

- ・シンプソンのパラドックス
- ・コインが公平だという帰無仮説が 5%の有意水準で棄却されたことは、「95%の確率で公平でない」という意味ではありません。
- ・信頼区間とは信頼係数の表す確率で真の値がそこに含まれる，という意味ではありません。
- ・全体で値が一番大きいところに到達する前に，その近所でだけ値が一番大きい間違っただけの解（これを専門家は「局所最適解」）

私は大学院生の頃，家庭教師のアルバイトで高校生に数学を教えていました。彼女は確率が特に苦手で，「同様に確からしい」と言うべきところでいつも，「どうやら確からしい」と言ってしまっていたのでした。

◆ベイズの公式

定義 1.1 ((初等的な)条件つき確率) 確率が正の二つの事象 A , B に対し, B を条件づけたときの A の条件つき確率 $P(A|B)$ を以下で定義する.

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

つまり, 事象 B の範囲内で事象 A が占める確率の割合です.

定理 1.1 (ベイズの公式)

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}.$$

条件つき確率 $P(A|B)$ に対し, その「逆の」条件つき確率 $P(B|A)$ が与えられることがポイントです. この公式を「今, $P(B)$ だとしている B の確率を, 情報 A を知ったことで, $P(B|A)$ に更新する手続き」だと読むのが, ベイズ推定の心です. 上式を一歩進めた以下の形もベイズの公式と呼ばれます.

定理 1.2 (ベイズの公式 2) 事象 A と, 全事象の分割であるような事象 H_1, \dots, H_n に対し,

$$P(H_i|A) = \frac{P(A|H_i)P(H_i)}{\sum_{j=1}^n P(A|H_j)P(H_j)}.$$

つまり, 各仮説 H_i のもとで観察結果 A が起きる確率を用いて, その逆に, 観察結果のもとで各仮説が正しい確率を計算します (第六原理).

では、コルモゴロフによる確率の定義を見てみましょう。皆さんが測度やルベーグ積分についてご存じなくても大丈夫です。まで、一つだけ準備として σ -加法族を定義し、続けて確率の定義を紹介します。

定義 3.1 (σ -加法族) 空でない集合 Ω の部分集合の集合 \mathcal{F} が以下の三つの性質を持つとき、 Ω 上の σ -加法族と言う。

- (1) $\Omega \in \mathcal{F}$,
- (2) $A \in \mathcal{F}$ ならば補集合 $\Omega \setminus A \in \mathcal{F}$,
- (3) 任意の可算個の $A_1, A_2, \dots \in \mathcal{F}$ に対し $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

ここで可算個とは、 $1, 2, 3, \dots$ と番号をつけられる程度の無限のことです。この限定が重要なポイントになります。

F の花文字。これで門前払い。わかる奴だけわかればいい世界。

~~~~~

$\mathcal{F}_t$  は、 $\forall \mathcal{F}_t$  とすることで出力可能

<https://mathlandscape.com/latex-mathcal/>

~~~~~

しかし、数学は「慣れが理解」の面がありますね。だんだんとそういうものかと感じられるようになり、またそこが醍醐味でもあり、腕の見せ所だと思えるようになりました。

確率とは何なのか、ランダムであるとはどういうことか、まだわからないこと、数学に落とし込めていないことがたくさんあります。コレクティヴ理論は本当に失敗作だったのでしょうか。コレクティヴの可能性はまだ汲み尽せてはいないのかも。そんなふうに感じるのも、天才の失敗の所以でしょうか。

◆コルモゴロフの複雑度

定理 5.1 (万能アルゴリズムの存在) 任意の部分帰納的関数 $A: \{0, 1\}^* \rightarrow \{0, 1\}^*$ に対し, ある定数 $c(A, A_0) \in \mathbb{N}$ を用いて

$$K_{A_0}(s) \leq K_A(s) + c(A, A_0), \quad s \in \{0, 1\}^*$$

となるような, (A によらない)部分帰納的関数 A_0 が存在する.

この A_0 を「万能アルゴリズム」と呼びます. この A_0 を用いて, いよいよコルモゴロフの複雑度を定義します.

定義 5.2 (コルモゴロフの複雑度) 万能アルゴリズムを一つ固定して A_0 と書く⁸⁾. このとき, $K(s) = K_{A_0}(s)$ のことを $s \in \{0, 1\}^*$ の(コルモゴロフ)複雑度と呼ぶ.

ただし, 具体的にある 0, 1 列を与えられて, そのコルモゴロフ複雑度を計算してみろ, と言われると困ります. 実際, よほど単純な場合に粗い評価を与えるのがせいぜいでしょう.

確率の三つの姿, 可能性, 複雑度, 戦略はこのように密接に関係しています. 他にもいろいろな見方がありますが, 一つの一般的な捉え方は, マルティン=レーフによるアルゴリズム的ランダムネスの定義を中心に, 他の概念との同値性や強弱を調べるというものです.

とは言え, ランダムネスとは何かという全体像と, 各側面の間関係には, わかっていないことがたくさんあります. また, これらのさまざまな顔を究極的に統一するような本質があるのでしょうか.

確率やランダムネスは, 一つの枠に収まり切らない多くの謎と, 大きな謎を抱えています. 分野を超えた研究は敷居が高くなりがちですが, こういう大きな問題意識を持ちながら, 確率の不思議な世界に挑戦してくれる若い方が増えると良いなあと思います.

◆主観確率と統計的推測

特に重要な応用は, ベイズ推定の首尾一貫した枠組みを提供できることです. つまり, ベイズ推定とはある人の主観をデータで更新していく手続きであり, ベイズの公式などの理論的根拠は一貫性の条件で与えることができます.

...

とは言え, 統計的推測の根拠を巡る議論は今も続いている, これからも決着しないでしょう. 統計的推測にはさまざまな立場が

あって、何を支持して勉強したら良いのかわからない、という人に対して、ある識者は以下のようにアドバイスしていますが 13)、私も現状、これに賛成したいと思います。

すなわち、現場の統計学者は四つの主要な理論に通じておくべきである。その四つとは、フィッシャー、ネイマン&ピアソン、ジェフリーズ、そしてデ・フィネッティ。

13) G・A・Barnard “Fragments of a statistical autobiography”, Student, pp. 257–268 (1996).

定理がわからない。

定理 7.1 (ブラウン運動の存在) 確率空間 (Ω, \mathcal{F}, P) で定義された実数値確率変数 $\{B_t(\omega)\} (t \in [0, \infty), \omega \in \Omega)$ であって、任意の自然数 n と $0 \leq t_1 < t_2 < \dots < t_n$ について $B_{t_1}, B_{t_2} - B_{t_1}, \dots, B_{t_n} - B_{t_{n-1}}$ が互いに独立で、任意の $0 < s < t$ に対し $B_t - B_s$ が平均 0 で分散 $t-s$ のガウス分布に従い、しかも、 $\omega \in \Omega$ ごとに B_t が t の連続関数であるものが存在する。

この定義(定理)から、ブラウン運動に関するいろいろな量や性質がガウス分布の計算で求められます。例えば、酔歩でも見た位置とその 2 乗の期待値は(原点出発 $B_0 = 0$ のとき)、 $E[B_t] = 0$, $E[B_t^2] = t$ と計算できます(演習問題)。

著者が面白いと思い、確率論の勉強をするきっかけになった式
ブラウン運動の経路の微分不可能性証明

$$\bigcup_{l \geq 1} \bigcup_{m \geq 1} \bigcap_{n \geq m} \bigcup_{0 < t \leq n+2} \bigcap_{t < h \leq t+3} \left\{ \omega \in \Omega : \left| \frac{B_h(\omega)}{n} - \frac{B_{h-1}(\omega)}{n} \right| < \frac{7l}{n} \right\}.$$

どうです？ 係数 5 と 7 が現れる理由はさておき、当時のウブな私はこの数式の姿にショックを受けると同時に、「おもしろいなあ!」と思ってしまったのです。やや複雑な「任意の」と「存在する」の関係を集合で書いただけなのですが。

こんなことに魅力を感じて、確率論を勉強してみようと思ったなんて、まさに若気の至りでお恥ずかしい限りです。邪道であることは間違いありませんが、意外にこんなことで進む分野が決まってしまうものなのかもしれません。

私の思い出話ばかりで、確率論の魅力を伝えられたか不安ですが、他の分野とはちょっと違うところや、他の分野との密接な関係を、少しは説明できたでしょうか。皆さんの中から、偶然の数学の面白さと不思議さを共有してくださる方が現れば、嬉しく思います。

コイン投げの結果は、PQ上を動くだけ。2乗和で考えると円周を動くことになる。発想が面白い。

これをグラフで図示しますと、以下図9.1の線分PQの上にコイン投げの確率の世界がすべて表現されていることになります。

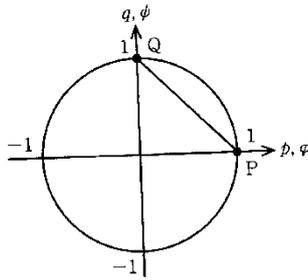


図9.1 確率とその平方根

これはどうも窮屈に見えます。そこで、マイナスかけるマイナスがプラスであることを思い出して、二乗すると確率になるものも考えてみましょう。

つまり、このコインは「内部状態」として、実数 φ と ψ を持っています¹⁾。そして、この二乗 $p = \varphi^2$ 、 $q = \psi^2$ が表と裏が出る確率 p, q であるようにふるまうのだ、と考えます。

すると、これら (φ, ψ) は $\varphi^2 + \psi^2 = 1$ を満たすもの全体ですから、確率の背後にある状態の世界は図9.1の円周になります。確率と違って、状態は負の値もとることに注意してください。

円は線分よりずっと対称性が高いので、なにかとても良いことが起こりそうな予感があります。このコインを「魔法のコイン」と呼ぶことにしましょう。

私の専門は確率論でしたが、やはり和風なのか、正直に言って、統計学をよく知らないことを白状します。どのくらいかと言うと、あなたは確率論の専門家なのだからと「確率・統計」や「数理統計」の講義を担当させられる羽目になるたび、入門書を読み直していたほどです（これは日本の確率論の専門家にはしょっちゅう起こる事態だと推察します）。

ちなみに、こういうときには同じ確率論分野の著者の「確率・統計」の教科書で勉強するのがコツで¹⁾、統計学の専門家が書いた本だと私の頭に入らないのですね。言葉遣いやセンスの違いでピンとこない。それくらい日本では確率論と統計学のコミュニティが遠いから……というわけでもないでしょうが。

とは言え、確率論と統計学が遠い関係で良いはずがないことは確かでしょう。また、確率論が数学で、統計学はその応用、という決めつけで良いわけもない。二つの分野は同じ知的関心から生まれ、互いに影響を与えあいながら育った兄弟なのですから。

1) 著者（私）のお気に入りには、『統計と確率の基礎』（服部哲弥，学術図書出版社）と『確率の基礎から統計へ』（吉田伸生，遊星社）。

シンプソンのパラドックス

10.3 シンプソンのパラドックス

われわれは確率論や統計学における「割合(比率)」という概念の価値を侮りすぎているのではないか、と思うことがあります。そもそも確率自体が割合ですし、平均値、期待値、スケーリング等々、確率と統計のほとんどすべての問題が、この概念を中心に回っていると言えるのではないのでしょうか。

このことを忘れて、つい軽視してしまうのは、それがただの割り算にすぎないからでしょう。しかし意外にも、統計においては割合がトリッキーなこともある、という例を挙げてみたいと思います。

以下の問題を考えてみてください。ちなみに「ブジョネ」とは、コルク栓の腐敗によるワインの不良品のことです。

問題 10.1 ワイン販売店を営むあなたは、二つの輸入業者 A 社と B 社からワインを仕入れている。赤ワインについては、A 社から仕入れたワインの方が B 社よりブジョネ率(ブジョネ本数/仕入れ本数)が低く、白ワインについても同じく、A 社の方がブジョネ率が低いことがわかった。

このことから、赤白あわせたワイン全体でも A 社の方がブジョネ率の低い、信頼できる業者であると考えすることは正しいか？

ほとんどの読者は、「その通り、もちろん A 社の方がブジョネ率が低い」とお考えになるのではないのでしょうか。実はこの推論は間違いです。私もこの「シンプソンのパラドックス」⁴⁾を理屈の上では理解していますが、どうしても正しく思えて仕方があり

ません。

実際に具体的な数字で、上の推論が成り立っていない例を挙げましょう(表 10.1)⁵⁾。各数字はブジョネ率にしては高すぎますが⁶⁾、本質とは無関係ですのでお気になさらず。

表 10.1 各輸入業者のブジョネ率(ブジョネ本数/総本数)

輸入業者	A 社	B 社	合計
赤ワイン	15/100	20/100	35/200
白ワイン	100/400	3/10	103/410
合計	115/500	23/110	138/610

この表で、赤ワインについては、15% 対 20% で A 社の方がブジョネ率が低く、白ワインについても $100/400 = 25\%$ 対 $3/10 = 30\%$ ですから同様です。しかし、赤白ワインを合計してみると、A 社 $115/500 = 23\%$ に対し B 社 $23/110 = \text{約 } 20.9\%$ となって、A 社の方がブジョネ率が高くなっています。

割合は特定のサンプルの特徴の指標にすぎませんが、平均の考え方は、全体の特徴への推論につながるからです。統計学では、この「全体」を母集団、**推測したいその特徴を母数**と言います。つまり、母集団全体を調べる代わりに、一部のサンプル（標本）の比率や割合だけを見ることで、**母数を推論することが統計学の基本的問題**です。

棄却するためのこの仮説を「帰無仮説」、この判断基準を「有意水準」と言います。この仮説は、あくまで捨てることが目的、捨てられてこそ意味をなす、**ギャンビット**（捨て駒）であることがポイントです。

~~~~~

#### gambit

(ポーンなどを捨てごまにする)さし初めの手、(交渉・会話などで、先まで計算したうえでの)手初め、切り出し

~~~~~

さて、実際にコインを10回投げてみたところ、表が9回出ました。帰無仮説、つまり $q=1/2$ の仮定のもと、現実には得られたデータが得られる確率を計算します。ちなみに、この値を「p-値」と呼びます。確率を計算すべき出来事は、上で述べた注意点より、「9回以上表が出るか、裏が出る」です。

このコインが公平ならば、10回中10回表が出る確率は $1/1024=0.000976\dots$ 、9回表が出る確率は $10/1024=0.00976\dots$ です。よって、表か裏が9回以上出る確率は、 $(1+10+1+10)/1024=22/1024=0.0214\dots$ になります。

この約2%という確率は有意水準5%以下なので、この実験結果は「統計的に有意」とであると判断し、仮説を棄却します。すなわち、「このコインは公平ではないだろう」が結論です。

一方、もし表が7回出るなどして、有意水準以下のp-値が得られなかった場合は、コインが公平でもたまたま起こりうる程度の出来事だと判断して、「このコインが公平でないとは（この証拠からは）言えない」という結論になります。

このように仮説検定のアイデア自体はやさしく、手続きもさほど難しくは見えません。しかし、**仮説検定から得られる結論は誤解されがち**です。

例えば、コインが公平だという帰無仮説が5%の有意水準で棄却されたことは、「**95%の確率で公平でない**」という意味ではあ

りません。もちろん、棄却できなかつた場合も、「95%の確率で公平だ」という意味ではありません。

11.4 区間推定の難しさ

区間推定も結構簡単だな、と思われたのではないでしょうか。実際、本質的には仮説検定の考え方をもう少し定量的にしたにすぎません。しかし、やはりこれはかなりトリッキーです。

まず第一に、**信頼区間とは信頼係数の表す確率で真の値がそこに含まれる、という意味ではありません。**仮説検定の場合とまったく同様に、信頼係数は出来事の確率に対する基準であって、 q の確率に対するものではないのですから。

この間違いは非常に多く、区間推定を日常的に用いている科学者すら、そう勘違いしている場合があります。おそらく、議論のトリッキーさの他に、私は「信頼区間」や「信頼係数」の用語、特に前者が誤解を招く一因ではないかと思っていますが、広く用いられてしまっているのでやむをえません。

第二の難しさは、この考え方を直接用いるのは煩雑なため、確率論を用いて便利で強力なメソッドを作ることが、頻度主義的または客観的な確率を用いる伝統的な（フィッシャー-ピアソン流の）統計学の主題であることです。

この理由で、ほとんどの統計学の教科書では、推定したい量が正規分布（ガウス分布）に従っていると仮定して、その推定手法を議論することを中心としています。正規分布は期待値と分散だけで決まりますから、この二つのパラメータを推定する問題です。

例えば、正規分布の期待値だけが未知である場合は、同じ形の正規分布が平行移動するだけなので、非常にすっきりした公式が書き下せます。また期待値と分散の両方が未知でも、それら未知の値が現れない確率分布が導けて、推測に用いることができます。

このように、**正規分布がとにかく数学的に扱いやすく、美しい構造と高い対称性を持つ**ため、これを縦横無尽に用いて精微な理論と、標準的な手続きと、便利な公式群が作り上げられているわけです。

これらは使うのが簡単であっても、その本質的な推理法を正規分布の特別な性質と複雑な理論とで覆ってしまうため、意味が誤解されやすく、間違った適用もされやすい原因になります。

また、そもそも**正規分布の仮定が適切か、という本質的な問題もあります。**この仮定がもっともらしい理由は、「どんな分布に従う量でも、その十分大きい標本の誤差の平均は、適当にスケールリングすれば正規分布にほぼ従う」という**中心極限定理**です。

しかし、この素晴らしい普遍性はあくまで数学の定理であって、

それが応用的に正しい意味を持つかは別問題です。実際、数学的に扱いやすい、というだけの理由から、あまりに広く正規分布を仮定してしまっているのが実情でしょう。

11.5 検定／推定理論の罪

以上のような誤解されやすさは、主観的な見積もりである確率を証拠でアップデートしていく、というベイズ推定の自然さに比べると、一際目立ちます。昨今では、**ベイズ推定の復権**と流行に歩調をあわせるように、この伝統的な統計手法である**検定／推定理論を攻撃する傾向**にあります。

実際、心理学の権威ある論文雑誌が、**仮説検定を禁止**する方針を公開して、大きな議論を呼ぶという事件がありました。また、統計学の専門家たちからも、科学論文においてこれらの手法が正しく機能していない、という批判が続出しています。

たしかに、多くの実験科学の論文では、**仮説検定の有意水準を特に理由もなく仮定**しています（多くの場合、自動的に5%）。これだけが「再現性の危機」の原因ではありませんが、このナイーブすぎる態度が要因の一つであることは間違いないでしょう。

実際、有意水準5%で統計的に有意なp値を得ることは、多くの場合にあまりにも容易です。実験結果を得てから有意水準をずらすようなインチキはさておき、**実験がうまく行ったときだけ論文にする**、というような無意識の態度からも、簡単に「統計的に有意な」結果が得られてしまいます。

また、これらの手法が単に誤解されやすかったり、正しく適用することが難しかったりするだけでなく、そもそも**統計的有意性の考え方自体**、その根本的な論理の弱さからして、主犯に名指しされても仕方がない一面もあります。

一方で、**ベイズ推定や主観確率に基づく手法にも**、理論と応用の両面に問題点があり、かつては、科学的な論拠にならないと考えられていたほどです。

マルコフ連鎖

これによって、通常の英文においては、例えば、文字“a”の直後に“b”が現れる確率 $P(a \rightarrow b)$ はいくらか、という基礎データが得られました。文章は文字が続いたものですから、文章 X を各文字の列で $X = x_1 x_2 \dots x_n$ と書きますと、この文章が現れる確率 $P(X)$ は

$$P(X) = P(x_1)P(x_1 \rightarrow x_2)P(x_2 \rightarrow x_3) \dots P(x_{n-1} \rightarrow x_n)$$

だと考えられます ($P(x_1)$ は 1 文字 x_1 の出現頻度)。

この計算は、ある文字の後にどんな文字が来るかが、前者の文字だけに依存して、しかも、純粋に確率的に決まると仮定してのことです。このような性質を確率論の専門家は「マルコフ性」、このように生成される記号の列を「マルコフ連鎖」と言います。

このように全体で値が一番大きいところに到達する前に、その近所だけで値が一番大きい間違っただけの解（これを専門家は「**局所最適解**」などと言います）に到達してしまう可能性をどう回避するか、という問題があることを覚えておいてください。

また、特に次の一步の確率を上のように決めることで局所最適解への落ち込みを避ける方法を**メトロポリス法**と言います。ちなみに、この「メトロポリス」は都市のことではなく、この手法の提唱者の人名です。ちなみに、「モンテカルロ法」というコードネームを命名したのは、このメトロポリスだそうです。