

# 数理統計学における検定の理解のための可視化

桃井 央† 斎藤 隆文‡

東京農工大学 工学部情報工学科†

東京農工大学大学院 生物システム応用科学府‡

## 1. 背景と目的

数理統計学における検定は、工学分野における定量的評価をはじめとし、さまざまな分野で利用される学問である。しかし式に数値を代入する作業となることが多く、各検定について本質的な理解に及んでいないことが多い。

そこで本研究では、数理統計学の本質的な意味がわかるように、様々な分布や検定における有意性等を視覚的に示し、教育へ利用することを目的とする。

## 2. 視覚的理解の有効性

数学を可視化して捉えることは、より深い理解を助け、数学の有用性や重要性も学ぶことができる効果が期待できる。また、「動き」は数学のイメージ造りに貢献する[1]。統計学をはじめ、数学は抽象的な学問であり、直感的に分かりにくい。そのため視覚情報があることによって理解しやすくなる。

## 3. カイ二乗検定

カイ二乗検定と呼ばれる検定の中で最も利用されているものが、ピアソンの提案した手法である。ピアソンのカイ二乗検定は、観察された事象の相対的頻度がカイ二乗分布に従うという帰無仮説を検定するものである。ここで検定統計量は事象の期待値を用いたカイ二乗であり、

$$\chi^2 = \sum_{i=1}^n \frac{(O_i - E)^2}{E} \quad (1)$$

で定義される。ただし、 $O_i$  : 測定値、 $E$  : 理論値である[2]。

このカイ二乗の値がカイ二乗分布に漸近的に従うと仮定したとき、有意水準に基づいて、帰無仮説の有意性が示される。

## 4. 既存手法

現在、統計学に関しての可視化が行われている例がいくつかある。本項では、その中の1つである、白田ら(2015)による「共分散に関する多変量解析の可視化教材[3]」について述べる。

統計データにおいて、標本サイズを任意で変更することが容易に可能であり、その変化の様子を視覚的に見せることで定理などの数学的意味を理解できるようになる。そのため統計学における可視化は有効である。

この教材では、回帰分析と主成分分析を用いた分散および共分散の意味の可視化を行う。スライダーを用いることでユーザが任意に標本サイズを変更できるようにし、変化の様子を容易に観察でき直感的な理解を促すことができる。

## 5. 提案手法

本研究では、各種検定に関しての新たな可視化手法を提案する。

### 5.1 確率分布の可視化

検定では、データが従う確率分布をもとに有意性を判断する。そこで確率分布がどのようにして変化するかを、アニメーションを用いて可視化する。アニメーションにすることで確率分布の変化の様子を観察できる。変化の様子を観察することは、学問の理解においては重要なことである。確率分布を提示する際に多量のデータを必要とするため、乱数を用いてデータを生成する。多量のデータを生成することなど手作業では扱いにくいことが、コンピュータの利用で容易になる[4]。乱数から生成したデータを標本ごとに順に出力し、図1にアニメーションの一部を示す。

### 5.2 理解のためのシステム

本研究では、検定の理解のために可視化を行うが、テキストも併用して検定の原理を解説、提示する。検定の確率分布に関しては前節で述べた。この様子を画面左にテキスト、画面右にアニメーションを並べて表示する。データの総数に対して横軸の各値の積み上げ量は正規化すると相対度数を示している。よって、グラフの

Visualization for understanding certification in mathematical statistics

†Department of Computer and Information Sciences, Tokyo University of Agriculture and Technology.

‡Graduate School of Bio-Applications and Systems Engineering, Tokyo University of Agriculture and Technology.

内部および周の面積が確率全体を表していることに留意させると，検定統計量の意味の理解が容易となる．図2にインタフェースの完成草案を示す．図2のように，5.2で述べた手法と解説テキストを併用することで，教材としての形式を取る．

## 6. おわりに

本研究では，検定の理解のために新たな可視化手法を提案した．本稿では手法の提案のみを行った．そのため，提案手法の適切な評価を行う必要があり，現在思案している．

また，可視化手法を用いた教材等，ユーザの理解のために利用できるシステムの作成等の課題が残っている．

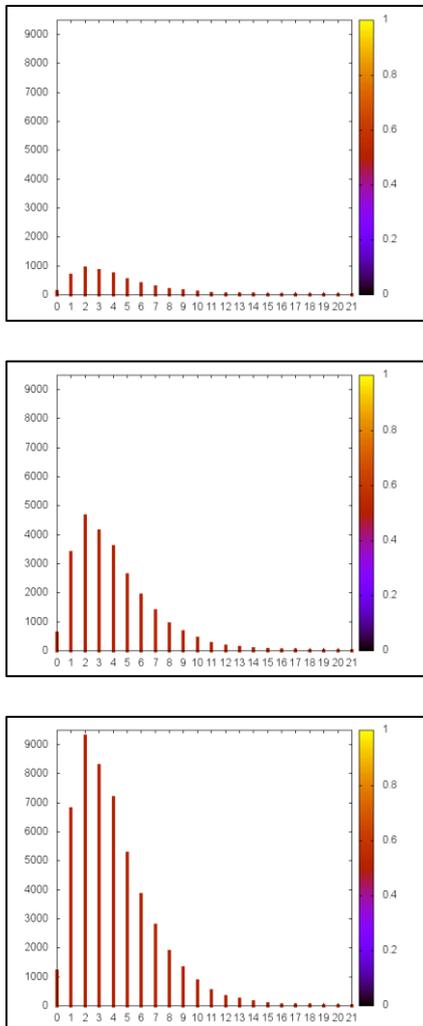


図1 確率分布曲線のアニメーションの様子

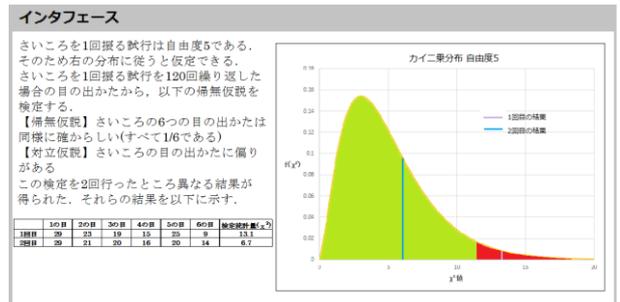
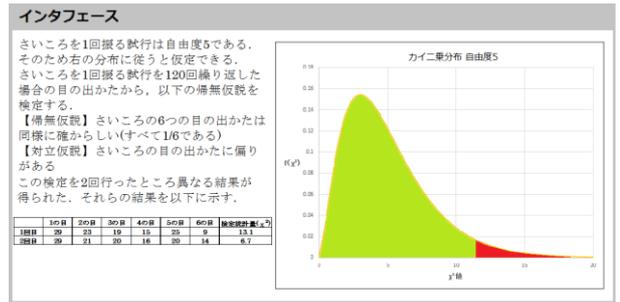
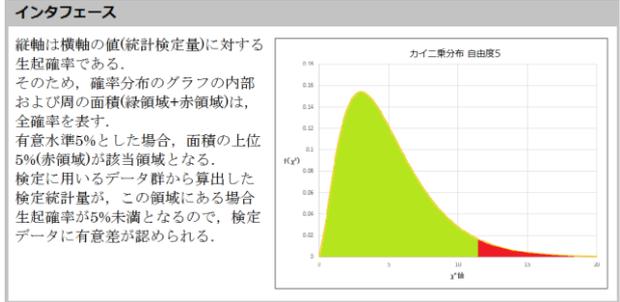


図2 インタフェースの草案

## 参考文献

- [1] 山本修一，“イメージ化がもたらす深い理解と数学の有用性”，数理解析研究所講究録，第1780巻，2012年，pp.221-231
- [2] Sarah Boslaugh，“統計クイックリファレンス第2版”，オーム社，2015年
- [3] 白田由香利，高橋裕，“共分散に関する多変量解析の可視化教材”，学習院大学経済論集，第52巻，第2号，2015年7月，pp.49-63
- [4] 鈴木将史，“確率統計における視覚化の効用”，イプシロン，Vol.4，2002年，pp.65-71