



統計と確率（2）

<https://l-hospitalier.github.io>

2021. 3



R.A. フィッシャー

感染対策の基礎知識

#280

【確率過程】世の中の現象には（古典）物理現象の弾丸の位置と速度のように時刻 t における状態 (x, y, h, v) が起点の時刻 t_0 における弾の速度 x と方向だけで決まる $y = f(x, t_0, t)$ のような**純決定論的モデル**とサイコロの目のように試行後の状態が y である確率が与えられる**確率論的モデル**の2つが考えられる。確率モデル（確率空間とも）には花粉のブラウン運動のように時刻によって変化する連続時間型確率過程とサイコロの目のように離散時間型確率モデルがある。**確率過程**（stochastic process）で次の状態が現在の状態のみで決定され、（離散時間では1つ前の）過去の状態の影響を受けないモデルを研究者¹にちなみ**マルコフ過程**と呼ぶ。代表は水に浮かぶ花粉が周辺の水分子の熱運動による衝突で位置を変えるブラウン運動（ランダムウォーク）。確率過程では推測が複雑になりすぎるのを避けるためマルコフ過程の仮定をすることが多い。**エルゴード性**は1個のブラウン運動する花粉のある時点の位置、方向、速度、加速度などの量を十分長い時間計測するのは大変なので、花粉の数を無数（可算無限個）に増やして同時に観測すれば両方のデータ集合はやがて一致するのでは？という仮説。エルゴードは**L ボルツマン**が1884年の論文でギリシャ語の *ergon*（仕事量）と *hodos*（経路）から作成。世界各地の海で波（高さ、周期、速度）を観測した結果と、一か所で数年観測した結果がほぼ等しいのでは？という仮説。この仮説は各分子の運動に関する物理量の（長）**時間平均**を**空間平均**（集合平均）で置き換え可能というもの。【記述統計】確率過程ではなく**決定論的過程**（deterministic process）であっても誤差によりランダムな現象に見えることがある。独立な観察を複数回行って現象の相関関係を見出す方法は、1919年 **RA フィッシャー**がロザムステッド農事研究所に赴任、農作物の収量²が①品種 ②肥料 ③畑（気温や日照）等の**要因による変動**と、確率過程の**誤差変動**、要因間の**交互作用の変動**を分離する方法を提唱して近代統計学が成立。分散比の分布（スネデカーによりフィッシャーを記念して**F分布**と命名）や小標本の場合の**W ゴセット**（筆名は **Student**）による **t分布**などが研究され、分散分析（ANOVA、**Analysis of Variance**）として成立。基本的手法は、データの分散成分の平方和を分解、誤差による変動から要因効果による変動を分離。次に平方和を自由度で割ることで平均平方を算出。そして要因効果（または、交互作用）によって説明される平均平方を分子、誤差によって説明される平均平方を分母とすることで**F値**を計算する（**F検定**）。各効果の有意性について有意水準を設けて判定。誤差は偶然による。これらは「**実験計画**」として実験科学者必須の教養となった。確率統計については事故³で早世（39歳）の京大助教授小針規宏の「確率・統計入門（岩波）」がお勧め。抜粋すると「確率というのは先験的にあるものではなく人間が主体的にきめるもの。《真理》などは言語の上にしか存在しない。何を仮定すれば何が結論されるか、その論理の連鎖が数学。ベルヌーイ試行（コイントス）の確率は $1/2$ ではない。わからないから $1/2$ としましょう、そうすると実験とよく合うようです。が正しい」などの記述。



L. ボルツマン



W ゴセット



小針規宏

¹アンドレイ・マルコフ(1856-1922) ²教養部の近代統計学で「ユリの花卉の長さ」と聞いた記憶があるが検索できず。Fisherは愛煙家で弟子たちが煙草と肺癌の明瞭な相関を示してもその因果関係を認めず、「たまたま癌の遺伝子とタバコ好きの遺伝子が同じなのかもしれない」と。ネオダーウィニズムの提唱者の一人で優生学者。 ³階段から転落とのことだが同僚の広中平祐は序文で学生運動に対する同情と共感が彼の命を縮めた？と。