統計Ⅱ（医療統計Ⅱ）
第1回　概要と統計Ⅰの復習

授業担当：徳永伸一

東京医科歯科大学教養部　数学講座

あらためて注意しておきたいこと
（前期のはじめに注意したこと＋α）

- 授業は今日を含め（たった）6回。
- 成績評価は前期試験＋後期試験で。
  - 後期の方が比重が大きいですが前期の出来が悪かった人はハンデがあると思ってください。
  - 後期試験にも一部前期授業の範囲（特に試験で出来が悪かったところ）を出題するので復習も怠りなく。
- 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
    おきたいと思っていますが、催促されないとサボりがちです。気軽に催促してください。
- 要点を効率よく押さえましょう。
  - 「計算方法」ではなく「統計的手法」を理解すること。
  - その前提となる概念・原理の理解も重視（当然だ！）。
  - 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：

- 授業全体（前期＋後期）のOverview
- 前期試験を復習しつつ、前期の内容を振り返る。

[再確認] 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」）

[再確認] 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（旧教科書p.1）
- 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・・・

[再確認] 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

[再確認] 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。

[再確認] 学部教員の声

某会議にて医学部T.Y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」

（隣にいらした歯学部○教授も頷く）

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。

[再確認] 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（電卓があれば）中学生でもできる．
  - 後期は多少公式が増えますが、前期の内容をしっかり理解していればむしろ前期より楽。
- 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．

[再確認] 教科書について

- 引き続き
  「臨床検査学講座 数学／統計学」（医歯薬出版）
  を使用します（13章途中から）。
- 徳永は後期の授業範囲（13章・14章）は執筆していませんが、
  - 第11章 確率変数と確率分布（18ページ）

  は後期の範囲を理解する上でも重要なので随時参照してください。
- でも教科書は絶対的なものではありません。
  - 残念ながらミスプリも多い（初版第1刷だし）
  - 判明しているミスプリについては、近日中に一覧表を作ってアップロードするので参照してください。また授業の進行に応じて順次指摘します。

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）（13章）
  - 仮説検定（14章）

第9章の概要

- 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

[復習] III. 確率の定義と性質

公理的確率

Ωの各部分集合（事象）に実数を対応させる写像(*) Pが以下の3条件（＝確率の公理）を満たすとき，Pを確率という．

（(*) の部分は修正しました。教科書も修正）

(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

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

[復習] IV.条件付き確率と事象の独立性（1）

- 「AのもとでのBの条件付き確率$P(B \mid A)$」の定義：

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

- 確率の乗法定理：

$$P(A \cap B) = P(A) \cdot P(B \mid A) \quad \cdots(1)$$

- 「AとBは（互いに）独立」（定義）

$$\Leftrightarrow 「P(A \cap B) = P(A) \cdot P(B)」 \quad \cdots(2)$$

のとき，であるという．

- (1)(2)より、AとBが独立のとき

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

[復習] 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」 ⇔ $P(A \cap B) = 0$

★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．

[復習] ベイズの定理 Bayes' Theorem

事象$A_1, A_2, \cdots A_r, B \in \Omega$ 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．

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

[復習] もういちど念押し

「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．

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（すなわち結論だけ）書く人は少なくない．

[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

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

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公式となる）。

平成17年度統計Ⅰ試験問題 [1]

問題文:

「4つの事象 X,Y,Z,W に対し、ベイズの定理を用いて確率 P(X | W) の値を計算したい。このとき以下の問いに答えよ。

(1) ベイズの定理の仮定 (X,Y,Z,Wの満たすべき条件) および結論を正確に述べよ。

(2) 条件付確率の定義に基づいて、ベイズの定理の結論を導く過程を説明せよ。式変形だけでなく、定理の仮定をどの部分で利用したで利用したかを明示すること。」

- 「ベイズの定理」に関する問題.
- 「ほぼ毎回試験に出している」と予告してあったので出来はまずまずだが、(2) は説明が不完全な人も多い。

[復習] 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.

[復習] <u>ベイズの定理</u>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 ）

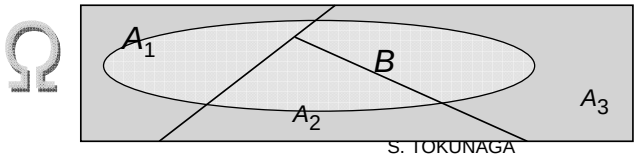
$$P(A_1|B) = \frac{P(A_1 \cap B)}{P(B)}$$

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

（↑ ここで仮定①②を使用）

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

（証明終わり）



[復習]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] III．データの特性値（3）

<u>1-散布度</u>

[1] 分散variance と 標準偏差standard deviation

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

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

標準偏差＝「σ^2の正の平方根」、すなわち

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

[復習] III．データの特性値（4）

[2] 不偏分散unbiased deviation

データ $x_1, x_2, \ldots, x_n$ 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）

[復習] III．データの特性値（5）

<u>不偏分散についての補足</u>

★本によっては

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。

★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。

★昨年度までの教科書でも「標本分散＝不偏分散」としていたので、再履修の人は混乱のないように！

(10章ここまで)

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

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

[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] II.確率変数の特性値（3）

2-確率変数の期待値と分散の性質

（以下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：

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

分散の公式：★教科書には載っていません.

$V(X) = E(X^2) - E(X)^2$

[復習] III.確率変数の独立性（1）

まず（確率変数ではなくて）事象の独立性について再確認

2つの事象A, Bの独立性は

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

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.

確率変数の独立性についても

確率変数X, Yが

「互いにもう一方の影響を受けない」

という状態を数式を用いて明確に定義したい.

[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 (離散型の場合)]
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.

【注意】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.

[復習] III.確率変数の独立性（5）

[定義 (一般の場合)]
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≦X≦b」と事象「c≦Y≦d」が独立』,
すなわち
『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) 』
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] 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

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

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.

平成17年度統計Ⅰ試験問題 [2]

「正四面体の各面に1～4の整数を対応させた「正四面体サイコロ」を考える。この正四面体サイコロを n 回振るとき、n回目に出た値を X_n とする。以下の問いに答えよ。ただし (1) ～ (4) においては、この正四面体サイコロの材質は均質であるとし、1回振るごとに1～4の各整数がそれぞれ等しい確率で出るとして考えよ。

(1) 各X_k (k=1,2,…) の期待値と分散を定義に従って求めよ。

(2) 「X_1+X_2が偶数」という事象 (Aとする) と「$X_1 \geqq 3$ かつ $X_2 \geqq 3$」という事象 (Bとする) は独立か。「事象の独立性」の定義に従って判定せよ。

(3) 確率変数Yを「X_1+X_2を2で割った余り」、確率変数Zを「X_1, X_2のうち3以上であるものの個数」でそれぞれ定義する。YとZは独立か。「確率変数の独立性」の定義に従って判定せよ (HINT: Yは0,1の2通り、Zは0,1,2の3通りだから、Y,Zの値の組合せは計2×3＝6通り)。

平成17年度統計Ⅰ試験問題 [2] 続き

(4) $W=(X_1+2X_2+3X_3)/6$とする。Wの期待値と分散を求めよ。

(5) 実は正四面体の材質が均質でなく、値によって出る確率が異なっていたとする。このとき各X_kが従う分布の期待値をWによって推定できるだろうか。Wの推定量としての妥当性について、不偏性などを踏まえつつ論じよ。

- (4) でのE (W) の計算過程においてWの不偏性がわかる。
  - 不偏性の定義を再確認！
- 平均 $(X_1+X_2+X_3)/3$との比較では？
  - 他の条件が同じなら、分散が小さい方が良い推定量 (ばらつきは小さい方が誤差が少ないはず)

[復習] 13章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．

[復習] 11章Ⅳ.代表的な確率分布（1）

1- 2項分布binomial distribution

X～ B（n,p）のとき

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

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

∵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*（λ）のとき

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

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

平成17年度統計Ｉ試験問題 [3]（2）

問題文：

[3]（2）インフルエンザの予防注射をした人のうち、5%は注射に対して望ましくない過敏な反応を示すという。注射をした200人中16人以上の人がこの反応を示す確率を厳密に表す式を、2項係数、Σ記号などを用いて記せ（計算しなくてよい）。

[復習] Ⅳ.代表的な確率分布（2）

3- 正規分布normal distribution

[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，$X \sim N(\mu, \sigma^2)$ 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．**

[2]正規分布の再生性

確率変数X_1, X_2が互いに独立で、
$X_1 \sim N(\mu_1, \sigma^2_1)$, $X_2 \sim N(\mu_2, \sigma^2_2)$ のとき，

$$X_1 + X_2 \sim N(\mu_1 + \mu_2, \sigma^2_1 + \sigma^2_2)$$

[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$X_1, X_2, \ldots, X_n$の分布が同一で，

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

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．

平成17年度統計Ｉ試験問題 [3]（1）

問題文：

[3]（1）2項分布が正規分布で近似できる理由を中心極限定理に基づいて説明せよ。ただし中心極限定理の仮定について具体的に言及すること。

[復習] 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．

[復習] Ⅴ.中心極限定理と正規近似（4）

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(図で確認)．
  - 再びp.97問題10で誤差の減少を確認．

平成17年度統計Ⅰ試験問題[3]（3）

問題文：

[3] (2) インフルエンザの予防注射をした人のうち、5%は注射に対して望ましくない過敏な反応を示すという。注射をした200人中16人以上の人がこの反応を示す確率を厳密に表す式を、2項係数、Σ記号などを用いて記せ

(3) (2)の確率の値を、正規近似によって求めよ。ただし半整数補正を行うこと。

[復習] 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 $X_1, X_2, \ldots, X_n$ はもちろん確率変数と見なすことができる．
- 標本平均 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・・・

[復習] 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，標本平均 X' について

$$E(X') = \mu \times n \times (1/n) = \mu$$

$$V(X') = \sigma^2 \times n \times (1/n)^2 = \sigma^2/n$$

（期待値・分散の加法性↑）　（↑積に関するE,Vの性質より）

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

正規母集団を仮定すると・・・

[復習] VI.標本分布（2）

定理（正規分布の性質より）教科書p.99

$X_1, X_2, \cdots, X_n$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

（★教科書p.99冒頭でN（μ, σ）となっているのはミスプリ）

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

さらにX'の標準化変数Zについて：

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

さらに・・・

[復習] VI.標本分布（3）

中心極限定理により，

nが十分大きければ，母集団分布が正規分布でなくても和$X_1 + X_2 + \ldots + X_n = n \times X'$（$X'$は標本平均）の分布は正規分布で近似できる．

→正規分布に従う確率変数はnで割っても正規分布に従うから

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

したがって・・・

[復習] VI.標本分布（4）

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大きさnの標本とするとき，
標本平均X'の分布は，nが十分大きければ，
正規分布N（μ，σ^2/n）で近似できる．

さらにX'の標準化変数$Z = (X' - \mu) / \sqrt{(\sigma^2/n)}$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！

[復習] 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（μ，σ^2）に従う母集団から無作為抽出した標本とすると

標本平均X' ～ N（μ，σ^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

標本平均X' ～ N（μ，σ^2/n）

となる．

★「3-その他の重要な標本分布」は後まわし
（推定・検定で利用）．

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
- 母分散が既知のとき

＊＊＊前期試験範囲ここまで＊＊＊

- 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

Ⅳ．母平均の区間推定

<u>母平均μの区間推定（母分散既知の場合）</u>

問題設定：

- 標本サイズn，標本平均 $\bar{X}$ の実測値が与えられており，母分散 σ^2 は既知とする．
- 母集団分布・・・正規分布なら好都合だが，任意の分布でもnが大きければ，標本平均の分布は中心極限定理により正規分布で近似可能．

以上の条件のもとで，

「μの 100γ％信頼区間」

を求める（γは信頼度＝信頼係数で，具体的には 0.95, 0.99, 0.90など）．

Ⅳ．母平均の区間推定

母平均μの区間推定（母分散既知）の解法

- $\bar{X} \sim N(\mu, \sigma^2/n)$ と近似（中心極限定理による）．
  - 注意：正規母集団を仮定すれば厳密．
- $Z = (\bar{X} - \mu) / (\sigma/\sqrt{n})$ と標準化すると $Z \sim N(0, 1)$
- $P(-z(\alpha/2) \leqq Z \leqq z(\alpha/2)) = \gamma$
  - ただし $\gamma = 1-\alpha$．$z(\alpha)$ は $P(Z \geqq z(\alpha)) = \alpha$ を満たす値．
  - たとえば $\gamma=0.95$ のとき $\alpha/2=0.025$，$z(\alpha/2) = z(0.025) = 1.96$
  - $z(\alpha/2)$ は「上側 $\alpha/2$ 点」と呼ばれる．
- ↑を同値変形すると

$$P\left(\bar{X} - z(\alpha/2)\ \sigma/\sqrt{n} \leqq \mu \leqq \bar{X} + z(\alpha/2)\ \sigma/\sqrt{n}\right) = \gamma$$

よって「μの（100×γ）%信頼区間」は

$$\left(\bar{X} - z(\alpha/2)\ \sigma/\sqrt{n}\ ,\ \bar{X} + z(\alpha/2)\ \sigma/\sqrt{n}\right)$$

平成17年度統計Ⅰ試験問題［4］

［4］あるビタミン剤のロット（同一の生産条件で作られた製品の集まり）における1錠当たりのビタミンC含有量について、大きさ 25 の無作為標本をとって検査を行ったところ、平均 X'=300 および分散 $s^2=729$ を得た。

ここで、含有量の母分散 σ^2 は母平均 μ の値と関係なく $\sigma^2=900$ で一定であるとする。

このとき以下の（a）～（f）にあてはまる数値、数式、文字または言葉を答えよ。

（1）標本の大きさ n が十分大きいときには、平均 X' は近似的に正規分布 N（（a），（b））に従う。

（2）n=25 を十分大きいと見なしたとき、μ は確率95%で（c）の範囲に収まる。この範囲のことを μ に対する（d）という。

（3）（2）で、「確率95%」を80%に下げると μ が収まるべき範囲は（e）となる。また「確率95%」はそのままで、μ が収まるべき範囲の幅を半分以下にしたければ、標本を最低あと（f）個追加する必要がある。

［注意］母分散既知なので「（標本の）分散 $s^2=729$」は使う必要なし。