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 2020.3

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、θ の尤度*1 (確率、尤度関数 L(θ)) は $\theta^2(1-\theta)^2$。 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(θ)) = 2 lnθ + 2 ln(1-θ)。 ln (L(θ)) = 2lnθ + 2ln(1-θ) を θ で微分して $\frac{d}{d\theta}$ ln L(θ) = $\frac{2}{\theta}-\frac{2}{1-\theta}$ が 0 になる θ=0.5 が最大の尤度となる。 あるいは実際の値を求めてみると θ=0.5 の時 2(-1.386-1.386) = -2.772。θ=0.4 の時 2(-0.916-0.510)=-2.652。θ=0.6 の時 2(-0.510-0.916) = -2.652 で θ=0.5 が最大値を与えるのがわかる。 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 とみなす。 3 回トスで表、裏、表の時は尤度は $\theta^2(1-\theta)^1$。 $\frac{d}{d\theta}$ ln L(θ) = $\frac{2}{\theta}-\frac{1}{1-\theta}$ が 0 になるのは 3(1-θ) = 1、3-3θ = 1 から 2=3θ で θ=$\frac{2}{3}$ の時、つまり表が出る確率は 2/3。 これは極めて良く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**を必要としない。 試行回数がある程度必要となる。 数学は仮定 (公理) を認めれば後は論理の連鎖。 事前確率 (あるいは**「理由不十分の原則」**すなわち事前確率が不明時は**等確率仮定**、例えば日本人の新型コロナ感染率の事前確率が判らない時は **50%**と仮定して始める) を公理と認めるのは数学としては不適切に思える。 **フィッシャー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。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が**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して真値に近づくことができる。 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 とする) 。 このベイズ推定法は確率や**条件付き確率**を P(B) > 0 の時 $\mathbf{P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}}$ (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。 今日本人の罹患率 3.3% (330 万) の糖尿病を考える。 検査方法は陽性と判定される確率が 95%、逆に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 の事象を B_1、罹患していない事象を B_2 とする。 ベイズの定理①を使うと、求める確率は $P(B_1|A)$。 各確率を以下に示す。 DM の確率：$P(B_1)$=0.033、DM でない確率：$P(B_2)$=1-0.033=0.967、DM 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：$P(A^c|B_2)$=0.80、健常者が陽性の確率：$P(A|B_2)$=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 0.05627= 5.6 %。 5~6 %なら DM をあまり心配する必要はない？ あなたが小学生ならそうかもしれない。 成人なら再検査！ このようにベイズ推定は事前確率の影響が大きい。 確率論の知識のない専門医が「**事前確率が大事**」と新型コロナ 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***2 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。 この時点で国内の PCR 陽性の標本抽出はなく**「理由不十分の原則」**に従うと等確率で 50%が感染者。 統計確率の数学的知識がないと本来知りたい陽性率の推定に「**事前確率 (=自分の思い込み)** 」が**大事**。 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。

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

*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) 数としている誤用が多い。 *2 誰も知らない頻度にこの表現が抵抗なく出来るのは自己の思考をチェックすることのない人。