Literatura

Koronacki J., Mielniczuk J., Statystyka dla kierunków technicznych i przyrodniczych, WNT 2001.

w sprawie oprogramowania Statistica proszę pisać na adres:
oprogramowanie@cri.agh.edu.pl

Statystyka jest nauką zajmującą się najogólniej mówiąc zbieraniem danych i wydobywaniem informacji zawartej w tych danych. Statystykę można podzielić na dwie części:

- statystykę opisową,
- statystykę matematyczną.

Statystyka opisowa zajmuje się gromadzeniem danych oraz informacji dotyczących sposobu ich uzyskania (np. informacji dotyczących sposobu przeprowadzenia eksperymentu laboratoryjnego) oraz ich wstępną obróbką, przez którą rozumiemy **sortowanie danych**, ich **prezentację graficzną** a także **wyznaczenie pewnych charakterystyk liczbowych**.

Statystyka opisowa nie używa aparatu probabilistycznego.

Jej celem jest sformułowanie pewnych hipotez badawczych mających intuicyjne uzasadnienie w zgromadzonych danych. W raz z powszechną dostępnością komputerów i wzrostem ich mocy obliczeniowej metody statystyki opisowej wzbogaciły się o nowe techniki zwane **eksploracyjną analizą danych**. Połączenie eksploracyjnej analizy danych z techniką systemów uczących i sztucznej inteligencji zaowocowały powstaniem nowej dziedziny: **inteligentnej analizy danych** (**data mining**).

Statystyka matematyczna jest częścią probabilistyki powstałą na gruncie rachunku prawdopodobieństwa. Istotna jest jednak różnica między zadaniami rachunku prawdopodobieństwa a zadaniami statystyki matematycznej. Sens tej różnicy ilustruje następujący przykład. Rzucamy 10 razy monetą. Zadanie z rachunku prawdopodobieństwa polega np. na obliczeniu prawdopodobieństwa uzyskania czterech orłów jeżeli wiadomo, że prawdopodobieństwo uzyskania orła w jednym rzucie jest równe 0,5. Statystykę matematyczną interesuje zagadnienie w pewnym sensie odwrotne. Jakie jest prawdopodobieństwo uzyskania orła w jednym rzucie, jeżeli w dziesięciu rzutach uzyskano np. cztery orły? **Zadaniem statystyki matematycznej jest więc określenie nieznanych prawdopodobieństw w przyjętym modelu doświadczenia.**

Cztery typy skal pomiarowych

Wyróżnia się 4 **podstawowe skale pomiarowe**

• nominalna	skale jakościowe
• porządkowa	
• interwałowa	skale ilościowe
• ilorazowa	

Przykłady zmiennych

- nominalnych: **płeć, wyznanie , grupa krwi**
- porządkowych: **poparcie dla partii rządzącej (**zdecydowanie nie, raczej nie, obojętny, raczej tak, zdecydowanie tak)
- interwałowych -pomiar temperatury na skali °C lub °F. Nie ma większego sensu mówić, że temperatura obiektu *A* jest 2 razy wyższa od temperatury obiektu *B*. Można jedynie mówić, że temperatura obiektu *A* jest wyższa od temperatury obiektu *B* o pewną liczbę jednostek.
- ilorazowych- pomiar temperatury na skali °K

Dla danego typu danych pomiarowych mamy do dyspozycji standardowe procedury statystyczne. Dla danych pomiarowych dostępnych na skali wyższego typu można stosować procedury analizy danych niższego typu. Oczywiście taka analiza wiąże się z częściową utratą informacji. Należy wyraźnie podkreślić, że stosowanie metod przeznaczonych dla wyższych skal pomiarowych do danych dostępnych na niższych skalach jest nieuprawnioną manipulacją.

Wstępna analiza danych jakościowych

Metody liczbowe

- tabele liczności,
- tabele wielodzielcze

	1 Płeć	2 Papierosy	3 Kawa
1	M	dużo	dużo
2	M	dużo	dużo
3	K	nigdy	dużo
4	M	dużo	dużo
5	K	mało	dużo
6	M	dużo	dużo

Metody graficzne:

- wykresy słupkowe i kołowe
- histogramy skategoryzowane,
- interakcje liczności,
- histogramy 3W
- wykresy obrazkowe (koła, gwiazdy, promienie, twarze Chernoffa, wielokąty, profile itd.)

Wstępna analiza danych ilościowych

Niech $x_1,...,x_n$ będzie ciągiem danych ilościowych. Oznaczmy przez $x_{(1)} \le ... \le x_{(n)}$ uporządkowane rosnąco dane.

Metody graficzne

- histogramy liczności i częstości - uwagi o wyborze długości przedziału i początku histogramu –jedna z możliwości: długość przedziału $h_0 = 2{,}64 \cdot IQR \cdot n^{-1/3}$ a początek wybieramy tak, aby najmniejsza obserwacja była środkiem pierwszego przedziału (*IQR* wyjaśnimy poniżej)

	Czas pracy wierteł
	1 Czas pracy
1	504
2	507
3	522
4	507

- łamane częstości i krzywe estymatora jądrowego
  Przykład. Załóżmy że zaobserwowano 4 wartości zmiennej losowej : 0, 1, 2 i 4. Narysować wykres funkcji będącej średnią arytmetyczną funkcji gęstości rozkładu normalnego z wartościami oczekiwanymi równymi podanym obserwacjom i odchyleniu standardowym 1. Na tym samym rysunku narysować analogiczne wykresy dla parametru σ=0,4 i σ=2.

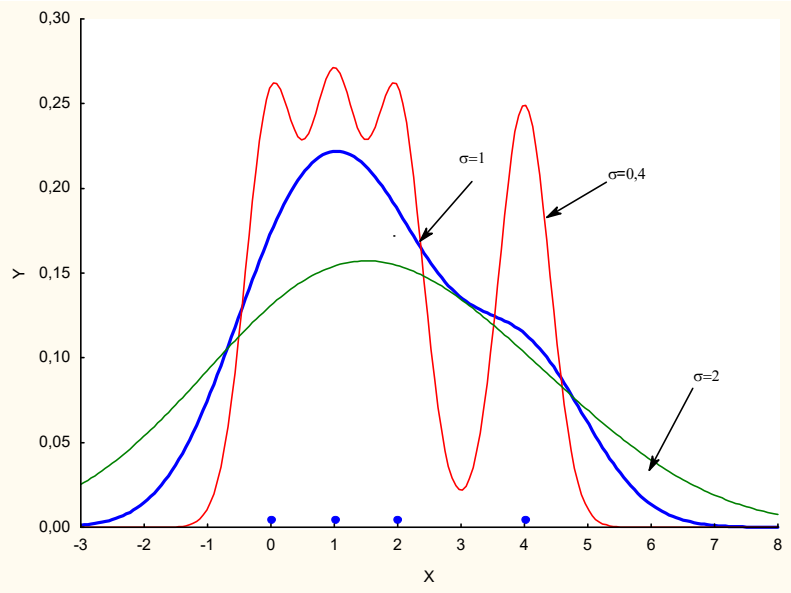
Wykonanie Wykresy/Wykresy 2W/ Wykresy funkcji użytkownika

Funkcja

Y= 1/4*(Normal(x;0;1)+Normal(x;1;1)+Normal(x;2;1)+Normal(x;4;1))

Podobne pozostałe funkcje

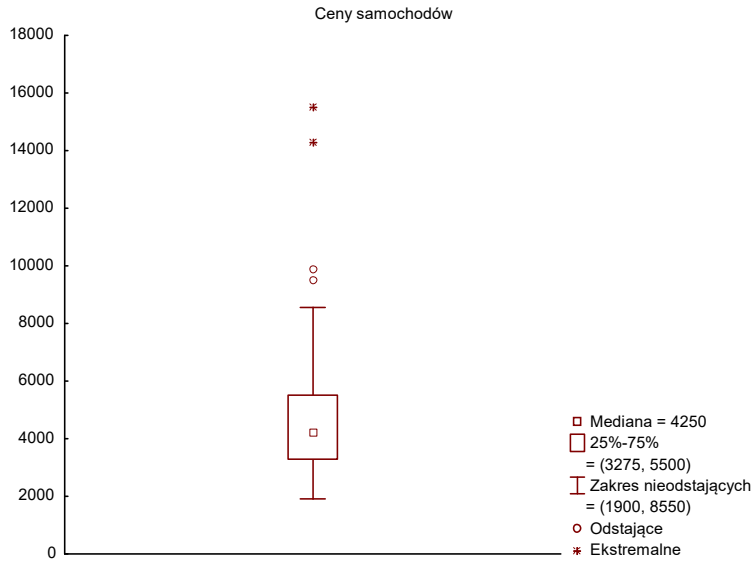
Dokonując stosownych zmian doprowadzić rysunek do poniższej postaci.



Narysowaliśmy wykres jądrowego estymatora funkcji gęstości dla trzech parametrów wygładzania 0,4 1 i 2.

- **Wykres ramkowy** (skrzynka z wąsami) .

Długość wąsa nie powinna przekraczać 1,5×IQR. Jeśli obserwacja jest odległa od brzegu skrzynki więcej niż 1,5×IQR, to jest traktowana jako obserwacja odstająca (outlier). Jeśli obserwacja jest odległa od brzegu skrzynki więcej niż 3×IQR, to jest traktowana jako ekstremalnie odstająca.



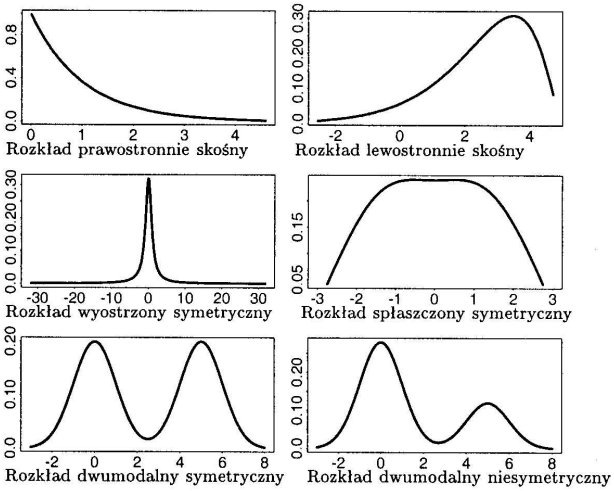
- wykresy przebiegu dla danych chronologicznych - trendy, okresowość
- wykresy kwantylowe

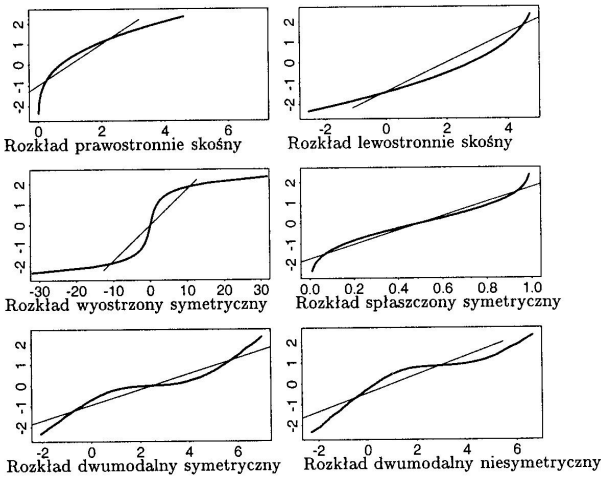
  $z_{i/n}$ - kwantyl rzędu i/n rozkładu $N(0,1)$ (zwykle kwantyl rzędu $(i\text{-}3/8)/(n+1/4)$)

  $x_{i/n}$ - kwantyl rzędu i/n badanego rozkładu

  Jeżeli badany rozkład jest normalny $N(m,\sigma^2)$ to punkty $(x_{i/n}, z_{i/n})$ leżą na prostej

$$z_{i/n} = \frac{x_{i/n} - m}{\sigma}$$





Wskaźniki sumaryczne

położenia

- wartość średnia w próbie (arytmetyczna, geometryczna, harmoniczna)
- mediana $x_{med} = \begin{cases} x_{((n+1)/2)} & n \text{ nieparzyste} \\ \frac{1}{2}(x_{(n/2)} + x_{(n/2+1)}) & n \text{ parzyste} \end{cases}$

- średnia ucinana (trimmed) $\bar{x}_{tk} = \frac{1}{n-2k}\sum_{i=k+1}^{n-k} x_{(i)}$
- średnia winsorowska $\bar{x}_{wk} = \frac{1}{n}\left((k+1)x_{(k+1)} + \sum_{i=k+2}^{n-k-1} x_{(i)} + (k+1)x_{(n-k)} \right)$ -*k* najmniejszych wartości $x_{(1)},...,x_{(k)}$ zastępujemy wartościami $x_{(k+1)}$ natomiast *k* największych wartości $x_{(n-k+1)},...,x_{(n)}$ zastępujemy wartościami $x_{(n-k)}$

rozproszenia (rozrzutu)

- rozstęp próby $R = x_{(n)} - x_{(1)}$
- wariancja w próbie $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ i odchylenie standardowe $s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2}$
- odchylenie przeciętne $d_1 = \frac{1}{n}\sum_{i=1}^{n} | x_i - \bar{x} |$
- kwantyle (w szczególności kwartyle)
- rozstęp międzykwartylowy $IQR = Q_3 - Q_1$

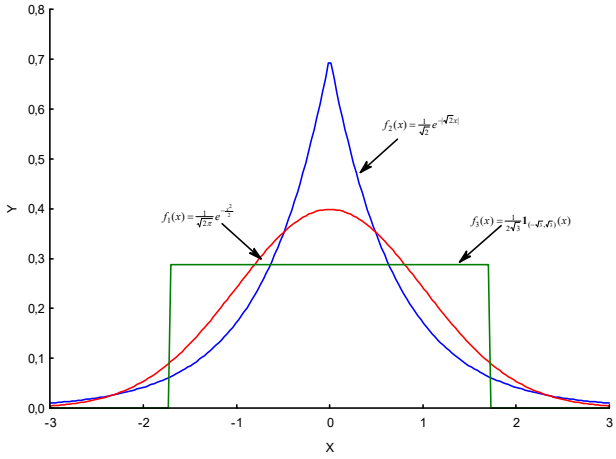
kształtu

- Skośność $= \frac{n\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)(n-2)s^3}$ (ocena $E\left(\frac{X-\mu}{\sigma}\right)^3$)
- Kurtoza $= \frac{n(n+1)\sum_{i=1}^{n}(x_i - \bar{x})^4}{(n-1)(n-2)(n-3)s^4} - 3\frac{(n-1)^2}{(n-2)(n-3)}$ (ocena $E\left(\frac{X-\mu}{\sigma}\right)^4 - 3$)

Rozważmy trzy symetryczne standaryzowane (o zerowej wartości oczekiwanej i jednostkowej wariancji) rozkłady: normalny, Laplace'a (dwustronny wykładniczy) i jednostajny o funkcjach gęstości odpowiednio:

$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$; $f_2(x) = \frac{1}{\sqrt{2}} e^{-|\sqrt{2}x|}$; $f_3(x) = \frac{1}{2\sqrt{3}} \mathbf{1}_{(-\sqrt{3},\sqrt{3})}(x)$

Kurtozy dla tych rozkładów są odpowiednio równe : 0; 3; $-\frac{6}{5}$.



Przestrzeń statystyczna

Z formalnego punktu widzenia eksperyment statystyczny jest opisywany za pomocą przestrzeni statystycznej $(\mathcal{X}, \mathcal{B}, \mathcal{P}=\{P_\theta : \theta\in\Theta\})$, gdzie

- $\mathcal{X}$ jest tzw. przestrzenią prób, czyli zbiorem możliwych wyników eksperymentu,
- $\mathcal{B}$ jest σ-ciałem podzbiorów przestrzeni prób $\mathcal{X}$
- $\mathcal{P}=\{P_\theta : \theta\in\Theta\}$, jest rodziną rozkładów na σ-ciele $\mathcal{B}$ parametryzowaną parametrem θ ze zbioru parametrów Θ.

Uwaga. Powyższy zapis nie ogranicza rodziny rozkładów, gdyż każda rodzina rozkładów może być sparametryzowana w trywialny sposób - parametrem może być sam rozkład (lub jego dystrybuanta).

Jeżeli zbiór parametrów Θ jest podzbiorem skończenie wymiarowej przestrzeni euklidesowej R^m, to rodzinę $\mathcal{P}=\{P_\theta : \theta\in\Theta\}$ nazywamy **rodziną parametryczną** a rozważane problemy statystyczne dotyczące tej rodziny nazywać będziemy **problemami parametrycznymi** np. estymacja parametryczna, testowanie hipotez parametrycznych. Jeżeli Θ nie jest podzbiorem żadnej skończenie wymiarowej przestrzeni euklidesowej, to rodzinę $\mathcal{P}=\{P_\theta : \theta\in\Theta\}$ nazywamy **rodziną nieparametryczną** a rozważane problemy statystyczne dotyczące tej rodziny nazywać będziemy **problemami nieparametrycznymi**.

Przykłady przestrzeni statystycznych

Niech $(X_1,...,X_n)$ będzie ciągiem niezależnych obserwacji zmiennej losowej X o rozkładzie P (czyli $X_1,...,X_n$ **iid** - independently identically distributed). Fakt ten wypowiadamy w statystyce matematycznej następująco: $(X_1,...,X_n)$ **jest próbą prostą z populacji o rozkładzie *P***. Wobec tego

$(x_1,...,x_n) = (X_1(\omega),...,X_n(\omega))$ dla pewnego ω

i $(x_1,...,x_n)$ traktujemy jako zaobserwowaną wartość (realizację) próby prostej $(X_1,...,X_n)$.

Przykład 1.Niech $\boldsymbol{X}=(X_1,...,X_n)$ będzie próbą prostą z populacji o rozkładzie $N(m,1)$. Przestrzenią statystyczną jest wówczas

$$(R^n,\mathcal{B}(R^n),\{f(x_1,...,x_n)=\frac{1}{(2\pi)^{n/2}}e^{-\frac{1}{2}\sum_{i=1}^{n}(x_i-m)^2}, m\in R\}),$$

która jest przestrzenią produktową i jest oznaczana również przez

$$(R,\mathcal{B}(R),\{f(x)=\frac{1}{(2\pi)^{1/2}}e^{-\frac{1}{2}(x-m)^2}, m\in R\})^n.$$

Jednoparametrowa rodzina rozkładów jest wyznaczona w tym przypadku przez rodzinę funkcji gęstości (względem miary Lebesgue'a). Przestrzenie produktowe otrzymujemy wówczas gdy mamy obserwacje typu **iid**. Zbiór parametrów Θ jest w tym przypadku równy R.

Przykład 2. Jeżeli w przykładzie 1 zastąpimy rozkład $N(m,1)$ rozkładem $N(m,\sigma^2)$, to otrzymamy następującą produktową przestrzeń statystyczną

$$(R,\mathcal{B}(R),\{f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}, (m,\sigma)\in R\times R_+\})^n,$$

Parametrem θ jest para (m,σ) a zbiorem parametrów Θ jest w tym przypadku $R\times R_+$.
Oczywiście w obu powyższych przykładach rodziny rozkładów są rodzinami parametrycznymi.

Przykład 3. Jeżeli w przykładzie 1 zastąpimy rozkład $N(m,1)$ rozkładem P_F o dystrybuancie F ze zbioru $\mathcal{F}$ wszystkich dystrybuant na prostej, to otrzymamy przestrzeń statystyczną
$(R,\mathcal{B}(R),\{P_F,F\in\mathcal{F}\})^n$. W tym przypadku $\theta=F$ a $\Theta=\mathcal{F}$. Zbiór $\mathcal{F}$ nie jest podzbiorem żadnej skończenie wymiarowej przestrzeni euklidesowej. Mamy więc do czynienia z nieparametryczną rodziną rozkładów.

Przykład 4. Wykonujemy ciąg niezależnych doświadczeń, z których każde kończy się sukcesem z nieznanym prawdopodobieństwem θ lub porażką z prawdopodobieństwem 1-θ. Doświadczenia te wykonujemy dopóty, dopóki nie uzyskamy m sukcesów. Sformułować model statystyczny tego eksperymentu.

Aby eksperyment zakończył się w k-tej próbie $k\geq m$ musimy uzyskać w próbie k sukces, a w poprzednich k-1 próbach dokładnie m-1 sukcesów. Oznaczając jak zwykle 1 - sukces, 0 - porażka przestrzeń prób można zapisać następująco:

$$\mathcal{X}=\{(x_1,...,x_k): k\geq m, x_i=1 \text{ lub } x_i=0, x_k=1, \sum_{i=1}^{k}x_i=m\}.$$

Ponieważ z prawdopodobieństwem 1 taki eksperyment zakończy się w skończonej liczbie prób pomijamy nieskończone ciągi zawierające mniej niż m jedynek. Jest ich przeliczalna ilość. Przestrzeń

prób składa się wiec z ciągów skończonych o długości co najmniej *m*, kończących się jedynką i zawierających dokładnie *m* jedynek.

Rodzina $\mathcal{B}$ jest w tym przypadku rodziną wszystkich podzbiorów zbioru $\mathcal{X}$, którą oznaczać będziemy $2^{\mathcal{X}}$. Rodzina rozkładów jest następująca:

$$\mathcal{P}=\{p_\theta(x_1,...,x_k)=\prod_{i=1}^{k}\theta^{x_i}(1-\theta)^{1-x_i}=\theta^m(1-\theta)^{k-m}\ \theta\in(0,1),\ k\geq m\ \}.$$

Tak skonstruowana przestrzeń **nie jest przestrzenią produktową**, ale rodzina rozkładów jest znów rodziną parametryczną.

Możliwa jest też konstrukcja innej przestrzeni

$$\mathcal{X}_1=\{m,m+1,...\},\ \mathcal{B}_1=2^{\mathcal{X}_1},\ \mathcal{P}_1=\{p_\theta(t)=\binom{t-1}{m-1}\theta^m(1-\theta)^{t-m}\ ,\ \theta\in(0,1)\}\ ,\ t=m,m+1...\ .$$

Powstaje pytanie: **Czy przeniesienie rozważań z przestrzeni** $(\mathcal{X},\mathcal{B},\mathcal{P})$ **do przestrzeni znacznie uboższej przestrzeni** $(\mathcal{X}_1,\mathcal{B}_1,\mathcal{P}_1)$ **pociąga za sobą utratę informacji**.

Odpowiedź w tym przypadku jest negatywna. Jej uzasadnienie prowadzi do pojęcia **statystyk dostatecznych**.

Pojęcie statystyki

Pojęcie **statystyki** w statystyce matematycznej jest odpowiednikiem pojęcia **zmiennej losowej** w rachunku prawdopodobieństwa. Niech $\boldsymbol{X}=(X_1,...,X_n)$ będzie próbą z pewnej populacji.

Definicja. Wektorową funkcję mierzalną $\boldsymbol{T}$**:** $\mathcal{X}\rightarrow\boldsymbol{T}(\boldsymbol{X})=(T_1(\boldsymbol{X}),...,T_k(X))\in R^k$ próby $\boldsymbol{X}$ nazywamy *k* wymiarową statystyką.

Należy koniecznie podkreślić, że **statystyka nie może zależeć od parametru** θ indeksującego rodzinę rozkładów.

Niech $\boldsymbol{X}=(X_1,...,X_n)$ będzie próbą prostą z populacji o rozkładzie $N(m,\sigma^2)$.Niech

$$\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i\ ,\ S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\ .$$

Wówczas $\boldsymbol{X}=(X_1,...,X_n)$, $\overline{X}$, S^2 są statystykami. Natomiast $\dfrac{\overline{X}-m}{S}$ i $\dfrac{\overline{X}-m}{\sigma}$ nie są statystykami.

Dystrybuanta empiryczna i jej podstawowe własności

Rozważmy przestrzeń statystyczną $(R,\mathcal{B},\{P_F,F\in\mathcal{F}\})^n$. Stawiamy pytanie:

Czy mając do dyspozycji ciąg niezależnych obserwacji $x_1,...,x_n$ zmiennej losowej o rozkładzie z nieznaną dystrybuantą F można choćby w przybliżeniu odtworzyć tą dystrybuantę ?

Aby odpowiedzieć na to pytanie definiujemy na R funkcję $\hat{F}_n(t)$ zwaną dystrybuantą empiryczną

$$\hat{F}_n(t) = \frac{\text{liczba obserwacji mniejszych niż } t}{\text{liczba wszystkich obserwacji}} = \frac{\#\{1 \le j \le n : x_j < t\}}{n}$$

Ponieważ obserwacja $(x_1,...,x_n)$ jest realizacją wektora losowego $(X_1(\omega),...,X_n(\omega))$, to dla każdego ustalonego t wartość dystrybuanty empirycznej $\hat{F}_n(t)$ traktujemy jako zaobserwowaną wartość zmiennej losowej $\hat{F}_n(t,\omega)$ zwanej również dystrybuantą empiryczną określonej wzorem

$$\hat{F}_n(t,\omega) = \frac{\#\{1 \le j \le n : X_j(\omega) < t\}}{n} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(-\infty,t)}(X_i(\omega))$$

gdzie $$\mathbf{1}_{(-\infty,t)}(x) = \begin{cases} 1, \text{ dla } x \in (-\infty, t] \\ 0, \text{ dla } x \notin (-\infty, t] \end{cases}.$$

Z powyższego wzoru widać, że dla ustalonego t dystrybuanta empiryczna jest sumą niezależnych zmiennych losowych $\mathbf{1}_{(-\infty,t)}(X_i(\omega))$ o rozkładzie dwupunktowym $B(1,F(t))$. Ogólnie, dystrybuanta empiryczna jest procesem stochastycznym na R. Z powyższych uwag wynika:

Twierdzenie. Dla każdego ustalonego t dystrybuanta empiryczna ma następujące własności:

1. $E_F(\hat{F}_n(t,\omega)) = F(t)$
2. $P_F\{\omega : \lim_{n\to\infty} \hat{F}_n(t,\omega) = F(t)\} = 1$
3. Rozkład zmiennej losowej $\sqrt{n}\frac{\hat{F}_n(t,\omega) - F(t)}{\sqrt{F(t)(1-F(t))}}$ dąży do rozkładu $N(0,1)$, gdy $n \to \infty$.

Dowód: Własność 1 wynika z liniowości operatora wartości oczekiwanej. Istotnie

$$E_F(\hat{F}_n(t,\omega)) = E_F[\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}_{(-\infty,t]}(X_i(\omega))] = \frac{1}{n}\sum_{i=1}^{n} E_F[\mathbf{1}_{(-\infty,t]}(X_i(\omega))] =$$

$$\frac{1}{n}\sum_{i=1}^{n} P_F(X_i(\omega) \le t) = \frac{1}{n}\sum_{i=1}^{n} F(t) = F(t)$$

Własność 2 wynika bezpośrednio z mocnego prawa wielkich liczb (MPWL) Kołmogorowa a własność 3 z centralnego twierdzenia granicznego (CTG) dla ciągu prób Bernoulliego.

Własności 1, 2 i 3 mają charakter lokalny. Warto zaznaczyć, że zbiór $\{\omega : \lim_{n\to\infty} \hat{F}_n(t,\omega) \neq F(t)\}$ o zerowym P_F prawdopodobieństwie jest zależny od t. Istotnym wzmocnieniem własności 2 jest następujące twierdzenie Gliwienki-Cantellego zwane także **podstawowym twierdzeniem statystyki matematycznej**.

Twierdzenie Gliwienki-Cantellego. Jeżeli $X_1,..., X_n$ jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie z dystrybuantą F, to

$$P_F\{\omega : \lim_{n\to\infty} D_n(\omega) = 0\} = 1, \text{ gdzie } D_n(\omega) = \sup_{-\infty<t<\infty} |\hat{F}_n(t,\omega) - F(t)|.$$

Twierdzenie to orzeka, że dystrybuanta empiryczna zbiega z prawdopodobieństwem 1 jednostajnie na R do dystrybuanty teoretycznej. Zatem, jeżeli rozmiar próby jest dostatecznie duży, to dystrybuanta empiryczna dowolnie dobrze przybliża nieznaną dystrybuantę F.

Bardzo ważnym i użytecznym jest następujące:

Twierdzenie Kołmogorowa. Jeżeli $X_1,..., X_n$ jest ciągiem niezależnych zmiennych losowych o tym samym rozkładzie z **ciągłą** dystrybuantą $F \in \mathcal{F}_c$, to

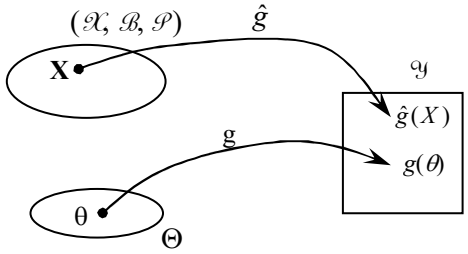
$$\lim_{n\to\infty} P_F(\sqrt{n} D_n \le x) = K(x) = \sum_{k=-\infty}^{\infty} (-1)^k e^{-2k^2x^2} \text{ dla } x>0.$$

Dystrybuanta $K(x)$ jest stablicowana. Rozkład zmiennej losowej D_n przy założeniu $F \in \mathcal{F}_c$ nie zależy od dystrybuanty F. Wynika to z faktu, że jeśli X jest zmienną losową o dystrybuancie F, to $F(X)$ jest zmienną losową o rozkładzie jednostajnym $U_{(0,1)}$.

Podstawowe problemy statystyki matematycznej

Niech $(\mathcal{X}, \mathcal{B}, \mathcal{P}=\{P_\theta : \theta\in\Theta\})$, będzie przestrzenią statystyczną indukowaną przez zmienną losową X (zwykle wektorową) o wartościach w przestrzeni $\mathcal{X}$. Podstawowe problemy statystyki matematycznej to:

Problem estymacji punktowej. Na podstawie obserwacji zmiennej losowej X oszacować $g(\theta)\in \mathcal{Y}$, gdzie g: $\Theta\to \mathcal{Y}$ jest znaną funkcją parametru θ o wartościach w pewnej przestrzeni metrycznej (zwykle euklidesowej). Ponieważ parametr θ jest nieznany wartość $g(\theta)$ jest nieznana. Rozwiązaniem tego problemu będzie pewna funkcja (statystyka, element losowy) $\hat{g}$:$\mathcal{X} \to \mathcal{Y}$ zwana estymatorem. Estymator może być uznany za dobry estymator, jeżeli funkcja $\hat{g}$ przyjmuje wartości bliskie wartościom $g(\theta)$ $\forall\theta\in\Theta$. Wprowadzenie do sformułowania problemu funkcji g poszerza klasę jednolicie rozważanych zagadnień, gdyż oprócz szacowania samego parametru θ (g jest wtedy identycznością) i różnych jego funkcji (np. θ^2) obejmuje szacowanie wartości pewnych funkcjonałów w przypadkach nieparametrycznych – $(R,\mathcal{B},\{P_F,F\in\mathcal{F}\}^n$ np. $g(\theta)=g(F)=\int x dF = E_\theta X$.



Problem estymacji przedziałowej. Dla przedstawionego powyżej problemu estymacji możemy konstruować inne rozwiązanie w postaci zbioru $\hat{g}(X) \subset \mathcal{Y}$ np. przedziału $(\hat{g}_1(X), \hat{g}_2(X)) \subset \mathcal{Y}$ w przypadku szacowania parametru liczbowego $g(\theta)$ tak, aby $\forall \theta \in \Theta \quad P_\theta(\hat{g}_1(X) \le g(\theta) \le \hat{g}_2(X)) \ge 1-\alpha$. Przedział $(\hat{g}_1(X), \hat{g}_2(X))$ jest oczywiście przedziałem losowym. Liczbę $1-\alpha$ (zwykle bliską jedności) nazywamy poziomem ufności. Zagadnienie estymacji przedziałowej można uogólnić zastępując przedział innym **zbiorem ufności** (np. kulą jeśli $\mathcal{Y}$ jest przestrzenią metryczną). Zazwyczaj przyjmuje się pewne założenia regularności (np. spójność zbioru ufności) i założenia dotyczące kształtu zbioru ufności.

Problem testowania hipotez. Niech $\Theta=\Theta_0 \cup \Theta_1$ i $\Theta_0 \cap \Theta_1 = \varnothing$. Na podstawie obserwacji zmiennej losowej X zweryfikować hipotezę H_0: $\theta \in \Theta_0$ wobec alternatywy H_1: $\theta \in \Theta_1$.

Rozwiązaniem problemu będzie pewna funkcja φ: $\mathcal{X} \to [0,1]$ zwana zrandomizowanym testem statystycznym. Dysponując testem statystycznym φ i obserwacją x zmiennej losowej X odrzucamy hipotezę H_0 z prawdopodobieństwem $\varphi(x)$. Aby podjąć konkretną decyzję należy więc użyć pewnego mechanizmu losowego, który produkuje dwa wyniki z prawdopodobieństwami $\varphi(x)$ i $1-\varphi(x)$ i na tej podstawie podjąć (wylosować) decyzję. Jeżeli zbiór wartości funkcji φ jest zbiorem dwuelementowym $\{0,1\}$, to test φ nazywamy niezrandomizowanym. Test taki dzieli przestrzeń prób na dwa rozłączne zbiory :

$$\varphi^{-1}(\{1\})=\{x \in \mathcal{X}: \varphi(x)=1\} \qquad \text{zwany zbiorem odrzucenia hipotezy } H_0$$

i

$$\varphi^{-1}(\{0\})=\{x \in \mathcal{X}: \varphi(x)=0\} \qquad \text{zwany zbiorem akceptacji } H_0.$$

Konstrukcja testu niezrandomizownego jest więc równoważna rozbiciu przestrzeni prób na dwa rozłączne podzbiory.

Problem klasyfikacji (zwany również problemem dyskryminacji.) jest uogólnieniem problemu testowania hipotez. Uogólnienie to polega na rozbiciu zbioru parametrów Θ (lub równoważnie rodziny $\mathcal{P}$ rozkładów prawdopodobieństwa na przestrzeni prób) na $k \ge 2$ rozłącznych podzbiorów tzn. $\Theta=\bigcup_{i=1}^{k}\Theta_i$, $i \ne j \Rightarrow \Theta_i \cap \Theta_j = \varnothing$. Dla dowolnej obserwacji x zmiennej losowej X należy zadecydować z której z k rozłącznych populacji ona pochodzi. Rozwiązaniem tego problemu będzie pewna funkcja

wektorowa $\varphi : \mathcal{X} \to \varphi(x)=(\varphi_1(x),..., \varphi_k(x))$ zwana zrandomizowaną funkcją dyskryminacyjną, gdzie $\varphi_i(x)$ jest prawdopodobieństwem zakwalifikowania obserwacji x do i-tej populacji i $\sum_{i=1}^{k} \varphi_i(x) = 1 \quad (\forall x)$. Zagadnienie testowania hipotez jest szczególnym przypadkiem zagadnienia klasyfikacji dla k=2. Warunek $\varphi_1(x)+\varphi_2(x)=1$ umożliwia używanie tylko jednej składowej funkcji wektorowej (φ_1,φ_2).

Każdy z przedstawionych problemów ma swój specyficzny aspekt, ale można też wyróżnić pewne wspólne cechy pozwalające traktować jednolicie te problemy jako tzw. **statystyczny problemem decyzyjny**, czyli grę dwóch osób: statystyka i natury.

Wybrane rozkłady prawdopodobieństwa użyteczne w statystyce

- Rozkład χ^2_n chi-kwadrat o n stopniach swobody - to rozkład sumy kwadratów n niezależnych zmiennych losowych o standaryzowanym rozkładzie normalnym $N(0,1)$ tzn.

  $X_1,...,X_n$ i.i.d. $N(0,1) \Rightarrow Y=\sum_{i=1}^{n} X_i^2$ ma rozkład χ^2_n o funkcji gęstości $f_Y(y) = \frac{1}{2^{n/2}\,\Gamma(n/2)} y^{n/2-1} e^{-y/2}$,

  $y>0$. Widać, że rozkład χ^2_n jest szczególnym przypadkiem rozkładu gamma.

  Uzasadnienie. Niech zmienna losowa X ma rozkład normalny $N(0,1)$. Wyznaczymy rozkład zmiennej losowej $Y=X^2$.

  Dla $y>0$; $F_Y(y) = P(Y<y) = P(X^2<y) = P(-\sqrt{y} < X < \sqrt{y}) = F_{N(0,1)}(\sqrt{y}) - F_{N(0,1)}(-\sqrt{y})$

  Stąd
  $$f_Y(y) = \frac{d}{dy} F_Y(y) = f_{N(0,1)}(\sqrt{y})\frac{1}{2\sqrt{y}} - f_{N(0,1)}(-\sqrt{y})(-\frac{1}{2\sqrt{y}}) =$$
  $$= f_{N(0,1)}(\sqrt{y})\frac{1}{\sqrt{y}} = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{y}} e^{-\frac{y}{2}} = \frac{(\frac{1}{2})^{\frac{1}{2}}}{\Gamma(\frac{1}{2})} y^{\frac{1}{2}-1} e^{-\frac{y}{2}}$$
  , czyli $Y \sim \mathcal{G}(\frac{1}{2},\frac{1}{2})$

  Z twierdzenia o dodawaniu dla rozkładu Gamma wynika, że $\chi^2_n = \mathcal{G}(\frac{1}{2},\frac{n}{2})$.

  - Rozkład t-Studenta o n stopniach swobody – niech X będzie zmienną losową o rozkładzie $N(0,1)$ a Y zmienną losową o rozkładzie χ^2_n. Jeżeli zmienne X i Y są niezależne, to zmienna losowa $T=\frac{X}{\sqrt{\frac{Y}{n}}}$ ma rozkład t Studenta o n stopniach swobody i funkcji gęstości

    $f_T(t) = \frac{\Gamma(n+1/2)}{\sqrt{n\pi}\,\Gamma(n/2)}(1+\frac{t^2}{n})^{-\frac{n+1}{2}}$

Rozważmy niezależne zmienne $X \sim N(0,1)$ i $Y \sim \chi^2_n$. Znajdziemy rozkład zmiennej losowej $T = \frac{X}{\sqrt{\frac{Y}{n}}}$

$$F_T(t) = P(T < t) = P(\tfrac{X}{\sqrt{\frac{1}{n}Y}} < t) = P(X < t\sqrt{\tfrac{1}{n}Y}) = \iint\limits_{x<t\sqrt{\frac{1}{n}y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \frac{1}{2^{n/2}\Gamma(n/2)} y^{\frac{n}{2}-1} e^{-\frac{y}{2}} dxdy =$$

$$\left\{ \begin{cases} u = y \\ v = \frac{x}{\sqrt{\frac{1}{n}y}} \end{cases} \Leftrightarrow \begin{cases} x = v\frac{\sqrt{u}}{\sqrt{n}}, \left|\frac{\partial(x,y)}{\partial(u,v)}\right| = \left\| \begin{matrix} \frac{v}{2\sqrt{u}\sqrt{n}} & \frac{\sqrt{u}}{\sqrt{n}} \\ 1 & 0 \end{matrix} \right\| = \left| -\frac{\sqrt{u}}{\sqrt{n}} \right| \\ y = u \end{cases} \right\}$$

$$= \int\limits_{-\infty}^{t} [\int\limits_{0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{uv^2}{2n}} \frac{1}{2^{n/2}\Gamma(n/2)} u^{\frac{n}{2}-1} e^{-\frac{u}{2}} \frac{\sqrt{u}}{\sqrt{n}} du]dv = \int\limits_{-\infty}^{t} [\frac{1}{\sqrt{n}2^{\frac{n+1}{2}}\Gamma(1/2)\Gamma(n/2)} \int\limits_{0}^{\infty} u^{\frac{n+1}{2}-1} e^{-\frac{1}{2}(1+\frac{v^2}{n})u} du]dv =$$

$$= \int\limits_{-\infty}^{t} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n}2^{\frac{n+1}{2}}\Gamma(1/2)\Gamma(n/2)(\frac{1}{2}(1+\frac{v^2}{n}))^{\frac{n+1}{2}}} [\int\limits_{0}^{\infty} \frac{(\frac{1}{2}(1+\frac{v^2}{n}))^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})} u^{\frac{n+1}{2}-1} e^{-\frac{1}{2}(1+\frac{v^2}{n})u} du]dv = \int\limits_{-\infty}^{t} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n}2^{\frac{n+1}{2}}\Gamma(1/2)\Gamma(n/2)(\frac{1}{2}(1+\frac{v^2}{n}))^{\frac{n+1}{2}}} dv =$$

$$= \int\limits_{-\infty}^{t} \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n}\Gamma(1/2)\Gamma(n/2)(1+\frac{v^2}{n})^{\frac{n+1}{2}}} dv$$

Stąd otrzymujemy funkcję gęstości $f_T(t) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma(n/2)(1+\frac{t^2}{n})^{\frac{n+1}{2}}}$.

Uwaga : $\int\limits_{0}^{\infty} \frac{(\frac{1}{2}(1+\frac{v^2}{n}))^{\frac{n+1}{2}}}{\Gamma(\frac{n+1}{2})} u^{\frac{n+1}{2}-1} e^{-\frac{1}{2}(1+\frac{v^2}{n})u} du = 1$ - całka z funkcji gęstości rozkładu $\mathcal{G}\left(\frac{1}{2}(1+\frac{v^2}{n}), \frac{n+1}{2}\right)$

- Rozkład $F_{n,m}$ Snedecora-Fishera- niech X i Y będą niezależnymi zmiennymi losowymi o rozkładach χ^2_n χ^2_m wówczas zmienna losowa $Z = \frac{X/n}{Y/m}$ ma rozkład $F_{n,m}$ Snedecora-Fishera z funkcją gęstości $f_Z(z) = \frac{\Gamma(\frac{n+m}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})} (\frac{m}{n})^{\frac{m}{2}} \frac{z^{n/2-1}}{(z+m/n)^{(n+m)/2}}$, $z>0$

Niecentralne rozkłady chi-kwadrat $\chi^2_{n,\delta}$, Studenta $t_{n,\delta}$ i Snedecora $F_{n_1,n_2,\delta}$

- **Niecentralny rozkład chi-kwadrat**

Niech $X_1,...,X_n$ będą niezależnymi zmiennymi losowymi o rozkładach $N(m_i,1)$ $i=1,...,n$. Wówczas zmienna losowa $U = \sum_{i=1}^{n} X_i^2$ ma rozkład niecentralny chi-kwadrat o n stopniach swobody i parametrze niecentralności $\delta = \sqrt{\sum_{i=1}^{n} m_i^2}$ oznaczany $\chi^2_{n,\delta}$.

- **Niecentralny rozkład *t*- Studenta**

Niech $X \sim N(\delta,1)$, $Y \sim \chi^2_n$, X i Y niezależne. Wówczas $T = \frac{X}{\sqrt{\frac{Y}{n}}} \sim t_{n,\delta}$

- **Niecentralny rozkład *F* Snedecora Fishera**

Niech $X \sim \chi^2_{n_1,\delta}$ $Y \sim \chi^2_{n_2}$, X i Y niezależne. Wówczas $F = \frac{\frac{X}{n_1}}{\frac{Y}{n_2}} \sim F_{n_1,n_2,\delta}$

Uzupełnienie. Niech $X_1,...,X_n$ będzie próbą prostą z rozkładu $N(m,\sigma^2)$. Niech

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \ , \ S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 .$$

Przekształćmy ortogonalnie wektor $\begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix} \sim N\left(\begin{bmatrix} m \\ \vdots \\ m \end{bmatrix}, \sigma^2 I\right)$ na wektor $\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix}$ według wzoru

$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \\ \vdots & \cdots & \vdots \\ \cdot & \cdots & \cdot \end{bmatrix} \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}$, gdzie pozostałe wiersze macierzy przekształcenia są aby macierz była ortonormalna (można to zrobić traktując pierwszy wiersz jako wektor w R^n, uzupełnić ten wektor n-1 wektorami tak, aby uzyskać bazę w R^n i zastosować procedurę ortogonalizacji Grama –Schmidta).

Przekształcony ortogonalnie wektor ma rozkład $\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} \sim N\left(\begin{bmatrix} \sqrt{n}\, m \\ \vdots \\ 0 \end{bmatrix}, \sigma^2 I\right)$, więc jego składowe są niezależne. Zauważmy, że $Y_1 = \sqrt{n}\overline{X}$, natomiast

$$\sum_{i=1}^{n}(X_i - \overline{X})^2 = \sum_{i=1}^{n} X_i^2 - n\overline{X}^2 = \sum_{i=1}^{n} Y_i^2 - Y_1^2 = \sum_{i=2}^{n} Y_i^2 .$$

Stąd statystyki $\sqrt{n}\overline{X}$ i $\sum_{i=1}^{n}(X_i - \overline{X})^2$ są niezależne . Zmienna losowa $\frac{\sqrt{n}(\overline{X} - m)}{\sigma} = \frac{Y_1 - \sqrt{n}m}{\sigma}$ ma rozkład N(0,1) natomiast $\frac{1}{\sigma^2}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{\sigma^2}\sum_{i=2}^{n} Y_i^2$ ma rozkład χ^2_{n-1}. Stąd

- $Q(X_1,...,X_n,m) = \dfrac{\sqrt{n}(\overline{X} - m)}{S} = \dfrac{\frac{\sqrt{n}(\overline{X} - m)}{\sigma}}{\frac{1}{\sigma}\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}}$ ma rozkład *t*-Studenta z n-1 stopniami swobody, natomiast funkcja
- $Q(X_1,...,X_n,\sigma^2) = \dfrac{(n-1)S^2}{\sigma^2}$ ma rozkład χ^2_{n-1} .

Estymacja punktowa

Podstawowe pojęcia estymacji punktowej

Niech $(\mathcal{X},\mathcal{B},\ \mathcal{P}=\{P_\theta:\theta\in\Theta\})$, będzie przestrzenią statystyczną indukowaną przez zmienną losową X o wartościach w przestrzeni $\mathcal{X}$.Na podstawie obserwacji x zmiennej losowej X oszacować $g(\theta)\in\mathcal{Y}$, gdzie $g:\Theta\to\mathcal{Y}$ jest znaną funkcją. Rozwiązaniem tego problemu będzie pewna funkcja $\hat{g}:\mathcal{X}\to\mathcal{Y}$ zwana estymatorem. Estymator może być uznany za dobry estymator, jeżeli funkcja $\hat{g}$ przyjmuje wartości bliskie wartościom $g(\theta)\ \forall\theta\in\Theta$. Oczywiście nie każdy estymator jest dobrym estymatorem. Wprowadza się więc pewne pojęcia umożliwiające porównywanie estymatorów i w konsekwencji wybór najlepszego z nich.
Załóżmy, że dana jest pewna funkcja L: $\mathcal{Y}\times\mathcal{Y}\to\mathcal{R}$ zwana funkcją strat, której wartość $L(g(\theta)),\hat{g}(x))$ określa stratę jaką ponosi statystyk przyjmując $\hat{g}(x)$ za oszacowanie nieznanej wielkości $g(\theta)$. Wobec tego $L(g(\theta),\hat{g}(X))$ jest zmienną losową określoną na przestrzeni prób $\mathcal{X}$. Określamy więc (o ile to możliwe) średnią (oczekiwaną) stratę będącą funkcją deterministyczną.

Def. Funkcję $R_L(\theta,\hat{g})=E_\theta\, L(g(\theta),\hat{g}(X))$ parametru θ dla dowolnego ustalonego estymatora $\hat{g}$ nazywamy funkcją ryzyka estymatora $\hat{g}$ indukowaną przez funkcję strat L.

Estymatory będziemy porównywać ze sobą porównując ich funkcje ryzyka przy czym estymator jest tym lepszy im jego funkcja ryzyka przyjmuje mniejsze wartości.

Def. Estymator $\hat{g}_1$ jest nie gorszy od estymatora $\hat{g}_2$ w sensie ryzyka R_L indukowanego przez funkcję straty L jeżeli $R_L(\theta,\hat{g}_1)\le R_L(\theta,\hat{g}_2)\ \ \forall\theta\in\Theta$.

Def. Estymator $\hat{g}_1$ jest lepszy $\hat{g}_2$ w sensie ryzyka R_L indukowanego przez funkcję straty L jeżeli $\hat{g}_1$ jest nie gorszy od estymatora $\hat{g}_2$ i $\exists\theta\in\Theta: R_L(\theta,\hat{g}_1)<R_L(\theta,\hat{g}_2)$.

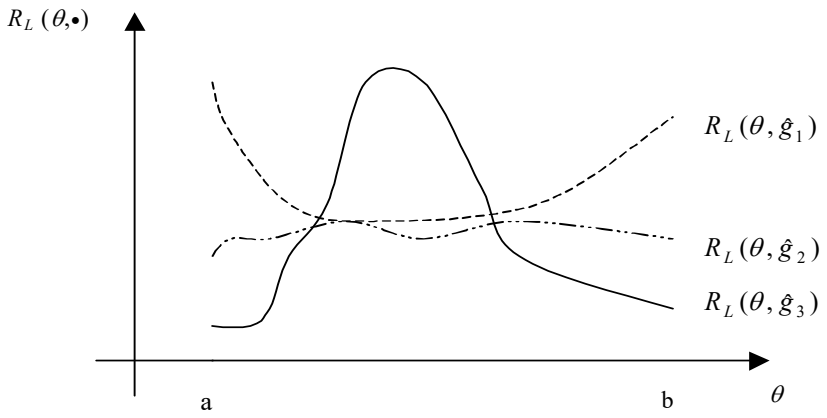
Niech $\mathcal{D}$ oznacza wyspecyfikowany zbiór estymatorów.

Def. Estymator $\hat{g}_1$ nazywamy niedopuszczalnym w $\mathcal{D}$ w sensie ryzyka R_L indukowanego przez funkcję strat L jeżeli istnieje w zbiorze $\mathcal{D}$ estymator $\hat{g}_2$ lepszy od $\hat{g}_1$.

Ze zbioru $\mathcal{D}$ rozważanych estymatorów można usunąć estymatory niedopuszczalne i ograniczyć rozważania jedynie do zbioru estymatorów dopuszczalnych $\mathcal{D}_{dop}$. Niestety, zwykle nie udaje się dla rozważanego problemu scharakteryzować klasy estymatorów dopuszczalnych. Czasami udaje się

udowodnić dopuszczalność konkretnego estymatora uzyskanego z rozważań optymalizacyjnych lub heurystycznych.

Niech Θ=<a,b>. Rozważmy trójelementowy zbiór estymatorów $\mathcal{D}$=$\{ \hat{g}_1, \hat{g}_2, \hat{g}_3\}$ pewnej wielkości g(θ) o funkcjach ryzyka przedstawionych na rysunku.



Widać, że estymator $\hat{g}_1$ jest niedopuszczalny, gdyż lepszym estymatorem jest $\hat{g}_2$.

Porównując estymatory poprzez porównywanie ich funkcji ryzyka możemy odrzucić pewne estymatory (niedopuszczalne). Pozostałe estymatory (dopuszczalne) są nieporównywalne w powyższym sensie, gdyż ich funkcje ryzyka wzajemnie się przecinają. Ponadto, praktyk wolałby otrzymać jakiś jeden estymator (najlepiej optymalny) zamiast zbioru dopuszczalnych estymatorów z którego i tak w końcu musi wybrać pewien konkretny estymator. Pokonać te trudności można na różne sposoby. Wymienić tu należy :

- podejście polegające na utrzymaniu kryterium porównywania estymatorów poprzez porównywanie ich funkcji ryzyka i ograniczaniu klasy rozważanych estymatorów np. estymatorów nieobciążonych, ekwiwariantnych, rekurencyjnych i innych. Na szczególną uwagę zasługuje tu teoria estymacji nieobciążonej o jednostajnie najmniejszej wariancji (ENJNW).

  Def. Estymator $\hat{g}$ wielkości $g(\theta)$ nazywamy nieobciążonym, gdy $\forall\ \theta\in\Theta \quad E_\theta(\hat{g}(X))= g(\theta)$ tzn. wartość oczekiwana estymatora $\hat{g}(X)$ jest równa (nieznanej) estymowanej wartości $g(\theta)$ przy każdym możliwym rozkładzie prawdopodobieństwa z założonej rodziny rozkładów.

  Nieobciążoność estymatora, która wyraża jego bezstronność (neutralność) wyrażającą się w braku skłonności do przeszacowywania bądź niedoszacowywania estymowanej wielkości przez estymator jest pozytywną cechą estymatora, której nie należy jednak demonizować.

Nieobciążoność jest szczególnie cenną własnością dopiero w przypadku gdy estymator (czyli zmienna losowa) ma niewielką wariancję. Okazuje się , że przy pewnych technicznych założeniach dotyczących regularności estymatora można podać oszacowanie od dołu (dokładnie kres dolny) wariancji estymatorów nieobciążonych. Estymatory nieobciążone o jednostajnie minimalnej wariancji, nazywane także estymatorami najefektywniejszymi (przy kwadratowej funkcji strat), potrafimy efektywnie konstruować w przypadku, gdy znana jest statystyka dostateczna, zupełna. Jeżeli nie potrafimy ustalić, czy istnieje estymator nieobciążony o jednostajnie minimalnej wariancji, to otwiera się inna możliwość. Możemy mianowicie porównywać wariancję badanego estymatora z wartością kresu dolnego wariancji, czyli szacować jego efektywność, którą definiujemy następująco.

Def. Jeżeli $\hat{g}$ jest ENJNW[$g(\theta)$]} o wariancji równej $Var_\theta(\hat{g})$, to efektywnością estymatora nieobciążonego $\hat{g}_1$ o wariancji $Var_\theta(\hat{g}_1)$ nazywamy wielkość

$$\text{eff}(\hat{g}_1)=\frac{Var_\theta(\hat{g})}{Var_\theta(\hat{g}_1)}.$$

Efektywność ENJMW jest więc równa 1. Może okazać się, że badany estymator ma wariancję niewiele większą od kresu dolnego wariancji wszystkich (regularnych) estymatorów nieobciążonych i wobec tego jest zadowalający z praktycznego punktu widzenia. Pewien rezultat związany z tym podejściem jest zawarty w następującym twierdzeniu.

Twierdzenie (**Cramer-Rao**).Niech $\mathcal{P}=\{P_\theta : \theta\in\Theta\}$ będzie rodziną rozkładów na przestrzeni prób $\mathcal{X}$ mających gęstości $p(x,\theta)$ względem pewnej ustalonej miary, i niech Θ będzie zbiorem otwartym w $\mathcal{R}$. Jeżeli są spełnione pewne warunki regularności, to wariancja każdego estymatora nieobciążonego $\hat{g}(X)$ wielkości $g(\theta)$ spełnia nierówność (Cramera-Rao)

$$Var_\theta[\hat{g}(X)] \geq \frac{\left[\frac{dg(\theta)}{d\theta}\right]^2}{E_\theta\left[\frac{\partial \ln p(X,\theta)}{\partial\theta}\right]^2}.$$

Wielkość $I_\theta = Var_\theta\left[\frac{\partial \ln p(X,\theta)}{\partial\theta}\right] = E_\theta\left[\frac{\partial \ln p(X,\theta)}{\partial\theta}\right]^2$ (mianownik w nierówności Cramera-Rao) nazywamy informacją w sensie Fishera o parametrze θ zawartą w obserwowanej zmiennej losowej (zwykle wektorowej) X. Jeżeli (regularny) estymator nieobciążony ma wariancję równą dolnemu ograniczeniu Cramera-Rao $D_{CR}=\frac{\left[\frac{dg(\theta)}{d\theta}\right]^2}{E_\theta\left[\frac{\partial \ln p(X,\theta)}{\partial\theta}\right]^2}$, to jest on estymatorem najefektywniejszym w klasie estymatorów regularnych. Efektywnością w sensie Cramera-Rao estymatora nieobciążonego $\hat{g}$ o wariancji $Var_\theta(\hat{g})$ nazywamy wielkość

$$\text{eff}_{CR}(\hat{g})=\frac{D_{CR}}{Var_\theta(\hat{g})}.$$

Przykład. Niech $X_1,...,X_n$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie (**iid**) $N(\theta,1)$ .Wówczas $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ jest nieobciążonym estymatorem parametru θ (w tym przypadku $g(\theta)=\theta$), gdyż $E_\theta(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n} E_\theta(X_i) = \frac{1}{n} n\theta = \theta$. Wariancja estymatora $\overline{X}$ jest równa

$$V_\theta(\overline{X}) = E_\theta(\frac{1}{n}\sum_{i=1}^{n} X_i - \theta)^2 = \frac{1}{n^2}\sum_{i,j=1}^{n} E_\theta(X_i-\theta)(X_j-\theta) = \frac{1}{n^2}\sum_{i,j=1}^{n} Cov(X_i, X_j) = \frac{1}{n}$$

i jest równa dolnemu ograniczeniu Cramera-Rao. Rzeczywiście

$$p(X_1,...,X_n,\theta) = \frac{1}{(2\pi)^{n/2}} e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i-\theta)^2} \quad , \quad \frac{\partial}{\partial\theta}(\ln p(X_1,...,X_n,\theta)) = \sum_{i=1}^{n}(X_i-\theta)$$

wiec $I_\theta = \sum_{i,j=1}^{n} E_\theta(X_i-\theta)(X_j-\theta) = \sum_{i,j=1}^{n} Cov(X_i,X_j) = n$, a stąd $Var_\theta(\overline{X}) = I_\theta^{-1} = \frac{1}{n}$.

Rozważmy pewien szczególny przypadek problemu estymacji punktowej. Niech g: $\Theta \ni \theta \to R$ będzie daną funkcją rzeczywistą, której wartość $g(\theta)\in R$ należy oszacować na podstawie obserwacji $X=(X_1,...,X_n)$. Przyjmijmy **kwadratową funkcję strat**

$L(u,v)=(v-u)^2$.

Wobec tego $L(g(\theta),\hat{g}(X))=(\hat{g}(X)- g(\theta))^2$ jest kwadratem błędu oszacowania $g(\theta)$ poprzez $\hat{g}(X)$ i jest wielkością losową. Funkcja ryzyka estymatora $\hat{g}$ jest równa

$$R_L(\theta,\hat{g}) = E_\theta\left(\hat{g}(X) - g(\theta)\right)^2$$

jest nazywana **błędem średniokwadratowym BSK** (ang.**MSE** mean square error) estymatora $\hat{g}$.

Łatwo zauważyć, że dla kwadratowej funkcji straty

$$R_L(\theta,\hat{g}) = E_\theta\left(\hat{g}(X) - g(\theta)\right)^2 = E_\theta\left(\hat{g}(X) - E_\theta(\hat{g}(X)) + E_\theta(\hat{g}(X)) - g(\theta)\right)^2 = V_\theta(\hat{g}) + b_\theta^2(\hat{g})$$

Ryzyko estymatora jest sumą jego wariancji i kwadratu obciążenia. Ta dekompozycja pokazuje, że czasami warto poszerzyć klasę estymatorów nieobciążonych o estymatory obciążone, gdyż niewielkie obciążenie może zostać zrekompensowane obniżką wariancji tak, że BSK estymatora obciążonego może być niższy od BSK najlepszego estymatora nieobciążonego.

- podejście minimaksowe polegające na porównywaniu estymatorów poprzez porównywanie maksimów globalnych ich funkcji ryzyka $\max_{\theta\in\Theta} R_L(\theta,\hat{g})$, przy czym estymator jest tym lepszy im ma mniejsze maksimum funkcji ryzyka.

Def. Estymator $\hat{g}_m$ jest estymatorem minimaksowym wielkości $g(\theta)$ w rozważanej klasie estymatorów $\mathcal{D}$, $\forall\, \hat{g} \in \mathcal{D}$ $\max_{\theta\in\Theta} R_L(\theta, \hat{g}_m) \le \max_{\theta\in\Theta} R_L(\theta, \hat{g})$.

Z uwagi na to, że nie ma generalnej potrzeby zakładania złośliwości natury, podejście minimaksowe nie cieszy się zbytnim powodzeniem wśród praktyków.

- podejście bayesowskie polegające na porównywaniu pewnych średnich wartości funkcji ryzyka dla poszczególnych estymatorów. Zakłada się tu, że statystyk posiada pewną wiedzę *a priori* o parametrze θ w postaci tak zwanego rozkładu *a priori* ν. Każdemu estymatorowi $\hat{g}$ przypisujemy wartość (liczbową) ryzyka bayesowskiego

$$r_\nu(\hat{g}) = E_\nu[R_L(\theta, \hat{g})]$$

względem rozkładu *a priori*, które jest średnią względem rozkładu *a priori* wartością funkcji ryzyka R_L.

Def. Estymator $\hat{g}_b$ jest estymatorem bayesowskim wielkości $g(\theta)$ względem rozkładu *a priori* ν w rozważanej klasie estymatorów $\mathcal{D}$, $\forall\, \hat{g} \in \mathcal{D}$ $r_\nu(\hat{g}_b) \le r_\nu(\hat{g})$.

Intuicyjnie wydaje się oczywiste, że jeżeli rozmiar n próby (czyli ilość pojedynczych obserwacji) rośnie, to należy się spodziewać coraz lepszych , czyli dokładniejszych wyników estymacji. Interesujące są więc pewne asymptotyczne własności estymatorów. Jedną z nich jest **zgodność**.

Def. Ciąg $\hat{g}_n$ estymatorów wielkości $g(\theta)$ nazywamy zgodnym gdy $\forall\, \theta \in \Theta$ $\hat{g}_n(X) \xrightarrow{wg\ P_\theta} g(\theta)$, gdy $n \to \infty$ a mocno zgodnym gdy $\forall\, \theta \in \Theta$ $\hat{g}_n(X) \xrightarrow{z\ pr.1(P_\theta)} g(\theta)$ gdy $n \to \infty$.

Przykład. Niech $X_1,\ldots,X_n$ będzie ciągiem niezależnych zmiennych losowych o tym samym rozkładzie (iid) P_θ. Wówczas Jeżeli istnieje $E_\theta(X_1)$, to $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ jest nieobciążonym i mocno zgodnym estymatorem $E_\theta(X_1)$. Jeżeli istnieje $Var_\theta(X_1)$, to $S_n^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$ jest mocno zgodnym estymatorem $Var_\theta(X_1)$ a $S_{n-1}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ jest nieobciążonym mocno zgodnym estymatorem $Var_\theta(X_1)$.

Metody estymacji

Metoda największej wiarygodności

Niech $(\mathcal{X}, \mathcal{B}, \mathcal{P}=\{P_\theta : \theta\in\Theta\})$, będzie przestrzenią statystyczną indukowaną przez zmienną losową X (zwykle wektorową) o wartościach w przestrzeni $\mathcal{X}$, przy czym wszystkie rozkłady prawdopodobieństwa P_θ: $\theta\in\Theta$ rodziny $\mathcal{P}$ mają gęstości $p(x,\theta)$ względem pewnej ustalonej miary.

Def. Dla ustalonego wyniku eksperymentu $x\in\mathcal{X}$ wielkość $L(\theta,x)=p(x,\theta)$ nazywamy wiarygodnością parametru θ, a funkcję $L(\cdot,x):\Theta \ni \theta \to L(\theta,x)$ określoną na przestrzeni parametrów Θ nazywamy funkcją wiarygodności.

Z definicji widać, że funkcja $L(\theta,x)$ dla każdego ustalonego θ jest funkcją gęstości $p(x,\theta)$ rozkładu prawdopodobieństwa P_θ na przestrzeni prób $\mathcal{X}$. Interpretacja funkcji wiarygodności jest szczególnie prosta, gdy rozważymy dyskretną przestrzeń statystyczną. Funkcja wiarygodności przypisuje parametrowi θ prawdopodobieństwo (wyznaczone z rozkładu P_θ) zaobserwowania danego wyniku eksperymentu $x\in\mathcal{X}$. W przypadku ciągłym interpretacja funkcji wiarygodności jest podobna. Prawdopodobieństwo zaobserwowania danego wyniku eksperymentu zastępujemy gęstością prawdopodobieństwa uzyskania danego wyniku eksperymentu. Ten sam wynik eksperymentu może mieć przypisane różne prawdopodobieństwa w zależności od wyboru parametru θ (czyli rozkładu P_θ). **Zasada największej wiarygodności sugeruje taki wybór parametru θ, przy którym zaobserwowany wynik eksperymentu $x\in\mathcal{X}$ jest najbardziej prawdopodobny**. W przypadku przestrzeni produktowej $(\mathcal{R}, \mathcal{B}(\mathcal{R}), \{p(x,\theta):\theta\in\Theta\})^n$ gdy wynik eksperymentu jest ciągiem $x=(x_1,\ldots,x_n)$, funkcja wiarygodności wyraża się wzorem $L(\theta,x_1,\ldots,x_n)=\prod_{i=1}^{n} p(x_i,\theta)$.

Def. Estymatorem największej wiarygodności parametru θ (ENW(θ)) (o ile istnieje) nazywamy estymator $\hat{\theta}:\mathcal{X}\ni x\to\hat{\theta}(x)=\arg\max_{\theta\in\Theta} L(\theta,x)$

Uwaga techniczna. Ze względu na monotoniczność funkcji logarytmicznej maksymalizacja funkcji wiarygodności $L(\theta,x)$ jest równoważna maksymalizacji jej logarytmu $l(\theta,x)=ln\ L(\theta,x)$.

Przy pewnych założeniach regularności (zapewniających różniczkowalność całek niewłaściwych) estymatory największej wiarygodności są:

- mocno zgodne
- asymptotycznie nieobciążone
- asymptotycznie najefektywniejsze

- asymptotycznie normalne

Szczególnie ta ostatnia własność jest bardzo cenna, gdyż pozwala konstruować asymptotyczne przedziały ufności dla estymowanego parametru w oparciu o twierdzenie

Zmienna losowa $\sqrt{n}\,(\hat{\theta}-\theta)$ ma asymptotycznie rozkład $N(0,\sigma_{as}^2)$, gdzie asymptotyczna wariancja σ_{as}^2 wyraża się wzorem $\sigma_{as}^2 = i_\theta^{-1}$, a $i_\theta = Var_\theta\left[\frac{\partial \ln p(X,\theta)}{\partial\theta}\right] = E_\theta\left[\frac{\partial \ln p(X,\theta)}{\partial\theta}\right]^2$ jest informacją w sensie Fishera dla pojedynczej obserwacji. Przy pewnych warunkach regularności $i_\theta = -E_\theta\left[\frac{\partial^2 \ln p(X,\theta)}{\partial\theta^2}\right]$.

Powyższe wyniki można uogólnić na przypadek estymacji parametru wektorowego $\theta = (\theta_1,...,\theta_k)^T$. Wówczas asymptotycznym rozkładem jest wielowymiarowy rozkład normalny z macierzą kowariancyjną będącą odwrotnością macierzy informacyjnej w sensie Fishera dla pojedynczej obserwacji

$$i_\theta = E_\theta\left[\frac{\partial \ln p(X,\theta_1,...,\theta_k)}{\partial\theta_i}\frac{\partial \ln p(X,\theta_1,...,\theta_k)}{\partial\theta_j}\right] = -E_\theta\left[\frac{\partial^2 \ln p(X,\theta_1,...,\theta_k)}{\partial\theta_i\partial\theta_j}\right].$$

Przykład Skonstruować estymator największej wiarygodności parametru θ i asymptotyczny przedział ufności na poziomie $1-\alpha = 0.95$ oparty na n niezależnych obserwacjach $X_1,X_2,...,X_n$ z rozkładu o gęstości $p(x,\theta) = \theta^2 x\, exp(-\theta x)$, $\theta > 0, x > 0$.

$$L(\theta,X_1,X_2,...,X_n) = \theta^{2n}\left(\prod_{i=1}^n X_i\right) exp\left(-\theta\sum_{i=1}^n X_i\right)$$

$$l(\theta,X_1,X_2,...,X_n) = \ln L(\theta,X_1,X_2,,,X_n) = 2n\ln\theta + \sum_{i=1}^n \ln X_i - \theta\sum_{i=1}^n \ln X_i$$

$$WK:\ \frac{2n}{\theta} - \sum_{i=1}^n X_i = 0 \text{ i WW} \Rightarrow \hat{\theta} = \frac{2}{\bar{X}} \text{ ENW}[\theta] \quad , \quad i_\theta = \frac{2}{\theta^2} \Rightarrow \sigma_{as}^2 = \frac{\theta^2}{2}$$

Korzystając z faktu, że asymptotyczny rozkład statystyki $\sqrt{n}\frac{\hat{\theta}-\theta}{\sigma_{as}}$ jest rozkładem $N(0,1)$ możemy z tablic tego rozkładu odczytać dla danego α wartość $u_{\alpha/2}$ taką, że

$$P(|\sqrt{n}\frac{\hat{\theta}-\theta}{\sigma_{as}}| < u_{\alpha/2}) = 1-\alpha \qquad \text{a stąd otrzymujemy} \qquad P(\hat{\theta} - \frac{\sigma_{as}u_{\alpha/2}}{\sqrt{n}} < \theta < \hat{\theta} + \frac{\sigma_{as}u_{\alpha/2}}{\sqrt{n}}) = 1-\alpha.$$

Po podstawieniu za σ_{as}^2 przybliżonej wartości $\frac{\hat{\theta}^2}{2}$ (wniosek z mocnej zgodności ENW) otrzymujemy asymptotyczny przedział ufności na poziomie $1-\alpha$ postaci:

$$P(\hat{\theta}(1-\frac{u_{\alpha/2}}{\sqrt{2n}}) < \theta < \hat{\theta}(1+\frac{u_{\alpha/2}}{\sqrt{2n}}) \approx 1-\alpha.$$

Przykład. Skonstruować estymator największej wiarygodności parametru θ oparty na n niezależnych obserwacjach $X_1,X_2,,,X_n$ z rozkładu jednostajnego $U[0,\theta]$.

W tym przypadku $L(\theta,X_1,X_2,...,X_n) = \begin{cases} \frac{1}{\theta^n}, \forall i: X_i \le \theta \\ 0, \exists i: X_i > \theta \end{cases} = \begin{cases} \frac{1}{\theta^n}, X_{\max} \le \theta \\ 0, X_{\max} > \theta \end{cases}$ skąd natychmiast otrzymujemy, że $X_{\max} = \max(X_1,X_2,...,X_n)$ jest ENW parametru θ. Jest to przypadek gdzie warunki regularności nie są

spełnione i nie można wykorzystać własności asymptotycznej normalności ENW do konstrukcji przedziału ufności dla θ.

Ciekawy przykład estymacji NW dla danych cenzurowanych.

Niech $X_1,...,X_{100}$ będzie próba losową z rozkładu wykładniczego o nieznanej wartości oczekiwanej a (tzn. $f(x)=\frac{1}{a}e^{-\frac{x}{a}}\mathbf{1}_{[0,\infty)}(x)$).Estymujemy a na podstawie częściowej informacji o próbce, a mianowicie na podstawie tego, iż

- 80 zmiennych (spośród wszystkich 100 z próbki) przybrało wartości poniżej 3,
- średnia arytmetyczna z tych wartości wynosi 2. Znaleźć ENW[a].

Rozwiązanie. Niech Y będzie zmienną losową zdefiniowaną wzorem

$$Y=X\cdot\mathbf{1}_{[0,3)}(X)+b\cdot\mathbf{1}_{[3,\infty)}(X)=\begin{cases}X,\text{gdy}\,X<3\\ b,\ \text{gdy}\,X\geq 3\end{cases}$$, gdzie liczba $b>3$ jest etykietą zdarzenia $X\geq 3$.

Zmienna Y ma dystrybuantę

$$F_Y(y)=\begin{cases}1-e^{-\frac{y}{a}}, y\leq 3\\ 1-e^{-\frac{3}{a}}, 3<y\leq b\\ 1,\quad y>b\end{cases}$$. Niech $\mu=\lambda+\delta_b$ gdzie λ oznacza miarę Lebesgue'a i $\forall\, A\in\mathcal{B}$ mamy

$$P(A)=\int_A \tfrac{1}{a}e^{-\frac{y}{a}}\mathbf{1}_{[0,3)}d\lambda(y)+e^{-\frac{3}{a}}\delta_b(A)=\int_A(\tfrac{1}{a}e^{-\frac{y}{a}}\mathbf{1}_{[0,3)}+e^{-\frac{3}{a}}\mathbf{1}_{\{b\}})d\mu\,.$$

Stąd $f_Y(y)=\frac{1}{a}e^{-\frac{y}{a}}\mathbf{1}_{[0,3)}(y)+e^{-\frac{3}{a}}\mathbf{1}_{\{b\}}(y)$ jest gęstością rozkłady zmiennej losowej Y względem μ.

Wobec tego funkcja wiarygodności jest postaci

$$L(a;Y_1,...,Y_{100})=\prod_{i=1}^{100}(\tfrac{1}{a}e^{-\frac{Y_i}{a}}\mathbf{1}_{[0,3)}(Y_i)+e^{-\frac{3}{a}}\mathbf{1}_{\{b\}}(Y_i))=\tfrac{1}{a^{80}}e^{-\frac{1}{a}\sum_i\{Y_i\mathbf{1}_{[0,3)}(Y_i)\}}e^{-\frac{3}{a}20}=\tfrac{1}{a^{80}}e^{-\frac{160}{a}}e^{-\frac{60}{a}}$$

$$l(a;Y_1,...,Y_{100})=\ln L(a;Y_1,...,Y_{100})=-80\ln a-\tfrac{220}{a}$$

WK: $\frac{dl}{da}=-\frac{80}{a}+\frac{220}{a^2}=0\Rightarrow\hat{a}=\frac{11}{4}$=ENW[$a$] (WW oczywisty)

Przykład. Niech $\mathbf{X}=(X_1,...,X_n)$ będzie próba prostą z rozkładu normalnego $N(m,\sigma^2)$ o funkcji

gęstości $p(x;m,\sigma^2)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-m)^2}{2\sigma^2}}$. Funkcja wiarygodności jest postaci

$$L(m,\sigma^2;\mathbf{X})=\frac{1}{(2\pi)^{\frac{n}{2}}\sigma^n}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i-m)^2}$$ a jej logarytm $l(m,\sigma^2;\mathbf{X})=-\frac{n}{2}\ln(2\pi)-n\ln\sigma-\frac{1}{2\sigma^2}\sum_{i=1}^n(X_i-m)^2$.

WK istnienia ekstremum $\begin{cases}\frac{\partial l}{\partial m}=0\\ \frac{\partial l}{\partial\sigma}=0\end{cases}\Leftrightarrow\begin{cases}\sum_{i=1}^n(X_i-m)=0\\ -\frac{n}{\sigma}+\frac{2}{2\sigma^3}\sum_{i=1}^n(X_i-m)^2=0\end{cases}$, stąd $\begin{cases}m=\overline{X}=\frac{1}{n}\sum_{i=1}^n X_i\\ \sigma^2=\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2\end{cases}$.

Pokażemy, że $(\hat{m},\hat{\sigma}^2)=(\overline{X},\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2)=ENW[(m,\sigma^2)]$, wykazując, że

$l(\hat{m},\hat{\sigma}^2;\mathbf{X})-l(m,\sigma^2;\mathbf{X})\geq 0 \quad \forall(m,\sigma^2)$ i równość zachodzi tylko dla $m=\hat{m}$ i $\sigma^2=\hat{\sigma}^2$.

$$l(\hat{m},\hat{\sigma}^2;\mathbf{X})-l(m,\sigma^2;\mathbf{X})=-\frac{n}{2}\ln\hat{\sigma}^2-\frac{n}{2}+\frac{n}{2}\ln\sigma^2+\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-m)^2\overset{(1)}{=}-\frac{n}{2}(\ln\frac{\hat{\sigma}^2}{\sigma^2}+1-\frac{\hat{\sigma}^2}{\sigma^2})+\frac{n}{2\sigma^2}(\overline{X}-m)^2\geq 0$$

Ostatnie wyrażenie jest sumą dwóch nieujemnych składników (bo $\ln x+1-x\leq 0$ i równość ma miejsce tylko dla $x=1$), które jednocześnie sie zerują tylko dla $m=\hat{m}$ i $\sigma^2=\hat{\sigma}^2$.

Uwaga: $\sum_{i=1}^{n}(X_i-m)^2\overset{(1)}{=}\sum_{i=1}^{n}(X_i-\overline{X}+\overline{X}-m)^2=\sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-m)^2=n\hat{\sigma}^2+n(\overline{X}-m)^2.$

Z własności niezmienniczości ENW widać, że $(\hat{m},\hat{\sigma})=(\overline{X},\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2})=ENW[(m,\sigma)]$.

Macierz informacji $\mathbf{i}_{(m,\sigma)}=-E_{(m,\sigma)}\begin{bmatrix}\frac{\partial^2}{\partial m^2}\ln p(X,m,\sigma) & \frac{\partial^2}{\partial\sigma\partial m}\ln p(X,m,\sigma)\\ \frac{\partial^2}{\partial m\partial\sigma}\ln p(X,m,\sigma) & \frac{\partial^2}{\partial\sigma^2}\ln p(X,m,\sigma)\end{bmatrix}=\begin{bmatrix}\frac{1}{\sigma^2} & 0\\ 0 & \frac{2}{\sigma^2}\end{bmatrix}$. Stąd

$\sqrt{n}\left(\begin{bmatrix}\hat{m}\\ \hat{\sigma}\end{bmatrix}-\begin{bmatrix}m\\ \sigma\end{bmatrix}\right)\sim as\,N\left(\begin{bmatrix}0\\ 0\end{bmatrix},\begin{bmatrix}\sigma^2 & 0\\ 0 & \frac{\sigma^2}{2}\end{bmatrix}\right)$, Podobnie $\sqrt{n}\left(\begin{bmatrix}\hat{m}\\ \hat{\sigma}^2\end{bmatrix}-\begin{bmatrix}m\\ \sigma^2\end{bmatrix}\right)\sim as\,N\left(\begin{bmatrix}0\\ 0\end{bmatrix},\begin{bmatrix}\sigma^2 & 0\\ 0 & 2\sigma^4\end{bmatrix}\right)$.

Przykład. Niech $\mathbf{X}=(X_1,...,X_n)$ będzie próba prostą z rozkładu Poissona $\mathcal{P}(\lambda)$ o funkcji gęstości $p(x;\lambda)=\frac{\lambda^x}{x!}e^{-\lambda}$, $x=0,1,...$ względem miary liczącej $\mu(\cdot)=\sum_{k=0}^{\infty}\delta_k(\cdot)$. Funkcja wiarygodności jest postaci

$L(\lambda;\mathbf{X})=\frac{\lambda^{\sum_{i=1}^{n}X_i}}{\prod_{i=1}^{n}X_i!}e^{-n\lambda}$ a jej logarytm $l(\lambda;\mathbf{X})=\sum_{i=1}^{n}X_i\ln\lambda-n\lambda-\ln\prod_{i=1}^{n}X_i!$.

WK : $\frac{\partial}{\partial\lambda}l(\lambda;\mathbf{X})=0\Leftrightarrow\frac{\sum_{i=1}^{n}X_i}{\lambda}-n=0\Rightarrow\lambda=\overline{X}$. Badając monotoniczność funkcji $l(\lambda;\mathbf{X})$ stwierdzamy, że $\hat{\lambda}=\overline{X}=ENW[\lambda]$. Widać , że $\frac{\partial^2}{\partial\lambda^2}\ln p(X,\lambda)=-\frac{X}{\lambda^2}$, stąd $i_\lambda=-E_\lambda(-\frac{X}{\lambda^2})=\frac{1}{\lambda}$, więc $\sqrt{n}(\overline{X}-\lambda)\sim as\,N(0,\lambda)$.

Kłopoty z metodą największej wiarygodności

- estymator największej wiarygodności może nie istnieć ,
- estymator największej wiarygodności może nie być określony jednoznacznie,
- efektywne wyznaczenie estymatora największej wiarygodności może być bardzo trudne.

Metoda momentów

Metodę momentów pochodzącą od K. Pearsona można krótko opisać w następujący sposób:

Niech $X_1,X_2,,,X_n$ będzie ciągiem **iid** o rozkładzie P_θ zależnym od wektorowego parametru $\theta=(\theta_1,..,\theta_k)$.

- wyznaczamy k pierwszych momentów zwykłych : $m_r=h_r(\theta_1,..,\theta_k)$.

- estymujemy momenty teoretyczne m_r momentami empirycznymi $\hat{m}_r = \frac{1}{n}\sum_{i=1}^{n} X_i^r$
- rozwiązujemy układ równań $\begin{cases} \hat{m}_1 = h_1(\theta_1,...,\theta_k) \\ \vdots \\ \hat{m}_k = h_k(\theta_1,...,\theta_k) \end{cases}$ i otrzymujemy estymatory $(\hat{\theta}_1,...,\hat{\theta}_k)$ **mocno zgodne**.

Przykład. Niech $X_1,X_2,,,X_n$ iid N(m,σ^2). Wiadomo, że $E(X_i)=m$ $E(X_i^2)=m^2+\sigma^2$. Z układu $\hat{m}_r = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\overline{X^2} = \frac{1}{n}\sum_{i=1}^{n} X_i^2 = m^2+\sigma^2$ otrzymujemy $\hat{m} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}$, $\hat{\sigma}^2 = S_n^2$.

Metoda Markowa

Metoda Markowa pochodząca z przełomu XIX i XX wieku polega na wyznaczaniu estymatorów

- nieobciążonych
- liniowych względem obserwowanych zmiennych losowych
- posiadających najmniejszą wariancje w klasie wszystkich estymatorów liniowych

Metodę Markowa można więc potraktować jako szczególny przypadek estymacji nieobciążonej o minimalnej wariancji, który ze względu na postulowaną liniowość jest szczególnie prosty i wygodny obliczeniowo. Wyjaśnimy to na przykładzie.

Przykład. Niech $X_1,...,X_n$ będzie ciągiem niezależnych zmiennych losowych o tych samych wartościach oczekiwanych $E(X_i)=m$ i znanych wariancjach $V(X_i)=\sigma_i^2$ .Znaleźć nieobciążony estymator liniowy dla m o najmniejszej wariancji.

Z uwagi na postulowaną liniowość estymator wartości oczekiwanej m ma postać $\hat{m} = \sum_{i=1}^{n} a_i X_i$.

Warunek nieobciążoności $m = E(\hat{m}) = \sum_{i=1}^{n} a_i E(X_i) = m\sum_{i=1}^{n} a_i$ prowadzi do równości $\sum_{i=1}^{n} a_i = 1$.

Wariancja estymatora

$$V(\hat{m}) = E(\sum_{i=1}^{n} a_i X_i - m)^2 = E(\sum_{i=1}^{n} a_i X_i - \sum_{i=1}^{n} a_i m)^2 = E(\sum_{i=1}^{n} a_i (X_i - m))^2 = \sum_{i=1}^{n} a_i^2 \sigma_i^2 .$$

Aby znaleźć estymator typu Markowa należy więc znaleźć współczynniki $a_1,...,a_n$ minimalizujące $\sum_{i=1}^{n} a_i^2 \sigma_i^2$ przy warunku $\sum_{i=1}^{n} a_i = 1$. Rozwiązując powyższy problem ekstremum warunkowego poprzez rozwikłanie ograniczeń i redukcję problemu do problemu ekstremum bezwarunkowego n-1 zmiennych lub stosując metodę mnożników Lagrange'a otrzymujemy $a_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^{n} \frac{1}{\sigma_i^2}}$.

Metoda najmniejszych kwadratów (MNK)

Obserwujemy zmienne losowe $Y_1,...,Y_n$ o których wiemy, że

$$E(Y_i)=g_i(\theta)\ ,\ i=1,...,n,$$

gdzie g_i: $\Re^k \supset \Theta \ni \theta \to \Re$ są znanymi funkcjami.

Jeżeli parametr θ przebiega zbiór Θ, to punkt $(g_1(\theta),...,g_n(\theta))$ przebiega pewien zbiór $\Gamma \subset \Re^n$. Zaobserwowany punkt $\boldsymbol{Y}=(Y_1,...,Y_n)$ również leży w $\Re^n$. Idea MNK polega na tym, żeby w zbiorze Γ znaleźć punkt $\boldsymbol{\gamma}(Y_1,...,Y_n)$ najbliższy zaobserwowanemu punktowi $\boldsymbol{Y}$ a następnie za oszacowanie parametru θ przyjąć taki punkt $\hat{\theta} \in \Theta$, któremu odpowiada wyznaczony punkt $\boldsymbol{\gamma}$, tzn. taki $\hat{\theta}$, że

$$(g_1(\hat{\theta}),...,g_n(\hat{\theta}))=\boldsymbol{\gamma}(Y_1,...,Y_n).$$

Zwykłe oba etapy łączy się w jeden i za estymator MNK przyjmuje się θ minimalizujące wielkość

$$\sum_{i=1}^{n}(Y_i - g_i(\theta))^2 .$$

MNK znajduje szczególne zastosowanie w tzw. liniowych modelach statystyki matematycznej i zostanie omówiona później.

Przykład. Niech X_1, X_2, X_3, X_4 będą wynikami lotniczych pomiarów kątów θ_1, θ_2, θ_3, θ_4 pewnego czworokąta na powierzchni ziemi. Załóżmy, że obserwacje te są obarczone błędami losowymi, które są niezależne, mają wartość oczekiwaną równą zeru i wspólną wariancję σ^2. Wyznaczyć metodą najmniejszych kwadratów estymatory wielkości kątów. Wyznaczyć nieobciążony estymator wariancji σ^2.

Przypuśćmy, że wiadomo, że dany czworokąt jest równoległobokiem, takim, że $\theta_1=\theta_3$ oraz $\theta_2=\theta_4$. Jaką postać mają wtedy estymatory kątów i jak można by szacować σ^2.

Oznaczając losowe błędy pomiarów przez ε_i możemy zapisać następujący model

$$X_i=\theta_i+\varepsilon_i,\quad E(X_i)=\theta_i,\quad V(X_i)=V(\varepsilon_i)=\sigma^2,\quad i=1,...,4,\quad \theta_1+\theta_2+\theta_3+\theta_4=2\pi.$$

Wyznaczenie estymatorów kątów sprowadza się do następującego zagadnienia ekstremum warunkowego:

Znaleźć argument minimalizujący $S(\theta_1,\theta_2,\theta_3,\theta_4)=\sum_{i=1}^{4}(X_i-\theta_i)^2$ przy warunku $\theta_1+\theta_2+\theta_3+\theta_4=2\pi$. Po rozwikłaniu ograniczeń otrzymujemy zagadnienie minimalizacji funkcji

$$S(\theta_1,\theta_2,\theta_3)=(X_1-\theta_1)^2+(X_2-\theta_2)^2+(X_3-\theta_3)^2+(X_4-2\pi+\theta_1+\theta_2+\theta_3)^2 .$$

Warunek konieczny istnienia ekstremum prowadzi do układu równań

$$\begin{bmatrix}2&1&1\\1&2&1\\1&1&2\end{bmatrix}\begin{bmatrix}\theta_1\\\theta_2\\\theta_3\end{bmatrix}=\begin{bmatrix}2\pi-X_4+X_1\\2\pi-X_4+X_2\\2\pi-X_4+X_3\end{bmatrix},$$

z którego otrzymujemy : $\hat{\theta}_i=\frac{1}{4}(2\pi-\sum_{i=1}^{4}X_i)+X_i$, $i=1,...,4.$

Warunek wystarczający potwierdza, że powyższe wzory rzeczywiście określają estymatory MNK kątów.

Łatwo zauważyć, że $\sum_{i=1}^{4}(X_i-\hat{\theta}_i)^2$ jest nieobciążonym estymatorem σ^2, gdyż

$$E(\sum_{i=1}^{4}(X_i-\hat{\theta}_i)^2)=E(\tfrac{1}{4}(2\pi-\sum_{i=1}^{4}X_i))^2=\tfrac{1}{4}E(\sum_{i=1}^{4}\varepsilon_i)^2=\tfrac{1}{4}\sum_{i,j=1}^{4}E(\varepsilon_i\varepsilon_j)=\sigma^2.$$

Rozwiązanie drugiej części zadania prowadzi do minimalizacji $S(\theta_1,\theta_2,\theta_3,\theta_4)=\sum_{i=1}^{4}(X_i-\theta_i)^2$ przy warunkach $\theta_1+\theta_2+\theta_3+\theta_4=2\pi$, $\theta_1=\theta_3$ oraz $\theta_2=\theta_4$. Po rozwikłaniu ograniczeń otrzymujemy zagadnienie minimalizacji

$$S(\theta_1)=(X_1-\theta_1)^2+(X_2-\pi+\theta_1)^2+(X_3-\theta_1)^2+(X_4-\pi+\theta_1)^2,$$

które prowadzi do estymatorów MNK

$$\hat{\theta}_1=\hat{\theta}_3=\tfrac{1}{4}(2\pi-X_2-X_4+X_1+X_3),\qquad \hat{\theta}_2=\hat{\theta}_4=\tfrac{1}{4}(2\pi-X_1-X_3+X_2+X_4).$$

$$E(\sum_{i=1}^{4}(X_i-\hat{\theta}_i)^2)=E(\left(X_1-\tfrac{1}{4}(2\pi-X_2-X_4+X_1+X_3)\right)^2+E(\left(X_2-\tfrac{1}{4}(2\pi+X_2+X_4-X_1-X_3)\right)^2+$$

$$+E(\left(X_3-\tfrac{1}{4}(2\pi-X_2-X_4+X_1+X_3)\right)^2+E(\left(X_4-\tfrac{1}{4}(2\pi+X_2+X_4-X_1-X_3)\right)^2=$$

$$\tfrac{1}{16}E\left(3(X_1-\theta_1)-(X_3-\theta_3)+(X_2-\theta_2)+(X_4-\theta_4)\right)^2+\tfrac{1}{16}E\left(3(X_2-\theta_2)-(X_4-\theta_2)+(X_3-\theta_3)+(X_4-\theta_4)\right)^2+$$

$$+\tfrac{1}{16}E\left(3(X_3-\theta_3)-(X_1-\theta_1)+(X_2-\theta_2)+(X_4-\theta_4)\right)^2+\tfrac{1}{16}E\left(3(X_4-\theta_4)-(X_2-\theta_2)+(X_1-\theta_1)+(X_3-\theta_3)\right)^2=$$

$$=4\tfrac{3}{4}\sigma^2=3\sigma^2.$$

Stąd $\tfrac{1}{3}\sum_{i=1}^{4}(X_i-\hat{\theta}_i)^2$ jest nieobciążonym estymatorem σ^2.

Zagadnienie ekstremum warunkowego można w obu przypadkach rozwiązać również metodą mnożników Lagrange'a.

Przedziały ufności –estymacja przedziałowa

Niech $\mathbf{X}=(X_1,...,X_n)$ będzie próbą prostą z rozkładu $P_\theta \in \mathcal{P}=\{P_\theta:\theta\in\Theta\}$.

Definicja. Losowy przedział $[L(\mathbf{X}),U(\mathbf{X})]$, gdzie $\forall\theta\in\Theta \quad P_\theta(L(\mathbf{X})<U(\mathbf{X}))=1$, taki, że dla zadanego $\alpha\in(0,1) \quad P_\theta(L(\mathbf{X})\le g(\theta)\le U(\mathbf{X}))=1-\alpha \quad \forall\theta\in\Theta$, nazywamy $100(1-\alpha)\%$ **przedziałem ufności** dla parametru $g(\theta)\in R$. Statystyki $L(\mathbf{X})$ i $U(\mathbf{X})$ nazywamy odpowiednio dolnym i górnym końcem przedziału a współczynnik $1-\alpha$ nazywamy **poziomem ufności**.

Interpretacja. Prawdziwa nieznana, nielosowa wartość $g(\theta)$ z prawdopodobieństwem $1-\alpha$ należy do losowego przedziału (jest pokryta losowym przedziałem) $[L(\mathbf{X}),U(\mathbf{X})]$. Nie można tu mówić o prawdopodobieństwie, że nieznany parametr będzie zawarty w jakimś stałym przedziale. Takie zdanie miałoby sens gdyby nieznany parametr był zmienną losową **a nie jest.**

Konstrukcja przedziałów ufności za pomocą funkcji centralnej (wiodącej)

Definicja. Niech $\mathbf{X}=(X_1,...,X_n)$ próbą prostą z rozkładu $P_\theta \in \mathcal{P}=\{P_\theta:\theta\in\Theta\}$. Funkcja $Q(\mathbf{X},g(\theta))$ (to nie jest statystyka) nazywa się **funkcją centralną** dla parametru $g(\theta)\in R$, jeżeli jej rozkład prawdopodobieństwa nie zależy od parametru θ.

Przykłady

- Niech $\mathbf{X}=(X_1,...,X_n)$ będzie próbą prostą z rozkładu $N(m,\sigma_0^2)$, gdzie σ_0^2 znane. Funkcja $Q(\mathbf{X},m)=\dfrac{\sqrt{n}(\overline{X}-m)}{\sigma_0}$ mająca rozkład $N(0,1)$ jest funkcją centralną dla m
- Niech Niech $\mathbf{X}=(X_1,...,X_n)$ będzie próbą prostą z rozkładu $N(m,\sigma^2)$. Oznaczmy $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ i $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$. Funkcja $Q(X_1,...,X_n,m)=\dfrac{\sqrt{n}(\overline{X}-m)}{S}$ mająca rozkład t_{n-1} (t - Studenta z n-1 stopniami swobody) jest funkcją centralną dla m.

  Funkcja $Q(X_1,...,X_n,\sigma^2)=\dfrac{(n-1)S^2}{\sigma^2}$ mająca rozkład χ^2_{n-1} jest funkcją centralną dla σ^2.

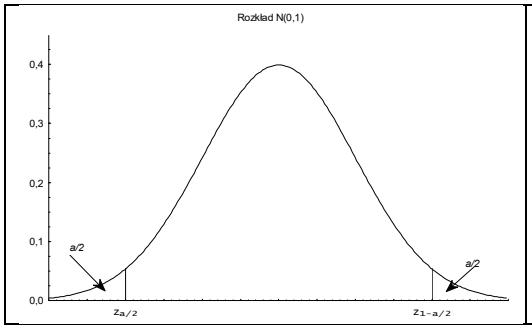
Załóżmy, że dysponujemy funkcją centralną $Q(\mathbf{X},g(\theta))$. Przedział ufności konstruujemy w następujący sposób:
Wybieramy liczby a i b tak aby spełniały nierówność $P_\theta(a\le Q(\mathbf{X},g(\theta))\le b)=1-\alpha \quad \forall\theta\in\Theta$. **Gdy funkcja centralna $Q(\mathbf{X},g(\theta))$ jest ciągłą i ściśle monotoniczną funkcją parametru** $g(\theta)\in R$, to

nierówność $a \le Q(\mathbf{X}, g(\theta)) \le b$ jest równoważna nierówności $[L(\mathbf{X},a,b) \le g(\theta) \le U(\mathbf{X},a,b)]$.

Stąd przedział $[L(\mathbf{X},a,b), U(\mathbf{X},a,b)]$ jest $100(1-\alpha)\%$ **przedziałem ufności** dla parametru $g(\theta) \in R$.

- **Przedział ufności dla m w rozkładzie $N(m,\sigma_0^2)$, gdzie σ_0^2 znane.**

Funkcja centralna $Q(X_1,...,X_n,m) = \dfrac{\sqrt{n}(\overline{X}-m)}{\sigma_0}$ ma rozkład $N(0,1)$. Wobec tego



$$\forall\ m \quad P_m\left(\left|\frac{\sqrt{n}(\overline{X}-m)}{\sigma_0}\right| \le z_{(1-\alpha/2)}\right) = 1-\alpha$$

co jest równoważne

$$P_m\left(\overline{X} - z_{(1-\alpha/2)}\frac{\sigma_0}{\sqrt{n}} \le m \le \overline{X} + z_{(1-\alpha/2)}\frac{\sigma_0}{\sqrt{n}}\right) = 1-\alpha$$

otrzymaliśmy wiec przedział ufności $[\overline{X} - z_{(1-\alpha/2)}\dfrac{\sigma_0}{\sqrt{n}}, \overline{X} + z_{(1-\alpha/2)}\dfrac{\sigma_0}{\sqrt{n}}]$ o stałej długości

$$l(n,\alpha) = 2\, z_{(1-\alpha/2)}\frac{\sigma_0}{\sqrt{n}}$$

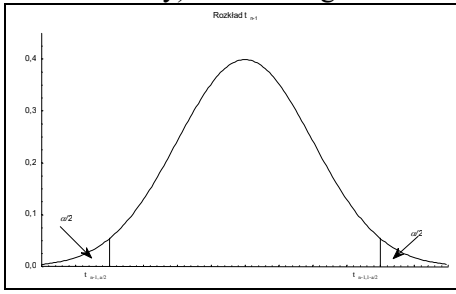
Tą konstrukcję można powielić konstruując **asymptotyczny przedział ufności** dla pojedynczego parametru estymowanego MNW

$$\forall\ \theta \quad P_\theta\left(\hat{\theta} - z_{(1-\alpha/2)}\frac{\sigma_{as}}{\sqrt{n}} \le \theta \le \hat{\theta} + z_{(1-\alpha/2)}\frac{\sigma_{as}}{\sqrt{n}}\right) \approx 1-\alpha$$

gdzie za asymptotyczną wariancję $\sigma_{as}^2 = i_\theta^{-1}$ należy przyjąć jej estymator. Dokładniej w informacji Fishera $i_\theta = i(\theta)$ za parametr θ należy przyjąć $\hat{\theta}$ =ENW[θ]. Poprawność tej konstrukcji jest prostym wnioskiem z asymptotycznych własności estymatorów NW i twierdzenia Słuckiego.

- **Przedział ufności dla m w rozkładzie $N(m,\sigma^2)$, gdzie σ^2 nieznane.**

Funkcja centralna $Q(X_1,...,X_n,m) = \dfrac{\sqrt{n}(\overline{X}-m)}{S}$ ma rozkład t_{n-1} (t - Studenta z n-1 stopniami swobody) . Wobec tego



$$\forall\ (m,\sigma^2) \quad P_{(m,\sigma)}\left(\left|\frac{\sqrt{n}(\overline{X}-m)}{S}\right| \le t_{(n-1,1-\alpha/2)}\right) = 1-\alpha$$

co jest równoważne

$$P_{(m,\sigma)}\left(\overline{X} - t_{(n-1,1-\alpha/2)}\frac{S}{\sqrt{n}} \le m \le \overline{X} + t_{(n-1,1-\alpha/2)}\frac{S}{\sqrt{n}}\right) = 1-\alpha$$

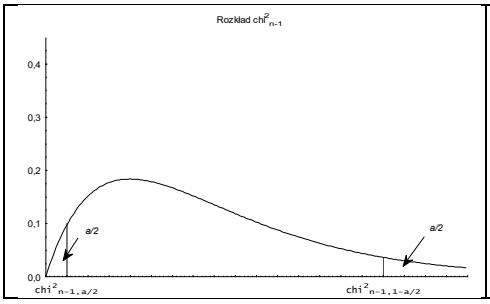
otrzymaliśmy wiec przedział ufności $[\overline{X} - t_{(n-1,1-\alpha/2)} \frac{S}{\sqrt{n}}, \overline{X} + t_{(n-1,1-\alpha/2)} \frac{S}{\sqrt{n}}]$ o losowej długości

$l(n,\alpha) = 2t_{(n-1,1-\alpha/2)} \frac{S}{\sqrt{n}}$ o wartości oczekiwanej długości

$$E_{(m,\sigma)} l(n,\alpha) = 2t_{(n-1,1-\alpha/2)} \frac{E_{(m,\sigma)}(S)}{\sqrt{n}} = 2t_{(n-1,1-\alpha/2)} \frac{\sigma E(\sqrt{\chi^2_{n-1}})}{\sqrt{n}} = 2t_{(n-1,1-\alpha/2)} \frac{const}{\sqrt{n}}$$

- **Przedział ufności dla σ^2 w rozkładzie $N(m,\sigma^2)$**

Funkcja $Q(X_1,...,X_n,\sigma^2) = \frac{(n-1)S^2}{\sigma^2}$ ma rozkład χ^2_{n-1} . Wobec tego



$$\forall\ (m,\sigma^2) \quad P_{(m,\sigma)}(\chi^2_{n-1,\alpha/2} \le \frac{(n-1)S^2}{\sigma^2} \le \chi^2_{n-1,1-\alpha/2}) = 1-\alpha$$

co jest równoważne

$$\forall\ (m,\sigma^2) \quad P_{(m,\sigma)}(\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}) = 1-\alpha$$

otrzymaliśmy wiec przedział ufności $[\frac{(n-1)S^2}{\chi^2_{n-1,1-\alpha/2}}, \frac{(n-1)S^2}{\chi^2_{n-1,\alpha/2}}]$ o losowej długości

$l(n,\alpha) = (n-1)S^2(\frac{1}{\chi^2_{n-1,1-\alpha/2}} - \frac{1}{\chi^2_{n-1,\alpha/2}})$ o wartości oczekiwanej długości

$$E_{(m,\sigma)} l(n,\alpha) = (n-1)E(S^2)(\frac{1}{\chi^2_{n-1,1-\alpha/2}} - \frac{1}{\chi^2_{n-1,\alpha/2}}) = (n-1)\sigma^2(\frac{1}{\chi^2_{n-1,1-\alpha/2}} - \frac{1}{\chi^2_{n-1,\alpha/2}})$$

W przypadku symetrycznego rozkładu funkcji centralnej wybór liczb *a* i *b* tak aby spełniały nierówność $P_\theta(a \le Q(X_1,...,X_n,\theta) \le b)=1-\alpha \ \forall \theta \in \Theta$ wydaje się oczywisty. W przypadku rozkładu niesymetrycznego wybór nie jest już oczywisty. Można postawić problem takiego doboru *a* i *b*, aby długość $l(a,b)$ przedziału $[L(X_1,...,X_n, a,b), U(X_1,...,X_n, a,b)]$ (gdy jest nielosowa) lub wartość oczekiwana długości była minimalna. Mówimy wtedy o **najkrótszych przedziałach ufności** na poziomie $1-\alpha$.. Jeżeli *F* oznacza **dystrybuantę rozkładu funkcji centralnej**, to należy rozwiązać następujące zagadnienie optymalizacyjne

Znaleźć *a* i *b* tak aby

$l(a,b) \to \min$

pod warunkiem

$F(b)-F(a)=1-\alpha$.

W rozważanym powyżej problemie

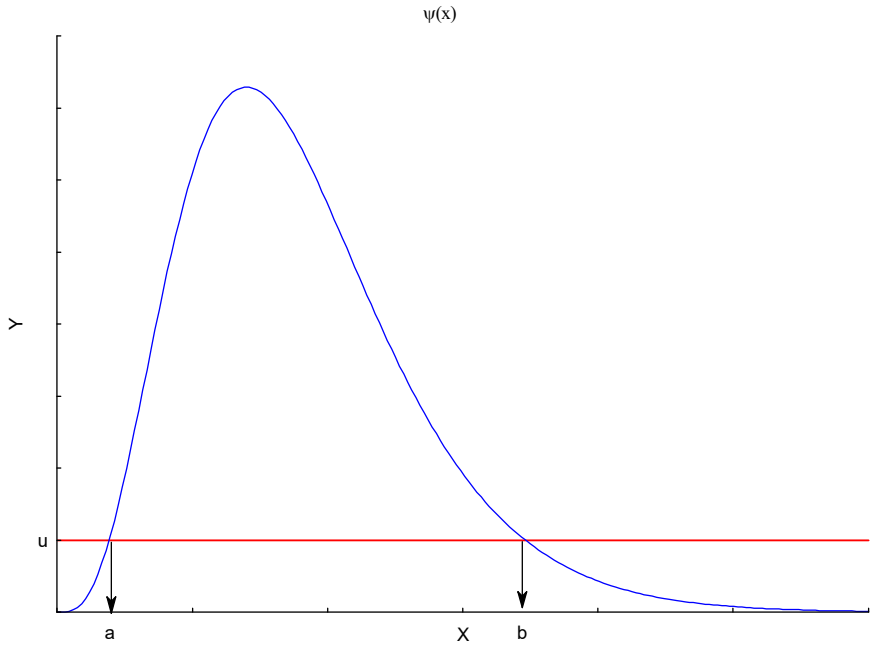
- $l(a,b) = (n-1)\sigma^2(\frac{1}{a} - \frac{1}{b})$
- $F_{\chi^2_{n-1}}(b) - F_{\chi^2_{n-1}}(a) = 1-\alpha$

Aby rozwiązać powyższy problem ekstremum warunkowego tworzymy funkcję Lagrange'a

$$L(a,b;\lambda)=(n-1)\sigma^2(\tfrac{1}{a}-\tfrac{1}{b})+\lambda(F_{\chi^2_{n-1}}(b)-F_{\chi^2_{n-1}}(a)-(1-\alpha)).$$

Z WK otrzymujemy układ $\begin{cases}\frac{\partial L}{\partial a}=-\frac{(n-1)\sigma^2}{a^2}-\lambda f_{\chi^2_{n-1}}(a)=0\\ \frac{\partial L}{\partial b}=\frac{(n-1)\sigma^2}{b^2}+\lambda f_{\chi^2_{n-1}}(b)=0\\ \frac{\partial L}{\partial \lambda}=F_{\chi^2_{n-1}}(b)-F_{\chi^2_{n-1}}(a)-(1-\alpha)=0\end{cases}\Leftrightarrow\begin{cases}-\frac{(n-1)\sigma^2}{\lambda}=a^2 f_{\chi^2_{n-1}}(a)\\ -\frac{(n-1)\sigma^2}{\lambda}=b^2 f_{\chi^2_{n-1}}(b)\\ F_{\chi^2_{n-1}}(b)-F_{\chi^2_{n-1}}(a)=1-\alpha\end{cases}$

Niech $\psi(x)=x^2 f_{\chi^2_{n-1}}(x)$



Należy tak dobrać poziom u aby para $\{a,b\}=\psi^{-1}[\{u\}]$ spełniała warunek $F_{\chi^2_{n-1}}(b)-F_{\chi^2_{n-1}}(a)=1-\alpha$

Testowanie hipotez

Niech $(\mathcal{X}, \mathcal{B}, \mathcal{P}=\{P_\theta: \theta\in\Theta\})$ będzie przestrzenią statystyczną, przy czym

$$\Theta = \Theta_0 \cup \Theta_1 \text{ i } \Theta_0 \cap \Theta_1 = \varnothing.$$

Problem testowania hipotez można sformułować następująco: na podstawie obserwacji $X\in\mathcal{X}$ zweryfikować hipotezę

$$H_0: \theta\in\Theta_0 \qquad \text{wobec alternatywy} \qquad H_1: \theta\in\Theta_1.$$

Problem testowania hipotezy nie jest problemem badawczym lecz decyzyjnym. Możemy podjąć jedną z dwóch możliwych decyzji

- d_o - akceptacja hipotezy H_0
- d_1 - odrzucenia H_0 i akceptacji H_1

Możemy więc podjąć poprawna decyzję lub popełnić jeden z dwóch rodzajów błędów

- **błąd pierwszego rodzaju** $d_1|H_0$ polegający na odrzuceniu H_0 , gdy w rzeczywistości jest ona prawdziwa,
- **błąd drugiego rodzaju** $d_0|H_1$ polegający na akceptacji H_0 , gdy w rzeczywistości jest ona fałszywa.

	d_0	d_1
H_0	Decyzja poprawna	Błąd I rodzaju $d_1\|H_0$
H_1	Błąd II rodzaju $d_0\|H_1$	Decyzja poprawna

Potrzebujemy pewnej procedury, która podpowie jaką podjąć decyzję gdy dysponujemy wynikiem eksperymentu X. Rozwiązaniem problemu (H_0, H_1) testowania hipotezy H_0 przeciwko alternatywie H_1 będzie pewna funkcja $\varphi: \mathcal{X}\to[0,1]$ zwana **zrandomizowanym testem statystycznym**. Dysponując testem statystycznym φ i X:

- z prawdopodobieństwem $1-\varphi(X)$ podejmujemy decyzję d_o - akceptacji hipotezy H_0
- z prawdopodobieństwem $\varphi(X)$ podejmujemy decyzję d_1 - odrzucenia H_0 i akceptacji H_1.

Aby więc podjąć konkretną decyzję należy więc użyć pewnego mechanizmu losowego, który produkuje dwa wyniki z prawdopodobieństwami $\varphi(X)$ i $1-\varphi(X)$ i na tej podstawie podjąć (wylosować) decyzję. Podejście takie jest krytykowane przez praktyków, jako stwarzające możliwość nadużyć w interpretacji danych. Zdecydowanie większe uznanie praktyków znalazły testy niezrandomizowane, dla których zbiór wartości funkcji φ jest zbiorem dwuelementowym $\{0,1\}$. Test taki dzieli przestrzeń prób na dwa rozłączne zbiory:

$C=\varphi^{-1}(\{1\})=\{X\in\mathcal{X}: \varphi(X=1\}$ zwany zbiorem odrzucenia hipotezy H_0 (lub zbiorem krytycznym

i

$\varphi^{-1}(\{0\})=\{X\in\mathcal{X}: \varphi(X)=0\}$ zwany zbiorem akceptacji H_0.

Konstrukcja testu niezrandomizownego jest więc równoważna rozbiciu przestrzeni prób na dwa rozłączne podzbiory : odrzucenia i akceptacji H_0. Jeżeli zaobserwowana wartość zmiennej losowej X

wpada do zbioru krytycznego $C=\varphi^{-1}(\{1\})$, to odrzucamy H_0, w przeciwnym przypadku akceptujemy H_0.

Zwykle w konstrukcji testu występuje etap pośredni polegający na wyznaczeniu tzw. statystyki testowej T: $\mathcal{X} \ni X \to T(X)$, która przenosi dalsze rozważania z przestrzeni $(\mathcal{X}, \mathcal{B}, \mathcal{P}=\{P_\theta: \theta \in \Theta\})$ do „prostszej" przestrzeni $(R, \mathcal{B}, \mathcal{P}^T=\{P_\theta^T : \theta \in \Theta\})$. Załóżmy, że duże wartości statystyki T świadczą przeciwko hipotezie zerowej. Mówimy wówczas o teście z prawostronnym obszarem krytycznym.

Niech t_{obs} oznacza zaobserwowaną wartość statystyki testowej T. Reguła decyzyjna jest następująca:

$$\text{jeżeli } t_{obs} \geq t_{kryt} \Rightarrow \text{odrzucamy } H_0.$$

Komputerowe realizacje zwykle podają tzw. p-wartość $p = \sup_{\theta \in \Theta_0} P(T(X) \geq t_{obs})$.

$$\text{Reguła decyzyjna jest następująca: jeżeli } p \leq \alpha \Rightarrow \text{odrzucamy } H_0.$$

Oczywiście nie każdy test statystyczny jest dobrym testem. Aby móc porównywać testy i w konsekwencji wybrać najlepszy można podobnie jak w problemie estymacji punktowej próbować określić stratę jaką ponosi statystyk podejmując błędną decyzję. Następnym krokiem jest wtedy wyznaczenie średniej straty dla danego testu, czyli wyznaczenie jego funkcji ryzyka. Testy, podobnie jak estymatory, można porównywać, porównując ich funkcje ryzyka. Można tu również wyróżnić :

- podejście globalne (porównywanie funkcji ryzyka) z zawężaniem klasy testów,
- podejście minimaksowe,
- podejście bayesowskie.

Największe uznanie zyskało jednak w teorii testów podejście **Neymana-Pearsona**. W ujęciu tym przypisuje się różną wagę błędom I i II rodzaju. Ważniejszy jest błąd pierwszego rodzaju. Dlatego nakłada się pewne ograniczenie na prawdopodobieństwo błędu I rodzaju i wśród dostępnych testów spełniających to ograniczenie poszukuje się testu, który minimalizuje prawdopodobieństwo błędu II rodzaju.

Def. Funkcją mocy testu φ nazywamy funkcję $\beta_\varphi(\theta)=E_\theta(\varphi(X))$ $(= P_\theta(C)$ dla testu niezrandomizowanego)

Łatwo zauważyć, że dla testu niezrandomizowanego mamy:

$\theta \in \Theta_0$ (czyli prawdziwa jest H_0) $\Rightarrow \beta_\varphi(\theta)=1 \cdot P_\theta(d_1|H_0)+0 \cdot P_\theta(d_0|H_0)=P_\theta(d_1|H_0)$ (pr. błędu I rodzaju)

$\theta \in \Theta_1$ (czyli prawdziwa jest H_1) $\Rightarrow \beta_\varphi(\theta)=1 \cdot P_\theta(d_1|H_1)+0 \cdot P_\theta(d_0|H_1)=P_\theta(d_1|H_1)=1-P_\theta(d_0|H_1)$

(pr. braku błędu II rodzaju- inaczej **moc testu** φ).

Def. **Rozmiarem testu** φ nazywamy liczbę $\sup_{\theta \in \Theta_0} \beta_\varphi(\theta)$ (lub $\sup_{\theta \in \Theta_0} P_\theta(C)$ dla testu niezrandomizowanego).

Def. Mówimy, że test φ hipotezy H_0 przeciwko H_1 jest **testem na poziomie istotności** α, jeżeli $\beta_\varphi(\theta) \leq \alpha$, czyli rozmiar testu nie przekracza α.

Wśród testów na poziomie istotności α poszukujemy najlepszego w sensie Neymana-Pearsona czyli najmocniejszego. Klasyczny lemat Neymana-Pearsona podaje sposób konstrukcji testu najmocniejszego dla testowania **hipotezy prostej** H_0: $\theta=\theta_0$ (czyli $\Theta_0=\{\theta_0\}$) przeciwko **prostej alternatywie** H_1: $\theta=\theta_1$ (czyli $\Theta_1=\{\theta_1\}$). Hipoteza prosta specyfikuje więc dokładnie jeden rozkład prawdopodobieństwa określony na przestrzeni prób. Hipoteza H_0 głosi, że prawdziwym rozkładem na przestrzeni prób jest rozkład $P_0=P_{\theta_0}$ a konkurencyjna hipoteza H_1 głosi, że prawdziwym rozkładem na przestrzeni prób jest rozkład $P_1=P_{\theta_1}$. Hipotezę, która nie jest prosta (tzn. specyfikuje przynajmniej dwuelementową rodzinę rozkładów prawdopodobieństwa) nazywamy **hipotezą złożoną**.

Lemat Neymana-Pearsona. Jeżeli rozkłady P_0 i P_1 mają gęstości p_0 i p_1 (w przypadku ciągłym względem miary Lebesguea a w przypadku dyskretnym względem miary liczącej) , to test najmocniejszy na poziomie α hipotezy H_0 przeciwko H_1 istnieje i ma postać

$$\varphi(x)=\begin{cases}1 & \text{gdy} \quad p_1(x)>kp_0(x)\\ \gamma & \text{gdy} \quad p_1(x)=kp_0(x)\\ 0 & \text{gdy} \quad p_1(x)<kp_0(x)\end{cases},$$

gdzie $k\geq 0$ i $\gamma\in<0,1)$ są tak dobranymi stałymi, aby rozmiar testu był równy α

Z postaci funkcji testowej φ wynika, że optymalny test H_0 przeciwko H_1 jest testem zrandomizowanym. W przypadku testowania hipotez dotyczących ciągłych rozkładów prawdopodobieństwa można stałą γ przyjąć równą 0 i otrzymać w ten sposób test niezrandomizowany. W przypadku dyskretnym zwykle nie można znaleźć testu niezrandomizowanego o zadanym rozmiarze. Można jednak znaleźć optymalny test niezrandomizowany o rozmiarze zbliżonym do zadanego. Lemat Neymana-Pearsona podaje porządek w jakim należy włączać poszczególne punkty przestrzeni prób do zbioru krytycznego. Porządek ten wyznacza wartość ilorazu $\frac{p_1(x)}{p_0(x)}$. Im wyższa wartość tego ilorazu tym szybciej punkt x trafia do zbioru krytycznego. Proces ten należy przerwać wówczas, gdy dołączenie następnych punktów do zbioru krytycznego powoduje przekroczenie założonego poziomu istotności.

Przykład. Niech $(X_1,...,X_n)$ (n=9) będzie ciągiem **iid** o rozkładzie P z dystrybuantą F. Znaleźć test najmocniejszy na poziomie $\alpha=0{,}05$ dla problemu testowania hipotezy H_0: $P=N(0,1)$ przeciwko alternatywie H_1: $P=N(1,1)$.

Zgodnie z lematem Neymana-Pearsona obszar krytyczny testu najmocniejszego ma postać

$$\{(X_1,...,X_n):\frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}\sum_{i=1}^{n}(X_i-1)^2}>k\frac{1}{(2\pi)^{\frac{n}{2}}}e^{-\frac{1}{2}\sum_{i=1}^{n}X_i^2}\}=\{(X_1,...,X_n):\ T(X)=\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i>k'\}$$ przy

czym $P_0(X>k')=\alpha$. Statystyką testową jest więc w tym przypadku $T(X)=\overline{X}$. Wiadomo, że w przypadku prawdziwości H_0 statystyka $\overline{X}$ ma rozkład $N(0,\frac{1}{3})$, $(\sigma=\frac{1}{3})$. Stąd $P_0(X>k')=\alpha)\Leftrightarrow$ $k'=F^{-1}_{N(0,\frac{1}{3})}(1-\alpha)$, w konsekwencji $k'=\text{vNormal}(0{,}95;0;\frac{1}{3})=0{,}55$.

W przypadku prawdziwości hipotezy H_1 statystyka $\overline{X}$ ma rozkład $N(1,\frac{1}{3})$. Wobec tego możemy dla zadanego poziomu istotności wyznaczyć moc uzyskanego testu:

$$\text{moc} = P_1(\overline{X} > k') = 1 - F_{N(1,\frac{1}{3})}(k') = 0{,}91 .$$

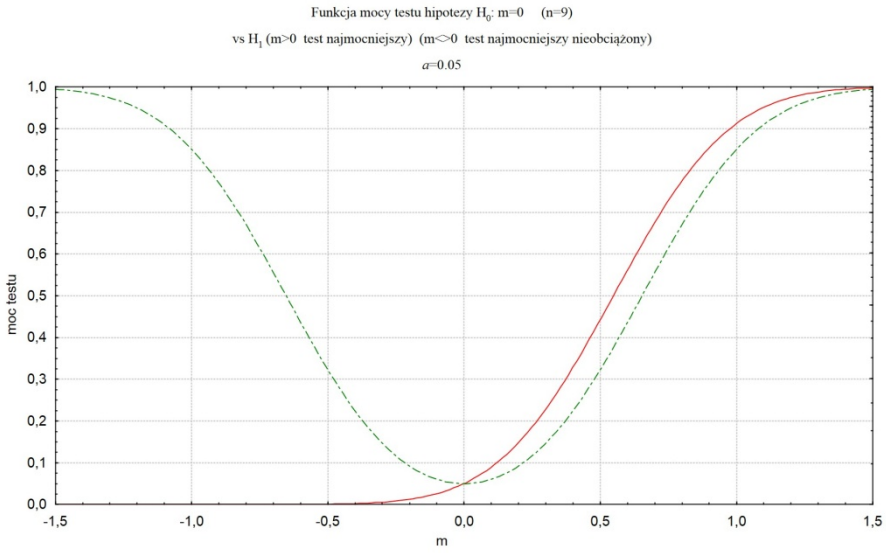
Ustalając poziom istotności α=0,05 możemy wyznaczyć funkcję mocy najmocniejszego testu na poziomie α=0,05 hipotezy H_0: m=0 vs. H_1: m>0 dla różnych alternatyw m: $\text{moc}(m) = 1 - F_{N(m,\frac{1}{3})}(F^{-1}_{N(0,\frac{1}{3})}(1-\alpha))$.

W języku programu STATISTICA

y1→ 1-iNormal(vNormal(0,95;0;1/3);x;1/3)

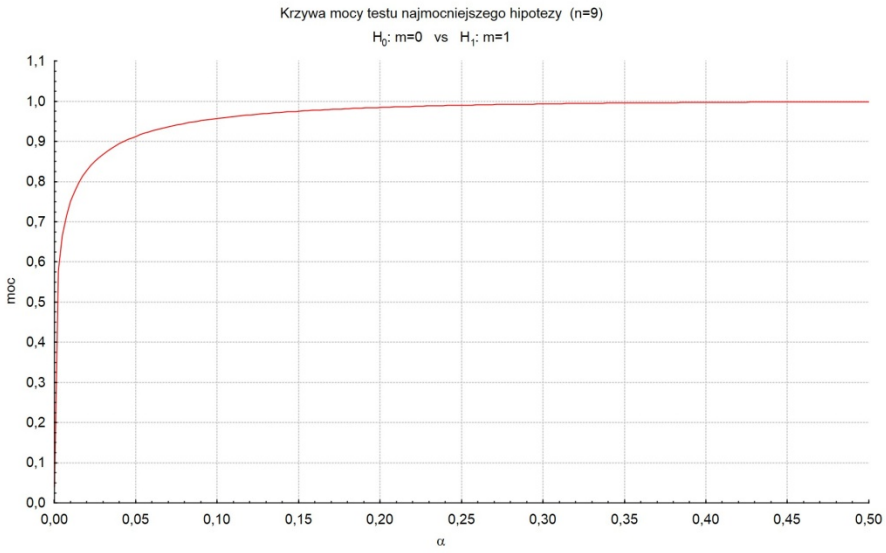
Możemy również wyznaczyć funkcje mocy najmocniejszego testu nieobciążonego hipotezy H_0: m=0 vs. H_1: m≠0

y2→1-iNormal(vNormal(0,975;0;1/3);x;1/3)+iNormal(vNormal(0,025;0;1/3);x;1/3)



Powtarzając obliczenia dla dowolnego $\alpha \in (0,1)$ możemy wyznaczyć **krzywą mocy** najmocniejszego testu hipotezy H_0: m=0 vs. H_1: m=1 w zależności od poziomu istotności moc(α).

W języku programu STATISTICA $\text{moc}(\alpha) = 1 - \text{iNormal}(\text{vNormal}(1-\alpha;0;1/3);1;1/3))$



Def. Test φ złożonej hipotezy H_0: $\theta \in \Theta_0$ przeciwko złożonej alternatywie H_1: $\theta \in \Theta_1$ nazywamy testem jednostajnie najmocniejszym na poziomie α, gdy dla każdego testu ψ H_0 przeciwko H_1 na poziomie α prawdziwy jest warunek: $\forall \theta \in \Theta_1 \ \beta_\varphi(\theta) \geq \beta_\psi(\theta)$.

W pewnych (nielicznych) przypadkach można z lematu Neymana Pearsona skonstruować test jednostajnie najmocniejszy złożonej hipotezy przeciwko złożonej alternatywie.

Testy nieobciążone

W sytuacji, gdy nie istnieje test jednostajnie najmocniejszy na poziomie α możemy próbować ograniczyć klasę testów, podobnie jak ograniczyliśmy klasę estymatorów do estymatorów nieobciążonych i być może w tej ograniczonej klasie znajdziemy test jednostajnie najmocniejszy.

Def. Test φ nazywamy testem nieobciążonym na poziomie istotności α, gdy

- $\sup_{\theta \in \Theta_0} \beta_\varphi(\theta) \leq \alpha$, czyli jest on testem na poziomie α
- $\forall \theta \in \Theta_1 \qquad \beta_\varphi(\theta) \geq \alpha$

czyli prawdopodobieństwo odrzucenia H_0 gdy jest ona **fałszywa** jest co najmniej takie jak prawdopodobieństwo jej odrzucenia, gdy jest ona **prawdziwa**. Inaczej moc testu ma być co najmniej tak duża jak rozmiar testu.

Def. Ciąg testów $\{\varphi_n\}$ dla problemu (H_0,H_1) nazywamy zgodnym, gdy dla każdego poziomu istotności α $\forall \theta \in \Theta_0 \ \forall n \ \beta_{\varphi_n}(\theta) \leq \alpha$ i $\forall \theta \in \Theta_1 \ \lim_{n \to \infty} \beta_{\varphi_n}(\theta) = 1$.

Test zgodny z prawdopodobieństwem 1 odrzuca hipotezę fałszywą, gdy liczność próby rośnie do nieskończoności.

Uwagi

W kontekście testowania istotności statystycznej spotykamy się z dwiema sytuacjami, odrzucająco-potwierdzającą (OP ang. RS Reject Support) i przyjmująco-potwierdzającą (PP ang. AS Accept Support). Przy testowaniu typu OP, hipoteza zerowa jest przeciwieństwem tego, co badacz chciałby wykazać. Np. lekarz testujący nową terapię oczekuje, że da ona wynik pozytywny wynik.

Przy testowaniu OP, błąd I-rodzaju oznacza błędne potwierdzenie tezy badacza. Z punktu widzenia reszty społeczeństwa, błąd ten jest szczególnie niekorzystny. Spowodować on może podjęcie inwestycji i wysiłków, a co najmniej dalszych badań, które nie mają szans powodzenia.Natomiast błąd II-rodzaju (przy teście OP) jest bardzo niekorzystny dla badacza, gdyż jego słuszne przewidywania nie zostały potwierdzone jedynie na skutek losowego błędu. Lekarz, który wymyślił nową terapię wpada w frustrację, gdyż był głęboko przekonany o swojej racji, a wynik testu statystycznego jednak jej nie potwierdził. Badania nad tą terapią nie będą kontynuowane i nikt się już nie dowie, jaka była prawda, a medycyna straciła cenny pomysł.

Testowanie typu OP (odrzucająco-potwierdzające) występuje najczęściej i dlatego związane z nim konwencje zdominowały popularne widzenie zagadnienia testów statystycznych. Przy testowaniu PP (przyjmująco-potwierdzającym) inaczej patrzy się na możliwość błędów obu rodzajów. Hipoteza H_0 wyraża tu pogląd (nadzieję) badacza. W takiej sytuacji błąd I-rodzaju jest fałszywym zaprzeczeniem przewidywań badacza, a błąd II-rodzaju fałszywym ich potwierdzeniem. Tak więc korzystne dla teorii przedstawianej przez badacza byłoby tu coś zupełnie nierealnego przy testowaniu OP, mianowicie przyjęcie bardzo niskiego α, na przykład 0,001.

W obu sytuacjach, OP i PP, łatwo podać przykłady trudnych badań i nierealistycznego testowania ich wyników. Rozpatrzmy najpierw przypadek OP. Czasami nie da się powiększyć próbki. Psycholog kliniczny może spędzić kilka dni na badaniu tylko jednego pacjenta; po roku będzie miał on N=50 obserwacji. A testy korelacji (częste w tego typu badaniach) mają małą moc przy małych licznościach próbek. W takiej sytuacji konieczne (i sensowne) może być wyjście ponad wartość α =0,05, jeżeli otrzyma się przy tym rozsądną moc testu. Z drugiej strony, możliwe są sytuacje za dużej mocy testu. Na przykład w testowaniu zgodności dwóch średnich (m_1=m_2) przy milionie pomiarów w każdej grupie (w przemyśle, przy automatycznych pomiarach N=1000000 nie jest od rzeczy). W takiej sytuacji, hipoteza zerowa prawie zawsze będzie odrzucona, przy najmniejszych, w jakimś sensie przypadkowych różnicach.

Sytuacja taka jest tym bardziej sztuczna przy testach PP. Jeżeli N jest bardzo duże, to badacz prawie zawsze zmuszony będzie odrzucić swoją teorię, nawet jeżeli "tak naprawdę" jest to dobra i wartościowa hipoteza i "całkiem" dobrze pasuje do danych. Zbyt duża dokładność doświadczenia działałaby tu na niekorzyść eksperymentatora.

Podsumowując testy OP (odrzucająco-potwierdzające):

1. Badacz cieszyłby się, gdyby miał podstawy do odrzucenia hipotezy H_0.
2. Społeczność, dla której badacz pracuje chce mieć kontrolę nad błędami I rodzaju.
3. W interesie badacza leży niska wartość błędu II rodzaju.
4. Duża liczność próby jest korzystna dla badacza.
5. Przy "za dużej" mocy testu nieważne efekty stają się "bardzo istotne".

Testy PP (przyjmująco-potwierdzające):

1. Badacz cieszyłby się, gdyby nie było podstaw do odrzucenia hipotezy H_0 i mógłby ją zaakceptować.
2. Społeczność, dla której badacz pracuje powinna mieć kontrolę nad błędami II rodzaju, choć nie zawsze sytuacja ta jest właściwie rozpoznawana i pozostaje się przy konwencjach stosowalnych przy testowaniu OP.
3. Badacz pilnie zwraca uwagę na błędy I rodzaju.
4. Duża liczność próby jest niekorzystna dla badacza.
5. Przy "za dużej" mocy testu teoria badacza zostanie raczej odrzucona w teście nawet jeśli "całkiem dobrze" pasuje do danych.

Testy oparte na ilorazie wiarygodności

Załóżmy ponadto, że rodzina rozkładów $\mathcal{P}=\{P_\theta : \theta\in\Theta\}$ jest dominowana przez pewną σ-skończoną miarę μ (zwykle miarę liczącą lub Lebesgue'a) i oznaczmy gęstość rozkładu P_θ względem miary μ przez $p_\theta = \frac{dP_\theta}{d\mu}$. Rodzinę $\mathcal{P}=\{P_\theta: \theta\in\Theta\}$ można więc utożsamić z rodziną gęstości $\mathcal{P}=\{p_\theta : \theta\in\Theta\}$.

Niech $L(\theta;X)$ będzie funkcją wiarygodności w przestrzeni $(\mathcal{X}, \mathcal{B}, \mathcal{P}=\{P_\theta : \theta\in\Theta\})$

Def. Wiarygodnością hipotezy $H_i: \theta\in\Theta_i$ gdy zaobserwowano $X\in\mathcal{X}$ nazywamy liczbę

$$L_{H_i}(X) = \sup_{\theta\in\Theta_i} L(\theta, X).$$

Uwaga. Wiarygodność hipotezy H_i możemy interpretować w przypadku dyskretnym jako największe prawdopodobieństwo, przy prawdziwości H_i , uzyskania zaobserwowanego $X\in\mathcal{X}$. Podobną interpretację (z oczywistymi modyfikacjami związanymi z interpretacją gęstości) mamy w przypadku ciągłym.

Jeżeli weryfikujemy hipotezę $H_0: \theta\in\Theta_0$ wobec alternatywy $H_1: \theta\in\Theta_1$, to należy zakwestionować hipotezę H_0, gdy bardziej wiarygodna jest hipoteza H_1 , tzn. gdy

$$\frac{L_{H_1}(X)}{L_{H_0}(X)} > \lambda_0 \geq 1,$$

gdzie λ_0 jest tak dobraną stałą, aby rozmiar testu nie przekraczał poziomu istotności α, czyli

$$\sup_{\theta\in\Theta_0} P_\theta\left(X : \frac{L_{H_1}(X)}{L_{H_0}(X)} > \lambda_0 \right) \leq \alpha .$$

Niech $\dot\theta$ będzie taką wartością parametru $\theta\in\Theta_0$ (o ile istnieje), że $L(\dot\theta, X) = L_{H_0}(X) = \sup_{\theta\in\Theta_0} L(\theta, X)$,

czyli zakładamy , że w punkcie $\dot\theta$ realizuje się wiarygodność hipotezy zerowej (czyli jest osiągany odpowiedni kres) i niech $\hat\theta$ będzie estymatorem największej wiarygodności parametru θ czyli

$$L(\hat\theta, X) = \sup_{\theta\in\Theta} L(\theta, X).$$

Estymator $\hat\theta$ nazywamy bezwarunkowym ENW a $\dot\theta$ nazywamy warunkowym (przy prawdziwości H_0) ENW parametru θ.

Widać , że dla $\lambda_0\geq 1$ mamy następującą równoważność

$$\frac{L_{H_1}(X)}{L_{H_0}(X)} > \lambda_0 \Leftrightarrow \frac{L(\hat\theta, X)}{L(\dot\theta, X)} > \lambda_0$$

bo $L_{H_1}(X) > L_{H_0}(X) \Leftrightarrow \sup_{\theta\in\Theta} L(\theta, X) = \sup_{\theta\in\Theta_1} L(\theta, X)$

Def. Wielkość $\lambda(X)=\dfrac{L(\hat{\theta},X)}{L(\dot{\theta},X)}=\dfrac{\sup\limits_{\theta\in\Theta}L(\theta,X)}{\sup\limits_{\theta\in\Theta_0}L(\theta,X)}$ nazywamy **ilorazem wiarygodności** a test H_0

przeciwko H_1 o obszarze krytycznym (obszarze odrzucenia H_0) postaci

$$C=\{X:\ \lambda(X)>\lambda_0\}$$

gdzie $\lambda_0\geq 1$ jest odpowiednio dobraną stałą nazywamy **testem opartym na ilorazie wiarygodności**.

Uwagi

- Testy oparte na ilorazie wiarygodności są szczególnie przydatne w przypadkach, gdy odpowiednie testy jednostajnie najmocniejsze nie istnieją (zwykle tak jest) lub nie potrafimy ich wyznaczyć a także wówczas, gdy obserwacje nie tworzą próby prostej (np. są zależne-tak jest w teorii szeregów czasowych).Czasami są jednak bezsensowne (przykład-Zielinski)
- W przypadku testowania hipotezy prostej przeciwko prostej alternatywie test ilorazu wiarygodności jest równoważny z testem najmocniejszym Neymana-Pearsona

Przykład 1. Test *t*-Studenta jako test oparty na ilorazie wiarygodności.

Niech $X=(X_1,...,X_n)$ będzie próbą prostą z rozkładu $N(m,\sigma^2)$ z nieznanymi parametrami m iσ^2.

Testujemy hipotezę

H_0: $m=m_0$ wobec alternatywy H_1: $m\neq m_0$.

Funkcja wiarygodności ma postać

$$L(m,\sigma;X)=(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-m)^2}.$$

Stąd $(\hat{m},\hat{\sigma})=(\overline{X},\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2})=ENW[(m,\sigma)]$. Wobec tego

$$L(\hat{m},\hat{\sigma};X)=(2\pi\hat{\sigma}^2)^{-\frac{n}{2}}e^{-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(X_i-\overline{X})^2}=(2\pi\hat{\sigma}^2)^{-\frac{n}{2}}e^{-\frac{n}{2}}$$

Podobnie przy prawdziwości hipotezy H_0 funkcja wiarygodności ma postać

$L(\sigma;X)=(2\pi\sigma^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-m_0)^2}$. Skąd łatwo widać, że $\dot{\sigma}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i-m_0)^2})=ENW[\sigma]$ i

$$L(\dot{\sigma};X)=(2\pi\dot{\sigma}^2)^{-\frac{n}{2}}e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(X_i-m_0)^2}=(2\pi\dot{\sigma}^2)^{-\frac{n}{2}}e^{-\frac{n}{2}}.$$

Wobec tego

$$C=\{X:\lambda(X)>k_1\}=\{X:\frac{\sum_{i=1}^{n}(X_i-m_0)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}>k_2\}=\{X:\frac{\sum_{i=1}^{n}((X_i-\overline{X})+(\overline{X}-m_0))^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}>k_2\}=$$

$$\{X:\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2+n(\overline{X}-m_0)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}>k_2\}=\{X:1+\frac{n(\overline{X}-m_0)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}>k_2\}=\{X:\frac{n(\overline{X}-m_0)^2}{\sum_{i=1}^{n}(X_i-\overline{X})^2}>k_3\}=$$

$$\{X:|t(X)|=\frac{\sqrt{n}\,|\overline{X}-m_0|}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}}>k\}$$

Wiadomo, **że przy prawdziwości $\mathbf{H_0}$** statystyka testowa $t(X)=\frac{\sqrt{n}(\overline{X}-m_0)}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}}$ ma rozkład t_{n-1}

Studenta o n-1 stopniach swobody, gdyż zmienne losowe $\frac{\sqrt{n}(\overline{X}-m_0)}{\sigma}$ i $\frac{\sum_{i=1}^{n}(X_i-\overline{X})^2}{\sigma^2}$ mają

odpowiednio rozkłady $N(0,1)$ i χ^2_{n-1} oraz są one niezależne (tw Fishera -wniosek z tw. Basu).

Wobec tego test oparty na ilorazie wiarygodności hipotezy H_0: $m=m_0$ wobec alternatywy H_1: $m\neq m_0$

ma obszar krytyczny postaci $C=\{X:|t(X)|=\frac{\sqrt{n}\,|\overline{X}-m_0|}{\sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2}}>t(1-\frac{\alpha}{2},n-1)\}$ gdzie $t(p,n)$ jest

kwantylem rzędu p rozkładu t-Studenta o n stopniach swobody.

Przykład 2 testowanie równości średnich w rozkładach normalnych.

Niech $X=(X_1,...,X_m)$ i $Y=(Y_1,...,Y_n)$ będą niezależnymi próbami prostymi z rozkładów $N(m_x,\sigma^2)$ i $N(m_y,\sigma^2)$ z nieznanymi parametrami m_x, m_y i σ^2. Testujemy hipotezę

H_0: $m_x=m_y=\mu$ wobec alternatywy H_1: $m_x\neq m_y$

Funkcja wiarygodności ma postać

$$L(m_x,m_y,\sigma;X,Y)=(2\pi\sigma^2)^{-\frac{m+n}{2}}e^{-\frac{1}{2\sigma^2}\{\sum_{i=1}^{m}(X_i-m_x)^2+\sum_{j=1}^{n}(Y_j-m_y)^2\}}.$$

a jej logarytm

$$l(m_x,m_y,\sigma;X,Y)=-\frac{m+n}{2}\ln(2\pi)-(m+n)\ln\sigma-\frac{1}{2\sigma^2}\{\sum_{i=1}^{m}(X_i-m_x)^2+\sum_{j=1}^{n}(Y_j-m_y)^2\}$$

Łatwo widać, że

$$(\hat{m}_x, \hat{m}_y, \hat{\sigma}) = (\overline{X}, \overline{Y}, \sqrt{\tfrac{1}{m+n}\{\sum_{i=1}^{m}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2\}}) = ENW[(m_x, m_y, \sigma)].$$

Stąd $L(\hat{m}_x, \hat{m}_y, \hat{\sigma}; X, Y) = (2\pi\hat{\sigma}^2)^{-\frac{m+n}{2}} e^{-\frac{m+n}{2}}$.

W przypadku prawdziwości H_0: $m_x = m_y = \mu$ funkcja wiarygodności przybiera postać

$$L(\mu, \sigma; X, Y) = (2\pi\sigma^2)^{-\frac{m+n}{2}} e^{-\frac{1}{2\sigma^2}\{\sum_{i=1}^{m}(X_i-\mu)^2 + \sum_{j=1}^{n}(Y_j-\mu)^2\}}.$$

Stąd

$$l(\mu, \sigma; X, Y) = -\tfrac{m+n}{2}\ln(2\pi) - (m+n)\ln\sigma - \tfrac{1}{2\sigma^2}\{\sum_{i=1}^{m}(X_i - \mu)^2 + \sum_{j=1}^{n}(Y_j - \mu)^2\}$$

i w konsekwencji

$$(\dot{\mu}, \dot{\sigma}) = (\tfrac{m}{m+n}\overline{X} + \tfrac{n}{m+n}\overline{Y}, \sqrt{\tfrac{1}{m+n}\{\sum_{i=1}^{m}(X_i - \dot{\mu})^2 + \sum_{j=1}^{n}(Y_j - \dot{\mu})^2\}}) = ENW[(\mu, \sigma)].$$

Wobec tego

$$L(\dot{\mu}, \dot{\sigma}; X, Y) = (2\pi\dot{\sigma}^2)^{-\frac{m+n}{2}} e^{-\frac{m+n}{2}}$$

Test oparty na ilorazie wiarygodności H_0 przeciwko H_1 ma obszar krytyczny postaci

$$C = \{(X,Y): \lambda(X,Y) > k_1\} = \{(X,Y): \frac{\sum_{i=1}^{m}(X_i - \dot{\mu})^2 + \sum_{j=1}^{n}(Y_j - \dot{\mu})^2}{\sum_{i=1}^{m}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2} > k_2\} =$$

$$\{(X,Y): \frac{\sum_{i=1}^{m}((X_i - \overline{X}) + (\overline{X} - \dot{\mu}))^2 + \sum_{j=1}^{n}((Y_j - \overline{Y}) + (\overline{Y} - \dot{\mu}))^2}{\sum_{i=1}^{m}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_i - \overline{Y})^2} > k_2\} =$$

$$\{(X,Y): 1 + \frac{m(\overline{X} - \dot{\mu})^2 + n(\overline{Y} - \dot{\mu})^2}{\sum_{i=1}^{m}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2} > k_2\} = \{(X,Y): \frac{\frac{mn}{m+n}(\overline{X} - \overline{Y})^2}{\sum_{i=1}^{m}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2} > k_3\} =$$

$$\{(X,Y): |t(X,Y)| = \frac{\sqrt{\frac{mn(n+m-2)}{m+n}}\,|\overline{X} - \overline{Y}|}{\sqrt{\sum_{i=1}^{m}(X_i - \overline{X})^2 + \sum_{j=1}^{n}(Y_j - \overline{Y})^2}} > k\}$$

Wiadomo, że **przy prawdziwości hipotezy H_0** statystyka $t(X,Y)$ ma rozkład t-Studenta o $n+m-2$ stopniach swobody t_{n+m-2}, gdyż:

$\overline{X} \sim N(\mu, \frac{\sigma^2}{m})$, $\overline{Y} \sim N(\mu, \frac{\sigma^2}{n})$, $\frac{1}{\sigma^2}\sum_{i=1}^{m}(X_i - \overline{X})^2 \sim \chi^2_{m-1}$, $\frac{1}{\sigma^2}\sum_{i=1}^{m}(X_i - \overline{X})^2 \sim \chi^2_{n-1}$ i wszystkie zmienne są niezależne. Wobec tego $\overline{X} - \overline{Y} \sim N(0, (\frac{1}{m} + \frac{1}{n})\sigma^2) = N(0, \frac{m+n}{mn}\sigma^2)$

$$\frac{1}{\sigma^2}\sum_{i=1}^{m}(X_i - \overline{X})^2 + \frac{1}{\sigma^2}\sum_{i=1}^{m}(X_i - \overline{X})^2 \sim \chi^2_{m+n-2}$$ (tw. o dodawaniu)

i obie zmienne są niezależne.

Stąd z definicji rozkładu *t*-Studenta $t(X,Y)=\dfrac{\sqrt{\frac{mn(n+m-2)}{m+n}}(\overline{X}-\overline{Y})}{\sqrt{\sum_{i=1}^{m}(X_i-\overline{X})^2+\sum_{j=1}^{n}(Y_j-\overline{Y})^2}} \sim t_{m+n-2}$.

Graniczny rozkład ilorazu wiarygodności

Tw. Przy pewnych warunkach regularności (regularny model Cramera-Rao) statystyka $2\ln\lambda(X)$ ma dla każdego $\theta\in\Theta_0$ (czyli przy prawdziwości H_0) graniczny rozkład χ^2_r o *r* stopniach swobody, gdzie *r* jest liczbą ograniczeń nałożonych na θ potrzebnych do wyspecyfikowania H_0.

Uwaga. W przykładach 1 i 2 statystyka $2\ln\lambda(X)$ ma graniczny rozkład χ^2_1

Uwaga. Aproksymacja $\sqrt{2\chi^2_n}-\sqrt{2n-1} \sim N(0,1)$ (dla dostatecznie dużych *n*)

Przegląd wybranych testów

Testy dotyczące wartości oczekiwanej w rozkładzie normalnym i problem testowania równości średnich w dwóch zależnych populacjach o rozkładzie normalnym.

Model 1. Niech $X=(X_1,...,X_n)$ będzie próbą prostą z rozkładu $N(m,\sigma^2)$ przy czym σ^2 jest **znane**. Testujemy hipotezę

1. H_{0a}: $m\geq m_0$ wobec alternatywy H_{1a}: $m<m_0$
2. H_{0b}: $m\leq m_0$ wobec alternatywy H_{1b}: $m>m_0$
3. H_{0c}: $m=m_0$ wobec alternatywy H_{1c}: $m\neq m_0$

Statystyką testową jest

$$T(X)=\sqrt{n}\,\frac{\overline{X}-m_0}{\sigma},$$

która przy ustalonym *m* ma rozkład $N(\sqrt{n}\,\frac{m-m_0}{\sigma},1)$. Zbiór krytyczny (odrzucenia H_0) *C* na poziomie α konstruujemy następująco:

1. $C=\{X: T(X)\leq u_\alpha\}$ dla alternatywy H_{1a}: $m<m_0$
2. $C=\{X: T(X)\geq u_{1-\alpha}\}$ dla alternatywy H_{1b}: $m>m_0$
3. $C=\{X: |T(X)|\geq u_{1-\frac{\alpha}{2}}\}$ dla alternatywy H_{1c}: $m\neq m_0$

gdzie u_α jest kwantylem rzędu α rozkładu $N(0,1)$

Uwaga. W przypadku 1 i 2 test jest jednostajnie najmocniejszy. W przypadku 3 test jest jednostajnie najmocniejszy w klasie testów nieobciążonych. Test jednostajnie najmocniejszy w tym przypadku nie istnieje.

Model 2. Niech $X=(X_1,...,X_n)$ będzie próbą prostą z rozkładu $N(m,\sigma^2)$ przy czym σ^2 jest **nieznane**. Testujemy hipotezę

1. H_{0a}: $m \geq m_0$ wobec alternatywy H_{1a}: $m < m_0$
2. H_{0b}: $m \leq m_0$ wobec alternatywy H_{1b}: $m > m_0$
3. H_{0c}: $m = m_0$ wobec alternatywy H_{1c}: $m \neq m_0$

Statystyką testową jest

$$T(X) = \sqrt{n}\,\frac{\overline{X} - m_0}{S_n}, \text{ gdzie } \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \;, \quad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2, \quad ,$$

która przy prawdziwości H_0 ma niecentralny rozkład t-Studenta o n-1 stopniach swobody i parametrze niecentralności $\lambda = \sqrt{n}\,\frac{m-m_0}{\sigma}$ (czyli $t_{n-1,\lambda}$). Zbiór krytyczny (odrzucenia H_0) C na poziomie α konstruujemy następująco:

1. $C = \{X : T(X) \leq u_\alpha\}$ dla alternatywy H_{1a}: $m < m_0$
2. $C = \{X : T(X) \geq u_{1-\alpha}\}$ dla alternatywy H_{1b}: $m > m_0$
3. $C = \{X : |T(X)| \geq u_{1-\frac{\alpha}{2}}\}$ dla alternatywy H_{1c}: $m \neq m_0$

gdzie u_α jest kwantylem rzędu α rozkładu centralnego t Studenta $t_{n\text{-}1}$.

Uwaga. W przypadku 1 i 2 test jest jednostajnie najmocniejszy. W przypadku 3 test jest jednostajnie najmocniejszy w klasie testów nieobciążonych. Test jednostajnie najmocniejszy w tym przypadku nie istnieje. Dla n >30 rozkład t-Studenta aproksymujemy rozkładem normalnym N(0,1).

Powyższe testy mogą być użyte do **porównywania wartości oczekiwanych w dwóch próbach zależnych o rozkładzie normalnym.**

Niech $(X_1,Y_1),...,(X_n,Y_n)$ będzie próbą prostą z dwuwymiarowego rozkładu normalnego $N\left(\begin{bmatrix} m_x \\ m_y \end{bmatrix}, \begin{bmatrix} V_x & C_{xy} \\ C_{yx} & V_y \end{bmatrix}\right)$. Chcemy testować hipotezę

H_0: $m_x = m_y$ przeciwko alternatywie H_0: $m_x \neq m_y$

Z powyższym problemem mamy do czynienia, gdy dla tego samego pacjenta rejestrujemy dwa pomiary pewnej wielkości przed i po zażyciu leku.

Definiując zmienną $Z=Y-X$, którą możemy interpretować jako poprawę spowodowaną zażyciem leku dostajemy próbę prostą $(Z_1,...,Z_n)$ z rozkładu $N(m_z,\sigma^2)$, gdzie

$$m_z = m_y - m_x \text{ i } \sigma^2 = \begin{bmatrix} -1 & 1 \end{bmatrix} \begin{bmatrix} V_x & C_{xy} \\ C_{yx} & V_y \end{bmatrix} \begin{bmatrix} -1 \\ 1 \end{bmatrix}$$

i problem sprowadza się do testowania hipotezy H_0: $m_z=0$ wobec alternatywy H_1: $m_z\neq 0$ (lub $m_z>0$

Testowanie równości średnich w dwóch niezależnych populacjach o rozkładzie normalnym.

Niech $X=(X_1,...,X_n)$ i $Y=(Y_1,...,Y_m)$ będą niezależnymi próbami prostymi z rozkładów $N(m_x,\sigma^2)$ i $N(m_y,\sigma^2)$ odpowiednio. Nieznana wariancja σ^2 jest **taka sama** w obu rozkładach.

Testujemy hipotezę

H_0: $m_x=m_y$ wobec jednej z alternatyw H_{1a}: $m_x<m_y$, H_{1b}: $m_x>m_y$, H_{1c}: $m_x\neq m_y$

Statystyką testową jest

$$T(X,Y) = \sqrt{\frac{nm(n+m-2)}{n+m}} \frac{\overline{X}-\overline{Y}}{\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2+\sum_{j=1}^{m}(Y_j-\overline{Y})^2}},$$

która przy prawdziwości H_0 ma rozkład t-Studenta o $n+m-2$ stopniach swobody (czyli t_{n+m-2}). Zbiór krytyczny (odrzucenia H_0) C na poziomie α konstruujemy następująco:

1. $C=\{X: T(X)\le u_\alpha\}$ dla alternatywy H_{1a}: $m_x<m_y$
2. $C=\{X: T(X)\ge u_{1-\alpha}\}$ dla alternatywy H_{1b}: $m_x>m_y$
3. $C=\{X: |T(X)|\ge u_{1-\frac{\alpha}{2}}\}$ dla alternatywy H_{1c}: $m_x\neq m_y$

gdzie u_α jest kwantylem rzędu α rozkładu t_{n+m-2} .

Uwaga. W przypadku 1 i 2 test jest jednostajnie najmocniejszy. W przypadku 3 test jest jednostajnie najmocniejszy w klasie testów nieobciążonych. Test jednostajnie najmocniejszy w tym przypadku nie istnieje. Dla $n>30$ rozkład t-Studenta aproksymujemy rozkładem normalnym N(0,1).

Nieparametryczne odpowiedniki powyższych modeli testowania hipotez -testy Wilcoxona i Manna-Whitneya.

Rozważmy jeszcze raz problem porównywania dwóch prób zależnych. Niech (X_1,Y_1), ..., (X_n,Y_n) będzie próbą prostą z pewnego dwuwymiarowego rozkładu ciągłego. Sytuacja taka odpowiada np. pomiarowi pewnej zmiennej dla tych samych jednostek eksperymentalnych przed i po zastosowaniu terapii. Definiując zmienną $Z=Y-X$, którą możemy interpretować jako poprawę spowodowaną terapią dostajemy próbę prostą $(Z_1,...,Z_n)$ z pewnego rozkładu ciągłego. Jeśli terapia jest nieskuteczna, czyli

zmienne X i Y mają taki sam rozkład, to zmienna Z ma rozkład symetryczny wokół 0. Oznacza to że zmienne Z (poprawa) i $-Z$ (pogorszenie) mają taki sam rozkład. Oznaczając przez $F_Z(t)$ dystrybuantę zmiennej Z widać, że $F_{-Z}(t)=P(-Z\le t)=1-P(-Z>t)=1-P(Z<-t)=1-F_Z(-t)$. Warunek symetryczności (wokół 0) rozkładu zmiennej Z przybiera postać $F_Z(t)+F_Z(-t)=1$ dla każdego $t\in R$. Jeżeli skutkiem terapii jest przesunięcie rozkładu, to zmienna Z ma dystrybuantę $F(t-\theta)$ gdzie F jest nieznaną dystrybuanta rozkładu symetrycznego względem 0 a θ nieznanym parametrem przesunięcia. Testowanie skuteczności terapii sprowadza się do testowania hipotezy o parametrze przesunięcia θ.

Niech $Z_1, ...,Z_n$ będzie próbą prostą z pewnego rozkładu $F(z-\theta)$ gdzie F jest ciągłą dystrybuantą rozkładu symetrycznego. Jest o oczywiście nieparametryczna rodzina rozkładów. Parametrem jest para $(F, \theta)\in\mathcal{F}_{sym}\times R$ ($\mathcal{F}_{sym}$ jest zbiorem symetrycznych absolutnie ciągłych dystrybuant na R.

Testujemy hipotezę

H_0: $\theta=0$ (terapia jest nieskuteczna)

wobec jednej z alternatyw

H_{1a}: $\theta<0$ albo H_{1b}: $\theta>0$ albo H_{1c}: $\theta\neq 0$

F jest w tym przypadku parametrem zakłócającym. Problem testowania jest niezmienniczy względem grupy wszystkich transformacji $z_i^{'}=f(z_i)$, $i=1,...,n$ takich, że f jest **ciągła**, **nieparzysta** i **ściśle rosnąca**. Transformacje powyższe zachowują znaki obserwacji i porządek bezwzględnych wartości obserwacji.

Oznaczmy przez R_i=ranga $|Z_i|$ wśród $|Z_1|,...,|Z_n|$

Można pokazać (Lehmann) , że maksymalnym niezmiennikiem jest zbiór rang $R_1,...,R_n$ .Redukcja przez statystyki dostateczne zastosowana do maksymalnego niezmiennika prowadzi do rang $R_1^+,...,R_k^+$, odpowiadających dodatnim obserwacjom $Z_1, ...,Z_n$ (których jest k).

Statystyka oparta na tym niezmienniku ma rozkład niezależny od dystrybuanty $F\in\mathcal{F}_{sym}$. Statystyką testową jest statystyka **Wilcoxona**

$$W=\sum_{i=1}^{k} R_i^+$$

Przy prawdziwości H_0 $E(W)=\frac{1}{2}\sum_{i=1}^{n} R_i=\frac{n(n+1)}{4}$ $V(W)=\frac{n(n+1)(2n+1)}{24}$.

W zależności od hipotezy alternatywnej obszar krytyczny konstruujemy lewostronny prawostronny, obustronny. Rozkład statystyki W jest stablicowany (Zieliński R, Zieliński W., Tablice statystyczne)

Dla $n>16$ stosujemy aproksymację gaussowską

Statystyka $\frac{W-\frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$ ma dla $n>16$ w przybliżeniu rozkład $N(0,1)$.

Test Manna-Whitneya-Wilcoxona jest nieparametrycznym odpowiednikiem testu *t* Studenta dla grup niezależnych

Niech $X_1,...,X_m$ oraz $Y_1,...,Y_n$ będą dwiema niezależnymi próbami prostymi z rozkładów o dystrybuantach $F(x- m_x)$ i $F(y- m_y)$ gdzie F jest pewną nieznaną dystrybuantą ze zbioru $\mathcal{F}_{acg}$ absolutnie ciągłych dystrybuant na R. Mamy tu do czynienia z rodziną rozkładów parametryzowaną parametrem (F,m_x,m_y) czyli z nieparametryczna rodziną rozkładów. Testujemy hipotezę

H_0: $m_x=m_y$

wobec jednej z alternatyw

H_{1a}: $m_x<m_y$ albo H_{1b}: $m_x>m_y$ albo H_{1c}: $m_x\neq m_y$

Powyższy problem jest niezmienniczy względem grupy transformacji $x_i^{'} = f(x_i)$, $y_j^{'} = f(y_j)$ ((i=1,...,m , j=1,...,n) gdzie f jest ciągłą i ściśle rosnącą bijekcją zbioru R na siebie. Niech $R_1,...,R_m$ oraz $S_1,...,S$ będą rangami (kolejnymi numerami) odpowiednio obserwacji $X_1,...,X_m$ oraz $Y_1,...,Y_m$ w tej połączonej próbie. Maksymalnym niezmiennikiem jest zbiór rang $R_1,...,R_m$, $S_1,...,S_n$. Redukcja przed statystyki dostateczne maksymalnego niezmiennika prowadzi do rang $S_1,...,S_n$ (czyli do rang Y-ków)

Statystyką testową jest statystyka Manna-Whitneya-Wilcoxona (MWW), która dla testowania H_0: $m_x=m_y$ przeciwko H_{1a}: $m_x<m_y$ albo H_{1b}: $m_x>m_y$ ma postać

$$W = \sum_{j=1}^{n} S_j$$ ((czyli suma rang Y-ków)

Duże wartości W świadczą przeciwko H_{1a}: $m_x<m_y$ a małe przeciwko H_{1b}: $m_x>m_y$. Z kolei duże wartości statystyki $|\sum_{j=1}^{n} S_j - \sum_{i=1}^{m} R_i|$ świadczą przeciwko H_{1c}: $m_x\neq m_y$.

Znane są rozkłady statystyk testowych dla małych m i n (które nie zależą od F).(zobacz ZielińskiR. Siedem wykładów...)) i aproksymacja (twierdzenie Hoeffdinga) normalna dla dużych m i n.

Statystyka $\frac{W - \frac{n(m+n+1)}{2}}{\sqrt{\frac{nm(m+n+1)}{12}}}$ ma rozkład zbieżny do rozkładu $N(0,1)$, gdy $\min(m,n)\to\infty$. Aproksymacja ta jest wystarczająco dokładna dla $\min(m,n)\geq 4$ i $m+n\geq 20$ (Plucińska)

Test zgodności Kołmogorowa

Niech $X=(X_1,...,X_n)$ będzie próba prosta z rozkładu o **ciągłej** dystrybuancie $F\in\mathcal{F}_c$. Niech $F_0\in\mathcal{F}_c$ będzie ustaloną ciągłą dystrybuantą. Testujemy hipotezę

H_0: $F=F_0$

wobec jednej z alternatyw

H_{1a}: $F<F_0$ albo H_{1b}: $F>F_0$ albo H_{1c}: $F\neq F_0$.

Oznaczmy przez F_n dystrybuantę empiryczną i rozważmy następujące statystyki Kołmogorowa:

$$D_n^+ = \sup_x (F_n(x) - F_0(x)),$$

$$D_n^- = \sup_x (F_0(x) - F_n(x)),$$

$$D_n = \sup_x |F_n(x) - F_0(x)|.$$

Niech $(X_{1:n},...,X_{n:n})$ będzie wektorem statystyk pozycyjnych (próbą uporządkowaną).

Dowodzi się, że $D_n^+ = \max_i (\frac{i}{n} - F_0(X_{i:n}))$, $D_n^- = \max_i (F_0(X_{i:n}) - \frac{i-1}{n})$, $D_n = \max\{D_n^+, D_n^-\}$. Przy prawdziwości H_0 rozkłady statystyk Kołmogorowa nie zależą od F_0 i są znane. Znane są również rozkłady graniczne wyżej wymienionych statystyk. Duże wartości statystyki D_n^- świadczą na korzyść H_{1a} (przeciwko H_0). Podobnie duże statystyki D_n^+ świadczą na korzyść H_{1b} (przeciwko H_0) a duże statystyki D_n świadczą na korzyść H_{1c} (przeciwko H_0) .

Jeżeli F_0 jest dystrybuantą rozkładu $N(m,\sigma^2)$, którego parametry m i σ^2 nie są znane lub dystrybuantą rozkładu wykładniczego $E(\lambda)$ z nieznanym parametrem λ , to dokładny rozkład statystyki D_n został wyznaczony przez Lillieforsa. Test Kolmogorowa Lillieforsa może być więc użyty do testowania hipotezy o normalności rozkładu.

Test zgodności chi- kwadrat

Rozważmy próbę z rozkładu wielomianowego

$$P(X_1 = n_1,..., X_k = n_k) = \frac{n!}{n_1!...n_k!} p_1^{n_1}...p_k^{n_k}; \qquad \sum_{j=1}^{k} p_j = 1 \ \sum_{j=1}^{k} n_j = n$$

Chcemy testować hipotezę

H_0: $(p_1,..., p_k) = (p_1^0,..., p_k^0)$ (hipoteza prosta)

przeciwko

H_1: $(p_1,..., p_k) \neq (p_1^0,..., p_k^0)$ (hipoteza złożona)

Pearson udowodnił, że statystyka

$\sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0}$ ma graniczny rozkład χ^2_{k-1}

Ponieważ statystyka Pearsona jest pewną miarą odstępstw liczności obserwowanych od oczekiwanych przy prawdziwości H_0 , "duże" wartości statystyki Pearsona świadczą przeciwko hipotezie H_0 . Wobec tego H_0 należy odrzucić na poziomie α, jeżeli $\sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0} > \chi^2_{k-1}(1-\alpha)$, gdzie $\chi^2_{k-1}(1-\alpha)$ oznacza kwantyl rzędu $1-\alpha$ rozkładu χ^2_{k-1}.

Testowanie złożonej hipotezy zgodności

H_0: $(p_1,...,p_k) = (p_1(\theta),...,p_k(\theta))$, gdzie $\theta \in R^s$; $s<k-2$ przeciwko H_1: ~H_0.

Niech $\hat{\theta}$ będzie estymatorem największej wiarygodności parametru θ. Oznaczmy $\hat{p}_i = p_i(\hat{\theta})$; $i=1,...,k$.

Wówczas statystyka $\sum_{i=1}^{k} \frac{(n_i - n\hat{p}_i)^2}{n\hat{p}_i}$ ma graniczny rozkład χ^2_{k-1-s}.

Dalsza procedura jest kopią powyższej z jedyną modyfikacją dotyczącą ilości stopni swobody granicznego rozkładu χ^2_{k-1-s}.

Test Shapiro-Wilka

Jest to powszechnie uznany uniwersalny (mocny przy szerokiej klasie alternatyw) test normalności.

Niech $X=(X_1,...,X_n)$ będzie próbą prostą z rozkładu o **ciągłej** dystrybuancie $F \in \mathcal{F}_c$.

Testujemy hipotezę

H_0: $F= F_0$, gdzie F_0 jest dystrybuantą rozkładu $N(m,\sigma^2)$, którego parametry m i σ^2 nie są znane, wobec alternatywy

H_1: $F \neq F_0$

Statystyka testowa jest postaci

$$W = \frac{\left(\sum_{i=1}^{[\frac{n}{2}]} a_{in} (X_{(n+1-i)} - X_{(i)}) \right)^2}{\sum_{i=1}^{n} (X_i - \overline{X})^2}$$

Gdzie współczynniki a_{in} są tablicowane dla $n \leq 50$. Dla $n > 50$ są dostępne programy komputerowe wyliczające te współczynniki. Test jest powszechnie dostępny w statystycznych programach komputerowych.