演習問題 1.32

連続変数のベクトル $\mathbf{x}$ を考え、それが分布 $p(\mathbf{x})$ とそれに対応するエントロピー $\mathrm{H}[\mathbf{x}]$ をもつとする。$\mathbf{x}$ に非特異な線形変換を行い、新たな変数 $\mathbf{y} = \mathbf{A}\mathbf{x}$ を得たとする。対応するエントロピーが $\mathrm{H}[\mathbf{y}] = \mathrm{H}[\mathbf{x}] + \ln|\det(\mathbf{A})|$ で与えられることを示せ。ただし、$|\det(\mathbf{A})|$ は、$\mathbf{A}$ の行列式の絶対値である。

[一変数関数の変数変換]

$$\int f(x)\,dx$$

で $x = g(y)$ とすると、$dx = g'(y)\,dy$ より、

$$\int f(g(y))\,g'(y)\,dy$$

[多変数関数の変数変換]

多変数関数の場合は、上記の $g'(y)$ の部分がヤコビ行列の行列式（ヤコビアン）になる。ヤコビアンは、変数変換した際のある点における微小区間の拡大率を意味する。

$x_i = g_i(y_1, \cdots, y_n)\ for\ i = 1, \cdots, n$ とすると、ヤコビアンは、

$$\det \mathbf{J} = \det\left(\frac{\partial x_i}{\partial y_i}\right) = \det\left(\frac{\partial(x_1, \cdots, x_n)}{\partial(y_1, \cdots, y_n)}\right) = \det\begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \cdots & \frac{\partial x_1}{\partial y_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial x_m}{\partial y_1} & \cdots & \frac{\partial x_m}{\partial y_n} \end{pmatrix}$$

よって、多変数関数の変数変換は、適当な条件の下で

$$\int \cdots \int f(x_1, \cdots, x_n)\,dx_1 \cdots dx_n$$

$$= \int \cdots \int f(g_1(y_1, \cdots, y_n), \cdots, g_n(y_1, \cdots, y_n)) \det \mathbf{J}\, dy_1 \cdots dy_n$$

となる。

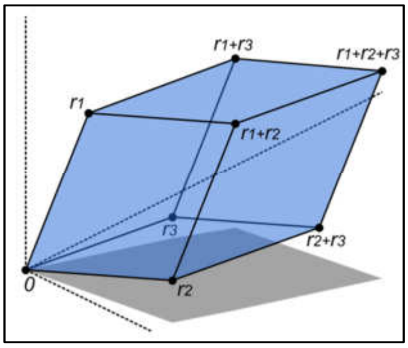
[多変数確率密度関数の変数変換]

$$p_{\mathbf{y}}(\mathbf{y}) = p_{\mathbf{x}}(\mathbf{x}) \left| \det\left(\frac{d\mathbf{x}}{d\mathbf{y}}\right) \right|$$

$$= p_{\mathbf{x}}(g(\mathbf{y}))\,|\det(g'(\mathbf{y}))|$$

$$\cdots (1.27)'$$

ただし、$\det\left(\frac{d\mathbf{x}}{d\mathbf{y}}\right)$ は行列式（ヤコビアン）を意味し、ここでの行列式は、線形変換によって、空間の体積要素が何倍に変わるかという概念を抽象化したものである。なお、行列式は、正方行列に対して定義される量である。



図：平行六面体

この立体の体積 V は、ベクトル r_1, r_2, r_3 の成す 3 × 3 行列の行列式

$$|\mathbf{r}|_\times = |\ r_1\quad r_2\quad r_3\ |_\times$$

$$= \det\begin{pmatrix} 0 & -r_3 & r_2 \\ r_3 & 0 & -r_1 \\ -r_2 & r_1 & 0 \end{pmatrix} = r_1\, r_2\, r_3$$

の絶対値に一致する。⇒　$V = |\ r_1\, r_2\, r_3\ |$

[エントロピー]

$$\mathbf{H}[\mathbf{x}] = -\int p(\mathbf{x}) \ln p(\mathbf{x})\, d\mathbf{x}$$

… (1.104)

[条件付きエントロピー]

$$\mathbf{H}[\mathbf{y}|\mathbf{x}] = -\int\int p(\mathbf{y},\ \mathbf{x}) \ln p(\mathbf{y}|\mathbf{x})\, d\mathbf{y}\, d\mathbf{x}$$

… (1.111)

[相互情報量 (カルバック – ライブラーダイバージェンス)]

$$\mathbf{I}[\mathbf{x},\mathbf{y}] = \mathrm{KL}(p(\mathbf{x},\mathbf{y}) \| p(\mathbf{x})p(\mathbf{y})) = -\int\int p(\mathbf{x},\mathbf{y}) \ln\left(\frac{p(\mathbf{x})p(\mathbf{y})}{p(\mathbf{x},\mathbf{y})}\right) d\mathbf{x}\, d\mathbf{y}$$

… (1.120)

[相互情報量の関係式]

$$\mathbf{I}[\mathbf{x},\mathbf{y}] = \mathbf{H}[\mathbf{x}] - \mathbf{H}[\mathbf{x}|\mathbf{y}] = \mathbf{H}[\mathbf{y}] - \mathbf{H}[\mathbf{y}|\mathbf{x}]$$

… (1.121)

[解]

連続変数のベクトル $\mathbf{x}$ を考え、それが分布 $p(\mathbf{x})$ とそれに対応するエントロピー $\mathrm{H}[\mathbf{x}]$ をもつとする。$\mathbf{x}$ に非特異な線形変換を行い、新たな変数 $\mathbf{y} = \mathbf{A}\mathbf{x}$ を得たとし、その対応するエントロピーが

$$\mathrm{H}[\mathbf{y}] = \mathrm{H}[\mathbf{x}] + \ln|\det(\mathbf{A})|$$

… ※

で与えられることを示す。上記の式の左辺は、式 (1.104) より、

$$式※の左辺 = -\int p(\mathbf{y}) \ln p(\mathbf{y})\, d\mathbf{y}$$

… ※※

となり、ここで式 (1.27)′ より、確率密度 $p(\mathbf{x})$ と確率密度 $p(\mathbf{y})$ の関係は、

$$p(\mathbf{x}) = p(\mathbf{y}) \left| \det\left(\frac{d\mathbf{y}}{d\mathbf{x}} \right) \right|$$

で与えられ、$\mathbf{y} = \mathbf{Ax}$ より、

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \mathbf{A}$$

であることから、

$$p(\mathbf{y}) = \frac{p(\mathbf{x})}{|\det(\mathbf{A})|}$$

式 ※※ は、

$$= -\int \frac{p(\mathbf{x})}{|\det(\mathbf{A})|} \left(\ln \frac{p(\mathbf{x})}{|\det(\mathbf{A})|} \right) |\det(\mathbf{A})| d\mathbf{x}$$

$$= -\int p(\mathbf{x}) \left(\ln \frac{p(\mathbf{x})}{|\det(\mathbf{A})|} \right) d\mathbf{x}$$

$$= -\int p(\mathbf{x}) (\ln p(\mathbf{x}) - \ln|\det(\mathbf{A})|) d\mathbf{x}$$

$$= -\int p(\mathbf{x}) \ln p(\mathbf{x}) d\mathbf{x} + \int p(\mathbf{x}) \ln|\det(\mathbf{A})| d\mathbf{x}$$

$$= \mathrm{H}[\mathbf{x}] + \ln|\det(\mathbf{A})|$$

= 式 ※ の右辺

となる。以上より、式 ※ が成り立つことが示せた。

[別解]

相互情報量 (1.121) の関係式より、

$$\mathbf{H}[\mathbf{x}] - \mathbf{H}[\mathbf{x}|\mathbf{y}] = \mathbf{H}[\mathbf{y}] - \mathbf{H}[\mathbf{y}|\mathbf{x}]$$

$$\mathbf{H}[\mathbf{y}] = \mathbf{H}[\mathbf{x}] - \mathbf{H}[\mathbf{x}|\mathbf{y}] + \mathbf{H}[\mathbf{y}|\mathbf{x}]$$

が得られ、条件付きエントロピーは式 (1.111) より、

$$\mathbf{H}[\mathbf{y}] = \mathbf{H}[\mathbf{x}] + \iint p(\mathbf{x},\mathbf{y}) \ln p(\mathbf{x}|\mathbf{y})\, d\mathbf{x}\, d\mathbf{y} - \iint p(\mathbf{y},\mathbf{x}) \ln p(\mathbf{y}|\mathbf{x})\, d\mathbf{y}\, d\mathbf{x}$$

$$\mathbf{H}[\mathbf{y}] = \mathbf{H}[\mathbf{x}] + \iint p(\mathbf{x},\mathbf{y}) \ln \frac{p(\mathbf{x}|\mathbf{y})}{p(\mathbf{y}|\mathbf{x})} d\mathbf{x}\, d\mathbf{y}$$

… ※※

と書き直せ、さらにベイズの定理より、

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

が成り立つので、式 ※※ は、

$$\mathbf{H}[\mathbf{y}] = \mathbf{H}[\mathbf{x}] + \int\int p(\mathbf{x},\mathbf{y}) \ln \frac{p(\mathbf{y}|\mathbf{x})\,p(\mathbf{x})}{p(\mathbf{y}|\mathbf{x})\,p(\mathbf{y})} d\mathbf{x}\, d\mathbf{y}$$

$$\mathbf{H}[\mathbf{y}] = \mathbf{H}[\mathbf{x}] + \int\int p(\mathbf{x},\mathbf{y}) \ln \frac{p(\mathbf{x})}{p(\mathbf{y})} d\mathbf{x}\, d\mathbf{y}$$

… ※※※

となる。ここで、式 (1.27)′ より、確率密度 $p(\mathbf{x})$ と確率密度 $p(\mathbf{y})$ の関係は、

$$p(\mathbf{x}) = p(\mathbf{y}) \left| \det\left(\frac{d\mathbf{y}}{d\mathbf{x}} \right) \right|$$

で与えられ、$\mathbf{y} = \mathbf{A}\mathbf{x}$ より、

$$\frac{d\mathbf{y}}{d\mathbf{x}} = \mathbf{A}$$

であることから、最終的に式 ※※※ は、

$$\mathbf{H}[\mathbf{y}] = \mathbf{H}[\mathbf{x}] + \int\int p(\mathbf{x},\mathbf{y}) \ln \frac{p(\mathbf{y})|\det(\mathbf{A})|}{p(\mathbf{y})} d\mathbf{x}\, d\mathbf{y}$$

$$\mathbf{H}[\mathbf{y}] = \mathbf{H}[\mathbf{x}] + \int\int p(\mathbf{x},\mathbf{y}) \ln|\det(\mathbf{A})|\, d\mathbf{x}\, d\mathbf{y}$$

$$\mathbf{H}[\mathbf{y}] = \mathbf{H}[\mathbf{x}] + \ln|\det(\mathbf{A})| \int\int p(\mathbf{x},\mathbf{y})\, d\mathbf{x}\, d\mathbf{y}$$

$$\mathbf{H}[\mathbf{y}] = \mathbf{H}[\mathbf{x}] + \ln|\det(\mathbf{A})|$$

となる。よって、題意が示せた。