イェンセンの不等式

エントロピーの関係を求めるときに出てくるイェンセンの不等式を導出します。

イェンセン (Jensen) の不等式は凸関数の関係なので、最初に凸関数を見ていきます。

凸関数 (とつかんすう、convex function) と凹関数（おうかんすう、concave function）の定義を与えます。視覚的に言えば、連続な関数 f があり、任意の 2 点 $(x_1, f(x_1))$, $(x_2, f(x_2))$ を直線で繋いだとき、$f(x)$ $(x_1 \leq x \leq x_2)$ がその直線より下の値を持つなら凸関数です。なので、凸は下に凸の意味になっています。凹関数は直線より上の値になっている場合で ($-f(x)$ が凸関数の場合)、凹は下に凹 (上に凸) の意味です。

定義を式にします。凸関数としていきますが、凹関数では $-f(x)$ に変えるだけです。$y = f(x)$ における 2 点 x_1, x_2 を繋ぐ直線 $g(x)$ は

$$g(x) = \frac{f(x_2) - f(x_1)}{x_2 - x_1}x + \frac{x_2 f(x_1) - x_1 f(x_2)}{x_2 - x_1}$$

これを変形すると

$$\frac{1}{x_2 - x_1}(xf(x_2) - x_1 f(x_2) - xf(x_1) + x_2 f(x_1)) = \frac{1}{x_2 - x_1}((x_2 - x)f(x_1) + (x - x_1)f(x_2))$$

ここで

$$\lambda = \frac{x_2 - x}{x_2 - x_1}\ ,\ \frac{x - x_1}{x_2 - x_1} = \frac{x_2 - x_1 - (x_2 - x)}{x_2 - x_1} = 1 - \lambda$$

とすれば

$$g(x) = \lambda f(x_1) + (1 - \lambda)f(x_2)$$

と書けます。λ は $x_1 \leq x \leq x_2$ から $0 \leq \lambda \leq 1$ です。

凸関数の定義から $f(x)$ は $g(x)$ 以下であればいいので、$f(x) \leq g(x)$ です。x は

$$x = -(x_2 - x_1)\lambda + x_2 = \lambda x_1 + (1 - \lambda)x_2$$

なので、凸関数は λ を使うと

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2)$$

と定義されます。

区間 I において関数 f が 2 回微分可能で、その 2 階微分 f'' が $f''(x) \geq 0$ $(x \in I)$ であるとき、その関数は凸関数となることを示します。$(x_1, f(x_1))$ と $(x_2, f(x_2))$ $(x_1, x_2 \in I)$ を繋いだ直線の関数 $g(x)$ と $f(x)$ $(x_1 \leq x \leq x_2)$ との差は

$$
\begin{aligned}
g(x)-f(x) &= \lambda f(x_1)+(1-\lambda)f(x_2)-f(x)\\
&= \lambda f(x_1)+(1-\lambda)f(x_2)-\frac{x-x_1+x_2-x}{x_2-x_1}f(x)\\
&= \frac{x_2-x}{x_2-x_1}f(x_1)+\frac{x-x_1}{x_2-x_1}f(x_2)-\frac{x-x_1}{x_2-x_1}f(x)-\frac{x_2-x}{x_2-x_1}f(x)\\
&= \frac{x_2-x}{x_2-x_1}(f(x_1)-f(x))+\frac{x-x_1}{x_2-x_1}(f(x_2)-f(x))\\
&= \frac{(x_2-x)(x_1-x)}{x_2-x_1}\frac{f(x_1)-f(x)}{x_1-x}+\frac{(x-x_1)(x_2-x)}{x_2-x_1}\frac{f(x_2)-f(x)}{x_2-x}
\end{aligned}
$$

平均値の定理

$$
f(b)-f(a)=f'(c)(b-a) \quad (a<c<b,\ f'(x)=\frac{df}{dx})
$$

から

$$
\begin{aligned}
\frac{f(x)-f(x_1)}{x-x_1} &= f'(c_1) \quad (x_1<c_1<x)\\
\frac{f(x_2)-f(x)}{x_2-x} &= f'(c_2) \quad (x<c_2<x_2)
\end{aligned}
$$

となるので

$$
\begin{aligned}
g(x)-f(x) &= \frac{(x_2-x)(x_1-x)}{x_2-x_1}f'(c_1)-\frac{(x_2-x)(x_1-x)}{x_2-x_1}f'(c_2)\\
&= \frac{(x_2-x)(x_1-x)}{x_2-x_1}(f'(c_1)-f'(c_2))\\
&= -\frac{(x_2-x)(x_1-x)(c_2-c_1)}{x_2-x_1}\frac{f'(c_2)-f'(c_1)}{c_2-c_1}
\end{aligned}
$$

平均値の定理をもう一度使えば

$$
\frac{f'(c_2)-f'(c_1)}{c_2-c_1}=f''(c_0) \quad (c_1<c_0<c_2)
$$

となるので

$$
g(x)-f(x)=\frac{(x_2-x)(x-x_1)(c_2-c_1)}{x_2-x_1}f''(c_0)
$$

$x_1 \leq x \leq x_2$ と $c_2-c_1>0$ なので、$f''(c_0) \geq 0$ のとき

$$
f(x) \leq g(x)
$$

となり、凸関数の定義になります。よって、2 回微分可能な関数 f が $f''(x) \geq 0$ なら凸関数となります。
　関数 f が凸関数であるなら $f''(x) \geq 0$ も言えます。f が凸関数なら $(x \pm h \in I)$

$$f(x) = f(\frac{1}{2}(x+h) + \frac{1}{2}(x-h)) = f(\frac{1}{2}x_1 + \frac{1}{2}x_2) \ \Leftrightarrow \ f(\lambda x_1 + (1-\lambda)x_2)$$

との対応から

$$f(x) = f(\frac{1}{2}(x+h) + \frac{1}{2}(x-h)) \leq \frac{1}{2}f(x+h) + \frac{1}{2}f(x-h)$$

これを $f''(x)$ は

$$f''(x) = \lim_{h\to 0}\frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

と書けることに使えば

$$\frac{1}{2}(f(x+h) - 2f(x) + f(x-h)) = \frac{1}{2}f(x+h) + \frac{1}{2}f(x-h) - f(x) \geq 0$$

と分かるので、$f''(x) \geq 0$ となります。
　対数 $\log_a x \ (x > 0)$ は

$$\frac{d}{dx}\log_a x = \frac{1}{\ln a}\frac{1}{x} \ , \ \frac{d^2}{dx^2}\log_a x = -\frac{1}{\ln a}\frac{1}{x^2}$$

なので、凹関数です。$x\log_a x$ では

$$\frac{d}{dx}x\log_a x = \log_a x + \frac{1}{\ln a} \ , \ \frac{d^2}{dx^2}x\log_a x = \frac{1}{\ln a}\frac{1}{x}$$

なので、$x\log_a x$ は凸関数です。
　ここからイェンセンの不等式を示していきます。$f(x)$ を凸関数とします。離散的な x_1, x_2 に対して $\lambda(x_1), \lambda(x_2)$ があり、$\lambda(x_1) + \lambda(x_2) = 1 \ (0 \leq \lambda(x_i) \leq 1)$ とします。これによって

$$F(x_1, x_2) = \lambda(x_1)f(x_1) + \lambda(x_2)f(x_2) = \lambda(x_1)f(x_1) + (1 - \lambda(x_1))f(x_2)$$

と作ります。このとき、$f(x)$ が凸関数なら

$$f(\lambda(x_1)x_1 + (1 - \lambda(x_1)x_2) \leq \lambda(x_1)f(x_1) + (1 - \lambda(x_1))f(x_2) \tag{1}$$

となるので

$$X_2 = \lambda(x_1)x_1 + (1 - \lambda(x_1))x_2$$

と書くことにすれば、不等式は

$$F(x_1, x_2) \leq f(X_2) \tag{2}$$

となっています。

x_1, x_2, x_3 として同じことをしてみます。ゴチャゴチャするので $\lambda(x_i), f(x_i)$ を λ_i, f_i と書くことにして

$$\lambda_1 + \lambda_2 + \lambda_3 = 1 \quad (0 \leq \lambda_i \leq 1)$$

これは

$$\frac{\lambda_1 + \lambda_2}{1 - \lambda_3} = 1$$

となっていることを利用して

$$\begin{aligned}
F(x_1, x_2, x_3) &= \sum_{i=1}^{3} \lambda_i f_i \\
&= \lambda_3 f_3 + \sum_{i=1}^{2} \lambda_i f_i \\
&= \lambda_3 f_3 + (1 - \lambda_3) \sum_{i=1}^{2} \frac{\lambda_i}{1 - \lambda_3} f_i \\
&= \lambda_3 f_3 + \overline{\lambda}_3 (\lambda'_1 f_1 + \lambda'_2 f_2) \quad (\overline{\lambda}_3 = 1 - \lambda_3,\ \lambda'_1 = \frac{\lambda_1}{\overline{\lambda}_3},\ \lambda'_2 = \frac{\lambda_2}{\overline{\lambda}_3}) \qquad (3)
\end{aligned}$$

第 2 項の括弧部分は 2 個の場合になっているので

$$F'(x_1, x_2) = \lambda'_1 f_1 + \lambda'_2 f_2\ ,\ f'(X_2) = f(\lambda'_1 x_1 + \lambda'_2 x_2) \quad (0 \leq \lambda'_i \leq 1)$$

とすれば、(2) から $F'(x_1, x_2) \geq f'(X_2)$ です。実際に

$$\lambda'_1 f_1 + \lambda'_2 f_2 = \lambda'_1 f_1 + (1 - \lambda'_1) f_2$$

と変形すれば、(1) が出てきます。これを (3) に入れることで

$$F(x_1, x_2, x_3) = \lambda_3 f_3 + \overline{\lambda}_3 F'(x_1, x_2) \geq \lambda_3 f_3 + (1 - \lambda_3) f'(X_2) = \lambda_3 f_3 + (1 - \lambda_3) f'(\lambda'_1 x_1 + \lambda'_2 x_2)$$

$0 \leq \lambda_3 \leq 1$ なので、凸関数の定義から

$$\begin{aligned}
F(x_1, x_2, x_3) &\geq f(\lambda_3 x_3 + (1 - \lambda_3)(\lambda'_1 x_1 + \lambda'_2 x_2)) \\
&= f(\lambda_3 x_3 + \lambda_1 x_1 + \lambda_2 x_2) \\
&= f(X_3)
\end{aligned}$$

となり、3 個ある場合でも同じ形の不等式になります。

この結果から、$n-1$ 個では

$$F(x_1,\ldots,x_{n-1}) \geq f(X_{n-1})$$

として成立しているとして、n 個でも成立するのか確かめます (帰納法)。n 個では

$$F(x_1,\ldots,x_n) = \sum_{i=1}^{n}\lambda_i f_i = \lambda_n f_n + (1-\lambda_n)\sum_{i=1}^{n-1}\frac{\lambda_i}{1-\lambda_n}f_i = \lambda_n f_n + (1-\lambda_n)\sum_{i=1}^{n-1}\lambda'_i f_i \quad (\lambda'_i = \frac{\lambda_i}{1-\lambda_n})$$

$$\lambda_1+\lambda_2+\cdots+\lambda_n = 1$$

第 2 項は

$$\sum_{i=1}^{n-1}\lambda'_i = \sum_{i=1}^{n-1}\frac{\lambda_i}{1-\lambda_n} = \frac{\lambda_1+\cdots+\lambda_{n-1}}{1-\lambda_n} = \frac{1-\lambda_n}{1-\lambda_n} = 1$$

なので

$$\sum_{i=1}^{n-1}\lambda'_i f_i = F(x_1,\ldots,x_{n-1}) \geq f(X_{n-1}) = f(\sum_{i=1}^{n-1}\lambda'_i x_i)$$

これを入れて、凸関数の定義を使えば

$$\begin{aligned}F(x_1,\ldots,x_n) &\geq \lambda_n f_n + (1-\lambda_n)f(\sum_{i=1}^{n-1}\lambda'_i x_i)\\ &\geq f(\lambda_n x_n + (1-\lambda_n)\sum_{i=1}^{n-1}\lambda'_i x_i)\\ &= f(\lambda_n x_n + \sum_{i=1}^{n-1}\lambda_i x_i)\\ &= f(X_n)\end{aligned}$$

よって、n 個でも成立しています。というわけで、離散的な $x_1, x_2, \ldots, x_n$ と $\lambda(x_1)+\lambda(x_2)+\cdots+\lambda(x_n)=1$ $(0\leq \lambda_i \leq 1)$ があり、f が凸関数のとき

$$\sum_{i=1}^{n}\lambda(x_i)f(x_i) \geq f(\sum_{i=1}^{n}\lambda_i x_i)$$

という不等式が成立し、これをイェンセンの不等式と言います。

$x\log x$ (底を省いています) が凸関数であることとイェンセンの不等式から、$a_i, b_i > 0$ $(i=1,2,\ldots,n)$ に対して

$$\sum_{i=1}^{n}a_i\log\frac{a_i}{b_i} \geq A\log\frac{A}{B} \quad (A=\sum_{i=1}^{n}a_i\ ,\ B=\sum_{i=1}^{n}b_i)$$

となることが分かります。これは対数和不等式 (log sum inequality) と呼ばれます。確かめるのは簡単です。左辺を

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} = \sum_{i=1}^{n} B \frac{b_i}{B} \frac{a_i}{b_i} \log \frac{a_i}{b_i} = B \sum_{i=1}^{n} \frac{b_i}{B} f(\frac{a_i}{b_i}) \quad (f(\frac{x}{y}) = \frac{x}{y} \log \frac{x}{y}\ ,\ \sum_{i=1}^{n} \frac{b_i}{B} = 1)$$

と変形して、イェンセンの不等式に入れれば

$$B \sum_{i=1}^{n} \frac{b_i}{B} f(\frac{a_i}{b_i}) \geq B f(\sum_{i=1}^{n} \frac{b_i}{B} \frac{a_i}{b_i}) = B f(\frac{A}{B}) = B \frac{A}{B} \log \frac{A}{B} = A \log \frac{A}{B}$$

よって

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq A \log \frac{A}{B}$$

c を定数として、$a_i = cb_i$ のとき

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} = \sum_{i=1}^{n} a_i \log \frac{a_i}{ca_i} = A \log \frac{1}{c}$$

$$A \log \frac{A}{B} = A \log \frac{A}{cA} = A \log \frac{1}{c}$$

となるので、等号になります。