PCR 検査における統計学的考察

2020 年 6 月 7 日
(改訂　2020 年 6 月 30 日)

杉本末雄
立命館大学 名誉教授

本測位技術振興会の活動も含め, 新型コロナウイルスの感染に伴う社会的影響は甚大である．感染拡大の克服が重要であることについては論を待たない．そのために，PCR 検査，抗体検査などが，注目を集めている．

しかし，この検査には，本当は感染している (陽性である) のに，陰性と検査結果が出る (偽陰性)，また逆に，本当は感染していない (陰性である) のに，陽性と検査結果が出る (偽陽性) という，検査に伴う誤りがある．このことを常に考慮しておくことが大変重要であることを示したい．

本論文では，統計学見地から PCR 検査 (抗体検査なども含め) について考察する．PCR 検査については，その検査結果を，誤り確率を明確にしたうえで，被検査者に伝えることが大切であることを示す．そのために，検査対象者にとって，PCR 検査結果がどの程度信用できるかについて，統計学での (S1) ベイズの定理を用いて検証する．

また，ランダムサンプリングにより，全国で 1 万人，大阪府では 5 千人と，PCR 検査対象者を増加させ，感染率を推定したいとの現下の要請がある．ここでは，(S2) 比率の最尤推定法について考察し，単なる比率の推定問題として取り扱うことでは，推定精度に大きな問題が生じることを示し，PCR 検査の誤り確率の推定値が非常に重要であることを示す．(S1)，(S2) のいずれの場合においても，PCR 検査の誤り確率を，統計学的に考慮することが重要である．

1．確率・統計学の基礎

1.1 条件付き確率

A, B を不規則現象での事象とし，$P(B) > 0$ と仮定する．この時，事象 B が生じたという条件のもとでの，事象 A が生じる確率を $P(A|B)$ と記す．この時，$P(A|B)$ （事象 B が起こったという条件のもとでの事象 A の起こる確率）は

$$P_B(A) \equiv P(A|B) \equiv \frac{P(A \cap B)}{P(B)} = \frac{P(A, B)}{P(B)}$$

と定義する，あるいは

$$P(A \cap B) = P(A|B)P(B)$$

と定義する．また $P(A|B)$ は事象 B が与えられたもとでの A の条件付確率 (conditional probability of A given B) という．

条件付き確率の計算例

サイコロを 1 回投げたとき，
(a) 4 以下の目が出る (事象 A) 確率を求めよ．
(b) 投げた結果が奇数の目である (事象 B) ことを知ったもとで，4 以下の目が出る (事象 A) 確率を求めよ．
ただし，1 の目 ～ 6 の目は等確率 $\frac{1}{6}$ で生じるものとする．

(解)

サイコロを 1 回投げたとき，見本空間は $\Omega = \{1\text{ の目}, 2\text{ の目}, \cdots, 6\text{ の目}\}$ $\equiv \{\omega_1, \cdots, \omega_6\}$ となり，いま,「奇数の目がでる」という事象を B,「4 以下の目がでる」という事象を A とすると，$\omega_i \cap \omega_j = \phi\ (i \neq j)$ であるので

(a)

$$P(A) = P(\{\omega_1, \cdots, \omega_4\}) = P(\omega_1) + P(\omega_2) + P(\omega_3) + P(\omega_4) = \frac{4}{6} = \frac{2}{3}$$

は明らか，また

$$P(B) = P(\{\omega_1, \omega_3, \omega_5\}) = P(\omega_1) + P(\omega_3) + \cdots + P(\omega_5) = \frac{3}{6} = \frac{1}{2}$$

(b)

$$P(A \cap B) = P(\{\omega_1, \omega_3\}) = P(\omega_1) + P(\omega_3) = \frac{2}{6} = \frac{1}{3}$$

であるので，

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

以上の条件付き確率の定義から，次の定理を示すことができる．

1.2 ベイズの定理

ベイズの定理 (Bayes' Theorem (Rule)

いま，事象 $A_1, A_2, \ldots, A_n$ が互いに排反，かつそれらの合併（和）集合が見本空間 Ω となるとき，すなわち，事象 $A_1, A_2, \ldots, A_n$ は見本空間 Ω の分割であるとき，事象 B のもとでの条件付確率について

$$P(A_k|B) = \frac{P(A_k \cap B)}{P(B)} = \frac{P(A_k)P(B|A_k)}{P(B \cap \Omega)} = \frac{P(A_k)P(B|A_k)}{\sum_{j=1}^{n} P(A_j)P(B|A_j)}$$

が成立する．

上記の定理は，与えられた条件 (原因) となる事象 B のもとでの，事象 A_k が起こる (結果) の確率が，条件 (原因) と結果が逆転した確率 $P(B|A_j)$ の計算により求められることを示しており，原因確率の定理 (a theorem of probability of cause) とも呼ばれる．

2. コロナ感染と PCR 検査

ある地区（または，あるグループ）での，病気 B の患者の割合は，$\beta = 1\%$ といわれている．

病気の検査を行うと，病気 B であるのに病気と判断されない事象 K_{NB}(偽陰性) が起こる確率は 0.3（第 2 種の過誤; のんびり者の過誤; 1-感度）である．

逆に病気 B でないのに病気と判断する事象 K_B(偽陽性) が起こる確率は 0.01（第 1 種の過誤; あわて者の過誤; 1-特異度）である．

R さんは，検査の結果 K_B と判断された．

R さんが，病気 B である，確率を求めよ．

また検査の結果 K_{NB} と判断されたとき，R さんが，NB である確率を求めよ．

(陽性の先験確率: $\beta = 0.01$)，

偽陰性者の確率: 0.3,　偽陽性者の確率: 0.01, と仮定する．

(解)　$P(B) \equiv \beta = 0.01, P(K_{NB}|B) = 0.3, P(K_B|NB) = 0.01$ であるので，

$$
\begin{aligned}
& P(B|K_B) \\
= \ & \frac{P(K_B, B)}{P(K_B)} = \frac{P(K_B|B)P(B)}{P(K_B)} \\
= \ & \frac{P(K_B|B)P(B)}{P(K_B|B)P(B) + P(K_B|NB)P(NB)} \\
= \ & \frac{0.7 \times \beta}{0.7 \times \beta + 0.01 \times (1-\beta)} \\
= \ & \frac{0.7\beta}{0.01 + 0.69\beta} \approx 0.414 \quad (\beta = 0.01)
\end{aligned}
$$

$$
\begin{aligned}
& P(NB|K_{NB}) \\
= \ & \frac{P(K_{NB}, NB)}{P(K_{NB})} = \frac{P(K_{NB}|NB)P(NB)}{P(K_{NB})} \\
= \ & \frac{P(K_{NB}|NB)P(NB)}{P(K_{NB}|NB)P(NB) + P(K_{NB}|B)P(B)} \\
= \ & \frac{0.99 \times (1-\beta)}{0.99 \times (1-\beta) + 0.3 \times \beta} \\
= \ & \frac{0.99 - 0.99\beta}{0.99 - 0.69\beta} \approx 0.997 \quad (\beta = 0.01)
\end{aligned}
$$

病気 B の患者の割合 ($\beta = P(B)$) は，$\beta = 0.1\% = 10^{-3}$, ではなく，$10^{-2}\% = 10^{-4}$ などの場合は，以下の表 1 となる．

表 1: 先験確率 : β と　真の確率

先験確率 β	真陽性確率 $P(B\|K_B)$	真陰性確率 $P(NB\|K_{NB})$
0.2	0.9459	0.9296
0.1	0.8861	0.9674
0.01	0.4142	0.9969
0.001	0.0655	0.9997
0.0001	0.0070	1.0000

3. PCR 検査によるコロナ感染者率の推定

日本全国で，ランダムサンプリング (無作為抽出) により，抗体検査，PCR 検査が行われる予定である．しかし品質管理のために，不良品の比率 (p) の推定検査のように，検査における，良品，不良品の決定に誤りがなければ，サンプルサイズ，n に対して，不良品数が X=x であれば，最尤推定法では，$X \sim B(n,p)$ であり，X は 2 項分布:

$$P(X=x) = {}_nC_x p^x (1-p)^{n-x}$$

に従い，

$$\arg\max_p \ln[{}_nC_x p^x (1-p)^{n-x}] = \frac{x}{n}$$

となり，最尤推定値 $\hat{p}$ は

$$\hat{p} = \frac{x}{n}$$

となる．次に，検査結果に誤りがある場合の，比率の推定問題について考察する．

実際には，検査対象者の感染率を p_o とするが，検査での誤り確率を考慮すると検査結果で陽性となる確率は

$$\begin{aligned} p_K &\equiv p_o(1-e_{NB}) + (1-p_o)e_B \quad = (1-e_{NB}-e_B)p_o + e_B \\ &\equiv ap_o + e_B; \quad a \equiv (1-e_{NB}-e_B) \end{aligned} \tag{1}$$

となる．ただし，e_{NB} は，陽性者に対する偽陰性の起こる誤り確率，e_B は，陰性者に対する偽陽性の起こる誤り確率とする．このとき，ランダムサンプリングにより n 人に対して，PCR 検査 (あるいは抗体検査など) を行い，検査結果が陽性とになった人数を $X=x$ とする．このデータから，p_o の最尤推定値 $\hat{p}_o$ は

$$\hat{p}_o = \arg\max_{p_o} P(X=x;p_o)$$

あるいは，対数尤度関数を用いて．

$$\begin{aligned} \hat{p}_o &= \arg\max_{p_o} \ln P(X=x;p_o) \\ &= \arg\max_{p_o} \ln[{}_nC_x p_K^x (1-p_K)^{n-x}]; \quad (p_K \equiv p_o(1-e_{NB}) + (1-p_o)e_B = ap_o + e_B) \\ &= \frac{\frac{x}{n} - e_B}{1-e_{NB}-e_B} \end{aligned}$$

となる事を以下に示す．(1) 式より

$$\begin{aligned} & \frac{d}{dp_o} \ln[{}_nC_x p_K^x (1-p_K)^{n-x}] \\ = \; & x\frac{a}{ap_o+e_B} + (n-x)\frac{-a}{1-p_o-e_B} = 0 \end{aligned}$$

したがって，

$$x(1-ap_o-e_B) - (n-x)(ap_o+e_B) = 0$$

すなわち，

$$x - nap_o + ne_B = 0$$

より

$$\begin{aligned} \hat{p}_o &= \frac{x-ne_B}{n(1-e_{NB}-e_B)} \\ &= \frac{\frac{x}{n}-e_B}{1-e_{NB}-e_B} \\ &= \frac{\text{検査陽性率} - \text{偽陽性率}}{\text{感度} - \text{偽陽性率}} \end{aligned} \tag{2}$$

したがって，上記の感染率推定値 $\hat{p}_o$ は，偽陽性率 (=1-特異度)，感度, の推定値がしっかりと，推定できなければ，検査での陽性率 $\frac{x}{n}$ は，ほとんど意味をもたないことになり，検査対象者群の感染率の推定精度，あるいは推定精度の信頼区間も求めることができない.

すなわち，(2) 式より，PCR 検査の感度 (sensitivity)，特異度 (specificity) の正確な値を知ることが，感染率の推定のために, 最も重要となる.