

正統派統計学の終焉

E T ジェインズ「確率理論：科学の論理」から

正統派統計学

推定値の誤差が3%以下である確率は0.95

推定値の誤差が5%以下である確率は0.99



推定値の誤差は4%

理論物理学者の使う論理としての確率理論

平成 15 年 6 月 10 日

EORC にて

宇宙開発事業団国際部

招聘開発部員 原 宣一

まえがき

宇宙システムで行われてきた信頼度計算は社会的に大きな誤解があったように思われます。この最たるものはアポロ計画と名付けられた米国の有人月面探査計画でしょう。まことしやかに「アポロ計画はシックス・ナインの信頼度で実行された」等という記述を市販の本で見ることがあります。これはもはや神話の域に達してしまった誤解としか言いようがありません。

西沢潤一教授は半導体や光ファイバーの応用研究では第一人者です。西沢教授ですら、ご自分の専門である半導体素子が今日見るような高集積度にするとは出来ないであろうと推測され、将来予測を間違ってしまったのです。現行の信頼度計算法が科学の発展の予測を狂わせた典型的な実例であると思われる。1)

現行の信頼度計算の問題点はどこに問題があったのでしょうか。大元をたどると確率に頻度概念を採用している正統派統計学に行き当たります。この正統派統計学によると、統計的推測の一般的な表現は次のようなものです。

「推定値の誤差が 3% 以下である確率は 90% である。」

神ならぬ人間は真の値を知ることができません。しかし、不確定ながらも確かさの度合いを付けて表現することはできます。この確かさの表現に使う数値が本来の確率です。材料の破壊強度（真の値）は分からないが、推定した破壊強度に確かさの度合いをつけて表現することができます。

命題「真の値は推定許容値以上である」は真であるか偽であるかのどちらか、です。そしてこの命題が真であることの確かさに対して割り当てた数値が確率なのです。この点に戻って考えてみると、正統派の持って回った表現は「誤差」とか「信頼水準」と言い分けても「確率」を 2 重に使っていることに気が付きます。

正統派統計学が採用している頻度概念の確率はその物が保有する物性値のように解釈し、それを推定することを目的とするので無理があるのです。この考え方を捨て去ることが必要です。代わりに、その物に対して人間が抱く確信の度合いを確率で表すという考え方に切り替えなければなりません。

本来は 200 年前に確立されたのに、正統派統計学に追いやられてしまった、ラプラス流の確率理論に戻すことで良いのです。実は、このことを詳細に説明している教科書兼参考書がジェインズにより書かれ、その後継者が仕上げた本があります。標題は「確率理論：科学の論理」です。しかし、この本は分厚く、難解な部分も多いものです。2)

そこで、本小冊子は「確率理論：科学の論理」の案内本であることを目指して書きました。これを読まれて、「おやっ」と思われ、オリジナルを読んで見ようと思われれば、この小冊子の役目は十分果たされたものと思います。

目次	
まえがき	2
1. なにが問題か	5
1.1 統計的推測の問題点	5
1.2 数学的確率の問題点	6
1.3 頻度確率の問題点	7
1.4 確率論における歴史的論争	9
1.5 信頼度計算の問題点	10
[演習問題 1] 真空管式コンピュータの信頼度	13
2. 確率と実世界	13
2.1 実世界の認識	13
2.2 演繹と帰納	15
2.3 拡張論理としての確率理論	16
2.4 ブール代数	17
2.5 関数方程式	19
3. 合理性と首尾一貫性	19
3.1 デシダレータ	19
3.2 積の法則	21
3.3 和の法則	24
3.4 ベイズの定理	25
4. ラプラスの確率	27
4.1 無差別の原理	27
4.2 ラプラスの確率の定義	28
[演習問題 2] 確率計算	28
[演習問題 3] 確率計算と勝ち目	29
5. 確率論の応用	30
5.1 初等標本抽出理論	30
5.2 連続則	32
6. 不確かさの測度	32
6.1 シャノンの条件と情報エントロピー	32

7 . 最大エントロピー法へ	34
7.1 不変量の原理	34
[演習問題 4] 誤差の分布形状	35
7.2 ガウスの誘導	35
7.3 最大エントロピー法	37
 8 . 信頼度計算の見直し	 37
8.1 統計データが利用できるとき	38
8.2 属性試験データ	38
8.3 パラメトリック・データ（大きな標本の場合）	38
8.4 パラメトリック・データ（小さな標本の場合）	39
 終わりに	 43
拡張論理としての確率理論	43
 文献	 44
 付録 - 演習問題と解答	 44

1. なにが問題か

1.1 統計的推測の問題点

まえがきで紹介した統計的推測の問題点を具体的な例で説明しましょう。

ある材料の強度を知りたいのだと想像してください。この強度を調べる目的で数個の試験片を作り引張り試験を実施します。首尾良く、破断応力のデータが得られたしましょう。これらデータの値とそのばらつき具合から、平均や標準偏差を求める計算を行って、前述のような結論、この場合は「破断許容応力の推定値を下回る誤差が3%以下である確率は90%である。」を得ます。

試験を行うたびにデータは異なるでしょうから、推定値はそのたびに異なるのが普通です。ばらつきの原因として、材料組成のばらつきや試験自体のばらつきもあるでしょう。従って、推定値に幅を持たせて考え、例えば97%はこの値以上であると考えられる数値を求めることにします。つまり、試験データの平均よりばらつきの標準偏差の2倍ほど小さい値を取って、誤差3%を許すと、この値以上であると推定したと表現するのです。

さらに、ここで使った推測方法は10回に9回は当たるが10回に1回は外れるので誤差が3%以下であるということは確率90%であると表現するのです。

「信頼水準90%で（破断応力がこの値以下である）誤差は3%以内である」と言う表現もありますが、これも同じ意味です。正確な表現はこのようなものになるとするのが正統派統計学です。3)

これまでの論法は、一見当然で、間違っていないように見えるでしょう。ここで使われている論法をもう一度復習します。

本当に知りたいのは真の破壊強度です。しかし材料自体がばらつくものであり、いくら幅を取って考えても絶対的なことは言えません。そこで、99%、あるいは97%、はこの値以上であるという値を知ることによって満足します。しかし、絶対的なことを言うことにこだわり、幅を広げすぎると無意味なことになります。「99.99%、つまり万に一つも間違いのない推定でいうと、・・・は0.01%以上である。」というような殆ど当たり前の表現になってしまうのです。

試験データから許容値を推定するのであれば、推定の確かさとして90%を取ると、「材料の97%は推定許容値以上である確率が90%」という表現にならざるを得ないと考えてしまうのです。何故、90%や97%なら納得できて、99.99%は過剰なのでしょうか。

現行の殆どの教科書・参考書で上述のように記載されています。従って、殆どの人が統計的表現とはこのようなものであると信じて疑わないでしょう。つまり、この論法でどこが悪いのかと思われるでしょう。

しかし、表現が不自然なものはやはりどこかおかしいのです。明らかにまずいことは次の点にあります。同じデータを使っても異なる信頼水準で推定した

ら、「推定値の誤差が 5 % 以下である確率が 95 % である」とも言えるかも知れないのです。もっと推定確度をあげて「推定値の誤差が 10 % 以下である確率が 99 % である」と言えるかも知れません。つまり、誤差と確率（信頼水準）の組み合わせは無限にあるということなのです。

信頼水準の数値を意図的に変えて結論を逆転できることが、可能な場合もあります。このことは、正統派統計学では首尾一貫性が保てないことを示しています。

それでもまだ次のように執拗に反論される方が多いでしょう。

材料自体に強度のばらつきがあるのだから、そのばらつきの中から取り出す不確定さと、測定自体のデータのばらつきに対する不確定さがあるのだから二重表現で良いのではないか。

試験の真の目的が「材料強度のばらつき」を知ることであれば、ばらつきに対して不確定さの程度を付けて表現することは当然です。しかし、目的はあくまでも材料の強度であった筈です。「材料の強度」を知ることに対して、不確定さを表す確率表現が 2 重になっていることが首尾一貫性を欠く原因です。材料のばらつきだけを考えても成分のばらつきや製造工程のばらつきなどいくらでもばらつきの要因は考えられます。より正確な表現をするためには、それぞれの不確定さを表現するために確率を使うとすると確率を何重につけても足りないことになるでしょう。

1.2 数学的確率の問題点

確率の歴史でコルモゴロフの公理による方法は必ずしも古くはありません。確率論争から少し離れた抽象の世界で数学理論として収まり、あたかも確率理論の基礎を与えているかのように見えます。このような雰囲気があることが実世界の人間にとって問題です。それを少し説明しましょう。

コルモゴロフは実世界の確率が頻度であるか、確信の度合いであるかに関係なくある性質を持ったものを確率と定義しました。ある性質とは、確率が持っているべき特質を表した 3 項目です。これを確率の公理としました。

(1) 確率 $P(A)$ は正の値または 0 である。

(2) 確実な事象の確率を 1 とする。

(3) もし、 A 、 B が相互背反（互いに素）であるならば、

$$P(A + B) = P(A) + P(B) \quad (1-1)$$

数学的確率理論は集合論の測度理論に基づいていて、数学的基盤は盤石です。しかし、公理の選択においてあまりにも抽象度を上げすぎていて、実世界から離れてしまったように思われます。つまり公理 (3) は、前提条件がついており、相互背反のときだけです。実世界では相互背反でない場合が多いのに、こ

のときのためにどうすべきかをこの公理の組では何も言っていません。これが第一の問題点です。数学的確率理論で展開された結果は如何に論理的に正しくても、それはあくまでも数学的モデルの世界であって、実世界のものからかけ離れてしまっているおそれがあります。

次に、実世界では確率を考えると、常に条件を明確にする必要がありますが、コルモゴロフはこのことを意に介していません。条件付き確率はベイズの定理を使って定義されますが、もはや実世界にうまく対応していません。

これらから、コルモゴロフ的確率理論の展開は数学的抽象理論としては欠陥がないのですが、現実と合っているか問題です。公理に基づく確率を使用して、確率モデルはいくらでも作れますが、それが実世界にどこまで合っているかが問題になってくるのです。

数学的確率の問題点

コルモゴロフの公理

事象Aの確率はこの事象に割り当てた数値 $P(A)$ で次の3つの公理を満たすものである。

$$P(A) \geq 0$$

1) $P(A)$ は正。

$$P(S) = 1$$

2) 確実な事象の確率は1。

3) もしA、Bが相互背反（互いに素）ならば、

$$P(A + B) = P(A) + P(B)$$

数学者は測度論に基づいた公理から確率論を展開している。

数学的基礎は確立されている。 $P(A)$ は抽象的对象で、

数学者は実世界でそれがどのように使われるか関心がない。

1.3 頻度確率の問題点

フォン・ミーゼスは確率の定義を次のように相対頻度の極限值としました。

意図する実験がN回行われるものとし、事象Aが N_a 回起こったとします。このとき、事象Aが起こる確率 $P(A)$ を次式で定義します。

$$P(A) = \lim_{N \rightarrow \infty} \frac{N_a}{N} \quad (1-2)$$

つまり、さいころで1が出る確率ならば、何度かそのさいころを振って見て、1が出た回数と振った回数の比を取り、その振った回数を無限に大きくした極限值とするものです。

この定義が実世界を離れて無理なことを要請していることは、無限の操作が出来ないことからすぐ分かります。しかし、十分試行回数が大きければその相

対頻度はある値に収束するであろうことは容易に想像がつきます。実際、正しく作られたさいころを多数回振ると1が出るのはおおむねその1 / 6であることはしばしば観察されることです。

頻度確率の問題点

フォン・ミーゼスの定義

意図する実験がN回行われるものとし、事象AがNa回起こったとする。事象Aが起こる確率P(A)は次式で定義する。

相対頻度の極限值
$$P(A) = \lim_{N \rightarrow \infty} \frac{Na}{N}$$

工学分野（信頼性工学）で採用（判りやすい）

確率の公理を満たす

大数の法則が成立するときには實際上、問題なし

しかし、定義に極限值を取ることで実世界から離れてしまっています。数学的にも相対頻度の極限值が存在することが証明できるわけでもありません。そして、実世界で1回しか生じない対象についてはこの定義ではまったく使えません。これを無理に多くの回数を想定して使うことから不首尾一貫性が生じたり、不合理な判断を招いたりしているのです。

頻度確率の問題点

ベイズの定理を条件確率の定義に使う

極限值が存在するという保証はない

P(A)を推測する以外に知る方法はない（MPF）

実世界で1回の事象には使えない（応用が限定される筈）

MPF：願望投影錯誤

アドホッカリ { 信頼水準
信頼区間
不偏推定子
パワー関数
末端領域の重要性検定

ラプラスの確率の定義は等しく起こり易い (Equally likely) という表現が嫌われたのです。このことを避けるために、フォン・ミーゼスは頻度概念による確率を考え、その一派が強力にラプラスの確率を古典派に追いやってしまいました。1900 年から 1970 年までの間支配した統計学が正統派とされたのです。この確率理論は生物学者の要請に応えた分かり易い概念のものでした。

ジェインズは願望投影錯誤 (MPF : Mind Projection Fallacy) と称して、このようなことを排除すべきだと力説しています。そして、MPF から正統派統計学が問題処理のためにその都度臨時に考案した概念に、信頼水準、信頼区間、不偏推定子、パワー関数、末端領域の重要性検定があるとし、これらはすべてアドホックリであると決めつけています。

冒頭で投げかけた質問、「何が問題か」に対して、答えはなんと一番重要な「確率の定義が問題」なのです！

1.4 確率理論における歴史的論争

近代の確率理論は 18 世紀にはじまったと言って良いでしょう。フランスの賭博師ル・シャブリエが二つの賭でどちらが有利かを数学者に聞いたことから始まったとされています。賭博師で数学史に名を残したのはこの人ぐらいではないでしょうか。

歴史的論争

論理としての確率

ベルヌーイ、ラプラス、ポアソン、ルジャンドル
ガウス、ボルツマン、マックスエル、ギブス
シャノン、ウイーバー、ジェフリー、コックス、ポーリャ
ケインズ、トリバス、ローゼン克蘭ツ
デ・フィネッティ、ジェインズ

抵抗勢力 $\Longleftrightarrow$ 正統派、頻度派

カルノー、エリス、ブール、ベン、フォン・ミーゼス
ネイマン、ピアソン、フィッシャー、クラメール

ベルヌーイからラプラスによる確率計算は本質的に場合の数を求める計算でした。さいころ 2 個同時に振った場合であれば、出目の種類としては 1 - 1 から 6 - 6 まで 15 種ですが、確率は $6 \times 6 = 36$ 種の出方がそれぞれ同じと見て計算すべきというようなことです。日本の江戸時代には丁半博打が広く行わ

れていました。2個のさいころを振った時の出目の和が丁（偶数）になる組み合わせは、は12通りありますが、半（奇数）になる組み合わせは9通りしかありません。そこで素人衆は丁の方が起こりやすいと錯覚したようです。

1960年代後半に岩波講座基礎工学、全13巻が発刊されています。この中で理論物理学者が書いた熱力学4)と統計学者の書いた確率統計現象5)では同じ講座でありながら確率の定義が異なっています。工学は役に立てば良いのであって、臨機応変に対処すれば良い、色んな考えがあって良いと思う人も多いかもしれません。

しかし、工学であっても背景としての考え方の論理は一つでなければおかしいのです。本当は両立しないことだからです。このため、正統派統計学の一派がラプラス流を攻撃した論争もすさまじいものでした。岩波講座基礎工学の確率について、個々の分冊では矛盾していなくても全体としてみると首尾一貫していないことは明かです。軍配は熱力学の確率に上げなければならないのです。

1.5 信頼度計算の問題点

開発段階で犠牲を伴ったもののアポロ計画の有人月探査はミッションを達成しました。このため、世間では大成功を納めたかのように見られてきたものに信頼度計算があります。この信頼度計算も、少し論理的に考察するとおかしいところにすぐ目に付くのですが、他に方法が無かったことから、唯一正統な方法であるとみなされて来ました。むしろ、このことが驚くべきことかもしれません。

いま簡単なモデルとして、2段式のロケットの信頼度を求めるものとします。このロケットは1段と2段で構成されているので、1段が成功して、その後分離した2段も成功して初めてこのロケットが成功したと言えます。このようなロケットの信頼度ブロック図は以下の図のように描けます。

今、命題A、B、Cを、以下のようにしますと、

A：1段が成功する

B：2段が成功する

C：ロケットが成功する

A が真でB が真であるとき、C が真ですから、 $C = A B$ の論理積になります。この場合物理現象的には、A が先でB が後との時間的な順序がありますが、論理的にはA とB の間で時間の順序はありません。

A が真である確かさを R_1 、B が真である確かさを R_2 とすると、C が真である確かさRは、

$$R = P(C | H) \quad (1-3)$$

Hはロケットが2段式であることなど、当該問題を設定した種々の条件を示します。P(C|H)は条件Hの下でCが真である確率を示します。つまり、Cの成立する確かさ、即ちロケットの信頼度です。

このとき、論理積の確率は、

$$P(C|H) = P(A|H)P(B|AH) \quad (1-4)$$

なのです。P(B|AH)は条件HのもとでAが真である条件付きでBが真であることの確率です。AとBが論理的に独立であるならば、

$$P(B|AH) = P(B|H), \quad (1-5)$$

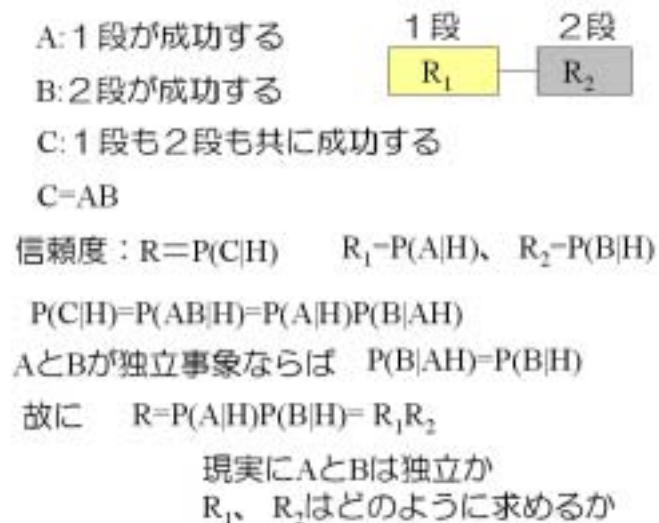
ですから、この場合に限り、

$$P(C|H) = P(A|H)P(B|H) \quad (1-6)$$

つまり信頼度は、 $R = R_1R_2$ の式が成立します。現実には1段と2段が論理的に独立であるとは言い難い場合が多いでしょう。

しかし、ブロック図を描いたときのそれぞれの構成要素はすべて独立であると仮定しないと計算のしようがないのです。

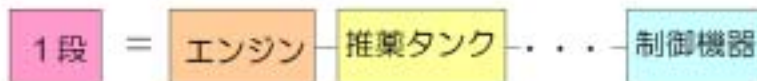
信頼度計算法の問題点



さて1段や2段の信頼度を見積もるために、それぞれの構成要素を同じ考えで1段の信頼度はその1段を構成するサブシステムの信頼度から、それぞれのサブシステムの信頼度はそのサブシステムを構成するコンポーネントから、そしてそれぞれのコンポーネントの信頼度はその構成部品の信頼度からというふうに考えます。最後の単位はトランジスタやコンデンサのレベルの部品になります。

信頼度計算法の問題点

サブシステムに分解



サブシステムはコンポーネントに分解

コンポーネントは部品に分解

部品の信頼度は？

部品の信頼度が信頼水準90%で0.99以上という表現からはコンポーネントの信頼度計算が出来ない

このぐらい詳細にまで部品のレベルを小さい単位で考えますと、部品の数は数万点から数十万点のオーダーです。ブロック図はこのように細かいレベルまで描ききれないことは明らかですから、同じような系統はなるべく代表的に描くことになります。多少の近似は仕方がないでしょう。そして、代表的なブロック図においてコンポーネントとそれを構成する部品で信頼度ブロック図が描けることでしょう。部品が信頼度計算の最終単位であるとする、それらの信頼度が分からなくてはならないことになります。ブロック図は必ずしも全部が直列系であるとは限りませんし、必要と思われるところは冗長系を組みますが、個々の部品の信頼度は極めて高くないと全体での信頼度は意味もないほど低くなってしまいます。

100 個の部品が一つでも故障したら全体の故障に繋がるとき、それぞれの部品の信頼度が 0.9 では $0.9^{100}=0.00003$ 、 $0.99^{100}=0.37$ 、そして $0.999^{100}=0.90$ ですから部品点数が 100 個から構成されるようなものであってもそれぞれの信頼度はせめて 0.999 ぐらいでないと、まず成功しそうなことがわかります。

ところが、一つの部品が 0.999 の信頼度があるとどうして言えるのでしょうか。客観的には 1000 個試験してみて全部成功すれば初めて 0.999 の信頼度があると言えるでしょうが、とても 1000 個もの試験をやれないの状況が普通です。まして、1 万個の部品が全部成功しないと全体の成功と言えない場合にはそれぞれの部品に要求される信頼度は極めて高いものになり、数が多すぎて試験では実施できなくなります。

試験個数が少ない場合にどうしたら高い信頼度であると言えるのかを考える

のではなく、その場合はそれなりの信頼度しか無いという表現が自然なのではないでしょうか。

[演習問題 1] 真空管式コンピュータの信頼度

真空管を 1 万本使ったコンピュータがある。真空管以外の箇所は故障しないが、真空管が 1 本でも故障したらコンピュータはまともに作動しないものとする。このコンピュータが完成後最初の 1 時間、まともに作動する信頼度を求めよ。

但し、このコンピュータに使われた真空管はどれも同じ型のもので、その寿命は過去の実績データを見ると故障までの時間の平均が 5000 時間で、分布の形状は指数分布に極めて近いことが判っているものとする。

ヒント：指数分布の密度関数は $f(x) = \lambda e^{-\lambda x}$

指数分布の期待値は $E(x) = \frac{1}{\lambda}$

指数分布の分散は $V(x) = \frac{1}{\lambda^2}$

真空管 1 本が 1 時間作動する信頼度を先に求め、1 万本の真空管が故障するのはそれぞれ独立であると考えよ。

2 . 確率と実世界

2.1 実世界の認識

現実の世界と抽象の世界はどこが違うのでしょうか。抽象世界は現実世界の近似であると考えた方が良さそうです。

現実世界には無限大も無限小もありません。無限大も無限小も抽象的概念に過ぎません。如何に大きなものでも有限で、小さい物でも最小単位があり量子的です。連続も抽象的概念です。これらのことを意識することは重要です。何故なら多くのパラドックスは消え、無用な議論を避けられるからです。

大きな物を考えて見ましょう。人間が実際に存在するものとして考えられる最大のものは宇宙全体でしょう。宇宙には 2000 億 (2×10^{11}) の銀河があり、一つの銀河は太陽程度の星を 2000 億個ぐらい含んでいます。太陽の質量は 2×10^{33} g です。従って、宇宙全体の質量は 8×10^{55} 程度です。宇宙にはダークマターの方が多いという説が有力になってきています。それで 10 倍してみても、宇宙全体の質量は 10^{57} g 程度でしょう。水素原子の質量が 1.6×10^{-24} g ですから、宇宙全体の原子の数は 10^{81} 個よりは少なそうです。

実世界の認識

現実世界は有限、不連続

一番小さいもの 素粒子 10^{-15}m

超ひも (まだ抽象?) 10^{-35}m

一番大きなもの 宇宙全体の質量 (含む、ダークマター)

一番小さいものを単位にして一番大きなものを表す

太陽の質量 $= 2 \times 10^{33}\text{g}$ 銀河は 2×10^{11} 個の星

水素原子の質量 $= 1.6 \times 10^{-24}\text{g}$

Google は 10^{100} π : 1 兆 2411 億桁まで計算

組み合わせの数は大きい。 $52! = 10^{68}$ $2^{128} = 10^{38}$

長さでも同じようなものです。現在の宇宙論では、宇宙の果ては 137 億光年と考えられていますが、この距離を m で表しても、

$137 \times 10^8 \times 365 \times 24 \times 60 \times 60 \times 30 \times 10^7 = 1.3 \times 10^{26}\text{m}$ です。この長さを一番小さな素粒子の大きさを単位として測ったとしても、素粒子の直径は 10^{-15}m ですから数字の大きさは 10^{41} 程度に過ぎません。将来、素粒子よりも小さい超ひも理論の超ひもが実在が確認されたとして、それを単位にすることにしても数値の大きさは 10^{20} 倍になる程度です。現在のところ超ひもはまだ抽象の世界です。つまり、現実世界ではどんなに小さな単位で計っても 10 進法で数字が 100 個もあれば済んでしまうのです。グーグルはインターネットのウェブ検索ソフトで最も有名なものですが、10 の 100 乗を意味する言葉だそうです。グーグルはゴーゴルから由来し、「ゴーゴル」は「天文学的な多さの」という意味がですが、天文学的数字と言ってもせいぜいこの程度です。

これに対して抽象の世界はもう桁違いに大きくも小さくもなります。超越数の一つである円周率パイの値は、平成 14 年末に東大の金田康正教授が 1 兆 2411 億桁まで計算しています。

これらのことから、現実の世界では無限集合ではなく、有限集合を考えれば十分であることがわかるでしょう。ただし、現実の世界でも、組み合わせの数はすぐに大きくなることに注意が必要です。

実世界の認識

現実世界

連続はあり得ない。 近似として考えるもの

無限大はあり得ない。 近似として考えるもの

無限小も無い。 ε 、 δ 理論は有効か？

パラドックスはあり得ない。

無限大を考える時に要注意。

有限操作の極限として無限大を考えるようにせよ。

殆どが抽象の世界

理論モデル

トランプのカードはジョーカーを抜きにして 52 枚あります。このカードは新品の時にはきれいに並んでいるのですが、これを何回も切ることによって乱雑になります。52 枚のカードの並び方は $52! = 10^{68}$ にもなります。しかし、人間の手で切る操作は有限の回数ですし、それを見る回数も知れています。決して使われることのなかった並び方の方がずっと多いのです。

インターネットで使われているアドレスは 2 進数で 32 桁です。これはだいたい 10 億ぐらいになります。しかし、世界の人口は 60 億を超えましたし、当初の予想に反してあまりにも急速にインターネットは進歩したため、既にアドレスが不足するようになってきました。そこで今、次世代のインターネットのプロトコルが考えられ、一部で試験使用が始まっています。今度は 2 進数 128 桁のアドレスです。これだと 10^{38} ぐらいの多さなので、人間が使う道具にまでアドレスを与えることが出来るほど十分余裕があります。

2.2 演繹と帰納

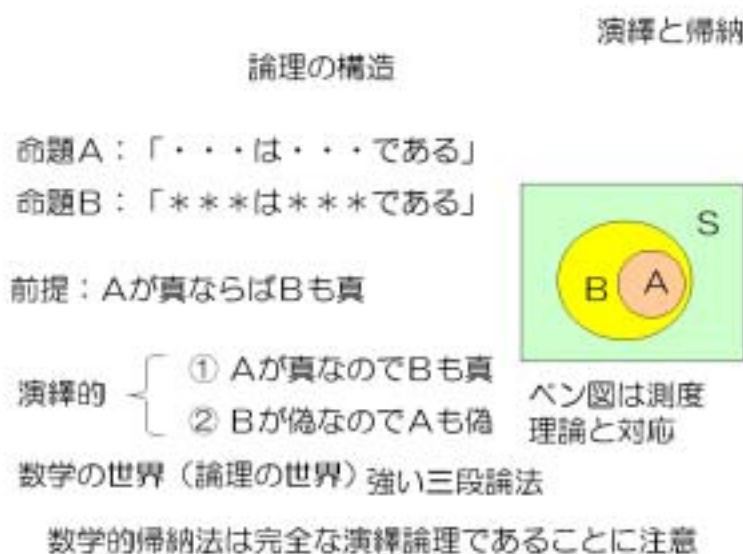
これから確率論を考える前に、論理の進め方を復習する必要があります。論理学というと堅苦しく難しそうですが、基礎的な部分だけですのでしばらくおつきあいください。

演繹 (deduction) とは論理的に否定できない推論を言います。その正しさに異議を唱えさせる隙もなく、有無を言わせない論理です。演繹に対応する言葉で帰納 (induction) ということがあります。帰納は妥当さ (plausibility) が増減するだけで、いつも絶対的に成り立つというわけではない推論を言います。ただし数学的帰納法で使っている論法は有無を言わずことなく正しい論理

ですので明らかに演繹的推論に入ります。

命題(proposition)とは真偽を主張する言明です。簡単な命題は「・・・は・・・である」という表現を取ります。否定の命題ですと「***は***でない」となります。

今、二つの異なる命題AとBがあって、Aが真ならBも真という状況があるものとします。例えばAとして「さいころを振って1が出る」で、Bが「さいころを振って奇数が出る」とした時に、この状況でしょう。



この条件のとき、演繹推論で言えることは Aが真なのでBが真、と Bが偽なのでAも偽の二つだけなのです。数学の理論は、殆どすべて、この演繹推論で成り立っています。これは強い三段論法（syllogism）と呼ばれます。純粋数学者が定理を証明するときに、この3段論法のみを使います。しかし、その最終結論に至るまでは、純粋数学者であっても次の弱い推論を使っているのです。

2.3 拡張論理としての確率理論

現実世界で私たちが経験する多くの場合は強い三段論法で押し通せません。これらには3通りの場合があり、弱い3段論法（epagoge）と呼ばれます。

まず、上述の状況と同じで、二つの異なる命題AとBがあったとし、Aが真ならBも真という状況です。

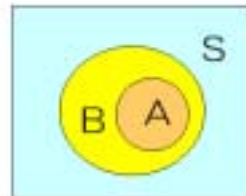
この条件のもとに、 Bが真なのでAの妥当さが増す、という時と、 Aが偽なのでBの妥当さが減る、という場合があります。これらの結論は帰納的推論と呼ばれます。

拡張論理

現実世界は殆どこの状況

前提：Aが真ならばBも真

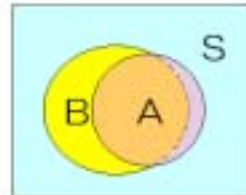
- ③ Bが真なのでAの妥当さが増す
- ④ Aが偽なのでBの妥当さが減る



前提：Aが真ならばBの妥当さが増す

- ⑤ Bが真なのでAの妥当さが増す

帰納的推論



帰納的推論と呼ばれるものにもう一つ別の状況があります。それは、Aが真ならばBの妥当さが増す、という状況のもとで、Bが真なのでAの妥当さが増す、と言う場合です。

強い3段論法は何段に重ねても強い結論ですが、弱い三段論法は重ねれば重ねるほど弱い結論になります。しかし、これらの弱い3段論法の間にも明確な規則があることを示したのが数学者ジョージ・ポーリャです。しかし、ポーリャは定性的な表現にとどまっています。これを定量的にしたものが、ジェインズの拡張論理としての確率理論なのです。

2.4 ブール代数

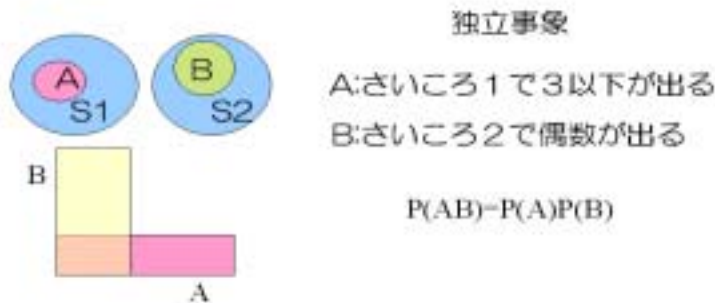
本論の核心に進むまえに、論理学の基礎を復習しておきます。いずれもベンダイアグラムを頭の中に描いて見れば簡単に納得できるものばかりです。

これらは、等冪律、交換律、結合律、分配率、双対律、です。これらの論理式が拡張論理においても基礎式であることに変わりはありません。

ブール代数

等幂律	$\begin{cases} AA = A \\ A + A = A \end{cases}$	交換律	$\begin{cases} AB = BA \\ A + B = B + A \end{cases}$
結合律	$\begin{cases} A(BC) = (AB)C = ABC \\ A + (B + C) = (A + B) + C = A + B + C \end{cases}$		
分配律	$\begin{cases} A(B + C) = AB + AC \\ A + (BC) = (A + B)(A + C) \end{cases}$		
双対律	$\begin{cases} \text{もし、} C = AB \text{ ならば、} \bar{C} = \bar{A} + \bar{B} \\ \text{もし、} D = A + B \text{ ならば、} \bar{D} = \bar{A}\bar{B} \end{cases}$		

ベン図で上手く表せないこと



ジェインズはベンダイアグラムを使うことを敢えて避けています。それはベンダイアグラムが包含関係を表す円などで表されているため、必然的に無限集合を基礎におくことになるからです。ジェインズの拡張論理としての確率理論は有限集合を基礎にすることで現実世界をより良く説明するからです。

また、ベンダイアグラムでうまく表せないものに、事象の独立があります。事象が相互に異なる次元であることを示す図でないと独立であることは描けません。

2.5 関数方程式

関数方程式とは $F(X + Y) = F(X) + F(Y)$ (コーシーの方程式) などのように、未知関数 $F(X)$ を含む方程式を言います。関数方程式は、その関数の持っている性質を表しています。したがって、関数方程式における興味の内容は、その性質を満たす関数をすべて求めたいということです。

関数方程式は、コーシーの方程式以外にも次のものがよく知られています。

(1) 全ての実数 X, Y に対して、 $F(X + Y) = F(X)F(Y)$

(2) 全ての正数 X, Y に対して、 $F(XY) = F(X) + F(Y)$

(3) 全ての正数 X, Y に対して、 $F(XY) = F(X)F(Y)$

そして、これらは全て、基本的にコーシーの方程式に帰着されるものです。論理としての確率理論でも情報理論の基礎にも関数方程式が出てきます。関数方程式は少し判りにくいことがラプラスの確率理論が排斥された原因の一つを作ったものと思われます。

3. 合理性と首尾一貫性

3.1 デシダレータ

弱い三段論法を定性的なものから定量的なものにするために誰もが認める基本的な約束事を定める必要があります。幾何学における公理のようなものです。ジェインズはこれらを、ラテン語で「必要不可欠な、または、切実な条件」を意味する言葉、デシダレータ(単数形はデシダレータム)と呼びました。公理と呼ばなかったのは、数学的な美しさに欠けるからだそうです。

第1のデシダレータムは、(1) 妥当さの程度は実数によって表される、というものです。妥当さの程度を他の物、例えば色彩で表すことが可能かも知れません。しかし、赤と青ではどちらがより妥当さが強いのか定義付けが難しいことでしょう。従って、実数で表すことが出来ることは合理的に表現するために必要不可欠なことです。

第2のデシダレータムは、(2) 常識との定性的一致です。これはより強い妥当さの方がそれを表す数値がより大きいとすることで方向を決めておくだけのものです。もし、逆に決めると人間は変な感覚に陥るでしょう。

第3のデシダレータムは首尾一貫性の要求です。(3) 一つの結論がいくつかの方法で得られるならば、それらの結果はすべて同じ結果でなければなりません。

デシダレータ

デシダレータ（切実な条件）

- (I) 妥当さの程度は実数によって表せる。
- (II) 常識との定性的一致を必要とする。
- (IIIa) もし一つの結論が多数の方法で推測出来る場合、それらは同じ結果にならなければならない。
- (IIIb) ロボットは質問に関係を持つ全ての証拠を考慮に入れなければならない。
- (IIIc) ロボットは常に等価な妥当さ割り当てに対して等価な知識の状態を示す。

構造要求：I、II、IIIa

インタフェース要求：IIIb、IIIc

デシダレータ

首尾一貫性

- (IIIa) もし一つの結論が一つ以上の方法で推測できるならばあらゆる可能な方法は同じ結果に至らなければならない。
- (IIIb) ロボットは常に質問に関係を持つ全ての証拠を考慮に入れなければならない。それは情報のいくつかを勝手に無視はしない。何が残っているかにのみ基づいて結論を出す。換言すると、ロボットは完全に悲観的である。
- (IIIc) ロボットは常に等価な妥当さ割り当てによって等価な知識の状態を表す。即ち、もし二つの問題でロボットの知識の状態が同じであれば、（おそらく、命題の呼び名を除いて）、ロボットは両方に同じ妥当さを割り当てなければならない。

三つのデシダレータをまとめて簡潔に記すと、

妥当さは実数で表現

常識との一致

首尾一貫性

になります。

これらの妥当さを示すのは人間なのですが、人間は不可解なもの、不合理な行動を取るもの、とかの主張をする人もいて議論が進まなくなりますので、合理的で首尾一貫した行動を取るロボットを想定します。私たちはこのロボットに対する設計条件を与えたことになります。

ロボットの首尾一貫性を確保するためにあと二つの条件を付け加えます。

- a 一つの結果がいくつかの方法で得られるならば、
それらの結果は同じでなければならない。
- b ロボットはすべての情報を考慮にいれなければならない。
- c ロボットは等価な妥当性によって、等価な知識の状態を表す。

初めてロボットの用語を作り出したのはアシモフです。アシモフは人間が作り出したロボットは人間のために働くべきだとして次の原則を与えました。

原則（ ）ロボットは人間に危害を加えてはならない。また人間に危害が加えられるのを見過ごしてはならない。

原則（ ）ロボットは人間の命令に従わなければならない。但し、命令が原則（ ）に反する場合はこの限りでない。

原則（ ）ロボットは原則（ ）及び（ ）に反しない限り自分の身を守らなければならない。

デシダレータと比較してみると、デシダレータはロボットの判断の仕方を与えているものであることが判ります。アシモフのロボットもデシダレータに従って判断し、人間とロボットの安全のために働くことになります。アシモフのロボットは何をすべきなのか、何故そうすべきなのか、何時、どこで、を教えられていません。ただロボットは如何にすべきかの条件だけを与えられています。

アシモフのロボット

原則（Ⅰ）ロボットは人間に危害を加えてはならない。
また人間に危害が加えられるのを見過ごしてはならない。

原則（Ⅱ）ロボットは人間の命令に従わなければならない。
但し、命令が原則（Ⅰ）に反する場合はこの限りでない。

原則（Ⅲ）ロボットは原則（Ⅰ）及び（Ⅱ）に反しない限り自分の身を守らなければならない。

3.2 積の法則

「確率理論は計算に落とされた常識に他ならない」はラプラスの 1812 年の言葉です。前述のデシダレータから定量的規則として、積の規則と和の規則が導か

れます。

最初に論理積 $A B$ の妥当さが、 A と B の妥当さに個別に関係している首尾一貫した規則を求めます。

$A B$ が真であることを決める手順は二通りあります。

(1) B が真であると決める。 $(B|C)$

(2) B が真であることを受け入れ、 A が真であると受け入れる。 $(A|BC)$

または

(1) A が真であると決める。 $(A|C)$

(2) A が真であることを受け入れ、 B が真であると受け入れる。 $(B|AC)$

最初の手続きを言葉で表すと、 $A B$ が真の命題であるために B が真であるということが必要ですから $B | C$ が組み込まれるべきです。加えて B が真であるならば、 A が真であることが必要ですから $A | B C$ も組み込まれるべきです。しかし、 $B | C$ と $A | B C$ が得られていればもはや $A | C$ は不要です。同様に $A | B$ も $B | A$ も不要です。

論理積は可換で、 $A B = B A$ ですから A と B を取り替えても全く同じことが言えます。即ち、 $A | C$ と $B | A C$ は $A B | C = B A | C$ を決めるのに等しく役立ちます。どちらの手続きからでも $A B | C$ のためにロボットは同じ値を得なければならないということが、デシダレータム (a) です。

積の規則

$$u = (AB|C) \quad \begin{array}{ll} v = (A|C), & w = (B|AC), \\ x = (B|C), & y = (A|BC) \end{array}$$

2個以上の組み合わせ (1 1 種) で
意味を持つのは2種のみ

$$\left\{ \begin{array}{l} u = F(x, y) \\ u = F(w, v) \end{array} \right.$$

$$(ABC|D) = F[(BC|D), (A|BCD)]$$

$$(ABC|D) = F\{F[(C|D), (B|CD)], (A|BCD)\}$$

$$(ABC|D) = F\{(C|D), F[(B|CD), (A|BCD)]\}$$

$$\text{関数方程式} \quad F[F(x, y), z] = F[x, F(y, z)]$$

従って、 $(A B | C)$ は $B | C$ と $A | B C$ のある関数になるでしょう。

$$(A B | C) = F [(B | C) (A | B C)] \quad (3-1)$$

他の関係式があり得るかを調べるために、

$$\begin{aligned} u &= (A B | C), \quad v = (A | C), \quad w = (B | A C) \\ x &= (B | C), \quad y = (A | B C) \end{aligned} \quad (3-2)$$

と置きます。uをv、w、x、yの二つ以上の関数として表すことが出来るならば、11の可能性がありますが、トリバス(1969)は論理的に可能性のあるのはただ二つだけであることを示しました。それは $u = F(x, y)$ と $u = F(v, w)$ だけでした。

次に三つの命題ABCが同時に真である妥当さ $(A B C | D)$ を見つけるものとします。ブール代数の結合律から $A B C = (A B) C = A (B C)$ です。つまり二つの異なった方法で求められ、それらが首尾一貫性要求から同じでなければなりません。

上述のように意味のある二つの表し方から、

$$(A B C | D) = F[(B C | D), (A | B C D)] \quad (3-3)$$

妥当さ $(B C | D)$ に再び(3-1)を適用して、

$$(A B C | D) = F\{F[(C | D), (B | C D)], (A | B C D)\}$$

別の表し方で、

$$\begin{aligned} (A B C | D) &= F[(C | D), (A B | C D)] \\ &= F\{(C | D), F[(B | C D), (A | B C D)]\} \end{aligned}$$

これらが同じでなければなりませんから、

$$F[F(x, y), z] = F[x, F(y, z)] \quad (3-4)$$

(3-4)式は関数を持つべき制約条件を表しており、関数方程式と呼ばれるものの一つです。

この関数方程式を解いてもう少し明確な表現の式にする途中経過は少し、いやかなり、煩雑です。結果は、

$$F(x, y) = w^{-1}[w(x), w(y)]$$

という関係が得られ、これから、

$w(A B | C) = w(A | B C) w(B | C) = w(B | A C) w(A | C)$ が得られます。これが積の法則です。

そして、Cの元にAが絶対に確実である場合には、

$$w(A | C) = 1、$$

Cの元にAが絶対に不可能である場合には、

$$w(A | C) = 0、$$

とすることがデシダレータに違反せずに可能なのです。ここでwの代わりにPを使うと確率の表現に同じです。

$$p(AB|C) = p(A|C)p(B|AC) = p(B|C)p(A|BC) \quad (3-5)$$

積の規則

デシダレータⅢaから

$$w(AB|C) = w(A|BC)w(B|C) = w(B|AC)w(A|C)$$

デシダレータⅡから

$w(x)$ は正の単調増加関数でなければならない

$0 \leq w(x) \leq 1$ として一般性を失わない。



$$P(AB|C) = P(A|BC)P(B|C) = P(B|AC)P(A|C)$$

3.3 和の法則

さて和の規則を誘導します。

$$p(AB|C) = p(A|C)p(B|AC) = p(B|C)p(A|BC) \quad (3-6)$$

$$p(A|B) + p(\bar{A}|B) = 1 \quad (3-7)$$

(3-7)式が和の規則なのですが、この式は殆ど自明でしょう。

和の規則

論理積 $A\bar{A}$ は常に偽

論理和 $A + \bar{A}$ は常に真

u と v に関数関係が必要

$$\begin{array}{l} u \equiv w(A|B) \\ v \equiv w(\bar{A}|B) \end{array} \quad v = S(u) \quad \left\{ \begin{array}{l} S(0) = 1, S(1) = 0 \\ 0 \leq u \leq 1 \end{array} \right. \quad \text{常識}$$

$$P(AB|C) = p(A|C)p(B|AC) = p(B|C)p(A|BC)$$

$$p(A|B) + p(\bar{A}|B) = 1$$



$$p(A + B|C) = p(A|C) + p(B|C) - p(AB|C)$$

これらの二つの式から、繰り返し適用して、

$$\begin{aligned}
p(A+B|C) &= 1 - p(\overline{AB}|C) = 1 - p(\overline{A}|C)p(\overline{B}|\overline{A}C) \\
&= 1 - p(\overline{A}|C)[1 - p(B|\overline{A}C)] = p(A|C) + p(\overline{A}B|C) \\
&= p(A|C) + p(B|C)p(\overline{A}|BC) = p(A|C) + p(B|C)[1 - p(A|BC)]
\end{aligned}$$

そして、

$$p(A+B|C) = p(A|C) + p(B|C) - p(AB|C) \quad (3-8)$$

(3-8)式が一般化した和の規則です。

3.4 ベイズの定理

さて、積の規則 (3-6) 式を変形すると、

$$p(A|BC) = p(A|C) \frac{p(B|AC)}{p(B|C)} \quad (3-9)$$

が得られます。このように簡単に得られますから、ベイズが自分では定理であると意識しなかったのも無理ありません。後生の人がベイズの論文を見てベイズの定理と名付けたものです。

ベイズの定理

X : 事前情報
H : 検定されるべきある仮説
D : データ

積の規則

$$P(DH|X) = P(D|HX)P(H|X) = P(H|DX)P(D|X)$$

$$\text{故に、} \quad P(H|DX) = P(H|X) \frac{P(D|HX)}{P(D|X)}$$

事象 A の確率 $p(A|C)$ が、事象 B が真であったという事実を得たあと $p(A|BC)$ に変わることを示しています。

X : 事前情報

H : 検定されるべきある仮説

D : データ

とすると、ベイズの定理は

$$p(H|DX) = p(H|X) \frac{p(D|HX)}{p(D|X)} \quad (3-10)$$

データを見た後の仮説の妥当さがこの事後確率計算により計算できます。ここ

で、 $p(D|HX)$ はデータの尤度です。また、事前事後は物理的なデータの発生前後でなく、データを見る前か見た後かなのです。言うならば「知前知後」というべきものです。

ベイズの定理

ベイズ統計学の骨格

事前分布 $P(H|X)$ の計算（確率割り当て）

観測データ D の入手

事後分布 $P(H|DX)$ の計算（事後確率計算）

$$P(H|DX) = P(H|X) \frac{P(D|HX)}{P(D|X)}$$

$P(D|HX)$: データの尤度

ベイズの定理を中心に据えたベイズ統計学と呼ばれるものがあります。論理としての確率理論もベイズの定理が主要な役割を演じます。しかし、世間に流布しているベイズ統計学に関しては、これを応用する立場で少し誤解があるかも知れません。

ベイズの定理

ベイズ統計学に対する誤解

誤解：事前分布は仮定するもの
主観確率は感覚的なもの

事前分布はどのように割り当てるべきか

デシダレータから無差別の原理で
（不十分理由の原理：ベルヌーイ）

最大エントロピー原理で

不変量の原理

周辺分布の原理？

それらは、事前分布は仮定するものとされているもの、または主観確率は感覚的なものと決めつけているものです。事前分布は仮定するものでなく、デシダレータに違反しないように割り当てるべきものなのです。そして、この割り当て方が個々の問題に応じて研究されるべきものなのです。これまでに、無差

別の原理、最大エントロピー原理、不変量の原理、周辺分布の原理などが知られています。

確率はデータを見ている人と見ていない人で異なるという意味では主観確率なのですが、データを見ている場合、見ていない場合でそれぞれに確率の割り当て方がデシダレータから決められていますので、この意味では客観的です。ベイズ統計学は主観確率であると言っても決して感覚的にいい加減に決めているわけではありません。

4. ラプラスの確率

4.1 無差別の原理

一番基本的な確率割り当ての方法が無差別の原理です。ベルヌーイが最初に使った原理ですが、彼は不十分理由の原理という奇妙な名前と呼びました。一般的には、この無差別の原理を少し変形した形で表現したラプラスの原理として知られています。

さて、命題 $\{A_1, A_2, \dots, A_n\}$ が相互背反で網羅的ならば、

$$\sum_{i=1}^n p(A_i | B) = 1$$

でなければなりません。そして、 A_i と A_j に関して特に差異を認める情報を持っていないならば、これらを入れ替えてもロボットが抱く妥当さは同じでなければなりません。つまり、デシダレータ b 、 c の要請です。従って、

$$p(A_i | B) = \frac{1}{n}, \quad (1 \leq i \leq n)$$

でなければならないのです。

無差別の原理

$$\sum_{i=1}^n p_i(A_i | B) = 1 \quad \text{排反事象で網羅的}$$

$\Downarrow$ デシダレータ III c から

$$p(A_i | B) = \frac{1}{n}, \quad (1 \leq i \leq n)$$

論理的推論の帰結として数式が導かれた！

エントロピー最大の原理に合致

これがケインズ(1921)によって名付けられた無差別の原理です。原理と言ってもデシダレータの要請から必然的な論理的帰結なのです。これが最も基本的な確率の割り当て方を決める式で、この場合は後で出てくる最大エントロピー原理に包含されるものです。そしてこの時点で始めて、論理的な推論の積み重ねから0、1以外の実数が出現しました。ジェインズはこの時点まで妥当さに与えた数値を、注意深く、「確率」と呼んでいません。それは妥当さに対して、どのように数値を与えるべきかが示されていなかったからです。

積の規則と和の規則に従う量としてこの確率 p を用いることが「正確」だからでなく、便利だからこれを用いるのです。このことは熱力学的に定義したケルビン温度を用いることが「正確である」という理由よりも便利だからこれを用いるのと同じです。

4.2 ラプラスの確率の定義

ベルヌーイ(1713)によって与えられ、その時点から 150 年間用いられた元の数学的確率の定義はラプラスの偉大な「確率の解析理論」(1812)に次のように定義されています。「事象の確率はなにもこれらの場合のどの一つが他より起こり易いと期待させないとき、それらが等しく可能であると示すとき、それに好ましい場合の数と全ての可能な場合の数に対する比である。」

ラプラスの確率の定義

ラプラス（ベルヌーイ）の定義（古典的）

$$\Longrightarrow p(A|B) = \frac{M}{N}$$

ラプラス（ベルヌーイ）の確率の定義

「事象の確率は何もこれらの場合のどの一つが他より起こりやすいと期待させないとき、それらが等しく可能であると示すとき、それに好ましい場合の数と全ての可能な場合の数に対する比である。」

-----確率の解析概論：ラプラス

物理学者が使っているが、統計学者から
Equally likelyに批判があった。

〔演習問題 2〕確率計算

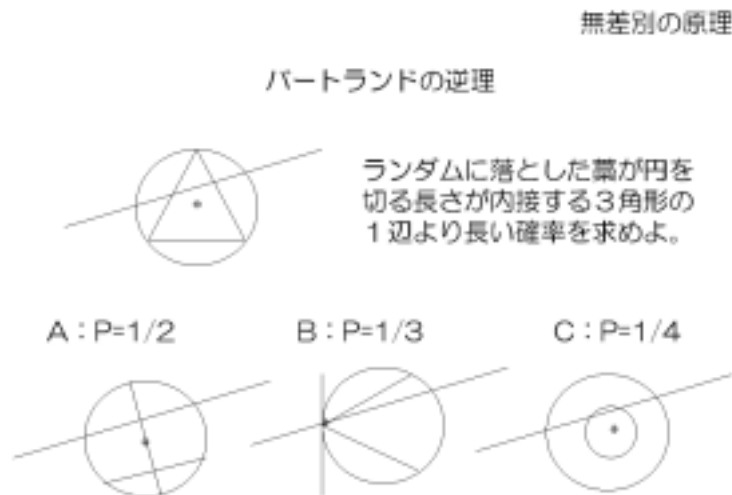
- (1) さいころ 1 個を 4 回振って 1 度は 6 が出る確率を求めよ。
- (2) さいころ 2 個を同時に 24 回振って、1 度は 6 - 6 が出る確率を求めよ。

〔演習問題 3〕確率計算と勝ち目

二人で交互にコインを投げ、最初に表が出た方が勝ちとする。先に投げる人が勝つ確率を求めよ。

統計物理学等の理論物理の学者は変わらずこの定義を使っていたのですが、生物学者を支援した統計学者の間から、この定義の「等しく可能である」という説明に猛烈な批判が起こりました。

この批判に力を与えたものに幾つかのパラドックスがあります。中でも最も有名で未だに解決されていないと信じられているおそれのあるものがバートランドのパラドックスです。



これは「ランダムに落とした藁（ストロー）が円を切る長さが内接する3角形の1辺より長い確率を求めよ。」というものです。藁は十分長く、かつ必ず円と交差するものという言外の前提があります。この確率をラプラスの定義で求めると、等しく起こり易いことの考え方で色々な答えが出てしまうというものです。

まず、円の中に入った部分のストローの中心が半径 $1/2$ の円内に落ちる確率に等しいと考えると確率は円の面積比に等しく $1/4$ になります。

次に、円を切るストローがその点での接線となす角度が 60 度から 120 度の範囲であると考えたとその確率は角度の比率を取って $1/3$ になります。

そして、円を切ったストローの中央部が直径上の中央部分にある範囲であると考えたと $1/2$ になります。

多くの教科書ではこのパラドックスの解は問題の定義が不十分なことによるとされています。確かに、問題をもっと具体的定義したらどれかを正解とする

問題ができるでしょう。

しかし、バートランドが与えたこのように簡単な問題に対して解が無いというのも変なことです。この問題は合理的に考えるとどう答えるべきかという観点に立てば解が決まります。それは不変量 (invariance) を考慮することなのです。

無差別の原理

不変量 (Invariance) を考慮

回転、尺度、並進の3種類の変数変換に対して
不変を保てるもの

$$p(x)dx = \frac{xdx}{(1-x^2)^{1/2}}, 0 \leq x \leq 1$$

ボレル (1909)

$$\int_{\sqrt{3}/2}^1 p(x)dx = 0.5$$

ボレル(1909)は回転、尺度、並進の3種類の変換に対して不変を保てる密度関数はどのようなものでなければならないかという推論から具体的な式の形を求めました。結果は、

$$p(x)dx = \frac{xdx}{(1-x^2)^{1/2}}, \quad 0 \leq x \leq 1$$

この関数を該当する範囲を積分しますと確率は 1/2 であると求められます。実際に実験した人がいて結果はこの結果に近かったそうです。6)

5 . 確率論の応用

5.1 初等標本抽出理論

標本抽出理論で必ず出てくる問題がベルヌーイの壺問題と呼ばれるものです。実際に壺から赤や白のボールを取り出したりすることは無いのですが、この問題が多くの実際的な問題の分かり易い理論モデルになっているのです。要は壺の中が見えないことです。日本の教科書では壺とボールの代わりに革袋と碁石のモデルで説明されることもあります。この場合も革袋であるから中が見えないことが肝心の要素です。

初等標本抽出理論として必要な基本式は、これまで出てきたものでまとめて並べても5つの式しかありません。たったこれだけから如何に多くの確率理論

が引き出せるか驚くほどです。

$$P(AB|C) = P(A|BC)P(B|C) = P(B|AC)P(A|C) \quad (5-1)$$

$$P(A|B) + P(\bar{A}|B) = 1 \quad (5-2)$$

$$P(A+B|C) = P(A|C) + P(B|C) - P(AB|C) \quad (5-3)$$

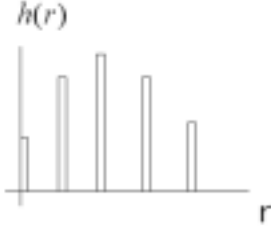
$$P(H_i|B) = \frac{1}{N}, \quad 1 \leq i \leq N \quad (5-4)$$

$$P(A|B) = \frac{M}{N} \quad (5-5)$$

初等標本抽出理論

A ≡ 「順序に関わらず n 回の取り出しで
r 個の赤玉が出る」 (戻し無し標本抽出)

$$h(r|N, M, n) = P(A|B)$$

$$h(r|N, M, n) = \frac{\binom{M}{r} \binom{N-M}{n-r}}{\binom{N}{n}} \Rightarrow$$


N : 総数

M : 赤玉総数 B : 実験条件

超幾何分布

一例として、ベルヌーイの壺問題から戻し無し標本抽出分布を求めてみます。N は壺の中にある赤いボールと白いボールの総数です。この中で赤いボールの総数は M であったとします。ここで、命題 A は「順序に関わらず n 回ボールを取り出したとき、r 回が赤玉である」とします。B は実験条件です。ここで、確率 $P(A|B)$ は r の関数になりますから、次の関数が定義できます。

$$h(r|N, M, n) = P(A|B) \quad (5-6)$$

これを上述の 5 つの式を使って具体的に求めると、

$$h(r|N, M, n) = \frac{\binom{M}{r} \binom{N-M}{n-r}}{\binom{N}{n}} \quad (5-7)$$

となって、横軸に r を取ってグラフにすると棒グラフ状になって、これは超幾何分布と呼ばれるものです。

5.2 連続則

ベルヌーイの壺を少し抽象的なものにします。ボールが非常に多く入っている壺ですが赤いボールがいくつ含まれているのか分かりません。そのような壺から n 個のボールを取りだしたら r 個のボールが赤でした。次に取り出すボールが赤である確率はどのようなものかという問題です。

このモデルは全く情報の発信され無い国が近々ロケットを打ち上げるという情報をつかんだとき、そのロケットが成功する確率はどうかと考える時と状況が似ていると思われるでしょう。そして、過去に n 回打ち上げて r 回成功したことが分かった後ではその確率はどうでしょうか。

ラプラスはこの確率は次式であることを導きました。

$$P = \frac{r+1}{n+2} \quad (5-8)$$

ラプラスはこの値が壺の中のボールで赤いボールの比の期待値として求めたものですが、この問題の場合、その期待値が求める確率でもあることは疑問のないところでしょう。

ラプラスはこの式の応用として、「明日も太陽が昇る」に対してどうなるかを記しました。但し書きで、この式を適用する時の条件を示したにも関わらず、条件を良く読まずに数値だけを聞いた人々の誤解を招いて、なおさらラプラスの確率定義が疎まれることになりました。あまり良い例でなかったとは言えるでしょう。しかし、この式が確率割り当てのための非常に基本的な式の一つであることには変わりがありません。

式の誘導は参考文献 7) に私が導いたものがあります。ラプラスが導いた方法は文献 2) に出ています。他にも 2 種類の誘導方法が出ていますが、最も簡単な方法は誘導と言うよりも解釈と言ったほうが合っているかも知れません。ラプラスはさらに一般化した式まで導いていますが、かなり難解です。

6 . 不確かさの測度

6.1 シャノンの条件と情報エントロピー

シャノンは確率分布でしか表すことが出来ない不確かな状態を合理的に表す方法を考えました。例えば、コインを投げて表が出るかどうかと言う状況と、さいころを投げて 1 が出るか出ないかという状況はどのように表されるべきかを考えたのです。

何かが二つに一つで決まる状況と六つに一つで決まる状況とは不確かさの程度が違ふと考えるわけです。当たりが出る状況が確率 $1/2$ の状況と確率 $1/6$ の状況とでは不確かさが異なります。宝くじでは、1 等が当たる不確かさの状況

は6等が当たる状況よりずっと大きいと考えられます。なかなか当たらないのは不確かさが大きいからです。

確率分布によって表された「不確かさの量」の測度が首尾一貫しているためにはどのような条件を満たせば良いのでしょうか。これがシャノンの条件と呼ばれるものです。

- (1) ある数値的な測度 $H_n(p_1, p_2, \dots, p_n)$ が存在すると仮定します。即ち、「不確かさの量」と実数の間にある種の間接関係を作ることが可能です。ここで、 $p_1 + p_2 + \dots + p_n = 1$ の関係があります。
- (2) 連続性特性を仮定します。 H_n は p_i の連続関数です。というのはそうでないとする確率分布の小さな変化が不確かさの量において、大きな変化に導くことになりかねないからです。
- (3) この測度は多くの可能性がある時には、それが少ないときよりも、より不確かでなければなりません。これは常識に定性的に対応しているものであることの要求です。この条件は p_i が全て等しい場合に次の形式を取ります。その量、

$$h(n) = H_n\left(\frac{1}{n}, \dots, \frac{1}{n}\right) \quad (6-1)$$

は n の単調増加関数です。これにより「方向感覚」が確立します。

これらから論理的に推論を重ねていくと、 $h(n)$ は次の関数方程式を満たさなければならないことが分かります。

$$h(mn) = h(m) + h(n) \quad (6-2)$$

この関数方程式の解は、

$$h(n) = K \log n \quad (6-3)$$

の形を取ります。

そして、途中を省略しますが、次のシャノンの定理が得られます。

シャノンの定理：「不確かさの量」の合理的な測度に課した条件を満足するただ一つの関数 $H(p_1, \dots, p_n)$ は、

$$H(p_1, \dots, p_n) = -\sum_{i=1}^n p_i \log p_i \quad (6-4)$$

である。

この式で計算された量がエントロピーと名付けられています。熱力学で定義されたエントロピーと区別するために、情報エントロピーとも呼ばれます。

なお、 $I(p) = -\log p$ は情報量と定義されますので、エントロピーは情報量の期待値になっています。情報量は加法性を持たせるため $I(p) = -\log p$ で定義するという説明がある本もあります。より小さな確率でしか発生しない事象が発生したという情報量はより大きいものであるように定義されています。

情報量の単位は、対数の底に 2 を取ったときはビット、自然対数のときはニットと呼びます。これらの単位はデシベルと同じような種類の単位であって、物性値の単位ではありません。エントロピーの単位もビット（またはニット）で表されることになります。

そして、常にエントロピーが最大になるように確率分布を決めることが合理的で首尾一貫したものになります。このことの理論的証明はショアとジョンソン(1980)によって与えられています。これが最大エントロピー原理です。デシダレータとシャノンの条件は極めて似通っていますので、最大エントロピー原理から確率を決めるとデシダレータに合致するのは当然であるとも言えるでしょう。

7. 最大エントロピー法へ

7.1 不変量の原理

今、的の中心に向かって矢を射る時の射手の技量を離れて散らばるものと想定しましょう。このとき、矢が的に当たる確率分布はどのようなものになるのでしょうか。もし、人間の制御できる範囲の外ではばらつくものとしても、合理的な理由もなく偏るわけではないとすれば、何らかの条件が付いてきます。

まず、的に中心を原点に取った直行座標で考えてみます。すると直角方向の誤差は相互に独立であるべきですから、確率密度を $\rho(x, y)$ とすると、

$$\rho(x, y) dx dy = f(x) dx \cdot f(y) dy \quad (7-1)$$

と書ける筈です。

この的と矢の当たる点の関係は極座標でも表わせますので

$$\rho(x, y) dx dy = g(r, \theta) r dr d\theta \quad (7-2)$$

とも書ける筈です。的は丸くて θ に依存しない筈ですから、 $g(r, \theta) = g(r)$ です。従って、

$$f(x) f(y) = g(\sqrt{x^2 + y^2}) \quad (7-3)$$

この式はハーシェル(1850)の不変量条件式と呼ばれているものです。

さて、 $y=0$ とおくと、この式は $g(x)=f(x)f(0)$ に縮小します。そこで上式は次の関数方程式に成ります。

$$\log \frac{f(x)}{f(0)} + \log \frac{f(y)}{f(0)} = \log \frac{f(\sqrt{x^2 + y^2})}{f(0)} \quad (7-4)$$

しかし、この一般解は明らかです。つまり x の関数 y の関数はただ $x^2 + y^2$ の関数です。唯一つの可能性は $\log[f(x)/f(0)] = ax^2$ ということです。もし a が負

であるときにのみ私たちは正規化出来る確率を持ちます。そしてそのとき、正規化は $f(0)$ を決めます。そこで一般解は次の形式のみを持つことが出来ます。

$$f(x) = \sqrt{\frac{\alpha}{\pi}} e^{-\alpha x^2}, \quad \alpha > 0 \quad (7-5)$$

この式には一つの未決定のパラメータがあります。ハーシェルの不変量条件式を満たすのは 2 次元確率密度のもので、円の対称ガウス分布です。

$$\rho(x, y) = \frac{\alpha}{\pi} \exp[-\alpha(x^2 + y^2)] \quad (7-6)$$

10 年後、ジェームス・クラーク・マックスエル(1860)はこの同じ論証の 3 次元版を気体の分子速度のための確率分布 $\rho(v_x, v_y, v_z) \propto \exp[-\alpha(v_x^2 + v_y^2 + v_z^2)]$ を見つけるために用いました。そしてこれは動力学理論と統計力学の基礎的な「マックスエルの速度分布則」として物理学者に良く知られるようになりました。

ハーシェル - マックスエル論証は、一般には両立しない二つの定性的条件がちょうど一つの定量的分布に対してのみ適合するものとなり、従ってそれらが唯一的に定まるので特に美しいのです。アインシュタイン(1905)はローレンツ変換則を相対理論での彼の二つの定性的仮定から演繹するために同じ種類の論証を用いました。

ハーシェル - マックスエルの誘導は、それが他の問題でも同じように上手く使えそうな幾何学的不変特性のみを考えることで成立しています。実際には確率理論を少しも使用していないということでも経済的です。

[演習問題 4] 誤差の分布形状

今、的の中心に向かってライフル銃を撃つものとする。特に癖がない銃で多くの人が撃つものとする、弾が的にあたる場所の分布形状を求めよ。((7-6) 式を導きなさい。)

7.2 ガウスの誘導

ガウス自身は尤度が最大になるような分布式を求める方法で有名なガウス分布を求めました。ここはジェインズの説明を第 7 章からそのまま引用して示します。

「最大尤度によって $(n+1)$ 個の観測 $(x_0 \dots x_n)$ からパラメータ θ の位置を推定します。もし標本抽出分布の要素が $p(x_0 \dots x_n | \theta) = f(x_0 | \theta) \dots f(x_n | \theta)$ ならば、尤度式は、

$$\sum_{i=0}^n \frac{\partial}{\partial \theta} \log f(x_i | \theta) = 0 \quad (7-8)$$

又は、次のように書いて、

$$\log f(x | \theta) = g(\theta - x) = g(u) \quad (7-9)$$

最大尤度推定値 $\hat{\theta}$ は次式を満足します。

$$\sum_i g'(\hat{\theta} - x_i) = 0 \quad (7-10)$$

さて、直感はその推定値は観察の数学的平均でもまたあるべきだと私たちに暗示するかもしれません。

$$\hat{\theta} = \bar{x} = \frac{1}{n+1} \sum_{i=0}^n x_i \quad (7-11)$$

しかし、(7-10)と(7-11)は一般的には両立しません[(7-11)は(7-10)の一つの根ではありません]。それにも拘わらず、たった一つの観測 x_0 が非負であるという、一つの可能な標本を考えましょう。(7-11)で次のように置くなれば、

$$x_0 = (n+1)u, \quad x_1 = x_2 = \cdots x_n = 0, (-\infty < u < \infty) \quad (7-12)$$

そのとき、 $\hat{\theta} = u, \hat{\theta} - x_0 = -nu$ で、ここで(7-10)は $g'(-nu) + ng'(u) = 0, n=1,2,3,\dots$ となります。 $n=1$ の場合は $g'(u)$ が反対称関数 $g'(-u) = -g'(u)$ でなければならないことを告げています。そこでこれは、

$$g'(nu) = ng'(u), \quad (-\infty < u < \infty), \quad n=1,2,3,\dots \quad (7-13)$$

明らかに、たった一つの可能性は線形関数です。

$$g'(u) = au, \quad g(u) = \frac{1}{2}au^2 + b \quad (7-14)$$

(7-9)によって後方に変換すると、正規化可能な分布は再び a が負であることを要求します。そして正規化はそのとき定数 b を定めます。標本抽出分布は次の形式を持たなくてはなりません。

$$f(x | \theta) = \sqrt{\frac{\alpha}{2\pi}} e^{-\frac{1}{2}\alpha(x-\theta)^2}, \quad (0 < \alpha < \infty) \quad (7-15)$$

(7-15)は特別な標本を仮定して誘導されましたので、(7-15)は最尤推定値と標本平均が等しいための必要条件であるとしてのみこのように先まで示しました。逆にもし(7-15)が満たされると、そのとき尤度式(7-8)は常に唯一の解(7-11)を持ちます。ですから(7-15)はこの協定に対して必要十分条件です。ただ一つの自由度はまだ明示していない尺度パラメータ α です。」

7.3 最大エントロピー法

今、エントロピー式を簡単な数値で当たってみます。

$$H(p_1, \dots, p_n) = - \sum_{i=1}^n p_i \log p_i \quad (7-16)$$

最も簡単な場合は $n=2$ の時です。対数の底は 2 に取ることになります。すると、

$p_1=p_2=1/2$ のときは $H=1$ 、

$p_1=1/4$ 、 $p_2=3/4$ のときは $H=0.811$ 、

$p_1=1/8$ 、 $p_2=7/8$ のときは $H=0.544$ 、が得られます。

確かに等確率のときがエントロピー最大になっていることが判ります。この結果は、エントロピー式は不確かさが最も大きい状態はどのようなものかと推論した結果得られた式ですから当然のことです。

さて、何も情報が無いときは、等確率の原理で確率を割り当てればエントロピーは最大になることが保証されます。つまり、等確率の原理で確率を決めることは最大エントロピー法の特別な場合です。この場合は取り得る状態の数を求めることで確率が決められます。さいころ振りや、コイン投げの問題は殆どこのような場合でしょう。

しかし、一般に少し関連情報を持っているときにエントロピー最大になるように確率分布を決める方法はやっかいです。問題ごとに事情が異なるからです。米国では、1978 年から毎年、最大エントロピー法のシンポジウムが開催されています 8)。このシンポジウムの設立当初の主メンバーは E.T. ジェインズでした。

8 . 信頼度計算の見直し

新しい信頼度計算の求め方は、パラメータごとに推定の信頼性水準を設定するのでなく、あくまでも最終命題に対して信頼度を割り当てるという態度を取ることです。分布の形状など、まだまだ多くの仮定は置かざるを得ないでしょう。この考え方は、正統派統計学の中にも見られなくはなかったのです。正規分布を仮定するが平均も分散も未知であるとき、利用できるデータが少ないときに t-分布を使います。t-分布は平均と分散を独立に推定するものではありませんでした。

信頼性ブロック図は独立性が保たれていると考えられる範囲に止めることが必要です。もし、独立性がないならば、細分化して考えるよりも全体で考えた方が良いでしょう。このことに注意して問題ごとに対処すべきでしょう。

この方向で現行方式を修正した信頼度割り当ての手順を私は機械工学便覧に提案しております 9)。これを以下に紹介します。

8.1 統計データが利用できるとき

民間機の事故発生確率などのように、既に多くのデータが利用可能な場合は、その統計データから、命題が真であった件数 n の全件数 N に対する比率（相対頻度）をその確率とします。

$$P = n / N \quad (8-1)$$

正確に求めようとするあまり、母集団の取り方を厳密にしすぎると利用できる標本数が小さくなってしまいますから命題は出来るだけ単純なものにすべきです。標本数はどのような条件を持って同一の母集団と見るかに左右されます。

8.2 属性試験データ

データが少ない場合は次のようにして求めます。まず、試験結果が成功か失敗か、取りうる状態の数が二つの場合しかないときその試験は属性試験と言います。全体の試験数 n に対して成功数が r であったという情報を得たとき、確率 P はラプラスの連続則で求めます。

$$P = (r + 1) / (n + 2) \quad (8-2)$$

この式は実質的に前の式を包含しています。

8.3 パラメトリック・データ（大きな標本の場合）

材料の破壊強度試験のようにデータの分布を仮定することが出来る場合には、属性試験に比べて等価的に大きな標本数の情報として確率を求めることが出来ます。（実は「仮定することが出来る場合には」が、ジェインズのいう MPF なのですが、急には習慣を変えられないものです。）

それは数値データから分布のパラメータを推定することにより任意の命題に対する確率を決めることができるからです。

ここでは、母集団として正規分布が仮定できる場合を例に取ります。

x を確率変数とすると、正規分布の密度式は次式で表されます。

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{(x - \mu)^2}{2\sigma^2}\right\} \quad (8-3)$$

命題： $\{x < x_0\}$ が真である確率は次式で定められます。

$$P = \int_{-\infty}^{x_0} f(x) dx \quad (8-4)$$

データ数が十分多い ($n > 30$) とき、平均 μ と分散 σ^2 の推定値として、それぞれ標本平均 $\bar{x}$ と標本分散 s^2 （不偏分散）をそのまま使います。

$$\bar{x} = (x_1 + x_2 + \cdots + x_n) / n \quad (8-5)$$

$$s^2 = \{1/(n-1)\} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (8-6)$$

つまり、分布が確定すると命題 $\{x < x_0\}$ に対する確信の度合いは上述の式で求められます。この確率値が計算または正規分布表を利用して求められます。必要ならば、正規分布表に限らず既存の数表はインターネットで容易に見つけることが出来るでしょう。

8.4 パラメトリック・データ（小さな標本の場合）

正規分布であるということと、データ $\{x_1, x_2, \dots, x_n\}$ を見たということから次に得る x が $\{x < x_0\}$ であることに対する確信の度合いを求めることとなります。

データ数が少ないとき、二つのパラメータ、平均 μ と分散 σ^2 に対する推定値 θ_1 と θ_2 について、事前分布として θ_1 と $\log \theta_2$ がそれぞれ独立に区間 $(-\infty, \infty)$ において一様分布であるとしします。正であることが判っているパラメータに対してその対数が一様分布であるとするのがデシダレータの要求を満足するものであることはポーリャにより示されています。我々が知りたいのは x に対する確信をどのような数値とすべきかです。そこでデータを見た後の x に対する事後分布について次の定理を導きます。

定理「データ $\{x_1, x_2, \dots, x_n\}$ が正規分布 $N(\theta_1, \theta_2)$ からの n 個の無作為標本であるとする、 x に対する事後分布は $\sqrt{n}(x - \bar{x})/(\sqrt{n+1}s)$ が自由度 $v = n-1$ の t -分布に従う。」

この定理から、目標とする任意の値 x_0 に対する確率 $P\{x < x_0\}$ を求めることが出来ます。ここで、 s は前述の式で計算します。但し、 n は 2 以上でなくてはなりません。

x は正規分布をすると仮定できるものの、平均 μ も分散 σ^2 も未知の場合を考えています。分布が仮定できることから何も情報がない事前分布として x でなく μ と σ^2 に対する事前分布が一様であるとするのです。平均値 μ は数直線上どのような値でもあり得ます。また、分散は正数であればどのような値でもあり得ます。このことから、上述のように事前分布を取ることが妥当なのです。

なお、同じ条件のもとに、 θ_1 の事後分布としては、 $n^{1/2}(\theta_1 - \bar{x})/s$ が自由度 $v = n-1$ の t -分布に従うこと、及び θ_2 の事後分布としては、 $(n-1)s^2/\theta_2$ が自由度 $v = n-1$ の χ^2 分布に従うことが同じ方法で証明されています 10)。

証明 θ_1 と θ_2 の同時事後密度はこれらが互いに独立であることから、 θ_1 、 θ_2 それぞれの密度の積で与えられる。ここで前者 θ_1 の密度は一定であり、後

者は θ_2^{-1} に比例する。よって同時事前密度は θ_2^{-1} に比例する。一方、尤度は、

$$P(X|\theta_1, \theta_2) = (2\pi\theta_2)^{-n/2} \exp\left[-\sum_{i=1}^n (x_i - \theta_1)^2 / 2\theta_2\right] \quad (8-7)$$

故に、ベイズの定理により、同時事後密度は同時事前密度と尤度の積に比例するから、

$$\begin{aligned} \pi(\theta_1, \theta_2|X) &\propto \theta_2^{-(n+2)/2} \exp\left[-\sum_{i=1}^n (x_i - \theta_1)^2 / 2\theta_2\right] \\ &= \theta_2^{-(n+2)/2} \exp\left[-\{(n-1)s^2 + n(\bar{x} - \theta_1)^2\} / 2\theta_2\right] \end{aligned} \quad (8-8)$$

$$\begin{aligned} \sum_{i=1}^n (x_i - \theta_1)^2 &= \sum_{i=1}^n (x_i - \bar{x} + \bar{x} - \theta_1)^2 \\ &= \sum_{i=1}^n (x_i - \bar{x})^2 + n(\bar{x} - \theta_1)^2 = (n-1)s^2 + n(\bar{x} - \theta_1)^2 \end{aligned} \quad (8-9)$$

X を見た後での x に対する密度 $\pi(x|X)$ は、 x 、 θ_1 、 θ_2 の同時分布の周辺分布として得られるから、

$$\pi(x|X) = \iint P(x|\theta_1, \theta_2) \cdot \pi(\theta_1, \theta_2|X) d\theta_1 d\theta_2 \quad (8-10)$$

$$\text{ここで、} \quad P(x|\theta_1, \theta_2) = (2\pi\theta_2)^{-1/2} \exp\left[-(x - \theta_1)^2 / 2\theta_2\right] \quad (8-11)$$

(8-8)、(8-11)式を(8-10)式に代入して計算する。積分する中身は、

$$P(x|\theta_1, \theta_2) \cdot \pi(\theta_1, \theta_2|X) = (2\pi)^{-1/2} \theta_2^{-\frac{n+3}{2}} \exp\left[-\{(x - \theta_1)^2 + (n-1)s^2 + n(\bar{x} - \theta_1)^2\} / 2\theta_2\right] \quad (8-12)$$

従って、定数項は省略して記すと、

$$\pi(x|X) \propto \int \theta_2^{-\frac{n+3}{2}} \exp\left[-(n-1)s^2 / 2\theta_2\right] \int \exp\left[-\{(x - \theta_1)^2 + n(\bar{x} - \theta_1)^2\} / 2\theta_2\right] d\theta_1 d\theta_2 \quad (8-13)$$

$$\text{ここで、} \quad (x - \theta_1)^2 + n(\bar{x} - \theta_1)^2 = (n+1)\left(\theta_1 - \frac{x + n\bar{x}}{n+1}\right)^2 + n(x - \bar{x})^2 / (n+1),$$

$$\begin{aligned} \pi(x|X) &\propto \int \theta_2^{-\frac{n+3}{2}} \exp\left[-\left\{(n-1)s^2 + \frac{n}{n+1}(x - \bar{x})^2\right\} / 2\theta_2\right] \times \\ &\quad \int \exp\left[-(n+1)\left(\theta_1 - \frac{x + n\bar{x}}{n+1}\right)^2 / 2\theta_2\right] d\theta_1 d\theta_2 \end{aligned} \quad (8-14)$$

内側の積分は通常の正規分布の積分であり、結局その値は $\theta_2^{1/2}$ に比例する。

$$\text{よって、 } \pi(x|X) \propto \int \theta_2^{-\frac{n+2}{2}} \exp\left[-\left\{(n-1)s^2 + \frac{n}{n+1}(x - \bar{x})^2\right\}/2\theta_2\right] d\theta_2 \quad (8-15)$$

ここで、次の公式を利用する。

$$\int_0^\infty e^{-\frac{A}{\theta}} \theta^{-m} d\theta = \frac{(m-2)!}{A^{m-1}} \quad (8-16)$$

$m = \frac{1}{2}(n+2)$ 、 $A = \frac{1}{2}\{(n-1)s^2 + \{n/(n+1)\}(x - \bar{x})^2\}$ とおくと、

$$\begin{aligned} \pi(x|X) &\propto \left[(n-1)s^2 + n(x - \bar{x})^2/(n+1)\right]^{-n/2} \\ &\propto \left[1 + \{n(x - \bar{x})^2\}/\{(n^2-1)s^2\}\right]^{-n/2} \end{aligned} \quad (8-17)$$

従って、 $t = \sqrt{n}(x - \bar{x})/(\sqrt{n+1}s)$ とおくと、 $v = n-1$ で x から t への変換のヤコビアンは定数だから、

$$\pi(x|X) \propto [1 + t^2/v]^{-(v+1)/2} \quad (8-18)$$

この式は t が自由度 $v = n-1$ の疑似 t -分布に従うことを示している。

(証明終わり)

この式的全積分値が 1 なるという条件から比例定数を定めます。

置換 $t^2/v = x/(1-x)$ を考えますと

$$\begin{aligned} \frac{dt}{dx} &= \frac{1}{2} v^{1/2} x^{-1/2} (1-x)^{-2/3} \\ 2 \int_0^\infty (1+t^2/v)^{-(v+1)/2} dt &= v^{1/2} \int_0^1 x^{-1/2} (1-x)^{v/2-1} dx \\ &= v^{1/2} \left(-\frac{1}{2}\right)! \left(\frac{1}{2}v-1\right)! / \left(\frac{1}{2}v-\frac{1}{2}\right)! \end{aligned} \quad (8-19)$$

この最後の結果は Beta 積分を利用しています。但し、 $m, n > -1$ です。なお、 $\left(-\frac{1}{2}\right)! = \sqrt{\pi}$ です。

$$\int_0^1 x^m (1-x)^n dx = \frac{m!n!}{(m+n+1)!} \quad (8-20)$$

このようにして、次の t-分布が得られます。

$$f_v(t) = \frac{\left[\frac{1}{2}(\nu-1)\right]!}{\sqrt{\nu\pi}\left(\frac{1}{2}\nu-1\right)!} \left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2} \quad (8-21)$$

事前分布の仮定の仕方がやや意図的に上記の定理が得られるようにした感がありますが、データの標本数が比較的多ければ事前分布の仮定はどのようなものであってもその影響は小さくなります。標本数が小さければ、分布形状の仮定とパラメータに対する事前分布の仮定とは同じように確率の決め方を左右します。いずれにしても標本数が極めて小さいときに、分布の両裾野域（小さな確率値）を求めるには無理があります。

大きな標本から推定したパラメータを使った分布であっても、その分布からあまりにも裾野域の確率を決めることは避けねばなりません。分布を仮定した時点で求める確率を仮定したようなものになるからです。上述の例では分布形状が正規分布であると仮定できる範囲を考慮する必要があります。要するに、30 個程度の標本数で 3 σ より離れた確率を云々することは出来ないでしょう。

表 - 1 t - 分布

ν (自由度)	α (確率)							両側 片側
	0.20	0.10	0.05	0.02	0.01	0.002	0.001	
1	3.08	6.31	12.7	31.8	63.7	318	637	
2	1.89	2.92	4.30	6.97	9.93	22.3	31.6	
3	1.64	2.35	3.18	4.54	5.84	10.2	12.9	
4	1.53	2.13	2.78	3.75	4.60	7.17	8.61	
5	1.48	2.02	2.57	3.37	4.03	5.89	6.87	
10	1.37	1.81	2.23	2.74	3.17	4.14	4.59	
15	1.34	1.75	2.13	2.60	2.94	3.73	4.07	
20	1.33	1.73	2.09	2.53	2.85	3.55	3.85	
30	1.31	1.70	2.04	2.46	2.75	3.39	3.65	
60	1.30	1.67	2.00	2.39	2.66	3.23	3.46	
120	1.29	1.66	1.98	2.36	2.62	3.16	3.37	
∞	1.28	1.65	1.96	2.33	2.58	3.09	3.29	

終わりに

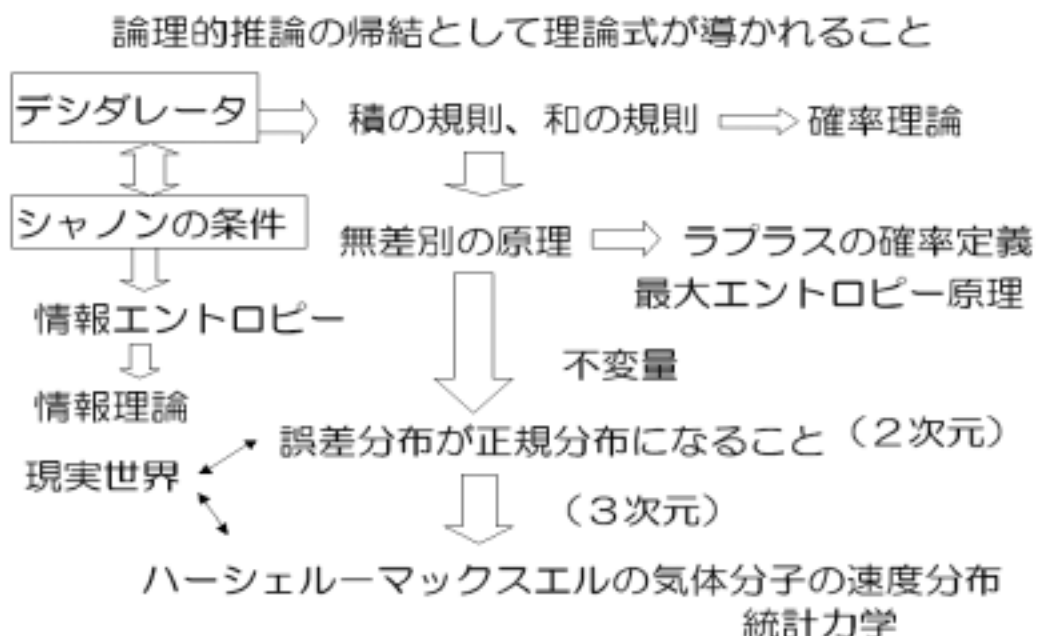
拡張論理としての確率理論

ベルヌーイ、ラプラスの確率理論は何故正統派統計理論に 150 年間も古典に追いやられてしまったのでしょうか。それは関数方程式が難しかったことと、いくつかのパラドックスに悩まされたからだと考えられます。パラドックスの多くは不用意に無限の概念を持ち込んだことにありました。

次頁の図は確率理論と情報理論が密接な関係を持ち、物理学の基礎とも流れの源が同じであることを示しています。デシダレータとシャノンの条件は見方を変えただけですが、いずれも人間にとって合理的であると理解できる方法で記述したものが自然界を説明することが、一見不思議なようにさえ思えます。

しかし、よく考えると自然界を説明する現行の物理学は人間が考えたものですから、人間が合理的で理解できる方法でのみ記述されたものに過ぎないから当然であるようにも思えます。宇宙のどこかにいるかも知れない異星人は、人類とは全く異なる方法で、例えば 4 次元世界で、物事を考え理解しているかも知れません。4 次元という概念ですら、人間が考えた人間にのみ理解できる程度の易しい概念かも知れません。

論理としての確率



文献

- 1) 西澤潤一、「独創性と信頼性技術」、第 28 回 信頼性品質保証シンポジウム、発表報文集、平成 10 年
- 2) E.T. Jaynes (2003), "Probability Theory: The Logic of Science",
- 3) P.G. ホーエル、初等統計学、浅井晃・村上正康共訳、培風館、改訂第 6 刷、昭和 46 年
- 4) 小野周、8 熱力学、岩波講座基礎工学 11、岩波書店、1969 年
- 5) 瀧保夫、3 確率統計現象、岩波講座基礎工学 1、岩波書店、1967 年
- 6) E.T. Jaynes (1973), "The Well-Posed Problem", Found Phys. 3, 477-493. Reprinted in Jaynes (1983).
- 7) 原 宣一、「少数の属性試験結果から得られる確信の度合い」、第 14 回信頼性シンポジウム、平成 13 年
- 8) <http://maxent23.org/>
- 9) 機械工学便覧（予稿）
- 10) D.V. リンドレー、確率統計入門 2、＜ベイズの方法による＞ 統計的推測、竹内啓、新家健精、共訳、培風館、1969

付録 - 演習問題の解答

1. 真空管を使ったコンピュータの信頼度

真空管を 1 万本使ったコンピュータがある。真空管以外の箇所は故障しないが、真空管が 1 本でも故障したらコンピュータはまともに作動しないものとする。このコンピュータが完成後最初の 1 時間、まともに作動する信頼度を求めよ。但し、このコンピュータに使われた真空管はどれも同じ型のもので、その寿命は過去の実績データを見ると故障までの時間の平均が 5000 時間で、分布の形状は指数分布に極めて近いことが判っているものとする。

ヒント：指数分布の密度関数は $f(x) = \lambda e^{-\lambda x}$

指数分布の期待値は $E(x) = \frac{1}{\lambda}$

指数分布の分散は $V(x) = \frac{1}{\lambda^2}$

真空管 1 本が 1 時間作動する信頼度を先に求め、1 万本の真空管が故障するのはそれぞれ独立であると考えよ。

（解答）

$$\frac{1}{\lambda} = 5000 \quad [h] \quad \text{故に、} \lambda = 2 \times 10^{-4} \quad [h^{-1}]$$

最初の 1 時間で真空管 1 本が故障する確率 F は、

$$F = \int_0^1 f(x) dx = \int_0^1 \lambda e^{-\lambda x} dx = \left[-e^{-\lambda x} \right]_0^1 = 1 - e^{-\lambda}$$

従って、最初の 1 時間まともに作動する信頼度 R は、

$$R = 1 - F = e^{-\lambda}$$

1 万本の真空管の故障は独立であるとする、全体の信頼度 R_{total} は

$$R_{\text{total}} = R^{10000} = (e^{-\lambda})^{10000} = e^{-2} = \frac{1}{2.718^2} = \frac{1}{7.39} = 0.14$$

答え 求める信頼度は 0.14

(検討)

- ・ 指数分布は最初の部分ほど故障する密度が高い。正規分布に比べると極めてばらつきが大きい分布であることに注意する必要がある。
- ・ 同じ会社で作った真空管は独立であると言えるであろうか。

2 . 確率計算

(1) さいころ 1 個を 4 回振って 1 度は 6 が出る確率を求めよ。

(解答)

方針： 4 回とも 1 が出ない確率を求めて 1 から引く。

4 回とも 1 が出ない確率は

$$\left(\frac{5}{6} \right)^4 = 0.482$$

$$1 - 0.482 = 0.518$$

答えは 0.518

(凡人の解答)

1 回目で 6 が出る確率 $p_1 = 1 / 6 = 0.1667$

2 回目で 6 が出る確率 $p_2 = (1 - p_1) \times 1 / 6 = 5 / 36 = 0.1389$

3 回目で 6 が出る確率 $p_3 = (1 - p_1 - p_2) \times 1 / 6 = 25 / 216 = 0.1157$

4 回目で 6 が出る確率 $p_4 = (1 - p_1 - p_2 - p_3) \times 1 / 6 = 125 / 1296 = 0.0965$

従って 4 回振って 1 度は 6 が出る確率は、 $p_1 + p_2 + p_3 + p_4 = 0.518$

答えは 0.518

(2) さいころ 2 個を同時に 24 回振って、1 度は 6 - 6 が出る確率を求めよ。

(解答)

方針：24 回とも 6-6 が出ない確率を求めて 1 から引く。

1 回振って 6-6 が出ない確率は $35/36$ である。

従って、24 回とも 6-6 が出ない確率は、

$$\left(\frac{35}{36}\right)^{24} = 0.509$$

$$1 - 0.509 = 0.491$$

答えは 0.491

(凡人の解答) ギブアップ！

3 . 確率計算と勝ち目

二人で交互にコインを投げ、最初に表が出た方が勝ちとする。先に投げる人が
勝つ確率を求めよ。

(解答)(凡人の答え)

1 回目で勝つ確率 = $1 / 2$

2 回目で勝つ確率 = $1 / 2 \times 1 / 2 \times 1 / 2 = 1 / 2 \times 1 / 4$

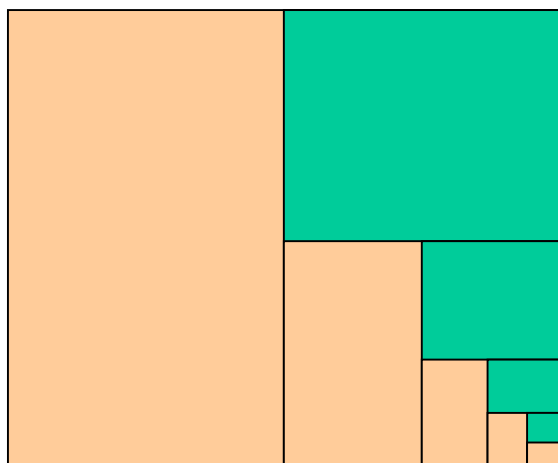
3 回目で勝つ確率 = $1 / 8 \times 1 / 2 \times 1 / 2 = 1 / 2 \times 1 / 16$

.....

従って、 $p = 1 / 2 (1 + 1 / 4 + 1 / 16 + \cdots)$

$$= 1 / 2 (1 / (1 - 1 / 4))$$

$$= 1 / 2 \times 4 / 3 = 2 / 3 \quad \text{答え} \quad 2 / 3$$



(解答)(頭の良い人)

方針：後者に投げる番が回って来たならば、その人が勝つ確率は最初に投げる

人が勝つ確率に等しいはずである。

最初に投げる人が勝つ確率を p と置き、後で投げる人の勝つ確率を q とすると、 $p + q = 1$ で、後者に投げるチャンスが巡ってくるのは $1 / 2$ の確率が存在する。

従って、 $q = 1 / 2 p$ の関係がある。

$$1 - p = 1 / 2 p$$

$$1 = 3 / 2 p$$

$$\text{故に、 } p = 2 / 3$$

$$\text{答え} \quad 2 / 3$$

勝ち目（オッズ）は前者と後者の勝つ確率の比を言う。

$$\text{従って、 } 2 / 3 : 1 / 3 = 2 : 1$$

先にコインを投げる人のオッズは $2 : 1$

4 . 誤差の分布形状

今、的の中心に向かってライフル銃を撃つものとする。特に癖がない銃で多くの人が撃つものとする、弾が的にあたる場所の分布形状を求めよ。

ヒント：(ハーシェルの方法)

まず、的の中心を原点に取った直交座標で、微小面積要素 $d x d y$ の領域に弾の中る確率は密度関数 $\rho(x, y)$ として、 x 方向と y 方向で異なる理由が無いから、

$$\rho(x, y) d x d y = f(x) d x f(y) d y \quad (1)$$

と書ける筈である。

一方で、同じく的の中心を原点に極座標で表わすと、微小面積要素 $r d r d \theta$ に中る確率は、

$$\rho(x, y) d x d y = g(r, \theta) r d r d \theta \quad (2)$$

ここで、 θ によって変わることはないので、

$$g(r, \theta) = g(r) \quad (3)$$

そして、 $r = \sqrt{x^2 + y^2}$ から、

$$f(x) f(y) = g(\sqrt{x^2 + y^2}) \quad (4)$$

ここで $y = 0$ とおくと、(4) 式は

$$g(x) = f(x) f(0) \quad (5)$$

従って、 $g(\sqrt{x^2 + y^2}) = f(\sqrt{x^2 + y^2}) f(0)$ ですから、

(4) 式の両辺を $f(0)$ の二乗で割って対数を取れば、

$$\log \frac{f(x)}{f(0)} + \log \frac{f(y)}{f(0)} = \log \frac{f(\sqrt{x^2 + y^2})}{f(0)} \quad (6)$$

この関数形状 f が満たすべき一般解は x の関数 z の関数が、ただ $x^2 + y^2$ の関数である。これを満たす関数は、以下のものしかないと考察しなさい。

$$\log \frac{f(x)}{f(0)} = ax^2 \quad (7)$$

$$\text{つまり、} \quad f(x) = f(0)e^{ax^2} \quad (8)$$

正規化条件から、定数を決めて密度関数 $\rho(x, y)$ を求めなさい。

(解答)

(8) 式は y についても言えるので、

$$f(y) = f(0)e^{ay^2}$$

(1) 式から、

$$\rho(x, y) = f(x)f(y) = f(0)^2 e^{a(x^2 + y^2)}$$

正規化条件は、

$$\iint \rho(x, y) dx dy = 1$$

$$\text{従って、} \quad \iint e^{a(x^2 + y^2)} dx dy = \frac{1}{f(0)^2}$$

$$x^2 + y^2 = r^2 \text{ から、} dx dy = r dr d\theta \quad \text{で}$$

$$a = -\alpha, \quad \alpha > 0 \text{ とすると}$$

$$\frac{1}{f(0)^2} = \int_0^\infty \int_0^{2\pi} e^{-\alpha r^2} r dr d\theta = 2\pi \int_0^\infty e^{-\alpha r^2} r dr = 2\pi \left[-\frac{e^{-\alpha r^2}}{2\alpha} \right]_0^\infty = \frac{\pi}{\alpha}$$

$$\text{従って、求める} \quad \rho(x, y) = f(x)f(y) = \frac{\alpha}{\pi} e^{-\alpha(x^2 + y^2)}$$

$$\text{または、} \quad \rho(x, y) = \frac{\alpha}{\pi} \exp[-\alpha(x^2 + y^2)] \quad \leftarrow \text{ 答え}$$

$\alpha > 0$ の任意常数が残っている。