第 1 回 情報量

たにむら よしのり

株式会社科学計算総合研究所

2020 年 01 月 20 日

Contents

Section 1

情報幾何学と統計問題

情報幾何学・統計問題

情報幾何学とは

統計問題に対する幾何学的なアプローチを調べる学問領域

統計問題とは

発生した現象から、確率測度を推測し、最適行動を決定する確率論の逆問題

コイントス

問題

コイントスで 8 回連続で表が出た。次は何が出る？

- 裏が出やすい：
- どっちも $\mathbf{1/2}$：
- 表が出やすい：

コイントス

問題

コイントスで 8 回連続で表が出た。次は何が出る？

- 裏が出やすい：怪しい宗教
- どっちも $\mathbf{1/2}$：
- 表が出やすい：

コイントス

問題

コイントスで 8 回連続で表が出た。次は何が出る？

- 裏が出やすい：怪しい宗教
- どっちも $\mathbf{1/2}$：確率論的な考え方
- 表が出やすい：

コイントス

問題

コイントスで 8 回連続で表が出た。次は何が出る？

- 裏が出やすい：怪しい宗教
- どっちも $\mathbf{1/2}$：確率論的な考え方
- 表が出やすい：統計学的な考え方

情報量

確率測度を動かして何が起こるか調べたい。

$\longrightarrow$ 確率測度を定量的に評価する量が必要になる。

$\longrightarrow$ そのような指標（のひとつ）が情報量（エントロピー）！

Section 2

平均エントロピー

自己エントロピー

確率 $\dfrac{1}{2^n}$ の事象が発生したとき、その情報量を kn (k: 正の定数) としたい。

- n 個の 0,1 符号で 2^n 通りの場合の数を表すことができる。
- 事象 E の情報量を $I(E)$ で表すとき、独立事象 A, B が次式を満たす。

$$I(A \cap B) = I(A) + I(B)$$

自己エントロピー

確率 $\dfrac{1}{2^n}$ の事象が発生したとき、その情報量を kn (k: 正の定数) としたい。

- n 個の 0,1 符号で 2^n 通りの場合の数を表すことができる。
- 事象 E の情報量を $I(E)$ で表すとき、独立事象 A, B が次式を満たす。

$$I(A \cap B) = I(A) + I(B)$$

Definition (自己エントロピー)

確率 p で発生する事象の自己エントロピーを $-\log p$ で定義する。

平均エントロピーの定義

Definition (自己エントロピー)

確率 p で発生する事象の自己エントロピーを $-\log p$ で定義する。

平均エントロピーの定義

Definition (自己エントロピー)

確率 p で発生する事象の自己エントロピーを $-\log p$ で定義する。

Definition (平均エントロピー)

離散確率空間 Ω の各要素 i が発生する確率を p_i とおく。
このとき、Ω の平均エントロピー H を次式で定義する。

$$H = -\sum_{i \in \Omega} p_i \log p_i$$

但し、右辺のシグマの中では $0 \log 0 = 0$ と解釈する。

自己エントロピーの平均が平均エントロピー

平均エントロピーの例

$$\boldsymbol{H = -\sum_{i \in \Omega} p_i \log p_i}$$

平均エントロピーの例

$$H = -\sum_{i \in \Omega} p_i \log p_i$$

Example

同様に確からしいコイントスの平均エントロピー

$$H = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = \log 2$$

平均エントロピーの例

$$H = -\sum_{i \in \Omega} p_i \log p_i$$

Example

同様に確からしいコイントスの平均エントロピー

$$H = -\left(\frac{1}{2}\log\frac{1}{2} + \frac{1}{2}\log\frac{1}{2}\right) = \log 2$$

Example

同様に確からしいサイコロの平均エントロピー

$$H = -\left(\frac{1}{6}\log\frac{1}{6} + \cdots + \frac{1}{6}\log\frac{1}{6}\right) = \log 6$$

平均エントロピーの例

$$H = -\sum_{i \in \Omega} p_i \log p_i$$

Example

離散空間 $\tilde{n} = \{0, \ldots, n-1\}$ の一様分布の平均エントロピー

$$H = -\sum_{i=0}^{n-1} \frac{1}{n} \log \frac{1}{n} = \log n$$

平均エントロピーの例

$$H = -\sum_{i \in \Omega} p_i \log p_i$$

Example

離散空間 $\tilde{n} = \{0, \ldots, n-1\}$ の一様分布の平均エントロピー

$$H = -\sum_{i=0}^{n-1} \frac{1}{n} \log \frac{1}{n} = \log n$$

Example

離散空間 $\tilde{n} = \{0, \ldots, n-1\}$ のひとつが確率 1 で発生する場合の平均エントロピー

$$H = -(1 \log 1 + (n-1) \cdot 0 \log 0) = -(0 + (n-1) \cdot 0) = 0$$

平均エントロピーの値域

Example

表が出る確率が $\mathbf{1/3}$ のコイントスの平均エントロピー

$$\boldsymbol{H = -\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right) = \log 3 - \frac{2}{3}\log 2 \fallingdotseq 0.2764}$$

これは同様に確からしい場合の平均エントロピー $\boldsymbol{\log 2 \fallingdotseq 0.3010}$ より小さい

平均エントロピーの値域

Example

表が出る確率が $1/3$ のコイントスの平均エントロピー

$$H = -\left(\frac{1}{3}\log\frac{1}{3} + \frac{2}{3}\log\frac{2}{3}\right) = \log 3 - \frac{2}{3}\log 2 \fallingdotseq 0.2764$$

これは同様に確からしい場合の平均エントロピー $\log 2 \fallingdotseq 0.3010$ より小さい

Proposition

離散空間 $\tilde{n} = \{0, \ldots, n-1\}$ の確率分布の平均エントロピーは 0 以上 $\log n$ 以下である。

Remark

- 一様分布は平均エントロピーの最大値 $\log n$ を実現する。
- ひとつの要素だけが確率 1 で発生する場合、平均エントロピーは最小値 0 を実現する。

平均エントロピーの値域

Proposition

離散空間 $\tilde{n}$ の確率分布の平均エントロピー H は 0 以上 $\log n$ 以下である。

証明

まず、$H \geq 0$ を示す。log は上に凸である。従って、$p_i \geq 0$ と $\sum_{i\in\tilde{n}} p_i = 1$ から、

$$-H = \sum_{i\in\tilde{n}} p_i \log p_i \leq \log\left(\sum_{i\in\tilde{n}} (p_i)^2\right)$$

となる。また、$0 \leq p_i \leq 1$ から $(p_i)^2 \leq p_i$ なので、log の単調増加性から、

$$\log\left(\sum_{i\in\tilde{n}} (p_i)^2\right) \leq \log\left(\sum_{i\in\tilde{n}} p_i\right) = \log 1 = 0$$

が成り立つ。従って、確かに $H \geq 0$ である。

平均エントロピーの値域

Proposition

離散空間 $\tilde{n}$ の確率分布の平均エントロピー H は 0 以上 $\log n$ 以下である。

証明

次に、$H \leq \log n$ を示す。log は上に凸である。従って、$p_i \geq 0$ と $\sum_{i \in \tilde{n}} p_i = 1$ から、

$$H = -\sum_{i \in \tilde{n}} p_i \log p_i = \sum_{i \in \tilde{n}} p_i \log \frac{1}{p_i} \leq \log \left(\sum_{i \in \tilde{n}} p_i \cdot \frac{1}{p_i} \right) = \log \left(\sum_{i \in \tilde{n}} 1 \right) = \log n$$

となる。

連続版平均エントロピー

Definition (離散版平均エントロピー)

離散確率空間 Ω の各要素 i が発生する確率を p_i とおく。
このとき、Ω の平均エントロピー H を次式で定義する。

$$H = -\sum_{i \in \Omega} p_i \log p_i$$

但し、右辺のシグマの中では $0 \log 0 = 0$ と解釈する。

連続版平均エントロピー

Definition (離散版平均エントロピー)

離散確率空間 Ω の各要素 i が発生する確率を p_i とおく。
このとき、Ω の平均エントロピー H を次式で定義する。

$$H = -\sum_{i \in \Omega} p_i \log p_i$$

但し、右辺のシグマの中では $0 \log 0 = 0$ と解釈する。

Definition (連続版平均エントロピー)

実数全体 $\mathbb{R}$ 上の確率分布 P の確率密度関数を p とおく。
このとき、P の平均エントロピー $H(P)$ を次式で定義する。

$$H(P) = -\int_{\mathbb{R}} p(x) \log p(x) \, \mathrm{d}x$$

但し、右辺の積分の中では $0 \log 0 = 0$ と解釈する。

平均エントロピーの問題点

離散版	連続版
$\boldsymbol{H} = -\sum_{i\in\boldsymbol{\Omega}} \boldsymbol{p}_i \log \boldsymbol{p}_i$	$\boldsymbol{H} = -\int_{\mathbb{R}} \boldsymbol{p}(\boldsymbol{x}) \log \boldsymbol{p}(\boldsymbol{x})\ \mathrm{d}\boldsymbol{x}$

平均エントロピーの問題点

離散版	連続版
$\boldsymbol{H = -\sum_{i \in \Omega} p_i \log p_i}$	$\boldsymbol{H = -\int_{\mathbb{R}} p(x) \log p(x)\ \mathrm{d}x}$

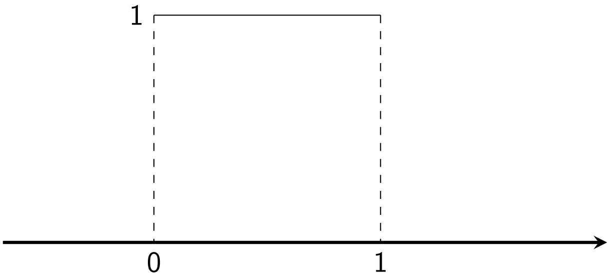
自然な連続化に見えるが、実際はあまり自然じゃない！

平均エントロピーの問題点

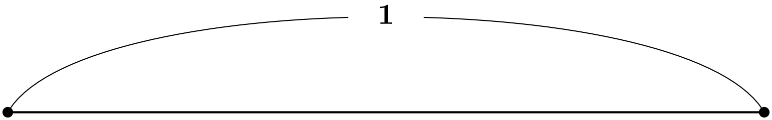
- 連続版平均エントロピーを “離散化” してから “連続化” すると発散する。
- 離散版と連続版を統一的に記述しようとすると、基準となる測度が必要になる。

離散化と連続化

$[\mathbf{0}, \mathbf{1}]$ 区間上の一様分布を考えてみる。

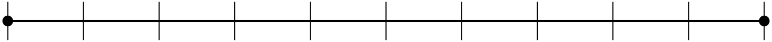


離散化と連続化



長さ 1 の線分

離散化と連続化



n 等分して

離散化と連続化

$$\frac{\bullet}{1/n}\quad \frac{\bullet}{1/n}\quad \frac{\bullet}{1/n}\quad \frac{\bullet}{1/n}\quad \frac{\bullet}{1/n}\quad \frac{\bullet}{1/n}\quad \frac{\bullet}{1/n}\quad \frac{\bullet}{1/n}\quad \frac{\bullet}{1/n}\quad \frac{\bullet}{1/n}$$

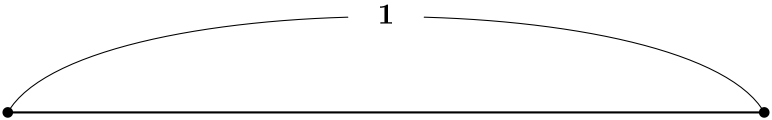
n 点集合に離散化

離散化と連続化

極限 $n \to \infty$ で連続化したら元に戻って欲しい。。。

離散化と連続化

元の空間



$$\begin{aligned} H &= -\int_{\mathbb{R}} p(x) \log p(x) \ \mathrm{d}x \\ &= -\left(\int_{\mathbb{R} \setminus [0,1]} 0 \cdot \log 0 \ \mathrm{d}x + \int_{[0,1]} 1 \cdot \log 1 \ \mathrm{d}x \right) = 0 \end{aligned}$$

離散化と連続化

離散化

$$\bullet_{1/n} \quad \bullet_{1/n} \quad \bullet_{1/n} \quad \bullet_{1/n} \quad \bullet_{1/n} \quad \bullet_{1/n} \quad \bullet_{1/n} \quad \bullet_{1/n} \quad \bullet_{1/n} \quad \bullet_{1/n}$$

$$\boldsymbol{H = -\sum_{i \in \tilde{n}} \frac{1}{n} \log \frac{1}{n} = \log n}$$

離散化と連続化

連続化

$$H = -\sum_{i \in \tilde{n}} \frac{1}{n} \log \frac{1}{n} = \log n \xrightarrow{n \to \infty} \infty$$

離散化と連続化

平均エントロピー

元の一様分布：	$\boldsymbol{H = 0}$
離散化：	$\boldsymbol{H = \log n}$
連続化：	$\boldsymbol{H \to \infty}$

離散化と連続化

平均エントロピー

元の一様分布：	$H = 0$
離散化：	$H = \log n$
連続化：	$H \to \infty$

離散化して連続化すると発散してしまう！

基準となる測度の存在

$\mathbb{R}$ 上の確率分布 $\boldsymbol{P}$ の確率密度関数を $\boldsymbol{p}$ とおくとき、

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

基準となる測度の存在

$\mathbb{R}$ 上の確率分布 $\boldsymbol{P}$ の確率密度関数を $\boldsymbol{p}$ とおくとき、

$$\boldsymbol{H(P) = -\int_{\mathbb{R}} p(x) \log p(x) \ \mathrm{d}x = -\int_{\mathbb{R}} \log {\color{orange}p(x)} \ \mathrm{d}P(x)}$$

式から確率密度関数を消せない！
Lebesgue 測度の寄与を無視できない！

平均情報量を一般化するのであれば。。。

可測空間 $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}})$ 上の測度 $\boldsymbol{\mu}$ をひとつ固定する。

Definition

$\boldsymbol{\mu}$ に絶対連続な $(\boldsymbol{\Omega}, \boldsymbol{\mathcal{F}})$ 上の確率測度 $\boldsymbol{P}$ の平均情報量 $\boldsymbol{H_\mu(P)}$ を以下で定義する。

$$\boldsymbol{H_\mu(P) := -\int_\Omega \log\left(\frac{\mathrm{d}P}{\mathrm{d}\mu}\right)\mathrm{d}P}$$

復習：絶対連続性

$(\Omega, \mathcal{F})$：可測空間

Definition

$\Phi\colon \mathcal{F} \to \mathbb{C}$ が完全加法的集合関数であるとは、次式が成り立つことをいう。

$$\Phi\left(\sum_{n\in\mathbb{N}} E_n\right) = \sum_{n\in\mathbb{N}} \Phi(E_n) \text{ where } \{E_n\}_{n\in\mathbb{N}} \subset \mathcal{F}\text{: disjoint}$$

Remark

0 以上の実数のみに値をとる完全加法的集合関数は有限測度に他ならない。

復習：絶対連続性

$(\Omega, \mathcal{F}, \mu)$：$\sigma$-有限な測度空間

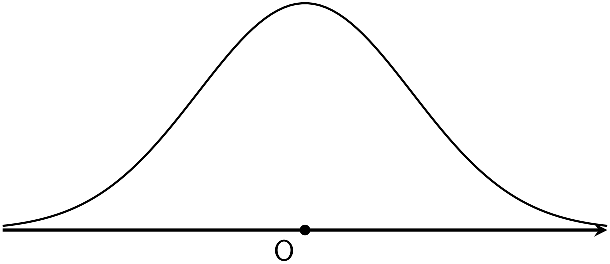
Definition

$\Phi : \mathcal{F} \to \mathbb{C}$：完全加法的集合関数

- Φ が μ に対して絶対連続であるとは、任意の $E \in \mathcal{F}$ に対して、
$$\mu(E) = 0 \Longrightarrow \Phi(E) = 0$$
が成り立つことを言いい、これを $\Phi \ll \mu$ で表す。
- Φ が μ に対して特異であるとは、ある $E \in \mathcal{F}$ が存在して、
$$\mu(E) = 0, \quad \Phi(E^c) = 0$$
が成り立つことを言う。

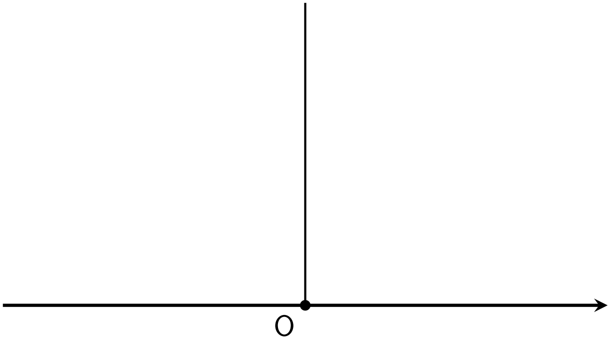
復習：絶対連続性

Lebesgue 測度に対して正規分布は絶対連続である。



復習：絶対連続性

Lebesgue 測度に対してデルタ分布は特異である。



復習：Radon-Nikodým 微分

$(\Omega, \mathcal{F}, \mu)$: σ-有限な測度空間　　$\Phi \colon \mathcal{F} \to \mathbb{C}$: 完全加法的集合関数

Theorem (Radon-Nikodým の定理)

Φ が μ に対して絶対連続であるとき、
次式を満たす $\frac{\mathrm{d}\Phi}{\mathrm{d}\mu} \in L^1(\Omega, \mathcal{F}, \mu)$ が唯一つ存在する。

$$\Phi(E) = \int_E \frac{\mathrm{d}\Phi}{\mathrm{d}\mu} \, \mathrm{d}\mu \quad (E \in \mathcal{F})$$

Definition (Radon-Nikodým 微分)

上の定理における $\frac{\mathrm{d}\Phi}{\mathrm{d}\mu}$ を Φ の Radon-Nikodým 微分と呼ぶ。

復習：Radon-Nikodým 微分

$(\Omega, \mathcal{F}, \mu)$: σ-有限な測度空間　　$\Phi \colon \mathcal{F} \to \mathbb{C}$: 完全加法的集合関数

Theorem (Radon-Nikodým の定理)

Φ が μ に対して絶対連続であるとき、
次式を満たす $\frac{\mathrm{d}\Phi}{\mathrm{d}\mu} \in L^1(\Omega, \mathcal{F}, \mu)$ が唯一つ存在する。

$$\Phi(E) = \int_E \frac{\mathrm{d}\Phi}{\mathrm{d}\mu} \, \mathrm{d}\mu \quad (E \in \mathcal{F})$$

Definition (Radon-Nikodým 微分)

上の定理における $\frac{\mathrm{d}\Phi}{\mathrm{d}\mu}$ を Φ の Radon-Nikodým 微分と呼ぶ。

Remark

確率密度関数は *Lebesgue* 測度に対する確率測度の *Radon-Nikodým* 微分である。

平均情報量を一般化するのであれば。。。

可測空間 $(\Omega, \mathcal{F})$ 上の測度 μ をひとつ固定する。

Definition

μ に絶対連続な $(\Omega, \mathcal{F})$ 上の確率測度 P の平均情報量 $H_\mu(P)$ を以下で定義する。

$$H_\mu(P) := -\int_\Omega \log\left(\frac{\mathrm{d}P}{\mathrm{d}\mu}\right) \mathrm{d}P$$

平均情報量を一般化するのであれば。。。

可測空間 $(\Omega, \mathcal{F})$ 上の測度 μ をひとつ固定する。

Definition

μ に絶対連続な $(\Omega, \mathcal{F})$ 上の確率測度 P の平均情報量 $H_\mu(P)$ を以下で定義する。

$$H_\mu(P) := -\int_\Omega \log\left(\frac{\mathrm{d}P}{\mathrm{d}\mu}\right)\mathrm{d}P$$

$\longrightarrow$ 測度に依らない、可測空間上の確率測度に定義できる指標が欲しい！

$\longrightarrow$ 相対エントロピー！

Section 3

相対エントロピー

相対エントロピー

Definition (離散版相対エントロピー)

Ω : 離散空間　　P, Q : Ω 上の確率測度　　$p_i := P(\{i\}), q_i := Q(\{i\})$ for $i \in \Omega$
このとき、相対エントロピー $D_{KL}(Q \| P)$ を次式で定義する。

$$D_{KL}(Q \| P) := -\sum_{i \in \Omega} q_i \log\left(\frac{p_i}{q_i}\right)$$

Definition (連続版相対エントロピー)

P, Q : 絶対連続な確率分布　　p, q : P, Q の確率密度関数
このとき、相対エントロピー $D_{KL}(Q \| P)$ を次式で定義する。

$$D_{KL}(Q \| P) := -\int_{\mathbb{R}} q(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x$$

離散版相対エントロピーの例

確率 $1/2$ で表が出ると思っていたコインが実は確率 $2/3$ で表が出るコインだった！

この場合の相対エントロピーは：

$$
\begin{aligned}
D_{KL}\left(\left(\frac{2}{3},\frac{1}{3}\right)\middle\|\left(\frac{1}{2},\frac{1}{2}\right)\right) &= -\frac{2}{3}\log\frac{1/2}{2/3}-\frac{1}{3}\log\frac{1/2}{1/3}\\
&= -\frac{2}{3}\log\frac{3}{4}-\frac{1}{3}\log\frac{3}{2}\\
&= \frac{5}{3}\log 2-\log 3 \fallingdotseq 0.02460
\end{aligned}
$$

連続版相対エントロピーの例

正規分布 $N(\mu, \sigma)$ と $N(\nu, \theta)$ の間の相対エントロピー

$$D_{KL}\left((\mu, \sigma) \| (\nu, \theta)\right) = D_{KL}\left((\mu, \theta) \| (\nu, \theta)\right) + D_{KL}\left((0, \sigma) \| (0, \theta)\right)$$

$$D_{KL}\left((\mu, \theta) \| (\nu, \theta)\right) = \frac{(\mu - \nu)^2}{2\theta^2}$$

$$D_{KL}\left((0, \sigma) \| (0, \theta)\right) = \frac{\sigma^2}{2\theta^2} + \log\frac{\theta}{\sigma} - \frac{1}{2}$$

相対エントロピーの意味

「自分が思っている情報量」−「真の情報量」　の　期待値

相対エントロピーの意味

確率 $\mathbf{1/2}$ で表が出ると思っていたコインが実は確率 $\mathbf{2/3}$ で表が出るコインだった！

相対エントロピーの意味

確率 $\mathbf{1/2}$ で表が出ると思っていたコインが実は確率 $\mathbf{2/3}$ で表が出るコインだった！

表が出た場合の情報量を考えると：

- 本当は表が出たら $-\log 2/3$ の情報量しか得られていないはず。。。
- 自分はこの事象には $-\log 1/2$ の情報量があると勝手に思っている

相対エントロピーの意味

確率 $1/2$ で表が出ると思っていたコインが実は確率 $2/3$ で表が出るコインだった！

表が出た場合の情報量を考えると：

- 本当は表が出たら $-\log 2/3$ の情報量しか得られていないはず。。。
- 自分はこの事象には $-\log 1/2$ の情報量があると勝手に思っている

$\longrightarrow$ 自分が思っている情報量と実際に得られた情報量の差は：

$$(-\log 1/2) - (-\log 2/3) = -\log \frac{1/2}{2/3}$$

相対エントロピーの意味

確率 $1/2$ で表が出ると思っていたコインが実は確率 $2/3$ で表が出るコインだった！

自分が思っている情報量と実際に得られた情報量の差

- 表が出た場合：$(-\log 1/2) - (-\log 2/3) = -\log \frac{1/2}{2/3}$
- 裏が出た場合：$(-\log 1/2) - (-\log 1/3) = -\log \frac{1/2}{1/3}$

自分が思っている情報量と実際に得られた情報量の差の期待値は：

$$\frac{2}{3} \cdot \left(-\log \frac{1/2}{2/3}\right) + \frac{1}{3} \cdot \left(-\log \frac{1/2}{1/3}\right) = D_{KL}\left(\left(\frac{2}{3}, \frac{1}{3}\right) \middle\| \left(\frac{1}{2}, \frac{1}{2}\right)\right)$$

相対エントロピーの意味

$\boldsymbol{\Omega}$：離散空間　　$\boldsymbol{Q}$：真の確率測度　　$\boldsymbol{P}$：自分の予想

事象 $\boldsymbol{i} \in \boldsymbol{\Omega}$ が発生した際に、自分が思っている情報量と実際に得られた情報量の差：

$$(-\log p_i) - (-\log q_i) = -\log \frac{p_i}{q_i}$$

自分が思っている情報量と実際に得られた情報量の差の期待値は：

$$\sum_{i \in \Omega} q_i \cdot \left(-\log \frac{p_i}{q_i} \right) = D_{KL}(Q \| P)$$

相対エントロピーの意味

$\boldsymbol{\Omega}$: 離散空間　　$\boldsymbol{Q}$: 真の確率測度　　$\boldsymbol{P}$: 自分の予想

事象 $\boldsymbol{i} \in \boldsymbol{\Omega}$ が発生した際に、自分が思っている情報量と実際に得られた情報量の差：

$$(-\log p_i) - (-\log q_i) = -\log \frac{p_i}{q_i}$$

自分が思っている情報量と実際に得られた情報量の差の期待値は：

$$\sum_{i \in \Omega} q_i \cdot \left(-\log \frac{p_i}{q_i}\right) = D_{KL}(Q \| P)$$

相対エントロピーとは

「自分が思っている情報量」−「真の情報量」　の　期待値

相対エントロピーの一般化

連続版相対エントロピーに対して：

$$D_{KL}(Q\|P) = -\int_{\mathbb{R}} q(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x = -\int_{\mathbb{R}} \log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}Q$$

相対エントロピーの一般化

連続版相対エントロピーに対して：

$$D_{KL}(Q\|P) = -\int_{\mathbb{R}} q(x)\log\left(\frac{p(x)}{q(x)}\right)\mathrm{d}x = -\int_{\mathbb{R}}\log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\mathrm{d}Q$$

Definition (相対エントロピー)

$(\Omega,\mathcal{F})$：可測空間　　P,Q：$(\Omega,\mathcal{F})$ 上の確率測度、互いに絶対連続
このとき、相対エントロピー $D_{KL}(Q\|P)$ を次式で定義する。

$$D_{KL}(Q\|P) := -\int_{\Omega}\log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right)\mathrm{d}Q$$

この定義は離散版と連続版の定義を両方含んでいる！

相対エントロピーの性質

$(\Omega, \mathcal{F})$：可測空間　　P, Q：$(\Omega, \mathcal{F})$ 上の確率測度、互いに絶対連続

Definition (相対エントロピー)

相対エントロピー $D_{KL}(Q\|P)$ を次式で定義する。

$$D_{KL}(Q\|P) := -\int_\Omega \log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}Q$$

相対エントロピーの性質

$(\Omega, \mathcal{F})$：可測空間　　P, Q：$(\Omega, \mathcal{F})$ 上の確率測度、互いに絶対連続

Definition (相対エントロピー)

相対エントロピー $D_{KL}(Q\|P)$ を次式で定義する。

$$D_{KL}(Q\|P) := -\int_{\Omega} \log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}Q$$

Proposition

$D_{KL}(Q\|P) \geq 0$ であり、これの等号成立条件は $P = Q$ である。

相対エントロピーの性質

$(\Omega, \mathcal{F})$: 可測空間　　P, Q : $(\Omega, \mathcal{F})$ 上の確率測度、互いに絶対連続

Definition (相対エントロピー)

相対エントロピー $D_{KL}(Q\|P)$ を次式で定義する。

$$D_{KL}(Q\|P) := -\int_{\Omega} \log\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}Q$$

Proposition

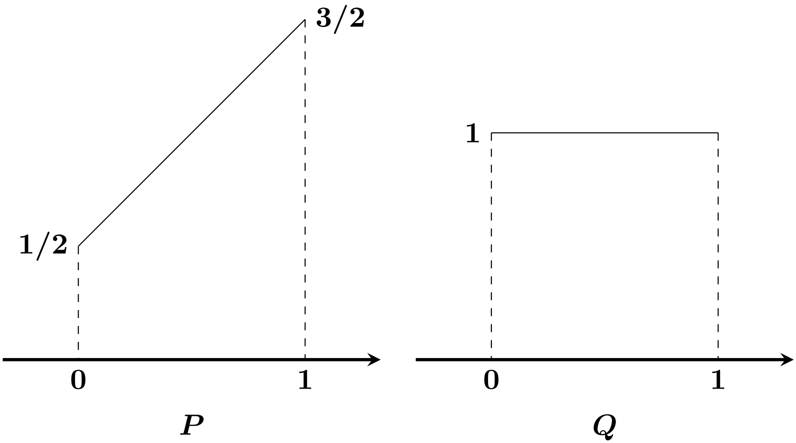
$D_{KL}(Q\|P) \geq 0$ であり、これの等号成立条件は $P = Q$ である。

Remark

$D_{KL}(P\|Q) = D_{KL}(Q\|P)$ とは限らない。

離散化と連続化

元の分布



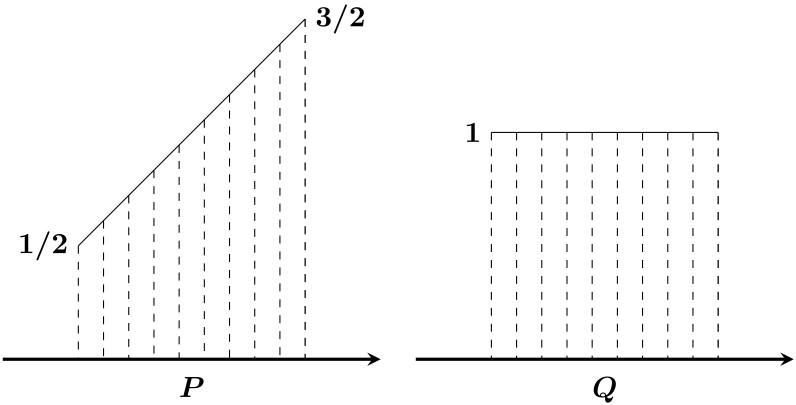
離散化と連続化

$$p(x) = x + \frac{1}{2} \qquad q(x) = 1 \qquad (x \in [0,1])$$

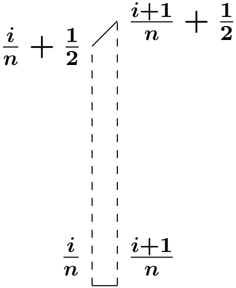
元の分布の相対エントロピー

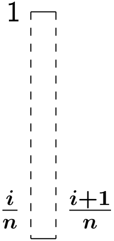
$$\begin{aligned}
D_{KL}(Q\|P) &= -\int_{\mathbb{R}} q(x) \log\left(\frac{p(x)}{q(x)}\right) \mathrm{d}x \\
&= -\int_0^1 1 \cdot \log\left(\frac{x+1/2}{1}\right) \mathrm{d}x \\
&= -\int_0^1 \log\left(x + \frac{1}{2}\right) \mathrm{d}x
\end{aligned}$$

離散化



離散化





$$P\left(\left[\frac{i}{n},\frac{i+1}{n}\right]\right)=\frac{1}{2}\left(\frac{2i+1}{n}+1\right)\frac{1}{n} \qquad Q\left(\left[\frac{i}{n},\frac{i+1}{n}\right]\right)=\frac{1}{n}$$

離散化と連続化

離散化された確率分布

$$p_i = \frac{n+2i+1}{2n^2} \qquad q_i = \frac{1}{n} \qquad (i \in \tilde{n})$$

離散化された分布の相対エントロピー

$$\begin{aligned} D_{KL}(Q_d \| P_d) &= -\sum_{i=0}^{n-1} q_i \cdot \log\left(\frac{p_i}{q_i}\right) \\ &= -\sum_{i=0}^{n-1} \frac{1}{n} \log\left(\frac{(n+2i+1)/(2n^2)}{1/n}\right) \\ &= -\sum_{i=0}^{n-1} \frac{1}{n} \log\left(\frac{i}{n} + \frac{1}{2} + \frac{1}{2n}\right) \end{aligned}$$

離散化と連続化

連続化

$$D_{KL}(Q_d \| P_d) = -\sum_{i=0}^{n-1} \frac{1}{n} \log\left(\frac{i}{n} + \frac{1}{2} + \frac{1}{2n}\right)$$
$$\xrightarrow{n\to\infty} -\int_0^1 \log\left(x + \frac{1}{2}\right)\,\mathrm{d}x = D_{KL}(Q \| P)$$

離散化と連続化

連続化

$$
\begin{aligned}
D_{KL}(Q_d \| P_d) &= -\sum_{i=0}^{n-1} \frac{1}{n} \log\left(\frac{i}{n} + \frac{1}{2} + \frac{1}{2n}\right) \\
&\xrightarrow{n\to\infty} -\int_0^1 \log\left(x + \frac{1}{2}\right)\,\mathrm{d}x = D_{KL}(Q\|P)
\end{aligned}
$$

離散化して連続化したら元に戻った！

Proposition

$(\Omega, \mathcal{F})$: 可測空間　　P, Q : $(\Omega, \mathcal{F})$ 上の確率測度、互いに絶対連続
このとき、$D_{KL}(Q\|P) \geq 0$ であり、これの等号成立条件は $P = Q$ である。

実は、これは $-\log x$ の凸性によってのみ証明される。

$\longrightarrow$ 別の凸関数を採用しても同様の性質がいえるはず！

f-divergence

$(\Omega, \mathcal{F})$: 可測空間　　P, Q : $(\Omega, \mathcal{F})$ 上の確率測度、互いに絶対連続
$f : \mathbb{R}_{>0} \to \mathbb{R}$: 狭義凸関数、$f(1) = 0$

Definition (f-divergence)

f-divergence $D_f(Q \| P)$ を次式で定義する。

$$D_f(Q \| P) := \int_\Omega f\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}Q$$

相対エントロピーを $-\log$-divergence とみたものを Kullback-Leibler divergence と呼ぶ。

f-divergence

$(\Omega, \mathcal{F})$: 可測空間　　P, Q : $(\Omega, \mathcal{F})$ 上の確率測度、互いに絶対連続

$f : \mathbb{R}_{>0} \to \mathbb{R}$: 狭義凸関数、$f(1) = 0$

Definition (f-divergence)

f-divergence $D_f(Q\|P)$ を次式で定義する。

$$D_f(Q\|P) := \int_\Omega f\left(\frac{\mathrm{d}P}{\mathrm{d}Q}\right) \mathrm{d}Q$$

相対エントロピーを $-\log$-divergence とみたものを Kullback-Leibler divergence と呼ぶ。

Proposition (f-divergence の基本性質)

$D_f(Q\|P) \geq 0$ であり、これの等号成立条件は $P = Q$ である。

Section 4

まとめ

まとめ：情報幾何学と統計問題

- 情報幾何学 ＝ 統計学 ＋ 幾何学
- 統計問題 ＝ 確率論の逆問題
- 確率測度を定量的に評価するために情報量を導入する。

まとめ：平均エントロピー

離散版	連続版
$$H = -\sum_{i \in \Omega} p_i \log p_i$$	$$H = -\int_{\mathbb{R}} p(x) \log p(x) \, \mathrm{d}x$$

平均エントロピーには以下の欠点がある。

- 離散版と連続版を確率測度の範疇だけで統一的に記述することができない。
- 連続版を “離散化” して再度 “連続化” しようとすると、発散してうまくいかない。

まとめ：相対エントロピー

離散版	連続版
$D_{KL}(Q\|P) = -\sum_{i\in\Omega} q_i \log\left(\frac{p_i}{q_i}\right)$	$D_{KL}(Q\|P) = -\int_{\mathbb{R}} q(x) \log\left(\frac{p(x)}{q(x)}\right)\,\mathrm{d}x$

- 相対エントロピー ＝ 自分の予想した情報量と実際の情報量の差の期待値
- 一般の可測空間上に一般化できる。
- 有限集合上の相対エントロピーの極限としても定式化することができる。
- 正の実数全体からの狭義凸関数 f が $f(1) = 0$ を満たせば f-divergence が定義できる。