

Prawdopodobieństwo i rozkłady

Pojedynczego rzutu monetą nie da się przewidzieć, ale tysiąca rzutów już tak: orłów będzie ok. 500, w ok. 95% przypadków między 470 a 530. Podobnie z rozpadem pojedynczego jądra atomowego i z liczbą zliczeń licznika Geigera w ciągu sekundy. Ta lekcja buduje język, w którym fizyka mówi o przypadku: prawdopodobieństwo, zmienna losowa, wartość oczekiwana i odchylenie standardowe. Na koniec dochodzi do trzech rozkładów, które wystarczają w większości zastosowań: dwumianowego, Poissona i normalnego. Pokazuje też, dlaczego krzywa Gaussa pojawia się wszędzie.

CZAS

ok. 3 h teorii + 2,5 h zadań

ZADANIA

18 zadań w treści, z odpowiedziami i odpowiedziami

ŹRÓDŁA

Feynmana
wykłady z fizyki, t. I, 6-1...6-5; M. L. Boas,
Mathematical Methods in the Physical Sciences, wyd. 3, rozdz. 15

W TEJ LEKCJI

- 0 Prolog: gry Feynmana w monetę
- 1 Prawdopodobieństwo i jego reguły
- 2 Zmienne losowe, średnia i wariancja
- 3 Rozkład dwumianowy i rozkład Poissona
- 4 Rozkład normalny i wzór Stirlinga
- 5 Błędne intuicje
- 6 Sprawdź się
- 7 Karta wzorów i co dalej

Prolog: gry Feynmana w monetę

W lekcji 0.5 Jednostki, analiza wymiarowa, szacowanie i niepewność pomiaru przytoczyliśmy doświadczenie Feynmana: sto gier po 30 rzutów monetą. Spodziewana liczba orłów to 15, ale w pojedynczej grze wychodziło raz mniej, raz więcej. Feynman pyta, jak często należy się spodziewać dokładnie 15, a jak często wyniku odległego od średniej. FLP I-6-2 Żeby odpowiedzieć liczbą, potrzebujemy rachunku prawdopodobieństwa.

Pytanie na start. Jakie jest prawdopodobieństwo dokładnie 15 orłów w 30 rzutach? A co najmniej 20? Odpowiedzi w przykładach 3.1 i 4.2.

Prawdopodobieństwo i jego reguły

Feynman definiuje prawdopodobieństwo zdarzenia jako ułamek prób, w których ono zachodzi, gdy prób jest bardzo dużo: $P(A) \approx N_A/N$. FLP I-6-1 Gdy wszystkie wyniki są równie możliwe (kostka, moneta), $P(A)$ to liczba wyników sprzyjających podzielona przez liczbę wszystkich. Z tego wynikają reguły:

REGUŁY RACHUNKU PRAWDOPODOBIEŃSTWA

- $0 \leq P(A) \leq 1$; zdarzenie pewne ma $P = 1$; $P(\text{nie } A) = 1 - P(A)$.
- suma: $P(A \text{ lub } B) = P(A) + P(B) - P(A \text{ i } B)$; dla zdarzeń wykluczających się $P(A) + P(B)$.
- prawdopodobieństwo warunkowe: $P(A | B) = \frac{P(A \text{ i } B)}{P(B)}$; zdarzenia niezależne: $P(A \text{ i } B) = P(A) P(B)$.
- wzór Bayesa: $P(A | B) = \frac{P(B | A) P(A)}{P(B)}$.

Zasady te przyjmuje się jako aksjomaty (Kołmogorow); interpretacja częstościowa mówi, jak je mierzyć. por. Boas §15.2-15.3

Przykład 1.1

Dwie kostki

Rzucamy dwiema kostkami. Jakie jest prawdopodobieństwo, że suma oczek wyniesie 7? Czy zdarzenia „na pierwszej szóstka” i „suma 7” są niezależne?

- 1 Wyników jest $6 \cdot 6 = 36$, równie możliwych. Sumę 7 dają $(1, 6)$, $(2, 5)$, ..., $(6, 1)$: 6 wyników. $P = \frac{6}{36} = \frac{1}{6}$.
- 2 $P(\text{szóstka na pierwszej}) = \frac{1}{6}$, a $P(\text{szóstka i suma 7}) = P((6, 1)) = \frac{1}{36} = \frac{1}{6} \cdot \frac{1}{6}$.

Wynik: $P(\text{suma 7}) = \frac{1}{6}$; te dwa zdarzenia są niezależne.

Sprawdzenie: to zaskakuje: wiedza o pierwszej kostce nie zmienia szansy na sumę 7, bo dla każdego wyniku pierwszej jest dokładnie jeden pasujący wynik drugiej ✓. Dla sumy 12 już tak nie jest: $P(12) = \frac{1}{36}$, ale $P(12 | \text{szóstka}) = \frac{1}{6}$.

Przykład 1.2

Dodatni wynik testu

Choroba dotyka 1% populacji. Test wykrywa ją u 99% chorych (czułość) i daje wynik ujemny u 95% zdrowych (swoistość). Twój wynik jest dodatni. Jakie jest prawdopodobieństwo, że jesteś chory?

- 1 $P(+) = P(+ | \text{chory})P(\text{chory}) + P(+ | \text{zdrowy})P(\text{zdrowy}) = 0,99 \cdot 0,01 + 0,05 \cdot 0,99 = 0,0594$.
- 2 Bayes: $P(\text{chory} | +) = \frac{0,99 \cdot 0,01}{0,0594} \approx 0,167$.

Wynik: ok. 17%, mimo „dokładności” testu 95–99%.

Sprawdzenie na liczbach: z 10 000 osób 100 jest chorych, 99 z nich ma wynik dodatni. Z 9900 zdrowych 495 ma wynik fałszywie dodatni. Dodatnich jest 594, chorych wśród nich 99: $\frac{99}{594} = \frac{1}{6}$ ✓. Rzadka choroba oznacza, że fałszywe alarmy przeważają.

- 1 Jakie jest prawdopodobieństwo, że suma oczek na dwóch kostkach wyniesie co najmniej 10?
■ ■ ■
- 2 Kawaler de Méré zakładał się, że w 4 rzutach kostką wypadnie co najmniej jedna szóstka, a w 24 rzutach dwiema kostkami co najmniej jedna para szóstek. Pierwszy zakład wygrywał częściej niż przegrywał, drugi nie. Oblicz oba prawdopodobieństwa.
■ ■ ■
- 3 Jakie jest prawdopodobieństwo, że wśród 23 osób co najmniej dwie mają urodziny tego samego dnia (pomiń 29 lutego, załóż równe szanse dni)?
■ ■ ■
- 4 Powtórz przykład 1.2 dla choroby występującej u 0,1% populacji. Jak zmieni się wynik, jeśli osoba z dodatnim wynikiem robi drugi, niezależny test i znowu wyjdzie dodatni?
■ ■ ■
- 5 Problem Monty'ego Halla: nagroda jest za jednym z trzech drzwi. Wybierasz drzwi, prowadzący (który wie, gdzie jest nagroda) otwiera inne, puste, i proponuje zmianę. Czy warto zmienić? Uzasadnij prawdopodobieństwem warunkowym.
■ ■ ■

■ ■ ■ podstawowe ■ ■ ■ średnie ■ ■ ■ feynmanowskie

2 Zmienne losowe, średnia i wariancja

Zmienna losowa X to liczba zależna od wyniku doświadczenia: liczba oczek, liczba orłów, czas do rozpadu. Opisuje ją **rozkład**: prawdopodobieństwa $p_k = P(X = x_k)$ dla zmiennej dyskretnej albo gęstość $f(x)$ dla ciągłej, gdzie $P(a < X < b) = \int_a^b f(x) dx$. Feynman wprowadza gęstość na przykładzie odległości po błędzeniu losowym: przy wielu krokach rozkład staje się gładką krzywą. (FLP I-6-4)

WARTOŚĆ OCZEKIWANA I WARIANCJA

$$\mu = E[X] = \sum_k x_k p_k \left(= \int x f(x) dx \right), \quad \sigma^2 = \text{Var } X = E[(X - \mu)^2] = E[X^2] - \mu^2.$$

$\sigma = \sqrt{\text{Var } X}$ to **odchylenie standardowe**. Wartość oczekiwana jest liniowa: $E[aX + bY] = aE[X] + bE[Y]$ zawsze. Dla zmiennych niezależnych wariancje się dodają: $\text{Var}(X + Y) = \text{Var } X + \text{Var } Y$. Stąd suma N niezależnych kopii ma odchylenie $\sigma\sqrt{N}$, a średnia z nich $\sigma/\sqrt{N}$: wzór z lekcji 0.5 Jednostki, analiza wymiarowa, szacowanie i niepewność pomiaru. (por. Boas §15.5–15.6)

Przykład 2.1

Rzut kostką

Oblicz wartość oczekiwaną, wariancję i odchylenie standardowe liczby oczek.

- 1 $\mu = \frac{1}{6}(1 + 2 + \dots + 6) = 3,5.$
- 2 $E[X^2] = \frac{1}{6}(1 + 4 + 9 + 16 + 25 + 36) = \frac{91}{6}; \sigma^2 = \frac{91}{6} - 3,5^2 = \frac{35}{12} \approx 2,917.$
- 3 $\sigma \approx 1,708.$

Wynik: $\mu = 3,5, \sigma^2 = \frac{35}{12}, \sigma \approx 1,71.$

Sprawdzenie: bezpośrednio: $\frac{1}{6}(2,5^2 + 1,5^2 + 0,5^2 + 0,5^2 + 1,5^2 + 2,5^2) = \frac{17,5}{6} = \frac{35}{12} \checkmark.$
Wartość oczekiwana 3,5 nigdy nie wypada: to średnia, a nie najbardziej prawdopodobny wynik.

ZADANIA · CZĘŚĆ 2

4 zadania · ok. 30 min

- 1 Zmienna X przyjmuje 1 z prawdopodobieństwem p i 0 z prawdopodobieństwem $1 - p$ (np. „orzeł”). Oblicz $E[X]$ i $\text{Var } X$. Dla jakiego p wariancja jest największa?
- 2 Oblicz wartość oczekiwaną i odchylenie standardowe sumy oczek na dwóch kostkach, korzystając z przykładu 2.1.
- 3 Zmienna ma rozkład jednostajny na $[0; 1]$ ($f = 1$). Oblicz jej odchylenie standardowe. Przyrząd cyfrowy zaokrągla do działki Δ ; jakie jest odchylenie standardowe błędu zaokrąglenia?
- 4 Czas do rozpadu jądra ma gęstość $f(t) = \lambda e^{-\lambda t}, t \geq 0$. Oblicz $E[T]$ i $\text{Var } T$. Pokaż, że $P(T > s + t | T > s) = P(T > t)$, i wyjaśnij, co to znaczy fizycznie.

□□□ podstawowe ■■■ średnie ■■■■ feynmanowskie

3 Rozkład dwumianowy i rozkład Poissona

Powtarzamy n razy niezależną próbę, która kończy się sukcesem z prawdopodobieństwem p . Konkretny ciąg z k sukcesami ma prawdopodobieństwo $p^k(1 - p)^{n-k}$, a takich ciągów jest $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ (liczba sposobów wybrania, które k prób było udanych).

ROZKŁAD DWUMIANOWY

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \mu = np, \quad \sigma^2 = np(1-p).$$

Średnia i wariancja wynikają z części 2: liczba sukcesów to suma n niezależnych zmiennych z zadania 2.1. (por. Boas §15.4, 15.7)

Przykład 3.1

Trzydzieści rzutów

Oblicz prawdopodobieństwo dokładnie 15 orłów w 30 rzutach, odchylenie standardowe liczby orłów i prawdopodobieństwo wyniku od 12 do 18.

- 1 $P(15) = \binom{30}{15} 2^{-30} = \frac{155\,117\,520}{1\,073\,741\,824} \approx 0,144.$
- 2 $\sigma = \sqrt{30 \cdot \frac{1}{2} \cdot \frac{1}{2}} \approx 2,74.$
- 3 Sumując $P(k)$ od 12 do 18: $\approx 0,80.$

Wynik: $P(15) \approx 14\%$; $\sigma \approx 2,74$; $P(12 \leq k \leq 18) \approx 80\%$.

Sprawdzenie: najbardziej prawdopodobny wynik zdarza się tylko w co siódmej grze ✓. Zakres 15 ± 3 , nieco szerszy niż $\pm\sigma$, obejmuje 80% gier, w zgodzie z odchyleniem $s = 2,8$ z jego tabeli (lekcja 0.5 Jednostki, analiza wymiarowa, szacowanie i niepewność pomiaru) ✓.

Gdy prób jest bardzo dużo, a sukces bardzo rzadki ($n \rightarrow \infty$, $p \rightarrow 0$ przy stałym $\lambda = np$: rozpady jąder w próbce, błędy w druku, telefony do centrali), rozkład dwumianowy przechodzi w prostszy. Wstawiamy $p = \lambda/n$: $\binom{n}{k} \approx \frac{n^k}{k!}$ dla $k \ll n$, a $p^k = \lambda^k/n^k$, a $(1 - \frac{\lambda}{n})^{n-k} \rightarrow e^{-\lambda}$ (czynniki $(1 - \lambda/n)^{-k} \rightarrow 1$) (lekcja 1.6 Szeregi liczbowe i potęgowe).

ROZKŁAD POISSONA

$$P(k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \mu = \sigma^2 = \lambda.$$

Liczba zdarzeń w ustalonym czasie, gdy zdarzenia są niezależne i zachodzą ze stałą średnią częstością. Odchylenie standardowe to $\sqrt{\lambda}$: przy 100 zliczeniach niepewność względna wynosi 10%, przy 10 000 już 1%. (por. Boas §15.9)

Przykład 3.2

Licznik Geigera i kopnięcia koni

Licznik rejestruje średnio 3 cząstki na sekundę. Jakie jest prawdopodobieństwo, że w danej sekundzie nie zarejestruje żadnej, dokładnie trzech i co najmniej pięciu?

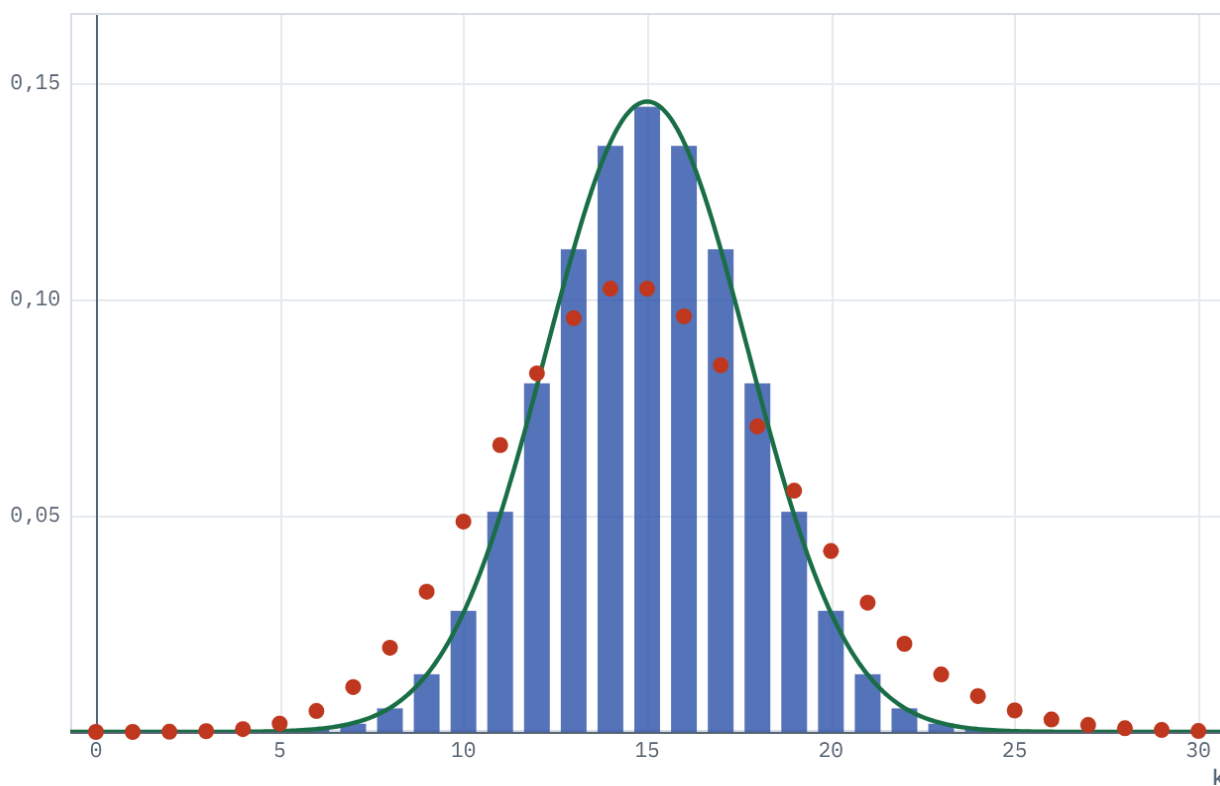
- 1 $P(0) = e^{-3} \approx 0,050$.
- 2 $P(3) = \frac{27}{6}e^{-3} \approx 0,224$.
- 3 $P(\geq 5) = 1 - \sum_{k=0}^4 P(k) \approx 1 - 0,815 = 0,185$.

Wynik: 5%; 22%; 18,5%.

Sprawdzenie historyczne: Bortkiewicz (1898) policzył żołnierzy armii pruskiej zabitych kopnięciem konia: 200 obserwacji (korpus $\times$ rok), średnio 0,61 ofiary. Poisson przewiduje 108,7; 66,3; 20,2; 4,1; 0,6 przypadków z 0, 1, 2, 3, 4 ofiarami, a obserwacje to 109, 65, 22, 3, 1 ✓.

SYMULACJA 1

Rozkład dwumianowy (słupki) w porównaniu z rozkładem Poissona o tej samej średniej (punkty) i rozkładem normalnym o tej samej średniej i wariancji (krzywa). Zmieniaj n i p : przy małym p i dużym n zgadza się Poisson, przy $np(1 - p)$ rzędu kilku i więcej Gauss.



— dwumianowy — Poisson ($\lambda = np$) — normalny ($\mu = np, \sigma^2 = np(1 - p)$)

$\mu = np$ 15,00

σ 2,739

max |dwumian – Poisson| 0,0420

max |dwumian – Gauss| 0,0012

Wariancja $np(1 - p)$ jest dość duża: słupki układają się w krzywą Gaussa (de Moivre–Laplace). Poisson jest za szeroki, bo jego wariancja to np , a nie $np(1 - p)$.

Stan na rysunku: $n = 30$, $p = 0,50$; $\mu = 15,00$, $\sigma = 2,739$; max różnica z Poissonem 0,0420, z Gaussem 0,0012.

ZADANIA · CZĘŚĆ 3

5 zadań · ok. 45 min

1 Oblicz prawdopodobieństwo dokładnie 5 orłów w 10 rzutach.



2 Rzucamy n razy monetą. Jak zmienia się z n prawdopodobieństwo dokładnie $n/2$ orłów, a jak odchylenie standardowe względnej liczby orłów k/n ?



- 3 W produkcji 2% sztuk jest wadliwych. W pudełku jest 100 sztuk. Oblicz prawdopodobieństwo, że wszystkie są dobre, rozkładem dwumianowym i Poissona.
- 4 Próbkę daje średnio 400 zliczeń na minutę. Jaka jest niepewność względna pojedynczego pomiaru minutowego? Ile minut trzeba mierzyć, żeby zejść do 1%?
- 5 Pokaż, że dla rozkładu Poissona $\sum P(k) = 1$, $E[k] = \lambda$ i $\text{Var } k = \lambda$.

podstawowe średnie feynmanowskie

4 Rozkład normalny i wzór Stirlinga

Słupki rozkładu dwumianowego dla dużego n układają się w gładką krzywą dzwonową. Żeby to pokazać, potrzebujemy przybliżenia silni, bo $n!$ rośnie szybciej niż każda wykładnicza.

WZÓR STIRLINGA

$$n! \approx \sqrt{2\pi n} \left(\frac{n}{e}\right)^n, \quad \text{błąd względny} \approx \frac{1}{12n}.$$

Skąd: $\ln n! = \sum_{k=1}^n \ln k \approx \int_1^n \ln x \, dx \approx n \ln n - n$, co daje $(n/e)^n$; czynnik $\sqrt{2\pi n}$ wychodzi z dokładniejszego rachunku, $n! = \int_0^\infty x^n e^{-x} \, dx$ przybliżone całką Gaussa wokół maksimum w $x = n$ (lekcja 3.3 Całki podwójne i potrójne).

Rozwijając $\ln P(k)$ rozkładu dwumianowego wokół maksimum $k = np$ do drugiego rzędu (wzór Taylora, lekcja 1.4 Rachunek różniczkowy w praktyce) i używając wzoru Stirlinga, dostaje się krzywą Gaussa: twierdzenie de Moivre'a-Laplace'a. Ten sam kształt pojawia się dużo ogólniej:

ROZKŁAD NORMALNY I CENTRALNE TWIERDZENIE GRANICZNE

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

Suma wielu niezależnych zmiennych o jednakowym rozkładzie i skończonej wariancji (ogólniej: gdy żaden składnik nie dominuje) ma rozkład bliski normalnemu, niezależnie od ich własnych rozkładów (centralne twierdzenie graniczne). Dlatego błędy pomiarów, które są sumą wielu drobnych przyczyn, mają często rozkład Gaussa. Reguła 68–95–99,7: w przedziałach $\mu \pm \sigma$, $\pm 2\sigma$, $\pm 3\sigma$ leży 68,3%, 95,4% i 99,7% prawdopodobieństwa. (por. Boas §15.8)

Przykład 4.1

Wzór Stirlinga w praktyce

Sprawdź wzór Stirlinga dla $n = 10$ i oszacuj $\binom{30}{15}$, a potem $P(15)$ z przykładu 3.1.

- 1 $\sqrt{20\pi} (10/e)^{10} \approx 3,599 \cdot 10^6$, a $10! = 3,629 \cdot 10^6$: błąd 0,83%, zgodnie z $\frac{1}{120}$.
- 2 Z wzoru Stirlinga $\binom{2m}{m} \approx \frac{4^m}{\sqrt{\pi m}}$, więc $P(m \text{ orłów z } 2m) \approx \frac{1}{\sqrt{\pi m}}$.
- 3 Dla $m = 15$: $\frac{1}{\sqrt{15\pi}} \approx 0,1457$.

Wynik: Stirling myli się o 0,8% dla 10!; $P(15) \approx 0,146$ wobec dokładnego 0,144.

Sprawdzenie: ta sama liczba wychodzi z gęstości normalnej: $\frac{1}{\sigma\sqrt{2\pi}}$ z $\sigma = \sqrt{7,5}$ daje 0,1457 ✓. Dwa przybliżenia, jeden wynik.

Przykład 4.2

Co najmniej 20 orłów

Oszacuj rozkładem normalnym prawdopodobieństwo co najmniej 20 orłów w 30 rzutach i porównaj z dokładnym.

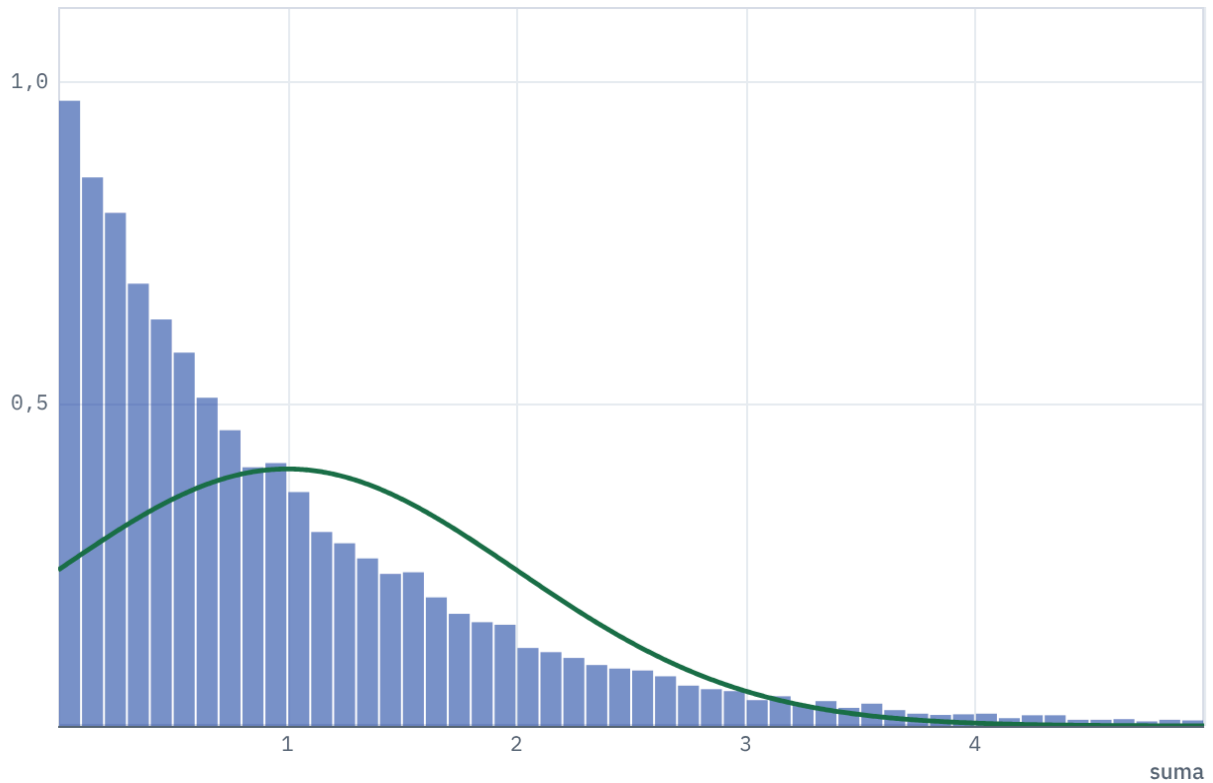
- 1 $\mu = 15$, $\sigma = \sqrt{7,5} \approx 2,739$.
- 2 Wartość dyskretna $k \geq 20$ odpowiada przedziałowi od 19,5 w górę (poprawka na ciągłość): $z = \frac{19,5-15}{2,739} \approx 1,64$.
- 3 Z tablic rozkładu normalnego $P(Z > 1,64) \approx 0,050$.

Wynik: $\approx 5,0\%$; dokładnie (suma dwumianowa) 4,94%.

Sprawdzenie: bez poprawki na ciągłość ($z = 1,83$) wyszłoby 3,4%, wyraźnie gorzej ✓. Wynik oddalony od średniej o ok. $1,6\sigma$ w jedną stronę zdarza się w co dwudziestą grze.

SYMULACJA 2

Centralne twierdzenie graniczne. Losujemy 20 000 razy sumę N niezależnych liczb z wybranego rozkładu (kostka, jednostajny, wykładniczy) i rysujemy histogram, a na nim krzywą Gaussa o tej samej średniej i wariancji. Dla kostki i rozkładu jednostajnego już kilka składników daje histogram bliski krzywej; dla wykładniczego potrzeba kilkudziesięciu, ale kształt dzwonu widać od razu.



— histogram sum (20 000 losowań) — krzywa Gaussa o tej samej średniej i wariancji

średnia (teoria / próba) 1,000 / 0,996

σ (teoria / próba) 1,000 / 0,985

skośność próby 1,923

Jeden składnik wykładniczy: rozkład bardzo niesymetryczny (skośność 2), w niczym nie przypomina Gaussa.

Stan na rysunku: wykładniczy ($\lambda = 1$), $N = 1$; średnia 0,996 (teoria 1,000), σ 0,985 (teoria 1,000), skośność 1,923.

ZADANIA · CZĘŚĆ 4

4 zadania · ok. 35 min

1 Wzrost dorosłych mężczyzn ma w przybliżeniu rozkład normalny ze średnią 178 cm i odchyleniem 7 cm. Jaki ułamek ma ponad 192 cm? A między 171 a 185 cm?

2 Sprawdź wzór Stirlinga dla $n = 5$ i $n = 100$. Jak błąd zależy od n ?

3 W 1000 rzutach monetą wypadło 530 orłów. Jakie jest prawdopodobieństwo co najmniej takiego nadmiaru przy uczciwej monecie? Czy to mocny argument, że moneta jest nieuczciwa?

4 Pokaż, że gęstość normalna jest unormowana i że jej średnia to μ , a wariancja σ^2 .
Skorzystaj z $\int_{-\infty}^{\infty} e^{-x^2} dx = \sqrt{\pi}$ (lekcja 3.3 Całki podwójne i potrójne).

▀▀▀ podstawowe ▀▀▀ średnie ▀▀▀ feynmanowskie

5 Błędne intuicje

× „Po pięciu orłach z rzędu reszka jest bardziej prawdopodobna.”

Rzuty są niezależne: moneta nie pamięta. Prawo wielkich liczb nie „wyrównuje” serii, tylko je rozcieńcza w coraz dłuższym ciągu.

× „Test o skuteczności 99% myli się w 1% przypadków, więc dodatni wynik oznacza 99% szans na chorobę.”

Liczy się też częstość choroby. Przy rzadkiej chorobie większość wyników dodatnich jest fałszywa (przykład 1.2).

× „Wartość oczekiwana to najbardziej prawdopodobny wynik.”

Wartość oczekiwana kostki to 3,5, a ten wynik nigdy nie wypada. Dla rozkładów niesymetrycznych średnia, mediana i wartość najczęstsza są różne.

× „Przy większej liczbie rzutów dokładna połowa orłów jest bardziej prawdopodobna.”

Mniej: $P(n/2) \sim 1/\sqrt{n}$. Bliżej połowy jest tylko ułamek orłów (zadanie 3.2).

× „Odchylenia standardowe dodają się.”

Dla niezależnych zmiennych dodają się wariancje; odchylenie sumy N składników rośnie jak $\sqrt{N}$.

Sprawdź się

- 1 Zapisz wzór Bayesa.

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

- 2 Jak obliczyć wariancję, znając $E[X]$ i $E[X^2]$?

$$\text{Var } X = E[X^2] - (E[X])^2.$$

- 3 Jak rośnie odchylenie standardowe sumy N niezależnych zmiennych?

$$\text{Jak } \sqrt{N}.$$

- 4 Podaj średnią i wariancję rozkładu dwumianowego.

$$np \text{ i } np(1-p).$$

- 5 Kiedy stosuje się rozkład Poissona i ile wynosi jego wariancja?

Dla liczby rzadkich, niezależnych zdarzeń w ustalonym czasie lub obszarze; wariancja równa średniej λ .

- 6 Zapisz wzór Stirlinga.

$$n! \approx \sqrt{2\pi n} (n/e)^n.$$

- 7 Co mówi centralne twierdzenie graniczne?

Suma wielu niezależnych zmiennych o jednakowym rozkładzie i skończonej wariancji ma w przybliżeniu rozkład normalny.

REGUŁY I MOMENTY

$$P(A \cup B) = P(A) + P(B) - P(A \cap B);$$

$$\text{niezależne: } P(A \cap B) = P(A)P(B)$$

$$P(A|B) = P(A \cap B)/P(B); \text{ Bayes:}$$

$$P(A|B) = P(B|A)P(A)/P(B)$$

$$\mu = \sum x_k p_k; \sigma^2 = E[X^2] - \mu^2; \text{Var}(X + Y) = \text{Var} X + \text{Var} Y \text{ (niezależne)}$$

$$\text{kostka: } 3, 5, \frac{35}{12}; \text{ jednostajny } [0; 1]: \frac{1}{2}, \frac{1}{12};$$

$$\text{wykładniczy: } \frac{1}{\lambda}, \frac{1}{\lambda^2}$$

ROZKŁADY

$$\text{dwumianowy: } \binom{n}{k} p^k (1-p)^{n-k}; np, np(1-p)$$

$$\text{Poissona: } \frac{\lambda^k}{k!} e^{-\lambda}; \mu = \sigma^2 = \lambda$$

$$\text{normalny: } \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}; 68-95-99,7\%$$

$$\text{Stirling: } n! \approx \sqrt{2\pi n} (n/e)^n; \binom{2m}{m} \approx 4^m / \sqrt{\pi m}$$

$$\text{dyskretny} \rightarrow \text{normalny: poprawka na ciągłość } \pm \frac{1}{2}$$

Do przeczytania u Feynmana

- **Obowiązkowo:** t. I, 6-1...6-4 (przypadek i prawdopodobieństwo, fluktuacje, błędzenie losowe, rozkład prawdopodobieństwa).
- **Dla chętnych:** 6-5 (zasada nieoznaczoności jako rozkład prawdopodobieństwa).

Do przeczytania w podręcznikach matematyki

- **M. L. Boas**, *Mathematical Methods in the Physical Sciences*, wyd. 3, rozdz. 15: prawdopodobieństwo, zmienne losowe, rozkłady dwumianowy, normalny i Poissona.
- **J. R. Taylor**, *An Introduction to Error Analysis*: rozdz. 5 (rozkład normalny), 10 (dwumianowy), 11 (Poissona). Przystępnie, z myślą o laboratorium.

W lekcji 5.2 Statystyka pomiarów i dopasowania użyjemy tych rozkładów do statystyki pomiarów: niepewności typu A i B, dopasowania prostej metodą najmniejszych kwadratów i testu χ^2 . W fizyce rozkład prędkości cząsteczek gazu (Maxwella) jest normalny w każdej składowej (Dział 8: Ciepło i teoria kinetyczna), a w mechanice kwantowej prawdopodobieństwo stanie się kwadratem modułu amplitudy (Dział 17: Amplitudy prawdopodobieństwa).



Odpowiedzi

Zadania · część 1

- Sumy 10, 11, 12 dają $3 + 2 + 1 = 6$ wyników z 36: $\frac{1}{6}$.
- Przez zdarzenie przeciwne: $1 - (5/6)^4 \approx 0,518$ i $1 - (35/36)^{24} \approx 0,491$. Ten problem w 1654 roku przedstawiono Pascalowi, a jego korespondencja z Fermatem zapoczątkowała rachunek prawdopodobieństwa.
- Wszystkie różne: $\frac{365}{365} \cdot \frac{364}{365} \cdot \dots \cdot \frac{343}{365} \approx 0,493$. Szukane: $\approx 0,507$. Par osób jest $\binom{23}{2} = 253$, dlatego wynik jest dużo większy, niż podpowiada intuicja.
- $\frac{0,99 \cdot 0,001}{0,99 \cdot 0,001 + 0,05 \cdot 0,999} \approx 0,019$, czyli 2%. Po drugim teście nowym „prawdopodobieństwem a priori” jest 0,019: $\frac{0,99 \cdot 0,019}{0,99 \cdot 0,019 + 0,05 \cdot 0,981} \approx 0,28$. Dopiero trzeci dodatni wynik daje ok. 89%.
- Warto: zmiana wygrywa z prawdopodobieństwem $\frac{2}{3}$. Twój pierwszy wybór trafia z prawdopodobieństwem $\frac{1}{3}$ i otwarcie pustych drzwi tego nie zmienia (prowadzący zawsze może je otworzyć). Z prawdopodobieństwem $\frac{2}{3}$ nagroda jest za jednym z dwóch pozostałych drzwi, a prowadzący wskazał, za którymi jej nie ma. Symulacja 200 000 gier daje 0,668.

Zadania · część 2

- $E[X] = p$, $E[X^2] = p$, $\text{Var } X = p - p^2 = p(1 - p)$. Maksimum $\frac{1}{4}$ dla $p = \frac{1}{2}$: najbardziej nieprzewidywalna jest uczciwa moneta.
- $E = 7$, $\text{Var} = 2 \cdot \frac{35}{12} = \frac{35}{6}$, $\sigma \approx 2,42$. Odchylenie rośnie $\sqrt{2}$ razy, nie 2 razy.
- $\mu = \frac{1}{2}$, $E[X^2] = \frac{1}{3}$, $\sigma^2 = \frac{1}{12}$, $\sigma \approx 0,289$. Błąd zaokrąglenia jest jednostajny na $[-\frac{\Delta}{2}; \frac{\Delta}{2}]$, więc $\sigma = \frac{\Delta}{\sqrt{12}} \approx 0,29\Delta$: tak liczy się niepewność typu B z rozdzielczości przyrządu (5.2 Statystyka pomiarów i dopasowania).
- Przez części: $E[T] = \frac{1}{\lambda}$, $E[T^2] = \frac{2}{\lambda^2}$, $\text{Var } T = \frac{1}{\lambda^2}$. $P(T > t) = e^{-\lambda t}$, więc $\frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t}$. Jądro „nie pamięta”, jak długo już przetrwało: nie starzeje się. Rozkład wykładniczy jest jedynym ciągłym rozkładem na $t \geq 0$ z tą własnością.

Zadania · część 3

- $\binom{10}{5} 2^{-10} = \frac{252}{1024} \approx 0,246$.
- $P(n/2) \approx \frac{1}{\sigma\sqrt{2\pi}} = \sqrt{\frac{2}{\pi n}}$ (przykład 4.1) maleje jak $\frac{1}{\sqrt{n}}$: dla 30 rzutów 0,14, dla 1000 ok. 0,025. Odchylenie k/n to $\frac{\sqrt{n/4}}{n} = \frac{1}{2\sqrt{n}}$, też maleje. Dokładnie połowa orłów jest coraz mniej prawdopodobna, a ułamek orłów coraz bliższy $\frac{1}{2}$: to nie sprzeczność.
- Dwumianowy: $0,98^{100} \approx 0,133$. Poisson z $\lambda = 2$: $e^{-2} \approx 0,135$. Dla co najwyżej dwóch wadliwych: 0,6767 i 0,6767, zgodność do czterech cyfr.
- $\sigma = \sqrt{400} = 20$, czyli 5%. Dla 1% potrzeba N zliczeń z $1/\sqrt{N} = 0,01$, czyli 10 000 zliczeń: 25 minut.
- $\sum e^{-\lambda} \frac{\lambda^k}{k!} = e^{-\lambda} e^{\lambda} = 1$. $E[k] = e^{-\lambda} \sum_{k \geq 1} \frac{\lambda^k}{(k-1)!} = \lambda e^{-\lambda} e^{\lambda} = \lambda$. $E[k(k-1)] = \lambda^2$ analogicznie, więc $E[k^2] = \lambda^2 + \lambda$ i $\text{Var} = \lambda$.

Zadania · część 4

1. $192 = \mu + 2\sigma$: $P \approx 2,3\%$. Między $\mu - \sigma$ a $\mu + \sigma$: 68% .
2. $n = 5$: 118,0 wobec 120, błąd 1,65%. $n = 100$: błąd 0,083%. Błąd maleje jak $\frac{1}{\sqrt{n}}$, choć sama silnia rośnie astronomicznie; dla $n = 100$ wynik ma 158 cyfr.
3. $\mu = 500$, $\sigma = \sqrt{250} \approx 15,8$. $z = \frac{529,5-500}{15,8} \approx 1,87$, $P \approx 3,1\%$ (dokładnie też 3,1%). Podejrzane, ale nie rozstrzygające: przy uczciwej monecie taki wynik zdarza się raz na ok. 32 serie. W lekcji 5.2 Statystyka pomiarów i dopasowania takie pytania zadamy systematycznie.
4. Z podstawieniem: $\int f dx = \frac{1}{\sqrt{\pi}} \int e^{-u^2} du = 1$. Średnia: $x - \mu$ daje funkcję nieparzystą w u , całka 0, więc $E = \mu$. Wariancja: $\frac{2\sigma^2}{\sqrt{\pi}} \int u^2 e^{-u^2} du = \frac{2\sigma^2}{\sqrt{\pi}} \cdot \frac{\sqrt{\pi}}{2} = \sigma^2$.

Matematyka · Prawdopodobieństwo i rozkłady. Źródła: R. P. Feynman, R. B. Leighton, M. Sands, *The Feynman Lectures on Physics*, t. I (Addison-Wesley 1963; New Millennium Edition 2011), 6-1...6-5; M. L. Boas, *Mathematical Methods in the Physical Sciences*, wyd. 3, rozdz. 15; J. R. Taylor, *An Introduction to Error Analysis*. Wszystkie liczby w przykładach policzone na nowo.