『確率・統計』ガイダンス

田中　勝人

１．授業の目標

初級レベルの確率と推測統計について講義する．

統計学： データを分析するための方法論的科学．大きく，**記述統計**と**推測統計**に分かれる．
記述統計： データを整理して見やすくまとめるための統計的方法.（100 番台の『統計学入門』）
推測統計： データの背後にある母集団の特性値（パラメータ）を推論するための統計的方法. この授業が対象とする分野.
確率の概念： 推測統計においては不可欠の概念.（観測値間のばらつき，実験結果の不確定性, 標本抽出における変動などは確率の概念を使って説明する）

２．授業の内容

(1) 確率の概念	⋯　２回
(2) 確率変数と確率分布	⋯　２－３回
(3) さまざまな母集団分布	⋯　２－３回
(4) 標本分布と正規近似	⋯　２回
(5) 確率モデルと推定	⋯　２－３回
(6) 検定	⋯　２－３回

３．テキスト

主として，次の本を使う．

田中　勝人著『統計学』新世社（第５章以降）

４．参考書

演習用の参考書は，以下の通り．

初級篇：立花・田川・成田 共著『確率・統計』共立出版
中級篇：野田・宮岡 共著『入門・演習 数理統計』共立出版
　　　　国沢清典 編『確率統計演習１，２』培風館

５．成績評価

宿題（４回）と筆記試験（２回）による．前者：後者＝１：３．

６．レジュメ

次の場所から，ダウンロードできる．

http://wakame.econ.hit-u.ac.jp/~tanaka/org.html

確率の基礎

試行：同じ条件のもとで繰り返すことのできる実験，観測，調査などの総称

全事象： ある試行により起こりうるすべての結果の集まり（集合：有限とは限らない）で，以下では総称的に Ω で表す.

基本事象： 全事象のそれぞれの要素のことで，総称的に ω で表す.

事象： 全事象 Ω の部分集合であり，記号 A, B などで表す（注： この定義は，Ω が非可算集合ならば，若干の修正が必要になる場合もある）.

さまざまな事象：全事象 Ω，基本事象 ω，和事象 $A \cup B$，差事象 $A - B$，積事象 $A \cap B$，
空事象 ϕ，余事象 $\bar{A}$，互いに排反な事象 $A \cap B = \phi$

確率は，事象（必ずしも有限でない集合）に対して定義される.

確率の公理：次の３つの条件をみたす集合関数 P を**確率測度 （Probability Measure）**あるいは単に確率という.

(P1) 任意の事象 A に対して $0 \leq P(A) \leq 1$
(P2) $P(\Omega) = 1$
(P3) $A_i \cap A_j = \phi \ (i \neq j)$ ならば

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

確率の性質： （(4), (5), (6) の性質は確率の連続性と呼ばれる.）

(1) $P(\bar{A}) = 1 - P(A)$
(2) $A \subset B$ ならば $P(A) \leq P(B)$
(3) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
(4) $A_1 \supset A_2 \supset \cdots$ （単調減少の事象列）ならば

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

(5) $A_1 \subset A_2 \subset \cdots$ （単調増加の事象列） ならば

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n)$$

(6) 単調でなくても，$\lim A_n = A$ ならば,

$$\lim_{n\to\infty} P(A_n) = P(A)$$

事象の従属性と独立性

事象 A が起きたという条件のもとで，事象 B が起きる確率を，

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

で表し，A のもとでの B の**条件付き確率**という．

独立性：

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

が成り立つとき，A と B は**互いに独立**であるという．同値な条件としては，

$$P(A|\bar{B}) = P(A), \quad P(B|\bar{A}) = P(B), \quad P(A \cap B) = P(A) \times P(B), \quad \cdots$$

独立性の意味： A が起きたという情報が B の生起に何ら影響しない（情報としての価値がない）．

(独立性の例) コイン投げ，サイコロ投げ，⋯

（自明でない例） 52 枚のトランプのカードから 1 枚をランダムに抜くとき，スペードを抜くという事象と絵札を抜くという事象は互いに独立である．

独立性のメリット： 同時確率 $P(A \cap B)$ の計算を回避して，周辺確率 $P(A)$, $P(B)$ の計算に帰着させる（単純化）．ただし，現実に独立性が成り立つかどうかは，先験的なことではなく，調べる（検定する）必要がある．

一般に，事象の無限列 $A_1, \cdots, A_n, \cdots$ が互いに独立とは，この中の任意の有限個の事象 $A_{i_1}, \cdots, A_{i_k}$ に対して，

$$P\left(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}\right) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

（注）ペアごとに独立でも，互いに独立になるとは限らない．

ベイズの定理

原因の確率，事後確率の計算方式を与える定理である．全事象 Ω が n 個の互いに排反な事象により分割されているとき，すなわち，

$$\Omega = \bigcup_{i=1}^{n} A_i, \qquad A_i \cap A_j = \phi \quad (i \neq j)$$

の場合，ある事象 B が起きたというもとで，A_i が起きる**条件付き確率**は

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i)\, P(B|A_i)}{\sum_{i=1}^{n} P(A_i)\, P(B|A_i)}$$

で計算される．

(応用例 1) ある病気の検査法は完全でなく，陰性でも感染している可能性が若干あり，陽性でも感染していないことがあり得る．このとき，陽性の結果が出た被検者が本当に感染している可能性は？

A: 検査の結果が陽性である事象　　B: 病気である事象

$$P(A|B) = 0.98 \qquad P(A|\bar{B}) = 0.05 \qquad P(B) = x$$

とするとき，

$$\begin{aligned}
P(B|A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)\,P(B)}{P(A|B)\,P(B) + P(A|\bar{B})\,P(\bar{B})} \\
&= \frac{0.98 \times x}{0.98 \times x + 0.05 \times (1-x)} = \frac{98x}{93x+5} \\
&= 0.17\,(x = 0.01), \quad 0.69\,(x = 0.1), \quad 0.95\,(x = 0.5)
\end{aligned}$$

(応用例 2) 入試合格者の中には，実力通りに合格した者の他に，実力がなくても幸運にも合格する者もいると言われている．このとき，合格者の中で本当に実力があって合格した者は何割？

A: 合格するという事象　　B: 実力があるという事象

$$P(A|B) = 0.6 \qquad P(A|\bar{B}) = 0.2 \qquad P(B) = x$$

とするとき，

$$\begin{aligned}
P(B|A) &= \frac{P(A \cap B)}{P(A)} = \frac{P(A|B)\,P(B)}{P(A|B)\,P(B) + P(A|\bar{B})\,P(\bar{B})} \\
&= \frac{0.6 \times x}{0.6 \times x + 0.2 \times (1-x)} = \frac{3x}{2x+1} \\
&= 0.43\,(x = 0.2), \quad 0.67\,(x = 0.4), \quad 0.82\,(x = 0.6)
\end{aligned}$$

確率変数の導入

事象を数値に変換することにより，数学的操作可能性を高めることを考える．そのために，確率変数を導入する．

確率変数 $X = X(\omega)$ の定義：全事象 Ω で定義され，実数の集合 R に値をとる関数である．

確率変数の利点:事象から数値に変換された確率変数は，数学的な操作性に優れている．すなわち，確率分布，分布関数，密度関数などの概念や，それらに基づく期待値，分散などを定義することができ，確率的な現象を多面的に考察することが可能となる．

分布関数

確率変数 X に対して，X が x 以下の値をとる確率を x の関数とみなしたもの，すなわち，

$$F(x) = P(X \leq x)$$

を確率変数 X の分布関数という．分布関数は X の性質を特徴付けるものであり，分布関数を推測するのが統計学の主要な目的の１つである．

分布関数の性質：

(D1) 任意の x に対して $0 \leq F(x) \leq 1$ であり，かつ，

$$F(-\infty) = 0, \quad F(+\infty) = 1$$

(D2) 単調非減少，すなわち，$x < y$ ならば $F(x) \leq F(y)$
(D3) 右連続，すなわち，任意の x に対して $F(x) = F(x+)$ である．しかし，必ずしも左連続ではない．

(D3) の証明：

$$A_n = \left\{\omega : X(\omega) \leq x + \frac{1}{n}\right\}$$

とおくと，$\{A_n\}$ は単調減少な集合列となり，かつ，

$$P(A_n) = F\left(x + \frac{1}{n}\right), \quad \lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n = \{\omega : X(\omega) \leq x\}$$

となる．したがって，確率の連続性より，

$$\begin{aligned} F(x) &= P\left(\lim_{n\to\infty} A_n\right) = \lim_{n\to\infty} P(A_n) \\ &= \lim_{n\to\infty} F\left(x + \frac{1}{n}\right) = F(x+) \end{aligned}$$

（問） 分布関数の左連続性は必ずしも成り立たないことを示せ．