データ解析

第12回: 統計的検定

渡辺澄夫

検定の光と影

実世界
観測データ
回帰・判別分析
推定と検定
解析方法
統計的検定
ベイズ法
・階層ベイズ法
時系列予測
因子・主成分
・クラスタ分析

復習:統計的検定とは

(1) 帰無仮説を決める。

(2) 対立仮説 を決める。

(3) データから実数への関数 $Y=f(X^n)$ を決める。

(4) 有意水準(危険率) ε を決める。

(5) 帰無仮説のもとで集合(棄却域) $\{ Y ; Y=f(X^n) \geqq a \}$ の確率が ε になるように a を決める。

(6) 対立仮説のもとでの棄却域の確率を検出力という。

(7) 実際のデータ X^n に対して $Y=f(X^n)$ を計算して、Yが棄却域に入ったら対立仮説を取り、そうでないときは帰無仮説を取る。

最強力検定

検定が強力であるとは

定義. データ X^n の二つの関数 $f(X^n)$ と $g(X^n)$ があるとする。
ふたつの関数によって作られる検定を F, G とする。
命題「F, G の有意水準が同じときは、F の検出力は G の検出力より大きい」が成り立つとき、F は G よりも強力であるという。

どんな関数から作られる検定よりも F が強力であるとき、F のことを最強力検定であるという。

(注意) 関係「A はBよりも強力である」は、検定の集合全体の上の順序関係を定めているが、この順序によって検定全体の集合は全順序ではない(二つの検定を比較できないこともある)。

尤度比検定は最協力検定

<u>定義</u>. $X^n=\{X_1,X_2,\ldots,X_n\}$ を確率変数の集合とする。

帰無仮説 :「X^n は $q(x)$ から独立に得られた」

対立仮説 :「X^n は $p(x)$ から独立に得られた」

このとき関数

$$L(X^n) = \Pi_i\, p(X_i) \;/\; \Pi_i\, q(X_i)$$

で定まる検定のことを尤度比検定という。

<u>定理(ネイマン・ピアソンの補題)</u>. 帰無仮説と対立仮説が上記のように１個の確率分布であるときには尤度比検定は最強力検定である。

（注意） $\log L(X^n)$ を用いても等価である（棄却域は同じ）。

証明

$\{L(X^n), a\}$ を尤度比検定とし $\{M(X^n), b\}$ を任意の検定とする。

棄却域を $A=\{X^n ; L(X^n) \geqq a\}$ および $B=\{X^n ; M(X^n) \geqq b\}$ とする。

集合 S の帰無仮説と対立仮説における確率を $Pr(S|H_0)$、$Pr(S|H_1)$ と書く。

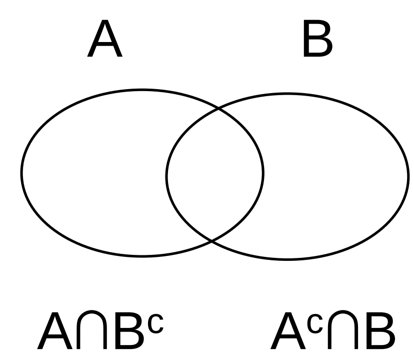
「 $Pr(A|H_0) = Pr(B|H_0)$ ならば $Pr(A|H_1) \geqq Pr(B|H_1)$ 」を示せばよい。

尤度比検定の定義から $S \subseteq A$ ならば $Pr(S|H_0) \geqq a\, Pr(S|H_1)$ であり、

$S \subseteq A^c$ ならば$Pr(S|H_0) \leqq a\, Pr(S|H_1)$ である(cは補集合)。従って

$$
\begin{aligned}
&Pr(A|H_1) - Pr(B|H_1) \\
&\quad = Pr(A \cap B^c|H_1) - Pr(A^c \cap B|H_1) \\
&\quad \geqq a\, Pr(A \cap B^c|H_0) - a\, Pr(A^c \cap B|H_0) \\
&\quad = a\, Pr(A|H_0) - a\, Pr(B|H_0) = 0
\end{aligned}
$$

（証明終）



尤度比検定を作ってみる

帰無仮説　$q(x) = 1/(2\pi)^{1/2} \exp(- x^2/2)$

対立仮説　$p(x) = 1/(2\pi)^{1/2} \exp(- (x-1)^2/2)$

尤度比は　$L(X^n) = \Pi_i \{ p(X_i) / q(X_i) \}$

$$= \Pi_i \exp(X_i - 1/2) = \exp(\Sigma_i X_i - n/2)$$

棄却域は、帰無仮説のもとで「$L(X^n) \geqq a$」の確率が有意水準になるように定める。「$L(X^n) \geqq a$」によって定まる集合は $(\Sigma_i X_i) / n^{1/2} \geqq a'$ によって定まる集合と等価なので有意水準から a' を決めればよい。

$(\Sigma_i X_i) / n^{1/2}$ は帰無仮説のもとで平均0 標準偏差1の正規分布に従うから有意水準 0.01 に対しては $(\Sigma_i X_i) / n^{1/2} \geqq 2.33$ が棄却域になる。

複合仮説では最強検定は存在しないことがある

定義. $X^n=\{X_1,X_2,\ldots,X_n\}$ を確率変数の集合とする。

帰無仮説 :「X^n は $p(x|w_0)$ から独立に得られた」

対立仮説:「X^n は $p(x|w_1)$ $(w_1\neq w_0)$ から独立に得られた」

このとき、対立仮説は複数の確率分布からできているので

最強力検定は存在するとは限らない。しかしながら、最強力とは

限らないが、最尤推定量 w^* を用いて

$$L(X^n) = \Pi_i\, p(X_i|w^*) / \Pi_i\, p(X_i|w_0)$$

を用いた検定がしばしば使われる。最強力でなくても検定を行うことは可能

である(最尤推定量は精度がよくないので検出力は小さいことが多いが)。

尤度比検定の例

これまでわが社の将棋ソフトは名人との対局で勝率が4割だった。このたび、わが精鋭チームが深層神経回路網を100億回自己対戦させ強化学習したところ、ついに名人との対戦で7勝3敗になった。わが社のソフトは本当に強くなったのだろうか。

帰無仮説： $q(x) = 0.4^{x} (1-0.4)^{1-x}$

対立仮説： $p(x|a) = a^{x} (1-a)^{1-x} \quad (a>0.4)$

2項分布の片側検定

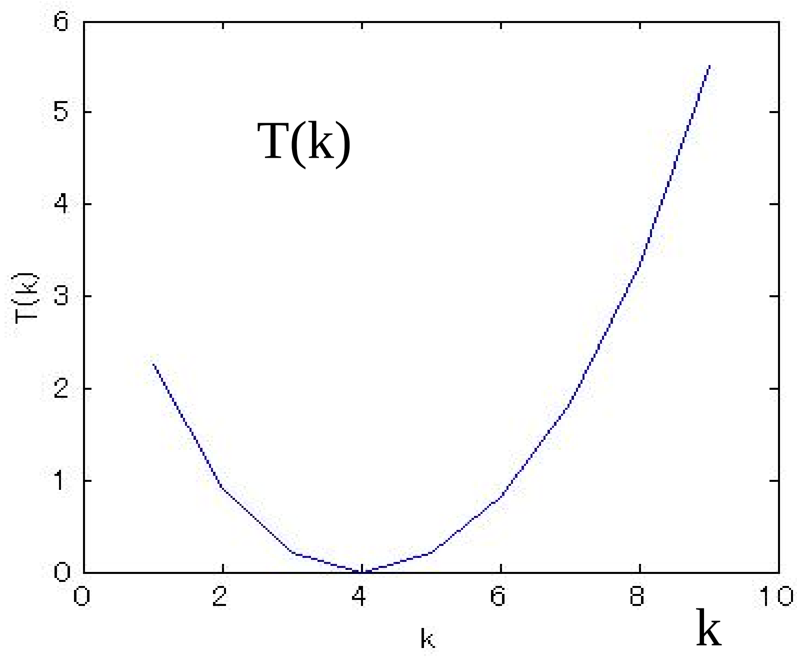
名人と n 回対戦して k 回勝ったとする。毎回独立だとする。 a^* を最尤推定量とすると

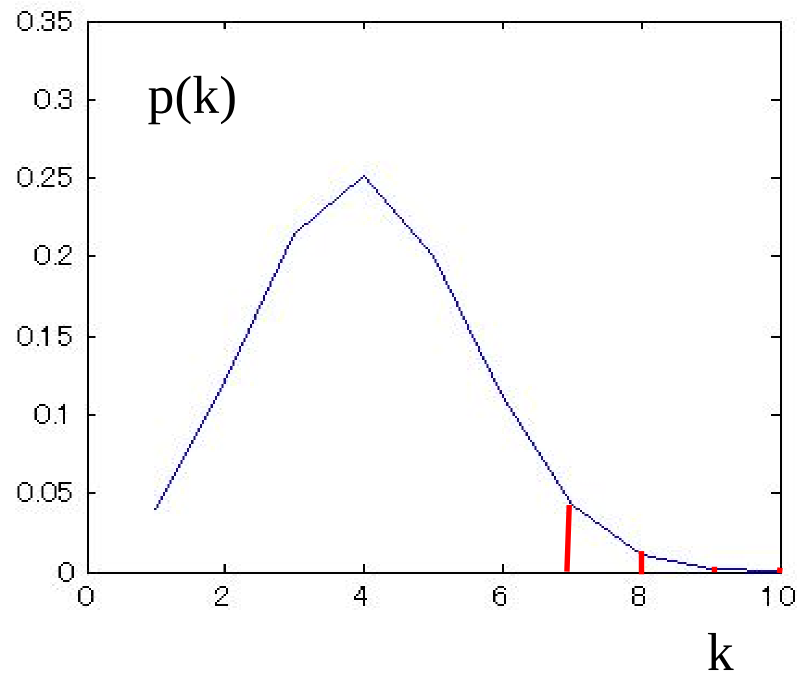
$$T = \log L(X^n) = \Sigma_i \{ \log p(X_i|a^*) - \log q(X_i) \}$$
$$= \Sigma_i \{ X_i \log (a^*/0.4) + (1-X_i) \log ((1-a^*)/0.6) \}$$
$$= k \log(a^*/0.4) +(n-k) \log ((1-a^*)/0.6)$$
$$= k \log((k/n)/0.4) + (n-k) \log ((1-k/n)/0.6)$$

最尤推定量は $a^*= k/n$ であることを用いた。T は k だけで定まる関数なので $T=T(k)$ と書く。これの$(1/n)$倍は KL 情報量なので、$T(k)$ は $k=0.4n$ で最小値 0 を取る。領域 $k \geqq 0.4n$ では $T(k)$ は単調非減少。

2項分布の片側検定

帰無仮説が正しいとすると k は 2項分布 $p(k) = {}_nC_k\ 0.4^k\ (1-0.4)^{n-k}$ に従う。棄却域は k の値を使って調べることができる。n=10 のとき、$Pr(k \geqq 7) = 0.054$ なので、有意水準 0.05 でも帰無仮説は棄却されない。すなわち、有意水準0.05 で「わが社のソフトは強くなったとはいえない」。





カイ二乗検定

データの数 n が十分に大きいときパラメータ w の次元を d とすると

$$Y = 2 \log \ L(X^{n}) = 2 \log \{ \Pi_i \, p(X_i | w^*) / \Pi_i \, p(X_i | w_0) \}$$

の確率分布は自由度 d のカイ二乗分布に法則収束するので、Y を用いた検定を行なうときの棄却域を $p(x) = \{ 1/(2^{d/2}\Gamma(d/2)) \} \ x^{d/2-1} \exp(-x/2)$ を用いて定めることができる。

（注）ただし Y がカイ二乗分布に法則収束することは機械学習のモデル例えば深層学習や混合正規分布やボルツマンマシンでは成り立たないのでこの方法は適用できない。そうしたモデルの検定に最尤推定量を適用すると検出力が非常に弱くなることが知られている。

ベイズ法の検定

定義. $X^n=\{X_1,X_2,\ldots,X_n\}$ を確率変数の集合とする。

帰無仮説 :「w は $\varphi(w)$ から得られ、X^n は $q(x|w)$ から独立に得られた」

対立仮説 :「w は $\psi(w)$ から得られ、X^n は $p(x|w)$ から独立に得られた」

このときは

$$L(X^{\mathbf{n}}) = \int \psi(w)\, \Pi_i\, p(X_i|w)\, dw \;/\; \int \varphi(w)\, \Pi_i\, q(X_i|w)\, dw$$

が最強力検定を与えることが知られている。ネイマン・ピアソンの補題はこの特殊な場合である。($\varphi(w)=\delta(w-w_0)$, $\psi(w)=\delta(w-w_1)$ に相当する)

（注意）ベイズ法では複合仮説を重み付けすることで最強力検定できる。

p値とは

p値の定義

帰無仮説　$q(x) = 1/(2\pi)^{1/2} \exp(-x^2/2)$

対立仮説　$p(x) = 1/(2\pi)^{1/2} \exp(-(x-m)^2/2)$ において $m>0$

帰無仮説のもとで$Y=(\Sigma_i X_i) / n^{1/2}$ は平均0標準偏差1の正規分布になる。実験で得られたデータが X^n に対して Y を計算したら y だった。このとき p値を次式で定義する。片側検定に対しては

$$p値 = \int_y^\infty p(x)\,dx$$

有意水準 0.05 のとき p値が 0.05 よりも小さいとき帰無仮説を棄却する。このとき「$p<0.05$」と書いたりする。

p値の問題は人間が原因

生物学や医学では、実験を行って p値を計算し、この値が小さくなることをもって「対立仮説を実験的に証明した」論文がたくさん掲載されていたが、その結果、再現性のない論文が異常に多くなったため、2016年、p値を用いて対立仮説を説明するという方法について否定的な意見が高まっている。

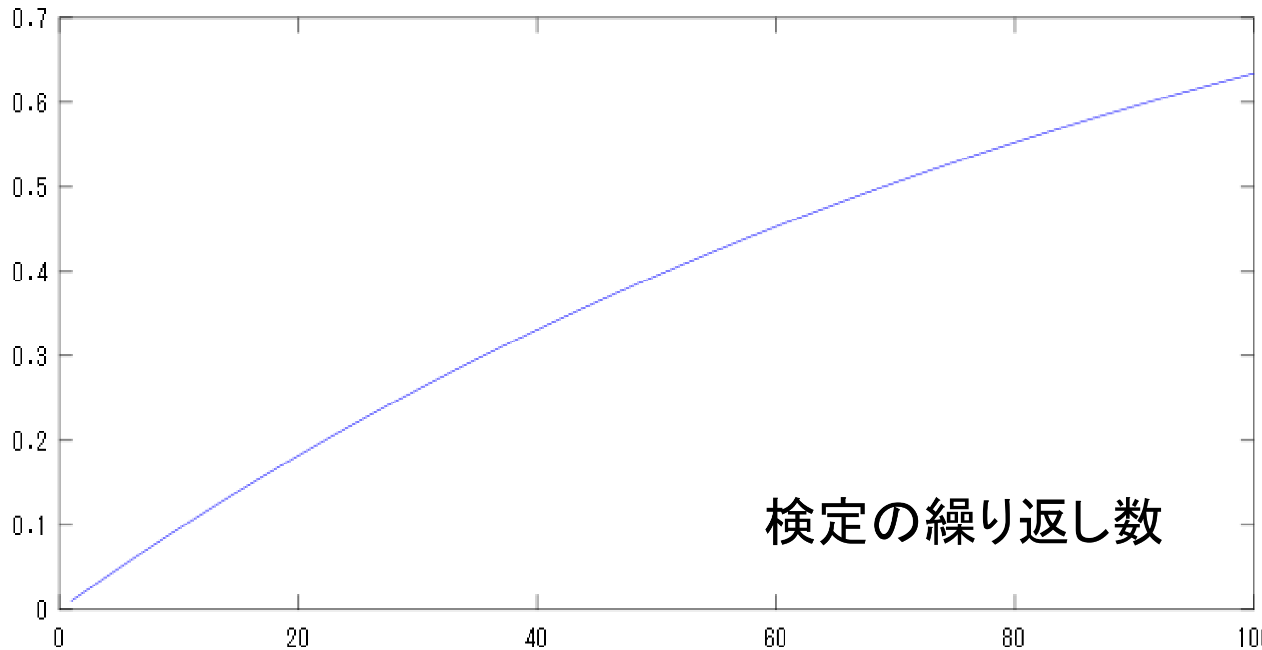
p値が 0.05 よりも小さくなった場合、帰無仮説のもとでその実験結果が得られる確率が 0.05 よりも小さいということ自体は正しい。p値が問題を生んでいる原因は統計学の誤りというよりも、実験上で正しい手続きが取られていない（たとえば、p値が0.05よりも小さくなるまで何度も実験を繰り返すなどの）誤りや、「基準さえみたせばOK」という人間の考え方の間違いである。

繰り返して検定すると

有意水準 ε の(独立なデータを取って)n 回検定を繰り返すと、その間に少なくても一回帰無仮説が棄却される確率 $P = 1-(1-\varepsilon)^n$

例 ε = 0.01 とする。

少なくても1回
帰無仮説が
棄却される
確率



検証をするには

１．統計的に誤った操作をしない（何度もデータを取って検定を通過するデータだけを採取するなど。

２．例えば有意水準 0.05 の検定は帰無仮説が正しくても20回に1回は間違うものであり、「統計的検定をすればOK」という考え方をしない。

３．現実の応用の場面では、検定だけに頼らず、様々な異なる観点から仮説の合理性を考察する。