

ベイズ推測（ベイズ推論）

Bayesian inference

正田 備也

masada@rikkyo.ac.jp

Contents

予測分布

正規分布の場合

多項分布の場合

まとめ

統計的推測 statistical inference

- ▶ 観測データとして、 d 次元ユークリッド空間 \mathbb{R}^d 上の n 個の点の集合 $\mathbf{x}_1, \dots, \mathbf{x}_N$ が与えられているとする
- ▶ 観測データをひとつの記号で \mathcal{D} と書くことにする
 - ▶ つまり、 $\mathcal{D} \equiv \{\mathbf{x}_1, \dots, \mathbf{x}_N\}$
- ▶ \mathbf{x}_i は、独立に同じ分布 $q(\mathbf{x})$ にしたがうと考えることにする
 - ▶ つまり、 $q(\mathcal{D}) = \prod_{i=1}^N q(\mathbf{x}_i)$
- ▶ しかし、この分布 $q(\mathbf{x})$ を直接知る方法はない
- ▶ そこで、 \mathcal{D} から $q(\mathbf{x})$ を推測することを、統計的推測、あるいは、統計的学習という

確率モデル probabilistic model

- ▶ 真の分布を推測するとき、私たちは、あるパラメータ η をもつ確率分布 $p(x|\eta)$ を準備する
 - ▶ $p(x|\eta)$ を確率モデルと呼ぶ
- ▶ また、パラメータを決めること自体にも不確かさがあるとする場合、パラメータがしたがう事前分布 $p(\eta)$ も準備する
 - ▶ 事前分布のパラメータ（ハイパーパラメータ）は、まだ書かずに置く
- ▶ このとき事後分布は以下のように書ける

$$p(\boldsymbol{\eta}|\mathcal{D}) = \frac{1}{Z_N} p(\boldsymbol{\eta}) p(\mathcal{D}|\boldsymbol{\eta}) = \frac{1}{Z_N} p(\boldsymbol{\eta}) \prod_{i=1}^N p(\boldsymbol{x}_i|\boldsymbol{\eta}) \quad (1)$$

- ▶ Z_N については次のスライドで

周辺尤度 marginal likelihood

- ▶ 式 (1) の事後分布の規格化定数 Z_N を、分配関数とも呼ぶ
 - ▶ 分配関数 (partition function) は物理方面から来ている用語
- ▶ Z_N は規格化定数なので、 $Z_N = \int p(\boldsymbol{\eta})p(\mathcal{D}|\boldsymbol{\eta})d\boldsymbol{\eta}$ を満たす
 - ▶ つまり $Z_N = p(\mathcal{D})$
- ▶ $p(\mathcal{D})$ を周辺尤度、または、エビデンス (evidence) と呼ぶ
- ▶ すなわち、事後分布は以下のようにも書ける

$$p(\boldsymbol{\eta}|\mathcal{D}) = \frac{p(\boldsymbol{\eta})p(\mathcal{D}|\boldsymbol{\eta})}{p(\mathcal{D})} \quad (2)$$

観測データを生成する分布の推定 (1/2)

- ▶ 観測データ \mathcal{D} を元に、それを生成する分布を推定する方法は、複数ある（推定された分布を $\hat{p}(x)$ と書くことにする）
- ▶ 観測データ $\mathcal{D} \equiv \{x_1, \dots, x_N\}$ が与えられているとき、パラメータ η の関数 $\prod_{i=1}^N p(x_i|\eta)$ を尤度関数と呼ぶ
- 1. 尤度関数を最大にするパラメータ η_{ML} を最尤推定量といい、 $p(x|\eta_{\text{ML}})$ を推測の結果 $\hat{p}(x)$ とする方法を、**最尤推定**という
- 2. 事後分布の最大値を与えるパラメータ η_{MAP} を事後確率最大化推定量といい、 $p(x|\eta_{\text{MAP}})$ を推測の結果 $\hat{p}(x)$ とする方法を、事後確率最大化推定（もしくは**MAP推定**）という

観測データを生成する分布の推定 (2/2)

3. ベイズ的なモデリングの課題は事後分布を求めることだったが、事後分布を使うと、以下で定義する予測分布 $p(\boldsymbol{x}|\mathcal{D})$ をもって推測の結果 $\hat{p}(\boldsymbol{x})$ とすることができる
- ▶ 事後分布 $p(\boldsymbol{\eta}|\mathcal{D})$ によって確率モデル $p(\boldsymbol{x}|\boldsymbol{\eta})$ を平均化したものの、つまり、 $p(\boldsymbol{\eta}|\mathcal{D})$ に関する $p(\boldsymbol{x}|\boldsymbol{\eta})$ の期待値を**予測分布**と呼ぶ

$$p(\boldsymbol{x}|\mathcal{D}) = \int p(\boldsymbol{x}|\boldsymbol{\eta})p(\boldsymbol{\eta}|\mathcal{D})d\boldsymbol{\eta} \quad (3)$$

ベイズ推測

- ▶ ベイズ推測 Bayesian inference とは、「真の分布は、おおよそ予測分布だろう」と推測することである

cf. 渡辺澄夫『ベイズ統計の理論と方法』コロナ社、p.5

- ▶ 以下、正規分布と多項分布の場合について、共役事前分布を使ったときに予測分布がどのような分布になるかを説明する

Contents

予測分布

正規分布の場合

多項分布の場合

まとめ

単変量正規分布を使ったベイズ的モデリング

- ▶ 観測データ $\mathcal{D} \equiv \{x_1, \dots, x_N\}$ が独立に同じ正規分布 $\mathcal{N}(\mu, \tau^{-1})$ にしたがうと仮定
- ▶ 平均 μ と精度 τ の事前分布として正規ガンマ分布 $\text{NG}(\mu, \tau; \mu_0, \lambda_0, \alpha, \beta)$ を使う
- ▶ 正規ガンマ分布の確率密度関数は

$$\begin{aligned} p(\mu, \tau; \mu_0, \lambda_0, \alpha, \beta) &= p(\mu|\tau; \mu_0, \lambda_0, \alpha, \beta)p(\tau; \alpha, \beta) \\ &= \frac{\beta^\alpha \sqrt{\lambda_0}}{\Gamma(\alpha)\sqrt{2\pi}} \tau^{\alpha-\frac{1}{2}} e^{-\beta\tau} e^{-\frac{\lambda_0\tau(\mu-\mu_0)^2}{2}} \end{aligned} \quad (4)$$

共役事前分布としての正規ガンマ分布 (1/2)

- ▶ 正規ガンマ分布は共役事前分布
- ▶ よって事後分布も正規ガンマ分布となる
- ▶ 事後分布 $p(\mu, \tau | \mathcal{D}; \mu_0, \lambda_0, \alpha, \beta)$ は以下のように書ける
 - ▶ 前回の講義資料を参照

$$\begin{aligned} p(\mu, \tau | \mathcal{D}; \mu_0, \lambda_0, \alpha, \beta) \\ \propto \tau^{\alpha + \frac{N}{2} - \frac{1}{2}} \exp \left[-\tau \left(\beta + \frac{Ns}{2} + \frac{\lambda_0 N (\bar{x} - \mu_0)^2}{2(\lambda_0 + N)} \right) \right] \\ \times \exp \left[-\frac{\tau}{2} (\lambda_0 + N) \left(\mu - \frac{\lambda_0 \mu_0 + N \bar{x}}{\lambda_0 + N} \right)^2 \right] \end{aligned} \quad (5)$$

共役事前分布としての正規ガンマ分布 (2/2)

- ▶ N を観測データの個数、 $\bar{x} = \frac{\sum_{i=1}^N x_i}{N}$ 、 $s = \frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N}$ とする。

cf. https://en.wikipedia.org/wiki/Normal-gamma_distribution

Table: 事前分布と事後分布とでのパラメータの値の対比

事前分布	事後分布
μ_0	$\frac{\lambda_0 \mu_0 + N \bar{x}}{\lambda_0 + N}$
λ_0	$\lambda_0 + N$
α	$\alpha + \frac{N}{2}$
β	$\beta + \frac{1}{2} \left(N s + \frac{\lambda_0 N (\bar{x} - \mu_0)^2}{\lambda_0 + N} \right)$

正規ガンマ事後分布に対応する予測分布

- ▶ 上記の設定のもとでの予測分布 $p(x|\mathcal{D})$ は

$$p(x|\mathcal{D}; \mu_0, \lambda_0, \alpha, \beta) = \int p(x|\mu, \tau) p(\mu, \tau|\mathcal{D}; \mu_0, \lambda_0, \alpha, \beta) d\mu d\tau$$

- ▶ 事後分布は $p(\mu, \tau|\mathcal{D}) = p(\mu|\tau, \mathcal{D}; \mu_0, \lambda_0) p(\tau|\mathcal{D}; \mu_0, \lambda_0, \alpha, \beta)$ と正規分布とガンマ分布の積で書ける

$$p(x|\mu, \tau) \propto \exp\left(-\frac{\tau(x-\mu)^2}{2}\right)$$

$$p(\mu|\tau, \mathcal{D}; \mu_0, \lambda_0) \propto \exp\left[-\frac{\tau(\lambda_0 + N)}{2} \left(\mu - \frac{\lambda_0\mu_0 + N\bar{x}}{\lambda_0 + N}\right)^2\right]$$

$$p(\tau|\mathcal{D}; \mu_0, \lambda_0, \alpha, \beta) \propto \tau^{\alpha + \frac{N}{2} - 1} \exp\left[-\tau\left\{\beta + \frac{Ns}{2} + \frac{\lambda_0 N(\bar{x} - \mu_0)^2}{2(\lambda_0 + N)}\right\}\right]$$

- ▶ 予測分布 $p(x|\mathcal{D}; \mu_0, \lambda_0, \alpha, \beta)$ はどういう分布になるか？ 13 / 30

t -分布 Student's t -distribution

- ▶ x_1, \dots, x_N を平均 μ 、標準偏差 σ の正規分布に独立にしたがう確率変数とする
- ▶ 標本平均を $\bar{x} = \frac{\sum_i x_i}{N}$ 、不偏分散を $s^2 = \frac{\sum_i (x_i - \bar{x})^2}{N-1}$ とする
- ▶ このとき、 $t = \frac{\bar{x} - \mu}{s/\sqrt{N}}$ と定義される値は、自由度 $\nu = N - 1$ の t -分布にしたがう
- ▶ t -分布の確率密度関数は、以下のようなになる

$$p(t; \nu) = \frac{\Gamma((\nu + 1)/2)}{\sqrt{\nu\pi}\Gamma(\nu/2)} (1 + t^2/\nu)^{-(\nu+1)/2} \quad (6)$$

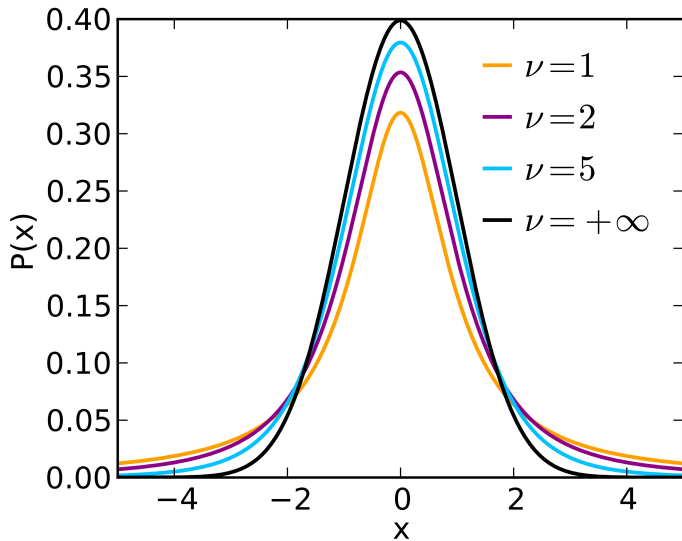


Figure: https://en.wikipedia.org/wiki/Student's_t-distribution

t location-scale distribution

- ▶ 平均が μ 、スケールが σ 、自由度が ν の t location-scale distribution の確率密度関数は、以下のとおり

$$t_{\nu}(x; \mu, \sigma^2) = \frac{\Gamma((\nu + 1)/2)}{\sigma\sqrt{\nu\pi}\Gamma(\nu/2)} \left(1 + \frac{(x - \mu)^2}{\nu\sigma^2}\right)^{-(\nu+1)/2} \quad (7)$$

- ▶ $\mu = 0$ 、 $\sigma = 1$ のとき、 t -分布に一致する
 - ▶ 英語版 Wikipedia 「 t -分布」の “Location-scale t-distribution” を参照
 - ▶ MATLAB の `tLocationScaleDistribution` の項も参照
- ▶ 以下、予測分布 $p(x|\mathcal{D}; \mu_0, \lambda_0, \alpha, \beta)$ が t location-scale 分布になることを示す

準備として、正規ガンマ分布 $\text{NG}(\mu, \tau | \mu_0, \lambda_0, \alpha, \beta)$ において $\alpha = \beta = \frac{\nu}{2}$ である場合、以下が成立することを確認しておく（証明はしない）。(cf. [CS340 \(Machine learning\) Fall 2007 @ UBC](#))

$$t_\nu(\mu; \mu_0, \lambda_0^{-1}) = \int_0^\infty \text{NG}(\mu, \tau; \mu_0, \lambda_0, \frac{\nu}{2}, \frac{\nu}{2}) d\tau = \int_0^\infty \mathcal{N}(\mu | \mu_0, (\lambda_0 \tau)^{-1}) \text{Ga}(\tau; \frac{\nu}{2}, \frac{\nu}{2}) d\tau \quad (8)$$

ただし、 $t_\nu(x | \mu, \sigma^2)$ は、平均 μ 、スケール σ^2 、自由度 ν の t location-scale distribution の密度関数である。

$x \sim \text{Ga}(\alpha, \beta)$ のとき、 $cx \sim \text{Ga}(\alpha, \frac{\beta}{c})$ となる。よって

$$\begin{aligned} \int_0^\infty \mathcal{N}(\mu | \mu_0, (\lambda_0 \tau)^{-1}) \text{Ga}(\tau; \alpha, \beta) d\tau &= \int_0^\infty \mathcal{N}(\mu | \mu_0, (\lambda_0 \tau)^{-1}) \text{Ga}(\tau; \alpha, \frac{\beta}{\alpha} \alpha) d\tau \\ &= \int_0^\infty \mathcal{N}(\mu | \mu_0, (\lambda_0 \tau)^{-1}) \text{Ga}(\frac{\alpha}{\beta} \tau; \alpha, \alpha) d\tau \\ &= \int_0^\infty \mathcal{N}(\mu | \mu_0, (\frac{\beta}{\alpha} \lambda_0 \tau')^{-1}) \text{Ga}(\tau'; \alpha, \alpha) d\tau' \\ &= t_{2\alpha}(\mu; \mu_0, \frac{\beta}{\alpha \lambda_0}) \end{aligned} \quad (9)$$

もうひとつの準備として、異なる二つの単変量正規分布の密度関数の積について、以下の結果を確認しておく。

この結果は、[このブログ記事](#)において提示されているものだが、[The Matrix Cookbook](#) の 7.2.6. で提示されている結果の特殊例でもある。

平均が m で標準偏差が s の単変量正規分布の密度関数を $f(x; m, s)$ と書くことにすると

$$f(x; \mu_1, \sigma_1) f(x; \mu_2, \sigma_2) = f\left(\mu_1; \mu_2, \sqrt{\sigma_1^2 + \sigma_2^2}\right) f(x; \mu, \sigma) \quad (10)$$

ただし、

$$\mu = \frac{\sigma_1^{-2} \mu_1 + \sigma_2^{-2} \mu_2}{\sigma_1^{-2} + \sigma_2^{-2}} \quad (11)$$

$$\sigma^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2} \quad (12)$$

$p(\mu|\tau, \mathcal{D}) \propto \exp[-\frac{\tau(\lambda_0+N)}{2}(\mu - \frac{\lambda_0\mu_0+N\bar{x}}{\lambda_0+N})^2]$ において $\mu_N = \frac{\lambda_0\mu_0+N\bar{x}}{\lambda_0+N}$, $\tau_N = \tau(\lambda_0 + N)$ とおく。
 前のスライドの結果を使うべく、 $\mu_1 = x, \sigma_1 = \tau^{-1/2}, \mu_2 = \mu_N, \sigma_2 = \tau_N^{-1/2}$ とおくと

$$\begin{aligned} p(x|\mu, \tau)p(\mu|\tau, \mathcal{D}) &\propto \exp\left[-\frac{\tau\tau_N}{2(\tau + \tau_N)}(x - \mu_N)^2\right] \exp\left[-\frac{\tau + \tau_N}{2}\left(\mu - \frac{\tau_N x + \tau\mu_N}{\tau + \tau_N}\right)^2\right] \\ &= \exp\left[-\frac{\tau(\lambda_0 + N)}{2(\lambda_0 + N + 1)}(x - \mu_N)^2\right] \exp\left[-\frac{\tau(\lambda_0 + N + 1)}{2}\left(\mu - \frac{(\lambda_0 + N)x + \mu_N}{\lambda_0 + N + 1}\right)^2\right] \end{aligned} \quad (13)$$

よって、 $\alpha_N = \alpha + \frac{N}{2}, \beta_N = \beta + \frac{Ns}{2} + \frac{\lambda_0 N(\bar{x} - \mu_0)^2}{2(\lambda_0 + N)}$ 、さらに $\lambda_N = \lambda_0 + N$ とおくと、

$$\begin{aligned} p(x|\mu, \tau)p(\mu|\tau, \mathcal{D})p(\tau|\mathcal{D}) &\propto \exp\left[-\frac{\tau(\lambda_0 + N)}{2(\lambda_0 + N + 1)}(x - \mu_N)^2\right] \\ &\times \exp\left[-\frac{\tau(\lambda_0 + N + 1)}{2}\left(\mu - \frac{(\lambda_0 + N)x + \mu_N}{\lambda_0 + N + 1}\right)^2\right] \times \tau^{\alpha_N-1}e^{-\tau\beta_N} \\ &= \tau^{\alpha_N-1}e^{-\tau\beta_N} \exp\left[-\frac{\tau\lambda_N}{2(\lambda_N + 1)}(x - \mu_N)^2\right] \times \exp\left[-\frac{\tau(\lambda_N + 1)}{2}\left(\mu - \frac{\lambda_N x + \mu_N}{\lambda_N + 1}\right)^2\right] \end{aligned} \quad (14)$$

μ を積分消去すると

$$p(x, \tau | \mathcal{D}) = \int_{-\infty}^{\infty} p(x | \mu, \tau) p(\mu | \tau, \mathcal{D}) p(\tau | \mathcal{D}) d\mu \propto \tau^{\alpha_N - \frac{1}{2}} e^{-\tau \beta_N} \exp \left[-\frac{\tau \lambda_N}{2(\lambda_N + 1)} (x - \mu_N)^2 \right] \quad (15)$$

この式は $p(x, \tau | \mathcal{D})$ が正規ガンマ分布の密度関数であることを示している。

そこで、 τ を積分消去するために式 (9) の結果を使うと、

$$p(x | \mathcal{D}) = \int p(x, \tau | \mathcal{D}) d\tau = t_{2\alpha_N} \left(x | \mu_N, \frac{\beta_N(\lambda_N + 1)}{\alpha_N \lambda_N} \right) \quad (16)$$

したがって、予測分布は、平均 μ_N 、スケール $\sqrt{\frac{\beta_N(\lambda_N + 1)}{\alpha_N \lambda_N}}$ 、自由度 $2\alpha_N$ の t location-scale distribution である。

但し、 $\mu_N, \lambda_N, \alpha_N, \beta_N$ は、元々の事前分布（正規ガンマ分布）のパラメータである $\mu_0, \lambda_0, \alpha, \beta$ を使った式として書かれていることに注意されたい。

つまり、 $\mu_0, \lambda_0, \alpha, \beta$ の値が定まってはじめて、予測分布のパラメータ（平均、スケール、自由度）の値も定まることに、注意されたい。

Contents

予測分布

正規分布の場合

多項分布の場合

まとめ

多項分布を使ったベイズ的モデリング

- ▶ 観測データ $\boldsymbol{x} = \{x_1, \dots, x_n\}$ が独立に同じカテゴリカル分布 $\text{Cat}(\phi)$ にしたがうと仮定
 - ▶ 例えば、 $x_{62} = \text{“apple”}$ は、62 番目に出現した単語が “apple” だ、という意味
- ▶ ϕ の事前分布としてディリクレ分布 $\text{Dir}(\beta)$ を使う
- ▶ ディリクレ分布の確率密度関数は

$$p(\phi; \beta) = \frac{\Gamma(\sum_{w=1}^W \beta_w)}{\prod_{w=1}^W \Gamma(\beta_w)} \prod_{w=1}^W \phi_w^{\beta_w - 1} \quad (17)$$

共役事前分布としてのディリクレ分布

- ▶ ディリクレ分布は共役事前分布
- ▶ つまり、事後分布もディリクレ分布となる
- ▶ 事後分布 $p(\phi|x; \beta)$ は以下のように書ける

$$p(\phi|x; \beta) = \frac{\Gamma(\sum_{w=1}^W (c_w + \beta_w))}{\prod_{w=1}^W \Gamma(c_w + \beta_w)} \prod_{w=1}^W \phi_w^{c_w + \beta_w - 1} \quad (18)$$

- ▶ c_w は w 番目の語彙が観測データ x に出現する回数

ディリクレ事後分布に対応する予測分布

- ▶ 多項分布の場合、予測分布についても、多数の単語出現からなる観測データの予測分布を考えることが多い
- ▶ その予測分布を求めたい単語列（つまり文書）を \mathbf{x}_0 とする
 - ▶ $\mathbf{x}_0 = \{x_{0,1}, \dots, x_{0,n_0}\}$ とする（ n_0 は \mathbf{x}_0 の文書長）
 - ▶ \mathbf{x}_0 において w 番目の語彙が出現する回数を $c_{0,w}$ と書く
- ▶ 上記の設定のもとでの予測分布 $p(\mathbf{x}_0|\mathbf{x};\boldsymbol{\beta})$ は

$$p(\mathbf{x}_0|\mathbf{x};\boldsymbol{\beta}) = \int p(\mathbf{x}_0|\boldsymbol{\phi})p(\boldsymbol{\phi}|\mathbf{x};\boldsymbol{\beta})d\boldsymbol{\phi} \quad (19)$$

によって求められる

- ▶ この予測分布がどういう分布になるかを、以下で示す

$$\begin{aligned}
p(\mathbf{x}_0|\mathbf{x};\boldsymbol{\beta}) &= \int p(\mathbf{x}_0|\boldsymbol{\phi})p(\boldsymbol{\phi}|\mathbf{x};\boldsymbol{\beta})d\boldsymbol{\phi} \\
&= \int \left(\frac{n_0!}{\prod_{w=1}^W c_{0,w}!} \prod_{w=1}^W \phi_w^{c_{0,w}} \times \frac{\Gamma(\sum_{w=1}^W (c_w + \beta_w))}{\prod_{w=1}^W \Gamma(c_w + \beta_w)} \prod_{w=1}^W \phi_w^{c_w + \beta_w - 1} \right) d\boldsymbol{\phi} \\
&= \frac{n_0!}{\prod_{w=1}^W c_{0,w}!} \frac{\Gamma(\sum_{w=1}^W (c_w + \beta_w))}{\prod_{w=1}^W \Gamma(c_w + \beta_w)} \int \prod_{w=1}^W \phi_w^{c_w + c_{0,w} + \beta_w - 1} d\boldsymbol{\phi} \\
&= \frac{n_0!}{\prod_{w=1}^W c_{0,w}!} \frac{\Gamma(\sum_{w=1}^W (c_w + \beta_w))}{\prod_{w=1}^W \Gamma(c_w + \beta_w)} \frac{\prod_{w=1}^W \Gamma(c_w + c_{0,w} + \beta_w)}{\Gamma(\sum_{w=1}^W (c_w + c_{0,w} + \beta_w))} \\
&= \frac{n_0! \Gamma(\sum_{w=1}^W (c_w + \beta_w))}{\Gamma(\sum_{w=1}^W (c_w + c_{0,w} + \beta_w))} \prod_{w=1}^W \frac{\Gamma(c_w + c_{0,w} + \beta_w)}{c_{0,w}! \Gamma(c_w + \beta_w)} \tag{20}
\end{aligned}$$

この x_0 の関数を、確率質量関数として持つ分布に、何か名前は付いているか？

ディリクレ多項分布 (Dirichlet-multinomial distribution)

- ▶ パラメータは $\alpha = (\alpha_1, \dots, \alpha_K)$ s.t. $\alpha_k > 0$ for $1 \leq k \leq K$
- ▶ 確率密度関数は

$$p(\mathbf{x}; \alpha) = \frac{\Gamma(\sum_k \alpha_k) \prod_k \Gamma(\alpha_k + n_k)}{\prod_k \Gamma(\alpha_k) \Gamma(\sum_k \alpha_k + n_k)} \quad (21)$$

- ▶ n_k は \mathbf{x} において第 k 番目のアイテムが出現する回数
- ▶ 導出の仕方

$$p(\mathbf{x}; \alpha) = \int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}; \alpha)d\boldsymbol{\theta} \quad (22)$$

- ▶ $p(\mathbf{x}|\boldsymbol{\theta})$ はパラメータが $\boldsymbol{\theta}$ の多項分布の確率質量関数
- ▶ $p(\boldsymbol{\theta}; \alpha)$ はパラメータが α のディリクレ分布の確率密度関数

ディリクレ事後分布に対応する予測分布（再）

- ▶ ディリクレ多項分布は式 (22) で導出される

$$p(\boldsymbol{x}; \boldsymbol{\alpha}) = \int p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}; \boldsymbol{\alpha})d\boldsymbol{\theta}$$

- ▶ 式 (19) も、式 (22) と同じ形をしている

$$p(\boldsymbol{x}_0|\boldsymbol{x}; \boldsymbol{\beta}) = \int p(\boldsymbol{x}_0|\boldsymbol{\phi})p(\boldsymbol{\phi}|\boldsymbol{x}; \boldsymbol{\beta})d\boldsymbol{\phi}$$

- ▶ 式 (22) の $p(\boldsymbol{\theta}; \boldsymbol{\alpha})$ を、事後分布 $p(\boldsymbol{\phi}|\boldsymbol{x}; \boldsymbol{\beta})$ で置き換えればよい。
- ▶ つまり、得られた予測分布は、ディリクレ多項分布である

Contents

予測分布

正規分布の場合

多項分布の場合

まとめ

まとめ：予測分布とは何だったか

- ▶ 事前分布 $p(\boldsymbol{\eta})$ と尤度 $p(\mathcal{D}|\boldsymbol{\eta})$ から、事後分布 $p(\boldsymbol{\eta}|\mathcal{D})$ を求める

$$p(\boldsymbol{\eta}|\mathcal{D}) = \frac{1}{Z_N} p(\boldsymbol{\eta}) p(\mathcal{D}|\boldsymbol{\eta}) = \frac{1}{Z_N} p(\boldsymbol{\eta}) \prod_{i=1}^N p(\boldsymbol{x}_i|\boldsymbol{\eta}) \quad (23)$$

- ▶ 事後分布 $p(\boldsymbol{\eta}|\mathcal{D})$ は、モデルパラメータが取りうるあらゆる値に重み付けをしている、と見なせる
- ▶ 予測分布 $p(\boldsymbol{x}|\mathcal{D})$ を使うと、事後分布 $p(\boldsymbol{\eta}|\mathcal{D})$ によるパラメータへの重み付けを反映するかたちで、未知データ \boldsymbol{x} の確率が得られる

$$p(\boldsymbol{x}|\mathcal{D}) = \int p(\boldsymbol{x}|\boldsymbol{\eta}) p(\boldsymbol{\eta}|\mathcal{D}) d\boldsymbol{\eta} \quad (24)$$

予測分布の別の見方

- ▶ 予測分布を求めることは、観測データがしたがる新しい分布をベイズ的な枠組みを使って作り出すこと、とも言える
 1. 正規分布と正規ガンマ分布から、 t location-scale 分布
 2. 多項分布とディリクレ分布から、ディリクレ多項分布
- ▶ ただし、今回の2つの例のように、予測分布の確率密度関数を、単に式の計算で求めることができってしまうケースは、実際には少ない。