2021年6月18日 統計数理研究所 オープンハウス

ベイズ事後分布から頻度論的な情報を取り出す

伊庭 幸人 モデリング研究系 教授

leave-one-out CVやWAICでは，1回のMCMCで得られる事後分布からのサンプルを用いて頻度論的なモデル選択を行うことができる．それらはどこまで一般化することができるのか？ベイズ統計と頻度主義的な統計を結び付ける仕組みは何なのか？

leave-one-out CV (LOOCV) とWAIC

$\mathrm{E}_{\mathrm{pos}}[\ ]$, $\mathrm{Var}_{\mathrm{pos}}[\ ]$, $\mathrm{Cov}_{\mathrm{pos}}[\ ]$: 事後分布による期待値，分散，共分散

平均対数損失 $-\mathrm{E}_{y_{new}}[\log p(y_{new}|y)] = -\mathrm{E}_{y_{new}}[\log \mathrm{E}_{\mathrm{pos}}[p(y_{new}|\theta)]]$ を推定（以下$2N$倍）

$\mathrm{E}_{y_{new}}[\ ]$:テストデータについての期待値, N:サンプルサイズ

$$\mathrm{LOOCV}/(-2) = \sum_i \left(\mathrm{E}_{\mathrm{pos}} \left[\frac{1}{p(y_i|\theta)} \right] \right)^{-1}$$

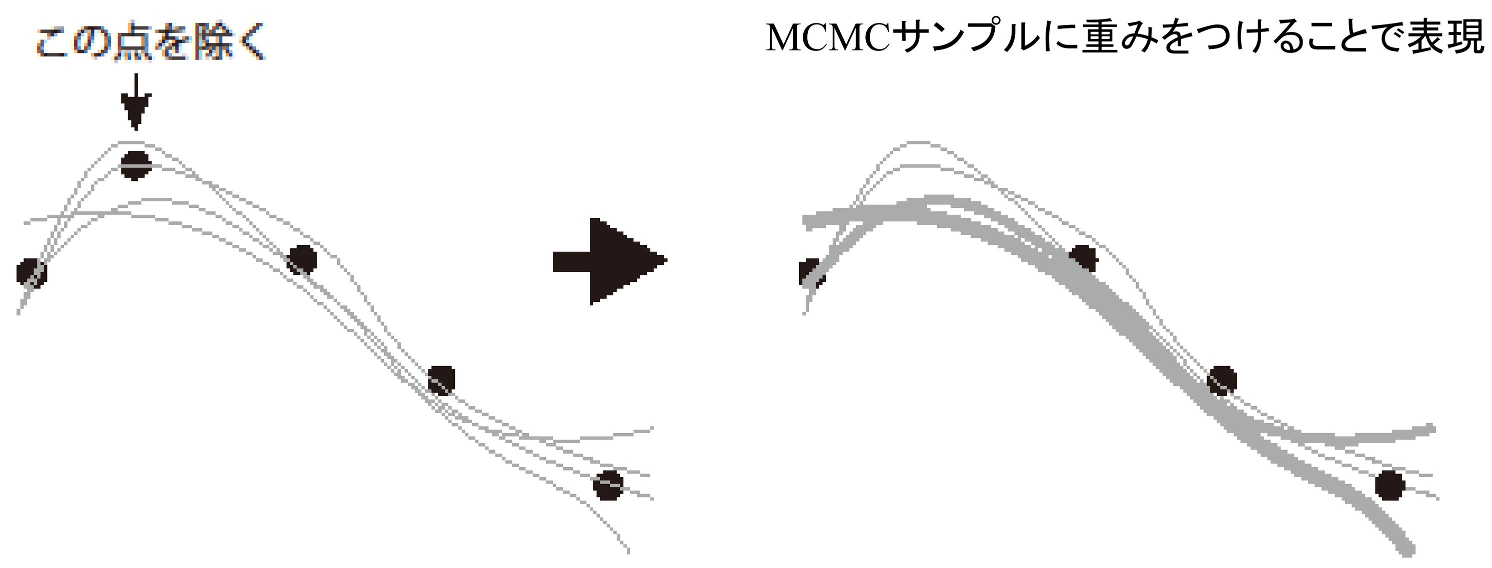
Geisser and Eddy (1979)
Gelfand and Dey (1994)

1個でなくε個除くとして1個除く場合に外挿
数学的にはキュムラント展開を利用

$$\mathrm{WAIC} = (-2) \left\{ \sum_i \log \mathrm{E}_{\mathrm{pos}}[p(y_i|\theta)] - \underbrace{\sum_i \mathrm{Var}_{\mathrm{pos}}[\log p(y_i|\theta)]}_{\text{事後分散}} \right\}$$

Watanabe (2010)

LOOCVでは観測を除いてMCMCを走らせる代わりに重みを付ける
→ 1回のMCMC計算で済む
WAICではそれを微分型に書き換える
→ 重みによるMCMC計算での分散の増大を抑止



なぜうまく行くのか？ ➔ 以下の3つの要素が本質的

- 各観測が独立，もしくは何らかの形でアプリオリに切り分けられている（ブートストラップやCVと同様の前提）：母集団の表現
- パラメータでなく，観測値の摂動で考える： 頻度論の数理的なとらえ方の工夫
- 有限の摂動でなく微分で考える→事後分散によるコンパクトな表現：WAICのうまい点

同様の考え方で事後分布と頻度論をより広く繋ぐことができるはず： 正則モデルの範囲でも下記の点で有用

- CVやブートストラップのような繰り返し計算がいらず1回のMCMCで計算できる．モデルごとの解析的計算も不要

提案手法：Posterior Covariance Information Criterion

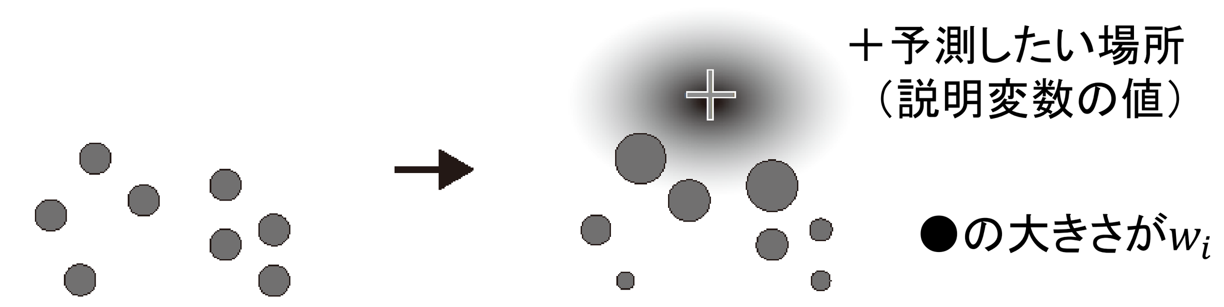
重み付き尤度$w_i^* \log p(y_i|\theta)$によって定義される事後分布
評価関数$h(y_{i,new}|\theta)$>0から定義される予測分布の重みw_i付きKL損失
に一般化 （特異モデルではhと尤度の間などに条件が必要）
重み付き平均対数損失 $-\sum_i \mathrm{E}_{y_{new}}[w_i \log \mathrm{E}_{\mathrm{pos}}[h(y_{i_{new}}|\theta)]]$を推定（以下2倍）

PCIC

$$= (-2) \left\{ \sum_i w_i \log \mathrm{E}_{\mathrm{pos}}[h(y_i|\theta)] - \underbrace{\sum_i w_i w_i^* \mathrm{Cov}_{\mathrm{pos}}[\log h(y_i|\theta), \log p(y_i|\theta)]}_{\text{事後共分散}} \right\}$$

たとえば，重み付き尤度による推論に応用がある（右側で説明）

共変量シフト下の推論，傾向スコア付きの因果推論などでは，尤度，損失関数に（しばしば別の）重みがつく
➔共変量シフトへの応用は矢野氏のポスター参照



傾向スコア付きの推論におけるモデル選択
Baba, Kanamori, Ninomiya (2017)
→ 同様のことがMCMCのサンプルを用いて行える

それ以外の一般化

ギブス損失に対する情報量規準（とりあえず正則モデルに限定）
任意の損失関数$\nu(y_{new}|\theta)$について
平均ギブス損失 $-\mathrm{E}_{y_{new}}[\mathrm{E}_{\mathrm{pos}}[\nu(y_{new}|\theta)]]$ を推定 （以下負号除きN倍）
（簡単のため重みなしの形で書いたが一般化は容易；右も同様）

$$\mathrm{PCIC}_G = \left\{ \sum_i \mathrm{E}_{\mathrm{pos}}[\nu(y_i|\theta)] - \sum_i \mathrm{Cov}_{\mathrm{pos}}[\nu(y_i|\theta), \log p(y_i|\theta)] \right\}$$

プラグイン推定に関する情報量規準
任意の損失関数$\nu(y_{new}|\theta)$について
パラメータの事後期待値$\mathrm{E}_{\mathrm{pos}}[\theta]$をプラグインしたときの平均損失
$-\mathrm{E}_{y_{new}}[\nu(y_{new}|\mathrm{E}_{\mathrm{pos}}[\theta])]$ を推定 （以下負号除きN倍）

＊まだ予想レベル
＊正則モデル限定 → νが依存するθによっては特異モデルでも成立？
＊一般化情報量規準GICをさらに一般化したものに対応

$$\mathrm{GPCIC} = \left\{ \sum_i \nu(y_i|\mathrm{E}_{\mathrm{pos}}[\theta]) - \sum_i \mathrm{Cov}_{\mathrm{pos}}[\nu(y_i|\theta), \log p(y_i|\theta)] \right\}$$

共同研究者・応用例・競争的資金

- 本研究は矢野恵佑氏との共同研究である
- ここでは着想に重点を置いて説明した →応用例は矢野氏のポスターを参照
- 科研費 基盤(C) 多様な予測に対応した情報量規準の開発：計算統計的アプローチ 伊庭（代表）/矢野（分担）