情報理論14

藤田 一寿

ver.2020

連続信号のエントロピー

連続信号の取扱の必要性

- これまでの資料では，信号は離散的な値を取るとしてきた．
- しかし，世にあるデータの殆どは連続的な値をとる．
- そこで，連続値をとる信号のエントロピーを考える必要がある．

連続信号のエントロピー

- 離散信号のエントロピーと同じように考えると，連続信号のエントロピーは次のようになると想像できる．
- x が生じる確率をp(x)とすると

$$\int p(x)dx = 1$$

- である．離散信号のエントロピーから類推した，連続信号のエントロピーは

$$H = -\int p(x) \log p(x) dx$$

離散信号のエントロピーの極限

- 離散信号のエントロピーの極限として連続信号のエントロピーを求める.
- 実数xをΔxで刻む. Xiと隣り合うxi+1との関係は

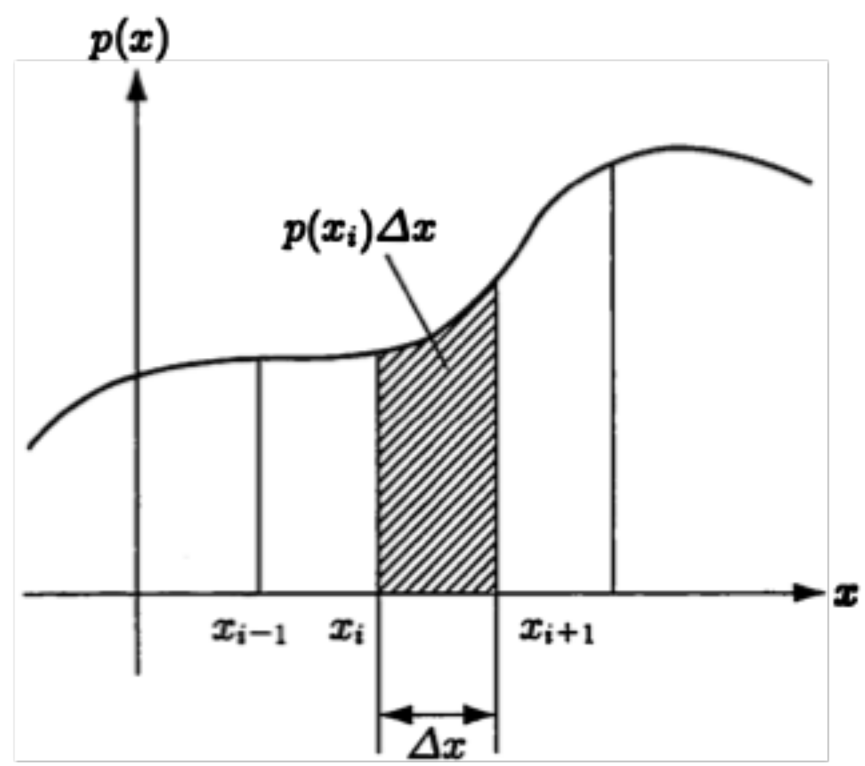
$$x_i = i\Delta x,\ \ x_{i+1} = (i+1)\Delta x$$

- となる.
- XがXiとxi+1との間にある確率p(xi)は

$$p(x_i)\Delta x$$

- よって, エントロピーは

$$H = -\sum_{i=-\infty}^{\infty} p(x_i)\Delta x \log\{p(x_i)\Delta x\}$$



離散信号のエントロピーの極限

$$H = -\sum_{i=-\infty}^{\infty} p(x_i)\Delta x \log\{p(x_i)\Delta x\}$$
$$= -\sum p(x_i)\Delta x \log p(x_i) - \sum p(x_i)\Delta x \log \Delta x$$

- ここで，Δx→0の極限をとる．
- 第1項は明らかに

$$-\int p(x)\log p(x)dx$$

- 第2項は

$$\sum p(x_i)\Delta x \to \int p(x)dx = 1$$

であるから，

$$-\log \Delta x$$

Δx→0のとき発散

よって

$$\lim_{\Delta x \to 0} H = -\int p(x)\log p(x)dx - \lim_{\Delta x \to 0} \log \Delta x \to \infty$$

離散信号のエントロピーの極限

$$\lim_{\Delta x \to 0} H = -\int p(x)\log p(x)dx - \lim_{\Delta x \to 0} \log \Delta x \to \infty$$

- 式の意味を考えると
- 第1項は，確率分布に関係した不確定度を表す．
- 第2項は，無限個の実数から一つの数を選び出すため生じる不確定度を表す．無限個の中から一つ選ぶため不確定度は無限大となってしまう．

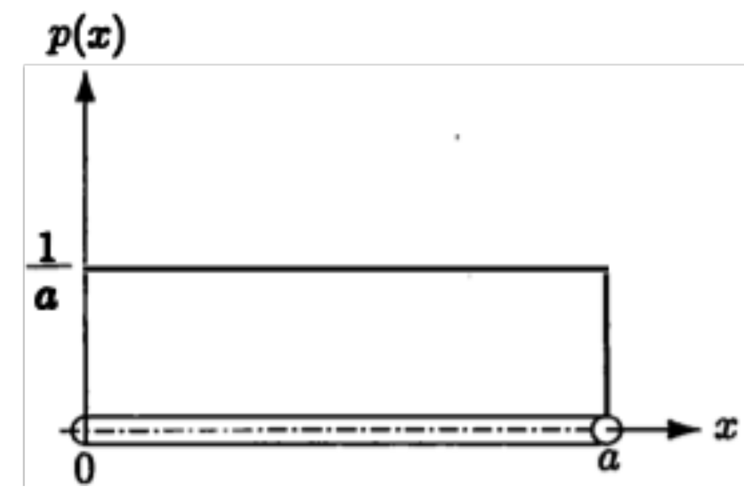
- 発散項をのぞいて，第1項のみでエントロピーを定義すると

$$H = -\int p(x)\log p(x)dx$$

連続信号のエントロピーの性質

- エントロピーは正の値も負の値も取りうり，エントロピーの値それ自体に意味はない．
  - 長さa [cm]の棒に，場所xに傷があるとする．傷がどこにできるかどうかの確率はそれぞれの場所で変わらないとすると，x [cm]に傷がある確率は

$$p(x) = \frac{1}{a},\ \ 0 \leqq x \leqq a$$



  - よって，エントロピーは

$$H = -\int_0^a \frac{1}{a} \log \frac{1}{a} dx = \log a$$

  - このエントロピーはaの値で変わってしまう．
  - a = 1のときH = 0，a < 1のときH < 0となる．だからといって，a=1のとき不確定度がまったくないわけではない． a < 1のときエントロピーは負になるが，負の値だからといって確実だと言うわけでもない．

連続信号のエントロピーの性質

- 値の単位を変えるとエントロピーも変わってしまう．
  - 先程の例では単位はcmで考えたが，mmに変えると次のようになる．
  - 場所x [mm]に傷がある確率は，

$$p(x) = \frac{1}{10a}, \ 0 \leqq x \leqq 10a$$

  - よってエントロピーは

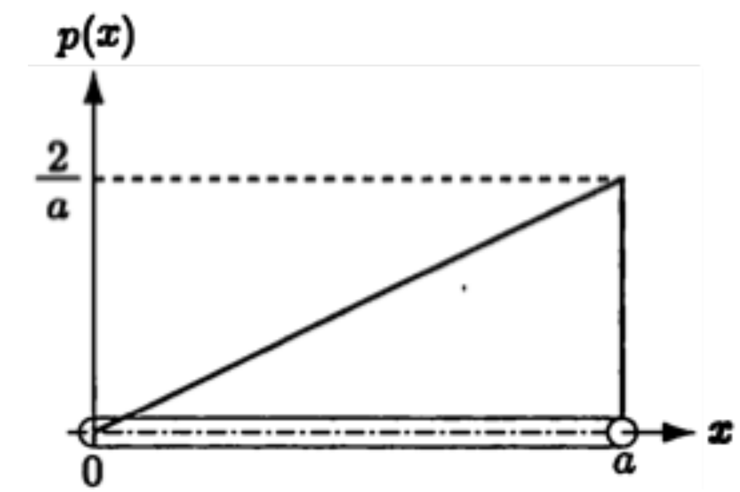
$$H = \log 10a$$

  - 単位を変えただけで，エントロピーが増加してしまう．

連続信号のエントロピーの性質

- **エントロピーの差で定義される情報量に意味がある．**
  - 図のように，各場所の傷のある確率が線形に増加する場合を考える．このときの確率密度関数は

$$p(x) = \frac{2}{a^2}x, \quad 0 \leqq x \leqq a$$



  - よって，エントロピーは

$$H' = -\int_0^a \frac{2}{a^2}x \log\left(\frac{2}{a^2}x\right) dx$$
$$= \log a - \log \frac{2}{\sqrt{e}}$$

  - 一様分布の場合とエントロピーを比較すると

$$I = H - H' = \log \frac{2}{\sqrt{e}} \fallingdotseq 0.3 \text{ ビット}$$

  - この値は，単位には依存しない．

n次元信号のエントロピー

- 信号がn次元ベクトル $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ で表されるとする．
- 信号xが現れる確率を $p(x_1, x_2, \cdots, x_n)$ とすると，エントロピーは

$$H = -\int p(x_1, x_2, \cdots, x_n) \log p(x_1, \cdots, x_n) dx_1 dx_2 \cdots dx_n$$

ある領域上でのエントロピー最大の信号分布

- 確率密度関数p(x)は $\int p(\boldsymbol{x})d\boldsymbol{x} = 1$ を満たさねばならない．この条件のもとでエントロピー

$$H = -\int_{R} p(\boldsymbol{x}) \log p(\boldsymbol{x}) d\boldsymbol{x}$$

- が最大となる確率密度関数を求める．ラグランジュの未定乗数をλとすると，

$$F = H - \lambda \int p(\boldsymbol{x}) d\boldsymbol{x}$$

- Fの偏微分は

$$\delta F = -\int (\log p(\boldsymbol{x}) + 1 + \lambda) \delta p(\boldsymbol{x}) d\boldsymbol{x}$$

- δp(x)が何であれ，δFが０でなければならないから

$$\log p(\boldsymbol{x}) + 1 + \lambda = 0,$$

- よってp(x)は定数である．

ある領域上でのエントロピー最大の信号分布

- つまり，p(x)は一様分布である.
- 領域の体積をAとすると，p(x)は

$$\boldsymbol{p}(\boldsymbol{x}) = \begin{cases} 0, & \boldsymbol{x} \notin \boldsymbol{R} \\ \dfrac{1}{A}, & \boldsymbol{x} \in \boldsymbol{R} \end{cases}$$

- このときのエントロピーは

$$H = \log A$$

平均エネルギー一定でエントロピーを最大にする分布

- １次元信号xの確率密度関数p(x)は

$$\int p(x)dx = 1.$$

- 信号xのエネルギーはx^2だから，この信号のエネルギーの期待値は

$$\int x^2 p(x)dx = \sigma^2$$

- これらの条件が与えられたとき，エントロピーが最大となる分布を求める．未定定数をλ，μとして

$$F = H - \lambda \int p(x)dx - \mu \int x^2 p(x)dx$$

- このp(x)に関する変分δFは

$$\delta F = -\int \{\log p(x) + 1 + \lambda + \mu x^2\} \delta p(x)dx.$$

平均エネルギー一定でエントロピーを最大にする分布

- よって，エントロピーを最大とする分布は次の方程式を満たす．

$$\log p(x) + 1 + \lambda + \mu x^2 = 0$$

- この方程式を解くと分布は

$$p(x) = e^{-(1+\lambda)} e^{-\mu x^2}$$

- λとμを求めると

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}}$$

- これは，分散がσ^2のガウス分布である．このときのエントロピーは

$$H = \log(\sqrt{2\pi e}\sigma)$$

条件付きエントロピー

- ２つの信号x, yがある．x, yが生じる同時確率をp(x,y)とする．xとyを組み合わせた信号のエントロピーは

$$H(X,Y) = -\int p(\boldsymbol{x},\boldsymbol{y}) \log p(\boldsymbol{x},\boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y}$$

- 信号xの確率密度関数p(x)は，yを周辺化することで

$$p(\boldsymbol{x}) = \int p(\boldsymbol{x},\boldsymbol{y}) d\boldsymbol{y}$$

- と求まる．同様に信号yの確率密度関数p(y)は

$$q(\boldsymbol{y}) = \int p(\boldsymbol{x},\boldsymbol{y}) d\boldsymbol{x}$$

条件付きエントロピー

- 信号xが分かっているという条件のもとでのyの条件付き確率は

$$q(\boldsymbol{y}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{y})}{p(\boldsymbol{x})}$$

- Xが何であるか分かったときの，yの条件付きエントロピーは

$$H_X(Y) = -\int p(\boldsymbol{x})q(\boldsymbol{y}|\boldsymbol{x}) \log q(\boldsymbol{y}|\boldsymbol{x}) d\boldsymbol{y} d\boldsymbol{x}$$

- yが何であるか分かったときの，yの条件付きエントロピーは

$$H_Y(X) = -\int q(\boldsymbol{y})p(\boldsymbol{x}|\boldsymbol{y}) \log p(\boldsymbol{x}|\boldsymbol{y}) d\boldsymbol{x} d\boldsymbol{y}$$

条件付きエントロピーと相互情報量

- 離散的な場合と同様，次の関係式が成立する．

i） $H(X,Y) \leqq H(X)+H(Y)$,

ii） $H_X(Y) \leqq H(Y)$,

iii） $$H(X,Y) = H(X)+H_X(Y) = H(Y)+H_Y(X).$$

iv） $$I(X,Y) = H(X)+H(Y)-H(X,Y) = H(Y)-H_X(Y) = H(X)-H_Y(X)$$

相互情報量

信号の変換とエントロピー

信号変換とエントロピー

- １次元信号xを次の式で変換する.

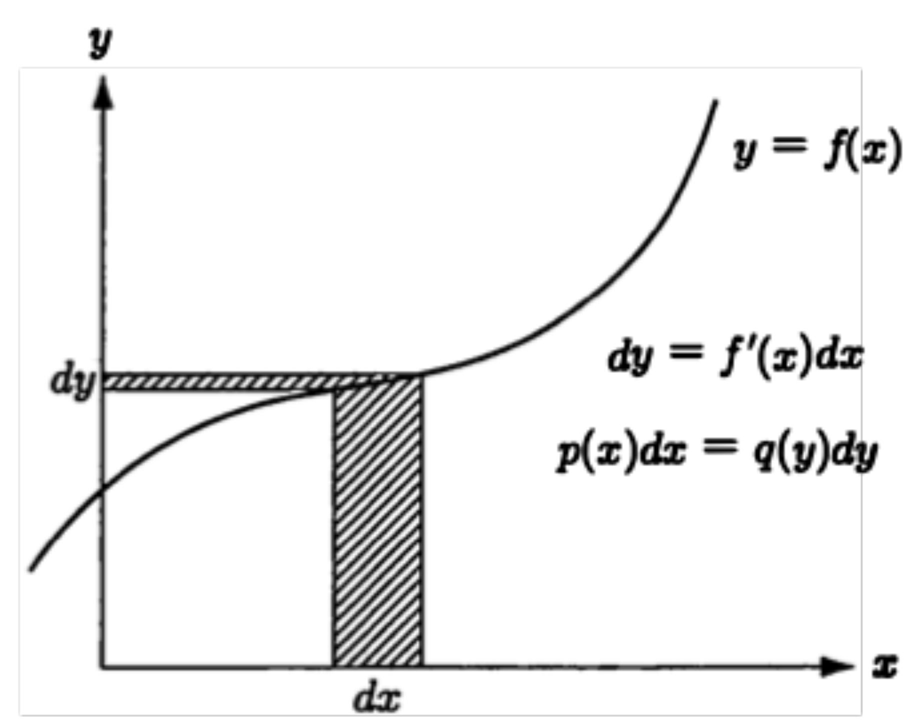
$$y = f(x)$$

- x〜x+dxの間の信号が出る確率はp(x)dxである.
- このとき信号は変換され，y 〜y+dy (f(x)〜f(x+dx))の信号となる.
  よって，このようなyが出る確率はq(y)dyだから，

$$dy = f'(x)dx$$

- また，$p(x)dx = q(y)dy$ だからこれを代入して

$$q(y) = \frac{p(x)}{f'(x)}$$



信号変換とエントロピー

- 信号yのエントロピーは

$$H(Y) = -\int q(y)\log q(y)dy$$

- 先程の式を代入すると

$$\begin{aligned} H(Y) &= -\int \frac{p(x)}{f'(x)}\left\{\log\frac{p(x)}{f'(x)}\right\} f'(x)dx \\ &= -\int p(x)\log p(x)dx + \int p(x)\log f'(x)dx \end{aligned}$$

- 第１項は信号xのエントロピーだから

$$H(Y) = H(X) + \int p(x)\log f'(x)dx$$

- これが，エントロピーの変換公式である．

例

- スケールを変えただけの

$$y = cx$$

- の場合は，f’(x)=cだから

$$H(Y) = H(X) + \log c$$

N次元信号の変換の場合

- N次元信号の場合も同様に考えれば，信号xは

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x})$$

- のように変換される．各成分について細かく書くと

$$y_1 = f_1(x_1, \cdots, x_n)$$
$$y_2 = f_2(x_1, \cdots, x_n)$$
$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots$$
$$y_n = f_n(x_1, \cdots, x_n)$$

- この場合も確率は次のよう計算できる．

$$p(\boldsymbol{x})d\boldsymbol{x} = q(\boldsymbol{y})d\boldsymbol{y},$$

N次元信号の変換の場合

- ただし，体積要素 $d\boldsymbol{x} = dx_1 dx_2 \cdots dx_n$ と $d\boldsymbol{y} = dy_1 \cdots dy_n$ との間には

$$d\boldsymbol{y} = \left| \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \right| d\boldsymbol{x}$$

- いう関係がある．
- $\left| \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \right|$ はヤコビアンと呼ばれ

$$\left| \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \right| = \begin{vmatrix} \frac{\partial f_1}{\partial x_1} & \cdots\cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots\cdots & \frac{\partial f_n}{\partial x_n} \end{vmatrix}$$

- である．したがって，エントロピーは

$$H(Y) = H(X) + \int p(\boldsymbol{x}) \log \left| \frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} \right| d\boldsymbol{x}$$

- と変換される．

例

- 例えば，線形変換

$$\boldsymbol{y} = A\boldsymbol{x}$$

- に対しては，

$$\left|\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}}\right| = |A|$$

- だから，

$$H(Y) = H(X) + \log|A|$$