ZESZYTY NAUKOWE POLITECHNIKI ŚLĄSKIEJ 1978

eria: Matematyka-Fizyka z. 29 Nr kol. 559

Damian BRUCKNER, Marian PODHORODYŃSKI
nstytut Matematyki - Uniwersytet Śląski

ogdan SKALMIERSKI
nstytut Mechaniki Teoretycznej - Politechnika Śląska

DSTAWY RACHUNKU PRAWDOPODOBIEŃSTWA W UJĘCIU DYSTRYBUCYJNYM

Streszczenie. W pracy przedstawiono możliwość zbudowania rachunku prawdopodobieństwa odstępując od tradycyjnego systemu Kołmogorowa. Autorzy chcą znaleźć wytłumaczenie pewnych faktów tej teorii na gruncie teorii dystrybucji. Ta dziedzina wiedzy matematycznej pozwala je stosunkowo łatwo spostrzegać i dzięki temu okazuje się istotna w zastosowaniu.

Wstęp

W pracy przedstawiono możliwość zbudowania rachunku prawdopodobieństwa odstępując od tradycyjnego systemu probabilistycznego (Kołmogorowa), opartego na pojęciach przestrzeni zdarzeń elementarnych Ω, σ-ciała $\mathcal{U}$, miary unormowanej P oraz zmiennej losowej - funkcji $\mathcal{U}$-mierzalnej.

Nie znaczy to jednak, że wyżej wymienione pojęcia zostaną tu odrzucone - pojawią się traktowane inaczej, nie jako pojęcia wyjściowe. Pokazana też będzie równoważność (w pewnym sensie) obu dróg dojścia do systemu probabilistycznego.

Przedstawiony w pracy sposób konstrukcji teorii prawdopodobieństwa nie jest nowy i, jak się wydaje, należy do Stratonowicza. W jego książce [5] ten aspekt jest wyraźnie poruszony.

Autorzy niniejszej pracy chcą znaleźć wytłumaczenie pewnych faktów tak potraktowanej tematyki na gruncie teorii dystrybucji. Ta dziedzina wiedzy matematycznej pozwala je stosunkowo łatwo spostrzegać. Dzięki temu i jej intuicyjnemu charakterowi okazuje się istotna w zastosowaniu.

Dlatego też wydaje się nam konieczne na początku artykułu zebrać podstawowe definicje i twierdzenia tej teorii [3], zwłaszcza te, które będą wykorzystywane dalej.

1. Podstawowe wiadomości z teorii dystrybucji [3]

Przyjmujemy następujące oznaczenia i umowy:

- O - podzbiór otwarty przestrzeni R^N, N - liczba naturalna
- I - przedział otwarty i ograniczony w R^N
- I - wewnątrz O, o ile $\mathrm{cl}I \subset O$, gdzie oznacza domknięcie przedziału I
- $x = (x^1,\dots,x^N)$, $x^i \in R$, $i = 1,\dots,N$, $\|x\| = \sqrt{(x^1)^2+\dots+(x^N)^2}$
- $k = (k_1,\dots,k_N)$, $0 \leq k_i \in Z$, $i = 1,\dots,N$, $|k| = k_1 + \dots + k_N$
- $\varphi^{(k)}(x) = \dfrac{\partial^{|k|}\varphi(x^1,\dots,x^N)}{\partial^{k_1}x^1 \dots \partial^{k_N}x^N}$, $\int\limits_{x_0}^{x} \varphi(t)dt^k$ - całka wielokrotna
- $\varphi(x) : R^N \to C$ gładka w O, o ile $\varphi^{(k)}(x)$ istnieje w O dla dowolnego rzędu k, C - zbiór liczb zespolonych
- $\varphi_n(x) \underset{I}{\Longrightarrow}$ - ciąg $\varphi_n(x)$ jest jednostajnie zbieżny w I.

Ciąg funkcji gładkich $\varphi_n(x)$ nazywamy ciągiem podstawowym w O, jeżeli dla każdego przedziału I wewnątrz O istnieje rząd k i ciąg funkcji gładkich $\Phi_n(x)$, dla których zachodzi:

$$\Phi_n^{(k)}(x) = \varphi_n(x), \quad x \in I \tag{1.1}$$

$$\Phi_n(x) \underset{I}{\Longrightarrow} \tag{1.2}$$

Ciągi podstawowe w O $\varphi_n(x)$, $\psi_n(x)$ nazywamy ciągami równoważnymi w O ($\varphi_n(x) \sim_O \psi_n(x)$), jeżeli ciąg przeplatany $\varphi_1(x)$, $\psi_1(x)$, $\varphi_2(x)$, $\psi_2(x)$,... - jest podstawowy w O.

Relacja $\sim_O$ okazuje się relacją równoważności, a zatem dzieli rodzinę wszystkich ciągów podstawowych w O na rozłączne klasy abstrakcji, które oznaczamy symbolem $[\varphi_n(x)]$, gdzie $\varphi_n(x)$ jest reprezentantem danej klasy.

Każdą klasę abstrakcji tej relacji nazywamy N-wymiarową dystrybucją, co zapisujemy:

$$f(x) = [\varphi_n(x)], \tag{1.3}$$

używając tego samego oznaczenia dla dystrybucji jak i dla funkcji, co uzasadnią *dalej podane fakty. Widoczna jest analogia między tą definicją a* definicją Cantora liczby rzeczywistej, polegającej na identyfikacji równoważnych ciągów podstawowych liczb wymiernych. Inaczej tylko określona jest podstawowość i równoważność ciągów.

Operację A określoną na m-krotnym produkcie zbioru funkcji N-zmiennych i przyjmującą wartości w zbiorze funkcji L-zmiennych nazywamy regularną, jeżeli zachodzi implikacja:

Jeżeli $\varphi_n^i(x)$ podstawowy w 0, $i=1,\dots,m$, to $A(\varphi_n^1(x),\dots,\varphi_n^m(x))$ podstawowy w 0. (1.4)

Operacja regularna ma następującą własność:

Jeżeli $\varphi_n^i(x) \sim_0 \psi_n^i(x)$, $i=1,\dots,m$; to $A(\varphi_n^1(x),\dots,\varphi_n^m(x)) \sim_0 A(\psi_n^1(x),\dots,$ $\dots,\psi_n^m(x))$, (1.5)

co łatwo sprawdzić, a zatem da się rozszerzyć na dystrybucje za pomocą wzoru:

$$A([\varphi_n^1(x)],\dots,[\varphi_n^m(x)]) = [A(\varphi_n^1(x),\dots,\varphi_n^m(x))]. \tag{1.6}$$

Podstawowość ciągu $A(\varphi_n^1(x),\dots,\varphi_n^m(x))$ wynika z własności (1.4), a niezależność od wyboru reprezentantów z warunku (1.5).

Sprawdzenie (1.4) dla odpowiednich operacji pozwala rozszerzyć je na dystrybucje za pomocą wzorów:

$$[\varphi_n(x)] + \lambda[\psi_n(x)] = [\varphi_n(x) + \lambda\psi_n(x)], \quad \lambda \in R \tag{1.7}$$

$$f(x+h) = [\varphi_n(x+h)], \; h \in R^N, \; f(x) = [\varphi_n(x)] \tag{1.8}$$

$$[\varphi_n(x)]^{(l)} = [\varphi_n^{(l)}(x)], \; l=(l_1,\dots,l_N), \; 0 \leqslant l_i \in Z, \; i=1,\dots,N \tag{1.9}$$

$$w(x)[\varphi_n(x)] = [w(x)\varphi_n(x)], \quad w(x) \text{ - funkcja gładka} \tag{1.7a}$$

$f(\sigma(x)) = [\varphi_n(\sigma(x))]$, $f(y) = [\varphi_n(y)]$, $y \in R^L$, $\sigma(x)$: $R^N \to R^L$ gładka, $L \leqslant N$, w każdym punkcie x macierz pochodnej $\sigma'(x)$ jest rzędu L (1.10)

$$[\varphi_n(x)]\,[\psi_n(y)] = [\varphi_n(x)\psi_n(y)] \tag{1.11}$$

$[\varphi_n(x)] \cdot w(x) = [\varphi_n(x) \cdot w(x)]$, $w(x)$ - funkcja gładka znikająca poza pewnym przedziałem. (1.12)

Następująca reguła: "Jeżeli obie strony pewnej równości są skończonymi złożeniami operacji regularnych i równość ta zachodzi dla wszystkich funkcji gładkich, to jest ona prawdziwa dla wszystkich dystrybucji" - ma duże znaczenie praktyczne - pozwala wykonywać rachunki, w których występują tylko operacje regularne na dystrybucjach tak samo jak na funkcjach gładkich. *Słuszność jej wynika z faktu*, iż skończone złożenie operacji regularnych jest operacją regularną.

Ciąg funkcji gładkich $\delta_n(x)$ nazywamy ciągiem deltowym, jeżeli:

1° $\delta_n(x) \geqslant 0$, dla $\|x\| < \alpha_n$ oraz $\delta_n(x) = 0$, dla $\|x\| \geqslant \alpha_n$, gdzie $0 < \alpha_n \to 0$;

$$\int_{-\infty}^{+\infty} \delta_n(x)\,dx = 1. \tag{1.13}$$

Każdy ciąg deltowy jest podstawowy w R^N oraz każde dwa ciągi deltowe są równoważne w R^N, bo utworzony z nich ciąg przeplatany jest deltowy. Tak więc ciąg deltowy wyznacza N-wymiarową dystrybucję

$$\delta(x) = [\delta_n(x)], \tag{1.14}$$

zwaną dystrybucją delta - Diraca.

Wzór

$$w(x)\delta(x-a) = w(a)\delta(x-a) \tag{1.15}$$

jest prawdziwy dla funkcji w(x) - gładkich. Można go jednak uogólnić na dowolne funkcje l(x), definiując:

$$l(x)\delta(x-a) = l(a)\delta(x-a), \tag{1.15a}$$

gdyż operacja mnożenia przez liczbę jest regularna.

Funkcję gładką $\varphi(x)$ utożsamiamy z dystrybucją $[\varphi(x)]$ generowaną przez ciąg stały $\varphi(x)$. Piszemy zatem:

$$\varphi(x) = [\varphi(x)]. \tag{1.16}$$

Funkcję ciągłą f(x) utożsamiamy z dystrybucją $f(x) * \delta_n(x)$ generowaną przez ciąg funkcji gładkich niemal jednostajnie zbieżny do f(x). Piszemy zatem:

$$f(x) = [f(x) * \delta_n(x)]. \tag{1.17}$$

Klasę funkcji lokalnie całkowalnych równych prawie wszędzie f(x) utożsamiany z dystrybucją będącą pochodną dystrybucyjną funkcji ciągłej $\int_{x_0}^{x} f(t)dt$. Piszemy zatem:

$$f(x) = \left(\int_{x_0}^{x} f(t)dt\right)' \tag{1.18}$$

Tu znajdujemy uzasadnienie używania tych samych symboli dla dystrybucji i funkcji.

Wszystkie zdefiniowane we wzorach (1.7)-(1.12) operacje, za wyjątkiem różniczkowania (1.9), wykonywane na funkcjach ciągłych (lokalnie całkowalnych), traktowanych jako dystrybucje, są zgodne ze zwykłymi operacjami na funkcjach ciągłych (lokalnie całkowalnych). W przypadku funkcji gładkich wniosek ten dotyczy także różniczkowania.

Każda dystrybucja określona w O jest w każdym przedziale wewnątrz O pochodną dystrybucyjną pewnego rzędu pewnej funkcji ciągłej.

Ciąg dystrybucji $f_n(x)$ nazywamy ciągiem zbieżnym dystrybucyjnie w 0 do dystrybucji $f(x)$, gdy $n \to \infty$, jeżeli dystrybucja $f(x)$ jest określona w 0 i jeżeli dla każdego przedziału I wewnątrz 0 istnieje rząd k oraz funkcje ciągłe $F_n(x)$, $n \in \mathcal{N}$ i $F(x)$, dla których zachodzi:

$$F_n^{(k)}(x) = f_n(x),\ n > n_0,\quad F^{(k)}(x) = f(x);\ x \in I \tag{1.19}$$

$$F_n(x) \underset{I}{\rightrightarrows} F(x). \tag{1.20}$$

Piszemy wtedy:

$$f_n(x) \underset{0}{\longrightarrow} f(x) \text{ bądź } \lim_{n\to\infty} f_n(x) = f(x) \text{ w } 0. \tag{1.21}$$

Dużą rolę odgrywają następujące, wynikające wprost z powyższej definicji, wnioski:

(i) Jeżeli ciąg funkcji ciągłych jest zbieżny niemal jednostajnie, to jest on także zbieżny dystrybucyjnie do tej samej granicy.

(ii) Jeżeli ciąg funkcji lokalnie całkowalnych jest zbieżny prawie wszędzie i jest ograniczony przez funkcję lokalnie całkowalną, to jest zbieżny dystrybucyjnie do tej samej granicy.

(iii) Ciąg funkcji gładkich jest dystrybucyjnie zbieżny do dystrybucji wtedy i tylko wtedy, gdy jest dla niej ciągiem podstawowym.

Ponadto zachodzą istotne związki:

$$\lim_{n\to\infty} (f_n(x) + \lambda g_n(x)) = \lim_{n\to\infty} f_n(x) + \lambda \lim_{n\to\infty} g_n(x) \tag{1.22}$$

$$\lim_{n\to\infty} f_n^{(m)}(x) = \left\{\lim_{n\to\infty}(f_n(x))\right\}^{(m)} \tag{1.23}$$

$w_n(x) \cdot f_n(x) \longrightarrow w_0(x) \cdot f(x)$, o ile $f_n(x) \longrightarrow f(x)$, $w_i(x)$ gładkie dla $i=0,1,2,\ldots$, $w_n^{(m)}(x) \rightrightarrows w_0^{(m)}(x)$ dla dowolnego rzędu m

(1.24)

$f_n(\theta_n(x)) \longrightarrow f(\theta_0(x))$, o ile $f_n(x) \longrightarrow f(x)$, $\theta_i(x)$ posiadają własności funkcji $\theta(x)$ ze wzoru (1.10). (1.25)

$$\lim_{n\to\infty} (f_n(x) \cdot g_n(y)) = \lim_{n\to\infty} f_n(x) \cdot \lim_{n\to\infty} g_n(y) \tag{1.26}$$

$\lim_{n\to\infty} (f_n(x) * w(x)) = (\lim_{n\to\infty} f_n(x)) * w(x)$, $w(x)$ - funkcja gładka, znikająca poza pewnym przedziałem. (1.27)

Bardzo pożyteczne do definiowania nowych dystrybucji (np.: $\delta(x^2 - a^2)$) jest następujące twierdzenie o sklejaniu dystrybucji:
Niech $O = \bigcup_{\lambda\in\Lambda} O_\lambda$ i niech $\{f_\lambda(x)\}_{\lambda\in\Lambda}$ będzie rodziną dystrybucji określonych odpowiednio na zbiorach O_λ, $\lambda\in\Lambda$, o własności $f_\lambda(x) = f_\mu(x)$ $x \in O_\lambda \cap O_\mu$. Wówczas istnieje dystrybucja $f(x)$ określona na O, mająca własność $f(x)=f_\lambda(x)$, $x \in O_\lambda$, $\lambda\in\Lambda$.

Dystrybucja $f_\alpha(x)$ (zależna od parametru rzeczywistego α) jest zbieżna dystrybucyjnie do dystrybucji $f(x)$ w O, gdy $\alpha \to \alpha_0$, jeżeli dla dowolnego ciągu $\alpha_n \to \alpha_0$, $n \to \infty$ zachodzi:

$$f_{\alpha_n}(x) \xrightarrow[O]{} f(x).$$

Piszemy wtedy:

$$f_\alpha(x) \xrightarrow[O]{} f(x), \alpha\to\alpha_0, \text{ bądź } \lim_{\alpha\to\alpha_0} f_\alpha(x)=f(x) \text{ w } O. \tag{1.28}$$

Podobnie jak w przypadku ciągów dystrybucji granica $\lim_{\alpha\to\alpha_0}$ jest przemienna z wszystkimi, zdefiniowanymi we wzorach (1.7) - (1.12) operacjami.

Przechodzimy do dystrybucji jednowymiarowych.

Punkt $x_0 \in O$ nazywamy punktem regularnym dystrybucji, $f(x)$ jeżeli istnieje:

$$\lim_{\alpha\to 0} f(\alpha x + x_0) = g(x). \tag{1.29}$$

Istnienie granicy $g(x)$ gwarantuje, że jest ona funkcją stałą.

Wartością dystrybucji $f(x)$ w punkcie regularnym x_0 nazywamy stałą wartość funkcji $g(x)$.

Zauważmy, że dystrybucja $\delta(x)$ ma wartość zero wszędzie, za wyjątkiem punktu $x = 0$. Jest ona pochodną dystrybucyjną dystrybucji (funkcji lokalnie całkowalnej Heaviside'a) $H(x)$, która ma wartość wszędzie poza zerem.

$$H(x) = \begin{cases} 0 & \text{dla } x<0 \\ 1 & \text{dla } x>0 \end{cases}$$

Istnienie granicy

$$\lim_{\beta\to\infty} f(x + \beta) = h(x) \tag{1.30}$$

gwarantuje, że jest ona funkcją stałą. Tak więc można przyjąć następującą definicję:

Wartością dystrybucji $f(x)$ w punkcie niewłaściwym $+\infty$ nazywamy stałą wartość funkcji $h(x)$. Podobnie definiuje się wartość dystrybucji w punkcie $-\infty$.

Całką nieoznaczoną dystrybucji $f(x)$, określonej w I, nazywamy każdą dystrybucję $\psi(x)$ określoną w I o własności $\psi'(x) = f(x)$ dla $x \in I$.

Całka nieoznaczona istnieje dla dowolnej dystrybucji i jest wyznaczona z dokładnością do stałej addytywnej. Tak więc dystrybucja

$$\int_a^b f(x+t)dt = \psi(x+b) - \psi(x+a), \quad a, b \in I \tag{1.31}$$

nie zależy od wyboru konkretnej całk. nieoznaczonej $\psi(x)$.

Całką oznaczoną dystrybucji $f(x)$ na przedziale $[a, b]$, gdzie a, b punkty regularne $f(x)$, nazywamy wartość dystrybucji $\int_a^b f(x+t)dt$ w punkcie $x = 0$. Oznaczamy ją symbolem $\int_a^b f(t)dt$.

Zauważmy, że zachodzi analogiczny do klasycznego związek pomiędzy całką oznaczoną a nieoznaczoną:

$$\int_a^b f(t)\,dt = \psi(b) - \psi(a). \tag{1.32}$$

Wzór (1.32) pomocny jest także dla definiowania całki oznaczonej w przypadku, gdy a lub b jest nieskończone oraz gdy istnieje wartość dystrybucji $f(x)$ w punkcie $+\infty$ bądź w punkcie $-\infty$. Mianowicie:

$$\int_a^{\infty} f(t)dt = \psi(\infty) - \psi(a)$$

$$\int_{-\infty}^{b} f(t)dt = \psi(b) - \psi(-\infty) \tag{1.32a}$$

$$\int_{-\infty}^{\infty} f(t)dt = \psi(\infty) - \psi(-\infty).$$

Słuszny jest także wzór na całkowanie przez części:

$$\int_a^b w(t)\,f'(t)dt + \int_a^b w'(t)\,f(t)dt = w(b)f(b) - w(a)f(a), \tag{1.33}$$

przy czym $w(x)$ jest funkcją gładką. Stąd i z (1.15) łatwo otrzymuje się wzory:

$$\int_a^b 1(t)\,\delta(t - x_0)dt = 1(x_0), \quad a < x_0 < b \tag{1.34}$$

$$\int_a^b 1(t)\delta(t - x_0)dt = 0, \quad x_0 < a < b \quad \text{lub} \quad a < b < x_0. \tag{1.35}$$

Wzór

$$\int_{\sigma(a)}^{\sigma(b)} f(t)dt = \int_a^b f(\sigma(t))\, \sigma'(t)dt \tag{1.36}$$

jest prawdziwy pod warunkiem, że całka nieoznaczone z $f(x)$ ma wartości w punktach $\sigma(a)$ i $\sigma(b)$ oraz, że $\sigma(x)$ jest funkcją jednej zmiennej spełniającą założenia ze wzoru (1.10).

Ponadto dla całek nieoznaczonych zachodzi implikacja:

Jeżeli $\quad f_n(x) \longrightarrow f(x)$, to $\int_a^b f_n(x+t)dt \longrightarrow \int_a^b f(x+t)dt.$ $\quad$ (1.37)

2. Podstawowe pojęcia rachunku prawdopodobieństwa

Idąc za myślą Stratonowicza [5], rozwiniętą w pracy [4], przedstawimy inne od tradycyjnego ujęcie podstaw rachunku prawdopodobieństwa. W ujęciu tym pojęciem wyjściowym jest zmienna losowa rozumiana inaczej niż zwykle [1], [2] i [6].

Rodzinę ξ ciągów liczbowych $\xi_1^\alpha, \xi_2^\alpha, \ldots, \xi_i^\alpha, \ldots$; $\alpha \in \Lambda$ takich, że dla każdego $x \in \bar{R} = R \cup \{-\infty\} \cup \{\infty\}$ granica

$$\langle H(x - \xi) \rangle \overset{df}{=} \lim_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} H(x - \xi_i^\alpha), \tag{2.1}$$

gdzie:

$$H(x) = \begin{cases} 0, & x \leq 0 \\ 1, & x > 0 \end{cases} \tag{2.2}$$

istnieje i nie zależy od wyboru ciągu ξ_i^α, $i \in \mathcal{N}$; (nie zależy od $\alpha \in \Lambda$), nazywamy zmienną losową.

Interpretując powyższą definicję przypuśćmy, że każdemu wynikowi doświadczenia przypisana jest pewna liczba. W zależności od przypadku możemy otrzymać różne wyniki doświadczeń, tzn. różne liczby. Przypuśćmy dalej, że w identycznych warunkach wykonamy n doświadczeń; otrzymamy wówczas n liczb $\xi_1, \ldots, \xi_n$. Gdyby liczba doświadczeń wzrastała nieograniczenie, otrzymalibyśmy nieskończony ciąg liczb $\xi_1, \ldots, \xi_i, \ldots$ - jest to jeden z ciągów w powyższej definicji zmiennej losowej.

Zauważmy jeszcze, że ze względu na postulat niezależności granicy (2.1) od $\alpha \in \Lambda$ wystarcza znajomość tylko jednego ciągu.

Jest rzeczą widoczną, że:

jeśli $$x_1 < x_2, \text{ to } \langle H(x_1 - \xi)\rangle \leq \langle H(x_2 - \xi)\rangle \tag{2.3}$$

oraz, że:

$$\langle H(-\infty - \xi)\rangle = 0, \quad \langle H(+\infty - \xi)\rangle = 1. \tag{2.4}$$

Ponadto z (2.3) wynika, że $\langle H(x - \xi)\rangle$ ma co najwyżej przeliczalną ilość punktów nieciągłości, co oznacza, że prawie wszędzie równa jest pewnej funkcji niemalejącej, lewostronnie ciągłej $F_\xi(x)$ o własnościach:

$$F_\xi(-\infty) = 0, \quad F_\xi(+\infty) = 1. \tag{2.4a}$$

W rezultacie zapiszemy:

$$F_\xi(x) = \langle H(x - \xi)\rangle \text{ prawie wszędzie.} \tag{2.5}$$

Funkcję $F_\xi(x)$ nazywamy dystrybuantą zmiennej losowej ξ.

Przypuśćmy teraz, że wykonujemy doświadczenie, w wyniku którego może zajść zdarzenie A. Wykonując je n-krotnie stwierdzimy, że zdarzenie A zaszło n_A razy. Stosunek $\frac{n_A}{n}$ nazywamy frekwencją. U podstaw rachunku prawdopodobieństwa leży obserwowalna doświadczalnie prawidłowość polegająca na tym, że frekwencja interesującego nas zdarzenia, w serii doświadczeń prowadzonych w identycznych warunkach, waha się wokół pewnej wartości. Tak więc przez prawdopodobieństwo zdarzenia A rozumie się liczbę $P(A) = \lim_{n\to\infty} \frac{n_A}{n}$.

W szczególności jeżeli zdarzenie A polega na tym, że w wyniku doświadczeń opisanych zmienną losową ξ otrzymujemy liczby z przedziału (a, b), to

$$P_\xi([a, b)) = \lim_{n\to\infty} \frac{n_{[a,b)}}{n}, \tag{2.7}$$

gdzie:

$$n_{[a,b)} = \sum_{i=1}^{n} (H(b - \xi_i) - H(a - \xi_i)). \tag{2.8}$$

Ze wzorów (2.7), (2.8) i (2.1) otrzymujemy:

$$P_\xi([a, b)) = \langle H(b - \xi)\rangle - \langle H(a - \xi)\rangle \tag{2.9}$$

Stąd i z (2.5)

$$P_\xi([a, b)) = F_\xi(b) - F_\xi(a), \tag{2.10}$$

gdzie a, b są punktami ciągłości $F_\xi(x)$.

Wzór (2.10) upoważnia nas do przyjęcia go za definicję w przypadku ogólnym (a, b - dowolne), tym bardziej, że $\langle H(x-\xi)\rangle = F_\xi(x)$ wszędzie, co uzasadnimy dalej na innej drodze,

$P_\xi([a, b))$ jest miarą przeliczalnie addytywną i unormowaną na ciele, złożonym ze skończonych sum przedziałów lewostronnie domkniętych - prawostronnie otwartych. Miarę tę można zatem rozszerzyć w sposób jednoznaczny do miary przeliczalnie addytywnej, unormowanej określonej na σ-ciele zbiorów borelowskich [1], [6]. Tak otrzymaną miarę nazywamy prawdopodobieństwem zmiennej losowej ξ, a zbiory borelowskie zdarzeniami. Zamiast $P_\xi(B)$ piszemy także $P(\xi \in B)$, gdzie B oznacza zbiór borelowski.

Ciąg funkcji lokalnie całkowalnych

$$S_n(x) = \frac{1}{n} \sum_{i=1}^{n} H(x - {}_i\xi_\alpha) \tag{2.11}$$

jest zbieżny do $F_\xi(x)$ oraz ograniczony przez funkcję lokalnie całkowalną (np. identycznie równą 1). Tak więc na podstawie twierdzenia (ii) i części pierwszej ciąg $S_n(x)$ jest zbieżny dystrybucyjnie do $F_\xi(x)$. Korzystając z (1.23) i (2.11) otrzymujemy:

$$\frac{d\,F_\xi(x)}{dx} = \lim_{n\to\infty} \frac{d\,S_n(x)}{dx} = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta(x - {}_i\xi_\alpha). \tag{2.12}$$

Oznaczając

$$\langle \delta(x-\xi)\rangle = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta(x - {}_i\xi_\alpha) \text{ (granica dystrybucyjna)} \tag{2.13}$$

oraz:

$$g_\xi(x) = \frac{d\,F_\xi(x)}{dx} \text{ (pochodna dystrybucyjna)} \tag{2.14}$$

otrzymujemy:

$$g_\xi(x) = \langle \delta(x-\xi)\rangle . \tag{2.15}$$

którą nazywamy gęstością rozkładu prawdopodobieństwa zmiennej losowej ξ.

Zmienna losowa jest wyznaczona przez swoją gęstość bądź dystrybuantę.

W celu porównania wyżej zdefiniowanej zmiennej losowej z klasyczną, wykorzystamy twierdzenie Kołmogorowa - Chińczyna z teorii rachunku prawdopodobieństwa, które głosi, że ciąg $X_1, X_2, \ldots$ niezależnych zmiennych losowych (funkcji $\mathcal{U}$ mierzalnych) o tych samych rozkładach i równych wartościach oczekiwanych $EX_n = m$ posiada własność:

$$\frac{1}{n} \sum_{i=1}^{n} X_i \longrightarrow m \tag{a}$$

prawie wszędzie względem miary P, określonej na σ-ciele $\mathcal{U}$ podzbiorów przestrzeni probabilistycznej Ω.

Niech dana będzie dystrybuanta $F(x)$ oraz ciąg X_n, $n \in \mathcal{N}$, niezależnych zmiennych losowych o tych samych rozkładach wyznaczonych przez $F(x)$. Ciąg

$$Y_n(x) = H(x - X_n), \tag{b}$$

gdzie $H(x)$, określona we wzorze (2.2), spełnia założenia (przy każdym ustalonym x) wspomnianego twierdzenia Kołmogorowa - Chińczyna. Jest mianowicie ciągiem niezależnych zmiennych losowych o tych samych rozkładach i wartościach oczekiwanych

$$E\, Y_n(x) = F(x). \tag{c}$$

Zatem na podstawie (a) dla każdego ustalonego x zachodzi:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} H(x - X_n(\omega)) = F(x) \tag{d}$$

prawie wszędzie względem miary P. Funkcja $F(x)$ jako lewostronnie ciągła jest wyznaczona jednoznacznie przez swoje wartości na zbiorze przeliczalnym i gęstym $\{x_k : k \in \mathcal{N}\}$. Granica (d) istnieje dla każdego x_k, $k \in \mathcal{N}$, i wszystkich $\omega \in \Omega \setminus B_{x_k}$, gdzie $P(B_k) = 0$. Oznaczając $B = \bigcup_{k \in \mathcal{N}} B_{x_k}$ otrzymujemy, że granica (d) istnieje dla każdego x będącego punktem ciągłości $F(x)$ i dla wszystkich $\omega \in \Omega \setminus B$, przy czym $P(B) = 0$.

Ciąg dystrybuant $F_n(x) = \frac{1}{n} \sum_{i=1}^{n} H(x - X_i(\omega))$ jest zbieżny na zbiorze gęstym $\{x_k : k \in \mathcal{N}\}$ do $F(x)$, a zatem jest zbieżny także w punktach ciągłości $F(x)$.

Oznaczając przez $\{y_k : k \in \mathcal{N}\}$ zbiór wszystkich punktów nieciągłości oraz przez B_{y_k} zbiór miary P-zero tych $\omega \in \Omega$, dla których zbieżność (d) w punkcie $x = y_k$ nie zachodzi, zauważamy, że zbieżność ta ma miejsce dla wszystkich x oraz $\omega \in \Omega \setminus B$, gdzie $B = B_1 \cup \bigcup_{k \in \mathcal{N}} B_{y_k}$ i $P(B) = 0$.

Ustalając $\omega \in \Omega \setminus B$ otrzymujemy ${}_i\xi_\omega = X_i(\omega)$. Rodzina ciągów ${}_1\xi_\omega, {}_2\xi_\omega, \ldots, {}_i\xi_\omega$ $\omega \in \Omega$ B, jest zmienną losową w sensie przyjętej przez nas definicji.

Często zdarza się, że wynikowi doświadczenia przypisana jest nie jedna liczba, lecz ich układ. Dlatego przyjmujemy następującą definicję (analogiczną do definicji zmiennej losowej).

Rodzinę $(\xi^1, \ldots, \xi^N)$ ciągów N-układów liczb $({}_1\xi^1_\alpha, {}_1\xi^2_\alpha, \ldots, {}_1\xi^N_\alpha)$, $({}_2\xi^1_\alpha, {}_2\xi^2_\alpha, \ldots, {}_2\xi^N_\alpha), \ldots\ldots, ({}_i\xi^1_\alpha, {}_i\xi^2_\alpha, \ldots, {}_i\xi^N_\alpha), \ldots$; $\alpha \in \Lambda$, N-liczba naturalna; takich,

że dla każdego układu $(x^1,\ldots\ldots x^N)$, gdzie $x^i \in \bar{R} = R \cup \{-\infty\} \cup \{+\infty\}$, $i=1,\ldots N$; granica

$$\langle H(x^1-\xi^1)\cdot \ldots \cdot H(x^N-\xi^N)\rangle = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \prod_{k=1}^{N} H(x^k-\xi^k_{\alpha_i}) \qquad (2.16)$$

istnieje i nie zależy od wyboru ciągu układów $(\xi^1_{\alpha_i},\ldots,\xi^N_{\alpha_i})$, $i\in\mathcal{N}$; (nie zależy od wyboru $\alpha\in\Lambda$), nazywamy N-wymiarowym wektorem losowym.

Ze względu na postulat niezależności granicy (2.16) od wyboru $\alpha\in\Lambda$ wystarczy znajomość tylko jednego ciągu układów.

W przypadku, gdy N=1, powyższa definicja pokrywa się z definicją zmiennej losowej podaną wcześniej.

Zauważmy, że jeżeli $x_i = -\infty$ dla pewnego $1\leqslant i\leqslant N$, to

$$\langle H(x^1-\xi^1), \ldots, H(-\infty-\xi^i), \ldots, H(x^N-\xi^N)\rangle = 0 \qquad (2.17)$$

Jeżeli mamy do czynienia ze zdarzeniem polegającym na tym, że w wyniku doświadczeń opisanych wektorem losowym $(\xi^1,\ldots,\xi^N)$ otrzymujemy układy liczb $(x^1,\ldots,x^N)$ z przedziału $[a, b)$, gdzie $a=(a^1,\ldots a^N)$, $b=(b^1,\ldots b^N)$, to

$$P_{\xi^1\ldots\xi^N}([a,b)) = \langle \prod_{k=1}^{N} (H(b^k-\xi^k) - H(a^k-\xi^k))\rangle \qquad (2.18)$$

istnieje na podstawie (2.16) i jest miarą przeliczalnie addytywną i unormowaną na ciele złożonym ze skończonych sum przedziałów N-wymiarowych lewostronnie domkniętych - prawostronnie otwartych. Miarę tę można rozszerzyć do miary przeliczalnie addytywnej, unormowanej, określonej na σ-ciele zbiorów borelowskich w sposób jednoznaczny [1], [6]. Tak otrzymaną miarę nazywamy prawdopodobieństwem wektora losowego $(\xi^1,\ldots,\xi^N)$, a zbiory borelowskie zdarzeniami. Zamiast $P_{\xi^1\ldots\xi^N}(B)$ piszemy także $P((\xi^1,\ldots,\xi^N)\in B)$, gdzie B oznacza zbiór borelowski.

Funkcję

$$F_{\xi^1\ldots\xi^N}(x^1\ldots x^N) = P_{\xi^1\ldots\xi^N}([-\infty, x^1)\times\ldots\times[-\infty, x^N)) \qquad (2.19)$$

nazywamy dystrybuantą wektora losowego $(\xi^1,\ldots,\xi^N)$.

Ze wzorów (2.19), (2.18) i (2.17) wynika, że:

$$F_{\xi^1\ldots\xi^N}(x^1,\ldots,x^N) = \langle H(x^1-\xi^1)\cdot \ldots \cdot H(x^N-\xi^N)\rangle . \qquad (2.20)$$

Podobnie jak w przypadku jednowymiarowym ciąg funkcji lokalnie całkowalnych

$$S_n(x^1,\dots,x^N) = \frac{1}{n}\sum_{i=1}^{n} (H(x^1 - \xi^1_{i\alpha})\cdot \dots \cdot H(x^N - \xi^N_{i\alpha})) \qquad (2.21)$$

jako zbieżny do $F_{\xi^1\dots\xi^N}(x^1,\dots,x^N)$ i ograniczony przez funkcję lokalnie całkowalną (np. identycznie równą 1) jest na podstawie twierdzenia (ii) z pierwszej części zbieżny dystrybucyjnie do $F_{\xi^1\dots\xi^N}(x^1\dots x^N)$. Tak więc na podstawie (1.23) i (2.21) otrzymujemy:

$$\frac{\partial^N F_{\xi^1,\dots,\xi^N}(x^1,\dots,x^N)}{\partial x^1,\dots,\partial x^N} = \lim_{n\to\infty} \frac{\partial^N S_n(x^1,\dots,x^N)}{\partial x^1\dots\partial x^N} =$$

$$= \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} (\delta(x^1 - \xi^1_{i\alpha})\cdot \dots \cdot \delta(x^N - \xi^N_{i\alpha})) =$$

$$= \langle \delta(x^1 - \xi^1)\cdot \dots \cdot \delta(x^N - \xi^N)\rangle . \qquad (2.22)$$

Oznaczając

$$g_{\xi^1,\dots,\xi^N}(x^1,\dots,x^N) = \frac{\partial^N F_{\xi^1,\dots,\xi^N}(x^1,\dots,x^N)}{\partial x^1\dots\partial x^N} \text{ (pochodna dystrybucyjna)} \qquad (2.23)$$

otrzymujemy:

$$g_{\xi^1,\dots,\xi^N}(x^1,\dots,x^N) = \langle \delta(x^1 - \xi^1)\cdot \dots \cdot \delta(x^N - \xi^N)\rangle , \qquad (2.24)$$

którą nazywamy gęstością rozkładu prawdopodobieństwa wektora losowego $(\xi^1,\dots,\xi^N)$.

Wektor losowy jest wyznaczony przez swoją gęstość bądź dystrybuantę.

Na zakończenie odnotujemy jeszcze kilka własności dystrybuanty i gęstości wektora losowego $(\xi^1,\dots,\xi^N)$.

Z definicji (2.19) wynika, że $F_{\xi^1,\dots,\xi^N}(x^1,\dots,x^N)$ jest funkcją nie malejącą i lewostronnie ciągłą, a ze wzoru (2.20) wynikają warunki symetrii:

$$F_{\xi^1,\dots,\xi^i,\dots,\xi^j,\dots,\xi^N}(x^1,\dots,x^i,\dots,x^j,\dots,x^N) =$$

$$= F_{\xi^1,\dots,\xi^j,\dots,\xi^i,\dots,\xi^N}(x^1,\dots,x^j,\dots,x^i,\dots,x^N) \text{ dla } 1\leq i\leq j\leq N \qquad (2.25)$$

i zgodności:

$$F_{\xi^1,\dots,\xi^k,\dots,\xi^N}(x^1,\dots,x^k,+\infty,\dots,+\infty) =$$

$$= F_{\xi^1,\dots,\xi^k}(x^1,\dots,x^k), \text{ dla } 1 \leqslant k \leqslant N, \tag{2.26}$$

oraz że:

$$F_{\xi^1,\dots,\xi^N}(+\infty,\dots,+\infty) = 1, \tag{2.27}$$

natomiast ze wzoru (2.17), iż:

$$F_{\xi^1,\dots,\xi^N}(x^1,\dots,\underset{(i)}{-\infty},\dots,x^N) = 0; \quad i \leqslant i \leqslant N. \tag{2.28}$$

Warunek symetrii dla gęstości

$$g_{\xi^1,\dots,\xi^i,\dots,\xi^j,\dots,\xi^N}(x^1,\dots,x^i,\dots,x^j,\dots x^N) =$$

$$= g_{\xi^1,\dots,\xi^j,\dots,\xi^i,\dots,\xi^N}(x^1,\dots,x^j,\dots,x^i,\dots,x^N) \text{dla } 1 \leqslant i \leqslant j \leqslant N \tag{2.29}$$

wynika ze wzoru (2.24).

Zauważmy jeszcze, że ważną rolę w rozważaniach odegrało pojęcie średniej po pewnej rodzinie ciągów realizacji:

$$\langle f(x-\xi)\rangle \overset{df}{=} \lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(x-{}_i\xi_\alpha),$$

jeśli istnieje i nie zależy od $\alpha \in \Lambda$.

Definicję tę można podać w przypadku, gdy $f(x)$ jest dystrybucją, a granica - granicą dystrybucyjną. Jest ona wówczas rozszerzeniem pojęcia średniej dla funkcji $f(x)$ - lokalnie całkowalnej, dla której ciąg $S_n(x) = \frac{1}{n}\sum_{i=1}^{n} f(x-{}_i\xi)$ jest zbieżny prawie wszędzie i ograniczony przez funkcję lokalnie całkowalną, co wynika z (ii) z części pierwszej.

Uwagi końcowe

Podjętą w artykule tematykę mamy zamiar kontynuować w następnych pracach. Będziemy starali się podawać i tłumaczyć z pomocą aparatu teorii

dystrybucji konsekwencje płynące z przedstawionego ujęcia rachunku prawdopodobieństwa. Dlatego też zamieściliśmy tak obszerny przegląd podstawowych faktów teorii dystrybucji, mimo iż nie wszystkie zostały wykorzystane.

Tematyka, którą rozpoczęliśmy i mamy zamiar przedstawiać nadal, była i będzie przedmiotem dyskusji seminarium Zakładu Mechaniki Teoretycznej w Instytucie Matematyki Uniwersytetu Śląskiego.

LITERATURA

[1] Gichman J.J., Skorochod A.W.: Wstęp do teorii procesów stochastycznych, PWN, Warszawa 1968.

[2] Łojew M.: Teorija wierojatnostiej, Izdatielstwo inostrannoj litieratury, Moskwa 1962.

[3] Mikusiński J., Sikorski R.: Elementarna teoria dystrybucji, PWN, Warszawa 1964.

[4] Skalmierski B.: Mechanika z wytrzymałością materiałów dla automatyków, PWN, Warszawa 1973.

[5] Stratonivič R.A.: Izbrannyje voprosy tieorii fluktuacji w radiotiechnikie, Sovietskoje Radio, Moskwa 1961.

[6] Zubrzycki S.: Wykłady z rachunku prawdopodobieństwa i statystyki matematycznej, PWN, Warszawa 1966.

ОСНОВЫ ТЕОРИИ ВЕРОЯТНОСТИ В ДИСТРИБУЦИОННОМ ПОДХОДЕ

Р е з ю м е

В работе представлена возможность построения теории вероятности,отступая от традиционной системы Колмогорова. Авторы хотят найти объяснение некоторых фактов этой теории на основании теории дистрибуции. Эта область математического знания позволяет их легко заметить и благодаря этому является существенной в применении.

THE FOUNDATIONS OF PROBABILITY, CALCULUS BASED ON THE THEORY OF DISTRIBUTION

S u m m a r y

The present study discusses the possibility of constructiong the calculus of probability, departing from the traditional Kołmogorow systems.The authors want to find an explanation of certain facts of this theory on the basis of the theory of distribution. This field of matematics makes it possible to notice them relatively easily and thanks to this turns out to be essential.