1-6　数学基礎（1-6-1から1-6-4）

東京大学 数理・情報教育研究センター

2021年5月7日

概要

- 数学や統計学は，データおよびAIの利活用のために必須の道具でもあります．この節では，その数学および統計学の基礎を学びます．

- 「ものの集まり」を表す集合の概念を学んだあと，統計学の基礎概念としてデータを要約・可視化する方法について学びます．

- また，統計学において重要な役割を果たす，確率を学びます．そして確率の考え方を用いて，一部のデータから全体の傾向を推測する「推測統計学」について学びます．

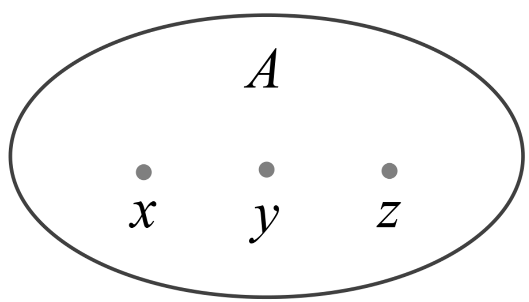
本教材の目次

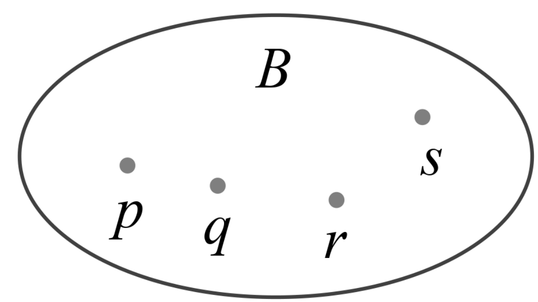
本教材の目次

1-6-1　集合と順列

集合の概念

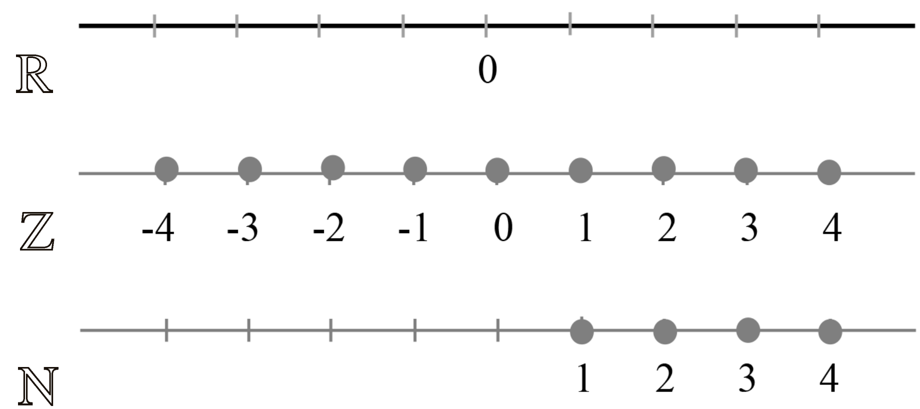
- 集合とは，ある特定の「もの」の集まり のことです.
- 集合 A を構成する「もの」のことを 要素または元(げん)とよびます.
    - x が A の要素であるとき，$x \in A$ または $A \ni x$ と書きます.
    - p が A の要素でないとき，$p \notin A$ または $A \not\ni p$ と書きます.
- 有限個の要素からなる集合を有限集合とよび，
  無限個の要素からなる集合を無限集合とよびます.





基本的な数の集合

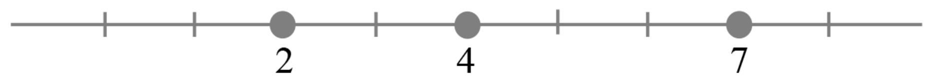
- 実数の集合 $\mathbb{R}$.
- 整数の集合 $\mathbb{Z}$.
- 自然数の集合 $\mathbb{N}$.
- 有理数の集合 $\mathbb{Q}$.



- これらは，すべて 無限集合です.
- 例： $16 \in \mathbb{N}$, $\quad 24.2 \notin \mathbb{N}$, $\quad 24.2 \in \mathbb{Q}$, $\quad \sqrt{5} \notin \mathbb{Q}$, $\quad \sqrt{5} \in \mathbb{R}$.

集合の表し方 (1/2)

- 例として，$\{2, 4, 7\}$ は,「3 つの実数 $2, 4, 7$ からなる集合」のことです.
  - 要素を書き並べる順番を変えても，集合としては同じものです.
    - つまり，$\{2, 4, 7\} = \{2, 7, 4\} = \{7, 4, 2\} = \{4, 7, 2\} = \cdots$.



- $\{5, 6, 7, \ldots, 30\}$ は,「5 以上 30 以下の自然数からなる集合」のことです.
- $\{1, 2, 3, \ldots\} = \mathbb{N}$.
- $\{0, \pm 1, \pm 2, \pm 3, \ldots\} = \mathbb{Z}$.

集合の表し方 (2/2)

- 「5 以上 30 以下の自然数からなる集合」は，$\{5, 6, 7, \ldots, 30\}$ と表しました.
  - これとは別の表し方として，

$$\{x \mid x \text{ は 5 以上 30 以下の自然数 }\},$$

または，$\{x \in \mathbb{N} \mid 5 \leqq x \leqq 30\}$ もあります.

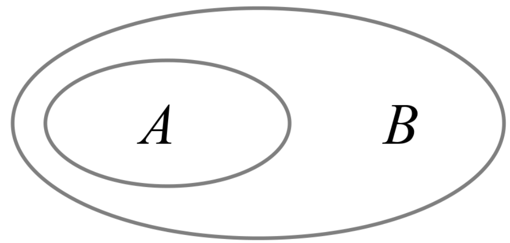
- $\{x \in \mathbb{R} \mid 5 \leqq x \leqq 30\}$ は,「5 以上 30 以下の実数からなる集合」です.
- $\{2i \mid i \in \mathbb{Z}\}$ は，偶数の集合です.
- $\{2i - 1 \mid i \in \mathbb{Z}\}$ は，奇数の集合です.

空集合

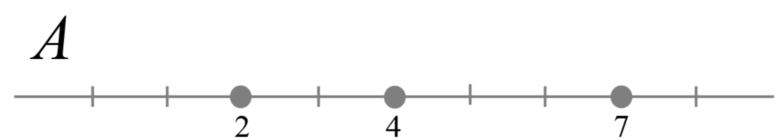
- 要素を 1 つももたない集合を，空集合とよびます．
  - 空集合は，$\emptyset$ や $\{\}$ などで表します．
- 例： $\{x \in \mathbb{R} \mid x^2 = -1\} = \emptyset$.
- 例： $\{x \in \mathbb{N} \mid x \leqq -5\} = \emptyset$.

部分集合

- 集合 A の要素がすべて集合 B の要素でもあるとき，A を B の部分集合とよびます.
  - このとき，$A \subset B$ または $B \supset A$ と書きます.



- $A \subset B$ かつ $A \supset B$ のとき，$A = B$ と書きます.

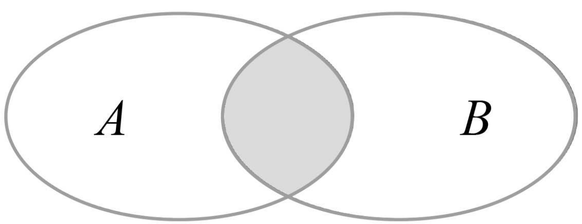


- 例： $A = \{2, 4, 7\}$ のとき，
  - $A \subset \{2, 4, 6, 7\}$,　$A \subset \mathbb{N}$,　$\{2, 4\} \subset A$.
  - $\{2, 5\}$ は，A の部分集合ではありません.
  - A の部分集合は，全部で 8 個あります：

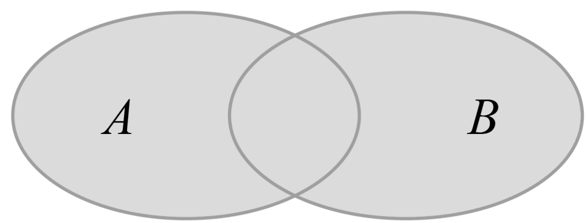
$$\emptyset\ ,\ \{2\}\ ,\ \{4\}\ ,\ \{7\}\ ,\ \{2, 4\}\ ,\ \{4, 7\}\ ,\ \{2, 7\}\ ,\ \{2, 4, 7\}\ .$$

共通部分，和集合，差集合 (1/2)

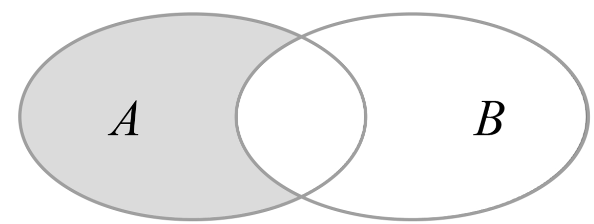
- 2 つの集合 A と B に対して，次の集合が定義できます.
- A と B の共通部分（または，交わり）$A \cap B$ とは，
  A と B の両方に含まれる要素をすべて集めた集合のことです.
- A と B の和集合 $A \cup B$ とは，
  A の要素と B の要素をすべて集めた集合のことです.
- A と B の差集合 $A - B$ とは，
  A に含まれるが B に含まれない要素をすべて集めた集合のことです.



$A \cap B$



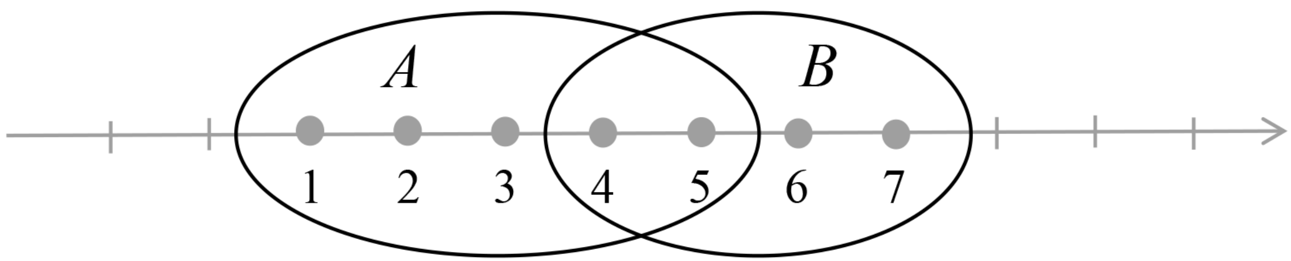
$A \cup B$



$A - B$

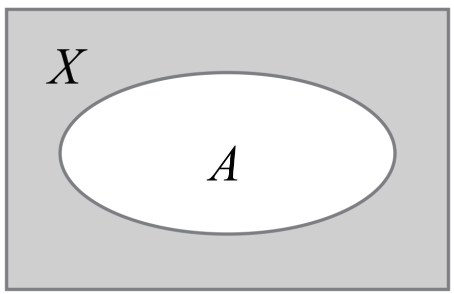
共通部分，和集合，差集合 (2/2)

- 例： $A = \{1, 2, 3, 4, 5\}$, $B = \{4, 5, 6, 7\}$ に対して，
  - $A \cap B = \{4, 5\}$.
  - $A \cup B = \{1, 2, 3, 4, 5, 6, 7, \}$.
  - $A - B = \{1, 2, 3\}$.
  - $B - A = \{6, 7\}$.



補集合

- 補集合
  - 集合 X と，その部分集合 A に対して，
    $X - A$ を A の（X における）補集合とよび，
    A^{c} や $X \setminus A$ などで表します．
  - このとき，X を全体集合とよびます．



（なお，このような図は，ベン図とよばれます．）

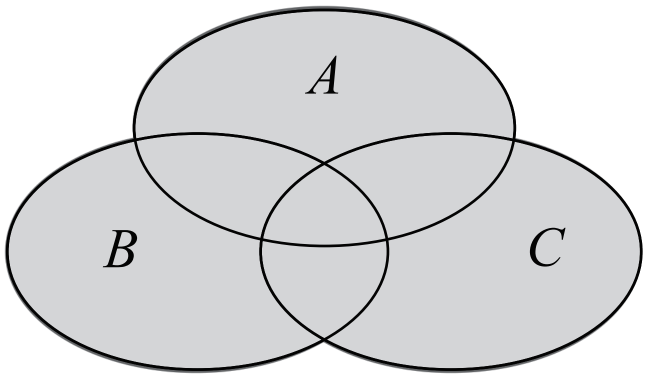
- 例： $X = \{1, 2, 3, 4, 5, 6\}$, $A = \{2, 3, 4\}$ とすると，$A^{\mathrm{c}} = \{1, 5, 6\}$ です．
- 例： $X = \mathbb{Z}$, $A = \{x \in \mathbb{Z} \mid x$ は 3 で割ると 1 余る $\}$ とすると，
  A^{c} は「3 の倍数と，3 で割ると 2 余る整数とを，すべて集めた集合」です．

結合法則と分配法則 (1/2)

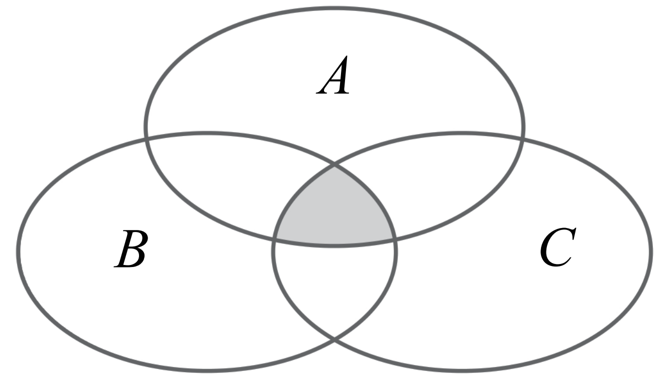
- 結合法則

(i) $(A \cup B) \cup C = A \cup (B \cup C)$.

(ii) $(A \cap B) \cap C = A \cap (B \cap C)$.



(i)



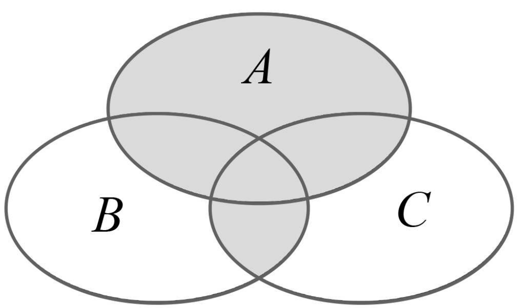
(ii)

結合法則と分配法則 (2/2)

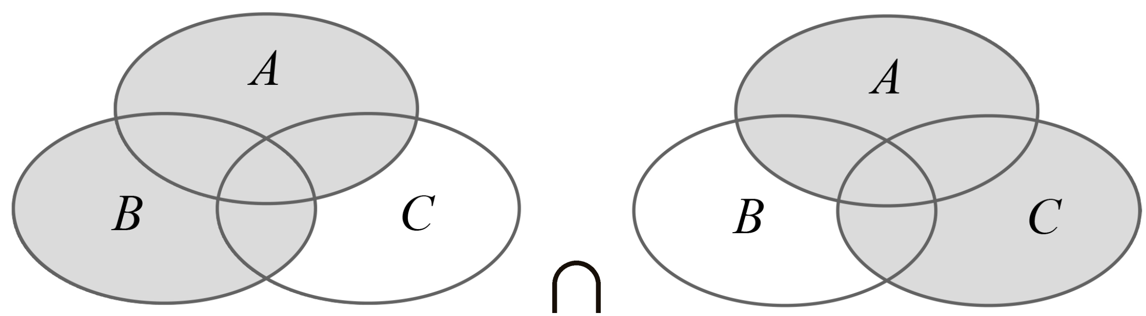
- 分配法則

(i) $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

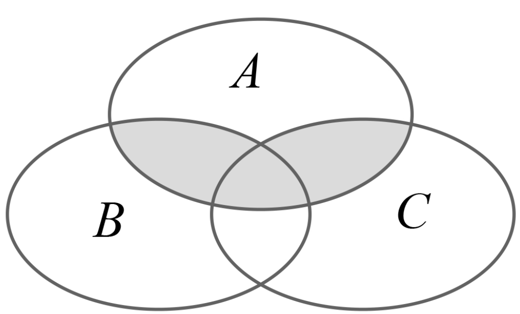
(ii) $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.



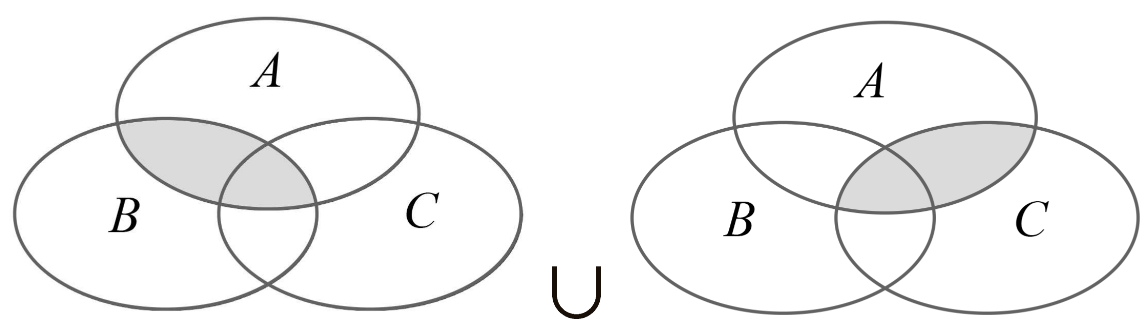
(i) の左辺



(i) の右辺



(ii) の左辺



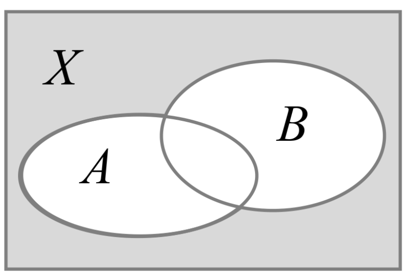
(ii) の右辺

ド・モルガンの法則

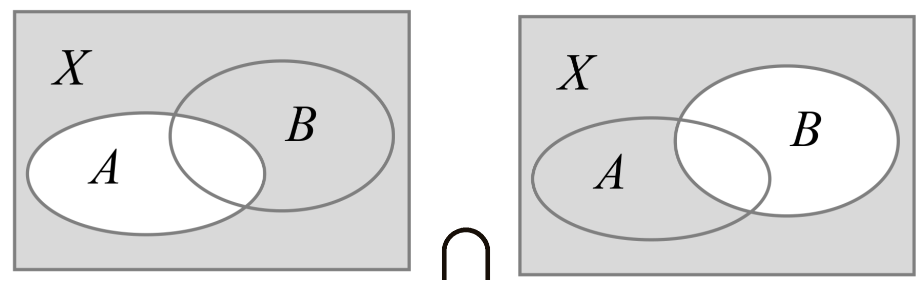
- 集合 X, A, B に対して,

(i) $X - (A \cup B) = (X - A) \cap (X - B)$.

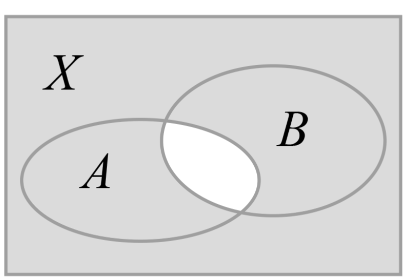
(ii) $X - (A \cap B) = (X - A) \cup (X - B)$.



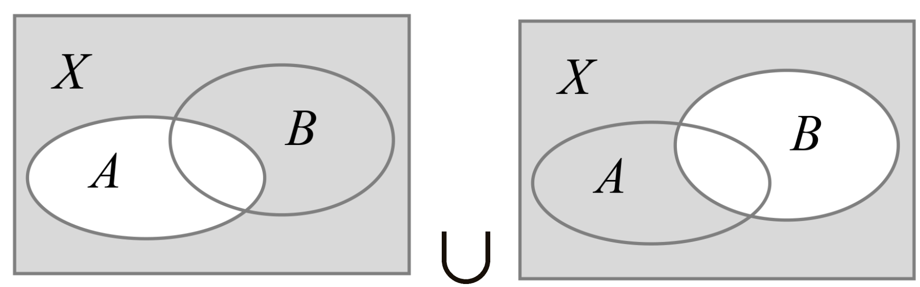
(i) の左辺



(i) の右辺



(ii) の左辺

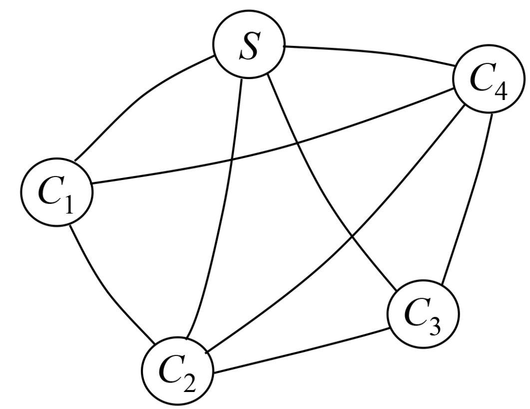


(ii) の右辺

順列：導入

- 例として，巡回セールスマン問題とよばれる問題をとりあげます.

- 都市 s から出発し，都市 $c_1, \ldots, c_4$ のすべてをちょうど 1 回ずつ訪れてから s に戻る経路のうち，移動距離が最小のものを求めて下さい.



- 候補となる経路は，

$$
\begin{array}{l}
s \to c_1 \to c_2 \to c_3 \to c_4 \to s \\
s \to c_1 \to c_2 \to c_4 \to c_3 \to s \\
s \to c_1 \to c_3 \to c_2 \to c_4 \to s \\
s \to c_1 \to c_3 \to c_4 \to c_2 \to s \\
s \to c_1 \to c_4 \to c_2 \to c_3 \to s \\
s \to c_1 \to c_4 \to c_3 \to c_2 \to s \\
s \to c_2 \to \cdots \\
\quad\quad \vdots
\end{array}
$$

  - 最初を c_1 と決めたら，6 通り.
  - 最初の決め方は 4 通り → 全部で $4 \cdot 3 \cdot 2 \cdot 1 = 24$ 通り.

順列とは

- いくつかのものを一列に並べるとき，その並べ方の１つ１つのことを，順列とよびます.
- 4 個の都市の順列の総数は $4 \cdot 3 \cdot 2 \cdot 1 = 24$ 通り.
- n 個の（相異なるものの）順列の総数は $n!$ です.
  - $n! = n \cdot (n-1) \cdots 3 \cdot 2 \cdot 1$ のことを，n の階乗とよびます.
    - $0! = 1$ と決めておく.
    - $n!$ は，n が大きくなると急激に大きくなります.
- n 個から r 個を選んで並べる場合は，順列の総数は

$$\underbrace{n \cdot (n-1) \cdots (n-r+2) \cdot (n-r+1)}_{r\text{ 個の積}}.$$

  - この数のことを，${}_n\mathrm{P}_r$ と書きます.
    - ${}_n\mathrm{P}_r = \dfrac{n!}{(n-r)!}$，　${}_n\mathrm{P}_n = n!$ です.

（狭義の）組合せとは

- n 個のものから r 個を選ぶときの，選び方の総数のことを，組合せとよび，${}_n\mathrm{C}_r$ や $\dbinom{n}{r}$ で表します.
  - 順列では順番を区別しますが，組合せでは順番を区別しません.
  - 選んできた r 個の順列の総数は $r!$ ですので，
$$ {}_n\mathrm{C}_r = \frac{{}_n\mathrm{P}_r}{r!} = \frac{n!}{(n-r)!r!} \,. $$
  - ${}_n\mathrm{C}_0 = 1\,,\quad {}_n\mathrm{C}_n = 1\,.$
  - ${}_{n+1}\mathrm{C}_r = {}_n\mathrm{C}_{r-1} + {}_n\mathrm{C}_r\,.$ （再帰的な定義）
    - $n+1$ 個目を選ぶ場合は，n 個目までからはあと $r-1$ 個を選ぶので，${}_n\mathrm{C}_{r-1}\,.$
    - $n+1$ 個目を選ばない場合は，n 個目までから r 個すべてを選ぶので，${}_n\mathrm{C}_r\,.$

2 項定理

- 母関数
  - 数列 $a_0, a_1, a_2, \ldots, a_n$ に対して，関数
  $$a_0 + a_1 x + a_2 x^2 + \cdots + a_n x^n$$
  のことを（x を変数とする）母関数とよびます.
- 数列 ${}_n\mathrm{C}_0, {}_n\mathrm{C}_1, {}_n\mathrm{C}_2, \ldots, {}_n\mathrm{C}_n$ の母関数は $(1+x)^n$ です，つまり，
$$(1+x)^n = {}_n\mathrm{C}_0 + {}_n\mathrm{C}_1 x + {}_n\mathrm{C}_2 x^2 + \cdots + {}_n\mathrm{C}_{n-1} x^{n-1} + {}_n\mathrm{C}_n x^n .$$
- これを一般化して，
$$(x+y)^n = {}_n\mathrm{C}_0 y^n + {}_n\mathrm{C}_1 x y^{n-1} + {}_n\mathrm{C}_2 x^2 y^{n-2} + \cdots + {}_n\mathrm{C}_{n-1} x^n y + {}_n\mathrm{C}_n x^n .$$
  - これを，2 項定理とよびます．また，${}_n\mathrm{C}_r$ を 2 項係数とよびます.
- 例：
  - $(x+y)^6$ を展開したときの x^2y^4 の係数は，${}_6\mathrm{C}_2 = 15$.
  - $(x-3y)^7$ を展開したときの x^4y^3 の係数は，${}_7\mathrm{C}_4 \cdot (-3)^3 = -945$.

2 項定理の応用

- $\displaystyle\sum_{k=0}^{n} {}_n\mathrm{C}_k = 2^n$.
  - $\because\ 2^n = (1+1)^n = \displaystyle\sum_{k=0}^{n} {}_n\mathrm{C}_k 1^k 1^{n-k}$.

    （n 個それぞれについて，選ぶか否かの 2 通りがある，と考えることもできます.）

- ${}_{m+n}\mathrm{C}_r = \displaystyle\sum_{k=0}^{r} {}_m\mathrm{C}_k {}_n\mathrm{C}_{r-k}$.
  - $\because\ (1+x)^{m+n} = (1+x)^m(1+x)^n = \displaystyle\sum_{k=0}^{m} {}_m\mathrm{C}_k x^k \sum_{l=0}^{n} {}_n\mathrm{C}_l x^l$ において，

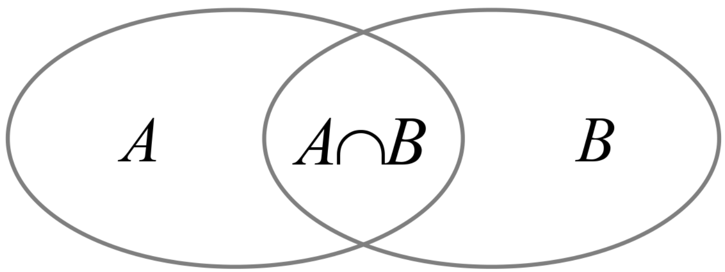
    x^r の係数を比較します.

（注）記号 $\because$ は,「なぜならば」を意味します.

包除原理 = 包含と排除の原理

- 数え上げに用いられる，集合の要素の数に関する原理です.
- S が有限集合のとき，S の要素の数を $|S|$ で表します.
- 有限集合 A, B に対して，

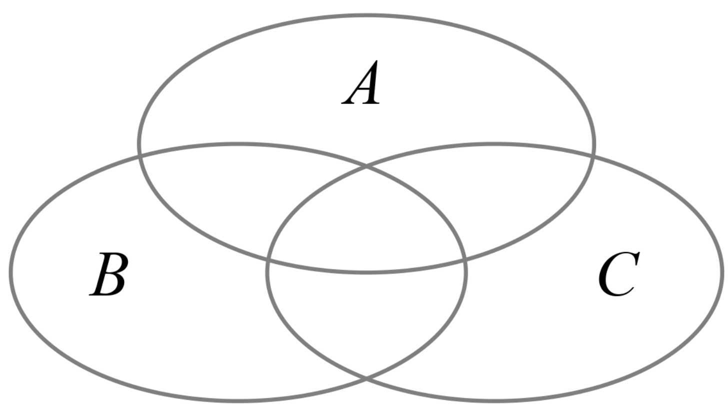
$$|A \cup B| = |A| + |B| - |A \cap B| .$$



包除原理 = 包含と排除の原理

- 数え上げに用いられる，集合の要素の数に関する原理です.
- S が有限集合のとき，S の要素の数を $|S|$ で表します.
- 有限集合 A, B, C に対して，

$$
\begin{aligned}
|A \cup B \cup C| = |A| + |B| + |C| \\
&- |A \cap B| - |B \cap C| - |C \cap A| \\
&+ |A \cap B \cap C| \, .
\end{aligned}
$$



包除原理の応用

- オイラー関数：自然数 n に対して，n と互いに素である 1 以上 n 以下の自然数の個数を $\varphi(n)$ で表します.
  - 例： $\varphi(6) = 2$ （6 と互いに素なのは，1, 5）.
    $\varphi(7) = 6$ （7 と互いに素なのは，$1, \ldots, 6$）.
    $\varphi(8) = 4$ （8 と互いに素なのは，1, 3, 5, 7）.

- $\varphi(504) =?$
  - $504 = 2^3 \times 3^2 \times 7$ なので，504 以下で，2 の倍数でも 3 の倍数でも 7 の倍数でもない自然数の個数を求めます.
  - A を 2 の倍数，B を 3 の倍数，C を 7 の倍数の集合とすると，
$$\phi(504) = 504 - |A \cup B \cup C|\,.$$
  - $|A| = 504/2 = 252\,,\quad |B| = 504/3 = 168\,,\quad |C| = 504/7 = 72\,.$
  - $|A \cap B| = 504/6 = 84$（6 の倍数の数），
    $|B \cap C| = 504/21 = 24\,,\quad |C \cap A| = 504/14 = 36\,,$
    $|A \cap B \cap C| = 504/42 = 12$ より，　$\phi(504) = 504 - 360 = 144\,.$

組合せは どこに現れるでしょうか

- 回帰分析や判別分析での変数選択
  - n 個の説明変数のうちの r 個で目的変数を説明するとすると，選び方は ${}_n\mathrm{C}_r$ 個あります.
- 文書要約
  - n 個の文からなる文書に対して，r 個の文を選んで要約とする．選び方は ${}_n\mathrm{C}_r$ 個あります.
- $2n$ 個の点からなるデータに 2-クラスタリングを適用するとき，解の候補は ${}_{2n}\mathrm{C}_1 + {}_{2n}\mathrm{C}_2 + \cdots + {}_{2n}\mathrm{C}_n$ 個あります.
- （注）以上の例では，すべての組合せを試してみることは，実際にはしません.
- 2 項分布の確率関数 $P(\{X = k\}) = {}_n\mathrm{C}_k p^k (1 - p)^k$.

  ➪ スライド 68

1-6-2 統計学の基礎

データを代表する数値：平均

- 「平均」は最もよく使われる代表値です。データを $X_1, \cdots, X_n$と 書くと、平均は$\bar{X}$またはμなどと書き，

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

と計算されます。

- 平均はデータの中心を表していると考えることができます。

＜データ＞				
3	7	4	1	8

平均値$= \frac{3+7+4+1+8}{5} = 4.6$

データを代表する数値：中央値

- 「中央値」は、データを小さい順に並べた時に真ん中に来る値です。
- データが偶数個なら真ん中の２つの値を足して２で割った値が中央値になります。

＜データ＞

3　7　4　1　8

↓　並べ替え

1　3　4　7　8

中央値：４

＜データ＞

5　4　7　2

↓　並べ替え

2　4　5　7

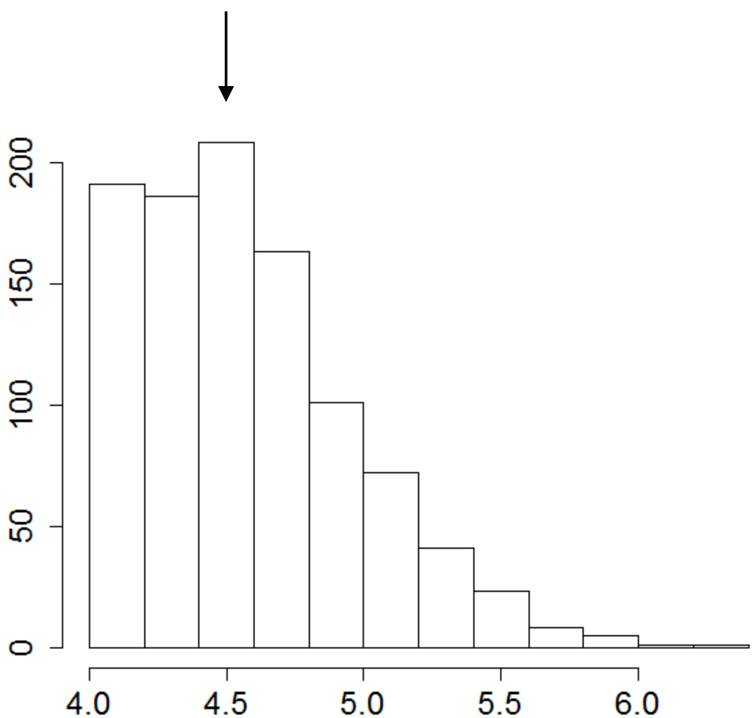
足して２で割る

中央値：４.5

データを代表する数値：最頻値

- 「最頻値」とはデータの中で最も頻繁に現れた値のことです。
- 連続した値のデータでは度数分布の中で最も回数の多い範囲の中央の値を最頻値とします。

最頻値：４．５

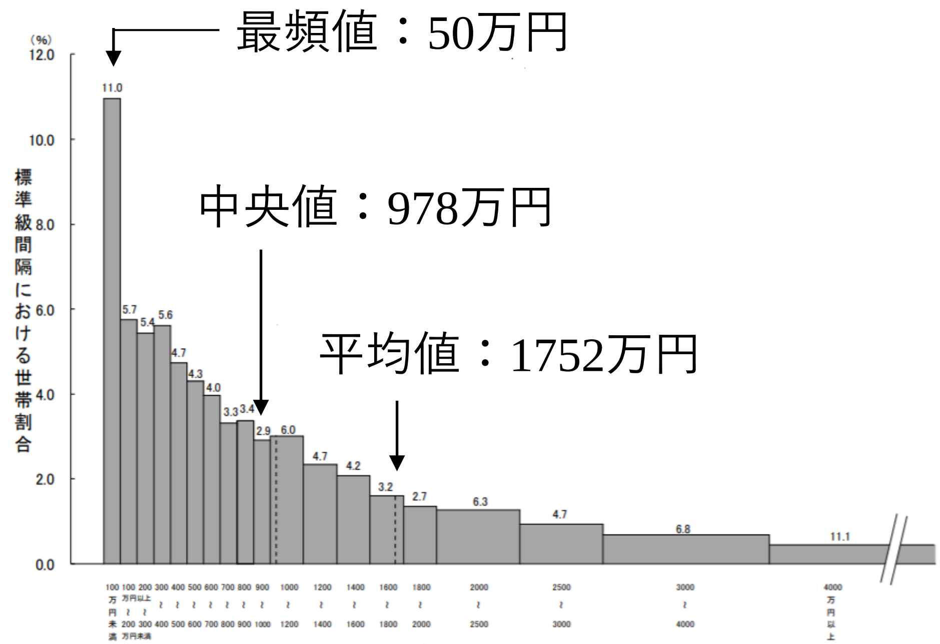


代表値の性質の違い

- ３つの代表値は、実際のデータでは値が異なることも多いです。
- 一部のデータが非常に大きな値となる時、平均値は高くなりやすいです。

- 右図は2018年の全国の二人以上の世帯の貯蓄額のヒストグラムです。

- 一部の裕福な世帯の影響を受け、平均値は中央値よりもかなり高い値になっています。（３分の２の世帯が平均を下回る）



「貯蓄現在高階級別世帯分布（二人以上の世帯）」（総務省統計局）を加工して作成

（https://www.stat.go.jp/data/sav/sokuhou/nen/pdf/2018_gai2.pdf）

データのばらつき（分散、標準偏差）

- データのばらつき度合いを測る指標として、「分散」「標準偏差」「偏差値」などがあります。
- 分散：データを$X_1, \cdots, X_n$として平均を$\bar{X}$とすると、分散σ^2は

$$\sigma^2 = \frac{(X_1 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$

  で与えられます。
  - 各データX_iに対し、$X_i - \bar{X}$の絶対値が大きい時に分散の値が大きくなるので、各データが平均からどの程度離れているかというばらつき度合いを測る指標となります。
- 標準偏差は$\sigma = \sqrt{\sigma^2}$と定義されます。
  - 分散はデータの２乗を計算し、例えば重さ(g)のデータであれば、単位が(g^2)となり、単位が変わってしまいますが、標準偏差の単位は元の単位(g)と同じになります。

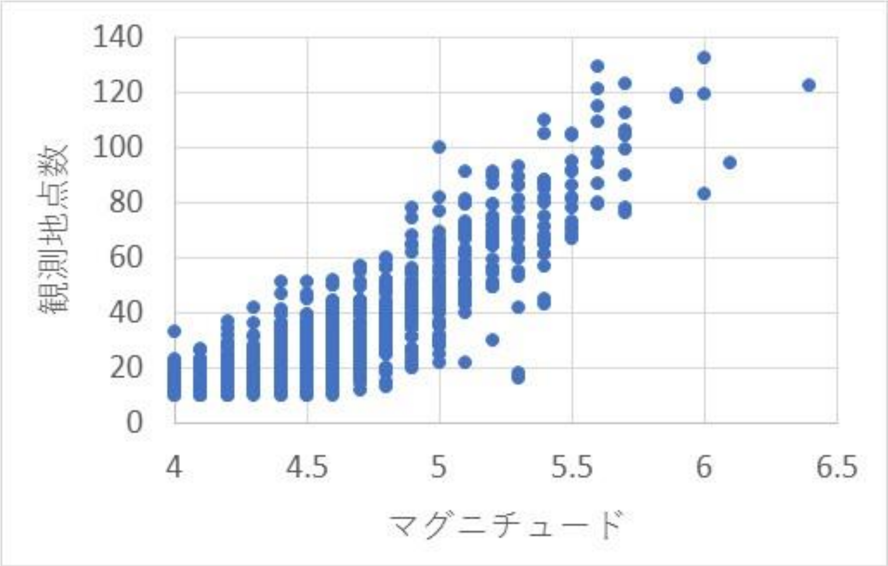
データのばらつき（偏差値）

- データを$X_1, \cdots, X_n$として、平均を$\bar{X}$、標準偏差をσとした時、
$$\frac{X_i - \bar{X}}{\sigma} \times 10 + 50$$
の値を偏差値といいます。
- データが平均値に等しい時（$X_i = \bar{X}$）、偏差値は50となります。
- 偏差値はデータのばらつきを補正した時の各データの位置づけを表していて、おおよそのデータの偏差値は30〜70程度の範囲に収まります。

➢ 各偏差値のおおよその目安は右の表のようになります。

偏差値	上位からの割合	偏差値	上位からの割合
70	2.3%	45	69.1%
65	6.7%	40	84.1%
60	15.9%	35	93.3%
55	30.9%	30	97.7%
50	50.0%		

散布図と相関係数



フィジーの地震データ

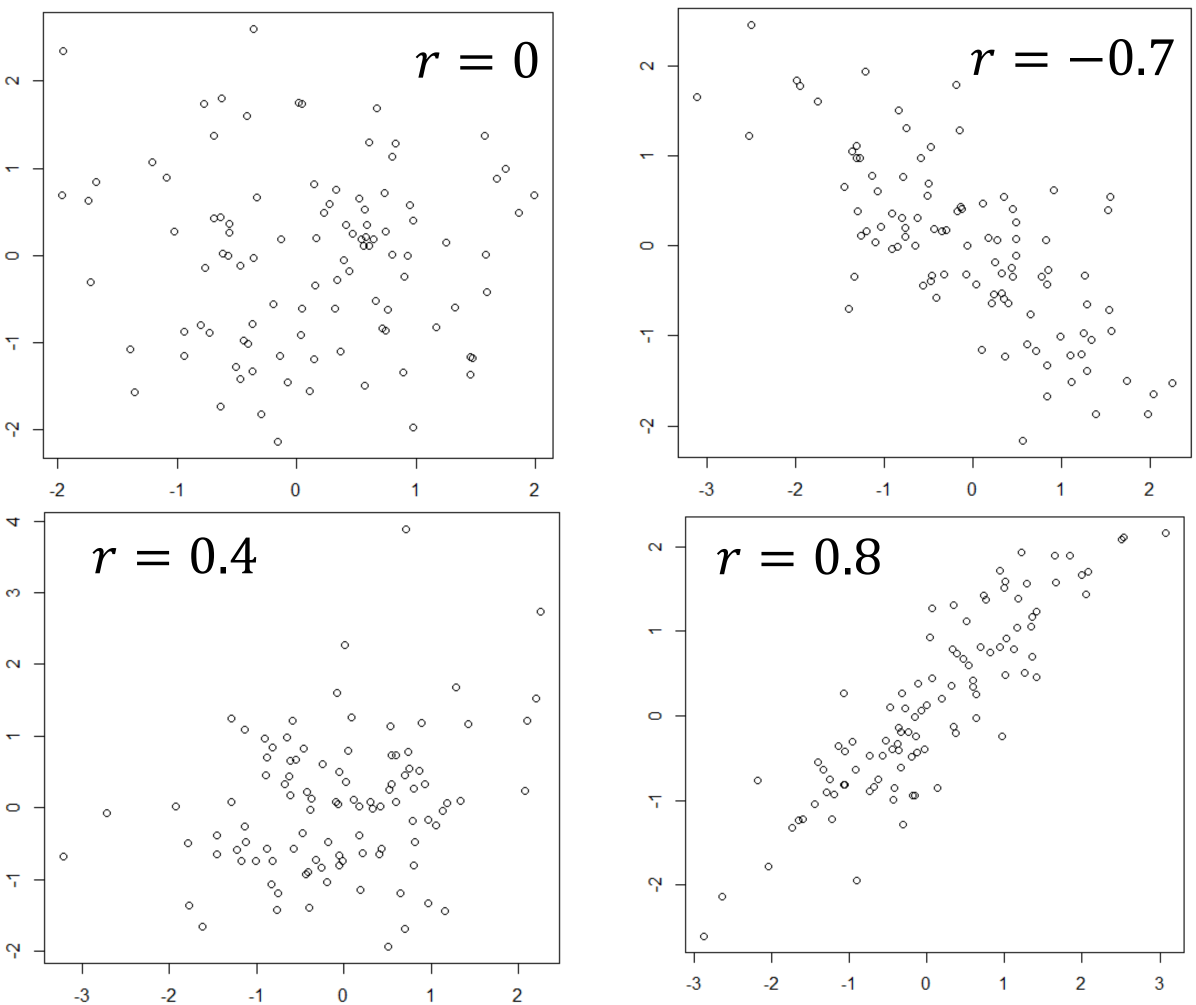
- 散布図
  - 右図のように、データの2種類の項目について2次元にプロットしたものを散布図といいます。
- 相関係数
  - ２種類のデータ$X_1,\cdots,X_n$と$Y_1,\cdots,Y_n$に対して、$X_1,\cdots,X_n$の標準偏差をσ_Xとし、 $Y_1,\cdots,Y_n$の標準偏差をσ_Yとすると、相関係数rは以下で定義されます。

$$r = \frac{(X_1 - \bar{X})(Y_1 - \bar{Y}) + \cdots + (X_n - \bar{X})(Y_n - \bar{Y})}{n\sigma_X\sigma_Y}$$

  - X_i, Y_iが平均から見て同じ方向に動くときにrの値は高くなるので、相関を計算することで連動性を測ることができます。
  - 分母にσ_X, σ_Yがあることで、 $-1 \leq r \leq 1$ となることが保証されます。

相関係数の例

- 相関係数rを変えた時の散布図は以下のようになります。
- rが大きい程散布図は右肩上がりで、-1に近いと右肩下がりになります。



相関行列

- データの全ての項目に対して、任意の２種類の相関をマトリックスで表示したものを相関行列といいます。

緯度	経度	地震の深さ（km）	マグニチュード	計測地点数
-20	182	562	4.8	41
-21	181	650	4.2	15
-26	184	42	5.4	43
-18	182	626	4.1	19
-20	182	649	4	11
-20	184	195	4	12
-12	166	82	4.8	43
-28	182	194	4.4	15
-29	182	211	4.7	35
-17	180	622	4.3	19
・	・	・	・	・
・	・	・	・	・
・	・	・	・	・

フィジーの地震データ

※赤字はマイナスのデータ。

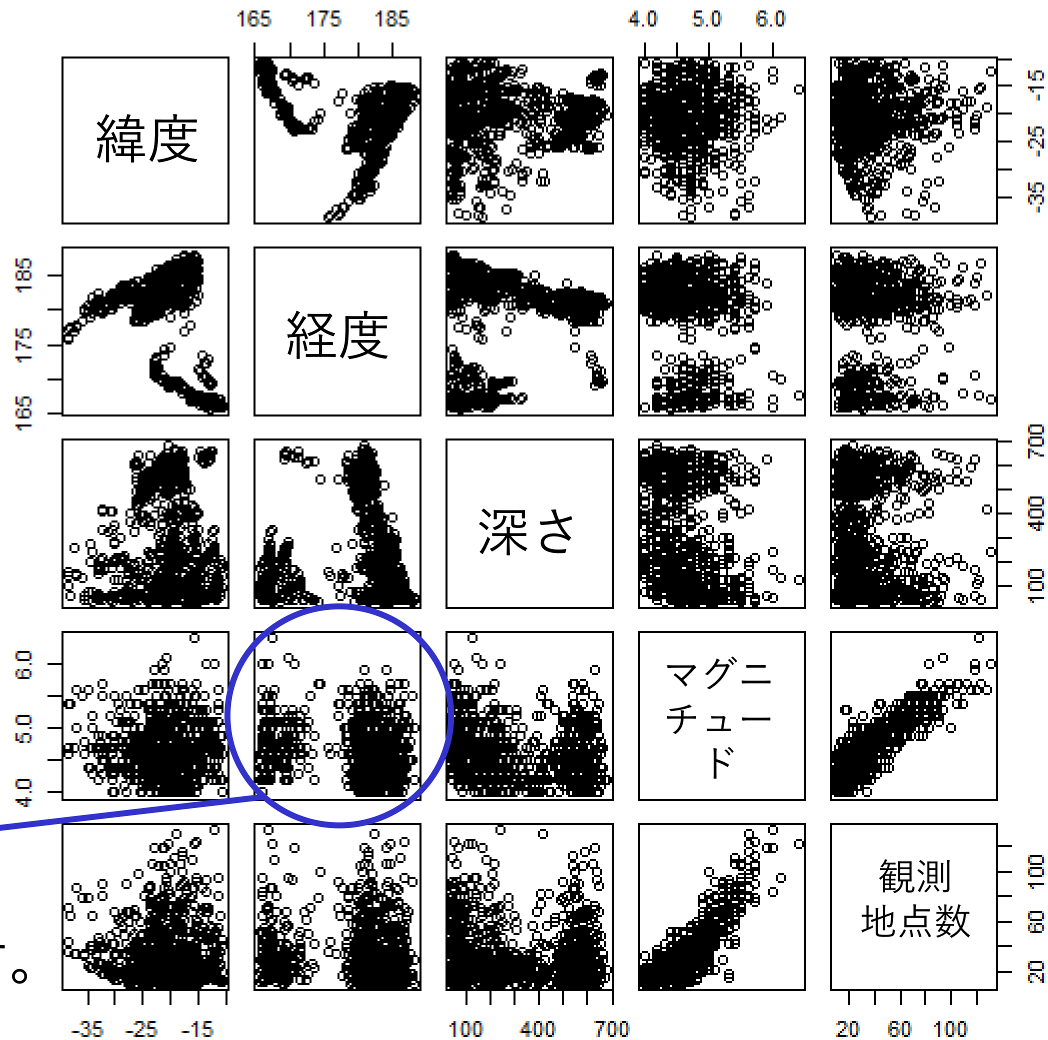
	緯度	経度	地震の深さ（km）	マグニチュード	計測地点数
緯度	1.00	-0.36	0.03	-0.05	-0.00
経度	-0.36	1.00	0.14	-0.17	-0.05
地震の深さ（km）	0.03	0.14	1.00	-0.23	-0.07
マグニチュード	-0.05	-0.17	-0.23	1.00	0.85
計測地点数	-0.00	-0.05	-0.07	0.85	1.00

「経度」と「地震の深さ」の相関が0.14ということになります。

同じ項目の相関は1になります

散布図行列

- データの全ての項目に対して、任意の２種類の散布図をマトリックスで表示したものを散布図行列と呼びます。



「マグニチュード」と
「経度」の散布図を表します。

相関と因果

- ある２つの項目の相関係数が高いからといって、その２つの項目に因果関係があると言えるわけではありません。
- 例えば、小学生の「足のサイズ」と「学力」のデータをとると相関係数が正になります。

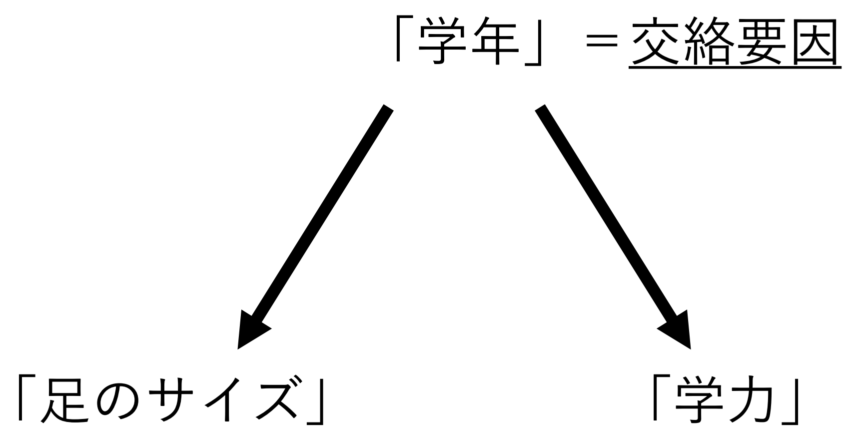
これはどう解釈すればよいでしょうか？

「足が大きくなると学力が高くなる？」

「勉強をすると足も伸びてくる？」

交絡要因と擬似相関

- 「足のサイズ」と「学力」の双方に影響を与える要因として「学年」という要因が考えられます。
  - 学年が高くなると、足のサイズも大きくなり、学力も高くなるので、結果として足のサイズと学力の相関が高くなると解釈できます。
  - このように因果のない２項目の相関が高くなることを擬似相関といい、２項目に影響を与えて相関を高くするような隠れた第三の要因のことを「交絡要因」と呼びます。



結果として、２項目の相関が高くなります

様々な尺度

- 統計学におけるデータは性質に応じて以下の４つの種類（尺度）に分けることができます。

- <u>**名義尺度**</u>・・・互いに区別はできますが順序に意味はない尺度です。
  - 例：氏名、職業、性別

- <u>**順序尺度**</u>・・・順序に意味がありますが、足したり引いたりすることはできない尺度です。
  - 例：大学の成績（優、良、可、不可）、ガンのステージ（I/II/III/IV）
  - 「優＞良＞可＞不可」といったように順序に意味はありますが、「優 ー 良」といった操作はできません。

様々な尺度

- **<u>間隔尺度</u>**・・・数値として表現され、間隔にも意味があり、足したり引いたりできますが、比率（２倍、３倍）には意味がないような尺度です。
  - 例：温度、西暦
  - 「60℃と50℃の差は10℃」「2010年の10年後は2020年」といった足し引きはできますが、「30℃の２倍は60℃」とは言えません。

- **<u>比例尺度</u>**・・・数値として表現され、足したり引いたり、比率を考えることができるような尺度です。
  - 例：身長、体重、金額
  - 「170cmと160cmの差は10cm」「60kgから5kg増えると65kg」「100円の２倍は200円」といった足し引きや比率にはすべて意味があります。

1-6-3　確率論

確率論：導入

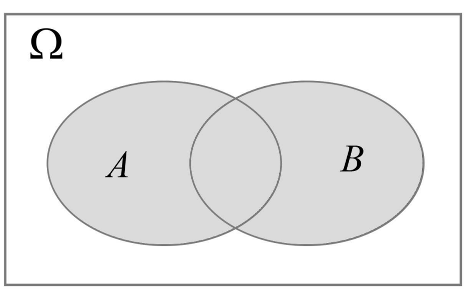
- 確率とは，ものごとが起きる「確からしさ」を 0 から 1 までの数で数値化したもの．偶然性・不規則性・予測不能性の程度の表現．
  - 例： 雨が降るかどうかの予測，疾患の生存率，地震の発生の予測，コインを投げたときに表と裏のどちらが出るかの予測．
- 先験的確率（組合せ確率）
  - 同等に起こりやすいと考えられる事柄は等しい確率をとる，として定めた確率．
- 経験的確率
  - 同じ実験や観測を繰り返して得たデータから定める確率．
- 公理論的確率
  - 現代の数学では，確率の意味については議論せず，ある公理系（約束ごと）を満たすものを確率とよびます．

試行，標本空間，事象

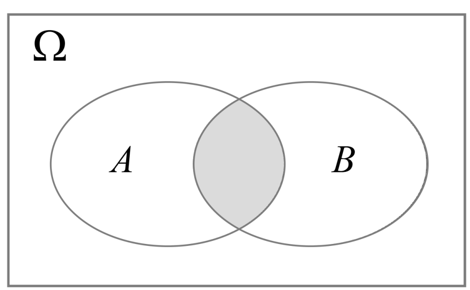
- 試行
  - 確率の考察の対象となる，再現性のある実験や観察を行うこと.
  - 例： サイコロを振ること.
- 標本点（または，標本） ω
  - 個々の試行によって得られた結果のこと.
  - 例： サイコロの例で，3 の目が出ること.
- 標本空間 Ω
  - 標本点すべての集合のこと.
  - 例： $\Omega = \{1\text{ の目が出る}, \ldots, 6\text{ の目が出る}\}$.
- 事象 $A, B, C, \ldots$
  - 標本空間の部分集合のこと.
  - 例： サイコロの例で，偶数の目が出ること.

事象の演算（$A, B, C \subseteq \Omega$ は事象）

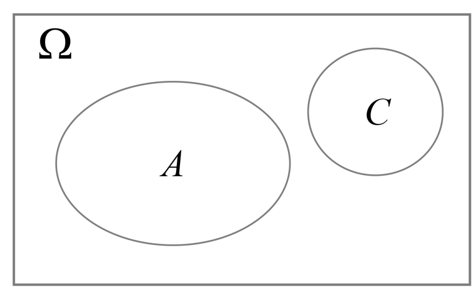
- A と B の和事象 $A \cup B$
  - A と B の少なくとも一方が起こること.
- A と B の積事象 $A \cap B$
  - A と B の両方が起こること.
- A と C が互いに排反であるとは，$A \cap C = \emptyset$（空事象，つまり，標本点を含まない事象）であること.
- A の余事象 A^{c}
  - A が起こらないという事象のこと.



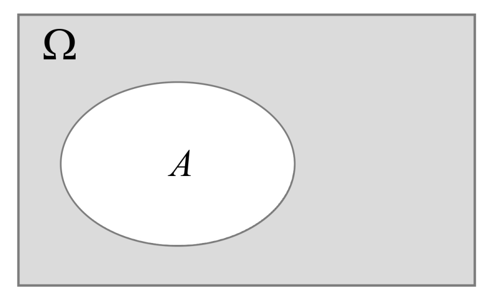
$A \cup B$



$A \cap B$



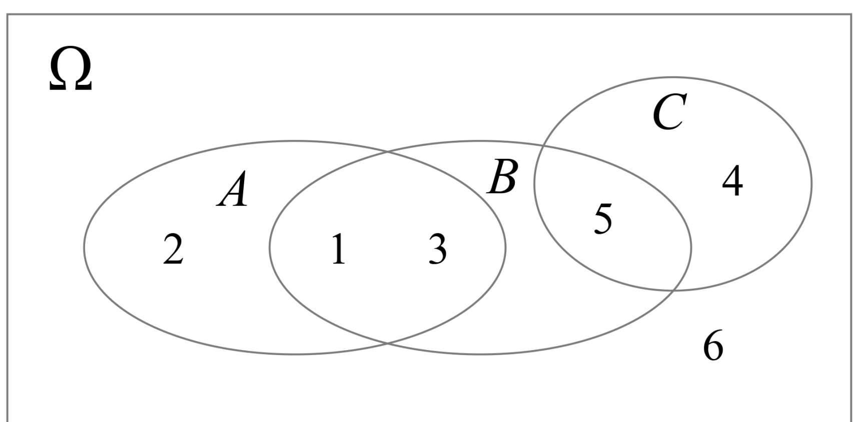
$A \cap C = \emptyset$



A^{c}

事象の演算の具体例

- サイコロを振って出る目を例として考えましょう．
  - 標本空間は，$\Omega = \{1$ の目が出る$, \ldots, 6$ の目が出る $\}$．
  - 3 以下の目のいずれかが出る事象を A とし，
    奇数の目のいずれかが出る事象を B とし，
    4 か 5 のいずれかの目が出る事象を C とすると ⋯
    - $A \cup B$ は，1, 2, 3, 5 のいずれかの目が出る事象．
    - $A \cap B$ は，1 か 3 のいずれかの目が出る事象．
    - A と C は互いに排反．
    - A^{c} は，4, 5, 6 のいずれかの目が出る事象．



確率の基本性質

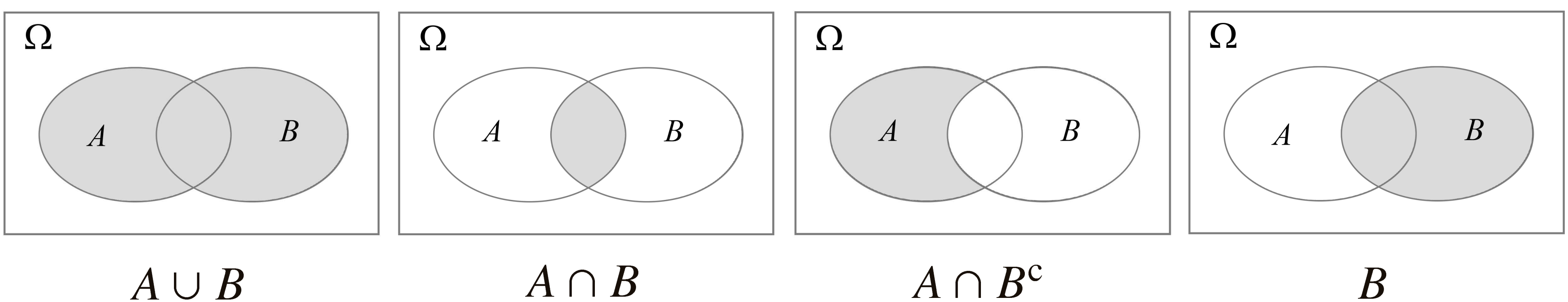
- 事象 A が起きる確率を，$P(A)$ で表します.
  - 任意の事象 A に対して，$0 \leqq P(A) \leqq 1$.
  - $P(\Omega) = 1\,, \quad P(\emptyset) = 0\,.$
  - 事象 A と事象 B が互いに排反（つまり，$A \cap B = \emptyset$）ならば，$P(A \cup B) = P(A) + P(B)\,.$
- 以上の性質は，約束ごとのように（確率の意味を議論することなく）決めておきます.

確率の基本性質に関する具体例

- サイコロを振って出る目を例として考えます.
  - $\Omega = \{1$ の目が出る$, \ldots, 6$ の目が出る $\}$.
  - それぞれの目が出る確率は $\dfrac{1}{6}$ であるとします.
  - 「1 の目が出ること」と「3 の目が出ること」が同時に起きることはありません（つまり，この 2 つの事象は排反です）ので,「1 または 3 の目が出る」確率は $\dfrac{1}{6} + \dfrac{1}{6} = \dfrac{1}{3}$.
  - 「1 または 3 の目が出ること」と「5 の目が出ること」が同時に起きることはありませんので,「奇数の目が出る」確率は $\dfrac{1}{3} + \dfrac{1}{6} = \dfrac{1}{2}$.
  - 同様にして,「偶数の目が出る」確率も $\dfrac{1}{2}$.
  - 「奇数の目が出ること」と「偶数の目が出ること」が同時に起きることはありませんので，$P(\Omega) = \dfrac{1}{2} + \dfrac{1}{2} = 1$.

加法定理

- 事象 A, B に対して，$P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- 証明：
  - A は，$A \cap B$ と $A \cap B^c$ にわけられます（この２つは排反）.
    - したがって，$P(A) = P(A \cap B) + P(A \cap B^c)$.
  - $A \cup B$ は，$A \cap B^c$ と B にわけられます（この２つは排反）.
    - したがって，$P(A \cup B) = P(A \cap B^c) + P(B)$.
  - あとは，２つの等式から $P(A \cap B^c)$ を消去すれば，証明が完成します.



条件付き確率と乗法定理

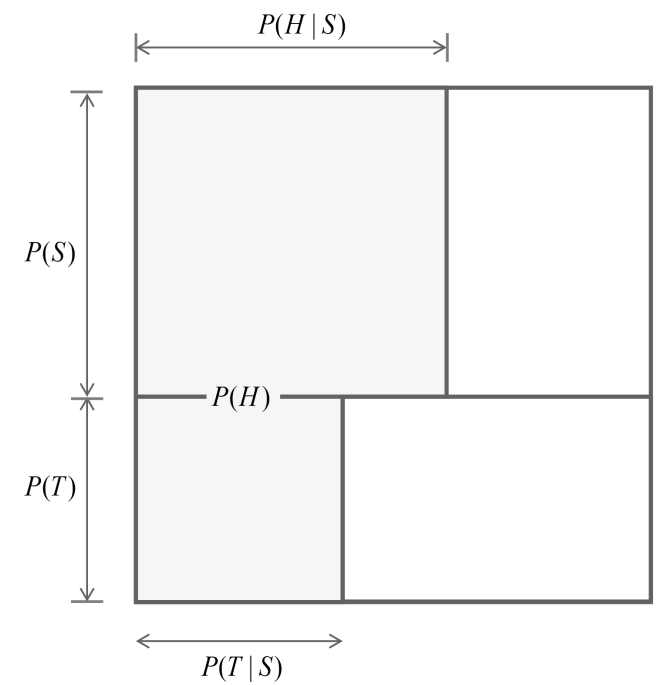
- 事象 A が起こった場合に事象 B が起きるという確率を，条件付き確率とよび，$P(B|A)$ で表します：
$$P(B|A) = \frac{P(A \cap B)}{P(A)}\,.$$
- 分母をはらった
$$P(A \cap B) = P(A)\,P(B|A)$$
の形（乗法定理）も，よく用いられます.

加法定理と乗法定理の具体例 (1/2)

- 10 円玉と 100 円玉を（別々に）投げたとき，表になる確率はそれぞれ $\dfrac{1}{2}$ であるとします．少なくとも一方が表になる確率は，次のようにして求められます．
- 10 円玉が表になる事象を A とし，
  100 円玉が表になる事象を B とします．
  - 「10 円玉が表」の確率は，$P(A) = \dfrac{1}{2}$.
  - 「10 円玉が表のときに 100 円玉が表」の確率は，$P(B|A) = \dfrac{1}{2}$.
- 乗法定理より,「10 円玉も 100 円玉も表」の確率は,
$$P(A \cap B) = P(A)\,P(B|A) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4} .$$
- 加法定理より,「少なくとも一方が表」の確率は,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{2} + \frac{1}{2} - \frac{1}{4} = \frac{3}{4} .$$

加法定理と乗法定理の具体例 (2/2)

- Ｓ型とＴ型のインフルエンザが流行しているとします.
- 患者がＳ型である事象を S とし，Ｔ型である事象を T とします．また，Ｓ型とＴ型の両方にかかっていることはないとします（つまり，S と T は互いに排反です）．このとき，$P(S) = 0.6$, $P(T) = 0.4$ であるとします.
- Ｓ型にかかった際に熱が 38 °C を超える確率は 0.7 であり，Ｔ型では 0.5 であるとします.
- インフルエンザにかかったときに熱が 38 °C を超える事象を H で表すと，その確率 $P(H)$ は次のように求められます.



- $P(S) = 0.6$, $P(H|S) = 0.7$ と乗法定理より，$P(S \cap H) = P(S)\,P(H|S) = 0.42$.
- $P(T) = 0.4$, $P(H|T) = 0.5$ と乗法定理より，$P(T \cap H) = P(T)\,P(H|T) = 0.2$.
- $P(H) = P(S \cap H) + P(T \cap H) = 0.62$.

独立性

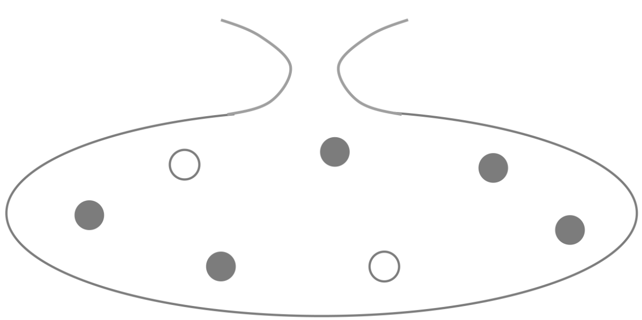
- 事象 A, B に対して条件
$$P(A \cap B) = P(A)\,P(B)$$
が成り立つとき，A と B は互いに独立であるといいます.

    - 乗法定理 $P(A \cap B) = P(A)\,P(B|A)$ より，A と B が互いに独立ならば
    $$P(B|A) = P(B)$$
    が成り立つことが分かります（つまり，B が起きる確率は，A が起きたかどうかに影響されません）.
    同様に,
    $$P(A|B) = P(A)$$
    も成り立ちます.

独立な事象の具体例

- 白い玉が２個と黒い玉が５個入った袋から，玉を取り出すことを考えましょう．どの玉も等しい確率で取り出されるとします.
- 玉を１つ取り出すと，袋に返してから，次の玉を取り出すことにします.
- A：最初の玉が白である事象，　B：２番目の玉が黒である事象.
  - $P(A) = \dfrac{2}{7}$　（どの玉も，取り出される確率は $\dfrac{1}{7}$ だから）.
  - $P(B|A) = P(B) = \dfrac{5}{7}$
    （最初の玉は返したので，最初に何が出ても次の試行に無関係）.
- 最初の玉が白で２番目の玉が黒である確率は，

$$P(A \cap B) = P(A)\,P(B) = \frac{10}{49}.$$



それでは，独立でない例は ⋯

- 白い玉が２個と黒い玉が５個入った袋から，玉を取り出すことを考えましょう．どの玉も等しい確率で取り出されるとします．
- 玉を１つ取り出すと，袋に返さずに，次の玉を取り出すことにします．
- A：最初の玉が白である事象，　B：２番目の玉が黒である事象．
  - $P(A) = \dfrac{2}{7}$　（どの玉も，取り出される確率は $\dfrac{1}{7}$ だから）．
  - $P(B|A) = \dfrac{5}{6}$　$\left(\neq P(B) = \dfrac{5}{7}\right)$
    （最初の白玉は返さないので，袋の中の玉の数は 6）．
- 最初の玉が白で２番目の玉が黒である確率は，乗法定理より
$$P(A \cap B) = P(A)\,P(B|A) = \frac{10}{42}.$$
  - 独立である場合（前頁）と結果が異なることに注意して下さい．

ベイズの定理

- 乗法定理より，$P(A \cap B)$ は
$$P(A)\,P(B|A) \quad \text{と} \quad P(B)\,P(A|B)$$
のいずれにも等しい．このことから，
$$P(B|A) = \frac{P(B)\,P(A|B)}{P(A)}.$$
- 一方，A は，$A \cap B$ と $A \cap B^{\mathrm{c}}$ にわけられます（2 つは排反）．
  - したがって，$P(A) = \underbrace{P(A \cap B)}_{=P(B)\,P(A|B)} + \underbrace{P(A \cap B^{\mathrm{c}})}_{=P(B^{\mathrm{c}})\,P(A|B^{\mathrm{c}})}$.
- 以上より，
$$\overbrace{P(B|A)}^{\text{事後確率}} = \frac{\overbrace{P(B)}^{\text{事前確率}}\ \ P(A|B)}{P(B)\,P(A|B) + P(B^{\mathrm{c}})\,P(A|B^{\mathrm{c}})}. \quad \text{（ベイズの定理）}$$
  - B が原因となって A が生じたと考えると，事後確率 $P(B|A)$ は結果 A を知った上で B が起きていた確率.

ベイズの定理の具体例（加法定理と乗法定理の具体例 (2) の続き）

- S 型と T 型のインフルエンザが流行しているとします．
- 患者が S 型である事象を S とし，T 型である事象を T とします．また，S 型と T 型の両方にかかっていることはないとします（つまり，S と T は互いに排反です）．このとき，$P(S) = 0.6$, $P(T) = 0.4$ であるとします．
- S 型にかかった際に熱が $38\,°\mathrm{C}$ を超える確率は 0.7 であり，T 型では 0.5 であるとします．
- インフルエンザにかかったときに熱が $38\,°\mathrm{C}$ を超える事象を H で表すと，$P(H) = P(S)\,P(H|S) + P(T)\,P(H|T) = 0.62$ でした．
- では，インフルエンザにかかって熱が $38\,°\mathrm{C}$ を超えたとき，S 型であった確率は

$$P(S|H) = \frac{P(S)\,P(H|S)}{P(H)} = \frac{0.6 \cdot 0.7}{0.62} \simeq 0.677\,.$$

ベイズの定理の具体例

- ある電子メールがスパムメール（迷惑メール）であるという事象を B とおき， $P(B) = 0.5$ としします（事前確率）.
- ある電子メールに「緊急」という単語が含まれるという事象を A とおきます.
  - スパムメールに「緊急」が含まれる確率を $P(A|B) = 0.75$, スパムでないメールに「緊急」が含まれる確率を $P(A|B^c) = 0.15$ とします.
- では，メールが「緊急」を含んでいるときにそれがスパムメールである確率（事後確率）は,

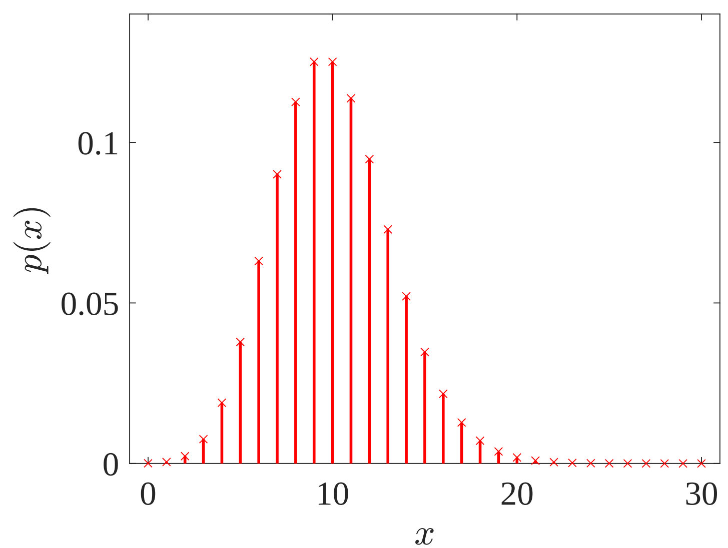
$$P(B|A) = \frac{P(B)\,P(A|B)}{P(B)\,P(A|B) + P(B^c)\,P(A|B^c)} = \frac{0.5 \cdot 0.75}{0.5 \cdot 0.75 + 0.5 \cdot 0.15} \simeq 0.833\,.$$

確率変数

- 例として，コイン投げにおいて，表が出ると $X = 1$ とし裏が出ると $X = 0$ とします．このように，標本点それぞれに対応して値が割り当てられた変数 X のことを，確率変数とよびます．
- より詳しく言うと，標本空間 Ω の要素を $\omega_1, \omega_2, \ldots$ のように列挙できる場合，各 $\omega_k \in \Omega$ に対してある数 x_k を $X(\omega_k) = x_k$ と割り当てます．この X を,（離散型の）確率変数とよびます．
    - Ω の要素がとびとびに数えることができない場合（たとえば，時刻 x までにある機械が故障する事象を考えるような場合）は，取り扱いがより複雑になるので，まずはより簡単な離散型の場合を考えましょう．

確率分布

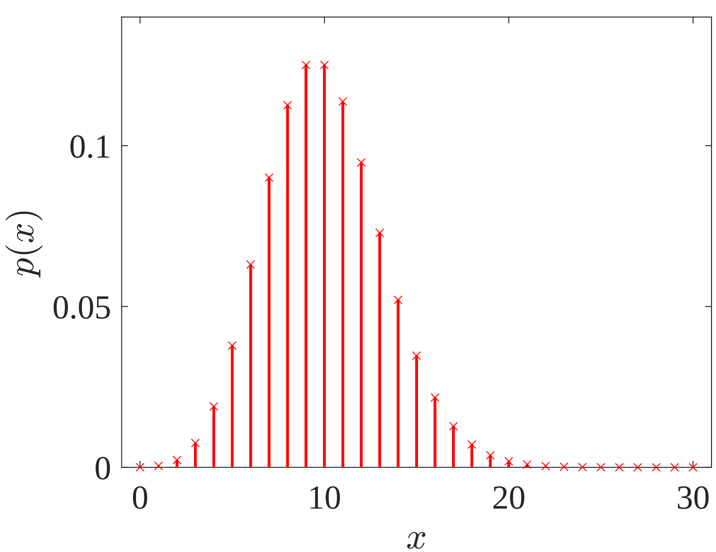
- $X(\omega)$ が値 x_k をとるという事象が起きる確率は，$P(\{\omega \in \Omega \mid X(\omega) = x_k\})$ です． これを，表記の簡単のため，$P(\{X = x_k\})$ と書くことにします．
- x_k と $P(\{X = x_k\})$ の組を，X の確率分布（または，分布）とよびます．
- $p(x) = P(\{X = x\})$ で定まる関数 p を，X の確率関数とよびます．



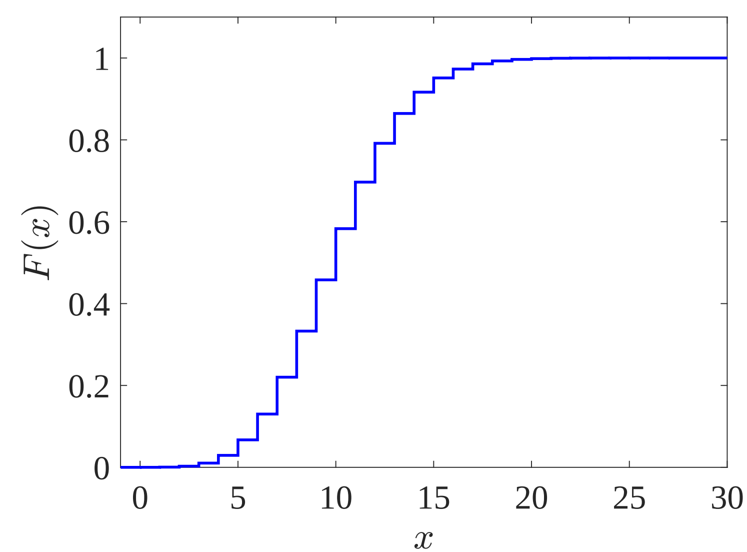
確率関数の例

確率関数と確率分布関数

- 確率関数は，$p(x) = P(\{X = x\})$ でした．
- 確率変数 X がある値 x 以下である確率
$$F(x) = P(\{X \leqq x\})$$
を，X の確率分布関数（または，分布関数，累積分布関数）とよびます．
  - $F(x)$ は，0 から 1 までの値をとります．また，$\lim_{x\to-\infty} F(x) = 0$，$\lim_{x\to+\infty} F(x) = 1$．
  - F は，単調増加関数です．



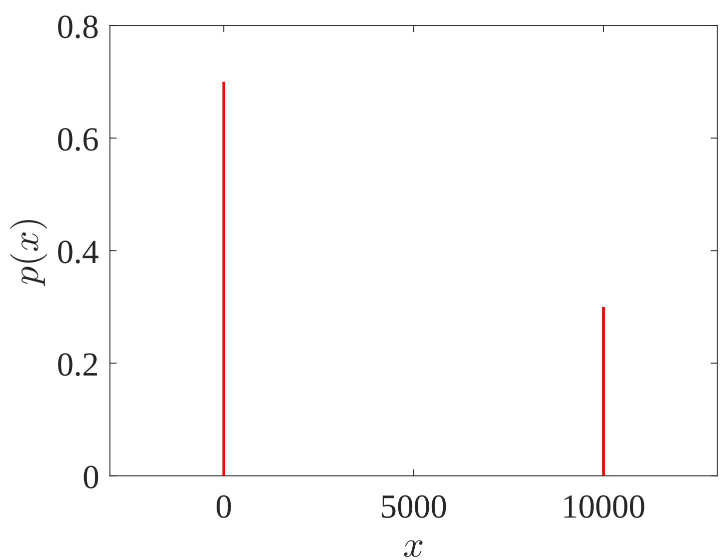
確率関数の例



対応する確率分布関数

確率変数の期待値

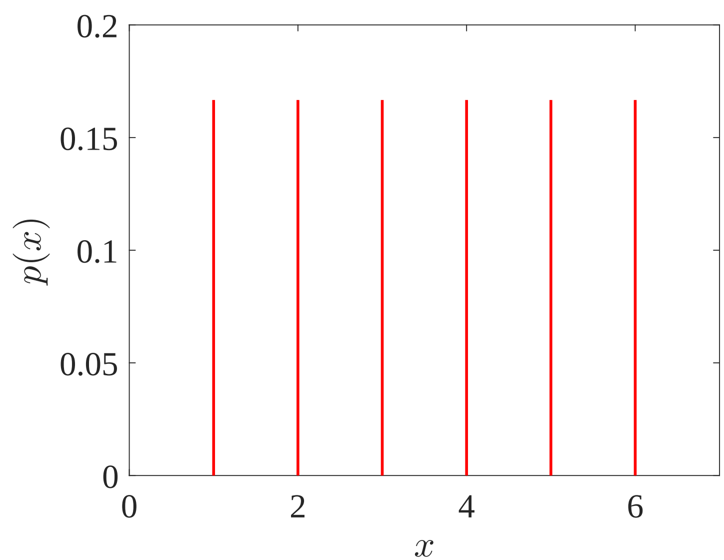
- 確率変数 X の期待値（または，平均）とは，
$$E[X] = x_1\,p(x_1) + x_2\,p(x_2) + \cdots\ .$$
（すべての標本点に関して和をとります）
  - μ で表すことが多いです．
- 例： 0.3 の確率で 1 万円が当たるクジがあるとしましょう．つまり，
$$P(\{X = 1\text{ 万円}\}) = 0.3\,,\quad P(\{X = 0\text{ 円}\}) = 0.7\,.$$
当選金の期待値は，
$$E[X] = 1\text{ 万円} \cdot 0.3 + 0\text{ 円} \cdot 0.7 = 3\text{ 千円}\,.$$



確率変数の期待値

- 確率変数 X の期待値（または，平均）とは，
$$E[X] = x_1\,p(x_1) + x_2\,p(x_2) + \cdots\,.$$
（すべての標本点に関して和をとります）
  - μ で表すことが多いです.

- サイコロを振って出る目の数を X とおくと，
$$P(\{X=1\}) = P(\{X=2\}) = \cdots = P(\{X=6\}) = \frac{1}{6}\,.$$



X の期待値は，
$$E[X] = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + \cdots + 6\cdot\frac{1}{6} = 3.5\,.$$

確率変数の分散

- 確率変数 X の分散とは,

$$V[X] = E[(X-\mu)^2] \qquad (\mu \text{ は } X \text{ の期待値})$$
$$= (x_1-\mu)^2\, p(x_1) + (x_2-\mu)^2\, p(x_2) + \cdots .$$

（すべての標本点に関して和をとります）

  - ばらつき具合の尺度．σ^2 で表すことが多いです.
  - $\sqrt{\sigma^2}$ を，標準偏差とよびます.

- 例．0.3 の確率で 1 万円が当たるクジがあるとします：

$$P(\{X = 1\text{ 万円}\}) = 0.3\,, \quad P(\{X = 0\text{ 円}\}) = 0.7\,.$$

当選金の期待値は,

$$E[X] = 1\text{ 万円}\cdot 0.3 + 0\text{ 円}\cdot 0.7 = 3\text{ 千円}\,.$$

当選金の分散は,

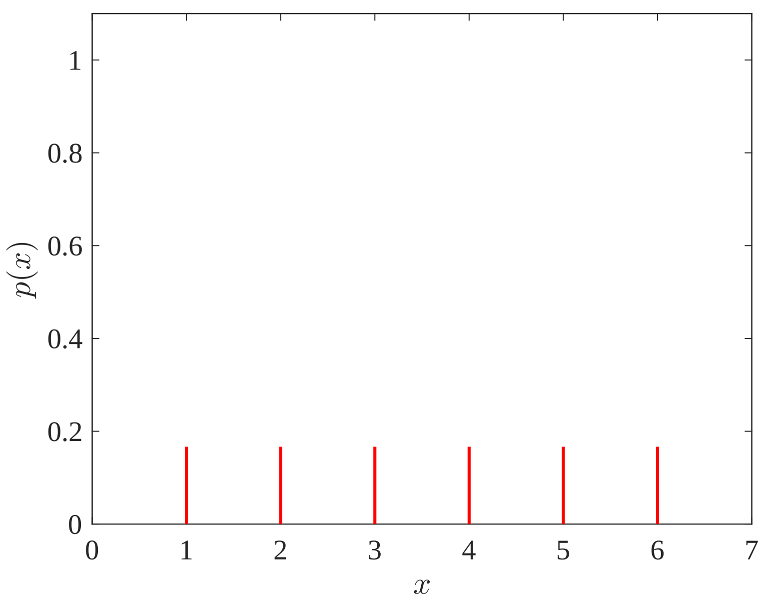
$$V[X] = (1\text{ 万円} - 3\text{ 千円})^2\cdot 0.3 + (0\text{ 円} - 3\text{ 千円})^2\cdot 0.7 = 2100\text{ 円}^2\,.$$

期待値と分散に関する具体例

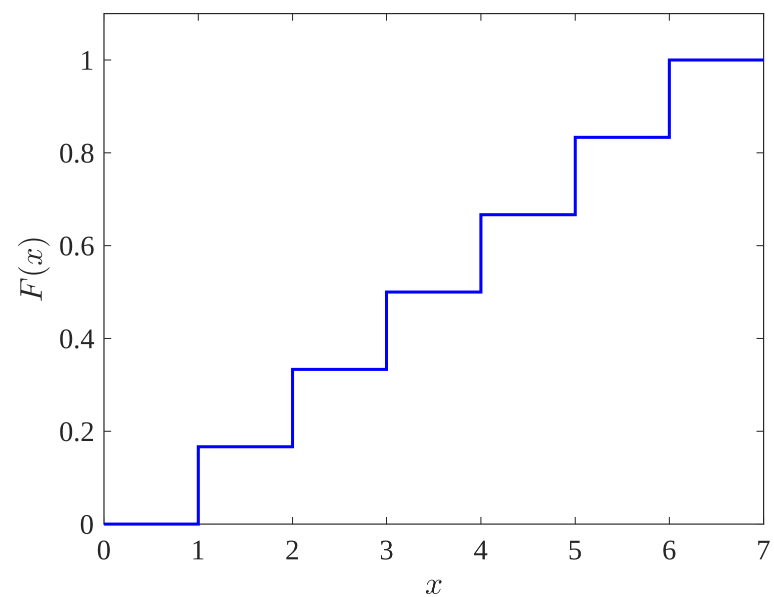
- 0.3 の確率で 1 万円が当たるクジ：
  $$P(\{X = 1\text{ 万円 }\}) = 0.3\,, \quad P(\{X = 0\text{ 円 }\}) = 0.7\,.$$
- 0.003 の確率で 100 万円が当たるクジ：
  $$P(\{Y = 100\text{ 万円 }\}) = 0.003\,, \quad P(\{Y = 0\text{ 円 }\}) = 0.997\,.$$
- X と Y では，期待値は同じだが分散が異なります：
  - $E[X] = 3$ 千円， $V[X] = 2100$ 円^2.
  - $E[Y] = 3$ 千円， $V[Y] = 299100$ 円^2.
- 両方のクジを 1 枚ずつ買った場合，当選金の合計の期待値は，
  $$E[X+Y] = E[X] + E[Y] = 6\text{ 千円}\,. \qquad \text{（期待値の加法性）}$$
- X のクジを α 枚だけ買った場合，当選金の合計の期待値と分散は，
  $$E[\alpha X] = \alpha E[X]\,,$$
  $$V[\alpha X] = \alpha^2 V[X]\,.$$

さまざまな確率分布：離散型

- 離散一様分布 DU(r)　（r は自然数）
  - $1, 2, \ldots, r$ の値をそれぞれ等確率 $1/r$ でとる確率分布です.
  - 例： サイコロを振ったときの目の確率分布は，DU(6) です.



DU(6) の確率関数

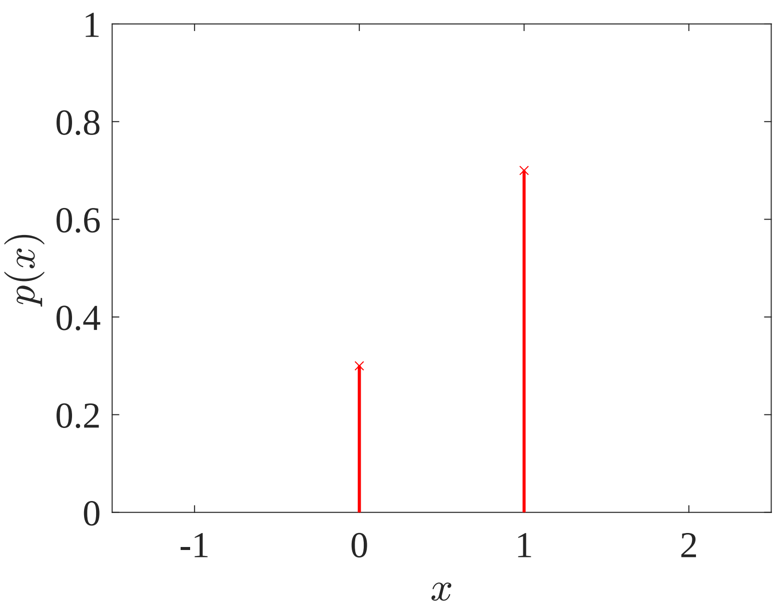


DU(6) の確率分布関数

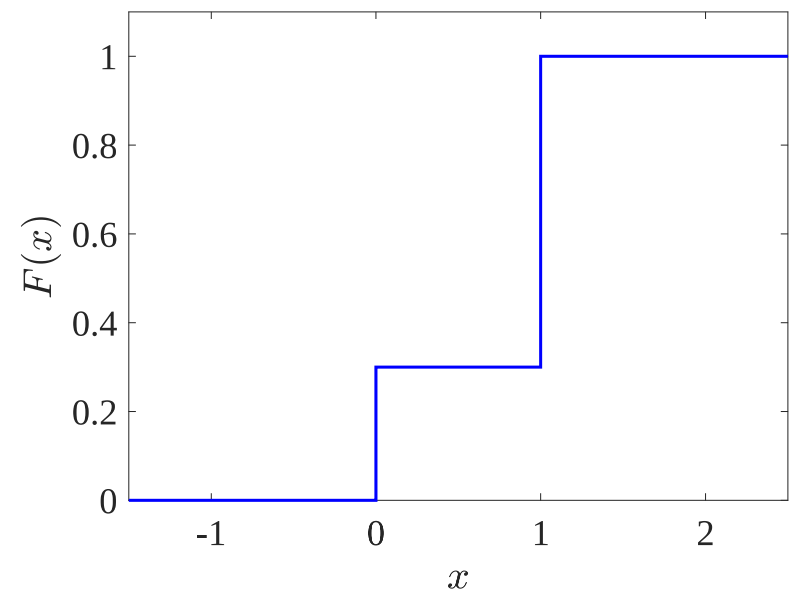
代表的な離散型の確率分布

- ベルヌーイ分布
  - 表の出る確率が p で裏の出る確率が $1-p$ のコインを投げ，表が出ると $X=1$ とし，裏が出ると $X=0$ とした確率分布のことです．
  - 確率関数は

$$P(\{X=1\})=p\,, \quad P(\{X=0\})=1-p\,.$$



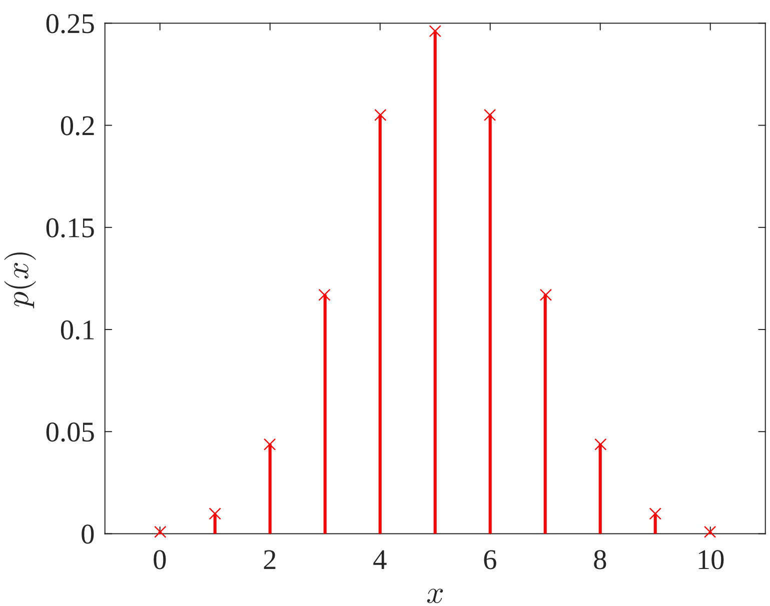
ベルヌーイ分布 ($p=0.7$) の確率関数



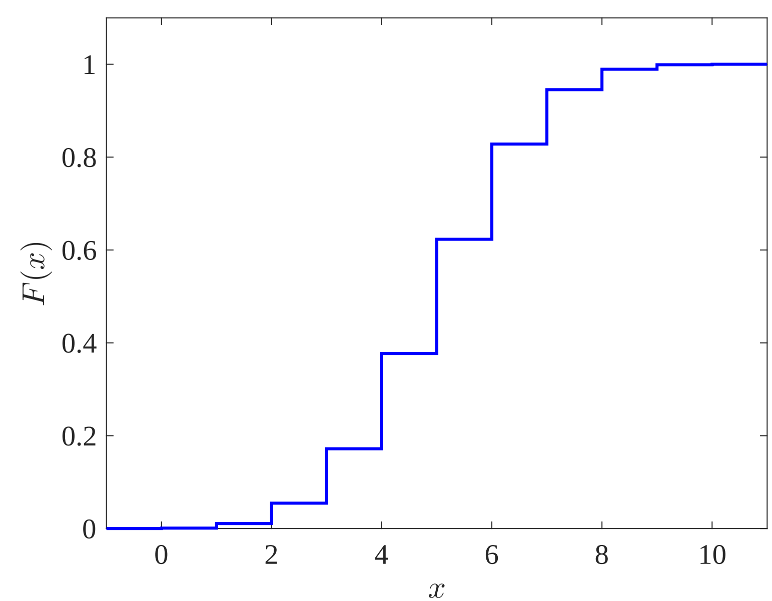
ベルヌーイ分布 ($p=0.7$) の確率分布関数

代表的な離散型の確率分布

- 2 項分布 Bɴ(n, p)
  - ベルヌーイ分布で考えたコインを n 回投げたとき，表の出る回数 X の確率分布のことです（つまり，ベルヌーイ分布は Bɴ$(1, p)$ です）.
  - 確率関数は $P(\{X = k\}) = {}_n\mathrm{C}_k p^k (1-p)^k \quad (k = 0, 1, \ldots, n)$.



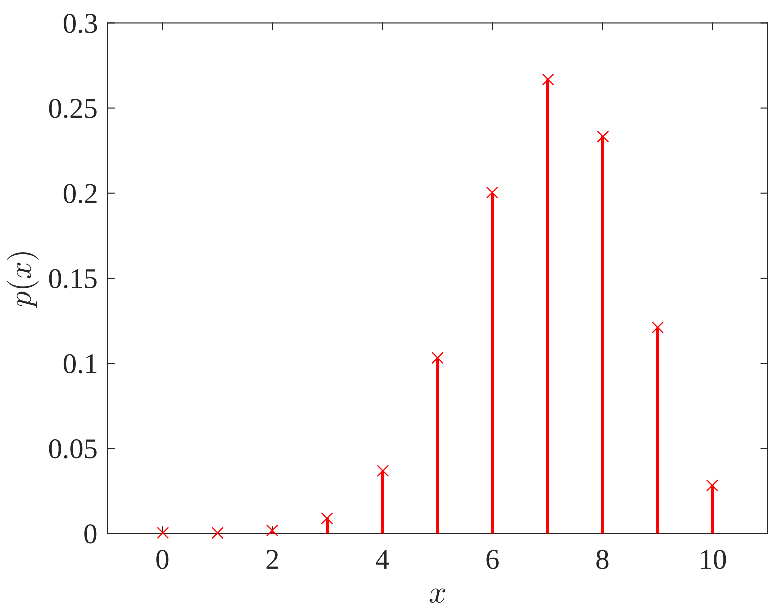
Bɴ$(10, 0.5)$ の確率関数



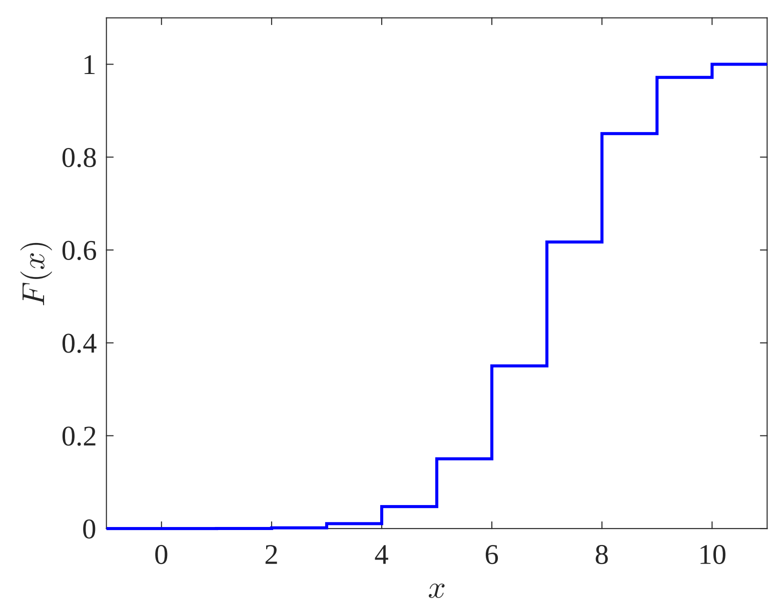
Bɴ$(10, 0.5)$ の確率分布関数

代表的な離散型の確率分布

- 2 項分布 $\mathrm{B\scriptsize N}(n, p)$
  - ベルヌーイ分布で考えたコインを n 回投げたとき，表の出る回数 X の確率分布のことです（つまり，ベルヌーイ分布は $\mathrm{B\scriptsize N}(1, p)$ です）.
  - 確率関数は $P(\{X = k\}) = {}_n\mathrm{C}_k p^k (1-p)^k \quad (k = 0, 1, \ldots, n)$.



$\mathrm{B\scriptsize N}(10, 0.7)$ の確率関数



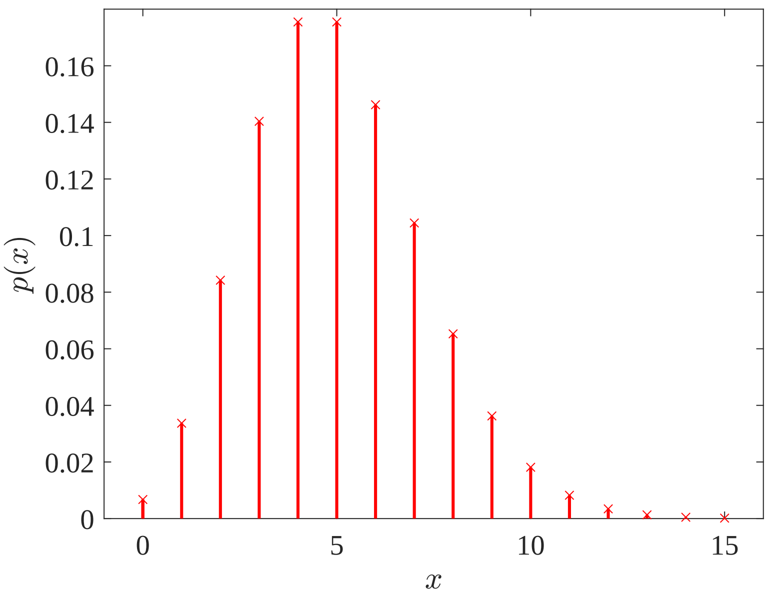
$\mathrm{B\scriptsize N}(10, 0.7)$ の確率分布関数

代表的な離散型の確率分布

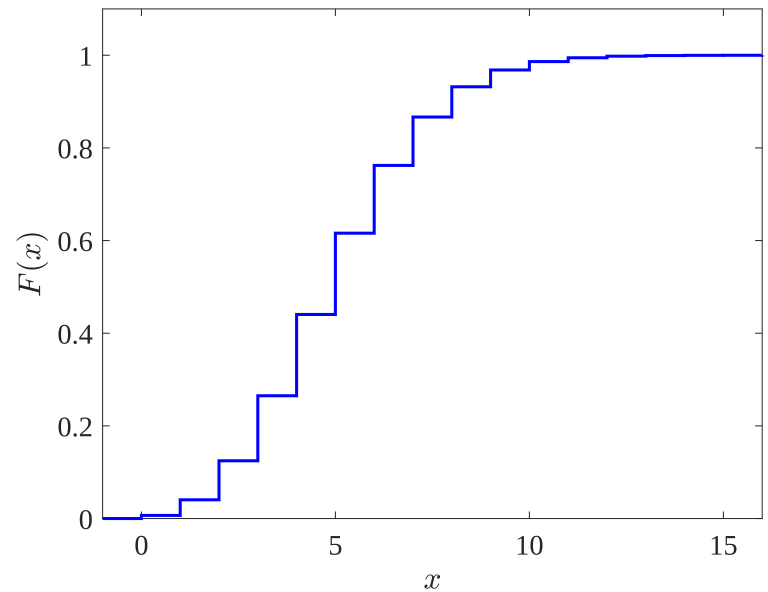
- ポアソン分布 Po(λ)
  - 2 項分布 Bn(n, p) で，$np = \lambda$（一定値）としたまま，コインを投げる回数 n を非常に多くし表の出る確率 p を 0 に近づけていくと得られる分布です.
  - 個々の生起確率は小さいけれど，分析対象の数が多い現象は，ポアソン分布がよくあてはまることが知られています．たとえば，
    - 一定期間に製造された製品のうち，不良品の個数，
    - 一定期間内の事故の件数，
    - 単位面積あたりの雨粒の数.
  - 確率関数は $P(\{X = k\}) = \dfrac{\lambda^k}{k!}e^{-\lambda} \quad (k = 0, 1, 2, \ldots)$.

代表的な離散型の確率分布

- ポアソン分布 Po(λ)
  - 確率関数は $P(\{X=k\}) = \dfrac{\lambda^k}{k!}e^{-\lambda} \quad (k=0,1,2,\ldots)$.
  - 期待値も分散も λ .



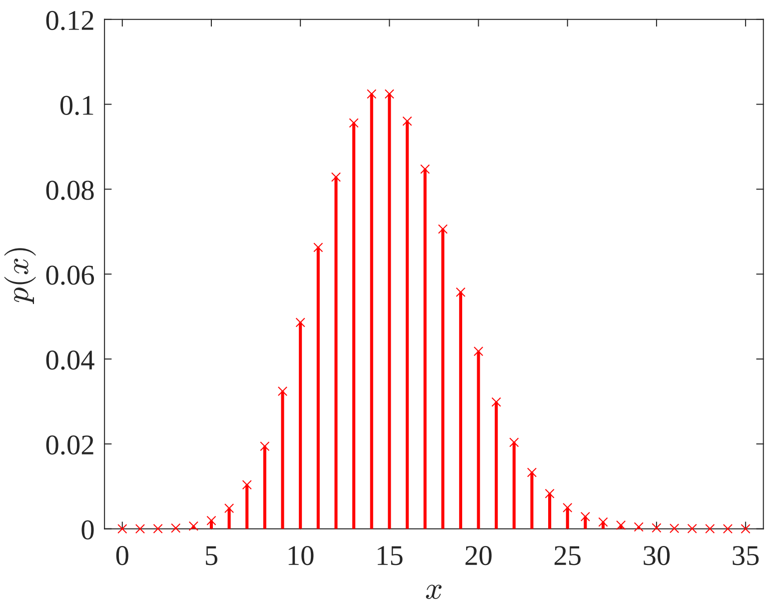
Po(5) の確率密度関数



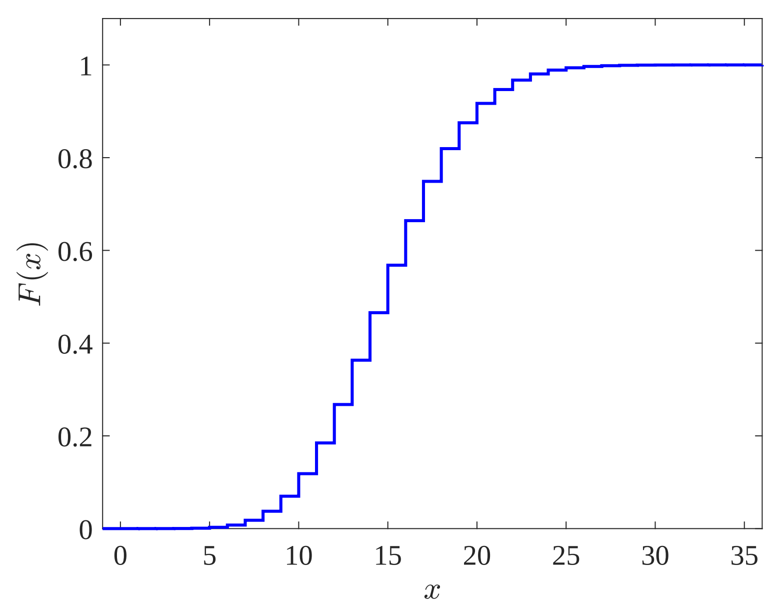
Po(5) の確率分布関数

代表的な離散型の確率分布

- ポアソン分布 Po(λ)
  - 確率関数は $P(\{X = k\}) = \dfrac{\lambda^k}{k!} e^{-\lambda} \qquad (k = 0, 1, 2, \ldots)$.
  - 期待値も分散も λ.



Po(15) の確率密度関数



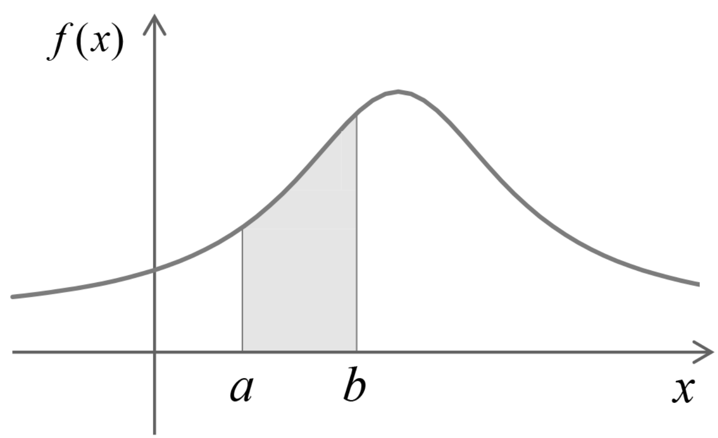
Po(15) の確率分布関数

連続型の確率分布

- 確率変数 X が，長さや時間など，連続値をとる変数である場合を考えましょう.
- X がとり得る値は無限個あるので，X がある 1 つの値 x をとる確率は $P(\{X = x\}) = 0$ ということになります.
- 一方，確率分布関数は，離散型の場合と同様に，$F(x) = P(\{X \leqq x\})$ で定義できます.

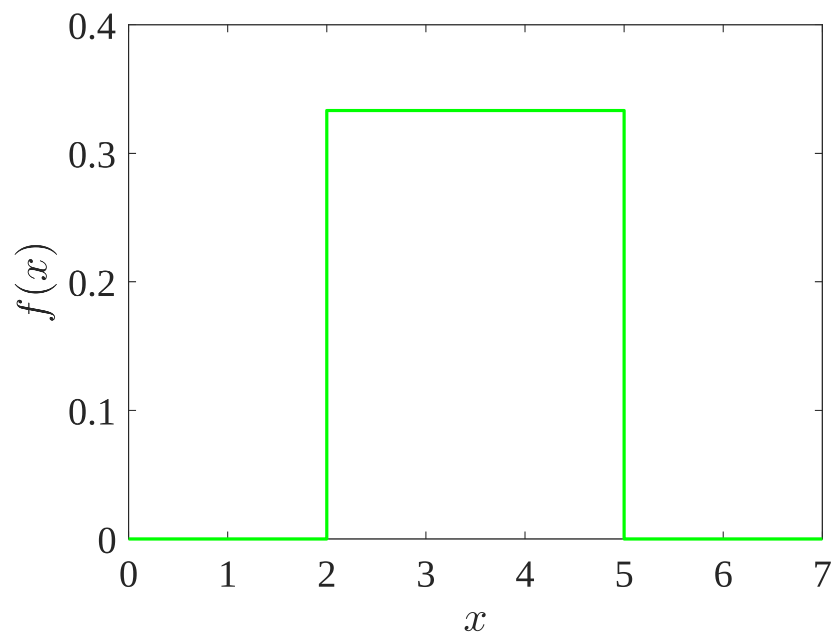
連続型の確率分布

- また，確率変数 X が a と b の間に入る確率が，
$$P(\{a \leqq X \leqq b\}) = \int_a^b f(x)\,dx$$
となる関数 f を確率密度関数と呼びます（ただし，a と b は定数で，$a < b$ とします）．
- 上の式の右辺は関数 f の積分を表していて，下図のように関数 f のグラフの $x = a$ から $x = b$ までの部分と x 軸で囲まれる色付き領域の面積を表しています（詳細は後述の積分の節を参照してください）．



代表的な連続型の確率分布

- 一様分布 $\mathrm{U}(a, b)$
  - a から b の間の区間（ただし，$a < b$）で等しい確率をとる確率分布です.



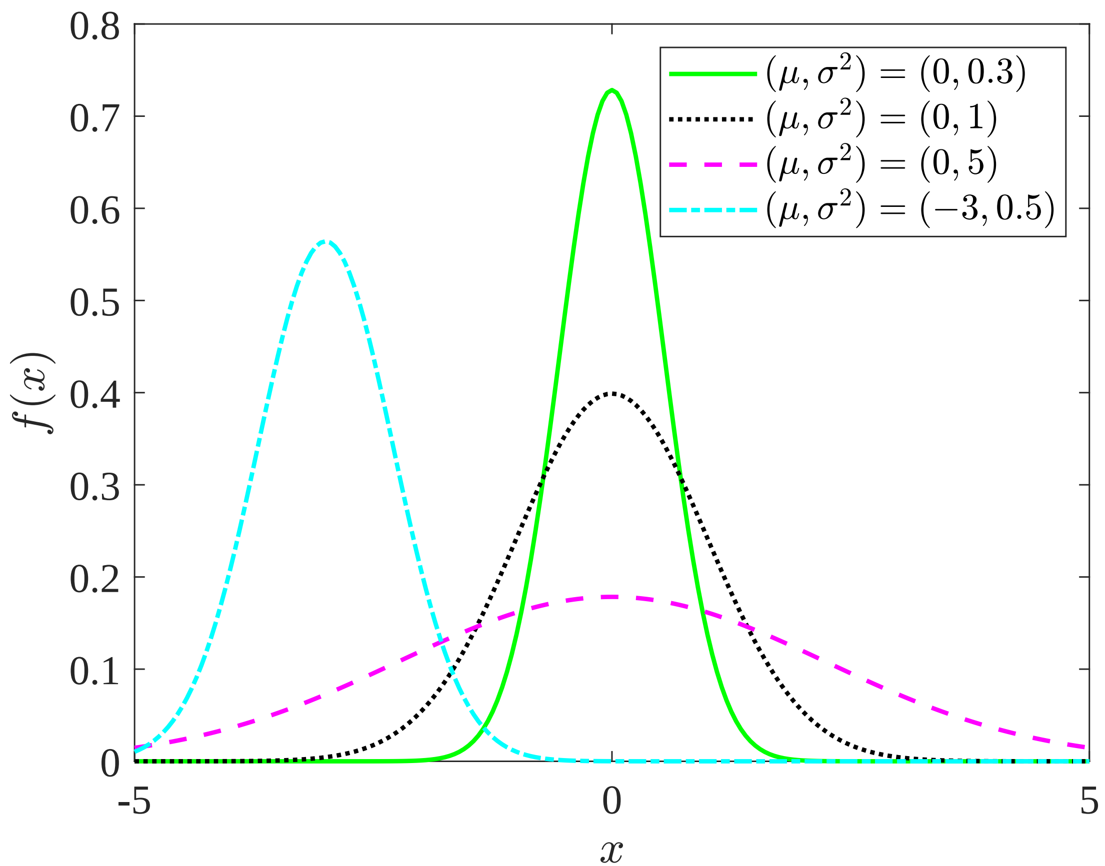
$\mathrm{U}(2, 5)$ の確率密度関数

代表的な連続型の確率分布

- 正規分布（または，ガウス分布） $\mathrm{N}(\mu, \sigma^2)$
  - 統計学で最もよく用いられる確率分布です.
  - 確率密度関数は $f(x) = \dfrac{1}{\sqrt{2\pi}\sigma} \exp\left(-\dfrac{(x-\mu)^2}{2\sigma^2}\right)$.
    （$\exp(x)$ は後述の指数関数を表す）
  - 期待値は μ で，分散は σ^2.
  - $\mathrm{N}(0, 1)$ を，標準正規分布とよびます.
    - X が $\mathrm{N}(\mu, \sigma^2)$ に従うならば，$\dfrac{X-\mu}{\sigma}$ は標準正規分布 $\mathrm{N}(0, 1)$ に従います.

代表的な連続型の確率分布

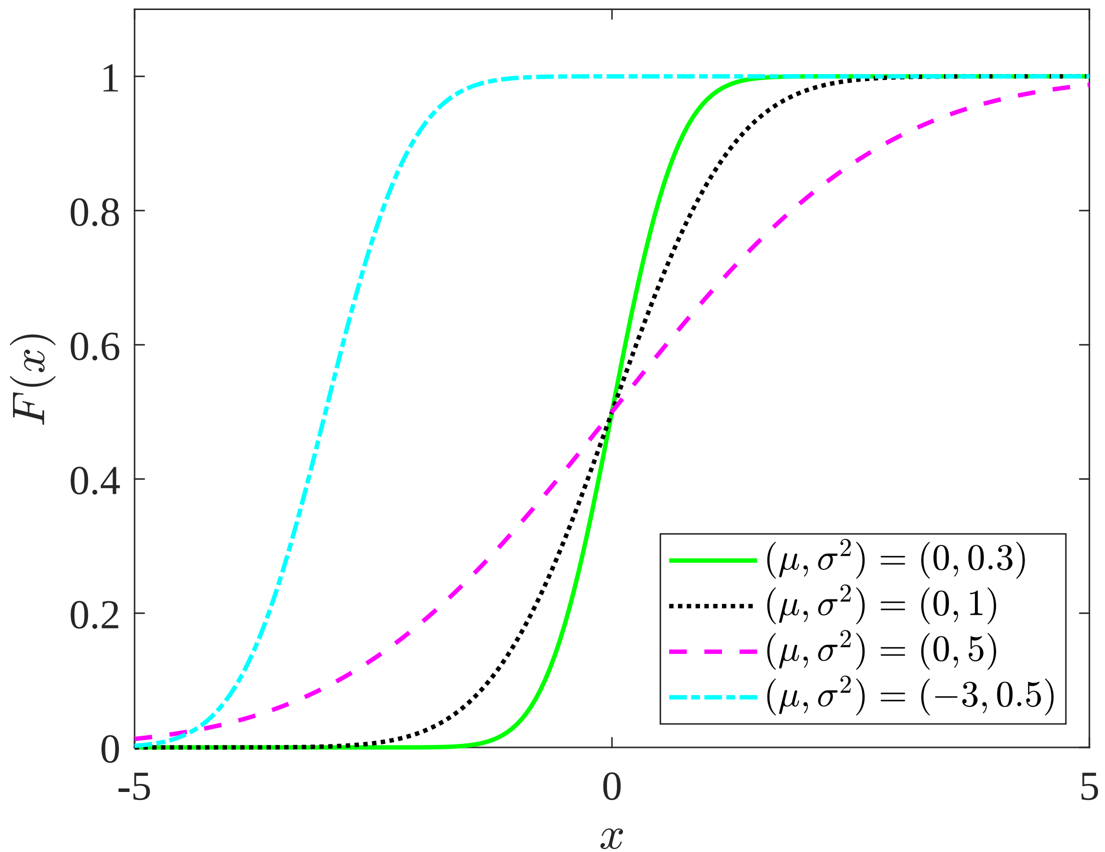
- 正規分布（または，ガウス分布） $\mathrm{N}(\mu, \sigma^2)$
  - 確率密度関数の値 $f(x)$ は，$x = \mu$ で最大になります．グラフは，左右対称で，単峰性をもちます．



$\mathrm{N}(\mu, \sigma^2)$ の確率密度関数

代表的な連続型の確率分布

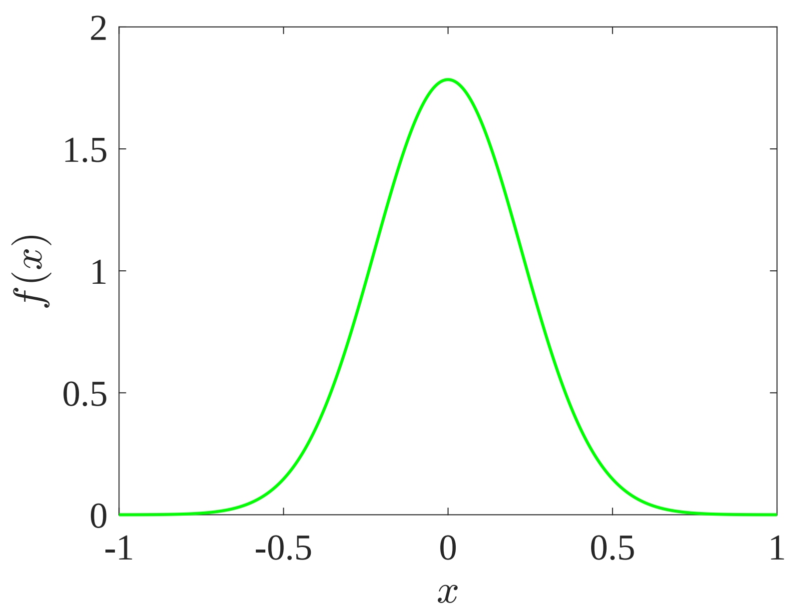
- 正規分布（または，ガウス分布） N(μ, σ^2)
  - 確率分布関数 $F(x)$ のグラフ：



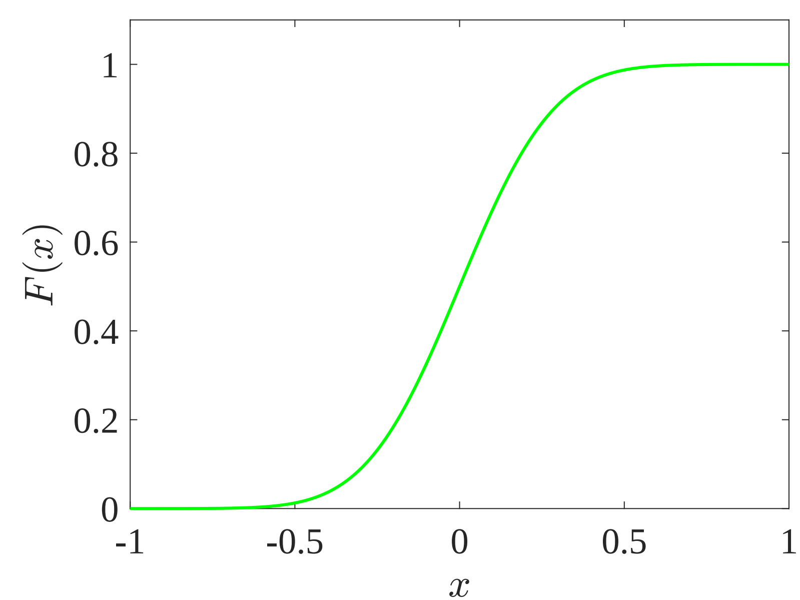
N(μ, σ^2) の確率分布関数

確率密度関数に関する注意

- 確率密度関数の値 $f(x)$ は,「確率変数 X が 値 x をとる確率」ではないことに注意して下さい.
- このため，分散 σ^2 が小さい場合，$f(x) > 1$ となり得ます.
  - あくまで，$\displaystyle\int_a^b f(x)\,dx$ が確率です.
- 確率分布関数の値 $F(x)$ は（確率 $P\{X \leqq x\}$ を表すので），必ず 1 以下です.



N(0, 0.05) の確率密度関数



N(0, 0.05) の確率分布関数

独立同一分布

- 確率変数の列 $X_1, X_2, \cdots$ に対して，$X_1, X_2, \cdots$ がそれぞれ独立に値が定まり，各 j に対して X_j の確率分布が j によらず同じものになっている時，独立同一分布であるといいます．
- 例えば，サイコロを何回も振ったり，同じ条件下で実験を繰り返すといった場合などがあります．サイコロを何回も振る場合，X_j は j 回目のサイコロの目に対応します．
- 繰り返し行われる試行のうち，独立同一分布は最も基本的で，数学的に扱いやすいものとなります．

大数の法則

- 「大数の法則」
  - $X_1, X_2, \cdots$ を独立同分布に従う確率変数列とします．この時，試行回数 n が大きくなるにつれて，データの平均

$$\frac{X_1 + \cdots + X_n}{n}$$

は X_1 の期待値 $E[X_1]$ に近づいていきます．

  - 試行回数 n が少ない時にはデータの平均は試行する度に違う値をとりますが，大数の法則によると試行回数 n が大きくなると，何度試行しても安定的に $E[X_1]$ に近い値をとるということがわかります．
  - ある人が一定期間の間に死亡するかどうかは，基本的には人によって独立で性別・年代別に見れば同じ分布に従っていると考えることができます．従って，大数の法則によって，例えば 40 代男性全体の死亡率は年によってほとんど変わらずに安定することがわかります．生命保険はこのような考え方に基づいて支払う保険金の合計額を予測し，月々に払う保険料を定めています．

1-6- 4　推測統計学

推測統計の目的

- n 個のデータ $x_1, x_2, \ldots, x_n$ が与えられているとします。
- 推測統計学では、データがある確率変数の「観測値」として得られているとモデル化し、その確率変数の従う確率分布に関する情報をデータから抽出することを目指します。
- より正確には、 n 個の確率変数 $X_1, X_2, \ldots, X_n$ があって、観測データ $x_1, x_2, \ldots, x_n$ が標本点ωが得られたときの実現値となっていると考えます。すなわち、$x_i = X_i(\omega), i = 1,2, \ldots, n$ とモデル化します。
- 最も基本的な状況として、 $X_1, X_2, \ldots, X_n$が独立同分布である場合を考えます。これらの確率変数が従う共通の確率分布をQとしたとき、Qに関する情報をデータ$x_1, x_2, \ldots, x_n$ から調べることがここでの目標です。
  - Qは母集団分布、あるいは単に母分布と呼ばれることがあります。

例：コイン投げ

- 表が出る確率がわからないコインを何回か投げて、表が出る確率θを調べることを考えます。

- ５回投げた際の結果が、「表、裏、表、表、表」だったとします。

- 表を１、裏を０とそれぞれ記録することにすれば、観測データとして
$$(x_1, x_2, x_3, x_4, x_5) = (1,0,1,1,1)$$
が得られたことになります。

- 推測統計では、これらを「確率θで１、確率$1-\theta$で０をとる５つの独立な確率変数X_1, X_2, X_3, X_4, X_5の観測値」としてモデル化します。

- このとき、母分布Qはベルヌーイ分布Bɴ$(1,\theta)$となります(67~69ページ参照)。

推測統計の基本概念

- 推定：データをもとに、母分布Qを特徴づけるパラメータ（上の例ではコインの表が出る確率 θ）の値を求めることが目標です。
  - 点推定：パラメータをある１つの値で指定する推定方式です。
    例）コインの表が出る確率θの推定値は0.8である
  - 区間推定：ある確率以上で真のパラメータを含むことが保証される範囲を求める推定方式です。
    例）コインの表が出る確率θは0.7~0.9の範囲に含まれる確率は95%以上である

- 仮説検定：母分布Qに関する仮説がデータと整合的か否かを確率計算で判断することが目標です。
  - 例）コインの表が出る確率θが0.5であると言えるか否かをデータから判断する

平均の点推定

- 確率分布を特徴づける最も基本的な特徴量は平均（期待値）と分散・標準偏差です。

- ここではまず、母分布 Q の平均 μ の点推定について考えます。
  - $X_1, X_2, \ldots, X_n$はいずれもQに従っていますので、すべての$i = 1,2,\ldots,n$について $\mu = E[X_i]$ です。

- μ の値を指定する「自然な」方法として、データの平均

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n}$$

が考えつきます。

平均の点推定

- $\bar{x}$ は、次の確率変数の観測値となっています。

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

- このように、推測統計の枠組みでは、点推定によって指定する値は、一般に何らかの確率変数の観測値となっています。
  - 前者を(点)推定値、後者を(点)推定量と区別することがあります。
  - 今の場合、$\bar{x}$ が推定値で、$\bar{X}$ が推定量です。$\bar{X}$ は標本平均と呼ばれることがあります。

- 推測統計では、推定量の確率的な性質を調べることで、点推定の「よさ」を判断します。

一致性と不偏性

- 標本平均 $\bar{X}$は、点推定量として次の 2 つのよい性質を持っています。

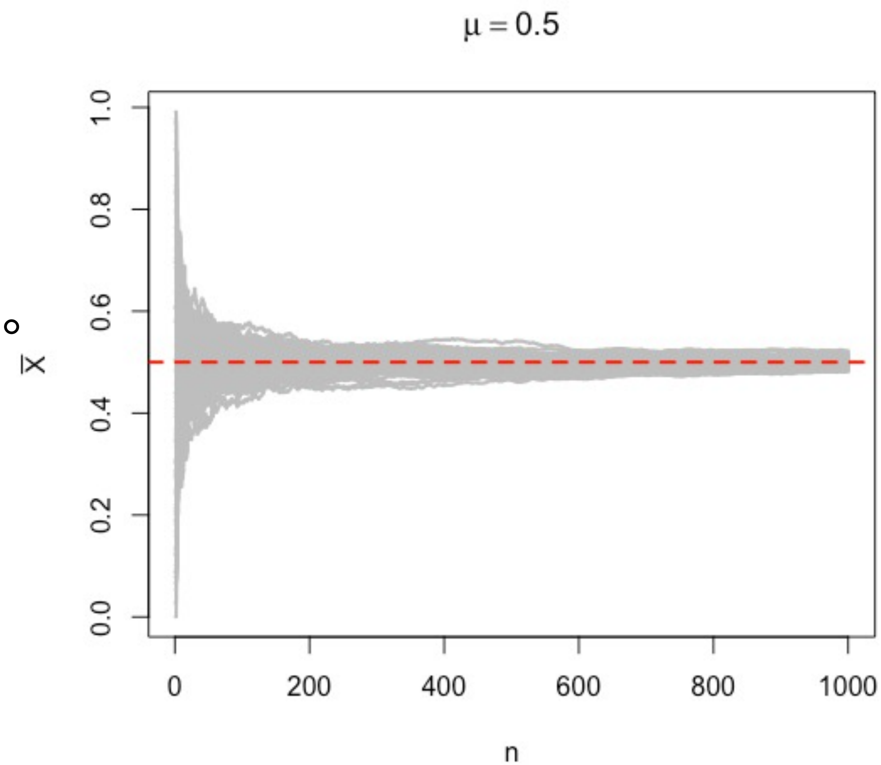
① 一致性：データ数nが大きくなるに従って、 $\bar{X}$の値が推定したい真のパラメーターμに近づいていく（大数の法則による）

② 不偏性：確率変数 $\bar{X}$ の期待値が、推定したい真のパラメーターμに一致する：$E[\bar{X}] = \mu$

一致性

- 母分布がベルヌーイ分布Bɴ(1,0.5)の場合に、「標本平均 $\bar{X}$をサンプル数を1から1000まで増やしながら計算し、その推移をプロットする」という試行を100回繰り返して得られたグラフを右図に示します。この場合の平均は$\mu = 0.5$です。

- 100回のどの試行に対するグラフも、サンプル数nが大きくなるに従って、最終的に$\mu = 0.5$に近づいていく様が見てとれます。



不偏性

- 一般に、未知パラメータθを推定量 $\hat{\theta}$ で推定した際の推定誤差$\hat{\theta} - \theta$は、以下のように「確率誤差」と「バイアス」に分解できます：

$$\hat{\theta} - \theta = \underbrace{\left(\hat{\theta} - E\left[\hat{\theta}\right]\right)}_{\text{確率誤差}} + \underbrace{\left(E\left[\hat{\theta}\right] - \theta\right)}_{\text{バイアス}}$$

  - 確率誤差は、ランダムな要因で発生する誤差で、完全に制御することはできませんが、確率分布の計算で見積もることが可能です(後述)。
  - バイアスは、推定値を一定方向に偏らせる、非ランダムな誤差です。

- 推定量の不偏性とは、バイアスが0であることを意味します。

$$E\left[\hat{\theta}\right] - \theta = 0 \Leftrightarrow E\left[\hat{\theta}\right] = \theta$$

分散の点推定

- 次に、母分布 Q の分散 σ^2 の点推定について考えます。
  - $X_1, X_2, \ldots, X_n$はいずれもQに従っていますので、すべての$i = 1,2,\ldots,n$について $\sigma^2 = V[X_i]$ です。

- 平均の点推定の類推から、σ^2の推定量としては、データの分散に対応する確率変数
$$S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n}$$
を考えるのが自然なように思えます。

- しかし、S^2は一致性は持ちますが、不偏性を持たないことが、S^2の期待値を計算すると以下のようになることからわかります。
$$E[S^2] = \frac{n-1}{n}\sigma^2 = \sigma^2 - \frac{\sigma^2}{n}$$

分散の点推定

- 最後の式から、 S^2は平均的には真の分散 σ^2をσ^2/n だけ過小推定することになります。

- 中央の式から、このバイアスを消すにはS^2に$n/(n-1)$をかけた推定量を考えればよいことになります。

$$s^2 = \frac{n}{n-1}S^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \cdots + (X_n - \bar{X})^2}{n-1}$$

  - s^2は一致性も持っています。
  - s^2は標本分散と呼ばれることもありますが、 S^2と区別するために不偏分散と呼ぶことも多いです。

区間推定

- 未知パラメータθを推定量 $\hat{\theta}$ で点推定する際、通常は「推定誤差」 $\hat{\theta} - \theta$が発生します。

- 推定値がどの程度信頼に足るか評価するのは、この推定誤差の大きさを評価する必要があります。

- 推定誤差も確率変数ですので、「推定誤差が必ず収まる範囲」では、しばしば保守的すぎて、実質的に役に立ちません。

- 例）コインの表が出る確率を、10回投げたうちに表が出た割合で推定する場合、たとえ表が出る確率が0.9であっても、10回連続で裏が出て推定値は0となるかもしれないです。この場合、「推定誤差が必ず収まる範囲」は「0以上1未満」となり、役に立ちません。

区間推定

- 上の例では、10回連続で裏が出るのは100億分の1の確率しかないため、推定誤差の大きさが0.9となるような状況はほぼ起きません。

- このことから、実用的な評価を得るには、ほとんど起きない事象を無視すればよいことが示唆されます。

- 上の議論を踏まえて、統計学では以下のようにして推定誤差の大きさを評価します。
  - (小さい)確率 α を 1 つ定めて、α 未満の確率でしか起きない事象は無視することにします($\alpha = 0.05$や$\alpha = 0.01$がよく使われます)。
  - 推定誤差が $1 - \alpha$ 以上の確率で収まる範囲を求めます。数式で述べると、

$$P\left(l \leq \hat{\theta} - \theta \leq u\right) \geq 1 - \alpha$$

を満たすl, uをデータから求めます。

区間推定

- 実際に興味があるのは推定誤差そのものではなく未知パラメータθの値ですので、上の式をθが収まる範囲で述べ直すと、
$$P\left(\hat{\theta} - u \leq \theta \leq \hat{\theta} - l\right) \geq 1 - \alpha$$
となって、「$1 - \alpha$ 以上の確率で θ は$\hat{\theta} - u$ から$\hat{\theta} - l$ の範囲にある」となります。

- 推定誤差は相対誤差$\hat{\theta}/\theta - 1$などで評価する方法もありますので、より一般的な形で、推定誤差を考慮した推定方式を述べると、
$$P(L \leq \theta \leq U) \geq 1 - \alpha$$
を満たすL, Uをデータから求める、ということになります。

- このような推定方式を区間推定と呼びます。「L 以上 U 以下の範囲」という区間は$100(1-\alpha)$%信頼区間と呼ばれます。

例：正規母集団の平均の区間推定（分散既知）

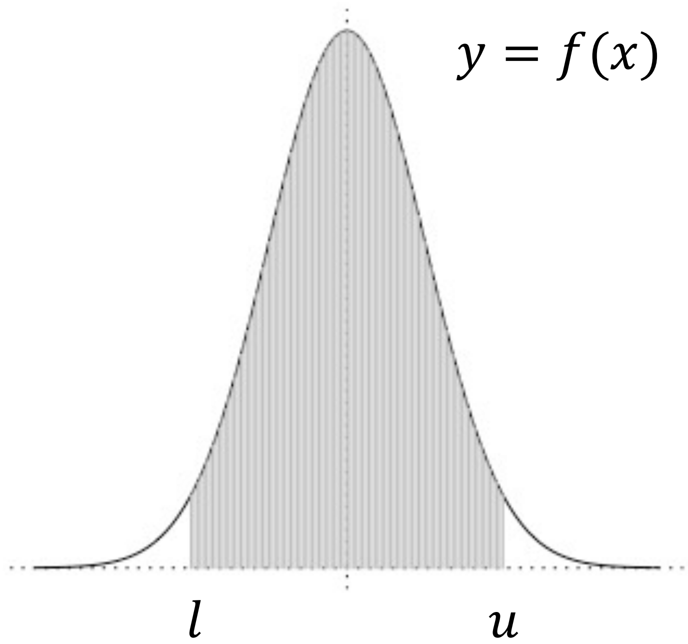
- 母分布が平均 μ 、分散 1の正規分布であるとします(76~78ページ参照)。
- 標本平均 $\bar{X}$ で μ を推定する場合、推定誤差を$\sqrt{n}$ 倍した量 $\sqrt{n}(\bar{X}-\mu)$ は平均0、分散1の正規分布（これを標準正規分布と呼びます）に従うことが知られています。
- 従って、$\sqrt{n}(\bar{X}-\mu)$ が l から u の範囲に含まれる確率は
$$P(l \leq \sqrt{n}(\bar{X}-\mu) \leq u) = \int_l^u f(x)dx$$
のように求めることができます。ここで、$f(x)$は標準正規分布の確率密度関数を表します：
$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

例：正規母集団の平均の区間推定（分散既知）

- 上の式は、確率 $P(l \leq \bar{X} - \mu \leq u)$が$y = f(x)$のグラフと$x$軸で l から u の範囲において囲まれる部分の面積と一致します。（下図斜線部）

- 右図の斜線部分の面積が$1-\alpha$以上になるように、l, uを決めてやればよいことになります。

- 通常、推定誤差がプラス側に出るのとマイナス側に出るのは等価に見ますので、右図の左右の白抜き部分の面積がそれぞれ$\alpha/2$となるようにl, uを決めます。そのようなl, uの値をそれぞれ標準正規分布の下側100α/2%点、上側100α/2%点と呼びます。



例：正規母集団の平均の区間推定（分散既知）

- 例えば、$\alpha = 0.05$に対応する標準正規分布の下側2.5%点、上側2.5%点は、それぞれおよそ-2、2です。（下表参照）

α	0.1	0.05	0.01
下側$100\alpha/2$%点	-1.64	-1.96	-2.58
上側$100\alpha/2$%点	1.64	1.96	2.58

- 従って、

$$P(-2 \leq \sqrt{n}(\bar{X} - \mu) \leq 2) \approx 0.95$$

となりますから、μの95%信頼区間は、およそ

$$\bar{X} - \frac{2}{\sqrt{n}} \text{ 以上 } \bar{X} + \frac{2}{\sqrt{n}} \text{ 以下の範囲}$$

となります。

例：正規母集団の平均の区間推定（分散既知）

- より一般に、μの$100(1-\alpha)$%信頼区間は、標準正規分布の下側$100\alpha/2$%点、上側$100\alpha/2$%点をそれぞれ l_α, u_αと書くと、
$$\bar{X} - \frac{u_\alpha}{\sqrt{n}} \text{ 以上 } \bar{X} - \frac{l_\alpha}{\sqrt{n}} \text{ 以下の範囲}$$
で与えられます。

- ここまでの議論では母分布の分散を1としていましたが、より一般に分散がσ^2であっても、その値がわかっていれば、$\sqrt{n}(\bar{X}-\mu)/\sigma$ が標準正規分布に従うという事実を利用して、μの信頼区間を構成できます。

- この場合、μの$100(1-\alpha)$%信頼区間は、
$$\bar{X} - \sigma\frac{u_\alpha}{\sqrt{n}} \text{ 以上 } \bar{X} - \sigma\frac{l_\alpha}{\sqrt{n}} \text{ 以下の範囲}$$
で与えられます。

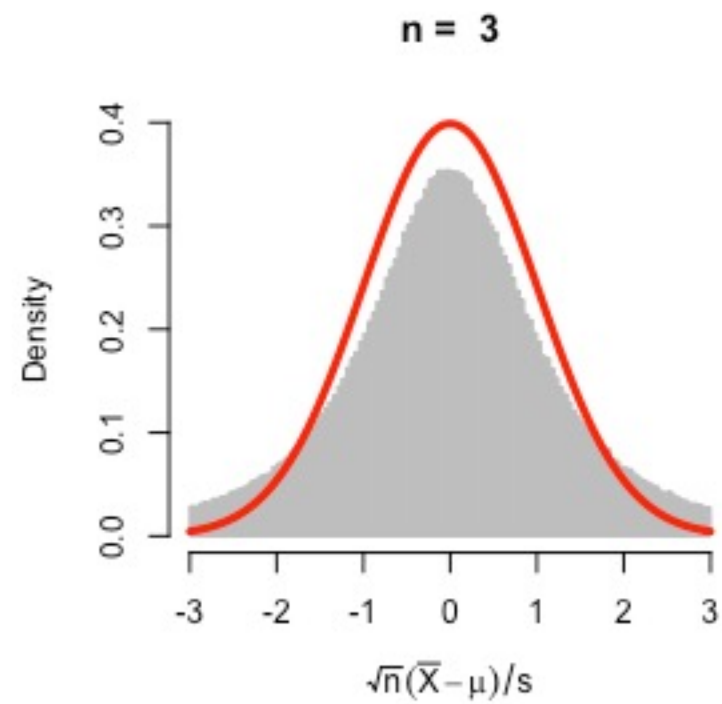
正規母集団の平均の区間推定（分散未知）

- 応用上は、母分布の分散σ^2の値がわかっているということはほとんどありません。

- その場合、 σ^2の値を不偏分散s^2 で推定するのが自然です。

- しかし、 $\sqrt{n}(\bar{X} - \mu)/\sigma$ のσをsで置き換えた量$\sqrt{n}(\bar{X} - \mu)/s$は標準正規分布に従いません。

- 右図に$n = 3$の場合に$\sqrt{n}(\bar{X} - \mu)/s$を100万個シミュレーションして描いたヒストグラム（灰）と標準正規分布の確率密度関数（赤）を描画していますが、明らかにずれています。



χ^2分布

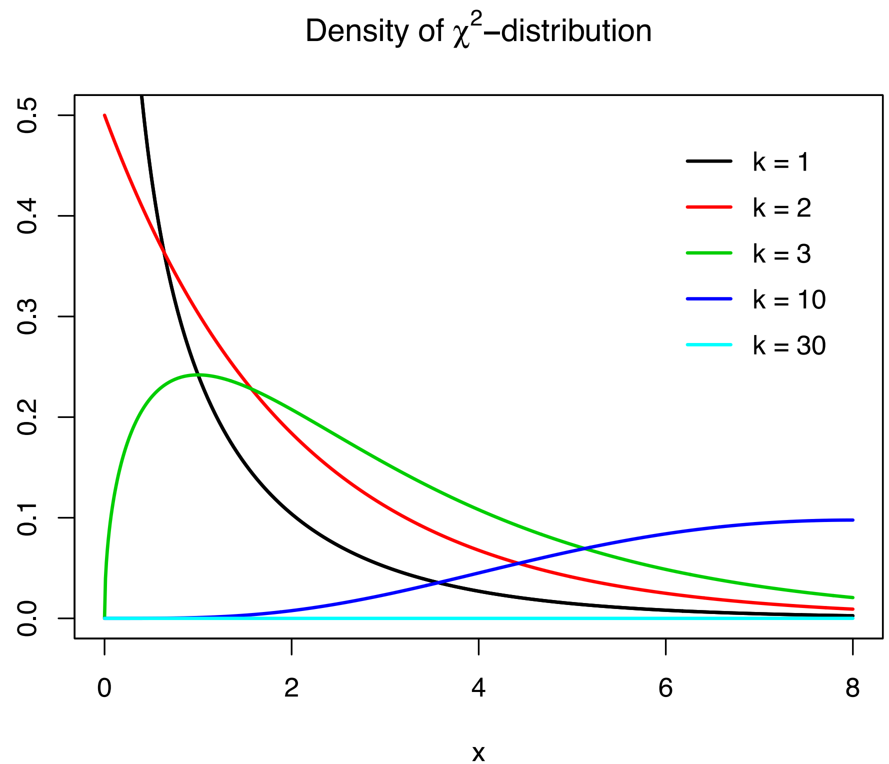
- $\sqrt{n}(\bar{X} - \mu)/s$ は自由度$n-1$のt分布と呼ばれる確率分布に従うことが知られています。t分布を定義するために、まずはχ^2分布と呼ばれる確率分布を定義します。

- $Z_1, Z_2, \ldots, Z_k$を標準正規分布に従う独立なk個の確率変数とします。

- このとき、確率変数
$$Y = Z_1^2 + Z_2^2 + \cdots + Z_k^2$$
の従う確率分布を自由度kのχ^2分布と呼びます。

χ^2分布

- χ^2分布は連続分布である(確率密度関数を持つ)ことが知られており、確率密度関数は下図のようなグラフを持っています。



t分布

- 自由度kのt分布は次のように定義されます。

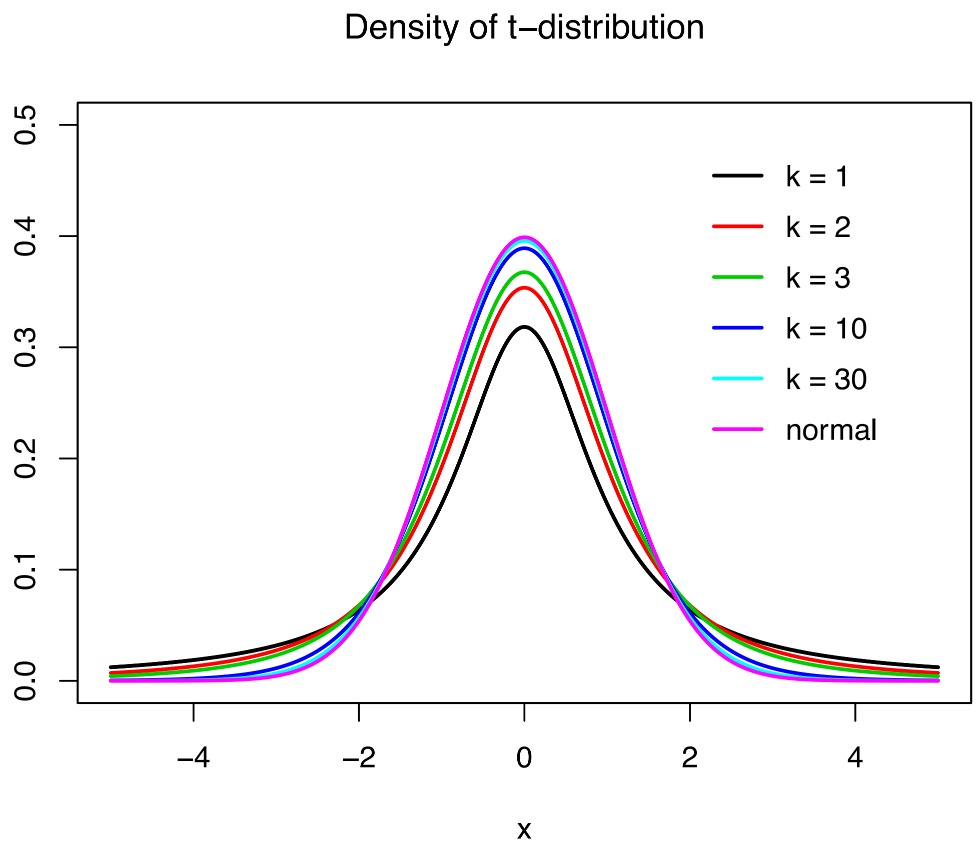
- Zを標準正規分布に従う確率変数、Yを自由度kのχ^2分布に従う確率変数とし、ZとYは独立であるとします。このとき、確率変数

$$t = \frac{Z}{\sqrt{Y/k}}$$

の従う確率分布を自由度kのt分布と呼びます。

t分布

- t分布は連続分布である(確率密度関数を持つ)ことが知られており、確率密度関数は下図のようなグラフを持っています。



正規母集団の平均の区間推定（分散未知）

- 母分布が平均μ、分散σ^2の正規分布ならば、$(n-1)s^2/\sigma^2$は自由度kのχ^2分布に従い、かつ$\bar{X}$と独立であることが知られています。

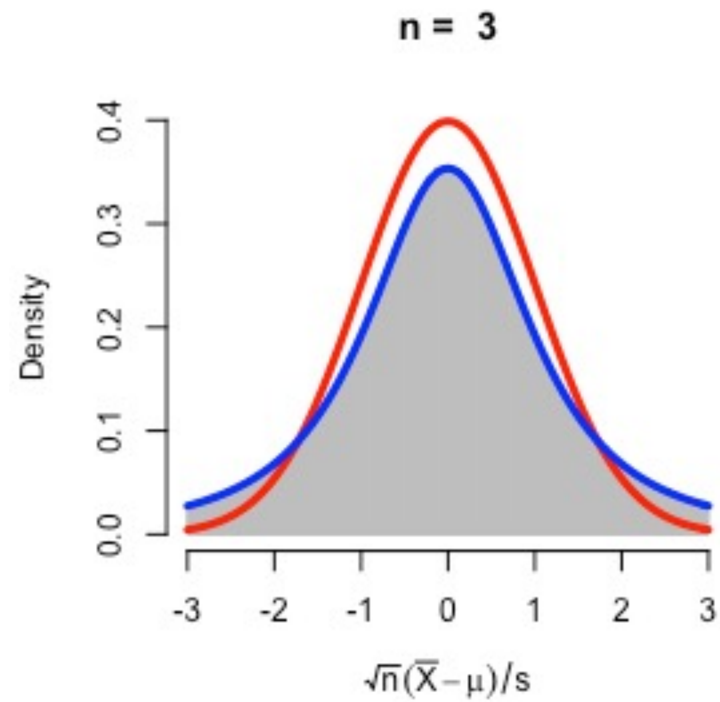
- $\sqrt{n}(\bar{X}-\mu)/\sigma$ は標準正規分布に従い、

$$\frac{\sqrt{n}(\bar{X}-\mu)}{s} = \frac{\sqrt{n}(\bar{X}-\mu)/\sigma}{\sqrt{\dfrac{(n-1)s^2/\sigma^2}{n-1}}}$$

と書き直せるので、 t分布の定義から、$\sqrt{n}(\bar{X}-\mu)/s$ は自由度$n-1$のt分布に従うことがわかります。

正規母集団の平均の区間推定（分散未知）

- 先ほどの$n = 3$の場合の$\sqrt{n}(\bar{X} - \mu)/s$のヒストグラムに自由度2の$t$分布の確率密度関数を上書き(青)すると、確かに重なることが確認できます。



- 従って、 μの$100(1-\alpha)$%信頼区間は、自由度$n-1$のt分布の下側$100\alpha/2$%点、上側$100\alpha/2$%点をそれぞれtl_α, tu_αと書くと、
$$\bar{X} - s\frac{tu_\alpha}{\sqrt{n}} \text{ 以上 } \bar{X} - s\frac{tl_\alpha}{\sqrt{n}} \text{ 以下の範囲}$$
で与えられます。

仮説検定：考え方

- （統計的）仮説検定は、母分布に対する仮説の真偽を、データの観測値に基づいて検証するための枠組です。

- 例）あるコインの表が出る確率が0.5であるか否か検証したいとします。コインを20回投げた結果は以下の通りでした。

1	2	3	4	5	6	7	8	9	10
表	表	裏	表	表	裏	表	裏	表	表
11	12	13	14	15	16	17	18	19	20
表	表	表	裏	裏	表	表	表	表	表

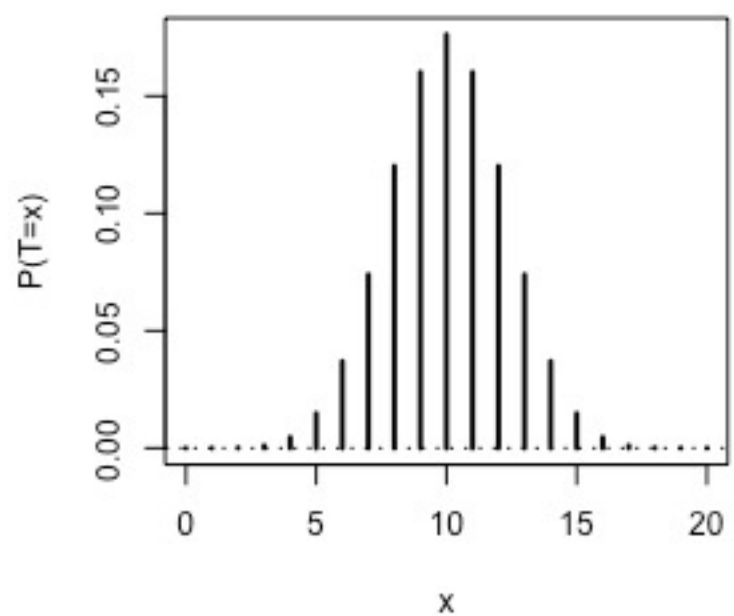
- この結果から、「このコインの表が出る確率は0.5である」という主張の信憑性を定量的に判断するにはどうすればよいでしょうか？

仮説検定：考え方

- 仮に、「このコインの表が出る確率は0.5である」という主張が正しかったとしましょう。

- この場合、コインを20回投げた際に表が出る総数をTとすると、Tは二項分布BN(20,0.5)に従う確率変数です。

- 従って、Tの期待値は10であり、Tがxという値をとる確率$P(T = x)$も右図のように計算できます。



仮説検定：考え方

- 特に、Tの観測値が期待値10からいくつかずれる確率を具体的に計算することができます。

- 下表に、$k = 1,2, \ldots, 6$に対して、Tの観測値が10から$\pm k$以上大きくずれる確率を示します。

k	1	2	3	4	5	6
確率	0.824	0.503	0.263	0.115	0.041	0.012

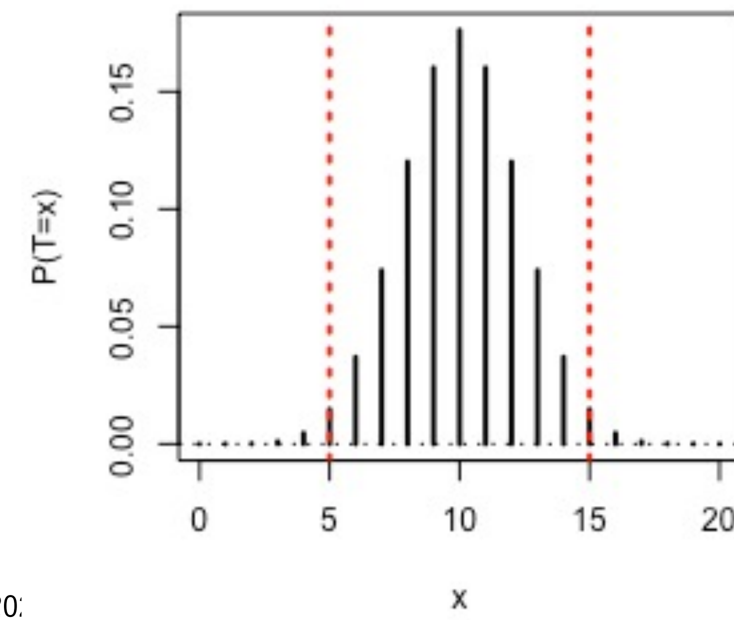
- ところで、今の例では、実際のデータから計算されるTの観測値は15ですので、「コインの表が出る確率は0.5である」という主張を信じる限り、上の議論からこのような結果が得られる確率は5%以下ということになります。

仮説検定：考え方

- 5%という確率を非常に小さいと考えると、このような結果が得られたのは、『運悪くそうなった』か、『そもそも「コインの表が出る確率は0.5である」という仮説が誤っている』のどちらかということになります。

- 運のせいにするのは非科学的なので、「コインの表が出る確率は0.5である」という仮説が誤っている、と判断するのが妥当だろうということになります。

赤点線から外側(赤点線含む)の値が観測される確率は5%以下です。



仮説検定：手順のまとめ

以上の議論をまとめると、以下のようになります。

1. 検証したい仮説 H_0 を設定します。
   - 上の例では「コインの表が出る確率は0.5である」が H_0 にあたります。
2. 「適当な」統計量 T を定め、仮説 H_0 の下で T が値をとる確率が低い領域 R を決めます。
   - 上の例では、T は「コインを20回投げた際に表が出る総数」で、R は「Tの値が、H_0が正しい場合の期待値10から±4より大きくずれる」となります。(5%を「低い確率」と考えています。)
3. T の実際の値 t_0 をデータから計算します。
   - 上の例では $t_0 = 15$ です。
4. t_0 が R に含まれる場合、H_0は誤りだと判断します。
   t_0 が R に含まれない場合、H_0はデータとは「矛盾しない」と判断します。

仮説検定：用語

- 手順１で設定する仮説 H_0 を帰無仮説と呼びます。
- 手順２で定める「適当な」統計量 T を検定統計量と呼びます。
- 手順２の「低い確率」を有意水準と呼びます。
- 手順２で決める領域 R を棄却域と呼びます。
- 手順４で
  - H_0 は誤りだと判断することを、H_0 を棄却すると言います。
  - H_0 はデータとは「矛盾しない」と判断することを、H_0 を受容すると言います。

仮説検定：用語

- 検定結果には２種類の誤った判断がありえます（下表参照）
  - 両者の発生はトレードオフの関係にあります。

- 仮説検定の方針：
  - 第一種の過誤が起きる確率を有意水準以下に抑えた下で、第二種の過誤が起きる確率をできる限り小さくすることを目標とします。

	帰無仮説が偽	帰無仮説が真
帰無仮説を棄却	正しい判断	第一種の過誤
帰無仮説を受容	第二種の過誤	正しい判断

対立仮説

- 検定統計量と棄却域は、上の目標を達成できるように定める必要があります。
  1. 第一種の過誤が起きる確率を有意水準以下となるように定める必要があります。このためには、帰無仮説が正しい場合の検定統計量の確率分布（帰無分布）の評価が必要です。
  2. 第二種の過誤が起きる確率をできる限り小さくする必要があります。このためには、 帰無仮説が誤りの場合に検定統計量が棄却しやすくなるように棄却域を設定する必要があります。

- 後者のためには、「帰無仮説が誤りの場合」に起こりうるシナリオを想定する必要があり、これを対立仮説と呼びます。
  - 上の例では、「コインの表が出る確率は0.5ではない」(帰無仮説の否定)が対立仮説でした。

対立仮説

- 仮説検定では、検定結果が積極的に支持されるのは、帰無仮説が棄却された場合だけです。
  - 「帰無仮説が正しい場合に誤って棄却してしまう」場合はまれであることは保証されていますが(有意水準以下の確率でしか起きない)、「帰無仮説が誤りだが受容してしまう」確率が小さいかはわからないからです。
  - この理由から、「立証したい主張」を対立仮説にとることが多いです。

- 何らかの理由で「帰無仮説が誤りの場合」に起こりうるシナリオに制限を加えることができる場合、その情報を対立仮説に反映させることで、第二種の過誤が起きる確率を小さくできることがあります。

対立仮説

- 例）先ほどのコイン投げの例で、検証の目的が「コインの表が出る確率が不正に高くないか確認する」ことであったとしましょう。
  - 例えば、コインを投げて表が出たら相手が勝利するようなゲームで、コインを用意したのが相手だった場合を想定します。
- この場合、「表が出る確率が0.5より小さいケース」は起きるシナリオとして想定する必要がありません。
  - もしそうであったとしても、こちらは不利益を被らないので、帰無仮説を棄却して「このコインは不正である」と主張する必要はないからです。
- 従って、対立仮説としては「コインの表が出る確率は0.5より大きい」をとればよいことになります。

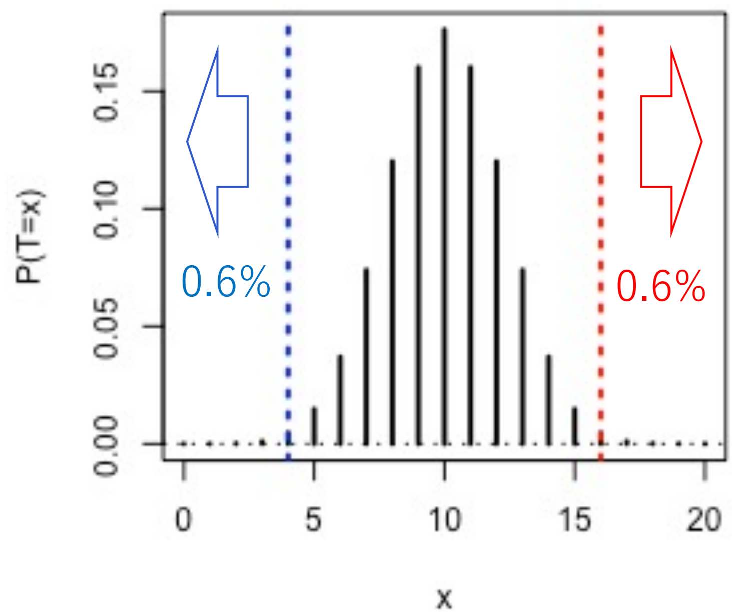
対立仮説

- この場合、対立仮説の下では、検定統計量 T (表の総数)は大きい値をとることが想定されるので、帰無仮説の下での期待値10より大きく外れた場合だけ帰無仮説を棄却すればよいことになります。
- 従って、棄却域は「T の値が$10 + k$ 以上となる」という形の領域に設定すればよいことになります。
    - kは帰無仮説の下でT の値が$10 + k$ 以上となる確率が有意水準以下となるように定めます。

対立仮説

- 例えば有意水準が1%の場合、棄却域は「T の値が16以上となる」です。
  - この場合、Tの観測値が16であれば帰無仮説を棄却できます。
  - 対立仮説が「コインの表が出る確率は0.5ではない」の場合、帰無仮説の下で「T の値が10から±6以上大きく外れる」確率は1.2%程度あるので、Tの観測値が16では帰無仮説を棄却できません。

赤点線から右側(赤点線含む)の値が観測される確率は0.6%程度ですが、赤・青点線の外側(点線含む)に値が観測される確率は1.2%程度です。



片側検定・両側検定

- 先の例のように、対立仮説の設定の仕方は、多くの場合棄却域の形状に影響を与えます。
  - 棄却域が左右両側にある検定は、両側検定と呼ばれます。（上の例では、対立仮説が「コインの表が出る確率は0.5ではない」の場合）
  - 棄却域が右側のみにある検定は、右片側検定と呼ばれます。（上の例では、対立仮説が「コインの表が出る確率は0.5より大きい」の場合）
  - 棄却域が左側のみにある検定は、左片側検定と呼ばれます。（上の例では、対立仮説が「コインの表が出る確率は0.5より小さい」の場合）

P値

- 検定統計量の値 t_0 がデータから計算されたとき、多くの場合、「帰無仮説を棄却できる有意水準の最小値」が計算できます。この値をP値と呼びます。
  - コイン投げの例では、「帰無仮説の下で、T が 10から$\pm|\mathrm{t}_0 - 10|$以上ずれる確率」がP値です。
  - 上の例の $t_0 = 15$ の場合は、P値 = 0.041 です。

- 検定のP値が計算できると、有意水準 α に対して以下の関係が成立します：

$$\mathrm{P値} < \alpha \Leftrightarrow \text{有意水準 } \alpha \text{ で帰無仮説は棄却される}$$

- P値を使うと、有意水準を決めずに検定結果を要約できるため、「有意水準を調整して検定結果を操作する」といった不正を防げます。そのため、通常は検定結果だけでなくP値も報告します。