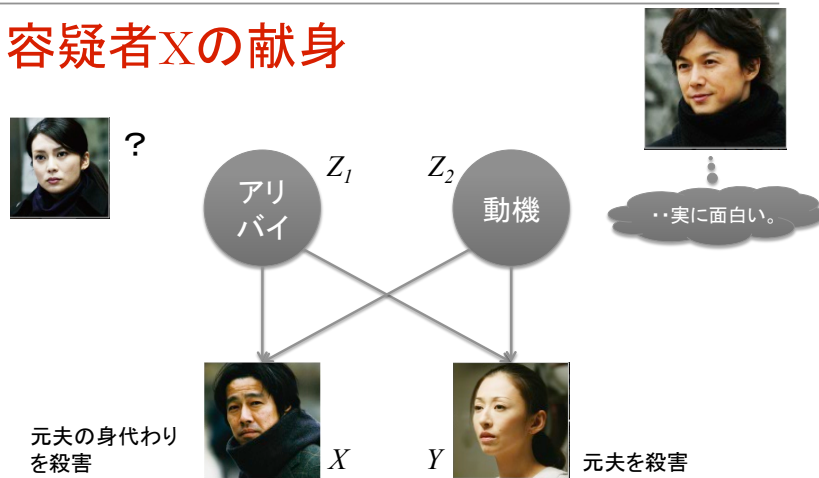


ベイズ統計

慶應義塾大学 古谷知之

容疑者Xの献身



- データ(アリバイと動機)が与えられたとき、容疑者Xが犯人である確証は？

写真の出演) <http://yougisha-x.com/>

講義概要

- 1限: ベイズ統計の基本
- 2限: ベイズ推論
- 3限: MCMCによるベイズ統計モデリング

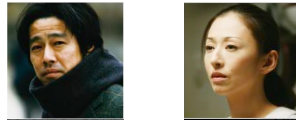
方法1) 帰無仮説の有意性検定

- 仮説を立てる
 - 帰無仮説 H_0 : 容疑者Xは犯人ではない
 - 対立仮説 H_1 : 容疑者Xは犯人である
- データを集める
 - アリバイ、動機、証拠、証言...
- 帰無仮説が正しいければ、犯人がXでない結果となる確率はどの程度か？ → p値
- p値が有意に小さければ、帰無仮説は棄却 → 容疑者Xが犯人である



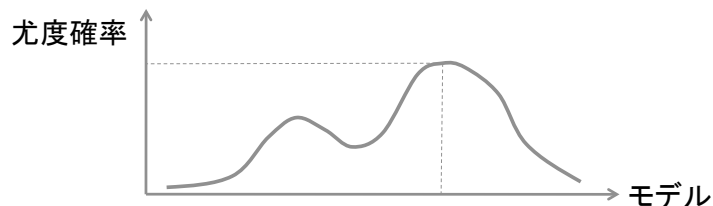
方法2) 情報理論的アプローチ

- 仮説を立てる
 - 仮説 H_X : 容疑者Xが犯人である
 - 仮説 H_Y : 容疑者Yが犯人である
- データを集める
 - アリバイ、動機、証拠、証言...
- 仮説1と2のもとで、データを観測する確率 $P(D|H_X)$ 、 $P(D|H_Y)$ はいくらか(=尤度)を計算する
- 各仮説の尤度比を検討することで、どの仮説が最も尤もらしいかを判断する



尤度

- 条件付き確率を尤度 (likelihood) という
- モデルが与えられた状態でデータが観測される **尤もらしさの度合い**を意味する。
- 頻度主義統計学で用いられる **最尤法**は、尤度を最大にするモデルを求める



頻度主義統計学

- データを観測するための「尤度」を推定する
- 帰無仮説の有意性検定
 - 選ばれた帰無仮説が真かどうかを有意性検定により判定し、データの成立の有無を判断
- 情報理論的アプローチ
 - 複数の競合する仮説に対するデータの整合性を評価

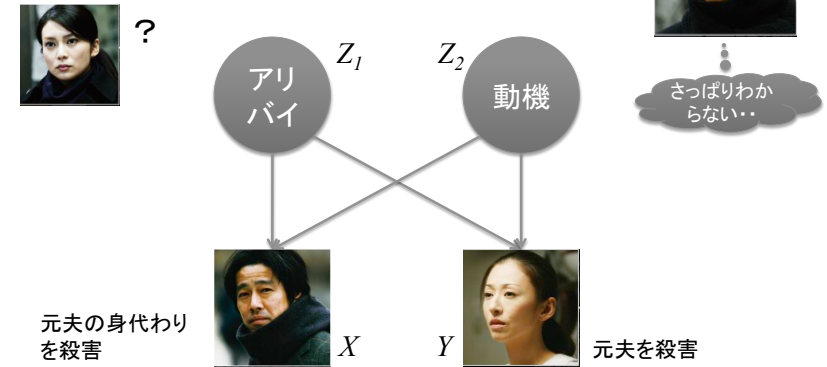
仮説検定に対する批判

- 与えられた(一つの)仮説を前提にしてデータが得られる確率を表そうとする
- 「 p 値 <0.05 だと統計的に有意」という根拠は人為的(恣意的)である
 - 帰無仮説はなお5%の確率で尤もらしい
 - 帰無仮説が真であれば p 値は十分に大きいですが、 p 値が大きい場合、帰無仮説が真であるという強い証拠にはならない

方法3) ベイズ的アプローチ

- 事前情報を確率として表現
 - $P(H_X)$: 容疑者Xが犯人である確率
 - $P(H_Y)$: 容疑者Yが犯人である確率
- それぞれの事前情報に対して、仮説が真である条件付きでデータが得られる確率＝尤度を計算
 - $P(D|H_X)$: 容疑者Xが犯人の時にアリバイ、動機がある確率
- 事前確率 $P(H_X)$ ×尤度 $P(D|H_X)$
＝事後確率 $P(H_X|D)$

容疑者Xの献身



- 容疑者Xが犯人の時、動機がある可能性は？

写真の出展) <http://yougisha-x.com/>

ベイズの定理による条件付き確率

- 容疑者Xが犯人のとき、動機が「ある」可能性は？
- 条件付き確率表

	あり=1	なし=0
$P(Z_1)$	0.4	0.6
$P(Z_2)$	0.9	0.1

Z_1	Z_2	過去の経験 $P(X Z_1, Z_2)$	
		犯人である=1	犯人でない=0
あり	あり	0.4	0.6
あり	なし	0.5	0.5
なし	あり	0.9	0.1
なし	なし	0.7	0.3

ベイズの定理の導き方

- より一般的に、二つの事象AとBがあるとする
- 同時確率＝条件付き確率×周辺確率
 $P(A \cap B) = P(A|B)P(B)$
 が成立する。
- 同様に、 $P(B \cap A) = P(B|A)P(A)$ が成立する
- $P(A \cap B) = P(B \cap A)$
 $\Leftrightarrow P(A|B)P(B) = P(B|A)P(A)$
 $\Leftrightarrow P(A|B) = P(B|A)P(A) / P(B)$
- これが「ベイズの定理」

ベイズの定理

- 事象 A と事象 B が起きる確率 $P(A)$ 、 $P(B)$
- 事象 A の排反事象 A^c

$$\begin{aligned}P(A|B) &= \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} \\ &= \frac{P(B|A)P(A)}{P(B)}\end{aligned}$$

ベイズの定理による条件付き確率

- 容疑者 X が犯人のとき、動機が「ある」可能性は？

$$P(Z_2=1|X=1) = \frac{P(X=1|Z_2=1)P(Z_2=1)}{P(X=1)}$$

$$\begin{aligned}P(X=1|Z_2=1) &= P(X=1|Z_2=1, Z_1=0)P(Z_1=0) \\ &\quad + P(X=1|Z_2=1, Z_1=1)P(Z_1=1) \\ &= 0.5 \times 0.6 + 0.4 \times 0.4 = 0.46\end{aligned}$$

$$\begin{aligned}P(X=1) &= \sum_{k=0}^1 \sum_{l=0}^1 P(X=1|Z_1=k, Z_2=l)P(Z_1=k)P(Z_2=l) \\ &= 0.7 \times 0.6 \times 0.1 + 0.9 \times 0.6 \times 0.9 \\ &\quad + 0.5 \times 0.4 \times 0.1 + 0.4 \times 0.4 \times 0.9 \\ &= 0.692\end{aligned}$$

$$P(Z_2=1|X=1) = \frac{0.41 \times 0.9}{0.692} = 0.598$$

ベイズの定理による条件付き確率

- 動機が「ある」場合に容疑者 X が犯人の可能性は？

$$\begin{aligned}P(X=1|Z_2=1) &= P(X=1|Z_2=1, Z_1=0)P(Z_1=0) \\ &\quad + P(X=1|Z_2=1, Z_1=1)P(Z_1=1) \\ &= 0.5 \times 0.6 + 0.4 \times 0.4 = 0.46\end{aligned}$$

- アリバイが「ない」場合に容疑者 X が犯人の可能性は？

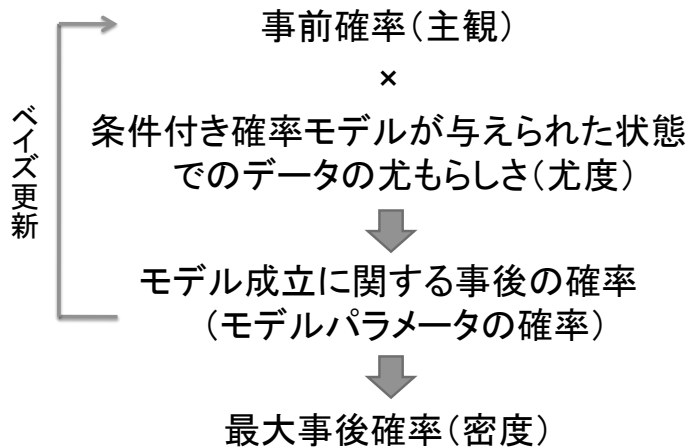
$$\begin{aligned}P(X=1|Z_1=0) &= P(X=1|Z_1=0, Z_2=0)P(Z_2=0) \\ &\quad + P(X=1|Z_1=0, Z_2=1)P(Z_2=1) \\ &= 0.7 \times 0.1 + 0.1 \times 0.9 = 0.16\end{aligned}$$

ベイズ的な考え方

- 新型インフルエンザの新薬: 1/2の確率で治癒
- この新薬が特効薬(=全員が治る)である可能性は？
- 古典統計的発想: p値 $\left(\frac{1}{2}\right)^5 = \frac{1}{32} = 0.03125$
- ベイズ的発想: 特効薬なら全員が治癒する確率 = 1 (事前確率1/2)

$$\frac{\frac{1}{2} \times 1}{\frac{1}{2} \times 1 + \frac{1}{2} \times \left(\frac{1}{2}\right)^5} \approx 96.97$$

ベイズ的アプローチ



古典統計とベイズ統計の主な相違点・共通点

古典統計	ベイズ統計
パラメータの真の値はわかる(存在する)	パラメータの真の値はわからない
パラメータは固定値をとる	パラメータは確率分布する
パラメータの統計的な確からしさを仮説検証する	平均値や中央値などでパラメータ値を代表させる
誤差ができるだけ小さくなるようにパラメータを推定	パラメータの確率分布をもとにパラメータを推定
母集団から抽出したサンプルデータをもとにパラメータを推計	

Thomas Bayes (1701?-1761)

- 英国非国教会派の牧師
- 死後に、初期の条件付き確率に関する論文が発見され、遺作として発表される
 - 「事後」で条件づけられた「事前」の確率を計算できること＝「ベイズの定理」を発見



<http://plus.maths.org/>

ラプラスによるベイズ確率の応用



<http://upload.wikimedia.org/>

- ナポレオン時代のフランス人数学者
 - 微分方程式、統計学、天体力学などで貢献
 - ナポレオン時代に内務大臣を務めたが更迭され、叙勲されるもナポレオン流刑の命令にサインする
- ベイズの定理の「予測確率」
- 土星の質量の予測にベイズの定理を応用
 - 質量の平均を用いたのが失敗

Year	\bar{X}_n	S_n^2	X_{n+1}	X_n	Laplace method	Bayes method	L.S. method
1800	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1801	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1802	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1803	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1804	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1805	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1806	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1807	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1808	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1809	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1810	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1811	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1812	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1813	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1814	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1815	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1816	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1817	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1818	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1819	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1820	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1821	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1822	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1823	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1824	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1825	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1826	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1827	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1828	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1829	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1830	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1831	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1832	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1833	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1834	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1835	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1836	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1837	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1838	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1839	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1840	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1841	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1842	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1843	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1844	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1845	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1846	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1847	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1848	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1849	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
1850	21.07	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000

Stigler (1975)

カール・ピアソン(1857-1936)



<http://upload.wikimedia.org>

- 政治学で学位。マルクス主義に傾倒した上流階級。
 - ユニバーシティ・カレッジ・ロンドン(UCL)教授
 - ピアソン=ネイマンの補題のピアソンは息子(エゴン)
- 相関係数、線形回帰、 χ^2 検定、確率分布の類型化などの業績で統計学の発展に寄与
 - 回帰・相関は、フランシス・ゴルトンとの共同研究成果
- ピアソンによる確率分布の類型化
 - B分布(タイプ1)、指数分布・ χ^2 分布・ Γ 分布(タイプ3)、スチューデントのt分布(タイプ4)、逆 Γ 分布(タイプ5)、F分布(タイプ6)

ピアソン(子)とネイマン

- エゴン・ピアソンはカール・ピアソンの息子
- イェッジ(イエルジー)ネイマンと共同研究者
 - 信頼区間・仮説検定などの方法を研究
- 推計統計学、頻度論を確立
- ベイズの定理では、母数に対して使われるときの確率が何を意味するのかが課題であった
- ピアソンは測定値の確率分布を定める母数の値を推定することに、科学調査の目的を移行させた
 - ベイズの定理への関心を低める結果に

R.A.フィッシャーによる推測統計学

- フィッシャー:小標本理論、実験計画法、分散分析、最尤法などの手法を開発
 - 父ピアソンに冷遇され、子ピアソンを攻撃
- p値を重要視し、小標本による推測統計学を推進
 - 「研究者のための統計的方法」を執筆
 - 相対的にベイズ統計学が急速に衰える

最近のベイズ統計学を巡る状況

- 計量経済学、マーケティング科学、計量政治学などの社会科学分野でメジャーになりつつある
 - 統計解析を行う欧米の社会科学者の半分くらいはベイズ統計を使うといわれている
- 人工知能やバイオ統計、計量生物学、認知心理学などでは普通に使われている
 - ベイズ統計を使っていないと論文として認めない分野もある
- 英語の教科書は増加傾向。日本語の教科書はまだ少ない

ベイズの定理と統計的推測

- 観測データ: D
- (統計モデル M が成立する)仮説: H

- ベイズの定理を用いると
データが与えられた条件で仮説 H (つまりモデル M)が成立する確率は、

$$P(H|D)=P(D|H)P(H)/P(D)$$

ベイズの定理と統計的推測

- モデル M が成立するという仮説を定める前に、
データ D が得られる周辺確率 $P(D)$

 - 同時確率(事前確率)、周辺確率ともに、明確に
定めることが非常に困難
- ↓
- 複数のモデルが成立すると仮定して、各モデル
が成立する仮説同士の相対的な確率を考える

ベイズの定理と統計的推測

- モデル M が成立するであろう仮説 H の確率 $P(H)$

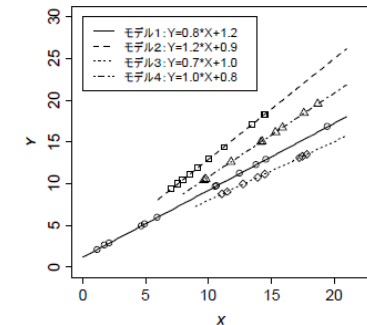
- データがわからないときの仮説の確率

- 一種の事前確率と見なすことができる

- 本当にこのような確率を定められるのか？

ベイズの定理と統計的推測

- 四つのモデルを考える
 - モデル1
 - モデル2
 - モデル3
 - モデル4



- 全くの主観で各モデルが成立する確率が等しい
と考える

$$P(\text{モデル1}) = P(\text{モデル2}) = P(\text{モデル3}) = P(\text{モデル4}) = 0.25$$

ベイズの定理と統計的推測

- 各モデルの条件付き確率
(仮説の下でデータが得られる尤もらしさ=尤度)

$$P(\text{観測データ}|\text{モデル 1}) = 12/40 = 0.30$$

$$P(\text{観測データ}|\text{モデル 2}) = 10/40 = 0.25$$

$$P(\text{観測データ}|\text{モデル 3}) = 8/40 = 0.20$$

$$P(\text{観測データ}|\text{モデル 4}) = 10/40 = 0.25$$

- 観測データが得られる確率

$$\begin{aligned} P(\text{観測データ}) &= P(\text{観測データ}|\text{モデル 1})P(\text{モデル 1}) \\ &\quad + P(\text{観測データ}|\text{モデル 2})P(\text{モデル 2}) \\ &\quad + P(\text{観測データ}|\text{モデル 3})P(\text{モデル 3}) \\ &\quad + P(\text{観測データ}|\text{モデル 4})P(\text{モデル 4}) \\ &= 0.30 \cdot 0.25 + 0.25 \cdot 0.25 + 0.20 \cdot 0.25 + 0.25 \cdot 0.25 = 0.25 \end{aligned}$$

事前確率の妥当性

- さきほど、全くの主観として事前確率(同時確率)を与えた
- 主観的に事前確率を与える方法は妥当か？
- 頻度主義統計学者がベイズ統計学に異論を唱える理由

ベイズの定理と統計的推測

- **ベイズの定理**より、データが与えられた場合のモデルが成立する確率

$$P(\text{モデル 1}|\text{観測データ}) = \frac{P(\text{観測データ}|\text{モデル 1}) \cdot P(\text{モデル 1})}{P(\text{観測データ})} = \frac{0.30 \cdot 0.25}{0.25} = 0.30$$

- 事前確率と条件付き確率(尤度)からモデルが成立する確率を求めたことに他ならない
- モデル2~4の確率も計算してみよう

継起の規則(ラプラス)

- 総・環 n 人の学生に、総合政策学部所属か環境情報学部所属かを尋ねる
- 任意に総合政策学部所属の学生にあう確率は $\frac{1}{n+1}$ となる
- 数学的帰納法により証明可能

継起の規則(ラプラス)

- $n=1$ のとき、明らか
- $n>1$ のとき、 $n+1$ 人中 m 人($m=0, \dots, n+1$)が総合である確率は、最初の n 人で m 人の総合の学生に会い、 $n+1$ 人目は環境の学生にあうことだから、 $[(n-m)+1]/[(n+1)(n+2)]$ となる
- それまで $m-1$ 人の総合の学生に会い、次にあう学生が総合の学生である確率は、 $[(m-1)+1]/[(n+1)(n+2)]$ となる
- この二つを足すと、 $1/(n+2)$ となる ■

継起の規則とベイズの定理

- ところで、総合の学生あるいは環境の学生に会う確率は、最初は五分五分(確率 $1/2$)と考えるのが一般的(経験・経験による事前確率)
 - 主観的に事前確率は均一と考える
- 事前分布は均一? どう設定すればいい? という疑問を継起の規則とベイズの定理を使って検証
 - ずっと総合の学生ばかりに会う状況を想定

理由不十分の原則

- ある事前確率を与えたとき、事前確率がそれ以外考えられない場合、当面それを事前確率と認めてよいこと
- 例:
 - 総合の学生に会う確率 $1/(n+1)$
 - 生まれてくる子供が男の子である確率 $1/2$

継起の規則とベイズの定理

- 最初にあう n 人が総合政策の学生である: 事象 A
- 次にあう m 人が総合政策の学生である: 事象 B
- ベイズの定理より、

$$\begin{aligned} P(B|A) &= \frac{P(A|B)P(B)}{P(A)} \\ &= \frac{1/(n+m+1)}{1/(n+1)} = \frac{(n+1)}{(n+m+1)} \end{aligned}$$

事前(主観)確率は重要

- $n \rightarrow \infty, m \rightarrow \infty$ なら、 $P(B|A) \rightarrow 0$
- だからといって、ずっと総合の学生にあったから、次は環境の学生に会うだろう、という推測は成立しない
- 事象 A に対するある程度の疑いは克服できない

- 事前確率の大小が問題ではない
- 事前確率が主観によることを認める意味で重要

確率変数に対するベイズの定理

- 観測値 Y とパラメータ θ の確率密度関数
 $p(\theta), p(y)$
- ベイズの定理

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

- 事前分布: $p(\theta)$
- 事後分布: $p(\theta|y)$

確率変数に対するベイズの定理

- 未知パラメータは確率的に変動する
- $p(y)$ は θ に依存しないため、定数と見なせる
→データが観測されると固定される

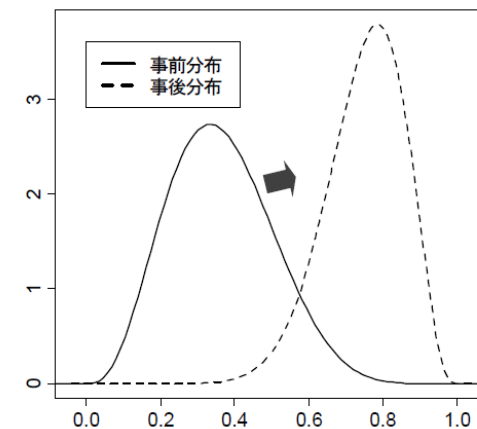
- 確率変数に対するベイズの定理

$$p(\theta|y) \propto p(y|\theta)p(\theta)$$

あるいは

$$p(\theta|y) \propto \ell(y|\theta)p(\theta)$$

事前分布と事後分布



正規分布とベイズの定理

$y \sim \mathcal{N}(\theta, 1)$, $\theta \sim \mathcal{N}(0, 10)$ のとき, $\theta|y$ は正規分布に従う.

$$\begin{aligned} p(\theta|y) &\propto \ell(y|\theta)p(\theta) \propto \exp\left(-\frac{(y-\theta)^2}{2} - \frac{\theta^2}{20}\right) \\ &\propto \exp\left(-\frac{11\theta^2}{20} + \theta y\right) \\ &\propto \exp\left(-\frac{11}{20}\left\{\theta - \left(\frac{10y}{11}\right)\right\}^2\right) \\ &\therefore \theta|y \sim \mathcal{N}\left(\frac{10}{11}y, \frac{10}{11}\right) \end{aligned}$$

線形回帰モデルのベイズ推定(2)

■ 線形回帰モデル

$$y_i = \beta x_i + \varepsilon_i \quad i = (1, \dots, n) \quad \varepsilon_i \sim \mathcal{N}(0, 1)$$

■ 誤差項の確率分布

$$p(\varepsilon) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}\varepsilon_i^2\right\} \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n \varepsilon_i^2\right\}$$

■ 被説明変数の事後確率

$$\begin{aligned} p(y|\beta, x) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(y_i - \beta x_i)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n (y_i - \beta x_i)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n \left(\beta - \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2}\right)^2 \sum_{i=1}^n x_i^2\right\} \end{aligned}$$

線形回帰モデルのベイズ推定(1)

■ 線形回帰モデル

$$y_i = \beta_0 + \beta_1 x_i, \varepsilon_i \sim \mathcal{N}(0, \sigma^2)$$

■ 説明変数 $X = (x_1, \dots, x_p, \dots, x_n)$ とパラメータ β_0, β_1 が既知の時、被説明変数は正規分布に従う

■ このとき、ベイズの定理を用いてパラメータを求めるには、以下の事後分布を計算する

$$p(\beta_0, \beta_1, \sigma^2 | Y, X) \propto \prod_{i=1}^n p(y_i, x_i | \beta_0, \beta_1, \sigma^2) \cdot p(\beta_0, \beta_1, \sigma^2)$$

線形回帰モデルのベイズ推定(3)

■ 尤度関数 $\ell(\beta|x, y) \propto \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^n \left(\beta - \frac{\sum_{i=1}^n y_i x_i}{\sum_{i=1}^n x_i^2}\right)^2\right\}$

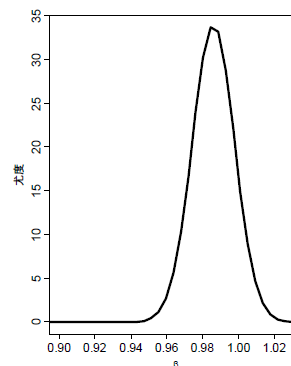


図 1.7 尤度関数の例

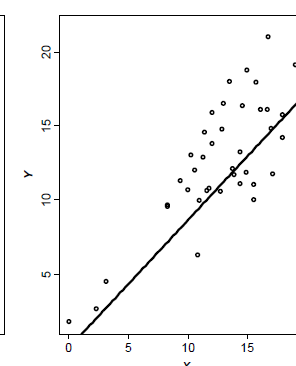


図 1.8 回帰直線

ベルヌーイ試行(1)



- n 回の試行で s 回成功し, その期待値 θ がわかっているとき, 実験が成功する確率

$$p(y|\theta, n) = \binom{n}{s} \theta^s (1 - \theta)^{n-s}$$

$$\binom{n}{s} = \frac{n!}{s!(n-s)!} = {}_n C_s$$

s 回成功
期待値 θ



$n-s$ 回成功
期待値 $1-\theta$

ベルヌーイ試行(2)

- $\theta^{a-1}(1-\theta)^{b-1}$ は以下のような性質を持つ

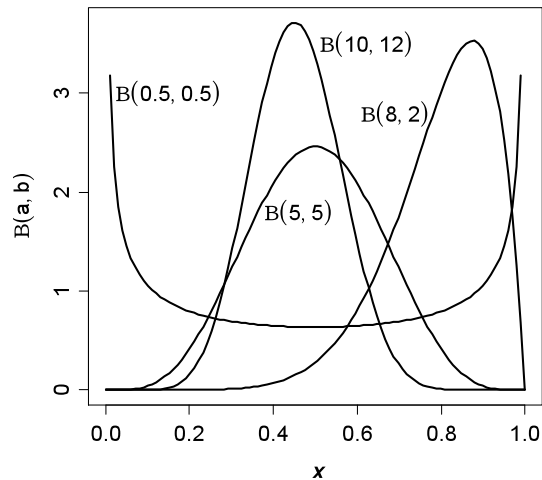
$$\int_0^1 \theta^{a-1} (1-\theta)^{b-1} d\theta = B(a, b) \quad \text{ベータ関数}$$

$$= \frac{(a-1)!(b-1)!}{(a+b-1)!}$$

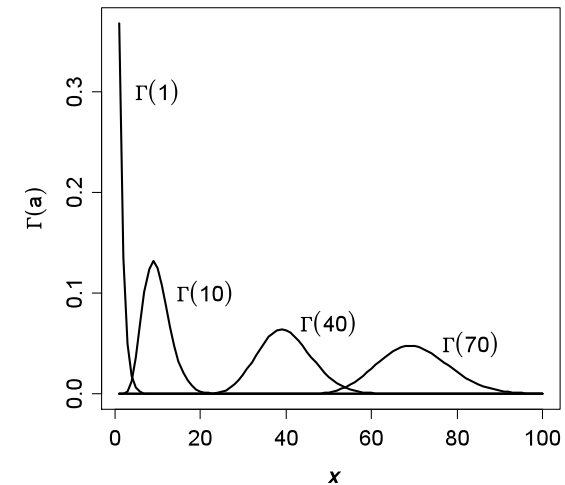
$$= \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx, a > 0 \quad \text{ガンマ関数}$$

ベータ関数の例



ガンマ関数の例



ベルヌーイ試行と尤度関数

- ベルヌーイ試行による尤度関数

$$p(y|\theta, n) = \binom{n}{s} \theta^s (1 - \theta)^{n-s}$$

- 区間[0, 1]で積分すると、事後確率は一意に得られる

$$\begin{aligned} \int_0^1 p(y|\theta, n) d\theta &= \int_0^1 \binom{n}{s} \theta^s (1 - \theta)^{n-s} d\theta \\ &= \frac{n!}{s!(n-s)!} \cdot \frac{s!(n-s)!}{(s+1+n-s+1-1)!} \\ &= \frac{1}{n+1} \end{aligned}$$

ベルヌーイ試行による尤度関数

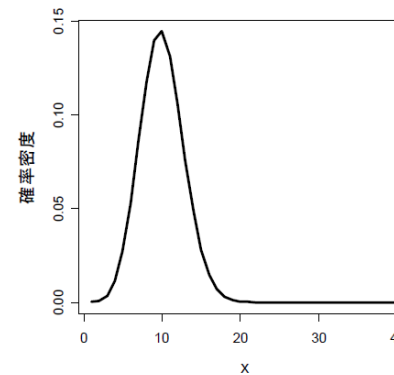


図 1.10 二項分布の例

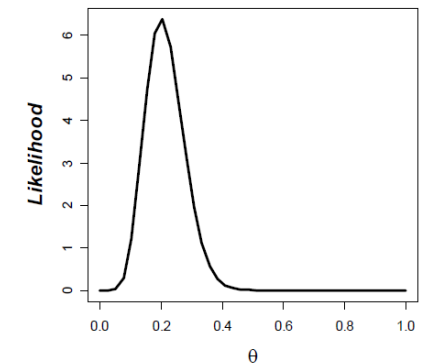


図 1.11 ベルヌーイ試行による尤度関数の例

尤度原理(1)

- モデル1では、40個の観測データのうち12個が観測されたとする
- このとき、期待値 θ に関する尤度関数は次式のよ
- うに表される

$$\ell_1(\theta|n=40, s=12) = \binom{40}{12} \theta^{12} (1 - \theta)^{28}$$

- 期待値 $\theta=0.25$ のとき、モデル1上にデータ観測される確率は、約0.1057となる

尤度原理(2)

- この問題を、別の観点から考える
- モデル1上に12個のデータが観測されるまでに、何個のデータを観測しなくてはならないのか
- 負の二項分布

$$p(n|\theta, s) = \binom{n-1}{s-1} \theta^s (1 - \theta)^{n-s}$$

$$\ell_2(\theta|n=40, s=12) = \binom{39}{11} \theta^{12} (1 - \theta)^{28}$$

$$\ell_1(\theta|n=40, s=12) \propto \ell_2(\theta|n=40, s=12)$$

尤度原理(3)

- 何回試行して何回成功したかという情報さえあれば(尤度が統計モデルに関するデータのすべての情報を持っていれば), **比例関係にある尤度関数からは同じ結論が導き出される**.
- 交換可能性の性質が証明されていることで, 尤度と事前情報が存在することが示される.

ドゥフィネッティの原理

- n 回の試行にたいして成功するかどうかの確率が, 次式で表せる
- 尤度と事前情報が存在することを証明

$$p(y_1, \dots, y_n) = \int_0^1 \theta^s (1 - \theta)^{n-s} p(\theta) d\theta$$

$$y \in \{0, 1\}, y = \sum_{i=1}^n y_i$$

かつ

$$\lim_{n \rightarrow \infty} \left(\frac{y}{n} \right) = p(\theta)$$

「同一性」と「可換性」

- 同一性 (identification)
 - すべての $y (\forall y \in \Omega)$ について $p(y|\theta_a) = p(y|\theta_b)$ が成立するとき, パラメータ θ は同一でなくてはならない
 - このとき, モデルは同一でなくてはならない
 - 現実には, 同一性が存在しないことも知られている
- 可換性(exchangeability)
 - n 回のベルヌーイ試行において, 成功する回数 s が同じであれば, 成功・失敗が生じる順序がどうであれ確率は同じである

ベイズ主義と頻度主義

- 頻度主義統計学では
 - 帰無仮説と対立仮説を立てて, 帰無仮説の統計的有意性を検定
 - モデルの期待値 $\theta = 0.25$ か $\theta \neq 0.25$ かが重要
- ベイズ主義統計学では
 - 帰無仮説に対して対立仮説が成立することも無視しない
 - θ をランダムな確率値として見なす

頻度主義統計学の場合

- 帰無仮説: $\theta=0.25$
- 対立仮説: $\theta>0.25$

$$\begin{aligned} p(y > 12|0.25,40) &= p(y = 12|0.25,40) + p(y = 13|0.25,40) + \dots + p(y = 40|0.25,40) \\ &= \binom{40}{12} 0.25^{12} (1 - 0.25)^{28} + \dots + \binom{40}{40} 0.25^{40} (1 - 0.25)^0 \\ &\cong 0.3 \end{aligned}$$

- p 値は十分に小さい確率値とはいえないため、帰無仮説を棄却
 - 「このモデルは統計的に有意でない」

ベイズ主義統計学の場合

- $\theta>0.25$ となる確率 $p(\theta>0.25) \doteq 0.7$
- 帰無仮説を棄却するほど十分に高い確率でない
と考える
- (同じことを言っているようだが) p 値を用いた仮説検定は、尤度原理に従っていないため、妥当ではないと考える

線形回帰モデル

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathcal{N}(0, \sigma^2)$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik} + \varepsilon_i,$$

$$\mathbf{y} = (y_1, \dots, y_i, \dots, y_n)^T,$$

$$X = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_i \\ \vdots \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{i1} & \vdots & x_{ik} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix},$$

$$\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^T, \boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T.$$

線形回帰モデルの尤度関数(1)

- 被説明変数の条件付き確率分布
 $\mathbf{y}|\boldsymbol{\beta}, \sigma^2, X \sim \mathcal{N}(X\boldsymbol{\beta}, \sigma^2)$
- 被説明変数の期待値と分散
 $\mathbb{E}(\mathbf{y}|\boldsymbol{\beta}, X) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_n x_{in}, \quad \mathbb{V}(\mathbf{y}|\sigma^2, X) = \sigma^2$
- 尤度関数

$$\begin{aligned} \ell(\boldsymbol{\beta}, \sigma^2 | \mathbf{y}, X) &= -\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - X\boldsymbol{\beta})^T (\mathbf{y} - X\boldsymbol{\beta})\right] \\ &= -\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_n x_{in})^2\right] \end{aligned}$$

線形回帰モデルの推定

- 最尤推定解 = 最小自乗解

$$\min_{\beta} (\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta) = \min_{\beta} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_{i1} - \dots - \beta_n x_{in})^2$$

- 不偏推定量と不偏分散

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T \mathbf{y} \\ \mathbb{V}(\mathbf{y}\hat{\beta} | \sigma^2, X) &= \sigma^2 (X^T X)^{-1} \\ \hat{\sigma}^2 = s^2 &= (\mathbf{y} - X\hat{\beta})^T (\mathbf{y} - X\hat{\beta})\end{aligned}$$

線形回帰モデルの尤度関数(2)

- 尤度関数

$$\begin{aligned}\ell(\beta, \sigma^2 | \mathbf{y}, X) &= -\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - X\hat{\beta})^T (\mathbf{y} - X\hat{\beta})\right. \\ &\quad \left. -\frac{1}{2\sigma^2} (\beta - \hat{\beta})^T (X^T X) (\beta - \hat{\beta})\right]\end{aligned}$$

- 逆ガンマ関数

$$\begin{aligned}\Gamma^{-1}(a, b) &= \frac{b^a}{\Gamma(a)} (\sigma^2)^{-a-1} \exp(-b(\sigma^2)^{-1}) \\ &\propto (\sigma^2)^{-a-1} \exp(-b(\sigma^2)^{-1})\end{aligned}$$

線形回帰モデルのベイズ推定

- β の事前情報と、 σ^2 の周辺事前情報

$$\begin{aligned}\beta | \sigma^2, X &\sim \mathcal{N}(\tilde{\beta}, \sigma^2 M^{-1}) \\ \sigma^2 | X &\sim \mathcal{IG}(a, b), a > 0, b > 0\end{aligned}$$

- ここで、 M は $k \times k$ の正定値対象行列
- \mathcal{IG} は逆ガンマ関数 Γ^{-1}

事後分布

- 事後情報の確率分布

$$\begin{aligned}p(\beta | \sigma^2, \mathbf{y}, X) &\sim \mathcal{N}((M + X^T X)^{-1} \{(X^T X)\hat{\beta} + M\tilde{\beta}\}, \sigma^2 (M + X^T X)^{-1}) \\ p(\sigma^2 | \mathbf{y}, X) &\sim \mathcal{IG}\left(\frac{n + 2a}{2}, \frac{2b + s^2 + (\tilde{\beta} - \hat{\beta})^T (M^{-1} + (X^T X)^{-1})^{-1} (\tilde{\beta} - \hat{\beta})}{2}\right)\end{aligned}$$

- 事後分布の平均

$$\begin{aligned}\mathbb{E}(\beta | \mathbf{y}, X) &= \mathbb{E}(\mathbb{E}(\beta | \sigma^2, \mathbf{y}, X) | \mathbf{y}, X) \\ &= (M + X^T X)^{-1} \{(X^T X)\hat{\beta} + M\tilde{\beta}\} \\ \mathbb{E}(\sigma^2 | \mathbf{y}, X) &= \frac{2b + s^2 + (\tilde{\beta} - \hat{\beta})^T (M^{-1} + (X^T X)^{-1})^{-1} (\tilde{\beta} - \hat{\beta})}{n + 2a - 2}, n \geq 2\end{aligned}$$

無情報事前分布

- 事前情報について何ら情報を持たない状況で、事前分布について適当にパラメータを設定して事後分布を得ようとする方法

今、我々は β , M , a , b について、何の情報ももっていない。

そこで、 $a = 2.1$, $b = 2$ とし、 σ^2 に関する事前平均と事前分散を、それぞれ $2/2.1 - 1 \cong 1.82$ 及び $2^2/(2.1 - 2)^2(2.1 - 2) \cong 33.06$ とする。また、 $\tilde{\beta} = 0$, $M = (1/c)I_k$ と定義する。このとき、 β の事後分布は、定数 c を用いて

$$\beta | \sigma^2, X \sim \mathcal{N}(0, c\sigma^2 I_k) \quad (2.46)$$

となる。

ジェフリーズの事前分布の適用

- 事前分布

$$p(\beta, \sigma^2 | X) \propto \sigma^{-2}$$

- 事後分布

$$\beta | \sigma^2, \mathbf{y}, X \sim \mathcal{N}(\hat{\beta}, \sigma^2 (X^T X)^{-1})$$

$$\sigma^2 | \mathbf{y}, X \sim \mathcal{IG}\left(\frac{n - k - 1}{2}, \frac{s^2}{2}\right)$$

信頼区間

- 得られた事後分布がどの程度信頼できるのか
- 100(1- α)信頼区間

$$p(\alpha/2 \leq \theta \leq 1 - \alpha/2) = \int_{\alpha/2}^{1-\alpha/2} p(\theta) d\theta = 1 - \alpha$$

- 最高事後密度
 - 100(1- α)信頼区間の最も短い区間

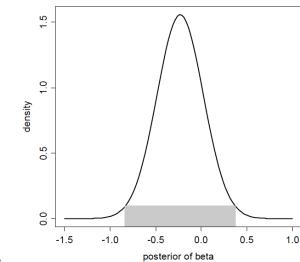


図 2.3 事後分布とその信頼区間

ゼルナーのG事前分布

- 未知パラメータに関する情報を明示的に扱う
 - 一般化自然共役事前分布

$$\beta | \sigma^2, X \sim \mathcal{N}(\tilde{\beta}, c\sigma^2 (X^T X)^{-1})$$

$$p(\sigma^2 | X) \propto \sigma^{-2}$$

- 事後分布 $p(\beta | \sigma^2, \mathbf{y}, X) \propto \ell(\mathbf{y} | \beta, \sigma^2, X) p(\beta, \sigma^2 | X)$

$$\propto (\sigma^{2s})^{-(n/2+1)} \exp\left[-\frac{1}{2\sigma^2} (\mathbf{y} - X\hat{\beta})^T (\mathbf{y} - X\hat{\beta})\right]$$

$$-\frac{1}{2\sigma^2} (\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta}) \Big] (\sigma^2)^{-k/2}$$

$$\times \exp\left[-\frac{1}{2c\sigma^2} (\beta - \hat{\beta})^T X^T X (\beta - \hat{\beta})\right]$$

ゼルナーのG事前分布

- 事後分布

$$\beta | \sigma^2, \mathbf{y}, X \sim \mathcal{N} \left(\frac{c}{c+1} \left(\frac{\tilde{\beta}}{c} + \hat{\beta} \right), \frac{c\sigma^2}{c+1} (X^T X)^{-1} \right)$$

$$\sigma^2 | \mathbf{y}, X \sim \mathcal{IG} \left(\frac{n}{2}, \frac{s^2}{2} + \frac{1}{2(c+1)} (\tilde{\beta} - \hat{\beta})^T X^T X (\tilde{\beta} - \hat{\beta}) \right)$$

- 事後平均

$$\mathbb{E}(\beta | \mathbf{y}, X) = \frac{c}{c+1} \left(\frac{\tilde{\beta}}{c} + \hat{\beta} \right)$$

$$\mathbb{E}(\sigma^2 | \mathbf{y}, X) = \frac{s^2 + (\tilde{\beta} - \hat{\beta})^T (M^{-1} + (X^T X)^{-1})^{-1} (\tilde{\beta} - \hat{\beta}) / (c+1)}{n-2}, n > 2$$

階層事前分布(2)

- 事後分布

$$p(\beta, \tau | \mathbf{y}, X) = \ell(\mathbf{y} | \beta, X) p(\beta, \tau)$$

$$\propto \tau^{n/2} \exp\{-(\tau/2)(\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta)\} \cdot (1/\tau)$$

$$\propto \tau^{n/2-1} \exp\{-(\tau/2)(\mathbf{y} - X\beta)^T (\mathbf{y} - X\beta)\}$$

$$\propto \tau^{n/2-1} \exp\{-(\tau/2)(\beta - b)^T X^T X (\beta - b)\} \times \exp\{-(\tau/2)v e^T e\}$$

- 事前情報はガンマ分布で表せる

$$p(\tau | \mathbf{y}, X) \propto \tau^{v/2-1} \exp\{-\tau v s^2 / 2\},$$

$$\tau = \Gamma(v/2, v s^2 / 2)$$

階層事前分布(1)

- 事前分布の影響を極力除去する

- 事前分布を階層的に与える

 - 例: 精度 $\tau = 1/\sigma^2$ を超パラメータとして与える

- 線形回帰モデルの誤差項に関する階層事前分布

$$p(\varepsilon | \tau) \propto \tau^{n/2} \exp\{-(\tau/2)\varepsilon^T \varepsilon\}$$

- 超パラメータの分布

$$p(\beta, \tau) \propto 1/\tau, -\infty < \beta < \infty, \tau > 0$$

階層事前分布(3)

- 1) 初期値である未知パラメータ β の事前分布と超パラメータ τ を与える

- 2) $\mathbf{y} | \beta$ を生成(sample)する

- 3) $\beta | \tau$ を生成する

線形回帰モデルのベイズ推定例

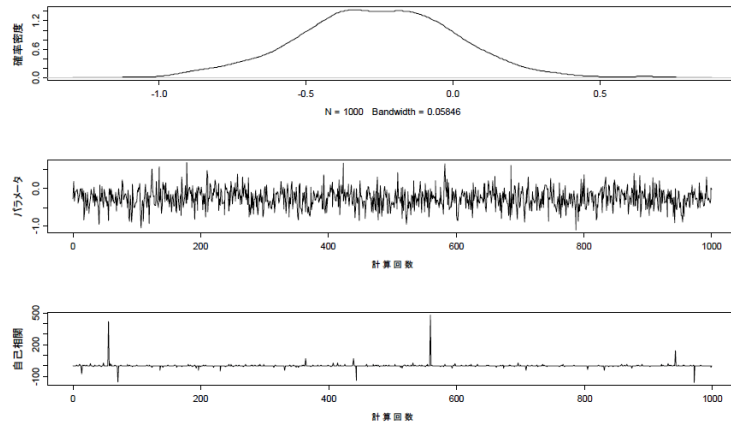


図 2.3 線形回帰モデルの 3 番目のパラメータの事後確率密度分布 (上)・
ndraw 回の計算結果 (中)・事後確率密度分布の自己相関 (下)

線形回帰モデルのベイズ推定例

- 平均、標準偏差、95%信頼区間をまとめた例

表 2.1 swiss データの線形回帰モデルのベイズ推定結果

変数	平均	標準偏差	97.5%	2.5%
定数項	67.208	1.286×10^{-2}	88.829	45.586
Agriculture	-0.172	5.319×10^{-3}	-0.030	-0.314
Examination	-0.246	6.682×10^{-2}	0.266	-0.759
Education	-0.883	3.260×10^{-2}	-0.513	-1.252
Catholic	0.105	1.193×10^{-3}	0.177	0.034
Infant.Mortality	1.059	1.520×10^{-1}	1.830	0.288

階層ベイズによる推定

- 誤差項の分散不均一を認める場合
- パラメータをサンプル(グループ)毎に推定する場合
 $y_i = \beta_i x + \varepsilon_i, \varepsilon_i \sim \mathcal{N}(0, \tau^{-1})$
- 階層事前情報に超パラメータを導入

$$p(\tau, \lambda | y) = \ell(y | \tau, \lambda) p(\tau | \lambda) p(\lambda)$$

- 事後分布

$$p(\beta_1, \dots, \beta_i, \dots, \beta_n, \tau, \lambda | y_i) = \prod_{i=1}^n \ell(y_i | \tau, \lambda) p(\beta_i | \tau) \times p(\tau | \lambda)$$

モデル選択

- パラメータ選択に関する判定方法
 - 事後オッズ
 - ベイズファクター
- モデル選択に関する判定方法
 - ベイズ情報量基準

モデルの事後分布

- ベイズ推定により、いくつかのモデル \mathcal{M}_i が推定されたとする
- あるモデルの事後分布確率を、他のすべてのモデル事後分布確率と比較することにより、どのモデルが確率的に生じやすいかを判定する
- 事後予測分布 $p(y)$ の下でモデル \mathcal{M}_i が生じる確率

$$P(\mathcal{M}_i|y) = \frac{p(y|\mathcal{M}_i)p(\mathcal{M}_i)}{p(y)}$$

$$p(y) = \sum_{i=1}^N p(y|\mathcal{M}_i)p(\mathcal{M}_i)$$

ベイズファクター(1)

- モデルパラメータの推定結果の妥当性を判断するために用いられる
- 事前分布が成立する仮説を仮説検定
 - 帰無仮説(モデル0) : $H_0: \theta \in \theta_0$
 - 対立仮説(モデル1) : $H_1: \theta \in \theta_1$

$$B_{10} = \frac{\int_{\theta_1} f(x)p_1(\theta)d\theta}{\int_{\theta_0} f(x)p_0(\theta)d\theta}$$

事後オッズとベイズファクター

- 二つのモデルについて、モデル0 \mathcal{M}_0 とモデル1 \mathcal{M}_1 の事後確率の比をとる = 事後オッズ

$$\text{事後オッズ} \quad \frac{P(\mathcal{M}_1|y)}{P(\mathcal{M}_0|y)} = \frac{\overbrace{p(y|\mathcal{M}_1)}^{\text{ベイズファクター}} \overbrace{p(\mathcal{M}_1)}^{\text{事前オッズ}}}{\overbrace{p(y|\mathcal{M}_0)}^{\text{ベイズファクター}} \overbrace{p(\mathcal{M}_0)}^{\text{事前オッズ}}}$$

ジェフリーズの判定方法

- 帰無仮説 H_0 を棄却するかどうか
- モデルを支持するかどうか

$\log B_{10}$ の値	帰無仮説 H_0	解釈
$\log B_{10} < 0$	非常に弱い	モデル \mathcal{M}_0 を支持
$0 < \log B_{10} < 0.5$	弱い	モデル \mathcal{M}_1 を弱く支持
$0.5 < \log B_{10} < 1$	相当なものである	モデル \mathcal{M}_1 を支持
$1 < \log B_{10} < 2$	強い	モデル \mathcal{M}_1 を強く支持
$2 < \log B_{10}$	決定的である	モデル \mathcal{M}_1 を決定的に支持

ベイズファクター(2)

- H0の事後分布

$$y|\beta^0, \sigma^2, X_0 \sim \mathcal{N}_n(X_0\beta^0, \sigma^2)$$

- H0の事前分布(ゼルナーのG事前分布)

$$\beta^0|\sigma^2, X_0 \sim \mathcal{N}_{k+1-q}(\tilde{\beta}^0, c_0\sigma^2(X_0^T X_0)^{-1})$$

- H0の事後分布

$$f(y|X_0, H_0) = (c+1)^{-(k+1-q)/2} (1/\sigma^2)^{-n/2} \Gamma(n/2)$$

$$\times \left[y^T y - \frac{c_0}{c_0+1} y^T X_0 (X_0^T X_0)^{-1} X_0^T y - \frac{1}{c_0+1} \tilde{\beta}_0^T X_0^T X_0 \tilde{\beta}_0 \right]^{-n/2}$$

ベイズファクター(3)

$$B_{10} = \frac{f(y|X, H_1)}{f(y|X_0, H_0)} = \frac{(c_0+1)^{-(k+1-q)/2}}{(c+1)^{-n/2}}$$

$$\times \left[\frac{y^T y - \frac{c_0}{c_0+1} y^T X_0 (X_0^T X_0)^{-1} X_0^T y - \frac{1}{c_0+1} \tilde{\beta}_0^T X_0^T X_0 \tilde{\beta}_0}{y^T y - \frac{c}{c+1} y^T X (X^T X)^{-1} X^T y - \frac{1}{c+1} \tilde{\beta}^T X^T X \tilde{\beta}} \right]^{-n/2}$$

情報量

- モデルパラメータ θ が与えられたとき、データ y が得られる情報量 $-\log p(y|\theta)$

- エントロピー=情報量の期待値

$$E\{-\log p(y|\theta)\} = -\sum_{\theta_m} \log p(y|\theta_m)$$

- 赤池の情報量基準(AIC)

$$AIC = -2\log p(y|\theta_m) + 2k$$

$$AIC' = -2\log p(y|\theta_m) + 2k(k+1)/(n-k-1)$$

ベイズ情報量基準(1)

- サンプル数 n が多いとき、ベイズファクターはベイズ情報量基準に近似できる
- パラメータの事後平均 $\hat{\theta}$ のとき、事後確率の対数は次式で表される

$$h(\theta) = \log[\ell(y|\theta)p(\theta)]$$

$$\int \exp\{h(\theta)\} d\theta \approx (2\pi)^{k/2} |\hat{H}(\hat{\theta})|^{1/2} \exp\{h(\hat{\theta})\}$$

$$\hat{H}(\hat{\theta}) = [h''(\hat{\theta})]^{-1}$$

ベイズ情報量基準(2)

- 事後分布

$$p(y) \approx (2\pi)^{k/2} |\hat{H}(\hat{\theta})|^{1/2} \ell(y|\hat{\theta}) p(\hat{\theta}) \quad h(\hat{\theta}) = \log[\ell(y|\hat{\theta}) p(\hat{\theta})]$$

- 両辺の対数をとると

$$\log p(y) \approx \log \ell(y|\hat{\theta}) + \log p(\hat{\theta}) + \frac{k}{2} \log(2\pi) + \frac{1}{2} \log |\hat{H}(\hat{\theta})|$$

- リンドレーのパラドックス

$$p(\hat{\theta}) = \frac{1}{\sqrt{(2\pi)^k}}$$

ベイズ情報量基準(3)

- 事後確率の対数

$$\log p(y) \approx \log \ell(y|\hat{\theta}) - \frac{k}{2} \log(n)$$

情報量が多いとき $\log |\hat{H}(\hat{\theta})| = -\log |h''(\hat{\theta})| \approx -\log(n^k)$

- ベイズ情報量基準

$$BIC = \frac{\ell(y|\hat{\theta}_1)}{\ell(y|\hat{\theta}_0)} n^{(k_1 - k_0)/2}$$

DIC: 偏差情報量基準

- モデルの当てはまりのよさ

$$D(y, \theta) = -2 \log[p(\theta|y)]$$

- モデルの複雑さ(有効なパラメータ数)

$$pD = \bar{D} - \hat{D},$$

\bar{D} : D の事後平均

$$\hat{D} = -2 \log[p(\hat{\theta}|y)]$$

- 偏差情報量基準(DIC)

$$DIC = pD + \bar{D} = \hat{D} + 2pD$$

- DICが最も小さいモデルがよいモデル

線形回帰モデルのベイズ推定例

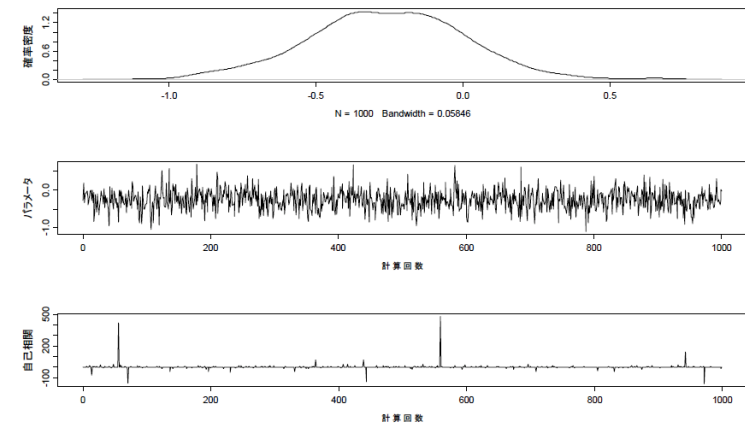
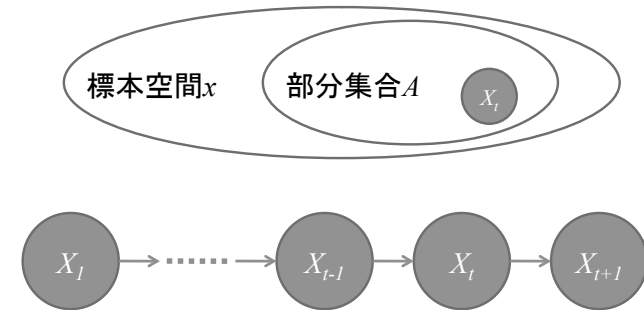


図 2.3 線形回帰モデルの3番目のパラメータの事後確率密度分布(上)・
ndraw 回の計算結果(中)・事後確率密度分布の自己相関(下)

ベイズ推定の繰り返し計算

- 与えられた事前情報に尤度関数を乗じて事後分布を得る
- これを非常にたくさんの回数を行う
 - シミュレーションによる事後分布の生成(サンプリング)
- 何回シミュレーションすればいいのか？
 - シミュレーション結果の信頼性は？
 - 収束するかどうかの判定方法は？

マルコフ連鎖



推移確率

$$\begin{aligned} P(X_{t+1} \in A | x_1, \dots, x_t) &= P(X_{t+1} \in A | x_t) \\ &= P(X_{t+1} | x_t) P(X_t | x_{t-1}) \cdots P(X_2 | x_1) P(x_1) \end{aligned}$$

定常分布

- 推移核

$$K(x, y) = P(X_{t+1} = y | X_t = x), \quad x, y \in A$$

- 定常分布

$$p(y) = \int K(x, y) p(x) dx$$

事後分布は、推移核に依存するが事前確率に依存せず

時系列回帰モデル

- AR(1)モデル

$$x_{t+1} = \rho x_t + \varepsilon_{t+1}, \varepsilon_{t+1} \sim \text{i.i.d. } \mathcal{N}(0, 1), t = 0, \dots, T$$

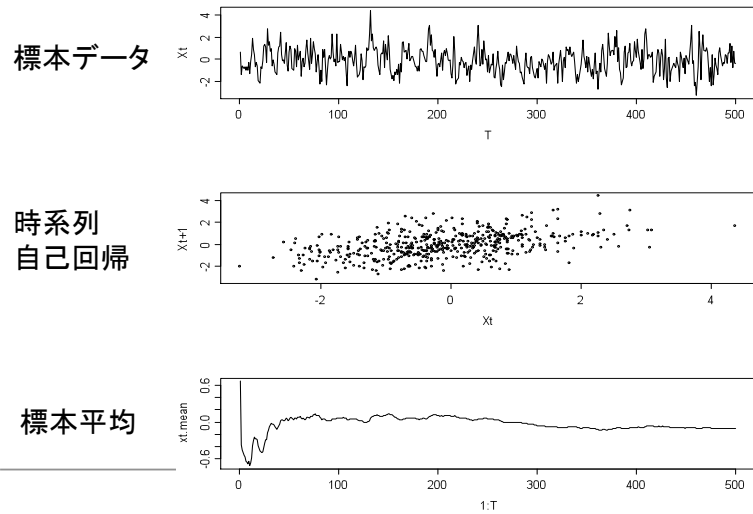
$$K(x_{t+1}, x_t) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}(x_{t+1} - \rho x_t)^2\right\}$$

- 期待値と分散

$$\mathbb{E}(X_{t+1}) = \rho \mathbb{E}(X_t)$$

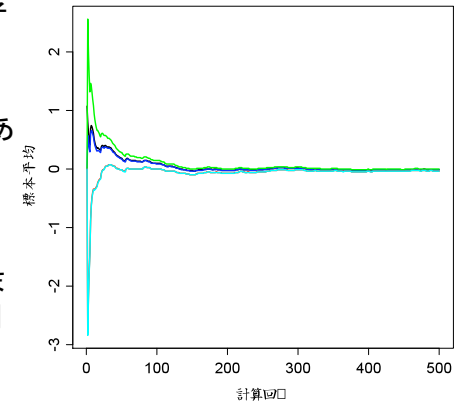
$$\mathbb{V}(X_{t+1}) = \rho^2 \mathbb{V}(X_t) + 1$$

AR(1)の標本データ



標本平均の収束過程

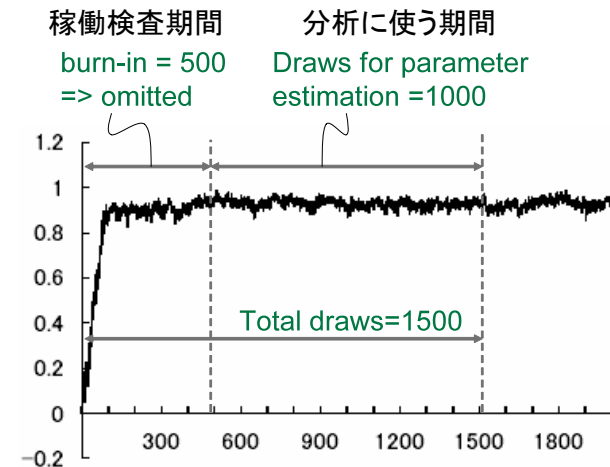
- 標本経路が初期値に依存
- 定常状態に至るまでの経路を棄却し、定常状態にある推定値を用いて議論することが必要
- サンプルングを棄却するまでの期間を稼働検査期間という



シミュレーション回数とパラメータとの関係

	$\rho = 0.0$		$\rho = 0.5$		$\rho = 0.9$		$\rho = 0.99$	
	\bar{x}_t	$sd(x_t)$	\bar{x}_t	$sd(x_t)$	\bar{x}_t	$sd(x_t)$	\bar{x}_t	$sd(x_t)$
$T=50$	-0.11414	1.10295	-0.34759	1.25428	-0.27350	2.01221	1.77826	1.79186
$T=500$	0.07139	0.99288	0.13185	1.23115	0.32551	2.71079	-1.43631	6.02930
$T=10^4$	-0.00584	1.00329	-0.00668	1.15513	-0.08652	2.26064	0.12971	7.92547
$T=10^6$	-0.00176	1.00105	0.00180	1.15448	-0.00794	2.29156	0.05549	7.11412

稼働検査期間と繰り返し回数



マルコフ連鎖の推移核を構築する方法

- メトロポリス・ヘイスティングス (M-H) 法
- ギブズ・サンプラー (M-H法の特殊タイプ)
- データ拡大法、データ補完アルゴリズム (欠損値に対処する上で有効な方法)
 - ギブズ・サンプリングの一種

ギブズ・サンプラー(2)

ギブズ・サンプラーのアルゴリズム

1. 繰り返し (draw) 回数を $s=(0, \dots, ndraw)$ とする
2. $s=0$ に対して, 初期値 $\mathbf{x}^s = (x_1^s, \dots, x_t^s)$ を決める
3. \mathbf{x}^s から \mathbf{x}^{s+1} を生成する
 - 3.1 $x_1^{s+1} \sim p(x_1 | x_2^s, \dots, x_t^s)$ から x_1^{s+1} を生成
 - 3.2 $x_2^{s+1} \sim p(x_2 | x_1^{s+1}, x_3^s, \dots, x_t^s)$ から x_2^{s+1} を生成
 - ⋮
 - 3.t $x_t^{s+1} \sim p(x_t | x_1^{s+1}, \dots, x_{t-1}^{s+1})$ から x_t^{s+1} を生成
4. $s < ndraw$ のとき, step 3 に戻る. $s = ndraw$ のとき, 計算終了

ギブズ・サンプラー(1)

- 目標分布とその確率密度分布がいくつかに分解できると考える
- 目標分布を $y = (y_1, y_2)$ のように分割
- 推移核 $K(x, y) = K(x_1, x_2, y_1, y_2)$
- 事後分布 $p_{Y_1|Y_2}(y_1|y_2)$ および $p_{Y_2|Y_1}(y_2|y_1)$

ギブズ・サンプラー(3)

- 二変量正規分布のギブズサンプラー
- 二変量正規分布 $\begin{cases} y_1 \sim \mathcal{N}(\mu_1, \rho) \\ y_2 \sim \mathcal{N}(\mu_2, \rho) \end{cases}$
- 事後分布 $\begin{cases} y_1|y_2 \sim \mathcal{N}\left(\mu_2 + \rho \frac{\sigma_2}{\sigma_1} (y_1 - \mu_1), \sigma_2^2(1 - \rho^2)\right) \\ y_2|y_1 \sim \mathcal{N}\left(\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (y_2 - \mu_2), \sigma_1^2(1 - \rho^2)\right) \end{cases}$

ギブズ・サンプラー(4)

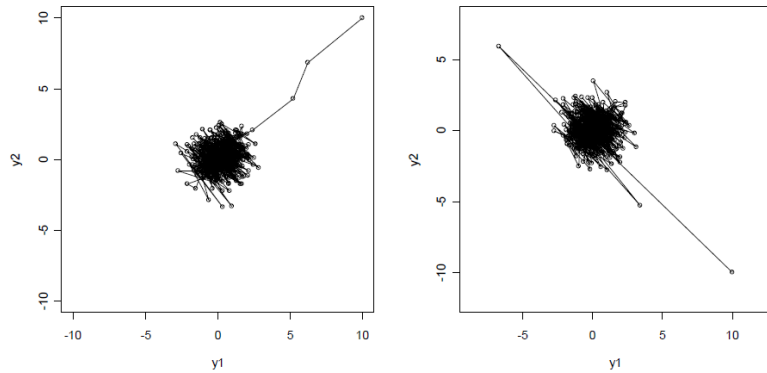


図 3.3 二変量正規分布ギブズ・サンプラーの収束過程
(左: 初期値(10, 10), 右: 初期値(10, -10))

線形回帰モデルのギブズ・サンプリング

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \sigma^2)$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i,$$

$$\mathbf{y} = (y_1, \dots, y_i, \dots, y_n)^T,$$

$$X = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_i \\ \vdots \\ \mathbf{x}_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{i1} & \vdots & x_{ik} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix},$$

$$\boldsymbol{\beta} = (\beta_0, \dots, \beta_k)^T, \boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^T.$$

線形回帰モデルのギブズ・サンプリング

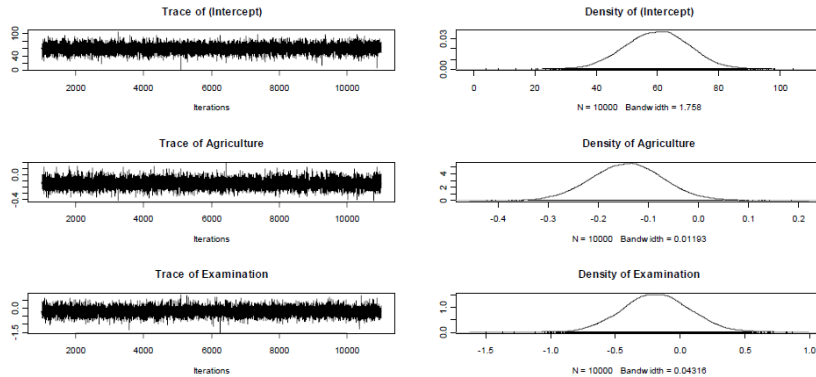
- 事前分布 $p(\boldsymbol{\beta}|\tau) \sim \mathcal{N}(b_0, (1/\tau)B_0)$
 $p(\tau|\mathbf{y}) \sim \mathcal{G}(v_0/2, v_0 S_0/2)$
 $B = X^T X \quad \tau = 1/\sigma^2$
- 尤度関数 $p(\mathbf{y}|\boldsymbol{\beta}, \tau) \sim \mathcal{N}(X\boldsymbol{\beta}, (1/\tau)I)$
- 事後分布 $v_1 = v_0 + v,$
 $v_1 S_1 = v_0 S_0 +$
 $p(\boldsymbol{\beta}|\tau, \mathbf{y}) \sim \mathcal{N}(b_1, B_1)$
 $p(\tau|\boldsymbol{\beta}, \mathbf{y}) \sim \mathcal{G}(v_1/2, v_1 S_1/2)$
 $(\mathbf{y} - Xb_1)^T (\mathbf{y} - X\boldsymbol{\beta}),$
 $p(\sigma^2|\boldsymbol{\beta}, \mathbf{y}) \sim \mathcal{IG}(v_1/2, v_1 S_1/2)$ ——— $b_1 = B_1(B_0^{-1}b_0 + \tau X^T \mathbf{y}),$
 $B_1^{-1} = B_0^{-1} + \tau B$

線形回帰モデルのギブズサンプリング

1. 繰り返し (draw) 回数を $s=(0, \dots, ndraw)$ とする
2. $s=0$ に対して, 初期値 $(\boldsymbol{\beta}^s, \tau^s)$ を決める
3. $(\boldsymbol{\beta}^s, \tau^s)$ から $(\boldsymbol{\beta}^{s+1}, \tau^{s+1})$ を生成する
 - i. $p(\boldsymbol{\beta}^{s+1}|\tau^s, \mathbf{y}) \sim \mathcal{N}(b_1, B_1)$ から $\boldsymbol{\beta}^{s+1}$ を生成
 ただし, $b_1 = B_1(B_0^{-1}b_0 + \tau^s X^T \mathbf{y}), B_1^{-1} = B_0^{-1} + \tau^s B$
 - ii. $p(\tau^{s+1}|\boldsymbol{\beta}^{s+1}, \mathbf{y}) \sim \mathcal{G}(v_1/2, v_1 S_1/2)$ から τ^{s+1} を生成
4. $s < ndraw$ のとき, step 3 に戻る. $s = ndraw$ のとき, 計算終了

線形回帰モデルのギブズサンプリング

- サンプルング11000回、稼働検査期間1000回



ギブズ・サンプラー

ギブズ・サンプラーのアルゴリズム

1. 繰り返し (draw) 回数を $s=(0, \dots, ndraw)$ とする
2. $s=0$ に対して, 初期値 $\mathbf{x}^s = (x_1^s, \dots, x_t^s)$ を決める
3. \mathbf{x}^s から \mathbf{x}^{s+1} を生成する
 - 3.1 $x_1^{s+1} \sim p(x_1 | x_2^s, \dots, x_t^s)$ から x_1^{s+1} を生成
 - 3.2 $x_2^{s+1} \sim p(x_2 | x_1^{s+1}, x_3^s, \dots, x_t^s)$ から x_2^{s+1} を生成
 - ⋮
 - 3.t $x_t^{s+1} \sim p(x_t | x_1^{s+1}, \dots, x_{t-1}^{s+1})$ から x_t^{s+1} を生成
4. $s < ndraw$ のとき, step 3 に戻る. $s = ndraw$ のとき, 計算終了

線形回帰モデルのギブズ・サンプリング

- 事前分布 $p(\beta | \tau) \sim \mathcal{N}(b_0, (1/\tau)B_0)$
 $p(\tau | \mathbf{y}) \sim \mathcal{G}(v_0/2, v_0 S_0/2)$
 $B = X^T X \quad \tau = 1/\sigma^2$
- 尤度関数 $p(\mathbf{y} | \beta, \tau) \sim \mathcal{N}(X\beta, (1/\tau)I)$
- 事後分布

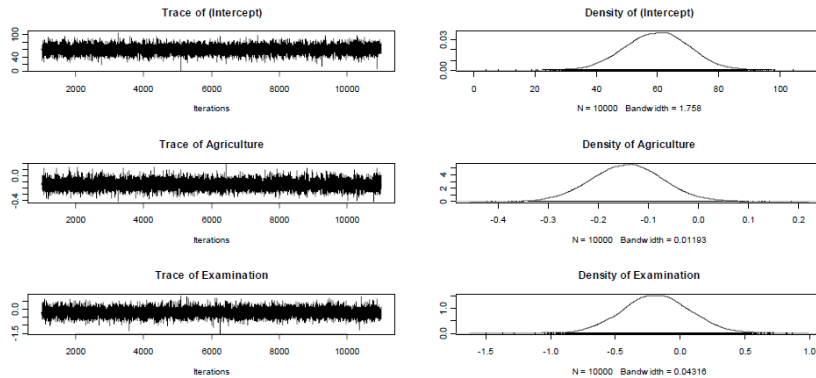
$p(\beta \tau, \mathbf{y}) \sim \mathcal{N}(b_1, B_1)$	$v_1 = v_0 + v,$
$p(\tau \beta, \mathbf{y}) \sim \mathcal{G}(v_1/2, v_1 S_1/2)$	$v_1 S_1 = v_0 S_0 +$
$p(\sigma^2 \beta, \mathbf{y}) \sim \mathcal{JG}(v_1/2, v_1 S_1/2)$	$(\mathbf{y} - Xb_1)^T (\mathbf{y} - X\beta),$
	$b_1 = B_1(B_0^{-1}b_0 + \tau X^T \mathbf{y}),$
	$B_1^{-1} = B_0^{-1} + \tau B$

線形回帰モデルのギブズサンプリング

1. 繰り返し (draw) 回数を $s=(0, \dots, ndraw)$ とする
2. $s=0$ に対して, 初期値 (β^s, τ^s) を決める
3. (β^s, τ^s) から $(\beta^{s+1}, \tau^{s+1})$ を生成する
 - i. $p(\beta^{s+1} | \tau^s, \mathbf{y}) \sim \mathcal{N}(b_1, B_1)$ から β^{s+1} を生成
 ただし, $b_1 = B_1(B_0^{-1}b_0 + \tau^s X^T \mathbf{y}), B_1^{-1} = B_0^{-1} + \tau^s B$
 - ii. $p(\tau^{s+1} | \beta^{s+1}, \mathbf{y}) \sim \mathcal{G}(v_1/2, v_1 S_1/2)$ から τ^{s+1} を生成
4. $s < ndraw$ のとき, step 3 に戻る. $s = ndraw$ のとき, 計算終了

線形回帰モデルのギブズサンプリング

- サンプリング11000回、稼働検査期間1000回



M-Hアルゴリズム

1. 繰り返し (draw) 回数を $s=(0, \dots, ndraw)$ とする
2. $s=0$ に対して, 任意の変数 $\mathbf{x}^s = (x_1^s, \dots, x_t^s)$ を初期値として与える
3. \mathbf{x}^s から $\tilde{\mathbf{x}} = q(\tilde{\mathbf{x}} | \mathbf{x}^s)$ を生成する
4. 受容確率 $\rho(\mathbf{x}^s, \tilde{\mathbf{x}})$ を計算する

$$\rho(\mathbf{x}^s, \tilde{\mathbf{x}}) = \min\left(\frac{p(\tilde{\mathbf{x}}) q(\mathbf{x}^s, \tilde{\mathbf{x}})}{p(\mathbf{x}^s) q(\tilde{\mathbf{x}}, \mathbf{x}^s)}, 1\right)$$

ただし, $p(\mathbf{x}^s)q(\tilde{\mathbf{x}}, \mathbf{x}^s) = 0$ のとき $\rho(\mathbf{x}^s, \tilde{\mathbf{x}}) = 1$ とおく

5. 一様乱数 $u \sim U(0,1)$ を発生させ, 次のように判定する

$$\mathbf{x}^s = \begin{cases} \tilde{\mathbf{x}}, & \text{if } u \leq \rho(\mathbf{x}^s, \tilde{\mathbf{x}}) \\ \mathbf{x}^s, & \text{otherwise} \end{cases}$$

6. $s < ndraw$ のとき, step 3 に戻る. $s = ndraw$ のとき, 計算終了

メトロポリス-ヘイスティングス法

- 提案分布 $q(\mathbf{x}^s | \mathbf{x})$ を生成
- 目標分布と提案分布との差を小さくするステップを取り入れる
- 次式のような受容確率を用いて, 一定の範囲内でMCMCを稼働させる

$$\rho(\mathbf{x}^s, \tilde{\mathbf{x}}) = \min\left(\frac{p(\tilde{\mathbf{x}}) q(\mathbf{x}^s, \tilde{\mathbf{x}})}{p(\mathbf{x}^s) q(\tilde{\mathbf{x}}, \mathbf{x}^s)}, 1\right)$$

Acceptance-Rejection (A-R) ステップ

- 提案密度の近似分布として, 正規分布を考える

$$\rho(\mathbf{x}^s, \tilde{\mathbf{x}}) = \min(q(\mathbf{x}^s, \tilde{\mathbf{x}}), ch(\mathbf{x}^s, \tilde{\mathbf{x}}))$$

$$h(\mathbf{x}^s, \tilde{\mathbf{x}}) \sim \mathcal{N}(m, V)$$

- ARステップのアルゴリズム

- a. $h(\mathbf{x}^s, \tilde{\mathbf{x}}) \sim \mathcal{N}(m, V)$ を生成する
- b. 一様分布 $u \sim U(0,1)$ に従う $u \sim U(0,1)$ を生成する
 - i. $u \leq q(\mathbf{x}^s, \tilde{\mathbf{x}}) / ch(\mathbf{x}^s, \tilde{\mathbf{x}})$ のとき $\tilde{\mathbf{x}}$ を受容する
 - ii. それ以外の場合, ステップ a. に戻る

AR-MHアルゴリズム

1. 繰り返し (draw) 回数を $s=(0, \dots, ndraw)$ とする
2. $s=0$ に対して, 任意の変数 $\mathbf{x}^s = (x_1^s, \dots, x_t^s)$ を初期値として与える
3. \mathbf{x}^s から $\tilde{\mathbf{x}} = q(\mathbf{x}^s, \mathbf{x})$ を生成する
4. AR ステップを実行する
 - a. $h(\mathbf{x}^s, \tilde{\mathbf{x}}) \sim \mathcal{N}(m, V)$ を生成する
 - b. 一様分布 $U(0,1)$ に従う $u \sim U(0,1)$ を生成する
 - i. $u \leq q(\mathbf{x}^s, \tilde{\mathbf{x}}) / \text{ch}(\mathbf{x}^s, \tilde{\mathbf{x}})$ のとき \mathbf{x}^s を受容する
 - ii. それ以外の場合, ステップ a. に戻る
5. 受容確率 $\rho(\mathbf{x}^s, \tilde{\mathbf{x}})$ を計算する

$$\rho(\mathbf{x}^s, \tilde{\mathbf{x}}) = \min\left(\frac{p(\tilde{\mathbf{x}}) q(\mathbf{x}^s, \tilde{\mathbf{x}})}{p(\mathbf{x}^s) q(\tilde{\mathbf{x}}, \mathbf{x}^s)}, 1\right)$$

ただし, $p(\mathbf{x}^s) q(\tilde{\mathbf{x}}, \mathbf{x}^s) = 0$ のとき $\rho(\mathbf{x}^s, \tilde{\mathbf{x}}) = 1$ とおく

6. 一様乱数 $u \sim U(0,1)$ を発生させ, 次のように判定する
7. $\mathbf{x}^s = \begin{cases} \tilde{\mathbf{x}}, & \text{if } u \leq \rho(\mathbf{x}^s, \tilde{\mathbf{x}}) \\ \mathbf{x}^s, & \text{otherwise} \end{cases}$
8. $s < ndraw$ のとき, step 3 に戻る. $s = ndraw$ のとき, 計算終了

Gelman-Rubin統計量

- Within chain variance (W)
- Between chain variance (B)

$$W = \frac{1}{m(n-1)} \sum_{j=1}^m \sum_{i=1}^n (\theta^{(i,j)} - \bar{\theta}^{(j)})^2$$

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

- Gelman-Rubin統計量

$$R = \sqrt{\hat{V}(\theta) / W}$$

$$\hat{V}(\theta) = (1 - 1/n)W + (1 - n)B$$

- 1.05以下なら収束しているとみなす

パラメータの収束判定方法

- 標本経路が初期値に依存しない定常状態であるかどうかを判定
- 稼働検査期間を除く標本経路を対象に判定を行う。
 - Gelman-Rubin統計量
 - Gewekeの方法
 - Raftery-Lewisの方法

```
# Gelman-Rubin 統計量
```

```
sample.gibbs.post2 <-
```

```
MCMCregress(Fertility~., data=swiss, burnin=10000, mcmc=100000,
```

```
marginal.likelihood="Chib95", b0=1, B0=0.05, c0=0.1, d0=0.1)
```

```
sample.gibbs.post.all <- mcmc.list(sample.gibbs.post1, sample.gibbs.post2)
```

```
gelman.diag(sample.gibbs.post.all)
```

```
> gelman.diag(sample.gibbs.post.all)
```

```
Potential scale reduction factors:
```

	Point est.	97.5% quantile
(Intercept)	1.16	1.56
Agriculture	1.06	1.23
Examination	1.02	1.11
Education	1.00	1.01
Catholic	1.00	1.00
Infant.Mortality	1.09	1.35
sigma2	1.00	1.02

```
Multivariate psrf
```

```
1.09+0i
```

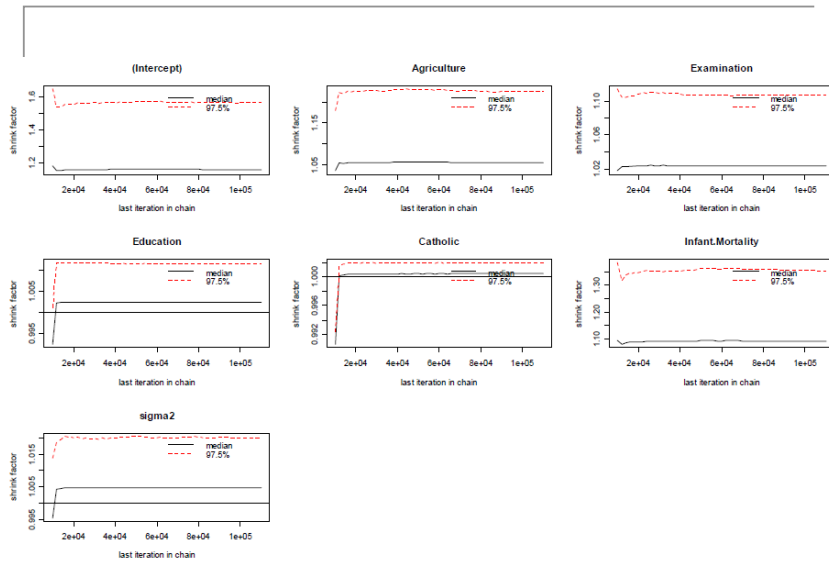



図 3.6 Gelman-Rubin 統計量の shrink factor

Gewekeの判定方法

- 平均値の差に関するZ検定を応用
- 最初の n_1 個 ($i = 1, \dots, n_1$) (最初の10%)
- 最後の $n - n_2 + 1$ 期間 (最後の50%)
- 二つの標本期間の平均値

$$\bar{g}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} g_1(\theta^{(i)}), \bar{g}_2 = \frac{1}{n_2} \sum_{i=n-n_2+1}^{n_2} g_2(\theta^{(i)})$$

- 平均値の差に関するZ検定量

$$Z = \frac{\bar{g}_1 - \bar{g}_2}{\sqrt{V(\bar{g}_1) + V(\bar{g}_2)}}$$

□ $|Z| > z_\alpha$ のとき収束していない

```
# Geweke の方法
geweke.diag(sample.gibbs.post1)
```

```
> geweke.diag(sample.gibbs.post2)
Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5
(Intercept)    Agriculture    Examination    Education    Catholic
0.7880         -0.5057         -0.7776         -0.1024         -0.8543
Infant.Mortality    sigma2
-0.2379         -0.1505
```

GEWEKE CONVERGENCE DIAGNOSTIC (Z-score)

```
=====
Iterations used = 10001:110000
Thinning interval = 1
Sample size per chain = 100000
$chain1
Fraction in 1st window = 0.1
Fraction in 2nd window = 0.5
(Intercept)    Agriculture    Examination    Education    Catholic
0.8127         -0.5216         -0.7695         -0.1145         -0.8412
Infant.Mortality    sigma2
-0.2598         -0.1150
```

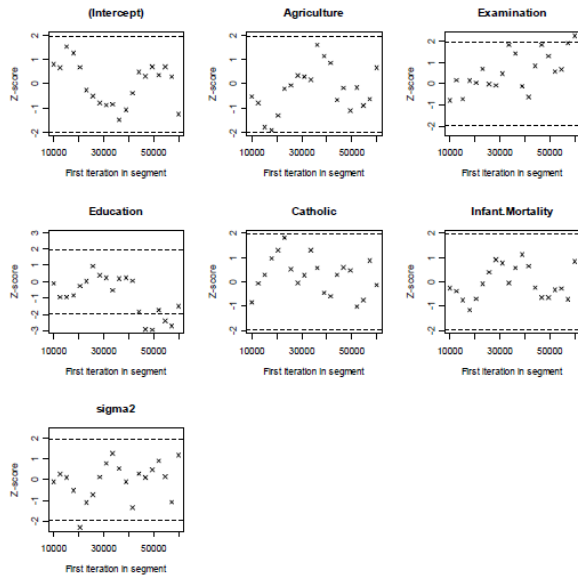
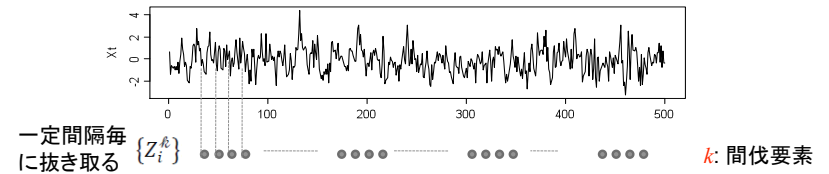


図 3.5 Geweke の判定方法による Z 値の表示

Raftery-Lewisの診断方法



繰り返し回数 $c_0 k$ が適切かを判断する方法
 一次マルコフ連鎖 $\{Z_i^k\}$ のうち、
 状態1 → 状態2 への変化率 α
 状態2 → 状態1 への変化率 γ

$$c_0 = \ln \left[\frac{(\alpha + \gamma)\epsilon}{\max\{\alpha, \gamma\}} \right] \cdot \frac{1}{\ln(|1 - \alpha - \gamma|)}$$

Raftery-Lewis の方法

```
raftery.diag(sample.gibbs.post1)
```

```
> raftery.diag(sample.gibbs.post1)
```

Quantile (q) = 0.025

Accuracy (r) = +/- 0.005

Probability (s) = 0.95

	Burn-in (M)	Total (N)	Lower bound (Nmin)	Dependence factor (I)
(Intercept)	2	3868	3746	1.03
Agriculture	1	3748	3746	1.00
Examination	2	3805	3746	1.02
Education	2	3832	3746	1.02
Catholic	2	3805	3746	1.02
Infant.Mortality	1	3755	3746	1.00
sigma2	2	3939	3746	1.05

おつかれさまでした