The Japanese Journal of Psychonomic Science
2014, Vol. 32, No. 2, 223–231

講演論文

ベイズ統計による情報仮説の評価は分散分析にとって代わるのか？[1]

岡　田　謙　介
専修大学

Does Bayesian evaluation of informative hypothesis outperform analysis of variance?

Kensuke Okada
Senshu University

The analysis of variance (ANOVA) has long held the status of being the most used (or I should say "abused") statistical technique in psychological research. Although there is no doubt about the usefulness of ANOVA, it is not free from disadvantages such as difficulty in accepting null hypothesis and multiplicity problem with multiple tests. In order to deal with these problems, the objective of this paper is to introduce Bayesian evaluation of informative hypothesis to psychonomic researchers. An informative hypothesis consists of inequality constraints between the parameters of interest. The relevance of informative hypothesis is evaluated thorough the Bayes factor against the unconstrained hypothesis. The calculation of Bayes factor is generally performed by means of Markov chain Monte Carlo techniques. This approach is illustrated by analysis of experimental data with mice having different nighttime light conditions.

Key words: informative hypothesis, unconstrained hypothesis, Bayes factor, posterior model probability, analysis of variance

はじめに

心理学，とくに基礎心理学領域において，分散分析はもっともよく利用され，またときに濫用されてきた統計手法と言えるであろう。Howell（2009）は心理学研究で分散分析が広く用いられていることの主な理由として，複数の水準間での比較ができること，複数の要因を考慮できることを挙げている。分散分析のこうした特徴は，とくに実験によって得られるデータと相性がよい。歴史的にみても，分散分析は実験計画法と密接な関係を持って発展してきた方法論である。

一般的な分散分析の応用においては，まず設定された要因の主効果や交互作用の有無が検討される。具体的には，各要因の主効果および交互作用が0であることを帰無仮説とした仮説検定（F検定）が行われる。この結果が有意であった場合には，続いてTukey法やBonferroni法など，各種の方法を用いて多重比較が行われる。そして，これら一連の分析結果に基づいて，研究者によって結果が解釈される。

しかしながら，こうした一般的な分散分析の方法論には様々な問題点が指摘できる。まず，仮説検定の枠組みを用いていることに由来する問題があり，ここでは2点を指摘する。第一に，仮説検定のロジックからして，帰無仮説H_0を積極的に支持することができない。仮説検定で用いられる検定統計量やp値は，帰無仮説H_0が正しいことを前提として導出されている。p値とは，帰無仮説H_0が正しいことを仮定したもとで，今回データから得られたよりも極端な，すなわち帰無仮説と整合的で

Corresponding address: Department of Psychology, Senshu University, 2–1–1, Higashimita, Tama-ku, Kawasaki-shi, Kanagawa 214–8580, Japan. E-mail: ken@psy.senshu-u.ac.jp

[1] 本研究は私立大学戦略的研究基盤形成支援事業（S11011013）および科学研究費補助金（23300310, 24730544, 25380871）の助成を受けています。本表題は狩野（2002）に依拠するものです。

ない検定統計量の値が得られる確率である。そのため，p値があらかじめ定めた有意水準よりも小さいことは，前提である帰無仮説を棄却するための証拠となり得る。しかしながら，仮説検定では対立仮説が正しい場合のもとでの議論は行われないため，p値が小さくない（有意水準よりも大きい）ことは帰無仮説を積極的に支持する証拠にはなり得ないのである。

第二に，仮説検定の結果は，標本サイズへの単調な依存関係がある。すなわち，母集団の効果量が同じ場合には，標本サイズNが大きければ大きいほど検定結果は有意になりやすくなる。p値は効果の大きさを表す量ではなく，検定で帰無仮説が棄却されることは，考えている効果が現実的に意味のあるものであることを含意しない。以上2点に代表されるような仮説検定の各種の問題点は，心理学における統計改革の流れで近年繰り返し指摘されている（Cumming, 2012; Kline, 2004; 大久保・岡田，2012）。

さらに，仮説検定全般というよりもこの分散分析の枠組みに特徴的な問題点も複数指摘できる。第一に，1段階目に行われる主効果・交互作用のF検定と，第2段階目に行われる多重比較とは一般に結果に整合性がない。たとえばF検定の結果ある主効果が有意になったにもかかわらず，多重比較の結果どの水準にも差がないといったことは実際に生じうる。これは本来別々の別々の分析（F検定と多重比較）をいわゆる「分散分析」という1つの方法論にまとめたために生じた問題であり，理論的には当然のことであるが，不一致な結果が出た場合にその解釈をどのように行えばよいのかは悩ましい問題である。第二に，分散分析のF検定が依って立つ平方和の分解は，釣り合い型（balanced design），すなわち各セルの標本サイズが等しい場合に成立する。現実的には脱落や欠測などによって，非釣り合い型，すなわち各セルの標本サイズが等しくないデータを分析しなければならないことが多い。この場合にはタイプI，タイプIIIといった各種平方和を利用することによる対応がなされるが，釣り合い型の場合のようにな単純な平方和の分解が成り立つわけではなく，あくまでいわば妥協の産物と言える。

さらに，いわゆる分散分析モデルがあまりに定着しているため，実験デザインやデータに統計モデルを合わせるのではなく，既存の分散分析モデルにあわせてデータを収集することになってしまっていることも指摘できる。このため，統計モデリングとしての観点からは考慮に入れられるべきデータの階層性や混合性などが，分散分析を用いたデータ分析では考慮に入れられない場合も多い。

本稿は分散分析の枠組みの有用性を疑うものではない。しかしながら，分散分析が万能ではないこともまた事実である。上記で見てきたように，分散分析は検定に由来する問題，分散分析の枠組みに由来する問題，応用上での問題といった各種問題を抱えている。

ここで，多要因・多群の平均値の比較に関する新しい枠組みとして，Hoijtink（2011），Hoijtink, Klugkist, & Boelen（2008）らの提唱する，ベイズ統計による情報仮説の評価（Bayesian evaluation of informative hypotheses）がある。ベイズ統計による情報仮説の評価では，帰無仮説を設定する古典的な仮説検定を用いない。また多くのベイズ統計の応用と同様に，多重性の問題への対処も容易である。そこで本稿では，ベイズ統計による情報仮説の評価の枠組みについて概説し，またそれをマウスの実験データに適用し知見を得ることを目的とする。第2節では，ベイズ統計学の枠組みと，モデル選択のためのベイズファクターについて述べる。第3節ではベイズ統計による情報仮説の評価の枠組みを述べる。第4節では，この枠組みをマウスの実データに適用した例を示す。第5節では近年の話題を扱い，最後にまとめの議論を行う。

ベイズ統計学とベイズファクター

ベイズ統計学は，データの情報を利用し，ベイズの定理によって母数に関する知識を更新していく枠組みである。それは公理から出発する演繹的な体系であり，矛盾なく推論の結果に到達することができる（繁桝，1985）。伝統的な統計学（しばしば頻度論frequentist theoryとも呼ばれる）においては，母数は未知の定数であり，標本データおよびそこから計算される標本統計量を確率変数と考える。これに対して，ベイズ統計学では，データは分析者の手元にあるのであるから定数であり，母数の値は分析者にとって未知であるから確率変数であると考える。このデータと母数のどちらを確率変数として扱うかの違いが，伝統的な統計学（頻度論）とベイズ統計学の違いの本質である（Table 1）。

頻度論の立場をとるかベイズ統計学の立場をとるかに関わらず，統計的データ分析において，仮説とは母数に関する仮説である。すなわち母数をθで表記すると，仮説は$H(\theta)$と表すことができる。通常これを略記して単にHで表す。ベイズ統計学においては母数θは確率変数であるので，仮説$H(\theta)$の適切さも確率によって評価することができる。

いま，データXを得る前の，ある仮説H_iが正しい確率（仮説H_iの事前確率）を$p(H_i)$で，またデータXを得た後で仮説H_iが正しい確率（仮説H_iの事後確率）を$p(H_i \mid X)$で表すことにする。このとき，事後確率$p(H_i \mid X)$はベ

Table 1.
The key difference between classical (frequentist) and Bayesian statistics

	頻度論 frequentist	ベイズ Bayesian
母数 parameter	定数 fixed	確率変数 random variable
データ data	確率変数 random variable	定数 fixed

イズの定理によって

$$p(H_i|X)=\frac{p(X|H_i)p(H_i)}{p(X)} \qquad (1)$$

と表すことができる。ここで（1)式分子の$p(X\mid H_i)$は，仮説H_iが真のときにデータXが得られる確率を表す。これはデータの発生メカニズムに対応するため，この項をデータ分布（data distribution）と呼ぶ。母数の関数と見なす場合には，これを尤度（likelihood）と呼ぶこともある。また，分母の$p(X)$は考えられるすべての仮説を考慮した（周辺化 marginalize した）もとでデータXが得られる確率である。この量は，分子の計算の可能なすべての組み合わせについて，合計が1になる（確率になる）ように基準化する項だと考えることができ，基準化定数と呼ばれる。しばしば，この項は直接計算しなくとも事後確率を求めることが可能である。

つまり，(1)式は次のように理解することができる：

$$[H_i\text{の事後確率}] =\frac{[\text{データ分布}]\times[H_i\text{の事前確率}]}{[\text{基準化定数}]} \qquad (2)$$

ベイズ統計学ではこの枠組みにしたがって，データXを得たもとで，関心のある母数および仮説に関する情報をアップデートしていく。

ここまでは1つの仮説H_iについての推論を見てきたが，実際には2つの仮説が存在し，仮説間でどちらをどれだけデータが支持するのかを調べたい場合が多い。こうしたモデル選択の問題は，ベイズ統計学においてはベイズファクター（Bayes factor, ベイズ因子とも）を用いて評価する。2つの仮説H_iとH_uを比較するベイズファクターは

$$BF_{iu}=\frac{\dfrac{p(H_i|X)}{p(H_u|X)}}{\dfrac{p(H_i)}{p(H_u)}} \qquad (3)$$

によって表される。ここで，(3)式分母の$p(H_i)/p(H_u)$は，データXを観測する前の2つの仮説の事前確率の比を表している。確率の比のことをオッズと呼ぶため，これを事前オッズ（prior odds）と呼ぶ。事前オッズは1のとき2つの仮説の確率は等しく，1より大きいときH_iの事前確率の方が大きく，1より小さいときH_uの事前確率の方が大きい。一方，分子の$p(H_i\mid X)/p(H_u\mid X)$）はデータ$X$を観測して更新された，2つの仮説の事後確率の比を表し，事後オッズ（posterior odds）と呼ぶ。解釈は同様である[2]。(3）式から，ベイズファクターは事後オッズと事前オッズの比として与えられることがわかる。したがって，ベイズファクターは，データによって与えられた，仮説H_uに比して仮説H_iを支持するオッズの変化率を表す（Lavine & Schervish, 1999)。$BF_{iu}=1$のとき，データを得ても事後オッズは事前オッズから変化しなかったことを意味する。$BF_{iu}>1$のとき，データを得ることによって事後オッズは事前オッズよりもH_iを支持するようになったことを意味する。$BF_{iu}<1$のとき，データを得ることによって事後オッズは事前オッズよりもH_uを支持するようになったことを意味する。

ベイズファクターの値を解釈する基準，いわゆる rule

Table 2.
Interpretations of Bayes factor according to Jeffreys (1961) and Kass & Raftery (1995)

(a) Jeffreys (1961)

BF_{iu}	Evidence against H_u
1 to 3.2	Not worth more than a bare mention
3.2 to 10	Substantial
10 to 100	Strong
＞100	Decisive

(b) Kass & Raftery (1995)

BF_{iu}	Evidence against H_u
1 to 3	Not worth more than a bare mention
3 to 20	Positive
20 to 150	Strong
＞150	Very strong

[2] 実際には，後に見るようにベイズ統計による情報仮説の評価の枠組みでは，事前オッズ・事後オッズともに1以下の値をとる。

of thumb（大ざっぱな経験則）としては，Jeffreys（1961）やKass & Raftery（1995）の基準がよく知られている。これをTable 2に示す。しかしながら，これらの基準はあくまで目安にすぎず，絶対的なものでないことには十分留意するべきである。この点をRosnow & Rosenthal（1989）が適切に述べているので，以下やや長くなるが拙訳の上引用する：「ベイズファクターは2つの対立する仮説について，データが支持する程度の比を直接数量化した量である。100倍支持するのであれば十分であり，1.04倍支持するのでは不十分だ，ということには多くの研究者が同意するだろう。しかしながら，文献中には明確なガイドラインはなく，また我々もそれを提供しない。なぜならば，恣意的な決定規則を与えたくはないからだ。p値についてのよく知られた警句を思い出すとよい：『神はp＜.05をp＜.06と等しく，そして同じくらい強く愛してくださる』」（p. 1277）

心理学研究において仮説の評価には長らくp値が第一義的に用いられてきたが，近年，心理学分野においてもベイズファクターを支持する文献が見られるようになっている（e.g., Massaro, Cohen, Campbell, & Rodriguez, 2001; Morey & Rouder, 2011; Rouder & Morey, 2011; Rouder, Morey, Speckman, & Province, 2012）。

情報仮説の評価

本節では，ベイズファクターによって研究者の持っている情報仮説をデータに基づいて評価する方法について述べる。例として，Fonken et al.（2010）による3条件下でのマウスの体重増についてのデータを利用する。ここでは，27匹のマウスを9匹ずつランダムに3つの群に割り当て，4週間後の体重増分を測定している。LD（light/dark）群では，通常の昼間は明るく夜は暗い照明が保たれた。LL（light/light）群では，昼夜を問わず照明は明るく保たれた。DM（dim light at night）群では，昼間は明るく，夜は薄明るい照明が保たれた。このデータは，RパッケージLock5DataにデータセットLightatNightとして収録されている（Lock, 2012）。

この3群における4週間後の体重増が従属変数である。各群の平均パラメータを順に$\mu_{LD}, \mu_{LL}, \mu_{DM}$で表すこととする。ここで，体重増に関して以下の3種類の仮説を考えることができる：

$$H_1: \mu_{LD} < \mu_{DM} < \mu_{LL} \tag{4}$$

$$H_2: \mu_{LD} < \{\mu_{DM}, \mu_{LL}\} \tag{5}$$

$$H_u: \mu_{LD}, \mu_{DM}, \mu_{LL} \tag{6}$$

最初の仮説H_1は，夜が暗い群の平均体重μ_{LD}よりも夜が薄明るい群の平均体重μ_{DM}が重く，さらにそれよりも昼夜関係なく明るい群の平均体重μ_{LL}が重くなるという仮説である。2番目の仮説H_2は，夜が暗い群（LD）よりも夜が暗くない群（DM, LL）の方で平均体重が重くなるという仮説である。この仮説では夜が少しでも明るければ絶対的な明るさの影響は考えないため，μ_{DM}とμ_{LL}の間には不等式制約が置かれていない点に注意する。(4)，(5)式の仮説H_1やH_2のように，母数（ここではμ）について不等式制約（inequality constraint）を入れることで表現される仮説を情報仮説（informative hypothesis）と呼ぶ。研究者が持っている仮説は，このように情報仮説によって表現される場合が多い。

一方，(6)式のH_uのように母数に制約をおかない仮説を無制約仮説（unconstrained hypothesis）と呼ぶ。ベイズ統計による情報仮説の評価の枠組みにおいては，情報仮説を無制約仮説と対比してそのよさを表現する。たとえば情報仮説H_1（4式）のよさを定量化する場合には，データを得たもとでH_1（4式）をH_u（6式）と比較するベイズファクターによって評価する。

ベイズファクターは情報仮説の評価の枠組みと相性がよい。その大きな理由の1つは，情報仮説H_iと無制約仮説H_uとを比較するベイズファクターが，以下の非常にシンプルな式で書けるからである（Klugkist, Laudy, & Hoijtink, 2005）

$$BF_{iu} = \frac{f_i}{c_i} \tag{7}$$

ここで，c_iは無制約仮説H_uの事前分布のうち，情報仮説H_iと整合的な確率密度の割合を表す。これを仮説H_iの複雑さ（complexity）と呼ぶ。一方，f_iは無制約仮説H_uの事後分布のうち，情報仮説H_iと整合的な確率密度の割合を表す。これをモデルの当てはまり（fit）と呼ぶ。上式の導出はHoijtink（2011, 2013）；Klugkist et al.（2005）などに見られる。

この過程を図示して例示するため，μ_1, μ_2という2つだけの母数に関心がある場合を考える。情報仮説として，1群の母平均μ_1が2群の母平均μ_2よりも大きいという仮説H_i: $\mu_1 > \mu_2$を考える。対する無制約仮説は，母平均に制約のないH_u: $\mu_1 > \mu_2$である。Figure 1(a)の左側は，H_uのもとでの母数空間$\{\mu_1, \mu_2\}$上に設定された事前分布を示している。事前の情報はとくに得られていない場合を想定しているため，裾の広い無情報的な事前分布が設定されている。一方，Figure 1(b)の左側は，そのうち情報仮説H_iと整合的な部分を示している。黒く塗られ

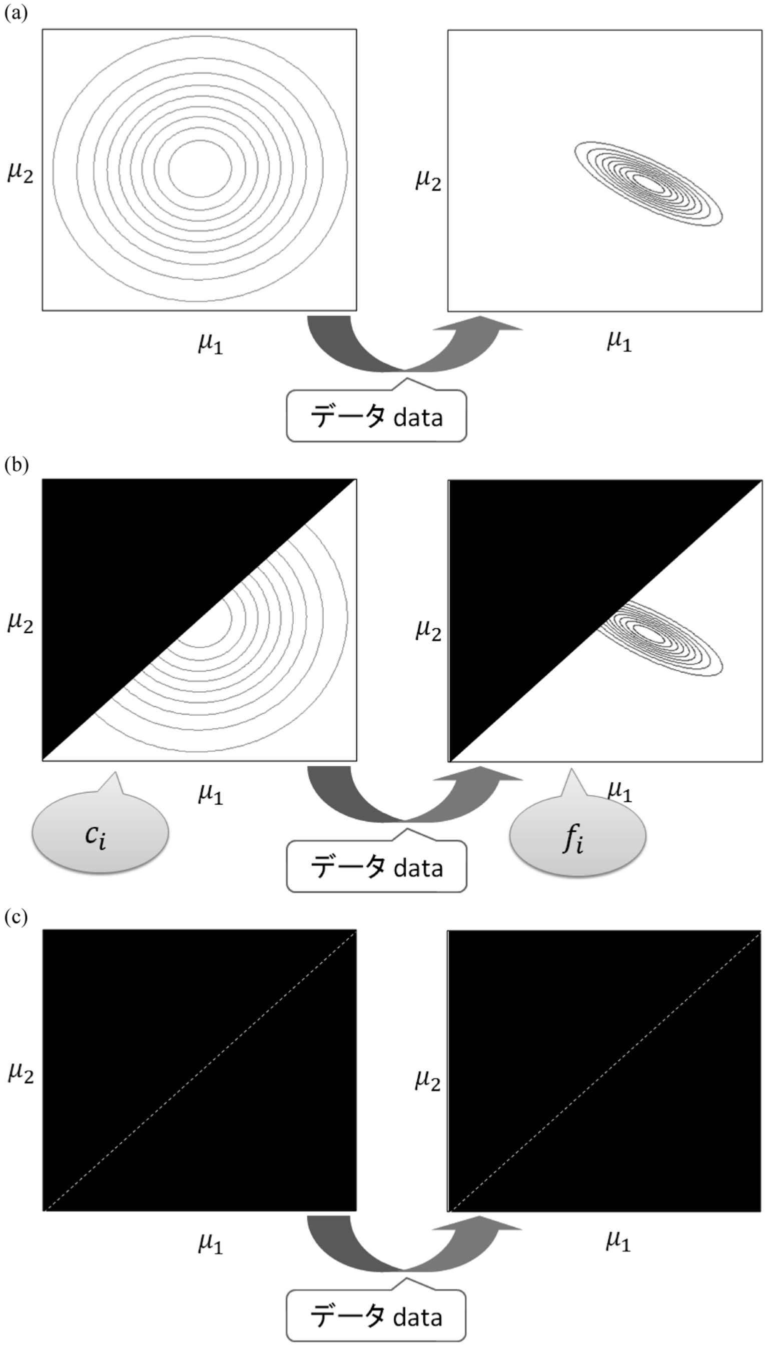

Figure 1.　Prior and posterior distributions under different hypotheses. (a) Prior (left) and posterior (right) distributions under H_u: μ_1, μ_2. (b) Prior (left) and posterior (right) distributions under H_i: $\mu_1 > \mu_2$. (c) Prior (left) and posterior (right) distributions under H_0: $\mu_1 = \mu_2$.

た左上半分は$\mu_1 < \mu_2$となる領域であり，情報仮説H_iと整合的でない。そこで，この情報仮説の複雑さc_iはちょうど$c_i = 1/2$となる。

一方，データを得たあとでのH_uのもとでの事後分布がFigure 1(a)の右側に示されている。データの持つ情報によって，事後分布の散布度は事前分布よりも大幅に

小さくなり，また$\mu_1 > \mu_2$である右下側に多くの確率密度を持つようになったことがわかる。同様に，仮説H_i: $\mu_1 > \mu_2$と整合的な領域をFigure 1(b) 右側に示す。事前分布に比べて事後分布では仮説H_iと整合的な領域に多くの確率密度が含まれるようになり，情報仮説H_iの当てはまりf_iは0.5よりも大きくなっていることがわかる。いまf_i＝0.9であったとすると，H_iをH_uと比較するベイズファクターは，(7)式よりB_{iu}＝0.9/0.5＝1.8と求められる。すなわち，データの情報を得たことによって，H_iをH_uと比較する際の事後オッズは，事前オッズよりも1.8倍に大きくなったことがわかる。

また，ベイズファクターは事後モデル確率（posterior model probability, PMP）へと変換することができる。一般に情報仮説がI個（$H_1, \ldots, H_i \ldots, H_I$）あり，これに加えて無制約仮説$H_u$がある場合において，$i$番目の情報仮説および無制約仮説の事後モデル確率はそれぞれ

$$PMP_i = \frac{BF_{iu}}{1 + \sum_i BF_{iu}} \quad \text{for } i = 1, \ldots, I \tag{8}$$

$$PMP_u = \frac{1}{1 + \sum_i BF_{iu}} \tag{9}$$

により与えられる。ベイズファクターと事後モデル確率は，表現は異なるが相互に変換可能な量である。

実データへの適用例

本節では，前節で導入したベイズ統計による情報仮説の評価の枠組みを実データに適用する。Fonken et al.（2010）のマウスの体重データを用いて，(4), (5)式の情報仮説H_1, H_2を（6)式の無制約仮説H_uと比較する。

情報仮説の評価を実行するためには専用のソフトウェアも開発されている。たとえば，Mulder, Hoijtink, & de Leeuw（2012）はBIEMS，Kuiper, Klugkist, & Hoijtink（2010）はConfirmatoryANOVAという，Fortranで書かれ，GUIを備えたソフトウェアをそれぞれ提供している。これらは，いずれも内部的には，規定の事前分布のもとでマルコフ連鎖モンテカルロ（Markov chain Monte Carlo, MCMC）法による推定を行い，事後分布およびベイズファクターを推定している。

一方，MCMC法を用いたより汎用的に利用できるソフトウェアとしては，WinBUGSやOpenBUGSがある（Lunn, Spiegelhalter, Thomas, & Best, 2009; Lunn, Thomas, Best, & Spiegelhalter, 2000）。van Rossum, van de Schoot, & Hoijtink（2013）はこのBUGSを用いた汎用的な方法を利用して情報仮説の評価を行っている。同様に本稿でもOpenBUGSを利用する。実際に利用したプログラムをAppendixに示す。

まず情報仮説H_1, H_2の複雑さを考える。今回の情報仮説では，複雑さを単純な計算によって求めることができる。(4)式の情報仮説H_1は3つの母数に2つの不等式制約をおくものであり，その場合の数は3!＝6通りある。情報仮説H_1はそのうちの1通りなので，その複雑さはc_1＝1/6となる。次に（5)式の情報仮説H_2を考えると，これは3つの母数に1つの不等式制約をおくものであり，その場合の数は3通りある。報仮説H_2はそのうちの1通りなので，複雑さはc_2＝1/3となる。

次に事後分布の推定を行う。今回のような単純な問題の場合には，事前分布を適切に選択することで事後分布に関する推論を解析的に行うことも可能であるが，汎用性のためにMCMCを用いた方法を利用する。ここでは無情報的な事前分布の設定を利用した。すなわち，各群の平均母数μ_1, μ_2, μ_3にはそれぞれ平均0，分散1000の正規分布を設定した。また，各データ分布の分散には，*IG*（0.01, 0.01）の逆ガンマ分布を設定した。これらはBUGSを用いた事前情報がない場合の無情報的な設定として標準的なものである（Ntzoufras, 2009)。

AppendixのBUGSプログラムを用い，1,000回のburn-inをとったのちの1,000,000回分のMCMC標本を推定に利用した。これだけの多数回のMCMC計算を行っても，所用CPU時間は手元のデスクトップPC（Intel Core i7, 3.4 GHz）で3秒ほどであり，モデルが単純でかつデータセットが大きくないこともあって計算時間は速いことがわかる。また，f_1とf_2の最初の3,000回分のMCMCトレースをFigure 2に示す。この図において，f_1が1の値をとった回のMCMC標本では情報仮説H_1が成立しており，0の値をとった回では成立していないことを意味する（f_2とH_2の組についても同様である)。

情報仮説H_1の当てはまりf_1の推定値は0.9277，情報仮説H_{i2}の当てはまりf_2の推定値は0.9349となった。したがって，ベイズファクターはそれぞれ

$$BF_{1u} = \frac{f_1}{c_1} = \frac{0.9277}{0.1667} = 5.57 \tag{10}$$

$$BF_{2u} = \frac{f_2}{c_2} = \frac{0.9349}{0.3333} = 2.80 \tag{11}$$

と求められる。これら結果の要約をTable 3に示す。以上より，情報仮説H_1とH_2の当てはまりにはほぼ差がないが，情報仮説H_1の方が複雑さの大きな仮説であったので，結果としてベイズファクターはH_2よりも約2倍大きいことがわかる。

また，事後モデル確率は（8), (9)式より

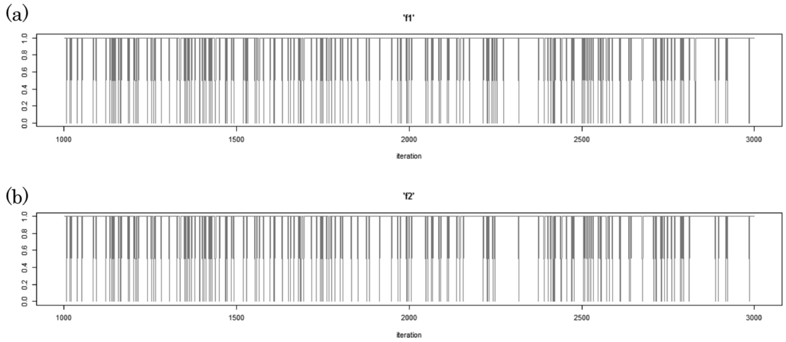

Figure 2.　MCMC traces of fit parameter for（a）H_1 and（b）H_2 for the first 3,000 iterations.

Table 3.
Summary of prior and posterior results for informative and unconstrained hypotheses.

	H_1	H_2	H_u
c_i（complexity）	0.1667	0.3333	—
f_i（fit）	0.9277	0.9349	—
BF_{iu}（Bayes factor）	5.57	2.80	1.00
PMP_i（Posterior model probability）	0.59	0.30	0.11

$$PMP_1 = \frac{5.57}{1+5.57+2.80} = 0.59 \tag{12}$$

$$PMP_2 = \frac{5.57}{1+5.57+2.80} = 0.30 \tag{13}$$

$$PMP_u = \frac{1}{1+5.57+2.80} = 0.11 \tag{14}$$

と得ることができる。

以上の分析結果をまとめると次のようになる。マウスの体重増のデータについて，(4)〜(6)式のH_1, H_2, H_uの3種類の仮説の間でのモデル選択問題を考えた。ベイズファクターの推定結果より，母数に何も制約がない無制約仮説H_uと比して，H_2は2.80倍，そしてH_1は5.57倍に，データによって事後オッズは事前オッズから変化した。また事後モデル確率を見ると，データを得た後でH_1が真である確率はほぼ60％であり，H_2が約30％，H_1が約10％であった。したがって，仮説を1つだけ選ぶならばそれはH_1である。また，そのデータからの支持度合いは，事後確率でみたときにH_2の約2倍，H_uの約6倍であることがわかった。

帰無仮説の扱いと客観ベイズ

前章で述べたとおり，ベイズ統計による情報仮説の評価の枠組みにおいては，研究者の持っている情報仮説H_iを，母数に制約のない無制約仮説H_uと比較して評価する。無制約仮説は関心のある母数に一切の情報仮説をおかない仮説であるため，情報仮説のよさを評価する基準として適切であると考えられる。一方で，伝統的な統計学の枠組みにおいて比較の基準となるのは，無制約仮説H_uではなく，母数に等式制約（equality constraint）の入った帰無仮説H_0である。前節のマウスの体重の例において，帰無仮説は

$$H_0\colon \mu_1 = \mu_2 = \mu_3 \tag{15}$$

で表すことができる。

Figure 1(c）において模式的に示されるように，こうした等式制約による帰無仮説H_0と整合的な確率密度は0である。したがって，前節で述べた方法をH_0の評価にそのまま適用することができない。こうした等式制約のみで表現されるH_0と無制約仮説H_uの比較には，Savage–Dickey密度比（Savagte-dickey density ratio; Dickey, 1971; Verdinelli & Wasserman, 1995）を利用することができる。これは，H_0の等式制約がおかれた点における事前分布と事後分布の確率密度の比である。ただし，3章で述べた情報仮説の評価の枠組みが，無情報的な事前分布を設定すれば事前分布の設定にほとんど依存しないのに対し，定義上Savage–Dickey密度比は事前分布の設定に大きく依存する。たとえば事前分布の分散を10^3とするか10^6とするかといったことが，結果として得られるベイズファクターに非常に大きな影響を与えるのである。したがって，母数空間が［0, 1］に制約される比率の推定のような状況を除いては，帰無仮説H_0を現在のベイズ統計による情報仮説の評価に直接持ち込むことは難しい。

このように等式制約を入れた仮説と入れない仮説，言い換えれば次元数の異なる仮説を比較する際に困難が生じることは古くから知られていた。こうした問題に対応するための枠組みとして，客観ベイズ（objective Bayes）の分野がある。客観ベイズの枠組みでは，データ分析に

際し，必要最小限度の情報はむしろ積極的に活用して，無情報的ではない事前分布を構成しようとする。本稿の範囲からは外れるが，Berger（1985）などを参照してほしい。

最近，Hoijtink（2013）は情報仮説の評価の枠組みにおける客観事前分布の議論を行っているが，その議論はいわゆる客観ベイズの枠組みとは異なる。客観ベイズとはH_0とH_uのように次元数の異なる2つの仮説を比較する際の無情報的ではない事前分布を，客観的に，すなわちデータに基づいて自動的に設定しようという考え方である。それに対しHoijtink（2013）の主張は，ある種の情報仮説（彼の言うところの同等集合equivalent setのうちの1つであるもの）については，事前分布を無情報的に設定することによって，ベイズファクターが本質的にデータの持つ情報だけから決められるということである。この議論は有用であるが，（1）すべての情報仮説に適用できるわけではないこと（同等集合を持つ情報仮説のみである）（2）等式制約によって表現されるH_0のような仮説に適用できるものではないこと に注意が必要である。

おわりに

本稿ではベイズ統計学による情報仮説の評価の枠組みを概説し，実データの分析を行うとともに最近の話題について述べた。統計学における仮説とは母数に関する仮説である。古典的な検定における帰無仮説・対立仮説も，情報仮説の評価で用いる情報仮説・無制約仮説も，母数についての仮説であることには変わりがない。ベイズ統計学は，頻度論と異なり，母数を確率変数として扱う。したがって，ベイズ統計学の立場からの分析を行うことにより，研究者の持っている情報仮説を確率的に評価することができる。

本稿では，用いたモデルは最も単純な，各群がそれぞれ平均母数が異なり分散母数の等しい独立な正規分布にしたがうというものであった。しかし，ベイズ統計による情報仮説の評価の枠組みは，これに限らず，広く一般のモデルで実行できるものである。分散が群間で異なってもよく，群間の共分散を考えてもよく，また階層モデルや混合モデルを設定してもよい。こうした様々なモデルは一般化線形混合モデルの枠組みで記述でき，本稿で行ったようにBUGSによって推定しベイズファクターを導出することができる（Gelman, 2007; Gelman et al., 2013)。このように，ベイズ統計による情報仮説の評価はいわゆる分散分析と比べ，非常に柔軟な枠組みである。

分散分析に相当する問題に情報仮説の評価を適用した事例としては，van Rossum et al.（2013）やvan de Schoot et al.（2011）がある。とくに，van de Schoot et al.（2011）では，2つ以上の要因を考え古典的には交互作用によって議論されるような複雑な設定においても，情報仮説の評価の枠組みが有用であることを発達の実データ分析によって示している。

研究者の持っている仮説は，情報仮説として適切に表現場合が多い。ベイズ統計による情報仮説の評価を利用することで，背理法的に棄却するために設定される帰無仮説をわざわざ考えたり，F検定と多重比較との結果の乖離に悩まされたりすることなく，研究者の仮説のよさを直接定量的に比較・評価することができる。この枠組みが分散分析にいますぐとって代わることはないかもしれない。しかし，少なくとも広く普及した分散分析の欠点を補う相補的な分析として，ベイズ統計による情報仮説の評価は，有用で，広く利用されてよい枠組みだと考える。

Appendix A

データ分析例で利用したOpenBUGSプログラム

f1の事後平均が情報仮説H_1の複雑さf_1の点推定値を，f2の事後平均がH_2の複雑さf_2の点推定値をそれぞれ表す。yが体重増分のデータであり，d1, d2, d3はそれぞれLD, DM LL群への所属を表すダミー変数である。

```
MODEL{
# Model and prior
for（i in 1:27）{
        mu[i] <- mu1*d1[i] + mu2*d2[i] + mu3*d3[i]
        y[i] ～ dnorm（mu[i],invsig2）
}
# Hyperprior
mu1~dnorm（0.0,0.001）
mu2~dnorm（0.0,0.001）
mu3~dnorm（0.0,0.001）
invsig2 ~ dgamma（0.01,0.01）
# Calculation of fit
f1a <- step(mu2-mu1)
f1b <- step(mu3-mu2)
f1 <- f1a*f1b
f2a <- step(mu2-mu1)
f2b <- step(mu3-mu1)
f2 <- f2a*f2b
}
```

引 用 文 献

Berger, J. O.（1985）. *Statistical decision theory and Bayesian analysis*. New York: Springer.

Cumming, G.（2012）. *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. New York: Routledge.

Dickey, J. M.（1971）. The weighted likelihood ratio, linear hypotheses on normal location parameters. *Annals of Mathematical Statistics*, **42**, 204–223.

Fonken, L. K., Workman, J. L., Walton, J. C., Weil, Z. M., Morris, J. S., Haim, A., & Nelson, R. J.（2010）. Light at night increases body mass by shifting the time of food intake. *Proceedings of the National Academy of Sciences of the United States of America*, **107**, 18664–18669.

Gelman, A.（2007）. *Data analysis using regression and multilevel/hierarchical models*: New York: Cambridge University Press.

Gelman, A., Carlin, B. P., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B.（2013）. *Bayesian data analysis*（3rd ed.）. Boca Raton, FL: Chapman & Hall/CRC.

Hoijtink, H.（2011）. *Informative hypotheses: Theory and practice for behavioral and social scientists*. Boca Raton, FL: Chapman and Hall/CRC.

Hoijtink, H.（2013）. Objective Bayes factors for inequality constrained hypotheses. *International Statistical Review*, **81**, 207–229.

Hoijtink, H., Klugkist, I., & Boelen, P.（2008）. *Bayesian Evaluation of Informative Hypotheses*. New York: Springer.

Howell, D. C.（2009）. *Statistical methods for psychology*（7th ed.）. Belmont, CA: Cengage Learning.

Jeffreys, H.（1961）. *Theory of probability*（3rd ed.）. Oxford, UK: Oxford University Press.

狩野　裕（2002）．構造方程式モデリングは，因子分析，分散分析，パス解析のすべてにとって代わるのか？　行動計量学，**29**, 138–159.
（Kano, Y.）

Kass, R. E., & Raftery, A. E.（1995）. Bayes factors. *Journal of the American Statistical Association*, **90**, 773–795.

Kline, R. B.（2004）. *Beyond significance testing*. Washington, DC: American Psychological Association.

Klugkist, I., Laudy, O., & Hoijtink, H.（2005）. Inequality constrained analysis of variance: A Bayesian approach. *Psychological Methods*, **10**, 477–493.

Kuiper, R. M., Klugkist, I., & Hoijtink, H.（2010）. A fortran 90 program for confirmatory analysis of variance. *Journal of Statistical Software*, **34**, 1–30.

Lavine, M., & Schervish, M. J.（1999）. Bayes factors: What they are and what they are not. *American Statistician*, **53**, 119–122.

Lock, R.（2012）. Lock5Data: R package version 2.6. Retrieved May 11, 2013, from http://CRAN.R-project.org/package＝Lock5Data

Lunn, D., Spiegelhalter, D., Thomas, A., & Best, N.（2009）. The BUGS project: evolution, critique and further directions. *Statistics in Medicine*, **28**, 3049–3067.

Lunn, D. J., Thomas, A., Best, N., & Spiegelhalter, D.（2000）. WinBUGS—a Bayesian modelling framework: concepts, structure, and extensibility. *Statistics and Computing*, **10**, 325–357.

Massaro, D. W., Cohen, M. M., Campbell, C. S., & Rodriguez, T.（2001）. Bayes factor of model selection validates FLMP. *Psychonomic Bulletin & Review*, **8**, 1–17.

Morey, R. D., & Rouder, J. N.（2011）. Bayes Factor approaches for testing interval null hypotheses. *Psychological Methods*, **16**, 406–419.

Mulder, J., Hoijtink, H., & de Leeuw, C.（2012）. BIEMS: A Fortran 90 program for calculating Bayes factors for inequality and equality constrained models. *Journal of Statistical Software*, **46**, 1–39.

Ntzoufras, I.（2009）. *Bayesian Modeling Using WinBUGS*. Hoboken, NJ: Wiley.

大久保街亜，岡田謙介（2012）．伝えるための心理統計：検定力・効果量・信頼区間　勁草書房
（Okubo, M. & Okuda, K.）

Rosnow, R. L., & Rosenthal, R.（1989）. Statistical procedures and the justification of knowledge in psychological science. *American Psychologist*, **44**, 1276–1284.

Rouder, J. N., & Morey, R. D.（2011）. A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review*, **18**, 682–689.

Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M.（2012）. Default Bayes factors for ANOVA designs. *Journal of Mathematical Psychology*, **56**, 356–374.

繁桝算男（1985）．ベイズ統計入門　東京大学出版会
（Shigemasu, K.）

van de Schoot, R., Hoijtink, H., Mulder, J., Van Aken, M. A., Orobio de Castro, B., Meeus, W., & Romeijn, J.-W.（2011）. Evaluating expectations about negative emotional states of aggressive boys using Bayesian model selection. *Developmental psychology*, **47**, 203.

van Rossum, M., van de Schoot, R., & Hoijtink, H.（2013）. “Is the hypothesis correct” or “is it not”: Bayesian evaluation of one informative hypothesis for ANOVA. *Methodology: European Journal of Research Methods for the Behavioral and Social Sciences*, **9**, 13.

Verdinelli, I., & Wasserman, L.（1995）. Computing Bayes factors using a generalization of the Savage-Dickey density ratio. *Journal of the American Statistical Association*, **90**, 614–618.