「ベイズ統計の理論と方法」
Chapter 2, p.46-47

@tmiya_

June 28, 2017

補題 6

$\mathcal{T}_n^{(1)}(\alpha)$, $\mathcal{T}_n^{(2)}(\alpha)$ もほぼ同様。
$\alpha = 0$ を代入すると $\ell_k(A) = \mathbb{E}_w[(\log p(A|w))^k]$ であるからキュムラントが求められる。特に

$$\mathcal{G}_n'(0) = \mathbb{E}_X[\mathbb{E}_w[\log p(X|w)]] = -L(w_0) - \mathbb{E}_w[K(w)]$$

$$\mathcal{T}_n'(0) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}_w[\log p(X_i|w)] = -L_n(w_0) - \mathbb{E}_w[K_n(w)].$$

$$\mathcal{G}_n''(0) = \mathbb{E}_X\left[\mathbb{E}_w[(\log p(X|w))^2] - \mathbb{E}_w[\log p(X|w)]^2\right],$$

$$\mathcal{T}_n''(0) = \frac{1}{n}\sum_{i=1}^{n}\left[\mathbb{E}_w[(\log p(X_i|w))^2] - \mathbb{E}_w[\log p(X_i|w)]^2\right].$$

（教科書のここの部分は符号が正しい）

定義 11，補題 7

定義 11

$$\mathcal{L}_k(A) = \frac{\mathbb{E}_w[(-f(A,w))^k \exp(-\alpha f(A,w))]}{\mathbb{E}_w[\exp(-\alpha f(A,w))]}$$

(分子の $(-f(A,w))^k$ **の符号が教科書は正しく無い)**

補題 7

$$\mathcal{G}_n^{(1)}(\alpha) = -L(w_0) + \mathbb{E}_X[\mathcal{L}_1(X)],$$

$$\mathcal{G}_n^{(2)}(\alpha) = \mathbb{E}_X[\mathcal{L}_2(X) - \mathcal{L}_1(X)^2],$$

$$\mathcal{T}_n^{(1)}(\alpha) = L_n(w_0) + \frac{1}{n}\sum_i \mathcal{L}_1(X_i),$$

$$\mathcal{T}_n^{(2)}(\alpha) = \frac{1}{n}\sum_i \{\mathcal{L}_2(X_i) - \mathcal{L}_1(X_i)^2\}.$$

教科書の (2.9)，(2.10) **式は符号が逆になってる。**

定義 11，補題 7

証明

$$
\begin{aligned}
\mathcal{G}_n(\alpha) &= \mathbb{E}_X[\log \mathbb{E}_w[\underbrace{p(X|w)}_{p_0 e^{-f}}{}^{\alpha}]] \\
&= \underbrace{\mathbb{E}_X[\log p(X|w_0)^{\alpha}]}_{-\alpha L(w_0)} + \mathbb{E}_x[\log \mathbb{E}_w[e^{-\alpha f}]] \\
&= -(2.9)\ \text{式} \qquad \Leftarrow \text{符号が逆では？} \\
\mathcal{G}_n'(\alpha) &= -L(w_0) + \frac{d}{d\alpha}\mathbb{E}_X[\log \mathbb{E}_w[e^{-\alpha f}]] \\
&= -L(w_0) + \mathbb{E}_X\left[\frac{\mathbb{E}_w[-fe^{-\alpha f}]}{\mathbb{E}_w[e^{-\alpha f}]}\right] \\
&= -L(w_0) + \mathbb{E}_X[\mathcal{L}_1(X)] \\
\mathcal{G}_n''(\alpha) &= -\frac{d}{d\alpha}L(w_0) + \frac{d}{d\alpha}\mathbb{E}_X\left[\frac{\mathbb{E}_w[-fe^{-\alpha f}]}{\mathbb{E}_w[e^{-\alpha f}]}\right] \\
&= \mathbb{E}_X\left[\frac{\mathbb{E}_w[f^2 e^{-\alpha f}]\mathbb{E}_w[e^{-\alpha f}] - \mathbb{E}_w[-fe^{-\alpha f}]^2}{\mathbb{E}_w[e^{-\alpha f}]^2}\right] \\
&= \mathbb{E}_X\left[\mathcal{L}_2(X) - \mathcal{L}_1(X)^2\right].
\end{aligned}
$$

補題 7

$\mathcal{T}_n'(\alpha)$, $\mathcal{T}_n''(\alpha)$ **の計算も同様。**

$$\begin{aligned}
\mathcal{T}_n(\alpha) &= \frac{1}{n}\sum_i \log \mathbb{E}_w[p(X_i|w)^\alpha] \\
&= \frac{1}{n}\sum_i \log p(X_i|w_0)^\alpha + \frac{1}{n}\sum_i \log \mathbb{E}_w[e^{-\alpha f}] \\
&= -\alpha L_n(w_0) + \frac{1}{n}\sum_i \log \mathbb{E}_w[e^{-\alpha f}] \\
&= -(2.10)\ \textbf{式} \qquad \Leftarrow \textbf{符号が逆では？} \\
\mathcal{T}_n'(\alpha) &= -L_n(w_0) + \frac{1}{n}\sum_i \mathcal{L}_1(X_i) \\
\mathcal{T}_n''(\alpha) &= \frac{1}{n}\sum_i \left\{\mathcal{L}_2(X_i) - \mathcal{L}_1(X_i)^2\right\}
\end{aligned}$$

補題 8

$c_2 = 2$, $c_3 = 6$, $c_4 = 26$ とすると、$k = 2, 3, 4$ において

$$\left| \left(\frac{d}{d\alpha} \right)^k \mathcal{G}_n(\alpha) \right| \leq c_k \mathbb{E}_X \left[\frac{\mathbb{E}_w[|f(X,w)|^k \exp(-\alpha f(X,w))]}{\mathbb{E}_w[\exp(-\alpha f(X,w))]} \right],$$

$$\left| \left(\frac{d}{d\alpha} \right)^k \mathcal{T}_n(\alpha) \right| \leq c_k \frac{1}{n} \sum_i \frac{\mathbb{E}_w[|f(X,w)|^k \exp(-\alpha f(X,w))]}{\mathbb{E}_w[\exp(-\alpha f(X,w))]}.$$

$\because$ **平均操作** $\mathbb{E}_w^{(\alpha)}[\ ]$ **を任意の関数** $g(\ ,w)$ **と確率変数** A **について**

$$\mathbb{E}_w^{(\alpha)}[g(A,w)] \equiv \frac{\mathbb{E}_w[g(A,w)\exp(-\alpha f(A,w))]}{\mathbb{E}_w[\exp(-\alpha f(A,w))]}$$

と定義すると、ヘルダーの不等式から $j \leq k$ **のとき、**

$$\mathbb{E}_w^{(\alpha)}[|f(A,w)|^j] \leq \mathbb{E}_w^{(\alpha)}[|f(A,w)|^k]^{\frac{j}{k}}$$

よって $|f(A,w)|$ **の最高次の項で抑えられる。（**$\mathcal{L}_j$ **の積は** j **の和が** k **になる）**
注意 18 **の係数の絶対値の和が** c_k **になっている。**

定理 1（ベイズ統計の基本定理）

条件

$$\left| \left(\frac{d}{d\alpha} \right)^3 \mathcal{G}_n(\alpha) \right| = o_p(\frac{1}{n}), \quad \left| \left(\frac{d}{d\alpha} \right)^3 \mathcal{T}_n(\alpha) \right| = o_p(\frac{1}{n}).$$

が成り立つと仮定する。この時、汎化誤差と経験損失はキュムラントから次の式で計算できる。

$$G_n = -\mathcal{G}_n(1) = -\mathcal{G}'_n(0) - \frac{1}{2}\mathcal{G}''_n(0) + o_p(\frac{1}{n}),$$
$$T_n = -\mathcal{T}_n(1) = -\mathcal{T}'_n(0) - \frac{1}{2}\mathcal{T}''_n(0) + o_p(\frac{1}{n}),$$

定理 1（ベイズ統計の基本定理）

$0 \leq \alpha \leq 1$ なる α が存在して

$$\mathcal{G}_n(1) = \mathcal{G}_n(0) + \mathcal{G}_n'(0) + \frac{1}{2}\mathcal{G}_n''(0) + \frac{1}{6}\mathcal{G}_n(\alpha)$$

よりなりたつ。T_n も同様。