感染対策の基礎知識

#234

ベイズ推定とベイズの定理

— 最尤法と新型コロナの PCR 検査 —

https://l-hospitalier.github.io 2020.4

【頻度主義の最尤（推定）法, Maximum Likelihood Estimation】 Ronald A Fisher により 1912~22 年に考案された頻度主義の確率推定法。 コイントスを 4 回したら表、裏、表、裏が出た。 表が出る事象を θ とすると、尤度[*1]（ゆうど）L の母数 θ による尤度関数は $L(\theta) = \theta^2 (1-\theta)^2$。 この式の最大値を与える θ をコイントスの確率とする方法。 計算を楽にするため対数をとると $\ln(L(\theta)) = 2 \ln\theta + 2 \ln(1-\theta)$。 この式を θ で微分して $\frac{d}{d\theta} \ln L(\theta) = \frac{2}{\theta} - \frac{2}{1-\theta}$。 傾きが 0 になる θ = 0.5 が最大の尤度となる。 または実際の値を求めてみると θ = 0.5 の時 $0.5^4 = 0.0625$。 θ = 0.4 と θ = 0.6 の時はいずれも $0.4^2 \times 0.6^2 = 0.0576$ で θ = 0.5 が L(θ)の最大値を与えることが分かる。 この時コイントスの確率は尤度関数 L(θ)の最大値を与える θ = 0.5 であると考える。 3 回トスで表、裏、表の時の尤度は $\theta^2(1-\theta)^1$。 $\frac{d}{d\theta} \ln L(\theta) = \frac{2}{\theta} - \frac{1}{1-\theta}$ が 0 になるのは $3(1-\theta) = 1$、$3-3\theta = 1$ から $2 = 3\theta$ で θ = 2/3 の時、つまり表が出る確率は 2/3。 これは極めて直感と一致する内容を数学的に記述したもの。 最尤法ではベイズ推定法と異なり事前確率設定はない。 試行回数がある程度必要となる。 数学は仮定（公理）を認めれば後は正しい論理の連鎖。 **【理由不十分の原則】**事前確率が不明の時は**等確率の仮定を採用する。** 例えば日本人の新型コロナ感染の事前確率が不明な時は 50%と仮定する。 理由不十分の原則を公理にするのは数学（論理）として不適切。 **Fisher はベイズ推定ではなく最尤推定を使うよう強く勧めた**ので、ベイズ推定が受け入れられたのは 21 世紀に入ってから。 **【ベイズの定理、主観確率】**は（事前確率関係の仮定を除けば）数学的に確立した方法でモンティ・ホール問題などの記述や解法は容易になる。 壺から球を 1 つ取り出す時、赤玉と白玉の確率（割合）を考える。 球を壺に入れる時全体に白が多いように見えた。これが事前確率。 ベイズ推定では適切な事前情報があれば試行回数を大幅に減らせる。 全体として**白っぽいなら事前確率 P(A)を赤 0.1 か 0.2 で始めようというのがベイズ推定（情報**がない時は「理由不十分の原則」で事前確率 0.5）。 このベイズ推定法は確率や条件付き確率を P(B) > 0 の時 $P(\mathbf{A}|\mathbf{B}) = \frac{P(B|A)\,P(A)}{P(B)}$ （ベイズの定理）が成立するのを使って推定精度を改善してゆく方法でラプラスによって紹介された。今日本人の罹患率 3.3%（330 万）の糖尿病[*2] を考える。 検査は DM で陽性と判定される確率が 95%、健常人が陰性と判定される確率は 80%。 ある人がこの検査を受けて陽性の時 DM の確率はいくら？ 陽性の事象を A、陰性の事象を事象 A^c（事象 A の余事象）、DM である事象を B_1、DM でない事象を B_2 とする。 ベイズの展開式①を使い**事後確率** $P(B_1|A)$を求める。 各確率を以下に示す。

$$P(B_1|A) = \frac{P(B_1)P(A|B_1)}{P(B_1)P(A|B_1) + P(B_2)P(A|B_2)} \cdots ①$$

DM の有病率 : $P(B_1) = 0.033$、DM でない確率 : $P(B_2) = 1-0.033 = 0.967$、DM 罹患で検査が陽性の確率 : $P(A|B_1) = 0.95$、健常者が検査陰性の確率 : $P(A^c|B_2) = 0.8$、健常者の検査が偽陽性の確率 : $P(A|B_2) = 1-0.80 = 0.2$。 これらの値を①に代入すると、ある人が**偶然**この検査を受けて陽性であった時 DM である確率は 0.1394 = **13.9 %**。 15 %ぐらいなら DM を気にする必要はない？ あなたが小学生ならそうかも。 成人なら再検査！ このようにベイズ推定は事前確率設定の影響が大きい[*2]。 確率論の知識のない**専門医**が新型コロナウイルス PCR 試験を「全く症状のない人や、数日続いているだけの風邪の症状がある人の検査をしても、今の流行状況？では[*3] 偽陽性が多くなる」という内容の記事の日付は 2020/3/6。 この時点で国内の新型コロナ？ウイルス有病率の無作為標本抽出はなく**「理由不十分の原則」**に従うと等確率で有病率 50%。 1 千万人検査すると 1 万人偽陽性が出るから「検査の増加は不要」という（自家撞着の）結論が導かれる。 統計（確率）学は検査データの数を議論しない。 得られたデータをどう整理して提示するかだけ。

← **#233 で P エルデシュの学生がパソコンの乱数発生とモンテカルロ法でモンティ・ホール問題を解いたのは最尤法と考えられる**

[*1] 尤度は確率をパラメータ（母数）で記述したもの。 母数を母集団の（標本）数とする誤用が多い。[*2] 何故か臨床検査の場合は有病率 0 に近い疾患がとりあげられ、適用すべき「理由不十分の原則」による等確率（有病率 50%）が適用されない。 有病率 0 なら事後確率も 0 なので全ての検査は無意味！ この恣意性がベイズ推定を疑問視する人に「学問ではない！」と言わせる。 [*3] 誰も知らない頻度にこの表現を抵抗なく使うのは自分で思考チェックができない人。