# 情報理論と情報エントロピー

## A. 確率論の基礎知識

### 1. 確率の定義

**「ある出来事が起きる確率（ある事象の確率）」には2つの見方がある**

#### 1-1. 頻度による定義

互いに相容れない事象が $N$ 個ある．$j$ 番目の事象を $x_j$ と呼ぼう（$j=1,\ldots,N$）．

たとえば，$x_1$:「晴れる」，$x_2$:「曇る」，$x_3$:「雨が降る」，$x_4$:「雪が降る」，$N=4$

$x_j$ が $n_j$ 回起きるとする，$(x_1,x_2,\ldots,x_N)$, $(n_1,n_2,\ldots,n_N)$

12 月 12 日の 12 時の小金井の天気の 12 年間の記録は，晴れ 6 回，曇り 3 回，雨 2 回，雪 1 回

定義： 事象 $x_j$ の確率は

$$P\left(x_j\right)=\frac{n_j}{\Sigma_{j=1,N}n_j}$$

この定義から直ちに $\Sigma_{j=1,N}P(x_j)=1$

#### 1-2. 先験的な確率

観測者がある事象の起きる確からしさについての数量（確率）に信念を持っている場合．人により異なる場合もあるし，同じ人でも人生を通して変わることもある．

#### 1-3. 現実によくあるのは，それらの中間

1-1 と 1-2 を関連づけることができる場合もある．1-2 は 1-1 を理想化した場合が多い．

### 2. 確率の基本的な性格

#### 2-1. 基本

$a,b$ を 2 つの異なる事象として

$$P(\overline{a})=1-P(a),\quad P(a\vee b)=P(a)+P(b)-P(a\wedge b)$$
$$P(a,b)\equiv P(a\wedge b)$$
$$P(a,b)=0 \Leftrightarrow a\text{と}b\text{は素}$$

$P$(晴れ) = 1－$P$(晴れではない),

$P$(「曇りか雨」または「雨か雪」)

=$P$(曇りか雨)+$P$(雨か雪)－$P$(雨)

$P$(雪，晴れ) = 0

#### 2-2. 条件付き確率

$P(a|b)$：すでに $b$ が起きたことを前提にして，$a$ の起きる確率

$$P(a,b)=P(a|\mathrm{b})P(b)=P(b,a)=P(b\,|\,a)P(a),$$

$$P(a\,|\,b)=\frac{P(b\,|\,a)P(a)}{P(b)}$$

#### 2-3. *a* と *b* が独立

$b$ が起きようが起きまいが $a$ が起きる確率は変わらない

$P(a|b)=P(a)$

このとき，この関係と $P(a\,|\,b)=\dfrac{P(b\,|\,a)P(a)}{P(b)}$ から

$$P(b\,|\,a)=P(b)$$

さらに

$$P(a,b)=P(a)P(b)$$

がなりたつ．

## B. 情報理論の基礎知識

### 1. 情報の数量化に必要なこと

#### 1-1 情報量

情報量

＝ その事象が起きたことを知ることによって得た量

「確率 $p$ で起きる事象が実際に起きたことを観測した」という情報を数量化して扱う．その情報の意義については論じない．単に起きたか否かだけを問題にする：情報はシンボルで表わせ、確率 $p$ を用い数量化できるとしその量を $I(p)$ と書く．

### 1-2 情報量が持つべき性質

・$I(p) \ge 0$

・$I(1)=0$：必ず起きることが確定しているものが起きたと聞いても情報は増えない．「太陽が東から昇った」

・$I(p_1 \times p_2) = I(p_1) + I(p_2)$：互いに独立な事象が同時に起きたことを知れば、得られた情報量は加算される．「今朝東京で雨が降った」「昨日のNYで停電があった」

・$I(p)$ は単調で連続な関数

⇒

$$I(p) = -\log_b p,\ \ b>1,\ 1 \ge p \ge 0$$

単位は $b$ の取り方による：

$b$=2 (bits と呼ぶ), 3(trits), e(nats), 10(Hartleys)

問　硬貨の表(h)と裏(t)が出る確率が等しく 1/2．1 回投げて（1 回の試行で）h が出たことを知る．その情報量を bit 単位で求めよ。$n$ 回の試行で、全部 h だったということの情報量は？

解：$I(1/2) = -\log_2\left(\frac{1}{2}\right) = 1,\ I((1/2)^n) = -\log_2\left(\frac{1}{2^n}\right) = n$

### 2. 情報エントロピー

情報源（対象を観測、次々に起きる事象をシンボルの列として発信）からシンボルの列 $\{a_1, a_2, ..., a_n\}$ が送られる（「事象 $a_1$ が起きた」「$a_2$ が起きた」…送られる順序は問題にしない）．各シンボル（事象）が起きる確率が $p_j$．$\{a_1, a_2, ..., a_n\}$ の各シンボルの現れ方は規則性がない（$j$ 番目にどのシンボルが現れるかは、履歴によらず独立に決まる）．このような仮定のもとで、平均として、シンボル 1 個あたりどれだけの情報量があるだろうか．

### 2-1　ある情報源から来る情報量の期待値 $\langle I \rangle$

$$\langle I \rangle = \sum_{j=1,n} p_j I(p_j) = \sum_{j=1,n} p_j \left(-\log_2 p_j\right) = \sum_{j=1,n} p_j \log_2(1/p_j)$$

問　確率が 0 の事象も上の計算に含めたい．$p \times \log(1/p)$ は $p \to 0$ の極限でどんな数に収束するか？

解：$\lim_{x\to\infty} \frac{\log x}{x} = \lim_{x\to\infty} \frac{\frac{d}{dx}\log x}{\frac{d}{dx}x} = \lim_{x\to\infty} \frac{1/x}{1} = 0$（ロピタルの法則を用いたが，本質的には「対数の発散速度はどのような「べき」より遅いので極限値は 0」ということである）．

**情報エントロピーの定義**：

$$H = \sum_{j=1,n} p_j \log(1/p_j) \to \int_{-\infty}^{\infty} p(x)\log(1/p(x))dx$$

問　$H$ を最大にする $\{p_j\}$ は？そのときの $H$ の値は？

解　ラグランジュの未定乗数法．$\sum p_j = 1$ のもとで

$H = \sum_{j=1,n} p_j \log(1/p_j)$ が極値となる $\{p_j\}$ は？

$$G \equiv H - \lambda\left(\Sigma p_j - 1\right)、\ \frac{\partial G}{\partial p_j} = \frac{\log(1/p_j) - 1}{\log 2} - \lambda = 0$$

より，どの $p_j$ も同じ値となるとき極値となる $p_j = 1/n$　、このとき

$$H = \sum_{j=1,n}\left(\tfrac{1}{n}\right)\log(n) = \log n$$

問　ある1つのことだけが起きるとき、$H$ の値は？

解　$H = 1\times\log 1 + 0\times\log 0 + 0\times\log 0 + \cdots = 0$

問　起こりうる $n$ 個の異なる事象が全くランダムに起きるので 1 度の試行について何が起きるか全く予想できないとき、情報エントロピーは？

解　最大の値、log n である．

問　文房具屋に鉛筆($E_1$=¥10)、ボールペン($E_2$=¥100)、万年筆($E_3$=¥1000)が各 1 種類ずつ在庫している。店に来る1人の客が鉛筆を買う確率を $p_1$ とし，ボールペンは $p_2$，万年筆は $p_3$ とする（1本しか買わない）．これまでの調査では1人が支払う金額が平均 $E$=¥200 であることが分かっているが，それ以外については全く不明である．三種類の在庫量の比を最適化するにはどうすればよいか？　ヒント：「顧客が何を買うかについて無知である」ことを「情報エントロピーが最大になる」として表す（最大エントロピー法）．ただし，支払い平均額は既知であるから，その条件下の最大化である．

解：条件 $\sum p_j = 1,\qquad \sum E_j p_j = E$　のもとで

$H = \sum_{j=1,3} p_j \log(1/p_j)$ が極値となる $\{p_j\}$ を求め，この確率に比例した数量を在庫すればよさそうである．

$$G = \sum p_j \log(1/p_j) - \alpha\left(\sum p_j - 1\right) - \beta\left(\sum E_j p_j - E\right)$$

$$\frac{\partial G}{\partial p_j} = \frac{\log(1/p_j) - 1}{\log 2} - \alpha - \beta E_j = 0 \to p_j = A \times 2^{-\beta E_j}$$

$$\sum p_j = \sum A \times 2^{-\beta E_j} = 1 \to 1/A = \sum 2^{-\beta E_j} \qquad \boxed{p_j = \frac{2^{-\beta E_j}}{\sum 2^{-\beta E_j}}}$$

$\sum E_j p_j = \dfrac{\sum E_j 2^{-\beta E_j}}{\sum 2^{-\beta E_j}} = E$ からβ が決まる．