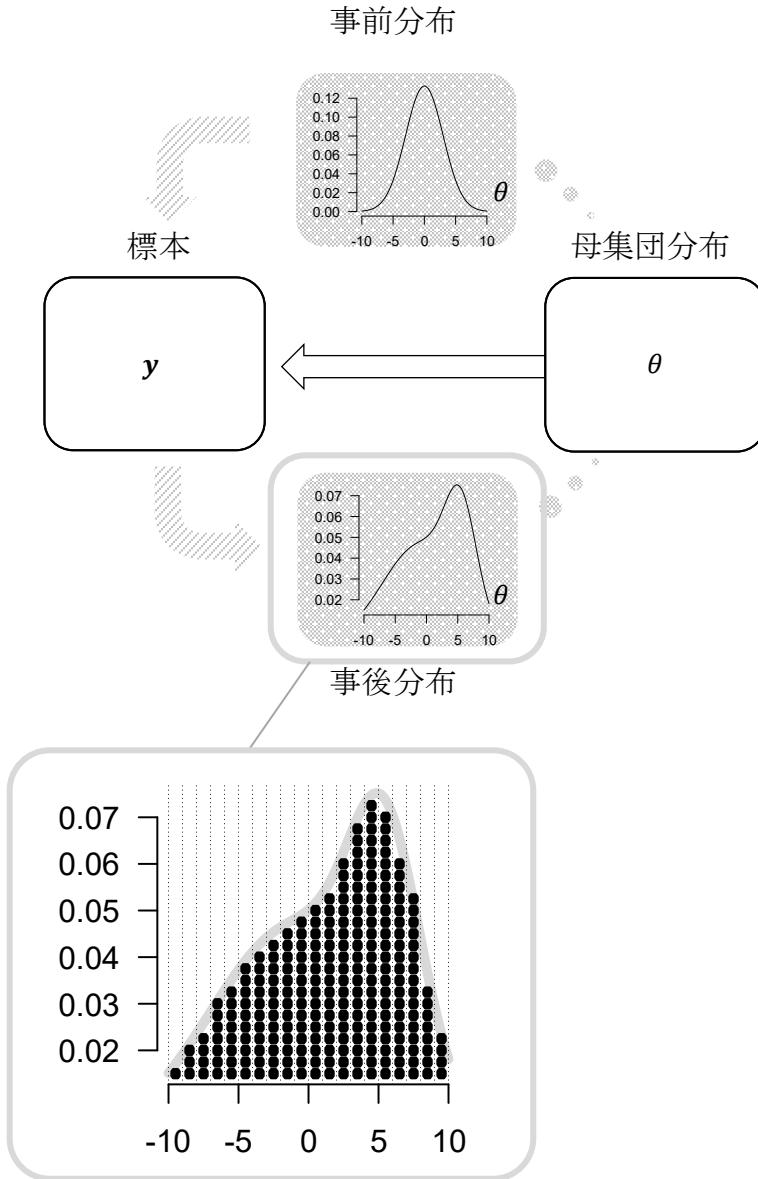


学びのポイント

- 事後分布に関わる各種統計量を算出する際に、コンピュータの力で事後分布を近似していることが分かり、この数値計算の代表的な方法に、モンテカルロ法やマルコフ連鎖モンテカルロ法（MCMC）と呼ばれる方法があることが理解できる。
- モンテカルロ法とマルコフ連鎖モンテカルロ法の違いが、サンプルを乱数で作成する際に、一つ前のステップで作成したサンプルの値を利用しているのかどうかという点にあることが説明でき、この性質をマルコフ性と呼ぶことが分かる。
- モンテカルロ法の代表的な手法に、逆変換法、棄却サンプリング、重点サンプリングと呼ばれるアルゴリズムがあることが分かり、それぞれの手法を説明できる。
- マルコフ連鎖モンテカルロ法の代表的な方法に、ギブス・サンプラー、メトロポリス法、メトロポリス・ヘイスティングス法、ハミルトニアン・モンテカルロ法と呼ばれるアルゴリズムがあることが分かり、それぞれの手法を説明できる。
- マルコフ連鎖モンテカルロ法で作成したサンプルたちが事後分布を代表した分布へと収束しているのかを確認する方法として、
- といった手法が提案されていることが分かり、実践で使用できる。

見取り図

【ベイズ統計学】



■ 目標

一つ前のベイズ統計学の基本的な考え方を説明した講義では、私たちの統計的推論は、得られた事後分布をもとに進められるということが説明されました。

ただし、前回の授業では、ざっくりと「事後期待値などの指標をもとに分布の中心を考える」という説明で終わっていて、それでは、具体的にそれをどのように求めるのかという実践的な推定のアルゴリズムについては何も語られていませんでした。

本講の目的は、この事後分布の推測の裏側でどのようなメカニズムが働いているのかという舞台裏をお見せすることにあります。それは実は「乱数を用いた試行錯誤」なのですが、なぜ、乱数を用いるのか、どういうふうに関数を用いるのかといった点で、様々な手法が提案されてきました。有名なアルゴリズムが基本的なものから発展的なものまで一つ一つ説明されるので、ベイズ統計学の事後分布の導出がどのような手順で進み、また、それがゆえに、どのような点に注意をしなければならないのかをしっかりと学び、実践で活かしてほしいと思います。



事後分布の解析に何で数値計算が必要ななの？

⇒ 期待値計算などに厄介な積分が登場するから！

(1) パラメータの値を求める方法

① 代数的方法

数式を変形して、答えを見つける方法

(例：中高で習ってきた解の出し方)

② 数値的方法

コンピュータの力で、近似的に答えを見つける方法

(例：最尤推定法で紹介されたニュートン・ラフソン法)

(方法1) モンテカルロ法

⇒ 毎回完全なランダムな乱数を使う。

(方法2) マルコフ連鎖モンテカルロ法 MCMC

⇒ 一つ前の過程で作った乱数に依存して、次の回の乱数を作る。



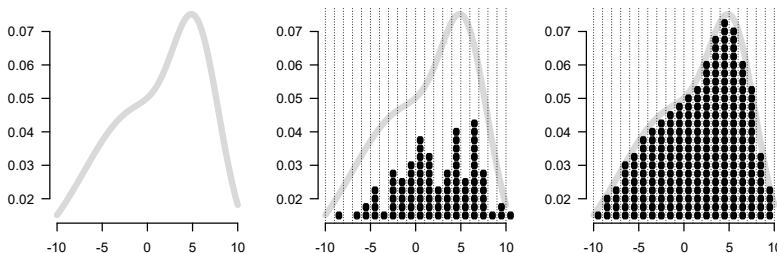
やりたいのはこんなこと：Galton Board を作る！

(ステップ1) 分割

近似させたい分布(左図)を点線で区切られているような微細な区間で分割する(中央図)。

(ステップ2) 充填

その分布の高さにそっくりな形状になるようにそれぞれの区間にビー玉を入れてやる(中央図)。

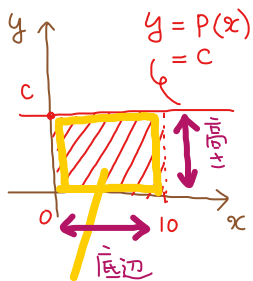


(ステップ3) 観察

ビー玉の位置で分布の性質を理解。例えば分布の中央値なら51個投入したビー玉の小さい順で26番目の位置を突き止めればよい(中央図)。

① 積分

① 長方形

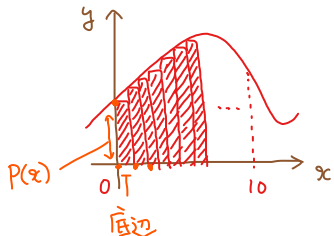


$$S = \text{高さ} \times \text{底辺}$$

$$= P(x) \times (10 - 0)$$

$$= c \times 10$$

② 一般的な場合



$S \hat{=}$ 長方形たちの合計

高さ \times 底辺

$$= \sum_{i=1}^N P(x_i) \times \text{底辺}$$

$$S = \int_0^{10} P(x) \times dx$$

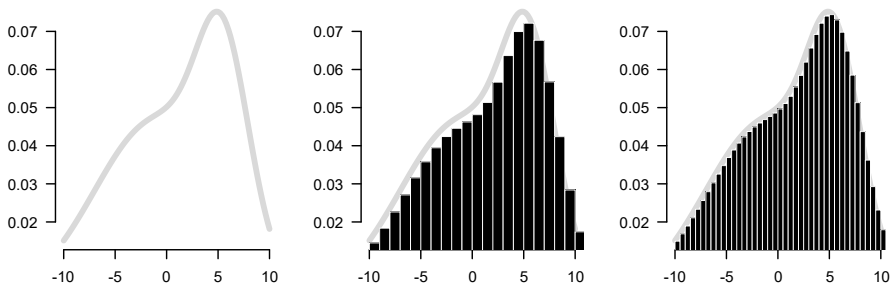
(2) 事後期待値の計算

$$\hat{\theta}_{EAP} = \int \theta \cdot p(\theta|y) d\theta$$



積分 Integration

小さな底辺 (dx) \times 高さ ($p(x)$) の長方形の面積を足し合わせる。



長方形の幅を狭めるとよくなる



期待値

重みづけ平均

$$\frac{1}{3} \times (2 + 3 + 4) = 3$$

$$\frac{1}{2} (2 + 4) = 3$$

$$\frac{1}{3} (2 + 4) = 2$$

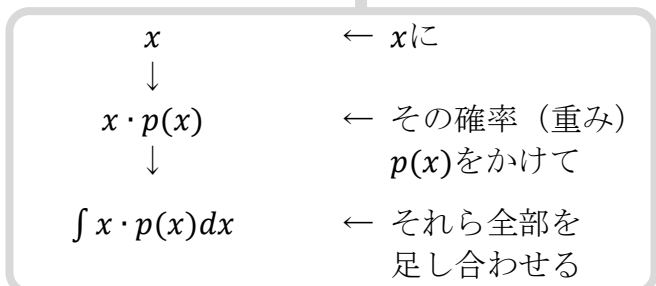
① 離散の場合

$$E(x) = \sum_{i=1}^n x_i \cdot p(x_i)$$

② 連続の場合

相対頻度 = 確率 (割合)

$$E(x) = \int_a^b x \cdot p(x) dx$$



① Σ と \int

合計 Sum $\sum_{i=1}^N P(x_i) \times \text{底辺}$

the sum of \sum

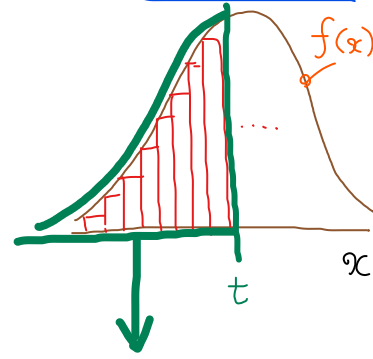
合計 $\int_a^b P(x) dx$

the sum of

■ 方法1：モンテカルロ法 Monte Carlo Methods

(1) 逆変換法 Inverse Transform Method

乱数 $u \sim Unif(0,1)$ を生成させ、累積分布関数 F の逆関数 F^{-1} で発生させた $F^{-1}(u)$ を F からの乱数と見なす方法。



左端からある点 t までのカーブ下の面積を教えてください関数 F が累積分布関数

$$F(t) = \int_{-\infty}^t f(x) dx$$

$$F(\infty) = 1$$

$$F(-\infty) = 0$$

💡 累積分布関数 Cumulative Distribution Function *cdf*

(1) 確率密度関数

$N(0,2)$

$Beta(0.5,2)$

(2) 累積分布関数

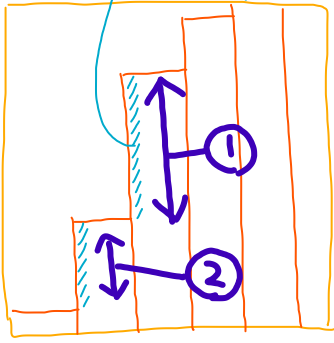
$N(0,2)$

$Beta(0.5,2)$

💡 累積分布関数の公式が知られていない場合

まず、確率密度の離散的な近似を作り、その高さを順に足して累積分布関数の近似を作る。

密度関数の高さ
に対応

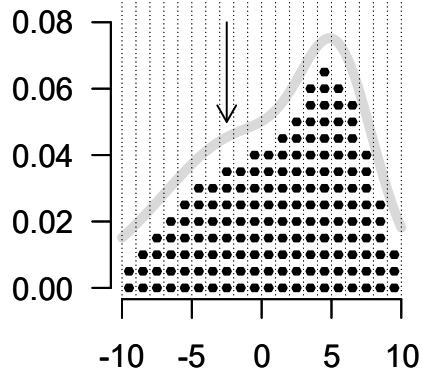
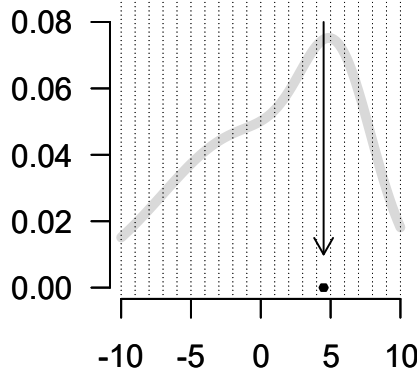
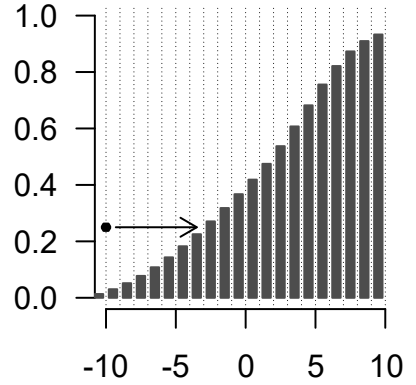
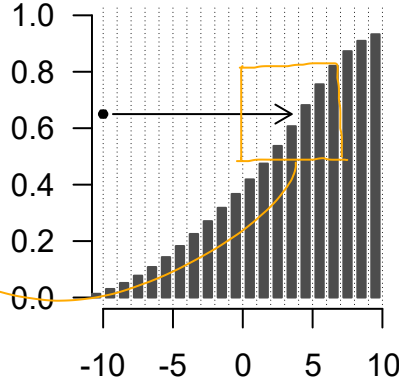


① = 入る確率
の"おこ"り
② = 入る確率
より大きい。



逆変換法の直感的な意味合い

左側の0から1の範囲でランダムに「狙撃位置」を定めて、そこからビー玉を右に向かって投げる（＝モンテカルロ法）。



ぶつかったところの棒の立つ位置にビー玉は落ちる。これを繰り返すと、左の棒との差が大きい棒の位置にビー玉はたくさん落ちてくる。ちょうどこの差分が密度なので、何度も試行を繰り返すと、目的とする密度分布に比例する形で、ビー玉はたまっていく。



逆変換法の弱点

(弱点1) 棒の間隔が緩いと精度が悪くなる

(弱点2) 高次元のパラメータの場合負荷が大きくなる
一つの次元に 1,000 個
⇒ p 次元だと、 $1,000^p$ 個必要!

■ 方法2：マルコフ連鎖モンテカルロ法 Markov-chain Monte Carlo Methods

(1) ギブス・サンプラー-Gibbs Sampler

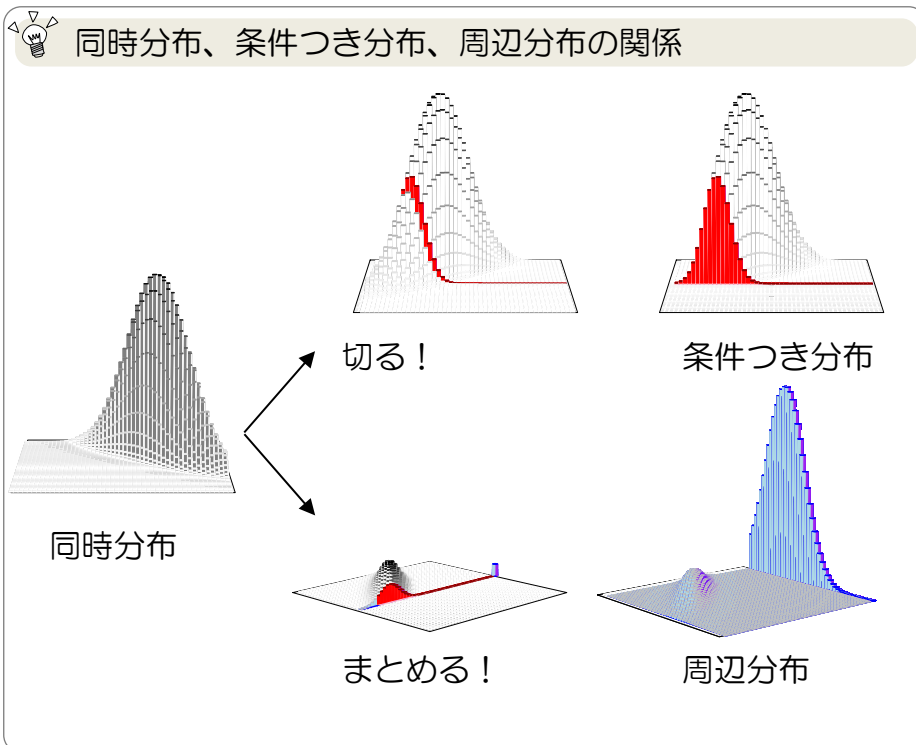
これは、同時分布を条件つき分布に分割し、各パラメータの更新をその条件つき分布から行うサンプリングである。

Step 1 初期値を決める

$$\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \theta_2^{(0)}, \dots, \theta_p^{(0)})$$

Step 2 $t = 1, \dots, T$ に対して次の工程を繰り返す

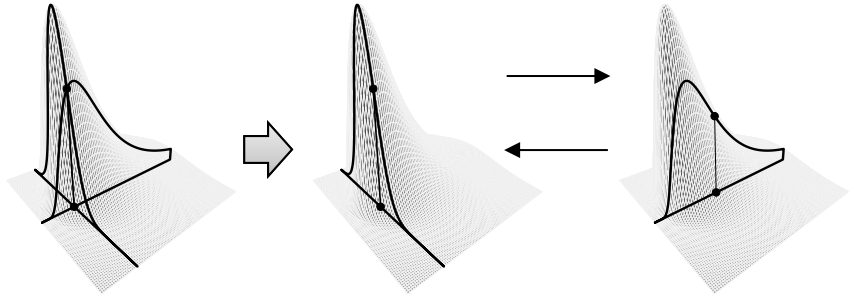
- a. $\theta_1^{(t)} \sim p(\theta_1 | \theta_2^{(t-1)}, \dots, \theta_p^{(t-1)})$
- b. $\theta_2^{(t)} \sim p(\theta_2 | \theta_1^{(t)}, \theta_3^{(t-1)}, \dots, \theta_p^{(t-1)})$
- c. $\theta_3^{(t)} \sim p(\theta_3 | \theta_1^{(t)}, \theta_2^{(t)}, \theta_4^{(t-1)}, \dots, \theta_p^{(t-1)})$
- ⋮
- d. $\theta_p^{(t)} \sim p(\theta_p | \theta_1^{(t)}, \theta_2^{(t)}, \dots, \theta_{p-1}^{(t)})$





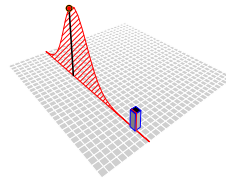
ギブス・サンプラーの直感的な理解

同時分布 (左図) を条件つき分布たちに分割 (中央図・右図)。各パラメータのサンプリングは、その条件つき分布にしたがうように実行していき、その工程を交互に繰り返す。

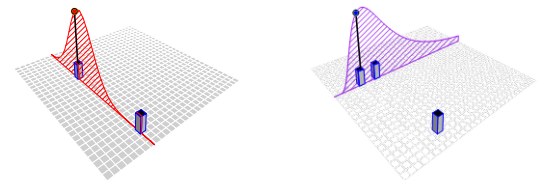


$$\theta_1^{(t)} \sim p(\theta_1 | \theta_2^{(t-1)}) \quad \theta_2^{(t)} \sim p(\theta_2 | \theta_1^{(t)})$$

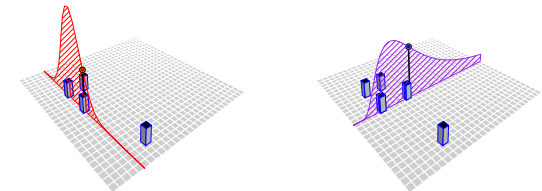
$t = 0$



$t = 1$

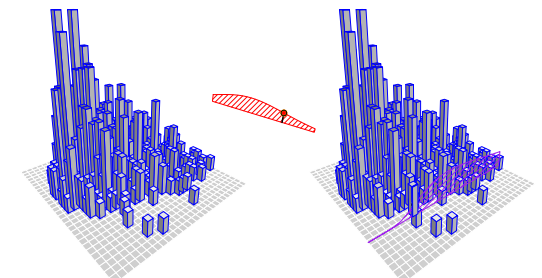


$t = 2$



⋮

$t = 500$



例題

ギブス・サンプラー

n 個のデータが観測された。これらは独立に同一の正規母集団からの標本である、つまり、次のモデルが正しいものと想定をし、分析を進めることにした。

$$y_i \sim N(\mu, \sigma^2)$$

まず、 μ と σ^2 の同時事前分布として、 $p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$ という確率分布を想定し、ギブス・サンプラーを用いて事後分布を数値的に近似してみよう。

① 尤度（密度）

i 番目のデータが得られる確率は次のようになる。

$$p(y_i | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(y_i - \mu)^2}{2\sigma^2}\right]$$

n 個のデータが同時に得られる確率は、標本が同一の正規分布から独立に採取されているので、次のようになる。

$$\begin{aligned} p(y_1, y_2, \dots, y_n | \mu, \sigma^2) &= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right] \\ &\propto \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2\right] \end{aligned}$$

② 事前分布

問題文の指示にあるように、次のように設定されている。

$$p(\mu, \sigma^2) \propto \frac{1}{\sigma^2}$$

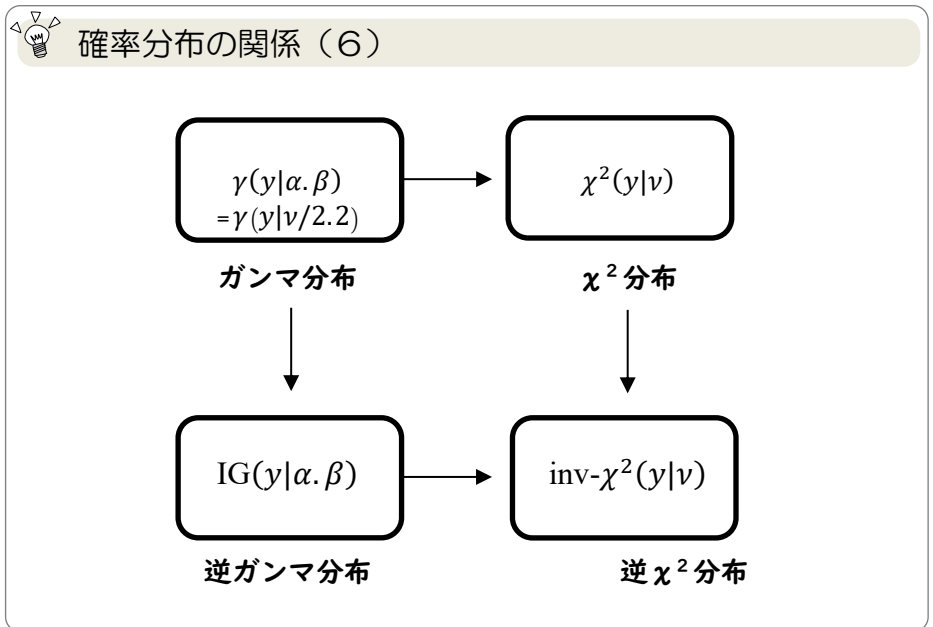
③ 事後分布

$$\begin{aligned}
 & p(\mu, \sigma^2 | y_1, y_2, \dots, y_n) \\
 & \propto p(y_1, y_2, \dots, y_n | \mu, \sigma^2) p(\mu, \sigma^2) \\
 & = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right] \times \frac{1}{\sigma^2}
 \end{aligned}$$

μ についてみると正規分布 $N(\bar{y}, \frac{\sigma^2}{n})$ のカーネル

$$\propto \{(\sigma^2)^{-1}\}^{\frac{n}{2}+1} \exp \left[-\frac{nv^2}{2} (\sigma^2)^{-1} \right] \exp \left[-\frac{1}{2(\sigma^2/n)} (\mu - \bar{y})^2 \right]$$

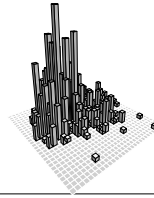
σ^2 についてみると逆ガンマ分布 $IG\left(\frac{n}{2}, \frac{nv^2}{2}\right)$ のカーネル



ノート2 MCMCのチェック事項(1): MCMCの収束



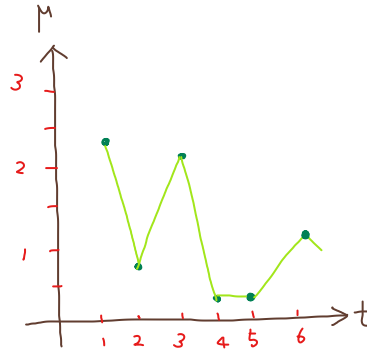
とりあえず1,000回やったけど、
得られたサンプルたちは
ちゃんと事後分布を
代表してくれているのかな？



■ 方法1: 視覚的な方法による判断

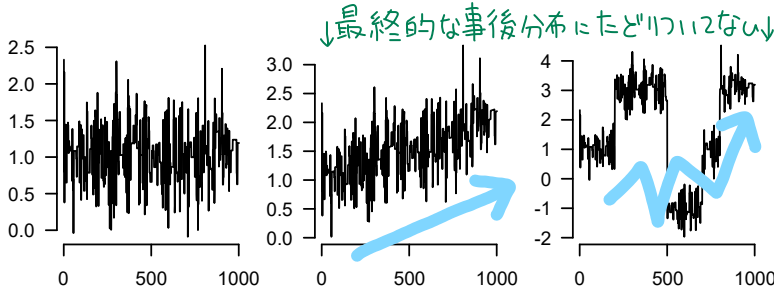
(1) 結果のファイル

t	μ	σ^2
1	2.33	0.39
2	0.80	1.72
3	2.16	1.01
4	0.37	0.56
5	0.37	0.56
⋮	⋮	⋮
T	0.41	1.59



(2) トレースプロット Trace Plot

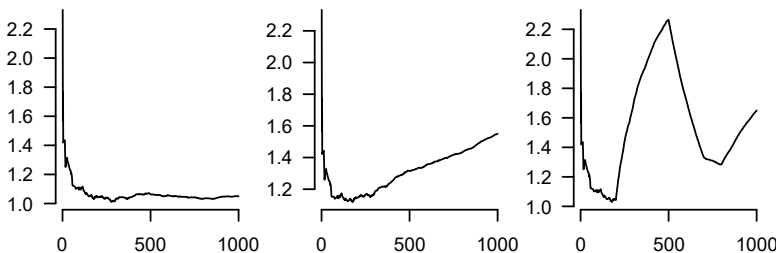
これは MCMC の結果を得られた時系列順に並べたもの。
トレンドや非ランダムな分布は未収束を示唆する。



※ただし、異常がないことは収束の証拠にはならない。

(3) 移動平均/中央値プロット Running Mean/Median Plot

これは、t番目までのサンプルの平均/中央値をプロットしたもの。トレンドやランダムさを検証する。



※ただし、異常がないことは収束の証拠にはならない。

(4) 複数回実施して結果を見比べる

一セットの MCMC サンプルの列をチェーン (chain) と呼ぶ。複数のチェーンの結果を比較して収斂を確かめる。

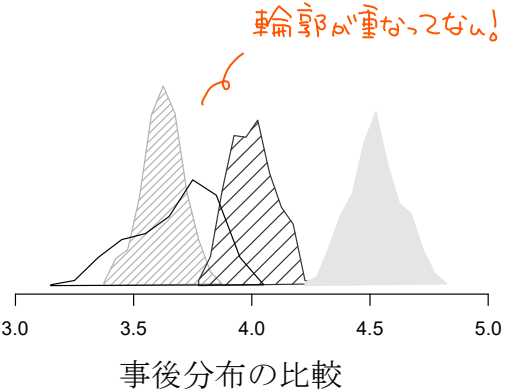
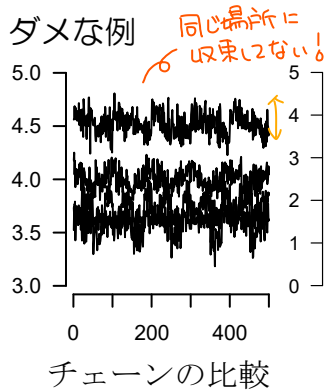
第一チェーン

t	μ	σ^2
1	2.33	0.39
2	0.80	1.72
3	2.16	1.01
4	0.37	0.56
5	0.37	0.56
\vdots	\vdots	\vdots
T	0.41	1.59

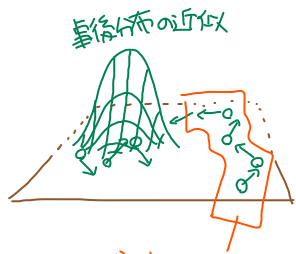
第二チェーン

t	μ	σ^2
1	0.71	0.64
2	0.71	0.64
3	0.99	1.21
4	1.42	0.78
5	1.62	0.91
\vdots	\vdots	\vdots
T	0.44	1.21


...



④ バーンイン期間



こいつらは、事後分布の近似計算から外す(切り捨てる)

 **バーンイン/ウォームアップ期間 Burn-in/warm-up Period**

これは、生成した MCMC サンプルの最初の一部で、実際の事後分布の計算には使用しない切り捨てる区間のこと。

t	μ	σ^2
1	2.33	0.39
2	0.80	1.72
3	2.16	1.01
4	0.37	0.56
5	0.37	0.56
\vdots	\vdots	\vdots
B-1	1.38	0.64

B	0.91	0.70
B+1	0.55	0.55
\vdots	\vdots	\vdots
T	0.41	1.59

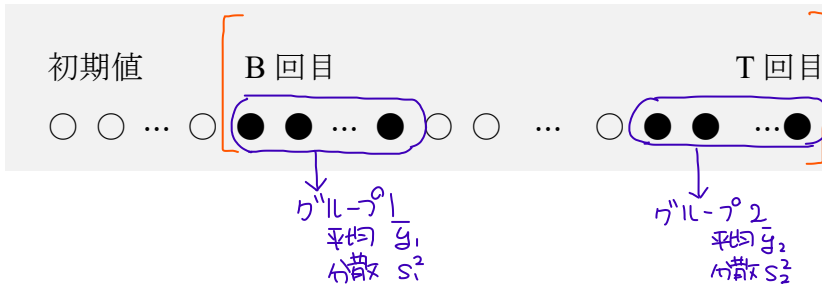
切り捨てる (B-1 and earlier)

使用 (B and later)

■ 方法2：統計量による判断

(1) Geweke の収束統計量 Geweke's convergence statistic

これは、前半と後半の列の標本平均を出し、その二群の差の平均に差がないか否かを z 検定するもの。



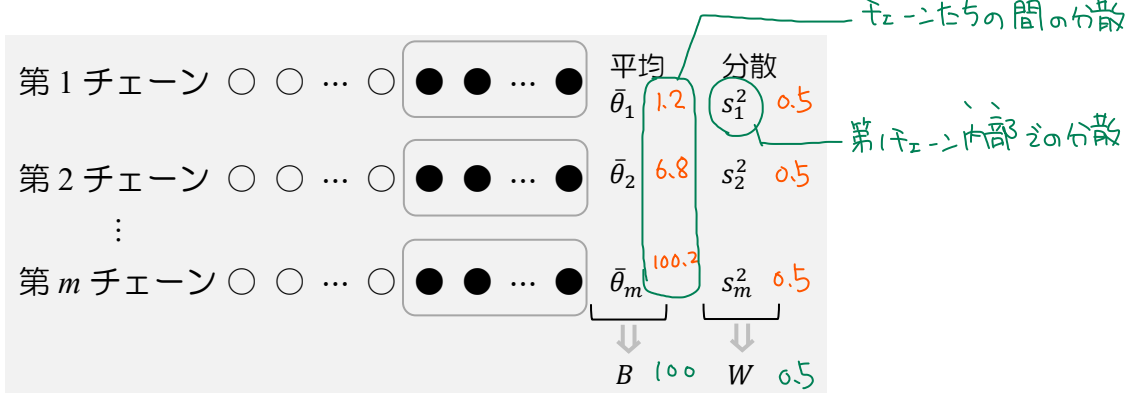
④ Geweke の発想

もし、「オレンジの区間のサンプルたちが、事後分布の近似(つまり事後分布からのランダムなサンプルだ)と見なせる」のなら、「グループ1とグループ2の平均には差がない」と予測。

→ z 検定
→ (z 検定)

(2) Gelman-Rubin 統計量 \hat{R} Gelman-Rubin Statistic

これは、「同一チェーン内での乱数の分散の平均値」と「異なるチェーンも含めたすべての乱数の分散」の比の値。



① 郡内 (チェーン内) 分散 Within-chain Variance

j 番目のチェーンの内部の不偏分散

$$s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (\theta_{ij} - \bar{\theta}_j)^2$$

m 個のチェーンの内部の不偏分散の平均値

$$W = \frac{1}{m} \sum_{j=1}^m s_j^2$$

② 郡間 (チェーン間) 分散 Between-chain Variance

j 番目のチェーンの平均

$$\bar{\theta}_j$$

j 番目のチェーンの平均の平均

$$\bar{\theta} = \frac{1}{m} \sum_{j=1}^m \bar{\theta}_j$$

m 個のチェーンの平均の不偏分散

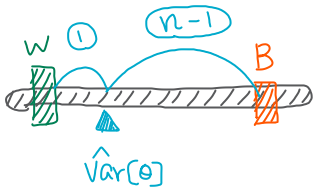
$$B = \frac{n}{m-1} \sum_{j=1}^m (\bar{\theta}_j - \bar{\theta})^2$$

① \hat{R} に対する閾値

昔は, $\hat{R} < 1.1$ や $\hat{R} < 1.05$ くらい
収束していると見なしていた時期も。

ただ, 近年では, もう少し高い
基準にした方が「良い」のでは,
という議論が起りつつある。

① B と W の関係



③ 異なるチェーンも含めたすべての乱数の分散

$$\widehat{Var}[\theta] = \frac{n-1}{n}W + \frac{1}{n}B$$

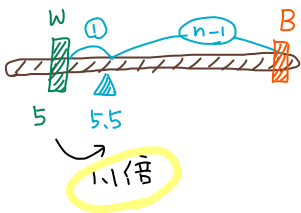
④ \hat{R} の計算

$$\hat{R} = \sqrt{\frac{\widehat{Var}[\theta]}{W}}$$

← チェーンを無視したときの分散 (分子)
← チェーン内分散の平均 (分母)
(分子) の何倍かを示す

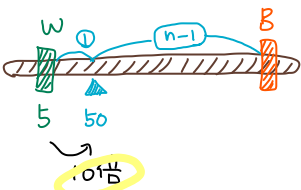
(1) 理想的な場合

$\widehat{Var}[\theta]$ (全2のの分散) と
 W (チェーン内の分散の平均) が
ほぼほぼ同じになるはず



(2) 非理想な場合

$\widehat{Var}[\theta]$ と W がだいぶ
ちがう



⇒ B が大きくなりすぎると,
支点がある $\widehat{Var}[\theta]$ も右にずれる。
 W の何倍も大きい値になる。

R における実装

```

> fit <- stan(file = "Stan01.stan", data = data1, iter = 25)
1: The largest R-hat is 1.37, indicating chains have not mixed.
    (あ! 収束していないぞ!)

> print(results3a)
Inference for Stan model: Class10Stan01.
4 chains, each with iter = 25; warmup = 12; thin = 1;
post-warmup draws per chain = 13, total post-warmup draws = 52.
    
```

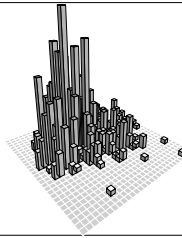
	mean	...	sd	2.5%	...	97.5%	n_{eff}	Rhat
beta0	3.29	...	0.32	2.68	...	3.66	19	1.17
beta1	0.42	...	0.32	-0.06	...	1.20	12	1.34
sigma	1.34	...	0.21	1.19	...	1.61	63	0.99

\hat{R} が 収束の値を示していないので, あてにならない。

📖 ノート2 MCMCのチェック事項(2): MCMCの効率性



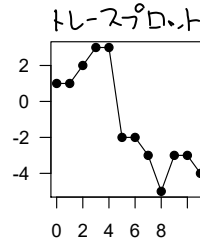
とりあえずサンプリングして
右のような感じになったけど
どのくらい効率的なサンプル
リングだったんだろうか？



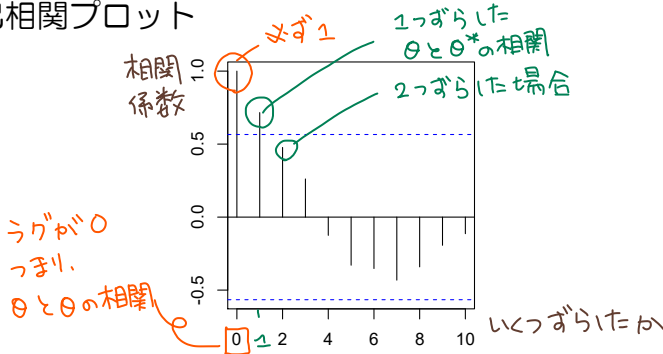
(1) 自己相関プロット Autocorrelation Function (ACF) Plot
これは、 $1 \leq k \leq n-1$ に対して $\theta = (\theta^{(0)}, \theta^{(1)}, \dots, \theta^{(n-k)})$ と
 $\theta = (\theta^{(k)}, \theta^{(k+1)}, \dots, \theta^{(n)})$ の相関係数を図示したもの。

① 時系列データ

t	0	1	2	3	4	5	6	7	8	9	10	11
θ	1	1	2	3	3	-2	-2	-3	-5	-3	-3	-4
θ^*		1	1	2	3	3	-2	-2	-3	-5	-3	

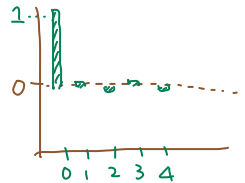


② 自己相関プロット



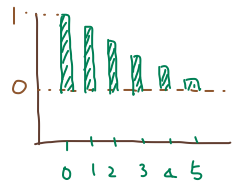
③ 自己相関プロット

(T-2,1) いきなり降下



時点 t と 時点 $t-1$
がずれては相関なし

(T-2,2) や、 θ と θ の降下

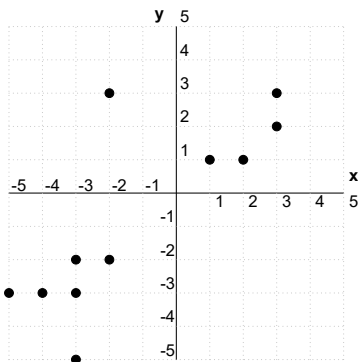


当然 MCMC 2"
採らしたサンプルたちは、
多かれ少なかれ
(T-2,2) のようになる。



復習：相関係数

$$\text{cor}(X, Y) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$



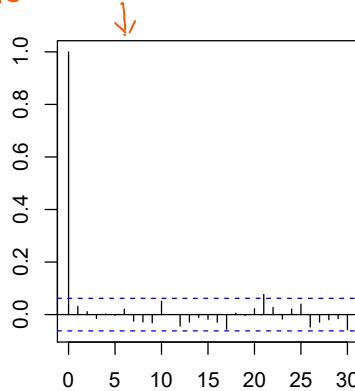
ID	x	y
0		1
1	1	1
2	2	1
3	3	2
4	3	3
5	-2	3
6	-2	-2
7	-3	-2
8	-5	-3
9	-3	-5
10	-3	-3
11	-4	-3
Mean	-1.0	-0.923

x と y の
1つずらしたものは
 y と x 。

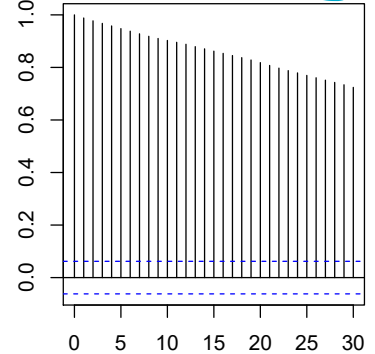
(2) **有効サンプルサイズ** Effective Sample Size

これは、 mn 個取得した MCMC サンプルが、互いに独立したサンプルであるなら何個分の情報に相当するかを表す。

理想的な状態



実際の場合



$$\hat{n}_{\text{eff}} = \frac{mn}{1 + 2 \sum_{k=1}^{\infty} ACF(k)}$$

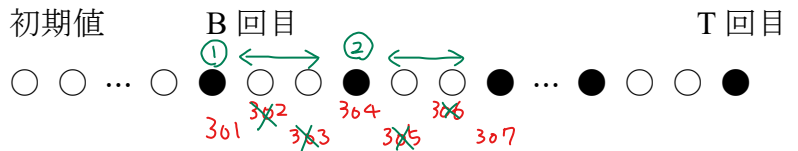
MCMCで実際に収集したサンプルの数 (m個のフェーズに各n個あるのzinn)

自己相関プロットがゆるやかだと(大)

分母については、実際の計算では、 $ACF(k) < 0.05$ などの閾値を超えた場合に中断される。

(3) **まびき** Thinning

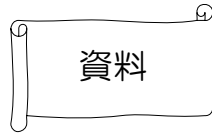
これは、効率性をいじらずに、自己相関を減少させる方法。パソコンのメモリの消費を抑えることにも有効。



収束診断 (Stan 公式 YouTube ビデオ)

Stan 開発チームは、積極的にチュートリアルを YouTube などに公開している。次のビデオは、収束に関する分かりやすい動画で、ぜひ参考にしてほしい。

https://www.youtube.com/watch?v=0FdMZwIbJ_4



資料6-1 事後分布の条件付事後分布への分解

「例題」で扱った棄却サンプリングの定数 M の求め方の詳細は以下の通りである。

$$p(\theta|y) \leq M \times q(\theta)$$

$$\frac{p(\theta|y)}{q(\theta)} \leq M$$

ここで、分母、分子の確率密度を代入する。

$$\begin{aligned} \frac{p(\theta|y)}{q(\theta)} &= \frac{\frac{2}{\sqrt{2\pi}} \exp\left[-\frac{\theta^2}{2}\right]}{\exp[-\theta]} \\ &= \frac{2}{\sqrt{2\pi}} \exp\left[-\frac{\theta^2}{2} + \theta\right] \\ &= \frac{2}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\theta^2 - 2\theta)\right] \\ &= \frac{2}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\theta^2 - 2\theta - 1 + 1)\right] \\ &= \frac{2}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}(\theta - 1)^2 + \frac{1}{2}\right] \\ &= \left(\frac{2}{\sqrt{2\pi}} \exp\frac{1}{2}\right) \times \exp\left[-\frac{1}{2}(\theta - 1)^2\right] \end{aligned}$$

ここで、常に、 $(\theta - 1)^2 \geq 0$ なので、 $-\frac{1}{2}(\theta - 1)^2 \leq 0$ であり、したがって、 $\exp\left[-\frac{1}{2}(\theta - 1)^2\right] \leq 1$ である。よって、

$$\left(\frac{2}{\sqrt{2\pi}} \exp\frac{1}{2}\right) \times \exp\left[-\frac{1}{2}(\theta - 1)^2\right] \leq \left(\frac{2}{\sqrt{2\pi}} \exp\frac{1}{2}\right)$$

$$\frac{p(\theta|y)}{q(\theta)} \leq \left(\frac{2}{\sqrt{2\pi}} \exp\frac{1}{2}\right)$$

資料6-2 事後分布の条件付事後分布への分解

「例題」で扱ったギブス・サンプラーを用いるための事後分布の分解の計算の詳細は以下の通りである。

$$\begin{aligned}
 p(\mu, \sigma^2 | y_1, y_2, \dots, y_n) & \\
 & \propto p(y_1, y_2, \dots, y_n | \mu, \sigma^2) p(\mu, \sigma^2) \\
 & = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right)^n \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right] \times \frac{1}{\sigma^2} \\
 & = \left(\frac{1}{\sqrt{2\pi}} \right)^n \times \left(\frac{1}{\sqrt{\sigma^2}} \right)^n \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right] \times \frac{1}{\sigma^2} \\
 & \propto \left(\frac{1}{\sqrt{\sigma^2}} \right)^n \frac{1}{\sigma^2} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right] \\
 & = \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}} \left(\frac{1}{\sigma^2} \right)^1 \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right] \\
 & = \left(\frac{1}{\sigma^2} \right)^{\frac{n}{2}+1} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right] \\
 & = (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \mu)^2 \right]
 \end{aligned}$$

実はこの事後分布は、二つの分布の積として分解できる。

$$\begin{aligned}
 & \propto (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \bar{y} + \bar{y} - \mu)^2 \right] \\
 & \propto (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \{(y_i - \bar{y}) - (\mu - \bar{y})\}^2 \right]
 \end{aligned}$$

$$\begin{aligned}
& \propto (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \sum_{i=1}^n \{(y_i - \bar{y})^2 - 2(y_i - \bar{y})(\mu - \bar{y}) + (\mu - \bar{y})^2\} \right] \\
& = (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^n (y_i - \bar{y})^2 - \sum_{i=1}^n 2(y_i - \bar{y})(\mu - \bar{y}) + \sum_{i=1}^n (\mu - \bar{y})^2 \right\} \right] \\
& = (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^n (y_i - \bar{y})^2 - 2(\mu - \bar{y}) \sum_{i=1}^n (y_i - \bar{y}) + n(\mu - \bar{y})^2 \right\} \right] \\
& = (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^n (y_i - \bar{y})^2 - 2(\mu - \bar{y}) \left(\sum_{i=1}^n y_i - \sum_{i=1}^n \bar{y} \right) + n(\mu - \bar{y})^2 \right\} \right] \\
& = (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^n (y_i - \bar{y})^2 - 2(\mu - \bar{y})(n\bar{y} - n\bar{y}) + n(\mu - \bar{y})^2 \right\} \right] \\
& = (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \left\{ \sum_{i=1}^n (y_i - \bar{y})^2 + n(\mu - \bar{y})^2 \right\} \right]
\end{aligned}$$

y_i の標本分散を $v^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})^2$ とおくと、 $\sum_{i=1}^n (y_i - \bar{y})^2$ は nv^2 と書き表せるので、式を見やすくするため赤い部分を置き換える。

$$\begin{aligned}
& = (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{1}{2\sigma^2} \{nv^2 + n(\mu - \bar{y})^2\} \right] \\
& = (\sigma^2)^{-\frac{n}{2}-1} \exp \left[-\frac{nv^2}{2\sigma^2} \right] \exp \left[-\frac{n}{2\sigma^2} (\mu - \bar{y})^2 \right] \\
& = \{(\sigma^2)^{-1}\}^{\frac{n}{2}+1} \exp \left[-\frac{nv^2}{2} (\sigma^2)^{-1} \right] \exp \left[-\frac{n/n}{2\sigma^2/n} (\mu - \bar{y})^2 \right]
\end{aligned}$$

μ についてみると正規分布 $N\left(\bar{y}, \frac{\sigma^2}{n}\right)$ のカーネル

$$= \underbrace{\{(\sigma^2)^{-1}\}^{\frac{n}{2}+1} \exp \left[-\frac{nv^2}{2} (\sigma^2)^{-1} \right]}_{\sigma^2 \text{ についてみると逆ガンマ分布 } IG\left(\frac{n}{2}, \frac{nv^2}{2}\right) \text{ のカーネル}} \overbrace{\exp \left[-\frac{1}{2(\sigma^2/n)} (\mu - \bar{y})^2 \right]}$$

σ^2 についてみると逆ガンマ分布 $IG\left(\frac{n}{2}, \frac{nv^2}{2}\right)$ のカーネル