Jan ŚNIADECKI

o nadziei, obawie i demoralizacji

[...] W grze pieniężnej, jeżeli summę, o którą się gra, rozmnożymy przez podobieństwo trafu, otrzymujemy to, co się nazywa nadzieja, czyli oczekiwanie; jeśli zaś tę summę pieniężną rozmnożymy przez podobieństwo chybienia, mamy to, co się nazywa obawa.
Więc rachunek ten daje nam wartość i cenę naszej nadziei i obawy nie tylko w grze, ale we wszystkich przedsięwzięciach losowych. Stosując podobieństwo trafu do summy w grze postawionej, mamy to, co się nazywa nadzieją matematyczną; stosując je zaś do całego majątku gracza, mamy nadzieję moralną, bo wtenczas dochodzić można niebezpieczeństw straty i zupełnego upadku gracza, a stąd wypadającej demoralizacji ludzi źle się rządzących.

Fragment z rozprawy: *O rachunku losów. Rzecz czytana na sessyi literackiej Uniwersytetu Wileńskiego* 15 *listopada* 1817 roku v.s. Cytowane wg: Jan Śniadecki, *Wybór Pism Naukowych*, PWN, Warszawa 1954.

Red.: Termin *nadzieja matematyczna* stosunkowo niedawno dopiero wyparty został przez *wartość oczekiwaną*; *obawa matematyczna* nigdy się nie przyjęła. *Podobieństwo* to oczywiście prawdopodobieństwo, a *rachunek losów* — rachunek prawdopodobieństwa.

Rozkłady prawdopodobieństwa

Pragniemy — jak co roku — ofiarować Czytelnikom prezent gwiazdkowy. Tym razem jest to przegląd ważniejszych rozkładów zmiennych losowych rozważanych w teorii prawdopodobieństwa i jej zastosowaniach.
Omówienie to dalekie jest od rygorystycznej poprawności formalnej. Odwołujemy się w nim przede wszystkim do intuicji i wyobraźni, bowiem zależy nam tu głównie na uwypukleniu związków teorii prawdopodobieństwa ze zjawiskami otaczającej nas rzeczywistości.

Zob. też artykuły R. Zielińskiego i L.I. Kubika.

Zmienna losowa i jej dystrybuanta

W badaniach przyrodniczych podstawą wszelkiego wnioskowania są na ogół *eksperyment* i *mierzenie*: w celu zdobycia informacji o jakimś zjawisku przeprowadza się badanie eksperymentalne, a wynikom tego badania przyporządkowuje się liczby, które stają się podstawą do dalszych wnioskowań i uogólnień.
Eksperyment badawczy na ogół ma charakter *losowy*. Sens bowiem prowadzenia badania eksperymentalnego kryje się w tym, że nie potrafimy przewidzieć, co się w nim wydarzy na pewno, potrafimy natomiast powiedzieć, jakie są możliwe wyniki. To, który z tych możliwych wyników w rzeczywistości się wydarzy, zależy często od czynników niemożliwych do uchwycenia, zwanych losem lub przypadkiem (tak jest zarówno przy rzucie monetą, jak i rejestrowaniu rozpadów cząstek elementarnych). Czasem zresztą badaniom eksperymentalnym świadomie nadaje się charakter losowy: przy statystycznej kontroli jakości dużej partii towaru bada się jakość przypadkowo wybranych egzemplarzy, na podstawie opinii wylosowanych rozmówców wnioskuje się o nastrojach społeczeństwa itd.
Eksperyment losowy (który może składać się z wielu drobniejszych doświadczeń) prowadzi się zawsze przy założeniu, że istnieją (lub będą istniały) podstawy do przyporządkowywania prawdopodobieństw niektórym przynajmniej jego wynikom. A ponieważ wyniki te na ogół jakoś się mierzy, to w efekcie wynikom mierzenia (liczbom lub zbiorom liczb) zostają przypisane pewne prawdopodobieństwa: prawdopodobieństwa zdarzeń, którym odpowiadają te liczby (lub zbiory liczb).
Abstrakcyjnym, matematycznym modelem tak rozumianego eksperymentu jest przestrzeń probabilistyczna. Składają się na nią:
— *zbiór zdarzeń elementarnych* (oznaczany zwykle symbolem Ω). Zdarzenia elementarne odpowiadają *możliwie najprostszym wynikom* eksperymentu, takim np. jak to, że przy rzucie kostką wypadła szóstka, lub to, że wzrost przypadkowo wybranego człowieka wyniósł dokładnie 187,9237134 cm.
— *zbiór zdarzeń* (oznaczany zwykle symbolem $\mathfrak{M}$). Zdarzenia odpowiadają *rzeczywiście interesującym nas wynikom doświadczenia*, takim np. jak to, że wzrost losowo wybranego pana zawarty był pomiędzy 187,5 cm a 188,5 cm. Zdarzenia są podzbiorami zbioru zdarzeń elementarnych.
— *prawdopodobieństwo* (oznaczane zwykle przez P), które jest funkcją przyporządkowującą zdarzeniom należącym do $\mathfrak{M}$ liczby z przedziału $\langle 0, 1\rangle$ w taki sposób, by spełnione były znane warunki (tzw. *aksjomaty Kołmogorowa*, zob. obok).
Prawdopodobieństwo jest odpowiednikiem *częstości* (lub *szansy*) występowania określonego wyniku.
Matematycznym odpowiednikiem mierzenia jest *zmienna losowa*, funkcja, która zdarzeniom elementarnym przypisuje liczby. Dla każdej zmiennej losowej można określić jej *dystrybuantę* — funkcję, która każdemu $x \in \mathrm{R}$ przypisuje prawdopodobieństwo tego, że zmienna ta przyjmie wartość mniejszą od x. Inaczej mówiąc, jeśli X jest zmienną losową, a F_X jej dystrybuantą, to

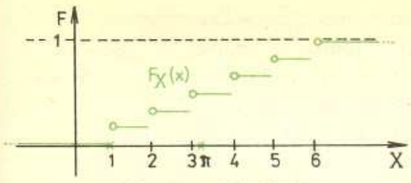
$$F_X(x) \stackrel{\mathrm{df}}{=} P(\{\omega \in \Omega: X(\omega) < x\}).$$

Aksjomatyczna definicja prawdopodobieństwa (Kołmogorowa):
1) $P: \mathfrak{M} \to \langle 0,1\rangle$,
2) $P(\Omega) = 1$,
3) Jeśli $A_1, A_2, \dots A_n, \dots \in \mathfrak{M}$ oraz dla każdych i, j $(i \neq j)$ jest $A_i \cap A_j = \emptyset$, to
$$P(A_1 \cup \dots \cup A_n \cup \dots) = $$
$$= P(A_1) + P(A_2) + \dots + P(A_n) + \dots.$$

W niektórych wykładach rachunku prawdopodobieństwa dystrybuantę definiuje się przez nierówność nieostrą:

$$F_X(x) = P(\{\omega \in \Omega: X(\omega \leqslant x\}),$$

co zmienia trochę własności dystrybuanty, ale w niczym nie zmienia ogólnej teorii.



DYSTRYBUANTA ILOŚCI OCZEK

Dla przykładu, jeśli X — ilość oczek w rzucie kostką, to wartość dystrybuanty tej zmiennej w punkcie 1 wynosi 0 (jest niemożliwe, by wypadło mniej niż 1 oczko), a $F_X(\pi) = 1/2$, bo zdarzenie „wypadło mniej niż π oczek" jest po prostu zdarzeniem „wypadło 1 oczko, 2 oczka lub 3 oczka". Trudniej zgadnąć, jaka byłaby dystrybuanta zmiennej Y: „wzrost losowo wybranego człowieka wyrażony w centymetrach", ale można na pewno twierdzić, że $F_Y(10) = 0$ i $F_Y(400) = 1$.

Niezależność zmiennych losowych

Układ zdarzeń $\{A_1, \ldots, A_n\}$ nazywamy niezależnym, jeśli zarówno dla tego układu, jak i dowolnego jego podukładu spełniony jest warunek: prawdopodobieństwo łącznego wystąpienia odpowiednich zdarzeń równe jest iloczynowi prawdopodobieństw tych zdarzeń.

Często zdarza się, że w jednym eksperymencie losowym wykonuje się szereg *doświadczeń niezależnych*: niezależnych w tym sensie, że wynik jednego doświadczenia nie może mieć wpływu na to, co wydarzy się w innych doświadczeniach. Formalnym odpowiednikiem wyników dwu niezależnych doświadczeń są *zdarzenia niezależne*, tzn. takie, że prawdopodobieństwo ich łącznego wystąpienia równe jest iloczynowi prawdopodobieństw każdego z nich z osobna (np. prawdopodobieństwo orła w każdym rzucie monety z osobna wynosi 1/2, a prawdopodobieństwo dwu orłów w dwu rzutach wynosi, wobec oczywistej niezależności wyników, 1/4). Zmienne losowe, odpowiadające pomiarom wyników niezależnych doświadczeń, nazywa się *zmiennymi niezależnymi*. Dla dwu zmiennych warunek niezależności formułuje się bardzo prosto: X i Y są niezależne wtedy i tylko wtedy, gdy dla dowolnych x i y ich dystrybuanty spełniają równość

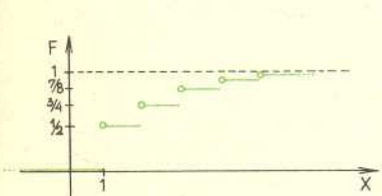
$$F_Y(x) \cdot F_Y(y) = P(X < x \text{ i } Y < y).$$

Możliwe jest też uogólnienie pojęcia niezależności na układy więcej niż dwu zmiennych losowych, tu poprzestaniemy na intuicyjnym rozumieniu niezależności zmiennych jako niezależności doświadczeń, w których dokonujemy odpowiednich pomiarów.

Typy zmiennych losowych

W praktyce są to przedziały oraz wszystkie takie zbiory, które można otrzymać z przedziałów przez tworzenie sum przeliczalnych i przeliczalnych iloczynów. Zbiory te nazywane są *zbiorami borelowskimi*.

Znajomość dystrybuanty zmiennej losowej wystarcza do określania prawdopodobieństw wszystkich praktycznie interesujących zbiorów liczb, w jakich może znaleźć się jej wartość. Mówi się, że dystrybuanta *wyznacza rozkład prawdopodobieństwa* zmiennej losowej. W praktyce jednak posługiwanie się dystrybuantą nie zawsze jest wygodne i w związku z tym wprowadza się inne jeszcze sposoby charakteryzowania rozkładu prawdopodobieństwa zmiennej losowej. Stosowane tu sposoby zależą jednak od własności dystrybuanty, czyli od *typu* zmiennej losowej. Mówi się, że zmienna losowa X jest typu skokowego (czasem, mniej poprawnie, typu dyskretnego), jeśli istnieje skończony lub przeliczalny ciąg liczb rzeczywistych $x_1, x_2, \ldots$ o tej własności, że dystrybuanta ma „skoki" w tych punktach i suma wszystkich tych skoków wynosi 1 (zob. rysunek). Wielkość „skoku" w punkcie x_i jest, jak łatwo stwierdzić, prawdopodobieństwem zbioru tych zdarzeń elementarnych, dla których zmienna X przyjmuje wartość x_i. W związku z tym wielkość tę oznacza się symbolem



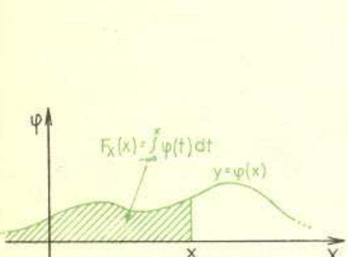
Np: $P(x=i) = \frac{1}{2^i}$ i $\sum_{i=1}^{\infty} \frac{1}{2^i} = 1$

$$P(X = x_i)$$

(„prawdopodobieństwo tego, że X przyjmie wartość x_i").
Ponieważ przy tym dla każdego $x \in \mathbf{R}$ mamy

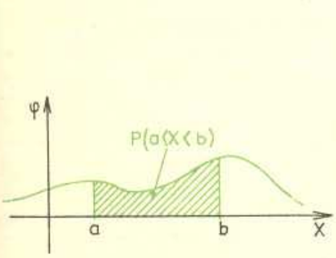
$$F_X(x) = \sum_{X < x_i} P(X = x_i)$$

(dlaczego?), to znajomość liczb $P(X = x_i)$ wyznacza rozkład prawdopodobieństwa. Przy tym empirycznym odpowiednikiem „skoku" w punkcie x_i jest częstość pojawiania się wyniku, któremu odpowiada miara x_i, i posługiwanie się wielkościami $P(X = x_i)$ dobrze przemawia do intuicji. Dlatego też rozkłady zmiennych skokowych charakteryzuje się w ten właśnie sposób. Mówi się, że zmienna losowa X jest *typu ciągłego*, jeśli istnieje taka nieujemna funkcja rzeczywista φ, że dla każdego $x \in \mathbf{R}$



$$F_X(x) = \int_{-\infty}^{x} \varphi(t)\,\mathrm{d}t.$$

Funkcja φ nazywa się gęstością rozkładu zmiennej X i oczywiście wyznacza rozkład zmiennej X (bo wyznacza jednoznacznie dystrybuantę). Jakie intuicje kryją się pod takim sposobem opisywania rozkładów? Wyobraźmy sobie pomiar masy losowo wybranego kamienia z Dunajca. Trudno tu z góry wyznaczyć wartości, jakie ta masa może przyjmować. Znacznie łatwiej pomyśleć, że masa może być dowolną liczbą rzeczywistą, czyli że nasza zmienna losowa może przyjmować każdą wartość. Ponieważ zaś czułość przyrządów pomiarowych jest i tak ograniczona, to zdarzeniami naprawdę nas interesującymi będą zdarzenia postaci „masa kamienia zawarta jest między a i b", a więc zdarzenia, których prawdopodobieństwo jest łatwe do obliczenia przy znajomości gęstości φ:

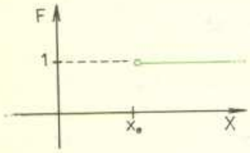


$$P(a < X < b) = F_X(b) - F_X(a) = \int_a^b \varphi(x)\,\mathrm{d}x.$$

Tak więc w przypadku zmiennej typu ciągłego też uzyskujemy interpretację „częstościową": występująca po prawej stronie całka jest teoretycznym odpowiednikiem częstości występowania takich wyników doświadczenia, których miara leży w przedziale (a, b). (Warto zauważyć, że jeśli różnica $b-a$ jest bardzo mała, a funkcja φ — ciągła, to częstość ta będzie w środku x przedziału (a, b) w przybliżeniu równa $\varphi(x) \cdot (b-a)$. Wynika stąd m.in., że istnieje możliwość przybliżenia rozkładów ciągłych rozkładami skokowymi.)
W dalszym ciągu będziemy zajmować się tylko rozkładami typu skokowego i typu ciągłego. Należy jednak pamiętać, że istnieją zmienne, które nie należą do żadnej z tych kategorii (tzw. *zmienne osobliwe* i zmienne, które dają się przedstawić jako sumy zmiennych różnych typów).

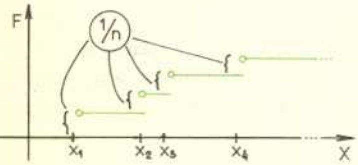
Zmienne o rozkładach dyskretnych

Rozkład jednopunktowy



Jest to rozkład takiej zmiennej, która na pewno, tj. z prawdopodobieństwem 1 przyjmuje pewną wartość x_0. Empiryczna interpretacja: mierzenie wyniku doświadczenia, w którym wszystko było z góry wiadome. („Eksperymenty" tego rodzaju bywają jednak urządzane i mogą mieć pewne walory pozanaukowe, np. szkoleniowe lub propagandowe.) Rozkład ten bywa użyteczny w rozważaniach teoretycznych.

Rozkład jednostajny (równomierny) dyskretny



Jest to rozkład takiej zmiennej X, która przyjmuje n różnych wartości $x_1, \ldots, x_n$ — każdą z tym samym prawdopodobieństwem:

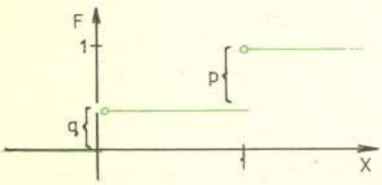
$$P(X = x_i) = \frac{1}{n}, \qquad i = 1, 2, \ldots, n.$$

Wprowadza się ją często przy opisie eksperymentu losowego, o którym wiadomo tyle tylko, że może się w nim wydarzyć każdy z n możliwych wyników, i nie ma jednocześnie podstaw do przypuszczenia, że któryś z wyników jest bardziej prawdopodobny, niż jakikolwiek inny (rzut kostką, monetą, wszelkie losowania ze zbioru skończonego ze zwracaniem, jak na przykład losowanie respondentów do badania ankietą socjologiczną itp.). W typowych zadaniach z rachunku prawdopodobieństwa sformułowania „wybór losowy", „na chybił-trafił", „przypadkowy" sugerują, iż w rozwiązaniu należy posłużyć się właśnie rozkładem tego typu. (Zob. też: *rozkład jednostajny ciągły.*)

Praktyka ta nazywana bywa „stosowaniem klasycznej definicji prawdopodobieństwa".

Rzucamy dwukrotnie kostką. Jaki jest rozkład zmiennej „suma oczek"
(Wsk.: Ponieważ nie ma podstaw do uznania, że któraś ze ścian kostki jest bardziej prawdopodobna niż inne, to przyjmujemy, że rozkład zmiennej „ilość oczek na kostce" jest jednostajny).

Rozkład dwupunktowy



Jest to rozkład zmiennej X, która z dodatnimi prawdopodobieństwami może przyjmować dwie wartości; najczęściej przyjmuje się, że są to 1 i 0. Skoki dystrybuanty w tych punktach oznacza się zazwyczaj odpowiednio przez p i q

$$P(X = 1) = p, \quad P(X = 0) = q, \quad p+q = 1.$$

Opisuje każdy taki eksperyment, w którym interesują nas tylko dwa wyniki, zwane umownie sukcesem lub porażką (wypadnięcie orła lub reszki, urodzenie chłopca lub dziewczynki itp.). Przyjęto „sukcesem" nazywać to zdarzenie, na którym zmienna X przyjmuje wartość 1.

Rozkład dwumianowy

Niech n będzie liczbą naturalną. Rozkładem dwumianowym nazywamy rozkład zmiennej S_n, która może przyjmować wartości $k = 0, 1, \ldots, n$ odpowiednio z prawdopodobieństwami

$$p_k = P(S_n = k) = \binom{n}{k} p^k q^{n-k}.$$

Powstaje przy opisywaniu tzw. *schematu Bernoulliego*, to jest ciągu niezależnych powtórzeń eksperymentu o rozkładzie dwupunktowym: S_n jest wtedy zmienną losową dającą się opisać jako „liczba sukcesów w n próbach". Przykłady doświadczeń i zjawisk, do opisu których wykorzystuje się rozkład dwumianowy:
— wielokrotny rzut monetą,
— wielokrotne losowanie (ze zwracaniem) z populacji, w której rozróżniamy tylko dwa rodzaje elementów,
— statystyczna kontrola jakości (np. żarówek: dobra — zła),
— testowanie wartości szczepionek i leków.

Omawiany w każdym podręczniku rachunku prawdopodobieństwa.

Wśród 20 żarówek losowo wybranych z dużej partii 10 przepaliło się w momencie pierwszego podłączenia do sieci. Producent twierdzi, że wypuszcza co najwyżej 5% braków. Jak wiarygodna jest wypowiedź producenta?
(Wiarygodnością nazwiemy — na użytek tego zadania — prawdopodobieństwo tego, co się nam przydarzyło, przy założeniu, że producent trafnie ocenia jakość swych wyrobów.)

Rozkład dwumianowy ważony

Powstaje przy opisywaniu ciągu n niezależnych prób typu *sukces — porażka*, jeśli prawdopodobieństwo sukcesu zależy od warunków, w jakich próby te były prowadzone, ale w ustalonych warunkach jest stałe. Załóżmy, że jest s możliwych warunków, przy czym i-te warunki występują z prawdopodobieństwem α_i ($i = 1, 2, \ldots, s$), a prawdopodobieństwo sukcesu w i-tych warunkach wynosi p_i.
Jeśli S_n jest liczbą sukcesów, to

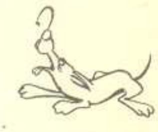
$$P(S_n = k) = \sum_{i=1}^{s} \alpha_i \binom{n}{k} p_i^k (1-p_i)^{n-k}$$

(jeśli wszystkie p_i są równe, to dostajemy rozkład dwumianowy). Rozkład ten bywa wykorzystywany przy analizie wyników diagnozy, o której wiadomo, że nie jest bezbłędna.

J. Neyman, *Zasady rachunku prawdopodobieństwa i statystyki matematycznej*, PWN, Warszawa 1969.

Gwiazdy klasyfikuje się ze względu na jasność jako karły (K), olbrzymy (O) lub superolbrzymy (S). Oznaczmy przez prawdopodobieństwo wystąpienia karła w danej rodzinie gwiazd, przez β-prawdopodobieństwo wystąpienia olbrzyma. Metoda klasyfikacji polega na badaniu widma gwiazd i nie jest bezbłędna. Prawdopodobieństwa *błędnej* klasyfikacji są następujące (dane fikcyjne):

w rzeczywistości \ Zakwalifikowany jako	K	O	S
K	×	$\frac{1}{4}$	$\frac{1}{8}$
O	$\frac{1}{4}$	×	$\frac{1}{5}$
S	$\frac{1}{10}$	$\frac{1}{5}$	×



Gwiazdę wybiera się losowo i przeprowadza trzy niezależne badania jej widma. Niech X będzie liczbą tych badań, w wyniku których gwiazdę zakwalifikowano jako olbrzyma. Jaki jest rozkład zmiennej X? (Wg J. Spławy-Neymana.)

Rozkład dwumianowy ujemny (rozkład Pascala)

Powstaje, podobnie jak dwumianowy, przy opisywaniu ciągu niezależnych powtórzeń doświadczenia o rozkładzie dwupunktowym, jeśli interesuje nas, ilu trzeba doświadczeń na to, aby pojawił się r-ty z kolei sukces. Oznaczmy przez X_r zmienną losową „łączna liczba porażek poprzedzających r-ty sukces". Zmienna ta może przyjmować każdą wartość $k = 0, 1, 2, \ldots$ i ma rozkład określony następująco:

$$P(X_r = k) = \binom{r+k-1}{k} p^r q^k.$$

Przykłady występowania:
— zadanie Banacha o matematyku, który nosił dwa pudełka zapałek,
— problemy czasu czekania, kolekcjonerstwa (zob. także: rozkład geometryczny).

W. Feller, *Wstęp do rachunku prawdopodobieństwa*, t. I, PWN, Warszawa 1971, oraz M. Fisz, *Rachunek prawdopodobieństwa i statystyka matematyczna*, PWN, Warszawa 1958.

(Zadanie Banacha o matematyku i zapałkach). Pewien matematyk nosi przy sobie dwa pudełka zapałek, jedno w prawej, drugie w lewej kieszeni. Kiedy potrzebuje zapałek, wybiera losowo z prawdopodobieństwem 1/2 zapałkę z jednej z kieszeni. Pudełka zawierają po N zapałek. Znaleźć prawdopodobieństwa:
a) że w chwili, w której z jednej kieszeni wyciągnięte zostanie puste pudełko, w drugiej będzie jeszcze r zapałek,
b) w chwili, w której jedno pudełko opróżni się (a nie, kiedy zostanie odkryte jako puste), w drugim będzie jeszcze r zapałek. Wykazać, że prawdopodobieństwo tego, iż pudełko, które opróżni się pierwsze, nie zostanie odkryte jako puste wynosi $\binom{2N}{N} 2^{-2N-1}$. (Wg W. Fellera.)

Rozkład geometryczny

Powstaje w sytuacji opisywanej powyżej dla $r = 1$. Wartość zmiennej losowej X_1 interpretuje się często jako „czas czekania na pojawienie się pierwszego sukcesu" w schemacie Bernoulliego. Jest oczywiście

$$P(X_1 = k) = q^k p.$$

Pojawia się w opisach bardzo wielu sytuacji empirycznych, w których interesującym wynikiem jest rzeczywisty czas czekania. Jest również rozkładem granicznym dla statystyki Bosego-Einsteina (por. artykuł L.T. Kubika).

Omawiany jest w większości podręczników rachunku prawdopodobieństwa. W szczególności zob. W. Feller, *Wstęp...*, t. I.

Ktoś ma w kieszeni 5 kluczy, z których tylko 1 pasuje do zamka. Po stwierdzeniu, że losowo z kieszeni wybrany klucz nie pasuje, odkłada go — z nie znanych nam bliżej powodów — z powrotem do tej samej kieszeni. Jakie jest prawdopodobieństwo tego, że otworzenie zamka będzie wymagało co najmniej 10 prób?

Dysponujemy 10 kluczami, z których 2 pasują do zamka. Po stwierdzeniu, że losowo z kieszeni wybrany klucz nie pasuje, odkładamy go na bok. Pokazać, że najbardziej prawdopodobna liczba prób wynosi 1 (tzn. że już w pierwszej próbie trafimy na dobry klucz!). (Wg L. T. Kubika; zadanie to nie ma żadnego związku z rozkładem geometrycznym, ale zamieszczamy je tu ze względu na pokrewieństwo tematyczne z poprzednim zadaniem.)

Rozkład hipergeometryczny

Rozkład ten powstaje przy opisywaniu następującego abstrakcyjnego eksperymentu: W urnie jest n_1 kul czarnych i $n_2 = n - n_1$ kul czerwonych. Losujemy bez zwracania r kul. Poszukujemy prawdopodobieństwa tego, że wśród wylosowanych będzie dokładnie k czarnych. Jeśli X oznacza zmienną „liczba wylosowanych kul czarnych", to

$$P(X = k) = \frac{\binom{n_1}{k}\binom{n-n_1}{r-k}}{\binom{n}{r}} \qquad k = 0, 1, \ldots, \min\,(n_1, r)$$

Schemat ten odpowiada wielu sytuacjom praktycznym, w których nie jest znane n lub n_1. W szczególności był on wykorzystywany do statystycznej kontroli jakości (nieznane n_1) i szacowania populacji zwierząt na podstawie liczby powtórnych złapań zwierząt oznaczonych przy pierwszym złapaniu (znane n_1, nieznane n).

W. Feller, *Wstęp...*, t. I; M. Fisz, *Rachunek...*

W zamkniętym stawie odłowiono 1000 ryb, które oznaczono czerwonymi plamami i wypuszczono. Po pewnym czasie dokonano nowego połowu 1000 ryb i znaleziono wśród nich 100 ryb oznaczonych. Oszacować prawdopodobieństwo tego zdarzenia przy założeniu, że w jeziorze jest a) 1900, b) 10^6 ryb. (Wsk. $n! \approx \sqrt{2\pi n} \cdot n^n \cdot e^{-n}$ — wzór Stirlinga.)

Oznaczmy przez $q(n)$ prawdopodobieństwo tego wydarzenia przy założeniu, że w jeziorze jest n ryb. Udowodnić, że $q(n)$ jest największe przy $n = 10\,000$.) Wsk. obliczyć iloraz $q(n)/q(n-1)$.) (Wg W. Fellera.)

Rozkład Pólyi

Rozważmy urnę napełnioną tak jak poprzednio, z tym, że po wyciągnięciu każdej kuli wrzucamy ją z powrotem, dodając jeszcze s kul tego samego koloru (s może być ujemne). Niech X będzie określona tak jak poprzednio: X przyjmuje wartość k, jeśli wśród wylosowanych r kul było k czarnych. Rozkład tej zmiennej nazywa się rozkładem Pólyi:

$$P(X = k) = \binom{r}{k} \frac{[n_1(n_1+s) \ldots (n_1+(k-1)s)] \cdot [n_2(n_2+s) \ldots (n_2+(r-k-1)s)]}{n(n+s) \cdot \ldots \cdot (n+(r-1)s)}$$

(dla $s = -1$ otrzymujemy rozkład hipergeometryczny, a dla $s = 0$ — jaki?). Schemat ten odpowiada przy $s > 0$ przybliżonemu opisowi rozprzestrzeniania się chorób zakaźnych, przy $s < 0$ — opisowi efektywności stosowania zasad BHP (wypadek zwiększa czujność i zmniejsza szansę powstania następnego, brak wypadku — na odwrót).

W. Feller, *Wstęp...*, t. I.

Rozkład Poissona

Jest to rozkład zmiennej X przyjmującej wartości naturalne $k = 0, 1, 2, \ldots$ odpowiednio z prawdopodobieństwami

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!},$$

gdzie λ jest pewną ustaloną liczbą dodatnią.
Jedną z możliwych do wyobrażenia realizacji tej zmiennej jest sytuacja następująca. Mamy nieskończoną objętość ciasta z rodzynkami w takiej ilości, że na ustaloną objętość V ciasta przypada średnio λ rodzynków, przy czym rodzynki nie oddziaływają na siebie wzajemnie (nie odpychają ani przyciągają). Wycinamy z ciasta losowo porcję o objętości V i pieczemy z niej placek: prawdopodobieństwo, że w placku znajdzie się k rodzynków wynosi właśnie $e^{-\lambda} \lambda^k/k!$. Inaczej mówiąc — zmienna losowa „liczba rodzynków w placku" ma rozkład Poissona.
Zmienne o rozkładzie Poissona odgrywają dużą rolę zarówno w rozważaniach teoretycznych jak i zastosowaniach praktycznych teorii prawdopodobieństwa. Wynika to między innymi z faktu, że rozkład Poissona bywa dobrym przybliżeniem rozkładu dwumianowego: jeśli w schemacie Bernoulliego prawdopodobieństwo sukcesu jest tak małe, że liczba $\lambda \stackrel{\mathrm{df}}{=} np$ jest też mała, to

$$P(S_n = k) = \binom{n}{k} p^k q^{n-k} \approx e^{-\lambda} \frac{\lambda^k}{k!},$$

przy czym przybliżenie jest tym lepsze, im większe n i im mniejsze λ. (Rozkład zmiennej „liczba rodzynków w placku" jest przy skończonej objętości ciasta dwumianowy, ale przy dużej objętości ciasta i małej domieszce rodzynków — w przybliżeniu poissonowski). Rozkład Poissona jest też rozkładem granicznym dla tzw. statystyk Maxwella-Boltzmanna (zob. artykuł L.T. Kubika).
Inne przykłady sytuacji, w opisie których wykorzystuje się zmienne o tym rozkładzie:
— rozpad radioaktywny (doświadczenia Rutherforda i Geigera: liczba rozpadów na jednostkę czasu ma rozkład Poissona),
— w telekomunikacji przewodowej (liczba zgłoszeń do centrali na jednostkę czasu, liczba błędnych połączeń, liczba prób potrzebnych do uzyskania połączenia — można opisywać jako zmienne o tym rozkładzie, zob. też *rozkład wykładniczy*),

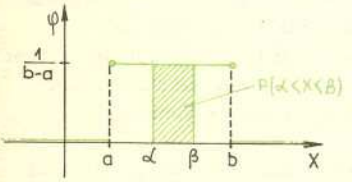
— analiza przestrzennego rozmieszczenia wybuchów bomb latających w Londynie w czasie II wojny światowej pozwoliła ustalić, że rozkład zmiennej „liczba wybuchów na jednostkę powierzchni" jest w przybliżeniu poissonowski, co z kolei pozwoliło odrzucić przypuszczenie, że pewne rejony miasta są celowo atakowane przez nieprzyjaciela intensywniej, niż inne,
— M. Fisz podaje też, że zmienna losowa „liczba zejść śmiertelnych spowodowanych kopnięciem przez konia w okresie 1 roku w 1 korpusie armii pruskiej" (dane z 10 korpusów kawalerii za okres 20 lat) ma rozkład dający się dobrze przybliżyć rozkładem Poissona. (Interpretację pozostawiamy Czytelnikowi.)

M. Fisz, *Rachunek...*, W. Feller, *Wstęp...*, t. I, R. Zieliński, *Rachunek prawdopodobieństwa z elementami statystyki matematycznej* PZWS, Warszawa 1973.

Do ciasta wrzucono 900 rodzynków i usmażono 1000 jednakowych pączków. Co jest prawdopodobniejsze — czy to, że w zakupionym przez nas pączku będzie 1 rodzynek, czy że nie będzie żadnego?
Obliczyć, ile co najmniej trzeba wrzucić rodzynków, by prawdopodobieństwo tego, że zakupiony pączek będzie miał co najmniej 1 rodzynek, było większe od 0,99.

Rozkłady ciągłe

Rozkład jednostajny (równomierny)



Najprostszy z rozkładów ciągłych. Mówi się, że zmienna losowa X ma rozkład jednostajny na przedziale (a, b), jeśli jej gęstość φ wyraża się wzorem

$$\varphi(x) = \begin{cases} \dfrac{1}{b-a} & \text{dla } x \in (a, b). \\ 0 & \text{dla } x \notin (a, b), \end{cases}$$

Jak łatwo zauważyć, dla dowolnego $I = (\alpha, \beta) \subset (a, b)$ jest

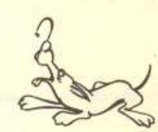
$$P(\alpha < X < \beta) = \frac{1}{b-a} \int_{\alpha}^{\beta} 1 \, dx = \frac{\beta - \alpha}{b-a},$$

Praktyka ta nazywana bywa „stosowaniem geometrycznej definicji prawdopodobieństwa"

skąd wynika, że prawdopodobieństwo iż zmienna X przyjmie wartość z pewnego przedziału I zawartego w (a, b) zależy tylko od długości tego przedziału (ale nie od położenia tego przedziału w (a, b)). (Tę samą własność mają przedziały domknięte i jednostronnie domknięte.) Rozkład ten bywa wykorzystywany w opisie takich eksperymentów, o których wiemy tyle tylko, że miary możliwych wyników zawarte są w pewnym przedziale skończonym i nie mamy powodu do przypuszczenia, że pewne rejony tego przedziału są bardziej uprzywilejowane niż inne. „Losowo wybrany punkt z przedziału (0, 1)", „przypadkowy kierunek na płaszczyźnie" (mierzony kątem, jaki tworzy z ustalonym kierunkiem) — to przykłady sformułowań sugerujących przyjęcie rozkładu jednostajnego odpowiednio na przedziałach $(0, 1)$, $\langle 0, 2\pi)$.

Omawiany w każdym podr. rachunku prawdopodobieństwa, zob. np.: R. Zieliński, *Rachunek...*

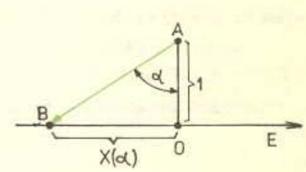
Z punktu A (źródła światła) wybiega w przypadkowym kierunku kwant światła padający na prostą E (ekran) w punkcie B. Niech O będzie rzutem prostokątnym A na E, $OA = 1$ i niech X oznacza odległość zorientowaną OB (w sytuacji przedstawionej na rysunku obok wielkość X jest ujemna). Znaleźć gęstość $\varphi(x)$ rozkładu zmiennej X.
Szkic rozwiązania: Sformułowanie zadania upoważnia do przyjęcia, że kąt zorientowany α między AO a AB ma rozkład jednostajny na przedziale $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Ponieważ $X(\alpha) = \operatorname{tg} \alpha$, to



$$F_X(x) = P(X(\alpha) < x) = P\left(-\frac{\pi}{2} < \alpha < \operatorname{arctg} x\right) = \frac{1}{\frac{\pi}{2} - \left(-\frac{\pi}{2}\right)} \int_{-\frac{\pi}{2}}^{\operatorname{arctg} x} 1 d\alpha = \frac{1}{\pi}\left(\operatorname{arctg} x + \frac{\pi}{2}\right).$$

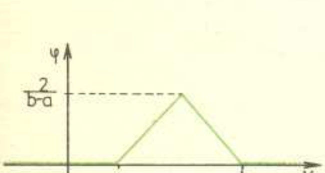
Stąd

$$\varphi(x) = F_X(x)' = \frac{1}{\pi}(\operatorname{arctg} x)' = \frac{1}{\pi} \frac{1}{1+x^2},$$



a więc zmienna X ma *rozkład Cauchy'ego* — zob. dalej. (Wg W. Fellera.)

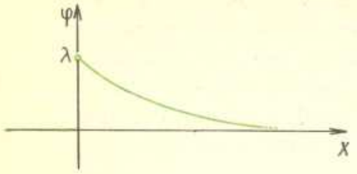
Rozkład trójkątny



Zmienna X ma rozkład trójkątny na przedziale (a, b), jeśli jej gęstość znika poza tym przedziałem, a wykres gęstości na (a, b) tworzy trójkąt równoramienny o polu 1 (tj. o wysokości $2/(b-a)$, zob. rys.). Powstaje przy rozpatrywaniu sumy dwu niezależnych zmiennych losowych o tym samym rozkładzie jednostajnym: jeśli np. X i Y są niezależne i każda z nich ma rozkład jednostajny na $(0, 1)$, to $X + Y$ ma rozkład trójkątny na $(0, 2)$. Podstawowe zastosowanie: do formułowania przykładów i zadań w podręcznikach rachunku prawdopodobieństwa.

Znaleźć rozkład sumy dwu niezależnych zmiennych losowych o rozkładach jednostajnych na (0,1). (Wsk.: skorzystać z definicji niezależności i definicji dystrybuanty.)

Rozkład wykładniczy



Jest to rozkład zmiennej X o gęstości

$$\varphi(x) = \begin{cases} \lambda e^{-\lambda x} & \text{dla} \quad x > 0, \\ 0 & \text{dla} \quad x \leqslant 0, \end{cases}$$

gdzie λ jest pewną liczbą dodatnią. Jedną z możliwych do wyobrażenia interpretacji tej zmiennej jest „czas czekania bez pamięci": jeśli to, na co czekamy, pojawia się zupełnie przypadkowo, i w taki sposób, że to, jak długo czekamy nie ma najmniejszego wpływu na to, ile jeszcze będziemy czekali, to zmienna „czas czekania" ma rozkład wykładniczy. Parametr λ jest odwrotnością „średniego czasu czekania". (Rozkład ten jest więc ciągłym odpowiednikiem rozkładu geometrycznego; przy dużej ilości prób rozkład wykładniczy jest dobrym przybliżeniem rozkładu geometrycznego.)

Rozkład wykładniczy wykorzystywany jest z powodzeniem do analizy wielu procesów przyrodniczych i społecznych, m.in. do

— opisu zjawisk promieniotwórczych (np. w promieniowaniu α czas czekania na emisję pojedynczej cząstki α ma rozkład wykładniczy),

— opisu zjawisk występujących w telekomunikacji przewodowej (czas czekania centrali na zgłoszenie się abonenta, czas trwania rozmów telefonicznych itp.), i ogólniej

— teorii obsługi masowej (teorii kolejek).

Różnorodność zastosowań rozkładu wykładniczego wynika również z jego związku z rozkładem Poissona: jeśli opisywany proces ma tę własność, że liczba interesujących nas zdarzeń w przedziałach czasu o ustalonej długości ma rozkład Poissona i przy tym średnia liczba zdarzeń jest proporcjonalna do tej długości, to czas czekania na takie zdarzenie ma rozkład wykładniczy.

Rozkład ten występuje również przy opisie wielu zjawisk innych, niż wyżej opisane, typów. Przykłady:

— rozkład zasięgu widzialności w rzadkim lesie, jeśli kierunek patrzenia jest losowy (a rozkład kierunków — jednostajny); jeśli zastąpimy drzewa gwiazdami, też będzie dobrze,

— rozkład prawdopodobieństwa tego, że nitka o długości t wytrzyma ustalone obciążenie (t jest tu zmienną losową).

R. Zieliński, *Rachunek*...; W. Feller, *Wstęp*..., t. I i II; i in.

Rutheford i Geiger stwierdzili, ze liczby cząstek α wypromieniowywanych przez pewną substancję w 7,5 sekundowych odcinkach czasu mają w przybliżeniu rozkład Poissona ze średnią 3,87 cząstek na okres. Zatem średni czas czekania na kolejną emisję wynosi

$$\frac{7,5}{3,87} = \frac{1}{0,516}$$

i wobec tego rozkład czasu czekania ma gęstość

$$\varphi(t) = \begin{cases} 0,516\, e^{-0,516t} & \text{dla } t > 0, \\ 0 & \text{dla } t \leqslant 0. \end{cases}$$

Jakie było prawdopodobieństwo tego, że między dwiema kolejnymi emisjami upłynie więcej niż 10 sekund (Wg R. Zielińskiego.)

Formalnie „brak pamięci" zmiennej losowej X można opisać następująco:

(*) $$P(X > t+s \mid X > t) = P(X > s) \quad \text{dla } s, t > 0.$$

Jeśli przyjąć

$$P(X > s) \stackrel{\mathrm{df}}{=} \varphi(s)$$

to okaże się, że z (*) wynika, iż

$$\begin{cases} \varphi \text{ jest monotoniczna}, \\ \varphi(s+t) = \varphi(s) \cdot \varphi(t) \quad \text{dla } s, t > 0. \end{cases}$$

Stąd z kolei łatwo wywnioskować, że $\varphi(t)$ jest dodatnia dla $t > 0$ i wobec tego $\psi \stackrel{\mathrm{df}}{=} \log \varphi$ jest też monotoniczna i spełnia

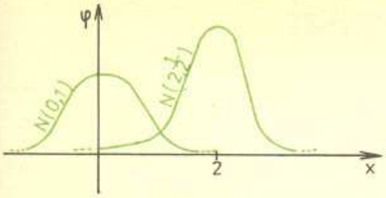
(**) $$\psi(s+t) = \psi(s) + \psi(t).$$

Jedynym monotonicznym rozwiązaniem (**) jest funkcja liniowa (por. art. M. Kuczmy, Delta 1/1977), skąd wynika, że *Rozkład wykładniczy jest jedynym rozkładem opisującym „procesy bez pamięci"*.

Rozkład normalny (Gaussa)

Zmienna losowa ma rozkład normalny o wartości średniej m i odchyleniu standardowym $\sigma > 0$ (ozn. $N(m, \sigma)$), jeśli jej gęstość wyraża się wzorem

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}}\, e^{-\frac{(x-m)^2}{2\sigma^2}}.$$



Dla $\sigma = 1$ i $m = 0$ otrzymujemy gęstość *standaryzowanego* rozkładu normalnego $N(0, 1)$:

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Wykres tej gęstości nazywany jest *krzywą Gaussa.*
Jest to — bez przesady — najważniejszy rozkład prawdopodobieństwa i znajduje zastosowanie praktyczne we wszystkich problemach, w których mamy do czynienia bądź ze średnimi z wielu pomiarów, bądź pomiarami zjawisk, na które oddziałuje wiele przypadkowych i niezależnych czynników. Teoretycznym uzasadnieniem tego faktu są tzw. *centralne twierdzenia graniczne*, których sens sprowadza się do tego, że *jeśli* $X_1, X_2, \dots X_n, \dots$ *jest* „dostatecznie porządnym" *ciągiem niezależnych zmiennych losowych, to dla dużych n rozkład średniej arytmetycznej*

$$Y_n = \frac{1}{n}(X_1 + \dots + X_n)$$

jest w przybliżeniu normalny. W szczególności np. jeśli Z_n jest średnią liczbą sukcesów w schemacie Bernoulliego podzieloną przez $\sqrt{npq}$, to ciąg dystrybuant F_{Z_n} zbiega do dystrybuanty rozkładu $N(p, 1)$ (jest to twierdzenie de Moivre'a-Laplace'a).

Sformułowanie to jest — delikatnie mówiąc — niezbyt precyzyjne. Z braku miejsca poprzestajemy jednak na nim, zapraszając Czytelnika do ew. lektury jakiegokolwiek z wymienionych tu podręczników.

Rozkład ten omawiany jest we wszystkich podręcznikach rachunku prawdopodobieństwa i statystyki i tam też można znaleźć liczne przykłady jego zastosowania (zob. też artykuł R. Zielińskiego).

Rozkład Cauchy'ego

Jest to rozkład zmiennej ciągłej o gęstości

$$\varphi_t(x) = \frac{1}{\pi}\,\frac{t}{t^2+x^2},$$

gdzie $t > 0$ jest dowolnym parametrem. Rozkład ten ma interesującą własność: jeśli $X_1, \dots, X_n$ są niezależnymi zmiennymi losowymi o gęstościach odpowiednio $\varphi_{t_1}(x), \dots, \varphi_{t_n}(x)$, to ich suma $X_1 + \dots + X_n$ ma też rozkład Cauchy'ego o gęstości $\varphi_T(x)$, gdzie $T = t_1 + \dots + t_n$. Konsekwencją tej własności jest to, że jeśli zmienne $X_1, \dots, X_n$ mają jednakowe rozkłady o gęstości $\varphi_t(x)$, to ich średnia arytmetyczna $Y = \frac{1}{n}(X_1 + \dots X_n)$ ma też rozkład o gęstości $\varphi_t(x)$.
Wynika stąd z kolei, że nie wszystko w przyrodzie jest normalne: ciąg zmiennych $X_1, \dots, X_n, \dots$ o jednakowym rozkładzie Cauchy'ego jest ciągiem, dla którego nie zachodzi centralne twierdzenie graniczne.

W. Feller, *Wstęp...*, t. II

(c.d. zadania o kwancie padającym na ekran).
Wykazać, że oświetlenie nieskończonego ekranu płaskiego punktowym źródłem światła umieszczonym w punkcie poza ekranem wyraża się wzorem

$$c \cdot \frac{1}{\pi} \cdot \frac{k}{k^2+x^2},$$

gdzie c, k — stałe charakteryzujące źródło światła i jego odległość od ekranu.
(Wsk. Źródło światła to źródło wielu niezależnych fotonów.)

Rozkłady χ^2 i rozkłady t Studenta

Listę naszą zamykają dwie klasy rozkładów, które odgrywają rolę szczególnie ważną w zastosowaniach rachunku prawdopodobieństwa do wnioskowania o pewnych charakterystykach liczbowych badanego zjawiska na podstawie wyników pomiaru, tj. w problematyce wnioskowania statystycznego.
Rozkładem χ^2 o k stopniach swobody nazywamy rozkład takiej zmiennej, która jest sumą kwadratów k ($k = 1, 2, \dots$) niezależnych zmiennych losowych o rozkładach normalnych $N(0, 1)$.
Inaczej mówiąc, jeśli $X_1, \dots, X_k$ mają rozkład normalny $N(0, 1)$ i są niezależne, to zmienna

$$Z = X_1^2 + \dots + X_k^2$$

ma rozkład χ^2 o k stopniach swobody.
Rozkładem *t Studenta* o k stopniach swobody nazywamy rozkład takiej zmiennej T, która jest postaci

$$T = \frac{X}{\sqrt{\frac{1}{k} Z}},$$

gdzie X i Z są zmiennymi losowymi niezależnymi, X ma rozkład normalny $N(0, 1)$, a Z ma rozkład χ^2 o k stopniach swobody. Dystrybuanty i gęstości tych rozkładów omawia w swym artykule R. Zieliński. W podanym tam przykładzie ich wykorzystania mamy $k = n - 1 = 4$ stopnie swobody. Jak łatwo sprawdzić, rozkład t o 1 stopniu swobody jest rozkładem Cauchy'ego.

Zob. R. Zieliński, *Rachunek...*; S. Zubrzycki, *Wykłady z rachunku prawdopodobieństwa i statystyki matematycznej*, PWN, Warszawa 1966.