2013年度修士論文

分割表データの様々な条件付き検定のp値（有意確率）の推定について

『グレブナー道場』第4章「マルコフ基底と実験計画法」の記述、特にマルコフ連鎖モンテカルロ法の問題点

早稲田大学大学院基幹理工学研究科

数学応用数理専攻

早稲田 太郎.

指導教員　楫元

Here is Copenhagen!
You and I are the largest!

はじめに

本論文は、実用上非常に重要な問題である「分割表データの様々な条件付き検定のｐ値（有意確率）の推定」（以下、主要問題と呼ぶことにする）について、『グレブナー道場』第４章「マルコフ基底と実験計画法」の記述を（問題のある記述だけでなく、あまり問題でない記述、問題ない記述も含めて（なのでそこは読み飛ばして頂いてもかまわない（読み返すと余計な話が所々挿入されているように感じる））適宜訂正補足解説することを通して、第４章（ないしは、この問題を扱っている論文などの多くの議論）におけるｐ値の推定法や議論の大きな問題点、特にマルコフ連鎖モンテカルロ法（以下、MCMC法）の問題点を浮かび上がらせること、および、それを通して主要問題に取り組むに当たって取り組むべき問題のいくつかを提案することを目的とした論文である。

以下、ページ数や定理番号、例番号などは『グレブナー道場』の初版第２刷のものである。また、読者は第４章などは既読であるとみなして記述する。

第４章の概要

第４章において、以下のような問題が扱われている。

主要問題（分割表の条件付き推(検)定）

・分割表の形で（ある）確率分布の観測値が表せるもの。（確率変数は分割表の各 cell）

・母数θを興味のある母数λと興味の無い母数（局外母数）ψに分解するパラメータ変換θ←→(λ,ψ)（通常１：１）を考えたときの

$$H0：\lambda=(0,0,0,...,0) \quad （帰無仮説）$$

$$H1：\lambda\neq(0,0,0,...,0)$$

の形の検定のｐ値の推定。（よって対立仮説が飽和モデルなのでこの検定は適合度検定）

ただし、パラメータ変換によって指数型分布族（この変換はこの問題を考えるときだけでなく、自然に使われている）として表示したものが「（帰無仮説 H0 の下での局外母数に関する）十分統計量」の固定がｎ元分割表の(n-1)次元周辺度数（以下、小計）の固定で

表されるもの。

4.1 節においてこの問題の具体例やいくつかの p 値の推定法について述べられているが、この p 値の推定に関する議論における問題点、及びそれが具体的に（特にＭＣＭＣ法において）どのように問題かを記述することが本論文の主目的である。

ちなみに 4.2 節は、ＭＣＭＣ法を用いる際に既約なマルコフ連鎖を構成する必要があるが、そのためにはこのマルコフ基底を構成すればよいこと、また、そのマルコフ基底とであることと同値な条件（定理 4.2.8）やそれを利用したマルコフ基底の求め方がグレブナー基底の理論を用いて示されることなどマルコフ基底についての話が述べられている。

4.3 節は実験計画法についてで、これも上記「主要問題」にあてはまるような分割表データ（例えば、2 水準 4 因子完全実施要因計画のデータは 2×2×2×2 分割表である）に関する検定であるため、4.1 節（4.2 節）とまったく同様の議論ができるということ（よって、4.1 節の議論における問題はこの節も射程に入れている）をいう。例 4.3.2 において、当然ながら各小計も総和も固定されていないことに注意。

4.4 節は研究課題について述べられている。

これらの部分においては、（誤記や説明不足の類いではない）本質的な問題点は特には無いかと思う。アルゴリズム 4.2.4 について正当性や停止性など私は証明しないで済ませており、アルゴリズム 4.2.5 は本当に大丈夫なのかよくわかっていないなどきちんと全て読んでいるわけではないのでわからないが。

主要問題の具体例及び重要性

4.1 節の議論の問題点を指摘する前にまず主要問題の重要さについて述べておきたい。

主要問題の具体例として 4.1節 では、次のような例が述べられている。

2×2 分割表		確率分布	興味の対象となる帰無モデル
例 4.1.5	行和固定	各行が独立な二項分布	2つの母比率が主題
例 4.1.6	総和固定	全体が一つの多項分布	行と列が独立
例 4.1.7	固定せず	各セルが独立なポアソン分布	2因子交互作用がない

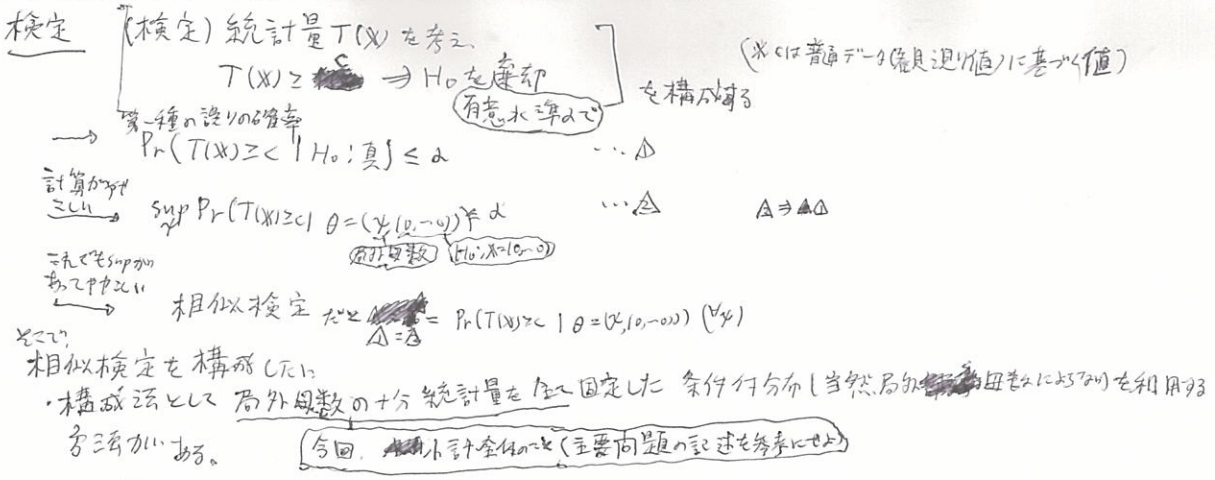
又、例4.1.5においても分割表ので例4.1.6と同様な例を出している。詳しくは4.1節に記載するとして、主要問題として細かな条件が書かれているが、上記の例を見ての通り、分割表で統計モデルを考えれば主要問題の条件も満たす。つまりかなりよくある、普遍的な問題であり、医学疫学などたいていの にかりても頻出するような重要な問題なのである。

もちろん、重要な問題であるということはよりよい「解答」をする意義が大きいことも意味する。

4.1 節の議論の問題点について

まずは「4.1.3 相似検定」での記述を見る。

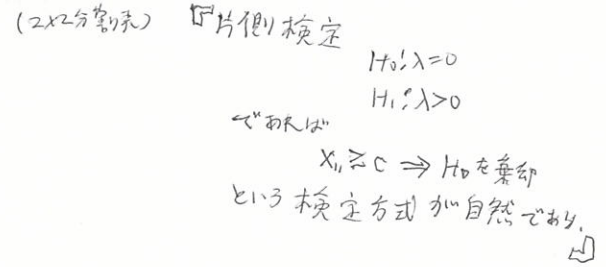
まず、検定について以下のようなことを述べている。



そこで、その具体例として Fisher の正確検定についての記述がはじまる。

正確検定がより具体的にどういったものであるかについてはここでは述べず適宜他の文献で補ってもらうこととする。

片側検定について、



とあるが、この記述が正しいかどうかわからない（ただし、この記述がたとえ誤りであっても大して問題ではない（のでまじめに調べたりもしない））が、通常？の片側検定について述べておく。

（検定統計量の、確率密度の値が正である定義域が 1 次元のときの）片側検定は、（通常は）統計量 T(X)（の値）を確率変数とする確率分布が 

のように単調増加→単調減少の形になっていることを前提として、観測値 T(Xo)が起きる確率より小さい T(X)たちのうち T(Xo)に対して頂点と反対側の確率の総計を p 値とする。

両側検定の場合は片側検定のような頂点との位置関係は関係なく集めた確率の総計を p 値とする。

検定統計量の定義域が2次元以上の場合は、両側検定に対応するものはあるが、片側検定に対応するものは一般には構成できない。

ここで、最初の大きな問題のある記述が出る。

この検定のp値は

$$p\text{値} = P_r(X_{11} \geq x_{11} \mid H_0) = \sum_{x=x_{11}}^{\min(x_{12}, x_{21})} p(x \mid x_{1\cdot}, x_{2\cdot}, x_{\cdot 1}, \lambda = 0)$$

と書ける。

とあるが、このうち左の等号はよいが、右の等号に大きな問題点がある。

まず、直感的にこのような等号は成立するとは思えない。

ついで、これは一種の「近似」ではないかという見方もありうるが、近似とみなすにはやっかいな点があり、ちゃんと右辺と左辺にどういった関係があるか調べてから誤った扱いをしないように使う必要がありそうだ。

さらには、"≠"であり、各値に対するp値も近似されてはいないが、右辺と左辺では別の検定をしているだけでより簡単な検定（相似検定）に持ち込んだだけではないか、とも読めなくもない。しかし、後述するようにこれも正しくなく、右辺により計算された擬似p値は通常p値が持つべき性質（後述）を持っておらず、何らかの検定のp値たり得ない。

そこで、本稿では"＝"が成立しないことを示し、そうだとしたら（"≠"なら）ではこの右辺と左辺はどういう関係なのか、について調べることとする。

さて、上記の記述の次に記述されている例4.1.11において、実際に小計が固定されていない確率分布においてFisherの正確検定（i.e.固定されている場合のみを考えて計算する）を適用し、「（擬似）p値」を計算しているが、この計算は、『この計算により求められた「（擬似）p値」の値から「Hoは棄却できない」という結論を出すこと』（つまり、この計算による「（擬似）p値」は（本来あるべき）p値より（個別にはともかく全体として（この表現の意味は次の「p値について」で説明する））常に大きく出るという性質（これがもしかしたらありえそうだ、ということも後述する）を持つとすると、とりあえず「帰無仮説の本来のp値での検定」より第1種の誤りを大きく見積もっているので、犯さなくていい第1種の誤りを犯さなくてすむ。ただし第2種の誤りを気にしないとして）はともかく、p値として扱うことは問題がある。そのことを述べる前にp値について少し解説をする。

P 値について

・各確率変数の p 値は検定（検定統計量）によって変わる。（恣意的に検定統計量を選べばまったく違った p 値になる）よって、それ単独ではあまり意味を持たない。恣意的に検定を選べばどうとでも操作できてしまう。（ただし、いい性質を持つ検定統計量については、でたらめに検定統計量を選んだ場合に比べて各検定の p 値はある程度近くなる傾向があることに注意したい）

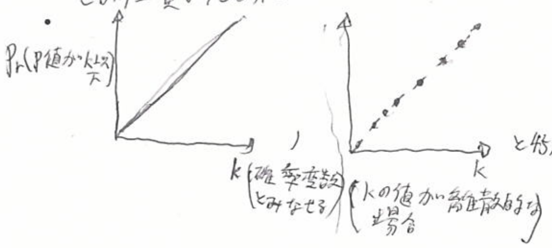
・（一つの一貫した検定における）p 値の重要な性質として、

まず pの値を（任意に）固定 p_{fix} すると。Pr（p値が p_{fix} 以下）$= p_{fix}$ が成立する。

というものがある。

・この性質は応用上ある程度重要で、事前に検定を決めておいてそれによって判断をするということを（一貫して実行することにより）事前に認める第一種の誤りの確率をコントロール（そしてそれは第二種の誤りも）することができる。医学や疫学などで第一種の誤りをコントロールできるというのは大きい。しかし、どうやって一貫した検定を実行してもらうことができるのか、たとえば学界などでこの場合はこの検定でないと受け入れない、などで不正や統計の誤用の一部（ごく一部だが）は防げるようになる。（実際は有意な水準が出ないと投稿されないなどどうしても（副作用なしに）一貫させるのは難しいですが）

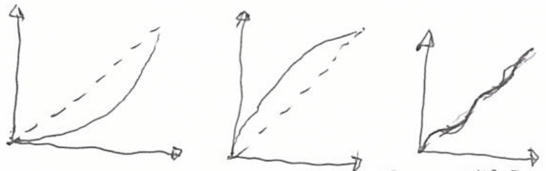
・この性質のために



と対角線にぴったり乗る場合（に限って）kはある検定のp値となりうる。

［問 このような時、kをp値とするような検定が常に（存在す）るか?］

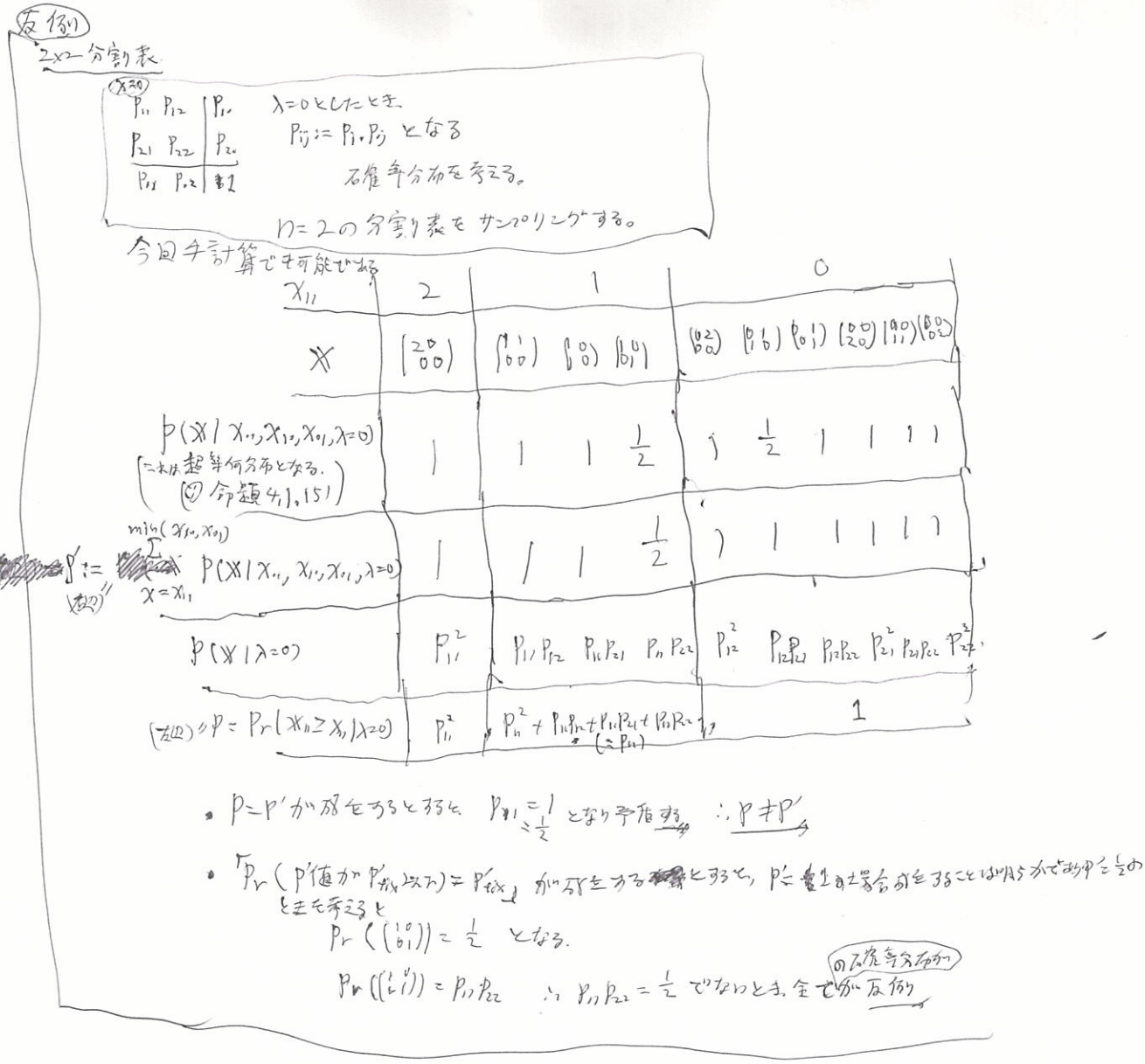
このグラフは



などいろいろな形をとりうる

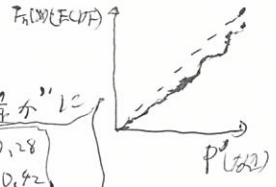
［問 さて、グラフの各場合において、kを「p値」のように扱って検定するとどうなるか?］

さて、先ほどの"="が成立しないこと、「（擬似）p値」はp値ならば持っている上記の重要な性質を持つとは限らない（通常は何らかの検定のp値とはなりえない）ことを反例を構成することによって示す。

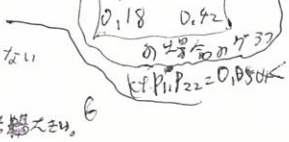


また、「証明」ではないが、

奥村晴彦氏のウェブページ"Fisherの正確検定かカイ二乗検定か"に



が載っており



ほぼ確実に「p値が満たすべき性質」を満たさないだろうことが分かる。

又、注目すべきは本来あるべきp値としての値との差は結構大きい。（数値の差で）

(例) 他の分布の場合のグラフ図はどういう形か？
数学的な考察，統計ソフトの使用など様々なアプローチで調べよ

つづけて 4.1 節の議論を追う。

「4.1.4 I×J 分割表」例 4.1.12（小計が固定されていない確率分布）

p.226「ここまでの議論は、…」ではじまる段落において、「2×2 分割表のときは、長期化分布をもとに、ただちに相似検定（Fisher の正確検定）を考えることができたが、一般の I×J 分割表では、条件付き分布が多次元であるため、検定統計量を定めなければならない。」との記述がある。

まず、2×2 分割表のケースの x_11 も検定統計量であることを指摘しておく。さて、I×J でも正確検定は存在する。これは（ただちに）相似検定であり（実際は、小計が固定されていない確率分布の場合、前述のように擬似 p 値で「正確検定」することになるため、そもそも検定と呼べるかどうか。）、前述の片側検定と両側検定について述べたところで述べたような「両側検定に相当するもの」がそれである。（ちなみに、この「両側検定相当」は p 値の定義が $P_r(T(X)\geq c \mid H_0)$ の 形で表されていないが、$p = P_r(T(X)\geq c \mid H_0)$ の等号が成立するように T(X)を取れるのは明らか(T( ):並び替え関数)）

前掲の引用文を読むと第 4 章の筆者達が正確検定についての概念が 2×2 分割表の場合以外無いと認識しているように読めるが、p.231 に「（Fisher の正確検定を I×J 分割表に拡張した検定問題に対するネットワークアルゴリズム([22])がなどが有名である）」との記述があるので話をややこしくしないために省いたのかも知れない。

いよいよ佳境である。

例 4.1.16 の次の段落において、検定の p 値を求める（主要な）方法（主要問題の解法）を以下のように 3 つに分類して述べている。本文をまとめつつ解説していく。

(a)統計検定量の漸近分布を利用する

(b)p 値を正確に計算する

(c)モンテカルロ法で p 値を推定する

(a)統計検定量の漸近分布を利用する

定理 4.1.17（適合度検定において適合度カイ二乗検定統計量、尤度比検定統計量はいずれも帰無仮説の下で、「興味のある母数λの次元を自由度に持つカイ二乗分布」に漸近的に従う（確率収束する））という定理を元に p 値を推定する方法。

この定理は小計を固定していない（固定した場合に議論を移していない）ことに注意してほしい。（超幾何分布なら超幾何分布、固定していない分布なら固定していない分布として扱える）

具体的には、H0 の下で検定統計量 χ^2,G^2 を上の定理 4.1.17 からカイ二乗分布に近似されるとみて、カイ二乗検定をする。

メリット・カイ二乗検定はＲに実装されているので手軽に使える。

・計算量が少ない

一方、精度が悪い場合がある。・サンプル数（総計）が少ない場合

・分割表が疎（例えば、あるセルの期待値が 10 以下）である場合

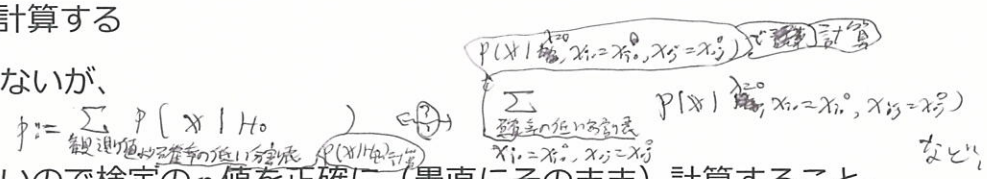
・行和や列和に偏りがある場合

などは収束が遅く、n がある程度あっても誤差が大きい。

また、本文に書いてないが、おそらく、確率分布が超幾何分布のように状態数が少ない場合も収束が遅い。

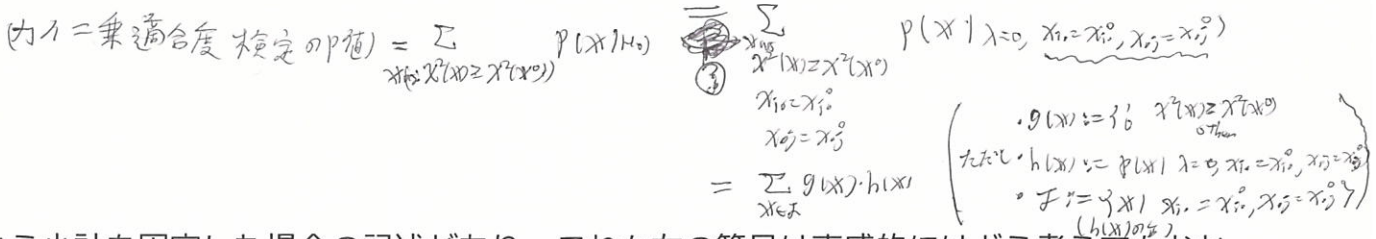
(b)p 値を正確に計算する

本には書いていないが、



など、なんでもよいので検定の p 値を正確に（愚直にそのまま）計算すること。

しかし、本にはまた



という小計を固定した場合の記述があり、これも右の等号は直感的にはどう考えてもおかしいが、実はこの"＝"も本論文で正確検定についてのところで前述したように、1.等号は成り立たず、2.擬似 p 値:= $\sum_{x \in \Omega} g(x) \cdot h(x)$ は p 値としての重要な性質を通常は持たない。このことを反例を挙げることによって示す。2 が成立すれば 1 は明らか（∵「等号が成立する⇒擬似 p 値は p 値としての重要な性質を持つ」の対偶）なので 2 のみを示せばよい。

今回は検定統計量がカイ二乗検定統計量なので、観測値の小計によって χ^2(*)は変わってしまい 2×2 正確検定においてのようには簡単には計算できない。

よって、各観測値 $\begin{array}{cc|c} x_{11} & x_{12} & x_{1\cdot} \\ x_{21} & x_{22} & x_{2\cdot} \\ \hline x_{\cdot 1} & x_{\cdot 2} & x_{\cdot\cdot} \end{array}$ に対して、H_0の下で $p_{ij} = \frac{x_{i\cdot} x_{\cdot j}}{x_{\cdot\cdot}^2}$ という確率分布を考えて

$p' = \sum_{x \in \Omega} g(x) h(x)$ を調べていくことにより示すことにする。

また、観測値の（あるいは最初に仮定する）小計のどれかが 0 だと当てはめ値の中に 0 であるものが表れるため、χ^2(x)=0/0 となり、どう扱えば①よいのか本文では規定されておらず p 値の計算ができない。

また、計算も手計算は厳しくなり、今回私はエクセルを使用して計算した。

まず、総計 n が 1 のときは、分割表は全て①である。

n が 2 のときは、分割表は全部で $_4H_2 = {}_5C_2 = 10$ 通りあり、①でない観測値 $\{(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}), (\begin{smallmatrix}0&1\\1&0\end{smallmatrix})\}$ の組がある。しかし、この各観測値により当てはめ値を計算し、$\chi^2(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}), \chi^2(\begin{smallmatrix}0&1\\1&0\end{smallmatrix})$ を計算すると $\chi^2(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}) = \chi^2(\begin{smallmatrix}0&1\\1&0\end{smallmatrix})$ となり、①でない観測値のどちらによる当てはめ値からも全ての分割表で p'値が 1 になり、これは自明な検定。

n が 3 のときは様子が変わる。分割表は $_4H_3 = {}_6C_3 = 20$ 通り で、①でない観測値 $\{(\begin{smallmatrix}1&0\\1&0\end{smallmatrix}), (\begin{smallmatrix}2&0\\0&1\end{smallmatrix})\}$, $\{(\begin{smallmatrix}1&1\\0&1\end{smallmatrix}), (\begin{smallmatrix}1&2\\0&0\end{smallmatrix})\}$, $\{(\begin{smallmatrix}0&1\\0&1\end{smallmatrix}), (\begin{smallmatrix}0&2\\1&0\end{smallmatrix})\}$, $\{(\begin{smallmatrix}0&1\\1&1\end{smallmatrix}), (\begin{smallmatrix}0&0\\2&1\end{smallmatrix})\}$ の組があり、この 8 つの観測値からそれぞれ（8 通りの）当てはめ値を計算し、それぞれの場合の p'値を計算（さらりと流しているがそれなりの計算が必要である）したところ、8 通りの全ての場合で

$$\left\{ \begin{array}{l} p'_{(\begin{smallmatrix}0&2\\1&0\end{smallmatrix})} = p'_{(\begin{smallmatrix}2&0\\0&1\end{smallmatrix})} = p'_{(\begin{smallmatrix}1&0\\0&2\end{smallmatrix})} = p'_{(\begin{smallmatrix}0&1\\2&0\end{smallmatrix})} = \frac{1}{3} \\ p'_{other} = 1 \end{array} \right\}$$

となった。

ここで、$x^0 = (\begin{smallmatrix}1&0\\1&0\end{smallmatrix})$ を前提としたときを考える。

P_r(p'値が p'_{fix} 以下) $= p'_{fix}$ が成立すると考えると $p'_{fix} = \frac{1}{3}$ のときを考えて

$$P_r(p'値が\tfrac{1}{3}以下) = 3\{(\tfrac{2}{9})^3 + (\tfrac{4}{9})^2 \tfrac{1}{9} + \tfrac{4}{9}(\tfrac{1}{9})^2 + (\tfrac{2}{9})^3\}$$
$$= \frac{4}{27}$$
$$\neq \frac{1}{3} = p'_{fix}$$

となり、P_r(p'値が p'_{fix} 以下) $\neq p'_{fix}$ が示された。

しかし、この議論は、$x^0 = \begin{array}{cc|c} x^0_{11} & x^0_{12} & x^0_{1\cdot} \\ x^0_{21} & x^0_{22} & x^0_{2\cdot} \\ \hline x^0_{\cdot 1} & x^0_{\cdot 2} & x^0_{\cdot\cdot} \end{array}$ に対して、生成確率は $p_{ij} = \frac{x^0_{i\cdot} x^0_{\cdot j}}{x^0_{\cdot\cdot}{}^2}$ となる場合の確率分布のみ考えている。

そこで、もう少し考えてみる。n=3、生成確率は $\begin{array}{cc|c} p_{11} & p_{12} & p_{1\cdot} \\ p_{21} & p_{22} & p_{2\cdot} \\ \hline p_{\cdot 1} & p_{\cdot 2} & \end{array}$ $p_{ij} = p_{i\cdot} p_{\cdot j}$ とする。2 を示すこれまでの議論において P_r(p'値が p'_{fix} 以下)の値 以外は生成確率は関係なく、

p'値は当てはめ値を決める観測値の小計によってのみで決まる。よって、p'は確定している。すると、2 が成り立つような生成確率であるためには、

$$Pr\left(\begin{pmatrix}0&3\\1&0\end{pmatrix}\right)+Pr\left(\begin{pmatrix}2&0\\0&1\end{pmatrix}\right)+Pr\left(\begin{pmatrix}1&0\\0&2\end{pmatrix}\right)+Pr\left(\begin{pmatrix}0&1\\2&0\end{pmatrix}\right)=\frac{1}{3}$$ ②

であればよいとわかる。これは一般的にはいえない。

問：「主観的な妥当な生成確率」なる概念は②=1/3 を満たすだろうか？

問：カイ二乗適合度検定において、①の場合、その小計が０ である部分を取り除くと、検定ができるが、その手法は有効か？

以上のように擬似検定 p'には問題がある。よって、p'を計算している例 4.1.19 も正確な p 値を計算できていない。

以下、「(b)p 値を正確に計算する」では実際に p 値を正確に計算しているものについての話とする。

メリット・常に正確な p 値が計算できる。

デメリット・n の増加により想定する分割表の総数が組み合わせ爆発し、計算量も爆発的に大きくなる。（また、p の計算量と p'の計算量では後者の方がずっと少ないことに注意したい）

参考・各検定統計量の性質を利用した高速な計算アルゴリズムを構成できることがある。

ついに第 4 章の主眼である(c)である。

(c)モンテカルロ法で p 値を推定する

ＭＣＭＣ法の問題点

ＭＣＭＣ法では

p' に $\sum_{x \in A} g(x) \cdot h(x)$ （状態空間 A の元は小計を固定した ある小計と 同じもののみ）を推定するから(a)で述べたような問題がある。

さらに、(b)では p'ではなく（計算量は多いけれども）p を計算することもできたが、ＭＣＭＣ法の場合、小計が固定されていることを前提とした手法であり、とても容易には修正できそうもない。

ＭＣＭＣ法はおおざっぱにはマルコフ連鎖を利用して適切なサンプルを発生させるモンテカルロ法であるが、他の適切なサンプルを発生させる方法を考えないといけない。

メリット・(b)がダメなほど n が大きく、(a)がダメなほど漸近近似精度が悪いときにも使える。

・p'値の不偏推定であり、サンプリング総数Nは自由に設定できる。

・(a)と違い、p'の推定量の確率的な評価ができる。

・(a)と違い、任意の検定統計量に対し適用可能。

欠点・p'を計算しているわけだが、p'とpとのあいだにどのような関係があるか、十分にわかっていない。近似としても優れているとは言えそうではない。

こうして、(a),(b),(c)を眺めると、非常によくある問題である主要問題に対してかなりのケースでp値をまともに推定しようとするとかなり大変であることがわかる。

この論文のここまでの記述ではpのかわりにp'を推定することに対してやや批判的な記述であったが、話をp'に移すことで得られたものも多くあったはずで、p'に移すことによって可能となる技法達を捨て去るのもまた問題であり、それらを有効に使うためにも、以下の問題が重要なのではないか、と思いここに提案する。

問題

・各検定において、pとp'との間にはどのような関係があるか。（関係式があるのか？近似と見れるのか？精度は？n→無限大ではどうなるか？確率分布や検定における違いはあるのか？など）

・p'の性質

また、

問：固定しない場合のＭＣ法（サンプリング法）の開発

もできたらいい。

また、

問：そもそもまともに計算することも難しいことを前提としたとき、p値とどうつきあうのが良いのか。

も重要に思われる。

終わりに（未完）

統計学の歴史の一面が計算可能性、計算量及び精度との戦いの歴史であったことをまた別の形で実感した。

また、統計の最初の入門で息の長い議論をしてでも導入するだけあってか、漸近分布論

の威力はさすがに凄い、と感じた。

いろいろと書きたいことは山ほどあるが、時間が無いのでここで切り上げる。

謝辞（仮）

多くの多くのいろいろなものに感謝！（時間が無い）

参考文献

- JST CREST 日比チーム(編)(2011) グレブナー道場 共立出版
- 奥村晴彦 統計データ解析 http://oku.edu.mie-u.ac.jp/~okumura/stat/ 2014年2月4日閲覧