統計（医療統計）
前期・第6回　推測統計への第一歩

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6回）について

- 授業の概要：
  - 前回（第5回）の復習
    ～第11章Ⅵ標本分布~13章（推定）のさわり
- 前期は今日でおしまいです.
- 試験は（たぶん）約1ヵ月後なのでよく復習を。
  - 授業スライドによく目を通すこと.
  - 試験は電卓持ち込み可です（ただし計算以外の余計な機能がないものに限る）．当日忘れても貸出しはしないので注意してください.
- 疑問点は早めに解決を（質問は随時受け付けるので、Eメールを活用すること）。

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（←前回この辺から）

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 （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

［復習］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.

[復習]事象の独立性と確率変数の独立性・まとめ

事象の独立性

「事象Aと事象Bが独立」

$$\Leftrightarrow \quad P(A \cap B) = P(A) \cdot P(B)$$

確率変数の独立性

「確率変数Xと確率変数Yが独立」

⇔『任意の実数a,b,c,dに対し

事象「a≦X≦b」と事象「c≦Y≦d」が独立』

⇔『任意の実数a,b,c,dに対し

P(a≦X≦b かつ c≦Y≦d)= P(a≦X≦b)P(c≦Y≦d)』

【注意1】「任意の」という条件は本質的.

【注意2】確率変数Xに対し,「P(X)」と書いたらアウト!

[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:

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

確率変数の標準化（上の性質の応用）

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

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

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

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

が成り立つ.

[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binomial distribution

X～ B(n,p)のとき

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

期待値，分散の公式：

$$E(X)=np\ ,\ V(X)=np(1-p)$$

∵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．

2‐ ポアソン分布Poisson distribution

X～ $Po(\lambda)$のとき

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

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

［復習］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.

[復習] Ⅳ.代表的な確率分布（7）

- 正規分布の密度関数:

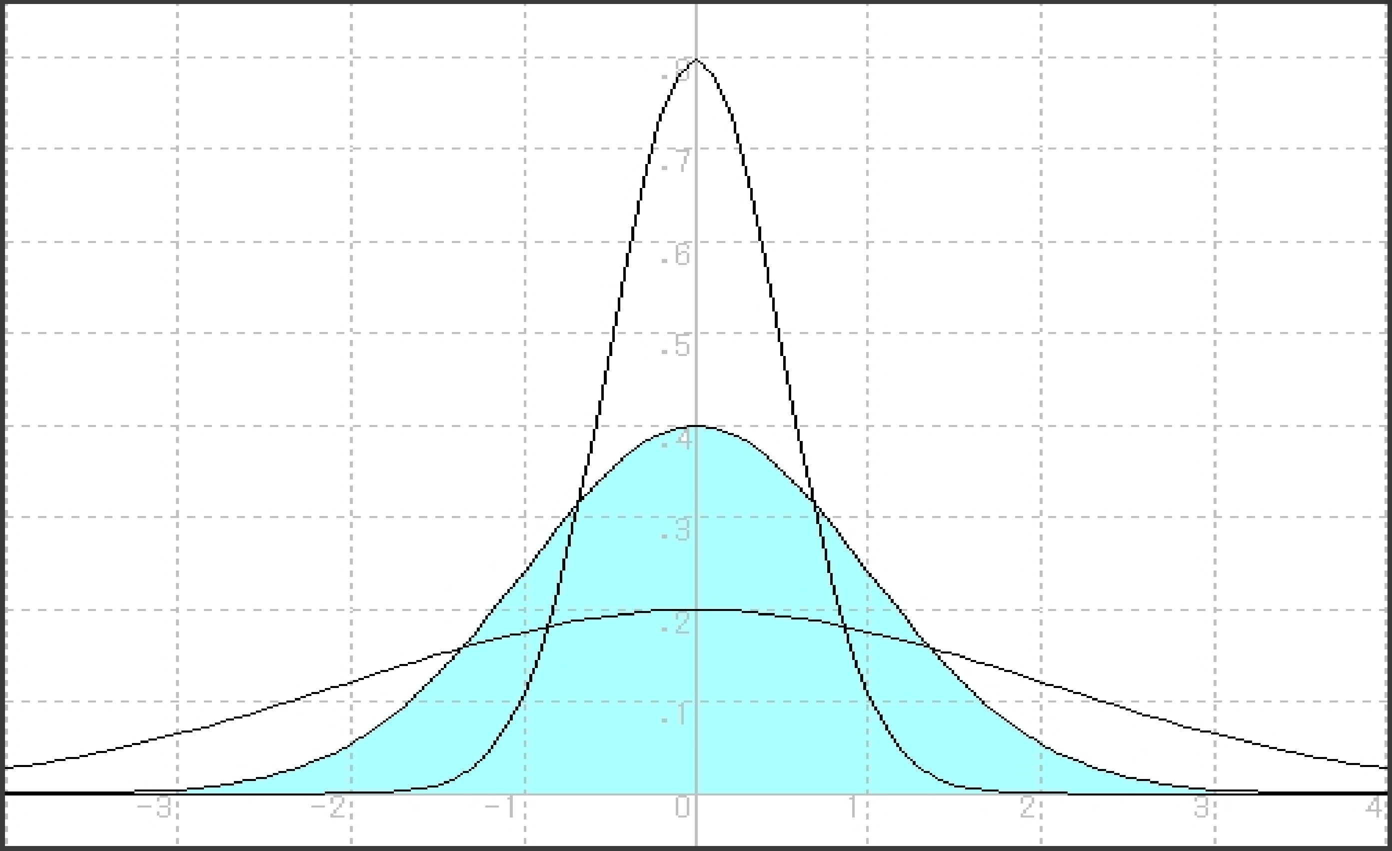
$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

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.

[復習] 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 $=\mu$ (=0)に関して対称.

(2)x $=\mu$ (=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 $=\pm\sigma$で変曲点.

［復習］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（μ , σ^2） のとき

Xの標準化変数Z＝（X－μ）/ σ

は 標準正規分布N（0,1） に従う.

［復習］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）

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

$= P(-1.645 \leqq Z \leqq 1.645) = 0.900$

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

$= P(-1.96 \leqq Z \leqq 1.96) = 0.950$

（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.

［復習］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:

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

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

[復習] 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！

［復習］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$N(n\mu, n\sigma^2)$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．

［復習］Ⅴ.中心極限定理と正規近似（3）

<u>二項分布の正規近似</u>

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.

（前回この辺まで）

Ｖ.中心極限定理と正規近似（4）

（前回の内容への補足）

半整数補正

- nが大きければかなり良い近似であると思われるが，nが小さいときはどのくらい誤差が出るのだろうか？
  - p.97問題10のケースで厳密値と正規近似の値を比較せよ．

- nが小さいときに少しでも誤差を減らす方法はないか？

⇒X～B(n,p)とする．整数a,bに対しP(a≦X≦b)を正規近似で求める際，P(a−0.5≦X≦b+0.5)と補正してから計算した方が誤差が減る．この補正を「不連続補正」ないし「半整数補正」といい，特にnが小さいときに効果的．

  - 区間を広げる方向に0.5ずらす（教科書p.97図11−7で確認 ）．
  - 再びp.97問題10で誤差の減少を確認．

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

Ⅴ．中心極限定理と正規近似　←ここまできた

Ⅵ．標本分布　←次ここ

Ⅵ.標本分布（1）

1−母集団分布と標本分布

KEYWORDS：母集団⇔標本，無作為抽出，母集団分布，統計量，標本分布

★母集団から無作為抽出した個々のデータの値を確率変数をみなして，確率分布の理論を適用することができる！

2−標本平均の分布

- 個々の標本データの値 $X_1, X_2, \cdots, X_n$ はもちろん確率変数と見なすことができる.
- 標本平均 $\bar{X}$ も1つの確率変数とみなすことができる！

（一定の大きさの標本を繰り返し抽出し，その度に標本平均の値を計算すれば，「標本平均の分布」を観察することができる）.

よって・・・

Ⅵ.標本分布（2）

標本平均の期待値と分散･標準偏差

$X_1, X_2, \cdots, X_n$を平均μ，分散σ^2である母集団から無作為抽出した標本とするとき，

$X_1, X_2, \cdots, X_n$はそれぞれ，期待値μ，分散σ^2の互いに独立な確率変数と見なせる．

よって標本平均$\bar{X}$について

$$E(\bar{X}) = \mu \times n \times (1/n) = \mu$$

$$V(\bar{X}) = \sigma^2 \times n \times (1/n)^2 = \sigma^2/n$$

（期待値･分散の加法性↑）　（↑積に関するE,Vの性質より）

$$\sigma(\bar{X}) = \sqrt{V(\bar{X})} = \sqrt{(\sigma^2/n)} = \sigma / \sqrt{n}$$

Ⅵ.標本分布（3）

正規母集団を仮定すると・・・

定理（正規分布の性質より）

$X_1, X_2, \cdots, X_n$ をN(μ, σ^2)に従う母集団から無作為抽出した標本とすると

- 和 $\Sigma X_k \sim N(n\mu, n\sigma^2)$
- 標本平均 $\bar{X} \sim N(\mu, \sigma^2/n)$

さらに $\bar{X}$ の標準化変数Zについて：

- $Z = (\bar{X} - \mu) / \sqrt{(\sigma^2/n)} \sim N(0,1)$

Ⅵ.標本分布（4）

さらに！

$$n \times \bar{X} = X_1 + X_2 + \cdots + X_n$$

であるから,

nが十分大きければ, 母集団分布が正規分布でなくても中心極限定理によって標本平均の分布を正規分布で近似できる！

注意:

- 同一分布性:同一の母集団から抽出したから
- 独立性:無作為抽出により保証される
- 正規分布に従う確率変数はnで割っても正規分布.

したがって・・・

Ⅵ.標本分布（5）

<u>定理（中心極限定理の系）</u>

$X_1, X_2, \cdots, X_n$を平均μ，分散σ^2である任意の母集団から無作為抽出した大きさnの標本とするとき，

標本平均$\bar{X}$の分布は，nが十分大きければ，正規分布$N(\mu, \sigma^2/n)$で近似できる．

さらに$\bar{X}$の標準化変数$Z=(\bar{X}-\mu)/\sqrt{(\sigma^2/n)}$は標準正規分布N(0,1)で近似できる．

【注意】母集団分布が任意でよいことにあらためて注目．これにより，（十分大きい標本さえ得られれば）未知の分布を持つ母集団の母平均を推定・検定する際，正規分布が利用できる！

標本平均の分布・まとめ（対比して再確認）

定理（正規分布の性質より）

$X_1, X_2, \cdots, X_n$を正規分布$N(\mu, \sigma^2)$に従う母集団から無作為抽出した標本とすると

$$\text{標本平均 } \bar{X} \sim N(\mu,\ \sigma^2/n)$$

定理（中心極限定理の系）

$X_1, X_2, \cdots, X_n$を平均μ，分散σ^2である任意の母集団から無作為抽出した標本とするとき，
標本サイズnが十分大きければ，近似的に

$$\text{標本平均 } \bar{X} \sim N(\mu,\ \sigma^2/n)$$

となる．

★「3−その他の重要な標本分布」は後回し（後期に）．

S. TOKUNAGA

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

Ⅴ．中心極限定理と正規近似
Ⅵ．標本分布　←ここまで終了

いよいよ13章 推定 へ

第13章 推定

Ⅰ．母集団と標本

Ⅱ．点推定

- 不偏性，不偏推定量

Ⅲ．区間推定

Ⅳ．母平均の区間推定

1. 母分散が既知のとき
2. 母分散が未知のとき

Ⅴ．母分散の区間推定

Ⅵ．母比率の区間推定

Ⅰ．母集団と標本

KEYWORDS：

- 母集団⇔標本sample，無作為標本
- 母平均，母分散
- 層別，層別抽出stratified sampling
- 乱数

Ⅱ．点推定（1）

KEYWORDS：

- 母数，母集団パラメータ
  - 母平均，母分散
- （標本）統計量，統計値（統計量の実現値）
  ・・・標本平均，標本分散，標本比率
- 推定量（母数の推定に用いる統計量），
  推定値（推定量の実現値）

点推定と区間推定

- 点推定：「無作為に選んだ標本から数値を得て，それから推定量の推定値を計算し，その推定値がイコール母数であるとする推定方法」
- 区間推定：「推定値から，ある確率である数値の区間の中にある，とする推定方法」

Ⅱ．点推定（2）

推定量が「不偏unbiasedである(偏りがない)」とは:

- 「対応する母数より大きい(or小さい)値が得られやすい」といった傾向がない.
- その推定量を繰り返し実測し、得られた値(推定値)の平均値は，繰り返しの回数を増やすほど対応する母数に近づく.

厳密には:

- ［定義］母数θの推定量θ'に対し,

$$E(\theta')=\theta$$

のときθ'をθの不偏推定量(不偏性を持つ推定量)であるという.

Ⅱ．点推定（3）

$X_1, X_2, \cdots, X_n$に対し

不偏分散 $U^2 := \{\sum(X_k - \bar{X})^2\} / (n-1)$

は，母分散σ^2の不偏推定量.

- すなわち，$E(U^2) = \sigma^2$ である(証明は割愛）．

ということは

- 分散$S^2 := \{\sum(X_k - \bar{X})^2\} / n$ については，

$$E(S^2) = E(((n-1)/n)U^2) = ((n-1)/n)\sigma^2$$

- つまりS^2の実測値は，母分散より小さめの値をとる傾向にある(すなわち不偏でない)．

不偏性に関する補足

- 標本平均 $\bar{X} := \Sigma X_k/n$も母平均μの不偏推定量.

$\because E(\bar{X}) = E(\Sigma X_k/n) = \Sigma E(X_k)/n = \mu$

- たとえばn=3のとき$X' = (X_1 + X_2 + 2X_3)/4$ とおいてもX'はμの不偏推定量(確認せよ).
- 不偏分散の平方根$U = \sqrt{(U^2)}$は, σの不偏推定量ではない!

$\because E(U) = E(\sqrt{(U^2)}) \neq \sqrt{(E(U^2))} = \sigma$

第13章 推定

Ⅰ．母集団と標本

Ⅱ．点推定

- 不偏性，不偏推定量

＊＊＊試験範囲ここまで＊＊＊

Ⅲ．区間推定

Ⅳ．母平均の区間推定

1. 母分散が既知のとき
2. 母分散が未知のとき

Ⅴ．母分散の区間推定

Ⅵ．母比率の区間推定