2000 後期 数理の世界　第１３回

思考の道具としての数学[1] 辻下　徹[2]

理学研究科数学専攻

目 次

8 確率の基礎概念と相互情報量

8.6 確率変数は新しい確率空間を決める

$X : \Omega{\to}V$ が確率変数のとき、新しい確率空間 (V, p_X) が

$$p_X(v) := p(X = v)$$

により定義される。つまり、$p_X(v)$ は「確率変数 X の値が v となる事象」の確率である。べつの書き方としては $p_X(v) = p(X^{-1}(v))$, $p_X(v) = \sum_{X(\omega)=v} p(\omega)$ などがある。

例　$\Omega = \{ 1, 2, 3, \cdots, 10 \}, p(i) = 0.1,\ V = \{ a, b, c, d \}$,

$$X(1) = X(2) = X(3) = X(4) = a,\ X(5) = X(6) = X(7) = b,\ X(8) = X(9) = c,\ X(10) = d$$

とすると、

$$p_X(a) = 0.4,\ p_X(b) = 0.3,\ p_X(c) = 0.2,\ p_X(d) = 0.1.$$

[1] URL:http://fcs.math.sci.hokudai.ac.jp/doc/announce/sw00.html
質問提出アドレス:sw00@fcs.math.sci.hokudai.ac.jp

[2] 研究室：理学部４号館４０３号室、011-706-3823
連絡は電子メールで：Email:tujisita@math.sci.hokudai.ac.jp,
Homepage:http://fcs.math.sci.hokudai.ac.jp/tjst/

例：頻度が定める確率空間 R を記録の集まりとし、その上の等確率の確率を与える：$p(r) = \frac{1}{n}$ (n は記録の個数)。$X(r)$ を記録 r の「値」とするとき、$p_X(\omega)$ は ω の出現頻度となる。

8.7 確率変数のエントロピー

確率変数 $X : \Omega \to V$ に対し、確率分布 p_X のエントロピーを確率変数 X のエントロピーといい $H(X)$ と書く：

$$H(X) := H(p_X) = -\sum_{v \in V} p(X = v) \log p(X = v).$$

8.8 条件付き確率（続き）

$X : \Omega \to V,\, Y : \Omega \to W \;\; v \in V,\, w \in W$ に対して、

$$p(v, w) := p(X = v \text{ かつ } Y = w),$$

$$p(v|w) := \frac{p(v, w)}{p(w)},\; p(w|v) := \frac{p(v, w)}{p(v)}.$$

[**問 34**] $\Omega = \{\ 1, 2, 3, \cdots, 10\ \}, p(i) = 0.1,\ V = \{\ a, b, c, d\ \}$,$W = \{$ あ, い, う, え $\}$

$$X(1) = X(2) = X(3) = X(4) = a,\ X(5) = X(6) = X(7) = b,\ X(8) = X(9) = c,\ X(10) = d$$

$X(1) = X(5) = X(8) = X(10) =$ あ, $X(2) = X(6) = X(9) =$ い, $X(3) = X(7) =$ う, $X(4) =$ え.

$p(v, w), p(v), p(w)$ はどうなるか。

基本的性質

1. $p(v|w) = 1 \Leftrightarrow Y = w$ ならば $X = v$. (当然 $u \neq v$ ならば $p(u|w) = 0$)
2. $p(v, w) = p(v)p(w) \Leftrightarrow X, Y$ は独立

[**問 35**] $p(v|w) = 1$ だが $p(w|v)$ がとても小さいということはどういうことか。

[**問 36**]

8.9 条件付きエントロピー

確率空間 (Ω, p) とその上の確率変数 $X : \Omega \to V$ を考えよう。$p_X(X = v) \neq 0$ のとき条件付き確率 $p_X(\bullet|v)$ が定義される。そのエントロピーのことを $H(X|v)$ と書き、これを v についての平均したものを、$H(p|X)$ と書く。このとき、次が簡単に示せる（証明してみよう）。

$$H(p|X) := \sum_{v \in V} H(X|v) p_X(v) = H(p) - H(X)$$

すなわち、$H(p|X)$ は、p のあいまいさの中で、観測 X によっては解消できない部分の大きさを表す。

8.10 相互情報量

確率空間 (Ω, p) 上の確率変数を 2 つ考える:$X : \Omega \to V, Y : \Omega \to W$ とすると「積確率変数」$X \times Y : \Omega \to V \times W$ が派生する：

$$(X \times Y)(\omega) := (X(\omega), Y(\omega)), \quad \omega \in \Omega.$$

今まで出てきたものとして次のエントロピーが定義される：

$$\begin{aligned} H(X) &:= H(p_X) \\ H(Y) &:= H(p_Y) \\ H(X,Y) &:= H(p_{X \times Y}) \\ H(X|Y) &:= H(X,Y) - H(Y) \\ H(Y|X) &:= H(X,Y) - H(X) \end{aligned}$$

この他に、

$$I(X,Y) := H(X) - H(X|Y)$$

が定義される。これは、X のあいまいさ $H(X)$ の中で、Y によって解消されない部分 $H(X|Y)$ を引いたもの。別の言い方をすると、X のあいまいさの中で Y によって解消できる部分を表す。つまり、Y がわかったときに X についてどれだけのことがわかるか、を表す。これを X と Y の相互情報量という。

相互情報量の基本的性質

定理 8.1 $I(X,Y) = I(Y,X)$.

定義より次がわかる

1. $I(X,Y) = H(X) + H(Y) - H(X,Y)$.
2. $H(X,Y) = I(X,Y) + H(X|Y) + H(Y|X)$.

これは、次のように図示するとわかりいやすい。

8.11 情報の流れ

脳の中を「情報が流れる」という言い方をすることがある。相互情報量を使って意味を与える。

A, B を脳内の 2 点とし、A,B を同時測定する。測定値は $\{ 0, 1 \}$ とする。十分長く（10000 sec）測定し、長さ 100 の $\{ 00, 01, 10, 11 \}$-列の上の頻度統計を作る（ここで、01 は A で 0,B で 1 という結果を表す。）a_i を時刻 i での A での値を表す変数、b_i を時刻 i での B での値を表す変数とする。このとき、$I(A, i|B, j) := I(a_i, b_j)$ は時刻 1 での A の状態と時刻 j での B の状態との間の共通情報量を表す。

$I(A, i \to B, j) := I(A, i, B, j)/H(a_j)$ と置くと、これは時刻 i での A-状態が時刻 j での B 状態がどのくらい「決めるか」、その度合いを表す。

$I(A, i \to B, j) = 1$ は、時刻 i での A-状態が時刻 j での B 状態を決めることがわかる。このとき $A, i \to B, j$ と書く。

$A, i \to B, j, \ j > i$ のとき、脳を使って情報を流すことができる。つまり、A に信号を入れると時刻 $j - i$ 遅れで B の所に信号が取りだせる。

$A, i \to B, j, \ j < i$ のときは、A での信号から B の時刻 $i - j$ 前の信号を取りだせる（推定できる）。

極端な場合には $I(A, i \to B, j) = I(B, j \to A, i) = 1$ ということも起こる。この場合には、A, i と B, j は interlocked しているという。A と B とは時間のずれがあるが直結しているかのように見え、$i < j$ ならば、A での信号が $j - i$ かかって B に到達した、と普通は言う状況になっている。ただ、A, B が離れていて、途中の部位については B のようなものがない場合も有りうるので、その場合に「情報が伝わっていった」という言い方は余り適切ではない。

9 順序集合

9.1 整除関係

自然数の整除関係 $x|y \overset{def}{\iff} y = x \times a$ となる自然数 a が存在する。は次の性質を満たす：

- $x|x,$
- $x|y,\ y|z$ ならば $x|z.$
- $x|y$ かつ $y|x$ ならば $x = y.$

同様に、$\{\,1, 2, 3, 4\,\}$ の部分集合の包含関係について

- $x \subseteq x,$
- $x \subseteq y,\ y \subseteq z$ ならば $x \subseteq z.$
- $x \subseteq y$ かつ $y \subseteq x$ ならば $x = y.$

が成り立つ。

このような２項関係のことを順序関係と呼ぶ。

9.2 順序関係の定義

集合 X 上の順序関係とは、２項関係 $x \preceq y$ で次の３条件を満たすもの：

(反射律) $x \preceq x.$

(推移律) $x \preceq y$ かつ $y \preceq z$ ならば $x \preceq z.$

(反対称律) $x \preceq y$ かつ $y \preceq x$ ならば $x = y.$

順序関係を与えた集合のことを順序集合 (partially ordered set 略して **poset**) という。

$a \preceq b$ または $b \preceq a$ のとき、a, b は**比較可能である**という。そうでないとき a, b は**比較できない**という。どの２元も比較可能であるとき、$\preceq$ は全順序 (total order, linear order) であるという。

a が b の**直前の元** であるとは $a \neq b$ であって $a \preceq c \preceq b$ かつ $a \neq c, b \neq c$ であるような c がないことをいう。このとき $a \lessdot b$ とかく。

有向グラフ $(V, \lessdot)$ を平面上に描くとき $a \lessdot b$ ならば a を b の下方に書くようにしたものを、順序関係 $(V, \preceq)$ の **Hasse 図**という。

[**問 37**] $V = \{\,1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\,\}$ として $a \preceq b \overset{def}{\iff} a|b$ と定義して得られる順序集合の Hasse 図を描け。

[**問 38**] $\{\,1, 2, 3\,\}$ の部分集合に包含関係を入れてできる順序集合の Hasse 図を描け。

9.3 最大と極大

$(X, \preceq)$ を順序集合とする。

どの元よりも大きいか等しい $m \in X$ を X **の最大元**(the greatest element) という。最大元は存在するとは限らない。

それより本当に大きな元がないような $m \in X$ を X **の 極大元**(a maximal element) という。極大元も存在するとは限らない。

補題 9.1 *1.* 2 個以上の最大元は存在しない。

2. 最大元は極大である.

双対概念として、最小元・極小元が定義される。

最大元を $\top$(トップと呼ぶ)、最小元を $\perp$(ボトムと呼ぶ) と書く。(各々を $1, 0$ と書く文献も多い。) 最大元はどの元とも比較可能である。

注意 (i). 最大元が存在しない理由は２通りある。

- 極大元が２つある。
- 極大元すら存在しない。

(ii). 極大元が存在しないとするとその理由は、

- いくらでも大きな元がある。$\forall x \exists y [x \preceq y \wedge x \neq y]$.

説明 社会に支配従属関係（話しを単純化して考えます）を入れて順序集合と考えます。極大元は、誰からも支配されない者を意味し、最大元は、すべての者を支配する者を意味します。一匹狼（誰からも支配されないが自分以外の誰も支配しない者）は極大元（かつ極小元）です。現代社会は他から支配されない者は沢山居るが、社会全体を支配する者（最大元）は居ません。

同様に、極小元は自分以外の誰も支配しない者を意味し、最小元は、すべての者に支配される者を意味します。

9.4 上限・上界

定義 $A \subseteq X$ とする。

上界 (upper bound) A のどの元よりも大きな u を A の上界という。

上限 (supremum) A の上界の中で最小のものがあるとき、それを A の上限といい $\bigvee A$ と書く。$\bigvee\{\, a, b \,\}$ は $a \vee b$ と書く。

$\bigvee A$ は次の性質により特徴付けられる：

(*) X の任意の元 x について $\boxed{\bigvee A \preceq x \iff \forall a \in A[a \preceq x].}$

注意 (i). A に上界が存在しないということは、

$$\forall x \in X \exists a \in A[a \preceq x \text{ でない}].$$

$a \preceq x$ でないというのは２通りの場合がある

- a, x は比較できない,
- $x \prec a$.

(ii). A に上限が存在しないということは、

- 上界が存在しない
- 上界はあるが、上界の中に極小なものが２つ以上ある。
- 上界はあるが、上界の中に極小なものがない。

下界と下限　上界と上限の双対概念が下界と下限である。下限は $\bigwedge A,\ a \wedge b$ と書く。

例：整除順序　自然数の集合に

$$n \preceq m \overset{def}{\iff} n|m(n \text{ は } m \text{ を割り切る})$$

という順序を入れる。このとき

- $n \vee m$ は n, m の最小公倍数.
- $n \wedge m$ は n, m の最大公約数.

例：べき集合　$(\mathbf{pow}\ (X), \subseteq)$ では

$$A \bigwedge B = A \bigcup B \qquad A \bigwedge B = A \bigcap B.$$

A　レポート問題割当

- **単位を必要な人は必ず提出して下さい。**
- 提出期限：２月１４日
- Ａ４版のレポート用紙を使用して下さい。
- 提出先：レポート提出用ボックス

問４と７ p1–7 宣教師と首刈族パズル（３人ずつの場合、４人ずつの場合）：組頭、石川

問５ p1–7 やきもちやきの３亭主：大貫

問６ p1–7 水差しパズルのバリエーション：石本、柴波

問１９ p5–1 sliding game ：多田

問２４ p6–2 １山石取りゲームのルールを「１個、３個、４個とれる」としたとき、先手必勝であるための条件を求めよ。：土子

問１１−１ p11–1 贋銅貨のパズルを、全体の数が１２の場合に解け。：木下

B 第 12 回質疑

[**Q12-1**](文学部) 確率でマイナスということはあるのですか。

[**A12-1**] ありません。

[**Q12-2**](法学部)1 より大きい確率は絶対にないのでしょうか。

(質問理由：よく絶対であることを強調するとき、120 % 保障するなどと言うが、それは言葉の綾なのかどうか知りたい。また、もし存在するとしたら、それを想像することができない。)

[**A12-2**] ありません。

[**Q12-3**](法学部) 周辺分布の値は、そのまま $P(x|y), P(y|x)$ の値に当てはめる事ができますか。

(質問理由：講義の例では、その数が一致していたから。)

[**A12-3**] $P(x) = P(x|y)$ ということは、y であることがわかったとしても、x となる確率に影響しない、ということです。このとき、x, y は独立であると言います。従って、x, y が独立ではない場合には、できません。

[**Q12-4**](文学部) 確率変数が何を意味するものなのかがよくわかりません。

(質問理由：確率変数の定義がちょっと理解できなかったので。また、何のために求めるのかもよくわからなかった。)

[**A12-4**] 「変数」(変量,varible)は、何かに応じて変化する量を表す言葉ですが、確率変数は、試行ごとに変化する量です。観測値は、観測(試行)ごとに変化い得るものですから変数ですが、変化の仕方が確率的にしかわからない場合に確率変数と言います。

[**Q12-5**](法学部)$P(x|y)$ といったような記号がよくわかりません。

(質問理由：y の中での x の割合というとこまではわかるのですが、$P(x, y)$ と比較してどうちがうかがわかりません。)

[**A12-5**] $P(x, y)$ は、全体の中で $X = x$ かつ $Y = y$ となる割合ですが、$P(x|y)$ は、$Y = y$ という場合の中で、$X = x$ かつ $Y = y$ (つまり $X = x$)となる割合です

[**Q12-6**](法学部) 有限確率があるのなら無限確率もあるのですか。

(質問理由：もしもあるのならどういうものか知りたいから。)

[**A12-6**] 現実にはたいてい無限確率です。例えば温度が T 度となる確率は、といった問題を考えるときは、Ω は温度の集まりですが、これは無限集合です。

[**Q12-7**](文学部) 今日やった考え方は、社会学をやる上で必要となるものでしょうか？

(質問理由：将来、社会学を研究したいと考えているから。)

[**A12-7**] 社会学で人々の色々な性格などを分析するときには基本的な役割を果たすものと思います。喫煙率はどれだけか、という調査は多分しないと思います。性別・年齢・職業などに喫煙率がどのように依存するかを考えると思います。その場合には P(喫煙者 | 男・２０代・会社員）や、逆に P(男 | 喫煙者) などにも関心を持つと思います。

2001.1.24

２０００年度後期　数理の世界 第１３回
学部：　　　　　学科：
学籍番号：　　　　　　　氏名：
質問：
質問理由：

アンケート

次の表にチェックしてください.

1： 良く知っている　2：知らなかった 3：よくわかった　4：大体わかった　5：まだよくわからない

用語	1	2	3	4	5
条件付き確率の定義	–	–			
確率変数	–	–			
相互情報量	–	–			
順序関係	–	–			
最大限と極大元の違い	–	–			

意見・希望などがあればどうぞ：

北大理数学 辻下　徹