

教育ノート

確率統計教育におけるコンピュータの活用

二宮 玲子

日本女子大学理学部数物科学科

(2002年9月9日受理)

要 旨 昨今、教員による情報機器を利用した授業改善、教育方法の研究が活発に行われている。“確率統計と情報処理”の授業を担当する機会を得たので、確率、および統計の概念をわかりやすく説明するため、動的な、ビジュアルな表示を使えないかを念頭に置き、授業、演習方法を検討してきた。本報では、講義内容の理解を支援してくれるマルチメディア教材を開発、使用した例、および演習時、学習内容習得にコンピュータを有効活用した例等を報告する。

キーワード：確率統計、インタラクティブ教材、JavaApplet、コンピュータ・シミュレーション

はじめに

教職課程「情報」に関する選択科目「確率統計と情報処理」が開講され、2001年度より筆者が担当することになった。“確率統計”の入門教育にコンピュータは切り離せない。科目名に“情報処理”も加えられたので確率および統計の基本、定理、標本分布等の理解を支援する目的で、また演習問題を解くツールとして、コンピュータを活用した。

下記の授業内容（前後期）の中で種々の場面でコンピュータを使用し、講義、演習をすすめてきた。取り上げる順序は、使う教科書により多少入れ替わる。講義時は、ノートパソコンの画面を表示できるプロジェクタのある教室を使用した。

【授業内容】

- ・データのまとめ方についての概観
度数分布とヒストグラム、データの要約、散布図、相関係数
- ・データの分布
母集団と標本、確率と確率変数、期待値と分散
- ・確率分布
離散分布、連続分布、多変数の確率分布

- ・標本平均と標本分布
大数の法則、中心極限定理、チェビシェフの不等式
- ・統計的推測
点推定、最尤法
区間推定、検定
- ・正規母集団に関する推測
 χ^2 乗分布、母分散の推定と検定
t分布、母平均の推定と検定
F分布、母分散比（等分散性）の検定
母平均の差の検定
- ・相関分析、回帰分析
母相関の推定、検定、単回帰分析、重回帰分析
- ・計数値データの推測
母比率の推定と検定、適合度検定

授業において用いたコンピュータの利用法は、だいたい4つに分けられる。

1. 確率統計の講義をする上で有効な補助教材を開発、講義時に使う。
JavaAppletを使い、画面から数値の入力、あるいは、数値を選択する等インタラクティブな操作ができるグラフィック・シミュレーション等の教材を開発し講義時に使用する。
2. 講義時に Excel の操作性の良さを生かし、表作成、グラフィック表示をする。

Contribution No.: MP 02-1

3. 演習にコンピュータ・シミュレーションを取り入れる。

授業受講生は、すでにC言語を学んでいるのでC言語プログラムを作成できる。数式による説明だけでは理解しにくい概念を自分で実験しているつもりでシミュレーション実験プログラムを作成し、結果をファイルに出力する。その出力ファイルを Excel の入力ファイルとし、Excel でグラフ化、考察することにより確率現象について理解を深める。

4. 演習に Excel を活用する。

Excel の操作性の良さを生かした表作成、グラフィック表示をする。

Excel を使うことにより、統計量等の数式の理解を深める。

Excel の関数を駆使し、統計計算を行う。

以上、それぞれの利用に関して、いくつかの例を紹介する。

1. JavaApplet を使った動的グラフィック表示教材

1-1 2項分布に関するシミュレーション (図1, 図2)

図1は、2項分布 $B(n, p)$ の n と p を入力して分布曲線を描く。 p が小さい場合でも n を大きくしていくと対称に近い形になることをみる。

図2は、2項分布 $B(n, p)$ の n は20に固定してある。 p を入力し、 p の値による分布の形をみる。

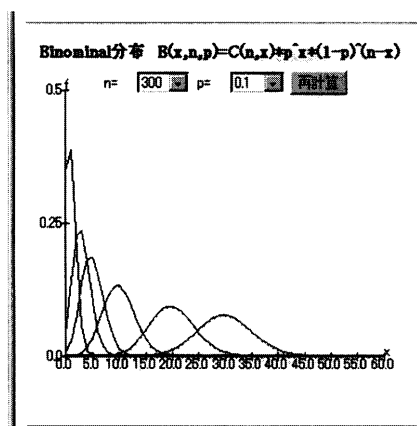


図1 nとpを与えたときの二項分布

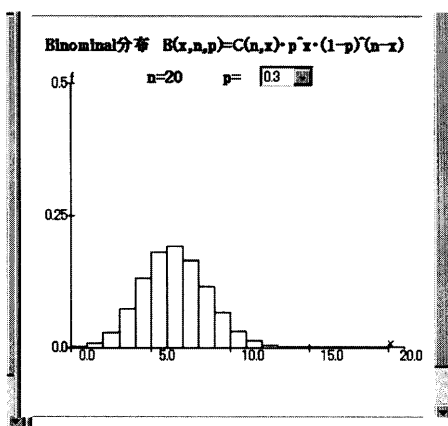


図2 n=20 pを与えたときの二項分布

1-2 中心極限定理に関するシミュレーション (図3)

母集団分布は、 $[0,1]$ の一様分布。標本数 n を入力し、一様分布から抽出された大きさ n の標本の平均値の度数分布と $N(0.5, 1/12n)$ を描く。

母集団が一様分布の場合、標本数が小さくても標本平均の分布は正規分布に近いことをみる。

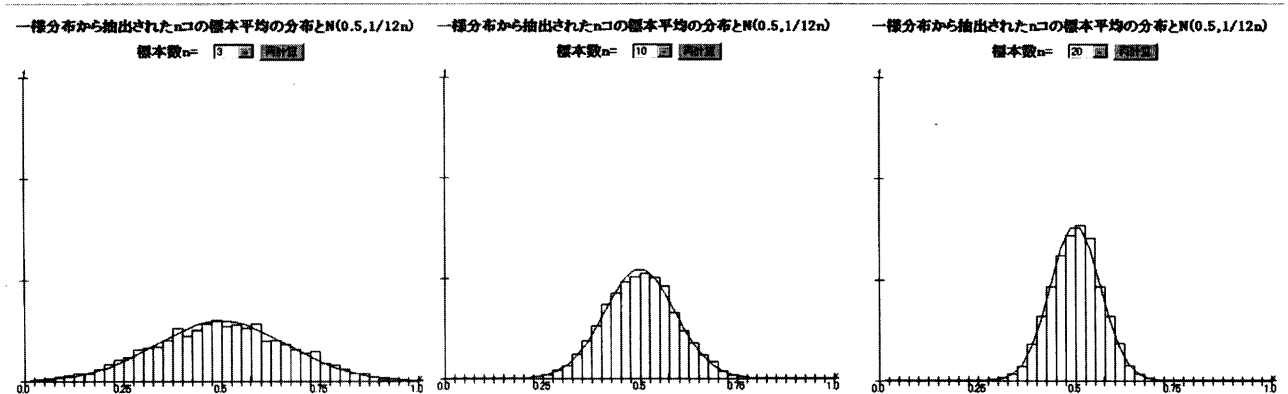


図3 中心極限定理のシミュレーション (区間 $[0,1]$ の一様分布からの標本平均の度数分布と正規分布)

1-3 χ^2 統計量の分布に関するシミュレーション

自由度を選択, χ^2 分布を描く。(図4)

標本数 n を入力する。母集団 $N(0,1)$ から大きさ n の標本を抽出, 値の2乗和の度数分布と自由度 n の χ^2 分布を描く。(図5)

1-4 T 統計量の分布に関するシミュレーション

自由度を選択, t 分布を描く。(図6)

標本数 n を入力する。母集団 $N(0,1)$ から大きさ n の標本を抽出, その T 統計量の度数分布と自由度 $n-1$ の t 分布を描く。(図7)

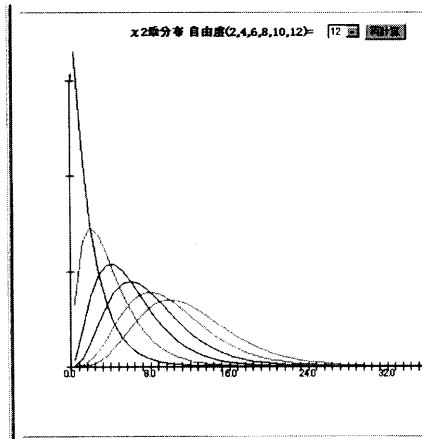


図4 自由度を与えたときの χ^2 分布

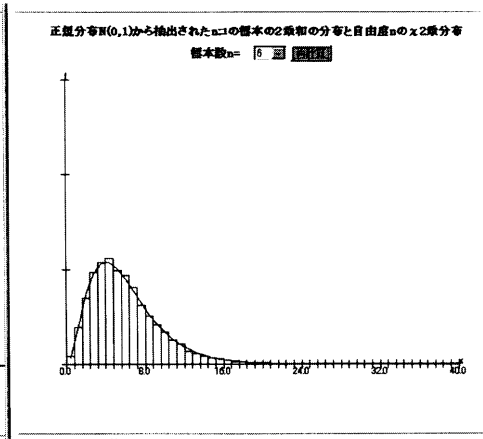


図5 シミュレーションによる χ^2 統計量の度数分布と χ^2 分布

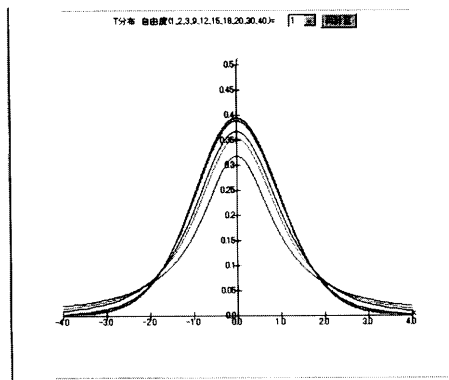


図6 自由度を与えたときの T 分布

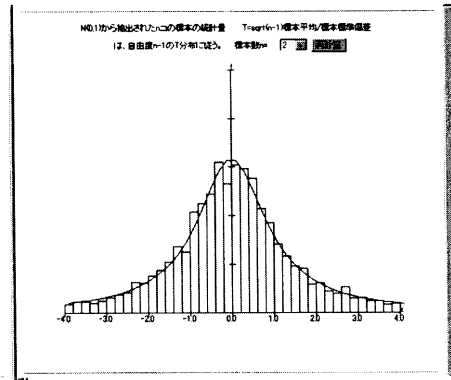


図7 シミュレーションによる T 統計量の度数分布と T 分布

1-5 F 統計量の分布に関するシミュレーション

F 分布の自由度 n_1, n_2 (ただし $n_1=n_2$) を選択し, F 分布を描く。(図8)

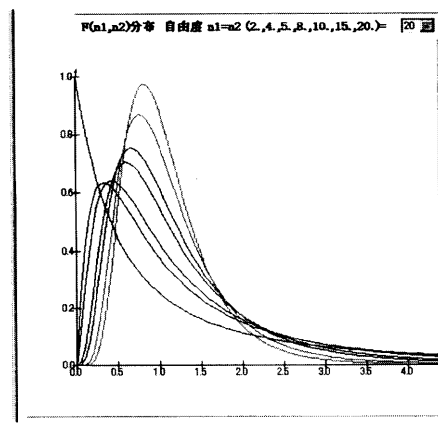


図8 自由度を与えたときの F 分布

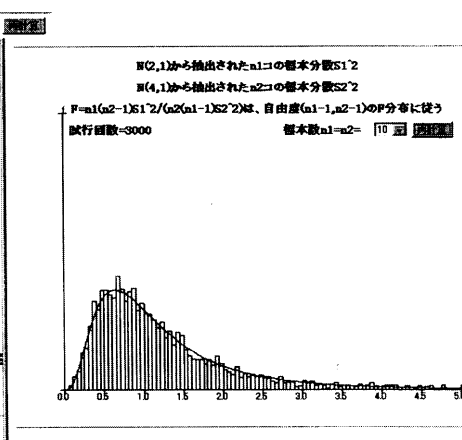


図9 シミュレーションによる F 統計量の度数分布と F 分布

2つの母集団から抽出する標本数 n_1, n_2 を入力する。母集団 $N(2,1), N(4,1)$ から抽出した n_1 個の標本, n_2 個の標本から得られる F 統計量の度数分布と自由度 (n_1-1, n_2-1) の F 分布を描く。(図 9)

1-6 母分散既知の場合, 母平均の区間推定シミュレーション

標本数 n を入力する。母集団 $N(165, 4^2)$ から n 個の標本を抽出, 平均値を求め, 母平均の 95% 信頼区間を求める。50 ケースの 95% 信頼区間を示す。

標本数 5 の場合の区間推定 (図10), 標本数15の場合の区間推定 (図11)

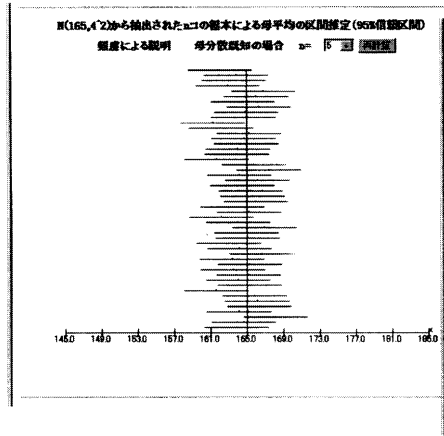


図10 $N(165, 4^2)$ からの標本平均による母平均の区間推定のシミュレーション実験
母分散既知 標本数: 5

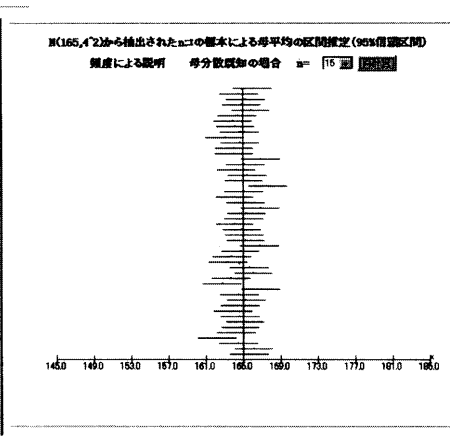


図11 $N(165, 4^2)$ からの標本平均による母平均の区間推定のシミュレーション実験
母分散既知 標本数: 15

1-7 母分散未知の場合, 母平均の区間推定シミュレーション

標本数 n を入力する。母集団 $N(165, 4^2)$ から n 個の標本を抽出, 平均値を求め, 母平均の 95% 信頼区間を求める。50 ケースの 95% 信頼区間を示す。

標本数 5 の場合の区間推定 (図12)

標本数15の場合の区間推定 (図13)

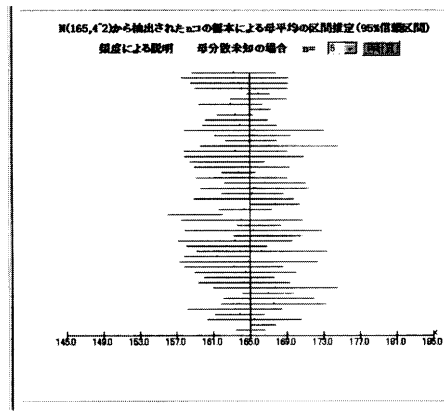


図12 $N(165, 4^2)$ からの標本平均による母平均の区間推定のシミュレーション実験
母分散未知 標本数: 5

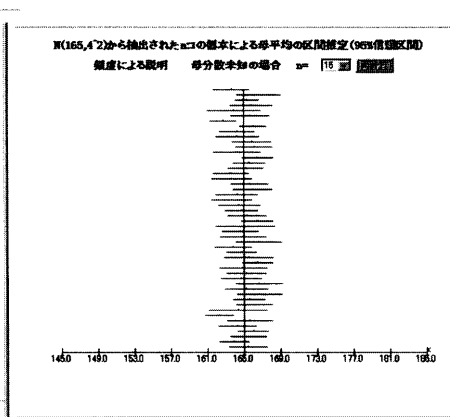


図13 $N(165, 4^2)$ からの標本平均による母平均の区間推定のシミュレーション実験
母分散未知 標本数: 15

1-8 相関に関するシミュレーション

母相関係数と標本数を入力し, 標本の散布図を描く。

図14は, 母相関係数 0, 標本数 200 の場合の標本散布図

図15は, 母相関係数 0.8, 標本数 200 の場合の標本散布図

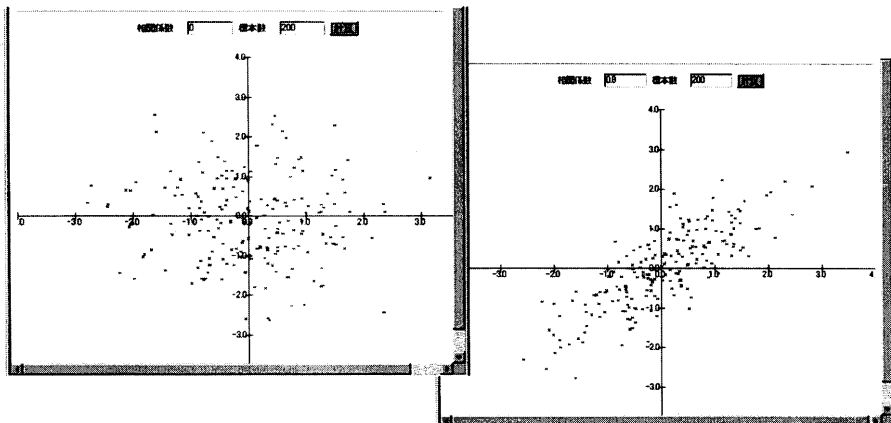


図14 母相関係数, 標本数を与えたときの抽出標本の分布 (母相関 0 標本数 200)

図15 母相関係数, 標本数を与えたときの抽出標本の分布 (母相関 0.8 標本数 200)

1-9 相関係数の分布に関するシミュレーション

母相関係数と標本数を入力し, 標本相関係数の度数分布を描く。

図16は, 母相関係数 0, 標本数 200 の場合の標本相関係数の度数分布

図17は, 母相関係数 0.8, 標本数 200 の場合の標本相関係数の度数分布

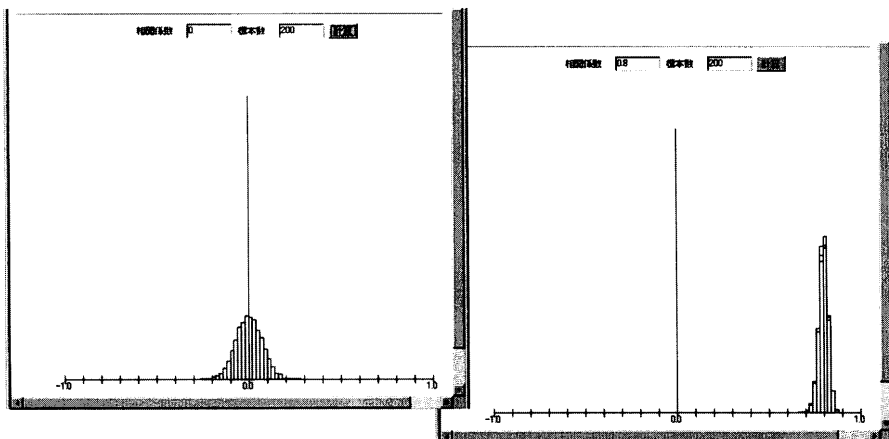


図16 母相関係数, 標本数を与えたときの標本相関係数の分布 (母相関 0 標本数 200)

図17 母相関係数, 標本数を与えたときの標本相関係数の分布 (母相関 0.8 標本数 200)

1-10 仮説検定における第1種の過誤と第2種の過誤

母集団分布 $N(\mu, \sigma^2)$, $\sigma = 1$

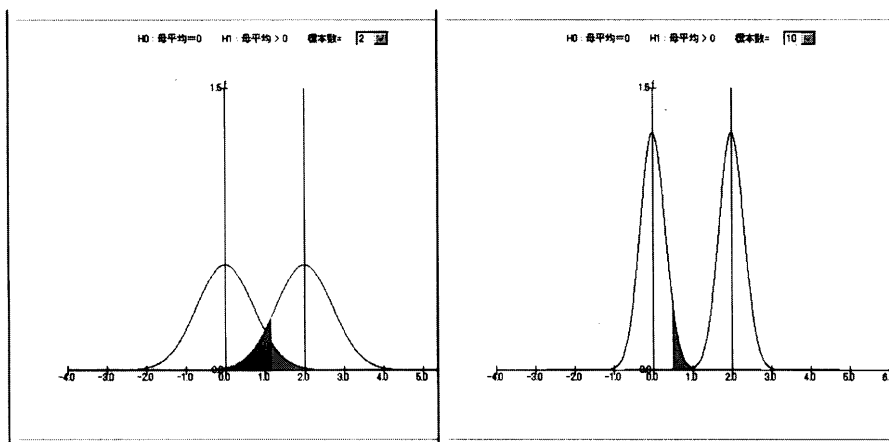


図18 仮説検定における第1種の過誤と第2種の過誤 (n=2)

図19 仮説検定における第1種の過誤と第2種の過誤 (n=10)

$H_0: \mu=0$ $H_1: \mu>0$ ($\mu=2$) を仮定, 標本数を入力し, 第1種の過誤と第2種の過誤を示す。
 図18に標本数2の場合, 図19に標本数10の場合

2. 講義時に Excel の操作性の良さを生かし, 表作成, グラフ表示を活用する。

Excel では, あるセルで数式を定義し, その数式をコピーすることで瞬時に表を作成できる。そして, グラフ化できる。この機能を活用, 確率分布の形状などを比較する。いくつか例をあげる。

2-1 2項分布 $B(10, p)$ の数表とグラフ (図20)
 2項分布の確率は, 次の漸化式で計算できる。

$$P(X=0) = (1-P)^n$$

$$P(X=x) = (n-x+1)/x * p / (1-p) * P(X=x-1)$$

($x=1,2,\dots,n$)

図20でA列に $x=0,1,2,\dots,n$ を2行に $p=0.1,0.2,\dots,0.9$ を入力する。

セルB3に “ $= (1-B\$2)^{10}$ ” を入力実行,
 セルB4に “ $= (10-\$A4+1)/\$A4*B\$2/(1-B\$2)*B3$ ” を入力, 実行する。後は, マウスをドラッグすることで数式をコピー, 表全体の数値が求められる。グラフも簡単に描くことができる。pと分布形状の対応をみることができる。

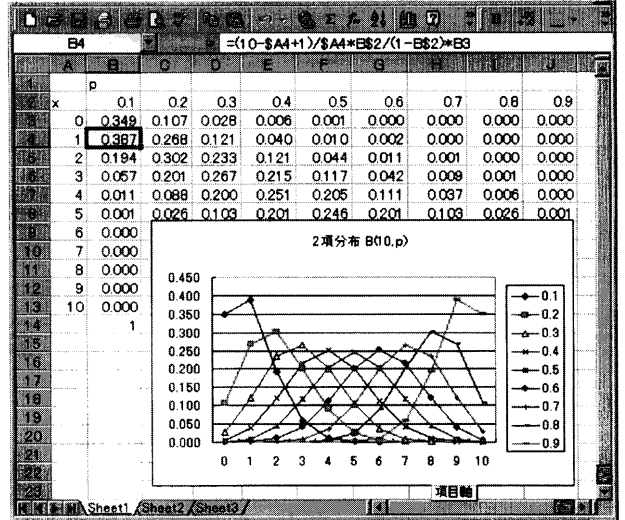


図20 2項分布 $B(10,p)$ の数表とグラフ

2-2 ポアソン分布の数表とグラフ (図21)

ポアソン分布の確率は, 次の漸化式で計算できる。

$$P(X=0) = e^{-\lambda}$$

$$P(X=x) = \lambda/x * P(X=x-1)$$

($x=1,2,3,\dots,n$)

図21でA列に x の値を, 2行に λ を入力する。

2-1の2項分布の例と同様, 2つのセルに数式を定義, 実行すれば, 後は, 数式のコピーにより表全体の値が求められる。グラフを描く。

λ と分布形状の対応をみることができる。平均値が λ であることもわかる。

2-3 2項分布で $np=一定$ (λ) な条件のもとで n を順次大きくしたとき, パラメータ λ のポアソン分布に従う。(図22)

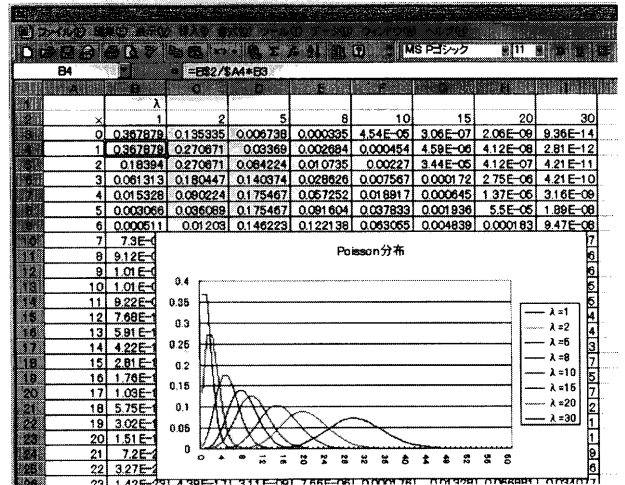


図21 パラメータ λ を変えたときのポアソン分布

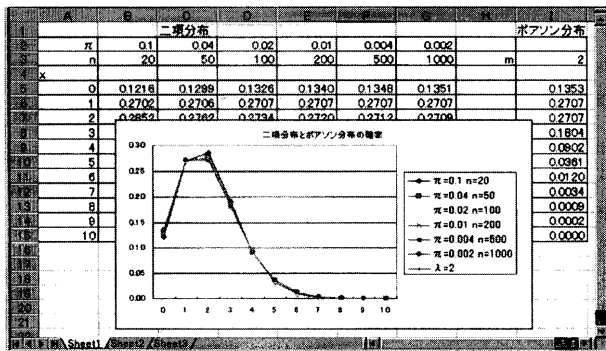


図22 母比率 $\pi \times n = 一定$ の二項分布と $\lambda = \pi \times n$ のポアソン分布

2-4 検定 $H_0: \mu=0$ $H_1: \mu>0$ に対する検定力曲線の数表とグラフ (図23)

検定における第2種の過誤の確率を β とおいたとき, $1-\beta$ を検定力という。

検定力と標本の大きさとの関係を片側検定 ($\alpha=0.05$) の場合を示す。

H_0 のもとでの母集団分布を $N(0,1)$ とする。

図23の表のA列に標本数, 1行に H_1 の μ の値を入力する。B列に各標本数の標本分布の上側 α 点を求める。Excel に用意されている2つの関数を使う。

正規累積分布の逆関数を返す NORMINV (確率, 平均, 標準偏差), 正規累積分布を返す NORMDIST (x, 平均, 標準偏差, 関数形式) を使う。

B2セルに “=NORMINV(0.95,0,SQRT(1/A2))” を入力, 実行し, B列にこの数式をコピーする。

C2セルに “=1-NORMDIST(\$B2,C\$1,SQRT(1/\$A2), TRUE)” を入力, 実行し, 表全体に数式をコピーする。

標本数ごとの検出力曲線を描く。

この表, グラフについては, 時間的な余裕があれば使用する。

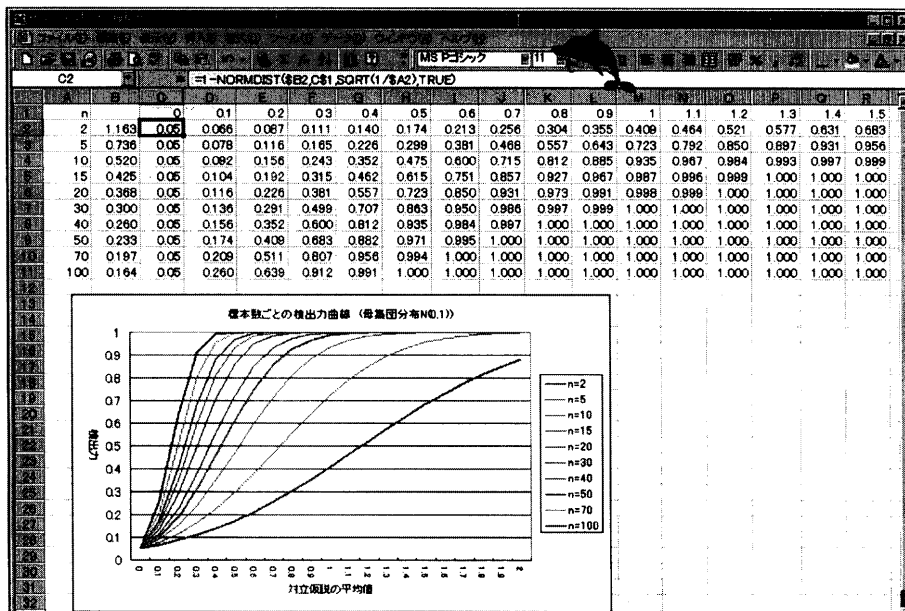


図23 標本数, 対立仮説の母平均ごとの検出力表と検出力曲線

3. 演習時に取り上げたコンピュータ・シミュレーション

標本分布を得るためのサンプリングの繰り返し等, 実際に試みるのが困難な試行をコンピュータという仮想空間で実現することができる。

次にあげる例は, 演習時, 学生にC言語プログラムを作成, Excelによるグラフ作成, 考察を課したものである。

3-1 相対頻度の収束で示す確率のシミュレーション

シミュレーションプログラムを作成する前にCに用意されている一様乱数発生関数 rand () で発生させた乱数が一様分布になることを度数分布のグラフ化により確かめた。

3-1-1 コイン投げシミュレーション

一様乱数 [0,1] を使って表0, 裏1を発生し, 100回ごとの相対度数をファイルに出力し, 出力結果をExcelでグラフ化する。(図24) [大数の法則] を理解する。

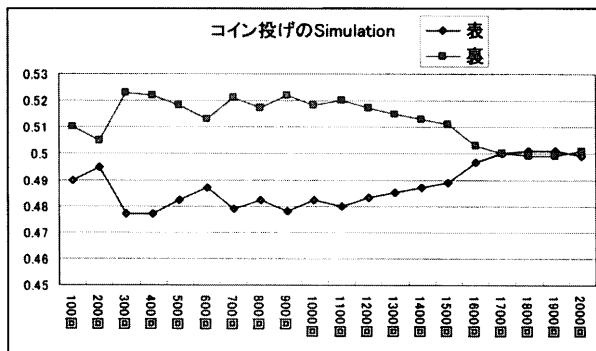


図24 コイン投げのシミュレーション

3-1-2 サイコロ投げのシミュレーション

整数 1～6 を発生し、100 回ごとの相対度数をファイルに出力し、出力結果を Excel でグラフ化する。(図25)

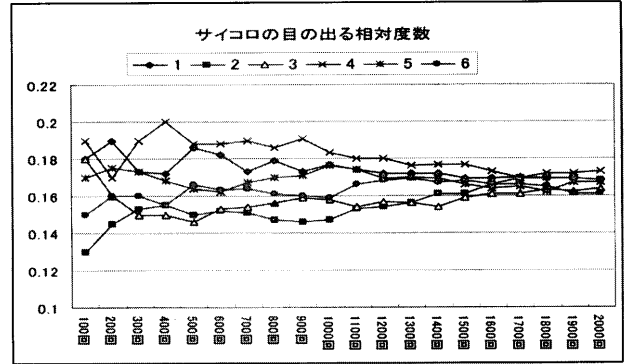


図25 サイコロ投げのシミュレーション

3-2 2項分布に関するシミュレーション

10個のコインを投げたとき、表の出る個数の分布を繰り返し、50, 100, 1,000, 2,000, 3,000, 5,000 回の相対度数分布と期待確率 $B(10, 0.5)$ の分布をファイル出力し、出力結果を Excel でグラフ化する。(図26)

繰り返し数を増やすに従い期待確率に収束することを確かめる。

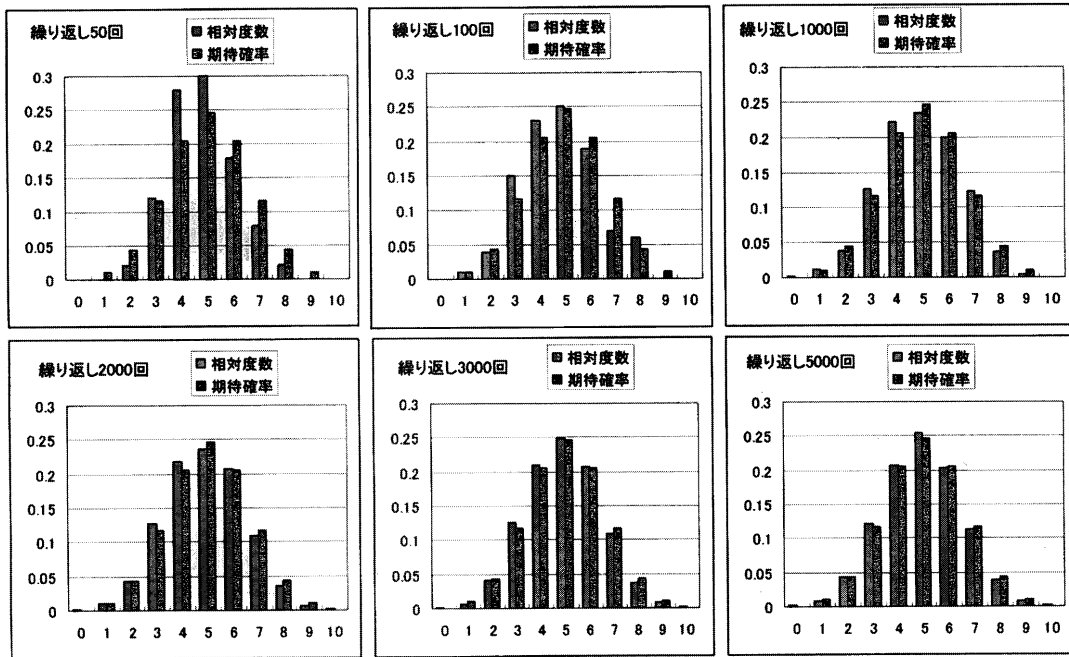


図26 二項分布のシミュレーション実験 (10個のコインを投げ、表の個数の分布と二項分布 $B(10, 0.5)$)

3-3 正規母集団から抽出された標本平均の分布に関するシミュレーション

中3男子身長を $N(165, 4^2)$ と仮定し、標本数を入力し、標本平均分布と期待確率を求めるシミュレーション

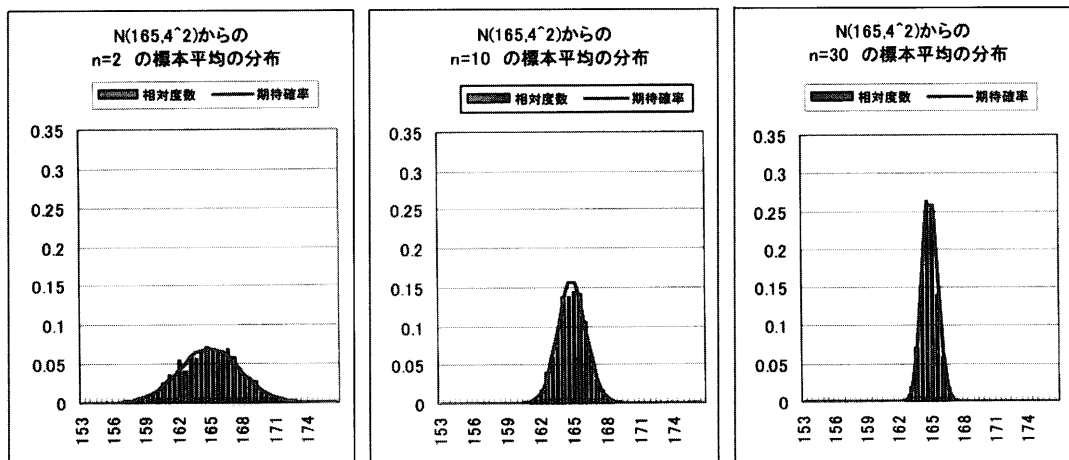


図27 シミュレーションによる母集団 $N(165, 4^2)$ からの標本平均の度数分布と $N(165, 4^2/n)$

ンプログラムを作成, 出力結果を Excel でグラフ化する。(図27)

この結果は, 母分散既知の場合の母平均の区間推定, 検定を考えるとときに用いられる。

3-4 中心極限定理に関するシミュレーション-1

歪のある母集団指数分布を取り上げる。

$$f(x) = \lambda e^{-\lambda x} \quad (x \geq 0)$$

指数分布母集団から抽出された n 個の標本の平均値の分布は, n が小さいときは, 歪がみられるが, n が十分大きければ近似的に正規分布 $N(1/\lambda, 1/\lambda^2 n)$ に従うことを確かめる。λ=1 とする。

出力結果を Excel でグラフ化する。(図28)

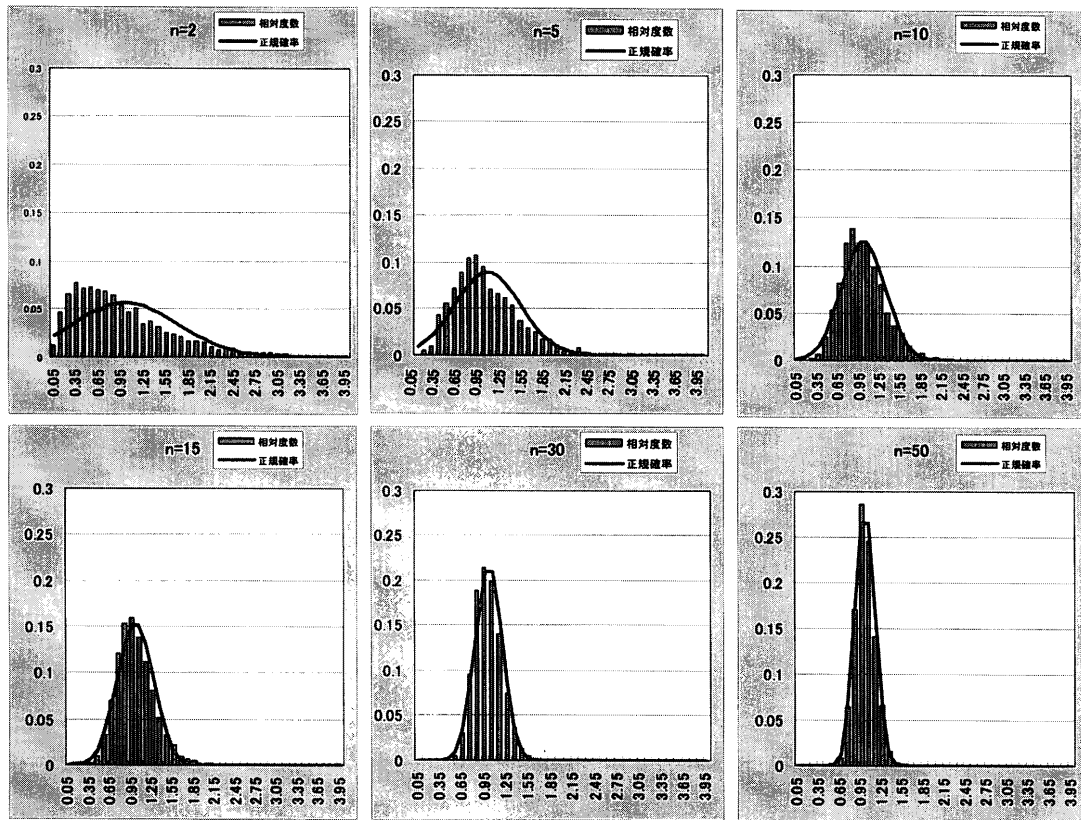


図28 中心極限定理のシミュレーション実験 (母集団分布が指数分布からの標本平均の分布)

3-5 中心極限定理に関するシミュレーション-2 (図29)

歪のない母集団一様分布を取り上げる。一様分布母集団からの標本平均の分布が, 正規分布に近づく度合いが早い

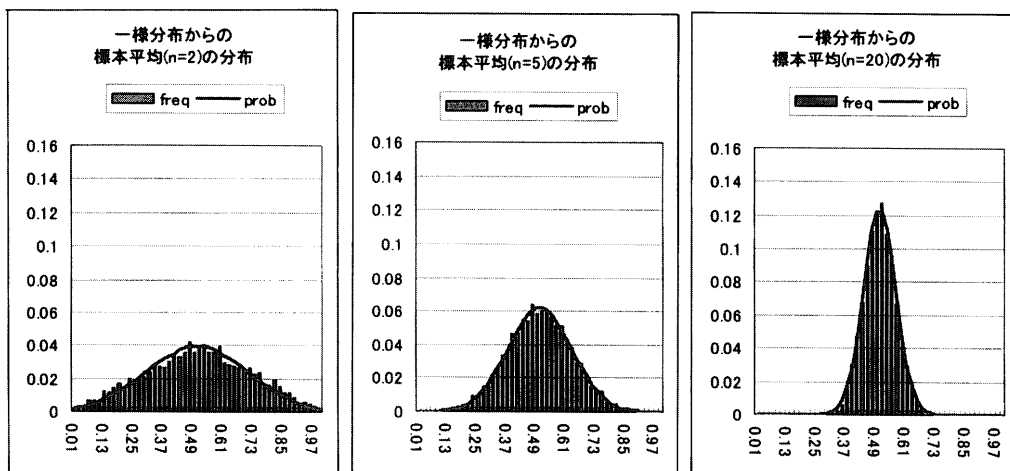


図29 中心極限定理のシミュレーション実験 (母集団分布: 一様分布からの標本平均の分布)

ことがわかる。

3-6 χ^2 統計量の分布に関するシミュレーション (図30)

正規母集団 $N(\mu, \sigma^2)$ から抽出した n 個の標本の

統計量 $\frac{\sum_{i=1}^n (X_i - \bar{X})^2}{\sigma^2}$ が自由度 $n-1$ の χ^2 分布に従うことを確かめる。

この結果は、母分散の区間推定、および検定を考えるときに用いられる。

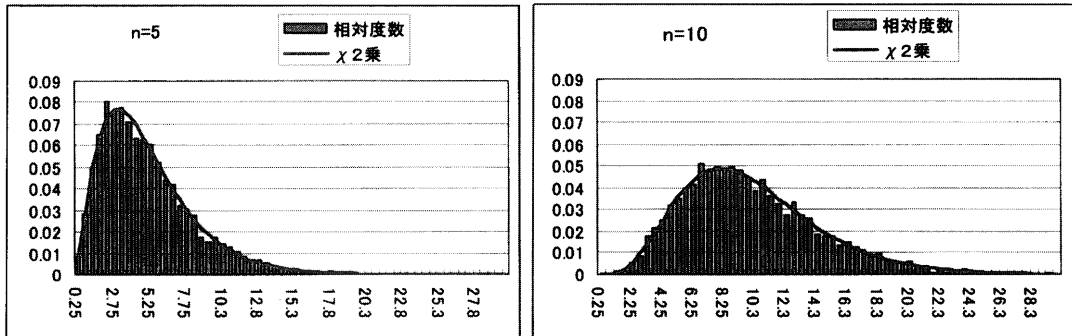


図30 シミュレーションによる χ^2 統計量の分布と期待確率分布

3-7 t 統計量の分布に関するシミュレーション (図31)

正規母集団 $N(\mu, \sigma^2)$ から抽出した n 個の標本の

統計量 $\frac{\bar{X} - \mu}{\sqrt{\frac{V}{n}}}$ が自由度 $n-1$ の t 分布に従うことを確かめる。ただし、 V は、不偏分散。

この結果は、母分散未知の場合の母平均の区間推定、および検定を考えるときにも用いられる。

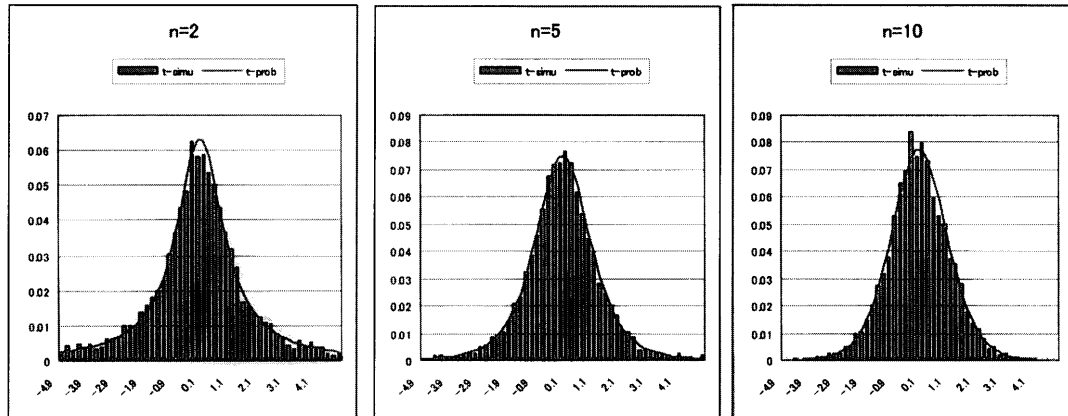


図31 シミュレーションによる t 統計量の度数分布と T 分布

3-8 F 統計量の分布に関するシミュレーション (図32)

正規母集団 $N(\mu_1, \sigma_1^2)$ から抽出した n_1 個の標本, 正規母集団 $N(\mu_2, \sigma_2^2)$ から抽出した n_2 個の標本から

得られる統計量 $\frac{V_1/V_2}{\sigma_1^2/\sigma_2^2}$ が自由度 (n_1-1, n_2-1) の F 分布に従うことを確かめる。 $\sigma_1^2 = \sigma_2^2$ とおく。

この結果は、母分散の比の検定を考えるときにも用いられる。

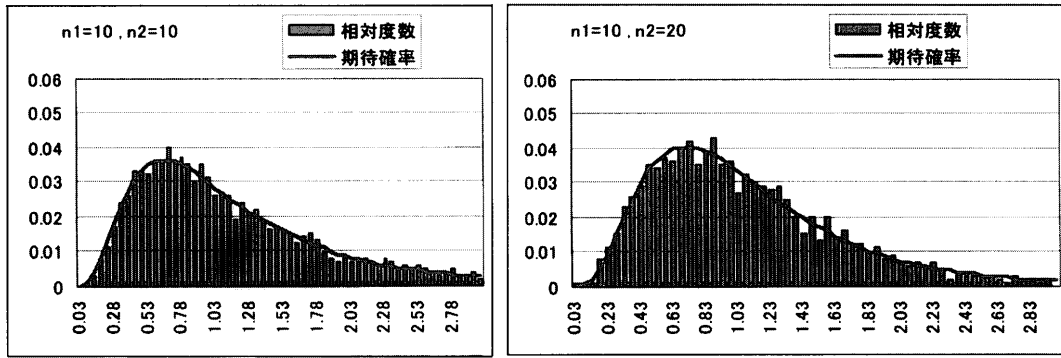


図32 シミュレーションによるf統計量の度数分布とF分布

4. 演習で Excel を活用する

1) Excel の操作性の良さを生かし、表作成やグラフィック表示に活用する。
先にあげた 2-1 の 2 項分布の数表とグラフ、2-2 のポアソン分布の数表とグラフ作成を演習でも課題とした。確率計算の漸化式により、2 つのセルの数式さえ正確に記述、実行できれば、瞬時に数表が求められ、グラ

フを書くことができる。考察にさく時間を充分とることができる。

4-1 稀な事象に対するポアソン近似

次に示す興味深い、稀な事象の観測データを取り上げポアソン近似を試みる。(図33)

- 4-1-1 東京郊外のある交差点で赤信号で止まった車の台数
- 4-1-2 ボルトケビッチの落馬死亡事故のデータ

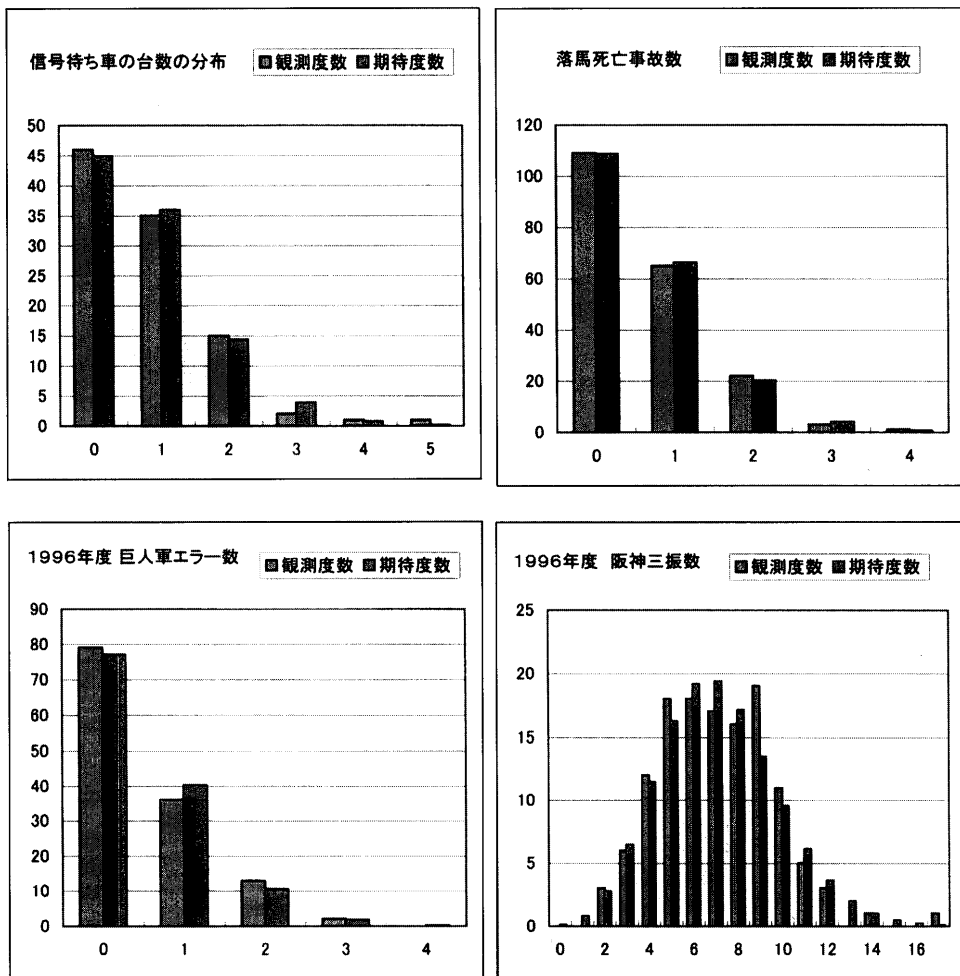


図33 4種の観測データとポアソン分布のあてはめ

4-1-3 1995年巨人軍のエラー数

4-1-4 1996年阪神の三振数

- 2) Excel を使うことにより, 統計量等の数式の理解を深める。
- 3) Excel の関数を駆使し, 統計計算を行う。推定, 検定に関する多くの演習問題を解く。

5. まとめ

演習では, いくつかのシミュレーション実験の課題が与えられた。学生に組んでもらうシミュレーションのプログラムは, インタラクティブにアクセスができる Java を採用することが望ましかった。しかし, 受講学生は C 言語を勉強中であり, Java 言語習得に時間をかけることはできなかったため, 実習では, C 言語を採用, グラフ化は, Excel でおこなった。1つの課題は, 1コマの時間内で仕上げ, 添付ファイルメールで提出された。標本分布のシミュレーション, 結果のグラフ化, 考察は, なかなか大変な作業であったが, 学生の集中力が上がってくることを感じ取ることができた。3. で示したグラフのように, 短時間で大変キメ細かく編集されたグラフが提出された。次週, 提出されたグラフを使って定理の説明, あるいは, 推定, 検定に発展させることができた。また, Excel を活用し, 多くの問題に触れることができた。

1. 授業時, 確率, 確率分布, 標本分布等統計の概念, あるいは定理の説明の補助教材としての JavaApplet 教材は, 画面から数値を入力, 実行というインタラクティブな操作により, ダイナミックにグラフィック表示ができ, 使いやすかった。統計的概念の理解を助ける目的に役立つ教材といえる。
2. 数式による説明だけでは理解しにくい概念を自分で実験しているつもりでシミュレーション実験プログラ

ムを作成し, グラフ化し, 考察することによって理解を助けることができる。

3. Excel の操作性のよさは, 効率的な数表作成, グラフ化に生かされ, 基本の把握に活用できる。
4. 推定, 検定等の演習問題処理に関しても Excel を活用することにより, 計算に時間をあまり費やすことなく, 考察に時間をかけることができる。

JavaApplet 教材は, コンピュータセンター NT サーバの授業用共有ファイルに置き, 参照できるようにした。今後, WBT サーバに置き, 授業支援, 学習支援に役立てたいと考えている。コンピュータ・シミュレーションにより標本分布を確かめた後, 得られたグラフをもとに推定, 検定をとりあげるとわかりやすい。ほとんどの教科書が標本分布, 推定, 検定と章立てが分かれているため, シミュレーション演習課題を生かした教科書の選択が必要とされる。

引用文献

- 1) 渡辺美智子, 櫻井尚子, 橋本紀子, 門間麻紀, 浅野美代子, 袴道守, 井上達紀, 山口和範: 論文誌 情報教育方法研究 2 37 (1999).
- 2) 渡辺美智子, 櫻井尚子, 山口和範, 井上達紀, 中川重和: 論文誌 情報教育方法研究 3 67 (2000).
- 3) Marlene Muller: Proceedings in Computational Statistics. 77 (1998).
- 4) 池田貞雄, 富田幸弘, 松井敬, 馬場善久: 統計学 データから現実をさぐる 内田老鶴園 東京 (1991).
- 5) 辻谷将明, 和田武夫: パワーアップ確率統計 共立出版 東京 (2001).
- 6) 杉山高一, 牛沢賢二: パソコンによる統計解析 朝倉書店 東京 (1992).
- 7) 桑原信也: プログラミング学習シリーズ Java 2 翔泳社 東京 (2001).

On the Teaching of Probability-Statistics by Using Computer

Reiko Ninomiya

Department of Mathematical and Physical Sciences, Faculty of Science, Japan Women's University

(Received September 9, 2002)

Abstract: The paper presents various methods using computer for teaching of Probability-Statistics.

- Interactive Graphical Simulations in Java applet are executed for explanation of statistical concepts.
- Students construct some simulation programs for understanding more easy statistical concepts (probability distributions, central limit theorem, sampling distributions and etc.).

Key words: Probability-Statistics, interactive graphic module, Java applet, computer simulation