3.2.4.3 を膨らませてみる。

そもそも、本にあるように、(3.39) のパラメータの定義をそのまま利用する。すると、

$$\mu = \frac{e^\eta}{1+e^\eta} \tag{1}$$

ベータ分布 (3.49) を (3.11) のように変数変換すると (3.50) にあるように、以下のようになる。

$$\begin{aligned}Beta_\eta(\eta|\lambda_1,\lambda_2) = Beta(\mu|\alpha,\beta)\frac{d\mu}{d\eta} = exp((\alpha-1)ln\mu + (\beta-1)ln(1-\mu) + ln\frac{\Gamma(\alpha+\beta}{\Gamma(\alpha)\Gamma(\beta)})(-\frac{e^\eta}{(1+e^\eta)^2} + \frac{e^\eta}{1+e^\eta})\\ = exp((\alpha-1)ln\frac{e^\eta}{1+e^\eta} + (\beta-1)ln\frac{1}{1+e^\eta} + ln\frac{\Gamma(\alpha+\beta}{\Gamma(\alpha)\Gamma(\beta)})(\frac{e^\eta}{(1+e^\eta)^2})\\ = exp(\eta(\alpha-1) - (\alpha+\beta-2)ln(1+e^\eta) + ln\frac{\Gamma(\alpha+\beta}{\Gamma(\alpha)\Gamma(\beta)})(\frac{e^\eta}{(1+e^\eta)^2}) = exp(\eta\alpha - (\alpha+\beta)ln(1+e^\eta) + ln\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)})\end{aligned} \tag{2}$$

(3.39) より、$a(\eta) = ln(1+e^\eta)$ なので、上記の式は (3.51) と一致する。

上記の式を (3.45) の式と見比べると、(3.51) の式が出てくる。

(3.52) に関して、(3.47) を踏まえると

$$\begin{aligned}p(\mu|\mathbf{X}) = Beta_\eta(\eta|\hat{\lambda_1},\hat{\lambda_2}) = Cexp(\eta(\alpha+\sum_{n=1}^N x_n) - a(\eta)(\alpha+\beta+N))\\ = Cexp((\alpha+\sum_{n=1}^N x_n - 1)ln(e^\eta) - ln(1+e^\eta)(\alpha+\beta+N-2))(\frac{e^\eta}{(1+e^\eta)^2})\\ = Cexp((\alpha+\sum_{n=1}^N x_n - 1)ln\frac{e^\eta}{1+e^\eta} + ln\frac{1}{1+e^\eta}(\beta+N-\sum_{n=1}^N x_n - 1))(\frac{e^\eta}{(1+e^\eta)^2})\\ = Cexp((\alpha+\sum_{n=1}^N x_n - 1)ln\mu + ln(1-\mu)(\beta+N-\sum_{n=1}^N x_n - 1))\frac{d\eta}{d\mu} = Beta(\mu|\hat{\alpha},\hat{\beta})\frac{d\eta}{d\mu}\end{aligned} \tag{3}$$

(自然パラメータのときは (3.47) で事後分布を計算できる。また、途中、a_c を定数 C とおいた。ベータ分布だとわかると正規化項は一意に定まる。)

よって、

$$Beta(\mu|\hat{\alpha},\hat{\beta}) = Cexp((\alpha+\sum_{n=1}^N x_n-1)ln\mu+ln(1-\mu)(\beta+N-\sum_{n=1}^N x_n-1)) = C\mu^{\alpha+\sum_{n=1}^N x_n-1}(1-\mu)^{\beta+N-\sum_{n=1}^N x_n-1} \tag{4}$$

これから、(3.54),(3.55) が求まる。

(3.53) に関して、(3.39) より、$h(x_*) = 1, t(x_*) = x_*$,(3.51) より、$a_c(\lambda) = -ln\frac{\Gamma(\lambda_1)\Gamma(\lambda_2-\lambda_1)}{\Gamma(\lambda_2)}$ なので、(3.48) に当てはめると、

$$p(x_*|\mathbf{X}) = \frac{exp(a_c(\hat{\lambda}_1+x_*,\hat{\lambda}_2+1))}{exp(a_c(\hat{\lambda}_1,\hat{\lambda}_2))} = \frac{\frac{\Gamma(\hat{\lambda}_1+x_*)\Gamma(\hat{\lambda}_2+1-\hat{\lambda}_1-x_*)}{\Gamma(\hat{\lambda}_2+1)}}{\frac{\Gamma(\hat{\lambda}_1)\Gamma(\hat{\lambda}_2-\hat{\lambda}_1)}{\Gamma(\hat{\lambda}_2)}} = \frac{\Gamma(\hat{\lambda}_1+x_*)\Gamma(\hat{\lambda}_2-\hat{\lambda}_1+1-x_*)}{\Gamma(\hat{\lambda}_2+1)}\frac{\Gamma(\hat{\lambda}_2)}{\Gamma(\hat{\lambda}_1)\Gamma(\hat{\lambda}_2-\hat{\lambda}_1)} \tag{5}$$

(3.47),(3.51) を用いて、式をまとめ、(3.54),(3.55) で置き換える。(3.25) にも注意すると、

$$
\begin{aligned}
p(x_*|\mathbf{X}) &= \frac{\Gamma(\lambda_1 + \sum_n x_n + x_*)\Gamma(\lambda_2 + N - \lambda_1 - \sum_n x_n + 1 - x_*)}{\Gamma(\lambda_2 + N + 1)} \frac{\Gamma(\lambda_2 + N)}{\Gamma(\lambda_1 + \sum_n x_n)\Gamma(\lambda_2 + N - \lambda_1 - \sum_n x_n)} \\
&= \frac{\Gamma(\alpha + \sum_n x_n + x_*)\Gamma(\beta + N - \sum_n x_n + 1 - x_*)}{\Gamma(\alpha + \beta + N + 1)} \frac{\Gamma(\alpha + \beta + N)}{\Gamma(\alpha + \sum_n x_n)\Gamma(\beta + N - \sum_n x_n)} \\
&= \frac{\Gamma(\hat{\alpha} + x_*)\Gamma(\hat{\beta} + 1 - x_*)}{\Gamma(\hat{\alpha} + \hat{\beta} + 1)} \frac{\Gamma(\hat{\alpha} + \hat{\beta})}{\Gamma(\hat{\alpha})\Gamma(\hat{\beta})} = \frac{\Gamma(\hat{\alpha} + x_*)\Gamma(\hat{\beta} + 1 - x_*)}{(\hat{\alpha} + \hat{\beta})\Gamma(\hat{\alpha} + \hat{\beta})} \frac{\Gamma(\hat{\alpha} + \hat{\beta})}{\Gamma(\hat{\alpha})\Gamma(\hat{\beta})} = \frac{\Gamma(\hat{\alpha} + x_*)\Gamma(\hat{\beta} + 1 - x_*)}{(\hat{\alpha} + \hat{\beta})\Gamma(\hat{\alpha})\Gamma(\hat{\beta})}
\end{aligned} \tag{6}
$$

この先は $x_* \in \{0, 1\}$ に注意し、ベイズ推論による機械学習 (緑本) の (3.19)-(3.21) のやり方を参考にすると[*1]、上記の式が平均、$\frac{\hat{\alpha}}{\hat{\alpha}+\hat{\beta}}$ となるベルヌーイ分布になることがわかり、(3.53) が導出される。

[*1] 購入をおすすめします。