広く使える情報量規準(WAIC)

東京工業大学　渡辺澄夫

要約：　事後分布が正規分布で近似できないときでもAIC, TIC, DIC が使えないときでも、いつでも汎化誤差を推定できる方法を作りました。

$X_1, X_2, \ldots, X_n$ ：　真の分布 $q(x)$ に従う独立な確率変数

$p(x|w)$:　モデル : パラメータ w を持つ x の確率分布

$\varphi(w)$:　パラメータ w の事前分布

$$E_w[\quad] = \frac{1}{Z} \int (\quad) \prod_{i=1}^{n} p(X_i|w)\, \varphi(w)\, dw$$ ：事後分布による平均

$p^*(x) = E_w[p(x|w)]$ ：　予測分布

$T = -(1/n) \sum_{i=1}^{n} \log p^*(X_i)$ ：　学習損失

$G = -\int q(x) \log p^*(x)\, dx$ ：　汎化損失

$V = \sum_{i=1}^{n} \{ E_w[(\log p(X_i|w))^2] - E_w[\log p(X_i|w)]^2 \}$: 汎関数分散

$w_0 = \mathrm{argmin}_w \{ -\int q(x) \log p(x|w) dx \}$: 一般にユニークでない

$L_n(w_0) = -(1/n) \sum_{i=1}^{n} \log p(X_i|w_0)$

$L(w_0) = -\int q(x) \log p(x|w_0)\, dx$

定義：　$\mathrm{WAIC} = T + V/n$ ：広く使える情報量規準

定理　：　G とWAIC は平均が同じです。

$$E[G] = E[\mathrm{WAIC}] + O(1/n^2)$$

定理　：　$G - L(w_0)$ と $\mathrm{WAIC} - L_n(w_0)$ は分散が同じです。

$$V[G - L(w_0)] = V[\mathrm{WAIC} - L_n(w_0)] + o(1/n^2)$$