行動科学演習・数理行動科学研究演習『社会科学のためのベイズ統計モデリング』第７章

Hiroshi Hamada
Tohoku University

Jun, 2020
at Tohoku University

確率モデルの情報量

例：ブログへのアクセス

確率モデルはポアソン分布

$$\mathrm{Poisson}(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda}$$

λは平均パラメータ

予測分布を $\mathrm{Poisson}(x|\lambda = 3)$ とおく

情報量

サンプルの実現値が x であるときの情報量

$$
\begin{aligned}
I(X = x) = -\log p(X = x) &= -\log \left(\frac{\lambda^x}{x!} e^{-\lambda} \right) \\
&= -x \log \lambda + \log x! + \lambda
\end{aligned}
$$

$\lambda = 3$ の場合，$x = 3$ の情報量は

$$-3 \log 3 + \log 3! + 3 \approx 1.49$$

$\lambda = 10$ の場合，$x = 3$ の情報量は

$$-3 \log 10 + \log 3! + 10 \approx 4.88$$

情報量は，データに近いモデルのほうが小さい

自由エネルギーの具体例

パラメータがガンマ分布に従うと仮定

$$
\begin{aligned}
X &\sim \mathrm{Poisson}(\lambda) \\
\lambda &\sim \mathrm{Gamma}(a, b)
\end{aligned}
$$

周辺尤度（共役事前分布の恩恵）

$$
p(x^n) = \frac{\Gamma(\sum x_i + a)}{\Gamma(a) \prod x_i!} \frac{b^a}{(n+b)^{\sum x_i + 1}}
$$

自由エネルギーは周辺尤度を対数化して符号を反転した値

$$
\begin{aligned}
F_n &= -\log p(X^n = x^n) \\
&= -\log \left(\frac{\Gamma(\sum x_i + a)}{\Gamma(a) \prod x_i!} \frac{b^a}{(n+b)^{\sum x_i + 1}} \right) \\
&= -\log \Gamma \left(\sum x_i + a \right) + \log \Gamma(a) + \sum \log x_i! \\
&\quad - a \log b + \left(\sum x_i + a \right) \log(n+b)
\end{aligned}
$$

サンプルの実現値 $(3, 4, 2, 7, 8)$

$\lambda = 3$ のポアソン分布の自由エネルギーは,

$$F_{M_0} = 13.43$$

$a = 3, b = 1$ のガンマ分布を λ の事前分布とするポアソン分布の自由エネルギーは

$$F_{M_1} = 12.60$$

ただしベイズファクターによると決定的な差ではない

モデル評価とオーバーフィッティング

$$M_0 : X \sim \mathrm{Poisson}(3)$$
$$M_1 : X \sim \mathrm{Poisson}(\lambda), \lambda \sim \mathrm{Gamma}(3, 1)$$

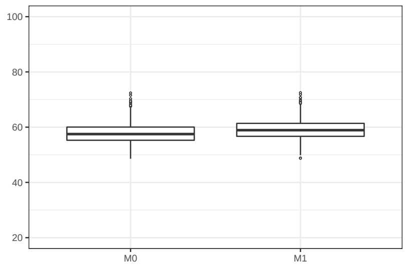

Figure: $\mathrm{Poisson}(\lambda^* = 3)$ が真の分布. M_0 と M_1 の自由エネルギーの分布の比較

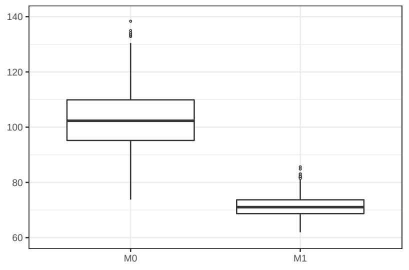

Figure: $\mathrm{Poisson}(\lambda^* = 6)$ が真の分布. M_0 と M_1 の自由エネルギーの分布の比較

自由エネルギーの推定値の計算

確率モデル:ポアソン分布，事前分布:ガンマ分布
逆温度 β の事後分布

$$
\begin{aligned}
p(\lambda|x^n)_\beta &= \frac{(n\beta+b)^{\sum x_i\beta+a}}{\Gamma\left(\sum x_i\beta+a\right)}\lambda^{\sum x_i\beta+a-1}e^{-\lambda(n\beta+b)} \\
&= \frac{b_n^{a_n}}{\Gamma(a_n)}\lambda^{a_n-1}e^{-b_n\lambda}
\end{aligned}
$$

$a_n = \sum x_i\beta + a, b_n = n\beta + b$ であるガンマ分布

WBIC の計算

$$\mathrm{WBIC} = \mathbb{E}_{p(\lambda|x^n)_\beta}[nL_n(\lambda)]$$

対数経験損失の事後分布による期待値

近似計算

$$\mathrm{WBIC}_{\mathrm{mcmc}} = \frac{1}{S}\sum_{s=1}^{S}\left(\sum_{i=1}^{n} - \log p(x_i|\theta_s)\right)$$

自由エネルギーのシミュレーション

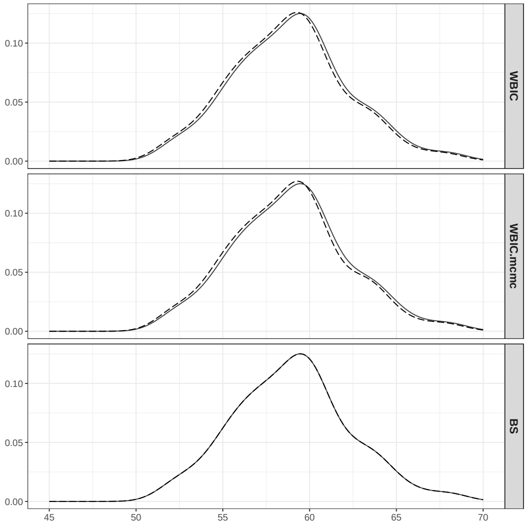

Figure: シミュレーション結果（実線が自由エネルギー，破線は各推定法による自由エネルギーの推定）

ブリッジ・サンプリング

周辺尤度の近似計算

- ナイーブ・モンテカルロ法

$$\hat{\theta}_i \sim \varphi(\theta), \qquad \hat{p}_{\mathsf{NM}}(x^n) = \frac{1}{R}\sum_{i=1}^{R} p(x^n|\hat{\theta}_i)$$

- 重点サンプリング法

$$\hat{\theta}_i \sim g(\theta), \qquad \hat{p}_{\mathsf{IS}}(x^n) = \frac{1}{R}\sum_{i=1}^{R} \left(p(x^n|\hat{\theta}_i)\frac{\varphi(\hat{\theta}_i)}{g(\hat{\theta}_i)} \right)$$

- 一般化調和平均サンプリング法

$$\hat{\theta}_j^* \sim p(\theta|x^n), \hat{p}_{GHM}(x^n) = \left(\frac{1}{R} \sum_{j=1}^{R} \left(\frac{1}{p(x^n|\theta_j^*)} \frac{g(\theta_j^*)}{\varphi(\theta_j^*)} \right) \right)$$

- ブリッジ・サンプリング法
  MCMC サンプルから計算可能

```
target += (1/log(N))*poisson_lpmf(X | lambda);
```

  逆温度を指定

汎化損失の推定

事後分布（事前分布はガンマ分布）

$$p(\lambda|x^n) = \frac{b_n^{a_n}}{\Gamma(a_n)} \lambda^{a_n - 1} e^{-b_n \lambda}$$

$a_n = \sum x_i + a, b_n = n + b$.

予測分布

$$p^*(x) = \int_0^\infty \underbrace{p(x|\lambda)}_{\text{確率モデル}} \underbrace{p(\lambda|x^n)}_{\text{事後分布}} d\lambda$$
$$= \frac{\Gamma(x + a_n)}{\Gamma(a_n) x!} \frac{b_n^{a_n}}{(1 + b_n)^{x + a_n}}.$$

真の分布が既知なので計算可能

$$G_n = \int -\log p^*(x) q(x) dx$$ 定義より

$$= \sum_{i=0}^{\infty} -\log p^*(x_i) \mathrm{Poisson}(x_i|\lambda^*)$$ ポアソン分布を代入

$$= \sum_{i=0}^{\infty} \frac{1}{x_i!} (\lambda^*)^{x_i} e^{-\lambda^*} [-\log \Gamma(x + a_n) + \log \Gamma(a_n) + \log x$$

$$- a_n \log b_n + (x + a_n) \log(1 + b_n)]$$

WAIC の計算

経験損失 T_n と汎関数分散 V_n の計算が困難な場合→
MCMC 利用

経験損失 T_n の MCMC による推定値

$$T_n \approx \frac{1}{n}\sum_{i=1}^{n}\left[-\log\left(\frac{1}{S}\sum_{s=1}^{S}p(x_i|\theta_s)\right)\right]$$

S は MCMC のサンプルの個数

汎関数分散 V_n の MCMC による推定値

$$V_n \approx \sum_{i=1}^{n} \left[\frac{1}{S-1} \sum_{s=1}^{S} \log p(x_i|\theta_s)^2 - \left(\frac{1}{S-1} \sum_{s=1}^{S} \log p(x_i|\theta_s) \right)^2 \right]$$

MCMC からの WAIC

$$\begin{aligned}
\mathrm{WAIC}_{\mathrm{mcmc}} =& \frac{1}{n}\sum_{i=1}^{n}\left[-\log\left(\frac{1}{S}\sum_{s=1}^{S}p(x_i|\theta_s)\right)\right] \\
&+\frac{1}{n}\sum_{i=1}^{n}\left[\frac{1}{S-1}\sum_{s=1}^{S}\log p(x_i|\theta_s)^2\right. \\
&\left.-\left(\frac{1}{S-1}\sum_{s=1}^{S}\log p(x_i|\theta_s)\right)^2\right]
\end{aligned}$$

Stan code

Stan 上で，以下の generated quantities ブロックを model ブロックの下にコピーして，データポイントごとに対数尤度を計算（poisson_WAIC.stan として保存）

```
generated quantities{
    real log_lik[N];
    for(n in 1:N){
        log_lik[n] = poisson_lpmf(X[n] | lambda);
        }
}
```

Stan code

Stan で生成した `log_lik` を `extract` 関数で取り出し,
関数 `waic_mcmc` の引数として渡す

```
waic_mcmc <- function(log_lik){
    T_n <- mean(-log(colMeans(exp(log_lik))))
    V_n_divide_n <- mean(apply(log_lik,2,var))
    waic <- T_n + V_n_divide_n
    return(waic)
}

model.waic <- stan_model("poisson_WAIC.stan")
fit.waic <- sampling(model.waic, data=list(N=n,X=x,a=a
      ,b=b))
log_lik <- rstan::extract(fit.waic)$log_lik
waic_mcmc(log_lik)
```

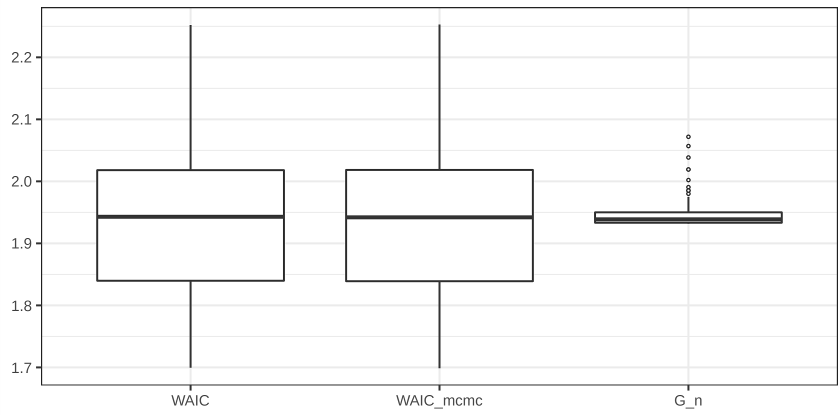

Figure: WAIC と汎化損失のシミュレーション結果

まとめ

- 自由エネルギーは複雑なモデルにペナルティが加わる
- ブリッジ・サンプリングによる推定は精度が良い
- モデルの良さの指標はサンプルの現れ方の変動の影響を受ける
- 解析的に分析できないモデルでも WAIC を MCMC サンプルから計算できる