

SĄ SZANSE ŻE...

WYKŁADY Z RACHUNKU PRAWDOPODOBIENSTWA
JESIEŃ 2008

Spis treści

1	Wstęp	5
1.1	Trochę filozofii (od Epikura do teorii kwantowej)	5
1.2	Trochę historii (od hazardu do aksjomatów)	6
1.3	Przykłady eksperymentów	8
2	Skończone przestrzenie probabilistyczne	9
2.1	Własności prawdopodobieństwa	9
3	Szczypta kombinatoryki	13
3.1	Prawa przeliczania	13
3.1.1	Prawo dodawania	13
3.1.2	Prawo mnożenia	13
3.1.3	Zasada bijekcji	13
3.2	Schematy losowania r spośród n elementów	13
4	Nieskończone przestrzenie probabilistyczne	17
4.1	Przestrzenie przeliczalne	17
4.2	Przykład przestrzeni nieprzeliczalnej: nieskończony ciąg rzutów monetą	18
4.3	Aksjomatyczna definicja przestrzeni probabilistycznej	19
4.4	Własności prawdopodobieństw monotonicznych ciągów zdarzeń	20
4.5	Powrót do odcinka $(0, 1]$	21
4.6	Prawdopodobieństwo geometryczne.	22
5	Prawdopodobieństwo warunkowe	23
6	Niezależność zdarzeń	29
6.1	Zastosowanie niezależności zdarzeń w teorii liczb	31
6.2	Przestrzenie produktowe	31
6.3	Schemat Bernoulliego	33
7	Zmienne losowe i ich rozkłady	35
7.1	Klasyfikacja rozkładów	38
7.1.1	Podział na rozkłady dyskretne i ciągłe	38
7.1.2	Podział na rozkłady absolutnie ciągłe i osobliwe	39
7.2	Przykłady rozkładów	40

7.2.1	Rozkłady dyskretne	40
7.2.2	Rozkłady ciągłe (absolutnie!)	41
8	Wektory losowe (X, Y)	43
8.1	Rozkłady brzegowe	44
8.2	Rozkłady warunkowe	45
9	Funkcje zmiennych losowych	47
9.1	Niezależność	47
9.2	Znajdowanie rozkładów funkcji zmiennych losowych	47
10	Nadzieja matematyczna	53
10.1	Mediana zmiennej losowej X	53
10.2	Wartość oczekiwana	53
10.3	Momenty wyższych rzędów	55
10.4	Nierówności Markowa i Czebyszewa	55
10.5	Słabe Prawo Wielkich Liczb	56
10.6	Nierówność Chernoffa	56
10.7	Kowariancja i korelacja	58
10.8	Warunkowa wartość oczekiwana	59
10.9	Regresja	61
10.10	Wielowymiarowy rozkład normalny	62
11	Funkcje tworzące i procesy gałązkowe	65
11.1	Funkcje tworzące	65
11.2	Procesy gałązkowe	66
12	Twierdzenia graniczne	69
12.1	Lematy Borela-Cantelliego	69
12.2	Mocne Prawo Wielkich Liczb	70
12.3	Funkcje charakterystyczne	71
12.4	Centralne Twierdzenie Graniczne	72
12.5	Typy zbieżności	73
13	Posmak statystyki matematycznej	75
13.1	Estymacja	75
13.2	Testowanie hipotez	77

Rozdział 1

Wstęp

Rachunek prawdopodobieństwa bada zjawiska i doświadczenia (eksperymenty) losowe, to znaczy takie, których skutku (wyniku) nie można przewidzieć w ramach posiadanej wiedzy. Liczba wypadków na drogach podczas najbliższego weekendu jest zjawiskiem losowym. Klasycznym eksperymentem losowym jest rzut monetą lub kostką (demonstracja rzutu kostką). Chociaż ruch kostki podlega prawom fizyki, to ich zastosowanie jest na tyle skomplikowane i niepraktyczne, że traktujemy ten rzut jak losowy. Również to, który tramwaj (linii nr 12, 14 czy 15) nadjedzie pierwszy na przystanek jest, dla osoby nie znającej rozkładu, zjawiskiem losowym.

1.1 Trochę filozofii (od Epikura do teorii kwantowej)

Rachunek prawdopodobieństwa jest próbą formalnego opisu praw rządzących przypadkiem. Można powiedzieć, że *szuka porządku w chaosie*. To tylko pozorny paradoks. Filozofowie od wieków toczą spór o to, czy świat jest przypadkowy czy zdeterminowany. Poglądy starożytnych były, jak zwykle, umiarkowane. Epikur twierdził, że *przypadek to odchylenie od konieczności*. Podobne stanowisko przyjmują wszystkie religie, dla których synonimem przypadku jest cud.

Filozofia nowożytna przyjęła początkowo stanowisko *determinizmu jednoznacznego*. Według jego wyznawców (Spinoza, Wolter, Hume, Humboldt) nic nie dzieje się przypadkowo; to tylko niedoskonałość naszej wiedzy powoduje, że niektóre zjawiska wydają nam się przypadkowe. Ujął to zwięźle Mill pisząc, że *przypadek to nieznaną przyczynę prawa*. Determinizm jednoznaczny obejmował również ludzką wolę, która miała być wolna tylko subiektywnie.

Wiek XIX podważył tezy determinizmu jednoznacznego, a przyczyniły się do tego nowe prawa fizyki o charakterze statystycznym (prawo Gay-Lussaca i prawo gazu doskonałego Clapeyrona). Hegel głosił, iż każde zjawisko jest zarazem przypadkowe i konieczne. Poglądy te potwierdziła współczesna teoria kwantowa. Z tych korzeni wyrósł *determinizm statystyczny*, którego

główne przesłanie można by ująć w stwierdzeniu, że *z chaosu w skali mikro zbudowany jest porządek w skali makro*. Pogląd ten jest zaskakująco zbieżny z teorią teleologiczną Platona. Głosiła ona, że determinizm świata wynika z jego celowości, której główną zasadą jest to, że *część jest podporządkowana całości*.

Taki też charakter, jak się przekonamy, mają twierdzenia rachunku prawdopodobieństwa. Wychodząc z niewiedzy na temat pojedynczych zjawisk, otrzymuje się wiedzę dotyczącą całych klas zjawisk.

1.2 Trochę historii (od hazardu do aksjomatów)

Przyjrzyjmy się teraz jak historycznie rozwijał się ten dział matematyki.

Starożytni Grecy rozczarują nas tym razem, bo chociaż namiętnie grali w kości, to nigdy nie zastanawiali się nad szansami. Podobnie Arabowie, choć słowo *hazard* pochodzi od arabskiego *al zar* co znaczy *kostka*. Pierwsze pytanie probabilistyczne opublikowano w 1477 roku w jednym z komentarzy do *Boskiej Komedi* Dantego. Dotyczyło ono zmienności szans na pojawienie się określonej sumy oczek przy rzucie trzema kostkami. Podobne pytania znalazły się w późniejszej korespondencji Pascala i Fermata (w. XVII). Pierwszą książkę poświęconą obliczaniu prawdopodobieństw wydał Montmort w 1708 w Paryżu. Prawdopodobieństwo jest w niej zdefiniowane jako *miara naszych oczekiwań*. Ale już w 1711 de Moivre wprowadził *prawdopodobieństwo klasyczne* jako odwrotność liczby wszystkich możliwych wyników przy założeniu, że są one *równoprawdopodobne*. Ta definicja spotkała się natychmiast z zarzutem, że opiera się na „błędym kole”. Wiek XVIII dał jeszcze jednego uczonego, który wielce zasłużył się dla rozwoju rachunku prawdopodobieństwa. Jakub Bernoulli jako pierwszy badał nieskończone ciągi prób losowych, wprowadził niezwykle ważne pojęcie niezależności zdarzeń i udowodnił pierwsze twierdzenia graniczne.

Ale naprawdę burzliwy rozwój nastąpił w wieku XIX. Najpierw, w pierwszej połowie wieku, Laplace, Poisson i Gauss (zmiennie losowe, rozkłady ciągłe), a następnie, w drugiej połowie, Rosjanie Czebyszew, Markow i Liapunow (momenty zmiennych losowych, centralne twierdzenie graniczne) sprawili, że rachunek prawdopodobieństwa stał się wyraźną i dynamiczną gałęzią matematyki. Wciąż jednak jej słabym punktem był brak formalnej definicji prawdopodobieństwa. Po niefortunnym prawdopodobieństwie klasycznym Moivra, Ellis w 1849 podjął kolejną próbę i wprowadził w zamian *definicję częstościową*, którą rozwinęli już w latach 20-tych XX wieku von Mises i Popper. Zgodnie z tym podejściem, aby mówić o prawdopodobieństwie trzeba wyobrazić sobie nieskończony ciąg wyników tego samego doświadczenia, powtarzanego w niezmiennych warunkach. Niech $a_n(\omega)$ oznacza liczbę tych spośród pierwszych n eksperymentów, które zakończyły się wynikiem ω .

Proponowano, by za prawdopodobieństwo zdarzenia ω przyjąć granicę

$$p(\omega) = \lim_{n \rightarrow \infty} \frac{a_n(\omega)}{n}$$

Ta definicja również jest trudna do zaakceptowania, bo wymaga wykonania nieskończonej liczby eksperymentów, a ponadto zakłada, że wspomniana granica istnieje.

Wyjściem z sytuacji stało się przyjęcie definicji aksjomatycznej. Nad dziełem aksjomatyzacji pracowali Bernstein, Borel i Kołmogorow w latach 20-tych i 30-tych XX wieku, a uwieńczył je Kołmogorow w 1933. W tym nowoczesnym ujęciu, prawdopodobieństwo, podobnie jak punkt w geometrii, jest obiektem niedefiniowalnym, który spełnia tylko pewne warunki (aksjomaty). Przy formułowaniu aksjomatów jednym ze źródeł inspiracji było powszechne już wtedy w użyciu *prawdopodobieństwo geometryczne*. Dotyczyło ono sytuacji, gdy zbiór wszystkich możliwych wyników eksperymentu był nieprzeliczalnym podzbiorem prostej czy płaszczyzny. Wtedy prawdopodobieństwo obliczało się jako stosunek pól (ogólniej miar) odpowiednich zbiorów. Efektem tej inspiracji jest to, że zaksjomatyzowane prawdopodobieństwo jest po prostu *miarą abstrakcyjną*. Rachunek prawdopodobieństwa bada własności tej miary. Nie wyznacza prawdopodobieństw, ale ustala związki między nimi. Na przykład, pozornie oczywiste stwierdzenie, że prawdopodobieństwo wyrzucenia na kostce parzystej liczby oczek wynosi $1/2$, tak naprawdę, zakłada, że kostka jest „uczciwa”. Gdyby jednak z jakichś powodów przyjąć, że szanse na 1, 2 i 3 oczka wynoszą po $1/4$, na 4 i 5 oczek po $1/8$ a 6 nigdy się nie pojawi, to powyższe prawdopodobieństwo wyniosłoby $3/8$.

Czyżby więc panowała tu straszliwa anarchia? Na szczęście, jak wiele innych teorii matematycznych, oprócz strony formalnej, rachunek prawdopodobieństwa ma swoje zaplecze intuicyjne, do którego powinien się odnosić. Tym zapleczem jest właśnie zaobserwowana częstość występowania danego zdarzenia (wyniku). Częstość jest tym dla prawdopodobieństwa, czym prosta narysowana na kartce papieru dla prostej w geometrii. W zastosowaniach należy być wiernym intuicji, i dlatego do problemów kombinatorycznych będziemy stosować prawdopodobieństwo klasyczne, czyli zwykle przeliczanie, a do geometrycznych użyjemy miary geometrycznej. Trzeba też pamiętać o statystycznym charakterze praw rachunku prawdopodobieństwa. Prawdopodobieństwo wyrzucenia orła na uczciwej monecie wynosi $1/2$, co wcale nie oznacza, że w dwóch kolejnych rzutach raz wypadnie orzeł a raz reszka. Znaczy to tylko tyle, że przy zwiększającej się liczbie rzutów, stosunek otrzymanych orłów i reszek będzie zbliżać się do jedności.

Podsumowując, rachunek prawdopodobieństwa bada zjawiska i eksperymenty losowe, które charakteryzują się brakiem deterministycznej regularności, ale za to wykazują pewną regularność statystyczną. W eksperymencie losowym można wyróżnić mechanizm losujący oraz zbiór możliwych wyników. Matematycznym modelem eksperymentu losowego jest przestrzeń probabilistyczna. Zanim ją zdefiniujemy, przyjrzyjmy się kilku przykładom takich ekspery-

mentów.

1.3 Przykłady eksperymentów

1. Rzut dwoma kostkami (demonstracja); nawet jeśli kostki wydają się być identyczne, to musimy je traktować jak obiekty rozróżnialne, bo takie w rzeczywistości są – druga opcja, to traktować je jak nierozróżnialne, ale wtedy przyjąć, że zdarzenia (i, i) , $i = 1, 2, \dots, 6$, są dwa razy bardziej prawdopodobne od pozostałych. Paradoks de Mere'a o szansach (równych, czy nie?) wyrzucenia sumy 11 i sumy 12 na 3 „identycznych” kostkach do gry, opierał się właśnie na niezrozumieniu tej subtelności;

2. Gra w 3 karty z wariantem;

3. Szybki tramwaj (w/g rozkładu).

Zbiór wyników może być:

- (a) skończony,
- (b) przeliczalny,
- (c) nieprzeliczalny.

Przypadki a) i b) określa się łącznie mianem *dyskretnych*. W związku z tym, rachunek prawdopodobieństwa dzieli się na dyskretny i ciągły, a metody na kombinatoryczne i miarowo-całkowe. Nasze wykłady rozpoczniemy od najprostszego i najbliższego intuicji przypadku, gdy zbiór wyników jest skończony.

Rozdział 2

Skończone przestrzenie probabilistyczne

Definicja. Skończona przestrzeń probabilistyczna to para

$$(\Omega, p), \quad \text{gdzie } p : \Omega \rightarrow [0, 1] \text{ oraz } \sum_{\omega \in \Omega} p(\omega) = 1$$

Elementy zbioru Ω nazywamy *zdarzeniami elementarnymi* (lub wynikami eksperymentu), a podzbiory zbioru Ω – *zdarzeniami*.

Prawdopodobieństwo zdarzenia $A \subseteq \Omega$ definiujemy jako wartość funkcji zbiorów $P : 2^\Omega \rightarrow [0, 1]$ określonej wzorem

$$P(A) = \sum_{\omega \in A} p(\omega) \tag{2.1}$$

dla $A \neq \emptyset$ oraz $P(\emptyset) = 0$.

O klasycznym prawdopodobieństwie mówimy, gdy $p \equiv \frac{1}{|\Omega|}$. Wtedy $P(A) = |A|/|\Omega|$.

Na razie (do czasu sformułowania aksjomatów) przyjmujemy, że p jest dane, a naszym celem jest obliczanie P na podstawie p , a nie samego p .

2.1 Własności prawdopodobieństwa

Podamy i udowodnimy 10 własności prawdopodobieństwa. Własności (4-10) wyprowadzimy z poprzednich, tak, by ich dowody nie zależały od definicji P (wzór (2.1)).

(1) $P(\Omega) = 1$ – z definicji

(2) $0 \leq P(A) \leq 1$ – z definicji

(3) Jeśli $AB = \emptyset$, to $P(A \cup B) = P(A) + P(B)$ (addytywność dla wykluczających się zdarzeń – prosty dowód z definicji; n zdarzeń przez indukcję)

(4) $P(B \setminus A) = P(B) - P(AB)$ – z (3)

(4a) Jeśli $A \subseteq B$, to $P(B \setminus A) = P(B) - P(A)$ – przypadek szczególny wzoru (4)

10 ROZDZIAŁ 2. SKOŃCZONE PRZESTRZENIE PROBABILISTYCZNE

- (5) Jeśli $A \subseteq B$, to $P(A) \leq P(B)$ – z (2) i (4a)
 (6) $P(A^c) = 1 - P(A)$ – z (3) i (1)
 (7) $P(A \cup B) = P(A) + P(B) - P(AB)$ – z (3) dla $n = 3$, oraz z (4)
 (8) $P(\bigcup_{i=1}^n A_i) \leq \sum_{i=1}^n P(A_i)$ – nierówność Boole'a dla n zdarzeń
 (9) Wzór Poincare z dowodem indukcyjnym

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n (-1)^{i-1} S_i,$$

gdzie

$$S_i^{(n)} = S_i = \sum_{J \in \binom{[n]}{i}} P\left(\bigcap_{j \in J} A_j\right).$$

Dowód. Wzór jest prawdziwy dla $n = 1$ (trywialnie) i dla $n = 2$ (7). Oznaczmy $B = \bigcup_{i=1}^{n-1} A_i$. Wtedy

$$P\left(\bigcup_{i=1}^n A_i\right) = P(B) + P(A_n) - P(BA_n)$$

Z założenia indukcyjnego dla $n - 1$

$$P(B) = \sum_{i=1}^{n-1} (-1)^{i-1} \sum_{J \in \binom{[n-1]}{i}} P\left(\bigcap_{j \in J} A_j\right) = \sum_{i=1}^{n-1} (-1)^{i-1} \sum_{J \in \binom{[n]}{i}, n \notin J} P\left(\bigcap_{j \in J} A_j\right)$$

oraz

$$\begin{aligned} P(BA_n) &= P\left(\bigcup_{i=1}^{n-1} (A_i A_n)\right) = \sum_{i=1}^{n-1} (-1)^{i-1} \sum_{J \in \binom{[n-1]}{i}} P\left(\bigcap_{j \in J \cup \{n\}} A_j\right) \\ &= \sum_{k=2}^n (-1)^{k-2} \sum_{J \in \binom{[n-1]}{k-1}} P\left(\bigcap_{j \in J \cup \{n\}} A_j\right) = \sum_{k=2}^n (-1)^k \sum_{J \in \binom{[n]}{k}, n \in J} P\left(\bigcap_{j \in J} A_j\right) \end{aligned}$$

Zatem

$$\begin{aligned} P\left(\bigcup_{i=1}^n A_i\right) &= \sum_{i=2}^{n-1} (-1)^{i-1} \sum_{J \in \binom{[n]}{i}} P\left(\bigcap_{j \in J} A_j\right) + \sum_{j=1}^{n-1} P(A_j) \\ &\quad + P(A_n) + (-1)^{n-1} P\left(\bigcap_{j=1}^n A_j\right) = \sum_{i=1}^n (-1)^{i-1} S_i \end{aligned}$$

□

(10) Zasada włączania i wyłączania (wzór Sylvestra) – wynika ze wzoru Poincare i prawa de Morgana.

$$P\left(\bigcap_{i=1}^n A_i^c\right) = \sum_{i=0}^n (-1)^i S_i \quad S_0 = 1$$

Jest to więc wzór na prawdopodobieństwo zdarzenia polegającego na tym, że nie zajdzie żadne ze zdarzeń A_1, \dots, A_n . Korzystając ze wzoru (2.1), można tę zasadę, a więc i wzór (9), dość szybko udowodnić kombinatorycznie.

Dowód. Rozpiszmy obie strony wzoru (10) stosując wzór (2.1):

$$\sum_{\omega \in \bigcap_{i=1}^n A_i^c} p(\omega) = \sum_{i=0}^n (-1)^i \sum_{J \in [n]^i} \sum_{\omega \in \bigcap_{j \in J} A_j} p(\omega). \quad (2.2)$$

Niech $\omega \in \Omega$ należy do dokładnie $r_\omega = r$ spośród zbiorów A_1, \dots, A_n , $0 \leq r \leq n$. Składnik $p(\omega)$ występuje po lewej stronie wzoru (2.2) wgdy $r_\omega = 0$. Policzmy teraz, ile razy ten składnik występuje po prawej stronie wzoru (2.2). Niech ω należy do zbiorów A_i o indeksach i_1, \dots, i_r . Wtedy p_ω wystąpi raz dla każdego zbioru $J \subseteq \{i_1, \dots, i_r\}$, a więc łącznie

$$\sum_{i=0}^r \binom{r}{i} (-1)^i = \begin{cases} 0 & \text{gdy } r > 0 \\ 1 & \text{gdy } r = 0, \end{cases} \quad (2.3)$$

co należało dowieść. \square

Problem listów, czyli roztargniona sekretarka, (Montmort 1708).

Wkładamy przypadkowo n listów do n zaadresowanych kopert. Dla ustalonego $0 \leq m \leq n$, obliczyć prawdopodobieństwo $P_m^{(n)}$ zdarzenia, że dokładnie m listów trafi do właściwych kopert.

Najpierw $m = 0$. Niech Ω będzie zbiorem wszystkich $n!$ permutacji liczb naturalnych od 1 do n . Niech, dla każdego $i = 1, \dots, n$, A_i będzie zdarzeniem, że i -ty list trafi do właściwej koperty. Mamy

$$S_i = \binom{n}{i} \frac{(n-i)!}{n!} = \frac{1}{i!}$$

skąd

$$P_0^{(n)} = \sum_{i=0}^n (-1)^i / i!$$

Dla dowolnego m , rozbijmy zdarzenie „dokładnie m punktów stałych”, na $\binom{n}{m}$ rozłącznych zdarzeń, według tego, gdzie znajdują się punkty stałe. Na podstawie wzoru (3),

$$P_m^{(n)} = P\left(\bigcup_{J \in \binom{[n]}{m}} B_J\right) = \sum_{J \in \binom{[n]}{m}} P(B_J),$$

gdzie

$$B_J = \bigcap_{j \in J} A_j \cap \bigcap_{j \notin J} A_j^c.$$

Na podstawie (2.1),

$$P(B_J) = \frac{D_{n-m}}{n!},$$

gdzie D_n jest liczbą nieporządków (tzn. permutacji bez punktów stałych) rzędu n . Ale $D_{n-m} = (n-m)!P_0^{(n-m)}$, więc stosując wzór na $P_0^{(n)}$, mamy ostatecznie

$$P_m^{(n)} = \binom{n}{m} \sum_{i=0}^{n-m} \frac{(-1)^i (n-m)!}{i!n!} = \sum_{i=0}^{n-m} \frac{(-1)^i}{i!m!} \rightarrow \frac{1}{em!},$$

gdy $n \rightarrow \infty$. Jak się później dowiemy, ciąg $a_m = 1/(em!)$, $m = 0, 1, \dots$, jest szczególnym przypadkiem rozkładu Poissona.

Rozdział 3

Szczypta kombinatoryki

Motywacja: klasyczna definicja prawdopodobieństwa prowadzi do przeliczania: $P(A) = |A|/|\Omega|$.

3.1 Prawa przeliczania

3.1.1 Prawo dodawania

Jeśli $A \cap B = \emptyset$, to $|A \cup B| = |A| + |B|$ (stosowane przy rozbijaniu na przypadki)

3.1.2 Prawo mnożenia

Dla dowolnych zbiorów A i B ,

$$|\{(a, b) : a \in A, b \in B\}| = |A||B|$$

(np. 3 pary spodni i 4 marynarki w szafie)

Przykład. (i) Ile jest palindromów wśród liczb pięciocyfrowych? (ii) ile liczb pięciocyfrowych zawiera dokładnie jedną cyfrę 3?

3.1.3 Zasada bijekcji

Jeśli istnieje bijekcja $f : A \rightarrow B$, to $|A| = |B|$.

Przykład. Jak pokazać, że jest 2^n podzbiorów zbioru mocy n ? Wystarczy wskazać naturalną bijekcję pomiędzy podzbiarami a binarnymi ciągami długości n , a te przeliczyć stosując prawo mnożenia (dla n 2-elementowych zbiorów).

3.2 Schematy losowania r spośród n elementów

Dwa pytania: I. Czy kolejność jest istotna? II. Czy dopuszczamy powtórzenia? Zgodnie z prawem mnożenia są 4 odpowiedzi na te pytania.

Kolejność istotna – wariacje (z powtórzeniami: n^r lub bez: $(n)_r$).

Kolejność nieistotna – kombinacje (zwykle: $\binom{n}{r}$ lub z powtórzeniami: $\binom{n+r-1}{r}$). Kombinacje z powtórzeniami nie przydają się zbytnio w rachunku prawdopodobieństwa. Służą bowiem do przeliczania obiektów nieoznaczonych, jak ma to miejsce przy rzucie dwoma identycznymi kośćmi do gry. Wyników jest 21, ale nie wszystkie są jednakowo prawdopodobne. Stosując ten model należałoby przypisać wynikom typu $\{i, i\}$ prawdopodobieństwo $1/36$ a pozostałym $1/18$. W takim razie lepiej od razu rozważać rzut dwoma różnymi kośćmi z prawdopodobieństwem klasycznym.

Zadanie Banacha. (Przykład zadania, gdzie przestrzeń nie jest jednakowo prawdopodobna, ale wzory kombinatoryczne bardzo się przydają.)

Matematyk (Banach) ma w kieszeniach dwa pudełka cukierków tic-tac (L w lewej i R w prawej), po N cukierków w każdym. Gdy przyjdzie mu ochota, bezwiednie sięga do przypadkowej kieszeni, wyjmując pudełko, a z niego cukierek, po czym chowa pudełko do tej samej kieszeni. Obliczyć prawdopodobieństwo tego, że gdy wyjmie po raz pierwszy puste pudełko, to w drugim będzie wtedy dokładnie k cukierków, $0 \leq k \leq N$.

Rozwiązanie: Ω składa się z ciągów binarnych (L i R) długości od $N+1$ do $2N+1$. Za prawdopodobieństwo $p(\omega)$ ciągu ω długości l należy przyjąć 2^{-l} . Można to tłumaczyć na kilka sposobów, lecz żaden nie jest w pełni przekonujący. Na przykład, że tak nakazuje intuicja i zdrowy rozsądek. Albo, że wszystkie ciągi tej samej długości powinny być równoprawdopodobne. Cóż z tego, skoro nie wszystkie ciągi długości $l > N+1$ należą do Ω . Można też stosować indukcję zapisując Ω w postaci drzewa.

Może najlepsze jest sztuczne wydłużenie eksperymentu i przyjęcie, że nawet po wyjęciu pustego pudełka matematyk chowa je z powrotem i ponownie losuje, tak by łączna liczba losowań wyniosła $2N+1$. W ten sposób wszystkie ciągi w $\omega \in \Omega$ są równej długości i nie ma powodu, by nie były równoprawdopodobne. Przypisujemy więc im wszystkim $p(\omega) = 2^{-2N-1}$. Zatem szukane prawdopodobieństwo wynosi

$$2 \binom{2N-k}{N} 2^k 2^{-2N-1} = \binom{2N-k}{N} 2^{-2N+k}.$$

Zauważmy, że przy okazji otrzymaliśmy ciekawą identyczność

$$\sum_{k=0}^N \binom{2N-k}{N} 2^k = 2^{2N}$$

Spójrzmy jeszcze, który przypadek jest bardziej prawdopodobny: $k=0$ czy $k=N$? Dla $k=N$, powyższe prawdopodobieństwo wynosi 2^{-N} . Natomiast dla $k=0$, stosując wzór Stirlinga, otrzymujemy przybliżony wynik $\frac{1}{\sqrt{\pi N}}$.

Na koniec, wzór na liczbę podziałów zbioru na podzbiory, który jest jednocześnie wzorem na liczbę permutacji z powtórzeniami. Jeśli $n = n_1 + \dots + n_k$, to liczba podziałów zbioru n -elementowego na podzbiory A_1, A_2, \dots, A_k o mocach, odpowiednio, n_1, n_2, \dots, n_k wynosi

$$\binom{n}{n_1, \dots, n_k}.$$

Permutacje k grup jednakowych elementów o mocach n_1, n_2, \dots, n_k , gdzie $n = n_1 + \dots + n_k$, można traktować jak podziały pustych (na razie) miejsc na których chcemy zapisać permutowane elementy.

Przykład : MATEMATYKA. Ile różnych ciągów można utworzyć z liter słowa MATEMATYKA? Jest to proste zadanie na permutacje z powtórzeniami. Jednocześnie dobrze widać tu odpowiadające im podziały. Dzielonym zbiorem jest zbiór 10 miejsc, na których zapisujemy litery. Podzbiory odpowiadają poszczególnym rodzajom liter (M, A, itd.)

Powyższy wzór można stosować tylko, gdy kolejność zbiorów w podziale jest istotna. W przeciwnym razie potrzebne są modyfikacje. Ilustrują to dwa kolejne przykłady.

Przykład. Przeliczyć podziały $2n$ -osobowej grupy studenckiej na 2-osobowe zespoły do (i) poloneza i (ii) w celu pracy nad tym samym zadaniem.

Drugi przykład pokazuje, że gdy nie wszystkie podzbiory są tej samej mocy, to problem staje się bardziej subtelny.

Przykład. Przeliczyć podziały 22-osobowej grupy na 3 zespoły 4-osobowe i 2 zespoły 5-osobowe w celu przydzielenia im (a) tego samego zadania; (b) 5 różnych zadań, przy czym wiadomo, które zadania są dla zespołów 4-osobowych, a które dla 5-osobowych (jesli tego nie zrobimy, wynik w części (b) trzeba przemnożyć przez $\binom{5}{3}$).

Odpowiedzi: (a) $\frac{22!}{4!^3 5!^2 3! 2!}$; (b) $\frac{22!}{4!^3 5!^2}$.

Rozdział 4

Nieskończone przestrzenie probabilistyczne

4.1 Przestrzenie przeliczalne

Czy ma sens pojęcie „losowej liczby naturalnej”? Nie ma, jeśli wymagać, by wszystkie liczby miały te same szanse.

Czy przestrzenie przeliczalne są do czegoś potrzebne? Często opis teoretyczny jakiegoś zjawiska losowego staje się prostszy, gdy przyjmiemy, że $\Omega = \mathcal{N} = \{1, 2, 3, \dots\}$, choć w rzeczywistości wynik eksperymentu można zawsze ograniczyć. Przykład: liczba wypadków na danym skrzyżowaniu w ciągu danego okresu czasu.

Przestrzenie przeliczalne, czyli takie gdzie $\Omega = \mathcal{N}$ oparte są na szeregach zbieżnych. Jeśli $\sum_{n=1}^{\infty} p_n = 1$, $p_n \geq 0$, to możemy przyjąć

$$p(\{n\}) = p_n$$

i oba warunki narzucone na funkcję p w przypadku skończonym (nieujemność i sumowanie do 1) są spełnione.

Przykłady:

1. Szereg geometryczny

$$\sum_{n=1}^{\infty} x^n = \frac{x}{1-x}$$

Zatem $p_n = (1-x)x^{n-1}$, $n = 1, 2, \dots$ jest rozkładem prawdopodobieństwa (rozkład geometryczny)

2. Szereg wykładniczy

$$\sum_{n=0}^{\infty} \frac{x^n}{n!} = e^x$$

Zatem $p_n = e^{-x} \frac{x^n}{n!}$, $n = 0, 1, 2, \dots$ jest rozkładem prawdopodobieństwa (rozkład Poissona) Znamy go już z problemu Montmorta (z $x = 1$).

Poza brakiem odpowiednika definicji klasycznej, przestrzenie przeliczalne nie różnią się niczym od skończonych. W szczególności, dla każdego podzbioru $A \subseteq \mathcal{N}$ można zdefiniować

$$P(A) = \sum_{n \in A} p_n$$

Jeśli A jest nieskończony, to $P(A)$ jest sumą szeregu zbieżnego.

4.2 Przykład przestrzeni nieprzeliczalnej: nieskończony ciąg rzutów monetą

Tutaj $\Omega = 2^{\mathcal{N}}$. Intuicja mówi, że $P(\omega) \equiv 0$, bo wszystkie ciągi mają tę samą szansę wystąpienia, a gdyby była ona dodatnia, to suma (na dowolnym przeliczalnym podzbiorze Ω) przekroczyłaby 1. A zatem nie możemy budować funkcji prawdopodobieństwa P jak poprzednio. Cegiełki, od których zaczniemy, muszą być znacznie większe niż atomy. Muszą być nieprzeliczalne! Żeby je lepiej określić, odwzorujemy powyższą przestrzeń w odcinek $[0, 1]$.

W tym celu skorzystamy z rozwinięć binarnych. Ponieważ jednak każda liczba o skończonym rozwinięciu ma drugie, nieskończone, tak jak $1/2 = 0,1 = 0,0111\dots$, to usuwamy z Ω wszystkie ciągi, które kończą się nieskończoną serią zer. Oznaczmy otrzymany podzbiór przez Ω' . Odwzorowanie $f : \Omega' \rightarrow (0, 1]$, przypisujące każdemu nieskończonemu ciągowi binarnemu z Ω' liczbę rzeczywistą z przedziału $[0, 1]$, której rozwinięciem binarnym jest ten ciąg, jest bijekcją.

Naturalnym kandydatem do przejęcia roli atomów są teraz przedziały. Naturalnym jest też, że powinniśmy przyjąć

$$P((a, b]) = b - a$$

Sprawdźmy, czy to jest zgodne z intuicją jaką mamy o rzutach monetą: niech A będzie zdarzeniem, że $d_i = u_i$, $i = 1, \dots, n$ (na przykład $d_1 = 1, d_2 = 0, d_3 = 1$). $P(A)$ powinno równać się 2^{-n} . Mamy

$$f(A) = \left[\sum_{i=1}^n \frac{u_i}{2^i}, \sum_{i=1}^n \frac{u_i}{2^i} + 2^{-n} \right]$$

więc $P(f(A)) = 2^{-n}$, i wszystko się zgadza (w naszym przykładzie $f(A) = (5/8, 3/4)$ i $P((5/8, 3/4)) = 1/8$).

Ale jak określić P na innych (niż przedziały) zdarzeniach? Do pytania tego powrócimy za chwilę, bo wykracza ono poza ramy tego konkretnego przykładu.

4.3 Aksjomatyczna definicja przestrzeni probabilistycznej

Okazuje się bowiem, że jeśli upierać się przy tym, by dla nieprzeliczalnej przestrzeni Ω i funkcji zbiorów $P : 2^\Omega \rightarrow [0, 1]$ zachodziły jednocześnie wszystkie trzy poniższe warunki:

$$(P1) \quad P(\Omega) = 1,$$

$$(P2) \quad P(\bigcup A_n) = \sum P(A_n) \text{ dla rozłącznych zdarzeń } A_1, A_2, \dots,$$

$$(P3) \quad P(\omega) \equiv 0,$$

to takie P , zakładając hipotezę continuum i aksjomat wyboru, NIE ZAWSZE ISTNIEJE (patrz Billingsley, str. 54, u góry).

Zadanie: Dla dowolnej przestrzeni Ω określić $P : 2^\Omega \rightarrow [0, 1]$, które spełnia tylko warunki (P1) i (P2).

Rozwiązanie: Ustalić element $\omega_0 \in \Omega$ i zdefiniować

$$P(A) = \begin{cases} 1 & \text{gdy } \omega_0 \in A \\ 0 & \text{w przeciwnym razie} \end{cases}.$$

W sposób oczywisty, warunek (P3) nie zachodzi.

Ponieważ jednak warunek (P3) jest w wielu przypadkach konieczny, to trzeba się ograniczyć do pewnej podrodziny $\mathcal{F} \subseteq 2^\Omega$. Rodzina ta powinna spełniać pewne naturalne aksjomaty:

$$(F1) \quad \emptyset, \Omega \in \mathcal{F},$$

$$(F2) \quad \text{Jeśli } A \in \mathcal{F}, \text{ to } A^c \in \mathcal{F},$$

$$(F3) \quad \text{Jeśli } A_1, A_2, \dots \in \mathcal{F}, \text{ to } \bigcup A_n \in \mathcal{F}.$$

Rodzinę \mathcal{F} spełniającą aksjomaty (F1-F3) nazywamy σ -ciałem, lub inaczej σ -algebrą. Jeśli warunek (F3) zastąpić słabszym warunkiem

$$(F3') \quad \text{Jeśli } A_1, A_2 \in \mathcal{F}, \text{ to } A_1 \cup A_2 \in \mathcal{F}$$

to rodzinę \mathcal{F} nazywamy ciałem. Zwróćmy uwagę, że to to samo ciało co poznane wcześniej na wykładach z algebry. Warunek (F1) można zastąpić przez (F1') $\Omega \in \mathcal{F}$ lub (F1'') $\mathcal{F} \neq \emptyset$.

Przykłady:

$$a) \quad \mathcal{F} = \{\emptyset, \Omega\},$$

$$b) \quad \mathcal{F} = 2^\Omega,$$

$$c) \quad \mathcal{F} = \{A \subseteq \Omega : A \text{ lub } A^c \text{ jest przeliczalny}\},$$

d) \mathcal{F} jest rodziną wszystkich skończonych sum parami rozłącznych przedziałów typu $(a, b]$, $0 \leq a \leq b \leq 1$,

e) \mathcal{F} jest rodziną wszystkich skończonych sum parami rozłącznych przedziałów postaci $(a, b]$, $-\infty < a \leq b < \infty$.

Przykłady (a-c) są σ -ciałami, (d) to tylko ciało w $\Omega = (0, 1]$ (nie jest σ -ciałem, bo nieskończona suma rozłącznych przedziałów $(\frac{1}{2n+1}, \frac{1}{2n})$, $n = 1, 2, \dots$, nie należy do tej rodziny). W przykładzie (e) $\Omega = (-\infty, \infty) \notin \mathcal{F}$, więc to nie jest nawet ciało (jest to pierścień).

Zauważmy, że część wspólna dowolnej ilości σ -ciał jest σ -ciałem. Dla dowolnej rodziny $\mathcal{A} \subseteq 2^\Omega$, oznaczmy przez $\sigma(\mathcal{A})$ najmniejsze σ -ciało zawierające (generowane przez) \mathcal{A} , to znaczy, część wspólna wszystkich *sigma*-ciał zawierających \mathcal{A} .

W przykładach (d) i (e), elementy rodziny $\mathcal{B} = \sigma(\mathcal{F})$ nazywamy zbiorami borelowskimi, odpowiednio, na odcinku $(0, 1]$ i na prostej.

Parę (Ω, \mathcal{F}) , gdzie Ω jest dowolnym zbiorem, a \mathcal{F} jest dowolnym σ -ciałem podzbiorów Ω , nazywa się w teorii miary przestrzenią mierzalną. Miarą określa się każdą nieujemną σ -addytywną funkcję zbiorów. Miarę, która na całej przestrzeni przyjmuje wartość 1, nazywamy *prawdopodobieństwem* i oznaczamy przez P . Zatem prawdopodobieństwo $P : \mathcal{F} \rightarrow [0, \infty)$ spełnia aksjomaty

1. (P1) $P(\Omega) = 1$
2. (P2) $P(\bigcup A_n) = \sum P(A_n)$ dla rozłącznych zdarzeń A_1, A_2, \dots

Z aksjomatu (P2) wynika, że $P(\emptyset) = 0$.

W ten sposób doszliśmy do aksjomatycznej definicji przestrzeni probabilistycznej.

Definicja. Przestrzenią probabilistyczną nazywamy trójkę (Ω, \mathcal{F}, P) , gdzie (Ω, \mathcal{F}) jest przestrzenią mierzalną, tzn. \mathcal{F} jest rodziną podzbiorów Ω i spełnia aksjomaty (F1-F3), a $P : \mathcal{F} \rightarrow [0, 1]$ jest prawdopodobieństwem, tzn. spełnia aksjomaty (P1) i (P2).

Wszystkie dotychczas udowodnione własności prawdopodobieństwa P (4-10) w przypadku skończonym przenoszą się na dowolne przestrzenie. Własności (1-3) są teraz aksjomatami i nie wymagają dowodu. Dodatkowo, addytywność (3) i nierówność Boole'a (8) rozciągają się na przeliczalny ciąg zdarzeń. Podamy teraz dwie dalsze własności prawdopodobieństwa o ważnym znaczeniu.

4.4 Własności prawdopodobieństw monotonicznych ciągów zdarzeń

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n) \quad (4.1)$$

dla $A_1 \subseteq A_2, \dots$, i

$$P\left(\bigcap_{n=1}^{\infty} A_n\right) = \lim_{n \rightarrow \infty} P(A_n) \quad (4.2)$$

dla $A_1 \supseteq A_2, \dots$

Dowód. Z aksjomatu (A2) i własności (4a) wynika, że

$$P\left(\bigcup_n A_n\right) = \sum_n (P(A_n) - P(A_{n-1})) = \lim_{n \rightarrow \infty} P(A_n),$$

co dowodzi (4.1). Wzór (4.2) można wywnioskować z (4.1), bowiem $P\left(\bigcap_n A_n\right) = 1 - P\left(\bigcup_n A_n^c\right)$. \square

4.5 Powrót do odcinka $(0, 1]$

Mając zdefiniowaną funkcję P na odcinkach rozszerzamy ją najpierw na wszystkie skończone sumy rozłącznych przedziałów, a następnie na wszystkie zbiory borelowskie stosując twierdzenie Lebesgue'a o jednoznacznym rozszerzaniu miary z ciała na σ -ciało generowane przez nie. Jest to więc konstrukcja czysto teoretyczna, a otrzymane prawdopodobieństwo P jest miarą Lebesgue'a λ . Żeby obliczyć $P(A)$ dla konkretnego zdarzenia A trzeba się sporo natrudzić, a czasem jest to wręcz niewykonalne.

Zadanie. Pokazać, że zbiór Cantora C jest borelowski i obliczyć $P(C)$.

Rozwiązanie Korzystając z rekurencyjnej definicji zbioru Cantora, oznaczymy przez A_n sumę 2^n przedziałów długości 3^{-n} otrzymanych po n krokach procedury definiującej zbiór Cantora. Mamy $C = \bigcap_n A_n$, więc jest to zbiór borelowski (bo jest przeliczalnym iloczynem skończonych sum przedziałów). Ponadto $A_n \supset A_{n+1}$, zatem z (12)

$$P(C) = \lim_{n \rightarrow \infty} P(A_n) = \lim_{n \rightarrow \infty} (2/3)^n = 0. \square$$

Podobnie wprowadza się prawdopodobieństwo na podzbiorach borelowskich dowolnego odcinka $(c, d]$ (trzeba po prostu przeskalować miarę Lebesgue'a, dzieląc ją przez $d - c$).

Trochę bardziej skomplikowany jest przypadek, gdy $\Omega = (-\infty, \infty)$. Tutaj trzeba skorzystać z miary Lebesgue'a-Stjeltjesa na prostej. Powrócimy do tego zagadnienia w rozdziale 7.

W „porządnym” przestrzeniach nieprzeliczalnych zwykle $P(\omega) \equiv 0$. Teoretycznie nie można wykluczyć atomów, tzn. takich ω , dla których $P(\omega) > 0$. Nie może być ich jednak zbyt wiele.

Własność 1 *Zbiór atomów jest co najwyżej przeliczalny.*

Dowód. Niech $T_i = \{\omega : P(\omega) > 1/i\}$. Wtedy $\bigcup_i T_i$ jest zbiorem wszystkich atomów, i jeśli ten byłby nieprzeliczalny, to dla pewnego i , T_i musiałby być

nieprzeliczalny. Niech T'_i będzie przeliczalnym podzbiorem takiego zbioru T_i . Wtedy otrzymujemy sprzeczność

$$1 \geq P(T'_i) = \sum_{\omega \in T'_i} P(\omega) > \sum_{\omega \in T'_i} 1/i = \infty . \square$$

4.6 Prawdopodobieństwo geometryczne.

Najpierw trochę teorii: miara Lebesgue'a uogólnia znaną ze szkoły miarę Jordana (długość, pole, objętość). Jeśli $\Omega \subset R^k$, $\lambda(\Omega) < \infty$, to $(\Omega, \mathcal{B}_\Omega, P)$, gdzie \mathcal{B}_Ω jest rodziną wszystkich zbiorów borelowskich zawartych w Ω , a

$$P(A) = \frac{\lambda(A)}{\lambda(\Omega)} ,$$

jest przestrzenią probabilistyczną. Poprawność tej definicji wynika stąd, że zbiory należące do rodziny \mathcal{B}_Ω są mierzalne w sensie Lebesgue'a. (Nie są to wszystkie zbiory mierzalne w sensie Lebesgue'a, ale po uzupełnieniu przestrzeni $(\Omega, \mathcal{B}_\Omega, P)$ o zbiory A takie, że $A \subset B$, $B \in \mathcal{B}_\Omega$, gdzie $\lambda(B) = 0$, obie rodziny pokrywają się.) Przestrzeń $(\Omega, \mathcal{B}_\Omega, P)$ opisuje sytuację, gdy losujemy punkt z Ω w sposób jednostajny, tzn. nie wyróżniając żadnego punktu. Jest to klucz do zadań na prawdopodobieństwo geometryczne.

Przykład (Paradoks Bertranda) Na koło rzucono losowo cięciwę. Obliczyć prawdopodobieństwo, że cięciwa jest dłuższa od boku trójkąta równobocznego wpisanego w to koło.

Mamy tu 3 różne odpowiedzi. W (a) i (b) położenie cięciwy jest wyznaczone przez jej środek. W (c) poprzez kąt środkowy jaki tworzy.

(a) Ω – zbiór wszystkich punktów koła

$$P = \frac{\pi(r/2)^2}{\pi r^2} = \frac{1}{4} ,$$

(b) Ω – zbiór wszystkich punktów ustalonego promienia

$$P = \frac{r/2}{r} = \frac{1}{2} ,$$

(c) $\Omega = [0, \pi]$

$$P = \frac{\pi/3}{\pi} = \frac{1}{3} .$$

Rozdział 5

Prawdopodobieństwo warunkowe

Przykład. Rzut kostką. Zdarzenia: A – co najmniej 3, B – co najmniej 5, C – parzysta liczba oczek. Wtedy $P(B) = \frac{1}{3}$ i $P(C) = \frac{1}{2}$. Przypuśćmy jednak, że ktoś, nie podając wyniku tego eksperymentu, ujawnił nam, że zaszło zdarzenie A . Przy tej częściowej informacji, zdarzenie B ma prawdopodobieństwo równe $\frac{1}{2}$. Onacza się je symbolem $P_A(B)$. Zatem

$$\frac{1}{3} = P(B) \neq P_A(B) = \frac{1}{2}.$$

Ale $P(C) = P_A(C)$. Stąd wniosek, że pod wpływem informacji, prawdopodobieństwa mogą się zmieniać lub nie.

Wzór na prawdopodobieństwo warunkowe można uzasadnić w oparciu o definicję częstościową. Przypuśćmy, że w wyniku N -krotnego wykonania tego samego eksperymentu losowego, n razy zaszło zdarzenie A , w tym k razy zaszło również zdarzenie B . Poszczególne częstości wynoszą więc:

$$f_A = \frac{n}{N} \quad f_{AB} = \frac{k}{N} \quad f_{B|A} = \frac{k}{n}$$

Stąd,

$$f_{B|A} = \frac{f_{AB}}{f_A}$$

Inne uzasadnienie ma charakter geometryczny (rysunek). Wiedząc, że zaszło zdarzenie A , cała masa prawdopodobieństwa (równa 1) przeszła z Ω na A , tak że $P_A(A) = 1$. Z kolei, dla dowolnego zdarzenia B , $P_A(B)$ powinno być proporcjonalne do $P(AB)$, to znaczy $P_A(B) = sP(AB)$, gdzie s jest współczynnikiem proporcjonalności. Wyznaczamy go z warunku $1 = P_A(A) = sP(A)$, czyli $s = 1/P(A)$. Ostatecznie, prawdopodobieństwo zdarzenia B względem (pod warunkiem) zdarzenia A definiujemy wzorem

$$P_A(B) = \frac{P(AB)}{P(A)}.$$

Często będziemy używać oznaczenia $P_A(B) = P(B|A)$.

Powstaje nowa (warunkowa) przestrzeń probabilistyczna (A, \mathcal{F}_A, P_A) , gdzie $\mathcal{F}_A = \{B' : B' = BA, B \in \mathcal{F}\}$.

Kolejny przykład wskaże na symetrię między $P(C|B)$ i $P(B|C)$.

Przykład. Rzut 2 kostkami. B – na pierwszej kostce co najwyżej 3 oczka, C – suma oczek na obu kostkach wynosi 6. Mamy $P(B) = 1/2$, $P(C) = 5/36$, $P(C|B) = 1/6$, $P(B|C) = 3/5$. Zatem

$$P(B)P(C|B) = P(C)P(B|C) .$$

Oznacza to, że obie „korelacje”, C względem B i odwrotnie, wyrażone stosunkami $P(C|B)/P(C)$ i $P(B|C)/P(B)$, wykazują ten sam trend.

Paradoks więźnia. Ten przykład pokazuje jak ostrożnie trzeba się obchodzić z prawdopodobieństwem warunkowym w zastosowaniach do sytuacji rzeczywistych.

Trzech więźniów X, Y i Z czeka w celi na egzekucję, która ma się odbyć następnego dnia rano. Wiedzą jednak, że potężny władca ułaskawi jednego z nich, wybierając szczęśliwca losowo. Losowanie miało miejsce po południu, a wieczorem strażnik, który zna wynik losowania, przynosi więźniom kolację. Nie wolno mu ujawnić nazwiska ułaskawionego. X prosi go jednak by wskazał, który z pozostałych współwięźniów będzie zgładzony. (Przecież w ten sposób nie zdradzi tajemnicy.) Strażnik wskazuje na więźnia Y . X nie może ukryć swej radości. Uważa bowiem, że w tym momencie jego szanse przeżycia wzrosły z $1/3$ do $1/2$. Z drugiej strony wydaje się to dziwne, bo przecież strażnik nie ujawnił żadnej istotnej informacji.

Pozornie rozumowanie więźnia X wydaje się poprawne. Jeśli przez X, Y, Z oznaczymy zdarzenia (elementarne), że poszczególni więźniowie będą ułaskawieni, to

$$P(X^c|Y^c) = \frac{P(X^cY^c)}{P(Y^c)} = \frac{P(Z)}{P(Y^c)} = \frac{1/3}{2/3} = \frac{1}{2}$$

Jednak w rzeczywistości nastąpiła zmiana przestrzeni probabilistycznej. W przypadku bowiem, gdy ułaskawionym jest X , strażnik ma do wyboru 2 odpowiedzi i możemy przyjąć, że wybiera jedną z nich z prawdopodobieństwem $1/2$. Zdarzenie elementarne X rozpada się na dwa atomy, X_Y i X_Z , które wyrażają wybór strażnika. Należy im przypisać prawdopodobieństwa $1/6$. Wtedy

$$P(X^c | \text{strażnik wskaże więźnia } Y) = \frac{P(Z)}{P(Z) + P(X_Y)} = \frac{1/3}{1/3 + 1/6} = \frac{2}{3} ,$$

a więc bez zmian.

Komentarz 1: Gdy strażnik nie preferuje żadnej odpowiedzi, to nie przekazuje więźniowi X żadnej istotnej informacji i prawdopodobieństwo nie może się zmienić. Warto przeanalizować ten przykład przy założeniu, że strażnik preferuje (p vs. $1 - p$) wskazanie więźnia Y , a w szczególności rozważyć skrajne przypadki $p = 0$ i $p = 1$.

Komentarz 2: Ta anegdota ma pewien związek z grą „w trzy karty” opisaną wcześniej. jak należałoby interpretować jej wariant, w którym grający ma możliwość zmiany swojej decyzji?

Definicja. Zdarzenia A_1, A_2, \dots tworzą *układ zupełny zdarzeń*, gdy są parami rozłączne (wykluczające się) oraz $\bigcup_{i=1}^{\infty} A_i = \Omega$.

Twierdzenie 1 *Jeśli A_1, A_2, \dots tworzą układ zupełny, to dla każdego $A \in \mathcal{F}$ zachodzi*

(a) **Wzór na prawdopodobieństwo całkowite:**

$$P(A) = \sum P(A_i)P(A|A_i)$$

(b) **Wzór Bayesa:** $P(A_i|A) = \frac{P(A|A_i)P(A_i)}{P(A)}$.

Dowód.

(a) Na podstawie σ -addytywności miary P , $P(A) = \sum_{i=1}^{\infty} P(AA_i)$.

(b) wynika z definicji prawdopodobieństwa warunkowego. \square

Oba wzory stosuje się w sytuacjach, gdy eksperyment losowy składa się w 2 fazy, z których druga zależy od pierwszej. Wzór (b) jest interesujący z tego względu, że wyznacza prawdopodobieństwo „przyczyny”, która spowodowała zaobserwowany „skutek”.

Przykład. (Rosyjska ruletka). Wśród 9 czekoladek 3 są zatrute (po spożyciu następuje natychmiast zgon). Wraz z przyjacielem bawicie się w taką oto grę: kolejno zjadacie po jednej czekoladce. Obliczyć prawdopodobieństwo, że

- (a) jeśli Ty zaczynasz, to przeżyjesz;
- (b) jeśli Ty zacząłeś i przeżyłeś, to i Przyjaciel przeżyje;
- (c) jeśli Ty zacząłeś i nie przeżyłeś, to Przyjaciel przeżyje;
- (d) jeśli Przyjaciel zaczyna, to Ty przeżyjesz;
- (e) jeśli Ty zaczynasz, to przeżyjesz, wiedząc, że Przyjaciel przeżyje.

Rozwiązanie: Niech A będzie zdarzeniem, że pierwsza osoba przeżyje, a B , że druga.

(a) $P(A) = \frac{2}{3}$

(b) $P(B|A) = \frac{(6 \cdot 5)/(9 \cdot 8)}{2/3} = \frac{5}{8}$

(c) $P(B|A^c) = \frac{(6 \cdot 3)/(9 \cdot 8)}{1/3} = \frac{3}{4}$

(d) $P(B) = P(B|A)P(A) + P(B|A^c)P(A^c) = \frac{5}{8} \cdot \frac{2}{3} + \frac{3}{4} \cdot \frac{1}{3} = \frac{2}{3}$

(e) $P(A|B) = P(B|A)P(A)/P(B) = \frac{5}{8}$

Pytanie (e) wydaje się nie mieć sensu, choć matematycznie jest całkiem poprawne. Można mu nadać praktyczny sens, wprowadzając Obserwatora, który znając wcześniej kolejność w grze, dzwoni chcąc się dowiedzieć o jej rezultat. Telefon odbiera Przyjaciół, po czym połączenie urywa się...

Odpowiedź na pytanie (d) oznacza, że kolejność w tej grze nie ma znaczenia - - szanse obu graczy są równe. To nie przypadek. Ta gra jest sprawiedliwa niezależnie od liczby czekoladek.

Przykład. Urna zawiera M czerwonych i $N - M$ białych kul. Oczywiście, losując jedną kulę, prawdopodobieństwo wylosowania kuli czerwonej wynosi M/N . Pokazać, że losując N -krotnie bez zwracania, prawdopodobieństwo wylosowania kuli czerwonej za k -tym razem, $k = 1, \dots, N$, również wynosi M/N .

Rozwiązanie: Dla $k = 2$, na podstawie wzoru na prawdopodobieństwo całkowite, szukane prawdopodobieństwo wynosi

$$\frac{M-1}{N-1} \frac{M}{N} + \frac{M}{N-1} \frac{N-M}{N} = \frac{M}{N}$$

Dla $k > 2$ można stosować indukcję, ale jest też inny sposób. Zauważmy, że w wszystkich k -elementowych wariacji bez powtórzeń z N -elementowego zbioru kul, w których na ostatnim miejscu jest kula czerwona, jest tyle samo, co k -elementowych wariacji, które rozpoczynają się od kuli czerwonej (bijekcja polega na odwróceniu ciągu).

Przykład. (Wesoły autobus) Autobus według rozkładu powinien podjeżdżać na przystanek o 8:30, 8:45, itd. W rzeczywistości zawsze jest 5 minut wcześniej lub później (obie możliwości są jednakowo prawdopodobne). Empirycznie ustalono, że prawdopodobieństwo, iż na przystanku nie ma nikogo w t minut po odjeździe ostatniego autobusu wynosi $e^{-t/5}$. Przychodzimy na pusty przystanek punktualnie o godz. 9:00. Obliczyć prawdopodobieństwo, że uciekł nam autobus.

Rozwiązanie: Niech A będzie zdarzeniem, że o 9:00 nie ma nikogo na przystanku, B zdarzeniem, że autobus z godz. 9:00 odjechał o 8:55. Wtedy, ze wzoru Bayesa

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)} = \frac{e^{-1/2}}{P(A)}.$$

Uwzględniając wszystkie 4 możliwości rzeczywistych odjazdów autobusów z godz. 8:45 i 9:00, obliczamy $P(A)$ na podstawie wzoru na prawdopodobieństwo całkowite:

$$P(A) = \frac{1}{4}(2e^{-1} + e^{-2} + e^{-4}).$$

Szacując $e^2(e-2) \geq 2$, szukane prawdopodobieństwo jest co najmniej równe 0,8 (a więc zgodne z naszym złym przecuciem).

Na zakończenie tego rozdziału przedstawię dwa dalsze użyteczne wzory:

Iterowane prawdopodobieństwo warunkowe Jak w przestrzeni warunkowej względem zdarzenia A_1 obliczać kolejne prawdopodobieństwo warunkowe, tym razem względem zdarzenia A_2 , tzn. $P(B|A_2|A_1)$? Kierując się analogią ze zwykłym prawdopodobieństwem warunkowym definiujemy je jako

$$P(B|A_2|A_1) = \frac{P(BA_2|A_1)}{P(A_2|A_1)} = P(B|A_1A_2) .$$

Okazuje się więc, że iterowane prawdopodobieństwo warunkowe można zastąpić zwykłym, w którym warunkiem jest koniunkcja obu warunków.

Wzór łańcuchowy

$$P(A_1 \dots A_n | B) = P(A_1 | B) P(A_2 | A_1 B) P(A_3 | A_1 A_2 B) \dots P(A_n | A_1 \dots A_{n-1} B) .$$

W szczególnym przypadku, gdy $B = \Omega$, otrzymujemy wzór

$$P(A_1 \dots A_n) = P(A_1) P(A_2 | A_1) P(A_3 | A_1 A_2) \dots P(A_n | A_1 \dots A_{n-1}) .$$

Rozdział 6

Niezależność zdarzeń

Jeśli prawdopodobieństwo warunkowe $P(A|B)$ i bezwarunkowe $P(A)$ są sobie równe, to w potocznym rozumieniu zdarzenia A i B są niezależne. Rozszerzmy tę definicję, tak by obejmowała również zdarzenia B , dla których $P(B) = 0$.

Mówimy, że zdarzenia A i B są *niezależne*, gdy $P(AB) = P(A)P(B)$.

Przy takiej definicji lepiej też widać jej symetrię.

Przykład. Losujemy rodzinę z trójką (z czwórka) dzieci. Zdarzenie A – co najwyżej 1 dziewczynka; zdarzenie B – co najmniej 1 dziewczynka i co najmniej 1 chłopiec. Sprawdzić niezależność zdarzeń A i B (w obu przypadkach).

Cztery własności.

- (a) Jeśli $P(A) = 0$ lub $P(A) = 1$, to A jest niezależne od każdego zdarzenia.
- (b) Jeśli $0 < P(A) < 1$, to A jest zależne od A .
- (c) Jeśli A i B są rozłączne oraz $P(A)P(B) > 0$, to A i B są zależne.
- (d) Jeśli A i B są niezależne, to również niezależne są pary zdarzeń A i B^c , A^c i B , oraz A^c i B^c .

Dowody:

- (a) Jeśli np. $P(A) = 1$, to $P(AB) = 1 + P(B) - P(A \cup B) \geq P(B)$, skąd $P(AB) = P(B) = P(A)P(B)$.
- (b) Równanie $x = x^2$ nie ma rozwiązania w przedziale $0 < x < 1$.
- (c) Oczywiste
- (d) Wystarczy pokazać tylko dla pierwszej pary:

$$P(AB^c) = P(A \setminus AB) = P(A) - P(AB) = P(A) - P(A)P(B) = P(A)P(B^c) \quad \square$$

Definicja Zdarzenia A_1, \dots, A_n nazywamy niezależnymi (łącznie), jeśli dla każdego $I \subset \{1, 2, \dots, n\}$, $I \neq \emptyset$, zachodzi wzór

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P(A_i) .$$

Musi więc zachodzić $2^n - n - 1$ równości.

Przykład. Rzut 2 kostkami. Zdarzenia: A – pierwsza kostka parzysta, B – druga kostka nieparzysta, C – obie kostki parzyste lub obie nieparzyste. Zdarzenia A, B, C są niezależne parami, ale nie łącznie.

Uwaga. Gdy $A_1 = \emptyset$, to

$$P(A_1 \dots A_n) = 0 = P(A_1) \dots P(A_n)$$

więc łatwo znaleźć przykład zdarzeń zależnych, które spełniają równość

$$P(A_1 \dots A_n) = P(A_1) \dots P(A_n) .$$

Trudniej jest znaleźć taki przykład przy założeniu, że $P(A_i) \neq 0$, $i = 1, \dots, n$.

Własność 2 (e) *Jeśli A_1, \dots, A_n są niezależne, to A_1^c, \dots, A_n^c też są niezależne.*

Dowód. Wystarczy pokazać, że $A_1, \dots, A_{n-1}, A_n^c$ są niezależne (a w efekcie otrzymamy mocniejszą tezę, że A_1^*, \dots, A_n^* są niezależne, gdzie $A_i^* \in \{A_i, A_i^c\}$).

Niech $I \subset [n-1]$. Oznaczmy $B_I = \bigcap_{i \in I} A_i$. Wtedy z niezależności A_1, \dots, A_n , która implikuje niezależność A_1, \dots, A_{n-1} , wynika, że

$$P(B_I) = \prod_{i \in I} P(A_i)$$

Trzeba więc tylko pokazać, że

$$P(B_I A_n^c) = \prod_{i \in I} P(A_i) P(A_n^c)$$

Ale B_I i A_n są niezależne (pokazać) i na podstawie własności (d), również B_I i A_n^c są niezależne. Zatem

$$P(B_I A_n^c) = P(B_I) P(A_n^c) = \prod_{i \in I} P(A_i) P(A_n^c)$$

□

Definicja Nieskończony ciąg zdarzeń nazywamy *niezależnym*, gdy każdy jego skończony podzbiór (równoważnie, każdy początkowy, skończony fragment) jest niezależny.

Zadanie. Pokazać, że w przestrzeni dyskretnej, dla żadnego $0 < \epsilon < 1/2$, nie istnieje nieskończony ciąg zdarzeń niezależnych, z których każde ma prawdopodobieństwo w przedziale $(\epsilon, 1 - \epsilon)$.

6.1. ZASTOSOWANIE NIEZALEŻNOŚCI ZDARZEŃ W TEORII LICZB31

Dowód: Niech $\omega \in \Omega$ i $p(\omega) = p_0 > 0$. Wybierzmy n tak, by $p_0 > (1-\epsilon)^n$. Przypuśćmy nie wprost, że istnieje nieskończony ciąg zdarzeń niezależnych A_1, A_2, \dots , taki, że $1 - \epsilon > P(A_i) > \epsilon$. Zdarzenia $A_1^* \cap \dots \cap A_n^*$, gdzie $A_i^* \in \{A_i, A_i^c\}$, $i = 1, \dots, n$, tworzą układ zupełny zdarzeń i mają prawdopodobieństwa nie większe niż $(1 - \epsilon)^n$. Zatem ω nie może należeć do żadnego z nich – sprzeczność. \square

6.1 Zastosowanie niezależności zdarzeń w teorii liczb

Funkcję Eulera definiujemy wzorem

$$\phi(n) = |\{x : 1 \leq x \leq n, (x, n) = 1\}|$$

Na przykład, $\phi(12) = 4$, $\phi(14) = 6$. Wyrazimy $\phi(n)$ poprzez dzielniki pierwsze liczby n . Niech $n = p_1^{\alpha_1} \dots p_k^{\alpha_k}$ będzie rozkładem liczby n na czynniki pierwsze. Wprowadźmy klasyczną przestrzeń probabilistyczną na zbiorze $\Omega = [n]$, którą interpretujemy jako wybór losowej liczby naturalnej z przedziału od 1 do n . Niech A_i będzie zdarzeniem, że wylosowana liczba dzieli się przez p_i , $i = 1, \dots, k$, i niech $A = \bigcap_{i=1}^k A_i^c$. Wtedy

$$P(A) = \frac{\phi(n)}{n}$$

Żeby wyprowadzić wzór na $\phi(n)$ wystarczy znaleźć $P(A)$. Jest to przykład tzw. metody probabilistycznej, bardzo ostatnio popularnej w kombinatoryce i teorii liczb. Pozwala ona rozwiązać problem czysto deterministyczny (nielosowy) przez wprowadzenie pomocniczej przestrzeni probabilistycznej i skorzystanie z własności prawdopodobieństwa.

Mamy $P(A_i) = \frac{n/p_i}{n} = \frac{1}{p_i}$, $P(A_i A_j) = \frac{n/p_i p_j}{n} = \frac{1}{p_i p_j}$, itd. Zatem zdarzenia A_1, \dots, A_k są niezależne i na podstawie własności (e) również A_1^c, \dots, A_k^c są niezależne. Ostatecznie,

$$\phi(n) = nP(A) = n \left(1 - \frac{1}{p_1}\right) \dots \left(1 - \frac{1}{p_k}\right)$$

6.2 Przestrzenie produktowe

Dla dwóch lub więcej eksperymentów losowych, które nie mają na siebie wpływu (są niezależne w potocznym sensie), a opisane są różnymi przestrzeniami probabilistycznymi, można zawsze stworzyć wspólną przestrzeń, w której zdarzenia związane z różnymi eksperymentami będą niezależne. Tę wspólną przestrzeń probabilistyczną nazywamy *produktową*.

Przykład: Kostka + moneta

Z przestrzeni $\{O, R\}$ i $\{1, 2, 3, 4, 5, 6\}$ tworzymy nową przestrzeń $\{(O, 1), \dots, (O, 6), (R, 1), \dots, (R, 6)\}$, gdzie $P(O, i) = P(R, i) = \frac{1}{12}$, $i = 1, \dots, 6$. Na przykład, zdarzenia O , że na monecie wypadł orzeł i A_1 , że na kostce wypadła jedynka, są niezależne, bowiem $P(O) = 6(\frac{1}{12}) = \frac{1}{2}$, $P(A_1) = 2(\frac{1}{12}) = \frac{1}{6}$ oraz $P(OA_1) = P(\{(O, 1)\}) = \frac{1}{12}$.

Przypadek skończony. Niech $(\Omega_1, p_1), \dots, (\Omega_n, p_n)$ będą przestrzeniami skończonymi. *Przestrzenią produktową* tych przestrzeni nazywamy parę (Ω, p) , gdzie $\Omega = \Omega_1 \times \dots \times \Omega_n$, a $p(\omega_1, \dots, \omega_n) = p_1(\omega_1) \dots p_n(\omega_n)$.

Wtedy, jeśli $A = A_1 \times \dots \times A_n$, gdzie $A_i \subseteq \Omega_i$, to $P(A) = P(A_1) \dots P(A_n)$.

Ta obserwacja jest punktem wyjścia definicji przestrzeni produktowej w przypadku ogólnym.

Jeśli dane są przestrzenie probabilistyczne $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, \dots, n$, to przestrzeń produktową tworzymy następująco:

$$\Omega = \Omega_1 \times \dots \times \Omega_n$$

$$\mathcal{F} = \sigma(\mathcal{A})$$

gdzie $\mathcal{A} = \{A_1 \times \dots \times A_n : A_i \in \Omega_i\}$. Miara P jest (jedynym) rozszerzeniem miary P' określonej na \mathcal{A} wzorem

$$P'(A_1 \times \dots \times A_n) = P_1(A_1) \dots P_n(A_n)$$

Przykład: prostokąt jednostkowy w R^n . Prostokąt jednostkowy w R^n jest przestrzenią produktową n kopii przestrzeni probabilistycznej zbudowanej na odcinku $[0, 1]$ przy pomocy miary Lebesgue'a.

W przypadku nieskończonego ciągu przestrzeni $(\Omega_i, \mathcal{F}_i, P_i)$, $i = 1, 2, \dots$, przestrzeń produktowa jest definiowana tak: $\Omega = \Omega_1 \times \Omega_2 \times \dots$;

$$\mathcal{F} = \sigma(\mathcal{A})$$

gdzie

$$\mathcal{A} = \{A_1 \times A_2 \times \dots : A_i \in \Omega_i \text{ oraz } A_i = \Omega_i \text{ poza skończoną liczbą wskaźników } i\}$$

jest rodziną cylindrów. Miara P jest (jedynym) rozszerzeniem miary P' określonej na \mathcal{A} wzorem

$$P'(A_1 \times A_2 \times \dots) = \prod_{i=1}^{\infty} P_i(A_i)$$

Przykład: Nieskończony ciąg rzutów niesymetryczną monetą. Tutaj $\Omega_i = \{0, 1\}$, $i = 1, 2, \dots$, 1 – orzeł, 0 – reszka, $p(1) = p$, $p(0) = q = 1 - p$. Wtedy prawdopodobieństwo, że pierwsze n rzutów da wynik $b_1 b_2 \dots b_n$ (to jest właśnie cylinder) wynosi $p^{\sum_{i=1}^n b_i} q^{n - \sum_{i=1}^n b_i}$. Wiemy już, że ta przestrzeń jest równoważna przestrzeni probabilistycznej zbudowanej na odcinku $[0, 1]$ przy pomocy miary Lebesgue'a.

6.3 Schemat Bernoulliego

$\Omega_i = \{0, 1\}$, $i = 1, \dots, n$, 0 – porażka, 1 – sukces, $p(1) = p$, $p(0) = q = 1 - p$.
 $\Omega = \Omega_1 \times \dots \times \Omega_n$; dla każdego $\omega = (\omega_1, \dots, \omega_n) \in \Omega$, mamy

$$p(\omega) = p^{\sum_{i=1}^n \omega_i} q^{n - \sum_{i=1}^n \omega_i}$$

Niech $A_k = \{\omega : \sum_{i=1}^n \omega_i = k\}$. tzn. A_k jest zdarzeniem, że zaszło dokładnie k sukcesów. Zauważmy, że $A_k \notin \mathcal{A}$. Mamy

$$P(A_k) = \sum_{\omega \in A_k} p(\omega) = \binom{n}{k} p^k q^{n-k}$$

Te liczby tworzą rozkład Bernoulliego (dwumianowy) i sumują się do 1 (wzór Newtona). Zdarzenia A_1, \dots, A_n tworzą układ zupełny.

Rozdział 7

Zmienne losowe i ich rozkłady

Zmienne losowe są pomostem pomiędzy dowolną przestrzenią probabilistyczną (Ω, \mathcal{F}, P) a światem lepiej nam znanym, tj. światem liczb. Są to funkcje postaci

$$X : \Omega \rightarrow R^n$$

przypisujące każdemu zdarzeniu elementarnemu wektor liczbowy, a w przypadku $n = 1$, liczbę. Jest to pierwszy, najważniejszy krok na drodze do odtworzenia przestrzeni (Ω, \mathcal{F}, P) na gruncie R^n . Tę nową przestrzeń probabilistyczną tworzy trójka (R, \mathcal{B}^n, P_X) , gdzie \mathcal{B}^n jest rodziną zbiorów borelowskich w R^n , a prawdopodobieństwo P_X jest wyznaczona poprzez wyjściowe prawdopodobieństwo P i zmienną losową X . W ogólności, NIE KAŻDA funkcja $X : \Omega \rightarrow R^n$ jest zmienną losową, ale gdy $\mathcal{F} = 2^\Omega$, to każda.

Zmienną losową, określoną na przestrzeni probabilistycznej (Ω, \mathcal{F}, P) , nazywamy dowolną funkcję $X : \Omega \rightarrow R^n$, spełniającą warunek, że dla każdego zbioru borelowskiego $B \in \mathcal{B}^n$, przeciwobraz $X^{-1}(B) \in \mathcal{F}$, to znaczy jest zdarzeniem w przestrzeni (Ω, \mathcal{F}, P) .

Warunek ten musi zachodzić, żeby móc zdefiniować miarę P_X na \mathcal{B} . Definiujemy ją bowiem następująco:

$$P_X(B) = P(X^{-1}(B)) = P(X \in B) = P(\{\omega : X(\omega) \in B\})$$

Jeśli $\mathcal{F} = 2^\Omega$, to warunek ten jest zawsze spełniony.

Miarę prawdopodobieństwa P_X nazywamy *rozkładem (prawdopodobieństwa) zmiennej losowej X* . *Atomem rozkładu P_X* nazywamy każdą liczbę rzeczywistą x , dla której $P_X(\{x\}) > 0$. Atomów może być co najwyżej przeliczalnie wiele.

W przypadku skończonym lub przeliczalnym, tzn. gdy $X(\Omega) = \{x_1, x_2, \dots\}$ (możemy tu przyjąć, że wszystkie x_i są atomami), ciąg $p_i = P_X(\{x_i\}) = P(\{\omega : X(\omega) = x_i\})$ wyznacza miarę P_X jednoznacznie wzorem

$$P_X(B) = \sum_{x_i \in B} P_X(\{x_i\}) . \tag{7.1}$$

Dlatego też ten ciąg (wraz z ciągiem (x_1, x_2, \dots)) nazywamy *rozkładem zmiennej losowej* X . Mamy wtedy $\sum_i p_i = 1$ oraz $p_i \geq 0$. Każdy nieujemny ciąg a_1, a_2, \dots taki, że $\sum_i a_i = 1$, jest rozkładem jakiejś zmiennej losowej.

rozkładem?

Zmienną losową X , której obraz $X(\Omega)$ jest skończony lub przeliczalny będziemy nazywać dyskretną. Wkrótce nieznacznie rozszerzymy to pojęcie.

Przykłady: (tam, gdzie można, z wykresami rozkładów)

1a Rzut kostką: $X(i \text{ oczek}) = i$;

1b Rzut kostką: $Y(i \text{ oczek}) = 1$, gdy i jest parzyste, $Y(i \text{ oczek}) = 0$ w przeciwnym razie;

2a Rzut dwoma kostkami: $X(i \text{ oczek}, j \text{ oczek}) = (i, j)$ – zmienna losowa dwuwymiarowa;

2b Rzut dwoma kostkami: $Y(i \text{ oczek}, j \text{ oczek}) = i + j$;

3a n rzutów monetą: $a_i = 1$, gdy w i -tym rzucie wypadnie orzeł, $a_i = 0$, w przeciwnym razie; $X(a_1, \dots, a_n) = (a_1, \dots, a_n)$ – zmienna losowa n -wymiarowa;

3b n rzutów monetą: $X(a_1, \dots, a_n) = a_1 + \dots + a_n$ (X to liczba orłów, $p_k = \frac{\binom{n}{k}}{2^n}$ – rozkład dwumianowy, symetryczny);

3c Schemat Bernoulliego (n prób): X jak wyżej, tyle, że prawdopodobieństwo sukcesu wynosi p , a porażki $q = 1 - p$; wtedy X to liczba sukcesów – rozkład dwumianowy, a $p_k = \binom{n}{k} p^k q^{n-k}$;

4 Nieskończony ciąg rzutów monetą ($a_i = 1$, gdy wypadł orzeł i $a_i = 0$, gdy reszka): $X(a_1, a_2, \dots) = \sum_i \frac{a_i}{2^i}$ – współrzędna punktu z odcinka $(0, 1]$. Ten rozkład nie ma atomów, więc trzeba go definiować inaczej (patrz poniżej).

5 Wybór punktu z kwadratu jednostkowego: $X(x, y) = \sqrt{x^2 + y^2}$ jest odległością punktu od początku układu współrzędnych.

6 Nieskończony ciąg rzutów monetą ($a_i = 1$, gdy wypadł orzeł i $a_i = -1$, gdy reszka) – tzw. błądzenie losowe po osi liczbowej:

$$X(a_1, a_2, \dots) = \begin{cases} \min\{n : \sum_1^n a_i = 0\} & \text{gdy takie } n \text{ istnieje} \\ \infty & \text{w przeciwnym razie} \end{cases}$$

– czas pierwszego powrotu; to przestrzeń nieprzeliczalna, ale obraz $X(\Omega)$ jest przeliczalny, więc X jest dyskretna.

W przypadku, gdy $X(\Omega)$ jest zbiorem nieprzeliczalnym (przykłady 4 i 5), najczęściej nie ma w ogóle atomów. Wtedy jednak wystarczy znać miarę P_X na przedziałach dowolnego typu, by w pełni ją zrekonstruować. W przestrzeniach wielowymiarowych rolę przedziałów pełnią prostopadłościany. Na razie ograniczymy się do przypadku 1-wymiarowego ($n = 1$) i przyjmiemy przedziały typu $(-\infty, x]$. Są one o tyle wygodniejsze od przedziałów typu (a, b) , że występuje w nich tylko jeden argument (x), a nie dwa.

Niech $F_X(x) = P_X((-\infty, x]) = P(\{\omega : X(\omega) \leq x\})$. Wtedy $P_X(a, b] = F_X(b) - F_X(a)$. Jeśli tylko funkcja F_X jest niemalejąca i różnica między jej granicami w $+\infty$ i $-\infty$ wynosi 1, to stosuje się twierdzenie Lebesgue'a o rozszerzeniu miary i można jednoznacznie określić miarę P_X na całej rodzinie borelowskiej \mathcal{B} – tzw. miara Lebesguea- Stieltjesa. Oczywiście, w przypadku dyskretnym, funkcja F_X wyznacza bezpośrednio rozkład P_X poprzez wzór (7.1) i równanie

$$p_i = P_X(x_i) = F_X(x_i) - F_X(x_{i-1})$$

To ostatnie równanie jest szczególnym przypadkiem wzoru, który pozwala wyznaczyć $P_X(x)$ dla każdego $x \in R$. Wzór ten ma postać

$$P_X(x) = F_X(x) - \lim_{h \rightarrow 0^+} F_X(x - h) \quad (7.2)$$

i zostanie udowodniony później.

Funkcję $F_X(x) = P_X((-\infty, x]) = P(\{\omega : X(\omega) \leq x\}) = P(X \leq x)$ nazywamy *dystrybuantą* (rozkładu (prawdopodobieństwa)) zmiennej losowej X .

Zatem dystrybuanta jednoznacznie wyznacza rozkład zmiennej losowej, ale nie ją samą.

Uwaga: Różne zmienne losowe mogą mieć ten sam rozkład. Rozważmy zmienną losową $X = 1$, gdy wypadnie orzeł i $X = 0$, gdy wypadnie reszka. Weźmy $Y = 1 - X$. Oczywiście $X \neq Y$, ale $F_X = F_Y$.

Rysując dystrybuanty do paru przykładów (powiedzmy 1 i 4) można zauważyć pewne wspólne własności: rośnie od 0 do 1, jest prawostronnie ciągła.

Twierdzenie 2 (własności dystrybuanty) *Dystrybuanta dowolnej zmiennej losowej jest*

- (a) *niemalejąca*
- (b) *prawostronnie ciągła*
- (c) *jej granice w $+\infty$ i $-\infty$ są równe odpowiednio 1 i 0.*

Dowód.

(a) Jeśli $x < y$, to zdarzenie $\{X \leq x\} \subseteq \{X \leq y\}$ i na podstawie własności (5) prawdopodobieństwa

$$F_X(x) = P(X \leq x) \leq P(X \leq y) = F_X(y)$$

(b) Weźmy ciąg $h_n \rightarrow 0$ w sposób malejący (np. $h_n = 1/n$) i oznaczmy przez B_n przedział $(x, x + h_n]$, $n = 1, 2, \dots$. Zauważmy, że

$$P_X(B_n) = F(x + h_n) - F(x)$$

oraz, że ciąg (B_n) tworzy zstępujący ciąg zdarzeń. Zatem, z jednej strony, na podstawie wzoru (12),

$$\lim_{n \rightarrow \infty} P_X(B_n) = P_X\left(\bigcap_{n=1}^{\infty} B_n\right) = P_X(\emptyset) = 0$$

a z drugiej strony

$$\lim_{n \rightarrow \infty} P_X(B_n) = \lim_{n \rightarrow \infty} F(x + h_n) - F(x) .$$

Ostatecznie,

$$\lim_{n \rightarrow \infty} F(x + h_n) = F(x) ,$$

co oznacza, że granica prawostronna w punkcie x zgadza się z wartością funkcji F_X w tym punkcie.

(c) Niech $x_n \rightarrow \infty$ w sposób rosnący. Wtedy, na podstawie wzoru (11)

$$\lim_{n \rightarrow \infty} F(x_n) = P_X\left(\bigcup_{n=1}^{\infty} (-\infty, x_n]\right) = P_X(R) = 1 .$$

Podobnie, jeśli $x_n \rightarrow -\infty$ w sposób malejący, to

$$\lim_{n \rightarrow \infty} F(x_n) = P_X\left(\bigcap_{n=1}^{\infty} (-\infty, x_n]\right) = P_X(\emptyset) = 0 .$$

Uwaga: Każda funkcja rzeczywista spełniająca warunki (a-c) jest dystrybuantą jakiegoś rozkładu.

Obserwacja: Atomy rozkładu P_X pokrywają się z punktami skokowymi dystrybuanty F_X , tzn. z tymi punktami $x \in R$, dla których granica lewostronna jest mniejsza od wartości funkcji F_X w tym punkcie. Rzeczywiście,

$$P_X(\{x\}) = P_X\left(\bigcap_{n=1}^{\infty} \left(x - \frac{1}{n}, x\right]\right) = \lim_{n \rightarrow \infty} P_X\left(\left(x - \frac{1}{n}, x\right]\right) = F_X(x) - \lim_{n \rightarrow \infty} F_X\left(x - \frac{1}{n}\right) ,$$

co jest równoważne ze wzorem (2).

7.1 Klasyfikacja rozkładów

7.1.1 Podział na rozkłady dyskretne i ciągłe

Rozkład zmiennej losowej X nazywamy *dyskretnym*, gdy istnieje przeliczalny zbiór atomów, taki że $\sum_{n=1}^{\infty} P_X(\{x_n\}) = 1$ (inaczej: dystrybuanta jest funkcją schodkową).

Rozkład zmiennej losowej X nazywamy *ciągłym*, gdy nie ma atomów (inaczej: dystrybuanta jest funkcją ciągłą).

Wzrost dystrybuanty od 0 do 1 odbywa się w przypadku dyskretnym wyłącznie skokowo, a w przypadku ciągłym, wyłącznie w sposób ciągły. Ale są też przypadki pośrednie.

Twierdzenie 3 (Twierdzenie Jordana) *Każdy rozkład (dystrybuanta) zmiennej losowej jest wypukłą kombinacją liniową rozkładu (dystrybuanty) dyskretnego i ciągłego, tzn. dla pewnego $0 \leq a \leq 1$*

$$P_X = aP_{X_d} + (1 - a)P_{X_c},$$

równoważnie,

$$F_X = aF_{X_d} + (1 - a)F_{X_c},$$

Dowód: Niech $a = \sum_k p_k \in (0, 1)$, gdzie sumowanie przebiega po wszystkich atomach x_1, x_2, \dots rozkładu P_X . Zdefiniujmy funkcje

$$F_d(x) = \frac{1}{a} \sum_{x_k \leq x} p_k$$

i

$$F_c(x) = \frac{F(x) - aF_d(x)}{1 - a}.$$

Wtedy F_d jest dystrybuantą dyskretną, a F_c ciągłą (pokazać na ćwiczeniach).

□

Przykład: rozkład mieszany. Strzał do tarczy, za który, jako wygraną, otrzymuje się -1 punkt, jeśli się chybi (z prawdopodobieństwem $1/2$), a w przeciwnym razie miarę ϕ kąta utworzonego przez ślad strzału z ustalonym promieniem, $\phi \in [0, 2\pi]$. Znaleźć dystrybuantę wygranej i przedstawić ją jako kombinację liniową zgodnie z tw. Jordana (ćwiczenia).

7.1.2 Podział na rozkłady absolutnie ciągłe i osobliwe

Rozkład zmiennej losowej X nazywamy *absolutnie ciągłym*, gdy dla każdego $B \in \mathcal{B}$, jeśli $\lambda(B) = 0$ to $P_X(B) = 0$.

Rozkład zmiennej losowej X nazywamy *osobliwym*, gdy istnieje zbiór $B \in \mathcal{B}$, taki że $\lambda(B) = 0$ a $P_X(B) = 1$.

Uwaga: każdy rozkład dyskretny jest osobliwy. Naprawdę osobliwe są rozkłady ciągłe i osobliwe zarazem.

Przykład: rozkład osobliwy, ciągły (szczegóły na ćwiczeniach).

Rozpatrzmy grę opartą o nieskończony ciąg rzutów monetą. Jeśli za k -tym razem wypadł orzeł, to wygrana wynosi $2/3^k$, w przeciwnym razie 0. Niech X będzie łączną wygraną w tej grze. Mamy $0 \leq X \leq 1$ (Sprawdzić, że to jest zmienna losowa). Narysować dystrybuantę: jest schodkowa w dopełnieniu zbioru Cantora, a więc na zbiorze Cantora ma miarę $P_X(C) = 1$, ale wiemy już, że $\lambda(C) = 0$. Jest to więc rozkład osobliwy. Z drugiej strony jest on ciągły, bo jeśli $0 \leq x - y \leq 3^{-n}$ to $F_X(x) - F_X(y) \leq 2^{-n}$, a to znaczy, że funkcja $F_X(x)$ jest ciągła.

Twierdzenie 4 (Twierdzenie Lebesgue'a) *Każdy rozkład (dystrybuanta) zmiennej losowej jest wypukłą kombinacją liniową rozkładu (dystrybuanty) osobliwego i absolutnie ciągłego, tzn. dla pewnego $0 \leq a \leq 1$*

$$P_X = aP_{X_o} + (1 - a)P_{X_{ac}}.$$

Nakładając na siebie oba twierdzenia otrzymujemy wniosek, że

$$P_X = aP_{X_d} + bP_{X_{ac}} + (1 - a - b)P_{X_{oc}} .$$

Od tej pory będziemy zajmować się prawie wyłącznie rozkładami czysto dyskretnymi lub czysto absolutnie ciągłymi (te drugie wkrótce zaczniemy nazywać ciągłymi, o ile nie będzie to prowadzić do nieporozumień).

7.2 Przykłady rozkładów

7.2.1 Rozkłady dyskretne

Rozkład dyskretny jest w pełni określony przez podanie ciągu p_1, p_2, \dots nieujemnych liczb rzeczywistych takich, że $\sum_n p_n = 1$. Wtedy $p_n = P_X(\{x_n\})$. Każdy taki ciąg jest (wyznacza) rozkładem jakiejś dyskretnej zmiennej losowej.

1) Zdegenerowany (parametr c): $P(X = c) = 1$; dystrybuanta ma jeden skok (w punkcie c)

2) 2-punktowy (3 parametry a, b, p): $P(X = a) = p$, $P(X = b) = q = 1 - p$, $a \neq b$; dwa skoki (w a i b); gdy $a = 1$ i $b = 0$, to czasem nazywa się go rozkładem Bernoulliego i oznacza przez $Be(p)$.

3) Dwumianowy (2 parametry n i p): $p_k = P(X = k) = \binom{n}{k} p^k q^{n-k}$, $k = 0, 1, \dots, n$; $n + 1$ skoków (X liczy sukcesy w schemacie Bernoulliego); oznaczenie $Bi(n, p)$; nazwa stąd, że $\sum_k p_k = (p + q)^n = 1$.

4) Poissona (parametr λ): $P(X = k) = e^{-\lambda} \lambda^k / k!$, $k = 0, 1, \dots$; przeliczalny zbiór wartości (skoków, atomów); jest granicą ciągu rozkładów $Bi(n, \lambda/n)$ (nazwa pochodzi od matematyka, który to pierwszy zauważył): interpretacja – X jest liczbą zgłoszeń w ciągu czasu λ , przy założeniu, że jeśli podzielić ten czas na dostatecznie wiele małych odcinków, to w każdym odcinku czasu ma miejsce co najwyżej jedno zgłoszenie i są to zdarzenia niezależne; oznaczenie $Po(\lambda)$;

5) Geometryczny (parametr p): $P(X = k) = q^{k-1} p$, $k = 1, 2, \dots$; X jest liczbą prób do pierwszego sukcesu (włącznie); nazwa stąd, że $\sum_k p_k = p \sum_k q^{k-1} = p/(1 - q) = 1$.

6) Pascala, inaczej ujemny dwumianowy (2 parametry p i r): $P(X = k) = \binom{k-1}{r-1} q^{k-r} p^r$, $k = r, r + 1, \dots$; X jest liczbą prób do r -tego sukcesu (włącznie); jest to uogólnienie (5) – tam $r = 1$; nazwa stąd, że $\sum_k p_k = p^r (1 - q)^{-r} = 1$.

7) Hipergeometryczny (3 parametry M , N i n):

$$P(X = k) = \frac{\binom{M}{k} \binom{N-M}{n-k}}{\binom{N}{n}}, \quad k = 0, \dots, \min\{M, n\}$$

X liczy czerwone kule wśród n kul wylosowanych bez zwracania spośród N kul, z których M jest czerwonych; nazwa stąd, że przy oznaczeniach $(r)_s =$

$r(r-1)\dots(r-s+1)$ i $(r)^{(s)} = r(r+1)\dots(r+s-1)$,

$$p_k = \frac{(N-M)_n}{(N)_n} a_k$$

gdzie

$$a_k = \frac{(n)_k (M)_k}{(N-M-n+1)^{(k)} k!} = \frac{(-n)^{(k)} (-M)^{(k)}}{(N-M-n+1)^{(k)} k!}$$

jest współczynnikiem w szeregu hipergeometrycznym Gaussa o ogólnej postaci

$$1 + \sum_k \frac{(\alpha)^{(k)} (\beta)^{(k)}}{(\gamma)^{(k)} k!} t^k$$

(z $\alpha = -n$, $\beta = -M$ i $\gamma = N - M - n = 1$). Ten szereg jest uogólnieniem szeregu geometrycznego (gdzie $\alpha = \beta = \gamma = 1$).

7.2.2 Rozkłady ciągłe (absolutnie!)

Dystrybuanta F zmiennej losowej o rozkładzie absolutnie ciągłym posiada prawie wszędzie pochodną $f = F'$, całkowalną w sensie Lebesgue'a i spełniającą dla wszystkich $a < b$ warunek

$$\int_a^b f(x) dx = F(b) - F(a) .$$

Zatem mamy też

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

Funkcję f nazywamy *gęstością* (rozkładu) zmiennej losowej X . Jej istnienie i jednoznaczność (z dokładnością do zbioru o mierze Lebesgue'a równej 0) wynika też z twierdzenia Radona-Nikodyma.

Dystrybuanta i gęstość dają dwa sposoby obliczania prawdopodobieństw; np.

$$P(1 < X < 2) = F_X(2) - F_X(1) = \int_1^2 f(x) dx.$$

Gęstość f spełnia warunki

- (i) $f \geq 0$
- (ii) $\int_{-\infty}^{+\infty} f(x) dx = 1$.

Każda całkowalna funkcja spełniająca te dwa warunki jest gęstością pewnego rozkładu absolutnie ciągłego.

Gęstość jest tym dla rozkładu ciągłego, czym ciąg (p_1, p_2, \dots) dla rozkładu dyskretnego. Pokazuje, jak masa prawdopodobieństwa (=1) rozkłada się w zbiorze R . Użycie w tym celu (w przypadku ciągłym) prawdopodobieństw $P_X(x)$ nie ma sensu, bo najczęściej są one wszystkie równe zero.

8) Jednostajny (2 parametry $a < b$): X jest losowym punktem odcinka $[a, b]$, dla dowolnych $a < c < d < b$, $P(c < X < d) = \frac{d-c}{b-a}$; stąd dystrybuanta

$$F_X(x) = \begin{cases} 0 & \text{gdy } x < a \\ \frac{x-a}{b-a} & \text{gdy } x \in [a, b] \\ 1 & \text{gdy } x > b \end{cases}$$

a gęstość

$$f_X(x) = \begin{cases} 0 & \text{gdy } x < a \\ \frac{1}{b-a} & \text{gdy } x \in [a, b] \\ 0 & \text{gdy } x > b \end{cases}$$

oznaczenie $J(a, b)$

9) Wykładniczy (parametr α) : ciągły analog rozkładu geometrycznego; obserwujemy skrzyżowanie od chwili $t = 0$ i niech X będzie momentem pierwszego wypadku, zakładając tzw. brak pamięci (inaczej, jednorodność czasowa), tzn., że dla wszystkich $x, y > 0$ zachodzi równość

$$P(X > x + y | X > x) = P(X > y)$$

oraz warunek początkowy $P(X = 0) = 0$; powyższa równość jest równoważna warunkowi

$$\frac{1 - F_X(x + y)}{1 - F_X(x)} = 1 - F_X(y)$$

lub, podstawiając $U(x) = 1 - F_X(x)$,

$$U(x + y) = U(x)U(y)$$

jest to równanie funkcyjne Cauchy'ego, którego ogólne rozwiązanie ma postać $U(x) = e^{-\alpha x}$; u nas $\alpha > 0$, bo $\lim_{x \rightarrow \infty} U(x) = 0$; zatem

$$F_X(x) = \begin{cases} 0 & \text{gdy } x \leq 0 \\ 1 - e^{-\alpha x} & \text{gdy } x \geq 0 \end{cases}$$

oraz

$$f_X(x) = \begin{cases} 0 & \text{gdy } x < 0 \\ \alpha e^{-\alpha x} & \text{gdy } x \geq 0 \end{cases};$$

gęstość f_X nie jest różniczkowalna w $x = 0$, więc w tym punkcie obieramy wartość gęstości dowolnie, np. $f_X(0) = 0$.

10) Normalny (2 parametry μ i σ) : najpopularniejszy rozkład ciągły, choć o jego doniosłości dowiemy się dopiero pod koniec wykładów; gęstość dana jest wzorem

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right);$$

jej wykres, zwany krzywą Gaussa, ma charakterystyczny kształt dzwonu i dlatego czasem w lit. ang. nazywa się ją „the Bell curve-- mnie kojarzy się bardziej z węzłem, który zjadł słonia; oznaczenie $N(\mu, \sigma^2)$. Rozkład $N(0, 1)$ nosi nazwę *standardowy normalny*.

Rozdział 8

Wektory losowe (X, Y)

Dystrybuantę (łącną) zmiennych losowych X i Y definiujemy wzorem

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) .$$

Jest niemalejąca i prawostronnie ciągła ze względu na każdą zmienną; jej granica równa się 1, gdy $x, y \rightarrow \infty$ oraz 0, gdy $x \rightarrow -\infty$ lub $y \rightarrow -\infty$; dla dowolnych $a < b$ i $c < d$

$$F(b, d) - F(a, d) - F(b, c) + F(a, c) \geq 0 ,$$

bo wyrażenie to jest równe prawdopodobieństwu $P(a < X \leq b, c < Y \leq d)$.

Rozkłady dyskretne: $p_{ij} = P(X = x_i, Y = y_j)$, $\sum_i \sum_j p_{ij} = 1$. Na przykład, przy rzucie 2 kostkami $p_{ij} \equiv \frac{1}{36}$.

Przykład: rozkład trójmianowy. n prób, 3 wyniki A, B, C , z prawdopodobieństwami p, r i $q = 1 - p - r$; $X = \#$ wyników A , $Y = \#$ wyników B . Wtedy

$$p_{ij} = \binom{n}{i, j, n-i-j} p^i r^j q^{n-i-j}$$

$i = 0, 1, \dots, n, j = 0, 1, \dots, n - i$.

Rozkłady (absolutnie) ciągłe:

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(t, u) du dt$$

gdzie $f(x, y)$ jest gęstością (łącną). Gęstość $f(x, y)$ można obliczyć biorąc drugą mieszaną pochodną cząstkową z dystrybuanty $F(x, y)$. Własność

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1$$

wraz z nieujemnością funkcji f charakteryzuje wszystkie gęstości 2- wymiarowe.

8.1 Rozkłady brzegowe

Często znając rozkład łączny wektora losowego (X, Y) chcemy znaleźć rozkład jednej (lub każdej) ze zmiennych z osobna, tzn. P_X lub/i P_Y . Wtedy taki rozkład nazywamy *brzegowym*. W przypadku dyskretnym korzystamy z oczywistego wzoru

$$p_i^X = P(X = x_i) = \sum_j P(X = x_i, Y = y_j) = \sum_j p_{ij} ,$$

gdzie sumowanie przebiega po wszystkich atomach drugiego rozkładu. W przypadku ciągłym, gęstość brzegową obliczamy analogicznie:

$$f_X(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

i konsekwentnie dystrybuanta brzegowa wyraża się wzorem

$$F_X(x) = F(x, \infty) = \int_{-\infty}^x \left(\int f(x, y) dy \right) dx .$$

Przykład. Rozpatrzmy N rzutów monetą, gdzie N ma rozkład Poissona $Po(\lambda)$. Niech X i Y będą liczbami orłów i reszek, odpowiednio. Wtedy, dla wszystkich i, j

$$p_{ij} = P(X = i, Y = j | N = i + j) P(N = i + j) = \binom{i + j}{i} p^i q^j e^{-\lambda} \frac{\lambda^{i+j}}{(i + j)!} .$$

Zatem

$$P(X = i) = \sum_{j=0}^{\infty} p_{ij} = \dots = \frac{(p\lambda)^i}{i!} e^{-p\lambda} .$$

Widać więc, że X ma rozkład Poissona $Po(p\lambda)$. Analogicznie, Y ma rozkład $Po(q\lambda)$.

Odwrotny kierunek : od rozkładów brzegowych do łącznego jest niemożliwy bez znajomości ich wzajemnych powiązań, tzn. stopnia zależności między X i Y . Do ich określenia użyjemy rozkładów warunkowych. Ale najpierw rozważmy szczególny przypadek zależności : niezależność. Można wtedy przyjąć, że X i Y są określone w różnych przestrzeniach, a wektor (X, Y) jest określony na ich przestrzeni produktowej. W takim przypadku, oczywiście, $P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y)$

Zmienne losowe X i Y nazywamy *niezależnymi* jeśli

$$F_{X,Y} = F_X F_Y .$$

(Rodzina zmiennych losowych jest niezależna, jeśli dystrybuanta łączna każdej skończonej podrodziny tych zmiennych losowych jest równa iloczynowi ich dystrybuant brzegowych.)

Zatem zmienne losowe wzięte z różnych przestrzeni i rozpatrywane łącznie w przestrzeni produktowej są zawsze niezależne. W przypadku dyskretnym i ciągłym mamy, odpowiednio,

$$p_{ij} = p_i p_j$$

i

$$f_{X,Y} = f_X f_Y .$$

Ten ostatni wzór pozwala łatwo stwierdzić czy zachodzi niezależność!

Przykład. Gołym okiem widać, że zmienne losowe o łącznej gęstości $f(x, y) = \frac{1}{y} \exp\{-y - \frac{x}{y}\}$, $x, y > 0$, nie są niezależne, bo nie da się jej rozłożyć na iloczyn postaci $g(x)h(y)$. Można się o tym przekonać znajdując rozkład brzegowy zmiennej losowej Y . Całkując względem X otrzymujemy, że $f_Y(y) = e^{-y}$ dla $y > 0$.

8.2 Rozkłady warunkowe

Gdy obie zmienne losowe X i Y są dyskretne, to rozkład warunkowy X względem Y wyraża się wzorem

$$p_{i|j} = P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)},$$

dla wszystkich $i, j = 1, 2, \dots$. Analogicznie, w przypadku ciągłym, gęstość warunkowa X względem Y dana jest wzorem

$$f_{X|Y}(x|y) = \frac{f(x, y)}{f_Y(y)},$$

pod warunkiem, że $f_Y(y) > 0$.

Przykład: X i Y są niezależne, każda o rozkładzie dwumianowym $Bi(n, p)$. Wtedy $Z = X + Y$ ma rozkład $Bi(2n, p)$. Obliczmy rozkład warunkowy X względem Z .

$$P(X = i | Z = j) = \frac{P(X = i)P(Y = j - i)}{P(Z = j)} = \frac{\binom{n}{i} \binom{n}{j-i}}{\binom{2n}{j}},$$

a więc jest to rozkład hipergeometryczny z parametrami $2n, n, j$.

Przykład: Wróćmy do rozkładu o gęstości $f(x, y) = \frac{1}{y} \exp\{-y - \frac{x}{y}\}$, $x, y > 0$. Wiemy już, że X i Y są zależne. Dla $x, y > 0$ gęstość warunkowa X względem Y wynosi

$$f_{X|Y}(x|y) = \frac{\frac{1}{y} \exp\{-y - \frac{x}{y}\}}{e^{-y}} = \frac{1}{y} e^{-\frac{x}{y}}$$

Rozdział 9

Funkcje zmiennych losowych

Funkcję $g : \mathfrak{R} \rightarrow \mathfrak{R}$ nazywamy *borelowską*, jeśli przeciwobraz każdego zbioru borelowskiego jest borelowski. Na przykład, każda funkcja ciągła lub monotoniczna jest borelowska. Jeśli g jest borelowska, a X jest zmienną losową, to funkcja $Y = g(X)$ też jest zmienną losową. Podobnie, jeśli $h : \mathfrak{R}^k \rightarrow \mathfrak{R}^l$ jest borelowska, a (X_1, \dots, X_k) jest k -wymiarową zmienną losową, to $h(X_1, \dots, X_k)$ jest l -wymiarową zmienną losową.

9.1 Niezależność

Funkcje zmiennych losowych zachowują ich niezależność.

Twierdzenie 5 *Twierdzenie* Jeśli X_1, \dots, X_n są niezależnymi zmiennymi losowymi a funkcje $g_i : \mathfrak{R} \rightarrow \mathfrak{R}$, $i = 1, \dots, n$ oraz $h : \mathfrak{R}^k \rightarrow \mathfrak{R}$ są borelowskie, to

- a) $g_1(X_1), \dots, g_n(X_n)$ są niezależne,
- oraz
- b) $h(X_1, \dots, X_k), X_{k+1}, \dots, X_n$ są niezależne.

Dowód (a):

$$\begin{aligned} P(g_1(X_1) \leq x_1, \dots, g_n(X_n) \leq x_n) &= P(X_1 \in g_1^{-1}(-\infty, x_1), \dots, X_n \in g_n^{-1}(-\infty, x_n)) \\ &= P(X_1 \in g_1^{-1}(-\infty, x_1)) \cdot \dots \cdot P(X_n \in g_n^{-1}(-\infty, x_n)) = F_{g(X_1)}(x_1) \cdot \dots \cdot F_{g(X_n)}(x_n) \square \end{aligned}$$

9.2 Znajdowanie rozkładów funkcji zmiennych losowych

Jeśli g jest monotoniczna, to łatwo wyprowadzić ogólny wzór na dystrybuantę i gęstość $Y = g(X)$. Konkretnie, oznaczając przez h funkcję odwrotną do g , mamy

$$F_Y(y) = \begin{cases} F_X(h(y)) & \text{gdy } g \text{ jest rosnąca} \\ 1 - F_X(h(y)) & \text{gdy } g \text{ jest malejąca} . \end{cases}$$

Zatem gęstość Y będzie w obu przypadkach równa

$$f_Y(y) = f_X(h(y)) \left| \frac{dh(y)}{dy} \right|.$$

Przykłady:

1. $Y = X^3$, $f(x) = e^{-x}$, $x > 0$. Wtedy $g(x) = x^3$, $h(y) = y^{1/3}$ i $F_Y(y) = F_X(y^{1/3})$. Ponadto,

$$f_Y(y) = \frac{1}{3} y^{-2/3} e^{-y^{1/3}}.$$

2. (Przykład dyskretny) $Y = \sin(\frac{\pi}{2}X)$, $P(X = k) = 2^{-k}$, $k = 1, 2, \dots$
Wtedy

$$Y = \begin{cases} 0 & \text{gdym } X = 2k \\ 1 & \text{gdym } X = 4k - 3 \\ -1 & \text{gdym } X = 4k - 1 \end{cases}$$

Stąd, $P(Y = 0) = 1/3$, $P(Y = 1) = 8/15$, $P(Y = -1) = 2/15$.

3. Jeśli $Y = aX + b$ i obie z.l. są ciągłe, to $f_Y(y) = |a|^{-1} f_X(\frac{y-b}{a})$.

4. Jeśli X ma rozkład normalny $N(\mu, \sigma^2)$, to dla $Y = aX + b$, $a \neq 0$,

$$f_Y(y) = f_X\left(\frac{y-b}{a}\right) \frac{1}{|a|} = \frac{1}{|a|\sigma\sqrt{2\pi}} \exp\left(-\frac{(y-b-a\mu)^2}{2a^2\sigma^2}\right)$$

a zatem Y ma rozkład normalny $N(a\mu + b, |a|^2\sigma^2)$.

5. $Y = X^2$, X ma rozkład standardowy normalny. Wprawdzie $g(x) = x^2$ nie jest monotoniczna, ale jest przedziałami monotoniczna, więc

$$F_Y(y) = F_X(\sqrt{y}) - F_X(-\sqrt{y})$$

oraz

$$f_Y(y) = f_X(\sqrt{y}) \frac{1}{2\sqrt{y}} + f_X(-\sqrt{y}) \frac{1}{2\sqrt{y}} = \frac{1}{\sqrt{2\pi y}} e^{-y/2}.$$

Jest to tzw. rozkład χ^2 o jednym stopniu swobody.

Przypadek dwuwymiarowy.

Jeśli X i Y są dyskretnie, a $Z = g(X, Y)$, to

$$P(Z = z_n) = \sum_{(k,l):g(x_k,y_l)=z_n} P(X = x_k, Y = y_l) \quad (9.1)$$

a w szczególnym przypadku, gdy X i Y są niezależne i przyjmują wartości całkowite, to

$$P(Z = n) = \sum_{(k,l):g(k,l)=n} P(X = k)P(Y = l).$$

6. Niech X i Y będą niezależne o rozkładach Poissona, odpowiednio, $Po(\lambda_1)$, $Po(\lambda_2)$. Wtedy

$$\begin{aligned} P(Z = n) &= \sum_{k=0}^n P(X = k)P(Y = n - k) \\ &= e^{-\lambda_1 - \lambda_2} \frac{\lambda_1^n}{n!} \sum_{k=0}^n \binom{n}{k} \left(\frac{\lambda_1}{\lambda_2}\right)^k = e^{-\lambda_1 - \lambda_2} \frac{(\lambda_1 + \lambda_2)^n}{n!} \end{aligned}$$

a więc Z ma rozkład $Po(\lambda_1 + \lambda_2)$.

6' (inny sposób). Niech $Z = (X, X + Y)$, gdzie X i Y jak w przykładzie 6. Oznaczmy $U = X$, $V = X + Y$, tak że teraz $Z = (U, V)$. Układ równań $u = x$, $v = x + y$ ma rozwiązanie $x = u$, $y = v - u$. Zatem

$$P(Z = (n, m)) = P(X = n)P(Y = m - n) = e^{-\lambda_1} \frac{\lambda_1^n}{n!} e^{-\lambda_2} \frac{\lambda_2^{m-n}}{(m-n)!} \quad m \geq n$$

Aby otrzymać rozkład brzegowy V sumujemy po wszystkich możliwych wartościach U , czyli

$$P(V = m) = \sum_{n=0}^m P(Z = (n, m)) = \dots$$

Wynik powinien być taki sam jak w prz. 6

W przypadku dwuwymiarowym ciągłym trzeba skorzystać z jacobianów. Niech $(U, V) = T(X, Y)$, gdzie $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ jest odwzorowaniem borelowskim. Niech $f(x, y)$ będzie gęstością wektora losowego (X, Y) , a $J(u, v)$ jacobianem odwzorowania odwrotnego $T^{-1} = (g, h)$, tzn. wyznacznikiem macierzy złożonej z pochodnych cząstkowych obu funkcji składowych g i h względem obu zmiennych u i v . Wtedy, dla każdego zbioru borelowskiego B

$$\begin{aligned} P((U, V) \in B) &= P((X, Y) \in T^{-1}(B)) = \int_{T^{-1}(B)} \int f_{X,Y}(x, y) dx dy \\ &= \int_B \int f_{X,Y}(g(u, v), h(u, v)) |J(u, v)| du dv, \end{aligned}$$

gdzie ostatnia równość wynika z zamiany zmiennych w całce. Zatem, gdy (u, v) należy do obrazu odwzorowania T , to

$$f_{U,V}(u, v) = f_{X,Y}(g(u, v), h(u, v)) |J(u, v)|.$$

W przeciwnym razie $f_{U,V}(u, v) = 0$.

7. Dana jest łączna gęstość $f_{X,Y}(x, y) = 4xy$ gdy $0 < x < 1$, $0 < y < 1$, $f_{X,Y}(x, y) = 0$ w przeciwnym razie. Niech $U = X^2$ i $V = Y^2$. Wtedy $g(u, v) = \sqrt{u}$ i $h(u, v) = \sqrt{v}$, a jacobian $J(u, v) = (uv)^{-1/2}/4$. Stąd $f_{U,V}(u, v) = 1$ gdy $0 < u < 1$, $0 < v < 1$ i 0 w przeciwnym razie.

8. Niech $X = aU + bV$, $Y = cU + dV$, przy czym wszystkie rozkłady są ciągłe. Mamy $J(u, v) = ad - bc$, a stąd $f_{U,V}(u, v) = |ad - bc|f_{X,Y}(au + bv, cu + dv)$.

9. Wyznaczmy gęstość iloczynu $U = XY$ dwóch ciągłych zmiennych losowych. Wprowadźmy dodatkową zmienną losową $V = X$. Wtedy $(U, V) = T(X, Y)$, gdzie T jest parą odwzorowań $u = xy$ i $v = x$. Odwzorowanie odwrotne T^{-1} składa się z funkcji $g(u, v) = v$ i $h(u, v) = u/v$. Kolejne pochodne cząstkowe wynoszą $0, 1, 1/v$ i $-u/v^2$. Stąd jacobian odwzorowania T^{-1} ma postać $J(u, v) = -1/v$, więc gęstość łączna U i V równa się

$$f_{U,V}(u, v) = f_{X,Y}(v, u/v)/|v|.$$

Ze wzoru na gęstość brzegową,

$$f_U(u) = \int_{-\infty}^{\infty} f_{X,Y}(v, u/v)/|v| dv.$$

10. Niech X i Y będą niezależnymi zmiennymi losowymi o rozkładzie wykładniczym z parametrem α . Znajdźmy rozkład łączny wektora (X, S) , gdzie $S = X + Y$. Jacobian odwzorowania odwrotnego wynosi -1 , więc, dla $0 \leq x \leq s$,

$$f_{X,S}(x, s) = \alpha e^{-\alpha x} \alpha e^{-\alpha(s-x)} = \alpha^2 e^{-\alpha s}$$

Czyżby zmienne losowe X i S były niezależne? Skądże znowu! Wprowadźmy ich łączną gęstość jest iloczynem funkcji zależnych od pojedynczych zmiennych, ale cały wzór jest prawdziwy tylko, gdy $0 \leq x \leq s$, i tu właśnie kryje się zależność między X i S . Można się o niej formalnie przekonać znajdując rozkład warunkowy X względem S :

$$f_{X|S}(x, s) = \frac{\alpha^2 e^{-\alpha s}}{\int_0^s \alpha^2 e^{-\alpha s} dx} = \frac{1}{s}.$$

Jest to rozkład jednostajny $J(0, s)$, co można interpretować w ten sposób, że ustalwszy wartość sumy $S = s$, wkład w nią jednego z niezależnych składników jest dowolną liczbą wybraną losowo z przedziału $[0, s]$.

Przykłady 6 i 10 dotyczyły sum zmiennych losowych. Odpowiadający im rozkład nosi nazwę splotu. W ogólności, splot w przypadku dyskretnym wyraża się wzorem (9.1) (gdzie sumowanie rozciąga się na wszystkie pary (k, l) , dla których $x_k + y_l = z_n$). Jego odpowiednik w przypadku ciągłym ma postać

$$f_S(s) = \int_{-\infty}^{\infty} f_{X,Y}(x, s-x) dx$$

a gdy X i Y są niezależne, to

$$f_S(s) = \int_{-\infty}^{\infty} f_X(x) f_Y(s-x) dx$$

9.2. ZNAJDOWANIE ROZKŁADÓW FUNKCJI ZMIENNYCH LOSOWYCH 51

11. Niech X i Y będą niezależne, każda o rozkładzie standardowym normalnym $N(0, 1)$. Obliczmy ich splot, czyli rozkład ich sumy $S = X + Y$. Zgodnie z powyższym wzorem

$$f_S(s) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}x^2 - \frac{1}{2}(s-x)^2\right) dx = \frac{1}{2\sqrt{\pi}} \exp\left(-\frac{1}{4}s^2\right)$$

(powyżej dokonano podstawienia $v = (x - s/2)/\sqrt{2}$ i skorzystano z faktu, że całka z gęstości wynosi 1). Zatem S ma rozkład $N(0, 2)$. Ogólnie, gdy X_i ma rozkład $N(\mu_i, \sigma_i^2)$, $i = 1, 2$, to $X_1 + X_2$ ma rozkład $N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$.

Uwaga: $X - Y$ też ma rozkład $N(0, 2)$, bo Y i $-Y$ mają ten sam rozkład $N(0, 1)$.

Rozdział 10

Nadzieja matematyczna

Są dwie koncepcje wartości średniej zmiennej losowej: mediana i nadzieja matematyczna, czyli wartość oczekiwana.

10.1 Mediana zmiennej losowej X

Jest to taka liczba m , że $P(X \leq m) \geq \frac{1}{2}$ i $P(X \geq m) \geq \frac{1}{2}$. Innymi słowy jest to środek rozkładu.

Gdy X ma rozkład ciągły, to mediana jest wyznaczona równością $F(m) = \frac{1}{2}$. Na przykład, dla rozkładu wykładniczego o gęstości $f(x) = e^{-x}$ mediana wynosi $m = \ln 2$. Dla rozkładów, których gęstości są funkcjami symetrycznymi, tzn. $f(x) = f(-x)$, $m = 0$. Tak jest na przykład dla standardowego rozkładu normalnego $N(0, 1)$ i rozkładu Cauchy'ego o gęstości $f(x) = \frac{1}{\pi(1+x^2)}$. W przypadkach dyskretnych mediana nie musi być określona jednoznacznie. Na przykład dla rzutu kostką, medianą jest dowolną liczbą rzeczywistą z przedziału $[3, 4]$. Dla rozkładu dwumianowego $Bi(4, \frac{2}{3})$ mediana wynosi $m = 3$.

Mediana jest szczególnym przypadkiem kwantyli. *Kwantylem rzędu p* nazywamy liczbę rzeczywistą x_p taką, że $P(X \leq x_p) \geq p$ i $P(X \geq x_p) \geq 1-p$. Zatem $m = x_{1/2}$.

Kwantyle są stosowane w statystyce matematycznej (przedziały ufności).

10.2 Wartość oczekiwana

Jest to uogólnienie pojęcia średniej ze zbioru liczb. Dla rozkładów równoprawdopodobnych (więc skończonych) jest to po prostu średnia arytmetyczna. Dla rzutu kostką średnia wynosi $\frac{1}{6}(1 + 2 + 3 + 4 + 5 + 6) = 3,5$. Dla rozkładu $Bi(4, \frac{2}{3})$ będzie to już średnia ważona $(1 \times 8 + 2 \times 24 + 3 \times 32 + 4 \times 16)/81 = \frac{8}{3}$

Wartością oczekiwaną dyskretnej zmiennej losowej X o rozkładzie $p_i = P(X = x_i)$, $i = 1, 2, \dots$, nazywamy liczbę

$$EX = \sum_{i=1}^{\infty} x_i p_i = \sum_{\omega \in \Omega} X(\omega) p(\omega) ,$$

o ile istnieje.

W przypadku ciągłym, przez analogię definiujemy

$$EX = \int_R x f(x) dx = \int_{\Omega} X dP ,$$

gdzie druga całka, to całka według miary P .

Wartość oczekiwana nie zawsze istnieje. Tak jest w przypadku rozkładu Cauchy'ego oraz, gdy $p_k = \frac{A}{k^2}$, $k = \pm 1, \pm 2, \dots$, bo $\sum_{k \neq 0} \frac{1}{k} = \infty - \infty$. Czasem istnieje, ale równa się ∞ . Nie musi pokrywać się z medianą. Na przykład, dla wspomnianego powyżej rozkładu wykładniczego, $m = \ln 2$, ale $EX = 1$.

Dla funkcji $Y = g(X)$ mamy, w przypadku ciągłym, prosty wzór $EY = \int_R g(x) f(x) dx$.

Podobnie dla funkcji wielu zmiennych, tzn. gdy $Y = g(X_1, \dots, X_n)$. Na przykład, gdy X i Y są ciągłe,

$$E(XY) = \int \int_{R \times R} xy f(x, y) dx dy .$$

Wzory w przypadku dyskretnym są analogiczne. Łatwo stąd wywnioskować, że jeśli X i Y są niezależne, to $E(XY) = EXEY$.

Własności wartości oczekiwanej

- (a) $P(X \leq Y) = 1$ to $EX \leq EY$, w szczególności, jeśli $X \geq 0$ to $EX \geq 0$
- (b) $E(aX + bY) = aEX + bEY$ (liniowość)
- (c) $P(X = c) = 1$ to $E(X) = c$
- (d) $E(I_A) = P(A)$, gdzie $A \in \mathcal{F}$, a I_A jest zmienną losową wskaźnikową zdarzenia A , tzn., $I_A = 1$ gdy zaszło zdarzenie A , a $I_A = 0$ w przeciwnym razie.

Zastosujmy wzory (b) i (d) do wyznaczenia EX , gdy X ma rozkład dwumianowy $Bi(n, p)$. Wtedy $X = I_1 + \dots + I_n$, gdzie każde I_i jest wskaźnikiem zdarzenia (sukcesu) o prawdopodobieństwie p . Zatem

$$EX = EI_1 + \dots + EI_n = nEI_1 = np .$$

10.3 Momenty wyższych rzędów

Liczbę $E(X^k)$ nazywamy k -tym momentem zwykłym, a $E((X - EX)^k)$ k -tym momentem centralnym. Ten drugi mierzy średnie odchylenie od średniej; 1. moment centralny jest bezużyteczny, bo zawsze równa się 0. Dlatego bardzo ważny jest 2. moment centralny, nazywany *wariancją zmiennej losowej* X i oznaczany przez

$$\sigma_X^2 = V(X) = \text{Var}(X) = E((X - EX)^2) .$$

Własności wariancji

(a) Wariancję łatwiej wyznaczać ze wzoru

$$V(X) = E(X^2) - (EX)^2 .$$

(b) Z definicji wynika, że $V(X - c) = V(X)$ oraz, że

(c) $V(X) \geq 0$, a $V(X) = 0$ wtedy i tylko wtedy, gdy $P(X = c) = 1$

(d) Ponadto, $V(aX) = a^2V(X)$, a jeśli X, Y są niezależne, to

$$V(X + Y) = V(X) + V(Y)$$

Dowodu wymaga jedynie własność (d), która mówi, że choć wariancja nie jest operatorem liniowym, to dla niezależnych zmiennych losowych jest addytywna.

Dowód (d):

$$V(X+Y) = E((X+Y)^2) - (E(X+Y))^2 = V(X) + V(Y) + 2E(XY) - 2EXEY \square$$

Uwaga. W powyższym dowodzie skorzystaliśmy z tego, że jeśli X, Y są niezależne, to $E(XY) = EXEY$ (co wynika z definicji wartości oczekiwanej). Odwrotnie być jednak nie musi i w ogólności własność (d) nie zachodzi w drugą stronę. Do wyjątków należy przypadek, gdy obie zmienne losowe są wskaźnikowe (dowód na kolokwium). Inny ważny wyjątek poznamy później.

Pierwiastek z wariancji $\sigma = \sqrt{V(X)}$ nazywamy *odchyleniem standardowym* zmiennej losowej X . Nazwę tę wyjaśnia nierówność Czebyszewa, którą wyprowadzamy poniżej z nierówności Markowa.

10.4 Nierówności Markowa i Czebyszewa

Uwaga: Markow był uczniem Czebyszewa, a twierdzenie odwrotne nie jest prawdziwe!

Nierówność Markowa. Dla $a, k > 0$

$$P(|X| \geq a) \leq \frac{1}{a^k} E(|X|^k)$$

Dowód: $|X|^k \geq a^k I_{|X| \geq a}$, więc z własności (a,b,d)

$$E(|X|^k) \geq a^k P(|X| \geq a)$$

Nierówność Czebyszewa. Dla $t > 0$

$$P(|X - EX| \geq t\sigma) \leq \frac{1}{t^2}$$

Dowód: stosujemy nierówność Markowa do $X = X - EX$, z $k = 2$ i $a = t\sigma$.

Dla $t = 3$ otrzymujemy tzw. *prawo trzech sigm*. Mówi ono, że szanse odchylenia się dowolnej zmiennej losowej od swojej wartości średniej o więcej niż 3σ nie przekraczają $1/9$, czyli są znikome (przynajmniej dla statystyków).

10.5 Słabe Prawo Wielkich Liczb

Niech X_1, X_2, \dots będzie nieskończonym ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie (tzw. i.i.d., czyli „independent, identically distributed”). Oznaczmy $\mu = E(X_i)$ i $\sigma^2 = V(X_i)$. Rozważmy średnią pierwszych n z nich, czyli

$$\hat{S}_n = \frac{1}{n} \sum_{i=1}^n X_i .$$

Łatwo sprawdzić, że $E(\hat{S}_n) = \mu$ oraz $V(\hat{S}_n) = \sigma^2/n$. Zatem kolejne średnie mają wciąż tę samą wartość oczekiwaną, ale są coraz bliżej niej skoncentrowane. Słabe Prawo Wielkich Liczb głosi, że z prawdopodobieństwem dążącym do 1 średnie \hat{S}_n zblizają się dowolnie blisko do μ .

Twierdzenie 6 *Twierdzenie* Dla dowolnego $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(|\hat{S}_n - \mu| > \epsilon) = 0 .$$

Dowód: Na podstawie nierówności Czebyszewa,

$$P(|\hat{S}_n - \mu| > \epsilon) \leq \frac{V(\hat{S}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \rightarrow 0 . \square$$

Uwaga. W dowodzie wystarcza niezależność parami.

10.6 Nierówność Chernoffa

Nierówność Czebyszewa pokazuje, że z dużym prawdopodobieństwem zmienna losowa jest skoncentrowana wokół swojej wartości oczekiwanej. Np. dla zmiennej losowej X o rozkładzie dwumianowym $Bi(n, p)$ mamy

$$P(|X - np| > 0,01np) = O(1/n)$$

Jest ona jednak za słaba, by dobrze oszacować prawdopodobieństwo, że powyższą nierówność będzie spełniać jednocześnie wiele zmiennych losowych, o których wzajemnych zależnościach nic nie wiemy. Powiedzmy, że mamy n^2 zmiennych losowych X_1, \dots, X_{n^2} , wszystkie o rozkładzie $Bi(n, p)$. Wtedy, na podstawie nierówności Boole'a

$$P(\exists i : |X_i - np| > 0,01np) \leq n^2 P(|X - np| > 0,01np) = O(n) ,$$

co jest bezużyteczne w świetle tego, że prawdopodobieństwo nigdy nie jest większe od 1. Potrzebna tu jest inna, mocniejsza nierówność. Łatwo ją otrzymać wzmacniając sprytnie nierówność Markowa. Mamy bowiem, dla dowolnego $u \geq 0$,

$$P(X \geq EX + t) = P(e^{uX} \geq e^{u(EX+t)}) \leq e^{-u(EX+t)} Ee^{uX}$$

i podobnie dla każdego $u \leq 0$,

$$P(X \leq EX - t) \leq e^{-u(EX-t)} Ee^{uX}$$

Jeśli X ma rozkład $Bi(n, p)$, to oznaczając $\lambda = np$, mamy

$$P(X \geq EX + t) \leq e^{-u(\lambda+t)} (1-p + pe^u)^n$$

Prawa strona osiąga minimum dla

$$e^u = \frac{(\lambda + t)(1 - p)}{(n - \lambda - t)p} ,$$

zatem

$$P(X \geq EX + t) \leq \left(\frac{\lambda}{\lambda + t} \right)^{\lambda+t} \left(\frac{n - \lambda}{n - \lambda - t} \right)^{n-\lambda-t} \leq \left(\frac{\lambda}{\lambda + t} \right)^{\lambda+t} e^t$$

Oznaczmy

$$\phi(x) = (1 + x) \log(1 + x) - x \quad x \geq -1$$

i zauważmy, że $\phi(0) = \phi'(0) = 0$ i

$$\phi''(x) = \frac{1}{1+x} \geq \frac{1}{(1+x/3)^3} = \left(\frac{x^2}{2(1+x/3)} \right)'' .$$

Zatem funkcja $f(x) = \phi(x) - \frac{x^2}{2(1+x/3)}$ ma niemalejącą pierwszą pochodną i $f'(0) = 0$, co oznacza, że $f'(x)$ jest nieujemna dla $x \geq 0$. W konsekwencji, $f(x)$ jest niemalejąca, a ponieważ również $f(0) = 0$, to $\phi(x) \geq \frac{x^2}{2(1+x/3)}$ dla $x \geq 0$. Ostatecznie,

$$P(X \geq EX + t) \leq \exp(-\lambda\phi(t/\lambda)) \leq \exp\left(-\frac{t^2}{2(\lambda + t/3)}\right) .$$

Podobnie, a nawet łatwiej, można oszacować tzw. dolny ogon rozkładu $Bi(n, p)$:

$$P(X \leq EX - t) \leq \exp\left(-\frac{t^2}{2\lambda}\right).$$

Stosując nierówność Chernoffa do naszego przykładu, mamy dla pewnej stałej $c > 0$,

$$P(|X - np| > 0,01np) \leq 2e^{-cn},$$

co pomnożone przez dowolną funkcję wielomianową n dąży do 0, gdy $n \rightarrow \infty$ (zakładamy tu, że p jest stałą niezależną od n).

10.7 Kowariancja i korelacja

Kowariancją zmiennych losowych X i Y nazywamy liczbę

$$Cov(X, Y) = E(XY) - EXEY = E((X - EX)(Y - EY))$$

Jest to miara zależności zmiennych losowych. Jeśli $Cov(X, Y) = 0$ to mówimy, że X i Y są *nieskorelowane*. W szczególnym przypadku $Y = X$ mamy $Cov(X, X) = var(X)$.

OSTRZEŻENIE: Nieskorelowane zmienne losowe mogą być zależne.

Kowariancja nie jest dobrą miarą zależności, bo może być dowolnie duża i na przykład jej wartość 10 nic nam nie mówi. Trzeba ją unormować po to, by uzyskać jakąś skalę porównawczą. W tym celu definiujemy kolejny parametr.

Współczynnikiem korelacji X i Y nazywamy liczbę

$$\rho(X, Y) = \frac{Cov(X, Y)}{\sqrt{V(X)V(Y)}}$$

Własności korelacji

(a) $\rho(X, X) = 1$

(b)

$$\rho(X, aZ + b) = \begin{cases} \rho(X, Z) & \text{gdy } a > 0 \\ -\rho(X, Z) & \text{gdy } a < 0 \end{cases}$$

(c) $\rho(X, Y) = \pm 1$ wtedy i tylko wtedy, gdy $P(Y = \pm aX + b) = 1$, gdzie $a > 0$, b -dowolne

(d) $|\rho(X, Y)| \leq 1$

Dowody:

(a) z definicji

(b) Tu potrzebna jest pomocnicza własność, że

$$\rho(X, Y) = E(\bar{X}\bar{Y})$$

gdzie $\bar{X} = \frac{X-EX}{\sigma_X}$ jest tzw. *standaryzacją* z.l. X . Własność (b) wynika teraz stąd, że $a\bar{Z} + b = \bar{Z}$.

(c) Jeśli np. $\rho(X, Y) = 1$, to $E(\bar{X}\bar{Y}) = 1$, skąd

$$V(\bar{X} - \bar{Y}) = E((\bar{X} - \bar{Y})^2) = 0 .$$

(Skorzystaliliśmy tu z tego, że $E(\bar{X}) = 0$ a $V(\bar{X}) = 1$.) Zatem, na podstawie własności 9c) wariancji $P(\bar{X} = \bar{Y}) = 1$ ($c = 0$, bo $E(\bar{X}) = E(\bar{Y}) = 0$), co kończy dowód z $a = \sigma_Y/\sigma_X$ i $b = \mu_Y - a\mu_X$. Dowód w drugą stronę opiera się na częściach (a) i (b). Dowód w przypadku, gdy $\rho(X, Y) = -1$ jest analogiczny.

(d) Ostatnia, najważniejsza własność, wynika z nierówności Schwarza

$$E((XY))^2 \leq E(X^2)E(Y^2),$$

na podstawie której,

$$|Cov(X, Y)| \leq E(|X - EX||Y - EY|) \leq \sqrt{V(X)V(Y)} .$$

Samą nierówność Schwarza łatwo udowodnić: Dla dowolnego $a > 0$ niech $Z = aX - Y$. Mamy

$$E(Z^2) = a^2E(X^2) - 2aE(XY) + E(Y^2) \geq 0 ,$$

co jest możliwe tylko, gdy wyróżnik powyższego trójmianu kwadratowego jest niedodatni. Rozpisanie wyróżnika kończy dowód. \square

Podsumowując, korelacja zawsze leży pomiędzy -1 a 1. Wartości skrajne osiąga, gdy Y maleje (rośnie) liniowo z X . Gdy X, Y są niezależne, to $\rho(X, Y) = 0$. Mówimy wtedy, że są nieskorelowane. Ale odwrotnie być nie musi.

10.8 Warunkowa wartość oczekiwana

Warunkowa wartość oczekiwana dyskretnej zmiennej losowej Y względem zdarzenia A dana jest wzorem

$$E(Y|A) = \sum_i y_i P(Y = y_i|A),$$

a warunkowa wartość oczekiwana zmiennej losowej Y względem zmiennej losowej X – wzorem

$$\Psi(x) = E(Y|X = x) = \begin{cases} \sum_i y_i P(Y = y_i|X = x) & \text{w przypadku dyskretnym} \\ \int_R y f_{Y|X}(y|x) dy & \text{w przypadku ciągłym} \end{cases} .$$

Jest to funkcja zmiennej losowej X , a więc jest to także zmienna losowa, którą oznaczamy przez $\Psi(X) = E(Y|X)$. Ta zmienna losowa przyjmuje wartości

$$\Psi(X)(\omega) = E(Y|X = X(\omega)) .$$

Jej wartość oczekiwana jest równa, zgodnie ze zdrowym rozsądkiem, $E(Y)$. Poniższa, bardzo użyteczna własność jest w swojej wersji dyskretnej szczególnym przypadkiem następującego uogólnienia wzoru na prawdopodobieństwo całkowite:

$$EY = \sum_i E(Y|A_i)P(A_i) ,$$

gdzie A_i tworzą zupełny układ zdarzeń. Rzeczywiście, biorąc $A_i = \{X = x_i\}$, prawa strona równa się $E(E(Y|X))$. Biorąc natomiast $Y = I_B$, otrzymujemy wzór na prawdopodobieństwo całkowite

Twierdzenie 7 $E(E(Y|X)) = E(Y)$.

Prosty dowód podajemy w przypadku dyskretnym.

$$\begin{aligned} E(\Psi(X)) &= \sum_i \Psi(x_i)P(X = x_i) = \sum_i \sum_j y_j P(Y = y_j|X = x_i)P(X = x_i) \\ &= \sum_i \sum_j y_j P(Y = y_j, X = x_i) = \sum_j y_j P(Y = y_j) = EY . \square \end{aligned}$$

Przykład: rozkład mieszany

To ciąg dalszy przykładu z rozdziału 7: strzał do tarczy, za który otrzymuje się -1 punkt, jeśli się chybi (z prawdopodobieństwem $q = 1 - p$), albo miarę ϕ kąta utworzonego przez ślad strzału z ustalonym promieniem, $\phi \in [0, 2\pi]$. Obliczyć wartość oczekiwaną zmiennej losowej X będącej wartością wygranej. Wprost byłoby to trudne. Skorzystamy z powyższej własności. Niech O i R będą zdarzeniami, że trafiliśmy w tarczę, bądź spudłowaliśmy. Wtedy

$$E(X) = E(E(X|I_O)) = E(X|O)P(O) + E(X|R)P(R) = \pi p - 1q = \pi p - q = p\pi - q$$

Przykład: rozkład ciągły

Niech $f(x, y) = 1/x$ dla $0 \leq y < x \leq 1$ i $f(x, y) = 0$ poza tym trójkątem. Naszym celem jest obliczenie EY . Można to zrobić na 3 sposoby.

(1) Korzystając z powyższej własności. Najpierw obliczamy $f_X(x) = \int_0^x \frac{1}{x} dy = 1$ oraz $f_{Y|X}(y|x) = \frac{1/y}{1} = \frac{1}{x}$, po to by wyznaczyć $E(Y|X = x) = \frac{x}{2}$ (co wynika stąd, że warunkowy rozkład Y względem $X = x$ jest jednostajny - $J(0, x)$). Teraz

$$EY = \int_0^1 E(Y|X = x)f_X(x)dx = \int_0^1 \frac{x}{2} dx = \frac{1}{4}$$

(2) Traktując Y jako funkcję wektora losowego (X, Y) :

$$EY = \int_0^1 \int_0^x y \frac{1}{x} dy dx = \int_0^1 \frac{x}{2} dx = \frac{1}{4}$$

(3) Obliczając najpierw rozkład brzegowy Y : $f_Y(y) = \int_y^1 \frac{1}{x} dx = -\ln y$, mamy

$$EY = \int_0^1 (-y \ln y) dy = \dots = \frac{1}{4}$$

Przykład: jaja i kurczęta

Kura składa N jaj, gdzie N jest zmienną losową o rozkładzie Poissona z wartością oczekiwaną λ . Z jaja wykluwa się pisklę z prawdopodobieństwem p , niezależnie od innych jaj (podobno p zależy od dyspozycji koguta). Niech K będzie liczbą piskląt. Najpierw obliczmy EK . Mamy $E(K|N = n) = pn$, bo rozkład warunkowy K względem N , to rozkład dwumianowy $Bi(n, p)$. Zatem $E(K|N) = pN$ i z twierdzenia,

$$EK = E(E(K|N)) = E(pN) = pEN = p\lambda .$$

Można również pokazać, że $E(N|K) = K + q\lambda$. Żmudne, choć proste rachunki pomijamy. Na koniec dokonajmy sprawdzenia:

$$E(E(N|K)) = EK + q\lambda = p\lambda + q\lambda = \lambda = EN .$$

10.9 Regresja

Zacznijmy od prostej obserwacji dotyczącej wariancji. Dla dowolnej liczby rzeczywistej A , na podstawie własności (b) wariancji,

$$V(Y) = V(Y - A) = E((Y - A)^2) - (EY - A)^2$$

skąd wynika, że

$$E((Y - A)^2) \geq V(Y) = E((Y - EY)^2) \quad (10.1)$$

Innymi słowy, jeśli chcemy przybliżyć Y jakąś stałą, to najlepszym przybliżeniem w sensie *średniego kwadratu* jest stała $A = EY$. Następne twierdzenie uogólnia ten fakt.

Twierdzenie 8 *Niech X i Y będą dowolnymi z.l., $V(Y) < \infty$, a $g : \mathfrak{R} \rightarrow \mathfrak{R}$ – dowolną funkcją borelowską. Wtedy wyrażenie $E((Y - g(X))^2)$ osiąga minimum dla $g(X) = E(Y|X) = \Psi(X)$.*

Szkic dowodu: Stosując (10.1) w kontekście warunkowej wartości oczekiwanej,

$$E((Y - g(X))^2|X) \geq V(Y|X) = E((Y - E(Y|X))^2|X)$$

Stosując obustronnie operator E i Tw. 7, otrzymujemy nierówność

$$E((Y - g(X))^2) \geq E((Y - E(Y|X))^2) . \square$$

Wykres funkcji $y = \Psi(x) = E(Y|X = x)$ (jak również funkcji $x = \Phi(y) = E(X|Y = y)$) nazywamy *krzywą regresji I rodzaju*.

Gdy X i Y są niezależne, to $\Psi(x) = EY$ i $\Phi(y) = EX$ są funkcjami stałymi. Gdyby wielokrotnie empirycznie wyznaczać wartości wektora losowego (X, Y) i nanosić w układzie kartezjańskim, to otrzymane punkty byłyby rozmieszczone chaotycznie wokół punktu (EX, EY) .

W drugim skrajnym przypadku, gdy $|\rho(X, Y)| = 1$ (tzn. $P(Y = aX + b) = 1$), krzywe $\Psi(x) = E(ax + b) = ax + b$ i $\Phi(y)$ są jedną i tą samą linią prostą, na której leżałyby wszystkie punkty empiryczne. Jednak, gdy $0 < |\rho(X, Y)| < 1$, to punkty empiryczne będą skupiały się wokół niewidocznej prostej – prostej, która najlepiej przybliża Y poprzez X . Tę prostą nazywamy *krzywą regresji II rodzaju*. Formalnie, jest to funkcja liniowa $y = ax + b$, która minimalizuje wielkość

$$e = E((Y - aX - b)^2),$$

zwaną *kwadratowym błędem liniowej aproksymacji*.

Teraz wyznaczmy tę prostą. Przy ustalonym a , na podstawie (10.1), należy przyjąć $b = E(Y - aX) = EY - aEX$. Zatem problem sprowadza się do minimalizacji funkcji jednej zmiennej

$$g(a) = e = V(Y) - 2aCov(X, Y) + a^2V(X)$$

co jest zadaniem łatwym. Ostatecznie, otrzymujemy następujący wzór na krzywą regresji II rodzaju:

$$y - EY = \rho(X, Y) \frac{\sigma_Y}{\sigma_X} (x - EX)$$

oraz na związany z nią błąd $e = (1 - \rho(X, Y)^2)\sigma_Y^2$, który, co zrozumiałe, wynosi 0 jeśli $|\rho(X, Y)| = 1$ (wtedy też krzywe regresji I i II rodzaju pokrywają się), a jest największy, gdy $\rho(X, Y) = 0$ (wtedy krzywe regresji II rodzaju są postaci $y = EY$ i $x = EX$ i, w przypadku niezależnych X i Y , także pokrywają się z tymi I rodzaju).

10.10 Wielowymiarowy rozkład normalny

Pamiętamy, że jednowymiarowa zmienna losowa X ma rozkład normalny $N(\mu, \sigma^2)$, gdy

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x - \mu)^2}{2\sigma^2}\right) .$$

Wiemy także, że jeśli $Y = aX + b$, to Y również ma rozkład normalny, ale z innymi parametrami – $N(a\mu + b, |a|^2\sigma^2)$. Można by więc zdefiniować klasę

rozkładów normalnych jako klasę rozkładów wszystkich zmiennych losowych będących funkcjami liniowymi $aZ + b$ zmiennej losowej Z o rozkładzie standardowym $N(0, 1)$. Z uwagi na dopuszczalny przypadek $a = 0$, klasa ta obejmowałaby także rozkłady zdegenerowane.

Dla dwuwymiarowych wektorów losowych, rozkład normalny można podobnie definiować na dwa sposoby. Zaczniemy od przypadku, gdy Z i W są niezależne i mają standardowe rozkłady normalne $N(0, 1)$. Wtedy gęstość łączna wyraża się wzorem

$$f_{Z,W}(z, w) = \frac{1}{2\pi} \exp\left(-\frac{1}{2} [z^2 + w^2]\right)$$

Otóż klasę dwuwymiarowych rozkładów normalnych definiujemy jako klasę wszystkich rozkładów, które otrzyma się biorąc dwuwymiarową funkcję liniową zmiennych Z i W postaci

$$\begin{cases} X = aZ + bW + c \\ Y = dZ + eW + f \end{cases}$$

Ogólną postać gęstości łącznej można teraz wyprowadzić stosując metode z rozdziału 9, przykład 8 (żmudne obliczenia pomijamy):

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sqrt{1 - \rho^2}\sigma_1\sigma_2} \times \exp\left(-\frac{1}{2(1 - \rho^2)} \left[\frac{(x - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{x - \mu_1}{\sigma_1} \frac{y - \mu_2}{\sigma_2} + \frac{(y - \mu_2)^2}{\sigma_2^2} \right]\right),$$

gdzie $\mu_1 = c$, $\mu_2 = f$, $\sigma_1^2 = a^2 + b^2$, $\sigma_2^2 = d^2 + e^2$, a $\rho = (ad - be)/(\sigma_1\sigma_2)$, przy czym zakładamy, że $\sigma_1, \sigma_2 > 0$ oraz $|\rho| \neq 1$. (Na ćwiczeniach okaże się, że można przyjąć $b = 0$.)

Wzór ten może służyć jako bezpośrednia definicja 2-wymiarowego rozkładu normalnego (jest nim każdy rozkład o powyższej gęstości). Ma on jak widać 5 parametrów $\mu_1, \mu_2, \sigma_1, \sigma_2$ i ρ . Za chwilę wszystkie je rozszyfrujemy. Wykresem tej dwuwymiarowej funkcji jest powierzchnia, której każdy przekrój pionową płaszczyzną jest krzywą Gaussa i wygląda tak, jakby ktoś przykrył wierzchołek Giewontu dużym prześcieradłem.

Nietrudno pokazać, że rozkłady brzegowe są wtedy normalne, odp. $N(\mu_i, \sigma_i^2)$, $i = 1, 2$, i w konsekwencji można wyznaczyć rozkłady warunkowe, które też są normalne, np. Y względem $X = x - N(\mu_2 + \rho \frac{\sigma_2}{\sigma_1}(x - \mu_1), \sigma_2^2(1 - \rho^2))$.

Przypadek $|\rho| = 1$ jest poniekąd zdegenerowany. Mamy bowiem wtedy z prawdopodobieństwem 1 zależność $Y = aX + b$ i łączna dystrybuanta jest tak naprawdę jednowymiarowa:

$$F_{X,Y}(x, y) = P\left(X \leq x, X \leq \frac{y - b}{a}\right) = P\left(X \leq \min\left\{x, \frac{y - b}{a}\right\}\right)$$

(Z podobną sytuacją mamy do czynienia również wtedy, gdy $\sigma_1\sigma_2 = 0$.)

Parametr ρ jest niczym innym jak współczynnikiem korelacji zmiennych losowych X i Y . Aby uniknąć zawiłych rachunków, pokażemy to tylko w szczególnym przypadku, gdy $\mu_1 = \mu_2 = 0$ i $\sigma_1 = \sigma_2 = 1$, a więc gdy obie zmienne losowe są standaryzowane. Wtedy

$$\begin{aligned} \rho(X, Y) &= E(XY) \\ &= \int \int xy f(x, y) dx dy = \frac{1}{2\pi\sqrt{1-\rho^2}} \int \int xy \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) dx dy \\ &= \int y \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \left(\int \frac{x}{\sqrt{2\pi(1-\rho^2)}} \exp\left(-\frac{(x-\rho y)^2}{2(1-\rho^2)}\right) dx \right) dy \\ &= \int y \frac{1}{\sqrt{2\pi}} e^{-y^2/2} \rho y dy = \rho \end{aligned}$$

Zauważmy, że jeśli $\rho = 0$, to łączna gęstość wynosi

$$f_{X,Y}(x, y) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{1}{2} \left[\frac{(x-\mu_1)^2}{\sigma_1^2} + \frac{(y-\mu_2)^2}{\sigma_2^2} \right]\right).$$

Jest to iloczyn gęstości brzegowych, a więc X i Y są niezależne. Zatem rozkład normalny należy do tych wyjątkowych przypadków kiedy z braku korelacji wynika niezależność. Drugą anomalią dwuwymiarowego rozkładu normalnego jest to, że krzywe regresji I i II rodzaju pokrywają się. Wynika to stąd, że krzywa regresji I rodzaju, czyli warunkowa wartość oczekiwana, powiedzmy Y względem $X = x$, $E(Y|X = x) = \mu_2 + \rho \frac{\sigma_2}{\sigma_1} (x - \mu_1)$, jest funkcją liniową.

Rozdział 11

Funkcje tworzące i procesy gałązkowe

11.1 Funkcje tworzące

Niech X będzie dyskretną zmienną losową przyjmującą tylko wartości $0, 1, 2, \dots$. Funkcją tworzącą (prawdopodobieństwa) zmiennej losowej X nazywamy funkcję

$$G_X(s) = E(s^X) = \sum_i P(X = i)s^i = \sum_i p_i s^i$$

Jak widać jest to szereg o promieniu zbieżności 1. Funkcja tworząca zawiera pełną informację o danym rozkładzie, więc określa go w sposób jednoznaczny.

Przykłady. Jeśli X ma rozkład geometryczny z parametrem p , to jego funkcja tworząca wynosi

$$G_X(s) = \sum_{i=1}^{\infty} p_i s^i = \frac{ps}{1 - qs}$$

Poniżej będzie nas jednak raczej interesować zmienna losowa $X - 1$ licząca tylko porażki do pierwszego sukcesu. Dla niej

$$G_{X-1}(s) = \frac{p}{1 - qs}.$$

Na ćwiczeniach sprawdzimy, że jeśli Y ma rozkład Poissona z parametrem λ , to $G_Y(s) = e^{\lambda(s-1)}$.

Twierdzenie 9 (a) $G_X(1) = 1$

(b) $G(0) = P(X = 0) = p_0$

(c) $E[X(X-1)\dots(X-k+1)] = G_X^{(k)}(1)$, $k = 1, 2, \dots$;
w szczególności $EX = G'_X(1)$

(d) Jeśli X i Y są niezależne, to $G_{X+Y} = G_X \cdot G_Y$.

Dowody: (a) i (b) są oczywiste; (c) także, bo po k -krotnym różniczkowaniu mamy

$$G_X^{(k)}(s) = \sum_{i \geq k} i(i-1)\dots(i-k+1)p_i s^{i-k};$$

Również (d) jest oczywiste, bowiem na podstawie niezależności z.l. s^X i s^Y

$$G_{X+Y}(s) = E(s^{X+Y}) = E(s^X)E(s^Y) \square$$

Twierdzenie 10 (Twierdzenie o superpozycji funkcji tworzących) *Jeśli X_1, X_2, \dots jest ciągiem i.i.d. zmiennych losowych o funkcji tworzącej G , a N jest niezależną od nich zmienną losową, to dla sumy*

$$S = X_1 + X_2 + \dots + X_N$$

losowej liczby kolejnych elementów ciągu X_1, X_2, \dots zachodzi wzór

$$G_S(s) = G_N(G(s))$$

Dowód:

$$\begin{aligned} G_S(s) &= E(E(s^S|N)) = \sum_n E(s^S|N=n)P(N=n) \\ &= \sum_n E(s^{X_1+\dots+X_n})P(N=n) = \sum_n (G(s))^n P(N=n) \square \end{aligned}$$

Przykład: jaja i kurczęta. Powracamy do naszego świątecznego przykładu. Przypisując i -temu jaju zmienną losową X_i będącą wskaźnikiem zdarzenia, że z niego wykluje się pisklę, mamy $G(s) = q + ps$ i $G_N(s) = e^{\lambda(s-1)}$, a zatem funkcja tworząca liczby kurcząt K wynosi

$$G_K(s) = G_N(G(s)) = e^{\lambda(q+ps-1)} = e^{\lambda p(s-1)}$$

co oznacza, że K ma rozkład $Po(\lambda p)$, a to już wiedzieliśmy wcześniej.

11.2 Procesy gałązkowe

Koronnym przykładem zastosowania funkcji tworzących jest analiza następującego procesu losowego. Proces rozwija się w generacjach (pokoleniach), których liczebność jest mierzona ciągiem zmiennych losowych Z_0, Z_1, Z_2, \dots . Na początku był z całą pewnością jeden osobnik („stwórca”), a więc $P(Z_0 = 1) = 1$, który wydał na świat pewną losową liczbę potomków Z_1 . Dla każdego $n \geq 1$, każdy osobnik, niezależnie od innych (zarówno współczesnych jemu, jak i tych z poprzednich i przyszłych pokoleń) wydaje na świat losową liczbę potomków, przy czym wszystkie te zmienne losowe mają taki sam rozkład

jak liczba dzieci „stwórcy”, czyli Z_1 . Przyjmijmy, że ten wspólny rozkład ma funkcję tworzącą $G = G_{Z_1}$. Naszym celem jest wyznaczenie prawdopodobieństwa wyginięcia całej populacji, które oznaczamy przez η .

Zacznijmy od tego, że ponieważ ciąg zdarzeń $Z_n = 0$ jest wstępujący, to

$$\eta = P(\exists n : Z_n = 0) = P\left(\bigcup_{n=1}^{\infty} \{Z_n = 0\}\right) = \lim_{n \rightarrow \infty} P(Z_n = 0) = \lim_{n \rightarrow \infty} G_{Z_n}(0)$$

W ramach przygotowań do wyznaczenia η podamy teraz dwa ważne fakty.

Twierdzenie 11 *Niech $G_n = G_{Z_n}$ (czyli $G_1 = G$) oraz $\mu = EZ_1$. Wtedy*

- (a) $G_{n+m}(s) = G_m(G_n(s))$
- (b) $E(Z_n) = \mu^n$

Dowody: $Z_{n+m} = X_1 + \dots + X_{Z_n}$, gdzie X_i jest liczbą potomków w $(n+m)$ -tym pokoleniu i -tego członka m -tej generacji. Wszystkie X_i są niezależne i mają taki sam rozkład jak Z_n . Wzór (a) wynika z tw. o superpozycji. Wzór (b) dowodzimy przez indukcję. Dla $n = 1$ jest on trywialnie spełniony. Z Własności 1(c) mamy $EZ_n = G'_n(1)$. Na podstawie dopiero co udowodnionej części (a)

$$G'_n(s) = [G(G_{n-1}(s))]' = G'(G_{n-1}(s))G'_{n-1}(s)$$

co dla $s = 1$ daje

$$G'(G_{n-1}(1))G'_{n-1}(1) = G'(1)G'_{n-1}(1) = \mu EZ_{n-1} \square$$

Na podstawie własności (b) z powyższego twierdzenia i nierówności Markowa, wnioskujemy, że jeśli $\mu < 1$, to $P(Z_n > 0) = P(Z_n \geq 1) \leq \mu^n \rightarrow 0$, gdy $n \rightarrow \infty$, a więc $\eta = 1$. Przypadek $\mu > 1$ jest jednak znacznie trudniejszy.

Przykład : geometryczny proces gałązkowy. Załóżmy, że Z_1 ma przesunięty o -1 rozkład geometryczny z zamienionymi ze sobą sukcesem i porażką, tzn. $p_i = qp^i$, $i = 0, 1, 2, \dots$, $0 < p < 1$, $q = 1 - p$ (Z_1 liczy sukcesy do pierwszej porażki, co znajduje naturalną interpretację w procesach gałązkowych). Łatwo pokazać (ćwiczenia), że wtedy

$$G_n(0) = \begin{cases} \frac{n}{n+1} & \text{gdy } p = q = \frac{1}{2} \\ \frac{q(p^n - q^n)}{p^{n+1} - q^{n+1}} & \text{gdy } p \neq q \end{cases}$$

Zatem, gdy $p \leq q$, to $\eta = \lim_{n \rightarrow \infty} G_n(0) = 1$, a gdy $p > q$, to $\eta = \frac{q}{p}$. Zauważmy, że $\mu = E(Z_1) = \frac{p}{q}$, a więc w tym przypadku

$$\eta = \begin{cases} 1 & \text{gdy } \mu \leq 1 \\ \frac{1}{\mu} & \text{gdy } \mu > 1 \end{cases}$$

Wyznaczenie $G_n(0)$ jest często kłopotliwe lub wręcz niemożliwe. Okazuje się jednak, że do określenia prawdopodobieństwa wyginięcia populacji η wystarczy jedynie znajomość samej funkcji G . Oprócz powyższych oznaczeń, niech $\sigma^2 = V(Z_1)$.

Twierdzenie 12 (Twierdzenie o wyginieciu) *Prawdopodobieństwo wyginiecia η jest najmniejszym, nieujemnym pierwiastkiem równania $s = G(s)$. W szczególności, gdy*

$$\mu \begin{cases} < 1 \text{ to } \eta = 1 \\ > 1 \text{ to } \eta < 1 \\ = 1 \text{ i } \sigma > 0 \text{ to } \eta = 1 \\ = 1 \text{ i } \sigma = 0 \text{ to } \eta = 0 \end{cases}$$

Przykład : geometryczny proces gałązkowy – c.d. Rozwiążmy równanie $s = \frac{q}{1-ps}$. Z dwóch rozwiązań $s = \frac{1 \pm |p-q|}{2p}$ wybieramy to mniejsze, ale dodatnie, czyli $\eta = \frac{1-|p-q|}{2p}$.

Dowód: Zauważmy najpierw, że każda funkcja tworząca $G(s)$ jest ciągła, niemalejąca i wypukła (tę ostatnią własność weryfikujemy licząc drugą pochodną). Oznaczmy $\eta_n = G_n(0)$. Mamy wtedy

$$\eta_n = G(G_{n-1}(0)) = G(\eta_{n-1}).$$

Ponieważ $\eta = \lim_{n \rightarrow \infty} \eta_n$, to z ciągłości funkcji $G(s)$ wnioskujemy, że

$$\eta = \lim_{n \rightarrow \infty} G(\eta_{n-1}) = G(\eta),$$

a zatem η jest rozwiązaniem równania $s = G(s)$. Przypuśćmy, że również $x = G(x)$ dla pewnego $x \geq 0$. Wtedy z monotoniczności $G(s)$,

$$\eta_1 = G(0) \leq G(x) = x,$$

$$\eta_2 = G(G(0)) = G(\eta_1) \leq G(x) = x,$$

itd. Można w ten sposób pokazać indukcyjnie, że dla każdego n $\eta_n \leq x$, a więc również $\eta = \lim_{n \rightarrow \infty} \eta_n \leq x$, co kończy dowód pierwszej części twierdzenia.

Aby wywnioskować z niej kolejne przypadki szczególne (w zależności od wartości μ i σ), wystarczy zauważyć, że $\mu = G'(1)$ jest współczynnikiem kierunkowym stycznej do wykresu funkcji $G(s)$ w punkcie $s = 1$. Stąd i z wypukłości funkcji $G(s)$ wynika, że gdy $\mu < 1$ to wykres $G(s)$ nigdy nie przetnie wykresu prostej $h(s) = s$ poza punktem $s = 1$. Jeśli natomiast $\mu > 1$, to przeciwnie, te dwie krzywe przetną się poniżej punktu $s = 1$.

Gdy $\mu = 1$, to wszystko zależy od tego czy $p_0 > 0$. Jeśli tak, a jest to możliwe wtedy i tylko wtedy, gdy $\sigma > 0$ (musi być jakaś losowa fluktuacja, by zmienna losowa o wartości oczekiwanej 1 przyjmowała z dodatnim prawdopodobieństwem wartość 0), to mamy sytuację taką jak dla $\mu < 1$. W końcu, gdy $\mu = 1$ a $\sigma = 0$, to Z_1 jest po prostu zmienną losową zdegenerowaną w 1. Oznacza to, że zawsze i na pewno, każdy osobnik będzie miał dokładnie jednego potomka i populacja nigdy nie wyginie, tzn $\eta = 0$. \square

Wniosek, że przetrwają tylko populacje zdegenerowane jest jednak fałszywy. Należy po prostu zadbać o to, by średnia liczba potomków była wyższa niż jeden, a dzieło „stwórcy” nigdy nie pójdzie w niepamięć (ten religijno-prodżinny wydzwitek nie był zamierzony).

Rozdział 12

Twierdzenia graniczne

W rozdziale tym zajmiemy się zbieżnością ciągów zmiennych losowych. Poznaliśmy już jedno twierdzenie graniczne – Słabe Prawo Wielkich Liczb. Teraz nadszedł czas na coś mocniejszego, czyli Mocne Prawo Wielkich Liczb. W ramach przygotowań udowodnimy najpierw ważny lemat, a właściwie parę lematów, z których tylko pierwszy będzie nam potrzebny.

12.1 Lematy Borela-Cantelliego

Niech dany będzie nieskończony ciąg zdarzeń A_1, A_2, \dots . Zdarzenie

$$\limsup A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_k$$

wyraża fakt, że zaszło nieskończenie wiele spośród zdarzeń A_n . Jeśli dodatkowo założyć niezależność zdarzeń A_n , $n = 1, 2, \dots$, to jest to tzw. zdarzenie ogonowe, bo nie zależy od dowolnej, skończonej ilości zdarzeń A_n . Zgodnie z prawem zero-jedynkowym Kołmogorowa, które wykracza poza zakres tych wykładów, prawdopodobieństwo takiego zdarzenia musi być równe 0 lub 1. Lematy Borela-Cantelliego podają warunki wystarczające na to by $P(\limsup A_n) = 0$ i 1, odpowiednio.

Twierdzenie 13 (Lematy Borela-Cantelliego) (a) *Jeśli $\sum_n P(A_n) < \infty$, to $P(\limsup A_n) = 0$;*

(b) *Jeśli $\sum_n P(A_n) = \infty$, i, dodatkowo, zdarzenia A_n , $n = 1, 2, \dots$ są niezależne, to $P(\limsup A_n) = 1$.*

Dowody: Oznaczmy $A = \limsup A_n$.

(a) Dla każdego n

$$P(A) \leq P\left(\bigcup_{k=n}^{\infty} A_k\right) \leq \sum_{k=n}^{\infty} P(A_k)$$

Prawa strona, jako reszta szeregu zbieżnego dąży do 0, więc $P(A) = 0$.

(b)

$$\begin{aligned}
P(A) &= 1 - P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k^c\right) \geq 1 - \sum_{n=1}^{\infty} P\left(\bigcap_{k=n}^{\infty} A_k^c\right) \\
&= 1 - \sum_{n=1}^{\infty} \lim_{N \rightarrow \infty} P\left(\bigcap_{k=n}^N A_k^c\right) \\
&\geq 1 - \sum_{n=1}^{\infty} \lim_{N \rightarrow \infty} \exp\left(-\sum_{k=n}^N P(A_k)\right) = 1 - \sum_{n=1}^{\infty} 0 = 1 \quad \square
\end{aligned}$$

Przykład : rzuty monetą

Jeśli wykonujemy nieskończony ciąg rzutów sprawiedliwą monetą, a przez A_n oznaczmy zdarzenie, że za n -tym razem wypadnie orzeł, to $\sum_n P(A_n) = \sum_n \frac{1}{2} = \infty$, więc z prawdopodobieństwem 1 orzeł wypadnie nieskończenie wiele razy (podobnie zresztą jak i reszka). Jeśli jednak zmniejszymy prawdopodobieństwo orła, uzależniając je od n , to sytuacja może być zgoła inna. Gdy $P(A_n) = \frac{1}{n}$, to szereg $\sum_n P(A_n)$ nadal jest rozbieżny i konkluzja jest taka jak w przypadku symetrycznym (choć w każdym skończonym segmencie będą przeważać reszki). Dopiero gdy obniżymy prawdopodobieństwo orła do wartości $P(A_n) = \frac{1}{n^2}$, to z prawdopodobieństwem 1 orzeł wypadnie tylko skończoną liczbę razy.

12.2 Mocne Prawo Wielkich Liczb

Twierdzenie 14 (Mocne Prawo Wielkich Liczb) *Przy oznaczeniach i założeniach takich samych jak w Słabym Prawie Wielkich Liczb*

$$P\left(\lim_{n \rightarrow \infty} \dot{S}_n = \mu\right) = 1.$$

Dowód: W poniższym dowodzie założymy technicznie, że istnieje czwarty moment z.l. X_1 , tzn., że $E(X_1^4) < \infty$. Najpierw zauważmy, że w dowodzie można się ograniczyć do przypadku, gdy $\mu = 0$. Rzeczywiście, podstawiając $X'_n = X_n - \mu$, mamy $\dot{S}'_n = \dot{S}_n - \mu$ i zbieżność $\dot{S}_n \rightarrow \mu$ jest równoważna zbieżności $\dot{S}'_n \rightarrow 0$.

Na podstawie definicji granicy ciągu, zastępując ϵ ciągiem $1/m$, zdarzenie przeciwne można zapisać następująco:

$$\left\{\lim_{n \rightarrow \infty} \dot{S}_n \neq 0\right\} = \bigcup_{m=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} \{|\dot{S}_k| \geq 1/m\}$$

Zatem, z nierówności Boole'a i Lematu (a) Borela-Cantelliego,

$$P\left(\lim_{n \rightarrow \infty} \dot{S}_n \neq 0\right) \leq \sum_{m=1}^{\infty} P(\limsup\{|\dot{S}_n| \geq 1/m\}) = 0$$

o ile pokażemy, że $\sum_n P(|\dot{S}_n| \geq 1/m) < \infty$.

W tym celu skorzystamy z nierówności Markowa z $a = 1/m$ i $k = 4$ ($k = 2$ nie wystarcza, a $k = 3$ z uwagi na nieparzystość jest niewygodne). Mamy

$$P(|\dot{S}_n| \geq 1/m) \leq m^4 E(\dot{S}_n^4).$$

Zmienna losowa \dot{S}_n^4 jest sumą n^4 iloczynów $X_a X_b X_c X_d$ podzieloną przez n^4 . Na szczęście, z uwagi na niezależność zmiennych losowych X_i i na to, że $\mu = 0$, dla większości składników mamy $E(X_a X_b X_c X_d) = 0$.

Pozostają tylko te postaci $E(X_a^4)$ i $E(X_a^2 X_b^2)$, a jest ich niewiele, bo $O(n^2)$. Zatem, dla pewnej stałej C ,

$$\sum_n P(|\dot{S}_n| \geq 1/m) \leq C \sum_n \frac{1}{n^2} < \infty$$

co kończy dowód. □

12.3 Funkcje charakterystyczne

W rozdziale 11 przekonaliśmy się jak silnym narzędziem są funkcje tworzące prawdopodobieństwa zmiennych losowych o całkowitych i nieujemnych wartościach. Ich odpowiednikiem dla dowolnych z.l. są funkcje charakterystyczne, które posłużą nam w dowodzie najważniejszego twierdzenia tego rozdziału – Centralnego Twierdzenia Granicznego.

Ich protoplastą są funkcje tworzące momenty, czyli $E(e^{tX})$ znane w analizie jako transformaty Laplace'a. Mają jednak tę wadę, że nie zawsze istnieją.

Funkcją charakterystyczną zmiennej losowej X nazywamy funkcję postaci

$$\Phi_X(t) = E(e^{itX}).$$

Jest to zatem transformata Fouriera. Ponieważ $|e^{itX}| = 1$, to całka definiująca powyższą wartość oczekiwaną zawsze istnieje.

Funkcja Φ_X zawiera pełną informację o rozkładzie P_X z.l. X . Ponieważ to samo można powiedzieć o dystrybuancie F_X (patrz r. 7), to istnieje bijekcja pomiędzy funkcjami charakterystycznymi a dystrybuantami, a więc i rozkładami z.l.

Twierdzenie 15 (a) $\Phi_X(t) = \sum_{j=0}^k \frac{E(X^j)}{j!} (it)^j + o(t^k)$, o ile istnieje k -ty moment z.l. X ; wyrażenie $o(t^k)$ oznacza dowolną wielkość zależną od t , która dąży do 0 szybciej niż t^k , gdy $t \rightarrow 0$.

(b) $\Phi_{aX+b}(t) = e^{itb} \Phi_X(at)$

(c) Jeśli X i Y są niezależne, to $\Phi_{X+Y} = \Phi_X \Phi_Y$

(d) Jeśli X ma rozkład $N(0, 1)$, to $\Phi_X(t) = e^{-t^2/2}$

Funkcje charakterystyczne są doskonałym narzędziem do wyznaczania rozkładów granicznych ciągów z.l. Teoretyczne podstawy tej metody daje następujące twierdzenie. (My będziemy korzystać tylko z części (b).)

Twierdzenie 16 *Twierdzenie o ciągłości* Niech F_1, F_2, \dots będzie ciągiem dystrybuant, a Φ_1, Φ_2, \dots odpowiadającym im ciągiem funkcji charakterystycznych.

(a) Jeśli $F_n(x) \rightarrow F(x)$, gdy $n \rightarrow \infty$, dla każdego punktu ciągłości pewnej dystrybuanty F o funkcji charakterystycznej Φ , to $\Phi_n(t) \rightarrow \Phi(t)$ dla każdego rzeczywistego t .

(b) I odwrotnie, jeśli dla każdego rzeczywistego t , funkcja

$$\Phi(t) = \lim_{n \rightarrow \infty} \Phi_n(t)$$

istnieje i jest ciągła w punkcie $t = 0$, to Φ jest funkcją charakterystyczną pewnej dystrybuanty F i $F_n(x) \rightarrow F(x)$ dla każdego punktu ciągłości F . \square

12.4 Centralne Twierdzenie Graniczne

Wiemy już, że \hat{S}_n jest bliskie, dla dużych n , swojej wartości oczekiwanej μ (patrz: MPWL). Teraz dowiemy się jak zachowuje się asymptotycznie różnica $\hat{S}'_n = \hat{S}_n - \mu$. Okazuje się, że w większości przypadków jest ona, po standaryzacji, zbieżna do tego samego rozkładu – standardowego rozkładu normalnego $N(0, 1)$. Fakt ten wyjaśnia nazwę tego rozkładu.

Twierdzenie 17 *Centralne Twierdzenie Graniczne* Niech X_1, X_2, \dots będą i.i.d. z.l. o skończonych paramerach $\mu = EX_1$ i $\sigma^2 = V(X_1)$. Niech ponadto $S_n = \sum_{i=1}^n X_i$ i $\hat{S}_n = S_n/n$. Wtedy dla każdego rzeczywistego x

$$\lim_{n \rightarrow \infty} P\left(\frac{\hat{S}_n - \mu}{\sigma/\sqrt{n}} \leq x\right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx$$

Dowód: Zauważmy, że $\bar{S}_n = \frac{S_n - \mu n}{\sigma\sqrt{n}} = \frac{\hat{S}_n - \mu}{\sigma/\sqrt{n}}$ oraz, że $\bar{S}_n = \bar{S}'_n$. Zatem dowód przeprowadzimy, dla wygody, w wersji dla \bar{S}_n i przy założeniu, że $\mu = 0$. Na podstawie twierdzenia o ciągłości i własności (d) z Twierdzenia 15 wystarczy pokazać, że dla każdego ustalonego t i dla $n \rightarrow \infty$,

$$\Phi_{\bar{S}_n}(t) \rightarrow e^{-t^2/2}$$

Zmienna losowa \bar{S}_n jest sumą n niezależnych składników postaci $X_i/(\sigma\sqrt{n})$ o jednakowym rozkładzie. Na podstawie, kolejno, własności (b), (c) i (a) z Twierdzenia 15 mamy zatem

$$\begin{aligned} \Phi_{\bar{S}_n}(t) &= (\Phi_{X_1/(\sigma\sqrt{n})}(t))^n = (\Phi_{X_1}(t/(\sigma\sqrt{n})))^n \\ &= \left(1 + \left(\frac{it}{\sigma\sqrt{n}}\right)^2 \sigma^2 + o\left(\frac{t^2}{n}\right)\right)^n = \left(1 - \frac{t^2}{n} + o\left(\frac{t^2}{n}\right)\right)^n \\ &\rightarrow e^{-t^2/2} \end{aligned}$$

Ostatni krok wynika stąd, że $o\left(\frac{t^2}{n}\right)n = o(1)$. \square

12.5 Typy zbieżności

W przedstawionych dotychczas twierdzeniach granicznych wystąpiły 3 różne typy zbieżności zmiennych losowych. Zdefiniujmy je teraz formalnie dorzucając jeszcze czwarty.

Mówimy, że X_n dąży do X *według prawdopodobieństwa*, co zapisujemy $X_n \xrightarrow{P} X$, gdy dla każdego $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P(\{\omega : |X_n(\omega) - X(\omega)| > \epsilon\}) = 0$$

Mówimy, że X_n dąży do X *z prawdopodobieństwem jeden* (lub *prawie na pewno*), co zapisujemy $X_n \xrightarrow{a.s.} X$, gdy

$$P(\{\omega : \lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)\}) = 1$$

Mówimy, że X_n dąży do X *według rozkładu*, co zapisujemy $X_n \xrightarrow{D} X$, gdy

$$\lim_{n \rightarrow \infty} F_{X_n} = F_X(x)$$

dla każdego punktu ciągłości x dystrybucyj F_X .

Mówimy, że X_n dąży do X *według r -tego momentu*, co zapisujemy $X_n \xrightarrow{r} X$, gdy

$$\lim_{n \rightarrow \infty} E(|X_n - X|^r) = 0$$

Przykład. Niniejszy przykład pokazuje, że w definicji zbieżności według rozkładu nie można żądać, by zbieżność zachodziła w każdym punkcie x . Niech $X_n = \frac{1}{n}$, a $Y_n = -\frac{1}{n}$, $n = 1, 2, \dots$. Oba te ciągi zdegenerowanych zmiennych losowych, zgodnie z intuicją powinny zbiegać do 0. Jednakże, oznaczając $F_n = F_{X_n}$ mamy

$$\lim_{n \rightarrow \infty} F_n(x) = \begin{cases} 0 & \text{gdy } x \leq 0 \\ 1 & \text{gdy } x > 0 \end{cases}$$

i, jak widać, funkcja graniczna nie jest w ogóle dystrybuantą. Nie ma tego problemu w przypadku Y_n . Tutaj dystrybuanty zbiegają do dystrybuanty z.l. zdegenerowanej w 0.

Poniższe twierdzenie zbiera wszystkie implikacje zachodzące między tymi czterema typami zbieżności.

Twierdzenie 18 (Twierdzenie 1 (o typach zbieżności))

$$\begin{aligned} X_n \xrightarrow{r} X &\stackrel{(1)}{\implies} X_n \xrightarrow{P} X \\ X_n \xrightarrow{a.s.} X &\stackrel{(2)}{\implies} X_n \xrightarrow{P} X \stackrel{(3)}{\implies} X_n \xrightarrow{D} X \\ X_n \xrightarrow{r} X &\stackrel{(4)}{\implies} X_n \xrightarrow{s} X \quad \text{dla } r \geq s \end{aligned}$$

Dowody: □

Podamy teraz kilka twierdzeń, które przy pewnych dodatkowych, nieraz bardzo ostrych założeniach, odwracają niektóre z powyższych implikacji.

Twierdzenie 19 *Jeśli $P(X = c) = 1$, tzn. X jest zdegenerowana w punkcie c , to*

$$X_n \xrightarrow{D} X \implies X_n \xrightarrow{P} X$$

Dowód: $P(|X_n - c| > \epsilon) = P(X_n < c - \epsilon) + P(X_n > c + \epsilon) \rightarrow 0$. □

Uwaga. Twierdzenie 19 pozwala podać dowód SPWL przy pomocy funkcji charakterystycznych. Po prostu wystarczy pokazać, że ciąg funkcji charakterystycznych z.l. \hat{S}_n jest zbieżny do $e^{i\mu t}$, czyli do funkcji charakterystycznej rozkładu zdegenerowanego w μ . Tw. o ciągłości implikuje teraz, że $\hat{S}_n \xrightarrow{D} \mu$, a powyższe twierdzenie stawia kropkę nad i.

Twierdzenie 20 *Jeśli zmienne losowe X_n , $n = 1, 2, \dots$ są określone na przeliczalanej przestrzeni probabilistycznej, to*

$$X_n \xrightarrow{P} X \implies X_n \xrightarrow{a.s.} X$$

Twierdzenie 21

$$X_n \xrightarrow{P} X \implies \exists(n_i) : X_{n_i} \xrightarrow{a.s.} X$$

Twierdzenie 22 (Twierdzenie Skorohoda o reprezentacjach) *Jeśli $X_n \xrightarrow{D} X$, to istnieje taka przestrzeń probabilistyczna oraz zdefiniowane na niej zmienne losowe Y_n , $n = 1, 2, \dots$ oraz Y , takie, że $F_{Y_n} = F_{X_n}$, $n = 1, 2, \dots$, $F_Y = F_X$, oraz $Y_n \xrightarrow{a.s.} Y$.*

Twierdzenie 23 *Jeśli $X_n \xrightarrow{D} X$ oraz $g : \mathfrak{R} \rightarrow \mathfrak{R}$ jest funkcją ciągłą, to $g(X_n) \xrightarrow{D} g(X)$.*

Dowód : Niech Y_n i Y będą jak w Tw. 22. Wtedy z ciągłości g mamy również $g(Y_n) \xrightarrow{a.s.} g(Y)$. Rzeczywiście, weźmy ω takie, że $\lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega)$. Dla każdego $\epsilon > 0$ istnieje takie $\delta > 0$, że jeśli $|x_1 - x_2| < \delta$, to $|g(x_1) - g(x_2)| < \epsilon$. Weźmy teraz tak duże n_0 , by dla każdego $n > n_0$ zachodziła nierówność $|Y_n(\omega) - Y(\omega)| < \delta$. Wtedy $|g(Y_n(\omega)) - g(Y(\omega))| < \epsilon$, co oznacza, że $\lim_{n \rightarrow \infty} g(Y_n(\omega)) = g(Y(\omega))$.

Zbieżność $g(Y_n) \xrightarrow{a.s.} g(Y)$, na podstawie Tw.18 (implikacje 2 i 3) implikuje zbieżność $g(Y_n) \xrightarrow{D} g(Y)$, a ta jest równoważna zbieżności $g(X_n) \xrightarrow{D} g(X)$, bo zbieżność według rozkładu nie zależy od samych zmiennych losowych, a tylko od ich rozkładów właśnie, które są takie same dla X_n i Y_n oraz dla X i Y . □

Rozdział 13

Posmak statystyki matematycznej

Statystyka matematyczna jest zastosowaniem rachunku prawdopodobieństwa do badania wybranych *cech populacji* na podstawie próby losowej. Cecha to zmienna losowa zdefiniowana na populacji (np. wzrost obywateli danego kraju), której rozkład jest (częściowo) nieznan.

Próba losowa to wektor liczbowy (x_1, \dots, x_n) będący wylosowaną wartością wektora losowego (X_1, \dots, X_n) , gdzie zmienne losowe $X_i, i = 1, \dots, n$, są i.i.d. o tym samym rozkładzie co badana cecha X (po prostu losuje się n obywateli, doprowadza do Urzędu i tam mierzy ich wzrost). Ponieważ pojęciem „próba losowa” określa się zarówno (x_1, \dots, x_n) , jak i (X_1, \dots, X_n) , to ten pierwszy wektor nazywa się też czasem, dla odróżnienia, *zaobserwowaną próbą* lub po prostu *próbką*.

Statystyka to zmienna losowa Y , będąca funkcją z próby:

$$Y = g(X_1, \dots, X_n)$$

Przykłady: Średnia

$$\hat{S} = \frac{1}{n} \sum_{i=1}^n X_i$$

Wariancja z próby

$$V^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \hat{S})^2$$

Statystyki służą do uzyskiwania prawdopodobnych informacji na temat badanej cechy X . Robi się to zasadniczo na dwa sposoby, zwane estymacją i testowaniem hipotez.

13.1 Estymacja

Estymacja polega na przybliżaniu nieznanego parametru θ rozkładu z.l. X na podstawie otrzymanych wartości odpowiednio dobranej statystyki. Taką specjalnie do tego celu przygotowaną statystykę nazywa się *estymatorem parametru θ* i oznacza przez $\hat{\theta}$.

Estymator jest *nieobciążony* gdy $E(\hat{\theta}) = \theta$.

Przykład. \hat{S} jest nieobciążonym estymatorem $\mu = EX$, bo jak wiemy $E(\hat{S}) = \mu$. Natomiast V^2 jest obciążonym estymatorem wariancji $\sigma^2 = V(X)$, bowiem

$$E(V^2) = \frac{n-1}{n}\sigma^2$$

Dlatego też częściej używa się skorygowanej wariancji z próby $V_1^2 = \frac{n}{n-1}V^2$.

Estymacja polegająca na przyjęciu za nieznaną wartość parametru θ obliczonej z próbki wartości estymatora $\hat{\theta}$ nosi nazwę *punktowej*. Znacznie lepsza jest jednak *estymacja przedziałowa*, w której przy pomocy dwóch estymatorów: dolnego $\bar{\theta}_d$ i górnego $\bar{\theta}_g$, określa się przedział, do którego wpada nieznaną wartość parametru z zadaniem prawdopodobieństwem:

$$P(\bar{\theta}_d \leq \theta \leq \bar{\theta}_g) = 1 - \alpha$$

Liczbę $1 - \alpha$ nazywamy *poziomą ufnością*, a przedział $(\bar{\theta}_d, \bar{\theta}_g)$ *przedziałem ufności*.

Przykład. Przypuśćmy, że wiemy, iż X ma rozkład $N(\mu, \sigma^2)$, σ jest nam znane, ale μ nie. Naturalnym estymatorem parametru μ jest średnia z próby $\hat{\mu} = \hat{S}_n$, o której wiemy w tym przypadku, że również ma rozkład normalny, $N(\mu, \frac{\sigma^2}{n})$. Standaryzacja $\hat{\mu}$ ma więc rozkład $N(0, 1)$. Niech λ_α będzie kwantylem rozkładu $N(0, 1)$, tzn. punktem odcinającym pod krzywą Gaussa obszar o polu α . Wtedy

$$P\left(\lambda_{\alpha/2} < \frac{\hat{\mu} - \mu}{\sigma/\sqrt{n}} < \lambda_{1-\alpha/2}\right) = 1 - \alpha$$

Rozwiązując powyższe nierówności względem μ i wykorzystując zależność $\lambda_{\alpha/2} = -\lambda_{1-\alpha/2}$, otrzymujemy przedział ufności na poziomie $1 - \alpha$:

$$P\left(\hat{\mu} - \lambda_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \hat{\mu} + \lambda_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Gdyby σ też było nieznaną, to wybieramy inną statystykę, a mianowicie

$$t = \frac{\hat{S}_n - \mu}{V} \sqrt{n-1}$$

Zmienna losowa t ma tzw. rozkład t -studenta. Niech t_α będzie kwantylem tego rozkładu. Rozumując podobnie jak poprzednio znajdujemy przedział ufności na poziomie $1 - \alpha$ postaci

$$\left(\hat{S}_n - t_{1-\alpha/2}\frac{V}{\sqrt{n-1}}, \hat{S}_n + t_{1-\alpha/2}\frac{V}{\sqrt{n-1}}\right).$$

13.2 Testowanie hipotez

Hipotezę nazywa się dowolne przypuszczenie na temat rozkładu bądź parametrów badanej cechy X . *Test* polega na weryfikacji lub odrzuceniu hipotezy na podstawie statystyki obliczonej z zaobserwowanej próby (zwanej testem). Idea jest taka, że jeśli test dał wynik, który przy założeniu prawdziwości hipotezy jest mało prawdopodobny, to należy hipotezę odrzucić. Przy okazji można popełnić jeden z dwóch rodzajów błędów: Błąd I rodzaju – odrzucenie prawdziwej hipotezy, Błąd II rodzaju – przyjęcie fałszywej hipotezy. Niestety nie da się zminimalizować obu tych błędów na raz. Gdy jeden maleje, drugi rośnie. Ponieważ, jak się wydaje, błąd II rodzaju jest poważniejszy, to zwykle, ustala się z góry wielkość błędu I rodzaju (α) i stara się zminimalizować błąd II rodzaju (β). W tym celu konstruuje się optymalnie tzw. *obszar krytyczny* R_α tak, by prawdopodobieństwo wpadnięcia testu do tego obszaru (zwykle jest to przedział lub suma przedziałów), przy założeniu, że hipoteza jest prawdziwa, wynosiło dokładnie α . Hipotezę odrzucamy, gdy rzeczywiście wartość testu wpadnie do obszaru R_α .

Testy dzielą się na parametryczne (czyli te dotyczące wartości parametrów) oraz nieparametryczne (dotyczące rodzaju rozkładów). Zwykle obok testowanej hipotezy (oznaczanej przez H_0) formułuje się tzw. *hipotezę alternatywną* H_1 , która zostaje automatycznie uznana za poprawną w przypadku, gdy w wyniku testu odrzucimy hipotezę H_0 .

Przykład. Porównujemy dwie populacje w aspekcie pewnej cechy (np. wzrost Czechów i Słowaków). Przyjmijmy, że obie zmienne losowe X i Y mają rozkłady normalne, odp. $N(\mu_1, \sigma_1^2)$ i $N(\mu_2, \sigma_2^2)$. Weryfikowana hipoteza H_0 głosi, że $\sigma_1 = \sigma_2$, podczas, gdy hipoteza alternatywna H_1 brzmi: $\sigma_1 > \sigma_2$ (nie dopuszcza się możliwości, że $\sigma_1 < \sigma_2$). Pobieramy dwie próbki o liczebnościach odp. n_1 i n_2 , i obliczamy na ich podstawie wartości v_1^2 i v_2^2 statystyk V_1^2 i V_2^2 (numerujemy je tak, by $v_1 \geq v_2$). Jako test przyjmujemy statystykę

$$\frac{V_1^2 n_1 (n_2 - 1)}{V_2^2 n_2 (n_1 - 1)}$$

Ma ona tak zwany rozkład F -Snedecora. Oznaczając jego kwantyl przez F_α , obszar krytyczny dla tego testu definiujemy jako przedział $R_\alpha = (F_{1-\alpha}, \infty)$.