最尤推定法とベイズ推定法

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

感染対策の基礎知識

#234

【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 4 回で表、裏、表、裏が出た（2 項分布）。 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 $\ln(L(\theta)) = 2 \ln\theta + 2 \ln(1-\theta)$。 θ で微分して $\frac{d}{d\theta} \ln L(\theta) = \frac{2}{\theta} - \frac{2}{1-\theta}$。 微分（傾き）が 0 になる θ = 0.5 が最大尤度となる。 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 であると推定する。 トス 3 回で表、裏、表の時の尤度は $\theta^2(1-\theta)^1$。 微分 $\frac{d}{d\theta} \ln L(\theta) = \frac{2}{\theta} - \frac{1}{1-\theta}$ が 0 になるのは 2(1-θ) = θ、2-2θ = θ から 2 = 3θ で θ = 2/3、つまり表が出る確率は 0.666…。 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が試行回数がある程度必要となる。 **【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）としては不適切。 **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 $P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$ （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。

← **#233** で **P** エルデシュの学生がパソコンで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に近いと考えられる

今日本人の糖尿病[*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。DM は 330 万人で 1 億の 3.3%。

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

罹患率：$P(B_1)$ = 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の時 DM である確率は①式に代入して 0.1394 = **13.9 %**。15 %以下だから DM の心配はない？ あなたが小学生ならそうかも。 成人なら当然再検査！ このようにベイズ推定は事前確率の影響が大きい[*2]。 確率論の知識のない**専門医**が新型コロナの 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（不明だから調査する）では偽陽性が多くなる」という内容の記事の日付は 2020/3/6。 この時点で都内の新型コロナウイルス感染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%。 有病率 0.044%[*3] と恣意的に低い事前確率を設定して多数の偽陽性者が出るので混乱するというのは自家撞着。 数学は仮定（公理）と正しい論理の連鎖。 統計・確率学は得られたデータをどう整理して提示するかの学問で特定の主張の擁護はしない。

[*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）数とするのは誤用。[*2] 何故か臨床検査の場合は有病率が低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%）は適用されない。 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] 都内 44 人を恣意的に 100 倍して事前確率を 4400/1000 万=0.044%。未知の有病率にこれを適用。