

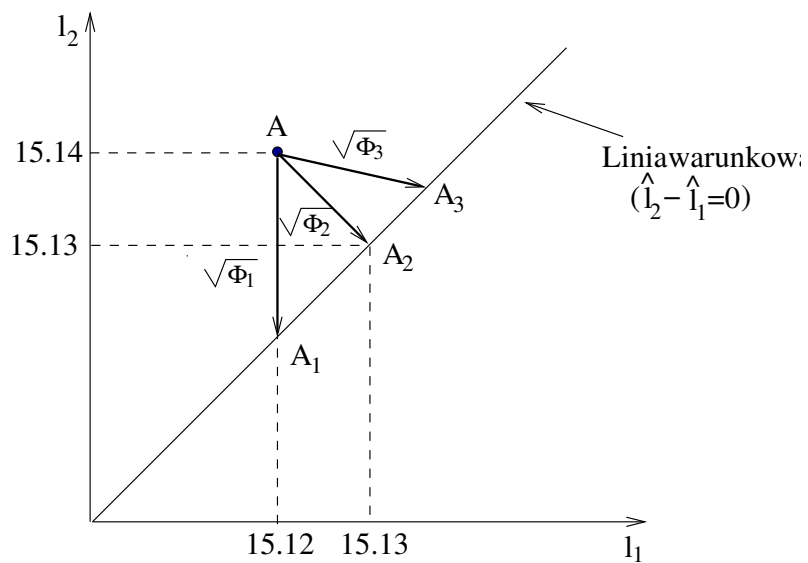
Wykład

METODY OPRACOWANIA OBSERWACJI

I rok studiów astronomii I-stopnia

Kompilacja i opracowanie *Tadeusz J. Jopek*

(Ad Usum Internum)



Poznań AD 8 maja 2016

Spis treści

1	Wstęp	5
1.1	Wielkości fizyczne i ich miary	6
1.2	Międzynarodowy Układ Jednostek SI	6
1.3	Pomiary wielkości fizycznych	7
1.4	Model obserwacji	9
2	Opracowanie obserwacji, koncepcja	12
2.1	Niepewności, błędy pomiarów, obserwacji	13
2.2	Ocena niepewności pomiarowych	14
2.2.1	Metoda A	15
2.2.2	Metoda B	16
2.3	Wyniki końcowe, obliczanie końcowej niepewności	18
2.4	Zaokrąglanie wyników pomiarów i ich zapis	19
2.4.1	Reguły zaokrąglania liczb przybliżonych	19
2.4.2	Zaokrąglanie i zapis wyników pomiarów	21
2.5	Ocena niepewności w pomiarach pośrednich	22
2.5.1	Pomiary pośrednie nieskorelowane	23
2.5.2	Pomiary pośrednie skorelowane	23
2.6	Statystyczna koncepcja rachunku wyrównawczego	24
2.7	Metoda najmniejszych kwadratów	25
2.8	Dygresja matematyczna: propagacja niepewności, linearyzacja funkcji	28
2.8.1	Przypadek jednowymiarowy	28
2.8.2	Przypadek wielowymiarowy	29
3	Prawdopodobieństwo, rozkłady, parametry	32
3.1	Wstęp	33
3.2	Prawdopodobieństwo 'definicje'	33
3.2.1	Prawdopodobieństwo matematyczne	33
3.2.2	Prawdopodobieństwo geometryczne	34
3.2.3	Prawdopodobieństwo częstościowe	34
3.3	Aksjomatyczne ujęcie prawdopodobieństwa	35
3.4	Własności prawdopodobieństwa	36
3.4.1	Prawdopodobieństwo sumy zdarzeń	37
3.4.2	Prawdopodobieństwo warunkowe	38
3.4.3	Niezależność zdarzeń	39
3.5	Zmienna losowa	40
3.5.1	Dystrybuanta	41
3.5.2	Funkcja gęstości	41
3.6	Rozkłady wielowymiarowe	42
3.6.1	Rozkłady brzegowe	44
3.6.2	Niezależność zmiennych losowych	45
3.6.3	Rozkłady warunkowe	45
3.7	Parametry rozkładu: nadzieja matematyczna, momenty i korelacja	46
3.7.1	Wartość oczekiwana	46
3.7.2	Wariancja	47
3.7.3	Kowariancja	48
3.7.4	Momenty	49

3.8	Modalna, mediana i kwantyle	51
3.8.1	Wartość modalna	51
3.8.2	Mediana	51
3.8.3	Kwantyle	52
3.9	Dygresja: odwracanie dystrybuanty	53
3.10	Dygresja: nierówność Czebyszewa	53
3.11	Zadanka na ćwiczenia	55
4	Przegląd rozkładów, obliczanie prawdopodobieństwa	56
4.1	Rozkłady Bernoulliego	57
4.1.1	Rozkład dwumianowy	57
4.1.2	Rozkład geometryczny	59
4.1.3	Rozkład wielomianowy	59
4.2	Prawo wielkich liczb	60
4.3	Rozkład hipergeometryczny	61
4.4	Rozkład Poissona	63
4.5	Rozkład jednostajny (jednorodny)	66
4.6	Dygresja: funkcja charakterystyczna rozkładu prawdopodobieństwa	68
4.7	Rozkład normalny Gaussa	70
4.8	Rozkład t studenta	72
4.9	Rozkład chi-kwadrat	73
4.10	Rozkład F Fishera-Snedecora	75
4.11	Rozkład normalny wielowymiarowy	76
4.11.1	Rozkład normalny dwuwymiarowy	76
4.11.2	Elipsy stałego prawdopodobieństwa	78
4.11.3	Rozkład 3-wymiarowy, elipsoidy stałego prawdopodobieństwa	81
4.12	Podsumowanie	82
4.13	Zadanka na ćwiczenia	82
5	Estymacja parametrów populacji	83
5.1	Wstęp	84
5.2	Sposoby opisu próby statystycznej	84
5.2.1	Histogramy, stereogramy	85
5.2.2	Statystyki z próby	89
5.2.2.1	Wartości centralne, miary położenia	89
5.2.2.2	Miary rozproszenia, rozrzutu	91
5.3	Estymacja punktowa	91
5.4	Metody estymacji punktowej	94
5.4.1	Metoda momentów	94
5.4.2	Metoda największej wiarygodności	94
5.4.3	Metoda najmniejszych kwadratów	96
5.5	Estymacja przedziałowa	96
5.5.1	Przedział ufności dla średniej	98
5.5.2	Przedział ufności dla wariancji	100
5.5.3	Przedziały ufności dla stosunku wariancji	101
5.6	Zadanka na ćwiczenia	101
6	Testowanie hipotez statystycznych	102
6.1	Wstęp	103
6.2	Hipotezy dotyczące zjawisk a hipotezy statystyczne	103
6.3	Hipoteza zerowa i alternatywna	105
6.4	Błędy pierwszego i drugiego rodzaju	105
6.5	Elementy statystycznego testu hipotezy zerowej	106
6.5.1	Wyznaczenie obszaru krytycznego	107
6.5.2	Wnioskowanie statystyczne	108
6.6	Przykłady weryfikacji hipotez, testy parametryczne	109
6.6.1	Schemat konstrukcji testu parametrycznego	109
6.6.2	Test parametryczny dla średniej z próby przy znanej wariancji	110
6.6.3	Test dla średniej z próby przy nieznannej wariancji	111
6.6.4	Test dla dwóch średnich z próby	112

6.6.5	Weryfikacja hipotezy dla wariancji	112
6.6.6	Porównanie dwóch wariancji	113
6.7	Moc testu	114
6.8	Zadanka na ćwiczenia	116
7	Koncepcja wag pomiarowych	117
7.1	Koncepcja wagi pomiarowej	118
7.2	Obserwacje zrównoważone	119
8	Podstawy techniki propagacji	120
8.1	Wstęp	121
8.2	Propagacja funkcji gęstości rozkładu prawdopodobieństwa	121
8.2.1	Przypadek wielowymiarowy	122
8.3	Propagacja średniej	123
8.3.1	Propagacja średnich dla funkcji liniowych	123
8.4	Propagacja wariancji i kowariancji	124
8.4.1	Propagacja wariancji i kowariancji dla funkcji liniowych	125
8.4.2	Propagacja wariancji i kowariancji dla funkcji nieliniowych	126
8.5	Propagacja w przypadku blokowych wektorów losowych	127
8.6	Dygresja - mnożenie symboliczne	128
8.7	Propagacja złożona z zastosowaniem macierzy	130
8.8	Propagacja błędów systematycznych i prawdziwych	131
9	Metoda najmniejszych kwadratów	133
9.1	Wstęp	134
9.2	Techniki pośredniego i bezpośredniego wyrównywania obserwacji	134
9.2.1	Bezpośrednie wyrównywanie obserwacji	134
9.2.2	Pośrednie wyrównywanie obserwacji	136
9.3	Równania warunkowe: notacja macierzowa	138
9.3.1	Technika pośrednia	138
9.3.2	Technika bezpośrednia	140
9.4	Macierzowy warunek najmniejszych kwadratów	141
9.4.1	Koncepcja wag pomiarów	141
9.4.2	Obserwacje zrównoważone	142
9.4.3	Macierzowa postać warunku najmniejszych kwadratów	142
9.5	Równania normalne: notacja macierzowa	143
9.5.1	Technika pośrednia	143
9.5.2	Technika bezpośrednia	145
9.6	Wyrównywanie z warunkami: przypadek ogólny	150
9.6.1	Techniki wyrównywania z warunkami	151
9.6.2	Równania normalne techniki wyrównywania z warunkami	152
9.6.2.1	Dygresja: oszacowanie wariancji odniesienia	155
9.6.3	Oszacowanie precyzji rezultatów wyrównania	155
9.6.4	Zależność funkcjonalna i posteriori macierze kofaktorowe	157
9.7	Interpretacja geometryczna metody NK	157
9.8	Wyrównanie z warunkami, przypadki specjalne	160
9.8.1	Wyrównywanie tylko obserwacji	161
9.8.2	Wyrównywanie obserwacji i parametrów	162
9.9	Dygresja: transformacja współrzędnych	163
9.9.1	Transformacje dwuwymiarowe	164
9.9.2	Transformacje trójwymiarowe	165
9.10	Przykład	165
9.10.1	Ogólny przypadek wyrównywania z warunkami	165
9.11	Dodatek A. Zadania	170

Zakończenie pracy

Rozdział 1

Wstęp

Streszczenie

Nie ma komu napisać I nie wiadomo czy będzie komu.

Słowa kluczowe: Wielkości fizyczne, jednostki miary, układ jednostek SI, pomiar, pomiar bezpośredni i pośredni, obserwacja, model funkcjonalny, model stochastyczny ^a ↗

^a[Modyfikowano AD 2011, Luty 08.]

1.1 Wielkości fizyczne i ich miary

Przedmiotem badań fizyki, astronomii jest występująca w różnej postaci i formie materia oraz generowane przez nią oddziaływania. Badania naukowe polegają na rozpoznaniu i ustaleniu cech fizycznych materii jako takiej, obiektów składających się z materii a także sposobu oddziaływania tych obiektów między sobą. W zakres zainteresowań fizyki i astronomii wchodzi jedynie to co dotyczy obiektów materialnych, co daje się badać metodami fizycznymi.

Cechy fizyczne obiektów materialnych jak i cechy zachodzących między nimi oddziaływań, zwykle nazywane są *wielkościami fizycznymi*, np.: masa, gęstość, szybkość, czas, napięcie elektryczne, siła Umownie, dowolną wielkość fizyczną będziemy oznaczali np. literą x .

Fizyka i astronomia interesują jedynie te wielkości, które dają się mierzyć, czyli takie wielkości fizyczne x , którym potrafią przypisać pewną liczbę — liczbową miarę, oznaczymy ją symbolem x .

Ponieważ miary x tej samej wielkości fizycznej x mogą być wyrażone w różny sposób, np. w różnych jednostkach, liczbie x musi towarzyszyć informacja o jednostkach w jakich została podana. Bez tej dodatkowej informacji moglibyśmy popaść w błędne przekonania porównując ze sobą masy tego samego obiektu zmierzone np. w kilogramach i tonach.

Dlatego pełna informacja o wielkości fizycznej x musi składać się z jej liczbowej miary x oraz jednostki $[x]$. Co zapisujemy zwięźle jako

$$x = x [x] \tag{1.1}$$

Np. Temperatura $T = 273.15 [K]$ ↗

1.2 Międzynarodowy Układ Jednostek SI

W celu ułatwienia współpracy naukowo technicznej, miary wielkości fizycznych należy podawać zgodnie z uzgodnionymi w skali światowej standardami. Opracowano różne standardy, z których najpowszechniej stosowanym jest Międzynarodowy Układ Jednostek SI (SI od francuskiego Le Système international d'unités). Instytucją uprawnioną do tworzenia, modyfikacji i ogłaszania Międzynarodowego Układu Jednostek SI jest **Międzynarodowe Biuro Wag i Miar w Paryżu** (BIPM).

Układ SI opracowano w roku 1960 w oparciu o starszy układ jednostek MKS (metr-kilogram-sekunda). Inne dawne układy jednostek jak układ cgs (centymetr-gram-sekunda) czy jego warianty nie odegrały większej roli w tworzeniu układu SI.

W układzie SI wprowadzono szereg nowych nazw jednostek miar, podano ich nowe definicje, ale nie oznacza to, że dzisiaj układ SI jest układem statycznym — jego jednostki nadal są tworzone i definiowane drogą międzynarodowej ugody w miarę postępu technologicznego.

Tablica 1.1: Jednostki podstawowe Układu SI.

Nazwa jednostki	Symbol jednostki	Wielkość fizyczna
metr	m	długość
kilogram	kg	masa
sekunda	s	czas
amper	A	natężenie prądu elektrycznego
kelwin	K	temperatura termodynamiczna
mol	mol	ilość substancji
kandela	cd	światłość (kierunkowa)

Międzynarodowy Układ SI obejmuje zbiór jednostek, ich definicje, nazwy oraz zbiór przedrostków do nazw jednostek. Zbiór jednostek dzieli się na dwa podzbiory:

- zbiór siedmiu jednostek podstawowych, nominalnie wymiarowo od siebie niezależnych (patrz tabela 1.1),
- zbiór jednostek pochodnych, definiowanych za pomocą jednostek podstawowych (tabela 1.2),

W praktyce metrologicznej napotykamy jeszcze dodatkowe jednostki nie należące formalnie do układu SI, ale których wykorzystanie wraz z jednostkami SI jest akceptowane.

Obok jednostek podstawowych, w układzie SI występują dwie jednostki bezwymiarowe, powstałe w wyniku podzielenia dwóch tych samych podstawowych jednostek SI i dlatego traktowanych przez BIPM jako jednostki pochodne. Formalnie ich jednostką SI są po prostu liczby 1, posiadają jednak nazwy by uniknąć nieporozumień w wypadku napotkania jedynie liczb bezwymiarowych.

W celu ułatwienia zapisu, do każdej nazwy jednostki można dodać przedrostki będące całkowitymi wielokrotnościami liczby 10. Np. przedrostek *kilo* oznacza mnożnik 1000, przedrostek *mili* oznacza mnożnik 0.001, stąd jednostka *milimetr* oznacza jedną tysięczną metra, *kilometr* oznacza tysiąc metrów. Przedrostki nie mogą być łączone, np. zamiast jedna milionowa kilograma możemy powiedzieć miligram a nie mikrokilogram. W tabeli 1.3 podano przedrostki akceptowane w Układzie Jednostek SI. ↗

1.3 Pomiary wielkości fizycznych

Słowa pomiar, obserwacja będą dalej traktowane jako posiadające to samo znaczenie. Mogą oznaczać zarówno ciąg operacji jak i jego rezultat. Ale znaczenie słowa pomiar jest szersze, bowiem wyniki procedur opracowania (wyrównywania) bezpośrednich pomiarów, w szczególności rezultaty numeryczne uzyskane w oparciu o wielkości zmierzone, także nazywane są pomiarami, obserwacjami. I właśnie takie obserwacje bezpośrednie jak i pośrednie (opracowane, wyrównane, przetworzone dane pomiarowe) mają fundamentalne znaczenie dla nauki i techniki — poznajemy otaczający nas świat poprzez pomiary, obserwacje wielkości charakteryzujące materię i zjawiska.

Tablica 1.2: Wybrane jednostki pochodne Układu SI posiadające własne nazwy.

Nazwa jednostki	Symbol jednostki	Wielkość fizyczna	Objaśnienie	
radian	rad	kąt	Jednostką kąta jest kat o wierzchołku w środku okręgu, rozpięty na łuku okręgu o długości równej promieniowi okręgu. Na pełnym okręgu rozpięty jest kąt 2π rad.	
steradian	sr	kąt bryłowy	Jednostką kąta bryłowego, jest kąt bryłowy o wierzchołku w środku sfery o promieniu r , rozpięty na części sfery o powierzchni r^2 . Na całej sferze rozpięty jest kąt 4π sr.	
		Wielkość fizyczna	Określenie za pomocą innych jednostek SI	Określenie za pomocą podst. jednostek SI
herc	Hz	częstotliwość, częstość	$1/s$	s^{-1}
newton	N	siła, ciężar	$mkgs^{-2}$	$mkgs^{-2}$
pascal	Pa	ciśnienie, naprężenie	N/m^2	$m^{-1}kgs^{-2}$
dżul	J	energia, praca, ciepło	Nm	m^2kgs^{-2}
watt	W	moc, moc promieniowania	J/s	m^2kgs^{-3}
coulomb	C	ładunek elektryczny, strumień elektryczny	sA	sA
volt	V	napięcie elektryczne, różnica potencjału, siła elektromotoryczna,	$W/A = J/C$	$m^2kgs^{-3}A^{-1}$
farad	F	pojemność elektryczna	C/V	$m^{-2}kg^{-1}s^4A^2$
ohm	Ω	opór elektryczny, impedancja, reaktancja	V/A	$m^2kgs^{-3}A^{-2}$
siemens	S	przewodność elektryczna	$1/\Omega$	$m^{-2}kg^{-1}s^3A^2$
weber	Wb	strumień indukcji magnetycznej	J/A	$m^2kgs^{-2}A^{-1}$
tesla	T	indukcja magnetyczna	$Vs/m^2 =$ $= Wb/m^2 = N/Am$	$kgs^{-2}A^{-1}$
henr	H	indukcyjność magnetyczna współczynnik indukcji własnej	$Vs/A = Wb/A$	$m^2kgs^{-2}A^{-2}$
Celsius	$^{\circ}C$	temperatura Celsjusza	$t_{^{\circ}C} = t_K - 273.15$	K
lumen	lm	strumień świetlny	$cdsr$	cd
lux	lx	natężenie oświetlenia	lm/m^2	$m^{-2}cd$
becquerel	Bq	radioaktywność (rozpad na jednostkę czasu)	$1/s$	s^{-1}
gray	Gy	dawka zaabsorbowana promieniowania jonizującego	J/kg	m^2s^{-2}
sievert	Sv	dawka równoważna promieniowania jonizującego	J/kg	m^2s^{-2}
katal	kat	aktywność katalitikczna	mol/s	$s^{-1}mol$

Tablica 1.3: Przedrostki akceptowane w Międzynarodowym Układzie Jednostek SI.

Nazwa	jotta	zetta	eksa	peta	tera	giga	mega	kilo	hekto	deka
Symbol	Y	Z	E	P	T	G	M	k	h	da
Mnożnik	10^{24}	10^{21}	10^{18}	10^{15}	10^{12}	10^9	10^6	10^3	10^2	10^1
Nazwa	decy	centy	mili	micro	nano	pico	femto	atto	zepto	jokto
Symbol	d	c	m	μ	n	p	f	a	z	y
Mnożnik	10^{-1}	10^{-2}	10^{-3}	10^{-6}	10^{-9}	10^{-12}	10^{-15}	10^{-18}	10^{-21}	10^{-24}

Nauką zajmującą się pomiarami jest **metrologia**. Przedmiotem metrologii są pomiary użytkowe (technika, handel ...) ale także pomiary o charakterze naukowym. Nie będziemy wkraczali na teren tej nauki, dla naszych potrzeb, ograniczymy się jedynie do prostego przykładu ukazującego, że z wykonywaniem pomiarów, obserwacji wiążą się niekiedy bardzo głębokie i trudne problemy.

Potoczne rozumienie koncepcji pomiaru nie stwarza problemów; gdy chcemy zmierzyć np. długość stołu, zgadzamy się, że w tym celu trzeba wziąć miarkę i zmierzyć co trzeba. Jednak uważniejsza analiza tego co rozumiemy przez pomiar długości prowadzi do następujących wniosków:

1. numeryczny odczyt, dokonany w efekcie pewnego procesu reprezentuje rezultat pomiaru,
2. pomiary wykonywane są poprzez odniesienie się do wzorców (wzorzec metra, ...), najczęściej za pomocą aparatów, przyrządów, w których wykorzystano odpowiednie zjawiska fizyczne (może poza przypadkami prostego zliczania występowania zjawisk),
3. mierzenie oznacza wykonanie ciągu fizycznych operacji, które mogą składać się także z ciągu operacji jak: ustawienie instrumentu, kalibracja, przykładanie przyrządu, nastawianie, porównywanie, etc.,
4. mierzenie dokonywane jest przez porównanie ze standardami, te zaś ustalane są mocą konwencji, nierzadko arbitralnej (jednostki, wymiar),
5. głębsze wniknięcie w pojęcie obserwacji ujawnia, że nierzadko obserwacje odnoszone są do teoretycznych koncepcji nie posiadających odpowiedników w świecie fizycznym. Np. są nimi abstrakcje geometryczne napotymane przy pomiarach odległości, kątów etc. Wybieramy daną koncepcję w celu opisu tych fragmentów natury, które są dla nas interesujące, jak np. koncepcje: położenie gwiazdy, jasność gwiazdy, komety,
6. wykonywanie pomiarów jest procesem, ciągiem operacji. Ich rezultat czerpie znaczenie jako pomiar jedynie na mocy powiązań z koncepcjami teoretycznymi, których dotyczy i na których się opiera.

Abstrakcje teoretyczne, do których odnoszą się obserwacje noszą nazwę modelu, który w nauce i technice ma niemal zawsze postać matematyczno fizyczną. Pomimo, że koncepcja modelu może wydawać się istotną jedynie dla teoretycznych rozważań, jej dobre zrozumienie wyraźnie usprawnia zaprojektowanie obserwacji i opracowywanie rezultatów pomiarowych. ↗

1.4 Model obserwacji

Będziemy interesowali się *modelem* obserwacji, tzn. teoretyczną koncepcją, za pomocą której opisywana jest jakaś sytuacja w świecie fizycznym — np. zmiana blasku gwiazdy, położenie komety. Opis nie musi być wyczerpujący, może dotyczyć jedynie tych aspektów przyrody, które stanowią przedmiot naszego zainteresowania. W zależności od tego co

zamiarujemy opisać, budowane są różne modele. Można też zbudować kilka modeli opisujących ten sam problem. Model zastępuje nam realnie istniejącą sytuację w celu łatwego dokonania pewnych oszacowań.

W modelu obserwacji będziemy wyróżniali dwie składowe : *model funkcjonalny* i *model stochastyczny*. Pierwsza składowa obejmuje ogólny opis własności deterministycznych rozważanego problemu, druga określa i opisuje własności stochastyczne zmiennych występujących w modelu funkcjonalnym, w szczególności tych zmiennych, które reprezentują wielkości pomiarowe. Dysponując modelem włączamy doń obserwacje. Obserwacje właśnie po to są robione aby ustalić wartości jednego bądź wielu parametrów modelu.

W astronomii i geodezji mamy do czynienia z różnymi modelami np. :

- w celu wyznaczenia odległości na powierzchni Ziemi, kształt Ziemi modelujemy za pomocą, sfery, elipsoidy obrotowej albo geoidy.
- do obliczenia poprawki refrakcyjnej w pozycyjnych obserwacjach ciał niebieskich wykorzystujemy tzw. model płaskiej atmosfery bądź model atmosfery radialnie symetrycznej,
- powstawanie obrazu w optyce astronomicznej — możemy posłużyć się tu modelem opartym na teorii optycznej Gaussa,
- w fotografii astronomicznej przyjmuje się, że obraz fotograficzny powstaje w wyniku projekcji centralnej,
- grawitacyjne pole Ziemi stanowi model sił w geodezji sateliternej,
- orbita keplerowska, orbita oskulacyjna to modele trajektorii ruchu ciał niebieskich.

Aby przekonać się do konstruowania modelu obserwacji, rozważmy prosty przykład. Mamy wyznaczyć parametry nieznanego trójkąta, interesuje nas kształt albo jego rozmiary. Matematyczny model problemu powinien ułatwić nam ustalenie minimalnej liczby elementów (zmiennych) potrzebnych do jednoznacznego rozwiązania postawionego zadania. W omawianym przykładzie, z modelu matematycznego (teoria trójkąta płaskiego) wiemy, że kształt trójkąta da się ustalić gdy znamy dwa kąty lub dwa boki i jeden kąt. Jeśli interesowałby nas rozmiar trójkąta, to nie zostanie on ustalony poprzez zmierzenie wyłącznie kątów. Rozważania tego typu pozwolą nam podjąć właściwą decyzję jeśli chodzi o zakup przyrządów niezbędnych w celu realizacji postawionego zadania.

Model funkcjonalny nie zawsze formułowany jest w sposób do końca jawny, dlatego może się zdarzyć, że zawiera on wiele elementów, których obecności trzeba się domyślać. A jeśli modele geometryczne nie są skomplikowane i dają się łatwo wizualizować, wówczas elementy modelu i świata fizycznego często przestają być odróżniane. Trzeba stale pamiętać, że w naturze nie ma takich obiektów jak punkty, kąty czy współrzędne. Są to jedynie elementy modelu funkcjonalnego stosowane do opisu odpowiednich obiektów naturalnych czy związków między nimi. Np. geodeta twierdzi, że mierzy odległość, a to oznacza, że odwołuje się do dwóch obiektów traktowanych jako dwa punkty geometryczne. Bywa, że zamiast odległości między nimi w bezpośrednim geometrycznym znaczeniu, geodeta może mierzyć jej rzut na płaszczyznę czy też na ziemską elipsoidę.

Modele funkcjonalne są zatem konstrukcją wygodną do dokonania analiz, a ze światem realnym połączone są za pomocą obserwacji.

W najprostszych przypadkach pomiarowych obserwacje odnoszą się w sposób bezpośredni do jakiegoś elementu modelu. Jednak często tak nie jest gdyż może się to okazać wysoce niepraktyczne. Np. odległość do sztucznego satelity Ziemi można mierzyć za pomocą laserów. W takich obserwacjach mierzony jest czas propagacji impulsu laserowego a nie odległość w sensie geometrycznym. I aby wykonane pomiary czasu połączyć z modelem wyznaczanej odległości, do prostego modelu wyznaczania odległości trzeba wprowadzić sporo dodatkowej teorii jak: propagacja promieniowania elektromagnetycznego w atmosferze Ziemi, kalibracja lasera. Pociąga to konieczność włączenia nowych zmiennych, nowych formuł matematycznych. Przykład ten ilustruje bardzo charakterystyczną cechę miernictwa, mianowicie rozbudowania-komplikacji modelu w efekcie połączenia go z obserwacjami. Powiększenie to sprawia, że zwykle czeka nas długi proces obliczeniowy zanim bezpośrednio rezultaty obserwacji zostaną powiązane z modelem.

W praktyce proces pomiarowy zawsze podlega wpływom, które nie mogą być w pełni kontrolowane. Podczas powtarzania obserwacji objawia się to drobną zmiennością uzyskiwanych wyników. Zmienności tej nie daje się powiązać z jakąś pojedynczą przyczyną natury fizycznej. Takie losowe, stochastyczne zmiany wyników obserwacji, czyli jak mówimy *fluktuacje* wyników, uważane są za podstawową własność precyzyjnych pomiarów i aby te własności należycie ująć, odwołujemy się do różnych koncepcji statystycznych. Ustalenie statystycznych własności elementów modelu najczęściej nie jest łatwym zadaniem. By poznać te własności można dokonać stosownej liczby powtórzeń obserwacji, co może być bardzo kosztowne, ale można też założyć, że pewne statystyczne prawidłowości (odkryte na podstawie wsześniejszych pomiarów dokonanych w podobnych warunkach), stosują się również do aktualnej sytuacji. Musimy wówczas starać się o utrzymanie odpowiednich warunków w jakich bieżące pomiary są dokonywane.

Bywa, że jako statystyczne własności obserwacji akceptujemy pewne bardzo grube przybliżenia. Np. niekiedy w astronomii i geodezji obserwacje traktowane są jako statystycznie niezależne, jako jednakowo dokładne, jako nieskorelowane. Czynimy tak ze względu na chęć uproszczenia problemu bądź ze względu na spore trudności w ustaleniu dokładności poszczególnych obserwacji, oszacowaniu korelacji etc.

Całość tego typu założeń o własnościach statystycznych interesujących nas zmiennych nazywana jest *modelem stochastycznym*. Wchodzą tu także parametry, którym przed procesem wyrównania obserwacji nadano z takich czy innych względów pewne wartości, a także te, których wartości zostaną dopiero wyznaczone. Wreszcie wchodzą tu czynniki związane z jakością wykorzystywanej aparatury czy też błędami osobowymi obserwatora. ↗

Rozdział 2

Opracowanie obserwacji, koncepcje

Streszczenie

Nie ma komu napisać I nie wiadomo czy będzie komu.

Słowa kluczowe: Pomiar, obserwacja, wyrównywanie obserwacji, model funkcjonalny, model stochastyczny, niepewności pomiarowe: przypadkowe, wzorcowania, eksperymentatora, danych z literatury, populacja, próba losowa, residua, metoda najmniejszych kwadratów, warunek najmniejszych kwadratów, reguły zaokrąglania, linearyzacja, metoda różniczki zupełnej, jacobian. ^a ↗

^a[Modyfikowano AD 2011, Marzec 16.]

2.1 Niepewności, błędy pomiarów, obserwacji

Przystępując do pomiaru danej wielkości fizycznej, np. ciśnienia atmosferycznego, jej model obserwacyjny obejmuje fizyczną koncepcję samej atmosfery, w szczególności ciśnienia jako jednego z parametrów określających stan atmosfery. Dokonując pomiaru ciśnienia, powszechnie oczekuje się, że przynajmniej w chwili pomiaru, ciśnienie atmosferyczne posiada ściśle określoną wartość, zwykle zwaną wartością *rzeczywistą* lub *prawdziwą*. Oznacza to przekonanie (wiarę), że bierzemy się za wyznaczenie tego co naprawdę istnieje.

Jednak najczęściej o istnieniu *rzeczywistych* wartości wielkości fizycznych jest sens mówić jedynie w ramach modelu obserwacyjnego. Np. gdy pytamy — ile *naprawdę* wynosi grubość miedzianego drutu? — zakładamy niejawnie, że drutu ma kształt walca. Tymczasem drut może posiadać w każdym miejscu inną grubość. Podobnie wygląda sprawa gdy wykonujemy pomiar grubości drutu w jakimś konkretnym miejscu — operowanie *rzeczywistą* grubością (tego miejsca) ma trudny do uchwycenia sens. Atomy drutu nieustannie drgają co pociąga ciągle chwilowe zmiany grubości drutu.

Ignorując trudności związane z ustaleniem tego co należy rozumieć przez *rzeczywistą* wartość x_r danej wielkości fizycznej, w wyniku wykonania pomiaru stajemy w obliczu kolejnej niewiadomej. Mianowicie, nigdy nie mamy pewności czy wynik x_p jest identyczny z wartością *rzeczywistą* x_r . Przyczyna tkwi w wielu czynnikach składających się na rezultat x_p , a których wpływu nie jesteśmy w stanie wyeliminować.

Uzyskana różnica

$$x_r - x_p = \Delta x_r$$

nazywana jest *rzeczywistą niepewnością pomiarową* wielkości fizycznej X .

W metrologii słowo niepewność bez dodatkowych określeń ma dwa znaczenia: ilościowej miary jak i parametru pewnej ogólnej koncepcji, stąd koniecznym jest uściślenie. O Δx_r ilościowej mierze różnicy $x_r - x_p$, mówimy jako o *niepewności pomiarowej*. Natomiast gdy mówimy *niepewność pomiaru* to będziemy przez to rozumieli parametr charakteryzujący rozrzut wartości wyników wielkości mierzonej. Uprzedzając, jego definicję — odchylenie standardowe jest właśnie takim parametrem, niepewnością pomiaru, charakteryzującym niepewność pomiarową mierzonej wielkości.

Pełna niepewność pomiarowa jest wypadkową:

- *niepewności przypadkowej*; o jej obecności w wynikach pomiarów świadczy występowanie tzw. statystycznego rozrzutu wyników (fluktuacje), którego źródła tkwią:
 - w naturze badanego zjawiska; np. w zjawisku rozpadu radioaktywnego, liczba rozpadów w jednostce czasu ma właśnie taki charakter,
 - w mierzonym przedmiocie; ten typ rozrzutu napotykamy podczas pomiarów grubości drutu uznawanego za cienki walec,
 - w czynnikach zewnętrznych przy jednoczesnym starannym zachowaniu odtwarzalności wyników; np. rozrzut spowodowany zmiennością reakcji zmysłów i przyrządów, zmiennością czynników uznanych za nieistotne,

- w *niepewności wzorcowania*; występuje ona zawsze i dominuje gdy nie ma rozrzutu wyników pomiaru, jej przyczyna tkwi w niepewności wzorcowania stosowanych mierników, niepewności wzorców stosowanych jako odniesienie,
- w *niepewności eksperymentatora*; chodzi tu o niepewności powodowane przyczynami znanymi eksperymentatorowi ale od niego niezależnymi jak np. drgania wskazówki miernika analogowego,
- w *niepewności danych wziętych z literatury*; dane zaczerpnięte z tablic matematycznych, z kalkulatora — najczęściej nie są dokładne.

Obok niepewności pomiarowej funkcjonuje jeszcze określenie *błąd pomiarowy*. Nazwa ta tradycyjnie obejmowała dwa składniki: błąd przypadkowy jak i błąd systematyczny. Obecnie (częściowo) odchodzimy od tej tradycji stosując w miejsce terminu błąd przypadkowy termin niepewność przypadkowa, wyjaśniony wyżej. Natomiast określenie błąd pomiarowy należy wykorzystywać w znaczeniu błędu zwanego dotąd błędem systematycznym, jako terminu rzeczywiście związanego z błędami w sztuce pomiarowej. Warto zapamiętać, że określenie błąd pomiarowy obejmuje:

- *błąd przeoczenia*; np. błąd występujący w trakcie pomiaru długości nienaprężonego przewodu,
- *błąd osobowy*; wynikający z niedoskonałości reakcji zmysłowych np. opóźnienie w odczycie momentu czasu przejścia gwiazdy przez nitki okularu,
- *pomyłki*; pomyłka eksperymentatora np. zamiast 21 odczytano 12, niepoprawna obsługa przyrządu, złe wyzerowanie miernika, niewłaściwe ustawienie koła podziałowego teleskopu,
- *błąd przybliżenia*; ma źródło np. w posługiwaniu się nieadekwatnym modelem pomiarowym, w rezultacie zastosowania przybliżonego wzoru.

Na błędy pomiarowe stosowane bywa określenie *niepewności systematyczne*. ↗

2.2 Ocena niepewności pomiarowych

Dokonując pomiaru nie otrzymujemy wartości rzeczywistej x_r , uzyskany rezultat pomiaru jest jedynie jej przybliżeniem. Jeśli wartość tego przybliżenia oznaczymy przez x_d to możemy zbudować przedział $x_d \pm U(x)$, w którym z dużym (niekiedy) ustalonym prawdopodobieństwem mieści się wartość rzeczywista x_r . Granice przedziału $U(x)$ będziemy określali za pomocą *niepewności standardowej* $u(x)$. To co powiedziano wyżej możemy wyrazić w jednej linii stosując notację probabilistyczną.

$$P[x_d - u(x) < x_r < x_d + u(x)] = P_{u(x)} \quad (2.1)$$

W dalszej części przedstawimy dwa sposoby (A i B) ilościowego oszacowania wyniku pomiaru i niepewności standardowej $u(x)$. W metodzie A niepewność standardową obliczana jest na podstawie rozkładu częstości wielkości mierzonej, a w metodzie B na podstawie

rozkładu prawdopodobieństwa założonego przez eksperymentatora. Metodę A stosujemy gdy mamy do czynienia ze statystycznym rozrzutem wyników pomiarów, w przeciwnym razie wykorzystujemy metodę B. ↗

2.2.1 Metoda A

Pojęcie niepewności przypadkowej pomiaru wykorzystywane jest do oceny poziomu rozrzutu wyników, przy czym, rozrzut musi mieć naturę statystyczną. Gdy w próbie pomiarowej kolejne wyniki (najczęściej) nieco się różnią, wówczas do ich opracowania można stosować statystykę matematyczną. Metodami statystycznymi obliczamy zarówno wynik końcowy (wartość średnią) jak i miarę rozrzutu pomiarów (odchylenie standardowe).

Rozmiar próby. Ze względu na *rozmiar próby*, czyli na liczbę n wykonanych pomiarów, próbę określamy jako:

- bardzo małą — gdy $n < 10$,
- małą — gdy $10 \leq n < 30$,
- dużą — gdy $30 \leq n < 100$,
- bardzo dużą — gdy $n \geq 100$.

Jeśli zbiór $\{x_1, x_2, x_3, \dots, x_n\}$ oznacza losową próbę n jednakowo dokładnych wyników pomiarów, tzn. n wyników wybranych losowo z np. nieskończonego zbioru wszystkich możliwych do wykonania pomiarów (populacja), to jako wynik końcowy wykonanego eksperymentu podajemy wartość średnią obliczoną jako średnia arytmetyczna z próby ¹

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.2)$$

gdzie x_i — wynik i -tego pomiaru, n — liczba wykonanych pomiarów (rozmiar próby),

Natomiast, jako miarę rozrzutu **poszczególnych** wyników, czyli jako miarę niepewności pojedynczego pomiaru przyjmujemy *odchylenie standardowe* s_x

$$s_x = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (2.3)$$

Ze względu na losowy charakter rozrzutu wyników, nie możemy ustalić znaku niepewności pomiarowej, dlatego należałoby pisać $\pm s_x$, ale, ponieważ w definicji (2.3) występuje pierwiastek kwadratowy, w samej definicji (2.3) znak \pm jest pomijany.

Zgodnie z wzorem (2.2), wartość średnia \bar{x} obliczana jest za pośrednictwem wielkości pomiarowych wykazujących rozrzut statystyczny, oznacza to, że gdybyśmy powtórzyli eksperyment i wyznaczyli \bar{x} za pomocą innej próby, uzyskalibyśmy inną wartość na \bar{x} . Powtarzając eksperyment wiele razy okazałoby się, że wartości \bar{x} również wykazują statystyczny rozrzut. Tyle, że byłby on wyraźnie mniejszy. Bez dowodu (chwilowo) podamy tu, że

¹Uzasadnienie takiego postępowania podane jest w podrozdziale 2.7



Rysunek 2.1: Przykład woltomierza analogowego klasy K=2.5, posiada on trzy zakresy pomiarowe.

jego miarą jest parametr $s_{\bar{x}}$ — o nazwie *odchylenie standardowe średniej*

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n(n-1)}} \quad (2.4)$$

lub zgodnie z (2.3) mamy

$$s_{\bar{x}} = \frac{s_x}{\sqrt{n}} \quad (2.5)$$

I właśnie odchylenie standardowe średniej $s_{\bar{x}}$ jest identyfikowane z niepewnością standardową $u(x)$ występująca w (2.1). Stąd, w wyniku zastosowania metody A, wyrażenie 2.1 uzyskuje formę

$$P[\bar{x} - s_{\bar{x}} < x_r < \bar{x} + s_{\bar{x}}] = P_{s_{\bar{x}}} = 0.68 \quad (2.6)$$

Obecność w tym wzorze liczby 0.68 zostanie uzasadnione w dalszym toku wykładu.

Metodę A obliczania niepewności standardowej stosujemy gdy przyczynek niepewności pochodzący od statystycznego rozrzutu jest przynajmniej o rząd wielkości większy od pozostałych. ↗

2.2.2 Metoda B

Jeśli wyniki pomiarów nie wykazują rozrzutu, tzn. wszystkie są sobie równe, i wynoszą $x_d = x_1 = x_2, \dots, = x_n$, niepewność standardową trzeba określić w inny sposób, zależny od tego jaki rodzaj niepewności ma wpływ dominujący.

Obecność niepewności wzorcowania. Przy braku rozrzutu, głównym przyczynkiem niepewności pomiarów może być niepewność wzorcowania $\Delta_d x$. Pomiary zawsze polegają na porównaniu wielkości mierzonej z wzorcem, którym najczęściej jest przyrząd pomiarowy, spełniający rolę wzorca.

Producenci, gwarantują taką dokładność przyrządu, by wynik uzyskany za jego pomocą nie różnił się od wzorca więcej niż o tzw. *działkę elementarną*, równą odstępowi $\Delta_d x$ pomiędzy sąsiednimi kreskami podziałki przyrządu. Oczywiście odstęp $\Delta_d x$ wyrażony jest w jednostkach podanych na podziałce przyrządu. Jest tak w przypadku przyrządów analogowych. W przypadku przyrządów cyfrowych działka elementarna jest równa jednostce

Tablica 2.1: Przykłady wartości działek elementarnych niektórych znanych przyrządów pomiarowych (za [3]).

miarka milimetrowa	$\Delta_d l = 1 \text{ mm}$
lekarski termometr rtęciowy	$\Delta_d \tau = 0.1^\circ\text{C}$
zegarek bez wskazówki sekundowej	$\Delta_d t = 1 \text{ min}$
zegarek cyfrowy wyświetlający sekundy	$\Delta_d t = 1 \text{ s}$
szybkościomierz samochodowy	$\Delta_d v = 5 \text{ km/h}$

wyświetlacza wskazującego najmniejszą wartość. Przykłady działek elementarnych podano w tabelce 2.1.

Jeśli posługujemy się przyrządem pomiarowym, np. miernikiem elektrycznym o znanej tzw. klasie pomiarowej, działkę elementarną możemy obliczyć za pomocą formuły

$$\Delta_d x = \frac{KZ}{100} \quad (2.7)$$

gdzie K, Z — oznaczają klasę pomiarową i zakres przyrządu, odpowiednio.

Ponieważ producent przyrządu gwarantuje, że odchylenie $x_r - x_d$ wielkości zmierzonej za pomocą tego przyrządu od wartości prawdziwej nie może przekroczyć wartości działki elementarnej $\Delta_d x$, stąd w kategoriach prawdopodobieństwa (stosujemy wzór (2.1) możemy twierdzić, że

$$P[x_d - \Delta_d x < x_r < x_d + \Delta_d x] = 1 \quad (2.8)$$

Jednak przyjęcie za niepewność standardową wartości $\Delta_d x$ byłoby rozwiązaniem zbyt asekuracyjnym. Niepewności standardowe powinny odpowiadać mniejszym wartościom prawdopodobieństwa, np. 0.50–0.70. W celu obliczenia niepewności standardowych odpowiadających takim wartościom prawdopodobieństwa, musimy dysponować rozkładem prawdopodobieństwa wykonywanych pomiarów. Jeśli nie znamy tego rozkładu, ratujemy się założeniem, że, np. każda wartość wyniku pomiaru wewnątrz przedziału $x_d - \Delta_d x < x_r < x_d + \Delta_d x$ jest jednakowo prawdopodobna. Co oznacza, że zakładamy jednostajny rozkład na wszystkie możliwe wyniki z wnętrza tego przedziału. Jak przekonamy się w dalszej części wykładu, dla rozkładu jednostajnego odchylenie standardowe wyraża się formułą

$$s_x = \frac{\Delta_d x}{\sqrt{3}}$$

Zatem gdy wyniki pomiarów zdominowane są przez niepewność wzorcowania, jako niepewność standardową możemy przyjąć

$$u(x) = \frac{\Delta_d x}{\sqrt{3}} \quad (2.9)$$

i wówczas prawdopodobieństwo, że wartość rzeczywista x_r mieści się w przedziale $(x_d \pm u(x))$ wynosi

$$P\left[x_d - \frac{\Delta_d x}{\sqrt{3}} < x_r < x_d + \frac{\Delta_d x}{\sqrt{3}}\right] = 0.57 \quad (2.10)$$

Uwaga! Gdyby zamiast jednostajnego założono inny rozkład prawdopodobieństwa, postać wzoru (2.10) byłaby inna.

Obecność niepewności eksperymentatora. Niepewność eksperymentatora $\Delta_e x$ definiujemy jako miarę niepewności pomiaru spowodowaną przyczynami niezależnymi od obserwatora; np. wahaniami wskazań miernika cyfrowego, drgania wskazówki przyrządu analogowego. Eksperymentator obserwuje wahaniami, drgania i w oparciu o swoje doświadczenie określa wielkość niepewności $\Delta_e x$, a następnie niepewność standardową $u(x)$. Gdy źródłem niepewności są drgania wskazówki, jako $\Delta_e x$ przyjmuje się połowę szerokości przedziału drgań. Podobnie jak w przypadku niepewności wzorcowania, niepewność standardowa eksperymentatora szacowana jest przy założeniu jednostajnego rozkładu wyników pomiarów w przedziale $(x_d - \Delta_e x < x_r < x_d + \Delta_e x)$. Pociąga to analogiczną do (2.9) formułę na niepewność standardową $u(x)$

$$u(x) = \frac{\Delta_e x}{\sqrt{3}} \quad (2.11)$$

Niepewność wielkości zaczerpniętych z literatury. Dane tablicowe lub inne zaczerpnięte z literatury niekonieczne są dokładne. Bywa, że obok interesującej nas wielkości w literaturze podano ich niepewność standardową $u(x)$. W takim wypadku nie musimy sami szacować jej wartości.

Gdy jest inaczej, najpierw ustalamy wartość samej niepewności $\Delta_t x$, jako równą 10-ciu jednostkom najmniej znaczącego miejsca dziesiętnego danej literaturowej. Zakładając jak poprzednio, że wartości wyników pomiaru w przedziale $(x_d - \Delta_t x < x_r < x_d + \Delta_t x)$ są jednakowo prawdopodobne, niepewność standardową obliczymy z wzoru

$$u(x) = \frac{\Delta_t x}{\sqrt{3}} \quad (2.12)$$

Na koniec przypominamy, że metodę B obliczania niepewności standardowej stosujemy gdy przyczynę niepewności od statystycznego rozrzutu pomiarów jest tego samego rzędu lub mniejszy od któregośkolwiek z pozostałych przyczynków. †

2.3 Wyniki końcowe, obliczanie końcowej niepewności

Gdy wyniki pomiarów wykazują statystyczny rozrzut, rezultat końcowy x_d obliczamy jako średnią arytmetyczną

$$x_k = \bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (2.13)$$

W wypadku przeciwnym, jako rezultat końcowy bierzemy dowolny wynik pomiaru $x_k = x_1 = x_2, \dots, = x_n$.

Całkowitą standardową niepewność rezultatu końcowego x_k obliczamy za pomocą formuły

$$u_k(x) = \sqrt{s_{\bar{x}}^2 + \frac{1}{3}(\Delta_d x)^2 + \frac{1}{3}(\Delta_e x)^2 + \frac{1}{3}(\Delta_t x)^2} \quad (2.14)$$

Tablica 2.2: Cyfry znaczące w wybranych liczbach, przykłady.

147	-	posiada 3 cyfry znaczące
14.78	-	posiada 4 cyfry znaczące
1.02	-	posiada 3 cyfry znaczące
0.01	-	posiada 1 cyfrę znaczącą
0.0120	-	posiada 3 cyfry znaczące
1200	-	posiada 4 cyfry znaczące
1.200	-	posiada 4 cyfry znaczące

Zauważmy, w formule (2.14) nie sumują się niepewności standardowe, lecz ich kwadraty. Wynika to z ogólnego prawa propagacji wariancji i kowariancji, o którym powiemy w dalszej części wykładu.

Formuła (2.14) może być okrojona o składniki o jeden rząd wielkości mniejsze od największego, lub o te, których w ogóle nie musimy uwzględniać; np. jeśli nie występuje rozrzut statystyczny wyników $s_{\bar{x}} = 0$. W laboratoriach uniwersyteckich najczęściej stosowaną postacią wzoru na całkowitą niepewność standardową jest

$$u_k(x) = \sqrt{s_{\bar{x}}^2 + \frac{1}{3}(\Delta_d x)^2} \quad (2.15)$$

Taka postać uwzględnia jedynie niepewność przypadkową i niepewność wzorcowania wyników pomiarów. ↗

2.4 Zaokrąglanie wyników pomiarów i ich zapis

Końcowe wyniki uzyskane za pośrednictwem wartości pochodzących z pomiaru mają charakter przybliżony i dlatego wymagają zaokrąglenia. Oczywiście moglibyśmy przedstawiać wyniki w takiej postaci w jakiej zostały one obliczone, np. z 16-toma cyframi. Takie postępowanie często pozbawione jest większego sensu, bo po cóż wypisywać stałą refrakcji w postaci 60.4567234560 skoro dokładność jej wyznaczenia wynosi zaledwie 0.01? Zaokrąglenia są konieczne i przy zaokrąglaniu wyników pomiarów bierzemy pod uwagę wartość odchylenia standardowego wyniku końcowego, korzystamy też z ogólnych zasad zaokrąglania liczb przybliżonych. ↗

2.4.1 Reguły zaokrąglania liczb przybliżonych

O każdej liczbie można powiedzieć, że składa się z pewnej liczby *cyfr znaczących*, o których wiadomo, że są *dokładne*, jak i pewnej liczby *cyfr wątpliwych*. Dla potrzeb naszego wykładu umówimy się, że cyfry znaczące to wszystkie cyfry danej liczby oprócz zer ustalających położenie kropki dziesiętnej liczby mniejszej od 1. Zastosujemy regułę przedstawiania liczb przybliżonych jedynie z jedną cyfrą wątpliwą, a więc jeśli w liczbie 127.882 cztery cyfry znaczące są pewne, to powinna być podana co najwyżej z czterema cyframi znaczącymi i z jedną cyfrą wątpliwą, czyli jako 127.88.

Liczby o wielu cyfrach wątpliwych należy zaokrąglić zgodnie z *regułami zaokrąglania* liczb przybliżonych. Dla naszych celów zastosujemy reguły poznane w szkole średniej.

Gdy w danej liczbie l wymaganych jest k cyfr znaczących, to od lewej strony tej liczby bierzemy $k + 1$ cyfr a resztę odrzucamy, a następnie stosujemy reguły zaokrąglania polegające na badaniu $k + 1$ cyfry, mianowicie:

1. jeśli $k + 1$ cyfra wynosi 0, 1, 2, 3, 4, to odrzucamy ją, a k -tą cyfrę pozostawiamy bez zmian;
2. jeśli $k + 1$ cyfra wynosi 5, 6, 7, 8, 9, wówczas odrzucamy ją, ale k -tą cyfrę powiększamy o jeden;

Przykład. $l = 13.3432$ i $k = 4$. Przycinamy l do postaci z 5-cima cyframi znaczącymi $l = 12.343$ i badamy 5-tą cyfrę od lewej. Ponieważ wynosi ona 3, to po zaokrągleniu będzie $l = 13.34$.

Przykład. $l = 13.2483$ i $k = 4$. Po przycięciu mamy $l = 13.248$, skoro 5-ta cyfra wynosi 8, zatem 4-tą cyfrę powiększamy o jeden, czyli po zaokrągleniu będzie $l = 13.25$,

Istnieją bardziej złożone od podanego wyżej algorytmy zaokrąglania, zainteresowanych czytelników odsyłamy do stron 98-99 w lit. [6], strony 10 w lit.[7], oraz do lit. [2] strona 20.

Wykonując obliczenia z liczbami przybliżonymi, rezultaty obliczeń również należy zaokrąglić podając w wyniku jedynie jedną cyfrę wątpliwą. I tak :

1. w dodawaniu (odejmowaniu) suma (różnica) powinna być zaokrąglona do tego miejsca dziesiętnego, które biorąc od prawej, jest najbardziej znaczące w dodawanych liczbach np.

$$165.21 + 121.1 + 8.2232 = 294.5332$$

po zaokrągleniu suma powinna mieć postać 294.5,

2. przy mnożeniu (dzieleniu) liczba cyfr dziesiętnych iloczynu (ilorazu) powinna być równa liczbie cyfr dziesiętnych czynnika z najmniejszą liczbą cyfr dziesiętnych (oprócz czynników dokładnych) np.

$$2.15 \cdot 11.1234 = 23.9$$

$$2 \cdot (2.15 \cdot 11.1234) = 47.8$$

3. przy wyciąganiu pierwiastka kwadratowego lub sześciennego z liczby przybliżonej, wynik winien mieć taką samą liczbę cyfr dziesiętnych jak liczba pierwiastkowana np.

$$\sqrt{7396} = 86.00$$

Podczas obliczeń bardziej złożonych wskazane jest utrzymywanie jednego dodatkowego miejsca dziesiętnego aż do zakończenia ciągu rachunków, po czym dokonać zaokrąglenia ostatecznego rezultatu. ↗

2.4.2 Zaokrąglenie i zapis wyników pomiarów

W przypadku liczb będących rezultatami pomiarów lub wartościami odchyłek standardowych postępujemy następująco (patrz lit. [3]):

- A – wynik pomiaru i jego niepewność wyrażamy w tych smych jednostkach SI,
- B – niepewność pomiarową zaokrąglamy do pierwszego miejsca albo do dwóch pierwszych miejsc dziesiętnych. Oznacza to, że niepewności pomiarowe wystarczy obliczać z dokładnością do trzech cyfr znaczących. Przed zaokrągleniem wynik niepewności standardowej $u_k(x)$ przedstawiamy w postaci

$$u(x) = 0.abc \cdot 10^m \quad (2.16)$$

gdzie a oznacza jedną z cyfr z przedziału $[1, 9]$, b i c cyfry z przedziału $[0, 9]$; m jest liczbą całkowitą.

Zaokrąglenie niepewności przebiega zależnie od wartości cyfry a :

- dla $a \leq 3$ — wynik zgodnie z regułami podanymi wcześniej zaokrąglamy do dwóch miejsc dziesiętnych:

$$u(x) = 0.ab' \cdot 10^m$$

gdzie b' oznacza cyfrę po zaokrągleniu.

- dla $a > 3$ — wynik zaokrąglamy do jednego miejsca dziesiętnego

$$u(x) = 0.a' \cdot 10^m$$

gdzie a' oznacza cyfrę po zaokrągleniu.

- C — końcowy wynik pomiaru zaokrąglamy na tym samym miejscu rozwinięcia dziesiętnego, na którym zaokrąglono wartość końcowej (całkowitej) niepewności standardowej.

W celu podania ostatecznego rezultatu pomiaru, wartość wielkości zmierzonej i jej niepewność zapisujemy w postaci wykładniczej lub stosujemy odpowiednie przedrostki przewidziane dla jednostek układu SI. Do kompletu podajemy także liczbę wykonanych pomiarów n . Wykładnik m (patrz wzór (2.16)) lub przedrostek dobieramy tak, by miejsca niepewne końcowej niepewności wystąpiły po przecinku, najlepiej na pierwszych dwóch miejscach, jak we wzorze (2.16).

W końcowym wyniku pomiarowym x_k po przecinku wypisujemy również zera, nawet jeżeli znajdują się one na miejscach obarczonych niepewnością (pierwsze lub pierwsze dwa miejsca po przecinku), patrz przykłady podane poniżej.

Przykłady. Poniżej podano kilka przykładów ([3]) zapisu wyników końcowych przed i po zaokrągleniu.

Przed zaokrągleniem	Po zaokrągleniu	Po zaokrągleniu
$R_k = 0.126333 \text{ k}\Omega$	$R_k = 126.3 \text{ }\Omega$	
$u_k(R) = 0.000906 \text{ k}\Omega$	$u_k(R) = 0.9 \text{ }\Omega$	
$n = 20$	$n=20$	
$C_k = 0.0002210045 \text{ F}$	$C_k = 221.0 \cdot 10^{-6} \text{ F}$	$C_k = 221.0 \text{ }\mu\text{F}$
$u_k(C) = 0.00000057 \text{ F}$	$u_k(C) = 0.6 \cdot 10^{-6} \text{ F}$	$u_k(C) = 0.6 \text{ }\mu\text{F}$
$n = 11$	$n=11$	$n=11$
$m_k = 17.350 \text{ g}$	$m_k = 17350.00 \cdot 10^{-6} \text{ kg}$	$m_k = 17350.00 \text{ mg}$
$\Delta m = 0.00010 \text{ g}$	$u_k(m) = 0.06 \cdot 10^{-6} \text{ kg}$	$u_k(m) = 0.06 \text{ mg}$
$n = 4$	$n=4$	$n=4$
$\tan 30^\circ = 0.57735027$	$\tan 30^\circ = 0.57735027$	
$\Delta(\tan 30^\circ) = 0.00000010$	$u_k(\tan 30^\circ) = 0.00000006$	
wynik z tablic	wynik z tablic	

↗

2.5 Ocena niepewności w pomiarach pośrednich

Podane dotąd sposoby obliczenia wyniku końcowego oraz jego niepewności można stosować jedynie do rezultatów pomiarów bezpośrednich. Gdy pomiary (w_1, w_2, \dots, w_m) nie stanowią końcowego celu eksperymentu i służą do obliczenia innej wielkości fizycznej z , końcowy rezultat oraz niepewność wielkości z obliczamy w inny sposób.

Założmy, że wielkość z wyznaczana jest za pomocą funkcji $z = f(w_1, w_2, \dots, w_m)$, do której podstawiamy m wielkości w_i uzyskanych z pomiarów. Np. stała refrakcji K wyznaczana jest za pośrednictwem ciśnienia P i temperatury T atmosfery ziemskiej: $K = f(P, T)$.

Wyróżnimy tu dwa przypadki:

- przypadek, w którym obserwacje $w_i, i = 1..m$ nie są skorelowane,²
- przypadek pomiarów skorelowanych, $w_i, i = 2..m$ są skorelowane.

Takie rozróżnienie jest ważne przy obliczaniu zarówno końcowego wyniku z_k jak i jej niepewności standardowej $u_k(z)$.

Przez korelację pomiędzy dwiema wielkościami pomiarowymi będziemy rozumieli miarę pewnej między nimi współzależności. Nie chodzi tu o jakąś zależność matematyczną (funkcjonalną), ale o statystyczną naturę tej zależności. ↗

²Przypadek ten obejmuje także sytuacje gdy mamy do czynienia z funkcjami zależnymi jedynie od jednej wielkości pomiarowej.

2.5.1 Pomiary pośrednie nieskorelowane

Wartość końcową wielkości \bar{z} obliczamy za pomocą tej samej zależności funkcyjnej, w której zamiast wartości pomiarowych w_i podstawiamy odpowiednie wartości końcowe \bar{w}_i , tych wielkości

$$\bar{z}_k = f(\bar{w}_1, \bar{w}_2, \bar{w}_3, \dots, \bar{w}_m) \quad (2.17)$$

Zakładając, że dla naszych potrzeb adekwatną jest zlinearyzowana postać zależności $z = f(w_1, w_2, \dots, w_m)$, niepewność końcową $u_k(z)$ (niepewność złożoną) wielkości z wyznaczamy za pomocą wzoru

$$u_k(z) = \sqrt{\sum_{i=1}^m \left(\frac{\partial f(w_1, w_2, \dots, w_i, \dots, w_m)}{\partial w_i} \right)^2 u_k^2(w_i)} \quad (2.18)$$

przy czym, wszystkie wartości pochodnych cząstkowych obliczamy podstawiając za w_i odpowiednie wartości końcowe (np. średnie) \bar{w}_i . Oznacza to, że zanim przystąpimy do obliczenia końcowych wartości \bar{z}_k oraz $u_k(z)$, wcześniej obliczamy wartości końcowe \bar{w}_i i ich niepewności $u_k^2(w_i)$. ↗

2.5.2 Pomiary pośrednie skorelowane

Podobnie jak poprzednio, w celu obliczenia końcowych wartości \bar{z}_k , $u_k(z)$ musimy dysponować wartościami końcowymi \bar{w}_i , $u_k^2(w_i)$ wszystkich wielkości służących do wyznaczenia wielkości z .

Wartość końcową \bar{z}_k obliczamy za pomocą formuły (2.17) natomiast by obliczyć końcowe odchylenie standardowe $u_k(z)$, musimy dodatkowo uwzględnić iloczyny pierwszych pochodnych liczonych względem każdej pary zmiennych w_i, w_j

$$u_k(z) = \left[\sum_{i=1}^m \left(\frac{\partial f(w_1, \dots, w_m)}{\partial w_i} \right)^2 u_k^2(w_i) + 2 \sum_{i=1}^{m-1} \sum_{j=i+1}^m \frac{\partial f(w_1, \dots, w_m)}{\partial w_i} \frac{\partial f(w_1, \dots, w_m)}{\partial w_j} u_k(w_i) u_k(w_j) r(w_i, w_j) \right]^{\frac{1}{2}}$$

gdzie współczynnik korelacji $r(w_i, w_j)$ zdefiniowany jest wzorem

$$r(w_i, w_j) = \frac{s(w_i, w_j)}{u(w_i)u(w_j)} \quad (2.19)$$

Współczynnik ten jest miarą korelacji pomiędzy zmiennymi w_i i w_j . Występująca we wzorze (2.19) kowariancja $s(w_i, w_j)$ wielkości pomiarowych w_i i w_j zdefiniowana jest formułą

$$s(w_i, w_j) = \frac{1}{(n-1)} \sum_{k=1}^n (w_{i,k} - \bar{w}_i)(w_{j,k} - \bar{w}_j) \quad (2.20)$$

gdzie $w_{i,k}$ w wyrażeniu pod znakiem sumy oznacza k -ty pomiar i -tej mierzonej wielkości fizycznej. ↗

2.6 Statystyczna koncepcja rachunku wyrównawczego

W rozdziale 2.2 w celu obliczenia końcowego rezultatu pomiarów zastosowano średnią arytmetyczną, uczyniono to bez jakiegokolwiek uzasadnienia, co u części czytelników wywołało uczucie silnego niedosytu. By ten niedosyt osłabić proponujemy rozważania przybliżające nas do uzasadnienia powszechnego stosowania średniej arytmetycznej jako wartości końcowych wielkości pomiarowych.

Rozpatrzmy prosty przypadek pomiaru jednej wartości pewnej wielkości fizycznej, np. pomiaru długości, temperatury, napięcia na zaciskach opornika, etc. W takim przypadku jest jasne, że wobec problemu — ile wynosi np. ta długość? — w celu uzyskania jednoznacznej odpowiedzi wystarczy dokonać jednego pomiaru. Tę informację zapisujemy jako równość $n_0 = 1$, bowiem w przypadku pomiaru długości, minimalna liczba pomiarów n_0 konieczna do jednoznacznego rozwiązania postawionego problemu wynosi jeden.³

Jeśli już wiemy czego nam potrzeba i co należy zrobić, czyli już dobrze znamy nasz model pomiarowy, możemy przystąpić do wykonania n_0 pomiarów. Jednak w praktyce zawsze wykonujemy ich więcej, choćby ze względu na możliwość wystąpienia pomyłki (np. błędu przeoczenia). Tak przynajmniej postępuje każdy dobry praktyk.

Zatem, po zakończeniu eksperymentu, dysponujemy n rezultatami obserwacji, a ponieważ $n > n_0$ to mamy do dyspozycji obserwacje nadliczbowe. Nadmiar pomiarów wyrażamy jako różnicę

$$r = n - n_0$$

zwaną *statystyczną liczbą stopni swobody*.

Powstaje teraz ciekawa sytuacja. Gdy obserwacje wykazują rozrzut, to z każdego zbioru n_0 obserwacji, wziętego spośród n obserwacji możemy otrzymać inną wartość, np. inną wartość długości zmierzonego przedmiotu. Zatem zbiór n rezultatów obserwacji nie opisuje jednoznacznie rozważanego problemu (modelu). Oczywistą przyczyną tego stanu rzeczy są fluktuacje pomiarowe.

Ta szczególna i nieprzyjemna sytuacja usuwana jest w następujący sposób. Każdy rezultat obserwacji l_i zastępujemy jego *estymatorem* oznaczanym symbolem \hat{l}_i , ale takim, że każdy podzbiór n_0 estymatorów określa model pomiarowy dokładnie w ten sam sposób.

Pozostaje jedynie podać metodę wyznaczenia estymatorów, np. taką — każdy z estymatorów otrzymamy za pomocą poprawki dodanej do wartości obserwowanej

$$\begin{aligned}\hat{l}_1 &= l_1 + \nu_1 \\ \hat{l}_2 &= l_2 + \nu_2 \\ \hat{l}_3 &= l_3 + \nu_3 \\ &\dots \\ \hat{l}_n &= l_n + \nu_n\end{aligned}$$

Poprawki ν_i tradycyjnie nazywane są przez astronomów i geodetów residuami.

³Oczywiście w innych problemach liczba ta może być większa.

Tworząc odpowiednie macierze kolumnowe możemy te związki ująć w formie

$$\hat{\mathbf{L}} = \mathbf{L} + \boldsymbol{\nu}$$

gdzie $\boldsymbol{\nu}$ jest macierzą *residuów* dodawanych do macierzy wartości obserwowanych \mathbf{L} .

Oczywiście residua $\boldsymbol{\nu}$ muszą być wyznaczone przed estymatorami $\hat{\mathbf{L}}$. Istnieje nieskończona liczba możliwych zbiorów $\boldsymbol{\nu}$, dla których obliczone estymatory $\hat{\mathbf{L}}$ określają model jednoznacznie. Logicznym⁴ jest jednak oczekiwać, że jedynie jeden z takich zbiorów da rozwiązanie optymalne.

Potrzebujemy więc warunku (kryterium) optymalności i co może nas zaskoczy, do dyspozycji mamy ich dość sporo. Jednym z najpowszechniej wykorzystywanych w astronomii i geodezji jest kryterium najmniejszych kwadratów zaproponowane w końcu XVIII stulecia przez K. Gaussa.

Procedurę znajdowania zbioru estymatorów $\hat{\mathbf{L}}$ w oparciu o pewne kryterium nazywamy opracowaniem (wyrównaniem) pomiarów.⁵

Oczywiście dane pomiarowe można opracowywać innymi prostszymi metodami. W przypadku wyznaczenia długości przedmiotu, posługujemy się w tym celu średnią arytmetyczną pomiarów. Nam jednak potrzeba ogólnej procedury, którą będziemy stosowali w problemach bardziej złożonych od wyznaczania rozmiaru przedmiotu, temperatury

Procedura wyrównawcza, w której jako kryterium optymalności zastosowano kryterium najmniejszych kwadratów, nosi nazwę *metody najmniejszych kwadratów* (MNK).[†]

2.7 Metoda najmniejszych kwadratów

W dalszej części tego rozdziału zakładamy, że interesujące nas obserwacje są nieskorelowane i o jednakowej precyzji.

Kryterium najmniejszych kwadratów wymusza by wyznaczone residua $\boldsymbol{\nu}$ czyniły zadość następującemu warunkowi

$$\Phi = \boldsymbol{\nu}^T \boldsymbol{\nu} = \sum_{i=1}^n \nu_i^2 \longrightarrow \text{minimum} \quad (2.21)$$

Zatem, w rachunku wyrównawczym, obok nałożonego na estymatory $\hat{\mathbf{L}}$ wymagania by jednoznacznie opisywały model pomiarowy, na residua $\boldsymbol{\nu}$ dodatkowo nakłada się warunek najmniejszych kwadratów.

Przykład wyrównywania obserwacji.

Problem. Wyznaczono odległość za pomocą pomiarów bezpośrednich. Zgodnie z modelem

⁴W dawnych dobrych czasach nikt w tym miejscu nie mówiłby o logice, zresztą całkiem słusznie. Po prostu odwołano by się do tzw. chłopskiego rozumu. Jednak dzięki słowu “logicznie” wypowiedziane zdania brzmią mądrzej i posiadają dostojny naukowy charakter.

⁵Równie dobrze można by powiedzieć: procedurę znajdowania wyników końcowych

tego zagadnienia, w celu jednoznacznego wyznaczenia odległości wystarczyło wykonać jeden pomiar ($n_0 = 1$). Tymczasem wykonano ich dwa:

$$l_1 = 15.12 \quad l_2 = 15.14$$

gdzie jednostką pomiarową był metr. Zatem mamy:

$$\begin{aligned} n &= 2 \\ r &= n - n_0 = 1 \end{aligned}$$

Rozwiązanie. Ponieważ pomiary l_1, l_2 różnią się, mamy do czynienia z niejednoznacznością, bowiem nie wiemy, który z pomiarów wybrać jako rezultat końcowy wyznaczanej odległości.

Ale jak powiedziano wyżej, w celu usunięcia niejednoznaczności, należy wyznaczyć wyrównane wartości pomiarów, czyli estymatory \hat{l}_1, \hat{l}_2 . Trzeba to tak zrobić by oba pomiary opisywały model jednoznacznie, co w tym przykładzie oznacza, że musi być spełniony warunek

$$\hat{l}_1 - \hat{l}_2 = 0 \tag{2.22}$$

Równanie to nazywamy *równaniem warunkowym*, a jego postać a także ich liczba (bo warunków może być więcej) zależą od konkretnego problemu pomiarowego.

Estymatory wyznaczone są za pośrednictwem poprawek (residuów)

$$\begin{aligned} \hat{l}_1 &= l_1 + \nu_1 \\ \hat{l}_2 &= l_2 + \nu_2 \end{aligned}$$

Powiedziano wcześniej, że residua, dla których estymatory spełniają równanie warunkowe (2.22) można wybrać na wiele sposobów, np.:

$$\begin{aligned} \nu_1 &= 0.0 \\ \nu_2 &= -0.02 \end{aligned}$$

lub

$$\begin{aligned} \nu_1 &= 0.01 \\ \nu_2 &= -0.01 \end{aligned}$$

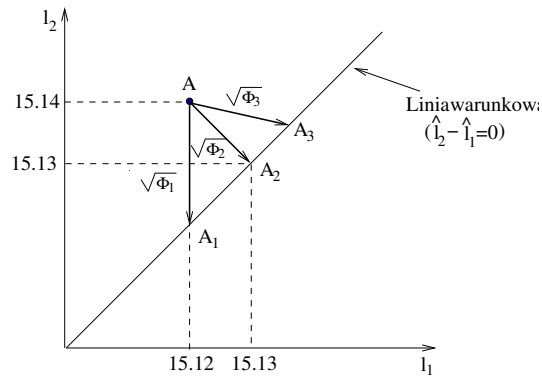
lub

$$\begin{aligned} \nu_1 &= 0.015 \\ \nu_2 &= -0.005 \end{aligned}$$

Wszystkie powyższe pary czynią zadość warunkowi (2.22), ale jeśli policzymy odpowiadające im wartości funkcji (2.21), otrzymamy w metrach kwadratowych:

$$\begin{aligned} \Phi_1 &== 4 \cdot 10^{-4} \\ \Phi_2 &== 2 \cdot 10^{-4} \\ \Phi_3 &== 2.5 \cdot 10^{-4} \end{aligned}$$

Są to różne wartości, a Φ_2 jest z nich najmniejsze, ale czy jest najmniejsze z wszystkich możliwych? Znalezienie odpowiedzi ułatwi nam rysunek (2.2) przedstawiający geometryczną interpretację zastosowania warunku najmniejszych kwadratów w naszym problemie. Osie układu współrzędnych odpowiadają pomiarowi numer 1 i 2, odpowiednio.



Rysunek 2.2: Interpretacja metody najmniejszych kwadratów zastosowanej w problemie wyrównywania dwóch pomiarów. Opis w tekście.

Położenie punktu A odpowiada wartościom obserwacyjnym l_1 i l_2 . Prosta o równaniu

$$\hat{l}_2 - \hat{l}_1 = 0$$

tzw. prosta warunkowa, reprezentuje warunek jaki muszą spełnić estymatory \hat{l}_1, \hat{l}_2 aby model pomiarowy (wyznaczona długość) był opisany jednoznacznie.

W omawianym przykładzie procedurę wyrównywania obserwacji można wyobrazić sobie jako zastępowanie punktu A innym punktem, byle tylko znajdującym się na prostej warunkowej. Jak widzimy można czynić to rzeczywiście na bardzo wiele sposobów, np. posługując się wektorem $A\vec{A}_1$ o długości $\sqrt{\Phi_1}$ przesuujemy punkt A do punktu A_1 . Ale równie dobrze możemy posłużyć się innymi wektorami $A\vec{A}_2, A\vec{A}_3, \dots$

Jednak ze wszystkich możliwych jedynie jedna para ν_1, ν_2 spełnia kryterium najmniejszych kwadratów (2.21). Są to residua, za pomocą których punkt A przemieścił się w punkt A_2 . Wektor $A\vec{A}_2$ jako prostopadły do linii warunkowej ma najmniejszą długość.

Residua, dla których punkt A przemieszcza się w A_2 spełniają warunek najmniejszych kwadratów, a *jednocześnie* punkt A_2 leży na prostej warunkowej, spełnia równanie warunkowe. W ten sposób model postawionego problemu jest określony optymalnie i jednoznacznie. Pamiętajmy jednak, że jest to *optymalnie* w sensie najmniejszych kwadratów, bo uzyskana przez nas odpowiedź niekoniecznie jest najdokładniejszą, najbliższą odpowiedzi prawdziwej, bo tej przecież nikt z nas nie zna.

Zatem ostatecznie w naszym przykładzie wyrównane wartości pomiarów l_1, l_2 wynoszą

$$\hat{l}_1 = \hat{l}_2 = 15.13 \quad [\text{m}]$$

Zauważmy jeszcze, że w przypadku dwóch lub więcej pomiarów odległości, jeśli pomiary mają jednakową precyzję oraz są nieskorelowane, w celu znalezienia "najlepszej" odległości wystarczy obliczyć średnią arytmetyczną

$$\hat{x} = \frac{l_1 + l_2}{2} = 15.13 \quad [\text{m}]$$

Rzeczywiście, dla wielu zagadnień szczegółowych można podać odpowiednie przepisy pozwalające na usunięcie niejednoznaczności opisu badanego modelu. Warto jednak opanować aparat pozwalający na wyrównanie obserwacji w każdym przypadku. I dlatego w dalszych wykładach powrócimy do metody najmniejszych kwadratów. ↗

2.8 Dygresja matematyczna: propagacja niepewności, linearyzacja funkcji

Niemal powszechnie rezultaty pomiarów służą do obliczenia innych wielkości. Ponieważ w dziedzinie wielkości pomiarowych mamy do czynienia z niepewnościami, naturalnym jest, że niepewności pojawiają się także po stronie rezultatów obliczeń. Oszacowanie niepewności wartości funkcji zależnych od wielkości pomiarowych, nazywamy *propagacją niepewności*.

2.8.1 Przypadek jednowymiarowy

Niech x będzie wielkością mierzoną, y natomiast będzie wielkością obliczoną za pomocą formuły liniowej

$$y = ax + b \quad (2.23)$$

gdzie parametry a, b są znane dokładnie.

Gdy x_p będzie wartością dokładną to prawdziwe y_p wynosi

$$y_p = ax_p + b$$

Jeśli za wartość zmierzoną położymy $x = x_p + dx$, to odpowiadającą jej “zmierzona” wartość y wynosi

$$\begin{aligned} y &= a(x_p + dx) + b = ax_p + b + adx \\ y &= y_p + adx = y_p + dy \end{aligned}$$

Stąd $dy = adx$. Z drugiej strony, z równania (2.23) mamy

$$\frac{dy}{dx} = a$$

Zatem, błąd obliczanej wielkości można wyznaczyć za pomocą *różniczki zupełnej* funkcji y , czyli

$$dy = \frac{dy}{dx} dx \quad (2.24)$$

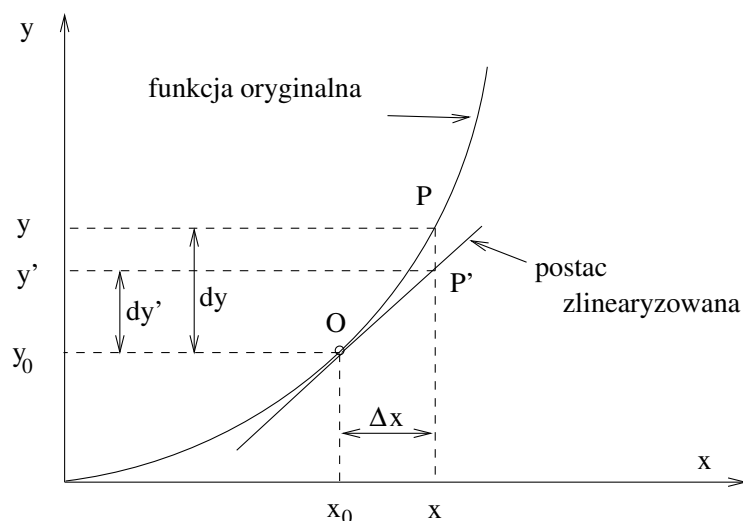
Przyrost dy obliczamy zawsze za pomocą (2.24) jeśli rozważana funkcja jest funkcją liniową (2.23).

Dla funkcji nieliniowych już tak dobrze nie będzie, np. dla funkcji kwadratowej $y = x^2$, błąd dx propaguje do wartości y zgodnie z formułą

$$dy = \left[\frac{dy}{dx} \right]_p dx + (dx)^2 \quad (2.25)$$

co przecież nie jest różniczką zupełną funkcji $y = x^2$.

W praktyce, w przypadku funkcji nieliniowych propagację niepewności szacuje się za pomocą formuł przybliżonych, np. w przypadku funkcji kwadratowej w równaniu (2.25)



Rysunek 2.3: Interpretacja geometryczna metody oszacowania propagacji niepewności Δx z pomocą zlinearyzowanej formy funkcji $y = f(x)$. Opis w tekście.

odrzucający jest człon kwadratowy $(dx)^2$. A taki zabieg jest równoważny zastosowaniu różniczki zupełnej do zlinearyzowanych form funkcji nieliniowych.

Podstawą *linearyzacji* funkcji jest szereg Taylora. Dla funkcji jednej zmiennej, $y = f(x)$, w otoczeniu pewnego x_0 szereg Taylora ma postać

$$y = y_0 + \left[\frac{dy}{dx} \right]_{x_0} \Delta x + \dots$$

gdzie $y_0 = f(x_0)$, $\Delta x = x - x_0$.

Ograniczając się w tym szeregu do formy liniowej i stosując do niej metodę różniczki zupełnej możemy dość dokładnie oszacować propagację niepewności x -sa w funkcji $f(x)$. Jak dokładnie? — zależy to od postaci odrzuconych wyrazów oraz od przyrostu Δx .

Interpretację geometryczną efektu zaniedbania wyrazów wyższych rzędów w szeregu Taylora przedstawiono na rysunku 2.3. Funkcja $f(x)$ zostaje zastąpiona prostą styczną do funkcji w punkcie O . Widzimy, że dla przyrostu Δx , zamiast $dy = y - y_0$, w oparciu o formę liniową bierzemy wartość $dy' = y' - y_0$.

2.8.2 Przypadek wielowymiarowy

W przypadku funkcji wielu zmiennych $y = f(x_1, x_2, \dots, x_n)$ jej formą zlinearyzowaną będzie

$$y = y_0 + \left[\frac{\partial y}{\partial x_1} \right]_{x_{i0}} \Delta x_1 + \left[\frac{\partial y}{\partial x_2} \right]_{x_{i0}} \Delta x_2 + \dots + \left[\frac{\partial y}{\partial x_n} \right]_{x_{i0}} \Delta x_n$$

która po wprowadzeniu oznaczenia $\left[\frac{\partial y}{\partial x_i} \right]_{x_{i0}} = j_i$ przyjmie postać

$$y = y_0 + j_1 \Delta x_1 + j_2 \Delta x_2 + \dots + j_n \Delta x_n$$

A tworząc odpowiednie macierze, wierszową i kolumnową otrzymamy

$$y = y_0 + [j_1, j_2, \dots, j_n] \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_n \end{bmatrix}$$

Po dalszej kompresji

$$y = y_0 + \mathbf{j} \Delta \mathbf{x} \quad (2.26)$$

Powyższą technikę można rozciągnąć na przypadek bardziej ogólny. Niech będą dane m funkcji y_1, \dots, y_m , z których każda jest funkcją n zmiennych x_1, \dots, x_n :

$$\begin{aligned} y_1 &= f_1(x_1, x_2, \dots, x_n) \\ y_2 &= f_2(x_1, x_2, \dots, x_n) \\ &\dots \\ y_m &= f_m(x_1, x_2, \dots, x_n) \end{aligned}$$

Każdą z tych funkcji można zlinearyzować w sposób dopiero co opisany

$$\begin{aligned} y_1 &= y_{10} + \left[\frac{\partial y_1}{\partial x_1} \right]_{x_{10}} \Delta x_1 + \dots + \left[\frac{\partial y_1}{\partial x_n} \right]_{x_{n0}} \Delta x_n \\ y_2 &= y_{20} + \left[\frac{\partial y_2}{\partial x_1} \right]_{x_{10}} \Delta x_1 + \dots + \left[\frac{\partial y_2}{\partial x_n} \right]_{x_{n0}} \Delta x_n \\ &\dots \\ y_m &= y_{m0} + \left[\frac{\partial y_m}{\partial x_1} \right]_{x_{10}} \Delta x_1 + \dots + \left[\frac{\partial y_m}{\partial x_n} \right]_{x_{n0}} \Delta x_n \end{aligned}$$

a dalej

$$\begin{aligned} y_1 &= y_{10} + j_{11} \Delta x_1 + j_{12} \Delta x_2 + \dots + j_{1n} \Delta x_n \\ y_2 &= y_{20} + j_{21} \Delta x_1 + j_{22} \Delta x_2 + \dots + j_{2n} \Delta x_n \\ &\dots \\ y_m &= y_{m0} + j_{m1} \Delta x_1 + j_{m2} \Delta x_2 + \dots + j_{mn} \Delta x_n \end{aligned}$$

a w postaci zwartej

$$\begin{aligned} y_1 &= y_{10} + \mathbf{j}_1 \Delta \mathbf{x} \\ y_2 &= y_{20} + \mathbf{j}_2 \Delta \mathbf{x} \\ &\dots \\ y_m &= y_{m0} + \mathbf{j}_m \Delta \mathbf{x} \end{aligned}$$

Tworząc kolumnowe macierze y-greków, j-tów otrzymamy

$$\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix} = \begin{bmatrix} y_{10} \\ \vdots \\ y_{m0} \end{bmatrix} + \begin{bmatrix} \mathbf{j}_1 \\ \vdots \\ \mathbf{j}_m \end{bmatrix} [\Delta \mathbf{x}]$$

i ostatecznie liniowa forma wektora \mathbf{y} ma postać

$$\mathbf{y} = \mathbf{y}_0 + \mathbf{J} \Delta \mathbf{x}$$

Macierz pochodnych \mathbf{J} , zwana jakobianem jest macierzą o elementach

$$\mathbf{J} = \frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_n} \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_n} \end{bmatrix}$$

Wartości tych pochodnych podobnie jak i wartości liczbowe składowych wektora \mathbf{y}_0 obliczyć należy w punktach x_{10}, \dots, x_{n0} . ↗

Rozdział 3

Prawdopodobieństwo, rozkłady, parametry

Streszczenie

Nie ma komu napisać

I nie wiadomo czy będzie komu.

Słowa kluczowe: Definicje prawdopodobieństwa, prawdopodobieństwo a statystyka, populacja, próba, rozkład prawdopodobieństwa, prawdopodobieństwo warunkowe, twierdzenie Bayesa, funkcja gęstości prawdopodobieństwa, dystrybuanta, rozkład wielowymiarowy, rozkład brzegowy, rozkład warunkowy, niezależność zdarzeń, niezależność zmiennych losowych, wartość oczekiwana, wariancja, odchylenie standardowe, kowariancja, korelacja, współczynnik korelacji, momenty centralne, momenty zwykłe, asymetria funkcji gęstości rozkładu prawdopodobieństwa, kurtoza, macierz wariancji-kowariancji, modalna, mediana, kwantyle, decyle, fraktyle, nierówność Czebyszewa. ^a ↗

^a[Modyfikowano AD 2011, Marzec 03.]

3.1 Wstęp

Przedstawimy tu niektóre pojęcia przydatne w opanowaniu metod opracowania obserwacji. Pojęcia te należą do dwóch pokrewnych działów matematyki: prawdopodobieństwa i statystyki.

Prawdopodobieństwo jest narzędziem, którym w statystyce posługujemy się w celu rozwiązania różnych problemów. Weźmy taki przykład. Pudełko zawiera 5 kul niebieskich, 5 czerwonych i 5 białych. Zadaniem rachunku prawdopodobieństwa jest udzielanie odpowiedzi np. na takie pytanie — jeśli na chybił trafił wybrano jedną kulę, jakie jest prawdopodobieństwo, że będzie to kula niebieska? Zauważmy, że wcześniej wiedziano ile i jakie kule są w pudełku.

W przypadku statystyki jest inaczej, nie wiemy co jest w pudełku i aby się tego dowiedzieć wyciągamy zeń coś na próbę. Na podstawie prób wnioskujemy na temat tego co znajduje się w pudełku po czym stosujemy rachunek prawdopodobieństwa.

Statystykę wykorzystujemy w celu uzyskania ocen, sformułowania wniosków opartych o zbiór danych pochodzących z obserwacji. Statystyka jest powiązana z metodami pozyskiwania, analizowania i interpretowania danych.

W oparciu o statystykę obserwator stara się dokonać właściwego osądu w obliczu niewiadomej. Jednak statystyka dostarcza jedynie narzędzi ułatwiających podjęcie decyzje, nigdy nie prowadzi wprost do decyzji, którą należy podjąć.

3.2 Prawdopodobieństwo 'definicje'

Użycie cudzysłowu w tytule tego podrozdziału ma na celu zwrócenie uwagi na trudności występujące w klasycznych definicjach prawdopodobieństwa. Czytelnik na pewno zauważy, że w matematycznej definicji prawdopodobieństwa tkwi założenie jednakowego prawdopodobieństwa zdarzeń elementarnych, pojęcia właśnie definiowanego.

3.2.1 Prawdopodobieństwo matematyczne

Klasyczna *matematyczna 'definicja' prawdopodobieństwa* korzysta z pojęcia układu zupełnego zdarzeń. Chodzi tu o zdarzenia elementarne, wzajemnie się wykluczające i jednakowo prawdopodobne.

Jeśli w układzie zupełnym N zdarzeń elementarnych, interesującemu nas zdarzeniu E_i sprzyja m_{E_i} zdarzeń to prawdopodobieństwo zajścia zdarzenia E_i można określić jako

$$P[E_i] = \frac{m_{E_i}}{N} \quad (3.1)$$

Tę definicję ¹ da się zastosować wszędzie tam gdzie potrafimy wyniki doświadczenia rozdzielić na jednakowo prawdopodobne zdarzenia elementarne, ustalić ich ogólną liczbę oraz liczbę zdarzeń sprzyjających.

¹A może należałoby powiedzieć — sposób obliczania prawdopodobieństwa.

3.2.2 Prawdopodobieństwo geometryczne

Prawdopodobieństwo możemy również obliczać jako stosunek pewnych miar. Rozważmy taki przykład. Strzelamy do wirującej okrągłej tarczy, której wycinek równy $1/6$ pomalowano na zielono a pozostałą część na biało. Strzały oddawane są z niewielkiej odległości, można więc przyjąć, że po każdym strzale trafienie w tarczę jest pewne oraz, że prawdopodobieństwo trafienia w dowolny punkt tarczy jest takie samo.² Interesuje nas prawdopodobieństwo trafienia w wycinek zielony.

Natychmiastowe zastosowanie definicji (3.1) jest niemożliwe, bowiem nie potrafimy obliczyć liczby zdarzeń sprzyjających i liczby wszystkich zdarzeń możliwych. Możemy jednak skorzystać z podszeptów naszej intuicji (chłopskiego rozumu) i przyjąć, że prawdopodobieństwo jest równe stosunkowi pola części zielonej do pola powierzchni całej tarczy

$$P[E_i] = \frac{S(m_{E_i})}{S(N)}$$

gdzie $S(m_{E_i})$ oznacza pole powierzchni zielonego wycinka tarczy, $S(N)$ pole powierzchni całej tarczy.

Prawdopodobieństwo jest tu określone jako stosunek miar: miary $S(m_{E_i})$ zdarzeń sprzyjających elementarnych i jednakowo prawdopodobnych m_{E_i} , do miary wszystkich zdarzeń elementarnych N .

Ogólny schemat takiej “definicji” prawdopodobieństwa ma postać

$$P[E_i] = \frac{\text{miara } m_{E_i}}{\text{miara } N} \quad (3.2)$$

gdzie miara m_{E_i} i miara N to miary zdarzeń sprzyjających i wszystkich możliwych.

Wiele problemów praktycznych dotyczących wyznaczenia prawdopodobieństwa sprowadza się do obliczenia stosunku miar geometrycznych pól, długości, kątów etc. Dlatego prawdopodobieństwo określone wzorem (3.2) często nazywane jest *prawdopodobieństwem geometrycznym*.

Stosowanie formuły (3.2) wymaga upewnienia się czy ma miejsce odpowiedniość (proporcjonalność) między daną miarą i liczbą możliwych wyników.

3.2.3 Prawdopodobieństwo częstościowe

Jeśli nie możemy zastosować formuł (3.1) i (3.2) możemy spróbować posłużyć się *statystyczną, częstościową* “definicją” prawdopodobieństwa. Jest to jedna z klasycznych definicji prawdopodobieństwa oparta o eksperyment, w której korzystamy z pojęcia *częstości*. Chodzi tu o prawdopodobieństwo rozumiane jako granica stosunku liczby n wystąpień interesującego nas zdarzenia E_0 do liczby N realizacji wszystkich zdarzeń E , tzn.

$$P[E_0] = \lim_{N \rightarrow \infty} \frac{n}{N} \quad (3.3)$$

↳

²Pomijamy rozmiary kuli i traktujemy ją jako punkt.

3.3 Aksjomatyczne ujęcie prawdopodobieństwa

W teorii prawdopodobieństwa podane wcześniej “definicje” prawdopodobieństwa, nie znajdują już zastosowania. W zamian, dobierany jest układ pewników określających niektóre podstawowe własności prawdopodobieństwa, czyli proponuje się aksjomatyczne ujęcie prawdopodobieństwa.

Pojęciem pierwotnym w aksjomatycznej teorii jest pojęcie zbioru zdarzeń elementarnych E , wzajemnie wykluczających się. Z elementów tego zbioru można utworzyć zdarzenia Z będące podzbiórami zbioru E , w szczególności zdarzenie niemożliwe, rozumiane jako zdarzenie nie zawiązające żadnego zdarzenia elementarnego (zbiór pusty), oraz zdarzenie pewne, zawierające zbiór E w całości.

Zbiór zdarzeń elementarnych E może być zbiorem skończonym, przeliczalnym albo mocy continuum. W tym ostatnim wypadku, zbiór zdarzeń Z nie zawiera wszystkich podzbiórów zbioru E , nie zawiera więc wszystkich możliwych zdarzeń. W teorii aksjomatycznej rachunku prawdopodobieństwa ograniczamy się do takiego zbioru Z , który jest *ciałem borelowskim* podzbiórów zbioru E .

Możemy teraz podać co w matematyce rozumiane jest przez pojęcie *zdarzenia losowego*. Zdarzeniem losowym nazywamy każdy element borelowskiego ciała Z .

Borelowskie ciało zbiorów, (σ ciało). Jest to zbiór Z , podzbiórów pewnego zbioru zdarzeń elementarnych E , mający podane niżej własności 1-5. Elementy zbioru Z nazywamy *zdarzeniami losowymi* (Patrz Lit.[4]).

1. Zbiór Z zdarzeń losowych zawiera — jako element — zbiór E zdarzeń elementarnych.
2. Zbiór Z zdarzeń losowych zawiera — jako element — zbiór pusty.
3. Jeżeli zdarzenia A_1 i A_2 należą do zbioru Z , to ich różnica również należy do zbioru Z .
4. Jeżeli zdarzenia A_1, A_2, \dots , w ilości skończonej lub przeliczalnej, należą do zbioru Z , to ich alternatywa również należy do zbioru Z .

$$\bigwedge_{A_i \in Z, i=1..} A = A_1 \cup A_2 \cup \dots \Rightarrow A \in Z$$

5. Jeżeli zdarzenia A_1, A_2, \dots , w ilości skończonej lub przeliczalnej, należą do zbioru Z , to koniunkcja tych zdarzeń również należy do zbioru Z .

$$\bigwedge_{A_i \in Z, i=1..} A = A_1 \cap A_2 \cap \dots \Rightarrow A \in Z$$

Własności 1-2 oznaczają, że zbiór Z zdarzeń losowych zawiera, jako elementy, zdarzenie *pewne* i zdarzenie *niemożliwe*.

W kolejnym kroku, na elementach zbioru $Z(\sigma$ ciała) wprowadzona jest miara P posiadająca określone własności w ramach pewnego układu pewników. Istnieją różne aksjomatyzacje (zestawy pawników) rachunku prawdopodobieństwa, za klasyczną uznano aksjomatyzację zaproponowaną w roku 1933 przez A. Kołmogorowa, podaną poniżej. Niech A, B będą elementami borelowskiego ciała Z , wówczas

1. każdemu zdarzeniu A odpowiada liczba nieujemna zwana jego prawdopodobieństwem

$$P[A] \geq 0$$

2. dla pary zdarzeń A i B (zdarzenia wykluczające się, rozłączne), prawdopodobieństwo zajścia zdarzenia (A lub B) jest równe

$$P[A \vee B] = P[A] + P[B]$$

3. zdarzeniu Ω odpowiada prawdopodobieństwo

$$P[\Omega] = 1$$

Zdarzenie Ω może być uważane za zdarzenie polegające na wystąpieniu dowolnego ze zdarzeń ciała σ .

Za pomocą tych aksjomatów można wydedukować inne bardziej lub mniej złożone twierdzenia i wnioski dotyczące prawdopodobieństw zdarzeń elementarnych i złożeń zdarzeń elementarnych. Zdarzenia złożone mogą być zdarzeniami niewykluczającymi się, czyli takimi, które stanowią przekrywające się zbiory zdarzeń elementarnych. ↗

3.4 Własności prawdopodobieństwa

Niech zdarzenie \bar{A} będzie *zdarzeniem przeciwnym* do zdarzenia A . ponieważ suma $\bar{A} \cup A$ tych zdarzeń wynosi Ω , to z aksjomatu 2 i 3 mamy

$$P[A \vee \bar{A}] = P[A] + P[\bar{A}] = 1$$

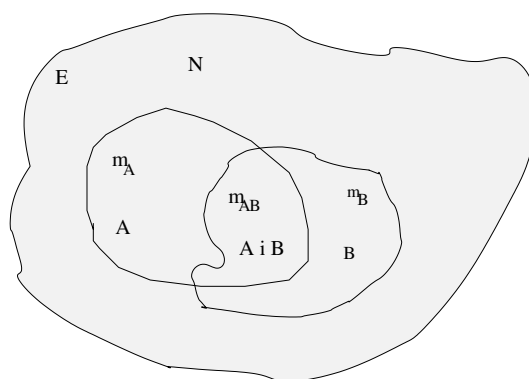
A korzystając jeszcze z aksjomatu 1 mamy, że dziedziną wartości prawdopodobieństwa jest przedział

$$0 \leq P[A] \leq 1$$

Z aksjomatu 2 możemy wydedukować bardziej ogólne twierdzenie dotyczące zdarzeń *wzajemnie rozłącznych* A, B, C, \dots

$$P[A \vee B \vee C \vee \dots] = P[A] + P[B] + P[C] + \dots \quad (3.4)$$

Jednak aksjomatyczne ujęcie prawdopodobieństwa nie pozwala na określenie wartości prawdopodobieństwa wystąpienia jakiegoś zdarzenia A . Do tego celu musimy posłużyć się np. klasycznym podejściem z podrozdziału 3.2. ↗



Rysunek 3.1: Ilustracja graficzna pomocna przy wyprowadzeniu praw dodawania i mnożenia prawdopodobieństw oraz przy wprowadzeniu koncepcji prawdopodobieństwa warunkowego.

3.4.1 Prawdopodobieństwo sumy zdarzeń

Podejście klasyczne wykorzystamy jeszcze do wyprowadzenia wzoru na prawdopodobieństwo sumy zdarzeń, które nie są rozłączne.

Zdarzenie $A \in Z$, czyli podzbiór składający się z pewnej liczby zdarzeń elementarnych E_i jest zdarzeniem nieelementarnym, zdarzeniem złożonym. Wystąpienie zdarzenia złożonego A będzie miało miejsce jeśli wystąpi co najmniej jedno ze zdarzeń elementarnych E_i wchodzących w skład podzbioru A . Prawdopodobieństwo zdarzenia A wynosi $P[A]$. Inne zdarzenie złożone $B \in Z$ będzie również podzbiorem zbioru zdarzeń elementarnych E_i , przy czym pewna ich liczba może także wchodzić do podzbioru definiującego zdarzenie A . Prawdopodobieństwo wystąpienia B wynosi $P[B]$.

Sytuację o jakiej mowa zilustrowano na rysunku 3.1. Zbiór E zawiera pewną skończoną liczbę N zdarzeń elementarnych E_i . Zdarzenie A składa się z m_A zdarzeń elementarnych, zdarzenie B zawiera ich m_B . Liczba punktów objętych częścią wspólną obu obszarów A i B wynosi $m_{A \cap B}$.

Zgodnie z definicją (3.1) prawdopodobieństwo zdarzenia A , czyli zdarzenia polegającego na wyborze dowolnego punktu z obszaru A wynosi

$$P[A] = \frac{m_A}{N}$$

Prawdopodobieństwo zdarzenia B wynosi

$$P[B] = \frac{m_B}{N}$$

Interesuje nas prawdopodobieństwo wystąpienia któregośkolwiek ze zdarzeń A lub B z rysunku (3.1), co zapisujemy jako $P[A \vee B]$. Ponieważ zdarzenia te nie są zdarzeniami rozłącznymi, nie możemy korzystać z aksjomatu 2. Ale gdybyśmy znali liczbę zdarzeń sprzyjających zdarzeniu $A \vee B$, wówczas w celu obliczenia $P[A \vee B]$ mogliśmybyśmy zastosować definicję (3.1). Nietrudno jest ustalić liczbę zdarzeń sprzyjających zdarzeniu $A \vee B$, wynosi ona

$$m_{A \vee B} = m_A + m_B - m_{A \cap B} \quad (3.5)$$

Od sumy liczb zdarzeń sprzyjających zdarzeniom A oraz B należy odjąć liczbę zdarzeń sprzyjających jednoczesnemu ich wystąpieniu, która została uwzględniona zarówno w m_A jak i m_B czyli dwukrotnie.

Gdy podzielimy obie strony równania (3.5) przez N , czyli przez liczbę wszystkich zdarzeń elementarnych w zbiorze E , na mocy definicji (3.1) będzie

$$P[A \vee B] = P[A] + P[B] - P[A \wedge B] \quad (3.6)$$

Jest to wzór na obliczanie sumy dwóch zdarzeń nierozłącznych.

3.4.2 Prawdopodobieństwo warunkowe

Dla zdarzeń złożonych nierozłącznych będących podzbiorami zdarzeń elementarnych, możemy pojęcie prawdopodobieństwa uogólnić do prawdopodobieństwa warunkowego. W tym celu prześledźmy następującą sytuację.

Niech będzie wiadomym, że zaszło zdarzenie A . Ile wówczas wynosi prawdopodobieństwo zajścia również zdarzenia B , nierozłącznego z A ?

Odpowiedź łatwo ustalić za pomocą definicji (3.1). Bowiemy wobec warunku — wystąpiło A — zamiast na całym zbiorze E , koncentrujemy się jedynie na zdarzeniach elementarnych z obszaru A ,³ i tylko przy takim warunku interesuje nas prawdopodobieństwo wystąpienia zdarzenia B . A to oznacza, że interesują nas jedynie te zdarzenia elementarne wchodzące do B , które należą także do A czyli interesują nas zdarzenia z obszaru wspólnego $A \wedge B$.

Biorąc stosunek liczby $m_{A \wedge B}$ zdarzeń elementarnych sprzyjających zdarzeniu $B \wedge A$ do liczby m_A wszystkich zdarzeń elementarnych w A mamy, że prawdopodobieństwo zdarzenia B pod warunkiem wystąpienia A wynosi

$$P[B|A] = \frac{m_{A \wedge B}}{m_A}$$

Podobnie, gdy pytamy o prawdopodobieństwo zajścia A pod warunkiem wystąpienia B , mamy

$$P[A|B] = \frac{m_{A \wedge B}}{m_B}$$

Gdybyśmy zapytali o prawdopodobieństwo $P[A \wedge B]$ wystąpienia zdarzenia $A \wedge B$ (bezw warunkowego), to zgodnie z definicją (3.1) będzie

$$P[A \wedge B] = \frac{m_{A \wedge B}}{N}$$

a rozszerzając ten ułamek o m_A mamy

$$P[A \wedge B] = \frac{m_A}{N} \frac{m_{A \wedge B}}{m_A}$$

³Zdanie — zaszło zdarzenie A oznacza, że wystąpiło któreś ze zdarzeń elementarnych należących do A . Nie interesują nas pozostałe zdarzenia w zbiorze E , bo na pewno żadne z nich nie zaszło. Ale możliwym jest, że jest to zdarzenie elementarne, które oznacza, że zaszło także zdarzenie B .

a więc wobec ustaleń wcześniejszych ostatecznie mamy

$$P[A \wedge B] = P[A]P[B|A] \quad (3.7)$$

$$P[B|A] = \frac{P[A \wedge B]}{P[A]} \quad (3.8)$$

W oparciu o definicję (3.8) możemy zdefiniować *regulę prawdopodobieństwa całkowitego*. Jeśli w pewnym eksperymencie wynik jest jednym z N możliwych zdarzeń A_i wzajemnie się wykluczających, dla których

$$\Omega = A_1 + A_2 + \dots + A_N$$

to jak wynika z równań (3.8) i (3.4) prawdopodobieństwo wystąpienia dowolnego zdarzenia $B \subset \Omega$ wynosi

$$P[B] = \sum_{i=1}^N P[A_i]P[B|A_i] \quad (3.9)$$

Twierdzeniem łączącym $P[A|B]$ i $P[B|A]$ jest *twierdzenie Bayesa*, zgodnie z którym dla zbiorów A i B (składających się ze zdarzeń elementarnych) zachodzi

$$P[A|B] = P[B|A]P[A]/P[B] \quad (3.10)$$

Wzór (3.10) nazywa się *wzorem Bayesa* albo *wzorem na prawdopodobieństwo a posteriori*. Bowiem, określa on prawdopodobieństwo zrealizowania zdarzenia A po wystąpieniu zdarzenia B . Wyrażenie $P[A]$ dotyczy tzw. *prawdopodobieństwa a priori* wystąpienia zdarzenia A .

↗

3.4.3 Niezależność zdarzeń

W oparciu o to co powiedziano wyżej możemy zdefiniować pojęcie *niezależności zdarzeń*. Dwa zdarzenia A i B uważamy za statystycznie niezależne, jeśli fakt wystąpienia zdarzenia A nie pociąga zmiany prawdopodobieństwa wystąpienia zdarzenia B , tzn. gdy

$$P[B|A] = P[B]$$

A korzystając z (3.7) dla zdarzeń niezależnych mamy

$$P[A \wedge B] = P[A]P[B] \quad (3.11)$$

↗

3.5 Zmienna losowa

Prawdopodobieństwo dotyczy zdarzeń losowych będących rezultatem rzeczywistych bądź hipotetycznych zjawisk losowych. Zjawiska te jak np. rzut kostką, pomiar kąta, kończą się w jakiś określony sposób, dając takie czy inne wyniki. Jeśli zjawisko może zakończyć się na wiele różnych sposobów, każdym z tych sposobów przypisujemy pewną liczbę — x zwaną *ex usu* zmienną losową (zmienną stochastyczną).

Od innych zmiennych, zmienna losowa różni się tym, że w stosunku do niej możemy pytać o dwie rzeczy: o to jaką przyjmuje wartość, np. 5 oczek w zjawisku rzutu kostką, ale możemy też zapytać czy wystąpienie 5-ciu oczek zdarza się częściej niż 4-rech, czyli możemy pytać o związane z tymi wartościami prawdopodobieństwa.

Przydziału (przyporządkowania) prawdopodobieństw wartościom zmiennej losowej dokonuje się w różny sposób, ale najwygodniej posłużyć się w tym celu formułą matematyczną, nazywaną funkcją rozkładu prawdopodobieństwa. I w tym sensie, *zmienna losowa jest funkcją*.

Rachunek prawdopodobieństwa i statystykę stosujemy w celu ujęcia losowych własności badanego fragmentu rzeczywistości. Ujęcie to stanowi matematyczny model badanej rzeczywistości określony przez pewien zestaw praw, zmiennych i parametrów. Zbiór wszystkich możliwych elementów badanego fragmentu rzeczywistości jaki należałoby przebadać nazywamy *populacją*. O populacji możemy również pomyśleć jako o zbiorze wszystkich możliwych wartości, które może przyjąć dana zmienna losowa. A więc, populacja to zbiór wszystkich możliwych wyników zjawiska statystycznego związanego z daną zmienną losową. Zbiór ten może być zbiorem skończonym (zmienna losowa dyskretna) jak i nieskończonym (zmienna losowa ciągła).

Tymczasem w praktyce zawsze mamy do czynienia ze skończoną liczbą danych obserwacyjnych, z tzw. *próbą*, która reprezentuje daną populację. Przebadanie wszystkich elementów populacji może być nieopłacalne lub po prostu niemożliwe. Dlatego wybieramy z populacji pewną liczbę elementów i w tak ograniczonej próbie dokonujemy szczegółowych analiz. Na ich podstawie wyciągamy wnioski, formułujemy zdania dotyczące populacji, z której próba pochodzi. Oznacza to, że na uzyskane wnioski, w pewnym stopniu wpływają: sposób pobrania próby i jej rozmiar. Im większy rozmiar próby ⁴ tym bardziej godne zaufania będą uzyskane rezultaty.

W procesie pobierania próby trzeba zapewnić by nie miały miejsca jakiegokolwiek systematyczności, gdyż rezultaty uzyskane z niewłaściwie wybranej próby mogą okazać się nieadekwatne dla całej populacji. By takie zagrożenie oddalić potrzeba pewności, że pobieramy próbę w sposób przypadkowy, tzn. każdy element populacji musi mieć jednakowe szanse zostania wybranym do próby, wybór każdego elementu próby powinien być niezależny od pozostałych wyborów.

Zbiór wszystkich możliwych wartości zmiennej losowej \tilde{x} , razem z odpowiadającymi im prawdopodobieństwami stanowi idealny punkt wyjścia do orzekania czegokolwiek losowego na temat populacji. ↗

⁴Na szczęście, rozmiar próby nie musi być bardzo duży by wnioskowanie w oparciu o próbę dało wiarygodne informacje o całej populacji.

3.5.1 Dystrybuanta

Rozkład prawdopodobieństwa zmiennej losowej określony jest w dwojaki sposób: za pomocą funkcji dystrybucji prawdopodobieństwa (*dystrybutanty*) lub za pomocą *funkcji gęstości* prawdopodobieństwa.

Dystrybuantę $F(x)$ dla zmiennej losowej \tilde{x} , definiuje się jako

$$P[\tilde{x} < x] = F(x) \quad (3.12)$$

Co należy czytać następująco: prawdopodobieństwo, że zmienna losowa \tilde{x} przyjmuje wartości mniejsze od x wynosi $F(x)$.⁵ Ponieważ *ex axiome*, $P[\tilde{x} \dots] \in [0, 1]$, dystrybutanta czyni zadość następującym warunkom brzegowym:

$$\begin{aligned} \lim_{x \rightarrow -\infty} F(x) &= 0 \\ \lim_{x \rightarrow +\infty} F(x) &= 1 \end{aligned} \quad (3.13)$$

Definicja (3.12) dotyczy zarówno dyskretnej jak i ciągłej zmiennej stochastycznej.[†]

3.5.2 Funkcja gęstości

Drugą funkcją wykorzystywaną do obliczenia wartości prawdopodobieństwa jest tzw. funkcja gęstości prawdopodobieństwa $f(x)$. Jest to koncepcja analogiczna do koncepcji gęstości stosowanej w fizyce. Funkcja gęstości pozwala na obliczenie prawdopodobieństwa odpowiadającego pewnemu przedziałowi Δx zmiennej losowej, patrz rysunek 3.2. Zakładając ciągłość i różniczkowalność obu funkcji $F(x)$ i $f(x)$, mamy następujące związki

$$F(x) = \int_{-\infty}^x f(r) dr \quad (3.14)$$

$$f(x) = \frac{\partial F(x)}{\partial x} \quad (3.15)$$

Z równań tych można wydedukować następujące własności:

1. prawdopodobieństwo, że $\tilde{x} \in [x_1, x_2)$ wynosi

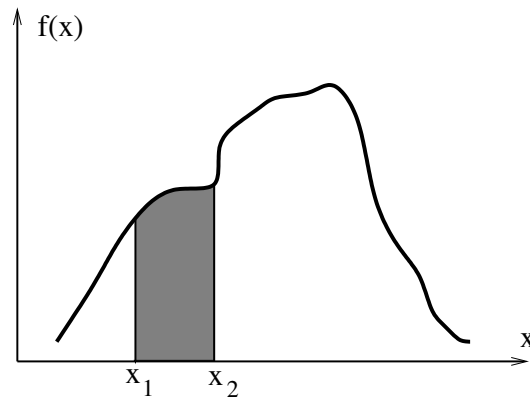
$$P[x_1 \leq \tilde{x} < x_2] = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(r) dr \quad (3.16)$$

Odpowiada ono zaczerpionemu obszarowi na rysunku 3.2.

2. prawdopodobieństwo, że $\tilde{x} < x_1$ wynosi

$$P[-\infty < \tilde{x} < x_1] = F(x_1) - F(-\infty) = \int_{-\infty}^{x_1} f(r) dr \quad (3.17)$$

⁵Symbol x bez tyldy oznaczać będzie wartość bieżącą zmiennej losowej, natomiast zmienną losową będziemy zawsze oznaczać z tyldą, np. \tilde{x} .



Rysunek 3.2: Wyznaczenie prawdopodobieństwa dla zmiennej losowej ciągłej za pomocą rozkładu danego za pomocą funkcji gęstości $f(x)$. Powierzchnia zaczerkionego obszaru odpowiada prawdopodobieństwu, że zmienna \tilde{x} przyjmie wartość z przedziału $[x_1, x_2]$

3. prawdopodobieństwo, że $\tilde{x} \geq x_1$ wynosi

$$P[x_1 \leq \tilde{x} < \infty] = F(\infty) - F(x_1) = \int_{x_1}^{\infty} f(r)dr \quad (3.18)$$

Ponadto o samej funkcji gęstości możemy powiedzieć, że:

1. funkcja gęstości jest nieujemna $f(x) \geq 0$ dla wszystkich x ,
2. funkcja gęstości jest unormowana do jedności

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

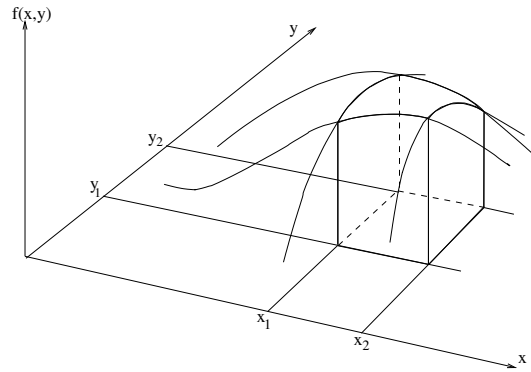
Dowolna funkcja, jeśli ma być funkcją gęstości prawdopodobieństwa musi spełniać oba powyższe warunki.

Funkcja gęstości dotyczy całej populacji i w praktyce określana jest za pomocą kilku liczb (zmiennych) często zwanych parametrami. Funkcja gęstości jest w pełni określona przez te parametry, np. przez wartość średnią i odchylenie standardowe.

3.6 Rozkłady wielowymiarowe

W wielu sytuacjach praktycznych napotykamy na większą liczbą zmiennych losowych. Np. rezultaty obserwacji położenia ciała na sferze (wyznaczenie pary liczb) stanowią zmienną losową dwuwymiarową.

Dla takiej zmiennej niezbędnym jest wprowadzenie nowych koncepcji. Każdą składową losowej pary możemy rozpatrywać niezależnie od pozostałej, bowiem istnieje związany tylko z nią rozkład prawdopodobieństwa. Ale do opisu prawdopodobieństwa łącznego



Rysunek 3.3: Wyznaczenie prawdopodobieństwa zdarzenia dla zmiennej losowej dwuwymiarowej \tilde{x}, \tilde{y} wymaga posłużenia się łączną funkcją gęstości $f(x, y)$. Prawdopodobieństwo jest równe objętości ograniczonej płatem powierzchni $f(x, y)$, płaszczyzną XY oraz czterema pionowymi płaszczyznami $x = x_1, x = x_2, y = y_1, y = y_2$.

wystąpienia danej pary współrzędnych potrzeba czegoś więcej. I tak, dla zmiennej dwuwymiarowej (\tilde{x}, \tilde{y}) mamy dwuwymiarową funkcję dystrybucji $F(x, y)$, pozwalającą na obliczenie łącznego prawdopodobieństwa zdarzenia $\tilde{x} < x$ oraz $\tilde{y} < y$, tzn.:

$$F(x, y) = P(\tilde{x} < x, \tilde{y} < y) \quad (3.19)$$

gdzie x, y stanowią bieżące wartości zmiennej losowej dwuwymiarowej (\tilde{x}, \tilde{y}) .

Funkcja $f(x, y)$ będzie dwuwymiarową łączną funkcją gęstości prawdopodobieństwa. Zakładając jej ciągłość i różniczkowalność mamy:

$$F(x, y) = \int_{-\infty}^x \int_{-\infty}^y f(u, v) du dv \quad (3.20)$$

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} \quad (3.21)$$

Ponieważ $P[\tilde{x}, \tilde{y}] \in [0, 1]$

$$\begin{aligned} \lim_{x, y \rightarrow -\infty} F(x, y) &= 0 \\ \lim_{x, y \rightarrow +\infty} F(x, y) &= 1 \end{aligned} \quad (3.22)$$

Natomiast prawdopodobieństwo, że $x \in [x_1, x_2)$ i $y \in [y_1, y_2)$ wynosi

$$P[x_1 \leq \tilde{x} < x_2, y_1 \leq \tilde{y} < y_2] = \int_{x_1}^{x_2} \int_{y_1}^{y_2} f(u, v) du dv \quad (3.23)$$

Rysunek 3.3 przedstawia geometryczną ilustrację dwuwymiarowego rozkładu prawdopodobieństwa. Prawdopodobieństwo dane równaniem (3.23) reprezentowane jest przez objętość ograniczoną płatem powierzchni $f(x, y)$, płaszczyzną XY oraz czterema pionowymi płaszczyznami $x = x_1, x = x_2, y = y_1, y = y_2$.

Dla przypadku n -wymiarowego, czyli zjawisk opisywanych przez n zmiennych losowych można utworzyć *wektor losowy* $\tilde{\mathbf{x}}$

$$\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n]^T$$

Jego łączną funkcją gęstości będzie

$$f(\mathbf{x}) = f(x_1, x_2, \dots, x_n)$$

Wielowymiarowa funkcja dystrybucji prawdopodobieństwa, czyli dystrybuanta ma postać

$$F(\mathbf{x}) = F(x_1, x_2, \dots, x_n)$$

Prawdopodobieństwo, że wektor losowy $\tilde{\mathbf{x}}$ będzie mniejszy od wektora \mathbf{x} , wynosi:

$$F(\mathbf{x}) = P[\tilde{\mathbf{x}} < \mathbf{x}] = P[\tilde{x}_1 < x_1, \tilde{x}_2 < x_2, \dots, \tilde{x}_n < x_n]$$

Przypadek jednowymiarowy i wielowymiarowy obok pewnych analogii różnią się na tyle, że dla $n \geq 2$ zachodzi konieczność sformułowania dodatkowych pojęć statystycznych jak: rozkład brzegowy, rozkład warunkowy, pojęcie niezależności zmiennych losowych.

3.6.1 Rozkłady brzegowe

Rozkłady brzegowe otrzymujemy z n -wymiarowego rozkładu ($n \geq 2$) poprzez pominięcie rozkładów jednej lub większej liczby składowych wektora losowego $\tilde{\mathbf{x}}$. Np. dwuwymiarowy łączny rozkład dla zmiennych losowych \tilde{x} i \tilde{y} daje się zredukować do jednowymiarowego rozkładu brzegowego dla zmiennej \tilde{x} , gdy zignorujemy jej stochastyczny związek ze zmienną \tilde{y} (o ile taki związek istnieje). Technicznie uzyskuje się to przez przyjęcie w formułach wartości górnego kresu dla ignorowanej zmiennej \tilde{y} , czyli w dystrybuancie $F(x, y)$ (równanie (3.20)) kładziemy $y = \infty$. W rezultacie, dystrybutanta rozkładu brzegowego zmiennej \tilde{x} dana jest jako łączne prawdopodobieństwo przyjęcia przez \tilde{x} wartości mniejszych niż x , a dla \tilde{y} wartości mniejszych od nieskończoności, czyli dowolnych wartości:

$$F_m(x) = F(x, \infty) = P[\tilde{x} < x, \tilde{y} < \infty] = \int_{-\infty}^{x_1} \int_{-\infty}^{\infty} f(u, v) du dv \quad (3.24)$$

Funkcja gęstości łącznego prawdopodobieństwa $f(x, y)$ redukuje się wówczas do formy brzegowej

$$f_m(x) = \int_{-\infty}^{\infty} f(u, v) dv \quad (3.25)$$

co dalej prowadzi do brzegowej dystrybutanty zmiennej \tilde{x} :

$$F_m(x) = \int_{-\infty}^x f_m(u) du \quad (3.26)$$

Rozszerzając tę koncepcję na przypadek n wymiarowy można otrzymać brzegowe rozkłady dla wszystkich kombinacji zmiennych losowych. \uparrow

3.6.2 Niezależność zmiennych losowych

Niech $F(x, y)$ oznacza dystrybutantę łącznego rozkładu prawdopodobieństwa dwuwymiarowej zmiennej losowej \tilde{x} i \tilde{y} . Niech $F_1(x)$ i $F_2(y)$ oznaczają odpowiednie dystrybutanty brzegowe. Zmienne losowe \tilde{x} i \tilde{y} uważamy za *niezależne* jeśli

$$F(x, y) = F_1(x)F_2(y) \quad (3.27)$$

co przy założeniu ciągłości funkcji $F_1(x)$, $F_2(y)$ prowadzi do analogicznego związku pomiędzy odpowiednimi funkcjami gęstości

$$f(x, y) = f_1(x)f_2(y) \quad (3.28)$$

Równanie (3.28) daje się łatwo rozszerzyć na przypadek n wymiarowy, mianowicie:

$$f(x_1, x_2, \dots, x_n) = f_1(x_1)f_2(x_2) \dots f_n(x_n) \quad (3.29)$$

Koncepcja niezależności dwóch lub większej liczby zmiennych losowych jest konsekwencją koncepcji niezależności zdarzeń losowych. Jak pamiętamy, dwa zdarzenia (zjawiska) są niezależne jeśli prawdopodobieństwo zajścia zdarzenia łącznego, równe jest iloczynowi prawdopodobieństw zajścia każdego ze zdarzeń z osobna, patrz równanie (3.11). Ale musimy tu rozróżnić niezależność stochastyczną (losową) od niezależności funkcjonalnej zmiennych. Dwie zmienne mogą być niezależne funkcjonalnie, ale nie muszą być niezależne stochastycznie. ↗

3.6.3 Rozkłady warunkowe

Koncepcja warunkowego rozkładu prawdopodobieństwa, definiowana jest analogicznie do prawdopodobieństwa warunkowego $P[B|A]$ — prawdopodobieństwa wystąpienia zdarzenia losowego B pod warunkiem zajścia zdarzenia A . Analogicznie rozkład warunkowy zmiennych $(\tilde{y} | \tilde{x})$ jest to rozkład zmiennej losowej \tilde{y} przy ustalonej wartości zmiennej \tilde{x} , utworzony z dwuwymiarowego rozkładu zmiennych losowych \tilde{x} i \tilde{y} oraz z funkcji $f_1(x)$ rozkładu brzegowego zmiennej \tilde{x} . Warunkowa funkcja gęstości dla zmiennej \tilde{y} przy ustalonej wartości $\tilde{x} = x$ ma postać

$$f(y | x) = \frac{f(x, y)}{f_1(x)} \quad (3.30)$$

Koncepcje niezależności zmiennych losowych i rozkładu warunkowego są ze sobą powiązane. Jeśli zmienne stochastyczne \tilde{x} i \tilde{y} są niezależne, wówczas warunkowy rozkład zmiennej \tilde{y} przy ustalonej wartości zmiennej \tilde{x} jest taki sam dla dowolnego x , i *vice versa*. Wykorzystując podane wyżej oznaczenia możemy napisać:

$$f(y | x) = \frac{f(x, y)}{f_1(x)} = \frac{f_1(x)f_2(y)}{f_1(x)} = f_2(y) \quad (3.31)$$

Zatem w przypadku niezależności statystycznej, przyjęcie przez zmienną \tilde{x} takiej czy innej wartości, nie ma wpływu na rozkład warunkowy zmiennej \tilde{y} , który jest identyczny z rozkładem brzegowym zmiennej \tilde{y} . ↗

3.7 Parametry rozkładu: nadzieja matematyczna, momenty i korelacja

Funkcja gęstości i funkcja dystrybucji prawdopodobieństwa dają się scharakteryzować za pomocą różnych parametrów, pozwalających na ustalenie ich własności. Jednym z takich parametrów jest wartość średnia zmiennej losowej, które w statystyce znane jest także pod takimi określeniami jak wartość oczekiwana, nadzieja matematyczna, wartość przeciętna.

↑

3.7.1 Wartość oczekiwana

Wartość oczekiwana $E[\tilde{x}]$ (jeśli istnieje), definiowana jest jako pewna typowa wartość μ_x zmiennej losowej \tilde{x} , obliczona z wszystkich możliwych wartości zmiennej \tilde{x} . Typowa w tym znaczeniu, że w danym zjawisku losowym powinniśmy się spodziewać rezultatu bliskiego właśnie tej wartości.

W przypadku zmiennej dyskretnej obliczamy ją jako sumę wszystkich możliwych wartości x_i zmiennej \tilde{x} , pomnożonych przez odpowiadające im prawdopodobieństwa $P[x_i]$

$$E[\tilde{x}] = \mu_x = \sum_{i=1}^n x_i P[x_i] \quad (3.32)$$

Wyrażenie (3.32) pozwala na ustalenie gdzie ma korzenie nazwa wartość średnia zmiennej \tilde{x} . Mianowicie, jeśli mamy n możliwych wartości x_i zmiennej \tilde{x} , o jednakowych prawdopodobieństwach $P[x_i] = 1/n$, wówczas wartość oczekiwana μ_x jest identyczna ze średnią arytmetyczną wartości x_i . Jeśli prawdopodobieństwa $P[x_i]$ nie są sobie równe, równanie (3.32) prowadzi do pojęcia średniej arytmetycznej ważonej. Wagi poszczególnych wartości x_i są w takim przypadku proporcjonalne do odpowiadających im prawdopodobieństw.

Dla zmiennej losowej ciągłej \tilde{x} o ciągłej funkcji gęstości $f(x)$, wartość oczekiwana dana jest za pomocą całki

$$E[\tilde{x}] = \int_{-\infty}^{\infty} x f(x) dx \quad (3.33)$$

Definicję (3.33) można uogólnić np. na przypadek, w którym interesuje nas wartość oczekiwana pewnej funkcji $g(\tilde{x})$ zmiennej losowej \tilde{x} , o funkcji gęstości $f(x)$. Wówczas typowa wartość funkcji $g(\tilde{x})$, jej wartość średnią obliczamy za pomocą wyrażenia

$$E[g(\tilde{x})] = \int_{-\infty}^{\infty} g(\tilde{x}) f(x) dx \quad (3.34)$$

Definicję (3.34) można rozszerzyć na przypadek dwóch i większej liczby zmiennych losowych. Jeśli dla wektora dwuwymiarowego $\tilde{\mathbf{x}} = [\tilde{x}_1, \tilde{x}_2]^T$ o rozkładzie łącznym danym ciągłą funkcją $f(x_1, x_2)$, dana jest funkcja $g(\tilde{x}_1, \tilde{x}_2)$, to jej wartością oczekiwaną jest całka podwójna

$$E[g(\tilde{x}_1, \tilde{x}_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\tilde{x}_1, \tilde{x}_2) f(x_1, x_2) dx_1 dx_2 \quad (3.35)$$

Dla przypadku n wymiarowego będzie,

$$E[g(\tilde{x}_1, \dots, \tilde{x}_n)] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(\tilde{x}_1, \dots, \tilde{x}_n) f(x_1, \dots, x_n) dx_1 \dots dx_n \quad (3.36)$$

Operator wartości oczekiwanej jest operatorem liniowym, stąd jego działanie podlega pewnym pożytecznym prawom, które podajemy bez dowodów:

$$E[c] = c \quad c = const \quad (3.37)$$

$$E[c\tilde{x}] = cE[\tilde{x}] \quad c = const \quad (3.38)$$

$$E[E[\tilde{x}]] = E[\tilde{x}] \quad (3.39)$$

$$E[(\tilde{x} + \tilde{y})] = E[\tilde{x}] + E[\tilde{y}] \quad (3.40)$$

$$E[\tilde{x}\tilde{y}] = E[\tilde{x}]E[\tilde{y}] \quad (\tilde{x} \text{ i } \tilde{y} \text{ niezależne}) \quad (3.41)$$

$$E[\tilde{x}^2] \neq (E[\tilde{x}])^2 \quad (\text{w ogólnosci}) \quad (3.42)$$

Przykład.

Jeżeli $\tilde{x}, \tilde{y}, \tilde{z}$ są niezależnymi zmiennymi stochastycznymi występującymi w funkcji postaci

$$\tilde{w} = 3\tilde{x} + 5\tilde{y}\tilde{z} - 2$$

to jaka jest wartość oczekiwana zmiennej \tilde{w} , jeśli znane są wartości średnie μ_x, μ_y, μ_z ?

Rozwiązanie. Do wartości oczekiwanej $E[\tilde{w}]$ stosujemy własności (3.39)–(3.42)

$$E[\tilde{w}] = \mu_w = E[3\tilde{x} + 5\tilde{y}\tilde{z} - 2] = 3\mu_x + 5\mu_y\mu_z - 2$$

3.7.2 Wariancja

Niech funkcja $g(\tilde{x})$ jest zdefiniowana następująco:

$$g(\tilde{x}) = (\tilde{x} - E[\tilde{x}])^2 = (\tilde{x} - \mu_x)^2 \quad (3.43)$$

Wartość oczekiwana tej funkcji nazywana jest *wariancją* zmiennej losowej \tilde{x}

$$var(\tilde{x}) = \sigma_x^2 = E[g(\tilde{x})] = E[(\tilde{x} - E[\tilde{x}])^2] = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx \quad (3.44)$$

gdzie $f(x)$ jest ciągłą funkcją gęstości zmiennej \tilde{x} .

Pierwiastek kwadratowy wariancji nazywany jest *odchyleniem standardowym*. Obydwa parametry są miarą dyspersji (rozrzutu) zmiennej losowej, co zresztą łatwo zauważyć przyglądając się równaniom (3.43), (3.44) — funkcja, której wartość oczekiwaną obliczamy jest miarą odległości zmiennej \tilde{x} od pewnej wartości ustalonej μ_x .

Korzystając z podanych wcześniej praw, można pokazać, że wariancję da się obliczyć za pomocą formuły

$$\begin{aligned}
 \text{var}(\tilde{x}) &= E[(\tilde{x} - E[\tilde{x}])^2] = E[(\tilde{x} - \mu_x)^2] = E[(\tilde{x}^2 - 2\tilde{x}\mu_x + \mu_x^2)] \\
 &= E[\tilde{x}^2] - 2\mu_x E[\tilde{x}] + E[\mu_x^2] \\
 &= E[\tilde{x}^2] - \mu_x^2 = E[\tilde{x}^2] - (E[\tilde{x}])^2 \\
 \text{var}(\tilde{x}) &= E[\tilde{x}^2] - \mu_x^2 = E[\tilde{x}^2] - (E[\tilde{x}])^2
 \end{aligned} \tag{3.45}$$

3.7.3 Kowariancja

W przypadku zmiennej wielowymiarowej pożytecznymi parametrami są nie tylko wartości średnie i miary dyspersji zmiennych losowych. Ważnymi parametrami są jeszcze kowariancja i korelacja.

Niech \tilde{x} i \tilde{y} będą zmiennymi losowymi o łącznym rozkładzie $f(x, y)$. Niech dalej będzie określona funkcja

$$h(\tilde{x}, \tilde{y}) = (\tilde{x} - E[\tilde{x}])(\tilde{y} - E[\tilde{y}]) = (\tilde{x} - \mu_x)(\tilde{y} - \mu_y)$$

Kowariancja między \tilde{x} i \tilde{y} to wartość oczekiwana tak zdefiniowanej funkcji $h(\tilde{x}, \tilde{y})$

$$\text{cov}(\tilde{x}, \tilde{y}) = \sigma_{xy} = E[h(\tilde{x}, \tilde{y})] = E[(\tilde{x} - \mu_x)(\tilde{y} - \mu_y)] \tag{3.46}$$

co w przypadku zmiennych losowych ciągłych oznacza

$$\sigma_{xy} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy \tag{3.47}$$

Wariancja zmiennej losowej dotyczy jej rozrzutu względem wartości oczekiwanej, kowariancja opisuje współzależność dwóch zmiennych losowych. Odzwierciedla wzajemne związki pomiędzy nimi, albo jak mówimy odzwierciedla wzajemną *korelację*.

Miarą korelacji jest *współczynnik korelacji* ρ definiowany jako

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} = E \left[\frac{(\tilde{x} - E[\tilde{x}])}{\sigma_x} \cdot \frac{(\tilde{y} - E[\tilde{y}])}{\sigma_y} \right] \tag{3.48}$$

gdzie σ_x, σ_y są odchyleniami standardowymi brzegowych rozkładów zmiennych \tilde{x} i \tilde{y} .

Korelacja opisuje współzależność zmiennych losowych. Trzeba jednak wyraźnie odróżnić ten typ współzależności od statystycznej zależności bądź niezależności, określonej za pomocą koncepcji rozkładu warunkowego (patrz równania (3.30), (3.31)). Korelacja i statystyczna zależność to nie to samo, mimo iż oba pojęcia bywają używane jako synonimy.

Można pokazać, że kowariancja (korelacja) zawsze wynosi zero gdy zmienne losowe \tilde{x} i \tilde{y} są statystycznie niezależne. Zgodnie z (3.47) i (3.28) będzie

$$\begin{aligned}
 \sigma_{xy} &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y) dx dy = \\
 &= \int_{-\infty}^{\infty} (x - \mu_x) f_1(x) dx \int_{-\infty}^{\infty} (y - \mu_y) f_2(y) dy = \\
 &= E[(\tilde{x} - \mu_x)] E[(\tilde{y} - \mu_y)] = 0
 \end{aligned} \tag{3.49}$$

Odwrotna własność w ogólnym przypadku nie jest prawdziwa. Zerowa kowariancja nie musi pociągać niezależności statystycznej.

Jednak warto zapamiętać, że dla wielowymiarowego rozkładu normalnego prawdopodobieństwa, zerowa kowariancja (brak korelacji) jest warunkiem wystarczającym dla niezależności statystycznej. ↗

3.7.4 Momenty

Koncepcja wartości oczekiwanej i kowariancji są szczególnymi przypadkami koncepcji ogólniejszych tzw. momentów statystycznych. I tak wartość oczekiwana funkcji $g(\tilde{x}) = (\tilde{x} - c)^k$, gdzie c jest stałą, zwana jest *momentem statystycznym* rzędu k zmiennej losowej \tilde{x} względem c , co piszemy w postaci

$$m_k(\tilde{x}) = E[(\tilde{x} - c)^k] \quad (3.50)$$

gdzie k oznacza rząd momentu.

W przypadku dyskretnej zmiennej losowej wykorzystując definicję (3.32) wartości oczekiwanej, dla $c = 0$ będziemy mieli

$$\bar{m}_k(\tilde{x}) = E[\tilde{x}^k] = \sum_{i=1}^n x_i^k P[x_i] \quad (3.51)$$

a dla zmiennej losowej ciągłej zdefiniowanej wyrażeniem (3.33), będzie

$$\bar{m}_k(\tilde{x}) = E[\tilde{x}^k] = \int_{-\infty}^{\infty} x^k f(x) dx \quad (3.52)$$

Klasa momentów określona wzorem (3.50) nosi miano *momentów centralnych*. Momenty określone formułami (3.51) i (3.52) nazywane są *momentami zwykłymi*.

W grupie momentów centralnych szczególnie ważne znaczenie praktyczne mają te momenty, dla których stała c wynosi

$$c = \mu_x = E[\tilde{x}]$$

czyli, momenty centralne definiowane są jako wartości oczekiwane potęg różnicy zmiennej \tilde{x} i jej wartości średniej

$$m_k(\tilde{x}) = E[(\tilde{x} - E[\tilde{x}])^k] \quad (3.53)$$

Widzimy, że równanie (3.43) jest szczególnym przypadkiem równania (3.53) dla $k = 2$. Wariancja σ_x^2 zmiennej losowej \tilde{x} jest identyczna z momentem centralnym rzędu 2, bowiem $\sigma_x^2 = m_2(\tilde{x})$.

Momenty zmiennej losowej \tilde{x} odzwierciedlają różne własności jej funkcji rozkładu. Jeśli funkcja rozkładu prawdopodobieństwa jest symetryczna względem pierwszego zwykłego momentu $\bar{m}_1(\tilde{x}) = E[\tilde{x}] = \mu_x$, wówczas dla tej zmiennej losowej znikają wszystkie centralne momenty nieparzystego rzędu, tzn. $m_k = 0$ gdy k jest nieparzyste. Z drugiej

strony, jeśli nieparzyste momenty centralne nie równają się zero, ich wartości odzwierciedlają stopień asymetrii lub ukośności funkcji rozkładu. Zatem trzeci moment centralny jest miarą asymetrii rozkładu. W zastosowaniach praktycznych wykorzystywana jest wielkość zwana *skośnością rozkładu* (współczynnikiem asymetrii) obliczana jako

$$\gamma_1 = \frac{m_3(\tilde{x})}{m_2^{1.5}(\tilde{x})} = \frac{m_3(\tilde{x})}{\sigma^3(\tilde{x})} \quad (3.54)$$

Dla rozkładu, w którym po prawej stronie maksimum mamy więcej prawdopodobieństwa niż po stronie lewej, skośność ma wartość dodatnią.

Czwarty moment centralny wykorzystywany jest jako miara *splaszczenia* γ_2 rozkładu

$$\gamma_2 = \frac{m_4(\tilde{x})}{m_2^2(\tilde{x})} - 3 = \frac{m_4(\tilde{x})}{\sigma^4(\tilde{x})} - 3 \quad (3.55)$$

Zamiast splaszczenia możemy napotkać inne określenia jak *kurtoza* (wydęcie), *eksces*. Liczba 3 występująca w definicji (3.55), w przypadku rozkładu Gaussa prowadzi do zerowej wartości kurtozy. Inne rozkłady, bardziej płaskie od rozkładu Gaussa mają splaszczenie ujemne, tzw. rozkłady platykurtyczne. Te z kurtozą dodatnią nazywamy leptokurtycznymi (za lit. [5]).

Definicje momentów można rozciągnąć na przypadki zmiennej wielowymiarowej. Niech (\tilde{x}, \tilde{y}) będzie dwuwymiarowym wektorem losowym. Wartość oczekiwana pewnej ogólnej funkcji $g(\tilde{x}^l, \tilde{y}^k)$ nazywana jest momentem rzędu $(l+k)$. Odpowiadający zmiennym (\tilde{x}, \tilde{y}) moment centralny rzędu $(l+k)$ dany jest jako wyrażenie

$$m_{lk}(\tilde{x}, \tilde{y}) = E[(\tilde{x} - E[\tilde{x}])^l (\tilde{y} - E[\tilde{y}])^k] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - E[\tilde{x}])^l (y - E[\tilde{y}])^k f(x, y) dx dy \quad (3.56)$$

gdzie $f(x, y)$ jest ciągłą funkcją gęstości.

Dla dwuwymiarowej funkcji rozkładu, mamy do dyspozycji trzy momenty centralne rzędu drugiego. Niech \bar{m}_{10} i \bar{m}_{01} będą pierwszymi zwykłymi momentami (wartości średnie) zmiennych \tilde{x} i \tilde{y} , odpowiednio. Wówczas momenty centralne rzędu drugiego tych zmiennych są postaci:

$$\begin{aligned} m_{20} &= E[(\tilde{x} - \bar{m}_{10})^2] = E[(\tilde{x} - \mu_x)^2] = \sigma_x^2 \\ m_{02} &= E[(\tilde{y} - \bar{m}_{01})^2] = E[(\tilde{y} - \mu_y)^2] = \sigma_y^2 \\ m_{11} &= E[(\tilde{x} - \bar{m}_{10})(\tilde{y} - \bar{m}_{01})] = E[(\tilde{x} - \mu_x)(\tilde{y} - \mu_y)] = \sigma_{xy} \end{aligned} \quad (3.57)$$

Momenty te grają szczególną rolę w statystycznych metodach opracowania obserwacji.

W przypadku n -wymiarowego wektora losowego $\tilde{\mathbf{x}}$, momenty centralne drugiego rzędu można tworzyć biorąc wszystkie kombinacje pomiędzy składowymi wektora. Zbiór wszystkich momentów wygodnie jest ująć w formie macierzy zwanej macierzą momentów centralnych drugiego rzędu lub *macierzą wariancji i kowariancji*. Macierz ta ma postać

$$\mathbf{M}_{xx} = \begin{bmatrix} m_{x_1x_1} & m_{x_1x_2} & \cdots & m_{x_1x_n} \\ m_{x_2x_1} & m_{x_2x_2} & \cdots & m_{x_2x_n} \\ \cdots & \cdots & \cdots & \cdots \\ m_{x_nx_1} & m_{x_nx_2} & \cdots & m_{x_nx_n} \end{bmatrix} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1x_2} & \cdots & \sigma_{x_1x_n} \\ \sigma_{x_2x_1} & \sigma_{x_2}^2 & \cdots & \sigma_{x_2x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \sigma_{x_nx_1} & \sigma_{x_nx_2} & \cdots & \sigma_{x_n}^2 \end{bmatrix} \quad (3.58)$$

Macierz (3.58) jest macierzą kwadratową symetryczną bowiem

$$m_{x_i x_j} = E[(\tilde{x}_i - E[\tilde{x}_i])(\tilde{x}_j - E[\tilde{x}_j])] = E[(\tilde{x}_j - E[\tilde{x}_j])(\tilde{x}_i - E[\tilde{x}_i])] = m_{x_j x_i}$$

Macierz \mathbf{M}_{xx} jest kwadratowa i symetryczna gdyż dotyczy jedynie n wymiarowego wektora losowego \tilde{x} . Jeżeli jednak skonstruujemy macierz \mathbf{M}_{xy} dla n wymiarowego wektora \tilde{x} i m wymiarowego wektora \tilde{y} , będzie to macierz o rozmiarach $n \times m$. Nie będziemy mieli tu ani jednej wariancji, wszystkie jej elementy będą kowariancjami,

Symbolem \mathbf{M} będziemy oznaczali macierz kowariancji, symbolem Σ będziemy oznaczali macierz wariancji-kowariancji. \uparrow

3.8 Modalna, mediana i kwantyle

Z matematycznego punktu widzenia, rozkład zmiennej losowej jest w pełni określony przez swoją wartość oczekiwaną $E[\tilde{x}]$, wariancję $var(\tilde{x})$ oraz momenty wyższego rzędu. Mimo to wygodnie jest wprowadzić definicję dalszych parametrów pozwalających lepiej ukazać pewne szczególne własności zmiennej losowej.

3.8.1 Wartość modalna

Wartość modalną x_m definiujemy jako wartość zmiennej losowej \tilde{x} , której odpowiada maksymalna wartości prawdopodobieństwa:

$$P[\tilde{x} = x_m] = \text{maksimum.} \quad (3.59)$$

Gdy funkcja gęstości rozkładu prawdopodobieństwa posiada pierwszą i drugą pochodną, to wartość modalna odpowiadająca maksimum funkcji rozkładu i określona jest przez warunki

$$\frac{df(x)}{dx} = 0, \quad \frac{d^2 f(x)}{dx^2} < 0$$

Bywają rozkłady o jednym maksimum, są to *rozkłady jednomodalne (unimodalne)*, ale bywają też rozkłady o kilku maksimach lokalnych, mówimy wówczas o rozkładach wielomodalnych.

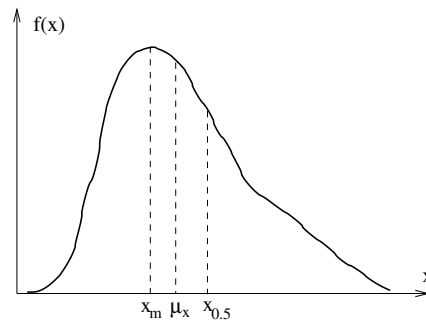
3.8.2 Mediana

Mediana $x_{0.5}$ rozkładu prawdopodobieństwa odpowiada takiej wartości zmiennej losowej \tilde{x} , dla której funkcja dystrybucji rozkładu przyjmuje wartość 0.5, czyli

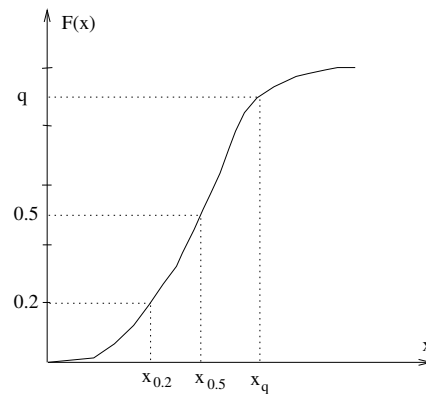
$$F(x_{0.5}) = P[\tilde{x} < x_{0.5}] = 0.5 \quad (3.60)$$

Dla zmiennej losowej ciągłej medianę określamy jako

$$\int_{-\infty}^{x_{0.5}} f(x) dx = 0.5$$



Rysunek 3.4: Asymetryczna funkcja rozkładu prawdopodobieństwa. Dla tego przypadku wartość średnia μ_x , wartość modalna x_m i mediana $x_{0.5}$ nie są sobie równe.



Rysunek 3.5: Zaznaczony na osi odciętych kwantyl (fraktyl) x_q odpowiada na osi rzędnych wartości dystrybuanty q .

Mediana dzieli dziedzinę zmiennej losowej na obszary o równym prawdopodobieństwie.

Z podanych dotąd definicji parametrów rozkładu prawdopodobieństwa wynika, że dla rozkładu jednomodalnego i symetrycznego, mamy

$$\mu_x = x_m = x_{0.5}$$

Własność ta nie jest słuszna dla rozkładów asymetrycznych, wówczas mamy sytuację taką jak na rysunku 3.4.

3.8.3 Kwantyle

Analogicznie jak medianę, można zdefiniować wartości zmiennych losowych odpowiadające innym ustalonym wartościom dystrybuanty $F(x)$, np.:

$$F(x_{0.25}) = P[\tilde{x} < x_{0.25}] = 0.25 \quad (3.61)$$

$$F(x_{0.75}) = P[\tilde{x} < x_{0.75}] = 0.75 \quad (3.62)$$

Wartości $x_{0.25}$ i $x_{0.75}$ nazywane są *dolnym* i *górnym kwantylem*. W podobny sposób możemy zdefiniować *decyle* $x_{0.1}, x_{0.2} \dots x_{0.9}$, *centyle* $x_{0.01}, x_{0.02} \dots x_{0.98}, x_{0.99}$. Wreszcie

jeszcze ogólniej, możemy określić dowolny *fraktyl* x_q jako

$$F(x_q) = \int_{-\infty}^{x_q} f(x)dx = q \quad (0 \leq q \leq 1) \quad (3.63)$$

↗

3.9 Dygresja: odwracanie dystrybuanty

W praktyce obliczeniowej napotykamy potrzebę generowania liczb, dla których rozkład prawdopodobieństwa podany jest np. w postaci histogramu, dystrybuanty (patrz rysunek 3.6). Wartości zmiennej losowej, dla których znamy odpowiadające im rozkłady prawdopodobieństwa (np. dystrybuanty), obliczane są za pomocą algorytmów zwanych odwracaniem dystrybuanty. Obliczenia polegają na znalezieniu dla ustalonego prawdopodobieństwa, odpowiadającej mu wartości kwantyla.

Podstawą tej metody jest następujące przekształcenie zmiennej losowej \tilde{x} w zmienną losową \tilde{y}

$$\tilde{y} = F(x) = \int_{-\infty}^x f(u)du \quad (3.64)$$

Funkcja gęstości prawdopodobieństwa $g(y)$ zmiennej losowej \tilde{y} , zgodnie z prawem propagacji rozkładu gęstości⁶ ma postać

$$g(y) = f(x = h(y)) \frac{dx}{dy} = f(x) \left(\frac{dy}{dx} \right)^{-1} = f(x)(f(x))^{-1} = 1$$

Widzimy, że \tilde{y} ma rozkład ciągły jednostajny. Zatem, wartości y zmiennej \tilde{y} możemy losować za pomocą generatora liczb o rozkładzie jednostajnym, po czym z zależności odwrotnej do (3.64) wyznaczyć odpowiadającą jej wartość x zmiennej losowej \tilde{x} .

↗

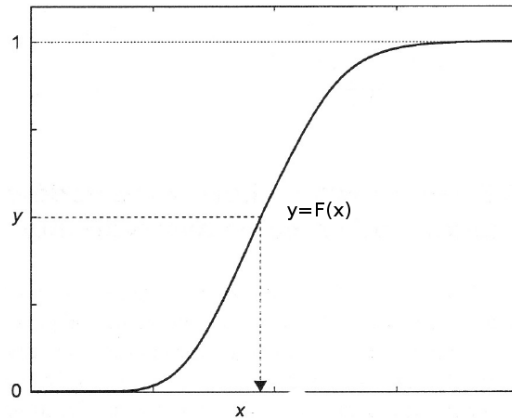
3.10 Dygresja: nierówność Czebyszewa

Wartości zmiennych losowych \tilde{x} grupują się w pobliżu wartości średniej μ_x . W kategoriach prawdopodobieństwa można tę tezę wypowiedzieć następująco: w porównaniu z odchyleniem standardowym σ_x duże odchylenia $(\tilde{x} - \mu_x)$ zmiennej losowej \tilde{x} od wartości średniej μ_x występują rzadziej, są mniej prawdopodobne. Ten łatwy do zauważenia w otaczającej nas rzeczywistości fakt wyraża ilościowo twierdzenie zwane *nierównością Czebyszewa*.

Niech $k \in \mathbb{R}^+$. Prawdopodobieństwo, że absolutne odchylenie $|\tilde{x} - \mu_x|$ przyjmie wartość większą od $k\sigma_x$ jest mniejsze od k^{-2}

$$P[|\tilde{x} - \mu_x| > k\sigma_x] < k^{-2} \quad (3.65)$$

⁶Poznany to prawo w dalszej części wykładu.



Rysunek 3.6: Zmienna losowa \tilde{y} ma rozkład jednostajny $U(0, 1)$. Wylosowanej wartości y odpowiada wartość x , którą można wyznaczyć za pomocą $F(y)^{-1}$, odwrotności dystrybuanty zmiennej losowej \tilde{x} .

Dowód. Mamy pokazać, że prawdziwa jest nierówność (3.65). W tym celu zamiast jej lewej strony wykorzystamy

$$P = P[|\tilde{x} - \mu_x| > k\sigma_x] = P[(\tilde{x} - \mu_x)^2 > k^2\sigma_x^2]$$

Niech $q(t)$ będzie funkcją gęstości prawdopodobieństwa zmiennej $t = (\tilde{x} - \mu_x)^2$, stąd

$$P = P[t > k^2\sigma_x^2] = \int_{k^2\sigma_x^2}^{\infty} q(t)dt$$

Z drugiej strony mamy

$$\sigma_x^2 = \int_{-\infty}^{\infty} tq(t)dt = \int_0^{k^2\sigma_x^2} tq(t)dt + \int_{k^2\sigma_x^2}^{\infty} tq(t)dt$$

Zmienna podcałkowa $t \geq 0$, podobnie jej funkcja gęstości $q(t)$ jest skończoną dodatnią funkcją prawdopodobieństwa, zatem jak poucza nas rachunek całkowy, każda z tych całek będzie co najmniej równa wartości otrzymanej przez podstawienie za t dolnej granicy całkowania. A więc

$$\sigma_x^2 > k^2\sigma_x^2 \int_{k^2\sigma_x^2}^{\infty} q(t)dt = k^2\sigma_x^2 P$$

I ostatecznie mamy

$$P = P[|\tilde{x} - \mu_x| > k\sigma_x] < k^{-2}$$

Co wyczerpuje dowód.

W praktyce powyższa nierówność jest słabsza i w wielu wypadkach będzie

$$P[|\tilde{x} - \mu_x| > k\sigma_x] \ll k^{-2}$$

Opierając się na nierówności Czebyszewa możemy sformułować następujący wniosek praktyczny: jeśli np. w zbiorze obserwacji położenia komety mamy obserwacje, dla których ich odchylenie od wartości obliczonych z orbity wyrównanej jest większe niż 3σ to mamy poważne podstawy sądzić, że obserwacje te są obciążone błędem grubym, a więc należy je z rozważań wykluczyć. [↗](#)

3.11 Zadanka na ćwiczenia

1. Dana jest zmienna losowa \tilde{x} przyjmująca wyłącznie dyskretne wartości:
0,1,2,3,4,5,6,7,8.

Fizyczny sens takiej zmiennej nie jest dla nas istotny, \tilde{x} może być np. błędem pomiarowym. Prawdopodobieństwo związane ze zmienną \tilde{x} dane jest za pomocą zależności funkcyjnej

$$p(x) = 2(x - 0.15)^2 e^{-1.5x}$$

Oblicz wartości prawdopodobieństwa i dystrybuanty odpowiadające poszczególnym niepewnościom pomiarowym. (MC)

↗

Rozdział 4

Przegląd rozkładów, obliczanie prawdopodobieństwa

Streszczenie

Nie ma komu napisać

I nie wiadomo czy będzie komu.

Słowa kluczowe: Rozkłady Bernoulliego: dwumianowy, wielomianowy, rozkład hipergeometryczny, rozkład Poissona, rozkład studenta, rozkład Fishera-Snedecora, rozkład χ^2 , rozkład jednostajny, rozkład normalny: jednowymiarowy, dwuwymiarowy, wielowymiarowy, elipsy i elipsoidy stałego prawdopodobieństwa, centralne twierdzenie graniczne, współczynnik skośności rozkładu, funkcja charakterystyczna. ^a

^a[Modyfikowano AD 2011, kwiecień 01.]

4.1 Rozkłady Bernouliego

4.1.1 Rozkład dwumianowy

Wyobraźmy sobie eksperyment, w którym możliwe są jedynie dwa wzajemnie wykluczające się rezultaty, dwa zdarzenia elementarne: A i \bar{A} , wówczas będzie

$$\Omega = A + \bar{A} \quad (4.1)$$

Niech prawdopodobieństwa tych zdarzeń wynoszą p i q :

$$\begin{aligned} P[A] &= p \\ P[\bar{A}] &= 1 - p = q \end{aligned} \quad (4.2)$$

Dla wygody, wyniki eksperymentu będziemy reprezentowali za pomocą zmiennej losowej \tilde{x}_i przyjmującej dwie wartości liczbowe: $x_1 = 1$ gdy zajdzie zdarzenie A i $x_2 = 0$ gdy zajdzie zdarzenie \bar{A} .

Wyznamy wartość średnią i wariancję zmiennej \tilde{x}_i . Zmienna ta przyjmuje wartość 1 z prawdopodobieństwem p i wartość 0 z prawdopodobieństwem q . Zgodnie z definicją wartości oczekiwanej, dla dyskretnej zmiennej losowej mamy

$$E[\tilde{x}_i] = 1 \cdot p + 0 \cdot q = p \quad (4.3)$$

$$\begin{aligned} var(\tilde{x}_i) &= E[(\tilde{x}_i - p)^2] = (1 - p)^2 \cdot p + (0 - p)^2 \cdot q = \\ &= p - 2p^2 + p^3 + p^2q = p - 2p^2 + p^3 + p^2(1 - p) = \dots \\ var[\tilde{x}_i] &= pq \end{aligned} \quad (4.4)$$

Po takich ustaleniach o naszym eksperymencie nie da się już nic ciekawego powiedzieć: mamy zmienne losowe, mamy wszystkie prawdopodobieństwa, wartość średnią, wariancję.

Skonstruujmy nowy eksperyment, polegający na n -krotnej realizacji doświadczenia wcześniejszego (schemat Bernouliego). W pojedynczym nowym eksperymencie wynik stanowią takie zestawy zdarzeń, n.p.:

$$AAA\bar{A}\bar{A}\bar{A} \dots \bar{A}A$$

W każdym takim zdarzeniu złożonym mamy n składników, n zdarzeń elementarnych, rezultat n prób Bernouliego.

Ekspiryment powtarzamy wielokrotnie i dla naszej wygody jego wynik końcowy przedstawimy w postaci zmiennej losowej

$$\tilde{x} = \sum_{i=1}^n \tilde{x}_i \quad (4.5)$$

gdzie każde \tilde{x}_i przyjmuje wartości 1 albo 0.

Nie jest oczywistym z jakim prawdopodobieństwem mogą zdarzać się konkretne realizacje tak określonego eksperymentu. Jaki rozkład ma zdefiniowana za pomocą wyrażenia (4.5) zmienna losowa \tilde{x} ?

W konkretnej realizacji schematu Bernouliego składającego się z n zdarzeń elementarnych, możliwe jest wystąpienie pewnej liczby zdarzeń A , np. k , i $(n - k)$ zdarzeń \bar{A} . Prawdopodobieństwo zajścia takiej serii zdarzeń, w której na początku wystąpiło k zdarzeń typu A wynosi

$$p^k q^{n-k} \quad (4.6)$$

Wynika to z prawa obliczania prawdopodobieństw zdarzeń niezależnych, a takimi są zdarzenia A i \bar{A} , bo przecież to, że na pierwszym miejscu danej serii wystąpi zdarzenie A nie ma żadnego wpływu na to co będzie na miejscu l -tym.

Oczywistym jest, że w danej serii n zdarzeń niekoniecznie k zdarzeń A musi wystąpić na początku. Mogą pojawiać od miejsca drugiego, trzeciego, \dots , nie muszą też występować jedno za drugim. Mamy tu zatem pewną liczbę możliwości równą liczbie kombinacji k elementów pośród n elementów. Wynosi ona

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

Skoro konkretne serie n zdarzeń są zdarzeniami wzajemnie się wykluczającymi, to prawdopodobieństwo wystąpienia serii z k zdarzeniami typu A jest sumą prawdopodobieństw określonych wyrażeniem (4.6), a więc wynosi ono

$$W_k^n = \binom{n}{k} p^k q^{n-k} \quad (4.7)$$

Pokazaliśmy, że zmienna \tilde{x} , zdefiniowana za pomocą wyrażenia (4.5) przyjmująca wartości z przedziału $0 \leq x \leq n$, ma wartości prawdopodobieństwa określone formułą (4.7). Formuła ta — czyli rozkład prawdopodobieństwa — nosi miano *rozkładu Bernouliego*, *rozkładu dwumianowego*.

Poszukamy wyrażen na wartość średnią i wariancję rozkładu Bernouliego. Za pomocą (4.3) i (4.4) otrzymamy

$$E[\tilde{x}] = E\left[\sum_{i=1}^n \tilde{x}_i\right] = E[\tilde{x}_1] + \dots + E[\tilde{x}_n] = np \quad (4.8)$$

$$\begin{aligned} \text{var}(\tilde{x}) &= E[(\tilde{x} - np)^2] = E\left[\sum_{i=1}^n \tilde{x}_i - np\right]^2 = \\ &= E[(\tilde{x}_1 + \dots + \tilde{x}_n - np)^2] = \\ &= E[((\tilde{x}_1 - p) + (\tilde{x}_2 - p) + \dots + (\tilde{x}_n - p))^2] = \\ &= E[(\tilde{x}_1 - p)^2] + \dots + E[(\tilde{x}_n - p)^2] + E[(\tilde{x}_1 - p)(\tilde{x}_2 - p)] + \\ &\quad E[(\tilde{x}_1 - p)(\tilde{x}_3 - p)] + \dots + E[(\tilde{x}_{n-1} - p)(\tilde{x}_n - p)] \end{aligned}$$

Ponieważ założono, że zmienne losowe \tilde{x}_i są losowo niezależne, stąd wszystkie kowariancje w wyrażeniu powyżej równają się zeru, i ostatecznie mamy

$$\text{var}(\tilde{x}) = npq \quad (4.9)$$

↗

4.1.2 Rozkład geometryczny

Próby Bernoulego możemy wykonywać w schemacie podanym wyżej ale możemy zapytać o prawdopodobieństwo sukcesu (wystąpienia zdarzenia A) w 1-wszej, 2- giej ... w k -tej próbie. Przestrzeń zdarzeń takiego zagadnienia składa się z rozłącznych zdarzeń: $A, \bar{A}A, \bar{A}\bar{A}A, \bar{A}\bar{A}\bar{A}A, \dots$. Łatwo sprawdzimy, że prawdopodobieństwo uzyskania pierwszego sukcesu w k -tej próbie dane jest wzorem

$$G_k = pq^{k-1}, \quad k = 1, 2, 3, \dots \quad (4.10)$$

Rozkład prawdopodobieństwa dany wzorem (4.10) nazywany jest *rozkładem geometrycznym*. Jego wartość oczekiwana i wariancja dane są wzorami

$$\begin{aligned} E[\tilde{k}] &= \frac{1}{p} \\ \text{var}[\tilde{k}] &= \frac{1-p}{p} E[\tilde{k}] \end{aligned} \quad (4.11)$$

↗

4.1.3 Rozkład wielomianowy

Jeśli możliwy jest eksperyment, w którym pojedynczy wynik może przyjąć jedną z $l > 2$ możliwości, np. 6 możliwych rezultatów rzutu kostką, to da się skonstruować zmienną losową Bernoulego o rozkładzie wielomianowym. Zbiór wszystkich zdarzeń elementarnych takiego eksperymentu tworzy układ zupełny postaci

$$\Omega = A_1 + A_2 + \dots + A_l$$

Są to zdarzenia wzajemnie wykluczające się o prawdopodobieństwach

$$P[A_j] = p_j \quad (4.12)$$

przy czym

$$\sum_{j=1}^l p_j = 1$$

Podobnie jak poprzednio, n krotnie powtarzamy eksperyment, którego wynikiem jest któreś ze zdarzeń elementarnych A_j . W rezultacie uzyskamy ciąg postaci $A_1 A_j A_4 A_j \dots A_5 A_l A_3$, w którym zdarzenie A_1 występuje 3 razy, zdarzenie A_2 5 razy etc. Prawdopodobieństwo, że w konkretnym rezultacie eksperymentu, czyli w ciągu n pojedynczych zdarzeń $A_1 A_j A_4 A_j \dots A_5 A_l A_3$ zdarzenie A_j wystąpi k_j razy, $j = 1, 2, \dots, l$ wynosi

$$W_{(k_1, \dots, k_l)}^n = \frac{n!}{\prod_{j=1}^l k_j!} \prod_{j=1}^l p_j^{k_j} \quad (4.13)$$

Rozkład ten nazywamy rozkładem wielomianowym.

Definiując zmienną losową \tilde{x}_{ij} , która w i -tym doświadczeniu ($i = 1, \dots, n$) przyjmuje wartości 1 albo 0, zależnie od tego czy wystąpiło czy nie wystąpiło zdarzenie A_j . Łączny rezultat — ciąg n elementowy $A_1 A_j A_4 A_j \dots A_5 A_l A_3$, w którym A_j wystąpiło k_j razy, możemy zastąpić zmienną losową postaci

$$\tilde{x}_j = \sum_{i=1}^n \tilde{x}_{ij}$$

Można pokazać, że wartość oczekiwana tej zmiennej losowej dana jest formułą

$$E[\tilde{x}_j] = \mu_{x_j} = np_j \quad (4.14)$$

a jej wariancja wynosi

$$\text{var}(\tilde{x}_j) = np_j(1 - p_j) \quad (4.15)$$

↑

4.2 Prawo wielkich liczb

Prawdopodobieństwo p_j wystąpienia elementarnego zdarzenia A_j zwykle nie jest znane i dlatego wyznaczane jest empirycznie. W eksperymencie polegającym na realizacji serii doświadczeń, których rezultatami są zdarzenia A_j , częstość wystąpienia zdarzenia A_j pośród n doświadczeń określona jest wyrażeniem

$$h = \frac{1}{n} \sum_{i=1}^n x_{ij} = \frac{1}{n} x_j \quad (4.16)$$

Zauważmy, że częstość h jest zmienną losową, zależy przecież od wyniku każdego z n losowych doświadczeń.¹

Wykorzystając równania (4.14) i (4.15) oraz własności operatora E , można pokazać, że

$$E[\tilde{h}] = E[n^{-1}\tilde{x}_j] = n^{-1}np_j = p_j \quad (4.17)$$

$$\text{var}[\tilde{h}] = \text{var}[n^{-1}\tilde{x}_j] = n^{-2}\text{var}[\tilde{x}_j] = n^{-1}p_j(1 - p_j) \quad (4.18)$$

Widzimy, że wartość oczekiwana częstości h danego zdarzenia A_j jest równa prawdopodobieństwu jego zajścia. Natomiast jej odchylenie standardowe osiąga małe wartości gdy n jest dostatecznie duże.² Dla skończonych n odchylenie standardowe σ_h jest rzędu $1/\sqrt{n}$, i to jest treścią *prawa wielkich liczb*.

Mamy tu ważne wnioski praktyczne. Często celem doświadczenia jest wyznaczenie prawdopodobieństwa wystąpienia zdarzeń określonego typu. Wzór (4.16) pozwala na wykorzystanie w tym celu częstości danego zdarzenia. Decydując się na takie rozwiązanie, natychmiast wiemy, że kwadrat błędu takiego oszacowania jest odwrotnie proporcjonalny do

¹Prawdopodobieństwa wystąpienia zdarzeń A_j nie są zmiennymi losowymi.

²Iloczyn $p_j(1 - p_j)$ jest zawsze mniejszy od $1/4$.

liczby przeprowadzonych prób. Tego rodzaju błąd jaki popełniamy ze względu na skończoną liczbę prób nosi miano *błędu statystycznego*. Jest to błąd bardzo istotny w tych zastosowaniach, w których zliczane są zdarzenia. ↗

4.3 Rozkład hipergeometryczny

Kolejny rozkład prawdopodobieństwa ustalimy dla zmiennej losowej opisującej zdarzenia polegające na ciągnięciu kul z urny o następującym początkowym składzie: w urnie znajduje się N kul, w tym K kul białych i $L = N - K$ kul czarnych. Interesuje nas prawdopodobieństwo wylosowania w n ciągnięciach dokładnie k białych i $l = n - k$ czarnych kul. W eksperymencie wybierana kula NIE wraca do urny, a zatem nie mamy tu możliwości zastosowania schematu dotyczącego rozkładu dwumianowego.

Wiadomo, że spośród N kul wybieramy n , a więc możemy uczynić to na $\binom{N}{n}$ równoważnych sposobów, bowiem kule danej barwy są nierozróżnialne. Zdarzenie to ma prawdopodobieństwo $1/\binom{N}{n}$. Pośród K kul białych możemy wybierać po k kul białych na $\binom{K}{k}$ sposobów. Podobnie mamy $\binom{L}{l}$ możliwości wyboru l kul z pośród L kul czarnych.

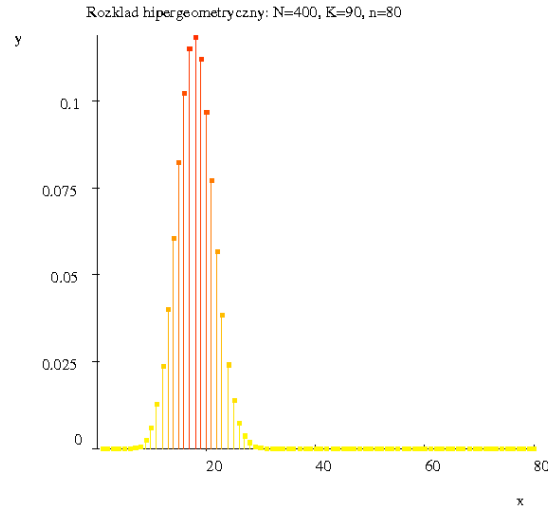
A zatem, skoro liczba możliwych zdarzeń wynosi $\binom{N}{n}$, zdarzeń sprzyjających mamy $\binom{K}{k} \binom{L}{l}$, bowiem każdy z $\binom{K}{k}$ zdarzeń towarzyszyć może któreś z $\binom{L}{l}$ zdarzeń, stąd poszukiwane prawdopodobieństwo wynosi

$$W_k = \frac{\binom{K}{k} \binom{L}{l}}{\binom{N}{n}} \quad (4.19)$$

Rozkład określony wyrażeniem (4.19) nazywany jest *rozkładem hipergeometrycznym*.

Zdefiniujmy zmienną losową $\tilde{x} = \sum_{i=1}^n \tilde{x}_i$, przy czym w i -tym ciągnięciu \tilde{x}_i przyjmuje wartość 1 gdy wylosowano kulę białą i 0 gdy wylosowano kulę czarną. Przechodzimy od koloru kul do zmiennej losowej \tilde{x} i szukamy jej wartości oczekiwanej posługując się definicją

$$E[\tilde{x}] = \sum_{i=1}^n x_i P[x_i]$$



Rysunek 4.1: Przykład rozkładu hipergeometrycznego

A więc korzystając z (4.19)

$$\begin{aligned}
 E[\tilde{x}] &= \sum_{i=1}^n i \cdot W_i = \frac{1}{\binom{N}{n}} \sum_{i=1}^n i \binom{K}{i} \binom{N-K}{n-i} \\
 &= \frac{(N-n)! n!}{N!} \sum_{i=1}^n \frac{i K(K-1)!(N-K)!}{(K-i)! i! (n-i)! (N-K-n+i)!} = \frac{n(n-1)!(N-n)!}{N(N-1)!} \cdot \\
 &\quad \cdot \sum_{i=1}^n \frac{K(K-1)!(N-K)!}{(i-1)! (K-1-(i-1))! (n-1-(i-1))! (N-K-n+1+(i-1))!} = \\
 &= n \frac{K}{N} \frac{(N-n)! (n-1)!}{N-1!} \sum_{j=0}^{n-1} \frac{(K-1)!(N-K)!}{j! (K-1-j)! (n-1-j)! (N-K-n+1+j)!} = \\
 &= n \frac{K}{N} \binom{N-1}{n-1}^{-1} \sum_{j=0}^{n-1} \binom{K-1}{j} \binom{N-K}{n-1-j}
 \end{aligned}$$

Korzystając z własności

$$\binom{n-1}{k} + \binom{n-1}{k-1} = \binom{n}{k}$$

można wykazać, że

$$\sum_{j=0}^{n-1} \binom{K-1}{j} \binom{N-K}{n-1-j} = \binom{N-1}{n-1}$$

A w takim razie

$$E[\tilde{x}] = n \frac{K}{N} \tag{4.20}$$

W podobny sposób można wyprowadzić wyrażenie na wariancję zmiennej \tilde{x}

$$\text{var}[\tilde{x}] = \frac{nK(N-K)(N-n)}{N^2(N-1)} \quad (4.21)$$

Powiedziano wcześniej, że w naszym eksperymencie, w przypadku rozkładu hipergeometrycznego procedura losowania jest inna aniżeli w wypadku eksperymentu z rozkładem dwumianowym. Tym razem wyniki kolejnego losowania wpływają na wyniki losowania następnego. Pobierając z urny kulę nie zwracamy jej, a więc przy następnym ciagnieniu skład urny jest inny. Jednak gdy $n \ll N$ to skład zmienia się nieznacznie i rozkład hipergeometryczny W_k upodobni się do rozkładu dwumianowego, w którym

$$p = \frac{K}{N} \quad q = \frac{N-K}{N} \quad (4.22)$$

Wówczas wartość oczekiwana

$$E[\tilde{x}] = n \frac{K}{N} = np \quad (4.23)$$

a wariancja (4.21), jeśli napiszemy ją w postaci

$$\text{var}[\tilde{x}] = \frac{npq(N-n)}{(N-1)}$$

dla $n \ll N$ da się wyrazić formułą

$$\text{var}[\tilde{x}] = npq \quad (4.24)$$

Podobnie jak to miało miejsce dla rozkładu dwumianowego, możliwym jest uogólnienie rozkładu hipergeometrycznego. Mianowicie, jeśli w populacji o N elementach, każdy element może posiadać jedną z l cech, np. jeden z l kolorów, będziemy wówczas mieli

$$N = N_1 + N_2 + \dots + N_l$$

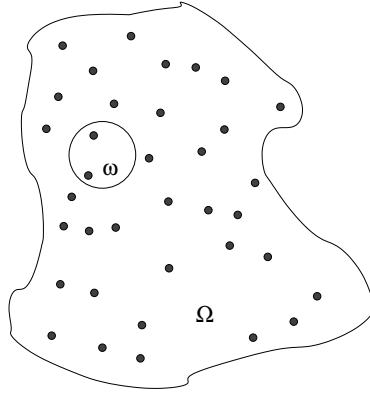
Gdy w takiej populacji wylosujemy $n_j, j = 1, 2, \dots, l$ elementów każdego typu, czyli n_1 elementów typu pierwszego, n_2 elementów typu drugiego etc., to przy n losowaniach prawdopodobieństwo takiego rezultatu wynosi

$$W_{n_1, n_2, \dots, n_l} = \frac{\binom{N_1}{n_1} \binom{N_2}{n_2} \cdots \binom{N_l}{n_l}}{\binom{N}{n}} \quad (4.25)$$

†

4.4 Rozkład Poissona

Rozważmy następujący eksperyment. W obszarze Ω z rysunku 4.2 w jakiś cudowny sposób co pewien czas, niezależnie jeden od drugiego, pojawia się n punktów, które po chwili, w jeszcze bardziej tajemniczy sposób znikają. Interesuje nas prawdopodobieństwo, że w obszarze $\omega \subset \Omega$ pojawi się na skutek takiego pojedynczego zjawiska dokładnie k punktów.



Rysunek 4.2: W obszarze Ω , niezależnie jeden od drugiego, pojawiło się n punktów. Jakie jest prawdopodobieństwo, że w pod-obszarze ω pojawiło się dokładnie k punktów?

Gdybyśmy założyli, że prawdopodobieństwo ukazania się każdego z punktów k w obszarze ω jest takie samo, to mielibyśmy do czynienia z problemem analogicznym do zagadnienia, w którym pytamy o prawdopodobieństwo, że k obiektów wylosowanych spośród n będzie miało pewną cechę, przy czym prawdopodobieństwo posiadania tej cechy przez każdy z k obiektów jest takie samo i wynosi p . Jeśli tak, to w oparciu o to co powiedziano o rozkładzie dwumianowym, poszukiwane przez nas prawdopodobieństwo daje się obliczyć za pomocą formuły

$$W_k^n = \binom{n}{k} p^k (1-p)^{n-k}$$

Tym razem poszukamy postaci tego rozkładu dla przypadku asymptotycznego, w którym $n \rightarrow \infty$ a jednocześnie iloczyn $np = \lambda = \text{const.}$ Zatem będziemy mieli

$$\begin{aligned} W_k^n &= \binom{n}{k} p^k (1-p)^{n-k} = \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \frac{\left(1 - \frac{\lambda}{n}\right)^n}{\left(1 - \frac{\lambda}{n}\right)^k} \\ &= \frac{\lambda^k n(n-1)(n-2)\dots(n-k+1)\left(1 - \frac{\lambda}{n}\right)^n}{k! n^k \left(1 - \frac{\lambda}{n}\right)^k} = \\ &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \underbrace{\frac{\left(1 - \frac{1}{n}\right)\left(1 - \frac{2}{n}\right)\dots\left(1 - \frac{k-1}{n}\right)n^k}{n^k \left(1 - \frac{\lambda}{n}\right)^k}}_{\text{uamek} \rightarrow 1} \end{aligned}$$

A wobec granicy

$$\lim_{n \rightarrow \infty} \left(1 - \frac{\lambda}{n}\right)^n = e^{-\lambda}$$

dla $n \rightarrow \infty$ będzie

$$W_k^n = f(k, \lambda) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (4.26)$$

Rozkład prawdopodobieństwa określony funkcją (4.26) nazywany jest *rozkładem Poissona*.

Rozkład (4.26) odpowiada zmiennej losowej *dyskretnej*, zdefiniowany jest dla całkowitych wartości k i dotyczy możliwości zaobserwowania pewnych konkretnych wartości k .

Prawdopodobieństwo całkowite, czyli prawdopodobieństwo zaobserwowania dowolnego k wynosi 1, i rzeczywiście

$$\sum_{k=0}^{\infty} f(k) = \sum_{k=0}^{\infty} \frac{e^{-\lambda} \lambda^k}{k!} = e^{-\lambda} (1 + \lambda + \frac{\lambda^2}{2!} + \frac{\lambda^3}{3!} + \dots) = e^{-\lambda} e^{\lambda} = 1$$

gdź wyrażenie w nawiasie jest rozwinięciem w szereg funkcji e^{λ} .

Wyznamy parametry rozkładu Poissona, wartość oczekiwaną zmiennej losowej $\tilde{x} = k$

$$E[\tilde{x}] = \sum_{k=0}^{\infty} \frac{k \lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \frac{k \lambda^k}{k!} e^{-\lambda} = \sum_{k=1}^{\infty} \frac{\lambda \lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \underbrace{\sum_{j=0}^{\infty} \frac{\lambda^j}{j!} e^{-\lambda}}_{\text{suma}=1}$$

zatem dla rozkładu Poissona

$$E[\tilde{x}] = \lambda \tag{4.27}$$

W celu wyznaczenia wariancji zmiennej o tym rozkładzie policzmy najpierw wartość oczekiwaną zmiennej \tilde{x}^2

$$\begin{aligned} E[\tilde{x}^2] &= \sum_{k=1}^{\infty} \frac{k^2 \lambda^k}{k!} e^{-\lambda} = \lambda \sum_{k=1}^{\infty} \frac{k \lambda^{k-1}}{(k-1)!} e^{-\lambda} = \lambda \sum_{j=0}^{\infty} (j+1) \frac{\lambda^j}{j!} e^{-\lambda} = \\ &= \lambda \left[\sum_{j=0}^{\infty} j \frac{\lambda^j}{j!} e^{-\lambda} + 1 \right] = \lambda(\lambda + 1) \end{aligned}$$

Za pomocą tego związku natychmiast widzimy, że wariancja

$$\text{var}[\tilde{x}] = E[\tilde{x}^2] - (E[\tilde{x}])^2 = \lambda(\lambda + 1) - \lambda^2 = \lambda \tag{4.28}$$

Współczynnik skośności rozkładu zdefiniowany jest jako stosunek

$$\gamma = \frac{m_3}{\sigma^3}$$

a ponieważ trzeci moment centralny rozkładu Poissona wynosi

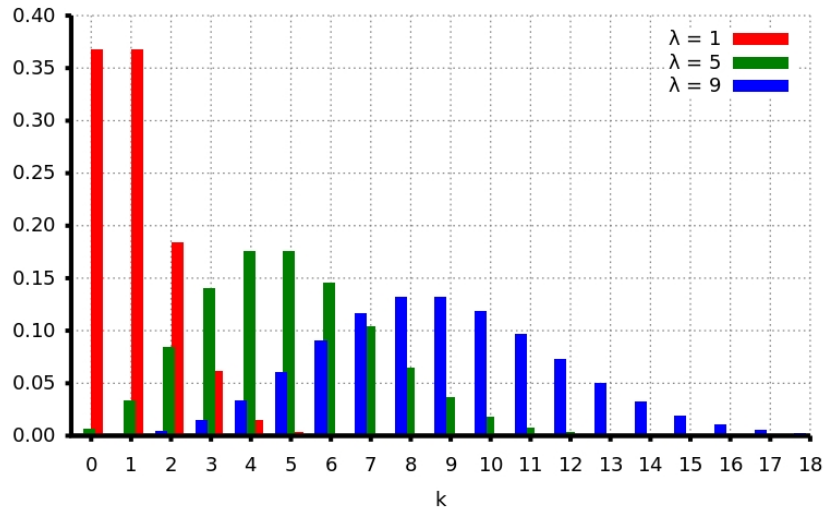
$$m_3 = E[(\tilde{x} - \mu_x)^3] = \lambda$$

wobec tego, współczynnik skośności rozkładu Poissona jest równy

$$\gamma = \frac{\lambda}{\lambda^{3/2}} = \lambda^{-1/2} \tag{4.29}$$

Rozkład Poissona nie jest rozkładem symetrycznym, jednak ze wzrostem wartości parametru λ , gdy $\lambda \gg 1$ staje się symetryczny asymptotycznie.

Ponieważ $\lambda = np$, a rozkład Poissona powstał jako przypadek graniczny z rozkładu dwumianowego przy stałej wartości λ , stąd, dla dużych n mamy małe p . Dlatego rozkład



Rysunek 4.3: Ilustracja rozkładu Poissona zmiennej losowej \tilde{k} dla kilku wartości parametru λ . Ze wzrostem tego parametru rozkład staje się coraz bardziej symetryczny.

ten stosowany jest do zjawisk, procesów, w których występuje dużo zdarzeń (zjawiska masowe) ale niewielki ich ułamek ma interesującą nas własność (interesujące nas zdarzenia zachodzą rzadko). Np. niech od pewnej gwiazdy w interwale czasu Δt , rejestrowanych jest średnio λ fotonów. Wszystkich fotonów wyemitowanych przez gwiazdę było n , prawdopodobieństwo zarejestrowania każdego pojedynczego fotonu wynosi $p \ll 1$. Jeśli zapytamy o prawdopodobieństwo zarejestrowania k fotonów w Δt , będziemy mogli je obliczyć za pomocą rozkładu Poissona $f(k, \lambda)$. ↗

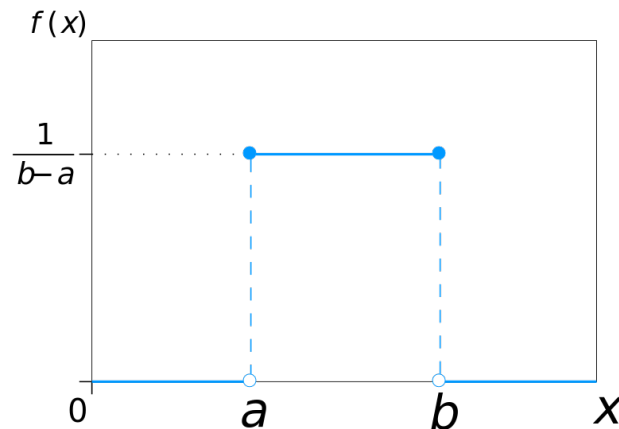
4.5 Rozkład jednostajny (jednorodny)

Omówimy go na przykładzie zmiennej losowej ciągłej. Jest to najprostszy typ rozkładu prawdopodobieństwa, definiujemy go następująco: w przedziale $a \leq x < b$, funkcja gęstości prawdopodobieństwa przyjmuje wartość $c = const$, poza tym przedziałem funkcja jest równa zero

$$\begin{aligned} f(x) &= c & a \leq x < b \\ f(x) &= 0 & x < a, \quad x \geq b \end{aligned}$$

Ponieważ całkowite prawdopodobieństwo wynosi 1, na to by funkcja $f(x)$ była funkcją gęstości prawdopodobieństwa, musi być unormowana do jedności. Osiągamy to poprzez nałożenie warunku

$$\int_{-\infty}^{\infty} f(x) dx = c \int_a^b dx = c(b - a) = 1$$

Rysunek 4.4: Ilustracja rozkładu jednostajnego ciągłej zmiennej losowej \tilde{x} .

A zatem rozkład jednostajny określony jest funkcją gęstości

$$\begin{aligned} f(x) &= \frac{1}{b-a}, & a \leq x < b \\ f(x) &= 0, & x < a, \quad x \geq b \end{aligned} \quad (4.30)$$

Dystrybuanta rozkładu jednorodnego dana jest jako

$$\begin{aligned} F(x) &= \int_a^x \frac{du}{b-a} = \frac{x-a}{b-a}, & a \leq x < b \\ F(x) &= 0, & x < a \\ F(x) &= 1, & x \geq b \end{aligned} \quad (4.31)$$

Natomiast wartość oczekiwana i wariancja wynoszą odpowiednio

$$E[\tilde{x}] = \mu_x = \frac{1}{b-a} \int_a^b x dx = \frac{1}{2} \frac{1}{b-a} [x^2]_a^b = \frac{b^2 - a^2}{2(b-a)} = \frac{b+a}{2} \quad (4.32)$$

$$\begin{aligned} \text{var}[\tilde{x}] &= E[(\tilde{x} - \mu_x)^2] = E[\tilde{x}^2] - 2\mu_x E[\tilde{x}] + \mu_x^2 = E[\tilde{x}^2] - \mu_x^2 = \\ &= \frac{1}{b-a} \int_a^b x^2 dx - \frac{(b+a)^2}{4} = \dots = \frac{(b-a)^2}{12} \end{aligned} \quad (4.33)$$

Z rozkładem jednostajnym rzadko mamy do czynienia w przyrodzie. Jest to jednak ważny rozkład, nawet bardzo ważny ze względu na jego rolę w eksperymentach komputerowych. Rozkład jednostajny transformowany jest w inne rozkłady, tym samym stanowi jądro techniki określanej mianem metoda Monte Carlo. ↗

4.6 Dygresja: funkcja charakterystyczna rozkładu prawdopodobieństwa

Zmienne losowe mogą przyjmować wartości nie tylko ze zbioru liczb rzeczywistych. Równie dobrze można skonstruować *zmienną losową zespoloną* składającą się z pary zmiennych losowych rzeczywistych \tilde{x}, \tilde{y} ,

$$\tilde{z} = \tilde{x} + i\tilde{y}$$

Jej wartość oczekiwaną definiujemy jako

$$E[\tilde{z}] = E[\tilde{x}] + iE[\tilde{y}]$$

Zmienna \tilde{z} jest zmienną losową niezależną gdy jej części rzeczywista i urojona są niezależnymi zmiennymi losowymi.

Niech \tilde{x} oznacza rzeczywistą zmienną losową posiadającą funkcje rozkładu $F(x)$ i $f(x)$. *Funkcję charakterystyczną* rozkładu zmiennej \tilde{x} definiujemy jako wartość oczekiwaną wyrażenia $\exp(it\tilde{x})$,

$$\varphi(t) = E[\exp(it\tilde{x})] \quad (4.34)$$

Gdy \tilde{x} jest zmienną ciągłą to funkcja charakterystyczna ma postać

$$\varphi(t) = \int_{-\infty}^{\infty} \exp(itx)f(x)dx \quad (4.35)$$

Jak widzimy, funkcja charakterystyczna jest *transformatą Fouriera* funkcji gęstości prawdopodobieństwa.

Gdy zmienna losowa \tilde{x} jest zmienną dyskretną wówczas funkcja charakterystyczna jest określona poprzez szereg

$$\varphi(t) = \sum_{k=1}^n \exp(itx_k)P[\tilde{x} = x_k] \quad (4.36)$$

Jak pamiętamy, zgodnie z definicją momenty zwykłe ciągłej zmiennej \tilde{x} wynoszą

$$\bar{m}_n = E[\tilde{x}^n] = \int_{-\infty}^{\infty} x^n f(x)dx$$

Okazuje się, że momenty można także otrzymać po n -krotnym różniczkowaniu funkcji charakterystycznej i obliczeniu wartości n -tej pochodnej w punkcie $t = 0$, mianowicie

$$\varphi^{(n)}(t) = \frac{d^n \varphi(t)}{dt^n} = i^n \int_{-\infty}^{\infty} x^n \exp(itx)f(x)dx$$

a podstawiając $t = 0$, mamy

$$\varphi^{(n)}(0) = i^n \bar{m}_n \quad (4.37)$$

Przesunięcie układu współrzędnych tak by jego początek znalazł się w μ_x , pozwoli nam na wprowadzenie nowej zmiennej \tilde{y} ,

$$\tilde{y} = \tilde{x} - \mu_x$$

Jej funkcją charakterystyczną będzie

$$\varphi_y(t) = \int_{-\infty}^{\infty} \exp(it(x - \mu_x)) f(x - \mu_x) dx$$

natomiast n -ta pochodna tej funkcji dla $t = 0$, z dokładnością do czynnika i^n równa jest n -temu momentowi zmiennej \tilde{x} względem μ_x ,

$$\varphi_y^{(n)}(0) = i^n \bar{m}_n = i^n E[\tilde{y}^n] = i^n E[(\tilde{x} - \mu_x)^n] \quad (4.38)$$

Dla $n = 2$

$$\sigma^2(x) = -\varphi_y''(0)$$

Transformatę Fouriera (4.35) możemy odwrócić, wówczas z funkcji charakterystycznej $\varphi(t)$ otrzymamy funkcję gęstości

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp(-itx) \varphi(t) dt \quad (4.39)$$

Wzór ten możemy stosować przy założeniu, że $\varphi(t)$ jest absolutnie całkowna.

Istnieje twierdzenie orzekające, że funkcja gęstości rozkładu prawdopodobieństwa jest wyznaczona jednoznacznie przez swoją funkcję charakterystyczną o ile funkcja ta jest całkowna absolutnie.

Również w wypadku zmiennej losowej dyskretnej, kiedy to funkcja gęstości nie jest zdefiniowana, rozkład prawdopodobieństwa jest jednoznacznie wyznaczony przez funkcję charakterystyczną. Odpowiednikiem równania (4.39) jest wówczas wyrażenie

$$F(b) - F(a) = \frac{i}{2\pi} \int_{-\infty}^{\infty} \frac{\exp(itb) - \exp(ita)}{t} \varphi(t) dt \quad (4.40)$$

Rozpatrzmy jeszcze sumę niezależnych zmiennych losowych \tilde{x} i \tilde{y}

$$\tilde{w} = \tilde{x} + \tilde{y}$$

Zgodnie z definicją, funkcja charakterystyczna zmiennej \tilde{w} ma postać

$$\varphi_w(t) = E[\exp(it\tilde{w})] = E[\exp(it(\tilde{x} + \tilde{y}))] = E[\exp(it\tilde{x}) \cdot \exp(it\tilde{y})]$$

Ponieważ w wypadku niezależnych zmiennych losowych zespolonych obowiązuje własność $E[\tilde{x} \cdot \tilde{y}] = E[\tilde{x}] \cdot E[\tilde{y}]$, otrzymamy

$$\varphi_w(t) = E[\exp(it\tilde{x})] \cdot E[\exp(it\tilde{y})] = \varphi_x(t) \cdot \varphi_y(t) \quad (4.41)$$

Zatem funkcja charakterystyczna sumy dwóch niezależnych zmiennych losowych jest identyczna z iloczynem ich funkcji charakterystycznych.

O funkcji charakterystycznej mówiliśmy ze względów praktycznych, bowiem w wielu zagadnieniach wygodniej jest posługiwać się tą funkcją aniżeli funkcją rozkładu prawdopodobieństwa. Może się to okazać korzystne np. ze względu na ułatwienia obliczeniowe.

↗

4.7 Rozkład normalny Gaussa

Rozkład Gaussa, inaczej zwany *rozkładem normalnym* jest bodaj najczęściej stosowanym rozkładem w teoriach i zastosowaniach statystycznych. Dla jednowymiarowej zmiennej losowej \tilde{x} , funkcja gęstości tego rozkładu ma postać

$$f(x) = \frac{1}{\sigma_x \sqrt{2\pi}} \exp \left\{ -\frac{(x - \mu_x)^2}{2\sigma_x^2} \right\} \quad (4.42)$$

gdzie wartość średnia μ_x i odchylenie standardowe σ_x są parametrami rozkładu.

W celu ułatwienia obliczeń, wygodnym pociągnięciem jest standaryzacja rozkładu normalnego. O zmiennej losowej \tilde{z}

$$\tilde{z} = \frac{\tilde{x} - \mu_x}{\sigma_x} \quad (4.43)$$

mówimy, że jest standaryzowaną zmienną losową o rozkładzie normalnym, posiada ona funkcję gęstości i dystrybuantę postaci

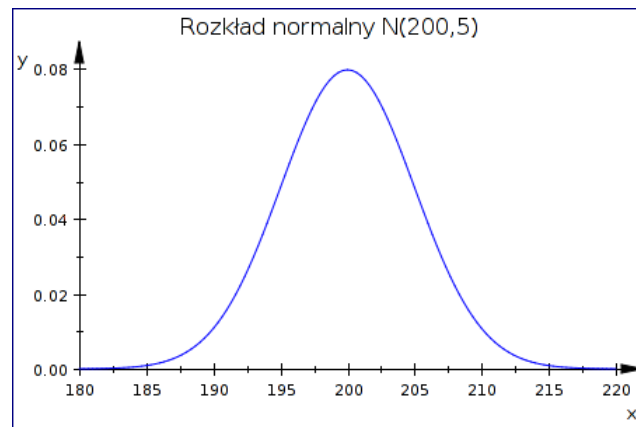
$$\begin{aligned} f(z) &= \frac{1}{\sqrt{2\pi}} \exp \left\{ -\frac{z^2}{2} \right\} \\ F(z) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp \left\{ -\frac{u^2}{2} \right\} du \end{aligned} \quad (4.44)$$

Uproszczona postać tych funkcji bierze się stąd, że dla zmiennej standaryzowanej \tilde{z} mamy $\mu_z = 0$ oraz $\sigma_z = 1$. Jest tak, bowiem

$$\begin{aligned} E[\tilde{z}] &= E\left[\frac{\tilde{x} - \mu_x}{\sigma_x}\right] = \frac{1}{\sigma_x} E[\tilde{x} - \mu_x] = \frac{1}{\sigma_x} (E[\tilde{x}] - \mu_x) = 0 \\ E[(\tilde{z} - \mu_z)^2] &= E\left[\frac{(\tilde{x} - \mu_x)^2}{\sigma_x^2}\right] = \frac{1}{\sigma_x^2} E[(\tilde{x} - \mu_x)^2] = 1 \end{aligned}$$

Funkcja gęstości rozkładu Gaussa (4.5) ma szereg własności :

1. $f(x)$ jest symetryczna względem średniej μ_x . Dlatego dla zmiennej o rozkładzie normalnym wszystkie nieparzyste momenty centralne wynoszą zero. Inne parametry położenia rozkładu jak mediana i modalna są równe średniej μ_x ,
2. maksymalna wartość dla rozkładu standaryzowanego wynosi $f(z) = 0.399$,
3. $f(x)$ dąży asymptotycznie do zera dla $x \rightarrow \pm\infty$,
4. $f(x)$ ma dwa punkty przegięcia w $x = \mu_x \pm \sigma_x$,
5. prawdopodobieństwo, że \tilde{x} przyjmie wartość z przedziału $[x_1, x_2]$ dane jest jako powierzchnia ograniczona krzywą $f(x)$, osią x oraz prostymi $x = x_1$ i $x = x_2$.



Rysunek 4.5: Wykres funkcji rozkładu normalnego o wartości średniej $\mu = 200$ i odchyleniu standardowym $\sigma = 5$. Powierzchnia pod krzywą wynosi 1.0; pomiędzy $\mu \pm \sigma$ wynosi 0.6827; pomiędzy $\mu \pm 2\sigma$ wynosi 0.9545; pomiędzy $\mu \pm 3\sigma$ wynosi 0.9973.

W praktyce szczególną rolę grają wartości prawdopodobieństwa dla odchyień od średniej o krotność wartości σ_x np.:

$$\begin{aligned} P[-\sigma_x < \tilde{x} - \mu_x < \sigma_x] &= 0.6827, \\ P[-2\sigma_x < \tilde{x} - \mu_x < 2\sigma_x] &= 0.9545, \\ P[-3\sigma_x < \tilde{x} - \mu_x < 3\sigma_x] &= 0.9973, \end{aligned} \quad (4.45)$$

bądź prawdopodobieństwa,

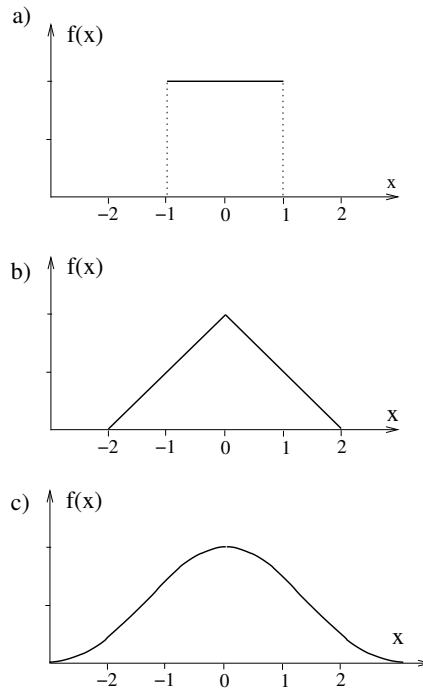
$$\begin{aligned} P[|\tilde{x} - \mu_x| < 1.645 \sigma_x] &= 0.90, \\ P[|\tilde{x} - \mu_x| < 1.960 \sigma_x] &= 0.95, \\ P[|\tilde{x} - \mu_x| < 2.576 \sigma_x] &= 0.99. \end{aligned} \quad (4.46)$$

Teoretyczne i praktyczne znaczenie rozkładu normalnego $\mathcal{N}(\mu_x, \sigma_x)$ ma uzasadnienie w *centralnym twierdzeniu granicznym* mówiącym, że suma losowych zmiennych

$$\tilde{x} = \lim_{n \rightarrow \infty} \sum_{i=1}^n \tilde{x}_i$$

gdzie \tilde{x}_i , $i = 1, 2, \dots, n$ są losowo niezależne, będzie miała asymptotyczny rozkład normalny.

Rysunek 4.6 przedstawia rozkłady kilku zmiennych losowych. W części górnej mamy przypadek pojedynczej zmiennej o rozkładzie jednostajnym, w środku narysowano rozkład zmiennej będącej sumą dwóch niezależnych zmiennych o identycznych rozkładach jednostajnych, u dołu mamy przypadek rozkładu dla sumy trzech takich zmiennych. Widoczna jest tendencja do upodobniania się kształtu tych rozkładów do rozkładu normalnego. ↗



Rysunek 4.6: Ilustracja rozkładu trzech zmiennych losowych: a) zmiennej o rozkładzie jednostajnym, b) sumy dwóch zmiennych, z których każda ma rozkład jednostajny, c) sumy trzech zmiennych o rozkładach jednostajnych. Widoczna jest tendencja do upodobniania się kształtu tych funkcji gęstości do kształtu funkcji rozkładu normalnego.

4.8 Rozkład t studenta

Rozkład zmiennej losowej \tilde{t} często jest stosowany do estymacji wartości określonych za pomocą danych uzyskanych z prób losowych oraz do testowania hipotez statystycznych. W szczególności korzystamy z rozkładu studenta gdy interesują nas odchylenia wartości średniej \bar{x} obliczonej z próby od wartości średniej μ populacji; stawiamy wtedy hipotezę o bliskości \bar{x} i μ .

Niech $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ będą niezależnymi zmiennymi stochastycznymi o identycznych rozkładach normalnych z parametrami μ i σ . *Zmienną losową studenta* \tilde{t} definiujemy jako wyrażenie

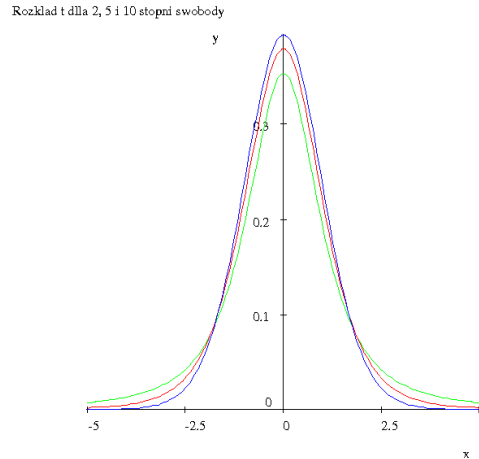
$$t = \frac{\bar{x} - \mu}{s} \sqrt{n} \quad (4.47)$$

gdzie

$$\bar{x} = \frac{1}{n} \sum_{k=1}^n x_k$$

$$s^2 = \frac{1}{n-1} \sum_{k=1}^n (x_k - \bar{x})^2$$

Statystyki \bar{x} oraz s określone są na podstawie próby losowej, są to odpowiednio, średnia



Rysunek 4.7: Przykład rozkładu studenta dla 2, 5, 10 stopni swobody. Dla 2 stopni swobody funkcja ma najniższe maksimum i najpóźniej znika.

próby i odchylenie standardowe próby.

Zmienna studenta \tilde{t} ma rozkład prawdopodobieństwa (*rozkład studenta*) określony funkcją gęstości postaci

$$f(t) = \frac{1}{\sqrt{(n-1)\pi}} \frac{\Gamma\left[\frac{n}{2}\right]}{\Gamma\left[\frac{(n-1)}{2}\right]} \frac{1}{\left(1 + \frac{t^2}{n-1}\right)^{n/2}} \quad (4.48)$$

gdzie Γ jest tzw *funkcją gamma* zdefiniowaną następująco

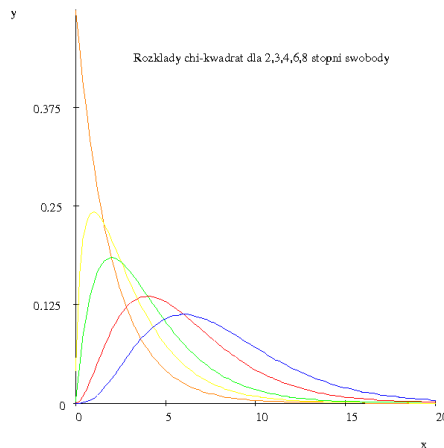
$$\begin{aligned} \Gamma[n] &= \int_0^\infty e^{-t} t^{n-1} dt \\ \Gamma[n+1] &= n \Gamma[n], & \Gamma[1] &= 1 \\ \Gamma[n] &= (n-1)!, & \text{dla } n &\in C^+ \end{aligned}$$

Równanie (4.48) definiuje funkcję gęstości o $(n-1)$ stopniach swobody. Funkcja jest symetryczna ze względu na wartość średnią, jej wykres przypomina wykres funkcji gęstości rozkładu normalnego $\mathcal{N}(\mu, \sigma)$. Podobieństwo to rośnie ze wzrostem liczby stopni swobody i przy $n > 30$ rozkład zmiennej \tilde{t} można po prostu zastąpić rozkładem normalnym. ↗

4.9 Rozkład chi-kwadrat

Niech $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ będą niezależnymi zmiennymi losowymi, każda o rozkładzie $\mathcal{N}(0, 1)$. Wówczas suma ich kwadratów jest zmienną losową oznaczaną tradycyjnie przez χ_n^2

$$\chi_n^2 = \tilde{x}_1^2 + \tilde{x}_2^2 + \dots + \tilde{x}_n^2 \quad (4.49)$$



Rysunek 4.8: Przykład rozkładu χ^2 dla kilku stopni swobody 2, 3, 4, 6, 8. Gdy $n > 7$ krzywe gęstości coraz bardziej przypominają kształtem funkcję gęstości rozkładu normalnego.

Rozkład zmiennej chi-kwadrat nosi miano *rozkładu chi-kwadrat* o n stopniach swobody. Jego funkcja gęstości ma postać

$$\begin{aligned}
 f(\chi_n^2) = f(x) &= c_n \cdot x^{(n-2)/2} e^{-x/2} && \text{dla } x > 0 \\
 f(x) &= 0 && \text{dla } x \leq 0 \\
 c_n &= \left(2^{n/2} \Gamma[n/2]\right)^{-1}
 \end{aligned} \tag{4.50}$$

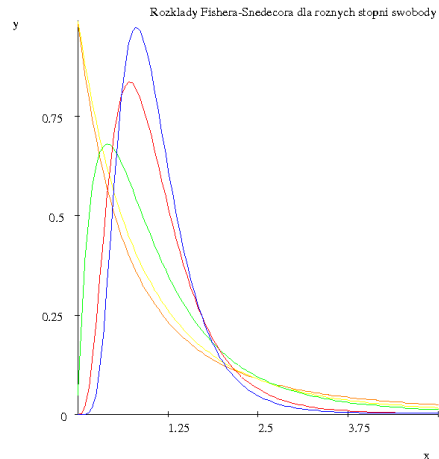
gdzie n jest liczbą stopni swobody.

Rozkład χ^2 ma szczególnie proste momenty:

$$\mu_{\chi_n^2} = E[\chi_n^2] = n \tag{4.51}$$

$$\text{var}[\chi_n^2] = \sigma_{\chi_n^2}^2 = 2n \tag{4.52}$$

Rysunek 4.8 przedstawia wykresy funkcji gęstości rozkładu χ^2 dla kilku stopni swobody. Dla małych n , funkcje te są wyraźnie asymetrycznie (ukośne), ale przy $n > 7$ przypominają funkcję gęstości rozkładu normalnego o średniej n i wariancji $2n$. ↗



Rysunek 4.9: Przykład rozkładu Fishera-Snedecora dla kilku stopni swobody n, m . Dla $n = 2$, krzywa ma kształt funkcji $f(F) = e^{-F}$. Dla $n > 2$ krzywa ma maksimum, którego wartość rośnie ze wzrostem m .

4.10 Rozkład F Fishera-Snedecora

Niech $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_m$ i $\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_n$ będą zbiorami niezależnych znormalizowanych zmiennych losowych o rozkładach $\mathcal{N}(0, 1)$. Rozważmy następujące sumy:

$$\chi_m^2 = \sum_{i=1}^m \tilde{x}_i^2$$

$$\chi_n^2 = \sum_{i=1}^n \tilde{y}_i^2$$

Zmienna losowa $F_{m,n}$

$$F_{m,n} = \frac{\frac{1}{m}\chi_m^2}{\frac{1}{n}\chi_n^2} \quad (4.53)$$

ma rozkład Fishera-Snedecora o m, n stopniach swobody. Jego funkcja gęstości $f_{m,n}(u)$ i dystrybuanta $F_{m,n}(u)$ dla $u > 0$, dane są wzorami

$$f_{m,n}(u) = \frac{\Gamma\left[\frac{m+2}{2}\right]}{\Gamma\left[\frac{m}{2}\right]\Gamma\left[\frac{n}{2}\right]} \binom{m}{n}^{m/2} \frac{u^{(m-2)/2}}{\left(1 + \frac{m}{n}u\right)^{(m+n)/2}} \quad (4.54)$$

$$F_{m,n}(u) = \int_0^u f_{m,n}(t) dt \quad (4.55)$$

Parametry rozkładu Fishera-Snedecora o m, n stopniach swobody mają postać wyrażeń

$$\mu_{F_{m,n}} = \frac{n}{n-2} \quad \text{dla } n > 2 \quad (4.56)$$

$$\sigma_{F_{m,n}}^2 = \frac{2n^2(m+n-2)}{m(n-2)^2(n-4)} \quad \text{dla } n > 4 \quad (4.57)$$

Rozkład F o m, n stopniach swobody, gdy $n \rightarrow \infty$ zbliża się asymptotycznie do rozkładu χ^2 o m stopniach swobody.

W zastosowaniach praktycznych rozkład $F_{m,n}$ jest wykorzystywany do porównania dwóch wariancji uzyskanych z dwóch różnych wyrównań metodą najmniejszych kwadratów. W szczególnych wypadkach porównania dokonuje się między wariancją otrzymaną z wyrównania i daną *a priori*, stosujemy wówczas rozkład $F_{m,\infty} = \chi_m^2$. ↗

4.11 Rozkład normalny wielowymiarowy

W metodzie najmniejszych kwadratów nie jest wymagana znajomość jakiegoś określonego rozkładu prawdopodobieństwa. Ale w następującym po wyrównaniu testowaniu statystycznym najczęściej mamy do czynienia z wielowymiarowym rozkładem normalnym n wymiarowego wektora statystycznego $\tilde{\mathbf{x}}$ o funkcji gęstości

$$f(\mathbf{x}) = \left[\frac{1}{(2\pi)^{n/2} \sqrt{\det[\Sigma]}} \right] \exp \left[-\frac{1}{2} (\mathbf{x} - \mu_{\mathbf{x}})^T \Sigma^{-1} (\mathbf{x} - \mu_{\mathbf{x}}) \right] \quad (4.58)$$

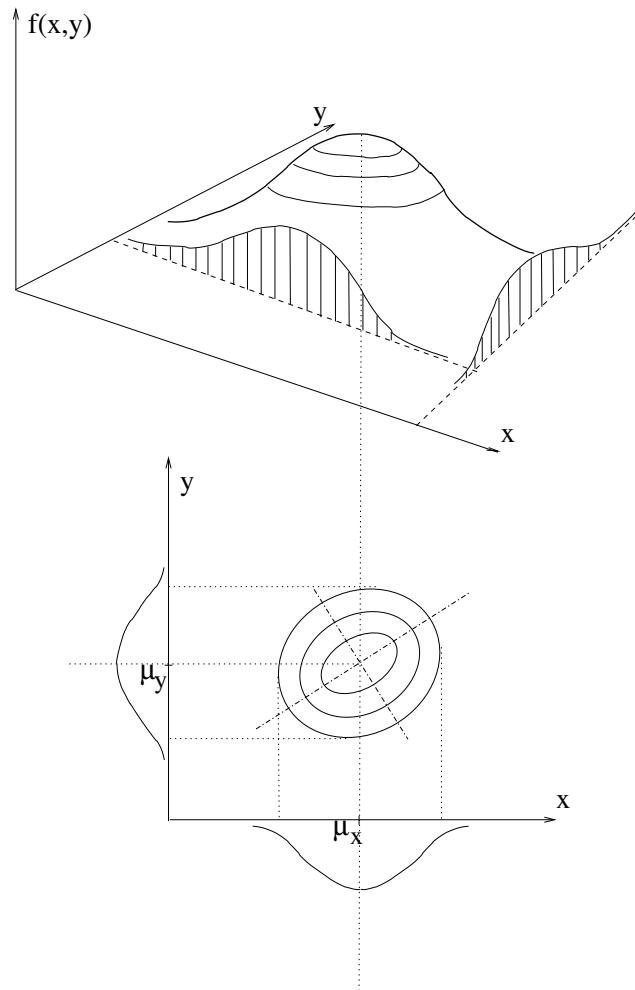
gdzie $\mu_{\mathbf{x}}$ jest wartością oczekiwaną wektora $\tilde{\mathbf{x}}$, a Σ jest macierzą kowariancji.

4.11.1 Rozkład normalny dwuwymiarowy

Dla wektora składającego się z dwóch zmiennych losowych (\tilde{x}, \tilde{y}) , równanie (4.58) w rozwiniętej formie ma postać

$$f(x, y) = \frac{1}{2\pi \sqrt{\sigma_x^2 \sigma_y^2 - \sigma_{xy}^2}} \exp \left\{ -\frac{\sigma_x^2 \sigma_y^2}{2(\sigma_x^2 \sigma_y^2 - \sigma_{xy}^2)} \left[\frac{(x - \mu_x)^2}{\sigma_x^2} - 2\sigma_{xy} \frac{(x - \mu_x)(y - \mu_y)}{\sigma_x^2 \sigma_y^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right] \right\} \quad (4.59)$$

gdzie μ_x, μ_y są średnimi rozkładu, σ_x^2, σ_y^2 są wariancjami zmiennych \tilde{x}, \tilde{y} , natomiast σ_{xy} jest ich kowariancją. Na rysunku 4.10 przedstawiono poglądową ilustrację dwuwymiarowego rozkładu normalnego. Przecięcia powierzchni $f(x, y)$ z płaszczyznami pionowymi prostopadłymi do osi \tilde{x} albo do osi \tilde{y} , przebiegają wzdłuż krzywych będących jednowymiarowymi funkcjami gęstości. Płaszczyzny równoległe do płaszczyzny $\tilde{x}\tilde{y}$ przecinają dzwon rozkładu wzdłuż elips.



Rysunek 4.10: Ilustracja dwuwymiarowego rozkładu normalnego. Płaszczyzny pionowe przecinają powierzchnię rozkładu dwuwymiarowego wzdłuż krzywych gęstości rozkładów jednowymiarowych. Cięcia płaszczyznami poziomymi przebiegają wzdłuż elips.

Dla rozkładu dwuwymiarowego można utworzyć warunkowe funkcje gęstości, n.p.

$$f(y|x) = \frac{1}{\bar{\sigma}\sqrt{2\pi}} \exp\left\{-\frac{(y - \mu(x))^2}{2\bar{\sigma}^2}\right\} \quad (4.60)$$

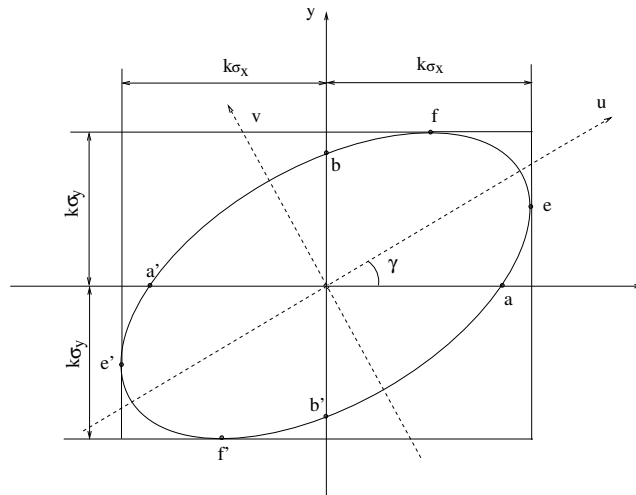
Jak widzimy, wartość średnia dla tego rozkładu jest funkcją zmiennej losowej \tilde{x} , funkcja ta jest postaci

$$\mu(x) = \mu_y + \frac{\rho\sigma_y}{\sigma_x}(x - \mu_x) = \mu_y + \frac{\sigma_{xy}}{\sigma_x^2}(x - \mu_x) \quad (4.61)$$

Jest to liniowa funkcja zmiennej \tilde{x} , zwana linią regresji \tilde{y} ze względu na \tilde{x} . Wariancja $\bar{\sigma}^2$ wynosi

$$\bar{\sigma}^2 = \sigma_y^2(1 - \rho^2) = \frac{\sigma_x^2\sigma_y^2 - \sigma_{xy}^2}{\sigma_x^2} \quad (4.62)$$

gdzie ρ jest współczynnikiem korelacji. ↗



Rysunek 4.11: Elipsa stałego prawdopodobieństwa $h(x, y) = k^2$.

4.11.2 Elipsy stałego prawdopodobieństwa

Przyjmując w równaniu (4.59) $f(x, y) = const$ otrzymamy równanie elipsy, wzdłuż której ma miejsce przecięcie funkcji gęstości z płaszczyzną poziomą. W rezultacie możemy uzyskać całą rodzinę elips o równaniu

$$h(x, y) = \frac{1}{(1 - \rho)^2} \left[\frac{(x - \mu_x)^2}{\sigma_x^2} + 2\rho \frac{(x - \mu_x)(y - \mu_y)}{\sigma_x \sigma_y} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right] = k^2 \quad (4.63)$$

Wspólny środek elips określony jest przez wartości średnie μ_x, μ_y wyznaczone z brzegowych rozkładów dla \tilde{x} i \tilde{y} . Dla uproszczenia dalszych analiz przesuniemy początek układu do punktu (μ_x, μ_y) .

Dla konkretnego k każde z równań $h(x, y) = k^2$ definiuje elipsę, którą da się zamknąć w prostokącie o rozmiarach $2k\sigma_x$ na $2k\sigma_y$ wzdłuż osi Ox i osi Oy , odpowiednio. Rysunek 4.11 przedstawia taką sytuację, pokazana elipsa nosi miano elipsy *stałej gęstości prawdopodobieństwa*. Wartość gęstości tego prawdopodobieństwa zależy od wybranej wartości k .

Jeśli przesuniemy początek układu współrzędnych, a ponadto weźmiemy $k = 1$ równanie elipsy stałego prawdopodobieństwa przyjmie postać

$$\left(\frac{x}{\sigma_x} \right)^2 - 2\rho \left(\frac{x}{\sigma_x} \right) \left(\frac{y}{\sigma_y} \right) + \left(\frac{y}{\sigma_y} \right)^2 = (1 - \rho)^2 \quad (4.64)$$

Elipsa o tym równaniu zwana jest *elipsą standardową*, jej kształt określają parametry σ_x, σ_y oraz ρ .

Kładąc do równania (4.64) $y = 0$, wówczas współrzędne punktów a, a' przecięcia elipsy z osią x-sową wynoszą

$$x = \pm \sigma_x \sqrt{1 - \rho^2}$$

a współrzędne punktów b, b' przecięcia z osią y-ek wynoszą

$$y = \pm \sigma_y \sqrt{1 - \rho^2}$$

Współrzędne te reprezentują pierwiastki kwadratowe odpowiednich warunkowych momentów rzędu drugiego: $m(x|y=0)$ i $m(y|x=0)$, odpowiednio.

Zamykający elipsę standardową prostokąt styka się z krzywą w punktach e, e' o współrzędnych $(\sigma_x, \rho\sigma_y)$ i $(-\sigma_x, -\rho\sigma_y)$, oraz f, f' o współrzędnych $(\rho\sigma_x, \sigma_y)$ i $(-\rho\sigma_x, -\sigma_y)$.

Korzystając z równania (4.64) i równania osiowego elipsy można pokazać, że jej półosie dane są formułami

$$a^2 = \frac{1}{2}(\sigma_x^2 + \sigma_y^2) + \sqrt{\frac{1}{4}(\sigma_x^2 - \sigma_y^2)^2 + \sigma_{xy}^2} \quad (4.65)$$

$$b^2 = \frac{1}{2}(\sigma_x^2 + \sigma_y^2) - \sqrt{\frac{1}{4}(\sigma_x^2 - \sigma_y^2)^2 + \sigma_{xy}^2} \quad (4.66)$$

Przy czym współczynnik korelacji ρ z równania (4.64) zastąpiono tu jego definicją $\rho = \sigma_{xy}/\sigma_x\sigma_y$.

Można pokazać, że półosie a, b z równań (4.65) i (4.66) są pierwiastkami kwadratowymi wartości własnych macierzy kowariancji zmiennych \tilde{x}, \tilde{y}

$$\Sigma = \begin{bmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{bmatrix} \quad (4.67)$$

Wielomianem charakterystycznym dla tej macierzy jest

$$(-\lambda)^2 + tr[\Sigma](-\lambda) + det[\Sigma] = 0$$

gdzie $tr[\Sigma]$ jest śladem macierzy Σ , wynosi on $(\sigma_x^2 + \sigma_y^2)$, $det[\Sigma]$ oznacza wyznacznik macierzy kowariancji Σ , jest on równy $(\sigma_x^2\sigma_y^2 - \sigma_{xy}^2)$.

Wielomian charakterystyczny ma zatem postać

$$\lambda^2 - (\sigma_x^2 + \sigma_y^2)\lambda + (\sigma_x^2\sigma_y^2 - \sigma_{xy}^2) = 0 \quad (4.68)$$

I właśnie pierwiastki tego równania dane są wzorami (4.65) i (4.66).

Kąt γ między półosią wielką a elipsy i osią Ox dany jest zależnością

$$\tan 2\gamma = \frac{2\sigma_{xy}}{\sigma_x^2 - \sigma_y^2} \quad (4.69)$$

Właściwa ćwiartka dla kąta 2γ daje się wyznaczyć w oparciu o badanie znaku wartości licznika i mianownika tego wyrażenia. Wartość $\sin 2\gamma$ ma taki sam znak jak σ_{xy} wartość $\cos 2\gamma$ ma znak identyczny z $\sigma_x^2 - \sigma_y^2$.

Na rysunku 4.11 współrzędne (u, v) , odpowiadają układowi obróconemu o kąt γ względem układu współrzędnych (x, y) . Osie Ou i Ov pokrywają się z osiami elipsy. Współrzędne (u, v) dają się wyznaczyć ze współrzędnych (x, y) za pomocą transformacji obrotu o kąt γ , wokół osi prostopadłej do płaszczyzny elipsy.

W nowym układzie współrzędnych zmienne losowe (\tilde{u}, \tilde{v}) posiadają brzegowe odchylenia standardowe σ_u, σ_v , poza tym jak wynika z wcześniejszej dyskusji, mamy, że $\sigma_u =$

$a, \sigma_v = b$, wówczas u, v są to zmienne stochastyczne nieskorelowane dla których $\sigma_{uv} = 0$, a ponieważ mamy do czynienia z rozkładem normalnym są to także zmienne statystycznie niezależne.

Zatem, jeśli tylko $\sigma_u \neq \sigma_v$, to za pomocą transformacji obrotu można zastąpić parę skorelowanych zmiennych losowych przez parę nieskorelowaną! Jest to własność ogólna zatem, zawsze można zastąpić zbiór skorelowanych zmiennych losowych przez inny zbiór (o tym samym wymiarze) zmiennych losowych nieskorelowanych.

Macierz tej transformacji można skonstruować przez wyznaczenie znormalizowanych wektorów własnych macierzy kowariancji oryginalnych zmiennych losowych i wstawienie ich jako kolumn nowej macierzy. Ta procedura jest równoważna z diagonalizacją oryginalnej macierzy kowariancji, bowiem otrzymana w rezultacie macierz kowariancji nowego nieskorelowanego zbioru zmiennych losowych zawsze jest diagonalna. Jej elementy są wartościami własnymi oryginalnej macierzy kowariancji.

Wobec braku korelacji $\rho = 0$, funkcja gęstości (4.59) może być napisana w postaci

$$\begin{aligned} f(x, y) &= \frac{1}{2\pi\sigma_x\sigma_y} \exp \left\{ -\frac{1}{2} \left(\frac{(x - \mu_x)^2}{\sigma_x^2} + \frac{(y - \mu_y)^2}{\sigma_y^2} \right) \right\} = \\ &= \left\{ \frac{1}{\sigma_x\sqrt{2\pi}} \exp \left(-\frac{(x - \mu_x)^2}{2\sigma_x^2} \right) \right\} \left\{ \frac{1}{\sigma_y\sqrt{2\pi}} \exp \left(-\frac{(y - \mu_y)^2}{2\sigma_y^2} \right) \right\} = \\ &= f_x(x) \cdot f_y(y) \end{aligned}$$

gdzie $f_x(x), f_y(y)$ są jednowymiarowymi rozkładami normalnymi o średnich μ_x, μ_y i wariancjach σ_x^2, σ_y^2 .

Czyli, gdy $\rho = 0$, łączna funkcja gęstości $f(x, y)$ może być wyrażona jako iloczyn brzegowych funkcji gęstości zmiennych \tilde{x} i \tilde{y} . Potwierdzamy w ten sposób wcześniejszą uwagę, że dla rozkładu normalnego brak korelacji oznacza także statystyczną niezależność.

Dysponując równaniem elipsy standardowej (4.64) wyznaczmy prawdopodobieństwo, że losowy wektor (\tilde{x}, \tilde{y}) opisuje punkt znajdujący się we wnętrzu elipsy. Obliczenie tego prawdopodobieństwa upraszcza się gdy pracujemy ze zmiennymi (\tilde{u}, \tilde{v}) , bowiem układ (\tilde{u}, \tilde{v}) pokrywa się z osiami elipsy. Jeśli jeszcze weźmiemy zmienne znormalizowane \tilde{u}/σ_u i \tilde{v}/σ_v , to zgodnie z określeniem (4.49), suma ich kwadratów ma rozkład χ^2_2 o dwóch stopniach swobody.

W rezultacie, prawdopodobieństwo znalezienia punktu określonego za pomocą wektora (\tilde{u}, \tilde{v}) wewnątrz elipsy o osiach $k\sigma_u$ i $k\sigma_v$ ma postać

$$P \left[\left(\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right) < k^2 \right] = P [\chi^2_2 < k^2] = 1 - \alpha \quad (4.70)$$

Dla $k = 1$ (elipsa standardowa) wartość $(1 - \alpha)$ wynosi 0.3935, zatem prawdopodobieństwo, że punkt "wpadnie" do elipsy jest równe 0.3935.

I odwrotnie, wychodząc z przedziału ufności, np. wybierając $\alpha = 0.05$, możemy wyznaczyć wartość mnożnika k .

$$P [\chi^2_2 < \chi_{0.05,2}] = P [\chi^2_2 < 5.94] = 0.95$$

co daje $k = \sqrt{5.94} = 2.447$, a elipsa odpowiadająca temu prawdopodobieństwu ma półosie $a = 2.447\sigma_u$, $b = 2.447\sigma_v$.

Zwróćmy jeszcze uwagę na fakt, iż dla brzegowych rozkładów normalnych jednowymiarowych, prawdopodobieństwo, że każda zmienna losowa \tilde{x} lub \tilde{y} z osobna były w obszarze $\pm\sigma_x, \pm\sigma_y$ wynosiło 0.683. Tymczasem prawdopodobieństwo łącznego zjawiska, punkt (x, y) znajduje się we wnętrzu standardowej elipsy wynosi tylko 0.394. \uparrow

4.11.3 Rozkład 3-wymiarowy, elipsoidy stałego prawdopodobieństwa

Biorąc wykładnik występujący w równaniu (4.58), warunek analogiczny do warunku (4.63) na elipsy odpowiadające stałej gęstości prawdopodobieństwa ma postać

$$(\tilde{\mathbf{x}} - \mu_x)^T \Sigma^{-1} (\tilde{\mathbf{x}} - \mu_x) = k^2 \quad (4.71)$$

Jest to dodatnio określona forma kwadratowa, reprezentująca rodzinę hiperelipsoid stałego prawdopodobieństwa.

Ważnym przypadkiem takich elipsoid jest przypadek trójwymiarowy, gdyż często pojawia się w zastosowaniach, np. przy wyznaczaniu położenia obiektu w przestrzeni. Wówczas równanie elipsoidy (dla prostoty kładziemy wektor wartości średnich $\mu_x = \mathbf{0}$) ma postać

$$h(x, y, z) = [x, y, z]^T \Sigma^{-1} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = k^2 \quad (4.72)$$

Dla $k = 1$, mamy tzw. elipsoidę standardową, a jej półosie (a, b, c) dają się wyznaczyć przez diagonalizację macierzy kowariancji Σ

$$\begin{bmatrix} \sigma_u^2 & 0 & 0 \\ 0 & \sigma_v^2 & 0 \\ 0 & 0 & \sigma_w^2 \end{bmatrix} = \begin{bmatrix} a^2 & 0 & 0 \\ 0 & b^2 & 0 \\ 0 & 0 & c^2 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \mathbf{T}^T \Sigma \mathbf{T}$$

gdzie \mathbf{T} jest ortogonalną macierzą, której kolumny są znormalizowanymi wektorami własnymi macierzy Σ , liczby $\lambda_1, \lambda_2, \lambda_3$ są wartościami własnymi macierzy Σ , natomiast osie Ou, Ov, Ow tworzą obrócony układ współrzędnych, taki, że zmienne losowe u, v, w są nieskorelowane.

W sposób podobny do przypadku dwuwymiarowego można pokazać na czym polega przydatność takiej elipsoidy. Mianowicie, prawdopodobieństwo, że punkt leży we wnętrzu elipsoidy o półosiach $a = k\sigma_u, b = k\sigma_v, c = k\sigma_w$ wyrazić można jako

$$P \left[\left(\frac{u^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} + \frac{w^2}{\sigma_w^2} \right) < k^2 \right] = P [\chi_3^2 < k^2] = 1 - \alpha \quad (4.73)$$

Dla elipsoidy standardowej $(1 - \alpha) = 0.199$. Tabela 4.1 ilustruje jak zmienia się to prawdopodobieństwo dla różnych wymiarów n , elipsoidy \uparrow

Tablica 4.1: Prawdopodobieństwo znalezienia punktu we wnętrzu hiperelipsoidy standardowej dla kilku wymiarów wektorów losowych.

n	1	2	3	4	5	6
P	0.683	0.394	0.199	0.090	0.037	0.014

4.12 Podsumowanie

Podsumowując dyskusję rozkładu normalnego wielowymiarowego, warto zwrócić szczególną uwagę na koncepcję korelacji. Jak powiedziano, opisuje ona pewną zależność między wartościami x i y zmiennych losowych \tilde{x} i \tilde{y} . Nie jest to jednak zależność ścisła, funkcyjna. Jest ona ścisła jedynie dla wartości średnich obu zmiennych losowych, kiedy to zależność dana jest w postaci linii regresji. Są to zawsze dwie linie gdyż wartość średnia y -ka w funkcji x jest inna niż wartość średnia x -sa w funkcji y .

Korelacja dla bieżących rzeczywistych wartości x i y , wyraża jedynie tendencję związku między nimi. Nie jest ona ścisła dla indywidualnych par wartości x i y . Jednak im wyższa korelacja, tym większa będzie tendencja na zrealizowanie związku regresyjnego. ↗

4.13 Zadanka na ćwiczenia

1. Dla zmiennej losowej \tilde{x} o rozkładzie jednostajnym $\mathcal{U}(0, 1)$ oblicz wartości prawdopodobieństwa przyjęcia przez tę zmienną wartości z rzedziału $E[\tilde{x}] \pm \sigma_x$, $E[\tilde{x}] \pm 2\sigma_x$ oraz $E[\tilde{x}] \pm 3\sigma_x$. Porównaj pierwszy wynik z prawą stroną równania (2.10) z rozdziału 2.

↗

Rozdział 5

Estymacja parametrów populacji

Streszczenie

Nie ma komu napisać

I nie wiadomo czy będzie komu.

Słowa kluczowe: Populacja a próba losowa, przedział klasowy, częstości klasowe, częstości względne, wykres słupkowy, histogram, stereogram, częstości brzegowe, korelacja, estymacja punktowa, kryteria dobrych estymatorów, dokładność a precyzja, sposoby wyznaczania dobrych estymatorów: metoda momentów, metoda największej wiarygodności, estymacja punktowa, centralne twierdzenie graniczne, estymacja przedziałowa, poziom ufności, przedział ufności, estymacja przedziałowa dla wartości średniej, dla wariancji i stosunku wariancji. ^a

^a[Modyfikowano AD 2010, Maj 11.]

5.1 Wstęp

Przyjmijmy, że mamy do ustalenia wartość pewnej mierzalnej cechy pewnej populacji, np. interesuje nas średnia liczba gwiazd podwójnych możliwych do obserwacji teleskopem Zeiss 20/3000 w ciągu jednego roku. Problem ma charakter statystyczny, zatem trzeba go rozwiązać posługując się metodami statystycznymi i tym celu daną sytuację fizyczną opiszemy za pomocą postulowanego modelu matematycznego. Elementy modelu traktowane są jako zmienne losowe, którym przypisano pewien łączny rozkład prawdopodobieństwa. Słowo “przypisano” oznacza, że rozkład ustalono na podstawie pewnych przesłanek np. teoretycznych, bądź w oparciu o wcześniej wykonane badania.

Wartości zmiennych losowych, np. liczba gwiazd obserwowana w polu widzenia teleskopu są konkretnymi realizacjami losowych zdarzeń. Takie realizacje można powtarzać wielokrotnie, jednak liczba powtórzeń jest zawsze ograniczona. Przecież nie możemy bez końca wykonywać np. obserwacje gwiazd podwójnych. Uzyskane w danej serii wykonywanego doświadczenia wartości zmiennych losowych tworzą tzw. *próbę losową*. Na podstawie analizy prób usiłujemy dociec jaki jest rozkład prawdopodobieństwa populacji, a przynajmniej niektóre jego parametry.

Po oszacowaniu (estymacji, obliczeniu) parametrów populacji należy jeszcze ustalić wiarygodność tego oszacowania. W tym celu, obok samych parametrów wyznaczane są odpowiadające im przedziały ufności. Inne formy oceny jakości parametrów np. statystyczne testy, dotyczą pytania czy rezultaty estymacji pozostają w zgodzie z początkowymi założeniami (hipotezami). ↗

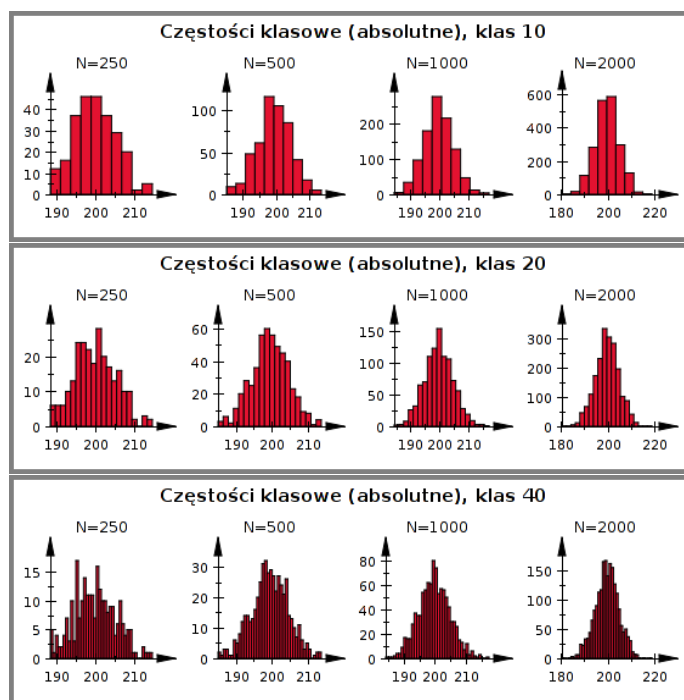
5.2 Sposoby opisu próby statystycznej

Dowolny skończony zbiór pomiarów (x_1, \dots, x_n) nazywamy próbą. Próba stanowi podzbiór wzięty z populacji n -wymiarowego losowego wektora \tilde{x} . Bywa, że próby pobierane są w celu ustalenia rozkładu prawdopodobieństwa wszystkich możliwych wartości wektora \tilde{x} . Jednak częściej za pomocą prób obliczane są jedynie parametry znanej albo postulowanej postaci tego rozkładu.

Dysponując losową próbą, w pierwszej kolejności warto poddać ją podstawowej analizie danych, polegającej na:

- konstruowaniu histogramów,
- wyznaczeniu wartości statystyk (estymatorów) położeniowych próby (średnia, mediana, modalna),
- obliczeniu wartości statystyk (estymatorów) rozproszeniowych (dyspersyjnych) próby (wariancje, kowariancje),
- obliczeniu statystycznych momentów wyższych rzędów.

↗



Rysunek 5.1: Histogramy częstości absolutnych rezultatów pomiaru pewnej wielkości fizycznej. Histogram w lewym górnym rogu odpowiada danym z Tabeli 5.1.

5.2.1 Histogramy, stereogramy

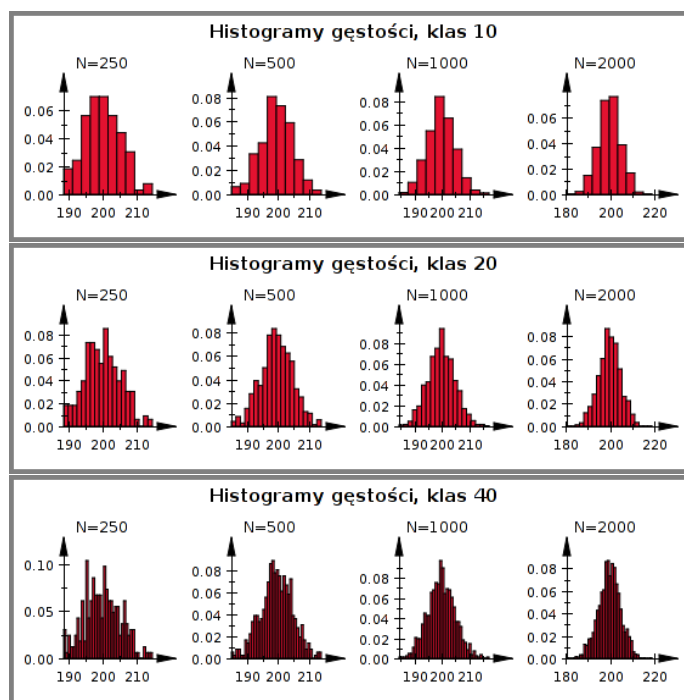
Budowanie histogramów i stereogramów ma sens jedynie dla dostatecznie dużych prób, szczególnie dużych w przypadku rozkładów wielowymiarowych. W celu skonstruowania histogramu dane pomiarowe grupowane są w przedziałach klasowych, a następnie w każdym z nich zliczane z osobna. W ten sposób uzyskujemy tzw. *częstości klasowe* (absolutne), czyli liczby rezultatów pomiarów o wartościach w danym przedziale klasowym. Zamiast absolutnych, możemy podać *częstości względne*, czyli stosunki częstości klasowych do liczby wszystkich pomiarów. Wreszcie za pomocą *gęstości częstości* (stosunek częstości względnej do szerokości przedziału klasowego) możemy oszacować gęstość prawdopodobieństwa wielkości pomiarowej, dla której wykonujemy histogram.

Tablica częstości klasowych określa *rozkład częstości*, który ilustrowany jest w postaci diagramów słupkowych, histogramów, a dla zmiennych losowych dwuwymiarowych w postaci diagramów dwuwymiarowych tzw. stereogramów.

Przykład jednowymiarowy przedstawiono w tabeli 5.1, w której zebrano rezultaty analizy 250 pomiarów pewnej wielkości fizycznej wynoszącej około 200. W oparciu o dane z tej tabeli oraz dane pochodzące z prób o innych rozmiarach, na rysunku 5.1 wykreślono histogramy częstości klasowych. Na rysunkach widzimy zmiany w wyglądzie histogramu. Przy ustalonej liczbie klas wartości podane na osiach pionowych silnie rosną wraz ze wzrostem rozmiaru próby. Dla prób o ustalonym rozmiarze wartości te silnie maleją ze wzrostem liczby przedziałów klasowych. Generalnie na tego typu histogramach wzrost liczby klas doprowadza do zanikania wysokości histogramów.

Tablica 5.1: Ilustracja rezultatów poczynañ, których celem była konstrukcja diagramu typu histogram. Wartości $n = 250$ wyników pomiarów mieszczą się w przedziale [188.52, 214.86]. Przedział ten podzielono na 10 klas (przedziałów klasowych) o jednakowej szerokości 2.634, w wierszu pierwszym podano odpowiadające każdej klasie wartości środkowe, w drugim wartości lewego i prawego końca przedziału klasowego. W celu zmniejszenia szerokości tabeli wartości te zmniejszono o 180.0. Dla każdej klasy wyznaczono liczbę pomiarów do niej wpadających (częstości klasowe, absolutne), a dzieląc te wartości przez wszystkie pomiarów obliczono częstości względne danej klasy. Wreszcie dzieląc częstości względne przez szerokość przedziału klasowego, obliczono odpowiadające im wartości gęstości, czyli częstości na jednostkę wielkości będącej przedmiotem pomiaru. Histogramy częstości klasowych oraz gęstości dla wartości podanych w tej tabeli pokazano w pierwszym wierszu po lewej na rysunkach 5.1 i 5.2.

Wartości środkowe	9.83	12.47	15.10	17.74	20.37	23.00	25.64	28.27	30.91	33.54	
Granice przedziałów klasowych	8.52-	11.15-	13.79-	16.42-	19.05-	21.69-	24.32-	26.96-	29.59-	32.22-	
	11.15	13.79	16.42	19.05	21.69	24.32	26.96	29.59	32.22	34.86	
Częstości klasowe	12	16	37	46	46	37	29	20	2	5	$\sum = 250$
Częstości względne	0.048	0.064	0.148	0.184	0.184	0.148	0.116	0.08	0.008	0.02	$\sum = 1.0$
Gęstość											
częstości względnej	0.0182	0.0243	0.0562	0.0699	0.0699	0.0562	0.0440	0.0304	0.0030	0.0076	



Rysunek 5.2: Histogramy gęstości częstości względnych rezultatów pomiaru pewnej wielkości fizycznej.

Możemy częściowo przewidzieć temu zjawisku wykreślając histogramy gęstości częstości względnych. Wygląd histogramów zmienia się jak poprzednio, ale na osiach pionowych mamy wartości bardzo do siebie zbliżone. Na takich histogramach powierzchnia każdego prostokąta jest równa prawdopodobieństwu wpadania pomiaru do danego przedziału klasowego. Ten typ histogramu przedstawiono na rysunku 5.2.

W tabeli 5.2 przedstawiono przypadek dwuwymiarowy ilustrujący grupowanie śladów po pociskach wystrzelonych w kierunku tarczy. Każdy punkt na tarczy reprezentowany jest przez parę współrzędnych kartezjańskich x, y . Rozmieszczenie punktów podlega prawom prawdopodobieństwa, tutaj dwuwymiarowemu rozkładowi zmiennych losowych \tilde{x}, \tilde{y} .

Pewna osoba oddała 400 strzałów, siatka pomiarowa miała kwadratowe oczka o bokach wynoszących 3 cm. Policzono liczbę punktów w każdym polu, wyniki zawarte są w górnej połowie kwadratów w tabeli 5.2.

Taka tablica reprezentuje dwuwymiarowy rozkład gęstości powtarzalności strzałów i może posłużyć do konstrukcji *stereogramu* czyli dwuwymiarowego histogramu, poprzez wzniesienie kolumn nad każdym kwadratem. Objętość kolumny jest proporcjonalna do względnej częstości liczby trafień w dany kwadrat. Np. częstość względna nad kwadratem (6, 6) wynosi $8/400 = 0.02$, 8 jest liczbą trafień w ten kwadrat, 400 jest ogólną liczbą strzałów. Częstość względna reprezentuje prawdopodobieństwo zajścia danego zdarzenia, dlatego możemy napisać $P[(6, 6)] = 0.02$. Określone w ten sposób prawdopodobieństwo trafienia w konkretny kwadrat tarczy, dotyczy obydwóch zmiennych losowych \tilde{x} i \tilde{y} łącznie.

Przypuśćmy jednak, że zainteresowani jesteśmy w celności trafienia jedynie ze względu na kierunek x , bez zwracania uwagi na kierunek y . Czyli pytamy jaki jest rozrzut strzałów

Tablica 5.2: Rezultaty badań celności strzałów do tarczy. Trafienia pogrupowano w kwadraty klasowe o rozmiarach 3×3 cm, liczba trafień w dany kwadrat podana jest w górnej części kwadratu. Opis szczegółowy podany jest w tekście.

	-18	-15	-12	-9	-6	-3	0	3	6	9	12	15	18	Rozkad brzegowy
18							$\frac{1}{20}$			$\frac{1}{20}$				2
15							$\frac{0}{3}$	$\frac{1}{20}$	$\frac{1}{20}$	$\frac{1}{20}$				5
12				$\frac{1}{20}$			$\frac{0}{4}$	$\frac{0}{3}$	$\frac{0}{4}$	$\frac{0}{2}$	$\frac{1}{20}$			12
9		$\frac{1}{20}$		$\frac{1}{20}$			$\frac{1}{3}$	$\frac{2}{20}$	$\frac{0}{4}$	$\frac{0}{2}$	$\frac{0}{3}$			21
6		$\frac{1}{20}$		$\frac{0}{2}$			$\frac{0}{4}$	$\frac{0}{3}$	$\frac{0}{4}$	$\frac{0}{5}$	$\frac{0}{3}$			42
3	$\frac{1}{20}$	$\frac{1}{20}$		$\frac{0}{3}$			$\frac{0}{11}$	$\frac{0}{9}$	$\frac{1}{15}$	$\frac{0}{7}$	$\frac{0}{4}$			67
0		$\frac{0}{2}$	$\frac{0}{4}$	$\frac{0}{3}$	$\frac{0}{25}$		$\frac{0}{14}$	$\frac{0}{15}$	$\frac{0}{20}$	$\frac{0}{15}$	$\frac{0}{4}$			67
		$\frac{0}{32}$	$\frac{0}{31}$	$\frac{0}{20}$	$\frac{0}{25}$		$\frac{0}{28}$	$\frac{0}{31}$	$\frac{0}{27}$	$\frac{0}{15}$	$\frac{0}{4}$			114
-3		$\frac{0}{17}$	$\frac{0}{23}$	$\frac{0}{20}$	$\frac{0}{20}$		$\frac{0}{10}$	$\frac{0}{11}$	$\frac{0}{5}$	$\frac{0}{6}$	$\frac{0}{4}$			54
		$\frac{0}{17}$	$\frac{0}{23}$	$\frac{0}{20}$	$\frac{0}{20}$		$\frac{0}{13}$	$\frac{0}{11}$	$\frac{0}{9}$	$\frac{0}{13}$	$\frac{0}{4}$			38
-6			$\frac{0}{15}$	$\frac{0}{20}$	$\frac{0}{11}$		$\frac{0}{9}$	$\frac{0}{10}$	$\frac{0}{6}$	$\frac{0}{4}$	$\frac{0}{3}$			23
				$\frac{0}{7}$	$\frac{0}{11}$		$\frac{0}{5}$	$\frac{0}{4}$	$\frac{0}{2}$	$\frac{0}{4}$	$\frac{0}{3}$			14
-9				$\frac{0}{7}$	$\frac{0}{4}$		$\frac{0}{7}$	$\frac{0}{6}$	$\frac{0}{4}$	$\frac{0}{8}$	$\frac{0}{10}$			23
				$\frac{0}{7}$	$\frac{0}{4}$		$\frac{0}{5}$	$\frac{0}{4}$	$\frac{0}{2}$	$\frac{0}{4}$	$\frac{0}{10}$			14
-12				$\frac{0}{7}$	$\frac{0}{4}$		$\frac{0}{5}$	$\frac{0}{3}$	$\frac{0}{2}$	$\frac{0}{4}$	$\frac{0}{10}$			6
-15							$\frac{0}{5}$	$\frac{0}{2}$	$\frac{0}{1}$	$\frac{0}{2}$	$\frac{0}{10}$			6
							$\frac{0}{1}$	$\frac{0}{3}$	$\frac{0}{2}$	$\frac{0}{1}$	$\frac{0}{10}$			2
							$\frac{0}{1}$	$\frac{0}{3}$	$\frac{0}{2}$	$\frac{0}{1}$	$\frac{0}{10}$			2
Częstosc brzegowa x	1	6	13	15	28	41	77	62	54	48	30	17	8	$\Sigma = 400$
Wzgl. częst. brzegowa x	0.002	0.015	0.033	0.038	0.070	0.103	0.192	0.155	0.135	0.120	0.075	0.042	0.029	$\Sigma = 1.000$
$\Sigma P[y x]$	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	

na lewo i na prawo od środka tarczy. Jest to przypadek jednowymiarowy ze zmienną losową \tilde{x} , a wartości w trzecim rzędzie od dołu w tabeli 5.2, odpowiadają częstościom trafień w odpowiednie 3 centymetrowe przedziały x . Wartości te uzyskano przez zsumowanie liczb z odpowiednich kolumn tabeli.

Podobny rozkład dla kierunku y otrzymamy sumując liczbę trafień w poszczególnych wierszach, podano go w ostatniej prawej kolumnie tabeli 5.2. Te dwa zbiory liczb stanowią *brzegowe częstości* trafień, a jeżeli podzielimy je przez całkowitą liczbę strzałów uzyskamy brzegowe częstości względne badanych zdarzeń. Np. prawdopodobieństwo, że x przyjmie wartość z przedziału $(7.5, 10.5)$ wynosi 0.120. Dla interwału $(-1.5, +1.5)$ wynosi 0.192.

Można określić jeszcze inny uniwersalny rozkład prawdopodobieństwa, np. ze względu na y przy ustalonym x . Przykładowo, dla $x = -6$, można odczytać podane w odpowiadającej mu kolumnie względne częstości dla y : 0.04, 0.11, 0.20, 0.25, ...¹. Analogicznie, biorąc dane z jednego wiersza, znajdziemy wartości częstości względnych dla x , przy ustalonych interwałach dla y . Liczby te odpowiadają warunkowym rozkładom częstości.

Zauważmy, że w poszczególnych wierszach mamy inne warunkowe częstości względne. Oznacza to, że jeśli zmienimy wartość jednej ze zmiennych, odpowiadający jej warunkowy rozkład dla drugiej zmiennej także się zmieni. Tego rodzaju “wpływ” jednej zmiennej losowej na drugą wskazuje na *korelację*. Takiej statystycznej korelacji czy zależności, nie należy mylić z zależnością algebraiczną. Przy zależności algebraicznej, wartość jednej ze zmiennych natychmiast określa drugą. Natomiast w naszym przykładzie jasne jest, że x i y nie są ze sobą związane funkcjonalnie.

Korelacja może być silna lub słaba. Oznacza to silne lub słabe zmiany rozkładów warunkowych przy zmianie warunku. Korelacja zerowa, oznacza że rozkład warunkowy nie zmienia się wraz ze zmianami warunku.

Z rozkładów częstości daje się wyprowadzić wnioski o odpowiadających im rozkładach prawdopodobieństw. Nie jest to wiarygodne gdy liczebność próby statystycznej jest niewielka, jak to często ma miejsce w zagadnieniach inżynierskich. W takich wypadkach typ rozkładu prawdopodobieństwa danej zmiennej losowej ustalony jest *a priori*, na mocy założenia. A z próby oszacowuje się jedynie pewne jego parametry. ↗

5.2.2 Statystyki z próby

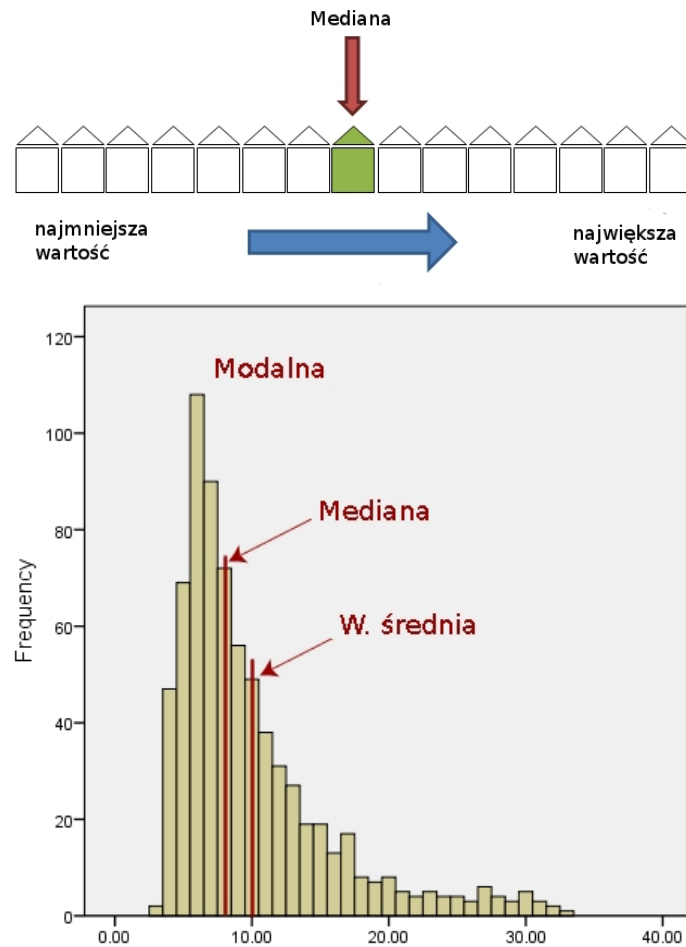
5.2.2.1 Wartości centralne, miary położenia

Średnia próby. Niech będzie dana próba (x_1, \dots, x_n) o rozmiarze n , zmiennej losowej \tilde{x} . Empiryczna wartość oczekiwana tej zmiennej, czyli średnia próby określona jest jako

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad (5.1)$$

Wartość oczekiwana zmiennej \tilde{x} , czyli $E[\tilde{x}]$, jest wartością średnią całej populacji, i jako taka jest wartością dokładną. Średnia empiryczna \bar{x} jest zmienną losową a jej wartością

¹Droga Czytelniczko, Czytelniku, myślę, że nie od rzeczy będzie jeśli zrobisz sobie teraz kawkę i w trakcie jej konsumpcji spróbujesz przez moment pomyśleć w jaki sposób policzono te wartości.



Rysunek 5.3: U góry, położenie mediany w uszeregowanej rosnąco próbie pomiarowej. U dołu, położenie modalnej, mediany i wartości średniej na histogramie.

oczekiwaną będzie

$$\begin{aligned}
 E[\bar{x}] &= E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} E[x_1 + \dots + x_n] = \\
 &= \frac{1}{n} (E[x_1] + \dots + E[x_n]) = \frac{1}{n} \cdot n\mu = \mu
 \end{aligned} \tag{5.2}$$

Średnia arytmetyczna \bar{x} , wyznaczona z próby, np. ze zbioru niezależnych obserwacji, jest jedynie estymatorem (przybliżeniem) wartości średniej μ zmiennej losowej \tilde{x} .

Mediana. Wyznaczenie mediany wymaga ustawienia wartości próby rosnąco albo malejąco. Mediana jest wartością środkowego pomiaru jeśli liczba pomiarów n jest nieparzysta, lub jest średnią arytmetyczną z dwóch wartości środkowych gdy n jest parzyste. Na lewo i prawo od mediany mamy więc zawsze identyczną liczbę obserwacji. Ze sposobu wyznaczenia mediany wynika, że jej wartość jest zupełnie nieczuła np. na wartości występujące po lewej bądź po prawej stronie szeregu pomiarowego. W żargonie statystyków, mówi się o niewrażliwości mediany na wartości występujące w ogonach rozkładu.

Modalna. Modalna próby jest wartością, która występuje w próbie najczęściej. W przypadku histogramu należy ona do przedziału klasowego, nad którym rozpina się prostokąt o największej wysokości. ↗

5.2.2.2 Miary rozproszenia, rozrzutu

Rozpiętość. Obliczana jest jako różnica pomiędzy największą i najmniejszą wartością z próby. Jest to najprostsza miara rozrzutu próby, ale nie jest ona tak intrygująca jak inne miary.

Odchylenie średnie (odchylenie przeciętne). Jest to jeszcze jedna gruba miara rozproszenia. Określana jest formułą

$$x_{od} = \frac{1}{n} \sum_{i=1}^n |x_i - \bar{x}| \quad (5.3)$$

gdzie \bar{x} jest średnią arytmetyczną. Zamiast średniej \bar{x} , można brać inne miary położenia. Moduł $|x_i - \bar{x}|$ jest odchyłką, x_{od} jest po prostu średnią arytmetyczną odchyłek.

Wariancja. Wariancja z próby określona jest wzorem

$$S_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (5.4)$$

gdzie \bar{x} — średnia z próby, n — rozmiar próby.

Powodem położenia w mianowniku różnicy $(n-1)$ jest żądanie, by S_x^2 miała wartość oczekiwaną równą wariancji populacji, czyli aby

$$E[S_x^2] = \sigma^2 \quad (5.5)$$

Równanie (5.5) oznacza, że zmienna losowa S_x^2 ma rozkład prawdopodobieństwa o wartości średniej σ^2 . Gdyby w równaniu (5.4) w mianowniku położył n , wówczas wartość oczekiwana tak zdefiniowanej wariancji wynosiłaby $(n-1) \cdot \sigma^2$. W tym wypadku wariancja z próby byłaby *obciążonym* estymatorem ponieważ jej wartość oczekiwana byłaby różna od wariancji populacji.

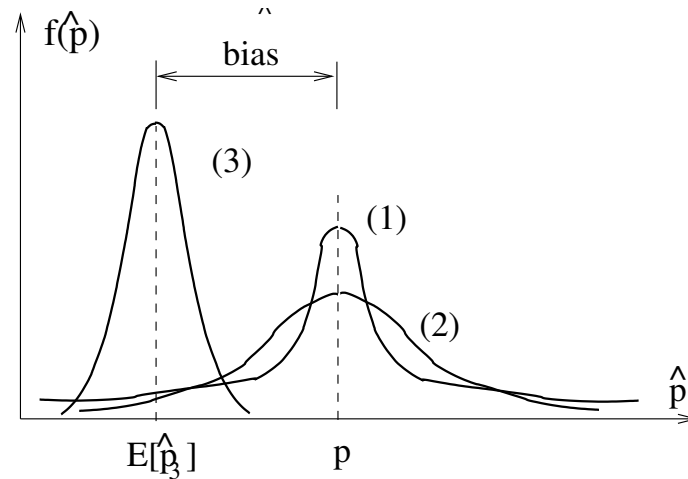
Kowariancja. Mając do dyspozycji n par $(x_1, y_1), (x_2, y_2), \dots$ wartości wektora losowego (\tilde{x}, \tilde{y}) oprócz S_x^2, S_y^2 można jeszcze obliczyć kowariancję z próby, mianowicie

$$S_{xy}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (5.6)$$

gdzie \bar{x}, \bar{y} są wartościami średnimi zmiennych losowych \tilde{x} i \tilde{y} . ↗

5.3 Estymacja punktowa

Oszacowanie parametrów rozkładu prawdopodobieństwa badanej populacji nazywamy *estymacją* parametrów, a do tego celu wykorzystujemy podane wyżej funkcje zwane statystyką, estymatorami. Ich wartości (tzw. estymaty) wyliczane są na podstawie próby pomiarowej reprezentującej daną populację.



Rysunek 5.4: Ilustracja trzech estymatorów parametru p . Estymatory (1), (2) są nieobciążone, obciążony estymator (3) ma najmniejszą wariancję.

Oszacowanie wartości jakiegoś parametru rozkładu populacji nazywane jest *estmacją punktową*.

Estymatorami punktowymi są średnia próby \bar{x} , wariancja S_x^2 , kowariancja S_{xy}^2 . Są one estymatorami parametrów $\mu, \sigma_x^2, \sigma_{xy}^2$ jakiegoś rozkładu prawdopodobieństwa.

Estymator \hat{p} daje dobre przybliżenie szacowanego parametru p populacji, jeśli spełnia określone warunki, kryteria: nieobciążoności, zgodności, minimalnej wariancji, efektywności.

Przypuśćmy, że na podstawie danych z próby o liczebności n obliczono estymatę \hat{p}_1 estymatora \hat{p} . Jeśli eksperyment zostanie powtórzony wielokrotnie, możliwym będzie obliczenie kolejnych estymat $\hat{p}_2, \hat{p}_3, \dots, \hat{p}_k$. Otrzymane estymaty na ogół będą się różniły między sobą. Estymator \hat{p} jest przecież zmienną losową a estymaty $\hat{p}_1, \hat{p}_2, \dots, \hat{p}_k$ są możliwymi wartościami tej zmiennej.

Estymator \hat{p} może być skonstruowany w różny sposób, np. niech będzie, że \hat{p} daje przybliżoną wartość parametru p zawsze z nadmiarem. Wówczas każda z wyznaczonych estymat $\hat{p}_i, (i = 1, \dots, k)$ będzie większa od rzeczywistej wartości parametru p . Również wartość oczekiwana zmiennej \hat{p} będzie większa od p , czyli $E[\hat{p}] > p$. Zatem posłużenie się estymatorem, którego wartość oczekiwana nie jest równa wartości szacowanego parametru prowadzi do błędów systematycznych.

Kryterium nieobciążoności. Na dobry estymator nakłada się warunek by był estymatorem nieobciążonym. Jest to równoważne żądaniu by dla dowolnego rozmiaru próby

$$E(\hat{p}) = p$$

Jeżeli jest inaczej mamy do czynienia z błędem systematycznym B_s (bias) postaci

$$B_s = E(\hat{p}) - p \quad (5.7)$$

Rysunek 5.4 przedstawia trzy estymatory. Dwa pierwsze (1) i (2) są nieobciążone, trzeci (3) wykazuje odchylenie systematyczne (bias).

Kryterium zgodności. Estymator jest estymatorem zgodnym jeśli przy rozmiarze próby $n \rightarrow \infty$ prawdopodobieństwo, że wartość estymatora \hat{p} coraz mniej różni się od wartości parametru p , dąży do jedności. Czyli dla dowolnie małego $\varepsilon > 0$ mamy

$$\lim_{n \rightarrow \infty} P(|\hat{p} - p| < \varepsilon) = 1 \quad (5.8)$$

Kryterium minimalnej wariancji. Nie można jednak sądzić, że estymator nieobciążony daje zawsze dobre przybliżenie szacowanego parametru. Może się zdarzyć, że rozproszenie możliwych wartości estymatora \hat{p} wokół wartości oczekiwanej jest duże, tzn. że ma on dużą wariancję.

Na rysunku 5.4 mamy przypadek, w którym wariancja estymatora (1) jest mniejsza aniżeli estymatora (2). Ale wariancja dla estymatora (3) jest najmniejsza. Stąd, biorąc pod uwagę jedynie kryterium minimalnej wariancji, estymator (3) jest najlepszym estymatorem parametru p populacji. Istotnie, gdyby wszystkie te estymatory były nieobciążone, byłby to wybór słuszny. Bowiem tylko w wypadku tego estymatora mamy najmniejsze ryzyko, że wyznaczona przez nas jego estymata wyraźnie różni się od wartości oczekiwanej tego estymatora.

Jeśli jednak (3) jest estymatorem obciążonym, a to właśnie rysunek 5.4 sugeruje, to trzeba starannie rozważyć korzyści płynące z nieobciążenia w stosunku do korzyści płynącej z minimalnej wariancji.

Kryterium efektywności. Estymator efektywny to estymator nieobciążony o minimalnej wariancji. Można wykazać, że jeśli wariancja estymatora nieobciążonego przy $n \rightarrow \infty$ dąży do zera, to taki estymator jest również estymatorem zgodnym.

Dygresja : dokładność - precyzja.

Miarą precyzji może być rozrzut (dyspersja) zbioru obserwacji. Precyzja jest miarą wewnętrznej zgodności zbioru obserwacji.

Na rysunku 5.4, estymator (2) ma mniejszą precyzję niż estymator (3). Estymatory (1) i (2) są jednakowo dokładne, ale nie tak precyzyjne jak estymator (3). Jednak estymator (3) jest mniej dokładny. Różnica pomiędzy precyzją i dokładnością wynika z obecności błędu systematycznego. Precyzja dotyczy jedynie efektów przypadkowych, dokładność natomiast zarówno przypadkowych jak i systematycznych.

Gauss jako miarę dokładności zaproponował tzw. *średni błąd kwadratowy*

$$m^2 = E[(\hat{p} - E[\hat{p}])^2] \quad (5.9)$$

co jak można pokazać jest równoważne równaniu

$$m^2 = \sigma_{\hat{p}}^2 + (bias)^2 \quad (5.10)$$

W celu ilustracji podanych wyżej koncepcji dobrego estymatora, rozważmy zmienną losową \tilde{x} , dla której wartość oczekiwana wynosi $E(\tilde{x}) = \mu$.

Oczywiście każda obserwacja x_i , ze zbioru n niezależnych pomiarów będzie nieobciążonym estymatorem średniej μ . Jak wiemy średnia arytmetyczna \bar{x} również jest nieobciążonym estymatorem μ . A zatem z punktu widzenia nieobciążoności są to jednakowo akceptowalne estymatory. Sytuacja zmienia się jeżeli przejdziemy do porównania wariancji.

Wariancja zmiennej \tilde{x} jest równa σ^2 wariancji populacji. Wariancję $\sigma_{\bar{x}}^2$ średniej arytmetycznej obliczymy ze wzoru

$$\begin{aligned}\sigma_{\bar{x}}^2 &= E[(\bar{x} - E(\bar{x}))^2] = E[(\bar{x} - \mu)^2] = \\ &= E\left[\left(\frac{x_1 + x_2 + \dots + x_n}{n} + \mu\right)^2\right] = \\ &= \frac{1}{n^2}E[(x_1 - \mu) + (x_2 - \mu) + \dots + (x_n - \mu)]^2 = \\ &= \frac{1}{n^2}E[(x_1 - \mu)^2 + (x_1 - \mu)(x_2 - \mu) + (x_1 - \mu)(x_3 - \mu) + \dots + (x_n - \mu)^2]\end{aligned}$$

a skoro elementy próby x_i są losowo niezależne, mamy, że wszystkie kowariancje, czyli wyrazy $E[(x_i - \mu)(x_j - \mu)]$, $i \neq j$ znikają. Zatem będziemy mieli

$$\sigma_{\bar{x}}^2 = \sigma^2/n \quad (5.11)$$

Ponieważ $\sigma_{\bar{x}}^2 < \sigma^2$, średnia \bar{x} czyni zadość kryterium najmniejszej wariancji. Estymator \bar{x} jest także estymatorem zgodnym. Estymator określany jest mianem *wystarczającego* (dostatecznego) jeśli skupia w sobie wszystkie informacje o estymowanym parametrze. \uparrow

5.4 Metody estymacji punktowej

Estymatory posiadające wszystkie wymienione wyżej własności określane są mianem "najlepszych" estymatorów. Ale pewne z tych własności bywają ważne tylko asymptotycznie.

W praktycznych zastosowaniach warto stosować metody wyznaczania estymatorów, dla których spełnione są wszystkie kryteria dobrego estymatora lub przynajmniej niektóre z nich. Są to następujące metody: metoda momentów, metoda największej wiarygodności, metoda najmniejszych kwadratów. \uparrow

5.4.1 Metoda momentów

Jako estymantę dla k -tego momentu rozkładu prawdopodobieństwa bierzemy po prostu k -ty moment z próby jednakowo dokładnych pomiarów

$$m_k = \frac{1}{n} \sum_{i=1}^n x_i^k \quad (5.12)$$

Przykład — wartość średnia z próby. Zauważmy, że czynnik $1/n$ w tym wzorze reprezentuje jednakowe prawdopodobieństwo każdej wartości pomiaru x_i \uparrow

5.4.2 Metoda największej wiarygodności

Jest to metoda szeroko wykorzystywana w statystyce. Stosowana jest dla oszacowania parametrów w taki sposób, by wyznaczone na podstawie próby ich wartości były najbardziej prawdopodobne.

Jeśli zmienne losowe x_i są niezależne i mają taki sam rozkład $f(x_i)$, to uzyskany z próby wektor $(x_1 \dots x_n)$ ma rozkład łączny postaci

$$L(x_1 \dots x_n; p_1 \dots p_m) = \prod_{i=1}^n f(x_i; p_1 \dots p_m) \quad (5.13)$$

gdzie $p_1 \dots p_m$ są parametrami funkcji rozkładu, które właśnie należy oszacować.

Równanie (5.13) nazywane jest *funkcją wiarygodności*. Gęstość $f(x_i)$ jest funkcją nieznanymi parametrów p_1, \dots, p_m , te z kolei są wyrażone przez wartości z próby $p_i = g_i(x_1 \dots x_n)$.

Wydaje się być wiarygodnym wyznaczenie estymatorów \hat{p}_i w taki sposób by łączna funkcja gęstości prawdopodobieństwa L miała wartość maksymalną. Warunek taki doprowadza do układu równań

$$\frac{\partial L}{\partial p} = 0 \quad \text{lub} \quad \frac{\partial \ln L}{\partial p} = 0 \quad (5.14)$$

Rozwiązanie (5.14) daje estymator parametru p zwany estymatorem *największej wiarygodności*. Ta metoda estymacji wymaga jednak *a priori* postaci funkcji rozkładu badanej zmiennej losowej.

Dla przykładu oszacujmy średnią rozkładu normalnego wykorzystując dane z próby $(x_1 \dots x_n)$. Ponieważ

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

funkcja wiarygodności ma postać

$$L = \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x_1 - \mu}{\sigma} \right)^2 \right] \cdot \dots \cdot \frac{1}{\sigma\sqrt{2\pi}} \exp \left[-\frac{1}{2} \left(\frac{x_n - \mu}{\sigma} \right)^2 \right]$$

$$L = \frac{1}{(2\pi\sigma^2)^{n/2}} \exp \left[-\frac{1}{2} \sum_{i=1}^n \frac{(x_i - \mu)^2}{\sigma^2} \right]$$

By ułatwić policzenie pochodnych, w warunku (5.14) opłaca się najpierw funkcję L zlogarytmować

$$\ln L = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu)^2$$

a po jej zróżniczkowaniu, pochodna ze względu na μ wynosi

$$\frac{\partial(\ln L)}{\partial \mu} = \frac{-2}{2\sigma^2} \sum_{i=1}^n (x_i - \mu) \cdot (-1)$$

a po przyrównaniu do zera, dla maximum będzie

$$\sum_{i=1}^n (x_i - \mu) = \sum_{i=1}^n x_i - n\mu = 0$$

$$\mu = \frac{1}{n} \sum_{i=1}^n x_i$$

Jak widać, estymator największej wiarygodności dla średniej z rozkładu normalnego jest równy średniej arytmetycznej z próby.

Poszukajmy za pomocą metody największej wiarygodności estymatora dla wariancji σ^2 . Różniczkując $\ln L$ po σ^2 otrzymamy

$$\frac{\partial(\ln L)}{\partial(\sigma^2)} = \frac{-n\pi}{2\pi\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^n (x_i - \mu)^2 = 0$$

Warunek maximum oznacza, że

$$\frac{1}{\sigma^2} \sum_{i=1}^n (x_i - \mu)^2 = n \Rightarrow \sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

co nie jest tym samym co definiowana równaniem (5.4) wariancja z próby S_x^2 . Widzimy że estymator σ^2 dany wzorem powyżej jest estymatorem obciążonym! ↗

5.4.3 Metoda najmniejszych kwadratów

Inną metodą znajdowania estymatorów jest metoda najmniejszych kwadratów (w skrócie MNK). W przeciwieństwie do metody największej wiarygodności, by zastosować MNK nie musimy znać rozkładu badanej zmiennej losowej. Dopiero w celu zbudowania przedziałów ufności, czy podczas testowania hipotez, informacja o rozkładzie jest wymagana. Jeżeli zmienne losowe (których obserwacje dotyczą) mają rozkład normalny, metoda najmniejszych kwadratów daje identyczny rezultat jak metoda największej wiarygodności.

Niech \mathbf{v} będzie wektorem residuów, tzn. $\mathbf{v} = \hat{\mathbf{l}} - \mathbf{l}$, gdzie $\hat{\mathbf{l}}$ jest estymatorem *a posteriori* wielkości obserwowanych. Niech \mathbf{v} ma rozkład normalny, oraz $E[\mathbf{v}] = \mathbf{0}$, zaś macierz Σ niech będzie macierzą kowariancji obliczoną za pomocą wartości wziętych z próby. Wówczas, skoro łączna funkcja gęstości rozkładu normalnego ma postać

$$\begin{aligned} f(\mathbf{v}) &= C \exp\left(-\frac{1}{2}(\mathbf{v} - E[\mathbf{v}])^T \Sigma^{-1}(\mathbf{v} - E[\mathbf{v}])\right) = \\ &= C \exp\left(-\frac{1}{2}\mathbf{v}^T \Sigma^{-1} \mathbf{v}\right) \end{aligned}$$

kryterium najmniejszych kwadratów pociąga by $\phi = \mathbf{v}^T \Sigma^{-1} \mathbf{v}$ osiągało minimum, co odpowiada maksymalnej wartości $f(\mathbf{v})$. A to jest równoważne warunkowi postawionemu w metodzie największej wiarygodności. ↗

5.5 Estymacja przedziałowa

Oszacowanie wartości parametru populacji za pomocą jednej liczby nazywamy estymacją punktową. Wszystkie omawiane wyżej oszacowania były estymacjami punktowymi. Tego typu oszacowanie może okazać się ryzykowne szczególnie gdy mamy do czynienia z próbami o niewielkich rozmiarach, bowiem nie zdając sobie z tego sprawy możemy wówczas

uzyskać oceny znacznie różniące się od szacowanego parametru. Dlatego gdy rozmiar próby jest nieduży, będzie bezpieczniej gdy zastosujemy estymację przedziałową.

Estymacją przedziałową nazywane jest oszacowanie parametru dokonane za pomocą dwóch liczb odpowiadających końcom tzw. przedziału ufności. Taka estymacja poza wartością parametru, umożliwia ustalenie wiarygodności oszacowania, odpowiadającej wybranemu poziomowi ufności (stopniowi pewności).

Przypuśćmy, że na podstawie próby obliczono wartość statystyki \hat{p} , która szacuje nieznaną wartość parametru p badanej populacji. O p wiemy, że jest wielkością stałą². Estymata \hat{p} tym lepiej określa parametr p , im mniejsza jest wartość bezwzględna różnicy $|p - \hat{p}|$ między tymi wielkościami.

Spostrzeżenie to można wyrazić nieco inaczej. Jeżeli $\varepsilon > 0$ oraz $|p - \hat{p}| < \varepsilon$, to im mniejsze jest ε tym oszacowanie parametru p będzie dokładniejsze. W ten sposób dodatnia liczba ε charakteryzuje jakość estymacji.

Niestety, metody statystyczne nie pozwalają na wyznaczenie przedziału, w którym z całą pewnością leży poszukiwany parametr p , czyli nigdy nie mamy pewności, że nierówność $|p - \hat{p}| < \varepsilon$ jest spełniona. W ramach statystyki można jedynie mówić o prawdopodobieństwie $\gamma = 1 - \alpha$ (tzw. *poziom ufności*) z jakim nierówność ta jest spełniona.

W przypadku gdy mamy do czynienia z rozkładami niesymetrycznymi nierówność ma postać $-\varepsilon_1 < p - \hat{p} < \varepsilon_2$, gdzie $\varepsilon_1, \varepsilon_2 > 0$.

W praktyce, poziom ufności jest zadany z góry, przyjmowane są tu wartości bliskie jedności, np.: 0.95, 0.99, 0.999, ale jednocześnie zależy nam na możliwie wąskim przedziale z dużą wiarygodnością pokrywającym nieznaną wartość parametru p .

W konstrukcji przedziału ufności korzystamy z wyrażenia pozwalającego na obliczenie prawdopodobieństwa z jakim zmienna losowa \tilde{x} przyjmuje wartości z przedziału (x_1, x_2) , tzn.

$$P(x_1 < \tilde{x} < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} f(x) \cdot dx$$

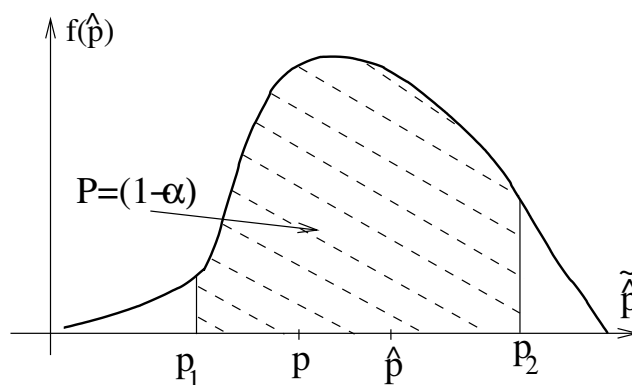
Analogicznie piszemy dla ustalonej wartości poziomu ufności $(1 - \alpha)$

$$P[\hat{p} - \varepsilon_1 < p < \hat{p} + \varepsilon_2] = 1 - \alpha, \quad (5.15)$$

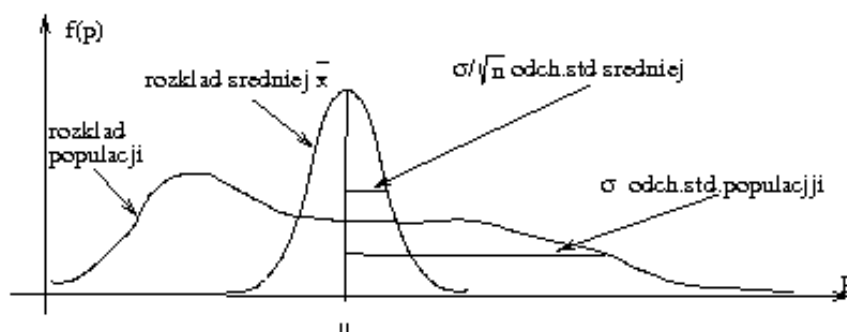
co interpretujemy następująco: przedział $(\hat{p} - \varepsilon_1, \hat{p} + \varepsilon_2)$ z prawdopodobieństwem $(1 - \alpha)$ pokrywa nieznaną wartość parametru p . Prawdopodobieństwo, że p nie wpada do tego przedziału wynosi α .

Mamy zatem, że *przedział ufności* to przedział $(\hat{p} - \varepsilon_1, \hat{p} + \varepsilon_2)$, który z zadaniem prawdopodobieństwem $(1 - \alpha)$ (poziom ufności), pokrywa nieznaną wartość parametru p .

²Możliwe jest uogólnienie w którym p jest zmienną losową.



Rysunek 5.5: Estymacja przedziałowa nieznanego parametru p populacji. Przedział p_1, p_2 z prawdopodobieństwem $(1 - \alpha)$ pokrywa nieznaną wartość p .



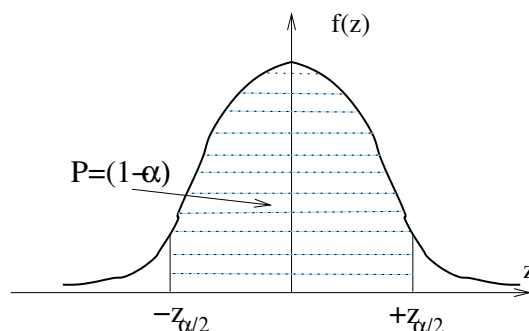
Rysunek 5.6: Ilustracja wniosku z centralnego twierdzenia granicznego: populacja o średniej μ i odchyleniu standardowym σ ma dowolny rozkład prawdopodobieństwa; średnia arytmetyczna \bar{x} próby wziętej z tej populacji ma w przybliżeniu rozkład normalny o średniej μ i odchyleniu standardowym σ/\sqrt{n} .

Dygresja. Końce przedziału $(\hat{p} - \varepsilon_1, \hat{p} + \varepsilon_2)$ są zmienne i mają charakter losowy, bowiem dla różnych prób branych z tej samej populacji uzyskamy różne wartości estymatora \hat{p} . Granice te są funkcjami wielkości pomiarowych i przyjętego stałego poziomu ufności. Dla ustalonej wartości prawdopodobieństwa $(1 - \alpha)$, szerokość przedziału ufności zmniejsza się ze wzrostem liczby pomiarów (ze wzrostem liczby stopni swobody). Ale jeśli z tej samej populacji pobieramy próby o jednakowej liczebności, wówczas dla wyższego poziomu ufności dostaniemy szerszy przedział ufności.

5.5.1 Przedział ufności dla średniej

Z *centralnego twierdzenia granicznego*³ wynika, że dla populacji o dowolnym rozkładzie $\mathcal{G}(\mu, \sigma)$, wartość średnia \bar{x} obliczona z pobranej z niej próby, w przybliżeniu ma rozkład $\mathcal{N}(\mu, \sigma/\sqrt{n})$ (rysunek 5.6). Przybliżenie jest tym lepsze im większy jest rozmiar n próby. Jeśli rozkład populacji jest normalny, to rozkład średniej też jest normalny, niezależnie od liczebności próby.

³Nazwę “centralne” nadano temu twierdzeniu z uwagi na rolę jaką ono odgrywa w statystyce i rachunku prawdopodobieństwa. Oczywiście istnieją zwolennicy nazywania tego twierdzenia “wielkim twierdzeniem granicznym”.



Rysunek 5.7: Estymacja przedziałowa wartości średniej. Zmienna standaryzowana z służy do ustalenia przedziału $-z_{\alpha/2}, z_{\alpha/2}$ pokrywającego z prawdopodobieństwem $1-\alpha$ nieznaną wartość z , dla której będziemy mieli, że $\mu = \bar{x} - z\sigma/\sqrt{n}$.

Rozważmy zmienną losową \tilde{x} mającą rozkład normalny, wyznaczoną jako średnia arytmetyczna z próby pomiarowej wziętej z populacji o nieznannej wartości oczekiwanej μ ale o znanej wariancji σ^2 . Za punktowy estymator wartości μ możemy wziąć wartość średnią \bar{x} . Obok oceny punktowej, chcemy wyznaczyć przedział ufności pokrywający przy poziomie ufności $(1-\alpha)$ nieznaną wartość parametru μ .

Teoria poucza nas, że zmienna losowa \tilde{z} , określona wyrażeniem $z = (\bar{x} - \mu)/(\sigma/\sqrt{n})$ ma rozkład normalny $\mathcal{N}(0, 1)$. Zatem jeszcze przed pobraniem próby możemy skonstruować przedział,

$$\left(\mu - z \frac{\sigma}{\sqrt{n}} < \bar{x} < \mu + z \frac{\sigma}{\sqrt{n}} \right)$$

o którym dla $z = 1.96$ moglibyśmy z prawdopodobieństwem 0.95, twierdzić, że wyznaczona za chwilę wartość \bar{x} będzie w nim się znajdowała.

Ponieważ odległość od \bar{x} do μ jest taka sama jak od μ do \bar{x} , stąd, \bar{x} wpada do przedziału $\mu \pm z\sigma/\sqrt{n}$ wtedy gdy μ znajduje się we wnętrzu przedziału $\bar{x} \pm z\sigma/\sqrt{n}$.

Gdybyśmy w populacji o znanej μ wielokrotnie realizowali nasz eksperyment przekonalibyśmy się, że dla $z = 1.96$, to co powiedziano wyżej będzie miało miejsce 95 razy na 100 przypadków, czyli w 95% przypadków.

W myśl tego co powiedziano wyżej przedział ufności dla średniej μ populacji można skonstruować następująco. Na podstawie standardowego rozkładu normalnego dobieramy taką wartość $z_{\alpha/2}$ by prawdopodobieństwo, że zmienna \tilde{z} przyjmie wartość na zewnątrz przedziału $(-z_{\alpha/2}, z_{\alpha/2})$ wynosiło α (patrz rysunek 5.7). Stwierdzenie probabilistyczne dotyczące symetrycznego przedziału ufności ma zatem postać

$$P \left[-z_{\alpha/2} < \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} < z_{\alpha/2} \right] = 1 - \alpha$$

lub

$$P \left[\bar{x} - z_{\alpha/2} \cdot \sigma/\sqrt{n} < \mu < \bar{x} + z_{\alpha/2} \cdot \sigma/\sqrt{n} \right] = 1 - \alpha \quad (5.16)$$

A stąd mamy, że przedział ufności dla μ ma postać

$$\bar{x} - z_{\alpha/2} \cdot \sigma/\sqrt{n} < \mu < \bar{x} + z_{\alpha/2} \cdot \sigma/\sqrt{n}$$

Przedział ten z prawdopodobieństwem $(1 - \alpha)$ pokrywa szacowaną przez nas średnią μ .

Dla $\alpha = 0.05$, będzie $z_{\alpha/2} = 1.96$ zatem równanie (5.16) przyjmie postać

$$P \left[\bar{x} - 1.96 \cdot \sigma / \sqrt{n} < \mu < \bar{x} + 1.96 \cdot \sigma / \sqrt{n} \right] = 0.95$$

a przedziałem ufności będzie

$$\bar{x} - 1.96 \cdot \sigma / \sqrt{n} < \mu < \bar{x} + 1.96 \cdot \sigma / \sqrt{n}$$

Był to przykład dwustronnego przedziału ufności. Dla jednostronnego przedziału ufności mamy

$$P \left[-z_{\alpha} < \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \right] = 1 - \alpha$$

$$P \left[\mu < \bar{x} + z_{\alpha} \cdot \sigma / \sqrt{n} \right] = 1 - \alpha \quad (5.17)$$

W przypadku gdy odchylenie standardowe σ populacji jest nieznanne, z konieczności trzeba je zastąpić odchyleniem standardowym s wyznaczonym z próby (równanie (5.4)). Estymator $(\bar{x} - \mu) / (s / \sqrt{n})$ ma teraz rozkład t o $(n - 1)$ stopniach swobody. Stwierdzenie probabilistyczne dotyczące przedziału ufności dla tego estymatora ma postać

$$P \left[-t_{\alpha/2, n-1} < \frac{\bar{x} - \mu}{s / \sqrt{n}} < t_{\alpha/2, n-1} \right] = 1 - \alpha$$

lub

$$P \left[\bar{x} - t_{\alpha/2, n-1} \cdot s / \sqrt{n} < \mu < \bar{x} + t_{\alpha/2, n-1} \cdot s / \sqrt{n} \right] = 1 - \alpha \quad (5.18)$$

Przyjmując $\alpha = 0.05$, z tablic rozkładu t dla np. $n - 1 = 7$ dostaniemy $t_{\alpha/2, 7} = 2.365$. Przedział ufności jest więc szerszy niżeli w przypadku gdy znamy wartość σ . Ze wzrostem liczby stopni swobody — czyli ze wzrostem wielkości próby różnica pomiędzy wartościami $t_{\alpha/2}$ i $z_{\alpha/2}$ maleje, o ile pozostałe wielkości nie zmieniają się. W praktyce gdy próba przekracza rozmiar 30, rozkład t często jest aproksymowany standaryzowanym rozkładem normalnym. ↗

5.5.2 Przedział ufności dla wariancji

Niech dana będzie wariancja s^2 otrzymana z próby zaczerpniętej z populacji o rozkładzie $N(\mu, \sigma)$. Zmienna losowa ms^2 / σ^2 (m jest liczbą stopni swobody z jaką policzono s^2), ma rozkład χ^2 dla m stopni swobody. Przy poziomie ufności $(1 - \alpha)$ zmienna ta z prawdopodobieństwem $(1 - \alpha)$ przyjmie wartości pomiędzy $\chi_{\alpha/2, m}^2$ i $\chi_{1-\alpha/2, m}^2$ czyli

$$P \left[\chi_{\alpha/2, m}^2 < \frac{ms^2}{\sigma^2} < \chi_{1-\alpha/2, m}^2 \right] = 1 - \alpha$$

lub

$$P \left[\frac{ms^2}{\chi_{1-\alpha/2, m}^2} < \sigma^2 < \frac{ms^2}{\chi_{\alpha/2, m}^2} \right] = 1 - \alpha \quad (5.19)$$

↗

5.5.3 Przedziały ufności dla stosunku wariancji

Niech dane są dwie niezależne próby losowe o rozmiarach n_1 i n_2 , pochodzące z normalnych populacji o wariancjach σ_1^2 i σ_2^2 . Wówczas każda ze zmiennych losowych $m_1 s_1^2 / \sigma_1^2$, $m_2 s_2^2 / \sigma_2^2$ gdzie $m_1 = n_1 - 1$, $m_2 = n_2 - 1$, ma rozkład χ^2 dla m_1 i m_2 stopni swobody, odpowiednio. Stosunek tych zmiennych

$$F = \frac{s_1^2 / \sigma_1^2}{s_2^2 / \sigma_2^2} = \frac{s_1^2 s_2^{-2}}{\sigma_1^2 \sigma_2^{-2}} \quad (5.20)$$

ma rozkład F o m_1 i m_2 stopniach swobody. Dla poziomu ufności $(1 - \alpha)$, zmienna losowa dana równaniem (5.20), z prawdopodobieństwem $(1 - \alpha)$ przyjmie wartości z przedziału $[F_{\alpha/2, m_1, m_2}, F_{1-\alpha/2, m_1, m_2}]$, a więc

$$P \left[F_{\alpha/2, m_1, m_2} < \frac{s_1^2 s_2^{-2}}{\sigma_1^2 \sigma_2^{-2}} < F_{1-\alpha/2, m_1, m_2} \right] = 1 - \alpha$$

lub

$$P \left[\frac{s_1^2 s_2^{-2}}{F_{1-\alpha/2, m_1, m_2}} < \sigma_1^2 \sigma_2^{-2} < \frac{s_1^2 s_2^{-2}}{F_{\alpha/2, m_1, m_2}} \right] = 1 - \alpha \quad (5.21)$$

↗

5.6 Zadanka na ćwiczenia

1. Dane są rezultaty 20 pomiarów odległości kątowej pomiędzy dwiema gwiazdami na sferze niebieskiej (patrz tabela 5.3). Narysuj histogram tych pomiarów, przyjmij szerokość przedziału klasowego 0.5", lewy brzeg histogramu ma odpowiadać wartości kąta 4". Oblicz: wartość średnią pomiarów (średnią z próby), odchylenie standardowe pojedynczego pomiaru, odchylenie standardowe średniej z próby. Ponadto wyznacz wartości: mediany, modalnej, rozpiętości oraz odchylenia przeciętnego.

Tablica 5.3: Rezultaty pomiarów odległości kątowej pomiędzy dwiema gwiazdami. Wartości kątów podano w sekundach łuku.

9.3	6.6	7.9	6.1	6.9	6.1	7.8	7.2	6.8	8.0
7.1	4.9	5.7	5.2	6.9	6.2	10.1	9.7	4.1	6.2

2. Dysponując rezultatami z poprzedniego zadania zbuduj przedziały ufności dla wartości średniej oraz wariancji rozkładu prawdopodobieństwa pomiarów. Przyjmij poziom ufności 95%.

↗

Rozdział 6

Testowanie hipotez statystycznych

Streszczenie

Rozstrzygnięcie między przynajmniej dwiema konkurencyjnymi hipotezami dotyczącymi zjawisk dokonywane jest w oparciu o adekwatne im hipotezy statystyczne. W problemie dwudecyzyjnym weryfikacja hipotezy zerowej H_0 polega obliczeniu wartości obserwowanej K_p odpowiedniej zmiennej losowej K , tzw. testu (sprawdzianu) oraz na sprawdzeniu do którego obszaru wartość ta należy. Zbiór wartości jakie test \tilde{K} może przyjmować składa się z dwóch rozłącznych podzbiorów: obszaru krytycznego i obszaru akceptacji, wyznaczanych przy ustalonym z góry poziomie istotności α . Gdy wartość K_p należy do obszaru krytycznego hipoteza H_0 jest odrzucana i jednocześnie akceptowana jest hipoteza alternatywna H_1 . Opisana metoda postępowania nie jest idealna bowiem wiąże się z ryzykiem popełnienia błędu I-go rodzaju (odrzuć hipotezy prawdziwej), albo błędu II-go rodzaju (przyjęć hipotezy fałszywej). Prawdopodobieństwo popełnienia błędu I-go rodzaju wynosi α , natomiast błędu II-go rodzaju wynosi β . Różnica $1 - \beta$ nazywana jest mocą testu, jest to prawdopodobieństwo trafienia testu \tilde{K}_p do obszaru krytycznego, gdy prawdziwa jest hipoteza H_1 . Dobra metoda weryfikacji hipotez powinna charakteryzować się małymi prawdopodobieństwami popełnienia błędu I-go i II-go rodzaju. Jest to równoważne żądaniu możliwie dużej mocy testu i małej wartości poziomu istotności α . Są to jednak wymagania nie do spełnienia jednocześnie, dlatego przy zadanym poziomie istotności obszar krytyczny testu wyznacza się w taki sposób, by moc testu była maksymalna. W niniejszych materiałach po wprowadzeniu w zagadnienie weryfikacji hipotez statystycznych przedstawiono kilka sposobów realizacji tzw. testów parametrycznych dla wartości średniej i dla wariancji rozkładu.

Słowa kluczowe: Testowanie hipotez statystycznych, poziom istotności, moc testu, błędy I-go i II-go rodzaju, testy parametryczne. ^a

^aModyfikowano AD 2010, maj, 11

6.1 Wstęp

Często przychodzi nam stawić czoła problemom, w których ze zbioru możliwych decyzji musimy dokonać wyboru jednej z nich z nich. Jasne jest, że w takich sytuacjach przydaje się metoda ułatwiająca podjęcie możliwie najlepszej decyzji.

Najprostszy przypadek podejmowania decyzji ma miejsce kiedy do wyboru mamy tylko dwie decyzje albo X albo Y — jest to tzw. problem dwu-decyzyjny. Przykładowo możemy mieć potrzebę rozstrzygnięcia:

- czy dana gwiazda jest, bądź nie jest członkiem określonej gromady gwiazd,
- czy prędkość radialna gwiazdy jest stała czy zmienna,
- czy hipoteza o elipsoidalnym rozkładzie prędkości gwiazd z otoczenia Słońca jest prawdziwa albo nie.

Tego typu problemy wymagające podejmowania decyzji określane są mianem *testowania hipotez*. Jedno z konkurencyjnych zdań nazywane jest *hipotezą* testowaną (*zerową*), drugie *hipotezą alternatywną*. Natomiast dwie alternatywne decyzje mają postać: odrzuć hipotezę testowaną albo zachowaj hipotezę testowaną. Np. przyjmijmy za hipotezę testowaną zdanie: Kastor jest gwiazdą pojedynczą, hipotezą alternatywną będzie: Kastor nie jest gwiazdą pojedynczą. Jeśli w rezultacie doświadczenia przekonamy się do pierwszego z tych zdań, wówczas naszą decyzją będzie: zachowaj hipotezę testowaną. Jeśli przeciwnie, zdecydujemy się traktować Kastora za gwiazdę wielokrotną, doprowadzi to do decyzji: odrzuć hipotezę testowaną. ↗

6.2 Hipotezy dotyczące zjawisk a hipotezy statystyczne

Ogólnie hipotezy o jakich tu mowa można podzielić na dwie grupy:

- hipotezy bezpośrednio dotyczące badanych zjawisk,
- hipotezy statystyczne.

Hipoteza dotycząca zjawiska jest zdaniem bezpośrednio związanym z pewnym rzeczywistym zjawiskiem. Zdanie "Kastor jest gwiazdą pojedynczą" jest hipotezą dotyczącą zjawiska. Podobnie zdanie "orbita komety Kiess ma kształt elipsy". W astronomii i w innych naukach empirycznych najczęściej interesują nas hipotezy dotyczące zjawisk. W wielu wypadkach hipotezy tego typu mogą być zweryfikowane w sposób ścisły (pewny). Np. dysponując dostatecznie dużą liczbą obserwacji możemy zweryfikować hipotezę: czy ciało niebieskie widoczne na płytach fotograficznych jest satelitą Jowisza? Jednak istnieje sporo przypadków kiedy krzyżowy test hipotezy dotyczącej zjawiska następuje z trudnością. Przykładowo, jest nim przypadek orbity komety wyznaczonej jedynie z trzech obserwacji, gdyż taką orbitę często można zaklasyfikować zarówno jako eliptyczną, paraboliczną bądź hiperboliczną. I dopiero kiedy napłyną nowe wystarczająco dokładne dane, początkowy werdykt o kształcie orbity będzie mógł być zweryfikowany.

Bywa, że podjęcie decyzji można bez problemu zawiesić nawet na dłuższy czas. Ale są i takie okoliczności, które wymagają decyzji natychmiastowej i w takich przypadkach możemy znaleźć się w trudnym położeniu, kiedy to np. pomimo dużej liczby obserwacji, ale przecież obciążonych niepewnościami pomiarowymi, krzyżowy test o kształcie orbity nie może być dokonany. Np. orbita NEA jest kolizyjna z Ziemią, mamy dwa dni do zderzenia. Co robić? Podjęcie decyzji jest wtedy związane jest ze szczególnym niebezpieczeństwem

popęlnienia błędu. I właśnie z tego powodu warto posługiwać się metodami wspierającymi proces podejmowania decyzji, zwłaszcza takimi, które zapewniają najmniejsze ryzyko błędu. Opracowanie m.in. tego rodzaju metod jest przedmiotem statystyki matematycznej.

Podejście statystyczne oznacza tu, że potrafimy zaproponować sposób na wyznaczenie względnych częstości występowania wszystkich wartości wielkości obserwowanej, np. postulując odpowiednią funkcję rozkładu prawdopodobieństwa. Dzięki temu, od rzeczywistego zjawiska możemy przejść do jego modelu stochastycznego, w którym rzeczywiste obserwacje mają swe odpowiedniki w postaci zmiennych losowych, a hipoteza dotycząca zjawiska \mathcal{H} , ma odpowiednik w hipotezie H dotyczącej modelu statystycznego.

Zatem hipoteza H nie dotyczy samego zjawiska ale zmiennej losowej, jej rozkładu postulowanego w ramach statystycznego modelu, stąd nazywana jest hipotezą statystyczną. Dowolne przypuszczenie dotyczące rozkładu zmiennej losowej, jego parametrów nazywane jest *hipotezą statystyczną*.

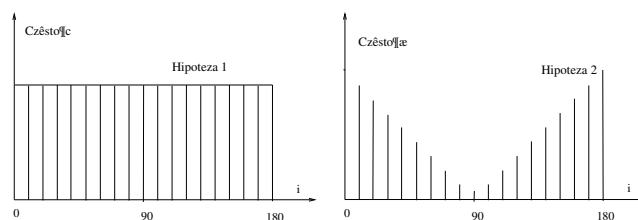
Hipotezę zawierającą tylko jedno przypuszczenie (założenie) nazywamy *hipotezą prostą*. Każda hipoteza, która nie jest prostą jest *hipotezą złożoną*, tzn. złożoną z dwóch lub większej liczby hipotez prostych. Np. gdy λ jest parametrem rozkładu wykładniczego, to hipoteza, że $\lambda = 5$ jest hipotezą prostą. Hipoteza, że $\lambda > 5$ składa się z nieskończonej liczby hipotez typu $\lambda = 5.5, \lambda = 6.0, \dots$

Przykład 1. Niech $(\alpha_i, \delta_i), t_i, i = 1, 2, \dots, n$, reprezentują obserwowaną rektascensję i deklinację, oraz moment czasu obserwacji komety. Niech hipoteza \mathcal{H} — kometa ma orbitę eliptyczną, będzie hipotezą dotyczącą zjawiska. Można zbudować model stochastyczny dla tych obserwacji i w ramach niego sformułować hipotezę statystyczną H — odpowiednik hipotezy \mathcal{H} . Np. w ramach hipotezy \mathcal{H} twierdzimy, że, elementy orbity komety mają wartości: $e = 0.9, q = 0.2, \omega = 123.0, \Omega = 232.1, i = 33.0, T = 2144323.5$.

W celu weryfikacji tej hipotezy moglibyśmy postulować model, w którym obserwowane współrzędne α_i, δ_i oraz wartości α_i^*, δ_i^* reprezentujące obliczone współrzędne komety na wspólne momenty t_i , różnią się o wartości podlegające rozkładowi normalnemu $\mathcal{N}(0, \sigma)$. Hipotezą statystyczną H , odpowiednikiem hipotezy dotyczącej zjawiska \mathcal{H} , byłoby zatem zdanie: wartości oczekiwane obserwowanej zmiennej losowej $(\alpha_i - \alpha_i^*, \delta_i - \delta_i^*)$ odpowiadają podanej orbicie eliptycznej.

Przykład 2. Istnieją sytuacje, w których więcej niż jedna hipoteza statystyczna może być traktowana jako adekwatny odpowiednik tej samej hipotezy dotyczącej zjawiska. W 1773 roku Laplace przedstawił pracę dotyczącą średniego nachylenia orbit komet. Laplace zamierzał rozstrzygnąć czy komety pochodzą z Układu Słonecznego. Przypuszczenie, że komety są obiektami spoza Układu Słonecznego będzie więc hipotezą \mathcal{H} , dla której trudno jest zrealizować test krzyżowy. Dlatego Laplace zbudował model omawianego zjawiska, w którym statystyczną hipotezę H można traktować jako adekwatny odpowiednik \mathcal{H} . Laplace wnioskował — jeśli komety są pochodzenia międzygwiazdowego wówczas ich orbity powinny być tak zorientowane, że o obserwowanych kometach moglibyśmy powiedzieć, że są “wrzucane” do Układu Planetarnego z dowolnego kierunku.

Obok hipotezy \mathcal{H} Laplace sformułował hipotezę statystyczną H_1 orzekającą, iż nachylenia orbit komet są zmiennymi losowymi o rozkładzie jednostajnym (patrz rysunek 6.1). Oprócz hipotezy H_1 , Laplace rozważał jeszcze drugą hipotezę statystyczną H_2 stwierdzającą, że prawdopodobieństwo ruchu prostego komety jest takie samo jak ruchu wstecznego i oba wynoszą 0.5.



Rysunek 6.1: Rysunki ilustrujące rozkłady nachyleń orbit komet, odpowiadające hipotezom H_1 i H_2 sformułowanym przez Laplace'a.

Jest to ciekawa sytuacja, gdyż pojedynczej hipotezie \mathcal{H} , w ramach tego samego modelu odpowiadają dwie hipotezy statystyczne. Która z nich lepiej odpowiada hipotezie \mathcal{H} ? Wydaje się, że w obliczu międzygwiazdowego pochodzenia komet lepszą jest hipoteza H_1 .

Rozróżnienie pomiędzy hipotezą dotyczącą zjawiska a jej statystycznym odpowiednikiem jest bardzo ważne. Jeśli dane zjawisko jest adekwatnie reprezentowane przez model stochastyczny wówczas, odrzucenie hipotezy statystycznej wzbudza poważne wątpliwości co do prawdziwości hipotezy \mathcal{H} . Jeśli hipoteza statystyczna nie jest odrzucona, mimo wszystko należy być ostrożnym w interpretowaniu tego faktu jako silnego potwierdzenia hipotezy dotyczącej zjawiska \mathcal{H} . ↗

6.3 Hipoteza zerowa i alternatywna

Przedstawimy krótki przegląd niektórych metod weryfikacji hipotez statystycznych dotyczących standardowych problemów, rozpracowanych poza astronomią, a które można stosować w obrębie astronomii. W wyborze metody pozwalającej na odrzucenie lub nie hipotezy statystycznej, musimy kierować się celem jaki stawiamy przed weryfikacją. Do testowania hipotez przystępujemy jedynie wtedy gdy nie mamy pewności jaką decyzję należy podjąć? Oczywiście bardzo chcielibyśmy podejmować jedynie decyzje poprawne.

Przystępując do testu statystycznego formułujemy *hipotezę zerową* H_0 której wiarygodność zostanie zweryfikowana. Np. “wariancje dwóch populacji o rozkładzie normalnym są sobie równe”.¹

Hipoteza zerowa ma konkurencyjną *hipotezę alternatywną*, o przeciwnej zawartości. Jeśli hipoteza zerowa wyraża przypuszczenie, że wartość oczekiwana zmiennej losowej o rozkładzie normalnym wynosi $\mu = 10$, to hipoteza alternatywna orzeka, że np. $\mu \neq 10$. Zapisujemy to w skrócie w następujący sposób

$$H_0 : \mu = 10, \quad H_1 : \mu \neq 10$$

↗

6.4 Błędy pierwszego i drugiego rodzaju

Postawiona hipoteza statystyczna może być prawdziwa albo fałszywa i aby to rozstrzygnąć musi zostać zweryfikowana. Ponieważ stosujemy tu metody statystyczne, w wyniku procesu weryfikacji narażeni jesteśmy na dwa rodzaje błędów:

¹Nie wszystkie stwierdzenia, hipotezy mają charakter statystyczny. Stwierdzenie: “ w roku 2222 będzie koniec świata” nie jest hipotezą statystyczną, gdyż niczego nam nie wiadomo o statystycznych własnościach tego zjawiska.

- *błąd pierwszego rodzaju*, który popełniamy gdy odrzucamy hipotezę prawdziwą,
- *błąd drugiego rodzaju*, który popełniamy w efekcie przyjęcia hipotezy fałszywej.

Następstwa tych błędów mogą być różne i muszą być starannie rozważone w każdym konkretnym przypadku. Ryzyko popełnienia błędu pierwszego rodzaju mierzone jest za pomocą *poziomu istotności* α równemu prawdopodobieństwu popełnienia tego właśnie błędu. Statystyczne testy zwykle są tak konstruowane, że z góry zakłada się wartość prawdopodobieństwa popełnienia błędu pierwszego rodzaju. ↗

6.5 Elementy statystycznego testu hipotezy zerowej

W celu weryfikacji hipotez statystycznych wykorzystywane są specjalnie dobrane (zależnie od badanego problemu) zmienne losowe o znanych rozkładach. Np. są to: zmienna \tilde{z} standardowego rozkładu normalnego, zmienna \tilde{F} o rozkładzie Fishera-Snedecora, zmienna \tilde{t} o rozkładzie studenta czy zmienna $\tilde{\chi}^2$ o rozkładzie chi-kwadrat. Ponieważ w tym paragrafie nie odwołujemy się do konkretnej postaci rozkładu, zmienne losowe wykorzystywane do weryfikacji hipotez oznaczать będziemy jako \tilde{K} .

Testem, sprawdzianem statystycznym (krótko *testem, sprawdzianem*) a niekiedy *statystyką*, nazywamy zmienną losową \tilde{K} , która służy do weryfikowania hipotezy H_0 . Przykładowo, jeśli weryfikujemy hipotezę zerową orzekającą o równości wariancji dwóch populacji o rozkładach normalnych, to jako sprawdzian \tilde{K} przyjmujemy stosunek wariancji uzyskanych z dwóch prób

$$F = \frac{S_1^2}{S_2^2}$$

Ponieważ w rezultacie kolejnych eksperymentów obliczone z prób wariancje przyjmują różne nieznanne z góry wartości, zmienna \tilde{F} jest zmienną losową o rozkładzie Fishera-Snedecora.

Weryfikacja badanej hipotezy odbywa się na podstawie danych pomiarowych z próby. Pomiar wykorzystujemy do obliczenia wartości zmiennych będących argumentami testu (jak np. wartość średnia, wartość wariancji), potrzebnych do wyznaczenia tzw. *wartości obserwowanej* K_p sprawdzianu \tilde{K} .

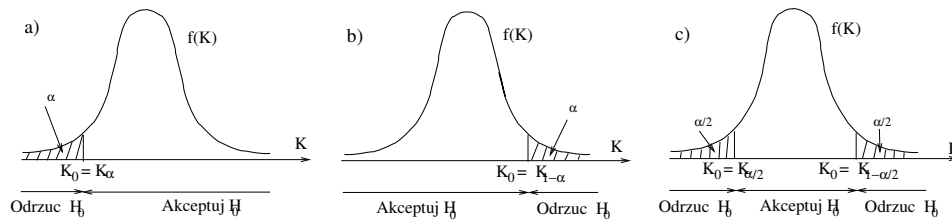
Przykład 3. Korzystając z danych pochodzących z prób wylosowanych z dwóch populacji o rozkładach normalnych, obliczono argumenty testu: wariancję $S_1^2 = 20$ i $S_2^2 = 5$. Z ich pomocą obliczono wartość obserwowaną testu \tilde{F}_p

$$F_p = \frac{S_1^2}{S_2^2} = 4$$

W przypadku dowolnego testu \tilde{K} , zbiór jego wszystkich możliwych wartości można rozbić na dwa rozłączne podzbiory:

1. podzbiór zawierający te wartości \tilde{K} , dla których hipoteza zerowa jest odrzucana,
2. podzbiór tych wartości testu, dla których hipoteza zerowa jest zachowana.

Pierwszy podzbiór służy do definicji *obszaru krytycznego*, którym jest zbiór wartości sprawdzianu \tilde{K} , dla których hipotezę zerową odrzucamy. Drugi podzbiór służy do definicji *obszaru zachowania* hipotezy, należą doń te wartości testu \tilde{K} , dla których hipoteza zerowa jest akceptowana.



Rysunek 6.2: Ilustracja obszarów krytycznych a) obszar lewostronny, b) obszar prawostronny, są to tzw. obszary jednoogonowe. c) obszar dwustronny (dwoogonowy). Położenia punktów krytycznych K_0 są identyczne z położeniami kwantyli, dla których dystrybuanty $F(K_0) = \alpha$, $F(K_0) = 1 - \alpha$, $F(K_{01}) = \alpha/2$, $F(K_{02}) = 1 - \alpha/2$, odpowiednio.

Możemy teraz sformułować podstawową zasadę weryfikacji hipotez statystycznych: jeżeli wartość zaobserwowana K_p sprawdzianu, NIE należy do obszaru krytycznego to hipotezę H_0 zachowujemy. W wypadku przeciwnym hipotezę H_0 odrzucamy.

Gdy test \tilde{K} jest jednowymiarową ciągłą zmienną losową, wszystkie możliwe wartości testu należą do pewnego przedziału. W takim przypadku obszar krytyczny, jak i obszar zachowania hipotezy są przedziałami, a zatem istnieją punkty, które te przedziały definiują. Nazywamy je *punktami krytycznymi* i oznaczamy przez K_0 .

Obszary krytyczne mogą być jednostronne i dwustronne. Obszar krytyczny *jednostronny* (jednoogonowy) jest to obszar prawostronny albo lewostronny (patrz rysunki 6.2a,b). Obszarem krytycznym *prawostronnym* nazywamy obszar określony nierównością $K > K_0$, obszarem krytycznym *lewostronnym* nazywamy obszar określony nierównością $K < K_0$.

Obszarem krytycznym *dwustronnym* (dwoogonowym) (patrz rysunek 6.2c) nazywamy obszar określony nierównościami: $K < K_{01} \wedge K > K_{02}$, gdzie $K_{01} < K_{02}$. W przypadku gdy punkty krytyczne są położone symetrycznie względem zera, obszar krytyczny dwustronny określony jest nierównością $|K| > K_0$, gdzie $K_0 > 0$. ↑

6.5.1 Wyznaczenie obszaru krytycznego

Wyznamy najpierw obszar krytyczny prawostronny określony nierównością

$$K > K_0$$

co wymaga ustalenia wartości punktu krytycznego K_0 . W tym celu przyjmujemy dostatecznie małe prawdopodobieństwo α (*poziom istotności*) popełnienia błędu I-rodzaju, po czym szukamy wartości kwantyla K_0 , dla którego

$$P[K > K_0] = \alpha \tag{6.1}$$

A zatem, wartość krytyczna K_0 obszaru prawostronnego jest to wartość kwantyla, identycznego z argumentem, dla którego dystrybuanta $F(\tilde{K} = K_0) = 1 - \alpha$ (patrz rysunek 6.2). Można ją łatwo wyznaczyć z funkcji odwrotnej do dystrybuanty danego rozkładu. Wyznaczenie obszaru krytycznego lewostronnego przebiega analogicznie, obszar ten spełnia nierówność

$$K < K_0$$

A położenie punktu krytycznego wyznaczamy z warunku: jeżeli hipoteza zerowa jest prawdziwa, to prawdopodobieństwo przyjęcia przez test \tilde{K} wartości mniejszej od K_0 jest równe przyjętemu

poziomowi istotności α

$$P[K < K_0] = \alpha \quad (6.2)$$

Wartość krytyczna K_0 odpowiada kwantylowi, dla którego mamy

$$F(K_0) = \alpha$$

Dwustronny obszar krytyczny określają nierówności

$$K < K_{01} \quad \text{i} \quad K > K_{02}$$

A punkty krytyczne K_1, K_2 wyznaczone są z warunku

$$P[K < K_{01}] + P[K > K_{02}] = \alpha \quad (6.3)$$

K_{01}, K_{02} odpowiadają kwantylom, dla których dystrybuanty wynoszą, odpowiednio

$$F(K_{01}) = \alpha/2 \quad \text{i} \quad F(K_{02}) = 1 - \alpha/2$$

Jeśli rozpartujemy test symetryczny względem zera to wyznaczamy punkty krytyczne symetryczne względem zera, tzn. wartości $-K_0$ i K_0 , gdzie $K_0 > 0$. Mamy tu

$$P[K < K_0] = P[K > K_0]$$

a uwzględniając równość (6.3) mamy

$$P[K > K_0] = \alpha/2 \quad (6.4)$$

przy czym

$$F(K_0) = 1 - \alpha/2$$

Wartości kwantyli odpowiadające punktom krytycznym obliczamy za pomocą odpowiedniego oprogramowania lub odczytujemy w zbudowanych w tym celu tablicach. ↗

6.5.2 Wnioskowanie statystyczne

W każdym z trzech rodzajów hipotez możemy wnioskować w bardzo podobny sposób. Dla hipotezy jednoogonowej, prawostronnej mielibyśmy takie rozumowanie. Jeśli hipoteza zerowa jest prawdziwa, to prawdopodobieństwo przyjęcia przez sprawdzian \tilde{K} wartości K większych od K_0 , czyli $P(K_p > K_0)$ jest równe przyjętemu poziomowi istotności α . Jest to małe prawdopodobieństwo, a z zasady “niemożliwości” zajścia zdarzeń małoprawdopodobnych wynika, że jeśli hipoteza zerowa jest prawdziwa, to w pojedynczym eksperymencie zdarzenie $K_p > K_0$ nie powinno zajść. Jeśli jednak zdarzenie to zaszło to najprawdopodobniej dlatego, że hipoteza zerowa jest fałszywa i powinna być odrzucona.

Jednak w wykonywanym teście, zaobserwowana wartość sprawdzianu K_p może być większa od K_0 niekoniecznie dlatego, że hipoteza zerowa jest fałszywa. Istnieją także inne powody jak: mała liczebność próby, słaba dokładność metody eksperymentu.

Ale mogło też po prostu wydarzyć się to, czego nikt nie oczekiwał — zaszło zdarzenie bardzo mało prawdopodobne. Wówczas, jeśli odrzucimy hipotezę prawdziwą popełniamy *błąd pierwszego rodzaju*. Prawdopodobieństwo popełnienia takiego błędu jest równe wartości poziomowi istotności α .

Jeśli hipotezy nie odrzucono *nie oznacza to*, że została ona udowodniona.² Dlatego właściwym będzie jedynie twierdzić, że “dane obserwacyjne zgadzają się z hipotezą zerową, a więc *nie ma podstaw* do jej odrzucenia”. W praktyce rozważaną hipotezę można dalej weryfikować przez sprawdzenie jej innymi sposobami, przez powtórzenie eksperymentu przy zwiększonej liczbie próby.

Odrzucenie hipotezy ma miejsce w sposób bardziej kategoryczny niż jej akceptacja. Być może ma to źródło w tym, że analogicznie jak w matematyce jeden kontrprzykład wystarczy aby obalić twierdzenie. Zatem gdy okaże się, że zaobserwowana wartość testu należy do obszaru krytycznego, to fakt ten jest kontrprzykładem dla hipotezy zerowej, a więc należy ją odrzucić. Oczywiście nie jest to analogia w ścisłym sensie. [†]

6.6 Przykłady weryfikacji hipotez, testy parametryczne

6.6.1 Schemat konstrukcji testu parametrycznego

Test nazywamy *parametrycznym*, jeżeli dotyczy hipotezy precyzującej wartości parametrów funkcji rozkładu o znanej postaci. Budowa testu hipotez parametrów rozkładu przebiega zgodnie z następującym schematem:

- w kroku pierwszym formułujemy hipotezę dotyczącą zjawiska. Jeśli nie ma ona charakteru hipotezy statystycznej (nie dotyczy bezpośrednio rozkładu zmiennej losowej), należy zaplanować eksperyment, którego wyniki (zmiennie losowe) w jakiś sposób są z nią związane,
- w następnym kroku, jeszcze przed pojawieniem się wyników eksperymentu, formułujemy hipotezę statystyczną H_0 ,
- wykonujemy eksperyment i w oparciu o wyniki pomiarów obliczamy wartość obserwowaną statystyki $\tilde{K} = K_p$, tzw. sprawdzianu hipotezy. Zmienna \tilde{K} musi mieć znany rozkład prawdopodobieństwa,
- ustalamy poziom istotności α testu, czyli wartość prawdopodobieństwa odrzucenia hipotezy zerowej wówczas, gdy jest ona prawdziwa (błąd I-go rodzaju),
- dla obranego poziomu istotności wyznaczamy punkt krytyczny K_0 , czyli obszar krytyczny wartości sprawdzianu \tilde{K} ,
- jeśli obserwowana wartość K_p leży w obszarze krytycznym, testowaną hipotezę H_0 odrzucamy. W wypadku przeciwnym nie mamy podstaw do jej odrzucenia, hipotezę H_0 zachowujemy.

²Dla matematyka podanie przykładu, dla którego rozważane twierdzenie jest prawdziwe, nie stanowi dowodu twierdzenia.

6.6.2 Test parametryczny dla średniej z próby przy znanej wariancji

Przykład 4. W znanym czasopiśmie “Astronomia dla inteligentnych Pań i Panów” autor rubryczki “Gwiazdy i gwiazdeczki” twierdzi, że średnice gwiazd z otoczenia Słońca mają rozkład $\mathcal{N}(\mu = 10.22, \sigma = 0.05)$, w jednostkach równych średnicy Słońca. Któraś z inteligentnych czytelniczek postanowiła to sprawdzić i w tym celu nabyła w firmie “SkyEyes-2003” doskonały teleskop “AT-10/50”, którym wyznaczyła rozmiary 22 gwiazd z otoczenia Słońca. Z wypiekami na twarzy obliczyła wartość średnią i okazało się, że $\bar{x} = 10.98$?!?! Hmmm! — pomyślała czytelniczka, czy to oznacza, że w moim ulubionym czasopiśmie wypisują bzdury? Przez moment była w kropce, ale na szczęście przypomniała sobie o testowaniu hipotez i przeprowadziła następujące rozumowanie.

“Niech będzie prawdą, że wartość średnia średnicy gwiazd rzeczywiście wynosi $\mu_0 = 10.22$. Jeśli tak naprawdę jest, to gdybym wybrała jakąś gwiazdę, tak na chybił trafił, jej średnica powinna być bliska tej wartości. A już na pewno średni rozmiar 22 gwiazd powinien niewiele różnić się od $\mu_0 = 10.22$. Słowa “powinien niewiele” muszę ująć językiem statystyki, czyli np. tak — jeśli świat jest taki jak twierdzi autor rubryczki, to niewielka, wyznaczona z obserwacji różnica $(\bar{x} - \mu_0)$ powinna być czymś typowym, o dużym prawdopodobieństwie wystąpienia. A ponieważ mamy do czynienia z rozkładami normalnymi (wartości średnie!), prawdopodobieństwo to mogę wyrazić za pomocą standardowego rozkładu normalnego:

$$P \left[z_1 < \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < z_2 \right] = 1 - \alpha$$

gdzie poziom ufności α jest małą liczbą dodatnią.

Potrzebuję teraz wartości z_1, z_2 zmiennej standardowej, odpowiadające prawdopodobieństwu $1 - \alpha$. Są to dwa kwantyle określające przedział, w którym jeśli autor rubryczki ma rację powinna znajdować się wyznaczona z obserwacji wartość z_p ”

$$z_p = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{10.98 - 10.22}{0.05/\sqrt{22}} = 71.3$$

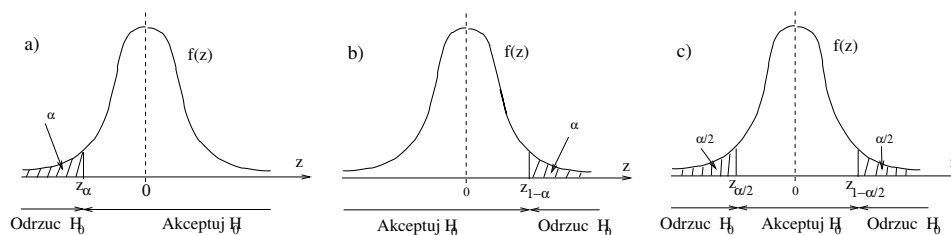
Dobierając $\alpha = 0.05$ nasza czytelniczka ze smutkiem stwierdziła, że $z_1 = -1.96, z_2 = 1.96$!. A to pozwala na stwierdzenie, że w jej ulubionym czasopiśmie można napotkać informacje wysrane z palca! Szok!

Powyższy przykład zawiera wszystkie elementy dwuogonowego testu wartości średniej μ populacji, przy znanej wariancji σ populacji. Mianowicie: testowano hipotezę zerową $H_0 : \mu = \mu_0$ w stosunku do hipotezy alternatywnej $H_1 : \mu \neq \mu_0$. Jako sprawdzian testu wykorzystano losową zmienną standardową z , test polegał na sprawdzeniu czy obserwowana wartość z_p tej zmiennej znajduje się w obszarze krytycznym wyznaczonym dla przyjętego poziomu istotności α . W ten sposób dostarczamy argumentów by udzielić odpowiedzi na pytanie czy średnia μ z populacji jest równa pewnej wartości μ_0 znanej *a priori*.

Zatem w celu weryfikacji hipotezy o wartości średniej w populacji, musimy obliczyć wartość obserwowaną testu, która tutaj przyjmuje wartość standardowej zmiennej \tilde{z}

$$z_p = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} \tag{6.5}$$

Następnie dla wybranego poziomu istotności α znajdujemy odpowiadające mu wartości kwantyli z_α i $z_{\alpha/2}$. Korzystamy ze stosownych tablic lub oprogramowania.



Rysunek 6.3: Testowanie hipotezy dla średniej z próby. a) i b) ilustrują przypadki testu jednoogonowego, c) dotyczy testu dwuogonowego.

Istnieją trzy warianty zerowej hipotezy H_0 : jednoogonowe hipotezy $H_1 : \mu > \mu_0$ albo $H_1 : \mu < \mu_0$, oraz hipoteza dwuogonowa $H_1 : \mu \neq \mu_0$ (patrz rysunek 6.3). Wówczas dla trzech wariantów można sformułować następujące testy:

1. $H_0 : \mu = \mu_0, H_1 : \mu < \mu_0$, (rysunek 6.3a).

Odrzucamy H_0 gdy $z_p < z_\alpha$, bowiem mamy

$$P \left[z_p = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < z_\alpha \right] = \alpha$$

2. $H_0 : \mu = \mu_0, H_1 : \mu > \mu_0$, (rysunek 6.3b).

Odrzucamy H_0 gdy $z_p > z_{1-\alpha}$, gdyż mamy

$$P \left[z_p = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} > z_{1-\alpha} \right] = \alpha$$

3. $H_0 : \mu = \mu_0, H_1 : \mu \neq \mu_0$, (rysunek 6.3c).

Odrzucamy H_0 gdy $z_p < -z_{\alpha/2}$ lub $z_p > z_{\alpha/2}$ bowiem mamy

$$P \left[z_{\alpha/2} < \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} < z_{1-\alpha/2} \right] = 1 - \alpha$$

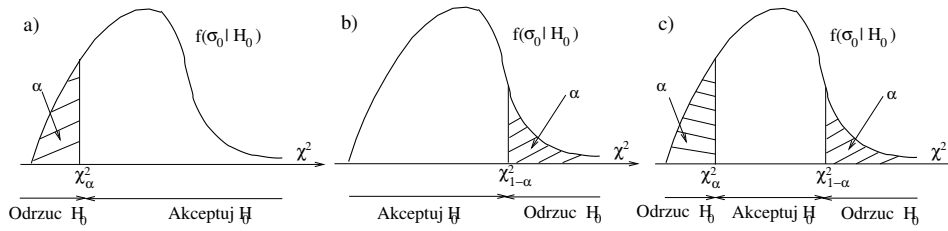
↗

6.6.3 Test dla średniej z próby przy nieznannej wariancji

Przy nieznanym odchyleniu standardowym σ populacji posługujemy się odchyleniem standardowym z próby s , a w miejsce zmiennej z jako test bierzemy zmienną losową studenta t

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \quad (6.6)$$

Możemy teraz wykonać analogiczne trzy testy jak w rozdziale 6.6.1 tyle, że kwantyle z_α i $z_{\alpha/2}$ zastępujemy kwantylami $t_{\alpha,m}$ i $t_{\alpha/2,m}$, gdzie $m = n - 1$ jest liczbą stopni swobody. ↗



Rysunek 6.4: Testowanie hipotezy o równości dwóch wariancji. a) i b) ilustrują przypadki testu jednoogonowego, c) dotyczy testu dwuogonowego.

6.6.4 Test dla dwóch średnich z próby

Tym razem interesuje nas problem porównania dwóch średnich z dwóch populacji, np. weryfikujemy hipotezy dotyczące różnicy pomiędzy dwiema średnimi. Zakładamy, że mamy do dyspozycji dwie niezależne próby pomiarowe o rozmiarach n_1 i n_2 , wzięte z dwóch populacji normalnych o średnich μ_1 i μ_2 , przy czym znane są jedynie wariancje σ_1^2 i σ_2^2 .

Nasze zadanie polega na testowaniu hipotezy zerowej $H_0 : \mu_1 - \mu_2 = \delta$, gdzie δ jest daną a priori stałą, wobec trzech hipotez konkurencyjnych: $H_1 : \mu_1 - \mu_2 \neq \delta$, $H_1 : \mu_1 - \mu_2 > \delta$, $H_1 : \mu_1 - \mu_2 < \delta$.

Weryfikacji hipotez można dokonać w oparciu o różnicę między średnimi $\bar{x}_1 - \bar{x}_2$, obliczonymi za pomocą pomiarów z obu prób, odpowiednio. Nie podamy tu metody konstrukcji tego testu, ograniczymy się jedynie do stwierdzenia, że jako wartość obserwowaną testu musimy policzyć zmienną standardową $z = z_p$, postaci

$$z_p = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{(\sigma_1^2/n_1 + \sigma_2^2/n_2)}} \quad (6.7)$$

a dalej postępujemy w sposób podany wcześniej dla pojedynczej średniej.

Jeżeli wartości wariancji σ_1^2 i σ_2^2 nie są znane, ale można uczynić założenie o ich równości, wówczas możemy je w testach zastąpić odpowiednimi wartościami S_1^2, S_2^2 obliczonymi z obu prób. Jednak zmienna losowa t , którą użyjemy do testowania hipotezy $H_0 : \mu_1 - \mu_2 = \delta$ będzie teraz postaci

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta}{\sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (6.8)$$

zmienna t jest zmienną losową o rozkładzie studenta o $m = n_1 + n_2 - 2$ stopniach swobody.

Hipotezy alternatywne formułujemy analogicznie do poprzednich biorąc w miejsce kwantyli z_α i $z_{\alpha/2}$ wartości $t_{\alpha,m}$ i $t_{\alpha/2,m}$. \uparrow

6.6.5 Weryfikacja hipotezy dla wariancji

Niech będzie dana próba o rozmiarze n wzięta z populacji normalnej. Możemy postawić hipotezę $H_0 : \sigma^2 = \sigma_0^2$, naprzeciw trzem alternatywom $H_1 : \sigma^2 \neq \sigma_0^2$, $H_1 : \sigma^2 > \sigma_0^2$, $H_1 : \sigma^2 < \sigma_0^2$, gdzie σ_0^2 jest określoną a priori (np. wziętą z pewnej teorii lub ulubionej gazetki) wartością wariancji.

Rozstrzygnięcia pośród tych hipotez dokonujemy w oparciu o wyznaczoną z próby wariancję S^2 , a jako sprawdzian bierzemy zmienną χ^2 . W tym celu obliczamy wartość wyrażenia

nia

$$\chi_m^2 = \frac{mS^2}{\sigma_0^2} \quad (6.9)$$

gdzie m jest liczbą stopni swobody z jaką policzono S^2 . Dla ustalonego poziomu istotności α możemy sformułować trzy testy (patrz rysunek (6.4)):

$$1. H_0 : \sigma^2 = \sigma_0^2, H_1 : \sigma^2 \neq \sigma_0^2.$$

Odrzucamy H_0 gdy $\chi_m^2 > \chi_{1-\alpha/2, m}^2$ lub $\chi_m^2 < \chi_{\alpha/2, m}^2$, bowiem

$$P \left[\chi_{\alpha/2, m}^2 < \frac{mS^2}{\sigma_0^2} < \chi_{1-\alpha/2, m}^2 \right] = 1 - \alpha$$

$$2. H_0 : \sigma^2 = \sigma_0^2, H_1 : \sigma^2 < \sigma_0^2.$$

Odrzucamy H_0 gdy $\chi_m^2 < \chi_{\alpha, m}^2$, bowiem

$$P \left[\frac{mS^2}{\sigma_0^2} < \chi_{\alpha, m}^2 \right] = \alpha$$

$$3. H_0 : \sigma^2 = \sigma_0^2, H_1 : \sigma^2 > \sigma_0^2.$$

Odrzucamy H_0 gdy $\chi_m^2 > \chi_{1-\alpha, m}^2$, bowiem

$$P \left[\frac{mS^2}{\sigma_0^2} > \chi_{1-\alpha, m}^2 \right] = \alpha$$

↑

6.6.6 Porównanie dwóch wariancji

Nieco inaczej przebiega test równości dwóch oszacowań wariancji na podstawie prób wziętych z tej samej populacji, czyli o tej samej średniej μ . Np. mierzymy odległość kątową dwóch gwiazd różnymi instrumentami pozbawionych błędów systematycznych. Czy pomiary mają jednakowe wariancje, czy oba instrumenty są jednakowo dokładne?

Testu dokonujemy w oparciu o stosunek dwóch empirycznych wariancji. Niech będą dane dwie niezależne próby o rozmiarach n_1 i n_2 wzięte z populacji normalnych o wariancjach σ_1^2, σ_2^2 . Stawiamy hipotezę zerową $H_0 : \sigma_1^2 = \sigma_2^2$. Weryfikacji hipotez dokonujemy za pomocą statystyki \tilde{F}

$$F_{m_1, m_2} = \frac{S_1^2}{S_2^2} \quad (6.10)$$

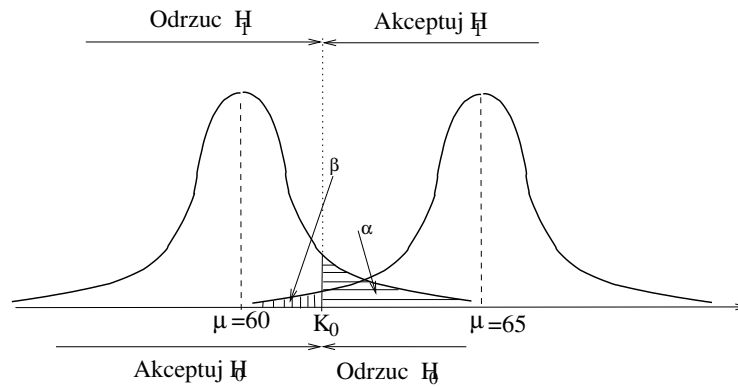
gdzie S_1^2, S_2^2 są wariancjami z próby, $m_1 = n_1 - 1, m_2 = n_2 - 1$ są liczbami stopni swobody. Jeśli obie wariancje są jednakowe, stosunek (??) powinien być bliski jedności

Możliwe trzy hipotezy testujemy w sposób następujący :

$$1. H_0 : \sigma_1^2 = \sigma_2^2, H_1 : \sigma_1^2 > \sigma_2^2.$$

Odrzucamy H_0 gdy $F_{m_1, m_2} > F_{1-\alpha, m_1, m_2}$, bowiem

$$P \left[\frac{S_1^2}{S_2^2} > F_{1-\alpha, m_1, m_2} \right] = \alpha$$



Rysunek 6.5: Ilustracja obszarów α i β odpowiadających prawdopodobieństwu popełnienia błędu pierwszego i drugiego rodzaju, odpowiednio. Obszary akceptacji i odrzucenia hipotezy H_0 odpowiadają sytuacji, w której jest ona prawdziwa czyli wartości średniej $\mu = 60$. Analogiczna uwaga dotyczy hipotezy H_1 . Widzimy, że prawdopodobieństwa α i β jednocześnie nie mogą być dowolnie małe.

$$2. H_0 : \sigma_1^2 = \sigma_2^2, H_1 : \sigma_1^2 < \sigma_2^2.$$

Odrzucamy H_0 gdy $F_{m_1, m_2} < F_{\alpha, m_1, m_2}$, bowiem

$$P \left[\frac{s_1^2}{s_2^2} > F_{\alpha, m_1, m_2} \right] = \alpha$$

$$3. H_0 : \sigma_1^2 = \sigma_2^2, H_1 : \sigma_1^2 \neq \sigma_2^2.$$

Odrzucamy H_1 gdy dla obliczonej z próby wartości testu mamy:

- dla prawego brzegu rozkładu mamy $F_{m_1, m_2} > F_{1-\alpha/2, m_1, m_2}$,
- dla lewego brzegu gdy $F_{m_1, m_2} < F_{\alpha/2, m_1, m_2}$.

gdzie α jest poziomem istotności testu. \uparrow

6.7 Moc testu

Obszar krytyczny budowaliśmy przy założeniu, że prawdopodobieństwo trafienia wartości statystyki \tilde{K} do tego obszaru, pod warunkiem, że hipoteza H_0 jest prawdziwa, jest równe α .

Chcąc ugruntować swe przekonanie o wiarygodności hipotezy nie warto wielokrotnie powtarzać tego samego testu. Warto natomiast ustalić prawdopodobieństwo trafienia testu \tilde{K}_p do obszaru krytycznego jeśli hipoteza zerowa jest fałszywa, a prawdziwa jest hipoteza alternatywna. Prawdopodobieństwo trafienia testu \tilde{K} do obszaru krytycznego gdy prawdziwa jest hipoteza alternatywna nazywane jest *mocą testu*. Można też powiedzieć, że moc testu jest prawdopodobieństwem odrzucenia hipotezy zerowej, gdy jest ona fałszywa.

Z powyższego wynika, że jeśli prawdopodobieństwo błędu drugiego rodzaju (przyjęcie hipotezy fałszywej) wynosi β , to moc testu (prawdopodobieństwo odrzucenia hipotezy fałszywej) jest równa $(1 - \beta)$. Jeśli moc $(1 - \beta)$ rośnie, to maleje prawdopodobieństwo β , im większa moc, tym mniejsze prawdopodobieństwo błędu drugiego rodzaju. Zatem testy powinny być tak skonstruowane by przy zadanym poziomie istotności α , obszar krytyczny sprawdzianu \tilde{K} odpowiadał maksymalnej mocy.

Jasne jest, że najlepszy obszar krytyczny uzyskamy jeśli oba prawdopodobieństwa błędów pierwszego i drugiego rodzaju są jak najmniejsze. Jednak przy danej liczebności próby jednoczesne zmniejszenie α i β jest niemożliwe, bowiem jeśli zmniejszamy α , to β nam rośnie. Wynika to z następującego rozumowania. Niech $\alpha = 0$, wówczas przyjęte są wszystkie hipotezy, a w tym i hipoteza fałszywa, czyli prawdopodobieństwo błędu drugiego rodzaju wynosi $\beta = 1$.

Prawdopodobieństwo popełnienia błędu pierwszego rodzaju znamy dokładnie, jest ono postulowane na początku testu. Inaczej wygląda sprawa z prawdopodobieństwem błędu drugiego rodzaju, jego obliczenie jest zawile, ponieważ jego wartość zależy od tego, którą z wartości dopuszczonych przez hipotezę alternatywną przyjmie badany przez nas parametr populacji. Oznacza to, że β zależy od wartości tego parametru, czyli w zasadzie mamy zależność $\beta(p)$.

Pokażemy to bliżej posługując się przykładem, niech $H_0 : \mu = 60$, $H_1 : \mu = 65$, liczebność próby $n = 100$. Załóżmy, że znamy odchylenie standardowe średniej μ i wynosi ono $\sigma = 20$. Sformułowany test jest testem prostym, prawostronnym: albo dojdziemy do przekonania, że średnia μ populacji wynosi 60, albo jest równa 65. W pierwszym wypadku zachowamy hipotezę zerową, w drugim ją odrzucimy. Dla poziomu istotności $\alpha = 0.05$ mamy, że wartość punktu krytycznego $K_0 = 61.29$. Na rysunku 6.5 po lewej stronie pokazano: krzywą gęstości dla estymatora \bar{x} odpowiadającego wartości średniej $\mu = 60$, punkt krytyczny K_0 oraz obszar krytyczny dla hipotezy H_0 . Ta sytuacja odpowiada założeniu, że hipoteza H_0 jest prawdziwa. Pole pod krzywą na prawo od punktu krytycznego $K_0 = 61,29$ wynosi 0.05 i jest to miara prawdopodobieństwa, że średnia z próby \bar{x} wypadnie po tej stronie punktu krytycznego, jeżeli średnia μ populacji rzeczywiście wynosi 60. Jak wiemy jest to prawdopodobieństwo błędu pierwszego rodzaju.

A w jaki sposób w tym przykładzie możemy oszacować prawdopodobieństwo błędu drugiego rodzaju. Z błędem tym mielibyśmy do czynienia tylko wtedy, gdybyśmy jako hipotezę H_0 uznali, że średnia w populacji wynosi 60, a tymczasem w rzeczywistości zgodnie z hipotezą H_1 wynosiłaby ona 65 (tzn. prawdziwą byłaby hipoteza alternatywna). A ponadto, wartość sprawdzianu, czyli tutaj \bar{x} wypadłaby na lewo od punktu krytycznego K_0 , a więc w obszarze przyjęcia hipotezy zerowej, patrz rysunek 6.5. Ile wynosi prawdopodobieństwo takiego zdarzenia, czyli prawdopodobieństwo β popełnienia błędu drugiego rodzaju? Jest ono równe polu na lewo od K_0 pod krzywą gęstości rozkładu odpowiadającą średniej $\mu = 65$. Te dwa prawdopodobieństwa moglibyśmy ująć jako:

$$\alpha = P(\bar{x} > K_0 | \mu = \mu_0)$$

$$\beta = P(\bar{x} < K_0 | \mu = \mu_1)$$

Prawdopodobieństwo α ustalono z góry. W omawianym przykładzie prawdopodobieństwo β możemy obliczyć za pomocą zmiennej normalnej standardowej:

$$\beta = P(\bar{x} < K_0 | \mu = \mu_1) = P\left(\frac{\bar{x} - \mu_1}{\sigma/\sqrt{n}} < \frac{K_0 - \mu_1}{\sigma/\sqrt{n}}\right) = P(z < -0.855) = 0.196$$

Mogliśmy oszacować β gdyż zgodnie z hipotezą H_1 twierdziliśmy, że $\mu = 65$. Jednak, gdybyśmy jako alternatywną postawili hipotezę złożoną $\mu > 60$, wówczas mamy sytuację, w której hipoteza zerowa może być fałszywa na nieskończenie wiele sposobów. Średnia populacji może przecież być równa 61, 61.5, 70.3 itd. Dla każdej z tych wartości mamy inną moc testu, rosnącą wraz ze wzrostem μ i zamiast o mocy testu, możemy tu mówić o funkcji mocy.

W jaki sposób należy dobierać najlepsze wartości α i β ? Odpowiedź zależy od ciężaru następstw obu błędów. Jeżeli błąd pierwszego rodzaju pociąga duże straty, a drugiego rodzaju małe, to należy przyjąć możliwie najmniejsze α . Jeżeli mamy α ustalone, to opierając się na twierdzeniach Neymana i Pearsona można obszar krytyczny wyznaczać tak aby wartość β była minimalna.

Na zakończenie wymienimy jeszcze czynniki wpływające na moc testu:

- moc zależy od odległości między wartością parametru zakładaną w hipotezie zerowej a jego wartością prawdziwą. Im większa odległość tym większa moc,
- moc zależy od odchylenia standardowego badanego parametru w populacji. Im mniejsze odchylenie, tym moc większa,
- moc zależy od liczebności próby, im liczniejsza próba, tym moc większa.
- moc zależy od poziomu istotności testu. Im niższy poziom istotności tym mniejsza moc testu.

↗

6.8 Zadanka na ćwiczenia

1. Zmierzono 10 razy deklinację pewnej gwiazdy. Pomiarów są niezależne jednakowo precyzyjne. W oparciu o pomiary obliczono wartość średnią oraz odchylenie standardowe $\bar{x} = 42^{\circ}12'14.4''$, $s = 3.7''$. Przy poziomie istotności 5% dokonaj testu hipotezy, że średnia populacji pomiarów deklinacji μ wynosi $42^{\circ}12'16''$, wobec hipotezy alternatywnej, że tak nie jest.

↗

Rozdział 7

Koncepcja wag popmmiarowych

Streszczenie

Nie ma komu napisać
I nie wiadomo czy będzie komu.

Słowa kluczowe: Pomiary niejednakowo dokładne, wagi pomiarów ^a ↗

^a[Modyfikowano AD 2015, maj 22.]

7.1 Koncepcja wagi pomiarowej

Niepewności pomiarowe mają różne własności, jedną z nich jest skupianie się wokół pewnej szczególnej wartości np. wartości średniej. Stopień skupienia określany jest za pomocą odchylenia standardowego σ , jego wartość jest miarą precyzji pomiarów. Jeśli σ jest małe, pomiary są silnie skupione wokół średniej, powiadamy, że precyzja pomiarów jest wysoka (nie mylić z dokładnością), i odwrotnie.

Istnieją inne miary precyzji, jedną z nich jest tzw. *waga pomiaru*. Można by powiedzieć, że za jej pomocą określa się dobroć danego pomiaru, stopień jego ważności, jego wagę w stosunku do innych pomiarów. Opracowując dane pomiarowe, ich wagę definiuje się w różny sposób. Dla naszych celów określimy ją jako wielkość odwrotnie proporcjonalną do wariancji pomiaru:

$$w = \frac{k}{\sigma^2}, \quad (7.1)$$

gdzie w to waga pomiaru a k jest stałą proporcjonalności.

W danej próbie pomiarowej dokonuje się wyboru jednego pomiaru jako *pomiar odniesienia*. Jego wariancję $\sigma^2 = \sigma_0^2$ przyjmuje się jako wariancję odniesienia, a jego wagę przyrównuje do jedności, czyli:

$$w = \frac{k}{\sigma_0^2} = 1, \quad \implies k = \sigma_0^2.$$

Wówczas wwagi pozostałych pomiarów można określać jako:

$$w = \frac{\sigma_0^2}{\sigma^2}. \quad (7.2)$$

Obserwacja o wadze równej jedności nazywana jest *obserwacją o wadze jednostkowej*.

Obok wariancji mamy w statystyce matematycznej pojęcie kowariancji. Pojawia się ono gdy wyznaczona wielkość jest zmienną losową dwu lub więcej wymiarową. Np. pomiar położenia ciał niebieskich wymaga wyznaczenia dwóch współrzędnych jest zatem zmienną losową dwuwymiarową. Kowariancja wyraża korelację (współzależność lub jej brak) między zmiennymi losowymi. W dalszym toku rozważań przyjmijmy, że obserwacje są niezależne, nieskorelowane, co oznacza, że ich kowariancja jest równa zeru.

Dla próby niezależnych pomiarów x_1, x_2, \dots, x_m o wariancjach $\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2$ (próba o rozmiarze m), praktycznym pociągnięciem jest pojęcie macierzy wariancji:

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \sigma_m^2 \end{bmatrix}. \quad (7.3)$$

Analogicznie można określić macierz wag:

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & w_m \end{bmatrix}. \quad (7.4)$$

Obierając w tej próbie wariancję któregoś z pomiarów jako wariancję odniesienia i stosując równanie (7.2) możemy (7.4) napisać w postaci:

$$\mathbf{W} = \sigma_0^2 \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_m^2} \end{bmatrix}, \quad (7.5)$$

co pociąga:

$$\mathbf{W} = \sigma_0^2 \sum^{-1}. \quad (7.6)$$

Uwaga!. Mimo poczynionych założeń o braku korelacji, równanie (7.7) jest prawdziwe także dla obserwacji skorelowanych.

↗

7.2 Obserwacje zrównoważone

Niech l_i , $i = 1, 2, \dots, n$, oznaczają zbiór rezultatów obserwacji o niejednakowej precyzji o odchyleniach standardowych σ_i .

Dokonajmy następującego zabiegu: pomnóżmy każdy pomiar l_i przez pewną liczbową stałą a_i . W ten sposób otrzymujemy nowy ciąg obserwacji $l_i^* = l_i a_i$.

Z prawa propagacji wariancji¹ wynika, że pomiary l_i^* mają odchylenia standardowe $\sigma_i^* = \sigma_i a_i$. Zatem ich wagi wynoszą

$$w_i^* = \frac{k}{\sigma_i^{*2}} = \frac{k}{\sigma_i^2 a_i^2} = \frac{w_i}{a_i^2}$$

Skoro tak, to zbiór takich obserwacji:

$$l_i \sqrt{w_i}, \quad (7.7)$$

jest zbiorem obserwacji jednakowo dokładnych (zrównoważonych), bowiem wszystkie obserwacje mają teraz jednakowe wagi, równe jedności.

↗

¹Będzie o nim mowa w następnym rozdziale

Rozdział 8

Podstawy techniki propagacji

Streszczenie

Nie ma komu napisać
I nie wiadomo czy będzie komu.

Słowa kluczowe: Propagacja funkcji gęstości zmiennej losowej, propagacja wartości oczekiwanej, wariancji i kowariancji, propagacja dla funkcji liniowych, propagacja dla blokowych wektorów losowych, mnożenie symboliczne. ^a ↗

^a[Modyfikowano AD 2015, maj 22.]

8.1 Wstęp

W świetnym podręczniku [1] określenie “propagacja błędów” zastąpiono zwrotem “zamiana zmiennych”, by zdaniem jego autorów uniknąć nieporozumień. W naszym skromnym odczuciu, do tego samego może prowadzić określenie “zamiana zmiennych”, dlatego pozwalamy sobie na zachowanie słowa “propagacja”, które będziemy wykorzystywali w szerszym i węższym znaczeniu. W znaczeniu szerszym propagacja dotyczy zależności funkcjonalnej zmiennych losowych i oznacza propagowanie statystycznych cech pomiędzy zmienną niezależną a zmienną zależną.

Niech $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)^T$ będzie n wymiarowym wektorem losowym o łącznej funkcji gęstości $f_x(x_1, x_2, \dots, x_n)$. Niech $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_m)^T$ będzie wektorem losowym którego składowe otrzymano za pomocą składowych wektora $\tilde{\mathbf{x}}$ i zależności:

$$\tilde{y}_k = u_k(\tilde{x}_1, \dots, \tilde{x}_n), \quad k = 1 \dots m$$

Zagadnienie propagacji polega na wyznaczeniu własności statystycznych wektora losowego $\tilde{\mathbf{y}}$ w oparciu o własności wektora losowego $\tilde{\mathbf{x}}$, w szczególności interesuje nas znalezienie funkcji gęstości $f_y(y_1, y_2, \dots, y_m)$. Ustalenie funkcji gęstości jest zadaniem najbardziej ogólnym, rzadko realizowanym w praktyce, najczęściej mamy do czynienia z zadaniami szczegółowymi jak:

- propagacja średnich (wartości oczekiwanych),
- propagacja błędów przypadkowych (wariancji i kowariancji),
- propagacja błędów systematycznych.

Poniżej omówimy wszystkie wymienione przypadki, ale ograniczając się jedynie do liniowych związków pomiędzy wektorami losowymi. \uparrow

8.2 Propagacja funkcji gęstości rozkładu prawdopodobieństwa

Jak wspomniano problem polega na znalezieniu funkcji gęstości $f_y(y_1, y_2, \dots, y_m)$ wektora losowego $\tilde{\mathbf{y}}$ w oparciu o funkcję gęstości $f_x(x_1, x_2, \dots, x_n)$ wektora losowego $\tilde{\mathbf{x}}$, przy danej funkcjonalnej zależności pomiędzy tymi wektorami.

Rozważmy najpierw przypadek jednowymiarowy. Niech będzie dana funkcja $\tilde{y} = g(\tilde{x})$, różniczkowalna, przy czym \tilde{x} ma rozkład prawdopodobieństwa określony przez funkcję gęstości $f_x(x)$. Zakładamy także istnienie funkcji odwrotnej $\tilde{x} = h(\tilde{y})$ jednoznacznej, ciągłej i różniczkowalnej.

Zatem, potrafimy każde zjawisko losowe \tilde{x}_i transformować w zjawisko $\tilde{y}_i = g(\tilde{x}_i)$ i odwrotnie. W tej sytuacji oczywistym jest, że

$$P(x_1 \leq \tilde{x} \leq x_2) = P(y_1 \leq \tilde{y} \leq y_2),$$

lub za pomocą funkcji gęstości, że:

$$\int_{x_1}^{x_2} f_x dx = \int_{y_1}^{y_2} f_y dy,$$

a ograniczając się do bardzo małych przedziałów, may też:

$$f_x(x) dx = f_y(y) dy,$$

A korzystając z reguł zamiany zmiennych pomiędzy \tilde{x} i \tilde{y} mamy:

$$f_x(x)dx = \left| f_x(h(y)) \frac{\partial h(y)}{\partial y} \right| dy = f_y(y)dy.$$

Zatem funkcja gęstości zmiennej losowej $\tilde{y} = g(\tilde{x})$ określona jest za pomocą:

$$f_y(y) = f_x(h(y)) \left| \frac{\partial h(y)}{\partial y} \right| \quad (8.1)$$

Przykład 1. Dana jest zależność $y = x^2$ oraz funkcja gęstości $f_x(x) = \exp\left(\frac{-x^2}{2}\right)$, wyznaczmy funkcję gęstości zmiennej y .

Skoro $g(x) = x^2$ to funkcja odwrotna ma postać:

$$h(y) = x = +\sqrt{y},$$

ponieważ bierzemy jedynie wartości dodatnie. Stąd:

$$\frac{\partial h(y)}{\partial y} = \frac{1}{2\sqrt{y}},$$

a zgodnie z równaniem (8.1), będzie:

$$f_y(y) = f_x(\sqrt{y}) \frac{1}{2\sqrt{y}}.$$

Kładąc do funkcji gęstości $f_x(x) = \exp\left(\frac{-x^2}{2}\right)$ zmienną $x = \sqrt{y}$ otrzymamy

$$f_y(y) = \exp\left(\frac{-y}{2}\right) \frac{1}{2\sqrt{y}}.$$

↗

8.2.1 Przypadek wielowymiarowy

Przejdziemy do zmiennej losowej wielowymiarowej. Niech będą dane m funkcji wektora losowego $\tilde{\mathbf{x}} = (\tilde{x}_1, \dots, \tilde{x}_n)$:

$$\tilde{y}_i = g_i(\tilde{x}_1, \dots, \tilde{x}_n) = g_i(\tilde{\mathbf{x}}) \quad i = 1 \dots m.$$

Pozwalają one na obliczenie składowych wektora losowego $\tilde{\mathbf{y}} = (\tilde{y}_1, \dots, \tilde{y}_m)$. Dalej założymy, że mamy funkcje odwrotne:

$$\tilde{x}_j = h_j(\tilde{y}_1, \dots, \tilde{y}_m) = h_j(\tilde{\mathbf{y}}) \quad j = 1 \dots n.$$

Jeżeli $f_x(\mathbf{x})$ jest łączną funkcją gęstości wektora losowego $\tilde{\mathbf{x}}$ to:

$$f_y(\mathbf{y}) = f_y(y_1, y_2, \dots, y_m) = f_x(h_1(y_1, y_2, \dots, y_m), h_2(y_1, y_2, \dots, y_m), \dots) |\mathbf{J}|.$$

jest łączną funkcją gęstości wektora losowego $\tilde{\mathbf{y}}$. Przy czym $|\mathbf{J}|$ jest wyznacznikiem macierzy Jacobiego transformacji odwrotnej $\tilde{x}_j = h_j(\tilde{\mathbf{y}})$. Jakobian ten oznaczany zwięźle jako:

$$\mathbf{J} = \frac{\partial \mathbf{x}}{\partial \mathbf{y}}$$

Propagacja rozkładów prawdopodobieństwa nie jest łatwym zagadnieniem, główne trudności dotyczą tu założenia o istnieniu funkcji odwrotnych h_j . W przypadku funkcji liniowych wiążących wektory losowe \tilde{y} i \tilde{x} sytuacja jest prostsza a w szczególności dla wektorów losowych \tilde{x} o rozkładzie normalnym, mianowicie, rezultat propagacji będzie miał także rozkład normalny. Z tego powodu propagacji najczęściej dokonuje się dla przypadków liniowych lub w oparciu o zlinearyzowane formy zależności pomiędzy wektorami losowymi \tilde{y} i \tilde{x} .

↗

8.3 Propagacja średniej

Średnia rozkładu jest ważnym parametrem statystycznym dlatego korzystając z definicji wartości oczekiwanej wyprowadzimy wyrażenie na bezpośrednią propagację wartości średniej.

Niech $\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n$ będą zmiennymi losowymi o znanych wartościach oczekiwanych $E[\tilde{x}_1], \dots, E[\tilde{x}_n]$. Utworzymy zbiór zmiennych $(\tilde{y}_1, \dots, \tilde{y}_m)$ będących funkcjami zmiennych $(\tilde{x}_1, \dots, \tilde{x}_n)$:

$$\tilde{y}_i = g_i(\tilde{x}_1, \dots, \tilde{x}_n) = g_i(\tilde{\mathbf{x}}) \quad i = 1 \dots m.$$

Wartości oczekiwane zmiennych \tilde{y}_i wynoszą:

$$E[\tilde{y}_i] = E[g_i(\tilde{x}_1, \dots, \tilde{x}_n)] \quad i = 1 \dots m.$$

a zgodnie z definicją wartości oczekiwanej:

$$E[\tilde{y}_i] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g_i(x_1, \dots, x_n) f_x(x_1, \dots, x_n) dx_1 \dots dx_n, \quad (8.2)$$

gdzie $f_x(x_1, \dots, x_n)$ jest funkcją gęstości zmiennych $(\tilde{x}_1, \dots, \tilde{x}_n)$. ↗

8.3.1 Propagacja średnich dla funkcji liniowych

W przypadku funkcji liniowych lub form zlinearyzowanych, formuły na propagację średnich są wyjątkowo proste. Dla zależności $\tilde{y} = a\tilde{x} + b$, wzór (8.2) przechodzi w:

$$E[\tilde{y}] = \int_{-\infty}^{\infty} (ax + b) f_x(x) dx = a \int_{-\infty}^{\infty} x f_x(x) dx + b \int_{-\infty}^{\infty} f_x(x) dx = aE[\tilde{x}] + b.$$

W przypadku większej liczby zmiennych będziemy mieli:

$$\tilde{y}_i = a_{0i} + a_{1i}\tilde{x}_1 + a_{2i}\tilde{x}_2 + \dots + a_{ni}\tilde{x}_n.$$

A wartości oczekiwane:

$$\begin{aligned} E[\tilde{y}_i] &= E[a_{0i} + a_{1i}\tilde{x}_1 + a_{2i}\tilde{x}_2 + \dots + a_{ni}\tilde{x}_n] + \\ &= E[a_{0i}] + E[a_{1i}\tilde{x}_1] + E[a_{2i}\tilde{x}_2] + \dots + E[a_{ni}\tilde{x}_n] + \\ &= a_{0i} + a_{1i}E[\tilde{x}_1] + a_{2i}E[\tilde{x}_2] + \dots + a_{ni}E[\tilde{x}_n], \end{aligned}$$

czyli

$$\mu_{yi} = a_{0i} + a_{1i}\mu_{x1} + a_{2i}\mu_{x2} \dots + a_{ni}\mu_{xn}. \quad (8.3)$$

Widzimy, że liniowy charakter zależności zachowany zostaje także dla wartości oczekiwanych. Warto też zauważyć, że liniowy charakter propagacji wartości średnich dla funkcji liniowych jest niezależny od rozkładu prawdopodobieństwa zmiennych $(\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_n)$.

Przykład 2. Dana jest zmienna losowa \tilde{x} i jej funkcja gęstości:

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

Jeżeli $\tilde{y} = \tilde{x}^2$ wyznacz μ_y .

Rozwiązanie sprowadza się do obliczenia całki

$$\mu_y = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} x^2 \exp \left[\frac{-1}{2} \left(\frac{x - \mu}{\sigma} \right)^2 \right] dx.$$

W tym celu podstawiamy

$$z = \frac{x - \mu}{\sigma\sqrt{2}},$$

skąd

$$z^2 = \frac{1}{2} \left[\frac{x - \mu}{\sigma} \right]^2,$$

$$x = \mu + \sigma\sqrt{2} z,$$

$$dx = \sigma\sqrt{2} dz.$$

Zatem

$$\begin{aligned} \mu_y &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} (\mu + \sigma\sqrt{2} z)^2 e^{-z^2} \sigma\sqrt{2} dz = \\ &= \frac{1}{\sqrt{\pi}} \left[\mu^2 \int_{-\infty}^{\infty} e^{-z^2} dz + 2\mu\sigma\sqrt{2} \int_{-\infty}^{\infty} z e^{-z^2} dz + 2\sigma^2 \int_{-\infty}^{\infty} z^2 e^{-z^2} dz \right] = \\ &= \frac{1}{\sqrt{\pi}} \left[\sqrt{\pi}\mu^2 + 0 + 2\sigma^2 \frac{2\sqrt{\pi}}{2} \right]. \end{aligned}$$

Ostatecznie

$$\mu_y = \mu^2 + \sigma^2.$$

↗

8.4 Propagacja wariancji i kowariancji

Niech \tilde{x}_1, \tilde{x}_2 będą zmiennymi losowymi o wartościach oczekiwanych μ_{x_1}, μ_{x_2} i łącznej funkcji gęstości $f(x_1, x_2)$. Rozważmy następujące funkcje:

$$\tilde{y}_1 = g_1(\tilde{x}_1, \tilde{x}_2), \quad \tilde{y}_2 = g_2(\tilde{x}_1, \tilde{x}_2).$$

Wariancje i kowariancja zmiennych \tilde{x}_1, \tilde{x}_2 mają postać:

$$\sigma_{x_1}^2 = E[(\tilde{x}_1 - \mu_{x_1})^2] = \int_{-\infty}^{\infty} (x_1 - \mu_{x_1})^2 f(x_1, x_2) dx_1 \quad (8.4)$$

$$\sigma_{x_2}^2 = E[(\tilde{x}_2 - \mu_{x_2})^2] = \int_{-\infty}^{\infty} (x_2 - \mu_{x_2})^2 f(x_1, x_2) dx_2 \quad (8.5)$$

$$\sigma_{x_1 x_2} = E[(\tilde{x}_1 - \mu_{x_1})(\tilde{x}_2 - \mu_{x_2})] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu_{x_1})(x_2 - \mu_{x_2}) f(x_1, x_2) dx_1 dx_2 \quad (8.6)$$

Analogicznie dla zmiennych \tilde{y}_1, \tilde{y}_2 mamy:

$$\sigma_{y_1}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_1(x_1, x_2) - \mu_{y_1})^2 f(x_1, x_2) dx_1 dx_2 \quad (8.7)$$

$$\sigma_{y_2}^2 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_2(x_1, x_2) - \mu_{y_2})^2 f(x_1, x_2) dx_1 dx_2 \quad (8.8)$$

$$\sigma_{y_1 y_2} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (g_1(x_1, x_2) - \mu_{y_1})(g_2(x_1, x_2) - \mu_{y_2}) f(x_1, x_2) dx_1 dx_2. \quad (8.9)$$

Wyrażenia (8.7)–(8.9) dla dwóch zmiennych losowych łatwo zmodyfikować na przypadek wielowymiarowy. ↗

8.4.1 Propagacja wariancji i kowariancji dla funkcji liniowych

Równania (8.7)–(8.9) nie są łatwe do wykorzystania w praktyce, całki po prawych stronach nierzadko następczą poważnych trudności. Dlatego pożądane są prostsze formuły, które uzyskuje się dla przypadku funkcji liniowych bądź form zlinearyzowanych zależności pomiędzy wektorami losowymi \tilde{y} i \tilde{x} .

Niech będą dane: zmienne losowe \tilde{x}_1, \tilde{x}_2 , ich wartości oczekiwane μ_{x_1}, μ_{x_2} oraz łączna funkcja gęstości $f(x_1, x_2)$. Rozważmy nowe zmienne \tilde{y}_1, \tilde{y}_2 związane z \tilde{x}_1, \tilde{x}_2 zależnościami liniowymi:

$$\begin{aligned} \tilde{y}_1 &= a_0 + a_1 \tilde{x}_1 + a_2 \tilde{x}_2 \\ \tilde{y}_2 &= b_0 + b_1 \tilde{x}_1 + b_2 \tilde{x}_2. \end{aligned} \quad (8.10)$$

Zgodnie z definicją wariancja zmiennej \tilde{y}_1 :

$$\begin{aligned} \sigma_{y_1}^2 &= E[(\tilde{y}_1 - \mu_{y_1})^2] = E[(a_0 + a_1 \tilde{x}_1 + a_2 \tilde{x}_2 - a_0 - a_1 \mu_{x_1} - a_2 \mu_{x_2})^2] = \\ &= E[(a_1(\tilde{x}_1 - \mu_{x_1}) + a_2(\tilde{x}_2 - \mu_{x_2}))^2] = \\ &= E[(a_1^2(\tilde{x}_1 - \mu_{x_1})^2 + a_2^2(\tilde{x}_2 - \mu_{x_2})^2 + 2a_1 a_2(\tilde{x}_1 - \mu_{x_1})(\tilde{x}_2 - \mu_{x_2}))] = \\ &= a_1^2 E[(\tilde{x}_1 - \mu_{x_1})^2] + a_2^2 E[(\tilde{x}_2 - \mu_{x_2})^2] + 2a_1 a_2 E[(\tilde{x}_1 - \mu_{x_1})(\tilde{x}_2 - \mu_{x_2})] \\ \sigma_{y_1}^2 &= a_1^2 \sigma_{x_1}^2 + a_2^2 \sigma_{x_2}^2 + 2a_1 a_2 \sigma_{x_1 x_2}. \end{aligned} \quad (8.11)$$

Wariancja drugiej zmiennej \tilde{y}_2 :

$$\sigma_{y_2}^2 = b_1^2 \sigma_{x_1}^2 + b_2^2 \sigma_{x_2}^2 + 2b_1 b_2 \sigma_{x_1 x_2}. \quad (8.12)$$

A ich kowariancja:

$$\sigma_{y_1 y_2} = a_1 b_1 \sigma_{x_1}^2 + a_2 b_2 \sigma_{x_2}^2 + (a_1 b_1 + a_2 b_2) \sigma_{x_1 x_2}. \quad (8.13)$$

Widzimy, że podobnie jak dla średnich, prawo propagacji wariancji i kowariancji dla funkcji liniowych *nie zależy od funkcji gęstości*.

Formuły (8.11)–(8.13) mają bardziej zwartą postać w notacji macierzowej, niech dane będą wektory losowe:

$$\tilde{\mathbf{y}} = [\tilde{y}_1 \ \tilde{y}_2]^T \quad i \quad \tilde{\mathbf{x}} = [\tilde{x}_1 \ \tilde{x}_2]^T.$$

Równania (8.10) w postaci macierzowej:

$$\tilde{\mathbf{y}} = \mathbf{c} + \mathbf{C} \tilde{\mathbf{x}},$$

gdzie

$$\mathbf{c} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} \quad \mathbf{C} = \begin{bmatrix} a_1 & a_2 \\ b_1 & b_2 \end{bmatrix}.$$

Niech:

$$\Sigma_{xx} = \begin{bmatrix} \sigma_{x_1}^2 & \sigma_{x_1 x_2} \\ \sigma_{x_1 x_2} & \sigma_{x_2}^2 \end{bmatrix} \quad \Sigma_{yy} = \begin{bmatrix} \sigma_{y_1}^2 & \sigma_{y_1 y_2} \\ \sigma_{y_1 y_2} & \sigma_{y_2}^2 \end{bmatrix}.$$

będą macierzami kowariancji wektorów losowych $\tilde{\mathbf{x}}$ i $\tilde{\mathbf{y}}$. Równaniom (8.11)–(8.13) opisującym propagację wariancji i kowariancji, można teraz nadać formę:

$$\Sigma_{yy} = \mathbf{C} \Sigma_{xx} \mathbf{C}^T. \quad (8.14)$$

↑

8.4.2 Propagacja wariancji i kowariancji dla funkcji nieliniowych

Łatwo pokazać, że macierz \mathbf{C} reprezentuje jacobian wektora $\tilde{\mathbf{y}}$ ze względu na $\tilde{\mathbf{x}}$:

$$\mathbf{J}_{y\mathbf{x}} = \frac{\partial \tilde{\mathbf{y}}}{\partial \tilde{\mathbf{x}}},$$

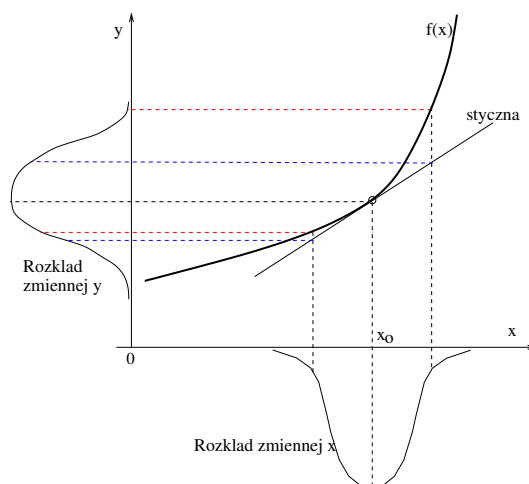
zatem, w formie bardziej ogólniej będzie:

$$\Sigma_{yy} = \mathbf{J}_{y\mathbf{x}} \Sigma_{xx} \mathbf{J}_{y\mathbf{x}}^T. \quad (8.15)$$

Liniowe równania nie występują często w praktyce, dlatego równania postaci (8.10) stanowią zwykle rezultat linearyzacji funkcji nieliniowych. Linearyzacja oznacza, że po rozwinięciu danej funkcji w szereg Taylora, z rozwinięcia bierzemy tylko dwa pierwsze wyrazy zaniedbując wszystkie wyrazy od rzędu drugiego począwszy.

Zauważmy, że występujące w rozwinięciu wyrazy zerowego rzędu nie są nam potrzebne, bowiem dla punktu x_1^0, x_2^0 (w jego otoczeniu dokonujemy rozwinięcia funkcji) wartości wzięte z form zlinearyzowanych odpowiadają wartościom a_0, b_0 , które nie propagują do $\sigma_{y_1}^2, \sigma_{y_2}^2, \sigma_{y_1 y_2}$. Propagują jedynie pochodne cząstkowe rozwinięcia, stąd równanie (8.15) jest przypadkiem ogólnym zarówno dla zależności liniowych jak i nieliniowych. W naszym przykładzie Jacobian \mathbf{J} zawiera wszystkie cztery pochodne obliczone w punkcie x_1^0, x_2^0 :

$$\mathbf{J}_{y\mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \frac{\partial y_1}{\partial x_2} \\ \frac{\partial y_2}{\partial x_1} & \frac{\partial y_2}{\partial x_2} \end{bmatrix}.$$



Rysunek 8.1: Zlinearyzowana forma funkcji $f(x)$ nie zawsze precyzyjnie propaguje własności zmiennej losowej \tilde{x} .

Stosowanie form zlinearyzowanych jest dopuszczalne jedynie dla dostatecznie małego przedziału dyspersji zmiennych \tilde{x}_1, \tilde{x}_2 w otoczeniu punktu x_1^0, x_2^0 . W obszarze naszych zainteresowań funkcje wyjściowe powinny być należycie aproksymowane przez formy liniowe. Na rysunku 9.1, ilustrujemy powyższe stwierdzenie w ramach przypadku jednowymiarowego. Zauważmy, że w przypadkach zlinearyzowanych funkcji, zmienne losowe przechodzą w postać zmiennych przyrostowych: Mianowicie po linearyzacji mamy:

$$\begin{aligned}x_i &= x_{i0} + \Delta x_i \\y_i &= y_{i0} + \Delta y_i.\end{aligned}$$

Zatem, własności np. błędów przypadkowych — a więc także ich rozkład, związane są teraz z przyrostami $\Delta x_i, \Delta y_i$. W obszarze liniowym rozważania geometryczne (patrz rys. 9.1) potwierdzają wcześniej otrzymany rezultat, że liniowa transformacja zmiennej losowej \tilde{x}_i o rozkładzie normalnym daje w rezultacie zmienną losową \tilde{y}_i również o rozkładzie normalnym.

Równanie (8.15) jest ogólnym równaniem propagacji macierzy wariancji dla wektora \tilde{y} nieliniowo zależnego od wektora \tilde{x} .

Wprowadzając macierze kofaktorowe zdefiniowane jako:

$$\mathbf{Q}_{xx} = \frac{1}{\sigma_0^2} \Sigma_{xx}, \quad \mathbf{Q}_{yy} = \frac{1}{\sigma_0^2} \Sigma_{yy},$$

równanie (8.15) można napisać w postaci:

$$\mathbf{Q}_{yy} = \mathbf{J}_{yx} \mathbf{Q}_{xx} \mathbf{J}_{yx}^T. \quad (8.16)$$

↑

8.5 Propagacja w przypadku blokowych wektorów losowych

Równanie (8.16) można uogólnić dalej, na przypadek wielokrotnej zależności pomiędzy wektorami losowymi. Niech wektory $\tilde{\mathbf{x}}_{n,1}, \tilde{\mathbf{t}}_{m,1}$ o n i m składowych, będą wektorami skorelowanymi o macierzach kofaktorowych $\mathbf{Q}_{xx}, \mathbf{Q}_{xt}, \mathbf{Q}_{tt}$. Niech wektory $\tilde{\mathbf{y}}_{q,1}, \tilde{\mathbf{z}}_{p,1}$ będą

funkcjami zmiennych wektorowych $\tilde{\mathbf{x}}_{n,1}, \tilde{\mathbf{t}}_{m,1}$:

$$\tilde{\mathbf{y}} = y(\tilde{\mathbf{x}}), \quad \tilde{\mathbf{z}} = z(\tilde{\mathbf{t}}). \quad (8.17)$$

Jakobiany tych funkcji mają postać:

$$\mathbf{J}_{yx} = \frac{\partial \tilde{\mathbf{y}}}{\partial \tilde{\mathbf{x}}}, \quad \mathbf{J}_{zt} = \frac{\partial \tilde{\mathbf{z}}}{\partial \tilde{\mathbf{t}}}. \quad (8.18)$$

Utwórzmy wektory blokowe (klatkowe):

$$\tilde{\mathbf{r}} = \begin{bmatrix} \tilde{\mathbf{y}} \\ \tilde{\mathbf{z}} \end{bmatrix}, \quad \tilde{\mathbf{s}} = \begin{bmatrix} \tilde{\mathbf{x}} \\ \tilde{\mathbf{t}} \end{bmatrix}. \quad (8.19)$$

Dla tych wektorów jacobian \mathbf{J}_{rs} będzie macierzą blokową:

$$\mathbf{J}_{rs} = \begin{bmatrix} \mathbf{J}_{yx} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{zt} \end{bmatrix}.$$

W przypadku blokowych wektorów losowych prawo propagacji ma postać identyczną z równaniem (8.16), mamy więc:

$$\mathbf{Q}_{rr} = \mathbf{J}_{rs} \mathbf{Q}_{ss} \mathbf{J}_{rs}^T. \quad (8.20)$$

A w postaci bardziej "rozpakowanej" będzie:

$$\begin{aligned} \begin{bmatrix} \mathbf{Q}_{yy} & \mathbf{Q}_{yz} \\ \mathbf{Q}_{zy} & \mathbf{Q}_{zz} \end{bmatrix} &= \begin{bmatrix} \mathbf{J}_{yx} & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{zt} \end{bmatrix} \begin{bmatrix} \mathbf{Q}_{xx} & \mathbf{Q}_{xt} \\ \mathbf{Q}_{tx} & \mathbf{Q}_{tt} \end{bmatrix} \begin{bmatrix} \mathbf{J}_{yx}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{J}_{zt}^T \end{bmatrix} = \\ &= \begin{bmatrix} \mathbf{J}_{yx} \mathbf{Q}_{xx} \mathbf{J}_{yx}^T & \mathbf{J}_{yx} \mathbf{Q}_{xt} \mathbf{J}_{zt}^T \\ \mathbf{J}_{zt} \mathbf{Q}_{tx} \mathbf{J}_{yx}^T & \mathbf{J}_{zt} \mathbf{Q}_{tt} \mathbf{J}_{zt}^T \end{bmatrix}. \end{aligned} \quad (8.21)$$

Możemy teraz wypisać zależności na wszystkie cztery kofaktory:

$$\begin{aligned} \mathbf{Q}_{yy} &= \mathbf{J}_{yx} \mathbf{Q}_{xx} \mathbf{J}_{yx}^T, \\ \mathbf{Q}_{yz} &= \mathbf{J}_{yx} \mathbf{Q}_{xt} \mathbf{J}_{zt}^T, \\ \mathbf{Q}_{zy} &= \mathbf{J}_{zt} \mathbf{Q}_{tx} \mathbf{J}_{yx}^T = \mathbf{Q}_{yz}, \\ \mathbf{Q}_{zz} &= \mathbf{J}_{zt} \mathbf{Q}_{tt} \mathbf{J}_{zt}^T. \end{aligned} \quad (8.22)$$

Dodajmy jeszcze, że łatwa do skonstruowania macierz ($m \times n$):

$$\mathbf{Q}_{xt} = \begin{bmatrix} q_{x1 t1} & \cdots & q_{x1 tm} \\ \cdots & \cdots & \cdots \\ q_{xn t1} & \cdots & q_{xn tm} \end{bmatrix},$$

nazywana jest macierzą kros-kofaktorową. \uparrow

8.6 Dygresja - mnożenie symboliczne

W celu otrzymania elementów macierzy kowariancji (8.22) oddzielnie, bez obliczania całej macierzy (8.21), Tienstra zaproponował pewien praktyczny schemat zwany "mnożeniem symbolicznym". W celu jego ilustracji rozważmy następujące równania:

$$\begin{aligned} u &= a_0 + a_1x + a_2y + a_3z \\ v &= b_0 + a_1x + b_2y + b_3z \end{aligned} \quad (8.23)$$

lub w postaci macierzowej

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} a_0 & a_1 & a_2 & a_3 \\ b_0 & b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \quad (8.24)$$

Analogicznie do równania (8.20) w sposób już nam znany, kofaktory zmiennych u, v otrzymać można po wymnożeniu macierzy:

$$\begin{bmatrix} Q_{uu} & Q_{uv} \\ Q_{vu} & Q_{vv} \end{bmatrix} = \begin{bmatrix} J_{ux} & J_{uy} & J_{uz} \\ J_{vx} & J_{vy} & J_{vz} \end{bmatrix} \begin{bmatrix} Q_{xx} & Q_{xy} & Q_{xz} \\ Q_{yx} & Q_{yy} & Q_{yz} \\ Q_{zx} & Q_{zy} & Q_{zz} \end{bmatrix} \begin{bmatrix} J_{ux}^T & J_{vx}^T \\ J_{uy}^T & J_{vy}^T \\ J_{uz}^T & J_{vz}^T \end{bmatrix}$$

$$\begin{bmatrix} q_{uu} & q_{uv} \\ q_{vu} & q_{vv} \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} q_{xx} & q_{xy} & q_{xz} \\ q_{yx} & q_{yy} & q_{yz} \\ q_{zx} & q_{zy} & q_{zz} \end{bmatrix} \begin{bmatrix} a_1 & b_1 \\ a_2 & b_2 \\ a_3 & b_3 \end{bmatrix}$$

a po wymnożeniu macierzy uzyskamy:

$$\begin{aligned} q_{uu} &= a_1^2 q_{xx} + a_2^2 q_{yy} + a_3^2 q_{zz} + 2a_1 a_2 q_{xy} + 2a_1 a_3 q_{xz} + 2a_2 a_3 q_{yz} \\ q_{vv} &= b_1^2 q_{xx} + b_2^2 q_{yy} + b_3^2 q_{zz} + 2b_1 b_2 q_{xy} + 2b_1 b_3 q_{xz} + 2b_2 b_3 q_{yz} \\ q_{uv} &= a_1 b_1 q_{xx} + a_2 b_2 q_{yy} + a_3 b_3 q_{zz} + (a_1 b_2 + a_2 b_1) q_{xy} + \\ &\quad + (a_1 b_3 + a_3 b_1) q_{xz} + (a_2 b_3 + a_3 b_2) q_{yz} \end{aligned} \quad (8.25)$$

Współczynniki kofaktorowe po prawych stronach równania (8.25) można otrzymać niemal natychmiast inną drogą — wykonując “mnożenie” równań (8.23) przez siebie, zanedbując a_0, b_0 . Przykładowy iloczyn typu (uu) ma postać:

$$(uu) = a_1^2 (xx) + a_2^2 (yy) + a_3^2 (zz) + 2a_1 a_2 (xy) + 2a_1 a_3 (xz) + 2a_2 a_3 (yz) \quad (8.26)$$

Porównując to równanie z pierwszym z równań (8.25) widzimy, że takie “mnożenie” faktycznie daje współczynniki kofaktorowe. Pamiętając, że iloczyny zmiennych np. (yz) należy traktować jako indeksy macierzy kofaktorowej, można tym sposobem otrzymać dowolny auto- bądź kros- kofaktor.

Metoda ta była nieoceniona w czasach rachunków ręcznych, zwłaszcza gdy potrzebny był pojedynczy element z całej macierzy kofaktorowej. Ale i obecnie nie straciła na znaczeniu bowiem ten sam schemat daje się zastosować do równań macierzowych. Jednakże w tym przypadku należy mieć baczenie na sekwencje macierzy, wprowadzając gdzie trzeba transpozycje. Przykładowo dla układu macierzowego:

$$\begin{aligned} \mathbf{y} &= \mathbf{Ax} + \mathbf{a} \\ \mathbf{z} &= \mathbf{Bt} + \mathbf{b} \end{aligned} \quad (8.27)$$

macierz kroskofaktorową \mathbf{Q}_{yz} można otrzymać drogą symbolicznego “mnożenia” wektorów \mathbf{y}, \mathbf{z} zanedbując wyrazy \mathbf{a} i \mathbf{b} , które są nieistotne z punktu widzenia propagacji:

$$(\mathbf{yz}^T) = (\mathbf{Ax})(\mathbf{Bt})^T = \mathbf{A}(\mathbf{xt}^T)\mathbf{B}^T$$

Traktując iloczyn wektorów (\mathbf{xt}^T) jako wskaźniki macierzy kofaktorowej \mathbf{Q}_{xt} , odrzucając symbol transpozycji, możemy napisać:

$$\mathbf{Q}_{yz} = \mathbf{A}\mathbf{Q}_{xt}\mathbf{B}^T$$

Kładąc $\mathbf{A} = \mathbf{J}_{yx}$ oraz $\mathbf{B} = \mathbf{J}_{zt}$ równanie to przechodzi w postać analogiczną z (8.16). ↗

8.7 Propagacja złożona z zastosowaniem macierzy

Rozpatrzmy następujące związki:

$$\begin{aligned} \mathbf{y} &= \mathbf{Ax} + \mathbf{a} \\ \mathbf{z} &= \mathbf{By} + \mathbf{b} \\ \mathbf{r} &= \mathbf{Cz} + \mathbf{c} \end{aligned} \quad (8.28)$$

Gdy dany jest wektor \mathbf{x} oraz jego kofaktor \mathbf{Q}_{xx} , możemy obliczyć odpowiednie macierze kofaktorowe dla pozostałych wektorów losowych \mathbf{y}, \mathbf{z} i \mathbf{r} . Dokonuje się tego w dwojaki sposób: przez podstawienie albo metodą krokową.

Podstawienie oznacza, że kładziemy kolejno równanie do równania np.:

$$\begin{aligned} \mathbf{y} &= \mathbf{Ax} + \mathbf{a} \\ \mathbf{z} &= \mathbf{By} + \mathbf{b} = \mathbf{B}(\mathbf{Ax} + \mathbf{a}) + \mathbf{b} = (\mathbf{BA})\mathbf{x} + (\mathbf{Ba} + \mathbf{b}) \\ \mathbf{r} &= \mathbf{Cz} + \mathbf{c} = \mathbf{C}(\mathbf{By} + \mathbf{b}) + \mathbf{c} = (\mathbf{CB})\mathbf{y} + (\mathbf{Cb} + \mathbf{c}) = \\ &= (\mathbf{CB})(\mathbf{Ax} + \mathbf{a}) + (\mathbf{Cb} + \mathbf{c}) = (\mathbf{CBA})\mathbf{x} + (\mathbf{CBA}\mathbf{a} + \mathbf{Cb} + \mathbf{c}) \end{aligned}$$

W celu wyznaczenia $\mathbf{Q}_{yy}, \mathbf{Q}_{zz}, \mathbf{Q}_{yz}, \dots$ możemy stosować podejście takie jak to uczyniono w przypadku równań (8.17), Ale możemy też zastosować mnożenie symboliczne Tienstry. W pierwszym wypadku musimy położyć:

$$\tilde{\mathbf{v}} = \begin{bmatrix} \tilde{\mathbf{y}} \\ \tilde{\mathbf{z}} \\ \tilde{\mathbf{r}} \end{bmatrix} \quad \tilde{\mathbf{s}} = [\tilde{\mathbf{x}}] \quad (8.29)$$

A dalej zgodnie z prawem propagacji będzie:

$$\begin{aligned} \mathbf{Q}_{vv} &= \mathbf{J}_{vs} \mathbf{Q}_{ss} \mathbf{J}_{vs}^T = \\ &= \begin{bmatrix} \mathbf{Q}_{yy} & \mathbf{Q}_{yz} & \mathbf{Q}_{yr} \\ \mathbf{Q}_{zy} & \mathbf{Q}_{zz} & \mathbf{Q}_{zr} \\ \mathbf{Q}_{ry} & \mathbf{Q}_{rz} & \mathbf{Q}_{rr} \end{bmatrix} = \begin{bmatrix} \mathbf{J}_{yx} \\ \mathbf{J}_{zx} \\ \mathbf{J}_{rx} \end{bmatrix} [\mathbf{Q}_{xx}] [\mathbf{J}_{yx}^T \quad \mathbf{J}_{zx}^T \quad \mathbf{J}_{rx}^T] = \\ &= \begin{bmatrix} \mathbf{A} \\ \mathbf{BA} \\ \mathbf{CBA} \end{bmatrix} [\mathbf{Q}_{xx}] [\mathbf{A}^T \quad \mathbf{A}^T \mathbf{B}^T \quad \mathbf{A}^T \mathbf{B}^T \mathbf{C}^T] = \end{aligned}$$

Po wykonaniu mnożenia macierzy

$$\begin{bmatrix} \mathbf{Q}_{yy} & \mathbf{Q}_{yz} & \mathbf{Q}_{yr} \\ \mathbf{Q}_{zy} & \mathbf{Q}_{zz} & \mathbf{Q}_{zr} \\ \mathbf{Q}_{ry} & \mathbf{Q}_{rz} & \mathbf{Q}_{rr} \end{bmatrix} = \begin{bmatrix} \mathbf{AQ}_{xx}\mathbf{A}^T & \mathbf{AQ}_{xx}\mathbf{A}^T\mathbf{B}^T & \mathbf{AQ}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \\ \mathbf{BAQ}_{xx}\mathbf{A}^T & \mathbf{BAQ}_{xx}\mathbf{A}^T\mathbf{B}^T & \mathbf{BAQ}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \\ \mathbf{CBAQ}_{xx}\mathbf{A}^T & \mathbf{CBAQ}_{xx}\mathbf{A}^T\mathbf{B}^T & \mathbf{CBAQ}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \end{bmatrix}$$

A poszukiwane kofaktory dane są przez następujące związki:

$$\begin{aligned} \mathbf{Q}_{yy} &= \mathbf{AQ}_{xx}\mathbf{A}^T \\ \mathbf{Q}_{zz} &= \mathbf{BAQ}_{xx}\mathbf{A}^T\mathbf{B}^T \\ \mathbf{Q}_{zz} &= \mathbf{CBAQ}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \\ \mathbf{Q}_{yz} &= \mathbf{AQ}_{xx}\mathbf{A}^T\mathbf{B}^T \\ \mathbf{Q}_{yr} &= \mathbf{AQ}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \\ \mathbf{Q}_{zr} &= \mathbf{BAQ}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \end{aligned} \quad (8.30)$$

Wyprowadzimy te same związki krokowo stosując prawo propagacji krok po kroku i wykorzystując podejście Tienstry, tzn symbolicznie wymnożymy stronami równania (8.28), w rezultacie otrzymamy:

$$\begin{aligned}
 \mathbf{Q}_{yy} &= \mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T \\
 \mathbf{Q}_{zz} &= \mathbf{B}\mathbf{Q}_{yy}\mathbf{B}^T = \mathbf{B}\mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T \\
 \mathbf{Q}_{rr} &= \mathbf{C}\mathbf{Q}_{zz}\mathbf{C}^T = \mathbf{C}\mathbf{B}\mathbf{A}\mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T\mathbf{C}^T \\
 \mathbf{Q}_{yz} &= \mathbf{A}\mathbf{Q}_{xy}\mathbf{B}^T \\
 \mathbf{Q}_{yr} &= \mathbf{A}\mathbf{Q}_{xz}\mathbf{C}^T \\
 \mathbf{Q}_{zr} &= \mathbf{B}\mathbf{Q}_{yz}\mathbf{C}^T
 \end{aligned} \tag{8.31}$$

Widzimy, że ostatnie trzy macierze kros- kofaktorowe nie są identyczne z odpowiednimi równaniami z zestawu równań (8.30), choćby z powodu braku w (8.30) macierzy \mathbf{Q}_{xy} , \mathbf{Q}_{xz} , \mathbf{Q}_{yz} . Można jednak otrzymać właściwą postać równań jeśli w podejściu krokowym, równania (8.28) uzupełnimy o oczywiste tożsamości, mianowicie:

$$\begin{aligned}
 \mathbf{x} &= \mathbf{I}\mathbf{x} \\
 \mathbf{y} &= \mathbf{A}\mathbf{x} + \mathbf{a} \\
 \mathbf{x} &= \mathbf{I}\mathbf{x} \\
 \mathbf{z} &= \mathbf{B}\mathbf{y} + \mathbf{b} \\
 \mathbf{y} &= \mathbf{I}\mathbf{y} \\
 \mathbf{z} &= \mathbf{B}\mathbf{y} + \mathbf{b}
 \end{aligned}$$

Zastosowane do tych równań mnożenie Tienstry daje:

$$\begin{aligned}
 \mathbf{Q}_{xy} &= \mathbf{I}\mathbf{Q}_{xx}\mathbf{A}^T = \mathbf{Q}_{xx}\mathbf{A}^T \\
 \mathbf{Q}_{xz} &= \mathbf{I}\mathbf{Q}_{xy}\mathbf{B}^T = \mathbf{Q}_{xx}\mathbf{A}^T\mathbf{B}^T \\
 \mathbf{Q}_{yz} &= \mathbf{Q}_{yy}\mathbf{B}^T
 \end{aligned}$$

Po podstawieniu tych równań do ostatnich trzech równań z zestawu (8.31), dostaniemy dokładnie równania (8.30) a więc oba podejścia — kolejnego podstawienia oraz propagacji krokowej dają identyczne rezultaty. Trzeba jednak pamiętać, że podejście krokowe wymaga zachowania pewnej ostrożności. ↗

8.8 Propagacja błędów systematycznych i prawdziwych

Niech \tilde{x} , \tilde{y} będą zmiennymi, dla których znamy pewną zależność funkcyjną $r = g(x, y)$:

Niech Δx , Δy będą przyrostami, o których naturze powiemy później. Z bezpośredniej propagacji wynika, że:

$$\begin{aligned}
 r + \Delta r &= g(x + \delta x, y + \delta y) \\
 \Delta r &= g(x + \delta x, y + \delta y) - g(x, y)
 \end{aligned}$$

Jeżeli przyrosty były odpowiednio małe możemy funkcję rozwinąć w szereg i brać pod uwagę wyłącznie człony liniowe rozwinięcia, oznacza to że:

$$\Delta r = \frac{\partial g(x, y)}{\partial x} \delta x + \frac{\partial g(x, y)}{\partial y} \delta y \tag{8.32}$$

Jak dotąd nie czyniliśmy żadnych uwag natury probabilistycznej, równanie (8.32) traktowaliśmy jako opisujące pewien problem obliczeniowy. W praktyce przyrosty $\Delta x, \Delta y$ można traktować jako wartości następujących błędów:

- błędy - poprawki wynikające z błędów rachunkowych,
- tzw. błędy prawdziwe, o których zakładamy, że są znane,
- błędy systematyczne.

We wszystkich tych przypadkach o przyrostach $\Delta x, \Delta y$ zakłada się, że są znane i że nie mają charakteru błędów przypadkowych. Dlatego z punktu widzenia statystyki można omawiany przypadek traktować jako specjalną postać propagacji średnich. Równanie (8.32) można rozważać jako graniczny przypadek, w którym zmienne losowe \tilde{x}, \tilde{y} traktowane są jako “stałe” tzn. ich statystyczne własności zostały zredukowane do zera.

Takie prawo propagacji z teoretycznego punktu widzenia ma niewielką wagę. Jednak w praktyce jest użyteczne zwłaszcza w badaniu pewnych efektów wynikających zastosowania uproszczonych modeli matematycznych. Błędy systematyczne w zasadzie są efektami zastosowania uproszczonych matematycznych modeli. Dlatego formuła (8.32) może być stosowana w analizie propagacji resztkowych błędów systematycznych. †

Rozdział 9

Metoda najmniejszych kwadratów: wyrównywanie z warunkami

Streszczenie. Nie ma komu napisać I nie wiadomo czy będzie komu. **Słowa kluczowe:**

Warunek najmniejszych kwadratów, równania warunkowe, równania normalne, technika bezpośrednia, technika pośrednia, przypadek ogólny wyrównywania z warunkami, formy dwuliniowe, różniczkowanie form dwuliniowych, metoda czynników Lagrange. ^a ↗

^a[Modyfikowano AD 2008, kwiecień, 23]

9.1 Wstęp

W rozdziale 2.7 przedstawiono jedynie zarys koncepcji wyrównywania obserwacji metodą najmniejszych kwadratów (MNK). Obecnie, rozwiemy ten temat rozpoczynając od przedstawienia dwóch podstawowych technik: bezpośredniego i pośredniego wyrównywania obserwacji.

Nie są to jedyne techniki, jest ich więcej, a konstruowano je w celu uproszczenia rozwiązań rozmaitych problemów, dlatego techniki te wzajemnie różnią się, ale pewne elementy są w nich podobne. Np. zawsze dotyczą modelu określonego przez n_0 wielkości, które wyznaczamy za pomocą obserwacji. Funkcjonalna część modelu wyrażona jest w formie równania, albo częściej, układu tzw. równań warunkowych¹, wiążących ze sobą wszystkie wielkości obserwowane. Jeśli obserwator zamiast n_0 wykonał $n > n_0$ obserwacji, mamy do czynienia z nadmiarem $r = n - n_0$ (liczba stopni swobody), a w rezultacie z niejednoznacznym rozwiązaniem interesującego nas problemu. Bowiemy w efekcie nieuchronnych niepewności pomiarowych, każdy podzbiór o liczebności n_0 wzięty spośród n obserwacji opisuje model w inny sposób. W celu usunięcia niejednoznaczności, wyznaczenia “najlepszego” zestawu n_0 interesujących nas wielkości posługujemy się kryterium najmniejszych kwadratów. ↗

9.2 Techniki pośredniego i bezpośredniego wyrównywania obserwacji

O konkretnej technice wyrównawczej stanowią: postać równań warunkowych, oraz sposób usuwania zaistniałej (wobec niezerowego nadmiaru r) niejednoznaczności. Jako pierwszą poznamy technikę bezpośredniego wyrównywania obserwacji. ↗

9.2.1 Bezpośrednie wyrównywanie obserwacji

Technika ta ma następujące cechy:

- liczba równań warunkowych wynosi $r = n - n_0$,
- poza wielkościami obserwowanymi bezpośrednio, równania warunkowe nie zawierają żadnych niewiadomych parametrów,
- w każdym równaniu warunkowym może wystąpić od 1 do n wielkości obserwowanych.

Przykład 1. Postanowiono wyznaczyć kształt trójkąta płaskiego. W tym celu zmierzono trzy kąty:

$$\alpha_1 = 41^\circ 33', \quad \alpha_2 = 78^\circ 57', \quad \alpha_3 = 59^\circ 27'$$

Wyznacz wyrównane wartości tych kątów.

Łatwo ustalimy, że liczba pomiarów $n = 3$ a minimalna liczba pomiarów koniecznych do jednoznacznego rozwiązania problemu $n_0 = 2$. Bo przecież w celu określenia kształtu

¹ Pośród astronomów zamiast równania warunkowe, częściej napotkamy na określenie równania obserwacyjne.

trójkąta wystarczyło zmierzyć jedynie dwa kąty. Stąd liczba stopni swobody $r = 1$ a zatem, zgodnie z tym co powiedziano wyżej, musimy napisać jedno równanie warunkowe.

Postać równania warunkowego stanowi zmienny element w MNK, a odnajdywanie jej wymaga od nas swego rodzaju zdolności kreatywnych. W naszym przykładzie nie muszą to być wybitne zdolności, dlatego napiszemy bez żadnych ceregieli, że model funkcjonalny rozważanego zagadnienia określony jest jednoznacznie, jeśli estymatory obserwacji czynią zadość równaniu:

$$\hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3 = 180^\circ$$

A zgodnie z definicją estymatorów wielkości obserwowanych $\hat{l}_i = l_i + \nu_i$:

$$(\alpha_1 + \nu_1) + (\alpha_2 + \nu_2) + (\alpha_3 + \nu_3) = 180^\circ$$

gdzie ν_i są niewielkimi najczęściej poprawkami zwanymi też residuami.

Po podstawieniu wartości zmierzonych kątów

$$\nu_1 + \nu_2 + \nu_3 = 180^\circ - (41^\circ 33' + 78^\circ 57' + 59^\circ 27')$$

$$\nu_1 + \nu_2 + \nu_3 = 3'$$

Zatem równanie warunkowe naszego zagadnienia może mieć postać:

$$\nu_3 = 3' - \nu_1 - \nu_2 \quad (9.1)$$

Wyrównywanie obserwacji α_i sprowadza się do znalezienia takich wartości residuów ν_i , które jednocześnie, spełniają równanie warunkowe oraz czynią zadość warunkowi najmniejszych kwadratów. Warunek ten oznacza, że funkcja skalarna:

$$\Phi = \nu_1^2 + \nu_2^2 + \nu_3^2 = \nu_1^2 + \nu_2^2 - (3' - \nu_1 - \nu_2)^2$$

dla poszukiwanych wartości residuów osiąga minimum, co pociąga:

$$\frac{\partial \Phi}{\partial \nu_1} = 2\nu_1 - 2(3' - \nu_1 - \nu_2) = 0$$

$$\frac{\partial \Phi}{\partial \nu_2} = 2\nu_2 - 2(3' - \nu_1 - \nu_2) = 0$$

Po uproszczeniach, jako warunek najmniejszych kwadratów, otrzymujemy równania zwane tradycyjnie *równanami normalnymi*:

$$2\nu_1 + \nu_2 = 3'$$

$$\nu_1 + 2\nu_2 = 3'$$

W wyniku rozwiązania tego układu, oraz skorzystania z równania warunkowego (9.1), mamy $\nu_1 = \nu_2 = \nu_3 = 1'$.

Ostatecznie wyrównane wartości kątów badanego trójkąta wynoszą:

$$\hat{\alpha}_1 = \alpha_1 + \nu_1 = 41^\circ 34'$$

$$\hat{\alpha}_2 = \alpha_2 + \nu_2 = 78^\circ 58'$$

$$\hat{\alpha}_3 = \alpha_3 + \nu_3 = 59^\circ 28'$$

Sprawdzając czy suma wyrównanych kątów wynosi 180° , mamy możliwość częściowej kontroli poprawności wykonanych obliczeń. ↗

9.2.2 Pośrednie wyrównywanie obserwacji

W tej technice wyrównywania danych obserwacyjnych:

- liczba równań warunkowych jest równa liczbie obserwacji n ,
- każde równanie warunkowe zawiera co najwyżej jedną obserwację ze współczynnikiem równym 1,
- obok pojedynczej obserwacji równania warunkowe zawierają dodatkowo u niewiadomych parametrów.

Ze względu na nadmiar $r = n - n_0$ zachodzi potrzeba sformułowania dokładnie r równań warunkowych. Gdy piszemy n takich równań, to mamy ich o $n - r$ za dużo. Dlatego w tej technice w równaniach warunkowych występuje $u = n - r$ dodatkowych parametrów, które nie posiadają żadnej wartości a priori, i które należy wyznaczyć wraz z residuami.

Przykład 2. Czterokrotnie zmierzono w stopniach odległość kątową między dwoma obiektami na sferze niebieskiej:

$$\begin{aligned} l_1 &= 32.51, & l_2 &= 32.48 \\ l_3 &= 32.52, & l_4 &= 32.53 \end{aligned}$$

Wyznacz wyrównaną wartość odległości kątowej α .

Mamy zatem $n = 4$, $n_0 = 1$, nadmiar oraz $r = 4 - 1 = 3$. Zgodnie z wymogami tej techniki należy ułożyć 4 równania warunkowe zawierające $u = n - r = 1$ parametrów. Równania mają postać :

$$\begin{aligned} l_1 + \nu_1 &= \alpha \\ l_2 + \nu_2 &= \alpha \\ l_3 + \nu_3 &= \alpha \\ l_4 + \nu_4 &= \alpha \end{aligned} \tag{9.2}$$

Albo

$$\begin{aligned} \nu_1 &= \alpha - l_1 \\ \nu_2 &= \alpha - l_2 \\ \nu_3 &= \alpha - l_3 \\ \nu_4 &= \alpha - l_4 \end{aligned}$$

Ich kształt wynika z oczywistego spostrzeżenia — gdyby wszystkie pomiary odległości kątowej dały taki sam rezultat α mielibyśmy jednoznaczną odpowiedź na postawiony problem. Podobnie jak w przykładzie poprzednim, resydua ν_i muszą czynić zadość warunkowi najmniejszych kwadratów, czyli szukamy takich ich wartości by funkcja:

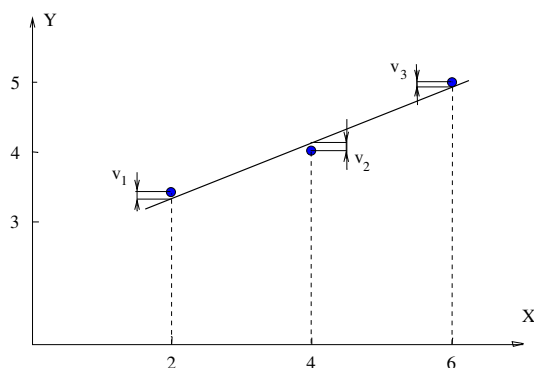
$$\Phi = \nu_1^2 + \nu_2^2 + \nu_3^2 + \nu_4^2 = (\alpha - 32.51)^2 + (\alpha - 32.48)^2 + 2(\alpha - 32.52)^2 + 2(\alpha - 32.53)^2$$

osiągała minimum, co pociąga warunek:

$$\frac{\partial \Phi}{\partial \alpha} = 2(\alpha - 32.51) + 2(\alpha - 32.48) + (\alpha - 32.52) + (\alpha - 32.53)$$

stad po rozwiązaniu wyrównana wartość odległości kątowej wynosi

$$\alpha = 32.51$$



Rysunek 9.1: Dopasowanie prostej do trzech punktów pomiarowych. Wartości współrzędnych X –sowych są dokładne, wartości współrzędnych Y –kowych wyznaczono za pomocą pomiarów.

Jak powiedziano technika pośrednia wyrównywania obserwacji pozwala na wyznaczenie pewnych parametrów tkwiących w modelu pomiarowym. Własność ta jest często wykorzystywana gdy interesują nas właśnie parametry a nie estymatory wielkości obserwowanej bezpośrednio.

Rozważmy kolejny przykład.

Przykład 3.

-	X [cm]	Y [cm]
P_1	2	3.2
P_2	4	4.0
P_3	6	5.0

Wyznaczono położenie trzech punktów, o których wiadomo, że leżą na jednej prostej. W tym celu zmierzono ich współrzędne prostokątne, patrz tabelka obok. Zakładając, że jedynie pomiary współrzędnych Y obciążone były niepewnościami (współrzędne X jakimś cudem określono w sposób dokładny), metodą najmniejszych kwadratów wyznaczyć współczynniki równania tej prostej.

Modelem funkcjonalnym tego problemu, jego równaniem warunkowym jest równanie prostej na płaszczyźnie spełnione przez estymatory, czyli wyrównane wielkości pomiarowe:

$$\hat{y} = ax + b$$

W rzeczywistości, w wyniku nieuchronnych niepewności pomiarowych punkty pomiarowe nie muszą należeć do wspólnej prostej. Oczekujemy jednak, że dla danych obserwacyjnych, spośród nieskończenie wielu linii prostych, istnieje pewna prosta najlepiej do nich dopasowana. Mianowicie taka, na której leżą punkty pomiarowe wyrównane metodą najmniejszych kwadratów (patrz rysunek ??). Parametry prostej można określić jednoznacznie za pomocą dwóch punktów, stąd w naszym zagadnieniu mamy $n_0 = 2$, a skoro liczba wykonanych pomiarów wynosi $n = 3$ to nadmiar $r = 1$. A ponieważ stosujemy technikę pośredniego wyrównywania obserwacji musimy napisać n równań warunkowych zawierających $u = n - r = 2$ parametry. Będą to równania:

$$\hat{y}_1 - ax_1 - b = 0$$

$$\hat{y}_2 - ax_2 - b = 0$$

$$\hat{y}_3 - ax_3 - b = 0$$

a po skorzystaniu z definicji estymatorów \hat{y}_i :

$$\begin{aligned}\nu_1 + y_1 - ax_1 - b &= 0 \\ \nu_2 + y_2 - ax_2 - b &= 0 \\ \nu_3 + y_3 - ax_3 - b &= 0\end{aligned}$$

Wprowadzając za y_i wartości liczbowe

$$\begin{aligned}\nu_1 = 2a + b - 3.2 &= 0 \\ \nu_2 = 4a + b - 4.0 &= 0 \\ \nu_3 = 6a + b - 5.0 &= 0\end{aligned}$$

Parametry a i b wyznaczymy z warunku najmniejszych kwadratów:

$$\Phi = \nu_1^2 + \nu_2^2 + \nu_3^2 = (2a + b - 3.2)^2 + (4a + b - 4.0)^2 + 2(6a + b - 5.0)^2$$

Minimalizacja tej formy kwadratowej pociąga warunki

$$\begin{aligned}\frac{\partial \Phi}{\partial a} &= 2(2a + b - 3.2) \cdot 2 + 2(4a + b - 4.0) \cdot 4 + 2(6a + b - 5.0) \cdot 6 = 0 \\ \frac{\partial \Phi}{\partial b} &= 2(2a + b - 3.2) + 2(4a + b - 4.0) + 2(6a + b - 5.0) = 0\end{aligned}$$

stąd

$$\begin{aligned}56a + 12b &= 52.4 \\ 12a + 3b &= 12.2\end{aligned}$$

i w rezultacie $a = 0.45$, $b = 2.27$ [cm]. ↗

9.3 Równania warunkowe: notacja macierzowa

Macierze pełnią w rachunku wyrównawczym poważną rolę, bowiem umożliwiają zwięźle sformułowanie bardzo złożonych zagadnień obserwacyjnych. Najpierw pokażemy w jaki sposób w formie macierzowej zapisywane są równania warunkowe dla obu poznanych technik wyrównawczych. ↗

9.3.1 Technika pośrednia

W przykładzie 2, w przypadku techniki pośredniego wyrównywania obserwacji, umacierzowanie równań warunkowych (9.2) może mieć następujący przebieg: najpierw przepisujemy je w postaci:

$$\begin{aligned}\nu_1 - \Delta &= -l_1 \\ \nu_2 - \Delta &= -l_2 \\ \nu_3 - \Delta &= -l_3 \\ \nu_4 - \Delta &= -l_4\end{aligned}\tag{9.3}$$

gdzie w przykładzie 2, symbol oznaczający kątem (parametr) α zastąpiliśmy symbolem Δ tradycyjnie oznaczającym jakiś dowolny parametr.

W drugim kroku utworzymy macierze:

$$\mathbf{v} = \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{bmatrix} \quad \mathbf{l} = \begin{bmatrix} l_1 \\ l_2 \\ l_3 \\ l_4 \end{bmatrix} \quad \mathbf{B} = \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix}\tag{9.4}$$

9.4 Macierzowy warunek najmniejszych kwadratów

Zanim pokażemy macierzową postać warunku najmniejszych kwadratów uczynimy drobną dygresję na temat macierzy wag wielkości pomiarowych. ▮

9.4.1 Koncepcja wag pomiarów.

Niepewności pomiarowe mają różne własności, jedną z nich jest tendencja przyjmowania wartości bliskich np. wartości średniej. Miarą tej tendencji jest odchylenie standardowe σ , jego wartość jest więc miarą precyzji metody pomiarowej. Jeśli σ jest małe, pomiary są silnie skupione wokół średniej, powiadamy, że precyzja pomiarów jest wysoka (nie mylić z dokładnością), i odwrotnie.

Istnieją inne miary precyzji, jedną z nich jest tzw. waga pomiaru. Potocznie można by powiedzieć, że z jej pomocą określa się dobroć danego pomiaru, stopień jego ważności, jego wagę w stosunku do innych. W zagadnieniach opracowywania danych pomiarowych, wagę obserwacji definiuje się w różny sposób, dla naszych celów określimy ją jako wielkość odwrotnie proporcjonalną do wariancji pomiaru

$$w = \frac{k}{\sigma^2} \quad (9.9)$$

gdzie w to waga pomiaru, k jest stałą proporcjonalności.

Pośród elementów próby pomiarowej wybiera się jeden z nich jako pomiar odniesienia. Jego wariancja $\sigma^2 = \sigma_0^2$ przyjmowana jest jako wariancja odniesienia, a jego wagę przyrównuje do jedności

$$w = \frac{k}{\sigma_0^2} = 1 \quad \implies k = \sigma_0^2$$

Wówczas wagi pozostałych pomiarów można określać jako

$$w = \frac{\sigma_0^2}{\sigma^2} \quad (9.10)$$

Obserwację, której waga wynosi jeden, nazywać będziemy obserwacją o *wadze jednostkowej*, jej wariancję, wariancją wagi jednostkowej.

Obok wariancji występuje w statystyce matematycznej pojęcie kowariancji. Pojawia się ono gdy wyznaczona wielkość jest zmienną losową dwu lub więcej wymiarową, np. pomiar położenia ciał niebieskich wymaga wyznaczenia dwóch współrzędnych. Kowariancja wyraża zależność (korelację) bądź niezależność między zmiennymi losowymi. Przyjmijmy na moment, że obserwacje są niezależne, nieskorelowane, co oznacza, że ich kowariancja jest równa zeru.

Dla m elementowej próby niezależnych pomiarów (x_1, x_2, \dots, x_m) o wariancjach $(\sigma_1^2, \sigma_2^2, \dots, \sigma_m^2)$, praktycznym pociągnięciem jest pojęcie macierzy wariancji

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & 0 & 0 \\ 0 & \sigma_2^2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \sigma_m^2 \end{bmatrix} \quad (9.11)$$

Analogicznie można określić macierz wag danej próby pomiarowej

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 & 0 & 0 \\ 0 & w_2 & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & w_m \end{bmatrix} \quad (9.12)$$

Obierając w próbie pomiarowej wariancję odniesienia i stosując równanie (9.10) możemy (9.12) napisać w postaci

$$\mathbf{W} = \sigma_0^2 \begin{bmatrix} \frac{1}{\sigma_1^2} & 0 & 0 & 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 & 0 \\ 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma_m^2} \end{bmatrix} \quad (9.13)$$

co pociąga

$$\mathbf{W} = \sigma_0^2 \mathbf{\Sigma}^{-1} \quad (9.14)$$

Mimo poczynionych założeń o braku korelacji, równanie (9.14) jest prawdziwe także dla obserwacji skorelowanych. ↗

9.4.2 Obserwacje zrównoważone

Niech l_i , $i = 1, 2, \dots, n$, oznaczają zbiór rezultatów obserwacji o niejednakowej precyzji o odchyleniach standardowych σ_i .

Dokonajmy następującego zabiegu: pomnóżmy wartość każdego pomiaru l_i przez pewną liczbową stałą. Otrzymujemy w ten sposób nowy ciąg obserwacji $l_i^* = l_i a_i$. Z prawa propagacji wariancji wynika, że pomiary l_i^* mają odchylenia standardowe $\sigma_i^* = \sigma_i a_i$. Zatem ich wagi wynoszą

$$w_i^* = \frac{k}{\sigma_i^{*2}} = \frac{k}{\sigma_i^2 a_i^2} = \frac{w_i}{a_i^2}$$

Skoro tak, to zbiór obserwacji $l_i \sqrt{w_i}$, jest zbiorem obserwacji jednakowo dokładnych (zrównoważonych), bowiem wszystkie obserwacje mają teraz jednakowe wagi, równe jedności. ↗

9.4.3 Macierzowa postać warunku najmniejszych kwadratów

Niech dany będzie ciąg rezultatów obserwacji l_i , $i = 1, 2, \dots, n$ o niejednakowej precyzji, określonej z pomocą wag w_i . Jeśli zamierzamy dokonać wyrównania tych obserwacji sposobami już poznanymi, w pewnym momencie musimy posłużyć się warunkiem najmniejszych kwadratów, który dla obserwacji nieskorelowanych, jednakowo dokładnych ma postać:

$$\Phi = \sum_{i=1}^n \nu_i^2 \quad (9.15)$$

Wiemy, że podstawienie $\tilde{l}_i = l_i \sqrt{w_i}$ pozwala na zrównoważenie obserwacji niejednakowo dokładnych. Analogiczne podstawienie dla residuów $\tilde{\nu}_i = \nu_i \sqrt{w_i}$ obserwacji nieskorelowanych i niejednakowo dokładnych, pozwoli na wykorzystanie warunku najmniejszych

kwadratów w postaci (9.15), mianowicie:

$$\Phi = \sum_{i=1}^n \tilde{v}_i^2 = \sum_{i=1}^n w_i v_i^2 \quad (9.16)$$

Za pomocą macierzy reszduów \mathbf{v} oraz macierzy wag \mathbf{W} , warunek ten możemy napisać jako

$$\Phi = \mathbf{v}^T \mathbf{W} \mathbf{v} \quad (9.17) \quad \uparrow$$

9.5 Równania normalne: notacja macierzowa

Dla obu omawianych technik pokażemy w jaki sposób w notacji macierzowej uzyskuje się MNK rozwiązanie postawionego zagadnienia. Zobaczmy, że dla danej techniki algorytm rozwiązujący jest zawsze taki sam, a więc można zaprogramować go w formie podprogramu obliczeniowego. W każdej bibliotece procedur numerycznych odnajdziemy gotowe takie podprogramy. \uparrow

9.5.1 Technika pośrednia

Rozpocznijmy od wyprowadzenia równań normalnych dla techniki pośredniej, w tym celu skorzystamy z równań warunkowych (9.6):

$$\begin{aligned} \mathbf{v} + \mathbf{B}\Delta &= \mathbf{d} - \mathbf{l} = \mathbf{f} \\ \mathbf{v} &= \mathbf{f} - \mathbf{B}\Delta \end{aligned}$$

i podstawimy wektor reszduów \mathbf{v} do warunku najmniejszych kwadratów (9.17):

$$\Phi = (\mathbf{f} - \mathbf{B}\Delta)^T \mathbf{W} (\mathbf{f} - \mathbf{B}\Delta)$$

Wykonując transpozycje w pierwszym nawiasie oraz odpowiednie wymnożenia, mamy

$$\Phi = \mathbf{f}^T \mathbf{W} \mathbf{f} - \Delta^T \mathbf{B}^T \mathbf{W} \mathbf{f} - \mathbf{f}^T \mathbf{W} \mathbf{B} \Delta + \Delta^T \mathbf{B}^T \mathbf{W} \mathbf{B} \Delta \quad (9.18)$$

Forma kwadrata Φ jest skalarem, składniki powyższej sumy również są skalarami a to oznacza, że drugi wyraz równy jest swej transpozycji:

$$\Delta^T \mathbf{B}^T \mathbf{W} \mathbf{f} = (\Delta^T \mathbf{B}^T \mathbf{W} \mathbf{f})^T = \mathbf{f}^T \mathbf{W} \mathbf{B} \Delta$$

Po podstawieniu tego rezultatu po prawej stronie (9.18) otrzymamy

$$\Phi = \mathbf{f}^T \mathbf{W} \mathbf{f} - 2\mathbf{f}^T \mathbf{W} \mathbf{B} \Delta + \Delta^T \mathbf{B}^T \mathbf{W} \mathbf{B} \Delta \quad (9.19)$$

W wyrażeniu tym elementy wszystkich macierzy i wektorów poza wektorem Δ zawierają jedynie stałe liczbowe. Stąd warunek na minimum Φ sprowadza się do przyrównania do zera pochodnej formy Φ po wektorze Δ . Korzystając z dodatku matematycznego ?? przekonamy się, że

$$\frac{\partial \Phi}{\partial \Delta} = -2\mathbf{f}^T \mathbf{W} \mathbf{B} + 2\Delta^T (\mathbf{B}^T \mathbf{W} \mathbf{B}) = 0$$

A po transpozycjach i uproszczeniach warunek na minimum formy Φ przybiera postać:

$$(\mathbf{B}^T \mathbf{W} \mathbf{B}) \Delta = \mathbf{B}^T \mathbf{W} \mathbf{f}$$

Wprowadzając podstawienia $\mathbf{N} = \mathbf{B}^T \mathbf{W} \mathbf{B}$ oraz $\mathbf{t} = \mathbf{B}^T \mathbf{W} \mathbf{f}$ uzyskamy:

$$\mathbf{N} \Delta = \mathbf{t} \quad (9.20)$$

Równania (9.20) nazywane są równaniami normalnymi, ich rozwiązanie otrzymane metodą odwracania macierzy \mathbf{N} daje natychmiast wartości poszukiwanych parametrów Δ :

$$\Delta = \mathbf{N}^{-1} \mathbf{t} \quad (9.21)$$

Przykład 4.

$l_1=32.51$	$w_1 = 1$
$l_2=32.48$	$w_2 = 2$
$l_3=32.52$	$w_3 = 1$
$l_4=32.53$	$w_4=0.5$

Rozważmy drobną modyfikację przykładu 2-giego, w którym wykonano $n = 4$ pomiary odległości kątowej, jednak tym razem nie były to pomiary jednakowo dokładne. Rezultaty pomiarów oraz ich wagi podane są w tabelce obok. Metodą najmniejszych kwadratów znaleźć wyrównane wartości \hat{l}_i : a) przy założeniu przypadku obserwacji o jednakowej precyzji, b) przy założeniu o niejednakowej precyzji pomiarów.

Z modelu zagadnienia wynika, że $n = 4$, $n_0 = 1$, $r = n - n_0 = 3$. Rozwiązując problem techniką pośrednią musimy napisać n równań obserwacyjnych zawierających $u = n - r = 1$ parametrów. Zgodnie z równaniami (9.3) oraz (9.5) mają one postać

$$\begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \end{bmatrix} + \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} \cdot [\Delta] = - \begin{bmatrix} -32.51 \\ -31.48 \\ -32.52 \\ -32.53 \end{bmatrix}$$

$$\mathbf{v} + \mathbf{B} \Delta = \mathbf{f}$$

Przypadek a) — obserwacji nieskorelowanych o jednakowej precyzji.

Mamy tu macierz wag $\mathbf{W} = \mathbf{I}$ identyczną z macierzą jednostkową, zatem macierze \mathbf{N} oraz \mathbf{t} w równaniach normalnych (9.20) policzymy jako:

$$\mathbf{N} = \mathbf{B}^T \mathbf{B} = [-1, -1, -1, -1] \begin{bmatrix} -1 \\ -1 \\ -1 \\ -1 \end{bmatrix} = [4]$$

$$\mathbf{t} = \mathbf{B}^T \mathbf{f} = [-1, -1, -1, -1] \begin{bmatrix} -32.51 \\ -32.48 \\ -32.52 \\ -32.53 \end{bmatrix} = [130.04]$$

W wyniku rozwiązania równań normalnych (9.20), Δ wynosi:

$$\Delta = \mathbf{N}^{-1} \mathbf{t} = (4^{-1}) \cdot (130.04) = 32.51^\circ$$

Przypadek b) — obserwacji nieskolerowanych o niejednakowej precyzji.

Tym razem macierz wag \mathbf{W} ma postać diagonalną

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix}$$

Z taką macierzą wag obliczamy macierze \mathbf{N} i \mathbf{t} :

$$\mathbf{N} = \mathbf{B}^T \mathbf{W} \mathbf{B} = \dots = [4.5]$$

$$\mathbf{t} = \mathbf{B}^T \mathbf{W} \mathbf{f} = \dots = [146.255]$$

A po rozwiązaniu macierzowego równania normalnego:

$$\Delta = \mathbf{N}^{-1} \mathbf{t} = (4.5^{-1}) \cdot (146.255) = 32.50^\circ$$

W porównaniu z wynikiem uzyskanym w przypadku a), jest to nieco inny rezultat. Wynik $\Delta = 32.50$ jest bliższy pomiarowi o najwyższej wadze, bowiem jego wpływ na wartość Δ jest największy z wszystkich pomiarów.

Na zakończenie zauważmy, że estymator $\hat{\alpha} = \Delta$ otrzymany MNK dla n nieskorelowanych pomiarów o różnych wagach, można było obliczyć za pomocą tzw. średniej ważonej:

$$\hat{\alpha} = \frac{\sum_{i=1}^n w_i l_i}{\sum_{i=1}^n w_i} = 32.50$$

Oznacza to, że metoda średniej ważonej jest szczegółowym rezultatem metody najmniejszych kwadratów. \uparrow

9.5.2 Technika bezpośrednia

W celu wyprowadzenia równań normalnych dla tej techniki, najpierw posłużymy się przykładem szczególnym. Dopiero później pokażemy jak uzyskać wyprowadzenie w przypadku ogólnym.

Niech będą dane równania warunkowe wydedukowane z modelu pewnego zagadnienia pomiarowego:

$$\begin{aligned} \nu_1 + \nu_2 - \nu_3 &= -(l_1 + l_2 - l_3) = f_1 \\ \nu_2 + \nu_4 - \nu_5 &= -(l_2 + l_4 - l_5) = f_2 \\ \nu_3 + \nu_5 + \nu_6 &= -(l_3 + l_5 + l_6) = f_3 \end{aligned} \quad (9.22)$$

a w postaci macierzowej:

$$\mathbf{A} \mathbf{v} = \mathbf{f} \quad (9.23)$$

W MNK residua \mathbf{v} winny spełniać to równanie ale także dać najmniejszą wartość wyrażenia:

$$\Phi = w_1 \nu_1^2 + w_2 \nu_2^2 + w_3 \nu_3^2 + w_4 \nu_4^2 + w_5 \nu_5^2 + w_6 \nu_6^2 \quad (9.24)$$

Do tego równania nie mamy możliwości podstawienia z osobna poszczególnych ν_i jak to miało miejsce w przypadku techniki pośredniej. By wpleść równania (9.22) do warunku najmniejszych kwadratów (9.24) zastosujemy inne podejście, zwane metodą czynników Lagrange.

W tym celu przepisujemy równania warunkowe (9.22) w postaci jednorodnej i dodatkowo, każde równanie odpowiednio, mnożymy przez tzw. czynnik Lagrangea $-2k_j$, $j = 1, 2, 3$:

$$\begin{aligned} -2k_1(\nu_1 + \nu_2 - \nu_3 - f_1) &= 0 \\ -2k_2(\nu_2 + \nu_4 - \nu_5 - f_2) &= 0 \\ -2k_3(\nu_3 + \nu_5 + \nu_6 - f_3) &= 0 \end{aligned} \quad (9.25)$$

Dodajmy iloczyny z lewych stron tych równań do warunku (9.24):

$$\begin{aligned}\Phi' = & w_1\nu_1^2 + w_2\nu_2^2 + w_3\nu_3^2 + w_4\nu_4^2 + w_5\nu_5^2 + w_6\nu_6^2 - 2k_1(\nu_1 + \nu_2 - \nu_3 - f_1) \\ & - 2k_2(\nu_2 + \nu_4 - \nu_5 - f_2) - 2k_3(\nu_3 + \nu_5 + \nu_6 - f_3)\end{aligned}\quad (9.26)$$

Ze względu na obecność zer w prawych stronach układu (9.25), warunek minimum nowej formy Φ' jest równoważny warunkowi na minimum formy Φ . Minimum osiągane przez Φ' dla residuów spełniających równania warunkowe ma miejsce w tym samym punkcie przestrzeni residuów co dla funkcji Φ .

Dzięki takiemu trikowi, zamiast formy Φ możemy minimalizować formę Φ' różniczkując ją po wszystkich niewiadomych. W ten sposób uzyskamy 6 + 3 równań z 6 + 3 niewiadomymi:

$$\begin{aligned}\frac{\partial\Phi'}{\partial\nu_1} &= 2w_1\nu_1 - 2k_1 = 0 \\ \frac{\partial\Phi'}{\partial\nu_2} &= 2w_2\nu_2 - 2(k_1 + k_2) = 0 \\ \frac{\partial\Phi'}{\partial\nu_3} &= 2w_3\nu_3 - 2(-k_1 + k_3) = 0 \\ \frac{\partial\Phi'}{\partial\nu_4} &= 2w_4\nu_4 - 2k_2 = 0 \\ \frac{\partial\Phi'}{\partial\nu_5} &= 2w_5\nu_5 - 2(-k_2 + k_3) = 0 \\ \frac{\partial\Phi'}{\partial\nu_6} &= 2w_6\nu_6 - 2k_3 = 0 \\ \frac{\partial\Phi'}{\partial k_1} &= -2(\nu_1 + \nu_2 - \nu_3 - f_1) = 0 \\ \frac{\partial\Phi'}{\partial k_2} &= -2(\nu_2 + \nu_4 - \nu_5 - f_2) = 0 \\ \frac{\partial\Phi'}{\partial k_3} &= -2(\nu_3 + \nu_5 + \nu_6 - f_3) = 0\end{aligned}\quad (9.27)$$

Jak widać, warunki na pochodne po k_i są identyczne z tymi tkwiącymi w równaniach warunkowych. W metodzie czynników Lagrange mamy więc do czynienia z rozszerzeniem układu równań na minimum wyjściowej formy Φ o wszystkie równania warunkowe.

Po uproszczeniach, układ (9.27) przechodzi w postać:

$$\begin{aligned}\nu_1 &= w_1^{-1}k_1 \\ \nu_2 &= w_2^{-1}(k_1 + k_2) \\ \nu_3 &= w_3^{-1}(-k_1 + k_3) \\ \nu_4 &= w_4^{-1}k_2 \\ \nu_5 &= w_5^{-1}(-k_2 - k_3) \\ \nu_6 &= w_6^{-1}k_3 \\ \nu_1 + \nu_2 - \nu_3 &= f_1 \\ \nu_2 + \nu_4 - \nu_5 &= f_2 \\ \nu_3 + \nu_5 + \nu_6 &= f_3\end{aligned}$$

Npiszmy pierwsze 6 równań w postaci macierzowej

$$\begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{bmatrix} = \begin{bmatrix} w_1^{-1} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_2^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & w_3^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & w_4^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & w_5^{-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_6^{-1} \end{bmatrix} \cdot \begin{bmatrix} k_1 \\ k_1 + k_2 \\ -k_1 + k_3 \\ k_2 \\ -k_2 - k_3 \\ k_3 \end{bmatrix}$$

Ostatnią macierz po prawej warto przedstawić w formie iloczynu dwóch macierzy, i wów-

czas

$$\begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \\ \nu_4 \\ \nu_5 \\ \nu_6 \end{bmatrix} = \begin{bmatrix} w_1^{-1} & 0 & 0 & 0 & 0 & 0 \\ 0 & w_2^{-1} & 0 & 0 & 0 & 0 \\ 0 & 0 & w_3^{-1} & 0 & 0 & 0 \\ 0 & 0 & 0 & w_4^{-1} & 0 & 0 \\ 0 & 0 & 0 & 0 & w_5^{-1} & 0 \\ 0 & 0 & 0 & 0 & 0 & w_6^{-1} \end{bmatrix} \cdot \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} k_1 \\ k_2 \\ k_3 \end{bmatrix} \quad (9.28)$$

Zinterpretujmy uzyskane wyniki:

- diagonalna macierz z elementami w_i^{-1} to odwrotność macierzy wag \mathbf{W} . W rachunku wyrównawczym nazywana jest macierzą kofaktorową $\mathbf{Q} = \mathbf{W}^{-1}$. Macierz ta zawiera względne wariancje pomiarów, a w tych przypadkach, w których wariancja odniesienia $\sigma_0 = 1$ mamy $\mathbf{Q} = \mathbf{\Sigma}$.
- wektor najbardziej na prawo, to wektor czynników Lagrange, oznaczamy go przez \mathbf{k} ,
- macierz na lewo od wektora czynników Lagrange jest równa \mathbf{A}^T , transponowanej macierzy współczynników równania warunkowego (9.23).

W nowych oznaczeniach równanie (9.28) przybiera postać:

$$\mathbf{v} = \mathbf{W}^{-1} \mathbf{A}^T \mathbf{k} = \mathbf{Q} \mathbf{A}^T \mathbf{k} \quad (9.29)$$

Wychodząc z równań warunkowych (9.22) dotyczących konkretnego problemu uzyskaliśmy ogólne macierzowe równanie (9.29) na residua wielkości obserwowanych. Poza wektorem residuów, nie znamy tu jedynie elementów wektora \mathbf{k} . Ale kładąc (9.29) do (9.23) dostaniemy macierzowe równanie, w którym jedyną niewiadomą będzie wektor \mathbf{k} :

$$\mathbf{A}(\mathbf{Q} \mathbf{A}^T \mathbf{k}) = (\mathbf{A} \mathbf{Q} \mathbf{A}^T) \mathbf{k} = \mathbf{f} \quad (9.30)$$

Kwadratową macierz współczynników ($\mathbf{A} \mathbf{Q} \mathbf{A}^T$) przyjęto oznaczać jako \mathbf{Q}_e , i w takich oznaczeniach zamiast (9.30) mamy:

$$(\mathbf{A} \mathbf{Q} \mathbf{A}^T) \mathbf{k} = \mathbf{Q}_e \mathbf{k} = \mathbf{f} \quad (9.31)$$

Dla techniki bezpośredniej macierzowe równanie (9.31) nazywane jest równaniem normalnym, jego rozwiązaniem jest wektor czynników Lagrange \mathbf{k} :

$$\mathbf{k} = \mathbf{Q}_e^{-1} \mathbf{f} = \mathbf{W}_e \mathbf{f} \quad (9.32)$$

gdzie podstawiono:

$$\mathbf{W}_e = \mathbf{Q}_e^{-1} = (\mathbf{A} \mathbf{Q} \mathbf{A}^T)^{-1} \quad (9.33)$$

Równanie normalne (9.31) mimo iż wyprowadzone za pomocą rozważań dotyczących konkretnego przykładu jest ogólnym równaniem normalnym techniki bezpośredniego wyrównywania obserwacji. Dla porządku przedstawimy jego wyprowadzenie bez odwoływania się do jakiegokolwiek przykładu.

Wyjściowym punktem wyprowadzenia jest macierzowy warunek najmniejszych kwadratów:

$$\Phi = \mathbf{v}^T \mathbf{W} \mathbf{v}$$

który zmodyfikujemy przez włączenie doń równań warunkowych (9.23) w postaci jednorodnej z czynnikami Lagrange:

$$\Phi' = \mathbf{v}^T \mathbf{W} \mathbf{v} - 2\mathbf{k}^t (\mathbf{A} \mathbf{v} - \mathbf{f}) \quad (9.34)$$

Pochodna formy Φ' po wektorze \mathbf{k} ma postać równań warunkowych (9.23):

$$\mathbf{A} \mathbf{v} - \mathbf{f} = \mathbf{0}$$

Pochodna po wektorze residuów \mathbf{v} wynosi:

$$\frac{\partial \Phi'}{\partial \mathbf{v}} = 2\mathbf{v}^T \mathbf{W} - 2\mathbf{k}^T \mathbf{A} = \mathbf{0}$$

Ponieważ macierz wag jest symetryczna $\mathbf{W} = \mathbf{W}^T$ mamy

$$\mathbf{W} \mathbf{v} = \mathbf{A}^T \mathbf{k}$$

skąd

$$\mathbf{v} = \mathbf{W}^{-1} \mathbf{A}^T \mathbf{k} = \mathbf{Q} \mathbf{A} \mathbf{k} \quad (9.35)$$

A po podstawieniu do równania warunkowego (9.23) łatwo dostaniemy równania normalne na wektor \mathbf{k} , postać identyczną z (9.31).

Przykład 5. Zmierzono wszystkie kąty trójkąta płaskiego. Oszacuj $\alpha_1 = 43^\circ 33' 45''$, $w_1 = 1$ wyrównane wartości kątów: a) przy założeniu, że pomiary $\alpha_2 = 78^\circ 57' 55''$, $w_2 = 0.67$ były nieskorelowane i jednakowo dokładne, b) przy założeniu, że $\alpha_3 = 59^\circ 27' 50''$, $w_3 = 0.50$ miały różne wagi, patrz tabela obok.

W przykładzie tym $n = 3$, $n_0 = 2$, $r = 1$. Stąd potrzebujemy jedynie jednego równania warunkowego:

$$(\alpha_1 + \nu_1) + (\alpha_2 + \nu_2) + (\alpha_3 + \nu_3) = 180^\circ$$

Po podstawieniu wartości pomiarowych:

$$\nu_1 + \nu_2 + \nu_3 = 30''$$

a w postaci macierzowej:

$$\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} = [30'']$$

$$\mathbf{A} \mathbf{v} = \mathbf{f}$$

Przypadek a), obserwacji nieskorelowanych o jednakowej precyzji.

Mamy tu, że $\mathbf{W} = \mathbf{Q} = \mathbf{I}$, a zatem z równania (9.31):

$$\mathbf{Q}_e = \mathbf{A} \mathbf{Q} \mathbf{A}^T = \mathbf{A} \mathbf{A}^T = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = [3]$$

Obliczając odwrotność macierzy \mathbf{Q}_e i podstawiając do równania (9.32) znajdujemy macierz czynników Lagrange:

$$\mathbf{k} = \mathbf{Q}_e^{-1} \mathbf{f} = [3^{-1}][30''] = [10'']$$

Znając elementy wektora \mathbf{k} łatwo już obliczyć macierz residuów, mianowicie z równania (9.29) mamy:

$$\mathbf{v} = \mathbf{Q}\mathbf{A}^T \mathbf{k} = \mathbf{A}^T \mathbf{k} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot [10''] = \begin{bmatrix} 10'' \\ 10'' \\ 10'' \end{bmatrix}$$

A stąd wyrównane wartości kątów wynoszą:

$$\hat{\alpha}_1 = \alpha_1 + \nu_1 = 41^\circ 33' 55''$$

$$\hat{\alpha}_2 = \alpha_2 + \nu_2 = 78^\circ 58' 05''$$

$$\hat{\alpha}_3 = \alpha_3 + \nu_3 = 59^\circ 28' 00''$$

Przypadek b), obserwacji nieskorelowanych o niejednakowej precyzji.

Tutaj macierz wag oraz macierz kofaktorów nie są macierzami jednostkowymi i wynoszą

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.67 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}, \quad \mathbf{W} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0.67 & 0 \\ 0 & 0 & 0.5 \end{bmatrix}, \quad \mathbf{Q} = \mathbf{W}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

Za pomocą (9.31) obliczymy \mathbf{Q}_e

$$\mathbf{Q}_e = \mathbf{A}\mathbf{Q}\mathbf{A}^T = [1 \ 1 \ 1] \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = [4.5]$$

Z równania (9.32)

$$\mathbf{k} = \mathbf{Q}_e^{-1} \mathbf{f} = [4.5^{-1}][30''] = [6, 7'']$$

Kładąc \mathbf{k} do równania (9.35) znajdujemy wartości residuów:

$$\mathbf{v} = \mathbf{Q}\mathbf{A}^T \mathbf{k} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1.5 & 0 \\ 0 & 0 & 2 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \cdot [6, 7''] = \begin{bmatrix} 7'' \\ 10'' \\ 13'' \end{bmatrix}$$

Wyrównane wartości kątów tym razem wynoszą:

$$\hat{\alpha}_1 = \alpha_1 + \nu_1 = 41^\circ 33' 52''$$

$$\hat{\alpha}_2 = \alpha_2 + \nu_2 = 78^\circ 58' 05''$$

$$\hat{\alpha}_3 = \alpha_3 + \nu_3 = 59^\circ 28' 03''$$

↗

9.6 Wyrównywanie z warunkami: przypadek ogólny

Metoda najmniejszych kwadratów (MNK), jest jedną z metod pozwalających na jednoznaczne wyznaczenie zbioru estymatorów danego zagadnienia pomiarowego. Możemy stosować ją wszędzie tam gdzie dysponujemy pewnym nadmiarem obserwacji $r = n - n_0$. Oprócz nadmiaru danych, musimy mieć do dyspozycji wagi poszczególnych pomiarów, natomiast nie istnieją (w przypadku metody najmniejszych kwadratów) żadne wymagania co do rozkładu prawdopodobieństwa wielkości pomiarowych. Algorytmy metody najmniejszych kwadratów, znalazły zastosowanie w bardzo wielu dziedzinach. Z rozwojem rachunku macierzowego oraz maszyn cyfrowych znaczenie tych algorytmów okrutnie wzrosło.

Dysponując modelem matematycznym danego problemu znajdujemy się w punkcie, od którego rozpoczyna się ustalenie konkretnej metodyki rachunku wyrównawczego. Przechodzimy do rozważań nad aspektami praktycznymi i rachunkowymi, przechodzimy do wyboru odpowiedniej techniki najmniejszych kwadratów. Obliczenia metodą najmniejszych kwadratów dostarczają nowych wartości estymatorów wszystkich zmiennych modelu a także ich macierze kowariancji. Ale to nie koniec, w następnym kroku należy metodami statystycznymi ocenić jakość uzyskanych rezultatów. Oceny te w zależności od stopnia wadliwości uzyskanych wyników mogą doprowadzić do przemodelowania wyrównywanego zagadnienia a nawet do jego całkowitej zmiany. Dla danego modelu, dla danego zbioru danych obserwacyjnych MNK daje rezultat jednoznaczny. Można jednak wykorzystać tu różne techniki najmniejszych kwadratów, ale niezależnie od zastosowanej techniki uzyskane rezultaty są zawsze takie same.

Poza danymi obserwacyjnymi model może zawierać inne dodatkowe zmienne lub stałe numeryczne. Rodzina tego typu zmiennych, które mają także charakter statystyczny określona jest mianem "parametrów", w celu odróżnienia ich od "obserwacji", dla których dysponujemy a priori wartościami próbnymi. Wartości parametrów nie są znane, a ich estymatory wyznaczone są w procesie wyrównawczym. Reprezentują one niewiadome typu funkcjonalnego, chociaż po wyrównaniu traktowane są jako zmienne losowe podlegające testom statystycznym podobnie do zmiennych reprezentujących obserwacje. Liczbę parametrów oznaczamy literą u , natomiast wektor Δ o rozmiarach $(u \times 1)$ reprezentuje je wszystkie razem.

Dana technika (algorytm) najmniejszych kwadratów działa w ramach pewnego zbioru matematycznych funkcji, równań. Opisują one funkcjonalny model wyrównywanego problemu. Rozróżniamy dwa typy równań: równania warunkowe i równania więzów. Dowolne równanie zawierające jedną lub więcej obserwacji nazywamy równaniem warunkowym. Każde zagadnienie wyrównawcze wymaga pewnej liczby równań warunkowych. Całkowitą ich liczbę oznaczamy przez c .

Z drugiej strony, równania nie zawierające żadnej obserwacji, nazywane są równaniami więzów lub krótko więzami. Z definicji równania te będą funkcjami jedynie parametrów i stałych.

Istnieją dwie zasadnicze grupy technik wyrównawczych, które można określić jako:

- technikę najmniejszych kwadratów z *warunkami*,
- technikę najmniejszych kwadratów z *warunkami i więzami*.

Taki podział technik wyrównawczych wynika z wprowadzenia rozdziału zmiennych na parametry i obserwacje. Możliwym jest jednak takie ujęcie, w którym wszystkie zmienne modelu traktowane są jako obserwacje, prowadzi to do łącznej techniki zwanej :

Tablica 9.1: Lista symboli oraz ich objaśnienia wykorzystywane w technikach najmniejszych kwadratów

Symbol	Objaśnienie
\mathbf{l}	wektor wartości pomiarowych badanych zmiennych losowych l ,
$\hat{\mathbf{l}}$	wektor estymatorów obserwacji (niewiadome),
\mathbf{v}	wektor residuów (niewiadome),
Δ	wektor parametrów (niewiadome),
n_0	minimalna liczba różnych zmiennych określających dany model,
n	liczba obserwacji (estymatorów),
r	nadmiar, statystyczna liczba stopni swobody,
u	liczba niewiadomych parametrów w Δ
c	liczba równań warunkowych,
s	liczba więzów w Δ),
\mathbf{Q}	macierz kofaktorowa wielkości obserwowanych l (inaczej \mathbf{Q}_{ll}),
\mathbf{W}	macierz wagowa obserwacji l (inaczej \mathbf{W}_{ll}).

- *zunifikowaną* techniką najmniejszych kwadratów.

Przypomnijmy raz jeszcze o konieczności rozdzielenia pojęć: modelu wyrównywanego i techniki wyrównywania. Wszystkie techniki dają taki sam rezultat. Natomiast jaką on będzie miał wartość, zależy silnie od wyboru modelu danego zagadnienia pomiarowego.

W tabeli 9.1 dla wygody podano listę symboli wykorzystywanych w różnych technikach najmniejszych kwadratów. [↑](#)

9.6.1 Techniki wyrównywania z warunkami

W tej grupie technik najmniejszych kwadratów parametry (o ile istnieją) są funkcjonalnie niezależne. Kiedy przystępujemy do danego zagadnienia wyrównawczego musimy najpierw zdefiniować jego model, a w szczególności jego część funkcjonalną. Umożliwi to ustalenie wartości n_0 , minimalnej liczby pomiarów określających model jednoznacznie. Dalej, dla danego zbioru n obserwacji trzeba sprawdzić czy w ramach ustalonego modelu rzeczywiście mamy nadmiar obserwacji $r = n - n_0 > 0$. Nadmiar można interpretować jako informację, że spośród n obserwacji istnieje r funkcji, warunków, którym trzeba zadośćuczynić. Dla $r = 0$, gdy $n = n_0$, wszystkie obserwacje spełniają model dokładnie. Gdy np. $r = 1$, należy napisać jedną funkcję wiążącą ze sobą wszystkie n obserwacji. Ale co będzie gdy w modelu, obok obserwacji, wystąpią dodatkowo niewiadome parametry? Wówczas, jeżeli przedtem musieliśmy napisać r równań warunkowych nie zawierających parametrów, to teraz gdy dodano do modelu jeden nieznaną parametr, trzeba będzie napisać $r + 1$ równań, potrzebne jest jedno równanie dla dodatkowej niewiadomej. Ale nie zmienia to liczby stopni swobody. Zatem jeżeli mamy u niewiadomych niezależnych parametrów, liczba koniecznych do spełnienia warunków c wynosi

$$c = r + u \tag{9.36}$$

Taką sytuację mamy w ogólnym przypadku omawianej techniki wyrównywania z warunkami. Przypadki szczególne wyróżniane są w zależności od liczby występujących w nich parametrów u . Najniższą wartością jest $u = 0$, najwyższą $u = n_0$, gdyż w tym wypadku

$c = n$. Przypadek $u > n_0$ jest niemożliwy, chyba, że parametry są w jakiś sposób od siebie zależne. Mamy więc dwie nierówności, które dla dowolnej techniki wyrównywania z warunkami muszą być jednocześnie spełnione

$$\begin{aligned} r &\leq c \leq n \\ 0 &\leq u \leq n \end{aligned} \quad (9.37)$$

Dysponując określoną dla danego problemu liczbą parametrów u musimy zestawić $c = r + u$ niezależnych równań warunkowych. Powinny to być równania liniowe lub ich formy zlinearyzowane o postaci ogólnej

$$\mathbf{A}(\mathbf{l} + \mathbf{v}) + \mathbf{B}\Delta = \mathbf{d} \quad (9.38)$$

gdzie \mathbf{A} , \mathbf{B} są macierzami współczynników danego problemu, \mathbf{d} jest kolumnowym wektorem stałych danego problemu wyrównawczego. Równanie (9.38) można nieco przekształcić co da mu postać przyjętą przez nas jako standardową, mianowicie

$$\mathbf{A}\mathbf{v} + \mathbf{B}\Delta = \mathbf{f} \quad (9.39)$$

gdzie

$$\mathbf{f} = \mathbf{d} - \mathbf{A}\mathbf{l} \quad (9.40)$$

Przy czym, w świetle nierówności (9.37) rozmiary poszczególnych macierzy wynoszą

$$\begin{aligned} \mathbf{A} &- c \times n & \text{rank}(\mathbf{A})=c \\ \mathbf{v} &- n \times 1 \\ \mathbf{l} &- n \times 1 \\ \mathbf{B} &- c \times u & \text{rank}(\mathbf{B})=c \\ \Delta &- u \times 1 \\ \mathbf{d} &- c \times 1 \\ \mathbf{f} &- c \times 1 \end{aligned}$$

↑

9.6.2 Równania normalne techniki wyrównywania z warunkami

Równanie (9.39) jest podstawową formą równań warunkowych dla łącznego wyrównywania obserwacji i parametrów. Reprezentuje ono c równań liniowych z $n + u$ niewiadomymi (wektory \mathbf{v} i Δ). Ponieważ liczba równań c jest mniejsza od $n + u$ istnieje wiele rozwiązań równania (9.39). Jednoznaczne rozwiązanie otrzymamy dołączając do warunku (9.39) dodatkowy warunek — kryterium najmniejszych kwadratów

$$\Phi = \mathbf{v}^T \mathbf{W} \mathbf{v} \rightarrow \textit{minimum}$$

W celu wymuszenia tego kryterium i jednocześnie otrzymania rozwiązania równania (9.39) stosuje się metodę ograniczonego minimum z mnożnikami Lagrange'a. Jeżeli \mathbf{k} ($c \times 1$) będzie wektorem niewiadomych mnożników Lagrange'a, to w celu uzyskania spłotu warunku najmniejszych kwadratów z równaniem warunkowym, musimy szukać minimum funkcji Φ'

$$\Phi' = \mathbf{v}^T \mathbf{W} \mathbf{v} - 2\mathbf{k}(\mathbf{A}\mathbf{v} + \mathbf{B}\Delta - \mathbf{f}) \quad (9.41)$$

Zauważmy, że jeżeli równanie (9.39) jest spełnione, wyrażenie w nawiasie znika.

Minimalizacja Φ' oznacza odnajdywanie miejsc zerowych pochodnych po niewiadomych \mathbf{v} oraz Δ

$$\begin{aligned}\frac{\partial \Phi'}{\partial \mathbf{v}} &= 2\mathbf{v}^T \mathbf{W} - 2\mathbf{k}^T \mathbf{A} = \mathbf{0}^T \\ \frac{\partial \Phi'}{\partial \Delta} &= -2\mathbf{k}^T \mathbf{B} = \mathbf{0}^T\end{aligned}$$

a po drobnych operacjach, będzie

$$\begin{aligned}-\mathbf{W}\mathbf{v} + \mathbf{A}^T \mathbf{k} &= \mathbf{0}^T \\ \mathbf{B}^T \mathbf{k} &= \mathbf{0}^T\end{aligned}\tag{9.42}$$

Do dyspozycji mamy zatem równania (9.39) i (9.42), w których łącznie występuje $(n + u + c)$ niewiadomych. Ale w (9.39) mamy c równań, w macierzowym układzie (9.42) mamy $n + u$ równań, co wystarcza do jednoznacznego rozwiązania problemu.

Macierzowe równania (9.39) i (9.42) można połączyć formę łączną

$$\begin{bmatrix} -\mathbf{W} & \mathbf{A}^T & \mathbf{0} \\ \mathbf{A} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{k} \\ \Delta \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{f} \\ \mathbf{0} \end{bmatrix}\tag{9.43}$$

gdzie \mathbf{W} jest $(n \times n)$, \mathbf{k} jest $(c \times 1)$.

Układ (9.43) określamy bywa mianem *całkowitego* układu równań normalnych. Jego macierz współczynników jest macierzą symetryczną rzędu $(n + c + u)$, jest ona zawsze nieosobliwa (jej rozmiar równy jest jej rzędowi) o ile tylko model badanego zagadnienia został określony poprawnie.

A zatem ogólny problem wyrównywania z warunkami rozwiązany zostanie przez odwrócenie całkowitego układu równań, tzn.

$$\begin{bmatrix} \mathbf{v} \\ \mathbf{k} \\ \Delta \end{bmatrix} = \begin{bmatrix} -\mathbf{W} & \mathbf{A}^T & \mathbf{0} \\ \mathbf{A} & \mathbf{0} & \mathbf{B} \\ \mathbf{0} & \mathbf{B}^T & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} \\ \mathbf{f} \\ \mathbf{0} \end{bmatrix}\tag{9.44}$$

W przypadku niedużych zagadnień, w których liczba równań normalnych nie jest zbyt wielka, równanie (9.44) może rzeczywiście stanowić drogę do bezpośredniego rozwiązania. Jednak w wielu problemach praktycznych mamy do czynienia z bardzo dużymi układami, a wówczas bezpośrednie rozwiązanie (9.44) nie jest dogodne czy też opłacalne.

Najczęściej interesują nas wektory \mathbf{v} lub Δ , rzadko natomiast poszukujemy wektora \mathbf{k} . Dlatego pożądane są jakieś alternatywne schematy rozwiązań, a gdy zauważmy, że układ (9.43) ma sporo podmacierzy zerowych, oznacza to, że jego rozwiązanie z podziałem na podmacierze będzie wyjątkowo proste.

Z pierwszego z równań (9.42) mamy

$$\mathbf{v} = \mathbf{W}^{-1} \mathbf{A}^T \mathbf{k} = \mathbf{Q} \mathbf{A}^T \mathbf{k}\tag{9.45}$$

Po podstawieniu tej równości do równania warunkowego (9.39) otrzymamy

$$\mathbf{A} \mathbf{Q} \mathbf{A}^T \mathbf{k} + \mathbf{B} \Delta = \mathbf{f}\tag{9.46}$$

Położmy

$$\mathbf{l}_e = \mathbf{A} \mathbf{l}\tag{9.47}$$

\mathbf{l}_e będziemy nazywali równoważnym zbiorem obserwacji. Stosując do (9.47) prawo propagacji, otrzymamy macierz kofaktorową \mathbf{Q}_e wektora \mathbf{l}_e

$$\mathbf{Q}_e = \mathbf{AQA}^T \quad (9.48)$$

którą wprowadzimy do równania (9.46),

$$\mathbf{Q}_e \mathbf{k} + \mathbf{B}\Delta = \mathbf{f} \quad (9.49)$$

a jego rozwiązanie ze względu na \mathbf{k} ma postać

$$\mathbf{k} = \mathbf{Q}_e^{-1}(-\mathbf{B}\Delta + \mathbf{f}) = \mathbf{W}_e(-\mathbf{B}\Delta + \mathbf{f}) \quad (9.50)$$

gdzie wprowadziliśmy macierz $\mathbf{W}_e = \mathbf{Q}_e^{-1} = (\mathbf{AQA}^T)^{-1}$.

Podstawiając (9.50) do drugiego z równań (9.42), po drobnej redukcji będzie

$$(\mathbf{B}^T \mathbf{W}_e \mathbf{B}) \Delta = (\mathbf{B}^T \mathbf{W}_e \mathbf{f}) \quad (9.51)$$

lub korzystając z definicji macierzy \mathbf{W}_e

$$(\mathbf{B}^T (\mathbf{AQA}^T)^{-1} \mathbf{B}) \Delta = \mathbf{B}^T (\mathbf{AQA}^T)^{-1} \mathbf{f} \quad (9.52)$$

Równania macierzowe (9.51), (9.52) reprezentują układ u równań z u niewiadomymi parametrami (elementami wektora Δ). Układy (9.51), (9.52) nazywane są *częściowo zredukowanymi* równaniami normalnymi.

Po podstawieniach

$$\begin{aligned} \mathbf{N} &= \mathbf{B}^T \mathbf{W}_e \mathbf{B} = \mathbf{B}^T (\mathbf{AQA}^T)^{-1} \mathbf{B} \\ \mathbf{t} &= \mathbf{B}^T \mathbf{W}_e \mathbf{f} = \mathbf{B}^T (\mathbf{AQA}^T)^{-1} \mathbf{f} \end{aligned} \quad (9.53)$$

kompaktowa forma równań (9.51), (9.52) ma postać

$$\mathbf{N}\Delta = \mathbf{t} \quad (9.54)$$

a jego rozwiązaniem będzie

$$\Delta = \mathbf{N}^{-1} \mathbf{t} \quad (9.55)$$

W powyższym wyprowadzeniu pojawiło się kilka odwrotności macierzy co wymaga pewnego komentarza:

- macierz \mathbf{W}^{-1} w równaniu (9.45) jest zawsze możliwa do policzenia, bowiem macierz wagowa \mathbf{W} jest nieosobliwa jeżeli obserwacje są niezależne funkcjonalnie,
- macierz \mathbf{Q}_e^{-1} występująca w równaniu (9.50) jest zawsze dostępna ponieważ $\text{rank}(\mathbf{Q}_e) = \text{rank}(\mathbf{A}) = c$,
- wreszcie macierz \mathbf{N} jest także możliwa do policzenia ponieważ \mathbf{N} ma rząd i rozmiar identyczny z wartością u .

Gdy wyznaczymy wartości parametrów Δ możemy z równania (9.50) policzyć elementy wektora \mathbf{k} i w dalszej kolejności za pomocą równania (9.45) obliczyć elementy wektora residuów \mathbf{v} .

Z kolei, wektor residuów możemy wykorzystać do obliczenia nieobciążonego estymatora wariancji odniesienia

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T \mathbf{W} \mathbf{v}}{r} \quad (9.56)$$

gdzie r oznacza liczbę stopni swobody, $r = c - u$.

W następnym kroku residua umożliwiają obliczenie estymatorów $\hat{\mathbf{I}}$ wielkości obserwowanych

$$\hat{\mathbf{I}} = \mathbf{1} + \mathbf{v} \quad (9.57)$$

9.6.2.1 Dygresja: oszacowanie wariancji odniesienia

W celu oszacowania estymatora a posteriori wariancji odniesienia $\hat{\sigma}_0^2$ musimy znać rezultat iloczynu macierzy $\mathbf{v}^T \mathbf{W} \mathbf{v}$. Poniżej podany efektywny sposób na znalezienie wartości tej formy kwadratowej. Z równania (9.45) mamy

$$\mathbf{v} = \mathbf{Q} \mathbf{A}^T \mathbf{k}$$

a za pomocą równania (9.50)

$$\mathbf{v} = \mathbf{Q} \mathbf{A}^T \mathbf{W}_e (-\mathbf{B} \Delta + \mathbf{f})$$

A zatem po podstawieniu tego wyniku do formy $\mathbf{v}^T \mathbf{W} \mathbf{v}$, po wykonaniu odpowiednich przekształceń

$$\mathbf{v}^T \mathbf{W} \mathbf{v} = \dots = \mathbf{f}^T \mathbf{W}_e \mathbf{f} - \Delta^T \mathbf{f} \quad (9.58)$$

Jeżeli wyrównujemy iteracyjnie problem nieliniowy, wówczas skoro i -ta poprawka do wektora Δ jest bliska zeru, ostatni wyraz w (9.58) pomijamy. A wtedy

$$\mathbf{v}^T \mathbf{W} \mathbf{v} = \mathbf{f}^T \mathbf{W}_e \mathbf{f} = \mathbf{f}^T (\mathbf{A} \mathbf{Q} \mathbf{A}^T)^{-1} \mathbf{f} \quad (9.59)$$

†

9.6.3 Oszacowanie precyzji rezultatów wyrównania

Drugim istotnym etapem procesu wyrównywania obserwacji jest obliczenie precyzji wielkości wyznaczonych podczas wyrównania. Precyzję określa się tu w formie macierzy kofaktorowych lub, co oczywiście jest równoważne informacji zawartej w macierzy kowariancji. Macierze te dają się otrzymać wprost z rozwiązania najmniejszych kwadratów, do których należy zastosować prawa propagacji omówione na jednym z poprzednich wykładów.

Rozwiązanie najmniejszych kwadratów można sprowadzić do wykonania wymienionych niżej operacji macierzowych :

$$\begin{aligned} \mathbf{l} &= \mathbf{I} \mathbf{l} \\ \mathbf{f} &= \mathbf{d} - \mathbf{A} \mathbf{l} \\ \Delta &= \mathbf{N}^{-1} \mathbf{B}^T \mathbf{W}_e \mathbf{f} = \mathbf{N}^{-1} \mathbf{B}^T \mathbf{W}_e (\mathbf{d} - \mathbf{A} \mathbf{l}) \\ \mathbf{k} &= \mathbf{W}_e (-\mathbf{B} \Delta + \mathbf{f}) = \mathbf{W}_e (-\mathbf{B} \mathbf{N}^{-1} \mathbf{B}^T \mathbf{W}_e + \mathbf{I}) (\mathbf{d} - \mathbf{A} \mathbf{l}) \\ \mathbf{v} &= \mathbf{Q} \mathbf{A}^T \mathbf{k} = \mathbf{Q} \mathbf{A}^T \mathbf{W}_e (-\mathbf{B} \mathbf{N}^{-1} \mathbf{B}^T \mathbf{W}_e + \mathbf{I}) (\mathbf{d} - \mathbf{A} \mathbf{l}) \\ \hat{\mathbf{I}} &= \mathbf{1} + \mathbf{v} \end{aligned} \quad (9.60)$$

Wówczas odpowiadające im macierze autokofaktorowe mają postać

$$\begin{aligned}
\mathbf{Q}_{ll} &= \mathbf{Q} \\
\mathbf{Q}_{ff} &= (-\mathbf{A})\mathbf{Q}(-\mathbf{A}^T) = \mathbf{Q}_e \\
\mathbf{Q}_{\Delta\Delta} &= (-\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}_e\mathbf{A})\mathbf{Q}(-\mathbf{A}^T\mathbf{W}_e\mathbf{B}\mathbf{N}^{-1}) = \mathbf{N}^{-1} \\
\mathbf{Q}_{kk} &= \dots = \mathbf{W}_e - \mathbf{W}_e\mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}_e \\
\mathbf{Q}_{vv} &= \dots = \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A}\mathbf{Q} - \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{B}\mathbf{Q}_{\Delta\Delta}\mathbf{B}^T\mathbf{W}_e\mathbf{A}\mathbf{Q} \\
\mathbf{Q}_{\hat{ll}} &= \mathbf{Q}_{ll} + \mathbf{Q}_{lv} + \mathbf{Q}_{vl} + \mathbf{Q}_{vv} = \mathbf{Q} - \mathbf{Q}_{vv}
\end{aligned} \tag{9.61}$$

a macierze kroskofaktorowe :

$$\begin{aligned}
\mathbf{Q}_{lf} &= -\mathbf{Q}\mathbf{A}^T \\
\mathbf{Q}_{l\Delta} &= -\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{B}\mathbf{N}^{-1} \\
\mathbf{Q}_{lk} &= -\mathbf{Q}\mathbf{A}^T(\mathbf{W}_e - \mathbf{W}_e\mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}_e) \\
\mathbf{Q}_{lv} &= \dots = -\mathbf{Q}_{vv} \\
\mathbf{Q}_{\hat{ll}} &= \mathbf{Q}_{ll} + \mathbf{Q}_{lv} = \mathbf{Q} - \mathbf{Q}_{vv} = \mathbf{Q}_{\hat{ll}} \\
\mathbf{Q}_{fv\Delta} &= \dots = \mathbf{B}\mathbf{N}^{-1} \\
\mathbf{Q}_{fk} &= \dots = (\mathbf{I} - \mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}_e) \\
\mathbf{Q}_{fv} &= \dots = (\mathbf{I} - \mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}_e)\mathbf{A}\mathbf{Q} \\
\mathbf{Q}_{f\hat{l}} &= \mathbf{Q}_{fl} + \mathbf{Q}_{fv} = \dots = \mathbf{B}\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}_e\mathbf{A}\mathbf{Q} \\
\mathbf{Q}_{\Delta k} &= \dots = \mathbf{0} \\
\mathbf{Q}_{\Delta v} &= \mathbf{Q}_{\Delta k}(\mathbf{Q}\mathbf{A}^T) = \mathbf{0} \\
\mathbf{Q}_{\Delta\hat{l}} &= \mathbf{Q}_{\Delta l} + \mathbf{Q}_{\Delta v} = -\mathbf{N}^{-1}\mathbf{B}^T\mathbf{W}_e\mathbf{A}\mathbf{Q} \\
\mathbf{Q}_{kv} &= \mathbf{Q}_{kk}(\mathbf{Q}\mathbf{A}^T)^T = \dots = \mathbf{Q}_{kl} \\
\mathbf{Q}_{k\hat{l}} &= \mathbf{Q}_{kl} + \mathbf{Q}_{kv} = \mathbf{Q}_{kl} - \mathbf{Q}_{kl} = \mathbf{0} \\
\mathbf{Q}_{v\hat{l}} &= \mathbf{Q}_{vl} + \mathbf{Q}_{vv} = -\mathbf{Q}_{vv} + \mathbf{Q}_{vv} = \mathbf{0}
\end{aligned} \tag{9.62}$$

Macierz kofaktorowa $\mathbf{Q}_{\Delta\Delta}$ dana trzecim z równań (9.61) ma szczególne znaczenie w MNK. Widzimy, że recyzja niewiadomych parametrów jest natychmiast określona jeśli układ równań normalnych (9.54) rozwiązywano metodą odwracania macierzy. Uwaga! Jeżeli wektor \mathbf{x}^0 jest przybliżonym wektorem wartości np. parametrów, to z punktu widzenia propagacji błędów jest wielkością stałą. Dlatego dla nieliniowych układów równań, dla których $\hat{\mathbf{x}} = \mathbf{x}^0 + \Delta$ mamy

$$\mathbf{Q}_{\hat{x}\hat{x}} = \mathbf{Q}_{\Delta\Delta} = \mathbf{N}^{-1} \tag{9.63}$$

W niektórych wypadkach procedury redukowania rozwiązania równań (9.43) warto urwać na kroku (9.46), wówczas pozostałe układy można napisać w postaci

$$\begin{bmatrix} \mathbf{Q}_e & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{k} \\ \Delta \end{bmatrix} = \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix} \tag{9.64}$$

z której wynika

$$\begin{bmatrix} \mathbf{k} \\ \Delta \end{bmatrix} = \begin{bmatrix} \mathbf{Q}_e & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{f} \\ \mathbf{0} \end{bmatrix} \tag{9.65}$$

Poszukiwany wektor Δ , dostaniemy biorąc z (9.65) odpowiedni podwektor.

Pozostanie jednak do rozwiązania problem obliczenia $\mathbf{Q}_{\Delta\Delta}$. W wyniku odwrócenia macierzy blokowej z równania (9.65) uzyskamy

$$\begin{bmatrix} \mathbf{Q}_e & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{W}_e - \mathbf{W}_e \mathbf{B} \mathbf{N}^{-1} \mathbf{B}^T \mathbf{W}_e & \mathbf{W}_e \mathbf{B} \mathbf{N}^{-1} \\ \mathbf{N}^{-1} \mathbf{B} \mathbf{W}_e^T & -\mathbf{N}^{-1} \end{bmatrix} \quad (9.66)$$

Widzimy, że \mathbf{Q}_{kk} dane czwartym z równań (9.61) jest identyczne z pierwszym elementem macierzy po prawej stronie (9.66). Jednak pozostałe podmacierze $\mathbf{Q}_{k\Delta}$, $\mathbf{Q}_{\Delta\Delta}$ nie są równe prawym stronom odpowiednich równań (9.61), bowiem mamy tam $\mathbf{Q}_{k\Delta} = \mathbf{0}$ oraz $\mathbf{Q}_{\Delta\Delta} = \mathbf{N}^{-1}$ a nie $-\mathbf{N}^{-1}$. Jeśli zatem “wyciągamy” macierz kofaktorową \mathbf{Q}_{Δ} z równania (9.66) musimy pamiętać o zmianie znaku. ↗

9.6.4 Zależność funkcjonalna i posteriori macierze kofaktorowe

Wspominaliśmy już, że dane a priori parametry i obserwacje są funkcjonalnie niezależne. Niezależność funkcjonalna obserwacji oznacza, że żadna z obserwacji nie może zostać wydedukowana czy obliczona z pozostałych $(n - 1)$ obserwacji. Zwróciliśmy też uwagę na różnicę między pojęciami zależności funkcjonalnej i zależności losowej. W przypadku korelacji między zmiennymi losowymi pokazano, że mimo iż mamy n obserwacji niezależnych funkcjonalnie, mogą one (lub nie) być ze sobą skorelowane.

Różnica między tymi koncepcjami wiąże się z macierzami kofaktorowymi lub macierzami kowariancji danych wektorów losowych. Gdy zmienne losowe są *funkcjonalnie niezależne*, wówczas rząd macierzy kowariancji lub macierzy kofaktorowej równy jest rozmiarowi tych macierzy, co jednocześnie informuje o nieosobliwości tych macierzy. Gdy zmienne te są *stochastycznie niezależne lub nieskorelowane*, wszystkie elementy pozadiagonalne macierzy kowariancji lub kofaktorowej muszą równać się zeru. To właśnie stanowi podstawową różnicę między tymi dwiema koncepcjami niezależności. (Na marginesie, ścisła stochastyczna korelacja równoważna jest liniowej funkcjonalnej zależności).

Elementy wektorów \mathbf{l} oraz Δ są a priori funkcjonalnie niezależne, zatem \mathbf{Q} oraz $\mathbf{Q}_{\Delta\Delta}$ są nieosobliwe, pierwsza ma rząd równy n , druga ma rząd u .

Inne wektory obliczone w procesie wyrównania zawierają residua \mathbf{v} lub estymatory obserwacji $\hat{\mathbf{I}}$, odpowiadają im macierze kofaktorowe \mathbf{Q}_{vv} i $\mathbf{Q}_{\hat{\mathbf{I}}\hat{\mathbf{I}}}$. Elementy wektorów \mathbf{v} oraz $\hat{\mathbf{I}}$ są funkcjonalnie zależne, a to pociąga, że macierze \mathbf{Q}_{vv} i $\mathbf{Q}_{\hat{\mathbf{I}}\hat{\mathbf{I}}}$ są osobliwe. Pokażemy to dla macierzy $\mathbf{Q}_{\hat{\mathbf{I}}\hat{\mathbf{I}}}$. W rezultacie wyrównania elementy wektora $\hat{\mathbf{I}}$ spełniają równania warunkowe. A więc skoro jest tych równań c , to przynajmniej c elementów wektora $\hat{\mathbf{I}}$ można obliczyć za pomocą pozostałych $(n - c)$ elementów. Rząd macierzy $\mathbf{Q}_{\hat{\mathbf{I}}\hat{\mathbf{I}}}$ jest zatem mniejszy od jej rozmiaru, i faktycznie

$$\begin{aligned} \text{rank}(\mathbf{Q}_{\hat{\mathbf{I}}\hat{\mathbf{I}}}) &= (n - r) = n_0 \\ \text{rank}(\mathbf{Q}_{vv}) &= r \end{aligned}$$

↗

9.7 Interpretacja geometryczna metody NK

Metoda najmniejszych kwadratów nie jest wyprowadzana w oparciu o rozważania geometryczne. W tym celu należałoby posłużyć się własnościami minimalnej wariancji. Ale geometryczna interpretacja MNK pozwala na lepsze z nią oswojenie się. Ciekawą interpretację podał Brown.

Rozważmy n -wymiarowy układ współrzędnych o ortogonalnych osiach $\mathbf{v} = (\nu_1, \nu_2, \nu_3, \dots, \nu_n)^T$. Wówczas forma kwadratowa $\Phi = \mathbf{v}^T \Sigma \mathbf{v}$ dodatnio określona, będzie reprezentowała n -wymiarową elipsoidę o środku w początku układu współrzędnych.

Jeżeli macierz Σ jest diagonalna, to osie elipsoidy pokrywają się z osiami układu współrzędnych. Jeżeli nie, elipsoida będzie obrócona względem osi układu. Posługując się transformacją obrotu, można elipsoidę doprowadzić do jej orientacji standardowej. Transformacja ta określona jest przez równanie $\mathbf{v}' = \mathbf{R}\mathbf{v}$, gdzie \mathbf{R} jest macierzą ($n \times n$), której wiersze lub kolumny zawierają znormalizowane wektory charakterystyczne (wektory własne) macierzy Σ .

Dzięki tej transformacji, problem dotyczący obserwacji skorelowanych można zredukować do problemu obserwacji nieskorelowanych. Wymiary hiper-elipsoidy, wskutek obrotu oczywiście nie ulegną zmianie. Długości jej osi są wprost proporcjonalne do pierwiastków kwadratowych wartości własnych macierzy Σ . Stałą proporcjonalności, pozwalającą na obliczenie absolutnych wymiarów hiper-elipsoidy jest $\sqrt{\Phi}$. Wynika stąd, że objętość hiper-elipsoidy jest wprost proporcjonalna do $(\sqrt{\Phi})^n$. Dlatego minimalizacja funkcji Φ jest równoważna minimalizacji objętości elipsoidy, co łącznie z żądaniem spełnienia równań warunkowych, oznacza, że uzyskane tą drogą residua opisują punkt leżący na hiper-elipsoidzie.

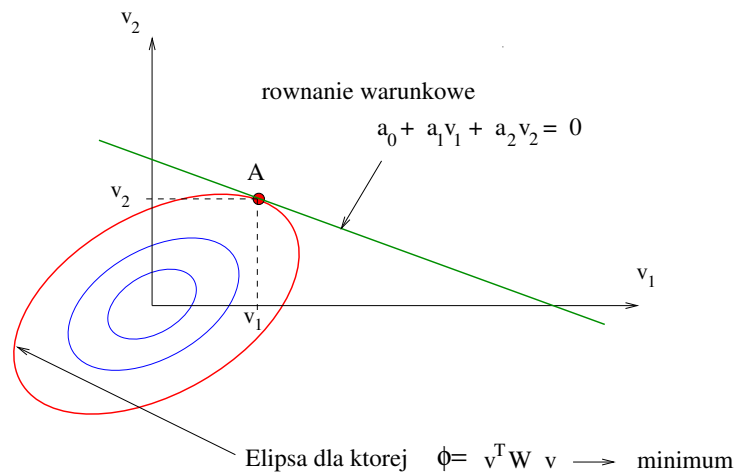
W celu uproszczenia wywodu możemy założyć, że z liniowych równań warunkowych usunięto wszystkie parametry. Mamy zatem r związków wyłącznie pomiędzy residuami. Wówczas każde równanie warunkowe reprezentuje hiper-płaszczyznę, a poszukiwane przez nas residua muszą należeć do hiper-prostej — miejsca geometrycznego przecięcia się r hiper-płaszczyzn.

Rozważmy rodzinę hiper-elipsoid określonych przez różne wartości Φ . Orientacja i rozmiary względne wszystkich takich elipsoid będą stałe, wszystkie będą miały środek w początku układu współrzędnych. Taką rodzinę można sobie wyobrazić jako podobną do nadmuchiwanej, rozszerzającego się "hiper-balonika", począwszy od jakiegoś infinitezimalnego maleństwa o kształcie hiper-elipsoidy. Można zatem wyobrazić sobie rozszerzającą się hiper-elipsoidę, aż do momentu kiedy nastąpi jej zetknięcie z hiper-prostą wyznaczoną poprzez przecięcie się hiper-płaszczyzn reprezentujących równania warunkowe. I właśnie dla tego punktu zetknięcia, spełnione są wszystkie równania warunkowe oraz objętość elipsoidy (a zatem i wartość Φ) jest najmniejsza. Współrzędne punktu styczności odpowiadają najbardziej prawdopodobnym residuom.

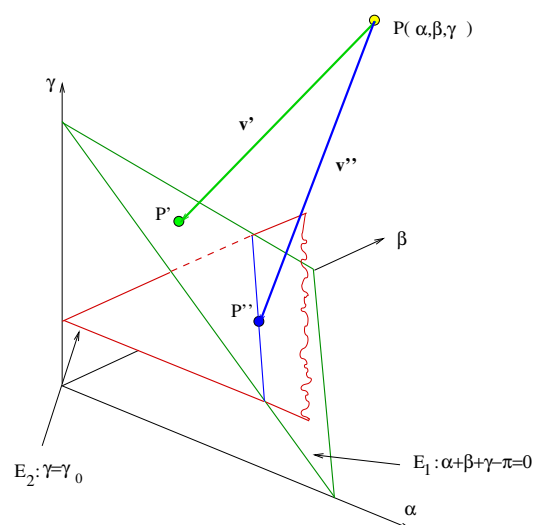
By zilustrować powyższy wywód, rozważmy prosty przypadek dwóch obserwacji z residuami ν_1, ν_2 . Na rysunku 9.2 układ współrzędnych reprezentuje dwuwymiarową przestrzeń residuów. Warunek który wiąże ν_1 z ν_2 może mieć ogólną postać $a - 0 + a_1\nu_1 + a - 2\nu - 2 = 0$, na rysunku 9.2 przedstawia go linia prosta. Natomiast forma kwadratowa $\Phi = \mathbf{v}^T \mathbf{W} \mathbf{v}$ reprezentuje rodzinę elips. Elipsa narysowana linią przerywaną, styczna do prostej warunkowej odpowiada minimalnej wartości Φ . Punkt styczności daje estymatory najmniejszych kwadratów na ν_1, ν_2 .

Przykład pokazany na rysunku 9.2 jest ogólnej natury i można zeń wyprowadzić przypadki szczegółowe. Np. jeżeli narzucimy warunek równości dwóch obserwacji, prosta warunkowa będzie nachylona pod kątem 45° do osi ν_1 i ν_2 . Dalej, gdyby obydwie obserwacje były jednakowej precyzji (σ_1, σ_2), wówczas elipsa przeszłaby w okrąg. W takim wypadku otrzymalibyśmy jednakowe estymatory ν_1, ν_2 .

Jeżeli zamiast residuów jako osie układu weźmiemy same obserwacje, możliwa jest nieco odmienna ilustracja MNK. Rozważmy przykładowo trzy kąty α, β, γ trójkąta płaskiego, które zmierzono z jednakową precyzją. Wobec nadmiaru $r = 1$, piszemy oczywisty



Rysunek 9.2: Ilustracja metody najmniejszych kwadratów, wyrównywanie z warunkami. Równanie warunkowe, łącznie z warunkiem najmniejszych kwadratów spełnione są dla punktu styczności A .



Rysunek 9.3: Ilustracja metody najmniejszych kwadratów, wyrównywanie z warunkami i więzami. Zmierzone wartości kątów w trójkącie odpowiadają punktowi P . Równanie warunkowe (płaszczyzna E_1), łącznie z warunkiem najmniejszych kwadratów (wektor reszduów v') spełnione są w punkcie P' . Równanie warunkowe z więzami (przecięcie płaszczyzn E_1 i E_2), łącznie z warunkiem najmniejszych kwadratów (wektor reszduów v'') spełnione są w punkcie P'' .

warunek $\alpha + \beta + \gamma = 0$, który musi być spełniony łącznie z warunkiem najmniejszych kwadratów. W trójwymiarowej przestrzeni α, β, γ (rysunek 9.3) warunek ten reprezentuje płaszczyzna E_1 .

Dane a priori rezultaty obserwacji α, β, γ ustalają położenie punktu P , niekoniecznie należącego do płaszczyzny E_1 . Procedura wyrównania tych kątów, dostarczy danych pozwalających na zastąpienie punktu P przez punkt P' . Przy czym, aby spełniony był warunek najmniejszych kwadratów, wektor residuów \mathbf{v}' musi mieć najmniejszą długość, a to oznacza, że odcinek $\overline{PP'}$ musi być prostopadły do płaszczyzny E_1 . W naszym przykładzie rozwiązanie najmniejszych kwadratów da residua

$$\nu_\alpha = \nu_\beta = \nu_\gamma = \frac{1}{3}(\pi - \alpha - \beta - \gamma)$$

bowiem płaszczyzna E_1 jest identycznie nachylona do wszystkich trzech osi. Residua będą miały takie same znaki — dodatnie, jeżeli P leży nad płaszczyzną, ujemne gdy P leży pod płaszczyzną E_1 .

Załóżmy teraz, że chcemy nałożyć pewien warunek (więzy) np. na kąt γ . Niech kąt ten ma być równy pewnej znanej skądinąd wartości γ_0 . Pozostałe kąty wyznaczone są doświadczalnie.

Wprowadzony dodatkowo warunek dotyczy modelu o nadmiarze $r = n - n_0 = 2$, bowiem skoro γ jest znane, to $n_0 = 1$. W trójwymiarowej przestrzeni wektorowej warunek aby $\gamma = \gamma_0$, reprezentowany jest przez płaszczyznę równoległą do płaszczyzny podstawowej układu α, β . Wyrównanie metodą najmniejszych kwadratów da w tym przypadku punkt P'' , dla którego spełnione są wszystkie warunki. Czyli P'' musi leżeć na przecięciu płaszczyzny E_1 z płaszczyzną o równaniu $\gamma = \gamma_0$, ponadto jest to taki punkt z tego przecięcia, dla którego długość odcinka $\overline{PP''}$ jest najmniejsza. †

9.8 Wyrównanie z warunkami, przypadki specjalne

Metoda najmniejszych kwadratów stosowana jest często w formie dwóch technik będących szczególnymi przypadkami ogólnej techniki wyrównywania z warunkami.

Pierwsza dotyczy sytuacji, w której w równaniach warunkowych mamy jedynie obserwacje a liczba parametrów $u = 0$, a więc mamy wówczas jedynie r równań warunkowych. Technika ta bywa stosowana w prostych problemach geometrycznych jak: wyrównywanie płaskiego trójkąta czy też mała sieć geodezyjna. Po rozwiązaniu równań normalnych, otrzymujemy tu wektor czynników Lagrange'a \mathbf{k} , który umożliwia obliczenie residuów \mathbf{v} , a te z kolei po dodaniu do wektora obserwacji \mathbf{l} dają estymatory $\hat{\mathbf{l}}$. Jeżeli bylibyśmy zainteresowani jakimś zbiorem zmiennych powiązanych z $\hat{\mathbf{l}}$ znanymi funkcjami, możemy je obliczyć po zakończeniu wyrównywania. Macierz kofaktorową tych zmiennych można również obliczyć o ile tylko w rezultacie wyrównania mamy do dyspozycji macierz $\mathbf{Q}_{\hat{\mathbf{l}}}$.

Druga technika stosowana bywa wówczas jeżeli możliwym jest sformułowanie równań warunkowych tak by zawierały wyłącznie jedną obserwację. Oznacza to, że w równaniach tych występują parametry, ponieważ liczba warunków jest wtedy równa liczbie obserwacji, dlatego ilość parametrów $u = n_0$. W sytuacji gdy mamy jedną obserwację na równanie warunkowe, macierz \mathbf{A} redukuje się do macierzy jednostkowej. Rozwiązanie równań normalnych daje wartości niewiadomych parametrów Δ oraz ich macierz kofaktorową $\mathbf{Q}_{\Delta\Delta}$. W dalszej kolejności możemy policzyć residua \mathbf{v} i estymatory $\hat{\mathbf{l}}$. †

9.8.1 Wyrównywanie tylko obserwacji

Ponieważ w tej technice nie występują żadne parametry, liniowe równania warunkowe mają postać

$$\mathbf{A}(\mathbf{l} + \mathbf{v}) = \mathbf{d} \quad (9.67)$$

lub

$$\mathbf{A}\mathbf{v} = \mathbf{d} - \mathbf{A}\mathbf{l} = \mathbf{f} \quad (9.68)$$

gdzie \mathbf{d} , \mathbf{f} są wektorami stałymi.

Zastosowanie warunku najmniejszych kwadratów oznacza tu wykonanie tych samych kroków co w przypadku techniki ogólnej, nie ma zatem potrzeby by je powtarzać. Dlatego podamy od razu postać całkowitego układu równań normalnych, mianowicie,

$$\begin{bmatrix} -\mathbf{W} & \mathbf{A}^T \\ \mathbf{A} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{v} \\ \mathbf{k} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{f} \end{bmatrix} \quad (9.69)$$

Po jego rozwiązaniu otrzymamy

$$\begin{aligned} \mathbf{k} &= \mathbf{W}_e \mathbf{f} = \mathbf{Q}^{-1} \mathbf{f} = (\mathbf{A} \mathbf{Q} \mathbf{A}^T)^{-1} \mathbf{f} \\ \mathbf{v} &= \mathbf{Q} \mathbf{A}^T \mathbf{k} = \mathbf{Q} \mathbf{A}^T \mathbf{W}_e \mathbf{f} \end{aligned} \quad (9.70)$$

A priori estymatę wariancji odniesienia obliczymy za pomocą formuły

$$\hat{\sigma}_0^2 = \frac{\mathbf{v}^T \mathbf{W} \mathbf{v}}{\mathbf{r}} = \frac{\mathbf{f}^T \mathbf{k}}{r} = \frac{\mathbf{k}^T \mathbf{f}}{r} \quad (9.71)$$

Pozostaje jeszcze do oszacowania precyzja uzyskanych rezultatów. W tym celu stosując prawo propagacji wariancji i kowariancji do występujących w tej technice zależności macierzowych :

$$\begin{aligned} \mathbf{f} &= \mathbf{d} - \mathbf{A}\mathbf{l} \\ \mathbf{k} &= (\mathbf{A} \mathbf{Q} \mathbf{A}^T)^{-1} \mathbf{f} = \mathbf{Q}_e^{-1} \mathbf{f} = -\mathbf{W}_e \mathbf{A} \mathbf{l} + \mathbf{W}_e \mathbf{d} = -\mathbf{W}_e \mathbf{A} \mathbf{l} + \mathbf{k}_0 \\ \mathbf{v} &= \mathbf{Q} \mathbf{A}^T \mathbf{k} = -\mathbf{Q} \mathbf{A}^T \mathbf{W}_e \mathbf{A} \mathbf{l} + \mathbf{Q} \mathbf{A}^T \mathbf{W}_e \mathbf{d} = -\mathbf{Q} \mathbf{A}^T \mathbf{W}_e \mathbf{A} \mathbf{l} + \mathbf{v}_0 \\ \hat{\mathbf{l}} &= \mathbf{l} + \mathbf{v} = (\mathbf{I} - \mathbf{Q} \mathbf{A}^T \mathbf{W}_e \mathbf{A}) \mathbf{l} + \mathbf{v}_0 \end{aligned}$$

metodą podstawień, dostaniemy następujące macierze autokofaktorowe i kros- kofaktorowe

:

$$\begin{aligned}
\mathbf{Q}_{ff} &= (-\mathbf{A})\mathbf{Q}(-\mathbf{A}^T) = \mathbf{Q}_e \\
\mathbf{Q}_{kk} &= \mathbf{W}_e\mathbf{A}\mathbf{Q}\mathbf{A}^T\mathbf{W}_e = \mathbf{W}_e \\
\mathbf{Q}_{vv} &= \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A}\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A}\mathbf{Q} = \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A}\mathbf{Q} \\
\mathbf{Q}_{\hat{ll}} &= (\mathbf{I} - \mathbf{Q}\mathbf{A}^T\mathbf{W}_e)\mathbf{Q}(\mathbf{I} - \mathbf{A}^T\mathbf{W}_e\mathbf{A}\mathbf{Q}) = (\mathbf{I} - \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A})^2\mathbf{Q} \\
\mathbf{Q}_{\hat{ll}} &= \mathbf{Q} - \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A}\mathbf{Q} = \mathbf{Q} - \mathbf{Q}_{vv} \\
\mathbf{Q}_{lf} &= \mathbf{I}\mathbf{Q}(-\mathbf{A})^T = -\mathbf{Q}\mathbf{A}^T \\
\mathbf{Q}_{lk} &= \mathbf{I}\mathbf{Q}(-\mathbf{W}_e\mathbf{A})^T = -\mathbf{Q}\mathbf{A}^T\mathbf{W}_e \\
\mathbf{Q}_{lv} &= \mathbf{I}\mathbf{Q}(-\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A})^T = -\mathbf{Q}_{vv} \\
\mathbf{Q}_{\hat{ll}} &= \mathbf{I}\mathbf{Q}(\mathbf{I} - \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A})^T = \mathbf{Q}_{\hat{ll}} \\
\mathbf{Q}_{fk} &= (-\mathbf{A})\mathbf{Q}(-\mathbf{W}_e\mathbf{A})^T = \mathbf{I} \\
\mathbf{Q}_{fv} &= (-\mathbf{A})\mathbf{Q}(-\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A})^T = \mathbf{A}\mathbf{Q} \\
\mathbf{Q}_{f\hat{l}} &= -\mathbf{A}\mathbf{Q}(\mathbf{I} - \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A})^T = \mathbf{0} \\
\mathbf{Q}_{kv} &= (-\mathbf{W}_e\mathbf{A})\mathbf{Q}(-\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A})^T = \mathbf{W}_e\mathbf{A}\mathbf{Q} \\
\mathbf{Q}_{k\hat{l}} &= (-\mathbf{W}_e\mathbf{A})\mathbf{Q}(\mathbf{I} - \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A})^T = \mathbf{0} \\
\mathbf{Q}_{v\hat{l}} &= (-\mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A})\mathbf{Q}(\mathbf{I} - \mathbf{Q}\mathbf{A}^T\mathbf{W}_e\mathbf{A})^T = \mathbf{0}
\end{aligned} \tag{9.72}$$

↑

9.8.2 Wyrównywanie obserwacji i parametrów

W tej technice mamy n równań warunkowych, każde zawiera tylko jedną obserwację

$$\mathbf{l} + \mathbf{v} + \mathbf{B}\Delta = \mathbf{d} \tag{9.73}$$

lub

$$\mathbf{v} + \mathbf{B}\Delta = -\mathbf{l} + \mathbf{d} = \mathbf{f} \tag{9.74}$$

Ponieważ residua \mathbf{v} w równaniu (9.74) są odseparowane od innych jego składników, rozwiązanie problemu najmniejszych kwadratów może być dokonane bardziej bezpośrednio, bez uciekania się do mnożników Lagrange. I tak, kładąc do warunku najmniejszych kwadratów residua wyznaczone z (9.74) dostaniemy

$$\Phi = \mathbf{v}\mathbf{W}\mathbf{v} = (\mathbf{f} - \mathbf{B}\Delta)^T(\mathbf{f} - \mathbf{B}\Delta) = \Delta^T\mathbf{B}^T\mathbf{W}\mathbf{B}\Delta - 2\mathbf{f}^T\mathbf{W}\mathbf{B}\Delta + \mathbf{f}^T\mathbf{W}\mathbf{f} \tag{9.75}$$

Warunek minimalizacji formy (9.75) oznacza

$$\frac{\partial\Phi}{\partial\Delta} = -2\Delta^T\mathbf{B}^T\mathbf{W}\mathbf{B} - 2\mathbf{f}^T\mathbf{W}\mathbf{B} = \mathbf{0} \tag{9.76}$$

lub

$$(\mathbf{B}^T\mathbf{W}\mathbf{B})\Delta = \mathbf{B}^T\mathbf{W}\mathbf{f} \tag{9.77}$$

A stosując oznaczenia jak w równaniu (9.20) będziemy mieli

$$\begin{aligned}
\mathbf{N}\Delta &= \mathbf{t} \\
\Delta &= \mathbf{N}^{-1}\mathbf{t}
\end{aligned} \tag{9.78}$$

Dysponując parametrami Δ obliczymy wyrównane wartości estymatorów $\hat{\mathbf{l}}$

$$\hat{\mathbf{l}} = \mathbf{l} + \mathbf{f} - \mathbf{B}\Delta \quad (9.79)$$

Ponieważ w tej technice wektor parametrów Δ jest różny od zera, w celu obliczenia a posteriori estymaty wariancji odniesienia $\hat{\sigma}_0^2$, kwadratową formę $\mathbf{v}^T \mathbf{W} \mathbf{v}$ musimy wyznaczyć z pełnego wyrażenia

$$\mathbf{v}^T \mathbf{W} \mathbf{v} = \mathbf{f}^T \mathbf{W} \mathbf{f} - \Delta \mathbf{t} \quad (9.80)$$

I podobnie jak dla poprzedniej techniki, proces wyrównywania obserwacji należy zakończyć oszacowaniem precyzji uzyskanych rezultatów. Stosując prawo propagacji wariancji i kowariancji do zależności macierzowych:

$$\begin{aligned} \mathbf{l} &= \mathbf{ll} \\ \mathbf{f} &= -\mathbf{l} + \mathbf{d} \\ \Delta &= \mathbf{N}^{-1} \mathbf{B} \mathbf{W} \mathbf{f} = -\mathbf{N}^{-1} \mathbf{B}^T \mathbf{W} \mathbf{l} + -\mathbf{N}^{-1} \mathbf{B}^T \mathbf{W} \mathbf{d} = -\mathbf{N}^{-1} \mathbf{B}^T \mathbf{W} \mathbf{l} + \Delta_0 \\ \mathbf{v} &= \mathbf{f} - \mathbf{B}\Delta = \mathbf{f} - \mathbf{B} \mathbf{N}^{-1} \mathbf{B} \mathbf{W} \mathbf{f} \\ \hat{\mathbf{l}} &= \mathbf{l} + \mathbf{v} \end{aligned}$$

otrzymamy macierze autokofaktorowe i kroskofaktorowe

$$\begin{aligned} \mathbf{Q}_{ll} &= \mathbf{Q} \\ \mathbf{Q}_{ff} &= \mathbf{Q} \\ \mathbf{Q}_{\Delta\Delta} &= \mathbf{N}^{-1} \mathbf{B}^T \mathbf{W} \mathbf{Q}_{ff} \mathbf{W} \mathbf{B} \mathbf{N}^{-1} = \mathbf{N}^{-1} \\ \mathbf{Q}_{vv} &= \mathbf{Q} - \mathbf{B} \mathbf{N}^{-1} \mathbf{B}^T \\ \mathbf{Q}_{\hat{\mathbf{l}}\hat{\mathbf{l}}} &= \mathbf{Q} - \mathbf{Q}_{vv} \\ \mathbf{Q}_{lf} &= -\mathbf{Q} \\ \mathbf{Q}_{l\Delta} &= -\mathbf{Q}(\mathbf{W} \mathbf{B} \mathbf{N}^{-1}) = -\mathbf{B} \mathbf{N}^{-1} \\ \mathbf{Q}_{lv} &= \mathbf{Q}_{lf} - \mathbf{Q}_{lf}(\mathbf{W} \mathbf{B} \mathbf{N}^{-1} \mathbf{B}^T) = -\mathbf{Q} + \mathbf{B} \mathbf{N}^{-1} \mathbf{B}^T = -\mathbf{Q}_{vv} \\ \mathbf{Q}_{\hat{\mathbf{l}}\hat{\mathbf{l}}} &= \mathbf{Q} + \mathbf{Q}_{lv} = \mathbf{Q} - \mathbf{Q}_{vv} = \mathbf{Q}_{\hat{\mathbf{l}}\hat{\mathbf{l}}} \\ \mathbf{Q}_{f\Delta} &= \mathbf{Q}(\mathbf{W} \mathbf{B} \mathbf{N}^{-1}) = \mathbf{B} \mathbf{N}^{-1} \\ \mathbf{Q}_{fv} &= \mathbf{Q}_{ff} - \mathbf{Q}_{f\Delta} \mathbf{B}^T = \mathbf{Q} - \mathbf{B} \mathbf{N}^{-1} \mathbf{B}^T \\ \mathbf{Q}_{f\hat{\mathbf{l}}} &= \mathbf{Q}_{fl} + \mathbf{Q}_{fv} = \mathbf{B} \mathbf{N}^{-1} \mathbf{B}^T \\ \mathbf{Q}_{\Delta v} &= \mathbf{Q}_{\Delta f} - \mathbf{Q}_{\Delta\Delta} \mathbf{B}^T = \mathbf{0} \\ \mathbf{Q}_{\Delta\hat{\mathbf{l}}} &= \mathbf{Q}_{\Delta l} + \mathbf{Q}_{\Delta v} = -\mathbf{N}^{-1} \mathbf{B}^T \\ \mathbf{Q}_{v\hat{\mathbf{l}}} &= \mathbf{Q}_{vl} + \mathbf{Q}_{vv} = \mathbf{0} \end{aligned} \quad (9.81)$$

↗

9.9 Dygresja: transformacja współrzędnych

Zagadnienie transformacji współrzędnych z jednego układu do drugiego jest operacją często stosowaną w geodezji i astrometrii. Transformacje te dotyczą przestrzeni dwu i trójwymiarowej, mogą opisywać zmiany początku układu, zmiany kierunku jego osi, bez zmian albo z jednostajną zmianą skali itd. ↗

9.9.1 Transformacje dwuwymiarowe

I. Od współrzędnych polarnych do prostokątnych:

$$\begin{aligned}x_1 &= r \cos \theta \\x_2 &= r \sin \theta\end{aligned}\tag{9.82}$$

II. Z układu kartezjańskiego do układu kartezjańskiego:

a) obrót układu

$$\begin{aligned}y_1 &= x_1 \cos \alpha - x_2 \sin \alpha \\y_2 &= x_1 \sin \alpha + x_2 \cos \alpha\end{aligned}\tag{9.83}$$

lub

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\tag{9.84}$$

b) obrót i zmiana skali

$$\begin{aligned}y_1 &= x_1(s \cos \alpha) - x_2(s \sin \alpha) \\y_2 &= x_1(s \sin \alpha) + x_2(s \cos \alpha)\end{aligned}\tag{9.85}$$

lub

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = s \cdot \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = s \cdot \begin{bmatrix} a & -b \\ b & a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\tag{9.86}$$

c) obrót, zmiana skali i przesunięcie

$$\begin{aligned}y_1 &= ax_1 - bx_2 + c \\y_2 &= bx_1 + ax_2 + d\end{aligned}\tag{9.87}$$

Transformacja ta ma różne nazwy : transformacja czteroparametrowa, transformacja podobieństwa, liniowa transformacja konformalna .

Wymienione wyżej transformacje nie dotyczą sytuacji, w których mielibyśmy do czynienia z deformacją przestrzeni. Jedyną zmianą w kształcie jest tu jednostajne skalowanie. Transformacje podane poniżej nie mają takiego ograniczenia.

d) transformacja 6-cio parametrowa

$$\begin{aligned}y_1 &= ax_1 + bx_2 + c \\y_2 &= dx_1 + ex_2 + f\end{aligned}\tag{9.88}$$

e) 8-mio parametrowa transformacja rzutowania (liniowa ułankowo)

$$\begin{aligned}y_1 &= \frac{a_1x_1 + a_2x_2 + a_3}{c_1x_1 + c_2x_2 + 1} \\y_2 &= \frac{b_1x_1 + b_2x_2 + b_3}{c_1x_1 + c_2x_2 + 1}\end{aligned}\tag{9.89}$$

f) ogólna transformacja wielomianowa

$$\begin{aligned}y_1 &= a_0 + a_1x_1 + a_2x_2 + a_3x_1x_2 + a_4x_1^2 + a_5x_2^2 \dots \\y_2 &= b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + b_4x_1^2 + b_5x_2^2 \dots\end{aligned}\tag{9.90}$$

f) dwuwymiarowa wielomianowa transformacja konformalna

$$\begin{aligned}y_1 &= A_0 + A_1x_1 + A_2x_2 + A_3(x_1^2 - x_2^2) + A_4(2x_1x_2) \dots \\y_2 &= B_0 - A_2x_1 + A_1x_2 - A_4(x_1^2 - x_2^2) + A_3(2x_1x_2) \dots\end{aligned}\tag{9.91}$$

↗

9.9.2 Transformacje trójwymiarowe

I. Najczęściej stosowaną jest transformacja siedmioparametrowa, pozwalająca na trzy rotacje, trzy translacje i jedną zmianę skali.

$$\mathbf{y} = s\mathbf{M}\mathbf{x} + \mathbf{k} \quad (9.92)$$

gdzie:

- $\mathbf{y} = (y_1, y_2, y_3)^T$ — wektor współrzędnych po transformacji,
- $\mathbf{x} = (x_1, x_2, x_3)^T$ — współrzędne przed transformacją,
- $\mathbf{k} = (k_1, k_2, k_3)^T$ — wektor trzech przesunięć,
- \mathbf{M} — ortogonalna macierz wyrażona za pomocą trzech niezależnych zmiennych, często traktowanych jako trzy niezależne kąty rotacyjne,
- — s czynnik zmiany skali.

Inna nazwa tej transformacji brzmi — trójwymiarowa liniowa transformacja konformalna.

II. Ogólna transformacja wielomianowa.

$$\begin{aligned} y_1 &= a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_1^2 + a_5x_2^2 + a_6x_3^2 + a_7x_1x_2 + a_8x_2x_3 + a_9x_3x_1 \dots \\ y_2 &= b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_1^2 + b_5x_2^2 + b_6x_3^2 + b_7x_1x_2 + b_8x_2x_3 + b_9x_3x_1 \dots \\ y_3 &= c_0 + c_1x_1 + c_2x_2 + c_3x_3 + c_4x_1^2 + c_5x_2^2 + c_6x_3^2 + c_7x_1x_2 + c_8x_2x_3 + c_9x_3x_1 \dots \end{aligned} \quad (9.93)$$

III. Wielomianowa transformacja konformalna w trzech płaszczyznach.

$$\begin{aligned} y_1 &= A_0 + A_1x_1 + A_2x_2 - A_3x_3 + A_5(x_1^2 - x_2^2 - x_3^2) + 0 + 2A_7x_3x_1 + 2A_6x_1x_2 \\ y_2 &= B_0 - A_2x_1 + A_1x_2 + A_4x_3 + A_6(-x_1^2 + x_2^2 - x_3^2) + 2A_7x_2x_3 + 0 + 2A_5x_1x_2 \\ y_3 &= C_0 + A_3x_1 - A_4x_2 + A_1x_3 + A_7(-x_1^2 - x_2^2 + x_3^2) + 2A_6x_2x_3 + 2A_5x_3x_1 + 0 \end{aligned} \quad (9.94)$$

↗

9.10 Przykład

↗

9.10.1 Ogólny przypadek wyrównywania z warunkami

Dana jest dwuwymiarowa transformacja — rotacja ze zmianą skali — pomiędzy układami współrzędnych X i Y :

$$\begin{aligned} y_{1i} &= ax_{1i} - bx_{2i} \\ y_{2i} &= bx_{1i} + ax_{2i} \end{aligned} \quad (9.95)$$

Para (x_{1i}, x_{2i}) odpowiada współrzędnym i -tego punktu w układzie X , para (y_{1i}, y_{2i}) odpowiada współrzędnym tego punktu w układzie Y , współczynniki (a, b) są parametrami transformacji.

W tabeli poniżej, dla trzech punktów podano wartości współrzędnych z obu układów

i	x_1	x_2	y_1	y_2
1	0.0	1.0	-2.1	1.1
2	1.0	0.0	1.0	2.0
3	1.0	1.0	-0.9	2.8

Dla wszystkich trzech punktów zakładamy, że macierze kowariancji pomiarów są równe macierzom kofaktorowym, zatem dla dowolnej pary współrzędnych typu X , mamy

$$\mathbf{Q}_{xx} = \Sigma_{xx} = \begin{bmatrix} 0.01 & 0 \\ 0 & 0.01 \end{bmatrix} = 0.01\mathbf{I} \quad (9.96)$$

Tym razem, przeciwnie aniżeli w przykładzie 3, wartości współrzędnych typu Y , przypisujemy status wartości dokładnych. Przy takim założeniu, metodą najmniejszych kwadratów należy określić estymatory parametrów a i b , a posteriori estymatę wariancji odniesienia $\hat{\sigma}_0^2$ oraz macierze kowariancji współrzędnych Y -kowych.

Rozwiązanie.

Funkcjonalny model tego zagadnienia dotyczy geometrii na płaszczyźnie — obrotu układu współrzędnych. Zamierzamy przetransformować współrzędne z układu X do układu Y .

W tabeli podano współrzędne trzech punktów, a więc do dyspozycji mamy $n = 6$ wielkości obserwowanych. Ponieważ interesują nas wyłącznie parametry transformacji a, b stąd $n_0 = 2$. W rezultacie nadmiar $r = 4$, a skoro $u = 2$ musimy napisać $c = 6$ równań warunkowych, postaci

$$\begin{aligned} \hat{x}_{1i}a - \hat{x}_{2i}b - y_{1i} &= 0 \\ \hat{x}_{2i}a + \hat{x}_{1i}b - y_{2i} &= 0, \quad i = 1, 2, 3 \end{aligned} \quad (9.97)$$

gdzie $\hat{x}_{ji} = x_{ij} + \nu_i$ oznaczają estymatory, czyli wyrównane obserwacje.

Ze względu na niewiadome, równania (9.97) są równaniami nieliniowymi bo przecież mamy np. $\hat{x}_{1i}a + \dots$. Zatem byśmy mogli skorzystać z poznanych technik MNK musimy te równania zlinearyzować.

Dygresja.

Niech \mathbf{F} będzie wektorem składającym się z m funkcji nieliniowych. W równaniu

$$\mathbf{F}(\mathbf{x}) = \mathbf{0} \quad (9.98)$$

\mathbf{x} oznacza wektor o p niewiadomych elementach.

W kontekście metody najmniejszych kwadratów, niewiadomymi mogą być parametry, obserwacje itp.. Gdybyśmy dysponowali wektorem \mathbf{x}^0 zawierającym przybliżone wartości elementów, można by w jego otoczeniu rozwinąć formę $\mathbf{F}(\mathbf{x})$ w szereg Taylora, ograniczając się do wyrazów liniowych:

$$\mathbf{F}(\mathbf{x}^0) + \left(\frac{\partial \mathbf{F}(\mathbf{x})}{\partial \mathbf{x}} \right)_{\mathbf{x}^0} \cdot \Delta \mathbf{x} = \mathbf{0} \quad (9.99)$$

Zbiór pochodnych cząstkowych ze względu na składowe wektora \mathbf{x} można zapisać w postaci macierzy \mathbf{U} o rozmiarach $(m \times p)$, a wektor $\Delta \mathbf{x}$ o rozmiarach $(p \times 1)$ można traktować jako wektor poprawek do przybliżonych wartości wektora niewiadomych \mathbf{x} .

Po linearyzacji, zamiast nieliniowego układu równań (9.98) możemy rozwiązywać jego liniowy odpowiednik (9.99). Można nadać mu ogólną postać

$$\mathbf{U} \cdot \Delta \mathbf{x} = \mathbf{u} \quad (9.100)$$

W wyniku wyrównania MNK, jako rozwiązanie otrzymamy wektor $\Delta \mathbf{x}$. Jeżeli wektor przybliżony \mathbf{x}^0 był dostatecznie "dobry", w tym sensie, że równanie (9.99) jest adekwatnym substytutem równania (9.98), a więc takim, że wyrazy wyższych rzędów rzeczywiście są do pominięcia, wówczas estymatą najmniejszych kwadratów wektora \mathbf{x} jest $(\mathbf{x}^0 + \Delta \mathbf{x})$. Jednak najczęściej \mathbf{x}^0 nie jest aż tak dobrym przybliżeniem i suma $(\mathbf{x}^0 + \Delta \mathbf{x})$ jest jedynie ulepszoną wartością przybliżoną. Ale za pomocą lepszego przybliżenia $(\mathbf{x}^0 + \Delta \mathbf{x})$ można ułożyć nowe równania (9.99) lub (9.100) i wyznaczyć nową poprawkę $\Delta \mathbf{x}$. Pozwoli ona na kolejne udokładnienie przybliżenia wektora \mathbf{x} i w ten sposób możemy rozwiązania iterować do momentu gdy np. obliczona po raz kolejny poprawka jest zanedbywalnie mała.

Powracamy do rozwiązania przykładu. Napiszmy raz jeszcze nasze równania warunkowe. Dla każdego punktu pomiarowego mamy po dwie ($m = 2$) funkcje F zależne od niewiadomego wektora $\hat{\mathbf{x}}_i = (\hat{x}_{1i}, \hat{x}_{2i}, a, b)^T$:

$$\begin{aligned} F_{1i}(\hat{\mathbf{x}}_i) &= \hat{x}_{1i}a - \hat{x}_{2i}b - y_{1i} = 0 \\ F_{2i}(\hat{\mathbf{x}}_i) &= \hat{x}_{2i}a + \hat{x}_{1i}b - y_{2i} = 0 \end{aligned} \quad (9.101)$$

W otoczeniu punktu $\mathbf{x}_i^0 = (x_{1i}, x_{2i}, a_0, b_0)^T$ rozwijamy każde z równań w szereg Taylora przestając na wyrazach liniowych:

$$\begin{aligned} (x_{1i}a_0 - x_{2i}b_0 - y_{1i}) + a_0\Delta x_{1i} - b_0\Delta x_{2i} + x_{1i}\Delta a - x_{2i}\Delta b &= 0 \\ (x_{2i}a_0 + x_{1i}b_0 - y_{2i}) + a_0\Delta x_{2i} + b_0\Delta x_{1i} + x_{2i}\Delta a + x_{1i}\Delta b &= 0 \end{aligned} \quad (9.102)$$

a po utożsamieniu przyrostów Δx_{ji} z residuami ν_{ji}

$$\begin{aligned} a_0\nu_{1i} - b_0\nu_{2i} + x_{1i}\Delta a - x_{2i}\Delta b &= y_{1i} - a_0x_{1i} + b_0x_{2i} \\ a_0\nu_{2i} + b_0\nu_{1i} + x_{2i}\Delta a - x_{1i}\Delta b &= y_{2i} - a_0x_{2i} - b_0x_{1i} \end{aligned} \quad (9.103)$$

po umacierzowaniu

$$\begin{bmatrix} a_0 & -b_0 \\ b_0 & a_0 \end{bmatrix} \cdot \begin{bmatrix} \nu_{1i} \\ \nu_{2i} \end{bmatrix} + \begin{bmatrix} x_{1i} & -x_{2i} \\ x_{2i} & x_{1i} \end{bmatrix} \cdot \begin{bmatrix} \Delta a \\ \Delta b \end{bmatrix} = \begin{bmatrix} f_{1i} \\ f_{2i} \end{bmatrix} \quad (9.104)$$

lub

$$\begin{matrix} \mathbf{A}_i \cdot \boldsymbol{\nu}_i & + & \mathbf{B}_i \cdot \boldsymbol{\Delta} & = & \mathbf{f}_i \\ (2 \times 2) (2 \times 1) & & (2 \times 2) (2 \times 1) & = & (2 \times 1) \end{matrix}$$

Dla trzech punktów będziemy mieli trzy równania macierzowe:

$$\begin{aligned} \mathbf{A}_1 \cdot \boldsymbol{\nu}_1 + \mathbf{B}_1 \cdot \boldsymbol{\Delta} &= \mathbf{f}_1 \\ \mathbf{A}_2 \cdot \boldsymbol{\nu}_2 + \mathbf{B}_2 \cdot \boldsymbol{\Delta} &= \mathbf{f}_2 \\ \mathbf{A}_3 \cdot \boldsymbol{\nu}_3 + \mathbf{B}_3 \cdot \boldsymbol{\Delta} &= \mathbf{f}_3 \end{aligned}$$

które można połączyć w jedno równanie

$$\begin{bmatrix} \mathbf{A}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{A}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{A}_3 \end{bmatrix} \cdot \begin{bmatrix} \nu_1 \\ \nu_2 \\ \nu_3 \end{bmatrix} + \begin{bmatrix} \mathbf{B}_1 \\ \mathbf{B}_2 \\ \mathbf{B}_3 \end{bmatrix} \cdot [\boldsymbol{\Delta}] = \begin{bmatrix} \mathbf{f}_1 \\ \mathbf{f}_2 \\ \mathbf{f}_3 \end{bmatrix} \quad (9.105)$$

a w skrócie

$$\begin{matrix} \mathbf{A} \cdot \boldsymbol{\nu} & + & \mathbf{B} \cdot \boldsymbol{\Delta} & = & \mathbf{f} \\ (6 \times 6)(6 \times 1) & & (6 \times 2)(2 \times 1) & & (6 \times 1) \end{matrix} \quad (9.106)$$

Możemy teraz przystąpić do obliczeń. Macierze \mathbf{A} , \mathbf{f} obliczymy jeśli tylko będą znane przybliżone wartości a_0 , b_0 . W tym celu biorąc z tabeli pomiarowej dane dla punktu nr 1 mamy

$$\begin{aligned} -2.1 &= 0 \cdot a_0 - b_0 \cdot 1 \\ +1.1 &= 0 \cdot b_0 + a_0 \cdot 1 \end{aligned} \quad (9.107)$$

stąd $a_0 \approx 1$, $b_0 \approx 2$.

Podstawiamy te wartości do macierzy \mathbf{A}_i , zauważmy, że w naszym przykładzie macierze te są sobie równe, stąd dla $i=1,2,3$ mamy

$$\mathbf{A}_i = \begin{bmatrix} a_0 & -b_0 \\ b_0 & a_0 \end{bmatrix} = \begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix} \quad (9.108)$$

Macierz \mathbf{f} wynosi

$$\mathbf{f} = \begin{bmatrix} y_{11} - x_{11}a_0 + x_{21}b_0 \\ y_{21} - x_{21}a_0 - x_{11}b_0 \\ y_{12} - x_{12}a_0 + x_{22}b_0 \\ y_{22} - x_{22}a_0 - x_{12}b_0 \\ y_{13} - x_{13}a_0 + x_{23}b_0 \\ y_{23} - x_{23}a_0 - x_{13}b_0 \end{bmatrix} = \begin{bmatrix} -0.1 \\ 0.1 \\ 0 \\ 0 \\ 0.1 \\ -0.2 \end{bmatrix} \quad (9.109)$$

W dalszej kolejności obliczymy macierz \mathbf{B} , która w całym procesie iteracyjnym nie zmienia się. Podstawiając odpowiednio wartości obserwowane uzyskamy:

$$\mathbf{B} = \begin{bmatrix} x_{11} & -x_{21} \\ x_{21} & x_{11} \\ x_{12} & -x_{22} \\ x_{22} & x_{12} \\ x_{13} & -x_{23} \\ x_{23} & x_{13} \end{bmatrix} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \quad (9.110)$$

Po obliczeniu macierzy \mathbf{B} , \mathbf{A} oraz \mathbf{f} możemy przystąpić do utworzenia układu równań normalnych i do rozwiązywania go. Korzystając z wyrażeń podanych wcześniej, w rezultacie nietrudnych obliczeń otrzymamy wektor Δ

$$\Delta = \begin{bmatrix} 0.0 \\ -0.05 \end{bmatrix}$$

pozwalający na obliczenie ulepszonych wartości współczynników poszukiwanej transformacji

$$a = a_0 + \Delta a = 1.00$$

$$b = b_0 + \Delta b = 1.95$$

Z tymi wartościami należałoby ponownie obliczyć elementy macierzy \mathbf{A} i \mathbf{f} , (równania ((9.108), (9.109))) po czym moglibyśmy obliczyć kolejne poprawki do współczynników transformacji a i b .

W naszym przykładzie porzucamy na pierwszej iteracji, a zatem przyjmujemy, że wyrównane wartości współczynników wynoszą

$$\hat{a} = 1.00, \quad \hat{b} = 1.95$$

Dysponując wektorem Δ obliczymy residua \mathbf{v}

$$\mathbf{v} = (0.01, 0.08, 0.02, 0.01, -0.05, -0.05)^T$$

W trakcie rozwiązywania równań normalnych obliczana jest macierz \mathbf{N}^{-1} , tym samym mamy oszacowanie precyzji wektora Δ , tzn. znamy jego macierz autokofaktorową

$$\mathbf{Q}_{\Delta\Delta} = \mathbf{N}^{-1} = 10^{-2} \begin{bmatrix} 1.25 & 0 \\ 0 & 1.25 \end{bmatrix}$$

jak widzimy, dla tej macierzy $rank = 2$.

Posługując się prawem propagacji wariancji i kowariancji możemy obliczyć autokofaktorową macierz estymatorów $\mathbf{Q}_{\hat{u}}$

$$\mathbf{Q}_{\hat{u}} = 10^{-3} \begin{bmatrix} 2.5 & 0 & 0 & -2.5 & 2.5 & -2.5 \\ 0 & 2.5 & 2.5 & 0 & 2.5 & 2.5 \\ 0 & 2.5 & 2.5 & 0 & 2.5 & 2.5 \\ -2.5 & 0 & 0 & 2.5 & -2.5 & 2.5 \\ 2.5 & 2.5 & 2.5 & -2.5 & 5.0 & 0 \\ -2.5 & 2.5 & 2.5 & 2.5 & 0 & 5.0 \end{bmatrix}$$

Dalej, zgodnie z żądaniem określonym w temacie przykładu, obliczamy a posteriori wariancję odniesienia $\hat{\sigma}_0^2$. Dla nadmiaru $r = 4$, po obliczeniu wartości formy kwadratowej $\mathbf{v}^T \mathbf{W} \mathbf{v}$ (patrz formuły (9.56), (9.58), (9.59)), otrzymamy

$$\hat{\sigma}_0^2 = 0.3123, \quad \hat{\sigma}_0 = 0.56$$

Gdybyśmy nie urwali obliczeń po pierwszym kroku i kontynuowali proces iteracyjny, otrzymalibyśmy nieco inne bardziej dokładne rezultaty. Podajemy je niżej dla ilustracji, oraz by z ich pomocą dokonać odpowiedniego testu statystycznego. I tak mielibyśmy odpowiednio:

$$\mathbf{Q}_{\Delta\Delta} = \mathbf{N}^{-1} = 10^{-2} \begin{bmatrix} 1.212 & 0 \\ 0 & 1.212 \end{bmatrix}$$

$$Q_{\hat{\beta}\hat{\beta}} = 10^{-3} \begin{bmatrix} 2.916 & 0 & 0.052 & -2.750 & 2.579 & -2.538 \\ 0 & 2.916 & 2.750 & 0.052 & 2.538 & 2.579 \\ 0.052 & 2.750 & 2.595 & 0 & 2.439 & 2.387 \\ -2.750 & 0.052 & 0 & 2.595 & -2.387 & 2.439 \\ 2.579 & 2.538 & 2.439 & -2.397 & 4.489 & 0 \\ -2.538 & 2.579 & 2.387 & 2.439 & 0 & 4.489 \end{bmatrix}$$

$$\hat{\sigma}_0^2 = 0.3114$$

$$\hat{\sigma}_0 = 0.56$$

A posteriori wariancja odniesienia $\hat{\sigma}_0^2$ może być testowana statystycznie w stosunku do jej wartości a priori. Jak pamiętamy, wartość ta wynosi $\sigma_0^2 = 1$ i odpowiada nieskończonej liczbie stopni swobody. Z naszego rachunku, dla czterech stopni swobody $r = 4$, otrzymaliśmy $\hat{\sigma}_0^2 = 0.3114$. Utwórzmy zmienną χ^2

$$\chi_m^2 = \frac{m \cdot \hat{\sigma}^2}{\sigma_0^2}$$

jej wartość odpowiadająca wielkościom obserwowanym wynosi

$$\chi_4^2 = \frac{4 \cdot 0.3114}{1.0} = 1.2441$$

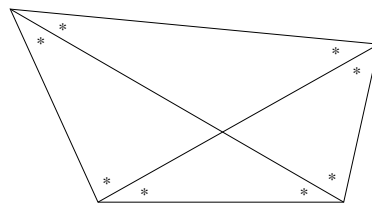
Hipotezą zerową będzie $H_0 : \sigma^2 = \sigma_0^2$, hipotezą przeciwną $H_1 : \sigma^2 > \sigma_0^2$. Zatem, hipoteza H_0 powinna zostać odrzucona jeżeli $\chi_4^2 > \chi_{\alpha,4}^2$.

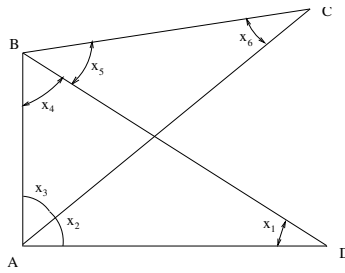
Przy poziomie istotności $\alpha = 0.05$, mamy $\chi_{0.05,4}^2 = 9.49$, stąd możemy twierdzić, że $\hat{\sigma}_0^2 = 0.3114$ — estymata a posteriori wariancji odniesienia nie różni się statystycznie od swej wartości a priori $\sigma_0^2 = 1.0$. \uparrow

9.11 Dodatek A. Zadania

Zadanie 1.

W czworoboku z rysunku po prawej stronie, zmierzono wszystkie kąty oznaczone gwiazdką. Określ nadmiar dla tego przypadku, jeśli celem było wyznaczenie kształtu czworokąta.



Zadanie 2.

W figurze geometrycznej przedstawionej na rysunku obok zmierzono sześć kątów. Otrzymano następujące wartości kątów:

$$x_1 = 48.88^\circ, \quad x_2 = 42.10^\circ, \quad x_3 = 44.52^\circ, \\ x_3 = 43.80^\circ, \quad x_5 = 46.00^\circ, \quad x_6 = 44.70^\circ.$$

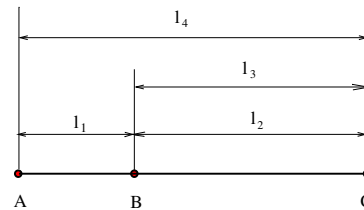
Metodą najmniejszych kwadratów wyznacz wyrównane wartości tych kątów.

Zadanie 3.

Wykonano pomiary odległości wg schematu pokazanego na rysunku obok. Pomiary mają identyczną precyzję i są nieskorelowane. Otrzymano następujące wartości odległości :

$$l_1 = 110.010[m], \quad l_2 = 200.050[m], \\ l_3 = 200.070[m], \quad l_4 = 300.090[m].$$

Posługując się metodą najmniejszych kwadratów wyznacz wyrównaną wartość odległości pomiędzy punktami *A* i *C*.

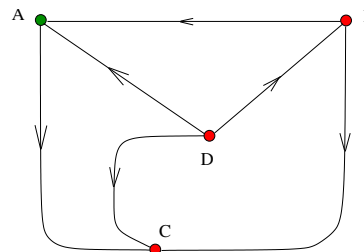
**Zadanie 4.**

Rysunek po prawej stronie przedstawia niewielką sieć geodezyjną, w której punkt *A* jest punktem odniesienia o znanej wysokości nad poziomem morza, równej 230.130 [m]. Dla każdej pary punktów tej sieci dokonano pomiarów różnic w wysokości, rezultaty zamieszczono w tabeli poniżej.

Punkt niższy	Punkt wyższy	Różnica w wysokości	Odległość pomiędzy punktami
B	A	$l_1=11.973$ [m]	20 [km]
D	B	$l_2=10.940$	12
D	A	$l_3=22.932$	15
B	C	$l_4=21.040$	28
D	C	$l_5=31.891$	20
A	C	$l_6= 8.983$	26

Ostatnia kolumna tabeli zawiera odległości pomiędzy punktami, dla których miały miejsce pomiary wysokości. Stosując metodę najmniejszych kwadratów oblicz wysokości nad poziomem morza punktów *B*, *C*, i *D* tej sieci, przy założeniu, że pomiary wysokości były nieskorelowane i niejednakowej dokładności, przy czym wagi pomiarów są odwrotnie proporcjonalne do podanych wzajemnych odległości. Obliczenia wykonać w formalizmie macierzowym.

↗



Bibliografia

- [1] Eadie W.T. Drijard D. James F.E. Roos M. Sadoulet B. *Metody statystyczne w fizyce doświadczalnej*. PWN Warszawa, 1989.
- [2] Szydlowski et al. *Teoria Pomiarów*. PWN Warszawa, 1974.
- [3] Szydlowski Henryk. *Niepewności w Pomiarach*. Wydawnictwo Naukowe UAM, 2001.
- [4] Fisz Marek. *Rachunek prawdopodobieństwa i statystyka matematyczna*. PWN Warszawa, 4 edition, 1967.
- [5] Nowak R. *Statystyka dla fizyków*. PWN, Warszawa, 1 edition, 2002.
- [6] Kozarski M. Szurmak Z. *Mini kalkulatory w obliczeniach naukowych i technicznych*. WNT Warszawa, 1980.
- [7] Strzałkowski A. Śliżyński A. *Matematyczne metody opracowywania wyników pomiarów*. PWN Warszawa, 2 edition, 1973. str 10-11.

↑