統計（医療統計）
前期・第1回　序説～確率

授業担当：徳永伸一

東京医科歯科大学教養部　数学講座

はじめに注意しておきたいこと

- 授業は前期６回＋後期６回
  - 前期末（または途中）に中間試験，後期末に本試験.
- 欠席した場合は次回までに要点の確認を.
  - 授業後３日以内（を目標）に授業スライドをｐｄｆファイルに変換してアップロードしておきます.
  - WebCTを利用した復習テストも実施するかもしれません。掲示や授業でのアナウンスに注意してください.
- 要点を効率よく押さえましょう.
  - 計算より概念・原理の理解を重視
  - 試験で出題のポイントとなる（＝合否の分かれ目となる）事項は授業でも強調しているつもり

（なのですが・・・）.

今日（第1回）の授業の概要：

- 統計学とは何か ～ なぜ統計学が必要か
- 授業全体のOverview
- 教科書第９章「順列・組合せと確率」ほぼ全部
  - 確率の基礎概念 ～ ベイズの定理

統計学(statistics)とは何か

- 「集団現象を観察し分析する方法を研究する学問」
（大辞林）
- *Statistics*： the science of using information discovered from studying numbers
（Cambridge Advanced Learner's Dictionary）
- 「数字データ（客観的な対象の性質・現象を数字で表したもの）をどのように分析し，どのような判断をくだしたらよいかを論ずる学問」
（どの本だか忘れました）
- 「検討したい事柄についてのデータを収集し、それが持っている有意味な情報を客観的・効率的に記述し、読み取るための道具」
（「本当にわかりやすいすごく大切なことが書いてあるごく初歩の統計の本」）

なぜ統計学が必要か

- 「個体差のある生物集団を対象とし、調査条件・実験条件を十分には管理することが出来ないような研究のための方法論」として必須（「メディカル・コメディカルの統計学」）
- 特定の薬や治療法が有効かどうかの合理的・客観的な判断を下すため

  （カンに頼られたんじゃ患者はたまらん！）
- そして・・・

なぜ統計学が必要か（続き）

［逆説］統計学は絶対的なものではない！

100%の客観性はあり得ない

恣意性の入る余地すらある

（いくら論理が正しくても、仮定・前提に誤りがあれば・・・）

↓

統計のウソ・誤用・悪用を見破るためにも

統計学の知識が必要

医歯学教育における「統計学」の位置づけ

担当者（徳永）の考え：

- 客観的・合理的判断を下すための基本的な技術であり、特に医療従事者には必須の学問。教養以前。
- 「統計学のわかってない医者・歯医者には絶対にかかりたくない」（社会的要請）
- よって情け容赦は無用。

学部教員の声

某会議にて某教授曰く

「統計できない人は進級させなくていいですから」

別の某教授からの怒りの電話：

「基本的なことがわかっていない学生・大学院生が多い．教養部のうちにちゃんと教育してもらわないと！」

重要性の認識は共通しており、平成15年度以降の新カリキュラムでも総コマ数は減少していない（学習内容は少し増えた）。

１７年度からは学士編入学生も必修に。

統計学は難しいか？

- 数学的に高級な定理（中心極限定理など）も用いるが，”現象”として理解すればよい．
- 数学的な抽象概念（確率変数など）の理解はある程度必要だが，高度な”数学的思考”を駆使するわけではない．
- 使いこなすべき公式はそれほど多くない．計算は（電卓を使えば）中学生でもできる．
- 最終的には多くの学生が「わかってしまえば簡単だった」という感想を漏らします．
  - 皆さんの先輩M君（現M４）の名言：

「自転車と同じで，いったん乗れるようになればあとは楽々ですよ」

教科書について

- 平成１７年度から新しい教科書に
  - （旧）「メディカル・コメディカルの統計学」（共立）
  - （新）「臨床検査学講座 数学／統計学」（医歯薬出版）

    【注意】最新版は「第４刷」
- 徳永は授業範囲のかなりの部分（主に前期の範囲）を執筆しています。
  - 第9章 順列・組合せと確率（6ページ）
  - 第10章 記述統計（4ページ）
  - 第11章 確率変数と確率分布（18ページ）
- でも教科書は絶対的なものではありません。
  - （最新版ではほぼ修正できたと思いますが）ミスプリもあるかも.

Overview

- 確率（9章）
- 記述統計（10章）・・・・情報の要約
  - 表やグラフで表す
  - 代表値（平均など）や散布度（分散など）を求める

↓ 確率モデル（11章）

- 推測統計（13章～）
  - 推定（点推定、区間推定）
  - 仮説検定

第9章「順列・組合せと確率」の概要

Ⅰ．順列と組合せ
Ⅱ．確率の基礎概念
Ⅲ．確率の定義と性質
Ⅳ．条件付き確率と事象の独立性
Ⅴ．ベイズの定理

- 大部分は高校数学（受験数学）の範囲です.
- とはいえ、高校数学であまり取り上げられない抽象的な概念をきちんと理解することはとても重要.
- 曖昧な理解のまま放っておくと後になって命取りとなります.

Ⅰ．順列と組合せ

ざっと確認のみ

- 順列 $_nP_r$ ＝ n(n-1)(n-2)・・・(n-r+1)
- 重複順列 $_n\Pi_r$ ＝ n^r
- 組合せ $\binom{n}{r}$ ＝ $_nC_r$ ＝ $_nP_r$ / r! ＝ n! / (r!(n-r)!)

  【注意】ほんとは$\binom{n}{r}$の形で統一したいが、パワーポイント上できれいに表示させるのはかなりやっかいなのでnCr型も併用します。

  - 11章Ⅳの「2項分布」で用いる.
  - 「2項定理」「パスカルの三角形」は知っておこう。
- 重複組合せ$_nH_r$ ＝ $_{n+r-1}C_r$

Ⅱ. 確率の基礎概念（1）

- 標本空間sample space Ω
  ・・・確率の対象となる「結果」の全体

- 事象event・・・「結果」の一部．　Ωの部分集合.
  - 根元事象elementary event
    ・・・「結果」の最小単位．それ以上分割できない事象
  - 全事象
    ・・・すべての事象を含む事象．すなわちΩと一致.
    （100％の確率で起こる）
  - 空事象φ
    ・・・空集合に対応する事象（起こる確率ゼロ）

Ⅱ. 確率の基礎概念（2）

- AとBの和事象 Union : $A \cup B$
- AとBの積事象 intersection : $A \cap B$
- A の余事象 complement : A^c または $\bar{A}$
- 事象Aの確率probabilityを　$P(A)$　で表す.
- AとBが排反・・・［定義］ $A \cap B = \phi$

Ⅲ. 確率の定義と性質

- 経験的確率

  実験や観測の繰り返しによって経験的に知る確率

- 数学的確率

  「根元事象等確率」などの仮定をおいて計算されるもの

数学的に扱いやすいものにするには定式化が必要
→最低限の約束ごと（公理）を定める

公理的確率（数学的に厳密な定式化）

Ωの事象Aに実数$P(A)$が対応し，以下の3条件（＝確率の公理）を満たすとき，PをΩ上の確率という．

（1）$0 \leqq P(A) \leqq 1$

（2）$P(\Omega)=1$，$P(\phi)=0$

（3）A，Bが互いに排反事象であるとき

$$P(A \cup B) = P(A) + P(B)$$

公理からただちに導けること:

（1） $P(A^c) = 1 - P(A)$

（2） $A \subset B \Rightarrow P(A) \leqq P(B)$

（3） $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

（確率の加法定理）

Ⅳ.条件付き確率と事象の独立性（1）

まず，
「AのもとでのBの条件付き確率$P(B \mid A)$」
を以下の式で定義：

$$P(B \mid A) := P(A \cap B) \,/\, P(A)$$

注：$P(B \mid A)$は結果的に，「Aが起こったという条件のもとでのBの起こる確率」と解釈できる．

（公理的確率の立場では，数式で明確に定義すべき）

Ⅳ.条件付き確率と事象の独立性（2）

$P(A)>0$のとき,

$$P(B \mid A)=P(A\cap B)/P(A)$$

の両辺に$P(A)$をかけることにより

$$P(A\cap B) = P(A)\cdot P(B \mid A)$$

が成立．これを確率の乗法定理という.

Ⅳ.条件付き確率と事象の独立性（3）

★事象の独立性

［定義］

$$P(A \cap B) = P(A) \cdot P(B) \cdots (\ast)$$

のとき，「AとBは（互いに）独立」であるという.

（＊）は独立事象の乗法定理とも呼ばれる.

注：（＊）を「定義」とする立場ではこれを「定理」と呼ぶのは不適切だが，習慣上の問題.

Ⅳ.条件付き確率と事象の独立性（4）

あらためて独立事象の乗法定理（本当は「定義」）

$$P(A\cap B) = P(A)\cdot P(B) \cdots (*)$$

と（一般の）確率の乗法定理

$$P(A\cap B) = P(A)\cdot P(B \mid A)$$

を見比べると，AとBが独立のときは

$$P(B \mid A) = P(B)$$

となっていることがわかる.

つまり，AとBが独立のときは，Aが起こっても起こらなくてもBが起こる確率は同じ.

すなわちAの成否がBに影響を与えていない

（同様にBもAに対して影響無し）.

これが‘独立’の意味するところ.

あと2つ，とても大事な注意

★試験のとき「独立」と「排反」を混同する人が毎回びっくりするほど多い．再確認しておこう．

「AとBが独立」⇔ $P(A \cap B) = P(A) \cdot P(B)$

「AとBが排反」⇔ $A \cap B = \phi$

$\Rightarrow P(A \cap B) = 0$

★後で出てくる「確率変数の独立性」を「事象の独立性」と混同する人も多い．関連はあるが，次元の異なる概念です．

Ⅴ. ベイズの定理（1）

どんな定理？

- 「原因」（病気など）と「結果」（症状など）の関係がある程度わかっているとき，「結果」からそれがある特定の「原因」によるものである確率を求める定理.
- 確率の乗法定理の簡単な応用.
- 試験によく出る定理
  （経験的確率は90％以上）

Ⅴ. ベイズの定理（2）

以下

Ω：標本空間，　A_1，　A_2，　・・・ A_r，B：Ωにおける事象とする。

定理の仮定（前提条件）

Ωが A_1，A_2，･･･ A_r によって分割されている、すなわち，

① $\bigcup_{1 \leqq k \leqq r} A_k = A_1 \cup A_2 \cup \cdots \cup A_r = \Omega$

かつ

② 各A_k は互いに排反

（注： A_1，　A_2，　・・・ A_r は「原因」に相当する事象）

Ⅴ. ベイズの定理（3）

定理の結論

このとき次式が成立する：

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{\sum_{k=1}^{r} P(A_k)P(B \mid A_k)}$$

Ⅴ. ベイズの定理（4）

注意

BがA_k の「結果」に相当する事象だとすると・・・

- $P(B \mid A_k)$ は「原因→結果」だから調べやすい
- 逆に $P(A_1 \mid B)$ は「結果→原因」なので直接調べることは難しいかもしれない.

ベイズの定理 Bayes' Theorem（証明の前に再確認）

事象A_1，A_2，・・・A_r，$B \in \Omega$ について

［仮定］① $\bigcup_{1 \leqq k \leqq r} A_k = \Omega$ かつ

② 各A_k は互いに排反

であるとき,

［結論］条件付確率$P(A_1|B)$ に関して，以下の公式が成立つ.

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{\sum_{k=1}^{r} P(A_k)P(B \mid A_k)}$$

ベイズの定理の証明

簡単のため，$r = 3$ とする.

仮定より， A_1，A_2，A_3 はΩの分割になっているとする.

すなわち，

① $A_1 \cup A_2 \cup A_3 = \Omega$ かつ

② A_1，A_2，A_3 は互いに排反事象

が成り立っている.

ベイズの定理の証明（続き）

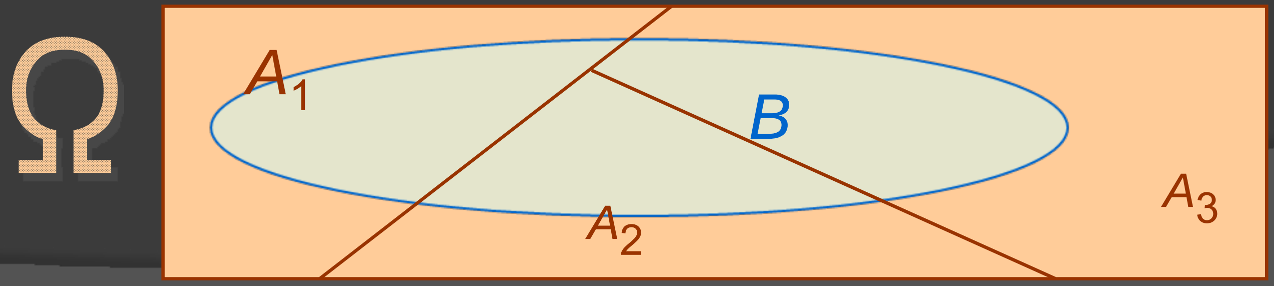
（ ①: $A_1 \cup A_2 \cup A_3 = \Omega$　②:A_1，A_2，A_3 は互いに排反 ）

$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)}$$

$$= \frac{P(A_1 \cap B)}{P(A_1 \cap B) + P(A_2 \cap B) + P(A_3 \cap B)}$$ ← ①②より

$$= \frac{P(A_1)\,P(B \mid A_1)}{P(A_1)\,P(B \mid A_1) + P(A_2)\,P(B \mid A_2) + P(A_3)\,P(B \mid A_3)}$$

（証明終わり）



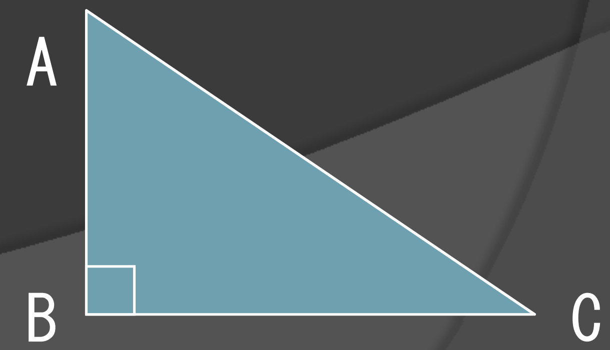
ちょっと横道・・・

「定理」というものは，必ず仮定と結論から成り立っています．

例：ピタゴラスの定理

［仮定］三角形ABCにおいて、$\angle B = 90^\circ$

［結論］$AB^2 + BC^2 = CA^2$



横道続き

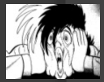
「ピタゴラスの定理ってなんですか」と尋ねられて，

「 **AB²＋BC²＝CA²** です」と答える人は，まずいない．

しかし！

「『ベイズの定理』の内容を書きなさい」

と試験で出題すると，



$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{\sum_{k=1}^{r} P(A_k)P(B \mid A_k)}$$

だけ書く人は

少なくない．

なのでしつこいけどもう一度 RECALL

ベイズの定理 Bayes' Theorem

（注：以下は記号の説明は省略してあります）

［仮定］ ① $\bigcup_{1\leqq k\leqq r} A_k = \Omega$ かつ

② 各 A_k は互いに排反

［結論］

$$P(A_1 \mid B) = \frac{P(A_1)P(B \mid A_1)}{\sum_{k=1}^{r} P(A_k)P(B \mid A_k)}$$

$r = 2$ の場合に関する補足

$r = 2$ のとき，仮定の条件は

「A_2はA$_1$の余事象」

と言っているのと同じ。よって

$A_1 = A, A_2 = \bar{A}$ として

$$P(A \mid B) = \frac{P(A)P(B \mid A)}{P(A)P(B \mid A) + P(\bar{A})P(B \mid \bar{A})}$$

と書ける（仮定は自動的に満たされるので一般に成り立つ式となる）.