



連載TOPへ

【書籍化決定！】本コンテンツの公開は2015年5月7日までとなります。

[SHARE]

Tweet

シェア

Bl

第6回

確率変数と確率分布をもって山門をくぐる

実験医学2014年10月号

はじめに

これまで2回の連載では、データのもつ“ばらつき”はどのように数値化され、同時に可視化されるかについて説明しました。

ある実験や観察を行なう際に仮定される母集団それ自体は（全数調査をしないかぎり）最後まで未知のままです。得られたデータと仮定されたモデルを手にした私たちは既知から未知への推論を行おうとしています。すなわち母集団から無作為抽出されたサンプルに基づいて、データの“平均”と“分散”に着目することにより、既知のサンプル情報から未知の母集団の属性に関する最良の推論（すなわちアブダクション）という、次に取り組むべき問題が浮上します。

しかし、参道の先によりやく見えてきたパラメトリック統計学の王国に入るには、そびえ立つ山門をくぐり抜けなければなりません。見上げれば苔むした石板には「証拠もなく言説を信ずることなかれ（Nullius in verba）」と刻まれています。ここはいたい…

§ 油断禁物・足元危険・頭上注意

さて、今回から4回にわたってパラメトリック統計学「王国」の様子に迫っていきます。“パラメトリック”という響きに、不穏なものを感じる読者もおられるでしょう。おどろおどろしい山門をくぐる前にひとつ注意事項があります。今回以降のキーワードとして確率変数（random variable）と確率分布（probabilistic distribution）という2つの言葉が頻繁に出てきます。母集団あるいは抽出されたサンプルのふるまいをモデル化するためにつくられたこの2つの概念は、パラメトリック統計学をしっかりと理解するうえでまたいで通り過ぎるわけにはいきません。

しかし、同時に一般的な統計ユーザーにとって、確率変数と確率分布は昔の嫌な記憶を思い出させる忌まわしさがまわりついていることも事実です。その理由は、統計学のカリキュラムは、たいていの場合、この確率変数と確率分布に関する数学理論からはじまるからです。実際のデータ解析の現場でそれらがどのように用いられるのかを知らないうちに、数学の小難しい理屈を叩き込まれることはさぞかし苦しい修行でしょう。以前、私も関係した農学研究者対象の統計研修会での質問票を紹介しましょう：

多くの確率分布が存在することはわかりましたが、一つひとつの分布が数式で説明されていて、なかなかイメージが湧きません。グラフや表を用いてイメージ化できないでしょうか。

この確率分布は実生活のこういう場面で使えますとか、こんな実験データに適用できますという具体的な説明ができないでしょうか。

受講生のいらだちが字面から立ち上ってくるような質問です。もともと数学がけっし

羊土社HP会員

English page

ログインしていません

羊土社HP会員とは？ ログイン

書籍検索

実験医学の定期購読

最新号がWEBでも読める！

国内送料無料

実験医学

月刊実験医学新刊

実験医学

次号予告

バックナンバー

連載一覧

掲載広告一覧

定期購読案内

詳細をみる

カートに入れる

実験医学増刊号新刊

実験医学

次号予告

バックナンバー

掲載広告一覧

定期購読案内

詳細をみる

カートに入れる

実験医学 電子バックナンバー発売中

DIGITAL ARCHIVE

新着情報 人材・セミナー案内

東京大学大学院 医学系研究科 疾患生命工学センター 分子病態医科学部門

平成28年度 大学院生（修士・博士）募集および説明会

詳細や他の情報はINFORMATIONコーナーをご覧ください

羊土社新刊・近刊

骨ペディア

サイトカイン増殖因子キーワード

Dr.北野の0から始めるシステムバイオロジー

詳細 購入

詳細 購入

詳細 購入

>>新刊一覧へ

て得意ではない彼らにとっては、数理統計学が当たり前のように用いる「数学」は時として越えられない“壁”として行く手を阻みます。

パラメトリック統計学は、いい意味でも悪い意味でも、数理統計学 (mathematical statistics) としてその学問的伝統をかたちづくってきました。現実世界を観察して、あるいは実験を通して得られたデータや知見にもとづいて、未知の物事に関する推論を行なうには、客観的かつ普遍的な「数値化」および「数学化」が不可欠であると創成期の理論統計学者たちは考えたわけです。なぜこのようなことになってしまったのか。その事情をかいま見るために統計学の歴史をさかのぼってみましょう。

§ 数値化というブッシュナイフが現実世界を切り拓く！

確率論と統計学の歴史を研究してきた科学史家セオドア・M・ポーターは、その著書『数値と客観性：科学と社会における信頼の獲得』<sup>1)</sup> のなかで、現実世界から得られた情報や知見を「数値化」することの意義を次のように述べています：

数値の力を基礎づけるのは、距離を越える技術、標準化された手続きである。それらはローカルノレッジ（注：狭い文化圏や社会の中だけで通用する知識）や信頼や知恵を前提としたものの考え方への依存度を小さくしてくれる（文献<sup>1</sup>, p.8より引用）

確率論的あるいは統計学的思考に基づく「数値化」もまた、ポーターの言うように、客観的かつ普遍的な知識体系の構築という歴史の大きな流れのなかに位置づけられるでしょう。現代に生きる私たちは、そのような定量的思考の系譜の最先端を統計学のさまざまな手法として学び、そして利用しているわけです。

一見したところ抽象的すぎる“数式”で書かれることが多いパラメトリック統計学の理論であっても、その歴史をさかのぼり、どのような状況のなかでそれが産声を上げたかを知れば、単なる無味乾燥な理屈としてではなく、現実世界を理解するため先人たちが試行錯誤して築き上げた、知的サバイバル技術の集大成であることがわかるでしょう。

本連載でこれまで説明してきたデータの“ふるまい”はまさに「数値化」の対象となりえます。私たちは観察や実験を通じて得たデータに基いて推定や推論をする際、平均や分散などの統計量を計算します。それらの数値は、単に目の前のデータを集計するだけの記述統計学的な意味にとどまらず、データが抽出された母集団に関する推測をも可能にします。

では、もとの母集団について何らかの数値化をすることはできないでしょうか。もちろん、母集団はとらえどころのない未知のものなので、確実なことは何もいえないでしょう。しかし、たとえそうであったとしても、母集団が従っているであろう規則性を仮定することはできそうです。確率論と統計学はこの問題に何世紀にもわたって取り組んできました。現代まで続く数学との密接なつながりもその長い歴史のなかで育まれてきました<sup>2)</sup>。

§ まずは二項分布と仲良くなろう

遡ること300年前、フランスの数学者ジャック・ベルヌーイ (Jacques Bernoulli, 1654-1705) の死後出版された主著『推測法 (Ars Conjectandi) 』をもって近代確率概念は確立したとされています<sup>3)</sup>。この本のなかで、彼はある頻度をともなつて生じる偶然的なできごと、すなわち事象 (event) のもつ確率とよばれる概念の数学的性質を明らかにしました。たとえばコインを投げて表が出るか裏が出るかというできごと、あるいはサイコロを投げたときにどの目が出るかというできごとは現実世界で私たちが会える確率的な事象です。

いま、ある<sup>1</sup>枚のコインをくり返し投げ上げるという事象を考えましょう。各回ごとに表の出る確率が $p$ 、裏の出る確率 $1-p$ であると仮定します ( $0 \leq p \leq 1$ )。今、コインを $n$ 回投げて、表が $x$ 回出れば、裏は残りの $n-x$ 回出ます。それぞれの試行の間に関連性がないならば、その確率は (表が $x$ 回出る確率)  $\times$  (裏が $n-x$ 回出る確率)  $= p^x \cdot (1-p)^{n-x}$ 。さらに全 $n$ 回の何回目に出るかは組み合わせの場合の数  $nCx = n! / (x!(n-x)!)x!$  だけあります。したがって、すべての場合を集計すれば、コインを全 $n$ 回

学会売行き良好書情報

● 第37回 日本分子生物学会 年会 (14/12/02)

>>過去の売行き情報はこちら

実験医学 550号 突破!

アンケートに答えて

ゴゴカレーを

多数のご回答ありがとうございました

実験医学 @Yodosha\_EM on Twitter

実験医学 jikkenigaku on Facebook

教科書・サブテキスト

をお探しの方へ

臨床医学系書籍 TOPページ(総合)

プライマリケアと救急を中心とした総合誌

レジデントノート

月刊 増刊

投げたうち表が出る回数 $x$ は次のように計算できます：

$$\begin{aligned} & (\text{全}n\text{回中}x\text{回表が出る場合の数}) \times (\text{表が}x\text{回出る確率}) \times (\text{裏が}n-x\text{回出る確率}) \\ &= nC_x \cdot p^x \cdot (1-p)^{n-x} \quad (1) \end{aligned}$$

次の図1は確率 $p=0.5$ と固定して、回数 $n$ を1, 5, 10, 20の4通りに設定したときの確率をグラフ化したものです。たとえば、 $n=1$ の場合は、コインを1回投げるだけなので、事象は「表が出る ( $x=1$ ) 」と「裏が出る ( $x=0$ ) 」の2つしかありません。それぞれの事象が生じる確率は等しく0.5となります。 $n=5$ ならば「すべて表が出る ( $x=5$ ) 」から「すべて裏が出る ( $x=0$ ) 」までの事象  $x=5, 4, 3, 2, 1, 0$  のそれぞれに対して確率の値 0.03125, 0.15625, 0.31250, 0.31250, 0.15625, 0.03125 が計算できます。

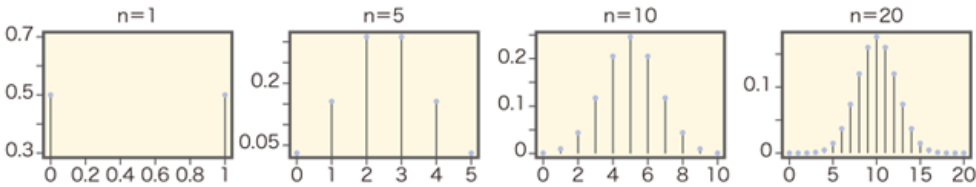


図1 表と裏の出る確率がそれぞれ0.5であるコインを $n$ 回 ( $n=1, 5, 10, 20$ ) 投げたときに表が出る回数 $x$ の確率を図示したグラフ

## § お守りの正体

このように、事象が偶然的に生じる可能性を「確率」として数値化することにより、私たちは不確かなできごとがもつ規則性に関する知見を得ることが出来ます。コイン投げ試行で表の出る回数のような、ある確率をとまって生じる変数をこれからは確率変数または変量とよぶことにしましょう。確率変数はそれぞれの値にある事象が生じる確率を対応させる規則をもっています。この規則のことを確率分布 (probabilistic distribution) とよびます。

数式 (1) と図1のグラフはコイン投げ試行における確率分布を示しています。 $n=1$ の場合の確率分布は発見者の名にちなんでベルヌーイ分布 (Bernoulli distribution) と名付けられていますが、一般の $n$ に対応する確率分布は二項分布 (binomial distribution) という名称が広く用いられています。

今回は、パラメトリック統計学の門をくぐるための“お守り”として確率変数と確率分布の2つをお渡ししました。確率変数と確率分布は母集団のもつ偶然的なふるまいをモデル化する手段と見なすことができます。まだ道のりは長いですが、ゆっくりしっかりと歩いていきましょう。その先に見えるのはもっと広大な確率分布の世界です。



## 文献

- 1) Theodore M Porter : Trust in Numbers: The Pursuit of Objectivity in Science and Public Life. Princeton University Press, 1996  
『数値と客観性：科学と社会における信頼の獲得』（セオドア・M・ポーター／著 藤垣裕子／訳）みすず書房，2013
- 2) Stephen M Stigler : The History of Statistics: The Measurement of Uncertainty before 1900. Harvard University Press, 1986
- 3) Ian Hacking : The Emergence of Probability: A Philosophical Study of Early Ideas about Probability, Induction and Statistical Inference, Second Edition, 2006  
『確率の出現』（イアン・ハッキング／著 広田すみれ，森元良太／訳），慶應義塾大学出版会，2013

[SHARE] [Tweet](#) [シェア](#) [B!](#)



[TOP](#)

「第7回 正規分布という王様が誕生する」は、本誌[2014年11月号](#)を御覧ください

本記事の掲載号



実験医学 2014年10月号 Vol.32 No.16  
**幹細胞ニッチが見えてきた！**

長澤丘司／企画

定価 2,000円＋税， 2014年9月発行

[▶詳細](#) [▶購入](#)

本連載に関する質問・感想、統計に関する具体的な悩みを編集部までお寄せください！

- 下記画像中の英数字をご入力ください



[画像を変更する](#)

おすすめ書籍



[▶詳細](#) [▶購入](#)



[▶詳細](#) [▶購入](#)



[▶詳細](#) [▶購入](#)



[▶詳細](#) [▶購入](#)



[▶詳細](#) [▶購入](#)

[会社案内](#) | [採用情報](#) | [個人情報取扱い](#) | [お問い合わせ](#) | [広告掲載について](#)

(C)2014 [YODOSHA CO., LTD.](#) All Rights Reserved.