GaP, NMF, and more

DMLA 2006-5-19 (Fri)

ATR 音声言語コミュニケーション研究所 持橋大地
daichi.mochihashi@atr.jp

* 準備

ポアソン分布：稀な事象の起こる離散確率分布

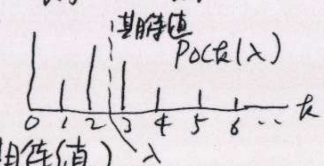
確率 p ($\ll 1$) で表が出る硬貨を n 回ふると、期待値は
$\lambda = np$. $\therefore p = \frac{\lambda}{n}$.

$$p(\text{表が}k\text{回出る}\mid n, p) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \binom{n}{k}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\cancel{n(n-1)\cdots(n-k+1)}}{k!}\frac{\lambda^k}{\cancel{n^k}}\left(1-\frac{\lambda}{n}\right)^n\left(1-\frac{\lambda}{n}\right)^{-k}$$

$$\xrightarrow{n\to\infty} \frac{\lambda^k}{k!}\cdot e^{-\lambda}\cdot 1$$



$$\therefore Po(k|\lambda) = \frac{\lambda^k}{k!}e^{-\lambda} \quad (\lambda: \text{期待値})$$

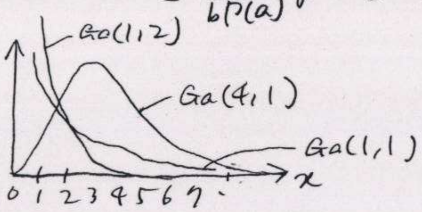
覚え方：e^λ のマクローリン展開 $e^\lambda = 1 + \frac{\lambda}{1} + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \cdots$ の各項.

ガンマ分布：ポアソン分布のパラメータ λ の共役事前分布

$$Ga(x|a,b) = \frac{1}{b^a\Gamma(a)} x^{a-1} e^{-\frac{x}{b}}$$

$$= \frac{1}{b\Gamma(a)}\left(\frac{x}{b}\right)^{a-1}\exp\left(-\frac{x}{b}\right). \quad \frac{x}{b} = y \text{ とおけば、}$$

$$= \frac{1}{b\Gamma(a)} y^{a-1} e^{-y} \propto \frac{y^{a-1}}{e^y}. \quad \left(\frac{\text{Polynomial}}{\text{Exponential}}\right)$$



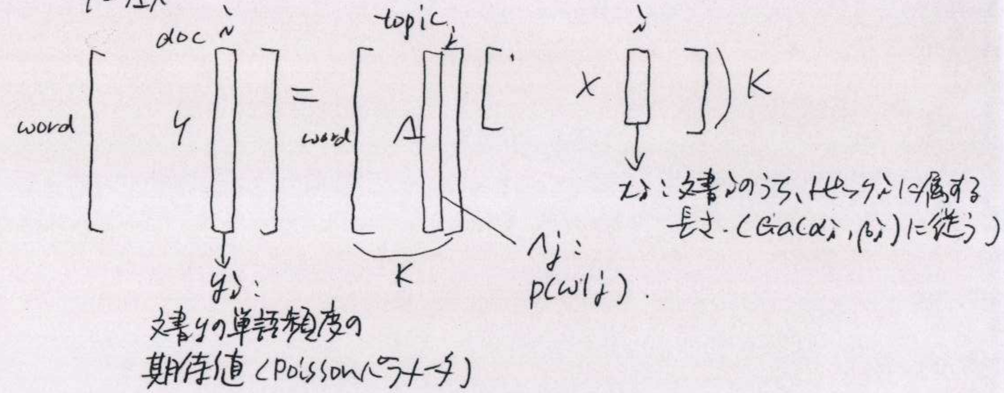
* $x_1 \sim Ga(a,1)$,
$x_2 \sim Ga(b,1)$ のとき、
$x_1 + x_2 \sim Ga(a+b, 1)$.

* GaP

$$\underset{\alpha,\beta}{\circ} \xrightarrow{\text{Gamma}} \underset{X}{\circ} \xrightarrow{\Lambda} \underset{Y}{\circ} \xrightarrow{\text{Poisson}} \underset{F}{\circledcirc}$$

(F：共起カウント行列)

$Y = \Lambda X$



$\langle x_{ij}\rangle = a_j b_j = c_j$ とおけば、$b_j = c_j / a_j$ ゆえ

$$p(F, Y, X) = p(X|a,b)\,p(F|Y=\Lambda X) = Ga(X|a,b)\,Po(F|\Lambda X)$$

$$= \prod_{j=1}^{n} \frac{y_j^{f_j}}{f_j!} \prod_{j=1}^{k} \frac{a_j^{a_j} x_{ij}^{a_j-1}\exp(-a_j x_{ij}/c_j)}{c_j^{a_j}\Gamma(a_j)}$$

$y_i = \sum_k \lambda_{ik} x_k$ なので、

$$\log p(F, X) = \sum_j \left[f_j \log \sum_k \lambda_{jk} x_k - \sum_k \lambda_{jk} x_k - \log f_j!\right]$$

$$+ \sum_j \left[(a_j - 1)\log x_{ij} - a_j x_{ij}/c_j + a_j \log \frac{a_j}{c_j} - \log\Gamma(a_j)\right]$$

これを最大化する. //

Lemma

ディリクレ分布 $\theta \sim Dir(\alpha_1, \alpha_2, \cdots, \alpha_k) = \frac{\Gamma(\sum_k \alpha_k)}{\prod_k \Gamma(\alpha_k)} \prod_k \theta_k^{\alpha_k - 1}$
からのサンプリング：

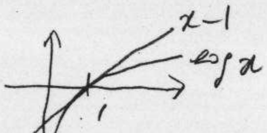
$x_j \sim Ga(\alpha_j, 1)$ $(j = 1, 2, \cdots, k)$
として、

$$\theta = \frac{1}{\sum_j x_j}(x_1, x_2, \cdots, x_k). \quad (\text{和を1に正規化する})$$

NMF

＊準備

$\log x \le x-1$ より、



$\log \frac{b}{a} \le \frac{b}{a} - 1$. 両辺に a をかけて

$a \log \frac{b}{a} \le b - a$. → $a \log \frac{a}{b} - a + b \ge 0$. (＊)

よって、非負測度 $a=(a_1, \cdots, a_n)$, $b=(b_1, \cdots, b_n)$ について

$\sum_i a_i \ne 1$, $\sum_i b_i \ne 1$ でも、I-divergence

$$D(a\|b) = \sum_{i=1}^{n} \left(a_i \log \frac{a_i}{b_i} - a_i + b_i\right) \quad \text{(NMF: (3)式)}$$

が定義できる.

$\sum_i a_i = \sum_i b_i = 1$ のときは、普通の KL-Divergence と同じ.

＊NMF

いま、NMF

$$V = WH \qquad (V: \text{データ})$$

について、I-ダイバージェンス

$$F(h) = D(V\|WH)$$

$$= \sum_i v_i \log \frac{v_i}{\sum_a w_{ia} h_a} - v_i + \sum_a w_{ia} h_a$$

ここが難しい.

を最小化する w, h を求めることを考える.

＊準備

$$\log \int p(x) f(x)\, dx \ge \int p(x) \log f(x)\, dx \quad \text{(Jensen)}$$

より.

$$\log \sum_a w_{ia} h_a = \log \sum_a \alpha_a \cdot \frac{w_{ia} h_a}{\alpha_a}$$

$$\ge \sum_a \alpha_a \log \frac{w_{ia} h_a}{\alpha_a}.$$

$\alpha_a = \dfrac{w_{ia} \hat{h}_a}{\sum_a w_{ia} \hat{h}_a}$ とおくと. 代入して

$$\log \sum_a w_{ia} h_a \ge \sum_a \frac{w_{ia} \hat{h}_a}{\sum_b w_{ib} \hat{h}_b} \left(\log w_{ia} h_a - \log \frac{w_{ia} \hat{h}_a}{\sum_b w_{ib} \hat{h}_b}\right).$$

よって、

$$F(h) = \sum_i (v_i \log v_i - v_i) + \sum_{ia} w_{ia} h_a - \sum_i v_i \log \sum_a w_{ia} h_a$$

$$\le \sum_i (v_i \log v_i - v_i) + \sum_{ia} w_{ia} h_a$$

$$- \sum_i v_i \sum_a \frac{w_{ia} \hat{h}_a}{\sum_b w_{ib} \hat{h}_b} \left(\log w_{ia} h_a - \log \frac{w_{ia} \hat{h}_a}{\sum_b w_{ib} \hat{h}_b}\right)$$

$$= G(h, \hat{h}). \quad \text{(upper bound)}$$

$$\frac{\partial}{\partial h_a} G(h, \hat{h}) = \sum_i w_{ia} - \sum_i v_i \frac{\cancel{w_{ia}}}{\cancel{w_{ia}} h_a} \frac{w_{ia} \hat{h}_a}{\sum_b w_{ib} \hat{h}_b}$$

$$= \sum_i w_{ia} - \sum_i v_i \cdot \frac{1}{h_a} \frac{w_{ia} \hat{h}_a}{\sum_b w_{ib} \hat{h}_b} = 0.$$

$$\therefore \frac{1}{h_a} \sum_i v_i \frac{w_{ia} \hat{h}_a}{\sum_b w_{ib} \hat{h}_b} = \sum_i w_{ia}$$

$$\rightarrow h_a = \frac{\sum_i v_i \frac{w_{ia} \hat{h}_a}{\sum_b w_{ib} \hat{h}_b}}{\sum_i w_{ia}} = \hat{h}_a \cdot \frac{\sum_i \frac{v_i w_{ia}}{\sum_b w_{ib} \hat{h}_b}}{\sum_i w_{ib}}. \quad //$$