参考資料: 対数関数、確率

工学部 情報エレクトロニクス学科

情報科学研究院 情報理工学部門
大規模知識処理研究室

堀山 貴史

対数計算について

- 情報理論（特に前半）では、
  対数関数(log)の計算が頻繁に出てきます。
- 教科書の付録Aで、対数の基本事項について
  まとめてあります。
  - 高校の数学でほとんど習っているはずですが、
    忘れている人も多いと思うので、慣れておいてください。
  - 定理には証明がついていますが、数学の講義ではないので、証明を暗記する必要はありません。
  - 定理を使って、対数を含む式の計算ができるようになっていれば十分です。
  - たとえば，定理A.4は便利なので，グラフを覚えるのがオススメです．

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．

つまり、a を
何乗したら x になるか

定義A.2:

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

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

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

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

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

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

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

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

※ Kenneth E. Iverson が1962年に導入

<u>例題</u>: 10進数で 225 は，二進数で何桁か？

$$\log_2 225 = \log_2 5^2 \cdot 3^2 = 2 \log_2 5 + 2 \log_2 3$$
$$\fallingdotseq 2(2.322 + 1.585) = 7.814$$

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

11100001

$y = \log_a x$ 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）

$x \to +\infty \Rightarrow \log_a x \to +\infty$

$x < y \iff \log_a x < \log_a y$

$0 < x < 1$ の間は負の値をとる

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

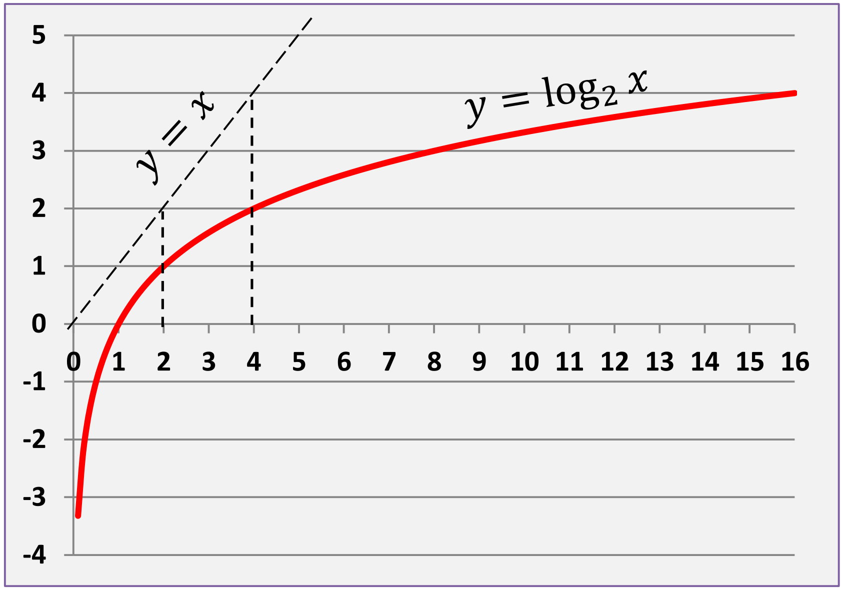


図 $y = \log_2 x$ のグラフ

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

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

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） $\ln x \leq x - 1$

Try 問3，問4

確率について

- 情報理論は確率を使って作られた理論です
（確率を避けて通ることはできません）
- 本講義で扱う確率は，高校で習う程度の知識で大丈夫
（本格的に情報理論に取り組むには，もう少し進んだ知識も）

- ここで復習しておきます
- <u>確率変数</u>：複数の値のうち，どの値になるか確率的に決定
  – 通常 X, Y などの大文字記号で表される
- <u>確率，確率分布</u>：確率変数のとるそれぞれの値が，どの程度の割合で実現するかを表す
  – 通常 $P(X = x), p_X(x)$ などの表現が使われる

確率について

- 同時確率分布（結合確率分布）：複数の確率変数があるとき，値の組合せそれぞれがどの程度の割合で実現するかを表す
  - 通常 $P(X=x,Y=y), p_{X,Y}(x,y)$ などの表現が使われる
- 周辺確率分布：複数の確率変数があるとき，その一部だけの確率分布
  - 通常 $P(X=x), p_X(x)$ などの表現が使われる
- 条件付き確率分布：複数の確率変数があるとき，一部が与えられたときの他の変数の確率分布
  - 通常 $P(Y=y|X=x), p_{Y|X}(y|x)$ などの表現が使われる
- 同時確率分布，周辺確率分布，条件付き確率分布の関係

$$p_X(x)=\sum_y p_{X,Y}(x,y),\ \ p_{X,Y}(x,y)=p_{Y|X}(y|x)p_X(x)$$

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

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

r ← 事象Aが起こる場合の数

n ← 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

まあ難しい話はおいといて･･･

「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 $E(X) = \sum_i x_i P_X(X = x_i)$

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

Try 問1

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,

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

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)

条件付き確率と独立性

$P(A) > 0$のとき,

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

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

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

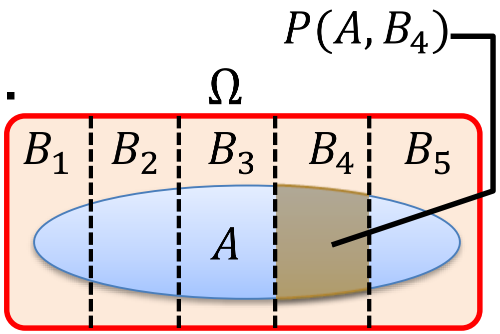
ベイズの定理

全事象 Ω 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,

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

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

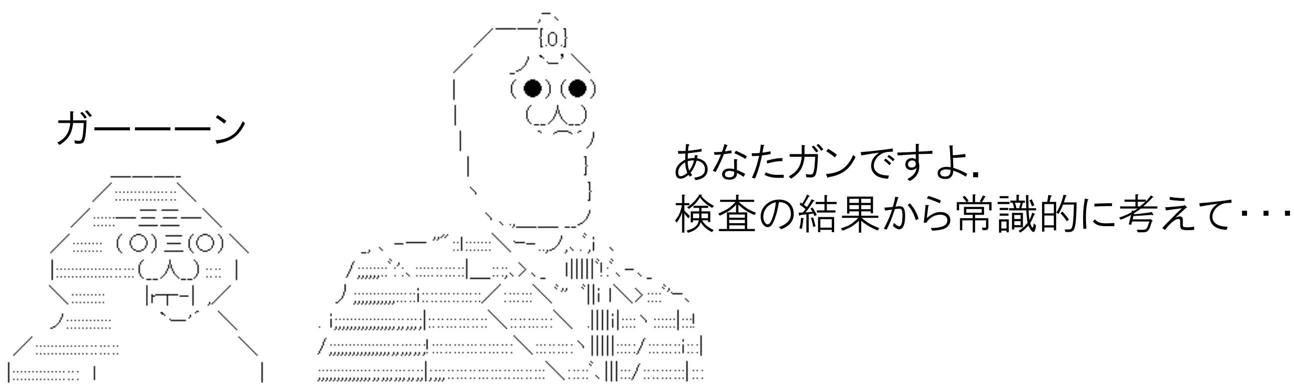
ベイズの定理を使う例題

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･･･



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,

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

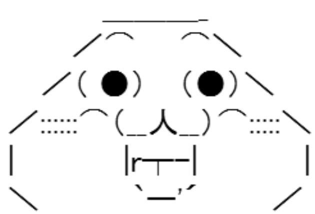
である. これに,

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

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

を代入すると,

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



たったの16%！
よかったお