統計（医療統計）

前期・第３回　確率変数と確率分布（１）

授業担当:徳永伸一

東京医科歯科大学教養部　数学講座

前回（第2回）の授業の概要：

- 第１回（教科書第９章「順列・組合せと確率」ほぼ全部）の復習.
- ベイズの定理に関する補足.
- 教科書第１０章「記述統計」最後まで.

Overview

- 確率(9章)　　　・・・第1回授業
- 記述統計(10章)　・・・第2回授業
  - 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 確率変数と確率分布)

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

S. TOKUNAGA 3

[復習] ベイズの定理 Bayes' Theorem

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.

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

S. TOKUNAGA 4

［復習］ 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

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

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）

［復習］例題（p.75）の解答と考察

P（A｜B）

=（P（A）P（B｜A））／（P（A）P（B｜A）＋P（A^C）P（B｜A^C））

= （0.01×0.99）／（0.01×0.99 + 0.99×0.07）

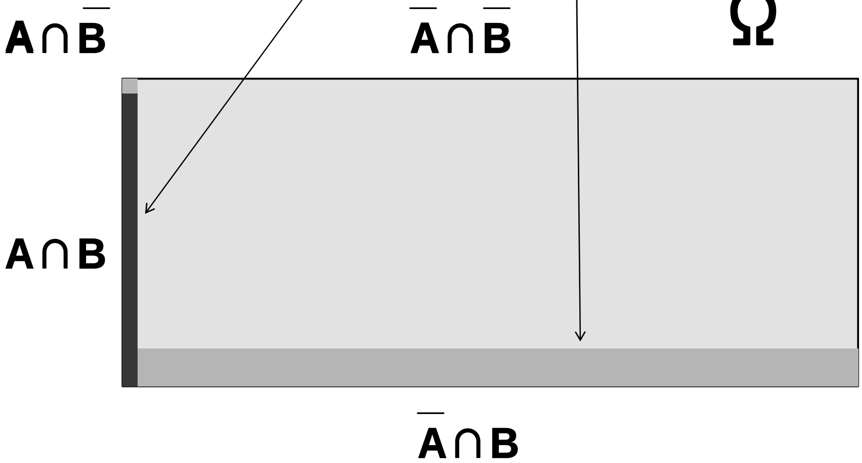
= <u>0.125</u> ・・・（答）

→意外と小さい？

<u>考察のポイント</u>

- 検診結果が陽性でも，実際には病気Xでない確率の方がずっと高い．
- しかし1%→12.5%だから確率は10倍以上．
- 使い方、結果の理解の仕方（患者への伝え方）が重要．

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



例題（p.75）の結果についての考察

<u>考察のポイント</u>

- 検診結果が陽性でも，実際には病気Xでない確率の方がずっと高い.
- しかし1%→12.5%だから確率は10倍以上.
- 使い方、結果の理解の仕方（患者への伝え方）が重要。

第9章の概要REPRISE

- Ⅰ．順列と組合せ
- Ⅱ．確率の基礎概念
  - 標本空間、事象
- Ⅲ．確率の定義と性質
  - 確率の公理
- Ⅳ．条件付き確率と事象の独立性
  - 「事象の独立性」の定義
- Ⅴ．ベイズの定理
  - 仮定と結論

では10章へ.

第10章 記述統計

Ⅰ．統計データの種類

Ⅱ．度数分布

1. 階級と度数，度数分布表
2. 度数分布表の視覚化（ヒストグラム）

Ⅲ．データの特性値

1. 代表値（平均・メディアン・モード）
2. 散布度（分散と標準偏差、**不偏分散**）

[復習] Ⅰ. 統計データの種類 & Ⅱ. 度数分布

Ⅰ. 統計データの種類

- 定性的データ（分類尺度）
- 定量的データ（順序尺度・間隔尺度）
  - 離散的discreteデータ
  - 連続的continuousデータ

  ★「離散的」か「連続的」かで数学的な扱い方が異なる

Ⅱ. 度数分布

KEYWORDS

- 度数frequence, 度数分布表, 階級class 、階級値
- スタージェスの公式
- 相対度数、累積度数、累積相対度数
- ヒストグラム

[復習] Ⅲ. データの特性値（2-3）

1-代表値 ・・・分布の中心的な位置を示す.

[1]平均mean

データ $x_1, x_2, \cdots, x_n$ に対し,

平均 $\bar{x} := (x_1 + x_2 + \cdots + x_n) / n = (1/n) \Sigma x_k$

と定義される。

度数分布表（階級数：m）が与えられているときは

階級値 $x'_1, x'_2, \cdots, x'_m$ と度数 $f_1, f_2, \cdots, f_m$ を用いて

$$\bar{x} := (1/n) \Sigma x'_k f_k$$

と計算（一種の近似計算）.

[2]メディアンmedian＝中央値（順位的に真ん中の値）

＊データが偶数個の場合は「真ん中の2つ」の平均.

[3]モードmode＝最頻値（度数が最大となる値or階級値）

［復習］ Ⅲ．データの特性値（4-5）

<u>2-散布度</u>･･･分布の広がり・ばらつきの度合いを示す.

［1］分散variance と 標準偏差standard deviation

データ $x_1,x_2,\cdots,x_n$ の平均 $\bar{x}$ に対し,

分散　$\sigma^2 := \{\Sigma(x_k - \bar{x})^2\}/n$

階級値 $x'_1, x'_2, \cdots, x'_m$ と度数 $f_1, f_2, \cdots, f_m$ を用いると

$$\sigma^2 := (1/n)\Sigma(x'_k - \bar{x})^2 f_k$$

標準偏差＝「σ^2の正の平方根」、すなわち

$$\sigma := \sqrt{(\sigma^2)}$$

S. TOKUNAGA　13

［復習］ Ⅲ．データの特性値（6）

［2］不偏分散unbiased variance

データ $x_1, x_2, \cdots, x_n$の平均 $\bar{x}$ に対し,

不偏分散 $U^2 := \{\Sigma(x_k - \bar{x})^2\}/(n-1)$

★nではなく(n－1)で割る理由：不偏性(→第13章Ⅱ)

★バラツキの度合いを表す指標としては同等.

★nが十分大きいときにはnで割っても(n−1)で割っても大差ない.

（たとえばn＝10000で有効数字3桁なら無視できる）

S. TOKUNAGA　14

［復習］ Ⅲ．データの特性値（7）

<u>【重要】不偏分散についての補足（定義の不統一）</u>

★文献によっては

①「分散」を不偏分散の形で定義

②「分散」は同じだが「**標本分散**」を不偏分散の形で定義

しているケースもあり、用語の使い方が統一されていない（以前使用していた教科書も②だった）.

★上記①②のケースでは、標準偏差ないし「標本標準偏差」を不偏分散の正の平方根$U=\sqrt{U^2}$で定義.

第10章 記述統計の概要

Ⅰ．統計データの種類

Ⅱ．度数分布

1. 階級と度数，度数分布表
2. 度数分布表の視覚化（ヒストグラム）

Ⅲ．データの特性値

1. 代表値（平均・メディアン・モード）
2. 散布度（分散と標準偏差，**不偏分散**）

次はいよいよ「１１章　確率変数と確率分布」

第11章 確率変数と確率分布

はじめに

確率変数は，確率・統計の学習において
もっとも基本的かつ重要な概念
であるが、きちんと理解するのは意外と難しい.
（一度わかってしまえば簡単だが）

ということを頭に留めておきましょう.

第11章 確率変数と確率分布

Ⅰ. 確率変数と確率分布の定義

Ⅱ. 確率変数の特性値
- 期待値（平均），分散など

Ⅲ. 確率変数の独立性

Ⅳ. 代表的な確率分布
- 2項分布，正規分布など

Ⅴ. 中心極限定理と正規近似

Ⅵ. 標本分布

Ⅰ．確率変数と確率分布の定義（1）

1−確率変数の定義

［定義］標本空間Ω上の実数値関数

（各根元事象に実数を対応させたもの）を

確率変数 random variable という.

- とり得る値が離散的→離散型確率変数
- とり得る値が連続的→連続型確率変数

S. TOKUNAGA 19

Ⅰ．確率変数と確率分布の定義（2）

教科書p.83例1

Ω：サイコロを振ったときの，目の出方で定まる 事象全体の集合.

- 「サイコロを振って1の目が出る」は 事象.
- 「サイコロを振って i の目が出る」という 事象ω_i に整数 i を対応させる 関数 を X（＝X(ω_i)） とおくと，Xは（離散型）確率変数 となる.
- 確率変数Xに対し，
  - 「X＝1」「X≦4」
  - 「Xは偶数」

  などは事象.

S. TOKUNAGA 20

Ⅰ．確率変数と確率分布の定義（3）

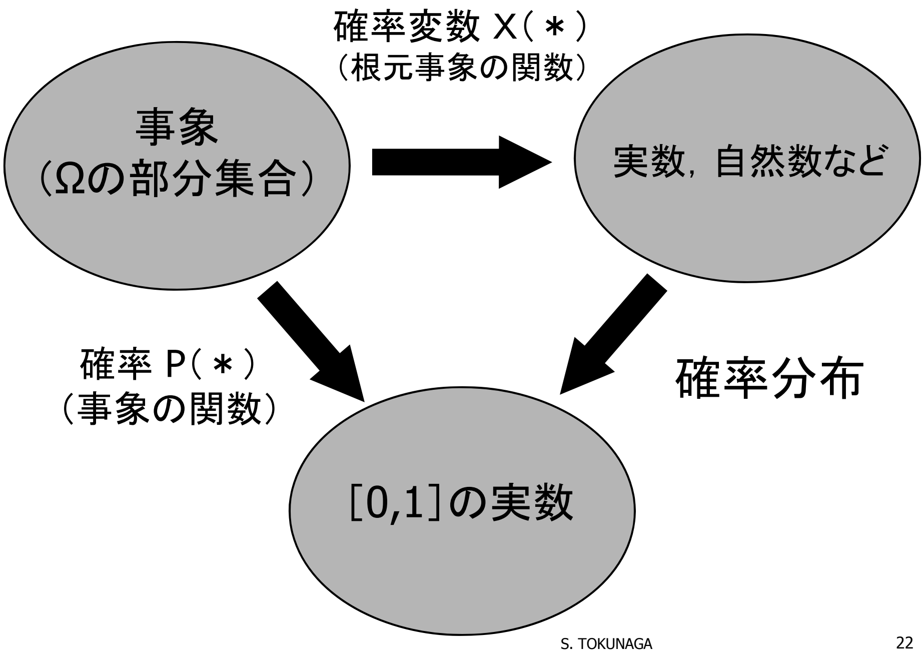
<u>2−離散型確率変数の確率分布</u>

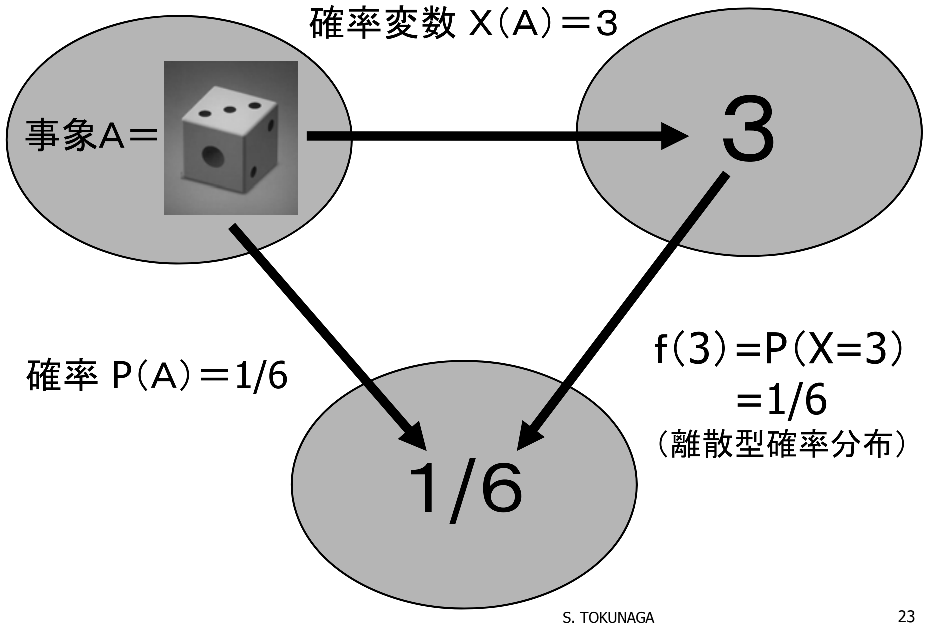
［定義］離散型確率変数Xのとる値xと，Xがその値をとる確率P(X=x)との対応関係を（Xの）確率分布という.

<u>教科書p.84例3</u>
X：サイコロを1回振ったときの目の値.
Xの確率分布（離散型）：

k	1	2	3	4	5	6
P(X=k)	1/6	1/6	1/6	1/6	1/6	1/6

★関数 f(x)=P(X=x) を「Xの確率分布」とよんで差し支えない。





Ⅰ．確率変数と確率分布の定義（4）

<u>離散型確率変数の性質:</u>

離散型確率変数Xの取り得る値を$x_1,x_2,\cdots$とする.

f(x) = P(X=x) とおくと, f は確率の性質(公理)より

$$f(x_k) \geqq 0 \ (k=1,2,\cdots) \text{ かつ } \sum f(x_k)=1$$

を満たすことがただちに導ける.

次に連続型確率変数へ

Ⅰ．確率変数と確率分布の定義（5）

<u>3−連続型確率変数の確率分布</u>

教科書p.83例2：

「ある短大の1年生から無作為に選んだ1名の身長」をXcmとすると，Xは連続型確率変数.

（とり得る値が連続的になっただけ）

では、

Xが連続型確率変数のとき，離散型の場合と同様に

「確率変数Xのとる値xと，確率P(X=x)との対応関係」

（もしくは関数 f(x)=P(X=x) そのもの）

を（連続型）確率分布と呼んで良いだろうか？

Ⅰ．確率変数と確率分布の定義（6）

そもそも

「連続型確率変数Xと確率との対応関係」

とは？

［注意］ Xが連続型確率変数のとき，

（特殊な例を除き）ほとんどすべての値xに対して

P(X=x)＝0である！

つまり

Ⅰ．確率変数と確率分布の定義（7）

連続型確率分布は
f(x)=P(X=x)のような関数で表すことはできない.

そこでこれに代わるものとして確率密度関数を導入.
［定義］

$f(x) \geqq 0$, $\int_{-\infty \leqq x \leqq \infty} f(x)dx = 1$であり,

$$P(a \leqq X \leqq b) = \int_{a \leqq x \leqq b} f(x)dx$$

であるような関数 f を，連続型確率変数Xの
確率密度関数という.
★すなわち連続型確率分布は，確率密度関数により表される.

連続型確率分布の例

<u>教科書p.85例4〈**一様分布**〉</u>

a,bを定数とするとき，密度関数

$f(x) = P(X=x) = 1/(b-a)$ （$a \leqq x \leqq b$）
$f(x) = P(X=x) = 0$ （$x < a$または$x > b$）

であらわされる確率分布を一様分布という.

- このときXは一様確率変数または**一様乱数**
- EXCEL課題で用いるRAND関数は a=0, b=1 とした一様乱数.

Ⅰ．確率変数と確率分布の定義（8）

［注意］

F(x)＝P(X≦x)

をXの累積分布関数という.

- 図11-1(b), 11-2(b)でイメージをつかんでください.
- 「累積」を省略して分布関数と呼ばれることも多く，紛らわしいので気をつけましょう.
- Excelの関数「BINOMDIST」で4つ目の引数を「TRUE」にした場合がこれに相当（→Excel実習の際に確認を）

Ⅱ．確率変数の特性値（1）

1-期待値と分散・標準偏差の定義

確率変数Xの平均（＝期待値expectation）E(X)

を次式で定義：

$E(X) := \sum x_k P(X = x_k)$ （Xが離散型）

$E(X) := \int x f(x) dx$ （Xが連続型）

（ただしf(x)はXの確率密度関数）

Xの値を繰り返し取り出したとき，それらの平均値は回数を増やすほど E(X) に近づくと考えられる

Ⅱ．確率変数の特性値（2）

μ ＝E(X)とするとき,
確率変数の分散varianceV(X)を

$$V(X) := E((X-\mu)^2)$$

で定義. すなわち,

- $V(X) = \Sigma(x_i - \mu)^2 P(X = x_i)$ (Xが離散型)
- $V(X) = \int (x-\mu)^2 f(x)\, dx$ (Xが連続型)

分散V(X)は, Xのばらつき, 変動の指標となる.
$V(X) = \sigma^2$と表すことも多い.

- Xの標準偏差standard deviation:

$$\sigma = \sigma(X) := \sqrt{\sigma^2}$$

Ⅱ．確率変数の特性値（3）

<u>期待値(平均)Eの性質:</u>

Xを確率変数,
a, b を定数とするとき,

$E(X+b) = E(X)+b$
$E(aX) = aE(X)$

が成り立つ.
以上合わせて

$E(aX+b) = aE(X)+b$

より一般には, 定数a,bと関数 f,g に対して

$E(af(X)+bg(X)) = aE(f(X)) + bE(g(X))$

(教科書には載っていません)

Ⅱ．確率変数の特性値（4）

<u>分散の性質:</u>（Xは確率変数，a，b は定数）

$$V(X+b) = E\left((X+b-E(X+b))^2\right)$$
$$=E\left((X+b-E(X)-b)^2\right)$$
$$=E\left((X-E(X))^2\right) = \mathbf{V(X)}$$

$$V(aX) = E\left((aX-E(aX))^2\right)$$
$$=E\left((aX-aE(X))^2\right)$$
$$=E\left(a^2(X-E(X))^2\right) = \mathbf{a^2V(X)}$$

以上合わせて　$V(aX+b) = a^2V(X)$

Ⅱ．確率変数の特性値（5）

★以下は有名な公式ですが，教科書には載っていません．
<u>分散の公式:</u>（$\mu=E(X)$とする）

$$V(X) = \mathbf{E(X^2) - E(X)^2}$$

［証明］

$$V(X) = E\left((X-\mu)^2\right)$$
$$= E\left((X^2-2X\mu+\mu^2)\right)$$
$$= E(X^2) - 2\mu E(X) + \mu^2 \quad \cdots(*)$$
$$= \mathbf{E(X^2) - E(X)^2}$$

<u>注意</u>:（＊）で公式

$$E(af(X)+bg(X)) = aE(f(X)) + bE(g(X))$$

を使ってます．

Ⅱ．確率変数の特性値（6）

教科書p.87例5
X：サイコロを1回振ったときの目の値　とする.

Xの確率分布（離散型）：

k	1	2	3	4	5	6
P(X=k)	1/6	1/6	1/6	1/6	1/6	1/6

$E(X) = \Sigma kP(X=k) = (1+2+\cdots+6)/6 = 7/2 = 3.5$
$V(X) = \Sigma (k-3.5)^2 P(X=k)$
$= ((1-3.5)^2+(2-3.5)^2+\cdots+(6-3.5)^2)/6$
$= 35/12 = 2.916666\cdots$

教科書p.87問題4

Z：サイコロを2回振ったときの目の和の値　とする.
このときZの確率分布（離散型）は：

k	2	3	4	･･･	7	8	･･･	12
P(X=k)	1/36	2/36	3/36	･･･	6/36	5/36	･･･	1/36

$E(Z) = \Sigma kP(Z=k)$
$= 2\cdot 1/36 + 3\cdot 2/36 + \cdots + 12/36$
$= 7 = 2\times 3.5$

$V(Z) = \Sigma (k-7)^2 P(Z=k)$
$= \cdots = 35/6 = 2\times 35/12$

期待値の加法性（その1）

実は・・・
任意の確率変数X，Yに対し

$$E(X+Y) = E(X)+E(Y)$$

が成り立っている！（期待値の加法性）

先の例2だと，サイコロを2回振ったとき
　X：1回目に出る目の値，Y：2回目に出る目の値
とすれば，

$$E(X)=E(Y) = 3.5$$

となり，Z＝X＋Yなので

$$E(Z) = 3.5+3.5 = 7$$

S. TOKUNAGA　37

期待値の加法性（その2）

Z_n：サイコロをn回振ったときの目の和
とすれば，

$$E(Z_n)=3.5n$$

も成り立つ．

さらに一般に，
　任意の定数$a_1,a_2,\cdots,a_n$と
　任意の確率変数$X_1,X_2,\cdots,X_n$に対し

$$E(\Sigma a_k X_k) = \Sigma a_k E(X_k)$$

が成り立つ（期待値の線形性）．

ところで，分散については？

S. TOKUNAGA　38

分散の加法性と確率変数の独立性

先のサイコロを2回振る例では，分散についても

$$V(Z) = 2 \times 35/12$$

が成り立っていた．
実は

Z_n：サイコロをn回振ったときの目の和

とすれば，

$$V(Z_n) = n \times (35/12)$$

も成り立っている．

しかし，「分散の加法性」

$$V(X+Y) = V(X) + V(Y)$$

は（「期待値の加法性」と違って）いつでも成り立つわけではない！

成り立つための（十分）条件：

→ **確率変数の独立性**

（詳しい説明は次回）

S. TOKUNAGA 39

第11章 確率変数と確率分布

- Ⅰ．確率変数と確率分布の定義
- Ⅱ．確率変数の特性値
  - 期待値（平均），分散など

＊＊＊今日はこの辺まで＊＊＊

- Ⅲ．確率変数の独立性
- Ⅳ．代表的な確率分布
  - 2項分布，正規分布など
- Ⅴ．中心極限定理と正規近似
- Ⅵ．標本分布

S. TOKUNAGA 40