W8PRML 演習問題 1.28, 1.29

大上 雅史　(@tonets)

2012/9/28

1.28

1.6 節で，エントロピー $h(x)$ のアイディアを，確率分布 $p(x)$ を持つ確率変数 x の値を観測することによって増える情報量として導入した．また，変数 x, y が $p(x, y) = p(x)p(y)$ となって独立なときは，エントロピーは加法的で，$h(x, y) = h(x) + h(y)$ となることを見た．この演習問題では，h と p の間の関数関係 $h(p)$ を導く．まず，$h(p^2) = 2h(p)$ となることを示し，数学的帰納法により，正の整数 n に対し $h(p^n) = nh(p)$ となることを示せ．さらに正の整数 m に対し，$h(p^{n/m}) = (n/m)h(p)$ が成り立つことを示せ．このことから x が正の有理数のとき，$h(p^x) = xh(p)$ が成り立つが，これは連続性により正の実数値の場合も成り立つ．最後にこのことから $h(p)$ が $h(p) \propto \ln p$ の形を取らなければならないことを示せ．

要するに**情報量の性質を満たすのは対数関数しかないことを証明する**問題である．利用して良い条件は，

- $h(xy) = h(x) + h(y)$
  ※ x, y が独立のとき，$h(x, y) = h(p(x, y)) = h(p(x)p(y)) = h(xy)$.
- $h(\cdot)$ の定義域は正の実数

の 2 つである．

(1) $h(p^2) = 2h(p)$ **の証明**

$$h(p^2) = h(p \cdot p) = h(p) + h(p) = 2h(p)$$

(2) $h(p^n) = nh(p) \quad (n \in \mathbb{N}^+)$ **の証明**

数学的帰納法で示す．$n = k$ のとき，$h(p^k) = kh(p)$ が成り立つとすると，$n = k + 1$ のとき，

$$h(p^{k+1}) = h(p^k p) = h(p^k) + h(p) = kh(p) + h(p) = (k + 1)h(p)$$

となって成り立つ．

(3) $h(p^{n/m}) = (n/m)h(p) \quad (n, m \in \mathbb{N}^+)$ **の証明**

$mh(p^{\frac{n}{m}})$ を考える．

$$mh\left(p^{\frac{n}{m}}\right) = h\left(\left(p^{\frac{n}{m}}\right)^m\right) = h(p^n) = nh(p)$$

$$\therefore h(p^{\frac{n}{m}}) = \frac{n}{m}h(p)$$

(4) $h(p^x) = xh(p) \quad (x \in \mathbb{R}, x > 0)$ **の証明**

$\forall x \in \mathbb{R}(x > 0)$ に対し，$\lim_{n\to\infty} q_n = x$ となる $q_n \in \mathbb{Q}, q_n > 0$ を取ってくると，

$$h(p^x) = h\left(p^{\lim_{n\to\infty} q_n}\right) = \lim_{n\to\infty} h(p^{q_n}) = \lim_{n\to\infty} q_n h(p) = xh(p)$$

(5) $h(p)$ が $h(p) \propto \ln p$ の形を取らなければならないことの証明

$g(q) = h(e^q)$ という q の関数 $g(q)$ を考えると，(4) より $\forall t \in \mathbb{R}$ に対して，

$$g(tq) = h(e^{tq}) = th(e^q) = tg(q)$$

となるので，$q = 1, c \triangleq g(1)$ とすると，$g(t) = ct$ となる．即ち，

$$g(t) = h(e^t) = ct$$

となる．$p = e^t$ とおくと，$t = \ln p$ なので，

$$\therefore h(p) = c \ln p$$

である．(ただし，$c = g(1) = h(e^1)$ は比例定数.)

コメント：普通は，情報量 $h(p)$ が (a) 連続関数，単調非増加で $p \in (0,1]$，(b) $h(p_1 p_2) = h(p_1) + h(p_2)$ を満たすという前提からこの証明をします．(4) は h の連続性を仮定していますが，連続性が満たされてなくても証明できるらしいです.)

1.29

M 状態の離散確率変数 x を考え，イェンセンの不等式 (1.115) を使って，確率分布 $p(x)$ のエントロピーが $H[x] \leq \ln M$ を満たすことを示せ.

エントロピーは,

$$H[x] = -\sum_{i=1}^{M} p(x_i) \ln p(x_i) = \sum_{i=1}^{M} p(x_i) \ln \frac{1}{p(x_i)}$$

となる．イェンセンの不等式を使うにあたり，関数 $f(x) = \ln x$ が凹関数であることを示す.
$f(x)$ の 2 階微分は,

$$\frac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} = -\frac{1}{x^2} < 0$$

より，関数の値は至る所で負であることから，$\ln x$ は凹関数である.
イェンセンの不等式は，凹関数 $f(x)$ に対して,

$$f\left(\sum_{i=1}^{M} \lambda_i x_i\right) \geq \sum_{i=1}^{M} \lambda_i f(x_i)$$

となるので，$\lambda_i \to p(x_i)$，$x_i \to 1/p(x_i)$，$f(x) \to \ln x$ と対応させて,

$$\begin{aligned} H[x] &= \sum_{i=1}^{M} p(x_i) \ln \frac{1}{p(x_i)} \\ &\leq \ln\left(\sum_{i=1}^{M} p(x_i) \frac{1}{p(x_i)}\right) \\ &= \ln\left(\sum_{i=1}^{M} 1\right) \\ &= \ln M \end{aligned}$$

を得る.

コメント：この不等式は，一様分布 $(p(x_i) = 1/M)$ のときにエントロピーが最大となることを示しています.