クラスタリングとノンパラメトリックベイズ

NIGG

Tokyo, Japan

June 8, 2016

Contents

ディリクレ過程混合モデル

無限混合モデルとも呼ばれる。

有限混合モデル

- クラスタ数を事前に定める必要がある。

ディリクレ過程混合モデル

- クラスタ数はディリクレ過程により決定される。
- 生成モデルは、

$$G \sim DP(\alpha, H_0)$$

$$\theta^{\mathbf{k}} \sim \mathbf{G}(\theta_{\mathbf{i}})$$

$$\mathbf{x_k}|\theta_{\mathbf{i}} \sim N(\theta_{\mathbf{i}}, \sigma_{\mathbf{0}}^{\mathbf{2}}\mathbf{I}$$

ディリクレ過程混合モデル

ディリクレ過程混合モデル

- ディリクレ過程混合モデルとは、θ の確率分布である $G(\theta)$ を、$G \sim p(G|\alpha, \eta)$ を用いて生成する。
- 次に、$G(\theta)$ を用いて n 個のパターンに対応する $\theta_{\mathbf{1}}, \ldots, \theta_{\mathbf{n}}$ を生成する。
- 最後に、これらをパラメータとする確率分布 $p(\mathbf{x}|\theta_{\mathbf{i}})$ を生成する。
- $G \sim p(G|\alpha, \eta)$ をディリクレ過程とよぶ。
- 上のモデルは $H_0(\theta|\eta)$ を用いて、

$$p(G|\alpha, \eta) = DP(\alpha, H_0)$$

  と表現する。
- α は集中度パラメータ、$H_0(\theta|\eta)$ は基底分布という。

クラスタリング

K 平均法

$$(z^*_{1:n}, \mu^*_{1:K}) = argmin_{z_{1:n}, \mu_{1:K}} \sum_{i=1}^{n} \sum_{k=1}^{K} \delta(z_i = k) ||x_i - \mu_k||^2$$

クラスタリング

有限混合モデル

$$(z^*_{1:n}, \mu^*_{1:K}) = argmin_{z_{1:n},\mu_{1:K}} log \Pi_{i=1}^{n} N(x_i|\mu_{z_i}, I)$$

（注）上の変形は、

$$\sum_{i=1}^{n}\sum_{k=1}^{K}\delta(z_i = k)||x_i - \mu_k||^2$$

$$\propto -\sum_{i=1}^{n}\sum_{k=1}^{K}\delta(z_i = k)(-\frac{1}{2}||x_i - \mu_k||^2)$$

$$= -(nDlog(\sqrt{2\pi} + \sum_{i=1}^{n}\sum_{k=1}^{K}\delta(z_i = k)logN(\mathbf{x_i}|\mu_\mathbf{k}, \mathbf{I}))$$

$$= -(nDlog(\sqrt{2\pi} + log\Pi_{i=1}^{n} N(\mathbf{x_i}|\mu_\mathbf{zi}, \mathbf{I}))$$

による。2 段目の変形は log(Aexp(B)) を考えればよい。

ディリクレ分布の仮定

仮定付きのディリクレ分布

$\alpha_k = \alpha/K$ を仮定する。

$$z_i \sim Multi(\pi)(i = 1, ..., n), \pi \sim Dir(\alpha), \alpha_k = \alpha/K$$

$Dir(\pi|\alpha/K)$ の二つの性質

(1) 各 k で α_k が同じ値であるため、事前分布としては k に区別はない.
(2) 次元 K が大きくなるにしたがって、ディリクレ分布のパラメータ α_k は小さくなる.

ディリクレ分布の説明

そもそもディリクレ分布とは。

例題

いま m 種の目 $v_1, \ldots, v_m$ をもつ仮想的なサイコロが一個与えられたとする。ここで k の目を v_k で表す。このサイコロは、投げて v_k が出る確率が θ_k となるような細工が施してあり、かつその値は未知とする。このサイコロを続けて n 回投げた結果が、$x^n = \{x_1, \ldots, x_t, \ldots, x_n\}$ で、その内容は v_k が n_k であったとする。ただし、$x_t \in \{v_1, \ldots, v_m\}$ である。この結果をもとに、θ_k の値をベイズ推定により推定せよ。

ディィリクレ分布の説明

- このサイコロを投げて v_k が n_k 回出る確率を $P(\mathbf{n};\theta)$ と記すと (k=1,...,m), $P(\mathbf{n};\theta)$ は、

$$P(\mathbf{n};\theta) = \frac{\mathbf{n}!}{\mathbf{n_1}!\ldots\mathbf{n_m}!}\theta_{\mathbf{1}}^{\mathbf{n_1}}\ldots\theta_{\mathbf{m}}^{\mathbf{n_m}}$$

- ベイズの定理より、

$$p(\theta|\mathbf{x}^{(\mathbf{n})}) = \frac{\Gamma(\mathbf{n}+\mathbf{m})}{\Gamma(\mathbf{n_1}+\mathbf{1})\ldots\Gamma(\mathbf{n_m}+\mathbf{1})}\theta_{\mathbf{1}}^{\mathbf{n_1}}\ldots\theta_{\mathbf{m}}^{\mathbf{n_m}}$$

- 上式が多項分布、下式がディィリクレ分布に対応。
- $p(\theta|\mathbf{x}^{(\mathbf{n})}) = \mathbf{Dir}(\mathbf{n_1}+\mathbf{1},\ldots,\mathbf{n_m}+\mathbf{1})$

ディリクレ分布と多項分布の定義

ディリクレ分布

N 次元確率ベクトルの集合を

$$\triangle^K = \left\{\pi = (\pi_1, \pi_2, \ldots, \pi_K) | \sum_{k=1}^{K} = 1, \pi_k >= 0 \quad \forall k\right\}$$

とする。ディリクレ分布はこのような $\triangle^K$ 上の確率分布としてしばしば使われ、その密度関数は、

$$Dir(\pi|\alpha) \equiv \frac{\Gamma(\sum_{k=1}^{K} \alpha_k)}{\prod_{k=1}^{K} \Gamma(\alpha_k)} \prod_{k=1}^{K} \pi_k^{\alpha_k - 1}$$

ディリクレ分布と多項分布の定義

多項分布

x を K 種類の値 1,2,...,K をとる確率変数とする．それぞれの値をとる確率を $\pi = (\pi_a, \pi_2, ..., \pi_K)(\sum_{k=1}^{K} \pi_k = 1)$ とする．n 回の独立した試行を考え、$x_i = k$ により、i 回目の試行における値が k であることを示すとする．n_k で k という値が出た回数を表現する．π が与えられたもとで、$x_i = k$ である確率は $p(x_i = k|\pi) = \pi_k$ となる．このとき、π が与えられたもとでの $x = x_1, x_2, ..., x_n$ の確率は、

$$p(x|\pi) = \Pi_{i=1}^{n} p(x_i|\pi) = \Pi_{k=1}^{K} \pi_k^{n_k}$$

と計算できる．

サンプリングと周辺化への動機

以上の仮定のもと、$z_{1:n}(=\{z_1, ..., z_n\}$ のサンプリングを考える．ただし、π は $\sum_{k=1}^{\infty} \pi_k = 1$ という制約があり、この制約を満たしたままサンプリングを行うことは難しい．→周辺化をしてしまえば π を直接扱う必要はなくなる．

周辺化とは

結合分布から特定の変数を積分消去すること.
(例)

$$p(x_1, x_3) = \int p(x_1, x_2, x_3) dx_2$$

ベイズ推定では、観測データ $x_{1:n}$ の尤度 $p(x_{1:n}|\phi)$ を事前分布 $p(\phi|\eta)$ を事前分布 $p(\phi|\eta)$ で周辺化した

$$p(x_{1:n}|\eta) = \int p(x_{1:n}|\phi)p(\phi|\eta)d\phi = \int p(x_{1:n}, \phi|\eta)d\phi$$

を周辺尤度と呼ぶ.

$p(z_i = k | z_{1:n}^{\backslash i}, \alpha)$

$p(z_{n+1} = k | \{z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n\}, \alpha)$ の導出

(3.11) 参照

$p(z_{n+1} = k | \{z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_n\}, \alpha)$

$$p(z_i = k | z_{1:n}^{\backslash i}, \alpha) = \frac{n_k^{\backslash i} + \alpha / K}{n - 1 + \alpha}$$

$p(z_i = k | z_{1:n}^{\setminus i}, \alpha)$

$\mathbf{z}_{1:n}^{\setminus \mathbf{i}}$ に現れている{1,2,...,K}の集合を $K^+(\mathbf{z}_{1:n}^{\setminus \mathbf{i}})$ と表現する

$$p(z_i = k | z_{1:n}^{\setminus i}, \alpha) = \begin{cases} \dfrac{n_k^{\setminus i} + \alpha/K}{n - 1 + \alpha} if \quad k \in K^+(z_{1:n}^{\setminus i}) & (1) \\ \dfrac{\alpha/K}{n - 1 + \alpha} otherwise & (2) \end{cases}$$

$p(z_i = k | z_{1:n}^{\setminus i}, \alpha)$

z_i として $k \notin K^+(\mathbf{z}_{1:n}^{\setminus \mathbf{i}})$ となるすべての値をとる確率

$$p(z_i \notin K^+(z_{1:n}^{\setminus i}) | z_{1:n}^{\setminus i}, \alpha) = \sum_{k \notin K^+} p(z_i = k | z_{1:n}^{\setminus i}, \alpha/K)$$

$$= (K - |K^+(z_{1:n}^{\setminus i})|) \frac{\alpha/K}{n - 1 + \alpha}$$

$$= (1 - \frac{|K^+(z_{1:n}^{\setminus i})|}{K}) \frac{\alpha}{n - 1 + \alpha}$$

$p(z_i = k|z_{1:n}^{\setminus i}, \alpha)$

以上より、z_i がすでに存在しているクラスタに入る時の確率（各クラスタに定義される）と、そうでない場合（すべてのありうるクラスタから既存のクラスタを除いた補集合に z_i が含まれる場合をひとまとめにする。つまり未登場のクラスタのいずれか）とで場合分けできる。

$p(z_i)$

$$\begin{cases} p(z_i = k|z_{1:n}^{\setminus i}, \alpha) = \dfrac{n_k^{\setminus i} + \alpha/K}{n-1+\alpha} & (3) \\ p(z_i \notin K^+(z_{1:n}^{\setminus i})|z_{1:n}^{\setminus i}, \alpha) = (1 - \dfrac{|K^+(z_{1:n}^{\setminus i})|}{K})\dfrac{\alpha}{n-1+\alpha} & (4) \end{cases}$$

$$p(z_i = k | z_{1:n}^{\backslash i}, \alpha)$$

$\mathrm{K} \to \infty$

$$\begin{cases} p(z_i = k | z_{1:n}^{\backslash i}, \alpha) = \dfrac{n_k^{\backslash i}}{n-1+\alpha} & (5) \\ p(z_i \notin K^+(z_{1:n}^{\backslash i}) | z_{1:n}^{\backslash i}, \alpha) = \dfrac{\alpha}{n-1+\alpha} & (6) \end{cases}$$

この式が意味することは、すでにサンプリングされた値は $\frac{n_k^{\backslash i}}{n-1+\alpha}$ の確率でサンプリングされ、それ以外の何らかの値は $\frac{\alpha}{n-1+\alpha}$ の確率でサンプリングされるということ。

$$\left(\frac{n_1^{\backslash i}}{n-1+\alpha}, \frac{n_2^{\backslash i}}{n-1+\alpha}, \frac{n_5^{\backslash i}}{n-1+\alpha}, \frac{\alpha}{n-1+\alpha}\right)$$

$z_{1:n}z_1, \ldots, z_n$ の値をサンプリングするアルゴリズム

数値の付け替え

- $(z_1, z_2, z_3, z_4) = (2, 5, 1, 2)$ が意味していることは、データ 1 とデータ 4 が同じクラスで、その他は個別のクラス。
- $(z_1, z_2, z_3, z_4) = (2, 5, 1, 2)$ でも
  $(z_1, z_2, z_3, z_4) = (1, 2, 3, 1)$ としてもよい。
- 潜在変数の数値の取り替えを考えると常に $z_{1:n}$ に出現している潜在変数の数値を $\{1, 2, \ldots, |K^+(z_{1:n}^{\setminus i})|\}$ のように連番にすることができる。

$$p(z_i = k|z_{1:n}^{\setminus i}, \alpha) \begin{cases} \dfrac{n_k}{n-1+\alpha} if \quad k \in K^+(z_{1:n}^{\setminus i}) & (7) \\ \dfrac{\alpha}{n-1+\alpha} if \quad k = |K^+(z_{1:n}^{\setminus i})| + 1 & (8) \end{cases}$$

ギブスサンプリング

ガウス分布の平均や分散をサンプリングしない場合

$$p(z_i = k | x_{1:n}, z_{1:n}^{\backslash i}, \mu_0, \rho_0, a_0, b_0, \alpha)$$

$$\propto p(x_i | z_i = k, x_{1:n}^{\backslash i}, z_{1:n}^{\backslash i}, \mu_0, \rho_0, a_0, b_0) \times p(z_i = k | z_{1:n}^{\backslash i}, \alpha)$$

$$= \begin{cases} \int p(x_i|\mu_k, \tau) p(\mu_k, \tau | x_{1:n}, z_{1:n}^{\backslash i}, \mu_0, & (9) \\ \rho_0, a_0, b_0) d\mu_k d\tau \times \frac{n_k^{\backslash i}}{n - 1 + \alpha} & (10) \\ if \quad k \in K^+(z_{1:n}^{\backslash i}) & (11) \\ \int p(x_i|\mu_k, \tau) p(\mu_k, \tau | \mu_0, \rho_0, a_0, b_0) d\mu_k d\tau & (12) \\ \times \frac{\alpha}{n - 1 + \alpha} & (13) \\ if \quad k = |K^+(z_{1:n}^{\backslash i})| + 1 & (14) \end{cases}$$

ギブスサンプリング

ガウス分布の平均や分散をサンプリングする場合

$$p(z_i = k|x_{1:n}, z_{1:n}^{\backslash i}, \mu_0, \rho_0, a_0, b_0, \alpha)$$

$$= \begin{cases} p(x_i|\mu_{\mathbf{k}}, \tau) \times \dfrac{\mathbf{n_k^{\backslash i}}}{\mathbf{n-1}+\alpha} & \mu_{\mathbf{k}} \sim \mathbf{N}(\mu_{\mathbf{0}}, (\tau\rho_{\mathbf{0}})^{\mathbf{-1}}\mathbf{I}) \quad (15) \\ & if \quad k \in K^+(z_{1:n}^{\backslash i}) \quad (16) \\ p(x_i|\mu_{\mathbf{k}}, \tau) \times \dfrac{\alpha}{\mathbf{n-1}+\alpha} & \mu_{\mathbf{k}} \sim \mathbf{N}(\mu_{\mathbf{0}}, (\tau\rho_{\mathbf{0}})^{\mathbf{-1}}\mathbf{I}) \quad (17) \\ & if \quad k = |K^+(z_{1:n}^{\backslash i})| + 1 \quad (18) \end{cases}$$

周辺尤度 $p(\mathbf{z_{1:n}}|\alpha) = \int \mathbf{p}(\mathbf{z_i} = \mathbf{k}|\mathbf{z_{1:n}^{\setminus i}}, \alpha)$

ディリクレ分布の無限次元化についてより詳細に分析するために、$K \to \infty$ とした場合の周辺尤度 $p(\mathbf{z_{1:n}}|\alpha)$ について分析する。以下は混合ガウスモデルのデータも含めた周辺尤度であるが、ここではディリクレ分布の性質を分析したいので扱わない。

$$p(\mathbf{x_{1:n}}, \mathbf{z_{1:n}}|\mu_0, \rho_0, \mathbf{a_0}, \mathbf{b_0}, \alpha)$$

$$= p(\mathbf{x_{1:n}}|\mathbf{z_{1:n}}, \mu_0, \rho_0, \mathbf{a_0}, \mathbf{b_0}, \alpha) \times \mathbf{p}(\mathbf{z_{1:n}}|\mathbf{z_{1:n}}, \mu_0, \rho_0, \mathbf{a_0}, \mathbf{b_0}, \alpha)$$

$$= p(\mathbf{x_{1:n}}|\mathbf{z_{1:n}}, \mu_0, \rho_0, \mathbf{a_0}, \mathbf{b_0}) \times \mathbf{p}(\mathbf{z_{1:n}}|\alpha)$$

周辺尤度 $p(\mathbf{z_{1:n}}|\alpha)$ は 0

$$p(\mathbf{z_{1:n}}|\alpha) = \frac{\Gamma(\alpha)}{\Gamma(\mathbf{n}+\alpha)}(\alpha/\mathbf{K})^{\mathbf{K^+}}\Pi_{\mathbf{k=1}}^{\mathbf{K^+}}\left[\left(\Pi_{\mathbf{j=1}}^{\mathbf{n_k-1}}(\mathbf{j}+\alpha/\mathbf{K})\right)\right]$$

ただし、この時、$K \to \infty$ により $(\alpha/K)^{K^+} \to 0$ となり、$\mathrm{p} = 0$ となってしまう。そこで $p(\mathbf{z_{1:n}}|\alpha)$ に $\frac{K!}{(K-K_+)!}$ をかけると、

$$\frac{K!}{(K-K_+)!}p(\mathbf{z_{1:n}}|\alpha) \to \frac{\Gamma(\alpha)}{\Gamma(\mathbf{n}+\alpha)}\alpha^{\mathbf{K^+}}\Pi_{\mathbf{k=1}}^{\mathbf{K^+}}(\mathbf{n_k}-\mathbf{1}) \quad \textbf{as } \mathbf{K} \to \infty$$

分割の確率

これは順列をすべて足しあわせたものの確率を表しており、例えば、(1, 2, 3, 1)、(3, 1, 5, 3)、(4, 2, 5, 4) のような 1 回目と 4 回目が同じで 2 回目と 3 回目がそれぞれ違うといった順列の集合の確率であり、このとき (1, 2, 3, 1) を代表して [1, 2, 3, 1] と表現することにする。より一般的には $z_{1:n}$ を $[z_{1:n}]$ とすることである。つまり、

$$p([z_{1:n}]|\alpha) = \frac{K!}{(K-K_+)!}p(\mathbf{z_{1:n}}|\alpha)$$

$$\to \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)}\alpha^{K^+}\prod_{k=1}^{K^+}(n_k-1) \quad as\ K \to \infty$$

まとめ

・$\alpha_k = \alpha/K$ としたディリクレ分布と多項分布を組み合わせることで、データの分割に対して確率分布を定義することができる（第 1 章第 5 編)。ここで $z_{1:n}$ は数値そのものに意味があるわけでなく、分割の仕方を現しているに過ぎない（第 3 章)。例えば、$(z_1, z_2, z_3, z_4) = (2, 5, 1, 2)$ が意味していることは、$\{\{z_1, z_4\}, \{z_2\}, \{z_3\}\}$ という分割を意味している。
・分割に対する生成モデルでは、分割数 K を $K \to \infty$ としても、分割に対して確率が計算可能（第 3 章、分割の仕方に確率を定義できるのであって、特定の組の確率測度は 0)。

中華料理店過程

t=i のときに z_i のとる値が k である確率

$$p(z_i = k|\mathbf{z_{1:i-1}}, \alpha) = \begin{cases} \frac{n_k}{n-1+\alpha} if & k \in K^+(z_{1:i-1}^{\setminus i}) \\ \frac{\alpha}{n-1+\alpha} if & k = |K^+(z_{1:i-1}^{\setminus i})| + 1 \end{cases} \tag{19}$$

(20)

※ CRP は確率過程であるため時系列的な確率変数※ n_k は $\mathbf{z_{1:i-1}}$ における k の出現回数

ディリクレ過程

定義

集合ϕとその部分集合を要素とする集合族 F からなる可測空間（Φ,F）の基底分布を G_0、集中度パラメータをα（>0）とする。確率測度 G が、Φのいかなる可測な排他的分割

$$U_{i=1}^{c} A_i = \phi and A_i \cap A_j = \phi (i \neq j)$$

に対しても、r 次元確率ベクトル $(G(A_1), \ldots, G(A_r))$ がディリクレ分布 $Dir(G_0(\alpha A_1), \ldots, \alpha G_0(A_r))$ に従うとき、すなわち、

$$(G(A_1), \ldots, G(A_r)) \sim Dir(\alpha G_0(A_1), \ldots, \alpha G_0(A_r))$$

のとき、G はディリクレ過程に従うといい

$$G \sim DP(\alpha, G_0)$$

と記す。

ディリクレ過程混合モデル

ディリクレ過程混合モデル

- クラスタ数はディリクレ過程により決定される。
- 生成モデルは、

$$G \sim DP(\alpha, H_0)$$

$$\theta^{\mathbf{k}} \sim \mathbf{G}(\theta_{\mathbf{i}})$$

$$\mathbf{x_k}|\theta_{\mathbf{i}} \sim N(\theta_{\mathbf{i}}, \sigma_{\mathbf{0}}^{\mathbf{2}}\mathbf{I}$$

ディリクレ過程混合モデル

ディリクレ過程混合モデル

- ディリクレ過程混合モデルとは、θ の確率分布である $G(\theta)$ を、$G \sim p(G|\alpha, \eta)$ を用いて生成する。
- 次に、$G(\theta)$ を用いて n 個のパターンに対応する $\theta_\mathbf{1}, \ldots, \theta_\mathbf{n}$ を生成する。
- 最後に、これらをパラメータとする確率分布 $p(\mathbf{x}|\theta_\mathbf{i})$ を生成する。
- $G \sim p(G|\alpha, \eta)$ をディリクレ過程とよぶ。
- 上のモデルは $H_0(\theta|\eta)$ を用いて、

$$p(G|\alpha, \eta) = DP(\alpha, H_0)$$

  と表現する。
- α は集中度パラメータ、$H_0(\theta|\eta)$ は基底分布という。

準備：ベータ分布

α をギブスサンプリングにより推定する方法について説明する。まず、準備としてベータ分布を考える。$\pi \in [0, 1]$ が a_1 および a_2 をパラメータとするベータ分布に従っているとすると、

$$1 = \frac{\Gamma(a_1 + a_2)}{\Gamma(a_1)\Gamma(a_2)} \int \pi^{a_1 - 1}(1 - \pi)^{a_2 - 1} d\pi$$

ここで、$a_1 = \alpha + 1, a_2 = n$ を代入すると、

$$1 = \frac{\Gamma(\alpha + 1 + n)}{\Gamma(\alpha + 1)\Gamma(n)} \int \pi^{\alpha}(1 - \pi)^{n-1} d\pi$$

$$\Leftrightarrow 1 = \frac{(\alpha + n)\Gamma(\alpha + n)}{\alpha\Gamma(\alpha)\Gamma(n)} \int \pi^{\alpha}(1 - \pi)^{n-1} d\pi$$

$$\Leftrightarrow \frac{\Gamma(\alpha)}{\Gamma(\alpha + n)} = \frac{(\alpha + n)}{\alpha\Gamma(\alpha)} \int \pi^{\alpha}(1 - \pi)^{n-1} d\pi$$

α の事後分布

次に、α の事前分布を $p(\alpha|c_1, c_2) = Ga(\alpha|c_1, c_2)$ と仮定すると、α の事後分布は、

$$p(\alpha|c_1, c_2) \propto p(\alpha, \mathbf{z_{1:n}}|\mathbf{c_1}, \mathbf{c_2}) = \mathbf{p}(\mathbf{z_{1:n}}|\alpha)\mathbf{p}(\alpha|\mathbf{c_1}, \mathbf{c_2})$$

$$= \frac{\Gamma(\alpha)}{\Gamma(\alpha+n)}\alpha^{K^+}\Pi_{k=1}^{K^+}(n_k-1)! \times \frac{c_2^{c_1}}{\Gamma(c_1)}\alpha^{c_1-1}exp(-c_2\alpha)$$

$$= \frac{(\alpha+n)}{\alpha\Gamma(\alpha)}\int \pi^{\alpha}(1-\pi)^{n-1}d\pi \times$$

$$\alpha^{K^+}\Pi_{k=1}^{K^+}(n_k-1)! \times \frac{c_2^{c_1}}{\Gamma(c_1)}\alpha^{c_1-1}exp(-c_2\alpha)$$

α の事後分布導出の注意

$p([z_{1:n}]|\alpha)$

$$p([z_{1:n}]|\alpha) = \frac{K!}{(K-K_+)!} p(\mathbf{z_{1:n}}|\alpha)$$

$$\rightarrow \frac{\Gamma(\alpha)}{\Gamma(n+\alpha)} \alpha^{K^+} \Pi_{k=1}^{K^+} (n_k - 1) \quad as\ K \rightarrow \infty$$

$p(\alpha|c_1, c_2)$

$$p(\alpha|c_1, c_2) = \frac{c_2^{c_1}}{\Gamma(c_1)} \alpha^{c_1 - 1} exp(-c_2 \alpha)$$

$p(\alpha, \pi | \mathbf{z_{1:n}}, \mathbf{c_1}, \mathbf{c_2})$

$p(\alpha | \mathbf{z_{1:n}}, \mathbf{c_1}, \mathbf{c_2}) = \int \mathbf{p}(\alpha, \pi | \mathbf{z_{1:n}}, \mathbf{c_1}, \mathbf{c_2}) \mathbf{d}\pi$ に着目すると、上式を π に関して微分すれば $p(\alpha, \pi | \mathbf{z_{1:n}}, \mathbf{c_1}, \mathbf{c_2})$ が得られるので、

$$p(\alpha, \pi | \mathbf{z_{1:n}}, \mathbf{c_1}, \mathbf{c_2})$$

$$= \frac{(\alpha + n)}{\alpha} \alpha^{K^+} \pi^{\alpha} (1 - \pi)^{n-1} \times \frac{c_2^{c_1}}{\Gamma(c_1)} \alpha^{c_1 - 1} exp(-c_2 \alpha)$$

定数項を無視すれば、

$$p(\alpha, \pi | \mathbf{z_{1:n}}, \mathbf{c_1}, \mathbf{c_2})$$

$$\propto \frac{(\alpha + n)}{\alpha} \alpha^{K^+} \pi^{\alpha} (1 - \pi)^{n-1} \alpha^{c_1 - 1} exp(-c_2 \alpha)$$

$$= \frac{(\alpha + n)}{\alpha} \pi^{\alpha} (1 - \pi)^{n-1} \alpha^{K^+} \alpha^{c_1 - 1} exp(-c_2 \alpha)$$

$$= (1 + \frac{n}{\alpha}) \pi^{\alpha} (1 - \pi)^{n-1} \alpha^{K^+} \alpha^{c_1 - 1} exp(-c_2 \alpha)$$

$p(\alpha, \pi, s|\mathbf{z}_{1:n}, \mathbf{c}_1, \mathbf{c}_2)$

さらに二値をとる確率変数 $s \in \{0, 1\}$ を導入して、

$$p(\alpha, \pi, s|\mathbf{z}_{1:n}, \mathbf{c}_1, \mathbf{c}_2) \propto (\frac{\mathbf{n}}{\alpha})^{\mathbf{s}}\pi^{\alpha}(\mathbf{1}-\pi)^{\mathbf{n}-\mathbf{1}}\alpha^{\mathbf{K}^+}\alpha^{\mathbf{c}_1-\mathbf{1}}\mathbf{exp}(-\mathbf{c}_2\alpha)$$

とすることができる。
（周辺化してみると、

$$p(\alpha, \pi|\mathbf{z}_{1:n}, \mathbf{c}_1, \mathbf{c}_2) = \boldsymbol{\Sigma}_{\mathbf{s}\in\{\mathbf{0},\mathbf{1}\}}\mathbf{p}(\alpha, \pi, \mathbf{s}|\mathbf{z}_{1:n}, \mathbf{c}_1, \mathbf{c}_2)$$

となることが確認できる)
この結合分布から各パラメータの条件付き確率分布が導出できる。

ギブスサンプリング

π のサンプリング

$$\pi \sim Beta(\pi|\alpha + 1, n)$$

s のサンプリング

$$s \sim Bernoulli\left(s|\frac{\frac{n}{\alpha}}{1 + \frac{n}{\alpha}}\right)$$

α のサンプリング

$$\alpha \sim Ga(\alpha|c_1 + K^+ - s, c_2 - log\pi)$$