- 数理統計学（火曜日３講時）
- 薬学部・農学部向け
（理学部・工学部からの受講不可）
- 確率と統計
- 担当者：尾畑伸明（情報科学研究科）

この授業について

目的

データ解析で必須である確率と統計を学び、
今後の実務や研究のためのリテラシーを養う。

教科書

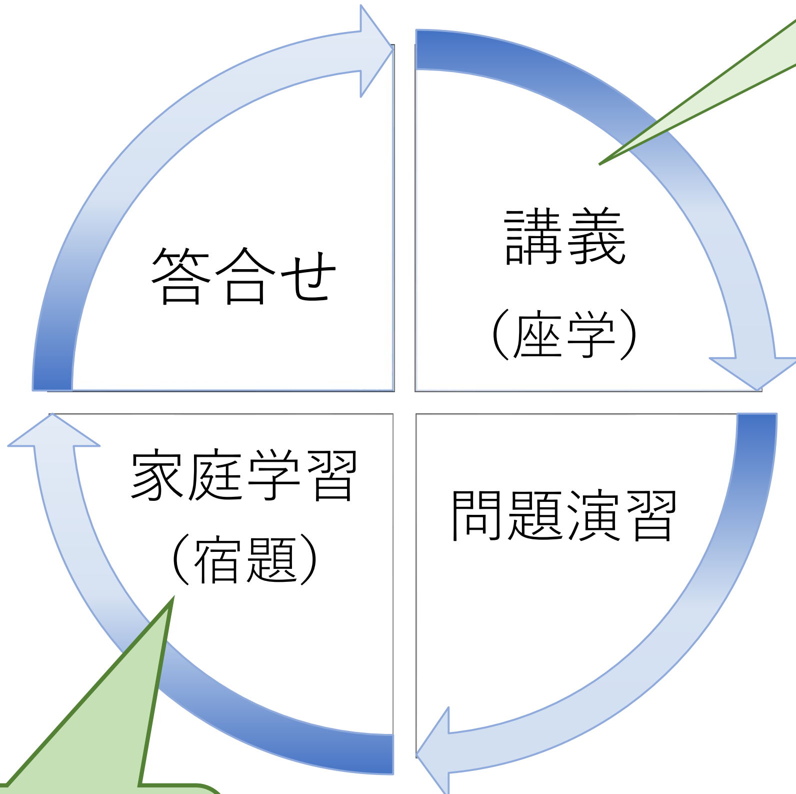
稲垣他：統計学講義, 裳華房

計算機

- 四則演算、平方根が必須
- 関数電卓・PC・統計ソフトは使わない

授業の進め方

教科書に準拠
追加が少しあり



予習 30分
宿題・復習 60分

パワポ資料

- 印刷配布しない
- 担当者のウェブサイトまたはISTUからダウンロード可能
- ただし、はじめの２回分と演習問題は配布する

www.math.is.tohoku.ac.jp/~obata

パワポ資料のほか、過年度の資料・試験問題、講義のビデオ（実験的に公開）は担当者のウェブサイトからダウンロード可能

ミニットペーパー

- 出席チェック（時には、家庭学習のチェックも）
- 質問を書く

  ⇨ 次回の授業時に回答またはウェブで回答
- 気の利いたことを書くと加点される
- 代筆は厳しく減点される

評価

- 期末試験 50% + 中間試験 30% + 平常点 20%
- 日程（暫定的）

  中間試験　　５月２８日（火）

  期末試験　　７月２３日（火）

参考書

ー本当は、教科書だけでは少し不満足

- **同程度の標準的なテキスト**

[1] 吾妻一興他：「概説 数理統計」共立出版, 1994.

[2] 宮川公男：「基本統計学 第4版」有斐閣, 2015.

[3] P.G.ホーエル（浅井・村上訳）：「初等統計学」培風館, 1981

[4] （演習書）白砂堤津耶：「例題で学ぶ初歩からの統計学」日本評論社, 2015.

[5]（少し易しいが） 鈴木義一郎：「初めて学ぶ基本統計学」森北出版, 2005.

- **本格的なテキスト**

[1] P.G.ホーエル（浅井・村上訳）：「入門数理統計学」培風館, 1978.

[2] 尾畑伸明：「数理統計学の基礎」共立出版, 2014.

[4] 打波守：「医・薬系のための統計入門」培風館, 2004.

[5] 丹後俊郎：「医学への統計学」朝倉書店, 2013.

[6] 階堂武郎：「医系の統計入門 (第2版)」森北出版, 2013.

ビデオ

- 授業はビデオ録画される（録画の再利用を検討中）

シラバスの授業予定と教科書の対応表

Lecture 番号と題目	教科書	備考
1　確率の基本的性質	第 1章 統計学と確率 1 統計学とは何か 1.1 記述統計 1.2 推測統計 2　確率とは何か 2.1 確率の定義 2.2 事象と確率	
2　条件付き確率とベイズの公式	2.2 事象と確率 2.3 ベイズの定理	
3　データの整理	第 2 章 データ処理 1 度数分布表とヒストグラム 2 データの特性値 2.1 データの位置 2.2 データの拡がり 3 データの変換 3.1 標準得点 3.2 簡便的計算法 4 2 次元データの整理 4.1 散布図と相関表 4.2 共分散と相関係数	

4　確率変数と確率分布	第 3 章　確率変数と確率分布 1 確率変数 2 平均と分散 3 離散型確率変数の分布 　3.1　2 項分布 　3.2　ポアソン分布	
5　期待値（平均値）・分散・大数の法則	4 連続型確率変数の分布 　4.1　指数分布 　4.2　正規分布	
	第 4 章 多変量確率変数 1　2 次元確率ベクトル 2 多変量確率変数	第 4 章は省略
6　正規分布と中心極限定理	第 5 章 母集団と標本 1 標本抽出 2 標本平均と標本分散 3 正規分布から導かれる標本分布 　3.1　カイ 2 乗分布 　3.2　ティー分布 　3.3　エフ分布 4 大数の法則と中心極限定理	3節は後回し
中間まとめと中間評価		

7　統計的推定とは	第 6 章推定 1 推定量とその性質	
8　正規分布にまつわる分布 （χ^2-分布、t-分布、F-分布）	第 5 章 母集団と標本 3 正規分布から導かれる標本分布 3.1　カイ 2 乗分布 3.2　ティー分布 3.3 エフ分布	
9　母比率・母平均の推定	第 6 章推定 2 平均の区間推定 2.1 母分散 σ^2 が既知の場合 2.2 母分散 σ^2 が未知の場合 3 分散の区間推定 4 比率の推定	3節は省略
10　仮説検定とは	第 7 章検定 1 検定の手順	
11　母比率・母平均の検定	2 平均の検定 2.1 母分散 σ^2 が既知の場合 2.2 母分散 σ^2 が未知の場合 3 分散の検定 4 比率の検定	3節は省略

12　母集団の比較	第 8 章 2 標本問題 1 平均の差について 　1.1 平均差の区間推定 　1.2 平均差の検定 2 分散比について 　2.1 分散比の区間推定 　2.2 等分散の検定 3 等分散性がない場合 　3.1 母分散 σ_1^2, σ_2^2 が既知の場合 　3.2 母分散 σ_1^2, σ_2^2 は未知であるが，標本数 n_1, n_2 が大きい場合 4 比率の 2 標本問題 　4.1 比率の差の区間推定 　4.2 比率の差の検定	2節と3節は省略
13　適合度検定・独立性検定	カイ2乗検定入門（記述なし）	資料配布
まとめと期末試験		

※ 1-2 回の休講が見込まれる

Lecture 1

確率の基本的性質

【教科書】

第１章 統計学と確率

1. 統計学とは何か
2. 確率とは何か（途中まで）

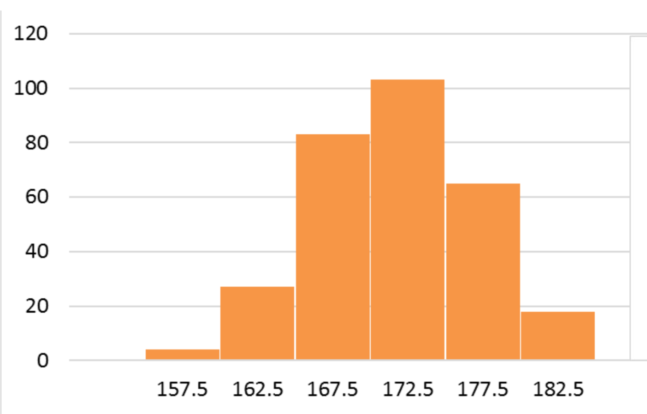
• 記述統計

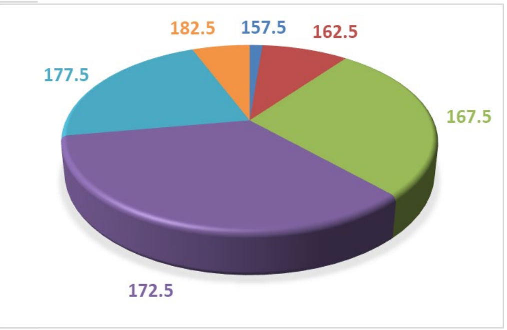
調査対象 → 数値データ → 表やグラフ

身長調べ

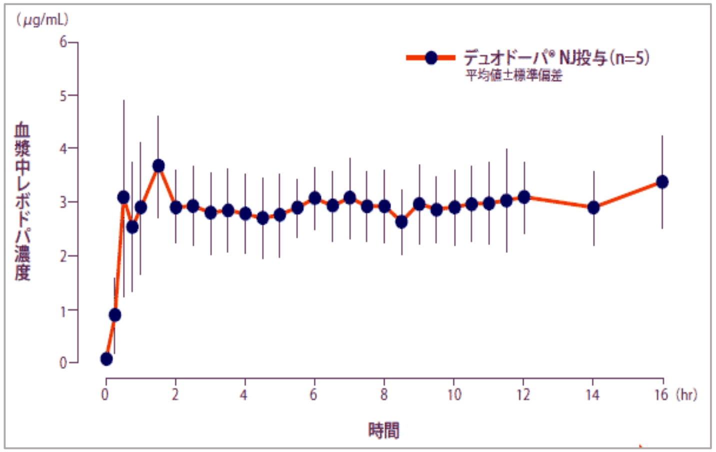
階級	155-160	160-165	165-170	170-175	175-180	180-185	合計
階級値	157.5	162.5	167.5	172.5	177.5	182.5	
度数	4	27	83	103	65	18	300
相対度数	0.013	0.090	0.277	0.343	0.217	0.060	1.000

- 国勢調査
- 世論調査
- 視聴率調査
- 様々な実験観察

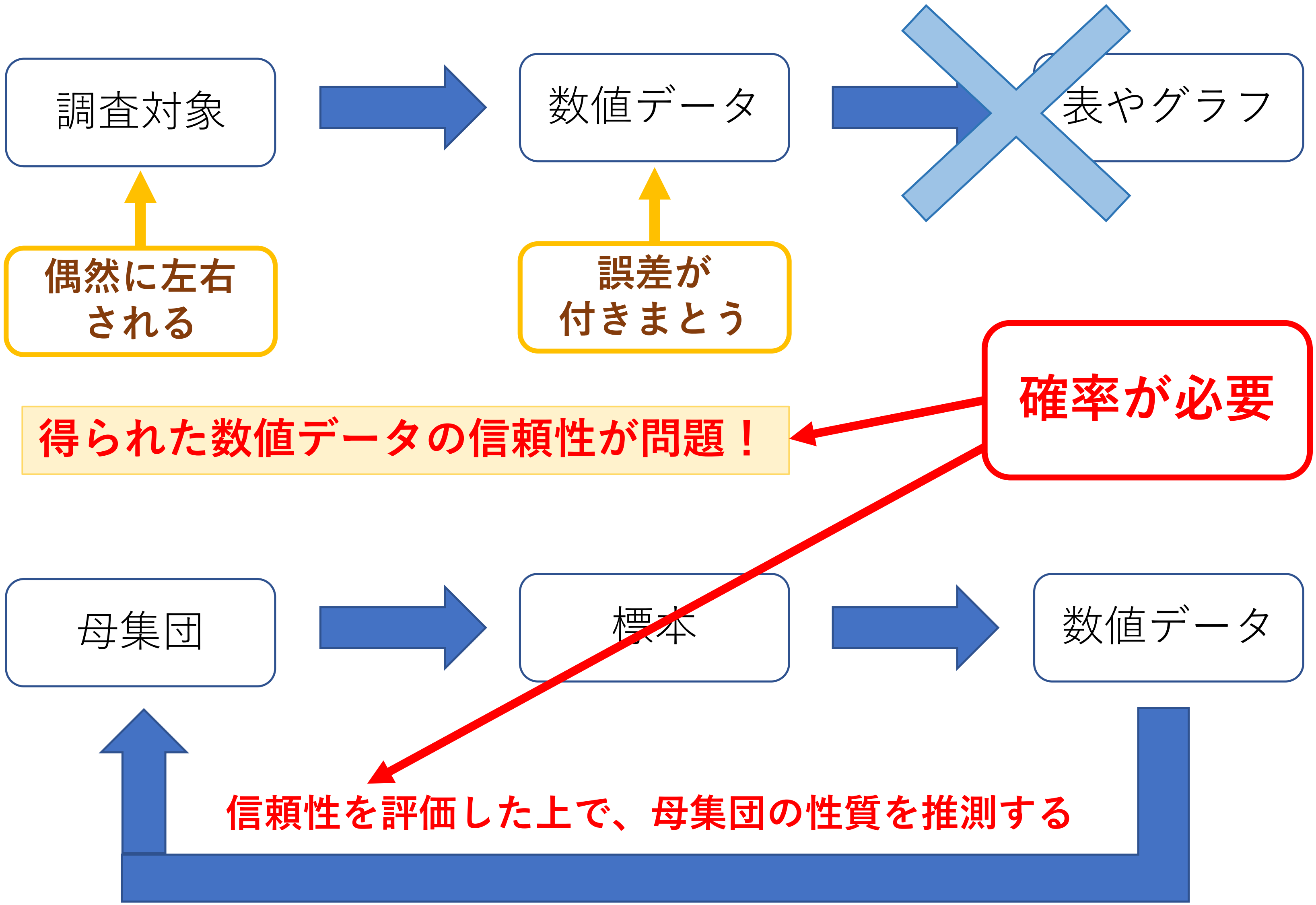




時間変化（時系列）



- 推測統計


調査対象
数値データ
表やグラフ
偶然に左右される
誤差が付きまとう
確率が必要
得られた数値データの信頼性が問題！
母集団
標本
数値データ
信頼性を評価した上で、母集団の性質を推測する

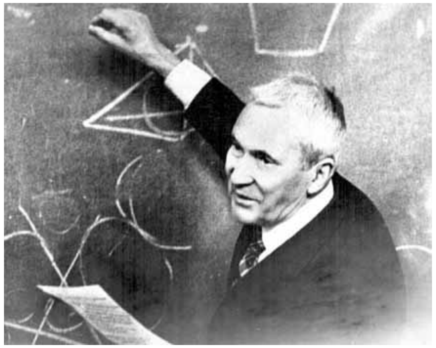
- 確率とは，　事柄に対する確からしさの程度を示す指標であって，　0 (= 0%) から 1 (= 100%) までの数値で表現される．
- 取り扱う対象はその生起が決定論的に確定しえない事象

$P(A)$　事象 A の起こる確率（Probability）

カルダーノ、フェルマ、パスカル、ベルヌーイ、ラプラス

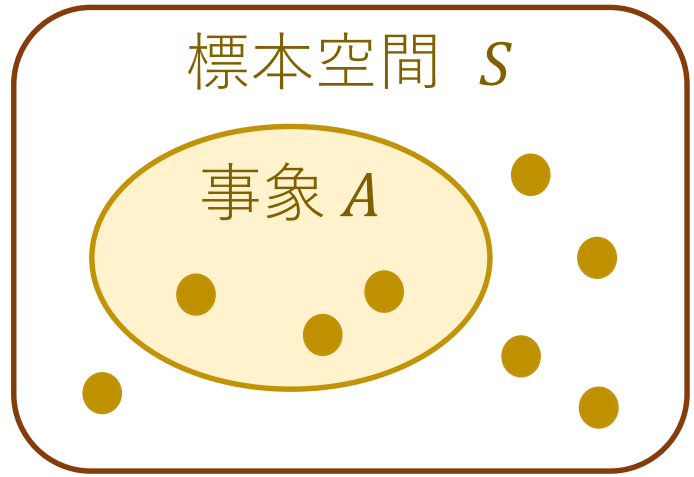
➡ コルモゴロフの公理論的確率

- 数学的に厳密な理論として出発する．
- 現代統計学の基礎になる．
- 確率モデルとして, 自然科学・生命科学・社会科学などあらゆる分野に波及している．


A. N. Kolmogorov (1903-1987)

S ：**全事象または標本空間**

- 根元事象を集めた集合
- 記号として　Ω（オメガ）も使われる



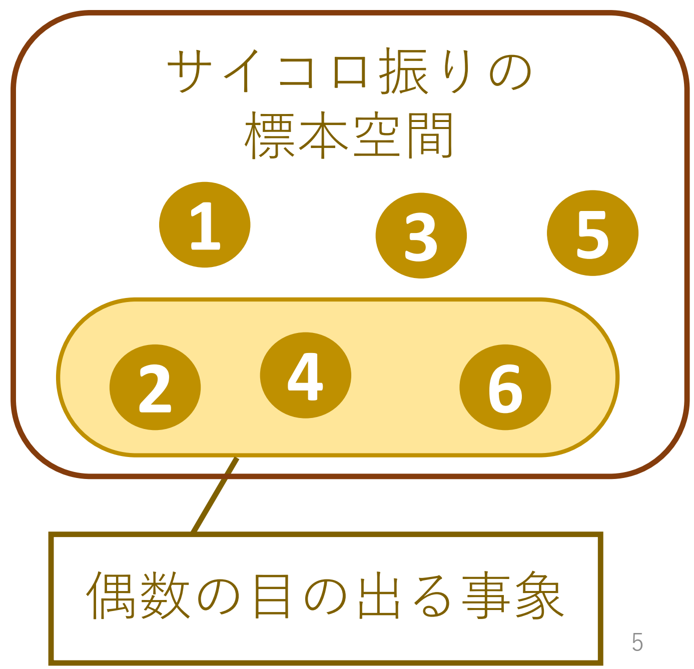
A ：**事象**

- 有限または無限個の根元事象からなり, 確率を与える対象となる
- 標本空間の部分集合

$P(A)$ ：**事象 A の起こる確率**

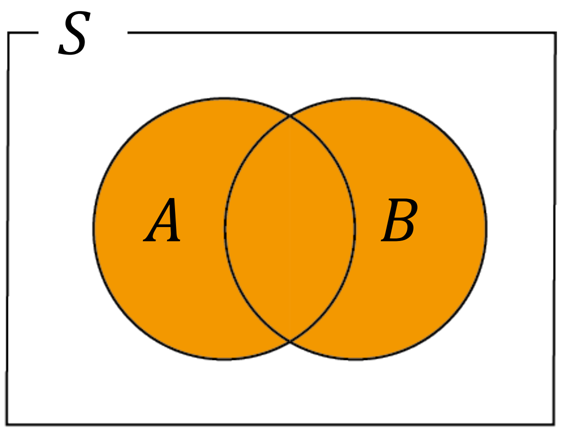
$\mathcal{A}$ ：**事象族**

- 扱う事象を限定する（数学的理由）

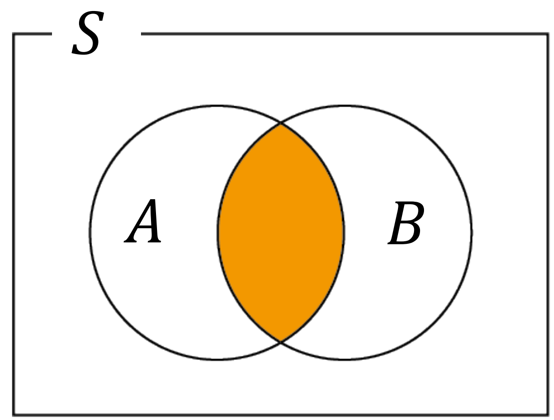


$(S, \mathcal{A}, P)$ ：確率空間

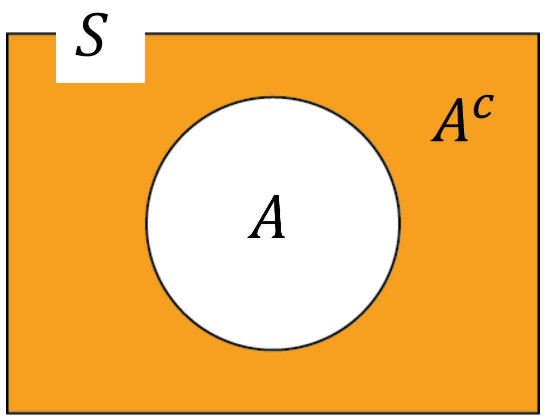
和事象　$A \cup B$



積事象　$A \cap B$



余事象　A^c



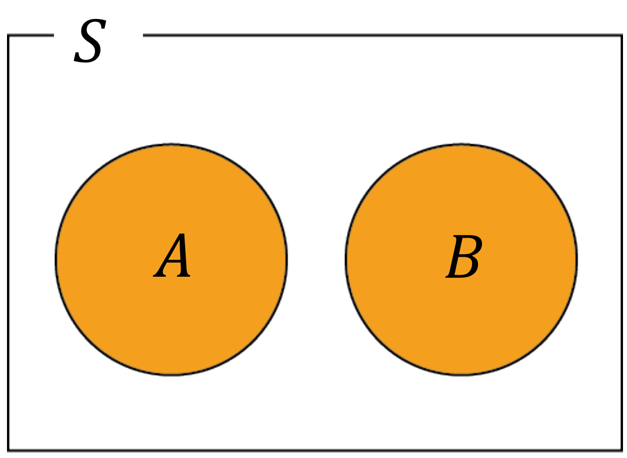
$\emptyset$: 空事象（起こりえない）

$$P(\emptyset) = 0$$

S : 全事象（必ず起こる事象）

$$P(S) = 1$$

排反事象　$A \cap B = \emptyset$



公理

(0) 事象族 $\mathcal{A}$ は $\boxed{\text{可算加法族}}$ をなす.

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

裏に高度な数学あり

(2) $P(\emptyset) = 0,\ P(S) = 1$

(3) $A_1, A_2, \cdots, A_n, \cdots$ が互いに排反であれば,

$$P(A_1 \cup A_2 \cup \cdots \cup A_n \cup \cdots) = P(A_1) + P(A_2) + \cdots + P(A_n) + \cdots$$

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n)$$

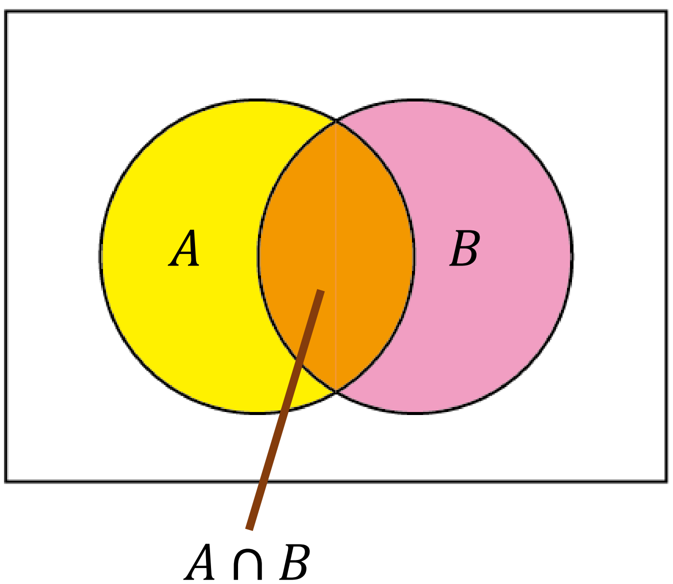
定理 1. 1

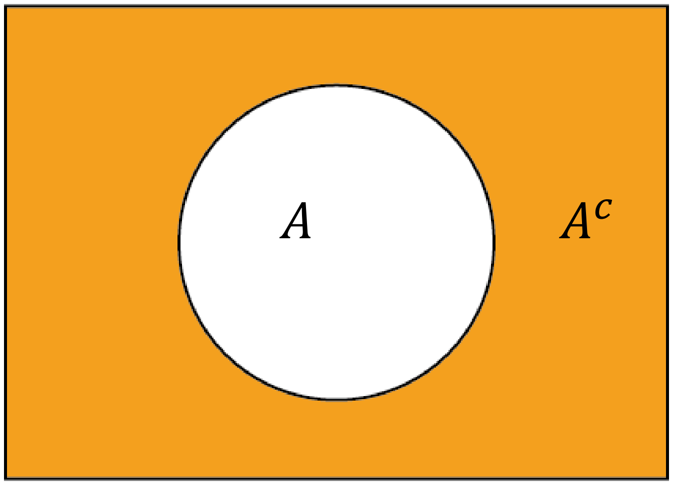
事象 A, B があるとき，A または B の事象の起こる確率 について

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

事象 A の余事象が起こる確率 について

$$P(A^c) = 1 - P(A)$$





例題 1. 1

2つのサイコロを投げ，出た目の差を D とする.
D が $0, 1, 2, 3, 4, 5$ の場合についてその確率を求めよ.

注意　（１）D はある一定の範囲を変化する（つまり変数）
（２）それぞれの値に対してその値になる確率が考えられる
（３）このような変数を**確率変数**という

考え方

（１）起こりうるすべての場合を考える

サイコロの目は 6 通りの出方があり，　2 つのサイコロでは 36 通りの出方がある

（２）記号を準備する

2 つのサイコロの目 を X ,Y とする. 試行の結果は (X, Y) となる.
したがって, その差は $D = |X - Y|$ となる.

（３）起こりうるそれぞれの場合の確率を考える

起こりうるすべての場合を書き上げる

X＼Y	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

根元事象

確率の決め方：公平なサイコロの場合は各根元事象は等確率

確率の決め方

公平なサイコロの場合は
各根元事象は等確率

X \ Y	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	$\boxed{(1,4)}$	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	$\boxed{(4,6)}$
5	(5,1)	(5,2)	(5,3)	(5,4)	$\boxed{(5,5)}$	(5,6)
6	(6,1)	(6,2)	(6,3)	$\boxed{(6,4)}$	(6,5)	(6,6)

確率計算の例

(1) $\mathrm{P}(X = 1, Y = 4)$

$$\mathrm{P}\big((X, Y) = (1,4)\big) = \frac{1}{36}$$

(2) $\mathrm{P}(X + Y = 10)$

- 等確率の根元事象３個分

$$= 3 \times \frac{1}{36} = \frac{3}{36}$$

- 根元事象へ分解する

$$\begin{aligned}
&\mathrm{P}(X + Y = 10)\\
&\quad = \mathrm{P}\big((X, Y) = (4,6)\big)\\
&\qquad + \mathrm{P}\big((X, Y) = (5,5)\big)\\
&\qquad + \mathrm{P}\big((X, Y) = (6,4)\big)
\end{aligned}$$

$$= \frac{1}{36} + \frac{1}{36} + \frac{1}{36} = \frac{3}{36}$$

例題1.1 の解

Y / X	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- 2個のサイコロの目の差

$$D = |X - Y|$$

- Dの取りうる値は

$$0, 1, 2, 3, 4, 5$$

- まず,

$$\begin{aligned} \mathrm{P}(D = 0) &= \mathrm{P}\big((X, Y) = (1,1)\big) \\ &\quad + \mathrm{P}\big((X, Y) = (2,2)\big) \\ &\quad + \cdots \\ &\quad + \mathrm{P}\big((X, Y) = (6,6)\big) \\ &= 6 \times \frac{1}{36} = \frac{6}{36} \end{aligned}$$

d	0	1	2	3	4	5
$P(D = d)$						

例題1.1 の解

Y \ X	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- 2個のサイコロの目の差

$$D = |X - Y|$$

- Dの取りうる値は

$$0, 1, 2, 3, 4, 5$$

- まず,

$$\begin{aligned} \mathrm{P}(D=0) &= \mathrm{P}\big((X,Y)=(1,1)\big) \\ &\quad + \mathrm{P}\big((X,Y)=(2,2)\big) \\ &\quad + \cdots \\ &\quad + \mathrm{P}\big((X,Y)=(6,6)\big) \\ &= 6 \times \frac{1}{36} = \frac{6}{36} \end{aligned}$$

d	0	1	2	3	4	5
$P(D=d)$	$\frac{6}{36}$					

例題1.1 の解

X \ Y	1	2	3	4	5	6
1	(1,1)	$\boxed{(1,2)}$	(1,3)	(1,4)	(1,5)	(1,6)
2	$\boxed{(2,1)}$	(2,2)	$\boxed{(2,3)}$	(2,4)	(2,5)	(2,6)
3	(3,1)	$\boxed{(3,2)}$	(3,3)	$\boxed{(3,4)}$	(3,5)	(3,6)
4	(4,1)	(4,2)	$\boxed{(4,3)}$	(4,4)	$\boxed{(4,5)}$	(4,6)
5	(5,1)	(5,2)	(5,3)	$\boxed{(5,4)}$	(5,5)	$\boxed{(5,6)}$
6	(6,1)	(6,2)	(6,3)	(6,4)	$\boxed{(6,5)}$	(6,6)

- 2個のサイコロの目の差

$$D = |X - Y|$$

- 次に,

$$\begin{aligned}\mathrm{P}(D = 1) &= \mathrm{P}\big((X, Y) = (1,2)\big) \\ &\quad + \mathrm{P}\big((X, Y) = (2,1)\big) \\ &\quad + \cdots \\ &\quad + \mathrm{P}\big((X, Y) = (6,5)\big) \\ &= 10 \times \frac{1}{36} = \frac{10}{36}\end{aligned}$$

d	0	1	2	3	4	5
$P(D = d)$	$\frac{6}{36}$					

例題1.1 の解

Y \ X	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- 2個のサイコロの目の差

$$D = |X - Y|$$

- 次に,

$$\begin{aligned}\mathrm{P}(D=1) &= \mathrm{P}\big((X,Y)=(1,2)\big)\\ &\quad + \mathrm{P}\big((X,Y)=(2,1)\big)\\ &\quad + \cdots\\ &\quad + \mathrm{P}\big((X,Y)=(6,5)\big)\\ &= 10 \times \frac{1}{36} = \frac{10}{36}\end{aligned}$$

d	0	1	2	3	4	5
$P(D=d)$	$\frac{6}{36}$	$\frac{10}{36}$				

例題1.1 の解

X \ Y	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)

- 2個のサイコロの目の差

$$D = |X - Y|$$

- Dの取りうる値は

$$0, 1, 2, 3, 4, 5$$

- 左の表で根元事象の個数を数えて

$$P(D = d)$$

を計算する

この表を D の確率分布という

d	0	1	2	3	4	5	合計
$P(D = d)$	$\frac{6}{36}$	$\frac{10}{36}$	$\frac{8}{36}$	$\frac{6}{36}$	$\frac{4}{36}$	$\frac{2}{36}$	1

宿題

演習問題 1 (page 13-14)

1.1　ロクロを半径 1 の円周上の一様な確率空間と考え，長さ s の弧には $\dfrac{s}{2\pi}$ の確率を与える．この円周を 25 等分して各区域に番号 1, 2, . . . , 25 を付ける．ロクロを回して偶数番号の区域が自分の前で止まる確率を計算せよ．

1.2　袋の中に同じ大きさの球が，赤 6，白 5，青 4 個入っている．ランダムに 2 個取り出すとき，それがともに赤である確率と，赤と白である確率を求めよ．

1.3　ジョーカーを除いた一組のトランプから 4 枚のカードを抜き取ったとき，スペードとハートのカードが 1 枚ずつ含まれている確率を求めよ．

1.4　xy 平面上の 4 点 $(0, 0)$, $(1, 0)$, $(1, 1)$, $(0, 1)$ を頂点とする正方形の上にランダムに 1 点をとる.

(1)　その点が $x = 0$, $y = 0$, $x + y = 1$ で囲まれる三角形に入る確率を計算せよ.

(2)　その点が $x = 0$, $y = 0$, $x + y = 1$ で囲まれる三角形にあることがわかっているとき, それがまた $y = 0$, $x = 1$, $x = y$ の三角形の中にある確率を計算せよ.

(3)　その点が $y = 0$, $y = 1$, $x = 0$, $x = \dfrac{1}{2}$ で囲まれる長方形にあることがわかっているとき, それがまた $y = 0$, $x = \dfrac{1}{2}$, $x + y = 1$ の 台形 の中にある確率を計算せよ.　$x = 0,$

1.5　半径 1 の円板上にランダムに 1 点をとる. その点が, 中心角が 0 から $\dfrac{\pi}{4}$ ラジアンまでの扇形部分に入る確率を求めよ.

Lecture 2

条件付き確率

【教科書】

第１章 統計学と確率

2. 確率とは何か（途中から）

• 独立性（定義1.2）

事象 A, B が独立である とは, 事象 A, B について

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

が成り立つことをいう．独立でないとき従属であるという.

【例】サイコロを２回投げる.

X ： １回目に出る目, Y ： ２回目の出る目

・事象 $A = \{X = 2\}$ と $B = \{Y = 5\}$ は独立である

$$P(A \cap B) = \frac{1}{36} \qquad P(A)P(B) = \frac{1}{6} \times \frac{1}{6}$$

ゆえに $P(A \cap B) = P(A)P(B)$ が成り立ち, A と B は独立.

例題 1. **

** 教科書に掲載されていない

52 枚のトランプから 1 枚をランダムに抜き取るとき，そのカードのスートを X, 数字を Y とする．X と Y は独立である．

考え方

（1）起こりうるすべての場合を考える

52枚のカードのうちの 1 枚なので 52 通りの起こり方がある

（2）記号を用いて適切に表示する

（3）起こりうるそれぞれの場合の確率を考える

起こりうるすべての場合をわかりやすく書く

♥ A	♥ 2									♥ J	♥ Q	♥ K
♠ A	♠ 2									♠ J	♠ Q	♠ K
♦ A	♦ 2									♦ J	♦ Q	♦ K
♣ A	♣ 2									♣ J	♣ Q	♣ K

X \ Y	1	2	3	4	5	6	7	8	9	10	11	12	13
1	♥ A	♥ 2									♥ J	♥ Q	♥ K
2	♠ A	♠ 2					♠ 7				♠ J	♠ Q	♠ K
3	♦ A	♦ 2									♦ J	♦ Q	♦ K
4	♣ A	♣ 2									♣ J	♣ Q	♣ K

確率計算の例　「スペードの７」を引く確率

X \ Y	1	2	3	4	5	6	7	8	9	10	11	12	13
1	♥ A	♥ 2									♥ J	♥ Q	♥ K
2	♠ A	♠ 2					♠ 7				♠ J	♠ Q	♠ K
3	♦ A	♦ 2									♦ J	♦ Q	♦ K
4	♣ A	♣ 2									♣ J	♣ Q	♣ K

確率計算の例　「スペードの７」を引く確率

$$P(\spadesuit 7) = P(X = 2, Y = 7) = \frac{1}{52}$$

考 察

$$P(\spadesuit) = P(X = 2) = \frac{13}{52} = \frac{1}{4}$$

$$P(7) = P(Y = 7) = \frac{4}{52} = \frac{1}{13}$$

ゆえに,

$$P(\spadesuit 7) = P(\spadesuit)P(7)$$

X \ Y	1	2	3	4	5	6	7	8	9	10	11	12	13
1	♥ A	♥ 2									♥ J	♥ Q	♥ K
2	♠ A	♠ 2					♠ 7				♠ J	♠ Q	♠ K
3	♦ A	♦ 2									♦ J	♦ Q	♦ K
4	♣ A	♣ 2									♣ J	♣ Q	♣ K

数学記号で表すのがよい

$$P(X=2, Y=7) = \frac{1}{52} \qquad P(X=2) = \frac{13}{52} = \frac{1}{4} \qquad P(Y=7) = \frac{4}{52} = \frac{1}{13}$$

$$P(X=2, Y=7) = P(X=2)\,P(Y=7)$$

Y X	1	2	3	4	5	6	7	8	9	10	11	12	13
1	♥ A	♥ 2									♥ J	♥ Q	♥ K
2	♠ A	♠ 2					$\boldsymbol{a}$ $\boldsymbol{b}$				♠ J	♠ Q	♠ K
3	♦ A	♦ 2									♦ J	♦ Q	♦ K
4	♣ A	♣ 2									♣ J	♣ Q	♣ K

数学記号で表すのがよい

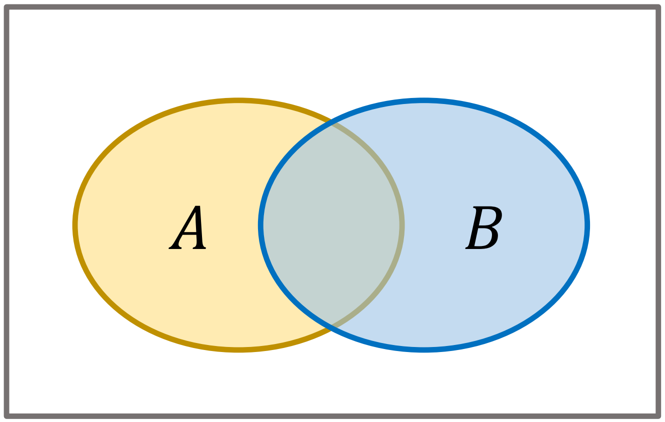
$$P(X=\boldsymbol{a},Y=\boldsymbol{b})=\frac{1}{52}\qquad P(X=\boldsymbol{a})=\frac{13}{52}=\frac{1}{4}\qquad P(Y=\boldsymbol{b})=\frac{4}{52}=\frac{1}{13}$$

$$P(X=\boldsymbol{a},Y=\boldsymbol{b})=P(X=\boldsymbol{a})\,P(Y=\boldsymbol{b})$$

確率変数 X,Y は独立である

- 条件付き確率

A, B : 事象



A の下での B の条件付確率

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

ただし，$P(A) > 0$

解釈：事象 A が起こったことを知った上で，事象 B の起こる確率

定理 1. 3 （乗法定理）

2 つの事象 A, B がともに起こる確率 $P\,(A \cap B)$ は, 条件付き確率を使って, 次のように表される.

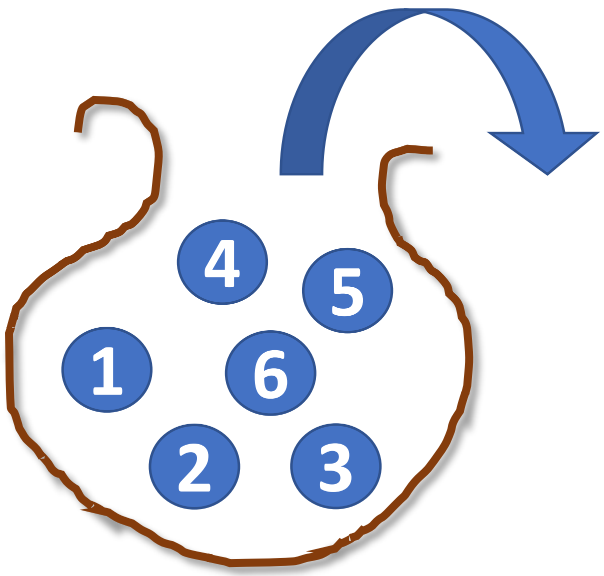
$$P(A \cap B) = P(A)P(B|A) = P(B)P(A|B)$$

例題 1. 2

袋の中に同じ形状の球 6 個が入っていて, それぞれに 1, 2, 3, 4, 5, 6 の数字が書いてあるとする. そのとき, 袋の中から次のような 2 通りの方法で 2 個の球を取り出すことを考える.

(1) 袋の中から球を 1 個取り出す. 1 回目の球の数字が i $(i = 1, \ldots, 6)$ である事象を A_i とする. 次に, 取り出した球を袋に戻し, よくかきまぜ, 改めて袋の中から球を 1 個取り出すとき(復元抽出), 球の数字が j $(j = 1, \ldots, 6)$である事象を B_j とする. そのとき取り出した 2 個の球の数字が i, j である確率はいくらか

(2) 1 回目に取り出した球をもとに戻さず, 続けて球を取り出すとき(非復元抽出), 取り出した 2 個の球の数字が i, j である確率はいくらか.



（１）6個の球は等確率で選ばれるから

$$P(A_i) = \frac{1}{6}$$

2回目の試行では, 袋の中や取り出し方は1 回目と同じである. したがって,

$$P(B_j) = \frac{1}{6}$$

1回目と２回目の試行は独立であるから

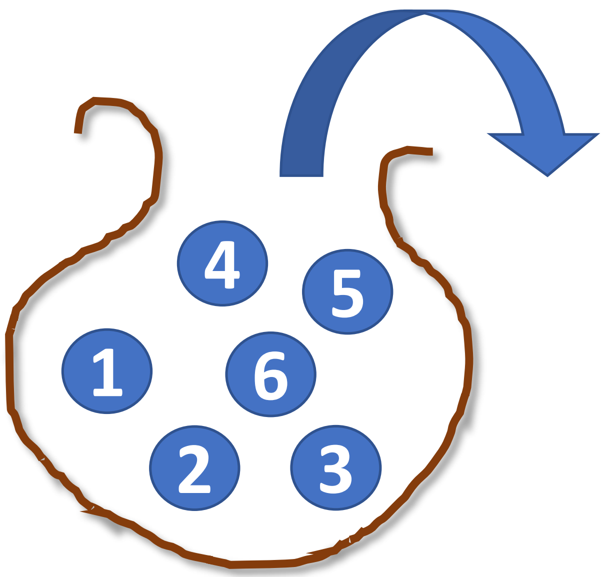
$$P(A_i \cap B_j) = P(A_i)P(B_j) = \frac{1}{36}$$

別解

１回目が２回目に影響しないので　$P(B_j) = P(B_j|A_i)$

乗法定理によって

$$P(A_i \cap B_j) = P(A_i)P(B_j|A_i) = P(A_i)P(B_j)$$



（２）乗法定理によって

$$P(A_i \cap B_j) = P(A_i)P(B_j|A_i)$$

題意から

$$P(B_j|A_i) = \begin{cases} 0, & i = j \\ \dfrac{1}{5}, & i \neq j \end{cases}$$

$i = j$ のとき,

$$P(A_i \cap B_j) = P(A_i)P(B_j|A_i) = 0$$

$i \neq j$ のとき,

$$P(A_i \cap B_j) = P(A_i)P(B_j|A_i) = \frac{1}{6} \times \frac{1}{5} = \frac{1}{30}$$

注意　当然、A_i と B_j は独立ではない.

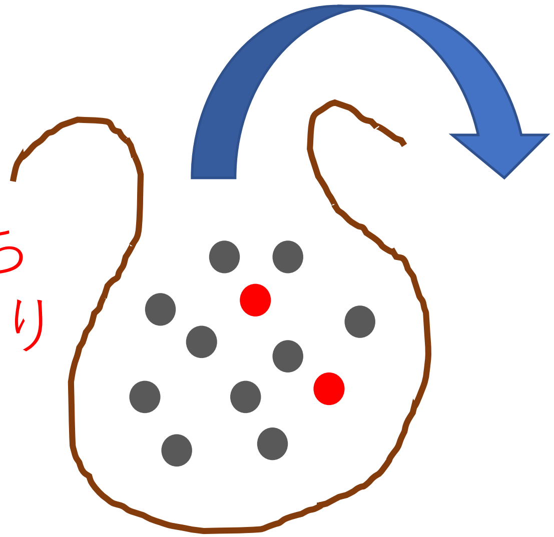
なぜなら $P(B_j) = \dfrac{1}{6}$ であるから（確認せよ）

例題 1. 3

20 本のうち，5 本が当たりであるクジがあり，A, B の 2 人が 1 本ずつクジを引くものとする.

（1）A が引いたクジをもとに 戻し, よくかきまぜて次に B が引 くとき(復元抽出)，2 人とも当たりである確率はいくらか.

（2）A が引いたクジをもとに戻さず，次に B が引くとき(非復元抽出), 2人とも当たりである確率はいくらか



- **まず記号を準備する**

A : A があたりを引く事象

B : B があたりを引く事象

A, B の 2 人ともあたりを引く事象

$$A \cap B$$

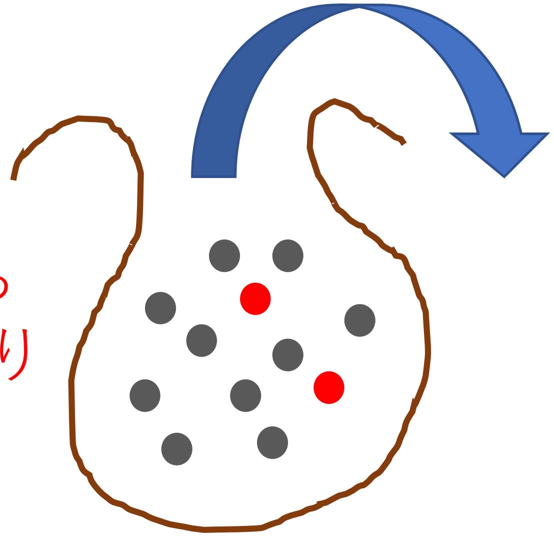
（1）A が引いたクジをもとに 戻し, よくかきまぜて次に B が引くとき(復元抽出), 2 人とも当たりである確率はいくらか.

$$P(A) = \frac{5}{20} \qquad P(B) = \frac{5}{20}$$

A と B は独立なので,

$$P(A \cap B) = P(A)P(B) = \frac{5}{20} \times \frac{5}{20} = \frac{25}{400} = \frac{1}{16}$$

20 本のうち
5 本が当たり



A：A があたりを引く事象

B：B があたりを引く事象

A, B の 2 人ともあたりを引く事象

$$A \cap B$$

（2）A が引いたクジをもとに戻さず，次に B が引くとき(非復元抽出),
2人とも当たりである確率はいくらか

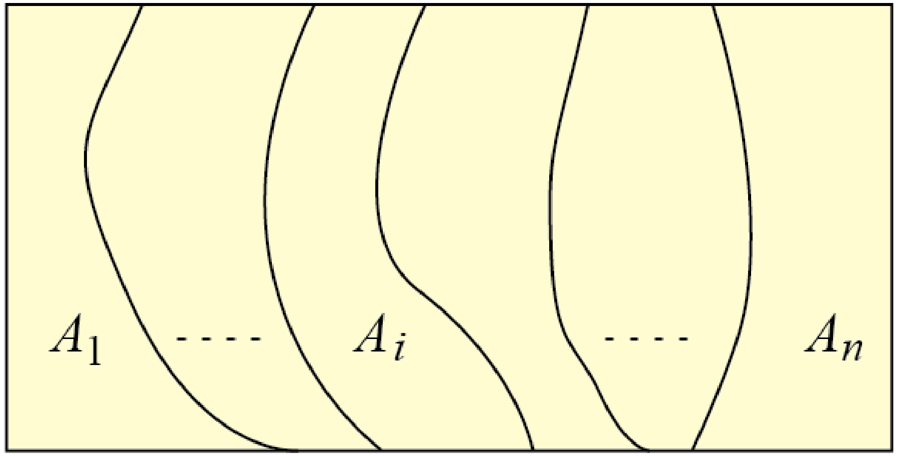
$$P(A) = \frac{5}{20} \qquad P(B|A) = \frac{4}{19}$$

乗法定理を用いて,

$$P(A \cap B) = P(A)P(B|A) = \frac{5}{20} \times \frac{4}{19} = \frac{20}{380} = \frac{1}{19}$$

• ベイズの定理

全事象 S の層別 (stratification)



$$A_1 \cup A_2 \cup \cdots \cup A_n = S$$
$$P(A_1) + P(A_2) + \cdots + P(A_n) = 1$$

例

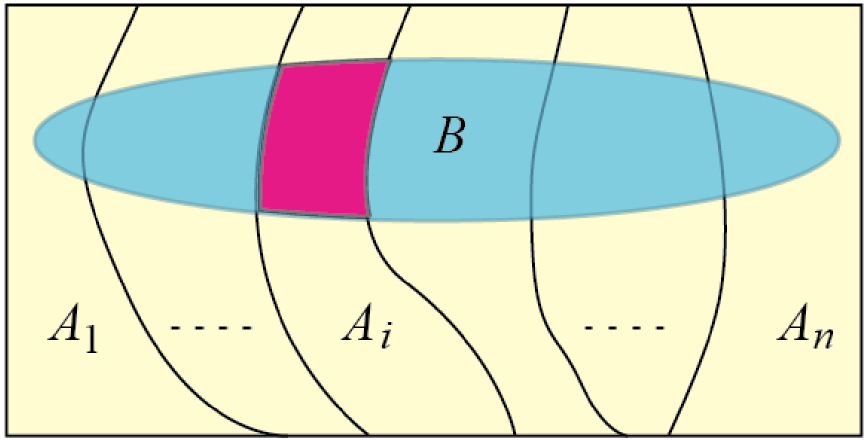
全人口を「10 代以下, 20 代, ..., 60 代, 70 代以上」と年代別に考える.

定理 1.4（全確率の公式）

全事象が A_i $(i = 1, \cdots, n)$ に層別されているとき, 事象 B の確率 $P(B)$ は, A_i の事前確率 $P(A_i)$ と条件付き確率 $P(B|A_i)$ を使って表される.

$$\boxed{P(B) = P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

$$\boxed{P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)}$$



$$B = \bigcup_{i=1}^{n} A_i \cap B$$ **（互いに排反）**

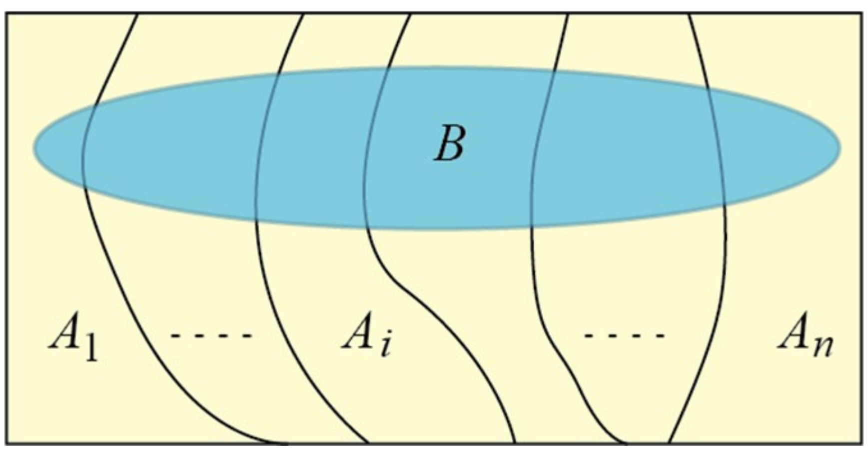
$$\therefore \quad P(B) = \sum_{i=1}^{n} P(A_i \cap B)$$

一方, $P(A_i \cap B) = P(A_i)P(B|A_i)$

定理 1. 5 （ベイズの定理）

全事象が A_i $(i = 1, \cdots, n)$ に層別されているとき, 事象 B が起こったときの各層の条件付き確率 $P(A_i|B)$ を事後確率といい, 事前確率 $P(A_i)$ と条件付き確率 $P(B|A_i)$ を使って表される.

$$\boxed{P(A_i|B) = \frac{P(A_i)\,P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}}$$



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

$$分子 = P(A_i \cap B) = P(A_i)\,P(B|A_i)$$

$$分母 = P(B) = \sum_{i=1}^{n} P(A_i)P(B|A_i)$$

これは全確率の公式

$$P(A_i|B) = \frac{P(A_i)\,P(B|A_i)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + \cdots + P(A_n)P(B|A_n)}$$

結果から原因を知る公式

原因 → 結果

$P(B|A_1)$

A_1 ：バスが遅れた

A_2 ：朝寝坊した

⋮

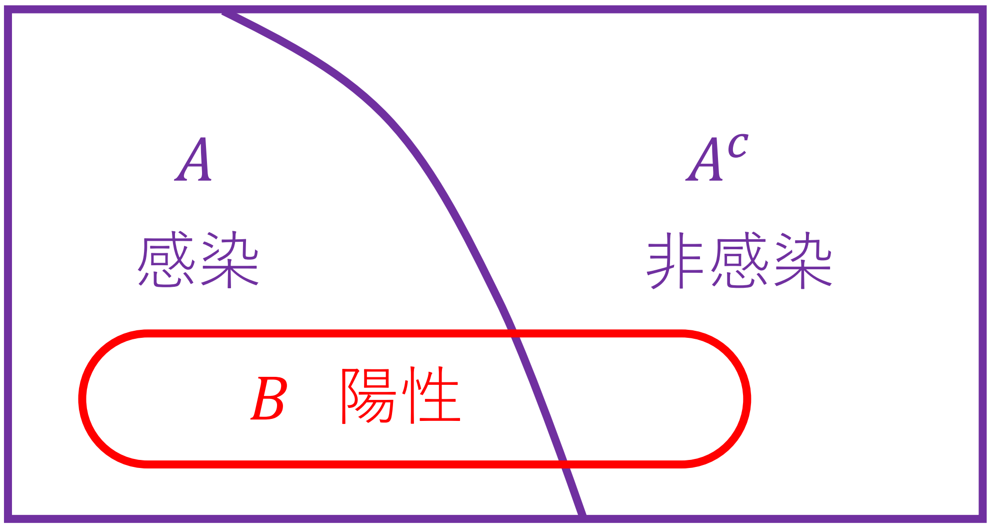
A_n ：時計が止まっていた

B : 講義に遅刻した

例題 1. ** （擬陽性の問題）

- 病気 A の感染者は 500 人に 2 人の割合であるという.
- 検査 B は, 感染者の 95% に陽性反応を示すが,
  非感染者の 2% にも陽性反応が出てしまう.

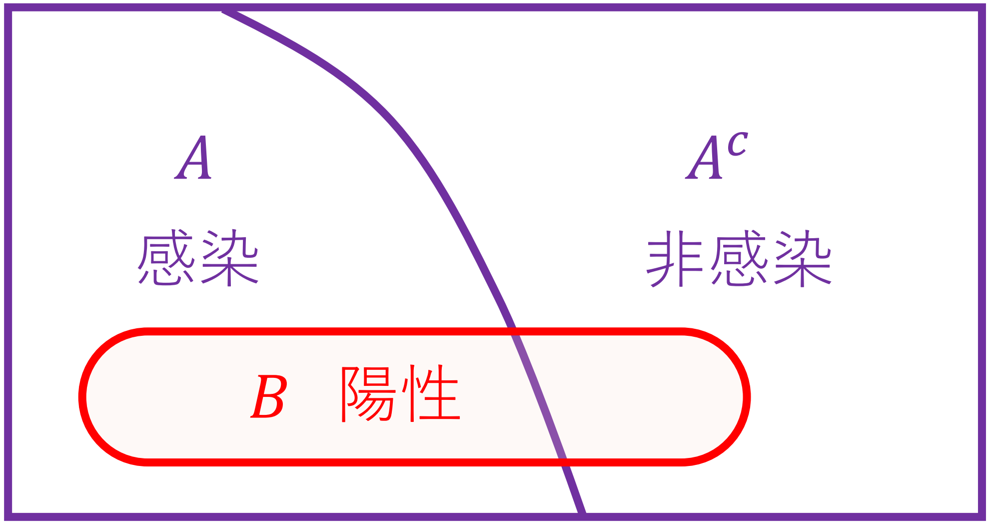
(1) 陽性反応が出れば感染しているか？
(2) 陰性反応なら非感染か？



$$P(A) = \frac{2}{500}$$

$$P(B|A) = 0.95$$

$$P(B|A^c) = 0.02$$



$$P(A) = \frac{2}{500}$$

$$P(B|A) = 0.95$$

$$P(B|A^c) = 0.02$$

(1) 陽性反応が出れば感染しているか？

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$

$$= \frac{\frac{2}{500} \times 0.95}{\frac{2}{500} \times 0.95 + \frac{498}{500} \times 0.02}$$

$$= \frac{1.9}{1.9 + 9.96} = 0.160$$

(2) 陰性反応なら非感染か？

$$P(A^c|B^c)$$

$$= \frac{P(A^c)P(B^c|A^c)}{P(A^c)P(B^c|A^c) + P(A)P(B^c|A)}$$

$$= \frac{\frac{498}{500} \times 0.98}{\frac{498}{500} \times 0.98 + \frac{2}{500} \times 0.05}$$

$$= \frac{488.04}{488.04 + 0.1} = 0.999795$$

演習問題 1（続き） (page 13-14)

1.6 半径 1 の円板を半径 $\frac{1}{4}$, $\frac{1}{2}$, $\frac{3}{4}$, 1 である 4 つの同心円に分けた標的がある．標的に向かってランダムにダーツ（投げ矢）を 10 回独立に投げる．

（1） 高々 1 本が半径 $\frac{1}{2}$ の円に囲まれる区域に当たる確率を計算せよ．

（2） 5 本が半径 $\frac{1}{2}$ の円の内側に当たったとき，少なくとも 1 本が半径 $\frac{1}{4}$ の円内に当たる確率を求めよ．

1.7 次の式が成立するための条件をそれぞれ調べよ．

（1） $P(A \mid B) + P(A^c \mid B^c) = 1$

（2） $P(A \mid B) = P(A \mid B^c)$

1.8　事象 A, B が互いに独立であるとき，A^c, B^c も互いに独立であることを示せ．

1.9　事象 A, B, C が

$$P(A \cap B \cap C) \neq 0, \quad P(C \mid A \cap B) = P(C \mid B)$$

を満たすとき，$P(A \mid B \cap C) = P(A \mid B)$ が成り立つことを示せ．

1.10　2 つの壺 U_1, U_2 があって，U_1 には赤球 5 個，白球 3 個，黒球 2 個；U_2 には赤球 2 個，白球 3 個，黒球 5 個が入っている．いま，U_1 から 1 個の球を取り出して U_2 に入れ，次に U_2 から 1 個の球を取り出したところ黒球であった．はじめに U_1 から取り出した球が黒球である確率を求めよ．

1.11　30 人が集まったパーティで，その中に，誕生日が同じ（同月同日）者が一組でもいる確率はいくらか．ただし，2 月 29 日は除くものとする．「いる」，「いない」の確率がほぼ五分五分となるのは参会者が何人のときか．

Lecture 3

データの整理

【教科書】

第 2 章 データ処理

新生児の体重データ （測定単位 g, 標本数 n = 100)

3110	2500	2770	3010	3000	3000	2740	3040	3060	3410
3100	2620	3910	3650	2840	2480	2790	3720	3520	2850
3140	2780	2270	2700	2830	3020	3160	4060	2620	3390
3050	3190	3710	3460	3200	3260	3040	3610	3360	3280
2480	3440	2970	3050	2590	3320	3580	3820	3450	4150
3300	3020	3360	3140	3300	3600	3330	3300	3300	3170
3340	3250	2880	3560	3060	3320	2740	2380	3590	2460
2960	3170	3000	3250	3140	3220	3160	3730	3460	3360
3160	3540	2890	3060	2900	3040	3220	3590	2680	3150
2770	3220	2970	3300	3560	3520	2760	2740	2820	4180

粗データ/ ローデータ (raw data)

粗データ/ ローデータ (raw data)

3110	2500	2770	3010	3000	3000	2740	3040	3060	3410
3100	2620	3910	3650	2840	2480	2790	3720	3520	2850
3140	2780	2270	2700	2830	3020	3160	4060	2620	3390
3050	3190	3710	3460	3200	3260	3040	3610	3360	3280
2480	3440	2970	3050	2590	3320	3580	3820	3450	4150
3300	3020	3360	3140	3300	3600	3330	3300	3300	3170
3340	3250	2880	3560	3060	3320	2740	2380	3590	2460
2960	3170	3000	3250	3140	3220	3160	3730	3460	3360
3160	3540	2890	3060	2900	3040	3220	3590	2680	3150
2770	3220	2970	3300	3560	3520	2760	2740	2820	4180

- わかりやすく整理
- 可視化
- 特徴を抽出

度数分布表 (frequency table)

いくつかの 階級（クラス）に
分類して表に整理する

① データの範囲

最大値 (max)　max = 4180
最小値 (min)　min = 2270
範囲 (range)　R = max - min = 1910

② 階級幅と階級数
目的に応じて決める
階級幅= 200　階級数 = 11 としよう

③ 各データはただ１つの階級に属する

④ 階級値

⑤ 度数の数え上げ

粗データ/ ローデータ (raw data)

3110	2500	2770	3010	3000	3000	2740	3040	3060	3410
3100	2620	3910	3650	2840	2480	2790	3720	3520	2850
3140	2780	2270	2700	2830	3020	3160	4060	2620	3390
3050	3190	3710	3460	3200	3260	3040	3610	3360	3280
2480	3440	2970	3050	2590	3320	3580	3820	3450	4150
3300	3020	3360	3140	3300	3600	3330	3300	3300	3170
3340	3250	2880	3560	3060	3320	2740	2380	3590	2460
2960	3170	3000	3250	3140	3220	3160	3730	3460	3360
3160	3540	2890	3060	2900	3040	3220	3590	2680	3150
2770	3220	2970	3300	3560	3520	2760	2740	2820	4180

- わかりやすく整理
- 可視化
- 特徴を抽出

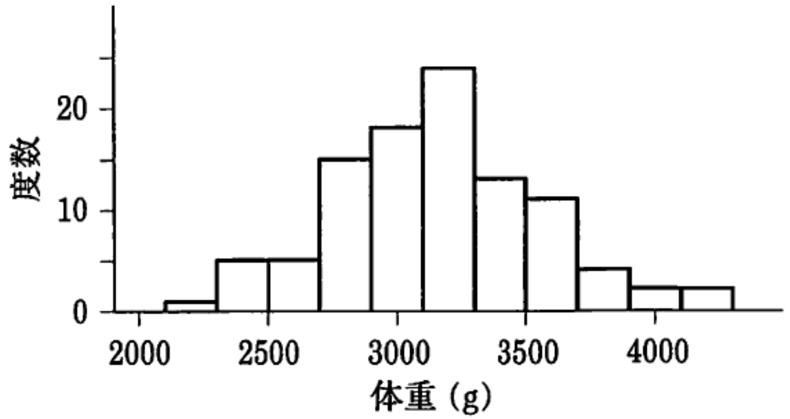
度数分布表 (frequency table)

階級番号	階級	階級値	度数
1	2100~2300	2200	1
2	2300~2500	2400	5
3	2500~2700	2600	5
4	2700~2900	2800	15
5	2900~3100	3000	18
6	3100~3300	3200	24
7	3300~3500	3400	13
8	3500~3700	3600	11
9	3700~3900	3800	4
10	3900~4100	4000	2
11	4100~4300	4200	2
計			100

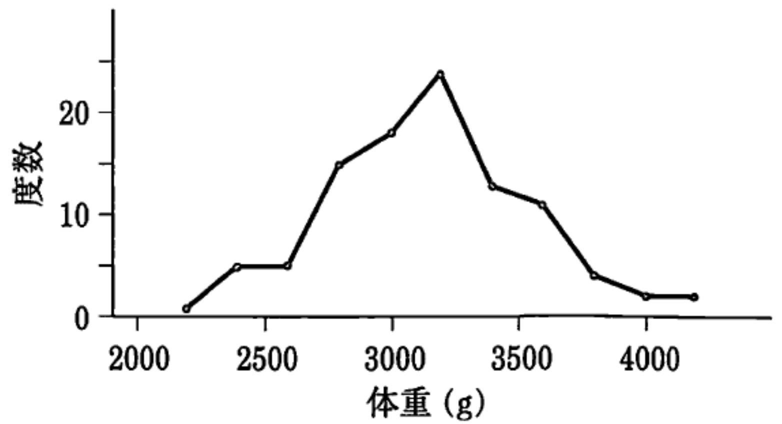
度数分布表 (frequency table)

階級番号	階級	階級値	度数
1	2100~2300	2200	1
2	2300~2500	2400	5
3	2500~2700	2600	5
4	2700~2900	2800	15
5	2900~3100	3000	18
6	3100~3300	3200	24
7	3300~3500	3400	13
8	3500~3700	3600	11
9	3700~3900	3800	4
10	3900~4100	4000	2
11	4100~4300	4200	2
計			100

ヒストグラム



度数多角形



• データの特性値

粗データ：$x_1, x_2, \cdots, x_n$

度数データ：

階級値	c_1	c_2	$\cdots$	c_j	$\cdots$	c_k	計
度数	f_1	f_2	$\cdots$	f_j	$\cdots$	f_k	n
相対度数	p_1	p_2	$\cdots$	p_j	$\cdots$	p_k	1

相対度数：　$p_j = \dfrac{f_j}{n}$

$$\sum_{i=1}^{n} x_i = x_1 + x_2 + \cdots + x_n$$

和の記号

明らかなときは
i の範囲を省略

平均値 (mean/average)

粗データの場合：　$\bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i = \dfrac{1}{n}\sum x_i$

度数データの場合：　$\bar{x} = \dfrac{1}{n}\displaystyle\sum_{j=1}^{k} c_j f_j = \dfrac{1}{n}\sum c_j f_j$

中央値 (median)

粗データの場合：データを小さいほうから並べて順位が中央の値

度数データの場合：多少の計算を要す（教科書参照）

最頻値 (mode)

度数データの場合のみ：最大度数をとる階級値

分散 (variance)　$s^2 = s_x^2$

粗データの場合：　$$s^2 = \frac{1}{n}\sum(x_i - \bar{x})^2$$

度数データの場合：　$$s^2 = \frac{1}{n}\sum_{j=1}^{k}\left(c_j - \bar{x}\right)^2 f_j$$

注意　**不偏分散**

後出

標準偏差 (standard variation)　$s = \sqrt{s^2}$

定理 2.1

（1）偏差の和はゼロである： $\displaystyle\sum_{i=1}^{n}(x_i - \bar{x}) = 0$

（2）分散公式

データの2乗の平均値 $= \overline{x^2}$

粗データの場合： $\displaystyle s^2 = \frac{1}{n}\sum_{i=1}^{n} {x_i}^2 - \bar{x}^2$

度数データの場合： $\displaystyle s^2 = \frac{1}{n}\sum_{j=1}^{k} {c_j}^2 f_j - \bar{x}^2$

$$\boxed{s^2 = \overline{x^2} - \bar{x}^2}$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 = \frac{1}{n}\sum\left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right)$$

$$= \frac{1}{n}\sum x_i^2 - 2\bar{x}\underbrace{\frac{1}{n}\sum x_i}_{\text{平均値 } \bar{x}} + \underbrace{\frac{1}{n}\sum \bar{x}^2}_{\bar{x}^2} = \frac{1}{n}\sum {x_i}^2 - \bar{x}^2$$

例題 2.3 （一部）

平均, 分散, 標準偏差を求めよ.

（単位 100 g）

階級値	c_j	22	24	26	28	30	32	34	36	38	40	42	計
度数	f_j	1	5	5	15	18	24	13	11	4	2	2	100
	$c_j f_j$	22	120	130	420	540	768	442	396	152	80	84	3154
	c_j^2	484	576	676	784	900	1024	1156	1296	1444	1600	1764	
	$c_j^2 f_j$	484	2880	3380	11760	16200	24576	15028	14256	5776	3200	3528	101068

$$\bar{x} = \frac{1}{n}\sum c_j f_j = \frac{1}{100} \times 3154 = 31.54$$

$$\overline{x^2} = \frac{1}{n}\sum c_j^2 f_j = \frac{1}{100} \times 101068 = 1010.68$$

$$s^2 = \overline{x^2} - \bar{x}^2$$

$$= 1010.68 - 31.54^2$$

$$= 15.91$$

$$s = \sqrt{15.91} = 3.99$$

演習問題 2.4（一部）

平均, 分散, 標準偏差を求めよ.

差	c_j	0	1	2	3	4	5	6	7	8	9	10	11	12	計
度数	f_j	5	16	13	9	13	10	9	10	7	5	2	1	0	100
	$c_j f_j$	**0**	**16**	**26**	**27**	**52**	**50**	**54**	**70**	**56**	**45**	**20**	**11**	**0**	**427**
	c_j^2	**0**	**1**	**4**	**9**	**16**	**25**	**36**	**49**	**64**	**81**	**100**	**121**	**144**	
	$c_j^2 f_j$	**0**	**16**	**52**	**81**	**208**	**250**	**324**	**490**	**448**	**405**	**200**	**121**	**0**	**2595**

$$\bar{x} = \frac{1}{n}\sum c_j f_j = \frac{1}{100} \times 427 = 4.27$$

$$\overline{x^2} = \frac{1}{n}\sum c_j^2 f_j = \frac{1}{100} \times 2529 = 25.29$$

$$\begin{aligned} s^2 &= \overline{x^2} - \bar{x}^2 \\ &= 25.29 - 4.27^2 \\ &= 7.72 \end{aligned}$$

$$s = \sqrt{7.72} = 2.78$$

英語と数学の成績の比較や日米の所得の比較を行うときのように測定の内容や測定単位が異 なるときには, 単にデータ（の数値）をそのまま比較することはできない．同じ土俵上で検討できるようにデータを変換する必要がある.

粗データ：$x_1, \cdots, x_i, \cdots, x_n$

平均値：$\bar{x}$

標準偏差：$s = s_x$

標準化（z 変換）

$$z_i = \frac{x_i - \bar{x}}{s}$$

重要な注意　標準化した変数については

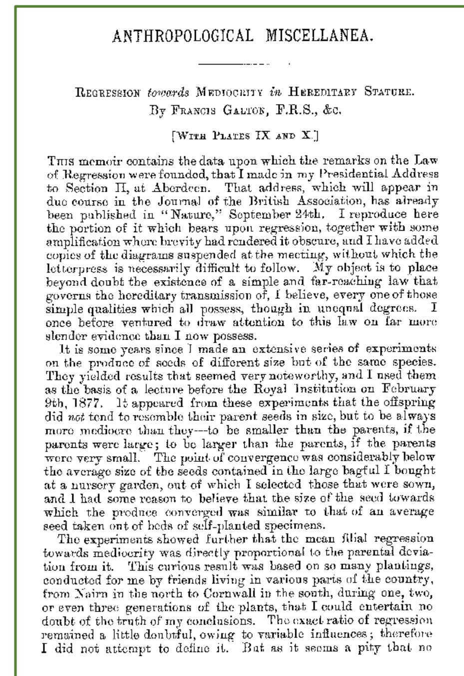
平均値 $= \bar{z} = 0$　　標準偏差 $= s_z = 1$

An example of data : 親の身長と子の身長 (x, y)

Mid-Heights of Parents (x) / Heights of Adult Children (y)

	below	64.5	65.5	66.5	67.5	68.5	69.5	70.5	71.5	72.5	above	sum
above							5	3	2	4		14
73.2						3	4	3	2	2	3	17
72.2			1		4	4	11	4	9	7	1	41
71.2			2		11	18	20	7	4	2		64
70.2			5	4	19	21	25	14	10	1		99
69.2	1	2	7	13	38	48	33	18	5	2		167
68.2	1		7	14	28	34	20	12	3	1		120
67.2	2	5	11	17	38	31	27	3	4			138
66.2	2	5	11	17	36	25	17	1	3			117
65.2	1	1	7	2	15	16	4	1	1			48
64.2	4	4	5	5	14	11	16					59
63.2	2	4	9	3	5	7	1	1				32
62.2		1		3	3							7
below	1	1	1			1		1				5
sum	14	23	66	78	211	219	183	68	43	19	4	928

F. Galton（ゴルトン）
Regression towards mediocrity in hereditary stature, Anthropological Miscellanea (1886)


ANTHROPOLOGICAL MISCELLANEA.

REGRESSION *towards* MEDIOCRITY *in* HEREDITARY STATURE.
By FRANCIS GALTON, F.R.S., &c.

[WITH PLATES IX AND X.]

THIS memoir contains the data upon which the remarks on the Law of Regression were founded, that I made in my Presidential Address to Section H, at Aberdeen. That address, which will appear in due course in the Journal of the British Association, has already been published in "Nature," September 24th. I reproduce here the portion of it which bears upon regression, together with some amplification where brevity had rendered it obscure, and I have added copies of the diagrams suspended at the meeting, without which the letterpress is necessarily difficult to follow. My object is to place beyond doubt the existence of a simple and far-reaching law that governs the hereditary transmission of, I believe, every one of those simple qualities which all possess, though in unequal degrees. I once before ventured to draw attention to this law on far more slender evidence than I now possess.

It is some years since I made an extensive series of experiments on the produce of seeds of different size but of the same species. They yielded results that seemed very noteworthy, and I used them as the basis of a lecture before the Royal Institution on February 9th, 1877. It appeared from these experiments that the offspring did *not* tend to resemble their parent seeds in size, but to be always more mediocre than they—to be smaller than the parents, if the parents were large; to be larger than the parents, if the parents were very small. The point of convergence was considerably below the average size of the seeds contained in the large bagful I bought at a nursery garden, out of which I selected those that were sown, and I had some reason to believe that the size of the seed towards which the produce converged was similar to that of an average seed taken out of beds of self-planted specimens.

The experiments showed further that the mean filial regression towards mediocrity was directly proportional to the parental deviation from it. This curious result was based on so many plantings, conducted for me by friends living in various parts of the country, from Nairn in the north to Cornwall in the south, during one, two, or even three generations of the plants, that I could entertain no doubt of the truth of my conclusions. The exact ratio of regression remained a little doubtful, owing to variable influences; therefore I did not attempt to define it. But as it seems a pity that no

回帰分析の始まり

表 2.4

2変量(2次元)データの例

新生児の身長，　体重のデータ $(n = 60)$

番号	身長	体重
1	46.0	2700
2	49.5	3220
3	50.0	3360
4	50.0	3500
5	49.0	3120
6	50.0	3160
7	53.0	4150
8	48.0	3310
9	49.0	2880
10	50.5	3090
11	49.5	3020
12	49.0	3360
13	50.0	3110
14	50.0	3560
15	47.5	2990
16	50.5	3440
17	48.0	2920
18	49.0	3060
19	49.0	3360
20	50.0	3400

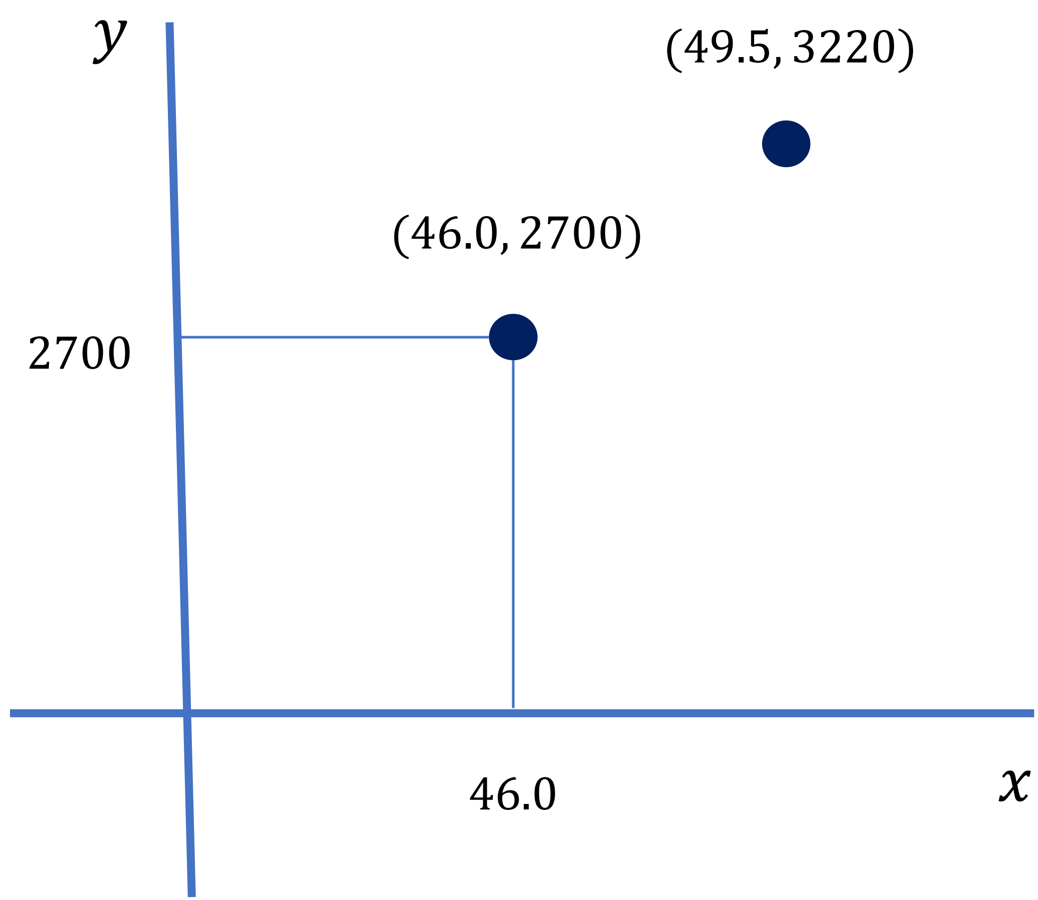
番号	身長	体重
21	48.0	3200
22	50.5	2940
23	48.5	2850
24	50.5	3220
25	48.5	2750
26	49.0	3020
27	48.5	2570
28	48.5	3030
29	45.0	2410
30	51.0	3280
31	50.5	3140
32	49.0	3040
33	52.0	3910
34	50.0	2770
35	46.5	2340
36	50.0	3140
37	50.5	3560
38	50.0	3390
39	50.0	3420
40	51.0	3450

番号	身長	体重
41	49.5	3590
42	48.5	2830
43	48.0	3120
44	51.0	3190
45	50.0	3600
46	47.0	2980
47	50.0	3090
48	51.0	3630
49	53.0	4060
50	50.0	3720
51	50.0	3400
52	50.5	3430
53	51.0	3250
54	48.0	2760
55	50.0	3320
56	49.0	2930
57	50.0	3320
58	48.0	2620
59	47.5	2860
60	48.0	2530

散布図 (Scatter Plot)

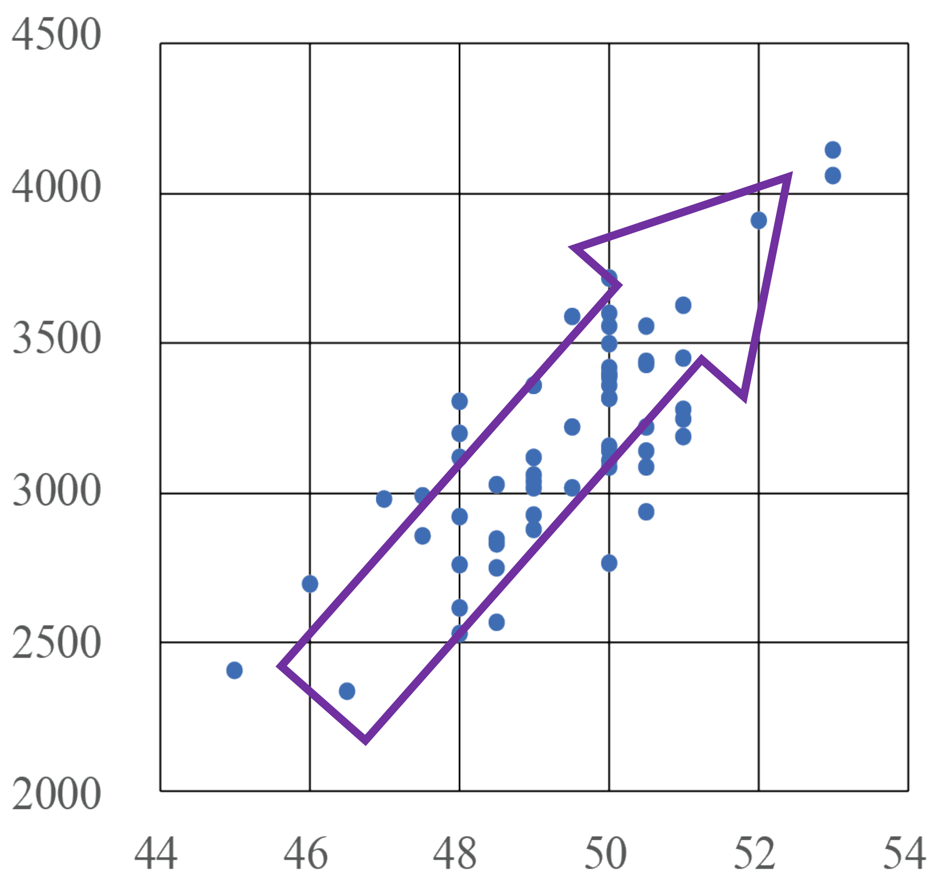
x 座標　y 座標

番号	身長	体重
1	46.0	2700
2	49.5	3220
3	50.0	3360
4	50.0	3500
5	49.0	3120
6	50.0	3160
7	53.0	4150
8	48.0	3310
⋮	⋮	⋮



番号	身長 x	体重 y
1	46.0	2700
2	49.5	3220
3	50.0	3360
⋮	⋮	⋮
i	x_i	y_i
⋮	⋮	⋮
60	48.0	2530

散布図



正の相関

2変量の統計量

番号	身長 x	体重 y
1	46.0	2700
2	49.5	3220
3	50.0	3360
⋮	⋮	⋮
i	x_i	y_i
⋮	⋮	⋮
60	48.0	2530

n：データの個数

x の平均値と分散

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

y の平均値と分散

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i \qquad s_y^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

x と y の共分散と相関係数

$$s_{xy} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

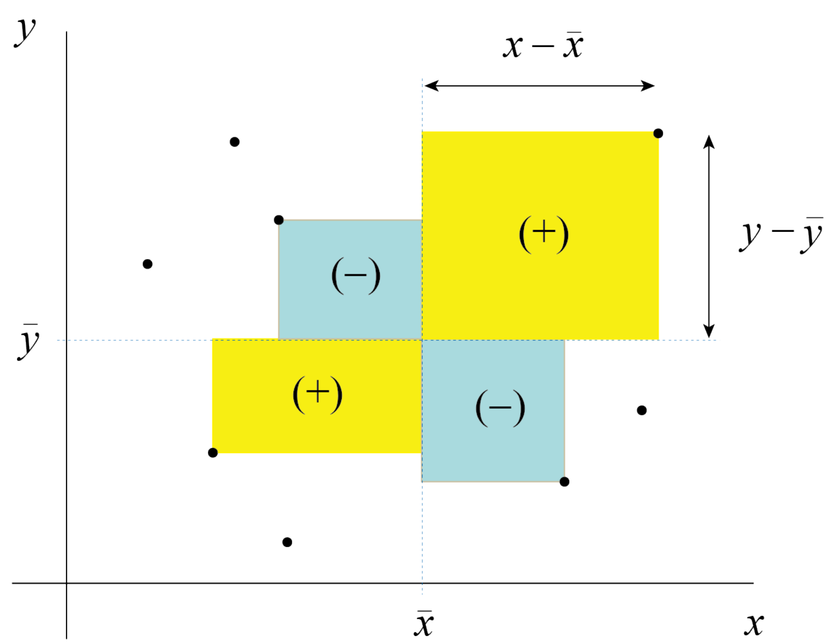
共分散・相関係数の意味

共分散

$$s_{xy} = \frac{1}{n}\sum(x_i - \bar{x})(y_i - \bar{y})$$

$s_{xy} > 0$：概ね右上がりの傾向

$s_{xy} < 0$：概ね右下がりの傾向



相関係数

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = \frac{1}{n}\sum \frac{x_i - \bar{x}}{\sigma_x}\,\frac{y_i - \bar{y}}{\sigma_y}$$

標準化された共分散といえる

相関係数の性質

定理 2.2　$-1 \leq r_{xy} \leq 1$

証明　$\sum\{t(x_i - \bar{x}) + (y_i - \bar{y})\}^2 \geq 0$ がすべての実数 t で成り立つ.

$$t^2 \sum (x_i - \bar{x})^2 + 2t \sum (x_i - \bar{x})(y_i - \bar{y}) + \sum (y_i - \bar{y})^2 \geq 0$$

$$t^2 \sigma_x^2 + 2t\sigma_{xy} + \sigma_y^2 \geq 0$$

$$\frac{D}{4} = \sigma_{xy}^2 - \sigma_x^2 \sigma_y^2 \leq 0$$

$$\sigma_{xy}^2 \leq \sigma_x^2 \sigma_y^2$$

$r_{xy} > 0$：正の相関

$r_{xy} < 0$：負の相関

$|r_{xy}| \approx 1$：強い相関

$|r_{xy}| \approx 0$：弱い相関

表 2.4

番号	身長 x	体重 y
1	46.0	2700
2	49.5	3220
3	50.0	3360
⋮	⋮	⋮
i	x_i	y_i
⋮	⋮	⋮
60	48.0	2530

相関表 = 2次元の度数分布表

y \ x	45	46	47	48	49	50	51	52	53	計
24	1	1	0	0	0	0	0	0	0	2
26	0	1	0	3	0	0	0	0	0	4
28	0	0	1	4	1	1	0	0	0	7
30	0	0	2	2	5	3	0	0	0	12
32	0	0	0	2	2	5	3	0	0	12
34	0	0	0	1	2	10	1	0	0	14
36	0	0	0	0	1	3	1	0	0	5
38	0	0	0	0	0	1	0	0	0	1
40	0	0	0	0	0	0	0	1	1	2
42	0	0	0	0	0	0	0	0	1	1
計	1	2	3	12	11	23	5	1	2	60

（y の本来の値は100倍する）

y \ x	45	46	47	48	49	50	51	52	53	計
24	1	1	0	0	0	0	0	0	0	2
26	0	1	0	3	0	0	0	0	0	4
28	0	0	1	4	1	1	0	0	0	7
30	0	0	2	2	5	3	0	0	0	12
32	0	0	0	2	2	5	3	0	0	12
34	0	0	0	1	2	10	1	0	0	14
36	0	0	0	0	1	3	1	0	0	5
38	0	0	0	0	0	1	0	0	0	1
40	0	0	0	0	0	0	0	1	1	2
42	0	0	0	0	0	0	0	0	1	1
計	1	2	3	12	11	23	5	1	2	60

$$\bar{x} = \frac{1}{n}\sum x_j f_j$$

$$= \frac{1}{60} \times 2956 = 49.3$$

$$\overline{x^2} = \frac{1}{n}\sum x_j^2 f_j$$

$$= \frac{1}{60} \times 145770 = 2429.5$$

$$s_x^2 = \overline{x^2} - \bar{x}^2$$

$$= 2429.5 - 49.3^2 = 2.296$$

$$s = \sqrt{2.296} = 1.52$$

x	45	46	47	48	49	50	51	52	53	計
f	1	2	3	12	11	23	5	1	2	60

y \ x	45	46	47	48	49	50	51	52	53	計
24	1	1	0	0	0	0	0	0	0	2
26	0	1	0	3	0	0	0	0	0	4
28	0	0	1	4	1	1	0	0	0	7
30	0	0	2	2	5	3	0	0	0	12
32	0	0	0	2	2	5	3	0	0	12
34	0	0	0	1	2	10	1	0	0	14
36	0	0	0	0	1	3	1	0	0	5
38	0	0	0	0	0	1	0	0	0	1
40	0	0	0	0	0	0	0	1	1	2
42	0	0	0	0	0	0	0	0	1	1
計	1	2	3	12	11	23	5	1	2	60

$$\bar{y} = \frac{1}{n}\sum y_j f_j$$

$$= \frac{1}{60} \times 1908 = 31.8$$

$$\overline{y^2} = \frac{1}{n}\sum y_j^2 f_j$$

$$= \frac{1}{60} \times 61504 = 1025.07$$

$$s_y^2 = \overline{y^2} - \bar{y}^2$$

$$= 1025.07 - 31.8^2 = 13.83$$

$$s = \sqrt{13.83} = 3.72$$

y	24	26	28	30	32	34	36	38	40	42	計
f	2	4	7	12	12	14	5	1	2	1	60

共分散の公式

$$s_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}$$

積の平均値

平均値の積

$$s_{xy} = \frac{1}{n}\sum (x_i - \bar{x})(y_j - \bar{y})$$

$$= \frac{1}{n}\sum x_i y_j - \bar{x}\frac{1}{n}\sum y_j - \bar{y}\frac{1}{n}\sum x_i + \bar{x} \cdot \bar{y}$$

$$= \overline{xy} - \bar{x} \cdot \bar{y}$$

y \ x	45	46	47	48	49	50	51	52	53	計
24	1	1	0	0	0	0	0	0	0	2
26	0	1	0	3	0	0	0	0	0	4
28	0	0	1	4	1	1	0	0	0	7
30	0	0	2	2	5	3	0	0	0	12
32	0	0	0	2	2	5	3	0	0	12
34	0	0	0	1	2	10	1	0	0	14
36	0	0	0	0	1	3	1	0	0	5
38	0	0	0	0	0	1	0	0	0	1
40	0	0	0	0	0	0	0	1	1	2
42	0	0	0	0	0	0	0	0	1	1
計	1	2	3	12	11	23	5	1	2	60

共分散のためには

$$\overline{xy} = \frac{1}{n}\sum x_j y_j f_{ij}$$

xyf の表

y \ x	45	46	47	48	49	50	51	52	53	計
24	1080	1104	0	0	0	0	0	0	0	2
26	0	1196	0	3744	0	0	0	0	0	4
28	0	0	1316	5376	1372	1400	0	0	0	7
30	0	0	2820	2880	7350	4500	0	0	0	12
32	0	0	0	3072	3136	8000	4896	0	0	12
34	0	0	0	1632	3332	17000	1734	0	0	14
36	0	0	0	0	1764	5400	1836	0	0	5
38	0	0	0	0	0	1900	0	0	0	1
40	0	0	0	0	0	0	0	2080	2120	2
42	0	0	0	0	0	0	0	0	2226	1
計	1	2	3	12	11	23	5	1	2	60

$$\overline{xy} = \frac{1}{n}\sum x_j y_j f_{ij}$$

$$= \frac{1}{60} \times 9426600$$

$$= 1571.1$$

$$s_{xy} = \overline{xy} - \bar{x} \cdot \bar{y}$$

$$= 1571.1 - 49.3 \times 31.8$$

$$= 4.42$$

$$r_{xy} = \frac{s_{xy}}{s_x s_y}$$

$$= \frac{4.42}{1.52 \times 3.72}$$

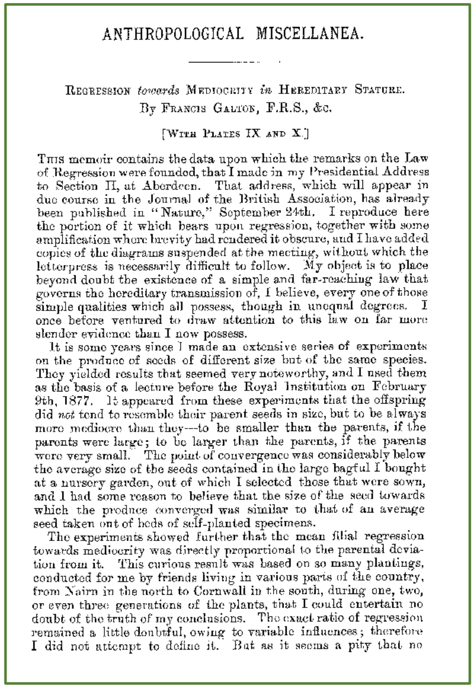
$$= 0.78$$

An example of data：親の身長と子の身長 (x, y)

Heights of Adult Children (y) / Mid-Heights of Parents (x)

	below	64.5	65.5	66.5	67.5	68.5	69.5	70.5	71.5	72.5	above	sum
above							5	3	2	4		14
73.2						3	4	3	2	2	3	17
72.2			1		4	4	11	4	9	7	1	41
71.2			2		11	18	20	7	4	2		64
70.2			5	4	19	21	25	14	10	1		99
69.2	1	2	7	13	38	48	33	18	5	2		167
68.2	1		7	14	28	34	20	12	3	1		120
67.2	2	5	11	17	38	31	27	3	4			138
66.2	2	5	11	17	36	25	17	1	3			117
65.2	1	1	7	2	15	16	4	1	1			48
64.2	4	4	5	5	14	11	16					59
63.2	2	4	9	3	5	7	1	1				32
62.2		1		3	3							7
below	1	1	1			1		1				5
sum	14	23	66	78	211	219	183	68	43	19	4	928

F. Galton（ゴルトン）
Regression towards mediocrity in hereditary stature, Anthropological Miscellanea (1886)



ANTHROPOLOGICAL MISCELLANEA.

REGRESSION *towards* MEDIOCRITY *in* HEREDITARY STATURE.
By FRANCIS GALTON, F.R.S., &c.

[WITH PLATES IX AND X.]

THIS memoir contains the data upon which the remarks on the Law of Regression were founded, that I made in my Presidential Address to Section H, at Aberdeen. That address, which will appear in due course in the Journal of the British Association, has already been published in "Nature," September 24th. I reproduce here the portion of it which bears upon regression, together with some amplification where brevity had rendered it obscure, and I have added copies of the diagrams suspended at the meeting, without which the letterpress is necessarily difficult to follow. My object is to place beyond doubt the existence of a simple and far-reaching law that governs the hereditary transmission of, I believe, every one of those simple qualities which all possess, though in unequal degrees. I once before ventured to draw attention to this law on far more slender evidence than I now possess.

It is some years since I made an extensive series of experiments on the produce of seeds of different size but of the same species. They yielded results that seemed very noteworthy, and I used them as the basis of a lecture before the Royal Institution on February 9th, 1877. It appeared from these experiments that the offspring did *not* tend to resemble their parent seeds in size, but to be always more mediocre than they—to be smaller than the parents, if the parents were large; to be larger than the parents, if the parents were very small. The point of convergence was considerably below the average size of the seeds contained in the large bagful I bought at a nursery garden, out of which I selected those that were sown, and I had some reason to believe that the size of the seed towards which the produce converged was similar to that of an average seed taken out of beds of self-planted specimens.

The experiments showed further that the mean filial regression towards mediocrity was directly proportional to the parental deviation from it. This curious result was based on so many plantings, conducted for me by friends living in various parts of the country, from Nairn in the north to Cornwall in the south, during one, two, or even three generations of the plants, that I could entertain no doubt of the truth of my conclusions. The exact ratio of regression remained a little doubtful, owing to variable influences; therefore I did not attempt to define it. But as it seems a pity that no

回帰分析の始まり

		Mid-height parents (x)									
Adult Children (y)		64.5	65.5	66.5	67.5	68.5	69.5	70.5	71.5	72.5	sum
	73.2					3	4	3	2	2	14
	72.2		1		4	4	11	4	9	7	40
	71.2		2		11	18	20	7	4	2	64
	70.2		5	4	19	21	25	14	10	1	99
	69.2	2	7	13	38	48	33	18	5	2	166
	68.2		7	14	28	34	20	12	3	1	119
	67.2	5	11	17	38	31	27	3	4		136
	66.2	5	11	17	36	25	17	1	3		115
	65.2	1	7	2	15	16	4	1	1		47
	64.2	4	5	5	14	11	16				55
	63.2	4	9	3	5	7	1	1			30
	62.2	1		3	3						7
	sum	22	65	78	211	218	178	64	41	15	892

$$\bar{x} = 68.3$$
$$\sigma_x^2 = 2.77$$
$$\sigma_x = 1.67$$

$$\bar{y} = 68.1$$
$$\sigma_y^2 = 5.62$$
$$\sigma_y = 2.37$$

$$\sigma_{xy} = 1.60$$
$$r_{xy} = 0.41$$

宿題

演習問題２　(page 37-39)

2.4 トランプ 52 枚を伏せ，任意の 2 枚をめくったときの，数字の差のデータ分布は次の通りであった．平均と分散を求めよ．また，理論値の度数を計算して比較せよ．

差	0	1	2	3	4	5	6	7	8	9	10	11	12
度数	5	16	13	9	13	10	9	10	7	5	2	1	0

2.7 次のデータはある高校での試験の成績である．相関係数を計算せよ．

英語	83	80	48	68	70	45	72	28	51	32	42	38	52	80	52	78	32	60	54	49
国語	55	42	32	71	67	60	63	51	49	51	64	15	73	71	32	68	42	55	62	31

2.4 トランプ 52 枚を伏せ，任意の 2 枚をめくったときの，数字の差のデータ分布は次の通りであった．平均と分散を求めよ．また，理論値の度数を計算して比較せよ．

差	0	1	2	3	4	5	6	7	8	9	10	11	12
度数	5	16	13	9	13	10	9	10	7	5	2	1	0

差(D)	0	1	2	3	4	5	6	7	8	9	10	11	12	和
度数(f)	5	16	13	9	13	10	9	10	7	5	2	1	0	100
Df	0	16	26	27	52	50	54	70	56	45	20	11	0	427
D^2	0	1	4	9	16	25	36	49	64	81	100	121	144	
D^2f	0	16	52	81	208	250	324	490	448	405	200	121	0	2595

$$\overline{D} = \frac{1}{100}\,427 = 4.27 \qquad \overline{D^2} = \frac{1}{100}\,2595 = 25.95$$

$$s_D^2 = \overline{D^2} - \overline{D}^2 = 25.95 - 4.27^2 = 7.72 \qquad s_D = \sqrt{7.71} = 2.78$$

2.4 トランプ52枚を伏せ，任意の2枚をめくったときの，数字の差のデータ分布は次の通りであった．平均と分散を求めよ．また，理論値の度数を計算して比較せよ．

差	0	1	2	3	4	5	6	7	8	9	10	11	12
度数	5	16	13	9	13	10	9	10	7	5	2	1	0

2枚抜き取る組合せ数 $= \binom{52}{2} = \dfrac{52 \times 51}{2 \times 1} = 1326$

$D = 0$ となる組み合わせ数 $= \binom{4}{2} \times 13 = 6 \times 13 = 78$

$D = 1$ となる組み合わせ数 $= 4 \times 4 \times 12 = 16 \times 12 = 192$

・・・

$D = 12$ となる組み合わせ数 $= 4 \times 4 \times 1 = 16 \times 1 = 16$

差(D)	0	1	2	3	4	5	6	7	8	9	10	11	12	和
度数(f)	78	192	176	160	144	128	112	96	80	64	48	32	16	1326

2.4 トランプ 52 枚を伏せ，任意の 2 枚をめくったときの，数字の差のデータ分布は次の通りであった．平均と分散を求めよ．また，理論値の度数を計算して比較せよ．

差	0	1	2	3	4	5	6	7	8	9	10	11	12
度数	5	16	13	9	13	10	9	10	7	5	2	1	0

差(D)	0	1	2	3	4	5	6	7	8	9	10	11	12	和
度数(f)	78	192	176	160	144	128	112	96	80	64	48	32	16	1326
Df	0	192	352	480	576	640	672	672	640	576	480	352	192	5824
D^2	0	1	4	9	16	25	36	49	64	81	100	121	144	
D^2f	0	192	704	1440	2304	3200	4032	4704	5120	5184	4800	3872	2304	37856

$$\overline{D} = \frac{1}{1326}\, 5824 = 4.39 \qquad \overline{D^2} = \frac{1}{1326}\, 37856 = 28.55$$

$$s_D^2 = \overline{D^2} - \overline{D}^2 = 28.55 - 4.39^2 = 9.28 \qquad s_D = \sqrt{9.28} = 3.05$$

2.7 次のデータはある高校での試験の成績である．相関係数を計算せよ．

英語	83	80	48	68	70	45	72	28	51	32	42	38	52	80	52	78	32	60	54	49
国語	55	42	32	71	67	60	63	51	49	51	64	15	73	71	32	68	42	55	62	31

データの大きさ= 20

																					和
英語 (e)	83	80	48	68	70	45	72	28	51	32	42	38	52	80	52	78	32	60	54	49	1114
国語 (j)	55	42	32	71	67	60	63	51	49	51	64	15	73	71	32	68	42	55	62	31	1054
e^2	6889	6400	2304	4624	4900	2025	5184	784	2601	1024	1764	1444	2704	6400	2704	6084	1024	3600	2916	2401	67776
j^2	3025	1764	1024	5041	4489	3600	3969	2601	2401	2601	4096	225	5329	5041	1024	4624	1764	3025	3844	961	60448
ej	4565	3360	1536	4828	4690	2700	4536	1428	2499	1632	2688	570	3796	5680	1664	5304	1344	3300	3348	1519	60987

$$\bar{e} = \frac{1}{20}\,1114 = 55.7 \qquad \bar{j} = \frac{1}{20}\,1054 = 52.7$$

$$\overline{e^2} = \frac{1}{20}\,67776 = 3388.87 \qquad \overline{j^2} = \frac{1}{20}\,60448 = 3022.4 \qquad \overline{ej} = \frac{1}{20}\,60987 = 3049.35$$

$$s_e^2 = \overline{e^2} - \bar{e}^2 = 286.31 \qquad s_j^2 = \overline{j^2} - \bar{j}^2 = 245.11 \qquad s_{ej} = \overline{ej} - \bar{e}\bar{j} = 113.96$$

2.7 次のデータはある高校での試験の成績である．相関係数を計算せよ．

英語	83	80	48	68	70	45	72	28	51	32	42	38	52	80	52	78	32	60	54	49
国語	55	42	32	71	67	60	63	51	49	51	64	15	73	71	32	68	42	55	62	31

データの大きさ= 20

																					和
英語 (e)	83	80	48	68	70	45	72	28	51	32	42	38	52	80	52	78	32	60	54	49	1114
国語 (j)	55	42	32	71	67	60	63	51	49	51	64	15	73	71	32	68	42	55	62	31	1054
e^2	6889	6400	2304	4624	4900	2025	5184	784	2601	1024	1764	1444	2704	6400	2704	6084	1024	3600	2916	2401	67776
j^2	3025	1764	1024	5041	4489	3600	3969	2601	2401	2601	4096	225	5329	5041	1024	4624	1764	3025	3844	961	60448
ej	4565	3360	1536	4828	4690	2700	4536	1428	2499	1632	2688	570	3796	5680	1664	5304	1344	3300	3348	1519	60987

$s_e^2 = \overline{e^2} - \bar{e}^2 = 286.31$ $s_j^2 = \overline{j^2} - \bar{j}^2 = 245.11$ $s_{ej} = \overline{ej} - \bar{e}\bar{j} = 113.96$

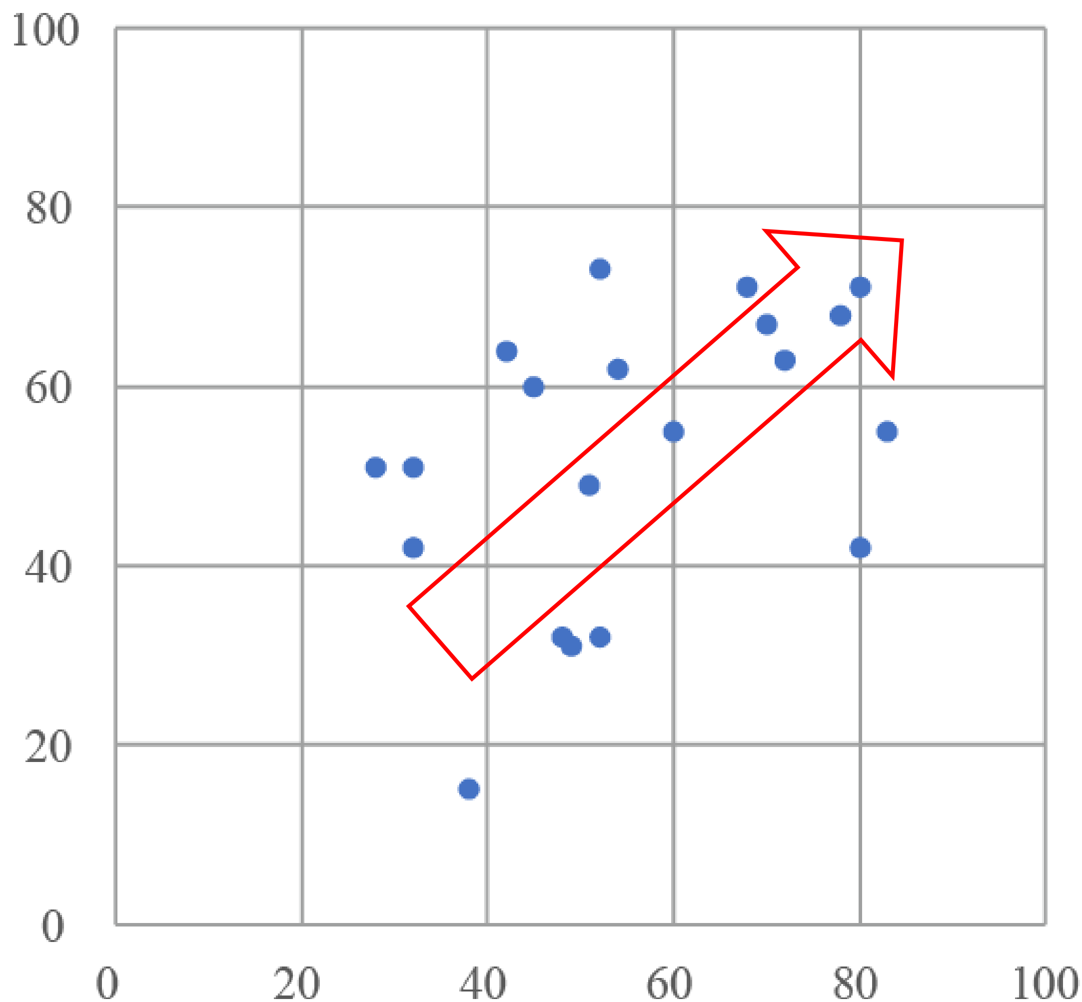
分散

共分散

$$r_{ej} = \frac{s_{ej}}{s_e s_j} = \frac{113.96}{\sqrt{286.31}\sqrt{245.11}} = 0.43$$

2.7　次のデータはある高校での試験の成績である．相関係数を計算せよ．

英語	83	80	48	68	70	45	72	28	51	32	42	38	52	80	52	78	32	60	54	49
国語	55	42	32	71	67	60	63	51	49	51	64	15	73	71	32	68	42	55	62	31



$$r_{ej} = 0.43$$

Lecture 4

確率変数と確率分布

【教科書】

第 3 章　確率変数と確率分布

1 確率変数

2 平均と分散

3 離散型確率変数の分布

変数：ある範囲の値を代表する．$x, y, z, t, \ldots$

確率変数：確率を伴ってある範囲の値をとる．$X, Y, Z, T, \ldots$

- 離散型 (discrete) ：とる値がトビトビである
  - ✓ サイコロをころがしたと きに 出る目の数
  - ✓ 一日に起こる交通事故の件数など
  - ✓ コイン 投げで, 表 = 1, 裏 = 0 としたとき（本来は結果が数値でないときの扱い）
- 連続型 (continuous) ：とる値が隙間なく並ぶ連続量になる
  - ✓ 大気の温度
  - ✓ 学生の身長
  - ✓ 商店の売上金額
  - ✓ 国民経済での総生産高

- 離散型確率変数 X

取りうる値　$a_1, a_2, \dots, a_i, \dots$

その確率　$p_1, p_2, \dots, p_i, \dots$

数式による表記

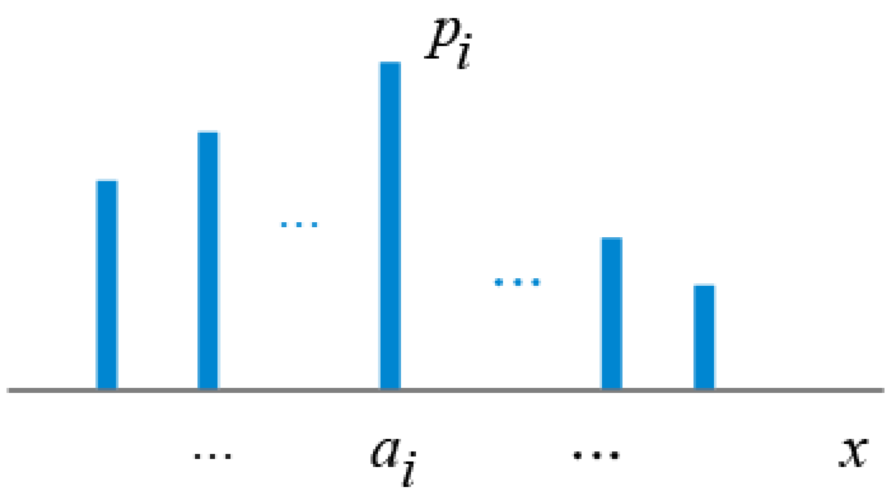
$$P(X = a_i) = p_i$$

x	a_1	$\cdots$	a_i	$\cdots$	合計
$P(X = x)$	p_1	$\cdots$	p_i	$\cdots$	1

基本的な性質

(1)　$p_i \geq 0$

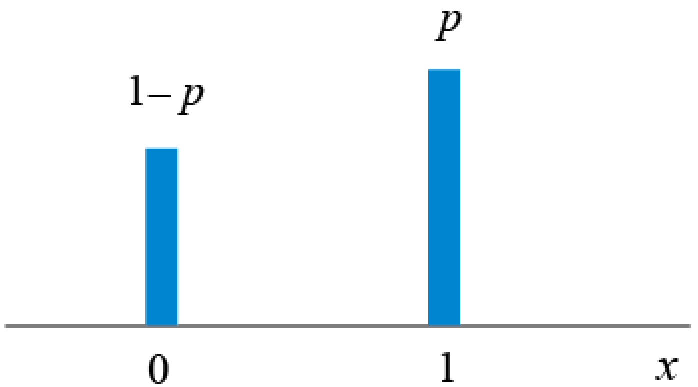
(2)　$\sum p_i = 1$



例題 3.1

もっとも簡単な確率変数は 2 値 $\{0,1\}$ をそれぞれ確率 $p, 1-p$ でとる **2 値確率変数** と呼ばれるものである.

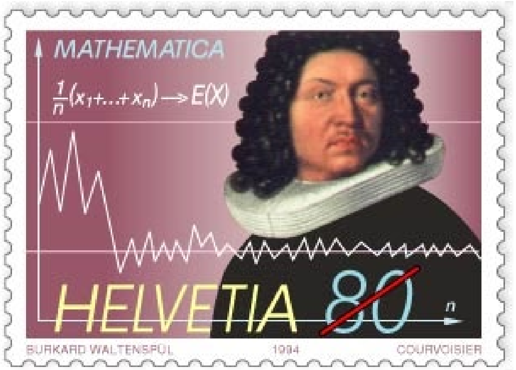
$$Z = \begin{cases} 1 & (\text{確率 } p \text{ で}) \\ 0 & (\text{確率 } 1-p \text{ で}) \end{cases}$$



$$P(Z=0) = 1-p,\quad P(Z=1) = p$$

x	0	1	計
$P(Z=x)$	$1-p$	p	1

Zの分布を**ベルヌーイ分布**といい,
記号 Ber (p) で表す.



Jacob Bernoulli (1654-1705)

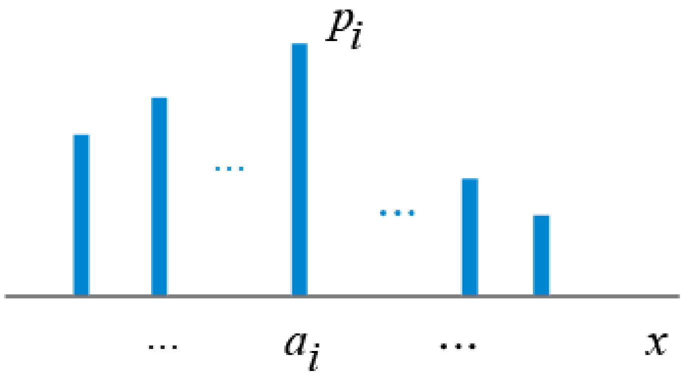
- 平均値 (mean) ＝ 期待値 (expectation)

分布が $P(X = a_i) = p_i$ で与えられている離散型

確率変数 X の平均値：

$$\mathbf{E}[X] = \mu_X = \sum a_i p_i = \sum a_i P(X = a_i)$$

x	a_1	$\cdots$	a_i	$\cdots$	合計
$P(X = x)$	p_1	$\cdots$	p_i	$\cdots$	1



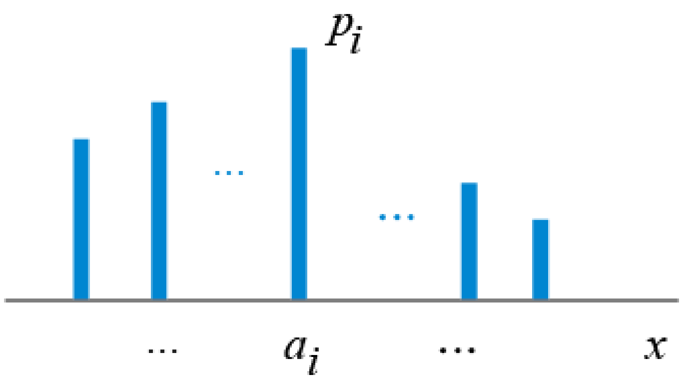
- 平均値 (mean) ＝ 期待値 (expectation)

分布が $P(X = a_i) = p_i$ で与えられている離散型

確率変数 X の平均値：

$$\mathbf{E}[X] = \mu_X = \sum a_i p_i = \sum a_i P(X = a_i)$$

x	a_1	$\cdots$	a_i	$\cdots$	合計
$P(X = x)$	p_1	$\cdots$	p_i	$\cdots$	1
	$a_1 p_1$		$a_i p_i$		



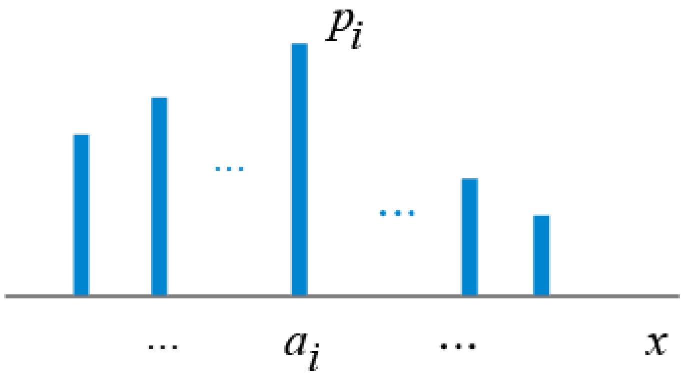
- 平均値 (mean) ＝ 期待値 (expectation)

分布が $P(X = a_i) = p_i$ で与えられている離散型

確率変数 X の平均値：

$$\mathbf{E}[X] = \mu_X = \sum a_i p_i = \sum a_i P(X = a_i)$$

x	a_1	$\cdots$	a_i	$\cdots$	合計
$P(X = x)$	p_1	$\cdots$	p_i	$\cdots$	1
	$a_1 p_1$		$a_i p_i$		μ_X



- 分散 (variance) と標準偏差 (standard variation)

分布が $P(X = a_i) = p_i$ で与えられている離散型

確率変数 X の分散と標準偏差：

$$\mathbf{V}[X] = \sigma_X^2 = \sum (a_i - \mu_X)^2 p_i$$

$$\sigma_X = \sqrt{\mathbf{V}[X]}$$

$$\sum (a_i - \mu_X)^2 p_i = \sum a_i^2 p_i - \mu_X^2 = \mathbf{E}[X^2] - \mathbf{E}[X]^2$$

分散公式：　$\mathbf{V}[X] = \mathbf{E}[X^2] - \mathbf{E}[X]^2$

- 平均値と分散の計算法

x	a_1	$\cdots$	a_i	$\cdots$	合計
$P(X=x)$	p_1	$\cdots$	p_i	$\cdots$	1

- 平均値と分散の計算法

x	a_1	$\cdots$	a_i	$\cdots$	合計
$P(X=x)$	p_1	$\cdots$	p_i	$\cdots$	1
$xP(X=x)$	a_1p_1		a_ip_i		

- 平均値と分散の計算法

x	a_1	$\cdots$	a_i	$\cdots$	合計
$P(X=x)$	p_1	$\cdots$	p_i	$\cdots$	1
$xP(X=x)$	a_1p_1		a_ip_i		$E[X]$

- 平均値と分散の計算法

x	a_1	$\cdots$	a_i	$\cdots$	合計
$P(X=x)$	p_1	$\cdots$	p_i	$\cdots$	1
$xP(X=x)$	a_1p_1		a_ip_i		$E[X]$
x^2					
$x^2P(X=x)$					

- 平均値と分散の計算法

x	a_1	$\cdots$	a_i	$\cdots$	合計
$P(X=x)$	p_1	$\cdots$	p_i	$\cdots$	1
$xP(X=x)$	a_1p_1		a_ip_i		$E[X]$
x^2	a_1^2		a_i^2		
$x^2P(X=x)$	$a_1^2p_1$		$a_i^2p_i$		

- 平均値と分散の計算法

x	a_1	$\cdots$	a_i	$\cdots$	合計
$P(X=x)$	p_1	$\cdots$	p_i	$\cdots$	1
$xP(X=x)$	a_1p_1		a_ip_i		$E[X]$
x^2	a_1^2		a_i^2		
$x^2P(X=x)$	$a_1^2p_1$		$a_i^2p_i$		$E[X^2]$

分散　$\mathbf{V}[X]=\mathbf{E}[X^2]-\mathbf{E}[X]^2$

標準偏差　$\sigma_X=\sqrt{\mathbf{V}[X]}$

例題 3.4　**二項母集団の分布**

ベルヌーイ分布 $\mathrm{Ber}(p)$ に従う確率変数 Z について

平均値　$\mu = \mathbf{E}[Z] = p$　　分散 $\sigma^2 = V[Z] = p(1-p)$

- 確率分布　$P(X=1) = p,\quad P(X=0) = 1-p$

x	0	1	計
$P(Z=x)$	$1-p$	p	1

- 平均値　$\mu = \mathbf{E}[X] = 0 \times (1-p) + 1 \times p = p$
- 分散　$\mathbf{E}[X^2] = 0^2 \times (1-p) + 1^2 \times p = p$

$$\sigma^2 = \mathbf{E}[X^2] - \boldsymbol{E}[X]^2 = p - p^2 = p(1-p)$$

例題 3.5

X をサイコロの出る目の数とするとき, その平均と分散を求めよ. また, 標準偏差はいくらか

X の確率分布を求める

x	1	2	3	4	5	6	合計
$P(X=x)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

$$\mathbf{E}[X] = m_X = \sum a_i p_i = 1 \times \frac{1}{6} + 2 \times \frac{1}{6} + \cdots + 6 \times \frac{1}{6} = \frac{21}{6} = 3.5$$

$$\mathbf{E}[X^2] = \sum a_i^2 p_i = 1^2 \times \frac{1}{6} + 2^2 \times \frac{1}{6} + \cdots + 6^2 \times \frac{1}{6} = \frac{91}{6}$$

$$\mathbf{V[X]} = \mathbf{E}[X^2] - \mathbf{E}[X]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12} = 2.92 \qquad \sigma_X = \sqrt{\mathbf{V}[X]} = 1.71$$

- 標準化（z 変換）

$$Z = \frac{X - E[X]}{\sqrt{V[X]}} = \frac{X - \mu}{\sigma}$$

  - $E[Z] = 0$
  - $V[Z] = 1$

定理 3.1

確率変数が $a + bX$ （ a, b は定数）と表されるとき,

(1) $E[a + bX] = a + bE[X] = a + b\mu$

(2) $V[a + bX] = b^2 V[X] = b^2 \sigma^2$

• 2 項分布 $B(n,p)$

ベルヌーイ試行列

- 成功確率 p の試行を独立に繰り返す（例：サイコロ振り）
- 独立な確率変数列の導入　$Z_1, Z_2, \ldots, Z_n$

$$Z_k = \begin{cases} 1, & \text{確率 } p \text{ で成功} \\ 0, & \text{確率} 1-p \text{ で失敗} \end{cases}$$

- 成功回数

$$X = \sum_{k=1}^{n} Z_k \qquad \{0,1,2,\cdots,n\} \text{ に値をとる離散型確率変数になる}$$

定義　X の分布を二項分布といい, $B(n,p)$ で表す.

$X \sim B(n,p)$ と書く.

$X \sim B(n,p)$ の確率分布の計算

x	0	1	$\cdots$	k	$\cdots$	n	合計
$P(X=x)$							1

- n 回中, ○（成功）が k 回, ×（失敗）が $n-k$ 回の確率

その一例　○×・・・○×・・・×○×

その確率　$p\ q\ \cdots\ p\ q\ \cdots\ q\ p\ q = p^k q^{n-k}$

（ただし, $q = 1-p$）

したがって,

$$\boxed{\begin{aligned} P(X=k) &= \frac{n!}{k!\,(n-k)!}p^k(1-p)^{n-k} \\ &= \binom{n}{k}\, p^k(1-p)^{n-k} \end{aligned}}$$

- 二項分布 $B(n,p)$ のまとめ

  - $X \sim B(n,p)$ とすると, X は成功確率 p のベルヌーイ試行を n 回行ったときの成功回数を意味する.

  - 確率分布

$$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$$

  - 平均値

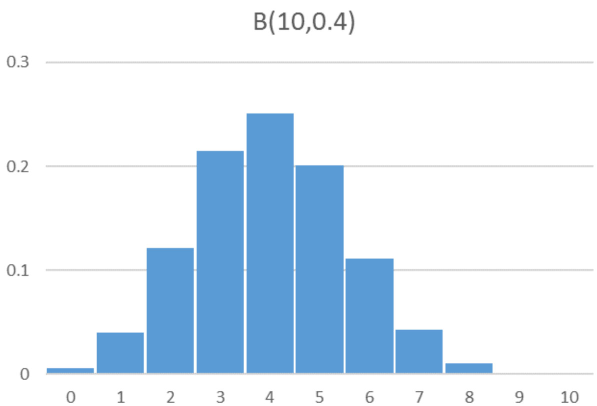
$$\mu = \mathbf{E}[X] = \sum_{k=0}^{n} k \binom{n}{k} p^k (1-p)^{n-k} = np$$

  - 分散

$$\mathbf{E}[X^2] = \sum_{k=0}^{n} k^2 \binom{n}{k} p^k (1-p)^{n-k} = np(1-p+np)$$

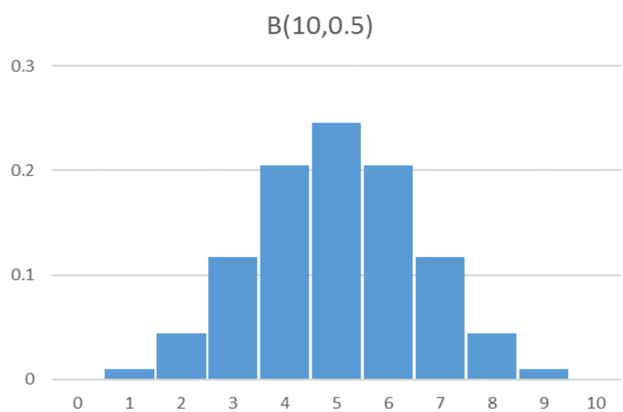
$$\sigma^2 = \mathbf{E}[X^2] - \boldsymbol{E}[X]^2 = np(1-p)$$

例 $B(10,0.4)$, $B(10,0.5)$, $B(10,0.65)$



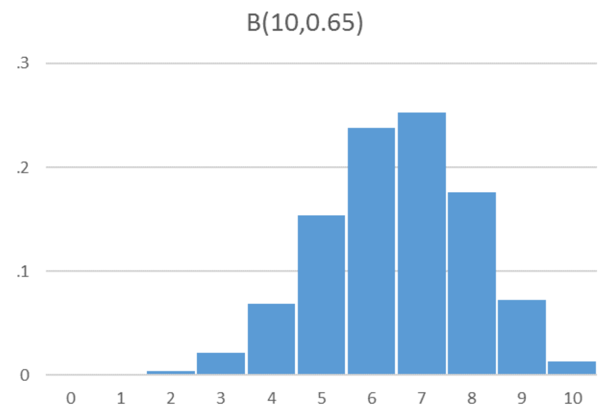
$\mu = 4$

$\sigma^2 = 2.4$



$\mu = 5$

$\sigma^2 = 2.5$

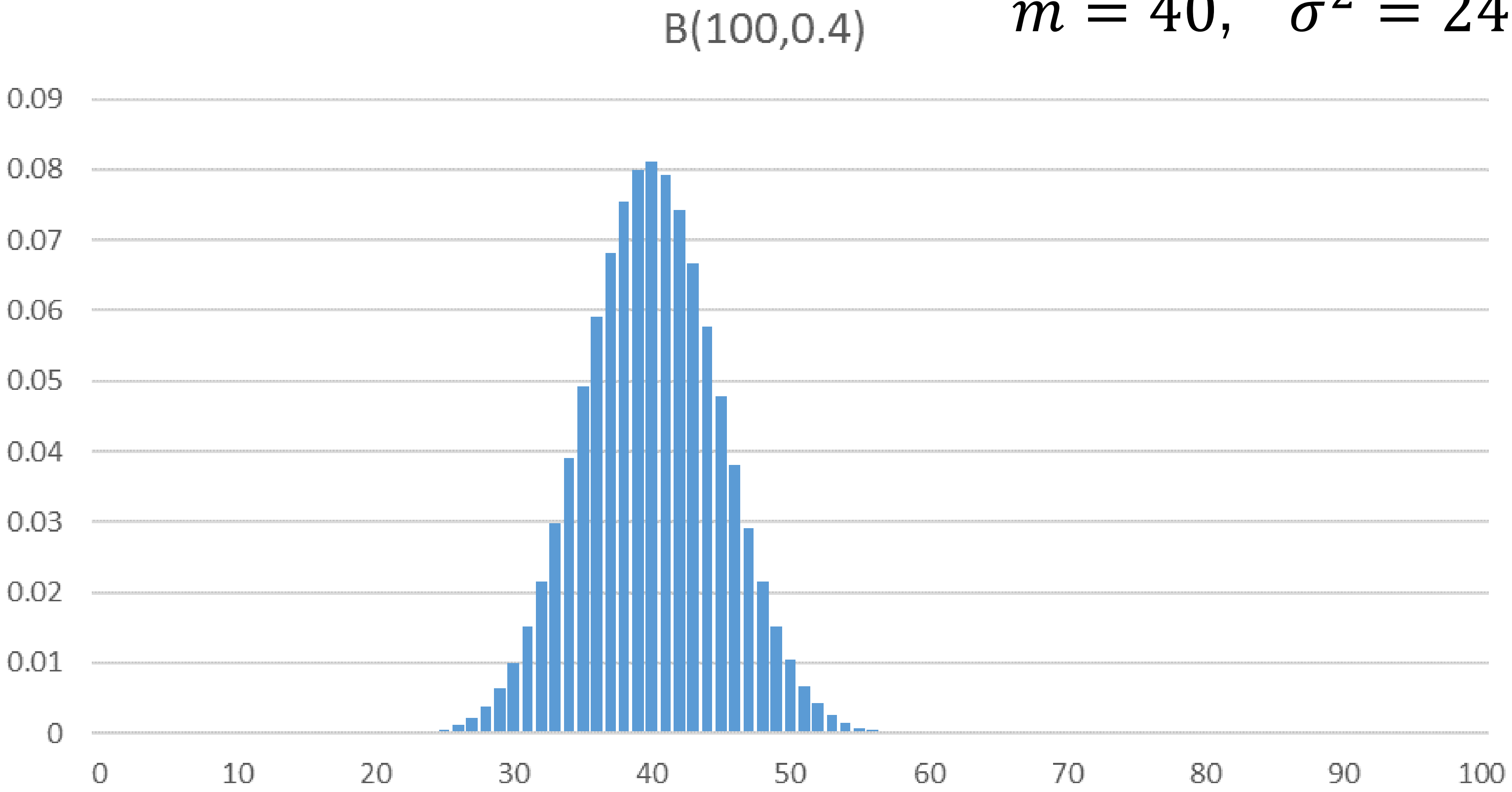


$\mu = 6.5$

$\sigma^2 = 2.275$

例 $B(100, 0.4)$

$m = 40, \quad \sigma^2 = 24$



例題 3.8

サイコロを 3 回投げて, 6 の目が 2 回出る確率を求めよ.

- サイコロを投げて6 の目が出たら成功（〇）, そうでないとき失敗（×）とみなす.

⇨ 成功確率 $p = \frac{1}{6}$ のベルヌーイ試行を 3 回行ったと考える.

- 成功回数 X は二項分布 $B(n,p) = B\left(3,\frac{1}{6}\right)$ に従う.　$X \sim B\left(3,\frac{1}{6}\right)$

- 求める確率は $P(X=2)$

$$P(X=2) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$= \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 = 3 \times \frac{5}{6^3} = \frac{15}{216}$$

付録：ポアソン分布

$\lambda > 0$ を定数とする.

確率変数 X がパラメータ λ のポアソン分布に従うとは,

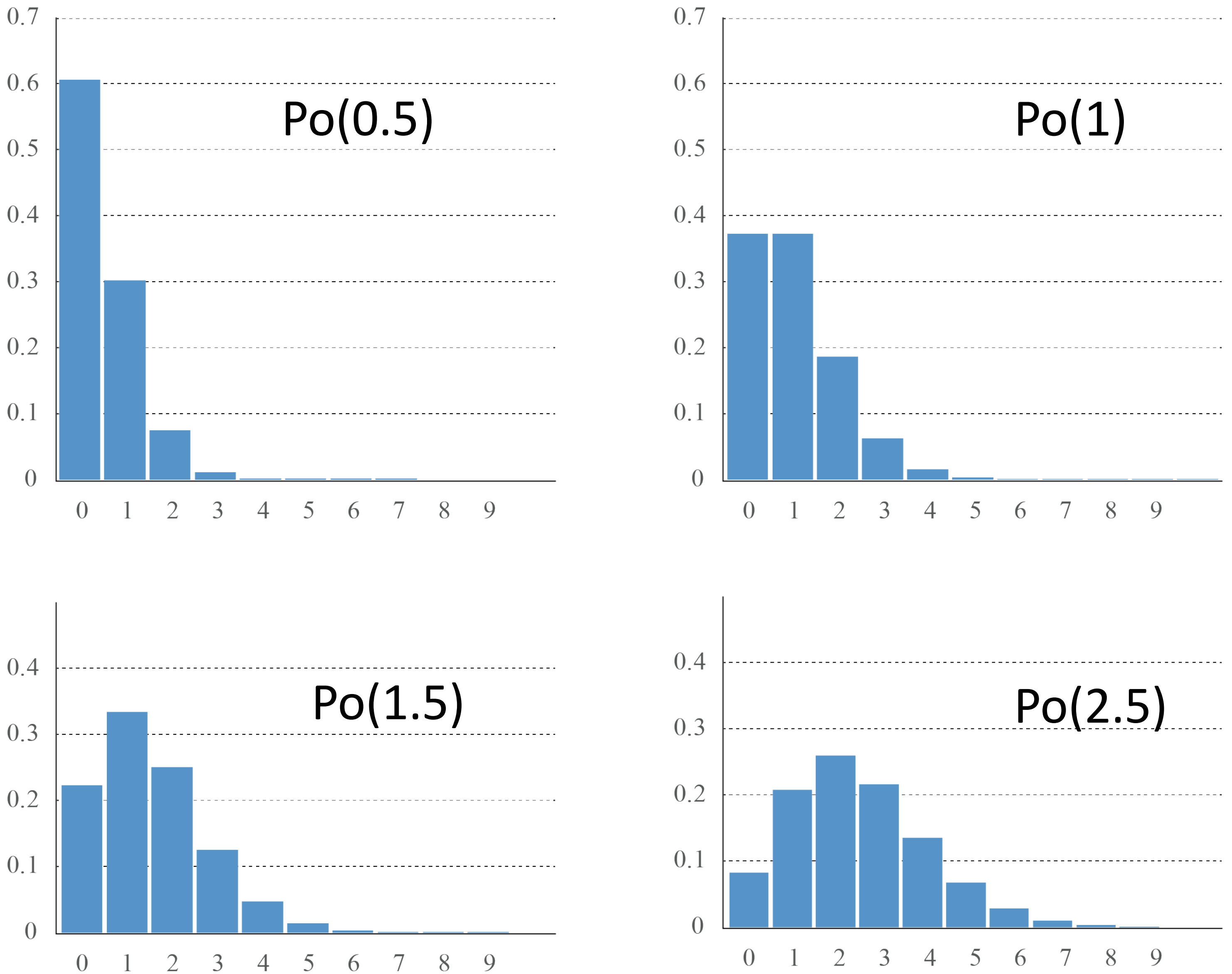
$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \qquad k = 0,1,2,\dots$$

$X \sim \mathrm{Po}(\lambda)$ と書く.

- 確かに確率分布になっている

指数関数のテーラー展開

$$e^{\lambda} = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} \quad \Longrightarrow \quad 1 = \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} e^{-\lambda}$$



0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0 1 2 3 4 5 6 7 8 9
Po(0.5)
Po(1)
Po(1.5)
Po(2.5)

• ポアソン分布 Po(λ) のまとめ

$$X \sim \mathrm{Po}(\lambda)$$

➢ 確率分布

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda} \qquad k=0,1,2,\ldots$$

➢ 平均値

$$\mu=\mathbf{E}[X]=\sum_{k=0}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\lambda$$

ポアソン分布の特徴

平均と分散が等しい

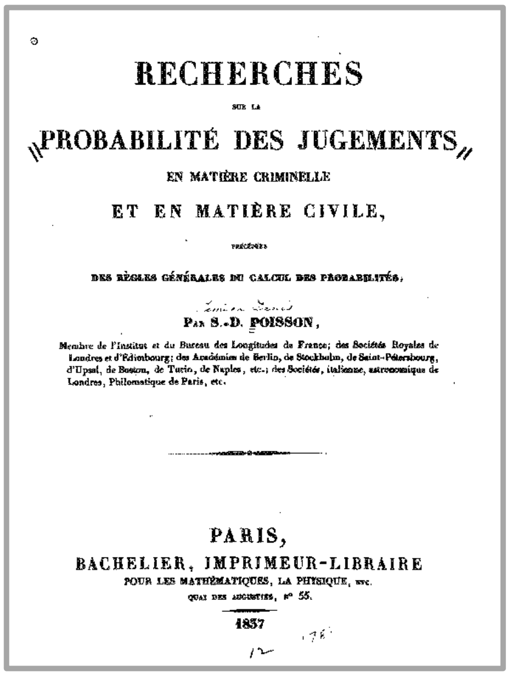
➢ 分散

$$\mathbf{E}[X^2]=\sum_{k=0}^{\infty}k^2\frac{\lambda^k}{k!}e^{-\lambda}=\lambda(\lambda+1)$$

$$\sigma^2=\mathbf{E}[X^2]-\boldsymbol{E}[X]^2=\lambda(\lambda+1)-\lambda^2=\lambda$$

Siméon Denis Poisson (1781-1840)

- エコール・ポリテクニクでラプラスらに学ぶ.
- 数学・物理学に多大な貢献
- ポアソン○○と名のついた概念多数



RECHERCHES

SUR LA

PROBABILITÉ DES JUGEMENTS

EN MATIÈRE CRIMINELLE

ET EN MATIÈRE CIVILE,

PRÉCÉDÉES

DES RÈGLES GÉNÉRALES DU CALCUL DES PROBABILITÉS;

PAR S.-D. POISSON,

Membre de l'Institut et du Bureau des Longitudes de France; des Sociétés Royales de Londres et d'Edimbourg; des Académies de Berlin, de Stockholm, de Saint-Pétersbourg, d'Upsal, de Boston, de Turin, de Naples, etc.; des Sociétés, italienne, astronomique de Londres, Philomatique de Paris, etc.

PARIS,

BACHELIER, IMPRIMEUR-LIBRAIRE

POUR LES MATHÉMATIQUES, LA PHYSIQUE, etc.

QUAI DES AUGUSTINS, n° 55.

1837

誤った有罪判決の回数について研究(1837)

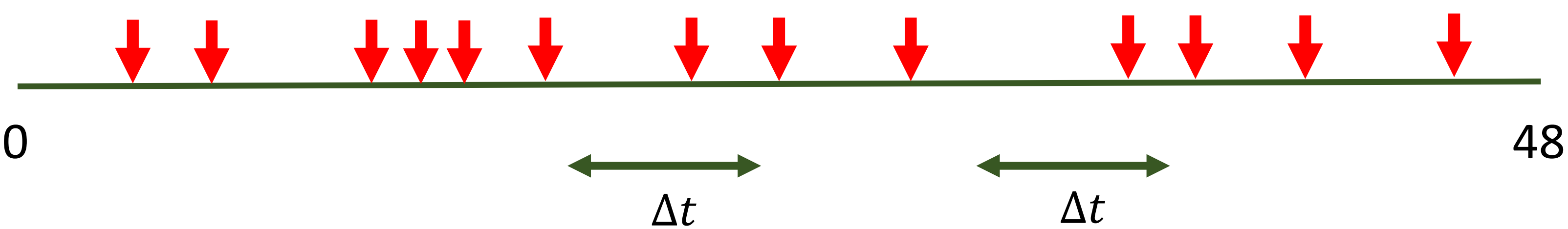
誤審が, ある一定期間に起こる回数 X の確率分布（⇒ ポアソン分布）を導出

稀なイベントの発生回数の確率分布として広い応用

- 馬に蹴られて死亡した兵士数（ボルトキーヴィッチ）
- 電話の呼び出し, メールの着信
- サッカーのゴール数, 野球のホームラン数

例 3.** メールの着信回数

メールの着信を観測して, 48時間に324回のメール着信があった. これをもとに, ある特定の１時間にメール着信が2回以下の確率を求めよ.

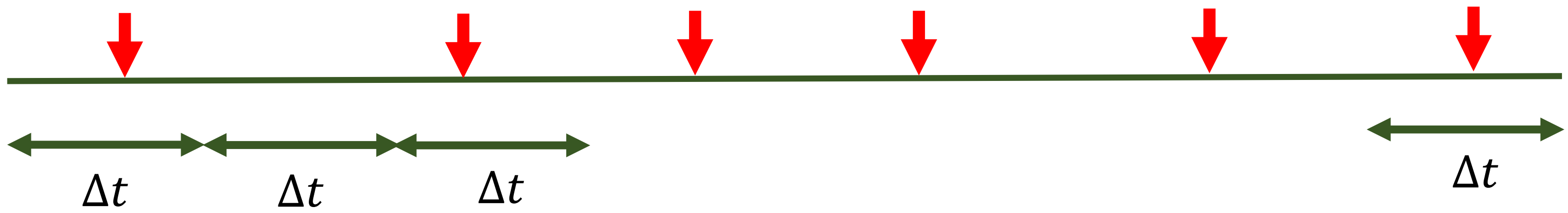


仮定

- 微小時間区間 Δt の中では着信は１回, または０回
- 微小時間区間 Δt がずれていれば着信の有無は独立

⇨ 微小時間区間 Δt 毎のメール着信をコイン投げとみなす

- 1時間当たりの着信回数 X をモデル化する



- 1時間を微小時間区間 Δt に N 分割

$$\Delta t = \frac{1}{N}$$

$$p = \mathrm{P}(\text{時間区間 } \Delta t \text{ にメール着信がある})$$

$$1 - p = \mathrm{P}(\text{時間区間 } \Delta t \text{ にメール着信がない})$$

- 1時間の間に N 回のコイン投げ

表の回数　$X \sim B(N, p)$

平均値 $\mu = Np$

- 観測から1時間当たり

$$\frac{324}{48} = 6.75 \quad \text{回の着信}$$

- p の推定

$$\mu = Np = 6.75$$

$$\Rightarrow \quad p = \frac{6.75}{N}$$

- こうして

$$X \sim B\left(N, \frac{6.75}{N}\right)$$

- ポアソン分布の導出（ポアソンの小数の法則）

$B\left(N, \frac{\lambda}{N}\right)$ は $N \to \infty$ とすると $\mathrm{Po}(\lambda)$ に収束する．

言い換え

$B(N, p)$ は, $N \to \infty, p \to 0, Np \to \lambda$ とすると $Po(\lambda)$ に収束する．

$X \sim B\left(N, \frac{\lambda}{N}\right)$ として

$$P(X = k) = \binom{N}{k} p^k (1 - p)^{N-k}$$

$$= \frac{N(N-1)\cdots(N-k+1)}{k!}\left(\frac{\lambda}{N}\right)^k \left(1 - \frac{\lambda}{N}\right)^N \left(1 - \frac{\lambda}{N}\right)^{-k}$$

$$\to \frac{\lambda^k}{k!} e^{-\lambda} \qquad (N \to \infty)$$

例 3.** メールの着信回数

メールの着信を観測して, 48時間に324回のメール着信があった. これをもとに, ある特定の１時間にメール着信が2回以下の確率を求めよ.

- X：1 時間当たりの着信回数
- 1時間を微小時間区間 Δt に N 分割してモデル化

$$X \sim B\left(N, \frac{6.75}{N}\right)$$

N は人為的なので消去したい

- **$N \to \infty$ として** $\boldsymbol{X \sim \mathrm{Po}(6.75)}$
- 答

$$P(X \leq 2) = \frac{6.75^0}{0!} e^{-6.75} + \frac{6.75^1}{1!} e^{-6.75} + \frac{6.75^2}{2!} e^{-6.75}$$

$$= (1 + 6.75 + 22.78) \times 0.00117 = 0.036$$

宿題

演習問題 3 (page 61-63)

3.2 1の目が出るまでサイコロを振るとき，次の値を求めよ．

（1）第６投目に初めて１の目が出る確率.

（2）初めて１の目が出る確率が少なくとも $\frac{1}{2}$ であるために必要な試行の回数

3.4 次の値を確率変数 X の平均 μ，分散 σ^2 を用いて表せ.

（1） $E\{X(X-1)\}$　　（2） $E\{X(X+5)\}$

3.5 確率変数 X の平均を μ，分散を σ^2 とするとき，$M(a) = E\{(X-a)^2\}$ を最小にする a の値は μ であり，その最小値は σ^2 であることを示せ.

3.6 X は正の整数値をとる確率変数であり，平均値が存在するとき，

$$E(X) = \sum_{x=1}^{\infty} P(X \geq x)$$

が成り立つことを示せ.

3.7 製薬会社が，ある病気にかかっている 20 人の患者に新しい薬を服用させた．もし，その薬が患者の各人を治す確率が 0.15 であり，ある患者に対する結果が他の患者に対する結果と独立ならば，20 人の患者のうち 3 人以上が治る確率はいくらか．

◆ **二項分布を用いて厳密値を求めよ. 次に, ポアソン分布による近似計算を行い, 値を比較せよ.**

3.11 標準偏差 σ を平均 μ で割ったものを**変動係数** (coefficient of variation) という．ポアソン分布で変動係数が 2 であるとき，その平均 μ を求めよ．

Lecture 5

平均値と分散

【教科書】

第 3 章 確率変数と確率分布

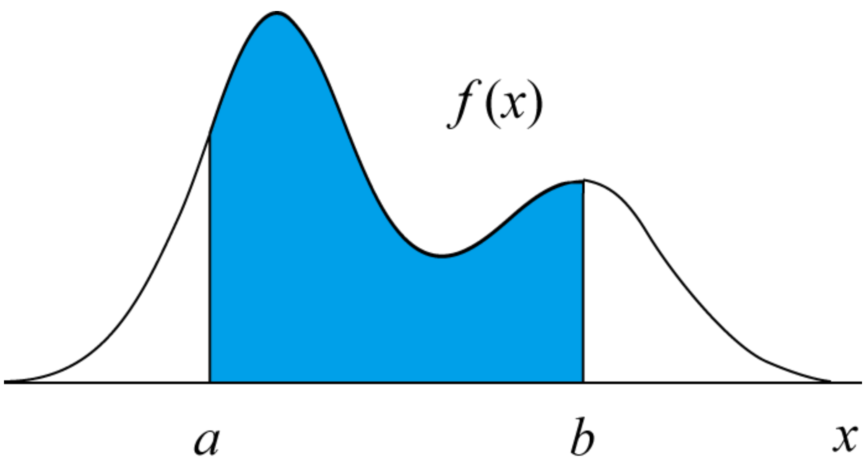
4 連続型確率変数の分布

- 連続型確率変数 X

< でも ≤ でも同じ

  - 特定の値 a をとる確率： $P(X = a) = 0$
  - ある範囲（a より大きく b 以下）の値をとる確率： $P(a < X \leq b)$

- 確率密度関数を用いて面積で確率を表す



基本的な性質

(1)　$f(x) \geq 0$

(2) $\int_{-\infty}^{\infty} f(x)dx = 1$

$$P(a \leq X \leq b) = \int_a^b f_X(x)dx$$

- 分布関数 (distribution function)
  ＝累積分布関数 (cumulative distribution function)

確率変数 X が x 以下の値をとる事象 $\{X \leq x\}$ の確率は，x の関数である：

$$F(x) = P(X \leq x) \quad -\infty < x < \infty$$

これを X の（累積）分布関数という

基本的な性質

(1) 単調増加関数である：$x \leq y$ ならば $F(x) \leq F(y)$

(2) 右側連続である ： $\lim_{\epsilon \to +0} F(x+\epsilon) = F(x)$

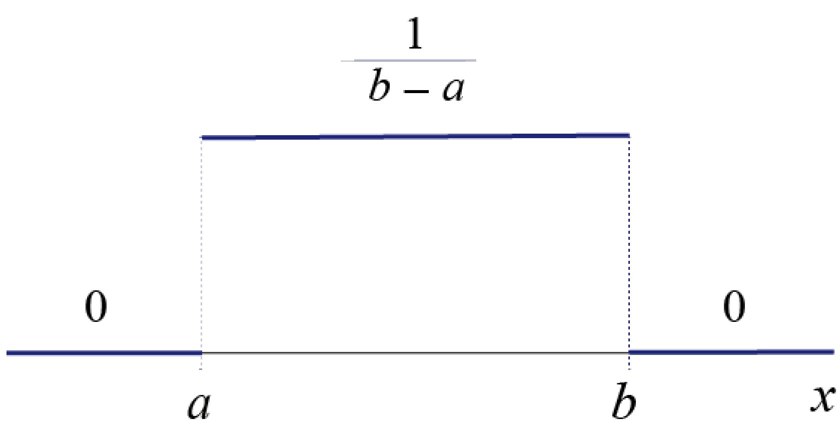
(3) $0 \leq F(x) \leq 1.\ F(-\infty) = 0,\ \mathrm{F}(\infty) = 1$

- 離散型の場合： 階段関数となる
- 連続型の場合： 分布関数は滑らかな曲線となる

$$f(x) = \frac{d}{dx}F(x) = \frac{d}{dx}\int_{-\infty}^{x} f(t)dt$$

例題 3.3

密度関数 $f(x)$ が区間 $[a, b]$ で与えられ, その値が一定であるような分布を**一様分布**という.



確率密度関数

$$f(x) = \begin{cases} \dfrac{1}{b-a}, & a \leq x \leq b \\ 0, & \text{その他} \end{cases}$$

区間 $[a, b]$ から1点を, どの点も同程度の確からしさで選ぶとき, 選ばれた点の座標を X とする.

X は連続型確率変数となり, その分布が一様分布

例題 3.7

確率変数 X が区間 $[a, b]$ 上の一様分布に従うとき，その平均と分散は次のようになることを示せ.

$$E[X] = \frac{a+b}{2} \qquad V[X] = \frac{(b-a)^2}{12}$$

注意　特定の点 c が選ばれる確率：$P(X = c) = 0$

平均値

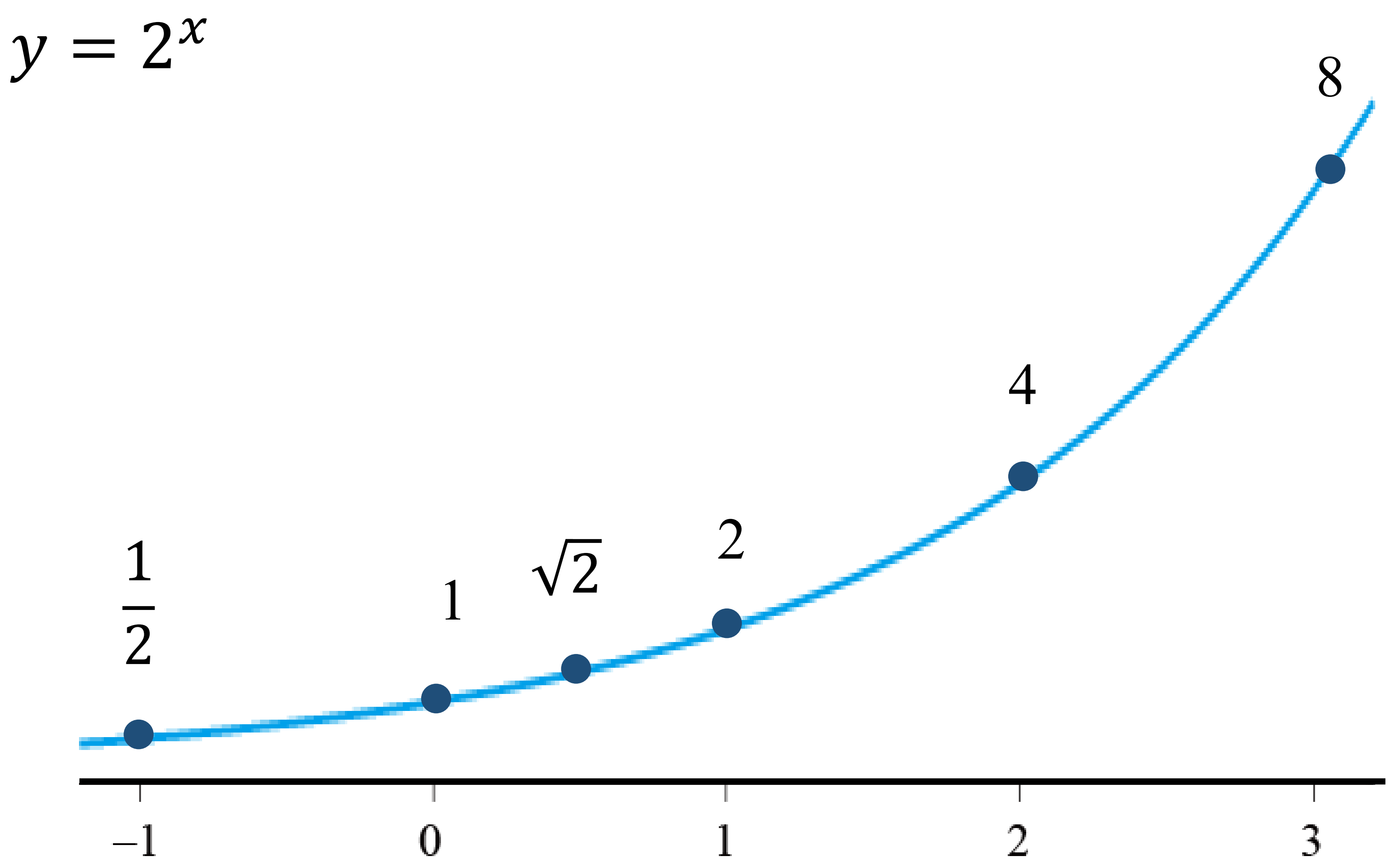
$$\mu = \mathbf{E}[X] = \int_a^b x \frac{dx}{b-a} = \frac{a+b}{2}$$

$$\mathbf{E}[X^2] = \int_a^b x^2 \frac{dx}{b-a} = \frac{a^2+ab+b^2}{3}$$

分散

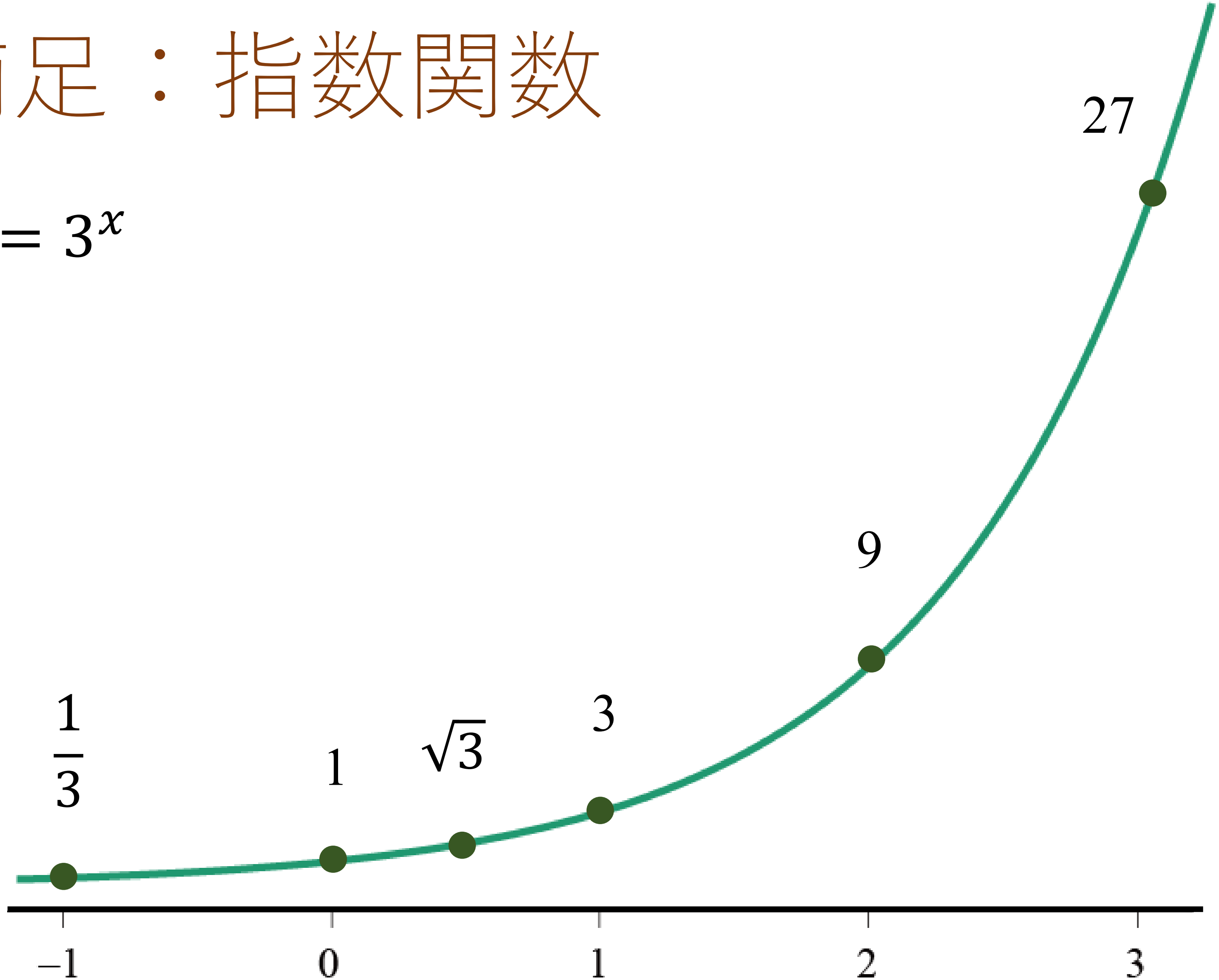
$$\sigma^2 = \mathbf{E}[X^2] - \boldsymbol{E}[X]^2 = \frac{(b-a)^2}{12}$$

補足：指数関数

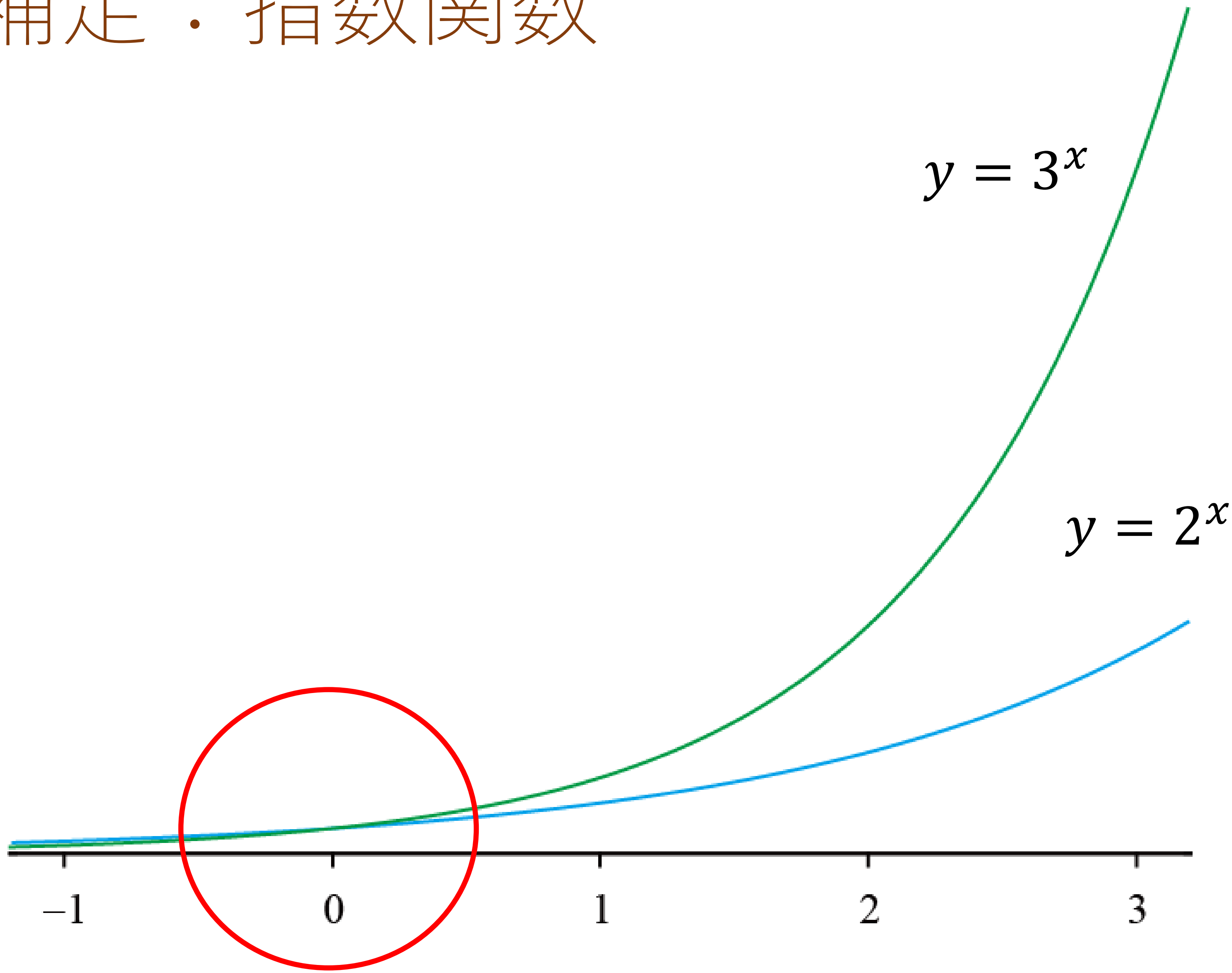


補足：指数関数

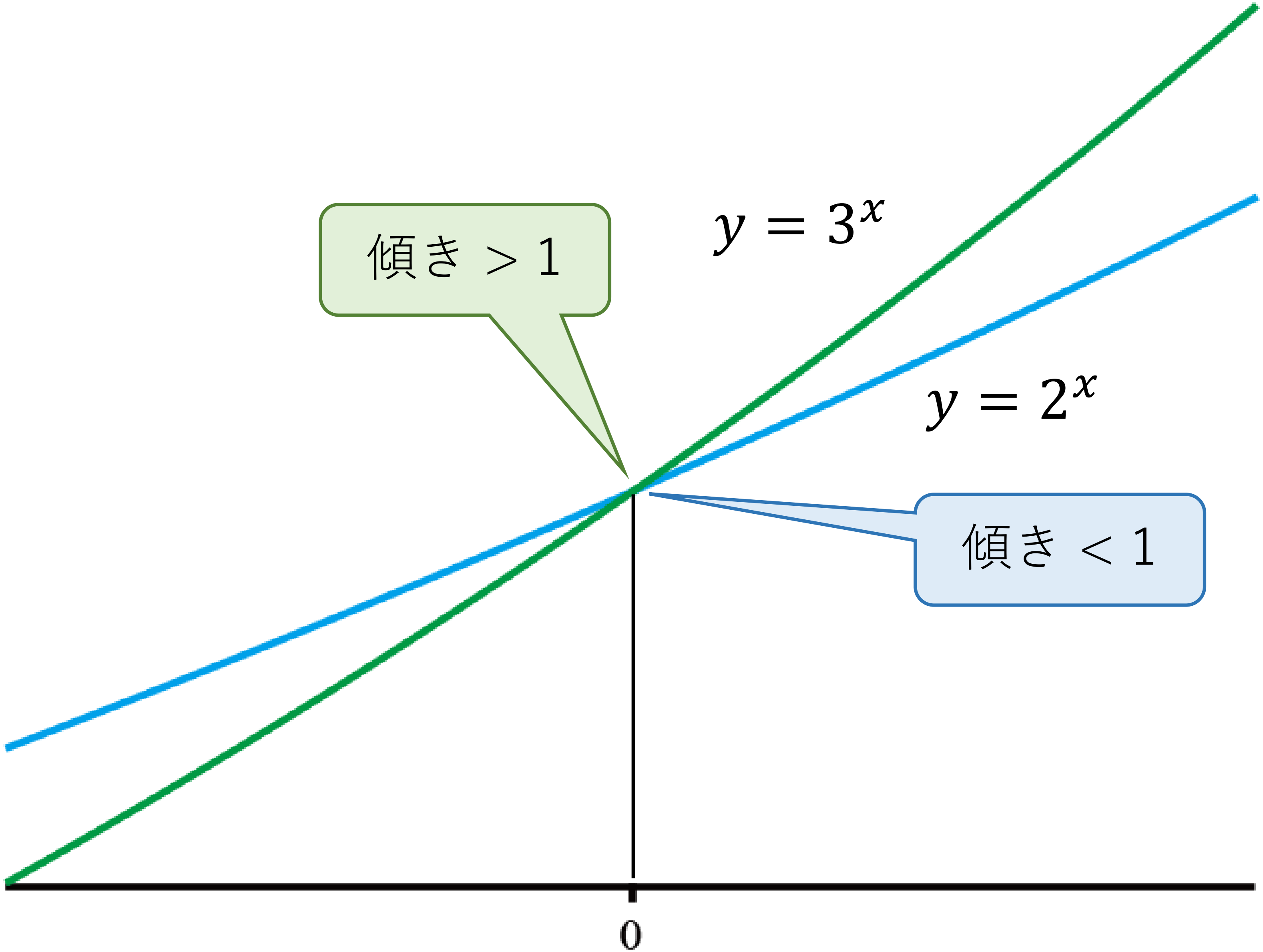
$y = 3^x$



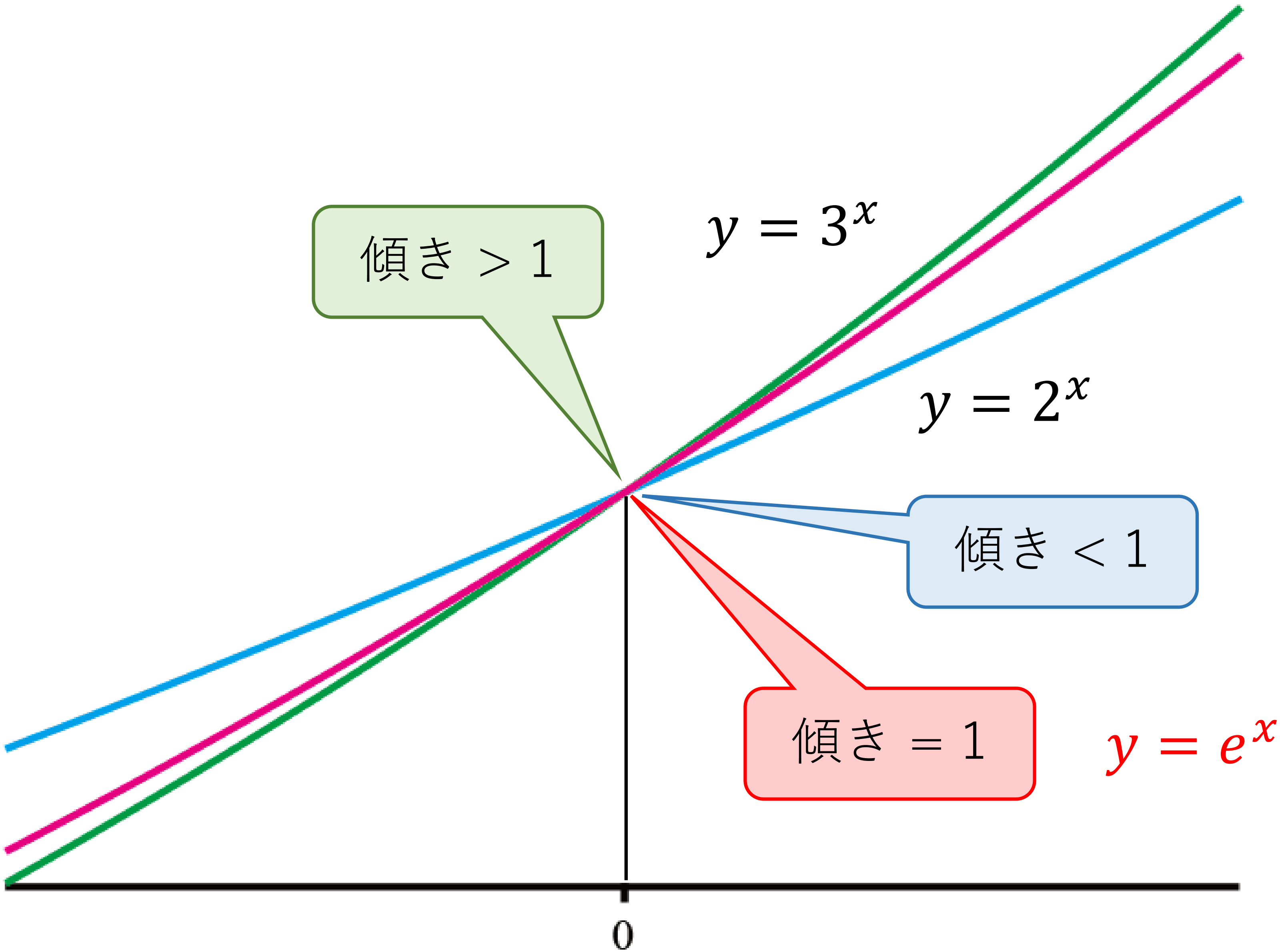
補足：指数関数



補足：指数関数

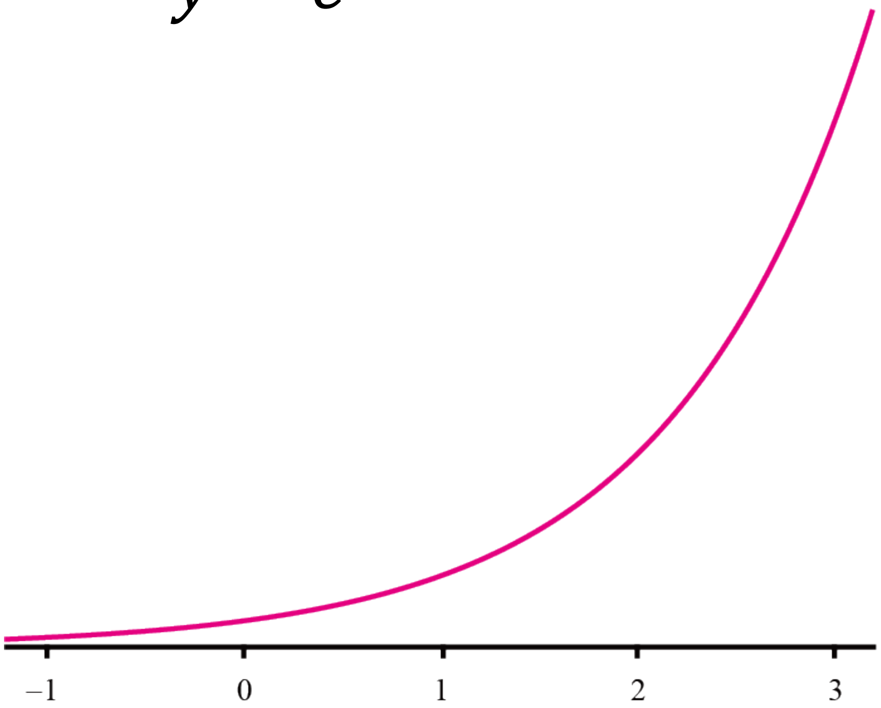


補足：指数関数



補足：指数関数

$$y = e^x$$



$e = 2.718\ldots$ （無理数）

e の呼び名

- 自然対数の底
- ネピアの数

テーラー展開

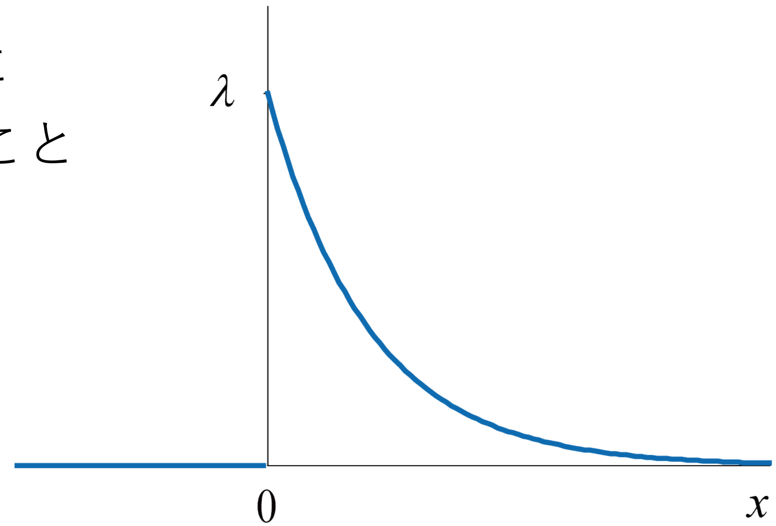
$$e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$$

指数法則

$$e^{x+y} = e^x e^y$$

- 指数分布

$\lambda > 0$ を定数とする.
確率変数 X がパラメータ λ の指数分布に
従うとは, X が次の確率密度関数をもつこと



密度関数

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geq 0 \\ 0, & \text{その他} \end{cases}$$

平均値

$$\mu = \mathbf{E}[X] = \int_0^\infty x\,\lambda e^{-\lambda x} dx = \frac{1}{\lambda}$$

$$\mathbf{E}[X^2] = \int_0^\infty x^2 \lambda e^{-\lambda x} dx = \frac{2}{\lambda^2}$$

計算は例題 3.9を参照

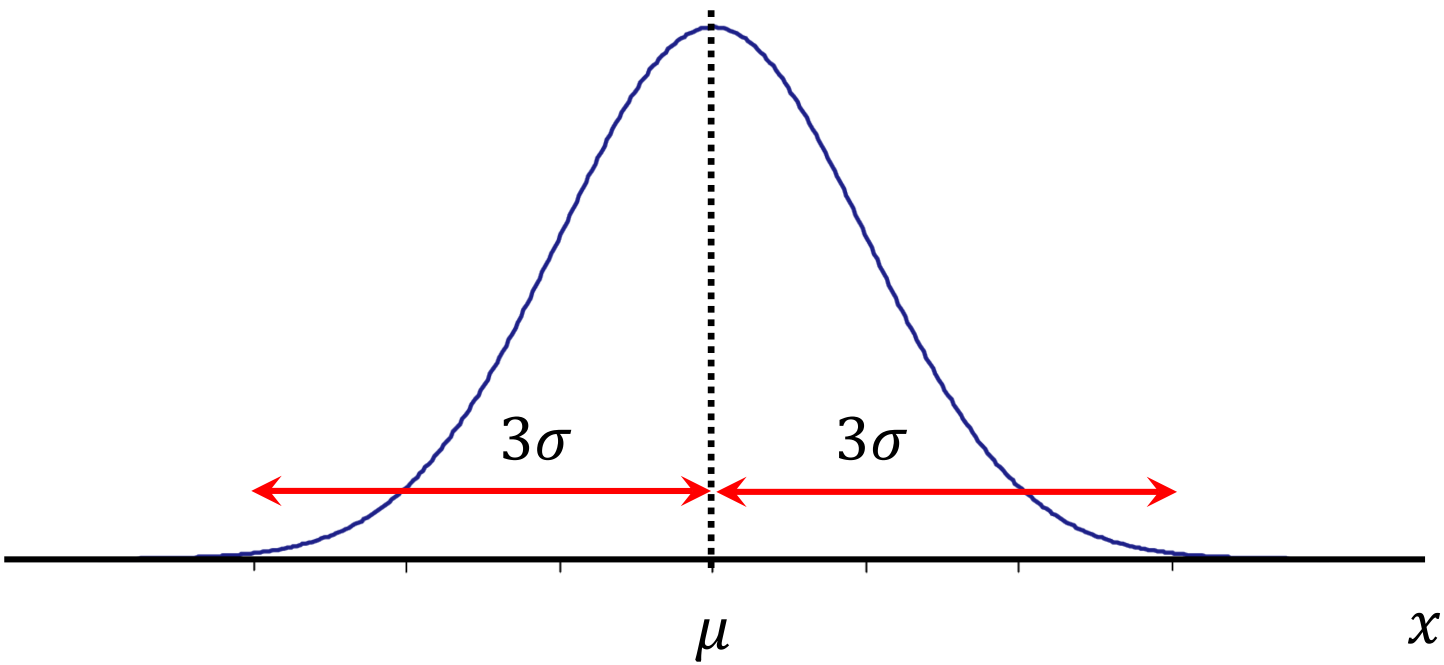
分散

$$\sigma^2 = \mathbf{E}[X^2] - \boldsymbol{E}[X]^2 = \frac{1}{\lambda^2}$$

- 正規分布＝ガウス分布 $N(\mu, \sigma^2)$

密度関数

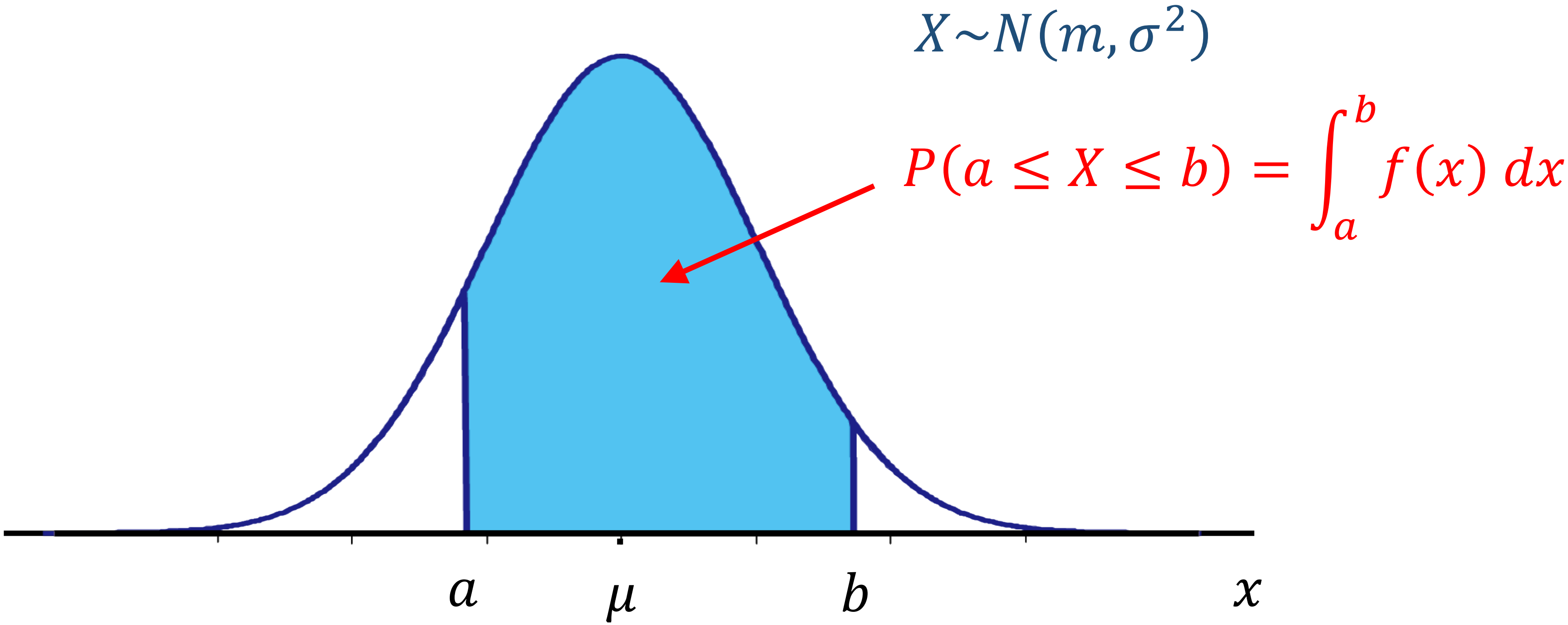
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



- 正規分布＝ガウス分布 $N(\mu, \sigma^2)$

密度関数

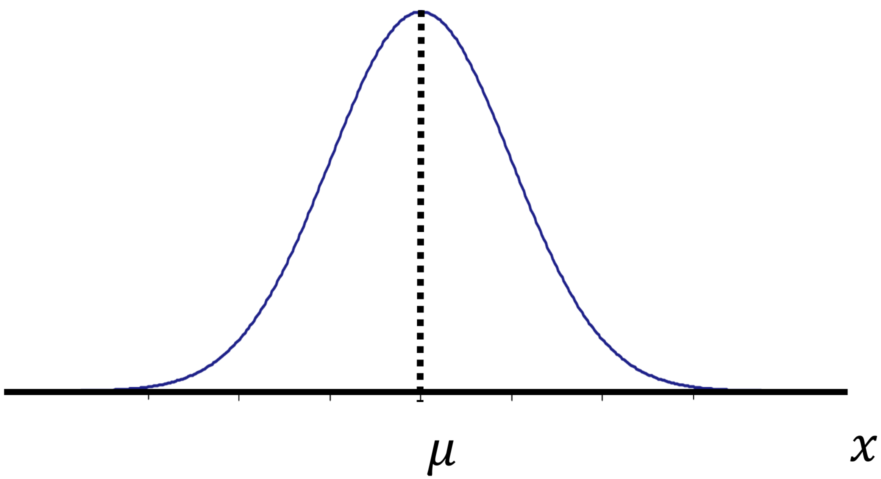
$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



- 正規分布＝ガウス分布 $N(\mu, \sigma^2)$

密度関数

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



$N(\mu, \sigma^2)$

平均値	μ
分散	σ^2

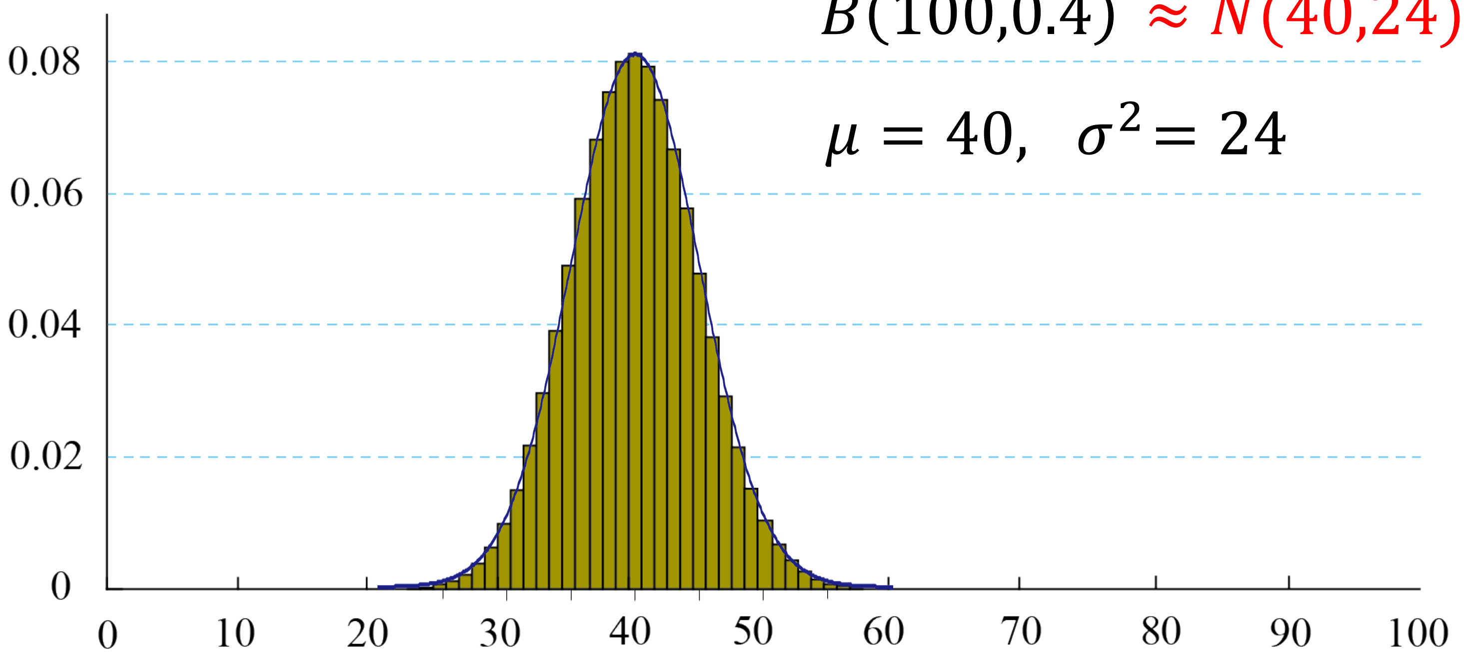
- ド・モアブルーラプラスの定理

二項分布 $B(n,p)$ は n が大きいとき,
同じ平均と分散をもつ正規分布に近い.

$$B(n,p) \approx N(\mu,\sigma^2) \qquad \mu = np, \quad \sigma^2 = np(1-p)$$

$$B(100,0.4) \approx N(40,24)$$

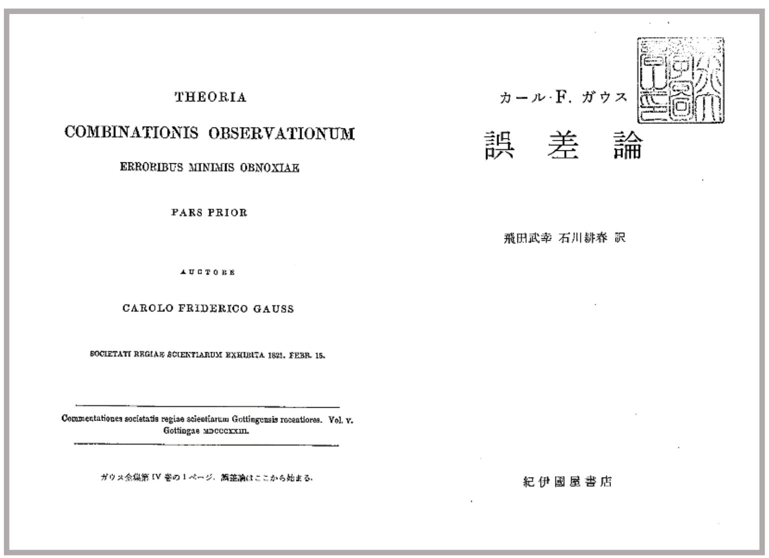
$$\mu = 40, \quad \sigma^2 = 24$$



Johann Carl Friedrich Gauss (1777-1855)

19世紀以降の近代数学のほとんどの分野に影響を与えている：

- 素数定理、平方剰余の相互法則
- 正十七角形の作図、代数学の基本定理、円分多項式(代数学)
- 最小二乗法(統計)
- 複素数、ガウス平面、複素関数論
- 電磁気などの物理学



THEORIA

COMBINATIONIS OBSERVATIONUM

ERRORIBUS MINIMIS OBNOXIAE

PARS PRIOR

AUCTORE

CAROLO FRIDERICO GAUSS

SOCIETATI REGIAE SCIENTIARUM EXHIBITA 1821. FEBR. 15.

Commentationes societatis regiae scientiarum Gottingensis recentiores. Vol. V. Gottingae MDCCCXXIII.

ガウス全集第 IV 巻の 1 ページ。誤差論はここから始まる。

カール・F. ガウス

誤　差　論

飛田武幸 石川耕春 訳

紀伊國屋書店

1807年～終生
ゲッティンゲンの天文台長（初代）

1809年 『天体運行論』
最小二乗法を用いたデータ補正、正規分布

1821-26年 『誤差を最小にする観測』

- 標準正規分布 (standard normal distribution)

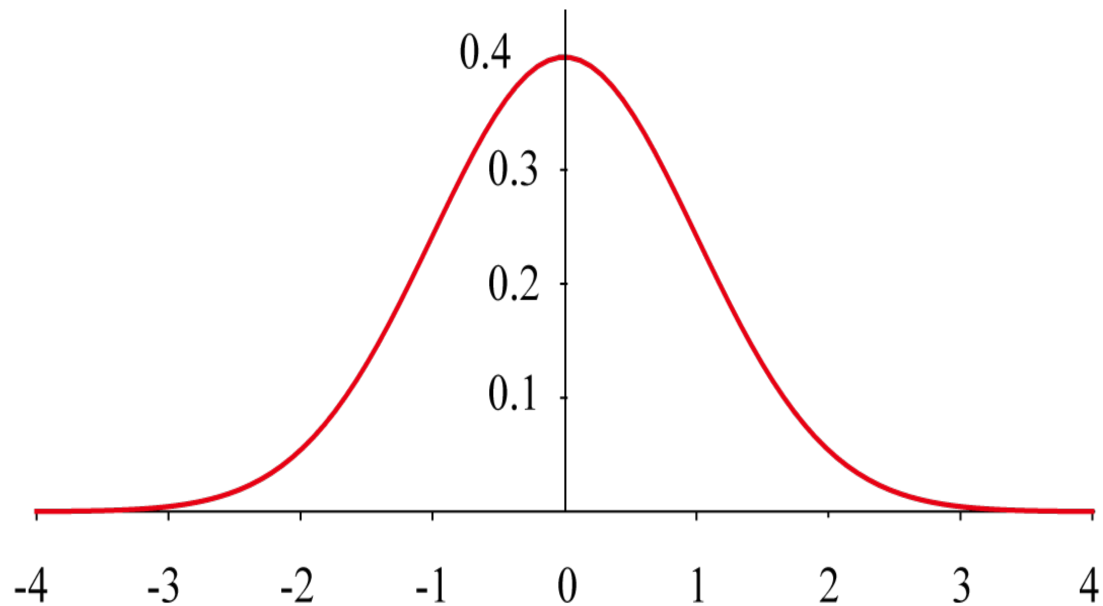
$$N(0,1)$$

平均値　$\mu = 0$

分散　$\sigma^2 = 1$

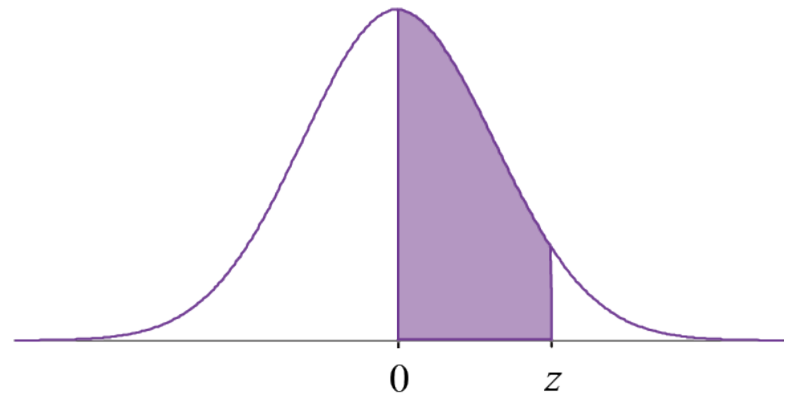
密度関数

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



標準正規分布表

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1311	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990



$$Z \sim N(0,1)$$

$$P(0 \le Z \le z) = \int_0^z \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx$$

例

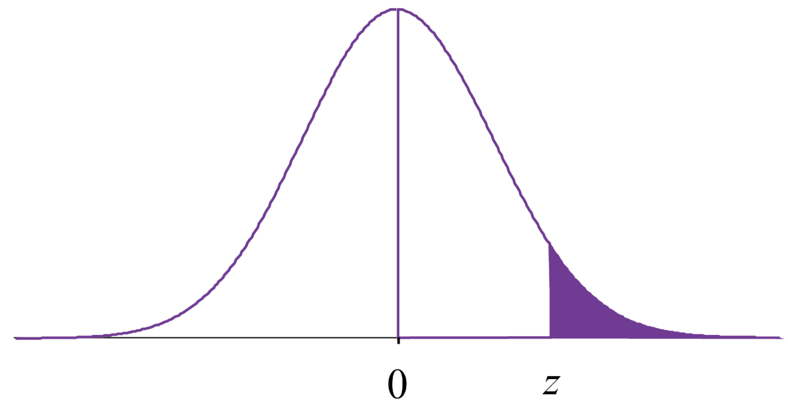
$$P(0 \le Z \le 1.15) = 0.3749$$

標準正規分布表

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1311	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

例

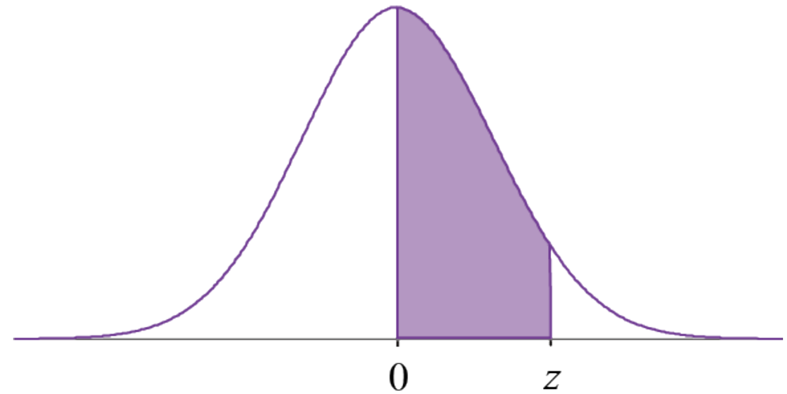
$P(Z \geq 2.18)$



$$\begin{aligned} P(Z \geq 2.18) &= 0.5 - P(0 \leq Z \leq 2.18) \\ &= 0.5 - 0.4854 \\ &= 0.0146 \end{aligned}$$

標準正規分布表

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1311	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990



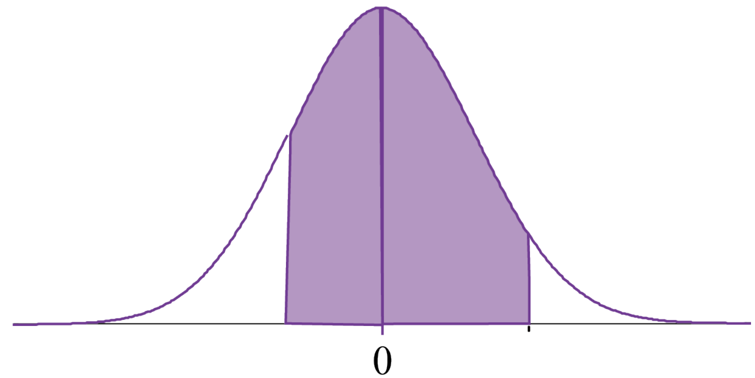
練習(5分)

$$P(-1.23 \leq Z \leq 2.06)$$

$$P(Z \leq -1.34)$$

標準正規分布表

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1311	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990



$$P(-1.23 \leq Z \leq 2.06)$$

$$= P(-1.23 \leq Z \leq 0)$$

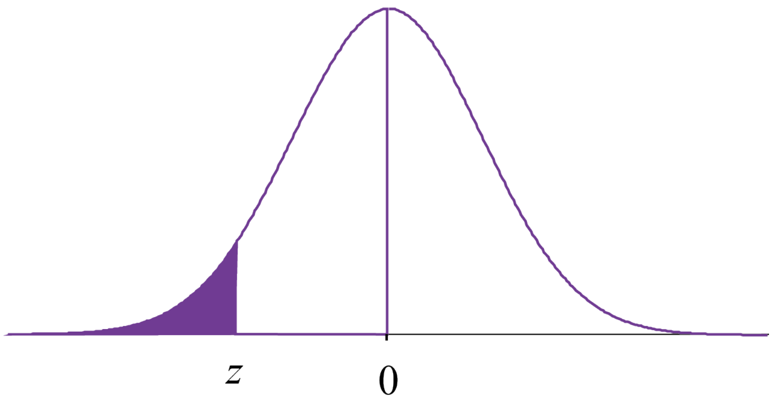
$$+P(0 \leq Z \leq 2.06)$$

$$= 0.3907 + 0.4803$$

$$= 0.8710$$

標準正規分布表

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1311	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990



$$P(Z \leq -1.34)$$

$$= 0.5 - 0.4099$$

$$= 0.0901$$

定理 3.2

確率変数 X が正規分布 $N(\mu,\sigma^2)$に従うとき, その 1 次変換

$Y = a + bX$ (a ,b は定数)も正規分布 $N(a + b\mu\,, b^2\sigma^2)$ に従う.

$$X \sim N(\mu,\sigma^2) \Rightarrow aX + b \sim N(a + b\mu\,, b^2\sigma^2)$$

標準化

$$X \sim N(\mu,\sigma^2) \Rightarrow Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

例

$X \sim N(2,5^2)$ を標準化すると

$$Z = \frac{X-2}{5} \sim N(0,1)$$

例題 3.10　X が正規分布 $N(-1, 2^2)$ に従うとき,

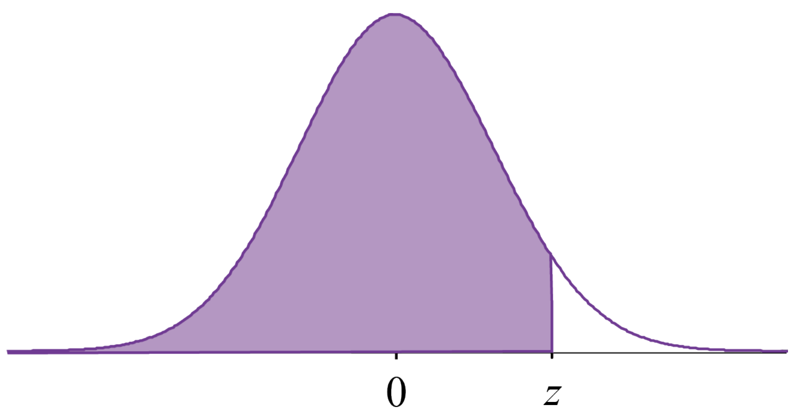
（1）$P(X \leq 2.29)$ を求めよ.

（2）$P(X > x) = 0.01$ であるような x の値を求めよ.

標準化　$X \sim N(\mu, \sigma^2)$ ⇒ $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$

（1）

$$P(X \leq 2.29) = P\left(\frac{X-(-1)}{2} \leq \frac{2.29-(-1)}{2}\right)$$



$$= P(Z \leq 1.645)$$

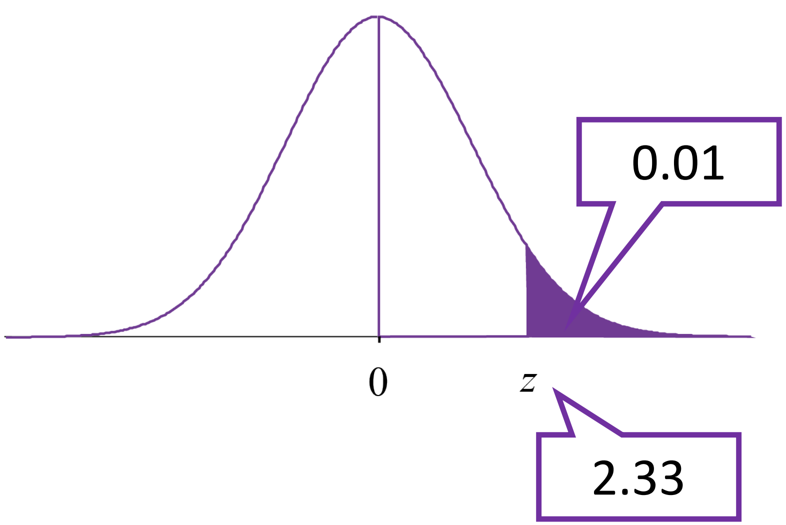
$$= 0.5 + 0.45 = 0.95$$

例題 3.10　X が正規分布 $N(-1, 2^2)$ に従うとき，

(1) $P(X \leq 2.29)$ を求めよ．
(2) $P(X > x) = 0.01$ であるような x の値を求めよ．

標準化　$X \sim N(\mu, \sigma^2)$ ➡ $Z = \dfrac{X-\mu}{\sigma} \sim N(0,1)$

(2)　$$P(X > x) = P\left(\frac{X-(-1)}{2} > \frac{x-(-1)}{2}\right) = P\left(Z > \frac{x+1}{2}\right)$$

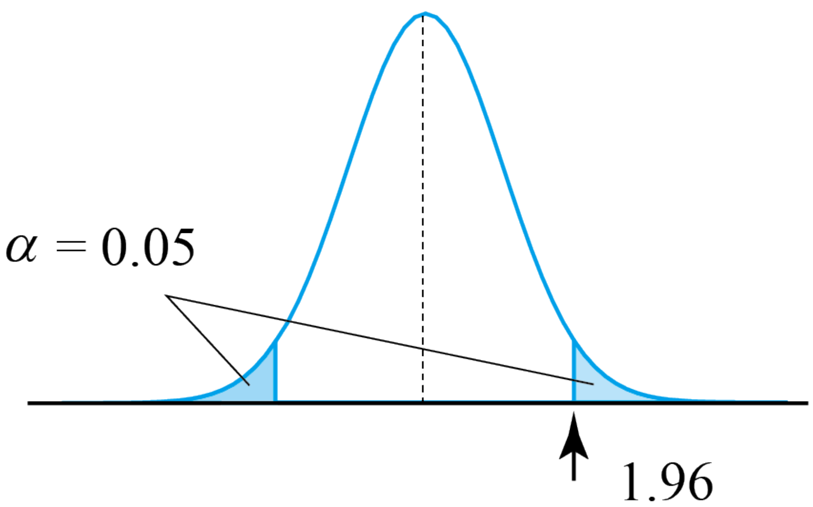


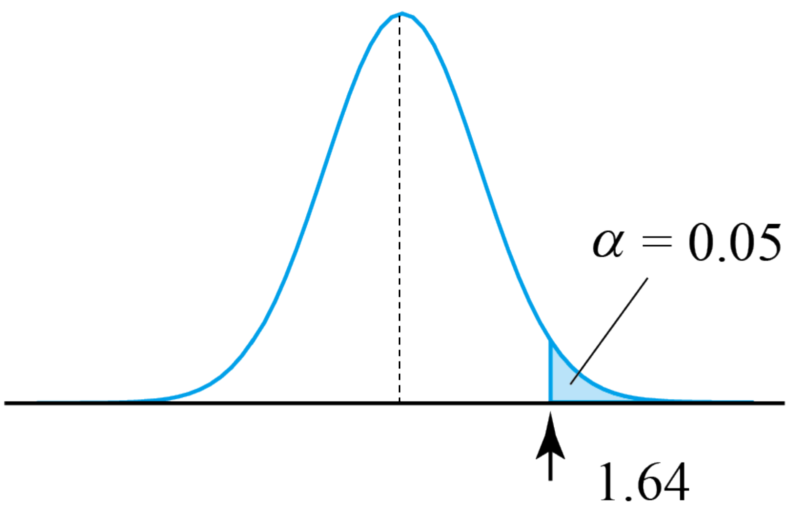
$$P(Z > 2.33) = 0.01$$

$$\frac{x+1}{2} = 2.33$$

$$x = 3.66$$

両側 α 点と上側 α 点





1.96 = 両側 5 % 点

= 上側 2.5 % 点

1.64 = 上側 5 % 点

= 両側 10 % 点

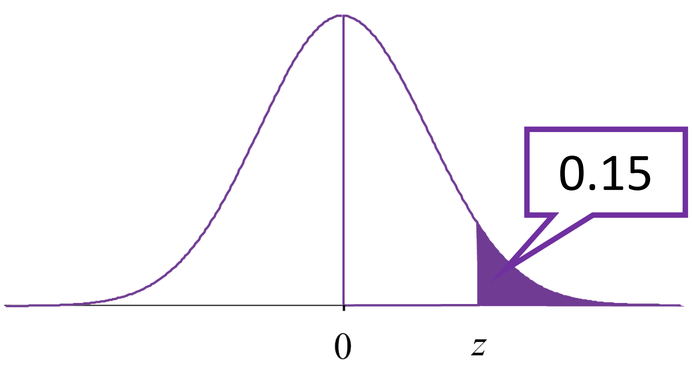
例題 3.11

大学 1 年生の統計学の試験（100 点満点）で，上位 15% の学生に成績「優」をつけようと思う．何点以上とすればよいか. ただし, 試験の得点は, 平均 55 点，標準偏差 15 点の正規分布 $N(55, 15^2)$ に従うものとする.

試験の得点を X とすると，$X \sim N(55, 15^2)$

求める得点を x とすると,題意より $P\,(X > x) = 0.15$

$$0.15 = P(X > x) = P\left(\frac{X-55}{15} > \frac{x-55}{15}\right) = P\left(Z > \frac{x-55}{15}\right)$$



$$P(Z > 1.04) = 0.15$$

$$\frac{x-55}{15} = 1.04 \quad \Rightarrow \quad x = 70.6$$

例題 3.12

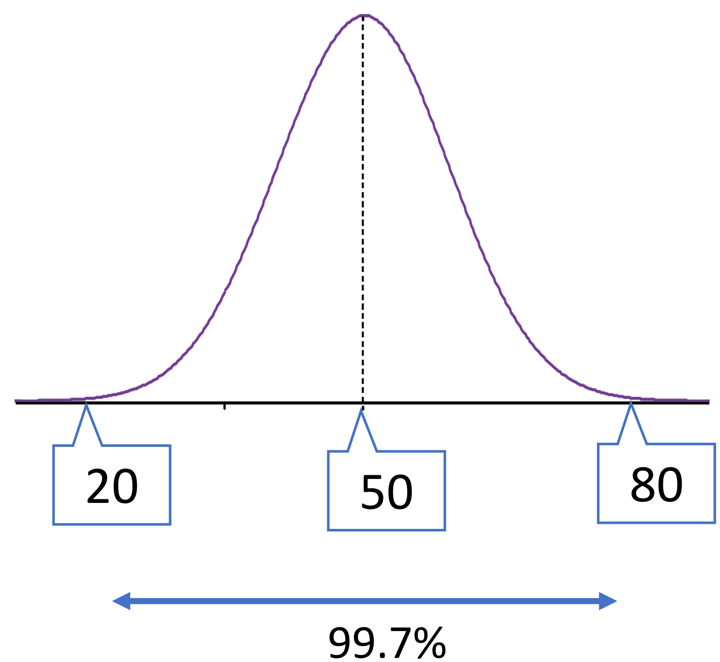
学力の偏差値は, テストの得点 X を正規分布 $N(50, 10^2)$ に従うように 1 次変換したものである. 例題 3.11 の統計学の試験の得点 X を偏差値 Y に変換する式を作れ.

例題 3.11 の統計学の試験の得点　$X \sim N(55, 15^2)$

z 変換は $Z = \dfrac{X-55}{15} \sim N(0,1)$

偏差値 Y は

$$
\begin{aligned}
Y &= 50 + 10Z \\
&= 50 + 10 \times \frac{X-55}{15} \\
&= \frac{2X+40}{3}
\end{aligned}
$$



宿題

演習問題 3（続き） (page 61-63)

3.1　確率変数 X が一様分布 $U(0, 2)$ に従うとき，次の確率を求めよ．

（1）　$P(X > 0.5)$　　（2）　$P(X < 1.2)$　　（3）　$P(X > 1.5 \mid X > 0.5)$

3.3 半径 R の円板内からランダムに点を選ぶ．X をこの選ばれた点と円板の中心との距離とする．

（1）　X の分布関数と密度関数を求めよ．

（2）　X の平均，分散を求めよ．

3.8　底辺の長さ l，高さ h の三角形の内部よりランダムに点を選ぶ．X をこの選ばれた点と底辺との距離を表すとする．

（1）　X の分布関数と密度関数を求めよ．

（2）　X の平均，分散を求めよ．

3.9 確率変数 X は分布関数

$$F(x) = \begin{cases} 0 & (x \leq 0) \\ \dfrac{x}{3} & (0 \leq x < 1) \\ \dfrac{x}{2} & (1 \leq x < 2) \\ 1 & (x \geq 2) \end{cases}$$

をもつとする．そのとき，次の確率を計算せよ．

(1) $P\left(\dfrac{1}{2} \leq X \leq \dfrac{3}{2}\right)$ (2) $P\left(\dfrac{1}{2} \leq X \leq 1\right)$ (3) $P\left(\dfrac{1}{2} \leq X < 1\right)$

(4) $P\left(1 \leq X \leq \dfrac{3}{2}\right)$ (5) $P(1 < X < 2)$

3.10 確率変数 X は分布関数

$$F(x) = \begin{cases} 0 & (x \leq 0) \\ \dfrac{x^2}{R^2} & (0 \leq x \leq R) \\ 1 & (x > R) \end{cases}$$

をもつとする．

（1） 密度関数 $f(x)$ を求めよ．また，その平均と分散を求めよ．

（2） $Y = X^2$ の密度関数を求めよ．Y は何分布に従うか．

$\boxed{\textbf{3.12}}$ 確率変数 X が正規分布 $N(\mu, \sigma^2)$ に従うとする．

（1） $\mu = 5$，$\sigma = 2$ のとき，$P(X \leq 7)$ を求めよ．

（2） $P(X \leq 6) = 0.9772$，$P(X \leq 4) = 0.8413$ であるとき，μ と σ の値を求めよ．

3.13 知能指数 IQ は正規分布 $N(100, 15^2)$ に従う．IQ が 150 以上の人は全体の何パーセントを占めるか．

3.14 料金徴収所での客の到着時間間隔 X（分）は指数分布に従い，平均の到着時間間隔は $\frac{1}{2}$ 分とする．

（1） X の確率密度関数を書け．

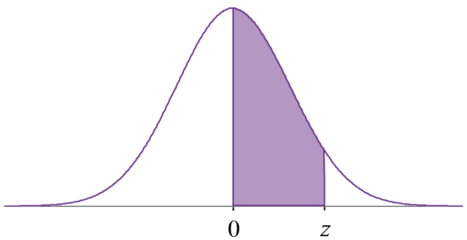
（2） X の分布関数は $1-e^{-2x}$ である．これを利用して，到着時間間隔が 0 と 1 の間にある確率を求めよ．

3.15 2 つの密度関数 $f_1(x)$，$f_2(x)$ はそれぞれ平均 μ_1, μ_2，分散 ${\sigma_1}^2, {\sigma_2}^2$ をもつとする．そのとき，

$$f(x) = \frac{1}{3}f_1(x) + \frac{2}{3}f_2(x)$$

もまた密度関数であることを示せ．この分布の平均と分散を求めよ．

标準正規分布表



z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1311	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2703	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Lecture 6

正規分布と中心極限定理

【教科書】

第 5 章　母集団と標本

1　標本抽出

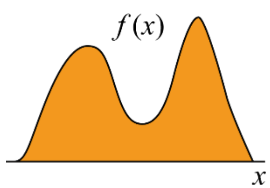
2　標本平均と標本分散

4　大数の法則と中心極限定理

標本抽出

母集団 (population) = 調査対象の全体　(有限または無限)
標本(sample) = 母集団から取り出した要素

母集団
Population



母集団分布
母数 θ

無作為
復元抽出

標本
Samples

$X_1, X_2, \dots, X_n$

計算

母数の推定

数理統計学の役割
合理的な計算公式を与える

$$T = T(X_1, X_2, \cdots, X_n)$$

母数 = 母集団の統計量
平均値、分散、そのほか

平均値の推定

母集団
母平均 μ
母分散 σ^2

無作為
復元抽出

標本
$X_1, X_2, \dots, X_n$

定義　標本平均

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

数学的には

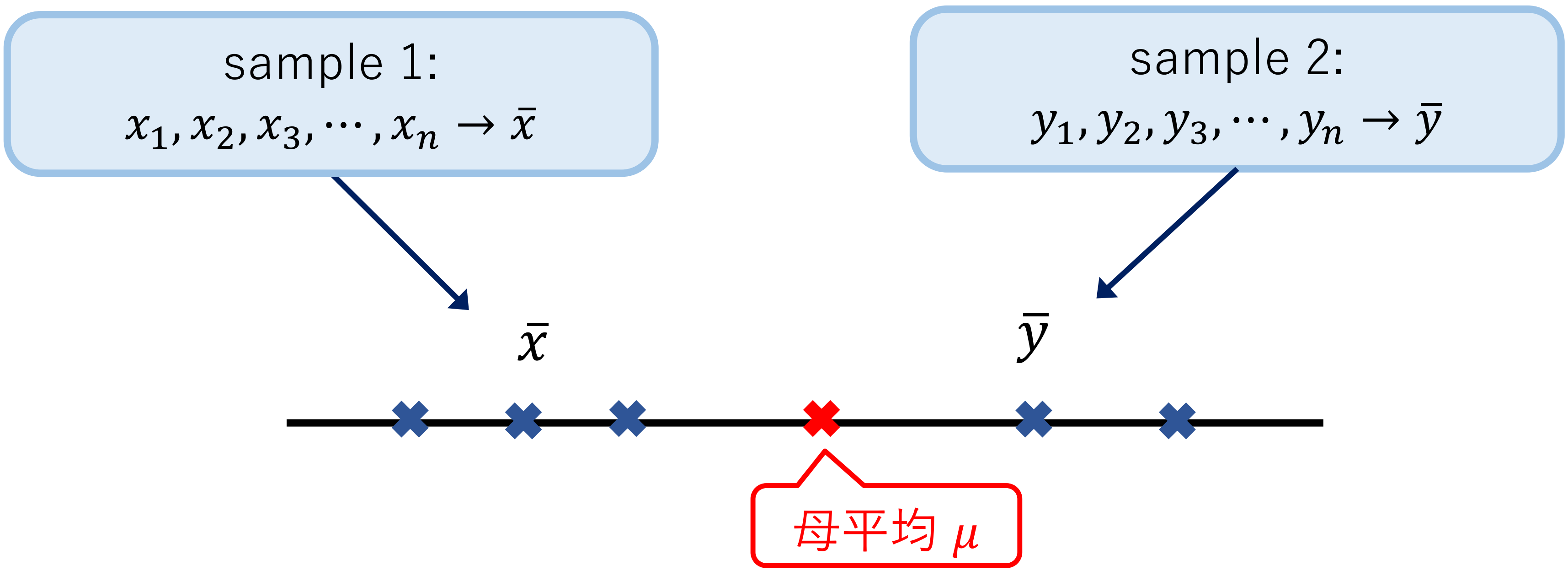
$X_1, X_2, \cdots, X_n$ は
母集団分布に従
う独立同分布(iid)
な確率変数列

母集団
母平均 μ
母分散 σ^2

無作為
復元抽出

標本
$X_1, X_2, \ldots, X_n$

標本平均　$\bar{X} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$　は確率変数である



定理 5.1　母平均を μ，母分散を σ^2 とする．

標本平均 $\bar{X} = \dfrac{1}{n}\sum_{k=1}^{n} X_k$ の平均値と分散は

$$E(\bar{X}) = \mu, \qquad V(\bar{X}) = \frac{\sigma^2}{n}$$

証明

[平均値の線形性] 確率変数 X, Y と 定数 a, b に対して

$$E[aX + bY] = aE[X] + bE[Y]$$

[分散の加法性] 独立な確率変数 X, Y と 定数 a, b に対して

$$V[aX + bY] = a^2\, V[X] + b^2\, V[Y]$$

定理 5.1 母平均を μ，母分散を σ^2 とする．

標本平均 $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} X_k$ の平均値と分散は

$$E(\bar{X}) = \mu, \qquad V(\bar{X}) = \frac{\sigma^2}{n}$$

証明

$$E(\bar{X}) = E\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n}\sum_{k=1}^{n} E(X_k) = \frac{1}{n}\sum_{k=1}^{n} \mu = \mu$$

$$V(\bar{X}) = V\left(\frac{1}{n}\sum_{k=1}^{n} X_k\right) = \frac{1}{n^2}\sum_{k=1}^{n} V(X_k) = \frac{1}{n^2}\sum_{k=1}^{n} \sigma^2 = \frac{\sigma^2}{n}$$

標本分散と不偏分散

母集団
母平均 μ
母分散 σ^2

無作為復元抽出

標本
$X_1, X_2, \ldots, X_n$

標本平均

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

標本分散

$$S^2 = \frac{1}{n}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

不偏分散

$$U^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

定理 5.3　母平均を μ，母分散を σ^2 とする．

(1) 標本分散　$S^2 = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n}(X_k - \bar{X})^2$　の平均値は

$$E(S^2) = \frac{n-1}{n}\sigma^2$$

(2) 不偏分散　$U^2 = \dfrac{1}{n-1}\displaystyle\sum_{k=1}^{n}(X_k - \bar{X})^2$　の平均値は

$$E(U^2) = \sigma^2$$

推定値として
優れている

証明　省略（教科書を見よ）

大数の法則 (Law of Large Numbers = LLN)

標本の大きさ n が十分大きければ，標本平均 $\bar{X}$ が

母平均 μ に近いということがいえる.

$$\boxed{\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k} \approx \mu$$

近い

左辺は確率変数!

それが定数 μ に近いとはどう表現すべきか？

コインンン投げのシミュレーション

コイン投げを繰り返す：　$X_1, X_2, \cdots, X_n, \cdots$

$X_n = 1$ (表のとき),　$X_n = 0$ (裏のとき),

$$\boxed{T_n = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad \text{初めの } n \text{ 回の内, 表の回数の相対頻度}}$$

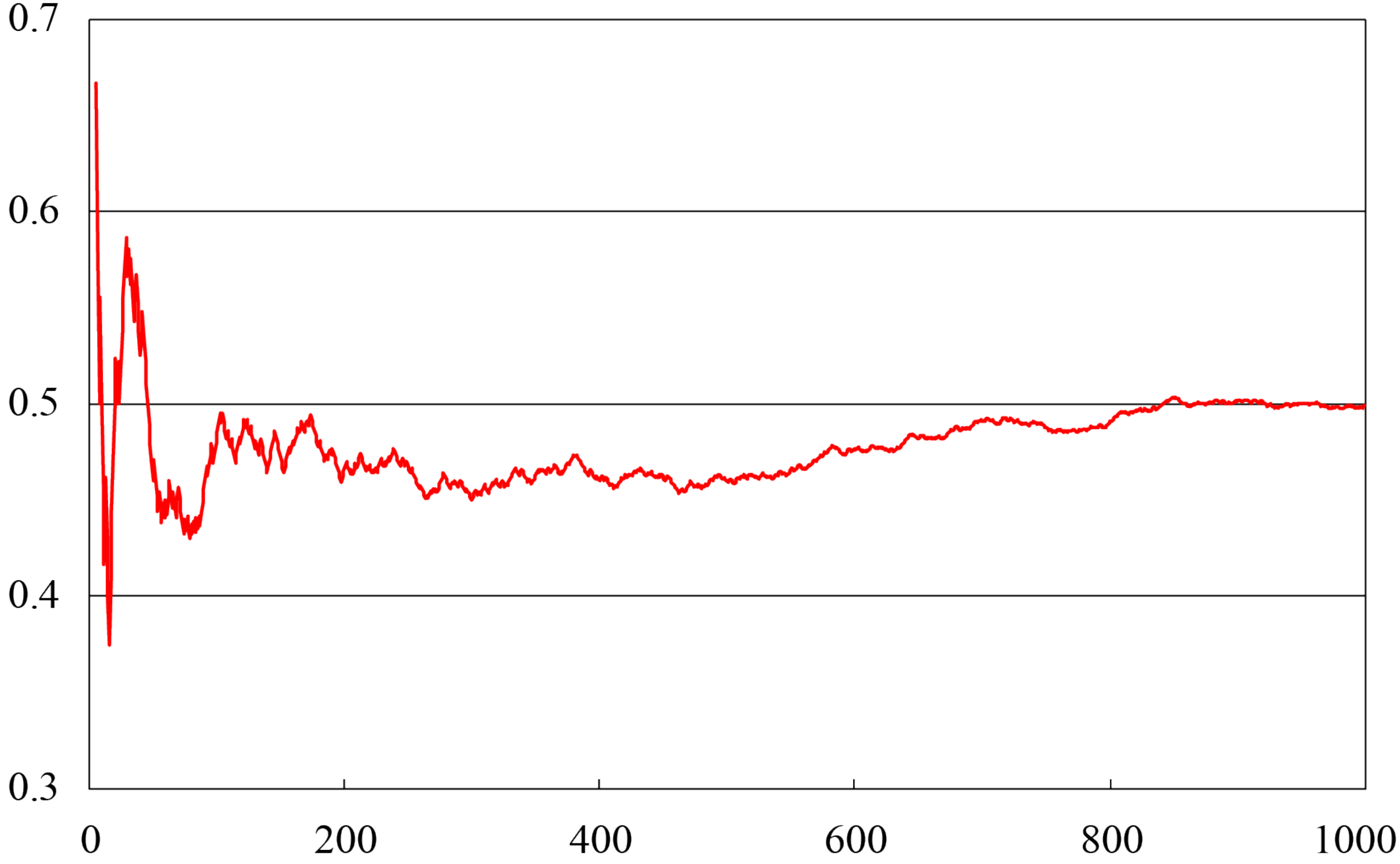
母平均（コイン投げ1回の平均値）

$$\mu = E(X_1) = 1 \times \frac{1}{2} + 0 \times \frac{1}{2} = \frac{1}{2}$$

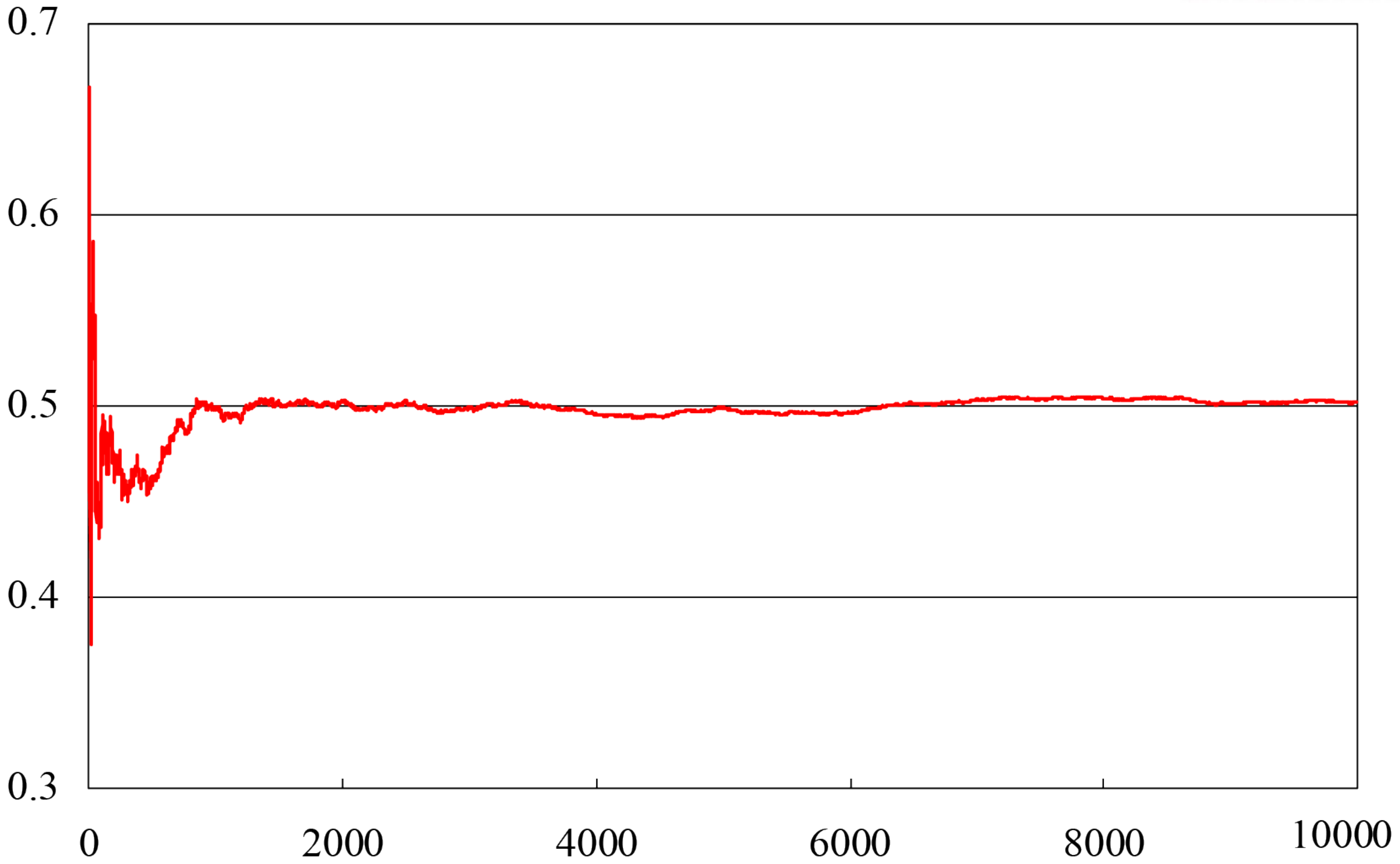
大数の法則

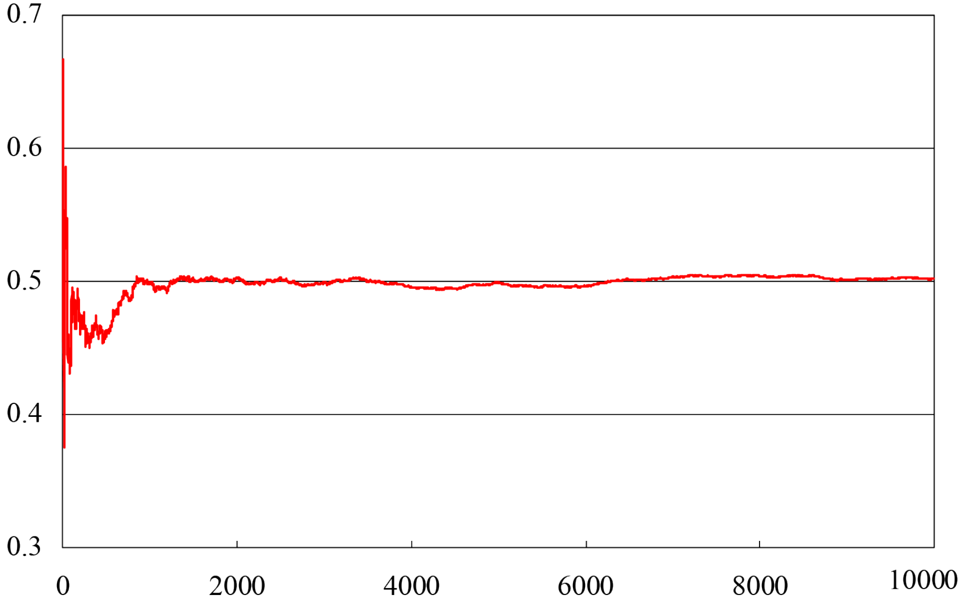
$T_n \approx \mu$

$$T_n = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad 1 \leq n \leq 1000$$



$$T_n = \frac{1}{n}\sum_{k=1}^{n} X_k \qquad 1 \le n \le 10000$$





$$T_n = \frac{1}{n}\sum_{k=1}^{n} X_k \to 0.5$$

だんだん
近づく

微積分の概念で「収束する」
と言ってよいか？

毎回表が出続けることもありうる（確率は小さいが）

そうすると, いつまでたっても $T_n = 1$ ⇔ 矛盾? ⇔ $T_n \to 0.5$

⇒ 確率を用いた表現が必要

大数の法則 (Law of Large Numbers = LLN)

標本の大きさ n が十分大きければ，標本平均 $\bar{X}$ が母平均 μ に近い．

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k \approx \mu$$

定理 5.8

母平均 μ をもつ分布からの無作為標本 $X_1, X_2, \cdots, X_n$ の標本平均

$$T_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$

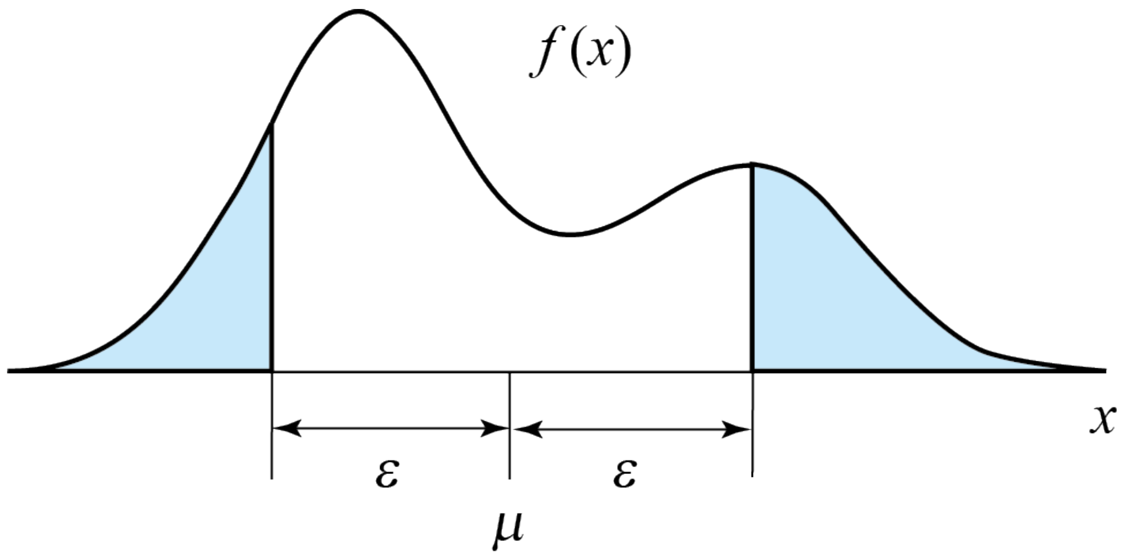
と任意の $\epsilon > 0$ に対して，次の式が成り立つ．

$$P(|T_n - \mu| > \epsilon) \to 0 \qquad (n \to \infty)$$

定理 5.7　（チェビシェフの不等式）

確率変数 X が平均 μ , 分散 σ^2 をもつとき，任意の $\epsilon > 0$ に対して，次の不等式が 成り立つ .

$$P(|X - \mu| > \epsilon) \leq \frac{\sigma^2}{\epsilon^2}$$



大数の法則の証明

※ ここでは，分布が分散をもつものとして，
チェビシェフの不等式を使って証明する

$$T_n = \frac{1}{n}\sum_{k=1}^{n} X_k$$ について, 定理5.1から

$$E(T_n) \ = \ \mu, \qquad V(T_n) = \frac{\sigma^2}{n}$$

チェビシェフの不等式を適用すれば,

$$P(|T_n - \mu| > \epsilon) \ \leq \ \frac{\sigma^2}{n\epsilon^2} \to \ 0$$

中心極限定理 (Central Limit Theorem = CLT)

定理 5.4

正規母集団 $N(\mu,\sigma^2)$ から取り出した標本 $X_1,X_2,\cdots,X_n$ に対して

$$\bar{X} \sim N\left(\mu,\frac{\sigma^2}{n}\right) \Leftrightarrow \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

標準化または z-変換

定理 5.9　中心極限定理(CLT)

平均値 μ，分散 σ^2 の<u>一般の母集団</u>でも，n が十分大きいとき，近似的に成り立つ．

CLTは実世界によく現れる

$$S_n = Z_1 + Z_2 + \cdots + Z_n = \sum_{k=1}^{n} Z_k$$

$$Z_1 \quad Z_2 \quad Z_3 \quad Z_4 \quad \cdots \quad Z_n$$

$Z_1, Z_2, \cdots, Z_n, \cdots$ ：独立同分布な確率変数列

それぞれの平均値 $= \mu$,　分散 $= \sigma^2$

$$E[S_n] = \sum_{k=1}^{n} E[Z_k] = n\mu$$

CLT　　$S_n - n\mu \sim N(0, n\sigma^2)$　　$(n \to \infty)$

・細かい誤差が集積すると全体の揺らぎは正規分布に従う

例題 5.4 （二項分布の正規近似）

二項分布 $B(n,p)$ は, n が大きいとき，同じ平均値と分散をもつ正規分布 $N(np\,,np(1-p))$ で近似できる.

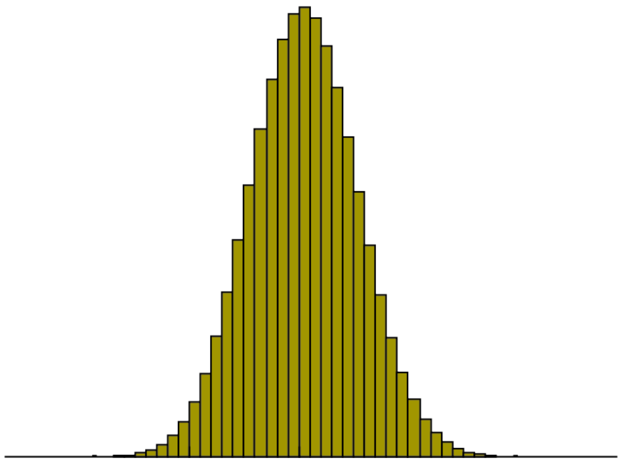
$Z_1, Z_2, \cdots, Z_n, \cdots$ ：ベルヌーイ試行列

それぞれの平均値 $= p$,　分散 $= p(1-p)$

$X_n = Z_1 + Z_2 + \cdots + Z_n \sim B(n,p)$

CLT により,　$$\frac{1}{n}\sum_{k=1}^{n} Z_k \approx N\left(p, \frac{p(1-p)}{n}\right)$$

したがって,　$$\sum_{k=1}^{n} Z_k \approx N(np, np(1-p))$$



宿題

演習問題５ (page 94-96)

5.2 確率変数 X が 2 項分布 $B_N(10, 0.4)$ に従うとき，X が母平均 μ から 2 以上離れる確率 $P(|X-\mu| \geq 2)$ を求めよ．その確率についてチェビシェフの不等式（p. 91）が成り立つことを確かめよ．

5.3 確率変数 X が正規分布 $N(5, 4)$ に従うとき，X が母平均 μ から 3 以上離れる確率 $P(|X-\mu| \geq 3)$ を求めよ．その確率についてチェビシェフの不等式が成り立つことを確かめよ．

5.4 確率変数 X が 2 項分布 $B_N(50, 0.1)$ に従うとき，確率 $P(X \leq 3)$ を計算せよ．また，この確率を 2 項分布のポアソン近似（2 項分布がポアソン分布に収束すること，p.53 参照）を用いて計算し，正確な確率と比較せよ．

5.5 確率変数 X が 2 項分布 $B_N(12, 0.4)$ に従うとき，確率 $P(3 < X \leq 6)$ を計算せよ．また，この確率を 2 項分布の正規近似を用いて計算し，正確な確率と比較せよ．

5.9 $X_1, \ldots, X_n$ は一様分布 $U(0, 1)$ に従う n 個の無作為標本とする．それらの値が t 以下の標本の個数を $\#\{X_i \leq t\}$ と表すとき，確率変数 $X = \#\{X_i \leq t\}$ はどのような分布に従うか．また，標本比率

$$F_n(t) = \frac{1}{n}\#\{X_i \leq t\}$$

を t の関数と見て，**経験分布関数** (empirical distribution function) という．そのとき，経験分布関数の平均と分散を求めよ．

5.10 (5.9 の続き) $0 \leq s < t \leq 1$ に対し，確率変数 $X_1 = \#\{X_i \leq s\}$，$X_2 = \#\{s < X_i \leq t\}$ を考えるとき，(X_1, X_2) はどのような分布に従うか．また，$F_n(s)$ と $F_n(t)$ の共分散を求めよ．

5.11 (5.10 の続き) 一様分布 $U(0, 1)$ の経験分布関数 $F_n(t)$ について，次の確率変数 $B_n(t)$ は，$n \to \infty$ のとき，正規確率変数 $B(t)$ に収束することを示せ．

$$B_n(t) = \sqrt{n}\,\{F_n(t) - t\} \to B(t) = N[0,\, t(1-t)]$$

5.12 ある家庭には固定電話 1 個と携帯電話 2 個があるとする．1 日に固定電話にかかってくる電話の回数 X は平均 $E(X) = 4$ のポアソン分布 $Po(4)$ に従う．1 日に 2 つの携帯電話にかかってくる電話の回数をそれぞれ Y, Z とし，これらはいずれも平均 $E(Y) = E(Z) = 3$ のポアソン分布 $Po(3)$ に従い，それらは独立とする．そのとき，次の問に答えよ．

(1) 1 日に携帯電話にかかってくる電話の回数の合計 $Y + Z$ はどのような分布に従うか．

(2) 1 日にその家庭にかかってくる電話の回数の合計 $X + Y + Z$ はどのような分布に従うか．また，1 日にその家庭にかかってくる電話の回数の合計が $X + Y + Z = 10$ である確率を求めよ．

(3) 1 日にその家庭にかかってくる電話の回数の合計が $X + Y + Z = 10$ であるとき，固定電話の回数 X はどのような分布に従うか．

Lecture 7

統計的推定

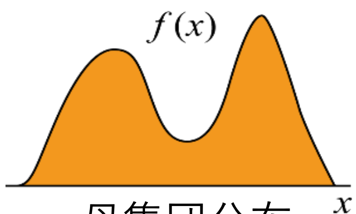
【教科書】

第 6 章 推　定

1 推定量とその性質

推定量（estimator）

母集団



母集団分布

母数 θ

無作為
復元抽出

標 本

$X_1, X_2, \ldots, X_n$

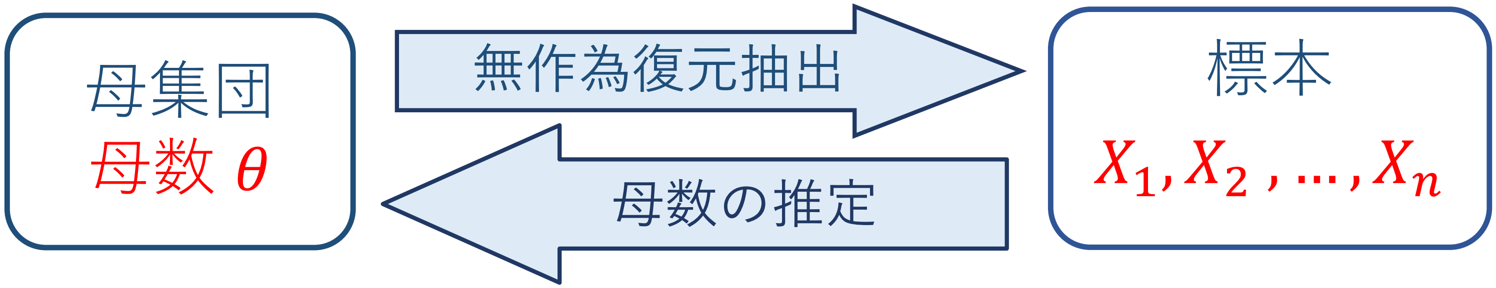
母平均 μ
母分散 σ^2
など

数学的には

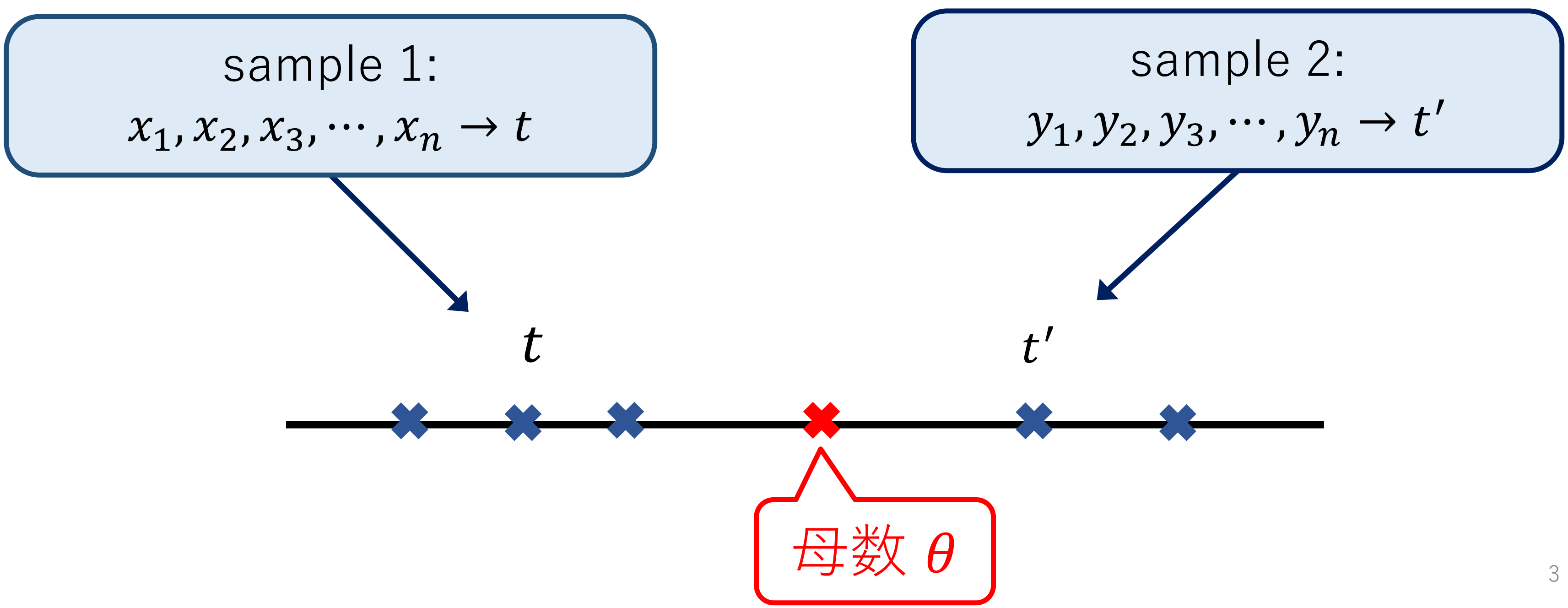
$X_1, X_2, \cdots, X_n$ は母集団分布に従う独立同分布な確率変数列

定義	推定量	$\hat{\theta} = T = T(X_1, X_2, \cdots, X_n)$

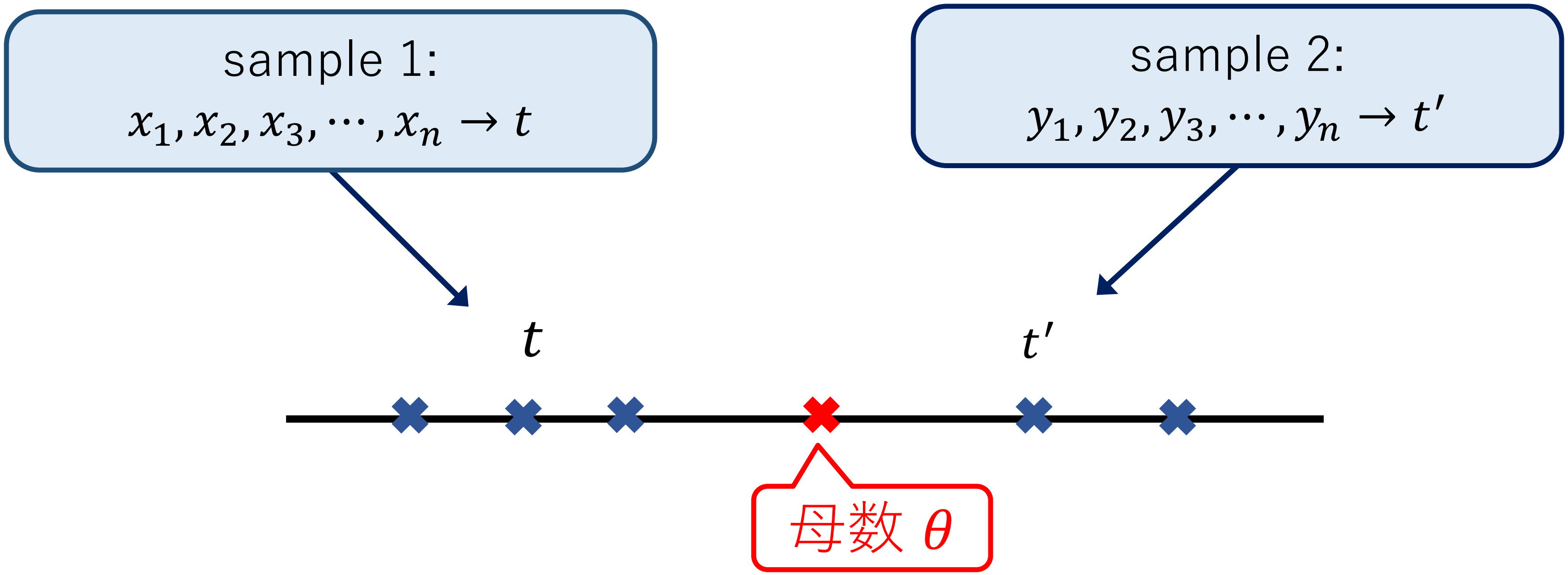
※ データから計算される $T(x_1, x_2, \cdots, x_n)$ は推定値という.



推定量 $T = T(X_1, X_2, \cdots, X_n)$ は確率変数である



不偏推定量（unbiased estimator）



推定量 $T = T(X_1, X_2, \cdots, X_n)$ は確率変数である

定 義　推定量 T が $\boldsymbol{E(T) = \theta}$ を満たすとき，その推定量を不偏推定量という．

例題 6.1 母平均 μ，母分散 σ^2 の母集団からとりだされた無作為標本を $X_1, X_2, \ldots, X_n$ とする．

(1) 標本平均 $\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$ は母平均 μ の不偏推定量である．

(2) 不偏分散 $U^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$ は母分散 σ^2 の

不偏推定量である．

---

(1) 定理 5.1 によって $E(\bar{X}) =$ μ は既知

(2) 定理 5.3 によって $E(U^2) = \sigma^2$ は既知.

演 習（10分）

6.3 $X_1, \ldots, X_n$ は独立で同一分布に従う観測で，その平均は $E(X_i) = \mu$，分散は $V(X_i) = \sigma^2$ であるとする．定数 $c_1, \ldots, c_n$ を係数とする観測の線形関数 $T = c_1X_1 + \cdots + c_nX_n$ を考える．

(1) T が μ の不偏推定量であるために，定数 $c_1, \ldots, c_n$ の満たすべき条件を求めよ．

(2) 線形不偏推定量の中で，分散を最小にする定数 $c_1, \ldots, c_n$ を求めよ．

演 習（10分）

6.3 $X_1, \ldots, X_n$ は独立で同一分布に従う観測で，その平均は $E(X_i) = \mu$，分散は $V(X_i) = \sigma^2$ であるとする．定数 $c_1, \ldots, c_n$ を係数とする観測の線形関数 $T = c_1X_1 + \cdots + c_nX_n$ を考える．

（1） T が μ の不偏推定量であるために，定数 $c_1, \ldots, c_n$ の満たすべき条件を求めよ．

（2） 線形不偏推定量の中で，分散を最小にする定数 $c_1, \ldots, c_n$ を求めよ．

[ヒント]

(1) [平均値の線形性] 確率変数 X, Y と 定数 a, b に対して

$$E[aX + bY] = aE[X] + bE[Y]$$

(2) [分散の加法性] 独立な確率変数 X, Y と 定数 a, b に対して

$$V[aX + bY] = a^2\,V[X] + b^2\,V[Y]$$

演 習（10分）

6.3 $X_1, \ldots, X_n$ は独立で同一分布に従う観測で，その平均は $E(X_i) = \mu$，分散は $V(X_i) = \sigma^2$ であるとする．定数 $c_1, \ldots, c_n$ を係数とする観測の線形関数 $T = c_1X_1 + \cdots + c_nX_n$ を考える．

（1） T が μ の不偏推定量であるために，定数 $c_1, \ldots, c_n$ の満たすべき条件を求めよ．

（2） 線形不偏推定量の中で，分散を最小にする定数 $c_1, \ldots, c_n$ を求めよ．

---

(1)

$$E(T) = E(c_1X_1 + \cdots + c_nX_n)$$
$$= c_1E(X_1) + \cdots + c_nE(X_n)$$
$$= c_1\mu + \cdots + c_n\mu$$
$$= (c_1 + \cdots + c_n)\mu$$

加重平均という

$E(T) = \mu$ となるのは $c_1 + \cdots + c_n$=1 のときである．

加重平均（重み付き平均）

普通の平均（算術平均）

$$\frac{X_1 + X_2}{2} = \frac{1}{2} X_1 + \frac{1}{2} X_2$$

重み付きの平均（加重平均）

$$\frac{1}{3} X_1 + \frac{2}{3} X_2$$

加重平均

$$T = \sum_{k=1}^{n} c_k X_k \quad \text{ただし} \quad \sum_{k=1}^{n} c_k = 1$$

等加重にすれば普通の算術平均　$\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$

演 習（10分）

6.3 $X_1, \ldots, X_n$ は独立で同一分布に従う観測で，その平均は $E(X_i) = \mu$，分散は $V(X_i) = \sigma^2$ であるとする．定数 $c_1, \ldots, c_n$ を係数とする観測の線形関数 $T = c_1 X_1 + \cdots + c_n X_n$ を考える．

（1） T が μ の不偏推定量であるために，定数 $c_1, \ldots, c_n$ の満たすべき条件を求めよ．

（2） 線形不偏推定量の中で，分散を最小にする定数 $c_1, \ldots, c_n$ を求めよ．

---

(2)
$$
\begin{aligned}
V(T) &= V(c_1 X_1 + \cdots + c_n X_n) \\
&= c_1^2 V(X_1) + \cdots + c_n^2 V(X_n) \\
&= (c_1^2 + \cdots + c_n^2)\sigma^2
\end{aligned}
$$

$c_1 + \cdots + c_n = 1$ のもとで $V(T)$ の最小値を考える．

※　$c_1+\cdots+c_n=1$ のもとで $c_1^2+\cdots+c_n^2$ の最小値を考える.

$$c_1^2+\cdots+c_n^2=(c_1-a)^2+\cdots+(c_n-a)^2+2a(c_1+\cdots+c_n)-na^2$$

$$=(c_1-a)^2+\cdots+(c_n-a)^2+2a-na^2$$

したがって，$c_1^2+\cdots+c_n^2$ は $c_1=\cdots=c_n=a$ のとき最小値をとる.

$c_1+\cdots+c_n=1$ を合わせて

$$c_1=\cdots=c_n=\frac{1}{n} \qquad (*)$$

算術平均
(普通の平均)

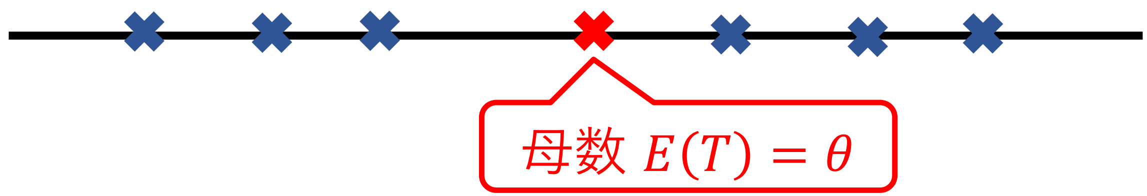
のとき，$c_1^2+\cdots+c_n^2$ は最小値をとる.

つまり，$T=c_1X_1+\cdots+c_nX_n$ の分散は（∗）のとき最小になる.

不偏推定量の比較

T：母数 θ の不偏推定量 $\Rightarrow$ $E(T) = \theta$

T の分散：$V(T) = E\{(T-\theta)^2\}$



ゆえに，不偏推定量の分散は推定量の良さの指標と考えることができる（分散が小さいほど推定量として優れている）

T_1, T_2：母数 θ の不偏推定量

$$V(T_1) \leq V(T_2)$$

T_2 よりも T_1 はより有効（efficient）であるという．

例＊＊ （ドイツ戦車問題）

1番から順に通し番号がついている戦車が街を巡回している．目撃された番号が X_1, X_2, X_3 であったとき，それらの最大値を $M = \max\{X_1, X_2, X_3\}$ とおく．このとき, $\hat{N} = \frac{4}{3}M - 1$ は戦車の総数 N の不偏推定量である．

$$P(M = k) = \frac{\binom{k-1}{2}}{\binom{N}{3}} = \frac{3!\,(k-1)(k-2)}{2!\,N(N-1)(N-2)} = \frac{3(k-1)(k-2)}{N(N-1)(N-2)}$$

$$E[M] = \sum_{k=1}^{N} kP(M = k) \quad = \frac{3}{N(N-1)(N-2)} \sum_{k=1}^{N} k(k-1)(k-2)$$

$$= \frac{3}{N(N-1)(N-2)} \times \frac{1}{4} \sum_{k=1}^{N} \{(k+1)k(k-1)(k-2) - k(k-1)(k-2)(k-3)\}$$

$$= \frac{3}{4}(N+1)$$

例＊＊　（ドイツ戦車問題）

目撃された番号が $X_1, X_2, \cdots, X_n$ であったとき，それらの最大値を

$$M = \max\{X_1, X_2, \cdots, X_n\}$$

とすれば，$\hat{N} = \left(1 + \frac{1}{n}\right) M - 1$ は戦車の総数 N の不偏推定量である．

Month	Statistical estimate	Conventional estimate*)	German records
June 1940	169	1,000	122
June 1941	244	1,550	271
August 1942	327	1,550	342

*) 連合軍情報本部の伝統手法

宿題

演習問題 6 (page 107-109)

6.1 確率変数 X_1, X_2 は独立で，平均 μ，分散 σ^2 をもつ同分布に従うとき，次のものを求めよ．

(1) $E(X_1 + X_2)$ (2) $E(X_1 X_2)$ (3) $V(X_1 + X_2)$ (4) $V(X_1 X_2)$

6.2 X, Y は独立な観測で，平均と分散はそれぞれ

$$E(X) = E(Y) = \mu, \quad V(X) = \sigma^2, \quad V(Y) = 3\sigma^2$$

であるとする．定数 a, b を係数とする観測の線形関数 $T = aX + bY$ に対して，次の問に答えよ．

（1） T が μ の不偏推定量であるためには，a, b はどのような条件を満たさねばならないか．

（2） T が μ の不偏推定量であるとき，分散を最小にする a, b の値を求めよ．

6.3 $X_1, \ldots, X_n$ は独立で同一分布に従う観測で，その平均は $E(X_i) = \mu$，分散は $V(X_i) = \sigma^2$ であるとする．定数 $c_1, \ldots, c_n$ を係数とする観測の線形関数 $T = c_1X_1 + \cdots + c_nX_n$ を考える．

（1） T が μ の不偏推定量であるために，定数 $c_1, \ldots, c_n$ の満たすべき条件を求めよ．

（2） 線形不偏推定量の中で，分散を最小にする定数 $c_1, \ldots, c_n$ を求めよ．

6.4 平均 μ_1, μ_2，分散 ${\sigma_1}^2, {\sigma_2}^2$ をもつ 2 つの異なる母集団から，無作為標本 $X_1, \ldots, X_{n_1}; Y_1, \ldots, Y_{n_2}$ の標本平均を $\bar{X}, \bar{Y}$ とするとき，次の加重平均 M の平均 $E(M)$ と分散 $V(M)$ を求めよ．

$$M = \frac{n_1\bar{X} + n_2\bar{Y}}{n_1 + n_2}$$

6.5 $X_1, X_2, \ldots, X_n$ を一様分布 $U(0, \theta)$ からの無作為標本とする.

(1) $Y = \max\{X_1, X_2, \ldots, X_n\}$ の密度関数を求めよ.

(2) 次の2つの推定量 T_1, T_2 はともに, θ の不偏推定量であることを示せ.

$$T_1 = \frac{n+1}{n}Y, \quad T_2 = 2\bar{X}$$

(3) (2)の2つの推定量 T_1, T_2 のうち, どちらがより有効な推定量であるか.

6.6 $X_1, X_2, \ldots, X_n$ は指数分布 $E_X\left(\frac{1}{\lambda}\right)$ からの無作為標本とする. 母数 λ の2つの推定量として $T_1 = \bar{X}$（標本平均）と標本最小値を使った推定量 $T_2 = n\min\{X_1, X_2, \ldots, X_n\}$ を考える. そのとき, 次の問に答えよ.

(1) これらの推定量は不偏であることを示せ.

(2) これらの推定量の分散を求め, 比較せよ.

Lecture 8

正規分布にまつわる分布

（ χ^2-分布、$\boldsymbol{t}$-分布、$\boldsymbol{F}$-分布）

【教科書】

第 5 章 母集団と標本

3 正規分布から導かれる標本分布

母集団
母平均 μ
母分散 σ^2

無作為復元抽出

標本
$X_1, X_2, \ldots, X_n$

標本平均

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

➢ 基礎理論のために, 正規母集団 $N(\mu, \sigma^2)$ を扱う

定理 5.4（復習）

正規母集団 $N(\mu, \sigma^2)$ から取り出した標本 $X_1, X_2, \cdots, X_n$ に対して

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

標準化または z-変換

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$$

※ 母分散 σ^2 が未知のときに使えない.

正規母集団
$N(\mu, \sigma^2)$

無作為
復元抽出

標本
$X_1, X_2, \dots, X_n$

標本平均

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

$$\boxed{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)}$$

σ^2 が未知のときは不偏分散で代替する　$E(U^2) = \sigma^2$

不偏分散

$$U^2 = \frac{1}{n-1}\sum_{k=1}^{n}(X_k - \bar{X})^2$$

$$\boxed{\frac{\bar{X}-\mu}{U/\sqrt{n}} \sim ???}$$

カイ 2 乗分布（χ^2-分布）

$Z_1, Z_2, \cdots, Z_\nu$ ：

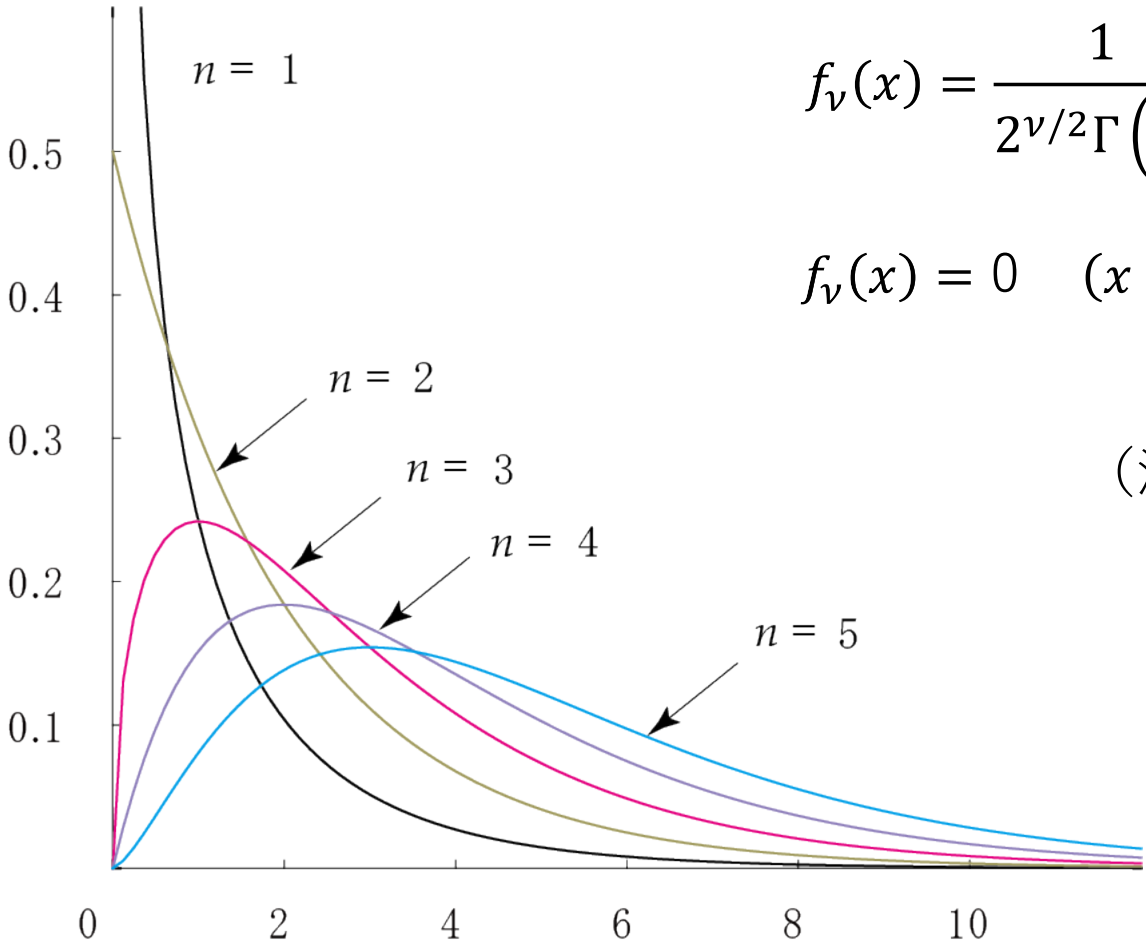
標準正規分布 $N(0,1)$ に従う ν 個の無作為標本

それらの 2 乗和

$$X = \sum_{i=1}^{\nu} Z_i^2$$

の分布を**自由度 ν のカイ 2 乗分布**と呼び, 記号 χ_ν^2 で表す.

自由度 ν のカイ 2 乗分布（χ^2_ν-分布）



$$f_\nu(x) = \frac{1}{2^{\nu/2}\Gamma\left(\frac{\nu}{2}\right)}\, x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} \qquad (x \geq 0),$$

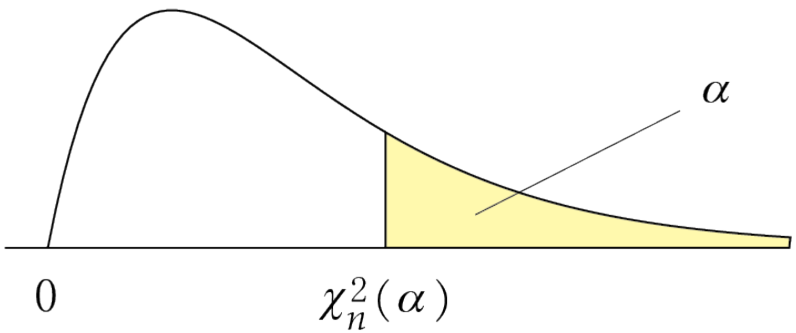
$$f_\nu(x) = 0 \quad (x < 0)$$

（注） $\Gamma(x)$ はガンマ関数

平均値 $m = \nu$

分散 $\sigma^2 = 2\nu$

自由度 ν のカイ 2 乗分布（χ^2_ν-分布）



上側 α 点

$\chi^2_\nu(\alpha)$

$n\backslash\alpha$	0.995	0.99	0.975	0.95	0.05	0.025	0.01	0.005
1	0.0^4393	0.0^3157	0.0^3982	0.0^2393	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	37.652	40.646	44.314	46.928

$\nu = 7$

$\alpha = 0.05$

$\chi^2_7(0.05) = 14.067$

演習（5分）

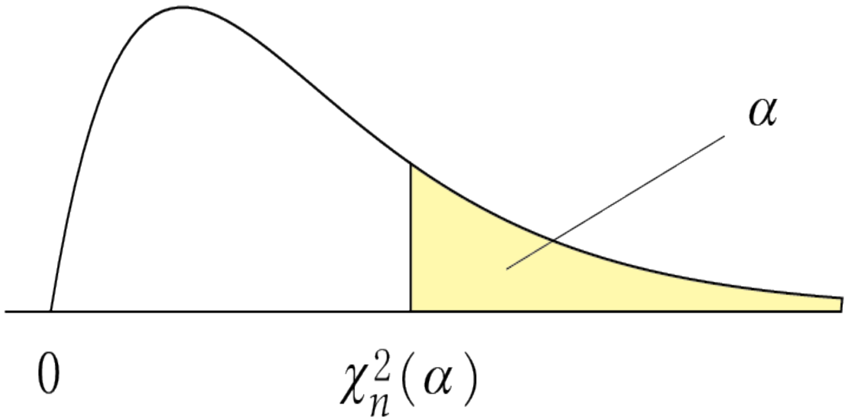
例 題 5.2

自由度が 10 のカイ 2 乗分布に従う確率変数 X がある値を超える確率が 5 % であるようにするには, その値をいくらにすればよいか?

演習（5分）

例 題 5.2

自由度が 10 のカイ 2 乗分布に従う確率変数 X がある値を超える確率が 5 % であるようにするには, その値をいくらにすればよいか?

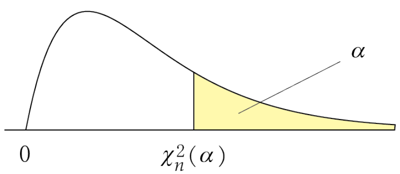


$\nu = 10, \qquad \alpha = 0.05$

上側 5% 点

$\chi_{10}^2(0.05) = 18.307$

カイ2乗分布



上側 α 点

$\chi_\nu^2(\alpha)$

ν \ α	.99	.975	.95	.90	.70	.50	.30	.10	.05	.025	.01
1	.000157	.00098	.00393	.0158	.148	.455	1.074	2.706	3.841	5.0238	6.635
2	.0201	.0506	.103	.211	.713	1.386	2408	4 605	5.991	7 3780	9.210
3	.115	.216	.352	.584	1.424	2.366	3.665	6 251	7.815	9 348	11.345
4	.297	.484	.711	1.064	2.195	3.357	4.878	7.779	9.488	11.243	13.277
5	.554	.831	1.145	1.610	3.000	4.351	6.064	9.236	11.070	12.832	15.086
6	.872	1.237	1.635	2.204	3.828	5.348	7.231	10 645	12.592	14.449	16.812
7	1.239	1.690	2.167	2.833	4.671	6.346	8.383	12.017	14.067	16 013	18.475
8	1.646	2.180	2.733	3.490	5.527	7.344	9.524	13.362	15.507	17.535	20.090
9	2.088	2.700	3.325	4.168	6.393	8.343	10.656	14.684	16.919	19.023	21.666
10	2.558	3.247	3.940	4.865	7.267	9.342	11.781	15.987	18.307	20.483	23.209
11	3.053	3.816	4.575	5.578	8.148	10.341	12.899	17.275	19.675	21.920	24.725
12	3 571	4.404	5.226	6.304	9.034	11.340	14.011	18.549	21.026	23 337	26.217
13	4.107	5.009	5.892	7.042	9.926	12.340	15 119	19.812	22.362	24.736	27.688
14	4.660	5.629	6 571	7.790	10 821	13.339	16.222	21.064	23.685	26.119	29.141
15	5.229	6.262	7.261	8.547	11.721	14.339	17.322	22.307	24.996	27.488	30.578
16	5.812	6.908	7.962	9.312	12.624	15.338	18.418	23.542	26.296	28.845	32.000
17	6.408	7.564	8.672	10.085	13.531	16.338	19.511	24 769	27.587	30.191	33.409
18	7.015	8.231	9.390	10.865	14.440	17.338	20.601	25.989	28.869	31.526	34.805
19	7.633	8.907	10.117	11.651	15.352	18.338	21.689	27.204	30.144	32.852	36.191
20	8.260	9.591	10.851	12.443	16.266	19.337	22.775	28.412	31.410	34.170	37.566
21	8.897	10.283	11.591	13.240	17.182	20.337	23.858	29.615	32.671	35.479	38.932
22	9.542	10.982	12.338	14.041	18.101	21.337	24.939	30.813	33.924	36.781	40.289
23	10.196	11.689	13.091	14.848	19.021	22.337	26.018	32 007	35.172	38 076	41.638
24	10 856	12.401	13.848	15.659	19.943	23 337	27.096	33 196	36.415	39 364	42980
25	11.524	13.120	14 611	16 473	20.867	24 337	28.172	34.382	37.652	40 646	44.314
26	12.198	13.844	15.379	17.292	21.792	25.336	29.246	35.563	38.885	41.923	45.642
27	12.879	14.573	16.151	18.114	22.719	26.336	30.319	36.741	40.113	43.194	46.963
28	13.565	15.308	16.928	18.939	23.647	27.336	31.391	37.916	41.337	44.461	48.278
29	14.256	16.047	17 708	19.768	24 577	28.336	32.461	39.087	42.557	45.722	49.588
30	14.953	16.791	18.493	20.599	25.508	29.336	33.530	40.256	43.773	46.979	50.892

定理 5.5（不偏分散の分布）

$X_1, X_2, \cdots, X_n$：標準正規分布 $N(0,1)$ に従う n 個の無作為標本

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ ：標本平均

$$Y = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(n-1)U^2}{\sigma^2}$$

は自由度 $v = n - 1$ のカイ 2 乗分布 χ^2_{n-1} に従う.

標準化すると， $Z_i = \dfrac{X_i - \mu}{\sigma} \sim N(0,1)$,　$Z = \dfrac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$

$$\underbrace{\sum_{i=1}^{n} Z_i^2}_{\chi_n^2} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \mu)^2 = \frac{1}{\sigma^2}\sum_{i=1}^{n} X_i^2 - \frac{2\mu}{\sigma^2}\sum_{i=1}^{n} X_i + \frac{n\mu^2}{\sigma^2}$$

$$= \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 + \frac{2\bar{X}}{\sigma^2}\sum_{i=1}^{n} X_i - \frac{n\bar{X}^2}{\sigma^2} - \frac{2\mu}{\sigma^2}\, n\bar{X} + \frac{n\mu^2}{\sigma^2} = Y + \overbrace{Z^2}^{\chi_1^2}$$

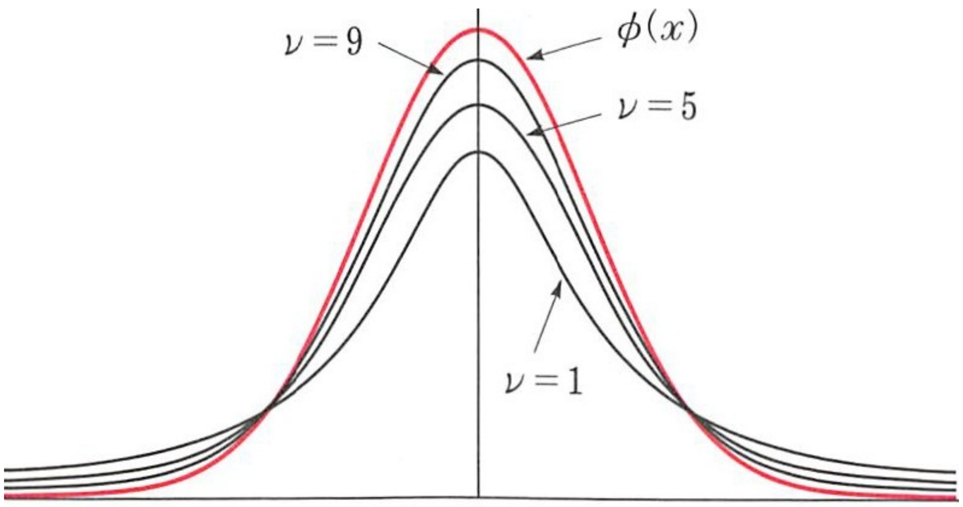
独立

t 分布

確率変数 X, Y が独立であり，$X \sim N(0,1)$，$Y \sim \chi_\nu^2$ であれば，

$$T = \frac{X}{\sqrt{Y/\nu}}$$

は**自由度 ν の t 分布**に従う．これを $T \sim t_\nu$ と書く．



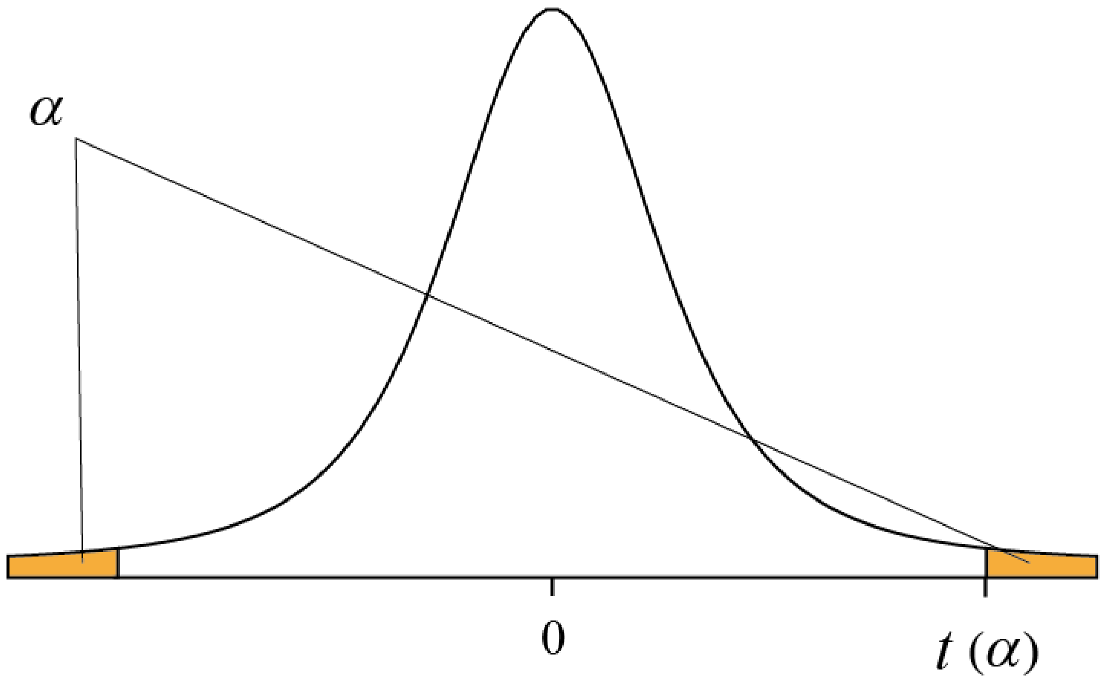
平均値

$$E(T) = 0 \quad (\nu \geq 2)$$

分散

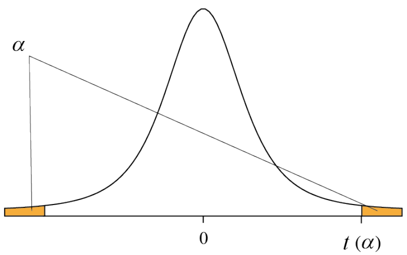
$$V(T) = \frac{\nu}{\nu - 2} \quad (\nu \geq 3)$$

両側 α 点



$$P\left(|T| \geq t_\nu(\alpha)\right) = \alpha$$

t-分布表



両側 α 点

$t_\nu\ (\alpha)$

ν \ α	.90	.80	.70	.60	.50	.40	.30	.20	.10	.05	.02	.01
1	.158	.325	.510	.727	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657
2	.142	.289	.445	.617	.816	1.061	1.386	1.886	2.920	4 303	6.965	9.925
3	.137	.277	.424	.584	.765	.978	1.250	1.638	2.353	3.182	4.541	5.841
4	.134	.271	.414	.569	741	.941	1.190	1.533	2.132	2.776	3.747	4.604
5	.132	.267	.408	.559	727	920	1.156	1.476	2.015	2.571	3.365	4.032
6	.131	.265	404	.553	.718	.906	1.134	1.440	1.943	2.447	3.143	3.707
7	.130	.263	.402	.549	.711	.896	1.119	1.415	1.895	2.365	2.998	3.499
8	.130	.262	.399	.546	.706	.889	1.108	1.397	1.860	2.306	2.896	3.355
9	.129	.261	.398	.543	.703	.883	1.100	1.383	1.833	2.262	2.821	3.250
10	.129	.260	.397	.542	.700	.879	1.093	1.372	1.812	2.228	2.764	3.169
11	.129	.260	.396	.540	.697	.876	1.088	1.363	1.796	2.201	2.718	3.106
12	.128	.259	.395	.539	.695	.873	1.083	1.356	1.782	2.179	2.681	3.055
13	.128	.259	.394	538	694	.870	1.079	1.350	1.771	2.160	2.650	3.012
14	.128	.258	.393	.537	.692	868	1.076	1.345	1 761	2.145	2.624	2.977
15	.128	.258	393	.536	.691	.866	1.074	1.341	1.753	2.131	2.602	2.947
16	.128	.258	.392	.535	690	.865	1.071	1.337	1.746	2 120	2.583	2.921
17	.128	.257	.392	.534	.689	.863	1.069	1.333	1.740	2 110	2.567	2.898
18	.127	.257	.392	.534	.688	.862	1.067	1.330	1.734	2 101	2.552	2.878
19	.127	.257	.391	.533	.688	.861	1.066	1.328	1.729	2.093	2.539	2.861
20	.127	.257	.391	.533	.687	.860	1.064	1.325	1.725	2.086	2.528	2.845
21	.127	.257	.391	.532	.686	.859	1.063	1.323	1.721	2.080	2.518	2.831
22	.127	.256	.390	.532	.686	.858	1.061	1.321	1.717	2.074	2.508	2.819
23	.127	.256	.390	.532	.685	.858	1.060	1.319	1.714	2.069	2.500	2.807
24	.127	.256	.390	.531	.685	.857	1.059	1.318	1.711	2.064	2.492	2.797
25	.127	.256	.390	531	.684	.856	1.058	1.316	1 708	2.060	2.485	2.787
26	.127	.256	.390	.531	.684	.856	1.058	1.315	1.706	2056	2.479	2.779
27	127	.256	.389	.531	.684	.855	1.057	1.314	1.703	2 052	2.473	2.771
28	.127	.256	.389	.530	.683	.855	1.056	1.313	1.701	2.048	2.467	2.763
29	.127	.256	.389	.530	.683	.854	1.055	1.311	1.699	2 045	2.462	2.756
30	.127	.256	.389	.530	.683	.854	1.055	1.310	1.697	2.042	2.457	2.750
40	.126	.255	.388	.529	.681	.851	1.050	1.303	1.684	2.021	2.423	2.704
60	.126	254	.387	.527	.679	.848	1.046	1.296	1.671	2 000	2.390	2.660
120	.126	.254	.386	.526	.677	.845	1.041	1.289	1.685	1.980	2.358	2.617
∞	.126	.253	.385	.524	.674	.842	1.036	1.282	1.645	1.960	2.326	2.576

演習（5分）

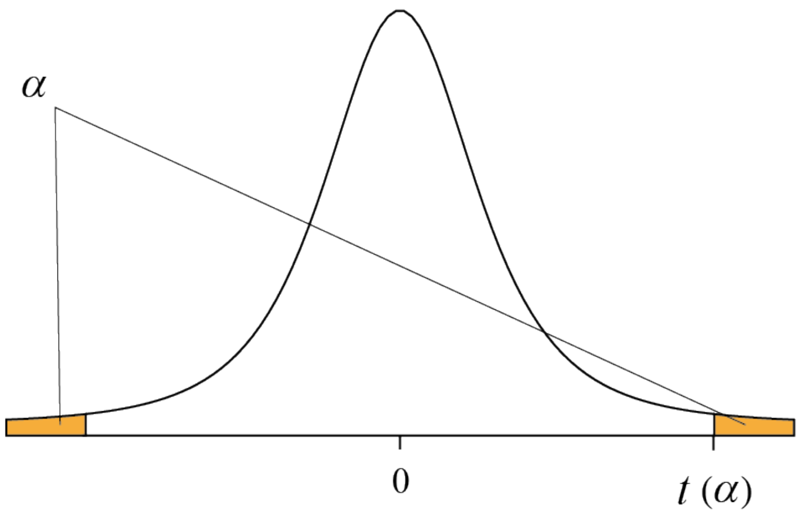
例 題 5.3

自由度が 20 の t 分布に従う確率変数 T の絶対値が, ある値を超える確率が 5% であるようにするには, その値をいくらにすればよいか?

演習（5分）

例 題 5.3

自由度が 20 の t 分布に従う確率変数 T の絶対値が, ある値を超える確率が 5% であるようにするには, その値をいくらにすればよいか?



$$P\left(|T| \geq t_\nu(\alpha)\right) = \alpha$$

$$\nu = 20$$

$$\alpha = 0.05$$

$$t_{20}(0.05) = 2.086$$

定 理 5.6（t-変換）　重要

$X_1, X_2, \cdots, X_n$：正規母集団 $N(\mu, \sigma^2)$ から取り出された n 個の無作為標本

標本平均： $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$　　不偏分散：$U^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

このとき，$T = \dfrac{\bar{X}-\mu}{U/\sqrt{n}}$ は自由度 $n-1$ の t-分布 t_{n-1} に従う．

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1), \qquad Y = \frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{(n-1)U^2}{\sigma^2} \sim \chi^2_{n-1}$$

Z と Y は独立なので，t-分布 t_{n-1} の定義から　$\dfrac{Z}{\sqrt{\dfrac{Y}{n-1}}} \sim t_{n-1}$

$$\frac{Z}{\sqrt{\dfrac{Y}{n-1}}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \times \frac{\sigma}{U} = \frac{\bar{X}-\mu}{U/\sqrt{n}}$$ なので　$\dfrac{\bar{X}-\mu}{U/\sqrt{n}} \sim t_{n-1}$

まとめ

正規母集団 $N(\mu,\sigma^2)$ → **無作為復元抽出** → 標本 $X_1, X_2, \dots, X_n$

標本平均： $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$　　　不偏分散：$U^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

定理 5.4　$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$　標準正規分布

定理 5.6　$\dfrac{\bar{X}-\mu}{U/\sqrt{n}} \sim t_{n-1}$　自由度 $n-1$ の t-分布

宿題

演習問題５（続き） (page 94-96)

5.1 確率変数 X が一様分布 $U(0,1)$ に従うとき，$Y = -2\log X$ は指数分布 $E_X\left(\frac{1}{2}\right)$ に従う，すなわち，自由度２のカイ２乗分布 ${\chi^2}_2$ に従うことを示せ.

5.6 Z は標準正規分布に従い，$z(\alpha)$ はその両側 α 点とする．T は自由度 ν のティー分布に従い，$t_\nu(\alpha)$ はその両側 α 点とする.

（１） 標準正規分布表により，$\alpha = 0.20,\ 0.10,\ 0.05$ に対して，$z(\alpha)$ の値を求めよ.

（２） ティー分布表により，$\alpha = 0.20,\ 0.10,\ 0.05$ と $\nu = 10,\ 20,\ 30$ に対して，$t_\nu(\alpha)$ の値を求めよ.

（３） これらの結果から，$z(\alpha) < t_\nu(\alpha)$ であることを確かめよ.

5.7 T が自由度 ν のティー分布に従うとき，T^2 はどのような分布に従うか.

Lecture 9

母比率・母平均の推定

【教科書】

第 6 章 推定

2 平均の区間推定

4 比率の推定

復習：母平均の推定

母集団

母平均 μ

無作為
復元抽出

標 本

$X_1, X_2, \ldots, X_n$

母平均の推定量として基本的なものは

標本平均　　$\bar{X} = \dfrac{1}{n}\sum_{k=1}^{n} X_k$

理由

（1）［不偏性］ $\mathbf{E}[\bar{X}] = \mu$

（2）［一致性］ $P\left(\lim_{n\to\infty} \bar{X} = \mu\right) = 1$

大数の法則

※ 標本平均をもって母平均の推定値とする＝点推定

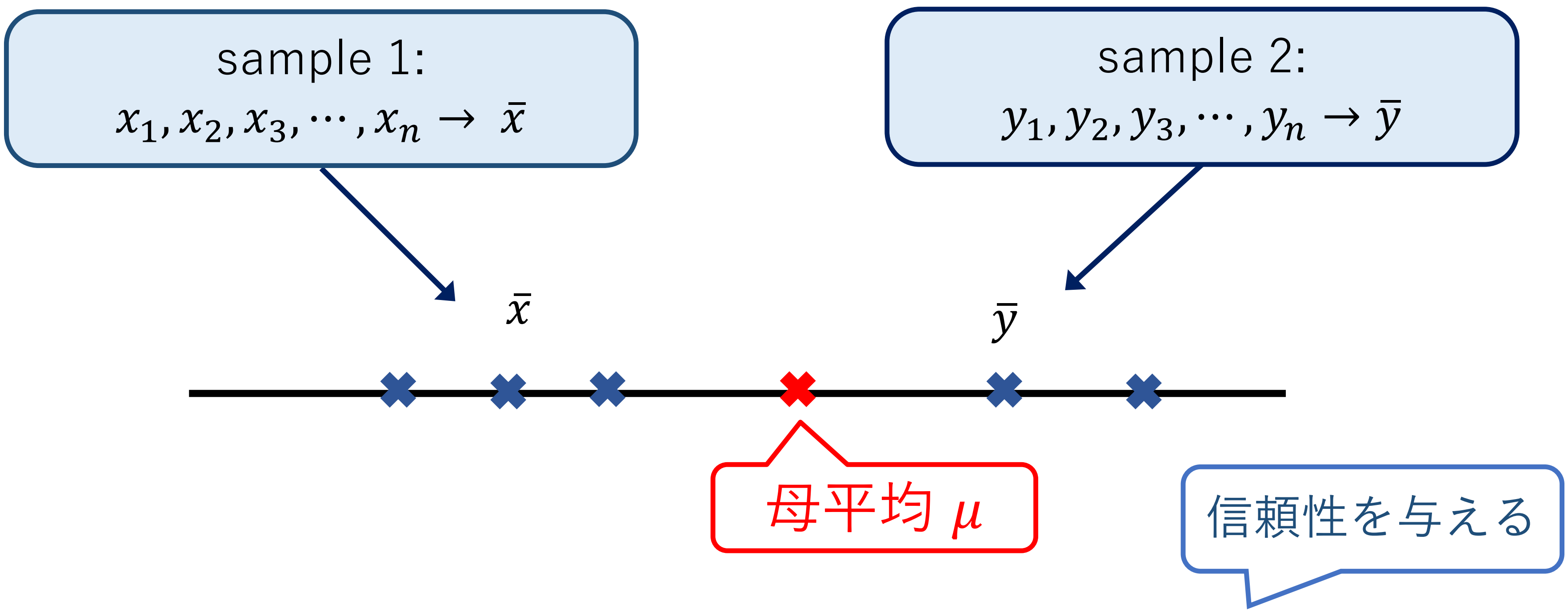
例えば，視聴率調査

ビデオリサーチ社：関東地区（ドラマ）

順位	番組タイトル	放送局	視聴率
1	連続テレビ小説・なつぞら （4/10）	NHK総合	23.1 %
2	特捜9 （4/10）	テレビ朝日	15.2 %
2	木曜ドラマ・緊急取調室 （4/11）	テレビ朝日	15.2 %
4	ラジエーションハウス・放射線科の診断レポート （4/8）	フジテレビ	12.7 %
5	日曜プライム「ドラマスペシャル アガサ・クリスティ　予告殺人」（4/14）	テレビ朝日	11.5 %
6	金曜ドラマ・インハンド （4/12）	TBS	11.3 %
7	白衣の戦士！ （4/10）	日本テレビ	10.3 %
8	緊急取調室（4/11）	テレビ朝日	9.9 %
9	いだてん～東京オリムピック噺～ （4/14）	NHK総合	9.6 %
10	特捜9[再]（4/10）	テレビ朝日	9.5 %

点推定の問題点

1回の標本調査で得られる標本平均 $\bar{x}$ が
母平均 μ に近いかどうか全く不明である.



- 得られた標本平均と母平均の差を確率的に評価

標本平均： $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$　　不偏分散： $U^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

標本平均に関する基本定理			
正規母集団	$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	標準正規分布	定理 5.4
正規母集団	$\dfrac{\bar{X}-\mu}{U/\sqrt{n}} \sim t_{n-1}$	自由度 $n-1$ の t-分布	定理 5.6
一般の母集団 n：大きい	$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ 近似的に成立	標準正規分布	定理 5.9 中心極限定理

母集団
母平均 μ
母分散 σ^2

無作為復元抽出

標 本
$X_1, X_2, \ldots, X_n$

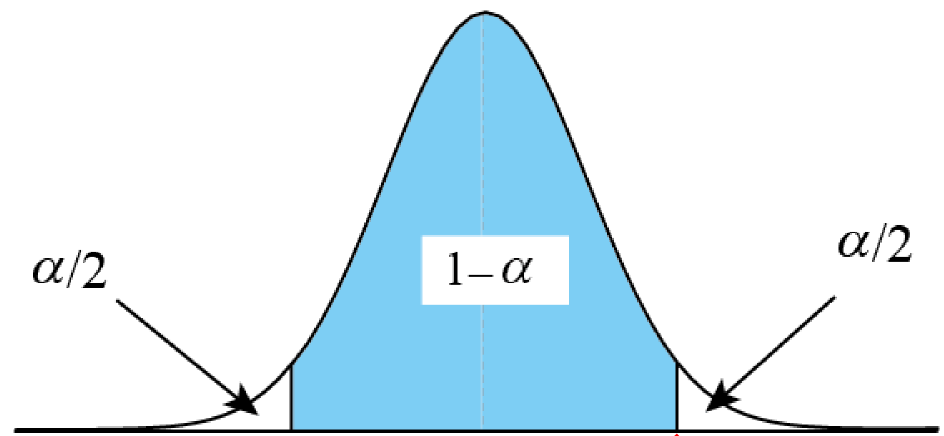
標本平均

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

前提

正規母集団または、
一般の母集団で n が大きい

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



両側 α点 $= z(\alpha)$

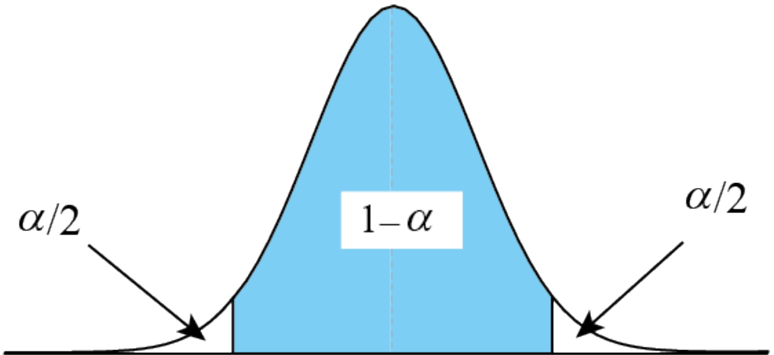
$$P\big(-z(\alpha) \leq Z \leq z(\alpha)\big) = 1 - \alpha$$

標本平均 $\bar{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 　標準化　 $Z=\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$

$$\underbrace{P\left(-z(\alpha)\le Z\le z(\alpha)\right)}=1-\alpha$$

$$-z(\alpha)\frac{\sigma}{\sqrt{n}}\le\bar{X}-\mu\le z(\alpha)\frac{\sigma}{\sqrt{n}}$$

$$\Leftrightarrow\ \bar{X}-z(\alpha)\frac{\sigma}{\sqrt{n}}\le\mu\le\bar{X}+z(\alpha)\frac{\sigma}{\sqrt{n}}$$



両側 α 点 $=z(\alpha)$

母平均 μ は区間 $\bar{X}\pm z(\alpha)\frac{\sigma}{\sqrt{n}}$ に確率 $1-\alpha$ で存在する

母平均 μ に対する信頼係数 $1-\alpha$ の信頼区間という

例 題 6.2

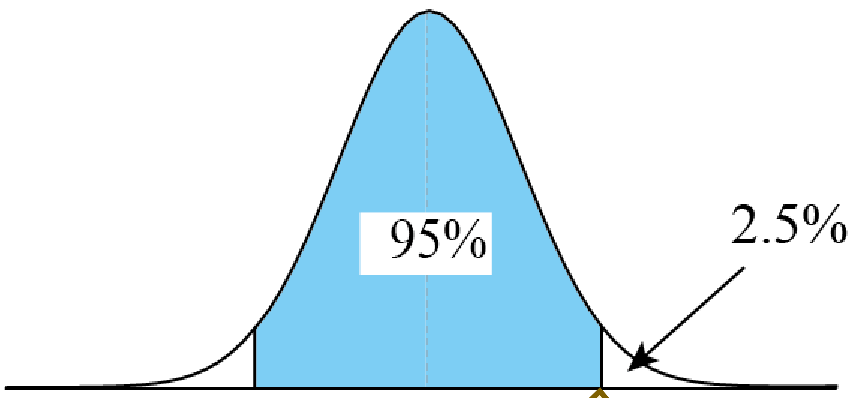
母分散が $\sigma^2 = 5^2$ である正規母集団から 10 個の無作為標本を抽出して標本平均値が $\bar{x} = 12.8$ であることを得た. 母平均の 95% 信頼区間を求めよ.

信頼係数 $1-\alpha$ の信頼区間は　$\bar{X} \pm z(\alpha)\dfrac{\sigma}{\sqrt{n}}$

95% 信頼区間 ⇒ $\alpha = 0.05$

$$12.8 \pm 1.96 \times \frac{5}{\sqrt{10}}$$

$$= 12.8 \pm 3.1$$



$z(0.05) = 1.96$

これまでのまとめ

母集団
母平均 μ
母分散 σ^2

無作為
復元抽出

標本

$X_1, X_2, \ldots, X_n$

母平均 μ

母平均の推定

標本平均 $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

正規母集団または、一般の母集団で n が大きいとき,
母平均 μ に対する信頼係数 $1-\alpha$ の信頼区間

$$\bar{X} \pm z(\alpha)\frac{\sigma}{\sqrt{n}}$$

σ が既知でないと使えない

母集団
母平均 μ
母分散 σ^2

無作為復元抽出 →

標本
$X_1, X_2, \ldots, X_n$

標本平均： $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$　　不偏分散： $U^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

標本平均に関する基本定理			
正規母集団	$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	標準正規分布	定理 5.4
正規母集団	$\dfrac{\bar{X}-\mu}{U/\sqrt{n}} \sim t_{n-1}$	自由度 $n-1$ の t-分布	定理 5.6
一般の母集団 n：大きい	$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ 近似的に成立	標準正規分布	定理 5.9 中心極限定理

母集団
母平均 μ
母分散 σ^2

無作為復元抽出 →

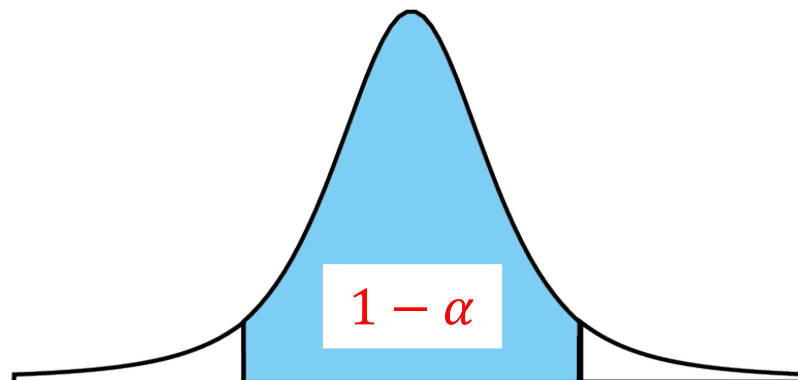
標本
$X_1, X_2, \ldots, X_n$

標本平均： $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$　　不偏分散： $U^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2$

正規母集団で母分散 σ^2 が未知なら
σ^2 の代わりに U^2 で置き換えて,

$$T = \frac{\bar{X} - \mu}{U/\sqrt{n}} \sim t_{n-1}$$

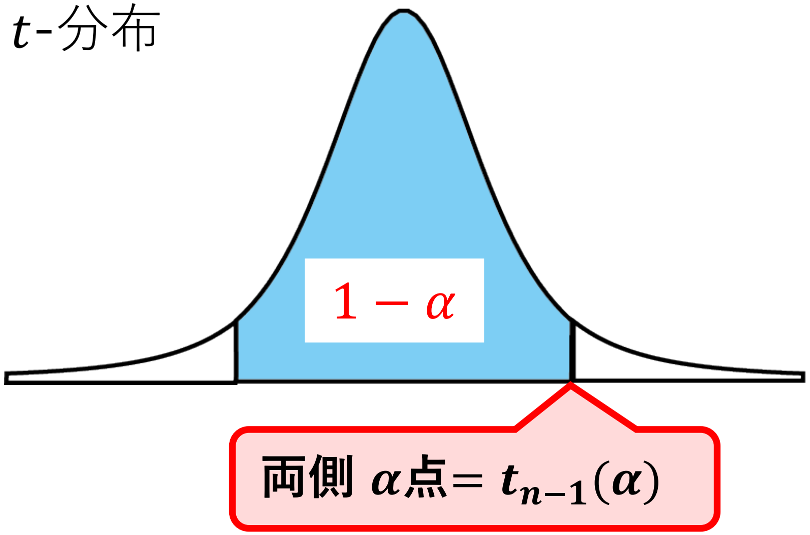
自由度 $n-1$ の t-分布



両側 α点$= t_{n-1}(\alpha)$

$$P\left(|T| \leq t_{n-1}(\alpha)\right) = 1 - \alpha$$

$$T = \frac{\bar{X} - \mu}{U/\sqrt{n}} \sim t_{n-1}$$ 自由度 $n-1$ の t-分布



$$P\Bigl(\underbrace{|T| \le t_{n-1}(\alpha)}\Bigr) = 1 - \alpha$$

$$-t_{n-1}(\alpha)\frac{U}{\sqrt{n}} \le \bar{X} - \mu \le t_{n-1}(\alpha)\frac{U}{\sqrt{n}}$$

$$\Leftrightarrow \ \bar{X} - t_{n-1}(\alpha)\frac{U}{\sqrt{n}} \le \mu \le \bar{X} + t_{n-1}(\alpha)\frac{U}{\sqrt{n}}$$

母平均 μ は区間 $\bar{X} \pm t_{n-1}(\alpha)\dfrac{U}{\sqrt{n}}$ に確率 $1-\alpha$ で存在する

母平均 μ に対する信頼係数 $1-\alpha$ の信頼区間という

まとめ

母集団		標本
母平均 μ 母分散 σ^2	無作為 復元抽出 →	$X_1, X_2, \ldots, X_n$

標本平均： $\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$　　不偏分散： $U^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

母平均 μ に対する信頼係数 $1-\alpha$ の信頼区間	
σ^2が既知	σ^2が未知
正規母集団または, 一般の母集団で n が大きい	正規母集団
$\dfrac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\dfrac{\bar{X}-\mu}{U/\sqrt{n}} \sim t_{n-1}$
$\bar{X} \pm z(\alpha)\,\dfrac{\sigma}{\sqrt{n}}$	$\bar{X} \pm t_{n-1}(\alpha)\,\dfrac{U}{\sqrt{n}}$

例 題 6.3

正規母集団から, 無作為標本を抽出して次のような 24 個のデータを得た. 母平均の 95% 信頼区間を求めよ.

35.9	43.9	51.2	35.3	36.7	49.4	39.5	59.6
43.8	32.9	36.0	43.0	41.9	44.6	47.2	56.2
45.6	47.7	38.1	51.8	42.3	46.6	35.5	32.4

信頼係数 $1-\alpha$ の信頼区間　$\bar{X} \pm t_{n-1}(\alpha)\,\dfrac{U}{\sqrt{n}}$

$$\bar{x} = \frac{1}{24}\sum x_i = 43.19 \qquad u^2 = \frac{1}{23}\sum (x_i - \bar{x})^2 = 54.231 = 7.36^2$$

95% 信頼区間　$43.19 \pm t_{23}(0.05) \times \dfrac{7.36}{\sqrt{24}} = 43.19 \pm 4.22$

問題 6.8（改）

練 習（10分）

ある製品の検査の所要時間は正規分布に従うといわれている. 大きさ 10 の無作為標本について, 次のデータを得た. 母平均の 95 % 信頼区間と90% 信頼区間を求めよ.

12.4　13.5　12.7　14.1　13.8　14.1　12.0　12.8　13.1　15.4

$$\bar{x} = \frac{1}{10}\sum x_i = 13.39$$

$$u^2 = \frac{1}{9}\sum (x_i - \bar{x})^2 = 1.0054 = 1.0027^2$$

$$\sum (x_i - \bar{x})^2 = 9.049$$

問題 6.8（改）

練 習（10分）

ある製品の検査の所要時間は正規分布に従うといわれている. 大きさ 10 の無作為標本について, 次のデータを得た. 母平均の 95 % 信頼区間と90% 信頼区間を求めよ.

12.4　13.5　12.7　14.1　13.8　14.1　12.0　12.8　13.1　15.4

$$\bar{x} = \frac{1}{10}\sum x_i = 13.39 \qquad u^2 = \frac{1}{9}\sum (x_i - \bar{x})^2 = 1.0054 = 1.0027^2$$

信頼係数 $1-\alpha$ の信頼区間は, $\bar{X} \pm t_{n-1}(\alpha)\ \frac{U}{\sqrt{n}}$

95 % 信頼区間

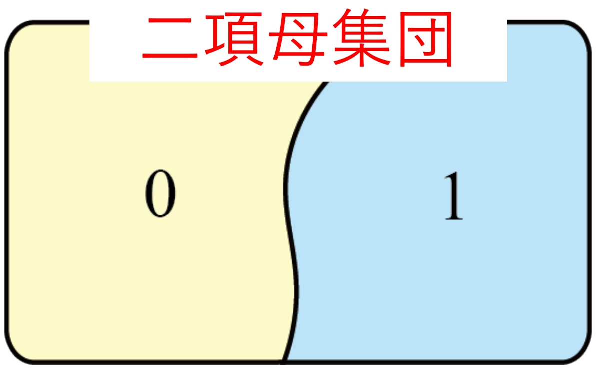
$$\bar{x} \pm t_9(0.05) \times \frac{1.0027}{\sqrt{10}} = 13.39 \pm 2.262 \times 0.317 = 13.39 \pm 0.717$$

90 % 信頼区間

$$\bar{x} \pm t_9(0.1) \times \frac{1.0027}{\sqrt{10}} = 13.39 \pm 1.833 \times 0.317 = 13.39 \pm 0.581$$

二項母集団の母比率の推定

- 二項母集団とは，０と１からなる母集団
- 母比率 p とは, 1 の割合



無作為復元抽出

母比率の推定

標 本

$X_1, X_2, \ldots, X_n$

母平均と母分散

$$m = p$$

$$\sigma^2 = p(1-p)$$

標本平均　＝標本比率

$$\bar{X} = \frac{1}{n}\sum_{k=1}^{n} X_k = \hat{p}$$

二項母集団
母比率 p

無作為復元抽出

標 本
$X_1, X_2, \ldots, X_n$

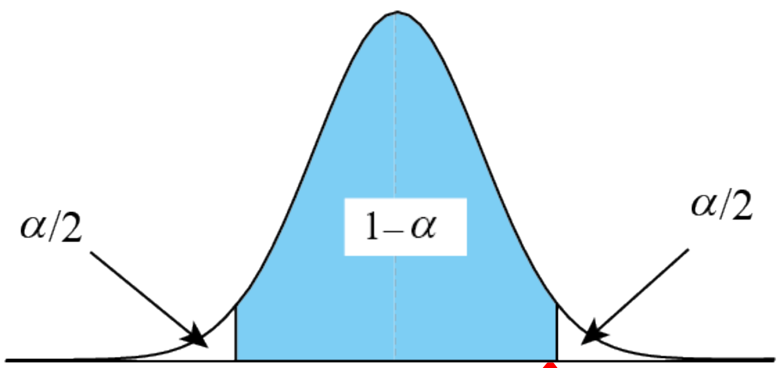
$\mu = p \quad \sigma^2 = p(1-p)$

標本平均＝標本比率

$$\bar{X} = \hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

n が大きいときは近似的に,

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



両側 α 点$= z(\alpha)$

$$P\big(-z(\alpha) \le Z \le z(\alpha)\big) = 1 - \alpha$$

二項母集団
母比率 p

無作為復元抽出

標 本
$X_1, X_2, \ldots, X_n$

$\mu = p \quad \sigma^2 = p(1-p)$

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

$$\bar{X} = \hat{p}$$

$$\underbrace{P\left(-z(\alpha) \le Z \le z(\alpha)\right)} = 1 - \alpha$$

$$\bar{X} - z(\alpha)\frac{\sigma}{\sqrt{n}} \le \mu \le \bar{X} + z(\alpha)\frac{\sigma}{\sqrt{n}}$$

いつも通り

$\sigma^2 = p(1-p) \approx \hat{p}(1-\hat{p})$ で近似（大数の法則）

$$\hat{p} - z(\alpha)\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}} \le p \le \hat{p} + z(\alpha)\frac{\sqrt{\hat{p}(1-\hat{p})}}{\sqrt{n}}$$

母比率の推定

二項母集団

母比率 p

無作為復元抽出

母比率の推定

標 本

$X_1, X_2, \ldots, X_n$

母平均と母分散　　$\mu = p$　　$\sigma^2 = p(1-p)$

標本平均＝標本比率　　$\hat{p} = \bar{X} = \dfrac{1}{n}\sum_{k=1}^{n} X_k$

母比率 p に対する信頼係数 $1-\alpha$ の信頼区間

$$\hat{p} \pm z(\alpha)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

例 関東地区の視聴率調査は 600 世帯を対象にしている.

順位	番組タイトル	放送局	視聴率
1	連続テレビ小説・なつぞら　(4/10)	NHK総合	23.1 %
2	特捜9　(4/10)	テレビ朝日	15.2 %
・・・・・			
9	いだてん～東京オリムピック噺～　(4/14)	NHK総合	9.6 %
10	特捜9[再]　(4/10)	テレビ朝日	9.5 %

95%信頼区間

$$\hat{p} \pm z(\alpha)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} = 0.231 \pm 1.96 \times \sqrt{\frac{0.231(1-0.231)}{600}} = 0.231 \pm 0.034$$

$$0.095 \pm 1.96 \times \sqrt{\frac{0.095(1-0.095)}{600}} = 0.095 \pm 0.023$$

例題 6.5　　練 習 （5分）

ある世論調査によると，500 人中 280 人が，ある候補者を支持するという結果を得た. この候補者の支持率の 95% 信頼区間を求めよ. また 90% 信頼区間を求めよ.

$\hat{p} = \frac{280}{500} = 0.56$　　信頼係数 $1-\alpha$ の信頼区間　$\hat{p} \pm z(\alpha)\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

95% 信頼区間

$$0.56 \pm 1.96 \times \sqrt{\frac{0.56(1-0.56)}{500}} = 0.56 \pm 1.96\sqrt{0.0004928} = 0.231 \pm 0.0435$$

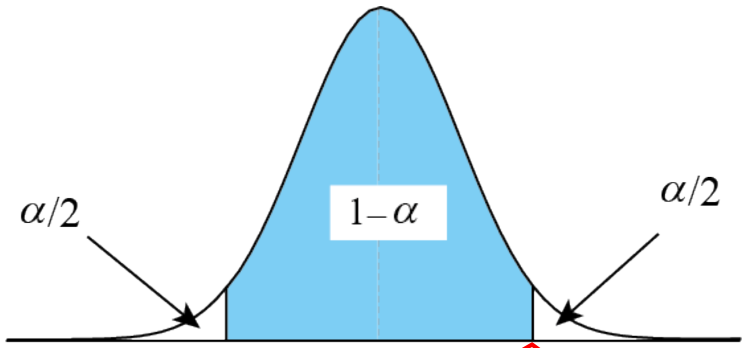
90% 信頼区間

$$0.56 \pm 1.64 \times \sqrt{\frac{0.56(1-0.56)}{500}} = 0.56 \pm 1.64\sqrt{0.0004928} = 0.231 \pm 0.0364$$

信頼係数と信頼区間の幅

$$\bar{X} \pm z(\alpha)\,\frac{\sigma}{\sqrt{n}}$$

$$\bar{X} \pm t_{n-1}(\alpha)\,\frac{U}{\sqrt{n}}$$



両側 $\boldsymbol{\alpha}$点
$= \boldsymbol{z(\alpha)}$ または $\boldsymbol{t_{n-1}(\alpha)}$

信頼係数 $1-\alpha$	0	小		大	1
α	1	大		小	0
信頼区間の幅	0	小		大	無限大 ∞

点推定

何も言わない

宿題

演習問題６（続き） (page 107-109)

6.7 正規母集団 $N(\mu, 0.5^2)$ から大きさ $n = 25$ の無作為標本を抽出し，標本平均 $\bar{x} = 18.26$ を得た．母平均 μ の 95% 信頼区間を求めよ．また，90% 信頼区間を求めよ．

6.8 ある製品の検査の所要時間は正規分布に従うといわれている．大きさ 10 の無作為標本について

12.4　13.5　12.7　14.1　13.8　14.1　12.0　12.8　13.1　15.4

のデータを得た（単位：分）．母平均 μ の 95% 信頼区間を求めよ．また，母分散 σ^2 の 95% 信頼区間を求めよ．

6.9 ある地区の 250 人において，あるテレビ番組に対する視聴率は 20% であった．視聴率の 90% 信頼区間を求めよ．また，95% 信頼区間を求めよ．

6.10 あるテレビ番組に対する視聴の有無を 180 名に聞いた．何名が視聴していれば，「視聴率が 20% であった」，ということができるか．

6.11 次のデータは鉛の融点を 12 回測定した結果である（単位：℃）．

327.1	325.5	336.8	324.2	328.5	321.0
332.2	321.8	317.1	337.8	324.0	326.8

（1） これらの測定値は正規分布 $N(\mu, 6.5^2)$ に従っているものとして，鉛の融点 μ の信頼区間を求めよ．ただし，信頼度は 95% とする．

（2） これらの測定値は正規分布 $N(\mu, \sigma^2)$ に従っているものとして，鉛の融点の信頼区間を求めよ．また，母分散 σ^2 の信頼区間を求めよ．

※ 母分散に対する信頼区間のためには χ^2 分布が必要なので省略

Lecture 10

仮説検定とは

【教科書】

第 7 章 検 定

1 検定の手順

2 平均の検定（途中まで）

典型例

コインを 100 回投げて, 表が 67 回出た. コインは公平といえるか?

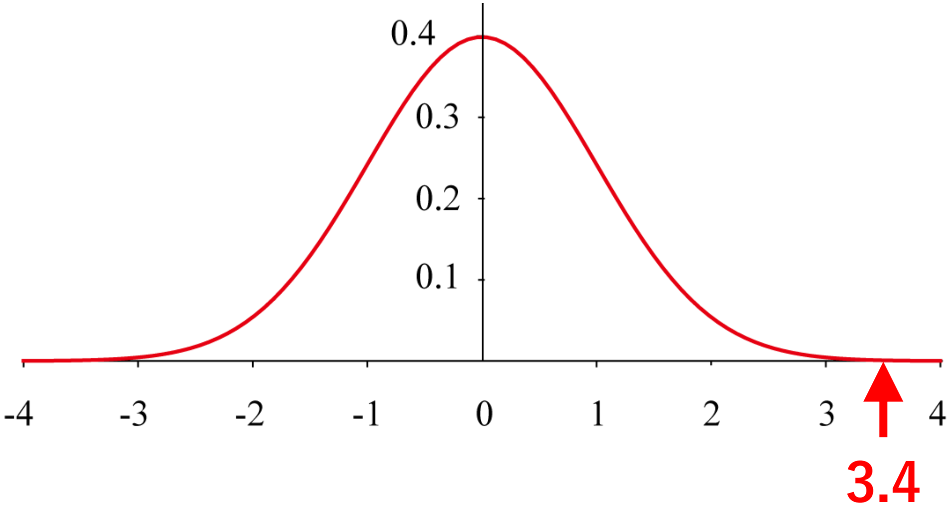
X ：100回投げて表の出る回数

$$X \sim B\left(100, \frac{1}{2}\right) \approx N(50, 5^2)$$

正規分布近似が便利

$$Z = \frac{X - 50}{5} \sim N(0,1)$$

実現値　$z = \dfrac{67 - 50}{5} = 3.4$



判断：かなり稀？

典型例

コインを 100 回投げて, 表が 54 回出た. コインは公平といえるか?

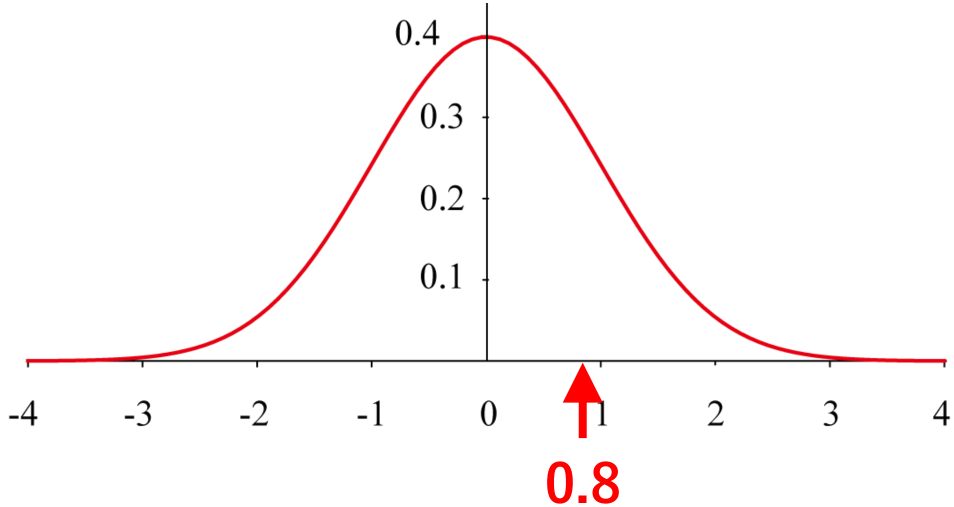
X ：100回投げて表の出る回数

$$X \sim B\left(100, \frac{1}{2}\right) \approx N(50, 5^2)$$

正規分布近似が便利

$$Z = \frac{X-50}{5} \sim N(0,1)$$

実現値　$z = \dfrac{54-50}{5} = 0.8$



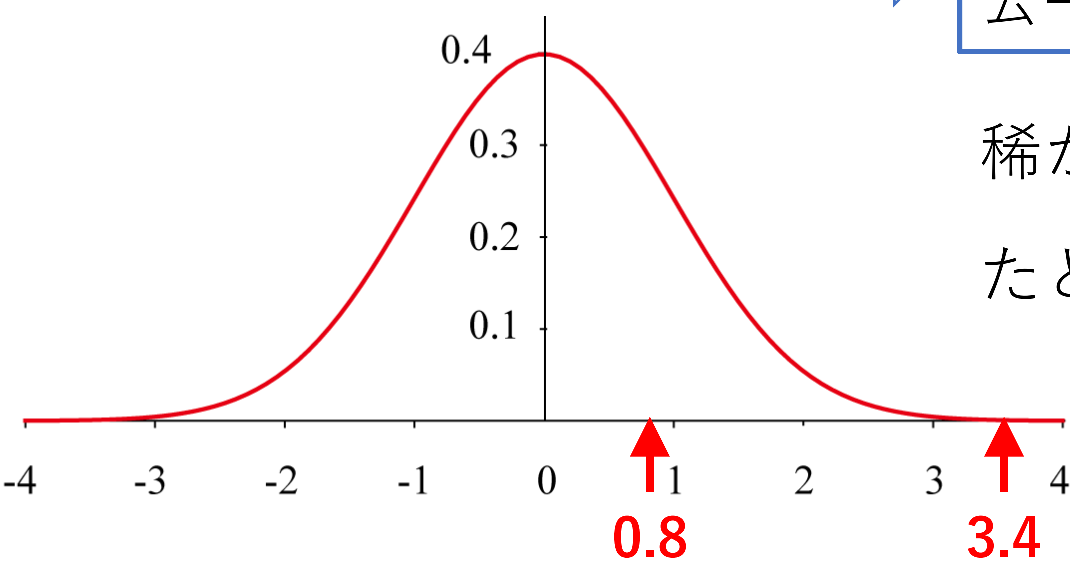
判断：ふつうに起こる？

仮説検定の考え方

表が 54 回　　$z = \frac{54-50}{5} = 0.8$　　偶然の揺らぎの範囲

表が 67 回　　$z = \frac{67-50}{5} = 3.4$　　稀なことが起こった？

いいえ、
公平なコインではなかった



稀かどうかの境目を決める

たとえば、$\alpha = 0.05$
または $\alpha = 0.01$

仮説検定の手順

(1) 母数に関する**帰無仮説**と**対立仮説**を決める.　H_0　H_1

(2) 関連する確率変数 T (**検定統計量**) を選び, H_0 の下で, この確率変数の分布を調べる

(3) **有意水準** $0 < \alpha < 1$ と**棄却域** W を決める.

(4) 標本から T の**実現値** t を計算する.

- $t \in W$ ⇒実現値は有意水準 α で**有意**である
  ⇒ H_0 を**棄却する**⇒ H_1 を採択する.
- $t \notin W$ ⇒実現値は有意水準 α で有意でない
  ⇒ H_0 を**棄却できない**
  (⇒ H_0 を採択する)

例　コインを 100 回投げて, 表が 67 回出た. コインは公正といえるか?

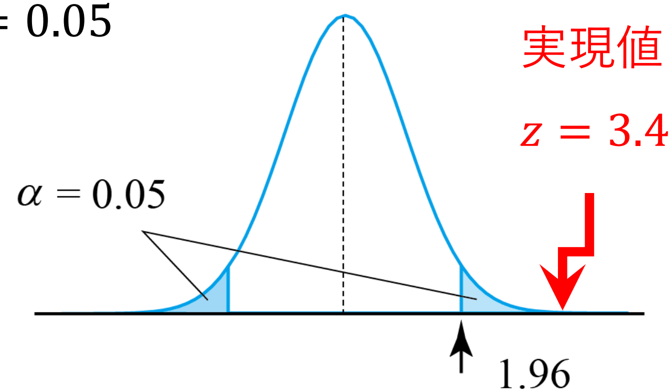
p ：表の出る確率

$$H_0 : p = \frac{1}{2} \qquad H_1 : p \neq \frac{1}{2}$$

X ：表の回数 $\sim B(100, 0.5) \approx N(50, 5^2)$

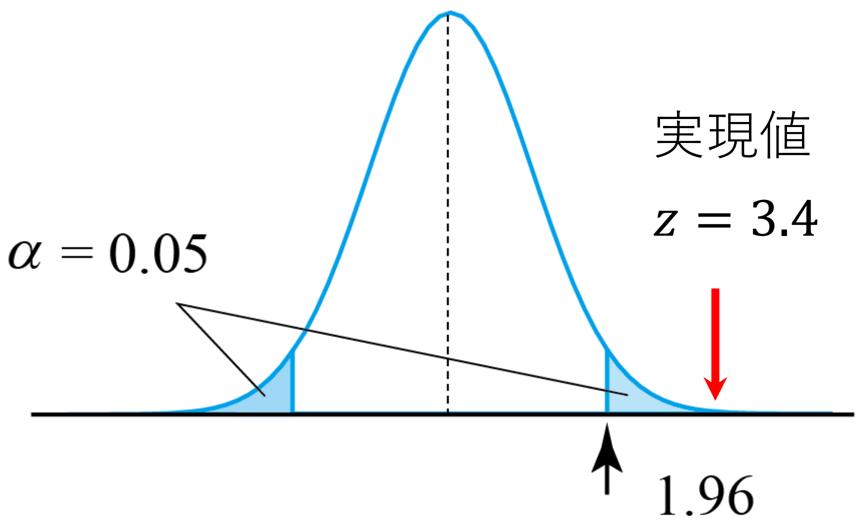
$$Z = \frac{X - 50}{5} \sim N(0,1)$$

$\alpha = 0.05$



$H_0 : p = \frac{1}{2}$ を棄却する

有意水準 α

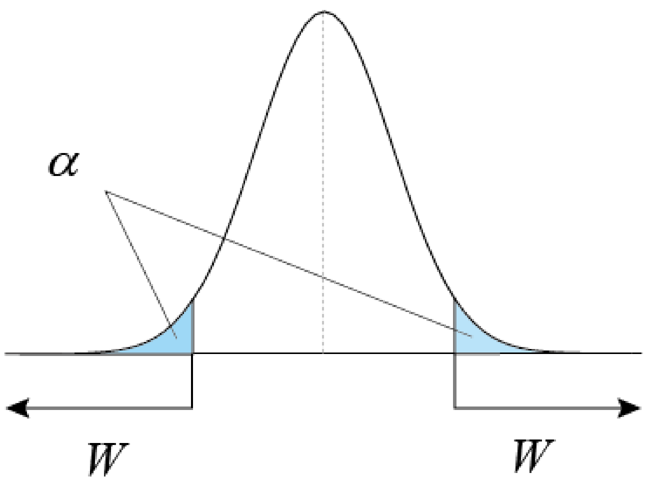


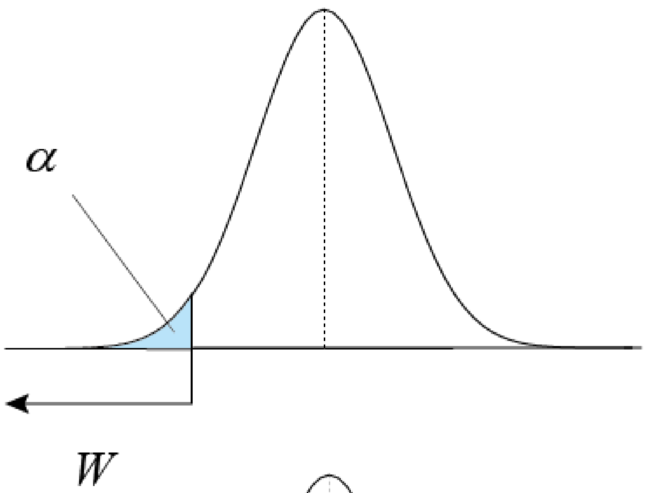
H_0: $p = \frac{1}{2}$ を棄却する

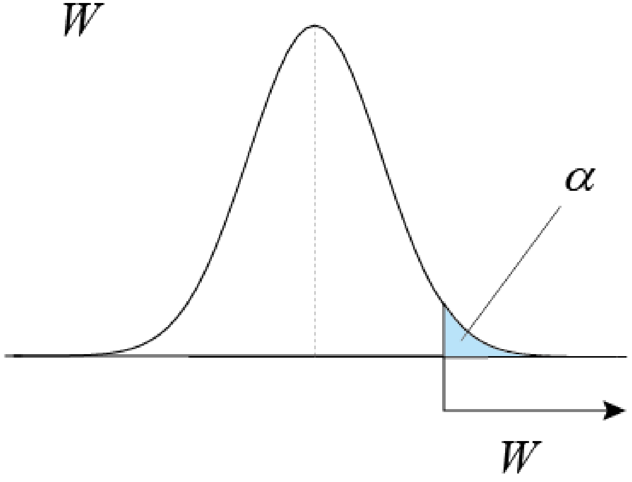
α は, 帰無仮説 H_0 が正しいのに, 検定の結論として帰無仮説を棄却してしまう誤り確率

※ 有意水準 α は文脈に応じて自分で設定する

棄却域の設定：両側検定と片側検定

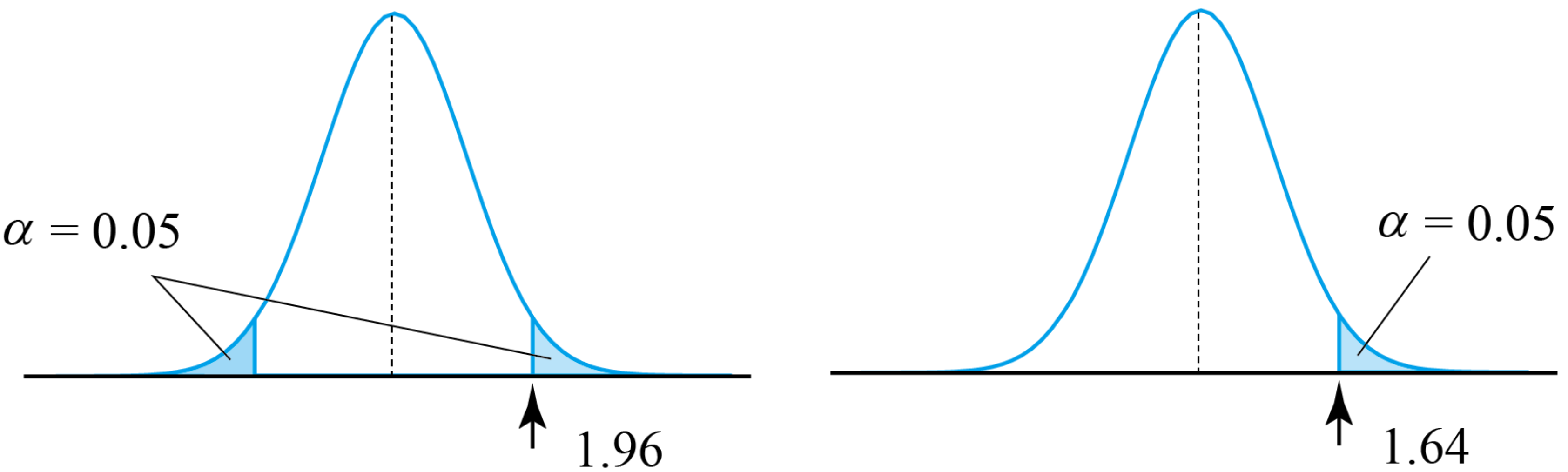






- 使い分けは文脈による
- 数理統計学の範疇ではない

両側 α 点と上側 α 点： $N(0,1)$ の場合



1.96 = 両側 5 % 点
　　= 上側 2.5 % 点

1.64 = 上側 5 % 点
　　= 両側 10 % 点

α	0.3173	0.1000	0.0500	0.0455	0.0100	0.0027	0.0010
$\alpha/2$	0.1587	0.0500	0.0250	0.0228	0.0050	0.0013	0.0005
z	1.000	1.645	1.960	2.00	2.576	3.000	3.290

母平均の検定

母集団

母平均 μ

母分散 σ^2

無作為
復元抽出

標本

大きさ：n

標本平均：$\bar{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$

不偏分散：$U^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

基本的な問題

母平均を$\mu = m_0$ とみなしてよいか？

帰無仮説と対立仮説

$H_0\colon \mu = m_0$　　$H_1\colon \mu \neq m_0$　（両側検定）

$H_1\colon \mu > m_0$　または　$H_1\colon \mu < m_0$　（片側検定）

母平均の分布（復習）

母集団

母平均 μ

母分散 σ^2

無作為
復元抽出

標本

大きさ：n

標本平均：$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

不偏分散：$U^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

母集団	基本定理	使う確率分布	参照
正規母集団	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	標準正規分布	定理 5.4
正規母集団	$\frac{\bar{X}-\mu}{U/\sqrt{n}} \sim t_{n-1}$	自由度 $n-1$ の t-分布	定理 5.6
一般の母集団 n：大きい	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ 近似的に成立	標準正規分布	定理 5.9 中心極限定理

例題 7.1

ある中学校で 1 年生 44 名に集団式知能検査を実施したところ，偏差値の平均は 52.4 であった．この学校の 1 年生は平均的な生徒といえるか．ただし，全国における知能検査の偏差値は $N(50, 10^2)$ に従うことが知られている．

母集団

母平均 μ

母分散 $\sigma^2 = 10^2$

標本

大きさ $n = 44$

標本平均
実現値

$\bar{x} = 52.4$

帰無仮説と対立仮説

H_0: $\mu = 50$　　生徒の知能は平均的である

H_1: $\mu \neq 50$　　生徒の知能は平均的ではない

例 7.1 （両側検定）

母集団
母平均 μ
母分散 $\sigma^2 = 10^2$

標本
大きさ $n = 44$

標本平均
実現値
$\bar{x} = 52.4$

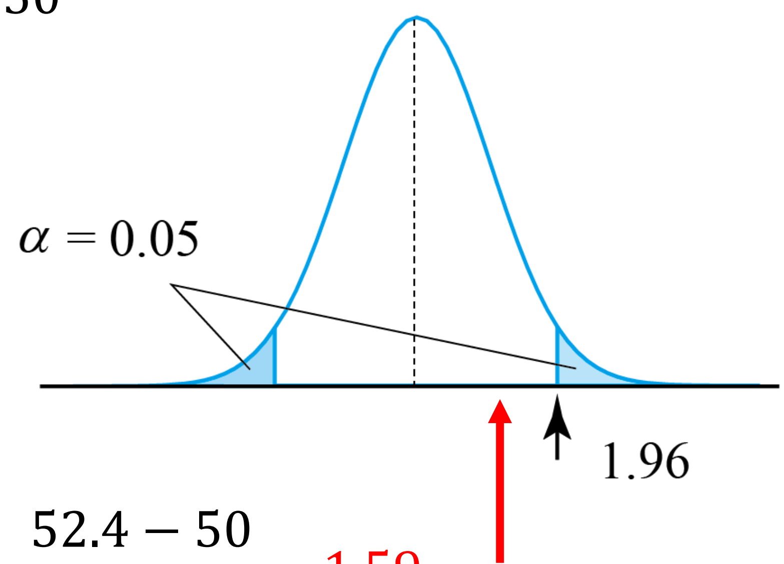
帰無仮説と対立仮説　$H_0\colon \mu = 50 \quad H_1\colon \mu \neq 50$

有意水準　$\alpha = 0.05$

検定統計量　H_0 の下で，

$$\bar{X} \sim N\left(50, \frac{10^2}{44}\right) = N(50, 1.51^2)$$

$$Z = \frac{\bar{X} - 50}{1.51} \sim N(0,1)$$



実現値　$z = \dfrac{52.4 - 50}{1.51} = 1.59$

結論　有意水準 5% の両側検定で H_0 は棄却されない．
$\mu = 50$ という判断である．

例題 7.2

あるメーカーの電化製品の寿命は, カタログによると平均 $\mu = 1200$ 時間, 標準偏差 $\sigma =\ 150$ 時間と書かれている. $n = 10$ 個のサン プルについてテスト したとき, 平均寿命が $\bar{x} = 1100$ 時間であった. カタログは偽りといえるか .

母集団

母平均 μ

母分散 $\sigma^2 = 150^2$

標本

大きさ $n = 10$

標本平均
実現値

$\bar{x} = 1100$

帰無仮説と対立仮説

H_0: $\mu = 1200$　カタログ通りである

H_1: $\mu < 1200$　カタログに偽りあり

$\mu > 1200$ でもカタログ通りでないといえるが、
実用の場面では問題ない
（まさに文脈の問題！）

例題 7.2 （片側検定）

母集団
母平均 μ
母分散 $\sigma^2 = 150^2$

標本
大きさ $n = 10$

標本平均
実現値
$\bar{x} = 1100$

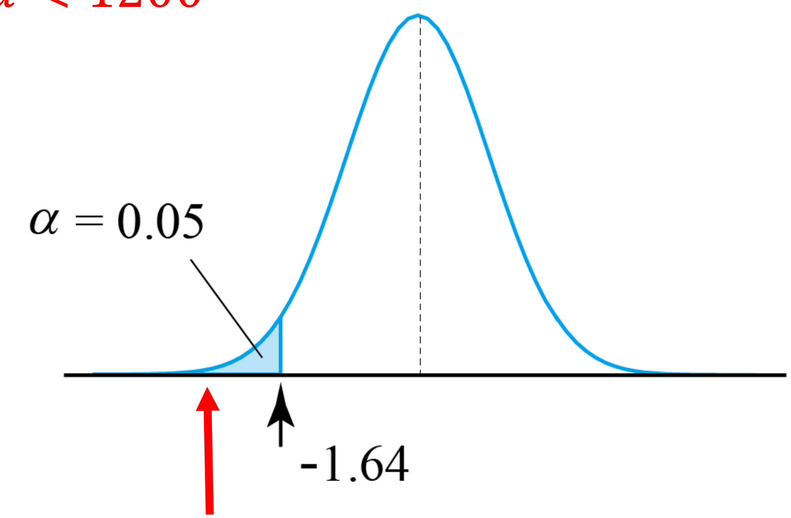
帰無仮説と対立仮説 $H_0 \colon \mu = 1200$　$H_1 \colon \mu < 1200$

有意水準 $\alpha = 0.05$

検定統計量 H_0 の下で,

$$\bar{X} \sim N\left(1200, \frac{150^2}{10}\right) = N(1200, 47.4^2)$$

$$Z = \frac{\bar{X} - 1200}{47.4} \sim N(0,1)$$



実現値 $$z = \frac{1100 - 1200}{47.4} = -2.11$$

結論 有意水準 5% の片側検定で H_0 は棄却される.

宿題

演習問題 7 (page 123-124)

7.1 母分散 $\sigma^2 = 15$ の正規分布に従うといわれる母集団から，標本数 30 の無作為標本を抽出し，標本平均 $\bar{x} = 56.75$ を得た．母平均を $\mu = 60$ とみなしてよいか．

7.2 正規母集団 $N(\mu, 6^2)$ から大きさ $n = 50$ の無作為標本を抽出したところ，標本平均は $\bar{x} = 28.4$ であった．帰無仮説 $H_0 : \mu = 30$ を次の 2 つの対立仮説に対してそれぞれ検定せよ．

（1） $H_1 : \mu \neq 30$　　（2） $H_1 : \mu < 30$

7.3 過去の経験から，ある製品の不良率は正規分布 $N(0.02, 0.04^2)$ に従うことがわかっている．製造方法を変更して，$n = 200$ の無作為標本について検査したところ，標本平均は $\bar{x} = 0.015$ に向上していた．効果があったといえるか．

Lecture 11

母比率・母平均の検定

【教科書】

第 7 章 検定
　2 平均の検定（途中から）
　4 比率の検定

母平均の検定

母集団

母平均 μ

母分散 σ^2

無作為
復元抽出

標本

大きさ：n

標本平均：$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

不偏分散：$U^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$

帰無仮説と対立仮説

H_0: $\mu = m_0$　　H_1: $\mu \neq m_0$（両側検定）

H_1: $\mu > m_0$　または　H_1: $\mu < m_0$（片側検定）

有意水準　$\alpha = 0.05$ など

検定統計量　$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$　または　$T = \frac{\bar{X} - \mu}{U/\sqrt{n}} \sim t_{n-1}$

例題 7.3

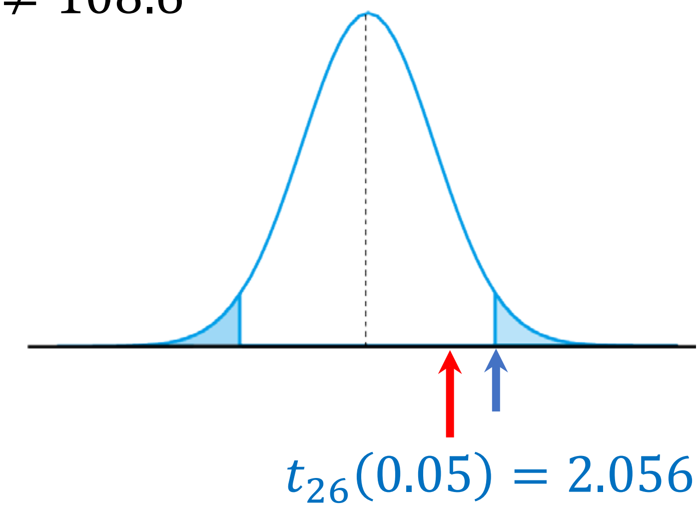
ある県での統計によると, 満 6 歳児の平均身長は 108.6 (cm) であるという. 同県のある小学校の 6 歳児 27 名について身長を調べ たところ, 平均 $\bar{x} = 109.7$ (cm), 不偏分散 $u^2 = 4.06^2$ (cm^2) であった．この結果から, 同校児童の身長は県平均に比べて高いといえるか.

帰無仮説と対立仮説　$H_0\colon \mu = 108.6 \quad H_1\colon \mu \neq 108.6$

有意水準　$\alpha = 0.05$

検定統計量　H_0 の下で,

$$T = \frac{\bar{X} - \mu}{U/\sqrt{n}} = \frac{\bar{X} - 108.6}{U/\sqrt{27}} \sim t_{26}$$



実現値　$$t = \frac{109.7 - 108.6}{4.06/\sqrt{27}} = 1.408$$

結論　有意水準 5% の両側検定で H_0 は棄却されない.

問題 7.4　　**練 習（10分）**

ある溶液に含まれる物質の濃度 (%) を測定して次のデータを得た.

12.6　13.4　14.1　12.4　11.2　12.5　10.9　11.8　11.6　13.1

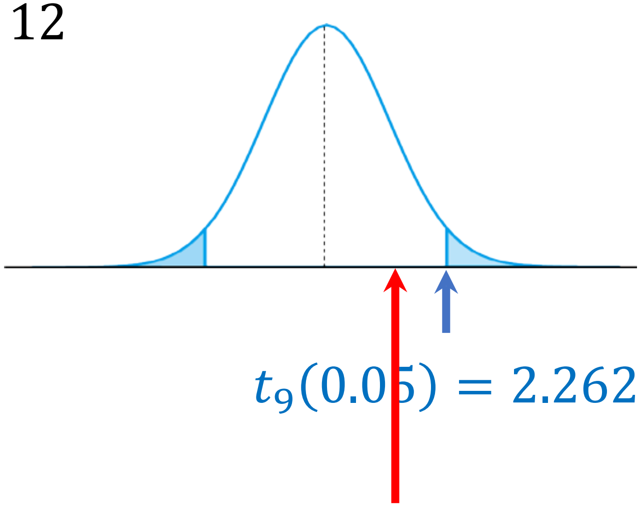
真の濃度を μ として，仮説 H_0 : $\mu =$ 12 を検定せよ.

帰無仮説と対立仮説　H_0: $\mu = 12$　　H_1: $\mu \neq 12$

有意水準　$\alpha = 0.05$

検定統計量　H_0 の下で,

$$T = \frac{\bar{X} - \mu}{U/\sqrt{n}} = \frac{\bar{X} - 12}{U/\sqrt{10}} \sim t_9$$

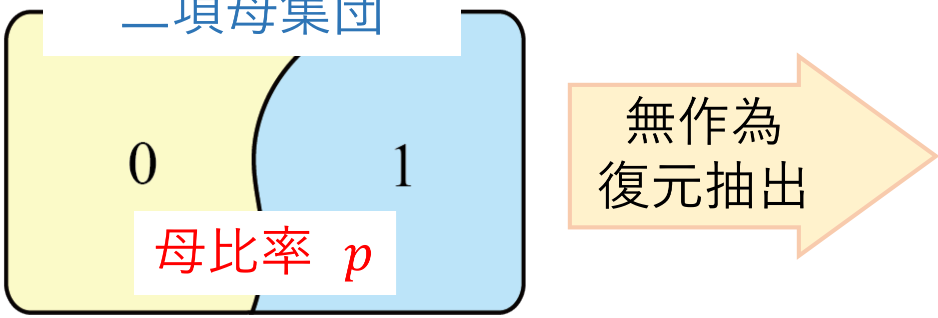


実現値　$\bar{x} = 12.36$　$u^2 = 1.0116 = 1.006^2$

$$t = \frac{12.36 - 12}{1.006/\sqrt{10}} = 1.132$$

結論　有意水準 5% の両側検定で H_0 は棄却されない.

二項母集団の母比率の検定



標 本

大きさ n

標本比率

$$\hat{p} = \frac{1}{n}\sum_{k=1}^{n} X_k$$

帰無仮説と対立仮説

$H_0\colon p = p_0$　　$H_1\colon p \neq p_0$　（両側検定）

$H_1\colon\ p > p_0$　または　$H_1\colon p < p_0$　（片側検定）

検定統計量

H_0 の下で,

$$n\hat{p} \sim B(n, p_0) \approx N(np_0, np_0(1-p_0))$$

$$\hat{p} \sim N\left(p_0, \frac{p_0(1-p_0)}{n}\right) \qquad Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1)$$

例題 7.5

あるクラスの平均出席率は 0.90 であるといわれている. ある日の欠席者は, 160 人中 25 人であった. この日は通常の出席ではないといえるか. 有意水準は $\alpha = 0.01$ とせよ.

帰無仮説と対立仮説　H_0: $p = 0.9$　H_1: $p \neq 0.9$（両側検定）

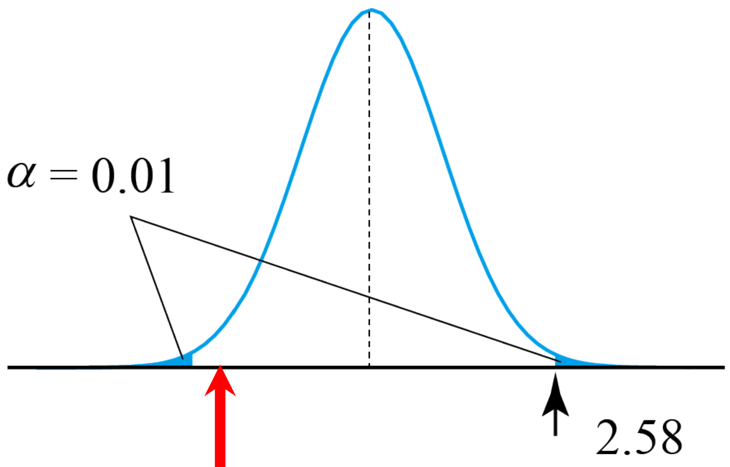
有意水準　$\alpha = 0.01$

検定統計量　H_0 の下で,

$$n\hat{p} \sim B(160, 0.9) \approx N(144, 3.79^2)$$

$$\hat{p} \sim N(0.9, 0.30^2)$$

$$z = \frac{\hat{p} - 0.9}{0.30/\sqrt{160}} = \frac{\hat{p} - 0.9}{0.0237} \sim N(0,1)$$



実現値　$z = \dfrac{0.8438 - 0.9}{0.0237} = -2.37$

結論　有意水準 1% の両側検定で H_0 は棄却されない.

問題 7.9

練 習（10分）

ある意見項目に対する賛成率を 30% は欲しいと思われていた．実際に，調査では 80 人中 23 人の賛成を得た．賛成率の目標を達成したと考えてよいか．

帰無仮説と対立仮説　H_0: $p = 0.3$　H_1: $p < 0.3$（片側検定）

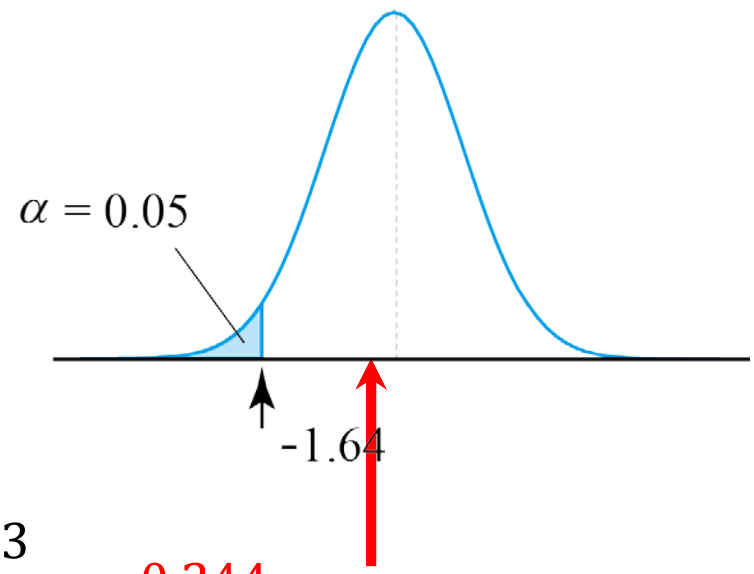
有意水準　$\alpha = 0.05$

検定統計量　H_0 の下で，

$$n\hat{p} \sim B(80, 0.3) \approx N(24, 4.10^2)$$

$$\hat{p} \sim N(0.3, 0.458^2)$$

$$z = \frac{\hat{p} - 0.3}{0.458/\sqrt{80}} = \frac{\hat{p} - 0.3}{0.0512} \sim N(0,1)$$



実現値　$z = \dfrac{0.2875 - 0.3}{0.0512} = -0.244$

結論　有意水準 5% の片側検定で H_0 は棄却されない．

2 種類の過誤

帰無仮説 H_0 をめぐって

採否\真偽	H_0 は真	H_0 は偽
H_0 を採択		
H_0 を棄却		

2 種類の過誤

帰無仮説 H_0 をめぐって

採否\真偽	H_0 は真	H_0 は偽
H_0 を採択	○	
H_0 を棄却		○

２種類の過誤

帰無仮説 H_0 をめぐって

採否\真偽	H_0 は真	H_0 は偽
H_0 を採択	○	第２種の誤り β
H_0 を棄却	第１種の誤り α	○

第１種の誤り＝生産者危険＝あわて者の間違い

第２種の誤り＝消費者危険＝ぼんやり者の間違い

第１種の誤り確率＝有意水準 α ← 自分で設定する

第２種の誤り確率 β ＝一般には算出できず，評価できない

宿題

演習問題７（続き） (page 123-124)

7.5 ある植物の生育は平均 $\boldsymbol{\mu} = \mathbf{15.4}$ の正規分布 $\boldsymbol{N}(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ に従うことがわかっている ある年，成長促進剤を施したところ，標本数 $\boldsymbol{n} = \mathbf{24}$ のデータで，標本平均 $\bar{\boldsymbol{x}} = \mathbf{18.4}$, 不偏分散 $\boldsymbol{u}^2 = \mathbf{6.8}^2$ であった. この促進剤の効果はあったといえるか.

7.6 正規母集団から無作為標本として

31.5	32.4	30.6	34.1	29.1	33.2	31.5	30.8	32.8	34.3

を得た．このデータより，次の仮説を検討せよ．ただし，対立仮説は両側仮説とする．

（1） $H_0 : \mu = 33$ （2） $H_0 : \sigma^2 = 2.80$ （分散の検定は省略）

7.9 ある意見項目に対する賛成率を 30% は欲しいと思われていた．実際に, 調査では 80 人中 23 人の賛成を得た．賛成率の目標を達成したと考えてよいか.

7.10 メンデルの法則によれば, ある花の栽培において, 2 種類の花が 3 : 1 の割合で生ずるという．実際に 217本 栽培した結果, 花が 156 : 61 の割合で発生した．この結果はメンデルの法則に従っているといえるか.

Lecture 12

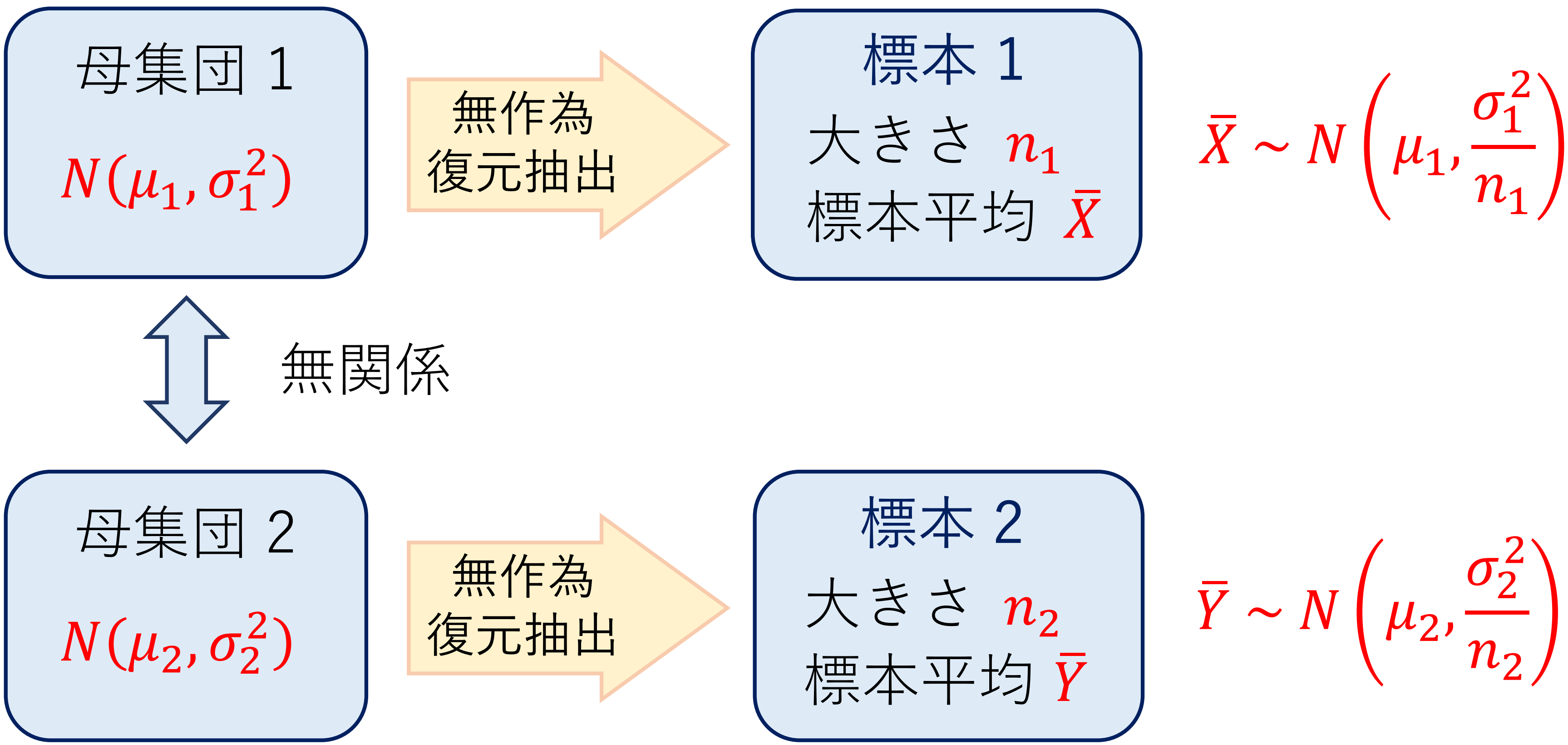
母集団の比較

【教科書】

第８章　２標本問題（一部だけを紹介）

母平均の差の検定

※ 2つの母集団の母平均に差があるかを検定する



確率変数の和（復習）

- 一般の確率変数 $X,\ Y$ と定数 a, b に対して

（1）［平均値の線形性］$E[aX + bY] = aE[X] + bE[Y]$

（2）分散は線形性をもたないが $V[aX] = a^2V[X]$

- 確率変数 $X,\ Y$ が独立ならば

（3）［平均値の乗法性］$E[XY] = E[X]E[Y]$

（4）［分散の加法性］$V[X + Y] = V[X] + V[Y]$

- 確率変数 $X \sim N(m_1, \sigma_1^2),\ Y \sim N(m_2, \sigma_2^2)$ が独立ならば

（5）線形結合 $aX + bY$ も正規分布に従い，

$$aX + bY \sim N(am_1 + bm_2, a^2\sigma_1^2 + b^2\sigma_2^2)$$

定理 （標本平均の差の分布）

正規母集団 $N(\mu_1, \sigma_1^2)$ から取り出した n_1 個の無作為標本の標本平均を $\bar{X}_1$，別の正規母集団 $N(\mu_2, \sigma_2^2)$ から取り出した n_2 個の無作為標本の標本平均を $\bar{X}_2$ とするとき，

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2\ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

母集団 1
$N(\mu_1, \sigma_1^2)$

標本 1
大きさ n_1
標本平均 $\bar{X}_1$

$$\bar{X}_1 \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$$

無関係

独立

母集団 2
$N(\mu_2, \sigma_2^2)$

標本 2
大きさ n_2
標本平均 $\bar{X}_2$

$$\bar{X}_2 \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$$

$\bar{X}_1 - \bar{X}_2$ について

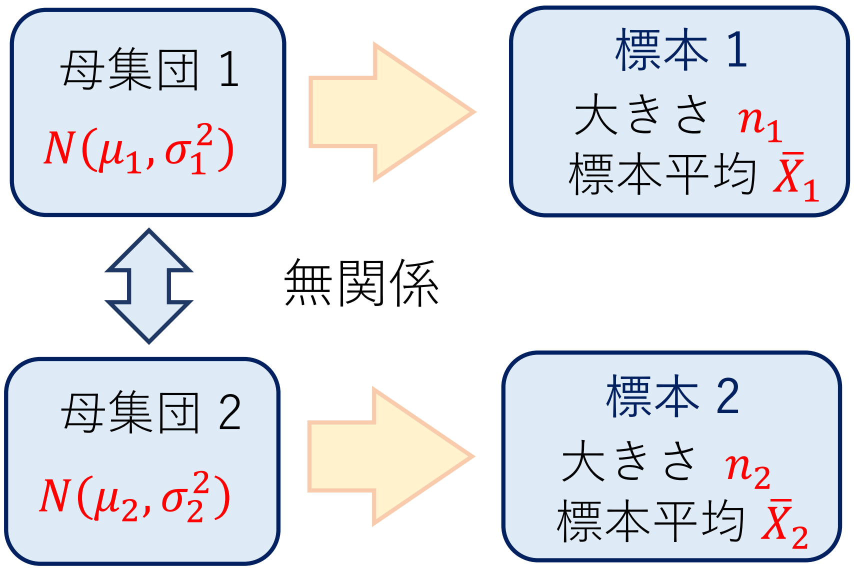
平均値

$$\mu_1 - \mu_2$$

分散

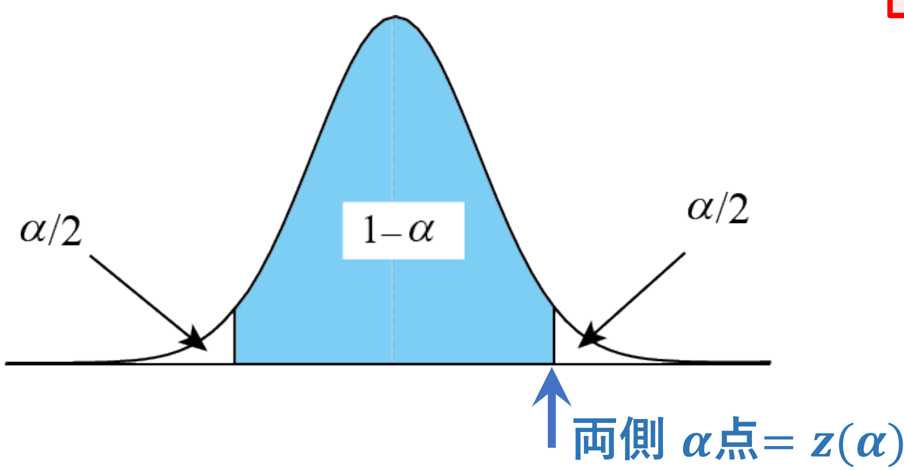
$$\frac{\sigma_1^2}{n_1} + (-1)^2 \frac{\sigma_2^2}{n_2}$$

標本平均の差の推定と検定（母分散既知）



$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2 \ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

$$Z = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

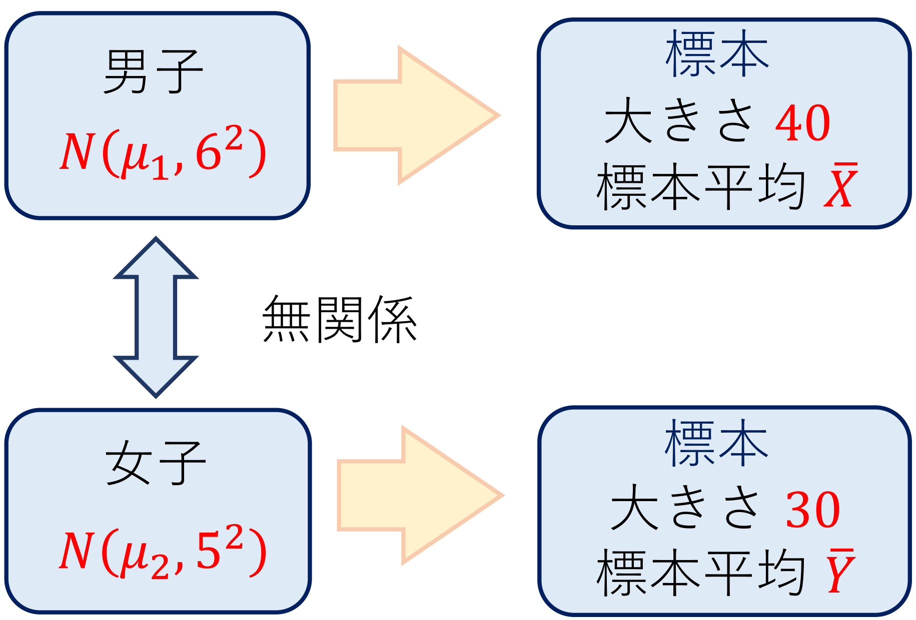


$$P\big(-z(\alpha) \leq Z \leq z(\alpha)\big) = 1 - \alpha$$

例題 8.6　（母分散既知の場合）

ある大学の男子学生 40 人と女子学生 30 人の身長を調べたところそれぞれ 171.47, 157.22 (cm) であった. それぞれの分散は既知であって 36 , 25 (cm^2)であるとして, 男女学生の身長差の 95% 信頼区間を求めよ.

男子学生, 女子学生の身長は正規分布に従うものとして, それぞれの平均値をそれぞれ μ_1, μ_2 とおく.



$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

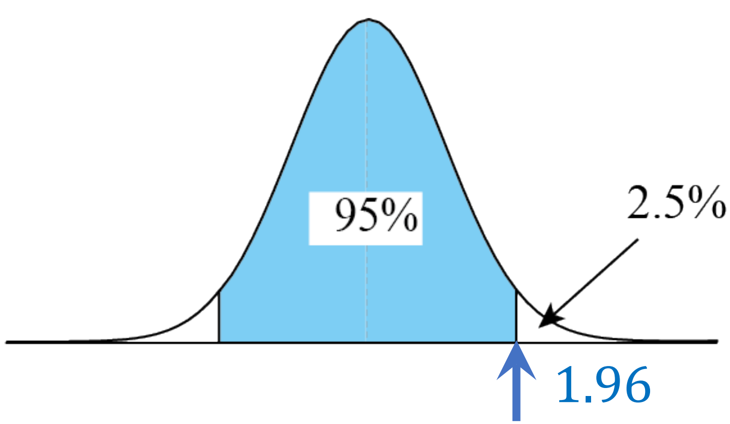
$$= \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{6^2}{40} + \dfrac{5^2}{30}}}$$

$$= \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{1.317} \sim N(0,1)$$

例題 8.6 （母分散既知の場合）

ある大学の男子学生 40 人と女子学生 30 人の身長を調べたところそれぞれ 171.47, 157.22 (cm) であった. それぞれの分散は既知であって 36 , 25 (cm^2)であるとして, 男女学生の身長差の 95% 信頼区間を求めよ.

$$Z = \frac{\bar{X} - \bar{Y} - (\mu_1 - \mu_2)}{1.317} \sim N(0,1)$$



$$P(-1.96 \le Z \le 1.96) = 0.95$$

$$-1.96 \le Z \le 1.96$$

$$\Updownarrow$$

$$\bar{X} - \bar{Y} - 1.96 \times 1.317$$
$$\le \mu_1 - \mu_2 \le \bar{X} - \bar{Y} + 1.96 \times 1.317$$

実現値　$\bar{x} = 171.47,\ \ \bar{y} = 157.22$

$\mu_1 - \mu_2$ に対する 95% 信頼区間

$$14.25 \pm 2.58$$

例題 8.7　（母分散既知の場合）

ある学年で知能指数を測定し, 男女別に集計したところ次の結果が得られた. 男女差があるといえるか．ただし, 知能指数の分布は $N\left(100, 15^2\right)$ といわれている.

	平均	人数
男生徒	103	40
女生徒	101	35

男生徒, 女生徒の平均値をそれぞれ μ_1, μ_2 とおく. 分散は 15^2 を用いる.

帰無仮説と対立仮説　$H_0\colon \mu_1 = \mu_2 \qquad H_1\colon \mu_1 \neq \mu_2$

有意水準　$\alpha = 0.05$

検定統計量　$\bar{X} \sim N\left(\mu_1, \dfrac{15^2}{40}\right) \qquad \bar{Y} \sim N\left(\mu_2, \dfrac{15^2}{35}\right)$

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{15^2}{40} + \frac{15^2}{35}\right) = N\left(0, 3.47^2\right)$$

男生徒, 女生徒の平均値をそれぞれ μ_1, μ_2 とおく. 分散は 15^2 を用いる.

帰無仮説と対立仮説　$H_0\colon \mu_1 = \mu_2 \qquad H_1\colon \mu_1 \neq \mu_2$

有意水準　$\alpha = 0.05$

検定統計量

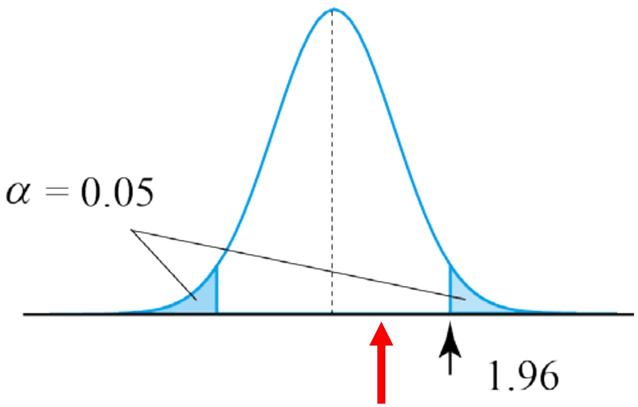
$$\bar{X} \sim N\left(\mu_1, \frac{15^2}{40}\right) \qquad \bar{Y} \sim N\left(\mu_2, \frac{15^2}{35}\right)$$

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{15^2}{40} + \frac{15^2}{35}\right) = N(0, 3.47^2)$$

$$Z = \frac{\bar{X} - \bar{Y}}{3.47} \sim N(0,1)$$

実現値

$$z = \frac{103 - 101}{3.47} = 0.576$$



結論

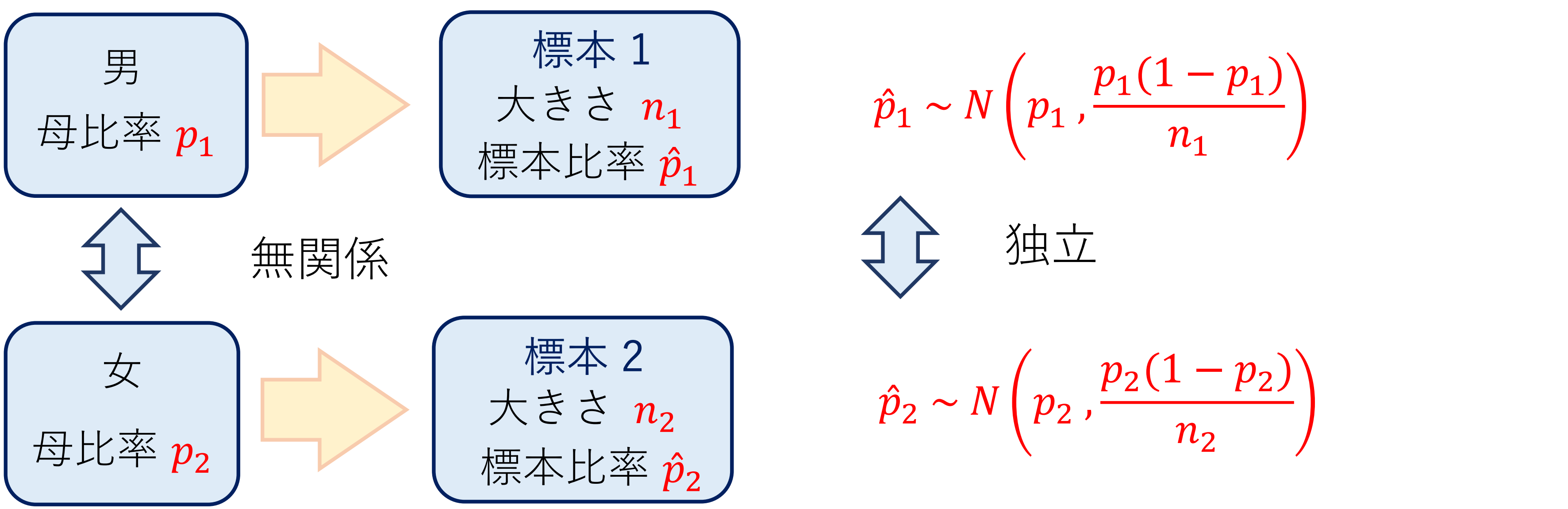
有意水準 $\alpha = 0.05$ の両側検定によって H_0 は棄却されない.

例題 8.10（母比率の差）

ある意見項目に対する賛否を男女別に集計したところ，次の結果を得た．賛成者の比率に男女差があるといえるか．

	賛成	反対	計
男	58 (0.592)	40 (0.408)	98 (1.000)
女	28 (0.394)	43 (0.606)	71 (1.000)

男, 女の賛成の母比率をそれぞれ p_1, p_2 とおく．

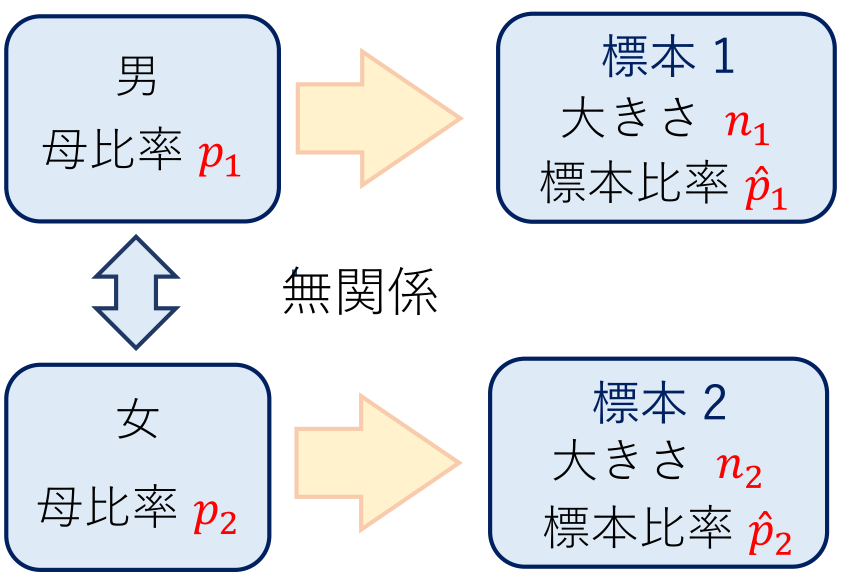


例題 8.10（母比率の差）

ある意見項目に対する賛否を男女別に集計したところ，次の結果を得た．賛成者の比率に男女差があるといえるか．

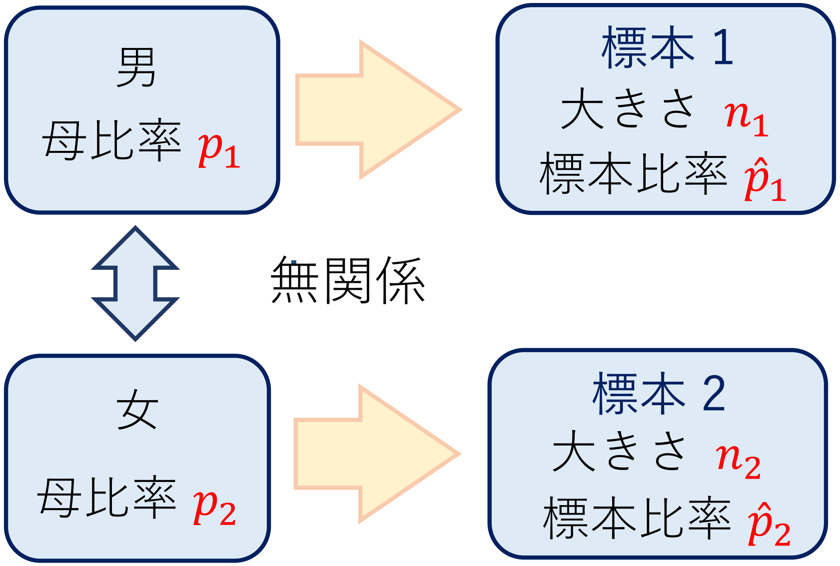
	賛成	反対	計
男	58 (0.592)	40 (0.408)	98 (1.000)
女	28 (0.394)	43 (0.606)	71 (1.000)

男, 女の賛成の母比率をそれぞれ p_1, p_2 とおく．



$$\hat{p}_1 - \hat{p}_2$$

$$\sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$



$$\hat{p}_1 - \hat{p}_2 \sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

教科書の記述は不適切

帰無仮説と対立仮説　$H_0\colon p_1 = p_2 = p$　　$H_1\colon p_1 \neq p_2$

有意水準　$\alpha = 0.05$

検定統計量　$\hat{p}_1 - \hat{p}_2 \sim N\left(0, \frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}\right) = N(0, 0.078^2)$

$$n_1 = 98 \quad n_2 = 71 \quad p = \frac{58+28}{98+71} = 0.509$$

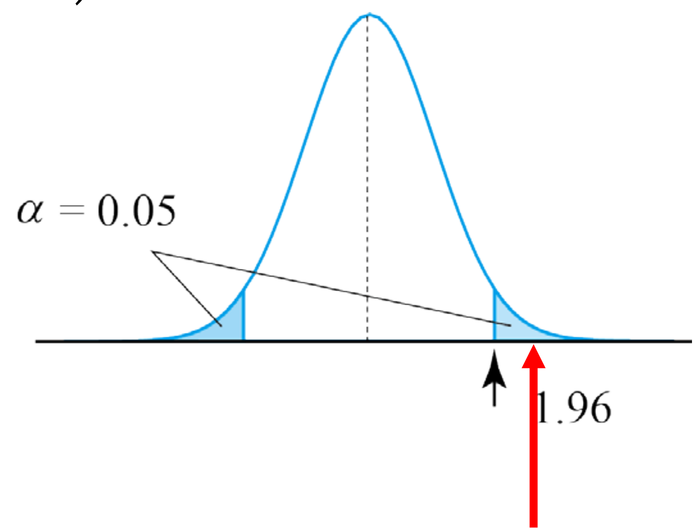
帰無仮説と対立仮説　$H_0\colon p_1 = p_2 = p \qquad H_1\colon p_1 \neq p_2$

有意水準　$\alpha = 0.05$

検定統計量

$$\hat{p}_1 - \hat{p}_2 \sim N\left(0, \frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}\right) = N(0, 0.078^2)$$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{0.078} \sim N(0,1)$$



実現値　$\hat{p}_1 = \frac{58}{98} = 0.592 \quad \hat{p}_2 = \frac{28}{71} = 0.394$

$$z = \frac{0.592 - 0.394}{0.078} = 2.54$$

結論　有意水準 $\alpha = 0.05$ の両側検定によって H_0 は棄却される.

標本平均の差の推定と検定（一般）

母集団 1
$N(\mu_1, \sigma_1^2)$

標本 1
大きさ n_1
標本平均 $\bar{X}_1$

無関係

母集団 2
$N(\mu_2, \sigma_2^2)$

標本 2
大きさ n_2
標本平均 $\bar{X}_2$

$$\bar{X}_1 - \bar{X}_2 \sim N\left(\mu_1 - \mu_2\ , \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right)$$

この式は一般の場合でも正しいが，
母分散未知の場合は使えない．

母分散未知

等分散の検定

F 検定

等分散の場合
$\sigma_1^2 = \sigma_2^2$

T 検定

等分散でない
$\sigma_1^2 \neq \sigma_2^2$

高度な議論

問題 8.8

練習（10分）

インフルエンザについて, 予防接種の有無と感染の有無のデータが次のようであった. 予防接種の効果はあるといえるか.

	感染	非感染
接種	18	67
非接種	45	65

接種, 非接種群の感染率（母比率）をそれぞれ p_1, p_2 とおく.

接種
母比率 p_1

標本 1
大きさ n_1
標本比率 $\hat{p}_1$

$$\hat{p}_1 \sim N\left(p_1, \frac{p_1(1-p_1)}{n_1}\right)$$

無関係

独立

非接種
母比率 p_2

標本 2
大きさ n_2
標本比率 $\hat{p}_2$

$$\hat{p}_2 \sim N\left(p_2, \frac{p_2(1-p_2)}{n_2}\right)$$

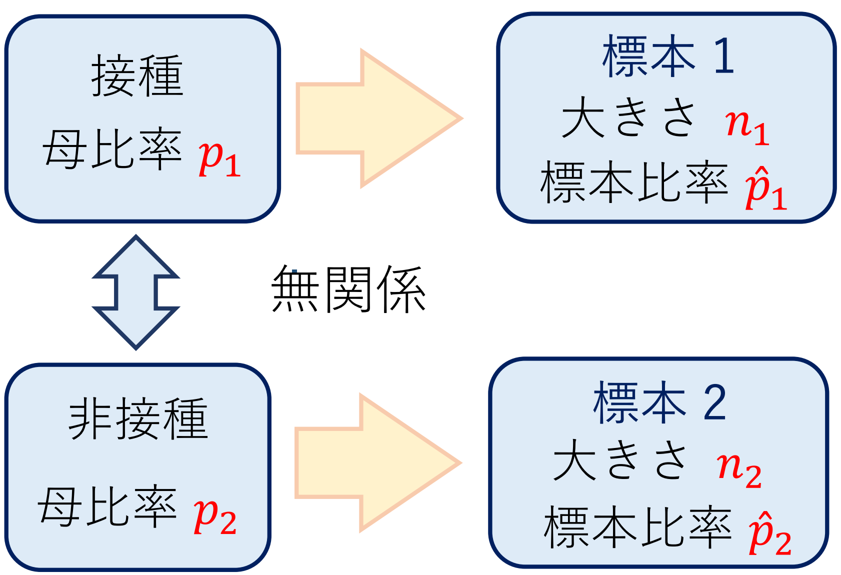
問題 8.8

練習（10分）

インフルエンザについて, 予防接種の有無と感染の有無のデータが次のようであった. 予防接種の効果はあるといえるか.

	感染	非感染
接種	18	67
非接種	45	65

接種, 非接種群の感染率（母比率）をそれぞれ p_1, p_2 とおく.



$$\hat{p}_1 - \hat{p}_2$$

$$\sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

接種
母比率 p_1

標本 1
大きさ n_1
標本比率 $\hat{p}_1$

無関係

非接種
母比率 p_2

標本 2
大きさ n_2
標本比率 $\hat{p}_2$

$$\hat{p}_1 - \hat{p}_2$$

$$\sim N\left(p_1 - p_2, \frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}\right)$$

帰無仮説と対立仮説　$H_0\colon p_1 = p_2 = p$　　$H_1\colon p_1 \neq p_2$

有意水準　$\alpha = 0.01$

検定統計量　$\hat{p}_1 - \hat{p}_2 \sim N\left(0, \frac{p(1-p)}{n_1} + \frac{p(1-p)}{n_2}\right) = N(0, 0.0675^2)$

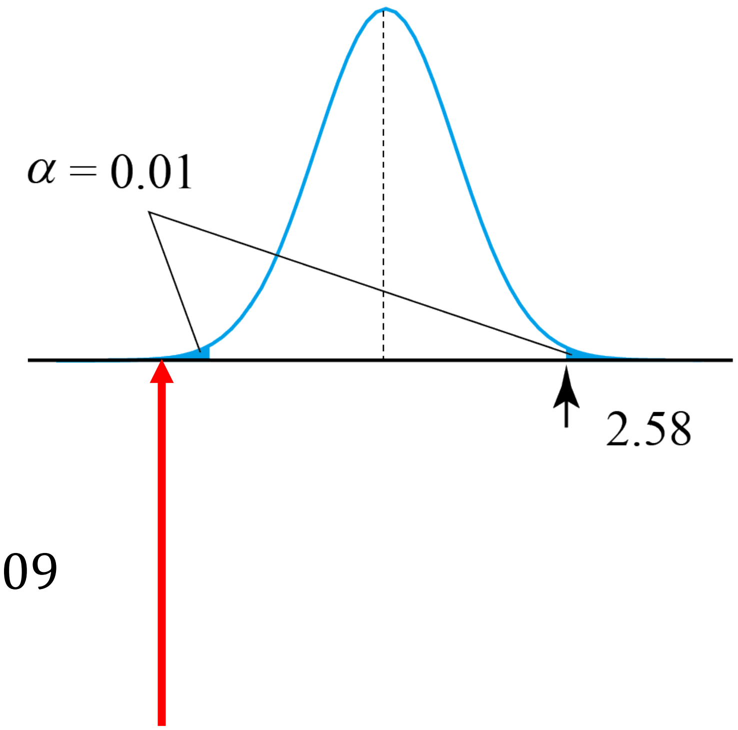
$$n_1 = 85 \qquad n_2 = 110 \qquad p = \frac{18+45}{85+110} = 0.323$$

帰無仮説と対立仮説　$H_0\colon p_1 = p_2 \qquad H_1\colon p_1 \neq p_2$

有意水準　$\alpha = 0.01$

検定統計量　$\hat{p}_1 - \hat{p}_2 \sim N(0, 0.0675^2)$

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{0.0675} \sim N(0,1)$$



実現値　$\hat{p}_1 = \dfrac{18}{85} = 0.212 \qquad \hat{p}_2 = \dfrac{45}{110} = 0.409$

$$z = \frac{0.212 - 0.409}{0.0675} = -2.91$$

結論　有意水準 $\alpha = 0.01$ の両側検定によって H_0 は棄却される.
高度に有意である.

演習問題８（抜粋, 一部改変） (page 141-143)

8.4 A 校から無作為に生徒 14 名を抽出し，同じく B 校から無作為に 12 名を選び，知能指数を比較したところ，A 校の平均 $\bar{x} = 115$，B 校の平均 $\bar{y} = 108$ を得た．両校に差があるといえるか. ただし, A校, B 校ともに知能指数は標準偏差 15 の正規分布に従うものとしてよい.

8.5 ある県下で大規模な一斉テストの結果は平均 54.2 であった．長年優秀受験校として実績のある A 校からは 54 人，普通校の B 校からは 38 人が受験しており，A 校は平均 61.2，B 校は平均 55.8 であった.，経験的に A 校の標準偏差は 8.1，B校の標準偏差は 12.8 であった.

(1) A 校の生徒は優秀 といえるか.

(2) A, B 両校の平均の差を検定せよ .

8.10 次のデータは 2 つの新聞 A, B の地域ごとの購読数である．両新聞の購読傾向はどのようなものであるか.

新聞	住宅地	商業地
A	28	47
B	37	68