研究論文

修正 2 次元統計量を用いた非条件付検定の性能について*

松　尾　精　彦

概　要

いくつかの二項確率変数の等確率性検定を有意水準を固定して行う場合を考える．この検定を正確法で行うには，条件付・非条件付の 2 つのアプローチがあり，それらの検出力比較が行われてきた．この検定で気を付けなければならないのが検定統計量分布の離散性である．条件付・非条件付によらず，サンプル数が少ないときやサンプル数が等しい時には，検定サイズが有意水準に遠く及ばず，その結果として低い検出力が懸念される．この問題を解決するため，私は 2 つの精密化のための方法を提案している．1 つは条件付検定の精密化であり (松尾, 2000a)，2 つめは適合度検定統計量の条件付分布から構成される修正統計量を用いた非条件付検定である (松尾, 2000b)．これらの方法を用いることにより，正確検定の問題点とされる，検定サイズが有意水準よりもはるかに小さくなるという問題を軽減することができた．これら 2 つの精密化法は，排他的なものではなく，同時に用いることができる．この論文では，従来のどのような方法を用いても高い検出力が望めない場面でも，2 つの精密化法を同時に用いれば効率的な検定が行えることを数値例を使って実証する．

キーワード: goodness of fit test, equality of binomial proportions.
経済学文献季報分類番号：16-10

1　紹介

ここでは統計の理論家，例えば Weerahandi(1995), の間で利用が推奨されている有意検定ではなく，有意水準を固定した検定問題をかんがえる．彼は「観測された有意水準を報告しさえすればよく，判断は専門家にゆだねればよい」と述べているいる．確かにこの主張が有効な場面も数多くあるだろうが，それでも新医薬品承認のときのように合否の決断を厳格に下さなければならない場面では有意水準を固定するのが妥当である．

ここで取り扱う，いくつかの 2 項分布の等確率性検定を行う際には，正確法よりはむしろ近似法に頼るのが一般的であった．その主な理由は 2 つあり，1 つは計算負荷の問題，もう 1 つは検出力の問題であった．前者は昨今のコンピュータ環境の発展により解消されつつある．また，後者は正確法が検定のルールを遵守しているために，逆に言えば，近似法が検定のルール

1)　この研究は平成 13 年度関西大学学部共同研究によって行った研究の一部である．

を破っているために生じる差であり正当な理由とは言い難い．

私は 2 つの論文で，コンピュータ資源を活用した正確検定法の精密化を提案してきた．これらは検定をルールを厳密に守りつつ検出力を上げるための方法である．松尾 (2000a) では条件付検定における離散性を緩和するために，2 次元統計量を考案し，それを用いた条件付検定の利用を提案した．また, 松尾 (2000b) では, 非条件付検定での局外パラメータ依存を軽減するために, 従来の適合度検定統計量の条件付分布から構成される修正統計量を提案し，修正統計量を用いた非条件付検定の利用を提案した．これら 2 つの正確検定の精密化法は，排他的なものではなく，同時に利用できることが判明した．このことは，これまで条件付・非条件付検定の保守性の原因とされてきた，離散性と局外パラメータへの依存という 2 つの問題を同時に解決できるということだ．この論文の目的は，従来別々に使われてきた 2 つの精密化法を同時に使った検定方式の高い検出力を実証することである．

第 2 節では検定の基礎的な説明を行う．第 3 節では, これまで提案してきた正確検定の精密化法についての説明を行い，その上で従来の 2 つの精密化の方法を同時に用いる新たな検定方式を提案する．第 4 節では，第 3 節で新たに紹介した検定方式のサイズ・検出力を，従来の非条件付検定と比較しその性能を数値的に評価する．

2　モデルと従来の検定法

この論文を通して，$Y_1, Y_2, \cdots, Y_k$ を独立な確率変数とし，各 Y_i は二項分布 $\mathrm{B}(n_i, \pi_i)$ に従うものとする． また $\boldsymbol{y} = (y_1, \cdots, y_k)$ は $\boldsymbol{Y} = (Y_1, \cdots, Y_k)$ の実現値を表わすものとする．なおベクトルは $\boldsymbol{y}, \boldsymbol{Y}$ のようにボールド体で表すことにする．このとき等確率性検定は,

$$\begin{cases} \mathrm{H}_o : \pi_1 = \pi_2 = \cdots = \pi_k = \pi_o \\ \mathrm{H}_A : \pi_i \neq \pi_j, \text{ for some } i \neq j \end{cases},$$

と書き表せる． この検定を，前節で述べた理由により近似法は用いずに，正確分布を用いて検定する．

帰無仮説を仮定するとき，共通のパラメータ π_o には十分統計量 $S = \sum_{i=1}^{k} Y_i$ が存在するため，この統計量の観測値 $s = \sum_{i=1}^{k} y_i$ で条件付けた検定統計量分布を用いるか，条件付けない検定統計量分布を用いるかの選択肢がある．観測変量が連続である場合には条件付検定を行えばよいが，離散の場合には離散性の問題もあり一般論では片付かない．α を検定の水準とし，$T(\boldsymbol{Y})$ を統計量として，条件付・非条件付検定の手順を簡単に説明しよう．なお，詳しい説明は，例えば Mehta and Hilton (1993) あるいは松尾 (2000b) を参照されたい．

まず条件付検定では，観測された十分統計量値 $s = \sum_{i=1}^{n} y_i$ を共有する全ての観測点からなる

集合 Γ_s (条件付参照集合と呼ぶ),

$$\Gamma_s = \{ \boldsymbol{y} \mid 0 \le y_i \le n_i, \ \sum_{i=1}^{k} y_i = s \},$$

を考え, 条件付参照集合 Γ_s 上の条件付分布をもとに棄却域を構成する．この条件付分布は，帰無仮説の下で，未知パラメータに依存しないという利点がある．このことが条件付検定の長所なのだが，条件付参照集合上の検定統計量分布の離散性が顕著な場合には，検定サイズが与えられた水準に遠く及ばなくなり，その結果として検出力が低くなる恐れがある．一方非条件付検定では，標本空間上の検定統計量分布により棄却域を構成する．この分布は未知局外パラメータ π_o に依存するため，最悪のパラメータ値においても検定サイズが水準を超えないようにしなければならない．このことが非条件付検定の保守性の原因となっている．

検定統計量は，適合度検定統計量として最もよく用いられてきた Pearson のカイ自乗統計量,

$$PX(\boldsymbol{y}) = \sum_{i=1}^{k} \frac{(y_i - n_i\hat{\pi})^2}{n_i\hat{\pi}(1-\hat{\pi})},$$

ここで $\hat{\pi} = \sum y_i / \sum n_i$, のみを扱うことにする．他にも一般化線形モデル (McCullagh and Nelder, 1989) の枠組みで利用される *Deviance* や Cressie and Read (1984) がその漸近的な性質から提唱した *Power divergence* があるが，ここでは取り上げない．なぜならば, 条件付検定においてはこれら 3 つの統計量はほとんど同じ振る舞いをすることが Matsuo (1999) で観察されており, ここで提案する精密化法においても同様の振る舞いが十分予想されること, また非条件付検定においては, *Power divergence* と Pearson のカイ自乗統計量 とは良く似た振る舞いをするのに対し, *Deviance* は未知局外パラメータ π_o の変化に対し極めて不安定な振る舞いを見せるため比較対象とならないためである．

3 正確検定の精密化

この節では，従来の正確検定法を精密化する方法を 2 つ紹介し，その上でそれらを同時に用いる検定方式を提案する．

3.1 2次元統計量を用いた条件付検定

この分節では，松尾 (2000a) が提案した，2 次元統計量を用いた条件付検定について説明する．2 次元統計量を考える動機付けとなったのは，サンプル数の設定により検定統計量の条件付分布の離散性がかなり変化するという観察である．サンプル数 n_i が全て等しいときには，条件付参照集合上で検定統計量がタイの値を取りやすくなるため離散性が高くなり，反対にサンプル数が互いに異なるときにはタイの値をとりにくくなるため離散性が軽減される．サンプル数が全て等しい時には検定サイズが水準に遠く及ばず，低い検出力が懸念される．サンプル数の

設定というあまり本質的でないことが，検定の効率を左右するのは不合理である．そこで，タイの値を共有する観測間に自然な順位を付けようというのが2次元統計量の考え方である．

順位付けのために必要なものは，各サンプルの結果と共に観測される何らかの説明変数である．これは必ずしも間隔尺度ある必要はなく，観測順あるいは予想される確率の大きさの順といった順位尺度でもよい．この説明変数の値を x_i として $\sum_{i=1}^{n} x_i Y_i$ を構成し，タイの値をとる観測間に順位付けをするのである．このことは，2次元統計量 $(T(\boldsymbol{Y}), \sum_{i=1}^{n} x_i Y_i)$ を用いて検定を行うことに対応する．

なお，サンプル数が互いに異なるときは条件付参照集合上で検定統計量がタイの値を取り難くなるため，2次元統計量を用いる効果はほとんどないことに注意されたい (松尾, 2000a).

3.2　修正統計量を用いた非条件付検定

この分節では，松尾 (2000b) が提案した修正統計量を用いた非条件付検定について説明する．この方法は，条件付検定と非条件付検定の長所を同時に持つものである．修正統計量は次のように構成される．$T(\boldsymbol{Y})$ をある適合度検定統計量とするとき，

$$T^{*}(\boldsymbol{y}) = \Pr_{H_{\circ}}\{\ T(\boldsymbol{Y}) < T(\boldsymbol{y}) \mid \sum_{i=1}^{k} y_i = s, \pi_{\circ}\ \}$$

により修正統計量を構成するのである．これは，統計量の値をそのまま用いるのではなく，条件付参照集合内の相対的な位置を表すものになっている．このため，非条件付で用いる場合，保守性の軽減が期待される．なお $T^{*}(\boldsymbol{Y})$ は $T(\boldsymbol{Y})$ の大小関係を保存するため，$T^{*}(\boldsymbol{Y})$ を検定統計量とする条件付検定は，$T(\boldsymbol{Y})$ を用いた条件付検定と同じ結果を与えることに注意されたい．このため, $T^{*}(\boldsymbol{Y})$ を用いた非条件付検定は，$T(\boldsymbol{Y})$ を用いた条件付検定よりも常に検出力が高くなる．

3.3　2つの精密化を同時に使う検定方式

前の2つの分節で紹介した正確検定の精密化法をそれぞれ単独で使いた場合，従来の非条件付検定よりも性能が劣る場面がしばしば見うけられた．この現象はサンプル数が大きくなるに従い消滅してゆくものの，無視しがたいものである．そこで，これら2つの方法を同時に使うことを考える．つまり，2次元統計量から修正統計量を構成し，それを用いて非条件付検定を行うというものである．この検定の棄却域は，2次元統計量を用いた条件付検定の棄却域を含み，さらに後者は元々の統計量を用いた条件付検定の棄却域を含む．この関係を考慮にいれるとき，この検定を実際に行う際には，まず元々の統計量を用いた条件付検定を行い，有意でなければ，2次元統計量を用いた条件付検定，さらに有意でなければ，修正2次元統計量を用いた非条件付検定を行えばよい．有意かどうか微妙なときだけ計算量が増えることになるが，これは精密さの代償と考えることができる．

4　サイズ・検出力比較

この節では，前節で紹介した 2 つの精密化を同時に用いる検定方式のサイズおよび検出力を数値例により示す．サンプルサイズの設定は，松尾 (2000a, 2000b) と同様に $k=3$ とし，最も離散性が心配される $n_1=n_2=n_3$ の組み合わせとして $\boldsymbol{n}=(15,15,15)$, $(20,20,20)$, $(25,25,25)$, $\cdots$, $(50,50,50)$ を取り上げ，次に離散性が心配される $n_1=n_2\neq n_3$ の組み合わせとして $\boldsymbol{n}=(15,15,16)$, $(20,20,21)$, $(25,25,26)$, $\cdots$, $(50,50,51)$ を取り上げる．なお，ここでは最も離散性が軽減される $n_1\neq n_2\neq n_3$ の組み合わせは取り上げないことに注意されたい．その理由は，この場合 2 次元統計量を導入する効果がないためである．松尾 (2000b) では $\boldsymbol{n}=(19,20,21)$,$\boldsymbol{n}=(29,30,31)$, $\boldsymbol{n}=(39,40,41)$, $\boldsymbol{n}=(49,50,51)$ を取り上げ，修正統計量を用いた非条件付検定の方が，元の統計量を用いた非条件付検定よりも検出力がほぼ一様に高いことを数値例により観察している．

ここで比較するのは，Pearson のカイ自乗検定統計量から 2 次元統計量を構成し，それをもとに作られる修正統計量 PX^{M2D} を用いた非条件付検定と，Pearson のカイ自乗統計量を用いた非条件付検定である．なおこの論文では，$x_i=i$ とし $\sum_{i=1}^{k} x_i y_i$ の大きい方から棄却域に入れることにした．

4.1　サイズ比較

まず，検定水準を $\alpha=0.05$ として，サイズについての比較を行う．次の図 1 ∼ 8 は $n_1=n_2=n_3$ の場合のサイズ関数を表している．各グラフの実線が PX, 破線が PX^{M2D} を用いた検定サイズ関数を表している．サイズ関数 $\alpha(\pi)$ は，対応する検定の棄却域を R とするとき，

$$\alpha(\pi)=\Pr_{H_0}\{\boldsymbol{y}\in R\}|_{\pi_0=\pi}$$

と表されるものであり，松尾 (2000a) で述べたように，$\pi=0.5$ について対称な関数である．

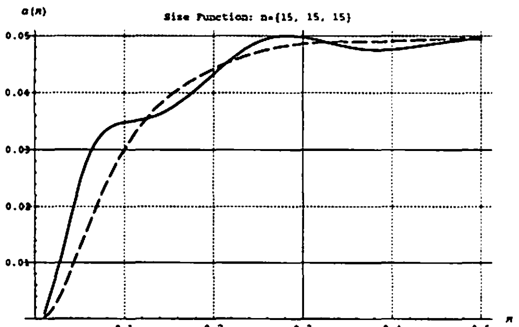

図 1: $\boldsymbol{n}=(15,15,15)$. PX:実線, PX^{M2D}:破線. 以下同様.

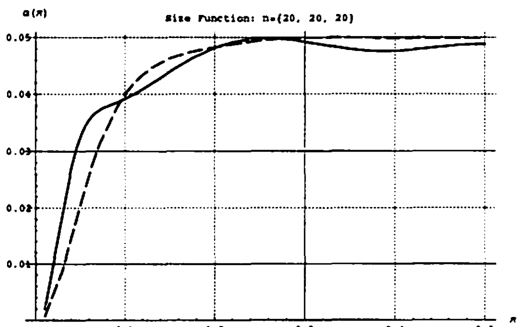

図 2: $\boldsymbol{n}=(20,20,20)$

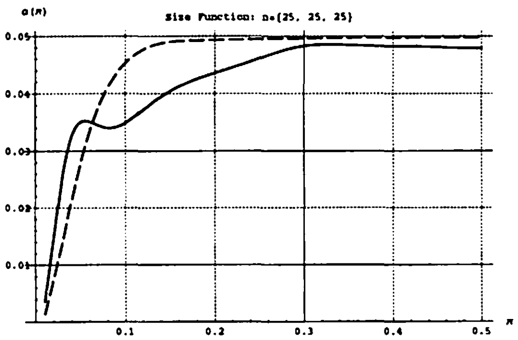

図 3: $\boldsymbol{n} = (25, 25, 25)$

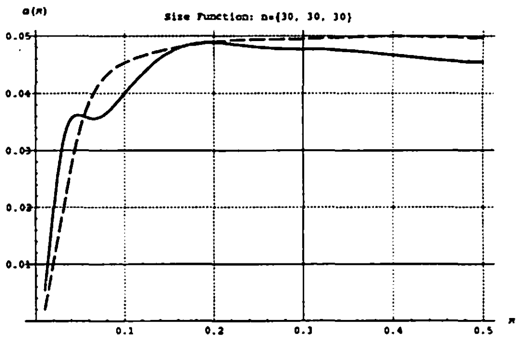

図 4: $\boldsymbol{n} = (30, 30, 30)$

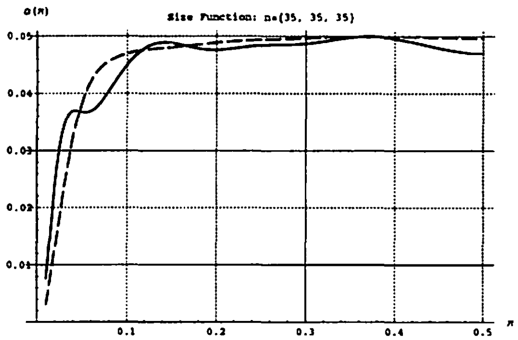

図 5: $\boldsymbol{n} = (35, 35, 35)$

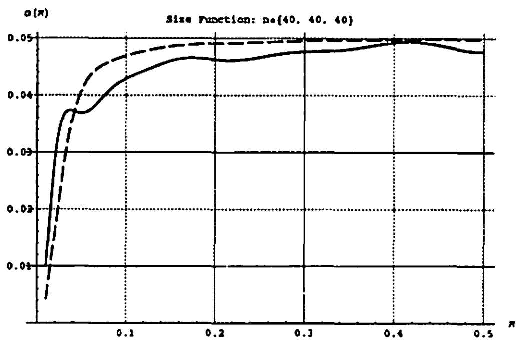

図 6: $\boldsymbol{n} = (40, 40, 40)$

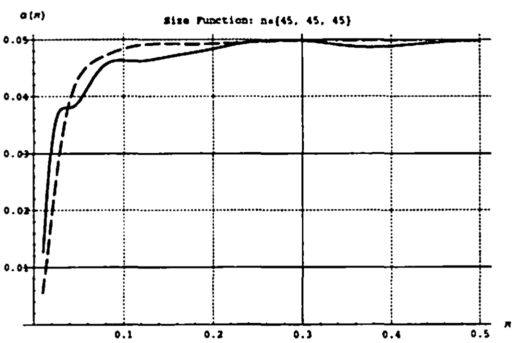

図 7: $\boldsymbol{n} = (45, 45, 45)$

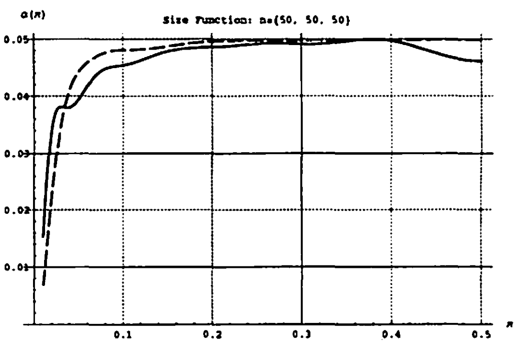

図 8: $\boldsymbol{n} = (50, 50, 50)$

これら 8 つの図を見ても分かるように，PX^{M2D} の非条件付分布はそのねらい通り，未知局外母数 π への依存度が低くなり同時に離散性が軽減されたものになっている．実際 PX^{M2D} を用いた非条件付検定のサイズ関数 (破線) は $0.1 < \pi < 0.9$ の範囲内で，検定水準 $\alpha = 0.05$ にかなり近い値を取っていることに注目されたい．

次に $n_1 = n_2 \neq n_3$ の場合に移ろう．図 9 ～ 16 も，また，PX^{M2D} を用いた非条件付検定のサイズ関数（破線）と PX を用いた非条件付検定のサイズ関数（実線）を比較するためのもので

ある．

この場合，サンプルサイズが大きくなるに従い，$n_1 = n_2 = n_3$ の場合よりも早く，サイズ関数が互いに接近してくる．またサンプルサイズが等しい場合に較べて，PX^{M2D} の優位性が顕著ではなくなる傾向が観察される．このことを言いかえれば，PX はそのままでもかなり良い性能を持っていることになる．このことは *Deviance* の非条件付分布が π に過度に依存することと対照的な特徴と言える（松尾, 2000b).

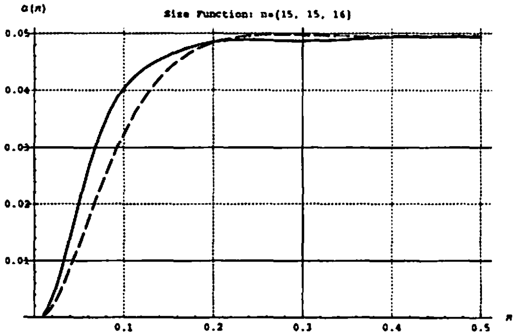

図 9: $\boldsymbol{n} = (15, 15, 16)$

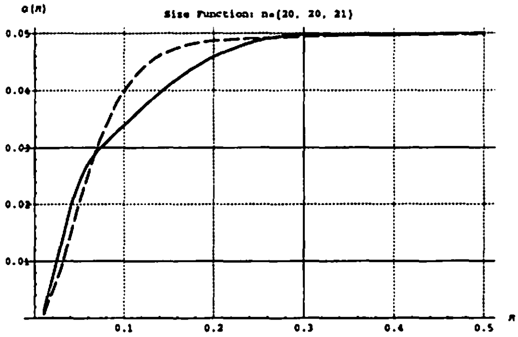

図 10: $\boldsymbol{n} = (20, 20, 21)$

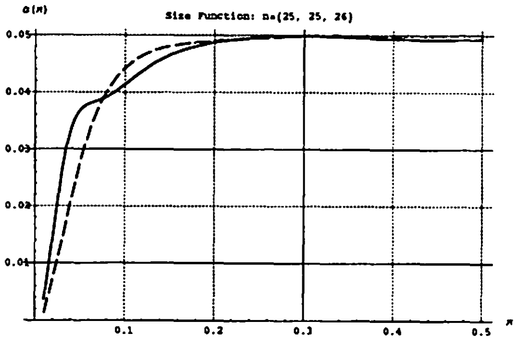

図 11: $\boldsymbol{n} = (25, 25, 26)$

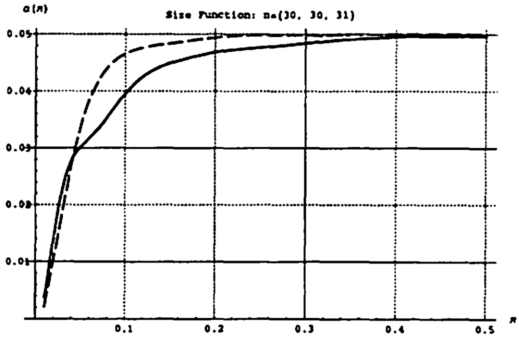

図 12: $\boldsymbol{n} = (30, 30, 31)$

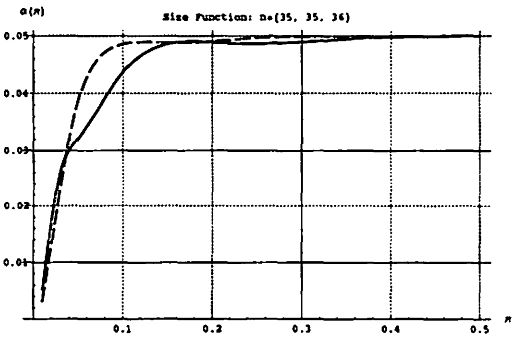

図 13: $\boldsymbol{n} = (35, 35, 36)$

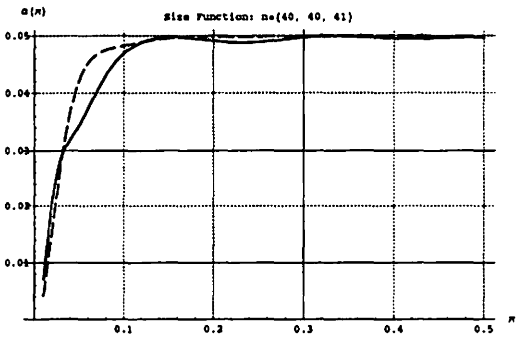

図 14: $\boldsymbol{n} = (40, 40, 41)$

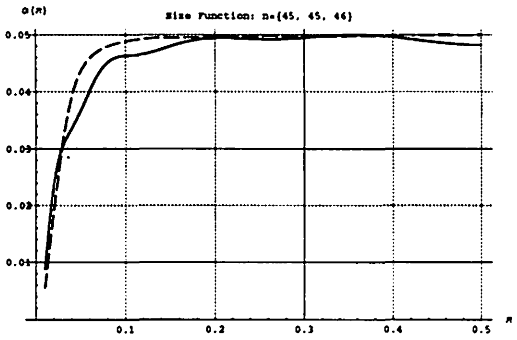

図 15: $\boldsymbol{n} = (45, 45, 46)$

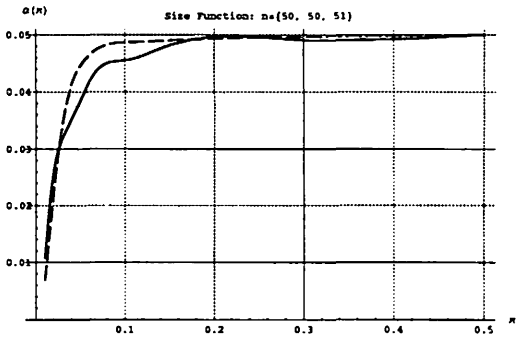

図 16: $\boldsymbol{n} = (50, 50, 51)$

4.2　検出力比較

最後に検出力比較を行おう．サンプルサイズの設定は **4.1** と同様とする．単純対立仮説は次の 156 の組み合わせを考える．これらはすべて $p_1 \le p_2 \le p_3$ を満たしており，2 次元統計量を用いて $\sum_{i=1}^{k} x_i y_i = \sum_{i=1}^{k} i y_i$ の大きいほうから棄却域に入れるという，この論文の計算方法に対応したものである．

(0.1, 0.1, 0.2), (0.1, 0.1, 0.3), (0.1, 0.1, 0.4), (0.1, 0.1, 0.5), (0.1, 0.1, 0.6),
(0.1, 0.1, 0.7), (0.1, 0.1, 0.8), (0.1, 0.1, 0.9), (0.1, 0.2, 0.2), (0.1, 0.2, 0.3),
(0.1, 0.2, 0.4), (0.1, 0.2, 0.5), (0.1, 0.2, 0.6), (0.1, 0.2, 0.7), (0.1, 0.2, 0.8),
(0.1, 0.2, 0.9), (0.1, 0.3, 0.3), (0.1, 0.3, 0.4), (0.1, 0.3, 0.5), (0.1, 0.3, 0.6),
(0.1, 0.3, 0.7), (0.1, 0.3, 0.8), (0.1, 0.3, 0.9), (0.1, 0.4, 0.4), (0.1, 0.4, 0.5),
(0.1, 0.4, 0.6), (0.1, 0.4, 0.7), (0.1, 0.4, 0.8), (0.1, 0.4, 0.9), (0.1, 0.5, 0.5),
(0.1, 0.5, 0.6), (0.1, 0.5, 0.7), (0.1, 0.5, 0.8), (0.1, 0.5, 0.9), (0.1, 0.6, 0.6),
(0.1, 0.6, 0.7), (0.1, 0.6, 0.8), (0.1, 0.6, 0.9), (0.1, 0.7, 0.7), (0.1, 0.7, 0.8),
(0.1, 0.7, 0.9), (0.1, 0.8, 0.8), (0.1, 0.8, 0.9), (0.1, 0.9, 0.9), (0.2, 0.2, 0.3),
(0.2, 0.2, 0.4), (0.2, 0.2, 0.5), (0.2, 0.2, 0.6), (0.2, 0.2, 0.7), (0.2, 0.2, 0.8),
(0.2, 0.2, 0.9), (0.2, 0.3, 0.3), (0.2, 0.3, 0.4), (0.2, 0.3, 0.5), (0.2, 0.3, 0.6),
(0.2, 0.3, 0.7), (0.2, 0.3, 0.8), (0.2, 0.3, 0.9), (0.2, 0.4, 0.4), (0.2, 0.4, 0.5),
(0.2, 0.4, 0.6), (0.2, 0.4, 0.7), (0.2, 0.4, 0.8), (0.2, 0.4, 0.9), (0.2, 0.5, 0.5),
(0.2, 0.5, 0.6), (0.2, 0.5, 0.7), (0.2, 0.5, 0.8), (0.2, 0.5, 0.9), (0.2, 0.6, 0.6),
(0.2, 0.6, 0.7), (0.2, 0.6, 0.8), (0.2, 0.6, 0.9), (0.2, 0.7, 0.7), (0.2, 0.7, 0.8),
(0.2, 0.7, 0.9), (0.2, 0.8, 0.8), (0.2, 0.8, 0.9), (0.2, 0.9, 0.9), (0.3, 0.3, 0.4),
(0.3, 0.3, 0.5), (0.3, 0.3, 0.6), (0.3, 0.3, 0.7), (0.3, 0.3, 0.8), (0.3, 0.3, 0.9),
(0.3, 0.4, 0.4), (0.3, 0.4, 0.5), (0.3, 0.4, 0.6), (0.3, 0.4, 0.7), (0.3, 0.4, 0.8),
(0.3, 0.4, 0.9), (0.3, 0.5, 0.5), (0.3, 0.5, 0.6), (0.3, 0.5, 0.7), (0.3, 0.5, 0.8),
(0.3, 0.5, 0.9), (0.3, 0.6, 0.6), (0.3, 0.6, 0.7), (0.3, 0.6, 0.8), (0.3, 0.6, 0.9),
(0.3, 0.7, 0.7), (0.3, 0.7, 0.8), (0.3, 0.7, 0.9), (0.3, 0.8, 0.8), (0.3, 0.8, 0.9),
(0.3, 0.9, 0.9), (0.4, 0.4, 0.5), (0.4, 0.4, 0.6), (0.4, 0.4, 0.7), (0.4, 0.4, 0.8),
(0.4, 0.4, 0.9), (0.4, 0.5, 0.5), (0.4, 0.5, 0.6), (0.4, 0.5, 0.7), (0.4, 0.5, 0.8),
(0.4, 0.5, 0.9), (0.4, 0.6, 0.6), (0.4, 0.6, 0.7), (0.4, 0.6, 0.8), (0.4, 0.6, 0.9),
(0.4, 0.7, 0.7), (0.4, 0.7, 0.8), (0.4, 0.7, 0.9), (0.4, 0.8, 0.8), (0.4, 0.8, 0.9),
(0.4, 0.9, 0.9), (0.5, 0.5, 0.6), (0.5, 0.5, 0.7), (0.5, 0.5, 0.8), (0.5, 0.5, 0.9),

(0.5, 0.6, 0.6), (0.5, 0.6, 0.7), (0.5, 0.6, 0.8), (0.5, 0.6, 0.9), (0.5, 0.7, 0.7),
(0.5, 0.7, 0.8), (0.5, 0.7, 0.9), (0.5, 0.8, 0.8), (0.5, 0.8, 0.9), (0.5, 0.9, 0.9),
(0.6, 0.6, 0.7), (0.6, 0.6, 0.8), (0.6, 0.6, 0.9), (0.6, 0.7, 0.7), (0.6, 0.7, 0.8),
(0.6, 0.7, 0.9), (0.6, 0.8, 0.8), (0.6, 0.8, 0.9), (0.6, 0.9, 0.9), (0.7, 0.7, 0.8),
(0.7, 0.7, 0.9), (0.7, 0.8, 0.8), (0.7, 0.8, 0.9), (0.7, 0.9, 0.9), (0.8, 0.8, 0.9),
(0.8, 0.9, 0.9).

まず $n_1 = n_2 = n_3$ の場合を考えよう．計算結果は次の図 17 ∼ 24 のようになる．各図の左側のグラフは，PX を用いた非条件付の検出力を x 軸に，PX^{M2D} を用いた非条件付の検出力を y 軸にとったものであり，右側のグラフは PX^{M2D} の検出力から PX の検出力を引いたものを上の単純対立仮説の順に並べたものである．

サンプルサイズが等しい場合には，PX^{M2D} 導入の効果が顕著に現われ，例外的なものを除きほとんど全ての単純対立仮説において検出力が上がっている．この傾向は，この論文で取り上げられた設定以外でも一貫して観察されるものである．

続いて $n_1 = n_2 \neq n_3$ の場合に移ろう．$n_1 = n_2 = n_3$ の場合と同様の計算を行った結果が次の図 25 ∼ 32 である．この場合，PX^{M2D} 導入の効果は $n_1 = n_2 = n_3$ の時ほど顕著ではないものの，全体的な傾向としてはより高い検出力を導いていることが分かる．

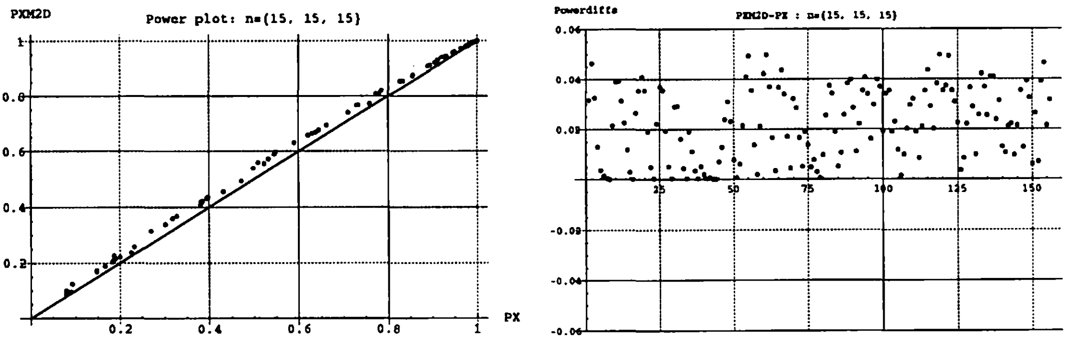

図 17: $\boldsymbol{n} = (15, 15, 15)$ のとき．左：検出力プロット (x 軸:PX, y 軸:PX^{M2D}), 右：（ PX^{M2D} の検出力）－（ PX の検出力）．以下同様.

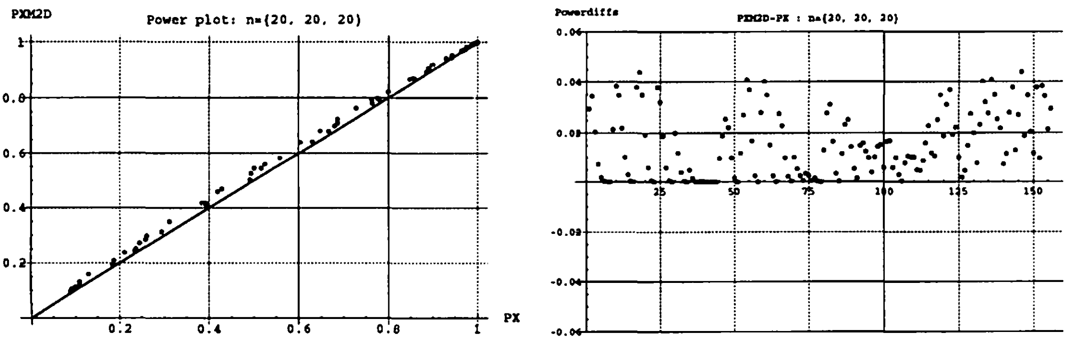

図 18: $\boldsymbol{n} = (20, 20, 20)$ のとき

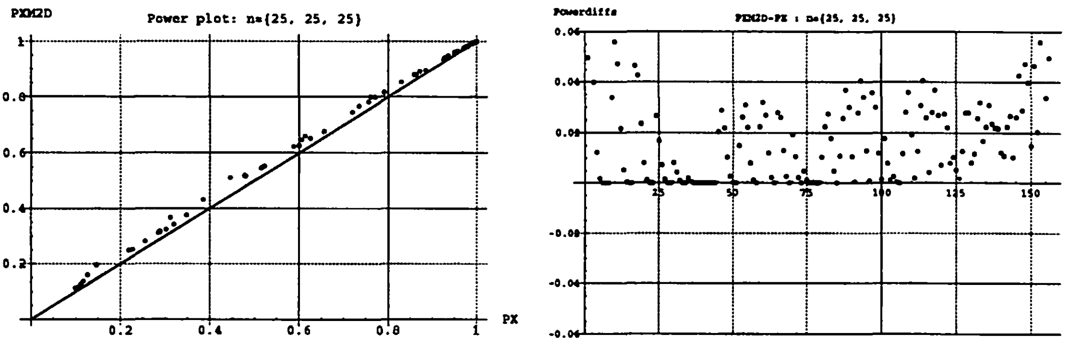

図 19: $n = (25, 25, 25)$ のとき.

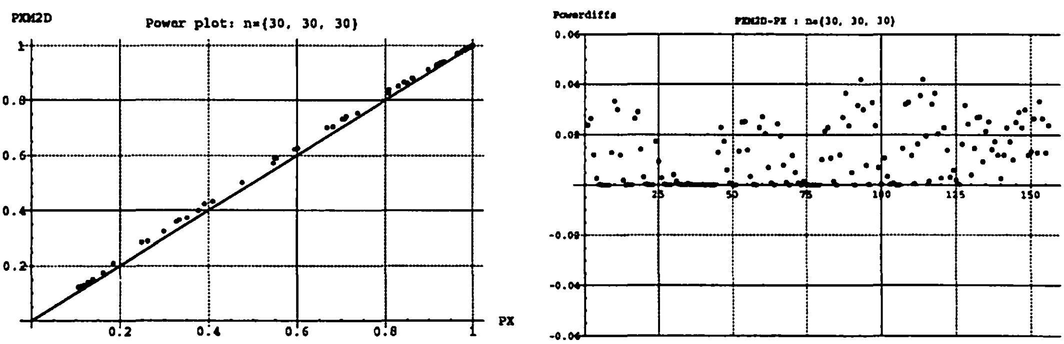

図 20: $n = (30, 30, 30)$ のとき.

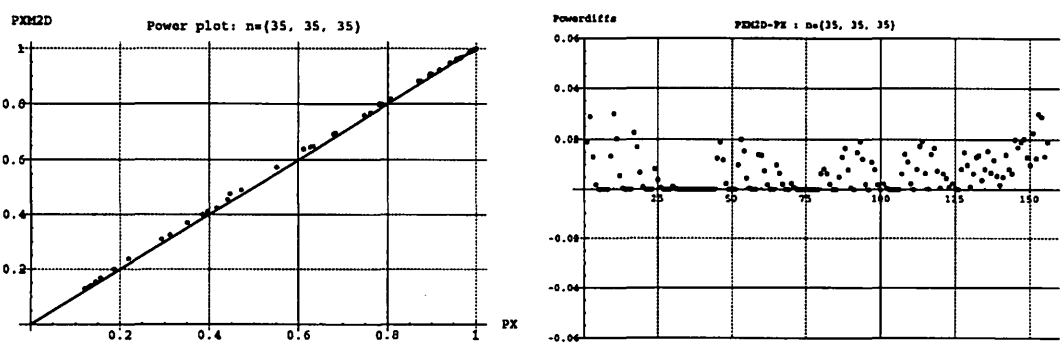

図 21: $n = (35, 35, 35)$ のとき.

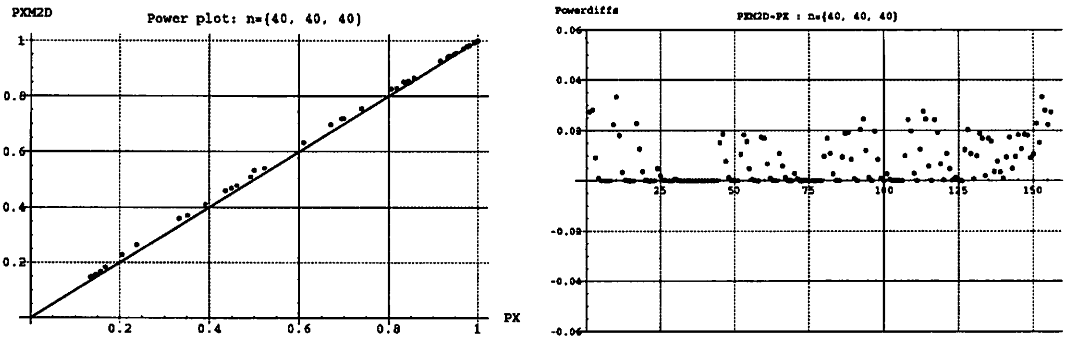

図 22: $n = (40, 40, 40)$ のとき.

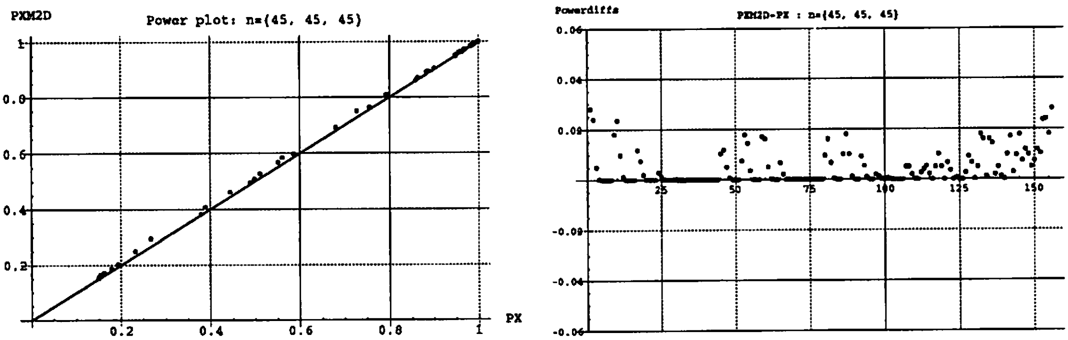

図 23: $n = (45, 45, 45)$ のとき.

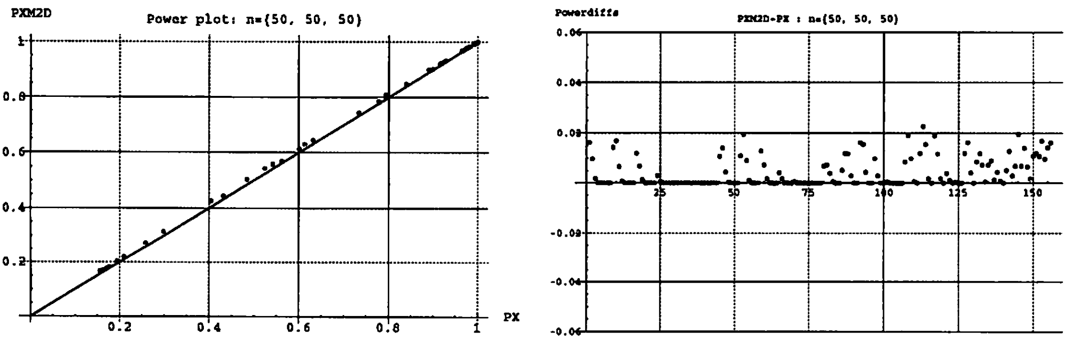

図 24: $n = (50, 50, 50)$ のとき.

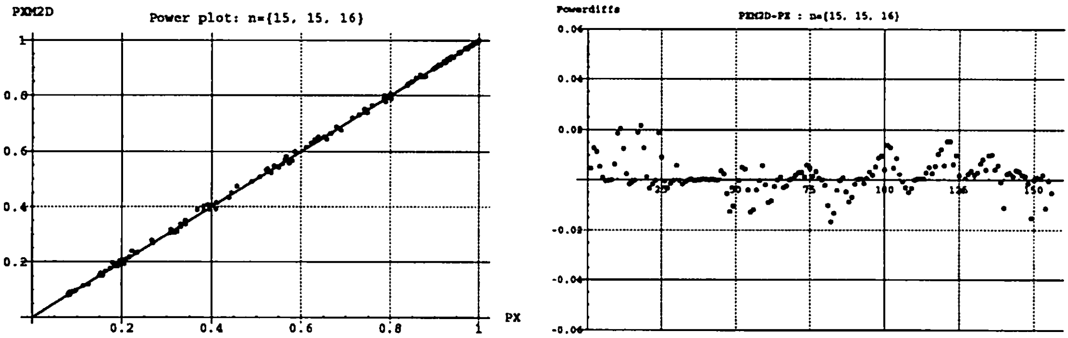

図 25: $n = (15, 15, 16)$ のとき.

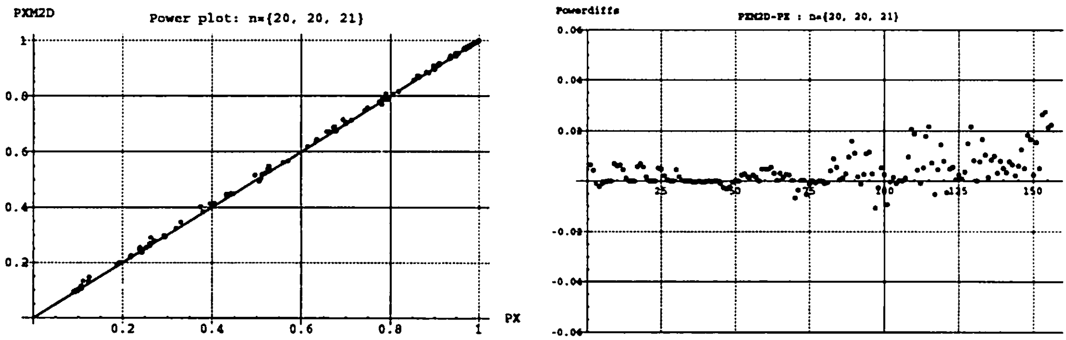

図 26: $n = (20, 20, 21)$ のとき.

5　終わりに

一般化線形モデルに属する連続変量モデルでは，局外パラメータの十分統計量で条件付けることにより最強力検定を実行できる．しかし，離散変量になると，その理論は成り立たない．なぜならば検定統計量の離散性のために，検定サイズが与えられた水準を達成できないからである．この問題は漸近論では全く問題にはならないのだが，実際のデータを扱う際には重大なものである．この論文では，いくつかの 2 項変量の等確率性検定問題に焦点を絞り，離散性を軽減する検定法として決定的な方式を提案できたと確信している．なお，この検定方式を用いるときのサンプルサイズの決定問題については，一般式で表すのは困難であり，アドホックな対応が必要になるだろう．

最後に強調しておきたいことは，実験計画の必要性である．出来ることならば，2 次元統計量を導入する必要のないように，サンプルサイズは互いに異なるようにしておく方がよい．そうしておいて元々の適合度検定統計量から修正統計量を構成し非条件付検定を行うことを推奨する．ここで紹介した方法が，偶発的にサンプルサイズが揃ってしまったときのためのものとなるよう，綿密に実験計画を行ってほしいと願っている．

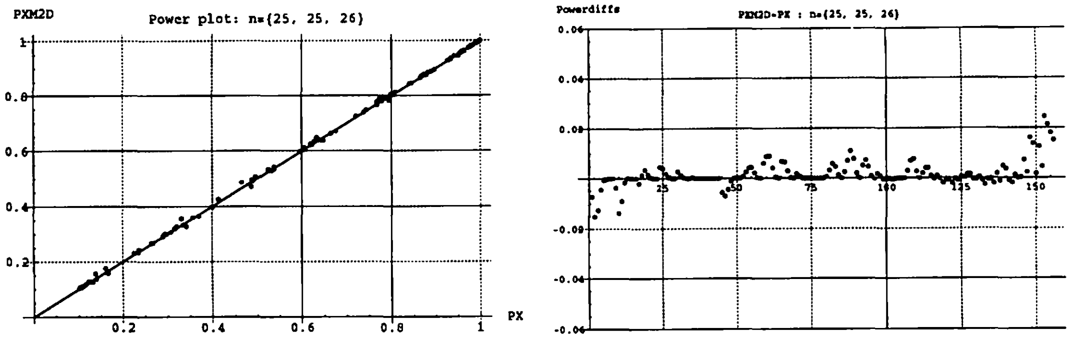

図 27: $\boldsymbol{n} = (25, 25, 26)$ のとき.

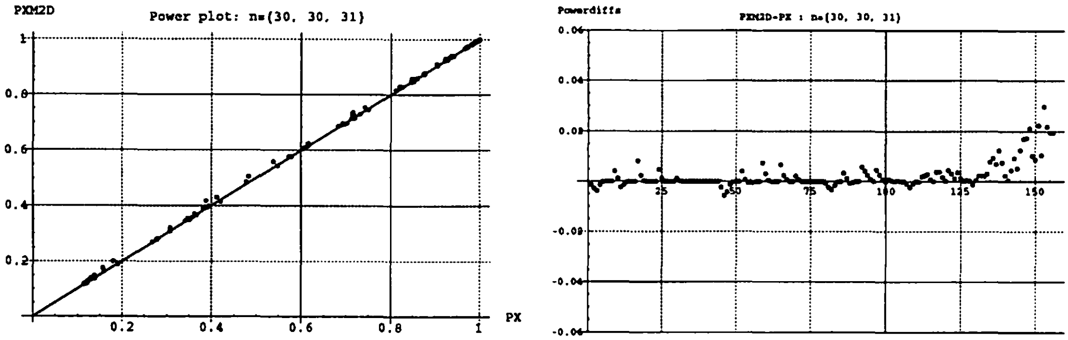

図 28: $\boldsymbol{n} = (30, 30, 31)$ のとき.

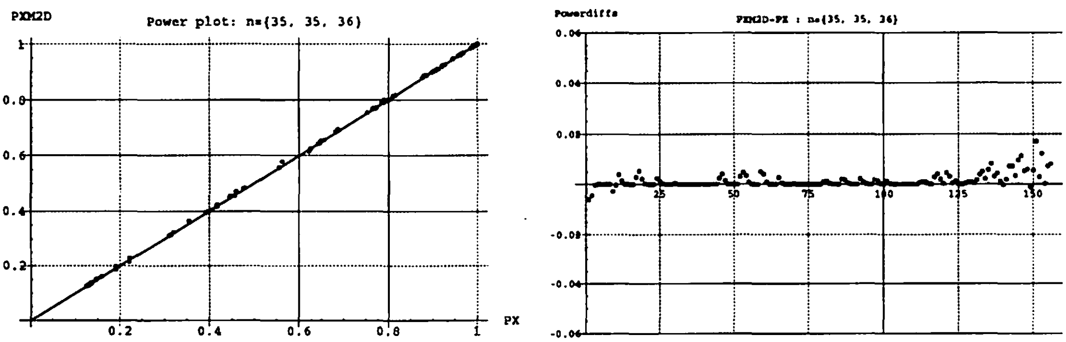

図 29: $\boldsymbol{n} = (35, 35, 36)$ のとき.

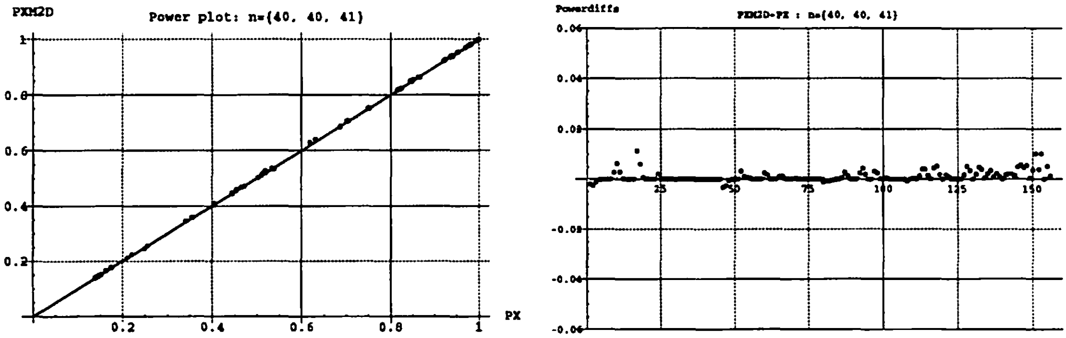

図 30: $\boldsymbol{n} = (40, 40, 41)$ のとき.

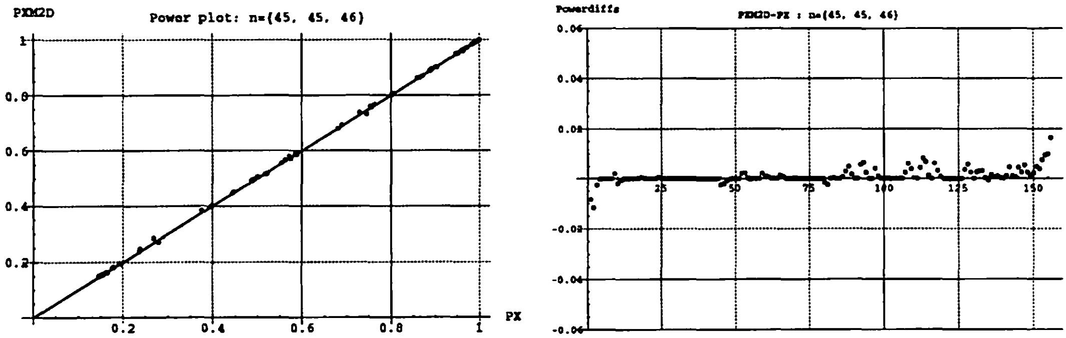

図 31: $\boldsymbol{n} = (45, 45, 46)$ のとき.

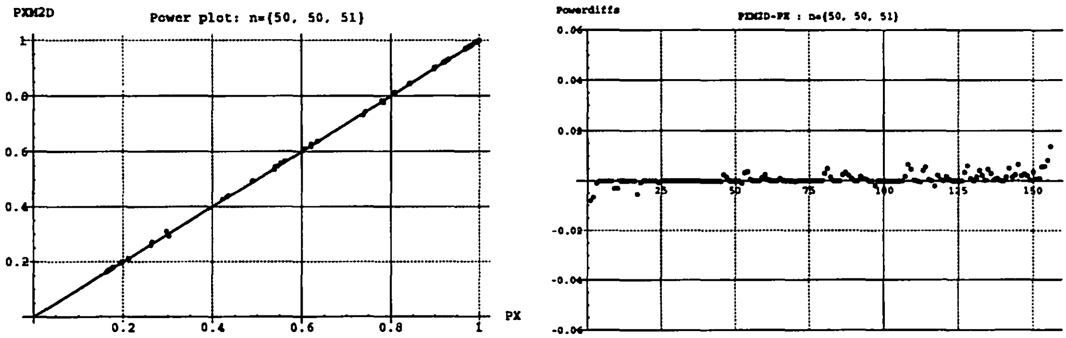

図 32: $\boldsymbol{n} = (50, 50, 51)$ のとき.

参考文献

[1] Matsuo, A. (1999). Exact conditional power comparison of three statistics in testing the equality of three binomial proportions. *J. Jpn. Soc. Comp. Statist.*.

[2] McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models* , 2nd edition. Chapman and Hall.

[3] Mehta, C. R. and Hilton, J. F. (1993). Exact power of conditional and unconditional tests: going beyond the 2×2 contingency table. *The American Statistician,* **47**, 91-98.

[4] Read, T. and Cressie, N. (1988). *Goodness-of-Fit Statistics for Discrete Multivariate Data.* Springer-Verlag.

[5] 松尾 精彦 (2000a). 二次元統計量を用いたロジスティック回帰モデルの条件付検定. 応用統計学, **29**, 1–25.

[6] 松尾 精彦 (2000b). 修正統計量を用いた非条件付正確検定法. 計算機統計学, **13**, 41–57.

[7] Weerahandi, S. (1995). *Exact Statistical Methods for data analysis.* Springer Verlag, Now York.