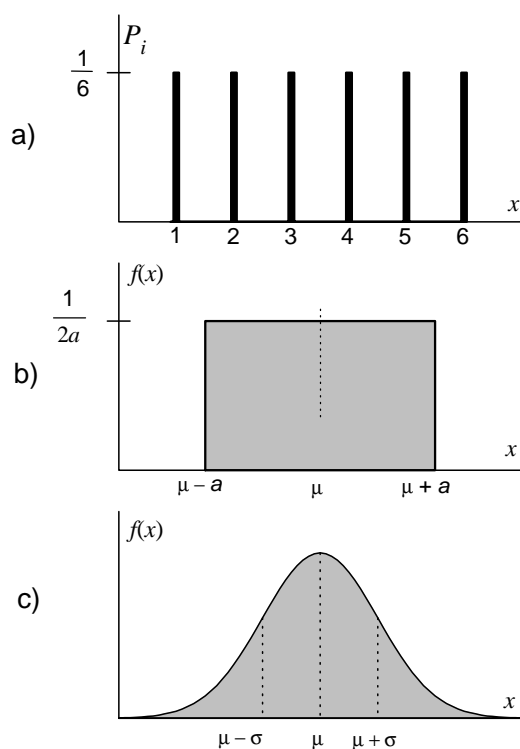


Dodatek A. Zmienna losowa

Pojęcie zmiennej losowej jest tak stare, jak sam rachunek prawdopodobieństwa. Ten dział matematyki został zapoczątkowany w XVII stuleciu, zaś impulsem do jego powstania było obliczanie prawdopodobieństwa wystąpienia różnych konfiguracji przy grze w kości¹.

A1. Pojęcie zmiennej losowej

Wynik rzutu kostką stanowi przykład *zmiennej losowej dyskretnej*. Liczby naturalne: 1, 2, 3, 4, 5 i 6 występują z jednakowym prawdopodobieństwem równym $1/6$ (rys. A1a).



Rys. A1. Funkcje rozkładu prawdopodobieństwa zmiennych losowych: a) dyskretna zmienna losowa (rezultaty rzutu kostką do gry); b) ciągła zmienna losowa o rozkładzie jednorodnym; c) ciągła zmienna losowa o rozkładzie Gaussa

Ważnym przykładem zmiennej losowej dyskretnej jest rozkład Poissona (Dodatek C).

¹ Więcej na ten temat w opracowaniu A. Lendy *Matematyczny groch ze statystyczną kapustą*, dostępnym na stronie WFiIS AGH.

Wartości *zmiennej losowej ciągłej* są liczbami rzeczywistymi. Ponieważ nawet najmniejszy przedział liczbowy o szerokości ε zawiera nieskończenie wiele liczb, więc ten typ zmiennej losowej określa się przez podanie *funkcji gęstości prawdopodobieństwa* $f(x)$, określonej jako stosunek prawdopodobieństwa znalezienia zmiennej losowej w przedziale $x, x + \varepsilon$ do szerokości przedziału ε w granicy ε dążącego do zera

$$f(x) = \lim_{\varepsilon \rightarrow 0} \frac{\text{prawdopodobieństwo, że } x \in (x, x + \varepsilon)}{\varepsilon}. \quad (\text{A1})$$

Z definicji (A1) wynika, że prawdopodobieństwo realizacji zmiennej losowej w przedziale $[a, b]$ jest dane całką

$$P(a \leq x \leq b) = \int_a^b f(x) dx. \quad (\text{A2})$$

Prawdopodobieństwo to w granicach $(-\infty, \infty)$ jest równe jedności, $\int_{-\infty}^{\infty} f(x) dx = 1$. Jest to tzw. warunek normalizacji funkcji gęstości prawdopodobieństwa.

Rysunek A1 pokazuje funkcje gęstości prawdopodobieństwa dla dwu najważniejszych typów zmiennej losowej o rozkładzie ciągłym. W przypadku *rozkładu jednostajnego* (nazywanego też *rozkładem prostokątnym*) funkcja $f(x)$ jest w określonym przedziale funkcją stałą (rys. A1b)

$$f(x) = \begin{cases} \frac{1}{2a} & |x - \mu| \leq a \\ 0 & |x - \mu| > a \end{cases}. \quad (\text{A3})$$

Rysunek A1b pokazuje, że μ oznacza środek rozkładu, a $2a$ – jego całkowitą szerokość.

Liczby losowe o rozkładzie jednorodnym z przedziału od 0 do 1 są potrzebne w wielu obliczeniach statystycznych. Tradycyjna metoda pozyskiwania tych liczb polegała na użyciu tablic liczb losowych. Obecnie do ich otrzymywania w wielkich ilościach służą *generatory liczb losowych* realizowane w komputerach i kalkulatorach.

Rozkład normalny zwany też *rozkładem Gaussa* (rys. A1c) definiuje funkcja

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[-\frac{(x - \mu)^2}{2\sigma^2} \right]. \quad (\text{A4})$$

Parametry μ oraz σ określają odpowiednio położenie środka i szerokość krzywej Gaussa. Teoretycznie wartość $f(x)$ jest niezerowa dla dowolnego x , ale w „ogonach” krzywej Gaussa maleje szybko do wartości bardzo małych (tab. A1).

Tabela A1
Wartości zestandaryzowanej funkcji Gaussa

$\frac{x - \mu}{\sigma}$	$f(x) \cdot \sigma$	$\frac{x - \mu}{\sigma}$	$f(x) \cdot \sigma$
0	0,399	2,00	0,054
0,25	0,387	2,25	0,032
0,50	0,352	2,50	0,018
0,75	0,301	2,75	0,009
1,00	0,242	3,00	0,0044
1,25	0,183	3,50	0,00087
1,50	0,130	4,00	0,00013
1,75	0,086	5,00	0,000015

A2. Wartość oczekiwana i odchylenie standardowe

Funkcja $f(x)$ jest różna od zera na ograniczonym obszarze zmiennej x , o ostrych (rozkład jednostajny) lub nieostrych (rozkład Gaussa) granicach. Dobrze jest znać liczbowe parametry określające środek i rozciągłość tego obszaru.

Wartość oczekiwana jest jedną z miar określających „środek” zmiennej losowej. W przypadku zmiennej losowej dyskretnej wartość μ definiuje suma

$$\mu = \sum_i x_i P_i . \quad (\text{A5a})$$

Wartość średnią zmiennej losowej ciągłej określa wzór całkowy

$$\mu = \int_{-\infty}^{+\infty} x f(x) dx . \quad (\text{A5b})$$

Dla rozkładów symetrycznych, takich jak rozkład jednostajny lub Gaussa, wartość oczekiwana μ pokrywa się ze środkiem symetrii² funkcji $f(x)$.

Odchylenie standardowe jest najpowszechniej używaną miarą rozrzutu zmiennej losowej wokół wartości średniej. W celu jej określenia (dla rozkładu ciągłego) definiujemy najpierw parametr zwany **wariancją**

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx, \quad (\text{A6a})$$

Odchylenie standardowe σ jest pierwiastkiem kwadratowym z wariancji,

$$\sigma = \sqrt{\sigma^2} . \quad (\text{A6b})$$

Przykładem zastosowania wzoru (A6b) jest obliczenie wariancji dla rozkładu jednostajnego

$$\sigma^2 = \int_{\mu-a}^{\mu+a} (x-\mu)^2 \frac{1}{2a} dx = \int_{-a}^a t^2 \frac{1}{2a} dt = \frac{a^2}{3} . \quad (\text{A7})$$

(Dla obliczenia całki użyliśmy podstawienia $t = x - \mu$.) Obliczając z uzyskanej wartości pierwiastek kwadratowy uzyskujemy odchylenie standardowe rozkładu jednostajnego $\sigma = a/\sqrt{3}$. Analogiczne obliczenie dla rozkładu Gaussa pokaże, że parametr σ we wzorze (A3) jest właśnie odchyleniem standardowym.

Ważną własnością funkcji $f(x)$ dla różnych rozkładów jest prawdopodobieństwo znalezienia wartości zmiennej losowej w przedziale³ $\mu \pm \sigma$. Dla rozkładu jednostajnego można je łatwo obliczyć

$$P(\mu - \sigma \leq x \leq \mu + \sigma) = \int_{-a/\sqrt{3}}^{a/\sqrt{3}} \frac{1}{2a} dx = \frac{1}{\sqrt{3}} \cong 0,58 .$$

² Środek symetrii funkcji $f(x)$ można zdefiniować jako wartość x_0 dla której $f(x_0 + \epsilon) = f(x_0 - \epsilon)$

³ Ten powszechnie stosowany skrótowy zapis oznacza przedział $[\mu - \sigma, \mu + \sigma]$.

Analogiczne obliczenie dla rozkładu Gaussa pokaże, że prawdopodobieństwo znalezienia liczby losowej o tym rozkładzie w przedziale $\mu \pm \sigma$ wynosi 0,68. Dla obydwu rozkładów w przybliżeniu co trzecia realizacja zmiennej losowej wyjdzie poza przedział $\mu \pm \sigma$. Dla zastosowań rozkładu Gaussa ważne jest ponadto prawdopodobieństwo jej wystąpienia w przedziałach $\mu \pm 2\sigma$ oraz $\mu \pm 3\sigma$ (tabela A2).

Tabela A2. Wartości całki z funkcji Gaussa

Przedział	Prawdopodobieństwo realizacji zmiennej losowej w przedziale	Przybliżone prawdopodobieństwo realizacji poza przedziałem
$\mu \pm \sigma$	0,683	1/3
$\mu \pm 2\sigma$	0,954	1/20
$\mu \pm 3\sigma$	0,9973	1/400

A3. Suma zmiennych losowych

Przez *sumę zmiennych losowych* rozumiemy nową zmienną losową y , której wartości uzyskuje się jako wynik dodawania liczb losowych u, v, w, \dots

$$y = u + v + w + \dots,$$

przy czym u, v, w są w ogólności realizacjami różnych zmiennych losowych, o różnych funkcjach rozkładu, wartościach oczekiwanych $\mu_u, \mu_v, \mu_w \dots$ i odchyleniach standardowych $\sigma_u, \sigma_v, \sigma_w \dots$. Interesuje nas odpowiedź na pytanie: jaka jest wartość oczekiwana, odchylenie standardowe i funkcja rozkładu dla zmiennej y ? Przedstawione poniżej twierdzenia obejmują ważny przypadek szczególny, gdy liczby u, v, w, \dots pochodzą z tego samego rozkładu – wartości μ oraz σ są wtedy identyczne.

Na podstawie definicji wartości oczekiwanej (A5b) łatwo wyprowadzić, że μ_y jest sumą algebraiczną wartości oczekiwanych składników,

$$\mu = \mu_u + \mu_v + \mu_w + \dots \quad (A8)$$

Wynik dla odchylenia standardowego jest mniej oczywisty. Wariancja (kwadrat odchylenia standardowego) sumy *nieskorelowanych* zmiennych losowych jest sumą wariancji składników

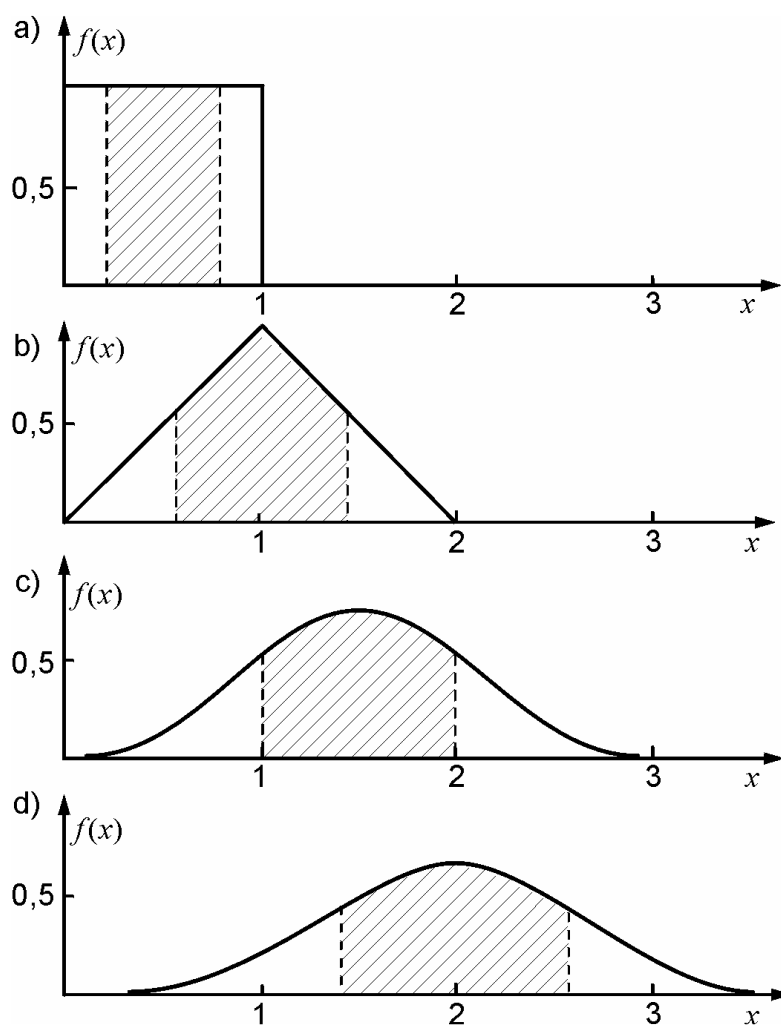
$$\sigma_y^2 = \sigma_u^2 + \sigma_v^2 + \sigma_w^2 + \dots \quad (A9)$$

Wzór (A9) jest jedną z przyczyn wyjątkowej roli odchylenia standardowego jako miary szerokości rozkładu. Definicja tego parametru (wzory (A6)) jest przecież nieoczywista, bardziej przemawiającą do wyobraźni miarą szerokości krzywej jest np. szerokość połówkowa 2Γ – odległość między punktami na zboczach krzywej $f(x)$, w których wartość funkcji maleje do połowy wartości maksymalnej. Ale dla szerokości połówkowej nie istnieje niezależny od postaci funkcji rozkładu wzór, który mógłby określić szerokość połówkową sumy zmiennych.

Najciekawszą z matematycznego punktu widzenia jest odpowiedź na pytanie, jaka jest *funkcja rozkładu* sumy zmiennych losowych. Jeżeli $f(x)$ i $g(x)$ są funkcjami gęstości prawdopodobieństwa dwu składników, to rozkład prawdopodobieństwa sumy określa operacja matematyczna zwana splotem lub konwolucją (symbol \otimes) i określona wzorem

$$f(x) \otimes g(x) = \int_{-\infty}^{\infty} f(t-x)g(t)dt. \quad (\text{A10})$$

Operację splatania funkcji można powtarzać, dzięki czemu obliczyć można funkcję rozkładu dla sumy dowolnej liczby składników. Rysunek A2 pokazuje, obliczone za pomocą wzoru (A10), funkcje gęstości prawdopodobieństwa dla sumy 2, 3 i 4 zmiennych losowych o rozkładzie jednostajnym w przedziale 0,1.



Rys. A2. Sumowanie zmiennych losowych: a) ciągła zmienna losowa o rozkładzie jednorodnym w przedziale (0, 1); b), c) i d) rozkłady gęstości prawdopodobieństwa dla sumy 2, 3 i 4 zmiennych losowych o ww. rozkładzie. Linie przerywane pokazują wartości $\mu - \sigma$ oraz $\mu + \sigma$ dla każdego rozkładu.

Zgodnie ze wzorem (A10) poszczególne krzywe są złożeniem kolejno: 2 odcinków prostej, 3 kawałków parabol zwykłych i 4 fragmentów parabol trzeciego stopnia. Ale te „skła-

danki” ze wzrostem liczby składników upodabniają się coraz bardziej do krzywej Gaussa. Nie jest to przypadek, lecz ilustracja jednego z najciekawszych twierdzeń statystyki matematycznej, tzw. **centralnego twierdzenia granicznego**. Zgodnie z tym twierdzeniem,

suma k zmiennych losowych ma w granicy $k \rightarrow \infty$ rozkład normalny (Gaussa) niezależnie od tego, jakie są rozkłady prawdopodobieństwa składników sumy.

Dodajmy, że twierdzenie jest prawdziwe przy założeniu, że wariancja dla każdego ze składników istnieje* i żaden składnik nie dominuje w sumie.

Jednym z zastosowań twierdzenia jest często praktykowany sposób generowania liczb o rozkładzie Gaussa, polegający na dodaniu do siebie kilkunastu liczb losowych o rozkładzie jednostajnym w przedziale (0,1). Centralne twierdzenie graniczne tłumaczy, dlaczego rozkład Gaussa jest często obserwowany w przyrodzie. Otóż w wielu przypadkach zjawisko losowe wynika z działania licznych przyczynków losowych. Jeżeli żaden z nich nie dominuje, suma ma rozkład Gaussa, niezależnie od (nieznanych w szczegółach) przyczyn przypadkowości danego procesu.

Dodatek B. Elementy teorii estymacji.

Doświadczalną informację o występujących w przyrodzie rozkładach prawdopodobieństwa uzyskujemy na podstawie znajomości zbioru n realizacji zmiennej losowej. Zbiór ten nazywamy **próbą losową**. Na tej podstawie staramy się obliczyć – czyli estymować – przybliżone wartości parametrów zmiennej (takich jak μ lub σ) a nawet określić w przybliżeniu funkcję $f(x)$. Własności, a zwłaszcza ograniczenia tej oceny podlegają prawidłowościom wynikającym z teorii prawdopodobieństwa.

B1. Estymowanie parametrów funkcji rozkładu

Prezentację własności funkcji losowych (Dodatek A) rozpoczęliśmy od przykładu rzutów kostką, którą można nazwać mechanicznym generatorem liczb losowych o dyskretnym rozkładzie jednostajnym. Wartość oczekiwana tej zmiennej (wzór (A5a)) wynosi

$$\mu = \sum_j P_j x_j = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = 3,5.$$

Przykład kostki do gry wykorzystać można również dla wprowadzenia pojęcia estymatora. W wyniku jedenastu rzutów uzyskano liczby 1, 3, 6, 5, 5, 2, 4, 4, 2, 1, 4 – jest to nasza próba losowa $\{x_i\}$ o liczebności $n = 11$. Jak z tych liczb obliczyć przybliżoną wartość parametru μ ?

* Przykładem funkcje rozkładu, dla której wariancja (i odchylenie standardowe) nie istnieje, jest funkcja Lorentza $f(x) = \Gamma^2 / (\Gamma^2 + x^2)$, co łatwo sprawdzić próbując obliczyć dla niej całkę (A6b).

Powszechnie używanym estymatorem wartości oczekiwanej jest *średnia arytmetyczna*

$$\bar{x} = \frac{1}{n} \sum x_i, \quad (\text{B2})$$

Dla podanych wyników rzutu kostką $\bar{x} = (1+3+6+5+5+2+4+4+2+1+4)/11 = 3,364$. Obliczenie średniej daje wynik zbliżony, ale nieidentyczny z wartością parametru $\mu = 3,5$. W zgodzie z terminologią *Przewodnika* rozróżniamy terminy:

estymator – algorytm służący do obliczenia przybliżonej wartości parametru, zwykle zdefiniowany wzorem algebraicznym (np. (B2)),

estymata – liczbową wartość estymatora dla rozpatrywanej próby losowej.

Wyraz estymata używa się zamiennie z terminem wartość estymatora. W sytuacjach, gdy znaczenie danej wielkości jako estymatora czy estymaty jest oczywiste, słowa te są pomijane.

Wróćmy jeszcze raz do przykładu z kostką do gry. Dla podanej próby losowej uzyskaliśmy $\bar{x} = 3,364$. Dla innej 11-elementowej próby losowej wyjdzie inna wartość, np. $\bar{x} = 3,875$, również różna od $\mu = 3\frac{1}{2}$. W ogólności: o ile parametr zmiennej losowej (np. μ) jest ustaloną liczbą, jego estymator (tu \bar{x}) fluktuuje, jest zatem *funkcją losową*.

Na przykładzie średniej omówimy własności estymatorów. Średnia jest estymatorem *zgodnym*; słowo to oznacza, że wartość estymatora dąży do wartości parametru x_0 w granicy $n \rightarrow \infty$. Dla skończonego n estymator, jako zmienną losową, charakteryzuje własna wartość średnia. Jeżeli pokrywa się ona z wartością parametru, estymator jest *nieobciążony*. Porównując dwa estymatory, estymator o mniejszej wariancji nazywamy bardziej *efektywnym*. Średnia arytmetyczna – jako estymator wartości oczekiwanej μ – jest dla każdej zmiennej losowej estymatorem zgodnym i nieobciążonym. Jeżeli zmienna x posiada rozkład normalny, średnia arytmetyczna jest ponadto estymatorem bardziej efektywnym od jakiegokolwiek innego.

Najpowszechniej używany **estymator wariancji** określony jest wzorem

$$s_x^2 = \frac{\sum (x_i - \bar{x})^2}{n-1}, \quad (\text{B3})$$

pierwiastek kwadratowy z (B3) definiuje **estymator odchylenia standardowego**, $s_x = \sqrt{s_x^2}$. Wprowadzamy symbol s_x by odróżnić estymator od samego odchylenia standardowego σ . Opisany wzorem (B3) estymator wariancji jest zgodny, nieobciążony i (dla rozkładu Gaussa) najbardziej efektywny. W mianowniku wzoru (B3) mamy $n-1$ właśnie dlatego, że bez odjęcia jedynki estymator byłby obciążony.

Estymator odchylenia standardowego średniej $s_{\bar{x}}$ jest \sqrt{n} razy mniejszy od estymatora s_x (wzór (1.7a) w rozdz. 1). Nietrudno udowodnić dlaczego tak jest, wykorzystując twierdzenia dotyczące sumy zmiennych losowych. Obliczanie średniej rozpoczyna się od sumowania n liczb o odchyleniu standardowym σ każda. Zgodnie wzorem (A9) wariancja sumy n liczb wynosi $n\sigma^2$, zatem odchylenie standardowe sumy jest równe $\sqrt{n}\sigma$. W celu obliczenia średniej sumę $\sum x_i$ dzielimy przez n . Przy tej operacji odchylenie standardowe

również zmniejsza się n razy, do wartości $\sigma_{\bar{x}} = \sigma/\sqrt{n}$. Ta sama relacja dotyczy estymatorów, $s_{\bar{x}} = s_x/\sqrt{n}$.

Względne odchylenie standardowe estymatorów $s_{\bar{x}}$ i s_x podaje tabela 1.1 w rozdziale 1. Z dobrym przybliżeniem wartości z tabeli dane są wzorem $1/\sqrt{2n-2}$.

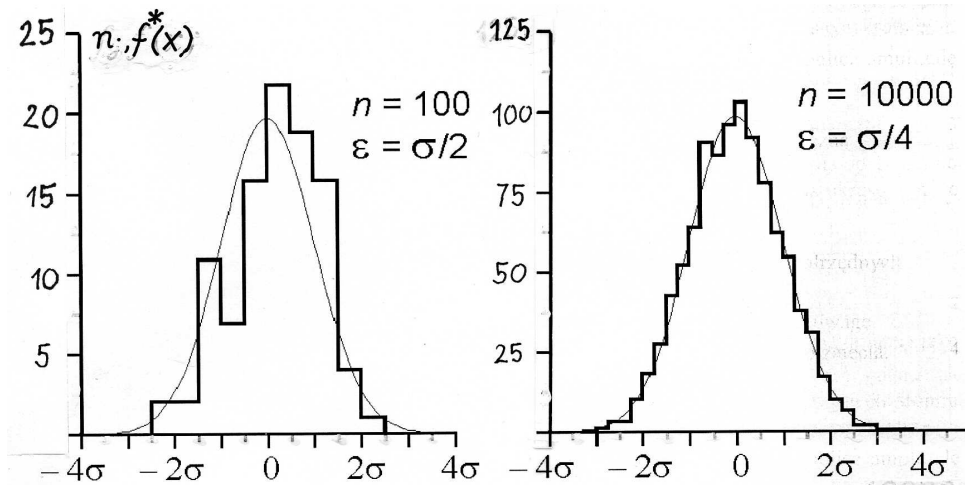
B2. Histogram funkcji rozkładu prawdopodobieństwa

Parametry rozkładu prawdopodobieństwa można obliczać nawet dla bardzo małych prób losowych. Dla zbadania funkcji rozkładu $f(x)$ nasza próba losowa musi być dość liczna, minimum to około setki elementów.

Przez **histogram doświadczalny** rozumiemy wykres słupkowy, gdzie na osi poziomej mamy badaną zmienną x podzieloną na równe przedziały o szerokości Δx , zaś wysokość słupków jest liczbą obserwacji n_j , jakie trafiły do kolejnego przedziału Δx .

a) fluktuacje liczby zliczeń w słupku histogramu

Rysunek B1 przedstawia uzyskane metodą symulacji komputerowej histogramy dla $n = 100$ oraz $n = 10000$ liczb losowych o rozkładzie Gaussa o parametrach $\mu = 0, \sigma = 1$. Charakterystyczną cechą pokazanych histogramów są fluktuacje wysokości słupka wynikające ze statystycznych fluktuacji liczb n_j .



Rys. B1. Przykładowe histogramy rozkładu Gaussa dla $n = 100$ oraz $n = 10\,000$ [prowizoryczny]

Wartości n_j podlegają dyskretnemu rozkładowi prawdopodobieństwa, tzw. rozkładowi Poissona (Dodatek C), dla którego odchylenie standardowe jest równe $\sqrt{n_j}$. Zatem, np. dla $n_j = 16$ fluktuacje są rzędu $\sqrt{16} = 4$, a więc bardzo duże. Fluktuacje liczby zliczeń są źródłem jakościowych anomalii, np. na lewym zboczu obydwu histogramów mamy niemonotoniczne zmniejszanie się wysokości słupków. Dziwić się temu nie należy, dla małej liczebności próby histogram pozbawiony takich czy innych anomalii jest wyjątkiem. Proces „wygładzania”

histogramu ze wzrostem n jest bardzo powolny, w celu k -krotnego zmniejszenia średniego pionowego i poziomego rozmiaru „schodka” trzeba k^3 -krotnie zwiększyć liczebność próby!

b) optymalna szerokość Δx słupka histogramu

Szerokość słupka jest najczęściej przedmiotem subiektywnego wyboru, przy czym niedoświadczeni często stosują zbyt małą wartość Δx , w wyniku czego histogram jest zdominowany przez statystyczne fluktuacje liczby zliczeń. Istnieje szereg wzorów na optymalną szerokość histogramu. Polecić można wzór Heada⁴ $\Delta x \approx s(\sqrt{8\pi}/n)^{1/5}$ jako dedykowany do problemu porównania histogramu z krzywą Gaussa. Do wykonania histogramu można brać wartość zaokrągloną Δx , bliską uzyskanej z ww. wzoru.

c) porównanie histogramu z krzywą teoretyczną

Histogram doświadczalny można porównać z *przeskalowaną krzywą Gaussa* opisaną równaniem

$$f^*(x) = n \Delta x \cdot \frac{1}{s\sqrt{2\pi}} \exp\left[-\frac{(x - \bar{x})^2}{2s^2}\right]. \quad (\text{B4})$$

Prawa część wzoru to nic innego jak teoretyczna funkcja gęstości prawdopodobieństwa dla rozkładu normalnego (wzór (A4)), w której nieznanne parametry μ oraz σ zostały zastąpione przez stosowne estymatory: wartość średnią \bar{x} oraz estymator odchylenia standardowego s , obliczone ze zmierzonych wartości x_i . W celu „dopasowania” do histogramu doświadczalnego krzywa teoretyczna jest pomnożona przez całkowitą liczbę pomiarów n oraz szerokość przedziału Δx . Takie przeskalowanie zapewnia, że powierzchnie pod linią histogramu i pod krzywą $f^*(x)$ są takie same.

Jakościowe porównanie histogramu doświadczalnego z krzywą teoretyczną może polegać na próbie odpowiedzi na pytania takie jak:

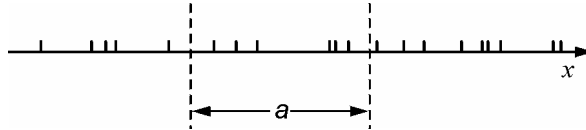
- czy zmierzony histogram wygląda na podobny do rozkładu normalnego?
- czy może być uznany za symetryczny?
- czy występują punkty odstające?
- co można powiedzieć o „ogonach” histogramu?

Półilościowa analiza „ogonów” eksperymentalnego rozkładu prawdopodobieństwa może polegać na określeniu liczby pomiarów, jakie nie mieszczą się w przedziale $(\bar{x} - 2s, \bar{x} + 2s)$. Wartość teoretyczna dla rozkładu normalnego wynosi około 5% (tabela A2), natomiast w przypadku rozkładu jednostajnego w przedziale $(\bar{x} - 2s, \bar{x} + 2s)$ winny się mieścić wszystkie pomiary.

⁴ Heald M.A.: *On choosing the bin width of a Gaussian histogram*. Am. J. Phys., **52**, 254 (1984). Inne wzory: patrz hasło „histogram” w ang. wersji Wikipedii.

Dodatek C. Rozkład Poissona

Rozpatrzmy następujące zagadnienie z rachunku prawdopodobieństwa. Na osi liczbowej „rozrzucamy” w sposób przypadkowy punkty (rys. C1). Prawdopodobieństwo, że na nieskończenie małym odcinku dx znajdziemy punkt, wynosi qdx , gdzie q jest stałą. Interesuje nas, ile punktów znajdziemy na odcinku osi liczbowej o skończonej długości a .

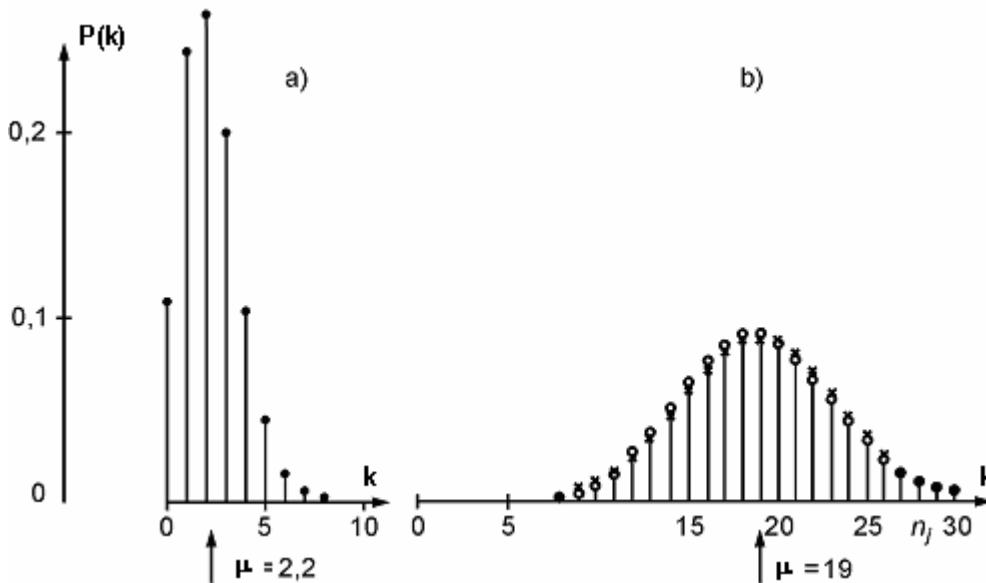


Rys. C1. Przypadkowe rozrzucenie punktów na osi liczbowej

Wartość oczekiwana dla liczby punktów, jakie znajdziemy na odcinku wynosi $\mu = qa$ i jest w ogólności liczbą *rzeczywistą*. Liczba punktów, jaką znajdziemy na odcinku przy kolejnym losowaniu jest liczbą *całkowitą* k , której wartości nie można przewidzieć. Można natomiast obliczyć prawdopodobieństwo uzyskania różnych wartości k . Wynosi ono

$$P(k) = \frac{\mu^k}{k!} \exp(-\mu). \quad (\text{C1})$$

Wzór (C1) określa rozkład Poissona. Wartości k zmiennej losowej Poissona są liczbami całkowitymi, dlatego rozkład ten jest rozkładem *dyskretnym*. Rysunek C2 przedstawia wykresy rozkładu Poissona dla dwóch wartości oczekiwanych, małej ($\mu = 2,2$) i większej ($\mu = 19$).



Rys. C2. Wykresy $P(k)$ dla rozkładów dyskretnych: a) Poissona dla $\mu = 2,2$ (kropki), b) Poissona dla $\mu = 19$ (kółka) i aproksymującego go dyskretnego rozkładu Gaussa (krzyżyki)

Parametr μ rozkładu Poissona jest jednocześnie wartością oczekiwaną i wariancją tego rozkładu. Zatem wartość odchylenie standardowe wynosi

$$\sigma = \sqrt{\mu}. \quad (\text{C2})$$

Najważniejszą konsekwencją wzoru (C2) jest łatwość oceny niepewności dla zmiennej podlegającej rozkładowi Poissona. Wystarczy obliczyć pierwiastek z liczby zliczeń!

Wzór (C1) definiujący rozkład Poissona jest słuszny dla dowolnych μ i k . Dla dużych wartości μ posługiwanie się nim jest utrudnione, gdyż funkcje μ^k i $k!$ gwałtownie rosną⁵ ze wzrostem k . Na szczęście ze wzrostem wartości oczekiwanej μ rozkład Poissona szybko upodabnia się do *dyskretnego rozkładu Gaussa* zdefiniowanego wzorem

$$P(k) = \frac{1}{\sqrt{2\pi\mu}} \exp\left[-\frac{(k-\mu)^2}{2\mu}\right]. \quad (\text{C3})$$

Wzór (C3) różni się od definicji (A4) tym, że zmienna k jest liczbą całkowitą, a za wartość odchylenia standardowego kładziemy $\sigma = \sqrt{\mu}$. Upodobnienie się rozkładu Poissona do rozkładu Gaussa dla dużych μ jest przykładem działania centralnego twierdzenia granicznego.

Głównym zastosowaniem rozkładu Poissona w fizyce jest opis statystycznych fluktuacji liczby impulsów z detektorów promieniowania. W „życiu codziennym” rozkład ten znajduje przybliżone zastosowanie wszędzie tam, gdzie interesuje nas *liczba zdarzeń przypadkowych* w określonej dużej populacji. Na przykład roczna liczba wypadków drogowych w Krakowie czy liczba uzyskanych tytułów profesora wśród pracowników AGH. Występowanie fluktuacji statystycznych rzędu pierwiastka kwadratowego z liczby zdarzeń utrudnia wnioskowanie o systematycznych zmianach wartości średniej, szczególnie wtedy, gdy liczba zdarzeń jest mała.

⁵ Sprawdź, dla jakiej maksymalnej liczby k używany przez Ciebie kalkulator lub program komputerowy potrafi obliczyć silnię. Przepelnienie pamięci nastąpi dla k mniejszego niż 100.