Część III

Testowanie hipotez statystycznych

Rozdział 7

Testy istotności

W tym rozdziale spróbujemy wyjaśnić, na czym polega zagadnienie testowania hipotez statystycznych. Pokażemy, jak konstruuje się tak zwane *testy istotności*. Skoncentrujemy się na kilku ważnych i typowych przykładach. Nasze rozważania będą miały charakter wstępny i heurystyczny. Bardziej systematyczną teorię przedstawimy w rozdziale następnym.

Przykład wstępny

Matematyczny model zjawiska losowego sprowadza się do określenia rozkładu prawdopodobieństwa $\mathbb{P}$ obserwacji X na przestrzeni $\mathcal{X}$. Przez *hipotezę statystyczną* rozumiemy, najogólniej mówiąc, pewną wypowiedź na temat rozkładu prawdopodobieństwa „rządzącego” zjawiskiem, które nas interesuje. Zajmiemy się najprostszą sytuacją. Rozważamy na razie jeden, ustalony rozkład $\mathbb{P}$. Pytamy, czy opisuje on poprawnie przebieg doświadczenia losowego. Stawiamy hipotezę: nasz model probabilistyczny jest zgodny z rzeczywistością. Ogólna zasada, wykraczająca daleko poza statystykę, jest taka: powinniśmy odrzucić hipotezę, jeśli rezultaty doświadczenia okażą się sprzeczne z przewidywaniami wynikającymi z modelu. Specyfika modeli probabilistycznych polega jednak na tym, że potrafimy przewidzieć tylko prawdopodobieństwo poszczególnych wyników. Zmodyfikujemy więc ogólną zasadę. Powinniśmy odrzucić hipotezę, *jeśli wynik doświadczenia jest bardzo mało prawdopodobny*, to znaczy model przewiduje taki wynik z małym prawdopodobieństwem. Innymi słowy, wybieramy pewien zbiór $K \subset \mathcal{X}$ taki, że $\mathbb{P}(X \in K) \leqslant \alpha$, gdzie α jest odpowiednio „małą” liczbą, zwaną *poziomem istotności*. Jeśli w wyniku doświadczenia stwierdzimy, że zaszło zdarzenie „$X \in K$”, to powinniśmy zwątpić w poprawność naszego modelu, czyli „odrzucić postawioną przez nas hipotezę statystyczną”. Zbiór K nazywamy *obszarem krytycznym* testu, zaś cała procedura nosi nazwę testu *istotności*.

7.0.1 Przykład (Czy prawdopodobieństwo „jedynki” jest równe 1/6?)**.** Wykonujemy 300 rzutów kostką do gry. Jeśli kostka jest symetryczna i rzuty wykonujemy rzetelnie, to zmienna losowa $X = N_1$ równa liczbie wyrzuconych „jedynek” ma rozkład dwumianowy $\mathrm{Bin}(300, 1/6)$. Mamy jednak pewne wątpliwości, czy kasyno nie używa obciążonej kostki, w której jedynka ma prawdopodobieństwo *mniejsze* niż 1/6. Jeśli w wyniku doświadczenia ani ani razu nie otrzymamy jedynki, musimy podejrzewać, że coś jest nie w porządku. Może kostka w ogóle nie ma ściany oznaczonej jedną kropką? Co prawda, zdarzenie „$X = 0$” w modelu symetrycznej kostki jest *możliwe* ale skrajnie mało prawdopodobne, $\mathbb{P}(X = 0) = (5/6)^{300} \approx 1.76{*}10^{-24}$. Nie jesteśmy skłonni wierzyć w tak wyjątkowy traf. Odrzucamy, bez wielkiego wahania, hipotezę o zgodności modelu. Jak postąpić jeśli otrzymamy, powiedzmy, 33 jedynki, nie jest aż tak jasne. Czy to wynik zgodny z hipotezą o rzetelności kostki, czy też jest podstawą do odrzucenia tej hipotezy?

Rozpatrzymy sprawę bardziej systematycznie. Stawiamy hipotezę

$$\mathrm{H}_0 : X \sim \mathrm{Bin}(300, 1/6).$$

Zdecydujemy się na taki przepis postępowania:

jeśli $X < c$ to odrzucamy H_0;

jeśli $X \geqslant c$ to pozostajemy przy H_0.

dla pewnej liczby „progowej” c, testu. Próg c ustalimy w następujący sposób. Wybierzmy *poziom istotności*, powiedzmy $\alpha = 0.01$. Uznamy, że zajście zdarzenia o prawdopodobieństwie mniejszym niż α jest dostatecznym powodem do odrzucenia H_0. Powinniśmy więc wybrać c tak, aby $\mathbb{P}(X < c) \leqslant 0.01$. Mamy przy tym na myśli prawdopodobieństwo obliczone przy założeniu, że H_0 jest prawdziwa. Łatwo przekonać się, że $\mathbb{P}(X < 36) = 0.00995 \leqslant 0.01$, a więc ostatecznie nasz test na poziomie istotności 0.01 jest następujący:

Odrzucamy H_0 jeśli $X < 36$.

Obszarem krytycznym naszego testu jest podzbiór $K = \{0, 1, \ldots, 35\}$ przestrzeni obserwacji $\mathcal{X} = \{0, 1, \ldots, 300\}$. Jeżeli wynikiem doświadczenia są 33 jedynki, to opisana powyżej reguła decyzyjna odrzuca H_0. Inne spojrzenie na ten sam test jest następujące. Obliczamy prawdopodobieństwo otrzymania wyniku takiego jak 33 lub gorszego (za „gorszy” uznajemy zbyt małą liczbę jedynek), przy założeniu H_0. Mówimy, że *p-wartość* (lub *poziom krytyczny testu*) jest w naszym przypadku równa $P = \mathbb{P}(X \leqslant 33) = 0.00379$. Porównujemy p-wartość z założonym poziomem istotności.

Odrzucamy H_0 jeśli $P < 0.01$.

Obie reguły są równoważne, to znaczy prowadzą do takiej samej decyzji. Niemniej obliczanie p-wartości stwarza możliwość (lub pokusę) „mierzenia stopnia przekonania” o nieprawdziwości H_0.

Zwróćmy uwagę, że nasz test „reaguje na odstępstwa od prawdopodobieństwa 1/6 tylko w dół”. Mówimy, że jest to test *hipotezy zerowej*

$$\mathrm{H}_0 : \text{prawdopodobieństwo jedynki } = 1/6$$

przeciwko hipotezie alternatywnej

$$\mathrm{H}_1 : \text{prawdopodobieństwo jedynki } < 1/6.$$

Jeśli rozważymy inną „jednostronną” alternatywę, H_1: prawdopodobieństwo jedynki $> 1/6$, to test powinien odrzucać H_0 jeśli $X > c$, dla odpowiednio dobranego c. Wybór $c = 65$ prowadzi do testu na poziomie istotności $\alpha = 0.01$, bo $\mathbb{P}(X > 65) = 0.00990$. Można również skonstruować test H_0 przeciwko „dwustronnej” alternatywie H_1: prawdopodobieństwo jedynki $\neq 1/6$. Do tej dyskusji jeszcze wrócimy. $\diamondsuit$

Podsumujmy rozważania z powyższego przykładu. Mieliśmy w nim do czynienia z hipotezą H_0, która precyzuje dokładnie jeden rozkład prawdopodobieństwa $\mathbb{P}$. Mówimy, że jest to *hipoteza prosta*. Test hipotezy H_0 na poziomie α spełnia postulat

$$\mathbb{P}(\text{ odrzucamy } \mathrm{H}_0) \leqslant \alpha,$$

gdzie prawdopodobieństwo jest obliczone przy założeniu prawdziwości H_0. Konstrukcja testu istotności dopuszcza dużą dowolność. Wybór obszaru krytycznego zależy od tego, „przed jaką ewentualnością chcemy się zabezpieczyć”, czyli jaka jest hipoteza alternatywna H_1.

Sposób obliczania p-wartości [1] też zależy od wyboru H_1 (bo musimy sprecyzować, jakie wyniki uznajemy za „gorsze”, bardziej świadczące przeciw H_0). Rozważmy statystykę $T = T(X)$ taką, że duże wartości T „świadczą przeciw H_0 na korzyść H_1”. Jeśli dla zaobserwowanej wartości $X = x$ otrzymaliśmy wartość statystyki $t = T(x)$, to p-wartość testu jest liczbą

$$P = \mathbb{P}(T(X) \geqslant t),$$

gdzie prawdopodobieństwo jest obliczone przy założeniu H_0. Innymi słowy, p-wartość jest prawdopodobieństwem, z jakim należałoby oczekiwać wyniku takiego jak faktycznie otrzymany lub „gorszego” jeśli hipoteza zerowa byłby prawdziwa. Oczywiście, mała p-wartość świadczy przeciw H_0. Jeśli c jest kwantylem rzędu $1 - \alpha$ statystyki testowej T, to zdarzenie $T > c$ zachodzi wtedy i tylko wtedy, gdy $p < \alpha$. Test na poziomie istotności α możemy sformułować dwojako.

Odrzucamy H_0 na rzecz H_1, jeśli $T > c$ lub równoważnie, jeśli $p < \alpha$.

[1] Używamy tu bardziej rozpowszechnionego terminu *p-wartość* (dosłowne tłumaczenie angielskiego „*p-value*”) zamiast poprawnej ale wychodzącej z użycia nazwy *poziom krytyczny*.

Ponieważ p-wartość jest obliczona na podstawie danych, jest więc statystyką, zmienną losową, i możemy mówić o jej rozkładzie prawdopodobieństwa. Niech $Q(t) = \mathbb{P}(T(X) \geqslant t)$ (prawdopodobieństwo obliczamy przy założeniu, że H_0 jest prawdziwa, czyli $X \sim \mathbb{P}$). Z definicji, p-wartość testu jest obliczana zgodnie ze wzorem

$$P = Q(T(X)).$$

Następujący fakt wynika bezpośrednio z określenia p-wartości.

7.0.2 Stwierdzenie. *Załóżmy, że* H_0 *jest prosta i rozkład statystyki testowej jest ciągły. Przy założeniu prawdziwości* H_0*, p-wartość testu jest zmienną losową o rozkładzie* $\mathrm{U}(0,1)$ *(jednostajnym na przedziale* $[0,1]$*).*

Dowód. Funkcja Q jest ciągła z założenia i oczywiście nierosnąca. Przeprowadzimy dowód zakładając dodatkowo, że Q jest na przedziale $\{x : 0 < Q(x) < 1\}$ ściśle malejąca (bez tego założenia rozumowanie wymaga tylko nieznacznych modyfikacji, ale pomińmy szczegóły). Niech Q^{-1} będzie funkcją odwrotną do Q. Obliczmy dystrybuantę zmiennej losowej $Q(T)$. Dla $0 < u < 1$ mamy

$$\begin{aligned}
\mathbb{P}(P \leqslant u) &= \mathbb{P}(Q(T) \leqslant u) && \text{(z definicji } P\text{)}\\
&= \mathbb{P}(T \geqslant Q^{-1}(u)) && \text{(bo } Q^{-1} \text{ jest funkcją malejącą)}\\
&= Q(Q^{-1}(u)) = u && \text{(z definicji funkcji } Q\text{)},
\end{aligned}$$

a to jest dystrybuanta rozkładu $\mathrm{U}(0,1)$. □

Oczywiście, jeśli H_0 nie jest prawdziwa, to rozkład p-wartości nie jest jednostajny. Jest pożądane, żeby był jak najbardziej skupiony na wartościach bliskich zeru (bo chcemy, żeby test „rozpoznawał” nieprawdziwość hipotezy zerowej).

7.1 Kilka wybranych testów istotności

Test proporcji

W sytuacji takiej jak w Przykładzie 7.0.1, powszechnie używa się testów opartych na przybliżeniu rozkładu dwumianowego rozkładem normalnym. Zakładamy, że obserwujemy liczbę sukcesów w schemacie Bernoulli'ego i stawiamy hipotezę

- H_0 : prawdopodobieństwo sukcesu $= p$.

Z Twierdzenia de Moivre'a - Laplace'a (CTG dla schematu Bernoulli'ego) wynika, że

$$\mathrm{Bin}(n,p) \simeq \mathrm{N}(np, np(1-p)),$$

przynajmniej dla „dostatecznie dużych n" (niektóre źródła zalecają stosowanie tego przybliżenia, jeśli $np \geqslant 5$ i $n(1-p) \geqslant 5$). Rozpatrzymy kolejno zagadnienia weryfikacji hipotezy zerowej przeciw alternatywie *lewostronnej, prawostronnej i dwustronnej*. Podamy testy na poziomie istotności (w przybliżeniu) α. Niech $\mathrm{z}_{1-\alpha}$ będzie kwantylem rozkładu $\mathrm{N}(0,1)$, to znaczy $\Phi(\mathrm{z}_{1-\alpha}) = 1-\alpha$.

- Test H_0 przeciwko alternatywie H_1 : prawdopodobieństwo sukcesu $< p$.

  Odrzucamy H_0, jeśli $X - np < -\mathrm{z}_{1-\alpha}\sqrt{np(1-p)}$.

  Reguła obliczania p-wartości jest następująca: $p \simeq \Phi\left(\dfrac{X-np}{\sqrt{np(1-p)}}\right)$.

- Test H_0 przeciwko alternatywie H_1 : prawdopodobieństwo sukcesu $> p$.

  Odrzucamy H_0, jeśli $X - np > \mathrm{z}_{1-\alpha}\sqrt{np(1-p)}$.

  Reguła obliczania p-wartości: $p \simeq \Phi\left(-\dfrac{X-np}{\sqrt{np(1-p)}}\right)$.

- Test H_0 przeciwko alternatywie H_1 : prawdopodobieństwo sukcesu $\neq p$:

  Odrzucamy H_0 jeśli $|X-np| > \mathrm{z}_{1-\alpha/2}\sqrt{np(1-p)}$ (zauważmy zamianę $\mathrm{z}_{1-\alpha}$ na $\mathrm{z}_{1-\alpha/2}$).

  Reguła obliczania p-wartości jest następująca: $p \simeq 2\Phi\left(-\dfrac{|X-np|}{\sqrt{np(1-p)}}\right)$.

Ze względu na późniejsze uogólnienia, napiszmy ten ostatni, dwustronny test w innej, równoważnej formie:

$$\text{Odrzucamy } \mathrm{H}_0 \text{ jeśli } \frac{(X-np)^2}{np(1-p)} > \chi^2_{1-\alpha}(1)$$

gdzie $\chi^2_{1-\alpha}(1) = \mathrm{z}^2_{1-\alpha/2}$ jest kwantylem rozkładu chi-kwadrat z jednym stopniem swobody.

Uwaga. Opisaliśmy test *przybliżony*, oparty na teorii asymptotycznej. Istnieje dość prosty test *dokładny*. Przybliżony test ma tę zaletę dydaktyczną, że w sposób naturalny uogólnia się na przypadek doświadczeń o $k > 2$ możliwych wynikach (zamiast dwóch: sukces/porażka). Tym uogólnieniem jest sławny test χ^2.

Test chi-kwadrat

Jest to jeden z klasycznych i najczęściej stosowanych testów statystycznych. Wyobraźmy sobie, że powtarzamy n-krotnie doświadczenie losowe, które ma k możliwych wyników. Wyniki kolejnych doświadczeń są zmiennymi losowymi $X_1, \dots, X_n$ o wartościach w zbiorze $\mathcal{X} = \{w_1, \dots, w_k\}$. Przypuśćmy, że te zmienne są niezależne i każda z nich ma ten sam rozkład prawdopodobieństwa dany tabelką

wynik	w_1	$\cdots$	w_i	$\cdots$	w_k
prawdopodobieństwo	p_1	$\cdots$	p_i	$\cdots$	p_k

gdzie $p_i = \mathbb{P}(X_j = w_i)$ (oczywiście, $p_1 + \cdots + p_k = 1$). Rezultaty doświadczeń podsumujmy w postaci „tabelki powtórzeń”:

wynik	w_1	$\cdots$	w_i	$\cdots$	w_k
liczba doświadczeń	N_1	$\cdots$	N_i	$\cdots$	N_k

wynik	w_1	$\cdots$	w_i	$\cdots$	w_k
liczba doświadczeń	N_1	$\cdots$	N_i	$\cdots$	N_k

gdzie $N_i = \sum_{j=1}^{n} \mathbb{1}(X_j = w_i)$ jest liczbą doświadczeń, których wynikiem jest w_i (oczywiście, $N_1 + \cdots + N_k = n$). Jeśli nasze przypuszczenia są słuszne, to zmienne losowe $N_1, \dots, N_k$ mają rozkład wielomianowy $\mathrm{Mult}(n, p_1, \dots, p_k)$, to znaczy

$$\mathbb{P}(N_1 = n_1, \dots, N_k = n_k) = \frac{n!}{n_1! \cdots n_k!} p_1^{n_1} \cdots p_k^{n_k}.$$

Najlepiej patrzeć na nasze doświadczenie jak na losowe wrzucanie n „kul” do k „komórek” („wielkość” i-tej komórki jest proporcjonalna do p_i). Test hipotezy

$$\mathrm{H}_0 : (N_1, \dots, N_k) \sim \mathrm{Mult}(n, p_1, \dots, p_k)$$

przeprowadzamy w następujący sposób. Obliczamy statystykę „chi-kwadrat”:

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - np_i)^2}{np_i}. \tag{7.1.1}$$

Dla wyznaczenia wartości krytycznej testu lub p-wartości, korzystamy z następującego faktu, którego dowód pominiemy.

7.1.2 Stwierdzenie. *Jeżeli* H_0 *jest prawdziwa i* $n \to \infty$*, to rozkład statystyki* χ^2 *danej wzorem* (7.1.1) *zmierza do rozkładu* $\chi^2(k-1)$ *(chi-kwadrat z* $k-1$ *stopniami swobody).*

Test na poziomie istotności α (w przybliżeniu) jest następującą regułą postępowania:

odrzucamy H_0 jeśli $\chi^2 > c$,

gdzie $c = \chi^2_{1-\alpha}(k-1)$ jest kwantylem rzędu $1-\alpha$ rozkładu $\chi^2(k-1)$.

7.1.3 Przykład (Czy kość jest rzetelna?)**.** Przypuśćmy, że wykonaliśmy 150 rzutów kostką. Wyniki doświadczenia podsumujmy w następującej tabelce:

liczba oczek	1	2	3	4	5	6
liczba rzutów	15	27	36	17	26	29

Stawiamy hipotezę, że kość jest rzetelna, czyli

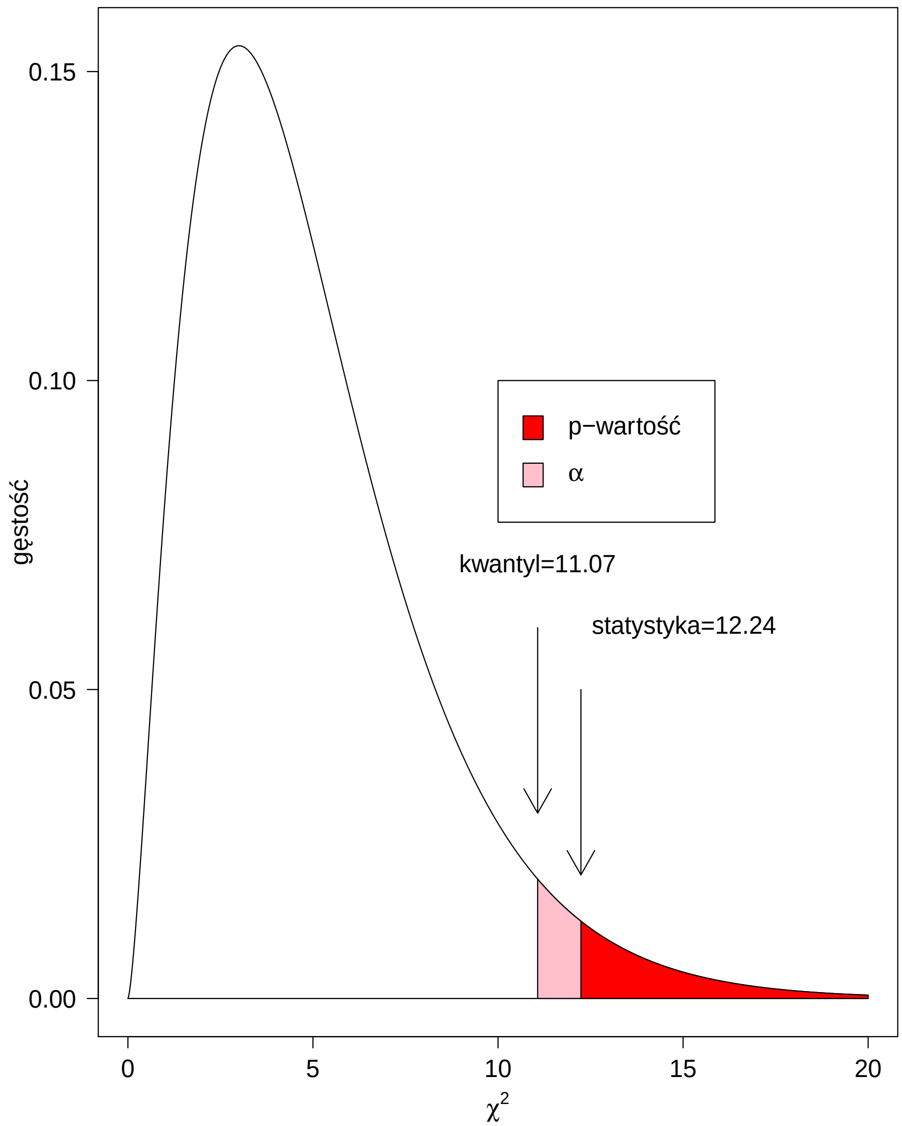
$$\mathrm{H}_0 : (N_1, N_2, N_3, N_4, N_5, N_6) \sim \mathrm{Mult}(150, 1/6, 1/6, 1/6, 1/6, 1/6, 1/6).$$

Wartość statystyki testowej jest równa

$$\begin{aligned} \chi^2 &= \frac{(15-25)^2}{25} + \frac{(27-25)^2}{25} + \frac{(36-25)^2}{25} \\ &+ \frac{(17-25)^2}{25} + \frac{(26-25)^2}{25} + \frac{(29-25)^2}{25} = 12.24. \end{aligned}$$

Przeprowadzimy test tej hipotezy na poziomie istotności 0.05. Odczytujemy z tablic kwantyl rzędu 0.95 rozkładu chi-kwadrat z $6-1=5$ stopniami swobody: $\chi^2_{0.95}(5) = 11.07$. Ponieważ $12.24 > 11.7$, odrzucamy H_0. Należy przypuszczać, że kość nie jest symetryczna. Zamiast odczytać z tablic kwantyl, można obliczyć p-wartość i zauważyć, że $P = 0.03164 < \alpha = 0.05$. Rysunek 7.1 pokazuje kwantyl $\chi^2_{0.95}(5) = 11.07$ i wartość statystyki testowej $\chi^2 = 12.24$ oraz gęstość rozkładu $\chi^2(5)$. Oba równoważne sposoby są zilustrowane na Rysunku 7.1. Pokazuje on kwantyl $\chi^2_{0.95}(5) = 11.07$ i wartość statystyki testowej $\chi^2 = 12.24$ oraz gęstość rozkładu $\chi^2(5)$.

$\Diamond$



Rysunek 7.1: p-wartość i poziom istotności testu.

7.1.4 Przykład (Rozkład statystyki testowej i p-wartości)**.** Korzystając z ułatwienia, jakie daje komputer wyposażony w R, wykonajmy test opisany w poprzednim przykładzie wiele razy. Powtórzmy $m = 1000$ razy serię $n = 150$ rzutów rzetelną kostką (H_0 prawdziwa). Rysunek 7.2 w górnej części pokazuje empiryczny rozkład statystyk testowych w $m = 1000$ powtórzeniach testu zgodności χ^2. Dane pochodziły z rozkładu spełniającego hipotezę zerową H_0. Poniżej pokazany jest empiryczny rozkład odpowiednich p-wartości.

Widzimy, że zgodnie ze Stwierdzeniem 7.1.2, wartości statystyki χ^2 mają rozkład zbliżony do $\chi^2(5)$ (gęstość tego rozkładu jest nałożona na histogram). Z kolei p-wartości mają rozkład zbliżony do rozkładu jednostajnego $\mathrm{U}(0, 1)$. To jest oczywiście związane ze Stwierdzeniem 7.0.2. Zauważmy przy tym, że założenia Stwierdzenia 7.0.2 są spełnione tylko w przybliżeniu. Statystyka χ^2 jest zmienną dyskretną i tylko w granicy ma rozkład ciągły $\chi^2(5)$. Pomimo tego, zgodność p-wartości z rozkładem jednostajnym jest uderzająca. ◊

Test chi-kwadrat ma wiele odmian, stosowanych w różnych okolicznościach. Z pewnymi wersjami tego testu spotkamy się jeszcze w tych wykładach. Postać statystyki, którą tu rozpatrzyliśmy najłatwiej zrozumieć i zapamiętać w takiej postaci:

$$\chi^2 = \sum \frac{(\text{wielkość obserwowana} - \text{wielkość oczekiwana})^2}{\text{wielkość oczekiwana}}.$$

Sumowanie rozciągnięte jest na wszystkie „komórki”, czyli wartości obserwacji. Oczywiście, w wersji tutaj omawianej „wielkość obserwowana” oznacza N_i, zaś „wielkość oczekiwana” (przy założeniu prawdziwości hipotezy) jest równa np_i.

Dyskretyzacja zmiennych ciągłych. Wspomnijmy jeszcze, że prosty chwyt pozwala stosować test zgodności χ^2 również do obserwacji „typu ciągłego”. Rozpatrzmy zmienne losowe $X_1, \dots, X_n$ i hipotezę

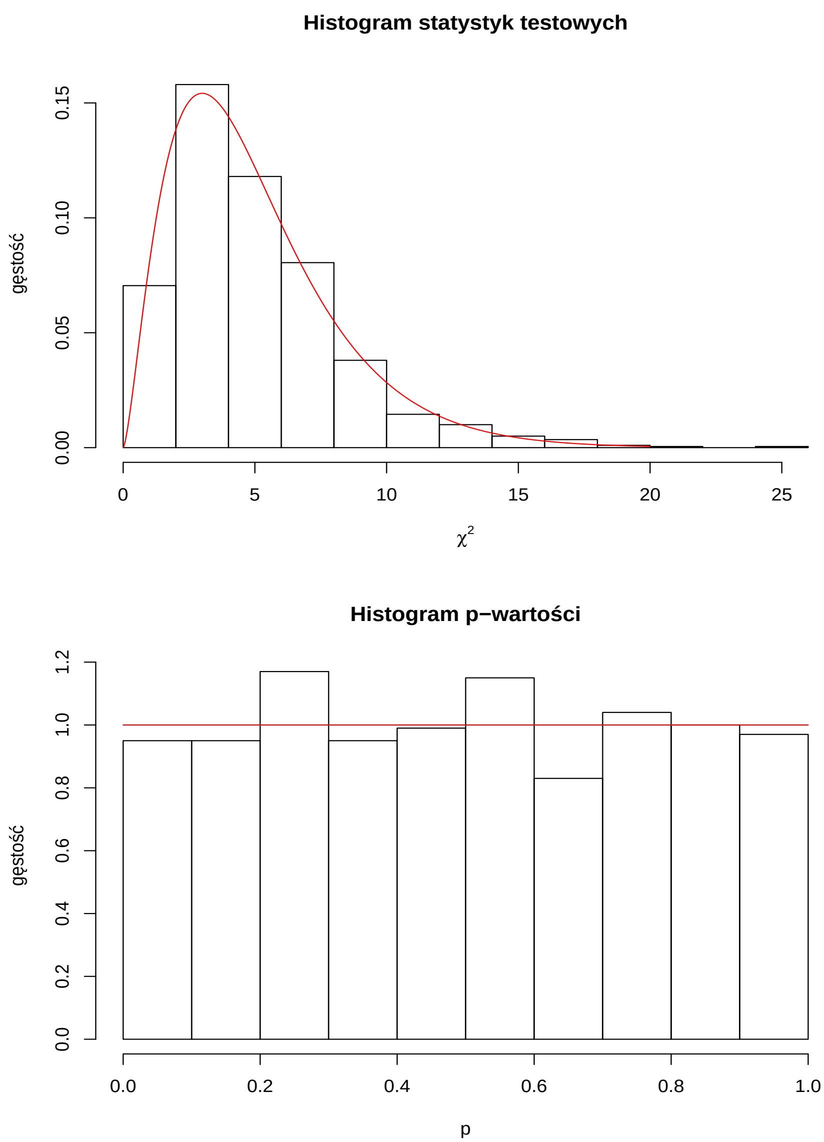
$$\mathrm{H}_0 : X_1, \dots, X_n \text{ jest próbką z rozkładu o dystrybuancie } F,$$

gdzie F jest ciągła. Wystarczy rozbić zbiór wartości zmiennych losowych na rozłączne „komórki”, na przykład wybrać punkty $-\infty = a_0 < a_< \cdots < a_k = \infty$ i zdefiniować

$$N_i = \sum_{j=1}^{n} \mathbb{1}(a_{i-1} < X_j \leqslant a_i), \qquad (i = 1, \dots, k).$$

Dalej postępujemy tak jak poprzednio, przyjmując $p_i = F(a_i) - F(a_{i-1})$. Taki test dopuszcza dużo dowolności w wyborze „komórek”.

Istnieją testy lepiej dostosowane do ciągłego typu danych. Należy do nich test Kołmogorowa-Smirnowa, który później omówimy.



Rysunek 7.2: Rozkład wartości statystyki testowej χ^2 i p-wartości testu.

Test chi-kwadrat hipotezy złożonej. Opiszemy modyfikację testu chi-kwadrat w sytuacji, gdy hipoteza zerowa precyzuje rozkład prawdopodobieństwa jedynie z dokładnością do nieznanego parametru. Podobnie jak dla „zwykłego” testu χ^2, rozważamy n niezależnych powtórzeń doświadczenia losowego o k możliwych wynikach. Hipoteza H_0 stwierdza, że rozkład prawdopodobieństwa pojedynczego doświadczenia jest dany tabelką:

wynik	w_1	$\cdots$	w_i	$\cdots$	w_k
prawdopodobieństwo	$p_1(\theta)$	$\cdots$	$p_i(\theta)$	$\cdots$	$p_k(\theta)$

gdzie $p_i(\theta)$ są, dla $i = 1, \cdots, k$, znanymi funkcjami nieznanego parametru θ. Ten parametr może być wektorem d-wymiarowym: $\theta = (\theta_1, \ldots, \theta_d)$. Statystykę χ^2 budujemy w dwóch etapach. Najpierw *estymujemy* θ metodą największej wiarogodności. Mamy

$$\begin{aligned} f_\theta(n_1, \ldots, n_k) &= \mathbb{P}_\theta(N_1 = n_1, \ldots, N_k = n_k) \\ &= \mathrm{const} \cdot p_1(\theta)^{n_1} \cdots p_k(\theta)^{n_k} \end{aligned}$$

(tak jak w „zwykłym” teście χ^2, zmienna N_i oznacza liczbę powtórzeń i-tego wyniku). Logarytm wiarogodności ma więc postać

$$l(\theta) = \mathrm{const} + \sum_{i=1}^{k} N_i \log p_i(\theta)$$

i estymator największej wiarogodności obliczamy rozwiązując układ równań

$$\frac{\partial l(\theta)}{\partial \theta_j} = \sum_{i=1}^{k} N_i \frac{1}{p_i(\theta)} \frac{\partial p_i(\theta)}{\partial \theta_j} = 0, \quad (j = 1, \ldots, d).$$

Statystykę χ^2 określamy tak jak w (7.1.1), wstawiając $\hat{\theta} = \mathrm{MLE}(\theta)$ w miejsce nieznanego θ:

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(N_i - np_i(\hat{\theta})\right)^2}{np_i(\hat{\theta})}.$$

Rozkład tej statystyki jest *inny*, niż statystyki obliczonej w przypadku ustalonych z góry prawdopodobieństw p_i. Intuicyjnie jest dość jasne, że statystyka z *estymowanymi* prawdopodobieństwami klatek jest średnio *mniejsza*, bo wartości oczekiwane są wyliczane na podstawie danych i mogą się „dopasować” do wartości obserwowanych. Okazuje się, że to „dopasowanie” zmniejsza *liczbę stopni swobody* granicznego rozkładu χ^2. Podamy ten fakt bez dowodu:

7.1.5 Stwierdzenie. *W opisanej wyżej sytuacji, rozkład statystyki χ^2 zmierza do rozkładu $\chi^2(k - d - 1)$, gdy $n \to \infty$.*

Przypomnijmy, że d jest liczbą współrzędnych wektora $(\theta_1, \ldots, \theta_d)$. Symbolicznie, stopnie swobody = liczba klatek − *liczba estymowanych parametrów* −1.

Oczywiście, wynika stąd następujący przepis na budowę testu na poziomie istotności α (w przybliżeniu, dla dużej liczby powtórzeń n):

odrzucamy H_0 gdy $\chi^2 > c$, gdzie $c = \chi^2_{1-\alpha}(k-d-1)$.

7.1.6 Przykład (Zgodność danych z modelem Hardy'ego-Weinberga)**.** Mówiliśmy już o tym, jak estymować nieznany parametr w tym modelu, patrz Przykłady 3.1.2 i 5.2.4. Teraz zajmiemy się sprawdzeniem hipotezy, że model poprawnie opisuje wyniki doświadczenia (genetycznego, w tym przypadku). Nasza hipoteza jest więc taka:

$$\mathrm{H}_0 : p_1 = \theta^2,\ \ p_2 = 2\theta(1-\theta),\ \ p_3 = (1-\theta)^2 \quad \text{dla pewnego } \theta \in (0,1).$$

Wykorzystamy estymator największej wiarogodności

$$\hat{\theta} = \mathrm{MLE}(\theta) = \frac{2N_1 + N_2}{2n}.$$

Statystyką testową jest

$$\chi^2 = \frac{\left(N_1 - n\hat{\theta}^2\right)^2}{n\hat{\theta}^2} + \frac{\left(N_2 - 2n\hat{\theta}(1-\hat{\theta})\right)^2}{2n\hat{\theta}(1-\hat{\theta})} + \frac{\left(N_3 - n(1-\hat{\theta})^2\right)^2}{n(1-\hat{\theta})^2}.$$

Proste, choć nieco żmudne przekształcenia pozwalają uprościć to wyrażenie:

$$\chi^2 = n\left(1 - \frac{N_2}{2n\hat{\theta}(1-\hat{\theta})}\right)^2.$$

Dla dużych n, ta statystyka ma w przybliżeniu rozkład $\chi^2(3-1-1) = \chi^2(1)$. ◊

Testowanie zgodności z typem rozkładu. Obserwujemy próbkę $X_1, \dots, X_n$ z pewnego nieznanego rozkładu o dystrybuancie F. Przypuszczamy, że jest to rozkład normalny. Stawiamy hipotezę

$$\mathrm{H}_0 : X_1, \dots, X_n \sim \mathrm{N}(\mu, \sigma^2) \quad \text{dla pewnych } \mu, \sigma.$$

Inaczej, H_0 mówi, że dystrybuanta ma postać $F(x) = \Phi\left((x-\mu)/\sigma\right)$ dla pewnych μ, σ, gdzie Φ jest *standardową* dystrybuantą normalną (dla rozkładu $\mathrm{N}(0,1)$).

Opiszemy sposób zastosowania testu chi-kwadrat w wersji dla hipotez złożonych. Najpierw dane ciągłe musimy zdyskretyzować, w sposób już wcześniej wyjaśniony. Wybieramy punkty $-\infty = a_0 < a_1 < \cdots < a_k = \infty$ i dzielimy prostą na rozłączne przedziały. Oznaczamy przez

N_i liczbę obserwacji wpadających do i-tego przedziału. Wartości obserwowane porównujemy z wartościami oczekiwanymi $n\hat{p}_i$, gdzie

$$\hat{p}_i = \Phi\left(\frac{a_i - \hat{\mu}}{\hat{\sigma}}\right) - \Phi\left(\frac{a_{i-1} - \hat{\mu}}{\hat{\sigma}}\right).$$

W zasadzie, należy za estymatory μ i σ przyjąć ENW obliczone dla *zdyskretyzowanych* zmiennych. Zgodnie z tym co powiedzieliśmy, te ENW są rozwiązaniem układu równań

$$\sum_i N_i \frac{1}{p_i}\frac{\partial p_i}{\partial \mu} = 0, \quad \sum_i N_i \frac{1}{p_i}\frac{\partial p_i}{\partial \sigma} = 0.$$

Statystyka χ^2 ma asymptotyczny rozkład $\chi^2(k-3)$ (bo estymujemy 2 parametry). W praktyce, rozwiązywanie napisanych wyżej równań jest zbyt kłopotliwe i przyjmuje się, że wstawienie „zwyczajnych” estymatorów $\hat{\mu} = \bar{X}$ i $\hat{\sigma} = S$ we wzory na p_i niewiele zmienia rozkład statystyki χ^2.

Dokładnie w ten sam sposób można testować inne hipotezy „zgodności z typem rozkładu”, czyli

$$\mathrm{H}_0 : F = F_\theta \quad \text{dla pewnego } \theta,$$

gdzie $\{F_\theta\}$ jest daną parametryczną rodziną dystrybuant (rozkładów wykładniczych, gamma itp.).

Test niezależności chi-kwadrat. Przypuśćmy, że w pojedynczym doświadczeniu obserwujemy *parę* zmiennych losowych (X, Y) o wartościach w zbiorze $\{1, \ldots, r\} \times \{1, \ldots, s\}$ (te wartości należy traktować jako umowne „etykietki”). Rozkład prawdopodobieństwa jest opisany dwuwymiarową tabelką $(p_{ij}; i = 1, \ldots, r; j = 1, \ldots, s)$, gdzie $p_{ij} = \mathbb{P}(X = i, Y = j)$. Prawdopodobieństwa brzegowe $\mathbb{P}(X = i)$ i $\mathbb{P}(Y = j)$ zapiszemy w postaci

$$p_{i\bullet} = \sum_{j=1}^{s} p_{ij}, \qquad p_{\bullet j} = \sum_{i=1}^{r} p_{ij}.$$

Jeśli powtarzamy doświadczenie niezależnie n-krotnie, mamy niezależne wektory losowe $(X_1, Y_1), \ldots, (X_n, Y_n)$, każdy o rozkładzie takim jak (X, Y). Możemy zbudować dwuwymiarową tabelkę (N_{ij}), gdzie

$$N_{ij} = \sum_{k=1}^{n} \mathbb{1}(X_k = i, Y_k = j),$$

$$N_{ij} = \text{ liczba doświadczeń, które zakończyły się parą wyników } (i, j).$$

Jest to tak zwana *tablica kontyngencji*. Wielkości „brzegowe” w tej tabelce oznaczymy

$$N_{i\bullet} = \sum_{j=1}^{s} N_{ij}, \qquad N_{\bullet j} = \sum_{i=1}^{r} N_{ij}.$$

Rozpatrzymy hipotezę, która stwierdza *niezależność* zmiennych X i Y:

$$\mathrm{H}_0 : p_{ij} = p_{i\bullet} p_{\bullet j}, \quad (i = 1, \ldots, r; j = 1, \ldots, s).$$

Może na pierwszy rzut oka tego nie widać, ale jest to specjalny przypadek schematu rozpatrzonego w poprzednim punkcie: prawdopodobieństwa „klatek” p_{ij} są znanymi funkcjami nieznanego parametru

$$\theta = (p_{1\bullet}, \ldots, p_{r-1\bullet}, p_{\bullet 1}, \ldots, p_{\bullet s-1}).$$

Zauważmy, że $p_{r\bullet} = 1 - p_{1\bullet} - \cdots - p_{r-1\bullet}$ i $p_{\bullet s} = 1 - p_{\bullet 1} - \cdots - p_{\bullet s-1}$, więc tych ostatnich prawdopodobieństw brzegowych nie włączyliśmy do wektora θ. W rezultacie θ ma wymiar $r + s - 2$. Zbudujemy statystykę chi-kwadrat, używając *estymowanych* prawdopodobieństw klatek

$$\hat{p}_{ij} = \frac{N_{i\bullet} N_{\bullet j}}{n^2}.$$

Przekonajmy się, że są to estymatory *największej wiarogodności* w modelu określonym hipotezą H_0. Wygodnie nam będzie potraktować wiarogodność jako funkcję *rozszerzonego* wektora parametrów $(p_{1\bullet}, \ldots, p_{r\bullet}, p_{\bullet 1}, \ldots, p_{\bullet s})$ i znaleźć maksimum *warunkowe*, przy ograniczeniach $\sum_{i=1}^r p_{i\bullet} = 1$ i $\sum_{j=1}^s p_{\bullet j} = 1$. Logarytm wiarogodności ma postać

$$\ell(p_{1\bullet}, \ldots, p_{r\bullet}, p_{\bullet 1}, \ldots, p_{\bullet s}) = \mathrm{const} + \sum_{i,j} N_{ij} [\log p_{i\bullet} + \log p_{\bullet j}].$$

Zgodnie z metodą Lagrange'a poszukujemy minimum funkcji $l - a - b$, gdzie

$$a = \lambda \left(\sum_i p_{i\bullet} - 1 \right), \quad b = \mu \left(\sum_j p_{\bullet j} - 1 \right),$$

λ i μ są „mnożnikami Lagrange'a”. Ponieważ

$$\frac{\partial}{\partial p_{i\bullet}} (l - a - b) = \frac{N_{i\bullet}}{p_{i\bullet}} - \lambda, \qquad \frac{\partial}{\partial p_{\bullet j}} (l - a - b) = \frac{N_{\bullet j}}{p_{\bullet j}} - \mu,$$

więc przyrównując pochodne do zera dostajemy $p_{i\bullet} = N_{i\bullet}/\lambda$ i $p_{\bullet j} = N_{\bullet j}/\mu$. Wykorzystując postać ograniczeń, obliczamy $\lambda = \mu = n$ i ostatecznie otrzymujemy „oczywiste” estymatory $\hat{p}_{i\bullet} = \mathrm{MLE}(p_{i\bullet}) = N_{i\bullet}/n$ i $\hat{p}_{\bullet j} = \mathrm{MLE}(p_{\bullet j}) = N_{\bullet j}/n$. Wartość „oczekiwana” (przy założeniu H_0) w „klatce” (i, j) jest równa $n \hat{p}_{i\bullet} \hat{p}_{\bullet j} = N_{i\bullet} N_{\bullet j}/n$. Z tego co zostało powiedziane wynika, że statystyka chi-kwadrat przeznaczona do testowania hipotezy o niezależności jest następująca:

$$\chi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(N_{ij} - N_{i\bullet} N_{\bullet j}/n)^2}{N_{i\bullet} N_{\bullet j}/n}.$$

Dla $n \to \infty$, rozkład tej statystyki zmierza do rozkładu

$$\chi^2 \left((r-1)(s-1) \right)$$

na mocy Stwierdzenia 7.1.5, bo $rs - (r-1) - (s-1) - 1 = (r-1)(s-1)$.

7.1.7 Przykład. [Zależność gustów muzycznych i poglądów politycznych] Rzecznik partii A twierdzi, że wśród zwolenników tej partii, miłośnicy muzyki disco-polo, rockowej i symfonicznej występują w tych samych proporcjach, co w całej populacji wyborców. Przeprowadzono sondaż. Wśród wylosowanych 100 wyborców (spośród osób o jednoznacznie sprecyzowanych preferencjach muzycznych), wyniki badania były następujące:

	Popieram A	Nie popieram A	Razem
Słucham Disco Polo	25	10	35
Słucham muzyki rockowej	20	20	40
Słucham Disco Polo	25	10	35
Razem	60	40	100

Sformułowana przez rzecznika hipoteza stwierdza niezależność dwóch „cech" wyborcy: stosunku do partii A i preferencji muzycznych. Przeprowadzimy test niezależności chi-kwadrat na poziomie istotności 0.05. Statystyka testowa ma wartość

$$\begin{aligned}\chi^2 &= \frac{(25-60*35/100)^2}{60*35/100}+\frac{(20-60*40/100)^2}{60*40/100}+\frac{(15-60*25/100)^2}{60*25/100}\\ &+\frac{(10-40*35/100)^2}{40*35/100}+\frac{(20-40*40/100)^2}{40*40/100}+\frac{(10-40*25/100)^2}{40*25/100}\\ &= 2.53.\end{aligned}$$

Wartość krytyczną odczytujemy z tablicy rozkładu chi-kwadrat z $(2-1)(3-1)=2$ stopniami swobody. Ponieważ $\chi^2_{0.95}(2) = 5.99 > 2.53$, więc zebrane dane nie dają podstaw, żeby zarzucić rzecznikowi kłamstwo. $\Diamond$

Test Kołmogorowa-Smirnowa

Niech F będzie ustaloną, ciągłą dystrybuantą. Rozpatrzmy hipotezę stwierdzającą, że obserwacje $X_1, \ldots, X_n$ są próbką z rozkładu o dystrybuancie F. Symbolicznie,

$$\mathrm{H}_0 : X_1, \ldots, X_n \sim F.$$

Niech $\hat{F}_n$ będzie dystrybuantą empiryczną. Statystyka testowa jest równa

$$D_n = \sup_{-\infty<x<\infty} |\hat{F}_n(x) - F(x)|.$$

Łatwo zauważyć, że

$$D_n = \max(D_n^+, D_n^-),$$

gdzie
$$D_n^+ = \max_{i=1,\dots,n}\left(\frac{i}{n} - F(X_{i:n})\right), \qquad D_n^- = \max_{i=1,\dots,n}\left(F(X_{i:n}) - \frac{i-1}{n}\right).$$
Idea testu jest prosta. Odrzucamy H_0, jeśli odchylenia dystrybuanty empirycznej od dystrybuanty teoretycznej (sprecyzowanej w hipotezie) są duże, czyli gdy $D_n > c$. Sposób doboru poziomu krytycznego c opiera się na dwóch faktach. Pierwszy z nich jest prosty.

7.1.8 Stwierdzenie. *Dla dowolnego rozkładu ciągłego F, rozkład statystyki D_n jest taki sam.*

Naturalnie, mamy tu na myśli rozkład statystyki D_n przy prawdziwej hipotezie H_0.

Dowód. Dla uproszczenia przeprowadzimy dowód przy założeniu, że istnieje dobrze zdefiniowana funkcja odwotna F^{-1} do dystrybuanty. Rozpatrzmy próbkę $U_1, \dots, U_n$ z rozkładu *jednostajnego* $\mathrm{U}(0,1)$. Zauważmy, że
$$\sup_{0<u<1}\left|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(U_i \leqslant u) - u\right| = \sup_{-\infty<x<\infty}\left|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(U_i \leqslant F(x)) - F(x)\right|$$
$$= \sup_{-\infty<x<\infty}\left|\frac{1}{n}\sum_{i=1}^n \mathbb{1}(F^{-1}(U_i) \leqslant x) - F(x)\right|.$$
Wiemy, że zmienne losowe $X_i = F^{-1}(U_i)$ mają rozkład o dystrybuancie $F(x)$. Ze wzoru napisanego powyżej wynika, że rozkład statystyki D_n jest taki sam dla rozkładu $\mathrm{U}(0,1)$ i dla rozkładu o dystrybuancie F. □

7.1.9 Uwaga. Stwierdzenie 7.1.8 pozwala budować tablice rozkładu statystyki D_n i obliczać p-wartości. Potrzebna jest, w zasadzie, oddzielna tablica dla każdego n. Teoria asymptotyczna pozwala na jeszcze dalsze uproszczenie. Przy $n \to \infty$ mamy zbieżność statystyk $\sqrt{n}D_n \to_d K$ do pewnego rozkładu prawdopodobieństwa (rozkładu Kołmogorowa), który nie zależy od dystrybuanty F. Można zatem wybierać stałą $c > 0$ w taki sposób, aby $\mathbb{P}(\sqrt{n}D_n > c) \to \alpha$ (przy prawdziwej H_0).

7.1.10 Przykład (Testowanie generatora liczb losowych)**.** Chcemy sprawdzić, czy ciąg produkowany przez R-owską funkcję `runif()` zachowuje się jak ciąg niezależnych zmiennych losowych o rozkładzie jednostajnym $\mathrm{U}(0,1)$. Test Kołmogorowa jest jednym z podstawowych sprawdzianów. Dla ilustracji rozważmy 10-cio elementową „próbkę”:

(*) $\quad 0.4085, 0.5267, 0.3751, 0.8329, 0.0846, 0.8306, 0.6264, 0.3086, 0.3662, 0.7952$

(w rzeczywistości bada się znacznie dłuższe ciągi). Przeprowadzimy test hipotezy

$$\mathrm{H}_0 : (*) \text{ jest próbką losową z } \mathrm{U}(0,1)$$

na poziomie istotności 0.1. Dystrybuanta rozkładu jednostajnego jest równa $F(x) = x$ dla $0 < x < 1$. Sposób obliczania statystyki D_{10} pokazuje Rysunek 7.3 i następująca tabelka:

$X_{i:10}$	$(i-1)/10$	$i/10$	$i/10 - X_{i:10}$	$X_{i:10} - (i-1)/10$
0.0846	0.0	0.1	0.0154	0.0846
0.3086	0.1	0.2	−0.1086	0.2086
0.3662	0.2	0.3	−0.0662	0.1662
0.3751	0.3	0.4	0.0249	0.0751
0.4085	0.4	0.5	0.0915	0.0085
0.5267	0.5	0.6	0.0733	0.0267
0.6264	0.6	0.7	0.0736	0.0264
0.7952	0.7	0.8	0.0048	0.0952
0.8306	0.8	0.9	0.0694	0.0306
0.8329	0.9	1.0	0.1671	−0.0671

Widać, że $D_{10}^{+} = 0.1671$ i $D_{10}^{-} = 0.2086$ (największe liczby w dwóch ostatnich kolumnach). Stąd $D_{10} = 0.2086$. Wartość krytyczna statystyki D_{10}, odczytana z tablic, jest równa 0.369. Test Kołmogorowa nie odrzuca hipotezy, że nasz ciąg jest próbką z rozkładu $\mathrm{U}(0,1)$. Doświadczenie nie dało powodów, by kwestionować generator. $\Diamond$

Dwupróbkowy test Kołmogorowa-Smirnowa. Rozpatrzmy dwa niezależne ciągi obserwacji, $X_1, \ldots, X_n$ i $Y_1, \ldots, Y_m$. Zakładamy, że są to próbki z rozkładów o dystrybuantach, odpowiednio, F i G:

$$X_1, \ldots, X_n \sim F, \qquad Y_1, \ldots, Y_m \sim G.$$

Stawiamy hipotezę, że obie próbki pochodzą z *tego samego* rozkładu prawdopodobieństwa:

$$\mathrm{H}_0 : F = G.$$

Sens tej hipotezy jest taki, że dwa doświadczenia losowe nie różnią się w istotny sposób. Najczęściej oba doświadczenia przeprowadzamy w nieco odmiennych warunkach. Chcemy skontrolować przypuszczenie, że ta odmienność nie wpływa na przebieg zjawiska.

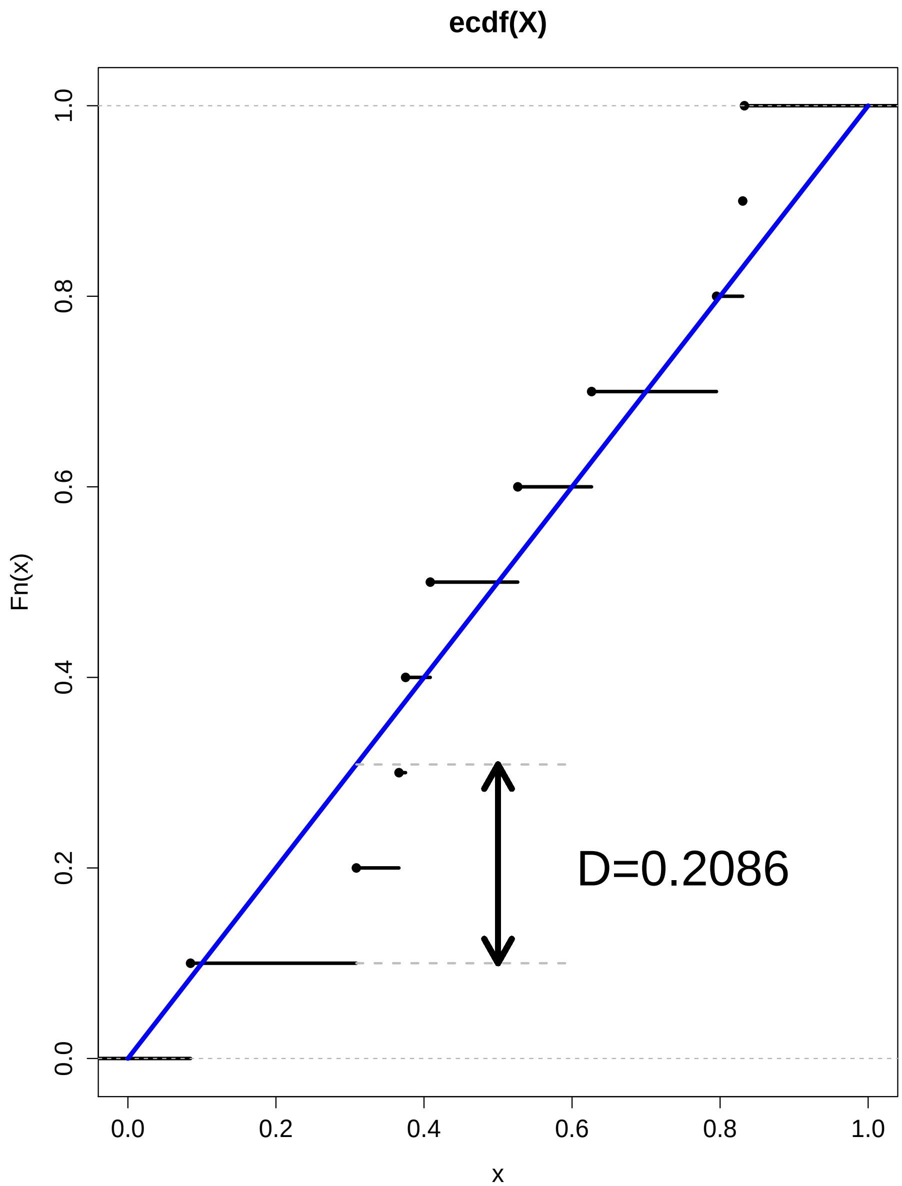
Idea dwupróbkowego testu Kołmogorowa-Smirnowa jest podobna do testu jednopróbkowego, omówionego poprzednio. Porównujemy teraz *dwie dystrybuanty empiryczne* – tę dla X-ów i tę dla Y-ów:

$$D_{n,m} = \sup_{-\infty < z < \infty} |\hat{F}_n(z) - \hat{G}_m(z)|,$$

gdzie

$$\hat{F}_n(z) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(X_i \leqslant z), \qquad \hat{G}_m(z) = \frac{1}{m} \sum_{i=1}^{m} \mathbb{1}(Y_i \leqslant z).$$

Jeśli hipoteza H_0 jest prawdziwa, to rozkład statystyki $D_{n,m}$ nie zależy od F pod warunkiem, że dystrybuanta F jest ciągła. Sposób postępowania jest podobny, jak dla testu jednopróbkowego. Odrzucamy H_0, jeśli $D_{n,m} > c$. W praktyce korzystamy z gotowej funkcji, która oblicza $D_{n,m}$ i (dokładną lub przybliżoną) p-wartość.



Rysunek 7.3: Dystrybuanta empiryczna i „hipotetyczna” dla 10-cio elementowej próbki.

7.1.11 Przykład. Zbadano iloraz inteligencji n uczniów szkoły „X” i m uczniów szkoły „Y”. Przykładowe dane, dla $n = 7$ i $m = 6$, wyglądają następująco:

X	110	112	115	98	130	123	141
Y	88	135	140	95	125	138	

Stawiamy hipotezę, że poziom inteligencji uczniów obu szkół jest taki sam. Matematycznie, hipoteza sprowadza się do przypuszczenia, że X-y i Y-i są próbkami losowymi *z tego samego* rozkładu prawdopodobieństwa.

Prześledźmy zastosowanie testu Kołmogorowa-Smirnowa na tych danych [2]. Sposób obliczania statystyki $D_{n,m}$ pokazuje następująca tabelka:

z		$\hat{F}_n(z)$	$\hat{G}_m(z)$	$\lvert\hat{F}_n(z) - \hat{G}_m(z)\rvert$
88	Y	$0/7$	$1/6$	$7/42$
95	Y	$0/7$	$2/6$	$14/42$
98	X	$1/7$	$2/6$	$8/42$
110	X	$2/7$	$2/6$	$2/42$
112	X	$3/7$	$2/6$	$4/42$
115	X	$4/7$	$2/6$	$10/42$
123	X	$5/7$	$2/6$	$16/42$
125	Y	$5/7$	$3/6$	$9/42$
130	X	$6/7$	$3/6$	$15/42$
135	Y	$6/7$	$4/6$	$8/42$
138	Y	$6/7$	$5/6$	$1/42$
140	Y	$6/7$	$6/6$	$6/42$
141	X	$7/7$	$6/6$	$0/42$

Widać, że $D_{7,6} = 16/42$. Test na poziomie istotności 0.05 odrzuca H_0 jeśli statystyka przekracza $30/42$ (to można odczytać z tablic). Konkluzja jest taka, że nie ma podstaw do odrzucenia hipotezy zerowej. Dane nie są sprzeczne z założeniem, że poziom inteligencji uczniów w obu szkołach ma ten sam rozkład prawdopodobieństwa. $\diamondsuit$

Test Kołmogorowa-Lilieforsa. Rozpatrzmy ponownie zagadnienie testowania zgodności z typem rozkładu. Hipotezę o *normalności* rozkładu,

$$\mathrm{H}_0 : F(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) \quad \text{dla pewnych } \mu, \sigma,$$

[2]W zasadzie, test Kołmogorowa-Smirnowa jest przeznaczony do stosowania dla danych pochodzących z rozkładów *ciągłych*. Nasze dane „szkolne” możemy *w przybliżeniu* uznać za ciągłe, bo nie zawierają powtarzających się wartości.

można testować przy pomocy zmodyfikowanej statystyki Kołmogorowa

$$D'_n = \sup_{-\infty<x<\infty} \left| \hat{F}_n(x) - \Phi\left(\frac{x-\bar{X}}{S}\right) \right|.$$

W tej formie, statystyka związana jest z nazwiskiem Lilieforsa. Rozpatrywana hipoteza zerowa jest złożona, to znaczy nie precyzuje jednego rozkładu prawdopodobieństwa, tylko całą rodzinę rozkładów. Dlatego rozkład statystyki jest inny, niż dla testu hipotezy prostej. Wartości krytyczne są stablicowane. Jak łatwo zgadnąć, są mniejsze od wartości krytycznych „zwykłego” testu Kołmogorowa-Smirnowa (tego dla prostej hipotezy zerowej). Sytuacja jest faktycznie znacznie gorsza, niż dla testu χ^2: rozkład statystyki Kołmogorowa-Lilieforsa zależy w istotny sposób od faktu, że obserwacje mają rozkład *normalny*. Porównaj, dla kontrastu, Stwierdzenie 7.1.5. Jeśli chcielibyśmy w podobny sposób testować (powiedzmy) hipotezę, że rozkład jest wykładniczy, to łatwo napisać odpowiednio zmodyfikowaną statystykę Kołmogorowa-Smirnowa – ale trudno ustalić wartości krytyczne.

7.2 Parametryczne testy istotności

Rozpatrzmy **model normalny**. Zakładamy, że $X_1, \ldots, X_n$ jest próbką z rozkładu $\mathrm{N}(\mu, \sigma^2)$ z nieznanymi paramerami μ i σ. Przedstawimy kilka typowych testów w tym modelu. Estymatory $\bar{X}$ i S^2 oznaczają to, co zwykle.

7.2.1. Test Studenta

Test $\mathrm{H}_0 : \mu = \mu_0$ przeciwko $\mathrm{H}_1 : \mu > \mu_0$, gdzie μ_0 jest ustaloną liczbą. Na poziomie istotności α, odrzucamy H_0, gdy

$$\sqrt{n}\frac{\bar{X}-\mu_0}{S} > t, \qquad t = \mathrm{t}_{1-\alpha}(n-1).$$

Test $\mathrm{H}_0 : \mu = \mu_0$ przeciwko $\mathrm{H}_1 : \mu \neq \mu_0$, gdzie μ_0 jest ustaloną liczbą. Na poziomie istotności α, odrzucamy H_0, gdy

$$\sqrt{n}\frac{|\bar{X}-\mu_0|}{S} > t, \qquad t = \mathrm{t}_{1-\alpha/2}(n-1).$$

Zajmiemy się teraz problemem porównania populacji, z których pochodzą dwie (niezależne) próbki. Niech $X_1, \ldots, X_n \sim \mathrm{N}(\mu_X, \sigma_X^2)$ i $Y_1, \ldots, Y_m \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$. Znaczenie symboli $\bar{X}$, $\bar{Y}$, S_X^2 i S_Y^2 jest oczywiste (i takie samo, jak w Przykładzie 2.2.6). △

7.2.2. Dwupróbkowy test Studenta.

Test $\mathrm{H}_0 : \mu_X = \mu_Y$ przeciwko $\mathrm{H}_1 : \mu_X > \mu_Y$. *Zakładamy*, że $\sigma_X^2 = \sigma_Y^2$. Testujemy więc hipotezę o równości wartości oczekiwanych, nie kwestionując założenia o równości wariancji. Odrzucamy H_0, gdy

$$\frac{\bar{X} - \bar{Y}}{\sqrt{(n-1)S_X^2 + (m-1)S_Y^2}} \sqrt{\frac{nm}{n+m}(n+m-2)} > t, \qquad t = \mathrm{t}_{1-\alpha}(n+m-2).$$

Pozostawiamy Czytelnikowi oczywistą modyfikację procedury testowania, gdy alternatywa jest dwustronna: $\mathrm{H}_0 : \mu_X = \mu_Y$ przeciwko $\mathrm{H}_1 : \mu_X \neq \mu_Y$. △

7.2.3. Testy dotyczące wariancji

Test $\mathrm{H}_0 : \sigma = \sigma_0$ przeciwko $\mathrm{H}_1 : \sigma > \sigma_0$, gdzie σ_0 jest ustaloną liczbą. Odrzucamy H_0, gdy

$$\frac{n-1}{\sigma_0^2} S^2 > c, \qquad c = \chi^2_{1-\alpha}(n-1).$$

Test $\mathrm{H}_0 : \sigma = \sigma_0$ przeciwko $\mathrm{H}_1 : \sigma \neq \sigma_0$, gdzie σ_0 jest ustaloną liczbą. Odrzucamy H_0, gdy

$$\frac{n-1}{\sigma_0^2} S^2 > c_2 \text{ lub } \frac{n-1}{\sigma_0^2} S^2 < c_1, \qquad c_1 = \chi^2_{\alpha/2}(n-1), c_2 = \chi^2_{1-\alpha/2}(n-1).$$

Zajmiemy się teraz problemem porównania populacji, z których pochodzą dwie (niezależne) próbki. Niech $X_1, \ldots, X_n \sim \mathrm{N}(\mu_X, \sigma_X^2)$ i $Y_1, \ldots, Y_m \sim \mathrm{N}(\mu_Y, \sigma_Y^2)$. Znaczenie symboli $\bar{X}$, $\bar{Y}$, S_X^2 i S_Y^2 jest oczywiste (i takie samo, jak w Przykładzie 2.2.6). △

7.3 Zadania

7.1. W Przykładzie 7.1.7 obliczyć p-wartość testu.

Wskazówka: Rozkład $\chi^2(2)$ jest rozkładem wykładniczym Ex(1/2).

7.2. Przeprowadzono sondaż wsród $n = 1000$ losowo wybranych pasażerów warszawskiego transportu publicznego (metra, tramwajów, autobusów). Pasażerowie oceniali, czy są zadowoleni z komfortu jazdy wybranym środkiem transportu. Otrzymano następujące dane:

$N_\mathrm{M} = 400$ osób wybrało metro i spośród nich $Z_\mathrm{M} = 300$ było zadowolonych;
$N_\mathrm{T} = 300$ osób wybrało tramwaj i spośród nich $Z_\mathrm{T} = 200$ było zadowolonych;
$N_\mathrm{A} = 300$ osób wybrało autobus i spośród nich $Z_\mathrm{A} = 150$ było zadowolonych.

(a) Zweryfikować hipotezę H_0 mówiącą, że frakcje osób niezadowolonych z komfortu jazdy są takie same dla wszystkich trzech środków transportu. Użyj testu niezależności χ^2.

(b) Wykazać, że statystyka testowa może być wyrażona następującym wzorem:

$$\chi^2 = \sum_i \frac{(Z_i - \hat{p})^2}{N_i \hat{p}(1-\hat{p})},$$

gdzie $\hat{p} = \sum_i Z_i/n$ i, oczywiście, $i \in \{\mathrm{M}, \mathrm{T}, \mathrm{A}\}$.

7.3. Interesuje nas procent absolwentów warszawskich wyższych uczelni, którzy znajdują pracę nie później niż pół roku po ukończeniu studiów. Zbadano reprezentatywną próbkę $n = 400$ absolwentów i okazało się, że wśród nich $S = 256$ znalazło pracę. Są to te same dane, co w Zadaniu 6.2 w Rozdziale 6.

(a) Weryfikujemy hipotezę H_0 mówiącą, że przynajmniej 70% absolwentów znajduje pracę przeciwko alternatywie, że mniej niż 70% absolwentów znajduje pracę. Przeprowadzić test na poziomie istotności $\alpha = 0.01$.

(b) Przeprowadzić test tej samej hipotezy zerowej przeciwko dwustronnej alternatywie, stwierdzającej, że procent absolwentów znajdujących pracę jest różny od 70%. Przyjąć poziom istotności $\alpha = 0.01$, jak poprzednio.

7.4. Zakładamy, że liczby wypadków w kolejnych tygodniach są niezależnymi zmiennymi losowymi o jednakowym rozkładzie (o możliwych wartościach $0, 1, 2, \ldots$). Chcemy zweryfikować hipotezę, mówiącą, że te zmienne losowe mają rozkład Poissona. Dane dotyczące 100 tygodni opisuje następująca tabelka:

liczba wypadków w tygodniu	0	1	2
Liczba tygodni	55	39	6

Przeprowadzić test χ^2 zgodności na poziomie istotności $\alpha = 0.05$.

Uwaga: Jest to zagadnienie „zgodności z typem rozkładu”. Potrzebna jest „dyskretyzacja” zbioru wartości obserwacji, który jest dla rozkładu Poissona nieskończony (mianowicie $\{0, 1, 2, \ldots\}$). Można podzielić ten zbiór na trzy podzbiory: $\{0\}$, $\{1\}$ i $\{2, 3, \ldots\}$. Odpowiadające temu prawdopodobieństwa są równe $p_0(\theta) = \mathrm{e}^{-\theta}$, $p_1(\theta) = \theta\mathrm{e}^{-\theta}$ i $p_{\geqslant 2}(\theta) = 1 - p_0(\theta) - p_1(\theta)$ (należy pamiętać, że w teście χ^2 prawdopodobieństwa teoretyczne muszą się sumować do 1).

7.5. Zważono 10 paczek masła i otrzymano następujące wyniki:

$$245;\ 248;\ 241;\ 251;\ 252;\ 244;\ 246;\ 248;\ 247;\ 248.$$

Są to te same dane co w Zadaniu 6.1 z Rozdziału 6. Zakładamy, że jest to próbka losowa z rozkładu normalnego $\mathrm{N}(\mu, \sigma^2)$ z nieznanymi parametrami μ i σ.

(a) Przeprowadzić test hipotezy H_0 stwierdzającej, że średnia waga μ masy paczki jest równa 250 przeciwko alternatywie, że jest mniejsza. Przyjmujemy poziom istotności $\alpha = 0.05$.

(b) Przeprowadzić test hipotezy H_0 stwierdzającej, że średnia waga μ masy paczki jest równa 250 przeciwko alternatywie, że jest inna. Przyjmujemy poziom istotności $\alpha = 0.05$.

(c) Przeprowadzić test hipotezy H_0 stwierdzającej, że odchylenie standardowe σ masy paczki nie przekracza 2 przeciwko alternatywie, że jest większe. Przyjmujemy poziom istotności $\alpha = 0.05$.

7.6. Porównywano średnie ceny 1 kg truskawek w dwu różnych miastach. W każdym z miast zebrano informacje z 6 losowo wybranych punktów sprzedaży. Po dokonaniu obliczeń otrzymano następujące wartości średnich cen i wariancji w obu miastach:

	średnia $\bar{x}$	estymator wariancji s^2
Miasto I	2.5	0.25
Miasto II	3.1	0.29

Estymatory wariancji s^2 są wyznaczone (dla każdej z dwóch próbek z osobna) według wzoru $s^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2/(n-1)$, gdzie n jest rozmiarem próbki. Zakładamy, że obserwowane zmienne losowe są niezależne i mają rozkłady normalne o tej samej wariancji.

(a) Przeprowadzić test dla weryfikacji hipotezy H_0 mówiącej, że przeciętne ceny są jednakowe w obu miastach, przy alternatywie H_1 mówiącej, że w mieście II przeciętna cena 1 kg truskawek jest wyższa. Przyjąć poziom istotności $\alpha = 0.05$.

(b) Przeprowadzić test dla weryfikacji hipotezy H_0 mówiącej, że przeciętne ceny są jednakowe w obu miastach, przy alternatywie H_1 mówiącej, że przeciętne są różne. Przyjąć poziom istotności $\alpha = 0.05$.

Rozdział 8

Teoria testowania hipotez

Przeanalizujemy zagadnienie testowania hipotez statystycznych w sposób bardziej formalny. Nasze rozważania będą miały dość ogólny i abstrakcyjny charakter. Punktem wyjścia jest przestrzeń statystyczna $(\mathcal{X}, \mathcal{F}, \{\mathbb{P}_\theta, \theta \in \Theta\})$.

8.1 Definicje

Hipotezy statystyczne, czyli wypowiedzi na temat rozkładu prawdopodobieństwa, utożsamiamy z *podzbiorami przestrzeni parametrów* Θ (tak, jak zdarzenia losowe utożsamiamy z podzbiorami przestrzeni próbkowej Ω). Mówiąc o zagadnieniu testowania, zawsze będziemy rozważali *hipotezę zerową*

$$\mathrm{H}_0 : \theta \in \Theta_0$$

i *hipotezę alternatywną*

$$\mathrm{H}_1 : \theta \in \Theta_1.$$

Zakładamy, że Θ_0 i Θ_1 są podzbiorami przestrzeni Θ takimi, że $\Theta_0 \cap \Theta_1 = \emptyset$. Znaczy to, że H_0 i H_1 wzajemnie się wykluczają. Obie „konkurencyjne" hipotezy traktujemy nierównoprawnie. Zasadniczo, interpretacja jest taka: H_0 jest założeniem obowiązującym do czasu, gdy pojawią się dane doświadczalne sprzeczne (lub raczej „bardzo trudne do pogodzenia") z tą hipotezą. Z kolei, H_1 jest „ewentualnością, z którą powinniśmy się liczyć", jeśli przyjdzie nam zrezygnować z hipotezy H_0. *Testem* hipotezy H_0 przeciw alternatywie H_1 nazywamy statystykę

$$\delta : \mathcal{X} \to \{0, 1\}.$$

Wartość „1" interpretujemy jako decyzję o odrzuceniu H_0, zaś „0" oznacza, że nie odrzucamy H_0. Zbiór $K = \{x \in \mathcal{X} : \delta(x) = 1\}$ nazywamy *obszarem krytycznym* testu. Przeważnie test ma postać $\delta(X) = \mathbb{1}(T(X) > c)$, gdzie $T(X)$ jest pewną „wygodną" statystyką, zwaną *statystyką testową*, zaś c jest liczbą zwaną *wartością krytycznym.* Tak więc,

$$\begin{aligned} &\text{jeśli } T(X) > c \text{ czyli } \delta(X) = 1 \text{ czyli } X \in K, \text{ to odrzucamy } \mathrm{H}_0; \\ &\text{jeśli } T(X) \leqslant c \text{ czyli } \delta(X) = 0 \text{ czyli } X \notin K, \text{ to nie odrzucamy } \mathrm{H}_0. \end{aligned}$$

Oczywiście, jest tylko sprawą wygody, czy określamy funkcję δ, T i c, czy zbiór K: są to trzy równoważne sposoby opisu testu. Istotne jest to, że test jest *regułą podejmowania decyzji* w zależności od wyniku obserwacji. Ponieważ obserwacje są losowe, musimy czasem popełniać błędy. Skutki działania testu przedstawimy tak:

stan rzeczy \ decyzja	$\delta = 0$	$\delta = 1$
H_0 prawdziwa	O.K.	błąd I rodzaju
H_1 prawdziwa	błąd II rodzaju	O.K.

Będziemy również mówić, że decyzja „0" oznacza akceptację H_0 a decyzja „1" – akceptację H_1. Niektórzy statystycy starannie unikają takich sformułowań. To jednak kwestia interpretacji decyzji. Statystyka nie zajmuje się skutkami błędnych decyzji, tylko bada *częstość* podejmowania takich decyzji. Niech

$$1 - \beta(\theta) = \mathbb{P}_\theta\left(\delta(X) = 1\right)$$

oznacza prawdopodobieństwo odrzucenia H_0. Nazwiemy $1 - \beta(\theta)$ *funkcją mocy* testu δ. Interpretacja tej funkcji jest odmienna dla $\theta \in \Theta_0$ i $\theta \in \Theta_1$:

Jeśli $\theta \in \Theta_0$ to

$$1 - \beta(\theta) = \mathbb{P}_\theta\left(\delta(X) = 1\right) \text{ jest prawdopodobieństwem błędu I rodzaju.}$$

Jeśli $\theta \in \Theta_1$ to

$$\beta(\theta) = \mathbb{P}_\theta\left(\delta(X) = 0\right) \text{ jest prawdopodobieństwem błędu II rodzaju.}$$

Za klasyczną teorią testowania stoi ważna idea metodologiczna. Jest to zasada konserwatyzmu: nie należy rezygnować z ustalonej teorii (hipotezy zerowej), jeśli nie ma po temu koniecznych lub przynajmniej bardzo wyraźnych powodów. Wobec tego staramy się przede wszystkim *kontrolować prawdopodobieństwo błędu I rodzaju.* Interesować nas będą testy, dla których prawdopodobieństwo błędu I rodzaju nie przekracza zadanej z góry, małej liczby. Spośród takich i tylko takich testów postaramy się wybrać te, które mają możliwie małe prawdopodobieństwo błędu II rodzaju.

Jest jeszcze inny powód kontroli błędu I rodzaju. Często w zastosowaniach za H_0 przyjmuje się hipotezę, której błędne odrzucenie ma poważniejsze skutki niż błędne jej przyjęcie. O przykłady takich sytuacji praktycznych nie jest trudno. Wyobraźmy sobie, że zadaniem statystyka jest porównanie skuteczności dwóch leków: A i B. Przypuśćmy, że lek A jest od dawna stosowany, jego działanie i skutki uboczne są dobrze znane. Lek B jest nowy i jego stosowanie wiąże się z pewnym ryzykiem. W takiej sytuacji statystyk powinien za hipotezę zerową przyjąć H_0: „lek A jest nie mniej skuteczny, niż B". W istocie, błąd I rodzaju polegający na pochopnym odrzuceniu H_0 może narazić pacjentów na niebezpieczeństwo – tego staramy się przede wszystkim uniknąć. Błąd II rodzaju może tylko trochę opóźnić postęp w metodach leczenia i jest mniej groźny w skutkach.

8.1.1 DEFINICJA. *Mówimy, że δ jest testem* ***na poziomie istotności*** *α, jeśli*

$$\sup_{\theta \in \Theta_0} \mathbb{P}_\theta\left(\delta(X) = 1\right) \leqslant \alpha.$$

Za poziom istotności α przyjmuje się zwykle „okrągłą, małą" liczbę, na przykład $\alpha = 0.01$ lub $\alpha = 0.05$. Zgodnie z tym co powiedzieliśmy wcześniej, wybór poziomu istotności odzwierciedla nasz „stopień konserwatyzmu": im bardziej przywiązani jesteśmy do wyjściowej hipotezy H_0, tym mniejsze wybieramy α.

Załóżmy, że test ma postać $\delta(X) = \mathbb{1}(T(X) > c)$. **Poziom krytyczny** testu lub inaczej p-**wartość** jest to n*ajmniejszy poziom istotności przy którym odrzucamy* H_0. Formalnie, jeśli zaobserwujemy $x \in \mathcal{X}$, to

$$P = \sup_{\theta \in \Theta_0} \mathbb{P}_\theta(T(X) > T(x)).$$

8.1.2 DEFINICJA. *Test δ^* jest* ***jednostajnie najmocniejszy*** *na poziomie istotności α, jeśli*

(i) δ^* jest testem na poziomie istotności α;

(ii) dla każdego testu δ na poziomie istotności α, mamy

$$\mathbb{P}_\theta\left(\delta^*(X) = 1\right) \geqslant \mathbb{P}_\theta\left(\delta(X) = 1\right)$$

dla każdego $\theta \in \Theta_1$.

W skrócie, δ^* jest UMPT (hipotezy $\mathrm{H}_0 : \theta \in \Theta_0$ przeciw alternatywie $\mathrm{H}_1 : \theta \in \Theta_1$ na poziomie istotności α).

8.2 Lemat Neymana-Pearsona

Niech θ_0 i θ_1 będą ustalonymi punktami przestrzeni parametrów Θ. Rozważymy dwie *hipotezy proste*:

$$\mathrm{H}_0 : \theta = \theta_0, \qquad \mathrm{H}_1 : \theta = \theta_1.$$

Znaczy to, że $\Theta_0 = \{\theta_0\}$ i $\Theta_1 = \{\theta_1\}$ są zbiorami jednopunktowymi. Niech $\mathbb{P}_0$ i $\mathbb{P}_1$ oznaczają rozkłady prawdopodobieństwa na przestrzeni próbkowej, odpowiadające wartościom parametru θ_0 i θ_1. Załóżmy, że te rozkłady mają gęstości f_0 i f_1. Niech X oznacza wektor obserwacji. Innymi słowy, obie hipotezy możemy sformułować tak:

$$\mathrm{H}_0 : X \sim f_0, \qquad \mathrm{H}_1 : X \sim f_1.$$

8.2.1 TWIERDZENIE (Lemat Neymana - Pearsona)**.** *Niech*

$$K^* = \left\{ x \in \mathcal{X} : \frac{f_1(x)}{f_0(x)} > c \right\}. \tag{8.2.2}$$

Załóżmy, że $\mathbb{P}_0(K^*) = \alpha$ i $\mathbb{P}_1(K^*) = \beta$. Jeśli $K \subset \mathcal{X}$ jest takim zbiorem, że $\mathbb{P}_0(K) \leqslant \alpha$, to $\mathbb{P}_1(K) \leqslant \beta$.

Dowód. Ponieważ $\int_{K^*} f_0 \geqslant \int_K f_0$, więc $\int_{K^* \setminus K} f_0 \geqslant \int_{K \setminus K^*} f_0$. Pomnóżmy tę ostatnią nierówność obustronnie przez c i skorzystajmy z faktu, że $f_1 > cf_0$ na zbiorze $K^* \setminus K$ i $cf_0 \geqslant f_1$ na $K \setminus K^*$. Otrzymujemy $\int_{K^* \setminus K} f_1 \geqslant \int_{K \setminus K^*} f_1$ i stąd

$$\int_{K^*} f_1 \geqslant \int_K f_1.$$

Dodajmy, że nierówność jest *ostra*, jeśli tylko $\int_{K^* \setminus K} f_0 > 0$ i $c > 0$. □

Popatrzmy na zbiory K^* i K jak na obszary krytyczne testów δ^* i δ. Z oczywistych względów, δ^* nazywamy *testem ilorazu wiarogodności*, po angielsku: *Likelihood Ratio Test* (LRT). Lemat Neymana - Pearsona stwierdza, że test ilorazu wiarogodności jest *najmocniejszy* na poziomie istotności α (przymiotnik „jednostajnie" możemy opuścić, bo hipoteza alternatywna jest prosta).

Uwaga. Przypuśćmy, że f_1 i f_2 oznaczają gęstości prawdopodobieństwa w „zwykłym" sensie, to znaczy obserwacja X jest zmienną losową „typu ciągłego". Jeśli α jest zadaną z góry liczbą $(0 < \alpha < 1)$, to *zazwyczaj* można dobrać poziom krytyczny c testu ilorazu wiarogodności tak, żeby $\mathbb{P}_0(f_1(X)/f_0(X) > c) = \alpha$. Nie będziemy tego stwierdzenia ani uściślać ani dowodzić, ale zobaczymy wkrótce na kilku przykładach jak to się robi. Z drugiej strony, „gęstości" w Twierdzeniu 8.2.1 mogą oznaczać dyskretne funkcje prawdopodobieństwa: $f_i(x) = \mathbb{P}_i(X = x)$. Dowód jest taki sam, tylko całki trzeba zamienić na sumy. Lemat Neymana - Pearsona pozostaje prawdziwy dla obserwacji typu dyskretnego, ale przy jego stosowaniu pojawiają się drobne trudności. Dla danego α, może nie istnieć c takie, że $\mathbb{P}_0(f_1(X)/f_0(X) > c) = \alpha$. Wtedy najmocniejszy test na poziomie istotności α może nie być testem ilorazu wiarogodności. Pewnym wyjściem jest użycie tak zwanych *testów zrandomizowanych*, w których uzależnia się decyzję od przypadku. W naszych rozważaniach, w szczególności w Definicji 8.1.2, ograniczyliśmy się dla uproszczenia do testów niezrandomizowanych. Dokładniejsza dyskusja wymagałaby więc bardziej ogólnej definicji testu. Nie będziemy się tym zajmować.

Wspomnijmy, że obszar krytyczny testu ilorazu wiarogodności można (i zazwyczaj wygodnie jest) napisać w równoważnej, „zlogarytmowanej” postaci $K^* = \{x : \log f_1(x) - \log f_0(x) > \log c = \text{const}\}$. Wprowadźmy umowę, że „const” jest „ogólnym” oznaczeniem poziomu krytycznego i może w trakcie rachunków reprezentować *różne* liczby.

8.2.3 Przykład (Model normalny, test „jednostronny”)**.** Załóżmy, że mamy próbkę $X_1, \dots, X_n$ z rozkładu $\mathrm{N}(\mu, \sigma^2)$, gdzie σ^2 jest znane. Rozważmy najpierw dwie hipotezy *proste*: $\mathrm{H}_0 : \mu = \mu_0$ przeciw $\mathrm{H}_1 : \mu = \mu_1$, gdzie $\mu_0 < \mu_1$. Zbudujemy test ilorazu wiarogodności. Niech

$$f_h(x_1, \dots, x_n) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{1}{2\sigma^2}(x_i - \mu_h)^2\right],$$

dla $h = 0, 1$. Mamy więc

$$\begin{aligned}\log f_1(X_1, \dots, X_n) &- \log f_0(X_1, \dots, X_n) \\ &= -\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu_1)^2 + \frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu_0)^2 \\ &= \frac{1}{2\sigma^2}\left[2\sum_{i=1}^{n} X_i(\mu_1 - \mu_0) + n(\mu_0^2 - \mu_1^2)\right].\end{aligned}$$

Test ilorazu wiarogodności odrzuca H_0 na rzecz H_1, jeśli powyższe wyrażenie przekracza pewną stałą. Ale tak się dzieje wtedy i tylko wtedy, gdy $\bar{X} > c$ dla pewnej (innej) stałej. Dopiero teraz dokończymy naszą konstrukcję, stosownie wybierając c. Niech $c = z\sigma/\sqrt{n}$, gdzie z jest kwantylem rzędu $1-\alpha$ standardowego rozkładu normalnego, czyli $\Phi(z) = 1 - \alpha$:

$$\text{odrzucamy } \mathrm{H}_0, \text{ jeśli } \bar{X} > \mu_0 + \frac{z\sigma}{\sqrt{n}}.$$

Niech $1-\beta(\mu)$ będzie funkcją mocy tego testu. Nasz wybór stałej c zapewnia, że $1-\beta(\mu_0) = \alpha$. Z lematu N-P wnioskujemy, że jest to *najmocniejszy* test $\mathrm{H}_0 : \mu = \mu_0$ przeciw $\mathrm{H}_1 : \mu = \mu_1$ na poziomie istotności α. Mówiliśmy na razie o dwóch hipotezach prostych, ale ten sam test jest również UMPT $\mathrm{H}_0 : \mu \leqslant \mu_0$ przeciw $\mathrm{H}_1 : \mu > \mu_0$, na poziomie istotności α. Żeby to uzasadnić, posłużymy się bardzo prostą argumentacją. Jeśli test ma poziom istotności α, to w szczególności $1 - \beta(\mu_0) \leqslant \alpha$. Dla *dowolnego* $\mu_1 > \mu_0$, spośród testów spełniających warunek $1 - \beta(\mu_0) \leqslant \alpha$, nasz test ma największą moc w punkcie μ_1. Ale to znaczy, że jest on *jednostajnie* najmocniejszym testem przeciw alternatywie $\mu > \mu_0$.

Na koniec przyjrzyjmy się funkcji mocy naszego UMPT:

$$1 - \beta(\mu) = 1 - \Phi\left(z - \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right). \tag{8.2.4}$$

◊

Rodziny z monotonicznym ilorazem wiarogodności. W Przykładzie 8.2.3, konstrukcja UMPT opierała się na spostrzeżeniu, że najmocniejszy test test hipotezy prostej μ_0 przeciw μ_1 na poziomie istotności α jest *taki sam* dla wszystkich alternatyw $\mu_1 > \mu_0$. To rozumowanie można uogólnić. Jeśli mamy rodzinę gęstości $\{f_\theta\}$ taką, że dla wszystkich $\theta_0 < \theta_1$ iloraz $f_{\theta_1}(x)/f_{\theta_0}(x)$ jest rosnącą funkcją pewnej statystyki $T(x)$ to mówimy o rodzinie z *monotonicznym ilorazem wiarogodności*. W zagadnieniu „jednostronnym" $\mathrm{H}_0 : \theta \leqslant \theta_0$ przeciw $\mathrm{H}_1 : \theta > \theta_0$, UMPT na poziomie istotności α ma postać $T(X) > c$, gdzie c jest stałą, dobraną tak żeby $\mathbb{P}_{\theta_0}(T(X) > c) = \alpha$.

Testy nieobciążone. Dla rodziny z monotonicznym ilorazem wiarogodności, w zagadnieniu „dwustronnym" $\mathrm{H}_0 : \theta = \theta_0$ przeciw $\mathrm{H}_1 : \theta \neq \theta_0$, UMPT po prostu *nie istnieje*. Rzeczywiście, test na poziomie istotności α *musi* być postaci $T(X) > c_1$, żeby być najmocniejszy w punkcie $\theta_1 > \theta_0$ i *musi* być postaci $T(X) < c_2$, żeby być najmocniejszy w punkcie $\theta_2 < \theta_0$. Powstaje podobny kłopot jak w teorii estymacji. Mamy do dyspozycji różne testy, z których żaden nie jest (uniwersalnie) najlepszy. Pewnym wyjściem jest ograniczenie rozważań do tak zwanych testów *nieobciążonych*.

8.2.5 DEFINICJA. *Rozważmy zagadnienie testowania* $\mathrm{H}_0 : \theta \in \Theta_0$ *przeciw* $\mathrm{H}_1 : \theta \in \Theta_1$. *Test* δ *nazywamy* ***nieobciążonym****, jeśli dla dowolnych* $\theta_0 \in \Theta_0$ *i* $\theta_1 \in \Theta_1$ *mamy*

$$\mathbb{P}_{\theta_1}(\delta(X) = 1) \geqslant \mathbb{P}_{\theta_0}(\delta(X) = 1).$$

Innymi słowy, dla testu nieobciążonego moc nigdzie nie spada poniżej poziomu istotności. Jest to bardzo naturalne wymaganie.

8.2.6 Przykład (Model normalny, test „dwustronny")**.** W modelu z Przykładu 8.2.3, rozważmy zagadnienie testowania

$$\mathrm{H}_0 : \mu = \mu_0 \text{ przeciw } \mathrm{H}_1 : \mu \neq \mu_0.$$

UMPT, jak wiemy, *nie istnieje*. Rozważmy test

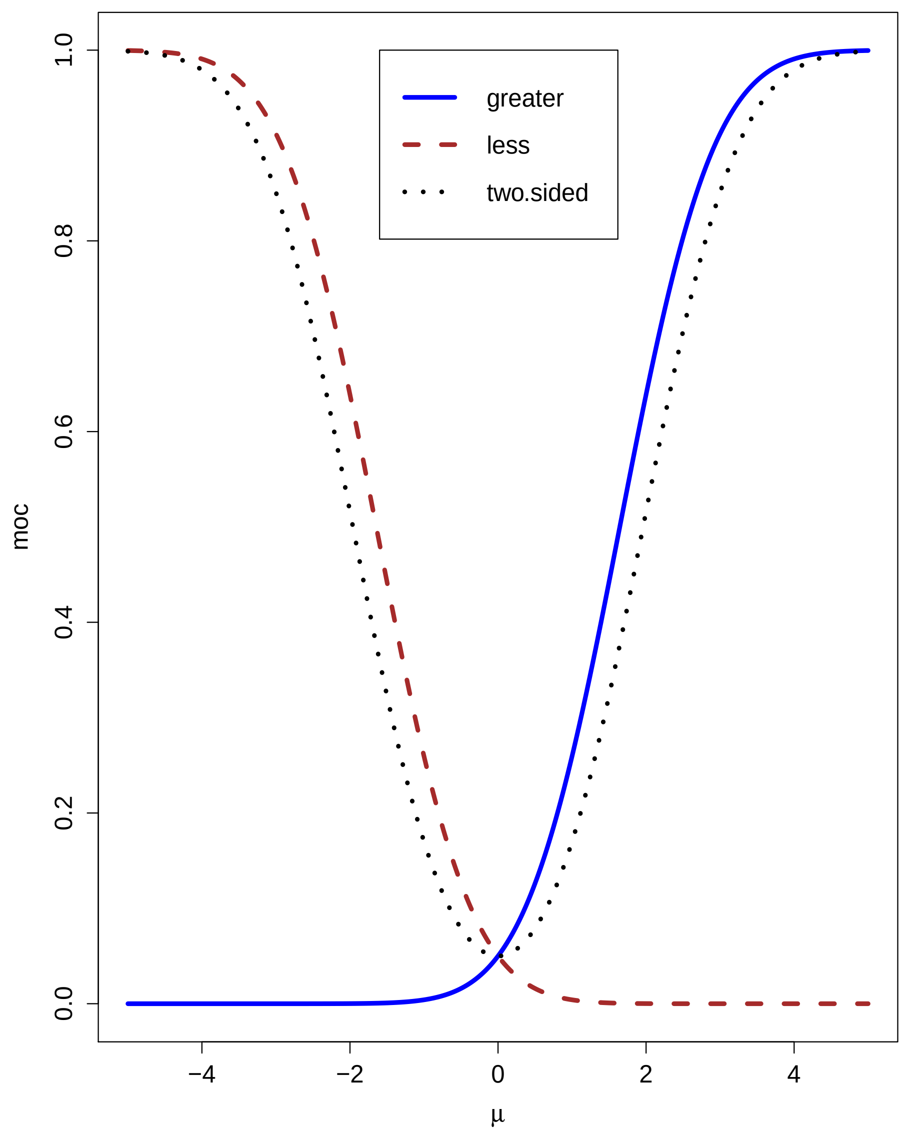
$$\text{odrzucamy } \mathrm{H}_0, \text{ jeśli } |\bar{X} - \mu_0| > \frac{z\sigma}{\sqrt{n}},$$

gdzie z jest kwantylem rzędu $1 - \alpha/2$, czyli $\Phi(z) = 1 - \alpha/2$. Funkcja mocy rozważanego teraz testu jest następująca:

$$1 - \beta(\mu) = 1 - \Phi\left(z - \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right) + \Phi\left(-z - \sqrt{n}\frac{\mu - \mu_0}{\sigma}\right).$$

Zauważmy, że $1 - \beta(\mu_0) = \alpha$ i $1 - \beta(\mu) > \alpha$ dla $\mu \neq \mu_0$. Zatem nasz test jest testem *nieobciążonym* $\mathrm{H}_0 : \mu \leqslant \mu_0$ przeciw $\mathrm{H}_1 : \mu > \mu_0$ na poziomie istotności α. W istocie, można pokazać więcej: jest to test najmocniejszy *spośród testów nieobciążonych* na ustalonym poziomie istotności.

Funkcje mocy dwóch testów jednostronnych $\mathrm{H}_0 : \mu = 0$ (przeciw $\mathrm{H}_1 : \mu > 0$ i przeciw $\mathrm{H}_1 : \mu < 0$, Przykład 8.2.3) oraz testu dwustronnego (przeciw $\mathrm{H}_1 : \mu \neq 0$, Przykład 8.2.6) są pokazane na Rysunku 8.1. Przyjęty poziom istotności jest równy $\alpha = 0.05$. $\Diamond$



Rysunek 8.1: Funkcje mocy testów jednostronnych i testu dwustronnego.

8.3 Testy ilorazu wiarogodności

Rozważmy dwa modele statystyczne (na tej samej przestrzeni obserwacji). W modelu pierwszym mamy do czynienia z rodziną rozkładów prawdopodobieństwa o gęstościach $f_0(\theta_0, x)$, gdzie $\theta_0 \in \Theta_0$, zaś w modelu drugim mamy gęstości $f_1(\theta_1, x)$, gdzie $\theta_1 \in \Theta_1$. Chcemy zdecydować, który z dwóch konkurujących modeli wybrać do opisu doświadczenia losowego, w którym pojawiła się obserwacja X. Zgodnie z klasycznym schematem testowania hipotez, oba modele nie będą traktowane równoprawnie: pierwszy z nich odgrywa rolę „hipotezy zerowej", z której nie mamy ochoty rezygnować bez wyraźnej konieczności. Rozważamy zatem zagadnienie testowania

$$\mathrm{H}_0 : X \sim f_0(\theta_0, \cdot) \textit{ dla pewnego } \theta_0 \in \Theta_0$$

przeciw

$$\mathrm{H}_1 : X \sim f_1(\theta_1, \cdot) \textit{ dla pewnego } \theta_1 \in \Theta_1.$$

W istocie, jesteśmy w podobnej sytuacji jak w Lemacie Neymana - Pearsona z tą różnicą, że dwa konkurencyjne modele są *statystyczne*, a nie *probabilistyczne*. Modele zawierają nieznane parametry (θ_0 i θ_1). Rzecz jasna, to bardzo komplikuje sprawę i nie możemy bezpośrednio odwołać się do Lematu N-P. Spróbujemy naśladować ideę testu ilorazu wiarogodności, z koniecznymi modyfikacjami. Za statystykę testową przyjmiemy

$$\lambda = \frac{\sup_{\theta_1 \in \Theta_1} f_1(\theta_1, X)}{\sup_{\theta_0 \in \Theta_0} f_0(\theta_0, X)}.$$

Innymi słowy,

$$\lambda = \frac{f_1(\hat{\theta}_1, X)}{f_0(\hat{\theta}_0, X)},$$

gdzie $\hat{\theta}_0 = \hat{\theta}_0(X)$ i $\hat{\theta}_1 = \hat{\theta}_1(X)$ są *estymatorami największej wiarogodności* odpowiednio w pierwszym i drugim modelu. Odrzucamy H_0, jeśli $\lambda > c$, gdzie c jest pewną stałą. Konstrukcję tego testu można nieformalnie uzasadnić tak: porównujemy „największą szansę otrzymania obserwacji X, gdy prawdziwa jest hipoteza H_1" i „największą szansę, gdy prawdziwa jest hipoteza H_0". Dostatecznie duża wartość ilorazu tych największych szans sugeruje odrzucenie H_0 na rzecz H_1.

8.3.1. Modele „zagnieżdżone". Rozważmy model statystyczny opisany przez rodzinę gęstości prawdopodobieństwa $\{f_\theta : \theta \in \Theta\}$ i „zanurzony" w nim mniejszy model $\{f_\theta : \theta \in \Theta_0\}$, gdzie $\Theta_0 \subset \Theta$. Interesuje nas pytanie: czy możemy założyć, że mniejszy model Θ_0 opisuje poprawnie doświadczenie losowe, czy potrzebny jest większy model Θ? Formalnie, zagadnienie polega na zbudowaniu testu hipotezy $\mathrm{H}_0 : \theta \in \Theta_0$ przeciw alternatywie H_1 :

$\theta \in \Theta \backslash \Theta_0 = \Theta_1$. Możemy wykorzystać test ilorazu wiarogodności opisany w poprzednim punkcie. Skupimy się na najbardziej typowej sytuacji, kiedy mniejszy model jest określony przez *nałożenie dodatkowych ograniczeń w postaci układu równań.* Załóżmy, że $\Theta \subset \mathbb{R}^d$, dana jest funkcja $h : \mathbb{R}^d \to \mathbb{R}^p$ i $\Theta_0 = \{\theta \in \Theta : h(\theta) = 0\}$. Zauważmy, że równanie $h(\theta) = 0$ jest faktycznie układem p równań z d „niewiadomymi": współrzędnymi wektora $\theta = (\theta_1, \ldots, \theta_d)$. Jest raczej jasne, że w „typowej" sytuacji Θ_0 jest podzbiorem „$d-p$-wymiarowym" zbioru „d-wymiarowego" Θ. W konsekwencji, zazwyczaj mamy $\sup_{\theta\in\Theta_1} f(\theta, X) = \sup_{\theta\in\Theta} f(\theta, X)$. Możemy statystykę ilorazu wiarogodności napisać w nieco prostszej postaci

$$\lambda = \frac{\sup_{\theta\in\Theta} f(\theta, X)}{\sup_{\theta\in\Theta_0} f(\theta, X)}.$$

W liczniku mamy maksimum *bezwarunkowe*, po całej przestrzeni parametrów Θ. To maksimum jest osiągane w punkcie $\hat{\theta}$, czyli w $\mathrm{MLE}(\theta)$ obliczonym dla większego modelu. W mianowniku jest maksimum *warunkowe*. Punkt, w którym to maksimum jest osiągane oznaczamy przez $\dot{\theta}$ i nazywamy *estymatorem największej wiarogodności z ograniczeniami*. Oczywiście $\dot{\theta}$ jest $\mathrm{MLE}(\theta)$ w mniejszym modelu i mamy $h(\dot{\theta}) = 0$. Reasumując, w zagadnieniu

$$\mathrm{H}_0 : h(\theta) = 0 \text{ przeciw } \mathrm{H}_1 : h(\theta) \neq 0.$$

statystyka ilorazu wiarogodności przybiera postać

$$\lambda = \frac{f(\hat{\theta}, X)}{f(\dot{\theta}, X)},$$

gdzie

$$\begin{aligned} f(\hat{\theta}) &= \sup_{\theta} f(\theta, X), \\ f(\dot{\theta}) &= \sup_{\theta : h(\theta)=0} f(\theta, X), \qquad (h(\dot{\theta}) = 0). \end{aligned}$$

$\triangle$

Faktycznie testy „dwustronne" w modelu normalnym, o których mówiliśmy w podrozdziale 7.2, są testami ilorazu wiarogodności dla modeli zagnieżdżonych. Rozpatrzmy jeden z przykładów.

8.3.2 Przykład. Test t-Studenta. Rozważmy próbkę $X_1, \ldots, X_n$ z rozkładu normalnego $\mathrm{N}(\mu, \sigma^2)$ z nieznanymi parametremi μ i σ. Dla danej liczby μ_0 testujemy hipotezę

$$\mathrm{H}_0 : \mu = \mu_0$$

przeciw alternatywie

$$\mathrm{H}_1 : \mu \neq \mu_0.$$

Zastosujmy schemat *modeli zagnieżdżonych*. Przestrzeń parametrów jest 2-wymiarowa: $\Theta = \{(\mu, \sigma) : \sigma > 0\}$. Podmodel, odpowiadający naszej hipotezie H_0 jest jest zadany przez jedno równańie i jest jednowymiarowy: $\Theta_0 = \{(\mu, \sigma) : \mu = \mu_0, \sigma > 0\}$. Zbudujemy test ilorazu wiarogodności. Wiarogodność jest dana wzorem

$$\mathcal{L}(\mu, \sigma) = \left(\frac{\mathrm{const}}{\sigma}\right)^n \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^n (X_i - \mu)^2\right].$$

Estymatory bez ograniczeń (w dużym modelu) są dobrze znane:

$$\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^n X_i, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2.$$

Podmodel jest jeszcze prostszy do przeanalizowania. Jeśli hipoteza H_0 jest prawdziwa, to mamy próbkę z rozkładu $\mathrm{N}(\mu_0, \sigma^2)$, więc

$$\dot{\mu} = \mu_0, \qquad \dot{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (X_i - \mu_0)^2.$$

Statystyką ilorazu wiarogodności jest, jak łatwo widzieć, $\mathcal{L}(\hat{\mu}, \hat{\sigma})/\mathcal{L}(\dot{\mu}, \dot{\sigma}) = (\dot{\sigma}/\hat{\sigma})^n$ (warto to sprawdzić). Test każe odrzucić hipotezę zerową, gdy ta statystyka przekracza odpowiednio ustalony poziom krytyczny. Przekształcimy ten test do równoważnej, powszechnie używanej postaci. Najpierw zauważmy, że $(\dot{\sigma}^2/\hat{\sigma})^2$ jest statystyką *równoważną*, bo jest rosnącą funkcją ilorazu wiarogodności. Następnie,

$$\frac{\dot{\sigma}^2}{\hat{\sigma}^2} = \frac{\sum_{i=1}^n (X_i - \bar{X})^2/n + (\bar{X} - \mu_0)^2}{\sum_{i=1}^n (X_i - \bar{X})^2/n} = 1 + \frac{(\bar{X} - \mu_0)^2}{\tilde{S}^2}.$$

Ostatecznie, ta statystyka jest równoważna

$$T^2 = \frac{(\bar{X} - \mu_0)^2}{S^2},$$

gdzie T jest statystyką t-Studenta w Punkcie 7.2.1. W istocie, otrzymaliśmy sławny test t-Studenta jako test ilorazu wiarogodności dla modeli zagnieżdżonych. ♢

Rozkład asymptotyczny testu ilorazu wiarogodności

Tak jak w punkcie 8.3.1, rozpatrujemy *dwa modele zagnieżdżone* i zagadnienie testowania

$$\mathrm{H}_0 : h(\theta) = 0 \text{ przeciw } \mathrm{H}_1 : h(\theta) \neq 0.$$

Zakładamy, że funkcja $h : \mathbb{R}^d \to \mathbb{R}^p$ jest „dostatecznie porządna", $\Theta \subset \mathbb{R}^d$ jest „zbiorem d-wymiarowym" i $\Theta_0 = \{\theta \in \Theta : h(\theta) = 0\}$ jest „zbiorem o wymiarze $d - p$":

$$\dim\Theta = d, \qquad \dim\Theta_0 = d - p < d.$$

Nie będziemy się starali uściślić tych sformułowań. W większości konkretnych przypadków i tak wiadomo, o co chodzi. Przy pewnych dodatkowych założeniach, prawdziwy jest następujący fakt.

8.3.3 TWIERDZENIE. *Jeśli* H_0 *jest prawdziwa, to przy* $n \to \infty$ *rozkład statystyki* $2\log\lambda$ *zmierza do rozkładu* $\chi^2(p)$ *(chi-kwadrat z* p *stopniami swobody).*

Mnemotechniczna regułka jest taka:

$$\text{liczba stopni swobody} = \text{liczba ograniczeń} = \dim\Theta - \dim\Theta_0.$$

Szkic dowodu Twierdzenia 8.3.3. Nie podamy dowodu w pełnej ogólności, ale naszkicujemy rozumowanie dla przypadku szczególnego, $d = 1$ i $p = 1$. Testowanie $\mathrm{H}_0 : \theta = \theta_0$ przeciw $\mathrm{H}_1 : \theta \neq \theta_0$ mieści się w rozpatrywanym schemacie modeli zagnieżdżonych, z *jednopunktowym* zbiorem $\Theta_0 = \{\theta_0\}$ i jednowymiarowym zbiorem Θ. W istocie, rozważamy model *probabilistyczny* zanurzony w większym modelu *statystycznym*. Oczywiście, „estymator z ograniczeniami" jest równy $\dot{\theta} = \theta_0$ (gdy „mały" model zawiera tylko jeden punkt, to nie ma co estymować).

Przyjmijmy, że spełnione są założenia Twierdzenia 5.2.5 o asymptotycznej normalności $\hat{\theta} = \mathrm{MLE}(\theta)$. Podobnie jak w dowodzie Twierdzenia 5.2.5, napiszemy $\ell'(\theta) = \sum_1^n (\log f)'(\theta, X_i)$ i wykorzystamy rozwinięcie Taylora, tym razem dla funkcji $\ell(\theta)$:

$$\log\lambda = \ell(\hat{\theta}) - \ell(\theta_0) \simeq \ell'(\theta_0)(\hat{\theta} - \theta_0) + \frac{1}{2}\ell''(\theta_0)(\hat{\theta} - \theta_0)^2.$$

Powołamy się teraz na przybliżoną równość $\hat{\theta} - \theta_0 \simeq \ell'(\theta_0)/\ell''(\theta_0)$. Otrzymujemy

$$(*) \qquad \log\lambda \simeq -\frac{1}{2}\frac{\left(\ell'(\theta_0)\right)^2}{\ell''(\theta_0)} = \frac{1}{2}\frac{\left(\sum(\log f)'(\theta_0, X_i)/\sqrt{n}\right)^2}{\sum(\log f)''(\theta_0, X_i)/n}.$$

Zarówno $\ell'(\theta_0)$, jak i $\ell''(\theta_0)$ są sumami niezależnych zmiennych losowych o jednakowym rozkładzie. Zastosujmy CTG do licznika i PWL do mianownika we wzorze $(*)$: mamy $\ell'(\theta_0)/\sqrt{n} \to_{\mathrm{d}} \mathrm{N}(0, I_1(\theta_0))$ i $\ell''(\theta_0)/n \to_{\mathrm{P}} -I_1(\theta_0)$, tak jak w dowodzie Twierdzenia 5.2.5. W rezultacie $2\log\lambda \to_{\mathrm{d}} \chi^2(1)$, co należało wykazać. □

8.4 Zgodność testów

Zgodność jest podstawowym pojęciem, dotyczącym asymptotycznych własności testów. Rozważmy zagadnienie testowania w najogólniejszej postaci:

$$\mathrm{H}_0 : \theta \in \Theta_0 \text{ przeciw } \mathrm{H}_1 : \theta \in \Theta_1.$$

Niech $X_1, \ldots, X_n, \ldots$ będzie nieskończonym ciągiem obserwacji o rozkładzie $\mathbb{P}_\theta$ i $\delta_n = \delta(X_1, \ldots, X_n)$ będzie ciągiem testów takich, że δ_n zależy od początkowych n obserwacji. Myślimy o $\delta(X_1, \ldots, X_n)$ jak o *jednym* teście, stosowanym do próbek o różnej liczności. Pod tym względem sytuacja jest podobna jak w Rozdziale 3, gdzie interesowaliśmy się granicznymi własnościami estymatorów.

8.4.1 DEFINICJA. *Mówimy, że test δ_n jest **zgodny** na poziomie istotności α, jeśli*

(i) $\lim_{n\to\infty} \mathbb{P}_\theta(\delta_n = 1) \leqslant \alpha$ *dla każdego* $\theta \in \Theta_0$*;*

(ii) $\lim_{n\to\infty} \mathbb{P}_\theta(\delta_n = 1) = 1$ *dla każdego* $\theta \in \Theta_1$*.*

Innymi słowy, test jest zgodny, jeśli przy zwiększającej się próbce utrzymuje się ustalony poziom istotności, a moc dąży do jedności. Chodzi o to, że test „w końcu z całą pewnością” odrzuci hipotezę zerową, jeśli jest fałszywa. Sens określenia zgodności dla testów i dla estymatorów jest podobny. Zgodny estymator „w końcu z całą pewnością” identyfikuje nieznany parametr. Zgodność jest raczej słabym, minimalnym wymaganiem (przynajmniej w sytuacji, gdy obserwacje $X_1, \ldots, X_n, \ldots$ są niezależnymi zmiennymi losowymi o jednakowym rozkładzie). Zadanie 8.1 dotyczy zgodności testów rozważanych w Przykładach 8.2.3 i 8.2.6. Oto inny, nieco mniej oczywisty przykład.

8.4.2 Przykład (Zgodność testu Kołmogorowa)**.** Zakładamy, że ciąg obserwacji $X_1, \ldots, X_n, \ldots$ jest próbką z rozkładu o nieznanej dystrybuancie F. Niech F_0 będzie ustaloną dystrybuantą. Rozważamy zagadnienie testowania

$$\mathrm{H}_0 : F = F_0 \text{ przeciw } \mathrm{H}_1 : F \neq F_0.$$

Napiszemy statystykę Kołmogorowa w postaci $D_n = D_n(F_0; X_1, \ldots, X_n) = \sup_x |\hat{F}_n(x) - F_0(x)|$. Przyjmijmy, że wartość krytyczną ustalamy przybliżoną metodą opisaną w Uwadze 7.1.9: test na poziomie istotności α odrzuca H_0 jeśli $\sqrt{n} D_n > c$, gdzie stała c nie zależy od n. Zgodność testu Kołmogorowa znaczy tyle, że dla $F \neq F_0$,

$$\mathbb{P}_F\left(\sqrt{n} D_n(F_0; X_1, \ldots, X_n) > c\right) \to 1, \qquad (n \to \infty).$$

Symbol $\mathbb{P}_F$ znaczy, że $X_1, \ldots, X_n \sim F$. Zauważmy, że Definicja 8.4.1 pracuje tu w niezwykle abstrakcyjnej sytuacji: za „parametr” rozkładu prawdopodobieństwa uważamy tu jego *dystrybuantę*. Zgodność testu Kołmogorowa wynika z twierdzenia Gliwienki-Cantelliego (Twierdzenie 1.1.6). $\diamondsuit$

8.5 Zadania

8.1. Obliczyć granicę $1-\beta(\mu)$ przy $n \to \infty$ dla testów rozważanych w Przykładach 8.2.3 i 8.2.6. Wskazać związek orzymanego wyniku z pojęciem zgodności testu.

8.2. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu $\mathrm{N}(\mu, \sigma^2)$ ze *znaną wartością oczekiwaną* μ. Pokazać, że UMPT $\mathrm{H}_0 : \sigma \leqslant \sigma_0$ przeciw $\mathrm{H}_1 : \sigma > \sigma_0$ na poziomie istotności α ma postać $\sum (X_i - \mu)^2 > c$. Jak wybrać poziom krytyczny c?

8.3. Obserwujemy pojedynczą zmienną losową X o rozkładzie o gęstości

$$f_\theta(x) = \begin{cases} (\theta+1)x^\theta & \text{jeśli } 0 \leqslant x \leqslant 1; \\ 0 & \text{w przeciwnym przypadku.} \end{cases}$$

Testujemy hipotezę $\mathrm{H}_0 : \theta = 0$ przeciwko $\mathrm{H}_1 : \theta > 0$.

(a) Skonstruować UMPT na poziomie istotności α.

(b) Obliczyć funkcję mocy, $1 - \beta(\theta)$, tego testu.

8.4. Niech $X_1, \ldots, X_n$ będzie próbką losową z rozkładu wykładniczego $\mathrm{Ex}(\lambda)$.

(a) Skonstruować test jednostajnie najmocniejszy na poziomie α do weryfikacji hipotezy $H_0 : \lambda = \lambda_0$ wobec hipotezy alternatywnej $H_1 : \lambda > \lambda_0$.

(b) W szczególnym przypadku, zakładając że $n = 1$, podać funkcję mocy tego testu.

8.5. Obserwujemy pojedynczą zmienną losową X z rozkładu o gęstości

$$f_\theta(x) = \begin{cases} \theta e^{-x} + 2(1-\theta)e^{-2x} & \text{dla } x \geqslant 0; \\ 0 & \text{dla } x < 0, \end{cases}$$

gdzie $\theta \in [0, 1]$ jest nieznanym parametrem.

(a) Skonstruować jednostajnie najmocniejszy test na poziomie istotności α do weryfikacji hipotezy zerowej $H_0 : \theta = 0$ przeciwko hipotezie alternatywnej $H_1 : \theta > 0$.

(b) Obliczyć funkcję mocy tego testu.

8.6. Obserwujemy pojedynczą zmienną losową X o rozkładzie gamma $\mathrm{Gam}(\theta, 1)$ o gęstości

$$f_\theta(x) = \begin{cases} \dfrac{1}{\Gamma(\theta)} x^{\theta-1} e^{-x} & \text{dla } x > 0; \\ 0 & \text{dla } x \leqslant 0, \end{cases}$$

gdzie $\theta > 0$ jest nieznanym parametrem. Testujemy hipotezę zerową

$$H_0 : \theta \leqslant \theta_0$$

przeciw hipotezie alternatywnej

$$H_1 : \theta > \theta_0.$$

(a) Wyznaczyć test jednostajnie najmocniejszy na poziomie istotności α.

(b) Obliczyć funkcję mocy tego testu, zakładając znajomość dystrybuanty i kwantyli rozkładów gamma.

8.7. Uzasadnić dokładnie fakt wspomniany w Przykładzie 8.4.2: zgodność testu Kołmogorowa wynika z twierdzenia Gliwienki-Cantelliego (Twierdzenie 1.1.6).

8.8. Niech $X_1, \ldots, X_n$ będzie próbką z rozkładu o dystrybuancie F. Przeprowadzamy test Kołmogorowa-Smirnowa hipotezy $H_0 : F = F_0$, gdzie F_0 jest dystrybuantą rozkładu wykładniczego Ex(1), przeciwko alternatywie $F \neq F_0$. Statystyką testową jest $D_n = \sup_x |\hat{F}_n(x) - F_0(x)|$. Naprawdę, próbka pochodzi z rozkładu Ex(2).

(a) Zbadać zbieżność ciągu zmiennych losowych D_n prawie na pewno.

(b) Zbadać zbieżność ciągu odpowiednich p-wartości prawie na pewno.

Wskazówka: Wykorzystać Twierdzenie Gliwienki-Cantelliego.