

» Idź do

- Spis treści
- Przykładowy rozdział

» Katalog książek

- Katalog online
- Zamów drukowany katalog

» Twój koszyk

- Dodaj do koszyka

» Cennik i informacje

- Zamów informacje o nowościach
- Zamów cennik

» Czytelnia

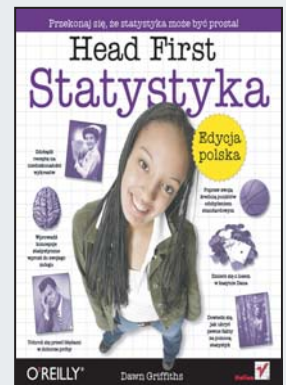
- Fragmenty książek online

» Kontakt

Helion SA
ul. Kościuszki 1c
44-100 Gliwice
tel. 032 230 98 63
e-mail: helion@helion.pl
© Helion 1991-2008

Head First. Statystyka. Edycja polska

Autor: Dawn Griffiths
Tłumaczenie: Przemysław Janicki
ISBN: 978-83-246-2065-4
Tytuł oryginału: [Head First Statistics](#)
Format: 200×230, stron: 452



Przekonaj się, że statystyka może być prosta!

Jeśli statystyka wydaje Ci się nudna i skomplikowana, to tylko dlatego, że nie korzystałeś dotąd z odpowiedniego podręcznika! Ta innowacyjna, przyjazna dla czytelnika książka z pewnością zmieni Twoją opinię. Wiedzę z zakresu statystyki podano tu w sposób uwzględniający najnowsze badania w zakresie efektywnego nauczania. Dzięki przyjęciu takiej formuły tekst dostosowano do sposobu funkcjonowania Twojego mózgu, aby w pełni wykorzystać jego możliwości i bezboleśnie wprowadzić Cię w świat skomplikowanych obliczeń.

Najważniejsze zagadnienia zostały tu zilustrowane za pomocą – nierzadko zabawnych – przykładów z życia codziennego, takich jak analiza statystyk sportowych, wyników gier hazardowych czy testów nowych leków. Dzięki tej książce dowiesz się m.in., jak wybrać optymalny wykres do wizualizacji określonych danych, szybko wskazać wartości reprezentatywne dla danego zbioru danych i za pomocą rachunku prawdopodobieństwa przewidywać skutki powtarzalnych zdarzeń w długich seriach. Z łatwością nie tylko przyswoisz zawartą tu wiedzę, ale i wykorzystasz ją w codziennym życiu!

- Wizualizacja danych
- Wykresy
- Praca z danymi zgrupowanymi
- Rozkład geometryczny, dwumianowy i Poissona
- Miary zróżnicowania
- Szacowanie parametrów populacji na podstawie próby
- Kwartyle i wariancje
- Prawdopodobieństwo zdarzeń
- Konstruowanie przedziału ufności
- Podstawy kombinatoryki
- Weryfikacja hipotez
- Korelacja i regresja

Z dobrym podręcznikiem nawet statystyka może być łatwa i ciekawa!

Spis treści (skrótowy)

	Wprowadzenie	27
1	Wizualizacja danych: <i>Pierwsze wrażenie</i>	39
2	Miary tendencji centralnej: <i>Droga Środka</i>	83
3	Miary zróżnicowania: <i>Potęga zmienności</i>	121
4	Prawdopodobieństwo zdarzeń: <i>Natura ryzyka</i>	165
5	Dyskretne rozkłady prawdopodobieństwa: <i>Zarządzamy oczekiwaniami</i>	235
6	Podstawy kombinatoryki: <i>Porządkujemy obiekty</i>	279
7	Poznajemy rozkłady: geometryczny, dwumianowy i Poissona: <i>Stajemy się dyskretni</i>	307
8	Poznajemy rozkład normalny: <i>Wybieramy normalność</i>	363
9	Poznajemy rozkład normalny (cd.): <i>Więcej niż normalność</i>	399
10	Przeprowadzamy losowanie: <i>Pobieramy próbkę</i>	453
11	Szacujemy parametry populacji na podstawie próby: <i>Dokonujemy ocen</i>	479
12	Konstruujemy przedziały ufności: <i>Wyrażamy przekonania</i>	525
13	Weryfikacja hipotez: <i>Oceniamy fakty</i>	559
14	Rozkład χ^2 : <i>Gdy sprawy idą nie po naszej myśli</i>	605
15	Korelacja i regresja: <i>Co z moją linią?</i>	643
A	Dodatek uzupełniający: <i>Dziesięć najważniejszych rzeczy, które pominęliśmy</i>	681
B	Tablice statystyczne: <i>Czasem trzeba coś sprawdzić</i>	695

Spis treści (z prawdziwego zdarzenia)



Wprowadzenie

Twój mózg a statystyka. Czytasz tę książkę, ponieważ chcesz się czegoś nauczyć. W tym czasie Twój mózg będzie Ci wyświadczał przysługę, dbając o to, byś się za bardzo nie przemęczał. Będzie Ci podpowiadał: „Zajmij się lepiej ważniejszymi sprawami, na przykład tym, jakich dzikich zwierząt trzeba się wystrzegać albo czy jeżdżenie na snowboardzie nago to aby na pewno dobry pomysł”. Jak w tej sytuacji mógłbyś przekonać swój mózg, że Twoje życie zależy od znajomości statystyki?

	Dla kogo przeznaczona jest ta książka?	28
	Wiemy, co sobie przed chwilą pomyślałeś	29
	Metapoznanie — myślenie o myśleniu	31
	Oto, co TY możesz zrobić, by pobudzić swój mózg	33
	Przeczytaj to	34
	Recenzenci merytoryczni	36
	Podziękowania	37

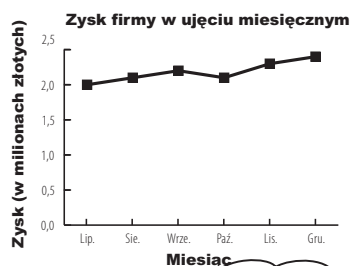
Wizualizacja danych

1

Pierwsze wrażenie

Czy masz problem ze zrozumieniem faktów zapisanych w danych?

Statystyki opisowe pomogą Ci zrozumieć znaczenie złożonych zbiorów danych. Dzięki nim **bardzo złożone kwestie** staną się **zupełnie proste**. Gdy już je zrozumiesz i będziesz chciał **podzielić się tą wiedzą z innymi**, z pewnością zainteresuje Cię potencjał licznych technik **wizualizacji danych**, jakie oferuje statystyka. Jeśli chcesz więc dobrać optymalny wykres dla swoich danych, weź w rękę swój płaszcz, spakuj ulubiony suwak logarytmiczny i ruszaj z nami po przygodę do Statsville.



Teraz widzisz, że miałam rację: zyski firmy są co miesiąc takie same.

Nieprawda, są wręcz zdumiewające. Spójrz, jak szybką w górę!



Statystyki są wszędzie	40
Co Ci dadzą statystyki?	41
Jak to z wykresami było	42
Prosty, lecz bardzo użyteczny wykres kołowy	46
Wykres słupkowy jest bardziej precyzyjny	48
Wykres kolumnowy	48
Wykres wierszowy	49
Wszystko jest kwestią odpowiedniej skali	50
Wykorzystanie skali bezwzględnej	51
Praca z bardziej złożonymi zbiorami danych	52
Kategorie a liczby	56
Praca z danymi zgrupowanymi	57
Konstruujemy histogram	58
Krok 1: Określ szerokość słupków histogramu	64
Krok 2: Dostosuj wysokość słupków histogramu	65
Krok 3: Wykreśl swój histogram	66
Czym są częstości skumulowane	72
Jak wykreślić częstości skumulowane	73
Jak wybrać odpowiedni typ wykresu	77

Miary tendencji centralnej

2 Droga Środka

Niekiedy trzeba po prostu dotrzeć do sedna sprawy. Czasami trudno ogarnąć ogrom informacji ukrytych w gąszczu danych. Pierwszym krokiem może być wówczas wyznaczenie średnich. Statystycy nazywają je miarami tendencji centralnej. Dzięki nim potrafią szybko wskazać wartości reprezentatywne dla danego zbioru danych i na tej podstawie wyciągnąć ważne wnioski. W tym rozdziale nauczysz się wyznaczać wartości kilku najważniejszych i najpopularniejszych statystyk — średniej, mediany i dominanty. Zobaczysz, jak łatwo i efektywnie można za ich pomocą dokonać **opisu danych**.



20 lat



20 lat



21 lat



20 lat



19 lat



Witamy w ośrodku odnowy	84
Najpopularniejszą średnią jest średnia arytmetyczna	85
W świecie symboli	86
Jak sobie radzić z niewiadomymi	87
Wracamy do średniej	88
Wróćmy do naszego klubu	91
Każdy ćwiczył kiedyś kung-fu	92
W naszych danych są wartości nietypowe	95
Czym są dane asymetryczne	96
Rozmowa przy dystrybutorze	98
Z pomocą przychodzi nam mediana	99
Jak znaleźć medianę w trzech prostych krokach	100
Nasz interes kwitnie	103
Nauka pływania dla... najmłodszych	104
Dlaczego średnia i mediana nie są miarodajne?	107
Jak możemy sobie radzić z danymi tego typu?	107
Cała prawda o średniej arytmetycznej	109
Przedstawiamy dominantę (modę)	111
Jak znaleźć dominantę w trzech prostych krokach	112

Miary zróżnicowania

3 Potęga zmienności

Nie wszystkiemu można wierzyć, ale jak się o tym przekonać?

Średnie pozwalają nam poznać typową wartość dla naszych danych, ale **nie mówią nam wszystkiego**. Umiemy już znajdować wartości centralne zbioru danych, ale średnia arytmetyczna, mediana czy dominanta nie zawsze wystarczają do wyciągnięcia głębszych wniosków. W tym rozdziale poszerzymy naszą wiedzę o narzędzia, dzięki którym będziemy mogli coś powiedzieć o **zróżnicowaniu** naszych danych.



Wszyscy trzej mają tę samą średnią punktów, więc potrzebuję innej wskazówki, która pomoże mi wybrać najlepszego. Czy masz jakiś pomysł?



W poszukiwaniu zawodnika	122
Musimy porównać wyniki kandydatów	123
O czym mówi rozstęp	124
Obserwacje nietypowe rodzą pewien problem	127
Musimy znaleźć sposób na pozbycie się obserwacji nietypowych	129
Na ratunek spieszą kwartyly	130
Rozstęp międzykwartyłowy wyklucza obserwacje nietypowe	131
Anatomia kwartyli	132
Nie musimy ograniczać się tylko do kwartyli	136
Czym są percentyle?	137
Wykres pudełkowy dobrze prezentuje rozproszenie danych	138
Zmienność to coś więcej niż tylko rozstęp	142
Jak obliczyć odchylenie od średniej	143
Zmienność możemy zmierzyć za pomocą wariancji...	144
...ale odchylenie standardowe jest miarą bardziej intuicyjną	145
Cała prawda o odchyleniu standardowym	146
Szybszy sposób na wariancję	151
A gdybyśmy potrzebowali punktu odniesienia dla porównań?	156
Standaryzacja danych sposobem na ich porównywanie	157
Jak interpretować dane wystandaryzowane	158
Nasza drużyna mistrzem!	163

Prawdopodobieństwo zdarzeń

4

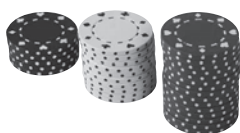
Natura ryzyka

Życie pełne jest niepewności. Czasami trudno jest nawet przewidzieć, co wydarzy się w ciągu najbliższych paru minut. Szanse zajścia pewnych zdarzeń są jednak większe niż innych, czego uczy nas **rachunek prawdopodobieństwa**. Szacowanie prawdopodobieństwa zdarzeń ułatwia **przewidywanie przyszłości**, ponieważ pozwala ocenić, jak duże są szanse ich wystąpienia. A to pozwala podejmować bardziej **świadome wybory**. W tym rozdziale dowiesz się, czym jest prawdopodobieństwo zdarzeń i jak może Ci ono pomóc zapanować nad przyszłością!



	00	3																				
	0	2	6																			
	1	4	5	8																		
	10	11	14	15	17	18																
	16	19	20	21	23	24																
	22	25	26	27	28	29	30															
	31	32	33	34	35	36																
	I TUZIN		II TUZIN				III TUZIN															
	1 – 18		PARZYSTE		♦		♠		NIEPARZYSTE				19 – 36									

Wielki Szlem	166
Wejdz do gry!	167
Jakie są moje szanse?	170
Znajdujemy prawdopodobieństwo wygranej w ruletkę	173
Do czego przydają się diagramy Venna	174
Możesz także dodać prawdopodobieństwa	180
Zdarzenia rozłączne	185
Gdy część wspólna sprawia problem	186
Trochę notacji	187
Znowu nieudany obrót...	193
Prawdopodobieństwo warunkowe	194
Obliczamy prawdopodobieństwa warunkowe	195
Prawdopodobieństwa warunkowe można przedstawić na drzewie stochastycznym	196
Drzewa są pomocne w obliczaniu prawdopodobieństw	197
Krok 1: Znajdujemy $P(\text{czarne} \cap \text{parzyste})$	205
Krok 2: Znajdujemy $P(\text{parzyste})$	207
Krok 3: Znajdujemy $P(\text{czarne} \text{parzyste})$	208
Wykorzystaj prawdopodobieństwo całkowite, by znaleźć $P(B)$	210
Twierdzenie Bayesa	211
Gdy zdarzenia wpływają na siebie, są zdarzeniami zależnymi	219
Jeśli zdarzenia nie wpływają na siebie, są niezależne	220
Kilka słów o liczeniu prawdopodobieństw dla zdarzeń niezależnych	221



Dyskretne rozkłady prawdopodobieństwa

5

Zarządzamy oczekiwaniami

Zdarzenia mało prawdopodobne czasem się zdarzają, ale jakie są ich konsekwencje?

W poprzednim rozdziale przekonaliśmy się, jak rachunek prawdopodobieństwa może pomóc nam ocenić szanse zajścia pewnych zdarzeń. Jednak nie powie nam on nic na temat **wplywu** tych zdarzeń na nasze życie. Choć przy stole ruletki czasem pada spora wygrana, to jednak — czy jest ona warta tych wszystkich pieniędzy, jakie przy okazji można stracić? W tym rozdziale pokażemy Ci, jak można posłużyć się rachunkiem prawdopodobieństwa do **przewidywania skutków powtarzalnych zdarzeń w długich seriach**, a także jak można **ocenić dokładność** takich prognoz.



Wracamy do kasyna Dana	236
Tworzymy rozkład prawdopodobieństwa wygranej na automacie	239
Wartość oczekiwana pozwala przewidzieć wynik...	242
...a wariancja mówi o tym, jak bardzo jest on zmienny	243
Wariancja a rozkład prawdopodobieństwa	244
Obliczamy wariancję dla naszego przykładu	245
Gdy ceny idą w górę	250
Między $E(X)$ i $E(Y)$ istnieje związek liniowy	255
Podsumujmy nasze rozważania	256
Ogólne wzory na przekształcenia liniowe	257
Każde pociągnięcie dźwigni jest niezależnym zdarzeniem	260
Przydatne skróty	261
Nowe automaty wchodzą do gry!	267
Dodaj $E(X)$ do $E(Y)$, by uzyskać $E(X+Y)$...	268
...lub odejmij $E(Y)$ od $E(X)$, by uzyskać $E(X-Y)$	269
Podobne operacje możesz wykonywać na zmiennych przekształcanych liniowo	270
Rozbiłeś bank!	276

Podstawy kombinatoryki

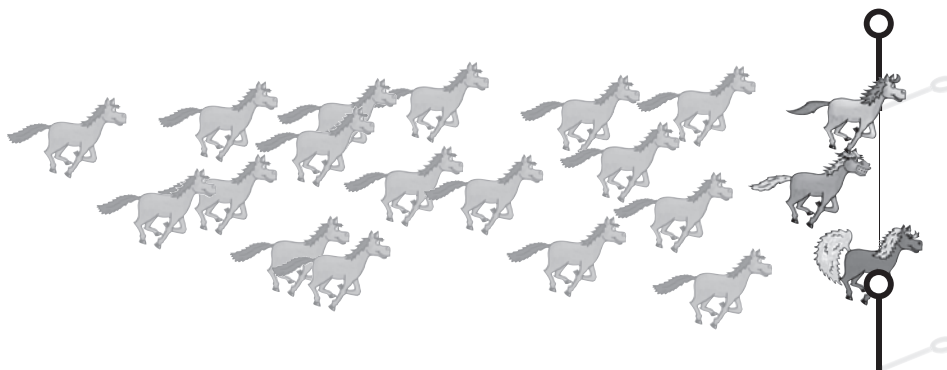
6

Porządkujemy obiekty

Czasami kolejność ma znaczenie. Policzenie **wszystkich możliwych sposobów** grupowania czy porządkowania pewnego zbioru obiektów może być niezwykle pracochłonne. Często jednak nie mamy wyjścia, bo takie informacje są **kluczowe** dla rachunku prawdopodobieństwa. W tym rozdziale poznamy **szybki i efektywny** sposób na zdobycie tego rodzaju informacji, który nie wymaga od nas znajomości wszystkich możliwych wyników doświadczenia losowego. Zostań więc z nami, a przekonasz się, jak łatwo można **zliczać wszystkie możliwości grupowania i porządkowania obiektów.**



Derby Statsville	280
Wyścig trójki koni	281
Na ile sposobów konie mogą przekroczyć linię mety?	283
Zliczamy możliwe ustawienia zwycięzców	284
Ustawiamy obiekty w okrąg	285
Czas na wyścig nowicjuszy	289
Porządkowanie klas to coś innego niż porządkowanie ich elementów	290
Porządkujemy zwierzęta według gatunku	291
Ogólna formuła na liczbę uporządkowań w przypadku powtórzeń	292
Czas na wyścig dwudziestu koni	295
Na ile sposobów możemy zapełnić trzy miejsca medalowe?	296
Obliczamy wariacje	297
Gdy kolejność nie ma znaczenia	298
Liczmy kombinacje	299
Cała prawda o kombinacjach	300
To już koniec zawodów	306

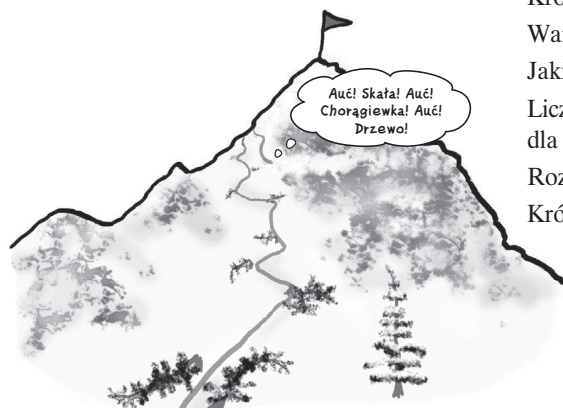


Poznajemy rozkłady: geometryczny, dwumianowy i Poissona

7 Stajemy się dyskretni

Wyznaczanie rozkładów prawdopodobieństwa zabiera sporo czasu.

Wiemy już, jak wyznaczać rozkłady prawdopodobieństwa, choć nie zawsze jest to łatwe. Pewnie więc zgodzisz się z nami, że byłoby dobrze, gdyby istniały ogólne, bardziej uniwersalne rozkłady prawdopodobieństwa, którymi **łatwo** można by się posługiwać na co dzień. W tym rozdziale zaprezentujemy kilka **standardowych rozkładów prawdopodobieństwa**, o ściśle określonych charakterystykach. Gdy zrozumiesz, jak to działa, będziesz w stanie **w rekordowo szybkim czasie obliczać prawdopodobieństwa, wartości oczekiwane i wariancje**. Przygotuj się więc na poznanie rozkładów: geometrycznego, dwumianowego i Poissona. Miłej lektury.



Znajdujemy rozkład prawdopodobieństwa dla osiągnięć Chada	311
Istnieje rozkład prawdopodobieństwa, który dobrze opisuje nasz problem	312
Prawdopodobieństwo możemy przedstawić za pomocą wzoru	315
Rozkład geometryczny pozwala operować także na nierównościach	317
Wartość oczekiwana dla rozkładu geometrycznego	318
Wartość oczekiwana wynosi $1/p$	319
Wariancja dla rozkładu geometrycznego	321
Krótki przewodnik po rozkładzie geometrycznym	322
Właśnie poznałeś rozkład geometryczny	325
Arcyfrajerzy	327
Lepiej grać czy jednak zrezygnować?	329
Uogólniamy rozkład na więcej niż trzy przypadki	331
Uogólniamy nasze wzory jeszcze bardziej	334
Ile wynosi wartość oczekiwana i wariancja dla tego rozkładu	336
Wartość oczekiwana i wariancja rozkładu dwumianowego	339
Krótki przewodnik po rozkładzie dwumianowym	340
Wartość oczekiwana i wariancja dla rozkładu Poissona	346
Jaki jest więc rozkład prawdopodobieństwa?	350
Liczmy prawdopodobieństwa zdarzeń złożonych dla rozkładu Poissona	351
Rozkład Poissona w przebraniu	354
Krótki przewodnik po rozkładzie Poissona	357

Poznajemy rozkład normalny

8

Wybieramy normalność

Dyskretne rozkłady prawdopodobieństwa nie w każdej sytuacji się sprawdzają.

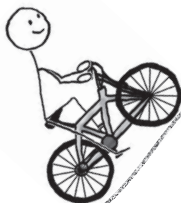
W poprzednim rozdziale poznaliśmy trzy rozkłady prawdopodobieństwa, dla których byliśmy w stanie wymienić wszystkie wartości, jakie może przyjąć zmienna losowa. Nie zawsze jednak jest to możliwe.

Niekiedy posiadane przez nas dane w ogóle **nie przystają** do żadnego z tych trzech rozkładów.

W tym rozdziale dowiemy się o istnieniu rozkładów innego typu, tak zwanych **ciągłych rozkładów prawdopodobieństwa**, oraz poznamy jeden z najważniejszych rozkładów w statystyce — **rozkład normalny**.



Zmienne dyskretne przyjmują wybrane wartości...	364
...ale nie wszystkie zmienne muszą być dyskretne	365
Ile będzie czekać Julie?	366
Musimy znaleźć rozkład prawdopodobieństwa dla danych ciągłych	367
Dla zmiennych ciągłych możemy wyznaczyć funkcję gęstości	368
Prawdopodobieństwo = pole powierzchni	369
Aby obliczyć prawdopodobieństwo, znajdź najpierw $f(x)$...	370
...a następnie oblicz prawdopodobieństwo, wyznaczając pole	371
Znaleźliśmy szukane prawdopodobieństwo	375
Szukanie bratniej duszy	376
Modelujemy wzrost mężczyzn	377
Rozkład normalny stanowi „idealny” model opisu danych ciągłych	378
Jak znajdować prawdopodobieństwa dla rozkładu normalnego?	379
Liczmy prawdopodobieństwo w trzech krokach	380
Krok 1: Wyznacz parametry definiujące rozkład	381
Krok 2: Dokonaj standaryzacji, by otrzymać $N(0, 1)$	382
Aby dokonać standaryzacji, najpierw przesuwamy środek rozkładu...	383
...a następnie zmieniamy jego szerokość	383
Obliczamy Z , dla której będziemy odczytywać prawdopodobieństwo	384
Krok 3: Odczytaj prawdopodobieństwo z tabeli	387



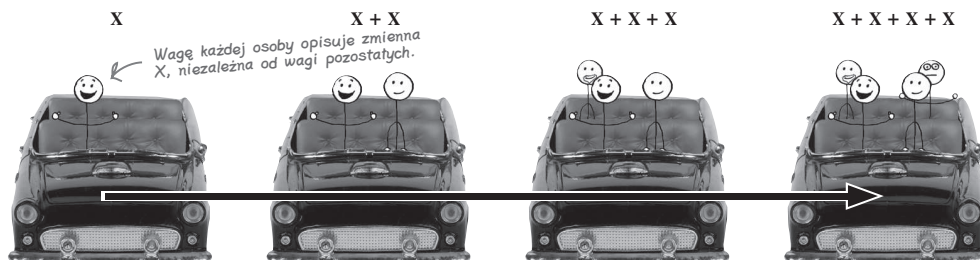
Poznajemy rozkład normalny (cd.)

9 Więcej niż normalność

Gdyby tak istniał jedynie rozkład normalny... Życie byłoby *o wiele prostsze*, gdyby wszystko dało się opisać rozkładem normalnym. Nie trzeba by poświęcać mnóstwa czasu na wyznaczanie różnych prawdopodobieństw w tak wielu rozkładach — można by przeznaczyć go na dużo przyjemniejsze rzeczy, na przykład na rozrywkę. Na szczęście są sposoby na to, by **najbardziej złożone problemy** rozwiązywać równie łatwo, jak w przypadku rozkładu normalnego. W tym rozdziale dowiesz się, kiedy jest możliwe **zastąpienie innego rozkładu** przez rozkład normalny i jak się to robi w praktyce.



Wszyscy na pokład Kolejki Miłości!	401
Sumujemy zmienne o rozkładzie normalnym	402
Nadal jest to jednak waga	403
Jaki jest więc rozkład wagi młodej pary?	405
Znajdujemy prawdopodobieństwo	408
Więcej ludzi chce skorzystać z Kolejki Miłości	413
Przekształcenia liniowe odnoszą się do zmian wartości...	414
...zmienne niezależne mówią o tym, ile różnych wartości posiadasz	415
Wartość oczekiwana i wariancja dla niezależnych zmiennych losowych	416
Wejść do gry czy zrezygnować?	421
Rozkład normalny przychodzi nam z pomocą	424
Kiedy stosować przybliżenie rozkładu dwumianowego rozkładem normalnym	427
Ponowny rzut oka na problem aproksymacji	432
Rozkład dwumianowy jest rozkładem dyskretnym, a normalny — ciągłym	433
Bierzemy poprawkę na ciągłość	434
Cała prawda o rozkładzie normalnym	442
Wszyscy na pokład!	443
Kiedy można aproksymować rozkład Poissona rozkładem normalnym	445
Olbrzymi sukces!	451



Przeprowadzamy losowanie

10

Pobieramy próbkę

Statystyka zajmuje się analizą danych, ale skąd właściwie bierze dane?

Czasami zebranie danych potrzebnych do analiz nie jest wcale trudne. Tak było wtedy, gdy potrzebowaliśmy informacji o przeciętnym wieku klientów klubu odnowy czy też danych o wielkości sprzedaży gier komputerowych. Ale co w sytuacji, gdy potrzebne nam dane nie są ogólnie dostępne i trzeba je w jakiś sposób zdobyć? Czasami ilość różnych informacji, jakie są nam potrzebne, jest na tyle duża, że nie wiemy nawet, jak się do ich gromadzenia zabrać. Od czego zacząć?

W tym rozdziale dowiesz się, jak **gromadzić rzeczywiste dane w sposób efektywny**, odpowiedni do potrzeb i możliwie jak najniższym kosztem. Witamy w świecie losowań!



Wielki test produktów Mighty Gumball	454
Firma traci z powodu zużytych gum	455
Przeprowadzamy testy na próbce, nie na całej populacji	456
Jak przebiega dobór próby	457
Kiedy próba nie jest reprezentatywna	458
Jak dobrać próbę	460
Definiujemy operat losowania	461
Czasami dostajemy próby obciążone	462
Źródła obciążenia próby	463
Jak właściwie dobrać próbę	468
Losowanie próby prostej	468
Jak uzyskać próbę prostą	469
Istnieją także inne schematy losowania	470
Możemy przeprowadzić losowanie warstwowe...	470
...losowanie zespolowe...	471
...a nawet losowanie systematyczne	471
Mighty Gumball dostał swoją próbę	477

Szacujemy parametry populacji na podstawie próby

11

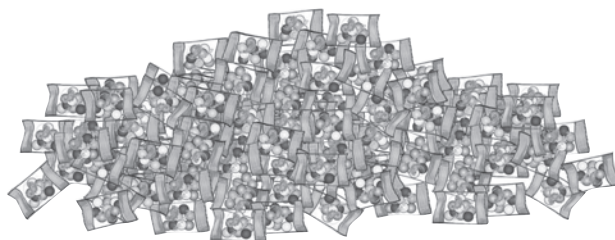
Dokonyjemy ocen

Czy nie byłoby wspaniale, gdybyśmy potrafili scharakteryzować populację generalną na podstawie informacji zawartych w pojedynczej próbie?

Zanim będziesz mógł świętować osiągnięcie **biegłości w prowadzeniu badań** statystycznych, musisz się nauczyć robić właściwy użytek z próbki pobieranej z populacji. Musisz więc posiadać umiejętność **przewidywania charakterystyk populacji** na podstawie informacji zawartych w próbie oraz nauczyć się, jak możesz ocenić **wiarygodność** swoich szacunków. W tym rozdziale pokażemy Ci, w jaki sposób posiadana próbka może być użyta jako użyteczne źródło informacji o badanej populacji i vice versa.

Jaka więc jest rzeczywista trwałość smaku tamtych gum?	480
Zacznijmy od oszacowania średniej w populacji	481
Estymatory punktowe pozwalają oszacować parametry populacji	482
Szacujemy wariancję populacji	486
Znajdujemy inny estymator niż wariancja z próby	487
Która formuła co oznacza?	489
Wszystko jest kwestią proporcji	492
Jaki ma to związek z estymacją parametrów?	497
Rozkład z próby estymatora p	498
Ile wynosi wartość oczekiwana P_s ?	500
A ile wynosi wariancja P_s ?	501
Ustalamy rozkład P_s	502
P_s ma rozkład normalny	503
Musimy znaleźć rozkład dla średniej z próby	509
Rozkład z próby estymatora średniej	510
Znajdujemy wartość oczekiwaną \bar{X}	512
A co z wariancją zmiennej \bar{X} ?	514
Jaki jest więc kształt rozkładu zmiennej \bar{X} ?	518
Jeśli n jest odpowiednio duże, rozkład \bar{X} jest zbliżony do rozkładu normalnego	519
Stosujemy centralne twierdzenie graniczne	520

Fantastycznie!
Mamy komplet bardzo
korzystnych dla nas
statystyk, które
możemy wykorzystać
w naszych
reklamach.



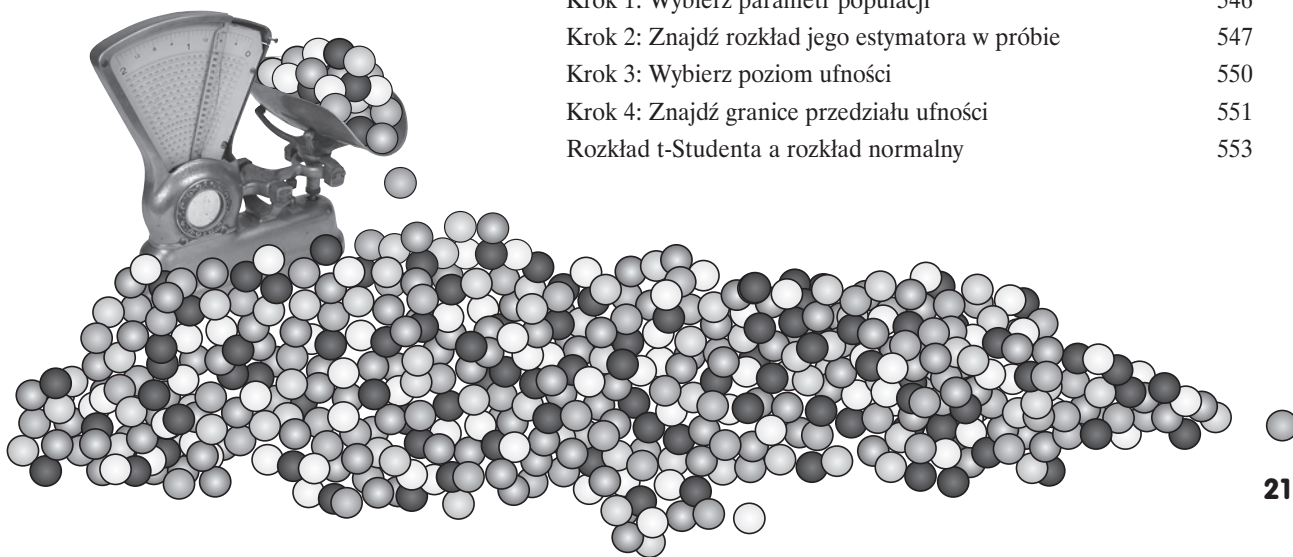
Konstruujemy przedziały ufności

12

Wyrażamy przekonania

Czasami estymacja punktowa daje nie do końca trafione wyniki. Wiesz już, jak za pomocą estymatorów punktowych uzyskać **dokładne oceny** parametrów populacji, takich jak wartość oczekiwana, wariancja czy wskaźnik struktury. Nie zawsze jednak ocena w postaci pojedynczej liczby zaspokoi w pełni Twoje oczekiwania. Bo jak ocenić, na ile jest ona dokładna? Bądź co bądź całe wnioskowanie o populacji generalnej opiera się na stosunkowo nielicznej próbie, która przecież nie zawsze musi w pełni odzwierciedlać charakterystyki populacji. W tym rozdziale poznasz **inną metodę szacowania nieznanymi wartości parametrów populacji**, która uwzględni pewien stopień **niepewności ocen** i — co więcej — pozwala ją zmierzyć. Czytaj dalej, a poznasz wszystkie tajemnice **przedziałów ufności**.

Mighty Gumball znów ma kłopot	526
Problemem pozostaje precyzja	527
Poznajemy przedziały ufności	528
Wyznaczamy przedział ufności w czterech krokach	529
Krok 1: Wybierz parametr populacji	530
Krok 2: Znajdź rozkład jego estymatora w próbie	530
Krok 3: Wybierz poziom ufności	532
Krok 4: Znajdź granice przedziału ufności	534
Zacniemy od wyznaczenia Z	535
Zapisujemy prawdopodobieństwo z użyciem \bar{X}	536
Znajdujemy ostatecznie wartość zmiennej \bar{X}	539
Znaleźliśmy poszukiwany przedział ufności	540
Podsumujmy wykonane kroki	541
Użyteczne skróty przy wyznaczaniu przedziałów ufności	542
Krok 1: Wybierz parametr populacji	546
Krok 2: Znajdź rozkład jego estymatora w próbie	547
Krok 3: Wybierz poziom ufności	550
Krok 4: Znajdź granice przedziału ufności	551
Rozkład t-Studenta a rozkład normalny	553



Weryfikacja hipotez

13

Oceniamy fakty

Nie wszystko, co do Ciebie dociera, musi być prawdą. Najgorsze jest jednak to, że trudno jest ocenić, kiedy ma się do czynienia z prawdą, a kiedy nie. **Weryfikacja hipotez**, drugi obok estymacji dział wnioskowania statystycznego, daje Ci narzędzie do oceny prawdziwości twierdzeń statystycznych. Za jej pomocą będziesz mógł ocenić, na ile takie, a nie inne charakterystyki próby mogą być efektem działalności określonych sił, stanowiących o kształcie całej populacji, a na ile są jedynie dziełem **czystego przypadku**. Z lektury tego rozdziału dowiesz się, w jaki sposób możesz potwierdzić lub obalić swoje przypuszczenia odnoszące się do otaczającej Cię rzeczywistości.

Cudowny lek na chrapanie	560
Ogólne spojrzenie na problem	564
Weryfikacja hipotez w sześciu krokach	565
Krok 1: Sformułuj hipotezę, którą chcesz zweryfikować	566
Krok 2: Wybierz statystykę testową (sprawdzian testu)	569
Krok 3: Określ obszar odrzuceń testowanej hipotezy	570
Krok 4: Znajdź prawdopodobieństwo p (p -wartość)	573
Krok 5: Sprawdź, czy sprawdzian testu wpada do obszaru odrzuceń	575
Krok 6: Podejmij decyzję	575
Co by się stało, gdyby próba była większa?	578
Przeprowadzamy kolejny test	581
Krok 1: Sformułuj hipotezę, którą chcesz zweryfikować	581
Krok 2: Wybierz statystykę testową (sprawdzian testu)	582
Przybliżamy rozkład statystyki testowej rozkładem normalnym	585
Krok 3: Określ obszar odrzuceń testowanej hipotezy	586
Zacznijmy od błędu I rodzaju	594
A co z błędem II rodzaju?	595
Znajdujemy prawdopodobieństwa α i β w naszym przykładzie	596
Znajdujemy zbiór wartości spoza obszaru krytycznego	597
Znajdujemy P (błąd II rodzaju)	598
Moc przybywa	599



Rozkład χ^2

14

Gdy sprawy idą nie po naszej myśli**Czasami sprawy toczą się zupełnie inaczej, niż się tego spodziewaliśmy.**

Kiedy decydujesz się na opis pewnego zjawiska za pomocą konkretnego rozkładu prawdopodobieństwa, zwykle masz jakieś wyobrażenia na temat tego, jak się ono rozwinie w dłuższym okresie. Czasem jednak te **wyobrażenia całkowicie rozmiijają się z rzeczywistością**. Co wtedy począć? Skąd masz wiedzieć, czy dostrzeżone różnice są jedynie dziełem przypadku, czy też może pierwszą oznaką błędnych założeń leżących u podstaw przyjętego przez Ciebie modelu? W tym rozdziale pokażemy Ci, jak możesz posłużyć się rozkładem χ^2 do **oceny rezultatów**, by móc wskazać wśród nich te najbardziej **podejrzane**.

Przed kasynem Dana rysują się kłopoty	606
Przeglądamy się automatom do gry	607
Rozkład χ^2 dobrze modeluje różnice	609
O czym więc mówi ta statystyka?	610
Główne zastosowania rozkładu χ^2	611
ν reprezentuje liczbę stopni swobody	612
Czym jest istotność statystyczna?	613
Testowanie hipotez z rozkładem χ^2	614
Rozwiązałeś tajemnicę wysokich wygranych w grach na automatach	617
Dan ma jeszcze jeden problem	623
Rozkład χ^2 sprawdza się również w testach niezależności	624
Częstości teoretyczne możemy wyznaczyć w oparciu o rachunek prawdopodobieństwa	625
Ile więc wynoszą częstości teoretyczne?	626
Musimy jeszcze poznać liczbę stopni swobody	629
Ogólna metoda wyznaczania liczby stopni swobody	634
A zatem formuła ma postać...	635
Uratowałaś kasyno Dana od bankructwa	637



Korelacja i regresja

15

Co z moją linią?

Czy zastanawiałeś się kiedyś, w jakim stopniu dwie rzeczy są ze sobą powiązane? W poprzednich rozdziałach przyglądaliśmy się użyciu statystyk, które opisywały zbiór danych z punktu widzenia wyłącznie jednej cechy — mówiliśmy na przykład o wzroście mężczyzn, punktach zdobytych przez zawodnika koszykówki czy też o trwałości smaku gum do żucia. Tymczasem istnieją statystyki, które pozwalają ocenić siłę **związku między większą liczbą zmiennych**. Ich znajomość dostarczy Ci znacznie bogatszych informacji na temat otaczającego Cię świata, które będziesz mógł wykorzystać we własnym interesie. W tym rozdziale pokażemy Ci, jak **wykrywać związki** między zmiennymi, korzystając z miar korelacji i regresji.



Przyjrzyjmy się danym na temat frekwencji i nasłonecznienia	645
Rzut oka na wymiary	646
Wykreślamy dane dwuwymiarowe	647
Wykresy rozrzutu pokazują trendy obecne w danych	650
Korelacja a przyczynowość	652
Wykorzystujemy do prognozowania linię o najlepszym dopasowaniu	656
Najlepsze dopasowanie jest nadal tylko dopasowaniem	657
Będziemy minimalizować odchylenia od wartości rzeczywistych	658
Wyznaczamy sumę kwadratów odchyleń	659
Znajdujemy wartości nieznanymi parametrów	660
Obliczamy nachylenie linii najlepszego dopasowania	661
Obliczamy nachylenie linii najlepszego dopasowania (cd.)	662
Znaleźliśmy b, ale co z a?	663
Znaleźliśmy związek między dwiema zmiennymi	667
Różne typy korelacji	668
Współczynnik korelacji mierzy siłę związku między zmiennymi	669
Poznajemy wzór na wartość współczynnika r	670
Obliczamy wartość r dla naszego zbioru danych	671
Obliczamy wartość r dla naszego zbioru danych (cd.)	672



Dodatek uzupełniający

A

Dziesięć najważniejszych rzeczy (które pominęliśmy)

Choć powiedzieliśmy już wiele, coś musimy dopowiedzieć. Jest jeszcze kilka rzeczy, o których, naszym zdaniem, powinieneś wiedzieć. Byłoby trochę nie w porządku wobec Ciebie, gdybyśmy je całkowicie przemilczeli. Nie ma przy tym znaczenia, że są to naprawdę sprawy, które wymagają tylko krótkiej wzmianki.



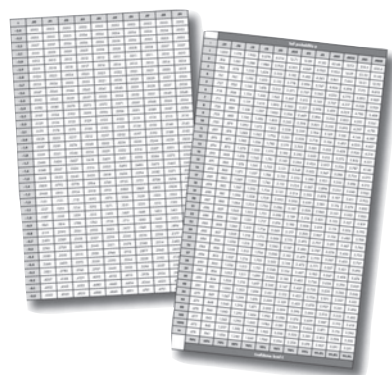
1. Inne techniki wizualizacji danych	682
2. Anatomia rozkładu prawdopodobieństwa	683
3. Eksperyment statystyczny	684
4. Metoda najmniejszych kwadratów w notacji alternatywnej	686
5. Współczynnik determinacji	687
6. Zależności nieliniowe	688
7. Przedział ufności dla współczynnika nachylenia prostej regresji	689
8. Rozkłady z próby — różnica między dwiema średnimi	690
9. Rozkłady z próby — różnica między wskaźnikami struktury	691
10. $E(X)$ i $\text{Var}(X)$ dla zmiennych ciągłych	692

Tablice statystyczne

B

Czasem trzeba coś sprawdzić

Co byśmy poczuli bez pocziwych tablic statystycznych? Nie wystarczy znać zastosowanie poszczególnych rozkładów. Bardzo często trzeba coś policzyć. Dobrze jest mieć wtedy pod ręką **tablice statystyczne**, zawierające **standardowe prawdopodobieństwa** dla typowych rozkładów. Dlatego ten załącznik prezentuje tablice dla rozkładów: **normalnego**, **t-Studenta** i χ^2 .



Standaryzowany rozkład normalny	696
Wartości krytyczne dla rozkładu t-Studenta	698
Wartości krytyczne dla rozkładu χ^2	699

S

Skorowidz

701

5. Dyskretne rozkłady prawdopodobieństwa

Zarządzamy oczekiwaniami



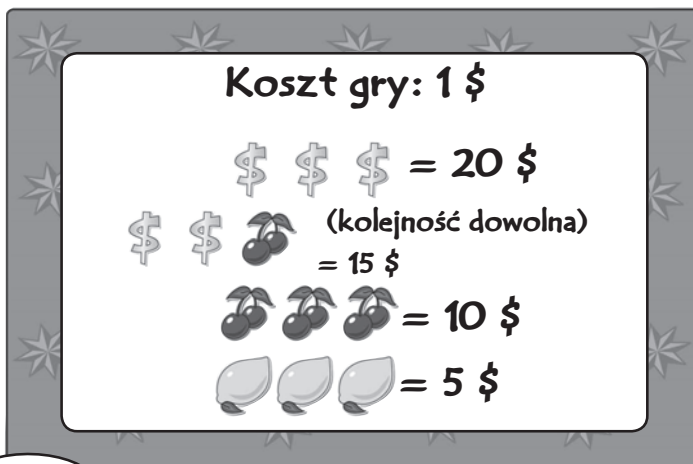
OK, upadek z tego drzewa był faktycznie niespodziewany, ale weź pod uwagę jego konsekwencje w dalszej perspektywie.

Zdarzenia mało prawdopodobne czasem się zdarzają, ale jakie są ich konsekwencje? W poprzednim rozdziale przekonaliśmy się, jak rachunek prawdopodobieństwa może pomóc nam ocenić szanse zajścia pewnych zdarzeń. Jednak nie powie nam on nic na temat **wpływu** tych zdarzeń na nasze życie. Choć przy stole ruletki czasem pada spora wygrana, to jednak — czy jest ona warta tych wszystkich pieniędzy, jakie przy okazji można stracić? W tym rozdziale pokażemy Ci, jak można posłużyć się rachunkiem prawdopodobieństwa do **przewidywania skutków powtarzalnych zdarzeń w długich seriach**, a także jak można **ocenić dokładność** takich prognoz.

Wracamy do kasyna Dana

Czy i Ty poddajesz się hipnotyzującemu wpływowi migających światełek automatu do gry zwanego „jednorękim bandytą”? A zatem jesteś prawdziwym szczęściarzem. W Fat Dan’s Casino znajduje się cały rząd automatów do gry, które tylko czekają na to, by ktoś na nich zagrał. Spróbujmy więc i my (pociągnięcie dźwigni kosztuje 1 dolara). Kto wie, może rozbijemy bank!

Jednoręki bandyta posiada trzy okienka, w których pojawiają się różne symbole. Gdy pojawią się one w odpowiedniej kombinacji, z umieszczonego poniżej otworu posypią się monety.



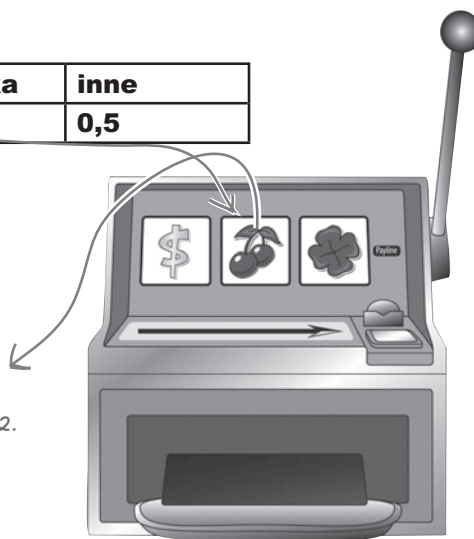
Kwota, jaką można wygrać, wygląda zachęcająco, ale najpierw chciałbym się dowiedzieć, jakie mam szanse na wygraną.

Cóż, wydaje się, że szanse te można by łatwo policzyć. Poniżej zamieszczono prawdopodobieństwo pojawienia się danego symbolu w okienku:

\$	wisienka	cytrynka	inne
0,1	0,2	0,2	0,5

Wszystkie trzy okienka są od siebie niezależne, zatem pojawienie się danego symbolu w jednym z nich nie ma wpływu na to, co pojawi się w pozostałych okienkach.

Prawdopodobieństwo pojawienia się wisiienki w tym okienku wynosi 0,2.



Wczuj się w rolę gracza

Spójrz na planszę z wygranymi zamieszczoną na poprzedniej stronie. Wyobraź sobie, że jesteś graczem, który chce poznać prawdopodobieństwo pojawienia się każdej z kombinacji gwarantujących wygraną. Uzupełnij tabelkę. Jakie jest prawdopodobieństwo tego, że nie wygrasz nic?



prawdopodobieństwo 

prawdopodobieństwo
(kolejność dowolna) 

prawdopodobieństwo 

prawdopodobieństwo 

prawdopodobieństwo braku wygranej

Wczuj się w rolę gracza: Rozwiązanie

Spójrz na planszę z wygranymi zamieszczoną na poprzedniej stronie. Wyobraź sobie, że jesteś graczem, który chce poznać prawdopodobieństwo pojawienia się każdej z kombinacji gwarantujących wygraną. Uzupełnij tabelkę. Jakie jest prawdopodobieństwo tego, że nie wygrasz nic?



prawdopodobieństwo



$$\begin{aligned} P(\$, \$, \$) &= P(\$) \times P(\$) \times P(\$) = \\ &= 0,1 \times 0,1 \times 0,1 = \\ &= 0,001 \end{aligned}$$

← Prawdopodobieństwo pojawienia się symbolu dolara wynosi 0,1.

prawdopodobieństwo
(kolejność dowolna)



Są trzy możliwości uzyskania takiego wyniku:

$$\begin{aligned} P(\$, \$, \text{wisienka}) + P(\$, \text{wisienka}, \$) + P(\text{wisienka}, \$, \$) &= \\ &= (0,1^2 \times 0,2) + (0,1^2 \times 0,2) + (0,1^2 \times 0,2) = \\ &= 0,006 \end{aligned}$$

prawdopodobieństwo



$$\begin{aligned} P(\text{cytrynka}, \text{cytrynka}, \text{cytrynka}) &= \\ &= P(\text{cytrynka}) \times P(\text{cytrynka}) \times P(\text{cytrynka}) = \end{aligned}$$

Pojawienie się cytrynki w jednym z okienek nie wpływa na jej pojawienie się w dwóch pozostałych. Dlatego mnożymy przez siebie te prawdopodobieństwa.

$$\begin{aligned} &= 0,2 \times 0,2 \times 0,2 = \\ &= 0,008 \end{aligned}$$

prawdopodobieństwo



$$\begin{aligned} P(\text{wisienka}, \text{wisienka}, \text{wisienka}) &= \\ &= P(\text{wisienka}) \times P(\text{wisienka}) \times P(\text{wisienka}) = \\ &= 0,2 \times 0,2 \times 0,2 = \\ &= 0,008 \end{aligned}$$

prawdopodobieństwo braku wygranej

To oznacza, że nie uzyskamy żadnej z wygrywających kombinacji.

$$\begin{aligned} P(\text{przegrana}) &= 1 - P(\$, \$, \$) - P(\$, \$, \text{wisienka (kolejność dowolna)}) - P(\text{wisienka}, \text{wisienka}, \text{wisienka}) + \\ &- P(\text{cytrynka}, \text{cytrynka}, \text{cytrynka}) \end{aligned}$$

$$\begin{aligned} &= 1 - 0,001 - 0,006 - 0,008 - 0,008 = \\ &= 0,977 \end{aligned}$$

← To są prawdopodobieństwa wygranej obliczone wyżej.

Zamiast wypisywać wszystkie kombinacje symboli prowadzące do przegranej, zauważamy, że: $P(\text{przegrana}) = 1 - P(\text{wygrana})$.

Tworzymy rozkład prawdopodobieństwa wygranej na automacie

Oto prawdopodobieństwa pojawienia się wszystkich kombinacji symboli gwarantujących wygraną:

To jest zestawienie prawdopodobieństw, które przed chwilą policzyliśmy.

Kombinacja	brak wygranej	cytrynki	wisienki	dolary/ wisienki	dolary
Prawdopodobieństwo	0,977	0,008	0,008	0,006	0,001

Pewnie takie zestawienie może się do czegoś przydać. Czy jednak nie moglibyśmy pójść o krok dalej? Dobrze, że znamy te prawdopodobieństwa, ale nadal nie wiemy, ile możemy wygrać.

Chcielibyśmy znać nie tylko prawdopodobieństwo wygranej, ale również to, jak wiele moglibyśmy wygrać.

Na razie znamy jedynie prawdopodobieństwa pojawienia się na automacie poszczególnych kombinacji symboli, ale nie wiemy nic o wysokości ewentualnej wygranej.

Dlatego nieco zmodyfikujemy naszą tabelkę, przyporządkowując kolejne prawdopodobieństwa do odpowiadającej im wygranej. Wysokość wygranej odczytaną z informacji umieszczonej na automacie pomniejszymy o kwotę 1 dolara, którą musimy zapłacić, by wziąć udział w grze.



Kombinacja	brak wygranej	cytrynki	wisienki	dolary/ dolary	dolary
Wyplata	1 \$	4 \$	9 \$	14 \$	19 \$
Prawdopodobieństwo	0,977	0,008	0,008	0,006	0,001

Tracimy 1 dolara, jeśli nie pojawi się żadna z kombinacji gwarantujących wygraną.

To te same prawdopodobieństwa co powyżej, ale przyporządkowane wygranej, jaka się z nimi wiąże.

Wyplata przy każdej kombinacji: kwota wygranej pomniejszona o 1 dolara opłaty za udział w grze.

W tabeli tej zawarto tak zwany *rozkład prawdopodobieństwa* wypłat z gry na automacie, bowiem każdej wygranej pomniejszonej o koszt wzięcia udziału w grze przyporządkowano prawdopodobieństwo, z jakim może ona paść.



Rozkłady prawdopodobieństwa z bliska

Aby stworzyć rozkład prawdopodobieństwa wypłat z gry na automacie, dla każdej możliwej kwoty wypłaty obliczyliśmy prawdopodobieństwo, z jakim może ona paść. Przy opisie tego rodzaju zdarzeń statystycy posługują się tak zwaną **zmienną losową**, czyli taką zmienną, która może przyjmować różne wartości z pewnego zbioru z określonym prawdopodobieństwem. W naszym przykładzie zmienna losowa reprezentuje wypłatę z gry, a więc wielkość, jaką mamy szansę zarobić, pociągając raz za dźwignię automatu.

Zmienne losowe oznaczamy zwykle wielkimi literami alfabetu łacińskiego, na przykład X czy Y . Poszczególne wielkości, jakie może przyjąć zmienna (tak zwane **realizacje**), oznaczamy odpowiadającymi im małymi literami, na przykład x czy y . Tak więc zapis $P(X = x)$ odczytujemy: „Prawdopodobieństwo, że zmienna X przyjmie wartość x ”.

Oto nasz rozkład prawdopodobieństwa zapisany przy wykorzystaniu tej notacji:

Kombinacja	brak wygranej	cytrynki	wisienki	dolary/wisienki	dolary
x	-1 \$	4 \$	9 \$	14 \$	19 \$
$P(X = x)$	0,977	0,008	0,008	0,006	0,001

Wartość wypłaty dla każdej kombinacji symboli jest reprezentowana przez x .

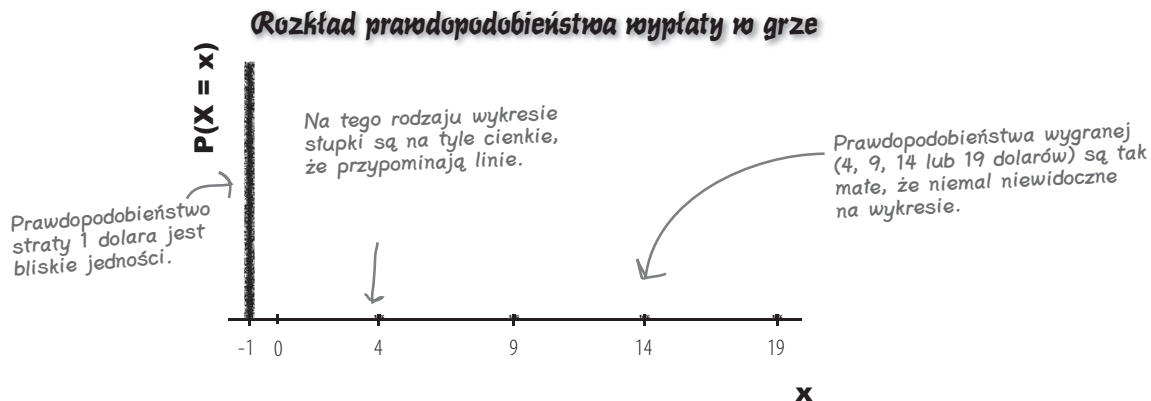
Tutaj x wynosi 19 dolarów.

X to nasza zmienna losowa.

To jest $P(X = 9)$, czyli prawdopodobieństwo tego, że na grze zyskamy 9 dolarów.

Ta zmienna losowa jest zmienną **dyskretną**. Oznacza to, że przyjmuje ona tylko wybrane, dokładne wartości. Rozkład Poissona też jest dyskretny, ale nie jest skończony.

Oprócz przedstawienia rozkładu prawdopodobieństwa w formie tabelarycznej możemy przedstawić go na wykresie, co znacząco ułatwia dalszą analizę rozkładu. Poniżej zamieszczono przykładowy wykres (słupkowy) dla rozkładu naszej zmiennej.





Dlaczego powinienam się przejmować rozkładami prawdopodobieństwa? Ja chcę tylko wiedzieć, ile mogę wygrać. Czy możemy to po prostu policzyć?

Gdy wyznaczymy już rozkład prawdopodobieństwa, możemy posłużyć się nim do oszacowania oczekiwanej realizacji zmiennej losowej.

W naszym przykładzie moglibyśmy wykorzystać wyznaczony rozkład prawdopodobieństwa do oszacowania, jakiej wypłaty moglibyśmy oczekiwać w długiej serii powtórzeń gry na automacie.

Nie istnieją głupie pytania

P: Dlaczego zamiast symboli do budowy rozkładu wykorzystujemy liczby? Czy naprawdę tak wiele na tym zyskujemy?

U: Oczywiście moglibyśmy posługiwać się różnymi symbolami, jednak mają one podstawową wadę — nie nadają się do obliczeń. Już niedługo przekonasz się, że na bazie rozkładu opisanego za pomocą liczb można wyznaczyć różne parametry zmiennej losowej. Nie byłoby to możliwe, gdybyśmy posługiwali się wyłącznie symbolami.

P: Czy mogę przedstawić rozkład prawdopodobieństwa na diagramie Venna?

U: Diagramy Venna raczej nie nadają się do tego celu. Zarówno one, jak i drzewa stochastyczne są pomocne przy obliczaniu prawdopodobieństw, niekoniecznie przy ich wizualizacji.

P: Czy mogę wykorzystać dowolną literę do oznaczenia zmiennej losowej?

U: Oczywiście że tak, pod warunkiem że nie będzie to prowadziło do nieporozumień. Najczęściej wykorzystuje się litery z końca alfabetu: X, Y czy Z.

P: Czy zmienną i jej realizację powinienem oznaczać tą samą literą?

U: W zasadzie mógłbyś użyć różnych liter, ale jak się jeszcze nieraz przekonasz, mogłoby to nadmiernie utrudnić rozeznanie w danych. Lepiej więc pozostać przy tej samej literze.

P: Stwierdziłicie, że zmienna dyskretna może przyjmować tylko wybrane, dokładne wartości. Czy nie jest to regułą?

U: Nie jest. W przykładzie z „jednorękim bandytą” wiemy z góry, jaką wypłatę gwarantują poszczególne kombinacje symboli. Nigdy nie uda nam się uzyskać kombinacji innej niż przewidziana przez konstruktora automatu.

Czasami mamy do czynienia ze zmiennymi losowymi, które mogą przyjmować dowolną wartość z pewnego przedziału.* Średnica jabłka może przyjmować dowolną wartość na przykład z przedziału 5 – 15 cm. Na razie nie musisz się jednak za bardzo tym przejmować. Wrócimy do tego zagadnienia w dalszej części książki. Każda zmienna, którą będziemy się zajmować w tym rozdziale, będzie zmienną dyskretną.

* Liczb naturalnych też jest nieskończenie wiele, ale zmienna losowa przyjmująca wartości naturalne jest nadal dyskretna.

Wartość oczekiwana pozwala przewidzieć wynik...

Znamy już rozkład prawdopodobieństwa wyplaty w grze na automacie, ale nie wiemy, jakiej tak naprawdę wygranej moglibyśmy oczekiwać. Odpowiedź na to pytanie da nam *wartość oczekiwana*. Jest to jeden z parametrów rozkładu prawdopodobieństwa, mówiący o tym, jakiej przeciętnej wartości zmiennej losowej należy się spodziewać w długiej serii jej realizacji.

Wartość oczekiwana zmiennej X przypomina nieco średnią arytmetyczną, tyle że jest wyznaczana na podstawie rozkładu prawdopodobieństwa. Liczymy ją w bardzo podobny sposób: wartość każdej realizacji x zmiennej losowej X mnożymy przez jej prawdopodobieństwo, a następnie sumujemy otrzymane iloczyny.

Wartość oczekiwana zmiennej losowej X oznaczamy zwykle symbolem $E(X)$, jednak równie często wykorzystuje się symbol μ , którym wcześniej oznaczaliśmy średnią arytmetyczną.

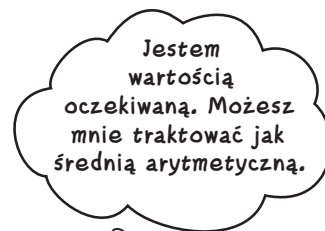
Oto wzór na policzenie wartości oczekiwanej $E(X)$:

Pomnóż każdą wartość x przez jej prawdopodobieństwo.

$E(X) = \sum xP(X = x)$

Gdy wykonasz mnożenie, dodaj do siebie otrzymane iloczyny.

$E(X)$ oznacza wartość oczekiwaną X .



$$E(X) = \mu$$

Posłużymy się tym wzorem do policzenia wartości oczekiwanej naszej zmiennej. Oto jej rozkład (pominęliśmy jednostki):

x	-1	4	9	14	19
P(X = x)	0,977	0,008	0,008	0,006	0,001

$$\begin{aligned}
 E(X) &= (-1 \times 0,977) + (4 \times 0,008) + (9 \times 0,008) + (14 \times 0,006) + (19 \times 0,001) = \\
 &= -0,977 + 0,032 + 0,072 + 0,084 + 0,019 = \\
 &= -0,77
 \end{aligned}$$

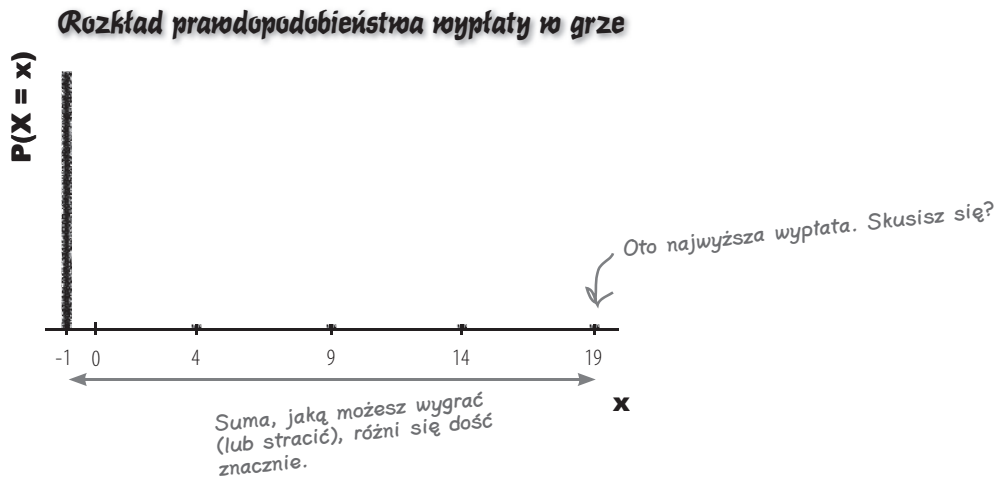
Oto wypłata (w dolarach), jakiej możesz oczekiwać, grając na naszym automacie — jak widzisz, jest ujemna!

Innymi słowy, przy dużej liczbie powtórzeń gry powinieneś oczekiwać, że w każdej z nich stracisz przeciętnie 0,77 dolara. Czyli grając 100 razy, możesz oczekiwać straty na poziomie 77 dolarów.

...a wariancja mówi o tym, jak bardzo jest on zmienny

Wartość oczekiwana mówi o tym, jakiej przeciętnie wypłaty powinniśmy oczekiwać w pojedynczej grze. Ponieważ w naszym przykładzie jest ona ujemna, zastanawiasz się zapewne, dlaczego ludzie decydują się na grę.

Odpowiedź jest prosta: to, że powinieneś oczekiwać raczej strat niż zysków z gry na automacie, nie oznacza jeszcze, że nie masz w ogóle szans na wygraną. Tak jak średnia arytmetyczna, wartość oczekiwana nie mówi nam wszystkiego o danym rozkładzie. Wyniki kolejnych podejść do gry mogą się dość znacząco różnić. Jak sądzisz, czy możemy zmierzyć tę zmienność?



Tak właśnie myślę... skoro wartość oczekiwana przypomina średnią, to może spróbowałibyśmy wyznaczyć również coś w rodzaju wariancji? Jak poprzednio...



Dla zmiennych losowych również wyznaczamy wariancję.

Wartość oczekiwana pozwala oszacować typową, przeciętną realizację zmiennej losowej, ale nie mówi nic o tym, jak bardzo jej wartości mogą się zmieniać. W naszym przykładzie wariancja pozwoli nam oszacować, jak bardzo mogą różnić się od siebie wypłaty w kolejnych podejściach do gry.

Podobnie jak w rozdziale 3., tak i tutaj posłużymy się **wariancją** jako miarą rozproszenia. Zobaczmy, jak to wygląda w praktyce.

Wariancja a rozkład prawdopodobieństwa

W rozdziale 3. wyznaczyliśmy wariancję dla pewnego zbioru liczb. W pierwszym kroku obliczyliśmy wartość wyrażenia $(x-\mu)^2$ dla każdej liczby ze zbioru danych, a następnie wyznaczyliśmy ich średnią arytmetyczną.

Podobnie możemy postąpić w przypadku zmiennej losowej. Nie będziemy jednak liczyć średniej wyrażenia $(X-\mu)^2$, ale jego wartość oczekiwaną, tak jak w poniższej formule:

Oto wariancja. Wariancję zmiennej X oznaczamy w skrócie przez $\text{Var}(X)$.

$$\text{Var}(X) = E(X - \mu)^2$$

μ jest innym oznaczeniem $E(X)$.

Musimy znaleźć wartość oczekiwaną wyrażenia $(X-\mu)^2$. Tylko jak?

Powstał jednak problem: jak policzyć wartość oczekiwaną $(X-\mu)^2$?

Jak więc wyznaczyć $E(X-\mu)^2$?

Obliczanie $E(X-\mu)^2$ przypomina bardzo liczenie zwykłej $E(X)$.

Aby policzyć $E(X)$, musimy przemnożyć każdą wartość zmiennej losowej przez odpowiadające jej prawdopodobieństwo, a następnie dodać do siebie otrzymane iloczyny. Innymi słowy, postępujemy według wzoru:

$$E(X) = \sum xP(X = x)$$

Aby policzyć wariancję zmiennej losowej X , wyznaczamy wartość wyrażenia $(x-\mu)^2$ dla każdej wartości x zmiennej X , mnożymy tę wartość przez odpowiadające jej prawdopodobieństwo, a następnie dodajemy otrzymane iloczyny:

Dla każdej wartości x wyznacz $(x-\mu)^2$. Następnie pomnóż to przez prawdopodobieństwo przypisane x ...

$$E(X - \mu)^2 = \sum (x - \mu)^2 P(X = x)$$

...i dodaj do siebie wyznaczone iloczyny.



Innymi słowy, zamiast mnożyć x przez odpowiadające mu prawdopodobieństwo, mnożymy przez nie wyrażenie o postaci $(x-\mu)^2$.

Obliczamy wariancję dla naszego przykładu

Zobaczymy, jak możemy wykorzystać tę formułę do wyznaczenia wariancji naszej zmiennej losowej. Najpierw odejmiemy wartość oczekiwaną tej zmiennej od każdej z jej realizacji, podniesiemy tę różnicę do kwadratu, a następnie pomnożymy ją przez prawdopodobieństwo przypisane każdej realizacji.

Dla przypomnienia: $E(X)$, czyli μ , wynosi $-0,77$ dolara.

← Na stronie 242 obliczyliśmy $E(X) = -0,77$.

↙ Oto rozkład prawdopodobieństwa dla naszego przykładu.

x	-1	4	9	14	19
P(X = x)	0,977	0,008	0,008	0,006	0,001

$$\text{Var}(X) = E(X - \mu)^2 =$$

$$= (-1 + 0,77)^2 \times 0,977 + (4 + 0,77)^2 \times 0,008 + (9 + 0,77)^2 \times 0,008 + (14 + 0,77)^2 \times 0,006 + (19 + 0,77)^2 \times 0,001 =$$

$$= (-0,23)^2 \times 0,977 + 4,77^2 \times 0,008 + 9,77^2 \times 0,008 + 14,77^2 \times 0,006 + 19,77^2 \times 0,001 =$$

$$= 0,0516833 + 0,1820232 + 0,7636232 + 1,3089174 + 0,3908529 =$$

$$= 2,6971$$

↖ ↗
 $(X - \mu)^2 \times P(X = x)$

Oznacza to, że wariancja zmiennej losowej X (wyплаты w grze) wynosi 2,6971.

A co z odchyleniem standardowym?
Czy także możemy je policzyć?



Podobnie jak wariancję, dla zmiennych losowych możemy także wyznaczyć odchylenie standardowe.

Pełni ono podobną rolę jak odchylenie standardowe, które wyznaczyliśmy w rozdziale 3. dla pewnego zbioru liczb. Określa ono, jak daleko od centrum rozkładu znajdują się przeciętnie wartości zmiennej.

Tak jak poprzednio, wyznaczamy je jako pierwiastek kwadratowy z wariancji:

$$\sigma = \sqrt{\text{Var}(X)}$$

← Odchylenie standardowe możemy oznaczyć tym samym symbolem co poprzednio.

Oznacza to, że odchylenie standardowe wypłaty w grze na naszym automacie wynosi $\sqrt{2,6971}$, czyli 1,642 dolara. Można powiedzieć, że wypłata w grze różni się przeciętnie o 1,642 dolara od wartości oczekiwanej, wynoszącej $-0,77$ dolara.



WYSIL SZARE KOMÓRKI

Czy wolałbyś grać na automacie, który charakteryzuje duża, czy mała zmienność wypłat?
Dlaczego?

P: Czyli wartość oczekiwana jest czymś w rodzaju średniej. Czy dla rozkładów prawdopodobieństwa istnieją również odpowiedniki mediany i dominanty?

U: Można wyznaczyć wartość zmiennej, która ma najwyższe prawdopodobieństwo realizacji, a więc coś w rodzaju dominanty. Zwykle jednak nie robi się tego. W przypadku rozkładów statystycy najczęściej posługują się wartością oczekiwaną jako miarą tendencji centralnej.

P: Czy wartość oczekiwana nie powinna znajdować się w zbiorze możliwych realizacji zmiennej losowej?

U: Niekoniecznie. Tak jak średnia nie musi znajdować się w zbiorze danych, dla którego została wyznaczona, tak i wartość oczekiwana nie musi znajdować się w zbiorze realizacji zmiennej losowej.

P: Czy wariancja i odchylenie standardowe to te same miary, które wyznaczyliśmy dla zbiorów liczb w rozdziale 3.?

U: W zasadzie tak, z wyjątkiem tego, że tym razem operujemy na rozkładach prawdopodobieństwa. I wariancja, i odchylenie standardowe pozostają jednak nadal miarami rozproszenia.

P: Nie do końca rozumiem, dlaczego liczymy $E(x - \mu)^2$. Czy to nie to samo, jak byśmy policzyli $E(x - \mu)$ i podnieśli wynik do kwadratu?

U: Nie, to dwie różne wartości. W przypadku $E(x - \mu)^2$ podnosimy do potęgi drugiej wyrażenie $(x - \mu)$ dla każdego x , a następnie wyznaczamy wartość oczekiwaną takich potęg. Gdy najpierw wyznaczymy $E(x - \mu)$, a następnie podniesiemy ją do potęgi drugiej, otrzymamy zupełnie inny wynik. Tak naprawdę w pierwszym przypadku liczymy więc $E((x - \mu)^2)$, ale upraszczamy zapis.

P: Jaka jest więc różnica między automatem o małej wariancji a tym o dużej wariancji wypłaty?

U: Automat o dużej wariancji wypłaca generalnie bardziej zróżnicowane kwoty. Sumę, jaką w ogólnym przypadku otrzymasz, trudniej jest przewidzieć. Zapamiętaj, że im mniejsza wariancja, tym bliżej oczekiwanej wypłaty będą się znajdować kwoty rzeczywiście wypłacone. Decydując się na grę na automacie o większej wariancji, będziesz musiał się liczyć z wynikiem, który trudniej przewidzieć.

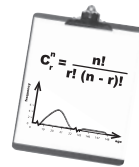


Podstawowe terminy

Wartość oczekiwana

Poniższą formułę wykorzystujemy do wyznaczania wartości oczekiwanej zmiennej losowej X :

$$E(X) = \sum xP(X = x)$$



Podstawowe terminy

Wariancja

Według tej formuły wyznaczamy wariancję zmiennej X :

$$\text{Var}(X) = E(X - \mu)^2$$



Ćwiczenie

Oto rozkład prawdopodobieństwa zmiennej losowej X :

x	1	2	3	4	5
$P(X = x)$	0,1	0,25	0,35	0,2	0,1

1. Ile wynosi $E(X)$?

2. Ile wynosi $\text{Var}(X)$?

Ćwiczenie: Rozwiązanie



Ćwiczenie Rozwiązanie

Oto rozkład prawdopodobieństwa zmiennej losowej X:

x	1	2	3	4	5
P(X = x)	0,1	0,25	0,35	0,2	0,1

1. Ile wynosi $E(X)$?

$$\begin{aligned} E(X) &= \sum xP(X = x) = \\ &= 1 \times 0,1 + 2 \times 0,25 + 3 \times 0,35 + 4 \times 0,2 + 5 \times 0,1 = \\ &= 0,1 + 0,5 + 1,05 + 0,8 + 0,5 = \\ &= 2,95 \end{aligned}$$

Pomnóż każdą wartość zmiennej przez prawdopodobieństwo jej realizacji i dodaj do siebie tak uzyskane iloczyny.

2. Ile wynosi $\text{Var}(X)$?

$$\begin{aligned} \text{Var}(X)^2 &= E(X - \mu)^2 = \\ &= \sum (x - \mu)^2 P(X = x) = \\ &= (1 - 2,95)^2 \times 0,1 + (2 - 2,95)^2 \times 0,25 + (3 - 2,95)^2 \times 0,35 + (4 - 2,95)^2 \times 0,2 + (5 - 2,95)^2 \times 0,1 = \\ &= (-1,95)^2 \times 0,1 + (-0,95)^2 \times 0,25 + 0,05^2 \times 0,35 + 1,05^2 \times 0,2 + 2,05^2 \times 0,1 = \\ &= 3,8025 \times 0,1 + 0,9025 \times 0,25 + 0,0025 \times 0,35 + 1,1025 \times 0,2 + 4,2025 \times 0,1 = \\ &= 0,38025 + 0,225625 + 0,000875 + 0,2205 + 0,42025 = \\ &= 1,2475 \end{aligned}$$

Wyznacz $(x - \mu)^2$ dla każdej wartości x. Następnie pomnóż tę wartość przez prawdopodobieństwo x i dodaj do siebie tak uzyskane iloczyny.



Na tropie
wielkiej
tajemnicy



Przypadek średnich ruchomych

Lokalna stacja telewizyjna w Statsville nadaje kilka popularnych teleturniejów, z których największą widownią cieszy się „Idź na całość”. Każdemu uczestnikowi zabawy prezentowane są skrzynki z umieszczonymi wewnątrz pieniędzmi. Jego zadaniem jest wybór jednej z nich bez zaglądanego do środka. Następnie skrzynki, które nie zostały wybrane przez zawodnika, są stopniowo otwierane. Po ujawnieniu zawartości każdej kolejno otwieranej skrzynki zawodnik musi zdecydować, czy chce pozostać przy swoim pierwotnym wyborze (dokonanym w ciemno), czy też wybrać inną ofertę, którą w międzyczasie złożył mu gospodarz programu. Każda wygrana zawodnika wiąże się z datkiem na miejscowe towarzystwo opieki nad fokami.

Uczestnik dzisiejszego teleturnieju jest statystykiem-amatorem. Zdaje sobie sprawę z tego, że znalazłby się w dużo lepszym położeniu, gdyby poznał wartość oczekiwaną kwot zamieszczonych w poszczególnych skrzynkach. Właśnie skończył ją obliczać, gdy podszedł do niego producent teleturnieju.

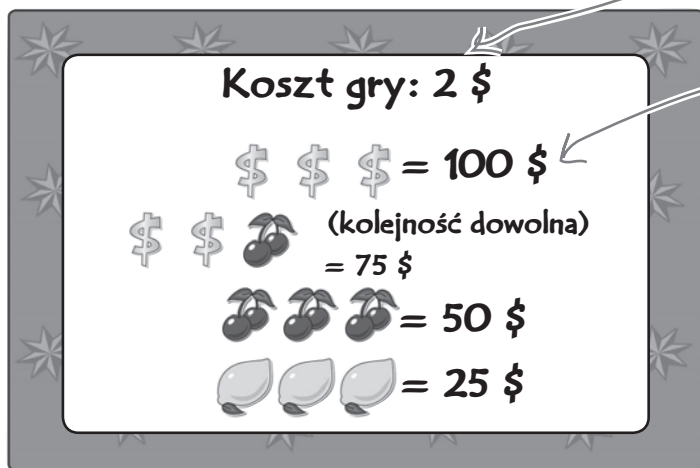
„Wchodzi pan na antenę za 3 minuty” — powiedział producent. „Zmieniliśmy zawartość poszczególnych skrzynek. Zawierają one teraz niemal dwa razy tyle pieniędzy co do tej pory. Bez 10 dolarów”.

Zawodnik spojrzał na producenta z wyrzutem. Czy to oznacza, że jego obliczenia poszły na marne? Wątpił w to, by w ciągu trzech minut udało mu się przeprowadzić wszystkie obliczenia od początku. Co więc powinien zrobić?

Czy nasz zawodnik mógłby w prostszy sposób wyznaczyć wartość oczekiwaną, tak by nie zajęło mu to więcej niż 3 minuty?

Gdy ceny idą w górę

W ciągu kilku ostatnich minut właściciel kasyna podniósł cenę zakładu w grze na automatach, zwiększając jednocześnie wygrane. Oto plansza z nowymi stawkami:



Udział w grze kosztuje 2 dolary, a nie 1 dolar, jak dotąd.

Stawki wygranych wzrosły pięciokrotnie.

**Teraz
płacę 5
razy
więcej!**



Koszt udziału w grze (pociągnięcia za dźwignię automatu) wzrósł do 2 dolarów, ale w ślad za tym poszły w górę stawki wygranych — są teraz pięć razy wyższe. Jeśli więc szczęście będzie nam sprzyjać, zarobimy dużo więcej niż poprzednio.

Oto nowy rozkład prawdopodobieństwa wypłat w tej grze:

y	-2	23	48	73	98
P(Y = y)	0,977	0,008	0,008	0,006	0,001

Tym razem zmienną losową oznaczyliśmy jako Y, nie X.



Gdybyśmy wiedzieli, jaka jest wartość oczekiwana i wariancja tej zmiennej, moglibyśmy oszacować wypłatę w długiej serii powtórzeń.

Zaostrz ołówek



Jaka jest wartość oczekiwana i wariancja zmiennej losowej Y ?
Jak się mają te wartości do wyznaczonych poprzednio: wartości oczekiwanej wynoszącej $-0,77$ dolara oraz wariancji równej $2,6971$?

y	-2	23	48	73	98
P(Y = y)	0,977	0,008	0,008	0,006	0,001

Zaostrz ołówek



Rozwiązanie

Jaka jest wartość oczekiwana i wariancja zmiennej losowej Y ?
Jak się mają te wartości do wyznaczonych poprzednio: wartości oczekiwanej wynoszącej $-0,77$ dolara oraz wariancji równej $2,6971$?

y	-2	23	48	73	98
$P(Y = y)$	0,977	0,008	0,008	0,006	0,001

$$\begin{aligned} E(Y) &= (-2) \times 0,977 + 23 \times 0,008 + 48 \times 0,008 + 73 \times 0,006 + 98 \times 0,001 = \\ &= -1,954 + 0,184 + 0,384 + 0,438 + 0,098 = \\ &= -0,85 \end{aligned}$$

$$\begin{aligned} \text{Var}(Y) &= E(Y - \mu)^2 = \\ &= \sum (y - \mu)^2 P(Y = y) = \\ &= (-2 + 0,85)^2 \times 0,977 + (23 + 0,85)^2 \times 0,008 + (48 + 0,85)^2 \times 0,008 + (73 + 0,85)^2 \times 0,006 + (98 + 0,85)^2 \times 0,001 = \\ &= (-1,15)^2 \times 0,977 + 23,85^2 \times 0,008 + 48,85^2 \times 0,008 + 73,85^2 \times 0,006 + 98,85^2 \times 0,001 = \\ &= 1,3225 \times 0,977 + 568,8225 \times 0,008 + 2386,3225 \times 0,008 + 5453,8225 \times 0,006 + 9771,3225 \times 0,001 = \\ &= 1,2920825 + 4,55058 + 19,09058 + 32,722935 + 9,7713225 = \\ &= 67,4275 \end{aligned}$$

Wartość oczekiwana jest nieco niższa, a zatem w długiej serii powtórzeń oczekujemy przeciętnej straty na poziomie 85 centów w każdej grze. Wariancja jest dużo większa niż poprzednio. Przeciętnie spodziewamy się więc większej straty w każdej grze, choć mamy co do tego mniejszą pewność.

Czy to oznacza,
że przy każdej zmianie stawek
przez Dana będziemy musieli
przeprowadzać te skomplikowane
obliczenia od początku?



Stare i nowe stawki wygranych są ze sobą powiązane.

Koszt udziału w grze zwiększył się do 2 dolarów, zaś stawki wygranych poszły w górę pięciokrotnie. Skoro jest między nimi tak prosta i bezpośrednia zależność, to warto się przekonać, czy podobne zależności istnieją również między parametrami rozkładów zmiennych X i Y .

Przekonajmy się zatem.

Łamigłówka

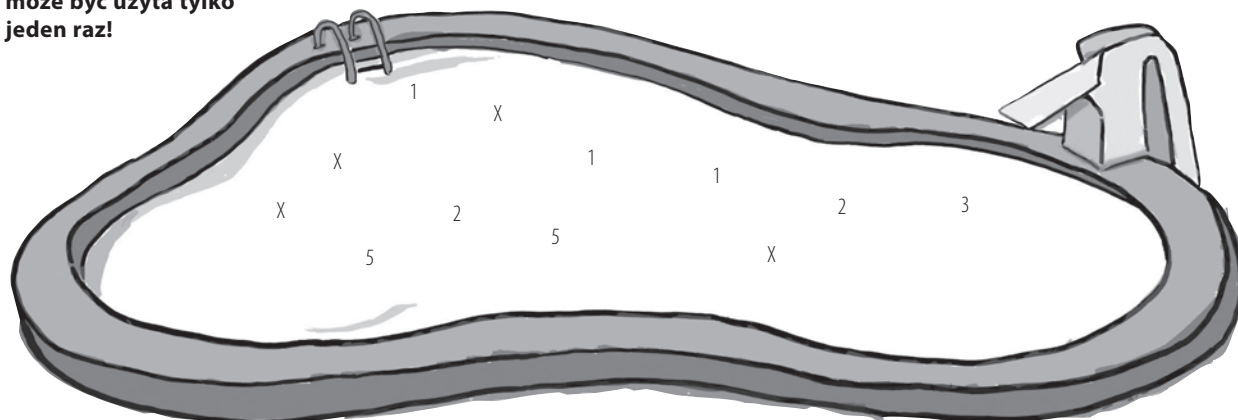


Czas na odrobinę algebry. Uzpełnij luki w poniższych obliczeniach odpowiednimi wartościami znajdującymi się w basenie. Każdą wartość możesz wykorzystać tylko **jeden** raz, choć niektóre nie będą Ci potrzebne w ogóle. Jeśli to zadanie wykonasz bezbłędnie, otrzymasz formułę pokazującą relację między starymi (X) i nowymi (Y) stawkami wyplat w grze na automacie w kasynie Dana.

$$\begin{aligned} X &= (\text{stara stawka wygranej}) - (\text{stary koszt gry}) = \\ &= (\text{stara stawka wygranej}) - \dots\dots\dots \\ (\text{stara stawka wygranej}) &= \dots\dots\dots + \dots\dots\dots \end{aligned}$$

$$\begin{aligned} Y &= 5 (\text{stara stawka wygranej}) - (\text{nowy koszt gry}) = \\ &= 5 (\dots\dots\dots + \dots\dots\dots) - \dots\dots\dots = \\ &= 5 \dots\dots\dots + \dots\dots\dots - \dots\dots\dots = \\ &= \dots\dots\dots \dots\dots\dots + \dots\dots\dots \end{aligned}$$

Uwaga: każda wartość może być użyta tylko jeden raz!



Łamigłówka: Rozwiązanie



Czas na odrobinę algebry. Uzupełnij luki w poniższych obliczeniach odpowiednimi wartościami znajdującymi się w basenie. Każdą wartość możesz wykorzystać tylko **jeden** raz, choć niektóre nie będą Ci potrzebne w ogóle. Jeśli to zadanie wykonasz bezbłędnie, otrzymasz formułę pokazującą relację między starymi (X) i nowymi (Y) stawkami wyplat w grze na automacie w kasynie Dana.

$$X = (\text{stara stawka wygranej}) - (\text{stary koszt gry}) = \leftarrow \text{Koszt gry wynosił przedtem 1 dolar.}$$

$$= (\text{stara stawka wygranej}) - \dots\dots\dots 1 \dots\dots\dots$$

$$(\text{stara stawka wygranej}) = \dots\dots\dots X \dots\dots\dots + \dots\dots\dots 1 \dots\dots\dots \leftarrow \text{W ten sposób przedstawiliśmy starą stawkę wyplaty jako funkcję X.}$$

Podstawiamy tutaj wartości starych stawek wygranej.

$$Y = 5 (\text{stara stawka wygranej}) - (\text{nowy koszt gry}) =$$

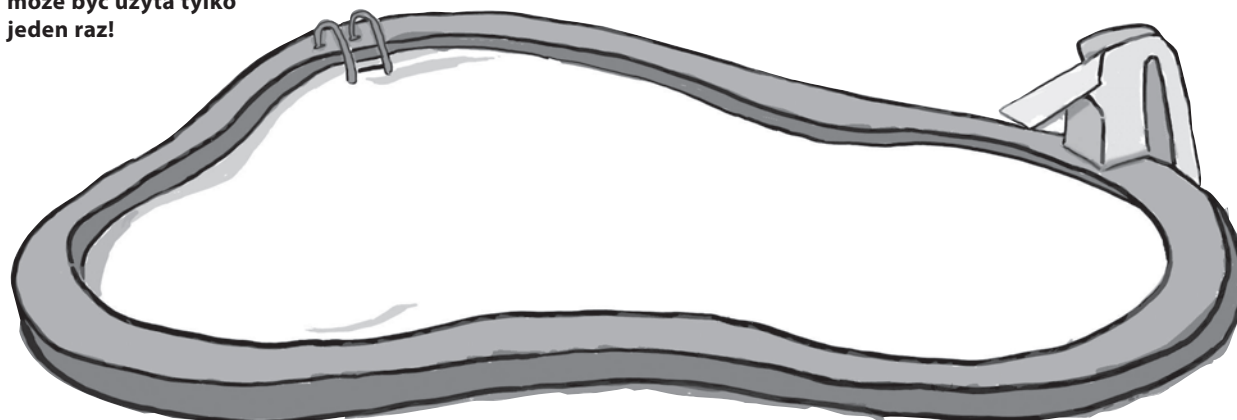
$$= 5 (\dots\dots\dots X \dots\dots\dots + \dots\dots\dots 1 \dots\dots\dots) - \dots\dots\dots 2 \dots\dots\dots =$$

$$= 5 \dots\dots\dots X \dots\dots\dots + \dots\dots\dots 5 \dots\dots\dots - \dots\dots\dots 2 \dots\dots\dots =$$

$$= \dots\dots\dots 5 \dots\dots\dots \dots\dots\dots X \dots\dots\dots + \dots\dots\dots 3 \dots\dots\dots \leftarrow$$

A zatem $Y = 5X + 3$. Między X i Y istnieje ściśle określony związek.

Uwaga: każda wartość może być użyta tylko jeden raz!



Między $E(X)$ i $E(Y)$ istnieje związek liniowy

Wiemy już, że między zmiennymi X (stara wypłata) i Y (nowa wypłata) istnieje zależność liniowa postaci $Y = 5X + 3$. Chcielibyśmy się teraz dowiedzieć, czy podobna zależność istnieje między $E(X)$ i $E(Y)$ oraz $\text{Var}(X)$ i $\text{Var}(Y)$.

Gdyby udało nam się potwierdzić istnienie takiej zależności, kolejne zmiany stawek w kasynie Dana nie byłyby dla nas problemem. Potrafilibyśmy bowiem łatwo wyznaczyć nową wartość oczekiwaną i wariancję na podstawie ich wartości sprzed zmiany.

Zaostrz ołówek



Spróbujmy sprawdzić, czy istnieje zależność między $E(X)$ i $E(Y)$ oraz $\text{Var}(X)$ i $\text{Var}(Y)$.

- $E(X) = -0,77$, zaś $E(Y) = -0,85$. Ile wynosi $5 \times E(X)$? A ile $5 \times E(X) + 3$? Jak się to ma do $E(Y)$?
- $\text{Var}(X) = 2,6971$, zaś $\text{Var}(Y) = 67,4275$. Ile wynosi $5 \times \text{Var}(X)$? A ile $5^2 \times \text{Var}(X)$? Jak się to ma do $\text{Var}(Y)$?
- Czy dałoby się uogólnić zaobserwowane zależności dla dowolnych zmiennych losowych pozostających w relacji: $Y = aX + b$?

Zaostrz ołówek



Rozwiązanie

Spróbujmy sprawdzić, czy istnieje zależność między $E(X)$ i $E(Y)$ oraz $\text{Var}(X)$ i $\text{Var}(Y)$.

1. $E(X) = -0,77$, zaś $E(Y) = -0,85$. Ile wynosi $5 \times E(X)$? A ile $5 \times E(X) + 3$? Jak się to ma do $E(Y)$?

$$5 \times E(X) = -3,85$$

$$5 \times E(X) + 3 = -0,85$$

$$E(Y) = 5 \times E(X) + 3$$

2. $\text{Var}(X) = 2,6971$, zaś $\text{Var}(Y) = 67,4275$. Ile wynosi $5 \times \text{Var}(X)$? A ile $5^2 \times \text{Var}(X)$? Jak się to ma do $\text{Var}(Y)$?

$$5 \times \text{Var}(X) = 13,4855$$

$$5^2 \times \text{Var}(X) = 67,4275$$

$$\text{Var}(Y) = 5^2 \times \text{Var}(X)$$

3. Czy dałoby się uogólnić zaobserwowane zależności dla dowolnych zmiennych losowych pozostających w relacji: $Y = aX + b$?

$$E(aX + b) = a E(X) + b$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X)$$

Podsumujmy nasze rozważania

Podsumujmy to, czego dowiedziałeś się na kilku ostatnich stronach.

Najpierw obliczyłeś wartość oczekiwaną i wariancję zmiennej losowej X , która reprezentuje wypłatę w pojedynczej grze na automacie.

Następnie chciałeś się przekonać, jaki wpływ na parametry rozkładu zmiennej X wywrze zmiana stawek zarządzana przez Dana, właściciela kasyna. Nie chciałeś jednak liczyć wszystkiego od początku, dlatego spróbowałeś znaleźć relację między parametrami starego i nowego rozkładu prawdopodobieństwa. Przekonałeś się, że:

$$E(5X + 3) = 5E(X) + 3$$

$$\text{Var}(5X + 3) = 5^2 \text{Var}(X)$$



Ogólne wzory na przekształcenia liniowe

Nasze rezultaty możemy uogólnić na dowolną zmienną losową.
Dla każdej zmiennej losowej X zachodzą poniższe równości:

$$E(aX + b) = aE(X) + b$$

Pomnóż wartość oczekiwaną przez a ,
a następnie dodaj b .

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

Pomnóż wariancję przez kwadrat
 a (pomijając b).

Tego rodzaju transformacje nazywamy przekształceniami liniowymi, ponieważ zmienna X występuje w pierwszej potęgce, a a o zmiennej Y mówimy, że jest funkcją X . Zmieniamy tu jedynie potencjalne wartości X , przekształcając je do postaci $aX + b$, lecz nie zmieniając prawdopodobieństwa ich realizacji.

————— Nie istnieją grupie pytania —————

P: Czy a i b muszą być stałymi liczbami?

U: Muszą. Jeśli a i b nie są stałe, powyższe wzory nie są prawdziwe.

P: Co się stało z b we wzorze na wariancję?

U: Dodanie ustalonej wartości do zmiennej losowej nie zmienia jej wariancji, a jedynie wartość oczekiwaną.

Spowoduje ono przesunięcie środka rozkładu w pewnym kierunku, ale nie zmieni ogólnego kształtu jego wykresu. Oznacza to, że zmieni się wartość oczekiwana, ale rozrzut danych, a więc i wariancja, pozostanie niezmieniony.

P: Zaskoczyło mnie to, że wariancję trzeba pomnożyć przez a^2 . Dlaczego nie przez a ?

U: Mnożąc zmienną losową przez stałą, każdą jej wartość mnożymy przez tą stałą.

We wzorze na wariancję wartości zmiennej losowej podnoszone są do potęgi drugiej. Dlatego gdy pomnożymy je przez a , wariancja zwiększy się a^2 razy.

P: Czy naprawdę muszę pamiętać wzory na przekształcenia liniowe parametrów?
Czy są one aż tak ważne?

U: Tak, są ważne. Dzięki nim możesz zaoszczędzić mnóstwo czasu, który musiałbyś przeznaczyć na wyznaczanie wartości parametrów rozkładu za każdym razem, gdy zmieniają się wartości zmiennej losowej. Zamiast więc wyznaczać rozkład nowej zmiennej, obliczać od nowa jej wartość oczekiwaną i wariancję, musisz podstawić jedynie do odpowiedniego wzoru stare wartości parametrów.

Znajomość tych wzorów może się też przydać w czasie egzaminów ze statystyki. Znajomość tego rodzaju skrótów pozwoli Ci oszczędzić sporo cennego czasu. Musisz też pamiętać, że na egzaminach nie zawsze podawane są pełne rozkłady prawdopodobieństwa, a jedynie pewne ich charakterystyki. Niekiedy więc możesz nie mieć wyjścia.

P: Próbowałem wyznaczyć oba parametry na piechotę, ale otrzymałem inne wartości. Dlaczego?

U: A zatem przekonałeś się, że liczenie na piechotę może być ryzykowne. Po drodze masz bowiem wiele okazji do popełnienia błędów. Zawsze, kiedy masz taką możliwość, powinieneś korzystać ze znanych Ci skrótów.

Przypadek średnich ruchomych: Rozwiązanie

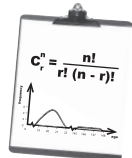
Czy nasz zawodnik mógłby w prostszy sposób wyznaczyć wartość oczekiwaną, tak by nie zajęło mu to więcej niż 3 minuty?

Przez krótką chwilę nasz zawodnik rozglądał się nerwowo dokoła, jednak szybko się uspokoił. Zmiana stawek wygranej to w końcu nie taki duży problem.

Już raz policzył wartość oczekiwaną kwot znajdujących się w kolejnych skrzynkach, co dało mu jakiś ogłęd tego, ile może wygrać w tym teleturnieju.

Producent programu powiedział, że nowe kwoty pieniędzy ukryte w skrzynkach są niemal dwa razy wyższe od dotychczasowych, z dokładnością do 10 dolarów. Oznacza to, że między nowymi (Y) i starymi (X) kwotami istnieje liniowy związek, który można by zapisać jako $Y = 2X - 10$.

To pozwala szybko obliczyć $E(Y)$ przy wykorzystaniu faktu, że $E(2X - 10) = 2E(X) - 10$. Zatem jedyne, co powinien zrobić nasz zawodnik, to pomnożyć obliczoną wcześniej wartość oczekiwaną wygranej przez 2 i od tego wyniku odjąć 10 dolarów.



Podstawowe terminy

Przekształcenia liniowe

Dla każdej zmiennej losowej X i stałych wartości a i b prawdziwe są równości:

$$E(aX + b) = aE(X) + b$$

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

CELNE SPOSTRZEŻENIA

- **Rozkład prawdopodobieństwa** przypisuje każdej wartości zmiennej losowej prawdopodobieństwo jej realizacji.
- **Wartość oczekiwana** informuje o przeciętnej realizacji zmiennej losowej w długiej serii powtórzeń. Oznacza się ją jako $E(X)$ lub μ i oblicza ze wzoru: $E(X) = \sum xP(X = x)$.
- **Wartość oczekiwana funkcji** zmiennej losowej X dana jest wzorem: $E(f(X)) = \sum f(x)P(X = x)$.
- **Wariancję** zmiennej losowej X wyznacza się według wzoru: $\text{Var}(X) = E(X - \mu)^2$.
- **Odchylenie standardowe** zmiennej losowej dane jest wzorem: $\sigma = \sqrt{\text{Var}(X)}$
- Z **przekształceniem liniowym** zmiennej losowej X mamy do czynienia wtedy, gdy jej wartości przekształcamy do ogólnej postaci: $aX + b$, gdzie a i b są wartościami ustalonymi. Wartość oczekiwana i wariancja przekształconej zmiennej są równe:
 $E(aX + b) = aE(X) + b$
 $\text{Var}(aX + b) = a^2\text{Var}(X)$



A zatem przedstawiając jedną zmienną losową jako funkcję innej zmiennej, będę mogła łatwo wyznaczyć parametry rozkładu także wtedy, gdy zagram więcej niż raz?

Przekształcenia liniowe nie sprawdzą się wtedy, gdy zmienia się prawdopodobieństwo realizacji zmiennej.

W wyniku liniowego przekształcenia zmiennej losowej zmieniają się jedynie wartości, jakie może ona przyjąć. Prawdopodobieństwa ich realizacji nie ulegają zmianie. Nie zmienia się również liczba potencjalnych wartości.

Gdy myślimy o rozegraniu kilku gier z rzędu, rozkład prawdopodobieństwa wypłaty będzie całkowicie inny: zmienią się zarówno wartości samej zmiennej, jak i prawdopodobieństwa ich realizacji. Nie wystarczy więc policzyć wyłącznie nowe wartości rozkładu, ale i ich prawdopodobieństwa, co nie musi być zadaniem łatwym.

Spójrzmy na przykład. Załóżmy, że gra toczy się na bardzo prostym automacie, który oferuje wypłaty według poniższego rozkładu:

x	-1	5
P(X = x)	0,9	0,1

Aby znaleźć rozkład prawdopodobieństwa zmiennej $2X$, wystarczy pomnożyć przez 2 wszystkie wartości zmiennej X :

2x	-2	10
P(2X = 2x)	0,9	0,1

Wartości zmiennej zostały pomnożone przez 2. Prawdopodobieństwa nie zmieniły się.



Jak wyglądałby rozkład prawdopodobieństwa wypłaty z gry przy założeniu, że zagralibyśmy dwukrotnie? Tym razem wszystkie prawdopodobieństwa musimy obliczyć od nowa, rozpatrując wszystkie możliwe kombinacje wygranej (przegranej) w obu grach:

$y = -2$, gdy przegrasz w obu grach.

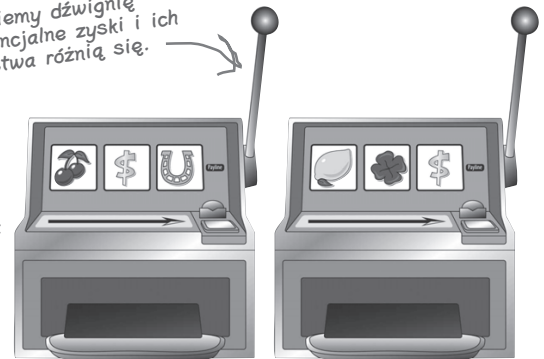
w	-2	4	10
P(Y = y)	0,81	0,18	0,01

Są to łączne wyniki uzyskane w obu grach

Tym razem ciągniemy dźwignię dwukrotnie. Potencjalne zyski i ich prawdopodobieństwa różnią się.

$y = 10$, gdy wygrasz w obu grach.

$y = 4$, gdy przegrasz w jednej grze, ale wygrasz w drugiej.



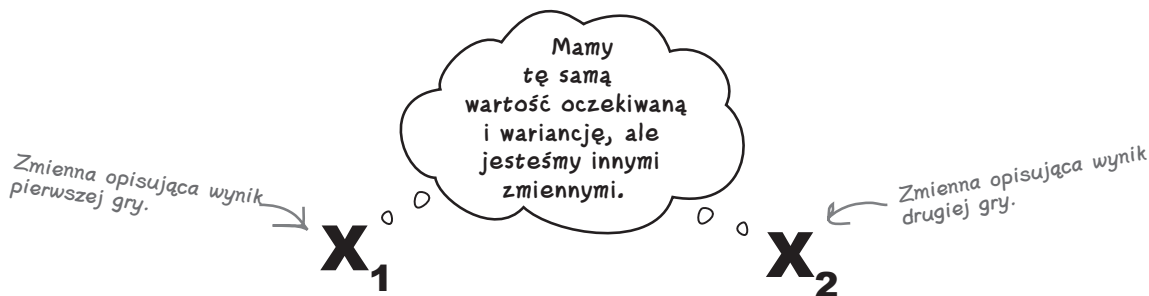
Tym razem zarówno wartości zmiennej, jak i ich prawdopodobieństwa są nieco inne. Czy istnieje jakiś sposób na skrócenie obliczeń w podobnych przypadkach?

Każde pociągnięcie dźwigni jest niezależnym zdarzeniem

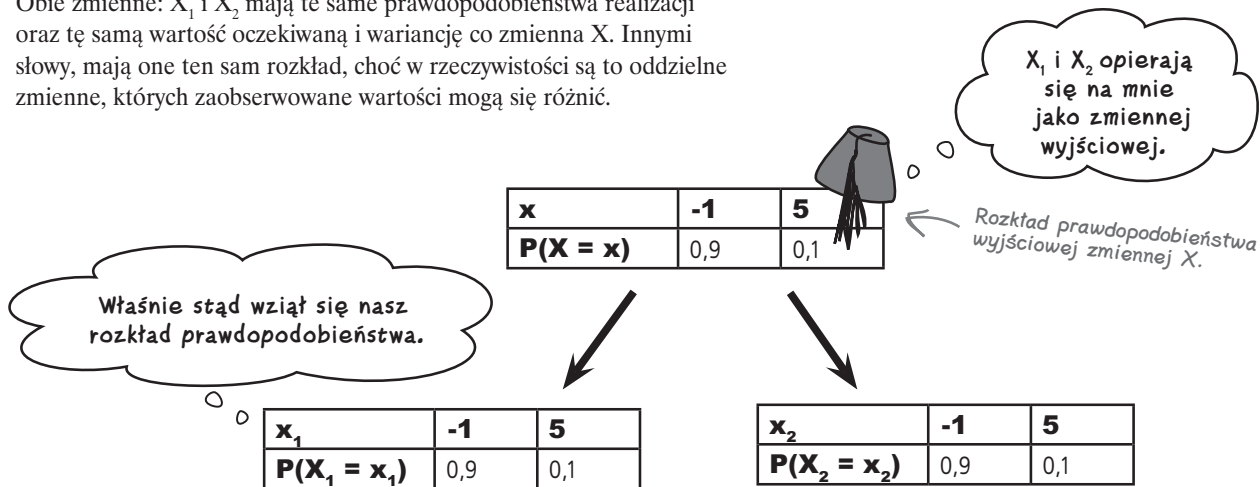
Każdą grę możemy traktować jako niezależne zdarzenie, opisywane za pomocą oddzielnej zmiennej losowej, a jej wynik jako oddzielną obserwację. Wszystkie obserwacje będące realizacjami zmiennych losowych o tym samym rozkładzie, mają tę samą wartość oczekiwaną i wariancję, choć same w sobie mogą się od siebie różnić — za każdym razem możemy uzyskać nieco inną kwotę.

Przydałby nam się jakiś sposób na to, by łatwo odróżnić od siebie poszczególne zdarzenia czy obserwacje. Jeśli wypłatę w pojedynczej grze oznaczyliśmy przez X , to zmienne opisujące wygrane w poszczególnych kolejkach moglibyśmy oznaczyć na przykład poprzez X_1, X_2 itd.

Każda gra to jedno zdarzenie. Wynik każdej gry to jedna obserwacja.



Obie zmienne: X_1 i X_2 mają te same prawdopodobieństwa realizacji oraz tę samą wartość oczekiwaną i wariancję co zmienna X . Innymi słowy, mają one ten sam rozkład, choć w rzeczywistości są to oddzielne zmienne, których zaobserwowane wartości mogą się różnić.



Jeśli więc chcemy znaleźć wartość oczekiwaną i wariancję łącznej wypłaty w obu grach, tak naprawdę musimy policzyć je dla sumy zmiennych $X_1 + X_2$. Spójrzmy, jak możemy tego dokonać jak najmniejszym kosztem.

Przydatne skróty

Spróbujmy znaleźć wartość oczekiwaną i wariancję dla sumy zmiennych $X_1 + X_2$.

Wartość oczekiwana

Policzmy najpierw $E(X_1 + X_2)$:

$$\begin{aligned} E(X_1 + X_2) &= E(X_1) + E(X_2) = \\ &= E(X) + E(X) = \\ &= 2E(X) \end{aligned}$$

Zarówno $E(X_1)$, jak i $E(X_2)$ są równe $E(X)$, ponieważ obie zmienne mają rozkład taki jak X .

Innymi słowy, aby wyznaczyć $E(X_1 + X_2)$, wystarczy pomnożyć $E(X)$ przez 2. Gdybyśmy więc dwukrotnie zegrali na automacie, dla którego $E(X) = -0,77$ dolara, moglibyśmy oczekiwać straty na poziomie $-0,77 \times 2$, czyli $-1,54$ dolara.

Wynik ten możemy uogólnić na dowolną liczbę zmiennych. Gdybyśmy chcieli wyznaczyć wartość oczekiwaną n zmiennych tego rodzaju, moglibyśmy wykorzystać wzór:

$$E(X_1 + X_2 + \dots + X_n) = nE(X)$$

Wariancja

A jak policzyć $\text{Var}(X_1 + X_2)$? Odpowiedź poniżej:

$$\begin{aligned} \text{Var}(X_1 + X_2) &= \text{Var}(X_1) + \text{Var}(X_2) = \\ &= \text{Var}(X) + \text{Var}(X) = \\ &= 2\text{Var}(X) \end{aligned}$$

$\text{Var}(X_1)$ i $\text{Var}(X_2)$ są takie same jak $\text{Var}(X)$, ponieważ obie zmienne mają rozkład taki jak X .

Dlatego wariancja wypłaty w przypadku dwukrotnej gry na automacie o wariancji równej 2,6971 wyniosłaby $2 \times 2,6971$, czyli 5,3942.

Wynik ten możemy uogólnić na dowolną liczbę niezależnych obserwacji. Gdybyśmy chcieli wyznaczyć wariancję n obserwacji tego rodzaju, moglibyśmy wykorzystać wzór:

$$\text{Var}(X_1 + X_2 + \dots + X_n) = n\text{Var}(X)$$

Podsumowując: aby znaleźć wartość oczekiwaną i wariancję w przypadku wielokrotnej gry, pomnóż $E(X)$ i $\text{Var}(X)$ przez liczbę obserwacji (kolejek gry).



$X_1 + X_2$ to nie to samo co $2X$.

Sumując zmienne X_1 i X_2 , musisz się oprzeć na realizacjach każdej z nich. $2X$ oznacza tylko jedną realizację, tyle że podwojoną w stosunku do X .

Gdy mamy n zmiennych, mnożymy przez n wartość $E(X)$.

Pomnóż $\text{Var}(X)$ przez liczbę obserwacji n .

Nie istnieją głupie pytania

P: Czy $E(X_1 + X_2)$ to nie to samo, co $E(2X)$?

U: Nie, choć rzeczywiście wyglądają podobnie.

$E(2X)$ jest wartością oczekiwaną zmiennej losowej, której wartości zostały podwojone względem zmiennej X . Mamy więc do czynienia z wartościami jednej zmiennej.

$E(X_1 + X_2)$ to wartość oczekiwana sumy dwóch różnych zmiennych losowych: X_1 i X_2 . Jeśli X_1 oznacza wygraną w jednej grze, a X_2 w drugiej, $X_1 + X_2$ oznaczać będzie łączną wygraną w obu grach.

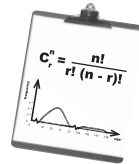
P: Rozumiem więc, że X_1 i X_2 są tym samym?

U: Niezupełnie. Obie zmienne mają ten sam rozkład prawdopodobieństwa, choć są to dwie różne zmienne. X_1 może oznaczać wygraną w jednej kolejce, a X_2 w drugiej. Choć mają ten sam rozkład prawdopodobieństwa, to wygrane w każdej kolejce mogą się od siebie znacząco różnić.

P: Dlaczego wariancja w przypadku n zmiennych wynosi $n\text{Var}(X)$, a nie $n^2\text{Var}(X)$, jak dla zmiennych przekształconych liniowo?

U: Tym razem mamy do czynienia z szeregiem niezależnych zmiennych losowych o tym samym rozkładzie prawdopodobieństwa. Całkowitą wariancję możemy wyznaczyć jako sumę wariancji poszczególnych zmiennych. Ponieważ mamy n zmiennych niezależnych, wynosi ona $n\text{Var}(X)$.

W przypadku zmiennej nX każdą z wartości zmiennej X mnożymy przez n . Ponieważ w formule na wariancję występują kwadraty tych wartości, stąd w wyniku $n^2\text{Var}(X)$, a nie $n\text{Var}(X)$.



Podstawowe terminy

Zmienne niezależne

Dla niezależnych zmiennych losowych prawdziwe są wzory:

$$E(X_1 + X_2 + \dots + X_n) = nE(X)$$

$$\text{Var}(X_1 + X_2 + \dots + X_n) = n\text{Var}(X)$$

CELNE SPOSTRZEŻENIA

- Rozkład prawdopodobieństwa opisuje prawdopodobieństwo realizacji każdej możliwej wartości zmiennej losowej.
- Wartość oczekiwana zmiennej losowej X opisuje przeciętny wynik możliwy do uzyskania w długiej serii powtórzeń. Oznaczamy ją jako $E(X)$ lub μ i obliczamy ze wzoru:

$$E(X) = \sum xP(X=x)$$

- Wariancja zmiennej losowej X dana jest wzorem:

$$\text{Var}(X) = E(X - \mu)^2$$

- Odchylenie standardowe σ jest pierwiastkiem kwadratowym z wariancji.
- Przekształcenie liniowe polega na przekształceniu zmiennej X w zmienną $aX + b$, gdzie a i b są stałymi. Wartość oczekiwaną i wariancję nowej zmiennej można obliczyć ze wzorów:

$$E(aX + b) = aE(X) + b$$

$$\text{Var}(aX + b) = a^2\text{Var}(X)$$

Zmienne przekształcone czy niezależne?

Poniżej zamieszczono opisy kilku sytuacji z życia codziennego. Zakładając, że znasz rozkład zmiennej X , zdecyduj, czy występujące w opisach zdarzenia można opisać za pomocą przekształcenia liniowego (funkcji) zmiennej X , czy też występują w nich zmienne niezależne.

	Przekształcenie liniowe	Zmienne niezależne
Ilość kawy, która składa się na dużą porcję; X — ilość kawy zawarta w standardowej porcji.	<input type="checkbox"/>	<input type="checkbox"/>
Spożywanie dodatkowego kubka kawy każdego dnia; X — ilość kawy w jednym kubku.	<input type="checkbox"/>	<input type="checkbox"/>
Wyplata możliwa do uzyskania po nabyciu 10 losów na loterii; X — wypłata możliwa do uzyskania po nabyciu jednego losu.	<input type="checkbox"/>	<input type="checkbox"/>
Wyplata możliwa do uzyskania po nabyciu pojedynczego losu, którego cena poszła w górę; X — wypłata możliwa do uzyskania po nabyciu pojedynczego losu w starej cenie.	<input type="checkbox"/>	<input type="checkbox"/>
Kupno dodatkowych kur znoszących jajka; X — liczba znoszonych jaj w zależności od gatunku kury.	<input type="checkbox"/>	<input type="checkbox"/>

Zmienne przekształcone czy niezależne?

Poniżej zamieszczono opisy kilku sytuacji z życia codziennego. Zakładając, że znasz rozkład zmiennej X , zdecyduj, czy występujące w opisach zdarzenia można opisać za pomocą przekształcenia liniowego (funkcji) zmiennej X , czy też występują w nich zmienne niezależne.

Przekształcenie liniowe

Zmienne niezależne

Ilość kawy, która składa się na dużą porcję;
 X — ilość kawy zawarta w standardowej porcji.



Spożywanie dodatkowego kubka kawy każdego dnia; X — ilość kawy w jednym kubku.



Wyplata możliwa do uzyskania po nabyciu 10 losów na loterii; X — wyplata możliwa do uzyskania po nabyciu jednego losu.



Wygrane z każdego losu są niezależne od siebie.

Wyplata możliwa do uzyskania po nabyciu pojedynczego losu, którego cena poszła w górę; X — wyplata możliwa do uzyskania po nabyciu pojedynczego losu w starej cenie.



Zmiana ceny losu wpływa na wartość wyplaty, ale nie na prawdopodobieństwo wygranej, dlatego mamy do czynienia z przekształceniem liniowym.

Kupno dodatkowych kur znoszących jajka;
 X — liczba znoszonych jaj w zależności od gatunku kury.





Ćwiczenie

Lokalna cukiernia włączyła do swojej oferty ciasteczka z niespodzianką. Ich cena to 0,5 dolara za sztukę. Większość ciasteczek zawiera pomyślną wróżbę na przyszłość, ale w niektórych kryją się pieniądze. Prawdopodobieństwo wygrania 2 dolarów wynosi 0,1, 5 dolarów — 0,07, zaś 10 dolarów — 0,03.

Niech X oznacza wypłatę (wygrana pomniejszona o koszt ciasteczka) z tej „gry”. Sporządź rozkład prawdopodobieństwa zmiennej X . Ile wynosi $E(X)$ i $\text{Var}(X)$?

Cukiernia zdecydowała się podnieść cenę ciasteczek do 1 dolara za sztukę. Ile teraz wynosi wartość oczekiwana i wariancja wypłaty?

Ćwiczenie: Rozwiązanie



Ćwiczenie Rozwiązanie

Lokalna cukiernia włączyła do swojej oferty ciasteczka z niespodzianką. Ich cena to 0,5 dolara za sztukę. Większość ciasteczek zawiera pomyślną wróżbę na przyszłość, ale w niektórych kryją się pieniądze. Prawdopodobieństwo wygrania 2 dolarów wynosi 0,1, 5 dolarów — 0,07, zaś 10 dolarów — 0,03.

Niech X oznacza wypłatę (wygrana pomniejszona o koszt ciasteczka) z tej „gry”. Sporządź rozkład prawdopodobieństwa zmiennej X . Ile wynosi $E(X)$ i $\text{Var}(X)$?

Oto rozkład prawdopodobieństwa zmiennej X :

x	-0,5	1,5	4,5	9,5
$P(X = x)$	0,8	0,1	0,07	0,03

$$\begin{aligned} E(X) &= (-0,5) \times 0,8 + 1,5 \times 0,1 + 4,5 \times 0,07 + 9,5 \times 0,03 = \\ &= -0,4 + 0,15 + 0,315 + 0,285 = \\ &= 0,35 \end{aligned}$$

$$\begin{aligned} \text{Var}(X) &= E(X - \mu)^2 = \\ &= \sum (x - \mu)^2 P(X = x) = \\ &= (-0,5 - 0,35)^2 \times 0,8 + (1,5 - 0,35)^2 \times 0,1 + (4,5 - 0,35)^2 \times 0,07 + (9,5 - 0,35)^2 \times 0,03 = \\ &= (-0,85)^2 \times 0,8 + 1,15^2 \times 0,1 + 4,15^2 \times 0,07 + 9,15^2 \times 0,03 = \\ &= 0,7225 \times 0,8 + 1,3225 \times 0,1 + 17,2225 \times 0,07 + 83,7225 \times 0,03 = \\ &= 0,578 + 0,13225 + 1,205575 + 2,511675 = \\ &= 4,4275 \end{aligned}$$

Cukiernia zdecydowała się podnieść cenę ciasteczek do 1 dolara za sztukę. Ile teraz wynosi wartość oczekiwana i wariancja wypłaty?

Cukiernia podniosła cenę ciasteczek o 0,5 dolara, a zatem wypłatę po podwyżce można przedstawić jako $X - 0,5$.

$$\begin{aligned} E(X - 0,5) &= E(X) - 0,5 = \\ &= 0,35 - 0,5 = \\ &= -0,15 \end{aligned}$$

$$\begin{aligned} \text{Var}(X - 0,5) &= \text{Var}(X) = \\ &= 4,4275 \end{aligned}$$

Nowe automaty wchodzą do gry!

Właściciel kasyna sprowadził całkiem nowy model automatów do gry. Każda gra kosztuje teraz więcej, ale jeśli szczęście będzie Ci sprzyjać, będziesz mógł liczyć na naprawdę spore pieniądze. Oto nowy rozkład prawdopodobieństwa:

x	-5	395
P(X = x)	0,99	0,01

Gra kosztuje więcej niż na innych automatach, ale spójrz na te wygrane!

Wiemy już, jak wyznaczyć wartość oczekiwaną i wariancję zarówno w pojedynczej grze, jak i w przypadku kilku gier rozegranych na tym samym automacie. Co by się jednak stało, gdybyśmy chcieli rozegrać kilka kolejek i na starych, i na nowych urządzeniach?

W tej sytuacji mamy do czynienia z dwoma niezależnymi rozkładami prawdopodobieństwa — dla maszyn starego i nowego typu:

x	-5	395
P(X = x)	0,99	0,01

Oto wygrane możliwe do uzyskania na automatach nowego typu.

y	-2	23	48	73	98
P(Y = y)	0,977	0,008	0,008	0,006	0,001

Oto wygrane w automatach starego typu.

Czy możemy łatwo policzyć wartość oczekiwaną i wariancję wypłaty w przypadku wielokrotnej gry na obu rodzajach automatów?

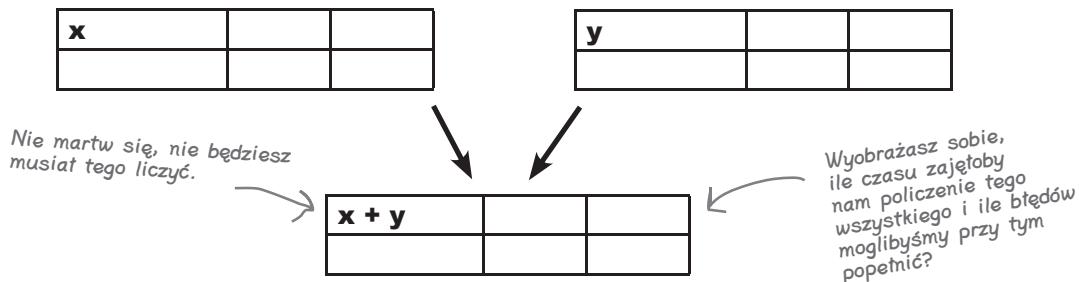
Moglibyśmy wyznaczyć rozkład sumy $X+Y$, ale byłoby to dość czasochłonne. Zastanawiam się, czy moglibyśmy pójść na skróty.



Dodaj $E(X)$ do $E(Y)$, by uzyskać $E(X+Y)$...

Chcemy znaleźć wartość oczekiwaną i wariancję wypłaty przy założeniu, że gramy raz na automacie starego typu i raz na automacie nowego typu. Innymi słowy, interesuje nas wartość $E(X+Y)$ i $\text{Var}(X+Y)$, gdzie X i Y są zmiennymi losowymi opisującymi grę na starych i nowych urządzeniach. Są to zmienne niezależne.

W tym celu moglibyśmy wyznaczyć rozkład prawdopodobieństwa sumy $X+Y$, a następnie znaleźć jego wartość oczekiwaną i wariancję.



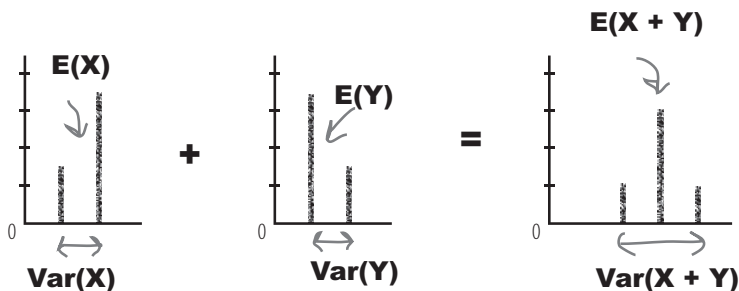
Na szczęście nie musimy tego robić. Aby znaleźć $E(X+Y)$, możemy po prostu dodać do siebie $E(X)$ i $E(Y)$.

$$E(X + Y) = E(X) + E(Y)$$

Przyznasz sam, że ma to sens. Gdybyś grał w dwie gry, przy czym w pierwszej oczekiwałbyś wygranej w wysokości 5 dolarów, a w drugiej 10 dolarów, łącznie spodziewałbyś się wygrać $5+10 = 15$ dolarów.

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

Podobnie możemy postąpić w przypadku wariancji. Aby znaleźć $\text{Var}(X+Y)$, możemy dodać do siebie wariancje obu zmiennych. Jednak tylko wtedy, gdy X i Y są zmiennymi niezależnymi.



Wariancja się zwiększa — rozkład prawdopodobieństwa jest bardziej rozproszony.



Uwaga!

Swobodnie można dodawać tylko wariancje zmiennych niezależnych.

Jeśli X i Y nie są zmiennymi niezależnymi, wtedy $\text{Var}(X+Y)$ nie jest równa $\text{Var}(X) + \text{Var}(Y)$.

...lub odejmij $E(Y)$ od $E(X)$, by uzyskać $E(X - Y)$

W podobny sposób możemy również wyznaczyć parametry zmiennej będącej różnicą dwóch zmiennych losowych, czyli $X - Y$.

Równie łatwo jak w przypadku sumy zmiennych można wyznaczyć $E(X - Y)$. Wystarczy odjąć $E(Y)$ od $E(X)$.

Formuła na wariancję jest nieco mniej intuicyjna. Aby znaleźć $\text{Var}(X - Y)$, musimy... dodać do siebie $\text{Var}(X)$ i $\text{Var}(Y)$.

$$E(X - Y) = E(X) - E(Y)$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$$

Tutaj wariancje sumujemy, więc bądź ostrożny!



Uwaga!

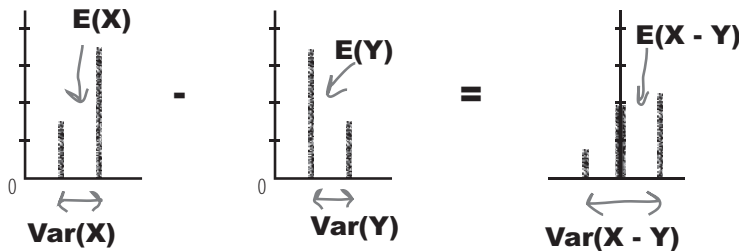
Wyznaczając wariancję różnicy zmiennych losowych, zsumuj ich wariancję.

Łatwo tu o pomyłkę, ponieważ w pierwszej chwili wydaje się to sprzeczne z intuicją. Zapamiętaj jednak, że jeśli X i Y są niezależnymi zmiennymi losowymi, to $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$.

Przecież to nie ma sensu. Dlaczego mielibyśmy dodawać wariancje?

Ponieważ w tym przypadku zmienność również się zwiększa.

Chociaż odejmujemy od siebie dwie zmienne losowe, to wariancja rozkładu wynikowego nie zmniejsza się, ale zwiększa.



Wariancja zwiększa się, choć odjęliśmy zmienne.

Wariancja zmiennej będącej różnicą dwóch niezależnych zmiennych losowych jest dokładnie taka sama, jak wariancja zmiennej będącej ich sumą. W obu przypadkach zmienność może jedynie wzrosnąć.

Odjęcie od siebie dwóch zmiennych losowych zwiększa wariancję.

Podobne operacje możesz wykonywać na zmiennych przekształconych liniowo

To jeszcze nie wszystko. Oprócz dodawania i odejmowania samych zmiennych losowych te same operacje można wykonywać na ich przekształceniach liniowych.

Zastanów się, co by się stało, gdyby właściciel kasyna zmienił stawki opłat i wygranych w maszynach starego i nowego typu. Ostatnią rzeczą, jaką chcielibyśmy robić, to tworzyć od początku jeszcze jeden rozkład prawdopodobieństwa i wyznaczać jego parametry.

Na szczęście istnieje wygodna droga na skróty.

Załóżmy, że Dan zmienił stawki wygranych na wszystkich maszynach dostępnych w kasynie, tak że wypłata z maszyn starego typu zmieniła się z X na aX , zaś z Y na bY , gdzie a i b są ustalonymi wartościami.

Aby znaleźć wartość oczekiwaną i wariancję liniowej kombinacji aX i bY , możemy postąpić w opisany niżej sposób.

$$X \rightarrow aX$$

$$Y \rightarrow bY$$

a i b mogą być dowolnymi liczbami.

Sumujemy aX i bY

Wartość oczekiwaną i wariancję zmiennej $aX + bY$ znajdziemy, korzystając ze wzorów:

$$E(aX + bY) = aE(X) + bE(Y)$$

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

W formule na wariancję występują kwadraty, ponieważ mamy tutaj do czynienia z przekształceniem liniowym.

Jest to rodzaj przekształcenia liniowego, stąd te kwadraty.

Odejmujemy aX i bY

Jeśli chcemy znaleźć $E(X-Y)$ lub $\text{Var}(X-Y)$, możemy skorzystać z poniższych formuł:

$$E(aX - bY) = aE(X) - bE(Y)$$

$$\text{Var}(aX - bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

Tak jak poprzednio, dodajemy wariancje obu zmiennych, choć szukamy wariancji ich różnicy.

Pamiętaj, że wariancje dodajemy.

Nie istnieją
głupie pytania

P: Skoro X i Y oznaczają wypłaty w dwóch grach, to czy $aX + bY$ oznacza łączną wygraną w a grach X i b grach Y ?

U: Zapis $aX + bY$ przedstawia właściwie sumę dwóch przekształceń liniowych zmiennych X i Y . Innymi słowy, zmianie ulegają zarówno wartości zmiennej X , jak i Y .

P: Nie wiem, czy kiedykolwiek przyda mi się znajomość rozkładu $X - Y$. Czy ma on w ogóle jakiś sens?

U: $X - Y$ przydaje się wówczas, gdy interesują nas różnice między dwiema wielkościami. $E(X - Y)$ daje odpowiedź na pytanie: „Jakiej różnicy między X i Y oczekujesz?”. $\text{Var}(X - Y)$ informuje o wariancji takiej różnicy.

P: Dlaczego przy liczeniu $\text{Var}(X - Y)$ musimy dodać wariancję zmiennych X i Y ? Czy nie powinniśmy ich odjąć?

U: Choć z pozoru może się to wydawać sprzeczne z intuicją, to jednak odejmując od siebie dwie zmienne losowe, zwiększamy zmienność wartości zmiennej wynikowej, dlatego jej wariancja rośnie. Wariancja sumy dwóch zmiennych niezależnych jest taka sama jak wariancja ich różnicy.

Można to wyjaśnić jeszcze inaczej: przy liczeniu wariancji podnosimy do kwadratu wartości zmiennych losowych. $\text{Var}(X + bY)$ jest równa $\text{Var}(X) + b^2\text{Var}(Y)$. Podstawiając $b = -1$, otrzymamy szukaną $\text{Var}(X - Y)$. Ponieważ $(-1)^2 = 1$, zatem $\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$.

P: Czy te same obliczenia możemy wykonywać dla zmiennych, które nie są niezależne?

U: Nie, formuły dla wariancji zachowują ważność tylko w przypadku zmiennych niezależnych. Jeśli chcesz wyznaczyć $\text{Var}(X + Y)$ dla zmiennych, które są zależne, musisz najpierw wyznaczyć rozkład prawdopodobieństwa zmiennej $X + Y$.

P: Wygląda na to, że te same zasady, które stosują się do $X + Y$, są również prawdziwe w przypadku $X_1 + X_2$. Mam rację?

U: Oczywiście. Możesz je stosować do wszystkich zmiennych losowych, pod warunkiem że są one niezależne.



CELNE SPOSTRZEŻENIA

■ Przez **niezależne obserwacje zmiennej X** rozumieć należy kolejne realizacje zmiennej X . Można je opisać oddzielnymi zmiennymi losowymi, z których każda ma ten sam rozkład prawdopodobieństwa co X .

■ Jeśli $X_1, X_2, X_3, \dots, X_n$ są zmiennymi niezależnymi o rozkładzie takim jak X , wówczas:

$$E(X_1 + X_2 + \dots + X_n) = nE(X)$$

$$\text{Var}(X_1 + X_2 + \dots + X_n) = n\text{Var}(X)$$

■ Jeśli X i Y są niezależnymi zmiennymi losowymi, prawdziwe są wzory:

$$E(X + Y) = E(X) + E(Y)$$

$$E(X - Y) = E(X) - E(Y)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$$

$$\text{Var}(X - Y) = \text{Var}(X) + \text{Var}(Y)$$

■ Wartość oczekiwaną i wariancję liniowych funkcji zmiennych losowych X i Y można wyznaczyć ze wzorów:

$$E(aX + bY) = aE(X) + bE(Y)$$

$$E(aX - bY) = aE(X) - bE(Y)$$

$$\text{Var}(aX + bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

$$\text{Var}(aX - bY) = a^2\text{Var}(X) + b^2\text{Var}(Y)$$

Ćwiczenie



Ćwiczenie

W poniższej tabeli zamieszczono wartości oczekiwane i wariancje różnych zmiennych losowych. Spróbuj podać najprostszy sposób na wyznaczenie każdej z tych wartości. W razie potrzeby przyjmij założenie o niezależności zmiennych losowych.

Parametr	Sposób obliczenia
$E(aX + b)$	
$\text{Var}(aX + b)$	
$E(X)$	
$E(f(X))$	
$\text{Var}(aX - bY)$	
$\text{Var}(X)$	
$E(aX - bY)$	
$E(X_1 + X_2 + X_3)$	
$\text{Var}(X_1 + X_2 + X_3)$	
$E(X^2)$	
$\text{Var}(aX - b)$	



Ćwiczenie

Pewna restauracja oferuje dwa menu: jedno przeznaczone na dni robocze, drugie ważne w weekendy. Każde z nich zawiera potrawy w czterech różnych kategoriach cenowych. Rozkłady prawdopodobieństw wydatków klientów zamieszczono w poniższych tabelkach:

Dni robocze:

x	10	15	20	25
P(X = x)	0,2	0,5	0,2	0,1

Weekend:

y	15	20	25	30
P(Y = y)	0,15	0,6	0,2	0,05

Kto, Twoim zdaniem, wyda więcej pieniędzy na posiłek w tej restauracji: grupa 20 klientów weekendowych czy 25 klientów odwiedzających restaurację w pozostałe dni?

Ćwiczenie: Rozwiązanie



Ćwiczenie Rozwiązanie

W poniższej tabeli zamieszczono wartości oczekiwane i wariancje różnych zmiennych losowych. Spróbuj podać najprostszy sposób na wyznaczenie każdej z tych wartości. W razie potrzeby przyjmij założenie o niezależności zmiennych losowych.

Parametr	Sposób obliczenia
$E(aX + b)$	$aE(X) + b$
$\text{Var}(aX + b)$	$a^2 \text{Var}(X)$
$E(X)$	$\sum x P(X = x)$
$E(f(X))$	$\sum f(x) P(X = x)$
$\text{Var}(aX - bY)$	$a^2 \text{Var}(X) + b^2 \text{Var}(Y)$
$\text{Var}(X)$	$E(X - \mu)^2 = E(X^2) - \mu^2$
$E(aX - bY)$	$aE(X) - bE(Y)$
$E(X_1 + X_2 + X_3)$	$3E(X)$
$\text{Var}(X_1 + X_2 + X_3)$	$3\text{Var}(X)$
$E(X^2)$	$\sum x^2 P(X = x)$
$\text{Var}(aX - b)$	$a^2 \text{Var}(X)$



Ćwiczenie Rozwiązanie

Pewna restauracja oferuje dwa menu: jedno przeznaczone na dni robocze, drugie ważne w weekendy. Każde z nich zawiera potrawy w czterech różnych kategoriach cenowych. Rozkłady prawdopodobieństw wydatków klientów zamieszczono w poniższych tabelkach:

Dni robocze:

x	10	15	20	25
P(X = x)	0,2	0,5	0,2	0,1

Weekend:

y	15	20	25	30
P(Y = y)	0,15	0,6	0,2	0,05

Kto, Twoim zdaniem, wyda więcej pieniędzy na posiłek w tej restauracji: grupa 20 klientów weekendowych czy 25 klientów odwiedzających restaurację w pozostałe dni?

Zacznijmy od wyznaczenia wartości oczekiwanej kwoty, jaką wyda przeciętny klient odwiedzający restaurację w weekend (Y) oraz w pozostałe dni (X).

$$\begin{aligned} E(X) &= 10 \times 0,2 + 15 \times 0,5 + 20 \times 0,2 + 25 \times 0,1 = \\ &= 2 + 7,5 + 4 + 2,5 = \\ &= 16 \end{aligned}$$

$$\begin{aligned} E(Y) &= 15 \times 0,15 + 20 \times 0,6 + 25 \times 0,2 + 30 \times 0,05 = \\ &= 2,25 + 12 + 5 + 1,5 = \\ &= 20,75 \end{aligned}$$

Wydatek każdego klienta można opisać za pomocą niezależnej obserwacji. Aby wyznaczyć łączne wydatki klientów zaliczonych do obu grup, pomnożymy $E(X)$ i $E(Y)$ przez liczebność każdej grupy.

$$25 \text{ klientów w dni robocze wyda razem: } 25 \times E(X) = 25 \times 16 = 400$$

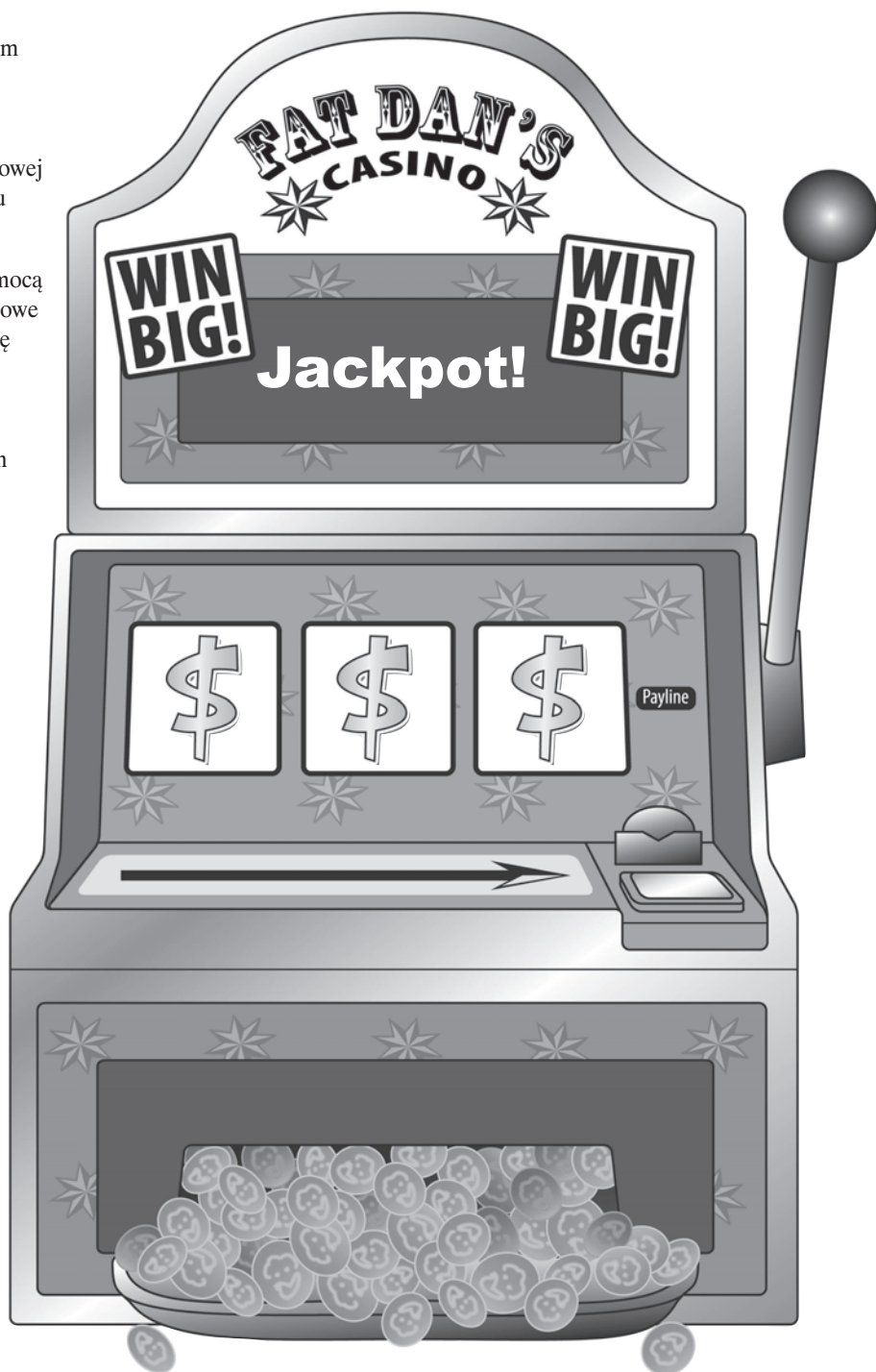
$$20 \text{ klientów weekendowych wyda zaś: } 20 \times E(Y) = 20 \times 20,75 = 415$$

Możemy więc oczekiwać, że 20 klientów odwiedzających restaurację w ciągu weekendu wyda więcej niż 25 klientów odwiedzających ją w tygodniu.

Rozbiłeś bank!

W tym rozdziale poruszyliśmy całkiem sporo zagadnień. Dowiedziałeś się, jak można posłużyć się rozkładem prawdopodobieństwa, wartością oczekiwaną i wariancją zmiennej losowej do przewidywania wygranej w starciu z „jednorożkim bandytą”.

Dowiedziałeś się również, jak za pomocą przekształceń liniowych wyznaczyć nowe wartości parametrów, gdy zmienia się profil wypłaty w grze. Dzięki wiedzy o tym, czym są zmienne niezależne, umiesz już wyznaczać parametry rozkładu w przypadku gier złożonych z wielu kolejek.





Ćwiczenie

Sam zwykł jadać w dwóch restauracjach. Restauracja A jest droższa niż B, ale serwuje znacznie lepsze jakościowo dania.

Poniższe tabelki prezentują rozkłady prawdopodobieństw wydatków Sama w obu restauracjach. Jak mógłbyś scharakteryzować różnicę w poziomie cen między restauracją A i B? Ile wynosi jej wariancja?

Restauracja A:	x	20	30	40	45
	P(X = x)	0,3	0,4	0,2	0,1

Restauracja B:	y	10	15	18
	P(Y = y)	0,2	0,6	0,2

Ćwiczenie: Rozwiązanie



Ćwiczenie Rozwiązanie

Sam zwykł jadać w dwóch restauracjach. Restauracja A jest droższa niż B, ale serwuje znacznie lepsze jakościowo dania.

Poniższe tabelki prezentują rozkłady prawdopodobieństw wydatków Sama w obu restauracjach. Jak mógłbyś scharakteryzować różnicę w poziomie cen między restauracją A i B? Ile wynosi jej wariancja?

Restauracja A:	x	20	30	40	45
	P(X = x)	0,3	0,4	0,2	0,1

Restauracja B:	y	10	15	18
	P(Y = y)	0,2	0,6	0,2

Wyznamy najpierw parametry rozkładów zmiennych X i Y :

$$\begin{aligned}E(X) &= 20 \times 0,3 + 30 \times 0,4 + 40 \times 0,2 + 45 \times 0,1 = \\ &= 6 + 12 + 8 + 4,5 = \\ &= 30,5\end{aligned}$$

$$\begin{aligned}\text{Var}(X) &= (20 - 30,5)^2 \times 0,3 + (30 - 30,5)^2 \times 0,4 + \\ &\quad + (40 - 30,5)^2 \times 0,2 + (45 - 30,5)^2 \times 0,1 = \\ &= (-10,5)^2 \times 0,3 + (-0,5)^2 \times 0,4 + 9,5^2 \times 0,2 + 14,5^2 \times 0,1 = \\ &= 110,25 \times 0,3 + 0,25 \times 0,4 + 90,25 \times 0,2 + 210,25 \times 0,1 = \\ &= 33,075 + 0,1 + 18,05 + 21,025 = \\ &= 72,25\end{aligned}$$

$$\begin{aligned}E(Y) &= 10 \times 0,2 + 15 \times 0,6 + 18 \times 0,2 = \\ &= 2 + 9 + 3,6 = \\ &= 14,6\end{aligned}$$

$$\begin{aligned}\text{Var}(Y) &= (10 - 14,6)^2 \times 0,2 + (15 - 14,6)^2 \times 0,6 + \\ &\quad + (18 - 14,6)^2 \times 0,2 = \\ &= (-4,6)^2 \times 0,2 + 0,4^2 \times 0,6 + 3,4^2 \times 0,2 = \\ &= 21,16 \times 0,2 + 0,16 \times 0,6 + 11,56 \times 0,2 = \\ &= 4,232 + 0,096 + 2,312 = \\ &= 6,64\end{aligned}$$

Różnicę między poziomami cen w obu restauracjach można przedstawić jako $X - Y$:

$$\begin{aligned}E(X - Y) &= E(X) - E(Y) = \\ &= 30,5 - 14,6 = \\ &= 15,9\end{aligned}$$

$$\begin{aligned}\text{Var}(X - Y) &= \text{Var}(X) + \text{Var}(Y) = \\ &= 72,25 + 6,64 = \\ &= 78,89\end{aligned}$$