統計（医療統計）
前期・第5回　確率変数と確率分布（3）

授業担当：徳永伸一

東京医科歯科大学教養部　数学講座

今日の授業（第5回）について

⊙授業の概要:
- 前回（第4回）の復習
- 第11章の続き（正規分布〜）

⊙前期は今日を含めてあと2回しかありません。
- 内容的にも大詰め。

もういちど Overview

- 確率（9章：6ページ）
- 記述統計（10章： 4 ページ）

確率モデル（11章：18ページ）

- 推測統計（13章：7ページ，14章：15ページ）
  - 推定（点推定、区間推定）
  - 仮説検定

（注：前期は推測統計のさわりまで）

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
- 1-確率変数の定義　・・・　離散型と連続型
- 2-離散型確率変数の確率分布
- 3-連続型確率変数の確率分布　・・・　確率密度関数

Ⅱ．確率変数の特性値
- 1-期待値と分散・標準偏差の定義
- 2-確率変数の期待値と分散の性質　・・・期待値の加法性
- 3-確率変数の標準化

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（←前回この辺で時間切れ）

Ⅴ．中心極限定理と正規近似

Ⅵ．標本分布

［復習］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）

［復習］Ⅱ．確率変数の特性値（3-5）

2−確率変数の期待値と分散の性質

（以下において Xは確率変数，a，b は定数）

期待値（平均）Eの性質：

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

分散の性質：

$$V(aX+b) = a^2V(X)$$

[復習]期待値の加法性

2−確率変数の期待値と分散の性質（続き）
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）

例:
Z_n: サイコロをn回振ったときの目の和
とすれば,

$$E(Z_n) = n \times E(Z_1) = 3.5n$$

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

[復習] Ⅱ. 確率変数の特性値 の続き

確率変数の標準化

$E(X)=\mu$, $V(X)=\sigma^2$ のとき

$$E(X-\mu)=0$$
$$V(X-\mu)=\sigma^2$$

さらに$Z=(X-\mu)/\sigma$ はとおけば

$$E(Z)=E((X-\mu)/\sigma)=0$$
$$V(Z)=V((X-\mu)/\sigma)=1$$

XからZへの変換を標準化変換,
ZをXの標準化変数という.

［復習］ Ⅲ.確率変数の独立性（1）

まず<u>（確率変数ではなくて）事象の独立性</u>
について再確認
［RECALL］
2つの事象A, Bの独立性は

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

と定義された.
これは
「AとBの成否が互いにもう一方の影響を受けない」
という状態を表していた.

［復習］ Ⅲ.確率変数の独立性（2）

では2つの確率変数X, Yの独立性はどのように定義すればよいだろうか？

すなわち
確率変数X, Yが
「互いにもう一方の影響を受けない」
とはどういうことか？

これを数式を使ってきちんと（明確に）表したい.

[復習] Ⅲ.確率変数の独立性（3）

<u>離散型確率変数の場合</u>

『X, Yのとり得るすべての値x, yについて
事象「X=x」と事象「Y=y」が独立』
ならば「XとYは独立」としよう.

すなわち言い換えると:

［定義］

X, Y: 離散型確率変数 のとき

『X, Yのとり得るすべての値x, yについて
P(X=x かつ Y=y)＝ P(X=x)P(Y=y)』
ならば「XとYは独立」という.

[復習] Ⅲ.確率変数の独立性（4）

しかし！

この定義を連続型確率変数の場合にそのまま当てはめることはできない.

なぜなら・・・

X, Yが連続型確率変数のとき
P(X=x かつ Y=y), P(X=x), P(Y=y)
はいずれも常に0なので,
P(X=x かつ Y=y)＝ P(X=x)P(Y=y)
は常に成り立つ！

そこで・・・

「値」ではなく「区間」で考える.

［復習］ Ⅲ.確率変数の独立性（5）

［定義］

確率変数X, Yについて

『任意の実数a,b,c,dに対し
事象「$a \leqq X \leqq b$」と事象「$c \leqq Y \leqq d$」が独立』,

すなわち

『任意の実数a,b,c,dに対し

$$P(a \leqq X \leqq b \text{ かつ } c \leqq Y \leqq d) = P(a \leqq X \leqq b)P(c \leqq Y \leqq d)$$』

ならば「XとYは独立」という.

【注意1】「任意の」という条件は本質的.

【注意2】上の定義は離散的確率変数にも適用できる.

［復習］ Ⅲ.確率変数の独立性（6）

<u>2-独立な確率変数の性質</u>

確率変数X, Yが互いに独立のとき

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

が成り立つ.

★左辺が＋でなく±となっていることに注意.

さらに一般には,

互いに独立な確率変数 $X_1, X_2, \cdots, X_n$ と

任意の定数 $a_1, a_2, \cdots, a_n$ 対し

$$V(\sum a_i X_i) = \sum a_i^2 V(X_i)$$

が成り立つ（分散の加法性）.

[復習] Ⅲ.確率変数の独立性（7）

つまり

Z_n：サイコロをn回振ったときの目の和

としたとき，

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

が成り立つのは，

「k回目に振ったのサイコロの目」により定まる
確率変数X_kが 互いに独立 で，分散の加法性：

$$V(Z_n) = V(X_1) + V(X_2) + \cdots + V(X_n)$$

が成り立つからでした！

<u>その他の性質</u>

X，Yが互いに独立のとき

$$E(X \cdot Y) = E(X)E(Y)$$

も成り立っている（この授業では使う機会がないが）．

[復習]余談

「確率変数と事象を混同しないように！！」
というのは毎年さんざん強調するのですが，
試験で

「2つの確率変数X，Yが独立である」
ことの定義を書け

という問題を出すと

$$P(X \cap Y) = P(X) \cdot P(Y)$$

と答案に書く人が後を断たない．
（「P(X)とかP(Y)って何よ？」と突っ込むところ）

これをやると，まず不合格です．
（いわゆる「地雷を踏んだ」状態）

[復習] 期待値と分散の性質まとめ１

（以下Xは確率変数, a, b 等は定数）

期待値（平均）Eの性質:

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

分散の性質:

$$V(aX+b) = a^2V(X)$$

<u>確率変数の標準化（上の性質の応用）</u>

$E(X)=\mu$, $V(X)=\sigma^2$ のとき

$Z=(X-\mu)/\sigma$ はXの標準化変数と呼ばれ,

$$E(Z)=0,\ V(Z)=1$$

[復習] 期待値と分散の性質まとめ２

期待値の加法性

任意の確率変数X, Yに対し

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

さらに一般には,

任意の定数$a_1,a_2,\cdots,a_n$と

任意の確率変数定数$X_1,X_2,\cdots,X_n$に対し

$$E(\Sigma a_kX_k)=\Sigma a_kE(X_k)$$

が成り立つ（期待値の線形性）.

分散の加法性

確率変数X, Yが互いに独立のとき

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

さらに一般には,

互いに独立な確率変数 $X_1,X_2,\cdots,X_n$ と

任意の定数 $a_1,a_2,\cdots,a_n$ 対し

$$V(\Sigma a_iX_i) = \Sigma a_i^2V(X_i)$$

が成り立つ .

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
1-期待値と分散・標準偏差の定義
2-確率変数の期待値と分散の性質
3-確率変数の標準化
Ⅲ. 確率変数の独立 ←ここまできた
Ⅳ. 代表的な確率分布 ←次ここ
- 2項分布, 正規分布など

＊＊＊前回この辺まで＊＊＊

Ⅴ. 中心極限定理と正規近似
Ⅵ. 標本分布

[復習] Ⅳ.代表的な確率分布 (0)

このセクションに出てくる確率分布
離散型確率分布:
・ 2項分布
・ ポアソン分布
連続型確率分布:
・ 正規分布

後で学ぶもの(「Ⅵ.標本分布」および13章以降)
・ χ^2分布, t分布, F分布(いずれも連続型)

[復習] Ⅳ.代表的な確率分布（1）

1- 2項分布

まず準備として，ベルヌーイ(Bernoulli)試行の概念を導入する.

ベルヌーイ試行（独立試行）の定義：

以下の3条件を満たす試行をベルヌーイ試行という.

1. 各試行において，A（成功）とA^C（失敗）の2通りの結果のみ起こる.
2. 各試行において，P（A）＝pは一定.
3. 各試行（の結果となる事象）は互いに独立.

[復習] Ⅳ.代表的な確率分布（2）

2項分布の定義：

1回ごとのA（成功）の確率がpであるようなn回のベルヌーイ試行において，n回中A（成功）が起こる回数をXとすると

$$P(X=x) = {}_nC_x \cdot p^x \cdot (1-p)^{(n-x)}$$

となる.

この式で表される（離散型）確率分布を

二項分布 binomial distribution といい，

記号B（n,p）で表す.

また確率変数XがB（n,p）に従うとき，

$$X \sim B(n,p)$$

と書く.

［復習］ Ⅳ.代表的な確率分布（3）

2項分布B（n,p）に従う確率変数Xは,
B（1,p）に従うn個の独立な確率変数$X_1, X_2, \cdots, X_n$
の和とみなすこともできる.

ここで，各X_iについて期待値と分散は

$E(X_i) = 0 \cdot (1-p) + 1 \cdot p = p$

$V(X_i) = (0-p)^2 (1-p) + (1-p)^2 p = p(1-p)$

よって期待値および分散の加法性より

$$E(X) = nE(X_i) = np$$

$$V(X) = nV(X_i) = np(1-p)$$

（前回このあたりまで）

Ⅳ.代表的な確率分布（4）

例題（平成15年度試験問題より）
サイコロを n 回振るとき、
「1または2の目が出る回数」
を X とおく。以下の問いに答えよ。

（1）X の分布を記号で表すとB(_______，_______)。
また E(X)=(　　　　）、V(X)=(　　　　）である。
（nを用いて表せ）

（2）n=5のとき、P(X ≦3)を求めよ。
［解答］
$P(X \leqq 3) = P(X=0)+P(X=1)+P(X=2)+P(X=3)$

$= {}_5C_0 (1/3)^0 (2/3)^5 + {}_5C_1 (1/3)^1 (2/3)^4 + \cdots$

Ⅳ.代表的な確率分布（5）

2- ポアソン分布

[定義]確率変数Xの分布が

$$f(X)=P(X=x)=e^{-\lambda}\lambda^{x}/x!$$

で表されるとき、Xの分布をポアソン分布Poisson distributionといい、*Po*(λ)で表す。

ポアソン分布の期待値と分散:

Xがポアソン分布*Po*(λ)に従うとき、

$$E(X)=V(X)=\lambda$$

25　S. TOKUNAGA

Ⅳ.代表的な確率分布（6）

3-正規分布normal distribution

- もっとも重要かつ役に立つ分布．単に「連続型確率分布の代表例」というだけではない．
- 後述の「中心極限定理」により，他の確率分布の問題も正規分布で近似して解決できるケースが少なくない．
- 別名ガウス分布．

26　S. TOKUNAGA

Ⅳ.代表的な確率分布（7）

◉ 正規分布の密度関数：

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

ここで，μ は期待値，σ^2は分散となっている.

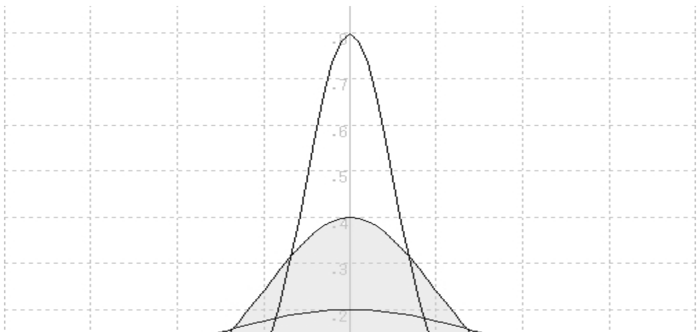
◉ 期待値μ，分散σ^2の正規分布をN(μ,σ^2)で表す.

◉ XがN(μ,σ^2)に従うとき，X～N(μ,σ^2)と書く.

◉ N(0, 1)を標準正規分布という.

Ⅳ.代表的な確率分布（8）

◉ N(0, 1)，N(0, 2)，N(0, 0.5)の密度関数のグラフ：



（1）直線x ＝μ（＝0）に関して対称.

（2）x ＝μ（＝0）で最大値$1/\sqrt{(2\pi\sigma^2)}$をとり、x ＝±∞で0.

（★分散が大きいほど,「山」は低くなり「裾野」が広がる）

（3）x ＝±σで変曲点.

Ⅳ.代表的な確率分布（9）

[1]正規分布の線形変換と標準化

Xが正規分布に従うとき，その線形変換（Y＝aX+bの形の変換）も正規分布に従う.

したがって，

X～ $N(\mu, \sigma^2)$ のとき

Xの標準化変数 $Z=(X-\mu)/\sigma$ は 標準正規分布 $N(0,1)$ に従う.

Ⅳ.代表的な確率分布（10）

Xの区間と確率の関係（教科書p.94図11-5参照）

$P(\mu-\sigma \leqq X \leqq \mu+\sigma) = P(-1 \leqq Z \leqq 1) = 0.6827$

$P(\mu-2\sigma \leqq X \leqq \mu+2\sigma) = P(-2 \leqq Z \leqq 2) = 0.9545$

$P(\mu-3\sigma \leqq X \leqq \mu+3\sigma) = P(-3 \leqq Z \leqq 3) = 0.9973$

★教科書p.135付表1で値を確認せよ.

（試験でこの表を使用するので慣れておく必要あり）

$P(\mu-1.645\sigma \leqq X \leqq \mu+1.645\sigma)$

$= P(-\mathbf{1.645} \leqq Z \leqq \mathbf{1.645}) = 0.900$

$P(\mu-1.96\sigma \leqq X \leqq \mu+1.96\sigma)$

$= P(-\mathbf{1.96} \leqq Z \leqq \mathbf{1.96}) = 0.950$

（特に最後の2つはよく用いられる値）

★教科書p.135付表1，p.136付表2で値を確認せよ.

Ⅳ.代表的な確率分布（11）

[2]正規分布の再生性

確率変数X_1, X_2が互いに独立で、それぞれ正規分布 $N(\mu_1, \sigma^2_1)$, $N(\mu_2, \sigma^2_2)$に従うとき，和X_1+X_2は

正規分布$N(\mu_1+\mu_2,\ \sigma^2_1+\sigma^2_2)$

に従う.

注意:

$E(X_1+X_2)=\mu_1+\mu_2$,　$V(X_1+X_2)=\sigma^2_1+\sigma^2_2$

はそれぞれ「期待値の加法性」「分散の加法性」から導けるので,

「和をとっても正規分布」

という部分のみ本質的.

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

1-期待値と分散・標準偏差の定義

2-確率変数の期待値と分散の性質

3-確率変数の標準化

Ⅲ. 確率変数の独立

Ⅳ. 代表的な確率分布

- 2項分布，正規分布など

＊＊＊ここまできた＊＊＊

Ⅴ. 中心極限定理と正規近似　　←次ここ

Ⅵ. 標本分布

Ⅴ.中心極限定理と正規近似（1）

中心極限定理1

[仮定]$X_1, X_2, \cdots, X_n$が

（任意の！）同じ分布に従う独立な確率変数

ならば,

[結論]$n \to \infty$のとき,

和$X_1 + X_2 + \cdots + X_n$の分布は

正規分布に収束する！

Ⅴ.中心極限定理と正規近似（2）

中心極限定理2（言い換え）

「互いに独立な確率変数$X_1, X_2, \cdots, X_n$の分布が同一で,

$$E(X_k) = \mu,\ V(X_k) = \sigma^2 \quad (k=1,2,\cdots,n)$$

であるとき，nが十分大きければ，和ΣX_kの分布は$N(n\mu, n\sigma^2)$にで近似できる」

【注意】仮定すべき条件は独立性と同一分布性のみ．元の分布は任意．

Ⅴ.中心極限定理と正規近似（3）

二項分布の正規近似

中心極限定理により，nが十分大きいとき，
B(n,p)はN(np, np(1−p))で近似できる.
よって標準化変数

$$Z = (X - E(X)) / \sqrt{V(X)}$$
$$= (X - np) / \sqrt{(np(1-p))}$$

は近似的にN(0,1)に従う.

∵ B(n,p)に従う確率変数は， B(1,p)に従う 独立なn個の確率変数の和と見なせるから.

平成16年度試験問題より

問題文

「野球における打率とは、安打（ヒット）数を打数で割ったものである。イチロー選手は今シーズン、9月27日までに673打数251安打の成績を残している。」

「残り7試合で30打数あるとし、これまでと同じ打率でヒットを打ち続けるとすると、1シーズン257安打の大リーグ記録を樹立する（あと6本以上ヒットを打つ）確率はいくらか。」

「2項分布の正規近似を利用して計算せよ。」

考え方

⦿「これまでと同じ打率」＝251/673≒0.373

⦿ X:残り30打数の安打数とすれば

$$X \sim B(30, 0.373)$$

よって

⦿ 期待値$\mu = E(X) = 30 \times 0.373 \fallingdotseq 11.2$

⦿ 分散$V(X) = 30 \times 0.373 \times 0.627 \fallingdotseq 7.0$

すなわち

⦿ $X \sim N(11.2,\ 7.0)$ と近似できる。

⦿ 標準化変数$Z = (X - 11.2)/\sqrt{7.0}$

より

$$P(X \geqq 6) \fallingdotseq P(Z \geqq (6 - 11.2)/\sqrt{7.0}) = \cdots$$

と計算.