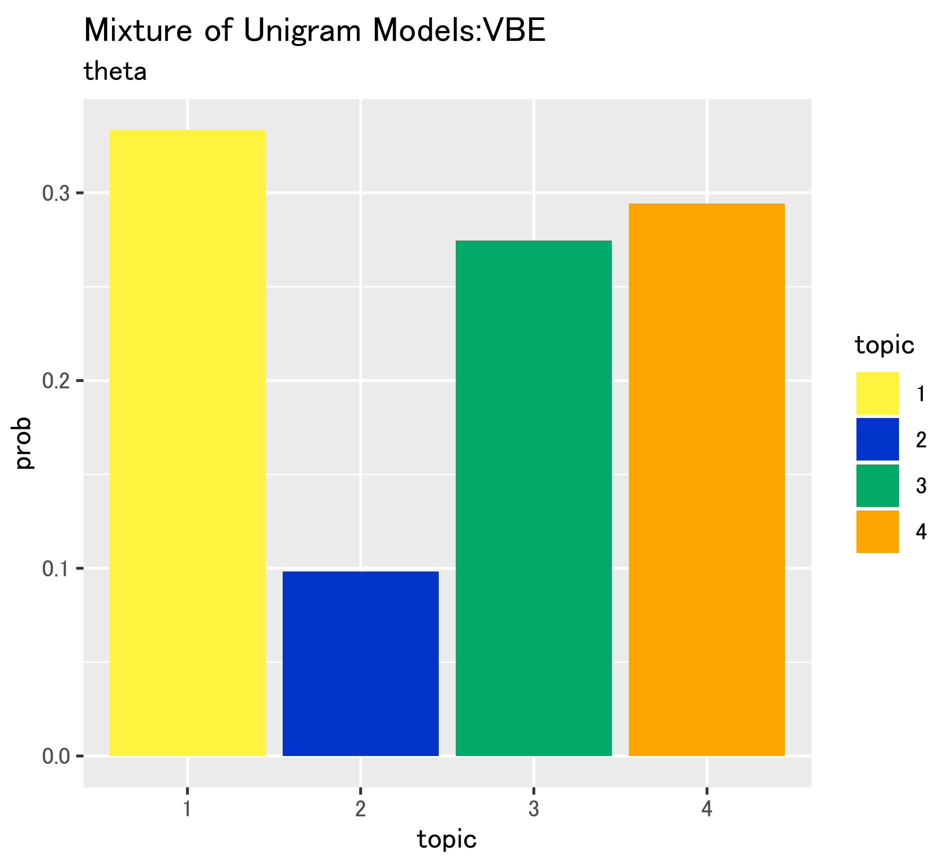
ディリクレ分布の平均値の計算式

$$\mathbb{E}[\phi_v] = \frac{\beta_v}{\sum_{v=1}^V \beta_v} \tag{1.15}$$

より、パラメータの平均値の分布を確認する。

・トピック分布

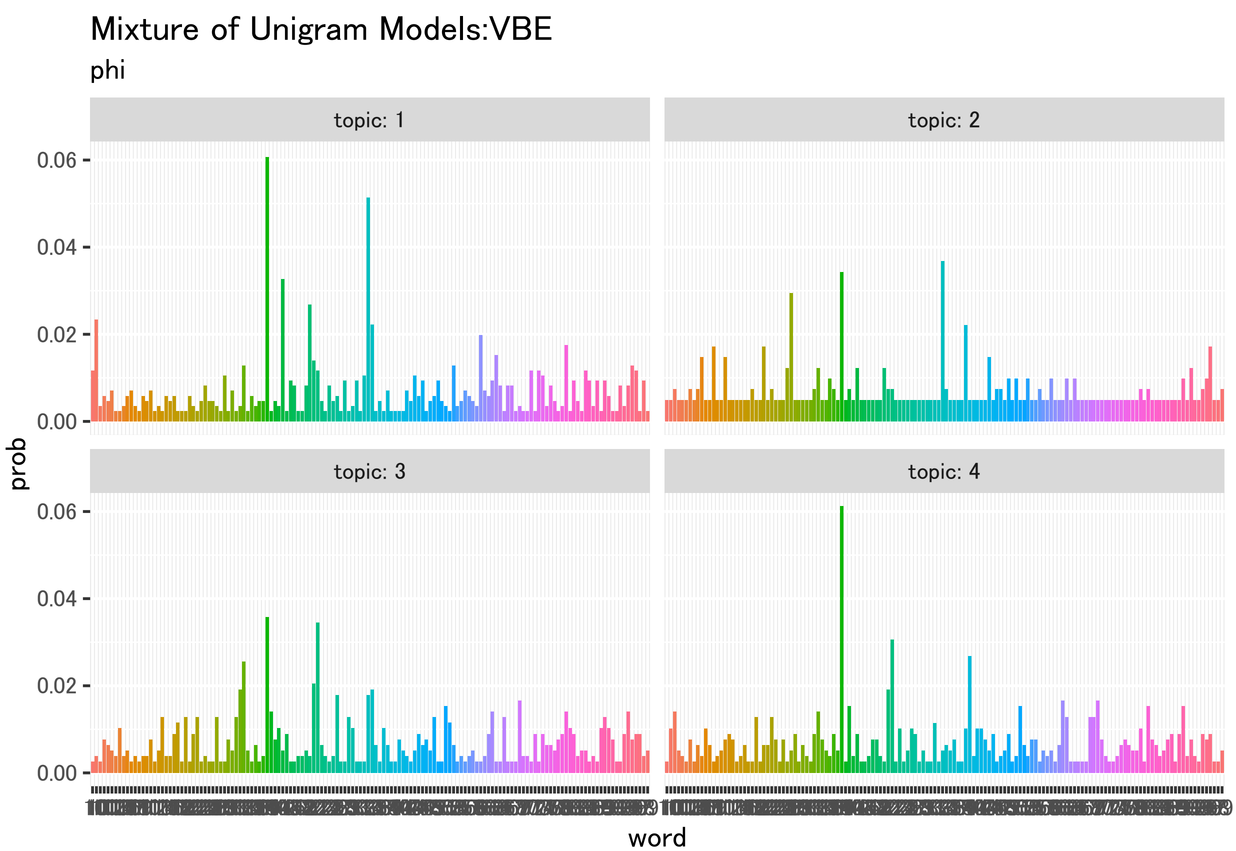
```
## トピック分布の推定結果（平均値）を確認
# パラメータの平均値を算出
theta_k  <- alpha_k / sum(alpha_k) # 平均値を計算

# データフレームを作成
theta_df <- data.frame(topic = as.factor(1:K),
                       prob = theta_k)

# 描画
ggplot(data = theta_df, mapping = aes(x = topic, y = prob, fill = topic)) + # データ
  geom_bar(stat = "identity", position = "dodge") +                  # 棒グラフ
  labs(title = "Mixture of Unigram Models:VBE", subtitle = "theta") # タイトル
```



・単語分布

```
## 単語分布の推定結果（平均値）を確認
# パラメータの平均値を算出
```

```r
phi_kv <- matrix(nrow = K, ncol = V)
for(k in 1:K) {
  phi_kv[k, ] <- beta_kv[k, ] / sum(beta_kv[k, ])
}

# データフレームを作成
phi_df_wide <- cbind(as.data.frame(phi_kv),
                     as.factor(1:K))

# データフレームを long 型に変換
colnames(phi_df_wide) <- c(1:V, "topic") # key 用の行名を付与
phi_df_long <- gather(phi_df_wide, key = "word", value = "prob", -topic) # 変換
phi_df_long$term <- phi_df_long$word %>%
                    as.numeric() %>%
                    as.factor()

# 描画
ggplot(data = phi_df_long, mapping = aes(x = word,  y = prob, fill = word)) + # データ
  geom_bar(stat = "identity", position = "dodge") +  # 棒グラフ
  facet_wrap(~ topic, labeller = label_both) +        # グラフの分割
  theme(legend.position = "none") +                   # 凡例
  labs(title = "Mixture of Unigram Models:VBE", subtitle = "phi") # タイトル
```



・推定推移の確認

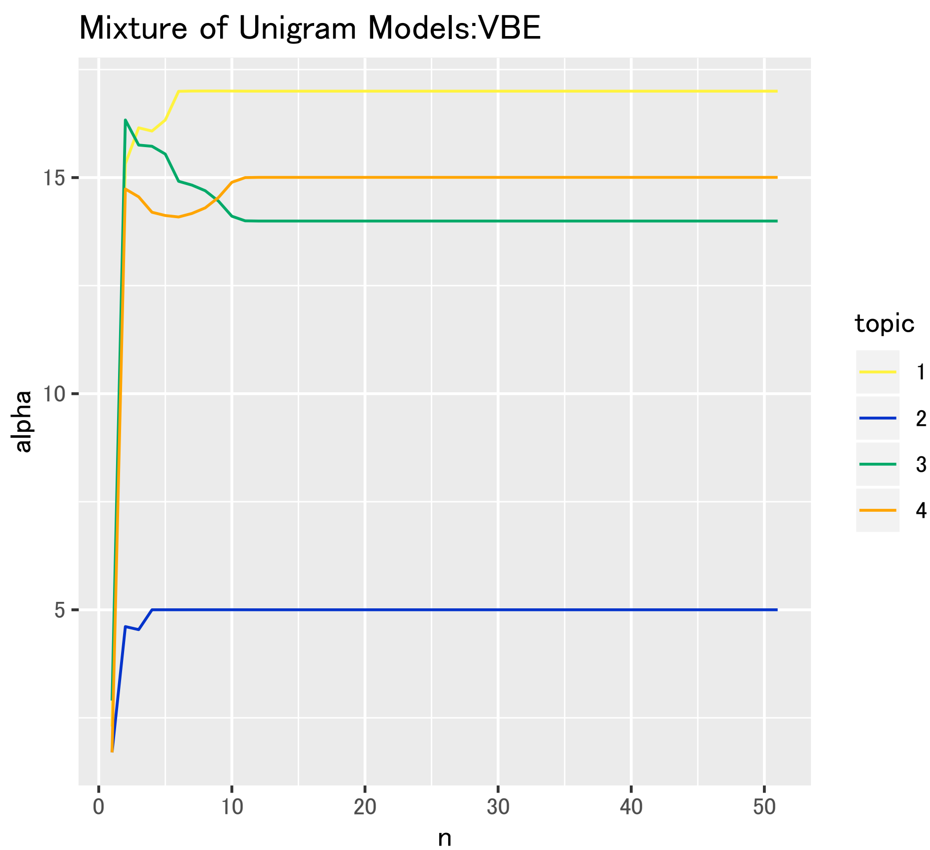
・推移の記録

```r
# 推移の確認用
trace_alpha <- as.matrix(alpha_k)
trace_beta  <- beta_kv[1, ]

# 推移の確認用
trace_alpha <- cbind(trace_alpha, alpha_k)
trace_beta  <- cbind(trace_beta, beta_kv[1, ])
```

推移を確認するために、ハイパーパラメータを更新する度に記録しておく。単語分布については、トピック 1 のみ確認する。

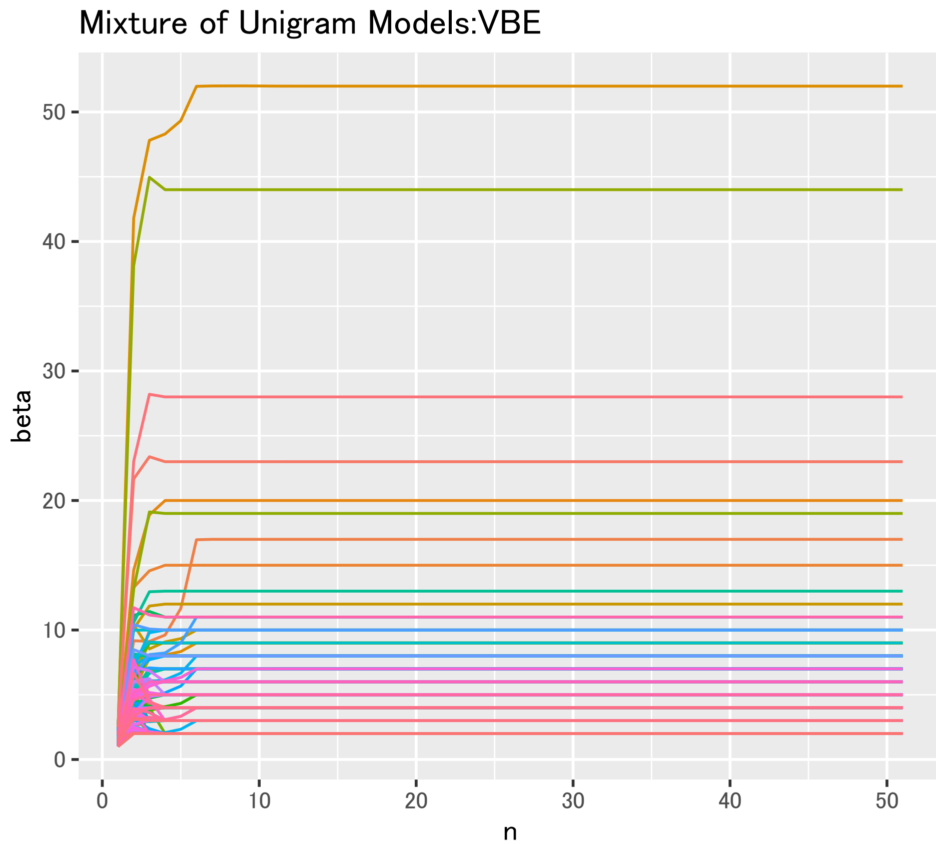
・トピック分布

```r
## トピック分布のパラメータ
# データフレームを作成
trace_alpha_df_wide <- cbind(t(trace_alpha),
                             1:ncol(trace_alpha)) %>%
                       as.data.frame()

# データフレームを long 型に変換
colnames(trace_alpha_df_wide) <- c(1:K, "n") # key 用の行名を付与
trace_alpha_df_long <- gather(trace_alpha_df_wide, key = "topic", value = "alpha", -n) # 変換

# 描画
ggplot(data = trace_alpha_df_long, mapping = aes(x = n, y = alpha, color = topic)) + # データ
  geom_line() +                                 # 棒グラフ
  labs(title = "Mixture of Unigram Models:VBE") # タイトル
```



・単語分布

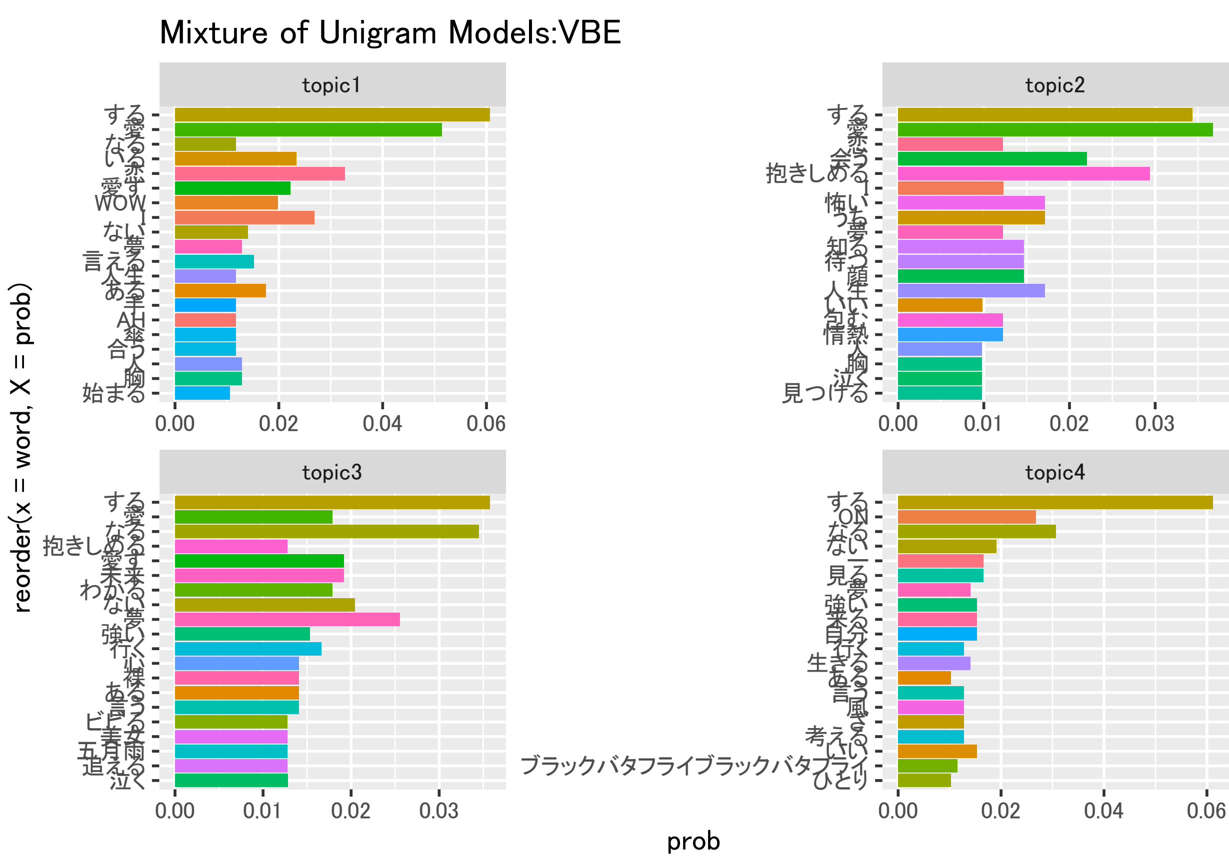
```
## 単語分布のパラメータ
# データフレームを作成
trace_beta_df_wide <- cbind(t(trace_beta),
                            1:ncol(trace_beta)) %>%
                      as.data.frame()

# データフレームを long 型に変換
colnames(trace_beta_df_wide) <- c(1:V, "n") # key 用の行名を付与
trace_beta_df_long <- gather(trace_beta_df_wide, key = "word", value = "beta", -n) # 変換
trace_beta_df_long$word <- trace_beta_df_long$word %>%  # 文字列になるため因子に変換
                          as.numeric() %>%
                          as.factor()

# 描画
ggplot(data = trace_beta_df_long, mapping = aes(x = n, y = beta, color = word)) + # データ
  geom_line() +                                   # 棒グラフ
  theme(legend.position = "none") +               # 凡例
  labs(title = "Mixture of Unigram Models:VBE") # タイトル
```



・各トピックの出現確率の上位語の確認

```
### 各トピックの出現確率の上位語

## 語彙インデックス (v)
# 指定した出現回数以上の単語の行番号
num <- mecab_df %>%
       select(-c(TERM, POS1, POS2)) %>%
       apply(1, sum) >= 5 # 抽出する総出現回数を指定する

v_index <- mecab_df[num, ] %>%
           filter(grepl("名詞|形容詞|^動詞", POS1)) %>%  # 抽出する品詞を指定する
           filter(grepl("一般|^自立", POS2)) %>%
           filter(!grepl(stop_words, TERM)) %>%
           .[, 1]  # 単語の列を抽出する

# データフレームを作成
phi_df_wide2 <- cbind(as.data.frame(t(phi_kv)),
                      v_index) %>%
                as.data.frame()
colnames(phi_df_wide2) <- c(paste0("topic", 1:K), "word") # key 用の行名を付与

# データフレームの整形
phi_df_long2 <- data.frame()
for(i in 1:K) {
```

```
  tmp_df <- phi_df_wide2 %>%
    select(paste0("topic", i), word) %>%  #
    arrange(-.[, 1]) %>%  # 降順に並べ替え
    head(20) %>%          # 任意で指定した上位 n 語を抽出
    gather(key = "topic", value = "prob", -word) # long 型に変換

  phi_df_long2 <- rbind(phi_df_long2, tmp_df)
}

# 描画
ggplot(data = phi_df_long2,
       mapping = aes(x = reorder(x = word, X = prob), y = prob, fill = word)) +  # データ
  geom_bar(stat = "identity", position = "dodge") + # 棒グラフ
  coord_flip() +                                    # グラフの向き
  facet_wrap( ~ topic, scales = "free") +           # グラフの分割
  theme(legend.position = "none") +                 # 凡例
  labs(title = "Mixture of Unigram Models:VBE")     # タイトル
```



3.5 ギブズサンプリング

3.5.1 マルコフ連鎖モンテカルロ法

変分ベイズ推定では周辺尤度の下限を最大化することで事後分布の近似分布を求めた。他にも混合ユニグラムモデルをベイズ推定する方法として、ギブズサンプリングがある。ギブズサンプリングはマルコフ連鎖モンテカルロ法の 1 種で、近似ではなく真の事後分布からサンプリングできる方法である。

例えば、$p(z_1, \cdots, z_D, \boldsymbol{W})$ を推定するとする。変数 z_d 以外の全ての変数が与えられたときの z_d の条件付き確率 $p(z_d|z_1, \cdots, z_{d-1}, z_{d+1}, \cdots, z_D, \boldsymbol{W})$ に従って z_d の値をサンプリングすることを全ての変数 $d = 1, \cdots, D$ に対して繰り返すことにより、目的の分布から事例を得る。十分多くの事例が得られれば目的の分布は次式の経験分布で近似できる。

$$p(z_1, \cdots, z_D, |\boldsymbol{W}) \simeq \frac{1}{S} \sum_{s=1}^{S} \delta \left((z_1, \cdots, z_D), (z_1^{(s)}, \cdots, z_D^{(s)}) \right)$$

ここで、S はサンプリングした事例数、$z_d^{(s)}$ は s 番目にサンプリングした変数 z_d の事例を表す。また、分布 $p(z_1, \cdots, z_D, |\boldsymbol{W})$ における関数 $f(z_1, \cdots, z_D)$ の期待値は

$$\mathbb{E}[f(z_1, \cdots, z_D)] \simeq \frac{1}{S} \sum_{s=1}^{S} f(z_1^{(s)}, \cdots, z_D^{(s)})$$

で近似できる。

3.5.2 パラメータの周辺化

混合ユニグラムモデルには $\boldsymbol{z}, \boldsymbol{\theta}, \boldsymbol{\Phi}$ の未知変数がある。本節では、パラメータ $\boldsymbol{\theta}, \boldsymbol{\Phi}$ を周辺化 (積分消去) し、トピック集合 $\boldsymbol{z}$ の事後分布を推定する。
パラメータを周辺化するギブズサンプリングは崩壊型ギブズサンプリングと呼ばれる。パラメータを周辺化してサンプリングする変数の数を減らすことで、効率的な推定が可能となる。

パラメータ $\boldsymbol{\theta}, \boldsymbol{\Phi}$ を周辺化したときの文書集合 $\boldsymbol{W}$ とトピック集合 $\boldsymbol{z}$ の同時分布は

$$p(\boldsymbol{W}, \boldsymbol{z}|\alpha, \beta) = p(\boldsymbol{z}|\alpha) p(\boldsymbol{W}|\boldsymbol{z}, \beta)$$

となる。ここで、トピック集合 $\boldsymbol{z}$ は α から ($\boldsymbol{\theta}$ を通して) 生成され、単語集合 $\boldsymbol{W}$ は $\boldsymbol{z}$ と β から ($\boldsymbol{\Phi}$ を通して) 生成されるという、混合ユニグラムモデルの生成過程を利用して同時分布を分解している。

この式の 1 つ目の因子は $\boldsymbol{\theta}$ を周辺化して求めている。

$$
\begin{aligned}
p(\boldsymbol{z}|\alpha) &= \int p(\boldsymbol{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\alpha)d\boldsymbol{\theta} && \text{(i)} \\
&= \int \prod_{k=1}^{K} \theta_k^{D_k} \frac{\Gamma(\sum_{k=1}^{K} \alpha)}{\prod_{k=1}^{K} \Gamma(\alpha)} \prod_{k=1}^{K} \theta_k^{\alpha-1} d\boldsymbol{\theta} && \text{(ii)} \\
&= \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \int \prod_{k=1}^{K} \theta_k^{D_k+\alpha-1} d\boldsymbol{\theta} && \text{(iii)} \\
&= \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \frac{\prod_{k=1}^{K} \Gamma(D_k+\alpha)}{\Gamma(\sum_{k=1}^{K} D_k+\alpha)} && \text{(iv)} \\
&= \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \frac{\prod_{k=1}^{K} \Gamma(D_k+\alpha)}{\Gamma(D+\alpha K)} && \text{(v, 3.24)} \\
&= \frac{\Gamma(\alpha K)}{\Gamma(D+\alpha K)} \prod_{k=1}^{K} \frac{\Gamma(D_k+\alpha)}{\Gamma(\alpha)} && \text{(3.24')}
\end{aligned}
$$

【途中式の途中式】

- (i)：

$p(\boldsymbol{z}|\boldsymbol{\theta}) = \mathrm{Categoical}(\boldsymbol{z}|\boldsymbol{\theta})$ であるため、1.2.2 節より

$$
\begin{aligned}
p(\boldsymbol{z}|\boldsymbol{\theta}) &= \mathrm{Categoical}(\boldsymbol{z}|\boldsymbol{\theta}) \\
&= \prod_{d=1}^{D} \mathrm{Categoical}(z_d = k|\boldsymbol{\theta}) \\
&= \prod_{d=1}^{D} \theta_{z_d} \\
&= \prod_{k=1}^{K} \theta_k^{D_k}
\end{aligned}
$$

となるので置き換える。

　これは例えば、文書 $d = 1, 2, 3$ のトピックがそれぞれ $k = 1, 1, 2$ のとき、トピック 1 の文書数は $D_1 = 2$ であり、トピック 2 の文書数は $D_2 = 1$ になる。このときの同時確率分布は

$$
\begin{aligned}
p(z_1 = 1, z_2 = 1, z_3 = 2|\boldsymbol{\theta}) &= \theta_1 * \theta_1 * \theta_2 \\
&= \theta_1^2 * \theta_k^1 \\
&= \prod_{k=1}^{K} \theta_k^{D_k}
\end{aligned}
$$

となる。
また、$p(\boldsymbol{\theta}|\alpha) = \mathrm{Dirichlet}(\boldsymbol{\theta}|\alpha)$ であるため、1.2.4 節より置き換える。

- (ii)：$\boldsymbol{\theta}$ と関係のない項を $\int$ の外に出し、θ_k を 1 つにまとめて、式を整理する。
- (iii)：ディリクレ分布の正規化項 (1.13) より、置き換える。
- (iv)：トピック k が割り振られた文書数 D_k の総和は、文書数 D である。
- (v)：分母分子を αK と α で揃えるために、分母を入れ替える。

同様に、2 つ目の因子は $\boldsymbol{\Phi}$ を周辺化して求めている。

$$
\begin{aligned}
p(\boldsymbol{W}|\boldsymbol{z},\beta) &= \int p(\boldsymbol{W}|\boldsymbol{z},\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\beta)d\boldsymbol{\Phi} && \text{(i)} \\
&= \int \prod_{k=1}^{K}\prod_{v=1}^{V} \phi_{kv}^{N_{kv}} \prod_{k=1}^{K} \frac{\Gamma(\sum_{v=1}^{V}\beta)}{\prod_{v=1}^{V}\Gamma(\beta)} \prod_{v=1}^{V} \phi_{kv}^{\beta-1} d\boldsymbol{\Phi} && \text{(ii)} \\
&= \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{KV}} \int \prod_{k=1}^{K}\prod_{v=1}^{V} \phi_{kv}^{N_{kv}+\beta-1} d\boldsymbol{\Phi} && \text{(iii)} \\
&= \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{KV}} \prod_{k=1}^{K} \frac{\prod_{v=1}^{V}\Gamma(N_{kv}+\beta)}{\Gamma(\sum_{v=1}^{V} N_{kv}+\beta)} && \text{(iv)} \\
&= \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{KV}} \prod_{k=1}^{K} \frac{\prod_{v=1}^{V}\Gamma(N_{kv}+\beta)}{\Gamma(N_k+\beta V)} && \text{(v, 3.25)} \\
&= \frac{\Gamma(\beta V)^K}{\Gamma(N_k+\beta V)} \prod_{k=1}^{K}\prod_{v=1}^{V} \frac{\Gamma(N_{kv}+\beta)}{\Gamma(\beta)^{KV}} && \text{(3.25')}
\end{aligned}
$$

【途中式の途中式】

- (i)：
  - $p(\boldsymbol{W}|\boldsymbol{z},\boldsymbol{\Phi}) = \mathrm{Categoical}(\boldsymbol{W}|\boldsymbol{z},\boldsymbol{\Phi})$ であるため、3.1 節より置き換える。
  - $p(\boldsymbol{\Phi}|\beta) = \mathrm{Dirichlet}(\boldsymbol{\Phi}|\beta)$ であるため、1.2.4 節より置き換える。
- (ii)：$\boldsymbol{\Phi}$ と関係のない項を $\int$ の外に出し、ϕ_{kv} を 1 つにまとめて、式を整理する。
- (iii)：ディリクレ分布の正規化項 (1.13) より、置き換える。
- (iv)：トピック k が割り振られた文書における語彙 v の出現回数 N_{kv} を $v = 1 \cdots V$ まで和をとると、トピック k が割り振られた文書における総単語数 N_k になる。
- (v)：分母分子を βV と β で揃えるために、分母を入れ替える。

3.5.3 サンプリング式

ギブズサンプリングに必要な条件付き確率は、ベイズの定理 (1.4) を用いて

$$
p(z_d = k|\boldsymbol{W},\boldsymbol{z}_{\backslash d},\alpha,\beta) \propto p(z_d = k|\boldsymbol{z}_{\backslash d},\alpha)p(\boldsymbol{w}_d|\boldsymbol{W}_{\backslash d},z_d = k,\boldsymbol{z}_{\backslash d},\beta)
$$

と計算できる。ここで $\boldsymbol{z}_{\backslash d} = (z_1,\cdots,z_{d-1},z_{d+1},\cdots,z_D)$ は、$\boldsymbol{z}$ から z_d のみを取り除いたトピックの集合を表す。

この式の 1 つ目の因子を求める。

$$
\begin{aligned}
p(z_d = k | \boldsymbol{z}_{\backslash d}, \alpha) &= \frac{p(z_d = k, \boldsymbol{z}_{\backslash d} | \alpha)}{p(\boldsymbol{z}_{\backslash d} | \alpha)} &\text{(i)} \\
&= \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \frac{\Gamma(D_{k\backslash d} + 1 + \alpha) \prod_{k' \neq k} \Gamma(D_{k'\backslash d} + \alpha)}{\Gamma(D + \alpha K)} \frac{\Gamma(\alpha)^K}{\Gamma(\alpha K)} \frac{\Gamma(D - 1 + \alpha K)}{\prod_{k'=1}^K \Gamma(D_{k'\backslash d} + \alpha)} &\text{(ii)} \\
&= \frac{(D_{k\backslash d} + \alpha)\Gamma(D_{k\backslash d} + \alpha) \prod_{k' \neq k} \Gamma(D_{k'\backslash d} + \alpha)}{(D - 1 + \alpha K)\Gamma(D - 1 + \alpha K)} \frac{\Gamma(D - 1 + \alpha K)}{\prod_{k'=1}^K \Gamma(D_{k'\backslash d} + \alpha)} &\text{(iii)} \\
&= \frac{D_{k\backslash d} + \alpha}{D - 1 + \alpha K} &(3.26)
\end{aligned}
$$

【途中式の途中式】

- (0)：$\frac{p(z_d=k,\boldsymbol{z}_{\backslash d}|\alpha)}{p(z_d=k,\boldsymbol{z}_{\backslash d}|\alpha)} = 1$ を掛ける。??
- (i)：式 (3.24) を用いる。

ここで D_k はトピック k を割り振られた文書数である。文書 d のトピック z_d が k のとき、トピック k を割り振られた文書 d 以外の文書数 $D_{k\backslash d}$ は $D_k - 1$ である。つまり

$$
D_{k\backslash d} = \begin{cases} D_k - 1 & \text{if} \quad z_d = k \\ D_k & \text{if} \quad z_d \neq k \end{cases}
$$

であることが分かる。これを D_k について解くと

$$
D_k = \begin{cases} D_{k\backslash d} + 1 & \text{if} \quad z_d = k \\ D_{k\backslash d} & \text{if} \quad z_d \neq k \end{cases}
$$

となる。これを式 (3.24) に代入して、$p(z_d = k, \boldsymbol{z}_{\backslash d}|\alpha)$ を求める。

$$
\begin{aligned}
p(\boldsymbol{z}|\alpha) &= \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \frac{\prod_{k=1}^K \Gamma(D_k + \alpha)}{\Gamma(D + \alpha K)} &(3.24) \\
\Rightarrow p(z_d = k, \boldsymbol{z}_{\backslash d}|\alpha) &= \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \frac{\Gamma(D_{k\backslash d} + 1 + \alpha) \prod_{k' \neq k} \Gamma(D_{k'\backslash d} + \alpha)}{\Gamma(D + \alpha K)}
\end{aligned}
$$

$z_d = k$ であるため、z_d(文書 d) に関する項は $\Gamma(D_{k\backslash d} + 1 + \alpha)$ として $\prod_{k=1}^K$ から取り出している。それ以外の $\boldsymbol{z}_{\backslash d}$ に関する項が、$\prod_{k' \neq k} \Gamma(D_{k'\backslash d} + \alpha)$ である。

同様に、分母 $p(\boldsymbol{z}_{\backslash d}|\alpha)$ は

$$
p(\boldsymbol{z}_{\backslash d}|\alpha) = \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \frac{\prod_{k=1}^K \Gamma(D_{k\backslash d} + \alpha)}{\Gamma(D - 1 + \alpha K)}
$$

となる。こちらの分母は、文書 d 以外の総文書数 $\sum_{k=1}^K D_{k\backslash d}$ となるため、$D - 1$ になる。

これらを用いて

$$
\begin{aligned}
p(z_d = k | \boldsymbol{z}_{\backslash d}, \alpha) &= \frac{p(z_d = k, \boldsymbol{z}_{\backslash d}|\alpha)}{p(\boldsymbol{z}_{\backslash d}|\alpha)} \\
&= \frac{\Gamma(\alpha K)}{\Gamma(D + \alpha K)} \frac{\Gamma(D_{k\backslash d} + 1 + \alpha) \prod_{k' \neq 1} \Gamma(D_{k'\backslash d} + \alpha)}{\Gamma(\alpha)^K} \frac{\Gamma(D + \alpha K)}{\Gamma(\alpha K)} \frac{\Gamma(\alpha)^K}{\prod_{k'=1}^K \Gamma(D_{k'\backslash d} + \alpha)}
\end{aligned}
$$

が得られる。

- (ii)：$\Gamma(x) = (x-1)\Gamma(x-1)$ であることから、変形する。
- (iii)：約分をして、式を整理する。ただし、$\prod_{k'=1}^K \Gamma(D_{k'\backslash d}+\alpha) = \Gamma(D_{k'\backslash d}+\alpha)\prod_{k'\neq k}\Gamma(D_{k'\backslash d}+\alpha)$ である。

同様に、2 つ目の因子を求める。

$$
\begin{aligned}
&p(\boldsymbol{w}_d|\boldsymbol{W}_{\backslash d}, z_d = k, \boldsymbol{z}_{\backslash d}, \beta) \\
&= \frac{p(\boldsymbol{W}|z_d = k, \boldsymbol{z}_{\backslash d}, \beta)}{p(\boldsymbol{W}_{\backslash d}|\boldsymbol{z}_{\backslash d}, \beta)} && \text{(i)} \\
&= \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{KV}} \frac{\prod_{v=1}^V \Gamma(N_{kv\backslash d} + N_{dv} + \beta)}{\Gamma(N_{k\backslash d} + N_d + \beta V)} \prod_{k'\neq k} \frac{\prod_{v=1}^V \Gamma(N_{k'v\backslash d} + \beta)}{\Gamma(N_{k'\backslash d} + \beta V)} \frac{\Gamma(\beta)^{KV}}{\Gamma(\beta V)^K} \prod_{k'=1}^K \frac{\Gamma(N_{k'\backslash d} + \beta V)}{\prod_{v=1}^V \Gamma(N_{k'v\backslash d} + \beta)} && \text{(ii)} \\
&= \frac{\prod_{v=1}^V \Gamma(N_{kv\backslash d} + N_{dv} + \beta)}{\Gamma(N_{k\backslash d} + N_d + \beta V)} \frac{\Gamma(N_{k\backslash d} + \beta V)}{\prod_{v=1}^V \Gamma(N_{kv\backslash d} + \beta)} && \text{(iii)} \\
&= \frac{\Gamma(N_{k\backslash d} + \beta V)}{\Gamma(N_{k\backslash d} + N_d + \beta V)} \prod_{v:N_{dv}>0} \frac{\Gamma(N_{kv\backslash d} + N_{dv} + \beta)}{\Gamma(N_{kv\backslash d} + \beta)}
\end{aligned}
$$

【途中式の途中式】

- (0)：
  - $p(\boldsymbol{w}_d) = \frac{p(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_d,\cdots,\boldsymbol{w}_D)}{p(\boldsymbol{w}_1,\cdots,\boldsymbol{w}_{d-1},\boldsymbol{w}_{d+1},\cdots,\boldsymbol{w}_D)}$ ということ ??
  - $z_d = k$ は $\boldsymbol{W}_{\backslash d}$ に関係ないことから省いている。
- (i)：式 (3.25) を用いる。

ここで N_{kv} はトピック k が割り振られた全文書における語彙 v の出現回数 (単語数) であり、N_{dv} は文書 d における語彙 v の出現回数である。文書 d のトピック z_d が k のとき、文書 d 以外のトピック k が割り当てられた単語数 $N_{kv\backslash d}$ は $N_{kv} - N_{dv}$ となる。

また、$z_d = k$ のとき、トピック k が割り当てられた文書 d 以外の総単語数 $N_{k\backslash d}$ は、トピック k が割り当てられた総単語数 N_k から文書 d の単語数 N_d を除いたものとなる。

これをまとめると

$$
N_{kv\backslash d} = \begin{cases} N_{kv} - N_{dv} & \text{if} \quad z_d = k \\ N_{kv} & \text{if} \quad z_d \neq k \end{cases}
$$

$$
N_{k\backslash d} = \begin{cases} N_k - N_d & \text{if} \quad z_d = k \\ N_k & \text{if} \quad z_d \neq k \end{cases}
$$

である。これをそれぞれ N_{kv} と N_k について解くと

$$
N_{kv} = \begin{cases} N_{kv\backslash d} + N_{dv} & \text{if} \quad z_d = k \\ N_{kv\backslash d} & \text{if} \quad z_d \neq k \end{cases}
$$

$$
N_k = \begin{cases} N_{k\backslash d} + N_d & \text{if} \quad z_d = k \\ N_{k\backslash d} & \text{if} \quad z_d \neq k \end{cases}
$$

となる。これを式 (3.25) に代入すると

$$
\begin{aligned}
p(\boldsymbol{W}|\boldsymbol{z},\beta) &= \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{KV}} \prod_{k=1}^{K} \frac{\prod_{v=1}^{V}\Gamma(N_{kv}+\beta)}{\Gamma(N_k+\beta V)} \qquad (3.25)\\
\Rightarrow p(\boldsymbol{W}|z_d=k,\boldsymbol{z}_{\backslash d},\beta) &= \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{KV}} \frac{\prod_{v=1}^{V}\Gamma(N_{kv\backslash d}+N_{dv}+\beta)}{\Gamma(N_{k\backslash d}+N_d+\beta V)} \prod_{k'\neq k} \frac{\prod_{v=1}^{V}\Gamma(N_{k'v\backslash d}+\beta)}{\Gamma(N_{k'\backslash d}+\beta V)}
\end{aligned}
$$

となる。$z_d = k$ であるため、z_d(文書 d) に関する項は 2 つ目の因子として $\prod_{k=1}^{K}$ から取り出している。それ以外の $\boldsymbol{z}_{\backslash d}$ に関する項が 3 つ目の因子である。

同様に分母は $p(\boldsymbol{W}_{\backslash d}|\boldsymbol{z}_{\backslash d},\beta)$ は

$$
p(\boldsymbol{W}_{\backslash d}|\boldsymbol{z}_{\backslash d},\beta) = \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{KV}} \prod_{k'=1}^{K} \frac{\prod_{v=1}^{V}\Gamma(N_{k'v\backslash d}+\beta)}{\Gamma(N_{k'\backslash d}+\beta V)}
$$

となる。これらを用いて

$$
\begin{aligned}
&\frac{p(\boldsymbol{W}|z_d=k,\boldsymbol{z}_{\backslash d},\beta)}{p(\boldsymbol{W}|\boldsymbol{z},\beta)}\\
&= \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{VK}} \frac{\prod_{v=1}^{V}\Gamma(N_{kv\backslash d}+N_{dv}+\beta)}{\Gamma(N_{k\backslash d}+N_d+\beta V)} \prod_{k'\neq k} \frac{\prod_{v=1}^{V}\Gamma(N_{k'v\backslash d}+\beta)}{\Gamma(N_{k'\backslash d}+\beta V)} \frac{\Gamma(\beta)^{VK}}{\Gamma(\beta V)^K} \prod_{k'=1}^{K} \frac{\Gamma(N_{k'\backslash d}+\beta V)}{\prod_{v=1}^{V}\Gamma(N_{k'v\backslash d}+\beta)}
\end{aligned}
$$

が得られる。

- (ii)：約分をして、式を整理する。
- (iii)：式を整理する。ただし、$\frac{\prod_{k'=1}^{K} x_{k'}}{\prod_{k'\neq k} x_{k'}} = \frac{x_1 * \cdots * x_{k'} * \cdots * x_K}{x_1 * \cdots * x_{k'-1} * x_{k'+1} * \cdots * x_K} = x_{k'}$ であることから、第 3 因子の $k' = k$ の項が残る。
- (iv)：
  - 分母分子で βV と β を揃えるために、分子を入れ替える。
  - $N_{dv} > 0$ でないと $z_d = k$ と $z_d \neq k$ を区別できないため、$\prod_{v:N_{dv}>0}$ とすることで条件を明示している。

これと式 (3.26) を用いてサンプリング式を求める。ただし、式 (3.26) の分母 $\Gamma(N_{k\backslash d}+N_d+\beta V)$ は k に依存しないため省く。

$$
\begin{aligned}
p(z_d=k|\boldsymbol{W},\boldsymbol{z}_{\backslash d},\alpha,\beta) &\propto p(z_d=k|\boldsymbol{z}_{\backslash d},\alpha) p(\boldsymbol{w}_d|\boldsymbol{W}_{\backslash d},z_d=k,\boldsymbol{z}_{\backslash d},\beta)\\
&= \frac{D_{k\backslash d}+\alpha}{D-1+\alpha K} \frac{\Gamma(N_{k\backslash d}+\beta V)}{\Gamma(N_{k\backslash d}+N_d+\beta V)} \prod_{v:N_{dv}>0} \frac{\Gamma(N_{kv\backslash d}+N_{dv}+\beta)}{\Gamma(N_{kv\backslash d}+\beta)}\\
&\propto (D_{k\backslash d}+\alpha) \frac{\Gamma(N_{k\backslash d}+\beta V)}{\Gamma(N_{k\backslash d}+N_d+\beta V)} \prod_{v:N_{dv}>0} \frac{\Gamma(N_{kv\backslash d}+N_{dv}+\beta)}{\Gamma(N_{kv\backslash d}+\beta)} \qquad (3.27)
\end{aligned}
$$

1 つ目の因子はトピック k が割り当てられた文書数に比例していて、割り当てられた文書数が多いトピックになりやすくなっている。また、2 つ目以降の因子は文書 d がトピック k に割り当てられたときの尤度 (ポリヤ分布) であり、文書 d がトピック k に割り当てられた文書集合と似ているほど、トピック k になりやすくなる。

・ハイパーパラメータの更新式

不動点反復法を用いて文書集合 $\boldsymbol{W}$ とトピック集合 $\boldsymbol{z}$ の同時確率分布 $p(\boldsymbol{W}, \boldsymbol{z}|\alpha, \beta)$ を最大化することでハイパーパラメータ α, β を推定する。

・**Tips**

$$
\begin{aligned}
&\frac{\Gamma(x)}{\Gamma(n+x)} \geq \frac{\Gamma(\hat{x}) \exp\{(\hat{x}-x)b\}}{\Gamma(n+\hat{x})} \\
&b = \Psi(n+\hat{x}) - \Psi(\hat{x})
\end{aligned}
$$

$$
\begin{aligned}
&\frac{\Gamma(n+x)}{\Gamma(x)} \geq cx^a \\
&a = \Big(\Psi(n+\hat{x}) - \Psi(\hat{x})\Big)\hat{x} \\
&c = \frac{\Gamma(n+\hat{x})}{\Gamma(\hat{x})}\hat{x}^{-a}
\end{aligned}
$$

この関係性 (2.7 節) を用いて、不動点反復法が使えるように式を変形する。

(3.24')(3.25') より

$$
\begin{aligned}
p(\boldsymbol{W}, \boldsymbol{z}|\alpha, \beta) &= \frac{\Gamma(\alpha K)}{\Gamma(D+\alpha K)} \prod_{k=1}^{K} \frac{\Gamma(D_k+\alpha)}{\Gamma(\alpha)} \frac{\Gamma(\beta V)}{\Gamma(N_k+\beta V)} \prod_{k=1}^{K} \prod_{v=1}^{V} \frac{\Gamma(N_{kv}+\beta)}{\Gamma(\beta)} \\
&\geq \frac{\Gamma(\hat{\alpha}K) \exp\{(\hat{\alpha}K - \alpha K)b_1\}}{\Gamma(D+\hat{\alpha}K)} \prod_{k=1}^{K} \frac{\Gamma(D_k+\hat{\alpha})}{\Gamma(\hat{\alpha})} \hat{\alpha}^{-a_1} \alpha^{a_1} \\
&\quad * \frac{\Gamma(\hat{\beta}V) \exp\Big\{(\hat{\beta}V - \beta V)b_2\Big\}}{\Gamma(N_k+\hat{\beta}V)} \prod_{k=1}^{K} \prod_{v=1}^{V} \frac{\Gamma(N_{kv}+\hat{\beta})}{\Gamma(\hat{\beta})} \hat{\beta}^{-a_2} \beta^{a_2} \equiv F
\end{aligned}
$$

ここで

$$
\begin{aligned}
a_1 &= \Big(\Psi(D_k+\hat{\alpha}) - \Psi(\hat{\alpha})\Big)\hat{\alpha} \\
b_1 &= \Psi(D+\hat{\alpha}K) - \Psi(\hat{\alpha}K) \\
a_2 &= \Big(\Psi(N_{kv}+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta} \\
b_2 &= \Psi(N_k+\hat{\beta}V) - \Psi(\hat{\beta}V)
\end{aligned}
$$

である。α に関する項を取り出して対数をとると

$$
\begin{aligned}
F(\alpha) = &\log\Gamma(\hat{\alpha}K) + (\hat{\alpha}K - \alpha K)\Big(\Psi(D+\hat{\alpha}K) - \Psi(\hat{\alpha}K)\Big) - \log\Gamma(D+\hat{\alpha}K) \\
&+ \sum_{k=1}^{K}\Big\{\log\Gamma(D_k+\hat{\alpha}) - \log\Gamma(\hat{\alpha}) - \Big(\Psi(D_k+\hat{\alpha}) - \Psi(\hat{\alpha})\Big)\hat{\alpha}\log\hat{\alpha} + \Big(\Psi(D_k+\hat{\alpha}) - \Psi(\hat{\alpha})\Big)\hat{\alpha}\log\alpha\Big\}
\end{aligned}
$$

となる。これを α 関して微分し $\frac{\partial F(\alpha)}{\partial \alpha} = 0$ となる α を求める。

$$
\begin{aligned}
\frac{\partial F(\alpha)}{\partial \alpha} &= -K\Big(\Psi(D+\hat{\alpha}K) - \Psi(\hat{\alpha}K)\Big) + \frac{1}{\alpha}\sum_{k=1}^{K}\Big(\Psi(D_k+\hat{\alpha}) - \Psi(\hat{\alpha})\Big)\hat{\alpha} = 0 \\
\alpha &= \hat{\alpha}\frac{\sum_{k=1}^{K}\Psi(D_k+\hat{\alpha}) - K\Psi(\hat{\alpha})}{K\Psi(D+\hat{\alpha}K) - K\Psi(\hat{\alpha}K)}
\end{aligned}
$$

この式の $\hat{\alpha}$ を現ステップのパラメータ α とし、左辺の α を更新後 (次ステップ) のパラメータ α^{new} とした

$$
\alpha^{\mathrm{new}} = \alpha\frac{\sum_{k=1}^{K}\Psi(D_k+\alpha) - K\Psi(\alpha)}{K\Psi(D+\alpha K) - K\Psi(\alpha K)} \tag{3.28}
$$

がハイパーパラメータ α の更新式となる。

同様に、F から β に関する項を取り出して対数をとると

$$
\begin{aligned}
F(\beta) = &\log\Gamma(\hat{\beta}V) + (\hat{\beta}V - \beta V)\Big(\Psi(N_k+\hat{\beta}V) - \Psi(\hat{\beta}V)\Big) - \log\Gamma(N_k+\hat{\beta}V) \\
&+ \sum_{k=1}^{K}\sum_{v=1}^{V}\Big\{\log\Gamma(N_{kv}+\hat{\beta}) - \log\Gamma(\hat{\beta}) - \Big(\Psi(N_{kv}+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\log\hat{\beta} + \Big(\Psi(N_{kv}+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\log\beta\Big\}
\end{aligned}
$$

となる。これを β に関して微分し $\frac{\partial F(\beta)}{\partial \beta} = 0$ となる β を求める。

$$
\begin{aligned}
\frac{\partial F(\beta)}{\partial \beta} &= -V\Big(\Psi(N_k+\hat{\beta}V) - \Psi(\hat{\beta}V)\Big) + \frac{1}{\beta}\sum_{k=1}^{K}\sum_{v=1}^{V}\Big(\Psi(N_{kv}+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta} = 0 \\
\beta &= \hat{\beta}\frac{\sum_{k=1}^{K}\sum_{v=1}^{V}\Psi(N_{kv}+\beta) - KV\Psi(\hat{\beta})}{V\sum_{k=1}^{K}\Psi(N_k+\hat{\beta}V) - KV\Psi(\hat{\beta}V)}
\end{aligned}
$$

この式の $\hat{\beta}$ を現ステップのパラメータ β とし、左辺の β を更新後 (時ステップ) のパラメータ β^{new} とした

$$
\beta^{\mathrm{new}} = \beta\frac{\sum_{k=1}^{K}\sum_{v=1}^{V}\Psi(N_{kv}+\beta) - KV\Psi(\beta)}{V\sum_{k=1}^{K}\Psi(N_k+\beta V) - KV\Psi(\beta V)} \tag{3.29}
$$

がハイパーパラメータ β の更新式となる。

・**R で組んでみる**

・コード全体

・テキスト処理

```
# 利用パッケージ
library(RMeCab)
library(tidyverse)

### テキスト処理

# 抽出しない単語を指定
stop_words <- "[a-z]" # 小文字のアルファベットを含む語

# 形態素解析
mecab_df <- docDF("フォルダ名", type = 1) # テキストファイルの保存先を指定する

# 文書 d の語彙 v の出現回数 (N_dv) の集合
N_dv <- mecab_df %>%
        filter(grepl("名詞|形容詞|^動詞", POS1)) %>% # 抽出する品詞（大分類）を指定する
        filter(grepl("一般|^自立", POS2)) %>%        # 抽出する品詞（細分類）を指定する
        filter(!grepl(stop_words, TERM)) %>%        # ストップワードを除く
        select(-c(TERM, POS1, POS2)) %>%            # 数値列のみを残す
        filter(apply(., 1, sum) >= 5) %>%           # 抽出する総出現回数を指定する
        t()                                         # 転置

# 確認用の行列名
dimnames(N_dv) <- list(paste0("d=", 1:nrow(N_dv)), # 行名
                       paste0("v=", 1:ncol(N_dv))) # 列名

# 文書 d の単語数 (N_d) のベクトル
N_d <- apply(N_dv, 1, sum) # 行方向に和をとる

# 文書数 (D)
D <- nrow(N_dv)

# 総語彙数 (V)
V <- ncol(N_dv)
```

・パラメータの初期設定

```
### パラメータの初期設定

# トピック数 (K)
K <- 4   # 任意の値を指定する

# ハイパーパラメータ (alpha, beta)
alpha <- 2 # 任意の値を指定する
beta  <- 2 # 任意の値を指定する

# トピック k が割り当てられた文書数 (D_{z_d}) のベクトルの初期化
D_k <- rep(0, K)

# トピック k が割り当てられた文書の語彙 v の出現回数 (N_{z_d,v}) の集合の初期化
```

```r
N_kv <- matrix(0, nrow = K, ncol = V,
               dimnames = list(paste0("k=", 1:K),
                               paste0("v=", 1:V)))

# トピック k が割り当てられた文書の単語数 (N_{z_d}) のベクトルの初期化
N_k <- rep(0, K)

# 文書 d のトピック (z_d) のベクトルの初期化
z_d <- rep(0, D)
```

・ギブスサンプリング

```r
### ギブスサンプリング

# 受け皿の用意
p <- NULL

# 結果の確認用
trace_alpha <- alpha
trace_beta  <- beta

for(i in 1:3000) { # 任意の回数を指定する

  # 新たに割り当られてるトピックに関するカウントを初期化
  new_D_k  <- rep(0, K)
  new_N_kv <- matrix(0, nrow = K, ncol = V,
                     dimnames = list(paste0("k=", 1:K),
                                     paste0("v=", 1:V)))
  new_N_k  <- rep(0, K)

  for(d in 1:D) { ## (各文書：1,...,D)

    # 現ステップの計算のために前ステップのカウントを移す
    tmp_D_k  <- D_k
    tmp_N_kv <- N_kv
    tmp_N_k  <- N_k

    if(z_d[d] > 0) { ## 初回を飛ばす処理

      # 前ステップで文書 d が与えられたトピックを `k`に代入
      k <- z_d[d]

      # 文書 d に関する値をカウントから除く
      tmp_D_k[k]     <- D_k[k] - 1
      tmp_N_kv[k, ] <- N_kv[k, ] - N_dv[d, ]
      tmp_N_k[k]     <- N_k[k] - N_d[d]
    }

    for(k in 1:K) { ## (各トピック：1,...,K)

      # サンプリング確率の計算
      tmp_p_alpha <- log(tmp_D_k[k] + alpha) # 第 1 因子
      tmp_p_beta1 <- lgamma(tmp_N_k[k] + beta * V) - lgamma(tmp_N_k[k] + N_d[d] + beta * V)  # 第 2 因子
      tmp_p_beta2 <- lgamma(tmp_N_kv[k, ] + N_dv[d, ] + beta) - lgamma(tmp_N_kv[k, ] + beta) # 第 3 因子
```

```
      p[k] <- exp(tmp_p_alpha + tmp_p_beta1 + sum(tmp_p_beta2))

    } ## (/各トピック：1,...,K)

    # サンプリング
    tmp_z_d <- rmultinom(n = 1, size = 1, prob = p) # カテゴリ分布によりトピックを生成
    z_d[d]  <- which(tmp_z_d == 1)                   # サンプリングの結果を抽出

    # 新たに割り当てられたトピックを`k`に代入
    k <- z_d[d]

    # 文書dに関する値をカウントに加える
    new_D_k[k]     <- new_D_k[k] + 1
    new_N_kv[k, ] <- new_N_kv[k, ] + N_dv[d, ]
    new_N_k[k]     <- new_N_k[k] + N_d[d]

  } ## (/各文書：1,...,D)

  # カウントを更新
  D_k  <- new_D_k
  N_kv <- new_N_kv
  N_k  <- new_N_k

  # ハイパーパラメータ（）を更新
  tmp_alpha_numer1 <- sum(digamma(D_k + alpha))  # 分子
  tmp_alpha_numer2 <- K * digamma(alpha)          # 分子
  tmp_alpha_denom1 <- K * digamma(D + alpha * K) # 分母
  tmp_alpha_denom2 <- K * digamma(alpha * K)     # 分母
  alpha <- alpha * (tmp_alpha_numer1 - tmp_alpha_numer2) / (tmp_alpha_denom1 - tmp_alpha_denom2)

  # ハイパーパラメータ（）を更新
  tmp_beta_numer1 <- sum(digamma(N_kv + beta))        # 分子
  tmp_beta_numer2 <- K * V * digamma(beta)            # 分子
  tmp_beta_denom1 <- V * sum(digamma(N_k + beta * V)) # 分母
  tmp_beta_denom2 <- K * V * digamma(beta * V)        # 分母
  beta <- beta * (tmp_beta_numer1 - tmp_beta_numer2) / (tmp_beta_denom1 - tmp_beta_denom2)

  # 結果の確認用
  trace_alpha <- c(trace_alpha, alpha)
  trace_beta  <- c(trace_beta, beta)
}
```

・コードの解説

・利用パッケージ

```
# 利用パッケージ
library(RMeCab)
library(tidyverse)
```

形態素解析のための `RMeCab::docDF()` と、グラフ作成時の `tidyr::gather()`、`ggplot2` 関連の関数以外にライブラリの読み込みが必要なのは `dplyr` のみ。(面倒なので `tidyvers` パッケージを読み込んでいる)

その他テキスト処理の部分は 3.3 節を参照のこと。

・パラメータの初期設定

・トピック数

```r
# トピック数 (K)
K <- 4   # 任意の値を指定する
```

任意のトピック数を指定する。

・パラメータの初期値の指定

```r
# ハイパーパラメータ (alpha, beta)
alpha <- 2 # 任意の値を指定する
beta  <- 2 # 任意の値を指定する
```

ハイパーパラメータ α, β の初期値を任意で指定する。

・トピック集合

```r
# 文書 d のトピック (z_d) のベクトルの初期化
z_d <- rep(0, D)
```

z_d は、各文書に割り当てられたトピックで、要素が D 個のベクトルである。初回はどの文書にもトピックが割り当てられていない状態として、全ての要素を 0 としておく。

・トピックごとの各カウントの初期化

```r
# トピック k が割り当てられた文書数 (D_{z_d}) のベクトルの初期化
D_k <- rep(0, K)

# トピック k が割り当てられた文書の語彙 v の出現回数 (N_{z_d,v}) の集合の初期化
N_kv <- matrix(0, nrow = K, ncol = V,
               dimnames = list(paste0("k=", 1:K),
                               paste0("v=", 1:V)))

# トピック k が割り当てられた文書の単語数 (N_{z_d}) のベクトルの初期化
N_k <- rep(0, K)
```

D_k は、トピックごとの割り当てられた文書の数で、要素が K 個のベクトルである。
N_kv は、各文書に割り当てられたトピックごとに語彙 v の出現回数を足し合わせた数で、K 行 V 列のマトリクスである。
N_k は、割り当てられたトピックごとに各文書の単語数 N_d を足し合わせた数で、要素が K 個のベクトルである。

初回はどの文書にもトピックが割り当てられていないため、初期値は全て 0 である。

次からは、for ループ内の処理である。

・ギブスサンプリング

・次ステップのカウントの初期化

```
# 新たに割り当てられてるトピックに関するカウントを初期化
new_D_k  <- rep(0, K)
new_N_kv <- matrix(0, nrow = K, ncol = V,
                   dimnames = list(paste0("k=", 1:K),
                                   paste0("v=", 1:V)))
new_N_k  <- rep(0, K)
```

`new_D_k`、`new_N_kv`、`new_N_k` は、それぞれ現ステップで割り当てられた文書に関するカウントを加算していく受け皿である。そのため、毎ステップの最初にカウントを 0 にリセットする。

次からは、`for(d in 1:D)` のループ内で、各文書順番に処理していく内容である。

・現ステップのカウント

```
# 現ステップの計算のために前ステップのカウントを移す
tmp_D_k  <- D_k
tmp_N_kv <- N_kv
tmp_N_k  <- N_k

if(z_d[d] > 0) { # 初回を飛ばす処理
  # 前ステップで文書 d が与えられたトピックを `k`に代入
  k <- z_d[d]

  # 文書 d に関する値をカウントから除く
  tmp_D_k[k]     <- D_k[k] - 1
  tmp_N_kv[k, ] <- N_kv[k, ] - N_dv[d, ]
  tmp_N_k[k]     <- N_k[k] - N_d[d]
}
```

`tmp_D_k`、`tmp_N_kv`、`tmp_N_k` は、それぞれ $D_{k\backslash d}$、$N_{kv\backslash d}$、$N_{k\backslash d}$ のことである。これらの変数は、現ステップでのサンプリング確率の計算に用いる。

文書ごとに処理していくので、毎回前ステップでのカウントを一度移した上で、`k` 番目トピックに関するの要素から `d` 番目の文書に関するカウントを除く。ただし、カウントの初期値は 0 であるため、初回の処理は行われないように `if(z_d[d] > 0)` として条件分岐を行う必要がある。

まず、`z_d` から前ステップで割り当てられたトピック番号を `k` に代入する。その `k` を添え字として使い、それぞれ `d` 番目の文書に関するカウントを引いた値を上書きする。

`D_k` は文書数であるため、どの文書のときであっても数は 1 なので、`k` 個目の要素から 1 を引く。

`N_kv` の `k` 行目 (番目のトピックに関する) の各語彙の出現回数から、`N_dv` から `d` 行目 (番目の文書に関する) の各語彙の出現回数を引く。

`N_k` の `k` 個目 (番目のトピックの) の単語数から、`d` 個目 (番目の文書の) 単語数を引く。

・サンプリング確率の計算

```r
for(k in 1:K) {
  # サンプリング確率の計算
  tmp_p_alpha <- log(tmp_D_k[k] + alpha) # 第 1 因子
  tmp_p_beta1 <- lgamma(tmp_N_k[k] + beta * V) - lgamma(tmp_N_k[k] + N_d[d] + beta * V)  # 第 2 因子
  tmp_p_beta2 <- lgamma(tmp_N_kv[k, ] + N_dv[d, ] + beta) - lgamma(tmp_N_kv[k, ] + beta) # 第 3 因子
  p[k] <- exp(tmp_p_alpha + tmp_p_beta1 + sum(tmp_p_beta2))
}
```

サンプリング確率の計算は次の式によって行う。

$$p(z_d = k|\boldsymbol{W}, \boldsymbol{z}_{\backslash d}, \alpha, \beta) \propto (D_{k\backslash d} + \alpha)\frac{\Gamma(N_{k\backslash d} + \beta V)}{\Gamma(N_{k\backslash d} + N_d + \beta V)} \prod_{v:N_{dv}>0} \frac{\Gamma(N_{kv\backslash d} + N_{dv} + \beta)}{\Gamma(N_{kv\backslash d} + \beta)} \tag{3.27}$$

因子ごとに分けて計算する。ただし、`gamma()` に渡す値が 171 より大きくなると発散してしまい扱えないため、対数化して計算する `lgamma()` を使う。そのため、他の値に関しても対数をとって計算を行い、最後に `exp()` で指数をとる。

トピックごとに 1 つずつ計算を行い、添え字を使って計算結果を `p` に代入していくため、最初に空のオブジェクトとして `p` を作っておく。

・サンプリング

```r
# サンプリング
tmp_z_d <- rmultinom(n = 1, size = 1, prob = p) # カテゴリ分布によりトピックを生成
z_d[d]  <- which(tmp_z_d == 1)                  # サンプリングの結果を抽出
```

上で求めた確率を基にカテゴリ分布に従ってトピックが決まる。多項分布の乱数発生関数 `rmultinom()` の引数に `n = 1`、`size = 1` を指定して試行回数を 1 回、サンプル数を 1 つとすることで、カテゴリ分布に従う乱数を発生させる。`rmultinom()` の結果は、サンプル確率を指定する引数 `prob` に渡す要素数の行、`n` に指定した試行回数の列のマトリクスで返ってくる。 この節の例では、$1, 2, \cdots, K$ の行を持つ 1 列のマトリクスが返ってくる。要素は列ごとに、割り当てられたトピックの行が 1 でそれ以外は 0 となる。

出力された K 行 1 列のマトリクスに `which()` を使い、要素が 1 の行番号を検索する。返ってきた値 (行番号) がトピック番号に相当するので、それを `z_d` の `d` 個目の要素に代入する。

・カウントの加算

```r
# 新たに割り当てられたトピックを`k`に代入
k <- z_d[d]

# 文書 d に関する値をカウントに加える
new_D_k[k]     <- new_D_k[k] + 1
new_N_kv[k, ] <- new_N_kv[k, ] + N_dv[d, ]
new_N_k[k]     <- new_N_k[k] + N_d[d]
```

`d` 番目の文書に関する値を除いたときと同様に今度は加算していく。割り当てられたトピック番号を `k` に代入して、`new_D_k`、`new_N_kv`、`new_N_k` にそれぞれ対応する値を順次加えていく。

ここまでが文書ごとに処理していく内容であった。ここでの結果を基に、パラメータを更新していく。

・カウントの更新

```
# カウントを更新
D_k  <- new_D_k
N_kv <- new_N_kv
N_k  <- new_N_k
```

全文書に関するカウントを集計できたので、その結果をそれぞれ D_k、N_kv、N_k に上書きする。

・ハイパーパラメータ α を更新

```
# ハイパーパラメータ（）を更新
tmp_alpha_numer1 <- sum(digamma(D_k + alpha))  # 分子
tmp_alpha_numer2 <- K * digamma(alpha)         # 分子
tmp_alpha_denom1 <- K * digamma(D + alpha * K) # 分母
tmp_alpha_denom2 <- K * digamma(alpha * K)     # 分母
alpha <- alpha * (tmp_alpha_numer1 - tmp_alpha_numer2) / (tmp_alpha_denom1 - tmp_alpha_denom2)
```

ハイパーパラメータ α の更新は、次の式によって行う。

$$\alpha^{\text{new}} = \alpha \frac{\sum_{k=1}^{K} \Psi(D_k + \alpha) - K\Psi(\alpha)}{K\Psi(D + \alpha K) - K\Psi(\alpha K)} \tag{3.28}$$

分かりやすくするために、各項を分けて計算した上で、最後にまとめて計算を行う。ディガンマ関数 $\Psi(x)$ の計算は digamma() を使う。

・ハイパーパラメータ β を更新

```
# ハイパーパラメータ（）を更新
tmp_beta_numer1 <- sum(digamma(N_kv + beta))        # 分子
tmp_beta_numer2 <- K * V * digamma(beta)            # 分子
tmp_beta_denom1 <- V * sum(digamma(N_k + beta * V)) # 分母
tmp_beta_denom2 <- K * V * digamma(beta * V)        # 分母
beta <- beta * (tmp_beta_numer1 - tmp_beta_numer2) / (tmp_beta_denom1 - tmp_beta_denom2)
```

ハイパーパラメータ β の更新は、次の式によって行う。

$$\beta^{\text{new}} = \beta \frac{\sum_{k=1}^{K} \sum_{v=1}^{V} \Psi(N_{kv} + \beta) - KV\Psi(\beta)}{V \sum_{k=1}^{K} \Psi(N_k + \beta V) - KV\Psi(\beta V)} \tag{3.29}$$

α と同様に、β も digamma() を使って各項をまず計算する。

・**推定結果の確認**

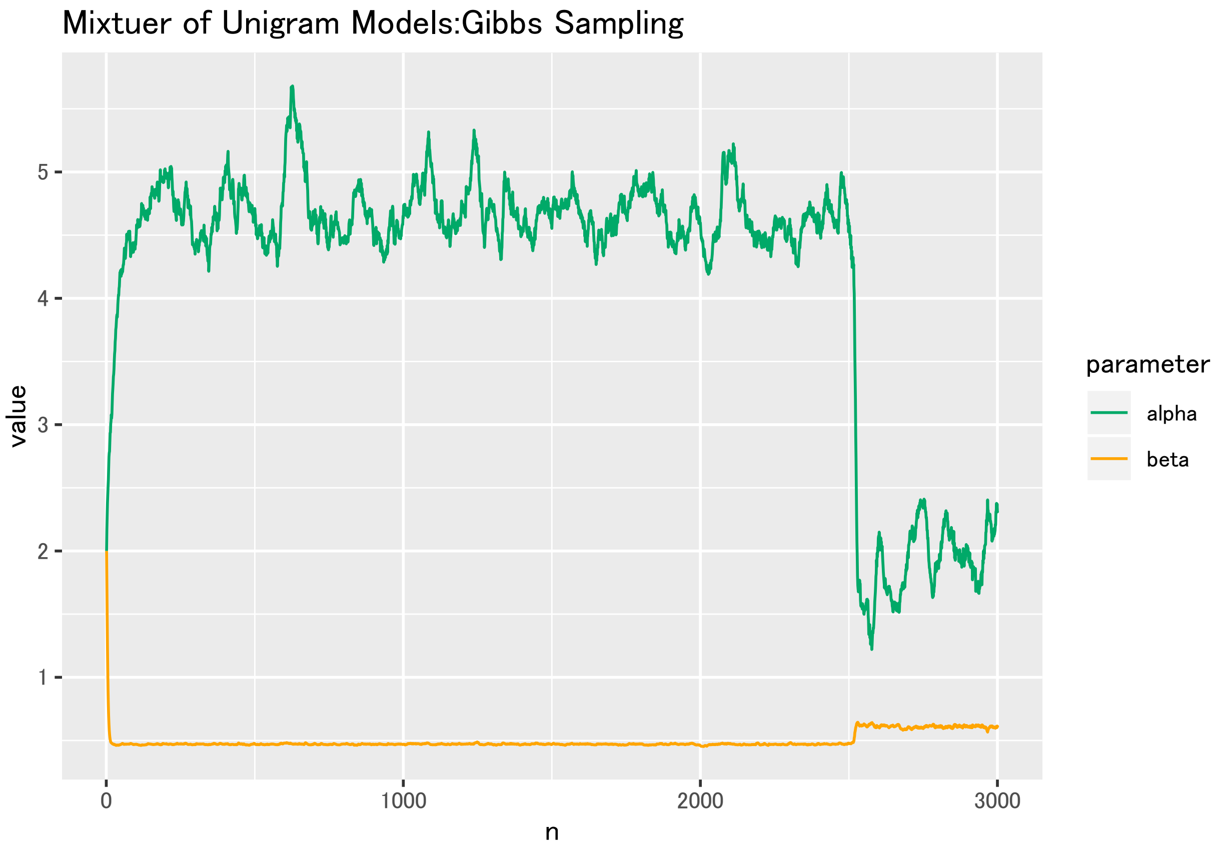
・推移の確認

・推移の記録

```
# 結果の確認用（初期値）
trace_alpha <- alpha
trace_beta  <- beta


# 結果の確認用（更新後）
  trace_alpha <- c(trace_alpha, alpha)
  trace_beta  <- c(trace_beta, beta)
```

・ハイパーパラメータ

```
### 推移の確認

# データフレームを作成
ab_df_wide <- data.frame(n = 1:length(trace_alpha),
                         alpha = trace_alpha,
                         beta = trace_beta)

# データフレームを long 型に変換
ab_df_long <- gather(ab_df_wide, key = "parameter", value = "value", -n)

# 描画
ggplot(ab_df_long, aes(n, value, color = parameter)) +     # データの指定
  geom_line() +                                            # 折れ線グラフ
  scale_color_manual(values = c("#00A968", "Orange")) +    # グラフの色
  labs(title = "Mixtuer of Unigram Models:Gibbs Sampling") # タイトル
```


Mixtuer of Unigram Models:Gibbs Sampling
value
5
4
3
2
1
0
1000
2000
3000
n
parameter
alpha
beta

4 トピックモデル

・記号一覧

記号	意味	制約・関係性など
D	文書数	
$d \in \{1, 2, \cdots, D\}$	文書インデックス	
$\boldsymbol{W} = (\boldsymbol{w}_1, \cdots, \boldsymbol{w}_d, \cdots, \boldsymbol{w}_D)$	文書集合	
$\boldsymbol{w}_d = (w_{d1}, \cdots, w_{dn}, \cdots, w_{dN_D})$	文書 d の単語集合	
$w_{dn} \in \{1, 2, \cdots, V\}$	文書 d の n 番目の単語	$w_{dn} \sim \mathrm{Categorical}(\boldsymbol{\phi}_{z_{dn}})$
N	文書全体での単語数	$N = \sum_{d=1}^{D} N_d = \sum_{v=1}^{V} N_v$
$n \in \{1, 2, \cdots, N_{dv}\}$	単語インデックス	
V	文書全体での語彙数	$N \geq V$
$v \in \{1, 2, \cdots, V\}$	語彙インデックス	
N_d	文書 d に含まれる単語数	$N_d = \sum_{v=1}^{V} N_{dv} = \sum_{k=1}^{K} N_{dk}$
N_v	文書全体での語彙 v の出現回数	$N_v = \sum_{d=1}^{D} N_{dv} = \sum_{k=1}^{K} N_{kv}$
N_{dv}	文書 d での語彙 v の出現回数	
N_k	文書全体でトピック k が割り当てられた単語数	$N_k = \sum_{d=1}^{D} N_{dk} = \sum_{v=1}^{V} N_{kv}$
N_{dk}	文書 d でトピック k が割り当てられた単語数	
N_{kv}	文書全体でトピック k が割り当てられた語彙 v の数	
K	トピック数	
$k \in \{1, 2, \cdots, K\}$	トピック	
$\boldsymbol{Z} = (z_{11}, \cdots, z_{1N_1}, z_{21}, \cdots, z_{DN_D})$	トピック集合	
$z_{dn} \in \{1, 2, \cdots, K\}$	文書 d の n 番目の単語に割り当てられたトピック	$z_{dn} \sim \mathrm{Categorical}(\boldsymbol{\theta}_d)$
$\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \cdots, \boldsymbol{\theta}_k, \cdots, \boldsymbol{\theta}_K)$	トピック分布集合	
$\boldsymbol{\theta}_d = (\theta_{d1}, \cdots, \theta_{dk}, \cdots, \theta_{dK})$	文書 d のトピック分布	$\boldsymbol{\theta}_d \sim \mathrm{Dirichlet}(\alpha)$
θ_{dk}	文書 d においてトピック k が割り当てられる確率	$p(k \vert \boldsymbol{\theta}_d) = \theta_{dk} \geq 0,\ \sum_{k=1}^{K} \theta_{dk} = 1$
$\boldsymbol{\Phi} = (\boldsymbol{\phi}_1, \cdots, \boldsymbol{\phi}_k, \cdots, \boldsymbol{\phi}_K)$	単語分布集合	
$\boldsymbol{\phi}_k = (\phi_{k1}, \cdots, \phi_{kv}, \cdots, \phi_{kV})$	トピック k の単語分布	$\boldsymbol{\phi}_k \sim \mathrm{Dirichlet}(\beta)$
ϕ_{kv}	トピックが k とき語彙 v が生成される確率	$p(v \vert \boldsymbol{\phi}_k) = \phi_{kv} \geq 0,\ \sum_{v=1}^{V} \phi_{kv} = 1$
q_{dnk}	負担率	
α	トピック分布のパラメータ (一様)	
$\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_k, \cdots, \alpha_K)$	トピック分布のパラメータ (多様)	
β	単語分布のパラメータ (一様)	
$\boldsymbol{\beta} = (\beta_1, \cdots, \beta_v, \cdots, \beta_V)$	単語分布のパラメータ (多様)	

4.1 トピックモデル

トピック分布 $\boldsymbol{\theta}_d$ と単語分布集合 $\boldsymbol{\phi}$ が与えられたときの文書 $\boldsymbol{w}_d$ の確率は以下となる。

$$
\begin{aligned}
p(\boldsymbol{w}_d|\boldsymbol{\theta}_d, \boldsymbol{\Phi}) &= \prod_{n=1}^{N_d} \sum_{k=1}^{K} p(z_{dn} = k|\boldsymbol{\theta}_d) p(w_{dn}|\boldsymbol{\phi}_k) \qquad \text{(i)} \\
&= \prod_{n=1}^{N_d} \sum_{k=1}^{K} \theta_{dk} \phi_{k w_{dn}}
\end{aligned}
$$

【途中式の途中式】

- (0)：
  - トピック分布 $\boldsymbol{\theta}_d$ に従って文書 d の n 番目の単語のトピック z_{dn} が決まることから,$p(z_{dn} = k|\boldsymbol{\theta}_d)$ となる。これを文書 d の全ての単語 $\boldsymbol{w}_d$ に対して行うので、$\prod_{n=1}^{N_d} p(z_{dn} = k|\boldsymbol{\theta}_d)$ となる。
  - 文書 d の n 番目の単語 w_{dn} は、割り当てられたトピックに基づき各トピックの単語分布 $\boldsymbol{\phi}_k$ に従って生成されることから、$p(w_{dn}|\boldsymbol{\phi}_k)$ となる。これも同様に、文書 d の全ての単語 $\boldsymbol{w}_d$ に対して行うので、$\prod_{n=1}^{N_d} p(w_{dn}|\boldsymbol{\phi}_k)$ となる。
  - また、これらをトピック $1, \cdots, K$ についても行う。ただし、各トピックに関しては同時に起こらないので足し合わせる。
- (i)：
  - 文書 d にトピック k が割り当てられる確率は θ_{dk} である。
  - トピックが k のとき、単語 w_{dn} が出現する確率は $\phi_{k w_{dn}}$ である。

4.2 グラフィカルモデル

グラフィカルモデル表現は、生成モデルを図示するためによく使われるものである。ノード (円) とエッジ (矢印) によって、モデル内の変数の依存関係が直感的に分かる。

色付きのノードは観測変数、白地のノードは未知変数を表している。また、四角は繰り返しを表し、四角内の数字はその繰り返し回数を表している。

・ユニグラムモデル

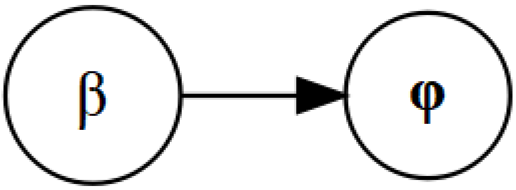
まずは、2 章で扱ったユニグラムモデルの生成モデルを例とする。

$$
p(\boldsymbol{W}|\boldsymbol{\phi}, \beta) = p(\boldsymbol{W}|\boldsymbol{\phi}) p(\boldsymbol{\phi}|\beta)
$$

ユニグラムモデルの事前分布 $p(\boldsymbol{\phi}|\beta)$ をグラフィカルモデルで表すと

```
library(DiagrammeR)

DiagrammeR::grViz("
  digraph dot{
    graph [rankdir = LR]

    node [shape = circle]
```

```
    beta [label = '&beta;']
    phi [label = <<B>&phi;</B>>]

    edge []
      beta -> phi;
  }
")
```

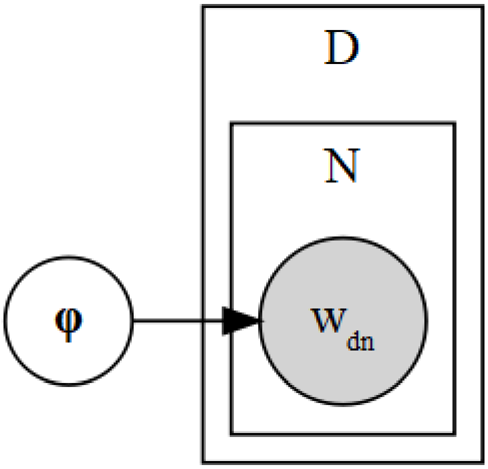


である。β という条件のもとで $\boldsymbol{\phi}$ が決まることが矢印で示されている。
続いて、尤度 $p(\boldsymbol{W}|\boldsymbol{\phi})$ について見る。尤度の式は

$$p(\boldsymbol{W}|\boldsymbol{\phi}) = \prod_{d=1}^{D}\prod_{n=1}^{N_d} p(w_{dn}|\boldsymbol{\phi})$$

と変形できる。この式をグラフィカルモデルで表すと

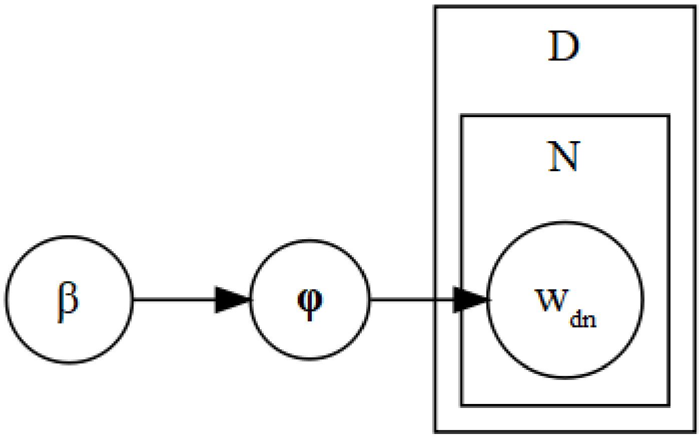
```
grViz("
  digraph dot{
    graph [rankdir = LR]

    node [shape = circle]

    phi [label = <<B>&phi;</B>>]

    subgraph cluster_D{
      label = D

      subgraph cluster_N{
        label = N

        w [label = 'w@_{dn}', style = filled, filledcolor = 'gray']
      }
    }

    edge []
      phi -> w;
  }
")
```



である。w_{dn} は観測した文書データであるため、色付きのノードとなっている。その w_{dn} は、ϕ を条件としてが決まることを矢印が示している。また、w_{dn} は $\prod$ が示す通り $1, \cdots, D$、$1, \cdots, N$ まで掛け合わせる。それを四角でそれぞれ D 回、N 回繰り返すことを表している。

2 つを組み合わせると、β によって ϕ が生成され、またその ϕ によって w_{dn} が生成されることから、生成モデル $\prod_{d=1}^D \prod_{n=1}^{N_d} p(w_{dn}|\phi)p(\phi|\beta)$ は

```
# ユニグラムモデル
grViz("
  digraph dot{
    graph [rankdir = LR]

    node [shape = circle]

    beta [label = '&beta;']
    phi [label = <<B>&phi;</B>>]

    subgraph cluster_D{
      label = D

      subgraph cluster_N{
        label = N

        w [label = 'w@_{dn}', style = filled, filledcolor = 'gray']
      }
    }

    edge []
      beta -> phi -> w
  }
")
```



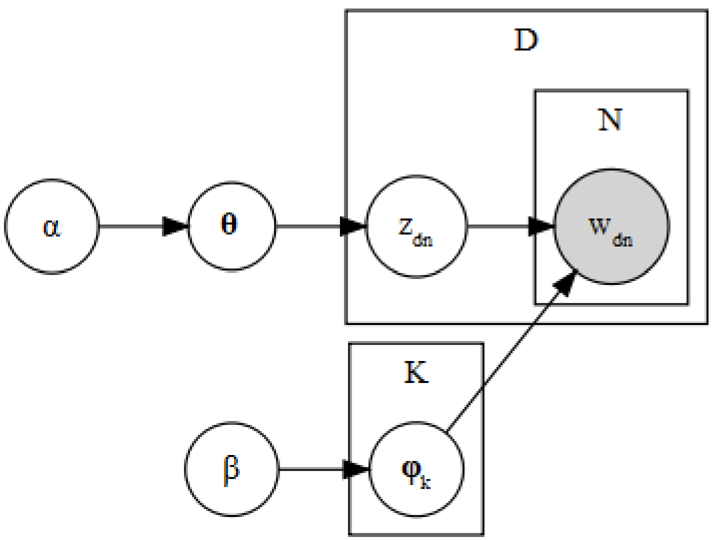
と表される。

・混合ユニグラムモデル

$$p(\boldsymbol{z}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\alpha)p(\boldsymbol{W}|\boldsymbol{z},\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\beta) = \prod_{d=1}^{D} p(z_d|\boldsymbol{\theta})p(\boldsymbol{\theta}|\alpha) \prod_{n=1}^{N_d} p(w_{dn}|z_d,\boldsymbol{\phi}_k) \prod_{k=1}^{K} p(\boldsymbol{\phi}_k|\beta)$$

```
# 混合ユニグラムモデル
grViz("
  digraph dot{
    graph [rankdir = LR]

    node [shape = circle]

    subgraph Cluster_alpha{
      alpha [label = '&alpha;']
      theta [label = <<B>&theta;</B>>]

      subgraph cluster_D{
        label = D

        z [label = 'z@_{dn}']

        subgraph cluster_N{
          label = N

          w [label = 'w@_{dn}', style = filled, filledcolor = 'gray']
        }
      }

      edge []
```

```
      alpha -> theta -> z -> w;
    }

    subgraph Cluster_beta{
      beta [label = '&beta;']

      subgraph cluster_K{
        label = K

        phi [label = <<B>&phi;</B>@_{k}>]
      }

      edge []
      beta -> phi -> w;
    }
  }
")
```

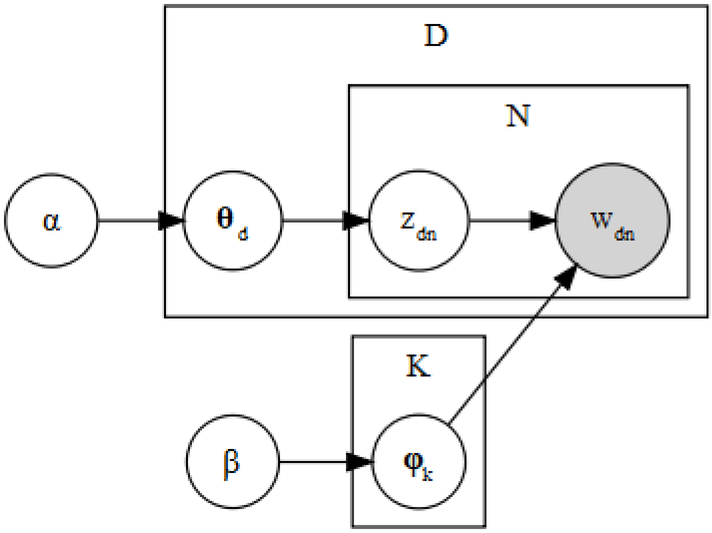


・トピックモデル

$$p(\boldsymbol{z}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\alpha)p(\boldsymbol{W}|\boldsymbol{z},\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\beta) = \prod_{d=1}^{D}\prod_{n=1}^{N_d} p(z_{dn}|\boldsymbol{\theta}_d)p(\boldsymbol{\theta}_d|\alpha)p(w_{dn}|z_{dn},\boldsymbol{\phi}_k)\prod_{k=1}^{K} p(\boldsymbol{\phi}_k|\beta)$$

```
# トピックモデル
grViz("
  digraph dot{
    graph [rankdir = LR]

    node [shape = circle]
```

```
    subgraph Cluster_alpha{
      alpha [label = '&alpha;']

      subgraph cluster_D{
        label = D

        theta [label = <<B>&theta;</B>@_{d}>]

        subgraph cluster_N{
          label = N

          z [label = 'z@_{dn}']
          w [label = 'w@_{dn}', style = filled, filledcolor = 'gray']
        }
      }

      edge []
        alpha -> theta -> z -> w;
    }

    subgraph Cluster_beta{
        beta [label = '&beta;']

      subgraph cluster_K{
        label = K

        phi [label = <<B>&phi;</B>@_{k}>]
      }

      edge []
        beta -> phi -> w;
    }
  }
")
```



各モデルの理解や比較の助けとなる。

4.3 最尤推定

トピックモデルを最尤推定する手法は、確率的潜在意味解析 (PLSA) と呼ばれる。

3.3 節の混合ユニグラムモデルの最尤推定と同様に、EM アルゴリズムを用いて対数尤度を最大にするパラメータ $\mathbf{\Theta}, \mathbf{\Phi}$ を求める。

$$
\begin{aligned}
L &= \log p(\boldsymbol{W}|\boldsymbol{\Theta}, \boldsymbol{\Phi}) && \text{(i)} \\
&= \log \prod_{d=1}^{D} p(\boldsymbol{w}_d|\boldsymbol{\theta}_d, \boldsymbol{\Phi}) && \text{(ii)} \\
&= \sum_{d=1}^{D}\sum_{n=1}^{N_d} \log \sum_{k=1}^{K} \theta_{dk}\phi_{kw_{dn}} && \text{(iii)} \\
&= \sum_{d=1}^{D}\sum_{n=1}^{N_d} \log \sum_{k=1}^{K} q_{dnk}\frac{\theta_{dk}\phi_{kw_{dn}}}{q_{dnk}} && \text{(iv)} \\
&\geq \sum_{d=1}^{D}\sum_{n=1}^{N_d}\sum_{k=1}^{K} q_{dnk} \log \left(\frac{\theta_{dk}\phi_{kw_{dn}}}{q_{dnk}}\right) && \text{(v)} \\
&= \sum_{d=1}^{D}\sum_{n=1}^{N_d}\sum_{k=1}^{K} q_{dnk}(\log \theta_{dk}\phi_{kw_{dn}} - \log q_{dnk}) && \text{(vi)} \\
&= \sum_{d=1}^{D}\sum_{n=1}^{N_d}\sum_{k=1}^{K} q_{dnk} \log \theta_{dk}\phi_{kw_{dn}} - \sum_{d=1}^{D}\sum_{n=1}^{N_d}\sum_{k=1}^{K} q_{dnk} \log q_{dnk} \equiv F
\end{aligned}
$$

【途中式の途中式】

- (i)：$\boldsymbol{W}, \boldsymbol{\Theta}$ の集合をそれぞれ展開する。
- (ii)：4.1 節より、置き換える。
- (iii)：求めたい負担率 q_{dnk} を $\frac{q_{dnk}}{q_{dnk}} = 1$ として分割して掛け合わせる。
- (iv)：この式が上に凸な対数関数であることから、イエンゼンの不等式 (1.17) を用いて、下限を求める。
- (v, vi)：式を展開する。

・**E ステップ**

E ステップでは、下限 F を最大にする文書 d の n 番目の単語がトピック k となる負担率 q_{dnk} を求める。

$\sum_{k=1}^{K} q_{dnk} = 1$ の制約の下での下限 F の最大化問題として、ラグランジュの未定乗数法を用いて解く。

$$
F(q_{dnk}) = \sum_{d=1}^{D}\sum_{n=1}^{N_d}\sum_{k=1}^{K} q_{dnk}(\log \theta_{dk}\phi_{kw_{dn}} - \log q_{dnk}) + \lambda \left(\sum_{k=1}^{K} q_{dnk} - 1\right)
$$

この式を q_{dnk} について微分して $\frac{\partial F(q_{dnk})}{\partial q_{dnk}} = 0$ となる q_{dnk} を求める。

$$
\begin{aligned}
\frac{\partial F(q_{dnk})}{\partial q_{dnk}} &= \log \theta_{dk}\phi_{kw_{dn}} - \log q_{dnk} - 1 + \lambda = 0 \\
\exp(\log q_{dnk}) &= \exp(\log \theta_{dk}\phi_{kw_{dn}} + \lambda - 1) \\
q_{dnk} &= \theta_{dk}\phi_{kw_{dn}} \exp(\lambda - 1) && (1)
\end{aligned}
$$

両辺を $k = 1$ から K まで和をとると、$\sum_{k=1}^{K} q_{dnk} = 1$ より

$$
\begin{aligned}
\sum_{k=1}^{K} q_{dnk} &= \sum_{k=1}^{K} \theta_{dk}\phi_{kw_{dn}} \exp(\lambda - 1) \\
1 &= \sum_{k=1}^{K} \theta_{dk}\phi_{kw_{dn}} \exp(\lambda - 1) \\
\exp(\lambda - 1) &= \frac{1}{\sum_{k=1}^{K} \theta_{dk}\phi_{kw_{dn}}}
\end{aligned}
$$

となる。これを式 (1) に代入すると

$$
q_{dnk} = \frac{\theta_{dk}\phi_{kw_{dn}}}{\sum_{k'=1}^{K} \theta_{dk'}\phi_{k'w_{dn}}} \tag{4.1}
$$

が得られる。これが負担率の計算式となる。
負担率 q_{dnk} はその文書のトピック分布 θ_{dk} とその単語のトピック k に比例していることが分かる。

・M ステップ

M ステップでは、今求めた負担率 q_{dnk} を用いて下限 F を最大にするパラメータを求める。

E ステップと同様に、M ステップでも $\sum_{k=1}^{K} \theta_{dk} = \sum_{kv}^{V} \phi_{kv} = 1$ の制約の下での下限 F の最大化問題として、ラグランジュの未定乗数法を用いて求める。まずは、θ_{dk} を求めていく。

$$
F(\theta_{dk}) = \sum_{n=1}^{N_d} q_{dnk}(\log\theta_{dk} + \log\phi_{kw_{dn}} - \log q_{dnk}) + \lambda\left(\sum_{k=1}^{K} \theta_{dk} - 1\right)
$$

この式を θ_{dk} に関して微分をして、$\frac{\partial F(\theta_{dk})}{\partial \theta_{dk}} = 0$ となる θ_{dk} を求める。

$$
\begin{aligned}
\frac{\partial F(\theta_{dk})}{\partial \theta_{dk}} = \frac{\sum_{n=1}^{N_d} q_{dnk}}{\theta_{dk}} + \lambda = 0& \\
\theta_{dk} = \frac{\sum_{n=1}^{N_d} q_{dnk}}{-\lambda}& \qquad (2)
\end{aligned}
$$

両辺を $k = 1$ から K まで和をとると、$\sum_{k=1}^{K} \theta_{dk} = 1$ より

$$
\begin{aligned}
\sum_{k=1}^{K} \theta_{dk} &= \frac{\sum_{k=1}^{K}\sum_{n=1}^{N_d} q_{dnk}}{-\lambda} \\
1 &= \frac{\sum_{k=1}^{K}\sum_{n=1}^{N_d} q_{dnk}}{-\lambda} \\
-\lambda &= \sum_{k=1}^{K}\sum_{n=1}^{N_d} q_{dnk}
\end{aligned}
$$

となる。これを式 (2) に代入すると

$$\theta_{dk} = \frac{\sum_{n=1}^{N_d} q_{dnk}}{\sum_{k'=1}^{K} \sum_{n=1}^{N_d} q_{dnk'}} \tag{4.2}$$

が得られる。この式が θ_{dk} の更新式となる。

θ_{dk} は、その文書内の単語のトピック k に関する負担率の和に比例していることが分かる。

ϕ_{kv} についても同様に

$$F(\phi_{kv}) = \sum_{d=1}^{D} q_{dnk}(\log\theta_{dk} + \log\phi_{kw_{dn}} - \log q_{dnk}) + \lambda\left(\sum_{v=1}^{V} \phi_{dw_{dn}} - 1\right)$$

を $\phi_{kw_{dn}}$ に関して微分して、$\frac{\partial F(\phi_{kw_{dn}})}{\partial \phi_{kw_{dn}}} = 0$ となる ϕ_{kv} を求める。

$$\begin{aligned}
\frac{\partial F(\phi_{kw_{dn}})}{\partial \phi_{kw_{dn}}} &= \frac{\sum_{d=1}^{D} q_{dnk}}{\phi_{kw_{dn}}} + \lambda = 0 \\
\phi_{kw_{dn}} &= \frac{\sum_{d=1}^{D} q_{dnk}}{-\lambda} \\
\sum_{n:w_{dn}=v} \phi_{kw_{dn}} &= \frac{\sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk}}{-\lambda} \\
\phi_{kv} &= \frac{\sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk}}{-\lambda} \qquad (3)
\end{aligned}$$

ここで $\sum_{n:w_{dn}=v}$ は、$w_{dn} = v$ である n に関する和を表す。トピック k のときの文書 d の n 番目の単語 w_{dn} について同一単語を足し合わせることで語彙 v の確率に変換している。

式 (3) の両辺を $v = 1$ から V までの和をとると、$\sum_{v=1}^{V} \phi_{kv}$ より

$$\begin{aligned}
\sum_{v=1}^{V} \phi_{kv} &= \frac{\sum_{v=1}^{V} \sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk}}{-\lambda} \\
1 &= \frac{\sum_{v=1}^{V} \sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk}}{-\lambda} \\
-\lambda &= \sum_{v=1}^{V} \sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk}
\end{aligned}$$

となる。これを式 (3) に代入すると

$$\phi_{kv} = \frac{\sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk}}{\sum_{v'=1}^{V} \sum_{d=1}^{D} \sum_{n:w_{dn}=v'} q_{dnk}} \tag{4.3}$$

が得られる。この式が ϕ_{kv} の更新式となる。

ϕ_{kv} は、全文書中における語彙 v に関する負担率の和に比例していることが分かる。

・**R** で組んでみる

・コード全体

・テキスト処理

```
# 利用パッケージ
library(RMeCab)
library(tidyverse)

### テキスト処理

# 抽出しない単語を指定
stop_words <- "[a-z]" # 小文字のアルファベットを含む語

# 形態素解析
mecab_df <- docDF(" フォルダ名", type = 1) # テキストファイルの保存先を指定する

# 文書 d の語彙 v の出現回数 (N_dv) の集合
N_dv <- mecab_df %>%
        filter(grepl(" 名詞 | 形容詞 |^ 動詞", POS1)) %>% # 抽出する品詞 (大分類) を指定する
        filter(grepl(" 一般 |^ 自立", POS2)) %>%        # 抽出する品詞 (細分類) を指定する
        filter(!grepl(stop_words, TERM)) %>%          # ストップワードを除く
        select(-c(TERM, POS1, POS2)) %>%              # 数値列のみを残す
        filter(apply(., 1, sum) >= 5) %>%             # 抽出する総出現回数を指定する
        t()                                           # 転置

# 確認用の行列名
dimnames(N_dv) <- list(paste0("d=", 1:nrow(N_dv)), # 行名
                       paste0("v=", 1:ncol(N_dv))) # 列名

# 文書 d の単語数 (N_d) のベクトル
N_d <- apply(N_dv, 1, sum) # 行方向に和をとる

# 文書数 (D)
D <- nrow(N_dv)

# 総語彙数 (V)
V <- ncol(N_dv)
```

・パラメータの初期設定

```
### パラメータの初期設定

# トピック数 (K)
K <- 4   # 任意の値を指定する

# 負担率 (q_dvk) の集合
q_dvk <- array(0, dim = c(D, V, K),
               dimnames = list(paste0("d=", 1:D),  # 確認用の次元名
                               paste0("v=", 1:V),
                               paste0("k=", 1:K)))
```

```
## トピック分布 (theta_dk) の集合
# 値をランダムに付ける
theta_dk <- sample(seq(0.1, 1, 0.01), D * K, replace = TRUE) %>%  # ランダムな値を生成
            matrix(nrow = D, ncol = K,
                   dimnames = list(paste0("d=", 1:D),  # 行名
                                   paste0("k=", 1:K))) # 列名

# 初期値の正規化
for(d in 1:D) {
  theta_dk[d, ] <- theta_dk[d, ] / sum(theta_dk[d, ])
}

## 単語分布 (phi_kv) の集合
# 値をランダムに付ける
phi_kv <- sample(seq(0, 1, 0.01), K * V, replace = TRUE) %>%  # ランダムな値を生成
          matrix(nrow = K, ncol = V,
                 dimnames = list(paste0("k=", 1:K),  # 行名
                                 paste0("V=", 1:V))) # 列名

# 初期値の正規化
for(k in 1:K) {
  phi_kv[k, ] <- phi_kv[k, ] / sum(phi_kv[k, ])
}
```

・最尤推定

```
# 推移の確認用
trace_theta <- theta_dk[1, ]
trace_phi   <- phi_kv[1, ]

for(i in 1:50) { # 任意の回数を指定する

  # パラメータを初期化
  next_theta_dk <- matrix(0, D, K,
                          dimnames = list(paste0("d=", 1:D),  # 行名
                                          paste0("k=", 1:K))) # 列名

  next_phi_kv <- matrix(0, K, V,
                        dimnames = list(paste0("k=", 1:K),  # 行名
                                        paste0("V=", 1:V))) # 列名

  for(d in 1:D) { ## (各文書)

    for(v in 1:V) { ## (各語彙 1)

      for(k in 1:K) { ## (各トピック)

        # 負担率を計算
        tmp_q_numer <- theta_dk[d, k] * (phi_kv[k, v] ^ N_dv[d, v])     # 分子
        tmp_q_denom <- sum(theta_dk[d, ] * (phi_kv[, v] ^ N_dv[d, v])) # 分母
        q_dvk[d, v, k] <- tmp_q_numer / tmp_q_denom

        # theta_dk を更新
        next_theta_dk[d, k] <- next_theta_dk[d, k] + q_dvk[d, v, k] * N_dv[d, v]
```

```
        for(v2 in 1:V) { ## （各語彙 2）

          # phi_kv を更新
          next_phi_kv[k, v2] <- next_phi_kv[k, v2] + q_dvk[d, v2, k] * N_dv[d, v2]

        } ## （/各語彙 2）
      } ## （/各トピック）
    } ## （/各語彙 1）
  } ## （/各文書）

  # パラメータを更新
  theta_dk <- next_theta_dk
  phi_kv   <- next_phi_kv

  # パラメータを正規化
  for(d in 1:D) {
    theta_dk[d, ] <- theta_dk[d, ] / sum(theta_dk[d, ])
  }

  for(k in 1:K) {
    phi_kv[k, ] <- phi_kv[k, ] / sum(phi_kv[k, ])
  }

  # 推移の確認用
  trace_theta <- rbind(trace_theta, theta_dk[1, ])
  trace_phi   <- rbind(trace_phi, phi_kv[1, ])
}
```

・コードの解説

・利用パッケージ

```
# 利用パッケージ
library(RMeCab)
library(tidyverse)
```

形態素解析のための RMeCab::docDF() と、グラフ作成時の tidyr::gather()、ggplot2 関連の関数以外にライブラリの読み込みが必要なのは dplyr のみ。(面倒なので tidyvers パッケージを読み込んでいる)

その他テキスト処理の部分は 3.3 節を参照のこと。

・パラメータの初期設定

```
# トピック数（K）
K <- 4    # 任意の値を指定する
```

任意のトピック数を指定する。

・負担率

```r
# 負担率 (q_dvk) の集合
q_dvk <- array(0, dim = c(D, V, K),
               dimnames = list(paste0("d=", 1:D),  # 確認用の次元名
                               paste0("v=", 1:V),
                               paste0("k=", 1:K)))
```

ここでは、単語レベルではなく語彙レベルで扱うため、q_{dnk} の変数名を q_dvk とする。
トピックモデルの負担率 q_{dvk} は 3 次元の変数なので、array() を使って配列型として扱う。初期値 0 を全て 0 として、引数 dim = c(D, V, K) で各次元の要素数を D, V, K 個とする。
変数に付ける行列名等は目で確認するためのもので、付けなくても処理に影響しない。他の変数についても同様である。

・トピック分布

```r
## トピック分布 (theta_dk) の集合
# 値をランダムに付ける
theta_dk <- sample(seq(0.1, 1, 0.01), D * K, replace = TRUE) %>%  # ランダムな値を生成
            matrix(nrow = D, ncol = K,
                   dimnames = list(paste0("d=", 1:D),  # 行名
                                   paste0("k=", 1:K))) # 列名

# 初期値の正規化
for(d in 1:D) {
  theta_dk[d, ] <- theta_dk[d, ] / sum(theta_dk[d, ])
}
```

まずは、ランダムな値を持つ D 行 K 列の行列を作成する。
$\sum_{k=1}^{K} \theta_{dk} = 1$ となるように正規化して初期値とする。

・単語分布

```r
## 単語分布 (phi_kv) の集合
# 値をランダムに付ける
phi_kv <- sample(seq(0, 1, 0.01), K * V, replace = TRUE) %>%  # ランダムな値を生成
          matrix(nrow = K, ncol = V,
                 dimnames = list(paste0("k=", 1:K),  # 行名
                                 paste0("V=", 1:V))) # 列名

# 初期値の正規化
for(k in 1:K) {
  phi_kv[k, ] <- phi_kv[k, ] / sum(phi_kv[k, ])
}
```

同様に、K 行 V 列の行列にランダムな値を割り振り、$\sum_{v=1}^{V} \phi_{kv} = 1$ となるように正規化して初期値とする。

・**最尤推定**

ここから for() ループ内の処理。

・パラメータの初期化

```
# パラメータを初期化
next_theta_dk <- matrix(0, nrow = D, ncol = K,
                        dimnames = list(paste0("d=", 1:D),  # 行名
                                        paste0("k=", 1:K))) # 列名

next_phi_kv <- matrix(0, nrow = K, ncol = V,
                      dimnames = list(paste0("k=", 1:K),  # 行名
                                      paste0("V=", 1:V))) # 列名
```

各パラメーターの更新時に用いる変数を作成する。
また、次ステップに入る前に初期化する。

・負担率の計算

```
# 負担率を計算
tmp_q_numer <- theta_dk[d, k] * (phi_kv[k, v] ^ N_dv[d, v])     # 分子
tmp_q_denom <- sum(theta_dk[d, ] * (phi_kv[, v] ^ N_dv[d, v])) # 分母
q_dvk[d, v, k] <- tmp_q_numer / tmp_q_denom
```

負担率は

$$q_{dnk} = \frac{\theta_{dk}\phi_{kw_{dn}}}{\sum_{k'=1}^{K}\theta_{dk'}\phi_{k'w_{dn}}} \tag{4.1}$$

の計算式に従って求める。ただし、この式は単語 w_{dn} を使っているので、これを語彙 v とするために各語彙の出現回数 N_{dv} を用いて変換する。次のトピック分布・単語分布についても同様である。

ϕ_{kv} は単語 v の出現確率のことである。N_{dv} 回出現する確率を考えると $\phi_{kv}^{N_{dv}}$ となる。従って、計算式は

$$q_{dvk} = \frac{\theta_{dk}\phi_{kv}^{N_{dv}}}{\sum_{k'=1}^{K}\theta_{dk'}\phi_{k'v}^{N_{dv}}}$$

になる。

・トピック分布の更新

```
# theta_dk を更新
next_theta_dk[d, k] <- next_theta_dk[d, k] + q_dvk[d, v, k] * N_dv[d, v]
```

トピック分布の更新式 (4.2) を語彙 v に変換するには、負担率を N_{dv} 回足し合わせればよい。すると、計算式は

$$\theta_{dk} = \frac{\sum_{v=1}^{V} q_{dvk}N_{dv}}{\sum_{k'=1}^{K}\sum_{v=1}^{V} q_{dvk'}N_{dv}}$$

になる。この式の $\sum_{v=1}^{V}$ の計算は `for(v in 1:V)` の処理によって、初期値 0 の `next_theta_dk` に V 回繰り返し `q_dvk` を足すことで行われる。また、分母の $\sum_{k'=1}^{K}$ については最後の正規化処理によって行う。

・単語分布の更新

```
for(v2 in 1:V) {
  # phi_kvを更新
  next_phi_kv[k, v2] <- next_phi_kv[k, v2] + q_dvk[d, v2, k] * N_dv[d, v2]
}
```

同様に、単語分布の更新式 (4.3) を語彙 v に変換すると、更新式は

$$\phi_{kv} = \frac{\sum_{d=1}^{D} q_{dvk} N_{dv}}{\sum_{v'=1}^{V} \sum_{d=1}^{D} q_{dv'k} N_{dv'}}$$

になる。初期値 0 の `next_phi_kv` に D 回 `q_{dvk}` を足すことで分子の計算を行う。分母の計算については次の正規化処理で行う。

・パラメータの正規化

```
# パラメータを更新
theta_dk <- next_theta_dk
phi_kv   <- next_phi_kv

# パラメータを正規化
for(d2 in 1:nrow(theta_dk)) {
  theta_dk[d2, ] <- theta_dk[d2, ] / sum(theta_dk[d2, ])
}

for(k2 in 1:nrow(phi_kv)) {
  phi_kv[k2, ] <- phi_kv[k2, ] / sum(phi_kv[k2, ])
}
```

各パラメータを次ステップの計算に用いるために移す。

更新された変数をそれぞれ $\sum_{k=1}^{K} \theta_{dk} = \sum_{v=1}^{V} \phi_{kv} = 1$ の条件に満たすように正規化する。

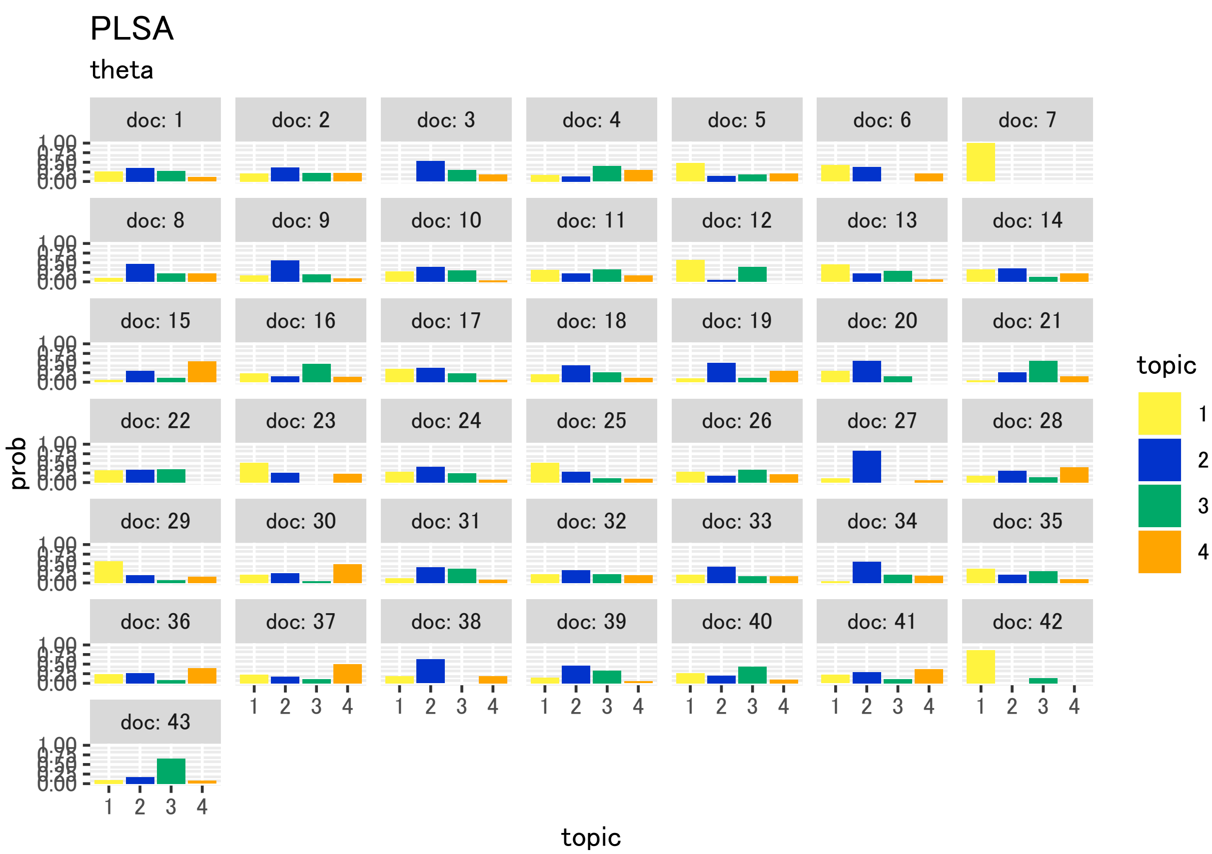
この推定処理を指定した回数繰り返す。(本当は収束条件を満たすまで繰り返すようにしたい…)

・推定結果の確認

・パラメータの推定結果の確認

・トピック分布

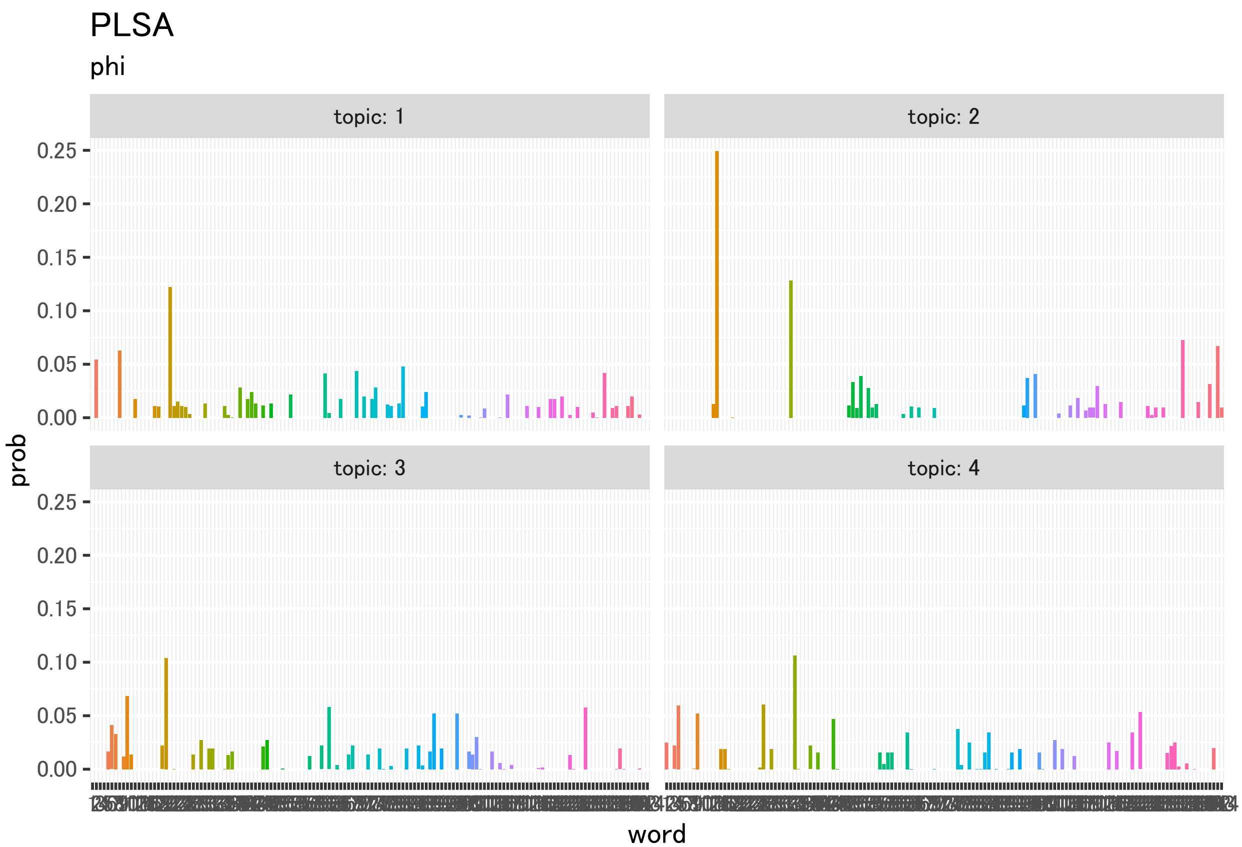
```r
## トピック分布
# データフレームを作成
theta_df_wide <- cbind(as.data.frame(theta_dk),
                       as.factor(1:D)) # 文書番号

# データフレームを long 型に変換
colnames(theta_df_wide) <- c(1:K, "doc") # key 用の行名を付与
theta_df_long <- gather(theta_df_wide, key = "topic", value = "prob", -doc) # 変換

# 描画
ggplot(data = theta_df_long, mapping = aes(x = topic, y = prob, fill = topic)) + # データ
  geom_bar(stat = "identity", position = "dodge") +       # 棒グラフ
  facet_wrap( ~ doc, labeller = label_both) +             # グラフの分割
  labs(title = "PLSA", subtitle = "theta")                # タイトル
```



・単語分布

```r
## 単語分布
# データフレームを作成
phi_df_wide <- cbind(as.data.frame(t(phi_kv)),
                     as.factor(1:V)) # 語彙番号

# データフレームを long 型に変換
colnames(phi_df_wide) <- c(1:K, "word") # key 用の行名を付与
phi_df_long <- gather(phi_df_wide, key = "topic", value = "prob", -word) # 変換
```

```
# 描画
ggplot(data = phi_df_long, mapping = aes(x = word, y = prob, fill = word)) + # データ
  geom_bar(stat = "identity", position = "dodge") +       # 棒グラフ
  facet_wrap( ~ topic, labeller = label_both) +           # グラフの分割
  theme(legend.position = "none") +                       # 凡例
  labs(title = "PLSA", subtitle = "phi")                  # タイトル
```



・推移の確認

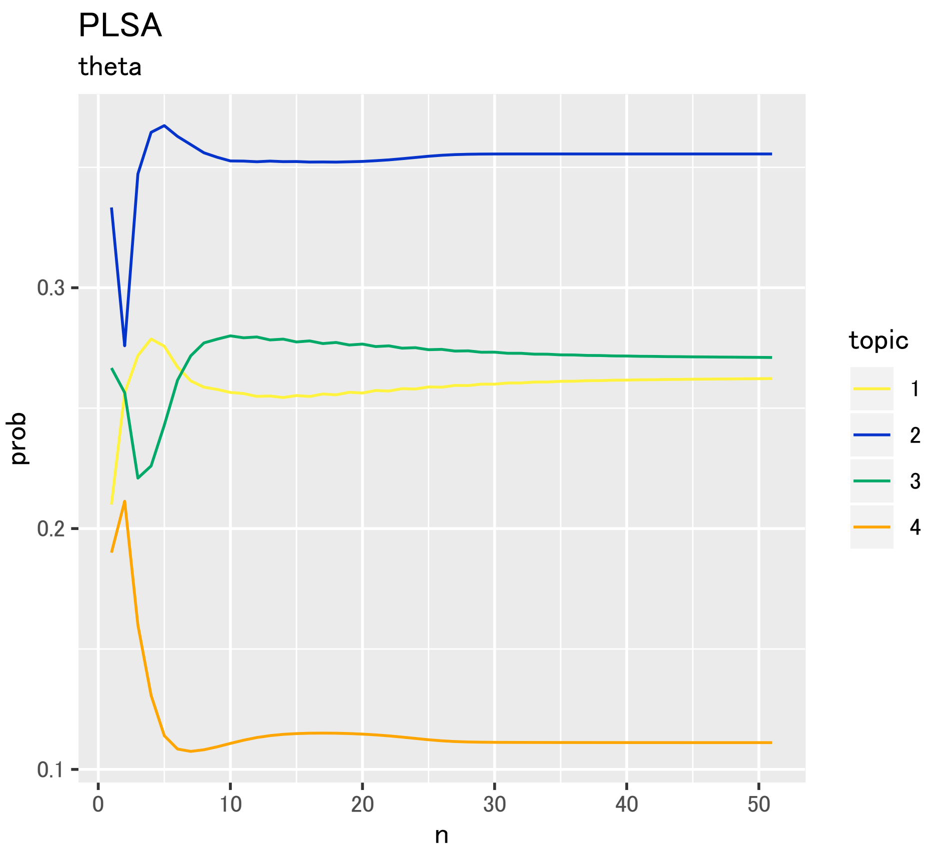
・推移の記録

```
# 推移の確認用（初期値）
trace_theta <- theta_dk[1, ]
trace_phi   <- phi_kv[1, ]


# 推移の確認用（更新後）
trace_theta <- rbind(trace_theta, theta_dk[1, ])
trace_phi   <- rbind(trace_phi, phi_kv[1, ])
```

推移を確認するために、パラメータを更新する度に記録しておく。トピック分布については文書 1、単語分布についてはトピック 1 のみ確認する。

・トピック分布

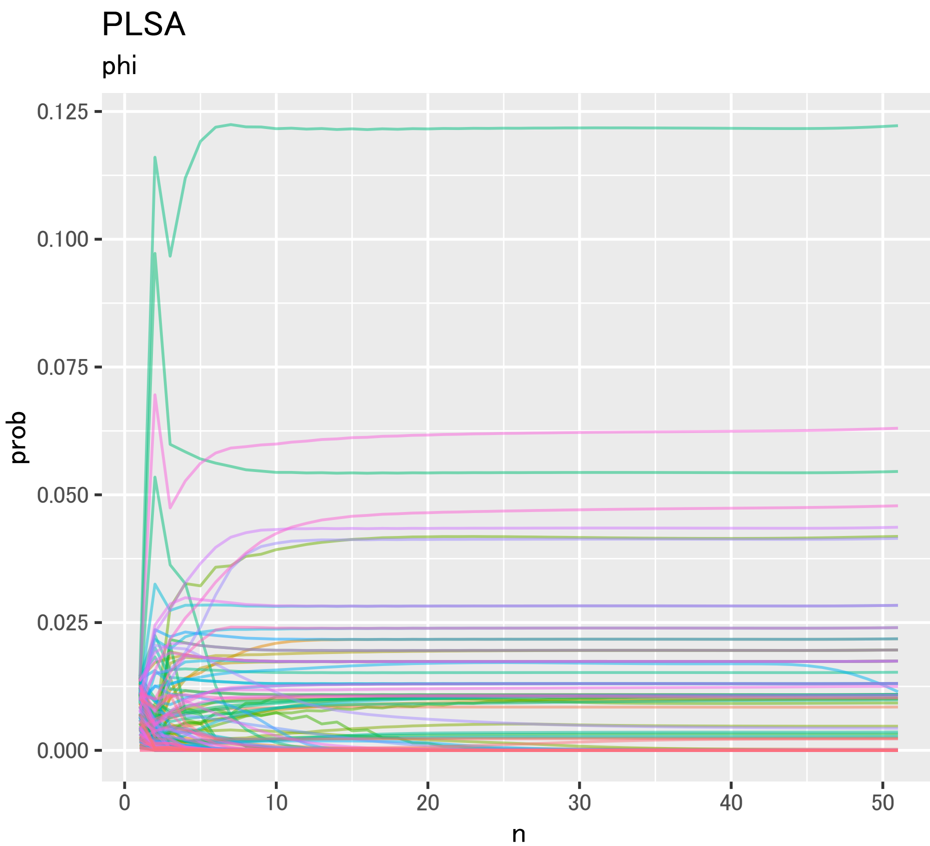
```r
## トピック分布
# データフレームを作成
theta_df_wide <- cbind(as.data.frame(trace_theta),
                       1:nrow(trace_theta)) # 試行回数

# データフレームを long 型に変換
colnames(theta_df_wide) <- c(1:K, "n") # key 用の行名を付与
theta_df_long <- gather(theta_df_wide, key = "topic", value = "prob", -n) # 変換

# 描画
ggplot(data = theta_df_long, mapping = aes(x = n, y = prob, color = topic)) + # データ
  geom_line() +                                        # 折れ線グラフ
  labs(title = "PLSA", subtitle = "theta") # タイトル
```



・単語分布

```r
## 単語分布
# データフレームを作成
phi_df_wide <- cbind(as.data.frame(trace_phi),
                     1:nrow(trace_phi)) # 試行回数

# データフレームを long 型に変換
colnames(phi_df_wide) <- c(1:V, "n") # key 用の行名を付与
phi_df_long <- gather(phi_df_wide, key = "word", value = "prob", -n) # 変換

# 描画
```

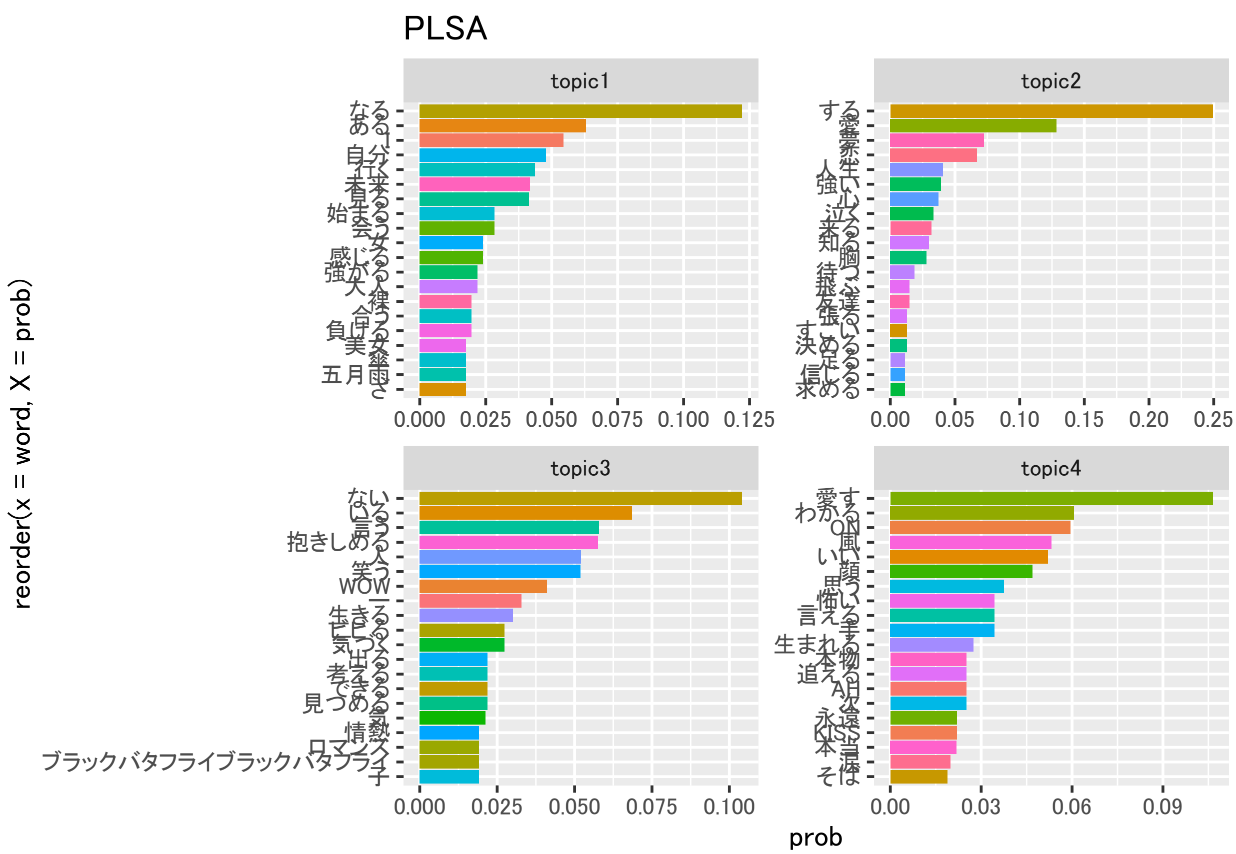
```r
ggplot(data = phi_df_long, mapping = aes(x = n, y = prob, color = word)) + # データ
  geom_line(alpha = 0.5) +                # 折れ線グラフ
  theme(legend.position = "none") +       # 凡例
  labs(title = "PLSA", subtitle = "phi") # タイトル
```



・各トピックの出現確率の上位語の確認

```r
### 各トピックの出現確率の上位語

## 語彙インデックス (v)
# 指定した出現回数以上の単語の行番号
num <- mecab_df %>%
       select(-c(TERM, POS1, POS2)) %>%
       apply(1, sum) >= 5 # 抽出する総出現回数を指定する

v_index <- mecab_df[num, ] %>%
           filter(grepl("名詞|形容詞|^動詞", POS1)) %>%  # 抽出する品詞を指定する
           filter(grepl("一般|^自立", POS2)) %>%
           filter(!grepl(stop_words, TERM)) %>%
           .[, 1]  # 単語の列を抽出する

# データフレームを作成
phi_df_wide2 <- cbind(as.data.frame(t(phi_kv)),
                      v_index) %>%
                as.data.frame()
```

```r
colnames(phi_df_wide2) <- c(paste0("topic", 1:K), "word") # key 用の行名を付与

# データフレームの整形
phi_df_long2 <- data.frame()
for(i in 1:K) {
  tmp_df <- phi_df_wide2 %>%
    select(paste0("topic", i), word) %>%
    arrange(-.[, 1]) %>%  # 降順に並べ替え
    head(20) %>%          # 任意で指定した上位 n 語を抽出
    gather(key = "topic", value = "prob", -word) # long 型に変換

  phi_df_long2 <- rbind(phi_df_long2, tmp_df)
}

# 描画
ggplot(data = phi_df_long2,
       mapping = aes(x = reorder(x = word, X = prob), y = prob, fill = word)) +  # データ
  geom_bar(stat = "identity", position = "dodge") + # 棒グラフ
  coord_flip() +                                    # グラフの向き
  facet_wrap( ~ topic, scales = "free") +           # グラフの分割
  theme(legend.position = "none") +                 # 凡例
  labs(title = "PLSA")                              # タイトル
```



4.4 変分ベイズ推定

4.4.1 周辺尤度の最大化

変分ベイズ推定では、未知変数 $\boldsymbol{Z}, \boldsymbol{\Theta}, \boldsymbol{\Phi}$ の事後分布を推定する。事後分布の近似である変分事後分布 $q(\boldsymbol{Z}, \boldsymbol{\Theta}, \boldsymbol{\Phi})$ を求める。

まずは、トピックモデルの対数周辺尤度の変分下限 F をイェンゼンの不等式を用いて求める。

$$
\begin{aligned}
\log p(\boldsymbol{W}|\alpha,\beta) &= \log \int\int \sum_{\boldsymbol{Z}} p(\boldsymbol{W},\boldsymbol{Z},\boldsymbol{\Theta},\boldsymbol{\Phi}|\alpha,\beta) d\boldsymbol{\Theta} d\boldsymbol{\Phi} && \text{(i)}\\
&= \log \int\int \sum_{\boldsymbol{Z}} p(\boldsymbol{Z}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\alpha)p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\beta) d\boldsymbol{\Theta} d\boldsymbol{\Phi} && \text{(ii)}\\
&= \log \int\int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z})q(\boldsymbol{\Theta},\boldsymbol{\Phi}) \frac{p(\boldsymbol{Z}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\alpha)p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\beta)}{q(\boldsymbol{Z})q(\boldsymbol{\Theta},\boldsymbol{\Phi})} d\boldsymbol{\Theta} d\boldsymbol{\Phi} && \text{(iii)}\\
&\geq \int\int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z})q(\boldsymbol{\Theta},\boldsymbol{\Phi}) \log \frac{p(\boldsymbol{Z}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\alpha)p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\beta)}{q(\boldsymbol{Z})q(\boldsymbol{\Theta},\boldsymbol{\Phi})} d\boldsymbol{\Theta} d\boldsymbol{\Phi} \equiv F
\end{aligned}
$$

【途中式の途中式】

- (i)：4.2 節の生成過程より分解する。
- (ii)：
  - 計算を容易にするため変分事後分布は、$q(\boldsymbol{Z},\boldsymbol{\Theta},\boldsymbol{\Phi}) = q(\boldsymbol{Z})q(\boldsymbol{\Theta},\boldsymbol{\Phi})$ と分解できると仮定して分解する。
  - 求めたい変分事後分布を $\frac{q(\boldsymbol{Z})q(\boldsymbol{\Theta},\boldsymbol{\Phi})}{q(\boldsymbol{Z})q(\boldsymbol{\Theta},\boldsymbol{\Phi})} = 1$ として、分割して掛け合わせる。
- (iii)：イェンゼンの不等式を用いて、下限を求める。

この式を整理する。

$$
\begin{aligned}
F &= \int\int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z})q(\boldsymbol{\Theta},\boldsymbol{\Phi}) \Big(\log p(\boldsymbol{Z}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\alpha)p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\beta) \\
&\qquad - \log q(\boldsymbol{Z})q(\boldsymbol{\Theta},\boldsymbol{\Phi}) \Big) d\boldsymbol{\Theta} d\boldsymbol{\Phi} \qquad (4.4)\\
&= \int\int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z})q(\boldsymbol{\Theta})q(\boldsymbol{\Phi}) \Big(\log p(\boldsymbol{Z}|\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}|\alpha) + \log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi}) + \log p(\boldsymbol{\Phi}|\beta) \\
&\qquad - \log q(\boldsymbol{Z}) - \log q(\boldsymbol{\Theta}) - \log q(\boldsymbol{\Phi}) \Big) d\boldsymbol{\Theta} d\boldsymbol{\Phi}
\end{aligned}
$$

次節では、この変分下限 F を最大にする各分布の変分事後分布を求めていく。

4.4.2 変分事後分布の推定

・トピック分布の変分事後分布

トピック分布の変分事後分布 $q(\boldsymbol{\Theta})$ を推定する。ラグランジュの未定乗数法を用いて、$\int q(\boldsymbol{\Theta})d\boldsymbol{\Theta} = 1$ の制約の下で最大となる $q(\boldsymbol{\Theta})$ を求めていく。

$$\begin{aligned}F(q(\boldsymbol{\Theta})) = \int\int\sum_{\boldsymbol{Z}} q(\boldsymbol{Z})q(\boldsymbol{\Theta})q(\boldsymbol{\Phi})\Big(&\log p(\boldsymbol{Z}|\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}|\alpha) + \log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi}) + \log p(\boldsymbol{\Phi}|\beta)\\ &- \log q(\boldsymbol{Z}) - \log q(\boldsymbol{\Theta}) - \log q(\boldsymbol{\Phi})\Big)d\boldsymbol{\Theta}d\boldsymbol{\Phi} + \lambda\left(\int q(\boldsymbol{\Theta})d\boldsymbol{\Theta} - 1\right)\end{aligned}$$

$F(q(\boldsymbol{\Theta}))$ を $q(\boldsymbol{\Theta})$ に関して変分する。

$$\begin{aligned}\frac{\partial F(q(\boldsymbol{\Theta}))}{\partial q(\boldsymbol{\Theta})} &= \int\sum_{\boldsymbol{Z}} q(\boldsymbol{Z})q(\boldsymbol{\Phi})\Big(\log p(\boldsymbol{Z}|\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}|\alpha) + \log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi}) + \log p(\boldsymbol{\Phi}|\beta)\\ &\qquad - \log q(\boldsymbol{Z}) - \log q(\boldsymbol{\Theta}) - \log q(\boldsymbol{\Phi}) - 1\Big)d\boldsymbol{\Phi} + \lambda \qquad \text{(i)}\\ &= \sum_{\boldsymbol{Z}} q(\boldsymbol{Z})\log p(\boldsymbol{Z}|\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}|\alpha) + \int\sum_{\boldsymbol{Z}} q(\boldsymbol{Z})q(\boldsymbol{\Phi})\log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})d\boldsymbol{\Phi} + \int q(\boldsymbol{\Phi})\log p(\boldsymbol{\Phi}|\beta)d\boldsymbol{\Phi}\\ &\qquad - \sum_{\boldsymbol{Z}} q(\boldsymbol{Z})\log q(\boldsymbol{Z}) - \log q(\boldsymbol{\Theta}) - q(\boldsymbol{\Phi})\log q(\boldsymbol{\Phi}) - 1 + \lambda \qquad \text{(ii)}\\ &= \mathbb{E}_{q(\boldsymbol{Z})}[\log p(\boldsymbol{Z}|\boldsymbol{\Theta})] + \log p(\boldsymbol{\Theta}|\alpha) - \log q(\boldsymbol{\Theta}) + \lambda - C\end{aligned}$$

【途中式の途中式】

- (i)：式を展開する。ただし、$\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) = \int q(\boldsymbol{\Phi})d\boldsymbol{\Phi} = 1$ である。
- (ii)：$q(\boldsymbol{\Theta})$ と関係のない項を C と置く。

$\frac{\partial F(q(\boldsymbol{\Theta}))}{\partial q(\boldsymbol{\Theta})} = 0$ となる $q(\boldsymbol{\Theta})$ を解く。$q_{dnk} \equiv q(z_{dn} = k)$ を文書 d の n 番目の単語のトピック z_{dn} が k となる変分事後確率とする。

$$\begin{aligned}\frac{\partial F(q(\boldsymbol{\Theta}))}{\partial q(\boldsymbol{\Theta})} =& \mathbb{E}_{q(\boldsymbol{Z})}[\log p(\boldsymbol{Z}|\boldsymbol{\Theta})] + \log p(\boldsymbol{\Theta}|\alpha) - \log q(\boldsymbol{\Theta}) + \lambda - C = 0 && \text{(i)}\\ \log q(\boldsymbol{\Theta}) =& \mathbb{E}_{q(\boldsymbol{Z})}[\log p(\boldsymbol{Z}|\boldsymbol{\Theta})] + \log p(\boldsymbol{\Theta}|\alpha) + \lambda - C && \text{(ii)}\\ q(\boldsymbol{\Theta}) =& \exp\left(\mathbb{E}_{q(\boldsymbol{Z})}[\log p(\boldsymbol{Z}|\boldsymbol{\Theta})] + \log p(\boldsymbol{\Theta}|\alpha) + \lambda - C\right) && \text{(iii)}\\ =& \exp\left\{\sum_{d=1}^D\sum_{n=1}^{N_d}\sum_{k=1}^K q_{dnk}\log\theta_{dk} + \log\left(\frac{\Gamma(\alpha D)^K}{\Gamma(\alpha)^{DK}}\prod_{d=1}^D\prod_{k=1}^K\theta_{dk}^{\alpha-1}\right) + \lambda - C\right\} && \text{(iv)}\\ =& \exp\left(\sum_{d=1}^D\sum_{k=1}^K\log\theta_{dk}^{\sum_{n=1}^{N_d} q_{dnk}} + \log\frac{\Gamma(\alpha D)^K}{\Gamma(\alpha)^{DK}} + \sum_{d=1}^D\sum_{k=1}^K\log\theta_{dk}^{\alpha-1} + \lambda - C\right) && \text{(v)}\\ \propto& \exp\left(\sum_{d=1}^D\sum_{k=1}^K\log\theta_{dk}^{\sum_{n=1}^{N_d} q_{dnk}+\alpha-1}\right) && \text{(vi)}\\ =& \prod_{d=1}^D\prod_{k=1}^K\theta_{dk}^{\alpha+\sum_{n=1}^{N_d} q_{dnk}-1} && \text{(vii)}\\ =& \prod_{d=1}^D\mathrm{Dirichlet}(\boldsymbol{\theta}_d|\alpha_{d1},\cdots,\alpha_{dK}) && (1)\end{aligned}$$

【途中式の途中式】

- (i)：式を整理する。
- (ii)：両辺で指数をとる。
- (iii)：
  - $q(\boldsymbol{Z}) = \prod_{d=1}^{D} \prod_{n=1}^{N_d} q(z_{dn})$ で、$q(z_{dn} = k) \equiv q_{dnk}$ より、置き換える。
  - $p(\boldsymbol{Z}|\boldsymbol{\Theta}) = \prod_{d=1}^{D} \prod_{n=1}^{N_d} p(z_{dn} = k|\boldsymbol{\theta}_d) = \prod_{d=1}^{D} \prod_{k=1}^{K} \theta_{dk}$ より、置き換える。
  - $p(\boldsymbol{\Theta}|\alpha) = \prod_{d=1}^{D} \mathrm{Dirichlet}(\boldsymbol{\theta_d}|\alpha)$ であるため、1.2.4 節より置き換える。
- (iv)：対数をとる。
- (v)：θ_{dk} の項をまとめる。また、$q(\boldsymbol{\Theta})$ に影響しない項を除く。
- (vi)：指数をとる。
- (vii)：$\alpha_{dk} = \alpha + \sum_{n=1}^{N_d} q_{dnk}$ としたとき、パラメータ $(\alpha_{d1}, \cdots, \alpha_{dK})$ を持つ正規化項のないディリクレ分布の形であることから。

トピック分布の変分事後分布 $q(\boldsymbol{\Theta})$ がディリクレ分布であることが分かった。また

$$\alpha_{dk} = \alpha + \sum_{n=1}^{N_d} q_{dnk}$$

がハイパーパラメータ α_{dk} の更新式となる。

・単語分布の変分事後分布

続いて、単語分布の変分事後分布 $q(\boldsymbol{\Phi})$ を推定する。トピック分布の変分事後分布の推定と同様に、ラグランジュの未定乗数法を用いて、$\int q(\boldsymbol{\Phi}) d\boldsymbol{\Phi} = 1$ の制約の下で最大となる $q(\boldsymbol{\Phi})$ を求めていく。

$$\begin{aligned}
F(q(\boldsymbol{\Phi})) = \int\int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) q(\boldsymbol{\Theta}) q(\boldsymbol{\Phi}) \Big(& \log p(\boldsymbol{Z}|\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}|\alpha) + \log p(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi}) + \log p(\boldsymbol{\Phi}|\beta) \\
& - \log q(\boldsymbol{Z}) - \log q(\boldsymbol{\Theta}) - \log q(\boldsymbol{\Phi}) \Big) d\boldsymbol{\Theta} d\boldsymbol{\Phi} + \lambda \left(\int q(\boldsymbol{\Phi}) d\boldsymbol{\Phi} - 1 \right)
\end{aligned}$$

$F(q(\boldsymbol{\Phi}))$ を $q(\boldsymbol{\Phi})$ に関して変分する。

$$\begin{aligned}
\frac{\partial F(q(\boldsymbol{\Phi}))}{\partial q(\boldsymbol{\Phi})} &= \int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) q(\boldsymbol{\Theta}) \Big(\log p(\boldsymbol{Z}|\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}|\alpha) + \log p(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi}) + \log p(\boldsymbol{\Phi}|\beta) \\
&\qquad - \log q(\boldsymbol{Z}) - \log q(\boldsymbol{\Theta}) - \log q(\boldsymbol{\Phi}) - 1 \Big) d\boldsymbol{\Theta} + \lambda \qquad \text{(i)} \\
&= \int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) q(\boldsymbol{\Theta}) \log p(\boldsymbol{Z}|\boldsymbol{\Theta}) d\boldsymbol{\Theta} + \int q(\boldsymbol{\Theta}) \log p(\boldsymbol{\Theta}|\alpha) d\boldsymbol{\Theta} + \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \log p(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi}) + \log p(\boldsymbol{\Phi}|\beta) \\
&\qquad - \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \log q(\boldsymbol{Z}) - \int q(\boldsymbol{\Theta}) \log q(\boldsymbol{\Theta}) d\boldsymbol{\Theta} - \log q(\boldsymbol{\Phi}) - 1 + \lambda \qquad \text{(ii)} \\
&= \mathbb{E}_{q(\boldsymbol{Z})}[\log p(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi})] + \log p(\boldsymbol{\Phi}|\beta) - \log q(\boldsymbol{\Phi}) + \lambda - C
\end{aligned}$$

【途中式の途中式】

- (i)：式を展開する。ただし、$\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) = \int q(\boldsymbol{\Theta}) d\boldsymbol{\Theta} = 1$ である。
- (ii)：$q(\boldsymbol{\Phi})$ と関係のない項を C と置く。

$\frac{\partial F(q(\boldsymbol{\Phi}))}{\partial q(\boldsymbol{\Phi})} = 0$ となる $q(\boldsymbol{\Phi})$ を解く。

$$
\begin{aligned}
\frac{\partial F(q(\boldsymbol{\Phi}))}{\partial q(\boldsymbol{\Phi})} &= \mathbb{E}_{q(\boldsymbol{Z})}[\log p(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi})] + \log p(\boldsymbol{\Phi}|\beta) - \log q(\boldsymbol{\Phi}) + \lambda - C = 0 && \text{(i)} \\
\log q(\boldsymbol{\Phi}) &= \mathbb{E}_{q(\boldsymbol{Z})}[\log p(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi})] + \log p(\boldsymbol{\Phi}|\beta) + \lambda - C && \text{(ii)} \\
q(\boldsymbol{\Phi}) &= \exp\big(\mathbb{E}_{q(\boldsymbol{Z})}[\log p(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi})] + \log p(\boldsymbol{\Phi}|\beta) + \lambda - C\big) && \text{(iii)} \\
&= \exp\left\{\sum_{d=1}^{D} \sum_{n=1}^{N_d} \sum_{k=1}^{K} q_{dnk} \log \prod_{v=1}^{V} \phi_{kv}^{N_{dv}} + \log\left(\frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{KV}} \prod_{k=1}^{K} \prod_{v=1}^{V} \phi_{kv}^{\beta-1}\right) + \lambda - C\right\} && \text{(iv)} \\
&= \exp\left(\sum_{k=1}^{K} \sum_{v=1}^{V} \log \phi_{kv}^{\sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk}} + \log \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{KV}} + \sum_{k=1}^{K} \sum_{v=1}^{V} \log \phi_{kv}^{\beta-1} + \lambda - C\right) && \text{(v)} \\
&\propto \exp\left(\sum_{k=1}^{K} \sum_{v=1}^{V} \log \phi_{kv}^{\sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk} + \beta - 1}\right) && \text{(vi)} \\
&= \prod_{k=1}^{K} \prod_{v=1}^{V} \phi_{kv}^{\beta + \sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk} - 1} && \text{(vii)} \\
&= \prod_{k=1}^{K} \mathrm{Dirichlet}(\boldsymbol{\phi}_k | \beta_{k1}, \cdots, \beta_{kV}) && (2)
\end{aligned}
$$

【途中式の途中式】

- (i)：式を整理する。
- (ii)：両辺で指数をとる。
- (iii)：
  - $q(\boldsymbol{Z}) = \prod_{d=1}^{D} \prod_{n=1}^{N_d} q(z_{dn})$ で、$q(z_{dn} = k) \equiv q_{dnk}$ より、置き換える。
  - $p(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi}) = \prod_{d=1}^{D} \prod_{n=1}^{N_d} q(w_{dn}|z_{dn} = k, \boldsymbol{\phi}_d) = \prod_{d=1}^{D} \prod_{k=1}^{K} \prod_{v=1}^{V} \phi_{kv}^{N_{dv}}$ より、置き換える。
  - $p(\boldsymbol{\Phi}|\beta) = \prod_{k=1}^{K} \mathrm{Dirichlet}(\boldsymbol{\phi}_k|\beta)$ であるため、1.2.4 節より置き換える。
- (iv)：対数をとる。
- (v)：ϕ_{kv} の項をまとめる。また、$q(\boldsymbol{\Phi})$ に影響しない項を除く。
- (vi)：指数をとる。
- (vii)：$\beta_{kv} = \beta + \sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk}$ としたとき、パラメータ $(\beta_{k1}, \cdots, \beta_{kV})$ を持つ正規化項のないディリクレ分布の形であることから。

単語分布の変分事後分布 $q(\boldsymbol{\Phi})$ がディリクレ分布であることが分かった。ここで $\sum_{n:w_{dn}=v}$ は、$w_{dn} = v$ である n に関する和を表す。文書 d に含まれている $w_{dn} = v$ である同一単語の数は、語彙 v の出現回数 N_{dv} と同じである。また

$$
\beta_{kv} = \beta + \sum_{d=1}^{D} \sum_{n:w_{dn}=v} q_{dnk}
$$

がハイパーパラメータ β_{kv} の更新式となる。

・トピックの変分事後分布

最後に、トピックの変分事後分布 $q(\boldsymbol{Z})$ を推定する。トピックの変分事後分布についても同様に、ラグランジュの未定乗数法を用いて、$\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) = 1$ の制約の下で最大となる $q(\boldsymbol{Z})$ を求めていく。

$$F(q(\boldsymbol{Z})) = \int\int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z})q(\boldsymbol{\Theta})q(\boldsymbol{\Phi})\Big(\log p(\boldsymbol{Z}|\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}|\alpha) + \log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi}) + \log p(\boldsymbol{\Phi}|\beta)$$
$$- \log q(\boldsymbol{Z}) - \log q(\boldsymbol{\Theta}) - \log q(\boldsymbol{\Phi})\Big) d\boldsymbol{\Theta} d\boldsymbol{\Phi} + \lambda\left(\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) - 1\right)$$

$F(q(\boldsymbol{Z}))$ を $q(\boldsymbol{Z})$ に関して変分する。

$$\begin{aligned}
\frac{F(q(\boldsymbol{Z}))}{q(\boldsymbol{Z})} &= \int\int q(\boldsymbol{\Theta})q(\boldsymbol{\Phi})\Big(\log p(\boldsymbol{Z}|\boldsymbol{\Theta}) + \log p(\boldsymbol{\Theta}|\alpha) + \log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi}) + \log p(\boldsymbol{\Phi}|\beta) \\
&\qquad - \log q(\boldsymbol{Z}) - \log q(\boldsymbol{\Theta}) - \log q(\boldsymbol{\Phi}) - 1\Big) d\boldsymbol{\Theta} d\boldsymbol{\Phi} + \lambda - 1 + \lambda \qquad \text{(i)} \\
&= \int q(\boldsymbol{\Theta}) \log p(\boldsymbol{Z}|\boldsymbol{\Theta}) d\boldsymbol{\Theta} + \int q(\boldsymbol{\Theta}) \log p(\boldsymbol{\Theta}|\alpha) d\boldsymbol{\Theta} + \int q(\boldsymbol{\Phi}) \log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi}) d\boldsymbol{\Phi} + \int q(\boldsymbol{\Phi}) \log p(\boldsymbol{\Phi}|\beta) d\boldsymbol{\Phi} \\
&\qquad - \log q(\boldsymbol{Z}) - \int q(\boldsymbol{\Theta}) \log q(\boldsymbol{\Theta}) d\boldsymbol{\Theta} - \int q(\boldsymbol{\Phi}) \log q(\boldsymbol{\Phi}) d\boldsymbol{\Phi} - 1 + \lambda \qquad \text{(ii)} \\
&= \mathbb{E}_{q(\boldsymbol{\Theta})}[\log p(\boldsymbol{Z}|\boldsymbol{\Theta})] + \mathbb{E}_{q(\boldsymbol{\Phi})}[\log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})] - \log q(\boldsymbol{Z}) + \lambda - C
\end{aligned}$$

【途中式の途中式】

- (i)：式を展開する。ただし、$\int q(\boldsymbol{\Theta})d\boldsymbol{\Theta} = \int q(\boldsymbol{\Phi})d\boldsymbol{\Phi} = 1$ である。
- (ii)：$q(\boldsymbol{Z})$ と関係のない項を C と置く。

$\frac{\partial F(q(\boldsymbol{Z}))}{\partial q(\boldsymbol{Z})} = 0$ となる $q(\boldsymbol{Z})$ を解く。

$$\begin{aligned}
\frac{\partial F(q(\boldsymbol{Z}))}{\partial q(\boldsymbol{Z})} &= \mathbb{E}_{q(\boldsymbol{\Theta})}[\log p(\boldsymbol{Z}|\boldsymbol{\Theta})] + \mathbb{E}_{q(\boldsymbol{\Phi})}[\log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})] - \log q(\boldsymbol{Z}) + \lambda - C = 0 \qquad \text{(i)} \\
\log q(\boldsymbol{Z}) &= \mathbb{E}_{q(\boldsymbol{\Theta})}[\log p(\boldsymbol{Z}|\boldsymbol{\Theta})] + \mathbb{E}_{q(\boldsymbol{\Phi})}[\log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})] + \lambda - C \qquad \text{(ii)} \\
q(\boldsymbol{Z}) &= \exp\left(\mathbb{E}_{q(\boldsymbol{\Theta})}[\log p(\boldsymbol{Z}|\boldsymbol{\Theta})] + \mathbb{E}_{q(\boldsymbol{\Phi})}[\log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})] + \lambda - C\right) \qquad \text{(iii)} \\
&= \exp\Bigg(\int \sum_{d=1}^{D} q(\boldsymbol{\theta}_d|\alpha_{d1},\cdots,\alpha_{dK}) \sum_{k=1}^{K} \log\theta_{dk} d\boldsymbol{\theta}_d \\
&\qquad + \int \sum_{k=1}^{K} q(\boldsymbol{\phi}_k|\beta_{k1},\cdots,\beta_{kV}) \sum_{v=1}^{V} N_{dv} \log\phi_{kv} d\boldsymbol{\phi}_k + \lambda - C\Bigg) \qquad \text{(iv)} \\
\prod_{d=1}^{D}\prod_{k=1}^{K} q_{dnk} &\propto \prod_{d=1}^{D}\prod_{k=1}^{K} \exp\left\{\Psi(\alpha_{dk}) - \Psi\Big(\sum_{k'=1}^{K}\alpha_{dk'}\Big) + N_{dv}\left(\Psi(\beta_{kv}) - \Psi\Big(\sum_{v=1}^{V}\beta_{kv}\Big)\right)\right\} \\
q_{dnk} &\propto \exp\left\{\Psi(\alpha_{dk}) - \Psi\Big(\sum_{k'=1}^{K}\alpha_{dk'}\Big) + N_{dv}\left(\Psi(\beta_{kv}) - \Psi\Big(\sum_{v=1}^{V}\beta_{kv}\Big)\right)\right\} \qquad \text{(4.8')}
\end{aligned}$$

【途中式の途中式】

- (i)：式を整理する。
- (ii)：両辺で対数をとる。
- (iii)：
  - 式 (1) より、$q(\mathbf{\Theta}) = \prod_{d=1}^D q(\boldsymbol{\theta}_d|\alpha_{d1}, \cdots, \alpha_{dK})$ で置き換える。
  - $p(\boldsymbol{Z}|\mathbf{\Theta}) = \prod_{d=1}^D \prod_{n=1}^{N_d} p(z_{dn} = k|\boldsymbol{\theta}_d) = \prod_{d=1}^D \prod_{k=1}^K \theta_{dk}$ より、置き換える。
  - 式 (2) より、$q(\mathbf{\Phi}) = \prod_{k=1}^K q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV})$ で置き換える。
  - $p(\boldsymbol{W}|\boldsymbol{Z}, \mathbf{\Phi}) = \prod_{d=1}^D \prod_{n=1}^{N_d} q(\boldsymbol{w}_d|z_{dn} = k, \boldsymbol{\phi}_d) = \prod_{d=1}^D \prod_{k=1}^K \prod_{v=1}^V \phi_{kv}^{N_{dv}}$ より、置き換える。
- (iv)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

この式が、トピックの変分事後分布 q_{dnk} の計算式となる。

4.4.3 変分下限

変分下限 F は次のように分解できる。

$$
\begin{aligned}
F = &\overbrace{\int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) q(\mathbf{\Theta}) \log p(\boldsymbol{Z}|\mathbf{\Theta}) d\mathbf{\Theta}}^{\text{1st term}} + \overbrace{\int q(\mathbf{\Theta}) \log p(\mathbf{\Theta}|\alpha) d\mathbf{\Theta}}^{\text{2nd term}} + \overbrace{\int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) q(\mathbf{\Phi}) \log p(\boldsymbol{W}|\boldsymbol{Z}, \mathbf{\Phi}) d\mathbf{\Phi}}^{\text{3rd term}} \\
&+ \overbrace{\int q(\mathbf{\Phi}) \log p(\mathbf{\Phi}|\beta) d\mathbf{\Phi}}^{\text{4th term}} - \overbrace{\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \log q(\boldsymbol{Z})}^{\text{5th term}} - \overbrace{\int q(\mathbf{\Theta}) \log q(\mathbf{\Theta}) d\mathbf{\Theta}}^{\text{6th term}} - \overbrace{\int q(\mathbf{\Phi}) \log q(\mathbf{\Phi}) d\mathbf{\Phi}}^{\text{7th term}}
\end{aligned}
$$

それぞれの項は以下のように計算できる。

・第 1 項

$$
\begin{aligned}
&\int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) q(\mathbf{\Theta}) \log p(\boldsymbol{Z}|\mathbf{\Theta}) d\mathbf{\Theta} && \text{(i)} \\
&= \sum_{d=1}^D \sum_{n=1}^{N_d} \sum_{k=1}^K q_{dnk} \int q(\boldsymbol{\theta}_d|\alpha_{d1}, \cdots, \alpha_{dK}) \log \theta_{dk} d\boldsymbol{\theta}_d && \text{(ii)} \\
&= \sum_{d=1}^D \sum_{n=1}^{N_d} \sum_{k=1}^K q_{dnk} \left(\Psi(\alpha_{dk}) - \Psi\Big(\sum_{k'=1}^K \alpha_{dk'} \Big) \right)
\end{aligned}
$$

【途中式の途中式】

- (i)：
  - $q(\boldsymbol{Z}) = \prod_{d=1}^D \prod_{n=1}^{N_d} q(z_{dn})$ で、$q(z_{dn} = k) \equiv q_{dnk}$ より、置き換える。
  - 式 (1) より、$q(\mathbf{\Theta}) = \prod_{d=1}^D q(\boldsymbol{\theta}_d|\alpha_{d1}, \cdots, \alpha_{dK})$ で置き換える。

- $p(\boldsymbol{Z}|\boldsymbol{\Theta}) = \prod_{d=1}^{D}\prod_{n=1}^{N_d} p(z_{dn}=k|\boldsymbol{\theta}_d) = \prod_{d=1}^{D}\prod_{k=1}^{K}\theta_{dk}$ より、置き換える。
- (ii)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・**第 2 項**

$$
\begin{aligned}
&\int q(\boldsymbol{\Theta})\log p(\boldsymbol{\Theta}|\alpha)d\boldsymbol{\Theta} && \text{(i)}\\
&= \int \sum_{d=1}^{D} q(\boldsymbol{\theta}_d|\alpha_{d1},\cdots,\alpha_{dK})\log\frac{\Gamma(\alpha K)^D}{\Gamma(\alpha)^{DK}}\prod_{k=1}^{K}\theta_{dk}^{\alpha-1}d\boldsymbol{\theta}_d && \text{(ii)}\\
&= D\log\Gamma(\alpha K) - DK\log\Gamma(\alpha) + (\alpha-1)\sum_{d=1}^{D}\sum_{k=1}^{K}\int q(\boldsymbol{\theta}_d|\alpha_{d1},\cdots,\alpha_{dK})\log\theta_{dk}d\boldsymbol{\theta}_d && \text{(iii)}\\
&= D\log\Gamma(\alpha K) - DK\log\Gamma(\alpha) + (\alpha-1)\sum_{d=1}^{D}\sum_{k=1}^{K}\left(\Psi(\alpha_{dk}) - \Psi\Big(\sum_{k'=1}^{K}\alpha_{dk'}\Big)\right)
\end{aligned}
$$

【途中式の途中式】

- (i)：
  - 式 (1) より、$q(\boldsymbol{\Theta}) = \prod_{d=1}^{D} q(\boldsymbol{\theta}_d|\alpha_{d1},\cdots,\alpha_{dK})$ で置き換える。
  - $p(\boldsymbol{\Theta}|\alpha) = \prod_{d=1}^{D}\mathrm{Dirichlet}(\boldsymbol{\theta}_d|\alpha)$ であるため、1.2.4 節より置き換える。
- (ii)：θ_{dk} と関係のない項を $\int$ の外に出して、式を整理する。
- (iii)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・**第 3 項**

$$
\begin{aligned}
&\int\sum_{\boldsymbol{Z}} q(\boldsymbol{Z})q(\boldsymbol{\Phi})\log p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})d\boldsymbol{\Phi} && \text{(i)}\\
&= \sum_{d=1}^{D}\sum_{n=1}^{N_d}\sum_{k=1}^{K} q_{dnk}\int q(\boldsymbol{\phi}_k|\beta_{k1},\cdots,\beta_{kw_{dn}})\log\phi_{kw_{dn}}d\boldsymbol{\phi}_k && \text{(ii)}\\
&= \sum_{d=1}^{D}\sum_{n=1}^{N_d}\sum_{k=1}^{K} q_{dnk}\left(\Psi(\beta_{kw_{dn}}) - \Psi\Big(\sum_{v=1}^{V}\beta_{kv}\Big)\right)
\end{aligned}
$$

【途中式の途中式】

- (i)：
  - $q(\boldsymbol{Z}) = \prod_{d=1}^{D}\prod_{n=1}^{N_d} q(z_{dn})$ で、$q(z_{dn}=k)\equiv q_{dnk}$ より、置き換える。
  - 式 (2) より、$q(\boldsymbol{\Phi}) = q(\boldsymbol{\phi}_k|\beta_{k1},\cdots,\beta_{kw_{dn}})$ で置き換える。
  - $p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi}) = \prod_{d=1}^{D}\prod_{n=1}^{N_d} q(w_{dn}|z_{dn}=k,\boldsymbol{\phi}_d) = \prod_{d=1}^{D}\prod_{n=1}^{N_d}\prod_{k=1}^{K}\phi_{kw_{dn}}$ より、置き換える。
- (ii)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

ただし、語彙レベルで考えると

$$
\begin{aligned}
&\int \sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) q(\boldsymbol{\Phi}) \log p(\boldsymbol{W}|\boldsymbol{Z}, \boldsymbol{\Phi}) d\boldsymbol{\Phi} \\
&= \sum_{d=1}^{D} \sum_{v=1}^{V} \sum_{k=1}^{K} q_{dvk} \int q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV}) \log \phi_{kv}^{N_{dv}} d\boldsymbol{\phi}_k \\
&= \sum_{d=1}^{D} \sum_{v=1}^{V} \sum_{k=1}^{K} q_{dvk} \int N_{dv} q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV}) \log \phi_{kv} d\boldsymbol{\phi}_k \\
&= \sum_{d=1}^{D} \sum_{v=1}^{V} \sum_{k=1}^{K} q_{dvk} N_{dv} \left(\Psi(\beta_{kv}) - \Psi\Big(\sum_{v'=1}^{V} \beta_{kv'} \Big) \right)
\end{aligned}
$$

となる。

・**第 4 項**

$$
\begin{aligned}
&\int q(\boldsymbol{\Phi}) \log p(\boldsymbol{\Phi}|\beta) d\boldsymbol{\Phi} && \text{(i)} \\
&= \int \sum_{k=1}^{K} q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV}) \log \left(\frac{\Gamma(\sum_{v=1}^{V} \beta)}{\prod_{v=1}^{V} \log \Gamma(\beta)} \prod_{v=1}^{V} \phi_{kv}^{\beta-1} \right) d\boldsymbol{\phi}_k && \text{(ii)} \\
&= K \log \Gamma(\beta V) - KV \log \Gamma(\beta) + (\beta - 1) \sum_{k=1}^{K} \sum_{v=1}^{V} \int q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV}) \log \phi_{kv} d\boldsymbol{\phi}_k && \text{(iii)} \\
&= K \log \Gamma(\beta V) - KV \log \Gamma(\beta) + (\beta - 1) \sum_{k=1}^{K} \sum_{v=1}^{V} \left(\Psi(\beta_{kv}) - \Psi\Big(\sum_{v'=1}^{V} \beta_{kv'} \Big) \right)
\end{aligned}
$$

【途中式の途中式】

- (i)：
  - 式 (2) より、$q(\boldsymbol{\Phi}) = \prod_{k=1}^{K} q(\boldsymbol{\phi}_k|\beta_{k1}, \cdots, \beta_{kV})$ で置き換える。
  - $p(\boldsymbol{\Phi}|\beta) = \prod_{k=1}^{K} \mathrm{Dirichlet}(\boldsymbol{\phi}_k|\beta)$ であるため、1.2.4 節より置き換える。
- (ii)：ϕ_{kv} と関係のない項を $\int$ の外に出して、式を整理する。
- (iii)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・**第 5 項**

$$
\begin{aligned}
&\sum_{\boldsymbol{Z}} q(\boldsymbol{Z}) \log q(\boldsymbol{Z}) && \text{(i)} \\
&= \sum_{d=1}^{D} \sum_{n=1}^{N_d} \sum_{k=1}^{K} q_{dnk} \log q_{dnk}
\end{aligned}
$$

【途中式の途中式】

- (i)：$q(\boldsymbol{Z}) = \prod_{d=1}^D \prod_{n=1}^{N_d} q(z_{dn})$ で、$q(z_{dn} = k) \equiv q_{dnk}$ より、置き換える。

・第 6 項

$$
\begin{aligned}
&\int q(\boldsymbol{\Theta}) \log q(\boldsymbol{\Theta}) d\boldsymbol{\Theta} && \text{(i)} \\
&= \sum_{d=1}^D \left\{ \int q(\boldsymbol{\theta}_d | \alpha_{d1}, \cdots, \alpha_{dK}) \log \left(\frac{\Gamma(\sum_{k=1}^K \alpha_{dk})}{\prod_{k=1}^K \Gamma(\alpha_{dk})} \prod_{k=1}^K \theta_{dk}^{\alpha_{dk}-1} \right) d\boldsymbol{\theta}_d \right\} && \text{(ii)} \\
&= \sum_{d=1}^D \left(\log \Gamma\Bigl(\sum_{k=1}^K \alpha_{dk} \Bigr) - \sum_{k=1}^K \log \Gamma(\alpha_{dk}) + \sum_{k=1}^K (\alpha_{dk} - 1) \int q(\boldsymbol{\theta}_d | \alpha_{d1}, \cdots, \alpha_{dK}) \log \theta_{dk} d\boldsymbol{\theta}_d \right) && \text{(iii)} \\
&= \sum_{d=1}^D \left\{ \log \Gamma\Bigl(\sum_{k=1}^K \alpha_{dk} \Bigr) - \sum_{k=1}^K \log \Gamma(\alpha_{dk}) + \sum_{k=1}^K (\alpha_{dk} - 1) \left(\Psi(\alpha_{dk}) - \Psi\Bigl(\sum_{k'=1}^K \alpha_{dk'} \Bigr) \right) \right\}
\end{aligned}
$$

【途中式の途中式】

- (i)：式 (1) より
  - $q(\boldsymbol{\Theta}) = \prod_{d=1}^D q(\boldsymbol{\theta}_d | \alpha_{d1}, \cdots, \alpha_{dK})$ で置き換える。
  - $q(\boldsymbol{\Theta}) = \prod_{d=1}^D \mathrm{Dirichlet}(\boldsymbol{\theta}_d | \alpha_{d1}, \cdots, \alpha_{dK})$ であるため、1.2.4 節より置き換える。
- (ii)：θ_{dk} と関係のない項を $\int$ の外に出して、式を整理する。
- (iii)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・第 7 項

$$
\begin{aligned}
&\int q(\boldsymbol{\Phi}) \log q(\boldsymbol{\Phi}) d\boldsymbol{\Phi} && \text{(i)} \\
&= \sum_{k=1}^K \left\{ \int q(\boldsymbol{\phi}_k | \beta_{k1}, \cdots, \beta_{kV}) \log \left(\frac{\Gamma(\sum_{v=1}^V \beta_{kv})}{\prod_{v=1}^V \log \Gamma(\beta_{kv})} \prod_{v=1}^V \phi_{kv}^{\beta_{kv}-1} \right) d\boldsymbol{\phi}_k \right\} && \text{(ii)} \\
&= \sum_{k=1}^K \left(\log \Gamma\Bigl(\sum_{v=1}^V \beta_{kv} \Bigr) - \sum_{v=1}^V \log \Gamma(\beta_{kv}) + \sum_{v=1}^V (\beta_{kv} - 1) \int q(\boldsymbol{\phi}_k) \log \phi_{kv} d\boldsymbol{\phi}_k \right) && \text{(iii)} \\
&= \sum_{k=1}^K \left\{ \log \Gamma\Bigl(\sum_{v=1}^V \beta_{kv} \Bigr) - \sum_{v=1}^V \log \Gamma(\beta_{kv}) + \sum_{v=1}^V (\beta_{kv} - 1) \left(\Psi(\beta_{kv}) - \Psi\Bigl(\sum_{v'=1}^V \beta_{kv'} \Bigr) \right) \right\}
\end{aligned}
$$

【途中式の途中式】

- (i)：式 (2) より
  - $q(\boldsymbol{\Phi}) = \prod_{k=1}^K q(\boldsymbol{\phi}_k | \beta_{k1}, \cdots, \beta_{kV})$ で置き換える。
  - $q(\boldsymbol{\Phi}) = \prod_{k=1}^K \mathrm{Dirichlet}(\boldsymbol{\phi}_k | \beta_{k1}, \cdots, \beta_{kV})$ であるため、1.2.4 節より置き換える。
- (ii)：ϕ_{kv} と関係のない項を $\int$ の外に出して、式を整理する。
- (iii)：ディリクレ分布に従う変数の対数の期待値 (1.17) より、置き換える。

・**R で組んでみる**

・コード全体

・テキスト処理

```
# 利用パッケージ
library(RMeCab)
library(tidyverse)

### テキスト処理

# 抽出しない単語を指定
stop_words <- "[a-z]" # 小文字のアルファベットを含む語

# 形態素解析
mecab_df <- docDF(" フォルダ名", type = 1) # テキストファイルの保存先を指定する

# 文書 d の語彙 v の出現回数 (N_dv) の集合
N_dv <- mecab_df %>%
        filter(grepl(" 名詞 | 形容詞 |^ 動詞", POS1)) %>% # 抽出する品詞 (大分類) を指定する
        filter(grepl(" 一般 |^ 自立", POS2)) %>%         # 抽出する品詞 (細分類) を指定する
        filter(!grepl(stop_words, TERM)) %>%           # ストップワードを除く
        select(-c(TERM, POS1, POS2)) %>%               # 数値列のみを残す
        filter(apply(., 1, sum) >= 5) %>%              # 抽出する総出現回数を指定する
        t()                                            # 転置

# 確認用の行列名
dimnames(N_dv) <- list(paste0("d=", 1:nrow(N_dv)), # 行名
                       paste0("v=", 1:ncol(N_dv))) # 列名

# 文書 d の単語数 (N_d) のベクトル
N_d <- apply(N_dv, 1, sum) # 行方向に和をとる

# 文書数 (D)
D <- nrow(N_dv)

# 総語彙数 (V)
V <- ncol(N_dv)
```

・パラメータの初期設定

```
### パラメータの初期設定

# トピック数 (K)
K <- 4   # 任意の値を指定する

# トピックの変分事後分布 (q_dvk) の集合
q_dvk <- array(0, dim = c(D, V, K),
               dimnames = list(paste0("d=", 1:D),  # 確認用の次元名
                               paste0("v=", 1:V),
                               paste0("k=", 1:K)))

# 事前分布のパラメータ (alpha, beta)
alpha <- 2 # 任意の値を指定する
```

```r
beta  <- 2 # 任意の値を指定する

# トピック分布の変分事後分布のパラメータ（alpha_dk）の集合
alpha_dk <- seq(1, 3, 0.01) %>%                   # 任意の範囲を指定する
            sample(D * K, replace = TRUE) %>%  # 値をランダムに生成
            matrix(nrow = D, ncol = K,          # 次元の設定
                   dimnames = list(paste0("d=", 1:D),  # 行名
                                   paste0("k=", 1:K))) # 列名

# 単語分布の変分事後分布のパラメータ（beta_kv）の集合
beta_kv <- seq(1, 3, 0.01) %>%                   # 任意の範囲を指定する
           sample(K * V, replace = TRUE) %>%  # 値をランダムに生成
           matrix(nrow = K, ncol = V,          # 次元の設定
                  dimnames = list(paste0("k=", 1:K),  # 行名
                                  paste0("v=", 1:V))) # 列名
```

・変分ベイズ推定

```r
### 変分ベイズ推定

# 推移の確認用
trace_alpha <- alpha_dk[1, ]
trace_beta  <- beta_kv[1, ]

for(i in 1:50) { # 任意の回数を指定する

  # パラメータを初期化
  new_alpha_dk <- matrix(data = alpha, nrow = D, ncol = K,
                         dimnames = list(paste0("d=", 1:D),
                                         paste0("k=", 1:K)))

  new_beta_kv <- matrix(data = beta, nrow = K, ncol = V,
                        dimnames = list(paste0("k=", 1:K),
                                        paste0("v=", 1:V)))

  for(d in 1:D) { ## （各文書）

    for(v in 1:V) { ## （各語彙）

      for(k in 1:K) { ## （各トピック）

        # 負担率を計算
        tmp_q_alpha <- digamma(alpha_dk[d, k]) - digamma(sum(alpha_dk[d, ]))
        tmp_q_beta  <- N_dv[d, v] * (digamma(beta_kv[k, v]) - digamma(sum(beta_kv[k, ])))
        q_dvk[d, v, k] <- exp(tmp_q_alpha + tmp_q_beta)

      }

      # 負担率を正規化
      q_dvk[d, v, ] <- q_dvk[d, v, ] / sum(q_dvk[d, v, ])

      for(k in 1:K) { ## （各トピック（続き））

        # alpha_dk を更新
```

```
        new_alpha_dk[d, k] <- new_alpha_dk[d, k] + q_dvk[d, v, k]

        # beta_kv を更新
        new_beta_kv[k, v] <- new_beta_kv[k, v] + q_dvk[d, v, k] * N_dv[d, v]

      } ## (/各トピック)
    } ## (/各語彙)
  } ## (/各文書)

  # パラメータを更新
  alpha_dk <- new_alpha_dk
  beta_kv  <- new_beta_kv

  # 推移の確認用
  trace_alpha <- rbind(trace_alpha, alpha_dk[1, ])
  trace_beta  <- rbind(trace_beta, beta_kv[1, ])
}
```

・コードの解説

・利用パッケージ

```
# 利用パッケージ
library(RMeCab)
library(tidyverse)
```

形態素解析のための `RMeCab::docDF()` と、グラフ作成時の `tidyr::gather()`、`ggplot2` 関連の関数以外にライブラリの読み込みが必要なのは `dplyr` のみ。(面倒なので `tidyvers` パッケージを読み込んでいる)

その他テキスト処理の部分は 3.3 節を参照のこと。

・パラメータの初期設定

・トピック数

```
# トピック数(K)
K <- 4   # 任意の値を指定する
```

任意のトピック数を指定する。

・負担率

```
# トピックの変分事後分布(q_dvk)の集合
q_dvk <- array(0, dim = c(D, V, K),
               dimnames = list(paste0("d=", 1:D),  # 確認用の次元名
                               paste0("v=", 1:V),
                               paste0("k=", 1:K)))
```

ここでは負担率 q_{dnk} を単語レベルではなく語彙レベルで扱うため、q_{dnk} を `q_dvk` として扱う。
`for()` 文の中で添え字で指定して 1 要素ずつ代入していくため、受け皿として事前に用意しておく。

トピックモデルの負担率 q_{dvk} は 3 次元の変数なので、array() を使って配列として扱う。初期値を 0 として、引数 dim = c(D, V, K) で各次元の要素数を D, V, K 個とする。
変数に付ける行列名等は目で確認するためのもので、付けなくても処理に影響しない。他の変数についても同様である。

・事前分布のパラメータ

```
# 事前分布のパラメータ (alpha, beta)
alpha <- 2 # 任意の値を指定する
beta  <- 2 # 任意の値を指定する
```

事前分布 $p(\boldsymbol{\theta}|\alpha), p(\boldsymbol{\Phi}|\beta)$ のパラメータ α, β の値を任意で設定する。
この値は更新パラメータの初期化時の値として用いる。

・トピック分布の変分事後分布のパラメータ

```
# トピック分布の変分事後分布のパラメータ (alpha_dk) の集合
alpha_dk <- seq(1, 3, 0.01) %>%                 # 任意の範囲を指定する
            sample(D * K, replace = TRUE) %>%  # 値をランダムに生成
            matrix(nrow = D, ncol = K,         # 次元の設定
                   dimnames = list(paste0("d=", 1:D),  # 行名
                                   paste0("k=", 1:K))) # 列名
```

seq(最小値，最大値，間隔) で初期値として取り得る値の範囲を任意で指定する。そこから、sample() でランダムに $D * K$ 個の値をとる。その数値ベクトルを matrix() で D 行 K 列の行列とする。

・単語分布の変分事後分布のパラメータ

```
# 単語分布の変分事後分布のパラメータ (beta_kv) の集合
beta_kv <- seq(1, 3, 0.01) %>%                 # 任意の範囲を指定する
           sample(K * V, replace = TRUE) %>%  # 値をランダムに生成
           matrix(nrow = K, ncol = V,         # 次元の設定
                  dimnames = list(paste0("k=", 1:K),  # 行名
                                  paste0("v=", 1:V))) # 列名
```

α と同様に、初期値をランダムに取る K 行 V 列の行列を作る。

・変分ベイズ推定

ここからは、for() ループ内の処理について説明する。

・ハイパーパラメータの初期化

```
# パラメータを初期化
new_alpha_dk <- matrix(data = alpha, nrow = D, ncol = K,
                       dimnames = list(paste0("d=", 1:D),
                                       paste0("k=", 1:K)))
```

```
new_beta_kv <- matrix(data = beta, nrow = K, ncol = V,
                      dimnames = list(paste0("k=", 1:K),
                                      paste0("v=", 1:V)))
```

各パラメーターの更新時に用いる変数を作成する。
また、次ステップに入る前に事前分布のパラメータとして設定した値でそれぞれ初期化する。

・負担率の計算

```
for(k in 1:K) { ## （各トピック）

  # 負担率（q_dvk）を計算
  tmp_q_alpha <- digamma(alpha_dk[d, k]) - digamma(sum(alpha_dk[d, ]))
  tmp_q_beta  <- N_dv[d, v] * (digamma(beta_kv[k, v]) - digamma(sum(beta_kv[k, ])))
  q_dvk[d, v, k] <- exp(tmp_q_alpha + tmp_q_beta)

}

# 負担率を正規化
q_dvk[d, v, ] <- q_dvk[d, v, ] / sum(q_dvk[d, v, ])
```

負担率の計算を

$$q_{dvk} \propto \exp\left\{\Psi(\alpha_{dk}) - \Psi\Bigl(\sum_{k'=1}^{K} \alpha_{dk'}\Bigr) + N_{dv}\left(\Psi(\beta_{kv}) - \Psi\Bigl(\sum_{v=1}^{V} \beta_{kv}\Bigr)\right)\right\} \tag{4.8'}$$

で行う。
ただしこの式がイコールの関係でないことから、そのままの値を用いずに $\sum_{k=1}^{K} q_{dk} = 1$ となるように正規化した値を用いる。正規化処理の時点で 1~K までの要素が必要であるため、`for(k in 1:K)` のループ処理を先に回し切り、正規化後のパラメータの更新では、もう一度 `for(k in 1:K)` のループ処理を行う。

・ハイパーパラメータ：α_{dk} の更新

```
# alpha_dk を更新
new_alpha_dk[d, k] <- new_alpha_dk[d, k] + q_dvk[d, v, k]
```

α の更新式

$$\alpha_{dk} = \alpha + \sum_{v=1}^{V} q_{dvk}$$

の計算を行い値を求める。$\sum_{v=1}^{V}$ については、`for(v in 1:V)` によって各要素に V 回繰り返し加算されることがこれに相当する。

・ハイパーパラメータ：β_{kv} の更新

```
# beta_kv を更新
new_beta_kv[k, v] <- new_beta_kv[k, v] + q_dvk[d, v, k] * N_dv[d, v]
```

β_{kv} も同様に、更新式

$$\beta_{kv} = \beta + \sum_{d=1}^D \sum_{v=1}^V q_{dvk} N_{dv}$$

の計算を行い値を求める。

・ハイパーパラメータの更新

```
# パラメータを更新
alpha_dk <- new_alpha_dk
beta_kv  <- new_beta_kv
```

各パラメータを次ステップでの計算に用いるために移す。

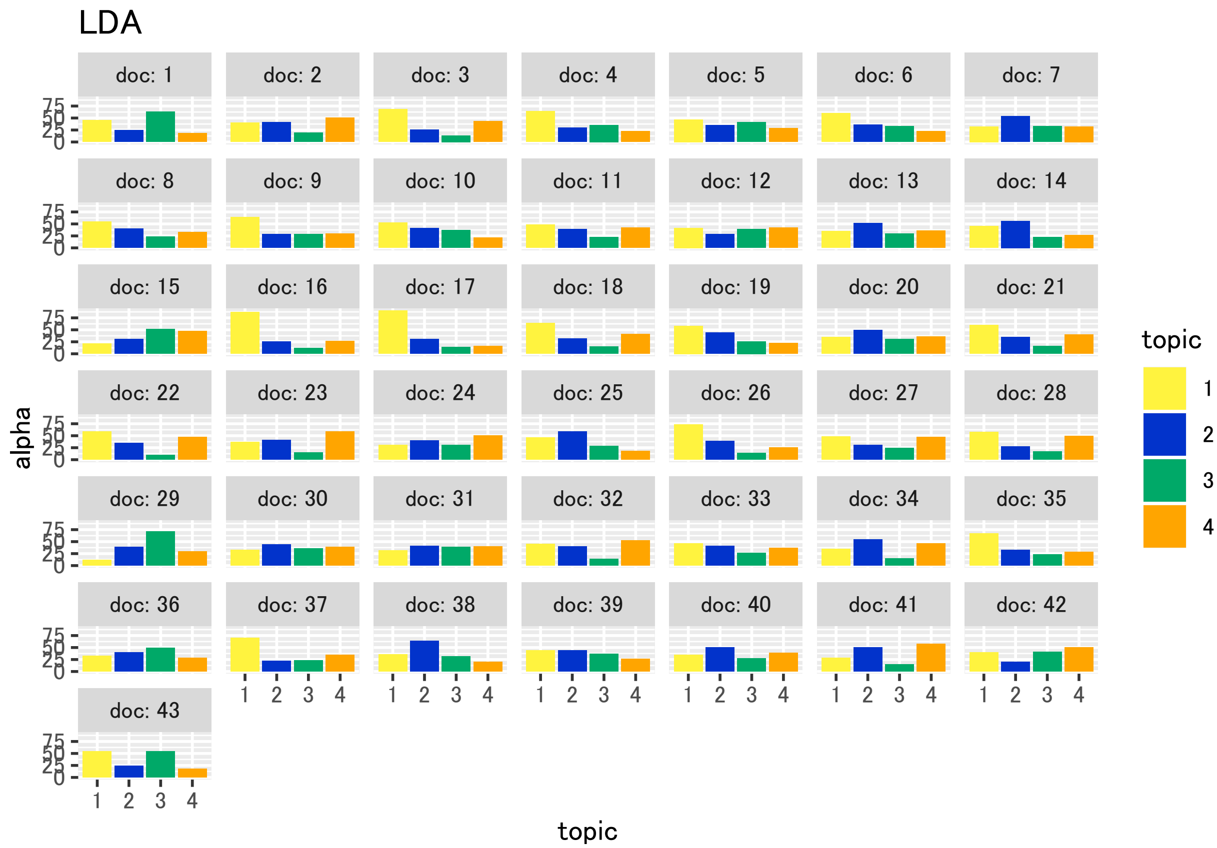
この推定処理を指定した回数繰り返す。(本当は収束条件を満たすまで繰り返すようにしたい…)

・推定結果の確認

・ハイパーパラメータの推定結果の確認

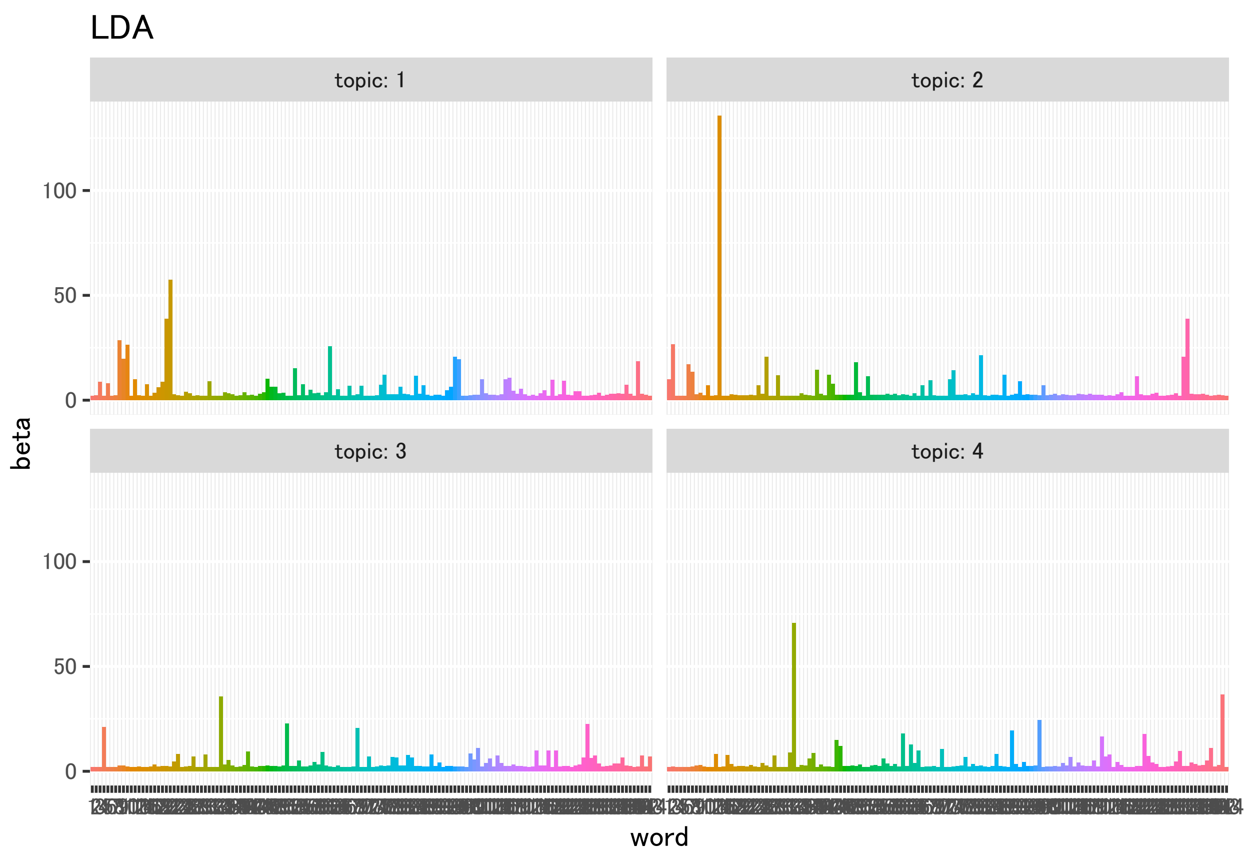
・トピック分布のパラメータ

```
## トピック分布のパラメータ
# データフレームを作成
alpha_df_wide <- cbind(as.data.frame(alpha_dk),
                       as.factor(1:D))

# データフレームを long 型に変換
colnames(alpha_df_wide) <- c(1:K, "doc")         # key 用に行名を付与
alpha_df_long <- gather(alpha_df_wide, key = "topic", value = "alpha", -doc) # 変換
alpha_df_long$topic <- alpha_df_long$topic %>%  # 文字列から因子に変換
                       as.numeric() %>%
                       as.factor()

# 作図
ggplot(data = alpha_df_long, mapping = aes(x = topic, y = alpha, fill = topic)) + # データ
  geom_bar(stat = "identity", position = "dodge") +  # 棒グラフ
  facet_wrap(~ doc, labeller = label_both) +         # グラフの分割
  labs(title = "LDA")                                # タイトル
```



・単語分布のパラメータ

```
## 単語分布のパラメータ
# データフレームを作成
beta_df_wide <- cbind(as.data.frame(beta_kv),
                      as.factor(1:K))

# データフレームを long 型に変換
colnames(beta_df_wide) <- c(1:V, "topic")    # key 用に行名を付与
beta_df_long <- gather(beta_df_wide, key = "word", value = "beta", -topic) # 変換
beta_df_long$word <- beta_df_long$word %>%  # 文字列から因子に変換
                     as.numeric() %>%
                     as.factor()

# 作図
ggplot(data = beta_df_long, mapping = aes(x = word, y = beta, fill = word)) + # データ
  geom_bar(stat = "identity", position = "dodge") +  # 棒グラフ
  facet_wrap(~ topic, labeller = label_both) +       # グラフを分割
  theme(legend.position = "none") +                  # 凡例
  labs(title = "LDA")                                # タイトル
```



最尤推定ではパラメータを点推定したが、変分ベイズ推定ではハイパーパラメータを推定することでパラメータを分布推定している。

・パラメータの推定結果 (平均値) の確認

ディリクレ分布の平均値の計算式

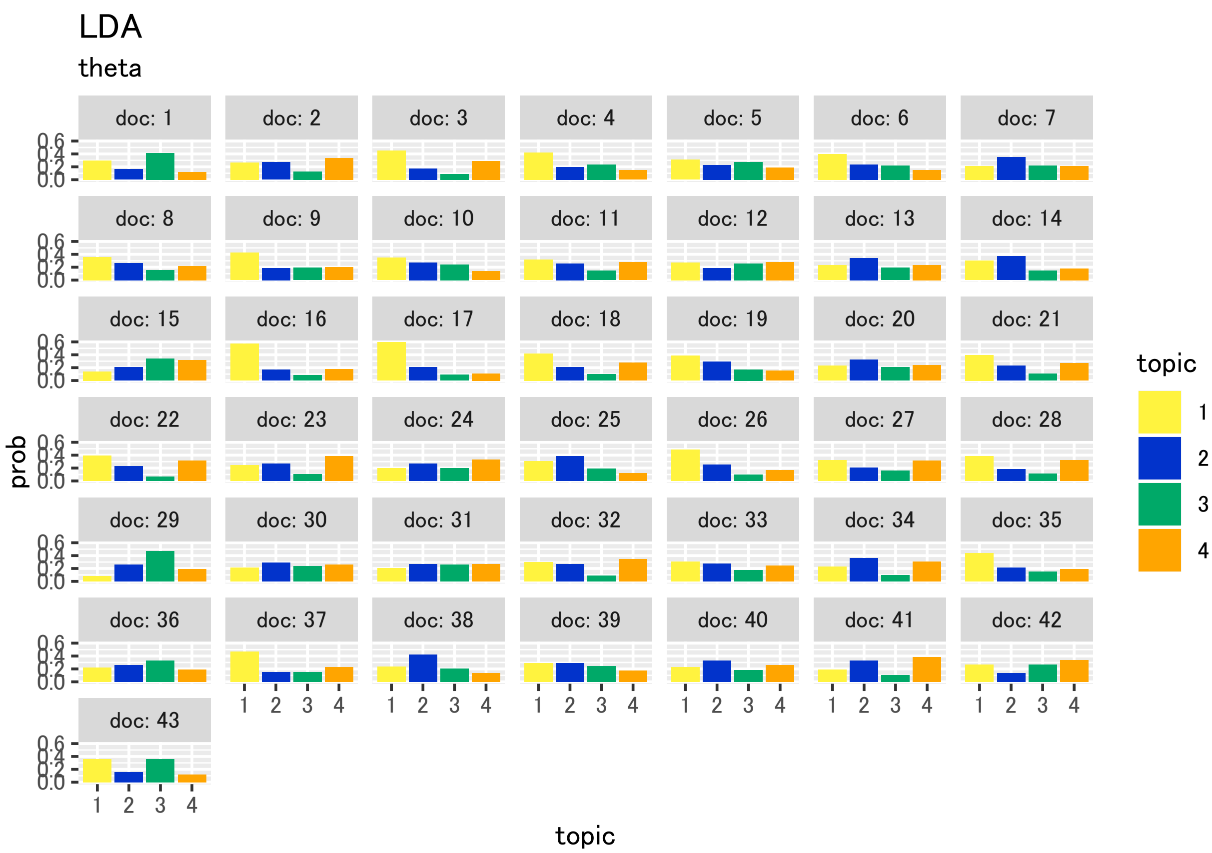
$$\mathbb{E}[\phi_v] = \frac{\beta_v}{\sum_{v=1}^{V} \beta_v} \tag{1.15}$$

より、パラメータの平均値の分布を確認する。

・トピック分布 (平均値)

```
## トピック分布（平均値）
#パラメータの平均値を算出
theta_dk <- matrix(nrow = D, ncol = K)
for(d in 1:D) {
  theta_dk[d, ] <- alpha_dk[d, ] / sum(alpha_dk[d, ])
}

# データフレームを作成
```

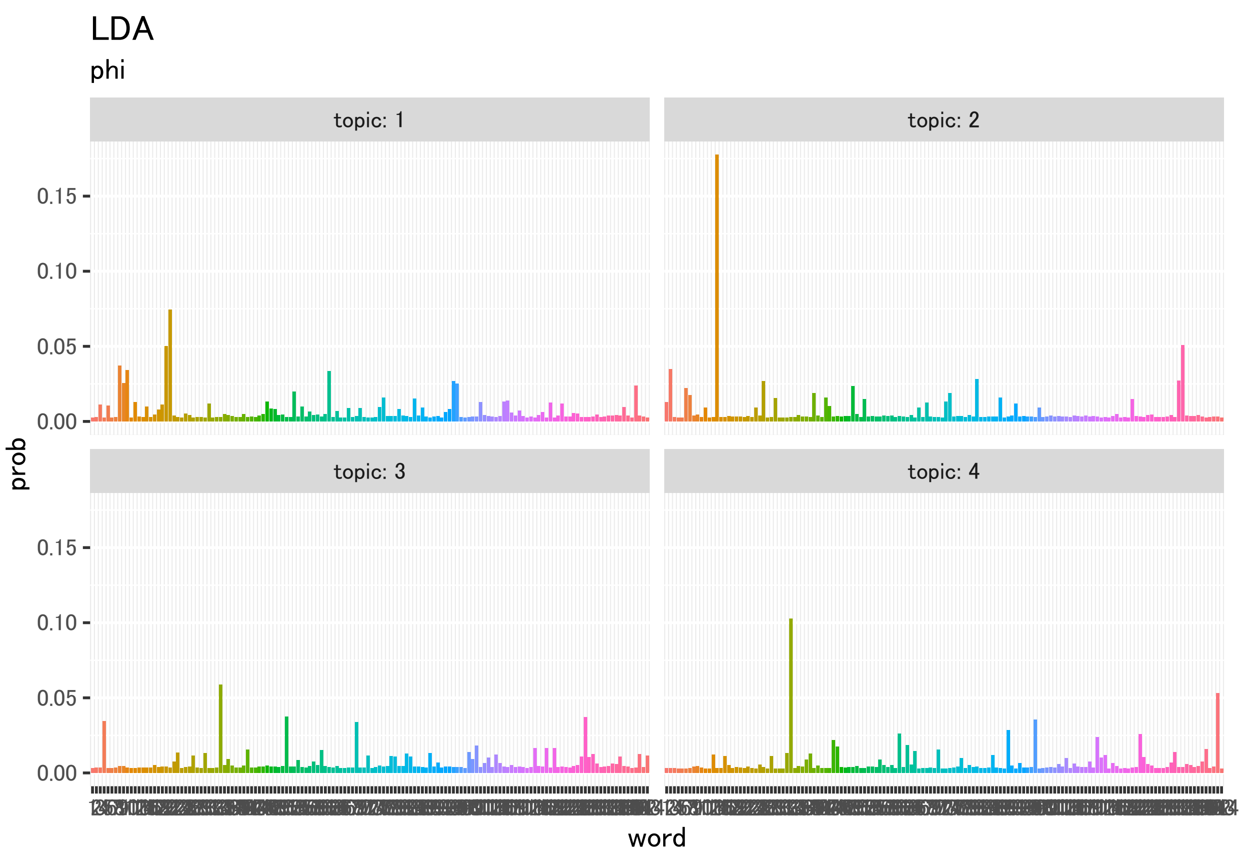
```
theta_df_wide <- cbind(as.data.frame(theta_dk),
                       as.factor(1:D)) # 文書番号の列を加える

# データフレームを long 型に変換
colnames(theta_df_wide) <- c(1:K, "doc")        # key 用に行名を付与
theta_df_long <- gather(theta_df_wide, key = "topic", value = "prob", -doc) # 変換
theta_df_long$topic <- theta_df_long$topic %>%  # 文字列から因子に変換
                       as.numeric() %>%
                       as.factor()

# 作図
ggplot(data = theta_df_long, mapping = aes(x = topic, y = prob, fill = topic)) + # データ
  geom_bar(stat = "identity", position = "dodge") +  # 棒グラフ
  facet_wrap(~ doc, labeller = label_both) +         # グラフの分割
  labs(title = "LDA")                                # タイトル
```



・単語分布 (平均値)

```
## 単語分布 (平均値)
#パラメータの平均値を算出
phi_kv <- matrix(nrow = K, ncol = V)
for(k in 1:K) {
  phi_kv[k, ] <- beta_kv[k, ] / sum(beta_kv[k, ])
}

# データフレームを作成
```

```r
phi_df_wide <- cbind(as.data.frame(phi_kv),
                     as.factor(1:K)) # トピック番号の列を加える

# データフレームを long 型に変換
colnames(phi_df_wide) <- c(1:V, "topic")  # key 用に行名を付与
phi_df_long <- gather(phi_df_wide, key = "word", value = "prob", -topic) # 変換
phi_df_long$word <- phi_df_long$word %>%  # 文字列から因子に変換
                    as.numeric() %>%
                    as.factor()

# 作図
ggplot(data = phi_df_long, mapping = aes(x = word, y = prob, fill = word)) + # データ
  geom_bar(stat = "identity", position = "dodge") +  # 棒グラフ
  facet_wrap(~ topic, labeller = label_both) +        # グラフを分割
  theme(legend.position = "none") +                   # 凡例
  labs(title = "LDA", subtitle = "phi")               # タイトル
```



・推移の確認

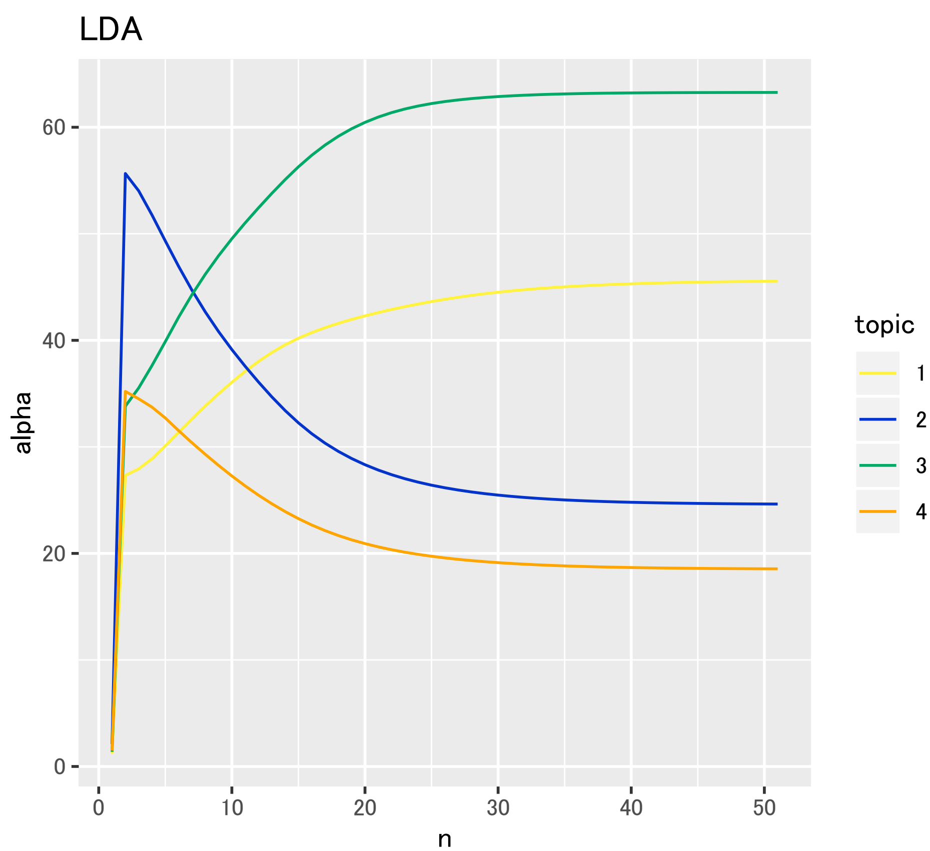
・推移の記録

```r
# 推移の確認用（初期値）
trace_alpha <- alpha_dk[1, ]
trace_beta  <- beta_kv[1, ]
```

```
# 推移の確認用（更新後）
trace_alpha <- rbind(trace_alpha, alpha_dk[1, ])
trace_beta  <- rbind(trace_beta, beta_kv[1, ])
```

・トピック分布のパラメータ

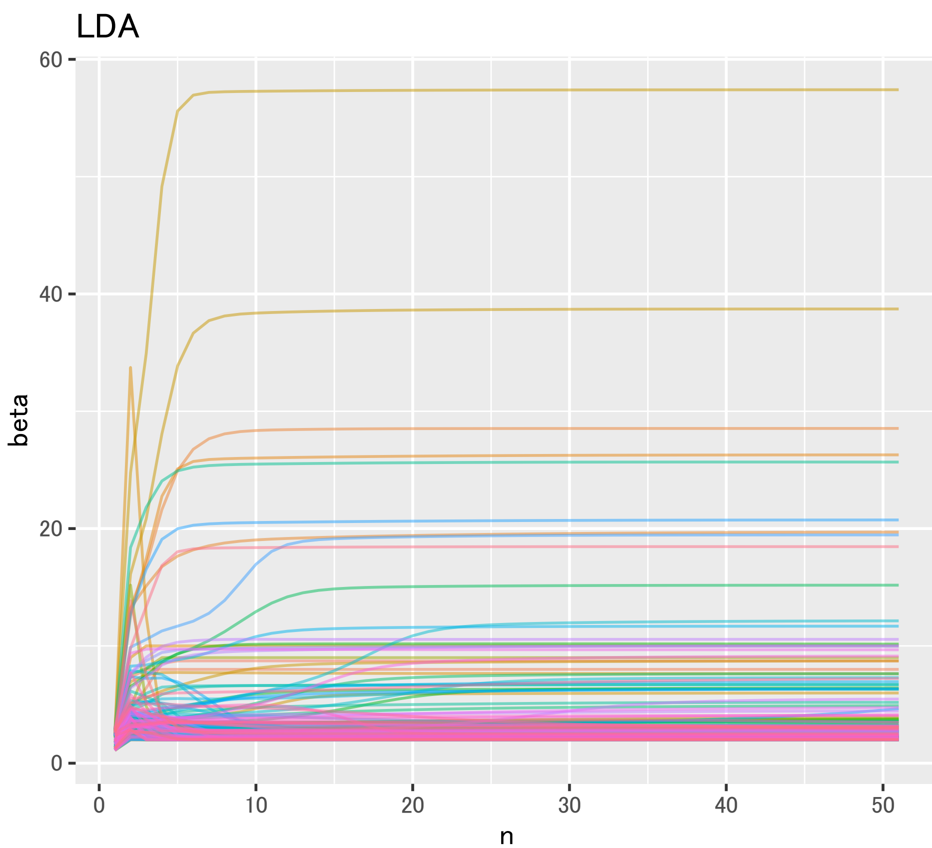
```
## トピック分布のパラメータ
# データフレームを作成
trace_alpha_df_wide <- cbind(as.data.frame(trace_alpha),
                             1:nrow(trace_alpha)) # 試行回数


# データフレームを long 型に変換
colnames(trace_alpha_df_wide) <- c(1:K, "n") # key 用の行名を付与
trace_alpha_df_long <- gather(trace_alpha_df_wide, key = "topic", value = "alpha", -n) # 変換

# 描画
ggplot(data = trace_alpha_df_long, mapping = aes(x = n, y = alpha, color = topic)) +  # データ
  geom_line() +                # 棒グラフ
  labs(title = "LDA") # タイトル
```



・単語分布のパラメータ

```
## 単語分布のパラメータ
# データフレームを作成
trace_beta_df_wide <- cbind(as.data.frame(trace_beta),
                            1:nrow(trace_beta)) # 試行回数
```

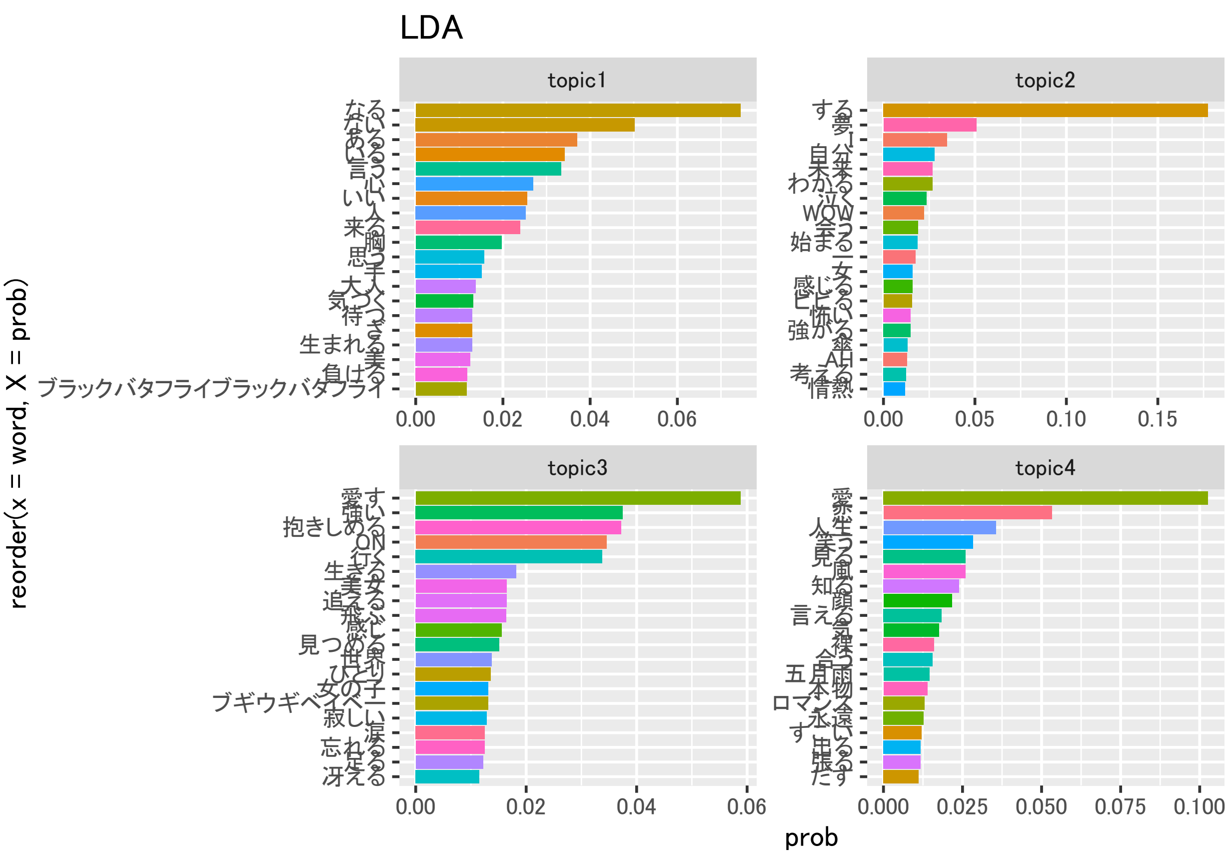
```r
# データフレームを long 型に変換
colnames(trace_beta_df_wide) <- c(1:V, "n")           # key 用の行名を付与
trace_beta_df_long <- gather(trace_beta_df_wide, key = "word", value = "beta", -n) # 変換
trace_beta_df_long$word <- trace_beta_df_long$word %>%  # 文字列になるため因子に変換
                           as.numeric() %>%
                           as.factor()

# 描画
ggplot(data = trace_beta_df_long, mapping = aes(x = n, y = beta, color = word)) +
  geom_line(alpha = 0.5) +           # 棒グラフ
  theme(legend.position = "none") +  # 凡例
  labs(title = "LDA")             # タイトル
```



・各トピックの出現確率の上位語

```r
### 各トピックの出現確率の上位語

## 語彙インデックス (v)
# 指定した出現回数以上の単語の行番号
num <- mecab_df %>%
  select(-c(TERM, POS1, POS2)) %>%
  apply(1, sum) >= 5 # 抽出する総出現回数を指定する

v_index <- mecab_df[num, ] %>%
```

```r
  filter(grepl("名詞|形容詞|^動詞", POS1)) %>%  # 抽出する品詞を指定する
  filter(grepl("一般|^自立", POS2)) %>%
  filter(!grepl(stop_words, TERM)) %>%
  .[, 1]  # 単語の列を抽出する

# データフレームを作成
phi_df_wide2 <- cbind(as.data.frame(t(phi_kv)),
                      v_index) %>%
  as.data.frame()
colnames(phi_df_wide2) <- c(paste0("topic", 1:K), "word") # key用の行名を付与

# データフレームの整形
phi_df_long2 <- data.frame()
for(i in 1:K) {
  tmp_df <- phi_df_wide2 %>%
    select(paste0("topic", i), word) %>%  #
    arrange(-.[, 1]) %>%  # 降順に並べ替え
    head(20) %>%          # 任意で指定した上位n語を抽出
    gather(key = "topic", value = "prob", -word) # long型に変換

  phi_df_long2 <- rbind(phi_df_long2, tmp_df)
}

# 描画
ggplot(data = phi_df_long2,
       mapping = aes(x = reorder(x = word, X = prob), y = prob, fill = word)) + # データ
  geom_bar(stat = "identity", position = "dodge") + # 棒グラフ
  coord_flip() +                                     # グラフの向き
  facet_wrap( ~ topic, scales = "free") +            # グラフの分割
  theme(legend.position = "none") +                  # 凡例
  labs(title = "LDA")                                # タイトル
```



4.5 ギブスサンプリング

4.5.1 パラメータの周辺化

トピックモデルを崩壊型ギブスサンプリングで推定する。混合ユニグラムモデルと同様に、パラメータ $\boldsymbol{\Theta}, \boldsymbol{\Phi}$ を積分消去し、トピック集合 $\boldsymbol{Z}$ の事後分布 $p(\boldsymbol{Z}|\boldsymbol{W}, \alpha, \beta)$ を推定する。

パラメータ $\boldsymbol{\Theta}, \boldsymbol{\Phi}$ を積分消去した文書集合 $\boldsymbol{W}$ とトピック集合 $\boldsymbol{Z}$ の同時分布は生成過程から

$$p(\boldsymbol{Z}, \boldsymbol{W}|\alpha, \beta) = p(\boldsymbol{Z}|\alpha)p(\boldsymbol{W}|\boldsymbol{Z}, \beta) \tag{4.9}$$

と分解できる。この式の 1 つ目の因子は

$$
\begin{aligned}
p(\boldsymbol{Z}|\alpha) &= \int p(\boldsymbol{Z}|\boldsymbol{\Theta}) p(\boldsymbol{\Theta}|\alpha) d\boldsymbol{\Theta} && \text{(i)} \\
&= \int \prod_{d=1}^{D} \prod_{k=1}^{K} \theta_{dk}^{N_{dk}} \prod_{d=1}^{D} \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \prod_{k=1}^{K} \theta_{dk}^{\alpha-1} d\boldsymbol{\theta}_d && \text{(ii)} \\
&= \frac{\Gamma(\alpha K)^D}{\Gamma(\alpha)^{KD}} \int \prod_{d=1}^{D} \prod_{k=1}^{K} \theta_{dk}^{N_{dk}+\alpha-1} d\boldsymbol{\theta}_d && \text{(iii)} \\
&= \frac{\Gamma(\alpha K)^D}{\Gamma(\alpha)^{KD}} \prod_{d=1}^{D} \frac{\prod_{k=1}^{K} \Gamma(N_{dk}+\alpha)}{\Gamma(N_d+\alpha K)} && \text{(iv, 4.10)} \\
&= \prod_{d=1}^{D} \frac{\Gamma(\alpha K)}{\Gamma(N_d+\alpha K)} \prod_{d=1}^{D} \prod_{k=1}^{K} \frac{\Gamma(N_{dk}+\alpha)}{\Gamma(\alpha)} && \text{(4.10')}
\end{aligned}
$$

【途中式の途中式】

- (0)：そもそも左辺 $p(\boldsymbol{Z}|\alpha)$ は、右辺の式の $\boldsymbol{\Theta}$ を積分消去したものであるため。
- (i)：
  - $p(\boldsymbol{Z}|\boldsymbol{\Theta}) = \prod_{d=1}^{D} \prod_{k=1}^{K} \mathrm{Categorical}(\boldsymbol{z}_d = k|\boldsymbol{\theta}_d)$ であることから、1.2.2 節より置き換える。
  - $p(\boldsymbol{\Theta}|\alpha) = \prod_{d=1}^{D} \mathrm{Dicichlet}(\boldsymbol{\theta}_d|\alpha)$ であることから、1.2.4 節より置き換える。
- (ii)：$\boldsymbol{\theta}_d$ と関係のない項を $\int$ の外に出し、θ_{dk} の項を 1 つにまとめる。
- (iii)：ディリクレ分布の正規化項 (1.13) より、置き換える。
- (iv)：分母分子の αK と α を揃えるために、分母を入れ替える。

となる。2 つ目の因子も同様に

$$
\begin{aligned}
p(\boldsymbol{W}|\boldsymbol{Z},\beta) &= \int p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi}) p(\boldsymbol{\Phi}|\beta) d\boldsymbol{\Phi} && \text{(i)} \\
&= \int \prod_{k=1}^{K} \prod_{v=1}^{V} \phi_{kv}^{N_{kv}} \prod_{k=1}^{K} \frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \prod_{v=1}^{V} \phi_{kv}^{\beta-1} d\boldsymbol{\phi}_K && \text{(ii)} \\
&= \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{VK}} \int \prod_{k=1}^{K} \prod_{v=1}^{V} \phi_{kv}^{N_{kv}+\beta-1} d\boldsymbol{\phi}_K && \text{(iii)} \\
&= \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{VK}} \prod_{k=1}^{K} \frac{\prod_{v=1}^{V} \Gamma(N_{kv}+\beta)}{\Gamma(N_k+\beta V)} && \text{(iv, 4.11)} \\
&= \prod_{k=1}^{K} \frac{\Gamma(\beta V)}{\Gamma(N_k+\beta V)} \prod_{k=1}^{K} \prod_{v=1}^{V} \frac{\Gamma(N_{kv}+\beta)}{\Gamma(\beta)} && \text{(4.11')}
\end{aligned}
$$

【途中式の途中式】

- (0)：そもそも左辺 $p(\boldsymbol{W}|\boldsymbol{Z},\beta)$ は、右辺の式の $\boldsymbol{\Phi}$ を積分消去したものであるため。
- (i)：
  - $p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi}) = \prod_{d=1}^{D} \mathrm{Categorical}(\boldsymbol{w}_d|\boldsymbol{z}_d = k, \boldsymbol{\phi}_k)$ であることから、1.2.2 節より置き換える。
  - $p(\boldsymbol{\Phi}|\beta) = \prod_{k=1}^{K} \mathrm{Dirichlet}(\boldsymbol{\phi}_k|\beta)$ であることから、1.2.4 節より置き換える。
- (ii)：$\boldsymbol{\phi}_K$ と関係のない項を $\int$ の外に出し、ϕ_{kv} の項を 1 つにまとめる。
- (iii)：ディリクレ分布の正規化項 (1.13) より、置き換える。

- (iv)：分母分子の βV と β を揃えるために、分母を入れ替える。

となる。

4.5.2 サンプリング式

・サンプリング式

ギブズサンプリングでは、単語ごとにトピックをサンプリングしていく。

文書 d の n 番目の単語のトピックをサンプリングする確率は、そのトピックを除いたトピック集合 $\boldsymbol{Z}_{\backslash dn} = (z_{11}, \cdots, z_{d,n-1}, z_{d,n+1}, \cdots, z_{DN_d})$ と文書集合 $\boldsymbol{W}$ が与えられたときの、トピック z_{dn} の条件付き確率は、ベイズの定理 (1.4) を用いて

$$
p(z_{dn} = k | \boldsymbol{W}, \boldsymbol{Z}_{\backslash dn}, \alpha, \beta) \propto p(z_{dn} = k | \boldsymbol{Z}_{\backslash dn}, \alpha) p(w_{dn} | \boldsymbol{W}_{\backslash dn}, z_{dn} = k, \boldsymbol{Z}_{\backslash dn}, \beta) \tag{4.12}
$$

で与えられる。この式の 1 つ目の因子は

$$
\begin{aligned}
p(z_{dn} = k | \boldsymbol{Z}_{\backslash dn}, \alpha) &= \frac{p(z_{dn} = k, \boldsymbol{Z}_{\backslash dn} | \alpha)}{p(\boldsymbol{Z}_{\backslash dn} | \alpha)} && \text{(i)} \\
&= \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \frac{\Gamma(N_{dk\backslash dn} + 1 + \alpha) \prod_{k' \neq k} \Gamma(N_{dk'\backslash dn} + \alpha)}{\Gamma(N_d + \alpha K)} \frac{\Gamma(\alpha)^K}{\Gamma(\alpha K)} \frac{\Gamma(N_d - 1 + \alpha K)}{\prod_{k'=1}^K \Gamma(N_{dk'\backslash dn} + \alpha)} && \text{(ii)} \\
&= \frac{(N_{dk\backslash dn} + \alpha) \Gamma(N_{dk\backslash dn} + \alpha) \prod_{k' \neq k} \Gamma(N_{dk'\backslash dn} + \alpha)}{(N_d - 1 + \alpha K) \Gamma(N_d - 1 + \alpha K)} \frac{\Gamma(N_d - 1 + \alpha K)}{\prod_{k'=1}^K \Gamma(N_{dk'\backslash dn} + \alpha)} && \text{(iii)} \\
&= \frac{N_{dk\backslash dn} + \alpha}{N_d - 1 + \alpha K} && (4.13)
\end{aligned}
$$

【途中式の途中式】

- (0) :??
- (i)：式 (4.10) を用いる。

$$
\begin{aligned}
p(\boldsymbol{Z} | \alpha) &= \frac{\Gamma(\alpha K)^D}{\Gamma(\alpha)^{KD}} \prod_{d=1}^D \frac{\prod_{k=1}^K \Gamma(N_{dk} + \alpha)}{\Gamma(N_d + \alpha K)} \\
\Rightarrow p(z_{dn} = k, \boldsymbol{Z}_{\backslash dn} | \alpha) &= \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \frac{\Gamma(N_{dk} + \alpha) \prod_{k' \neq k} \Gamma(N_{dk'} + \alpha)}{\Gamma(N_d + \alpha K)} \\
\Rightarrow p(\boldsymbol{Z}_{\backslash dn} | \alpha) &= \frac{\Gamma(\alpha K)}{\Gamma(\alpha)^K} \frac{\prod_{k=1}^K \Gamma(N_{dk} + \alpha)}{\Gamma(N_d + \alpha K)}
\end{aligned}
$$

ここでは文書 d についてのみ扱うので、$\prod_{d=1}^D$ は不要である。また、分子の 1~K までの掛け合わせから k についての項のみ取り出しておく。

ここで、N_{dk} は文書 d においてトピック k を割り当てられた単語数である。文書 d の n 番目の単語のトピック z_{dn} が k のとき、n 番目の単語を除いた文書 d のトピック k が割り当てられた単語数 $N_{dk\backslash dn}$ は

$N_{dk} - 1$ である。しかし z_{dn} が k 以外のトピックであれば、N_{dk} に文書 d の n 番目の単語の分のカウントはそもそも含まれていないため、$N_{dk\backslash dn} = N_{dk}$ である。これをまとめると

$$N_{dk\backslash dn} = \begin{cases} N_{dk} - 1 & \text{if} \quad z_{dn} = k \\ N_{dk} & \text{if} \quad z_{dn} \neq k \end{cases}$$

という関係である。これを N_{dk} について解くと

$$N_{dk} = \begin{cases} N_{dk\backslash dn} + 1 & \text{if} \quad z_{dn} = k \\ N_{dk\backslash dn} & \text{if} \quad z_{dn} \neq k \end{cases}$$

となる。これを代入する。また、分母 $p(\boldsymbol{Z}_{\backslash dn}|\alpha)$ の分母である n 番目の単語を除いたときの文書 d の総単語数 $\sum_{k=1}^K N_{dk\backslash dn} = N_{d\backslash dn}$ も同様にカウントを 1 つ減らして $N_d - 1$ になる。

- (ii)：ガンマ関数の性質 $\Gamma(x+1) = (x)\Gamma(x)$ を用いて変形する。
- (iii)：約分して式を整理する。ただし、$\Gamma(N_{dk\backslash dn} + \alpha) \prod_{k' \neq k} \Gamma(N_{dk'\backslash dn} + \alpha) = \prod_{k'=1}^K \Gamma(N_{dk'\backslash dn} + \alpha)$ である。

となる。2 つ目の因子は

$$\begin{aligned}
&p(w_{dn}|\boldsymbol{W}_{\backslash dn}, z_{dn} = k, \boldsymbol{Z}_{\backslash dn}, \beta) \\
&= \frac{p(w_{dn}, \boldsymbol{W}_{\backslash dn}|z_{dn} = k, \boldsymbol{Z}_{\backslash dn}, \beta)}{p(\boldsymbol{W}_{\backslash dn}|z_{dn} = k, \boldsymbol{Z}_{\backslash dn}, \beta)} && \text{(i)} \\
&= \frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \frac{\Gamma(N_{kw_{dn}\backslash dn} + 1 + \beta) \prod_{v \neq w_{dn}} \Gamma(N_{kv\backslash dn} + \beta)}{\Gamma(N_{k\backslash dn} + 1 + \beta V)} \frac{\Gamma(\beta)^V}{\Gamma(\beta V)} \frac{\Gamma(N_{k\backslash dn} + \beta V)}{\prod_{v=1}^V \Gamma(N_{kv\backslash dn} + \beta)} && \text{(ii)} \\
&= \frac{(N_{kw_{dn}\backslash dn} + \beta)\Gamma(N_{kw_{dn}\backslash dn} + \beta) \prod_{v \neq w_{dn}} \Gamma(N_{kv\backslash dn} + \beta)}{(N_{k\backslash dn} + \beta V)\Gamma(N_{k\backslash dn} + \beta V)} \frac{\Gamma(N_{k\backslash dn} + \beta V)}{\prod_{v=1}^V \Gamma(N_{kv\backslash dn} + \beta)} && \text{(iii)} \\
&= \frac{N_{kw_{dn}\backslash dn} + \beta}{N_{k\backslash dn} + \beta V} && (4.14)
\end{aligned}$$

【途中式の途中式】

- (0) :??
- (i)：式 (4.11) を用いる。

$$\begin{aligned}
p(\boldsymbol{W}|\boldsymbol{Z}, \beta) &= \frac{\Gamma(\beta V)^K}{\Gamma(\beta)^{VK}} \prod_{k=1}^K \frac{\prod_{v=1}^V \Gamma(N_{kv} + \beta)}{\Gamma(N_k + \beta V)} \\
\Rightarrow p(w_{dn}, \boldsymbol{W}_{\backslash dn}|z_{dn} = k, \boldsymbol{Z}_{\backslash dn}, \beta) &= \frac{\Gamma(\beta V)}{\Gamma(\beta)^V} \frac{\Gamma(N_{kw_{dn}} + \beta) \prod_{v \neq w_{dn}} \Gamma(N_{kv} + \beta)}{\Gamma(N_k + \beta V)} \\
\Rightarrow p(\boldsymbol{W}_{\backslash dn}|z_{dn} = k, \boldsymbol{Z}_{\backslash dn}, \beta) &= \frac{\Gamma(\beta V)}{\Gamma(\beta) V} \frac{\prod_{v=1}^V \Gamma(N_{kv} + \beta)}{\Gamma(N_k + \beta V)}
\end{aligned}$$

ここではトピック k についてのみ扱うので、$\prod_{k=1}^K$ は不要である (たぶん)。また、分子の 1~V までの掛け合わせから $v = w_{dn}$ である項を取り出しておく。

　ここで、$N_{kw_{dn}}$ は語彙 w_{dn}(ここでは文書 d の n 番目の単語と同じ単語のことで、つまり語彙 v と同じニュアンス) の内トピック k が割り当てられた数である。文書 d の n 番目の単語のトピック z_{dn} が k のとき、文書 d の n 番目の単語以外のトピック k が割り当てられた語彙数 $N_{k\backslash dn}$ は $N_{kv} - 1$ である。しかし

z_{dn} が k 以外のトピックであれば、N_{kv} に文書 d の n 番目の単語の分のカウントはそもそも含まれていないため、$N_{kv\backslash dn} = N_{kv}$ である。これをまとめると

$$N_{kv\backslash dn} = \begin{cases} N_{kv} - 1 & \text{if} \quad z_{dn} = k \\ N_{kv} & \text{if} \quad z_{dn} \neq k \end{cases}$$

である。これを N_{kv} について解くと

$$N_{kv} = \begin{cases} N_{kw_{dn}\backslash dn} + 1 & \text{if} \quad z_{dn} = k \\ N_{kv\backslash dn} & \text{if} \quad z_{dn} \neq k \end{cases}$$

となる。これを代入する。また、分子 $p(w_{dn}, \boldsymbol{W}_{\backslash dn} | z_{dn} = k, \boldsymbol{Z}_{\backslash dn}, \beta)$ の分母である文書 d の n 番目の単語以外のトピック k が割り当てられた総単語数 $N_{k\backslash dn}$ に、w_{dn} の分のカウントを 1 つ加えて $N_{k\backslash dn} + 1$ になる。

- (ii)：ガンマ関数の性質 $\Gamma(x+1) = (x)\Gamma(x)$ を用いて変形する。
- (iii)：約分して式を整理する。ただし、$\Gamma(N_{kw_{dn}\backslash dn} + \beta) \prod_{v \neq w_{dn}} \Gamma(N_{kw_{dn}\backslash dn} + \beta) = \prod_{v=1}^{V} \Gamma(N_{kv\backslash dn} + \beta)$ である。

となる。式 (4.13) と (4.14) を式 (4.12) に代入して

$$p(z_{dn} = k | \boldsymbol{W}, \boldsymbol{Z}_{\backslash dn}, \alpha, \beta) \propto (N_{dk\backslash dn} + \alpha) \frac{N_{kw_{dn}\backslash dn} + \beta}{N_{k\backslash dn} + \beta V} \tag{4.15}$$

が得られる。式 (4.13) の分母は k に依存しないため省いている。

1 つ目の因子は、n 番目の単語を除いた文書 d でトピック k が割り当てられた単語数を表しており、その文書で多く割り当てられているトピックになりやすいことが分かる。また 2 つ目の因子は、語彙 w_{dn} がトピック k に割り当てられた数に比例していて、語彙 w_{dn} が全文書集合で割り当てられているトピックになりやすいことが分かる。

・ハイパーパラメータの更新式

ハイパーパラメータ α, β は、周辺同時尤度 (4.9) に不動点反復法を用いて最大化することで求める。

・**Tips**

$$
\begin{aligned}
\frac{\Gamma(x)}{\Gamma(n+x)} &\geq \frac{\Gamma(\tilde{x})\exp\{(\tilde{x}-x)b\}}{\Gamma(n+\tilde{x})} \\
b &= \Psi(n+\tilde{x}) - \Psi(\tilde{x})
\end{aligned}
$$

$$
\begin{aligned}
\frac{\Gamma(n+x)}{\Gamma(x)} &\geq cx^a \\
a &= \Big(\Psi(n+\tilde{x}) - \Psi(\tilde{x})\Big)\tilde{x} \\
c &= \frac{\Gamma(n+\tilde{x})}{\Gamma(\tilde{x})}\tilde{x}^{-a}
\end{aligned}
$$

この関係性を用いる。詳細は 2.7 節にて。

式 (4.9) に式 (4.10') と (4.11') を代入する。そして、$\frac{\Gamma(\alpha K)}{\Gamma(N_d+\alpha K)}$ と $\frac{\Gamma(\beta V)}{\Gamma(N_k+\beta V)}$ を $\frac{\Gamma(x)}{\Gamma(n+x)}$、$\frac{\Gamma(N_{dk}+\alpha)}{\Gamma(\alpha)}$ と $\frac{\Gamma(N_{kv}+\beta)}{\Gamma(\beta)}$ を $\frac{\Gamma(n+x)}{\Gamma(x)}$ に対応させて、式を不動点反復法行えるように変形する。

$$
\begin{aligned}
p(\boldsymbol{Z},\boldsymbol{W}|\alpha,\beta) &= \prod_{d=1}^{D}\frac{\Gamma(\alpha K)}{\Gamma(N_d+\alpha K)}\prod_{d=1}^{D}\prod_{k=1}^{K}\frac{\Gamma(N_{dk}+\alpha)}{\Gamma(\alpha)}\prod_{k=1}^{K}\frac{\Gamma(\beta V)}{\Gamma(N_k+\beta V)}\prod_{k=1}^{K}\prod_{v=1}^{V}\frac{\Gamma(N_{kv}+\beta)}{\Gamma(\beta)} \\
&\geq \prod_{d=1}^{D}\frac{\Gamma(\hat{\alpha}K)\exp\{(\hat{\alpha}K-\alpha K)b_1\}}{\Gamma(N_d+\hat{\alpha}K)}\prod_{d=1}^{D}\prod_{k=1}^{K}\frac{\Gamma(N_{dk}+\hat{\alpha})}{\Gamma(\hat{\alpha})}\hat{\alpha}^{-a_1}\alpha^{a_1} \\
&\quad * \prod_{k=1}^{K}\frac{\Gamma(\hat{\beta}V)\exp\{(\hat{\beta}V-\beta V)b_2\}}{\Gamma(N_k+\hat{\beta}V)}\prod_{k=1}^{K}\prod_{v=1}^{V}\frac{\Gamma(N_{kv}+\hat{\beta})}{\Gamma(\hat{\beta})}\hat{\beta}^{-a_2}\beta^{a_2} = F
\end{aligned}
$$

ここで

$$
\begin{aligned}
a_1 &= \Big(\Psi(N_{dk}+\hat{\alpha}) - \Psi(\hat{\alpha})\Big)\hat{\alpha} \\
b_1 &= \Psi(N_d+\hat{\alpha}K) - \Psi(\hat{\alpha}K) \\
a_2 &= \Big(\Psi(N_{kv}+\hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta} \\
b_2 &= \Psi(N_k+\hat{\beta}V) - \Psi(\hat{\beta}V)
\end{aligned}
$$

である。ここから α に関する項を取り出して対数をとると

$$
\begin{aligned}
F(\alpha) &= \sum_{d=1}^{D}\Big\{\log\Gamma(\hat{\alpha}K) + (\hat{\alpha}K-\alpha K)\Big(\Psi(N_d+\hat{\alpha}K) - \Psi(\hat{\alpha}K)\Big) - \log\Gamma(N_d+\hat{\alpha}K)\Big\} \\
&\quad + \sum_{d=1}^{D}\sum_{k=1}^{K}\Big\{\log\Gamma(N_{dk}+\hat{\alpha}) - \log\Gamma(\hat{\alpha}) - \Big(\Psi(N_{dk}+\hat{\alpha}) - \Psi(\hat{\alpha})\Big)\hat{\alpha}\log\hat{\alpha} + \Big(\Psi(N_{dk}+\hat{\alpha}) - \Psi(\hat{\alpha})\Big)\hat{\alpha}\log\alpha\Big\}
\end{aligned}
$$

となる。これを α に関して微分し、$\frac{\partial F(\alpha)}{\partial\alpha} = 0$ となる α を求める。

$$
\frac{\partial F(\alpha)}{\partial \alpha} = -\sum_{d=1}^{D} K\Big(\Psi(N_d + \hat{\alpha}) - \Psi(\hat{\alpha})\Big) + \sum_{d=1}^{D}\sum_{k=1}^{K} \frac{1}{\alpha}\Big(\Psi(N_{dk} + \hat{\alpha}) - \Psi(\hat{\alpha})\Big)\hat{\alpha} = 0
$$

$$
\alpha = \hat{\alpha}\frac{\sum_{d=1}^{D}\sum_{k=1}^{K}\Psi(N_{dk} + \hat{\alpha}) - DK\Psi(\hat{\alpha})}{K\sum_{d=1}^{D}\Psi(N_d + \hat{\alpha}K) - DK\Psi(\hat{\alpha}K)}
$$

この式の $\hat{\alpha}$ を現ステップの α、左辺の α を更新後 (次ステップ) のパラメータ α^{new} とした次の式が更新式となる。

$$
\alpha^{\mathrm{new}} = \alpha\frac{\sum_{d=1}^{D}\sum_{k=1}^{K}\Psi(N_{dk} + \alpha) - DK\Psi(\alpha)}{K\sum_{d=1}^{D}\Psi(N_d + \alpha K) - DK\Psi(\alpha K)} \tag{4.16}
$$

β の更新についても同様に、F から β に関する項を取り出して対数をとり

$$
\begin{aligned}
F(\beta) = \sum_{k=1}^{K}&\Big\{\log\Gamma(\hat{\beta}V) + (\hat{\beta}V - \beta V)\Big(\Psi(N_k + \hat{\beta}V) - \Psi(\hat{\beta}V)\Big) - \log\Gamma(N_k + \hat{\beta}V)\Big\} \\
&+ \sum_{k=1}^{K}\sum_{v=1}^{V}\Big\{\log\Gamma(N_{kv} + \hat{\beta}) - \log\Gamma(\hat{\beta}) - \Big(\Psi(N_{kv} + \hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\log\hat{\beta} + \Big(\Psi(N_{kv} + \hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta}\log\beta\Big\}
\end{aligned}
$$

これを β に関して微分し、$\frac{\partial F(\beta)}{\partial \beta} = 0$ となる β を求める。

$$
\frac{\partial F(\beta)}{\partial \beta} = -\sum_{k=1}^{K} V\Big(\Psi(N_k + \hat{\beta}V) - \Psi(\hat{\beta}V)\Big) + \sum_{k=1}^{K}\sum_{v=1}^{V} \frac{1}{\beta}\Big(\Psi(N_{kv} + \hat{\beta}) - \Psi(\hat{\beta})\Big)\hat{\beta} = 0
$$

$$
\beta = \hat{\beta}\frac{\sum_{k=1}^{K}\sum_{v=1}^{V}\Psi(N_{kv} + \hat{\beta}) - KV\Psi(\hat{\beta})}{V\sum_{k=1}^{K}\Psi(N_k + \hat{\beta}V) - KV\Psi(\hat{\beta}V)}
$$

この式の $\hat{\beta}$ を現ステップの β、左辺の β を更新後 (次ステップ) のパラメータ β^{new} とした次の式が更新式となる。

$$
\beta^{\mathrm{new}} = \beta\frac{\sum_{k=1}^{K}\sum_{v=1}^{V}\Psi(N_{kv} + \beta) - KV\Psi(\beta)}{V\sum_{k=1}^{K}\Psi(N_k + \beta V) - KV\Psi(\beta V)} \tag{4.17}
$$

ここまでは、ディリクレ事前分布のパラメータが全て一様の場合を考えてきた。次は一様でない場合の更新式を求めていく。

・ハイパーパラメータが一様でない場合

ディリクレ事前分布が一様でないパラメータ $\boldsymbol{\alpha} = (\alpha_1, \cdots, \alpha_k, \cdots, \alpha_K)$ と $\boldsymbol{\beta} = (\beta_1, \cdots, \beta_v, \cdots, \beta_V)$ をそれぞれ持つとする。これを一様の場合と同様の手順で更新式を求めていく。

・パラメータの周辺化

文書集合 $\boldsymbol{W}$ とトピック集合 $\boldsymbol{Z}$ の同時分布を生成過程より

$$p(\boldsymbol{Z},\boldsymbol{W}|\boldsymbol{\alpha},\boldsymbol{\beta}) = p(\boldsymbol{Z}|\boldsymbol{\alpha})p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\beta}) \tag{1}$$

と分割する。右辺の 2 つの因子はそれぞれ

$$\begin{aligned}
p(\boldsymbol{Z}|\boldsymbol{\alpha}) &= \int p(\boldsymbol{Z}|\boldsymbol{\Theta})p(\boldsymbol{\Theta}|\boldsymbol{\alpha})d\boldsymbol{\Theta} \\
&= \int \prod_{d=1}^{D}\prod_{k=1}^{K}\theta_{dk}^{N_{dk}} \prod_{d=1}^{D}\frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\prod_{k=1}^{K}\Gamma(\alpha_k)}\prod_{k=1}^{K}\theta_{dk}^{\alpha_k-1}d\boldsymbol{\theta}_d \\
&= \frac{\Gamma(\sum_{k=1}^{K}\alpha_k)^D}{\prod_{k=1}^{K}\Gamma(\alpha_k)^D}\int\prod_{d=1}^{D}\prod_{k=1}^{K}\theta_{dk}^{N_{dk}+\alpha_k-1}d\boldsymbol{\theta}_d \\
&= \frac{\Gamma(\sum_{k=1}^{K}\alpha_k)^D}{\prod_{k=1}^{K}\Gamma(\alpha_k)^D}\prod_{d=1}^{D}\frac{\prod_{k=1}^{K}\Gamma(N_{dk}+\alpha_k)}{\Gamma(N_d+\sum_{k=1}^{K}\alpha_k)} &(2) \\
&= \prod_{d=1}^{D}\frac{\Gamma(\sum_{k=1}^{K}\alpha_k)}{\Gamma(N_d+\sum_{k=1}^{K}\alpha_k)}\prod_{d=1}^{D}\prod_{k=1}^{K}\frac{\Gamma(N_{dk}+\alpha_k)}{\Gamma(\alpha_k)} &(2')
\end{aligned}$$

$$\begin{aligned}
p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\beta}) &= \int p(\boldsymbol{W}|\boldsymbol{Z},\boldsymbol{\Phi})p(\boldsymbol{\Phi}|\boldsymbol{\beta})d\boldsymbol{\Phi} \\
&= \int \prod_{k=1}^{K}\prod_{v=1}^{V}\phi_{kv}^{N_{kv}} \prod_{k=1}^{K}\frac{\Gamma(\sum_{v=1}^{V}\beta)}{\prod_{v=1}^{V}\Gamma(\beta_v)}\prod_{v=1}^{V}\phi_{kv}^{\beta_v-1}d\boldsymbol{\phi}_K \\
&= \frac{\Gamma(\sum_{v=1}^{V}\beta_v)^K}{\prod_{v=1}^{V}\Gamma(\beta_v)^K}\int\prod_{k=1}^{K}\prod_{v=1}^{V}\phi_{kv}^{N_{kv}+\beta_v-1}d\boldsymbol{\phi}_K \\
&= \frac{\Gamma(\sum_{v=1}^{V}\beta_v)^K}{\prod_{v=1}^{V}\Gamma(\beta_v)^K}\prod_{k=1}^{K}\frac{\prod_{v=1}^{V}\Gamma(N_{kv}+\beta_v)}{\Gamma(N_k+\sum_{v=1}^{V}\beta_v)} &(3) \\
&= \prod_{k=1}^{K}\frac{\Gamma(\sum_{v=1}^{V}\beta_v)}{\Gamma(N_k+\sum_{v=1}^{V}\beta_v)}\prod_{k=1}^{K}\prod_{v=1}^{V}\frac{\Gamma(N_{kv}+\beta_v)}{\Gamma(\beta_v)} &(3')
\end{aligned}$$

になる。

・サンプリング式

サンプリングの確率は

$$p(z_{dn}=k|\boldsymbol{W},\boldsymbol{Z}_{\backslash dn},\boldsymbol{\alpha},\boldsymbol{\beta}) \propto p(z_{dn}=k|\boldsymbol{Z}_{\backslash dn},\boldsymbol{\alpha})p(w_{dn}|\boldsymbol{W}_{\backslash dn},z_{dn}=k,\boldsymbol{Z}_{\backslash dn},\boldsymbol{\beta}) \tag{4}$$

となる。左辺の 2 つの因子は、それぞれ式 (2),(3) を用いて

$$
\begin{aligned}
&p(z_{dn}=k|\boldsymbol{Z}_{\backslash dn},\boldsymbol{\alpha})\\
&=\frac{p(z_{dn}=k,\boldsymbol{Z}_{\backslash dn}|\boldsymbol{\alpha})}{p(\boldsymbol{Z}_{\backslash dn}|\boldsymbol{\alpha})}\\
&=\frac{\Gamma(\sum_{k=1}^K\alpha_k)}{\prod_{k=1}^K\Gamma(\alpha_k)}\frac{\Gamma(N_{dk\backslash dn}+1+\alpha_k)\prod_{k'\neq k}\Gamma(N_{dk'\backslash dn}+\alpha_k)}{\Gamma(N_d+\sum_{k=1}^K\alpha_k)}\frac{\prod_{k=1}^K\Gamma(\alpha_k)}{\Gamma(\sum_{k=1}^K\alpha_k)}\frac{\Gamma(N_d-1+\sum_{k=1}^K\alpha_k)}{\prod_{k'=1}^K\Gamma(N_{dk'\backslash dn}+\alpha_k)}\\
&=\frac{(N_{dk\backslash dn}+\alpha_k)\Gamma(N_{dk\backslash dn}+\alpha_k)\prod_{k'\neq k}\Gamma(N_{dk'\backslash dn}+\alpha_{k'})}{(N_d-1+\sum_{k=1}^K\alpha_k)\Gamma(N_d-1+\sum_{k=1}^K\alpha_k)}\frac{\Gamma(N_d-1+\sum_{k=1}^K\alpha_k)}{\prod_{k'=1}^K\Gamma(N_{dk'\backslash dn}+\alpha_k)}\\
&=\frac{N_{dk\backslash dn}+\alpha_k}{N_d-1+\sum_{k'=1}^K\alpha_{k'}}
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
&p(w_{dn}|\boldsymbol{W}_{\backslash dn},z_{dn}=k,\boldsymbol{Z}_{\backslash dn},\boldsymbol{\beta})\\
&=\frac{p(w_{dn},\boldsymbol{W}_{\backslash dn}|z_{dn}=k,\boldsymbol{Z}_{\backslash dn},\boldsymbol{\beta})}{p(\boldsymbol{W}_{\backslash dn}|z_{dn}=k,\boldsymbol{Z}_{\backslash dn},\boldsymbol{\beta})}\\
&=\frac{\Gamma(\sum_{v=1}^V\beta_v)}{\prod_{v=1}^V\Gamma(\beta_v)}\frac{\Gamma(N_{kw_{dn}\backslash dn}+1+\beta_v)\prod_{v'\neq w_{dn}}\Gamma(N_{kv'\backslash dn}+\beta_{v'})}{\Gamma(N_{k\backslash dn}+1+\sum_{v=1}^V\beta_v)}\frac{\prod_{v=1}^V\Gamma(\beta_v)}{\Gamma(\sum_{v=1}^V\beta_v)}\frac{\Gamma(N_{k\backslash dn}+\sum_{v=1}^V\beta_v)}{\prod_{v=1}^V\Gamma(N_{kv\backslash dn}+\beta_v)}\\
&=\frac{(N_{kw_{dn}\backslash dn}+\beta_v)\Gamma(N_{kw_{dn}\backslash dn}+\beta_v)\prod_{v'\neq w_{dn}}\Gamma(N_{kv'\backslash dn}+\beta_{v'})}{(N_{k\backslash dn}+\sum_{v=1}^V\beta_v)\Gamma(N_{k\backslash dn}+\sum_{v=1}^V\beta_v)}\frac{\Gamma(N_{k\backslash dn}+\sum_{v=1}^V\beta_v)}{\prod_{v=1}^V\Gamma(N_{kv\backslash dn}+\beta_v)}\\
&=\frac{N_{kw_{dn}\backslash dn}+\beta_v}{N_{k\backslash dn}+\sum_{v'=1}^V\beta_{v'}}
\end{aligned}
\tag{6}
$$

になる。式 (5),(6) を式 (4) に代入して

$$
p(z_{dn}=k|\boldsymbol{W},\boldsymbol{Z}_{\backslash dn},\boldsymbol{\alpha},\boldsymbol{\beta})\propto(N_{dk\backslash dn}+\alpha_k)\frac{N_{kw_{dn}\backslash dn}+\beta_v}{N_{k\backslash dn}+\sum_{v'=1}^V\beta_{v'}}
$$

が得られる。式 (5) の分母は k に依存しないため省いている。

・不動点反復法

式 (2'),(3') を式 (1) に代入する。その式を Tips の関係性を用いて不動点反復法を行えるように変形する。

$$
\begin{aligned}
p(\boldsymbol{Z},\boldsymbol{W}|\boldsymbol{\alpha},\boldsymbol{\beta})&=\prod_{d=1}^D\frac{\Gamma(\sum_{k=1}^K\alpha_k)}{\Gamma(N_d+\sum_{k=1}^K\alpha_k)}\prod_{d=1}^D\prod_{k=1}^K\frac{\Gamma(N_{dk}+\alpha_k)}{\Gamma(\alpha_k)}\prod_{k=1}^K\frac{\Gamma(\sum_{v=1}^V\beta_v)}{\Gamma(N_k+\sum_{v=1}^V\beta_v)}\prod_{k=1}^K\prod_{v=1}^V\frac{\Gamma(N_{kv}+\beta_v)}{\Gamma(\beta_v)}\\
&\geq\prod_{d=1}^D\frac{\Gamma(\sum_{k=1}^K\hat{\alpha}_k)\exp\{(\sum_{k=1}^K\hat{\alpha}_k-\alpha_k)b_1\}}{\Gamma(N_d+\sum_{k=1}^K\hat{\alpha}_k)}\prod_{d=1}^D\prod_{k=1}^K\frac{\Gamma(N_{dk}+\hat{\alpha}_k)}{\Gamma(\hat{\alpha}_k)}\hat{\alpha}_k^{-a_1}\alpha_k^{a_1}\\
&\quad*\prod_{k=1}^K\frac{\Gamma(\sum_{v=1}^V\hat{\beta}_v)\exp\{(\sum_{v=1}^V\hat{\beta}_v-\beta_v)b_2\}}{\Gamma(N_k+\sum_{v=1}^V\hat{\beta}_v)}\prod_{k=1}^K\prod_{v=1}^V\frac{\Gamma(N_{kv}+\hat{\beta}_v)}{\Gamma(\hat{\beta}_v)}\hat{\beta}_v^{-a_2}\beta_v^{a_2}=F
\end{aligned}
$$

ここで

$$
\begin{aligned}
a_1 &= \Big(\Psi(N_{dk}+\hat{\alpha}_k)-\Psi(\hat{\alpha}_k)\Big)\hat{\alpha}_k \\
b_1 &= \Psi\left(N_d+\sum_{k=1}^{K}\hat{\alpha}_k\right)-\Psi\left(\sum_{k=1}^{K}\hat{\alpha}_k\right) \\
a_2 &= \Big(\Psi(N_{kv}+\hat{\beta}_v)-\Psi(\hat{\beta}_v)\Big)\hat{\beta}_v \\
b_2 &= \Psi\left(N_k+\sum_{v=1}^{V}\hat{\beta}_v\right)-\Psi\left(\sum_{v=1}^{V}\hat{\beta}_v\right)
\end{aligned}
$$

である。ここから α_k に関する項を取り出して、対数をとった式を α_k に関して微分し $\frac{\partial F(\alpha_k)}{\partial \alpha_k}=0$ となる α_k を求める。

$$
\begin{aligned}
F(\alpha_k) =& \sum_{d=1}^{D}\left\{\log\Gamma\Big(\sum_{k=1}^{K}\hat{\alpha}_k\Big)+\Big(\sum_{k=1}^{K}\hat{\alpha}_k-\alpha_k\Big)\left(\Psi\Big(N_d+\sum_{k=1}^{K}\hat{\alpha}_k\Big)-\Psi\Big(\sum_{k=1}^{K}\hat{\alpha}_k\Big)\right)-\log\Gamma\Big(N_d+\sum_{k=1}^{K}\hat{\alpha}_k\Big)\right\} \\
&+\sum_{d=1}^{D}\sum_{k=1}^{K}\Big\{\log\Gamma(N_{dk}+\hat{\alpha}_k)-\log\Gamma(\hat{\alpha}_k)-\Big(\Psi(N_{dk}+\hat{\alpha}_k)-\Psi(\hat{\alpha}_k)\Big)\hat{\alpha}_k\log\hat{\alpha}_k \\
&+\Big(\Psi(N_{dk}+\hat{\alpha}_k)-\Psi(\hat{\alpha}_k)\Big)\hat{\alpha}_k\log\alpha_k\Big\} \\
\frac{\partial F(\alpha_k)}{\partial \alpha_k} =& -\sum_{d=1}^{D}\left(\Psi\Big(N_d+\sum_{k'=1}^{K}\hat{\alpha}_{k'}\Big)-\Psi\Big(\sum_{k'=1}^{K}\hat{\alpha}_{k'}\Big)\right)+\sum_{d=1}^{D}\frac{1}{\alpha_k}\Big(\Psi(N_{dk}+\hat{\alpha}_k)-\Psi(\hat{\alpha}_k)\Big)\hat{\alpha}_k=0 \\
\alpha_k =& \hat{\alpha}_k\frac{\sum_{d=1}^{D}\Psi(N_{dk}+\hat{\alpha}_k)-D\Psi(\hat{\alpha}_k)}{\sum_{d=1}^{D}\Psi(N_d+\sum_{k'=1}^{K}\hat{\alpha}_{k'})-D\Psi(\sum_{k'=1}^{K}\hat{\alpha}_{k'})}
\end{aligned}
$$

この式の $\hat{\alpha}$ を現ステップの α、左辺の α を更新後のパラメータ α^{new} とした次の式が更新式となる。

$$
\alpha_k^{\text{new}}=\alpha_k\frac{\sum_{d=1}^{D}\Psi(N_{dk}+\alpha_k)-D\Psi(\alpha_k)}{\sum_{d=1}^{D}\Psi(N_d+\sum_{k'=1}^{K}\alpha_{k'})-D\Psi(\sum_{k'=1}^{K}\alpha_{k'})}
$$

β_v^{new} についても同様に、まずは F から β_v に関する項を取り出して、対数をとった式を β_v に関して微分し $\frac{\partial F(\beta_v)}{\partial \beta_v}=0$ となる β_v を求める。

$$
\begin{aligned}
F(\beta_v) =& \sum_{k=1}^{K}\left\{\log\Gamma\Big(\sum_{v=1}^{V}\hat{\beta}_v\Big)+\Big(\sum_{v=1}^{V}\hat{\beta}_v-\beta_v\Big)\left(\Psi\Big(N_k+\sum_{v=1}^{V}\hat{\beta}_v\Big)-\Psi\Big(\sum_{v=1}^{V}\hat{\beta}_v\Big)\right)-\log\Gamma\Big(N_k+\sum_{v=1}^{V}\hat{\beta}_v\Big)\right\} \\
&+\sum_{k=1}^{K}\sum_{v=1}^{V}\Big\{\log\Gamma(N_{kv}+\hat{\beta}_v)-\log\Gamma(\hat{\beta}_v)-\Big(\Psi(N_{kv}+\hat{\beta}_v)-\Psi(\hat{\beta}_v)\Big)\hat{\beta}_v\log\hat{\beta}_v \\
&+\Big(\Psi(N_{kv}+\hat{\beta}_v)-\Psi(\hat{\beta}_v)\Big)\hat{\beta}_v\log\beta_v\Big\} \\
\frac{\partial F(\beta_v)}{\partial \beta_v} =& -\sum_{k=1}^{K}\left(\Psi\Big(N_k+\sum_{v'=1}^{V}\hat{\beta}_{v'}\Big)-\Psi\Big(\sum_{v'=1}^{V}\hat{\beta}_{v'}\Big)\right)+\sum_{k=1}^{K}\frac{1}{\beta_v}\Big(\Psi(N_{kv}+\hat{\beta}_v)-\Psi(\hat{\beta}_v)\Big)\hat{\beta}_v=0 \\
\beta =& \hat{\beta}_v\frac{\sum_{k=1}^{K}\Psi(N_{kv}+\hat{\beta}_v)-K\Psi(\hat{\beta}_v)}{\sum_{k=1}^{K}\Psi(N_k+\sum_{v'=1}^{V}\hat{\beta}_{v'})-K\Psi(\sum_{v'=1}^{V}\hat{\beta}_{v'})}
\end{aligned}
$$

この式の $\hat{\beta}$ を現ステップの β、左辺の β を更新後のパラメータ β^{new} とした次の式が更新式となる。

$$\beta^{\text{new}} = \beta_v \frac{\sum_{k=1}^{K} \Psi(N_{kv} + \beta_v) - K\Psi(\beta_v)}{\sum_{k=1}^{K} \Psi(N_k + \sum_{v'=1}^{V} \beta_{v'}) - K\Psi(\sum_{v'=1}^{V} \beta_{v'})}$$

・**R** で組んでみる

・コード全体

・テキスト処理

```r
# 利用パッケージ
library(RMeCab)
library(tidyverse)

### テキスト処理

# テキストファイルの保存先を指定する
file_path <- " フォルダ名"

# 抽出しない単語を指定
stop_words <- "[a-z]" # 小文字のアルファベットを含む語

# 形態素解析
mecab_df <- docDF(file_path, type = 1)

# 文書 d の語彙 v の出現回数 (N_dv) の集合
N_dv <- mecab_df %>%
        filter(grepl(" 名詞 | 形容詞 |^ 動詞", POS1)) %>% # 抽出する品詞 (大分類) を指定する
        filter(grepl(" 一般 |^ 自立", POS2)) %>%        # 抽出する品詞 (細分類) を指定する
        filter(!grepl(stop_words, TERM)) %>%          # ストップワードを除く
        select(-c(TERM, POS1, POS2)) %>%              # 数値列のみを残す
        filter(apply(., 1, sum) >= 5) %>%             # 抽出する総出現回数を指定する
        t()                                           # 転置

# 確認用の行列名
dimnames(N_dv) <- list(paste0("d=", 1:nrow(N_dv)), # 行名
                       paste0("v=", 1:ncol(N_dv))) # 列名

# 文書 d の単語数 (N_d) のベクトル
N_d <- apply(N_dv, 1, sum) # 行方向に和をとる

# 文書数 (D)
D <- nrow(N_dv)

# 総語彙数 (V)
V <- ncol(N_dv)
```

・パラメータの初期設定

```r
### パラメータの初期設定

# トピック数 (K)
K <- 4   # 任意の値を指定する
```

```r
# ハイパーパラメータ (alpha, beta)
alpha_k <- rep(2, K) # 任意の値を指定する
beta    <- 2         # 任意の値を指定する

# 文書 d の語彙 v に割り当てられたトピック (z_dn) の集合
z_dn <- array(0, dim = c(D, V, max(N_dv)),
              dimnames = list(paste0("d=", 1:D),
                              paste0("v=", 1:V),
                              paste0("N_dv=", 1:max(N_dv))))

# 文書 d においてトピック k が割り当てられた単語数 (N_dk) の集合
N_dk <- matrix(0, nrow = D, ncol = K,
               dimnames = list(paste0("d=", 1:D),
                               paste0("k=", 1:K)))

# 文書全体でトピック k が割り当てられた語彙 v の出現回数 (N_kv) の集合
N_kv <- matrix(0, nrow = K, ncol = V,
               dimnames = list(paste0("k=", 1:K),
                               paste0("v=", 1:V)))

# 全文書でトピック k が割り当てられた単語数 (N_k) のベクトル
N_k <- rep(0, K)
```

・ギブスサンプリング

```r
### ギブスサンプリング

# 受け皿の用意
p <- NULL

# 結果の確認用
trace_alpha <- as.matrix(alpha_k)
trace_beta  <- beta

for(i in 1:1000) { # 任意の回数を指定する

  ## 新たに割り当られたトピックに関するカウントを初期化
  new_N_dk <- matrix(0, nrow = D, ncol = K,
                     dimnames = list(paste0("d=", 1:D), paste0("k=", 1:K)))
  new_N_kv <- matrix(0, nrow = K, ncol = V,
                     dimnames = list(paste0("k=", 1:K), paste0("v=", 1:V)))
  new_N_k  <- rep(0, K)

  for(d in 1:D) { ## (各文書)

    for(v in 1:V) { ## (各語彙)

      if(N_dv[d, v] > 0) { ## (条件分岐：N_dv > 0)

        for(ndv in 1:N_dv[d, v]) { ## (各語彙の出現回数)

          # 現ステップの計算のためにカウントを移す
          tmp_N_dk <- N_dk
          tmp_N_kv <- N_kv
```

```r
        tmp_N_k  <- N_k

        if(z_dn[d, v, ndv] > 0) { # 初回を飛ばす処理

          # 前ステップで文書 d の語彙 v に割り当てられたトピックを `k`に代入
          k <- z_dn[d, v, ndv]

          # 文書 d の語彙 v の分のカウントを引く
          tmp_N_dk[d, k] <- N_dk[d, k] - 1
          tmp_N_kv[k, v] <- N_kv[k, v] - 1
          tmp_N_k[k]     <- N_k[k] - 1
        }

        for(k in 1:K) { ## (各トピック)

          # サンプリング確率を計算
          tmp_p_alpha      <- tmp_N_dk[d, k] + alpha_k[k]
          tmp_p_beta_numer <- tmp_N_kv[k, v] + beta
          tmp_p_beta_denom <- tmp_N_k[k] + beta * V
          p[k] <- tmp_p_alpha * tmp_p_beta_numer / tmp_p_beta_denom

        } ##  (/各トピック)

        # サンプリング
        tmp_z_dn <- rmultinom(n = 1, size = 1:K, prob = p)
        z_dn[d, v, ndv] <- which(tmp_z_dn == 1)

        # 新たに割り当てられたトピックを `k`に代入
        k <- z_dn[d, v, ndv]

        # 文書 d の語彙 v の分のカウントを加える
        new_N_dk[d, k] <- new_N_dk[d, k] + 1
        new_N_kv[k, v] <- new_N_kv[k, v] + 1
        new_N_k[k]     <- new_N_k[k] + 1

      } ## (/各語彙の出現回数)
    } ## (/条件分岐:N_dv > 0)
  } ## (/各語彙)
} ## (/各文書)

# トピック集合とカウントを更新
N_dk <- new_N_dk
N_kv <- new_N_kv
N_k  <- new_N_k

# ハイパーパラメータ (alpha) の更新
tmp_alpha_numer1 <- apply(digamma(t(N_dk) + alpha_k), 1, sum) # 分子
tmp_alpha_numer2 <- D * digamma(alpha_k)                      # 分子
tmp_alpha_denom1 <- sum(digamma(N_d + sum(alpha_k)))          # 分母
tmp_alpha_denom2 <- D * digamma(sum(alpha_k))                 # 分母
alpha_k <- alpha_k * (tmp_alpha_numer1 - tmp_alpha_numer2) / (tmp_alpha_denom1 - tmp_alpha_denom2)

# ハイパーパラメータ (beta) の更新
```

```
  tmp_beta_numer1 <- sum(digamma(N_kv + beta))         # 分子
  tmp_beta_numer2 <- K * V * digamma(beta)             # 分子
  tmp_beta_denom1 <- V * sum(digamma(N_k + beta * V)) # 分母
  tmp_beta_denom2 <- K * V * digamma(beta * V)         # 分母
  beta <- beta * (tmp_beta_numer1 - tmp_beta_numer2) / (tmp_beta_denom1 - tmp_beta_denom2)

  # 結果の確認用
  trace_alpha <- cbind(trace_alpha, as.matrix(alpha_k))
  trace_beta  <- c(trace_beta, beta)
}
```

・コードの解説

・利用パッケージ

```
# 利用パッケージ
library(RMeCab)
library(tidyverse)
```

形態素解析のための `RMeCab::docDF()` と、グラフ作成時の `tidyr::gather()`、`ggplot2` 関連の関数以外にライブラリの読み込みが必要なのは `dplyr` のみ。(面倒なので `tidyvers` パッケージを読み込んでいる)

その他テキスト処理の部分は 3.3 節を参照のこと。

・パラメータの初期設定

・トピック数

```
# トピック数 (K)
K <- 4    # 任意の値を指定する
```

任意のトピック数を指定する。

・ハイパーパラメータ

```
# ハイパーパラメータ (alpha, beta) : 一様
alpha <- 2 # 任意の値を指定する
beta  <- 2 # 任意の値を指定する

# ハイパーパラメータ (alpha, beta) : 多様
alpha_k <- rep(2, K) # 任意の値を指定する
beta_v  <- rep(2, V) # 任意の値を指定する
```

任意のハイパーパラメータの初期値を指定する。

各パラメータが一様であるか多様であるかを決める。一様を仮定するなら、上のスカラーを作成する。多様な値を取ると仮定するなら、下のベクトルを作成する。

ここでは参考書に従って、トピック分布のパラメータ α を多様に、単語分布のパラメータ β を一様とする。

・トピック集合

```
# 文書 d の語彙 v に割り当てられたトピック (z_dn) の集合
z_dn <- array(0, dim = c(D, V, max(N_dv)),
              dimnames = list(paste0("d=", 1:D),
                              paste0("v=", 1:V),
                              paste0("N_dv=", 1:max(N_dv))))
```

`z_dv` は、文書ごとの各単語に割り当てられたトピックの集合である。しかしここでは (~~参考書に逆らって~~)、単語レベル (n) ではなく語彙レベル (v) で扱い、D 行 V 列の行列としたい。そこで、`z_dv` を D 行 V 列の行列からなる配列とする。

トピックモデルでは、同じ単語であっても別のトピックが割り当てられる。そこで、配列の 3 次元目を出現回数とする。文書 d の語彙 v のトピックを 3 次元の方向に N_{dv}(出現回数) 個格納されていくことになる。そのため、3 次元方向の要素の数は出現回数の最大値 `max(N_dv)` を用意しておく。
初回はどの文書にもトピックが割り当てられていない状態として、全ての要素を 0 としておく。出現回数以上の要素については 0 のままで推定が行われる。

・トピックごとの各カウント

```
# 文書 d においてトピック k が割り当てられた単語数 (N_dk) の集合
N_dk <- matrix(0, nrow = D, ncol = K,
               dimnames = list(paste0("d=", 1:D),
                               paste0("k=", 1:K)))

# 文書全体でトピック k が割り当てられた語彙 v の出現回数 (N_kv) の集合
N_kv <- matrix(0, nrow = K, ncol = V,
               dimnames = list(paste0("k=", 1:K),
                               paste0("v=", 1:V)))

# 全文書でトピック k が割り当てられた単語数 (N_k) のベクトル
N_k <- rep(0, K)
```

`N_dk` は、文書 d においてトピック k が割り当てられた単語数の集合で、D 行 K 列の行列である。
`N_kv` は、語彙 v においてトピック k が割り当てられた語数の集合で、K 行 V 列の行列である
`N_k` は、トピック k が割り当てられてられた総単語数のベクトルで、要素が K 個である。

初回はどの文書にもトピックが割り当てられていないため、初期値は全て 0 である。

次からは、for ループ内の処理である。

・ギブスサンプリング

・次ステップのカウントの初期化

```r
## 新たに割り当られたトピックに関するカウントを初期化
new_N_dk <- matrix(0, nrow = D, ncol = K,
                   dimnames = list(paste0("d=", 1:D), paste0("k=", 1:K)))
new_N_kv <- matrix(0, nrow = K, ncol = V,
                   dimnames = list(paste0("k=", 1:K), paste0("v=", 1:V)))
new_N_k  <- rep(0, K)
```

new_N_dk、new_N_kv、new_N_k は、それぞれ現ステップでトピックを割り当られた単語に関するカウントを加算していく受け皿である。そのため、毎ステップの最初にカウントを 0 にリセットする。

次からは、各文書と各文書の語彙に対してループで処理を行う内容である。
for(d in 1:D) と for(v in 1:V) に加えて、for(ndv in 1:N_dv[d, v]) により各語彙の出現回数分繰り返し回す。それによって、単語レベルでトピックを割り当てることができる。ただし出現回数が 0 の語彙も含まれているため、if(N_dv[d, v] > 0) により、0 なら以降の処理を行わないようにしておく。

・現ステップのカウント

```r
# 現ステップの計算のためにカウントを移す
tmp_N_dk <- N_dk
tmp_N_kv <- N_kv
tmp_N_k  <- N_k

if(z_dn[d, v, ndv] > 0) { # 初回を飛ばす処理

  # 前ステップで文書 d の語彙 v に割り当てられたトピックを `k`に代入
  k <- z_dn[d, v, ndv]

  # 文書 d の語彙 v の分のカウントを引く
  tmp_N_dk[d, k] <- N_dk[d, k] - 1
  tmp_N_kv[k, v] <- N_kv[k, v] - 1
  tmp_N_k[k]     <- N_k[k] - 1
}
```

tmp_N_dk、tmp_N_kv、tmp_N_k は、それぞれ $N_{dk\backslash dn}N_{kv\backslash dn}N_{k\backslash dn}$ のことである。これらの変数は、現ステップのサンプリング確率の計算に用いる。
これらは、文書 d の n 番目の単語 (ここでは語彙 v) を除いたカウントである。つまり、それぞれ該当する要素から 1 を引けばよい。
そのため、まず前ステップのカウントをそれぞれ移して、それぞれ該当する要素を添え字で指定して、そこから 1 を引く。ただし、初回については割り当てられたトピックがないためカウントを引く必要がない。そのため、初回にの処理に関しては if() を使って、処理を行わない。

・サンプリング確率の計算

```r
for(k in 1:K) { ## (各トピック)

  # サンプリング確率を計算
  tmp_p_alpha      <- tmp_N_dk[d, k] + alpha_k[k]
  tmp_p_beta_numer <- tmp_N_kv[k, v] + beta
  tmp_p_beta_denom <- tmp_N_k[k] + beta * V
  p[k] <- tmp_p_alpha * tmp_p_beta_numer / tmp_p_beta_denom
```

```
} ## （/各トピック）
```

サンプリング確率の計算は次の式によって行う。

$$
p(z_{dn} = k | \boldsymbol{W}, \boldsymbol{Z}_{\backslash dn}, \alpha, \beta) \propto (N_{dk \backslash dn} + \alpha_k) \frac{N_{k w_{dn} \backslash dn} + \beta}{N_{k \backslash dn} + \beta V}
$$

ただし、ハイパーパラメータ α, β を一様とする場合は

$$
p(z_{dn} = k | \boldsymbol{W}, \boldsymbol{Z}_{\backslash dn}, \alpha, \beta) \propto (N_{dk \backslash dn} + \alpha) \frac{N_{k w_{dn} \backslash dn} + \beta}{N_{k \backslash dn} + \beta V} \tag{4.15}
$$

であり、また多様とする場合は

$$
p(z_{dn} = k | \boldsymbol{W}, \boldsymbol{Z}_{\backslash dn}, \boldsymbol{\alpha}, \boldsymbol{\beta}) \propto (N_{dk \backslash dn} + \alpha_k) \frac{N_{k w_{dn} \backslash dn} + \beta_v}{N_{k \backslash dn} + \sum_{v'=1}^{V} \beta_{v'}}
$$

の式になる。

この計算をトピックごとに 1 つずつ行い、添え字を使って計算結果を p に代入していくため、最初に空のオブジェクトとして p を作っておく。

・サンプリング

```
# サンプリング
tmp_z_dn <- rmultinom(n = 1, size = 1:K, prob = p)
z_dn[d, v, ndv] <- which(tmp_z_dn == 1)
```

上で求めた確率を基に、カテゴリ分布に従ってトピックを割り当てる。多項分布の乱数発生関数 rmultinom() の引数に n = 1、size = 1 を指定して試行回数を 1 回、サンプル数を 1 つとすることで、カテゴリ分布に従う乱数を発生させる。rmultinom() の結果は、サンプル確率を指定する引数 prob に渡す要素数の行、n に指定した試行回数の列のマトリクスで返ってくる。　この節の例では、$1, 2, \cdots, K$ の行を持つ 1 列のマトリクスが返ってくる。要素は列ごとに、割り当てられたトピックの行が 1 でそれ以外は 0 となる。

出力された K 行 1 列のマトリクスに which() を使い、要素が 1 の行番号を検索する。返ってきた値 (行番号) がトピック番号に相当するので、それを z_d の d 個目の要素に代入する。

・カウントの加算

```
# 新たに割り当てられたトピックを `k`に代入
k <- z_dn[d, v, ndv]

# 文書 d の語彙 v の分のカウントを加える
new_N_dk[d, k] <- new_N_dk[d, k] + 1
new_N_kv[k, v] <- new_N_kv[k, v] + 1
new_N_k[k]     <- new_N_k[k] + 1
```

カウントを除いたときと同様に、現ステップで割り当てられたトピックについて該当する要素に加えていく。ループ処理の中で全ての単語分が加算される。

ここまでが、各文書・各語彙についてのループ処理の内容であった。次からは、この結果を基にパラメータを更新していく。

・カウントの更新

```
# トピック集合とカウントを更新
N_dk <- new_N_dk
N_kv <- new_N_kv
N_k  <- new_N_k
```

全ての単語に関するカウントを集計できたので、その結果をそれぞれ N_dk、N_kv、N_k に上書きする。

・ハイパーパラメータ α の更新

```
# ハイパーパラメータ (alpha) の更新：一様
tmp_alpha_numer1 <- sum(digamma(N_dk + alpha))        # 分子
tmp_alpha_numer2 <- D * K * digamma(alpha)            # 分子
tmp_alpha_denom1 <- K * sum(digamma(N_d + alpha * K)) # 分母
tmp_alpha_denom2 <- D * K * digamma(alpha * K)        # 分母
alpha <- alpha * (tmp_alpha_numer1 - tmp_alpha_numer2) / (tmp_alpha_denom1 - tmp_alpha_denom2)


# ハイパーパラメータ (alpha) の更新：多様
tmp_alpha_numer1 <- apply(digamma(t(N_dk) + alpha_k), 1, sum) # 分子
tmp_alpha_numer2 <- D * digamma(alpha_k)                      # 分子
tmp_alpha_denom1 <- sum(digamma(N_d + sum(alpha_k)))          # 分母
tmp_alpha_denom2 <- D * digamma(sum(alpha_k))                 # 分母
alpha_k <- alpha_k * (tmp_alpha_numer1 - tmp_alpha_numer2) / (tmp_alpha_denom1 - tmp_alpha_denom2)
```

α の更新式 (一様) は

$$\alpha^{\text{new}} = \alpha \frac{\sum_{d=1}^{D} \sum_{k=1}^{K} \Psi(N_{dk} + \alpha) - DK\Psi(\alpha)}{K \sum_{d=1}^{D} \Psi(N_d + \alpha K) - DK\Psi(\alpha K)} \tag{4.16}$$

である。また多様の場合は

$$\alpha_k^{\text{new}} = \alpha_k \frac{\sum_{d=1}^{D} \Psi(N_{dk} + \alpha_k) - D\Psi(\alpha_k)}{\sum_{d=1}^{D} \Psi(N_d + \sum_{k'=1}^{K} \alpha_{k'}) - D\Psi(\sum_{k'=1}^{K} \alpha_{k'})}$$

である。

分かりやすくするために、各項を分けて計算した上で、最後にまとめて計算を行う。

・ハイパーパラメータ β の更新

```
# ハイパーパラメータ (beta) の更新：一様
tmp_beta_numer1 <- sum(digamma(N_kv + beta))        # 分子
tmp_beta_numer2 <- K * V * digamma(beta)            # 分子
tmp_beta_denom1 <- V * sum(digamma(N_k + beta * V)) # 分母
```

```r
tmp_beta_denom2 <- K * V * digamma(beta * V)        # 分母
beta <- beta * (tmp_beta_numer1 - tmp_beta_numer2) / (tmp_beta_denom1 - tmp_beta_denom2)


# ハイパーパラメータ (beta) の更新：多様
tmp_beta_numer1 <- apply(digamma(t(N_kv) + beta_v), 1, sum) # 分子
tmp_beta_numer2 <- K * digamma(beta_v)                      # 分子
tmp_beta_denom1 <- sum(digamma(N_k + sum(beta_v)))          # 分母
tmp_beta_denom2 <- K * digamma(sum(beta_v))                 # 分母
beta_v <- beta_v * (tmp_beta_numer1 - tmp_beta_numer2) / (tmp_beta_denom1 - tmp_beta_denom2)
```

β の更新式 (一様) は

$$\beta^{\text{new}} = \beta \frac{\sum_{k=1}^{K} \sum_{v=1}^{V} \Psi(N_{kv} + \beta) - KV\Psi(\beta)}{V \sum_{k=1}^{K} \Psi(N_k + \beta V) - KV\Psi(\beta V)} \tag{4.17}$$

である。また多様の場合は

$$\beta^{\text{new}} = \beta_v \frac{\sum_{k=1}^{K} \Psi(N_{kv} + \beta_v) - K\Psi(\beta_v)}{\sum_{k=1}^{K} \Psi(N_k + \sum_{v'=1}^{V} \beta_{v'}) - K\Psi(\sum_{v'=1}^{V} \beta_{v'})}$$

である。

こちらも同様に、各項分けて計算する。

・推定結果の確認

・ハイパーパラメータの推定結果の確認

```r
## トピック分布の推定結果の確認
# データフレームを作成
alpha_df <- data.frame(topic = as.factor(1:K),
                       alpha = alpha_k)

# 描画
ggplot(data = alpha_df, mapping = aes(x = topic, y = alpha, fill = topic)) +  # データ
  geom_bar(stat = "identity", position = "dodge") +  # 折れ線グラフ
  labs(title = "LDA:Gibbs Sampling")                 # タイトル
```

・パラメータ推定結果 (平均値) の確認

```r
## トピック分布 (平均値) の確認
# データフレームを作成
theta_df <- data.frame(topic = as.factor(1:K),
                       prob = alpha_k / sum(alpha_k))

# 描画
ggplot(data = theta_df, mapping = aes(x = topic, y = prob, fill = topic)) +  # データ
```

```
  geom_bar(stat = "identity", position = "dodge") +  # 折れ線グラフ
  labs(title = "LDA:Gibbs Sampling")                 # タイトル
```

・**推定推移の確認**

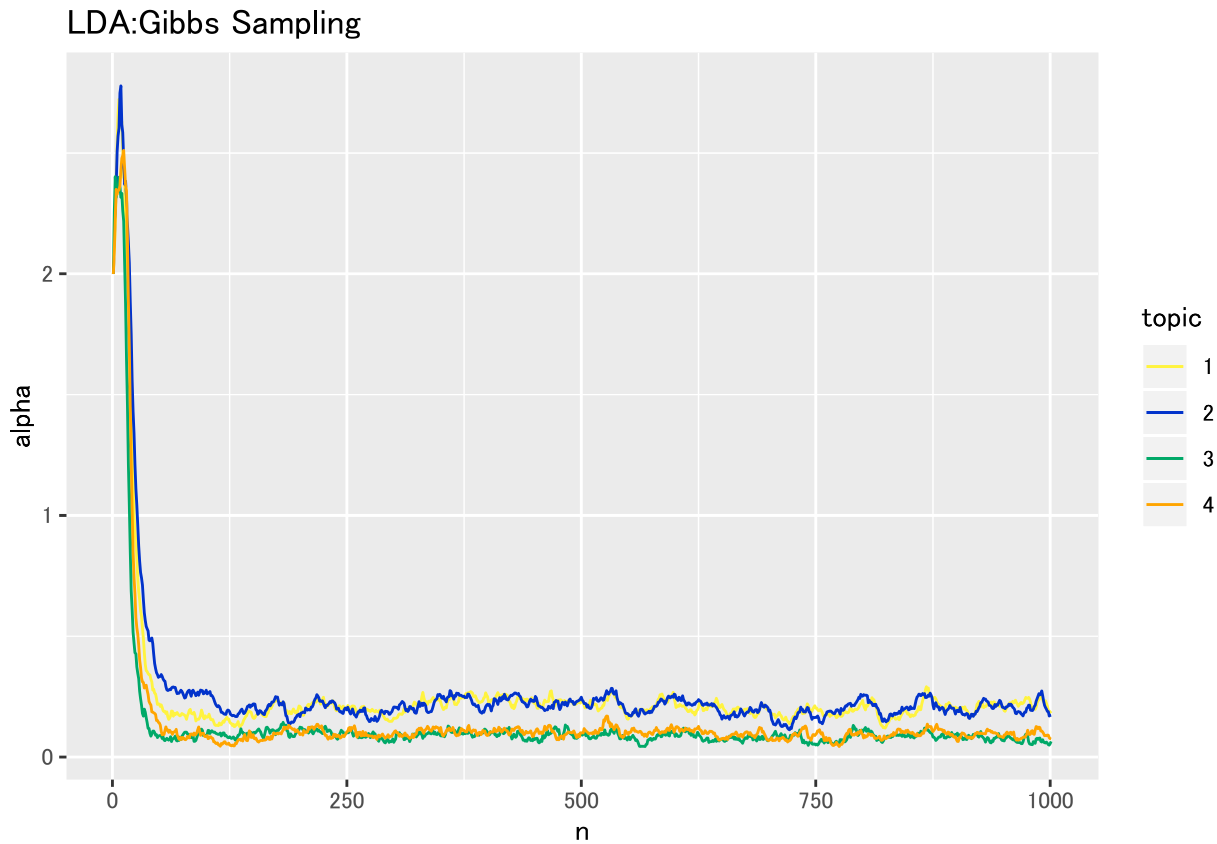
・推移の記録

```
# 結果の確認用（初期値）
trace_alpha <- as.matrix(alpha_k)
trace_beta  <- beta


# 結果の確認用（更新後）
trace_alpha <- cbind(trace_alpha, as.matrix(alpha_k))
trace_beta  <- c(trace_beta, beta)
```

推移を確認するために、ハイパーパラメータを更新する度に記録しておく。

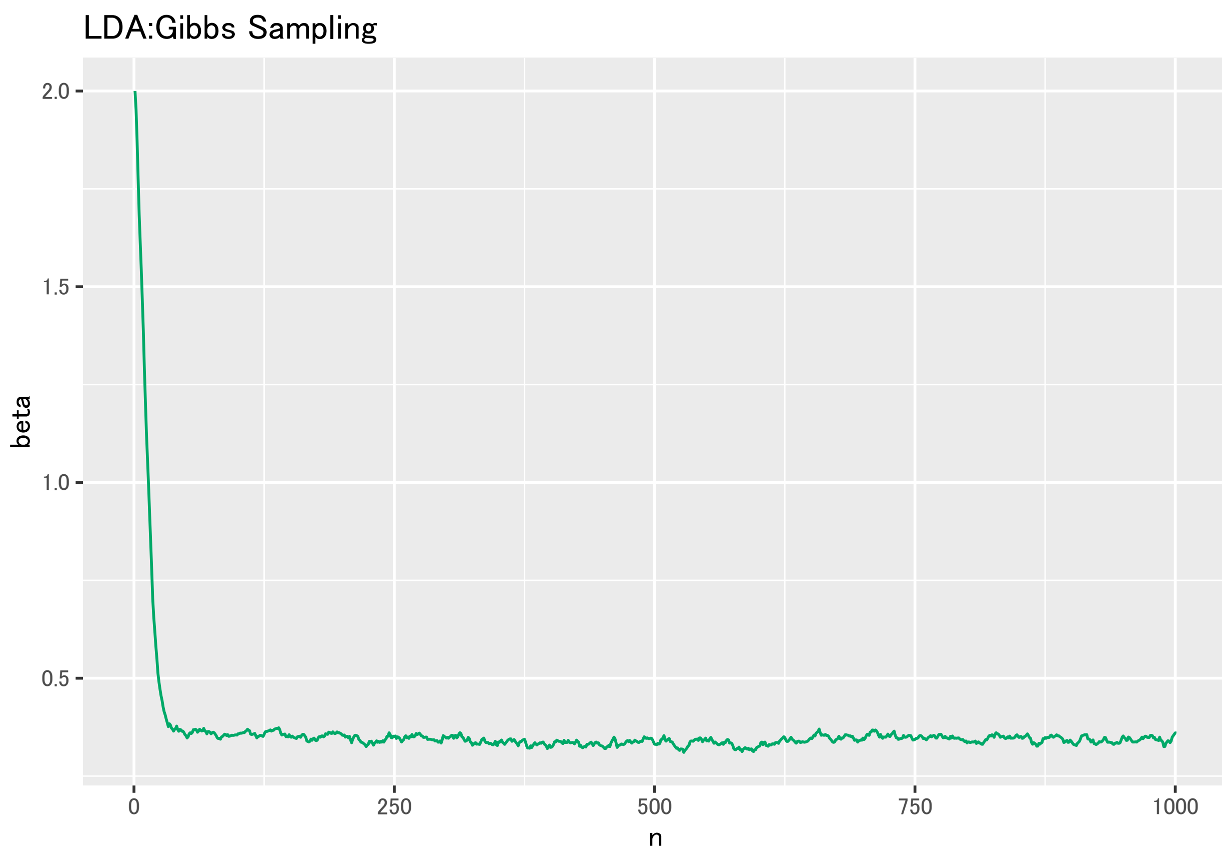
・トピック分布のパラメータ (多様)

```
## トピック分布のパラメータの推移の確認
# データフレームを作成
alpha_df_wide <- cbind(as.data.frame(t(trace_alpha)),
                       as.factor(1:ncol(trace_alpha))) # 推定回数

# データフレームを long 型に変換
colnames(alpha_df_wide) <- c(1:K, "n") # key 用の行名を付与
alpha_df_long <- gather(alpha_df_wide,
                        key = "topic", value = "alpha", -n) # 変換
alpha_df_long$n <- as.numeric(alpha_df_long$n)

# 描画
ggplot(data = alpha_df_long, mapping = aes(x = n, y = alpha, color = topic)) +  # データ
  geom_line() +                     # 折れ線グラフ
  labs(title = "LDA:Gibbs Sampling") # タイトル
```



・単語分布のパラメータ (一様)

```
## 単語分布のパラメータの推移の確認
# データフレームを作成
beta_df <- data.frame(n = 1:length(trace_beta),  # 推定回数
                      beta = trace_beta)

# 描画
ggplot(data = beta_df, mapping = aes(x = n, y = beta)) +  # データ
  geom_line(color = "#00A968") +      # 折れ線グラフ
  labs(title = "LDA:Gibbs Sampling") # タイトル
```



取り得る値が多様の場合のコードは以下になる。

```
# 結果の確認用（初期値）
trace_beta  <- as.matrix(beta_v)


# 結果の確認用（更新後）
trace_beta  <- cbind(trace_beta, as.matrix(beta_v))


## 単語分布のパラメータの推移の確認
# データフレームを作成
beta_df_wide <- cbind(as.data.frame(t(trace_beta)),
                      as.factor(1:ncol(trace_beta))) # 推定回数

# データフレームを long 型に変換
colnames(beta_df_wide) <- c(1:V, "n") # key 用の行名を付与
beta_df_long <- gather(beta_df_wide,
                       key = "topic", value = "beta", -n) # 変換
beta_df_long$n <- as.numeric(beta_df_long$n)

# 描画
ggplot(data = beta_df_long, mapping = aes(x = n, y = beta, color = topic)) +  # データ
  geom_line(alpha = 0.5) +            # 折れ線グラフ
  theme(legend.position = "none") +  # 凡例
  labs(title = "LDA:Gibbs Sampling") # タイトル
```

両方を一様とするの場合のコードは以下になる。

```
# 結果の確認用（初期値）
trace_alpha <- alpha
trace_beta  <- beta


# 結果の確認用（更新後）
trace_alpha <- c(trace_alpha, alpha)
trace_beta  <- c(trace_beta, beta)


## ハイパーパラメータの推移の確認
# データフレームを作成
ab_df_wide <- data.frame(n = seq_along(trace_alpha),  # 推定回数
                         alpha = trace_alpha,
                         beta = trace_beta)

# データフレームを long 型に変換
ab_df_long <- gather(ab_df_wide, key = "parameter", value = "value", -n)

# 描画
ggplot(data = ab_df_long, mapping = aes(x = n, y = value, color = parameter)) +  # データ
  geom_line() +                          # 折れ線グラフ
  labs(title = "LDA:Gibbs Sampling") # タイトル
```

テキストやトピック数に関わらず、同じ様な推移を辿ったり値が低かったりなど、プログラムの出来に不安が残ります。

Chapter4.6-4.9 はパスしました。

参考書籍

- 岩田具治 (2015)『トピックモデル』(機械学習プロフェッショナルシリーズ) 講談社
- 須山敦志 (2017)『ベイズ推論による機械学習入門』(機械学習スタートアップシリーズ) 講談社

おわりに

この資料は、青本輪読会 (参加者 2 名) のレジュメをまとめたものです。勉強会にてブレイン氏からトピックモデルの核となる部分の解説を受けつつ、全編私が執筆いたしました。それ故、誤りなど至らぬ点が多々あるかと思われます。ご了承ください。
何か気になる点があれば、全てご指摘いただけるととても嬉しいです。よろしくお願いいたします。

また資料作成において、NLPer や R おじさんのブログや助言に大変助けられました。ありがとうございます。この資料もそのように誰かの役に立てば幸いです。

という訳で、ここまで読むくらい楽しんでいただけたのでしたら、是非とも青本の方をご購入くださいね。ほら色々と著作権的な問題もあるでしょうし、知らんけど。勿論良い本ですし、何より応用編 Chapter5~8 が残っていますしね。

あと、私に興味を持っていただけたのでしたら、備忘録ブログ https://www.anarchive-beta.com/ なんぞを覗いてくれたら、私のやる気に繋がります。

では、@anemptyarchive の次回作にご期待ください。明日もがんばりまなかん゜。\(*'v'*)/゜

Topic Models(AO)

2019 年 07 月 09 日　第 1 版第 1 刷発行
2019 年 08 月 24 日　第 2 版第 1 刷発行

執筆者　@anemptyarchive
監修者　勉強会仲間
製作者　RStudio