



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](#) [B!](#)

第8回 ピアソンが築いたパラメトリック統計学の礎石

実験医学2014年12月号

はじめに

前回 は、正規分布という、ある確率分布のルーツをたどることを通して、現実世界での確率的に生じる不確定事象がどのように数値化・数理化されてきたのかを振り返りました。そのなかで、パラメトリック統計学の共通言語である確率変数は、私たちの日常生活にみられる具体的な出来事を数理の目でモデル化するために編み出された考え方であるご紹介しました。

確率変数や確率分布は、統計学を学ぶ初学者にとって最初の関門かもしれません。誰だって勉強しはじめるなり数式の洗礼を浴びるのはごめん こうむりたいはずです。しかし、前回話したように、確率論や統計学の数学理論は他ならない現実の日常世界から生まれたことを思い出してください。現在の数理統計学がどれほどいかつい顔つきで私たちに迫ってきたとしても、元をたどれば身の回りでごく普通に起きている出来事への素朴な関心が出発点であることに 違いありません。

今回の物語は19世紀のヴィクトリア朝ロンドンが舞台です。その主役は稀代の統計学者カール・ピアソンです。相手かまわず学問論争をふっかける強気なピアソンはいたるところに敵がいました。しかし、正規分布に代表される確率分布の世界を見わたした、20世紀のパラメトリック統計学が築き上げることになる城の礎石を敷いたというピアソンの業績は、いくら強調してもし過ぎることはありません。

§ ピアソンは現実の世界から出発した

ピアソンはどういった研究をしてきた人だったんですか？

1894年、ロンドン王立協会理学紀要に出版されたピアソンの論文「進化の数学理論への貢献¹⁾」を例にとって説明しましょう。この論文には、ピアソンの兄弟弟子であるウォルター・F・R・ウェルドン (Walter F. R. Weldon ; 1860-1906) との共同研究による数多くのデータが使われています。自然界の生物に関する観察データに対して、数理統計学のアプローチがいかに効果的にあてはまるかを具体的かつ詳細に論じている点では、この論文はいま読んで印象的な内容をもっています。

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

月刊実験医学新刊

- 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

- 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

広島大学大学院「放射線災害復興を推進するフェニックスリーダー育成プログラム」

平成27年10月入学者募集のお知らせ

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

[詳細](#) [購入](#) [詳細](#) [購入](#) [詳細](#) [購入](#)

>>新刊一覧へ

学会売行き良好書情報

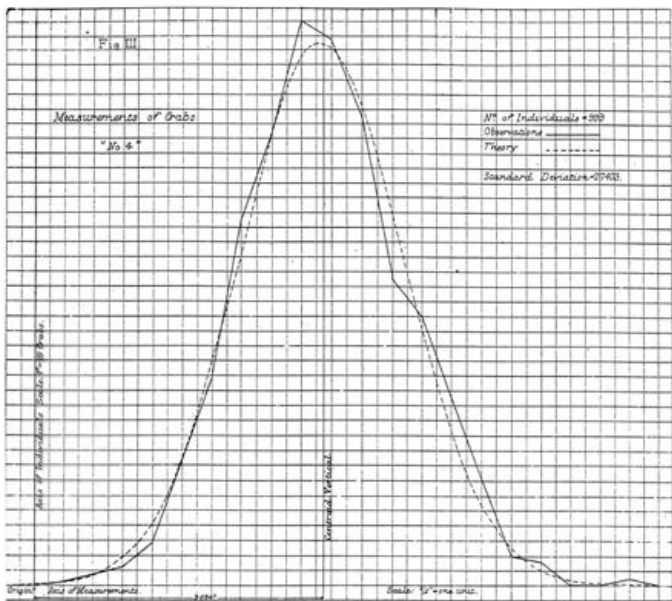


図1 ウェルドンの観測データにピアソンが正規分布曲線をあてはめた例
文献1より転載。

図1はこの論文に添付された「図版III」です。ピアソンは、ウェルドンがイタリアのナポリに生息するあるカニの個体群からサンプリングした999個体のデータを用いて解析を進めました。図の横軸はカニの甲羅サイズ、縦軸はその頻度をあらわしています。実線の折れ線で表示されているのは観察されたデータのヒストグラムです。このヒストグラムと重なるように破線の曲線が描かれています。この曲線は観察データから計算された正規分布曲線です。

ピアソンが示したのは、このカニのデータに対しては、正規分布曲線をうまく当てはめることができるという点でした。彼は他にも自然界や人間社会で観察されるさまざまなデータを取り上げ、それらを確率分布曲線によってどのように近似すればいいのかという問題を論じました。

§ 正規分布を解剖する一パラメーターとは何か

ピアソンが図1でデータの近似式として用いた正規分布曲線の一般形は次の関数によって与えられます：

$$\frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

この関数は自然対数の底「e」に関する指数関数として定義され、平均（mean）μと分散（variance）σ²という2つのパラメーター（parameter）をもちます。分散の平方根σは標準偏差（standard deviation）とよべれます。ここでいうパラメーターとは確率分布の形を決める定数という意味です。平均μは確率分布の「位置」を決定し、分散σ²あるいは標準偏差σは確率分布の「広がり」を決めています。

サンプルされた標本のデータからいかにして平均と分散を計算するかは前回までに詳しく説明しました。これに対して、正規分布関数に含まれる平均と分散は、母集団に関する未知のパラメーターを意味します。推測統計学の観点からいえば、母集団が正規分布に従うと仮定したとき、平均と分散という未知パラメーターをその母集団からサンプルされた標本によって推定するということになります。データから計算された算術平均値は母集団の平均μの推定値であり、同様にデータから計算された分散値すなわち平方和／自由度は母集団の分散σ²の推定値ということになります。数学的にこれらのパラメーターを定義することができます。確率分布の平均とは、確率変数がどれくらいの値をとるかの期待値（expectation）と定義され、確率変数の値xにその確率密度f（x）を乗じて全定義域にわたって積分した値です。また、分散σ²は確率変数のも

- 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
ポイントカードを
多数のご回答ありがとうございました

実験医学 @Yodsha_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

つ偏差平方 ($x-\mu$) の期待値として定義され、平均と同じく偏差平方を全定義域にわたって積分した値です。

もっと直観的にパラメーターの意味を知りたいのです。

たしかに、このように数式をいくら並べ立てても確率変数や確率分布の具体的なイメージはなかなか湧いてきません。そこで、正規分布の2つのパラメーターを変化させるとどのように見えるかをヴィジュアルに示しましょう。図2Aは標準偏差を0.5に固定し、平均だけを0.0から2.0まで0.5刻みに変化させたときの正規分布曲線の様子です。平均というパラメーターを変化させると、曲線の山の「位置」は左右に移動しますが、山の「かたち」そのものは変わりません。次の図2Bは平均を0.0に固定し、今度は標準偏差だけを0.5から2.5まで0.5刻みで変化させます。標準偏差が小さい値のときは平均を中心として尖った分布形状になりますが、標準偏差が大きくなるにしたがって裾野がなだらかに広がる分布形状になります。分散あるいは標準偏差は確率変数が平均からどのくらい遠くまでばらつくかの尺度にほかなりませんので、その値が小さければ平均値のごく近くの狭い範囲に高い確率で集中するために分布形状は尖り、逆に大きくなるほど平均から遠く離れた値でもそれなりの大きな確率で生じるために分布形状はなだらかになると考えればわかりやすいでしょう。つまり、分散というパラメーターを変化させると、正規分布曲線の山の「位置」は変わりませんが、その「かたち」は変化するということになります。

§ 良き統計ユーザーであるために

さて、正規分布の数学的性質はこのようにいくらでも詳細に説明することができます。しかし、読者の関心はこの正規分布を現実のデータに適用することによってどのような利点があるのかにあるでしょう。ピアソンが強調したのはまさにそこでした。彼は正規分布の確率密度関数がきれいに当てはまる実例をいくつも挙げることで、現実の生物現象にみられるデータのばらつき（ここでは生物個体群の形態変異）が正規分布という数式によってうまく近似できることを読者に示しました。

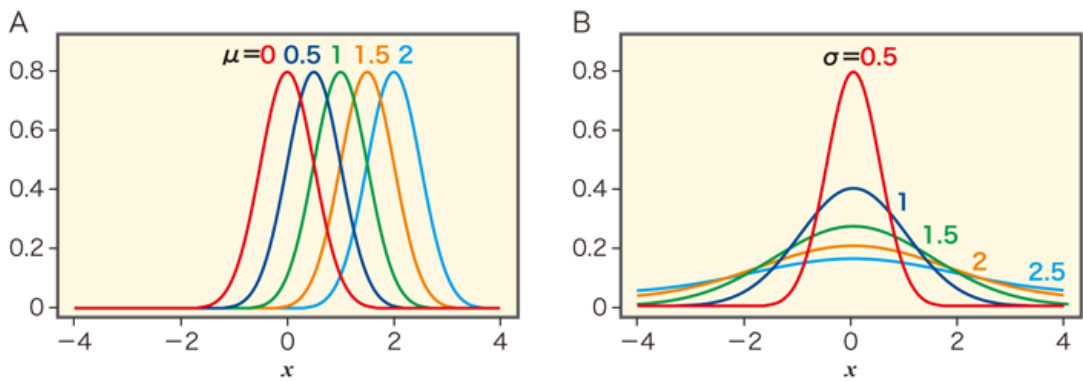


図2 正規分布曲線の位置／形状変化とパラメーターの関係

母集団をある確率分布によってモデル化するとき、確率分布曲線を定めるパラメーターが重要な役割を果たします。このように、パラメーターを含むモデルを立てることにより母集団を数値化し抽出された標本に基づく推論を行なう統計学の立場がパラメトリック統計学 (parametric statistics) にほかなりません。

数学としてのパラメトリック統計学には厳密な形式化と過度の抽象化という傾向が内在する点是否定できません。医学・生物学系あるいは農学系の統計ユーザーはときにそれが苦痛になることもあるでしょう。しかし、確率変数や確率分布に関する数学理論は、現実世界の母集団をいかにきちんと記述できるか、観察されたデータのふるまいをどれほど正確にモデル化できるかを念頭に置いて発展してきた点も同時に強調すべきでしょう。現代の数理統計学の礎を築いたカール・ピアソンは少なくともそういう姿勢で研究をしてきたからです。私たち統計ユーザーは現実世界の具体的な応用とそのデータに足場を置き続けるべきです。そのうえで、パラメトリック統計学が提供するツールをいつどのように使うべきかあるいは使わざるべきかはユーザーの賢明な判断に委ねられています。

パラメトリック統計学の世界には数々の確率分布が存在します。正規分布はその一つに過ぎません。では、なぜ正規分布はこの世界の中で大きな顔をしているのでしょうか。正規分布は単なるはだかの王様にすぎないのでしょうか。それとも正規分布は誰もがひれ伏す実力をもっているのでしょうか。次回の「確率分布曼荼羅」では、それについて話をしましょう。

文献

- 1) Pearson K：Phil R Soc Lond A, 185：71-110, 1894
- 2) Theodore M Porter, Karl Pearson：The Scientific Life in a Statistical Age. Princeton University Press, 2004

[SHARE]  Tweet  シェア 

[Prev](#) [8](#) [Next](#)

[TOP](#)

「第9回 確率分布曼荼羅」は、本誌[2015年1月号](#)を御覧ください

本記事の掲載号



実験医学 2014年12月号 Vol.32 No.19
原因か？結果か？がんのエピゲノム異常

鈴木 拓、今井浩三／企画
定価 2,000円＋税、2014年11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](#)

[▶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.