

# 講義「情報理論」

## 第2回 事前知識の準備(確率, 対数計算)

情報理工学部門 情報知識ネットワーク研究室  
喜田拓也

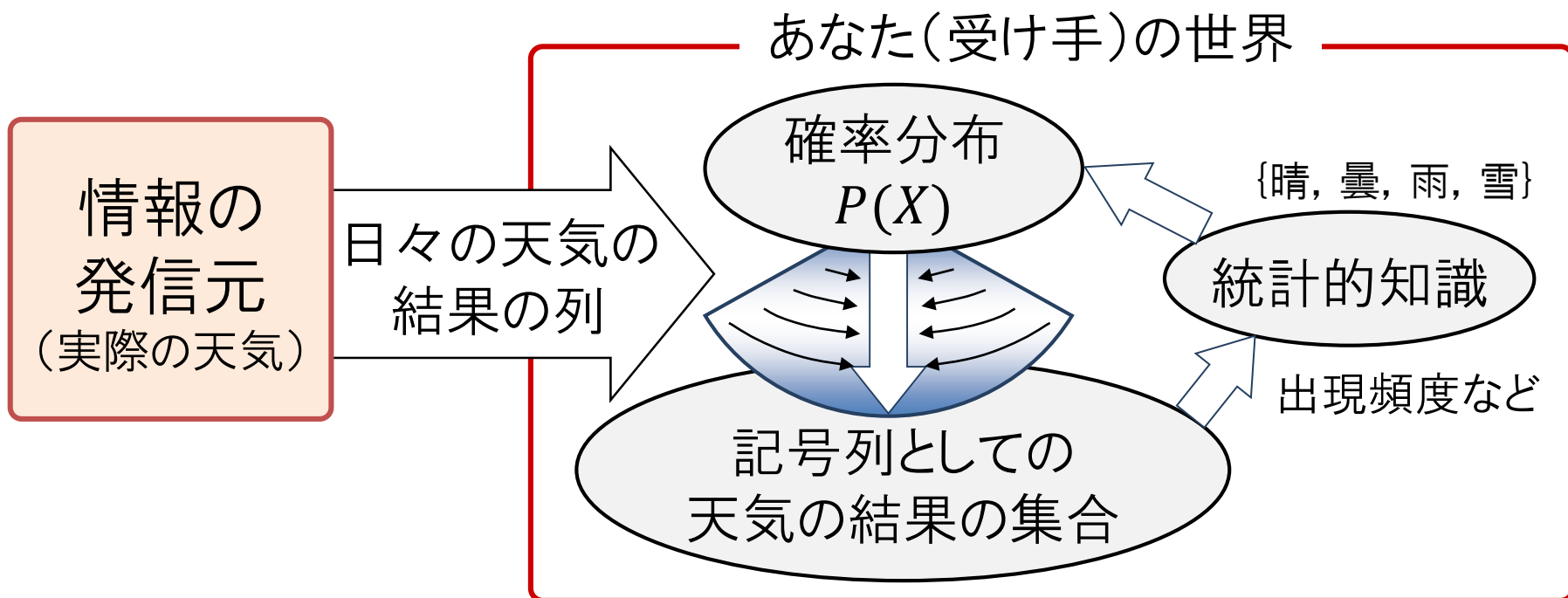
# 情報理論での情報伝達のモデル(おさらい)

情報を だと割り切って取り扱う

その意味内容には立ち入らない

送信される情報は記号 {晴, 曇, 雨, 雪} の列

あなた(受け手)の世界は, 記号(列)の集合とその統計的知識



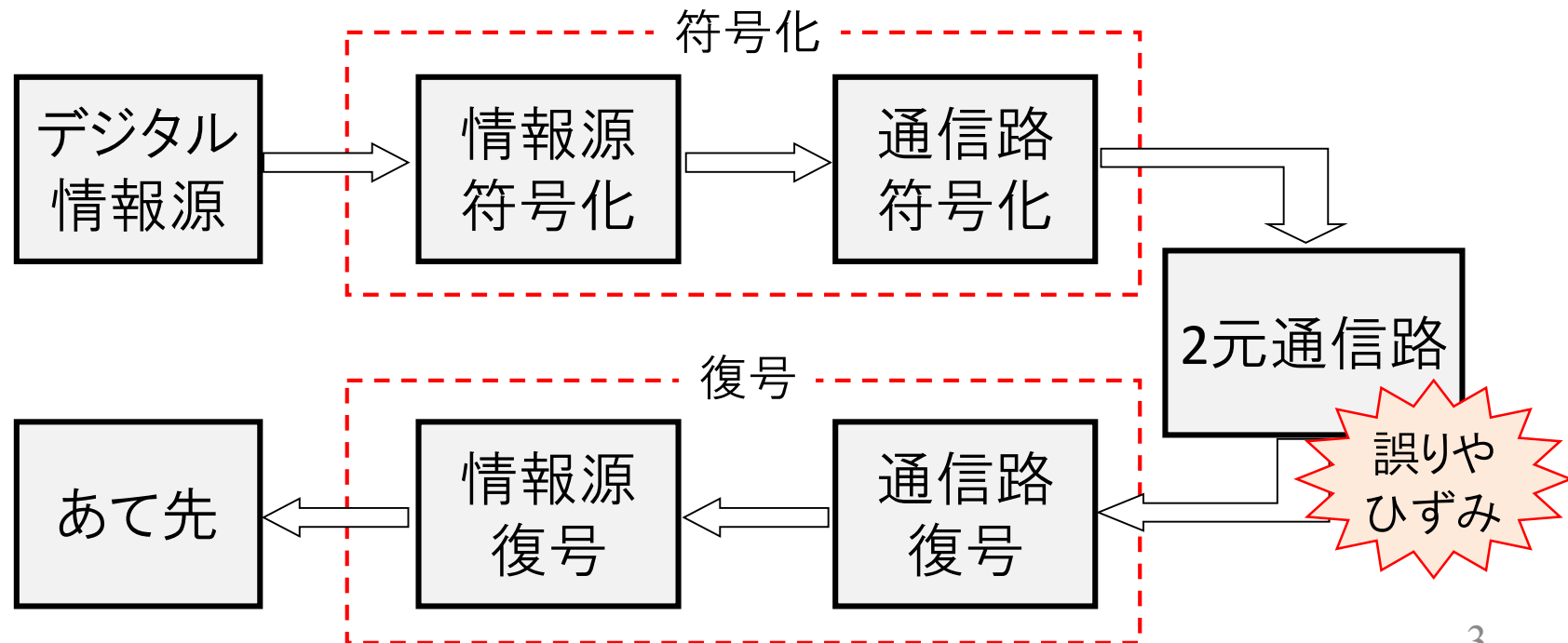
# 情報理論が取り組む4つの問題(おさらい)

【問題1】

【問題2】

【問題3】

【問題4】



# 今日の内容

付録A 対数とその性質

確率について基本的なお話

# 対数の定義

## 定義A.1:

$a \neq 1$  を正の実数とする. このとき, 任意の実数  $x > 0$  に対し

$$x = a^p$$

を満たす実数  $p$  が唯一に決まる. これを  $p = \log_a x$  と書き,  $a$  を底とする  $x$  の**対数**(logarithm)と呼ぶ.

## 定義A.2:

$$e = \lim_{t \rightarrow 0} (1 + t)^{\frac{1}{t}} = 2.71828182845904 \dots$$

で定義される定数  $e$  をネイピア数(またはオイラー数)という.

- 底が10のときを**常用対数**( $\text{Log } x$ )
- 底が  $e$  のときを**自然対数**( $\ln x$ )
- 底が2のときを**二進対数**( $\lg x, \log x$ )

情報理論では  
二進対数が  
良く用いられる

# 対数の基本的な定理

定理A.1:  $a \neq 1$  を正の実数とする. 任意の正の実数  $x, y$  と任意の実数  $b$  に対し, 以下が成り立つ.

$$(1) a^{\log_a x} = x$$

$$(2) \log_a a = 1$$

$$(3) \log_a 1 = 0$$

掛け算が足し算に!

割り算が引き算に!

$$(4) \log_a xy = \log_a x + \log_a y \quad (4') \log_a \frac{x}{y} = \log_a x - \log_a y$$

$$(5) \log_a x^b = b \log_a x \quad (5') \log_a \frac{1}{x} = \log_a x^{-1} = -\log_a x$$

$$(6) a > 1 \text{ かつ } x < y \text{ のとき, } \log_a x < \log_a y$$

$$(7) a > 1 \text{ のとき, } \lim_{x \rightarrow +\infty} \log_a x = +\infty$$

定理A.2: 正の実数  $a, b$  ( $\neq 1$ ) および任意の正の実数  $x$  に対し,

$$\log_a x = \frac{\log_b x}{\log_b a}$$

底の変換!

が成り立つ.

# ノート A.1

$x$ の $a$ を底とした対数 $\log_a x$ は、 $x$ を $a$ 進数表示したときの桁数 (記述長)に対応する。

例えば、ある自然数 $n$ の10進数での桁数を $m$ とする。このとき、

$$10^{m-1} \leq n < 10^m .$$

各辺の対数をとると

$$\begin{aligned} m - 1 &\leq \log_{10} n < m , \\ \therefore \log_{10} n < m &\leq \log_{10} n + 1 . \end{aligned}$$

したがって、

$$m = \lfloor \log_{10} n + 1 \rfloor .$$

床関数  $\lfloor x \rfloor$  :  
 $x$  以下の最大の整数

例題: 10進数で225は、二進数で何桁か？

※ Kenneth E. Iverson が1962年に導入

$$\begin{aligned} \log_2 225 &= \log_2 5^2 \cdot 3^2 = 2 \log_2 5 + 2 \log_2 3 \\ &\doteq 2(2.322 + 1.585) = 7.814 \end{aligned}$$

11100001

すなわち、 $7 \leq \log_2 225 < 8$ なので、8桁である。

# $y = \log_a x$ の性質

$a > 1$  のとき,  $\log_a x$  は単調増加  
( $x$  に対して **なだらかに増加**する)

$$x \rightarrow +\infty \Rightarrow \log_a x \rightarrow +\infty$$

$$x < y \Leftrightarrow \log_a x < \log_a y$$

$0 < x < 1$  の間は負の値をとる

$$x \rightarrow 0 \Rightarrow \log_a x \rightarrow -\infty$$

定理A.3:  $\frac{d}{dx} \log_e x = \frac{1}{x}$  .

一般的には,  $\frac{d}{dx} \log_a x = \frac{\log_a e}{x}$  .

定理A.4: 正の実数  $a \neq 1$  に対して, 次が成り立つ.

- (1)  $y = \log_a x$  の曲線は  $y = a^x$  の曲線と直線  $y = x$  に関して対称
- (2)  $\ln x \leq x - 1$

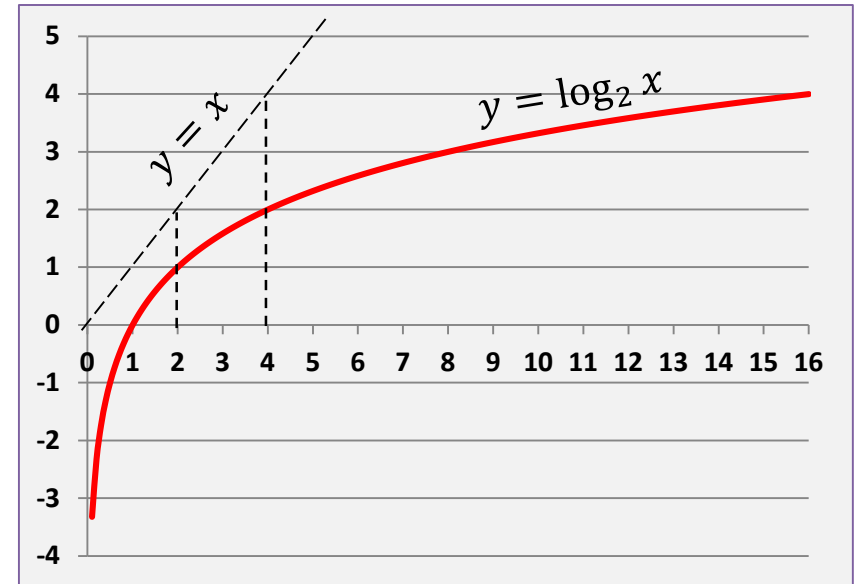


図  $y = \log_2 x$  のグラフ

Try 問3, 問4



# ちよつと休憩

# 確率とは

「ある**事象**についての**起こりやすさの指標**」  
もとは、賭け事の戦略や、金銭の公平な分配を  
考えるために用いられた概念

「1個のサイコロを振って、奇数だったら私の勝ち、偶数だったらあなたの勝ち。参加費は100円。勝ったら180円をあげるよ。勝負する？」



# 事象，起こりやすさの指標って？

偶然現象において，起こりうる事柄を**事象**という  
それぞれの事象が起こりうる可能性を，すべての  
起こりうる事象に対する割合として表すとき，その  
値を**確率**という

古典的な  
確率の定義

$$P(A) = \frac{r}{n}$$

← 事象Aが起こる場合の数  
← すべての場合の数



偶然現象って何？

サイコロ振りだって物理現象でしょ？  
ロボットが振ったらいつも同じじゃん

現代の確率の定義は末尾の資料を参照 11

# まあ難しい話はおいて…

「1個のサイコロを振って、奇数だったら私の勝ち、偶数だったらあなたの勝ち. 参加費は100円. 勝ったら180円をあげるよ. 勝負する？」

標本集合  $\Omega = \{1,2,3,4,5,6\}$ , あなたが勝つ事象  $A = \{2,4,6\}$

事象Aが起こる確率  $P(A) = |A|/|\Omega| = \frac{3}{6} = \frac{1}{2}$  確率を表す関数

得られる賞金額を確率変数  $X$  とすると, その確率分布  $P_X$  は

$$P_X(X = 0) = \frac{1}{2}, P_X(X = 180) = \frac{1}{2}.$$

$X$ は確率によって取る値が決まる変数

得られる賞金額の期待値  $E(X)$  は

確率変数の「見込み」の値  
 $E(X) = \sum_i x_i P_X(X = x_i)$

$$E(X) = \sum_{x \in \{0,180\}} x \cdot P_X(X = x) = 0 \times \frac{1}{2} + 180 \times \frac{1}{2} = 90.$$

# 確率の統計的な定義

標本集合のどの要素(根源事象)も同程度に確からしく起こる場合は古典的定義で問題ない

この仮定が適用できない場合は困る

根源事象の  
生起にかたよ  
りがある場合

試行を $n$ 回繰り返し行ったときに事象 $A$ が起こった回数を $n(A)$ とする。このとき、

$$P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n}$$

を事象 $A$ の確率とする。

(現実に無限回の試行は不可能なので極限を取らないことも)

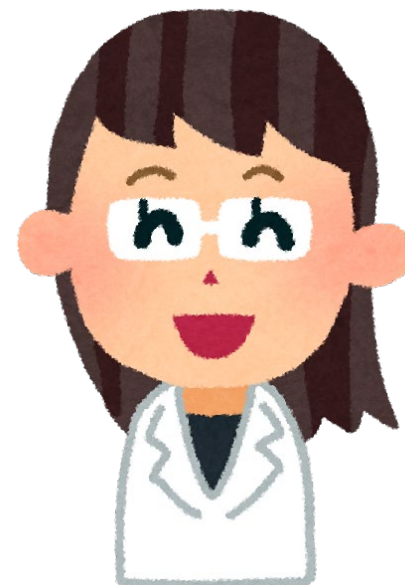
# とある病院での、患者と女医の会話

えっ！そんなに  
少ないの……



大門か……

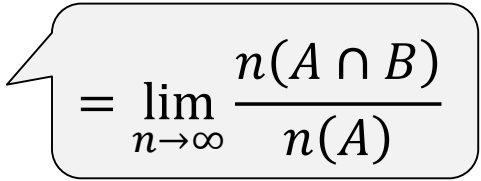
私は失敗しない  
のでっ！



# 条件付き確率と独立性

$P(A) > 0$ のとき,

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


$$= \lim_{n \rightarrow \infty} \frac{n(A \cap B)}{n(A)}$$

を, 事象 $A$ が起こったもとでの事象 $B$ の条件付き確率と定義する.

このとき, 明らかに次式が成り立つ.

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

事象 $B$ の起こる確率が事象 $A$ の生起に無関係な場合, すなわち

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

が成り立つとき, 事象 $A$ と事象 $B$ は独立であるという.

このとき, 明らかに次式が成り立つ.

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

# 条件付き確率の簡単な例

1個のサイコロを振る試行を考える。

サイコロの出目が偶数である事象を $A$ 、  
サイコロの出目が4以上である事象を $B$ とする

事象 $A$ の確率 $P(A)$ は  $P(A) = \frac{3}{6} = \frac{1}{2}$  偶数である確率

事象 $B$ の確率 $P(B)$ は  $P(B) = \frac{3}{6} = \frac{1}{2}$  4,5,6である確率

事象 $A \cap B$ の確率は  $P(A \cap B) = \frac{2}{6} = \frac{1}{3}$  4か6である確率

このとき、事象 $A$ で条件付けた事象 $B$ の確率 $P(B|A)$ は

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{1}{3}}{\frac{1}{2}} = \frac{2}{3}$$

出目が偶数であることは分かっている場合で4以上である確率



# ベイズの定理

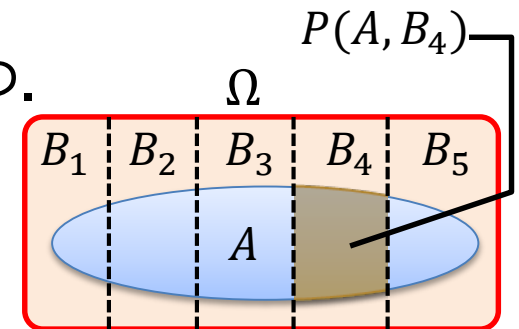
## ベイズの定理

全事象  $\Omega$  を互いに排反な  $n$  個の事象  $B_1, B_2, \dots, B_n$  に分割する。  
すなわち,

$$B_i \cap B_j = \emptyset \quad (i \neq j),$$
$$B_1 \cup B_2 \cup \dots \cup B_n = \Omega.$$

このとき、任意の事象  $A$  に対して次が成り立つ。

$$P(B_k|A) = \frac{P(B_k)P(A|B_k)}{\sum_{i=1}^n P(B_i)P(A|B_i)}$$



次の関係は  $\Omega = B \cup \bar{B}$  の場合の簡略版

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(\bar{B})P(A|\bar{B})} = \frac{P(A, B)}{P(A)}.$$

$P(A, B)$  は  
 $P(A \cap B)$   
と同じ意味

# ベイズの定理を使う例題

ガンを診断するための検査法があり、あなたがそれを受けると仮定しよう。

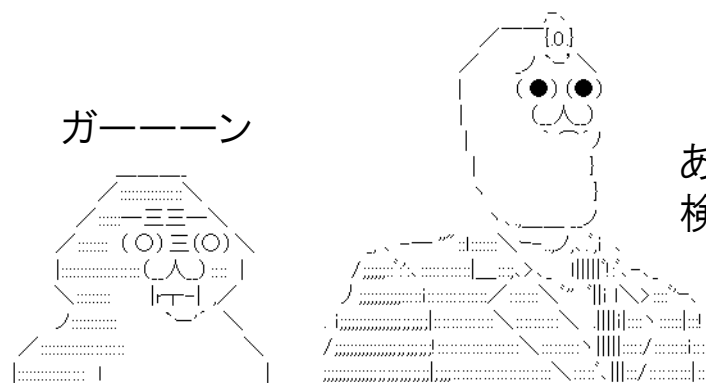
$C$  を、被検査者がガンであるという事象、

$A$  を、検査の結果が被検査者はガンであると示す事象とする。

(つまり検査結果が陽性となる事象)

検査法の認識力が95% ( $P(A|C) = 0.95$  かつ  $P(\bar{A}|\bar{C}) = 0.95$ ) であるとし、  
検査を受ける人の中で実際にガンである割合が  $P(C) = 0.01$  であるとする。

このとき・・・



あなたガンですよ。  
検査の結果から常識的に考えて・・・

検査の結果が陽性だったとき、あなたが本当にガンである確率はいくらか？  
(つまり、 $P(C|A)$ を求めなさい)

# ベイズの定理を使う例題の答え

ベイズの定理より,

$$P(C|A) = \frac{P(C)P(A|C)}{\sum_{i=1}^2 P(C_i)P(A|C_i)} = \frac{P(C)P(A|C)}{P(C)P(A|C) + P(\bar{C})P(A|\bar{C})}$$

である. これに,

$$P(C) = 0.01, \quad P(\bar{C}) = 1 - P(C) = 0.99,$$

$$P(A|C) = 0.95, \quad P(A|\bar{C}) = 1 - P(\bar{A}|\bar{C}) = 0.05$$

を代入すると,

$$P(C|A) = \frac{0.01 \times 0.95}{0.01 \times 0.95 + 0.99 \times 0.05} = \frac{19}{118} \approx 0.161.$$



たったの16%!  
よかったお

Try 問2

# 今日のまとめ

確率の考え方について学んだ

確率の定義

標本集合, 事象, 平均(期待値)

対数とその性質について復習した

次回テーマ

情報量とエントロピー

# 現代の確率の考え方

「サイコロの目がどう出るかは最初のサイコロの持ち方とその振り方によって決まる. そこには何の偶然性もない. ただし, その因果関係はあまりにも複雑なので, 結果を予想するのはほぼ不可能である. . . . (中略). このような予測しにくいこと, 制御しにくいことを「偶然である」とみなすのが確率統計の出発点である. 」

(金谷健一著, 「確率統計を学ぶにあたって」岡山大学工学部講義資料)

現代の確率論は, 「確率」が何を意味するかは問わない. 「確率」が満たすべき数学的な性質を規定し, その性質から導かれる定理を論じる.

公理主義

# 確率の公理による定義

by コルモゴロフ(Kolmogorov)

ある空でない集合  $\Omega$  (結果の集合, **標本集合**) について,  
その任意の部分集合を**事象**という.

事象全体の集合を  $\mathcal{F} (= 2^\Omega)$  とする.

写像  $P: \mathcal{F} \rightarrow [0,1]$  が任意の  $A \in \mathcal{F}$  について次を満たす  
とき,  $P(A)$  を事象  $A$  の**確率**という.

1.  $0 \leq P(A)$

2.  $P(\Omega) = 1$

3.  $A_1, A_2, \dots \in \mathcal{F}$  について  $A_i \cap A_j = \emptyset$  ( $i \neq j$ ) ならば,  
$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

$\emptyset$ : (ヌル)  
空集合

$Pr(A)$  と  
書くことも

**確率の三公理**

# 公理から導かれる確率の基本的性質

1.  $P(\emptyset) = 0$

2.  $P(\bar{A}) = 1 - P(A)$

$\bar{A}$ は事象 $A$ が起こらない事象(余事象)

3. 事象 $A, B \in \mathcal{F}$ に対して

$$A \subset B \text{ ならば } P(A) \leq P(B)$$

4. 任意の事象 $A \in \mathcal{F}$ に対して $P(A) \leq 1$

5. 事象 $A, B \in \mathcal{F}$ に対して

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

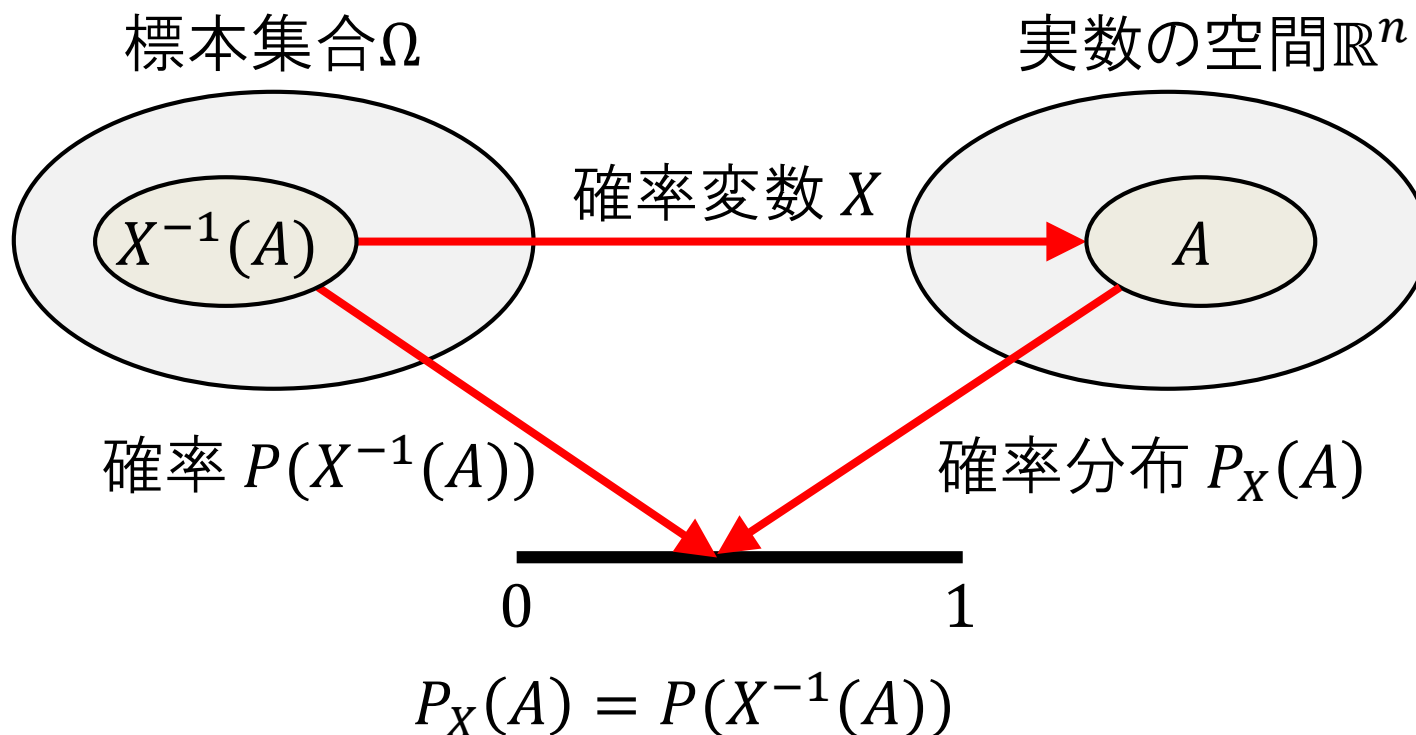
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

確率の加法定理

# 確率変数と確率分布

確率変数  $X$  とは、標本集合  $\Omega$  から実数空間  $\mathbb{R}(\mathbb{R}^n)$  への (確率と関連付いた) 写像である

確率変数  $X$  がある値になる確率を **確率分布** という





# なんでこんなに面倒な議論をするの？

A. 抽象的な概念を抽出して議論することで、  
様々なケースに適用できるから

離散的な集合だけでなく、連続的な集合の上でも  
同じようにして確率を定義できる

確率的な性質さえ満たすなら、何でもいい

→ 主観的な確信の度合いとして確率を用いても  
よかろうもん(ベイズ主義)

直接に統計情報を取れない  
事象を推測するのに用いる