確率密度と統計教育の研究
~ポアソン分布の教育効果を中心に~

新堀友太[*1]・高數学[*2]
Email: a090131k@st.u-gakugei.ac.jp

*1：東京学芸大学教育学部初等教員養成課程
*2：東京学芸大学

◎Key Words　ポアソン分布, 統計学, 確率論

1　はじめに

確率論において学習者の理解を困難にする要因の中には「確率」と「確率密度」の違いがある。本研究は、ポアソン分布を中心とした教育方法について検討を行うものである。教育対象は統計・確率の専門知識獲得を目指す学習者とする。

1.1 研究背景

近年、統計やそれを学ぶ学問である統計学は非常に注目されている。これはITの進歩により様々なデータを保存・管理することが容易となったことが要因である。それにより、多くの分野の研究・ビジネスが、今までのプロセスを変えることとなった。そうした管理・保存されたデータを価値あるものとして見出していく際、統計手法による解析・分析は非常に効果的である。つまり、それを学ぶ統計学はあらゆる分野に対して必要となる重要な学問となっている。

また統計において確率は重要なものである。なぜなら実際の事象を統計手法で分析することは難しく、事象を数量として扱うための測度が必要だからである。その際に確率は非常に有用な測度であるといえる。過去において統計は近代確率論の結果を受けて発展していった。現代の統計の基礎は、そういった過去の統計学の蓄積の上で成り立っている。このことからも、統計を学ぶ上で確率の理解が重要であるということがわかる。

しかしながら、確率論において学習者の理解を困難にする要因がいくつかある。中でも「確率」と「確率密度」の違いについては、高度な数学の知識を要求するため、学習者に明快に理解されていない。しかしこれの理解は、確率分布の特性を理解することにつながり、統計と確率の相互の関連を深く理解するために重要である。また離散型分布において「確率」と「確率密度」が同じ値を取るため、それぞれを混同させてしまうといった点でも、学習者の理解をより困難にしてしまっている。

1.2　研究目的

そこで本研究ではポアソン分布に着目し、この２つの相違についての教育方法の検討を行う。ポアソン分布とは指数分布族に属する離散型分布である。また、ある事象の単位時間あたりの生起確率を示す分布である。このポアソン分布は二項分布から求めることができ、指数分布と深く関係している分布である。また統計分析モデルの中でも、一般化線形モデルにおいて、誤差項の従う分布はポアソン分布を含む指数型分布族である。その中でもポアソン回帰分析は、平均値が負の値を取らないカウントデータを分析する際に有効な手法である。そのため、交通事故の発生数や一定期間内に疾患を発症した例数といったデータの分析に用いられる。

1.3　研究意義

本研究では統計・確率の専門知識獲得を目指す学習を対象とし、「確率」と「確率密度」の相違について、ポアソン分布を中心としコンピュータを用いた統計学を援用することで教育していく。これによりこの２つの相違を学習者に理解させるとともに、一般化線形モデルまで教育を拡張し、確率・統計の関係性について視覚的に理解を深める教育効果が望めると考える。

2　統計と確率の定義

統計は一般に次のように言える。(3)

「統計(statistics)という用語の基本的な意味は、集団を記述する数量ということである」

集団とは、何らかの意味で同質とみなされ、同時に諸特徴・属性は均一でなく不規則に変動しているような個体の集まりのことを指す。

また確率は一般に以下のように定義する。(5)

(Ⅰ)$A \in \mathfrak{F}$に対し$P(A) \geq 0$

(Ⅱ)$P(\Omega)=1$

(Ⅲ)$A_1, A_2, A_3, \cdots \in \mathfrak{F}$で$A_i \cap A_j = \varphi(i = j)$ならば

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

ここで$\mathfrak{F}$は完全加法族である。[1]

確率は確率測度の一般的な呼称であり、何らかの事象を数量的に測る際に用いられるものである。

3　問題提起

3.1　「確率」と「確率密度」

前節で確率を定義したが、確率密度は以下のように説明できる。確率密度とは、特性関数と呼ばれるものの中の一つである確率密度関数によってもとめられる

[1] 完全加法族とは集合族の一つであり、特徴として$\mathfrak{F}$の元Aに可算無限が許されることがあげられる。(5)

値である。確率密度関数とは分布関数を微分することで得られる。一般に一つの確率変数 X に対し一つの確率密度関数を定めることができる。特定の密度関数をもつ確率変数の性質については、確率変数を介在させることなく統一的に論ずることができる。これにより中心極限定理による分布の収束など統計における重要な演算が可能となり、統計分析が可能となる。

3.2 「確率」と「確率密度」の相違

前節のとおり「確率」と「確率密度」は違うものであるが、既存の教育方法では学習者はこの２つの違いについて明快に理解できていない。理由として

①高度な数学の知識を網羅的に扱っている
②離散型分布において「確率」と「確率密度」が同じ値を取るためそれぞれを混同してしまう

という２点が挙げられる。

まず問題点①について、「確率」・「確率密度」をそれぞれ別々に教育する場合、「確率」の理解のためには完全加法族、集合、「確率密度」の理解のためには特性関数、微分といった、数学の知識が必要である。さらに、それらの関連についても理解することが必要である。既存の教育方法では、これらの高度な数学の知識の獲得を目指す場合、網羅的に扱っているため、それぞれの関連へのアプローチが欠如してしまう。また、「確率」・「確率密度」それぞれを理解する前に数学の部分で学習者の理解が困難になってしまっていると考えられる。

また問題点②は先ほどの高度な数学の知識に対し、説明を省略し部分的な理解のみを促している場合に起こる問題である。学習者が「確率」と「確率密度」それぞれについて、部分的にのみ理解してしまっている。そのため、離散型分布での「確率」・「確率密度」について、同じ値という結果について理論を説明することができない。

つまり「確率」と「確率密度」の違いを理解する場合、高度な数学の知識が学習者の明快な理解を妨げていると言える。そこで本研究ではこれらの問題点を解決するための教育方法を検討していく。既存の教育方法はそれぞれ

①'高度な数学の知識を前提とし「確率」・「確率密度」を教授する
②'高度な数学の知識のうち一部を教授し、それによって「確率」・「確率密度」それぞれの部分的な教授にとどまる

のどちらかであったが、本研究では

③「確率」・「確率密度」の相違そのものについて分布を用いて教授する

という教育のプロセスについて提案する。これにより、学習者は高度な数学の知識を前提とせず、相互の関連について、明快に理解することが可能であると考える。

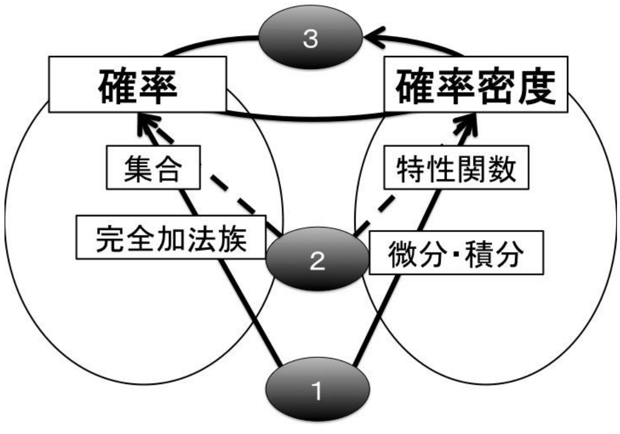

図 1　教育方法イメージ比較図

また、ここで用いる分布としてポアソン分布に着目する。本論文では、４にてポアソン分布の概要について述べ、５にてポアソン分布に着目する意義について検討し、６にて教育方法の提案を行う。

4　ポアソン分布の概要

4.1 ポアソン分布について

ポアソン分布は

$$P_{\gamma}(r) = \frac{e^{-\lambda}\lambda^r}{r!}$$

で定義される０および自然数上の分布と定義される。

これは二項分布$B_{n,p}(r)$について$np = \lambda$（λは一定）とし、$n\to\infty$（つまり $p\to 0$）にすることで得られる。

$$B_{n,p}(r) = nC_r p^r q^{n-r}$$

$$= \frac{n!}{r!\,(N-R)!}\left(\frac{\lambda}{n}\right)^r\left(1-\frac{\lambda}{n}\right)^{n-r}$$

$$= \frac{\lambda^r n(n-1)\ldots(n-r+1)}{r!\,nn\ldots n}\left(1-\frac{\lambda}{n}\right)^{-r}\left(1-\frac{\lambda}{n}\right)^{n}$$

ここで$\frac{n-1}{n}, \frac{n-2}{n}, \ldots, \frac{n-r+1}{n}, \left(1-\frac{\lambda}{n}\right)^{-r}$は n を限りなく大にするときすべて１に収束する。

また$\left(1-\frac{\lambda}{n}\right)^{n} = \left(1-\frac{\lambda}{n}\right)^{(-n/\gamma)(-\lambda)}$であり、$\left(1-\frac{\lambda}{n}\right)^{(-n/\lambda)}$はnを無限に大にするとき e に収束することから

$$B_{n,p}(r) \to \frac{e^{-\lambda}\lambda^r}{r!} = P_{\lambda}(r)$$

となる。

4.2 ポアソン分布の再生性

ポアソン分布は再生性をもつ。それは以下のように証明できる。

独立な確率変数 X,Y がそれぞれポアソン分布$P(\lambda_1), P(\lambda_2)$に従うとする。それらの確率変数の和を Z とすると

$$P(Z = k) = \sum_{j=0}^{k} P(X = j)\,P(Y = k - j)$$

$$= \sum_{j=0}^{k} \frac{\lambda_1{}^j}{j!} e^{-\lambda_1} \frac{\lambda_2{}^{k-j}}{(k-j)!} e^{-\lambda_2}$$
$$= \frac{e^{-(\lambda_1+\lambda_2)}}{k!} \sum_{j=0}^{k} \frac{k!}{j!\,(k-j)!} \lambda_1{}^j \lambda_2{}^{k-j}$$
$$= e^{-(\lambda_1+\lambda_2)} \frac{(\lambda_1+\lambda_2)^k}{k!}$$

となり、ポアソン分布$P(\lambda_1+\lambda_2)$に従う。

4.3　統計学におけるポアソン分布

ポアソン分布はある事象の単位時間当たりの生起確率を示す平均・分散がλの分布である。このことから平均値が負の値を取らない。そのため、現代の統計において一定の時間内に何人の客が訪れるかといったデータなどに対し用いられる。統計では、いかなる分布に従う母集団から標本を取り出しても、標本数が十分大きければ標本統計量は正規分布に近似できるという中心極限定理を利用した様々な統計分析手法がある。しかし、これには大きな問題点がある。それは上記のようなデータに対して、正規分布では平均値が負の値をとることを防ぐことはできないことである。このことからポアソン分布は現代の統計において、正規分布よりも実際を反映しやすいことも多く、重要な分布であると言える。

5　ポアソン分布に着目する意義

5.1　パラメータ

一般に確率分布を特徴づける定数のことを、パラメータと呼ぶ。たとえば正規分布において、パラメータは平均・分散の２つである。この２つが決定することで分布の形が決まる。指数型分布族においてパラメータは基本的に２つである。つまり２つのパラメータの関連を理解することは、その分布の特性を理解することである。

ポアソン分布において、平均・分散はλである。パラメータが一つの値で決定されるという点において、学習者は正規分布や二項分布より分布の特性を理解しやすいと言える。λが大きくなるにつれ平均・分散も大となる。つまり平均と分散は相関関係である。またλを極限に大、つまり無限にすることで確率は 0 に収束する。

5.2　ポアソン分布の理解

前節のとおりポアソン分布の特性を理解するためには、パラメータである平均・分散を決定するλの理解が必要である。λが平均と分散を決定する定数であることを理解するためには、3.1 にて触れたとおり、$n \to \infty$つまり極限について理解することが必要である。また、極限を理解することは収束を理解することである。

6　教育方法の提案

6.1　教育の対象・プロセスの整理

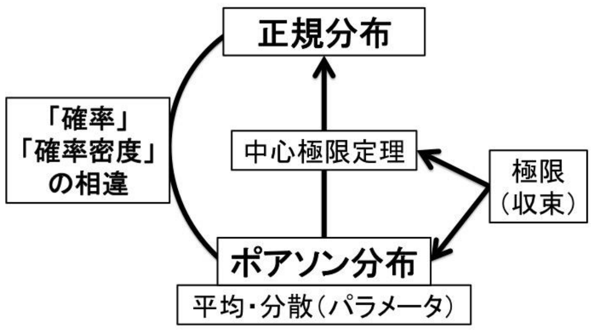

図 2　教育方法のプロセス

前節までを踏まえ、教育方法を検討していく。図 2 は、ポアソン分布から「確率」と「確率密度」の違いについて教育するプロセスをまとめたものである。教育対象は統計・確率の専門知識獲得を目指す学習者とし、平均・分散などの統計学・確率論の基本的な知識は有しているものとする。図 2 に従い教育の手順を整理すると

i. 二項分布からポアソン分布を教育する。（極限の概念の理解）
ii. ポアソン分布と極限の概念から、中心極限定理を教育する。
iii. ポアソン分布の再生性と中心極限定理から正規分布を教育する。（離散型分布と連続型分布の理解）
iv. 離散型分布と連続型分布の違いから、「確率」と「確率密度」の違いを教育する

となる。以降 6.2 でⅰ.を、6.3 でⅱ.ⅲ.を、6.4 でⅳ.についてそれぞれ述べていく。

6.2 ポアソン分布の教育

5.2 で述べたように、ポアソン分布の教育において、平均と分散という２つのパラメータがλという一つの定数で決定するという点を理解することが重要である。これを理解するためには学習者が極限の概念を理解する必要がある。その際、二項分布からポアソン分布への教育の拡張は学習者にとっても理解しやすいものと考える。

二項分布はコインの表裏やくじの当たり外れなど、学習者が確率・統計の専門的知識や高度な数学の知識をもっていない場合でも体感として理解しやすい。またポアソン分布は 4.3 にて述べたように統計において実際を反映しやすいため、実際のデータを当てはめて考えやすい。

そのため実際の事例を通して、二項分布における n が大きくなっていく極限の概念の基礎的な部分を教育することができる。同時に二項分布では膨大になってしまう演算処理を体感することもできる。そのためポ

アソン分布の統計における有用性を学習者は理解することができると考える。

ポアソン分布を教育することで、学習者は離散型の確率分布を理解することができる。また、ポアソン分布では平均と分散という２つのパラメータがλという一つの値で決定する。ここから、他の指数型分布族よりもシンプルな分布と言える。そのため、学習者はポアソン分布について数理的な操作を行う場合でも、２つのパラメータについてバラバラに扱う他の指数型分布族よりも理解しやすいといえる。

6.3　ポアソン分布を利用した確率論の教育方法

ポアソン分布の再生性から、ポアソン分布が正規分布に近似できることを導き出せる。これを利用し学習者に正規分布の教育を行う。これにより、学習者は連続型分布を理解することができる。ポアソン分布の正規近似は、以下のように証明できる。

$X_i, i = 1,2, \ldots$をポアソン分布$P(\lambda)$に従う独立な確率変数とすると、

$$S_n = X_1 + X_2 + \cdots + X_n, n = 1,2, \ldots$$

はポアソン分布$P(n\lambda)$に従う。ここで、中心極限定理より

$$\frac{s_n - n\lambda}{\sqrt{n\lambda}}$$

は標準正規分布$N(0,1)$に収束する。このことから、nが大きければ$P(n\lambda)$は正規分布$N(n\lambda, n\lambda)$で近似できることがわかる。λは任意の定数なため、$n\lambda$をλに置き換え次のように言い換えられる。

λが十分に大きいとき、ポアソン分布$P(\lambda)$は正規分布$N(\lambda, \lambda)$で近似できる。

この際、中心極限定理という重要な極限定理を扱うことになる。中心極限定理をはじめとした極限定理は、収束の概念を理解していない学習者にとっては非常に理解し難い馴染みのないものであると言える。しかし本研究の教育のプロセスでは、学習者はポアソン分布を教育行う際に極限・収束の概念を獲得している。そのため、中心極限定理を理解しやすいといえる。

6.4　「確率」と「確率密度」の相違の教育

ここまでの教育により、学習者は離散型分布と連続型分布の２つを理解することができた。これらを用いることで「確率」・「確率密度」の違いを教育することが可能であると考える。離散型分布と連続型分布の違いについて、それぞれの分布を用いて教育する。すなわち離散型分布において面積が存在しない（あるいは面積となる区間が離散している）ことを学習者に教育する。これにより「確率」が分布において面積で表されることから、学習者は離散型分布において「確率」と「確率密度」が同じ値を取ることを、感覚的に理解できると考える。

よって離散型分布と連続型分布の違いについての理解を、「確率」と「確率密度」の違いについての理解までスライドさせて教育することができる。従来の教育では、目に見えない「確率」と「確率密度」に対し学習者も高度な数学の知識を活用することで理解するというアプローチだった。しかし、この教育方法では分布という視覚による理解というアプローチをとることができる。そのため、学習者も感覚的に理解できる。

7　おわりに

本研究は、「確率」と「確率密度」の違いについて、ポアソン分布を中心とした教育方法の検討を行った。「確率」と「確率密度」について、別々ではなく関連させてその違いを教育することで、学習者は統計・確率についてより深く理解できると考え、教育方法の提案を行った。これにより従来の教育方法とは違い、高度な数学の知識について「確率」と「確率密度」の双方を用いてトップダウン式に教育することが可能であると考える。

また、ポアソン分布は現代統計においても活用されている。そのため、ポアソン回帰モデルなどを含む一般化線形モデルへの教育の拡張も期待できる。あるいは、パラメータの推定やデータのフィッティングについても、ポアソン分布は扱い易く実践統計への応用も可能であると考える。

今後は従来の教育の分類・整理について評価するとともに、この教育方法についてそれをもとにより細かな教育内容の検討を行う。

参考文献

(1)　小針アキ宏：“確率・統計入門”、岩波書店(2012)

(2)　佐伯胖・松原望：“実践としての統計学”,東京大学出版会（2000）

(3)　田栗正章：“統計学とその応用”,財団法人放送大学教育振興会（2005）

(4)　竹内啓：“統計学大辞典”,東洋経済新報社（1989）

(5)　松原望：“統計の考え方”, 団法人放送大学教育振興会（2003）.

(6)　松原望：“入門確率過程”,東京図書（2011）

(7)　盛山和夫：“統計学入門”，財団法人放送大学教育振興会（2004）.

(8)　森棟公夫：“統計学入門第２版”,新生社（2000）