

確率と統計 (0)

「確率と統計の基礎 (第4章)」

■担当教員: 杉山 将 (計算工学専攻)

■居室: W8E-505

■電子メール: sugi@cs.titech.ac.jp

■授業のウェブサイト:

<http://sugiyama-www.cs.titech.ac.jp/~sugi/>

連絡事項

■ テキスト:

- 統計学入門、東京大学出版会

■ その他参考書

- 渡辺澄夫・村田昇、確率と統計—情報学への架橋—、コロナ社,2005年
- 学習システムの理論と実現、渡辺澄夫編、森北出版、2004年
- 自然科学の統計学、東京大学出版会

■ 成績評価:宿題, 中間試験, 期末試験

統計学とは

■ 統計学(statistics)

データから現象の法則などの有用な情報を取り出すための方法論を議論する学問。

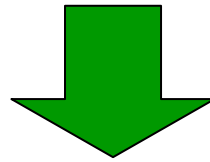
- 記述統計学(descriptive statistics)
- 数理統計学(mathematical statistics)

記述統計学

- 全てのデータを丹念に調べ, 規則性などを見つけ出す.
 - **国勢調査**
5年に一度, 日本に居住するすべての人の性別, 年齢, 国籍, 就業状態, 仕事の種類などを調べる.
 - **入試の成績**
毎年, 全員の試験の点数を調べる.

数理統計学

- 記述統計学はデータを整理・要約する方法論を議論する学問である。
- しかし、現実にはすべてのデータを採集するのが大変な場合や、原理的にデータを採取できない場合もある。



- **数理統計学**(mathematical statistics)
確率論に基づいて、一部のデータから論理性のある推測で全体の法則を見出すための方法論を議論する。

数理統計学が役立つ例

- **全数調査**: 国勢調査のように全員の情報を調べるのは時間もお金もかかる.
- **破壊検査**: 携帯電話を1メートルの高さから落としたときに壊れる確率がどの程度か知りたいとき, 実際に全ての携帯電話を落として壊れる確率を調べることはできない.
- **予測**: 明日の降水量を知りたいとき, 明日にならないと明日の降水量のデータは得られない.
- **学習するコンピュータ**: 人間が幾つかの例を教えれば, あとはコンピュータが自分で考えて自動的に処理してくれる.

基礎用語 (1)

- **標本点**(sample point) ω

起こりえる可能な結果.

- **標本空間**(sample space) Ω

標本点の全体の集合.

- **例) さいころ**

$$\omega = 1, 2, 3, 4, 5, 6$$

$$\Omega = \{1, 2, 3, 4, 5, 6\}$$

基礎用語 (2)

- **事象(event)**

起こりえる事柄. 標本空間の部分集合で定義される

- **空事象(empty event)** ϕ

標本点を一つも含まない起こりえない事柄

- **根元事象(elementary event)**

ただ一つの標本点からなる事象

- **複合事象(composite event)**

複数の標本点を含む事象

- **例)** さいころを振って奇数が出るという事象 A

$$A = \{1, 3, 5\}$$

基礎用語 (3)

■ **和事象**(union of events) $A \cup B$

AとBの少なくともどちらか一つが起こるという事象

■ **積事象**(intersection of events) $A \cap B$

AとB両方が同時に起こるという事象

■ **例)** さいころを振って奇数が出るという事象 A

$$A = \{1, 3, 5\}$$

さいころを振って3以下の目が出るという事象 B

$$B = \{1, 2, 3\}$$

- $A \cup B = \{1, 3, 5\} \cup \{1, 2, 3\} = \{1, 2, 3, 5\}$
- $A \cap B = \{1, 3, 5\} \cap \{1, 2, 3\} = \{1, 3\}$

基礎用語 (4)

- 事象AとBが同時に起こりえないとき, AとBは**排反事象**(disjoint events)であるといい, 次式で表す

$$A \cap B = \phi$$

例) A :さいころを振って奇数が出る

B :さいころを振って偶数が出る

- 事象Aが起こらないという事象をAの**補事象**(complementary event)と呼び, 次式で表す

$$A^c$$

例) A :さいころを振って奇数が出る

A^c :さいころを振って偶数が出る

基礎用語 (5)

■ 分配法則(distributive law)

$$(A \cup B) \cap C = (A \cap C) \cup (B \cap C)$$

$$(A \cap B) \cup C = (A \cup C) \cap (B \cup C)$$

■ ド・モルガンの法則(de Morgan's law)

$$(A \cup B)^c = A^c \cap B^c$$

$$(A \cap B)^c = A^c \cup B^c$$

確率

- **確率**: 事象の起こりやすさを定量的に示すもの
- 事象Aの起こる確率を次式で表す

$$P(A)$$

- 高校で習う確率は、例えばさいころの出た目を数え上げるなど、**計算**するものであった。
- 大学以降で習う確率は、数学的に**定義**するもの。

確率の公理

■ コルモゴロフ(Kolmogorov) の公理

1. 任意の事象 A_i に対して

$$0 \leq P(A_i) \leq 1$$

2. 全確率は1:

$$P(\Omega) = 1$$

3. 互いに排反な事象 A_i に対して

$$P\left(\bigcup_i A_i\right) = \sum_i P(A_i)$$

■ 以後, 全ての確率計算は上記の3つの公理のみに基づいて行なわれる.

加法定理

■ 事象AとBに対して

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

■ 証明: 図より, $A \cap B^c$, $A^c \cap B$, $A \cap B$ は互いに排反なので, 公理3より.

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

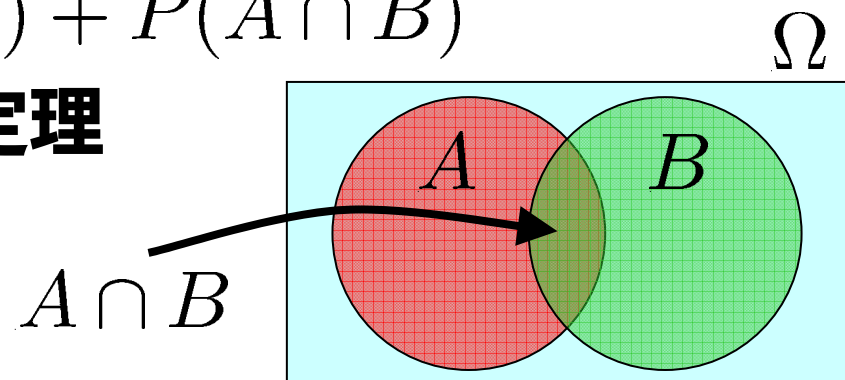
また, $A \cap B^c$, $A \cap B$ も互いに排反なので

$$P(A) = P(A \cap B^c) + P(A \cap B)$$

同様に, $A^c \cap B$, $A \cap B$ も互いに排反なので

$$P(B) = P(A^c \cap B) + P(A \cap B)$$

これらを合わせれば, 加法定理
が得られる



加法定理 (つづき)

- 同様に, 3つの事象A, B, Cに対して, 次式が成り立つ

$$\begin{aligned} P(A \cup B \cup C) \\ &= P(A) + P(B) + P(C) \\ &\quad - P(A \cap B) - P(A \cap C) - P(B \cap C) \\ &\quad + P(A \cap B \cap C) \end{aligned}$$

条件付き確率

- 事象Bが起こったとわかっている場合に事象Aが起こる確率を、Bを条件とするAの**条件付き確率** (conditional probability) とよぶ

- Aの条件付き確率を $P(A|B)$ で表し、次式で定義

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

例) A : さいころの目が奇数

B : さいころの目が3以下

$P(A|B)$: 3以下のものの中で奇数の確率

- **乗法定理:** $P(A \cap B) = P(A|B)P(B)$
 $= P(B|A)P(A)$

独立性

■ 事象AとBが

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

を満たすとき、AとBは**独立**(independent)であるという

例) A :さいころの目が奇数

B :さいころの目が4以下

■ 3つの事象A, B, Cの独立性は次のように定義

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

$$P(A \cap C) = P(A)P(C)$$

$$P(B \cap C) = P(B)P(C)$$

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

排反性と独立性

- A と B が排反ならば, コルモゴロフの公理3より

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

- A と B が独立ならば, 定義より

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

- 加法定理より

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

なので, A と B が独立でも, 排反とは限らない:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A)P(B) \\ &\neq P(A) + P(B) \end{aligned}$$

- 同様に, A と B が排反でも, 独立とは限らない:

$$\begin{aligned} P(A \cap B) &= P(A) + P(B) - P(A \cup B) = 0 \\ &\neq P(A)P(B) \end{aligned}$$

因果関係

- **結果が得られたとき, その原因を知りたい**
- **例) 結果 A : お腹が痛い**
原因 $\{H_i\}_{i=1}^k$: 風邪, 食中毒, 仮病?
- **原因を知るためには, $\{P(H_i|A)\}_{i=1}^k$ を調べればよい**
- **しかし, $\{P(H_i|A)\}_{i=1}^k$ は未知である事が多い**

ベイズの定理

- $\{H_i\}_{i=1}^k$ は互いに排反で $\bigcup_{i=1}^k H_i = \Omega$ と仮定

- **ベイズの定理**(Bayes' theorem):

$$P(H_j|A) = \frac{P(H_j)P(A|H_j)}{\sum_{i=1}^k P(H_i)P(A|H_i)}$$

- $P(H_j)$: H_j の**事前確率**(prior probability)
- $P(H_j|A)$: H_j の**事後確率**(posterior probability)
- **事前確率と条件付き確率** $\{P(A|H_i)\}_{i=1}^k$ が分かれば, 事後確率が計算できる

宿題：×切は次回の授業開始時²³

- 確率と統計の授業が好きな人, 嫌いな人の確率が

$$P(\text{好})=0.9, \quad P(\text{嫌})=0.1$$

であるとする. また, 確率と統計の授業が好きな人の中で授業中眠たい人, および, 確率と統計の授業が嫌いな人の中で授業中眠たい人の確率が, それぞれ

$$P(\text{眠} \mid \text{好})=0.2, \quad P(\text{眠} \mid \text{嫌})=0.8$$

であるとする.

- 次の確率を計算せよ. (1) $P(\text{好} \cap \text{眠})$, (2) $P(\text{眠})$,
(3) $P(\text{好} \cup \text{眠})$, (4) $P(\text{好} \mid \text{眠})$
- (5) 「確率と統計の好き嫌い」と「授業中眠たいこと」は独立か?
- (6) ベイズの定理を証明せよ.