

CAS Lecture Notes Numer 5

Ryszard Zieliński

**Statystyka matematyczna
stosowana**

Elementy

Ryszard Zieliński

Instytut Matematyczny
Polska Akademia Nauk
Śniadeckich 8
00-956 Warszawa
E-mail: R.Zielinski@impan.pl

Korekta: **Henryka Walas**

Redaktor merytoryczny: **Stanisław Janeczko**

Skład redakcji: **Małgorzata Zielińska, Anna Żubrowska**

Projekt graficzny i skład okładki: **Emilia Bojańczyk / Podpunkt**

© Copyright by Centrum Studiów Zaawansowanych Politechniki Warszawskiej,
Warszawa 2011

Informacje o innych wydawnictwach tej serii dostępne pod adresem www.csz.pw.edu.pl

ISBN: 978-83-61993-03-2

Wydrukowano w Polsce

Spis treści

1. Wprowadzenie	1
2. Modele gaussowskie.....	4
2.1. Model pomiaru ze znaną precyzją.....	4
2.1.1. Problem i oznaczenia.....	4
2.1.2. Estymacja parametru μ	4
2.1.3. Testowanie hipotez o parametrze μ	8
2.2. Model pomiaru z nieznaną precyzją	14
2.2.1. Problem i oznaczenia.....	14
2.2.2. Estymacja punktowa σ i σ^2	15
2.2.3. Estymacja przedziałowa σ i σ^2	17
2.2.4. Testowanie hipotez o wariancji σ^2	19
2.2.5. Estymacja parametru μ	23
2.2.6. Testowanie hipotez o parametrze μ	25
3. Modele parametryczne.....	29
3.1. Problem i oznaczenia	29
3.2. Statystyki pozycyjne. Rozkład beta	31
3.3. Estymacja mediany μ	33
3.3.1. Estymacja punktowa. Obciążenie estymatora	33
3.3.2. Estymacja przedziałowa	35
3.4. Testowanie hipotez o medianie μ	35
3.5. Zastosowania.....	36
3.5.1. Rozkład Cauchy'ego. Parametr położenia i parametr skali.....	36
3.5.2. Rozkład Levy'ego. Parametr skali.....	38
3.5.3. Rozkład Pareto. Parametr kształtu	39
4. Modele nieparametryczne	41
4.1. Problem i oznaczenia	41
4.2. Estymacja mediany μ	42
4.2.1. Estymacja punktowa mediany μ	42
4.2.2. Przedział ufności dla mediany μ	42
4.3. Testowanie hipotez o medianie μ	45



5. Retrospekcja.....	48
Oznaczenia.....	51
Literatura.....	52
Indeks	53



*Not everything that can be counted counts,
and
not everything that counts can be counted.*

Albert Einstein

1. Wprowadzenie

Odkąd komputery zblądziły pod strzechy, także w statystyce można więcej uwagi przeznaczyć na myślenie niż na liczenie. Prezentowany wykład nie jest jeszcze jednym podręcznikiem statystyki matematycznej; jest już ich dostatecznie dużo — w Literaturze są wymienione te najnowsze i najłatwiej dostępne w języku polskim. W naszym wykładzie będziemy mówili o sprawach, które w tych podręcznikach rzadko są poruszane.

Gdy będziemy mówili o konstrukcji testu do weryfikacji hipotezy statystycznej, zasadniczą rolę będziemy przypisywali hipotezie alternatywnej. Już na początku wprowadzimy pojęcie *mocy testu statystycznego* i pokażemy, jak rzut oka na funkcję mocy pozwala wybrać, na przykład, odpowiedni test Studenta (test prawostronny, lewostronny lub dwustronny). Rozważając estymatory punktowe natychmiast będziemy podnosili kwestię oceny ich dokładności; za odpowiednie do tego narzędzie wybierzemy *przedziały ufności*. Żeby nie zaciemniać wykładu bardziej złożonymi modelami statystycznymi, dla demonstracji najważniejszych idei statystyki matematycznej zajmiemy się prostym i intuicyjnie dla każdego oczywistym *statystycznym modelem pomiaru*. Rozumiemy przez to następującą sytuację. Mamy zmierzyć pewną nieznaną wielkość liczbową μ . Wykonując pomiar otrzymujemy wynik X , który różni się od μ o błąd losowy ε ; mamy więc wynik $X = \mu + \varepsilon$ i na tej podstawie chcemy odpowiedzieć na różne pytania dotyczące μ . Sytuacja jest bardzo ogólna. W tym schemacie mieści się zarówno pomiar długości lub ciężaru jakiegoś fizycznego obiektu, jak też przeciętny czas życia lub przeciętna cena danego papieru na giełdzie; wówczas „błąd losowy” ε jest po prostu odchyłką wielkości indywidualnego obiektu od wielkości przeciętnej (na przykład średniej lub mediany) dla całej populacji.

Rutynowe postępowanie polega na wielokrotnym powtórzeniu pomiaru X i wnioskowaniu o nieznanym parametrze μ na podstawie jakiegoś uśrednienia otrzymanych wyników pomiarów. Liczbę powtórzeń pomiarów w całym wykładzie będziemy oznaczali przez n , a poszczególne pomiary przez X_1, \dots, X_n . Mamy więc obserwacje

$$X_j = \mu + \varepsilon_j, \quad j = 1, \dots, n,$$



gdzie $\varepsilon_1, \dots, \varepsilon_n$ jest ciągiem niezależnych błędów losowych o takim samym rozkładzie, powiedzmy o dystrybuancie F . Jeżeli o rozkładzie F nic nie wiemy, to oczywiście na podstawie obserwacji X_1, \dots, X_n nic nie potrafimy powiedzieć o mierzonej wielkości μ . W odpowiednim miejscu powiemy o sytuacji, w której zwielokrotnianie liczby n pomiarów nie ma sensu lub nawet może psuć wyniki wnioskowania.

W wykładzie omawiamy cztery, coraz bardziej ogólne modele statystyczne pomiaru. W pierwszym z nich rozważamy wnioskowanie o parametrze μ , gdy wiadomo, że błąd losowy ε ma standardowy rozkład normalny o średniej równej zeru i znanej wariancji σ^2 (znanym odchyleniu standardowym σ). Oznaczamy ten rozkład przez $N(0, \sigma)$. Odchylenie standardowe σ reprezentuje tu dokładność pomiaru (rozrzut, rozproszenie wokół wartości μ , precyzję przyrządu pomiarowego). Jest to elementarna część wykładu, raczej wszystkim znana, ale bardzo istotna, bo właśnie tu, na tym elementarnym technicznie poziomie, wprowadzamy oznaczenia i pojęcia, którymi będziemy się później intensywnie posługiwać. Drugi model różni się tylko tym od pierwszego, że odchylenie standardowe σ nie jest znane i dla wnioskowania o wielkości μ będziemy najpierw musieli oszacować σ na podstawie obserwacji X_1, \dots, X_n . Trzeci z rozważanych dalej modeli to model ze znanym rozkładem F , już niekoniecznie rozkładem normalnym. Motywacja wynika ze współczesnych zastosowań statystyki matematycznej w finansach, ubezpieczeniach i ekologii; występują tu rozkłady, w których prawdopodobieństwo pojawienia się bardzo dużych obserwacji jest znacznie większe niż jest to realnie możliwe w modelach gaussowskich. Rozkłady te mogą nie mieć wartości oczekiwanej, a średnia arytmetyczna obserwacji może być bardziej rozproszona wokół estymowanej wielkości μ niż pojedyncza obserwacja. W takich modelach posługiwanie się standardowymi miarami, takimi jak średnia i wariancja, nie znajduje żadnego uzasadnienia. Czwarty model to model z nieznanym rozkładem F . W tym przypadku o rozkładzie F będziemy zakładali tylko tyle, że jest to rozkład z ciągłą i ściśle rosnącą dystrybuantą.

To, co do tej pory mówiliśmy o rozkładzie prawdopodobieństwa, dotyczy tylko rozkładu prawdopodobieństwa błędu losowego ε . Faktycznie interesuje nas rozkład prawdopodobieństwa obserwacji X , a ten różni się od rozkładu F błędu ε „przesunięciem” μ . Rozkład obserwacji X będziemy oznaczali przez F_μ . W rozważanym modelu mamy $F_\mu(x) = F(x - \mu)$. Wielkości μ nie znamy — właśnie mamy ją oszacować. Wobec tego rozkład prawdopodobieństwa obserwacji X także nie jest znany. Jeżeli interesuje nas jakieś zdarzenie losowe związane z obserwacją X , na przykład zdarzenie $\{X > 0\}$ polegające na tym, że w wyniku pomiaru otrzymamy wartość dodatnią, to na pytanie o prawdopodobieństwo tego zdarzenia nie ma jednoznacznej odpowiedzi, dopóki nie sprecyzujemy wielkości μ . Formalnie ujmujemy to zapisem $P_\mu\{X > 0\}$, pamiętając o tym, że w statystyce matematycznej nie ma jednego prawdopodobieństwa; w danym



przypadku jest ich tyle, ile różnych wartości μ w naszym modelu bierzemy pod uwagę. Mówiąc ogólnie, w statystyce matematycznej mamy całą rodzinę prawdopodobieństw i trzeba zawsze wyjaśnić, z którym (z którymi) z nich mamy właśnie do czynienia. Służy temu dodatkowy indeks przy symbolu P lub, gdy mówimy o wartościach oczekiwanych, przy symbolu E . Indeks staje się niepotrzebny, gdy wiadomo, z jakim rozkładem mamy do czynienia. Na przykład, gdy mówimy o prawdopodobieństwie zdarzenia polegającego na tym, że zmienna losowa chi-kwadrat o ν stopniach swobody (oznaczamy ją przez χ_ν^2) nie przekroczy x , napiszemy $P\{\chi_\nu^2 \leq x\}$, bo ta zmienna ma jednoznacznie określony rozkład (patrz rozdz. 2.2.2).

Najważniejszy w całym wykładzie jest rozdział 2.1. Wprowadzamy w nim i komentujemy wszystkie potrzebne później pojęcia (estymator, przedział ufności, test statystyczny, moc testu, poziom krytyczny testu). W istocie rzeczy w dalszej części wykładu powtarzamy rozumowania z tego rozdziału dla coraz bardziej złożonych (i bliższych życiu) modeli statystycznych.

W wykładzie nie ma dużo rachunków. Rozkład normalny, rozkład beta, niecentralny rozkład t Studenta i in. traktujemy jak pojęcia, dla których w licznych i łatwo dostępnych pakietach komputerowych istnieją oprogramowania za pomocą funkcji standardowych danego pakietu. Jeżeli mówimy, na przykład, że $B(x; p, q)$ jest wartością dystrybuanty rozkładu beta o parametrach p, q w punkcie x , to zakładamy, że Czytelnik w razie potrzeby potrafi dowiedzieć się od swojego komputera, ile to wynosi dla danych x, p, q .

Zakładamy, że Czytelnik zna podstawowe fakty z teorii prawdopodobieństwa i statystyki na poziomie przynajmniej jednej z książek wymienionych w Literaturze. Na przykład w rozdz. 2.1.2 piszemy, bez żadnego dodatkowego komentarza: „ponieważ obserwacja X ma rozkład $N(\mu, \sigma)$, to średnia \bar{X}_n jest zmienną losową o rozkładzie normalnym $N(\mu, \sigma/\sqrt{n})$, czyli $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ jest zmienną losową o rozkładzie normalnym $N(0, 1)$ i dla dowolnie wybranego $\gamma \in (0, 1)$ mamy

$$P_\mu \{|\sqrt{n}(\bar{X}_n - \mu)/\sigma| \leq z_{(1+\gamma)/2}\} = \gamma,$$

gdzie z_α jest kwantylem rzędu α rozkładu normalnego $N(0, 1)$ ”. Zakładamy, że żadne dodatkowe wyjaśnianie nie jest potrzebne.



2. Modele gaussowskie

2.1. Model pomiaru ze znaną precyzją

2.1.1. Problem i oznaczenia

Rozważamy ciąg X_1, \dots, X_n obserwacji postaci

$$X_j = \mu + \varepsilon_j, \quad j = 1, \dots, n, \quad (2.1)$$

gdzie μ jest ustaloną, nieznaną liczbą oraz $\varepsilon_1, \dots, \varepsilon_n$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie normalnym $N(0, \sigma)$. Dystrybuantę standardowego rozkładu normalnego $N(0, 1)$ oznaczamy przez Φ , jej wartość w punkcie x przez $\Phi(x)$, i odpowiednio gęstość prawdopodobieństwa przez $\varphi(x)$.

Zakładamy, że odchylenie standardowe σ (wariancja σ^2) jest znane. Zadanie polega na wnioskowaniu o wielkości μ na podstawie pomiarów X_j zaburzonych błędem losowym ε_j . To wnioskowanie może polegać na udzieleniu odpowiedzi na pytanie „Ile wynosi μ ”, albo na pytanie „Czy prawdą jest, że $\mu = \mu_0$ ” dla danej liczby μ_0 . W pierwszym przypadku mówimy o *estymacji*, w drugim o *testowaniu hipotez*. Ponieważ pomiary są obciążone błędami losowymi, odpowiedź, na przykład, $\mu = 7$ na pierwsze pytanie może budzić zastrzeżenia. Spodziewamy się raczej bardziej rzeczowej odpowiedzi, na przykład $\mu = 7 \pm 0.1$ albo $\mu = 7 \pm 0.01$, sugerując tym samym, że odpowiedź *punktowa* jest obciążona błędem statystycznym, rzędu 0.1 w pierwszym lub 0.01 w drugim przypadku. Bardziej adekwatna dla naszego problemu jest raczej odpowiedź *przedziałowa*. Nawet popularne media uczą nas takiego patrzenia na wyniki pomiarów i mówią, na przykład, że „Pan X ma poparcie 37% wyborców z błędem statystycznym 2%”.

2.1.2. Estymacja parametru μ

Standardowe rozwiązanie polega na oszacowaniu parametru μ przez wartość średnią z obserwacji

$$\bar{X}_n = \frac{1}{n} \sum_{j=1}^n X_j. \quad (2.2)$$



Jak dokładne jest to oszacowanie? Oto standardowe rozumowanie: ponieważ obserwacja X ma rozkład $N(\mu, \sigma)$, to średnia \bar{X}_n jest zmienną losową o rozkładzie normalnym $N(\mu, \sigma/\sqrt{n})$, czyli $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ jest zmienną losową o rozkładzie normalnym $N(0, 1)$ i dla dowolnie wybranego $\gamma \in (0, 1)$ mamy

$$P_\mu \{ |\sqrt{n}(\bar{X}_n - \mu)/\sigma| \leq z_{(1+\gamma)/2} \} = \gamma,$$

gdzie z_α jest kwantylem rzędu α rozkładu normalnego $N(0, 1)$. Możemy to zapisać w postaci

$$P_\mu \left\{ \bar{X}_n - z_{(1+\gamma)/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X}_n + z_{(1+\gamma)/2} \frac{\sigma}{\sqrt{n}} \right\} = \gamma$$

i interpretować w następujący sposób: z wybranym prawdopodobieństwem γ , losowy przedział

$$\left(\bar{X}_n - z_{(1+\gamma)/2} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{(1+\gamma)/2} \frac{\sigma}{\sqrt{n}} \right) \quad (2.3)$$

zawiera nieznaną, szacowaną wartość parametru μ .

W pakiecie R wartość współczynnika $z_{(1+\gamma)/2}$ oblicza się za pomocą komendy `qnorm((1 + γ)/2)`

Przedział (2.3) nazywa się dwustronnym *przedziałem ufności* na poziomie ufności γ . Długość przedziału ufności, równa $2 z_{(1+\gamma)/2} \frac{\sigma}{\sqrt{n}}$, charakteryzuje dokładność (błąd) estymacji parametru μ . Możemy tym manipulować przez odpowiedni wybór γ lub/i n (jeżeli mamy czas i pieniądze na wykonanie odpowiednio dużej liczby pomiarów).

Analogicznie możemy skonstruować jednostronne przedziały ufności na poziomie ufności γ :

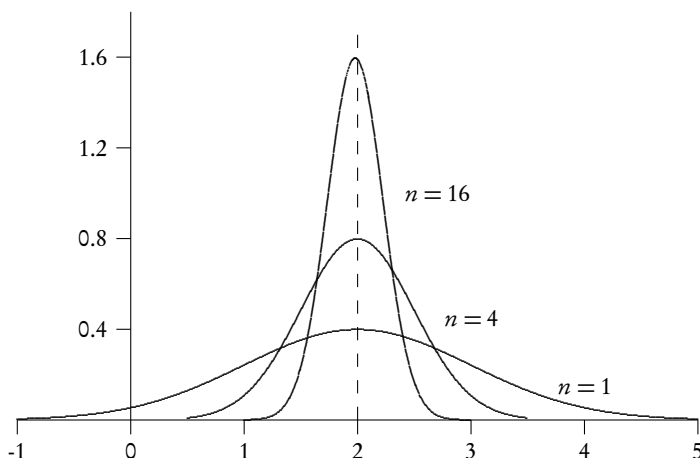
$$\left(\bar{X}_n + z_{1-\gamma} \frac{\sigma}{\sqrt{n}}, +\infty \right) \quad (2.4)$$

lub

$$\left(-\infty, \bar{X}_n + z_\gamma \frac{\sigma}{\sqrt{n}} \right). \quad (2.5)$$

Zwróćmy uwagę na rolę liczności n próby X_1, \dots, X_n i zauważmy, że wnioskowanie o parametrze μ staje się bardziej dokładne, gdy wielkość n próby rośnie. Wtedy rozkład prawdopodobieństwa średniej \bar{X}_n coraz bardziej koncentruje się wokół interesującej nas wartości parametru; gęstość rozkładu prawdopodobieństwa średniej dla $\mu = 2$, $\sigma = 1$ oraz $n = 1, 4, 16$ pokazano na rys. 2.1.





Rysunek 2.1. Gęstość rozkładu średniej \bar{X}_n z próby X_1, \dots, X_n pochodzącej z rozkładu normalnego $N(2, 1)$

Tyle o estymacji parametru μ za pomocą średniej arytmetycznej \bar{X}_n . Wprowadzenie do rozważań innych estymatorów rozpoczniemy od odpowiednich oznaczeń i definicji.

Uporządkujmy obserwacje X_1, \dots, X_n w ciąg niemalejący

$$X_{1:n}, \dots, X_{n:n}, \quad X_{1:n} \leq \dots \leq X_{n:n}. \quad (2.6)$$

Jest to ciąg *statystyk pozycyjnych*. W tym ciągu $X_{1:n}$ jest najmniejszą wartością zaobserwowaną w próbie. Istnieje dla niej nieco wygodniejsze oznaczenie X_{min} , podobnie jak dla największej wartości $X_{n:n}$ oznaczenie X_{max} . Jeżeli n jest liczbą nieparzystą, to obserwacja $X_{\frac{n+1}{2}:n}$ jest środkowa w tym sensie, że zarówno na lewo od niej, jak i na prawo od niej mamy w próbie taką samą liczbę $\frac{n-1}{2}$ obserwacji. Nazywa się ona *medianą*, czasami, dla uniknięcia nieporozumień, *medianą z próby* lub *medianą próbkową*; będziemy ją oznaczali przez M_n .

Mediana M_n zastępuje we wnioskowaniu średnią \bar{X}_n tam, gdzie pojawiają się podejrzenia, że skrajne obserwacje są obciążone jakimiś dużymi błędami i „lepiej jest nie brać ich pod uwagę”. Medianę traktuje się wtedy jako *odporny* estymator parametru μ . Dokładność oszacowania parametru μ za pomocą mediany M_n ocenia się, podobnie jak w przypadku estymatora \bar{X}_n , za pomocą przedziałów ufności; odpowiednie wzory otrzymuje się ze wzorów (2.3), (2.4) oraz (2.5), zastępując w nich \bar{X}_n przez M_n oraz współczynniki z_α/\sqrt{n} współczynnikami odpowiednimi dla rozkładu prawdopodobieństwa mediany M_n . O tym, jak obliczać te współczynniki mówimy dokładnie w rozdz. 3; tutaj odnotujmy tylko, że dwustronny przedział ufności na poziomie ufności γ ma postać



$$\left(M_n - \Phi^{-1} \left(B^{-1} \left(\frac{1+\gamma}{2}; \frac{n+1}{2}, \frac{n+1}{2} \right) \right) \sigma, \right. \\ \left. M_n + \Phi^{-1} \left(B^{-1} \left(\frac{1+\gamma}{2}; \frac{n+1}{2}, \frac{n+1}{2} \right) \right) \sigma \right), \quad (2.7)$$

gdzie $B^{-1}(\alpha; p, q)$ jest kwantylem rzędu α rozkładu beta z parametrem (p, q) .

W pakiecie R wartość współczynnika $\Phi^{-1} \left(B^{-1} \left(\frac{1+\gamma}{2}; \frac{n+1}{2}, \frac{n+1}{2} \right) \right)$ oblicza się za pomocą komendy `qnorm(qbeta((1 + γ)/2, (n + 1)/2, (n + 1)/2))`

W zastosowaniach zdarzają się również inne sytuacje. Rozważmy, na przykład, zadanie polegające na oszacowaniu średniej temperatury dobowej za pomocą termometru, który dokonuje pomiaru co pewien ustalony czas (np. co minutę), ale zapamiętuje tylko temperaturę najniższą i najwyższą w danej dobie. Estymatorem parametru μ może być teraz *środek rozrzutu* $(X_{min} + X_{max})/2$; dla oznaczenia tego estymatora użyjemy tutaj ogólnego symbolu $\hat{\mu}$. Jeżeli próba X_1, \dots, X_n pochodzi z rozkładu normalnego $N(\mu, \sigma)$, to rozkład zmiennej losowej $(\hat{\mu} - \mu)/\sigma$ jest taki sam jak rozkład środka rozrzutu $\hat{\mu}_\Phi$ z próby pochodzącej z rozkładu normalnego $N(0, 1)$. Gdybyśmy umieli obliczyć kwantyle τ_γ rozkładu zmiennej losowej $\hat{\mu}_\Phi$, umielibyśmy także obliczać przedziały ufności dla μ na podstawie środka rozrzutu $\hat{\mu}$. Na przykład dwustronny przedział ufności na poziomie ufności ma wtedy łatwą do wydedukowania postać

$$(\hat{\mu} - \tau_{(1+\gamma)/2} \sigma, \hat{\mu} + \tau_{(1+\gamma)/2} \sigma)$$

(por. przedział ufności (2.3) oparty na średniej \bar{X}_n). Jeżeli próba X_1, \dots, X_n pochodzi z rozkładu o dystrybuancie $F(x)$ i gęstości $f(x)$, to dystrybuanta rozkładu środka rozrzutu $\hat{\mu}$ wyraża się wzorem

$$P_F\{\hat{\mu} \leq t\} = n \int_{-\infty}^t [F(2t - x) - F(x)]^{n-1} f(x) dx. \quad (2.8)$$

Wzór ten można łatwo otrzymać, wypisując łączny rozkład statystyk pozycyjnych (X_{min}, X_{max})

$$f_{X_{min}, X_{max}}(x, y) = n(n-1)[F(y) - F(x)]^{n-2} f(x)f(y), \quad -\infty < x < y < +\infty,$$

i całkując tę gęstość po obszarze $\{(x, y) : (x + y)/2 \leq t\}$ obliczyć prawdopodobieństwo zdarzenia $\{(X_{min} + X_{max})/2 \leq t\}$. Ewentualnie można ten wzór po prostu wygooglować. W celu wyznaczenia współczynnika $\tau_{(1+\gamma)/2}$ wystarczy w jednym z licznych pakietów komputerowych rozwiązać względem t równanie

$$P_F\{\hat{\mu} \leq t\} = \frac{1+\gamma}{2}.$$



W pakiecie R całkę we wzorze (2.8) oblicza się za pomocą komendy `function(t, n){integrate(F, -Inf, t, t, n)}$value` po zdefiniowaniu dla rozkładu normalnego funkcji $F < -function(x, t, n){n * (pnorm(2 * t - x) - pnorm(x))^{n-1} * dnorm(x)}$

Dokładność oszacowania μ za pomocą mediany M_n lub środka rozrzutu $\hat{\mu}$ jest mniejsza niż dokładność tego oszacowania za pomocą średniej \bar{X}_n . Intuicja podpowiada, że tak powinno być, gdyż w omawianym przypadku wykorzystujemy mniej informacji z próby. Ilustruje to następująca tabelka obliczona dla poziomu ufności $\gamma = 0.9$:

n	$z_{(1+\gamma)/2}/\sqrt{n}$	$\Phi^{-1}\left(B^{-1}\left(\frac{1+\gamma}{2}; \frac{n+1}{2}, \frac{n+1}{2}\right)\right)$	$\tau_{(1+\gamma)/2}$
5	0.7356	0.8806	0.8398
25	0.3290	0.4087	0.5987

W bardziej zaawansowanych podręcznikach statystyki matematycznej dowodzi się, że oszacowanie parametru μ w naszym modelu za pomocą średniej \bar{X}_n z próby X_1, \dots, X_n jest najdokładniejsze.

2.1.3. Testowanie hipotez o parametrze μ

Rozpoczniemy od komentarza na temat testowania hipotez. Opiera się ono na następującym rozumowaniu: *jeżeli hipoteza H jest prawdziwa, to pewne określone zdarzenie Z jest niemożliwe*. Wykonuję badanie skierowane na obserwację zdarzenia Z . Jeżeli zaobserwuję zdarzenie Z , hipotezę uważam za nieprawdziwą. Ale jeżeli, mimo wszelkich możliwych wysiłków, nie uda mi się zaobserwować zdarzenia Z , to nie mam podstaw do kwestionowania hipotezy H ; nie mam także żadnych podstaw do uznania jej za prawdziwą. Uznanie za prawdziwą hipotezę H , gdy nie stwierdziłem Z , może mieć tragiczne skutki. Kontynuujemy jeszcze przez chwilę to rozumowanie. Mamy szklankę wody i wielką ochotę na to, żeby ją wypić. Jeżeli ta woda jest zdrowa, to z pewnością nie pachnie siarkowodorem. Obwąchuję więc wodę w szklance, nawet proszę o pomoc licznych znajomych, zgodnie nie stwierdzamy zapachu siarkowodoru, więc wodę wypijam: czy na pewno przeżyję to doświadczenie? W praktyce często tak bardzo wierzymy w prawdziwość jakiejś hipotezy, na przykład hipotezy, że nasze obserwacje pochodzą z rozkładu normalnego, że jeżeli za pomocą jakiegoś testu statystycznego nie odrzucimy jej, akceptujemy ją jako prawdziwą. Szczególnie w sytuacji gdyby odrzucenie hipotezy sprawiło nam różne kłopoty: jak, na przykład, obejść się w statystyce bez słynnego testu t Studenta gdyby okazało się, że założenia o normalności rozkładu obserwacji nie da się utrzymać? Interesujące, kto z Czytelników rozpozna źródło następującej uwagi: „Wystarczy mieć jakąś trzymającą



się kupy teorii, jakiś wyraźny pogląd na rzecz i chcieć tym poglądom wierzyć, aby wszystkie fakty układały się w łańcuch dowodów zupełnie zdawałoby się niezbitych. Lecz wystarczy zmienić pogląd, aby te fakty świadczyły o czymś wręcz odmiennym”.

W przypadku hipotez statystycznych rozumowanie ulega pewnej modyfikacji. Wróćmy na początek poprzedniego akapitu. Teraz powiemy, że *jeżeli hipoteza H jest prawdziwa, to zdarzenie Z jest praktycznie niemożliwe*. „Praktycznie niemożliwe” jest tym samym co „mało prawdopodobne” lub nawet „bardzo mało prawdopodobne”. Możemy to skwantyfikować, wybierając małą (bardzo małą) liczbę $\alpha > 0$ i formułując rozumowanie w następujący sposób: *jeżeli hipoteza H jest prawdziwa, to prawdopodobieństwo zdarzenia Z jest nie większe niż α ; jeżeli zaobserwujemy to zdarzenie, to na poziomie istotności α odrzucam H* . Odrzucenie hipotezy H , gdy jest ona prawdziwa, jest oczywiście błędem. W statystyce nazywa się *błędem pierwszego rodzaju (błędem I rodzaju)*. W opisanej konstrukcji wnioskowania, prawdopodobieństwo popełnienia takiego błędu nie przekracza danej, z góry zaakceptowanej, małej liczby α . Mówimy wtedy o odrzuceniu lub nieodrzuconiu weryfikowanej hipotezy H na zadanym poziomie istotności α .

Po tych ogólnych uwagach wracamy do modelu statystycznego (2.1) i rozpoczynamy od testowania hipotezy $H : \mu = \mu_0$ dla pewnej ustalonej liczby μ_0 . Spośród wszystkich możliwych rozkładów F_μ , $\mu \in \mathcal{R}$, hipoteza wyróżnia jeden rozkład F_{μ_0} i zadanie polega na sprawdzeniu, czy wyniki obserwacji X_1, \dots, X_n nie kwestionują przypuszczenia, że rozkładem prawdopodobieństwa, z którego pochodzą, jest właśnie ten wyróżniony rozkład. W zastosowaniach taki problem pojawia się na przykład przy sprawdzaniu przyrządu pomiarowego na danym obiekcie o znanej wartości μ .

Gdyby hipoteza była prawdziwa, to zdarzenie losowe polegające na tym, że średnia z próby \bar{X}_n różni się istotnie od wyspecyfikowanej przez hipotezę wartości μ_0 , byłoby mało prawdopodobne. Gdybyśmy zaobserwowali istotną różnicę, uznalibyśmy hipotezę za mało prawdopodobną i odrzucilibyśmy ją jako nieprawdziwą. Co w rozważanym teraz modelu statystycznym znaczy „różni się istotnie” i co „mało prawdopodobne”?

Wiemy, że jeżeli H jest prawdziwa, to zmienna losowa $\sqrt{n}(\bar{X}_n - \mu_0)/\sigma$ ma rozkład normalny $N(0, 1)$. Zdarzenie losowe $\{\sqrt{n}(\bar{X}_n - \mu_0)/\sigma < z_\alpha\}$ ma prawdopodobieństwo równe α . Jeżeli α jest małe, to wystąpienie tego zdarzenia losowego, czyli zdarzenia

$$D_\alpha^- := \left\{ \bar{X}_n < \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \right\} \quad (2.9)$$

prowadzi do odrzucenia hipotezy H na poziomie istotności α . Interpretacja: zaobserwowana średnia jest zbyt mała, aby uwierzyć w prawdziwość weryfikowanej



hipotezy. W typowych zastosowaniach przemysłowych, biologicznych i medycznych, ekonometrycznych itp. przyjmuje się zwykle $\alpha = 0.05$ ($z_{0.05} = -1.645$) lub $\alpha = 0.01$ ($z_{0.01} = -2.326$), ale są sytuacje gdy przyjmuje się np. $\alpha = 10^{-6}$. Zauważmy, że w naszym modelu poziom istotności α testu jest równy prawdopodobieństwu odrzucenia weryfikowanej hipotezy H , gdy jest ona prawdziwa. Zbiór tych wartości statystyki \bar{X}_n , które realizują zdarzenie D_α^- , nazywa się *obszarem krytycznym testu*.

Chwila zastanowienia wystarczy żeby zauważyć, że podobnie jak zdarzenie losowe D_α^- , również zdarzenia losowe

$$D_\alpha^+ := \left\{ \bar{X}_n > \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} \right\} \quad (2.10)$$

oraz

$$D_\alpha := \left\{ |\bar{X}_n - \mu_0| > z_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right\} \quad (2.11)$$

mają prawdopodobieństwo równe α , czyli tak jak zdarzenie D_α^- falsyfikują hipotezę H na przyjętym poziomie istotności α . Faktycznie więc dla rozpatrywanej hipotezy $H: \mu = \mu_0$ mamy teraz już trzy różne testy na danym poziomie istotności α : takich testów możemy skonstruować dowolnie dużo. Powstaje problem, który z nich wybrać. Sensowna odpowiedź wymaga zastanowienia się nad konsekwencjami zastosowania testu i wprowadzenia nowego elementu do naszych rozważań. Tym nowym elementem jest *moc testu* oraz *funkcja mocy testu*.

Rozpatrzmy dokładniej test z obszarem krytycznym D_α^- . Przypominamy: weryfikujemy hipotezę $H: \mu = \mu_0$. Przypuśćmy, że w rzeczywistości $\mu < \mu_0$. Jakie jest wtedy prawdopodobieństwo odrzucenia hipotezy H ? Intuicja podpowiada, że test jest tym lepszy, im to prawdopodobieństwo jest większe. Błąd polegający na tym, że hipoteza nie zostanie odrzucona, gdy nie jest prawdziwa, nazywa się *błędem drugiego rodzaju* (*błędem II rodzaju*).

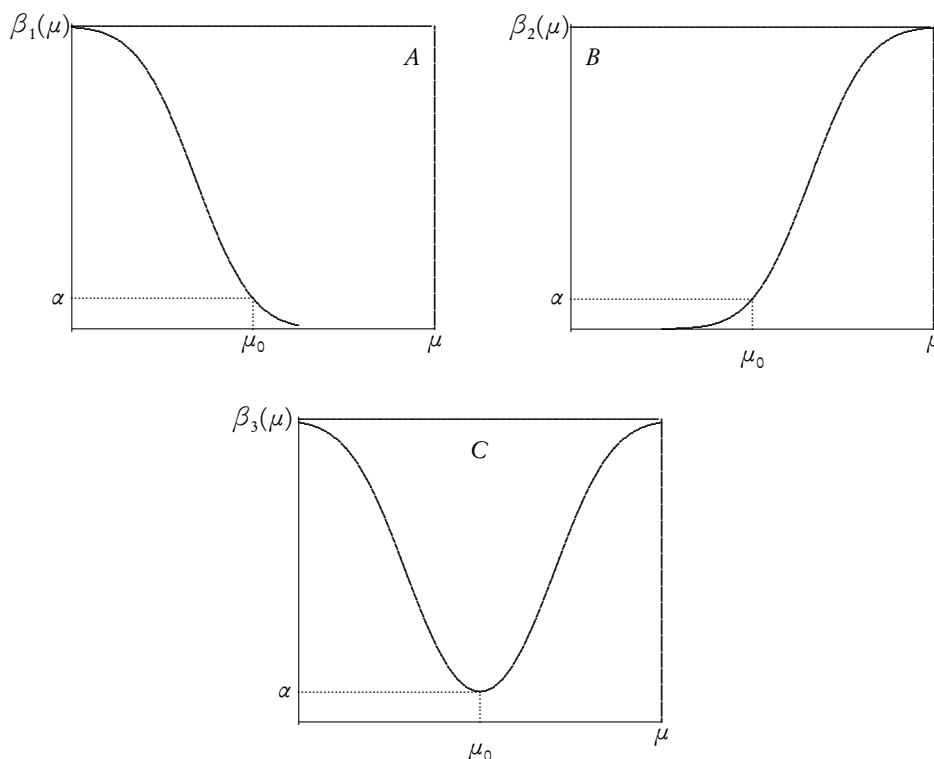
Prawdopodobieństwo odrzucenia hipotezy H , gdy parametr ma wartość μ , $\mu \in \mathcal{R}$, wyraża się wzorem

$$\begin{aligned} P_\mu \left\{ \bar{X}_n < \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \right\} &= P_\mu \left\{ \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} < \frac{\mu_0 - \mu}{\sigma} \sqrt{n} + z_\alpha \right\} \\ &= \Phi \left(\frac{\mu_0 - \mu}{\sigma} \sqrt{n} + z_\alpha \right). \end{aligned}$$

Wartość tego prawdopodobieństwa dla ustalonego μ nazywa się *mocą testu na hipotezie alternatywnej* μ , a funkcja $\mu \rightarrow \beta_1(\mu)$, określona wzorem

$$\beta_1(\mu) = \Phi \left(\frac{\mu_0 - \mu}{\sigma} \sqrt{n} + z_\alpha \right), \quad \mu \in \mathcal{R}, \quad (2.12)$$





Rysunek 2.2. Moc testu (A) D_{α}^{-} , (B) D_{α}^{+} , (C) D_{α}

nazywa się *funkcją mocy* rozważanego testu. Przebieg tej funkcji przedstawia wykres A na rys. 2.2.

W pakiecie R funkcję mocy $\beta_1(\mu)$ oblicza się za pomocą `function(mu, mu_0, sigma, n, alpha) {pnorm(((mu_0 - mu) * sqrt(n)) / sigma + qnorm(alpha))}`
Podobnie $\beta_2(\mu)$ oraz $\beta_3(\mu)$

Mamy oczywiście $\beta_1(\mu_0) = \alpha$, bo to jest test na poziomie istotności α . Rozumując w analogiczny sposób, otrzymujemy moce testów D_{α}^{+} oraz D_{α} ; przedstawiają je, odpowiednio, wykresy B i C na rys. 2.2. Funkcje mocy β_2 i β_3 dla tych testów wyrażają się oczywistymi wzorami

$$\beta_2(\mu) = 1 - \Phi\left(\frac{\mu_0 - \mu}{\sigma} \sqrt{n} + z_{1-\alpha}\right), \quad (2.13)$$

$$\beta_3(\mu) = 1 - \Phi\left(\frac{\mu_0 - \mu}{\sigma} \sqrt{n} + z_{1-\alpha/2}\right) + \Phi\left(\frac{\mu_0 - \mu}{\sigma} \sqrt{n} + z_{\alpha/2}\right). \quad (2.14)$$



Wybór testu w konkretnym problemie zależy od tego, o co w tym problemie chodzi. Przypuśćmy, że od pewnego producenta regularnie kupujemy pewien smakołyk w μ_0 -gramowych opakowaniach. Chcemy sprawdzić, czy nie przepłacamy (czy opakowania nie zawierają mniejszej niż nominalna ilość smakołyku). Po każdym zakupie ważymy opakowanie i otrzymujemy ciąg obserwacji X_1, \dots, X_n . Naturalnym w tej sytuacji jest wybór testu D_α^- : prawdopodobieństwo zakwestionowania uczciwości producenta (tzn. odrzucenia hipotezy $H: \mu = \mu_0$) jest tym większe, im bardziej aktualna średnia zawartość opakowania jest mniejsza od nominalnej. Nie mamy jednak powodu do zmartwienia, gdy $\mu > \mu_0$. Test jest tak skonstruowany, że prawdopodobieństwo zakwestionowania hipotezy maleje ze wzrostem μ . W opisaney sytuacji mówimy o weryfikacji hipotezy $H: \mu = \mu_0$ wobec hipotezy alternatywnej $K: \mu < \mu_0$. Interes producenta jest oczywiście przeciwny: musi on zabezpieczać się przed tym, żeby nie wkładać do opakowania zbyt dużo smakołyku. Producent wybierze więc raczej test D_α^+ , będzie więc weryfikował hipotezę $H: \mu = \mu_0$ wobec hipotezy alternatywnej $K: \mu > \mu_0$.

A jeżeli martwimy się zarówno o interes producenta, jak i konsumenta, albo jeżeli producent i konsument w drodze negocjacji dojdą do porozumienia, zdecydują się zapewne na dwustronny test D_α o mocy przedstawionej na rys. 2.2 (wykres C): test hipotezy $H: \mu = \mu_0$ wobec hipotezy alternatywnej $K: \mu \neq \mu_0$. Jeżeli obie strony dojdą do porozumienia, że test powinien być bardziej czuły i z większym prawdopodobieństwem wykrywać mniejsze odchylenia od wartości nominalnej, mogą zwiększyć moc testu, decydując się na większą liczbę n obserwacji w teście (rys. 2.3).

Można wręcz postawić pytanie o wielkość n próby potrzebnej do wykrycia, z zadanyym z góry prawdopodobieństwem β , odchyłki $k\sigma$ od wartości nominalnej μ_0 . W przypadku testu (2.11) odpowiedzią jest n , które jest rozwiązaniem równania (por. wzór (2.14) dla mocy β_3)

$$1 - \Phi(k\sqrt{n} + z_{1-\alpha/2}) + \Phi(k\sqrt{n} + z_{\alpha/2}) = \beta. \quad (2.15)$$

W pakiecie R moc (2.15) definiuje się następująco:

```
moc <- function(n, k, alpha, beta)
```

```
1 - pnorm(k * sqrt(n) + qnorm(1 - alpha/2))
```

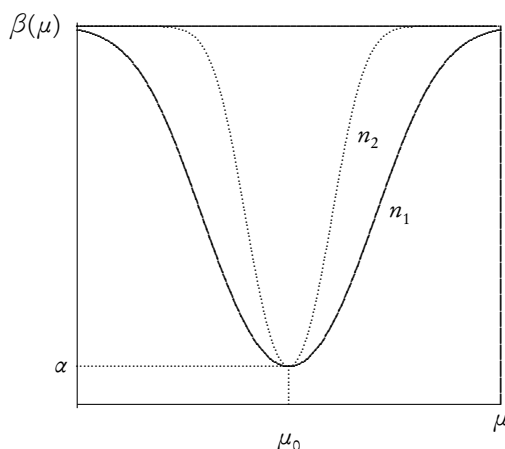
```
+ pnorm(k * sqrt(n) + qnorm(alpha/2)) - beta
```

```
i rozwiązuje równanie, np. dla k = 1, alpha = 0.1, beta = 0.9:
```

```
uniroot(moc, c(1, 100), k = 1, alpha = 0.1, beta = 0.9)$root
```

W zastosowaniach pewien problem stanowi wybór poziomu istotności testu. Można spojrzeć na ten problem nieco inaczej. Zanim dokładnie opiszemy to inne spojrzenie, wprowadzimy pewne uogólnienie tego, co do tej pory mówiliśmy.



Rysunek 2.3. Moc testu D_α dla $n_1 < n_2$

Weryfikujemy hipotezę $H : \mu = \mu_0$. W tym celu definiujemy odpowiednią statystykę $T = T(X_1, \dots, X_n)$ i wybieramy ją w taki sposób, żeby jej duże wartości były traktowane jako argument przeciwko hipotezie. Dla danego poziomu istotności α wyznaczamy taką *wartość krytyczną* $k(\alpha)$, żeby zdarzenie $\{T > k(\alpha)\}$ prowadziło do odrzucenia weryfikowanej hipotezy na poziomie istotności α . Na przykład w teście D_α^- mieliśmy $T = -\bar{X}_n$ oraz $k(\alpha) = -(\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}})$, w teście D_α^+ mieliśmy $T = \bar{X}_n$ oraz $k(\alpha) = \mu_0 + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$, a w dwustronnym teście D_α przyjęliśmy $T = |\bar{X}_n - \mu_0|$ oraz $k(\alpha) = z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}$. Uzmienimy teraz α i zdefiniujemy funkcję $k(\alpha)$, $\alpha \in (0, 1)$, wzorem

$$P_{\mu_0} \{T > k(\alpha)\} = \alpha. \quad (2.16)$$

Wielkość $k(\alpha)$ jest oczywiście wartością krytyczną testu hipotezy $H : \mu = \mu_0$, na poziomie istotności α .

Funkcja $k(\alpha)$ jest malejącą funkcją argumentu α : jeżeli hipoteza zostaje odrzucona na poziomie istotności α , to zostaje również odrzucona na każdym poziomie istotności $\alpha' > \alpha$ (por. rys. 2.4).

Dla danej liczby k , ze wzoru (2.16) można wyznaczyć taką liczbę $\alpha = \alpha(k)$, że

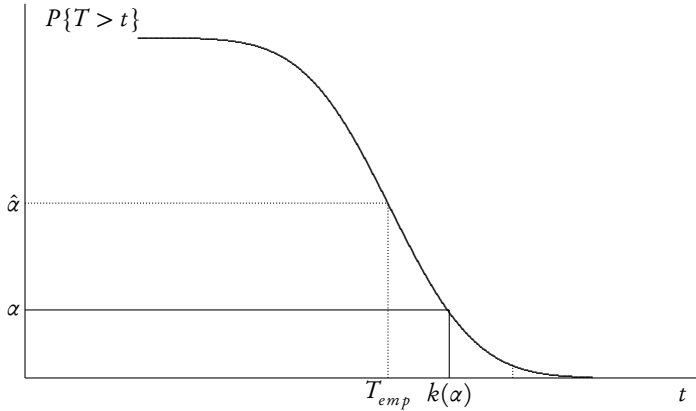
$$P_{\mu_0} \{T > k\} = \alpha(k).$$

Funkcja $k \rightarrow \alpha(k)$ jest oczywiście funkcją odwrotną funkcji $\alpha \rightarrow k(\alpha)$. Jeżeli dla danej próby X_1, \dots, X_n przez $T_{emp} = T(X_1, \dots, X_n)$ oznaczymy wartość statystyki T w tej próbie, to

$$\hat{\alpha} = \alpha(T_{emp})$$

nazywa się *poziomem krytycznym testem*. Używa się także nazwy *empiryczny poziom istotności testu*, czasami *prawdopodobieństwo krytyczne*. W literaturze angiel-





Rysunek 2.4. Wartość krytyczna $k(\alpha)$ i poziom krytyczny $\hat{\alpha}$

skojęzycznej i w angielskojęzycznych pakietach komputerowych jest używana nazwa *p-value*, co w niektórych tekstach polskich jest tłumaczone jako *p-wartość*, ale ponieważ „Polacy nie gęsi”, więc pozostajmy przy „poziomie krytycznym”. Oto niektóre interpretacje poziomu krytycznego $\hat{\alpha}$:

- najmniejszy poziom istotności, przy którym następuje odrzucenie weryfikowanej hipotezy;
- jak mało prawdopodobny jest wynik, który otrzymaliśmy, gdy H jest prawdziwa;
- największy poziom istotności, przy którym jeszcze nie odrzucamy sprawdzanej hipotezy;
- gdyby zaobserwowana wartość statystyki testu była wartością krytyczną, to taki byłby poziom istotności tego testu.

Definicja (2.16) nawiązuje do intuicji związanych z pojęciem poziomu krytycznego testu $\hat{\alpha}$, ale w celu wyznaczenia tej wartości warto zauważyć, że jeżeli t_{emp} jest zaobserwowaną wartością statystyki testowej T w konkretnej próbie, to w tej próbie mamy po prostu $\hat{\alpha} = \hat{\alpha}(t_{emp}) = P_{\mu_0}\{T > t_{emp}\}$. Podkreślmy ponadto, że poziom krytyczny jest związany z konkretnym testem. Jeżeli za pomocą jednego z testów (2.9), (2.10) lub (2.11) testujemy np. hipotezę $H : \mu = 7$ przy znanym $\sigma/\sqrt{n} = 1$ i zaobserwujemy średnią $\bar{x} = 8$, to poziom krytyczny $P_{\mu_0}\{T > t\} = \Phi(1) = 0.84$ w przypadku testu D_{α}^{-} , dla $T = -\bar{X}_n$ oraz $t_{emp} = -\bar{x}$. Natomiast dla dwóch pozostałych testów 0.16 i 0.32.

2.2. Model pomiaru z nieznaną precyzją

2.2.1. Problem i oznaczenia

Podobnie jak w rozdz. 2.1 rozważamy model obserwacji

$$X_j = \mu + \varepsilon_j, \quad j = 1, \dots, n, \quad (2.1)$$



gdzie μ jest ustaloną, nieznaną liczbą oraz $\varepsilon_1, \dots, \varepsilon_n$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie normalnym $N(0, \sigma)$, ale teraz wariancja σ^2 błędów losowych $\varepsilon_1, \dots, \varepsilon_n$ nie jest znana. Nadal oczywiście obserwacja X ma rozkład normalny $N(\mu, \sigma)$, zmienna losowa $(X - \mu)/\sigma$ ma rozkład normalny $N(0, 1)$, ale z tej wiedzy nie mamy żadnego pożytku, gdyż nie znamy σ . Jeżeli za to nieznanne σ podstawimy jakieś oszacowanie $\tilde{\sigma}$, to zmienna losowa $(X - \mu)/\tilde{\sigma}$ nie musi mieć rozkładu normalnego. W celu oszacowania parametru μ (jest to w naszych wykładach zasadnicze zadanie) wykonamy następujący program. Po pierwsze, oszacujemy za pomocą $\tilde{\sigma}$ nieznanne odchylenie standardowe σ , po drugie, wyznaczmy rozkład prawdopodobieństwa zmiennej losowej $(X - \mu)/\tilde{\sigma}$ i w końcu, znając ten rozkład, będziemy, tak jak w rozdz. 2.1, szacowali interesujące nas μ oraz testowali o nim interesujące nas hipotezy.

W modelu ze znanym odchyleniem standardowym i nieznanym parametrem μ rozkład obserwacji oznaczaliśmy przez F_μ . Faktycznie mieliśmy do czynienia z jednoparametrową rodziną rozkładów prawdopodobieństwa. Teraz nasz model jest większy i rozkłady prawdopodobieństwa są indeksowane dwoma parametrami $\mu \in \mathcal{R}$ oraz $\sigma \in \mathcal{R}^+$; dla oznaczania tych rozkładów będziemy używali symbolu $F_{\mu, \sigma}$, a dla wartości oczekiwanej różnych funkcji obserwacji o tym rozkładzie symbolu $E_{\mu, \sigma}$.

2.2.2. Estymacja punktowa σ i σ^2

Przypomnijmy, że z definicji

$$\begin{aligned} \sigma^2 &= \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{+\infty} (x - \mu)^2 \exp\left\{-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right\} dx \\ &= E_{\mu, \sigma}(X - \mu)^2. \end{aligned}$$

Próbkowym odpowiednikiem tego parametru jest

$$S^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2. \quad (2.17)$$

Z elementarnego kursu statystyki matematycznej wiadomo, że jeżeli obserwacje X_1, \dots, X_n mają rozkład normalny $N(\mu, \sigma)$, to zmienna losowa $n S^2 / \sigma^2$ ma rozkład chi-kwadrat o $(n - 1)$ stopniach swobody. Taką zmienną losową o ν stopniach swobody będziemy oznaczali przez χ_ν^2 , gęstość jej rozkładu prawdopodobieństwa przez g_ν , dystrybuantę przez G_ν , a jej kwantyl rzędu α przez $\chi_\nu(\alpha)$ (rys. 2.5).

Zajmiemy się najpierw punktową estymacją odchylenia standardowego σ i wariancji σ^2 . Korzystając ze wzoru dla gęstości $g_\nu(x)$ rozkładu chi-kwadrat o ν



stopniach swobody,

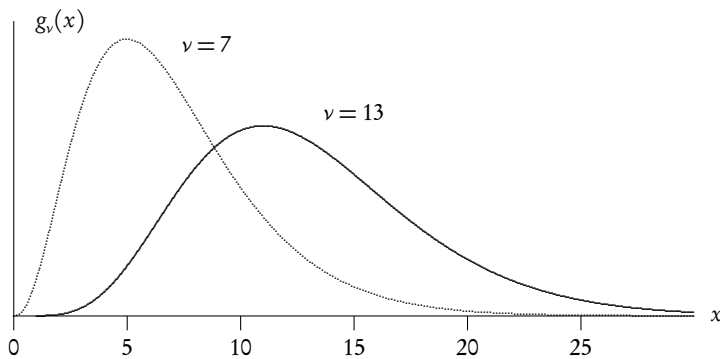
$$g_\nu(x) = \frac{1}{2^{\nu/2} \Gamma(\frac{\nu}{2})} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}}, \quad x > 0 \quad (2.18)$$

i spostrzegając, że

$$\int_0^{+\infty} x^{\frac{\nu}{2}-1} e^{-\frac{x}{2}} = 2^{\nu/2} \Gamma\left(\frac{\nu}{2}\right),$$

dla zmiennej losowej χ_ν^2 łatwo otrzymujemy pożyteczny wzór:

$$E(\chi_\nu^2)^k = 2^k \frac{\Gamma(k + \frac{\nu}{2})}{\Gamma(\frac{\nu}{2})}.$$



Rysunek 2.5. Gęstość rozkładu chi-kwadrat $g_\nu(x)$

W szczególności mamy

$$E(\chi_\nu^2) = \nu, \quad E(\chi_\nu^2)^2 = \nu(\nu + 2), \quad E\sqrt{\chi_\nu^2} = \sqrt{2} \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})}.$$

Stąd wnioskujemy, że

$$E_{\mu, \sigma} S^2 = \frac{n-1}{n} \sigma^2,$$

a więc że S^2 , określone wzorem (2.17), jest obciążonym estymatorem wariancji σ^2 . Obciążenie wynosi $-\sigma^2/n$ i dla dużych licznosci n prób może nie odgrywać istotnej roli. Estymatorem nieobciążonym wariancji σ^2 jest

$$S_0^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2,$$

a estymatorem nieobciążonym odchylenia standardowego σ jest

$$\sqrt{\frac{n}{2}} \frac{\Gamma(\frac{n-1}{2})}{\Gamma(\frac{n}{2})} S. \quad (2.19)$$



W pakiecie R wartość współczynnika we wzorze (2.19) oblicza się za pomocą komendy
`sqrtn/2)*(gamma((n-1)/2)/gamma(n/2))`

W zastosowaniach powstaje pytanie, który z estymatorów S^2 czy S_0^2 wariancji powinien być wykorzystany w danym problemie lub, inaczej, który z nich jest „lepszy”. Nie istnieje żadna odpowiedź na to pytanie, dopóki nie zostanie sformułowane kryterium oceny jakości estymatora. Spójrzmy na to trochę inaczej. Weźmy pod uwagę klasę estymatorów

$$\hat{\sigma}_c^2 = c \sum_{i=1}^n (X_i - \bar{X}_n)^2, \quad c > 0,$$

i za kryterium wyboru najlepszego współczynnika c przyjmijmy ryzyko estymatora definiowane jako średni względny błąd kwadratowy

$$\begin{aligned} R_{\hat{\sigma}_c^2}(\sigma) &= E_\sigma \left(\frac{\hat{\sigma}_c^2 - \sigma^2}{\sigma^2} \right)^2 \\ &= c^2(n^2 - 1) - 2c(n - 1) + 1. \end{aligned}$$

Okazuje się, że estymatorem o jednostajnie najmniejszym średnim względnym błędzie kwadratowym jest $(n + 1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Jak widzimy, istnieje duża swoboda w wyborze estymatora wariancji; jest istotne żeby zawsze jasno określić, który z estymatorów wybraliśmy, a przy studiowaniu literatury i korzystaniu z pakietów komputerowych, który z estymatorów został wybrany przez autora. Jak się za chwilę przekonamy, dla przedziałowej estymacji wariancji wybór współczynnika c nie będzie miał żadnego znaczenia: będziemy korzystali tylko z faktu, że zmienna losowa $\sum_{i=1}^n (X_i - \bar{X}_n)^2 / \sigma^2$ ma rozkład chi-kwadrat.

2.2.3. Estymacja przedziałowa σ i σ^2

Jak już wiemy, jeżeli próba X_1, \dots, X_n pochodzi z rozkładu normalnego $N(0, 1)$, to zmienna losowa nS^2/σ^2 ma rozkład chi-kwadrat o $(n - 1)$ stopniach swobody. Przez $\chi_{n-1}(\gamma)$ oznaczyliśmy kwantyl rzędu γ tego rozkładu. Zatem, dla każdego $\gamma \in (0, 1)$

$$P_{\mu, \sigma} \left\{ \frac{nS^2}{\sigma^2} \leq \chi_{n-1}(\gamma) \right\} = \gamma,$$

więc

$$\left(\frac{nS^2}{\chi_{n-1}(\gamma)}, +\infty \right)$$

jest (jednostronnym) przedziałem ufności dla wariancji σ^2 , na poziomie



ufności γ . Oczywiście jest, że

$$\left(\sqrt{\frac{n}{\chi_{n-1}(\gamma)}} S, +\infty \right)$$

jest przedziałem ufności dla odchylenia średniego σ , na poziomie ufności γ .

Budowa dwustronnego przedziału ufności nie jest już tak jednoznaczna. Ponieważ dla każdego $\alpha, \beta \geq 0$ takich, że $\alpha + \beta = 1 - \gamma$, mamy

$$P_{N(\mu, \sigma)} \{ \chi_{n-1}(\alpha) \leq n S^2 / \sigma^2 \leq \chi_{n-1}(1 - \beta) \} = \gamma,$$

więc otrzymujemy całą rodzinę przedziałów ufności na poziomie ufności γ :

$$\left(\frac{n S^2}{\chi_{n-1}(1 - \beta)}, \frac{n S^2}{\chi_{n-1}(\alpha)} \right), \quad \alpha, \beta \geq 0, \quad \alpha + \beta = 1 - \gamma. \quad (2.20)$$

W elementarnym wykładzie statystyki sugeruje się najczęściej proste rozwiązanie: $\alpha = \beta = (1 - \gamma)/2$. Otrzymuje się wtedy przedział ufności symetryczny w tym sensie, że po jego obu stronach mieści się taka sama masa prawdopodobieństwa. Długość dwustronnego przedziału ufności (2.20) dla ustalonych n oraz γ jest równa $\Delta(\alpha)nS^2$, gdzie

$$\Delta(\alpha) = \frac{1}{\chi_{n-1}(\alpha)} - \frac{1}{\chi_{n-1}(\gamma + \alpha)} \quad (2.21)$$

i najkrótszy przedział ufności otrzymamy wtedy, gdy za α wybierzemy to, które minimalizuje funkcję $\Delta(\alpha)$, $0 \leq \alpha \leq \gamma$. Skutki ilustruje numerycznie poniższa tabelka, w której D oraz G oznaczają, odpowiednio, dolny oraz górny współczynnik $n/\chi_{n-1}(\alpha)$ oraz $n/\chi_{n-1}(\gamma + \alpha)$ przedziału ufności (2.20), natomiast Δ oznacza jego długość; liczba α w ostatniej kolumnie tabelki oznacza optymalną wartość rzędu tego kwantyla.

W pakiecie R w celu obliczenia wartości optymalnego α definiuje się funkcję (2.21)

$\Delta <- \text{function}(\alpha, \gamma, n) \{ 1/qchisq(\alpha, n - 1) - 1/qchisq(\gamma + \alpha, n - 1) \}$
i komendą `nlm(Delta, (1 - gamma)/2, gamma = 0.9, n = 5)$estimate`
wyznacza się α . UWAGA NUMERYKA!

Przedziały ufności dla $\gamma = 0.9$

n	Symetryczny			Najkrótszy			
	D	G	Δ	D	G	Δ	α
5	0.5270	7.0351	6.5081	0.2762	4.7345	4.4583	0.0988
25	0.6865	1.8053	1.1188	0.6083	1.6500	1.0417	0.0837
50	0.7537	1.4736	0.7199	0.7108	1.4055	0.6947	0.0756



2.2.4. Testowanie hipotez o wariancji σ^2

Testowanie, dla ustalonego σ_0 , hipotezy prostej $H : \sigma = \sigma_0$ wobec jednostronnej hipotezy alternatywnej $K : \sigma > \sigma_0$ jest typowe na przykład wtedy, gdy przez dłuższy czas jest używany przyrząd pomiarowy o znanej precyzji charakteryzowanej wariancją σ^2 i w badaniach okresowych mamy sprawdzić, czy jego dokładność nie pogorszyła się; jeżeli w pewnym badaniu kontrolnym stwierdzimy, że dokładność przyrządu uległa pogorszeniu, oddajemy go do naprawy lub zastępujemy innym. Konstrukcja testu nie nastrocza żadnych trudności: hipotezę H odrzucamy na poziomie istotności α , gdy

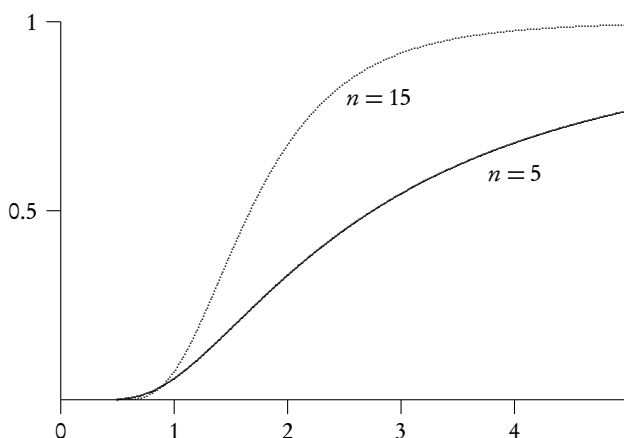
$$S^2 > \frac{\sigma_0^2}{n} \chi_{n-1}(1-\alpha),$$

gdzie jak zwykle $\chi_{n-1}(1-\alpha)$ jest kwantylem rzędu $1-\alpha$ rozkładu chi-kwadrat o $(n-1)$ stopniach swobody. Czasami używa się nazwy *wartość krytyczna rzędu α* , a niektóre tablice statystyczne lub programy komputerowe przedstawiają właśnie te wartości.

Moc rozważanego testu wyraża się wzorem

$$\begin{aligned} \beta(\sigma) &= P_{N(\mu, \sigma)} \left\{ S^2 > \frac{\sigma_0^2}{n} \chi_{n-1}(1-\alpha) \right\} \\ &= P \left\{ \chi_{n-1}^2 > \left(\frac{\sigma_0}{\sigma} \right)^2 \chi_{n-1}(1-\alpha) \right\} \\ &= 1 - G_{n-1} \left(\left(\frac{\sigma_0}{\sigma} \right)^2 \chi_{n-1}(1-\alpha) \right). \end{aligned}$$

Wykres mocy tego testu jednostronnego przedstawia rys. 2.6.

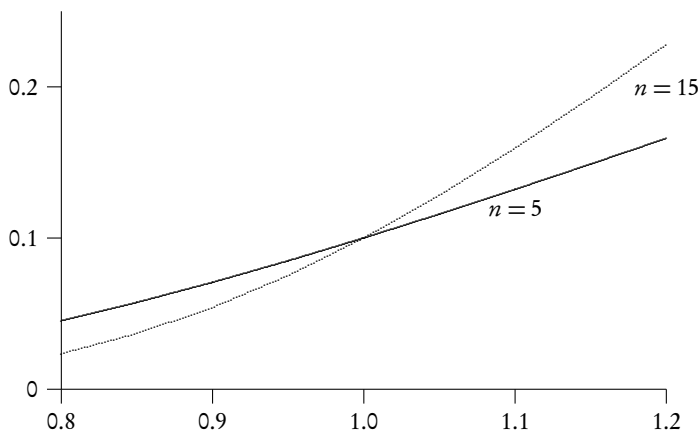


Rysunek 2.6. Moc testu $H : \sigma = 1, K : \sigma > 1$ dla $\alpha = 0.1$



W pakiecie R moc $\beta(\sigma)$ oblicza się za pomocą funkcji
 $\text{moc} <- \text{function}(\sigma, \sigma_0, \alpha, n)$
 $\{1 - \text{pchisq}((\sigma_0/\sigma)^2 * \text{qchisq}(1 - \alpha, n - 1), n - 1)\}$

Wykresy funkcji mocy wszystkich testów hipotezy σ_0 , na poziomie istotności α , przechodzą przez punkt (σ_0, α) ; w naszym przykładzie jest to punkt $(1, 0.1)$ (rys. 2.7).



Rysunek 2.7. Moc testu $H : \sigma = 1, K : \sigma > 1$ dla $\alpha = 0.1$ w otoczeniu H

W przypadku dwustronnej hipotezy alternatywnej $K : \sigma \neq \sigma_0$ pojawia się problem niejednoznaczności, analogiczny do problemu konstrukcji dwustronnego przedziału ufności dla σ . Hipotezę H kwestionujemy, gdy zaobserwowana wariancja próbkowa S^2 zbyt daleko odbiega od wariancji σ_0^2 określonej przez weryfikowaną hipotezę H , tzn. gdy $S^2 < k_1$ lub $S^2 > k_2$, dla k_1, k_2 wybranych tak, żeby test miał założony z góry poziom istotności α , tzn. żeby

$$P_{\sigma_0} \{S^2 < k_1\} + P_{\sigma_0} \{S^2 > k_2\} = \alpha$$

czyli tak, aby

$$G_{n-1} \left(\frac{n}{\sigma_0^2} k_1 \right) + 1 - G_{n-1} \left(\frac{n}{\sigma_0^2} k_2 \right) = \alpha.$$

Istnieje nieskończenie wiele testów spełniających ten warunek: wystarczy wybrać $\alpha_1 \geq 0$ i $\alpha_2 \geq 0$ w taki sposób, żeby $\alpha_1 + \alpha_2 = \alpha$ i zdefiniować

$$k_1 = \frac{\sigma_0^2}{n} \chi_{n-1}(\alpha_1), \quad k_2 = \frac{\sigma_0^2}{n} \chi_{n-1}(1 - \alpha_2).$$

Moc tego testu wyraża się wzorem

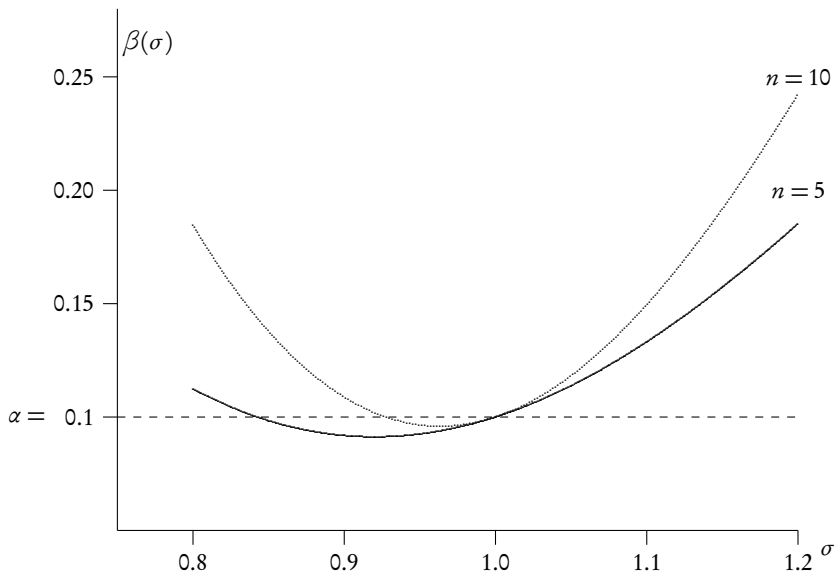


$$\begin{aligned}\beta(\sigma; \alpha_1, \alpha_2) &= P_\sigma \left\{ S^2 < \frac{\sigma_0^2}{n} \chi_{n-1}(\alpha_1) \right\} + P_\sigma \left\{ S^2 > \frac{\sigma_0^2}{n} \chi_{n-1}(1 - \alpha_2) \right\} \\ &= P \left\{ \chi_{n-1}^2 < \left(\frac{\sigma_0}{\sigma} \right)^2 \chi_{n-1}(\alpha_1) \right\} + P \left\{ \chi_{n-1}^2 > \left(\frac{\sigma_0}{\sigma} \right)^2 \chi_{n-1}(1 - \alpha_2) \right\},\end{aligned}$$

czyli

$$\begin{aligned}\beta(\sigma; \alpha_1, \alpha_2) &= G_{n-1} \left(\left(\frac{\sigma_0}{\sigma} \right)^2 \chi_{n-1}(\alpha_1) \right) \\ &\quad + 1 - G_{n-1} \left(\left(\frac{\sigma_0}{\sigma} \right)^2 \chi_{n-1}(1 - \alpha_2) \right)\end{aligned}\quad (2.22)$$

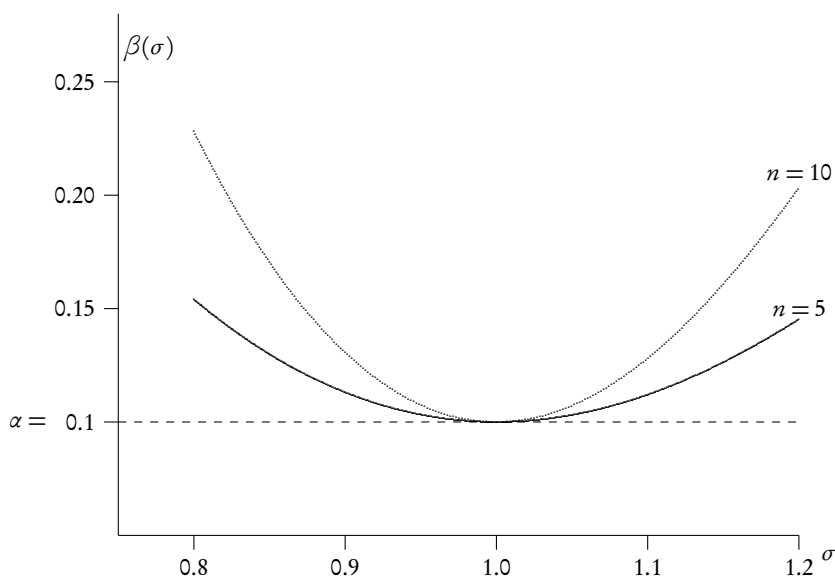
Standardowy wybór „symetryczny” $\alpha_1 = \alpha_2 = \alpha/2$ nie jest w tym przypadku poprawny: moc testu dla niektórych $\sigma \neq \sigma_0$ może być mniejsza niż moc na testowanej hipotezie σ_0 , tzn. mniejsza od założonego poziomu istotności α testu. Oznacza to, że prawdopodobieństwo odrzucenia weryfikowanej hipotezy H , gdy jest ona fałszywa, może być czasami mniejsze niż prawdopodobieństwo jej odrzucenia, gdy jest prawdziwa. Takie testy nazywają się „obciążone”. Przykładem jest moc testu symetrycznego ($\alpha_1 = \alpha_2 = 0.05$) dla weryfikacji $H : \sigma = 1$ wobec $H : \sigma \neq 1$ przedstawiona na rys. 2.8.



Rysunek 2.8. Moc symetrycznego testu hipotezy $H : \sigma = 1$ w otoczeniu hipotezy H

Testami nieobciążonymi na poziomie istotności $\alpha = 0.1$ w naszej sytuacji są testy z parametrami $\alpha_1 = 0.0730$ dla $n = 5$ oraz $\alpha_1 = 0.0658$ dla $n = 10$ (por. rys. 2.9). Parametr α_1 testu nieobciążonego wyznacza się w następujący sposób.





Rysunek 2.9. Moc nieobciążonego testu hipotezy $H: \sigma = 1$ w otoczeniu hipotezy H

Moc (2.22) testu dwustronnego jest ciągłą i różniczkowalną funkcją argumentu σ . Dla uproszczenia zapisów wprowadzimy oznaczenie $t = (\sigma_0/\sigma)^2$. Wtedy moc testu można zapisać w postaci

$$\beta(t; \alpha_1, \alpha_2) = G_{n-1}(t\chi_{n-1}(\alpha_1)) + 1 - G_{n-1}(t\chi_{n-1}(1 - \alpha_2)).$$

Po zróżniczkowaniu względem argumentu t otrzymujemy

$$\beta'(t; \alpha_1, \alpha_2) = \chi_{n-1}(\alpha_1) \cdot g_{n-1}(t\chi_{n-1}(\alpha_1)) - \chi_{n-1}(1 - \alpha_2) \cdot g_{n-1}(t\chi_{n-1}(1 - \alpha_2)).$$

Po przekształceniach otrzymujemy

$$\beta'(t; \alpha_1, \alpha_2) = \frac{t^{(n-3)/2} (\chi_{n-1}(1 - \alpha_2) \text{brigt})^{(n-1)/2}}{2^{(n-1)/2} \Gamma((n-1)/2)} \cdot \left[\left(\frac{\chi_{n-1}(\alpha_1)}{\chi_{n-1}(1 - \alpha_2)} \right)^{(n-1)/2} - \exp \left\{ -\frac{1}{2} t \cdot [\chi_{n-1}(1 - \alpha_2) - \chi_{n-1}(\alpha_1)] \right\} \right].$$

Przyrównując tę pochodną do zera stwierdzamy, że moc testu osiąga minimum dla

$$t = (n-1) \frac{\log \chi_{n-1}(1 - \alpha_2) - \log \chi_{n-1}(\alpha_1)}{\chi_{n-1}(1 - \alpha_2) - \chi_{n-1}(\alpha_1)}.$$

Ponieważ moc naszego testu powinna osiągać minimum dla $\sigma = \sigma_0$, czyli dla



$t = 1$, uwzględniając, że $1 - \alpha_2 = 1 - \alpha + \alpha_1$, otrzymujemy następujące równanie dla α_1

$$(n-1) \frac{\log \chi_{n-1}(1 - \alpha + \alpha_1) - \log \chi_{n-1}(\alpha_1)}{\chi_{n-1}(1 - \alpha + \alpha_1) - \chi_{n-1}(\alpha_1)} = 1. \quad (2.23)$$

Rozwiązanie tego równania w jednym z licznych i powszechnie znanych pakietów komputerowych nie przedstawia żadnej trudności. Przy innym wyborze α_1 i $\alpha_2 = \alpha - \alpha_1$ test jest obciążony.

Rozwiązanie równania (2.23) dla $\alpha = 0.1, n = 10$ w pakiecie R:
`r <- function(a,alfa,n)`
`{(log(qchisq(1-alfa+a,n-1))-log(qchisq(a,n-1)))/`
`(qchisq(1-alfa+a,n-1)-qchisq(a,n-1))-1/(n-1)}`
`uniroot(r,c(0.00001,0.09999),alfa=0.1,n=10)$root`
 UWAGA NUMERYKA!

2.2.5. Estymacja parametru μ

Punktem wyjścia do estymacji parametru μ w modelu gaussowskim ze znaną precyzją, przy wyborze średniej \bar{X}_n jako estymatora, był fakt, że zmienna losowa $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ ma rozkład normalny $N(0, 1)$. Teraz zamiast znanym odchyleniem średnim σ dysponujemy tylko jego oszacowaniem S (2.17). Pojawia się nowa zmienna losowa określona wzorem

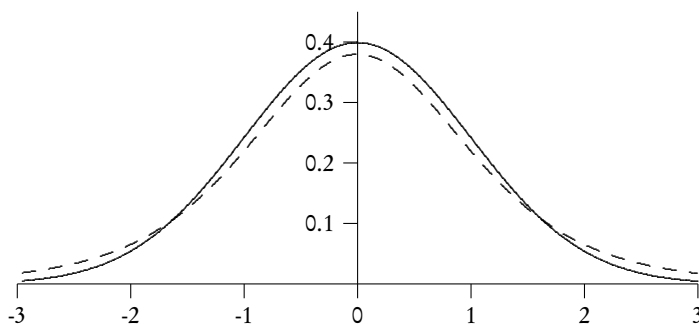
$$t_\nu = \frac{\xi}{\sqrt{\chi_\nu^2/(\nu-1)}},$$

gdzie zmienna losowa ξ ma rozkład normalny $N(0, 1)$, zmienna losowa χ_ν^2 ma rozkład chi-kwadrat o ν stopniach swobody i obie zmienne losowe są niezależne. Rozkład tej zmiennej losowej nazywa się *rozkładem t Studenta o ν stopniach swobody*. W rozważanym teraz przypadku mamy zatem do czynienia ze zmienną losową

$$\frac{\frac{\bar{X}_n - \mu}{\sigma} \sqrt{n}}{\sqrt{\frac{nS^2}{\sigma^2}/(n-1)}} = \frac{\bar{X}_n - \mu}{S} \sqrt{n-1}$$

o rozkładzie t Studenta z $(n-1)$ stopniami swobody. Możliwość wnioskowania o parametrze μ ulega zmianie, bo zmienna losowa $\frac{\bar{X}_n - \mu}{S} \sqrt{n-1}$ o rozkładzie t Studenta jest bardziej rozproszona wokół zera niż zmienna losowa $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ o rozkładzie normalnym (rys. 2.10).





Rysunek 2.10. Rozkład normalny (—) i rozkład t Studenta (- - -)

W naturalny sposób otrzymujemy dla estymowanego parametru μ przedziały ufności na danym poziomie ufności γ : przedział dwustronny

$$\left(\bar{X}_n - t_{n-1} \left(\frac{1+\gamma}{2} \right) \frac{S}{\sqrt{n-1}}, \bar{X}_n + t_{n-1} \left(\frac{1+\gamma}{2} \right) \frac{S}{\sqrt{n-1}} \right) \quad (2.24)$$

i przedziały jednostronne

$$\left(\bar{X}_n + t_{n-1} (1-\gamma) \frac{S}{\sqrt{n-1}}, +\infty \right),$$

$$\left(-\infty, \bar{X}_n + t_{n-1} (\gamma) \frac{S}{\sqrt{n-1}} \right),$$

(por. (2.3), (2.4), (2.5)), gdzie $t_\nu(\alpha)$ jest kwantylem rzędu α rozkładu t Studenta o ν stopniach swobody.

Gdy znaleźliśmy odchylenie standardowe σ , długość przedziału ufności na poziomie ufności γ (2.3) mogliśmy wyrazić wzorem $2 z_{(1+\gamma)/2} \frac{\sigma}{\sqrt{n}}$ i za pomocą odpowiednich manipulacji liczbą n obserwacji w próbie X_1, \dots, X_n mogliśmy uzyskać potrzebną dokładność estymacji parametru μ . Jeżeli teraz nieznanne odchylenie standardowe σ zastąpimy jego oszacowaniem S , to tak obliczona długość przedziału będzie losowa. Problem polega na takim doborze n , żeby ta zmienna losowa nigdy nie przekraczała zadanej z góry liczby $2d$. Znane są różne rozwiązania tego zadania: przedstawię najprostsze i intuicyjnie najbardziej przejrzyste, tzw. *dwuetapową procedurę Steina*.

Idea rozwiązania polega na tym, żeby w pierwszym etapie oszacować nieznanne odchylenie standardowe, a następnie „doszacować” parametr μ tak, żeby uzyskać odpowiednią dokładność jego oszacowania. Najpierw ustalmy dowolnie licznosc $n_1 > 2$ próby wstępnej X_1, \dots, X_{n_1} i policzmy na jej podstawie oszacowanie wariancji

$$S_{n_1}^2 = \frac{1}{n_1} \sum_{i=1}^{n_1} (X_i - \bar{X}_{n_1})^2.$$



Oznaczamy

$$\Delta = t_{n_1-1} \left(\frac{1+\gamma}{2} \right) \frac{S_{n_1}}{\sqrt{n_1-1}}.$$

Jeżeli $\Delta \leq d$, konstrukcja jest zakończona; przedział ufności obliczony na podstawie tej wstępnej próby ma długość nie większą od założonej. Jeżeli tak nie jest, dobieramy dodatkowe $n - n_1$ obserwacji i średnią \bar{X}_n liczymy dla całej rozszerzonej próby X_1, \dots, X_n . Mamy teraz

$$\sqrt{\frac{n_1 S_{n_1}^2}{\sigma^2}} \sim \chi_{n_1-1}^2, \quad \frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} \sim N(0, 1)$$

$$\frac{(\bar{X}_n - \mu) \sqrt{(n_1 - 1)n}}{S_{n_1} \sqrt{n_1}} \sim t_{n_1-1}$$

więc

$$\left(\bar{X}_n - \sqrt{\frac{n_1}{n}} \Delta, \bar{X}_n + \sqrt{\frac{n_1}{n}} \Delta \right)$$

jest przedziałem ufności na poziomie ufności γ . Będzie to przedział ufności o długości co najwyżej d na poziomie ufności γ , jeżeli n wybierzemy w taki sposób, żeby

$$\sqrt{\frac{n_1}{n}} \Delta \leq d,$$

czyli

$$n \geq n_1 \left(\frac{\Delta}{d} \right)^2.$$

W pakiecie R definiuje się funkcję
 $n < -function(n_1, S_{n_1}, \gamma, d)$
 $\{(n_1 * (qt((1 + \gamma)/2, n_1 - 1))^2 * S_{n_1}^2) / ((n_1 - 1) * d^2)\}$
i oblicza n jako $ceiling(n(n_1, S_{n_1}, \gamma, d))$

Odnotujmy, że teraz liczność n próby jest zmienną losową. Można pomyśleć o optymalizowaniu procedury (np. minimalizacji wartości oczekiwanej zmiennej losowej n), ale ta problematyka przekracza ramy naszego wykładu.

2.2.6. Testowanie hipotez o parametrze μ

Punktem wyjścia do testowania hipotez o parametrze μ w modelu gaussowskim ze znaną precyzją był fakt, że zmienna losowa $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ ma rozkład normalny $N(0, 1)$. Teraz, podobnie jak w rozdz. 2.2.5, rozkład normalny zastąpimy rozkładem t Studenta i koncepcyjnie nic się nie zmienia w stosunku do tego, co



mówiliśmy przy okazji modelu gaussowskiego o znanej wariancji σ^2 . Na przykład testując hipotezę $H : \mu = \mu_0$ przy alternatywie $K : \mu < \mu_0$, na poziomie istotności α odrzucimy H , gdy zaobserwujemy zdarzenie

$$\bar{X}_n < \mu_0 + \frac{S}{\sqrt{n-1}} t_{n-1}(\alpha)$$

to znaczy, gdy

$$\left\{ \frac{\bar{X}_n - \mu_0}{S} \sqrt{n-1} < t_{n-1}(\alpha) \right\}.$$

Moc tego testu wyraża się oczywistym wzorem

$$\beta(\mu, \sigma) = P_{N(\mu, \sigma)} \left\{ \frac{\bar{X}_n - \mu_0}{S} \sqrt{n-1} < t_{n-1}(\alpha) \right\}.$$

Zauważmy, że moc nie zależy od nieznanego σ . Jeżeli próba X_1, \dots, X_n pochodzi z rozkładu $N(\mu, \sigma)$, to

$$\frac{\bar{X}_n - \mu_0}{S} \sqrt{n-1} = \frac{\frac{\bar{X}_n - \mu}{\sigma} \sqrt{n} + \frac{\mu - \mu_0}{\sigma} \sqrt{n}}{\sqrt{\frac{nS^2}{\sigma^2} / (n-1)}}$$

jest ilorazem zmiennej losowej o rozkładzie normalnym $N(\frac{\mu - \mu_0}{\sigma} \sqrt{n}, 1)$ i niezależnej od niej zmiennej losowej $\sqrt{\chi_{n-1}^2 / (n-1)}$. Ten iloraz nie jest już zmienną losową o rozkładzie t Studenta. Pojawia się nowa zmienna losowa określona wzorem

$$t_{v, \lambda} = \frac{\xi + \lambda}{\sqrt{\chi_v^2 / v}},$$

gdzie zmienna losowa ξ ma rozkład normalny $N(0, 1)$, zmienna losowa χ_v^2 ma rozkład chi-kwadrat o v stopniach swobody, obie zmienne losowe są niezależne, natomiast λ jest dowolną liczbą rzeczywistą. Rozkład tej zmiennej losowej nazywa się *niecentralnym rozkładem t Studenta o v stopniach swobody i parametrze niecentralności λ* . Dystrybucję tej zmiennej losowej oznaczamy przez $H_{v, \lambda}$, a jej gęstość przez $h_{v, \lambda}$. Jest to bogata rodzina rozkładów, zawierająca także rozkłady niesymetryczne. Przykłady różnych gęstości pokazano na rys. 2.11.

Moc rozważanego testu jednostronnego możemy zapisać w postaci

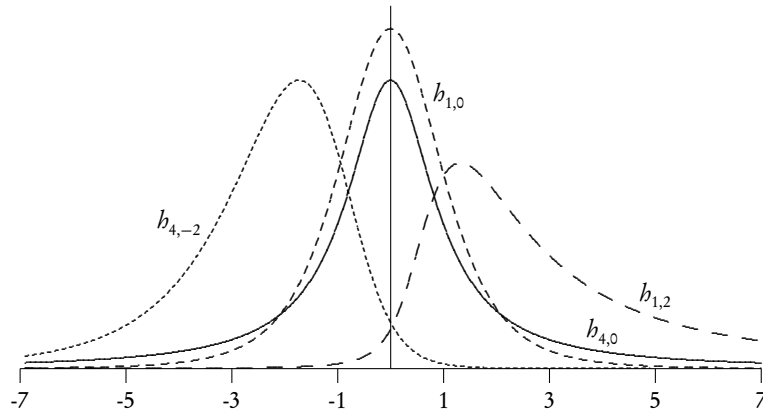
$$\beta(\mu, \sigma) = H_{n-1, \frac{\mu - \mu_0}{\sigma} \sqrt{n}}(t_{n-1}(\alpha)) \quad (2.25)$$

lub w terminach $k = (\mu - \mu_0) / \sigma$ w postaci

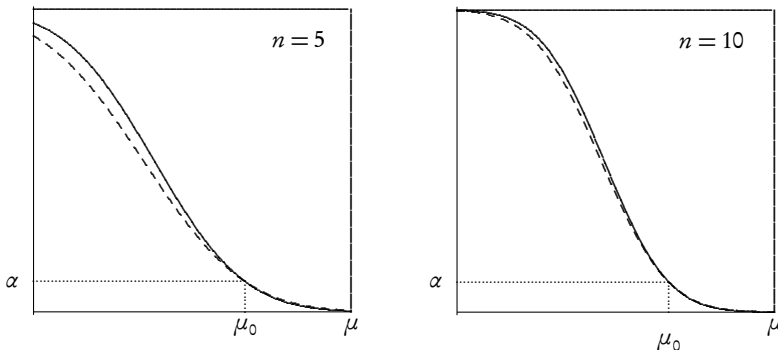
$$\beta(k) = H_{n-1, k \sqrt{n}}(t_{n-1}(\alpha)), \quad k = (\mu - \mu_0) / \sigma, \quad (2.26)$$

co można łatwo liczyć i rysować w standardowych pakietach komputerowych.



Rysunek 2.11. Gęstości $h_{1,0}, h_{4,0}, h_{1,2}, h_{4,-2}$ niecentralnego rozkładu t Studenta

W pakiecie R moc $\beta(k)$ ma postać funkcji
 $\beta < -function(k, \alpha, n)pt(qt(\alpha, n - 1), n - 1, k * sqrt(n))$

Rysunek 2.12. Moc w rozkładzie normalnym (—) i t Studenta (---)

Moc w teście t Studenta jest oczywiście mniejsza niż moc w teście opartym na rozkładzie normalnym (rys. 2.12). W przypadku rozkładu normalnego korzystamy z tego, że znamy wariancję σ^2 , natomiast w teście t Studenta szacujemy ją z obserwacji X_1, \dots, X_n . Dla dużych licznosci n próby różnice mogą okazać się mało istotne, bo wtedy wariancja próbkowa S^2 dobrze przybliża nieznaną wariancję σ^2 .

Test hipotezy $H : \mu = \mu_0$, na poziomie istotności α , przy dwustronnej alternatywie $H : \mu \neq \mu_0$, ma obszar krytyczny

$$\left\{ \bar{X}_n < \mu_0 + t_{n-1} \left(\frac{\alpha}{2} \right) \frac{S}{\sqrt{n-1}} \right\} \cup \left\{ \bar{X}_n > \mu_0 + t_{n-1} \left(1 - \frac{\alpha}{2} \right) \frac{S}{\sqrt{n-1}} \right\},$$



a jego moc wyraża się oczywistym wzorem

$$\beta(\mu, \sigma) = H_{n-1, \frac{\mu - \mu_0}{\sigma} \sqrt{n}} \left(t_{n-1} \left(\frac{\alpha}{2} \right) \right) + 1 - H_{n-1, \frac{\mu - \mu_0}{\sigma} \sqrt{n}} \left(t_{n-1} \left(1 - \frac{\alpha}{2} \right) \right).$$

Pewnym problemem jest to, że moc testu zależy od nieznanego odchylenia standardowego σ . Być może lepiej byłoby od samego początku sformułować weryfikowaną hipotezę w równoważnej postaci $H : \mu - \mu_0 = 0$, a hipotezę alternatywną w postaci $K : (\mu - \mu_0 = k\sigma, k \neq 0)$ i interesować się mocą testu jako funkcją argumentu k . Jest to sprawa wyboru jednostek, w jakich mierzymy obserwacje X ; przy proponowanym podejściu jednostką jest σ . Łatwo jest teraz odpowiedzieć na przykład na pytanie: jak duża ma być próba, aby dla danego $k \neq 0$ moc była równa β : wystarczy rozwiązać względem n równanie

$$H_{n-1, k\sqrt{n}} \left(t_{n-1} \left(\frac{\alpha}{2} \right) \right) + 1 - H_{n-1, k\sqrt{n}} \left(t_{n-1} \left(1 - \frac{\alpha}{2} \right) \right) = \beta.$$

Wtedy, ze względu na monotoniczność (!) funkcji mocy, moc jest nie mniejsza od β dla każdego $k' > k$.



3. Modele parametryczne

3.1. Problem i oznaczenia

Tak jak poprzednio, i jak w całym wykładzie, rozważamy ciąg X_1, \dots, X_n obserwacji postaci

$$X_j = \mu + \varepsilon_j, \quad j = 1, \dots, n,$$

gdzie μ jest ustaloną, nieznaną liczbą oraz $\varepsilon_1, \dots, \varepsilon_n$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie. Ten rozkład i jego dystrybuantę będziemy oznaczali przez F , a jego gęstość prawdopodobieństwa przez f . Wprowadzamy oznaczenie \mathcal{F} dla rodziny wszystkich rozkładów prawdopodobieństwa o ciągłej dystrybuancie, ściśle rosnącej na swoim nośniku. Zakładamy, że rozkład $F \in \mathcal{F}$ jest znany. Wtedy obserwacja X ma rozkład F_μ z jednym nieznanym parametrem μ i zadanie polega na wnioskowaniu (estymacji i weryfikacji hipotez) o tym parametrze na podstawie próby X_1, \dots, X_n .

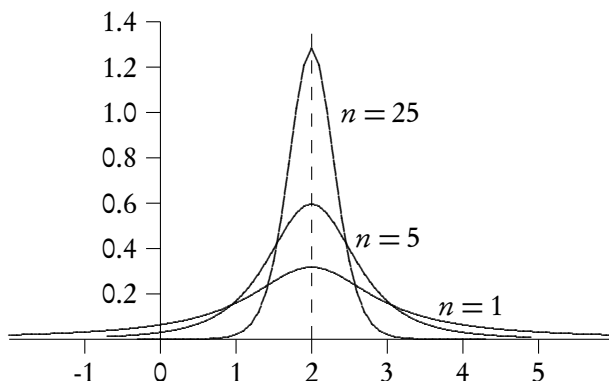
W modelu gaussowskim zakładaliśmy, że błąd ε ma wartość oczekiwaną równą zero, a więc że obserwacja ma wartość oczekiwaną równą μ . Zadanie polegało na wnioskowaniu o wspólnej wartości oczekiwanej niezależnych obserwacji X_1, \dots, X_n . W modelu, którym teraz będziemy się zajmowali, zakładamy, że błąd ε ma medianę równą zero, a więc że obserwacja ma medianę równą μ . Zadanie będzie polegało na wnioskowaniu o wspólnej medianie niezależnych obserwacji X_1, \dots, X_n . Medianę rozkładu F oznaczamy przez $m(F)$.

Dla rozkładów z rozważanej przez nas modelowej rodziny \mathcal{F} mediana zawsze istnieje, jest określona jednoznacznie i ma własność

$$P_F\{X \leq m(F)\} = P_F\{X < m(F)\} = P_F\{X \geq m(F)\} = P_F\{X > m(F)\} = \frac{1}{2}.$$

Można oczekiwać, że w rozważanym teraz modelu mediana z próby może spełniać taką samą rolę, jak średnia z próby w modelu gaussowskim. Zajmiemy się tym szczegółowo w dalszym ciągu wykładu, a teraz pokażemy tylko zachęcającą ilustrację. Na rys. 3.1 przedstawiono gęstości rozkładu mediany M_n z próby X_1, \dots, X_n , pochodzącej z rozkładu Cauchy'ego, który — jak wiadomo — nie ma wartości oczekiwanej.





Rysunek 3.1. Rozkład mediany M_n w modelu z rozkładem Cauchy'ego

Rozważany teraz model zapiszemy formalnie w postaci

$$X_j = \mu + \varepsilon_j, \quad j = 1, \dots, n, \quad \varepsilon \sim F, \quad F \in \mathcal{F}, \quad m(F) = 0; \quad (3.1)$$

(por. model (2.1)). Teraz obserwacja X ma rozkład $F_\mu(x) = F(x - \mu)$ o medianie μ . Rozkład F jest znany. Zadanie polega na wnioskowaniu (estymacja i weryfikacja hipotez) o nieznanym parametrze μ .

Chociaż odejście od rozkładu normalnego sprawia moc trudności technicznych i koncepcyjnych, jest to konieczne ze względu na rozszerzający się zakres zastosowań statystyki matematycznej: w matematyce finansowej, ekologii, ubezpieczeniach typowe rozkłady to rozkłady „o grubszych ogonach”: duże wartości obserwowanych zmiennych losowych pojawiają się częściej, niż by to wynikało z modelu gaussowskiego. Tak jest na przykład wtedy, gdy interesuje nas rozmiar finansowej odpowiedzialności ubezpieczyciela w związku z wypadkami losowymi jego klientów przy ubezpieczeniu OC, AC oraz od wypadków przy pracy; gdy interesuje nas strumień plików przesyłanych w internecie (bardzo dużo małych plików i od czasu do czasu bardzo duży plik), pojemność złóż ropy naftowej, rozmiary osiedli ludzkich lub tzw. zwroty w operacjach giełdowych. W szerokiej klasie \mathcal{F} rozkładów, które teraz musimy brać pod uwagę, znajdują się także rozkłady, dla których średnia \bar{X}_n z próby jest bardziej rozproszona wokół interesującego nas parametru μ , niż pojedyncza obserwacja (rys. 3.2).

W celu dokładniejszego wyjaśnienia tego zjawiska rozpatrzmy jednoparametrową klasę rozkładów α -stabilnych, symetrycznych względem zera. Jest to podklasa rozważanej przez nas klasy \mathcal{F} . Jeżeli zmienna losowa X ma symetryczny rozkład α -stabilny z parametrem $\lambda > 0$, to jej funkcja charakterystyczna wyraża się wzorem

$$E \exp\{itX\} = \exp\{-(\lambda|t|)^\alpha\}, \quad \lambda > 0. \quad (3.2)$$

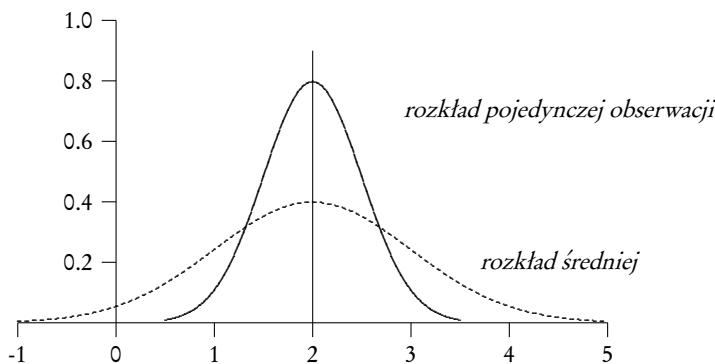
Próba X_1, \dots, X_n jest ciągiem niezależnych zmiennych losowych, takich jak X ,



więc dla funkcji charakterystycznej średniej \bar{X}_n z tej próby mamy

$$\begin{aligned} E \exp\{it\bar{X}_n\} &= E \exp\left(it \frac{1}{n} \sum_{j=1}^n X_j\right) \\ &= \prod_{j=1}^n E \exp\left(i \frac{t}{n} X_j\right) \\ &= \left(\exp\left(-\lambda \frac{|t|}{n}\right)\right)^\alpha \\ &= \exp\{-\lambda(n^{1/\alpha-1})|t|^\alpha\}. \end{aligned}$$

Porównując tę funkcję charakterystyczną z funkcją charakterystyczną (3.2), widzimy, że rozkład średniej arytmetycznej różni się od rozkładu pojedynczej obserwacji tylko parametrem skali. Dla rozkładu normalnego mamy $\alpha = 2$, więc parametr skali rozkładu średniej różni się od parametru skali pojedynczej obserwacji czynnikiem $1/\sqrt{n}$ (co już doskonale wiemy z elementarnej statystyki matematycznej). W przypadku rozkładu Cauchy'ego mamy $\alpha = 1$, więc rozkłady pojedynczej obserwacji i średniej są takie same. Dla $\alpha < 1$ mamy sytuację jak na rys. 3.2: uśrednianie pogarsza dokładność wnioskowania.



Rysunek 3.2. Patologia?

3.2. Statystyki pozycyjne. Rozkład beta

W dalszej części wykładu, zarówno w bieżącym rozdziale o modelach parametrycznych jak i w następnym rozdziale o modelach nieparametrycznych, istotną rolę będą odgrywały statystyki pozycyjne. Wygodnym narzędziem w tych rozważaniach będzie *rozkład beta*.

Zmienna losowa X ma rozkład beta z parametrem (p, q) , $p, q > 0$, jeżeli jej gęstość prawdopodobieństwa wyraża się wzorem

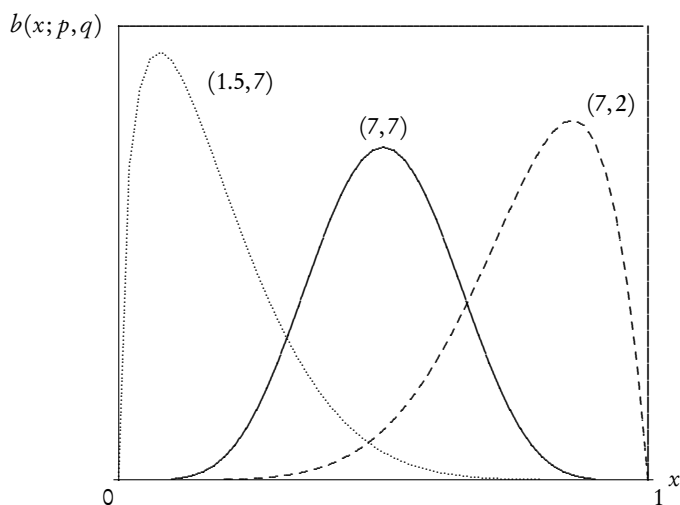
$$b(x; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} x^{p-1}(1-x)^{q-1}, \quad x \in (0, 1), \quad p, q > 0.$$



Dystrybuantę tej zmiennej losowej będziemy oznaczali przez $B(x; p, q)$:

$$B(x; p, q) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} \int_0^x t^{p-1}(1-t)^{q-1} dt; \quad (3.3)$$

jest to znana z analizy matematycznej *niekompletna funkcja beta*. Wartość funkcji odwrotnej w punkcie α , oznaczamy ją przez $B^{-1}(\alpha; p, q)$, jest oczywiście kwantylem rzędu α tego rozkładu (tej zmiennej losowej). Funkcje $b(x; p, q)$, $B(x; p, q)$ oraz $B^{-1}(x; p, q)$ są funkcjami standardowymi w wielu pakietach komputerowych, więc nie będziemy się tu zajmowali problemem „jak to się liczy”. Przykładowe gęstości rozkładu beta pokazano na rys. 3.3.



Rysunek 3.3. Gęstości rozkładu beta (przykłady)

Zauważmy, że dla $p = q$ gęstość jest symetryczna względem środka przedziału $(0, 1)$, więc dla każdego $p > 0$ mamy $B(\frac{1}{2}; p, p) = \frac{1}{2}$ i podobnie dla kwantyla $B^{-1}(\frac{1}{2}; p, p) = \frac{1}{2}$. Odnotujmy jeszcze pożyteczny wzór wiążący rozkład dwumianowy z rozkładem beta:

$$\sum_{j=k}^n \binom{n}{j} x^j (1-x)^{n-j} = B(x; k, n-k+1). \quad (3.4)$$

Dowód tego wzoru polega na obliczeniu przez części całki (3.3), dla całkowitych n oraz k .

Niech X_1, \dots, X_n będzie próbą z rozkładu F . Utwórzmy, jak w rozdz. 2.1.1, ciąg statystyk porządkowych

$$X_{1:n}, \dots, X_{n:n}, \quad X_{1:n} \leq \dots \leq X_{n:n}. \quad (2.6)$$



Łatwo jest wyznaczyć dystrybuantę rozkładu prawdopodobieństwa k -tej statystyki pozycyjnej $X_{k:n}$, gdy próba pochodzi z rozkładu o dystrybuancie $F \in \mathcal{F}$:

$$F_{k,n}(x) = \sum_{j=k}^n \binom{n}{j} F^j(x) (1-F(x))^{n-j}. \quad (3.5)$$

Najprostsze uzasadnienie (wyprowadzenie) tego wzoru jest następujące. Próba X_1, \dots, X_n jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie prawdopodobieństwa z dystrybuantą F . Nawiązując do powszechnie znanego schematu Bernoulliego, nazwijmy sukcesami zdarzenia polegające na tym, że $X_j \leq x$, $j = 1, \dots, n$. Prawdopodobieństwo sukcesu jest oczywiście równe $F(x)$. Zdarzenie polegające na tym, że $X_{k:n} \leq x$ jest równoważne ze zdarzeniem polegającym na tym, że w próbie X_1, \dots, X_n pojawi się co najmniej k sukcesów i stąd wzór (3.5). Gęstość rozkładu prawdopodobieństwa k -tej statystyki pozycyjnej dana jest więc wzorem

$$f_{k,n}(x) = \frac{n!}{(k-1)!(n-k)!} F^{k-1}(x) (1-F(x))^{n-k} f(x).$$

Na mocy (3.4) i (3.5) mamy dla dystrybuanty rozkładu k -tej statystyki pozycyjnej wzór

$$F_{k,n}(x) = B(F(x); k, n-k+1). \quad (3.6)$$

Wykorzystamy później ten wzór do konstrukcji przedziału ufności dla mediany. Ogólniej, można ten wzór wykorzystać do konstrukcji przedziału ufności dla kwantyla dowolnego rzędu $q \in (0, 1)$.

3.3. Estymacja mediany μ

3.3.1. Estymacja punktowa. Obciążenie estymatora

Medianę μ rozkładu obserwacji będziemy szacowali za pomocą mediany M_n z próby X_1, \dots, X_n . Powiedzieliśmy już na ten temat kilka słów w rozdz. 2.1.1, teraz zajmiemy się tym dokładniej. Zgodnie z ogólnie przyjętą umową, medianę M_n z próby definiujemy wzorem

$$M_n = \begin{cases} \frac{1}{2}(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n}), & \text{jeżeli } n \text{ jest parzyste,} \\ X_{\frac{n+1}{2}:n}, & \text{jeżeli } n \text{ jest nieparzyste.} \end{cases} \quad (3.7)$$

Zauważmy że, w przypadku parzystego n liczbą, która dzieli próbę na pół, może być każda liczba postaci $\lambda X_{\frac{n}{2}:n} + (1-\lambda)X_{\frac{n}{2}+1:n}$, $\lambda \in (0, 1)$; wybór $\lambda = \frac{1}{2}$ jest w pełni arbitralny i, jak się później przekonamy, może sprawiać różne kłopoty. Z tego powodu w dalszym wykładzie będziemy wyraźnie rozgraniczali przypadki „ n nieparzyste” i „ n parzyste”.



Zajmiemy się przede wszystkim sprawą obciążoności estymatora M_n . Zwykła definicja, w której estymator T nazywa się estymatorem nieobciążonym parametru θ , jeżeli $E_\theta T = \theta$ dla każdego θ , nie znajduje tutaj zastosowania, bo mediana M_n może nie mieć wartości oczekiwanej. Z kolei nie mamy powodów do ograniczenia rozważanej klasy \mathcal{F} rozkładów o ciągłych i ściśle rosnących dystrybucjach do mniejszej klasy rozkładów, dla których mediana z próby i ewentualnie inne statystyki pozycyjne mają wartość oczekiwaną. Musielibyśmy wtedy wykluczyć m.in. rozkłady z tłustymi ogonami, tak ważne we współczesnych zastosowaniach probabilistyki w finansach, ubezpieczeniach, ekologii itp. Z drugiej strony, całkowita rezygnacja z nieobciążoności estymatora byłaby zbyt daleko posuniętą decyzją: wszak musimy w jakiś sposób kontrolować, jak rozkład danego estymatora T sytuuje się w stosunku do tego, co T ma estymować. Wprowadzimy pojęcie nieobciążoności medianowej; będziemy mówili, że estymator T jest estymatorem medianowo-nieobciążonym (lub nieobciążonym w sensie mediany) parametru θ , jeżeli dla każdego θ jego mediana $Med_\theta T = \theta$. Innymi słowami, T jest estymatorem medianowo-nieobciążonym parametru θ , jeżeli

$$P_\theta\{T \leq \theta\} = P_\theta\{T \geq \theta\} = \frac{1}{2}, \quad \text{dla każdego } \theta$$

(to oczywiście przy założeniu, że podobnie jak rozkład obserwacji X , także rozkład estymatora T ma ciągłą i ściśle rosnącą dystrybucję, a więc jednoznaczną medianę).

Łatwo jest sprawdzić, że jeżeli próba X_1, \dots, X_n ma nieparzystą liczbę elementów n , to mediana M_n z próby jest medianowo-nieobciążonym estymatorem mediany μ rozkładu F_μ obserwacji X . Na mocy (3.5) dystrybuanta mediany M_n wyraża się wzorem

$$P_\mu\{M_n \leq x\} = B\left(F(x - \mu); \frac{n+1}{2}, \frac{n+1}{2}\right), \quad (3.8)$$

więc

$$\begin{aligned} P_\mu\{M_n \leq \mu\} &= B\left(F(0); \frac{n+1}{2}, \frac{n+1}{2}\right) \\ &= B\left(\frac{1}{2}; \frac{n+1}{2}, \frac{n+1}{2}\right) = \frac{1}{2}. \end{aligned} \quad (3.9)$$

W przypadku próby X_1, \dots, X_n o parzystej liczbie elementów mediana M_n , zdefiniowana wzorem (3.7), nie jest medianowo-nieobciążonym estymatorem mediany μ i dla niektórych rozkładów F_μ obserwacji X różnica pomiędzy medianą estymatora M_n i medianą μ może być bardzo duża. Nie podajemy tutaj dowodu tego interesującego faktu.



3.3.2. Estymacja przedziałowa

Rozważamy tylko przypadek nieparzystej liczby obserwacji n w próbie; wtedy rozkład mediany próbkowej jest określony wzorem (3.8). Niech $x_\gamma(M_n)$ będzie kwantylem rzędu γ estymatora M_n , tzn. liczbą taką, że

$$P_\mu\{M_n \leq x_\gamma(M_n)\} = \gamma.$$

Na mocy (3.8) otrzymujemy

$$x_\gamma(M_n) = \mu + F^{-1}\left(B^{-1}\left(\gamma; \frac{n+1}{2}, \frac{n+1}{2}\right)\right)$$

i stąd jednostronny przedział ufności na poziomie ufności γ :

$$\left(M_n - F^{-1}\left(B^{-1}\left(\gamma; \frac{n+1}{2}, \frac{n+1}{2}\right)\right), +\infty\right).$$

W podobny sposób, wychodząc z relacji

$$P_\mu\{|M_n| \leq x_{\frac{1+\gamma}{2}}(M_n)\} = \gamma,$$

otrzymujemy dwustronny przedział ufności na poziomie ufności γ :

$$\left(M_n - F^{-1}\left(B^{-1}\left(\frac{1+\gamma}{2}; \frac{n+1}{2}, \frac{n+1}{2}\right)\right), M_n + F^{-1}\left(B^{-1}\left(\frac{1+\gamma}{2}; \frac{n+1}{2}, \frac{n+1}{2}\right)\right)\right). \quad (3.10)$$

Por. wzór (2.7), gdzie F jest rozkładem normalnym $N(0, \sigma)$.

Opisane konstrukcje przedziałów ufności w modelach z parametrem położenia μ są bardzo ogólne i ze względu na to, że klasa \mathcal{F} rozważanych rozkładów jest bardzo szeroka, znajdują liczne zastosowania. Przykłady zastosowań prezentujemy dalej w rozdz. 3.5.

3.4. Testowanie hipotez o medianie μ

Dla testowania hipotezy $H: \mu = \mu_0$ przy jednostronnej alternatywie $K: \mu > \mu_0$, wyznaczając $x_{1-\alpha}(M_n)$ z równania

$$P_{\mu_0}\{M_n > x_{1-\alpha}(M_n)\} = \alpha,$$

otrzymujemy wartość krytyczną

$$x_{1-\alpha}(M_n) = \mu_0 + F^{-1}\left(B^{-1}\left(1-\alpha; \frac{n+1}{2}, \frac{n+1}{2}\right)\right).$$

Moc tego testu wyraża się wzorem

$$\beta(\mu) = 1 - B\left(F\left[\mu_0 - \mu + F^{-1}\left(B^{-1}\left(1-\alpha; \frac{n+1}{2}, \frac{n+1}{2}\right)\right)\right]; \frac{n+1}{2}, \frac{n+1}{2}\right)$$



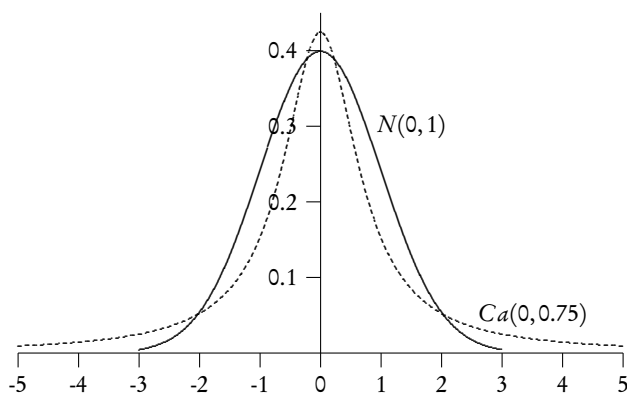
W świetle tego co do tej pory mówiliśmy, konstrukcja testu hipotezy $H : \mu = \mu_0$ przy dwustronnej alternatywie $K : \mu \neq \mu_0$ nie nastęca żadnych nowych problemów koncepcyjnych.

3.5. Zastosowania

3.5.1. Rozkład Cauchy'ego. Parametr położenia i parametr skali

Rozkład Cauchy'ego wspominaliśmy już na początku rozdz. 3. Ilustrowaliśmy tam fakt, że w przypadku rozkładu Cauchy'ego, który nie ma wartości oczekiwanej i którego średnia próbkowa ma taki sam rozkład jak pojedyncza obserwacja, mediana próbkowa może spełniać taką rolę, jaką w przypadku gaussowskim spełnia średnia próbkowa (rys. 3.1).

Pozornie różnica między modelem statystycznym z rozkładem normalnym i modelem statystycznym z rozkładem Cauchy'ego jest niewielka (rys. 3.4).



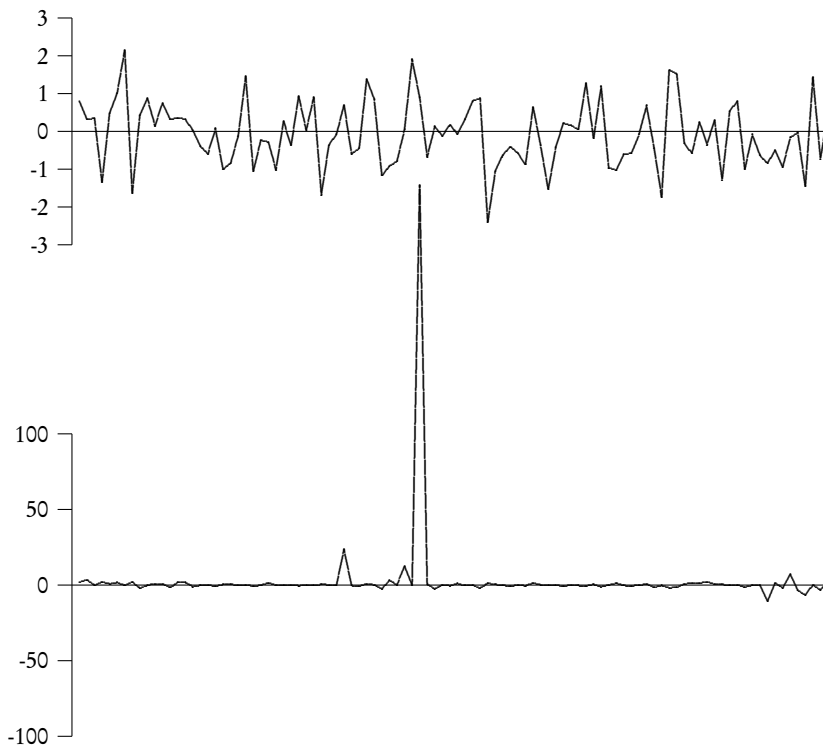
Rysunek 3.4. Gęstość rozkładu normalnego i Cauchy'ego

To jednak, co się dzieje na ogonach obu rozkładów, istotnie zmienia zakres sensownych zastosowań modelu gaussowskiego i modelu Cauchy'ego. Spójrzmy na przykładowe realizacje ciągów niezależnych zmiennych losowych o jednokowym rozkładzie normalnym $N(0, 1)$ i Cauchego $Ca(0, 0.75)$ na rys. 3.5 (rozkłady takie same jak na rys. 3.4. Zwróćmy uwagę na różnice skali na osi rzędnych:

Wydaje się oczywiste, że rozkład Cauchy'ego lepiej nadaje się do modelowania różnych procesów na rynkach finansowych, w ubezpieczeniach oraz w ekologii, czyli tam, gdzie od czasu do czasu mogą pojawiać się jakieś wyjątkowo duże odchylenia. Wybór modelu innowacji w szeregach czasowych oczywiście nie musi być modelem Cauchy'ego; dalej powiemy o jeszcze innych modelach.

Zadanie estymacji mediany rozkładu Cauchy'ego oraz zadanie weryfikacji hipotez o medianie rozwiązujemy tak, jak to w ogólnym przypadku dowolnego, ale znanego rozkładu F , przedstawiliśmy w rozdz. 3.3 i 3.4.





Rysunek 3.5. Szereg czasowy $N(0,1)$ oraz $Ca(0,0.75)$

W przypadku rozkładu Cauchy'ego ze znanym parametrem skali λ i nieznanym parametrem położenia μ , o dystrybuancie i kwantylach postaci

$$F_{\mu,\lambda}(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg}\left(\frac{x - \mu}{\lambda}\right), \quad F_{\mu,\lambda}^{-1}(\gamma) = \mu + \lambda \operatorname{tg}\left(\left(\gamma - \frac{1}{2}\right)\pi\right),$$

otrzymujemy dwustronny przedział ufności na poziomie ufności γ :

$$(M_n - \lambda C(\gamma, n), M_n + \lambda C(\gamma, n)),$$

gdzie

$$C(\gamma, n) = \operatorname{tg}\left(\left(B^{-1}\left(\frac{1+\gamma}{2}; \frac{n+1}{2}, \frac{n+1}{2}\right) - \frac{1}{2}\right)\pi\right).$$

Obliczanie tych granic realizuje się łatwo w każdym standardowym pakiecie komputerowym.

Jeżeli rozpatrujemy rozkład Cauchy'ego jako rozkład innowacji w szeregach czasowych (por. rys. 3.5), to zwykle przyjmujemy, że mediana tego rozkładu jest równa zero (rozkład jest symetryczny względem zero), a zmienność rozważanego procesu jest opisywana za pomocą parametru skali tego rozkładu. Dystrybu-



anta innowacji X wyraża się wtedy wzorem

$$F_\lambda(x) = \frac{1}{2} + \frac{1}{\pi} \operatorname{arctg}\left(\frac{x}{\lambda}\right). \quad (3.11)$$

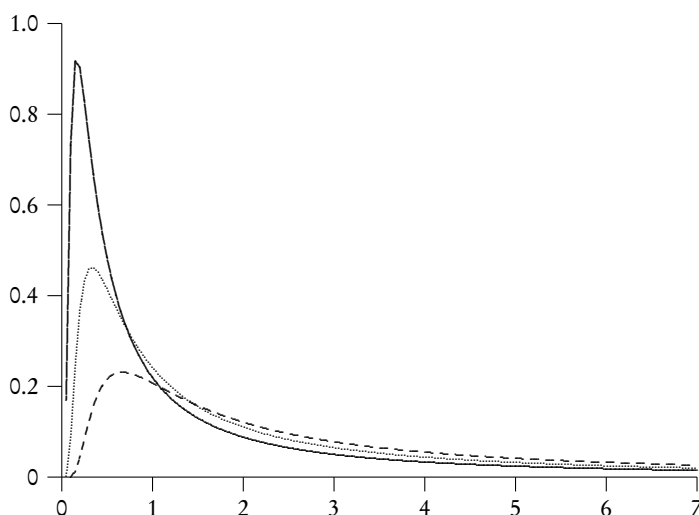
Problem estymacji parametru skali λ może być częściowo rozwiązany za pomocą przedstawionych wyżej narzędzi. Mianowicie, jeżeli zmienna losowa X ma rozkład (3.11), to zmienna losowa $|X|$ ma rozkład o dystrybuancie $(2/\pi) \operatorname{arctg}(x/\lambda)$, którego mediana jest równa λ . Mediana $M_n = |X|_{\frac{n+1}{2}:n}$ z próby $|X|_1, \dots, |X|_n$ jest medianowo nieobciążonym estymatorem parametru λ . Przedstawiona wyżej konstrukcja przedziału ufności i testów dla parametru położenia μ wymaga odpowiedniej modyfikacji w przypadku parametru skali λ .

3.5.2. Rozkład Levy'ego. Parametr skali

Oprócz rozkładu normalnego ($\alpha = 2$) i rozkładu Cauchy'ego ($\alpha = 1$), rozkład Levy'ego ($\alpha = \frac{3}{2}$) jest jednym z trzech rozkładów α -stabilnych, dla których gęstość prawdopodobieństwa wyraża się w jawny sposób łatwym wzorem. Ma ona postać

$$f_\lambda(x) = \sqrt{\frac{\lambda}{2\pi}} x^{-3/2} e^{-\lambda/2x}, \quad \lambda > 0, x > 0.$$

W odróżnieniu od obu wspomnianych rozkładów, nośnikiem tego rozkładu jest dodatnia półoś (rys. 3.6). Rozkład Levy'ego, obok takich rozkładów jak rozkład gamma, może modelować, na przykład, wielkości roszczeń w ubezpieczeniach.



Rysunek 3.6. Gęstość rozkładu Levy'ego $\lambda = 0.5$ (—), $\lambda = 1$ (.....), $\lambda = 2$ (- -)



Dystrybuanta tego rozkładu wyraża się wzorem

$$F_\lambda(x) = 2 \left(1 - \Phi \left(\sqrt{\frac{\lambda}{x}} \right) \right).$$

Dla funkcji kwantylowej $F^{-1}(q)$ łatwo otrzymujemy

$$F_\lambda^{-1}(q) = \frac{\lambda}{z_{1-q/2}^2}.$$

Dla mediany mamy więc

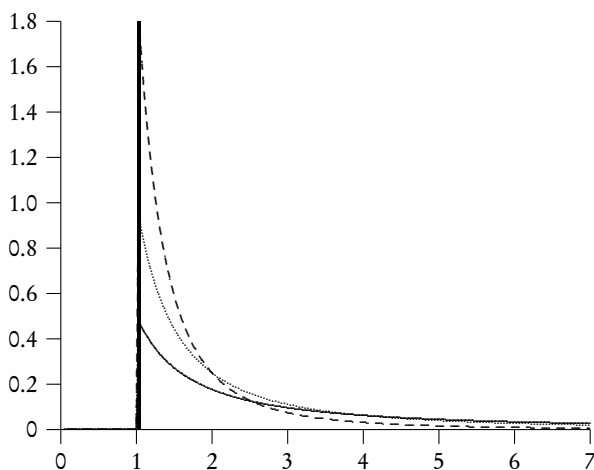
$$F_\lambda^{-1}\left(\frac{1}{2}\right) = 2.1981\lambda,$$

zatem $M_n/2.1981$ jest medianowo nieobciążonym estymatorem parametru skali λ .

3.5.3. Rozkład Pareto. Parametr kształtu

Standardowy rozkład Pareto ma gęstość prawdopodobieństwa oraz dystrybuantę określone wzorami

$$f_\alpha(x) = \alpha \left(\frac{1}{x}\right)^{\alpha+1}, \quad F_\alpha(x) = 1 - \left(\frac{1}{x}\right)^\alpha, \quad x \geq 1.$$



Rysunek 3.7. Gęstość rozkładu Pareto $\lambda = 0.5$ (-.-), $\lambda = 1$ (.....), $\lambda = 2$ (—)

Rozkład ten bywa stosowany w naukach społecznych, geofizyce i aktuariacie. Poza ekonomią jest czasem nazywany rozkładem Bradforda. Pareto oryginalnie używał tego rozkładu do opisu alokacji dóbr w społeczeństwie, gdyż, jak zauważył, większa część bogactwa dowolnego społeczeństwa jest w posiadaniu



niewielkiego procenta jego członków. W informatyce modeluje losowe wielkości plików przesyłanych w internecie (dużo małych plików, mało dużych plików) oraz losowe czasy wykonywania procesu obliczeniowego przez superkomputer (niewiele długich procesów, dużo krótkich). Tutaj zajmujemy się nim, bo jest interesującym przykładem medianowo nieobciążonej estymacji parametru kształtu α . Funkcja kwantylowa tego rozkładu ma postać

$$F_{\alpha}^{-1}(q) = (1 - q)^{-1/\alpha},$$

więc mediana jest równa $\sqrt[\alpha]{2}$. Korzystając z faktu, że mediana próbkowa M_n z próby o nieparzystej liczności n jest medianowo nieobciążonym estymatorem mediany rozkładu, z którego pochodzi próba, otrzymujemy

$$\frac{1}{2} = P_{\alpha}\{M_n \leq \sqrt[\alpha]{2}\} = P_{\alpha}\left\{\alpha \leq \frac{\log 2}{\log M_n}\right\},$$

skąd wynika, że $\log 2 / \log M_n$ jest medianowo nieobciążonym estymatorem parametru kształtu α .



4. Modele nieparametryczne

4.1. Problem i oznaczenia

W dalszym ciągu zajmujemy się ciągiem X_1, \dots, X_n obserwacji postaci

$$X_j = \mu + \varepsilon_j, \quad j = 1, \dots, n,$$

gdzie μ jest ustaloną, nieznaną liczbą oraz $\varepsilon_1, \dots, \varepsilon_n$ jest ciągiem niezależnych zmiennych losowych o jednakowym rozkładzie $F \in \mathcal{F}$. W tym rozdziale nie zakładamy, że rozkład $F \in \mathcal{F}$ jest znany; zakładamy jedynie (jak do tej pory w całym wykładzie), że jego mediana jest równa zero, tak że obserwacja X ma rozkład $F_\mu(x) = F(x - \mu)$ o nieznanym kształcie F i nieznanym medianie μ . Tak jak w całym naszym wykładzie, zadanie polega na estymacji mediany μ i testowaniu o niej hipotez. Teraz o rozkładzie F , z którego pochodzą obserwacje X_1, \dots, X_n , wiemy tak mało, że jedyny sposób wnioskowania na jego temat polega na zastąpieniu go rozkładem empirycznym i budowanie estymatorów i testów w oparciu o ten rozkład. Rozkład empiryczny rozkładu F będziemy oznaczali przez F_n . Formalnie definiujemy go wzorem

$$F_n(x) = \frac{1}{n} \sum_{j=1}^n 1_{(-\infty, x]}(X_j),$$

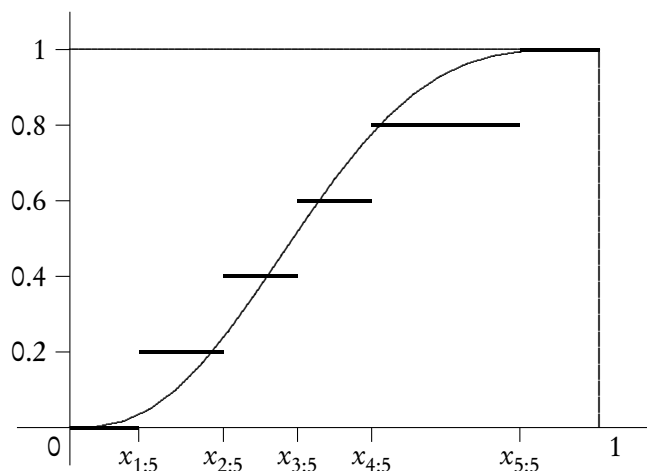
gdzie

$$1_{(-\infty, x]}(X_j) = \begin{cases} 1, & \text{jeżeli } X_j \leq x, \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Jest to funkcja schodkowa o skokach wysokości $1/n$ w punktach $X_{j:n}$ (rys. 4.1).

Istnieje oczywiście wzajemnie jednoznaczna odpowiedniość między dystrybuantą empiryczną F_n a ciągiem $X_{1:n}, \dots, X_{n:n}$ statystyk pozycyjnych. Wnioskowanie o parametrze μ będziemy opierali na tych statystykach.



Rysunek 4.1. Dystrybuanta F i dystrybuanta empiryczna F_5

4.2. Estymacja mediany μ

4.2.1. Estymacja punktowa mediany μ

Jeżeli liczba n elementów w próbie jest nieparzysta, to za estymator mediany μ przyjmujemy, zgodnie z definicją (3.7), $M_n = X_{\frac{n+1}{2}:n}$. Jest to estymator medianowo nieobciążony; dowód jest identyczny jak w (3.9), gdyż nie korzystaliśmy tam z faktu, że rozkład F jest znany.

Jeżeli liczba n elementów w próbie jest parzysta, to estymator M_n , określony wzorem (3.7), jest obciążony. W modelach parametrycznych ze znanym rozkładem F mogliśmy zbadać obciążenie estymatora i ewentualnie dokonać jego modyfikacji na drodze badań symulacyjnych. Teraz nie mamy takiej możliwości i musimy się liczyć „z najgorszym”. A to najgorsze wygląda w następujący sposób. Niech $\text{Med}_F(M_n)$ oznacza medianę rozkładu estymatora M_n , gdy próba X_1, \dots, X_n pochodzi z rozkładu F , i niech $m(F)$ oznacza medianę rozkładu F (wprowadziliśmy już dawniej to oznaczenie). Można udowodnić, że dla każdej, dowolnie dużej liczby $C > 0$ w rozważanej klasie rozkładów \mathcal{F} znajdzie się rozkład F taki, że $|\text{Med}_F(M_n) - m(F)| > C$. Medianowe obciążenie estymatora $M_n = \frac{1}{2}(X_{\frac{n}{2}:n} + X_{\frac{n}{2}+1:n})$ można usunąć przez dodatkową randomizację: wystarczy za M_n przyjąć jedną z dwóch, wybranych losowo, z jednakowym prawdopodobieństwem, statystyk pozycyjnych $X_{\frac{n}{2}:n}$ lub $X_{\frac{n}{2}+1:n}$.

4.2.2. Przedział ufności dla mediany μ

Zacznijmy od konstrukcji przedziału ufności dla kwantyla $x_q = F^{-1}(q)$ dowolnego rzędu $q \in (0, 1)$; wtedy przedział ufności dla mediany będzie szczególnym



przypadkiem dla $q = \frac{1}{2}$. Zajmiemy się najpierw przedziałem jednostronnym postaci $(X_{i:n}, +\infty)$. Skoro ma to być przedział ufności na zadanym poziomie ufności γ , to powinniśmy wybrać wskaźnik $i \in \{1, 2, \dots, n\}$ w taki sposób, żeby $P_F\{X_{i:n} \leq x_q\} \geq \gamma$ dla każdego $F \in \mathcal{F}$. Ponieważ $X_{i:n} < X_{j:n}$, gdy $i < j$, więc rozsądnie jest oczywiście wybrać największą liczbę $i = i(n, \gamma)$ spełniającą podany warunek. Korzystając z wzoru (3.5) na dystrybuantę i -tej statystyki pozycyjnej z próby X_1, \dots, X_n , mamy

$$\begin{aligned} P_F\{X_{i:n} \leq x_q\} &= P_F\{X_{i:n} \leq F^{-1}(q)\} \\ &= \sum_{j=i}^n \binom{n}{j} (F(F^{-1}(q)))^j (1 - F(F^{-1}(q)))^{n-j} \\ &= \sum_{j=i}^n \binom{n}{j} q^j (1 - q)^{n-j}. \end{aligned}$$

Zatem rozwiązaniem jest największe $i = i(n, q)$ takie, że

$$\sum_{j=i(n, \gamma)}^n \binom{n}{j} q^j (1 - q)^{n-j} \geq \gamma.$$

Przedział ufności na poziomie ufności γ dla kwantyla rzędu $q \in (0, 1)$ nie zawsze istnieje; istnieje tylko wtedy, gdy

$$\sum_{j=1}^n \binom{n}{j} q^j (1 - q)^{n-j} \geq \gamma,$$

tzn. gdy $(1 - q)^n \leq 1 - \gamma$: na podstawie zbyt małej próby nie da się wyznaczyć przedziału ufności na zbyt dużym poziomie ufności.

Jako wniosek otrzymujemy jednostronny przedział ufności dla mediany $(X_{i:n}, +\infty)$, gdzie $i = i(n, \frac{1}{2}) \in \{1, \dots, n\}$ jest największą liczbą taką, że

$$2^{-n} \sum_{s=i(n, \gamma)}^n \binom{n}{s} \geq \gamma.$$

Ze względu na dyskretność rozkładu faktyczny poziom ufności

$$\gamma^* = 2^{-n} \sum_{j=i(n, \gamma)}^n \binom{n}{j}$$

może oczywiście być większy od założonego γ . Dla $\gamma = 0.9$ i kilku wybranych licznosci n próby X_1, \dots, X_n , ilustruje to poniższa tabelka:



n	$i(n, \gamma)$	γ^*
5	1	0.969
10	3	0.945
20	7	0.942
50	20	0.940

Dwustronny przedział ufności konstruujemy w postaci $(X_{i:n}, X_{j:n})$. Mamy

$$\begin{aligned} P_F\{X_{i:n} \leq F^{-1}(q) \leq X_{j:n}\} &= P_F\{X_{i:n} \leq F^{-1}(q)\} - P_F\{X_{j:n} > F^{-1}(q)\}, \\ &= \sum_{s=i}^{j-1} \binom{n}{s} q^s (1-q)^{n-s} \end{aligned}$$

i powstaje problem wyboru wskaźników (i, j) w taki sposób, żeby

$$\sum_{s=i}^{j-1} \binom{n}{s} q^s (1-q)^{n-s} \geq \gamma. \quad (4.1)$$

Z problemem niejednoznaczności dwustronnego przedziału ufności spotkaliśmy się już w rozdz. 2 przy okazji konstrukcji przedziału ufności dla wariancji w modelu gaussowskim. Tutaj naturalną miarą długości przedziału ufności jest różnica $j - i$. Konstrukcję najkrótszego przedziału ufności dla mediany zilustrujemy przykładem próby o licznosci $n = 7$; łatwo wyłoni się stąd ogólna, niezbyt trudna dla komputerowego zaprogramowania, konstrukcja.

We wzorze (4.1) wprowadźmy oznaczenie $bin(s; n, q) = \binom{n}{s} q^s (1-q)^{n-s}$. Poniższa tabelka podaje wartości $bin(s; 7, \frac{1}{2})$:

s	$bin(s; 7, \frac{1}{2})$
0	0.0078
1	0.0547
2	0.1641
3	0.2734
4	0.2734
5	0.1641
6	0.0547
7	0.0078

Wynika z niej na przykład, że prawdopodobieństwo pokrycia mediany μ przez przedział ufności $(X_{2:7}, X_{5:7})$ wynosi $0.1641 + 0.2734 + 0.2734 = 0.7109$. Dla pozostałych możliwych przedziałów ufności otrzymujemy prawdopodobieństwa pokrycia tak, jak to pokazuje kolejna tabelka:



$(X_{3:7}, X_{4:7})$	0.2734
$(X_{4:7}, X_{5:7})$	0.2734
$(X_{3:7}, X_{5:7})$	0.5468
$(X_{2:7}, X_{5:7})$	0.7109
$(X_{3:7}, X_{6:7})$	0.7109
$(X_{2:7}, X_{6:7})$	0.8750
$(X_{1:7}, X_{6:7})$	0.9297
$(X_{2:7}, X_{7:7})$	0.9297
$(X_{7:7}, X_{7:7})$	0.9844

W konsekwencji, dla przykładowego $n = 7$ oraz wybranych poziomów ufności γ , otrzymujemy przedziały ufności dla mediany oraz faktyczny poziom ufności γ^* :

γ	Przedział ufności	γ^*
0.90	(X_1, X_6) lub (X_2, X_7)	0.9297
0.95	(X_1, X_7)	0.9844
0.99	nie ma	—

4.3. Testowanie hipotez o medianie μ

Weryfikujemy hipotezę $H : \mu = \mu_0$ na podstawie próby X_1, \dots, X_n z rozkładu F_μ , o którym nie wiemy nic więcej ponad to, że ma ciągłą i ściśle rosnącą dystrybuantę. Równoważne sformułowanie hipotezy ma postać $H : F_\mu(\mu_0) = \frac{1}{2}$. Z takiego sformułowania weryfikowanej hipotezy wynika w sposób naturalny następująca metoda jej testowania. Przekształćmy próbę X_1, \dots, X_n w próbę Y_1, \dots, Y_n za pomocą wzorów

$$Y_j = 1_{[-\infty, \mu_0]}(X_j) = \begin{cases} 1, & \text{jeżeli } X_j \leq \mu_0, \\ 0, & \text{w przeciwnym przypadku.} \end{cases}$$

Jeżeli próba X_1, \dots, X_n pochodzi z rozkładu F_μ , to zero-jedynkowa próba Y_1, \dots, Y_n może być traktowana jak próba z rozkładu dwumianowego, w którym prawdopodobieństwo jedynki jest

$$\theta = F_\mu(\mu_0). \quad (4.2)$$

Zadanie weryfikacji hipotezy o medianie zmiennej losowej X zamienia się w równoważne zadanie weryfikacji hipotezy o prawdopodobieństwie sukcesu θ w rozkładzie dwumianowym.

Jeżeli weryfikujemy $H : \mu = \mu_0$ przy jednostronnej hipotezie alternatywnej $K : \mu < \mu_0$, to równoważne zadanie dla rozkładu dwumianowego polega na weryfikacji hipotezy $H' : \theta = \frac{1}{2}$ przy jednostronnej hipotezie alternatywnej



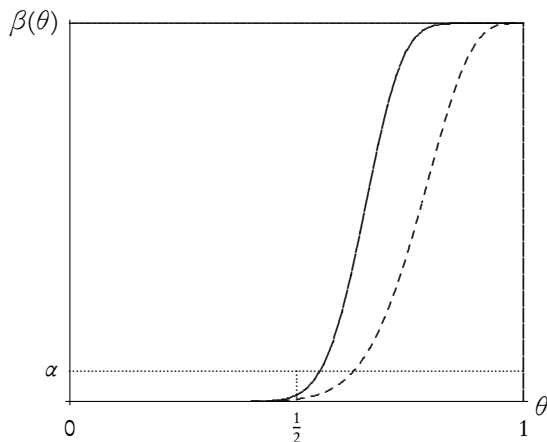
$K' : \theta > \frac{1}{2}$. Tu i dalej symbole primowane dotyczą problemu testowania hipotez o parametrze θ i są odpowiednikami symboli w problemie testowania hipotez o parametrze μ . Odpowiednią (w bardziej zaawansowanych wykładach statystyki matematycznej dowodzi się, że faktycznie jedyną) statystyką dla testowania hipotezy H' jest liczba sukcesów $S_n = \sum_{j=1}^n Y_j$, a zdarzeniem losowym, które na zadanym poziomie istotności α dyskwalifikuje hipotezę H' , jest $\{S_n > k(\alpha)\}$, gdzie $k(\alpha)$ jest liczbą wybraną w taki sposób, żeby prawdopodobieństwo odrzucenia H' , gdy jest ona prawdziwa, nie przekraczało α :

$$P_{H'}\{S_n > k(\alpha)\} \leq \alpha. \quad (4.3)$$

Moc $\beta'(\theta)$ tego testu wyraża się wzorem

$$\beta'(\theta) = P_{\theta}\{S_n > k(\alpha)\}.$$

Jeżeli $k_1 > k_2$, to $P_{\theta}\{S_n > k_1\} < P_{\theta}\{S_n > k_2\}$, więc dla uzyskania możliwie największej mocy testu, $k(\alpha)$ powinno być najmniejszą liczbą spełniającą (4.3). Ponieważ rozkład dwumianowy jest dyskretny, może nie istnieć takie $k(\alpha)$, żeby $P_{H'}\{S_n > k(\alpha)\} = \alpha$. Na przykład dla próby o liczebności $n = 50$, dla wartości krytycznej testu na poziomie istotności $\alpha = 0.05$ mamy $k(0.05) = 32$, przy czym faktyczny poziom istotności testu wynosi 0.0325. Podobnie dla $n = 20$ mamy $k(0.05) = 15$ i faktyczny poziom istotności 0.0206. Wykresy mocy obu testów przedstawia rys. 4.2.



Rysunek 4.2. Moc testu $H' : \theta = \frac{1}{2}, K' : \theta > \frac{1}{2}$

Oryginalna hipoteza $H : \mu = \mu_0$ dotyczy zmiennej losowej X o nieznanym rozkładzie F_{μ} z medianą μ . W terminach oryginalnej zmiennej losowej X obszar krytyczny testu ma postać $\{\sum_{j=1}^n 1_{[\mu_0, +\infty)}(X_j) > k(\alpha)\}$, więc moc tego testu



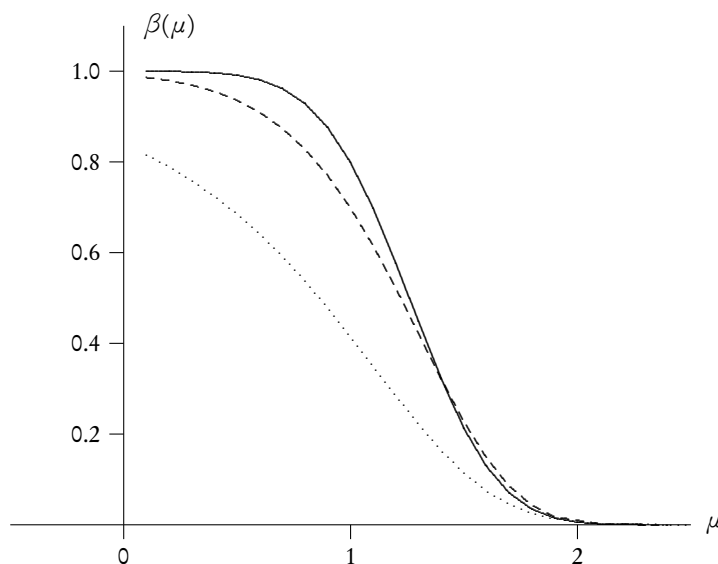
w terminach mediany μ wyraża się wzorem

$$\beta(\mu) = P_{\mu} \left\{ \sum_{j=1}^n 1_{[\mu_0, +\infty)}(X_j) > k(\alpha) \right\}, \quad (4.4)$$

co jednak zależy nie tylko od mediany μ , ale także od nieznanego rozkładu F obserwacji X . Na mocy (4.2) moc testu hipotezy $H : \mu = \mu_0$ przy alternatywie $H : \mu < \mu_0$ może być przedstawiona w postaci

$$\beta(\mu) = \beta'(F_{\mu}(\mu_0)).$$

Dla trzech różnych rozkładów F (rozkładu normalnego, rozkładu Laplace'a o gęstości proporcjonalnej do $\exp\{-|x|\}$ oraz rozkładu Cauchy'ego) moc jest przedstawiona na rys. 4.3.



Rysunek 4.3. Moc dla rozkładu Gaussa (—), Laplace'a (- - -), Cauchy'ego (.....)

Przypadek dwustronnej hipotezy alternatywnej jest technicznie trochę bardziej złożony, ale w stosunku do wszystkiego, o czym dotychczas mówiliśmy w naszym wykładzie, nie stwarza żadnych nowych problemów merytorycznych.



5. Retrospekcja

W całym wykładzie rozwiązywaliśmy tylko jedno zadanie: w celu oszacowania lub zweryfikowania hipotezy o pewnym nieznanym parametrze rzeczywistym μ , wykonywaliśmy odpowiednie obserwacje X_1, \dots, X_n tej wielkości. Każda obserwacja $X_j, j = 1, \dots, n$, różniła się od μ o pewną losową wielkość ε_j (statystyczny błąd obserwacji). Jeżeli o naturze błędu ε nic nie wiemy, to o wielkości μ nie potrafimy nic powiedzieć. Jeżeli potrafimy opisać błąd losowy ε w terminach teorii prawdopodobieństwa, tzn. jeżeli potrafimy coś powiedzieć o rozkładzie prawdopodobieństwa tego błędu losowego, wówczas w tych samych terminach potrafimy odpowiedzieć na różne pytania na temat parametru μ . W ten sposób wnioskowanie o μ staje się wypadkową naszej wiedzy *a priori* o tym parametrze i wiedzy uzyskanej z obserwacji X_1, \dots, X_n .

Rozkład prawdopodobieństwa błędu losowego μ oznaczaliśmy ogólnie przez F (dystrybuanta); wtedy obserwacja miała rozkład F_μ taki, że $F_\mu(x) = F(x - \mu)$. Kolejno rozpatrywaliśmy cztery, coraz bardziej ogólne modele naszych obserwacji X_1, \dots, X_n :

- *Model I*: F jest rozkładem normalnym $N(0, \sigma)$ o znanym odchyleniu standardowym σ .
- *Model II*: F jest rozkładem normalnym $N(0, \sigma)$ o nieznanym odchyleniu standardowym σ .
- *Model III*: F jest znanym rozkładem o ciągłej i ściśle rosnącej dystrybuancie.
- *Model IV*: F jest nieznanym rozkładem o ciągłej i ściśle rosnącej dystrybuancie. W tym przypadku wydaje się, że „faktycznie nic nie wiemy”, ale okazuje się, że wiedza o tym, że jest to dystrybuanta ciągła i ściśle monotoniczna jest już wystarczająca, żeby w połączeniu z informacją z obserwacji X_1, \dots, X_n można było powiedzieć coś niebanalnego na temat interesującego nas parametru μ .

Spójrzmy jeszcze raz na uzyskane wyniki i przyjrzyjmy się dokładności wnioskowania w kolejnych modelach. Dokładność wnioskowania opisujemy za



pomocą szerokości przedziału ufności dla μ . Zależy ona oczywiście od rozkładu F błędu i od liczności n próby X_1, \dots, X_n . Poniżej przedstawiamy porównania numeryczne dla rozkładu normalnego $N(0, 1)$ z dystrybuantą Φ oraz dla $n = 25$. Tak jak w całym wykładzie przez γ oznaczamy poziom ufności.

W modelu I otrzymaliśmy w rozdz. 2 przedział ufności (2.3):

$$\left(\bar{X}_n - z_{(1+\gamma)/2} \frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{(1+\gamma)/2} \frac{\sigma}{\sqrt{n}} \right). \quad (5.1)$$

W modelu II otrzymaliśmy w tym samym rozdz. przedział ufności (2.24):

$$\left(\bar{X}_n - t_{n-1} \left(\frac{1+\gamma}{2} \right) \frac{S}{\sqrt{n-1}}, \bar{X}_n + t_{n-1} \left(\frac{1+\gamma}{2} \right) \frac{S}{\sqrt{n-1}} \right). \quad (5.2)$$

W modelu III otrzymaliśmy w rozdz. 3 przedział ufności (3.10):

$$\left(M_n - \Phi^{-1} \left(B^{-1} \left(\frac{1+\gamma}{2}; \frac{n+1}{2}, \frac{n+1}{2} \right) \right), \right. \\ \left. M_n + \Phi^{-1} \left(B^{-1} \left(\frac{1+\gamma}{2}; \frac{n+1}{2}, \frac{n+1}{2} \right) \right) \right). \quad (5.3)$$

W modelu IV otrzymaliśmy przedział ufności postaci

$$(X_{i:n}, X_{j:n}). \quad (5.4)$$

Przedziały ufności (5.1) oraz (5.3) mają deterministyczną długość, zależną tylko od n . Połowę ich długości oznaczamy, odpowiednio, przez $D(5.1)$ oraz $D(5.3)$. Przedziały (5.2) oraz (5.4) mają losową długość, więc do dalszych porównań weźmiemy oczekiwane wartości ich długości, które oznaczmy, odpowiednio, przez $D(5.2)$ oraz $D(5.4)$. Otrzymujemy

$$D(5.2) = t_{n-1} \left(\frac{1+\gamma}{2} \right) \frac{E(S)}{\sqrt{n-1}}$$

i na mocy wzoru (2.19) mamy

$$E(S) = \sqrt{\frac{2}{\pi}} \frac{\Gamma(\frac{n}{2})}{\Gamma(\frac{n-1}{2})}.$$

Dla $D(5.4)$ otrzymujemy

$$D(5.4) = \frac{1}{2} (E_{N(0,1)} X_{j:n} - E_{N(0,1)} X_{i:n}),$$

gdzie przez $E_{N(0,1)} X_{j:n}$ oznaczyliśmy wartość oczekiwaną j -ej statystyki pozycyjnej z próby X_1, \dots, X_n , gdy ta próba pochodzi ze standardowego rozkładu normalnego $N(0, 1)$.



Po wykonaniu obliczeń otrzymamy:

γ	$D(5.1)$	$D(5.2)$	$D(5.3)$	$D(5.4)$
0.90	0.3290	0.3386	0.4087	0.4086
0.95	0.3920	0.4085	0.4872	0.5193
0.99	0.5152	0.5536	0.6411	0.7641

Tabela dobitnie ilustruje znaczenie wyboru modelu statystycznego i statystyk (średnia? mediana? inne statystyki pozycyjne?), na których ma być oparte wnioskowanie o interesujących nas wielkościach.



Oznaczenia

$N(\mu, \sigma)$	– rozkład normalny o średniej μ i odchyleniu standardowym σ
Φ, φ	– dystrybuanta i gęstość standardowego rozkładu normalnego $N(0,1)$
$z_\alpha = \Phi^{-1}(\alpha)$	– kwantyl rzędu α rozkładu $N(0, 1)$
χ_ν^2	– zmienna losowa o rozkładzie chi-kwadrat z ν stopniami swobody
$\chi_\nu(\alpha)$	– kwantyl rzędu α rozkładu chi-kwadrat o ν stopniach swobody
G_ν, g_ν	– dystrybuanta i gęstość rozkładu chi-kwadrat o ν stopniach swobody
t_ν	– zmienna losowa o rozkładzie t Studenta z ν stopniami swobody
$t_\nu(\alpha)$	– kwantyl rzędu α rozkładu t Studenta z ν stopniami swobody
$t_{\nu, \lambda}$	– zmienna losowa o niecentralnym rozkładzie t Studenta z ν stopniami swobody i parametrze niecentralności λ
$H_{\nu, \lambda}, h_{\nu, \lambda}$	– dystrybuanta i gęstość zmiennej losowej $t_{\nu, \lambda}$
$B(x; p, q), b(x; p, q)$	– dystrybuanta i gęstość rozkładu beta z parametrem (p, q)
$B^{-1}(\alpha; p, q)$	– kwantyl rzędu α rozkładu beta z parametrem (p, q)
\mathcal{F}	– rodzina wszystkich rozkładów o ciągłej i ściśle rosnącej dystrybuancie



Literatura

- J. Bartoszewicz, *Wykłady ze statystyki matematycznej*. PWN, Warszawa, 1996.
- D. Bobrowski i K. Maćkowiak-Łybacka, *Wybrane metody wnioskowania statystycznego*. Wydawnictwo Politechniki Poznańskiej, Poznań, 2006.
- L. Gajek i M. Kałużska, *Wnioskowanie statystyczne. Modele i metody*. WNT, Warszawa, 1996.
- W. Klonecki, *Statystyka dla inżynierów*. PWN, Warszawa, 1999.
- J. Koronacki i J. Mielniczuk, *Statystyka dla studentów kierunków technicznych i przyrodniczych*. WNT, Warszawa, 2001.
- M. Krzyśko, *Statystyka matematyczna*. Wydawnictwo Naukowe UAM, Poznań, 2004.
- R. Magiera, *Modele i metody statystyki matematycznej. Część I. Rozkłady i symulacja stochastyczna*. Oficyna Wydawnicza GiS, Wrocław, 2005.
- R. Magiera, *Modele i metody statystyki matematycznej. Część II. Wnioskowanie statystyczne*. Oficyna Wydawnicza GiS, Wrocław, 2007.
- W. Niemirow, *Rachunek prawdopodobieństwa i statystyka matematyczna*. Biblioteka Szkoły Nauk Ścisłych, Warszawa, 1999.
- A. Plucińska i E. Pluciński, *Rachunek prawdopodobieństwa. Statystyka matematyczna. Procesy stochastyczne*. WNT, Warszawa, 2000.
- M. Startek, *Podstawy rachunku prawdopodobieństwa z elementami statystyki matematycznej*. Oficyna Wydawnicza Politechniki Rzeszowskiej, Rzeszów, 2005.
- R. Zieliński, *Siedem wykładów wprowadzających do statystyki matematycznej*, PWN, Warszawa, 1990 (www.impan.pl/~rziel/7ALL.pdf).



Indeks

- błąd I rodzaju, 9
- błąd II rodzaju, 10
- błąd losowy, 1
- długość przedziału ufności, 5, 24
- dokładność estymacji, 5
- dwustronny przedział ufności, 5
- dystrybuanta empiryczna, 41
- empiryczny poziom istotności, 13
- estymacja, 4
- estymacja parametru skali, 38
- estymator medianowo nieobciążony, 34
- estymator nieobciążony odchylenia standardowego, 16
- estymator nieobciążony wariancji, 16
- estymator odporny, 6
- funkcja charakterystyczna, 30
- funkcja mocy, 10, 11
- hipoteza alternatywna, 12
- innowacje, 36, 37
- jednostronny przedział ufności, 5
- mediana, 1, 6, 33
- mediana próbkowa, 6
- mediana z próby, 6
- moc testu, 10
- najkrótszy przedział ufności dla wariancji, 18
- niecentralny rozkład t Studenta, 26
- niekompletna funkcja beta, 32
- obszar krytyczny testu, 10
- p-value, 14
- p-wartość, 14
- parametr kształtu, 39
- parametr niecentralności, 26
- poziom istotności, 9
- poziom krytyczny testu, 13
- poziom ufności, 5
- próba, 4
- próba zero-jedynkowa, 45
- prawdopodobieństwo krytyczne, 13
- procedura Steina, 24
- przedział ufności, 5
- przedział ufności dla kwantyla, 42
- przedział ufności dla mediany, 35, 42
- przedział ufności dwustronny, 5
- przedział ufności jednostronny, 5
- przedział ufności najkrótszy, 18, 44
- przedział ufności symetryczny, 18
- rozkład α -stabilny, 30
- rozkład k -tej statystyki pozycyjnej, 33
- rozkład t Studenta, 23
- rozkład t Studenta niecentralny, 26
- rozkład beta, 31
- rozkład Cauchy'ego, 29, 36
- rozkład chi-kwadrat, 15
- rozkład dwumianowy, 32, 45
- rozkład empiryczny, 41
- rozkład Laplace'a, 47
- rozkład Levy'ego, 38
- rozkład Pareto, 39
- ryzyko estymatora, 17
- statystyki pozycyjne, 6, 31, 41
- stopnie swobody, 23



średni względny błąd kwadratowy, 17

średnia, 4

środek rozrzutu, 7

test, 1

test nieobciążony, 21

test obciążony, 21

testowanie hipotez, 4, 8

wariancja, 4

wartość krytyczna, 13

wynik pomiaru, 1

zmienna losowa χ^2 , 15

