Bridgeサンプリングを用いたベイズモデルの評価(仮)

専修大学大学院 文学研究科 M2 **北條大樹**

2017/10/15
IRT勉強会@専修大学

あなたの心に *Bridge Sampling* 私とあなたの Bayes Factor を求めよう

- Bridgeサンプリングを用いたベイズモデルの評価 –

専修大学大学院 文学研究科 M2 **北條大樹**

目次

- ベイズファクター
- 様々なBF算出方法
- Bridge Sampling
- WARP-III
- 実際の心理学研究での応用例
- まとめ

本スライドの主な文献 (上から順に多く引用)

- Gronau, Q. F., Sarafoglou, A., Matzke, D., Ly, A., Boehm, U., Marsman, M., ... & Steingroever, H. (2017). A tutorial on bridge sampling. *arXiv preprint arXiv:1703.05984v2*.

- Gronau, Q. F., Wagenmakers, E. J., Heck, D. W., & Matzke, D. (2017). A Simple Method for Comparing Complex Models: Bayesian Model Comparison for Hierarchical Multinomial Processing Tree Models using Warp-III Bridge Sampling.

- Wang, L., & Meng, X. L. (2016). Warp bridge sampling: the next generation. *arXiv preprint arXiv:1609.07690*.

- Meng, X. L., & Schilling, S. (2002). Warp bridge sampling. *Journal of Computational and Graphical Statistics*, *11*(3), 552-586.

ベイズファクター

- **パラメータ推定**の文脈で、我々の興味のあるパラメータθの不確かさを観測データyから求めるとき

事後分布　尤度　事前分布

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

周辺尤度

と表される。

- **モデル比較**の文脈では、**m** 個の候補モデルがあり、データ **y** を所与としたときの **i** 番目のモデルの相対的な最もらしさを表すモデル事後確率は、

$$p(M_i|y) = \frac{p(y|M_i)p(M_i)}{\sum_{j=1}^{m} p(y|M_j)p(M_j)}$$

事後モデル確率　**Σ 周辺尤度×事前モデル確率**

事後モデル確率 = 周辺尤度×事前モデル確率 / 候補すべての周辺尤度×事前モデル確率

ベイズファクター

- もしここでm=2だったら… (比較したいモデルが2つのみ)

$$p(M_1|y) = \frac{p(y|M_1)p(M_1)}{\sum_{j=1}^{m} p(y|M_j)p(M_j)} \qquad p(M_2|y) = \frac{p(y|M_2)p(M_2)}{\sum_{j=1}^{m} p(y|M_j)p(M_j)}$$

同じ

- つまりオッズで表すことが可能

事後オッズ **ベイズファクター** **事前オッズ**

$$\frac{p(M_1|y)}{p(M_2|y)} = \frac{p(y|M_1)}{p(y|M_2)} \times \frac{p(M_1)}{p(M_2)}$$

- 次のようにも

ベイズファクター

$$BF_{12} = \frac{p(y|M_1)}{p(y|M_2)} = \frac{\frac{p(M_1|y)}{p(M_2|y)}}{\frac{p(M_1)}{p(M_2)}}$$

事後オッズ

事前オッズ

ベイズファクターとは、データによって与えられたモデルM_2に比してモデルM_1を支持する程度（オッズ）の変化を表す

ベイズファクターって何?

- ある者は､､､

“**standard Bayesian solution** to the hypothesis testing and model selection problems” (Lewis & Rftery, 1997, p648)

- またある者は､､､

“**the primary tool used in Bayesian inference** for hypothesis testing and model selection” (Berger, 2006, p.378)

- またある**OKADA**は､､､

“ベイズの枠組みでのモデル比較を考える際に、ベイズファクターは最も順当な考え方である。”

それなのになぜBFを使わないのか?

Q. なぜBFを使わないのか?

A. 使わないというよりは、**使えない(解析的に周辺尤度を計算できない)**ケースが多い。✌('ω'✌) ∫∫∫ ∫∫∫ ∫∫∫ ∫∫∫ ∫∫∫(✌'ω')✌多重積分ワッショイ

どんなケースで使えないのか?

- 非線形モデル
- パラメータ数が多い場合(特に階層モデル)

⇒むしろ、ベイズファクターを使えるモデルを探すほうが、使えないモデルを挙げていくより早いかも。。。

じゃあ、結局BFは求められないのか?

我々の最後の武器は何だろう。。。

- そう、MCMC(;´Д｀)ﾊｧﾊｧ

Mr.Unadon @MrUnadon · 7月8日
ああ、、あ！あああああああ。僕の終電！！あああああああ。あ。MCMC(*´Д`)ﾊｧﾊｧ

Muto, Hiroyuki @mutopsy · 7月21日
MCMCはなぜこんなに(*´Д`)ﾊｧﾊｧするのだろうか。

Hiroshi Shimizu @simizu706 · 9月6日
合宿行けなくて悔しいから部屋でMCMC(*´Д`)ﾊｧﾊｧする

Hiroshi Shimizu @simizu706 · 9月5日
今日はついポテチとコーラを買ってしまったんだ。Stanよ，罪深い僕をどうぞMCMC(*´Д`)ﾊｧﾊｧ

kazutan v3.4.2さんがリツイートしました
numba shushi @NSushi · 2016年12月15日
病気でもMCMC(*´Д`)ﾊｧﾊｧ！

Nさんと他2人さんがリツイートしました
Hiroshi Shimizu @simizu706 · 2016年12月12日
Stan(*´Д`)ﾊｧﾊｧ　（もはやMCMCを走らせなくてもコード書いてるだけで興奮できる体に

平川　真 @hirakawamakoto · 2016年12月3日
明日はMCMC(*´Д`)ﾊｧﾊｧしたいから明日までの依頼仕事をあげたら、明日の仕事が降ってきたでござる

kosugitti(こむぎっち) @kosugitti · 9月28日
MCMC(；´Д｀)ﾊｧﾊｧ　次の次元へのソースを手に入れたぞ

kosugitti(こむぎっち) @kosugitti · 6月20日
MCMC(；´Д｀)ﾊｧﾊｧ！答えは2000秒後！

Muto, Hiroyukiさんと他1人さんがいいねしました
kosugitti(こむぎっち) @kosugitti · 5月28日
MCMC(；´Д｀)ﾊｧﾊｧ

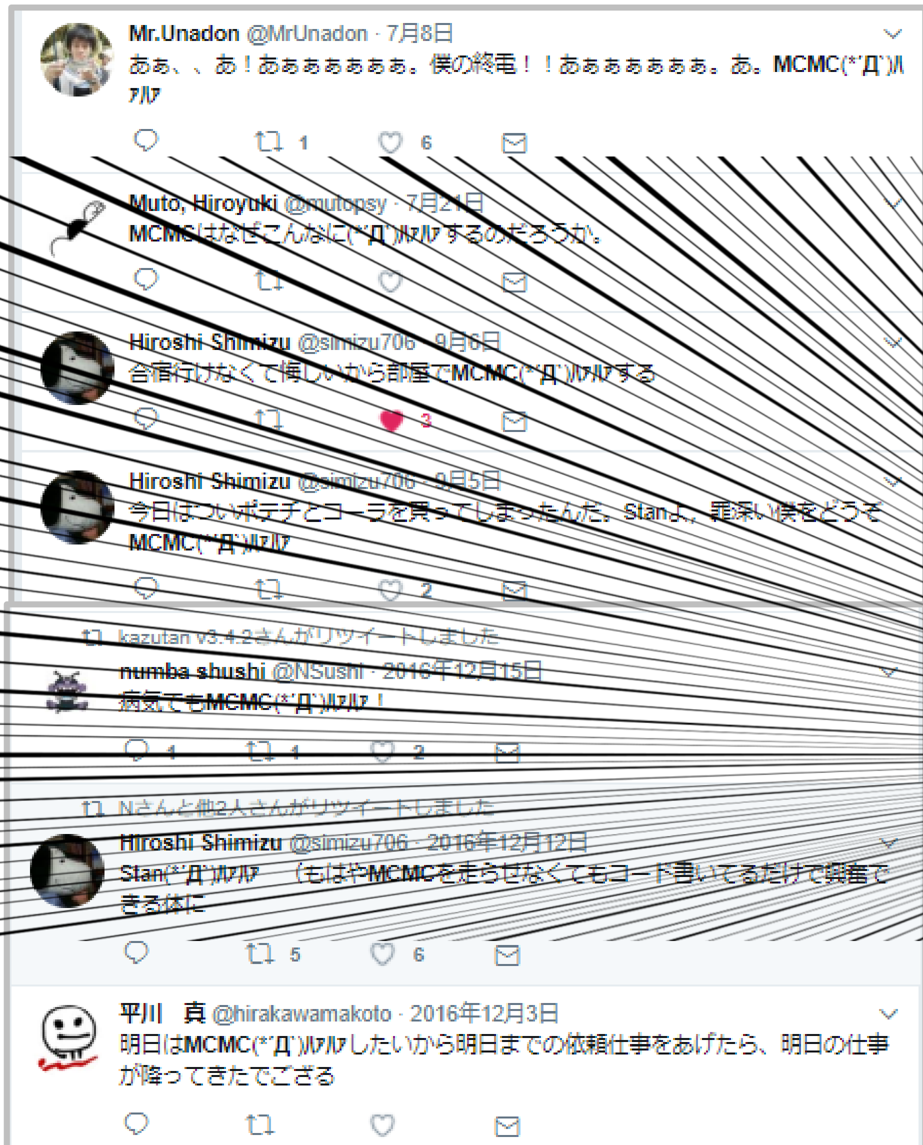
Hiroshi Shimizuさんと他1人さんがリツイートしました
kosugitti(こむぎっち) @kosugitti · 3月14日
MCMC(；´Д｀)ﾊｧﾊｧを言い出したのは私です。念のため。

kosugitti(こむぎっち) @kosugitti · 3月14日
会議をしながらMCMC(；´Д｀)ﾊｧﾊｧ(；´Д｀)ﾊｧﾊｧ

MCMCでこれを解決しよう。

我々の最後の武器は何だろう。。。

- そう、**MCMC(;´Д｀)ﾊｧﾊｧ**



MCMCでこれを解決しよう。

MCMCによる様々なBF(周辺尤度)への近似法

1. Naive Monte Carlo Estimator
2. Importance Sampling Estimator (重点サンプリング法)
3. Generalized Harmonic Mean Estimator (一般化調和平均法??)
4. **Bridge Sampling** (ブリッジサンプリング)
5. **WARP-III** bridge sampling or **WARP-U** bridge sampling

今回は、4. 5. の手法を重点的に紹介

1,2,3の詳しい話については、Gronau et al. (2017)を参照し、その後原著へ

再確認: 求めたいもの

- **BFを求めたい、すなわち、周辺尤度を求めれば勝ち**

$$BF_{12} = \frac{p(y|M_1)}{p(y|M_2)}$$

- 周辺尤度って?

$$p(y|M) = \int p(y|\theta, M)p(\theta|M)d\theta$$

周辺尤度 **尤度** **事前分布**

興味のあるモデルで
yが観測される確率

= 事前分布を通した尤度の積分

= θを所与とするときの尤度の重み付き平均
(重みというのは事前分布θの最もらしさのこと)

- 次のようにも…

$$p(y|M) = \mathbb{E}_{prior}\left(p(y|\theta, M)\right)$$

$\mathbb{E}_{prior}$は事前分布の期待値を表す。

速習 Naive Monte Carlo Estimator

- 周辺尤度を求めたい。ただ、ここでどのモデルのとかはない

$$p(y|M) = \mathbb{E}_{prior}\big(p(y|\theta, M)\big) \quad \Rightarrow \quad p(y) = \mathbb{E}_{prior}\big(p(y|\theta)\big)$$

- これを近似するには、**θにおける事前分布からのN個のサンプルで尤度を評価**し、**その結果を平均**すればよい。

$$p(y) = \mathbb{E}_{prior}\big(p(y|\theta)\big)$$

$$\Downarrow$$

$$\hat{p}_1(y) = \frac{1}{N}\Sigma_{i=1}^{N} p\left(y\middle|\tilde{\theta}_i\right) \qquad \tilde{\theta}_i \sim p(\theta)$$

尤度の平均　　**事前分布からのサンプル**

- コインの例で考えた具体例は、Gronau et al. (2017) p8-9 にて

この方法が使えるのは、事前分布と事後分布の形が似ていて、オーバーラップしていることが条件。

速習 Importance Sampling Estimator

- 事後分布と'似ていない'事前分布を使うのではなく、importance density(**重点密度**) $g_{IS}(\theta)$を導入。

- 重点密度の特徴は、**尤度の高いところのθを重点的にサンプリングし、尤度の低いところはあまりサンプリング**しない。
- 以下のようにして$g_{IS}(\theta)$を導入できる。

$$
\begin{aligned}
p(y) &= \int p(y|\theta)p(\theta)d\theta \\
&= \int p(y|\theta)p(\theta)\frac{g_{IS}(\theta)}{g_{IS}(\theta)}d\theta \\
&= \int \frac{p(y|\theta)p(\theta)}{g_{IS}(\theta)}g_{IS}(\theta)\,d\theta \\
&= \mathbb{E}_{g_{IS}(\theta)}\left(\frac{p(y|\theta)p(\theta)}{g_{IS}(\theta)}\right)
\end{aligned}
$$

速習 Importance Sampling Estimator

- 実際に推定するときは、、、

$$\hat{p}_2(y) = \frac{1}{N}\Sigma_{i=1}^{N}\frac{p(y|\tilde{\theta}_i)p(\tilde{\theta}_i)}{g_{IS}(\tilde{\theta}_i)} \qquad \tilde{\theta}_i \sim g_{IS}(\theta)$$

調整された尤度の平均　**重点密度からのサンプル**

- 適切な重点密度はどのような特徴をもつべき?
  - 評価が簡単
  - 事後分布と同一の定義域をもつ
  - 事後分布によく似ている
  - **事後分布よりも厚い裾(fatter tail)を持つこと**

速習 Importance Sampling Estimator

- どのように重点密度を決めるか?
  - コインの例では、事後分布はβ分布で表せる。そのため適切な重点密度を**β分布と一様分布の混合分布で定義**しようとしている。
  - この際、混合分布におけるベータ分布は、事後分布からのサンプルを使ってモーメント推定をして得たモーメントを用いている。
    - モーメント:β分布でいえばα、βパラメータを指し、正規分布でいえば平均と標準偏差パラメータをさす。
  - 最後に、ベータ分布と一様分布の混合具合を決めるチューニング(重み)パラメータをrを用意し、これによってどっちの分布を多めにチューニングするかを決めて、重点密度分布からのサンプリングを行う。

※ rの調節が大変。

速習 Generalized Harmonic Mean Estimator

- Importance Samplingとは対照的で、thinner tailになるようにする。
- そして、事後分布からのサンプルを使って計算するように変更。

$$
\begin{aligned}
\frac{1}{p(y)} &= \int \frac{1}{p(y)} g_{IS}(\theta) d\theta \\
&= \int \frac{p(\theta|y)}{p(y|\theta)p(\theta)} g_{IS}(\theta) d\theta \\
&= \int \frac{g_{IS}(\theta)}{p(y|\theta)p(\theta)} p(\theta|y) d\theta \\
&= \mathbb{E}_{\boldsymbol{post}}\left(\frac{g_{IS}(\theta)}{p(y|\theta)p(\theta)}\right)
\end{aligned}
$$

$$
p(y) = \left(\mathbb{E}_{\boldsymbol{post}}\left(\frac{g_{IS}(\theta)}{p(y|\theta)p(\theta)}\right)\right)^{-1}
$$

参考:Importance Sampling

$$
\mathrm{p(y)} = \mathbb{E}_{g_{IS}(\theta)}\left(\frac{p(y|\theta)p(\theta)}{g_{IS}(\theta)}\right)
$$

- 逆数になっている
- 事後分布からのサンプルを使う

速習 Generalized Harmonic Mean Estimator

- 実際に推定するときは、、、

注: θ_j^*と$\tilde{\theta}_i$は、違う分布からサンプリングされていることに注意

重点密度

$$\hat{p}_3(y) = \left(\frac{1}{N}\Sigma_{j=1}^{N}\frac{g_{IS}\left(\theta_j^*\right)}{p\left(y\middle|\theta_j^*\right)p\left(\theta_j^*\right)}\right)^{-1} \qquad \theta_j^* \sim p(\theta|y)$$

尤度　事前分布

事後分布からのサンプル

- 適切な重点密度はどのような特徴をもつべき?
  - 評価が簡単
  - 事後分布と同一の定義域をもつ
  - 事後分布によく似ている
  - **事後分布よりも薄い裾(thinner tail)を持つこと**

- コインの例では、事後サンプルが0-1の範囲しか取らないので、サンプルをまず、プロビット変換して、-∞~∞(正規分布)の範囲で考えられるようにして、Importance Sampling 同様にモーメント推定を行う。

速習 Generalized Harmonic Mean Estimator

重点密度

$$\hat{p}_3(y) = \left(\frac{1}{N} \Sigma_{j=1}^{N} \frac{g_{IS}\left(\theta_j^*\right)}{p\left(y \middle| \theta_j^*\right) p\left(\theta_j^*\right)} \right)^{-1}$$

尤度　事前分布

$\theta_j^* \sim p(\theta|y)$

事後分布からのサンプル

注: θ_j^*と$\tilde{\theta}_i$は、違う分布からサンプリングされていることに注意

- プロビット変換をして考える場合

重点密度

$$\hat{p}_3(y) = \left(\frac{1}{N} \Sigma_{j=1}^{N} \frac{\frac{1}{\hat{\sigma}} \phi\left(\frac{\xi_j^* - \hat{\mu}}{\hat{\sigma}}\right)}{p\left(y \middle| \Phi(\xi_j^*)\right) \phi\left(\xi_j^*\right)} \right)^{-1}$$

尤度　事前分布

$\xi_j^* = \Phi^{-1}\left(\theta_j^*\right)\ and\ \theta_j^* \sim p(\theta|y)$

事後分布からのサンプルθ_j^*をプロビット変換したサンプルξ_j^*

と表現できる

ここまでまとめ

- Importance Samplingでは、**重点密度分布からのサンプル**$\tilde{\theta}_i$**を利用**して、周辺尤度を求めようとする。

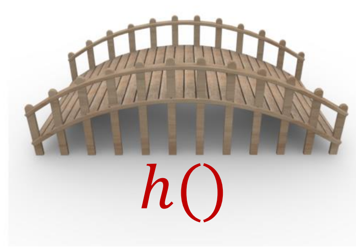
- Generalized Harmonic Mean Estimatorでは、上記の逆数を考えることで**事後分布からのサンプル**θ_j^*から最適な重点密度分布を探して周辺尤度を求めようとした。

- ただ、2つの方法の欠点は、分布の裾に強い仮定(厚いか・薄いか)を置いており、これが高次元空間では満たすことができない。

- どうする?

⇒ようやく Bridge Sampler が登場。

$\tilde{\theta}_i$



θ_j^*

Bridge Sampling Estimator

WARP-I

WARP-II

WARP-III

WARP-U

Bridge Sampling

- ここで適当に比を書いてみる。

尤度　事前分布 Bridge関数 提案分布

$$1 = \frac{\int p(y|\theta)p(\theta)h(\theta)g(\theta)d\theta}{\int p(y|\theta)p(\theta)h(\theta)g(\theta)d\theta}$$

尤度　事前分布 Bridge関数 提案分布

- なんとなく周辺尤度$p(y)$を両辺に掛けてみる。

$$p(y) = \frac{\int p(y|\theta)p(\theta)h(\theta)g(\theta)d\theta}{\int \frac{p(y|\theta)p(\theta)}{p(y)} h(\theta)g(\theta)d\theta}$$

$$= \frac{\int p(y|\theta)p(\theta)h(\theta) \overbrace{g(\theta)}^{\text{提案分布}} d\theta}{\int h(\theta)g(\theta) \underbrace{p(\theta|y)}_{\text{事後分布}} d\theta}$$

Bridge Sampling

- つづき

$$p(y)=\frac{\int p(y|\theta)p(\theta)h(\theta)g(\theta)d\theta}{\int \frac{p(y|\theta)p(\theta)}{p(y)}h(\theta)g(\theta)d\theta}$$

提案分布

$$=\frac{\int p(y|\theta)p(\theta)h(\theta)\quad g(\theta)d\theta}{\int h(\theta)g(\theta)\quad p(\theta|y)d\theta}$$

事後分布

$$=\frac{\mathbb{E}_{g(\theta)}(p(y|\theta)p(\theta)h(\theta))}{\mathbb{E}_{post}(h(\theta)g(\theta))}$$

- 実際の推定では、

Bridge関数

$$\hat{p}(y)=\frac{\frac{1}{N_2}\Sigma_{i=1}^{N_2}\left(p(y|\tilde{\theta}_i)p(\tilde{\theta}_i)h(\tilde{\theta}_i)\right)}{\frac{1}{N_1}\Sigma_{i=1}^{N_1}\left(h(\theta_j^*)g(\theta_j^*)\right)}$$

Bridge関数　提案分布

$\tilde{\theta}_i \sim g(\theta)$

提案分布からのサンプル

$\theta_j^* \sim p(\theta|y)$

事後分布からのサンプル

Bridge Sampling

- つづき

$$p(y)=\frac{\int p(y|\theta)p(\theta)h(\theta)g(\theta)d\theta}{\int \frac{p(y|\theta)p(\theta)}{p(y)}h(\theta)g(\theta)d\theta}$$

提案分布

$$=\frac{\int p(y|\theta)p(\theta)h(\theta)\quad g(\theta)d\theta}{\int h(\theta)g(\theta)\quad p(\theta|y)d\theta}$$

事後分布

$$=\frac{\mathbb{E}_{g(\theta)}\left(p(y|\theta)p(\theta)h(\theta)\right)}{\mathbb{E}_{post}\left(h(\theta)g(\theta)\right)}$$

- 実際の推定では、

Bridge関数

$$\hat{p}(y)=\frac{\frac{1}{N_2}\sum_{i=1}^{N_2}\left(p(y|\tilde{\theta}_i)p(\tilde{\theta}_i)h(\tilde{\theta}_i)\right)}{\frac{1}{N_1}\sum_{i=1}^{N_1}\left(h(\theta_j^*)g(\theta_j^*)\right)}$$

Bridge関数　提案分布

$\tilde{\theta}_i \sim g(\theta)$

提案分布からのサンプル

$\theta_j^* \sim p(\theta|y)$

事後分布からのサンプル

おわかりいただけただろうか?

提案分布はImportance Samplingのあれ

- **Importance Sampling**

$$\hat{p}_2(y) = \frac{1}{N}\Sigma_{i=1}^{N}\frac{p(y|\tilde{\theta}_i)p(\tilde{\theta}_i)}{g_{IS}(\tilde{\theta}_i)} \qquad \tilde{\theta}_i \sim g_{IS}(\theta)$$

調整された尤度の平均　重点密度からのサンプル

- **Generalized Harmonic Mean Estimator**

重点密度

$$\hat{p}_3(y) = \left(\frac{1}{N}\Sigma_{j=1}^{N}\frac{g_{IS}(\theta_j^*)}{p(y|\theta_j^*)p(\theta_j^*)}\right)^{-1} \qquad \theta_j^* \sim p(\theta|y)$$

尤度　事前分布

事後分布からのサンプル

- **Bridge Sampling**

Bridge関数

$$\hat{p}(y) = \frac{\frac{1}{N_2}\Sigma_{i=1}^{N_2}\left(p(y|\tilde{\theta}_i)p(\tilde{\theta}_i)h(\tilde{\theta}_i)\right)}{\frac{1}{N_1}\Sigma_{i=1}^{N_1}\left(h(\theta_j^*)g(\theta_j^*)\right)}$$

Bridge関数　提案分布

$\tilde{\theta}_i \sim g(\theta)$　$\theta_j^* \sim p(\theta|y)$

提案分布からのサンプル　事後分布からのサンプル

最適なBridge関数(Meng & Wong, 1996)

- これが最適。相対的なMSE(平均二乗誤差)を最小にできる

$$h(\theta) = C \cdot \frac{1}{s_1 p(y|\theta)p(\theta) + s_2 p(y) g(\theta)}$$

尤度　事前分布　周辺尤度 提案分布

ここで、$s_1 = \frac{\mathrm{N}_1}{N_2+N_1}, s_2 = \frac{\mathrm{N}_2}{N_2+N_1}$, Cは定数

- ここで問題がある。
  - 周辺尤度$p(y)$を評価するため$h(\theta)$を導入したのに、その最適関数に周辺尤度$p(y)$が入っている。

- **周辺尤度を求めたいのに周辺尤度を求めないと周辺尤度がわからない。**

⇒どうする?

何回もこれを計算して収束させたものを使おう

- t回目の最適関数$h(\theta)$をt+1回目の周辺尤度$p(\hat{y})$の式に代入

$$h(\theta) = C \cdot \frac{1}{s_1 p(y|\theta)p(\theta) + s_2 \hat{p}(y)^t g(\theta)}$$

$$\hat{p}(y)^{t+1} = \frac{\frac{1}{N_2}\Sigma_{i=1}^{N_2} \frac{p(y|\tilde{\theta}_i)p(\tilde{\theta}_i)}{s_1 p(y|\tilde{\theta}_i)p(\tilde{\theta}_i) + s_2 \hat{p}(y)^t g(\tilde{\theta}_i)}}{\frac{1}{N_1}\Sigma_{i=1}^{N_1} \frac{g(\theta_j^*)}{s_1 p(y|\theta_j^*)p(\theta_j^*) + s_2 \hat{p}(y)^t g(\theta_j^*)}}$$

- そして、整理した結果

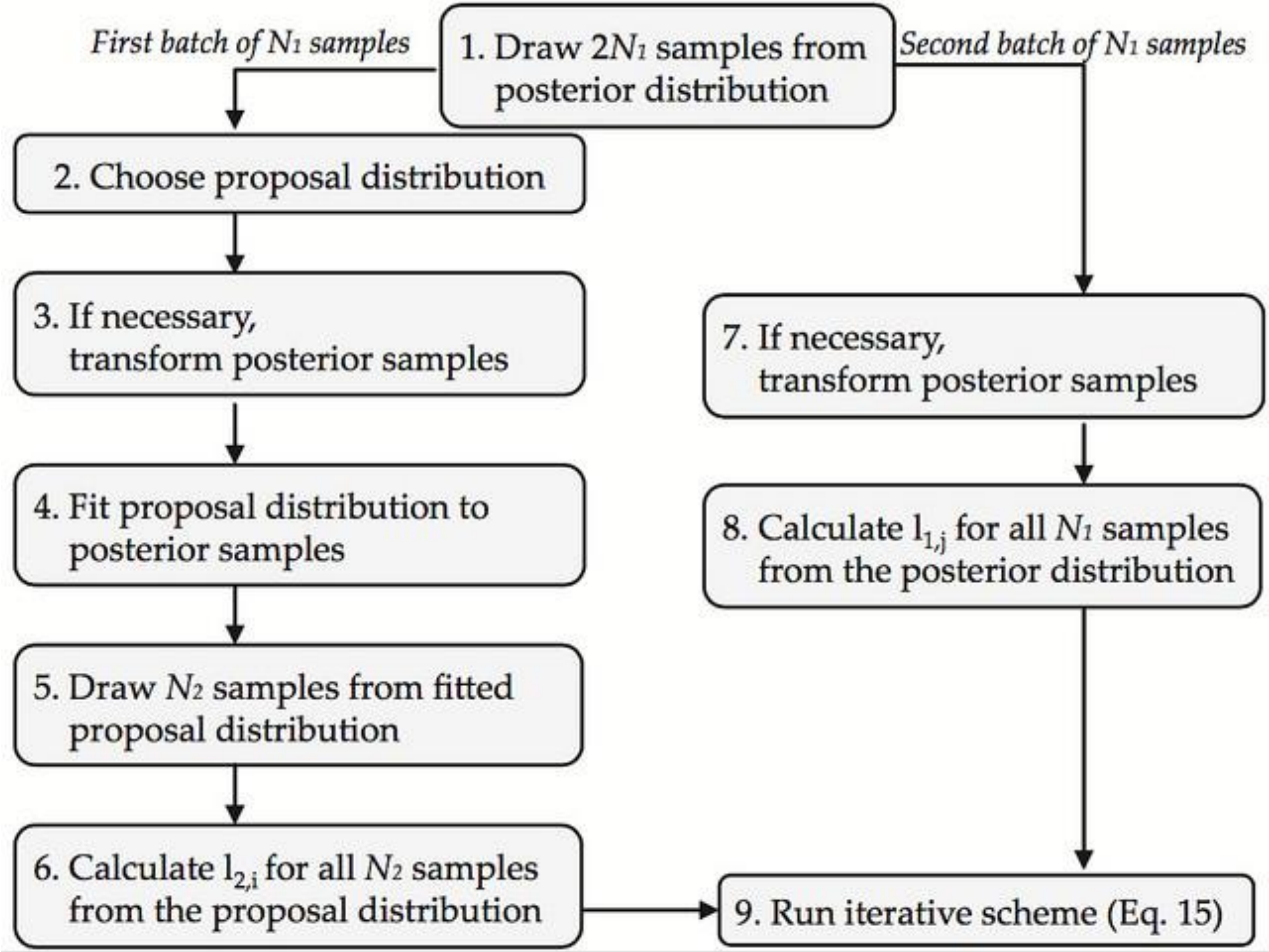
$$s_1 = \frac{\mathrm{N}_1}{N_2 + N_1}, s_2 = \frac{\mathrm{N}_2}{N_2 + N_1} \quad l_{1,j} = \frac{p(y|\theta_j^*)p(\theta_j^*)}{g(\theta_j^*)} \quad l_{2,i} = \frac{p(y|\tilde{\theta}_i)p(\tilde{\theta}_i)}{g(\tilde{\theta}_i)}$$

$$\hat{p}_4(y)^{t+1} = \frac{\frac{1}{N_2}\Sigma_{i=1}^{N_2} \frac{l_{2,i}}{s_1 l_{2,i} + s_2 \hat{p}_4(y)^t}}{\frac{1}{N_1}\Sigma_{i=1}^{N_1} \frac{1}{s_1 l_{1,j} + s_2 \hat{p}_4(y)^t}}$$

事後分布から
のサンプル
$\theta_j^* \sim p(\theta|y)$

提案分布から
のサンプル
$\tilde{\theta}_i \sim g(\theta)$

計算の手順


First batch of N1 samples
1. Draw 2N1 samples from posterior distribution
Second batch of N1 samples
2. Choose proposal distribution
3. If necessary, transform posterior samples
4. Fit proposal distribution to posterior samples
5. Draw N2 samples from fitted proposal distribution
6. Calculate l2,i for all N2 samples from the proposal distribution
7. If necessary, transform posterior samples
8. Calculate l1,j for all N1 samples from the posterior distribution
9. Run iterative scheme (Eq. 15)

手順

1. 推定した事後分布から2×N個のサンプルを抽出してくる。
   - ex) コインを投げた結果が12個あったなら24個サンプルを抽出
   - 半分(奇数番目)をN_1、半分(偶数番目)をN_1に分けておく。
2. 提案分布を選ぶ
   - ex) (多変量)正規分布
3. N_1サンプルを正規分布(提案分布)に合うように(今回は)プロビット変換する
4. 3.で変換した者を利用して正規分布のモーメント(平均･標準偏差)を推定する。
5. 4.の正規分布からN_2サンプルを生成する。
6. そのサンプルから$l_{2,i}$を計算する(R等で簡単にできる)。
7. 一番はじめに使っていない方のN_1のサンプルを(今回は)プロビット変換する。
8. 事後分布を用いて、$l_{1,j}$を計算する(6.と同様に)。
9. ブリッジサンプリングの更新を始める。

Bridge Samplingの弱点

1. 推定した事後分布から2×N個のサンプルを抽出してくる。
   - ex) コインを投げた結果が12個あったなら24個サンプルを抽出
   - 半分(奇数番目)をN_1、半分(偶数番目)をN_1に分けておく。
2. **提案分布を選ぶ**
   - **ex) (多変量)正規分布**
3. N_1サンプルを正規分布(提案分布)に合うように(今回は)プロビット変換する
4. 3.で変換した者を利用して正規分布のモーメント(平均・標準偏差)を推定する。
5. 4.の正規分布からN_2サンプルを生成する。
6. そのサンプルから$l_{2,i}$を計算する(R等で簡単にできる)。
7. 一番はじめに使っていない方のN_1のサンプルを(今回は)プロビット変換する。
8. 事後分布を用いて、$l_{1,j}$を計算する(6.と同様に)。
9. ブリッジサンプリングの更新を始める。

Bridge Samplingの弱点

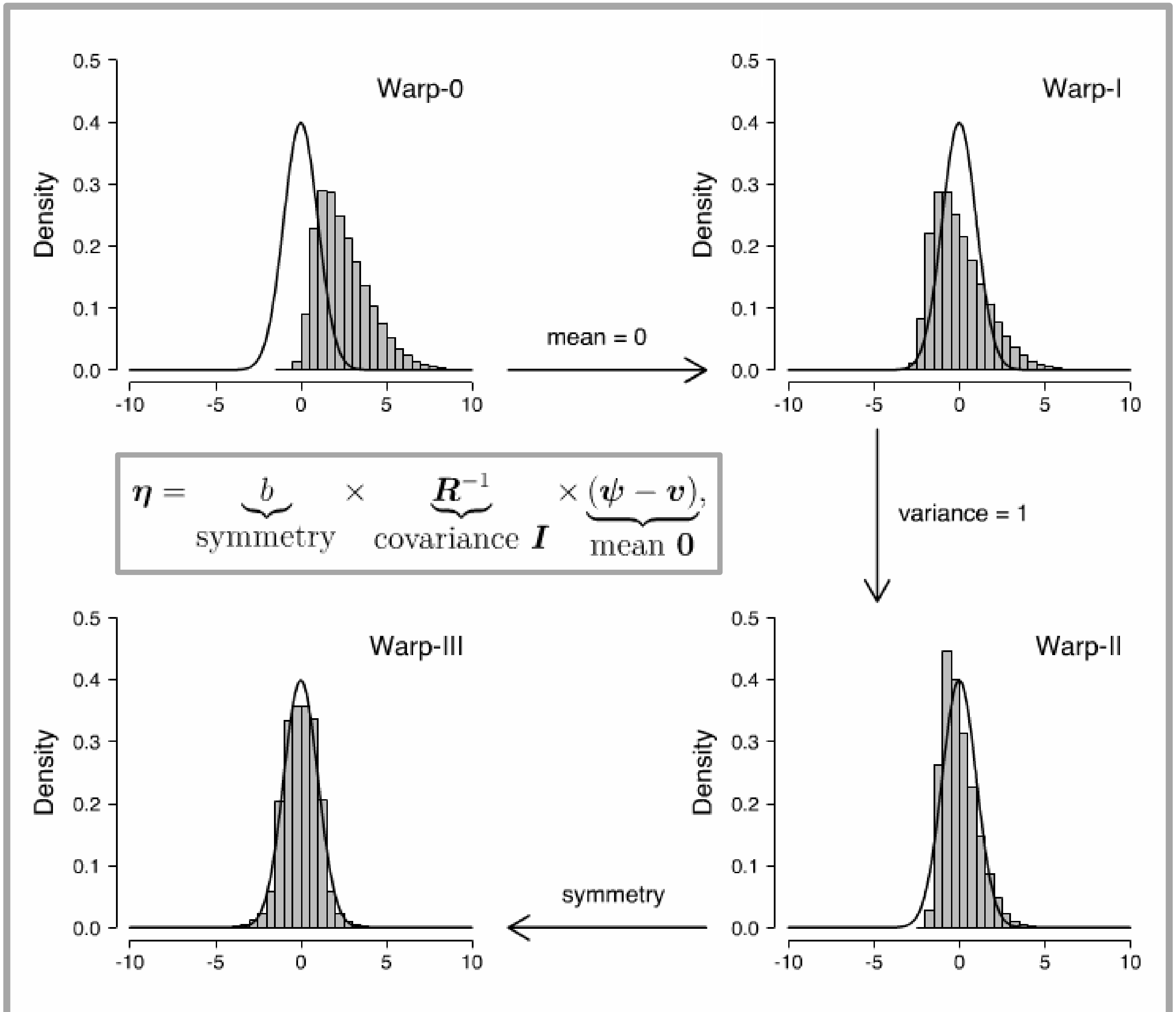
- 他の近似方法よりも仮定は緩いが、**事後分布と提案分布が似ていて、オーバーラップしている必要**がある。

- これがかなり大事で、ダメだと推定できなくなってしまう。

- どうしようか?
  - **WARP-III Bridge Sampling**
    ⇒mean/scale/skewness(歪度)を調整する方法
  - **WARP-U Bridge Sampling**
    ⇒混合分布のようなmulti-modal(多峰)な分布でも上記のような調整を行う方法

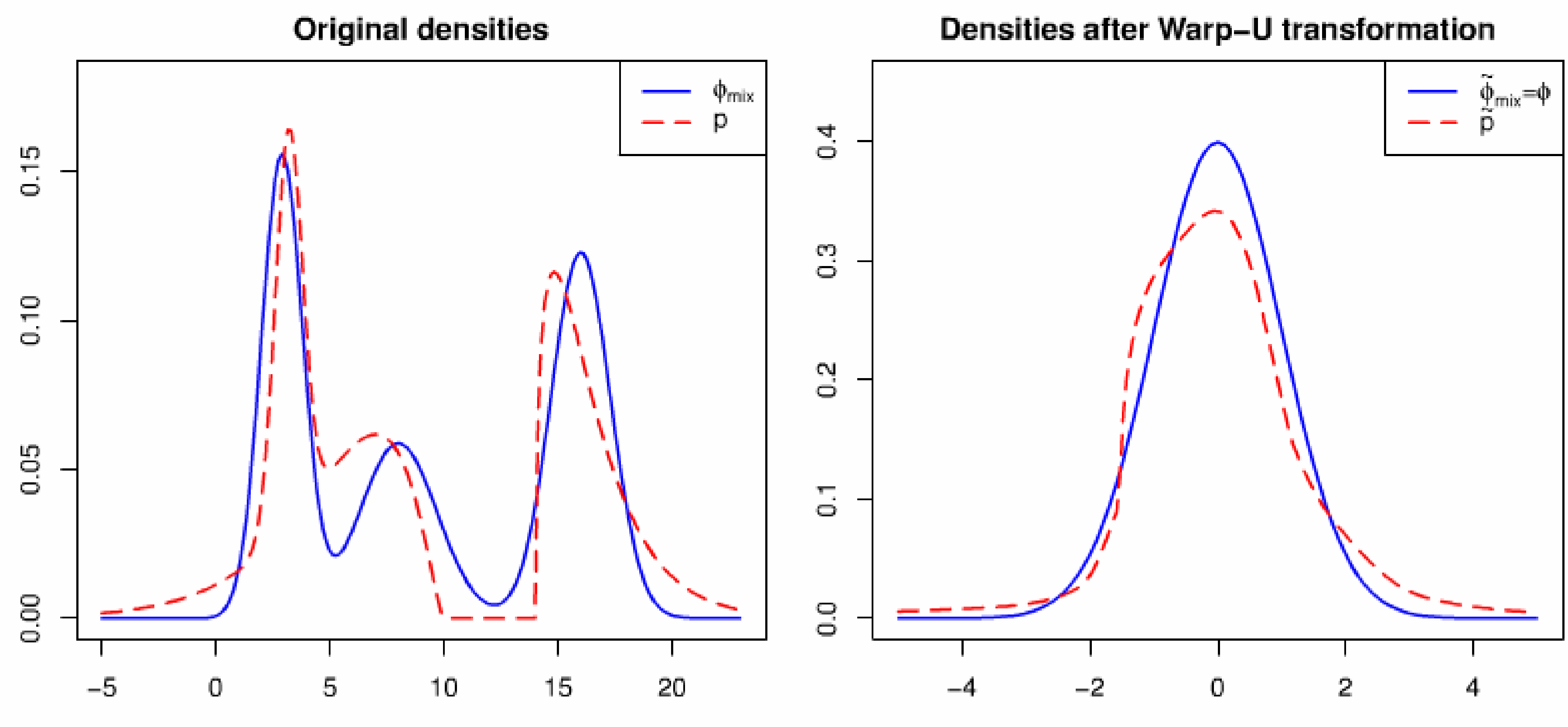
WARP-III Gronau et al. (2017) Fig.2

- ヒストグラムが事後分布。実線が標準正規分布(提案分布)



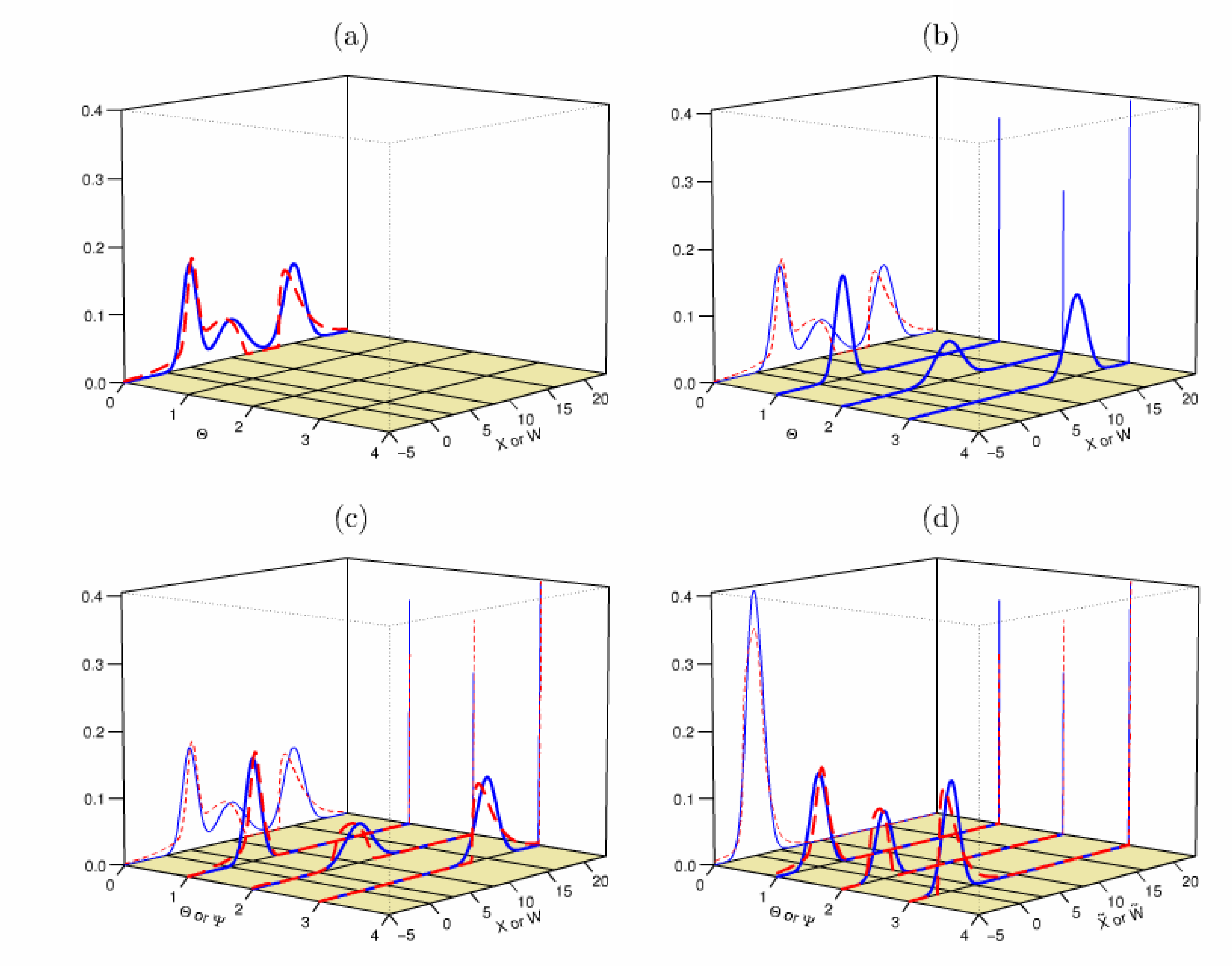
この方法で大事なことは正規化定数を維持したまま変換を行っている。

WARP-U Wang & Meng (2016) Fig.3-4



この方法で大事なことは正規化定数を維持したまま変換を行っている。

WARP-U Wang & Meng (2016) Fig.3-4



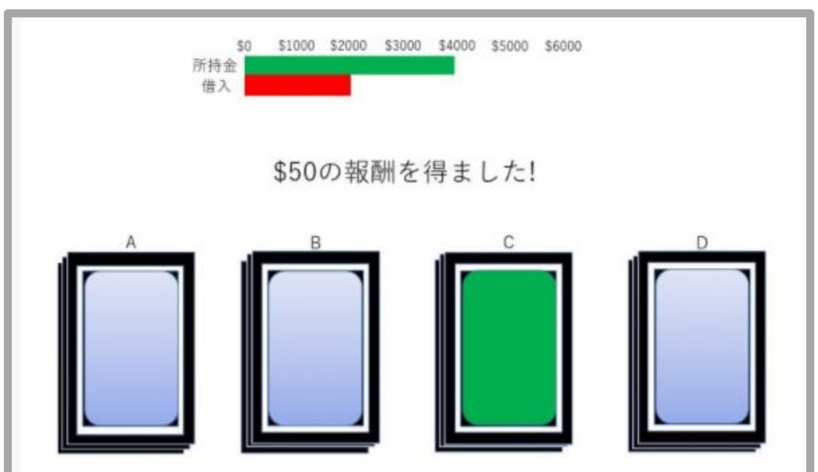
この方法で大事なことは正規化定数を維持したまま変換を行っている。

Bridge Samplingを用いた実際の研究

BridgeSamplingを用いた心理系論文

- IGT課題でのモデル比較 Gronau et al.(2017)
  - Importance Samplingで算出した対数周辺尤度とBridgeSamplingで算出した対数周辺尤度の一致率を示した研究
  - そして、BridgeSamplingを用いて、個人差を考慮した階層モデルにおける周辺尤度を算出した。

- MPTモデルでのモデル比較 Gronau et al.(2017)
  - 人の認知過程をモデリングしたMPTモデルで、BridgeSamplingを用いてBFを算出した研究。
  - これまで不可能だった階層MPTモデルにおけるBFを算出し、異なるパラメータ制約のモデルを8つ作り、比較を行った。



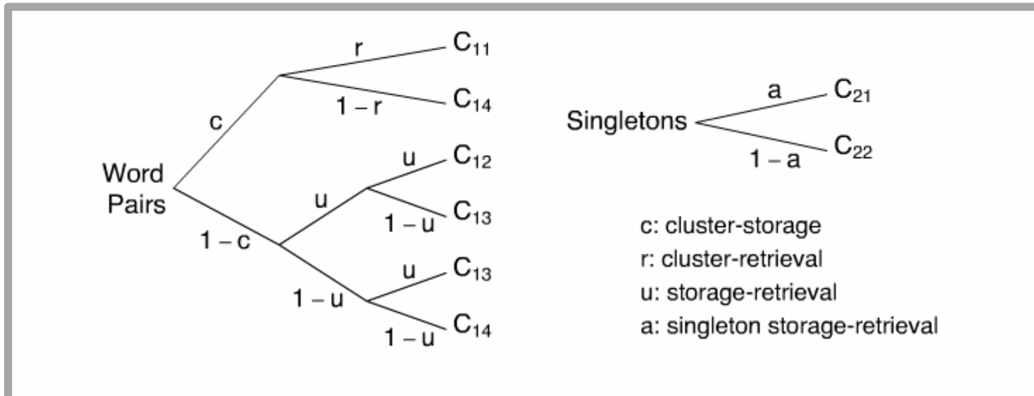


Figure 1. The pair-clustering MPT. Available at https://tinyurl.com/yb7bma4e under

↑申し訳程度のIRTっぽい要素

今日のまとめ

- BridgeSamplingは昔からある方法だが、WARPという計算方法?によってオーバーラップ力を挙げて、BFの計算が正確にできるようになった。

- StanとJAGSでWARP-III Bridge Samplingを行うパッケージ"BridgeSampling"がこの前公開された(WARP-Uは搭載されていない模様)。

- 今回は、資料作成が遅くなってしまい、実装方法がまとめられませんでしたが、興味がある方がいればご説明いたします。

- パッケージの開発者達の講演をこの前の国際学会で聞きましたが、**Stanコードが書ける人なら誰でもOK**といっていたので、皆さんならBridge Samplingができないわけがない。

- BF出してみたいデータがある。なんかやり方よくわからねぇ…

⇒いつでも共同研究のご相談お待ちしています。

2017/10/15
ベイズIRT勉強会@専修大学

あなたの心に *Bridge Sampling* 私とあなたの **Bayes Factor** を求めよう

- Bridgeサンプリングを用いたベイズモデルの評価 –

専修大学大学院 文学研究科 M2 **北條大樹**

時間があれば、Rで実践します!!!!!