2022 年 07 月 28 日

多次元確率分布と独立性

新潟工科大学 基礎教育・教養系 竹野茂治

1 はじめに

講義では 2 次元確率分布と、2 つの確率変数の独立性について紹介したが、一般の n 次元確率分布、n 個の確率変数の独立性については詳しくは説明しなかったので、補足としてここにまとめておく。

なお、現代的な公理的確率論ではなく、古典的確率論の範疇で考える。

2 離散確率分布の場合

まずは、離散確率分布の場合を考える。

2.1 多次元確率分布

古典的確率論では、離散確率分布は、確率変数 x、その値の集合である標本空間 $\Omega=\{a_1,a_2,\ldots\}$、および Ω 上の確率関数 $p:\Omega\rightarrow[0,1]$ によって決定する。ここで、Ω は有限集合かまたは可算集合。確率関数 $p(a)$ の値は、$x=a$ である確率、すなわち

$$\mathrm{Prob}\{x=a\}=p(a)$$

を意味するもので、よって $p(x)$ は

$$\sum_{x\in\Omega}p(x)=1 \tag{1}$$

を満たす必要がある。なお、以後 $\mathrm{Prob}\{A\}$ で「A である確率」を表すことにする。

本稿では、この (Ω, p, x) の組を「離散確率分布」と呼ぶことにする。

n 個の離散確率分布 (Ω_j, p_j, x_j) $(j = 1, 2, \ldots, n)$ に対して、

$$r(\alpha_1, \alpha_2, \ldots, \alpha_n) = \mathrm{Prob}\{(x_1, x_2, \ldots, x_n) = (\alpha_1, \alpha_2, \ldots, \alpha_n)\} \tag{2}$$

のように「$x_1 = \alpha_1$ かつ ... かつ $x_n = \alpha_n$」となる確率を考えることができるとき、直積集合

$$\Omega = \Omega_1 \times \ldots \Omega_n = \{(a_1, \ldots, a_n) \mid a_j \in \Omega_j,\ 1 \le j \le n\} \tag{3}$$

を標本空間、$r(x_1, \ldots, x_n)$ を n 変数の確率関数とする n 次元確率変数 $(x_1, \ldots, x_n)$ を考えることができる。この組 $(\Omega, r, (x_1, \ldots, x_n))$ を「n 次元離散確率分布」と呼ぶ。

当然 r は、

$$\sum_{(x_1, \ldots, x_n) \in \Omega} r(x_1, \ldots, x_n) = 1 \tag{4}$$

を満たす必要があるが、(Ω_j, p_j, x_j) との整合性として、すべての j に対して、

$$\mathrm{Prob}\{x_j = \alpha_j\} = \sum_{x_1 \in \Omega_1} \cdots \sum_{x_{j-1} \in \Omega_{j-1}} \sum_{x_{j+1} \in \Omega_{j+1}} \cdots \sum_{x_n \in \Omega_n} r(x_1, \ldots, x_n) = p_j(x_j) \tag{5}$$

も満たす必要がある。

逆に、直積集合 Ω (3) と、(4) を満たす n 変数関数 $r : \Omega \to [0, 1]$ を取り、それに対し (5) の和によって $p_j(x_j)$ という関数を定義すれば、条件 (4) により p_j は (1) の条件を満たすので、n 個の離散確率分布 (Ω_j, p_j, x_j) $(j = 1, 2, \ldots, n)$ が作られ、$(\Omega, r, (x_1, \ldots, x_n))$ をはその n 次元離散確率分布となる。この場合、各離散確率分布 (Ω_j, p_j, x_j) を、$(\Omega, r, (x_1, \ldots, x_n))$ の「周辺分布」と呼ぶことがある。

2.2 独立性

$(\Omega, r, (x_1, \ldots, x_n))$ を n 次元離散分布、(Ω_j, p_j, x_j) $(1 \le j \le n)$ をその周辺分布とする。すべての $(\alpha_1, \ldots, \alpha_n) \in \Omega$ に対し

$$\mathrm{Prob}\{(x_1, \ldots, x_n) = (\alpha_1, \ldots, \alpha_n)\} = \mathrm{Prob}\{x_1 = \alpha_1\} \times \cdots \times \mathrm{Prob}\{x_n = \alpha_n\} \tag{6}$$

が成り立つとき、確率変数 $x_1,\ldots,x_n$ は「独立」であるという。

(6) は、

$$r(\alpha_1,\ldots,\alpha_n)=p_1(\alpha_1)\times\cdots\times p_n(\alpha_n) \tag{7}$$

と表すこともできる。

逆に、n 個の確率分布 (Ω_j,p_j,x_j) に対して、(7) で n 次元確率関数 r を定義すれば、これは当然 (4),(5) を満たすので、ひとつの n 次元確率分布が決定し、その元ではこの確率変数 $x_1,\ldots,x_n$ は独立になる。

つまり、確率変数 $x_1,\ldots,x_n$ の独立性は、n 次元確率分布 (n 次元確率関数) の元で決まるものであり、また n 個の確率変数が独立となるような n 次元確率分布はいつでも作ることができる。

2.3 多変数関数による確率分布

以後、$(x_1,\ldots,x_n)=\vec{x}$ のようなベクトル表記も用いることにする。

n 次元離散分布 $(\Omega,r,\vec{x})$ と、Ω 上の実数値関数 $\phi:\Omega\rightarrow\mathrm{R}$ に対して、ϕ の像を

$$\Lambda=\phi(\Omega)=\{\phi(\vec{x})|\ \vec{x}\in\Omega\} \tag{8}$$

とすると Λ は有限集合か可算集合で、$y\in\Lambda$ に対して

$$q(y)=\sum_{\{\vec{x}|\ \phi(\vec{x})=y\}} r(\vec{x}) \tag{9}$$

によって $\phi(\vec{x})=y$ となる確率を与える関数を定めることができる。ここで、和は $\phi(\vec{x})=y$ となるすべての $\vec{x}$ に対する $r(\vec{x})$ の和を意味する。これにより、$q(y)$ を確率関数とする確率変数 $y=\phi(\vec{x})$、すなわち確率分布 (Λ,q,y) が定まる。

このようなやり方で、例えば $z=xy$ や $y=x_1+x_2+\cdots+x_n$ のような複数の確率変数の式で表される確率変数が作られることになる。

さて、離散確率分布 (Ω,p,x) に対する平均 $E[x]$ は、

$$E[x]=\sum_{x\in\Omega} xp(x) \tag{10}$$

と定義される。多変数関数による確率変数 $y=\phi(\vec{x})$ の、この平均 $E[y]$ の計算について考える。(9) より、

$$E[y]=\sum_{y\in\Lambda} yq(y)=\sum_{y\in\Lambda} y\sum_{\{\vec{x}|\phi(\vec{x})=y\}} r(\vec{x})=\sum_{y\in\Lambda}\sum_{\{\vec{x}|\phi(\vec{x})=y\}}\phi(\vec{x})r(\vec{x})=\sum_{\vec{x}\in\Omega}\phi(\vec{x})r(\vec{x})$$

となり、結局 $E[y]$ の計算を、Λ の代わりに Ω 上の値で

$$E[y]=\sum_{\vec{x}\in\Omega}\phi(\vec{x})r(\vec{x}) \tag{11}$$

のように計算できることになる。この右辺を $E[\phi(\vec{x})]$ のように書く。

3 連続確率分布の場合

次に、連続確率分布の場合を考える。

3.1 多次元確率分布

古典的確率論では、連続確率変数 x の値の集合 Ω は通常実数全体 R で、その確率は、R の部分集合 $A\subset\mathrm{R}$ に対して、x の値が A に含まれる確率 $\mathrm{Prob}\{x\in A\}$ を考え、1 点の値に対する確率 $\mathrm{Prob}\{x=a\}$ は 0 とする。x の確率分布は、分布関数 $F(x)$、またはその導関数である密度関数 $f(x)=F'(x)$ によって決定する。分布関数 (累積分布関数) $F(x)$ は、

$$F(t)=\mathrm{Prob}\{x\leq t\}=\mathrm{Prob}\{x\in(-\infty,t]\} \tag{12}$$

と定義され、1 点の確率が 0 であれば非減少な連続関数となる。

(12) により $a\leq x\leq b$ となる確率は

$$\mathrm{Prob}\{a\leq x\leq b\}=F(b)-F(a) \tag{13}$$

と表される。また、密度関数 $f(x)=F'(x)$ は、(13) より、

$$\mathrm{Prob}\{a\leq x\leq b\}=\int_a^b f(x)dx \tag{14}$$

となり、さらに一般に $A \subset \mathrm{R}$ に対して

$$\mathrm{Prob}\{x \in A\} = \int_A f(x)dx \tag{15}$$

となる。

分布関数 $F(x)$ は、

$$\begin{cases} F(x) \text{ は非減少で連続、} \\ F(-\infty) = \lim_{x\to-\infty} F(x) = 0,\ \ F(\infty) = \lim_{x\to\infty} F(x) = 1 \end{cases} \tag{16}$$

を満たす必要があり、密度関数 $f(x)$ は、

$$f(x) \geq 0,\ \ \int_{\mathrm{R}} f(x)dx = 1 \tag{17}$$

を満たす必要がある。

逆に、(16) を満たす $F(x)$ を取れば、$f(x) = F'(x)$ により密度関数が定まり、それにより (15) で x の確率が求まるので、これで一つの連続確率分布が決定する。

または、(17) を満たす $f(x)$ を取れば、(15) から x の確率が求まるので、一つの連続確率分布が決定し、その分布関数 F も

$$F(t) = \mathrm{Prob}\{x \leq t\} = \int_{-\infty}^{t} f(x)dx \tag{18}$$

により得られる。よって、連続分布を定めるには、分布関数 $F(x)$、密度関数 $f(x)$ のいずれかを設定すればよいので、とりあえず本稿では (R, f, x) の組を「連続確率分布」と呼ぶことにする。

n 個の連続確率分布 (R, f_j, x_j) $(j = 1, 2, \ldots, n)$ において、R の部分集合 $A_j \subset \mathrm{R}$ $(1 \leq j \leq n)$ に対して、

$$\mathrm{Prob}\{(x_1, x_2, \ldots, x_n) \in A_1 \times \cdots \times A_n\} \tag{19}$$

を考えることができるとき、n 次元連続確率分布を構成できる。なお、x_j の分布関数を $F_j(x_j)$ とする。

n 次元確率変数 $\vec{x} = (x_1, \ldots, x_n)$ の分布関数 $F(\vec{x})$ を、

$$F(\vec{t}) = F(t_1, \ldots, t_n) = \mathrm{Prob}\{\vec{x} \in (-\infty, t_1] \times \cdots \times (-\infty, t_n]\} \tag{20}$$

すなわち、$x_1 \leq t_1$ かつ $\ldots x_n \leq t_n$ となる確率によって定義する。

このとき、

$$\begin{aligned}
&\mathrm{Prob}\{\vec{x} \in (a_1, b_1] \times (-\infty, t_2] \times \cdots \times (-\infty, t_n]\} \\
&= \; [F(t_1, t_2, \ldots, t_n)]_{t_1=a_1}^{t_1=b_1} = \int_{a_1}^{b_1} F_{x_1}(\vec{t}) dt_1, \\
&\mathrm{Prob}\{\vec{x} \in (a_1, b_1] \times (a_2, b_2] \times (-\infty, t_3] \times \cdots \times (-\infty, t_n]\} \\
&= \; \left[\int_{a_1}^{b_1} F_{x_1}(\vec{t}) dt_1\right]_{t_2=a_2}^{t_2=b_2} = \int_{a_1}^{b_1} dt_1 \int_{a_2}^{b_2} F_{x_1 x_2}(\vec{t}) dt_2, \\
&\mathrm{Prob}\{\vec{x} \in (a_1, b_1] \times (a_2, b_2] \times (a_3, b_3] \times (-\infty, t_4] \times \cdots \times (-\infty, t_n]\} \\
&= \; \left[\int_{a_1}^{b_1} dt_1 \int_{a_2}^{b_2} F_{x_1 x_2}(\vec{t}) dt_2\right]_{t_3=a_3}^{t_3=b_3} = \int_{a_1}^{b_1} dt_1 \int_{a_2}^{b_2} dt_2 \int_{a_3}^{b_3} F_{x_1 x_2 x_3}(\vec{t}) dt_3
\end{aligned}$$

等となり、これを繰り返すと、

$$\mathrm{Prob}\{\vec{x} \in (a_1, b_1] \times \cdots \times (a_n, b_n]\} = \int_{a_1}^{b_1} dt_1 \cdots \int_{a_n}^{b_n} F_{x_1 \ldots x_n}(\vec{t}) dt_n \tag{21}$$

が得られる。よって、$\vec{x}$ の密度関数 $f(\vec{x})$ をこの分布関数の n 階導関数

$$f(\vec{x}) = F_{x_1 \ldots x_n}(\vec{x}) = \frac{\partial^n}{\partial x_1 \cdots \partial x_n} F(\vec{x}) \tag{22}$$

と定義すれば、

$$\mathrm{Prob}\{\vec{x} \in (a_1, b_1] \times \cdots \times (a_n, b_n]\} = \int_{a_1}^{b_1} dx_1 \cdots \int_{a_n}^{b_n} f(\vec{x}) dx_n \tag{23}$$

となり、より一般に $A \subset \mathrm{R}^n = \mathrm{R} \times \cdots \times \mathrm{R}$ に対して

$$\mathrm{Prob}\{\vec{x} \in A\} = \int_A f(\vec{x}) d\vec{x} \tag{24}$$

となる。

n 次元分布関数 $F(\vec{x})$ は、

$$\begin{cases} F(\vec{x}) \text{ は各 } x_j \text{ に対して非減少で連続,} \\ F(x_1, \ldots, x_{j-1}, -\infty, x_{j+1}, \ldots, x_n) = \lim\limits_{x_j \to -\infty} F(\vec{x}) = 0 \ (1 \leq j \leq n), \\ F(\infty, \ldots, \infty) = \lim\limits_{x_1 \to \infty} \cdots \lim\limits_{x_n \to \infty} F(\vec{x}) = 1 \end{cases} \tag{25}$$

を満たし、n 次元密度関数 $f(\vec{x})$ は

$$f(\vec{x}) \geq 0, \ \int_{\mathrm{R}^n} f(\vec{x}) d\vec{x} = 1 \tag{26}$$

を満たす。

さらに周辺分布として、$F(\vec{x})$ は $j = 1, \ldots, n$ に対して

$$\lim_{t_1 \to \infty} \cdots \lim_{t_{j-1} \to \infty} \lim_{t_{j+1} \to \infty} \cdots \lim_{t_n \to \infty} F(\vec{t}) = \mathrm{Prob}\{x_j \leq t_j\} = F_j(t_j) \tag{27}$$

を満たし、$f(\vec{x})$ は $j = 1, \ldots, n$ に対して

$$\int_{\mathrm{R}} dx_1 \cdots \int_{\mathrm{R}} dx_{j-1} \int_{\mathrm{R}} dx_{j+1} \cdots \int_{\mathrm{R}} f(\vec{x}) dx_n = f_j(x_j) \tag{28}$$

を満たす。この (28) は、この左辺を $\bar{f}(x_j)$ とすると、(23) より

$$\int_{-\infty}^{t} \bar{f}(x_j) dx_j = \mathrm{Prob}\{x_j \leq t\} = F_j(t)$$

となるので、両辺を t で微分すれば $\bar{f}(t) = f_j(t)$ となることから得られる。

逆に (25) を満たす $F(\vec{x})$ によって (22) と (24) から一つの n 次元連続確率分布が決定し、(27) によって周辺分布が決定する。

そしてそれらは F の代わりに (26) を満たす $f(\vec{x})$ によっても決定する。この n 次元連続確率分布を $(\mathrm{R}^n, f, \vec{x})$ と書くことにする。

3.2 独立性

n 次元連続分布 $(\mathrm{R}^n, f, \vec{x})$、およびその周辺分布 (R, f_j, x_j) $(1 \leq j \leq n)$ に対し、

$$\mathrm{Prob}\{\vec{x} \in A_1 \times \cdots \times A_n\} = \mathrm{Prob}\{x_1 \in A_1\} \times \cdots \times \mathrm{Prob}\{x_n \in A_n\} \tag{29}$$

が成り立つとき、確率変数 $x_1, \ldots, x_n$ は「独立」であるという。

(29) で $A_j = (-\infty, t_j]$ とすれば、分布関数の関係式

$$F(\vec{t}) = F_1(t_1) \cdots F_n(t_n) \tag{30}$$

が得られ、またこの (30) の $\vec{t}$ を $\vec{x}$ に置き換えて $x_1, \ldots, x_n$ に関して微分すれば密度関数の関係式

$$f(\vec{x}) = f_1(x_1) \cdots f_n(x_n) \tag{31}$$

が得られる。逆に、(31) が成り立てば、これを $A_1 \times \cdots \times A_n$ で積分することで (29) が得られるので、結局 (29),(30),(31) は同値であることがわかる。

3.3 多変数関数による確率分布

n 次元連続分布 $(\mathrm{R}^n, f, \vec{x})$ に対し、R^n 上の実数値関数 $\phi : \mathrm{R}^n \to \mathrm{R}$ に対して $y = \phi(\vec{x})$ によって連続確率変数 y が決まるかを考える。

y の分布関数 $G(y)$ は、(24) より

$$G(t) = \mathrm{Prob}\{y \le t\} = \mathrm{Prob}\{\phi(\vec{x}) \le t\} = \int_{\{\vec{x}|\ \phi(\vec{x}) \le t\}} f(\vec{x}) d\vec{x} \tag{32}$$

となる。

y が連続分布となるためには、一点の確率 $\mathrm{Prob}\{y = t\}$ は 0、すなわち

$$\mathrm{Prob}\{y = t\} = \mathrm{Prob}\{\phi(\vec{x}) = t\} = \int_{\{\vec{x}|\ \phi(\vec{x}) = t\}} f(\vec{x}) d\vec{x} = 0 \tag{33}$$

となる必要があるが、これはすべての関数 $\phi(\vec{x})$ に対して成り立つわけではない。例えば、$n = 2$ の場合に、ある面積を持った領域の上で $\phi(x_1, x_2)$ が定数 $(= c)$ であれば、$t = c$ のとき (33) の積分領域はその領域を含み、よって (33) の積分値は正となりうる。

つまり、$y = \phi(\vec{x})$ が連続確率変数となるためには、すべての $t \in \mathrm{R}$ に対して (33) が成り立つことが必要条件となる。

逆にそれを満たしていれば、(32) で定まる分布関数 $G(y)$ が条件 (16) を満たすことは、連続性を除いては容易にわかり、連続性についても積分論のやや難しい定理 (ルベーグ収束定理) と (33) から示すことができる。これにより密度関数 $g(y) = G'(y)$ も決定し、連続分布 (R, g, y) が確かに決定することになる。

なお、より細かいことを言えば、$G(y)$ が連続というだけではその微分可能性は得られないが、$G(y)$ は単調なので、「ほとんどの点で」微分可能であることが保証される。

この場合も、離散分布の場合と同様に、平均の計算が $g(y)$ の代わりに $f(\vec{x})$ の方で計算できることを示す。

連続分布 (R, f, x) の平均は、

$$E[x] = \int_{\mathrm{R}} x f(x) dx \tag{34}$$

と定義されるので、よって、(R, g, y) の平均は、

$$E[y] = \int_{\mathrm{R}} y g(y) dy \tag{35}$$

となるが、これが $f(\vec{x})$ による計算

$$E[\phi(\vec{x})] = \int_{\mathrm{R}^n} \phi(\vec{x}) f(\vec{x}) d\vec{x} \tag{36}$$

に一致することを、次の節で詳細に示すが、おおざっぱな「説明」を以下に紹介する。

非常に小さい正数 Δy を取り、積分を

$$E[y] = \int_{\mathrm{R}} y g(y) dy = \sum_{j=-\infty}^{\infty} \int_{j\Delta y}^{(j+1)\Delta y} y g(y) dy$$

と分けると、$j\Delta y < y < (j+1)\Delta y$ では $y \fallingdotseq j\Delta y$ なので (32) より

$$\begin{aligned} E[y] &\fallingdotseq \sum_{j=-\infty}^{\infty} j\Delta y \int_{j\Delta y}^{(j+1)\Delta y} g(y) dy = \sum_{j=-\infty}^{\infty} j\Delta y \{G((j+1)\Delta y) - G(j\Delta y)\} \\ &= \sum_{j=-\infty}^{\infty} j\Delta y \int_{\{\vec{x} | j\Delta y < \phi(\vec{x}) \leq (j+1)\Delta y\}} f(\vec{x}) d\vec{x} \end{aligned}$$

と書ける。ここで、$j\Delta y < \phi(\vec{x}) \le (j+1)\Delta y$ では $j\Delta y \doteqdot \phi(\vec{x})$ なので、

$$E[y] \doteqdot \sum_{j=-\infty}^{\infty} \int_{\{\vec{x}|j\Delta y<\phi(\vec{x})\le(j+1)\Delta y\}} \phi(\vec{x})f(\vec{x})d\vec{x} = \int_{\mathrm{R}^n} \phi(\vec{x})f(\vec{x})d\vec{x}$$

となる。近似をした部分が 2 箇所あるが、それは $\Delta y \to +0$ とすることで等号となる、といった具合である。

3.4 平均に関する命題の証明

前節の平均に関する命題 $E[y] = E[\phi(\vec{x})]$ を示す。まず仮定と結論を明示する。

命題 1

$f : R^n \to \mathrm{R}$ が n 次元密度関数、$\phi : R^n \to \mathrm{R}$ は $\phi(\vec{x})f(\vec{x})$ が R^n 上可積分、すなわち

$$f(\vec{x}) \ge 0, \ \int_{\mathrm{R}^n} f(\vec{x})d\vec{x} = 1, \ \int_{\mathrm{R}} |\phi(\vec{x})|f(\vec{x})d\vec{x} < \infty \tag{37}$$

を満たし、かつすべての $y \in \mathrm{R}$ に対し

$$\int_{\{\vec{x}|\phi(\vec{x})=y\}} f(\vec{x})d\vec{x} = 0 \tag{38}$$

であるとする。また、$y = \phi(\vec{x})$ の分布関数

$$G(y) = \int_{\{|\vec{x}| \ \ \phi(\vec{x})\le y\}} f(\vec{x})d\vec{x} \tag{39}$$

の導関数を $g(y) = G'(y) \ (\ge 0)$ とする。

このとき、$yg(y)$ は R 上可積分、すなわち

$$\int_{\mathrm{R}} |y|g(y)dy < \infty \tag{40}$$

で、

$$\int_{\mathrm{R}} yg(y)dy = \int_{\mathrm{R}^n} \phi(\vec{x})f(\vec{x})d\vec{x} \tag{41}$$

が成り立つ。

証明

正数 Δy を任意に取る。また、$A \subset \mathrm{R}$ に対して

$$\chi_A(x) = \begin{cases} 1 & (x \in A \text{ のとき}) \\ 0 & (x \notin A \text{ のとき}) \end{cases} \tag{42}$$

とする。$y \geq 0$ の範囲を Δy 幅の区間 $A_j = [j\Delta y, (j+1)\Delta y)$ $(j \geq 0)$ に分けると、$y \geq 0$ に対して

$$\sum_{j=0}^{\infty} j\Delta y \chi_{A_j}(y) \leq y \leq \sum_{j=0}^{\infty} (j+1)\Delta y \chi_{A_j}(y) \tag{43}$$

が成り立つので (両辺は Δy 幅の階段関数)、$g(y) \geq 0$ より、この左辺を $g(y)$ 倍して $[0, \infty)$ で積分したものを I_1, 右辺を $g(y)$ 倍して $[0, \infty)$ で積分したものを I_2 とすると

$$I_1 \leq \int_0^{\infty} yg(y)dy \leq I_2 \tag{44}$$

となる。I_1 の積分を分けて変形すると、

$$\begin{aligned} I_1 &= \int_0^{\infty} \sum_{j=0}^{\infty} j\Delta y \chi_{A_j}(y) g(y) dy \; = \; \sum_{k=0}^{\infty} \int_{k\Delta y}^{(k+1)\Delta y} \sum_{j=0}^{\infty} j\Delta y \chi_{A_j}(y) g(y) dy \\ &= \sum_{k=0}^{\infty} \int_{k\Delta y}^{(k+1)\Delta y} k\Delta y g(y) dy \; = \; \sum_{k=0}^{\infty} k\Delta y \{G((k+1)\Delta y) - G(k\Delta y)\} \\ &= \sum_{k=0}^{\infty} k\Delta y \int_{\{\vec{x} | k\Delta y < \phi(\vec{x}) \leq (k+1)\Delta y\}} f(\vec{x}) d\vec{x} \\ &\geq \sum_{k=0}^{\infty} \int_{\{\vec{x} | k\Delta y < \phi(\vec{x}) \leq (k+1)\Delta y\}} (\phi(\vec{x}) - \Delta y) f(\vec{x}) d\vec{x} \end{aligned}$$

$$
\begin{aligned}
&= \int_{\{\vec{x}|\phi(\vec{x})>0\}} (\phi(\vec{x}) - \Delta y) f(\vec{x}) d\vec{x} \\
&= \int_{\{\vec{x}|\phi(\vec{x})>0\}} \phi(\vec{x}) f(\vec{x}) d\vec{x} - \Delta y \int_{\{\vec{x}|\phi(\vec{x})>0\}} f(\vec{x}) d\vec{x}
\end{aligned}
$$

と下から評価される。同様にして、I_2 は

$$
\begin{aligned}
I_2 &= \int_0^\infty \sum_{j=0}^\infty (j+1)\Delta y \chi_{A_j}(y) g(y) dy \\
&= \sum_{k=0}^\infty (k+1)\Delta y \int_{\{\vec{x}|k\Delta y<\phi(\vec{x})\leq(k+1)\Delta y\}} f(\vec{x}) d\vec{x} \\
&\leq \sum_{k=0}^\infty \int_{\{\vec{x}|k\Delta y<\phi(\vec{x})\leq(k+1)\Delta y\}} (\phi(\vec{x}) + \Delta y) f(\vec{x}) d\vec{x} \\
&= \int_{\{\vec{x}|\phi(\vec{x})>0\}} (\phi(\vec{x}) + \Delta y) f(\vec{x}) d\vec{x} \\
&= \int_{\{\vec{x}|\phi(\vec{x})>0\}} \phi(\vec{x}) f(\vec{x}) d\vec{x} + \Delta y \int_{\{\vec{x}|\phi(\vec{x})>0\}} f(\vec{x}) d\vec{x}
\end{aligned}
$$

と上から評価される。よって、(44) より

$$
\begin{aligned}
&\int_{\{\vec{x}|\phi(\vec{x})>0\}} \phi(\vec{x}) f(\vec{x}) d\vec{x} - \Delta y \int_{\{\vec{x}|\phi(\vec{x})>0\}} f(\vec{x}) d\vec{x} \ \leq \ \int_0^\infty y g(y) dy \\
&\quad \leq \ \int_{\{\vec{x}|\phi(\vec{x})>0\}} \phi(\vec{x}) f(\vec{x}) d\vec{x} + \Delta y \int_{\{\vec{x}|\phi(\vec{x})>0\}} f(\vec{x}) d\vec{x} \qquad (45)
\end{aligned}
$$

となる。仮定 (37) よりこの右辺は有限値なので、(44) より

$$
\int_0^\infty y g(y) dy < \infty \qquad (46)
$$

が得られ、また Δy は任意なので、$\Delta y \to +0$ とすれば (45) より

$$
\int_0^\infty y g(y) dy = \int_{\{\vec{x}|\phi(\vec{x})>0\}} \phi(\vec{x}) f(\vec{x}) d\vec{x} \qquad (47)
$$

が得られる。

$y < 0$ に対しても同様に、

$$\sum_{j=-\infty}^{-1} j\Delta y \chi_{A_j}(y) \le y \le \sum_{j=-\infty}^{-1} (j+1)\Delta y \chi_{A_j}(y) \tag{48}$$

となるので、

$$I_3 = \int_{-\infty}^{0} \sum_{j=-\infty}^{-1} j\Delta y \chi_{A_j}(y) g(y) dy, \ \ I_4 = \int_{-\infty}^{0} \sum_{j=-\infty}^{-1} (j+1)\Delta y \chi_{A_j}(y) g(y) dy$$

に対して

$$I_3 \le \int_{-\infty}^{0} y g(y) dy \le I_4 \tag{49}$$

となり、

$$\begin{aligned}
I_3 &= \sum_{k=-\infty}^{-1} \int_{A_k} \sum_{j=-\infty}^{-1} j\Delta y \chi_{A_j}(y) g(y) dy \ = \ \sum_{k=-\infty}^{-1} \int_{A_k} k\Delta y g(y) dy \\
&= \sum_{k=-\infty}^{-1} k\Delta y \{G((k+1)\Delta y) - G(k\Delta y)\} \\
&= \sum_{k=-\infty}^{-1} k\Delta y \int_{\{\vec{x} | k\Delta y < \phi(\vec{x}) \le (k+1)\Delta y\}} f(\vec{x}) d\vec{x} \\
&\ge \sum_{k=-\infty}^{-1} \int_{\{\vec{x} | k\Delta y < \phi(\vec{x}) \le (k+1)\Delta y\}} (\phi(\vec{x}) - \Delta y) f(\vec{x}) d\vec{x} \\
&= \int_{\{\vec{x} | \phi(\vec{x}) \le 0\}} (\phi(\vec{x}) - \Delta y) f(\vec{x}) d\vec{x} \\
&= \int_{\{\vec{x} | \phi(\vec{x}) \le 0\}} \phi(\vec{x}) f(\vec{x}) d\vec{x} - \Delta y \int_{\{\vec{x} | \phi(\vec{x}) \le 0\}} f(\vec{x}) d\vec{x}, \\
I_4 &= \sum_{k=-\infty}^{-1} \int_{A_k} \sum_{j=-\infty}^{-1} (j+1)\Delta y \chi_{A_j}(y) g(y) dy \\
&= \sum_{k=-\infty}^{-1} (k+1)\Delta y \int_{\{\vec{x} | k\Delta y < \phi(\vec{x}) \le (k+1)\Delta y\}} f(\vec{x}) d\vec{x} \\
&\le \sum_{k=-\infty}^{-1} \int_{\{\vec{x} | k\Delta y < \phi(\vec{x}) \le (k+1)\Delta y\}} (\phi(\vec{x}) + \Delta y) f(\vec{x}) d\vec{x} \\
&= \int_{\{\vec{x} | \phi(\vec{x}) \le 0\}} \phi(\vec{x}) f(\vec{x}) d\vec{x} + \Delta y \int_{\{\vec{x} | \phi(\vec{x}) \le 0\}} f(\vec{x}) d\vec{x}
\end{aligned}$$

となるので、$y \geq 0$ の場合同様、仮定 (37) と I_3 の評価より

$$\int_{-infty}^{0} |y|g(y)dy = -\int_{-infty}^{0} yg(y)dy < \infty \tag{50}$$

が得られ、$\Delta y \to +0$ により

$$\int_{-\infty}^{0} yg(y)dy = \int_{\{\vec{x}|\phi(\vec{x})\leq 0\}} \phi(\vec{x})f(\vec{x})d\vec{x} \tag{51}$$

が得られる。

よって、(46),(50) より (40) が、(47), (51) より (41) が得られる。■

なお、この証明と同じ手法により、任意の $t \in \mathrm{R}$ に対して

$$\int_{-\infty}^{t} yg(y)dy = \int_{\{\vec{x}|\phi(\vec{x})\leq t\}} \phi(\vec{x})f(\vec{x})d\vec{x} \tag{52}$$

が成り立つこともわかる。

4 最後に

講義では、3.3 節については $n = 2$ の場合だけプリントで紹介したが、(33) の条件については降れなかった。それは、実は今回これをまとめるにあたり気づいたものなので、講義の説明はやや手落ちであったというべきだろう。なお、この条件 (33) は、$n = 1$ でも必要なものである。

また、3.4 節の命題の証明は講義では説明していなかったが、手近な確率の教科書をいくつか見てみたがそれに触れている本は多くはなく、[1], [2] はむしろ $E[\phi(\vec{x})]$ の方を平均の定義としている。[3] には証明が書いてあるのだが、その本は公理論的確率論の範疇で書いているので、本稿のような証明ではない。

よって、本稿の記述はあまり目にすることはないようなので、それなりに意味があるのではないかと思う。

参考文献

[1] 服部哲也「理工系の確率・統計入門 (第 4 版)」、学術図書出版社 (2019)

[2] E. クライツィグ (田栗正章訳) 「技術者のための高等数学 7 確率と統計 (原書第 8 版)」、培風館 (2004)

[3] 西尾真喜子「確率論」、実教出版 (1978)